



COMISIÓN DE LA  
VERDAD



# Informe metodológico del proyecto conjunto JEP-CEV-HRDAG de integración de datos y estimación estadística\*

2 de agosto de 2022

## Índice

<b>1. Introducción</b>	<b>4</b>
1.1. ¿Cuál es nuestra meta?	4
1.2. Panorama de los datos	6
1.3. Flujo de trabajo	8
1.4. Resumen de este reporte	9
<b>2. Hallazgos estadísticos</b>	<b>9</b>
<b>3. Vinculación de registros</b>	<b>23</b>
3.1. Preparación de datos	23
3.2. Generación de bloques de registros	23
3.2.1. Generación de datos de entrenamiento	23
3.2.2. Generación de bloques	24
3.3. Generación de características	26
3.4. Modelo por pares	26
3.4.1. Evaluación del modelo	26
3.5. Agrupamiento	29
3.6. Unión y exportación	30
3.6.1. Reclasificación	30
3.7. Resultados	31

\*En este proyecto participaron las y los siguientes investigadores cuantitativos de cada organización. CEV: Paula Amado, William Acero, Camilo Argoty, Giovanni Babativa, Luz Karime Bernal, Alejandro Castro, María Juliana Durán, Esteban Espinosa, Valentina Gómez, Andrea González y María Alejandra Ortíz. HRDAG: Patrick Ball, Maria Gargiulo, Valentina Roza y Tarak Shah. JEP: Juan Guillermo Dueñas, Luis Ernesto Jaimes, Camila Lozano, Mary Julieth Murillo y Harbey Rodríguez.

3.8. Auditoría . . . . .	32
3.8.1. Errores humanos . . . . .	32
3.8.2. Esquema . . . . .	32
3.8.3. Expectativas . . . . .	35
3.8.4. Identificación de pares correferentes . . . . .	35
3.8.5. Resultados finales de la vinculación de registros . . . . .	36
<b>4. Imputación estadística de campos faltantes</b>	<b>36</b>
4.1. Campos faltantes . . . . .	36
4.2. Pertenece al conflicto . . . . .	36
4.2.1. Datos faltantes de la variable “pertence al conflicto” . . . . .	37
4.3. Es desaparición forzada . . . . .	38
4.4. Variables de soporte para imputación estadística de datos faltantes . . . . .	38
4.4.1. Ajuste del modelo. . . . .	39
4.5. Imputación estadística de campos faltantes para sexo, etnia, edad, responsable, municipio, pertenece al conflicto, y es desaparición forzada . . . . .	40
<b>5. Patrones de documentación</b>	<b>42</b>
5.1. Homicidio . . . . .	42
5.2. Desaparición . . . . .	42
5.3. Reclutamiento de niños, niñas y adolescentes . . . . .	43
5.4. Secuestro . . . . .	44
5.5. Desplazamiento forzado . . . . .	44
<b>6. Estimación por sistemas múltiples</b>	<b>45</b>
6.1. Detalles técnicos . . . . .	45
6.2. Implementación con datos imputados . . . . .	47
<b>7. Sobre las estimaciones</b>	<b>49</b>
7.1. Sesgo . . . . .	49
7.1.1. Detalles sobre sesgo por sobreestratificación . . . . .	51
7.2. Varianza . . . . .	51
<b>8. Limitaciones y trabajo futuro</b>	<b>53</b>
8.1. Datos . . . . .	53
8.1.1. Sobre los datos de la Fiscalía General de la Nación . . . . .	54
8.2. Vinculación de registros . . . . .	55
8.2.1. Campos mínimos . . . . .	55
8.2.2. Falsos positivos vs. falsos negativos . . . . .	55

8.2.3. Casos de ambigüedad . . . . .	56
8.3. Imputación estadística . . . . .	56
8.3.1. Bases especializadas vs. bases no especializadas . . . . .	56
8.3.2. Extensiones de la imputación . . . . .	56
8.3.3. Imputación guerrilla sin especificar . . . . .	57
8.3.4. Modelo de clasificación . . . . .	59
8.4. Estimación por sistemas múltiples . . . . .	59
8.4.1. Método . . . . .	59
8.4.2. Estratificación . . . . .	60
8.4.3. Método para estimar la varianza . . . . .	60
8.5. Trabajo futuro . . . . .	61
<b>9. Anexo</b>	<b>62</b>
9.1. Agradecimientos . . . . .	62
9.2. Sobre las organizaciones que participaron en el proyecto . . . . .	62
9.2.1. Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición (CEV)	62
9.2.2. Jurisdicción Especial para la Paz (JEP) . . . . .	62
9.2.3. Human Rights Data Analysis Group . . . . .	62
9.3. Datos . . . . .	63
9.3.1. Agencia para la Reincorporación y la Normalización (ARN) . . . . .	63
9.3.2. Asociación Codebac . . . . .	64
9.3.3. Asociación Colombiana de Víctimas de Secuestro y Desaparición Forzada (ACOMIDES)	64
9.3.4. Asociación Nacional de Usuarios Campesinos (ANUC) . . . . .	65
9.3.5. Asociación de Víctimas Nuevo Amanecer . . . . .	65
9.3.6. Cancillería de Colombia . . . . .	66
9.3.7. Centro de Cooperación al Indígena (CECOIN) . . . . .	66
9.3.8. Centro Nacional de Memoria Histórica (CNMH) . . . . .	66
9.3.9. Cifras y Conceptos . . . . .	67
9.3.10. Colectivo de Abogados “José Alvear Restrepo” CAJAR . . . . .	67
9.3.11. Comisión Colombiana de Juristas (CCJ) . . . . .	67
9.3.12. Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición (CEV)	70
9.3.13. Fundación Comité de Solidaridad con los Presos Políticos (CSPP) . . . . .	70
9.3.14. Comunidades Negras de la cuenca del río Domingodó . . . . .	70
9.3.15. Consejería Presidencial para los Derechos Humanos . . . . .	71
9.3.16. Consejo Nacional de Lucha contra el Secuestro y demás Atentados contra la Libertad Personal (CONASE) . . . . .	71
9.3.17. Coordinación Colombia Europa Estados Unidos (CCEEU) . . . . .	72

9.3.18. Corporación Caribe afirmativo . . . . .	73
9.3.19. Corporación para la Defensa y Promoción de los Derechos Humanos Reiniciar . . . . .	73
9.3.20. Corporación Regional para la Defensa de los Derechos Humanos (CREDHOS) . . . . .	74
9.3.21. Departamento Administrativo de Seguridad (DAS) . . . . .	74
9.3.22. Defensoría del Pueblo de Colombia . . . . .	75
9.3.23. Escuela Nacional Sindical (ENS) . . . . .	75
9.3.24. Federación Colombiana de Trabajadores de la Educación (FECODE) . . . . .	76
9.3.25. Fiscalía General de la nación (FGN) . . . . .	76
9.3.26. Fundación Forjando Futuros . . . . .	79
9.3.27. Fundación Lazos de Dignidad (FLD) . . . . .	79
9.3.28. Fundación País Libre . . . . .	80
9.3.29. Grupo Docentes Universidades: Víctimas de Misión Médica de Norte de Santander . . . . .	80
9.3.30. Instituto Colombiano de Bienestar Familiar (ICBF) . . . . .	81
9.3.31. Instituto de Estudios para el Desarrollo y la Paz (INDEPAZ) . . . . .	81
9.3.32. Instituto Nacional de Medicina Legal y Ciencias Forenses (INML) . . . . .	82
9.3.33. Jurisdicción Especial para la Paz (JEP) . . . . .	83
9.3.34. Justicia Penal Militar (JPM) . . . . .	85
9.3.35. Ministerio de Defensa Nacional . . . . .	86
9.3.36. Observatorio Surcolombiano de Derechos Humanos, Paz y Territorio (OBSURDH) . . . . .	87
9.3.37. Observatorio de Tierras . . . . .	87
9.3.38. Organización Nacional Indígena de Colombia (ONIC) . . . . .	87
9.3.39. Personería de Medellín . . . . .	88
9.3.40. Procuraduría General de la Nación (PGN) . . . . .	88
9.3.41. Policía Nacional (PONAL) . . . . .	89
9.3.42. Sistema de monitoreo de riesgos de la Unidad de Investigación Y Acusación - Jurisdicción Especial Para la Paz . . . . .	89
9.3.43. Somos Defensores . . . . .	90
9.3.44. Unidad Administrativa Especial para la Atención y Reparación Integral a las Víctimas . . . . .	90
9.4. Códigos . . . . .	91
9.4.1. Códigos utilizados para la variable “pertenece al conflicto” . . . . .	91
9.4.2. Códigos y Excel utilizados para la variable “es desaparición forzada” . . . . .	151
9.4.3. Filtros adicionales . . . . .	190

## 1. Introducción

Desde mediados de 2020 la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición (CEV), la Jurisdicción Especial para la Paz (JEP) y la organización Human Rights Data Analysis Group (HRDAG) se unieron en el proyecto *Integración de datos y estimaciones estadísticas de víctimas en el marco del conflicto armado*. El objetivo de este proyecto es dotar a las entidades del Sistema Integral de Verdad, Justicia, Reparación y No Repetición (SIVJRNR) de argumentos científicos sólidos para develar la magnitud de la violencia, mediante la estimación del subregistro y la identificación de patrones de victimización de homicidio,<sup>1</sup> desaparición forzada, secuestro, desplazamiento forzado y reclutamiento de niños, niñas y adolescentes.

Este proyecto consta de cuatro componentes:

1. Vinculación de registros de bases de datos disponibles sobre el conflicto armado en Colombia.
2. Generación de estimaciones y análisis del subregistro relacionado con violaciones de Derechos Humanos y Derecho Internacional Humanitario.
3. Asesoramiento en la implementación del análisis de datos relacionados con estas estimaciones.
4. Fortalecimiento de capacidades en las entidades del SIVJRNR para la contrastación de datos e implementación de modelos estadísticos que fortalezcan el análisis de datos sobre derechos humanos.

El proyecto se desarrolló en cuatro fases. La Fase 1 se concluyó el 18 de diciembre de 2020 y la Fase 2 el 14 de abril de 2021. En septiembre de 2021 se presentó el informe de la Fase 3 con un detalle de las metodologías implementadas hasta el momento. Finalmente, este informe presenta las actualizaciones del informe a la Fase 3, incluyendo nuevas bases de datos, decisiones metodológicas y análisis.

### 1.1. ¿Cuál es nuestra meta?

La base del conocimiento sobre violaciones de derechos humanos es la documentación que, en las últimas décadas, ha sido hecha en bases de datos. Es posible elegir qué casos registrar y cada proyecto tiene un universo de interés definido por los criterios que elige. Estos pueden incluir un foco en un grupo étnico o en un lugar geográfico, en registrar solo cadáveres o casos reportados a la policía, casos que hayan sido investigados, casos reportados a los medios, casos de personas solicitando compensación por sus pérdidas en el conflicto, o casos investigados detalladamente para estudiar su relación con el conflicto armado. Estas son solo algunas formas en las que los creadores de las bases de datos definen su universo de interés. En los últimos 40 años muchas personas, organizaciones e instituciones del Estado han creado cientos de bases de datos con información de violencia en Colombia.

A partir de datos normalmente queremos conocer los patrones: ¿hubo más violencia en 2001 o en 2002? ¿qué grupo armado cometió la mayor proporción de homicidios? Sin embargo, cualquier base de datos, incluso las mejores, contienen solo una imagen del mundo según los criterios definidos para ese proyecto. Una base de datos puede decirnos si las personas *reportaron* más homicidios a esta base en 2001 o 2002, pero los patrones de la base no tienen necesariamente una relación con los patrones del mundo. Tal vez, los investigadores en el 2001 eran muy activos y en el 2002 fueron reemplazados por otros con menos interés en hacer trabajo de campo, por lo que se vio el pico en 2001. Este pico nos diría más del proyecto, no de la violencia en el mundo. Cada base de datos es finita y limitada por el presupuesto del proyecto, su habilidad de alcanzar a las comunidades afectadas y ganar su confianza y por la capacidad de sus investigadores de registrar toda la información que reciben. Las bases de datos son necesariamente parciales, incluyendo solo algunos eventos que son parte del universo de interés. Esta no es una crítica a ninguna base de datos. Todas las bases de

---

<sup>1</sup>Se entienden “homicidio” como la muerte intencional de personas por parte de alguno de los actores del conflicto armado, así como aquella que se produce en personas civiles en desarrollo de hostilidades entre actores armados, ya sea en medio de combates o ataques indiscriminados, o por la activación de artefactos explosivos de guerra, armas trampa, minas antipersona u otro artefacto. En este proyecto no se pudo discriminar todavía entre las muertes en combate de civiles y combatientes. Está pendiente realizarlo en futuros proyectos.

datos tienen subregistro y responden a una pregunta clave: ¿qué hemos documentado? Una buena base de datos nos permite saber qué hay en ella.

Los argumentos estadísticos requieren algo más. Vemos a las estadísticas para que nos digan algo del mundo, no de los patrones en los datos. Los argumentos estadísticos requieren que vayamos más allá de los datos para estimar qué es cierto en el mundo. Con datos limitados no podemos saber la verdad del mundo, entonces tenemos que aproximarnos usando los datos que tenemos y modelos de probabilidad creados y testeados en los últimos cientos de años. Podemos estimar el resultado más reciente. Tenemos que tener en cuenta la información faltante de todos los registros disponibles y considerar el universo entero de casos, tanto documentados como no documentados. De nuevo, nuestros datos están limitados, por lo que nuestro resultado no es un único número, es un rango en el que es probable que esté la respuesta verdadera. Una estimación renuncia a la definición precisa de lo que contiene una base de datos para aproximarse a la verdad en el mundo. Ambos métodos tienen roles importantes para entender el mundo: las buenas bases de datos nos informan sobre los casos que contienen, mientras que las buenas estimaciones nos informan sobre el mundo. Necesitamos ambas.

Nuestra meta con este proyecto es crear un análisis estadístico de los patrones del conflicto armado en Colombia. Por análisis estadístico nos referimos a que queremos tomar en consideración dos componentes fundamentales de medidas matemáticas rigurosas: sesgo y varianza. El sesgo, en sentido técnico, es la diferencia entre nuestra estimación y la realidad. No podemos medirlo porque no sabemos exactamente qué es la verdad. Sin embargo, tenemos más de un siglo de matemáticas, teoría estadística y práctica que nos muestra cómo usar datos en modelos estadísticos para minimizar el sesgo. La varianza es la medida de qué tan precisas son nuestras estimaciones. La varianza es una medida formal y matemática de la incertidumbre, que se expresa en un intervalo. Por ejemplo, en el proyecto encontramos que hubo más de 450.000 homicidios en el conflicto armado. Más formalmente, encontramos que tenemos registros de más de 450.000 víctimas con nombre y apellido en una o más bases de datos, que estimamos fueron víctimas del conflicto armado. Esta última parte, “estimamos fueron víctimas del conflicto armado,” no es precisa porque hay miles de víctimas para las que no sabemos si sus homicidios ocurrieron o no en el marco del conflicto. Entonces, la expresión formal de nuestra estimación se basa en la información que tenemos y en su uso para considerar otra información que no tenemos sobre exactamente qué casos están o no relacionados con el conflicto armado. Estimamos que hay entre 436.747 y 464.585 víctimas de homicidio relacionadas con el conflicto armado.

Incluir la información de si un registro es o no parte del conflicto es solo el primer paso. Hay muchas más víctimas que no fueron documentadas o que fueron documentadas solo con información parcial. Por ejemplo, tenemos casos sin el nombre de la víctima (como “campesino en la finca X”) o cadáveres que no fueron identificados. Sabemos que hay muchos más cuerpos escondidos en fosas clandestinas, quemados o en ríos. Algunos de esos homicidios son clasificados erróneamente como accidentes. Y muchos no son documentados porque las familias de las víctimas o los testigos deciden, seguramente con razón, que reportar la muerte podría generar un problema para sus vidas. Teniendo en cuenta toda la información que tenemos y usando modelos estadísticos que han sido utilizados para estimar poblaciones humanas desde mediados del siglo XX, estimamos que hubo más de 800.000 homicidios en el marco del conflicto armado. Como es una estimación tiene un intervalo de 777.852–852.756.

¿Qué queremos decir con una “estimación?” Hemos descrito dos tipos de estimación que usamos en nuestro proyecto. La primera se llama “imputación estadística” (no tiene nada que ver con imputación en términos jurídicos). Muchos de nuestros registros tienen información faltante sobre un campo importante. Continuando con el ejemplo anterior, tenemos 374.567 para los que estamos seguros de que su homicidio está relacionado con el conflicto armado y 75.752 de las que estamos seguros que su homicidio no se relaciona con el conflicto armado. Sin embargo, tenemos registro de otras 104.050 para quienes no sabemos si son parte o no del conflicto.

¿Por qué deberíamos hacer estimaciones? Podríamos decidir calcular las estadísticas solo con las 374.567 víctimas para las que tenemos certeza. Pero, ignorar a las 104.050 para las que no tenemos certeza es hacer un supuesto muy fuerte. En estadística a este supuesto se le conoce como “datos faltantes **completamente** al azar” y significa que usando solo los 374.567 homicidios asume que la proporción de los 104.050 casos desconocidos es idéntica a la proporción que observamos. Esto muestra que todo análisis, sin excepción, que

hacemos de una estadística se basa en un modelo, aunque los analistas no digan explícitamente qué supone el modelo.

El problema de asumir que los datos son faltantes completamente al azar es aun más complicado cuando analizamos una segunda variable. Por ejemplo, si queremos entender la relación de la violencia en el tiempo comparando los homicidios en 2002 con los homicidios en 2015 e ignoramos los 104.050 casos que no sabemos si son parte o no del conflicto, estamos suponiendo que si incluyéramos a los casos de estos 104.050 que sí ocurrieron en el marco del conflicto armado, nuestro entendimiento de la relación entre estos dos años no cambiaría. Este supuesto no es correcto, como mostramos en la Sección 6.1. Mientras que en 2002 casi todos los casos que eran parte del conflicto armado fueron documentados como tal, en 2015 encontramos que una fracción sustancial de los homicidios que no tienen información de su relación con el conflicto armado están probablemente relacionados con el conflicto. Estimar el subregistro es fundamental porque los casos con menor probabilidad de ser documentados son los casos de las personas probablemente más excluidas de los datos: de zonas rurales, de grupos étnicos minoritarios, con bajos ingresos y con familias que tienen preocupaciones sobre su seguridad. Los modelos estadísticos tratan de reconocer estos casos faltantes, como mostramos en los hallazgos. Si no podemos documentar a cada víctima deberíamos al menos tratar de estimar cuántas son y que su sufrimiento no sea totalmente silenciado.

Distintos proyectos tienen distintos criterios para incluir datos. Estos criterios reflejan las ideas del proyecto sobre lo que tratan de aprender y expresar y, por lo tanto, las estadísticas que vienen de cada proyecto suelen ser diferentes (ver por ejemplo la comparación del RUV y el CNMH en [Suárez 2018](#)). En nuestro proyecto tratamos de crear análisis estadísticos que incluyen a todas las víctimas del conflicto armado para homicidio, secuestro, desaparición forzada y reclutamiento de niños, niñas y adolescentes. Queremos presentar nuestras estimaciones científicamente rigurosas, lo que significa que tenemos que usar un modelo para reducir el sesgo y describir la varianza. Hay dos tipos de información faltante que tenemos que manejar. El primer tipo de información faltante son los campos faltantes, que no están registrados en cada caso, como si el hecho ocurrió o no en el marco del conflicto o qué grupo armado es el presunto responsable. Este vacío lo trabajamos usando imputación estadística, explicada en la Sección 4. El segundo tipo de vacío de información son las víctimas que no fueron registradas. Estos son los registros faltantes y para ellos estimamos el subregistro, como describimos en la Sección 6.

Tenemos muchas bases de datos y muchas víctimas aparecen en más de una base de datos. Esta deduplicación es quizás la parte más grande del proyecto. Los registros que se refieren a una misma víctima son llamados “correferentes.” Algunas bases de datos también contienen más de un registro para la misma víctima, lo que puede ser resultado de cómo se extrajo la información de la base de datos original o de que el proyecto no podía identificar a las mismas víctimas. Manejamos las dos formas de deduplicación al tiempo, como explicamos en la Sección 3. Una observación es que las fuentes tienden a agruparse. Es decir, grupos de fuentes muchas veces documentan los mismos casos, con relativamente pocos casos en común con otras fuentes. A este análisis le llamamos “patrones de documentación” y lo mostramos en la Sección 5.

### 1.2. Panorama de los datos

La CEV y la JEP recibieron cientos de conjuntos de datos de múltiples entidades u organizaciones sociales y de víctimas. Dado que este proyecto requiere información a nivel de víctima, después de revisar y validar, se implementaron técnicas determinísticas y probabilísticas, que permitieron consolidar una base de datos que integra los registros de un total de 112 conjuntos de datos. Estos aportaron información sobre homicidio, desaparición forzada, secuestro, reclutamiento de niños, niñas y adolescentes, desplazamiento forzado y exilio<sup>2</sup> de organizaciones de la sociedad civil, de víctimas e instituciones oficiales. Nos limitamos a estos hechos de violencia porque nos enfocamos en aquellos que las personas suelen sufrir solo una vez. Desplazamiento es la excepción, que incluimos por la alta cantidad de datos.

Este es el primer proyecto que integra esta cantidad de bases de datos con un enfoque de derechos humanos. Con un análisis rápido, se puede ver que las bases de datos abarcan diferentes periodos, regiones, poblaciones

---

<sup>2</sup>Solamente la base de datos de la CEV aportó información sobre víctimas de exilio, por lo cual no fue posible realizar un proceso de integración de datos ni de estimaciones de subregistro.

de víctimas, y por lo cual cuentan diferentes historias. Cada base de datos tiene información parcial, y por tanto también sesgos de cobertura. Esto no es una crítica a las bases de datos, por el contrario, es una característica básica de todo proceso de recolección de información. Las entidades u organizaciones que contribuyeron con información al SIVJRN y por tanto al proyecto son:

1. Agencia para la Reincorporación y la Normalización (ARN);
2. Asociación Codebac;
3. Asociación Colombiana de Víctimas de Secuestro y Desaparición Forzada (ACOMIDES);
4. Asociación Nacional de Usuarios Campesinos (ANUC);
5. Asociación de Víctimas Nuevo Amanecer;
6. Cancillería de Colombia;
7. Centro de Cooperación al Indígena (CECOIN);
8. Centro Nacional de Memoria Histórica (CNMH);
9. Cifras y Conceptos;
10. Colectivo de Abogados José Alvear Restrepo (CAJAR);
11. Comisión Colombiana de Juristas (CCJ);
12. Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición (CEV);
13. Comité de Solidaridad con los Presos Políticos (CSPP);
14. Comunidades Negras de la cuenca del río Domingodó;
15. Consejería Presidencial para los Derechos Humanos y Asuntos Internacionales (SIDDHH);
16. Consejo Nacional de Lucha contra el Secuestro y demás Atentados contra la Libertad Personal (CONASE);
17. Coordinación Colombia Europa Estados Unidos (CCEEU)
18. Corporación Caribe Afirmativo;
19. Corporación para la Defensa y Promoción de los Derechos Humanos Reiniciar;
20. Corporación Regional para la Defensa de los Derechos Humanos (CREDHOS);
21. Departamento Administrativo de Seguridad (DAS).
22. Defensoría del Pueblo;
23. Escuela Nacional Sindical;
24. Federación Colombiana de Trabajadores de la Educación (FECODE);
25. Fiscalía General de la Nación (FGN);
26. Fundación Forjando Futuros;
27. Fundación Lazos de Dignidad;
28. Fundación País Libre;
29. Grupo Docentes Universidades;
30. Instituto Colombiano de Bienestar Familiar (ICBF);
31. Instituto de Estudios para el Desarrollo y la Paz (INDEPAZ);
32. Instituto Nacional de Medicina Legal y Ciencias Forenses (INML);
33. Jurisdicción Especial para la Paz (JEP);
34. Justicia Penal Militar (JPM);
35. Ministerio de Defensa Nacional;
36. Observatorio Surcolombiano de Derechos Humanos, Paz y Territorio (OBSURDH);
37. Observatorio de Tierras;
38. Organización Nacional Indígena de Colombia (ONIC);
39. Personería de Medellín;
40. Procuraduría General de la Nación (PGN);
41. Policía Nacional (PONAL);
42. Sociedad Internacional para los Derechos Humanos (SIDDHH);
43. Somos Defensores;
44. Unidad Administrativa Especial para la Atención y Reparación Integral a las Víctimas

La importancia de este proyecto radica en que los registros de violaciones de derechos humanos sufren de dos tipos de vacíos de información: campos faltantes y subregistro.<sup>3</sup> Por lo tanto, no es correcto analizar

---

<sup>3</sup>Más detalle sobre esto se puede ver en este [video](#).



patrones de violencia a partir de bases de datos que registren a víctimas. Ya que los datos observados no siempre reflejan la realidad, sino la documentación realizada por cierto proyecto.

El primer tipo de vacío de información son los campos faltantes y se limita a los datos registrados. Todas las bases de datos cuentan con distintos campos: algunos están relacionados con el hecho (como el año, municipio, departamento y responsable), mientras que otros están relacionados con las personas (nombre, apellido, edad, sexo y etnia, entre otros). Sin embargo, cuando las organizaciones o instituciones documentan los hechos de violencia, no siempre tienen la información completa y es normal que algunos campos queden vacíos para algunos registros. Hay muchas situaciones en las que esto puede ocurrir. Por ejemplo, puede suceder que se tenga registro de la persona, pero no se sepa el día exacto en el que ocurrió el hecho; o tal vez se tiene el nombre y apellido, pero no la cédula; mientras que algunas ocasiones no se conoce la etnia; entre otras. Además de que el campo esté vacío, puede suceder que tenga información incorrecta, como un error de digitación de la cédula o un error ortográfico en el nombre o apellido.

El segundo tipo de vacío de información es el subregistro y va más allá de los datos registrados. El subregistro se explica porque hay víctimas que no están documentadas, lo que puede ocurrir por distintos motivos. Por ejemplo, puede deberse a que la víctima o sus familiares tengan miedo de hacer la denuncia y opten por el silencio; también puede ocurrir que el hecho de violencia haya tenido lugar en un sitio muy lejano, al que no lleguen las organizaciones o instituciones; podría ser porque aunque fueron víctimas dentro del marco del conflicto armado, se registró una violación distinta a la que sufrieron; sus cuerpos fueron encontrados, pero no han sido identificados; o sus cuerpos fueron tirados a ríos o fosas comunes y no haya ningún registro. A diferencia del primer tipo de vacío de información, en este caso no se conoce cuántos registros faltan.

A pesar de unir las 112 fuentes de información estos vacíos están presentes. Por lo tanto, pueden ser una fuente de sesgo, bien sea por los campos faltantes o por subregistro. Entonces, es necesario encontrar la forma de tratar estos vacíos, lo que se convierte en un reto para la investigación.

Para corregir los dos tipos de vacío de información es necesario seguir tres grandes pasos. Primero, integrar y deduplicar la información existente. Segundo, imputar estadísticamente los campos faltantes. Es decir, utilizar métodos estadísticos que permitan inferir valores plausibles para las características de las víctimas que tienen campos faltantes. Tercero, estimar el universo de víctimas, incluyendo el subregistro, usando estimación por sistemas múltiples.

### 1.3. Flujo de trabajo

En el proyecto se usó el flujo de trabajo “cuántico” definido por HRDAG. Según este, todas las tareas deben ser auto contenidas y auto documentadas. El proyecto cuenta con seis grandes tareas, como se observa en la Figura 1. Se inicia con la tarea de “individual” en la que se hace el procesamiento de los datos para cada una de las bases de cada organización. Allí se importan, limpian, estandarizan, filtran, exportan y demás tareas para trabajar con los datos en el resto del proyecto. Una vez las variables y el contenido de las distintas bases son estandarizadas, en el sentido de tener los mismos nombres para las mismas variables, mismo uso de fechas, etc., hay dos tareas paralelas. Por un lado, está la tarea de “vinculación de registros.” Como su nombre lo indica, en esta tarea se hace la vinculación de registros o la deduplicación. Por otro lado, está la tarea de “variables de soporte,” en la que se entrena a la red neuronal para generar estas variables. Cuando se tienen estas dos tareas, es decir, los registros deduplicados y las variables de soporte, se pasa a la imputación estadística. Con los resultados de la imputación es posible hacer las estimaciones. Finalmente, a partir de todas las tareas anteriores, llegamos a los resultados.

Los datos originales e intermedios reposan en un servidor de la CEV y fueron manejados con control de versionamiento en un repositorio de GitHub. Vale la pena resaltar que el repositorio contiene solo el código y no los datos.

- En el primer nivel se tienen las tareas. Por ejemplo, vinculación de registros, imputación de campos faltantes y estimación por sistemas múltiples (ESM)
- En el segundo nivel, dentro de cada una de estas tareas, hay al menos tres directorios: input, src y output. En input se incluyen los archivos que van a ser leídos. En src se incluyen los scripts que van a

ser ejecutados en los archivos de entrada. En output se escriben los resultados de los scripts de src. La información de input y output solo reposó en el servidor de la CEV, mientras que los de src estaban disponibles en el GitHub. Además, en cada tarea hay un Makefile, que garantiza la replicabilidad.

Gracias al hecho de tener las tareas en estos tres directorios la tarea está auto contenida. También, la tarea está auto documentada porque todo lo necesario para entenderla está en sus archivos. Más información sobre el flujo de trabajo está disponible en [Ball \(2016\)](#) y [Shah \(2019\)](#).

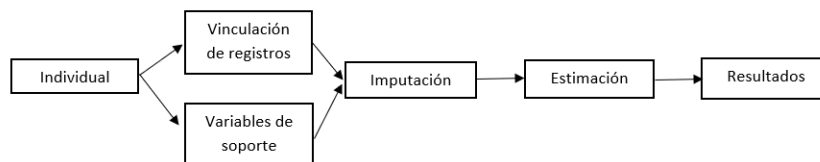


Figura 1: Flujo de trabajo utilizado en el proyecto.

## 1.4. Resumen de este reporte

Este informe está dividido en 9 secciones, incluyendo esta introducción. En la segunda sección presentamos los hallazgos estadísticos del proyecto. En la tercera sección detallamos la vinculación o deduplicación de registros. Esta tarea es la que integra la información de los 112 conjuntos de datos y está dividida en 6 pasos, que son presentados según su respectivo orden. En la cuarta sección exponemos la imputación estadística de campos faltantes, que corrige el primer tipo de vacío de información. Además, allí presentamos cómo trabajamos con las bases especializadas y no especializadas en conflicto armado, así como con las desapariciones forzadas y no forzadas. En la quinta sección mostramos los patrones de documentación después de la imputación estadística. Es decir, las víctimas del conflicto documentadas por las distintas organizaciones. En la sexta sección presentamos el método implementado para estimar el subregistro de poblaciones por sistemas múltiples. Este es el método que utilizamos para subsanar el segundo tipo de vacío de información, es decir, el subregistro. En la séptima sección ahondamos en dos características de las estimaciones: el sesgo y la varianza. En la octava sección mostramos las limitaciones y trabajo futuro. Por último, en la novena sección, disponibilizamos los anexos.

## 2. Hallazgos estadísticos

En esta sección presentamos los hallazgos del proyecto. Cada uno de los hechos de violencia tiene un periodo distinto de tiempo, según la disponibilidad de datos, de la siguiente forma: desaparición forzada (1985–2016), desplazamiento forzado (1985–2019), homicidio (1985–2018), reclutamiento de niños, niñas y adolescentes (1990–2017) y secuestro (1990–2018).

Vamos a presentar tres formas distintas de entender los datos. Primero, los datos observados, que se limitan por los dos tipos de vacíos de información (campos faltantes y registros faltantes). Segundo, los datos imputados estadísticamente, que solucionan el problema de los campos faltantes con una medida de la incertidumbre que viene de este vacío de información. Tercero, estimaciones de la población, que solucionan los dos tipos de vacíos de información (campos faltantes y registros faltantes) y nos permiten calcular la incertidumbre que viene de estos. Note que para desplazamiento no se incluyen imputaciones estadísticas ni estimaciones, pues dada la dominancia del Registro Único de Víctimas (RUV) en la documentación de víctimas de este hecho, no es recomendable. A lo largo de los capítulos del Informe Final se ha presentado el dato más probable para la imputación estadística. En la Sección 4 describimos el proceso en el que hacemos 10 réplicas, que después son combinadas. El dato más probable es la moda de las 10 réplicas.

Para comenzar, concentrémonos en los datos observados. Es decir, en la segunda columna de cada tabla. Lo primero que se evidencia es que los homicidios son el principal hecho de violencia documentado ocurrido en el conflicto armado (excluyendo desplazamiento forzado) con 374.567 víctimas en la Tabla 2, seguido por desaparición forzada con 110.351 víctimas en la Tabla 1, secuestro con 46.319 víctimas en la Tabla 4 y reclutamiento de niños, niñas y adolescentes con 14.669 en la Tabla 3.

Ahora bien, los únicos de estos hechos de violencia que por definición ocurrieron en el marco del conflicto armado son el reclutamiento de niños, niñas y adolescentes y el desplazamiento forzado. Entonces, para los demás hechos fue necesario imputar la variable de “pertenece al conflicto.” Además, para desaparición forzada fue necesario imputar estadísticamente la variable “es desaparición forzada” y para reclutamiento de niños, niñas y adolescentes fue necesario imputar estadísticamente la variable edad.

Cuando hacemos la imputación estadística para solucionar el problema de campos faltantes, el total de víctimas documentadas cambia para algunos hechos. Por ejemplo, en el caso de desaparición forzada, observamos 14.539 casos de desaparición identificados como desaparición forzada en el marco del conflicto armado, mientras que 14.539 desapariciones fueron documentadas sin información de si fueron forzadas o si pertenecen al conflicto armado.<sup>4</sup> En la imputación estadística, el modelo de la imputación asigna valores a cada uno de los campos faltantes en los registros. Después de llenar los valores faltantes en todos los registros, podemos usarlos para recalcular las cantidades de interés. Estimamos que hay un promedio de 121.768 víctimas de desaparición forzada relacionadas al conflicto armado que corresponde al número de víctimas documentadas más probable. La imputación estadística también mide la incertidumbre en el número de desapariciones forzadas, entonces las tablas también muestran el rango (con los límites inferior y superior). En el caso de desaparición forzada, la Tabla 1 muestra el rango: 119.777–123.759.

Al hacer la imputación estadística de campos faltantes, en la tercera columna de las tablas, es interesante que el límite inferior que más se distancia del valor observado es el de homicidio. Esto indicaría que este es el hecho de violencia al que más se le atribuyen víctimas de las que no se sabía si estaban o no relacionadas con el conflicto armado. Le siguen la desaparición forzada y el secuestro. Con base en la imputación estadística, en Colombia habría 450.666 víctimas de homicidio, 121.768 víctimas de desaparición forzada, 50.770 víctimas de secuestro y 16.238 víctimas de reclutamiento de NNA.

En el grupo de columnas al lado derecho de cada tabla mostramos las estimaciones del número de víctimas total, incluso las no documentadas en las fuentes integradas en este proyecto. La mejor manera de entender estas estimaciones es considerar su rango. Por ejemplo, para homicidios, el intervalo mostrado en la Tabla 2 es entre 777.852 y 852.756 víctimas. En cuanto a la desaparición forzada, la Tabla 1 muestra que en Colombia habría entre 204.395 y 225.410 víctimas. Las Tablas 3 y 4 muestran las estimaciones del número de víctimas total para secuestro y reclutamiento de niños, niñas y adolescentes.

Finalmente, al estudiar las estimaciones se evidencia que los homicidios son el principal hecho de violencia del conflicto con entre 777.852 y 852.756 víctimas. En cuanto a la desaparición forzada, en Colombia habría entre 204.395 y 225.410 víctimas. Por su parte, las víctimas de reclutamiento serían entre 27.101 y 40.828. En cuanto a las víctimas de secuestro, estas oscilarían entre 74.768 y 92.849. Por último, al hacer la deduplicación de registros en Colombia habría 7.752.964 de desplazamiento forzado que cuentan con nombre, apellido, lugar y fecha.

---

<sup>4</sup>Ver la Sección 4.3 para una discusión de nuestro proceso de identificar estas condiciones.

Tabla 1: Distribución del número de víctimas de desaparición forzada de acuerdo con su relación con el conflicto armado (1985–2016)<sup>a</sup>

Relación al conflicto armado	Observado	Imputado			Estimado		
	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Víctimas vinculadas al conflicto armado	110.351	119.777	121.768	123.759	204.395	214.418	225.410
Víctimas sin relación al conflicto armado	47.207	48.337	50.328	52.319	-	-	-
Sin Información	14.539	-	-	-	-	-	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado - de acuerdo con las reglas definidas. Además, es importante aclarar que las palabras: inf, sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); y sup: límite superior (ya sea de la imputación o la estimación.)

Tabla 2: Distribución del número de víctimas de homicidio de acuerdo con su relación con el conflicto armado (1985–2018)<sup>a</sup>

Relación al conflicto armado	Observado	Imputado			Estimado		
	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Víctimas vinculadas al conflicto armado	374.567	436.747	450.666	464.585	777.852	813.707	852.756
Víctimas no vinculadas al conflicto armado	75.752	89.784	103.703	117.622	-	-	-
Sin información	104.050	-	-	-	-	-	-

<sup>a</sup> Es importante aclarar que las palabras: inf, sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); y sup: límite superior (ya sea de la imputación o la estimación.)

Tabla 3: Distribución del número de víctimas de reclutamiento de niños, niñas y adolescentes de acuerdo con su relación con el conflicto armado (1990–2017)<sup>a</sup>

Relación al conflicto armado	Observado	Imputado			Estimado		
	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Víctimas vinculadas al conflicto armado	14.669	16.027	16.238	16.449	27.101	32.806	40.828

<sup>a</sup> Por definición, todas las víctimas de reclutamiento están vinculadas al conflicto armado, sin embargo, esta tabla incluye la imputación de la variable EDAD, ya que este análisis incluye únicamente las víctimas menores de 18 años en el momento del reclutamiento.

<sup>b</sup> Es importante aclarar que las palabras: inf, sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); y sup: límite superior (ya sea de la imputación o la estimación.)

Tabla 4: Distribución del número de víctimas de secuestro de acuerdo con su relación con el conflicto armado (1990–2018)<sup>a</sup>

Relación al conflicto armado	Observado	Imputado			Estimado		
	N.	N(inf)	N(imp)	N(sup)	N(inf)	N(est)	N(sup)
Víctimas vinculadas al conflicto armado	46.319	50.234	50.770	51.306	74.768	82.765	92.849
Víctimas sin relación al conflicto armado	4.647	4.647	5.102	5.638	-	-	-
Sin información	4.906	-	-	-	-	-	-

<sup>a</sup> Es importante aclarar que las palabras: inf, sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); y sup: límite superior (ya sea de la imputación o la estimación.)

En las gráficas de la Figura 2 encontramos las víctimas observadas (línea negra), imputadas estadísticamente (línea azul) y estimadas (área verde) para los 4 hechos de violencia a los que se les hizo estimación por sistemas múltiples. Los puntos verdes muestran que la varianza es tan amplia que no cabe en la gráfica. Lo primero que encontramos es que las estimaciones siguen las tendencias de las víctimas documentadas. Al analizar cada uno de los hechos de violencia se deben notar las escalas del eje vertical. Como mencionamos anteriormente, las magnitudes de homicidio son muy superiores a las de los demás hechos de violencia (excepto desplazamiento).

Para todos los hechos de violencia se evidencia que hay un pico de violencia a inicios del 2000. Este pico podría

Tabla 5: Distribución del número de víctimas de desplazamiento forzado de acuerdo con su relación con el conflicto armado (1985–2019)

Relación al conflicto armado	Observado
	N.
Víctimas vinculadas al conflicto armado	7.752.964

haber sido particularmente alto para desaparición forzada, reclutamiento de niños, niñas y adolescentes y secuestro, en los que la varianza es tan amplia que no se presenta en la gráfica. También, vale la pena observar que los casos de reclutamiento y secuestro en los últimos años llegan casi a cero y el subregistro se reduce. No obstante, persisten los casos y, especialmente en homicidios, el subregistro.

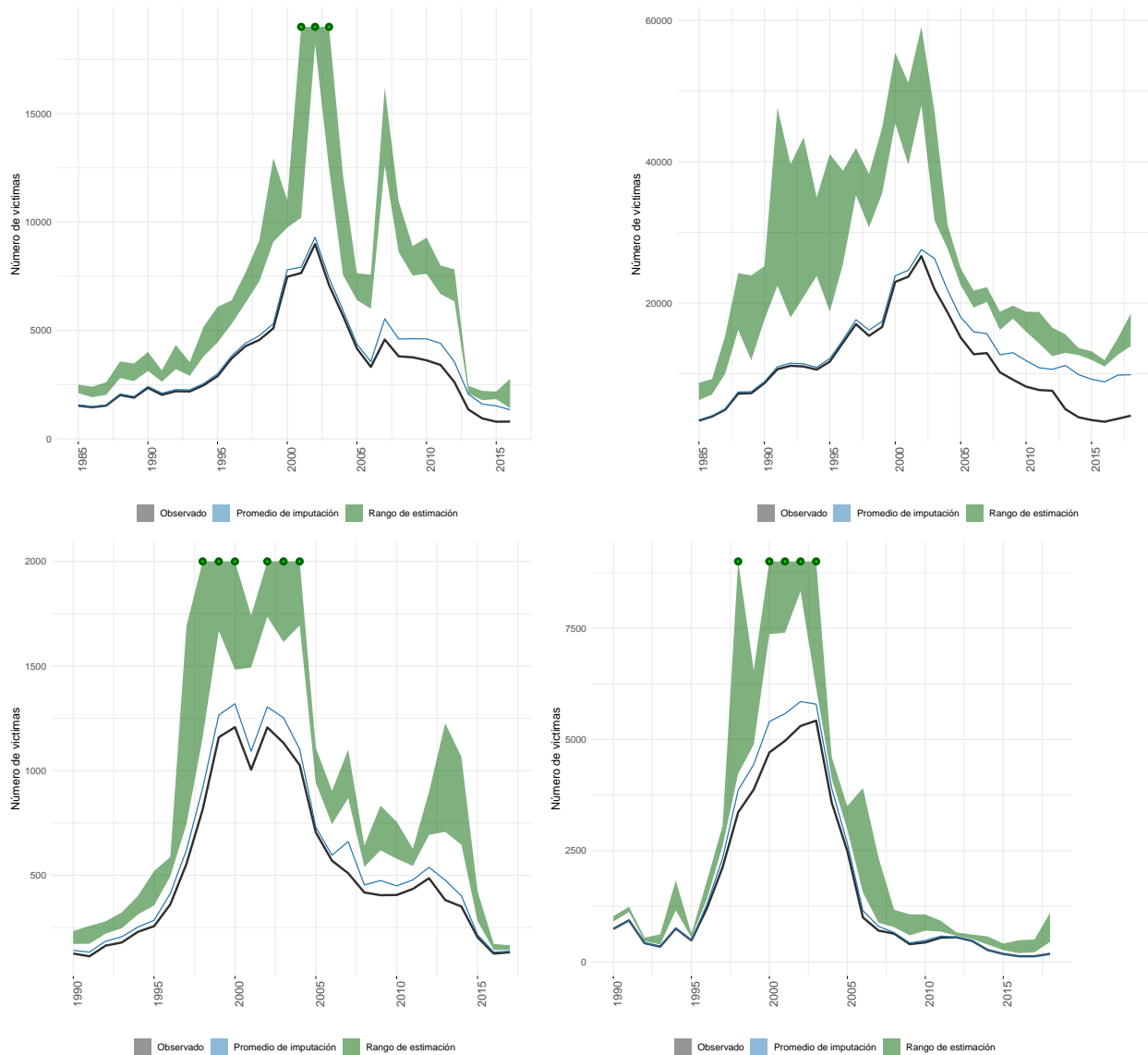


Figura 2: Víctimas observadas, imputadas estadísticamente y estimadas para desaparición forzada (sup-izq), homicidio (sup-der), reclutamiento de niños, niñas y adolescentes (inf-izq) y secuestro (inf-der) desagregados por presunto responsable

La Figura 3 muestra los 10 principales departamentos para cada hecho de violencia. Se evidencia que Antioquia ocupó el primer lugar en tres de los hechos de violencia (desaparición forzada, homicidio y secuestro). Llama la atención que, además de Antioquia; Cesar, Meta, Nariño, Norte de Santander y Valle del Cauca están dentro de los 10 departamentos con más víctimas para los cuatro hechos de violencia. Además, para cada hecho el orden de los departamentos es aproximadamente igual sin importar los datos que usemos (observados, imputados estadísticamente o estimados). Por ejemplo, si consideramos solo los datos observados (barra negra), Antioquia tiene el mayor número de homicidios. También es el caso con los datos de la imputación estadística (barra azul) y las estimaciones de las víctimas no documentadas (barra verde).

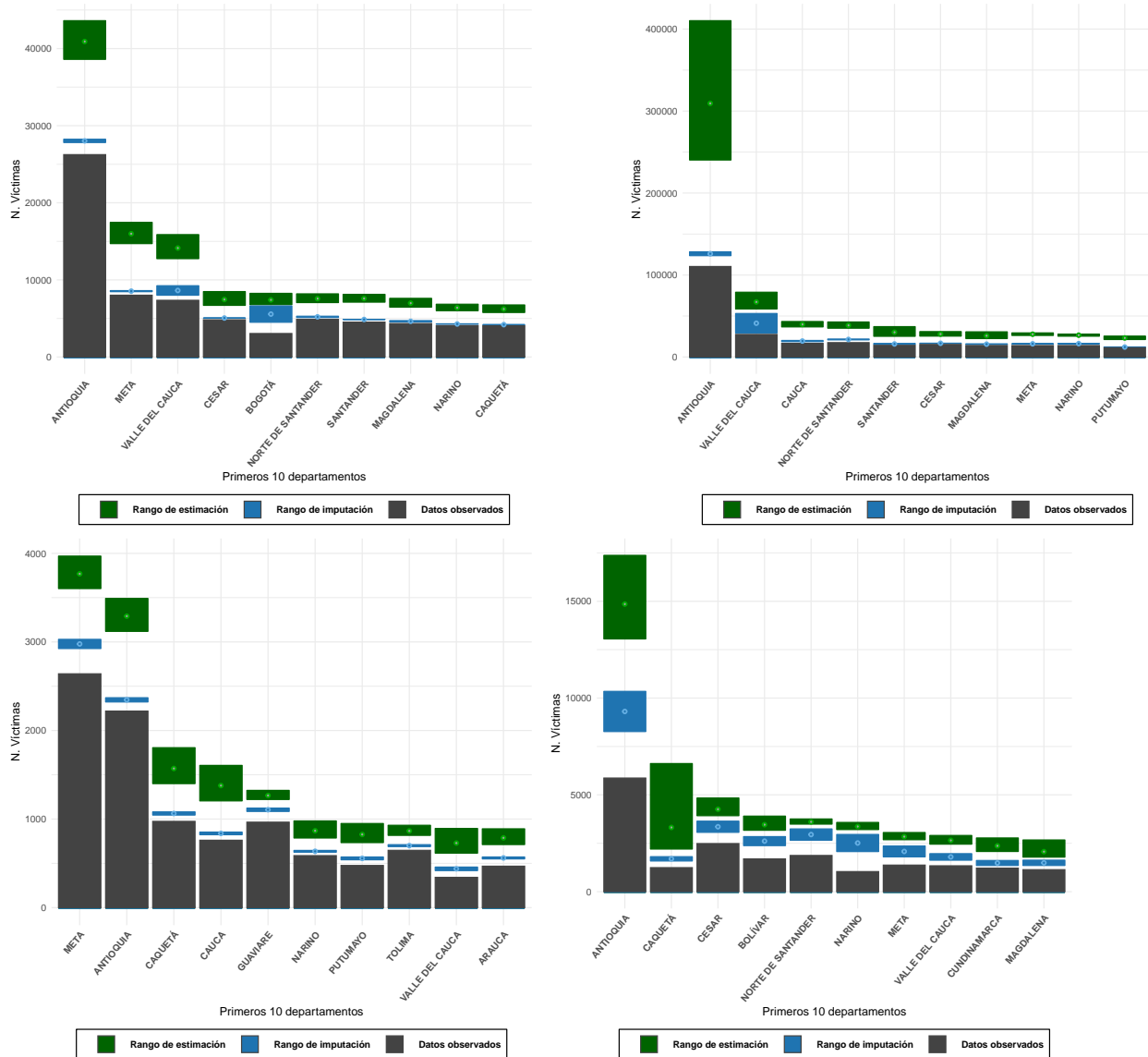


Figura 3: 10 principales departamentos para desaparición forzada (sup-izq), homicidio (sup-der), reclutamiento de niños, niñas y adolescentes (inf-izq) y secuestro (inf-der) desagregados por presunto responsable

En las siguientes tablas observamos el sexo de las víctimas para los distintos hechos de violencia. Se evidencia que los hombres fueron las principales víctimas directas para todos los hechos. Para los datos observados se tienen los siguientes resultados. La proporción de víctimas hombres de desplazamiento forzado es de 0,48. Frente a desaparición forzada, su proporción sería 0,85 de las víctimas, mientras que en homicidio sería 0,85, en reclutamiento 0,69 y en secuestro 0,78.

En cuanto a la imputación estadística, hallamos lo siguiente. En desaparición forzada, la proporción de hombres estaría entre 0,82 y 0,84 de las víctimas, mientras que en homicidio serían entre 0,88 y 0,84, en reclutamiento 0,68 y 0,72, y secuestro entre 0,77 y 0,79. En el caso de la integración de datos de desplazamiento, la proporción de las víctimas hombres para las víctimas imputadas estadísticamente sería de entre 0,48 y 0,48

Ahora bien, para las estimaciones encontramos que en el caso de desaparición forzada, la proporción de hombres estaría entre 0,79 y 0,87 de las víctimas, mientras que en homicidio serían entre 0,88 y 0,87, en reclutamiento 0,55 y 1, y secuestro entre 0,69 y 0,87. En el caso de la integración de datos de desplazamiento, la proporción de las víctimas hombres para las víctimas imputadas estadísticamente sería de 0,48.

Tabla 6: Sexo de la víctima desplazamiento (1985–2019)<sup>a</sup>

Sexo	Observado		
	N.	P(NA)	P
MUJER	3.994.536	0,52	0,52
HOMBRE	3.750.968	0,48	0,48
Obs. con NAs	7.460	0,00	-
Total	7.752.964	1,00	1,00

<sup>a</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1.

Tabla 7: Sexo de la víctima desaparición forzada (1985–2016)

Sexo	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf )	N(est)	N(sup )	P(inf )	P(est )	P(sup )
HOMBRE	92.934	0,84	0,85	100.088	101.292	102.496	0,82	0,83	0,84	161.406	169.535	178.545	0,79	0,83	0,87
MUJER	16.662	0,15	0,15	19.591	20.477	21.363	0,16	0,17	0,18	32.310	34.922	37.994	0,16	0,17	0,19
Obs. con NAs	755	0,01	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	110.351	1,00	1,00	-	121.769	-	-	1,00	-	-	-	-	-	1,00	-

<sup>a</sup> Esta tabla incluye únicamente observaciones que han sido imputadas como pertenecientes al conflicto armado, y como desapariciones forzadas (IS CONFLICT e IS FORCED DIS).

<sup>b</sup> Es importante aclarar que las columnas: inf , med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(media) media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable SEXO como de la variable IS CONFLICT

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

Tabla 8: Sexo de la víctima homicidio (1985–2018)<sup>a</sup>

Sexo	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf )	N(est)	N(sup )	P(inf )	P(est )	P(sup )
HOMBRE	333.633	0,89	0,91	396.135	408.680	421.225	0,88	0,91	0,93	684.144	716.851	752.722	0,88	0,92	0,97
MUJER	33.509	0,09	0,09	40.178	41.986	43.794	0,09	0,09	0,10	54.390	60.692	69.424	0,07	0,08	0,09
Obs. con NAs	7.425	0,02	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	374.567	1,00	1,00	-	450.666	-	-	1,00	-	-	-	-	-	1,00	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado.

<sup>b</sup> Es importante aclarar que las columnas: inf , med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable SEXO como de la variable IS CONFLICT

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1



Tabla 9: Sexo de la víctima reclutamiento de niños, niñas y adolescentes (1990–2017)

Sexo	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf )	N(est)	N(sup )	P(inf )	P(est )	P(sup )
HOMBRE	8.489	0,58	0,69	10.977	11.314	11.651	0,68	0,7	0,72	16.270	21.262	29.915	0,55	0,72	1,00
MUJER	3.825	0,26	0,31	4.558	4.924	5.290	0,28	0,3	0,33	7.326	8.290	9.547	0,25	0,28	0,32
Obs. con NAs	2.355	0,16	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	14.669	1,00	1,00	-	16.238	-	-	1,0	-	-	-	-	-	1,00	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado.

<sup>b</sup> Es importante aclarar que las columnas: inf , med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable SEXO como de la variable IS CONFLICT

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

Tabla 10: Sexo de la víctima secuestro (1990–2018)

Sexo	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf )	N(est)	N(sup )	P(inf )	P(est )	P(sup )
HOMBRE	35.526	0,77	0,78	38.989	39.479	39.969	0,77	0,78	0,79	53.484	59.415	67.296	0,69	0,76	0,87
MUJER	10.033	0,22	0,22	11.086	11.291	11.496	0,22	0,22	0,23	16.514	18.309	20.570	0,21	0,24	0,26
Obs. con NAs	760	0,02	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	46.319	1,01	1,00	-	50.770	-	-	1,00	-	-	-	-	-	1,00	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado.

<sup>b</sup> Es importante aclarar que las columnas: inf , med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable SEXO como de la variable IS CONFLICT

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

Finalmente, presentamos los hallazgos por actor armado. Lo primero que se evidencia es que los paramilitares son los principales responsables tanto de desaparición forzada como de homicidio, mientras que las FARC son las principales responsables de secuestro y reclutamiento de niños, niñas y adolescentes. Al hacer el agregado, se evidencia que estos dos grupos armados ocupan el primer y segundo lugar en cada uno de los hechos de violencia y los rangos de las estimaciones no se superponen. Es decir, para desaparición forzada y homicidio los paramilitares fueron el principal actor armado, seguido por las FARC. Mientras que para reclutamiento de niños, niñas y adolescentes y secuestro las FARC fueron el principal actor armado, seguido por los paramilitares. Se debe notar que el orden relativo de los responsables es consistente entre los datos observados, la imputación estadística y la estimación del población de víctimas total, incluso las no documentadas.

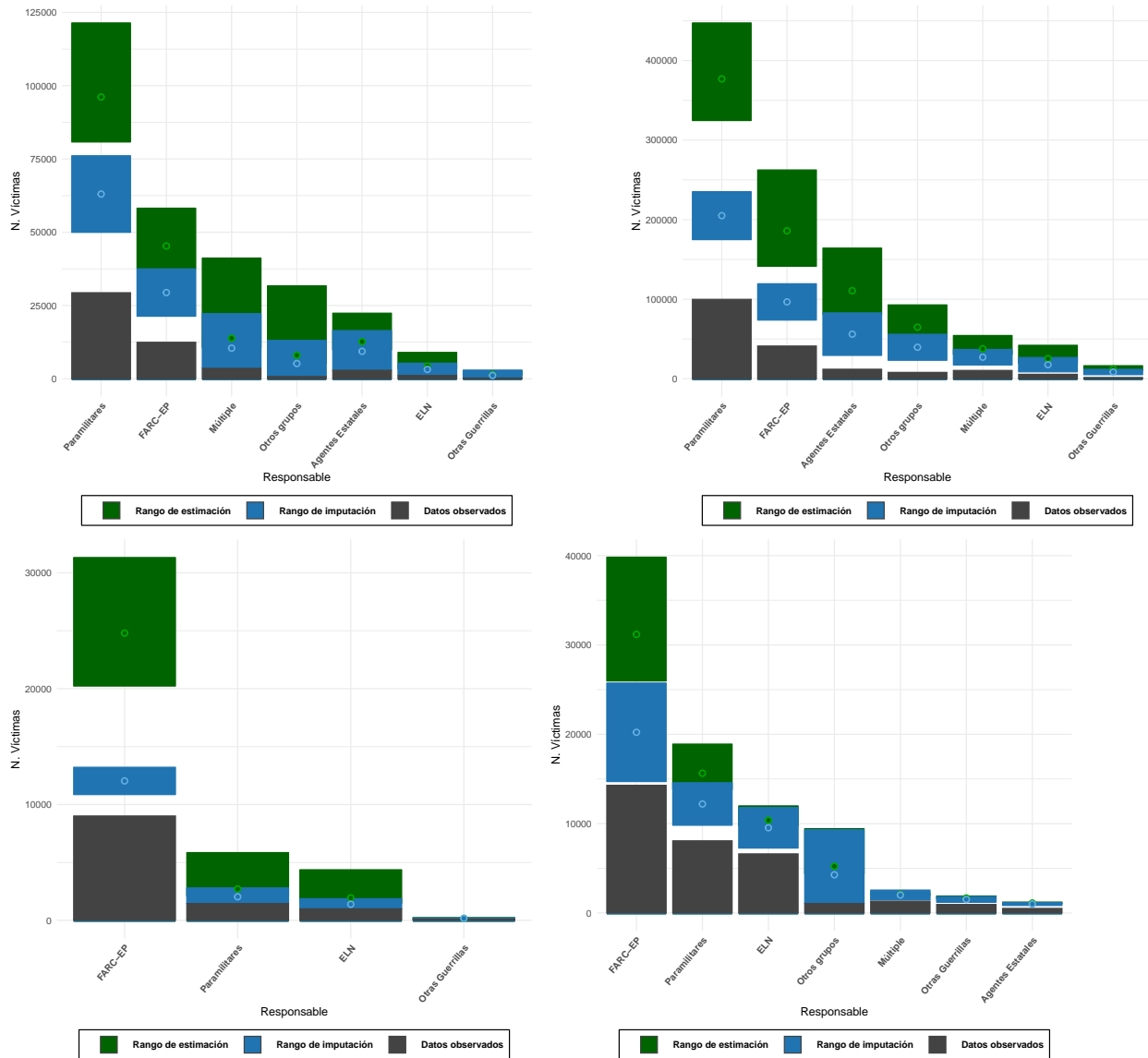


Figura 4: Presunto responsable para desaparición forzada (sup - izq), homicidio (sup - der), reclutamiento de niños, niñas y adolescentes (inf - izq) y secuestro (inf - der) para víctimas observadas, imputadas estadísticamente y estimadas

En las siguientes gráficas, al analizar la línea de tiempo de cada uno de los hechos por presunto responsable, llama la atención que parecería haber una relación directa y positiva entre los distintos actores armados.

Es decir, siguen tendencias similares a lo largo del tiempo. Para homicidio se evidencia un pico de violencia cometida presuntamente por agentes estatales en 2007, mientras que la violencia de los paramilitares y las FARC estaba en descenso. En cuanto a reclutamiento de niños, niñas y adolescentes, el pico de violencia de las FARC se habría dado previo al de los paramilitares, ELN y otras guerrillas. Aunque la varianza es amplia, parecería que hubo un pico por parte de los paramilitares y las FARC entre el 2010 y 2013. Finalmente, en secuestro se observan unos picos en los años 90 por las FARC y el ELN. Desde el 96 se evidencia un aumento en los casos, que alcanza el pico a inicios del 2000 para distintos actores. Este es el hecho de violencia en el que más se ve el papel de los otros actores, que tienen una alta varianza, pero están presentes en los distintos años. Como fue el caso con otros análisis, el patrón de los datos observados sigue el patrón de la imputación estadística y la estimación del población de víctimas entero.

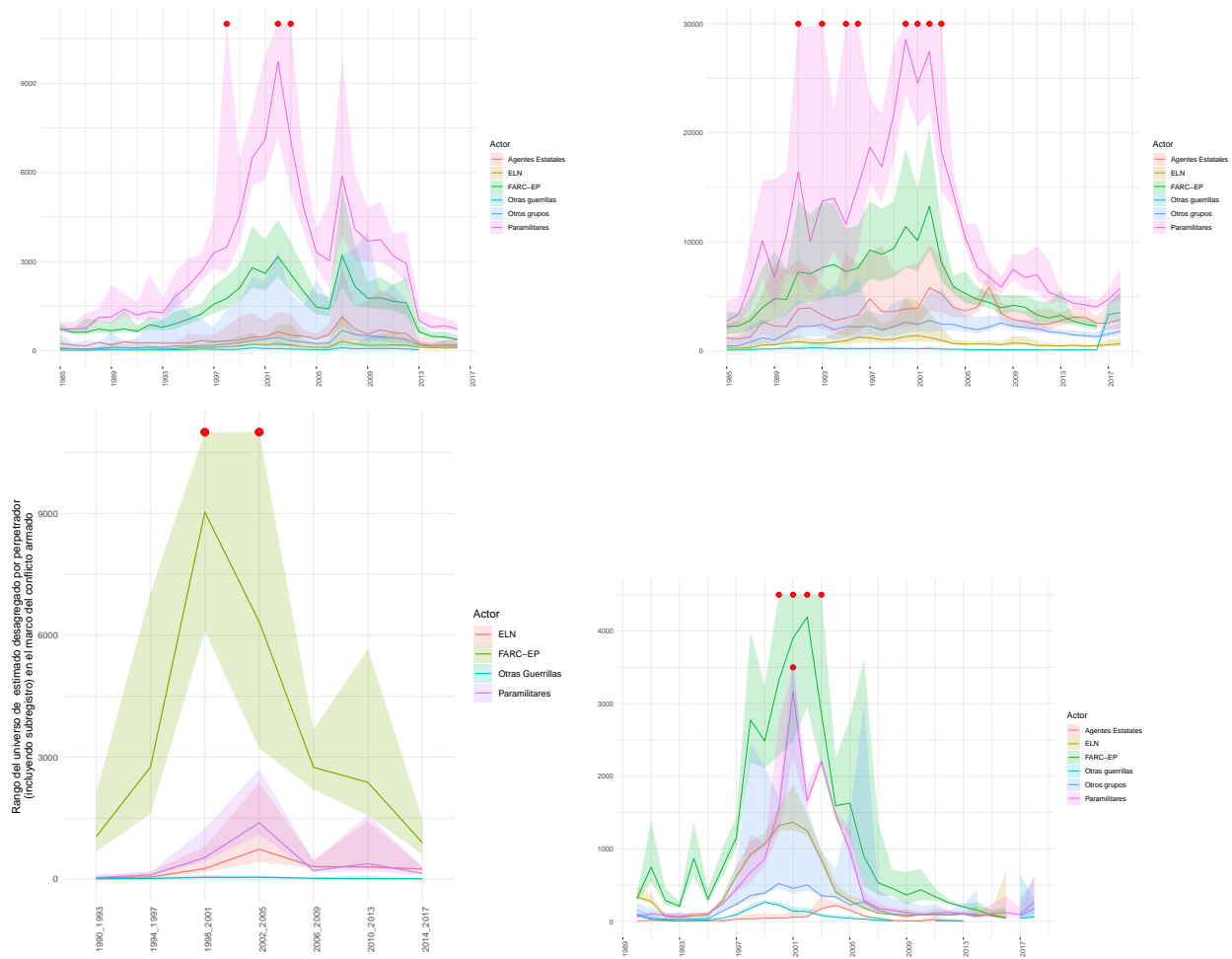


Figura 5: Estimaciones para desaparición forzada (sup-izq), homicidio (sup-der), reclutamiento de niños, niñas y adolescentes (inf-izq) y secuestro (inf-der) desagregados por presunto responsable

En las siguientes tablas se observan las presuntas responsabilidades por actor. Al hacer la integración de datos, en desplazamiento las guerrillas sin especificar un grupo serían las principales actores con una proporción de 0,51 si se consideran solo los registros con responsable, seguidas por paramilitares con 0,37.

Es interesante que para desapariciones forzadas y homicidio la proporción de responsabilidades observada es, respectivamente, 0,58 y 0,56 para paramilitares, 0,25 y 0,23 para FARC y 0,06 y 0,07 para agentes estatales. Es decir, la proporción en estos dos hechos en cuanto a datos observados es similar. En cambio, para reclutamiento de niños, niñas y adolescentes la responsabilidad de FARC es 0,76, de paramilitares 0,12 y de ELN 0,08. Mientras que para secuestro la responsabilidad de FARC es de 0,44, de paramilitares 0,25 y de ELN 0,2. Por su parte, para desplazamiento las otras guerrillas son las presuntas responsables en una proporción de 0,51, seguidas por paramilitares con 0,37.

Los datos observados muestran la importancia de utilizar la estadística. En las tablas es claro el alto porcentaje de datos sin información del presunto responsable. Este es un campo faltante de manera crítica en desplazamiento (0,67), desaparición forzada (0,54) y homicidio (0,52), mientras que en secuestro alcanza 0,29.

En cuanto a la imputación estadística para desaparición forzada, el principal responsable continuarían siendo los paramilitares (0,41–0,62) con más del doble que las FARC (0,18–0,31). Al hacer la estimación para desaparición forzada, el principal responsable continuarían siendo los paramilitares (0,44–0,67) con más del doble que las FARC (0,21–0,32).

Vale la pena notar cómo aumenta la varianza en las responsabilidades al estimar. Por ejemplo, la responsabilidad del múltiples actores y Estado aumentan entre los datos observados y estimados, pasando de 0,07 a entre 0,06 y 0,23 y entre 0,05 y 0,12, respectivamente. Por su parte, la de “otros grupos” aumenta de 0,02 a entre 0,03 y 0,17.

En cuanto al análisis de presunta responsabilidad de homicidio con datos imputados estarían los paramilitares con (0,39–0,52), las FARC con (0,16–0,27) y el Estado con (0,07–0,18). Mientras que al hacer la estimación tenemos que los paramilitares serían los principales responsables (0,4–0,55), seguidos por las FARC (0,17–0,32) y el Estado (0,1–0,2).

Hay dos puntos para notar. Primero, el orden de los grupos responsables de mayor a menor es igual para la imputación estadística y la estimación de la población de todas las víctimas: primero los paramilitares, seguidos por las FARC y agentes de Estado. Ni la imputación estadística ni la estimación cambia este hallazgo principal. Segundo, comparando los intervalos de la imputación estadística y la estimación, observamos que los intervalos no se cruzan. Por ejemplo, el intervalo de la imputación estadística para los paramilitares no se cruza con el intervalo de las FARC. Esto significa que las dos estimaciones son estadísticamente significativamente diferentes, es decir, hay evidencia estadística para afirmar que son distintos.

Respecto al reclutamiento, las FARC son el principal actor tanto en los datos observados como imputados estadísticamente y estimados. Mientras que su responsabilidad en los datos observados es 0,76. Al hacer la imputación estadística la proporción oscilaría entre 0,67 y 0,82. Al hacer la estimación la proporción oscilaría entre 0,67 y 1. Les seguirían los paramilitares con entre 0,07 y 0,19.

Frente al secuestro, nuevamente serían las FARC el principal responsable. En los datos observados serían responsables del 0,44. Mientras que en los datos imputados estadísticamente su responsabilidad sería de entre 0,29 y 0,51. En cuanto a la estimación, la responsabilidad de las FARC tendría un rango de entre 0,39 y 0,59, seguido por los paramilitares con entre 0,21 y 0,28, y el ELN con entre 0,15 y 0,18.

Tabla 11: Presunto responsable desplazamiento forzado (1985–2019)<sup>a</sup>

Responsable	Observado		
	N.	P(NA)	P
Obs. con NAs	5.208.856	0,67	-
Guerrillas sin especificar	1.309.028	0,17	0,51
Paramilitares	947.073	0,12	0,37
Múltiple	126.922	0,02	0,05
Otros grupos	114.883	0,01	0,05
FARC-EP	37.542	0,00	0,01
Agentes Estatales	7.326	0,00	0,00
ELN	1.334	0,00	0,00
Total	7.752.964	0,99	0,99

<sup>a</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1.

<sup>b</sup> Como se mostró en los patrones de documentación, la mayoría de los registros de desplazamiento forzado provienen del RUV. En muchos casos se tiene guerrilla sin especificar

Tabla 12: Presunto responsable desaparición forzada (1985–2016)<sup>a</sup>

Responsable	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf )	N(est)	N(sup )	P(inf )	P(est )	P(sup )
Paramilitares	29.368	0,27	0,58	49.960	63.029	76.098	0,41	0,52	0,62	80.789	96.131	121.418	0,44	0,53	0,67
FARC-EP	12.451	0,11	0,25	21.330	29.410	37.490	0,18	0,24	0,31	37.643	45.301	58.213	0,21	0,25	0,32
Múltiple	3.566	0,03	0,07	3.566	10.448	22.239	0,00	0,09	0,18	10.523	13.844	41.199	0,06	0,08	0,23
Agentes Estatales	2.907	0,03	0,06	2.907	9.359	16.442	0,02	0,08	0,14	9.888	12.662	22.384	0,05	0,07	0,12
Otros grupos	770	0,01	0,02	770	5.200	13.089	0,00	0,04	0,11	5.284	8.009	31.723	0,03	0,04	0,17
ELN	1.120	0,01	0,02	1.120	3.194	5.320	0,01	0,03	0,04	3.288	4.304	8.993	0,02	0,02	0,05
Otras Guerrillas	275	0,00	0,01	275	1.128	2.949	0,00	0,01	0,02	1.143	1.437	2.788	0,01	0,01	0,02
Obs. con NAs	59.894	0,54	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	110.351	1,00	1,01	-	121.768	-	-	1,01	-	-	-	-	-	1,00	-

<sup>a</sup> Esta tabla incluye únicamente observaciones que han sido imputadas como pertenecientes al conflicto armado, y como desapariciones forzadas (IS CONFLICT e IS FORCED DIS).

<sup>b</sup> Es importante aclarar que las columnas: inf , med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable RESPONSABLE como de las variable IS CONFLICT e IS FORCED DIS.

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

<sup>e</sup> La estimación por responsable tiene un pequeño sesgo hacia abajo.

Tabla 13: Presunto responsable homicidio (1985–2018)

Responsable	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf )	N(est)	N(sup )	P(inf )	P(est )	P(sup )
Paramilitares	99.837	0,27	0,56	174.867	205.028	235.189	0,39	0,45	0,52	324.602	376.928	447.276	0,40	0,46	0,55
FARC-EP	41.161	0,11	0,23	73.746	96.592	119.438	0,16	0,21	0,27	141.374	185.934	262.475	0,17	0,23	0,32
Agentes Estatales	12.008	0,03	0,07	29.364	56.094	82.824	0,07	0,12	0,18	80.926	110.652	164.263	0,10	0,14	0,20
Otros grupos	8.048	0,02	0,05	23.252	39.606	55.960	0,05	0,09	0,12	50.265	64.799	92.879	0,06	0,08	0,11
Múltiple	10.508	0,03	0,06	17.347	27.124	36.901	0,04	0,06	0,08	30.503	37.695	54.384	0,04	0,05	0,07
ELN	5.635	0,02	0,03	8.668	17.725	26.782	0,02	0,04	0,06	19.482	25.306	42.107	0,02	0,03	0,05
Otras Guerrillas	1.473	0,00	0,01	5.097	8.496	11.895	0,01	0,02	0,03	10.233	12.391	16.439	0,01	0,02	0,02
Obs. con NAs	195.897	0,52	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	374.567	1,00	1,01	-	450.665	-	-	0,99	-	-	-	-	-	1,01	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado.

<sup>b</sup> Es importante aclarar que las columnas: inf , med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable RESPONSABLE como de la variable IS CONFLICT

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

Tabla 14: Presunto responsable reclutamiento de niños, niñas y adolescentes (1990–2017)

Responsable	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
FARC-EP	9.013	0,61	0,76	10.857	12.038	13.219	0,67	0,75	0,82	20.227	24.798	31.322	0,67	0,82	1,00
Grupos Paramilitares	1.457	0,10	0,12	1.457	2.038	2.804	0,09	0,13	0,17	2.101	2.714	5.851	0,07	0,09	0,19
ELN	997	0,07	0,08	997	1.391	1.876	0,06	0,09	0,12	1.441	1.938	4.373	0,05	0,06	0,14
Otras Guerrillas	116	0,01	0,01	131	174	217	0,01	0,01	0,01	142	172	242	0,00	0,01	0,01
Obs. con NAs	2.777	0,19	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	14.636	0,99	1,00	-	16.129	-	-	1,00	-	-	-	-	-	1,00	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado.

<sup>b</sup> Es importante aclarar que las columnas: inf, med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable RESPONSABLE como de la variable EDAD

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

<sup>e</sup> El total observado no es igual a la suma de los presuntos responsables porque algunos casos tenían como presunto responsable al Estado. Estos no fueron tenidos en cuenta.

Tabla 15: Presunto responsable secuestro (1990–2018)<sup>a</sup>

Responsable	Observado			Imputado						Estimado					
	N.	P(NA)	P	N(inf)	N(imp)	N(sup)	P(inf)	P(imp)	P(sup)	N(inf)	N(est)	N(sup)	P(inf)	P(est)	P(sup)
FARC-EP	14.291	0,31	0,44	14.709	20.223	25.737	0,29	0,40	0,51	25.968	31.180	39.826	0,39	0,46	0,59
Paramilitares	8.077	0,17	0,25	9.818	12.200	14.582	0,19	0,24	0,29	13.825	15.644	18.905	0,21	0,23	0,28
ELN	6.618	0,14	0,20	7.264	9.538	11.812	0,14	0,19	0,23	9.774	10.378	11.988	0,15	0,15	0,18
Otros grupos	1.059	0,02	0,03	1.059	4.276	9.323	0,00	0,08	0,18	4.380	5.235	9.440	0,07	0,08	0,14
Múltiple	1.317	0,03	0,04	1.450	2.009	2.568	0,03	0,04	0,05	2.029	2.076	2.178	0,03	0,03	0,03
Otras Guerrillas	958	0,02	0,03	1.195	1.519	1.843	0,02	0,03	0,04	1.569	1.673	1.903	0,02	0,02	0,03
Agentes Estatales	514	0,01	0,02	840	1.005	1.170	0,02	0,02	0,02	1.053	1.114	1.222	0,02	0,02	0,02
Obs. con NAs	13.485	0,29	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	46.319	0,99	1,01	-	50.770	-	-	1,00	-	-	-	-	-	0,99	-

<sup>a</sup> Esta tabla incluye únicamente las víctimas vinculadas al conflicto armado.

<sup>b</sup> Es importante aclarar que las columnas: inf, med y sup, indican el intervalo de confianza, es decir, inf: límite inferior (ya sea de la imputación o de la estimación); P(imp): media de la imputación y P(est): media de la estimación y sup: límite superior (ya sea de la imputación o la estimación.)

<sup>c</sup> La imputación incluye valores faltantes tanto de la variable RESPONSABLE como de la variable IS CONFLICT

<sup>d</sup> La fila con el total evidencia las columnas que se pueden sumar y cuyo resultado debe ser igual a 1

### 3. Vinculación de registros

Dado que el insumo del proyecto son distintas bases de datos y que una misma víctima puede estar registrada en más de una de ellas, es necesario hacer una vinculación de los registros. A este proceso también se le conoce como “deduplicación” y consiste en dejar un solo registro de cada víctima para evitar dobles conteos. Un resumen de la extensa investigación sobre este tema está disponible en [Herzog, Scheuren, and Winkler \(2007\)](#).

El proceso que llevamos a cabo para deduplicar dentro del proyecto sigue los pasos descritos en [Christen \(2012\)](#). En específico, iniciamos con la importación de datos. Segundo, generamos bloques de registros y, tercero, generamos características. Cuarto, hacemos un modelo por pares. Quinto, agrupamos. Sexto, unimos y exportamos el resultado. Por último, hacemos una auditoría para probar que el resultado de la integración sea correcto. A continuación explicamos cada uno de estos pasos en detalle.

#### 3.1. Preparación de datos

Si bien muchos de los campos de las bases de datos descritas en el anexo son compartidos, es necesario hacer la tarea de estandarización. Por ejemplo, es posible que algunas bases tengan los departamentos en mayúsculas, otros en minúsculas, algunos con tildes, etc. También, los nombres de las variables pueden ser distintos. Por ejemplo, algunas pueden referirse al código del municipio como *divipola*, otras como *coddane*, etc. Con la estandarización de datos mantenemos la información original, pero hacemos los cambios necesarios en los campos para que sean los mismos entre bases. Es decir, que las variables que se refieren a lo mismo tengan un mismo nombre y la información en un mismo formato.

Una vez los datos están estandarizados el siguiente paso para poder hacer la deduplicación es tener la información mínima necesaria para identificar si dos registros corresponden o no a la misma persona. Es importante notar que no todos los registros tienen los campos completos. Por ejemplo, aunque podría pensarse que una solución sería hacer la unión por cédula, este dato no siempre está disponible. Entonces, debemos definir cuál es la información mínima necesaria para identificar si dos registros corresponden o no a una misma víctima. En este caso, definimos como campos mínimos al nombre de la víctima, apellido de la víctima, departamento del hecho, y año del hecho. Además, el hecho registrado debe ser homicidio, desaparición, secuestro, reclutamiento de niños, niñas y adolescentes, desplazamiento forzado y/o exilio. Entonces, nos quedamos únicamente con los registros que tengan la información mínima necesaria para poder hacer la deduplicación y que sean víctimas directas. Al finalizar este paso el archivo consolidado a partir de todos los conjuntos de datos contiene 12.863.977 registros.

