Identificación de factores asociados a la violencia intrafamiliar en Colombia mediante técnicas de reducción de dimensiones y clustering difuso

Andres Felipe Useche Hernandez

Tutoras:

Lina Lucia Hernandez Velasco

Lina Maria Acosta Avena



Pontificia Universidad Javeriana

Facultad de Ciencias

Programa académico de Ciencia de Datos

Bogotá D.C.

2026

Resumen

La violencia intrafamiliar en Colombia constituye una problemática social que a pesar de los esfuerzos normativos vigentes, ha mostrado una tendencia creciente en los últimos años. Este trabajo surge ante la necesidad de complementar los análisis descriptivos que se suelen emplear para comprender las posibles asociaciones entre factores sociodemográficos y contextuales que configuran el fenómeno. El objetivo principal es analizar las dimensiones geográficas, sociodemográficas y contextuales de la violencia intrafamiliar en Colombia mediante técnicas avanzadas que permitan generar insumos para la toma de decisiones orientadas a su prevención y atención. En primer lugar, se aplicaron análisis de tasas de incidencia estandarizadas para describir las diferencias territoriales del fenómeno. Posteriormente, se empleó el Análisis de Correspondencias Múltiple (ACM) con el fin de identificar asociaciones estructurales entre variables y reducir la dimensionalidad de los datos. Finalmente, se implementó la técnica de Fuzzy C-Means para segmentar los casos en perfiles de violencia con fronteras flexibles. El análisis se realizó a partir de reportes medicolegales oficiales correspondientes al año 2024.

Los resultados evidencian la existencia de cuatro tipologías principales: violencia de pareja actual y expareja (54,61%), un perfil familiar mixto articulado por la intolerancia social (27,63%), un perfil de victimización de niños, niñas y adolescentes altamente especializado (4,05%) y violencia en convivencia asociada al consumo de sustancias (13,71%). Estos hallazgos permiten identificar configuraciones de violencia que profundizan las clasificaciones oficiales, facilitando una comprensión de los grupos conformados y de la pertenencia de los reportes a cada clúster en función de sus características sociodemográficas y contextuales.

Palabras clave: Violencia intrafamiliar, análisis de correspondencia múltiple, clustering difuso y Fuzzy C-Means.

Agradecimientos

Con el presente trabajo de grado culmino mis estudios en la Pontificia Universidad Javeriana, etapa en la que agradezco profundamente las experiencias, aprendizajes y oportunidades que la institución me brindó, así como a las personas que hicieron parte de este proceso académico y personal.

En primer lugar, expreso mi agradecimiento a Dios, a mis padres y a mi abuela, por su acompañamiento constante, su apoyo incondicional y por ser el pilar fundamental de mi formación. Sin ellos, este logro no habría sido posible.

De manera especial, agradezco la memoria de mi compañero de vida de cuatro patas, Zeus, quien me acompañó durante gran parte de este camino. Su amor, lealtad y compañía incondicional han sido una fuente permanente de motivación y fortaleza.

Asimismo, extiendo mi gratitud a la Selección de Rugby de la universidad, que me acogió con las puertas abiertas. Este deporte y las personas que conocí en este espacio contribuyeron significativamente a mi desarrollo personal, enseñándome resiliencia, disciplina y trabajo en equipo.

Finalmente, agradezco a mis tutoras Lina Maria Acosta y Lina Lucia Hernandez por su acompañamiento durante este proceso, por su formación de alta calidad y por su apoyo no solo como docentes, sino también como personas. Les expreso mis mejores deseos en su trayectoria profesional y personal.