#### 3.2. Generación de bloques de registros

Intuitivamente el objetivo de la vinculación de registros consiste en analizar cada uno de los posibles pares de registros y decidir si se tratan o no de la misma persona. A esto se le denomina ser “correferentes.” Sin embargo, hay demasiados registros para comparar todos los pares posibles. Tenemos  $\binom{n}{2}$  posibles combinaciones. Como en este caso nuestro  $n$  es 12.863.977 registros, necesitamos generar trillones de pares de registros. Esta cifra es aproximadamente el número de células en el cuerpo humano y es muchos billones de veces mayor de lo que un supercomputador puede procesar. Para reducir el problema a un tamaño tratable debemos comenzar identificando solo los pares que tienen alguna posibilidad de ser correferentes, por lo cual describiremos la generación de datos de entrenamiento.

##### 3.2.1. Generación de datos de entrenamiento

Dado que no es posible ni para un humano ni para un computador analizar todos los posibles pares de registros, generamos bloques con base en distintos criterios y a partir de estos identificamos los registros que son similares y podrían corresponder a una misma persona para tener los datos de entrenamiento.



Los datos de entrenamiento en este caso fueron generados con base en la experticia humana. Específicamente, a esta persona se le llama el “oráculo” (Christen, Vatsalan, and Wang 2015). Para este proyecto, el oráculo fue la experta en vinculación de registros de HRDAG, Michelle Dukich, quien revisó cientos de bloques a partir de los cuales se generaron los pares de registros.<sup>5</sup>

Se seleccionaron grupos de 50 a 200 registros que comparten algunas características para reducir la magnitud de la tarea. Por ejemplo, registros de personas que tengan como primer nombre “MARÍA,” que hayan sufrido alguno de los hechos de violencia en Antioquia en marzo de 2001; o registros de personas que sufrieron homicidio, cuyo primer apellido comienza por “LOP.” El oráculo separó estos registros en grupos de registros que, según su criterio de experta, se refieren a la misma persona. Esto lo hizo basada en la intuición humana y contemplando el panorama completo. Es decir, no se limitó a que los nombres sea iguales, sino que incluyó un análisis de los nombres, lugar de los hechos, fecha, etc.

A partir de estos grupos, se extrajeron todos los pares de registros que potencialmente se refieren a la misma persona, llamados “pares positivos.” En total se identificaron 1.082.479 pares positivos y 1.717.201 pares negativos. Estos son los datos de entrenamiento para el modelo de generación de bloques.

Una posible preocupación sería que los datos de entrenamiento estén equivocados. Para probar esto le pedimos en diciembre de 2021 a 5 integrantes del Grupo de Análisis de la Información en la JEP que hicieran el ejercicio del oráculo.<sup>6</sup> Las decisiones tomadas por cada uno de ellos fueron comparadas con la asignación que realizó el oráculo. Los resultados mostraron que el acuerdo intersubjetivo (*intersubjective agreement* en inglés), es decir qué tanto coincidieron con el oráculo, fue superior a 0,9 en 4 de los casos.<sup>7</sup> Al analizar los registros en los que había diferencias, se encontró que el oráculo había identificado a pares de personas que los analistas de la JEP no. Los analistas reconocieron que en estos casos sí se trataría de la misma persona.

### 3.2.2. Generación de bloques

El modelo de bloques utiliza los datos de entrenamiento (aquellos entregados por el oráculo) para encontrar cuáles son las “reglas” que definen si dos registros corresponden o no a la misma persona. Es decir, aprender a imitar las decisiones de la experta. Encontrar estos bloques de potenciales pares es un problema desafiante en estadística y ciencias de la computación.<sup>8</sup>

A partir de los datos de entrenamiento se generan las “reglas” para tener bloques de registros que son muy similares entre sí. Por ejemplo, porque comparten las tres letras iniciales del apellido, el hecho de violencia es el mismo y ocurrió en la misma fecha. Esta regla combina tres campos (apellido, hecho de violencia y fecha), que deben coincidir para que un par de registros corresponda potencialmente a la misma persona. A esta combinación de campos se le conoce como *conjunción lógica*.

Por supuesto, cada regla cubre por sí misma solo unos pocos de todos los posibles pares. Por lo tanto, se adiciona más de una regla. Al hecho de tener más de una regla se le conoce como *disyunción lógica*. Las reglas se adicionan hasta que dentro de la disyunción lógica se encuentran todos (o casi todos) los pares positivos de los datos de entrenamiento. Es decir, hasta que la mayoría de los pares positivos que fueron identificados por el oráculo en los datos de entrenamiento son identificados con las reglas.

La tarea de generación de reglas debe hacerse de manera cuidadosa, pues de lo contrario podrían incluir a demasiados registros. Por ejemplo, si consideramos pares definidos por tener el mismo primer nombre y primer apellido, habría más de 1,5 billones de pares posibles para una capacidad computacional de aproximadamente 30 millones de pares. Entonces, se debe encontrar un balance entre reglas que incluyan al mayor número de posibles pares de manera eficiente para ser analizados con la capacidad computacional. La disyunción final de conjunciones se muestra abajo.

---

<sup>5</sup>Para más detalles ver [esta presentación](#).

<sup>6</sup>Analizaron 230.582, 185.187, 409.301, 197.983 y 354.279, pares de registros, respectivamente.

<sup>7</sup>El acuerdo intersubjetivo se calcula con el Kappa de Cohen, que toma 1 como valor máximo. Los Kappas fueron 0,968, 0,975, 0,972, 0,971 y 0,763. Con la quinta persona se alcanzó 0,763, pero se reconoció que no había realizado el ejercicio con suficiente dedicación.

<sup>8</sup>Para una explicación más amplia del modelo de bloques (a veces llamado indexación) ver Christen (2012), Capítulo 4, Herzog, Scheuren, and Winkler (2007), Bilenko, Kamath, and Mooney (2006), Michelson and Knoblock (2006) y este [blog](#) por HRDAG.

```

(apellido_1_meta1 ^ nombre_1_meta2 ^ synth_hecho_ymd) ∨
(cens_apellido_1 ^ cens_nombre_1 ^ nombre_2 ^ sort_nm_1) ∨
(dept_hecho_desaparicion ^ desaparicion_ymd ^ sort_nm_1 ^ sort_nm_1_last_3) ∨
(cens_sort_nm_1 ^ dept_hecho_homicidio ^ homicidio_ymd) ∨
(muni_hecho_reclutamiento ^ nombre_2_last_3 ^ sexo ^ sort_nm_n ^ sort_nm_n_meta2) ∨
(dept_hecho_homicidio ^ homicidio_yrmo ^ sort_nm_n_meta1 ^ sort_nm_n_meta2) ∨
(secuestro_ymd) ∨
(apellido_1 ^ homicidio_ymd) ∨
(cens_nombre_1 ^ desaparicion_yrmo) ∨
(apellido_2_first_3 ^ nombre_2_last_3 ^ reclutamiento_yr ^ sort_nm_n_meta2) ∨
(homicidio_ymd ^ nombre_2_first_3) ∨
(apellido_2_first_3 ^ desaparicion_ymd ^ etnia ^ nombre_1_first_3 ^ nombre_1_last_3 ^ nombre_2_meta2) ∨
(cens_sort_nm_1 ^ sort_nm_n_meta2 ^ synth_hecho_ymd) ∨
(desaparicion_ymd ^ dob_ym ^ nombre_2_meta2) ∨
(apellido_1 ^ cens_apellido_1 ^ nombre_2 ^ secuestro_yrmo ^ sort_nm_1_meta1) ∨
(apellido_2_first_3 ^ desaparicion_ymd ^ nombre_2_meta1) ∨
(apellido_2_last_3 ^ cens_dob_ymd ^ etnia ^ muni_hecho_desaparicion ^ nombre_2_first_3 ^ nombre_2_last_3) ∨
(apellido_1_meta1 ^ muni_hecho_desaparicion ^ nombre_2_meta2) ∨
(cens_nombre_1 ^ homicidio_ymd ^ nombre_1_last_3) ∨
(apellido_2 ^ cens_dob_ym ^ dob_ymd ^ nombre_2 ^ sort_nm_n) ∨
(cens_apellido_1 ^ cens_nombre_1 ^ cens_sort_nm_1 ^ homicidio_yrmo) ∨
(muni_hecho_secuestro ^ nombre_2_first_3 ^ sort_nm_n) ∨
(apellido_2 ^ cens_apellido_1 ^ cens_sort_nm_n ^ nombre_2 ^ sort_nm_1) ∨
(cens_sort_nm_n ^ desaparicion_yrmo ^ sort_nm_1) ∨
(apellido_1 ^ cens_apellido_1 ^ desaparicion_ymd) ∨
(apellido_1_last_3 ^ desaparicion_ymd ^ nombre_2) ∨
(cens_apellido_1 ^ homicidio_yrmo) ∨
(apellido_1_first_3 ^ sort_nm_1 ^ synth_hecho_ymd ^ synth_muni_code) ∨
(apellido_2 ^ cens_nombre_1 ^ desplazamiento_yr ^ nombre_2 ^ sort_nm_1) ∨
(apellido_2_meta1 ^ homicidio_ymd) ∨
(muni_hecho_desaparicion ^ sort_nm_1_first_3 ^ sort_nm_n) ∨
(apellido_1_first_3 ^ apellido_2_last_3 ^ cens_nombre_1 ^ homicidio_yrmo) ∨
(apellido_1 ^ cens_dept_hecho_homicidio ^ cens_nombre_1 ^ muni_hecho_homicidio) ∨
(cens_sort_nm_n ^ desaparicion_ymd) ∨
(apellido_1_last_3 ^ desaparicion_ymd ^ synth_muni_code) ∨
(apellido_1_last_3 ^ apellido_2_meta2 ^ desaparicion_yrmo ^ sexo) ∨
(synth_id) ∨
(apellido_1 ^ apellido_2 ^ nombre_1) ∨
(apellido_1 ^ dob_ymd ^ nombre_1) ∨
(nac_sorted)

```

Para entender las reglas, vale la pena tomar algunos ejemplos. La última regla (`nac_sorted`) considera todos los pares de registros que comparten el primer y segundo nombre, y primer y segundo apellido, con los *tokens* ordenados. Por ejemplo, si hubiese una víctima Juan Gabriel Pérez Ramírez, lo ordenamos como Pérez Ramírez Juan Gabriel. Mientras que Ana Rodríguez sería Rodríguez NA Ana NA. Ahora bien, la penúltima regla considera aquellos con primer nombre y primer apellido iguales, así como con la misma fecha de nacimiento. Por su parte, la regla `synth_id`, considera todos los pares de registros que comparten el número de cédula o el número de algún otro documento de identificación.<sup>9</sup>

Aunque hay muchas reglas en esta lista, es necesario recordar que un par de registros tiene que satisfacer sólo una de estas reglas para ser incluido entre los pares candidatos. Cada vez que se encuentra un nuevo par positivo, se genera nuevamente la lista de reglas. Con el algoritmo de búsqueda, se abarca 99,7% de los pares positivos etiquetados. Vale la pena enfatizar que a partir de esto tenemos bloques de potenciales registros que pertenecen a una misma persona, pero no se ha estudiado si esos registros son efecto o no la misma persona. Hasta el momento hay entonces un total de 105.734.852 pares de registros a ser analizados.

<sup>9</sup>Muchos números de cédula fueron reportados como error. Por ejemplo, porque los números en un par de registros coinciden exactamente salvo por un carácter en la cédula. Del mismo modo, números de cédulas fueron reportados con frecuencia en el campo de “otro documento” y viceversa. También hay muchos registros que carecen completamente de información de cédula.

### 3.3. Generación de características

Una vez se tienen bloques más pequeños es necesario analizar los posibles pares de registros dentro de cada bloque. De nuevo, para una persona sería sencillo identificar si dos registros son o no correferentes. Por ejemplo, porque los nombres, hechos y fecha coincidan. También puede ser porque las fechas de nacimiento sean muy cercanas. O porque los nombres suenen similar, aunque se escriban distinto.<sup>1011</sup>

Se generaron 116 medidas de comparación (por ejemplo, comparar las 3 primeras letras del nombre) de un total de 45 campos (o variables). Ninguna de estas medidas es suficiente en sí misma, pero al combinarlas, proporcionan un punto de inicio suficiente para estimar la probabilidad que dos registros se refieran a la misma persona.

### 3.4. Modelo por pares

En esta etapa, creamos un modelo que predice una ponderación de que un par de registros se refiera a una misma persona. Las características generadas en la tarea anterior son las variables independientes del modelo, y la variable dependiente es tomada de los datos de entrenamiento que el oráculo etiquetó.

#### 3.4.1. Evaluación del modelo

En este caso la pregunta que se está resolviendo es: ¿los registros X y Y se refieren a la misma persona? Para responderla podemos usar un modelo de clasificación binaria, como un bosque aleatorio (*random forest* en inglés) o una regresión logística (James et al. 2013). Se ha encontrado que el algoritmo XGBoost (Chen and Guestrin 2016) consistentemente genera los mejores resultados para esta aplicación. XGBoost es un clasificador de regresión basado en árboles potenciados, y debido a que la variable a predecir es binaria, se decidió usar una función logística como nuestra función objetivo.

El modelo es entrenado con los datos de entrenamiento. Es decir, el algoritmo aprende las ponderaciones relativas (o importancia) de las características que mejor predicen las etiquetas positivas y negativas en el subconjunto de datos seleccionado para entrenar. Estas ponderaciones son llamadas por el modelo. Así, se utiliza el modelo para predecir la probabilidad de que cada par se refiera a la misma persona. La importancia relativa para las diez variables identificadas por el algoritmo como más informativas son mostradas en la Figura 6. Tenga en cuenta que el gráfico es ilustrativo para entender la importancia relativa de las variables, pero la escala no es proporcional a la misma. Es decir, `nombre_apellido_completo_1v` no es 7 veces más importante que `nac_sorted_jaccard3`, pero sí es más importante.

---

<sup>10</sup>Ver [explicación de wikipedia](#) sobre metáfonos.

<sup>11</sup>Ver explicaciones de wikipedia de la [distancia de Levenshtein](#) y la [distancia de Jaro-Winkler](#).

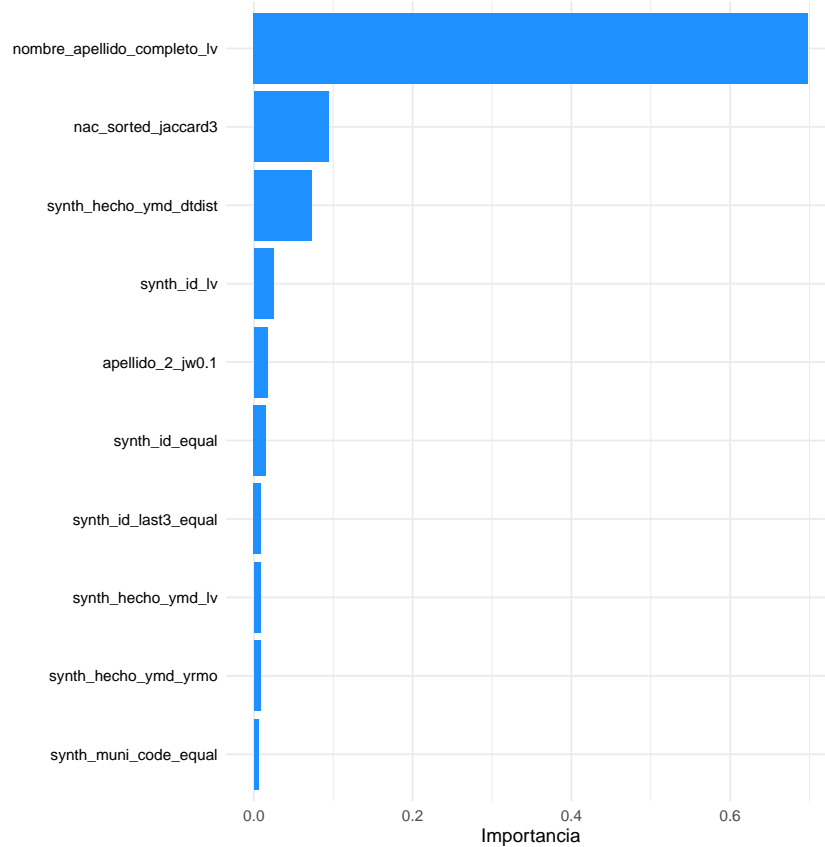


Figura 6: Top 10 de las variables más importantes en el ajuste del modelo con los datos de entrenamiento.

En este modelo, la mayoría del poder explicativo del mismo viene de la variable `nombre_apellido_completo_lv`. Esta variable mide la diferencia entre las cadenas de texto de los nombres completos del primer registro respecto al segundo en el par considerado (la diferencia en esta medida es expresada por el número de caracteres que se necesitan para transformar una cadena de texto a otra). Le sigue el índice de Jaccard, que mide la similitud entre dos conjuntos, del nombre y apellidos completo. En tercer lugar, la coincidencia de las fechas. También se incluye la similaridad de la cédula, apellido y municipio.

¿Qué tan buena es la clasificación? El modelo es testeado usando el conjunto de datos de prueba. Estos datos no fueron usados en el proceso de entrenamiento, y por tanto permiten medir el rendimiento del modelo sobre datos que nunca ha visto. A través de diferentes métricas de aprendizaje automático, este modelo se comporta muy bien. Las métricas  $F_1$ , recall, precisión, exactitud y el coeficiente de correlación de Matthews son superiores a 0,98.

Estas medidas se calculan para mostrar una evaluación completa del modelo, pero preferimos dos métodos gráficos para evaluar el modelo: separación y calibración, mostradas en la Figura 7.

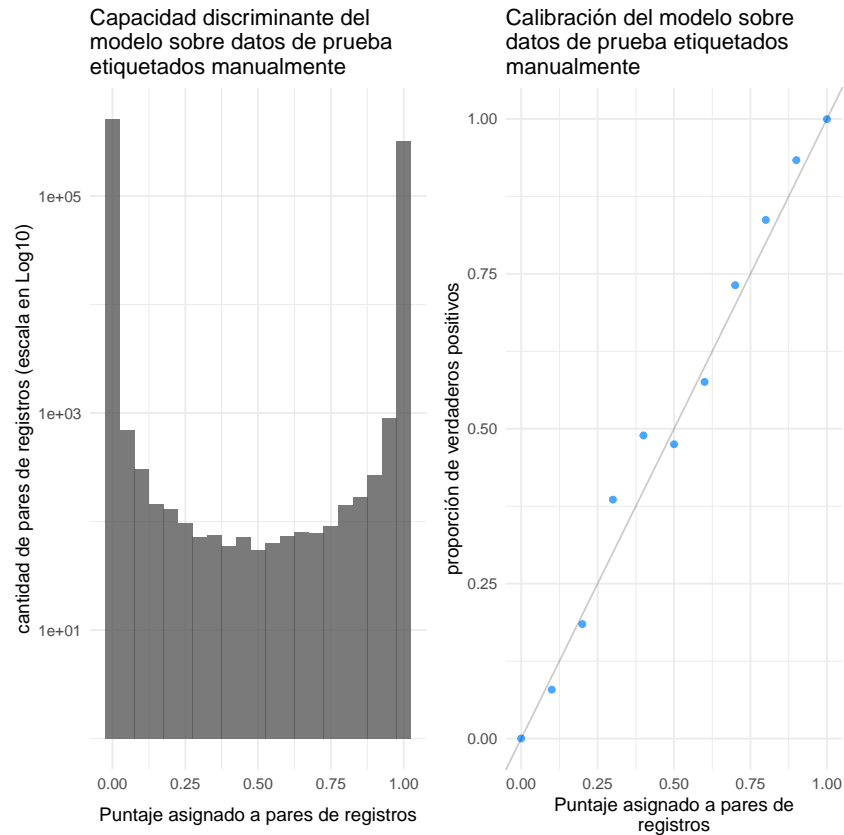


Figura 7: Evaluación del rendimiento del modelo a través de un análisis de separación de clases (0 para no coincidente y 1 para coincidente) y calibración.

El panel izquierdo de la Figura 7, la gráfica de separación, muestra el número de pares del conjunto de entrenamiento para cada uno de los valores predichos por el modelo. Vale la pena notar que el eje  $y$  está en escala logarítmica. Las dos barras más altas están en 0 y 1, lo cual indica que el modelo es capaz de discriminar bien la mayoría de pares, ya que le asigna 0 a los pares de registros que no se refieren a la misma persona o 1 en caso contrario. Por lo tanto, el modelo es confiable. Por su parte, para la mayoría de pares con una probabilidad entre 0 y 1 hay alrededor de 100 pares. Entonces, el modelo está realizando una buena separación.

El panel derecho de la Figura 7, la gráfica de calibración, muestra la extensión a la cual los valores predichos por el modelo son una representación exacta de la certeza o incertidumbre del modelo. La figura muestra la proporción de pares con cada valor que el oráculo etiquetó como positivos. Un valor de 0,5 indica que el modelo no tiene forma de saber si un par con este valor es positivo o negativo. En general, se espera que en un modelo bien calibrado, el puntaje asignado a pares de registros y la proporción de verdaderos positivos sean iguales, por lo cual se presenta la recta  $y = x$  en la gráfica. Allí se observa que los puntos están muy cerca a la línea indicando una buena calibración.

La combinación de una buena separación y una buena calibración significa que el modelo está reproduciendo las decisiones del oráculo. Es decir, el modelo aprendió a imitar las decisiones humanas sobre si un par de registros son o no la misma persona. Una vez se tiene el modelo entrenado, se aplica al conjunto entero de datos de 105.734.852 candidatos de pares de registros, y se guardan los valores de las probabilidades para la tarea de agrupamiento.

### 3.5. Agrupamiento

Para los datos de entrenamiento seleccionamos componentes conectadas en el primer paso del método de agrupamiento para que fueran revisadas por el oráculo. De forma similar al trabajo hecho en el modelo de bloques, en este paso el oráculo revisó agrupaciones de registros,<sup>12</sup> como aquellos mostrados en el dendrograma. Las agrupaciones son componentes conectadas, y por tanto el oráculo las separa en grupos de coincidencia que se refieren a la misma persona. Esta forma de generar datos de entrenamiento permite generar muchos pares positivos (cada par en un grupo de coincidencia es un par positivo) y muchos pares negativos (cada par de registros en una componente conectada que no está en el mismo grupo de coincidencia es un par negativo). Este es un uso eficiente de la cantidad limitada de datos de entrenamiento.

En este momento hemos identificado si cada par de registros dentro de cada bloque corresponde o no a la misma persona. Sin embargo, una misma persona podría estar en dos pares de registros o más. Entonces, es necesario agrupar a los pares de registros que correspondan a una misma persona.

El agrupamiento consiste en dos pasos. Primero, se agrupan los registros que tienen alguna conexión, considerando una probabilidad de coincidencia muy baja (hemos usado  $cc_t = 0,2$ ). Si un registro A tiene un vínculo con un registro B  $\geq 0,2$ , y B tiene un vínculo con un registro C, entonces A, B, y C forman un “componente conectado” (ver Hagberg, Schult, and Swart (2008)), usando el lenguaje de la teoría de grafos. Este paso separa nuestros registros en particiones relativamente pequeñas de agrupaciones posibles.

La vinculación de registros se logra gracias a una serie de criterios en el algoritmo. Por ejemplo, comparar las tres primeras letras de los registros. Sin embargo, el modelo no siempre tiene certeza sobre si un par de registros pertenecen o no a la misma persona y, en muchos casos, los asigna con una probabilidad cercana a 0,5. De acuerdo a la teoría de la percolación (ver Berchenko et al. (2009) y Ming Li and Zhang (2021)), los posibles pares eventualmente se fusionan y se conectan en una estructura enorme, conocida como grupo de expansión. Entonces, en algún momento al tener tantos pares que “tal vez” son la misma persona, la cantidad de bloques crece exponencialmente. Sin embargo, dadas las limitaciones computacionales, solo se puede considerar un número limitado de bloques.

Después, se separa cada agrupación usando un [agrupamiento jerárquico aglomerativo](#), por medio del algoritmo y el software `fastcluster` (Müllner 2013). Este método organiza los registros dentro de una agrupación en subagrupaciones más pequeñas que sean cercanas entre sí.

Luego es posible particionar las subagrupaciones en grupos de registros coincidentes que se refieren a una persona. Hemos usado un vínculo “promedio” con un umbral de vínculo “plano” de  $f_t = 0,65$ .<sup>13</sup> El hiperparámetro  $f_t$  es una distancia, no una probabilidad de coincidencia, así este es el complemento  $(1 - p)$  de los puntajes arrojados por el modelo. Así, si el modelo indicó un puntaje de 0,8 entre dos pares, que al acercarse a 1 indica que es altamente probable que este par se refiera a la misma persona, la distancia entre ellos para el agrupamiento será de 0,2.

El método de agrupamiento se entiende mejor con una visualización, por lo cual se presenta una estas agrupaciones abajo en una gráfica denominada “[dendrograma](#)”. La Figura 8 muestra la distancia entre pares de registros y cómo las subagrupaciones son similares entre ellas. Entonces, el eje horizontal muestra la distancia entre subagrupaciones, mientras que las líneas muestran la distancia requerida para unir las subagrupaciones.

En este caso, a la izquierda está la cédula. Las cajas de colores muestran si la cédula es o no igual para dos o más registros. En este caso, solo el antepenúltimo y el último registro tienen la misma cédula. Le sigue el nombre 1, tapado por las cajas negras, que muestran que en este caso el nombre 1 coincide para todos los registros. El dendrograma nos muestra que los registros 5 y 6 son muy similares, están conectados por la línea roja, así como los registros 1, 2 y 3. Dado que permitimos un  $f_t$  de 0,65, en este caso estos registros serían los únicos vinculados.

---

<sup>12</sup>Para una discusión de generación de datos de entrenamiento con base en agrupaciones, y una comparación con los datos de entrenamiento por pares, ver [esta presentación](#) (ver cerca al minuto 17:30) dada en 2016 en un congreso sobre vinculación de registros en el Instituto Isaac Newton para Ciencias Matemáticas en la Universidad de Cambridge.

<sup>13</sup>Usamos la [función `fcluster` de `scipy`](#) para particionar las agrupaciones.

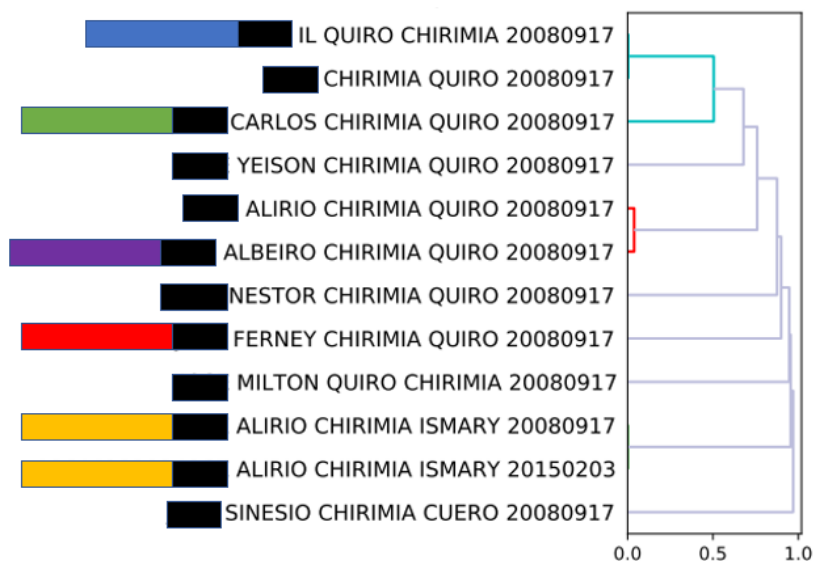


Figura 8: Dendrograma

### 3.6. Unión y exportación

El método de agrupamiento agrupa todos los registros que se refieren a una persona dentro de grupos de coincidencia. Es posible que una misma víctima tenga información distinta según la fuente. Por esto, es necesario establecer una serie de decisiones de reclasificación.

#### 3.6.1. Reclasificación

**3.6.1.1. Hechos de violencia** Una vez se ha realizado el trabajo de identificar a los pares de registros que corresponden a una misma persona es necesario establecer qué hacer cuando una misma víctima tiene dos o más hechos de violencia registrados. Con base en la experticia de investigadores sobre el conflicto se tomó una serie de decisiones de reclasificación.

De acuerdo con el conocimiento de los y las investigadoras se tomaron las siguientes decisiones. Primero, si una vez deduplicamos una misma víctima tiene el mismo hecho, se dejó dicho hecho. Segundo, si en la deduplicación una misma víctima está registrada como víctima de homicidio y otro hecho de violencia, se mantienen los dos hechos. Es decir, en los casos de homicidio y secuestro, homicidio y desaparición, homicidio y reclutamiento, y homicidio y desplazamiento, se mantienen los dos hechos.

La mayor ambigüedad se encuentra en las combinaciones de secuestro, desaparición forzada y reclutamiento, que se observan en la Tabla 16. Allí vemos las proporciones, tomando el primer hecho de violencia como el denominador. Por ejemplo, vemos que el 0,08 de las desapariciones también son secuestro. Mientras que el 0,19 de los secuestros también son desapariciones. Hay muchas más desapariciones que secuestros, por lo que las proporciones son distintas. Contamos a una víctima solo una vez en cada uno de estos tres hechos, por lo que una víctima está solo una vez en una de estas tres categorías.

Las reglas de reclasificación fueron hechas en este caso según la cantidad de información que se requiere para reportar cada uno de los hechos como uno particular. En específico, para hacer el reporte de un reclutamiento es viable suponer que hay información de que el o la menor fueron llevados por un grupo armado ilegal. Mientras que para hacer el reporte de secuestro es viable pensar que se pidió una compensación a cambio. En cambio, para desaparición forzada no se tiene ningún tipo de información. Según estos criterios, es posible suponer que si un registro está documentado como reclutamiento o secuestro es porque hay más información que en el caso de la desaparición forzada. Entonces, en los casos de ambigüedad se hizo la reclasificación de reclutamiento de niños, niñas y adolescentes, seguido por secuestro y por desaparición forzada.

Tabla 16: Proporción de dos hechos de violencia registrados en una misma víctima

Hechos de violencia	Proporción
desaparicion con secuestro	0,08
desaparicion con reclutamiento	0,01
secuestro con desaparicion	0,19
secuestro con reclutamiento	0,01
reclutamiento con desaparicion	0,11
reclutamiento con secuestro	0,04

Por último, es importante enfatizar que también hay víctimas que tienen homicidio y otro hecho de violencia. En estos casos, mantenemos los dos hechos de violencia. La proporción de desaparición y homicidio es 0,16, la de secuestro y homicidio es 0,08 y la de reclutamiento y homicidio es 0,08.

**3.6.1.2. Etnia** Se decidió priorizar a los grupos históricamente discriminados. En específico, si un registro tiene a una etnia históricamente discriminada asociada y, además, a “mestizo,” se mantiene la históricamente discriminada. A partir de esta reglas hicimos la reclasificación de 490 registros de mestizo a indígena, 357 registros de mestizo a afrocolombiano<sup>14</sup> y 4 registros de mestizo a Rrom.

**3.6.1.3. Otras variables** Si se tienen muchos registros en un mismo grupo y los valores para una variable distinta al hecho de violencia o la etnia son distintos, tomamos la moda. Es decir, es valor más común. En caso de que haya un empate, elegimos aleatoriamente.

### 3.7. Resultados

La Tabla 17 muestra la matriz de confusión de los datos de entrenamiento. Un “verdadero positivo” es un par de registros que tanto el oráculo como el modelo identificaron como coincidencia. Mientras que un “verdadero negativo” es un par que tanto el oráculo como el modelo identificaron como una no-coincidencia. Por su parte, un “falso positivo” es un par que el oráculo identificó como no correferentes, pero el modelo agrupó. En cambio, un “falso negativo” es un par que el oráculo identificó como correferentes, pero el modelo no.

Los hiperparámetros del método de agrupamiento  $(cc_t, f_t)$  pueden ser ajustados para cambiar el balance entre falsos positivos y falsos negativos. Nosotros optamos por tener un balance entre falsos positivos y falsos negativos. Sin embargo, dado que nuestra prioridad es identificar a las distintas víctimas del conflicto estamos dispuesto a aceptar una tasa ligeramente más alta de falsas coincidencias, en vez de dejar de identificar a distintas personas. Lo anterior se observa en la 17, donde vemos que contamos con un poco más de 200 falsos positivos más que falsos negativos. Sin embargo, vemos que los registros falso negativo o falso positivo son pocos en comparación a los verdadero positivo o verdadero negativo. Estos resultados indican que el modelo logra clasificar bien a los registros.

Tabla 17: Matriz de confusión final de datos etiquetados manualmente

Medida	Conteo
Falso Negativo	714
Falso Positivo	962
Verdadero Negativo	1.716.239
Verdadero Positivo	1.081.765

<sup>14</sup>A lo largo del texto nos referiremos a personas negras, afrocolombianas, raizales y/o palenqueras como afrocolombianas.



### 3.8. Auditoría

El modelo de vinculación de registros funciona bajo una premisa: todos los pares correferentes, es decir, que se refieren a una misma persona, fueron cubiertos por un bloque. Pero, sería posible que esta premisa se incumpla. Específicamente, realizamos cuatro tests de verificación: posibles errores humanos, el esquema de los datos, las expectativas y la correcta identificación de pares correferentes. A continuación, exponemos cada uno de ellos.

#### 3.8.1. Errores humanos

La tarea de vinculación de registros tiene lugar después de haber hecho la limpieza de datos de cada una de las bases. Podría haber errores humanos en los scripts de procesamiento de las bases. Uno de los errores podría estar relacionado con seleccionar las variables equivocadas. En múltiples ocasiones las bases de datos traen más de una variable con información relacionada con nombre o apellido, que son los campos fundamentales para la vinculación de registros. Por ejemplo, esto puede ser porque las bases incluyan información de otras personas, como quienes denunciaron o algunos testigos. En este caso, sería posible que se hubiese tomado una variable equivocada. Para esto, identificamos todas las columnas con información de nombre/apellido y creamos un archivo en el que expresamos para cada base de datos qué variables existen con esta información y cuál debe ser la correcta. Una vez se tuvo este archivo, hicimos una auditoría de todos los scripts y verificamos que no hubiese errores en el código original y que se estuviesen tomando las variables correctas.

#### 3.8.2. Esquema

Es posible que los nombres de las variables no tengan en su contenido lo que una persona supondría. Por ejemplo, podría ser que una base tenga la variable “nombre” pero que esta tenga un código. Para esto, generamos histogramas con distintas medidas para evaluar el contenido de las variables. En específico, analizamos la cantidad de caracteres y la distribución por letras del nombre 1, nombre 2, apellido 1, apellido 2 y nombre completo de los registros.

En las siguientes figuras vemos los histogramas para el largo del nombre 1 y apellido 1 para 20 fuentes de información que tienen a más del 90% de las víctimas únicas. Observamos que ninguna tiene un comportamiento extraño, lo que sugiere que las variables utilizadas sí tienen un nombre o apellido.

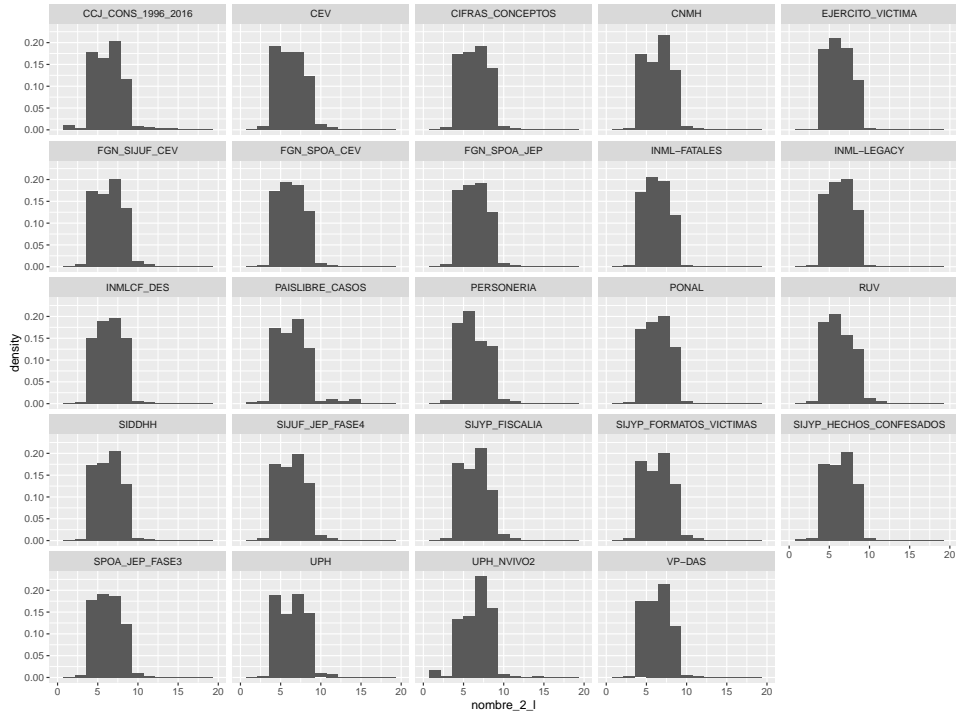


Figura 9: Distribución de cantidad de caracteres de nombre 1

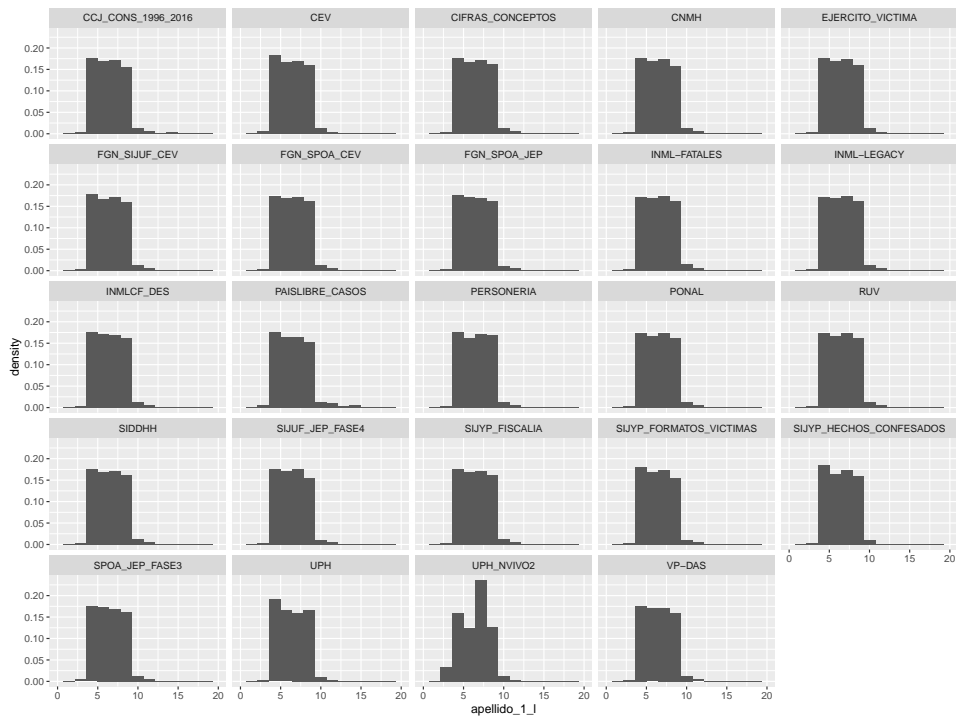


Figura 10: Distribución de cantidad de caracteres de apellido 1

Similarmente, en los siguientes histogramas vemos la distribución por letras de los campos nombre 1 y apellido 1 para 20 fuentes de información que tienen a más del 90% de las víctimas únicas. Nuevamente, no

encontramos comportamientos atípicos.

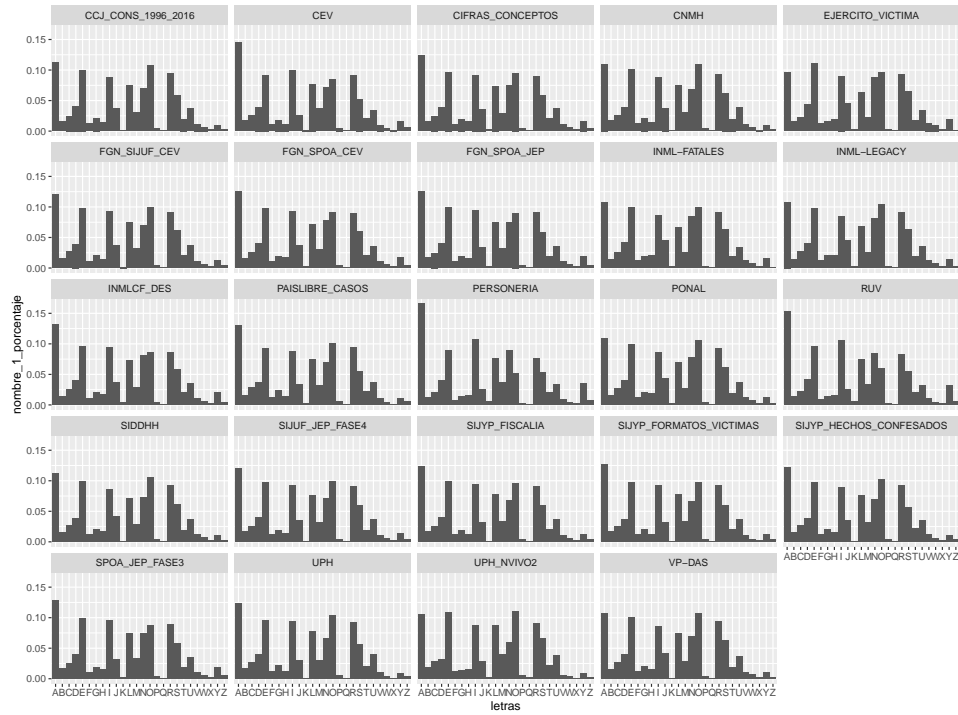


Figura 11: Distribución de caracteres de campo nombre 1 por letra

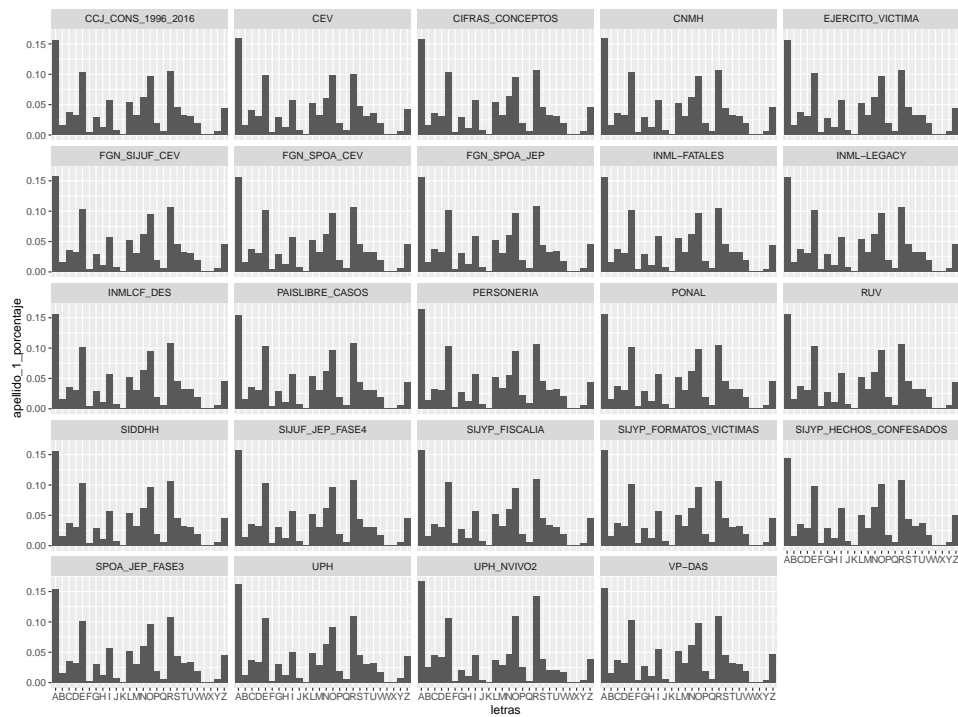


Figura 12: Distribución de caracteres en apellido 1 por letra

### 3.8.3. Expectativas

Algunas de las bases de datos están construidas con otras bases de datos como insumo. Entonces, esperamos algunas sobreposiciones mínimas de las que sabemos ex-ante. En específico, tenemos que:

$$\text{sobreposición}(A, B) = \frac{A \cap B}{|A|}$$

Esta sobreposición no es simétrica, pues es posible que la organización A haya tomado como fuente de información a la organización B, pero lo inverso no es necesariamente cierto.

Gracias a la documentación o conocimiento de las bases podemos conocer algunos casos en los que los proyectos toman como fuente a otras organizaciones. A partir de esto, definimos una sobreposición esperada y testeamos si se cumple o no.

Los resultados se muestran en la Tabla 18, donde evidenciamos que se cumplen las expectativas de sobreposición.

Tabla 18: Expectativas de sobreposición

Hecho	Base 1	Base 2	Expectativa de sobreposición mínima	Sobreposición
Desaparición	INML	CNMH	0,3	0,64
Desplazamiento	FGN	RUV	0,5	0,75
Homicidio	INML	FGN	0,5	0,92
Reclutamiento	ARN	ICBF	0,5	0,66
Secuestro	País Libre	CNMH	0,3	0,64

### 3.8.4. Identificación de pares correferentes

Si los registros correferentes parecieran lo suficientemente diferentes no estarían siendo cubiertos por los bloques. Entonces, sería posible que no se vincularan a una misma víctima. Una posible causa de que los registros correferentes parezcan diferentes es que el procesamiento de datos haya corrompido la información. Por ejemplo, si tuviéramos un error de codificación que intercambié el nombre de pila con el apellido de algunos registros. Este problema afectaría no solo al bloque, sino también al proceso de muestreo que usamos para generar datos de entrenamiento para la canalización de vinculación de registros. Por lo tanto, sería difícil evaluar si se están produciendo tales errores en la generación de bloques.

Gracias a que almacenamos datos intermedios en cada etapa del procesamiento de datos, podemos recuperar versiones no transformadas de los datos de entrada. O sea, las originales. Entonces, entregamos una muestra de enlaces potenciales al oráculo, y ella identificó 3.665 pares de correferencias, de los cuales 83 no fueron descubiertos en los bloques.

Entonces, generamos datos de entrenamiento adicionales para la tarea de bloques. Primero, generamos pares con base en bloques de campos no estandarizados y luego filtramos aquellos que se descubrieron previamente y que nuestro clasificador de pares etiquetó como coincidencias probables. Una innovación aquí fue nuestra capacidad para considerar múltiples representaciones del mismo registro durante la generación de pares. Algunos registros tienen varias versiones del nombre de la víctima, a veces deletreadas de distintas formas, y durante la estandarización de datos seleccionamos una de ellas como canónica.

Finalmente, volvimos a ejecutar la búsqueda de reglas de generación de bloques con los datos de entrenamiento enriquecidos. Las reglas resultantes continuaron funcionando bien en nuestros datos de entrenamiento existentes y, además, cubrieron los 83 pares descubiertos anteriormente que identificó el oráculo. En síntesis, dados todos los test y los resultados del modelo tenemos una alta confianza en la capacidad de nuestro modelo de identificar a las víctimas únicas.

### 3.8.5. Resultados finales de la vinculación de registros

Empezamos la vinculación con 12.863.977 registros y concluimos con 12.862.014 víctimas individuales. Algunas víctimas en la base deduplicada sufrieron más que un hecho de violencia (ver la Sección 3.6.1.1). Otras víctimas en esta base son filtradas (ver la Sección 8.1.1) y por lo tanto no son consideradas para ningún hecho de violencia. Los cálculos para el número de víctimas en la base deduplicada para cada hecho de violencia son presentados en la Sección 2.

## 4. Imputación estadística de campos faltantes

El reto de la imputación estadística de campos faltantes es conocer la probabilidad de que cada campo faltante pertenezca a cierta categoría o tenga cierto valor (como la edad). Los campos que se dejaron para cada una de las bases de datos integradas de los 4 hechos de violencia<sup>15</sup> son: sexo, etnia, edad, responsable y municipio. Sin embargo, dado que también contamos con bases especializadas y no especializadas en conflicto armado fue necesario imputar dos variables adicionales: “pertenece al conflicto” y “es desaparición forzada.” A continuación, presentamos estas imputaciones estadísticas en detalle.

### 4.1. Campos faltantes

En la Tabla 19 vemos para cada variable la proporción de registros que tienen campos faltantes. Podemos observar que la variable más crítica es “responsable,” pues los campos faltantes llegan al 70 % de los registros. Por su parte, la variable de etnia tiene el 39% de campos faltantes en reclutamiento y alcanza 59 % en desaparición. Vemos que las variables de sexo y edad suelen estar más completas. Es necesario recordar que este es el resultado del total de bases, que no se limitan al conflicto armado.

Tabla 19: Proporción de datos faltantes de edad, etnia, responsable y sexo según tipo de violencia.

Hecho de violencia	Edad	Etnia	Responsable	Sexo
Desaparición	0,06	0,59	0,7	0
Homicidio	0,1	0,43	0,66	0,04
Reclutamiento	0,1	0,39	0,23	0,19
Secuestro	0,23	0,44	0,33	0,01

<sup>a</sup> Esta tabla incluye los registros para todos los años, sin limitarse a un periodo determinado.

<sup>a</sup> En todas las variables a imputar hay campos faltantes. Por los dos decimales pueden ser aproximadas a cero.

### 4.2. Pertenece al conflicto

En el proyecto contamos tanto con bases especializadas como con bases no especializadas en conflicto armado. Las bases especializadas son todas aquellas que se limitan a documentar lo ocurrido en el marco del conflicto. Dentro de ellas se encuentra el Observatorio de Memoria y Conflicto del Centro Nacional de Memoria Histórica, el Registro Único de Víctimas de la Unidad para la Atención y Reparación Integral a las Víctimas, y la base de datos de las entrevistas realizadas por la CEV, entre otras. Entonces, construimos una variable instrumental que permite clasificar la pertenencia de los registros al conflicto armado. Esta variable es “pertenece al conflicto” y toma el valor de 1 para las bases especializadas, porque se tiene la certeza de que ocurrieron en el marco del conflicto armado.

Por otro lado, están las bases no especializadas. Dentro de estas se encuentran las del Instituto Nacional de Medicina Legal y Ciencias Forenses y las de la Fiscalía General de la Nación, entre otras. En este caso,

<sup>15</sup>Homicidio, desaparición, secuestro y reclutamiento. No se incluye exilio por los pocos datos. Tampoco se incluye a desplazamiento dada la alta influencia del RUV, como se muestra en la Sección 5.5.

podríamos saber que algunos de los registros ocurrieron en el marco del conflicto armado y la variable tomaría el valor de 1. Por ejemplo, porque el presunto responsable sea un grupo armado. También, podríamos saber que algunos no ocurrieron en el marco del conflicto armado y la variable tomaría el valor de 0. Por ejemplo, porque están definidos como delincuencia común. Sin embargo, hay miles de casos para los que no es claro si ocurrieron o no en el marco del conflicto armado.

Tiene sentido asumir que algunos de los registros de PONAL, FGN y el INML ocurrieron en el marco del conflicto armado, pero ¿cuántos? Para poder responder a esta pregunta, podemos usar al aprendizaje automático. Podemos entonces tener a un algoritmo que aprenda a identificar las características de similitud entre registros que sí ocurrieron en el conflicto armado y que no ocurrieron en el conflicto armado para estimar si un registro sin esta información hace parte o no del conflicto.

En la siguiente subsección presentamos los datos faltantes de la variable “pertenece al conflicto” y cuáles fueron las decisiones que tomamos para generar los datos de entrenamiento y así poder utilizar bases tanto especializadas como no especializadas.

#### 4.2.1. Datos faltantes de la variable “pertenece al conflicto”

En la Tabla 20 vemos los registros que no contaban con esta información. Encontramos que los hechos de violencia de desaparición forzada y homicidio son los más críticos. En el caso de la desaparición, no tenemos información sobre si el 30 % de los registros ocurrieron en el marco del conflicto. Mientras que en el caso de homicidio este porcentaje es del 20 %.

Podemos entonces utilizar el aprendizaje automático para clasificar los registros para los que no tenemos conocimiento. Para esto necesitamos datos de entrenamiento. Es decir, datos que cumplan y que no cumplan una condición. En este caso, la condición es “pertenecer al conflicto armado.” Específicamente, se tomaron las bases de ACOMIDES, CCJ, FGN, JPM, País Libre y el INML y se hizo una serie de criterios a juicio de experto para definir si hace o no parte del conflicto. Así, construimos datos de entrenamiento tomen el valor de 0 si no están asociadas al conflicto y 1 si lo están. Los datos de entrenamiento surgen de la experticia humana. En este caso, acudimos a expertos y expertas en conflicto armado de la CEV, quienes crearon una serie de reglas para los datos de entrenamiento.

Estas reglas son de dos tipos. Por un lado, están los casos para los que no hay duda que el hecho ocurrió o no en el marco del conflicto armado. Entonces, se asigna directamente el valor de 0 o 1. Este es el caso, por ejemplo, de la delincuencia común o de los hechos cometidos por actores armados, respectivamente. Aunque estas reglas cubren la mayor parte de los registros, no son suficientes. Esto, pues necesitamos más datos de entrenamiento, especialmente para los casos que no ocurrieron en el marco del conflicto armado. Es decir, necesitamos más registros con valores conocidos para poder entrenar el modelo.

Así pues, los expertos identificaron una serie de reglas probabilísticas para los homicidios consumados del Sistema Penal Oral Acusatorio (SPOA) de la Fiscalía. Estas reglas nos permiten incluir la incertidumbre en nuestro análisis y reconocer que existe. A continuación, las enunciamos:<sup>16</sup>

- Poco probable que haya ocurrido en el marco del conflicto armado. Por ejemplo, en el caso en que el delito sea “abuso de autoridad por omisión de denuncia art. 417 del Código Penal.” Se asigna una probabilidad de 0.75 a que “pertenece al conflicto” sea falso. Es decir, `is_conflict == 0`.
- No es claro si ocurrió o no en el marco del conflicto armado. Por ejemplo, los “delitos contra la vida y la integridad personal.” Se asigna 0,5 a que “pertenece al conflicto” sea verdadero. Es decir, `is_conflict == 1`.
- Muy probable que haya ocurrido en el marco del conflicto armado. Por ejemplo, “actos de barbarie art. 145 del código penal.” Se asigna 0,75 a que “pertenece al conflicto” sea verdadero. Es decir, `is_conflict == 1`.

La asignación probabilística la hacemos una única vez y gracias a ella es posible convertir algunos valores de NA a 0 y otros a 1. Si el registro no cumple con ninguna de las reglas, permanece como NA. Vale la

---

<sup>16</sup>Están disponibles de forma detallada en el anexo en 9.4.1.

pena recordar que es valioso preservar tantos registros como sea posible de aquellos para los que no hay una claridad sobre su vinculación o no con el conflicto armado para que la imputación estadística pueda explorar el impacto de esta incertidumbre en las estimaciones.

Tabla 20: Proporción de datos faltantes de pertenece al conflicto según tipo de violencia.

Hecho de violencia	Proporción NA
Desaparición	0,3
Homicidio	0,2
Reclutamiento	0
Secuestro	0,08

<sup>a</sup> Por definición el reclutamiento de niños, niñas y adolescentes ocurrió en el marco del conflicto armado, por lo que no hay campos faltantes en la variable pertenece al conflicto

### 4.3. Es desaparición forzada

Similar al caso de haber ocurrido o no en el marco del conflicto armado, es necesario que conozcamos si las desapariciones son o no forzadas. Nuevamente, necesitamos datos de entrenamiento. Entonces, acudimos a la experticia de los investigadores, quienes identificaron casos para los que la asignación resulta inmediata. Por ejemplo, porque el presunto responsable sea un actor armado (toma el valor de 1) o porque sea un caso de delincuencia común (toma el valor de 0). No obstante, necesitamos más datos de entrenamiento para poder clasificar a los registros que no toman el valor de 0 o 1 inicialmente.

Nuevamente, acudimos a una serie de reglas probabilísticas. Este ejercicio se hizo para las bases de Instituto Nacional de Medicina Legal y Ciencias Forenses con base en una serie de hitos en departamentos y años específicos, así como la pertenencia grupal de la persona. Según estas, se asignaron probabilidades de 0, 0,5, 0,6, 0,7, 0,8, 0,9 y 1 de que las desapariciones hayan sido forzadas o no.<sup>17</sup>

Al igual que en el caso de pertenecer o no al conflicto, buscamos reconocer que existe la incertidumbre. La asignación probabilística se hizo una única vez, convirtiendo algunos valores de NA a 0 y otros a 1. Es decir, si un registro está en un departamento  $x$  año definido por las personas expertas como “1 - 0,75” y este registro selecciona un valor aleatorio menor a 0,75, lo asignamos a “es desaparición forzada” = 1. Pero, si el registro selecciona un valor aleatorio mayor a 0,75 permanece como NA. En contraste, considere un registro en un departamento  $x$  año definido por los investigadores como “0 - 0,5.” Si este registro toma un valor aleatorio menor a 0,5, lo asignamos a “es desaparición forzada” = 0. Pero, de lo contrario, permanece como NA. Es importante notar que en cada regla procuramos dejar tantos registros como NA como sea posible para que el modelo pueda explorar la variación creada por la imputación. Nuestra meta es definir registros suficientes con 0 o 1 para entrenar los modelos de soporte e imputación estadística.

Tenga en cuenta que nos interesan las desapariciones forzadas que ocurrieron en el marco del conflicto armado. Si tenemos estas dos condiciones en cuenta, los registros a imputar son los que tienen valores faltantes en “pertenece al conflicto” y/o en “es desaparición forzada.” En este caso, la proporción de datos a imputar estadísticamente es de 0,31.

### 4.4. Variables de soporte para imputación estadística de datos faltantes

Imputar los campos faltantes únicamente con base en las variables disponibles en las bases de datos de los hechos de violencia sería muy difícil. Sería necesario completar estos campos únicamente con base en etnia, edad, sexo, departamento, año, municipio, pertenece al conflicto y es desaparición forzada.

<sup>17</sup>El detalle está disponible en el anexo en 9.4.2.

Sin embargo, en las bases originales hay mucha más información. Algunas bases documentan, por ejemplo, la profesión de la víctima. Otras el arma homicida o la vereda exacta. Y además, muchas tienen campos con texto en los que describen los hechos. Toda esta información le ayudaría muchísimo al modelo. Si él pudiera conocer toda la información que hay de un mismo registro, seguro tomaría mejores decisiones. Para esto, se crean variables de soporte.

Las variables de soporte capturan la información heterogénea y no estandarizada de las bases originales de las que vienen los registros. Resultan de especial riqueza los campos narrativos y descriptivos. Todos los campos se transforman a una secuencia de palabras.

En general, las variables que son útiles para la imputación son las que cumplen con dos características: (i) están fuertemente correlacionadas con las variables a imputar, pero no tienen colinealidad; y (ii) no están faltantes para todos los registros del conjunto de datos completo (Van Buuren, Boshuizen, and Knook (1999)). Nuestras variables de soporte cumplen con estas dos características.

Dado que los datos originales tienen distintos formatos, es necesario generar una estructura estándar que permita utilizarlos en un solo modelo. Para esto, combinamos todos los valores de un registro en una sola secuencia. A esta secuencia le hicimos un proceso de limpieza eliminando los números, puntuación y palabras vacías y convertimos cada palabra en su lema.<sup>18</sup> Después, tomamos una lista de las 2000 palabras más comunes y convertimos cada palabra en una posición de índice de esta lista.

A partir de la conversión de las palabras es posible entrenar una red neuronal recurrente de tipo *memoria a largo-corto plazo* [LSTM por sus siglas en inglés; Hochreiter and Schmidhuber (1997)]. Usamos Tensorflow y Keras para entrenar el modelo. La capa LSTM del modelo extrae 32 neuronas, las cuales son funciones que incorporan los cuatro procesos de LSTM en cada iteración (o “época”) del entrenamiento del modelo (la neurona consta de cuatro procesos, que incluye una celda, una puerta de entrada, una puerta de salida y una puerta de olvido). Las neuronas se alimentan de una función sigmoidea, que es esencialmente una regresión logística, que se ajusta a una variable dependiente binaria.

Para entrenar la red neuronal para una variable en particular se inicia con los registros que no tienen campos faltantes en dicha variable. Dado que el modelo sigmoideo funciona mejor con datos equilibrados, tomamos aleatoriamente una misma cantidad de registros que cumplen y que no cumplen la condición. Por ejemplo, en el sexo tomamos un mismo número de registros que cumplen con la condición de ser hombre y de no ser hombre. A partir de esta selección, aleatoriamente seleccionamos el 75% de los registros para entrenar el modelo y reservamos el 25% para probar el ajuste y el rendimiento del modelo.

Esta información va a permitir en la imputación estadística buscar los registros más “parecidos” a aquellos que tienen campos faltantes y, de esta forma, tomar mejores decisiones sobre el grupo al que pertenecen. En palabras simples, con las variables de soporte se identifican “pistas” adicionales para la imputación estadística, como la descripción del presunto responsable, la vereda exacta, etc. para identificar la categoría a la que pertenecerían los registros. Estas “pistas” no son creadas por humanos, sino aprendidas por la red neuronal. En la siguiente subsección mostramos el ajuste del modelo de las variables de soporte y a continuación la imputación estadística de los campos faltantes.

#### 4.4.1. Ajuste del modelo.

Como se observa en la Tabla 21, el AUC es superior a 0,93 en todas las variables de soporte, mientras que el coeficiente de Pearson supera 0,79. Estas medidas indican que el modelo para crear variables de soporte es confiable para representar la probabilidad observada de que cada variable sea (o no) relevante para cada registro.<sup>19</sup>

---

<sup>18</sup>Consulte [este repositorio de Github](#) para un diccionario de lematización para el español.

<sup>19</sup>El AUC es un indicador del desempeño de un modelo de clasificación, que toma valores entre 0 (el modelo tiene la peor capacidad posible de separación) y 1 (tiene la mejor capacidad de separación). El coeficiente de Pearson toma valores entre -1 (las variables están perfectamente inversamente relacionadas) y 1 (las variables están perfectamente relacionadas).



Tabla 21: Valores de ajuste del modelo, coeficiente de correlación de las variables de soporte con los valores originales, y proporción de datos faltantes imputados.

Variable	AUC	Coficiente Pearson	Proporción faltante
ADULTO	0,942	0,799	0,534
CONFLICT	0,999	0,994	0,091
ELN	0,998	0,983	0,819
EST	0,999	0,995	0,819
EST_PAR	0,981	0,895	0,819
FARC	0,999	0,992	0,819
FEM	0,999	0,991	0,424
FORCED_DIS	0,982	0,893	0,978
GUE	0,999	0,996	0,819
GUE_OTRA	0,937	0,848	0,819
INDIGENA	0,999	0,999	0,552
MASC	0,999	0,991	0,424
MENOR	0,942	0,802	0,534
MESTIZO	0,997	0,987	0,552
AFROCOLOMBIANO	0,999	0,994	0,552
NINO	0,994	0,975	0,534
OTRO	0,993	0,941	0,819
PARA	0,999	0,993	0,819
POSDES	0,999	0,996	0,819

En síntesis, la variable de soporte permite que no haya valores faltantes para ninguno de los registros (pues la variable de soporte toma un valor). Esta es utilizada para soportar a la imputación estadística de datos faltantes, pero no determina su resultado.

#### 4.5. Imputación estadística de campos faltantes para sexo, etnia, edad, responsable, municipio, pertenece al conflicto, y es desaparición forzada

Las variables a las que se les debe completar los campos faltantes, es decir, las variables a imputar son: sexo, etnia, edad, responsable, municipio, pertenece al conflicto, y es desaparición forzada. Una posible solución para lidiar con los campos faltantes sería eliminar los registros que no tienen la información completa. Sin embargo, esto tiene al menos dos problemas. Primero, significaría asumir que los casos sin información tienen la misma distribución que los que sí tienen información. Podemos pensar las consecuencias de esto con una variable como la etnia y el hecho de violencia de homicidio. Este supuesto implicaría que el 43% de las víctimas de homicidio que no tienen etnia registrada tienen la misma distribución que las que sí la tienen. Esto no es necesariamente cierto, pues puede ocurrir que por distintos motivos las personas mestizas sean más reportada que otras. Segundo, esto significaría perder un alto porcentaje de la información disponible, especialmente para homicidio y desaparición. Otra solución podría ser mantener los registros, pero excluir aquellos que no tengan la información completa para un análisis particular. Sin embargo, esto desconocería que es posible que los registros con campos faltantes tuvieran información distinta. Por ejemplo, si los presuntos responsables de los campos desconocidos fueran distintos a los conocidos, la historia que conoceríamos del conflicto sería distinta. Se hace necesario entonces imputar estadísticamente.

Una intuición sería que los valores faltantes de una variable sigan un patrón similar a los valores observados de la variable condicionada a todos los valores observados. A eso se le conoce como “valores faltantes aleatorios” (MAR en inglés). Esta es una suposición bastante plausible, que indicaría que los registros similares en otras variables se parecen entre sí en el valor que deban tomar en el campo faltante.

Para poder imputar el campo faltante de una variable condicionada a todos los valores en un registro se utiliza la combinación de todas las otras variables de ese registro, incluyendo las variables de soporte. Esto se

realiza para todos los campos faltantes del registro y se le llama *imputación múltiple* (Rubin 1987, 1996).<sup>20</sup> La aproximación descrita aquí es una *especificación completamente condicionada*.<sup>21</sup>

Cuando es usado correctamente, el método de imputación múltiple produce estimaciones no sesgadas para las variables imputadas e incorpora la incertidumbre introducida por la imputación estadística de datos faltantes en las estimaciones finales de las estadísticas de interés (van Buuren 2018, Capítulo 2).<sup>22</sup>

Este método comienza completando los valores faltantes por algún valor. Luego, el valor para cada variable se predice usando todas las otras columnas. Hay muchos modelos posibles para la predicción. En este caso, utilizamos un método conocido como *predictive mean matching*. Este método calcula el valor predicho del campo faltante con base en la selección aleatoria de un registro cercano a él. Este es un método robusto, que tiene mejor desempeño que otros como árboles de decisión.

El ejercicio de imputación estadística de datos faltantes es secuencial, pues se completan todos los campos faltantes de todas las variables. Al proceso de predecir y completar cada variable se le conoce como una réplica. Cada una de estas réplicas tiene un componente aleatorio para ayudar al modelo a computar. En este caso, el *donante* del que se toma el valor. Por lo tanto, cada vez que se realiza una réplica el resultado será un poco distinto, reflejando la incertidumbre. Entonces, se realiza el ejercicio más de una vez. A este método se le conoce como “imputación múltiple” y fue propuesto por Rubin (1978). Años después, Van Buuren, Boshuizen, and Knook (1999) propusieron la imputación múltiple con ecuaciones encadenadas.<sup>2324</sup> Este es el método estándar utilizado en la ciencia para lidiar con campos faltantes y se usa en campos como la ciencia política y epidemiología, entre otros.

En la Tabla 22 ejemplificamos el impacto de la imputación estadística con el caso de los homicidios y la variable pertenece al conflicto. La columna “Registros totales” nos muestra los homicidios totales (sin limitarse al conflicto) documentados por la Instituto Nacional de Medicina Legal (INML), la Policía Nacional (PONAL), la Comisión Colombiana de Juristas (CCJ) y la Fiscalía General de la Nación (FGN). Ahora, como para algunos registros no sabemos si ocurrieron o no dentro del marco del conflicto armado, utilizamos el modelo descrito previamente en esta sección. Los resultados de aplicar nuestro modelo para medir e incluir la incertidumbre de los homicidios registrados se muestran en la tercera columna. Según nuestras estimaciones, de los 155.997 homicidios registrados por el INML 92.757 habrían ocurrido en el marco del conflicto armado. Es decir, un 59%. Para PONAL, la CCJ y la FGN estos porcentajes serían 55%, 90% y 64%, respectivamente. Vale la pena recordar que tanto en los “Registros totales” como en los “Registros relacionados con el conflicto” hay víctimas que son comunes en las 4 bases de datos. Es decir, hay coincidencias entre las víctimas registradas por el INML, PONAL, la CCJ y la FGN.<sup>25</sup>

Tabla 22: Registros totales y relacionados con el conflicto armado

Base de datos	Registros totales	Registros relacionados con el conflicto
INML	155.997	92.757
PONAL	151.407	83.431
CCJ	38.131	34.266
FGN	145.068	92.439

<sup>a</sup> Estos registros no están limitados en el tiempo, por lo que pueden diferir de la sección de hallazgos.

<sup>20</sup>Se utiliza el paquete de R `mice` para los cálculos realizados (van Buuren and Groothuis-Oudshoorn 2011).

<sup>21</sup>Ver van Buuren (2018), específicamente la Sección 4.5.1. También ver Kropko et al. (2014), quien muestra que para variables categóricas como las usadas aquí, las especificaciones condicionadas (como las usadas aquí) tienden a producir mejores imputaciones que los modelos normales multivariados. Ver también a Cro (2017) y Heymans and Eekhout (2019).

<sup>22</sup>Este libro está disponible en línea aquí.

<sup>23</sup>En específico se hacen 10 réplicas White, Royston, and Wood (2011).

<sup>24</sup>Una vez están las réplicas, estas son combinadas de acuerdo con las reglas de Rubin (1987).

<sup>25</sup>La imputación estadística la hacemos 10 veces. A cada una de esas veces se le conoce como “réplica.” Por ejemplo, en la primera réplica el INML y PONAL tienen 39.725 registros que coinciden, mientras que el INML y la FGN tienen 27.976 y la FGN y PONAL tienen 17.809 coincidencias.

## 5. Patrones de documentación

Como se ha mencionado previamente, es probable que una misma víctima esté registrada en una o más bases de datos. Esto va a ser clave para realizar la estimación por sistemas múltiples. Previo a ahondar en las estimaciones, es necesario comprender cuáles son las coincidencias entre las distintas bases de datos para cada uno de los hechos de violencia.

Para esto, en esta sección se utilizan unas gráficas llamadas **UpSet Plots**. Dado que nos interesa estudiar los patrones de documentación de las víctimas relacionadas con el conflicto armado, en esta sección presentamos los UpSet después de la imputación de la variable “pertenece al conflicto.” Entonces, estas gráficas permiten identificar cuántas personas están reportadas en distintas combinaciones de bases de datos para la primera réplica de la imputación estadística de campos faltantes. También, es importante resaltar que se presentan las bases o combinaciones que más registros reportaron, pero no por esto son las únicas que posean información sobre un hecho de violencia en particular.

Vale la pena recordar que hay una diferencia entre las dinámicas de *violencia* y los patrones de *documentación*. Las muestras que se tienen en las bases de datos son muestras por conveniencia, que muestran lo observable. Sin embargo, puede haber violencias que no son documentadas y que hacen parte de las dinámicas de violencia (Price and Ball 2015). Los siguientes UpSet presentan los patrones de *documentación*.

### 5.1. Homicidio

Cada una de las filas muestra una organización y los puntos negros muestran cuando una organización está “activa.” En este caso, la primera columna muestra que el RUV es la fuente que más víctimas únicas reportó y que no estaban en otra. En la segunda columna están activos el RUV, CNMH y SIJyP, reflejando que el segundo mayor número de víctimas únicas lo registran estas tres organizaciones. Es decir, las tres tienen documentadas a estas víctimas. El tercer lugar lo ocupa la FGN, que reiteramos muestra en este caso los homicidios consumados.

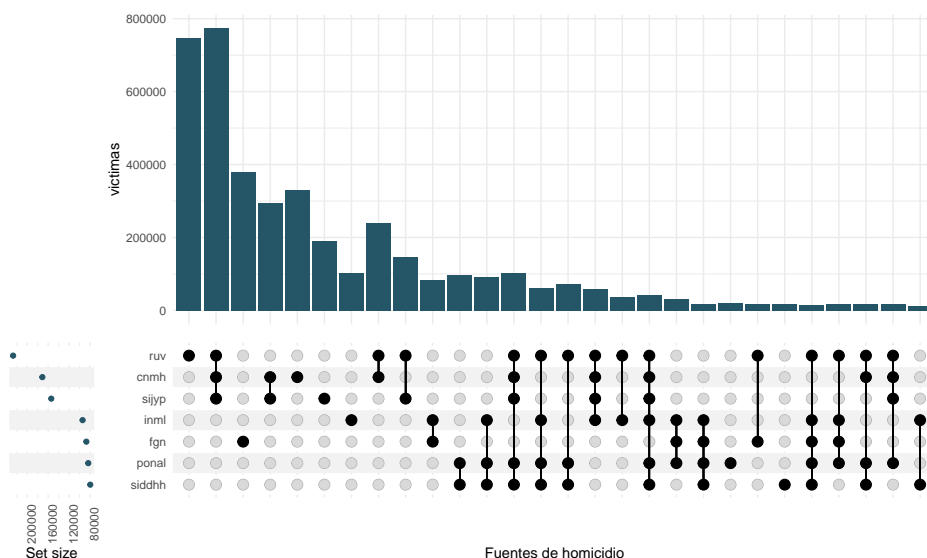


Figura 13: Patrón de documentación entre fuentes para homicidios

### 5.2. Desaparición

El INML es la fuente que más víctimas únicas reportó, seguida por las coincidencias entre INML, CNMH, RUV y SIJyP. El tercer lugar lo ocupan CNMH y RUV.

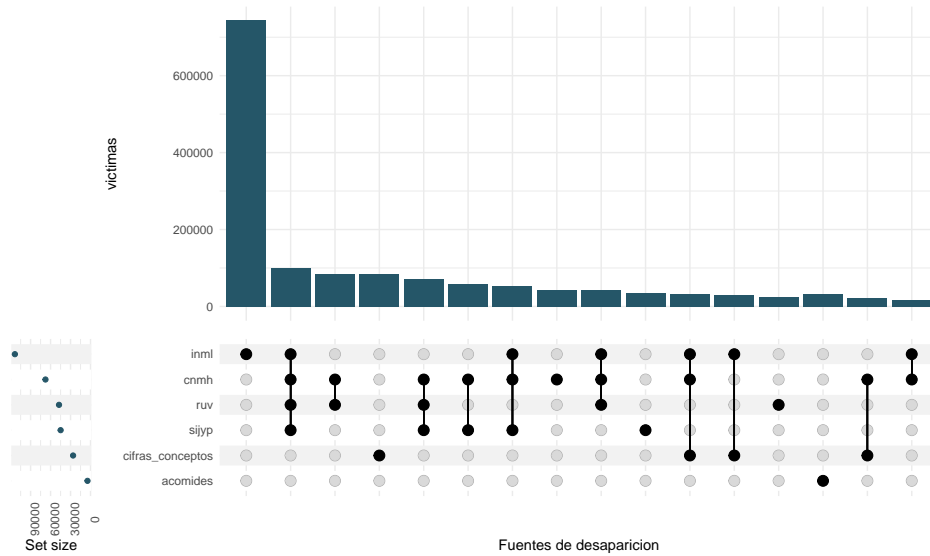


Figura 14: Patrón de documentación entre fuentes para desaparición

### 5.3. Reclutamiento de niños, niñas y adolescentes

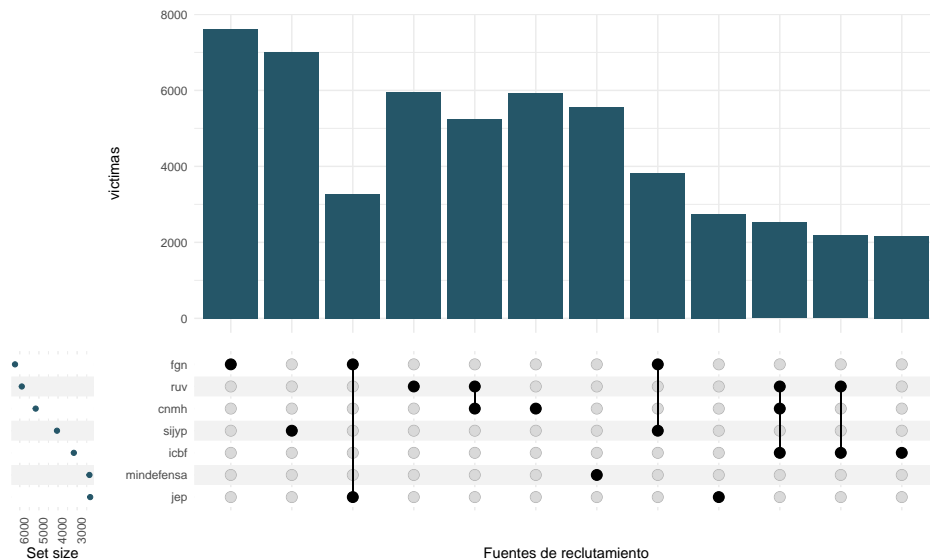


Figura 15: Patrón de documentación entre fuentes para reclutamiento

En el caso de reclutamiento de niños, niñas y adolescentes, la FGN es la fuente que más víctimas reportó de manera única. Le sigue SIJyP y, a continuación, las coincidencias entre FGN y JEP. Vale la pena reiterar que los datos de la FGN son los documentados por esta institución, que fueron revisados por analistas de la JEP.

### 5.4. Secuestro

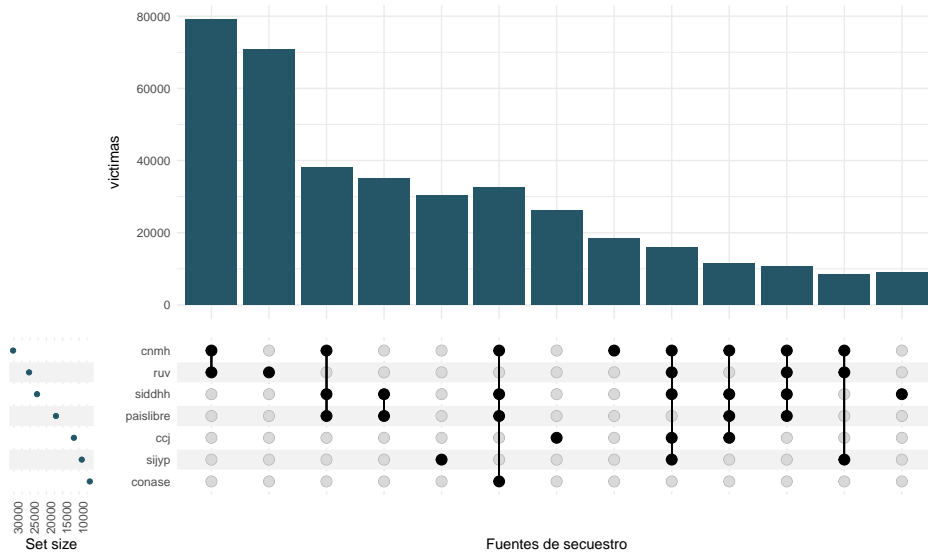


Figura 16: Patrón de documentación entre fuentes para secuestro

En el caso de secuestro el primer lugar lo ocupa el CNMH y el RUV. Le siguen las coincidencias entre CNMH, País Libre y la base de datos del sistema Información en Derechos Humanos IDH de la Consejería Presidencial para los Derechos Humanos.

### 5.5. Desplazamiento forzado

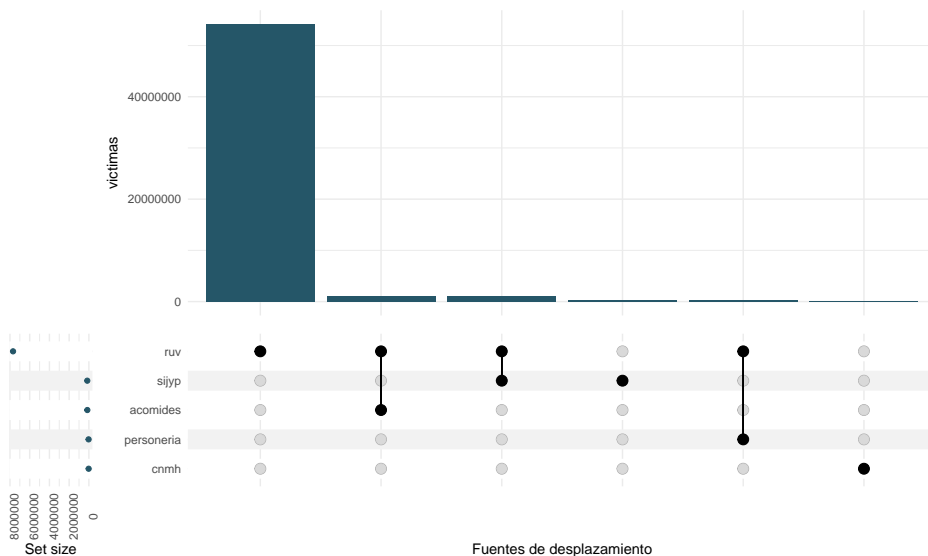


Figura 17: Patrón de documentación entre fuentes para desplazamiento forzado

En la deduplicación de desplazamiento forzado es claro el peso enorme que tiene la base datos del RUV. Aquí se tomaron sólo víctimas directas del RUV que hayan sufrido al menos un desplazamiento forzado y se

incluye únicamente las observaciones donde la víctima reporta un departamento del hecho, año del hecho, y al menos un nombre y un apellido que identifique a una persona en específico (no un grupo colectivo). En el caso de múltiples desplazamiento solo se toma la información de un desplazamiento tomando la moda sobre todos los registros asociados a una misma víctima. En víctimas únicas, al RUV le siguen las coincidencias entre el RUV y COMIDES, y el RUV y SIJyP.

## 6. Estimación por sistemas múltiples

Para estimar la población total de víctimas de varias violaciones de derechos humanos y, por lo tanto, el subregistro en la documentación de la violencia, utilizamos una técnica llamada *estimación por sistemas múltiples* (ESM), también conocida como captura-recaptura. La ESM es una clase de modelos estadísticos que se utiliza para estudiar poblaciones humanas y animales desde finales de la década de 1890. [Chao \(2001\)](#) ofrece una revisión del desarrollo técnico de este método, mientras que [Bird and King \(2018\)](#) proporcionan una mirada histórica de las aplicaciones de ESM a poblaciones humanas. [Lum, Price, and Banks \(2013\)](#), [Ball and Price \(2019\)](#), [Ball and Price \(2018\)](#) y [Rozo Ángel and Ball \(2019\)](#) muestran aplicaciones a cuestiones de derechos humanos, incluso en el contexto de conflictos armados.

La intuición detrás de este método es la siguiente:<sup>26</sup> Imagine dos cuartos oscuros. Queremos saber sus tamaños, pero no podemos ver dentro de ellos y la única herramienta que tenemos para explorar los tamaños es un puñado de pelotas de goma. Las pelotas de goma no hacen ningún sonido cuando se estrellan con las paredes, el techo o el piso, pero hacen un pequeño ruido—*click*—cuando se estrellan entre ellas. Lanzamos las pelotas de goma en la primera habitación y escuchamos muchos *clicks* seguidos. Tomamos nuevamente las pelotas y las lanzamos en la segunda habitación con la misma fuerza. Ahora escuchamos *clicks* pero con menos frecuencia. Nuestra intuición es que la segunda habitación es más grande porque las pelotas de goma se dispersan más y por lo tanto se estrellan con menor frecuencia.

En el lenguaje estadístico, el tamaño del “cuarto” es el tamaño de la población de víctimas de un tipo de hecho que queremos estimar y estamos “lanzando” las fuentes a la población de víctimas. Cuando dos o más de las fuentes documentan a la misma víctima es como si se “estrellaran” haciendo un *click*. Utilizamos estos patrones de documentación para estimar el tamaño de la población total de víctimas de un hecho específico, incluso las que nunca fueron documentadas en nuestras fuentes (el subregistro).

### 6.1. Detalles técnicos

La ESM fue implementada originalmente en 1783 por el matemático Frances Pierre-Simon Laplace con dos fuentes de datos ([Amorós 2014](#)) y cuatro supuestos fundamentales ([Scheuren 2004](#)). Aunque los métodos modernos de la ESM no necesariamente requieren de estos supuestos, resulta útiles conocerlos para entender cómo funciona la metodología. A continuación presentamos los supuestos:

- 1. La población estimada es “cerrada,” es decir, miembros de la población no son creados ni eliminados durante el periodo de documentación.

En este análisis el objetivo es estimar la población total de víctimas de desaparición forzada, desplazamiento forzado, homicidio, reclutamiento de niños, niñas y adolescentes y secuestro. Para los hechos que pueden ocurrir más de una vez a una víctima (a excepción de homicidio) se incluye sólo una instancia del hecho. Entonces, es posible afirmar que las poblaciones de víctimas de los cinco hechos no cambian retroactivamente. Con la excepción de casos raros en los que se documenta que fue asesinado y luego se encuentra vivo, no hay forma de salir de la población de víctimas una vez sea ingresada.

- 2. El vínculo de los registros es preciso.

---

<sup>26</sup>HRDAG usa esta analogía en la mayoría de los reportes que utilizan ESM.

La ESM requiere una base de datos consolidada que incluya a cada una de las víctimas indicando qué fuente la registró. Esta base de datos es el resultado del proceso de vinculación de registros. Aunque no es posible asegurar que se hizo la vinculación perfecta de todos los registros como sugiere el Supuesto (2), el proceso semi-supervisado permite la evaluación de desempeño del vínculo. En la Sección 3, se mostró que la precisión en la calibración del modelo implementado es suficiente para tener una tasa baja de falsos positivos y negativos, lo cual permite tener confianza en el proceso de vinculación de registros. Además, se ha mostrado que incluso con una pobre vinculación de registros, los resultados de la ESM son confiables (J. Johndrow, Lum, and Dunson 2018).

- 3. Estar documentado en un sistema o fuente no afecta la probabilidad de estar documentado en cualquier otro (“independencia de listas o fuentes”).

Resulta útil usar un ejemplo para comprender qué es la independencia de fuentes. Imaginemos que tenemos una población en un municipio donde ocurrió una serie de hechos de violencia en el marco del conflicto armado. Cada una de las víctimas de esta población tiene muchas características: su raza, sexo, etnia, entre otras. Una de sus características es la ideología política: probablemente algunas de las víctimas de un mismo municipio tengan una ideología más cercana a la derecha y otras a la izquierda. Ahora bien, una vez ocurren los ataques, distintas organizaciones llegan a documentar lo sucedido. Empecemos con una organización que tiene una ideología política de derecha o a la que se le asocia más con esta corriente política: es posible que algunas víctimas con una ideología cercana a la izquierda no se sientan cómodas compartiendo con la organización lo sucedido. Así pues, es probable que el listado de esta organización refleje más la victimización que sufrieron las personas de derecha que las de izquierda, no por un prejuicio de la organización sino por procesos de confianza. En este caso, la correlación de la probabilidad de observación entre las dos fuentes sería negativa.

Este supuesto puede flexibilizarse usando tres o más fuentes de datos y usando modelos log-lineales para estimar la interacción entre subgrupos de fuentes (Bishop, Fienberg, and Holland 1975; Cormack 1989; Chao 1987). Para el enfoque bayesiano ver Madigan and York (1997). Sin embargo, el enfoque log-lineal está limitado a unas pocas fuentes de datos. También requiere que el analista seleccione un modelo específico, o bien calcule un promedio ponderado sobre todos los posibles modelos, lo cual contribuye a controlar el Supuesto (4) sólo en la medida en que la heterogeneidad de las probabilidades de captura esté correlacionada con aparecer registrado en una fuente de datos específica.

- 4. La probabilidad de ser documentado en un sistema o fuente particular es igual para todos los miembros de la población (“homogeneidad de captura”).

La probabilidad de captura (probabilidad de estar registrado en una fuente) es constante entre las fuentes. Las desviaciones de este supuesto son llamadas “heterogeneidad de captura” o “capturabilidad variable” en la literatura estadística (Chao 1987). En realidad, esto no suele suceder con todas las víctimas, pues algunas tienen más probabilidad de ser registradas que otras. Por ejemplo, es más probable que se registre el homicidio de un abogado defensor de derechos humanos en Bogotá al asesinato de una campesina en una vereda. También, es posible que los proyectos no tengan la misma probabilidad de recopilar datos en distintos años. Por ejemplo, una fuente específica puede haber tenido un recorte de presupuesto en un año y haber parado la documentación de violaciones de DDHH. Entonces, la probabilidad de documentación de una víctima ese año en esa fuente sería menor que en un año distinto. Si no se incluyera el año de recolección, habría una heterogeneidad de captura dentro de la fuente.

Esto puede controlarse de distintas formas. El enfoque clásico de Sekar and Deming (1949) reduce la heterogeneidad de captura agrupando registros similares. A este proceso se le conoce como “estratificación.” En el proyecto siempre estratificamos por año y en algunos casos por año y departamento u otras variables. Con la estratificación se crean pequeñas listas, por ejemplo año1-depto1, año1-depto2, ... añoX-depto32, que le brindan al modelo más información.

Cuando el modelo no puede manejar la heterogeneidad de captura se observan distribuciones posteriores multimodales. Es decir, el modelo tiene dificultad en elegir una respuesta sin más información. Muchos

estudios de la teoría de probabilidad de la ESM han encontrado que esto puede deberse a una configuración en los datos, que lleva a la “no identificabilidad” de una respuesta consistente.<sup>27</sup> Para una discusión y una propuesta para resolver este aspecto, que no utilizamos en el proyecto, ver a [J. E. Johndrow, Lum, and Manrique-Vallier \(2019\)](#). Para una discusión de cómo seleccionar los estratos para tratar con estimación multimodal, ver [Hoover Green and Ball \(2019\)](#).

Con tres o más fuentes de datos independientes hay varias maneras de reducir el efecto de la heterogeneidad de captura. El enfoque clásico señala que la heterogeneidad de captura normalmente genera dependencia entre las fuentes. Un enfoque para controlar la heterogeneidad de captura durante los años 90 fue usar los modelos de Rasch ([Darroch et al. 1993](#); [Fienberg, Johnson, and Junker 1999](#)). En este método, una variable latente recoge la variación en la capturabilidad individual para cada registro  $i$  (expresada como  $\theta_i$ ) y el efecto individual de aparecer en la lista  $j$  expresado como  $\text{logit}(p_{ij}) = \beta_j + \theta_i$ . Aunque este enfoque fue un gran avance, es todavía limitado debido a que asume que los efectos entre las fuentes o listas y los efectos individuales no interactúan.

En este análisis usamos un método que cubre los supuestos (3) y (4) simultáneamente superando las limitaciones del enfoque log-lineal. Usamos un tipo de ESM bayesiano no paramétrico llamado “*Modelos de Clases Latentes Múltiples para Captura-Recaptura*” [LCMCR por sus siglas en inglés; [Manrique-Vallier \(2016\)](#)].<sup>28</sup> El modelo LCMCR estima una distribución de probabilidad posterior de los valores probables para la cantidad total de víctimas (los observados más nuestra estimación del tamaño del subregistro) en el estrato de interés.

## 6.2. Implementación con datos imputados

La varianza de los modelos encadenados se propaga. En nuestro caso, las imputaciones múltiples propagan su error a las estimaciones por sistemas múltiples. Cuando la imputación estadística de datos faltantes tiene una alta incertidumbre, la varianza es mayor, entonces la varianza de las estimaciones también lo será.<sup>29</sup>

En general, la varianza de las imputaciones es mayor que la de la ESM. Esto se debe a que la varianza de la imputación estadística de datos faltantes se basa en el supuesto de que las réplicas tienen una distribución normal (lado izquierdo), mientras que las estimaciones tienen una distribución log-normal (lado derecho). Lo anterior, pues en las estimaciones conocemos el mínimo, que es igual a las víctimas documentadas.

---

<sup>27</sup>Ver a [Link \(2004\)](#).

<sup>28</sup>Utilizamos el paquete [LCMCR](#) desarrollado por [Profesor Daniel Manrique-Vallier](#) en R para todas las estimaciones en nuestro análisis.

<sup>29</sup>Ver [Zhou and Reiter \(2010\)](#).



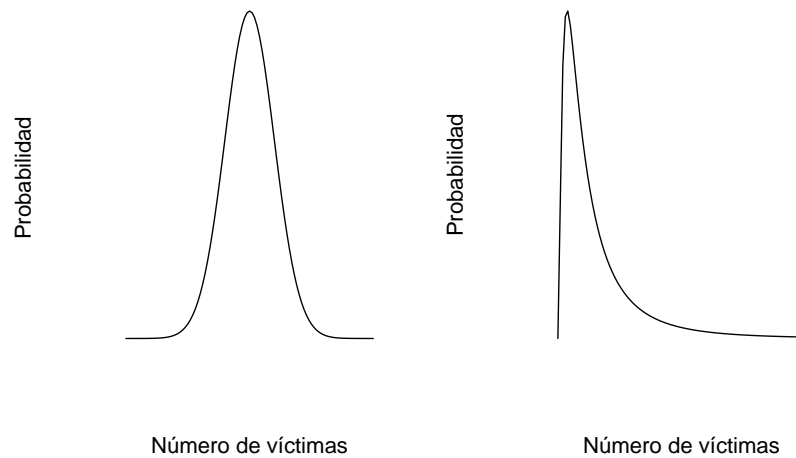


Figura 18: Distribución distribución normal y log-normal.

Para el modelo LCMCR se usa como insumo una tabla con tantas filas como cantidad de víctimas y tantas columnas como fuentes. Si una víctima fue documentada por una fuente toma el valor de 1, si no, toma el valor de cero. A partir de esta tabla de ceros y unos se hace la estimación. Pero, previamente a estimar se estratifican los datos para reconocer que puede haber particularidades de ciertas variables, que afecten la heterogeneidad de captura. En cualquier caso, es necesario que en cada estrato haya más de una víctima observada en al menos tres fuentes. Entonces, solo estimamos estratos para los que hubiera al menos 3 listas y corrimos nuestro muestreador, que genera muestras de 1.000, itera 500 veces y quema 10.000 datos para generar las estimaciones.

En caso de que los estratos satisfagan esta condiciones, se estima el estrato utilizando LCMCR. Para esto se generan 1000 muestras aleatorias de la distribución posterior de la cantidad de víctimas probables para las 10 réplicas de la imputación estadística de datos faltantes. Realizamos la estimación de las 10 réplicas y combinamos las distribuciones posteriores para llegar a la estimación y el intervalo. Si bien para esto se necesita que las estimaciones posteriores se aproximen a una normal, en general las estimaciones de ESM tienen una cola pesada a la derecha.

Siguiendo a [Chao \(1987\)](#), transformamos nuestros datos según  $\theta(N) = \log(N - n)$ , donde  $n$  es el número de víctimas observadas en el estrato. Hacemos esta transformación por dos razones. Primero, el límite inferior del número de víctimas en un estrato es delimitado por el número de víctimas observadas en dicho estrato. Los cálculos de la incertidumbre para las estimaciones no se deben tener límites inferiores menos que el número de víctimas documentadas en el estrato. Usar la cantidad  $N - n$  se garantiza que el intervalo de credibilidad para la estimación es siempre mayor que el número de víctimas documentadas en el estrato.<sup>30</sup> Segundo, para calcular las estimaciones finales y los intervalos de credibilidad, utilizamos las reglas de combinación descrito en [Rubin \(1987\)](#). Estas reglas se suponen que la distribución de  $N$  es normal, pero como ya explicamos las distribuciones posteriores tienen una cola pesada al lado derecho. La distribución de  $N$  se aproxima una distribución log-normal, así  $\log(N)$  se aproxima una distribución normal. Combinando los dos: la distribución de  $\theta(N)$  se garantiza que el intervalo de credibilidad siempre es igual a o mayor que el número de víctimas documentadas en el estrato y es generalmente aproximadamente normal así podemos aplicar las reglas descritas en [Rubin \(1987\)](#) sin violar los supuestos.

Aplicamos esta transformación a las muestras de las distribuciones posteriores de  $N$  para cada una de las diez réplicas. Finalmente, combinamos las diez estimaciones de  $\theta$ ,  $\hat{\theta}_{IM}$ , y sus errores estándar, s.e. ( $\hat{\theta}_{IM}$ ) siguiendo

<sup>30</sup>Ver [Chao \(2015\)](#) p. 12 nota 2 en el código.

las reglas de combinación descrito en Rubin (1987) y deshaciendo la transformación para obtener resultados en la escala original.<sup>31</sup> Como resultado tenemos,  $\hat{N}_{IM} = n + \exp(\hat{\theta}_{IM})$  y  $IC_{95\%} = n + \exp(\hat{\theta}_{IM} \pm t \times s.e.(\hat{\theta}_{IM}))$ , donde  $IC_{95\%}$  es el intervalo de credibilidad de probabilidades iguales de 95% y  $t$  es el percentil 97,5 de una distribución  $t$  con grados de libertad calculados según las reglas descritas en Rubin (1987).

## 7. Sobre las estimaciones

Hay dos características principales de las estimaciones: el sesgo y la varianza. Describimos la estimación de esta última en la Sección 6.2. El sesgo representa la distancia que existe entre el valor verdadero y el valor esperado con el procedimiento de estimación. Éste se da por dos motivos principales: i) la variación en la probabilidad de observación o heterogeneidad de captura y ii) la sobre-estratificación. Por su parte, la varianza (o incertidumbre) significa que, al no contar con todos los registros, el proceso de estimación conlleva una incertidumbre. Es decir, que los resultados podrían cambiar dependiendo de las fuentes de variación (o de incertidumbre). Entonces, los métodos aplicados permiten estimar esta variación generando un rango donde puede encontrarse el verdadero valor. Así pues, las estimaciones no brindan una cifra certera. Existen 4 fuentes de incertidumbre, que explicamos en detalle en esta sección: i) la relación entre víctimas documentadas y el universo de víctimas, ii) la densidad de información en las listas, iii) la variación de la probabilidad de observación y iv) la imputación estadística de campos faltantes. Ahondaremos en cada una de las características y sus fuentes a continuación.

### 7.1. Sesgo

El sesgo ocurre cuando existe una diferencia sistemática entre la realidad y los resultados de la estimación. Pero ¿por qué habría sesgo? Esto sucedería si la información que tenemos estuviera omitiendo a algunas poblaciones. Por ejemplo, si ninguna de las bases de datos registrara hechos de violencia contra población afrocolombiana, nuestras estimaciones estarían sesgadas: tampoco mostrarían la violencia que ha sufrido esta población. Sistemáticamente, nuestras estimaciones excluirían la violencia que han vivido las personas afrocolombianas, por lo que no nos mostrarían la realidad de la violencia. Otro ejemplo sería una población exterminada de la que no hay registro: nuestras estimaciones estarían sesgadas porque estarían omitiendo a estas víctimas. Hay dos fuentes que generan el sesgo: la variación en la probabilidad de observación y la sobre-estratificación. A continuación explicamos cada una de ellas.

- 1. Variación en la probabilidad de observación/heterogeneidad de captura: La variación en la probabilidad de observación se da cuando distintos individuos tienen diferente probabilidad de ser documentados en un listado. Por ejemplo, hay dos individuos: un abogado reconocido en una capital y una campesina en una vereda. Si los dos son asesinados, es más probable que los proyectos documenten el asesinato del abogado (podría ser con una probabilidad de 0.9), mientras que el homicidio de la campesina tendría una probabilidad más baja de ser registrada (por ejemplo, 0.05). Entonces, es posible que por aleatoriedad en las bases de datos no haya ningún registro de campesinas asesinadas. Al hacer las estimaciones, fallaríamos en estimar el subregistro para campesinas, pero lo haríamos bien para abogados. Nuestras estimaciones estarían sesgadas al no incluir a las víctimas campesinas.

La heterogeneidad de captura no se da solo por la profesión de una víctima. Por ejemplo, es probable que en los municipios más urbanos haya un mayor registro de la victimización. Asimismo, es probable que en los periodos más recientes haya un menor subregistro gracias a la conectividad o a la concientización de los derechos humanos o facilidad de denunciar. También, es posible que en los lugares con más seguridad haya más registro, entre otras. Estas diferencias pueden

<sup>31</sup>Ver Gelman et al. (2013) p. 453 para un resumen breve de las reglas que utilizamos para combinar los resultados de las diez réplicas.

llevar a que las estimaciones fallen y estén sesgadas al subestimar a las poblaciones con menor probabilidad de captura.

Para entender de dónde proviene la variación en la probabilidad de observación resulta útil comprender el proceso de recolección de los datos. Las listas se construyen con base en al menos tres características únicas de cada institución u organización i) su mandato, ii) su presupuesto, y iii) su capacidad logística. Esto lleva a que los proyectos documenten lo que está dentro de sus capacidades y misión, por lo que habrá víctimas que no estén incluidas. Por ejemplo, algunas pueden especializarse en hechos de violencia cometidos contra un sector específico de la sociedad (sindicalistas, mujeres, afrocolombianos, etc.) y otras pueden concentrarse en una región del país (el Pacífico, un departamento concreto, una ciudad, etc.).

La variación en la probabilidad de observación crea sesgo, pues dadas las características de las organizaciones e instituciones que documentan a las víctimas, tendrán mayor o menor registro de algún tipo de víctimas según las características sociodemográficas, temporales o geográficas. Así, si hiciéramos estimaciones basadas en listas con características similares, nuestros resultados estarían sesgados: estaríamos mostrando la realidad de las víctimas que tienen mayor probabilidad de ser registradas por las organizaciones. No obstante, en este proyecto contamos con más de 100 bases de datos de más de 40 organizaciones e instituciones diversas, por lo que consideramos que logran cubrir a gran parte de los tipos de víctima y permite mitigar el sesgo. Sin embargo, esto no elimina el sesgo que podría haber por variables temporales o regionales.

- 2. “Sobre-estratificación”: Una de las formas de lidiar con la heterogeneidad de captura es por medio de la estratificación. Una estratificación, según [Särndal \(2003\)](#), es una partición de la población. Es decir, la unión de todos los conjuntos reproduce la población y la intersección entre cualquier par de conjuntos es vacía. Esto se hace para lograr homogeneidad de la característica de interés en cada estrato y así disminuir la incertidumbre. En otras palabras, la estratificación es una estrategia por medio de la cual se toma un listado inicial y a partir de él se crean listados más pequeños con registros que compartan características, por ejemplo sexo y etnia. Se tienen entonces listados de “mujeres-afro,” “hombres-afro,” “mujeres-indígenas,” y así sucesivamente para todas las posibles combinaciones de las categorías de las dos características. Al separar la gran lista en pequeñas listas, el modelo tiene más información (más listas) para hacer la estimación.

En este proyecto siempre estratificamos por al menos el año para reconocer las diferencias que podrían llevar a que en algunos años haya mayor o menor probabilidad de observar a las víctimas. Si bien intuitivamente podría pensarse que sería mejor utilizar todas las variables disponibles para estratificar y así “garantizar” que el modelo lidie mejor con la variación en la probabilidad de observación, en caso de que sobre-estratifiquemos estaremos teniendo estimaciones sesgadas. Pensemos un momento qué significa sobre-estratificar: significa que utilizamos tantas variables para hacer listas pequeñas de todas las combinaciones, por ejemplo “mujer-afro-en el departamento X-de edad Y-en el año Z- victimizada por el actor J,” que estas listas terminarían teniendo muy pocos registros. No solo porque pocas víctimas cumplirían con esas características, sino porque no todos los registros tienen información en todos los campos, por lo que muchos se quedarían por fuera de estas mini listas. Al dejar registros por fuera, es decir, a una parte de la verdad por fuera, tendríamos un resultado sesgado: sería similar al ejemplo de no incluir a los afrocolombianos o a la población masacrada. En la siguiente subsección detallamos el por qué esto ocurre.

Aunque existe la posibilidad de que la estimación no esté completamente libre de sesgos, como ocurre en cualquier investigación cuantitativa, las ventajas frente a un análisis descriptivo son evidentes por al menos dos motivos. Primero, hay formas de identificar cuándo una estimación puede estar sesgada. Se puede comparar el número de registros previos a la estratificación vs. después de la estratificación. Así, es posible identificar si se está dejando un alto porcentaje de registros por fuera y si se podría estar cayendo en una situación de sesgo. Como se mencionó previamente, la estratificación es una tarea que requiere encontrar un balance: mientras que crear pequeñas listas según características que podrían ayudarle al modelo a reconocer e integrar la variación en la probabilidad de captura para reducir el sesgo, crear demasiadas de estas listas le

quitaría información al modelo y aumentaría el sesgo. Es necesario probar con distintas estratificaciones hasta encontrar el balance óptimo.

La segunda ventaja está relacionada con la pregunta clave del sesgo: si se está en esta situación ¿se estaría subestimando o sobrestimando la cantidad de víctimas? En el caso específico del modelo utilizado en el proyecto, la ciencia ha mostrado que las estimaciones suelen tener un sesgo hacia abajo. Es decir, suelen subestimar la cantidad de víctimas.<sup>32</sup>

### 7.1.1. Detalles sobre sesgo por sobreestratificación

A lo largo del proyecto observamos que en algunos casos la estimación del subregistro tendía a ser muy pequeña. Es decir, que el modelo sugería que casi todos los casos estaban siendo documentados en al menos una lista. Pero, al agregar varios estratos, el subregistro aumentaba. Por ejemplo, si estimábamos por separado los años entre 1986 y 1989 encontrábamos un bajo subregistro. No obstante, al hacer la estimación para el periodo entre 1986 y 1989 encontrábamos un mayor subregistro.

Como mencionamos en la Sección 6.1 el algoritmo de ESM que utilizamos es el de “Modelos de Clases Latentes Múltiples para Captura-Recaptura” (Manrique-Vallier 2016). Nos contactamos con el profesor Manrique-Vallier para discutir los hallazgos que obteníamos al estimar los años de manera individual y de manera conjunta. Manrique-Vallier nos explicó que, en efecto, el modelo LCMCR, tal como se implementa actualmente, tiende a producir estimaciones conservadoras siempre que se ajusten con pocos registros y varias listas.

El motivo de este comportamiento es la especificación a priori de las probabilidades de captura a nivel de listas. Estas probabilidades se modelan para que las a priori sean independientes y uniformes en un intervalo  $(0, 1)$  para representar la ignorancia de su valor verdadero. Sin embargo, cuando se toman muchas listas en conjunto, esta misma especificación implica que algunas de estas listas pueden tener una alta probabilidad de capturar a individuos. Entonces, cuando hay un alto número de listas, la probabilidad a priori de la mayoría de individuos de ser capturados por al menos una lista tiende a ser considerable. Esto significa que es probable considerar de manera a priori que haya una enumeración completa.

Este efecto no es problemático siempre que haya suficientes datos para superar la especificación de la probabilidad a priori. Sin embargo, cuando hay un gran número de listas (lo que significaría que la a priori sugeriría una enumeración completa) y pocos registros, la a priori puede ser demasiado fuerte. Entonces, la a priori sería difícil de superar, lo que resultaría en un estimación conservadora cercana al valor observado.

El profesor Manrique-Vallier nos explicó que por eso a veces nos encontrar diferencias importantes entre estimaciones finamente estratificadas y estimaciones menos estratificadas. Si consideramos el mismo número de listas, la especificación a priori sugerirá una enumeración casi completa con la misma fuerza, independientemente del número de registros en los subestratos. Por lo tanto, si tenemos un gran número de listas (que sugiere una enumeración completa), y algunos subestratos con muy pocos registros, las estimaciones en estos subestratos podrían ser más conservadoras que en el agregado, lo que resulta en disparidades.

Hemos sido conscientes de este desafío y hemos creado estratos que parecen tener suficientes datos para que podamos evitar esta limitación. Aunque no siempre es posible, para la mayoría de los estratos, hemos sido capaces de hacer estimaciones adecuadas.

En un trabajo futuro, el profesor Manrique-Vallier sugiere que creará mecanismos para establecer probabilidades a priori más específicas en  $\lambda_j$  en lugar de asumir la a priori “plana.”

## 7.2. Varianza

Una estimación parte de información parcial de la realidad y tiene un grado de incertidumbre (o “varianza” en términos estadísticos) ¿Qué significa “incertidumbre” en el contexto de la estadística? Significa que existe una parte de los resultados que es consecuencia de la aleatoriedad debida a los registros documentados, es decir, si

---

<sup>32</sup>Ver Rivest (2011) y la sección de limitaciones.

los registros documentados fueran otros de algunas víctimas que no se tienen entonces el resultado esperado sería diferente. Pero ¿qué tanto podrían cambiar los resultados dependiendo de los registros documentados que son analizados? Esto se puede medir a partir de la varianza y con ella determinar la incertidumbre de la estimación, de esta manera el resultado no es un valor específico sino un rango donde se espera que se encuentre la verdad.<sup>33</sup> La incertidumbre nos permite entonces conocer todos los escenarios posibles con un grado de certeza.

Muchas veces las personas desconfían de un rango por no responder a la pregunta de cuál es *la* cifra. Sin embargo, esta desconfianza desconoce que si hiciéramos estadística descriptiva solo basados en las observaciones, no se estaría conociendo la verdad sino lo que ha sido documentado. Entonces, no habría varianza, pues hay certeza absoluta de qué es lo que se documentó. Sin embargo, existiría un sesgo inmensurable: lo que se documentó no refleja la realidad, pero no habría forma de calcular qué tan distante es lo ocurrido en el conflicto de lo documentado. En cambio, gracias a las estimaciones es posible reducir la incertidumbre de cuántas víctimas quedaron por fuera a un rango medible e interpretable. Este rango nos permite conocer patrones y tendencias.

En el caso específico de este proyecto, una de las estimaciones de mayor interés es el universo de víctimas. Es decir, tanto las víctimas documentadas por las organizaciones y entidades como las que no fueron registradas. Al igual que cualquier estimación, aquí tenemos incertidumbre (nuestras estimaciones se representan en rangos). Pero, ¿por qué existe la incertidumbre? Hay 4 motivos para ello:

- 1. La relación entre víctimas documentadas y el universo de víctimas: Entre más registro haya de las víctimas frente al universo (el total de víctimas tanto las documentadas como las no documentadas) menor será la incertidumbre. Podemos pensar esto intuitivamente: el hecho de que las organizaciones e instituciones tengan la capacidad, en agregado, de mostrar una situación más cercana a la realidad, menor será la incertidumbre sobre cómo es la realidad. Esto mismo le sucede al modelo: a medida que tiene menos información faltante sobre la realidad, pues las organizaciones lograron capturar un alto porcentaje de los casos totales (dejaron un porcentaje bajo sin registrar), tendrá mayor certeza sobre cuántos casos podrían haberse quedado sin documentar. Por lo tanto, la varianza es menor entre mayor es la relación entre víctimas registradas y total de víctimas.
- 2. La densidad de información en las listas: Entre mayor coincidencia haya en las víctimas registradas por distintas organizaciones, menor será la varianza. Para comprender la intuición detrás de esto, podemos imaginar que la estimación busca responder a qué tan grande es un cuarto oscuro.<sup>34</sup> La forma de saber el tamaño es a través de pelotas (equivalentes a los registros de cada lista), que son lanzados en el cuarto. A medida que escuchamos más golpes entre las pelotas, esto significará que el cuarto es más pequeño: hay menos espacio para que se muevan sin encontrarse con otra. Esta intuición es trasladable al caso en cuestión. A medida que hay más coincidencia entre las listas, en otras palabras mayor densidad, menor incertidumbre tendrá la estimación.<sup>35</sup>
- 3. La variación de la probabilidad de observación: Aunque nuestro modelo es el mejor en la actualidad para incluir la variación en la probabilidad de observación, es posible que con la información documentada, como sexo, etnia, año y municipio, no sea suficiente para mostrar las diferencias en la probabilidad de observación. Para estimar el universo de víctimas, tomamos la intersección de todas las listas. Es decir, sabemos con exactitud en qué listas está documentada la víctima “Juan Pérez, asesinada el 2 de mayo de 1990 en Medellín, Antioquia.” Según la cantidad de intersecciones entre listas, es decir, de coincidencias entre registro, habrá más o menos incertidumbre. Si tenemos más coincidencias, la varianza será menor: intuitivamente las listas están capturando a la misma

---

<sup>33</sup>Esta afirmación se basa en la interpretación de la estadística bayesiana de “intervalo de credibilidad” Los frequentistas utilizan “intervalos de confianza,” que dan una interpretación distinta de la incertidumbre.

<sup>34</sup>HRDAG utiliza esta analogía en la mayoría de los estudios en los que se usa estimación por sistemas múltiples.

<sup>35</sup>Podemos conocer en qué bases de datos está cada víctima gracias a que en el proceso de deduplicación generamos una única base de datos que tiene en las filas a cada una de las víctimas registradas y en las columnas, además de su información, una columna para cada base de datos del proyecto. Estas celdas toman el valor de 0 si la víctima no estaba registrada en la base y 1 en caso contrario.

población, por lo que es probable que el universo de víctimas sea similar al documentado. En este caso habrá menor incertidumbre. Por lo tanto, la incapacidad de capturar en su totalidad la variación de la probabilidad de observación genera incertidumbre.

- 4. Imputación estadística de campos faltantes: Para conocer el universo de víctimas tenemos que partir de una gran base de datos, que es el cruce de todas las disponibles eliminando registros duplicados. En esa gran base tenemos que completar la información faltante de los registros. Es decir, asignar la raza, etnia, sexo, municipio y si el hecho ocurrió o no dentro del marco del conflicto armado a los registros que no tengan información en estos campos. A esta tarea de completar los campos faltantes se le conoce como “imputación” fue explicada en detalle en la Sección 4. A medida que haya más registros con información completa frente a los registros con campos faltantes, menor será la incertidumbre.

En resumen, existen 4 fuentes de incertidumbre que llevan a que las estimaciones se muevan dentro de un rango. Si bien es tentador optar por utilizar los datos de los que hay registro y limitarse a la estadística descriptiva, al hacer esto dejaríamos por fuera a todas las víctimas no documentadas, incurriendo así en un sesgo. Las cuatro fuentes de incertidumbre comprenden a la relación entre víctimas documentadas y el universo de víctimas, la densidad de listas con las que contamos, la variación en la probabilidad de observación y la imputación estadística de campos faltantes. Aunque la incertidumbre no nos permite responder a cuál es *la cifra*, sí nos permite identificar tendencias y patrones. Además, los rangos tienen la característica de que los números más cercanos al centro son más probables.

## 8. Limitaciones y trabajo futuro

*“En todas las ficciones, cada vez que un hombre se enfrenta con diversas alternativas, opta por una y elimina las otras (...) Crea, así, diversos porvenires, diversos tiempos, que también proliferan y se bifurcan.” (Jorge Luis Borges, 1944)*

En el cuento “Los jardines que se bifurcan” de Jorge Luis Borges los personajes discuten sobre un posible laberinto creado por los ancestros del protagonista. El laberinto representa la idea de que cada vez que los personajes llegan a un punto en el que deben elegir dos caminos, la historia se separa en dos copias y los personajes siguen ambos caminos.

Este cuento ha sido usado por grandes estadísticos, como [Gelman and Loken \(2014\)](#), para describir el proceso de la investigación científica. Cada vez que los investigadores toman una decisión sobre el diseño del proyecto se sigue un camino, aunque existe otro. Somos conscientes de que hay muchos caminos que no tomamos en nuestro trabajo. Cada vez que lo hicimos dejamos senderos sin explorar, que podrían haber llevado a otros resultados.

A lo largo de este documento hemos mostrado las decisiones que tomamos y hemos justificado el por qué consideramos que era el mejor camino. Consideramos que elegimos los caminos que tienden a influenciar menos en los resultados y, por lo tanto, nuestros resultados deberían permanecer tan cercanos a los datos originales como los modelos lo permitan. Nuestros resultados reflejan la combinación de supuestos y decisiones que tomamos. Creemos que los supuestos son posibles y plausibles y que las decisiones son las mejores. Sin embargo, hay otros caminos posibles.

En esta sección compilamos esos momentos y presentamos las limitaciones del proyecto a partir de los datos, la vinculación de registros, la imputación estadística y la estimación por sistemas múltiples. Para cada uno de estos elementos, retomamos las decisiones claves, el por qué, los otros posibles caminos y las consecuencias de haber optado por una decisión en particular.

### 8.1. Datos

Cualquier investigación cuantitativa se ve limitada por los datos que posea. Partimos de un supuesto fundamental e identificamos cinco limitaciones relacionadas con los datos.



El supuesto del que partimos en el proyecto es que la gran mayoría de los registros que recibimos son hechos de violencia que efectivamente ocurrieron. No estamos preocupados de que no lo sean porque no hayan sucedido, sino porque no representen lo que la base nos dice que representan. Por ejemplo, porque podría haber registros dentro de bases de homicidio que hayan sido tentativas de homicidio o porque el año que se dejó en la base sea el año de denuncia y no de ocurrencia, entre otros posibles errores sistemáticos de los datos.

En cuanto a las limitaciones, lo primero es que no fue posible incluir los datos de SIJUF (de la Fiscalía) en los análisis. Tampoco pudimos incluir los datos del SPOA para hechos de violencia distintos a “homicidio consumado.” Por lo tanto, por el diseño del sistema de la Fiscalía, fueron excluidas dos de las fuentes de información más grandes no especializadas del país. Esto lo explicaremos en detalle a continuación en la subsección 8.1.1.

Segundo, como se mostró en la sección de violencia registrada, los datos de desplazamiento se concentran en el RUV. Dado que el modelo de estimación por sistemas múltiples se basa en las coincidencias (y no coincidencias) entre listas, las estimaciones se ven afectadas por la alta importancia del RUV. Entonces, recomendamos trabajar con las imputaciones estadísticas de esta violencia. Es decir, no fue posible calcular el subregistro.

Tercero, aunque uno de los hechos de violencia que se buscó incluir en el proyecto es el exilio, no hay suficientes datos. Entonces, este hecho de violencia fue excluido de la investigación.

Cuarto, si bien desde la CEV se está estudiando el conflicto desde 1958, los datos previos a 1985 son escasos. Aunque es posible realizar la imputación estadística de campos faltantes y las estimaciones, la varianza es muy alta, por lo que no es recomendable utilizar los análisis del proyecto para este periodo de tiempo. Esta es la primera decisión que tomamos e implica que, cada hecho de violencia tiene datos adecuados para distintos periodos. Como mostramos en las tablas, homicidio tiene la mayor parte de datos disponibles entre 1985 y 2018, desaparición forzada entre 1985 y 2016, secuestro entre 1990 y 2018, reclutamiento entre 1985 y 2017, y desplazamiento entre 1985 y 2019, aunque sin imputación ni estimación. Los datos fuera de este periodo llevaban a imputaciones y estimaciones no estables y por lo tanto poco confiables.

Por último, una de las hipótesis que se sostienen sobre la victimización es que las actividades a las que se dedica la víctima; como ser defensora de DDHH, periodista o sindicalista; es un elemento fundamental y diferenciador. Sin embargo, las bases de datos con esta información son escasas, por lo que no es posible hacer la imputación estadística de campos faltantes de esta variable. No obstante, hay algunas bases especializadas en cierta población. Este es el caso, por ejemplo, del sindicalismo. Para las situaciones en las que hay bases especializadas en ciertos sectores, pudimos realizar análisis específicos. Pero, este no es el caso para todas las poblaciones de interés y hay grupos de importante relevancia que no pudieron ser analizados.

### 8.1.1. Sobre los datos de la Fiscalía General de la Nación

Dos de las grandes bases de datos oficiales en el país que incluyen hechos de violencia son el Sistema Penal Oral Acusatorio (SPOA) y el Sistema de Información de Justicia de la Fiscalía (SIJUF) de la Fiscalía General de la Nación. Sin embargo, estos sistemas utilizan como fuente de observación los **casos** y no las **personas** ni las **violaciones**. Es decir, en un mismo caso hay más de una persona. Por ejemplo, la víctima y sus padres, la víctima y los denunciantes, etc. Entonces, no se puede responder a la pregunta de “¿quién le hizo qué a quién?” (Ball 2008).

No obstante, la Fiscalía sí cuenta con una variable de “homicidio consumado” dentro del SPOA. Esto nos permitió usar los homicidios consumados. Es decir, no tenemos datos de SPOA para hechos de violencia diferentes a homicidio. En cambio, por las cuestiones de diseño del SIJUF, no pudimos incluirlo en el análisis.

La Fiscalía cuenta con otras bases de datos que fueron entregadas al proyecto aparte del SPOA y SIJUF. En específico, se trata de bases de datos relacionadas con reclutamiento, que fueron incautadas por la institución. Estas bases de datos sí fueron utilizadas dentro del proyecto.

Además, la JEP ha realizado una tarea de verificación de los casos de Unión Patriótica (UP) y de líderes sociales registrados por la Fiscalía, por lo que estos también fueron incluidos.

En síntesis, el SPOA y el SIJUF están diseñados para estudiar “casos” y no “víctimas.” Por lo tanto, utilizamos la variable “homicidio consumado” para identificar a las víctimas de homicidio consumado del SPOA. Entonces, de Fiscalía incluimos únicamente a las víctimas de homicidio consumado, y las bases de reclutamiento, líderes y UP.

## 8.2. Vinculación de registros

El resultado de la vinculación de registros, descrita en la Sección 3, es la base a partir de la cual se hacen la imputación estadística de campos faltantes y la estimación por sistemas múltiples. Identificamos dos decisiones particulares que tomamos, que podrían haber sido distintas: los campos mínimos a utilizar y la relación entre faltos positivos y negativos. A continuación, explicamos cada una de ellas.

### 8.2.1. Campos mínimos

Para poder hacer la vinculación de registros es necesario tener unos campos mínimos. En específico, nosotros optamos por el nombre1, apellido1, año, departamento y tipo del hecho de violencia. No obstante, no todos los registros cuentan con estos campos y, por lo tanto, hay víctimas del conflicto que no son incluidas de manera explícita dentro del proyecto. Sin embargo, estas están incluidas en el subregistro al hacer la ESM.

Podríamos haber tomado otros campos mínimos, como nombre, apellido y año, o nombre, departamento y año. Sin embargo, consideramos que dado el gran número de registros era necesario contar con más información para que el oráculo pudiera generar los datos de entrenamiento certeros. Creemos que con las cinco variables que escogimos hay un buen balance entre la cantidad de información con la que cuenta el oráculo y las víctimas que son excluidas por no contar con todos los campos mínimos.

### 8.2.2. Falsos positivos vs. falsos negativos

Este trade-off es específico para el modelo de vinculación de registros, que es de clasificación. Como su nombre lo indica, el algoritmo busca clasificar a un objeto de interés en un grupo. En el caso de la vinculación de registros, buscamos responder a si un registro forma parte o no de un mismo bloque.

En la ciencia de datos existen distintas medidas útiles para estudiar el desempeño de un modelo. Como expusimos en la Sección 3 entre ellas, se encuentran la exhaustividad (*recall* en inglés) y la precisión (*precision* en inglés). La exhaustividad nos muestra el rendimiento de un clasificador frente a los falsos negativos. Es decir, los registros que son identificados como víctimas distintas, aunque sean la misma víctima. Por su parte, la precisión nos muestra el rendimiento frente a los falsos positivos. Es decir, los registros que son identificados como una misma víctima aunque no lo sean.

Es posible priorizar a una de las dos medidas. Por un lado, se puede decidir enfocarse en minimizar los falsos negativos (maximizando la exhaustividad) o se puede minimizar a los falsos positivos (maximizando la precisión). En el caso de la vinculación de registros, optamos por un balance entre las dos, aceptando una tasa ligeramente más alta de falsas coincidencias, en vez de dejar de identificar a distintas personas. Elegimos esto para tener un balance entre el riesgo de sobre-identificar y sub-identificar. Es decir, entre no reconocer que dos o más registros pueden ser la misma víctimas, y que dos o más registros son la misma víctima. En la Tabla 17 presentamos el resultado del modelo, donde se evidencia el balance entre falsos negativos y falsos positivos, con un poco más de falsos positivos. Esta es otra de las decisiones que podría haber sido distinta. Aquí los posibles caminos que se habrían serían enormes según la proporción entre falsos negativos y falsos positivos. Entonces, existe una gran cantidad de posibles escenarios de datos deduplicados, según este balance. Desde el proyecto creemos que la mejor decisión era permitir una tasa ligeramente de falsas coincidencias sin irnos a los extremos. De lo contrario, podríamos haber terminado con un doble conteo de víctimas o un subconteo. Sin embargo, a partir del uso de otra deduplicación los resultados serían distintos. Por último, es importante mencionar que la literatura ha mostrado que incluso si la vinculación de registros tiene un desempeño pobre, los resultados de la ESM son confiables (J. Johndrow, Lum, and Dunson 2018).



### 8.2.3. Casos de ambigüedad

Uno de los temas clave de la imputación estadística de campos faltantes son las decisiones por tomar en los casos de ambigüedad. Como mostramos en la Sección 3, el último paso de la vinculación de registros es la unión y exportación. Ahí es necesario decidir qué hacer con los registros que tienen ambigüedad en alguna variable. Es decir, que tienen distinta información en un mismo campo. La decisión de la priorización de los hechos de violencia la tomamos con base en el análisis de los expertos en conflicto. En específico, se decidió que si al hacer la deduplicación tenemos para un mismo registro homicidio y desplazamiento, se toma homicidio. Si un mismo registro tiene reclutamiento de niños, niñas y adolescentes, desaparición y secuestro, se toma reclutamiento. Mientras que se tiene desaparición forzada y secuestro, se toma desaparición forzada.

Por su parte, la priorización de la etnia la hicimos con el objetivo de priorizar a los grupos históricamente discriminados. En caso de que un registro tuviese dos etnias distintas se priorizó indígena, seguido por afrocolombiano y por Rrom. A partir de la experticia de investigadores en cuestiones étnicas aprendimos que en algunos casos las personas indígenas, afrocolombianas o Rrom descubren que han sido registradas en bases de datos como “mestizo.” Entonces, decidimos priorizar las referencias a indígena, afrocolombiano y Rrom cuando ocurrieron en pares correferentes. La ambigüedad no solo puede estar en los hechos de violencia o etnia. Aplica también para otras variables, en las que tomamos el valor más frecuente. Mientras que, si había un empate, elegimos aleatoriamente. Sin embargo, estas reglas podrían haber sido distintas y los resultados de la imputación estadística de campos faltantes también lo serían. No obstante, creemos que tomamos las mejores decisiones en línea con el conocimiento sobre conflicto, con el reconocimiento a los pueblos étnicos y la probabilidad.

## 8.3. Imputación estadística

Una vez se hizo la deduplicación, algunos registros seguirán teniendo campos faltantes. Entonces, es necesario hacer la imputación estadística de campos faltantes. Las decisiones que tomar en este paso se pueden dividir en tres grandes grupos: el uso de bases especializadas y no especializadas, las relacionadas con las variables a utilizar para dar información en la imputación estadística y las relacionadas con el algoritmo a utilizar.

### 8.3.1. Bases especializadas vs. bases no especializadas

Una de las grandes decisiones a tomar en el proyecto fue si limitarnos al uso de bases especializadas en conflicto armado o utilizar también las no especializadas. Decidimos incluir los dos tipos de bases para poder incluir una cantidad importante de datos oficiales. En específico, los provenientes del INML-CF, la FGN y PONAL. La consecuencia directa de esto es que tuvimos que imputar estadísticamente la variable “pertenece al conflicto,” como explicamos en la próxima subsección.

### 8.3.2. Extensiones de la imputación

Como mostramos en la Sección 6.2 distintas variables tienen distintas proporciones de campos faltantes. Hay tres que resultan especialmente críticas: pertenece al conflicto; es desaparición forzada; y el presunto responsable.

Lo primero a decidir se relaciona con los datos de entrenamiento. Como explicamos en la Sección 6.2, para los casos de pertenece al conflicto y es desaparición forzada nosotros elegimos reglas probabilísticas con base en el criterio de expertos y expertas. Sin embargo, podría haberse hecho de manera determinística o con otras reglas. Pensamos que tomamos las mejores decisiones por dos motivos. Primero, porque las reglas fueron creadas con base en el conocimiento de las dinámicas del conflicto por investigadores expertos y expertas de la CEV. Segundo, porque al ser probabilísticas reconocen que no necesariamente todos los hechos que cumplen con una regla cumplen o no con la condición. Entonces, aunque confiamos en el criterio de los y las expertas que definieron las reglas, debemos reconocer que, si tomáramos otras reglas, la imputación estadística de campos faltantes también cambiaría.

El otro gran reto se relaciona con cómo lidiar con el presunto responsable. En nuestro caso, decidimos dividir a las guerrillas en grupos: FARC-EP, ELN y “otras.” Sin embargo, podríamos haberlas agrupado como “guerrillas” y los resultados serían distintos. Dada la magnitud de la violencia registrada de las FARC-EP y del ELN en los años de estudio respecto a los otros grupos guerrilleros, consideramos que es importante analizar sus casos específicos como un ejercicio de contribución a la verdad. En la siguiente subsección damos más detalles de esta decisión.

Decidimos utilizar todas las variables posibles para la imputación. De hecho, solo eliminamos las variables que fueran constantes o tuvieran colinealidad. La imputación estadística de datos faltantes la podríamos haber hecho únicamente a partir de ciertas variables. Pero, creemos que, al darle información más completa al modelo, sin caer en la maldición de la dimensionalidad agregando un gran número de variables gracias a las variables de soporte, podemos llegar a mejores resultados.

Entonces, una de las decisiones fundamentales que tomamos fue la de crear variables de soporte para ayudar la imputación estadística de campos faltantes (ver Sección 4.4). Nosotros optamos por esta aproximación para poder tomar toda la información de las bases originales y aprender a identificar pistas que nos puedan ayudar a responder si un hecho de violencia ocurrió o no en el marco del conflicto armado, si una desaparición es o no forzada, el sexo de la víctima, si es o no menor de edad, el presunto responsable y el municipio en el que ocurrió el año. En el caso del conflicto armado o de la desaparición forzada, una pista por ejemplo podría ser que en el relato se incluya que la persona tenía fuertes posiciones políticas. También puede ser la mención a enfrentamientos específicos o el hecho de referirse a los grupos como “subversivo” o “terrorista.” Queremos enfatizar que estas palabras o frases no son elegidas por nosotros. Estas no son unas reglas, tampoco son identificadas por humanos. Son identificadas por el modelo porque las palabras y frases se correlacionan con la variable a predecir. Creemos que esta es la mejor decisión porque permite no perder información valiosa de los registros originales, especialmente el texto, pero la imputación estadística de campos faltantes podría hacerse sin ayuda de las variables de soporte.

### 8.3.3. Imputación guerrilla sin especificar

Como mencionamos previamente, algunos registros tienen como presunto responsable a “guerrilla” sin especificar qué grupo. Distinguir entre diferentes grupos guerrilleros es esencial para el estudio de patrones de violencia, que podría ser distinto para cada grupo armado inclusive entre los varios grupos guerrilleros (Gutiérrez-Sanín and Wood 2017). Entonces, utilizamos el modelo de imputación estadística con la ayuda de las variables de soporte detalladas en la Sección 4 para asignar estos casos a un grupo armado específico. Dado el periodo de tiempo de análisis del proyecto, nos limitamos a 3 categorías: FARC-EP, ELN y otras guerrillas.

Previo a hacer la imputación del presunto responsable para los casos que tienen “guerrilla” sin especificar, estudiamos los registros antes de exportar los resultados del proceso de vinculación. En ese punto conocemos a todos los registros de las varias fuentes que pertenecen a la misma víctima. Encontramos que el comportamiento de las víctimas con por lo menos un registro de “guerrilla” sin especificar como el presunto responsable difiere de lo de las víctimas con por lo menos un registro que pertenece a un grupo guerrillero específico. Al estudiar todos los registros que pertenecen a una misma víctima, en el 0,44 de los casos en los que por lo menos uno de los registros tiene “guerrilla” sin especificar hay por lo menos otro registro que pertenece a una misma víctima con los paramilitares, el Estado, “múltiple,” u otro grupo (como bandas criminales) como el presunto responsable. Es decir, la información del presunto responsable contenida en los varios registros que refieren a la misma persona es contradictoria. En cambio, esta contradicción aparece en el 0,15 de las víctimas con por lo menos un registro con “guerrilla-FARC,” y 0,2 de las víctimas con por lo menos un registro con “guerrilla-ELN.” Es decir, los registros de “guerrilla” sin especificar están vinculados a grupos distintos a guerrillas (principalmente paramilitares) con una tasa más del doble que los registros que sí tienen a un grupo guerrillero específico. Esta alta tasa de contradicción nos hace ser escépticos de la calidad de la información de la variable “guerrilla” sin especificar. Entonces, incluimos esta información en las variables de soporte para que los registros pudieran ser imputados estadísticamente.

Hay varios mecanismos que pueden explicar la inexactitud de estos registros. Por ejemplo, puede ser que al reportar la violencia, el denunciante no sabe con certeza quién fue el grupo responsable o confunde un

grupo con otro. También puede ser una decisión consciente para reducir el riesgo de represalia por denunciar la violencia de un grupo específico. Estas son especulaciones, no sabemos por qué hay esta deficiencia en estos registros, solo sabemos que la información “guerrilla” sin especificar en estos casos no es consistente con otros registros vinculados a la misma víctima.

Aunque elegimos imputar el presunto responsable con el modelo estadístico detallado en la Sección 4, podríamos haber tomado otra aproximación, como asignar los casos con “guerrilla” sin especificar aleatoriamente a FARC, ELN u otro grupo guerrillero, siguiendo la proporción de los datos observados. Es decir, 0,8 a FARC-EP, 0,15 a ELN y 0,05 a otras guerrillas. A partir de ahí es posible hacer la imputación de los homicidios. Encontramos que la relación de responsables cambia de la siguiente forma: la estimación de la proporción atribuida a los paramilitares se reduce de 0,45 a 0,4<sup>36</sup> y la estimación para FARC aumenta de 0,21 a 0,27.<sup>37</sup> El alto nivel de registros con “guerrilla” sin especificar y presuntos responsables distintos a guerrilla afecta especialmente a la categoría de “múltiple,” aumentándola de 0,06 to 0,12;<sup>38</sup> mientras que los demás presuntos responsables se reducen un poco. Observamos que utilizando esta imputación proporcional, las proporciones de la presunta responsabilidad estimada para los paramilitares y las FARC-EP están dentro de los intervalos de confianza estimados sin usar “guerrilla” sin especificar. Entonces, el impacto principal de usar “guerrilla” sin especificar es que se aumenta la categoría de “múltiple.”

Para ser claros, este experimento no es un hallazgo confiable. Es un análisis de sensibilidad para mostrar el impacto máximo posible de la asignación de la información de “guerrilla” sin especificar. Usar esta imputación no cambiaría el hallazgo sustancial de que los paramilitares son los principales presuntos responsables de homicidios, con el FARC-EP en segundo lugar. Lo que este experimento enfatiza es que usar “guerrilla” sin especificar de manera directa amplifica la contradicción en el campo de presunto responsable, llevando a una estimación mayor de responsables múltiples.

Vale la pena estudiar dos problemas claves con esta aproximación. Primero, la asignación de las proporciones no incluye el alto nivel de contradicción que observamos. Este experimento asume que la probabilidad de un grupo específico dentro de “guerrilla” sin especificar es la misma que en los datos observados de los grupos que sí tienen una guerrilla específica asignada, aunque el detalle esté faltando. No sabemos la distribución verdadera, pero es muy probable que una fracción sustancial de “guerrilla” sin especificar deba ser asignada a múltiple o a paramilitares.

Segundo, como mostramos, los datos de la variable “guerrilla” sin especificar tienen incertidumbre. Estamos seguros de que algunos registros deberían ser atribuidos a FARC-EP y una fracción menor a ELN. Sin embargo, otros registros deberían ser atribuidos a los paramilitares, otros grupos y agentes del Estado. No sabemos la proporción correcta para los distintos actores. Determinar estas proporciones y determinar su certeza es exactamente la tarea de la imputación estadística. En particular, si hay incertidumbre en los datos el mejor resultado es que las proporciones estimadas tengan intervalos de confianza lo suficientemente amplios para representar el rango de posibles valores. Nosotros le damos la información al modelo de imputación por medio de las variables de soporte y de estratificación y el modelo nos muestra el rango de variación. Algunas veces necesitamos asignar algunos casos no certeros con reglas de probabilidad a partir de expertos, como lo hicimos para “pertenece al conflicto” y “es desaparición forzada.” En estos casos tratamos de asignarle un tamaño a los intervalos. Si no conocemos el valor verdadero de una variable, es mejor dejarlo como desconocido para que pueda contribuir a la incertidumbre, medida por el intervalo de confianza.

Hicimos otros experimentos con “guerrilla” sin especificar, como asignar la proporción a las tres categorías de guerrilla y algunas a múltiple, dejando otras para el modelo de imputación estadística. Todos estos experimentos llevaron a resultados similares a los que mostramos en la Tabla 13. Es decir, en cada uno de nuestros experimentos, las estimaciones para los grupos paramilitares y las FARC-EP están dentro del intervalo de confianza mostrado en la Tabla 13. Esto significa que para los resultados de paramilitares y FARC-EP, en un sentido estadístico, los experimentos no fueron significativamente distintos que los resultados de la Tabla 13. Hay dos diferencias consistentes entre nuestros experimentos y la Tabla 13: la categoría múltiple siempre es mayor que en la Tabla 13 y los intervalos de confianza son más pequeños. Ninguno de estos resultados nos parece probable, por lo que preferimos la aproximación de las variables de soporte.

<sup>36</sup>Con un intervalo de confianza del 95 % de 0,4 (0,35–0,45).

<sup>37</sup>Con un intervalo de confianza del 95 % de 0,27 (0,23–0,32).

<sup>38</sup>Con un intervalo de confianza del 95 % de 0,12 (0,09–0,14).

La mejor aproximación sería un modelo de imputación estadística más avanzado en el que la información de guerrilla sin especificar guiara la imputación estadística sin sobredeterminar el resultado. Usamos las variables de soporte para esto y consideramos que es la mejor manera de hacerlo. Con los variables de soporte, no estamos quitando la información de “guerrilla” de los registros, sino utilizándola con el resto de la información contenido en el registro para hacer una predicción probabilística del presunto responsable. También el supuesto sobre el comportamiento de los datos faltantes en nuestro modelo de imputación—“valores faltantes aleatorios”—es mucho más justificable que el supuesto improbable requerido para hacer la imputación de estos casos de una manera proporcional y no ignora que algunos de estos registros probablemente corresponden a los paramilitares, el estado u otro grupos armados. Sin tener más información sobre estos registros que sólo reporten “guerrilla” como el presunto responsable, nuestra aproximación a la imputación es el más riguroso técnicamente que es posible con la información y la tecnología que tenemos a nuestra disposición.

#### 8.3.4. Modelo de clasificación

Existen distintos modelos de clasificación que se pueden utilizar para la especificación condicional completa de la imputación estadística de campos faltantes. Hasta la Fase 3 del proyecto se había utilizado CART, un árbol de decisión que tiene distintas ventajas. Entre ellas se encuentra que es robusto a la colinealidad entre las variables predictoras y la capacidad de ajustar datos de diferentes tipos (binarios, categóricos, o continuos). CART también ajusta relaciones no lineales e interacciones complejas sin asumir hipótesis sobre los parámetros (Burgette and Reiter 2010; Akande 2015). Sin embargo, su costo computacional es elevado. En la Fase 4 del proyecto optamos por utilizar PMM, tal y como mencionamos en la Sección 6. Esto, porque tiene distintas ventajas sobre CART. Primero, preserva la varianza mejor que CART. Entonces, podemos confiar más en él al conocer el tamaño del error. Segundo, es cerca de 15 veces más rápido, lo que computacionalmente es importante dada la altísima cantidad de registros que estamos utilizando en el proyecto. Tercero, PMM tiene las ventajas de CART de capacidad de ajuste a diferentes tipos de datos y ser robusto a modelos pobremente especificados. Entonces, logramos tener resultados más eficientes. Sin embargo, los resultados variarían de acuerdo con el algoritmo que se seleccione.

### 8.4. Estimación por sistemas múltiples

El tercer y último paso es la estimación por sistemas múltiples. Nuevamente, acá habría una serie de posibilidades y nosotros tomamos las decisiones que consideramos las más acertadas. En específico, se relacionan con el método, la estratificación y el método para estimar la varianza. Además, hay una limitación relacionada con la varianza.

#### 8.4.1. Método

Existen distintos métodos de estimación por sistemas múltiples. Nosotros utilizamos el método de LCMCR (Bayesiano No Paramétrico de Captura-Recaptura). Podríamos haber utilizado otro método, como Promedio de Modelo Bayesiano (DGA). DGA permite tener una probabilidad a priori más específica e interpretable (ver Hoover Green and Ball 2019), pero LCMCR nos permite usar más fuentes, pues DGA está limitado a 5 fuentes.<sup>39</sup>

El costo de usar LCMCR es que, como mostramos en la Sección 7, este método tiene un sesgo hacia abajo. Entonces, nuestros resultados subestiman la violencia. No obstante, creemos que era mejor asumir este sesgo para poder aprovechar la rica información que han documentado decenas de organizaciones.

---

<sup>39</sup>También podríamos haber usado un método frecuentista, como [Rcapture](#), pero los métodos frecuentistas están limitados en el número de listas que pueden usar, como DGA, o están limitados en el número de interacciones. Creemos que es difícil justificar esto en la práctica por la cantidad de listas que tenemos. No habría forma de usarlas todas y necesitaríamos justificar la selección de algunas.

### 8.4.2. Estratificación

Entre más variables para estratificar utilicemos, menor heterogeneidad de captura habrá. Pero, el costo de una mayor estratificación es un mayor sesgo. Volvamos al ejemplo utilizado en la Sección 7 en la que había un abogado en Bogotá con probabilidad de captura 0,9 y una campesina en una vereda con probabilidad de captura 0,05. Aquí podríamos estratificar por distintas variables: el año, el departamento, el sexo y, hasta el municipio. Estaríamos entonces creando pequeñas listas de registros que cumplan con las posibles combinaciones de estas variables. Sin embargo, si nos limitamos por ejemplo a los municipios de solo un departamento, podríamos dejar por fuera a víctimas como la campesina con muy baja probabilidad de ser capturadas. Entonces, el modelo no tendrá cómo estimar cuántas tienen probabilidad de no ser registradas. Así, hay una relación entre el tamaño de la muestra (que depende de los estratos) y qué tan bien el modelo estima a todas las víctimas, incluyendo a las que tienen poca probabilidad de ser documentadas.

Por último, es importante aclarar sobre la estratificación que esta debe servir las necesidades analíticas del proyecto. Por ejemplo, si necesitáramos comparar el número de víctimas estimadas por grupo armado, debemos estratificar por grupos armados. Elegir la estratificación siempre es un balance entre la cantidad de datos en el estrato, el control de la heterogeneidad de captura y las necesidades analíticas del proyecto.

### 8.4.3. Método para estimar la varianza

Como mostramos en la Sección 6 seguimos a [Chao \(1987\)](#) para hacer la transformación de los datos y esta es la transformación que aplicamos a las muestras de las distribuciones posteriores para cada una de las diez réplicas. Sin embargo, LCMCR lleva a una estimación bayesiana y, por lo tanto, es posible en teoría hacer una estimación completamente bayesiana de la varianza. Siguiendo a [Zhou and Reiter \(2010\)](#) sería posible combinar todas las muestras de las distribuciones posteriores. Esto permitiría, por ejemplo, tener intervalos de credibilidad en vez de intervalos de confianza. Sin embargo, [Zhou and Reiter \(2010\)](#) proponen un mínimo de 30 réplicas y, en el mejor de los casos, proponen 100 réplicas.

Pero, en promedio, las 10 réplicas que tenemos en cada estrato requieren de entre 3 y 5 horas de CPU para estimar con LCMCR. Si tuviéramos que estimar entre 3 y 10 veces más réplicas, había entre 9 y 50 horas de CPU en cada estrato. En el caso de sexo, si hiciéramos una estratificación por año y sexo entre 1990 y 2015 tendríamos 50 estratos. Esto sugeriría entre 450 y 2.500 horas de CPU solo para este análisis. Con nuestra tecnología actual y presupuesto, esto no es posible.

Con mejoras computacionales sería aconsejable utilizar el método de [Zhou and Reiter \(2010\)](#). En la Figura 19 mostramos la comparación de los intervalos de confianza usando el método de [Chao \(1987\)](#) contra el intervalo de credibilidad usando el método de [Zhou and Reiter \(2010\)](#). En el lado izquierdo observamos las 10 distribuciones posteriores de las 10 réplicas para un mismo estrato. Las líneas negras muestran el intervalo de confianza siguiendo a [Chao \(1987\)](#), mientras que las azules muestran el intervalo de credibilidad siguiendo a [Zhou and Reiter \(2010\)](#). Entonces, vemos que con [Chao \(1987\)](#) la varianza estimada es mayor que la varianza siguiendo a [Zhou and Reiter \(2010\)](#). Por su parte, en el lado derecho vemos todas las posteriores de las imputaciones juntas.

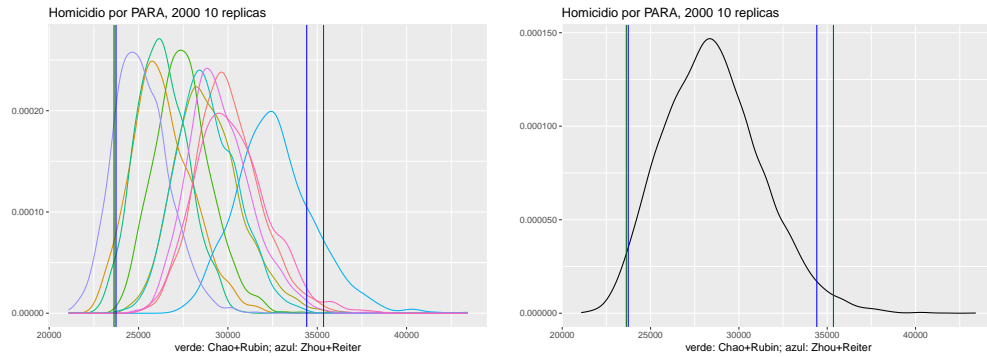


Figura 19: Ejemplo de intervalos de confianza y credibilidad para paramilitares 2002.

La diferencia que vemos entre estas dos aproximaciones podría estar relacionada con una idea de [Rubin \(1996\)](#) llamada “supereficiencia.” Rubin observó que cuando un modelo de imputación usa más información que el modelo de estimación, la varianza estimada por sus reglas tenderá a ser conservadoras. “Conservadoras” en este contexto significa que la varianza estimada podría ser más grande que la varianza verdadera. En nuestro caso, el modelo de imputación usa muchas variables que no están en la estimación. Específicamente, las variables de soporte se usan solo en la imputación. Entonces, las estimaciones de ESM serán “supereicientes.” Entonces, la varianza estimada por el método de Rubin para la estimación tenderá a sobrestimar la varianza, lo que podría explicar por qué el intervalo de credibilidad de [Zhou and Reiter \(2010\)](#) es un poco más pequeño que el intervalo de confianza estimado por el método de Chao y Rubin. Se puede ver a [Cro \(2017\)](#) para más explicaciones de este efecto.

## 8.5. Trabajo futuro

En esta sección hemos enfatizado en las limitaciones actuales y otros caminos por los que podríamos haber optado. Como mostramos, algunas de las limitaciones se dan por capacidad computacional, otras porque la ciencia aun no ha hecho los respectivos desarrollos. Identificamos tres posibles mejoras para un trabajo futuro.

Primero, se podría modificar la distribución a priori en el método LCMCR. Esto, permitiría incorporar la heterogeneidad de captura. Por ejemplo, podría incluirse la información sobre la probabilidad que tienen los registros de ser o no documentados.

Segundo, se podría incorporar la correlación de las observaciones en los estratos. Con el método actual se hace la estimación para cada estrato por separado, suponiendo su independencia. En la vida real, es probable que los estratos estén correlacionados. Esto también nos permitiría tomar información “prestada” entre los estratos para hacer estimaciones en estratos más pequeños. Por ejemplo, si una fuente en particular tuviera una baja probabilidad de observación de homicidios en Casanare en el 2003, es probable que esta fuente también tuviera una baja probabilidad de observación de homicidios en Arauca en 2003 y en Casanare en 2002. Si no hubiera suficiente información para Casanare en 2003, incorporar la correlación de las observaciones nos permitiría tomar información entre los estratos para hacer las estimaciones. Un modelo de este tipo también nos permitiría hacer estimaciones para estratos más pequeños, como municipios o meses específicos.

Tercero, se podrían integrar la imputación estadística de campos faltantes y la estimación en un solo modelo. Como mostramos previamente, el hecho de tener los modelos encadenados lleva a que se propague la varianza de la imputación estadística de campos faltantes en la estimación. Esto podría modificarse al integrar los modelos.

En el artículo [Manrique-Vallier, Ball, and Sadinle \(2022\)](#) se puede ahondar en los avances y trabajo futuro de la estimación por sistemas múltiples.



## 9. Anexo

### 9.1. Agradecimientos

Le agradecemos a los donantes cuyo apoyo hizo posible este trabajo, incluido el apoyo básico del John D. y Catherine T. MacArthur Foundation, The Oak Foundation y el Foundation to Promote Open Society, y el apoyo para este proyecto específico de la División de Seguridad Humana de la Embajada de Suiza, el Secretario de Estado de Asuntos Exteriores, Commonwealth y Desarrollo del Reino Unido y la Universidad Justus Liebig de Giessen, con fondos otorgados por el Ministerio Federal de Relaciones Exteriores de Alemania.

### 9.2. Sobre las organizaciones que participaron en el proyecto

#### 9.2.1. Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición (CEV)

En el marco del Acuerdo Final para la terminación del conflicto y la construcción de una paz estable y duradera, suscrito entre el Gobierno de Colombia y las FARC-EP, mediante el Acto Legislativo 01 de 2017 y el Decreto 588 de 2017, se creó la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición. Esta es un mecanismo de carácter temporal y extrajudicial del Sistema Integral de Verdad, Justicia, Reparación y No Repetición (SIVJRNR), para conocer la verdad de lo ocurrido en el marco del conflicto armado y contribuir al esclarecimiento de las violaciones e infracciones cometidas durante el mismo y ofrecer una explicación amplia de su complejidad a toda la sociedad.

#### 9.2.2. Jurisdicción Especial para la Paz (JEP)

La Jurisdicción Especial para la Paz (JEP), es el componente de justicia del Sistema Integral de Verdad, Justicia, Reparación y no Repetición, creado por el Acuerdo de Paz entre el Gobierno Nacional y las FARC-EP. La JEP tiene la función de administrar justicia transicional y conocer los delitos cometidos en el marco del conflicto armado antes del 1 de diciembre de 2016. La existencia de la JEP no podrá ser superior a 20 años.

La JEP fue creada para satisfacer los derechos de las víctimas a la justicia, ofrecerles verdad y contribuir a su reparación, con el propósito de construir una paz estable y duradera. El trabajo de la JEP se enfoca en los delitos más graves y representativos del conflicto armado, de acuerdo con los criterios de selección y priorización que se han definidos por la ley y los magistrados. En particular, podrá conocer de los delitos que hubieren cometido excombatientes de las FARC-EP, miembros de la Fuerza Pública, otros agentes del Estado y terceros civiles. Sobre estos dos últimos, la Corte Constitucional aclaró que su participación en la JEP sería voluntaria.

#### 9.2.3. Human Rights Data Analysis Group

Human Rights Data Analysis Group (HRDAG) es una organización sin ánimo de lucro y sin afiliaciones políticas que aplica métodos científicos al análisis de violaciones de derechos humanos alrededor del mundo. Su trabajo inició en 1991 cuando Patrick Ball empezó a desarrollar bases de datos para grupos de derechos humanos en El Salvador. HRDAG creció en la American Association for the Advancement of Science entre 1994 y 2003, y en la Benetech Initiative entre 2003 y 2013. En febrero de 2013 HRDAG se convirtió en una organización independiente en San Francisco, California. Información adicional y de contacto está disponible en su [página web](#), [Twitter](#) y [Facebook](#).

HRDAG está compuesta por estadísticos, científicos de la computación, demógrafos y científicos sociales. HRDAG apoya las protecciones establecidas por la Declaración Universal de los Derechos Humanos, la Convención Internacional de Derechos Civiles y Políticos, y otros tratados e instrumentos internacionales de

derechos humanos. Los científicos de HRDAG proveen resultados no sesgados y científicos a defensores de derechos humanos para aclarar violaciones de derechos humanos.

### 9.3. Datos

Desde el Sistema Integral de Verdad, Justicia, Reparación y no Repetición (SIVJRR) y la organización Human Rights Data Analysis Group (HRDAG) fueron enviados distintos conjuntos de datos que se utilizaron para el procesamiento de información que involucra la primera y segunda fase del proyecto. Las bases de datos provienen de las siguientes entidades: Agencia para la Reincorporación y la Normalización (ARN), Asociación Codebac, Asociación Colombiana de Víctimas de Secuestro y Desaparición Forzada (ACOMIDES), Asociación Nacional de Usuarios Campesinos (ANUC), Asociación de Víctimas Nuevo Amanecer, Cancillería de Colombia, Centro de Cooperación al Indígena (CECOIN), Centro Nacional de Memoria Histórica (CNMH), CIFRAS y CONCEPTOS, Comisión Colombiana de Juristas (CCJ), Colectivo de Abogados “José Alvear Restrepo” CAJAR, Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición (CEV), Comité de Solidaridad con los Presos Políticos (CSPP), Comunidades Negras de la cuenca del río Domingodó, Consejería Presidencial para los Derechos Humanos y Asuntos Internacionales (SIDDDH), Coordinación Colombia Europa Estados Unidos (CCEEU), Consejo Nacional de Lucha contra el Secuestro y demás Atentados contra la Libertad Personal (CONASE), Corporación Caribe afirmativo, Corporación para la Defensa y Promoción de los Derechos Humanos Reiniciar, Corporación Regional para la Defensa de los Derechos Humanos (CREDHOS), Defensoría del pueblo Departamento Administrativo de Seguridad (DAS), Escuela Nacional Sindical (ENS), Federación Colombiana de Trabajadores de la Educación (FECODE), GRUPO DOCENTES UNIVERSIDADES, Fiscalía General de la nación (FGN), Fundación Forjando Futuros, Fundación Lazos de Dignidad (FLD), Fundación País Libre, Instituto Colombiano de Bienestar Familiar (ICBF), Instituto de Estudios para el Desarrollo y la Paz (INDEPAZ), Instituto Nacional de Medicina Legal (INML), Jurisdicción Especial para la Paz (JEP), Justicia Penal Militar (JPM), Ministerio de Defensa, Observatorio Surcolombiano de Derechos Humanos, Paz y Territorio (OBSURDH), OBSERVATORIO DE TIERRAS, Organización Nacional Indígena de Colombia (ONIC), Personería Medellín, Policía Nacional de Colombia (PONAL), Procuraduría General de la Nación (PGN), Sistema de monitoreo de riesgos de la Unidad de Investigación Y Acusación (MEDIOS), Somos Defensores y Unidad Administrativa Especial para la Atención y Reparación Integral a las Víctimas (UARIV).