“En todo amar y servir.”
— San Ignacio de Loyola

Índice

1. Introducción 4

1.1 Problemática 4

1.2 Justificación 4

2. Objetivos 5

2.1 Objetivo general 5

2.2 Objetivos específicos 5

3. Marco teórico 6

3.1 Antecedentes 6

3.2 Marco teórico conceptual 7

3.3 Marco teórico técnico 10

4. Metodología 15

5. Resultados 19

5.1 Distribución espacial y demográfica de las tasas de violencia intrafamiliar en Colombia . 25

5.2 Identificación de factores asociados a la violencia intrafamiliar 27

5.3 Segmentación y caracterización de perfiles mediante clustering difuso 37

6. Conclusiones 45

7. Limitaciones y trabajos futuros 46

8. Referencias 47

9. Anexos 49

1. Introducción

1.1 Problemática

La violencia intrafamiliar es un problema social de gran impacto que incide en diferentes contextos a nivel mundial. Organismos internacionales como la ONU y OMS han mencionado la importancia de combatir este fenómeno debido a sus consecuencias físicas, psicológicas y sociales (Organización Mundial de la Salud [OMS],2022; Organización de las Naciones Unidas [ONU], 2020). En Colombia diferentes entidades mencionan y tratan de tomar medidas para contrarrestar la violencia intrafamiliar como lo es la LEY 294 DE 1996 creada por el Congreso de la República y modificada por distintas leyes a lo largo del tiempo, con el fin de establecer mecanismos de protección, prevención y sanción frente a la violencia intrafamiliar. Esta se entiende como cualquier acto de dominación, sometimiento, control o agresión física, psíquica, sexual, patrimonial o económica por parte de un miembro familiar hacia otro. Esto puede darse fuera o dentro del domicilio familiar (ICBF, 2017).

A pesar de los esfuerzos normativos y de las campañas institucionales, en Colombia la violencia intrafamiliar ha aumentado en los últimos años (Ministerio de Justicia y Derecho, 2024). Este fenómeno afecta de diferentes maneras a hombres, mujeres, niños y adultos mayores; sus efectos no solo se limitan al ámbito individual, sino que también tienen un efecto sobre la estabilidad familiar y el tejido social. No obstante, todavía hay vacíos en la comprensión de los factores que influyen en su ocurrencia y de cómo estos interactúan entre sí en diversos entornos geográficos y socioeconómicos del país. Esta carencia de información impide crear políticas públicas más centradas y eficaces, así que es preciso examinar las características de los contextos en los que ocurren los hechos, las víctimas y los agresores, para proporcionar evidencia empírica que ayude a prevenir y abordar la violencia intrafamiliar de manera integral.

1.2 Justificación

Actualmente existen estudios, cifras oficiales y campañas de prevención sobre el tema. Sin embargo, la mayoría de estas investigaciones se enfocan en describir el fenómeno cualitativamente o, en el caso de los estudios cuantitativos, se limitan al uso de la estadística descriptiva, lo que impide identificar perfiles de riesgo específicos y comprender cómo interactúan las variables sociodemográficas con el contexto del hecho. Esta falta de

profundidad analítica dificulta el diseño de estrategias de prevención que sean realmente focalizadas y eficientes Ejemplo de ello son trabajos como *Factores de riesgo que limitan la denuncia oportuna de la violencia intrafamiliar en mujeres víctimas* (Losada et al., 2020), los reportes elaborados por el Instituto nacional de Medicina Legal y Ciencias Forenses (2006), o el trabajo de grado de la UNAD (Cervantes, 2017) sobre *Factores Psicosociales asociados a la violencia intrafamiliar en sectores de Barranquilla*, esas limitaciones dificultan una gestión preventiva y focalizada por parte de las autoridades y organizaciones que luchan contra este tipo de violencia. Por ende, este trabajo se enfoca en identificar de manera integral los factores asociados a la violencia intrafamiliar a nivel nacional.

De este modo el presente trabajo, a partir de datos oficiales del año 2024 que contienen variables sobre las características de las víctimas, los agresores y el contexto del hecho, se realiza un análisis cuantitativo y territorial a nivel nacional. Con el fin de superar el enfoque descriptivo basado en el conteo de casos, se plantea la aplicación de técnicas de análisis multivariado y aprendizaje de máquina difuso, orientadas a la identificación de factores asociados a la violencia intrafamiliar y la construcción de tipologías empíricas. Este enfoque permite capturar la ambigüedad de las relaciones familiares donde los límites de la violencia no son estrictamente definidos y cuantificar el peso estadístico de los factores que configuran distintos perfiles de agresión.

2. Objetivos

2.1 Objetivo general

Analizar las dimensiones geográficas, sociodemográficas y contextuales de la violencia intrafamiliar en Colombia mediante técnicas avanzadas que permitan generar insumos para la toma de decisiones orientadas a su prevención y atención.

2.2 Objetivos específicos

- Describir la distribución geográfica y demográfica de la violencia intrafamiliar en Colombia.
- Identificar los factores sociodemográficos, contextuales y territoriales asociados a la ocurrencia de violencia intrafamiliar a través de técnicas estadísticas.

- Segmentar y caracterizar a la población estudiada con características similares en torno a la violencia intrafamiliar a través de técnicas de aprendizaje de máquina.

3. Marco teórico

3.1 Antecedentes

Desde una perspectiva institucional, el Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF, 2006) caracterizó inicialmente el fenómeno al evaluar 72.849 víctimas, identificando que el 59,4% de los casos correspondían a violencia de pareja, el 21,9% a otros familiares y el 18,5% a menores de edad. Esta dinámica se actualiza en el boletín *Datos para el cambio* del Instituto Colombiano de Bienestar Familiar (ICBF, 2023), el cual reporta que para 2023 las valoraciones contra la niñez ascendieron a 8.796 casos (23% del total), destacando una vulnerabilidad creciente en adolescentes de 12 a 17 años y señalando a los municipios de Rondón, Busbanzá y Cácota con las tasas de incidencia más altas (pp. 50-55). Pese a la relevancia de estas cifras, ambos documentos mantienen un enfoque descriptivo tradicional que se limita al reporte de frecuencias, omitiendo la exploración de los factores de riesgo multidimensionales que esta investigación busca identificar mediante el uso de modelos analíticos avanzados.

Por otra parte, el Ministerio de Justicia y del Derecho (MinJusticia, 2024) presentó el informe Violencia intrafamiliar en Colombia (2016-2023), en el cual la Policía Nacional exhibe un incremento en los casos de incidentes, pasando de 207,4 incidentes por cada 100.000 habitantes en 2016 a 228,8 en 2023. El reporte destaca que las mujeres representan entre el 70% - 77% de las víctimas, subrayando la urgencia de intervenciones preventivas, lo cual refleja la necesidad de prevención e intervención ante este delito. No obstante, al igual que los documentos anteriores, este informe mantiene un enfoque descriptivo y no profundiza en el análisis de las relaciones o asociaciones entre las variables.

En el ámbito del análisis espacial de la criminalidad en Colombia, Arango et al. (2016) emplearon técnicas de agrupamiento como Fuzzy C-Means (FCM) para analizar trece variables delictivas, entre ellas la violencia intrafamiliar y el homicidio, con el fin de identificar patrones de criminalidad en jóvenes y adultos en la ciudad de Medellín. Su investigación permitió identificar “microcorredores” territoriales donde la alta criminalidad se consolida

progresivamente y evidenció una marcada variabilidad espacial en las dinámicas delictivas. Por ende, el trabajo respalda el uso del clustering difuso al demostrar su capacidad para capturar la complejidad de fenómenos sociales en los que las fronteras entre grupos no son rígidas, aportando así el rigor necesario para avanzar desde la estadística descriptiva tradicional hacia la identificación de tipologías empíricas basadas en la ambigüedad inherente a las características sociodemográficas y de territorio.

Finalmente, se encuentra la investigación realizada en el marco del Programa Orquídeas del Ministerio de Ciencia, Tecnología e Innovación, la cual emplea técnicas de aprendizaje de máquina y análisis multivariado tales como el análisis de correspondencia múltiple(ACM), para detectar patrones de riesgo en casos de suicidio, intoxicación y violencia de género en San Andres de Tumaco, Nariño, Aunque este estudio se centra en una población específica y no en el fenómeno de violencia intrafamiliar a nivel nacional, es un antecedente relevante en el uso de herramientas estadísticas y computacionales para el estudio y prevención de fenómenos sociales (Bernal et al., 2025).