A continuación se describen las bases de datos recibidas indicando su propósito dentro del proyecto, la fecha de recibo, fecha de actualización, la temporalidad de los hechos, la cantidad de registros recibidos y la cantidad de registros resultantes después del procesamiento.

#### 9.3.1. Agencia para la Reincorporación y la Normalización (ARN)

La Agencia para la Reincorporación y la Normalización (ARN) es la entidad de Presidencia de la República que desde 2003 acompaña y brinda asesoría permanente a quienes le apuestan a la paz y hacen tránsito a la legalidad, generando oportunidades que fortalezcan sus capacidades y puedan desenvolverse plenamente como ciudadanos.

**9.3.1.1. Archivo 1** Contiene información sobre reclutamiento de niños, niñas y adolescentes en Colombia durante 1979 a 2017.

- **Fecha de última actualización:** 5-05-2021
- **Periodo documentado:** 1979–2017
- **Cantidad de registros del archivo:** 3.319
- **Cantidad de registros procesados:** 425
- **Clasificación:** Conflicto armado



### 9.3.2. Asociación Codebac

Entidad sin ánimo de lucro, de carácter permanente, independiente y autónomo y de utilidad común, defensora de derechos humanos, de la población víctima del conflicto armado, de la biodiversidad y de las comunidades más vulnerables. El informe presentado ante la JEP documenta las graves violaciones de derechos humanos y del Derecho Internacional Humanitario cometidos por parte de los diferentes grupos armados que delinquieron en los departamentos de Cundinamarca, Boyacá, Meta, Guaviare, Vichada y Casanare. En especial, el informe se refiere a los hechos de violencia sucedidos en la región de San Luis de Gaceno, en Boyacá, entre ellos, la masacre de el páramo de la Sarna. A su vez, presenta información de contexto frente al origen del paramilitarismo en Casanare y Boyacá, como su consolidación en la región de San Luis de Gaceno.

#### 9.3.2.1. Archivo 1 Parte 1 de la base de datos anexa al informe

- **Fecha de última actualización:** 07-04-2021
- **Periodo documentado:** 1986–2016
- **Cantidad de registros del archivo:** 522
- **Cantidad de registros procesados:** 292
- **Clasificación:** Conflicto armado

#### 9.3.2.2. Archivo 2 Parte 1 de la base de datos anexa al informe

- **Fecha de última actualización:** 07-04-2021
- **Periodo documentado:** 1986–2016
- **Cantidad de registros del archivo:** 518
- **Cantidad de registros procesados:** 311
- **Clasificación:** Conflicto armado

Los dos archivos mencionados fueron procesados y arrojaron una cantidad de 603 registros.

### 9.3.3. Asociación Colombiana de Víctimas de Secuestro y Desaparición Forzada (ACOMIDES)

Es una organización que se caracteriza por defender los derechos de miembros de la Fuerza Pública y sus familias víctimas del conflicto armado.

#### 9.3.3.1. Archivo 1 Contiene registros de miembros del ejército que sufrieron hechos relacionados con desaparición forzada en Colombia entre 1992–2013.

- **Fecha de última actualización:** 26-11-2018
- **Periodo documentado:** 1992–2013
- **Cantidad de registros del archivo:** 112
- **Cantidad de registros procesados:** 112
- **Clasificación:** Conflicto armado

#### 9.3.3.2. Archivo 2 Contiene registros de miembros del ejército que sufrieron hechos relacionados con secuestro en Colombia entre 1990–2013. Esta base de datos incluye variables de fecha del hecho, presunto autor, reporte médico y datos asociados de la salud de la víctima.

- **Fecha de última actualización:** 05-12-2018

- **Periodo documentado:** 1990-2013
- **Cantidad de registros del archivo:** 312
- **Cantidad de registros procesados:** 309
- **Clasificación:** Conflicto armado

**9.3.3.3. Archivo 3** Contiene registros de miembros del Ejército sobre hechos relacionados con homicidio, desaparición forzada, desplazamiento, despojo, secuestro, tortura y amenazas en Colombia entre 1946-2018. Esta base de datos registra casos que hacen parte del anexo al informe sobre militares víctimas en el marco del conflicto y el informe preliminar de ACOMIDES. Incluye variables de tiempo y lugar, así como del grado militar de la víctima y pertenencia a las unidades militares, sexo, pertenencia étnica, tipo de hecho de violencia, entre otros.

- **Fecha de última actualización:** 26-11-2018
- **Periodo documentado:** 1946-2018
- **Cantidad de registros del archivo:** 207645
- **Cantidad de registros procesados:** 183332
- **Clasificación:** Conflicto armado

Los tres archivos mencionados fueron procesados y arrojaron una cantidad de 183.753 registros.

#### **9.3.4. Asociación Nacional de Usuarios Campesinos (ANUC)**

La ANUC es una asociación social gremial, de derecho privado, sin ánimo de lucro con carácter de confederación que agremia los campesinos de Colombia y a las diferentes formas asociativas integradas por ellos.

**9.3.4.1. Archivo 1** La base de datos contiene la documentación de cerca de 400 víctimas, líderes y lideresas integrantes de esa organización y los hechos cometidos por diversos actores, que según el patrón de macro criminalidad argumentado buscaba atacar de forma directa la ANUC.

- **Fecha de última actualización:** 17-12-2020
- **Periodo documentado:** 1986-2016
- **Cantidad de registros del archivo:** 998
- **Cantidad de registros procesados:** 179
- **Clasificación:** Conflicto armado

#### **9.3.5. Asociación de Víctimas Nuevo Amanecer**

Fundación Humanitaria de Desplazados Nuevo Amanecer. Informe titulado “A orillas del río Caquetá, una historia de resistencia por la defensa de la vida” entregado a la JEP.

**9.3.5.1. Archivo 1** La base de datos documenta las víctimas en el conflicto armado en Curillo (Caquetá) y su zona de influencia presuntamente atribuibles al Frente Sur de las FARC-EP entre el periodo de 1985-2016.

- **Fecha de última actualización:** 16-02-2021
- **Periodo documentado:** 1985-2016
- **Cantidad de registros del archivo:** 1457
- **Cantidad de registros procesados:** 293
- **Clasificación:** Conflicto armado

### 9.3.6. Cancillería de Colombia

Es la entidad encargada de dirigir y coordinar la política exterior y las relaciones diplomáticas de Colombia.

**9.3.6.1. Archivo 1** Contiene registros sobre hechos relacionados con tortura, desaparición forzada y homicidio contra miembros de la Unión Patriótica en Colombia entre 1982–2006.

- **Fecha de última actualización:** 02-01-2019
- **Periodo documentado:** 1982–2006
- **Cantidad de registros del archivo:** 4.568
- **Cantidad de registros procesados:**
- **Clasificación:** Conflicto armado

### 9.3.7. Centro de Cooperación al Indígena (CECOIN)

Es una organización no gubernamental que se ha enfocado en la consolidación de los derechos territoriales de los pueblos indígenas, el seguimiento a las políticas públicas relacionadas con este sector de la población, la difusión de sus derechos, el apoyo a la capacitación para la consolidación de sus organizaciones autónomas y la publicación de trabajos sobre diversos problemas que afectan su territorialidad.

**9.3.7.1. Archivo 1** Contiene datos de hechos de violencia en contra de comunidades indígenas en el marco del conflicto armado en Colombia.

- **Fecha de última actualización:** 14-08-2018
- **Periodo documentado:** 1899-2014
- **Cantidad de registros del archivo:** 41.091
- **Cantidad de registros procesados:** 3.462
- **Clasificación:** Conflicto armado

### 9.3.8. Centro Nacional de Memoria Histórica (CNMH)

Es una organización gubernamental cuyo fin es contribuir a la reparación integral y al derecho a la verdad, a través de la recuperación, conservación y divulgación de las memorias plurales de las víctimas, así como del deber de memoria del Estado y de todos los victimarios con ocasión de las violaciones ocurridas en el marco del conflicto armado.

**9.3.8.1. Archivo 1** Esta base de datos se encuentra información sobre víctimas y las organizaciones de víctima y también de las instituciones públicas y privadas a las que el Observatorio ha acudido para documentar los casos de violencia. \* **Fecha de última actualización:** 06-11-2019 \* **Periodo documentado:** 1933–2018 \* **Cantidad de registros del archivo:** 432.493 \* **Cantidad de registros procesados:** 315.361 \* **Clasificación:** Conflicto armado

**9.3.8.2. Archivo 2** Filtro de la base de dato del CNMH a partir del campo población vulnerable que cumpliera con criterios de miembro de comunidad religiosa, líder campesino, periodista, autoridad indígena, líder comunitario o militante político.

- **Fecha de última actualización:** 24-05-2021
- **Periodo documentado:** 1982–2012
- **Cantidad de registros del archivo:** 275
- **Cantidad de registros procesados:** 0
- **Clasificación:** Conflicto armado

### 9.3.9. Cifras y Conceptos

Cifras y Conceptos es una empresa privada dedicada a la consultoría estadística y en análisis de datos.

#### 9.3.9.1. Archivo 1 Base de datos víctimas de desaparición forzada.

- **Fecha de última actualización:** 18-05-2021
- **Periodo documentado:** 1955–2013
- **Cantidad de registros del archivo:** 29.013
- **Cantidad de registros procesados:** 28.083
- **Clasificación:** Conflicto armado

### 9.3.10. Colectivo de Abogados “José Alvear Restrepo” CAJAR

Es una Organización No Gubernamental sin ánimo de lucro, defensora de Derechos Humanos en Colombia. En el año de 1980 adquiere su personería jurídica convirtiéndose en una de las primeras organizaciones defensoras de Derechos Humanos en Colombia compuesta por abogados y abogadas. Cuenta con estatus consultivo ante la Organización de Estados Americanos, OEA, ante la ONU.

**9.3.10.1. Archivo 1** El informe describe actuaciones bajo la presunta responsabilidad del del Batallón de Artillería No. 2 “La Popa” para el periodo diciembre de 2007 a junio de 2009. Documenta 23 casos en los que fueron asesinadas 39 víctimas.

- **Fecha de última actualización:** 03-07-2019
- **Periodo documentado:** 2007–2009
- **Cantidad de registros del archivo:** 111
- **Cantidad de registros procesados:** 90
- **Clasificación:** Conflicto armado

### 9.3.11. Comisión Colombiana de Juristas (CCJ)

Es una organización no gubernamental que busca contribuir al desarrollo del derecho internacional de los derechos humanos y del derecho internacional humanitario, de conformidad con los propósitos y principios de la Carta de las Naciones Unidas y, por esa vía, a la plena vigencia del Estado social y democrático de derecho en Colombia.

**9.3.11.1. Archivo 1** Contiene datos de víctimas de homicidio selectivo y desaparición forzada en Colombia entre 1986–2010.

- **Fecha de última actualización:** 23-01-2017
- **Periodo documentado:** 1986–2010
- **Cantidad de registros del archivo:** 9.164
- **Cantidad de registros procesados:** 2.952
- **Clasificación:** Violencia sociopolítica

**9.3.11.2. Archivo 2** Contiene datos de sindicalistas víctimas de homicidios selectivos y desaparición forzada en Colombia entre 1986–2011.

- **Fecha de última actualización:** 27-07-2020
- **Periodo documentado:** 1986–2011
- **Cantidad de registros del archivo:** 3.079
- **Cantidad de registros procesados:** 2.951
- **Clasificación:** Violencia sociopolítica

**9.3.11.3. Archivo 3** Base de datos entregada por la CCJ a la JEP que contiene los casos de representación de esta ONG ante diferentes instancias judiciales.

- **Fecha de última actualización:** 03-10-2017
- **Periodo documentado:** 1987–2008
- **Cantidad de registros del archivo:** 55
- **Cantidad de registros procesados:** 41
- **Clasificación:** Conflicto armado

**9.3.11.4. Archivo 4** Contiene datos de víctimas por amenazas, homicidio, desplazamiento forzado, secuestro, tortura, desaparición forzada y otros delitos en Colombia entre 1996–2016.

- **Fecha de última actualización:** 17-10-2020
- **Periodo documentado:** 1996–2016
- **Cantidad de registros del archivo:** 154.040
- **Cantidad de registros procesados:** 54.415
- **Clasificación:** Violencia sociopolítica

**9.3.11.5. Archivo 5** Contiene datos de víctimas de desplazamiento forzado que también fueron afectadas por homicidios selectivos y desaparición forzada en Colombia entre 2007–2011.

- **Fecha de última actualización:** 23-01-2017
- **Periodo documentado:** 2007–2011
- **Cantidad de registros del archivo:** 289
- **Cantidad de registros procesados:** 248
- **Clasificación:** Violencia sociopolítica

**9.3.11.6. Archivo 6** Contiene datos de funcionarios del poder judicial víctimas de homicidios selectivos, ejecuciones extrajudiciales y desaparición forzada en Colombia entre 2002–2008.

- **Fecha de última actualización:** 16-10-2020
- **Periodo documentado:** 2002–2008
- **Cantidad de registros del archivo:** 181
- **Cantidad de registros procesados:** 179
- **Clasificación:** Violencia sociopolítica

**9.3.11.7. Archivo 7** Contiene datos de comunidades indígenas y afrodescendientes víctimas de homicidio y desaparición forzada en Colombia entre 2004–2008.

- **Fecha de última actualización:** 11-01-2017

- **Periodo documentado:** 2004–2008
- **Cantidad de registros del archivo:** 1.291
- **Cantidad de registros procesados:** 494
- **Clasificación:** Violencia sociopolítica

**9.3.11.8. Archivo 8** Contiene datos de mujeres víctimas de homicidio selectivo, desaparición forzada y secuestro en Colombia entre 2002–2008.

- **Fecha de última actualización:** 11-01-2017
- **Periodo documentado:** 2002–2008
- **Cantidad de registros del archivo:** 1.736
- **Cantidad de registros procesados:** 1.248
- **Clasificación:** Violencia sociopolítica

**9.3.11.9. Archivo 9** Contiene datos de niños, niñas y adolescentes víctimas de homicidio selectivo, desaparición forzada y secuestro en Colombia entre 2007–2008.

- **Fecha de última actualización:** 22-09-2008
- **Periodo documentado:** 2007-2008
- **Cantidad de registros del archivo:** 251
- **Cantidad de registros procesados:** 82
- **Clasificación:** Violencia sociopolítica

**9.3.11.10. Archivo 10** Contiene datos de víctimas de homicidio selectivo y desaparición forzada perpetradas por paramilitares en Colombia entre 2002–2008.

- **Fecha de última actualización:** 11-01-2017
- **Periodo documentado:** 2002–2008
- **Cantidad de registros del archivo:** 17.867
- **Cantidad de registros procesados:** 13.177
- **Clasificación:** Conflicto armado

**9.3.11.11. Archivo 11** Contiene datos de funcionarios del poder judicial víctimas de amenazas, secuestro y tortura en Colombia entre 1996–2008.

- **Fecha de última actualización:** 17-10-2020
- **Periodo documentado:** 1996–2008
- **Cantidad de registros del archivo:** 542
- **Cantidad de registros procesados:** 137
- **Clasificación:** Violencia sociopolítica

**9.3.11.12. Archivo 12** Contiene datos de defensoras y defensores de derechos humanos víctimas de homicidio selectivo, ejecuciones y desaparición forzada en Colombia entre 1996–2012.

- **Fecha de última actualización:** 17-10-2020
- **Periodo documentado:** 1996–2012
- **Cantidad de registros del archivo:** 290
- **Cantidad de registros procesados:** 279
- **Clasificación:** Violencia sociopolítica

Los doce (12) archivos mencionados fueron procesados y arrojaron una cantidad de 76.203 registros.

### **9.3.12. Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No repetición (CEV)**

Es una Entidad de Estado que busca el esclarecimiento de los patrones y causas explicativas del conflicto armado interno que satisfaga el derecho de las víctimas y de la sociedad a la verdad, promueva el reconocimiento de lo sucedido, la convivencia en los territorios y contribuya a sentar las bases para la no repetición, mediante un proceso de participación amplio y plural para la construcción de una paz estable y duradera.

**9.3.12.1. Archivo 1** La CEV entregó tres archivos que por su contenido se consolidaron en uno solo, dado que el primero contenía la caracterización de la víctima, el segundo contenía la caracterización del hecho y el tercer archivo se refería a la relación entre la víctima y los hechos. Por lo anterior el archivo consolidado englobó datos históricos víctimas de 18 tipos de violencia junto con información de caracterización de las víctimas, de los hechos perpetrados y sus respectivos responsables en el marco del conflicto armado entre 1930 y 2020.

- **Fecha de última actualización:** 31-08-2020
- **Periodo documentado:** 1930–2020
- **Cantidad de registros del archivo:** 29.147
- **Cantidad de registros procesados:** 18.909
- **Clasificación:** Conflicto armado

### **9.3.13. Fundación Comité de Solidaridad con los Presos Políticos (CSPP)**

La Fundación CSPP, se propone contribuir a la exigencia, promoción y difusión del respeto y las garantías de los Derechos Humanos de todas las personas en Colombia, en la búsqueda de la verdad, la justicia y la reparación; en especial el derecho a la Vida, la Libertad, la Integridad Física y Moral, a un trato digno, a un juicio justo e imparcial y los demás derechos de las personas privadas de la libertad, procesadas por delitos políticos y judicializadas por participar en la protesta social.

**9.3.13.1. Archivo 1** Contiene registros sobre homicidio y desaparición forzada en Colombia entre 1990–2016.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 1990–2016
- **Cantidad de registros del archivo:** 1.124
- **Cantidad de registros procesados:**
- **Clasificación:** Conflicto armado

### **9.3.14. Comunidades Negras de la cuenca del río Domingodó**

Es el conjunto de familias de ascendencia afrocolombiana pertenecientes a los consejos comunitarios ubicados en la cuenca del río Domingodó, en Carmen del Darién, Chocó.

**9.3.14.1. Archivo 1** Contiene datos de víctimas del Consejo Comunitario de la Cuenca del Río Timba y Mary López afectadas por homicidios selectivos, desaparición forzada y desplazamiento forzado entre 1998–2007.

- **Fecha de última actualización:** 20-05-2020
- **Periodo documentado:** 1998–2007
- **Cantidad de registros del archivo:** 32
- **Cantidad de registros procesados:** 19
- **Clasificación:** Conflicto armado

**9.3.14.2. Archivo 2** Contiene datos de víctimas de las comunidades negras de la cuenca del río Dominodó afectadas por desplazamiento forzado entre 1993–2013.

- **Fecha de última actualización:** 20-05-2020
- **Periodo documentado:** 1993–2013
- **Cantidad de registros del archivo:** 185
- **Cantidad de registros procesados:** 111
- **Clasificación:** Conflicto armado

**9.3.14.3. Archivo 3** Contiene datos de víctimas del Consejo Comunitario de la Cuenca del Río Timba y Mary López afectadas por homicidios selectivos, amenaza, desaparición y desplazamiento forzados entre 1997–2008.

- **Fecha de última actualización:** 20-05-2020
- **Periodo documentado:** 1997–2008
- **Cantidad de registros del archivo:** 2.123
- **Cantidad de registros procesados:** 819
- **Clasificación:** Conflicto armado

**9.3.14.4. Archivo 4** Contiene datos de víctimas del Consejo Comunitario de la Cuenca del Río Salaqui afectadas por homicidios selectivos, tortura, secuestro, amenaza, desaparición y desplazamiento forzados entre 1947–2012.

- **Fecha de última actualización:** 20-05-2020
- **Periodo documentado:** 1947–2012
- **Cantidad de registros del archivo:** 418
- **Cantidad de registros procesados:** 105
- **Clasificación:** Conflicto armado

### **9.3.15. Consejería Presidencial para los Derechos Humanos**

La Consejería Presidencial para los Derechos Humanos y Asuntos Internacionales es la entidad presidencial que coordina e impulsa las acciones dirigidas a la promoción, protección y defensa de los derechos humanos, con las diferentes entidades gubernamentales que tienen competencia en la materia. El objetivo principal de esta Consejería Presidencial es el de mejorar los niveles de respeto, protección y garantía de los derechos humanos, así como la aplicación del DIH. Además, contribuir a que el Estado en su conjunto desarrolle una política integral de promoción y respeto a los derechos humanos.

**9.3.15.1. Archivo 1** Base de datos del sistema Información en Derechos Humanos IDH entre 1982 y 2014.

- **Fecha de última actualización:** 18-05-2021
- **Periodo documentado:** 1982–2014
- **Cantidad de registros del archivo:** 624.816
- **Cantidad de registros procesados:** 158.836
- **Clasificación:** Conflicto armado

### **9.3.16. Consejo Nacional de Lucha contra el Secuestro y demás Atentados contra la Libertad Personal (CONASE)**

Es un órgano asesor, consultivo y de coordinación en la lucha contra los delitos contra la libertad individual, en especial el secuestro y la extorsión creado en mediante la ley 282 de 1996.



**9.3.16.1. Archivo 1** Contiene datos históricos de secuestro en Colombia entre 1996–2015.

- **Fecha de última actualización:** 27-08-2020
- **Periodo documentado:** 1996-2015
- **Cantidad de registros del archivo:** 7.257
- **Cantidad de registros procesados:** 7.212
- **Clasificación:** Conflicto armado

### **9.3.17. Coordinación Colombia Europa Estados Unidos (CCEEU)**

Es la plataforma de organizaciones de derechos humanos y sociales, en cuyo mandato se reconoce como un actor colectivo de incidencia internacional y nacional, que mediante una agenda política de trabajo busca contribuir a la promoción, divulgación y defensa de los derechos humanos con énfasis en derechos civiles y políticos, la lucha contra la impunidad, la vigencia del Estado social de Derecho, la construcción de paz y la búsqueda de una solución política negociada al conflicto armado.

**9.3.17.1. Archivo 1** El propósito de este conjunto de datos entregado a la CEV y a la JEP por la CCEEU es documentar los casos de víctimas de ejecuciones extrajudiciales entre 1990-2015 por parte de Agentes Estatales. Este conjunto de datos es generado por la mesa de trabajo sobre ejecuciones extrajudiciales. Relaciona casos de víctimas de ejecuciones extrajudiciales del país, con las variables de fecha, municipio, departamento, nombres y apellidos de las víctimas, edad, género, si es menor de edad, condición o actividad de la víctima, documento de identidad, tipo de ejecución, contexto, otros hechos de violencia, unidad, brigada, división e institución a la cual pertenecía el responsable de la ejecución, jurisdicción de la unidad militar identificada, seguimiento judicial, registro, si tiene ficha en el sistema de la coordinación, fuentes de información, estado del proceso en la Procuraduría y en la Fiscalía, el grado militar al momento de los hechos, delito por el cual fue condenado el responsable y estado del proceso en la Justicia Penal Militar.

- **Fecha de última actualización:** 15-01-2021
- **Periodo documentado:** 1990–2015
- **Cantidad de registros del archivo:** 6.885
- **Cantidad de registros procesados:** 5.517
- **Clasificación:** Conflicto armado

**9.3.17.2. Archivo 2** Este conjunto de datos entregado por la JEP contiene datos de víctimas de homicidio selectivo, desaparición forzada, secuestro y tortura en Colombia entre 2002-2013. La mayoría de los registros de este archivo estaban incluidos en el Archivo 1.

- **Fecha de última actualización:** 14-08-2015
- **Periodo documentado:** 2002–2013
- **Cantidad de registros del archivo:** 5.515
- **Cantidad de registros procesados:** 7
- **Clasificación:** Conflicto armado

**9.3.17.3. Archivo 3** Este conjunto de datos entregado por la JEP contiene datos de víctimas de homicidios selectivos en Colombia entre 2010-2015. La mayoría de los registros de este archivo estaban incluidos en el Archivo 1.

- **Fecha de última actualización:** 14-08-2015
- **Periodo documentado:** 2010–2015
- **Cantidad de registros del archivo:** 279

- **Cantidad de registros procesados:** 2
- **Clasificación:** Conflicto armado

Los tres (3) archivos mencionados fueron procesados y arrojaron una cantidad de 5.526 registros.

### **9.3.18. Corporación Caribe afirmativo**

Es una organización no gubernamental cuyo propósito es incidir en la construcción y posicionamiento de agendas de diversidad sexual y de género en el Caribe, que permitan el reconocimiento de los derechos y el acceso a la justicia de las personas LGBTI en un escenario de posconflicto.

**9.3.18.1. Archivo 1** Contiene datos de víctimas de distintos repertorios de violencia relacionados con la persecución por su orientación sexual, identidad y expresión de género diverso que sufrieron las personas LGBT en el marco conflicto armado en la región Caribe entre el periodo 1988–2016.

- **Fecha de última actualización:** 26-09-2020
- **Periodo documentado:** 1988–2016
- **Cantidad de registros del archivo:** 69
- **Cantidad de registros procesados:** 3
- **Clasificación:** Conflicto armado

**9.3.18.2. Archivo 2** Contiene datos sobre delitos contra la libertad e integridad sexual de mujeres en el marco conflicto armado en la región Caribe entre el periodo 1986–2016.

- **Fecha de última actualización:** 26-09-2020
- **Periodo documentado:** 1986–2016
- **Cantidad de registros del archivo:** 7
- **Cantidad de registros procesados:** 8
- **Clasificación:** Conflicto armado

Los dos (2) archivos mencionados fueron procesados y arrojaron una cantidad de 11 registros.

### **9.3.19. Corporación para la Defensa y Promoción de los Derechos Humanos Reiniciar**

Es una organización no gubernamental, sin ánimo de lucro, con estatus consultivo ante la Organización de Estados Americanos y el Consejo Económico y Social de la Organización de las Naciones Unidas.

Fue fundada en Bogotá en 1992 por un grupo de personas defensoras de derechos humanos de la región del Magdalena Medio colombiano, quienes se vieron obligadas a desplazarse por hostigamientos y amenazas contra sus vidas. Representa la decisión de dar continuidad a la defensa y promoción de los derechos humanos y el Derechos Internacional Humanitario en cualquier lugar o circunstancia.

**9.3.19.1. Archivo 1** Contiene registros de desaparición forzada, homicidio, tortura, desplazamiento forzado y amenazas selectivas en contra de personas pertenecientes al movimiento político Unión Patriótica (UP) entre 1984–2006.

- **Fecha de última actualización:** 23-09-2020
- **Periodo documentado:** 1984–2006
- **Cantidad de registros del archivo:** 6.487
- **Cantidad de registros procesados:** 5.196
- **Clasificación:** Conflicto armado

### 9.3.20. Corporación Regional para la Defensa de los Derechos Humanos (CREDHOS)

Es una organización no gubernamental comprometida con la defensa, promoción y protección de los Derechos Humanos y la divulgación del Derecho Internacional Humanitario. Dentro de las políticas de CREDHOS, está la búsqueda de caminos, que lleven a la reconciliación nacional y a la paz con justicia social en la Región del Magdalena Medio.

**9.3.20.1. Archivo 1** Contiene datos de homicidios selectivos, desaparición forzada, desplazamiento forzado, amenaza y tortura en el Magdalena Medio entre 1991–2015.

- **Fecha de última actualización:** 09-12-2019
- **Periodo documentado:** 1991–2015
- **Cantidad de registros del archivo:** 424
- **Cantidad de registros procesados:** 225
- **Clasificación:** Conflicto armado

### 9.3.21. Departamento Administrativo de Seguridad (DAS)

Departamento Administrativo de Seguridad (DAS), suprimido en el año 2011, fue un organismo estatal encargado de realizar la inteligencia y contrainteligencia en Colombia. Los datos obtenidos provienen de *boletines* diarios del DAS, que circularon dentro de las oficinas del gobierno durante los años 2003 y 2009.

Estos archivos fueron recopilados por el Observatorio de Derechos Humanos de la Vicepresidencia y compartidos con la Oficina del Alto Comisionado para los Derechos Humanos de la ONU como parte de un proyecto de HRDAG realizado entre el año 2013 y 2014. Esta base de datos se usó en los análisis publicados en [Ball and Reed Hurtado \(2015\)](#) y [Ball and Reed Hurtado \(2016\)](#). Cada boletín constaba de narraciones que describían la confrontación directa entre agentes del estado y varios actores como la delincuencia común, las guerrillas, bandas criminales organizadas y grupos paramilitares. Cuando hubo arrestos de víctimas, las personas a menudo fueron identificadas por su nombre y cédula. En 2014 y en la segunda mitad del 2017, los analistas de HRDAG leyeron aproximadamente 70.000 boletines, que no necesariamente se refieren a personas específicas. Los cuatro archivos se refieren al mismo set de datos procesados.

**9.3.21.1. Archivo 1** VP-DAS-04-09

- **Fecha de última actualización:** 2013
- **Periodo documentado:** 2004–2009
- **Cantidad de registros del archivo:** 11.660
- **Cantidad de registros procesados:** 13.356
- **Clasificación:** Violencia sociopolítica

**9.3.21.2. Archivo 2** Combates-DAS

- **Fecha de última actualización:** 2013
- **Periodo documentado:** 2003–2007
- **Cantidad de registros del archivo:** 7.649
- **Cantidad de registros procesados:** 13.356
- **Clasificación:** Conflicto armado

#### **9.3.21.3. Archivo 3** Violaciones-DAS

- **Fecha de última actualización:** 2013
- **Periodo documentado:** 2002–2007
- **Cantidad de registros del archivo:** 46.742
- **Cantidad de registros procesados:** 13.356
- **Clasificación:** Violencia sociopolítica

#### **9.3.21.4. Archivo 4** VS-DAS-handprocessed

- **Fecha de última actualización:** 2013
- **Periodo documentado:** 2002–2007
- **Cantidad de registros del archivo:** 11.593
- **Cantidad de registros procesados:** 13.356
- **Clasificación:** Violencia sociopolítica

Los cuatro (4) archivos mencionados fueron procesados y arrojaron una cantidad de 13.356 registros.

### **9.3.22. Defensoría del Pueblo de Colombia**

La Defensoría del Pueblo de Colombia es un órgano constitucional y autónomo creado por la Constitución Política de Colombia de 1991 en su Capítulo II: Del ministerio público, con el propósito de “velar por la promoción, el ejercicio y la divulgación de los derechos humanos”

#### **9.3.22.1. Archivo 1** Base de datos de violaciones de los DDHH contra activistas de DDHH hasta 2009

- **Fecha de última actualización:** 20-05-2021
- **Periodo documentado:** 1985–2009
- **Cantidad de registros del archivo:** 443
- **Cantidad de registros procesados:** 54
- **Clasificación:** Conflicto armado

#### **9.3.22.2. Archivo 2** Base de datos de violaciones de los DDHH contra líderes sociales desde 2000 hasta 2021

- **Fecha de última actualización:** 20-05-2021
- **Periodo documentado:** 2000–2021
- **Cantidad de registros del archivo:** 2.849
- **Cantidad de registros procesados:** 54
- **Clasificación:** Conflicto armado

Los dos (2) archivos mencionados fueron procesados y arrojaron una cantidad de 3.292 registros.

### **9.3.23. Escuela Nacional Sindical (ENS)**

Organización más importante de Colombia dedicada a la defensa los derechos de los trabajadores y trabajadoras del país para que sean protagonistas de la sociedad.

**9.3.23.1. Archivo 1** Base de datos que contiene información sobre violaciones de DDHH contra sindicalistas en Colombia.

- **Fecha de última actualización:** 27-07-2020
- **Periodo documentado:** 1971–2020
- **Cantidad de registros del archivo:** 15.160
- **Cantidad de registros procesados:** 4.673
- **Clasificación:** Conflicto armado

#### **9.3.24. Federación Colombiana de Trabajadores de la Educación (FECODE)**

Es una organización de carácter sindical que está comprometida con la lucha por la satisfacción plena de las necesidades y expectativas educativas del pueblo, mediante el logro del derecho a la educación integral de calidad, con cobertura total y financiada por el Estado.

**9.3.24.1. Archivo 1** El conjunto de datos entregado por FECODE a la CEV y a la JEP tiene como propósito la documentación de la violencia ejercida contra los trabajadores de la educación en el marco del conflicto armado con especial énfasis en las violaciones de sindicalistas miembros de FECODE víctimas de homicidio y desaparición forzada entre 1982–2016.

- **Fecha de última actualización:** 15-10-2020
- **Periodo documentado:** 1982–2016
- **Cantidad de registros del archivo:** 1.104
- **Cantidad de registros procesados:** 1.095
- **Clasificación:** Conflicto armado

#### **9.3.25. Fiscalía General de la nación (FGN)**

Es un organismo del Estado independiente adscrito al poder judicial en Colombia. Por tal motivo, ejerce la acción penal y de extinción de dominio en el marco del derecho constitucional al debido proceso; participa en el diseño y la ejecución de la política criminal del Estado; garantiza el acceso efectivo a la justicia, la verdad y la reparación de las víctimas de los delitos; y genera confianza en la ciudadanía.

**9.3.25.1. Archivo 1** Base de datos aportada por la JEP por medio del auto 100 de la SRVR de 26 de junio de 2019. Contiene información de procesos penales posiblemente relacionados con el conflicto armado interno por conductas criminales cometidas entre 2000–2016 antes del 1 de diciembre de 2016 por miembros de la Fuerza Pública y las FARC-EP, así como por grupos sin identificar, en el marco de la Ley 906 de 2004. Se diferencia de las otras bases de datos en que los procesos a los que se refiere incluyen personas que no estaban en el inventario del conflicto armado entregado por la FGN a la JEP pero cuyos presuntos responsables correspondían a personas sometidas a la JEP.

- **Fecha de entrega:** 18-02-2020
- **Periodo documentado:** 2000–2016
- **Cantidad de registros del archivo:** 99.578
- **Cantidad de registros procesados:** 2.667
- **Clasificación:** Conflicto armado

**9.3.25.2. Archivo 2** Base de datos que contiene registros sobre desaparición forzada, desplazamiento forzado, homicidio y reclutamiento en la clasificación conflicto armado entre 1978–2019.

- **Fecha de última actualización:** 23-08-2020
- **Periodo documentado:** 1978–2019
- **Cantidad de registros del archivo:** 2.017.220
- **Cantidad de registros procesados:** 918.237
- **Clasificación:** Violencia sociopolítica

**9.3.25.3. Archivo 3** Base de datos aportada por la JEP mediante auto 100 de la Sala de Reconocimiento de Verdad, de Responsabilidad y de Determinación de los Hechos y Conductas (SRVR) del 26 de junio de 2019 y contiene registros de víctimas de delitos contra la integridad sexual, terrorismo, amenazas, homicidio selectivo, secuestro, desaparición forzada, entre otros, en Colombia entre 1980–2017, por miembros de la Fuerza Pública y las FARC-EP, así como por grupos sin identificar, cuyos procesos se encuentran en el marco de la Ley 906 de 2004 y de la Ley 600 de 2000. Hace parte de los anexos al informe No. 1 entregado por la FGN a la JEP sobre el inventario del conflicto armado en Colombia.

- **Fecha de última actualización:** 30-06-2020
- **Periodo documentado:** 1980–2017
- **Cantidad de registros del archivo:** 733.368
- **Cantidad de registros procesados:** 113.457
- **Clasificación:** Conflicto armado

**9.3.25.4. Archivo 4** Registros de la FGN sobre personas presuntamente reclutadas por el bloque oriental de las FARC-EP.

- **Fecha de última actualización:** 4-06-2019
- **Periodo documentado:** 1975–2007
- **Cantidad de registros del archivo:** 4.495
- **Cantidad de registros procesados:** 3.194
- **Clasificación:** Conflicto armado

**9.3.25.5. Archivo 5** Registros del SPOA y el SIJUF de la FGN cuya víctima era líder social o defensor activista de DDHD.

- **Fecha de última actualización:** 24-06-2021
- **Periodo documentado:** 1986–2020
- **Cantidad de registros del archivo:** 6.100
- **Cantidad de registros procesados:** 665
- **Clasificación:** Conflicto armado

**9.3.25.6. Archivo 6** Registros de procesos penales en ley 906 y ley 600 que la FGN consideró relacionados al conflicto armado. Contiene información de víctimas, victimarios y hechos delictivos, así como variables de tiempo, modo y lugar, entre 1978–2019.

- **Fecha de última actualización:** 30-06-2020
- **Periodo documentado:** 1978–2019
- **Cantidad de registros del archivo:** 2.098.671
- **Cantidad de registros procesados:** 289.717
- **Clasificación:** Conflicto armado

**9.3.25.7. Archivo 7** Registros de la dirección de Justicia Transicional de la FGN de delitos cometidos contra miembros de la Unión Patriótica.

- **Fecha de última actualización:** 25-10-2021
- **Periodo documentado:** 1984–2005
- **Cantidad de registros del archivo:** 1.113
- **Cantidad de registros procesados:** 875
- **Clasificación:** Conflicto armado

**9.3.25.8. Archivo 8** Base de datos que contiene registros sobre reclutamiento en Colombia entre 1971–2016 y corresponde al anexo del informe No. 4 de la FGN entregado a la JEP denominado Vinculación y Utilización de niños, niñas, adolescentes por parte de las FARC-EP.

- **Fecha de última actualización:** 25-07-2017
- **Periodo documentado:** 1971–2016
- **Cantidad de registros del archivo:** 8.293
- **Cantidad de registros procesados:** 6.017
- **Clasificación:** Conflicto armado

**9.3.25.9. Archivo 9** SIJYP\_FORMATOS\_VICTIMAS - Contiene datos de hechos de violencia entre 1983–2017 del territorio colombiano.

- **Fecha de última actualización:** 15-10-2020
- **Periodo documentado:** 1983–2017
- **Cantidad de registros del archivo:** 204.993
- **Cantidad de registros procesados:** 149.218
- **Clasificación:** Conflicto armado

**9.3.25.10. Archivo 10** Contiene datos de hechos de violencia entre 1988–2011 del territorio colombiano, correspondiente a hechos confesados por miembros de las FARC desmovilizados en el marco de la ley 975 de 2005 o ley de Justicia y Paz.

- **Fecha de última actualización:** 30-05-2018
- **Periodo documentado:** 1988–2011
- **Cantidad de registros del archivo:** 38.989
- **Cantidad de registros procesados:** 19.852
- **Clasificación:** Conflicto armado

**9.3.25.11. Archivo 11** SIJYP - Contiene información sobre desaparición forzada, homicidio y reclutamiento forzado en Colombia entre 1930–2020.

- **Fecha de última actualización:** 06-10-2020
- **Periodo documentado:** 1930–2020
- **Cantidad de registros del archivo:** 449.445
- **Cantidad de registros procesados:** 433.813
- **Clasificación:** Conflicto armado

**9.3.25.12. Archivo 12** Contiene datos sobre desaparición forzada, homicidio, reclutamiento forzado, tortura, desplazamiento forzado, amenazas y delitos contra la integridad sexual entre 1988–2013 en Colombia.

- **Fecha de última actualización:** 15-10-2020
- **Periodo documentado:** 1988–2013
- **Cantidad de registros del archivo:** 1481
- **Cantidad de registros procesados:** 1200
- **Clasificación:** Conflicto armado

**9.3.25.13. Archivo 13** Registros de procesos penales de la FGN asociados a víctimas militantes del partido político de la Unión Patriótica

- **Fecha de última actualización:** 22-10-2021
- **Periodo documentado:** 1984–2016
- **Cantidad de registros del archivo:** 1.721
- **Cantidad de registros procesados:** 1.472
- **Clasificación:** Conflicto armado

Los 13 archivos mencionados fueron procesados y arrojaron una cantidad de 1.506.571 registros.

### **9.3.26. Fundación Forjando Futuros**

Es una organización no gubernamental que asesora y representa judicialmente a las víctimas en los procesos de reclamación para la restitución de las tierras despojadas y/o abandonadas con motivo del conflicto armado y en el sistema integral de verdad, justicia y reparación y garantías de no repetición.

**9.3.26.1. Archivo 1** Contiene registros de desplazamiento forzado en Colombia entre 1982–2018.

- **Fecha de última actualización:** 02-09-2020
- **Periodo documentado:** 1982–2018
- **Cantidad de registros del archivo:** 9.690
- **Cantidad de registros procesados:** 8.825
- **Clasificación:** Conflicto armado

### **9.3.27. Fundación Lazos de Dignidad (FLD)**

Es una organización no gubernamental, sin ánimo de lucro, dedicada a la Defensa, Promoción y Educación en Derechos Humanos; conformada por personas con distintas profesiones y oficios, unidas por el sentido humanista, el principio de solidaridad, el enfoque crítico de los derechos humanos, el compromiso de aportar a la construcción de un mundo más justo y una sociedad en dignidad.

**9.3.27.1. Archivo 1** Contiene registros sobre homicidio contra líderes en Colombia entre 2016–2020.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 2016–2020
- **Cantidad de registros del archivo:** 246
- **Cantidad de registros procesados:** 147
- **Clasificación:** Violencia sociopolítica



### 9.3.28. Fundación País Libre

Hasta 2017 fue una organización no gubernamental sin ánimo de lucro que trabajaron en la prevención contra el secuestro, la extorsión, la desaparición forzada y otras privaciones ilegales de la libertad en Colombia.

**9.3.28.1. Archivo 1** Contiene registros sobre secuestro y extorsión entre 1996–2006.

- **Fecha de última actualización:** 02-11-2018
- **Periodo documentado:** 1996–2006
- **Cantidad de registros del archivo:** 22.518
- **Cantidad de registros procesados:** 22.343
- **Clasificación:** Conflicto armado

**9.3.28.2. Archivo 2** Contiene registros sobre secuestro entre 1982–2013 atribuidos a la guerrilla de las FARC-EP.

- **Fecha de última actualización:** 11-07-2018
- **Periodo documentado:** 1982–2013
- **Cantidad de registros del archivo:** 284
- **Cantidad de registros procesados:** 272
- **Clasificación:** Conflicto armado

**9.3.28.3. Archivo 3** Contiene registros sobre desaparición forzada, secuestro y desplazamiento forzado atribuidos a la guerrilla de las FARC-EP entre 1989–2007.

- **Fecha de última actualización:** 11-07-2018
- **Periodo documentado:** 1989–2007
- **Cantidad de registros del archivo:** 202
- **Cantidad de registros procesados:** 209<sup>40</sup>
- **Clasificación:** Conflicto armado

**9.3.28.4. Archivo 4** Contiene registros sobre secuestro y desaparición forzada en Colombia entre 1985–2015.

- **Fecha de última actualización:** 11-07-2018
- **Periodo documentado:** 1985–2015
- **Cantidad de registros del archivo:** 554
- **Cantidad de registros procesados:** 176
- **Clasificación:** Conflicto armado

Los cuatro (4) archivos mencionados fueron procesados y arrojaron una cantidad de 23.000 registros.

### 9.3.29. Grupo Docentes Universidades: Víctimas de Misión Médica de Norte de Santander

Las razones de creación del grupo, obedece a la necesidad que tienen los integrantes de Misión Médica, de que se reconozca de manera pública como víctimas de las graves infracciones al DIH cometidas por los actores del conflicto, la satisfacción de los derechos de las víctimas, en materia de verdad, justicia y medidas retributivas que el grupo quiere proponer a la JEP y ante todo se busca otorgarle dignidad a la labor del personal de salud en los lugares apartados del país. En el informe participan 23 personas de salud.

---

<sup>40</sup>Algunos de los registros originales contienen información sobre más de una víctima en un sólo registro. Hemos separado estos registros para tener un registro para cada víctima y por eso la cantidad de registros procesados es mayor que la cantidad original.

**9.3.29.1. Archivo 1** Base de datos de hechos de violencia cometidos contra personal médico en algunos municipios de Norte de Santander y Cesar (parte 1)

- **Fecha de última actualización:** 28-10-2020
- **Periodo documentado:** 1984–2019
- **Cantidad de registros del archivo:** 76
- **Cantidad de registros procesados:** 34
- **Clasificación:** Conflicto armado

**9.3.29.2. Archivo 2** Base de datos de hechos de violencia cometidos contra personal médico en algunos municipios de Norte de Santander y Cesar (parte 2)

- **Fecha de última actualización:** 20-10-2020
- **Periodo documentado:** 1984–2015
- **Cantidad de registros del archivo:** 55
- **Cantidad de registros procesados:** 34
- **Clasificación:** Conflicto armado

Los dos (2) archivos mencionados fueron procesados y arrojaron una cantidad total de registros de 34.

### **9.3.30. Instituto Colombiano de Bienestar Familiar (ICBF)**

Es la entidad del Estado colombiano que trabaja por la prevención y protección integral de la primera infancia, la niñez, la adolescencia y el bienestar de las familias en Colombia.

**9.3.30.1. Archivo 1** Contiene registros sobre reclutamiento en Colombia entre 1992–2015.

- **Fecha de última actualización:** 31-01-2019
- **Periodo documentado:** 1992–2015
- **Cantidad de registros del archivo:** 6.615
- **Cantidad de registros procesados:** 3.382
- **Clasificación:** Conflicto armado

### **9.3.31. Instituto de Estudios para el Desarrollo y la Paz (INDEPAZ)**

Es una organización no gubernamental integrada por personas que han construido relaciones de colaboración en la permanente actividad por la paz de Colombia.

**9.3.31.1. Archivo 1** Contiene registros sobre homicidio contra líderes en Colombia entre 2016–2020.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 2016–2020
- **Cantidad de registros del archivo:** 1.000
- **Cantidad de registros procesados:** 1.000
- **Clasificación:** Violencia sociopolítica

**9.3.31.2. Archivo 2** Contiene registros sobre homicidio contra líderes en Colombia entre 2016–2020.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 2016–2020
- **Cantidad de registros del archivo:** 1.080
- **Cantidad de registros procesados:** 1.000
- **Clasificación:** Violencia sociopolítica

Los dos (2) archivos mencionados fueron procesados y arrojaron una cantidad total de registros de .

### **9.3.32. Instituto Nacional de Medicina Legal y Ciencias Forenses (INML)**

Es una entidad del Estado de referencia técnico científica que dirige y controla el sistema de Medicina Legal y Ciencias Forenses en Colombia.

**9.3.32.1. Archivo 1** Contiene registros sobre homicidios selectivos en Colombia del sistema de información de red de desaparecidos y cadáveres (SIRDEC) entre 1989–2018.

- **Fecha de última actualización:** 03-10-2018
- **Periodo documentado:** 1989–2018
- **Cantidad de registros del archivo:** 56.121
- **Cantidad de registros procesados:** 50.474
- **Clasificación:** Conflicto armado

**9.3.32.2. Archivo 2** El Instituto Nacional de Medicina Legal y Ciencias Forenses (INML) compartió estos registros con la Oficina del Alto Comisionado para los Derechos Humanos de las Naciones Unidas en 2013. Esta oficina compartió los registros con HRDAG en el marco del proyecto de vinculación de registros para las Naciones Unidas y publicado en la revista del INML *Forensis* (Ball and Reed Hurtado 2015). Los registros incluidos reportan muertes violentas y accidentales entre 2007–2013.

- **Fecha de última actualización:** 2013
- **Periodo documentado:** 2007–2013
- **Cantidad de registros del archivo:** 226.855
- **Cantidad de registros procesados:** 130.807
- **Clasificación:** Violencia sociopolítica

**9.3.32.3. Archivo 3** Contiene datos sobre desaparición forzada en Colombia entre 1974–2015.

- **Fecha de última actualización:** 03-10-2018
- **Periodo documentado:** 1974–2015
- **Cantidad de registros del archivo:** 1.475
- **Cantidad de registros procesados:** 1.089
- **Clasificación:** Conflicto armado

**9.3.32.4. Archivo 4** Archivo plano con variables pertenecientes al sistema de información SIRDEC, módulo de desaparecidos relacionados con el conflicto armado. Contiene información sobre desaparecidos en Colombia entre 1938 a 2016.

- **Fecha de última actualización:** 18-05-2021

- **Periodo documentado:** 1938–2016
- **Cantidad de registros del archivo:** 122.457
- **Cantidad de registros procesados:** 120.080
- **Clasificación:** Conflicto armado

Los cuatro (4) archivos mencionados fueron procesados y arrojaron una cantidad de 302.450 registros.

### **9.3.33. Jurisdicción Especial para la Paz (JEP)**

Es el componente de justicia del Sistema Integral de Verdad, Justicia, Reparación y no Repetición, creado por el Acuerdo de Paz entre el Gobierno Nacional y las FARC-EP. La JEP tiene la función de administrar justicia transicional y conocer de los delitos cometidos en el marco del conflicto armado que se hubieran cometido antes del 1 de diciembre de 2016.

**9.3.33.1. Archivo 1** Tabla de sistematización de informes recibidos por la SRVR de la JEP del Despacho relator del caso 07 de la JEP.

- **Fecha de última actualización:** 12-02-2021
- **Periodo documentado:** 1973–2016
- **Cantidad de registros del archivo:** 4.459
- **Cantidad de registros procesados:** 399
- **Clasificación:** Conflicto armado

**9.3.33.2. Archivo 2** Tabla de víctimas acreditadas en el caso 07 de la JEP.

- **Fecha de última actualización:** 10-05-2021
- **Periodo documentado:** 1979–2014
- **Cantidad de registros del archivo:** 772
- **Cantidad de registros procesados:** 165
- **Clasificación:** Conflicto armado

**9.3.33.3. Archivo 3** Contiene información sobre víctimas directas e indirectas, contactadas telefónicamente desde la Unidad de Investigación y Acusación (UIA) y comparecientes asociados a estas víctimas. La información cualitativa registrada incluye los datos de las Salas y Secciones donde se generan la necesidad de contacto, Fiscales a cargo, datos de identificación, ubicación y contacto de las víctimas, hechos de violencia, fecha y respuesta al contacto realizado, entre otros.

- **Fecha de última actualización:** 17-05-2020
- **Periodo documentado:** 1997–2007
- **Cantidad de registros del archivo:** 2.441
- **Cantidad de registros procesados:** 1.216
- **Clasificación:** Conflicto armado

**9.3.33.4. Archivo 4** Contiene registros de información tanto cualitativa como cuantitativa sobre muertes ilegítimamente presentadas como bajas en combate incluidas en el Sistema Integrado de Gestión de la Información (SIGI) en el módulo ejecuciones extrajudiciales.

- **Fecha de última actualización:** 17-05-2020
- **Periodo documentado:** 1964–2018
- **Cantidad de registros del archivo:** 167
- **Cantidad de registros procesados:** 105
- **Clasificación:** Conflicto armado

**9.3.33.5. Archivo 5** Base de datos construida por el despacho relator del macrocaso 06 sobre víctimas de la Unión Patriótica a partir de la revisión de expedientes judiciales.

- **Fecha de última actualización:** 22-10-2021
- **Periodo documentado:** 1984–2016
- **Cantidad de registros del archivo:** 1.168
- **Cantidad de registros procesados:** 784
- **Clasificación:** Conflicto armado

**9.3.33.6. Archivo 6** Datos obtenidos del módulo de alistamiento en el cual se desarrollan actividades de recolección, estructuración, consolidación, almacenamiento y custodia de la información, con unas respectivas categorías de análisis de las violaciones en Colombia.

- **Fecha de última actualización:** 17-05-2020
- **Periodo documentado:** 1988–2019
- **Cantidad de registros del archivo:** 24.659
- **Cantidad de registros procesados:** 8.498
- **Clasificación:** Conflicto armado

**9.3.33.7. Archivo 7** Contiene registros sobre homicidio, desaparición forzada, terrorismo, secuestro y delitos contra la integridad sexual en Colombia entre 1990–2008 sistematizados por el GRAI de la JEP a partir de información suministrada por el Ministerio de Defensa Nacional.

- **Fecha de última actualización:** 04-04-2019
- **Periodo documentado:** 1990–2008
- **Cantidad de registros del archivo:** 831
- **Cantidad de registros procesados:** 467
- **Clasificación:** Conflicto armado

**9.3.33.8. Archivo 8** Contiene registros oficiales sobre homicidio, secuestro y extorsión contra la integridad sexual en Colombia entre 2005–2008 sistematizados por el GRAI de la JEP a partir de información suministrada por el Ministerio de Defensa Nacional.

- **Fecha de última actualización:** 26-04-2019
- **Periodo documentado:** 2005–2008
- **Cantidad de registros del archivo:** 165
- **Cantidad de registros procesados:** 27
- **Clasificación:** Conflicto armado

**9.3.33.9. Archivo 9** Contiene la base de datos del universo provisional de hechos (UPH) compartida por el Grupo de análisis de la información (GRAI) de la JEP sobre desaparición forzada, homicidio, tortura, desplazamiento forzado, amenazas, violencia de género, entre otros, en Colombia entre 1962–2020.

- **Fecha de última actualización:** 28-12-2020
- **Periodo documentado:** 1962–2020
- **Cantidad de registros del archivo:** 64.810
- **Cantidad de registros procesados:** 15.027
- **Clasificación:** Conflicto armado

**9.3.33.10. Archivo 10** Registros sistematizados por el GRAI de la JEP a partir de los informes allegados a la JEP hasta junio de 2020 (parte 1).

- **Fecha de última actualización:** 31-10-2020
- **Periodo documentado:** 1960–2019
- **Cantidad de registros del archivo:** 23.407
- **Cantidad de registros procesados:** 5.178
- **Clasificación:** Conflicto armado

**9.3.33.11. Archivo 11** Registros sistematizados por el GRAI de la JEP a partir de los informes allegados a la JEP hasta junio de 2021 (parte 2).

- **Fecha de última actualización:** 31-10-2021
- **Periodo documentado:** 1965–2021
- **Cantidad de registros del archivo:** 5.488
- **Cantidad de registros procesados:** 658
- **Clasificación:** Conflicto armado

**9.3.33.12. Archivo 12** Registros sistematizados por el GRAI de la JEP a partir de los informes allegados a la JEP hasta junio de 2020 (parte 3).

- **Fecha de última actualización:** 31-10-2020
- **Periodo documentado:** 1970–2012
- **Cantidad de registros del archivo:** 33.233
- **Cantidad de registros procesados:** 15.811
- **Clasificación:** Conflicto armado

**9.3.33.13. Archivo 13** Registros sistematizados por el GRAI de la JEP a partir de los informes allegados a la JEP hasta junio de 2021 (parte 4).

- **Fecha de última actualización:** 31-20-2020
- **Periodo documentado:** 1997–1998
- **Cantidad de registros del archivo:** 9.471
- **Cantidad de registros procesados:** 8.436
- **Clasificación:** Conflicto armado

**9.3.33.14. Archivo 14** Registros sistematizados por el GRAI de la JEP a partir de los informes allegados a la JEP hasta junio de 2021 de delitos conexos con la violencia sexual y la violencia basada en genero.

- **Fecha de última actualización:** 01-30-2021
- **Periodo documentado:** 1954–2018
- **Cantidad de registros del archivo:** 3.919
- **Cantidad de registros procesados:** 310
- **Clasificación:** Conflicto armado

Los catorce (14) archivos mencionados fueron procesados y arrojaron una cantidad de 57.081 registros.

### **9.3.34. Justicia Penal Militar (JPM)**

Es aquella justicia especializada que se encarga de la investigación y juzgamiento de los delitos cometidos en actos relacionados con el servicio por miembros uniformados de la Fuerza Pública.

**9.3.34.1. Archivo 1 JPM 1**

- **Fecha de última actualización:** 11-06-2020
- **Periodo documentado:** 1956–2009
- **Cantidad de registros del archivo:** 7.200
- **Cantidad de registros procesados:** 3.005
- **Clasificación:** Conflicto armado

**9.3.34.2. Archivo 2 JPM 2**

- **Fecha de última actualización:** 28-09-2020
- **Periodo documentado:** 1933–2012
- **Cantidad de registros del archivo:** 14.164
- **Cantidad de registros procesados:** 3.005
- **Clasificación:** Conflicto armado

**9.3.34.3. Archivo 3 JPM 3**

- **Fecha de última actualización:** 18-12-2020
- **Periodo documentado:** 1947–2020
- **Cantidad de registros del archivo:** 15.151
- **Cantidad de registros procesados:** 3.005
- **Clasificación:** Conflicto armado

Los tres (3) archivos mencionados fueron procesados y arrojaron una cantidad total de registros de .

**9.3.35. Ministerio de Defensa Nacional**

Es la máxima autoridad en materia de defensa, seguridad y asuntos militares de la República de Colombia; formula, diseña, desarrolla y ejecuta las políticas de defensa y seguridad nacionales; conduce la Fuerza Pública, conformada por las Fuerzas Militares, y la Policía Nacional.

**9.3.35.1. Archivo 1** Información asociada a personas desmovilizados que ingresaron siendo menores de edad al grupo armado FARC-EP. Contiene información de nombres, apellidos, bloque y frente, sexo, fecha de nacimiento, nivel educativo, causa de desmovilización, departamento de desmovilización, el cargo que desempeñó, años de permanencia en las FARC; edad y departamento de reclutamiento, así como del nombre del reclutador.

- **Fecha de última actualización:** 05-05-2021
- **Periodo documentado:** 1994–2018
- **Cantidad de registros del archivo:** 2.538
- **Cantidad de registros procesados:** 1.680
- **Clasificación:** Conflicto armado

**9.3.35.2. Archivo 2** Registros de presuntos reclutamientos asociados a las FARC con categorías de datos personales, edad, grupo armado y estructura, departamento de afectación, información de ingreso al programa de desmovilización del Ministerio de Defensa, sexo, Información de la “entrega,” motivo de la entrega, nivel cultural, cargo general y específico al interior del grupo armado, duración al interior del programa y alias.

- **Fecha de última actualización:** 05-05-2021
- **Periodo documentado:** 1994–2018
- **Cantidad de registros del archivo:** 6.758
- **Cantidad de registros procesados:** 812
- **Clasificación:** Conflicto armado

Los dos (2) archivos mencionados fueron procesados y arrojaron una cantidad de 2.492 registros.

### **9.3.36. Observatorio Surcolombiano de Derechos Humanos, Paz y Territorio (OBSURDH)**

La Corporación OBSURDH es una ONG de Derechos Humanos que trabaja en la investigación, divulgación, defensa y promoción de los DDHH y del DIH en Colombia, así como la denuncia de las violaciones e infracciones de éstos.

#### **9.3.36.1. Archivo 1** Contiene registros sobre “ejecuciones extrajudiciales” en Colombia entre 1984–2012.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 1984–2012
- **Cantidad de registros del archivo:** 237
- **Cantidad de registros procesados:** 237
- **Clasificación:** Conflicto armado

### **9.3.37. Observatorio de Tierras**

El Observatorio de Tierras constituye una red de investigación e intervención pensada para constituir una interfaz intelectual y técnica que permita a las víctimas, el Estado, la opinión pública y los demás sectores de la sociedad analizar y tomar decisiones frente al campo colombiano. La red de investigadores del Observatorio pretende construir y consolidar una red académica que se convierta en un centro técnico y analítico de reflexión sobre las políticas y problemáticas agrarias en el país, se den alertas tempranas y se construyan indicadores sobre la ejecución de las políticas, entre otros.

#### **9.3.37.1. Archivo 1** Contiene registros sobre homicidio y desaparición forzada en Colombia entre 2005–2020.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 2005–2020
- **Cantidad de registros del archivo:**
- **Cantidad de registros procesados:**
- **Clasificación:** Conflicto armado

### **9.3.38. Organización Nacional Indígena de Colombia (ONIC)**

Es una autoridad de gobierno, justicia, legislación y representación de los pueblos indígenas de Colombia. Su objetivo es fortalecer y apoyar el gobierno propio de los pueblos indígenas y su ejercicio de autoridad para que asuman con unidad, autonomía y dignidad, el control de sus territorios y la realización y defensa de sus derechos humanos y colectivos.



**9.3.38.1. Archivo 1** Contiene datos de hechos de violencia en contra de comunidades indígenas en el marco del conflicto armado en Colombia entre 1959–2007.

- **Fecha de última actualización:** 11-10-2019
- **Periodo documentado:** 1959–2007
- **Cantidad de registros del archivo:** 7.541
- **Cantidad de registros procesados:** 3.636
- **Clasificación:** Conflicto armado

### **9.3.39. Personería de Medellín**

Es una entidad del Estado que protege y promueve los derechos humanos, vigila la conducta oficial de quienes desempeñen funciones públicas, protege el interés público y el medio ambiente, aporta a la solución alternativa de conflictos; al fortalecimiento, respeto y garantía de la diversidad y grupos poblacionales, buscando estar más cerca de la comunidad.

**9.3.39.1. Archivo 1** Contiene datos de víctimas de desplazamiento forzado entre 1974–2020.

- **Fecha de última actualización:** 04-03-2021
- **Periodo documentado:** 1974–2020
- **Cantidad de registros del archivo:** 96.644
- **Cantidad de registros procesados:** 40.133
- **Clasificación:** Conflicto armado

### **9.3.40. Procuraduría General de la Nación (PGN)**

Es la entidad que representa a los ciudadanos ante el Estado. Es el máximo organismo del Ministerio Público, conformado además por la Defensoría del Pueblo y las personerías. La principal responsabilidad de la Procuraduría es “prevenir antes que sancionar,” vigilar el actuar de los servidores públicos y advertir cualquier hecho que pueda ser violatorio de las normas vigentes, sin que ello implique coadministración o intromisión en la gestión de las entidades estatales. Los archivos que a continuación se describen son anexos al informe que la PGN entregó a la JEP por mandato de los acuerdos de paz.

**9.3.40.1. Archivo 1** Contiene información de procesos activos en la PGN sobre homicidios selectivos, violencia basada en género y otros delitos en Colombia entre 1991–2016.

- **Fecha de última actualización:** 03-09-2020
- **Periodo documentado:** 1991–2016
- **Cantidad de registros del archivo:** 7.985
- **Cantidad de registros procesados:** 846
- **Clasificación:** Conflicto armado

**9.3.40.2. Archivo 2** Contiene información de procesos archivados en la PGN sobre homicidios selectivos y violencia basada en género y otros delitos en Colombia entre 1989–2016.

- **Fecha de última actualización:** 03-09-2020
- **Periodo documentado:** 1989–2016
- **Cantidad de registros del archivo:** 3.038
- **Cantidad de registros procesados:** 558
- **Clasificación:** Conflicto armado

**9.3.40.3. Archivo 3** Contiene información de sentencias de la PGN sobre homicidios selectivos y violencia basada en género y otros delitos en Colombia entre 2001–2012.

- **Fecha de última actualización:** 03-09-2020
- **Periodo documentado:** 2001–2012
- **Cantidad de registros del archivo:** 1.892
- **Cantidad de registros procesados:** 246
- **Clasificación:** Conflicto armado

Los tres (3) archivos mencionados fueron procesados y arrojaron una cantidad de 1.650 registros.

### **9.3.41. Policía Nacional (PONAL)**

Es un cuerpo armado de naturaleza civil, a cargo de la seguridad pública de la República de Colombia. El fin primordial de la Policía Nacional es el mantenimiento de la convivencia como condición necesaria para el ejercicio de los derechos y libertades públicas y para asegurar que los habitantes de Colombia convivan en paz fundamentada en el código de ética policía. Estos archivos hacen parte de la memoria institucional de HRDAG.

**9.3.41.1. Archivo 1** La oficina de las Naciones Unidas compartió la base de datos que recibió de la Policía Nacional con HRDAG como parte del proyecto de vinculación de registros llevado a cabo por las Naciones Unidas y publicado en la revista académica de la Policía Nacional, *Revista Criminalidad* (Ball and Reed Hurtado 2016). Los registros incluyen reportes de muertes violentas entre 1994–2013.

- **Fecha de última actualización:** 2013
- **Periodo documentado:** 1994–2013
- **Cantidad de registros del archivo:** 164.581
- **Cantidad de registros procesados:** 151.883
- **Clasificación:** Violencia sociopolítica

### **9.3.42. Sistema de monitoreo de riesgos de la Unidad de Investigación Y Acusación - Jurisdicción Especial Para la Paz**

Es una herramienta web de la Unidad de Investigación y Acusación que permite organizar diariamente las noticias publicadas por los medios de comunicación, con relación a los eventos del conflicto armado, la afectación a civiles, y las acciones desplegadas por la Fuerza Pública y otras agencias del Estado colombiano, para combatir a los Grupos Armados Organizados e implementar el Acuerdo de Paz.

**9.3.42.1. Archivo 1** Medios de comunicación: Contiene registro de las noticias que sean indicativas de situaciones de riesgo para los usuarios de la JEP, como las víctimas, los testigos y los demás intervinientes.

- **Fecha de última actualización:** 29-04-2021
- **Periodo documentado:** 2016–2020
- **Cantidad de registros del archivo:** 6.092
- **Cantidad de registros procesados:** 4.434
- **Clasificación:** Conflicto armado

**9.3.42.2. Archivo 2** Redes sociales: Contiene registros de información procedente de las redes sociales sobre noticias de riesgos para los usuarios de la JEP.

- **Fecha de última actualización:** 29-04-2021
- **Periodo documentado:** 2016–2020
- **Cantidad de registros del archivo:** 1.600
- **Cantidad de registros procesados:** 936
- **Clasificación:** Conflicto armado

### **9.3.43. Somos Defensores**

Organizaciones no gubernamentales de derechos humanos Asociación MINGA, Benposta Nación de Muchachos y la Comisión Colombiana de Juristas (CCJ), que busca desarrollar una propuesta integral para prevenir agresiones y proteger la vida de las personas que corren riesgos por su labor como defensores de derechos humanos, cuando resguardan los intereses de grupos sociales y comunidades afectadas por la violencia en Colombia.

**9.3.43.1. Archivo 1** Contiene registros sobre homicidios en Colombia entre 2016–2020.

- **Fecha de última actualización:** 02-01-2021
- **Periodo documentado:** 2016–2020
- **Cantidad de registros del archivo:** 885
- **Cantidad de registros procesados:**
- **Clasificación:** Conflicto armado

### **9.3.44. Unidad Administrativa Especial para la Atención y Reparación Integral a las Víctimas**

Es un instrumento del Estado que sirve como registro nacional de las víctimas de violencia ocurrida en Colombia. Sirve para que el Estado pueda planificar y aplicar mejor las políticas, planes y proyectos que benefician a las víctimas del conflicto armado.

**9.3.44.1. Archivo 1** Contiene registros oficiales de la Unidad para las Víctimas sobre múltiples repertorios de violencia en el marco del conflicto armado en Colombia durante 1940–2019

- **Fecha de última actualización:** 01-19-2020
- **Periodo documentado:** 1940–2019
- **Cantidad de registros del archivo:** 11.671.007
- **Cantidad de registros procesados:** 9.286.794
- **Clasificación:** Conflicto armado

**9.3.44.2. Archivo 2** Base de datos filtrada del RUV por el despacho relator del macrocaso 06 de la JEP relacionada con víctimas identificadas como militantes o miembros de la Unión Patriótica.

- **Fecha de última actualización:** 22-10-2021
- **Periodo documentado:** 1982–2020
- **Cantidad de registros del archivo:** 3.799
- **Cantidad de registros procesados:** 1.723
- **Clasificación:** Conflicto armado

Los dos (2) archivos mencionados fueron procesados y arrojaron una cantidad de 9.288.517 registros.

## 9.4. Códigos

### 9.4.1. Códigos utilizados para la variable “pertenece al conflicto”

```
cat(readLines(files$lideres), sep = '\n')
```

#### 9.4.1.1. FGN - Líderes

```
## #
## # Authors:      Valentina Gómez
## # Maintainers  Valentina Gómez, PB
## # Copyright    2022, HRDAG,
## # =====
## # CO-SIVJRN-data/individual/FGN/is-ca/src/is-ca-fgn_lideres.R
## #
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
## #
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
## #
## stopifnot(endsWith(getwd(),"is-ca"))
## #
## parser <- ArgumentParser()
## parser$add_argument("--input_fgn_lideres",
##                      default=here::here("individual/FGN/clean/output/fgn_lideres.parquet"))
## parser$add_argument("--examples",
##                      default = "individual/FGN/is-ca/hand/example_fgn_lideres.csv")
## parser$add_argument("--output",
##                      default = "output/fgn_lideres.parquet")
## args <- parser$parse_args()
## #
## # ----- function
## #
## # ----- main
## #
## # NB: we are only retaining homicidios for consumado=SI; all is_conflict
## data_fgn_lideres <- read_parquet(args$input_fgn_lideres) %>%
##   mutate(is_conflict = 1)
## #
## # names(data_fgn_lideres)
## # nrow(data_fgn_lideres)
## #
## # table(data_fgn_lideres$conflicto_armado_caso, useNA = "always")
## # table(data_fgn_lideres$vinculacion_persona, useNA = "always")
## # table(data_fgn_lideres$grupo_delito, useNA = "always")
## #
## # caso_is_ca = c("FARC-EP", "PARAMILITARES", "OTRAS GUERRILLAS", "SI")
## #
## # viol_is_ca = c("DESAPARICION FORZADA", "HOMICIDIO DOLOSO", "DESPLAZAMIENTO")
```

```

## #
## # viol_na_ca = c("AMENAZAS", "FEMINICIDIO", "LESIONES PERSONALES")
## #
## #
## # conflict_fgn_lideres <- data_fgn_lideres %>%
## #   mutate(is_conflict = case_when(grupo_delito %in% viol_is_ca &
## #     conflicto_armado_caso %in% caso_is_ca ~ as.integer(1),
## #     grupo_delito %in% viol_na_ca ~ NA_integer_,
## #     grupo_delito %in% viol_is_ca &
## #     conflicto_armado_caso == "NO" ~ as.integer(0),
## #     is.na(conflicto_armado_caso) &
## #     grupo_delito == "DESPLAZAMIENTO" ~ as.integer(1),
## #     conflicto_armado_caso == "SIN GRUPO IDENTIFICADO" &
## #     grupo_delito %in% viol_is_ca ~ as.integer(1),
## #     TRUE ~ NA_integer_))
## #
## # table(conflict_fgn_lideres$is_conflict, useNA = "always")
## #
## # # 0    1  <NA>
## # #566 348  5186
## #
## # nrow <- nrow(conflict_fgn_lideres)
## #
## # sample <- conflict_fgn_lideres %>%
## #   select(recordid, is_conflict)
## #
## # sample <- sample_frac(sample, 0.4)
## #
## # write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
## #
## # write_parquet(data_fgn_lideres, args$output)
## #
## # #done!

```

```
cat(readLines(files$jtr), sep = '\n')
```

#### 9.4.1.2. FGN - UP

```

## #
## # Authors:      SAC
## # Maintainers: SAC, PB
## # Copyright:   2022, HRDAG, GPL v2 or later
## # =====
## # CO-SIVJNRN-data/individual/FGN/is-ca/src/is-ca-fgn-jtr-up.R
## #
## # ----- setup
## # Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## # set.seed(19481210)
## #
## # require(pacman)
## # p_load(argparse, dplyr, here, arrow,assertr, stringr, lubridate)

```

```

##
## stopifnot(str_detect(Sys.getlocale(), "LC_CTYPE=en_US.UTF-8"))
## stopifnot(str_ends(getwd(), "FGN/is-ca"))
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                       default=here::here("individual/FGN/clean/output/up-fgn-transicional.parquet"))
## parser$add_argument("--output",
##                       default = "output/up-fgn-transicional.parquet")
##
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## up_fgn_jtr <- read_parquet(args$input)
##
## up_fgn_jtr <- up_fgn_jtr %>%
##   mutate(is_conflict = 1)
##
## glimpse(up_fgn_jtr)
##
## up_fgn_jtr %>% write_parquet(args$output)
##
## #done

```

```
cat(readLines(files$sijuf_cv), sep = '\n')
```

#### 9.4.1.3. FGN - SIJUF - CEV

```

## #
## # Authors:      VG
## # Maintainers  VG, PA, PB, JGD
## # Copyright    2022, HRDAG,
## # =====
## # CO-SIVJNRN-data/individual/FGN/is-ca/src/is-ca-FISCALIA_datos_sijuf_cv.R
##
## # following the email from Folco Zaffalon (8 abr 2022, 17:16 PDT),
## # and the accompanying file
## # 'SIJUF_casos_perpetrador_ocupacion victima_delitos.docx'
## #
## # Verde = SI CONFLICTO (1)
## # Azul = MUY PROBABLE 0.75~(is_conflict == 1)
## # Amarillo = NO SE SABE 0.5~(is_conflict == 1)
## # Naranja = POCO PROBABLE 0.75~(is_conflict == 0)
## # Rojo = NO CONFLICTO (0)
##
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")

```

```

## set.seed(19481210) # <-- the seed is really important in this run!
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input_SIJUF_CEV",
##                      default=here::here("individual/FGN/clean/output/FISCALIA_datos_sijuf_cv.parquet"),
##                      parser$add_argument("--examples",
##                      default = "individual/FGN/is-ca/hand/example_SIJUF.csv")
## parser$add_argument("--output",
##                      default = "output/FISCALIA_datos_sijuf_cv.parquet")
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## data_SIJUF_CEV <- read_parquet(args$input_SIJUF_CEV)
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##               "AGENTES_ESTATALES","ELN - EJERCITO DE LIBERACION NACIONAL",
##               "EPL - EJERCITO POPULAR DE LIBERACION",
##               "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##               "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##               "GUERRILLA", "OTROS GRUPOS PARAMILITARES", "PARAMILITARES")
##
## perp_no_ca = c("COMUNIDADES INDIGENAS", "DELINCUENCIA COMUN")
##
## data_SIJUF_CEV$fecha_hechos <- as.Date(data_SIJUF_CEV$fecha_hechos, "%d/%m/%Y")
##
##
## #IS CA
##
## casos_verde <- c("CASOS AUTO 92 CORTE CONSTITUCIONAL", "CASOS DEMANDADOS CIDH",
##                 "CASOS EN EL SIDH", "CASOS OIT",
##                 "CASOS SAN JOSE DE APARTADO", "CASOS UP",
##                 "COMPULSAS - JUSTICIA Y PAZ", "CONFLICTO ARMADO",
##                 "ELN", "FARC-EP",
##                 "GENERO, NINEZ Y ADOLESCENCIA HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                 "MARCHA PATRIOTICA", "MASACRES",
##                 "PARAPOLITICA", "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO",
##                 "RELACION DE CIVILES CON ACTORES DE CONFLICTO", "RESTITUCION DE TIERRAS",
##                 "VICTIMAS DE MINAS ANTIPERSONALES",
##                 "VICTIMAS MIEMBROS DE LAS ONG'S", "VICTIMAS",
##                 "VICTIMAS SINDICALISTAS",
##                 "VIOLENCIA SEXUAL CONFLICTO ARMADO",
##                 "VIXCTIMAS DEFENSORES DE DDHH")
##
## casos_amarillo <- c("AFRODESCENDIENTES", "BANDAS EMERGENTES", "VIOLENCIA DE GENERO")
##
## casos_azul <- c("VICTIMAS INDIGENAS", "VICTIMAS LGBT", "PERIODISTAS", "VINCULO FUNC. PUBLICOS CON GRU")

```

```
##
## casos_rojo <- c("TRATA TRANSNACIONAL")
##
## perp_verde <- c("AGENTE DEL ESTADO", "AGUILAS NEGRAS", "AUC", "CORDILLERA ELN", "GAO", "GUERRILLA",
##              "INDEPENDIENTES", "LOS RASTROJOS", "M19", "MACHETEROS", "PARAMILITARES",
##              "ELN", "EPL", "ERP", "ERG", "FARC", "GAO")
## perp_azul <- c("FUERZA PUBLICA", "AGENTE DEL ESTADO")
## perp_naranja <- c("INDEPENDIENTES")
##
## prof_verde <- c("ACTIVIDADES RELACIONADAS CON EL SINDICALISMO",
##               "ACTIVIDADES RELACIONADAS ORG. CIVICAS Y CAMPESINAS", "ACTIVISTA DERECHOS HUMANOS",
##               "ACTIVISTA SINDICAL", "ALCALDE", "ALMIRANTE (FUERZA PUBLICA)",
##               "DESMOVILIZADO", "DESPLAZADO",
##               "DIRIGENTE ORGANIZACION DERECHOS HUMANOS", "DIRIGENTE SINDICAL", "INFANTE DE MARINA",
##               "INTEGRANTE AUTODEFENSAS", "INTEGRANTE GUERRILLA",
##               "JUEZ PENAL CIRCUITO", "JUEZ PENAL MUNICIPAL",
##               "JUEZ PROMISCUO", "LIDER ORGANIZACION CAMPESINA", "LIDER ORGANIZACION COMUNITARIA",
##               "LIDER ORGANIZACION INDIGENA", "LIDER ORGANIZACION POLITICA", "MARINERO ARMADA NACIONAL",
##               "MIEMBRO, AFILIADO, ACTIVISTA DE LA UNION PATRIOTICA", "SECRETARIO DE SINDICATO" )
##
## prof_azul <- c("AEROTECNICO SUBJEFE FUERZA AEREA", "AGENTE POLICIA NACIONAL",
##              "BRIGADIER GENERAL DEL EJERCITO NACIONAL", "BRIGADIER GENERAL INFANTERIA DE MARINA",
##              "BRIGADIER GENERAL POLICIA NACIONAL", "CABO PRIMERO DEL EJERCITO NACIONAL",
##              "CABO PRIMERO POLICIA NACIONAL", "CABO SEGUNDO DEL EJERCITO NACIONAL",
##              "CABO SEGUNDO POLICIA NACIONAL", "CABO TERCERO DEL EJERCITO NACIONAL",
##              "CAPITAN DE CORBETA ARMADA NACIONAL", "CAPITAN DE FRAGATA ARMADA NACIONAL",
##              "CAPITAN DE LA FUERZA AEREA", "CAPITAN DE LA MARINA",
##              "CAPITAN DE NAVIO ARMADA NACIONAL", "CAPITAN DEL EJERCITO NACIONAL",
##              "CAPITAN POLICIA NACIONAL", "CONCEJAL", "CONTRA ALMIRANTE ARMADA NACIONAL",
##              "CONTRALOR MUNICIPAL", "CORONEL DEL EJERCITO NACIONAL",
##              "CORONEL INFANTERIA DE MARINA", "CORONEL POLICIA NACIONAL",
##              "DEFENSOR DEL PUEBLO", "DEFENSOR DEL PUEBLO DELEGADO", "DIPUTADO",
##              "EMPLEADOS AREA DE FISCALIA", "EMPLEADOS DEL AREA DEL CTI DE LA FISCALIA",
##              "ENFERMERAS", "ENFERMERAS JEFES", "ESTUDIANTES UNIVERSITARIOS", "EXALCALDE",
##              "FISCAL DELEG. TRIBUNAL SUP.", "FISCAL ESPECIALIZADO", "FISCAL LOCAL",
##              "FISCAL SECCIONAL", "GANADERO",
##              "GENERAL DE LA FUERZA AEREA", "GENERAL DE LA POLICIA NACIONAL",
##              "GENERAL DE LA REPUBLICA", "GENERAL DE LA REPUBLICA (FUERZA PUBLICA)",
##              "GENERAL DEL EJERCITO NACIONAL", "GENERAL INFANTERIA DE MARINA",
##              "GOBERNADOR", "GOBERNADOR DEPARTAMENTO", "INDIGENA", "INDIGENTE",
##              "INGENIEROS DE PETROLEOS", "INTENDENTE POLICIA NACIONAL", "INVESTIGADOR JUDICIAL C.T.",
##              "INVESTIGADOR JUDICIAL D.A.S.", "INVESTIGADOR JUDICIAL POLICIA NACIONAL",
##              "MAGISTRADO CONSEJO DE ESTADO", "MAGISTRADO CONSEJO SECC. JUDICATURA",
##              "MAGISTRADO CONSEJO SUP. JUDICATURA", "MAGISTRADO TRIBUNAL ADMINISTRATIVO",
##              "MAGISTRADO TRIBUNAL SUPERIOR", "MAGISTRADOS DE TRIBUNALES", "MAYOR DE LA FUERZA AEREA",
##              "MAYOR DEL EJERCITO NACIONAL", "MAYOR GENERAL DEL EJERCITO NACIONAL",
##              "MAYOR GENERAL POLICIA NACIONAL", "MAYOR INFANTERIA DE MARINA",
##              "MAYOR POLICIA NACIONAL", "MIEMBROS CIVILES ARMADA NACIONAL",
##              "MIEMBROS CIVILES DE LA MARINA", "MIEMBROS CIVILES EJERCITO NACIONAL",
##              "MIEMBROS CIVILES POLICIA NACIONAL", "MIEMBROS FUERZAS MILITARES Y POLICIAS",
##              "MIEMBROS RETIRADOS ARMADA NACIONAL", "MIEMBROS RETIRADOS EJERCITO NACIONAL",
##              "MIEMBROS RETIRADOS FUERZA AEREA MIEMBROS RETIRADOS FUERZAS MILITARES Y DE POLICIA",
##              "MIEMBROS RETIRADOS POLICIA NACIONAL", "OCUPACIONES EXCLUSIVAS DE LAS FUERZAS MILITARES",
##              "OFICIALES DE CUBIERTA", "OFICIALES DE LAS FUERZAS MILITARES", "OFICIALES DE POLICIA",
```



```

## "OTROS MIEMBROS ARMADA NACIONAL", "OTROS MIEMBROS DE LA MARINA", "OTROS MIEMBROS EJERCITO NACIONAL",
## "OTROS MIEMBROS FUERZA AEREA", "OTROS MIEMBROS POLICIA NACIONAL",
## "PRESIDENTE DE SINDICATO", "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA",
## "PROFESORES UNIVERSITARIOS", "PROMOTORES DE SALUD", "REINSERTADO",
## "REPRESENTANTE A LA CAMARA", "SACERDOTE", "SARGENTO MAYOR DEL EJERCITO NACIONAL",
## "SARGENTO MAYOR POLICIA NACIONAL", "SARGENTO PRIMERO DEL EJERCITO NACIONAL",
## "SARGENTO PRIMERO POLICIA NACIONAL", "SARGENTO SEGUNDO DEL EJERCITO NACIONAL",
## "SARGENTO SEGUNDO POLICIA NACIONAL", "SARGENTO VICE PRIMERO POLICIA NACIONAL",
## "SARGENTO VICEPRIMERO DEL EJERCITO NACIONAL", "SARGENTO VICEPRIMERO INFANTERIA DE AVIACION",
## "SENADOR DE LA REPUBLICA", "SOLDADO EJERCITO NACIONAL", "SOLDADO FUERZA AEREA",
## "SUB COMISARIO POLICIA NACIONAL", "SUB INTENDENTE POLICIA NACIONAL",
## "SUB OFICIAL DE LA MARINA", "SUB OFICIAL JEFE ARMADA NACIONAL",
## "SUB OFICIAL PRIMERO ARMADA NACIONAL", "SUB OFICIAL SEGUNDO ARMADA NACIONAL",
## "SUB TENIENTE POLICIA NACIONAL", "SUBOFICIALES DE LA POLICIA",
## "SUBOFICIALES DE LAS FUERZAS MILITARES", "SUBTENIENTE DEL EJERCITO NACIONAL", "TECNICOS",
## "TENIENTE CORONEL DE LA FUERZA AEREA", "TENIENTE CORONEL DEL EJERCITO NACIONAL",
## "TENIENTE CORONEL INFANTERIA DE MARINA", "TENIENTE CORONEL POLICIA NACIONAL",
## "TENIENTE DE CORBETA ARMADA NACIONAL", "TENIENTE DE FRAGATA ARMADA NACIONAL",
## "TENIENTE DE LA FUERZA AEREA", "TENIENTE DE LA MARINA",
## "TENIENTE DE NAVIO ARMADA NACIONAL", "TENIENTE DEL EJERCITO NACIONAL",
## "TENIENTE POLICIA NACIONAL", "TRABAJADORES EN SERVICIO SOCIAL Y COMUNITARIO", "TRABAJADORES EN SERVICIO SOCIAL Y COMUNITARIO",
##
## prof_amarillo <- c("AEROTECNICO JEFE FUERZA AEREA", "AEROTECNICO PRIMERO FUERZA AEREA",
## "AGRICULTOR", "AGRONOMOS Y ESPECIALISTAS AGRICOLAS", "ALBANIL", "ARTESANOS",
## "ASISTENTES ADMINISTRATIVOS", "ASISTENTES DE CONTABILIDAD",
## "ASISTENTES DE JUZGADOS, TRIBUNALES",
## "AUXILIARES DE ENFERMERIA", "AUXILIARES DE ODONTOLOGIA",
## "AUXILIARES DE TRIBUNALES, JUZGADOS",
## "BIOLOGOS, BOTANICOS, ZOOLOGOS Y RELACIONADOS",
## "BRIGADIER GENERAL DE LA FUERZA AEREA", "CAJEROS",
## "CAJEROS DE SERVICIOS FINANCIEROS", "CARNICEROS", "COMERCIANTE",
## "COMERCIANTE DE GANADO", "COMISARIOS, INSPECTORES DE POLICIA",
## "CONDUCTOR O AUXILIAR DE TRANSPORTE", "CONTRALOR DEPARTAMENTA",
## "COTERO", "DESEMPLEADO",
## "DIRECTORES Y ADMINISTRADORES DE EDUCACION BASICA Y DIRECTORES Y GERENTES GENERALES DE EMPRESAS",
## "DIRECTORES Y GERENTES GENERALES PRODUCCION DE BIENES", "DIRECTORES Y GERENTES GENERALES DE EMPRESAS",
## "EBANISTA", "INVESTIGADORES Y ANALISTAS DE POLITICA", "EDILES",
## "ELECTRICISTA", "EMBAJADOR", "EMPLEADOS DE BANCA, SEGUROS Y OTROS SERVICIOS FINANCIEROS",
## "ESTUDIANTE PRIMARIA", "ESTUDIANTES SECUNDARIA", "FARMACEUTICOS",
## "FOTOGRAFOS", "FUNCIONARIOS DE PROGRAMAS EXCLUSIVOS DE LA ADMINIS",
## "GEOLOGOS, GEOQUIMICOS Y GEOFISICOS", "GERENTES DE ARQUITECTURA Y CIENCIAS",
## "GUARDIANES DE PRISION", "INGENIERO DE VIAS Y TRANSPORTE", "INGENIERO SANITARIO Y AMBIENTAL",
## "INGENIEROS CIVILES", "INGENIEROS DE MINAS", "INGENIEROS DE SISTEMAS",
## "INGENIEROS ELECTRICOS Y ELECTRONICOS", "INGENIEROS INDUSTRIALES",
## "INGENIEROS MECANICOS", "INGENIEROS METALURGICOS", "INGENIEROS QUIMICOS",
## "INSPECTOR DE CONSTRUCCION", "INSPECTORES DE SANIDAD, SEGURIDAD Y SALUD OCUPACIONAL",
## "INSTRUMENTADORES QUIRURGICOS", "INVESTIGADORES Y CONSULTORES, DESARROLLO ECONOMICO",
## "JEFE DE MISION DIPLOMATICA", "JEFES DE BIBLIOTECA, PUBLICACIONES Y EMPLEADOS DE BIBLIOTECA",
## "JEFES DE OFICINA EN GENERAL", "JEFES DE REGISTRO, DISTRIBUCION Y PROGRAMACION",
## "JUEZ CIVIL CIRCUITO", "JUEZ CIVIL MUNICIPAL", "JUEZ DE FAMILIA", "JUEZ DE MENORES",
## "JUEZ DE PAZ", "JUEZ LABORAL", "MATEMATICOS, ESTADISTICOS Y ACTUARIOS",
## "MECANICO AUTOMOTRIZ", "MEDICOS GENERALES", "MERCADERISTAS", "MEDICOS ESPECIALISTAS",
## "MIEMBROS DEL PODER EJECUTIVO LEGISLATIVO Y JUDICIAL", "MINERO",
## "MINISTRO DEL DESPACHO", "MOTOTAXISTA", "NOTARIO PUBLICO", "OBRERO",

```

```

## "OCUPACIONES PROFESIONALES EN ORGANIZACION Y ADMINI", "ODONTOLOGOS",
## "ORIENTADORES EDUCATIVOS", "OTRAS ACTIVIDADES DE LA CONSTRUCCION", "OTRAS ACTIVIDAD
## "OTRAS OCUPACIONES RELIGIOSAS", "OTRO", "OTROS ARTISTAS", "OTROS ESTUDIANTES",
## "OTROS INVESTIGADORES JUDICIALES", "OTROS JUECES", "OTROS SUPERVISORES", "PANADERO
## "PELUQUEROS, ESTILISTAS Y AFINES", "PENSIONADO", "PERSONAL DIRECTIVO DE LA ADMINI
## "PATRULLERO POLICIA NACIONAL", "PERIODISTAS", "PERSONERO", "POLICIAS", "PRESIDI
## "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA", "POLITOLOGO",
## "PROFESORES DE EDUCACION MEDIA", "PROFESORES DE PREESCOLAR", "PSICOLOGOS",
## "RECICLADOR", "REGISTRADOR NACIONAL ESTADO CIVIL", "SECRETARIAS", "SOLDADOR",
## "SUPERVISORES DE VIGILANCIA", "TECNOLOGOS Y TECNICOS AGROPECUARIOS", "TERAPEUTAS O
## "TRABAJADORA SEXUAL", "VICE ALMIRANTE ARMADA NACIONAL", "VIGILANTES Y GUARDIANES DI
##
## prof_naranja <- c("ABOGADO", "ADMINISTRADOR DE EMPRESAS", "ADMINISTRADORES DE COMERCIO AL POR MENOR"
## "ADMINISTRADORES DE EDUCACION SUPERIOR Y FORMACION",
## "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
## "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES",
## "ARBITROS", "ARQUITECTOS", "ASEADORES Y SERVICIO DOMESTICO",
## "ASISTENTES DE PERSONAL Y SELECCION", "ADMINISTRADOR DE EMPRESAS",
## "ADMINISTRADORES DE COMERCIO AL POR MENOR ADMINISTRADORES DE EDUCACION SUPERIOR Y
## "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
## "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES", "ASEADORES Y SERVICIO DOMESTIC
## "ASISTENTES DE PERSONAL Y SELECCION", "AUXILIARES ADMINISTRATIVOS",
## "AUXILIARES DE ALMACEN Y BODEGA", "AUXILIARES DE CONTABILIDAD", "AUXILIARES DE OFI
## "AUXILIARES DE PERSONAL Y NOMINA", "AUXILIARES DE SERVICIO A PASAJEROS",
## "AYUDANTES DE COCINA Y CAFETERIA", "CARTEROS Y MENSAJEROS", "CHEFS", "COCINERO",
## "COMERCIANTE DE ESMERALDA", "COMPRADORES", "DEPORTISTA",
## "DIRECTOR DEPARTAMENTO ADMINISTRATIVO",
## "DIRECTORES DE FISCALIAS DIRECTORES DE PROGRAMAS DE ESPARCIMIENTO Y DEPORT",
## "DIRECTORES Y GERENTES SERVICIOS FINANCIEROS",
## "DISENADORES DE TEATRO, MODA Y EXHIBICION Y OTROS C",
## "DISENADORES GRAFICOS Y DIBUJANTES ARTISTICOS", "DISENADORES INDUSTRIALES",
## "DOMESTICADORES Y TRABAJADORES DEL CUIDADO DE ANIMA",
## "ECONOMISTAS", "EMPLEADOS DE INFORMACION Y SERVICIO AL CLIENTE",
## "EMPLEADOS DE PUBLICACION Y AFINES", "EMPLEADOS DE VENTAS Y SERVICIOS DE AREOLINEA
## "FISICOS Y ASTRONOMOS", "FISIOTERAPEUTAS",
## "GERENTES DE COMERCIO AL POR MENOR", "GERENTES DE EMPRESAS DE TELECOMUNICACIONES",
## "GERENTES DE MEDIOS DE COMUNICACION Y ARTES ESCENIC", "GERENTES DE OTROS SERVICIOS
## "GERENTES DE OTROS SERVICIOS A LAS EMPRESAS", "GERENTES DE OTROS SERVICIOS ADMINIS
## "GERENTES DE PROGRAMAS DE POLITICA DE DESARROLLO EC GERENTES DE SEGUROS, BIENES RA
## "GERENTES DE SERVICIOS DE COMERCIO EXTERIOR", "GERENTES DE VENTAS, MERCADEO Y PUBL
## "GERENTES FINANCIEROS", "GUIAS DE VIAJES Y TURISMO", "HIGIENISTAS DENTALES",
## "MPULSADORES Y DEMOSTRADORES", "LAVACARROS", "MECANOGRAFOS", "MESEROS", "MUSICOS Y
## "OFICIALES DE BOMBEROS", "ORNAMENTADOR", "OTROS INGENIEROS", "OTROS INSPECTORES",
## "OTROS INSTRUCTORES", "PARTICULAR", "PATRONISTAS PRODUCTOS TEXTILES, CUERO Y PIEL"
## "PILOTOS, INGENIEROS E INSTRUCTORES DE VUELO", "PINTORES, ESCULTORES Y OTROS ARTIS
## "PLOMERO", "PRACTICANTES DE LA MEDICINA ALTERNATIVA", "PROGRAMADORES DE SISTEMAS",
## "QUIMICOS", "RECEPCIONISTAS Y OPERADORES DE CONMUTADOR", "RECREACIONISTAS",
## "REPRESENTANTES DE VENTAS NO TECNICAS", "SUPERINTENDENTE DE SALUD", "SUPERINTENDEN
## "SUPERINTENDENTE Y GERENTE DE ESTABLECIMIENTOS PUBL", "SUPERVISORES DE VENTAS",
## "TESORERO", "TOPOGRAFOS", "VENDEDORES DE MOSTRADOR", "VENDEDORES, VENTAS TECNICAS"
##
## prof_rojo <- c("ACTIVIADES RELACIONADAS CON LA ZAPATERIA", "ACTIVIDADES DE LATONERIA Y PINTURA AUTOM
## "ACTIVIDADES DE PESCA", "ACTIVIDADES RELACIONADAS CON EL HOGAR",
## "ACTIVIDADES RELACIONADAS CON LA CARPINTERIA", "ACTIVIDADES RELACIONADAS CON LA TAUROM

```

```

##          "AGENTES DE ADUANA Y OTROS AGENTES", "BARMAN", "BOMBEROS",
##          "CONSEJEROS DE EMPLEO", "CONTADORES Y AUDITORES FINANCIEROS",
##          "CONTRALOR GENERAL DE LA REPUBLICA", "CONTROLADORES DE TRAFICO AEREO", "DIGITADORES",
##          "OPERADORES DE JUEGOS MECANICOS Y DE SALON", "ORGANIZADORES DE EVENTOS",
##          "OTRAS OCUPACIONES DE SERVICIOS PERSONALES", "OTRAS OCUPACIONES ELEMENTALES DE LAS VEN
##          "OTRAS OCUPACIONES ELEMENTALES DE LOS SERVICIOS OTRAS OCUPACIONES PROFESIONALES EN TER
##          "PRESIDENTE DE LA REPUBLICA",
##          "TECNICOS DENTALES",
##          "TECNICOS EN GRABACION DE AUDIO Y VIDEO", "TECNICOS EN TRANSMISION",
##          "TECNICOS OPTICOS",
##          "TECNICOS Y MECANICOS DE INSTRUMENTOS DE AERONAVEGA  TECNICOS Y MECANICOS DE INSTRUMEN
##          "TECNOLOGOS Y TECNICOS DE LABORATORIO MEDICO Y PATO",
##          "TECNOLOGOS Y TECNICOS EN ARQUITECTURA", "TECNOLOGOS Y TECNICOS EN CIENCIAS BIOLGICAS"
##          "TECNOLOGOS Y TECNICOS EN ELECTROCARDIOGRAFIA Y ELE", "TECNOLOGOS Y TECNICOS EN GEOLOG
##          "TECNOLOGOS Y TECNICOS EN INGENIERA CIVIL TECNOLOGOS Y TECNICOS EN INGENIERA ELECTRICA
##          "TECNOLOGOS Y TECNICOS EN QUIMICA APLICADA TECNOLOGOS Y TECNICOS FORESTALES Y DE RECUR
##          "TRABAJADORES DE ESTACION DE SERVICIO")
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_SIJUF_CEV <- data_SIJUF_CEV %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     vinculacion == "PERPETRADOR" ~ zero,
##     perpetrador %in% perp_verde ~ one,
##     perpetrador %in% perp_azul & prob < 0.75 ~ one,
##     perpetrador %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador1 %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador2 %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador3 %in% perp_is_ca & vinculacion == "VICTIMA" ~ one,
##     perpetrador %in% perp_no_ca ~ zero,
##     perpetrador1 %in% perp_no_ca ~ zero,
##     perpetrador2 %in% perp_no_ca ~ zero,
##     perpetrador3 %in% perp_no_ca ~ zero,
##     profesion %in% prof_verde ~ one,
##     profesion %in% prof_azul & prob < 0.75 ~ one,
##     profesion %in% prof_amarillo & prob < 0.5 ~ one,
##     profesion %in% prof_naranja & prob < 0.75 ~ zero,
##     profesion %in% prof_rojo ~ zero,
##     calidad_profesion %in% prof_naranja & prob < 0.75 ~ zero,
##     categoria %in% casos_verde ~ one,
##     categoria %in% casos_azul & prob < 0.75 ~ one,
##     categoria %in% casos_amarillo & prob < 0.5 ~ one,
##     categoria %in% casos_rojo ~ zero,
##     vinculacion == "OTRA" ~ NA_integer_,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob)
##
##
## table(conflict_SIJUF_CEV$is_conflict, useNA = "always")
##
## # before Folco's changes

```

```

## # 0      1      <NA>
## #70429  1444 112509
##
## # after Folco's changes
## # 0      1      <NA>
## # 72386  2669 109327
##
## nrows <- nrow(conflict_SIJUF_CEV)
##
## sample <- conflict_SIJUF_CEV %>%
##   filter(is_conflict ==0) %>%
##   select(recordid, is_conflict)
##
## sample <- sample_frac(sample, 0.4)
##
## # write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
## write_parquet(conflict_SIJUF_CEV, args$output)
##
## # -----
##
## # names(data_SIJUF_CEV)
## # nrow(data_SIJUF_CEV)
## #
## # table(data_SIJUF_CEV$grupo_delito, useNA = "always")
## # table(data_SIJUF_CEV$clasificacion_conflicto_armado, useNA = "always")
## #
## # #Filtrando base: contiene y no contiene el hecho de desaparición
## # desa <- data_SIJUF_CEV %>%
## #   filter(grupo_delito == "DESAPARICION FORZADA")
## #
## # no_desa <- data_SIJUF_CEV%>%
## #   filter(!grupo_delito == "DESAPARICION FORZADA")
## #
## # #recategorizando variable
## #
## # desaparicion <- desa %>%
## #   mutate(car_per_cat=case_when(
## #     str_detect(calidad_profesion, "AUTODEFENSAS") ~ "AUTODEFENSAS",
## #     str_detect(calidad_profesion, "COMERCIANTE") ~ "COMERCIANTE",
## #     str_detect(calidad_profesion,"HOGAR|AMA DE CASA|DOMESTICO")~ "HOGAR",
## #     str_detect(calidad_profesion, "AGRICULTOR|AGROPECUARIO|AGRO") ~ "AGRICULTOR",
## #     str_detect(calidad_profesion,"VIGILANCIA|SEGURIDAD") ~ "SEGURIDAD",
## #     str_detect(calidad_profesion, "DESMOVILIZADO|REINSERTADO|EXCOMBATIENTE") ~ "DESMOVILIZADO",
## #     str_detect(calidad_profesion, "AEROTECNICO|TECNICO ARMADA NACIONAL$|ARMADA|INFANTERIA|FUERZA A")
## #     str_detect(calidad_profesion,"SARGENTO|CAPITAN|BRIGADIER|POLICIA|CABO|SOLDADO BACHILLER|TENIEN")
## #     str_detect(calidad_profesion, "VENDEDOR|VENTAS") ~ "TRABAJO INDEPENDIENTE",
## #     calidad_profesion == "ALBANIL" ~ "TRABAJO INDEPENDIENTE",
## #     calidad_profesion == "CONSTRUCCION" ~ "TRABAJO INDEPENDIENTE",
## #     str_detect(calidad_profesion, "AFRODESCENDIENTE|INDIGENA") ~ "GRUPOS ETNICOS",
## #     calidad_profesion == "MINERO" ~ "MINERO",
## #     str_detect(calidad_profesion, "EBANISTA|CARNICERO|GANADERO|OBRERO|CONSTRUCCION|VENDEDOR|VENTAS")
## #     str_detect(calidad_profesion, "CAMPELINAS|JORNALERO|PESCADOR|CAMPO|CAMPELINO|PESCA") ~ "CAMPELINO"

```

```

## #   str_detect(calidad_profesion,"DESPLAZADO")~ "VICTIMA CONFLICTO",
## #   str_detect(calidad_profesion, "JUDICIAL|JUEZ|FISCAL|MAGISTRADO|PODER EJECUTIVO") ~ "FUNCIONARIO",
## #   str_detect(calidad_profesion, "FUNCIONARIO|CONTRALOR|PROCURADOR|PERSONERO|DIRECTOR DE CARCEL|N",
## #   str_detect(calidad_profesion, "EDIL|UP|POLITICO|MARCHA PATRIOTICA|CANDIDATO|DIPUTADO|CONCEJAL|",
## #   str_detect(calidad_profesion, "ESTUDIANTE") ~ "ESTUDIANTE",
## #   str_detect(calidad_profesion,"DERECHO|TRIBUNAL|JUZGADO")~ "SECTOR JUDICIAL",
## #   str_detect(calidad_profesion,"INDIGENTE") ~ "HABITANTE DE CALLE",
## #   str_detect(calidad_profesion,"TRABAJADORA SEXUAL") ~ "TRABAJO SEXUAL",
## #   str_detect(calidad_profesion, "EDUCATIVO|PROFESOR|EDUCADOR|EDUCACION|DOCENTES") ~ "EDUCACION",
## #   str_detect(calidad_profesion, "SINDICAL|SINDICATO")~ "SINDICALISTA",
## #   str_detect(calidad_profesion, "COMUNIDAD LGBT") ~ "COMUNIDAD LGBT",
## #   str_detect(calidad_profesion, "CONDUCTOR|TRANSPORTE|TRANSITO") ~ "TRANSPORTE",
## #   str_detect(calidad_profesion, "MIEMBRO GUERRILLA|GUERRILLERO|FARC|ELN|GUERRILLA") ~ "GUERRILLERO",
## #   str_detect(calidad_profesion, "BIBLIOTECA|ADMINISTRADOR DE EMPRESAS") ~ "PROFESIONALES",
## #   str_detect(calidad_profesion,"MEDICO") ~ "SALUD",
## #   calidad_profesion == "OTRO" ~ NA_character_,
## #   calidad_profesion == "OTRAS OCUPACIONES ELEMENTALES DE LOS SERVICIOS" ~ NA_character_,
## #   calidad_profesion == "OTRAS ACTIVIDADES U OFICIOS"~ NA_character_,
## #   calidad_profesion == "PARTICULAR" ~ NA_character_,
## #   calidad_profesion == "EMPLEADOS DE PUBLICACION Y AFINES" ~ NA_character_,
## #   TRUE ~ calidad_profesion))
##
## #done!

```

```
cat(readLines(files$spoa_cv), sep = '\n')
```

#### 9.4.1.4. FGN - SPOA - CEV

```

## #
## # Authors:      Maria Ortiz, Valentina Gómez, Paula Amado, PB
## # Maintainers  Maria Ortiz
## # Copyright    2022, HRDAG,
## # =====
## # CO-SIVJRRR-data/individual/FGN/is-ca/src/is-ca-FISCALIA_datos_spoa_cv.R
##
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow,assertr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input_SPOA_CEV",
##                      default=here::here("individual/FGN/clean/output/FISCALIA_datos_spoa_cv.parquet"),
##                      )
## parser$add_argument("--examples",
##                      default = "individual/FGN/is-ca/hand/example_SPOA.csv")
## parser$add_argument("--output",

```

```

##                                     default = "individual/FGN/is-ca/output/FISCALIA_datos_spoa_cv.parquet")
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## data_SPOA_CEV <- read_parquet(args$input_SPOA_CEV)
##
##
## # #IS -CA
##
## viol_is_ca = c("DESAPARICION FORZADA", "DESPLAZAMIENTO", "HOMICIDIO DOLOSO",
##              "RECLUTAMIENTO Ilicito")
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##              "AGENTES_ESTATALES", "ELN - EJERCITO DE LIBERACION NACIONAL",
##              "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##              "FUERZA PUBLICA - DELITOS", "OTRAS GUERRILLAS")
##
## cat_caso_is_ca = c("AUTO 092 (LIGA DE MUJERES DESPLAZADAS", "CASOS UP",
##                  "CONFLICTO ARMADO", "CRIMEN DE GUERRA",
##                  "ELN", "FARC-EP", "HOMICIDIO PRESENTADO COMO BAJA POR FUERZA PUBLICA",
##                  "HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                  "LESA HUMANIDAD", "PARAPOLITICA", "POLICIA NACIONAL VINCULOS GRUPOS ILEGALES",
##                  "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO", "UNION PATRIOTICA",
##                  "MILITARES VINCULOS GRUPOS ILEGALES", "RECLUTAMIENTO DE MENORES",
##                  "VINCULO FUNC. PUBLICOS CON GRUPOS ILEGALES",
##                  "RELACION DE CIVILES CON ACTORES DE CONFLICTO",
##                  "VINCULOS F.P CON GRUPOS AL MARGEN DE LA LEY", "VIOLACION DD.HH",
##                  "VIOLENCIA SEXUAL CONFLICTO ARMADO", "VIXCTIMAS DEFENSORES DE DDHH")
##
## # is -ca colores
##
##
## perp_verde <- c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##               "AUTODEFENSAS",
##               "AUTODEFENSAS UNIDAS DE COLOMBIA",
##               "BANDAS EMERGENTES",
##               "EJERCITO DE LIBERACION NACIONAL",
##               "ELN - EJERCITO DE LIBERACION NACIONAL",
##               "EPL - EJERCITO POPULAR DE LIBERACION",
##               "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##               "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##               "FARC",
##               "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##               "FARC-EP",
##               "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##               "FARC.",
##               "FUERZA PUBLICA",
##               "MJBC - MOVIMIENTO JAIME BATEMAN CAYON",
##               "OTROS GRUPOS PARAMILITARES",
##               "AGENTES_ESTATALES",
##               "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",

```

```
## "AUTODEFENSAS",
## "AUTODEFENSAS UNIDAS DE COLOMBIA",
## "BANDAS EMERGENTES",
## "CAP - COMANDO ARMADO POPULAR",
## "EJERCITO DE LIBERACION NACIONAL",
## "ELN - EJERCITO DE LIBERACION NACIONAL",
## "EPL - EJERCITO POPULAR DE LIBERACION",
## "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
## "FARC",
## "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
## "FARC-EP",
## "FARC.",
## "FP - DELITOS",
## "FUERZA PUBLICA",
## "GUERRILLA",
## "MPLICADO FUNCIONARIO PUBLICO",
## "MOVIMIENTO 26 DE ABRIL CARLOS PIZARRO LEON",
## "OTROS GRUPOS PARAMILITARES",
## "PARAMILITARES",
## "AGENTES_ESTATALES",
## "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
## "AUTODEFENSAS",
## "BANDAS EMERGENTES",
## "EJERCITO DE LIBERACION NACIONAL",
## "ELN - EJERCITO DE LIBERACION NACIONAL",
## "EPL - EJERCITO POPULAR DE LIBERACION",
## "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
## "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
## "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
## "MOVIMIENTOS POLITICOS",
## "PARAMILITARES")
##
## perp_rojo <- c("DELINCUENCIA COMUN","COMUNIDADES INDIGENAS")
##
## perp_amarillo <- c("CARTEL DEL NORTE DEL VALLE","NARCOTRAFICO",
## "CMR - COMANDO MILICIANO REVOLUCIONARIO",
## "FP - DELITOS",
## "IMPLICADO FUNCIONARIO PUBLICO")
##
## perp_naranja <- c("MOVIMIENTOS POLITICOS")
##
## #a35_titulo
## titulo_rojo <- c("DE LA PROTECCION DE LA INFORMACION Y DE LOS DATOS",
## "DE LOS DELITOS CONTRA LA SALUD PUBLICA",
## "DE LOS DELITOS CONTRA LOS ANIMALES",
## "DE LOS DELITOS CONTRA LOS DERECHOS DE AUTOR",
## "DECRETOS",
## "DELITOS CONTRA EL ORDEN ECONOMICO SOCIAL",
## "DELITOS CONTRA EL ORDEN ECONOMICO Y SOCIAL",
## "DELITOS CONTRA LA ADMINISTRACION DE PUBLICA",
## "DELITOS CONTRA LA EFICAZ Y RECTA IMPARTICION DE JUSTICIA",
## "DELITOS CONTRA LA FAMILIA",
## "DELITOS CONTRA LA FE PUBLICA",
## "DELITOS CONTRA LA INTEGRIDAD MORAL",
```

```
##           "DELITOS CONTRA LA SALUD PUBLICA")
##
## titulo_verde <- c("DE LOS DELITOS CONTRA EL REGIMEN CONSTITUCIONAL Y LEGAL",
##                 "DELITOS CONTRA LAS PERSONAS Y BIENES PROTEGIDOS POR EL DERECH",
##                 "DELITOS CONTRA PERSONAS Y BIENES PROTEGIDOS POR EL D.I.H.")
##
## titulo_amarillo <- c("DE LOS DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                     "DELITOS CONTRA LA LIBERTAD INDIVIDUAL Y OTRAS GARANTIAS",
##                     "DELITOS CONTRA LA LIBERTAD, INTEGRIDAD Y FORMACION SEXUALES",
##                     "DELITOS CONTRA LA VIDA Y LA INTEGRIDAD PERSONAL",
##                     "LEYES",
##                     "SIN DELITO")
##
## titulo_naranja <- c("DELITOS CONTRA EL PATRIMONIO ECONOMICO",
##                     "DELITOS CONTRA LA EXISTENCIA Y SEGURIDAD DEL ESTADO",
##                     "DELITOS CONTRA LA SEGURIDAD PUBLICA",
##                     "DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                     "DELITOS CONTRA MECANISMOS DE PARTICIPACION DEMOCRATICA")
##
## #tipo_hecho
## delito_rojo <- c("ABANDONO ART 128 C.P.",
##                 "ABANDONO ART. 127 C.P.",
##                 "ABANDONO DE HIJO FRUTO DE ACCESO CARNAL VIOLENTO ABUSIVO O DE INSEMINACION ARTIFICI",
##                 "ABUSO DE CONFIANZA CALIFICADO. ART. 250 C.P.",
##                 "ABUSO DE CONFIANZA. ART. 249 C.P.",
##                 "ACAPARAMIENTO. ART. 297 C.P.",
##                 "ADOPCION IRREGULAR. ART. 232",
##                 "AGIOTAJE. ART. 301",
##                 "APLICACION FRAUDULENTE DE CREDITO OFICIALMENTE REGULADO ART. 311 C.P.",
##                 "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO. ART. 252 C.P.",
##                 "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P.",
##                 "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P. INCISO 1",
##                 "ASOCIACION PARA LA COMISION DE UN DELITO CONTRA LA ADMINISTRACION PUBLICA ART. 434",
##                 "CALUMNIA. ART. 221 C.P.",
##                 "CAPTACION MASIVA Y HABITUAL DE DINEROS ART. 316 C.P.",
##                 "CAZA ILEGAL ART. 336",
##                 "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 172 C.P.",
##                 "CIRCULACION Y USO DE EFECTO OFICIAL O SELLO FALSIFICADO. ART. 281 C.P.",
##                 "CIRCUNSTANCIA DE AGRAVACION PUNIBLE ART 377B CP LEY 1453 DE 2011",
##                 "COHECHO IMPROPIO ART. 406 C.P.",
##                 "COHECHO POR DAR U OFRECER ART. 407 C.P.",
##                 "COHECHO PROPIO ART. 405 C.P.",
##                 "CONTRABANDO ART. 319 C.P.",
##                 "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 410 C.P.",
##                 "CORRUPCION DE ALIMENTOS, PRODUCTOS MEDICOS O MATERIAL PROFILACTICO ART. 372 C.P.",
##                 "CORRUPCION PRIVADA ART. 250A C.P.",
##                 "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##                 "DAÑO INFORMatico ART 269D LEY 1273 DE 2009",
##                 "DE LA PRESTACION, ACCESO O USO ILEGALES DE LOS SERVICIOS DE TELECOMUNICACIONES ART",
##                 "DEFRAUDACION A LOS DERECHOS PATRIMONIALES DE AUTOR. ART. 271 C.P.",
##                 "DEFRAUDACION DE FLUIDOS. ART. 256 C.P.",
##                 "DENEGACION DE INSCRIPCION ART. 396 C.P.",
##                 "DESTRUCCION, SUPRESION U OCULTAMIENTO DE DOCUMENTO PUBLICO. ART. 292 C.P.",
##                 "DESTRUCCION, SUPRESION Y OCULTAMIENTO DE DOCUMENTO PRIVADO. ART. 293 C.P.",
```



## "EJERCICIO ARBITRARIO DE LA CUSTODIA DE HIJO MENOR DE EDAD ART. 230A C.P. AD. LEY 89  
## "EJERCICIO ILICITO DE ACTIVIDAD MONOPOLISTICA DE ARBITRIO RENTISTICO ART. 312 C.P."  
## "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUE. ART. 248 C.P.",  
## "EMISIONES ILEGALES. ART. 276 C.P.",  
## "ENRIQUECIMIENTO ILICITO ART. 412 C.P.",  
## "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P.",  
## "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P. INFERIOR 100 SMLM",  
## "ESTAFA. ART. 246 C.P.",  
## "ESTIMULO AL USO ILICITO ART. 378 C.P.",  
## "EXPORTACION O IMPORTACION FICTICIA ART. 310 C.P.",  
## "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",  
## "FALSA AUTOACUSACION ART. 437 C.P.",  
## "FALSA DENUNCIA ART. 435 C.P.",  
## "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 436 C.P.",  
## "FALSEDAD EN DOCUMENTO PRIVADO. ART. 289 C.P.",  
## "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO. ART. 286 C.P.",  
## "FALSEDAD MARCARIA. ART. 285 C.P.",  
## "FALSEDAD MARCARIA. ART. 285 C.P. INCISO 1",  
## "FALSEDAD MATERIAL EN DOCUMENTO PUBLICO. ART. 287 C.P.",  
## "FALSEDAD PARA OBTENER PRUEBA DE HECHO VERDADERO. ART. 295 C.P.",  
## "FALSEDAD PERSONAL. ART. 296 C.P.",  
## "FALSIFICACION DE EFECTO OFICIAL TIMBRADO. ART. 280 C.P.",  
## "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA. ART. 273 C.P.",  
## "FALSIFICACION O USO FRAUDULENTO DE SELLO OFICIAL. ART. 279 C.P.",  
## "FALSO TESTIMONIO ART. 442 C.P.",  
## "FAVORECIMIENTO ART. 446 C.P.",  
## "FAVORECIMIENTO DE CONTRABANDO ART. 320 C.P.",  
## "FECUNDACION Y TRAFICO DE EMBRIONES ART. 134 C.P.",  
## "FEMINICIDIO ART. 104A C.P.",  
## "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",  
## "FRAUDE PROCESAL ART. 453 C.P.",  
## "FUGA DE PRESOS ART. 448 C.P.",  
## "GESTION INDEBIDA DE LOS RECURSOS SOCIALES. ART. 260 C.P.",  
## "HOMICIDIO CULPOSO ART. 109 C.P.",  
## "HOMICIDIO POR PIEDAD ART. 106 C.P.",  
## "HURTO CALIFICADO. ART. 240 C.P.",  
## "HURTO POR MEDIOS INFORMATICOS Y SEMEJANTES ART. 269I LEY 1273 DE 2009",  
## "HURTO. ART. 239 C.P.",  
## "ILICITA EXPLOTACION COMERCIAL ART. 303",  
## "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",  
## "IMITACION O SIMULACION DE ALIMENTOS, PRODUCTOS O SUSTANCIAS ART. 373 C.P.",  
## "IMPEDIMENTO O PERTURBACION DE LA CELEBRACION DE AUDIENCIAS PUBLICAS ART. 454C C.P.",  
## "INASISTENCIA ALIMENTARIA ART. 233 C.P.",  
## "INCENDIO ART. 350 C.P.",  
## "INCESTO. ART. 237 C.P.",  
## "INDUCCION A AYUDA AL SUICIDIO ART. 107 C.P.",  
## "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 445 C.P.",  
## "INJURIA. ART. 220 C.P.",  
## "INTERES INDEBIDO EN LA CELEBRACION DE CONTRATOS ART. 409 C.P.",  
## "INTERVENCION EN POLITICA ART. 422 C.P.",  
## "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",  
## "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",  
## "IRRESPECTO A CADAVERES. ART. 204 C.P.",  
## "LESIONES AL FETO ART. 125 C.P.",

## "LESIONES CULPOSAS AL FETO ART. 126 C.P.",  
## "LESIONES CULPOSAS ART. 120 C.P.",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O M  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCI  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL TRANSITORIA ART.114 INC  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA PERMANENTE ART.115",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA TRANSITORIA ART.115",  
## "LESIONES CULPOSAS ART. 120 C.P. INCISO 1",  
## "LESIONES CULPOSAS ART. 120 C.P.AGRAVADO POR NO LICENCIA O LICENCIA SUSPENDIDA ART.  
## "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 D  
## "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 90 DIAS ART. 112 C  
## "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.  
## "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA AFECTA ROSTRO ART.  
## "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA PERMANENTE ART. 11  
## "LESIONES PERSONALES CULPOSAS ART. 120 C.P. CON DEFORMIDAD FISICA TRANSITORIA ART. 1  
## "MALTRATO ANIMAL ART. 339A C.P.",  
## "MALTRATO MEDIANTE RESTRICCIÓN A LA LIBERTAD FISICA. ART. 230 C.P.",  
## "MALTRATO POR DESCUIDO O ABANDONO EN PERSONA MAYOR DE 60 AÑOS",  
## "MALVERSACION Y DILAPIDACION DE BIENES FAMILIARES. ART. 236 C.P.",  
## "MANIPULACION FRAUDULENTE DE ESPECIES INSCRITAS EN EL REGISTRO NACIONAL DE VALORES I  
## "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",  
## "OBTENCION DE DOCUMENTO PUBLICO FALSO ART. 288 C.P.",  
## "OCULTAMIENTO, ALTERACION O DESTRUCCION DE ELEMENTO MATERIAL PROBATORIO ART.454B C.  
## "OCULTAMIENTO, RETENCION Y POSESION ILICITA DE CEDULA ART. 395 C.P.",  
## "OFRECIMIENTO ENGAÑOSO DE PRODUCTOS Y SERVICIOS. ART. 300 C.P.",  
## "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P.",  
## "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P. MODIFICADO ART. 18 LEY 1121 DE 200  
## "OMISION DEL AGENTE RETENEDOR O RECAUDADOR ART. 402 C.P.",  
## "PANICO ECONOMICO ART. 302 C.P.",  
## "PARTO O ABORTO PRETERINTENCIONAL ART. 118 C.P.",  
## "PECULADO CULPOSO ART. 400 C.P.",  
## "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 399 C.P.",  
## "PECULADO POR APROPIACION ART. 397 C.P.",  
## "PECULADO POR USO ART. 398 C.P.",  
## "PESCA ILEGAL ART. 335",  
## "PORNOGRAFIA CON MENORES ART. 218 C.P.",  
## "PORTE DE SUSTANCIAS ART. 383 C.P.",  
## "PREVARICATO POR ACCION ART. 413 C.P.",  
## "PREVARICATO POR OMISION ART. 414 C.P.",  
## "PROPAGACION DE EPIDEMIA ART. 369 C.P.",  
## "PROPAGACION DEL VIRUS DE INMUNODEFICIENCIA HUMANA O DE LA HEPATITIS B ART. 370 C.P  
## "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P.",  
## "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P. ADICIONADO LEY 1329 DE 2009",  
## "RECEPTACION ART. 327 C C.P.",  
## "RECEPTACION ART. 447 C.P.",  
## "REPETIBILIDAD DEL SER HUMANO ART. 133 C.P.",  
## "SABOTAJE. ART. 199 C.P.",  
## "SIN DELITO",  
## "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",  
## "SOBORNO ART. 444 C.P.",  
## "SOBORNO EN LA ACTUACION PENAL ART. 444A C.P. AD. LEY 890/2004 ART.10",  
## "SUMINISTRO A MENOR ART. 381 C.P.",  
## "SUPRESION, ALTERACION O SUPOSICION DEL ESTADO CIVIL. ART. 238 C.P.",  
## "SUSTRACCION DE BIEN PROPIO. ART. 254 C.P.",

```
## "SUSTRACCION DE COSA PROPIA AL CUMPLIMIENTO DE DEBERES CONSTITUCIONALES O LEGALES A
## "TRAFICO DE MIGRANTES ART. 188 C.P. MOD. LEY 747/2002 ART.1",
## "TRAFICO DE MONEDA FALSIFICADA. ART. 274 C.P.",
## "TRAFICO DE NIÑAS, NIÑOS Y ADOLESCENTES ART 188C LEY 1453 DE 2011",
## "TRAFICO, ELABORACION Y TENENCIA DE ELEMENTOS DESTINADOS A LA FALSIFICACION DE MONEDAS
## "TRANSFERENCIA NO CONSENTIDA DE ACTIVOS VALIENDOSE DE ALGUNA MANIPULACION INFORMATICA
## "TRATA DE PERSONAS ART. 188A C.P.",
## "TRATA DE PERSONAS TRANSNACIONAL ART. 188A C.P.",
## "TURISMO SEXUAL. ART.219 C.P. MOD. POR ART 23 L, 1336 DE 2009",
## "URBANIZACION ILEGAL ART. 318 C.P.",
## "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
## "USO DE SOFTWARE MALICIOSO ART 269E LEY 1273 DE 2009",
## "USO ILEGITIMO DE PATENTES ART. 307 C.P.",
## "USO ILEGITIMO DE PATENTES ART. 307 C.P. INCISO 1. EL QUE FABRIQUE PRODUCTO SIN AUTORIZACION
## "USURA ART. 305 C.P.",
## "USURA ART. 305 C.P. INCISO 1 A CAMBIO DE PRESTAMO DE DINERO O POR CONCEPTO DE VENTA
## "USURPACION DE AGUAS. ART. 262 C.P.",
## "USURPACION DE DERECHOS DE PROPIEDAD INDUSTRIAL Y DERECHOS DE OBTENTORES DE VARIEDADES
## "UTILIZACION O FACILITACION DE MEDIOS DE COMUNICACION PARA OFRECER ACTIVIDADES SEXUALES CON MENORES
## "VIOLACION A LA LIBERTAD RELIGIOSA. ART. 201 C.P.}VIOLACION A LOS DERECHOS MORALES DE AUTOR
## "VIOLACION A LOS DERECHOS MORALES DE AUTOR. ART. 270 C.P. N.1",
## "VIOLACION A LOS DERECHOS PATRIMONIALES DE AUTOR Y DERECHOS CONEXOS ART. 271 C.P. M
## "VIOLACION AL REGIMEN LEGAL O CONSTITUCIONAL DE INHABILIDADES E INCOMPATIBILIDADES
## "VIOLACION DE HABITACION AJENA. ART. 189 C.P.",
## "VIOLACION DE LA RESERVA INDUSTRIAL O COMERCIAL ART. 308 C.P.",
## "VIOLENCIA INTRAFAMILIAR ART. 229 C.P.")
##
## delito_naranja <- c("ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P. INCISO 1",
## "ACCESO ABUSIVO A UN SISTEMA INFORMatico ART 269A LEY 1273 DE 2009",
## "ACCESO ABUSIVO A UN SISTEMA INFORMatico. ART. 195 C.P.",
## "ACOSO SEXUAL ART. 210A C.P. ADICIONADO LEY 1257 DE 2008",
## "ACTO SEXUAL VIOLENTO CON MENOR DE CATORCE AÑOS ART. 209 C.P.",
## "ACTO SEXUAL VIOLENTO. ART. 206 C.P.",
## "DEMANDA DE EXPLOTACION SEXUAL COMERCIAL MENOR DE 18 AÑOS ADICIONADO LEY 1329 DE 2009",
## "DISPARO DE ARMA DE FUEGO ART. 356A CP LEY 1453 DE 2011",
## "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
## "EXPLOTACION DE MENORES DE EDAD",
## "FRAUDE A RESOLUCION JUDICIAL ART. 454 C.P.",
## "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
## "HOMICIDIO ART. 103 C.P.",
## "INSTIGACION A DELINQUIR ART. 348 C.P.",
## "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
## "INTERCEPTACION DE DATOS INFORMATICOS, ART 269C LEY 1273 DE 2009",
## "LESIONES ART. 111 C.P.",
## "LESIONES ART. 113 C.P.",
## "LESIONES CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANNO O MIEMBRO ART.116",
## "LESIONES CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCISO 2",
## "LESIONES CON PERTURBACION PSIQUICA PERMANENTE ART.115",
## "LESIONES CON PERTURBACION PSIQUICA TRANSITORIA ART.115",
## "LESIONES PERSONALES AGRAVADAS POR CIRCUNSTANCIAS ART 104 C.P. ART. 119 C.P.",
## "LESIONES PERSONALES ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.P. INCISO 1",
## "LESIONES PERSONALES CON DEFORMIDAD FISICA AFECTA ROSTRO ART. 113 C.P.",
## "LESIONES PERSONALES CON DEFORMIDAD FISICA PERMANENTE ART. 113 C.P. INCISO 2",
## "LESIONES PERSONALES CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 DIAS ART. 112 C.P. INCISO 1"
```

```
## "OBSTACULIZACION ILEGITIMA DEL SISTEMA INFORMATICO O RED DE TELECOMUNICACION ART
## "OMISION DE APOYO ART. 424 C.P.",
## "OMISION DE SOCORRO ART. 131 C.P.",
## "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
## "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
## "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
## "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
## "SECUESTRO SIMPLE ART. 168 C.P.",
## "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO CUANDO SE TRAFIQUE CON LA PERSONA SECUES
## "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO SOBREVENIR MUERTE O LESIONES PERSONALES
## "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
## "SUPLANTACION DE SITIOS WEB PARA CAPTURAR DATOS PERSONALES ART 269G LEY 1273 DE 1
## "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.
## "TRAFICO DE INFLUENCIAS DE SERVIDOR PUBLICO ART. 411 C.P.",
## "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C.",
## "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
## "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
## "UTILIZACION INDEBIDA DE INFLUENCIAS DERIVADAS DEL EJERCICIO DE FUNCION PUBLICA
## "VIOLACION DE INMUNIDAD DIPLOMATICA ART. 465 C.P.",
## "VOTO FRAUDULENTO ART. 391 C.P")
##
##
## delito_verde <- c("ABORTO FORZADO EN PERSONA PROTEGIDA ART. 139E",
## "ACCESO CARNAL ABUSIVO EN PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 138",
## "ACCESO CARNAL VIOLENTO EN PERSONA PROTEGIDA ART. 138 C.P.",
## "ACTOS DE TERRORISMO ART. 144 C.P.",
## "ACTOS SEXUALES CON PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 139A",
## "ACTOS SEXUALES VIOLENTOS EN PERSONA PROTEGIDA ART. 139 C.P.",
## "ADIC.L.579/2002 ART.2.EMPLEO,PRODUCCION,COMERCIALIZACION Y ALMACENAMIENTO DE MINA
## "ADIC.L.579/2002 ART.3.AYUDA E INDUCCION AL EMPLEO, PRODUCCION Y TRANSFERENCIA DE
## "ADMINISTRACION DE RECURSOS RELACIONADOS CON ACTIVIDADES TERRORISTAS ART. 345 C.P.
## "AMENAZAS CONTRA DEFENSORES DE DERECHOS HUMANOS Y SERVIDORES PUBLICOS ART. 188E",
## "APOLOGIA DEL GENOCIDIO ART. 102 C.P.",
## "ASONADA ART. 469 C.P.",
## "ATAQUE CONTRA OBRAS E INSTALACIONES QUE CONTIENEN FUERZAS PELIGROSAS ART. 157 C.P
## "ATENTADOS A LA SUBSISTENCIA Y DEVASTACION ART. 160 C.P.",
## "CONSPIRACION ART. 471 C.P.",
## "CONSTREÑIMIENTO A APOYO BELICO ART. 150 C.P.",
## "DEPORTACION, EXPULSION, TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART.
## "DESCONOCIMIENTO DE HABEAS CORPUS ART. 177 C.P.",
## "DESNUDEZ FORZADA EN PERSONA PROTEGIDA ART. 139D",
## "DESPLAZAMIENTO FORZADO ART. 180 C.P.",
## "DESPOJO EN EL CAMPO DE BATALLA ART. 151 C.P.",
## "DESTRUCCION DE BIENES E INSTALACIONES DE CARACTER SANITARIO ART. 155 C.P.",
## "DESTRUCCION DEL MEDIO AMBIENTE ART. 164 C.P.",
## "DESTRUCCION O UTILIZACION ILICITA DE BIENES CULTURALES Y DE LUGARES DE CULTO ART.
## "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDOS ART. 154 C.P.",
## "DETENCION ARBITRARIA ESPECIAL ART. 176 C.P.",
## "DETENCION ILEGAL Y PRIVACION DEL DEBIDO PROCESO ART. 149 C.P.",
## "EMBARAZO FORZADO EN PERSONA PROTEGIDA ART. 139C",
## "ESCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
## "EXACCION O CONTRIBUCIONES ARBITRARIAS ART. 163 C.P.",
```

```

## "EXCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
## "GENOCIDIO ART. 101 C.P.",
## "HOMICIDIO EN PERSONA PROTEGIDA ART. 135 C.P.",
## "HOSTILIDAD MILITAR ART. 456 C.P.",
## "LESIONES EN PERSONA PROTEGIDA ART. 136 C.P.",
## "OBSTACULIZACION DE TAREAS SANITARIAS Y HUMANITARIAS ART. 153 C.P.",
## "OMISION DE MEDIDAS DE PROTECCION A LA POBLACION CIVIL ART. 161 C.P.",
## "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIAS ART. 152 C.P.",
## "PERFIDIA ART. 143 C.P.",
## "PRIVACION ILEGAL DE LA LIBERTAD ART. 174 C.P.",
## "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 175 C.P.",
## "PROSTITUCION FORZADA EN PERSONA PROTEGIDA ART. 141",
## "PROSTITUCION FORZADA O ESCLAVITUD SEXUAL ART. 141 C.P.",
## "REBELION ART. 467 C.P.",
## "RECLUTAMIENTO ILICITO ART. 162 C.P.",
## "REPRESALIAS ART. 158 C.P.",
## "SEDICION ART. 468 C.P.",
## "SEDUCCION, USURPACION Y RETENCION ILEGAL DEL MANDO ART. 472 C.P.",
## "TOMA DE REHENES ART. 148 C.P.",
## "TORTURA CONTRA PERSONA PROTEGIDA ART. 137 C.P.",
## "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART
## "UTILIZACION DE MEDIOS Y METODOS DE GUERRA ILICITOS ART. 142 C.P.",
## "UTILIZACION ILEGAL DE UNIFORMES E INSIGNIAS ART. 346 C.P.")
##
## delito_amarillo <- c("ABORTO ART. 122 C.P.",
## "DAÑO EN BIEN AJENO. ART. 265 C.P.",
## "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
## "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
## "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE AÑOS. ART. 208 C.P.",
## "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
## "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
## "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 416",
## "ABUSO DE CONDICIONES DE INFERIORIDAD. ART. 251 C.P.",
## "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 1
## "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 1
## "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
## "ACTOS SEXUALES CON MENOR DE CATORCE AÑOS. ART. 209 C.P.",
## "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
## "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
## "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
## "ALZAMIENTO DE BIENES. ART. 253 C.P.",
## "CONCUSION ART. 404 C.P.",
## "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
## "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
## "ABUSO DE FUNCION PUBLICA ART. 428 C.P.",
## "ACTOS DE DISCRIMINACION RACIAL ART. 147 C.P.",
## "ACTOS DE RACISMO O DISCRIMINACIÓNART. 134 A",
## "CONSTREÑIMIENTO A LA PROSTITUCION ART. 214 C.P.",
## "CONSTREÑIMIENTO AL SUFRAGANTE ART. 387 C.P.",
## "CONSTREÑIMIENTO ILEGAL ART. 182 C.P.",
## "CONTAMINACION AMBIENTAL ART. 332 C.P.",
## "CONTAMINACION AMBIENTAL POR EXPLOTACION DE YACIMIENTO MINERO O HIDROCARBURO AR
## "CORRUPCION DE SUFRAGANTE ART. 390 C.P.",
## "DESTINACION ILEGAL DE COMBUSTIBLE ART. 327D C.P.",

```

```

## "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
## "EXTORSION. ART. 244 C.P.",
## "FABRICACION, IMPORTACION, TRAFICO, POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICAS
## "FABRICACION, TRAFICO Y PORTE DE ARMAS DE FUEGO O MUNICIONES ART. 365 C.P.",
## "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.
## "FAVORECIMIENTO DE LA FUGA ART. 449 C.P.",
## "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART.
## "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P.",
## "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P. INCISO 1",
## "OMISION DE CONTROL ART. 325 C.P.",
## "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
## "SECUESTRO EXTORSIVO ART. 169 C.P.",
## "TERRORISMO ART. 343 C.P.",
## "TESTAFERRATO ART. 326 C.P.",
## "TESTAFERRATO ART. 326 C.P. MENOR 100 SALARIOS",
## "TRAFICO, TRANSPORTE Y POSESION DE MATERIALES RADIOACTIVOS O SUSTANCIAS NUCLEARES
## "USO DE MENORES DE EDAD LA COMISION DE DELITO ART 188D LEY 1453 DE 2011",
## "USURPACION DE TIERRAS. ART. 261 C.P.",
## "UTILIZACION INDEBIDA DE INFORMACION OFICIAL PRIVILEGIADA. ART. 420 C.P.",
## "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.",
## "VIOLACION DE LA LIBERTAD DE TRABAJO. ART. 198 C.P.",
## "FRAUDE DE SUBVENCIONES",
## "HOMICIDIO PRETERINTENCIONAL ART. 105 C.P.",
## "INDUCCION A LA PROSTITUCION ART. 213 C.P.",
## "LAVADO DE ACTIVOS ART. 323 C.P.",
## "USURPACION DE MARCAS Y PATENTES ART. 306 C.P.",
## "DAÑO")
##
## delito_azul <- c("CONSTREÑIMIENTO PARA DELINQUIR ART. 184 C.P.",
## "ACTOS DE BARBARIE ART. 145 C.P.",
## "AMENAZAS A TESTIGOS ART. 454A C.P. AD. LEY 890 DE 2004 ART.13",
## "AMENAZAS ART. 347 C.P.",
## "APODERAMIENTO DE AERONAVES, NAVES O MEDIOS DE TRANSPORTE COLECTIVO. ART. 173 C.P."
## "APODERAMIENTO DE LOS HIDROCARBUROS, SUS DERIVADOS, BIOCOMBUSTIBLES O MEZCLAS QUE L
## "CONCIERTO PARA DELINQUIR ART. 340 C.P.",
## "DAÑO EN LOS RECURSOS NATURALES ART. 331 C.P.",
## "DAÑO EN OBRAS DE UTILIDAD SOCIAL ART. 351 C.P.",
## "DAÑO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUSTIBLES
## "DESAPARICION FORZADA ART. 165 C.P.",
## "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
## "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
## "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 423 C.P.",
## "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 359 C.P.",
## "ENTRENAMIENTO PARA ACTIVIDADES ILICITAS ART. 341 C.P.",
## "ESPIONAJE ART. 463 C.P.",
## "FABRICACION, TRAFICO Y PORTE DE ARMAS DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART.
## "HOSTIGAMIENTO POR MOTIVOS DE RAZA, RELIGIÓN, IDEOLOGÍA, POLÍTICA, U ORIGEN NACIONAL
## "OBSTRUCCION A VIAS PUBLICAS QUE AFECTAN EL ORDEN PUBLICO. ART.353A C.P. LEY 1453 D
## "TORTURA ART. 178 C.P.",
## "UTILIZACION DE ASUNTO SOMETIDO A SECRETO O RESERVA ART. 419 C.P.",
## "VIOLACION DE DATOS PERSONALES ART 269F LEY 1273 DE 2009",
## "VIOLACION DE HABITACION AJENA POR SERVIDOR PUBLICO ART. 190 C.P.",
## "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION. ART. 200 C.P.",
## "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL. ART. 19

```

```
##          "VIOLACION ILICITA DE COMUNICACIONES. ART. 192 C.P.",
##          "VIOLENCIA CONTRA SERVIDOR PUBLICO ART. 429 C.P.")
##
## delito_na <- c("ABORTO ART. 122 C.P.",
##              "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##              "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##              "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##              "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##              "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART. 207 C.P.",
##              "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##              "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##              "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##              "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##              "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##              "CONCUSION ART. 404 C.P.",
##              "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##              "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
##              "DAÑO EN BIEN AJENO. ART. 265 C.P.",
##              "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##              "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##              "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##              "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##              "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##              "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##              "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.",
##              "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART. 6 DECRE",
##              "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##              "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##              "FUGA DE PRESOS ART. 448 C.P.",
##              "HOMICIDIO ART. 103 C.P.",
##              "HURTO CALIFICADO. ART. 240 C.P.",
##              "HURTO. ART. 239 C.P.",
##              "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##              "INCENDIO ART. 350 C.P.",
##              "INSTIGACION A DELINQUIR ART. 348 C.P.",
##              "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
##              "INTERVENCION EN POLITICA ART. 422 C.P.",
##              "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
##              "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
##              "IRRESPETO A CADAVERES. ART. 204 C.P.",
##              "LAVADO DE ACTIVOS ART. 323 C.P.",
##              "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",
##              "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
##              "OMISION DE APOYO ART. 424 C.P.",
##              "OMISION DE SOCORRO ART. 131 C.P.",
##              "PERFIDIA ART. 143 C.P.",
##              "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
##              "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
##              "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
##              "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
##              "RECEPTACION ART. 447 C.P.",
##              "RECEPTACION ART. 327 C C.P.",
##              "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
##              "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
```

```

## "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
## "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
## "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",
## "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
## "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
## "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
## "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.")
##
## # Rules is conflict.
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_spoa <- data_SPOA_CEV %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     tipo_vinculacion == "PERPETRADOR" ~ zero,
##     perpetrador %in% perp_verde ~ one,
##     perpetrador1 %in% perp_verde ~ one,
##     perpetrador2 %in% perp_verde ~ one,
##     perpetrador3 %in% perp_verde ~ one,
##     perpetrador4 %in% perp_verde ~ one,
##     perpetrador %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador1 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador2 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador3 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador4 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador1 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador2 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador3 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador4 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador %in% perp_rojo ~ zero,
##     perpetrador1 %in% perp_rojo ~ zero,
##     perpetrador2 %in% perp_rojo ~ zero,
##     perpetrador3 %in% perp_rojo ~ zero,
##     perpetrador4 %in% perp_rojo ~ zero,
##     perpetrador %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador1 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador2 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador3 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     perpetrador4 %in% perp_is_ca & tipo_vinculacion == "VICTIMA" ~ one,
##     delito %in% delito_na ~ NA_integer_,
##     delito %in% delito_verde ~ one,
##     delito %in% delito_azul & prob < 0.75 ~ one,
##     delito %in% delito_amarillo & prob < 0.5 ~ one,
##     delito %in% delito_naranja & prob < 0.75 ~ zero,
##     delito %in% delito_rojo ~ zero,
##     titulo %in% titulo_verde ~ one,
##     titulo %in% titulo_amarillo & prob < 0.5 ~ one,
##     titulo %in% titulo_naranja & prob < 0.75 ~ zero,
##     titulo %in% titulo_rojo ~ zero,

```



```

##      tipo_vinculacion == "OTRA" ~ NA_integer_,
##      TRUE ~ NA_integer_)) %>%
##    select(-prob)
##
##
## sample <- conflict_spoa %>%
##   filter(is_conflict ==0) %>%
##   select(recordid, is_conflict)
##
## # ----- training data
##
## sample <- sample_frac(sample, 0.2)
##
## write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
## write_parquet(conflict_spoa, args$output)
## #done

```

```
cat(readLines(files$is_ca_sijuf_jep_fase4), sep = '\n')
```

#### 9.4.1.5. FGN - SIJUF - JEP FASE 4

```

## #
## # Authors:      PA
## # Maintainers  VG, PA, PB, JGD
## # Copyright    2022, HRDAG,
## # =====
## # CO-SIVJNR-data/individual/FGN/is-ca/src/is-ca-sijuf_jep_fase4.R
##
## # following the email from Folco Zaffalon (8 abr 2022, 17:16 PDT),
## # and the accompanying file
## # 'SIJUF_casos_perpetrador_ocupacion victima_delitos.docx'
## #
## # Verde = SI CONFLICTO (1)
## # Azul = MUY PROBABLE 0.75~(is_conflict == 1)
## # Amarillo = NO SE SABE 0.5~(is_conflict == 1)
## # Naranja = POCO PROBABLE 0.75~(is_conflict == 0)
## # Rojo = NO CONFLICTO (0)
##
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210) # <-- the seed is really important in this run!
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow,assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--sijuf_jep_fase4",
##                      default=here("individual/FGN/clean/output/sijuf_jep_fase4.parquet"))

```

```
## parser$add_argument("--examples",
##                       default = "individual/FGN/is-ca/hand/example_sijuf_jep.csv")
## parser$add_argument("--output",
##                       default = "output/sijuf_jep_fase4.parquet")
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## data_sijuf_jep_fase4 <- read_parquet(args$sijuf_jep_fase4)
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##               "AGENTES_ESTATALES", "ELN - EJERCITO DE LIBERACION NACIONAL",
##               "EPL - EJERCITO POPULAR DE LIBERACION",
##               "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##               "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##               "GUERRILLA", "OTROS GRUPOS PARAMILITARES", "PARAMILITARES")
##
## perp_no_ca = c("COMUNIDADES INDIGENAS", "DELINCUENCIA COMUN")
##
## perp_verde <- c("AGENTE DEL ESTADO", "AGUILAS NEGRAS", "AUC", "CORDILLERA ELN", "GAO", "GUERRILLA",
##               "INDEPENDIENTES", "LOS RASTROJOS", "M19", "MACHETEROS", "PARAMILITARES",
##               "ELN", "EPL", "ERP", "ERG", "FARC", "GAO")
##
## perp_azul <- c("FUERZA PUBLICA", "AGENTE DEL ESTADO")
##
## perp_naranja <- c("INDEPENDIENTES")
##
## delito_verde <- c("ACCESO CARNAL VIOLENTO EN PERSONA PROTEGIDA ART. 138 C.P.",
##                 "ACTOS DE TERRORISMO ART. 144 C.P.",
##                 "ACTOS SEXUALES VIOLENTOS EN PERSONA PROTEGIDA ART. 139 C.P.",
##                 "ADMINISTRACION DE RECURSOS CON FINES TERRORISTAS ART. 17 DEC180 DE 1988",
##                 "ADMINISTRACION DE RECURSOS RELACIONADOS CON ACTIVIDADES TERRORISTAS ART. 345 C.P.",
##                 "APOLOGIA DEL GENOCIDIO ART. 102 C.P.",
##                 "ASONADA ART. 128 C.P.",
##                 "ASONADA ART. 469 C.P.",
##                 "ATAQUE CONTRA OBRAS E INSTALACIONES QUE CONTIENEN FUERZAS PELIGROSAS ART. 157 C.P.",
##                 "ATENTADOS A LA SUBSISTENCIA Y DEVASTACION ART. 160 C.P.",
##                 "CONSPIRACION ART. 130 C.P.",
##                 "CONSPIRACION ART. 471 C.P.",
##                 "CONSTRENIMIENTO A APOYO BELICO ART. 150 C.P.",
##                 "DEPORTACION, EXPULSION , TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART.",
##                 "DEPORTACION, EXPULSION, TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART.",
##                 "DESPLAZAMIENTO FORZADO ART. 180 C.P.",
##                 "DESPLAZAMIENTO FORZADO LEY 589/2000",
##                 "DESPOJO EN EL CAMPO DE BATALLA ART. 151 C.P.",
##                 "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDAS ART. 154 C.P.",
##                 "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDOS ART. 154 C.P.",
##                 "ENTRENAMIENTO PARA ACTIVIDADES SICARIALES ART. 3 DEC.1194 DE 1989",
##                 "GENOCIDIO ART. 101 C.P.",
##                 "HOMICIDIO EN PERSONA PROTEGIDA ART. 135 C.P.",
##                 "HOMICIDIO EN PERSONAS PROTEGIDAS ART. 135 C.P.",
##                 "INSTIGACION A LA GUERRA ART. 458 C.P.",
```

```
## "INSTIGACION AL TERRORISMO ART. 8 DEC. 180 DE 1988",
## "INSTIGACION O CONSTRENTAMIENTO PARA INGRESO A GRUPOS TERRORISTAS ART. 6 DEC. 180 DE 1988",
## "INSTRUCCION Y ENTRENAMIENTO AL TERRORISMO ART. 15 DEC.180 DE 1988",
## "LESIONES EN PERSONA PROTEGIDA ART. 136 C.P.",
## "LESIONES PERSONALES FINES TERRORISTAS ART. 31 DEC. 180 DE 1988",
## "OMISION DE INFORMES SOBRE ACTIVIDADES TERRORISTAS ART. 4 DEC. 180 DE 1988",
## "OMISION DE MEDIDAS DE PROTECCION A LA POBLACION CIVIL ART. 161 C.P.",
## "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIA ART. 152 C.P.",
## "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIAS ART. 152 C.P.",
## "REBELION ART. 125 C.P. MOD. ART. 1 DEC. 1857 DE 1989",
## "REBELION ART. 467 C.P.",
## "RECLUTAMIENTO Ilicito ART. 162 C.P.",
## "REPRESALIAS ART. 158 C.P.",
## "SEDICION ART. 126 C.P.",
## "SEDICION ART. 468 C.P.",
## "SEDUCCION, USURPACION Y RETENCION ILEGAL DEL MANDO ART. 472 C.P.",
## "TERRORISMO ART. 343 C.P.",
## "TERRORISMO ART.187 C.P.",
## "TOMA DE REHENES ART. 148 C.P.",
## "TORTURA CONTRA PERSONA PROTEGIDA ART. 137 C.P.",
## "TORTURA EN PERSONA PROTEGIDA ART. 137 C.P.",
## "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART. 142 C.P.",
## "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART. 142 C.P.",
## "UTILIZACION DE MEDIOS Y METODOS DE GUERRA Ilicitos ART. 142 C.P.")
##
## delito_azul <- c("ACTOS DE BARBARIE ART. 145 C.P.",
## "ACTOS DE DISCRIMINACION RACIAL ART. 147 C.P.",
## "AMENAZAS ART. 347 C.P.",
## "AMENAZAS PERSONALES O FAMILIARES ART. 26 DEC. 180 DE 1988",
## "DESAPARICION FORZADA ART. 165 C.P.",
## "DESAPARICION FORZADA LEY 589/2000",
## "DESTRUCCION O UTILIZACION Ilicita DE BIENES CULTURALES Y DE LUGARES DE CULTO ART. 142 C.P.",
## "DETENCION ARBITRARIA ESPECIAL ART. 176 C.P.",
## "DETENCION ARBITRARIA ESPECIAL ART. 274 C.P.",
## "DETENCION ILEGAL Y PRIVACION DEL DEBIDO PROCESO ART. 149 C.P.",
## "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
## "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART.195 C.P.",
## "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 159 C.P.",
## "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 423 C.P.",
## "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 198 C.P.",
## "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 359 C.P.",
## "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS CON FINES TERRORISTAS ART. 142 C.P.",
## "EMPLEO, PRODUCCION, COMERCIALIZACION Y ALMACENAMIENTO DE MINAS ANTIPERSONAL ART. 359 C.P.",
## "ENRIQUECIMIENTO Ilicito DERIVADO DE SECUESTRO ART. 6 LEY 40 DE 1993",
## "ENTRENAMIENTO PARA ACTIVIDADES Ilicitas ART. 341 C.P.",
## "EXACCION O CONTRIBUCIONES ARBITRARIAS ART. 163 C.P.",
## "EXTORSION ART. 355 C.P.",
## "EXTORSION. ART. 244 C.P.",
## "FABRICACION Y TRAFICO DE ARMAS DE FUEGO O MUNICIONES ART. 201 C.P.",
## "FABRICACION Y TRAFICO DE ARMAS Y MUNICIONES DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART. 201 C.P.",
## "FABRICACION, IMPORTACION, TRAFICO POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICAS Y NUCLEARES ART. 365 C.P.",
## "FABRICACION, IMPORTACION, TRAFICO, POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICAS Y NUCLEARES ART. 365 C.P.",
## "FABRICACION, TRAFICO Y PORTE DE ARMAS DE FUEGO O MUNICIONES ART. 365 C.P.",
## "FABRICACION, TRAFICO Y PORTE DE ARMAS DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART. 201 C.P.")
```

```
## "INCITACION A LA INSUBORDINACION POR MEDIOS PUBLICOS ART. 28 LEY 29 DE 1944",
## "OMISION DE INFORMES EN SECUESTRO ART. 9 LEY 40 DE 1993",
## "OMISION DE SOCORRO ART. 131 C.P.",
## "PERTENENCIA A BANDAS DE SICARIOS ART. 2 DEC-LEY 1194789 ACLP ART. 6 DEC 2266791",
## "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P.",
## "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
## "PROSTITUCION FORZADA O ESCLAVITUD SEXUAL ART. 141 C.P.",
## "SECUESTRO EXTORSIVO AGRAVADO ART. 170 C.P.",
## "SECUESTRO EXTORSIVO ART. 169 C.P.",
## "SECUESTRO EXTORSIVO ART. 268 C.P.",
## "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
## "SINIESTRO O DANO DE NAVE ART. 193 C.P.",
## "SINIESTRO O DANO DE NAVE ART. 354 C.P.",
## "SUPLANTACION DE AUTORIDAD ART. 20 DEC. 180 DE 1986",
## "TESTAFERRATO ART. 326 C.P.",
## "TESTAFERRATO ART.6 DEC. LEY 1856 DE 1989",
## "TORTURA ART. 178 C.P.",
## "TORTURA ART. 279 C.P.",
## "UTILIZACION ILEGAL DE UNIFORMES E INSIGNIAS ART. 346 C.P.",
## "UTILIZACION ILEGAL DE UNIFORMES O INSIGNIAS ART. 19 DEC. 180 DE 1988",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTOR ART. 16 DEC. 180 DE 1988",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES ART. 197 C.P.",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
## "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION 292 C.P.",
## "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION. ART. 200 C.P.")
##
## delito_naranja <- c("ACCESO ABUSIVO A UN SISTEMA INFORMATICO ART. 195 C.P.",
## "ACCESO ABUSIVO A UN SISTEMA INFORMATICO. ART. 195 C.P.",
## "ACCESO CARNAL MEDIANTE ENGAÑO ART. 301 C.P.",
## "ASOCIACION PARA LA COMISION DE UN DELITO CONTRA LA ADMINISTRACION PUBLICA ART. 4",
## "CONTRABANDO ART. 15 LEY 383/97",
## "CONTRABANDO ART. 319 C.P.",
## "CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS DTO 1900/02",
## "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 146 C.P.",
## "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 410 C.P.",
## "DE LA EXTINCION DE DOMINIO",
## "DESTINACION DE MUEBLES E INMUEBLES ART. 34 LEY 30 DE 1986",
## "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
## "HOMICIDIO ART. 103 C.P.",
## "HURTO ART. 349 C.P.",
## "HURTO ART. 349 C.P.MOD. ART. 1 N.11 LEY 23 DE 1991",
## "HURTO. ART. 239 C.P.",
## "PISTAS DE ATERRIZAJE ILEGALES (RESPONSABILIDAD DEL DUENO POSEEDOR O ARRENDATARIO)",
## "REINGRESO ILEGAL AL PAIS ART. 185 C.P.",
## "REVELACION DE SECRETO ART. 154 C.P.",
## "REVELACION DE SECRETO ART. 418 C.P.",
## "SECUESTRO SIMPLE ART. 168 C.P.",
## "SECUESTRO SIMPLE ART. 269 C.P.",
## "SEDUCCION USURPACION Y RETENCION ILEGAL DEL MANDO ART. 131 C.P.",
## "SOBORNO ART. 174 C.P.",
## "SOBORNO ART. 174 C.P.",
## "SOBORNO ART. 444 C.P.",
## "SOBORNO EN LA ACTUACION PENAL ART 444A C.P.",
## "TRAFICO DE INFLUENCIAS DE SERVIDOR PUBLICO ART. 411 C.P.",
```

```
## "TRAFICO DE MONEDA FALSIFICADA ART. 208 C.P.",
## "TRAFICO DE MONEDA FALSIFICADA. ART. 274 C.P.",
## "TRATA DE PERSONAS ART. 311 C.P",
## "TRATA DE PERSONAS. ART. 215 C.P.",
## "USURPACION DE AGUAS. ART. 262 C.P.",
## "USURPACION DE FUNCIONES PUBLICAS ART. 161 C.P.",
## "VIOLACION DE HABITACION AJENA ART. 284 C.P.",
## "VIOLACION DE HABITACION AJENA. ART. 189 C.P.",
## "VIOLACION DE INMUNIDAD DIPLOMATICA ART. 465 C.P.")
##
## delito_rojo <- c("ABANDONO ART. 127 C.P.",
## "ABANDONO ART. 346 C.P.",
## "ABANDONO DEL CARGO ART. 156 C.P.",
## "ABUSO DE CONFIANZA ART. 358 C.P.",
## "CONFIANZA ART. 358 C.P.MOD. ART. 1 N.16 LEY 23 DE 1991",
## "ABUSO DE CONFIANZA CALIFICADO ART. 250 C.P.",
## "ABUSO DE CONFIANZA. ART. 249 C.P.",
## "ACCESO O PRESTACION ILEGAL DE LOS SERVICIOS DE TELECOMUNICACIONES ART.6 LEY 422/98",
## "ACTO SEXUAL MEDIANTE ENGAÑO ART. 302 C.P.",
## "ADOPCION ILEGAL DE MENORES ART. 267 DEC. 2737 DE 1989",
## "AGIOTAJE. ART. 301",
## "ALTERACION Y MODIFICACION DE CALIDAD CANTIDAD PESO O MEDIDA ART. 231 C.P.",
## "ALTERACION Y MODIFICACION DE CALIDAD, CANTIDAD, PESO O MEDIDA. ART. 299 C.P.",
## "APLICACION FRAUDULENTA DE CREDITO OFICIALMENTE REGULADO ART. 241 C.P.",
## "APLICACION FRAUDULENTA DE CREDITO OFICIALMENTE REGULADO ART. 311 C.P.",
## "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO ART. 361 C.P.",
## "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO. ART. 252 C.P.",
## "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P.",
## "BIGAMIA ART. 260 C.P.",
## "CALUMNIA ART. 314 C.P.",
## "CALUMNIA INDIRECTA ART. 315 C.P.",
## "CALUMNIA. ART. 221 C.P.",
## "CAPTACION MASIVA Y HABITUAL ART. 208 N.3 DEC. 663 DE 1993",
## "CAZA ILEGAL ART. 336",
## "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 12 LEY 40 DE 1993",
## "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 172 C.P.",
## "CELEBRACION INDEBIDA DE CONTRATOS DE SEGUROS ART. 172 C.P.",
## "CIRCULACION Y USO DE EFECTO OFICIAL O SELLO FALSIFICADO. ART. 281 C.P.",
## "COHECHO IMPROPIO ART. 142 C.P.",
## "COHECHO IMPROPIO ART. 406 C.P.",
## "COHECHO POR DAR U OFRECER ART. 143 C.",
## "COHECHO POR DAR U OFRECER ART. 407 C.P.",
## "COHECHO PROPIO ART. 141 C.P.",
## "COHECHO POR DAR U OFRECER ART. 407 C.P.",
## "COHECHO PROPIO ART. 141 C.P.",
## "COHECHO PROPIO ART. 405 C.P.",
## "CONSTANCIAS Y ESTADOS FINANCIEROS FALSOS ART. 43 LEY 222/95",
## "CORRUPCION DE ALIMENTOS Y MEDICINAS ART. 206 C.P.",
## "CORRUPCION DE ALIMENTOS, PRODUCTOS MEDICOS O MATERIAL PROFILACTICO ART. 372 C.P.",
## "DEFRAUDACION A LAS RENTAS DE ADUANA ART.17 LEY 383/97",
## "DEFRAUDACION A LOS DERECHOS PATRIMONIALES DE AUTOR. ART. 271 C.P.",
## "DEFRAUDACION DE FLUIDOS ART. 256 C.P.",
## "DEFRAUDACION DE FLUIDOS. ART. 256 C.P.",
## "DEL ACCESO ILEGAL O PRESTACION ILEGAL DE LOS SERVICIOS DE TELECOMUNICACIONES. ART.
```