3.2 Marco teórico conceptual

Para el ICBF, la violencia intrafamiliar se entiende como cualquier acto de someter, controlar, dominar, agredir física, sexual, psíquica, patrimonial o económicamente por un miembro familiar hacia otro. Esto puede darse fuera o dentro del domicilio familiar (ICBF, 2017).

Estos actos constituyen distintas formas de violencia intrafamiliar. La violencia física comprende toda acción que atente contra la integridad física de la persona, como empujones, mechoneadas y golpes, que pueden generar morados, fracturas, sangrado, incapacidad física o incluso la muerte. La violencia patrimonial incluye acciones como ocultar, robar, destruir o retener documentos personales, objetos laborales, bienes o valores. Por su parte, la violencia económica se refiere a formas de control mediante la restricción de recursos, como quitar las ganancias económicas a un miembro de la familia o impedirle contar con un ingreso propio. La violencia sexual se manifiesta cuando se fuerza a una persona a sostener contacto o conductas de tipo sexual, ya sea de manera verbal o física. Finalmente, la violencia psicológica surge a partir del desprecio, el autoritarismo y la imposición, afectando el bienestar emocional y mental de los miembros de la familia (MinJusticia et al., s.f.).

A partir de estas bases conceptuales, se presentan las definiciones operativas de las variables del conjunto de datos. Estas variables permiten caracterizar sociodemográficamente a las víctimas y contextualizar los hechos, facilitando la identificación de las tipologías y determinantes de riesgo propuestos en este estudio.

Ciclo vital

Esta variable corresponde al establecimiento de la etapa de desarrollo vital en la que se encontraba la víctima al momento de realizarse el examen medicolegal. Su análisis es fundamental en esta investigación, ya que permite identificar grupos poblacionales específicos y focalizar los resultados hacia la prevención en los sectores más vulnerables. Las categorías que integran esta dimensión son: (00 a 05) Primera Infancia, (06 a 11) Infancia, (12 a 17) Adolescencia, (18 a 28) Juventud, (29 a 59) Adultez y (Más de 60) Adulto Mayor (INMLCF, 2025).

Circunstancia del hecho detallada

Esta dimensión establece la relación entre la víctima y la situación del entorno en el momento de los hechos, la cual está determinada fundamentalmente por la posibilidad de reconocer la presencia o ausencia de intervención humana. Se define como el acontecimiento o condición específica que se vincula con el acto violento, permitiendo comprender el escenario en el que se produce la agresión. Para efectos de este estudio, las categorías analizadas incluyen: Conflicto relacional, Intolerancia social, Consumo de sustancias, Violencia estructural, Conflicto familiar, Negligencia/abandono, Condición de salud, Criminal/extrema y registros catalogados como Sin información (INMLCF, 2025).

Pertenencia grupal

La pertenencia grupal se comprende como las cualidades de individuos o comunidades que aumentan su vulnerabilidad estructural. En este estudio, dicha variable integra el reconocimiento de grupos étnicos (afrocolombianos, indígenas, Rrom, etc.) junto con otros factores de vulnerabilidad como lo son: pertenencia a grupos indígenas o minoritarios, la migración, la victimización y el desplazamiento forzado, el género, la pobreza y la privación de libertad. La identificación precisa de las personas en situación de vulnerabilidad en cada nación variará según sus particularidades, e incluso su grado de avance social y económico (INMLCF, 2025).

En este sentido, el conjunto de poblaciones clasificadas en condición de vulnerabilidad abarca a mujeres cabeza de hogar, grupos étnicos, campesinos y trabajadores del campo, personas en situación de desplazamiento o de calle, y miembros del sector social LGBTI (OSIGD). Asimismo, se integran categorías como personas con capacidades diferentes, adultos mayores en hogares de cuidado, individuos bajo custodia o protección del ICBF, personas en situación de prostitución, desmovilizados y tribus urbanas. Finalmente, se consideran grupos específicos por su rol o entorno, tales como conductores de servicio público, educadores, funcionarios judiciales, defensores de derechos humanos, periodistas, líderes políticos y personal de salud o religioso en zonas de conflicto.