## "DEL URBANIZADOR ILEGAL ART. 367A CP.",  
## "DESTRUCCION SUPRESION OCULTA DOCUMENTO PRIVADO ART. 224 C.P.",  
## "DESTRUCCION SUPRESION OCULTA DOCUMENTO PUBLICO ART. 223 C.P.",  
## "DESTRUCCION, SUPRESION U OCULTAMIENTO DE DOCUMENTO PUBLICO. ART. 292 C.P.",  
## "DESTRUCCION, SUPRESION Y OCULTAMIENTO DE DOCUMENTO PRIVADO. ART. 293 C.P.",  
## "DISPOSICION DE BIEN PROPIO GRAVADO CON PRENDA ART. 364 C.P.",  
## "DISPOSICION DE BIEN PROPIO GRAVADO CON PRENDA. ART. 255 C.P.",  
## "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS ART. 289 C.P.",  
## "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",  
## "EJECUCION O REPRODUCCION DE OBRAS ARTISTICAS SIN AUTORIZACION DEL TITULAR DEL DERECHO  
## "EJERCICIO ARBITRARIO DE LA CUSTODIA DE HIJO MENOR DE EDAD ART 230A C.P.",  
## "EJERCICIO ILICITO DE ACTIVIDAD MONOPOLISTICA DE ARBITRIO RENTISTICO ART. 312 C.P.",  
## "EJERCICIO ILICITO DE ACTIVIDADES MONOPOLISTICAS DE ARBITRIO RENTISTICO ART. 241A C.P.",  
## "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUE. ART. 248 C.P.",  
## "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUES ART. 357 C.P.",  
## "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUES ART. 357 C.P. MOD. ART. 1 N.15 LEY 23 DE 1991",  
## "EMISIONES ILEGALES. ART. 276 C.P.",  
## "ENRIQUECIMIENTO ILICITO ART. 412 C.P.",  
## "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 1 DEC. 1895 DE 1989",  
## "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P.",  
## "ENRIQUECIMIENTO ILICITO DE SERVIDOR PUBLICO ART. 148 C.P.",  
## "ESTAFA ART. 356 C.P. MOD. ART. 4 N.1 LEY 23 DE 1991",  
## "ESTAFA. ART. 246 C.P.",  
## "ESTIMULO A LA PROSTITUCION DE MENORES ART. 312 C.P.",  
## "ESTIMULO A LA PROSTITUCION DE MENORES. ART. 217 C.P.",  
## "ESTIMULO AL USO ILICITO ART. 378 C.P.",  
## "EXISTENCIA, CONSTRUCCION Y UTILIZACION ILEGAL DE PISTAS DE ATERRIZAJE ART. 385 C.P.",  
## "EXPERIMENTACION ILEGAL EN ESPECIES ANIMALES O VEGETALES ART. 334 C.P.",  
## "EXPLORACION O EXPLORACION ILICITA MINERA O PETROLERA ART. 244 C.P. MOD. ART. 21 LEY 23 DE 1991",  
## "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",  
## "FALSA AUTOACUSACION ART. 168 C.P.",  
## "FALSA AUTOACUSACION ART. 437 C.P.",  
## "FALSA DENUNCIA ART. 166 C.P.",  
## "FALSA DENUNCIA ART. 435 C.P.",  
## "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 167 C.P.",  
## "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 436 C.P.",  
## "FALSEDAD DERECHOS DE AUTOR ART. 51 LEY 44 DE 1993",  
## "FALSEDAD EN DOCUMENTO PRIVADO ART. 221 C.P.",  
## "FALSEDAD EN DOCUMENTO PRIVADO. ART. 289 C.P.",  
## "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO ART. 219 C.P.",  
## "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO. ART. 286 C.P.",  
## "FALSEDAD MARCARIA ART. 217 C.P.",  
## "FALSEDAD MARCARIA. ART. 285 C.P.",  
## "FALSEDAD MATERIAL EMPLEADO OFICIAL DOCUMENTO PUBLICO ART.218 CP",  
## "FALSEDAD MATERIAL EN DOCUMENTO PUBLICO. ART. 287 C.P.",  
## "FALSEDAD MATERIAL PARTICULAR DOCUMENTO PUBLICO ART. 220 C.P.",  
## "FALSEDAD PERSONAL ART. 227 C.P.",  
## "FALSEDAD PERSONAL PARA OBTENCION DOCUMENTO PUBLICO ART. 226 CP",  
## "FALSEDAD PERSONAL. ART. 296 C.P.",  
## "FALSIFICACION DE EFECTO OFICIAL TIMBRADO. ART. 280 C.P.",  
## "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA ART. 207 C.P.",  
## "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA. ART. 273 C.P.",  
## "FALSO TESTIMONIO ART. 172 C.P.",  
## "FALSO TESTIMONIO ART. 442 C.P.",

## "FAVORECIMIENTO ART. 176 C.P. MOD. ART. 6 LEY 356/97",  
## "FAVORECIMIENTO ART. 446 C.P.",  
## "FAVORECIMIENTO CONTRABANDO DE HIDROCARBUROS - DERIVADOS DTO 1900/02",  
## "FAVORECIMIENTO DE CONTRABANDO ART. 320 C.P.",  
## "FAVORECIMIENTO DE CONTRABANDO ART.16 LEY 383/97",  
## "FECUNDACION Y TRAFICO DE EMBRIONES ART. 134 C.P.",  
## "FRAUDE A RESOLUCION JUDICIAL ART. 454 C.P.",  
## "FRAUDE A RESOLUCION JUDICIAL ART.184 C.P.",  
## "FRAUDE PROCESAL ART. 182 C.P.",  
## "FRAUDE PROCESAL ART. 453 C.P.",  
## "INDEBIDA DE LOS RECURSOS SOCIALES. ART. 260 C.P.",  
## "GESTION INDEBIDA DE RECURSOS SOCIALES ART. 260 C.P.",  
## "HOMICIDIO CULPOSO ART. 109 C.P.",  
## "HOMICIDIO CULPOSO ART. 329 C.P.",  
## "HOMICIDIO POR PIEDAD ART. 106 C.P.",  
## "HOMICIDIO POR PIEDAD ART.326 C.P.",  
## "HOMICIDIO PRETERINTENCIONAL ART. 105 C.P.",  
## "IMITACION O SIMULACION DE ALIMENTOS, PRODUCTOS O SUSTANCIAS ART. 373 C.P.",  
## "INASISTENCIA ALIMENTARIA ART. 233 C.P.",  
## "INASISTENCIA ALIMENTARIA ART. 263 C.P.",  
## "INCESTO ART. 259 C.P.",  
## "INCESTO. ART. 237 C.P.",  
## "INDUCCION A AYUDA AL SUICIDIO ART. 107 C.P.",  
## "INDUCCION A LA PROSTITUCION ART. 213 C.P.",  
## "INDUCCION A LA PROSTITUCION ART. 308 C.P.",  
## "INDUCCION O AYUDA AL SUICIDIO ART. 327 C.P.",  
## "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 175 C.P.",  
## "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 445 C.P.",  
## "INJURIA ART. 313 C.P.",  
## "INJURIA POR VIAS DE HECHO ART. 319 C.P.",  
## "INJURIA. ART. 220 C.P.",  
## "INSEMINACION ARTIFICIAL O TRANSFERENCIA DE OVULO FECUNDADO NO CONSENTIDAS. ART. 18",  
## "INTERES ILICITO EN LA CELEBRACION DE CONTRATOS ART. 145 C.P.",  
## "INTERES INDEBIDO EN LA CELEBRACION DE CONTRATOS ART. 409 C.P.",  
## "LESIONES AL FETO ART. 125 C.P.",  
## "LESIONES ART. 111 C.P.",  
## "LESIONES CULPOSAS AL FETO ART. 126 C.P.",  
## "LESIONES CULPOSAS ART. 120 C.P.",  
## "LESIONES PERSONALES CULPOSAS ART. 340 C.P.",  
## "LESIONES PERSONALES CULPOSAS ART. 340 C.P.MOD. ART. 12 LEY 228 DE 1995",  
## "MALVERSACION Y DILAPIDACION DE BIENES FAMILIARES. ART. 236 C.P.",  
## "MATRIMONIO ILEGAL ART. 261 C.P.",  
## "MENDICIDAD Y TRAFICO DE MENORES. ART. 231 C.P.",  
## "MODALIDAD CULPOSA DEL FAVORECIMIENTO DE LA FUGA ART. 180 C.P.",  
## "MODALIDAD CULPOSA DEL FAVORECIMIENTO DE LA FUGA ART. 450 C.P.",  
## "MUERTE DE HIJO FRUTO DE ACCESO CARNAL VIOLENTO, ABUSIVO O DE INSEMINACION ARTIFICI",  
## "OBTENCION DE DOCUMENTO PUBLICO FALSO ART. 288 C.P.",  
## "OCULTAMIENTO RETENCION Y POSESION ILICITA DE CEDULA ART. 257 C.P.",  
## "OCULTAMIENTO, ALTERACION O DESTRUCCION DE ELEMENTO MATERIAL PROBATORIO ART. 454B",  
## "OCULTAMIENTO, RETENCION Y POSESION ILICITA DE CEDULA ART. 395 C.P.",  
## "OMISION DE DENUNCIA DE PARTICULAR",  
## "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P.",  
## "OMISION DEL AGENTE RETENEDOR O RECAUDADOR ART. 402 C.P.",  
## "PANICO ECONOMICO ART. 302 C.P.",

```

##          "PARTO O ABORTO PRETERINTENCIONAL ART. 118 C.P.",
##          "PECULADO CULPOSO ART. 137 C.P.",
##          "PECULADO CULPOSO ART. 400 C.P.",
##          "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 136 C.P.",
##          "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 399 C.P.",
##          "PECULADO POR APROPIACION ART. 133 C.P.",
##          "PECULADO POR APROPIACION ART. 397 C.P.",
##          "PECULADO POR EXTENSION ART. 138 C.P.",
##          "PECULADO POR USO ART. 134 C.P.",
##          "PECULADO POR USO ART. 398 C.P.",
##          "PERFIDIA ART. 143 C.P.",
##          "PORNOGRAFIA CON MENORES ART. 218 C.P.",
##          "PREVARICATO POR ACCION ART. 149 C.P.",
##          "PREVARICATO POR ACCION ART. 413 C.P.",
##          "PREVARICATO POR ASESORAMIENTO ILEGAL ART. 151 C.P.",
##          "PREVARICATO POR OMISION ART. 150 C.P.",
##          "PREVARICATO POR OMISION ART. 414 C.P.",
##          "PROPAGACION DEL VIRUS DE INMUNODEFICIENCIA HUMANA O DE LA HEPATITIS B ART. 370 C.P.",
##          "REPETIBILIDAD DEL SER HUMANO ART. 13 C.P.",
##          "REPETIBILIDAD DEL SER HUMANO ART. 133",
##          "ALTERACION O SUPOSICION DEL ESTADO CIVIL. ART. 238 C.P.",
##          "SUSTRACCION DE BIEN PROPIO. ART. 254 C.P.",
##          "TRAFICO INFLUENCIAS OBTENER FAVOR SERVIDOR PUBLICO ART. 147 C.P.",
##          "TURISMO SEXUAL. ART. 219 C.P.",
##          "URBANIZACION ILEGAL ART. 318 C.P.",
##          "USURA ART. 235 C.P.",
##          "USURA ART. 305 C.P.",
##          "USURPACION DE MARCAS Y PATENTES ART. 236 C.P.",
##          "USURPACION DE MARCAS Y PATENTES ART. 306 C.P.",
##          "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.",
##          "VIOLACION A LOS DERECHOS MORALES DE AUTOR ART. 270 C.P.",
##          "VIOLACION A LOS DERECHOS MORALES DE AUTOR. ART. 270 C.P.",
##          "VIOLACION A LOS MECANISMOS DE PROTECCION DE LOS DERECHOS PATRIMONIALES DE AUTOR Y O",
##          "VIOLACION A LOS MECANISMOS DE PROTECCION DE LOS DERECHOS PATRIMONIALES DE AUTOR Y O",
##          "VIOLACION AL REGIMEN LEGAL O CONSTITUCIONAL DE INHABILIDADES E INCOMPATIBILIDADES",
##          "VIOLACION DE FRONTERAS PARA EXPLOTACION DE RECURSOS NATURALES ART. 123 C.P.",
##          "VIOLACION DE MEDIDAS SANITARIAS ART. 203",
##          "VIOLACION REGIMEN LEGAL INHABILIDADES INCOMPATIBILIDADES ART.144 CP.",
##          "VIOLENCIA INTRAFAMILIAR ART. 22 LEY 294 DE 1996",
##          "VIOLENCIA INTRAFAMILIAR ART. 229 C.P.",
##          "VIOLENCIA SEXUAL ENTRE CONYUGES ART. 25 LEY 294 DE 1996")
##
## delito_amarillo <- c("ABORTO ART. 122 C.P.",
##                    "ABORTO ART. 343 C.P.",
##                    "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##                    "ABORTO SIN CONSENTIMIENTO ART. 344 C.P.",
##                    "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 152 C.P.",
##                    "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 416",
##                    "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 153 C.P.",
##                    "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##                    "ABUSO DE CIRCUNSTANCIAS DE INFERIORIDAD ART. 360 C.P.",
##                    "ABUSO DE CONDICIONES DE INFERIORIDAD. ART. 251 C.P.",
##                    "ABUSO DE FUNCION PUBLICA ART. 162 C.P.",
##                    "ABUSO DE FUNCION PUBLICA ART. 428 C.P.",

```



## "CARNAL ABUSIVO CON INCAPAZ DE RESISTIR ART. 304 C.P/97",  
## "ACCESO CARNAL ABUSIVO CON MENOR ART. 303 C.P",  
## "CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",  
## "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",  
## "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART.  
## "ACCESO CARNAL VIOLENTO ART. 298 C.P.",  
## "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",  
## "ACOSO SEXUAL ART. 210 A",  
## "ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 300 C.P. 97",  
## "ACTO SEXUAL VIOLENTO ART. 299 C.P.",  
## "ACTO SEXUAL VIOLENTO. ART. 206 C.P.",  
## "ACTOS CONTRARIOS A LA DEFENSA DE LA NACION ART. 460 C.P.",  
## "ACTOS SEXUALES CON MENOR DE 14 ANOS ART. 305 C.P",  
## "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",  
## "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",  
## "ALTERACION DE RESULTADOS ELECTORALES ART.256 C.P.",  
## "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO ART.354 C.P.",  
## "ALTERACION, DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",  
## "ALZAMIENTO DE BIENES ART. 362 C.P.",  
## "ALZAMIENTO DE BIENES. ART. 253 C.P.",  
## "AMENAZAS A TESTIGO ART. 454A C.P.",  
## "APODERAMIENTO DE AERONAVES, NAVES O MEDIOS DE TRANSPORTE COLECTIVO. ART. 173 C  
## "APODERAMIENTO DE HIDROCARBUROS, SUS DERIVADOS, BIOCOMBUSTIBLES O MEZCLAS QUE L  
## "APODERAMIENTO Y DESVIO DE AERONAVE ART. 28 DEC. 180 DE 1988",  
## "CONCIERTO PARA DELINQUIR ART. 186 C.P",  
## "CONCIERTO PARA DELINQUIR ART. 340 C.P.",  
## "CONCIERTO PARA DELINQUIR LEY 589/2000",  
## "CONCUSION ART. 404 C.P.",  
## "CONCUSION ART.140 C.P.",  
## "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",  
## "CONSTRENIAMIENTO A LA PROSTITUCION ART. 214 C.P.",  
## "CONSTRENIAMIENTO A LA PROSTITUCION ART. 309 C.P.",  
## "CONSTRENIAMIENTO AL ELECTOR Y VIOLENCIA Y FRAUDE ELECTORALES ART. 249 Y ART. 25  
## "CONSTRENIAMIENTO AL SUFRAGANTE ART. 387 C.P.",  
## "CONSTRENIAMIENTO ILEGAL ART. 182 C.P.",  
## "CONSTRENIAMIENTO ILEGAL ART. 276 C.P.",  
## "CONSTRENIAMIENTO PARA DELINQUIR ART. 184 C.P.",  
## "CONSTRENIAMIENTO PARA DELINQUIR ART. 277 C.P.",  
## "CONTAMINACION AMBIENTAL ART. 247 C.P. MOD. ART. 24 LEY 491/99",  
## "CONTAMINACION AMBIENTAL ART. 332 C.P.",  
## "CONTAMINACION AMBIENTAL CULPOSA POR EXPLOTACION DE YACIMIENTO MINERO O HIDROCAI  
## "CONTAMINACION AMBIENTAL CULPOSA POR EXPORTACION DE YACIMIENTO MINERO O HIDROCAI  
## "CONTAMINACION DE AGUAS ART. 205 C.P.",  
## "CONTAMINACION DE AGUAS ART. 371 C.P.",  
## "CORRUPCION DE ELECTOR ART. 251 C.P.",  
## "CORRUPCION DE SUFRAGANTE ART. 390 C.P.",  
## "CULTIVO CONSERVACION FINANCIACION PLANTACIONES ART. 32 LEY 30/86",  
## "DANO EN BIEN AJENO ART. 370 C.P.",  
## "DANO EN BIEN AJENO ART. 370 C.P.MOD. ART. 1 N.19 LEY 23 DE 1991",  
## "DANO EN BIEN AJENO. ART. 265 C.P.",  
## "DANO EN LOS RECURSOS NATURALES ART. 246 C.P.",  
## "DANO EN LOS RECURSOS NATURALES ART. 331 C.P.",  
## "DANO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",  
## "DANO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUS"

## "DANO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUSTIBLES ART. 296 C.P. MOD. ART. 296 C.P.",  
## "DANO O AGRAVIOS A PERSONAS O COSAS DESTINADAS AL CULTO ART. 296 C.P. MOD. ART. 296 C.P.",  
## "DESTRUCCION DE BIENES E INSTALACIONES DE CARACTER SANITARIO ART. 155 C.P.",  
## "DESTRUCCION DEL MEDIO AMBIENTE ART. 164 C.P.",  
## "EXTRADICION",  
## "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",  
## "FAVORECIMIENTO DE LA FUGA ART. 179 C.P.",  
## "FAVORECIMIENTO DE LA FUGA ART. 449 C.P.",  
## "FAVORECIMIENTO DEL VOTO FRAUDULENTO ART. 253 C.P.",  
## "FRAUDE AL SUFRAGANTE ART. 388 C.P.",  
## "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",  
## "FUGA DE PRESOS ART. 178 C.P.",  
## "FUGA DE PRESOS ART. 448 C.P.",  
## "HOMICIDIO ART. 323 C.P.",  
## "HURTO AGRAVADO ART. 11 LEY 228 DE 1995",  
## "HURTO CALIFICADO ART. 350 C.P.",  
## "HURTO CALIFICADO. ART. 240 C.P.",  
## "HURTO DE HIDROCARBUROS O SUS DERIVADOS DTO 1900/02",  
## "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",  
## "ILICITO APROVECHAMIENTO DE RECURSOS BIOLÓGICOS ART. 242 C.P. MOD. ART. 19 LEY 4 DE 1993",  
## "IMPEDIMENTO Y PERTURBACION DE CEREMONIA RELIGIOSA. ART. 202 C.P.",  
## "INCENDIO ART. 189 C.P.",  
## "INCENDIO ART. 350 C.P.",  
## "INSTIGACION A DELINQUIR ART. 188 C.P.",  
## "INSTIGACION A DELINQUIR ART. 348 C.P.",  
## "INSTIGACION A DELINQUIR LEY 589/2000",  
## "INTERVENCION EN POLITICA ART. 158 C.P.",  
## "INTERVENCION EN POLITICA ART. 422 C.P.",  
## "INVASION DE AEREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 243 C.P. MOD. ART. 243 C.P.",  
## "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",  
## "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",  
## "INVASION DE TIERRAS O EDIFICIOS ART. 367 C.P.",  
## "IRRESPECTO A CADAVERES ART. 297",  
## "IRRESPECTO A CADAVERES. ART. 204 C.P.",  
## "LAVADO DE ACTIVOS ART. 247A CP.",  
## "LAVADO DE ACTIVOS ART. 323 C.P.",  
## "LESIONES PERSONALES ART. 331 C.P.",  
## "MALTRATO CONSTITUTIVO DE LESIONES PERSONALES ART. 23 LEY 294 DE 1996",  
## "MALTRATO MEDIANTE RESTRICCION A LA LIBERTAD FISICA. ART. 230 C.P.",  
## "MALTRATO MEDIANTE RESTRICCION A LA LIBERTAD FISICA ART. 230 C.P.",  
## "MALTRATO MEDIANTE RESTRICCION DE LA LIBERTAD FISICA ART. 24 LEY 294 DE 1996",  
## "OBSTACULIZACION DE TAREAS SANITARIAS Y HUMANITARIAS ART. 153 C.P.",  
## "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",  
## "OMISION DE ACTO PROPIO EN DELITOS DE EXTORSION O SECUESTRO ART. 33 LEY 40 DE 1990",  
## "OMISION DE APOYO ART. 160 C.P.",  
## "OMISION DE APOYO ART. 424 C.P.",  
## "OMISION DE CONTROL ART. 247B CP.",  
## "PERTURBACION DE ACTOS OFICIALES ART. 165 C.P.",  
## "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",  
## "PERTURBACION DE LA POSESION SOBRE INMUEBLES ART. 368 C.P.",  
## "PERTURBACION DE LOS SERVICIOS DE COMUNICACION, ENERGIA Y DE COMBUSTIBLES ART. 296 C.P. MOD. ART. 296 C.P.",  
## "PERTURBACION ELECTORAL 248 C.P.",  
## "PERTURBACION EN EL SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 192 C.P.",  
## "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",

```

## "POSESION DE SUSTANCIAS PARA EL PROCESAMIENTO DE NARCOTICOS ART. 43 LEY 30 DE 1990",
## "PRIVACION ILEGAL DE LA LIBERTAD ART. 174 C.P.",
## "PRIVACION ILEGAL DE LA LIBERTAD ART. 272 C.P.",
## "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 175 C.P.",
## "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 273 C.P.",
## "RECEPTACION ART. 177.- C.P.",
## "RECEPTACION ART. 447 C.P.",
## "SIMULACION DE INVESTIDURA O CARGO ART. 163 C.P.",
## "TENENCIA FABRICACION Y TRAFICO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 197 C.P.",
## "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
## "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 33 LEY 30/86",
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C.P.",
## "TRAFICO, TRANSPORTE Y POSESION DE MATERIALES RADIOACTIVOS O SUSTANCIAS NUCLEARES ART. 382 C.P.",
## "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
## "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
## "USO DE DOCUMENTO PUBLICO FALSO ART. 222 C.P.",
## "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
## "USURPACION DE TIERRAS ART. 365 C.P.",
## "USURPACION DE TIERRAS. ART. 261 C.P.",
## "VIOLACION DE HABITACION AJENA POR SERVIDOR PUBLICO ART. 190 C.P.",
## "VIOLACION DE LA LIBERTAD DE TRABAJO ART. 290 C.P.",
## "VIOLACION DE LA LIBERTAD DE TRABAJO. ART. 198 C.P.",
## "VIOLACION ILICITA DE COMUNICACIONES ART. 288 C.P.",
## "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL ART. 192 C.P.",
## "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL. ART. 192 C.P.",
## "VIOLACION ILICITA DE COMUNICACIONES. ART. 192 C.P.",
## "VIOLENCIA CONTRA EMPLEADO OFICIAL ART. 164 C.P.",
## "VIOLENCIA CONTRA SERVIDOR PUBLICO ART. 429 C.P.")
##
##
## delito_na <- c("ABORTO ART. 122 C.P.",
## "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
## "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
## "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
## "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
## "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART. 207 C.P.",
## "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
## "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
## "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
## "ALTERACION, DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
## "ALZAMIENTO DE BIENES. ART. 253 C.P.",
## "CONCUSION ART. 404 C.P.",
## "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
## "CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS DTO 1900/02",
## "DANO EN BIEN AJENO. ART. 265 C.P.",
## "DANO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
## "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
## "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
## "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART.195 C.P.",
## "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
## "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS ART. 289 C.P.",
## "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
## "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",

```

## "FAVORECIMIENTO CONTRABANDO DE HIDROCARBUROS - DERIVADOS DTO 1900/02",  
## "FRAUDE AL SUFRAGANTE ART. 388 C.P.",  
## "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",  
## "FUGA DE PRESOS ART. 178 C.P.",  
## "FUGA DE PRESOS ART. 448 C.P.",  
## "HOMICIDIO ART. 103 C.P.",  
## "HURTO CALIFICADO. ART. 240 C.P.",  
## "HURTO CALIFICADO ART. 350 C.P.",  
## "HURTO. ART. 239 C.P.",  
## "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",  
## "INCENDIO ART. 350 C.P.",  
## "INCENDIO ART. 189 C.P.",  
## "INSTIGACION A DELINQUIR ART. 348 C.P.",  
## "INSTIGACION A DELINQUIR ART. 188 C.P.",  
## "INTERVENCION EN POLITICA ART. 158 C.P.",  
## "INTERVENCION EN POLITICA ART. 422 C.P.",  
## "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",  
## "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",  
## "INVASION DE TIERRAS O EDIFICIOS ART. 367 C.P.",  
## "IRRESPETO A CADAVERES ART. 297",  
## "IRRESPETO A CADAVERES. ART. 204 C.P.",  
## "LAVADO DE ACTIVOS ART. 323 C.P.",  
## "LAVADO DE ACTIVOS ART. 247A CP",  
## "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",  
## "MALTRATO MEDIANTE RESTRICCION A LA LIBERTAD FISICA ART. 230 C.P.",  
## "MALTRATO MEDIANTE RESTRICCION DE LA LIBERTAD FISICA ART. 24 LEY 294 DE 1996",  
## "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",  
## "OMISION DE APOYO ART. 424 C.P.",  
## "OMISION DE APOYO ART. 160 C.P.",  
## "OMISION DE SOCORRO ART. 131 C.P.",  
## "PERFIDIA ART. 143 C.P.",  
## "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P.",  
## "PERTURBACION DE ACTOS OFICIALES ART. 165 C.P.",  
## "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",  
## "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",  
## "PERTURBACION DE LA POSESION SOBRE INMUEBLES ART. 368 C.P.",  
## "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",  
## "PERTURBACION EN EL SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 192 C.P.",  
## "RECEPTACION ART. 447 C.P.",  
## "RECEPTACION ART. 177.- C.P.",  
## "REVELACION DE SECRETO ART. 418 C.P.",  
## "REVELACION DE SECRETO ART. 154 C.P.",  
## "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",  
## "SIMULACION DE INVESTIDURA O CARGO ART. 163 C.P.",  
## "SINIESTRO O DANO DE NAVE ART. 354 C.P.",  
## "SINIESTRO O DANO DE NAVE ART. 193 C.P.",  
## "TENENCIA FABRICACION Y TRAFICO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 197 C.P.",  
## "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",  
## "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",  
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",  
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 33 LEY 30/86",  
## "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",  
## "USO DE DOCUMENTO FALSO. ART. 291 C.P.",  
## "USO DE DOCUMENTO PUBLICO FALSO ART. 222 C.P.",

```

##          "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
##          "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES ART. 197 C.P.",
##          "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTOR ART. 16 DEC. 180 DE 1988",
##          "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
##          "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.")
##
## prof_verde <- c("ACTIVIDADES RELACIONADAS CON EL SINDICALISMO",
##               "ACTIVIDADES RELACIONADAS ORG. CIVICAS Y CAMPESINAS", "ACTIVISTA DERECHOS HUMANOS",
##               "ACTIVISTA SINDICAL", "ALCALDE", "ALMIRANTE (FUERZA PUBLICA)",
##               "DESMOVLIZADO", "DESPLAZADO",
##               "DIRIGENTE ORGANIZACION DERECHOS HUMANOS", "DIRIGENTE SINDICAL", "INFANTE DE MARINA"
##               "INTEGRANTE AUTODEFENSAS", "INTEGRANTE GUERRILLA",
##               "JUEZ PENAL CIRCUITO", "JUEZ PENAL MUNICIPAL",
##               "JUEZ PROMISCOUO", "LIDER ORGANIZACION CAMPESINA", "LIDER ORGANIZACION COMUNITARIA",
##               "LIDER ORGANIZACION INDIGENA", "LIDER ORGANIZACION POLITICA", "MARINERO ARMADA NACI
##               "MIEMBRO, AFILIADO, ACTIVISTA DE LA UNION PATRIOTICA", "SECRETARIO DE SINDICATO" )
##
## prof_azul <- c("AEROTECNICO SUBJEFE FUERZA AEREA", "AGENTE POLICIA NACIONAL",
##               "BRIGADIER GENERAL DEL EJERCITO NACIONAL", "BRIGADIER GENERAL INFANTERIA DE MARINA",
##               "BRIGADIER GENERAL POLICIA NACIONAL", "CABO PRIMERO DEL EJERCITO NACIONAL",
##               "CABO PRIMERO POLICIA NACIONAL", "CABO SEGUNDO DEL EJERCITO NACIONAL",
##               "CABO SEGUNDO POLICIA NACIONAL", "CABO TERCERO DEL EJERCITO NACIONAL",
##               "CAPITAN DE CORBETA ARMADA NACIONAL", "CAPITAN DE FRAGATA ARMADA NACIONAL",
##               "CAPITAN DE LA FUERZA AEREA", "CAPITAN DE LA MARINA",
##               "CAPITAN DE NAVIO ARMADA NACIONAL", "CAPITAN DEL EJERCITO NACIONAL",
##               "CAPITAN POLICIA NACIONAL", "CONCEJAL", "CONTRA ALMIRANTE ARMADA NACIONAL",
##               "CONTRALOR MUNICIPAL", "CORONEL DEL EJERCITO NACIONAL",
##               "CORONEL INFANTERIA DE MARINA", "CORONEL POLICIA NACIONAL",
##               "DEFENSOR DEL PUEBLO", "DEFENSOR DEL PUEBLO DELEGADO", "DIPUTADO",
##               "EMPLEADOS AREA DE FISCALIA", "EMPLEADOS DEL AREA DEL CTI DE LA FISCALIA",
##               "ENFERMERAS", "ENFERMERAS JEFES", "ESTUDIANTES UNIVERSITARIOS", "EXALCALDE",
##               "FISCAL DELEG. TRIBUNAL SUP.", "FISCAL ESPECIALIZADO", "FISCAL LOCAL",
##               "FISCAL SECCIONAL", "GANADERO",
##               "GENERAL DE LA FUERZA AEREA", "GENERAL DE LA POLICIA NACIONAL",
##               "GENERAL DE LA REPUBLICA", "GENERAL DE LA REPUBLICA (FUERZA PUBLICA)",
##               "GENERAL DEL EJERCITO NACIONAL", "GENERAL INFANTERIA DE MARINA",
##               "GOBERNADOR", "GOBERNADOR DEPARTAMENTO", "INDIGENA", "INDIGENTE",
##               "INGENIEROS DE PETROLEOS", "INTENDENTE POLICIA NACIONAL", "INVESTIGADOR JUDICIAL C.T
##               "INVESTIGADOR JUDICIAL D.A.S.", "INVESTIGADOR JUDICIAL POLICIA NACIONAL",
##               "MAGISTRADO CONSEJO DE ESTADO", "MAGISTRADO CONSEJO SECC. JUDICATURA",
##               "MAGISTRADO CONSEJO SUP. JUDICATURA", "MAGISTRADO TRIBUNAL ADMINISTRATIVO",
##               "MAGISTRADO TRIBUNAL SUPERIOR", "MAGISTRADOS DE TRIBUNALES", "MAYOR DE LA FUERZA AERE
##               "MAYOR DEL EJERCITO NACIONAL", "MAYOR GENERAL DEL EJERCITO NACIONAL",
##               "MAYOR GENERAL POLICIA NACIONAL", "MAYOR INFANTERIA DE MARINA",
##               "MAYOR POLICIA NACIONAL", "MIEMBROS CIVILES ARMADA NACIONAL",
##               "MIEMBROS CIVILES DE LA MARINA", "MIEMBROS CIVILES EJERCITO NACIONAL",
##               "MIEMBROS CIVILES POLICIA NACIONAL", "MIEMBROS FUERZAS MILITARES Y POLICIAS",
##               "MIEMBROS RETIRADOS ARMADA NACIONAL", "MIEMBROS RETIRADOS EJERCITO NACIONAL",
##               "MIEMBROS RETIRADOS FUERZA AEREA MIEMBROS RETIRADOS FUERZAS MILITARES Y DE POLICIA",
##               "MIEMBROS RETIRADOS POLICIA NACIONAL", "OCUPACIONES EXCLUSIVAS DE LAS FUERZAS MILITARI
##               "OFICIALES DE CUBIERTA", "OFICIALES DE LAS FUERZAS MILITARES", "OFICIALES DE POLICIA"
##               "OTROS MIEMBROS ARMADA NACIONAL", "OTROS MIEMBROS DE LA MARINA", "OTROS MIEMBROS EJER
##               "OTROS MIEMBROS FUERZA AEREA", "OTROS MIEMBROS POLICIA NACIONAL",
##               "PRESIDENTE DE SINDICATO", "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA",

```

```

## "PROFESORES UNIVERSITARIOS", "PROMOTORES DE SALUD", "REINSERTADO",
## "REPRESENTANTE A LA CAMARA", "SACERDOTE", "SARGENTO MAYOR DEL EJERCITO NACIONAL",
## "SARGENTO MAYOR POLICIA NACIONAL", "SARGENTO PRIMERO DEL EJERCITO NACIONAL",
## "SARGENTO PRIMERO POLICIA NACIONAL", "SARGENTO SEGUNDO DEL EJERCITO NACIONAL",
## "SARGENTO SEGUNDO POLICIA NACIONAL", "SARGENTO VICE PRIMERO POLICIA NACIONAL",
## "SARGENTO VICEPRIMERO DEL EJERCITO NACIONAL", "SARGENTO VICEPRIMERO INFANTERIA DE AVIACION",
## "SENADOR DE LA REPUBLICA", "SOLDADO EJERCITO NACIONAL", "SOLDADO FUERZA AEREA",
## "SUB COMISARIO POLICIA NACIONAL", "SUB INTENDENTE POLICIA NACIONAL",
## "SUB OFICIAL DE LA MARINA", "SUB OFICIAL JEFE ARMADA NACIONAL",
## "SUB OFICIAL PRIMERO ARMADA NACIONAL", "SUB OFICIAL SEGUNDO ARMADA NACIONAL",
## "SUB TENIENTE POLICIA NACIONAL", "SUBOFICIALES DE LA POLICIA",
## "SUBOFICIALES DE LAS FUERZAS MILITARES", "SUBTENIENTE DEL EJERCITO NACIONAL", "TECNICO",
## "TENIENTE CORONEL DE LA FUERZA AEREA", "TENIENTE CORONEL DEL EJERCITO NACIONAL",
## "TENIENTE CORONEL INFANTERIA DE MARINA", "TENIENTE CORONEL POLICIA NACIONAL",
## "TENIENTE DE CORBETA ARMADA NACIONAL", "TENIENTE DE FRAGATA ARMADA NACIONAL",
## "TENIENTE DE LA FUERZA AEREA", "TENIENTE DE LA MARINA",
## "TENIENTE DE NAVIO ARMADA NACIONAL", "TENIENTE DEL EJERCITO NACIONAL",
## "TENIENTE POLICIA NACIONAL", "TRABAJADORES EN SERVICIO SOCIAL Y COMUNITARIO", "TRABAJADORES",
##
## prof_amarillo <- c("AEROTECNICO JEFE FUERZA AEREA", "AEROTECNICO PRIMERO FUERZA AEREA",
## "AGRICULTOR", "AGRONOMOS Y ESPECIALISTAS AGRICOLAS", "ALBANIL", "ARTESANOS",
## "ASISTENTES ADMINISTRATIVOS", "ASISTENTES DE CONTABILIDAD",
## "ASISTENTES DE JUZGADOS, TRIBUNALES",
## "AUXILIARES DE ENFERMERIA", "AUXILIARES DE ODONTOLOGIA",
## "AUXILIARES DE TRIBUNALES, JUZGADOS",
## "BIOLOGOS, BOTANICOS, ZOOLOGOS Y RELACIONADOS",
## "BRIGADIER GENERAL DE LA FUERZA AEREA", "CAJEROS",
## "CAJEROS DE SERVICIOS FINANCIEROS", "CARNICEROS", "COMERCIANTE",
## "COMERCIANTE DE GANADO", "COMISARIOS, INSPECTORES DE POLICIA",
## "CONDUCTOR O AUXILIAR DE TRANSPORTE", "CONTRALOR DEPARTAMENTA",
## "COTERO", "DESEMPLEADO",
## "DIRECTORES Y ADMINISTRADORES DE EDUCACION BASICA Y DIRECTORES Y GERENTES GENERALES",
## "DIRECTORES Y GERENTES GENERALES PRODUCCION DE BIENES", "DIRECTORES Y GERENTES GENERALES",
## "EBANISTA", "INVESTIGADORES Y ANALISTAS DE POLITICA", "EDILES",
## "ELECTRICISTA", "EMBAJADOR", "EMPLEADOS DE BANCA, SEGUROS Y OTROS SERVICIOS FINANCIEROS",
## "ESTUDIANTE PRIMARIA", "ESTUDIANTES SECUNDARIA", "FARMACEUTICOS",
## "FOTOGRAFOS", "FUNCIONARIOS DE PROGRAMAS EXCLUSIVOS DE LA ADMINIS ",
## "GEOLOGOS, GEOQUIMICOS Y GEOFISICOS ", "GERENTES DE ARQUITECTURA Y CIENCIAS",
## "GUARDIANES DE PRISION", "INGENIERO DE VIAS Y TRANSPORTE", "INGENIERO SANITARIO Y EPIDEMIOLOGIA",
## "INGENIEROS CIVILES", "INGENIEROS DE MINAS", "INGENIEROS DE SISTEMAS",
## "INGENIEROS ELECTRICOS Y ELECTRONICOS", "INGENIEROS INDUSTRIALES",
## "INGENIEROS MECANICOS", "INGENIEROS METALURGICOS", "INGENIEROS QUIMICOS",
## "INSPECTOR DE CONSTRUCCION", "INSPECTORES DE SANIDAD, SEGURIDAD Y SALUD OCUPACIONAL",
## "INSTRUMENTADORES QUIRURGICOS", "INVESTIGADORES Y CONSULTORES, DESARROLLO ECONOMICO",
## "JEFE DE MISION DIPLOMATICA", "JEFES DE BIBLIOTECA, PUBLICACIONES Y EMPLEADOS DE BIBLIOTECA",
## "JEFES DE OFICINA EN GENERAL", "JEFES DE REGISTRO, DISTRIBUCION Y PROGRAMACION",
## "JUEZ CIVIL CIRCUITO", "JUEZ CIVIL MUNICIPAL", "JUEZ DE FAMILIA", "JUEZ DE MENORES",
## "JUEZ DE PAZ", "JUEZ LABORAL", "MATEMATICOS, ESTADISTICOS Y ACTUARIOS",
## "MECANICO AUTOMOTRIZ", "MEDICOS GENERALES", "MERCADERISTAS", "MEDICOS ESPECIALISTAS",
## "MIEMBROS DEL PODER EJECUTIVO LEGISLATIVO Y JUDICIAL", "MINERO",
## "MINISTRO DEL DESPACHO", "MOTOTAXISTA", "NOTARIO PUBLICO", "OBRERO",
## "OCUPACIONES PROFESIONALES EN ORGANIZACION Y ADMINIS", "ODONTOLOGOS",
## "ORIENTADORES EDUCATIVOS", "OTRAS ACTIVIDADES DE LA CONSTRUCCION", "OTRAS ACTIVIDADES",
## "OTRAS OCUPACIONES RELIGIOSAS", "OTRO", "OTROS ARTISTAS", "OTROS ESTUDIANTES",

```

```

## "OTROS INVESTIGADORES JUDICIALES", "OTROS JUECES", "OTROS SUPERVISORES", "PANADERO
## "PELUQUEROS, ESTILISTAS Y AFINES", "PENSIONADO", "PERSONAL DIRECTIVO DE LA ADMINI
## "PATRULLERO POLICIA NACIONAL", "PERIODISTAS", "PERSONERO", "POLICIAS", "PRESI
## "PROCURADOR DELEGADO", "PROFESORES DE EDUCACION BASICA", "POLITOLOGO",
## "PROFESORES DE EDUCACION MEDIA", "PROFESORES DE PREESCOLAR", "PSICOLOGOS",
## "RECICLADOR", "REGISTRADOR NACIONAL ESTADO CIVIL", "SECRETARIAS", "SOLDADOR",
## "SUPERVISORES DE VIGILANCIA", "TECNOLOGOS Y TECNICOS AGROPECUARIOS", "TERAPEUTAS
## "TRABAJADORA SEXUAL", "VICE ALMIRANTE ARMADA NACIONAL", "VIGILANTES Y GUARDIANES
##
## prof_naranja <- c("ABOGADO", "ADMINISTRADOR DE EMPRESAS", "ADMINISTRADORES DE COMERCIO AL POR MENOR"
## "ADMINISTRADORES DE EDUCACION SUPERIOR Y FORMACION",
## "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
## "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES",
## "ARBITROS", "ARQUITECTOS", "ASEADORES Y SERVICIO DOMESTICO",
## "ASISTENTES DE PERSONAL Y SELECCION", "ADMINISTRADOR DE EMPRESAS",
## "ADMINISTRADORES DE COMERCIO AL POR MENOR ADMINISTRADORES DE EDUCACION SUPERIOR Y
## "ADMINISTRADORES DE INMUEBLES", "ADMINISTRADORES Y OPERADORES DE SISTEMAS",
## "ANALISTAS DE SISTEMAS", "ANUNCIADORES Y LOCUTORES", "ASEADORES Y SERVICIO DOMESTICO
## "ASISTENTES DE PERSONAL Y SELECCION", "AUXILIARES ADMINISTRATIVOS",
## "AUXILIARES DE ALMACEN Y BODEGA", "AUXILIARES DE CONTABILIDAD", "AUXILIARES DE OFI
## "AUXILIARES DE PERSONAL Y NOMINA", "AUXILIARES DE SERVICIO A PASAJEROS",
## "AYUDANTES DE COCINA Y CAFETERIA", "CARTEROS Y MENSAJEROS", "CHEFS", "COCINERO",
## "COMERCIANTE DE ESMERALDA", "COMPRADORES", "DEPORTISTA",
## "DIRECTOR DEPARTAMENTO ADMINISTRATIVO",
## "DIRECTORES DE FISCALIAS DIRECTORES DE PROGRAMAS DE ESPARCIMIENTO Y DEPORT",
## "DIRECTORES Y GERENTES SERVICIOS FINANCIEROS",
## "DISENADORES DE TEATRO, MODA Y EXHIBICION Y OTROS C",
## "DISENADORES GRAFICOS Y DIBUJANTES ARTISTICOS", "DISENADORES INDUSTRIALES",
## "DOMESTICADORES Y TRABAJADORES DEL CUIDADO DE ANIMA",
## "ECONOMISTAS", "EMPLEADOS DE INFORMACION Y SERVICIO AL CLIENTE",
## "EMPLEADOS DE PUBLICACION Y AFINES", "EMPLEADOS DE VENTAS Y SERVICIOS DE AREOLINEAS
## "FISICOS Y ASTRONOMOS", "FISIOTERAPEUTAS",
## "GERENTES DE COMERCIO AL POR MENOR", "GERENTES DE EMPRESAS DE TELECOMUNICACIONES",
## "GERENTES DE MEDIOS DE COMUNICACION Y ARTES ESCENIC", "GERENTES DE OTROS SERVICIOS
## "GERENTES DE OTROS SERVICIOS A LAS EMPRESAS", "GERENTES DE OTROS SERVICIOS ADMINIS
## "GERENTES DE PROGRAMAS DE POLITICA DE DESARROLLO EC GERENTES DE SEGUROS, BIENES RA
## "GERENTES DE SERVICIOS DE COMERCIO EXTERIOR", "GERENTES DE VENTAS, MERCADEO Y PUBL
## "GERENTES FINANCIEROS", "GUIAS DE VIAJES Y TURISMO", "HIGIENISTAS DENTALES",
## "MPULSADORES Y DEMOSTRADORES", "LAVACARROS", "MECANOGRAFOS", "MESEROS", "MUSICOS Y
## "OFICIALES DE BOMBEROS", "ORNAMENTADOR", "OTROS INGENIEROS", "OTROS INSPECTORES",
## "OTROS INSTRUCTORES", "PARTICULAR", "PATRONISTAS PRODUCTOS TEXTILES, CUERO Y PIEL"
## "PILOTOS, INGENIEROS E INSTRUCTORES DE VUELO", "PINTORES, ESCULTORES Y OTROS ARTIS
## "PLOMERO", "PRACTICANTES DE LA MEDICINA ALTERNATIVA", "PROGRAMADORES DE SISTEMAS",
## "QUIMICOS", "RECEPCIONISTAS Y OPERADORES DE CONMUTADOR", "RECREACIONISTAS",
## "REPRESENTANTES DE VENTAS NO TECNICAS", "SUPERINTENDENTE DE SALUD", "SUPERINTENDEN
## "SUPERINTENDENTE Y GERENTE DE ESTABLECIMIENTOS PUBL", "SUPERVISORES DE VENTAS",
## "TESORERO", "TOPOGRAFOS", "VENDEDORES DE MOSTRADOR", "VENDEDORES, VENTAS TECNICAS"
##
## prof_rojo <- c("ACTIVIADES RELACIONADAS CON LA ZAPATERIA", "ACTIVIDADES DE LATONERIA Y PINTURA AUTOM
## "ACTIVIDADES DE PESCA", "ACTIVIDADES RELACIONADAS CON EL HOGAR",
## "ACTIVIDADES RELACIONADAS CON LA CARPINTERIA", "ACTIVIDADES RELACIONADAS CON LA TAURON
## "AGENTES DE ADUANA Y OTROS AGENTES", "BARMAN", "BOMBEROS",
## "CONSEJEROS DE EMPLEO", "CONTADORES Y AUDITORES FINANCIEROS",
## "CONTRALOR GENERAL DE LA REPUBLICA", "CONTROLADORES DE TRAFICO AEREO", "DIGITADORES",

```

```

## "OPERADORES DE JUEGOS MECANICOS Y DE SALON", "ORGANIZADORES DE EVENTOS",
## "OTRAS OCUPACIONES DE SERVICIOS PERSONALES", "OTRAS OCUPACIONES ELEMENTALES DE LAS VE
## "OTRAS OCUPACIONES ELEMENTALES DE LOS SERVICIOS OTRAS OCUPACIONES PROFESIONALES EN TE
## "PRESIDENTE DE LA REPUBLICA",
## "TECNICOS DENTALES",
## "TECNICOS EN GRABACION DE AUDIO Y VIDEO", "TECNICOS EN TRANSMISION",
## "TECNICOS OPTICOS",
## "TECNICOS Y MECANICOS DE INSTRUMENTOS DE AERONAVEGA TECNICOS Y MECANICOS DE INSTRUME
## "TECNOLOGOS Y TECNICOS DE LABORATORIO MEDICO Y PATO",
## "TECNOLOGOS Y TECNICOS EN ARQUITECTURA", "TECNOLOGOS Y TECNICOS EN CIENCIAS BIOLGICAS
## "TECNOLOGOS Y TECNICOS EN ELECTROCARDIOGRAFIA Y ELE", "TECNOLOGOS Y TECNICOS EN GEOLO
## "TECNOLOGOS Y TECNICOS EN INGENIERA CIVIL TECNOLOGOS Y TECNICOS EN INGENIERA ELECTRIC
## "TECNOLOGOS Y TECNICOS EN QUIMICA APLICADA TECNOLOGOS Y TECNICOS FORESTALES Y DE RECU
## "TRABAJADORES DE ESTACION DE SERVICIO")
##
## categoria_ca <- c("PARAPOLITICA" ,"FARC-EP","VICTIMAS PERIODISTAS",
## "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO","CONFLICTO ARMADO","CASOS UP" ,
## "HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
## "CASOS EN EL SIDH","VIXCTIMAS DEFENSORES DE DDHH",
## "VINCULO FUNC. PUBLICOS CON GRUPOS ILEGALES","MASACRES",
## "CASOS DEMANDADOS CIDH","CASOS SAN JOSE DE APARTADO",
## "RESTITUCION DE TIERRAS",
## "VICTIMAS INDIGENAS","CASOS AUTO 92 CORTE CONSTITUCIONAL",
## "BANDAS EMERGENTES" ,
## "RELACION DE CIVILES CON ACTORES DE CONFLICTO",
## "COMPULSAS - JUSTICIA Y PAZ","VICTIMAS SINDICALISTAS" ,
## "ELN","CASOS OIT",
## "VICTIMAS DE MINAS ANTIPERSONALES","VIOLENCIA SEXUAL CONFLICTO ARMADO",
## "VICTIMAS MIEMBROS DE LAS ONG'S","MARCHA PATRIOTICA")
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_sijuf_jep <- data_sijuf_jep_fase4 %>%
## mutate(prob = runif(nrow(.))) %>%
## mutate(is_conflict = case_when(
## a29_tipo_persona == "PERPETRADOR" ~ zero, # irrelevant, will be filtered
## perpetrador1 %in% perp_no_ca ~ zero,
## perpetrador2 %in% perp_no_ca ~ zero,
## perpetrador3 %in% perp_no_ca ~ zero,
## a67_delito_unico %in% delito_rojo ~ zero,
## a63_delito %in% delito_rojo ~ zero,
## a23_delito_ppal %in% delito_rojo ~ zero,
## perpetrador1 %in% perp_verde ~ one,
## perpetrador2 %in% perp_verde ~ one,
## perpetrador3 %in% perp_verde ~ one,
## a63_delito %in% delito_verde ~ one,
## a67_delito_unico %in% delito_verde ~ one,
## a24_categoria_caso %in% categoria_ca ~ one,
## a44_categoria %in% categoria_ca ~ one,
## a23_delito_ppal %in% delito_verde ~ one,
## perpetrador1 %in% perp_is_ca & a29_tipo_persona == "VICTIMA" ~ one,
## perpetrador2 %in% perp_is_ca & a29_tipo_persona == "VICTIMA" ~ one,
## perpetrador3 %in% perp_is_ca & a29_tipo_persona == "VICTIMA" ~ one,

```



```

##   perpetrador1 %in% perp_naranja & prob < 0.75 ~ zero,
##   perpetrador2 %in% perp_naranja & prob < 0.75 ~ zero,
##   perpetrador3 %in% perp_naranja & prob < 0.75 ~ zero,
##   a23_delito_ppal %in% delito_naranja & prob < 0.75 ~ zero,
##   a67_delito_unico %in% delito_naranja & prob < 0.75 ~ zero,
##   a63_delito %in% delito_naranja & prob < 0.75 ~ zero,
##   perpetrador1 %in% perp_azul & prob < 0.75 ~ one,
##   perpetrador2 %in% perp_azul & prob < 0.75 ~ one,
##   perpetrador3 %in% perp_azul & prob < 0.75 ~ one,
##   a23_delito_ppal %in% delito_azul & prob < 0.75 ~ one,
##   a23_delito_ppal %in% delito_amarillo & prob < 0.5 ~ one,
##   a63_delito %in% delito_azul & prob < 0.75 ~ one,
##   a63_delito %in% delito_amarillo & prob < 0.5 ~ one,
##   a67_delito_unico %in% delito_azul & prob < 0.75 ~ one,
##   a67_delito_unico %in% delito_amarillo & prob < 0.5 ~ one,
##   a63_delito %in% delito_na ~ NA_integer_,
##   a29_tipo_persona == "OTRA" ~ NA_integer_,
##   a67_delito_unico %in% delito_na ~ NA_integer_,
##   a23_delito_ppal %in% delito_na ~ NA_integer_,
##   TRUE ~ NA_integer_) %>%
##   select(-prob)
##
## sample <- conflict_sijuf_jep %>%
##   filter(is_conflict == 0) %>%
##   select(recordid, is_conflict)
##
## # ----- training data
##
## sample <- sample_frac(sample, 0.2)
##
## write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
## write_parquet(conflict_sijuf_jep, args$output)

```

```
cat(readLines(files$is_ca_spoa_jep_fase3), sep = '\n')
```

#### 9.4.1.6. FGN - SPOA - JEP FASE 3

```

## #
## # Authors:      PA, VG
## # Maintainers  VG, PA, PB, JGD
## # Copyright    2022, HRDAG,
## # =====
## # CO-SIVJNRN-data/individual/FGN/is-ca/src/is-ca-spoa_jep_fase3.R
## #
## # following the email from Folco Zaffalon (8 abr 2022, 17:16 PDT),
## # and the accompanying file
## # 'SIJUF_casos_perpetrador_ocupacion victima_delitos.docx'
## #
## # Verde = SI CONFLICTO (1)
## # Azul = MUY PROBABLE 0.75~(is_conflict == 1)

```

```
## # Amarillo = NO SE SABE 0.5~(is_conflict == 1)
## # Naranja = POCO PROBABLE 0.75~(is_conflict == 0)
## # Rojo = NO CONFLICTO (0)
##
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210) # <-- the seed is really important in this run!
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow, assertr, stringr, lubridate, readr)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--spoa_jep_fase3",
##                      default=here("individual/FGN/clean/output/spoa_jep_fase3.parquet"))
## parser$add_argument("--examples",
##                      default = "individual/FGN/is-ca/hand/example_spoa_jep.csv")
## parser$add_argument("--output",
##                      default = "output/spoa_jep_fase3.parquet")
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## data_spoa_jep_fase3 <- read_parquet(args$spoa_jep_fase3)
##
##
## # #IS -CA
##
## viol_is_ca = c("DESAPARICION FORZADA", "DESPLAZAMIENTO", "HOMICIDIO DOLOSO",
##               "RECLUTAMIENTO Ilicito")
##
## perp_is_ca = c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##               "AGENTES_ESTATALES","ELN - EJERCITO DE LIBERACION NACIONAL",
##               "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##               "FUERZA PUBLICA - DELITOS", "OTRAS GUERRILLAS")
##
## cat_caso_is_ca = c("AUTO 092 (LIGA DE MUJERES DESPLAZADAS)", "CASOS UP",
##                   "CONFLICTO ARMADO", "CRIMEN DE GUERRA",
##                   "ELN", "FARC-EP", "HOMICIDIO PRESENTADO COMO BAJA POR FUERZA PUBLICA",
##                   "HOMICIDIO PRESENTANDO COMO BAJA POR FUERZA PUBLICA",
##                   "LESA HUMANIDAD", "PARAPOLITICA", "POLICIA NACIONAL VINCULOS GRUPOS ILEGALES",
##                   "PRESUNTOS HOMICIDIOS AGENTES DEL ESTADO", "UNION PATRIOTICA",
##                   "MILITARES VINCULOS GRUPOS ILEGALES","RECLUTAMIENTO DE MENORES",
##                   "VINCULO FUNC. PUBLICOS CON GRUPOS ILEGALES",
##                   "RELACION DE CIVILES CON ACTORES DE CONFLICTO",
##                   "VINCULOS F.P CON GRUPOS AL MARGEN DE LA LEY", "VIOLACION DD.HH",
##                   "VIOLENCIA SEXUAL CONFLICTO ARMADO", "VIXCTIMAS DEFENSORES DE DDHH")
##
## # is -ca colores
##
##
```

```
## perp_verde <- c("AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##              "AUTODEFENSAS",
##              "AUTODEFENSAS UNIDAS DE COLOMBIA",
##              "BANDAS EMERGENTES",
##              "EJERCITO DE LIBERACION NACIONAL",
##              "ELN - EJERCITO DE LIBERACION NACIONAL",
##              "EPL - EJERCITO POPULAR DE LIBERACION",
##              "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##              "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##              "FARC",
##              "FARC - FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA",
##              "FARC-EP",
##              "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##              "FARC.",
##              "FUERZA PUBLICA",
##              "MJBC - MOVIMIENTO JAIME BATEMAN CAYON",
##              "OTROS GRUPOS PARAMILITARES",
##              "AGENTES_ESTATALES",
##              "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##              "AUTODEFENSAS",
##              "AUTODEFENSAS UNIDAS DE COLOMBIA",
##              "BANDAS EMERGENTES",
##              "CAP - COMANDO ARMADO POPULAR",
##              "EJERCITO DE LIBERACION NACIONAL",
##              "ELN - EJERCITO DE LIBERACION NACIONAL",
##              "EPL - EJERCITO POPULAR DE LIBERACION",
##              "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##              "FARC",
##              "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##              "FARC-EP",
##              "FARC.",
##              "FP - DELITOS",
##              "FUERZA PUBLICA",
##              "GUERRILLA",
##              "MPLICADO FUNCIONARIO PUBLICO",
##              "MOVIMIENTO 26 DE ABRIL CARLOS PIZARRO LEON",
##              "OTROS GRUPOS PARAMILITARES",
##              "PARAMILITARES",
##              "AGENTES_ESTATALES",
##              "AUC - AUTODEFENSAS UNIDAS DE COLOMBIA",
##              "AUTODEFENSAS",
##              "BANDAS EMERGENTES",
##              "EJERCITO DE LIBERACION NACIONAL",
##              "ELN - EJERCITO DE LIBERACION NACIONAL",
##              "EPL - EJERCITO POPULAR DE LIBERACION",
##              "ERG - EJERCITO REVOLUCIONARIO GUEVARISTA",
##              "ERP - EJERCITO REVOLUCIONARIO DEL PUEBLO",
##              "FARC - FUERZAS ARMADAS REVOLUCIONARIAS COLOMB",
##              "MOVIMIENTOS POLITICOS",
##              "PARAMILITARES")
##
## perp_rojo <- c("DELINCUENCIA COMUN","COMUNIDADES INDIGENAS")
##
## perp_amarillo <- c("CARTEL DEL NORTE DEL VALLE","NARCOTRAFICO",
```

```
##           "CMR - COMANDO MILICIANO REVOLUCIONARIO",
##           "FP - DELITOS",
##           "IMPLICADO FUNCIONARIO PUBLICO")
##
## perp_naranja <- c("MOVIMIENTOS POLITICOS")
##
## #a35_titulo
## titulo_rojo <- c("DE LA PROTECCION DE LA INFORMACION Y DE LOS DATOS",
##                "DE LOS DELITOS CONTRA LA SALUD PUBLICA",
##                "DE LOS DELITOS CONTRA LOS ANIMALES",
##                "DE LOS DELITOS CONTRA LOS DERECHOS DE AUTOR",
##                "DECRETOS",
##                "DELITOS CONTRA EL ORDEN ECONOMICO SOCIAL",
##                "DELITOS CONTRA EL ORDEN ECONOMICO Y SOCIAL",
##                "DELITOS CONTRA LA ADMINISTRACION DE PUBLICA",
##                "DELITOS CONTRA LA EFICAZ Y RECTA IMPARTICION DE JUSTICIA",
##                "DELITOS CONTRA LA FAMILIA",
##                "DELITOS CONTRA LA FE PUBLICA",
##                "DELITOS CONTRA LA INTEGRIDAD MORAL",
##                "DELITOS CONTRA LA SALUD PUBLICA")
##
## titulo_verde <- c("DE LOS DELITOS CONTRA EL REGIMEN CONSTITUCIONAL Y LEGAL",
##                  "DELITOS CONTRA LAS PERSONAS Y BIENES PROTEGIDOS POR EL DERECH",
##                  "DELITOS CONTRA PERSONAS Y BIENES PROTEGIDOS POR EL D.I.H.")
##
## titulo_amarillo <- c("DE LOS DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                      "DELITOS CONTRA LA LIBERTAD INDIVIDUAL Y OTRAS GARANTIAS",
##                      "DELITOS CONTRA LA LIBERTAD, INTEGRIDAD Y FORMACION SEXUALES",
##                      "DELITOS CONTRA LA VIDA Y LA INTEGRIDAD PERSONAL",
##                      "LEYES",
##                      "SIN DELITO")
##
## titulo_naranja <- c("DELITOS CONTRA EL PATRIMONIO ECONOMICO",
##                    "DELITOS CONTRA LA EXISTENCIA Y SEGURIDAD DEL ESTADO",
##                    "DELITOS CONTRA LA SEGURIDAD PUBLICA",
##                    "DELITOS CONTRA LOS RECURSOS NATURALES Y EL MEDIO AMBIENTE",
##                    "DELITOS CONTRA MECANISMOS DE PARTICIPACION DEMOCRATICA")
##
## #tipo_hecho
## delito_rojo <- c("ABANDONO ART 128 C.P.",
##                 "ABANDONO ART. 127 C.P.",
##                 "ABANDONO DE HIJO FRUTO DE ACCESO CARNAL VIOLENTO ABUSIVO O DE INSEMINACION ARTIFIC",
##                 "ABUSO DE CONFIANZA CALIFICADO. ART. 250 C.P.",
##                 "ABUSO DE CONFIANZA. ART. 249 C.P.",
##                 "ACAPARAMIENTO. ART. 297 C.P.",
##                 "ADOPCION IRREGULAR. ART. 232",
##                 "AGIOTAJE. ART. 301",
##                 "APLICACION FRAUDULENTO DE CREDITO OFICIALMENTE REGULADO ART. 311 C.P.",
##                 "APROVECHAMIENTO DE ERROR AJENO O CASO FORTUITO. ART. 252 C.P.",
##                 "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P.",
##                 "ASESORAMIENTO Y OTRAS ACTUACIONES ILEGALES ART. 421 C.P. INCISO 1",
##                 "ASOCIACION PARA LA COMISION DE UN DELITO CONTRA LA ADMINISTRACION PUBLICA ART. 434",
##                 "CALUMNIA. ART. 221 C.P.",
##                 "CAPTACION MASIVA Y HABITUAL DE DINEROS ART. 316 C.P.",
```

## "CAZA ILEGAL ART. 336",  
## "CELEBRACION INDEBIDA DE CONTRATO DE SEGURO ART. 172 C.P.",  
## "CIRCULACION Y USO DE EFECTO OFICIAL O SELLO FALSIFICADO. ART. 281 C.P.",  
## "CIRCUNSTANCIA DE AGRAVACION PUNIBLE ART 377B CP LEY 1453 DE 2011",  
## "COHECHO IMPROPIO ART. 406 C.P.",  
## "COHECHO POR DAR U OFRECER ART. 407 C.P.",  
## "COHECHO PROPIO ART. 405 C.P.",  
## "CONTRABANDO ART. 319 C.P.",  
## "CONTRATO SIN CUMPLIMIENTO DE REQUISITOS LEGALES ART. 410 C.P.",  
## "CORRUPCION DE ALIMENTOS, PRODUCTOS MEDICOS O MATERIAL PROFILACTICO ART. 372 C.P.",  
## "CORRUPCION PRIVADA ART. 250A C.P.",  
## "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",  
## "DAÑO INFORMatico ART 269D LEY 1273 DE 2009",  
## "DE LA PRESTACION, ACCESO O USO ILEGALES DE LOS SERVICIOS DE TELECOMUNICACIONES ART  
## "DEFRAUDACION A LOS DERECHOS PATRIMONIALES DE AUTOR. ART. 271 C.P.",  
## "DEFRAUDACION DE FLUIDOS. ART. 256 C.P.",  
## "DENEGACION DE INSCRIPCION ART. 396 C.P.",  
## "DESTRUCCION, SUPRESION U OCULTAMIENTO DE DOCUMENTO PUBLICO. ART. 292 C.P.",  
## "DESTRUCCION, SUPRESION Y OCULTAMIENTO DE DOCUMENTO PRIVADO. ART. 293 C.P.",  
## "EJERCICIO ARBITRARIO DE LA CUSTODIA DE HIJO MENOR DE EDAD ART. 230A C.P. AD. LEY 8  
## "EJERCICIO ILICITO DE ACTIVIDAD MONOPOLISTICA DE ARBITRIO RENTISTICO ART. 312 C.P."  
## "EMISION Y TRANSFERENCIA ILEGAL DE CHEQUE. ART. 248 C.P.",  
## "EMISIONES ILEGALES. ART. 276 C.P.",  
## "ENRIQUECIMIENTO ILICITO ART. 412 C.P.",  
## "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P.",  
## "ENRIQUECIMIENTO ILICITO DE PARTICULARES ART. 327 C.P. INFERIOR 100 SMLM",  
## "ESTAFA. ART. 246 C.P.",  
## "ESTIMULO AL USO ILICITO ART. 378 C.P.",  
## "EXPORTACION O IMPORTACION FICTICIA ART. 310 C.P.",  
## "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",  
## "FALSA AUTOACUSACION ART. 437 C.P.",  
## "FALSA DENUNCIA ART. 435 C.P.",  
## "FALSA DENUNCIA CONTRA PERSONA DETERMINADA ART. 436 C.P.",  
## "FALSEDAD EN DOCUMENTO PRIVADO. ART. 289 C.P.",  
## "FALSEDAD IDEOLOGICA EN DOCUMENTO PUBLICO. ART. 286 C.P.",  
## "FALSEDAD MARCARIA. ART. 285 C.P.",  
## "FALSEDAD MARCARIA. ART. 285 C.P. INCISO 1",  
## "FALSEDAD MATERIAL EN DOCUMENTO PUBLICO. ART. 287 C.P.",  
## "FALSEDAD PARA OBTENER PRUEBA DE HECHO VERDADERO. ART. 295 C.P.",  
## "FALSEDAD PERSONAL. ART. 296 C.P.",  
## "FALSIFICACION DE EFECTO OFICIAL TIMBRADO. ART. 280 C.P.",  
## "FALSIFICACION DE MONEDA NACIONAL O EXTRANJERA. ART. 273 C.P.",  
## "FALSIFICACION O USO FRAUDULENTO DE SELLO OFICIAL. ART. 279 C.P.",  
## "FALSO TESTIMONIO ART. 442 C.P.",  
## "FAVORECIMIENTO ART. 446 C.P.",  
## "FAVORECIMIENTO DE CONTRABANDO ART. 320 C.P.",  
## "FECUNDACION Y TRAFICO DE EMBRIONES ART. 134 C.P.",  
## "FEMINICIDIO ART. 104A C.P.",  
## "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",  
## "FRAUDE PROCESAL ART. 453 C.P.",  
## "FUGA DE PRESOS ART. 448 C.P.",  
## "GESTION INDEBIDA DE LOS RECURSOS SOCIALES. ART. 260 C.P.",  
## "HOMICIDIO CULPOSO ART. 109 C.P.",  
## "HOMICIDIO POR PIEDAD ART. 106 C.P.",

## "HURTO CALIFICADO. ART. 240 C.P.",  
## "HURTO POR MEDIOS INFORMATICOS Y SEMEJANTES ART. 269I LEY 1273 DE 2009",  
## "HURTO. ART. 239 C.P.",  
## "ILICITA EXPLOTACION COMERCIAL ART. 303",  
## "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",  
## "IMITACION O SIMULACION DE ALIMENTOS, PRODUCTOS O SUSTANCIAS ART. 373 C.P.",  
## "IMPEDIMENTO O PERTURBACION DE LA CELEBRACION DE AUDIENCIAS PUBLICAS ART. 454C C.P.",  
## "INASISTENCIA ALIMENTARIA ART. 233 C.P.",  
## "INCENDIO ART. 350 C.P.",  
## "INCESTO. ART. 237 C.P.",  
## "INDUCCION A AYUDA AL SUICIDIO ART. 107 C.P.",  
## "INFIDELIDAD A LOS DEBERES PROFESIONALES ART. 445 C.P.",  
## "INJURIA. ART. 220 C.P.",  
## "INTERES INDEBIDO EN LA CELEBRACION DE CONTRATOS ART. 409 C.P.",  
## "INTERVENCION EN POLITICA ART. 422 C.P.",  
## "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",  
## "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",  
## "IRRESPETO A CADAVERES. ART. 204 C.P.",  
## "LESIONES AL FETO ART. 125 C.P.",  
## "LESIONES CULPOSAS AL FETO ART. 126 C.P.",  
## "LESIONES CULPOSAS ART. 120 C.P.",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O M.",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCI.",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION FUNCIONAL TRANSITORIA ART.114 INC.",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA PERMANENTE ART.115",  
## "LESIONES CULPOSAS ART. 120 C.P. CON PERTURBACION PSIQUICA TRANSITORIA ART.115",  
## "LESIONES CULPOSAS ART. 120 C.P. INCISO 1",  
## "LESIONES CULPOSAS ART. 120 C.P.AGRAVADO POR NO LICENCIA O LICENCIA SUSPENDIDA ART.",  
## "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 D.",  
## "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MAYOR 90 DIAS ART. 112 C.",  
## "LESIONES PERSONALES CULPOSAS ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.",  
## "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA AFECTA ROSTRO ART.",  
## "LESIONES PERSONALES CULPOSAS ART. 120 C.P CON DEFORMIDAD FISICA PERMANENTE ART. 11.",  
## "LESIONES PERSONALES CULPOSAS ART. 120 C.P. CON DEFORMIDAD FISICA TRANSITORIA ART.",  
## "MALTRATO ANIMAL ART. 339A C.P.",  
## "MALTRATO MEDIANTE RESTRICCION A AL LIBERTAD FISICA. ART. 230 C.P.",  
## "MALTRATO POR DESCUIDO O ABANDONO EN PERSONA MAYOR DE 60 AÑOS",  
## "MALVERSACION Y DILAPIDACION DE BIENES FAMILIARES. ART. 236 C.P.",  
## "MANIPULACION FRAUDULENTA DE ESPECIES INSCRITAS EN EL REGISTRO NACIONAL DE VALORES I",  
## "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",  
## "OBTENCION DE DOCUMENTO PUBLICO FALSO ART. 288 C.P.",  
## "OCULTAMIENTO, ALTERACION O DESTRUCCION DE ELEMENTO MATERIAL PROBATORIO ART.454B C.",  
## "OCULTAMIENTO, RETENCION Y POSESION ILICITA DE CEDULA ART. 395 C.P.",  
## "OFRECIMIENTO ENGAÑOSO DE PRODUCTOS Y SERVICIOS. ART. 300 C.P.",  
## "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P.",  
## "OMISION DE DENUNCIA DE PARTICULAR ART. 441 C.P. MODIFICADO ART. 18 LEY 1121 DE 200",  
## "OMISION DEL AGENTE RETENEDOR O RECAUDADOR ART. 402 C.P.",  
## "PANICO ECONOMICO ART. 302 C.P.",  
## "PARTO O ABORTO PRETERINTENCIONAL ART. 118 C.P.",  
## "PECULADO CULPOSO ART. 400 C.P.",  
## "PECULADO POR APLICACION OFICIAL DIFERENTE ART. 399 C.P.",  
## "PECULADO POR APROPIACION ART. 397 C.P.",  
## "PECULADO POR USO ART. 398 C.P.",  
## "PESCA ILEGAL ART. 335",

```

##          "PORNOGRAFIA CON MENORES ART. 218 C.P.",
##          "PORTE DE SUSTANCIAS ART. 383 C.P.",
##          "PREVARICATO POR ACCION ART. 413 C.P.",
##          "PREVARICATO POR OMISION ART. 414 C.P.",
##          "PROPAGACION DE EPIDEMIA ART. 369 C.P.",
##          "PROPAGACION DEL VIRUS DE INMUNODEFICIENCIA HUMANA O DE LA HEPATITIS B ART. 370 C.P.",
##          "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P.",
##          "PROXENETISMO CON MENOR DE EDAD ART. 213A C.P. ADICIONADO LEY 1329 DE 2009",
##          "RECEPTACION ART. 327 C C.P.",
##          "RECEPTACION ART. 447 C.P.",
##          "REPETIBILIDAD DEL SER HUMANO ART. 133 C.P.",
##          "SABOTAJE. ART. 199 C.P.",
##          "SIN DELITO",
##          "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
##          "SOBORNO ART. 444 C.P.",
##          "SOBORNO EN LA ACTUACION PENAL ART. 444A C.P. AD. LEY 890/2004 ART.10",
##          "SUMINISTRO A MENOR ART. 381 C.P.",
##          "SUPRESION, ALTERACION O SUPOSICION DEL ESTADO CIVIL. ART. 238 C.P.",
##          " SUSTRACCION DE BIEN PROPIO. ART. 254 C.P.",
##          "SUSTRACCION DE COSA PROPIA AL CUMPLIMIENTO DE DEBERES CONSTITUCIONALES O LEGALES A",
##          "TRAFICO DE MIGRANTES ART. 188 C.P. MOD. LEY 747/2002 ART.1",
##          "TRAFICO DE MONEDA FALSIFICADA. ART. 274 C.P.",
##          "TRAFICO DE NIÑAS, NIÑOS Y ADOLESCENTES ART 188C LEY 1453 DE 2011",
##          "TRAFICO, ELABORACION Y TENENCIA DE ELEMENTOS DESTINADOS A LA FALSIFICACION DE MONE",
##          "TRANSFERENCIA NO CONSENTIDA DE ACTIVOS VALIENDOSE DE ALGUNA MANIPULACION INFORMATI",
##          "TRATA DE PERSONAS ART. 188A C.P.",
##          "TRATA DE PERSONAS TRANSNACIONAL ART. 188A C.P.",
##          "TURISMO SEXUAL. ART.219 C.P. MOD. POR ART 23 L, 1336 DE 2009",
##          "URBANIZACION ILEGAL ART. 318 C.P.",
##          "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
##          "USO DE SOFTWARE MALICIOSO ART 269E LEY 1273 DE 2009",
##          "USO ILEGITIMO DE PATENTES ART. 307 C.P.",
##          "USO ILEGITIMO DE PATENTES ART. 307 C.P. INCISO 1. EL QUE FABRIQUE PRODUCTO SIN AUT",
##          "USURA ART. 305 C.P.",
##          "USURA ART. 305 C.P. INCISO 1 A CAMBIO DE PRESTAMO DE DINERO O POR CONCEPTO DE VENT",
##          "USURPACION DE AGUAS. ART. 262 C.P.",
##          "USURPACION DE DERECHOS DE PROPIEDAD INDUSTRIAL Y DERECHOS DE OBTENTORES DE VARIEDAD",
##          "UTILIZAC.O FACILITAC.MEDIOS DE COMUNICAC.PARA OFRECER ACTIV. SEXUALES CON MENORES I",
##          "VIOLACION A LA LIBERTAD RELIGIOSA. ART. 201 C.P.}VIOLACION A LOS DERECHOS MORALES I",
##          "VIOLACION A LOS DERECHOS MORALES DE AUTOR. ART. 270 C.P. N.1",
##          "VIOLACION A LOS DERECHOS PATRIMONIALES DE AUTOR Y DERECHOS CONEXOS ART. 271 C.P. M",
##          "VIOLACION AL REGIMEN LEGAL O CONSTITUCIONAL DE INHABILIDADES E INCOMPATIBILIDADES",
##          "VIOLACION DE HABITACION AJENA. ART. 189 C.P.",
##          "VIOLACION DE LA RESERVA INDUSTRIAL O COMERCIAL ART. 308 C.P.",
##          "VIOLENCIA INTRAFAMILIAR ART. 229 C.P.")
##
## delito_naranja <- c("ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P. INCISO 1",
##                    "ACCESO ABUSIVO A UN SISTEMA INFORMatico ART 269A LEY 1273 DE 2009",
##                    "ACCESO ABUSIVO A UN SISTEMA INFORMatico. ART. 195 C.P.",
##                    "ACOSO SEXUAL ART. 210A C.P. ADICIONADO LEY 1257 DE 2008",
##                    "ACTO SEXUAL VIOLENTO CON MENOR DE CATORCE AÑOS ART. 209 C.P.",
##                    "ACTO SEXUAL VIOLENTO. ART. 206 C.P.",
##                    "DEMANDA DE EXPLOT.SEX. COMERC. MENOR DE 18 AÑOS ADIC. ART 217A C.P. LEY 1329 DE",
##                    "DISPARO DE ARMA DE FUEGO ART. 356A CP LEY 1453 DE 2011",