Escolaridad

Se refiere al nivel educativo máximo alcanzado por la persona según los niveles del sistema educativo formal: preescolar, educación básica en primaria, secundaria, media y superior (INMLCF, 2025).

Sexo de la víctima

Hace referencia al género de la persona que haya sufrido de violencia intrafamiliar. Desde la óptica biológica, se refiere a las cualidades genéticas, hormonales y estructurales del organismo. Las clases empleadas para ordenar estas características en los humanos son masculino, femenino (INMLCF, 2025).

Estado civil

Es la posición de cada individuo frente a las normas o tradiciones vinculadas al matrimonio que se encuentran en el país. En un principio solo había un estado civil, el de soltero, en el período en que el matrimonio no tenía reconocimiento legal. Con su formalización aparece el estado de matrimonio que solo se perdía con el fallecimiento, obteniendo el sobreviviente la condición de viudo. En esa época solo existían tres estados civiles: soltero, casado o viudo. Al legalizarse el divorcio se añade el estado de divorciado, de igual manera ocurre con la separación de cuerpos para el separado, y, por último, con la aceptación de la convivencia, con el conviviente (INMLCF, 2025).

Contexto de violencia

Se define como el conjunto de circunstancias, vínculos y condiciones sociales en las que se produce una situación de agresión, hacia un individuo o grupo particular. El contexto ayuda a

reconocer no solo el acto violento, sino también a quién se dirige y en qué condiciones ocurre, lo que es importante para entender su impacto desigual. Las condiciones sociales son: “Violencia contra el adulto mayor”, “Violencia de pareja”, “Violencia entre familiares” y “Violencia contra niños, niñas y adolescentes” (INMLCF, 2025).

3.3 Marco teórico técnico

En este apartado se presentan y describen las técnicas estadísticas y computacionales que se implementarán en el desarrollo de este trabajo.

Indicadores de incidencia y riesgo poblacional

Para el análisis territorial y comparativo, la investigación transita del conteo de frecuencias absolutas hacia el cálculo de tasas de incidencia. A diferencia de las proporciones internas, las tasas permiten ajustar los datos según el tamaño de la población de referencia, facilitando la comparación del riesgo de victimización entre entidades territoriales con diferentes estructuras demográficas.

Tasa bruta de incidencia (TBI)

Representa la frecuencia relativa global con la que ocurre un evento en la población total de un área geográfica durante un periodo determinado. Es un indicador resumen que mide el riesgo promedio de la población total. La fórmula es la siguiente:

$$TBI = \frac{C}{P} \times K$$

donde C es la cantidad total de casos reportados en el área y periodo de estudio, P es la población total (proyectada) del área y K es la constante de proporcionalidad (normalmente 100,000 habitantes).

Tasa específica de incidencia (TEI)

Permite desagregar el fenómeno según subgrupos poblacionales con características comunes, como el ciclo vital. Este indicador es fundamental para identificar grupos vulnerables y focalizar políticas públicas. La fórmula empleada es:

$$TEI_i = \left(\frac{C_i}{P_i}\right) \times K$$

donde C_i son los casos ocurridos en el grupo o categoría i, P_i corresponde a la población perteneciente al mismo grupo o categoría i y K es la constante de proporcionalidad (100,000 habitantes).

Tasa Estandarizada de Incidencia (TAI)

La tasa estandarizada es un indicador teórico el cual permite comparar poblaciones eliminando los efectos demográficos estructurales. La fórmula es la siguiente:

$$TAI = \sum_{i=1}^{k} (TEI_i \times w_i)$$

donde k es la cantidad de grupos, TEI_i es la tasa específica del grupo i en la localidad evaluada y $w_i$es el peso o proporción de la población del grupo i en la población estándar de referencia. Se calcula como la población del grupo i de la ciudad dividida por la población total de la misma.