```

```
## "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
## "EXPLOTACION DE MENORES DE EDAD",
## "FRAUDE A RESOLUCION JUDICIAL ART. 454 C.P.",
## "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
## "HOMICIDIO ART. 103 C.P.",
## "INSTIGACION A DELINQUIR ART. 348 C.P.",
## "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
## "INTERCEPTACION DE DATOS INFORMATICOS, ART 269C LEY 1273 DE 2009",
## "LESIONES ART. 111 C.P.",
## "LESIONES ART. 113 C.P.",
## "LESIONES CON PERDIDA ANATOMICA O FUNCIONAL DE UN ORGANO O MIEMBRO ART.116",
## "LESIONES CON PERTURBACION FUNCIONAL PERMANENTE ART.114 INCISO 2",
## "LESIONES CON PERTURBACION PSIQUICA PERMANENTE ART.115",
## "LESIONES CON PERTURBACION PSIQUICA TRANSITORIA ART.115",
## "LESIONES PERSONALES AGRAVADAS POR CIRCUNSTANCIAS ART 104 C.P. ART. 119 C.P.",
## "LESIONES PERSONALES ART 120 C.P. CON INCAPACIDAD MENOR 30 DIAS ART.112 C.P. INC.",
## "LESIONES PERSONALES CON DEFORMIDAD FISICA AFECTA ROSTRO ART. 113 C.P.",
## "LESIONES PERSONALES CON DEFORMIDAD FISICA PERMANENTE ART. 113 C.P.INCISO 2",
## "LESIONES PERSONALES CON INCAPACIDAD MAYOR 30 DIAS MENOR 90 DIAS ART. 112 C.P. I.",
## "OBSTACULIZACION ILEGITIMA DEL SISTEMA INFORMATICO O RED DE TELECOMUNICACION ART",
## "OMISION DE APOYO ART. 424 C.P.",
## "OMISION DE SOCORRO ART. 131 C.P.",
## "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
## "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
## "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
## "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
## "SECUESTRO SIMPLE ART. 168 C.P.",
## "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO CUANDO SE TRAFIQUE CON LA PERSONA SECUES",
## "SECUESTRO SIMPLE ART. 168 C.P. AGRAVADO SOBREVENIR MUERTE O LESIONES PERSONALES",
## "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
## "SUPLANTACION DE SITIOS WEB PARA CAPTURAR DATOS PERSONALES ART 269G LEY 1273 DE 1",
## "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
## "TRAFICO DE INFLUENCIAS DE SERVIDOR PUBLICO ART. 411 C.P.",
## "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C.",
## "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
## "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
## "UTILIZACION INDEBIDA DE INFLUENCIAS DERIVADAS DEL EJERCICIO DE FUNCION PUBLICA",
## "VIOLACION DE INMUNIDAD DIPLOMATICA ART. 465 C.P.",
## "VOTO FRAUDULENTO ART. 391 C.P")
##
##
## delito_verde <- c("ABORTO FORZADO EN PERSONA PROTEGIDA ART. 139E",
## "ACCESO CARNAL ABUSIVO EN PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 138",
## "ACCESO CARNAL VIOLENTO EN PERSONA PROTEGIDA ART. 138 C.P.",
## "ACTOS DE TERRORISMO ART. 144 C.P.",
## "ACTOS SEXUALES CON PERSONA PROTEGIDA MENOR DE CATORCE AÑOS ART. 139A",
## "ACTOS SEXUALES VIOLENTOS EN PERSONA PROTEGIDA ART. 139 C.P.",
## "ADIC.L.579/2002 ART.2.EMPLEO,PRODUCCION,COMERCIALIZACION Y ALMACENAMIENTO DE MINAS",
## "ADIC.L.579/2002 ART.3.AYUDA E INDUCCION AL EMPLEO, PRODUCCION Y TRANSFERENCIA DE I",
## "ADMINISTRACION DE RECURSOS RELACIONADOS CON ACTIVIDADES TERRORISTAS ART. 345 C.P.",
## "AMENAZAS CONTRA DEFENSORES DE DERECHOS HUMANOS Y SERVIDORES PUBLICOS ART. 188E",
## "APOLOGIA DEL GENOCIDIO ART. 102 C.P.",
```



```
## "ASONADA ART. 469 C.P.",
## "ATAQUE CONTRA OBRAS E INSTALACIONES QUE CONTIENEN FUERZAS PELIGROSAS ART. 157 C.P."
## "ATENTADOS A LA SUBSISTENCIA Y DEVASTACION ART. 160 C.P.",
## "CONSPIRACION ART. 471 C.P.",
## "CONSTREÑIMIENTO A APOYO BELICO ART. 150 C.P.",
## "DEPORTACION, EXPULSION, TRASLADO O DESPLAZAMIENTO FORZADO DE POBLACION CIVIL ART."
## "DESCONOCIMIENTO DE HABEAS CORPUS ART. 177 C.P.",
## "DESNUDEZ FORZADA EN PERSONA PROTEGIDA ART. 139D",
## "DESPLAZAMIENTO FORZADO ART. 180 C.P.",
## "DESPOJO EN EL CAMPO DE BATALLA ART. 151 C.P.",
## "DESTRUCCION DE BIENES E INSTALACIONES DE CARACTER SANITARIO ART. 155 C.P.",
## "DESTRUCCION DEL MEDIO AMBIENTE ART. 164 C.P.",
## "DESTRUCCION O UTILIZACION ILICITA DE BIENES CULTURALES Y DE LUGARES DE CULTO ART."
## "DESTRUCCION Y APROPIACION DE BIENES PROTEGIDOS ART. 154 C.P.",
## "DETENCION ARBITRARIA ESPECIAL ART. 176 C.P.",
## "DETENCION ILEGAL Y PRIVACION DEL DEBIDO PROCESO ART. 149 C.P.",
## "EMBARAZO FORZADO EN PERSONA PROTEGIDA ART. 139C",
## "ESCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
## "EXACCION O CONTRIBUCIONES ARBITRARIAS ART. 163 C.P.",
## "EXCLAVITUD SEXUAL EN PERSONA PROTEGIDA ART. 141A",
## "GENOCIDIO ART. 101 C.P.",
## "HOMICIDIO EN PERSONA PROTEGIDA ART. 135 C.P.",
## "HOSTILIDAD MILITAR ART. 456 C.P.",
## "LESIONES EN PERSONA PROTEGIDA ART. 136 C.P.",
## "OBSTACULIZACION DE TAREAS SANITARIAS Y HUMANITARIAS ART. 153 C.P.",
## "OMISION DE MEDIDAS DE PROTECCION A LA POBLACION CIVIL ART. 161 C.P.",
## "OMISION DE MEDIDAS DE SOCORRO Y ASISTENCIA HUMANITARIAS ART. 152 C.P.",
## "PERFIDIA ART. 143 C.P.",
## "PRIVACION ILEGAL DE LA LIBERTAD ART. 174 C.P.",
## "PROLONGACION ILICITA DE LA PRIVACION DE LA LIBERTAD ART. 175 C.P.",
## "PROSTITUCION FORZADA EN PERSONA PROTEGIDA ART. 141",
## "PROSTITUCION FORZADA O ESCLAVITUD SEXUAL ART. 141 C.P.",
## "REBELION ART. 467 C.P.",
## "RECLUTAMIENTO ILICITO ART. 162 C.P.",
## "REPRESALIAS ART. 158 C.P.",
## "SEDICION ART. 468 C.P.",
## "SEDUCCION, USURPACION Y RETENCION ILEGAL DEL MANDO ART. 472 C.P.",
## "TOMA DE REHENES ART. 148 C.P.",
## "TORTURA CONTRA PERSONA PROTEGIDA ART. 137 C.P.",
## "TRATOS INHUMANOS Y DEGRADANTES Y EXPERIMENTOS BIOLOGICOS EN PERSONA PROTEGIDA ART."
## "UTILIZACION DE MEDIOS Y METODOS DE GUERRA ILICITOS ART. 142 C.P.",
## "UTILIZACION ILEGAL DE UNIFORMES E INSIGNIAS ART. 346 C.P.")
##
## delito_amarillo <- c("ABORTO ART. 122 C.P.",
## "DAÑO EN BIEN AJENO. ART. 265 C.P.",
## "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
## "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
## "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE AÑOS. ART. 208 C.P.",
## "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
## "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
## "ABUSO DE AUTORIDAD POR ACTO ARBITRARIO O INJUSTO ART. 416",
## "ABUSO DE CONDICIONES DE INFERIORIDAD. ART. 251 C.P.",
## "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 210",
## "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR ART. 210")
```

```

## "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
## "ACTOS SEXUALES CON MENOR DE CATORCE AÑOS. ART. 209 C.P.",
## "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
## "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
## "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
## "ALZAMIENTO DE BIENES. ART. 253 C.P.",
## "CONCUSION ART. 404 C.P.",
## "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
## "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
## "ABUSO DE FUNCION PUBLICA ART. 428 C.P.",
## "ACTOS DE DISCRIMINACION RACIAL ART. 147 C.P.",
## "ACTOS DE RACISMO O DISCRIMINACIÓNART. 134 A",
## "CONSTREÑIMIENTO A LA PROSTITUCION ART. 214 C.P.",
## "CONSTREÑIMIENTO AL SUFRAGANTE ART. 387 C.P.",
## "CONSTREÑIMIENTO ILEGAL ART. 182 C.P.",
## "CONTAMINACION AMBIENTAL ART. 332 C.P.",
## "CONTAMINACION AMBIENTAL POR EXPLOTACION DE YACIMIENTO MINERO O HIDROCARBURO ART. 332 C.P.",
## "CORRUPCION DE SUFRAGANTE ART. 390 C.P.",
## "DESTINACION ILEGAL DE COMBUSTIBLE ART. 327D C.P.",
## "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
## "EXTORSION. ART. 244 C.P.",
## "FABRICACION, IMPORTACION, TRAFICO, POSESION Y USO DE ARMAS QUIMICAS, BIOLOGICAS, NUCLEARES O RADIOACTIVAS ART. 365 C.P.",
## "FABRICACION, TRAFICO Y PORTE DE ARMAS DE FUEGO O MUNICIONES ART. 365 C.P.",
## "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.",
## "FAVORECIMIENTO DE LA FUGA ART. 449 C.P.",
## "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART. 320-1 C.P.",
## "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P.",
## "FAVORECIMIENTO POR SERVIDOR PUBLICO ART. 322 C.P. INCISO 1",
## "OMISION DE CONTROL ART. 325 C.P.",
## "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
## "SECUESTRO EXTORSIVO ART. 169 C.P.",
## "TERRORISMO ART. 343 C.P.",
## "TESTAFERRATO ART. 326 C.P.",
## "TESTAFERRATO ART. 326 C.P. MENOR 100 SALARIOS",
## "TRAFICO, TRANSPORTE Y POSESION DE MATERIALES RADIOACTIVOS O SUSTANCIAS NUCLEARES O BIOLOGICAS ART. 365 C.P.",
## "USO DE MENORES DE EDAD LA COMISION DE DELITO ART 188D LEY 1453 DE 2011",
## "USURPACION DE TIERRAS. ART. 261 C.P.",
## "UTILIZACION INDEBIDA DE INFORMACION OFICIAL PRIVILEGIADA. ART. 420 C.P.",
## "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.",
## "VIOLACION DE LA LIBERTAD DE TRABAJO. ART. 198 C.P.",
## "FRAUDE DE SUBVENCIONES",
## "HOMICIDIO PRETERINTENCIONAL ART. 105 C.P.",
## "INDUCCION A LA PROSTITUCION ART. 213 C.P.",
## "LAVADO DE ACTIVOS ART. 323 C.P.",
## "USURPACION DE MARCAS Y PATENTES ART. 306 C.P.",
## "DAÑO")
## delito_azul <- c("CONSTREÑIMIENTO PARA DELINQUIR ART. 184 C.P.",
## "ACTOS DE BARBARIE ART. 145 C.P.",
## "AMENAZAS A TESTIGOS ART. 454A C.P. AD. LEY 890 DE 2004 ART.13",
## "AMENAZAS ART. 347 C.P.",
## "APODERAMIENTO DE AERONAVES, NAVES O MEDIOS DE TRANSPORTE COLECTIVO. ART. 173 C.P.",
## "APODERAMIENTO DE LOS HIDROCARBUROS, SUS DERIVADOS, BIOCOMBUSTIBLES O MEZCLAS QUE LOS CONTIENAN ART. 319-1 C.P.",
## "CONCIERTO PARA DELINQUIR ART. 340 C.P.",

```

```

##      "DAÑO EN LOS RECURSOS NATURALES ART. 331 C.P.",
##      "DAÑO EN OBRAS DE UTILIDAD SOCIAL ART. 351 C.P.",
##      "DAÑO EN OBRAS O ELEMENTOS DE LOS SERVICIOS DE COMUNICACIONES, ENERGIA Y COMBUSTIBLES ART. 352 C.P.",
##      "DESAPARICION FORZADA ART. 165 C.P.",
##      "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##      "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##      "EMPLEO ILEGAL DE LA FUERZA PUBLICA ART. 423 C.P.",
##      "EMPLEO O LANZAMIENTO DE SUSTANCIAS U OBJETOS PELIGROSOS ART. 359 C.P.",
##      "ENTRENAMIENTO PARA ACTIVIDADES ILICITAS ART. 341 C.P.",
##      "ESPIONAJE ART. 463 C.P.",
##      "FABRICACION, TRAFICO Y PORTE DE ARMAS DE USO PRIVATIVO DE LAS FUERZAS ARMADAS ART. 358 C.P.",
##      "HOSTIGAMIENTO POR MOTIVOS DE RAZA, RELIGIÓN, IDEOLOGÍA, POLÍTICA, U ORIGEN NACIONAL ART. 357 C.P.",
##      "OBSTRUCCION A VIAS PUBLICAS QUE AFECTAN EL ORDEN PUBLICO. ART.353A C.P. LEY 1453 DE 2009",
##      "TORTURA ART. 178 C.P.",
##      "UTILIZACION DE ASUNTO SOMETIDO A SECRETO O RESERVA ART. 419 C.P.",
##      "VIOLACION DE DATOS PERSONALES ART 269F LEY 1273 DE 2009",
##      "VIOLACION DE HABITACION AJENA POR SERVIDOR PUBLICO ART. 190 C.P.",
##      "VIOLACION DE LOS DERECHOS DE REUNION Y ASOCIACION. ART. 200 C.P.",
##      "VIOLACION ILICITA DE COMUNICACIONES O CORRESPONDENCIA DE CARACTER OFICIAL. ART. 191 C.P.",
##      "VIOLACION ILICITA DE COMUNICACIONES. ART. 192 C.P.",
##      "VIOLENCIA CONTRA SERVIDOR PUBLICO ART. 429 C.P.")
##
## delito_na <- c("ABORTO ART. 122 C.P.",
##              "ABORTO SIN CONSENTIMIENTO ART. 123 C.P.",
##              "ABUSO DE AUTORIDAD POR OMISION DE DENUNCIA ART. 417 C.P.",
##              "ACCESO CARNAL ABUSIVO CON MENOR DE CATORCE ANOS. ART. 208 C.P.",
##              "ACCESO CARNAL O ACTO SEXUAL ABUSIVOS CON INCAPAZ DE RESISTIR. ART. 210 C.P.",
##              "ACCESO CARNAL O ACTO SEXUAL EN PERSONA PUESTA EN INCAPACIDAD DE RESISTIR. ART. 207 C.P.",
##              "ACCESO CARNAL VIOLENTO. ART. 205 C.P.",
##              "ACTOS SEXUALES CON MENOR DE CATORCE ANOS. ART. 209 C.P.",
##              "ALTERACION DE RESULTADOS ELECTORALES ART. 394 C.P.",
##              "ALTERACION DESFIGURACION Y SUPLANTACION DE MARCAS DE GANADO. ART. 243 C.P.",
##              "ALZAMIENTO DE BIENES. ART. 253 C.P.",
##              "CONCUSION ART. 404 C.P.",
##              "CONSERVACION O FINANCIACION DE PLANTACIONES ART. 375 C.P.",
##              "CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 319-1 C.P.",
##              "DAÑO EN BIEN AJENO. ART. 265 C.P.",
##              "DAÑO EN MATERIA PRIMA, PRODUCTO AGROPECUARIO O INDUSTRIAL. ART. 304 C.P.",
##              "DESTINACION ILICITA DE MUEBLES O INMUEBLES ART. 377 C.P.",
##              "DISPARO DE ARMA DE FUEGO CONTRA VEHICULO ART. 356 C.P.",
##              "DIVULGACION Y EMPLEO DE DOCUMENTOS RESERVADOS. ART. 194 C.P.",
##              "EXPLOTACION ILICITA DE YACIMIENTO MINERO Y OTROS MATERIALES ART. 338 C.P.",
##              "FABRICACION Y COMERCIALIZACION DE SUSTANCIAS NOCIVAS PARA LA SALUD ART. 374 C.P.",
##              "FAVORECIMIENTO DE CONTRABANDO DE HIDROCARBUROS Y SUS DERIVADOS ART. 320-1 C.P.",
##              "FAVORECIMIENTO DEL DELITO DE CONTRABANDO DE HIDROCARBUROS O SUS DERIVADOS ART. 6 DE LA LEY 1712 DE 2014",
##              "FRAUDE AL SUFRAGANTE ART. 388 C.P.",
##              "FRAUDE EN LA INSCRIPCION DE CEDULAS ART. 389 C.P.",
##              "FUGA DE PRESOS ART. 448 C.P.",
##              "HOMICIDIO ART. 103 C.P.",
##              "HURTO CALIFICADO. ART. 240 C.P.",
##              "HURTO. ART. 239 C.P.",
##              "ILICITO APROVECHAMIENTO DE LOS RECURSOS NATURALES RENOVABLES ART. 328 C.P.",
##              "INCENDIO ART. 350 C.P.",
##              "INSTIGACION A DELINQUIR ART. 348 C.P.",

```

```

## "INSTIGACION A DELINQUIR ART. 348 C.P. INCISO 1",
## "INTERVENCION EN POLITICA ART. 422 C.P.",
## "INVASION DE AREAS DE ESPECIAL IMPORTANCIA ECOLOGICA ART. 337 C.P.",
## "INVASION DE TIERRAS O EDIFICACIONES. ART. 263 C.P.",
## "IRRESPETO A CADAVERES. ART. 204 C.P.",
## "LAVADO DE ACTIVOS ART. 323 C.P.",
## "MALTRATO MEDIANTE RESTRICCIÓN A AL LIBERTAD FISICA. ART. 230 C.P.",
## "OBSTRUCCION DE OBRAS DE DEFENSA Y ASISTENCIA ART. 364 C.P.",
## "OMISION DE APOYO ART. 424 C.P.",
## "OMISION DE SOCORRO ART. 131 C.P.",
## "PERFIDIA ART. 143 C.P.",
## "PERTURBACION DE ACTOS OFICIALES ART. 430 C.P. INCISO 1",
## "PERTURBACION DE CERTAMEN DEMOCRATICO ART. 386 C.P.",
## "PERTURBACION DE LA POSESION SOBRE INMUEBLE. ART. 264 C.P.",
## "PERTURBACION EN SERVICIO DE TRANSPORTE COLECTIVO U OFICIAL ART. 353 C.P.",
## "RECEPTACION ART. 447 C.P.",
## "RECEPTACION ART. 327 C C.P.",
## "REVELACION DE SECRETO ART. 418 C.P. INCISO 1",
## "SIMULACION DE INVESTIDURA O CARGO ART. 426 C.P.",
## "SINIESTRO O DAÑO DE NAVE ART. 354 C.P.",
## "TENENCIA, FABRICACION Y TRAFICO DE SUSTANCIA U OBJETOS PELIGROSOS ART. 358 C.P.",
## "TRAFICO DE SUSTANCIAS PARA PROCESAMIENTO DE NARCOTICOS ART. 382 C.P.",
## "TRAFICO FABRICACION O PORTE DE ESTUPEFACIENTES ART. 376 C..",
## "ULTRAJE A EMBLEMAS Y SIMBOLOS PATRIOS ART. 461 C.P.",
## "USO DE DOCUMENTO FALSO. ART. 291 C.P.",
## "USURPACION DE FUNCIONES PUBLICAS ART. 425 C.P.",
## "UTILIZACION ILICITA DE EQUIPOS TRANSMISORES O RECEPTORES. ART. 197 C.P.",
## "UTILIZACION INDEBIDA DE INFORMACION PRIVILEGIADA. ART. 258 C.P.")
##
## # Rules is conflict.
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## conflict_spoa <- data_spoa_jep_fase3 %>%
##   mutate(prob = runif(nrow(.))) %>%
##   mutate(is_conflict = case_when(
##     a02_tipo_persona == "PERPETRADOR" ~ zero,
##     perpetrador1 %in% perp_verde ~ one,
##     perpetrador2 %in% perp_verde ~ one,
##     perpetrador3 %in% perp_verde ~ one,
##     perpetrador1 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador2 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador3 %in% perp_amarillo & prob < 0.5 ~ one,
##     perpetrador1 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador2 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador3 %in% perp_naranja & prob < 0.75 ~ zero,
##     perpetrador1 %in% perp_rojo ~ zero,
##     perpetrador2 %in% perp_rojo ~ zero,
##     perpetrador3 %in% perp_rojo ~ zero,
##     perpetrador1 %in% perp_is_ca & a02_tipo_persona == "VICTIMA" ~ one,
##     perpetrador2 %in% perp_is_ca & a02_tipo_persona == "VICTIMA" ~ one,
##     perpetrador3 %in% perp_is_ca & a02_tipo_persona == "VICTIMA" ~ one,
##     a70_delito_a %in% delito_na ~ NA_integer_,

```

```

##   a70_delito_a %in% delito_verde ~ one,
##   a70_delito_a %in% delito_azul & prob < 0.75 ~ one,
##   a70_delito_a %in% delito_amarillo & prob < 0.5 ~ one,
##   a70_delito_a %in% delito_naranja & prob < 0.75 ~ zero,
##   a70_delito_a %in% delito_rojo ~ zero,
##   a35_titulo %in% titulo_verde ~ one,
##   a35_titulo %in% titulo_amarillo & prob < 0.5 ~ one,
##   a35_titulo %in% titulo_naranja & prob < 0.75 ~ zero,
##   a35_titulo %in% titulo_rojo ~ zero,
##   a02_tipo_persona == "OTRA" ~ NA_integer_,
##   TRUE ~ NA_integer_)) %>%
##   select(-prob)
##
## table(conflict_spoa$is_conflict, useNA = "always")
##
## # 0          1      <NA>
## # 1137614  923465  37592
##
## sample <- conflict_spoa %>%
##   filter(is_conflict ==0) %>%
##   select(recordid, is_conflict)
##
## # ----- training data
##
## sample <- sample_frac(sample, 0.2)
##
## write.table(sample, file = args$examples, sep = "|", quote = FALSE, row.names = FALSE)
##
##
## write_parquet(conflict_spoa, args$output)

```

```
cat(readLines(files$is_ca), sep = '\n')
```

#### 9.4.1.7. FGN - IS\_CA

```

## #
## # Authors:      Maria Ortiz
## # Maintainers  Maria Ortiz
## # Copyright    2020, HRDAG,
## # =====
## # CO-SIVJNRN-data/individual/FGN/is-ca/src/is-ca.R
##
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow,assertr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))

```

```
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                       default=here::here("individual/FGN/clean/output/fgn_lideres.parquet"))
## parser$add_argument("--output",
##                       default = "output/fgn_lideres.parquet")
##
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## data <- read_parquet(args$input)
##
## data <- data %>%
##   mutate(is_conflict = 1)
##
## glimpse(data)
##
## write_parquet(data, args$output)
##
## #done
```

```
cat(readLines(files$is_ca_desaparecido), sep = '\n')
```

#### 9.4.1.8. INML - Desaparecidos

```
## #
## # Authors:      JG
## # Maintainers  JG
## # Copyright    2020, HRDAG,
## # =====
## # CO-SIVJNRN-data/individual/INML/is-ca/src/is-ca-legacy.R
##
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr,assertr, dplyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                       default=here::here("individual/INML/clean/output/inml-desaparecidos.parquet"))
## parser$add_argument("--examples",
##                       default = here::here("individual/INML/is-ca/hand/examples-desaparecidos.csv"))
## parser$add_argument("--output",
```

```

##           default = "individual/INML/is-ca/output/inml-desaparecidos.parquet")
## args <- parser$parse_args()
##
## # ----- main
##
## inml_str <- read_parquet(args$input) %>%
##   mutate(perpetrator = organizacion_responsable,
##          tipo_hecho = clasificacion_de_la_desaparicion,
##          #contexto = circunstancia_de_los_hechos,
##          #mecanismo = causa_de_muerte_definitiva,
##          social_group = grupo_vulnerable)
##
## inml_str <- inml_str %>%
##   mutate_at(vars(tipo_hecho),
##             ~case_when(. == "SIN INFORMACION" ~ NA_character_,
##                       . == "PARA VERIFICACION DE IDENTIDAD" ~NA_character_,
##                       TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
##   mutate_at(vars(social_group),
##             ~case_when(. == "NINGUNA" ~ NA_character_,
##                       . == "OTROS" ~ NA_character_,
##                       . == "SIN INFORMACION" ~ NA_character_,
##                       . == "NO DILIGENCIADO" ~ NA_character_,
##                       TRUE ~ .)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
## perp_vars <- inml_str %>%
##   group_by(perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(n))
##
## conflict_perps <- c("AUTODEFENSAS INDEPENDIENTES", "AUTODEFENSAS UNIDAS DE COLOMBIA",
##                   "DISIDENCIAS DE LOS PRINCIPALES GRUPOS SUBVERSIVOS", "EJERCITO DE LIBERACION NAC
##                   "FUERZAS ARMADAS REVOLUCIONARIAS DE COLOMBIA", "EJERCITO NACIONAL",
##                   "POLICIA NACIONAL")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(
##     perpetrator %in% c("EJERCITO NACIONAL", "POLICIA NACIONAL") & is.na(tipo_hecho) ~ NA_integer_,
##     perpetrator %in% conflict_perps ~ as.integer(1),
##     is.na(perpetrator) & tipo_hecho == "DESAPARICION PRESUNTAMENTE FORZADA" ~ as.integer(1),
##     TRUE ~ NA_integer_))
##
## is_conflict_dist <- inml_str %>%
##   group_by(is_conflict, perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## # there is no records with is_conflict = 0
##
## write_parquet(inml_str, args$output)

```

```
cat(readLines(files$is_ca_fatales), sep = '\n')
```

#### 9.4.1.9. INML - Fatales

```
## #
## # Authors:      JG
## # Maintainers  JG
## # Copyright    2020, HRDAG,
## # =====
## # CO-SIVJRN-data/individual/INML/is-ca/src/is-ca-legacy.R
## #
## # ----- setup
## #
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr, assertr, dplyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                      default=here::here("individual/INML/clean/output/inml-fatales.parquet"))
## parser$add_argument("--examples",
##                      default = here::here("individual/INML/is-ca/hand/examples-fatales.csv"))
## parser$add_argument("--output",
##                      default = "individual/INML/is-ca/output/inml-fatales.parquet")
## args <- parser$parse_args()
##
## # ----- function
## #
## # ----- main
## #
## inml_str <- read_parquet(args$input) %>%
##   mutate(perpetrator = presunto_agresor,
##          tipo_hecho = manera_de_muerte_definitiva,
##          contexto = circunstancia_de_los_hechos,
##          mecanismo = causa_de_muerte_definitiva,
##          social_group = condicion_de_vulnerabilidad)
##
## inml_str <- inml_str %>%
##   mutate_at(vars(tipo_hecho),
##             ~case_when(. == "INDETERMINADA" ~ NA_character_,
##                       . == "EN ESTUDIO" ~ NA_character_,
##                       . == "SIN INFORMACION" ~ NA_character_,
##                       . == "POR DETERMINAR" ~ NA_character_,
##                       . == "NS/NR-SININFORMACION" ~ NA_character_,
##                       . == "POR DETERMINAR" ~ NA_character_,
##                       . == "VIOLENTA - SIN DETERMINAR" ~ NA_character_,
##                       TRUE ~.)) %>%
##   verify(nrow(.) == nrow(inml_str)) %>%
```



```

## mutate_at(vars(contexto),
##           ~case_when(. == "NS/NR - SIN INFORMACION" ~ NA_character_,
##                       . == "NO APLICA" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str)) %>%
## mutate_at(vars(mecanismo),
##           ~case_when(. == "NS/NR - SININFORMACION" ~ NA_character_,
##                       . == "INDETERMINADA" ~ NA_character_,
##                       . == "EN ESTUDIO" ~ NA_character_,
##                       . == "SIN INFORMACION" ~ NA_character_,
##                       . == "POR DETERMINAR" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str)) %>%
## mutate_at(vars(social_group),
##           ~case_when(. == "NINGUNA" ~ NA_character_,
##                       . == "OTROS" ~ NA_character_,
##                       . == "SIN INFORMACION" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str))
##
## perp_vars <- inml_str %>%
##   group_by(perpetrator, contexto, tipo_hecho, mecanismo) %>%
##   count() %>%
##   arrange(desc(n))
##
## # validando categorias por hecho y grupo social para perpetradores relacionados con fuerza pública
## ffmm_perps <- c("SERVICIOS DE INTELIGENCIA", "FUERZAS MILITARES", "POLICIA",
##               "FUERZAS IRREGULARES")
##
## ffmm_social <- inml_str %>%
##   filter(perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA")) %>%
##   group_by(perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(n))
##
## perps_is_ca <- c("SERVICIOS DE INTELIGENCIA", "FARC",
##                 "OTRAS GUERRILLAS", "PARAMILITARES - AUTODEFENSAS", "ELN",
##                 "FUERZAS MILITARES", "POLICIA", "FUERZAS IRREGULARES")
##
## perps_not_ca <- c('MADRASTA', 'ABUELO (A)', 'HERMANO (A)', 'VECINO', 'AMIGO',
##                  'COMPANERO (A) PERMANENTE', 'EX ESPOSO (A)', 'COMPANERO (A) DE TRABAJO',
##                  'EX-NOVIO (A)', 'MADRE', 'DELINCUENCIA COMUN',
##                  'CLIENTE', 'PADRE', 'PANDILLAS', 'OTROS FAMILIARES CIVILES O CONSANGUINEOS',
##                  'AMANTE', 'ARRENDADOR', 'ARRENDATARIO', 'CONOCIDO SIN NINGUN TRATO',
##                  'CUNADO (A)', 'EMPLEADO (A)', 'EMPLEADOR', 'ENCARGADO DEL MENOR', 'ESPOSO (A)',
##                  'EX-AMANTE', 'HIJO (A)', 'MADRASTRA', 'NOVIO (A)',
##                  'PADRASTRO', 'PRIMO (A)', 'PROFESOR (A)', 'PROVEEDOR', 'SUEGRO (A)', 'TIO (A)')
##
## contexto_is_ca <- c("ACCION GUERRILLERA", "ACCION PARAMILITAR", "CONFLICTO ARMADO",
##                    "ENFRENTAMIENTO ARMADO", "ASESINATO POLITICO", "CONFLICTO POLITICO",
##                    "ACCION MILITAR", "SECUESTRO", "TERRORISMO")
##
## contexto_not_ca <- c("RINA", "ATRACO CALLEJERO", "MALTRATO DE PAREJA",
##                     "MALTRATO ENTRE OTROS FAMILIARES",

```

```

##           "ROBO ENTIDAD BANCARIA O COMERCIAL", "ROBO RESIDENCIA - MORADA", "ROBO VEHICULO
##           "VIOLENCIA CONTRA GRUPOS DESCALIFICADOS O MARGINALES",
##           "RESPONSABILIDAD MEDICA / ENFERMERIA", "RESPONSABILIDAD ODONTOLOGICA")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(
##     # validado con el campo social_group: varios registros son posteriormente asignados como
##     # NA cuando no hay grupo social identificado
##     contexto == "ACCIONMILITAR" & is.na(social_group) ~ NA_integer_,
##     contexto == "ENFRENTAMIENTOARMADO" & is.na(social_group) ~ NA_integer_,
##     contexto %in% contexto_is_ca ~ as.integer(1),
##     contexto %in% contexto_not_ca ~ as.integer(0),
##     contexto == "VENGANZA-AJUSTEDECUENTAS" &
##       str_detect(social_group, "(GREMIAL|CAMPEÑO|DESMOVILIZADOS|FUNCIONARIOS JUDICIALES|GRUPOS ARM
##     contexto == "VENGANZA-AJUSTEDECUENTAS" ~ as.integer(0),
##     contexto == "VENGANZA-AJUSTEDECUENTAS" & is.na(social_group) ~ NA_integer_,
##     tipo_hecho == "NATURAL" ~ as.integer(0),
##     tipo_hecho == "VIOLENTA-ACCIDENTAL" | contexto == "AUTOLESIONINVOLUNTARIA(ACCIDENTE)" ~ as.integ
##     tipo_hecho == "VIOLENTA-SUICIDIO" | contexto == "AUTOLESIONINVOLUNTARIA(SUICIDIO)" ~ as.integer(0)
##     tipo_hecho == "VIOLENTA-TRANSITO" ~ as.integer(0),
##     perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA") & is.na(social_group) ~ NA_integer
##     perpetrator %in% perps_is_ca ~ as.integer(1),
##     perpetrator %in% perps_not_ca ~ as.integer(0),
##     perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" ~ as.integer(0),
##     perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" & social_group == "PRESUNTO COLABORADOR GRUPO ILEGAL
##     perpetrator %in% c('NARCOTRAFICANTES', 'GRUPOS DE SEGURIDAD PRIVADA') ~ NA_integer_,
##     TRUE ~ NA_integer_)) %>%
##   verify(nrow(.) == nrow(inml_str))
##
## conflict_perps <- c("ELN", "FARC", "OTRAS GUERRILLAS", "PARAMILITARES - AUTODEFENSAS")
##
## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(is_conflict == 0 & perpetrator %in% conflict_perps ~ as.integer(1),
##     TRUE ~ is_conflict))
##
## is_conflict_dist <- inml_str %>%
##   group_by(is_conflict, perpetrator, contexto, social_group) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## inconsistent_recs <- inml_str %>%
##   filter(is_conflict == 0 & perpetrator %in% conflict_perps) %>%
##   verify(nrow(.) == 0)
##
## sample_no_conflict <- inml_str %>%
##   filter(is_conflict == 0) %>%
##   group_by(contexto) %>%
##   slice_sample(n = 100) %>%
##   ungroup() %>%
##   select(recordid, is_conflict) %>%
##   distinct()
##
## write_delim(sample_no_conflict, args$examples, delim = "|")
## write_parquet(inml_str, args$output)

```

```
## #done
```

```
cat(readLines(files$is_ca_legacy), sep = '\n')
```

#### 9.4.1.10. INML - Legacy

```
## #
## # Authors:      JG
## # Maintainers  JG
##
## # Copyright    2020, HRDAG,
## # =====
## # CO-SIVJNRN-data/individual/INML/is-ca/src/is-ca-legacy.R
##
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr,assertr, dplyr, tidyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(),"is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                      default = here::here("individual/INML/clean/output/inml-legacy.parquet"))
## parser$add_argument("--examples",
##                      default = here::here("individual/INML/is-ca/hand/examples-legacy.csv"))
## parser$add_argument("--output",
##                      default = here::here("individual/INML/is-ca/output/inml-legacy.parquet"))
## args <- parser$parse_args()
##
## inml <- read_parquet(args$input) %>%
##   mutate(N_str = str_count(violation, ",") +1)
##
## data_viol_0 <- filter(inml, is.na(N_str))
## data_viol_1 <- filter(inml, N_str ==1) # tipo_hecho
## data_viol_2 <- filter(inml, N_str ==2) # tipo_hecho, mecanismo
## data_viol_3 <- filter(inml, N_str ==3) # contexto, tipo_hecho, mecanismo
##
## data_viol_0 <- data_viol_0 %>%
##   mutate(contexto = NA_character_,
##          tipo_hecho = NA_character_,
##          mecanismo = NA_character_)
##
## data_viol_1 <- data_viol_1 %>%
##   mutate(contexto = NA_character_,
##          tipo_hecho = violation,
##          mecanismo = NA_character_)
##
## data_viol_2 <- data_viol_2 %>%
```

```

## mutate(contexto = NA_character_) %>%
## separate(violation, c("tipo_hecho", "mecanismo"), sep = ",", remove = FALSE)
##
## data_viol_3 <- data_viol_3 %>%
## separate(violation, c("contexto", "tipo_hecho", "mecanismo"), sep = ",", remove = FALSE)
##
## inml_str <- bind_rows(list(data_viol_0, data_viol_1, data_viol_2, data_viol_3)) %>%
## verify(nrow(.) == nrow(inml))
##
## inml_str <- inml_str %>%
## mutate_at(vars(tipo_hecho),
##           ~case_when(. == "INDETERMINADA" ~ NA_character_,
##                       . == "ENESTUDIO" ~ NA_character_,
##                       . == "SININFORMACION" ~ NA_character_,
##                       . == "POREDETERMINAR" ~ NA_character_,
##                       . == "NS/NR-SININFORMACION" ~ NA_character_,
##                       . == "PORESTABLECER" ~ NA_character_,
##                       . == "VIOLENTA-SINDETERMINAR" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str)) %>%
## mutate_at(vars(contexto),
##           ~case_when(. == "NS/NR-SININFORMACION" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str)) %>%
## mutate_at(vars(mecanismo),
##           ~case_when(. == "NS/NR-SININFORMACION" ~ NA_character_,
##                       . == "INDETERMINADA" ~ NA_character_,
##                       . == "ENESTUDIO" ~ NA_character_,
##                       . == "SININFORMACION" ~ NA_character_,
##                       . == "12" ~ NA_character_,
##                       . == "PORESTABLECER" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str)) %>%
## mutate_at(vars(social_group),
##           ~case_when(. == "NINGUNA" ~ NA_character_,
##                       . == "OTROS" ~ NA_character_,
##                       . == "SIN INFORMACION" ~ NA_character_,
##                       TRUE ~ .)) %>%
## verify(nrow(.) == nrow(inml_str))
##
##
## N_suicidio <- inml_str %>%
## filter(violation == "VIOLENTA-SUICIDIO" | contexto == "AUTOLESIONVOLUNTARIA(SUICIDIO)") %>%
## nrow(.)
##
## N_accidental <- inml_str %>%
## filter(violation == "VIOLENTA-ACCIDENTAL" | contexto == "AUTOLESIONINVOLUNTARIA(ACCIDENTE)") %>%
## nrow(.)
##
## N_transito <- filter(inml, violation == "VIOLENTA-TRANSITO") %>% nrow(.)
##
## perp_vars <- inml_str %>%
## group_by(perpetrator, contexto, tipo_hecho, mecanismo) %>%
## count() %>%

```

```

## arrange(desc(n))
##
## perp_contexto <- inml_str %>%
##   group_by(perpetrator, contexto) %>%
##   count() %>%
##   arrange(desc(n))
##
## perp_tipo_hecho <- inml_str %>%
##   group_by(perpetrator, tipo_hecho) %>%
##   count() %>%
##   arrange(desc(n))
##
## perp_mecanismo <- inml_str %>%
##   group_by(perpetrator, mecanismo) %>%
##   count() %>%
##   arrange(desc(n))
##
## # validando categorias por hecho y grupo social para perpetradores relacionados con fuerza pública
## ffmm_perps <- c("SERVICIOS DE INTELIGENCIA", "FUERZAS MILITARES", "POLICIA",
##               "FUERZAS IRREGULARES")
##
## ffmm_social <- inml_str %>%
##   filter(perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA")) %>%
##   group_by(perpetrator, tipo_hecho, social_group) %>%
##   count() %>%
##   arrange(desc(n))
##
## perps_is_ca <- c("SERVICIOS DE INTELIGENCIA", "FARC",
##                "OTRAS GUERRILLAS", "PARAMILITARES - AUTODEFENSAS", "ELN",
##                "FUERZAS MILITARES", "POLICIA", "FUERZAS IRREGULARES")
##
## perps_not_ca <- c('MADRASTA', 'ABUELO (A)', 'HERMANO (A)', 'VECINO', 'AMIGO',
##                  'COMPANERO (A) PERMANENTE', 'EX ESPOSO (A)', 'COMPANERO (A) DE TRABAJO',
##                  'EX-NOVIO (A)', 'MADRE', 'DELINCUENCIA COMUN',
##                  'CLIENTE', 'PADRE', 'PANDILLAS', 'OTROS FAMILIARES CIVILES O CONSANGUINEOS',
##                  'AMANTE', 'ARRENDADOR', 'ARRENDATARIO', 'CONOCIDO SIN NINGUN TRATO',
##                  'CUNADO (A)', 'EMPLEADO (A)', 'EMPLEADOR', 'ENCARGADO DEL MENOR', 'ESPOSO (A)',
##                  'EX-AMANTE', 'HIJO (A)', 'MADRASTRA', 'NOVIO (A)',
##                  'PADRASTRO', 'PRIMO (A)', 'PROFESOR (A)', 'PROVEEDOR', 'SUEGRO (A)', 'TIO (A)')
##
## contexto_is_ca <- c("ACCIONGUERRILLERA", "ACCIONPARAMILITAR", "CONFLICTOARMADO",
##                   "ENFRENTAMIENTOARMADO", "ASESINATOPOLITICO", "CONFLICTOPOLITICO",
##                   "ACCIONMILITAR", "SECUESTRO", "TERRORISMO")
##
## contexto_not_ca <- c("RINA", "ACCIDENTEDETRABAJO", "ACCIDENTEDETRANSPORTE", "AGRESIONPORANIMALES",
##                    "ATRACOCALLEJERO", "AUTOLESIONINVOLUNTARIA (ACCIDENTE)", "AUTOLESIONVOLUNTARIA (S
##                    "DESASTRENATURAL", "DESASTRENONATURALOACC.MASIVO", "ECONOMICO",
##                    "EMBRIAGUEZ (ALCOHOLICAYNOALCOHOLICA)", "EXCESOVELOCIDAD", "MALTRATODEPAREJA",
##                    "MALTRATOENTREOTROSFAMILIARES", "POSIBLESFALLASMECANICAS",
##                    "ROBOENTIDADBANCARIAOCOMERCIAL", "ROBORESIDENCIA-MORADA", "ROBOVEHICULO",
##                    "VIOLACIONNORMASDETRANSITOPEATONES", "VIOLACIONOTRASNORMASDETRANSITO",
##                    "VIOLENCIACONTRAGRUPOSDESCALIFICADOSOMARGINALES",
##                    "RESPONSABILIDADMEDICA/ENFERMERIA", "RESPONSABILIDADODONTOLOGICA")
##
##

```

```

## inml_str <- inml_str %>%
##   mutate(is_conflict = case_when(
##     # validado con el campo social_group: varios registros son posteriormente asignados como
##     # NA cuando no hay grupo social identificado
##     contexto == "ACCIONMILITAR" & is.na(social_group) ~ NA_integer_,
##     contexto == "ENFRENTAMIENTOARMADO" & is.na(social_group) ~ NA_integer_,
##     contexto %in% contexto_is_ca ~ as.integer(1),
##     contexto %in% contexto_not_ca ~ as.integer(0),
##     contexto == "VENGANZA-AJUSTEDECUENTAS" &
##       str_detect(social_group, "(GREMIAL|CAMPELINO|DESMOVLIZADOS|FUNCIONARIOS JUDICIALES|GRUPOS AR
##     contexto == "VENGANZA-AJUSTEDECUENTAS" ~ as.integer(0),
##     contexto == "VENGANZA-AJUSTEDECUENTAS" & is.na(social_group) ~ NA_integer_,
##     tipo_hecho == "NATURAL" ~ as.integer(0),
##     tipo_hecho == "VIOLENTA-ACCIDENTAL" | contexto == "AUTOLESIONINVOLUNTARIA(ACCIDENTE)" ~ as.inte
##     tipo_hecho == "VIOLENTA-SUICIDIO" | contexto == "AUTOLESIONVOLUNTARIA(SUICIDIO)" ~ as.integer(0
##     tipo_hecho == "VIOLENTA-TRANSITO" ~ as.integer(0),
##     perpetrator %in% c(ffmm_perps, "GRUPOS DE SEGURIDAD PRIVADA") & is.na(social_group) ~ NA_intege
##     perpetrator %in% perps_is_ca ~ as.integer(1),
##     perpetrator %in% perps_not_ca ~ as.integer(0),
##     perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" ~ as.integer(0),
##     perpetrator == "GRUPOS DE SEGURIDAD PRIVADA" & social_group == "PRESUNTO COLABORADOR GRUPO ILEG
##     perpetrator %in% c('NARCOTRAFICANTES', 'GRUPOS DE SEGURIDAD PRIVADA') ~ NA_integer_,
##     TRUE ~ NA_integer_) %>%
##   verify(nrow(.) == nrow(inml_str))
##
##
## is_conflict_dist <- inml_str %>%
##   group_by(is_conflict, perpetrator, contexto, social_group, tipo_hecho) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## sample_no_conflict <- inml_str %>%
##   filter(is_conflict == 0) %>%
##   group_by(contexto) %>%
##   slice_sample(n = 100) %>%
##   ungroup() %>%
##   select(recordid, is_conflict) %>%
##   distinct()
##
## write_delim(sample_no_conflict, args$examples, delim = "|")
## write_parquet(inml_str, args$output)
##
## #done

```

```
cat(readLines(files$is_ca_inmlcs), sep = '\n')
```

#### 9.4.1.11. INML - Es conflicto

```

## #
## # Authors:      JG
## # Maintainers  JG

```

```

## # Copyright 2020, HRDAG,
## # =====
## # CO-SIVJNR-data/individual/INMLCF/is-ca/src/is-ca.R
##
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, here, arrow, readr, assertr, dplyr, stringr, lubridate)
##
## stopifnot(endsWith(getwd(), "is-ca"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                      default=here::here("individual/INMLCF/is-forced-dis/output/inmlcf-des.parquet"))
## parser$add_argument("--examples",
##                      default = here::here("individual/INML/is-ca/hand/examples-desaparicion.csv"))
## parser$add_argument("--output",
##                      default = "output/inmlcf.parquet")
## args <- parser$parse_args()
##
## #--- funciones
##
## # ----- main
##
## inmlcf <- read_parquet(args$input)
##
## inmlcf <- inmlcf %>%
##   mutate(is_conflict = case_when(
##     # Related to conflict
##     ancestro_racial %in% c("INDIGENA", "NEGRO") ~ as.integer(1),
##     pertenencia_grupal == "GRUPOS ETNICOS" & is_forced_dis == 1 ~ as.integer(1),
##     pertenencia_grupal == "DESPLAZADOS" ~ as.integer(1),
##     pertenencia_grupal == "LIDERES" ~ as.integer(1),
##     pertenencia_grupal == "CAMPEsinOS" ~ as.integer(1),
##     pertenencia_grupal == "SERVIDOR PUBLICO" ~ as.integer(1),
##     pertenencia_grupal == "SINDICALISTA" ~ as.integer(1),
##     pertenencia_grupal == "PERIODISTA" ~ as.integer(1),
##     pertenencia_grupal == "POLITICO - DIRIGENTE POLITICO" ~ as.integer(1),
##     pertenencia_grupal == "MISION MEDICA - TRABAJADORES SALUD" ~ as.integer(1),
##     pertenencia_grupal == "HERIDO Y/O ENFERMO BAJO PROTECCION SANITARIA O MEDICA" ~ as.integer(1),
##     pertenencia_grupal == "RELIGIOSOS" & is_forced_dis == 1 ~as.integer(1),
##     pertenencia_grupal == "DESMOVILIZADOS" & is_forced_dis == 1 ~as.integer(1),
##     pertenencia_grupal == "ADICTOS" & is_forced_dis == 1 ~as.integer(1),
##     pertenencia_grupal == "EDUCADORES" & is_forced_dis == 1 ~as.integer(1),
##     clasificacion_desaparicion == "SECUESTRO" ~ as.integer(1),
##     clasificacion_desaparicion == "RECLUTAMIENTO" ~ as.integer(1),
##     (clasificacion_desaparicion == "DESAPARICION PRESUNTAMENTE FORZADA" | is_forced_dis ==1) & is.na(
## #pertenencia_grupal == "TRABAJADORA SEXUAL" & clasificacion_desaparicion == "DESAPARICION PRESUN
## perpetrator %in% c("GUERRILLAS", "ESTADO", "PARAMILITARES", "BANDAS EMERGENTES") ~ as.integer(1)
##     # not related to conflict
##     clasificacion_desaparicion == "DESASTRE NATURAL" ~ as.integer(0),

```

```

## pertenencia_grupal == "TRABAJADORA SEXUAL" & is.na(clasificacion_desaparicion) ~ as.integer(0),
## clasificacion_desaparicion == "TRATA DE PERSONAS" ~ as.integer(0),
## perpetrator %in% c('PANDILLAS') ~ as.integer(0),
## perpetrator %in% c('FAMILIARES') ~ as.integer(0),
## pertenencia_grupal == "RELCUSO" ~ NA_integer_,
## TRUE ~ NA_integer_)
##
## is_conflict_dist <- inmlcf %>%
##   group_by(is_conflict, perpetrator) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
##
## is_conflict_desap <- inmlcf %>%
##   group_by(is_conflict, perpetrator, clasificacion_desaparicion) %>%
##   count() %>%
##   arrange(desc(is_conflict, n))
##
## sample_no_conflict <- inmlcf %>%
##   filter(is_conflict == 0) %>%
##   group_by(perpetrator) %>%
##   slice_sample(n = 50) %>%
##   ungroup() %>%
##   select(recordid, is_conflict) %>%
##   distinct()
##
## write_delim(sample_no_conflict, args$examples, delim = "|")
## write_parquet(inmlcf, args$output)
##
## #done

```

#### 9.4.2. Códigos y Excel utilizados para la variable “es desaparición forzada”

```
cat(readLines(files$is_forced_dis_fgn), sep = '\n')
```

##### 9.4.2.1. FGN - Es desaparición forzada

```

## #
## # Authors:      Maria Ortiz
## # Maintainers  Maria Ortiz
## # Copyright    2021, HRDAG,
## # =====
## # CO-SIVJNRN-data/individual/FGN/is-forced-dis/src/is-forced-dis.R
##
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, dplyr, here, arrow,assertr, stringr, lubridate)

```



```

##
## stopifnot(endsWith(getwd(),"is-forced-dis"))
##
## parser <- ArgumentParser()
## parser$add_argument("--input",
##                       default=here::here("individual/FGN/clean/output/vinculacion.parquet"))
## parser$add_argument("--output",
##                       default = "output/fgn.parquet")
##
## args <- parser$parse_args()
##
## # ----- function
##
## # ----- main
##
## data <- read_parquet(args$input)
##
## if (str_detect(args$input, "hv_bloque_oriental")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(tipo_hecho == "DESAPARICION FORZADA" ~ as.integer(1),
##                                     TRUE ~ as.integer(0)))
## } else if (str_detect(args$input, "lideres")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(grupo_delito == "DESAPARICION FORZADA" ~ as.integer(1),
##                                     TRUE ~ as.integer(0)))
## } else if (str_detect(args$input, "jep_consolidado")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(hecho == "DESAPARICION FORZADA" ~ as.integer(1),
##                                     TRUE ~ as.integer(0)))
## } else if (str_detect(args$input, "up-fgn-transicional")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(afectacion_1_adj == "DESAPARICION FORZADA" ~ as.integer(1),
##                                     TRUE ~ as.integer(0)))
## } else if (str_detect(args$input, "vinculacion")){
##   data <- data %>%
##     mutate(is_forced_dis = case_when(hecho == "DESAPARICION FORZADA" ~ as.integer(1),
##                                     TRUE ~ as.integer(0)))
## }
##
## glimpse(data)
## table(data$is_forced_dis, useNA = "always")
##
## write_parquet(data, args$output)
##
## #done

```

```
cat(readLines(files$is_forced_dis_fgn_sijuf_cv), sep = '\n')
```

## 9.4.2.2. FGN - SIJUF - CEV

```

## # -----
## # Authors:      PA
## # Maintainers  PA, VG, PB
## # Copyright    2022, HRDAG, GPL-2 or newer
## # -----
## # CO-SIVJRN-data/individual/FGN/is-forced-dis/src/is-forced-dis-FISCALIA_datos_sijuf_cv.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## pacman::p_load(argparse, readxl, dplyr, here, arrow, assertr,
##                 stringr, stringi, lubridate, logger)
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default="./clean/output/FISCALIA_datos_sijuf_cv.parquet")
## parser$add_argument("--dane", default=here("share/DANE/output/dipo.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i-
## parser$add_argument("--output", default = "output/FISCALIA_datos_sijuf_cv.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ----- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
##   rename(depto_hechos = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(grupo_delito == "DESAPARICION FORZADA") %>%
##   left_join(probs, by = c("depto_hechos", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data initially read")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION", "INVESTIGACION", "JUICIO", "EJECUCION DE PENAS", "ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE", "INDAGACION", "TERMINACION ANTICIPADA", "ETAPA DE INVESTIGACION PRELIM
## hecho_no_desp <- c("HOMICIDIO DOLOSO", "DESPLAZAMIENTO", "RECLUTAMIENTO ILICITO", "FEMINICIDIO")
##

```

```
## log_info("starting big case_when")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     etapa %in% etapa_caso_des ~ one,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero,
##     TRUE ~ NA_integer_)) %>%
##   select(-prob_forced_dis, -prob, -direction) %>%
##   write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## #done
```

```
cat(readLines(files$is_forced_dis_fgn_spoa_cv), sep = '\n')
```

#### 9.4.2.3. FGN - SPOA - CEV

```
## # -----
## # Authors:      PA
## # Maintainers  PA, VG, PB
## # Copyright    2022, HRDAG, GPL-2 or newer
## # -----
## # CO-SIVJNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-FISCALIA_datos_spoa_cv.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## pacman::p_load(argparse, readxl, dplyr, here, arrow, assertr,
##               stringi, stringr, lubridate, logger)
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here("individual/FGN/clean/output/FISCALIA_datos_spoa_cv.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-is-forced-dis-probs.csv"))
## parser$add_argument("--output", default = "output/FISCALIA_datos_spoa_cv.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ----- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
##   rename(dpto_hechos = departamento_desaparicion) %>%
```

```

## mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
## mutate(prob = as.numeric(prob)) %>%
## mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
## rename(prob_forced_dis = prob) %>%
## mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
## filter(grupo_delito == "DESAPARICION FORZADA") %>%
## left_join(probs, by=c("dpto_hechos", "yy_hecho")) %>%
## mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION", "INVESTIGACION", "JUICIO", "EJECUCION DE PENAS", "ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE", "INDAGACION", "TERMINACION ANTICIPADA", "ETAPA DE INVESTIGACION PRELIM
## hecho_no_desp <- c("HOMICIDIO DOLOSO", "DESPLAZAMIENTO", "RECLUTAMIENTO Ilicito", "FEMINICIDIO")
##
## log_info("beginning case_when")
## data <- data %>%
## mutate(is_forced_dis = case_when(
##   etapa_caso %in% etapa_caso_des ~ one,
##   (direction == 1) & (prob_forced_dis > prob) ~ one,
##   (direction == 0) & (prob_forced_dis > prob) ~ zero,
##   TRUE ~ NA_integer_)) %>%
## select(-prob_forced_dis, -prob, -direction) %>%
## write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## #done

```

```
cat(readLines(files$is_forced_dis_sijuf_jep), sep = '\n')
```

#### 9.4.2.4. FGN - SIJUF - JEP

```

## # -----
## # Authors:      PA
## # Maintainers  PA, VG, PB
## # Copyright    2022, HRDAG, GPL-2 or newer
## # -----
## # CO-SIVJNRN-data/individual/FGN/is-forced-dis/src/is-forced-dis-sijuf-jep-fase4.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ----- setup
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##

```

```

## pacman::p_load(argparse, readxl, dplyr, here, arrow, assertr,
##               stringi, stringr, lubridate, logger)
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here("individual/FGN/clean/output/sijuf_jep_fase4.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i
## parser$add_argument("--output", default = "output/sijuf_jep_fase4.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ----- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
##   rename(a18_depto_hechos = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(str_detect(a23_delito_ppal, "DESAPARICION FORZADA")) %>%
##   left_join(probs, by = c("a18_depto_hechos", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read initially")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION", "INVESTIGACION", "JUICIO", "EJECUCION DE PENAS", "ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE", "INDAGACION", "TERMINACION ANTICIPADA", "ETAPA DE INVESTIGACION PRELIM
##
## log_info("beginning case_when")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     a10_etapa %in% etapa_caso_des ~ one,
##     (direction == 1) & (prob_forced_dis > prob) ~ one,
##     (direction == 0) & (prob_forced_dis > prob) ~ zero)) %>%
##   select(recordid, is_forced_dis) %>%
##   write_parquet(args$output)
## log_info("made is_forced_dis, write complete")
## print(table(data$is_forced_dis, useNA="always"))
##
## #done

cat(readLines(files$is_forced_dis_spoa_jep3), sep = '\n')

```

#### 9.4.2.5. FGN - SPOA - JEP FASE 3

```
## # -----
```

```

## # Authors:      PA
## # Maintainers  PA, VG, PB
## # Copyright    2022, HRDAG,
## # -----
## # CO-SIVJNRN-data/individual/FGN/is-forced-dis/src/is-forced-dis-spoa-jep-fase3.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, readxl, dplyr, here, arrow, assertr,
##         stringr, stringi, lubridate, logger)
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default="./clean/output/spoa_jep_fase3.parquet")
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i
## parser$add_argument("--output", default = "output/spoa_jep_fase3.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ----- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
##   rename(a43_departamento = departamento_desaparicion) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
## log_info("probs read")
##
## data <- read_parquet(args$input) %>%
##   filter(str_detect(a39_descripcion, "DESAPARICION FORZADA")) %>%
##   left_join(probs, by = c("a43_departamento", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read initially")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION", "INVESTIGACION", "JUICIO", "EJECUCION DE PENAS", "ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE", "INDAGACION", "TERMINACION ANTICIPADA", "ETAPA DE INVESTIGACION PRELIM
##
## log_info("case_when begins")
## data <- data %>%
##   mutate(is_forced_dis = case_when(
##     a24_etapa %in% etapa_caso_des ~ one,

```

```
## (direction == 1) & (prob_forced_dis > prob) ~ one,
## (direction == 0) & (prob_forced_dis > prob) ~ zero,
## TRUE ~ NA_integer_) %>%
## select(-prob_forced_dis, -prob, -direction) %>%
## write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## #done
```

```
cat(readLines(files$is_forced_dis_spoa_jep), sep = '\n')
```

#### 9.4.2.6. FGN - SPOA - JEP

```
## # -----
## # Authors: PA
## # Maintainers PA, VG, PB
## # Copyright 2022, HRDAG,
## # -----
## # CO-SIVJNR-data/individual/FGN/is-forced-dis/src/is-forced-dis-spoa-jep.R
##
## # note that this script outputs only the recs that are disappearances.
## # the output should not be used for anything except to feed the is-forced-dis
## # model.
## # ----- setup
##
## Sys.setlocale("LC_CTYPE", "en_US.UTF-8")
## set.seed(19481210)
##
## require(pacman)
## p_load(argparse, readxl, dplyr, here, arrow, assertr,
## stringr, stringi, logger, lubridate)
##
##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here("individual/FGN/clean/output/spoa_jep.parquet"))
## parser$add_argument("--dept_anio_probs", default=here("individual/INMLCF/is-forced-dis/hand/inmlcf-i
## parser$add_argument("--output", default = "output/spoa_jep.parquet")
## args <- parser$parse_args()
##
## one <- as.integer(1)
## zero <- as.integer(0)
##
## # ----- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
## mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
## rename(departamento = departamento_desaparicion) %>%
## mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
## mutate(prob = as.numeric(prob)) %>%
```

```

## mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
## rename(prob_forced_dis = prob) %>%
## mutate(yy_hecho = year)
## log_info("probs read")
##
## # ----- main
##
## data <- read_parquet(args$input) %>%
## filter(str_detect(descripcion, "DESAPARICION FORZADA")) %>%
## left_join(probs, by = c("departamento", "yy_hecho")) %>%
## mutate(prob = runif(nrow(.)))
## rm(probs)
## log_info("data read initially")
##
## etapa_caso_des <- c("ETAPA DE INSTRUCCION", "INVESTIGACION", "JUICIO", "EJECUCION DE PENAS", "ETAPA JUIC
## etapa_na_des <- c("QUERELLABLE", "INDAGACION", "TERMINACION ANTICIPADA", "ETAPA DE INVESTIGACION PRELIM
##
## data <- data %>%
## mutate(is_forced_dis = case_when(
##   etapa %in% etapa_caso_des ~ as.integer(1),
##   (direction == 1) & (prob_forced_dis > prob) ~ one,
##   (direction == 0) & (prob_forced_dis > prob) ~ zero,
##   TRUE ~ NA_integer_)) %>%
## select(-prob_forced_dis, -prob, -direction) %>%
## write_parquet(args$output)
##
## print(table(data$is_forced_dis, useNA="always"))
## log_info("done.")
##
## # done

```

```
cat(readLines(files$is_forced_dis_inml), sep = '\n')
```

#### 9.4.2.7. INML - Es desaparición forzada

```

## #
## # Authors:      Valentina Gómez
## # Maintainers  Valentina Gómez, Paula Amado, PB
## # Copyright    2022, HRDAG, GPL-2 or better
## # =====
## # CO-SIVJNRN-data/individual/INMLCF/is-forced-dis/src/is-forced-dis.R
##
## # ----- setup
## pacman::p_load(argparse, dplyr, here, tidyr,
##               arrow, assertr, stringr, stringi,
##               lubridate, readr, readxl, yaml)
##
## stopifnot(str_detect(Sys.getlocale(), "LC_CTYPE=en.US.UTF-8"))
## set.seed(19481210)
## one <- as.integer(1)
## zero <- as.integer(0)

```



```

##
## parser <- ArgumentParser()
## parser$add_argument("--input", default=here::here("individual/INMLCF/clean/output/inmlcf-des.parquet")
## parser$add_argument("--dept_anio_probs", default="hand/inmlcf-is-forced-dis-prob.xlsx")
## parser$add_argument("--dane", default=here::here("share/DANE/output/dipo.parquet"))
## parser$add_argument("--output", default = "output/inmlcf.parquet")
## args <- parser$parse_args()
##
##
## # ----- main
## # Federico & Liliana's work is in this file
## probs <- read_excel(args$dept_anio_probs) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
##   mutate(direction = ifelse(is.na(direction), 1, direction)) %>%
##   mutate(prob = as.numeric(prob)) %>%
##   mutate(prob = ifelse(is.na(prob), 0.5, prob)) %>%
##   rename(prob_forced_dis = prob) %>%
##   mutate(yy_hecho = year)
##
## inmlcf <- read_parquet(args$input) %>%
##   mutate(departamento_desaparicion = stri_trans_general(departamento_desaparicion, "Latin-ASCII")) %>%
##   mutate(yy_hecho = yy_fecha_desaparicion) %>%
##   mutate(perpetrator =
##     case_when(str_detect(presuncion_responsabilidad, "(AUTODEFENSAS|BLOQUE)") ~ "PARAMILITARES",
##               str_detect(presuncion_responsabilidad, "(DISIDENCIAS|FARC|SECRETARIADO|FRENTE|C") ~ "FRENTE",
##               str_detect(presuncion_responsabilidad, "(NACIONAL|ESTADO|DIVISION|POLICIA|GAULA") ~ "POLICIA",
##               str_detect(presuncion_responsabilidad, "CARTEL") ~ "NARCOTRAFICANTES",
##               str_detect(presuncion_responsabilidad, "BANDA") ~ "BANDAS EMERGENTES",
##               str_detect(presuncion_responsabilidad, "(PADRE|MADRE|OTROS FAMILIARES)") ~ "FAMILIAR",
##               str_detect(presuncion_responsabilidad, "NINGUNA") ~ NA_character_,
##               TRUE ~ presuncion_responsabilidad)) %>%
##   left_join(probs, by = c("departamento_desaparicion", "yy_hecho")) %>%
##   mutate(prob = runif(nrow(.)))
##
## inmlcf <- inmlcf %>%
##   mutate(is_forced_dis = case_when(
##     clasificacion_desaparicion == "DESAPARICION PRESUNTAMENTE FORZADA" ~ one,
##     pertenencia_grupal == "DESMOVILIZADOS" & is.na(clasificacion_desaparicion) ~ one,
##     pertenencia_grupal == "FUNCIONARIOS JUDICIALES" ~ one,
##     pertenencia_grupal == "SERVIDOR PUBLICO" ~ one,
##     pertenencia_grupal == "POLITICO - DIRIGENTE POLITICO" ~ one,
##     pertenencia_grupal == "SINDICALISTA" ~ one,
##     pertenencia_grupal == "RECLUSO" ~ one,
##     pertenencia_grupal == "MISION MEDICA - TRABAJADORES SALUD" ~ one,
##     pertenencia_grupal == "PERIODISTA" ~ one,
##     pertenencia_grupal == "GRUPOS ETNICOS" &
##       is.na(clasificacion_desaparicion) &
##       between(yy_fecha_desaparicion, 1998, 2004) ~ one,
##     clasificacion_desaparicion == "SECUESTRO" ~ one,
##     clasificacion_desaparicion == "RECLUTAMIENTO" ~ one,
##     clasificacion_desaparicion == "DESASTRE NATURAL" ~ zero,
##     str_detect(estado_desaparicion, "ANULADO") ~ zero,
##     str_detect(estado_desaparicion, "MUERTO") & (0.8 > prob) ~ one,
##     str_detect(estado_desaparicion, "VIVO") & (0.8 > prob) ~ zero,

```

```
## is.na(prob_forced_dis) ~ NA_integer_,
## (direction == 1) & (prob_forced_dis > prob) ~ one,
## (direction == 0) & (prob_forced_dis > prob) ~ zero,
## TRUE ~ NA_integer_)) %>%
## select(-prob_forced_dis, -prob, -direction, -yy_hecho) %>%
## write_parquet(args$output)
##
## print(table(inmlcf$forced_dis, useNA="always"))
## #done
```

```
library(readxl)
df <- read_excel(files$excel)
knitr::kable(df, "latex", longtable = T, booktabs = T)
```

#### 9.4.2.8. INML - EXCEL

dept_code_hecho	year	departamento_desaparicion	prob	direction
5	1980	ANTIOQUIA	1	-
8	1980	ATLÁNTICO	1	-
11	1980	BOGOTÁ D.C	1	-
13	1980	BOLÍVAR	1	-
15	1980	BOYACÁ	1	-
17	1980	CALDAS	1	-
18	1980	CAQUETÁ	1	-
85	1980	CASANARE	1	-
20	1980	CESAR	1	-
27	1980	CHOCÓ	1	-
25	1980	CUNDINAMARCA	1	-
41	1980	HUILA	1	-
47	1980	MAGDALENA	1	-
50	1980	META	1	-
52	1980	NARIÑO	1	-
54	1980	NORTE DE SANTANDER	1	-
63	1980	QUINDIO	1	-
68	1980	SANTANDER	1	-
-	1980	SIN INFORMACIÓN	1	0
70	1980	SUCRE	1	-
73	1980	TOLIMA	1	-
76	1980	VALLE DEL CAUCA	1	-
5	1981	ANTIOQUIA	1	-
11	1981	BOGOTÁ D.C	1	-
15	1981	BOYACÁ	1	-
85	1981	CASANARE	1	-
19	1981	CAUCA	1	-
41	1981	HUILA	1	-
50	1981	META	1	-
68	1981	SANTANDER	1	-

76	1981	VALLE DEL CAUCA	1	-
5	1982	ANTIOQUIA	1	-
81	1982	ARAUCA	1	-
8	1982	ATLÁNTICO	1	-
11	1982	BOGOTÁ D.C	1	-
13	1982	BOLÍVAR	1	-
15	1982	BOYACÁ	1	-
17	1982	CALDAS	1	-
18	1982	CAQUETÁ	1	-
19	1982	CAUCA	1	-
20	1982	CESAR	1	-
27	1982	CHOCÓ	1	-
23	1982	CÓRDOBA	1	-
25	1982	CUNDINAMARCA	1	-
95	1982	GUAVIARE	1	-
41	1982	HUILA	1	-
44	1982	LA GUAJIRA	0.5	-
47	1982	MAGDALENA	1	-
50	1982	META	1	-
52	1982	NARIÑO	1	-
54	1982	NORTE DE SANTANDER	1	-
86	1982	PUTUMAYO	1	-
63	1982	QUINDIO	1	-
66	1982	RISARALDA	1	-
68	1982	SANTANDER	1	-
-	1982	SIN INFORMACIÓN	1	0
-	1982	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1982	SUCRE	1	-
73	1982	TOLIMA	1	-
76	1982	VALLE DEL CAUCA	1	-
81	1983	ARAUCA	1	-
11	1983	BOGOTÁ D.C	1	-
18	1983	CAQUETÁ	1	-
19	1983	CAUCA	1	-
25	1983	CUNDINAMARCA	1	-
47	1983	MAGDALENA	1	-
50	1983	META	1	-
86	1983	PUTUMAYO	1	-
76	1983	VALLE DEL CAUCA	1	-
5	1984	ANTIOQUIA	1	-
81	1984	ARAUCA	1	-
11	1984	BOGOTÁ D.C	1	-
15	1984	BOYACÁ	1	-
18	1984	CAQUETÁ	1	-
19	1984	CAUCA	1	-
23	1984	CÓRDOBA	1	-
25	1984	CUNDINAMARCA	1	-
95	1984	GUAVIARE	1	-
41	1984	HUILA	1	-

47	1984	MAGDALENA	1	-
50	1984	META	1	-
54	1984	NORTE DE SANTANDER	1	-
63	1984	QUINDIO	1	-
68	1984	SANTANDER	1	-
76	1984	VALLE DEL CAUCA	1	-
99	1984	VICHADA	1	-
91	1985	AMAZONAS	1	-
5	1985	ANTIOQUIA	0.8	-
81	1985	ARAUCA	1	-
8	1985	ATLÁNTICO	1	-
11	1985	BOGOTÁ D.C	1	-
13	1985	BOLÍVAR	1	-
15	1985	BOYACÁ	1	-
17	1985	CALDAS	1	-
18	1985	CAQUETÁ	1	-
85	1985	CASANARE	1	-
19	1985	CAUCA	1	-
20	1985	CESAR	1	-
27	1985	CHOCÓ	1	-
23	1985	CÓRDOBA	1	-
25	1985	CUNDINAMARCA	1	-
94	1985	GUAINÍA	1	-
95	1985	GUAVIARE	1	-
41	1985	HUILA	1	-
44	1985	LA GUAJIRA	1	-
47	1985	MAGDALENA	1	-
50	1985	META	1	-
52	1985	NARIÑO	1	-
54	1985	NORTE DE SANTANDER	1	-
86	1985	PUTUMAYO	1	-
63	1985	QUINDIO	1	-
66	1985	RISARALDA	1	-
68	1985	SANTANDER	1	-
-	1985	SIN INFORMACIÓN	1	0
-	1985	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1985	SUCRE	1	-
73	1985	TOLIMA	0.8	-
76	1985	VALLE DEL CAUCA	1	-
99	1985	VICHADA	1	-
5	1986	ANTIOQUIA	0.8	-
81	1986	ARAUCA	0.8	-
11	1986	BOGOTÁ D.C	1	-
13	1986	BOLÍVAR	1	-
15	1986	BOYACÁ	0.8	-
17	1986	CALDAS	1	-
18	1986	CAQUETÁ	1	-
85	1986	CASANARE	1	-
19	1986	CAUCA	1	-

20	1986	CESAR	1	-
23	1986	CÓRDOBA	1	-
25	1986	CUNDINAMARCA	1	-
94	1986	GUAINÍA	1	-
95	1986	GUAVIARE	1	-
41	1986	HUILA	1	-
44	1986	LA GUAJIRA	1	-
-	1986	LARA	1	0
47	1986	MAGDALENA	1	-
50	1986	META	1	-
52	1986	NARIÑO	1	-
54	1986	NORTE DE SANTANDER	1	-
86	1986	PUTUMAYO	1	-
63	1986	QUINDIO	1	-
66	1986	RISARALDA	1	-
68	1986	SANTANDER	1	-
-	1986	SIN INFORMACIÓN	1	0
-	1986	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1986	SUCRE	1	-
73	1986	TOLIMA	1	-
76	1986	VALLE DEL CAUCA	1	-
99	1986	VICHADA	1	-
5	1987	ANTIOQUIA	0.8	-
81	1987	ARAUCA	1	-
8	1987	ATLÁNTICO	1	-
11	1987	BOGOTÁ D.C	1	-
13	1987	BOLÍVAR	1	-
15	1987	BOYACÁ	0.8	-
17	1987	CALDAS	1	-
18	1987	CAQUETÁ	1	-
85	1987	CASANARE	1	-
19	1987	CAUCA	1	-
20	1987	CESAR	1	-
27	1987	CHOCÓ	1	-
23	1987	CÓRDOBA	1	-
25	1987	CUNDINAMARCA	1	-
95	1987	GUAVIARE	1	-
41	1987	HUILA	1	-
44	1987	LA GUAJIRA	1	-
47	1987	MAGDALENA	1	-
50	1987	META	1	-
52	1987	NARIÑO	1	-
54	1987	NORTE DE SANTANDER	1	-
86	1987	PUTUMAYO	1	-
63	1987	QUINDIO	1	-
66	1987	RISARALDA	1	-
68	1987	SANTANDER	1	-
-	1987	SIN INFORMACIÓN	1	-
70	1987	SUCRE	1	-

73	1987	TOLIMA	1	-
76	1987	VALLE DEL CAUCA	1	-
97	1987	VAUPÉS	1	-
99	1987	VICHADA	1	-
5	1988	ANTIOQUIA	0.7	-
81	1988	ARAUCA	1	-
8	1988	ATLÁNTICO	1	-
11	1988	BOGOTÁ D.C	1	-
13	1988	BOLÍVAR	1	-
15	1988	BOYACÁ	1	-
17	1988	CALDAS	0.8	-
18	1988	CAQUETÁ	1	-
85	1988	CASANARE	1	-
19	1988	CAUCA	1	-
20	1988	CESAR	1	-
27	1988	CHOCÓ	1	-
23	1988	CÓRDOBA	1	-
25	1988	CUNDINAMARCA	0.8	-
95	1988	GUAVIARE	1	-
41	1988	HUILA	1	-
44	1988	LA GUAJIRA	1	-
47	1988	MAGDALENA	1	-
50	1988	META	1	-
52	1988	NARIÑO	1	-
54	1988	NORTE DE SANTANDER	1	-
86	1988	PUTUMAYO	1	-
63	1988	QUINDIO	1	-
66	1988	RISARALDA	1	-
68	1988	SANTANDER	0.8	-
-	1988	SIN INFORMACIÓN	1	0
-	1988	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1988	SUCRE	1	-
73	1988	TOLIMA	0.8	-
76	1988	VALLE DEL CAUCA	0.8	-
99	1988	VICHADA	1	-
91	1989	AMAZONAS	1	-
5	1989	ANTIOQUIA	0.8	-
81	1989	ARAUCA	1	-
8	1989	ATLÁNTICO	0.8	-
11	1989	BOGOTÁ D.C	1	-
13	1989	BOLÍVAR	1	-
15	1989	BOYACÁ	0.8	-
17	1989	CALDAS	0.8	-
18	1989	CAQUETÁ	1	-
85	1989	CASANARE	1	-
19	1989	CAUCA	1	-
20	1989	CESAR	1	-
27	1989	CHOCÓ	1	-
23	1989	CÓRDOBA	1	-

25	1989	CUNDINAMARCA	1	-
94	1989	GUAINÍA	1	-
95	1989	GUAVIARE	1	-
41	1989	HUILA	1	-
44	1989	LA GUAJIRA	1	-
47	1989	MAGDALENA	1	-
-	1989	MÉRIDA	1	0
50	1989	META	1	-
52	1989	NARIÑO	1	-
54	1989	NORTE DE SANTANDER	1	-
86	1989	PUTUMAYO	1	-
63	1989	QUINDIO	1	-
66	1989	RISARALDA	1	-
68	1989	SANTANDER	1	-
-	1989	SIN INFORMACIÓN	1	0
-	1989	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1989	SUCRE	1	-
73	1989	TOLIMA	1	-
76	1989	VALLE DEL CAUCA	0.8	-
97	1989	VAUPÉS	1	-
99	1989	VICHADA	1	-
5	1990	ANTIOQUIA	0.8	-
81	1990	ARAUCA	1	-
8	1990	ATLÁNTICO	0.8	-
11	1990	BOGOTÁ D.C	0.8	-
13	1990	BOLÍVAR	0.8	-
15	1990	BOYACÁ	0.8	-
17	1990	CALDAS	0.8	-
18	1990	CAQUETÁ	1	-
85	1990	CASANARE	1	-
19	1990	CAUCA	1	-
20	1990	CESAR	1	-
27	1990	CHOCÓ	1	-
23	1990	CÓRDOBA	1	-
25	1990	CUNDINAMARCA	0.8	-
-	1990	FLORIDA	1	0
95	1990	GUAVIARE	-	-
41	1990	HUILA	1	-
44	1990	LA GUAJIRA	0.7	-
47	1990	MAGDALENA	1	-
50	1990	META	1	-
52	1990	NARIÑO	1	-
54	1990	NORTE DE SANTANDER	1	-
-	1990	PORTUGUESA	1	0
86	1990	PUTUMAYO	1	-
63	1990	QUINDIO	1	-
66	1990	RISARALDA	0.8	-
68	1990	SANTANDER	-	-
-	1990	SIN INFORMACIÓN	1	0

70	1990	SUCRE	1	-
73	1990	TOLIMA	0.8	-
76	1990	VALLE DEL CAUCA	0.7	-
97	1990	VAUPÉS	1	-
99	1990	VICHADA	1	-
5	1991	ANTIOQUIA	0.7	-
81	1991	ARAUCA	1	-
8	1991	ATLÁNTICO	0.7	-
11	1991	BOGOTÁ D.C	0.8	-
13	1991	BOLÍVAR	0.8	-
15	1991	BOYACÁ	0.7	-
17	1991	CALDAS	0.7	-
18	1991	CAQUETÁ	1	-
85	1991	CASANARE	1	-
19	1991	CAUCA	1	-
20	1991	CESAR	1	-
27	1991	CHOCÓ	1	-
23	1991	CÓRDOBA	1	-
25	1991	CUNDINAMARCA	0.8	-
94	1991	GUAINÍA	1	-
95	1991	GUAVIARE	1	-
41	1991	HUILA	1	-
44	1991	LA GUAJIRA	0.7	-
47	1991	MAGDALENA	1	-
50	1991	META	0.8	-
52	1991	NARIÑO	1	-
54	1991	NORTE DE SANTANDER	1	-
86	1991	PUTUMAYO	1	-
63	1991	QUINDIO	1	-
66	1991	RISARALDA	0.8	-
68	1991	SANTANDER	1	-
-	1991	SIN INFORMACIÓN	1	0
70	1991	SUCRE	1	-
73	1991	TOLIMA	0.8	-
76	1991	VALLE DEL CAUCA	0.7	-
0	1992	-	1	0
91	1992	AMAZONAS	1	-
5	1992	ANTIOQUIA	0.8	-
81	1992	ARAUCA	1	-
8	1992	ATLÁNTICO	0.7	-
11	1992	BOGOTÁ D.C	0.7	-
13	1992	BOLÍVAR	1	-
15	1992	BOYACÁ	0.8	-
17	1992	CALDAS	0.7	-
18	1992	CAQUETÁ	1	-
85	1992	CASANARE	1	-
19	1992	CAUCA	1	-
20	1992	CESAR	1	-
27	1992	CHOCÓ	1	-



23	1992	CÓRDOBA	1	-
25	1992	CUNDINAMARCA	0.8	-
-	1992	ESMERALDAS	1	0
95	1992	GUAVIARE	1	-
41	1992	HUILA	0.8	-
44	1992	LA GUAJIRA	0.8	-
47	1992	MAGDALENA	1	-
50	1992	META	0.8	-
52	1992	NARIÑO	1	-
54	1992	NORTE DE SANTANDER	1	-
86	1992	PUTUMAYO	1	-
63	1992	QUINDIO	0.8	-
66	1992	RISARALDA	0.8	-
88	1992	SAN ANDRES Y PROVIDENCIA	1	0
68	1992	SANTANDER	0.8	-
-	1992	SIN INFORMACIÓN	1	0
70	1992	SUCRE	1	-
73	1992	TOLIMA	0.7	-
76	1992	VALLE DEL CAUCA	0.7	-
97	1992	VAUPÉS	1	-
99	1992	VICHADA	1	-
5	1993	ANTIOQUIA	0.8	-
81	1993	ARAUCA	1	-
8	1993	ATLÁNTICO	0.8	-
-	1993	BARINAS	1	0
11	1993	BOGOTÁ D.C	0.7	-
13	1993	BOLÍVAR	1	-
15	1993	BOYACÁ	0.7	-
17	1993	CALDAS	0.7	-
18	1993	CAQUETÁ	1	-
85	1993	CASANARE	1	-
19	1993	CAUCA	1	-
20	1993	CESAR	1	-
27	1993	CHOCÓ	1	-
23	1993	CÓRDOBA	1	-
25	1993	CUNDINAMARCA	0.8	-
-	1993	FLORIDA	1	0
95	1993	GUAVIARE	1	-
41	1993	HUILA	0.8	-
44	1993	LA GUAJIRA	0.8	-
47	1993	MAGDALENA	1	-
50	1993	META	1	-
52	1993	NARIÑO	1	-
54	1993	NORTE DE SANTANDER	1	-
-	1993	PANAMÁ	1	0
86	1993	PUTUMAYO	1	-
63	1993	QUINDIO	0.8	-
66	1993	RISARALDA	0.7	-
68	1993	SANTANDER	1	-

-	1993	SIN INFORMACIÓN	1	0
70	1993	SUCRE	1	-
-	1993	TÁCHIRA	1	0
73	1993	TOLIMA	0.7	-
76	1993	VALLE DEL CAUCA	0.8	-
99	1993	VICHADA	1	-
-	1993	ZULIA	1	0
5	1994	ANTIOQUIA	0.7	-
81	1994	ARAUCA	1	-
8	1994	ATLÁNTICO	0.7	-
11	1994	BOGOTÁ D.C	0.6	-
13	1994	BOLÍVAR	1	-
15	1994	BOYACÁ	0.8	-
17	1994	CALDAS	0.7	-
18	1994	CAQUETÁ	1	-
85	1994	CASANARE	1	-
19	1994	CAUCA	1	-
20	1994	CESAR	1	-
27	1994	CHOCÓ	1	-
23	1994	CÓRDOBA	1	-
25	1994	CUNDINAMARCA	0.7	-
-	1994	DISTRITO CAPITAL	-	-
94	1994	GUAINÍA	1	-
95	1994	GUAVIARE	1	-
41	1994	HUILA	0.8	-
44	1994	LA GUAJIRA	0.8	-
47	1994	MAGDALENA	1	-
50	1994	META	0.8	-
52	1994	NARIÑO	1	-
54	1994	NORTE DE SANTANDER	1	-
86	1994	PUTUMAYO	1	-
63	1994	QUINDIO	0.8	-
66	1994	RISARALDA	0.8	-
68	1994	SANTANDER	0.8	-
-	1994	SIN INFORMACIÓN	1	0
70	1994	SUCRE	1	-
73	1994	TOLIMA	0.7	-
76	1994	VALLE DEL CAUCA	0.8	-
97	1994	VAUPÉS	1	-
99	1994	VICHADA	1	-
5	1995	ANTIOQUIA	0.6	-
-	1995	APURE	1	0
81	1995	ARAUCA	1	-
8	1995	ATLÁNTICO	0.7	-
11	1995	BOGOTÁ D.C	0.6	-
13	1995	BOLÍVAR	1	-
15	1995	BOYACÁ	0.7	-
17	1995	CALDAS	0.7	-
18	1995	CAQUETÁ	1	-

85	1995	CASANARE	1	-
19	1995	CAUCA	1	-
20	1995	CESAR	1	-
27	1995	CHOCÓ	1	-
23	1995	CÓRDOBA	1	-
25	1995	CUNDINAMARCA	0.8	-
94	1995	GUAINÍA	1	-
95	1995	GUAVIARE	1	-
41	1995	HUILA	0.8	-
44	1995	LA GUAJIRA	0.8	-
47	1995	MAGDALENA	1	-
50	1995	META	0.8	-
52	1995	NARIÑO	1	-
54	1995	NORTE DE SANTANDER	1	-
86	1995	PUTUMAYO	1	-
63	1995	QUINDIO	1	-
66	1995	RISARALDA	0.7	-
68	1995	SANTANDER	0.8	-
-	1995	SIN INFORMACIÓN	1	0
-	1995	SIN INFORMACIÓN (VENEZUELA)	1	0
70	1995	SUCRE	1	-
73	1995	TOLIMA	0.7	-
76	1995	VALLE DEL CAUCA	0.8	-
99	1995	VICHADA	1	-
0	1996	(NA)	1	0
91	1996	AMAZONAS	1	-
5	1996	ANTIOQUIA	0.6	-
-	1996	APURE	1	0
81	1996	ARAUCA	1	-
8	1996	ATLÁNTICO	0.7	-
11	1996	BOGOTÁ D.C	0.6	-
13	1996	BOLÍVAR	1	-
15	1996	BOYACÁ	0.7	-
17	1996	CALDAS	0.7	-
18	1996	CAQUETÁ	1	-
85	1996	CASANARE	1	-
19	1996	CAUCA	1	-
20	1996	CESAR	1	-
27	1996	CHOCÓ	1	-
23	1996	CÓRDOBA	1	-
25	1996	CUNDINAMARCA	0.8	-
-	1996	DISTRITO CAPITAL	1	-
94	1996	GUAINÍA	1	-
95	1996	GUAVIARE	1	-
41	1996	HUILA	0.8	-
44	1996	LA GUAJIRA	1	-
47	1996	MAGDALENA	1	-
50	1996	META	0.8	-
52	1996	NARIÑO	1	-

54	1996	NORTE DE SANTANDER	1	-
86	1996	PUTUMAYO	1	-
63	1996	QUINDIO	1	-
66	1996	RISARALDA	0.7	-
68	1996	SANTANDER	0.8	-
-	1996	SIN INFORMACIÓN	1	0
70	1996	SUCRE	1	-
-	1996	TÁCHIRA	1	0
73	1996	TOLIMA	0.7	-
76	1996	VALLE DEL CAUCA	0.8	-
-	1996	VARGAS	1	0
97	1996	VAUPÉS	1	-
99	1996	VICHADA	1	-
91	1997	AMAZONAS	1	-
5	1997	ANTIOQUIA	0.7	-
-	1997	APURE	1	0
81	1997	ARAUCA	1	-
8	1997	ATLÁNTICO	0.7	-
11	1997	BOGOTÁ D.C	0.6	-
13	1997	BOLÍVAR	1	-
15	1997	BOYACÁ	0.7	-
17	1997	CALDAS	0.6	-
18	1997	CAQUETÁ	1	-
85	1997	CASANARE	1	-
19	1997	CAUCA	1	-
20	1997	CESAR	0.8	-
27	1997	CHOCÓ	0.6	-
23	1997	CÓRDOBA	0.8	-
25	1997	CUNDINAMARCA	0.7	-
95	1997	GUAVIARE	0.8	-
41	1997	HUILA	0.8	-
44	1997	LA GUAJIRA	0.7	-
47	1997	MAGDALENA	1	-
50	1997	META	0.8	-
52	1997	NARIÑO	1	-
54	1997	NORTE DE SANTANDER	1	-
86	1997	PUTUMAYO	1	-
63	1997	QUINDIO	0.7	-
66	1997	RISARALDA	0.7	-
68	1997	SANTANDER	0.8	-
-	1997	SIN INFORMACIÓN	1	0
70	1997	SUCRE	1	-
-	1997	TÁCHIRA	1	0
73	1997	TOLIMA	0.6	-
76	1997	VALLE DEL CAUCA	0.8	-
97	1997	VAUPÉS	1	-
99	1997	VICHADA	1	-
-	1997	ZULIA	1	0
0	1998	-	1	0

5	1998	ANTIOQUIA	0.7	-
81	1998	ARAUCA	1	-
8	1998	ATLÁNTICO	0.7	-
11	1998	BOGOTÁ D.C	0.6	-
13	1998	BOLÍVAR	1	-
15	1998	BOYACÁ	0.7	-
17	1998	CALDAS	0.6	-
18	1998	CAQUETÁ	0.8	-
85	1998	CASANARE	1	-
19	1998	CAUCA	1	-
20	1998	CESAR	1	-
27	1998	CHOCÓ	1	-
23	1998	CÓRDOBA	1	-
25	1998	CUNDINAMARCA	0.7	-
94	1998	GUAINÍA	1	-
95	1998	GUAVIARE	1	-
41	1998	HUILA	0.7	-
44	1998	LA GUAJIRA	0.7	-
47	1998	MAGDALENA	0.8	-
50	1998	META	0.7	-
52	1998	NARIÑO	1	-
54	1998	NORTE DE SANTANDER	1	-
86	1998	PUTUMAYO	1	-
63	1998	QUINDIO	0.8	-
66	1998	RISARALDA	0.6	-
68	1998	SANTANDER	0.8	-
-	1998	SIN INFORMACIÓN	1	0
70	1998	SUCRE	1	-
73	1998	TOLIMA	0.6	-
76	1998	VALLE DEL CAUCA	0.8	-
97	1998	VAUPÉS	1	-
99	1998	VICHADA	1	-
5	1999	ANTIOQUIA	0.6	-
-	1999	APURE	1	0
81	1999	ARAUCA	1	-
8	1999	ATLÁNTICO	0.7	-
11	1999	BOGOTÁ D.C	0.7	0
13	1999	BOLÍVAR	0.8	-
15	1999	BOYACÁ	0.7	-
17	1999	CALDAS	0.6	-
18	1999	CAQUETÁ	1	-
85	1999	CASANARE	0.8	-
19	1999	CAUCA	1	-
20	1999	CESAR	1	-
27	1999	CHOCÓ	1	-
23	1999	CÓRDOBA	1	-
25	1999	CUNDINAMARCA	0.7	-
-	1999	DISTRITO CAPITAL	1	-
94	1999	GUAINÍA	1	-

95	1999	GUAVIARE	1	-
41	1999	HUILA	0.7	-
44	1999	LA GUAJIRA	0.8	-
47	1999	MAGDALENA	1	-
50	1999	META	0.8	-
-	1999	MIRANDA	1	0
52	1999	NARIÑO	1	-
54	1999	NORTE DE SANTANDER	0.7	-
86	1999	PUTUMAYO	0.8	-
63	1999	QUINDIO	0.6	-
66	1999	RISARALDA	0.6	-
88	1999	SAN ANDRES Y PROVIDENCIA	1	0
68	1999	SANTANDER	0.8	-
-	1999	SIN INFORMACIÓN	1	0
70	1999	SUCRE	1	-
73	1999	TOLIMA	0.7	-
76	1999	VALLE DEL CAUCA	0.7	-
-	1999	VARGAS	1	0
97	1999	VAUPÉS	1	-
99	1999	VICHADA	1	-
0	2000	-	1	0
91	2000	AMAZONAS	1	-
5	2000	ANTIOQUIA	0.5	0
-	2000	APURE	1	0
81	2000	ARAUCA	1	0
8	2000	ATLÁNTICO	0.6	-
11	2000	BOGOTÁ D.C	0.7	0
13	2000	BOLÍVAR	0.7	0
15	2000	BOYACÁ	0.7	0
17	2000	CALDAS	0.7	0
18	2000	CAQUETÁ	0.8	-
85	2000	CASANARE	1	-
19	2000	CAUCA	0.8	-
20	2000	CESAR	0.8	-
27	2000	CHOCÓ	1	-
23	2000	CÓRDOBA	0.7	-
25	2000	CUNDINAMARCA	0.7	0
-	2000	CURAÇAO	1	-
94	2000	GUAINÍA	1	-
95	2000	GUAVIARE	1	-
41	2000	HUILA	0.7	-
44	2000	LA GUAJIRA	0.6	-
47	2000	MAGDALENA	0.7	-
50	2000	META	0.8	-
-	2000	MONTEVIDEO	1	0
52	2000	NARIÑO	1	-
54	2000	NORTE DE SANTANDER	0.8	-
-	2000	PUNTARENAS	1	0
86	2000	PUTUMAYO	0.8	-

63	2000	QUINDIO	0.7	-
66	2000	RISARALDA	0.6	-
68	2000	SANTANDER	0.8	-
-	2000	SIN INFORMACIÓN	1	0
-	2000	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2000	SUCRE	1	-
73	2000	TOLIMA	0.5	0
76	2000	VALLE DEL CAUCA	0.5	0
97	2000	VAUPÉS	1	-
99	2000	VICHADA	1	-
0	2001	-	1	0
5	2001	ANTIOQUIA	0.7	0
-	2001	APURE	1	0
81	2001	ARAUCA	1	-
8	2001	ATLÁNTICO	0.7	-
11	2001	BOGOTÁ D.C	0.8	0
13	2001	BOLÍVAR	0.7	-
15	2001	BOYACÁ	0.7	0
17	2001	CALDAS	0.8	0
18	2001	CAQUETÁ	0.8	-
85	2001	CASANARE	0.8	-
-	2001	CATALUÑA	1	0
19	2001	CAUCA	0.8	-
20	2001	CESAR	0.8	-
27	2001	CHOCÓ	1	-
23	2001	CÓRDOBA	0.8	-
25	2001	CUNDINAMARCA	0.7	0
95	2001	GUAVIARE	1	-
41	2001	HUILA	0.6	0
44	2001	LA GUAJIRA	0.7	-
47	2001	MAGDALENA	1	-
50	2001	META	0.7	-
52	2001	NARIÑO	0.8	-
-	2001	NEW YORK	1	0
54	2001	NORTE DE SANTANDER	0.8	-
86	2001	PUTUMAYO	0.8	-
63	2001	QUINDIO	0.7	-
66	2001	RISARALDA	0.6	-
68	2001	SANTANDER	0.8	-
-	2001	SIN INFORMACIÓN	1	0
70	2001	SUCRE	1	-
-	2001	SUCUMBÍOS	1	0
-	2001	TÁCHIRA	1	0
73	2001	TOLIMA	0.7	0
76	2001	VALLE DEL CAUCA	0.8	-
97	2001	VAUPÉS	1	-
99	2001	VICHADA	1	-
0	2002	-	1	0
91	2002	AMAZONAS	1	-

5	2002	ANTIOQUIA	0.5	-
-	2002	APURE	1	0
81	2002	ARAUCA	0.7	-
8	2002	ATLÁNTICO	0.7	0
11	2002	BOGOTÁ D.C	0.8	0
13	2002	BOLÍVAR	0.8	-
15	2002	BOYACÁ	0.7	0
17	2002	CALDAS	0.8	0
18	2002	CAQUETÁ	0.8	-
85	2002	CASANARE	0.8	-
19	2002	CAUCA	0.7	-
20	2002	CESAR	0.8	-
27	2002	CHOCÓ	1	-
23	2002	CÓRDOBA	1	-
25	2002	CUNDINAMARCA	0.7	0
94	2002	GUAINÍA	1	-
95	2002	GUAVIARE	0.8	-
41	2002	HUILA	0.6	0
44	2002	LA GUAJIRA	0.7	0
47	2002	MAGDALENA	0.7	-
-	2002	MÉRIDA	1	0
50	2002	META	0.7	-
-	2002	MONAGAS	1	0
52	2002	NARIÑO	0.7	-
54	2002	NORTE DE SANTANDER	0.7	0
86	2002	PUTUMAYO	0.7	-
63	2002	QUINDIO	0.7	0
66	2002	RISARALDA	0.8	0
-	2002	SAN JOSÉ	1	0
-	2002	SANTA CRUZ	1	0
68	2002	SANTANDER	1	-
-	2002	SIN INFORMACIÓN	1	0
-	2002	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2002	SUCRE	1	-
-	2002	TÁCHIRA	1	0
73	2002	TOLIMA	0.8	0
76	2002	VALLE DEL CAUCA	0.7	-
97	2002	VAUPÉS	1	-
99	2002	VICHADA	1	-
-	2002	ZULIA	1	0
0	2003	-	1	0
91	2003	AMAZONAS	0.8	-
5	2003	ANTIOQUIA	0.7	-
-	2003	APURE	1	0
81	2003	ARAUCA	1	-
8	2003	ATLÁNTICO	0.8	0
11	2003	BOGOTÁ D.C	0.8	0
13	2003	BOLÍVAR	0.8	-



15	2003	BOYACÁ	0.7	0
17	2003	CALDAS	0.8	0
18	2003	CAQUETÁ	0.8	-
-	2003	CARABOBO	1	0
85	2003	CASANARE	0.7	-
19	2003	CAUCA	1	-
20	2003	CESAR	0.8	-
27	2003	CHOCÓ	1	-
23	2003	CÓRDOBA	1	-
25	2003	CUNDINAMARCA	0.8	0
-	2003	DISTRITO CAPITAL	1	-
94	2003	GUAINÍA	1	-
95	2003	GUAVIARE	0.8	-
41	2003	HUILA	0.6	-
44	2003	LA GUAJIRA	0.7	0
47	2003	MAGDALENA	0.8	-
-	2003	MÉRIDA	1	0
50	2003	META	0.7	-
52	2003	NARIÑO	0.8	-
54	2003	NORTE DE SANTANDER	0.8	-
-	2003	PANAMÁ	1	0
86	2003	PUTUMAYO	0.8	-
63	2003	QUINDIO	0.7	0
66	2003	RISARALDA	0.8	0
68	2003	SANTANDER	0.8	-
-	2003	SIN INFORMACIÓN	1	0
-	2003	SOLOLÁ	1	0
70	2003	SUCRE	1	-
-	2003	TÁCHIRA	1	0
73	2003	TOLIMA	0.8	0
76	2003	VALLE DEL CAUCA	0.5	0
97	2003	VAUPÉS	1	-
99	2003	VICHADA	1	-
0	2004	-	1	0
5	2004	ANTIOQUIA	0.5	-
-	2004	APURE	1	0
81	2004	ARAUCA	1	-
8	2004	ATLÁNTICO	0.6	0
11	2004	BOGOTÁ D.C	0.8	0
13	2004	BOLÍVAR	0.7	-
15	2004	BOYACÁ	0.6	-
17	2004	CALDAS	0.7	0
18	2004	CAQUETÁ	0.8	-
85	2004	CASANARE	0.8	-
19	2004	CAUCA	0.8	-
20	2004	CESAR	0.8	-
27	2004	CHOCÓ	0.8	-
23	2004	CÓRDOBA	0.8	-
25	2004	CUNDINAMARCA	0.6	-

-	2004	DISTRITO CAPITAL	1	-
-	2004	ESMERALDAS	1	0
94	2004	GUAINÍA	1	-
95	2004	GUAVIARE	0.8	-
41	2004	HUILA	0.6	-
44	2004	LA GUAJIRA	0.7	-
-	2004	LARA	1	0
47	2004	MAGDALENA	0.8	-
-	2004	MÉRIDA	1	0
50	2004	META	0.8	-
52	2004	NARIÑO	0.8	-
54	2004	NORTE DE SANTANDER	0.7	-
-	2004	PANAMÁ	1	0
-	2004	PICHINCHA	1	-
86	2004	PUTUMAYO	0.8	-
63	2004	QUINDIO	0.7	-
66	2004	RISARALDA	0.8	-
88	2004	SAN ANDRES Y PROVIDENCIA	1	0
68	2004	SANTANDER	0.8	-
-	2004	SIN INFORMACIÓN	1	0
-	2004	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2004	SUCRE	1	-
-	2004	TÁCHIRA	1	0
73	2004	TOLIMA	0.8	0
76	2004	VALLE DEL CAUCA	0.6	-
97	2004	VAUPÉS	1	-
99	2004	VICHADA	1	-
0	2005	-	1	0
91	2005	AMAZONAS	1	-
5	2005	ANTIOQUIA	0.7	-
-	2005	APURE	1	0
81	2005	ARAUCA	1	-
8	2005	ATLÁNTICO	0.7	0
11	2005	BOGOTÁ D.C	0.8	0
13	2005	BOLÍVAR	1	-
15	2005	BOYACÁ	0.8	0
17	2005	CALDAS	0.7	-
18	2005	CAQUETÁ	1	-
-	2005	CARABOBO	1	0
85	2005	CASANARE	0.8	-
19	2005	CAUCA	1	-
20	2005	CESAR	1	-
27	2005	CHOCÓ	1	-
23	2005	CÓRDOBA	1	-
25	2005	CUNDINAMARCA	0.8	0
94	2005	GUAINÍA	1	-
95	2005	GUAVIARE	0.8	-
41	2005	HUILA	0.7	0
-	2005	IMBABURA	1	0

44	2005	LA GUAJIRA	0.7	0
-	2005	LOJA	1	0
47	2005	MAGDALENA	1	-
50	2005	META	0.8	-
52	2005	NARIÑO	0.8	-
54	2005	NORTE DE SANTANDER	0.8	-
86	2005	PUTUMAYO	0.8	-
63	2005	QUINDIO	0.6	0
66	2005	RISARALDA	0.6	0
88	2005	SAN ANDRES Y PROVIDENCIA	1	0
68	2005	SANTANDER	1	-
-	2005	SIN INFORMACIÓN	1	0
70	2005	SUCRE	1	-
-	2005	TÁCHIRA	1	0
73	2005	TOLIMA	0.7	0
76	2005	VALLE DEL CAUCA	0.7	-
97	2005	VAUPÉS	1	-
99	2005	VICHADA	1	-
-	2005	ZULIA	1	0
0	2006	-	1	0
91	2006	AMAZONAS	1	-
5	2006	ANTIOQUIA	0.7	-
-	2006	APURE	1	0
-	2006	ARAGUA	1	0
81	2006	ARAUCA	1	-
-	2006	ASUNCIÓN	1	0
8	2006	ATLÁNTICO	0.7	-
11	2006	BOGOTÁ D.C	0.8	0
13	2006	BOLÍVAR	0.8	-
15	2006	BOYACÁ	0.7	0
17	2006	CALDAS	0.7	0
18	2006	CAQUETÁ	1	-
85	2006	CASANARE	0.8	-
19	2006	CAUCA	0.8	-
20	2006	CESAR	0.8	-
27	2006	CHOCÓ	0.8	-
23	2006	CÓRDOBA	0.8	-
25	2006	CUNDINAMARCA	0.7	0
-	2006	DISTRITO CAPITAL	1	-
94	2006	GUAINÍA	1	-
95	2006	GUAVIARE	1	-
41	2006	HUILA	0.7	0
-	2006	IMBABURA	1	-
44	2006	LA GUAJIRA	0.7	0
-	2006	LAS PALMAS	1	0
47	2006	MAGDALENA	1	-
50	2006	META	1	-
52	2006	NARIÑO	1	-
54	2006	NORTE DE SANTANDER	1	-

-	2006	PICHINCHA	1	0
86	2006	PUTUMAYO	1	-
63	2006	QUINDIO	0.7	-
66	2006	RISARALDA	0.7	-
68	2006	SANTANDER	1	-
-	2006	SIN INFORMACIÓN	1	0
-	2006	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2006	SUCRE	1	-
-	2006	TÁCHIRA	1	0
73	2006	TOLIMA	0.7	0
76	2006	VALLE DEL CAUCA	0.7	-
97	2006	VAUPÉS	1	-
99	2006	VICHADA	1	-
-	2006	ZULIA	1	0
0	2007	-	1	0
91	2007	AMAZONAS	0.8	-
5	2007	ANTIOQUIA	0.8	-
-	2007	APURE	1	0
81	2007	ARAUCA	1	-
8	2007	ATLÁNTICO	0.7	0
-	2007	BARINAS	1	0
11	2007	BOGOTÁ D.C	0.8	0
13	2007	BOLÍVAR	0.8	-
15	2007	BOYACÁ	0.7	0
17	2007	CALDAS	0.6	-
18	2007	CAQUETÁ	0.6	0
85	2007	CASANARE	0.7	-
19	2007	CAUCA	0.6	0
20	2007	CESAR	0.8	-
27	2007	CHOCÓ	1	-
23	2007	CÓRDOBA	0.8	-
25	2007	CUNDINAMARCA	0.7	-
-	2007	DISTRITO CAPITAL	1	-
-	2007	ESMERALDAS	1	0
94	2007	GUAINÍA	1	-
95	2007	GUAVIARE	0.6	0
41	2007	HUILA	0.7	0
44	2007	LA GUAJIRA	0.7	0
47	2007	MAGDALENA	1	-
50	2007	META	0.6	0
52	2007	NARIÑO	0.6	0
54	2007	NORTE DE SANTANDER	0.8	-
-	2007	PANAMÁ	1	0
-	2007	PICHINCHA	1	0
86	2007	PUTUMAYO	0.6	0
63	2007	QUINDIO	0.7	-
66	2007	RISARALDA	0.7	-
68	2007	SANTANDER	0.8	-
-	2007	SIN INFORMACIÓN	1	0

-	2007	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2007	SUCRE	1	-
-	2007	TÁCHIRA	1	0
73	2007	TOLIMA	0.7	0
76	2007	VALLE DEL CAUCA	0.8	0
97	2007	VAUPÉS	1	-
99	2007	VICHADA	0.6	-
91	2008	AMAZONAS	0.8	-
5	2008	ANTIOQUIA	0.8	-
-	2008	APURE	1	0
81	2008	ARAUCA	1	-
8	2008	ATLÁNTICO	0.7	0
11	2008	BOGOTÁ D.C	0.8	0
13	2008	BOLÍVAR	1	-
15	2008	BOYACÁ	0.7	0
17	2008	CALDAS	0.7	0
18	2008	CAQUETÁ	1	-
85	2008	CASANARE	0.8	-
19	2008	CAUCA	0.8	-
20	2008	CESAR	1	-
27	2008	CHOCÓ	1	-
23	2008	CÓRDOBA	1	-
25	2008	CUNDINAMARCA	0.7	0
-	2008	DISTRITO FEDERAL	1	0
94	2008	GUAINÍA	1	-
95	2008	GUAVIARE	1	-
41	2008	HUILA	0.7	0
44	2008	LA GUAJIRA	0.7	0
47	2008	MAGDALENA	1	-
50	2008	META	0.8	-
52	2008	NARIÑO	0.8	-
54	2008	NORTE DE SANTANDER	1	-
-	2008	PICHINCHA	1	-
86	2008	PUTUMAYO	1	-
63	2008	QUINDIO	0.7	-
66	2008	RISARALDA	0.6	-
88	2008	SAN ANDRES Y PROVIDENCIA	1	0
68	2008	SANTANDER	0.8	-
-	2008	SIN INFORMACIÓN	1	0
-	2008	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2008	SUCRE	1	-
-	2008	TÁCHIRA	1	0
73	2008	TOLIMA	0.7	0
76	2008	VALLE DEL CAUCA	0.7	0
97	2008	VAUPÉS	1	-
99	2008	VICHADA	0.7	-
91	2009	AMAZONAS	1	-
5	2009	ANTIOQUIA	0.7	0
-	2009	APURE	1	0

81	2009	ARAUCA	1	-
8	2009	ATLÁNTICO	0.8	0
11	2009	BOGOTÁ D.C	0.9	0
13	2009	BOLÍVAR	1	-
15	2009	BOYACÁ	0.8	0
17	2009	CALDAS	0.7	0
18	2009	CAQUETÁ	1	-
85	2009	CASANARE	0.8	-
19	2009	CAUCA	1	-
20	2009	CESAR	1	-
27	2009	CHOCÓ	1	-
23	2009	CÓRDOBA	1	-
25	2009	CUNDINAMARCA	0.7	-
-	2009	DISTRITO CAPITAL	1	-
-	2009	DISTRITO FEDERAL	1	0
94	2009	GUAINÍA	1	-
95	2009	GUAVIARE	1	-
41	2009	HUILA	0.7	0
-	2009	ÎLE-DE-FRANCE	1	0
44	2009	LA GUAJIRA	0.7	-
-	2009	LORETO	1	0
47	2009	MAGDALENA	1	-
-	2009	MÉRIDA	1	0
50	2009	META	0.8	-
-	2009	METROPOLITANA	1	0
52	2009	NARIÑO	1	-
54	2009	NORTE DE SANTANDER	0.7	-
-	2009	PANAMÁ	1	0
86	2009	PUTUMAYO	1	-
63	2009	QUINDIO	0.7	-
66	2009	RISARALDA	0.7	-
88	2009	SAN ANDRES Y PROVIDENCIA	1	0
68	2009	SANTANDER	0.8	-
-	2009	SIN INFORMACIÓN	1	0
-	2009	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2009	SUCRE	1	-
-	2009	TÁCHIRA	1	-
73	2009	TOLIMA	0.7	0
-	2009	TRUJILLO	1	0
76	2009	VALLE DEL CAUCA	0.7	-
97	2009	VAUPÉS	1	-
99	2009	VICHADA	0.8	-
-	2009	ZULIA	1	-
91	2010	AMAZONAS	0.8	-
5	2010	ANTIOQUIA	0.7	0
81	2010	ARAUCA	1	-
8	2010	ATLÁNTICO	0.8	0
11	2010	BOGOTÁ D.C	0.9	0
13	2010	BOLÍVAR	0.8	-

15	2010	BOYACÁ	0.7	0
17	2010	CALDAS	0.7	0
18	2010	CAQUETÁ	1	-
-	2010	CARABOBO	1	0
85	2010	CASANARE	0.7	-
19	2010	CAUCA	0.8	-
20	2010	CESAR	0.8	-
27	2010	CHOCÓ	1	-
23	2010	CÓRDOBA	0.8	-
-	2010	CORTÉS	1	-
25	2010	CUNDINAMARCA	0.8	0
-	2010	DISTRITO CAPITAL	1	-
-	2010	FALCÓN	1	0
94	2010	GUAINÍA	1	-
95	2010	GUAVIARE	1	-
41	2010	HUILA	0.7	-
-	2010	IMBABURA	1	0
44	2010	LA GUAJIRA	0.8	-
47	2010	MAGDALENA	0.8	-
-	2010	MÉRIDA	1	0
50	2010	META	0.7	-
-	2010	MONAGAS	1	-
52	2010	NARIÑO	1	-
-	2010	NEW YORK	1	0
54	2010	NORTE DE SANTANDER	0.7	-
-	2010	PANAMÁ	1	0
-	2010	PICHINCHA	1	0
86	2010	PUTUMAYO	1	-
63	2010	QUINDIO	0.7	0
66	2010	RISARALDA	0.8	0
88	2010	SAN ANDRES Y PROVIDENCIA	1	0
68	2010	SANTANDER	0.7	-
-	2010	SIN INFORMACIÓN	1	0
70	2010	SUCRE	1	-
-	2010	SUCUMBÍOS	1	0
-	2010	TÁCHIRA	1	0
73	2010	TOLIMA	0.7	0
-	2010	TRUJILLO	1	0
76	2010	VALLE DEL CAUCA	0.6	0
97	2010	VAUPÉS	1	-
99	2010	VICHADA	1	-
-	2010	ZULIA	1	-
0	2011	(NA)	1	-
91	2011	AMAZONAS	0.8	-
5	2011	ANTIOQUIA	0.8	0
81	2011	ARAUCA	1	-
8	2011	ATLÁNTICO	0.9	0
11	2011	BOGOTÁ D.C	0.9	0
13	2011	BOLÍVAR	1	-

15	2011	BOYACÁ	0.7	0
17	2011	CALDAS	0.7	0
18	2011	CAQUETÁ	1	-
-	2011	CARABOBO	1	-
85	2011	CASANARE	1	-
19	2011	CAUCA	1	-
20	2011	CESAR	0.8	-
27	2011	CHOCÓ	1	-
-	2011	COJEDES	1	0
23	2011	CÓRDOBA	0.8	-
25	2011	CUNDINAMARCA	0.8	0
-	2011	DISTRITO FEDERAL	1	-
94	2011	GUAINÍA	1	-
95	2011	GUAVIARE	1	-
-	2011	GUAYAS	1	0
41	2011	HUILA	0.7	0
-	2011	IMBABURA	1	0
44	2011	LA GUAJIRA	0.7	-
-	2011	LIMA	1	0
47	2011	MAGDALENA	0.8	-
50	2011	META	0.7	0
-	2011	MÉXICO	1	0
52	2011	NARIÑO	0.8	-
54	2011	NORTE DE SANTANDER	0.8	-
-	2011	PANAMÁ	1	0
-	2011	PICHINCHA	1	0
86	2011	PUTUMAYO	1	-
63	2011	QUINDIO	0.7	0
66	2011	RISARALDA	0.7	0
88	2011	SAN ANDRES Y PROVIDENCIA	1	0
-	2011	SANTA CRUZ	1	0
68	2011	SANTANDER	0.8	-
-	2011	SIN INFORMACIÓN	1	0
70	2011	SUCRE	1	-
-	2011	TÁCHIRA	1	0
73	2011	TOLIMA	0.8	0
76	2011	VALLE DEL CAUCA	0.6	0
97	2011	VAUPÉS	1	-
99	2011	VICHADA	0.8	-
-	2011	ZULIA	1	-
91	2012	AMAZONAS	0.8	-
5	2012	ANTIOQUIA	0.8	0
-	2012	APURE	1	0
-	2012	ARAGUA	1	0
81	2012	ARAUCA	1	-
8	2012	ATLÁNTICO	0.8	0
11	2012	BOGOTÁ D.C	0.9	0
13	2012	BOLÍVAR	0.7	-
15	2012	BOYACÁ	0.7	0



-	2012	BRITISH COLUMBIA	1	-
17	2012	CALDAS	0.7	0
18	2012	CAQUETÁ	1	-
-	2012	CARCHI	1	-
85	2012	CASANARE	0.7	-
19	2012	CAUCA	1	-
20	2012	CESAR	0.8	-
27	2012	CHOCÓ	0.8	-
23	2012	CÓRDOBA	1	-
-	2012	CORTÉS	1	0
25	2012	CUNDINAMARCA	0.8	0
-	2012	DISTRITO CAPITAL	1	-
-	2012	DISTRITO FEDERAL	1	0
-	2012	ESMERALDAS	1	0
94	2012	GUAINÍA	1	-
95	2012	GUAVIARE	1	-
41	2012	HUILA	0.8	0
-	2012	ÎLE-DE-FRANCE	1	0
-	2012	JALISCO	1	0
44	2012	LA GUAJIRA	0.7	-
47	2012	MAGDALENA	0.7	-
-	2012	MÉRIDA	1	0
50	2012	META	0.8	-
-	2012	MÉXICO	1	0
-	2012	MICHOACÁN	1	0
52	2012	NARIÑO	0.8	-
54	2012	NORTE DE SANTANDER	0.8	-
-	2012	PANAMÁ	1	0
-	2012	PICHINCHA	1	0
86	2012	PUTUMAYO	1	-
63	2012	QUINDIO	0.8	0
66	2012	RISARALDA	0.8	0
88	2012	SAN ANDRES Y PROVIDENCIA	1	0
68	2012	SANTANDER	1	0
-	2012	SIN INFORMACIÓN	1	0
-	2012	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2012	SUCRE	1	0
-	2012	TÁCHIRA	1	0
73	2012	TOLIMA	0.8	-
-	2012	TRUJILLO	1	0
76	2012	VALLE DEL CAUCA	0.8	0
97	2012	VAUPÉS	1	-
99	2012	VICHADA	1	-
-	2012	VIRGINIA	1	0
-	2012	YARACUY	1	0
-	2012	ZULIA	1	0
91	2013	AMAZONAS	0.8	-
5	2013	ANTIOQUIA	0.7	-
81	2013	ARAUCA	1	0

8	2013	ATLÁNTICO	0.8	0
11	2013	BOGOTÁ D.C	0.9	0
13	2013	BOLÍVAR	0.7	0
-	2013	BONAIRE	1	0
15	2013	BOYACÁ	0.8	0
17	2013	CALDAS	0.8	0
18	2013	CAQUETÁ	1	-
85	2013	CASANARE	0.8	-
19	2013	CAUCA	1	-
20	2013	CESAR	0.8	-
27	2013	CHOCÓ	1	-
23	2013	CÓRDOBA	1	-
25	2013	CUNDINAMARCA	0.8	0
-	2013	ESMERALDAS	1	0
-	2013	FLORIDA	1	0
94	2013	GUAINÍA	1	-
95	2013	GUAVIARE	1	0
-	2013	GUAYAS	1	0
41	2013	HUILA	0.7	0
44	2013	LA GUAJIRA	0.7	-
-	2013	LA PAZ	1	0
-	2013	LARA	1	0
-	2013	LIMA	1	0
47	2013	MAGDALENA	0.7	0
50	2013	META	0.7	-
-	2013	METROPOLITANA	-	0
52	2013	NARIÑO	1	-
54	2013	NORTE DE SANTANDER	0.8	-
-	2013	NORTH CAROLINA	1	0
-	2013	PANAMÁ	1	0
-	2013	PICHINCHA	1	0
86	2013	PUTUMAYO	1	-
63	2013	QUINDIO	0.8	0
66	2013	RISARALDA	0.8	0
88	2013	SAN ANDRES Y PROVIDENCIA	1	0
68	2013	SANTANDER	1	-
-	2013	SÃO PAULO	1	0
-	2013	SIN INFORMACIÓN	1	0
-	2013	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2013	SUCRE	1	-
-	2013	TÁCHIRA	1	0
73	2013	TOLIMA	0.8	0
76	2013	VALLE DEL CAUCA	0.8	0
97	2013	VAUPÉS	1	-
99	2013	VICHADA	0.7	-
0	2014	(NA)	1	0
91	2014	AMAZONAS	1	-
5	2014	ANTIOQUIA	0.8	0
81	2014	ARAUCA	1	-

8	2014	ATLÁNTICO	0.8	0
-	2014	BARINAS	1	0
11	2014	BOGOTÁ D.C	0.9	0
13	2014	BOLÍVAR	0.7	0
15	2014	BOYACÁ	0.8	0
17	2014	CALDAS	0.8	0
18	2014	CAQUETÁ	1	-
85	2014	CASANARE	0.7	0
19	2014	CAUCA	1	-
20	2014	CESAR	0.8	0
27	2014	CHOCÓ	1	-
-	2014	COCLÉ	1	0
23	2014	CÓRDOBA	1	-
25	2014	CUNDINAMARCA	0.8	0
-	2014	CUSCO	1	0
-	2014	DISTRITO FEDERAL	1	0
-	2014	ESMERALDAS	1	0
94	2014	GUAINÍA	1	-
95	2014	GUAVIARE	1	-
-	2014	GUAYAS	1	0
-	2014	HESSEN	1	-
41	2014	HUILA	0.7	0
44	2014	LA GUAJIRA	0.7	0
-	2014	LAMBAYEQUE	1	0
-	2014	LIMA	1	0
47	2014	MAGDALENA	0.7	0
-	2014	MANABÍ	1	0
50	2014	META	0.7	-
-	2014	MÉXICO	1	0
-	2014	MIRANDA	1	0
52	2014	NARIÑO	0.7	-
54	2014	NORTE DE SANTANDER	0.5	0
-	2014	PICHINCHA	1	0
86	2014	PUTUMAYO	1	-
63	2014	QUINDIO	0.8	0
66	2014	RISARALDA	0.8	0
88	2014	SAN ANDRES Y PROVIDENCIA	1	0
-	2014	SANTA CRUZ	1	0
68	2014	SANTANDER	1	-
-	2014	SIN INFORMACIÓN	1	0
-	2014	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2014	SUCRE	1	-
-	2014	TÁCHIRA	1	0
-	2014	TARAPACÁ	1	0
73	2014	TOLIMA	0.8	0
76	2014	VALLE DEL CAUCA	0.7	0
99	2014	VICHADA	0.7	-
-	2014	ZULIA	1	-
91	2015	AMAZONAS	0.8	-

5	2015	ANTIOQUIA	0.7	0
81	2015	ARAUCA	1	-
8	2015	ATLÁNTICO	0.8	0
-	2015	BARINAS	1	-
11	2015	BOGOTÁ D.C	0.9	0
13	2015	BOLÍVAR	0.7	-
15	2015	BOYACÁ	0.8	0
17	2015	CALDAS	0.8	0
18	2015	CAQUETÁ	1	-
85	2015	CASANARE	0.8	-
19	2015	CAUCA	1	-
20	2015	CESAR	1	-
27	2015	CHOCÓ	1	-
23	2015	CÓRDOBA	1	-
25	2015	CUNDINAMARCA	0.8	0
-	2015	DISTRITO CAPITAL	1	-
-	2015	DISTRITO FEDERAL	1	0
-	2015	FALCÓN	1	0
-	2015	FLORIDA	1	0
95	2015	GUAVIARE	1	-
-	2015	GUERRERO	1	-
41	2015	HUILA	0.8	0
-	2015	ÎLE-DE-FRANCE	1	-
-	2015	IMBABURA	1	0
44	2015	LA GUAJIRA	0.7	-
-	2015	LIMA	1	0
47	2015	MAGDALENA	0.6	0
-	2015	MÉRIDA	1	0
50	2015	META	0.8	-
-	2015	METROPOLITANA	1	0
52	2015	NARIÑO	1	0
54	2015	NORTE DE SANTANDER	0.8	-
-	2015	PANAMÁ	1	0
86	2015	PUTUMAYO	1	-
63	2015	QUINDIO	0.7	0
-	2015	RIO DE JANEIRO	1	-
66	2015	RISARALDA	0.8	0
88	2015	SAN ANDRES Y PROVIDENCIA	1	-
68	2015	SANTANDER	1	0
-	2015	SIN INFORMACIÓN	1	-
-	2015	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2015	SUCRE	1	-
-	2015	TÁCHIRA	1	0
73	2015	TOLIMA	0.8	0
76	2015	VALLE DEL CAUCA	0.8	0
99	2015	VICHADA	0.8	-
-	2015	ZULIA	1	0
91	2016	AMAZONAS	0.8	-
5	2016	ANTIOQUIA	0.8	0

-	2016	APURE	1	0
81	2016	ARAUCA	1	-
8	2016	ATLÁNTICO	0.8	0
11	2016	BOGOTÁ D.C	0.9	0
13	2016	BOLÍVAR	1	-
15	2016	BOYACÁ	0.8	0
17	2016	CALDAS	0.8	0
18	2016	CAQUETÁ	1	-
85	2016	CASANARE	0.8	-
19	2016	CAUCA	1	-
20	2016	CESAR	1	-
27	2016	CHOCÓ	1	-
23	2016	CÓRDOBA	1	-
25	2016	CUNDINAMARCA	0.8	0
95	2016	GUAVIARE	1	-
-	2016	GUAYAS	1	-
41	2016	HUILA	0.8	0
44	2016	LA GUAJIRA	0.6	-
-	2016	LIMA	1	0
-	2016	LOJA	1	0
47	2016	MAGDALENA	0.7	-
50	2016	META	1	-
-	2016	METROPOLITANA	1	0
-	2016	MIRANDA	1	0
-	2016	MORELOS	1	0
52	2016	NARIÑO	1	-
54	2016	NORTE DE SANTANDER	0.8	-
-	2016	NUEVO LEÓN	1	0
-	2016	PANAMÁ	1	0
-	2016	PICHINCHA	1	0
86	2016	PUTUMAYO	1	0
63	2016	QUINDIO	0.8	0
-	2016	RIO DE JANEIRO	1	0
66	2016	RISARALDA	0.9	0
88	2016	SAN ANDRES Y PROVIDENCIA	1	0
-	2016	SAN MARCOS	1	0
68	2016	SANTANDER	1	0
-	2016	SÃO PAULO	1	0
-	2016	SIN INFORMACIÓN	1	0
-	2016	SIN INFORMACIÓN (VENEZUELA)	1	0
70	2016	SUCRE	1	-
-	2016	TÁCHIRA	1	0
73	2016	TOLIMA	0.8	0
76	2016	VALLE DEL CAUCA	0.8	0
97	2016	VAUPÉS	1	-
99	2016	VICHADA	0.8	-
5	2017	ANTIOQUIA	1	-
81	2017	ARAUCA	1	-
11	2017	BOGOTÁ D.C	1	-

13	2017	BOLÍVAR	1	-
15	2017	BOYACÁ	1	-
17	2017	CALDAS	1	-
18	2017	CAQUETÁ	1	-
19	2017	CAUCA	1	-
27	2017	CHOCÓ	1	-
23	2017	CÓRDOBA	1	-
95	2017	GUAVIARE	1	-
41	2017	HUILA	1	-
44	2017	LA GUAJIRA	1	-
47	2017	MAGDALENA	1	-
54	2017	NORTE DE SANTANDER	1	-
86	2017	PUTUMAYO	1	-
63	2017	QUINDIO	1	-
66	2017	RISARALDA	1	-
68	2017	SANTANDER	1	-
73	2017	TOLIMA	1	-
76	2017	VALLE DEL CAUCA	1	-
5	2018	ANTIOQUIA	1	-
8	2018	ATLÁNTICO	1	-
15	2018	BOYACÁ	1	-
85	2018	CASANARE	1	-
41	2018	HUILA	1	-
50	2018	META	1	-
52	2018	NARIÑO	1	-
66	2018	RISARALDA	1	-
68	2018	SANTANDER	1	-
-	2018	TÁCHIRA	1	0
76	2018	VALLE DEL CAUCA	1	-
5	2019	ANTIOQUIA	1	-
81	2019	ARAUCA	1	-
11	2019	BOGOTÁ D.C	1	-
95	2019	GUAVIARE	1	-
-	2019	MÉRIDA	1	0
52	2019	NARIÑO	1	-
63	2019	QUINDIO	1	-
68	2019	SANTANDER	1	-
-	2019	SIN INFORMACIÓN	1	0
73	2019	TOLIMA	1	1
76	2019	VALLE DEL CAUCA	1	1
5	2020	ANTIOQUIA	1	1
15	2020	BOYACÁ	1	1
17	2020	CALDAS	1	1
18	2020	CAQUETÁ	-	-
19	2020	CAUCA	1	1
20	2020	CESAR	1	1
23	2020	CÓRDOBA	1	1
50	2020	META	1	1
54	2020	NORTE DE SANTANDER	1	1

68	2020	SANTANDER	1	1
76	2020	VALLE DEL CAUCA	1	1

### 9.4.3. Filtros adicionales

```
cat(readLines(files$merge), sep = '\n')
```

```
## #!/usr/bin/env Rscript --vanilla
## # vim:set expandtab ts=4 sw=4 ai fileencoding=utf-8:
## #
## # Author: PB
## # Maintainer(s): PB
## # License: (c) HRDAG 2022, GPL v2 or newer
## #
## # -----
## # CO-SIVJNR-data/match/fase4/merge/src/merge.R
## #
## pacman::p_load(argparse, arrow, logger, tidyr, assertr, stringr,
##                dplyr, here, data.table, nnet)
## #
## stopifnot(str_detect(Sys.getlocale(), "en_US.UTF-8"))
## stopifnot(str_detect(getwd(), "merge$"))
## set.seed(19481210)
## log_threshold(DEBUG)
## #
## vln_types <- c("exilio", "homicidio", "secuestro",
##               "desaparicion", "desplazamiento", "reclutamiento")
## #
## perps <- c("AGENTES_ESTATALES", "GRUPOS_POSDESMV_PARAMILITAR",
##            "GUERRILLA", "GUERRILLA_ELN", "GUERRILLA_FARC",
##            "GUERRILLA_OTRA", "OTRO", "PARAMILITARES")
## #
## to_rename <- c("yy_hecho", "mm_hecho", "dd_hecho",
##               "dept_code_hecho", "muni_code_hecho",
##               paste0("perp_", perps))
## #
## recflds <- c(
##   "recordid",
##   "sexo",
##   "edad",
##   "edad_categoria",
##   to_rename)
## #
## #
## getmode <- function(x) {
##   # from StackOverflow, w nnet::which.is.max to break ties at random
##   # * prioritizes
##   #   full munis (e.g., 5012) over dptos w no muni (e.g., 5000)
##   #   full dates (e.g., 20010626) over ymds w no month-day (e.g., 20010000)
##   if (length(x) == 1) return(x[1])
##   x <- x[!is.na(x)]
## }
```

```

##   if (length(x) == 1) return(x[1])
##   x2 <- str_subset(x, pattern="000$", negate=TRUE)
##   if (length(x2) == 1) return(x2[1])
##   if (length(x2) > 1) x <- x2
##   if (length(x) == 2) return(sample(x, size = 1))
##   ux <- unique(x)
##   if (length(ux) == 1) return(ux[1])
##   if (length(ux) == 2) return(sample(ux, size = 1))
##   ux[which.is.max(tabulate(match(x, ux)))]
## }
##
##
##
## getargs <- function() {
##   # NB: we only need to pass the output.
##   parser <- argparse::ArgumentParser()
##   parser$add_argument("--input_records",
##                       default = here("match/fase4/import/output/input-records.parquet"))
##   parser$add_argument("--match_groups",
##                       default = here("match/fase4/cluster-py/output/match-groups.parquet"))
##   parser$add_argument("--fgn_to_filter",
##                       default = here::here("individual/FGN/homcon/output/fgn-to-filter.parquet"))
##   parser$add_argument("--fgn",
##                       default = here::here("individual/FGN/export/output/fgn.parquet"))
##   parser$add_argument("--cnmh_to_filter",
##                       default = here::here("individual/CNMH/filter-archivado/output/cnmh-to-filter"))
##   parser$add_argument("--rpv",
##                       default = here("match/fase4/merge/output/recordid-p-violation.parquet"))
##   parser$add_argument("--vtype",
##                       default = "promoted")
##   parser$add_argument("--output",
##                       default = here("match/fase4/merge/output/mrd-promoted.parquet"))
##   args <- parser$parse_args()
##
##   args$vtype <- str_sub(str_extract(args$output, "-[a-z]+\\.\\.\\."), 2, -2)
##   args$log <- paste0("output/mrd-", args$vtype, ".log")
##   args
## }
##
##
##
## get_ir <- function(args) {
##   log_info("data read for {args$vtype} begins")
##   rpv <- read_parquet(args$rpv) %>%
##     filter(vtype == args$vtype) %>%
##     select(recordid)
##   log_debug("got rpv with {nrow(rpv)} recs")
##   ir <- read_parquet(args$input_records) %>%
##     filter(recordid %in% rpv$recordid) %>%
##     left_join(read_parquet(args$match_groups), by='recordid') %>%
##     assert(not_na, match_group_id) %>%
##     verify(nrow(.) == nrow(rpv))
##   log_info("retained {nrow(ir)} input-recs for vtype={args$vtype}")
##   rm(rpv)
##   ir

```



```

## }
##
## rename_and_select <- function(args, ir, vln) {
##   # rename here xforms the cols from their vln_specific names to canonical
##   # colnames
##   vlnt <- paste0("-", vln)
##   cnmh_to_filter <- read_parquet(args$cnmh_to_filter)
##   fgn_to_filter <- read_parquet(args$fgn_to_filter)
##   rns <- ir %>%
##     anti_join(fgn_to_filter) %>%
##     anti_join(cnmh_to_filter) %>%
##     rename_with(~ str_replace(., vlnt, ""), col=all_of(to_rename)) %>%
##     select(all_of(recflds), is_conflict, match_group_id) %>%
##     mutate(ymd_hecho = str_c(
##       str_pad(str_replace_na(yy_hecho, "0000"), width=4, pad="0"),
##       str_pad(str_replace_na(mm_hecho, "00"), width=2, pad="0"),
##       str_pad(str_replace_na(dd_hecho, "00"), width=2, pad="0")) %>%
##     mutate(ymd_hecho = if_else(startsWith(ymd_hecho, "0"),
##                               paste0(yy_hecho, "0000"),
##                               ymd_hecho) %>%
##     select(-yy_hecho, -mm_hecho, -dd_hecho) %>%
##     mutate(ymd_hecho = na_if(ymd_hecho, "NA0000")) %>%
##     mutate(dept_code_hecho = replace_na(dept_code_hecho, 0)) %>%
##     mutate(muni_code_hecho = if_else(is.na(muni_code_hecho),
##                                     paste0(dept_code_hecho, "000"),
##                                     as.character(muni_code_hecho))) %>%
##     mutate(muni_code_hecho = na_if(muni_code_hecho, "0000")) %>%
##     select(-dept_code_hecho, -recordid) %>%
##     mutate(across(c(edad), as.character))
##   log_debug("rename_and_select returning {nrow(rns)} for {vln}")
##   rns
## }
##
## merge_mg <- function(irs) {
##   log_debug("starting merge")
##   irm <- irs %>%
##     mutate(across(starts_with("perp_"), ~ replace_na(., -1))) %>%
##     group_by(match_group_id) %>%
##     summarize(
##       sexo = getmode(sexo),
##       edad = getmode(edad),
##       edad_categoria = getmode(edad_categoria),
##       muni_code_hecho = getmode(muni_code_hecho),
##       ymd_hecho = min(ymd_hecho, na.rm=TRUE),
##       is_conflict = max(is_conflict, na.rm=TRUE),
##       perp_AGENTES_ESTATALES = max(perp_AGENTES_ESTATALES),
##       perp_GRUPOS_POSDESMV_PARAMILITAR = max(perp_GRUPOS_POSDESMV_PARAMILITAR),
##       perp_GUERRILLA = max(perp_GUERRILLA),
##       perp_GUERRILLA_ELN = max(perp_GUERRILLA_ELN),
##       perp_GUERRILLA_FARC = max(perp_GUERRILLA_FARC),
##       perp_GUERRILLA_OTRA = max(perp_GUERRILLA_OTRA),
##       perp_OTRO = max(perp_OTRO),
##       perp_PARAMILITARES = max(perp_PARAMILITARES)) %>%
##     ungroup() %>%

```

```

##       mutate(across(starts_with("perp_"), ~ na_if(., -1)))
##       log_debug("summarize complete")
##       irm
## }
##
## #---main-----
## args <- getargs()
## log_info("path & locale ok, seed set, log open, run begins.")
##
## ir <- get_ir(args)
## log_info("read {nrow(args)} recs from input-records")
##
## if (args$vttype == "promoted") {
##   # NB: this pulls the same input-recs 4x, once for each promoted vln type
##   # it has to xfrom the vln-specific cols into canonical cols: wide->long
##   # yy_hecho_{vln}, dept_code_hecho_{vln}, perp_AGENTES_ESTATALES_{vln} etc.
##   # once the vln-specific cols are in long form, they can be collapsed on mgi
##   irs <- bind_rows(
##     rename_and_select(args, ir, "homicidio"),
##     rename_and_select(args, ir, "reclutamiento"),
##     rename_and_select(args, ir, "secuestro"),
##     rename_and_select(args, ir, "desaparicion")
##   ) else {
##     irs <- rename_and_select(args, ir, args$vttype)
##   }
##   log_info("irs has {nrow(irs)} recs.")
##
##   # irs has canonically named cols so can do a collapse w merge_mg()
##   irm <- merge_mg(irs) %>%
##     mutate(muni_code = as.numeric(muni_code_hecho)) %>%
##     mutate(dept_code_hecho = as.integer(muni_code/1000)) %>%
##     mutate(psum = rowSums(across(starts_with("perp_")))) %>%
##     mutate(ymd_int = as.numeric(ymd_hecho)) %>%
##     mutate(yy_hecho = as.integer(ymd_int/10000)) %>%
##     write_parquet(args$output)
##   log_info("done with merge, output written")
##   log_info("running tests now")
##
##   # For testing only
##   # test that the perps were not conflated
##   stopifnot(sum(irm$perp_AGENTES_ESTATALES, na.rm=T) != sum(irm$perp_PARAMILITARES, na.rm=T))
##   # test that >80% of irm recs (w known perp) have 1 perp
##   stopifnot(between(mean(irm$psum, na.rm=T), 1.01, 2))
##   stopifnot(0 <= min(irm$psum, na.rm=T) & max(irm$psum, na.rm=T) <= 8)
##   log_info("perp tests ok: are separate, not conflated, no overcounts")
##
##   x <- irm %>%
##     assert(within_bounds(1000, 99999), muni_code) %>%
##     select(-muni_code) %>%
##     assert(within_bounds(19000000, 20200000), ymd_int) %>%
##     select(-ymd_int) %>%
##     assert(not_na, yy_hecho, dept_code_hecho)
##   rm(x)

```

```
## log_info("test ok: muni_code ok, ymd ok, not_na yy+dept")
##
## log_info("done.")
## # done.
```

```
cat(readLines(files$makei), sep = '\n')
```

```
## #!/usr/bin/env Rscript --vanilla
## # vim: set expandtab ts=4 sw=4 ai fileencoding=utf-8
## #
## # Author: PB
## # Maintainer(s): PB
## # License: (c) HRDAG 2021, GPL v2 or newer
## #
## # -----
## # CO-SIVJNRN-data/match/fase4/merge/src/make-i-vecs.R
## #
##
## pacman::p_load(argparse, arrow, logger, tidyr, assertr, stringr, yaml,
##               purrr, dplyr, glue, here, data.table)
## stopifnot(str_detect(Sys.getlocale(), "en_US.UTF-8"))
## stopifnot(str_detect(getwd(), "merge$"))
## set.seed(19481210)
##
##
## vln_types <- c("exilio", "desplazamiento",
##               "homicidio", "secuestro", "desaparicion", "reclutamiento")
## promoted_vlns <- vln_types[3:length(vln_types)]
##
## getargs <- function() {
##   parser <- argparse::ArgumentParser()
##   parser$add_argument("--recordid_p_violation",
##                       default = "output/recordid-p-violation.parquet")
##   parser$add_argument("--match_groups",
##                       default = here("match/fase4/cluster-py/output/match-groups.parquet"))
##   parser$add_argument("--match_group_violation",
##                       default = here::here("match/fase4/merge/output/match-group-violation.parquet")
##   parser$add_argument("--cnmh_to_filter",
##                       default = here::here("individual/CNMH/filter-archivado/output/cnmh-to-filter")
##   parser$add_argument("--fgn_to_filter",
##                       default = here::here("individual/FGN/homcon/output/fgn-to-filter.parquet"))
##   parser$add_argument("--fgn",
##                       default = here::here("individual/FGN/export/output/fgn.parquet"))
##   parser$add_argument("--src2in",
##                       default = here::here("match/fase4/merge/hand/src2in.yaml"))
##   parser$add_argument("--output",
##                       default = "output/inclusion-vecs-desaparicion.parquet")
##   args <- parser$parse_args()
##   args$vln <- str_sub(str_extract(args$output, "-[a-z]+\\."), 2, -2)
##   stopifnot(args$vln %in% vln_types)
##   args$log <- paste0("output/make-i-vecs-", args$vln, ".log")
##   args
## }
##
```

```

##
## # --- main -----
## args <- getargs()
##
## log_appender(appender_tee(args$log))
## log_threshold(DEBUG)
## log_info("path & locale ok, seed set, log open, {args$vln} run begins.")
##
## #---setup-----
## in_sources <- yaml.load_file(args$src2in)
## # need to test that all sources in input_records are mapped in in_sources!
##
## if (args$vln %in% promoted_vlns) {
##   target_vln <- "promoted"
## } else {
##   target_vln <- args$vln
## }
##
## # filter for unusable recs here
## cnmh_to_filter <- read_parquet(args$cnmh_to_filter)
## log_info("read {nrow(cnmh_to_filter)} rows to filter from CNMH")
## fgn_to_filter <- read_parquet(args$fgn_to_filter)
## log_info("read {nrow(fgn_to_filter)} rows to filter from FGN")
##
## # recordids, filtered to the appropriate violation level, with source
## input_records <- read_parquet(args$recordid_p_violation) %>%
##   filter(vtype == target_vln) %>%
##   mutate(source = str_replace(source, regex("-"), "_")) %>%
##   select(recordid, source) %>%
##   assert(is_uniq, recordid)
## log_info("read {nrow(input_records)} input-records of type {target_vln}")
##
## input_records <- input_records %>%
##   anti_join(fgn_to_filter)
## log_info("after FGN filter, {nrow(input_records)} input-records retained")
##
## input_records <- input_records %>%
##   anti_join(cnmh_to_filter)
## log_info("after CNMH filter, {nrow(input_records)} input-records retained")
##
## stopifnot(input_records %>% filter(recordid %in% fgn_to_filter$recordid) %>% nrow(.) == 0)
## stopifnot(input_records %>% filter(recordid %in% cnmh_to_filter$recordid) %>% nrow(.) == 0)
## log_info("input_records has no recs that should be filtered")
##
## # match-group-ids, filtered for vln == match-group's _already promoted_vln
## # vln is in {homicidio, reclutamiento, secuestro, desaparicion}
## # match_group_violation filters match_groups for the appropriate vln type
## match_groups <- read_parquet(args$match_group_violation) %>%
##   filter(violation == args$vln) %>%
##   select(match_group_id) %>%
##   distinct() %>%
##   left_join(read_parquet(args$match_groups)) %>%
##   distinct()
## n_mgs <- length(unique(match_groups$match_group_id))

```

```

## log_info("read {nrow(match_groups)} recordids in {n_mgs} match-groups for {args$vlN}")
##
## # filtering. Note that when we're filtering the promoted recs, the number of recs
## # filtered might be greater than the number expected for a specific violation.
## match_groups <- match_groups %>%
##   anti_join(fgn_to_filter)
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("after FGN filter, {nrow(match_groups)} recordids and {n_mgs} match_groups retained")
## match_groups <- match_groups %>%
##   anti_join(cnmh_to_filter)
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("after CNMH filter, {nrow(match_groups)} recordids and {n_mgs} match_groups retained")
## n_mgs <- length(unique(match_groups$match_group_id))
## log_info("read {nrow(match_groups)} recordids in {n_mgs} match-groups after FGN- and CNMH-to-filter :")
## stopifnot(match_groups %>% filter(recordid %in% fgn_to_filter$recordid) %>% nrow(.) == 0)
## stopifnot(match_groups %>% filter(recordid %in% cnmh_to_filter$recordid) %>% nrow(.) == 0)
## log_info("match_groups has no recs that should be filtered")
##
## # in this match, some match-groups include recordids that need to be filtered
## # because they're not of the correct type. This is a consequence of the crude
## # filtering in the match_groups step: recall that (match_group_id, violation)
## # in match-group-violation is *not unique*. Within a match_group_id there
## # may be several violations. the filter in the step below removes the irrelevant
## # recordids.
## in_vecs <- match_groups %>%
##   inner_join(input_records, by="recordid") %>%
##   filter(!is.na(source))
## log_info("in_vecs initialized w {nrow(in_vecs)}, before in_ vars and summary")
##
## stopifnot(in_vecs %>% filter(recordid %in% fgn_to_filter$recordid) %>% nrow(.) == 0)
## stopifnot(in_vecs %>% filter(recordid %in% cnmh_to_filter$recordid) %>% nrow(.) == 0)
## log_info("in_vecs has no recs that should be filtered")
## in_vecs <- in_vecs %>% select(-recordid)
## rm(input_records, match_groups)
##
##
## log_info("starting str_detect:")
## for (i in seq(length(in_sources))) {
##   in_srcs <- in_sources[[i]]
##   in_srcs_char <- str_c(in_srcs, collapse = ', ')
##   in_name <- paste0("in_", names(in_sources)[i])
##   log_info("{in_srcs_char} to {in_name}")
##   in_vecs <- in_vecs %>%
##     select(source) %>%
##     mutate(in_ = as.integer(source %in% in_srcs)) %>%
##     select(-source) %>%
##     rename(!in_name := in_) %>%
##     bind_cols(in_vecs)
## }
##
## in_vecs <- in_vecs %>% select(-source) %>% as.data.table
## log_info("as.data.table.")
## # if (args$vlN != "homicidio") {
## #   in_vecs <- in_vecs %>% filter(in_FGN == 0) %>% select(-in_FGN)

```

```
## # }
##
## log_info("starting sum by match_group_id")
## in_vecs <- in_vecs[, lapply(.SD, sum), by=match_group_id]
##
## print(glue("sum(in_FGN) == {sum(in_vecs$in_FGN)}"))
## # if (args$vlm == "homicidio") {
## #   stopifnot("in_FGN" %in% colnames(in_vecs))
## # } else {
## #   print(sum(in_vecs$in_FGN))
## #   stopifnot(! "in_FGN" %in% colnames(in_vecs))
## # }
##
## in_vecs %>% write_parquet(args$output)
## log_info("wrote {nrow(in_vecs)} to {args$output}; done.")
## # done.
```

## Referencias

- Akande, Olanrewaju Michael. 2015. "A Comparison of Multiple Imputation Methods for Categorical Data." Master's thesis, Duke University.
- Amorós, Jaume. 2014. "Recapturing Laplace." *Significance*. <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2014.00754.x>.
- Ball, Patrick. 2008. *Quién Le Hizo Qué a Quién*. Benetech.
- . 2016. "The Task Is a Quantum of Workflow." <https://hrdag.org/2016/06/14/the-task-is-a-quantum-of-workflow/>.
- Ball, Patrick, and Megan Price. 2018. "The Statistics of Genocide." *CHANCE* 31 (1): 38–45.
- . 2019. "Using Statistics to Assess Lethal Violence in Civil and Inter-State War." *Annual Review of Statistics and Its Application* 6: 63–84.
- Ball, Patrick, and Michael Reed Hurtado. 2015. "Exploración y Análisis de Los Datos Para Comprender La Realidad." *Forensis 2014* 16 (1): 529–45.
- . 2016. "El Registro y La Medición de La Criminalidad. El Problema de Los Datos Faltantes y El Uso de La Ciencia Para Producir Estimaciones En Relación Con El Homicidio En Colombia, Demostrado a Partir de Un Ejemplo: El Departamento de Antioquia (2003-2011)." *Revista Criminalidad* 58 (1): 9–23.
- Berchenko, Yakir, Yael Artzy-Randrup, Mina Teicher, and Lewi Stone. 2009. "Emergence and Size of the Giant Component in Clustered Random Graphs with a Given Degree Distribution." *Phys. Rev. Lett.* 102 (March): 138701. <https://doi.org/10.1103/PhysRevLett.102.138701>.
- Bilenko, Mikail, Beena Kamath, and Raymond J Mooney. 2006. "Adaptive Blocking: Learning to Scale up Record Linkage." In. Proceedings of the 6th IEEE International Conference on Data Mining. <http://bilenko.com/papers/06-icdm.pdf>.
- Bird, Sheila M, and Ruth King. 2018. "Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy." *Annual Review of Statistics and Its Application* 5: 95–118.
- Bishop, Yvonne M M, Stephen E Fienberg, and Paul W Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Springer.
- Burgette, Lane F, and Jerome P Reiter. 2010. "Multiple Imputation for Missing Data via Sequential Regression Trees." *American Journal of Epidemiology* 172 (9): 1070–6.
- Chao, Anne. 1987. "Estimating the Population Size for Capture-Recapture Data with Unequal Capturability." *Biometrics* 43: 783–91.
- . 2001. "An Overview of Closed Capture-Recapture Models." *Journal of Agricultural, Biological, and Environmental Statistics* 6 (2): 158–75.
- . 2015. "Capture-Recapture for Human Populations." *Wiley StatsRef: Statistics Reference Online*, 1–16.
- Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." *CoRR*

- abs/1603.02754. <http://arxiv.org/abs/1603.02754>.
- Christen, Peter. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. New York: Springer.
- Christen, Peter, Dinusha Vatsalan, and Qing Wang. 2015. “Efficient Entity Resolution with Adaptive and Interactive Training Data Selection.” In *2015 IEEE International Conference on Data Mining*. <http://users.cecs.anu.edu.au/~u5170295/papers/icdm-christen-2015.pdf>.
- Cormack, R M. 1989. “Log-Linear Models for Capture-Recapture.” *Biometrics* 45 (2): 395–413. <http://www.jstor.org/stable/2531485>.
- Cro, S. 2017. “Relevant Accessible Sensitivity Analysis for Clinical Trials with Missing Data.” PhD thesis, London School of Hygiene & Tropical Medicine.
- Darroch, John N, Stephen E Fienberg, Gary F V Glonek, and Brian W Junker. 1993. “A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability.” *Journal of the American Statistical Association* 88 (423): 1137–48. <http://www.jstor.org/stable/2290811>.
- Fienberg, Stephen E, M S Johnson, and B W Junker. 1999. “Classical multilevel and Bayesian approaches to population size estimation using multiple lists.” *Journal of the Royal Statistical Society Series A* 162 (3): 383–405. <https://doi.org/10.1111/1467-985X.00143>.
- Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis, Third Edition*. Chapman; Hall/CRC.
- Gelman, Andrew, and E. Loken. 2014. “The Statistical Crisis in Science.” *American Scientist*.
- Gutiérrez-Sanín, F., and E. Wood. 2017. “What Should We Mean by ‘Pattern of Political Violence?’ Repertoire, Targeting, Frequency, and Technique.” *Cambridge Core* 15.
- Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart. 2008. *Exploring Network Structure, Dynamics, and Function Using NetworkX*. Edited by Ga.<sup>el</sup> Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA.
- Herzog, Thomas N, Fritz J Scheuren, and Wiliam E Winkler. 2007. *Data Quality and Record Linkage Techniques*. Springer. <http://www.springer.com/computer/database+management+%26+information+retrieval/book/978-0-387-69502-0>.
- Heymans, M, and I Eekhout. 2019. *Applied Missing Data Analysis with SPSS and (r)studio*. Heymans; Eekhout.
- Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. “Long Short-Term Memory.” *Neural Computation* 9 (8): 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hoover Green, Amelia, and Patrick Ball. 2019. “Civilian Killings and Disappearances During Civil War in El Salvador (1980–1992).” *Demographic Research* 41: 781–814. <https://www.jstor.org/stable/26850667>.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.
- Johndrow, James E, Kristian Lum, and Daniel Manrique-Vallier. 2019. “Low-risk population size estimates in the presence of capture heterogeneity.” *Biometrika* 106 (1): 197–210. <https://doi.org/10.1093/biomet/asy065>.
- Johndrow, James, Kristian Lum, and David Dunson. 2018. “Theoretical Limits of Microclustering for Record Linkage.” *Biometrika* 105: 431–46.
- Kropko, Jonathan, Ben Goodrich, Andrew Gelman, and Jennifer Hill. 2014. “Multiple Imputation for Continuous and Categorical Data: Comparing Joint Multivariate Normal and Conditional Approaches.” *Political Analysis* 22 (4): 497–519. <https://doi.org/10.1093/pan/mpu007>.
- Link, William A. 2004. “Individual Heterogeneity and Identifiability in Capture Recapture Models.” *Animal Biodiversity and Conservation* 27 (1): 87–91.
- Lum, Kristian, Megan Price, and David Banks. 2013. “Applications of Multiple Systems Estimation in Human Rights Research.” *The American Statistician* 67 (4): 191–200.
- Madigan, David, and Jeremy C York. 1997. “Bayesian Methods for Estimation of the Size of a Closed Population.” *Biometrika* 84 (1): 19–31. <http://www.jstor.org/stable/2337552>.
- Manrique-Vallier, Daniel. 2016. “Bayesian Population Size Estimation Using Dirichlet Process Mixtures.” *Biometrics* 72 (4): 1246–54.
- Manrique-Vallier, Daniel, Ball P., and M. Sadinle. 2022. “Capture-Recapture for Casualty Estimation and Beyond: Recent Advances and Research Directions.”
- Michelson, Matthew, and Craig A Knoblock. 2006. “Learning Blocking Schemes for Record Linkage.” In

- Proceedings of the 21st National Conference on Artificial Intelligence (AAAI).*
- Ming Li, Linyuan Lü, Run-Ran Liu, and Yi-Cheng Zhang. 2021. “Percolation on Complex Networks: Theory and Application.” *Physics Reports* 907: 1–68.
- Müllner, Daniel. 2013. “Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for r and Python.” *Journal of Statistical Software* 53 (9): 1–18.
- Price, Gohdes, M., and P. Ball. 2015. “Documents of War: Understanding the Syrian Conflict.” *Royal Statistical Society* 12: 14–19.
- Rivest, Louis-Paul. 2011. “A Lower Bound Model for Multiple Record Systems Estimation with Heterogeneous Catchability.” *The International Journal of Biostatistics* 7 (1). <https://doi.org/10.2202/1557-4679.1283>.
- Rozo Ángel, Valentina, and Patrick Ball. 2019. “Asesinatos de Líderes Sociales En Colombia: Una Estimación Del Universo-Actualización 2018.” Human Rights Data Analysis Group.
- Rubin, Donald B. 1978. “Multiple Imputations in Samples Surveys - a Phenomenological Bayesian Approach to Nonresponse.” *American Statistical Association Proceedings of the Survey Research Methods Section*: 20–34.
- . 1987. *Multiple Imputation for Nonresponse in Surveys*. Wiley.
- . 1996. “Multiple Imputation After 18 Plus Years.” *Journal of the American Statistical Association* 91 (434): 473–89.
- Särndal, Swensson, K. E. 2003. *Model Assisted Survey Sampling*. Springer.
- Scheuren, Fritz. 2004. “History Corner.” *The American Statistician*. <https://www.tandfonline.com/doi/abs/10.1198/0003130042926>.
- Sekar, C Chandra, and W Edwards Deming. 1949. “On a Method of Estimating Birth and Death Rates and the Extent of Registration.” *Journal of the American Statistical Association* 44 (245). <https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1949.10483294>.
- Shah, T. 2019. “rproj Considered Harmful.” <https://hrdag.org/tech-notes/harmful.html>.
- Suárez, Andrés Felipe. 2018. “Dos Registros, Dos Versiones. Los Usos Políticos de Los Registros Oficiales En Las Luchas Por La Verdad de La Guerra En Colombia.” *Nuevo Mundo Mundos Nuevos*.
- van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. Boca Raton: Chapman; Hall/CRC. <https://stefvanbuuren.name/fimd>.
- van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. “Mice: Multivariate Imputation by Chained Equations in r.” *Journal of Statistical Software, Articles* 45 (3): 1–67. <https://doi.org/10.18637/jss.v045.i03>.
- Van Buuren, S., H. C. Boshuizen, and D. L. Knook. 1999. “Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis.” *Statistics in Medicine* 18 (6): 681–94. [https://doi.org/https://doi.org/10.1002/\(SICI\)1097-0258\(19990330\)18:6%3C681::AID-SIM71%3E3.0.CO;2-R](https://doi.org/https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6%3C681::AID-SIM71%3E3.0.CO;2-R).
- White, Ian R, Patrick Royston, and Angela M Wood. 2011. “Multiple Imputation Using Chained Equations: Issues and Guidance for Practice.” *Statistics in Medicine* 30 (4): 377–99. <https://doi.org/https://doi.org/10.1002/sim.4067>.
- Zhou, X, and J. Reiter. 2010. “A Note on Bayesian Inference After Multiple Imputation.” *The American Statistician* 64 (2).