Chi cuadrado de independencia de Pearson

La prueba Chi Cuadrado de Independencia es una técnica estadística la cual permite identificar si existe una asociación significativa entre dos variables categóricas. Esta prueba contrasta la independencia entre las variables comparando las frecuencias observadas en una tabla de contingencia con las frecuencias esperadas bajo el supuesto de independencia. (Agresti, 2013).

El estadístico de la prueba es el siguiente:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

donde O_{ij} es la frecuencia observada en la celda *(ij),* E_{ij} es la frecuencia esperada bajo independencia, r es el número de filas y c es el número de columnas.

Para poder emplear esta prueba, deben cumplirse ciertos supuestos: las observaciones deben ser independientes entre sí, las variables analizadas deben ser de frecuencia categórica y las frecuencias esperadas en cada celda de la tabla de contingencia deben ser lo suficientemente grandes, es decir, que al menos el 80 % de las celdas tenga frecuencias esperadas mayores o iguales a 5 y que ninguna celda tenga una frecuencia esperada menor a 1 (Moore et al., 2015; Agresti, 2002).

Para la prueba se plantean las hipótesis H_0: Las variables son independientes (no existe asociación) y H_1: Las variables no son independientes (existe asociación). Se utiliza un nivel de significancia de α. El criterio de decisión se basa en el valor p asociado al estadístico de prueba. En este sentido, se rechaza la hipótesis nula cuando $p < \alpha$, concluyendo la existencia de una asociación estadísticamente significativa entre las variables analizadas. En caso contrario, no se rechaza la hipótesis nula.

Chi cuadrado de tendencia

La prueba de tendencia de Cochran-Armitage es una extensión de la prueba de Pearson, diseñada específicamente para evaluar la existencia de una tendencia lineal en la asociación entre una variable binaria y una variable ordinal. A diferencia de la prueba de independencia general, esta técnica permite identificar si la proporción de una categoría de la variable binaria aumenta o disminuye de manera sistemática a medida que progresan los niveles de la variable ordinal (Cochran et al., 1955).

El estadístico de la prueba es:

$$Z = \frac{\sum_{j=1}^{k}(O_{1j} - E_{1j})}{\sqrt{n_1 n_0 [\sum_{j=1}^{k} w_j^2 p_j \; - \; (\sum_{j=1}^{k} \sum_j w_j p_j)^2]}}$$

En la formulación del estadístico, k es el número de categorías, w_j representa el puntaje ordinal asociado a la categoría j. Por su parte, O_{1j} corresponde al número de observaciones en el grupo 1 para dicha categoría, mientras que E_{1j} denota el número de observaciones esperadas bajo la hipótesis nula de independencia. Adicionalmente, p_j representa la proporción marginal de la categoría j, y n_1 y n_0 corresponden a los tamaños muestrales de los grupos comparados.

Bajo la hipótesis nula de ausencia de tendencia, el cuadrado del estadístico Z^2 se distribuye aproximadamente como una variable $\mathcal{X}^2$ con un grado de libertad ($gl = 1$). Es importante notar que el hallazgo de una tendencia significativa indica un patrón sistemático en los datos, pero no establece una relación de causalidad (Cochran et al., 1955).

Para la prueba se plantean las hipótesis H_0: No existe tendencia monotónica entre las variables y H_1: Existe una tendencia monotónica (creciente o decreciente) entre las variables. Se utiliza un nivel de significancia de α. El criterio de decisión se basa en el valor p asociado al estadístico de prueba. En este sentido, se rechaza la hipótesis nula cuando $p < \alpha$, concluyendo la

existencia de una tendencia monotónica estadísticamente significativa entre las variables analizadas. En caso contrario, no se rechaza la hipótesis nula.

Cramer's V

La técnica de Cramer's V es una medida de fuerza de asociación entre dos variables categóricas y se utiliza como complemento de la prueba Chi Cuadrado de Independencia. Una vez que ésta prueba permite establecer la existencia de una asociación estadísticamente significativa entre las variables, el estadístico Cramer's V cuantifica qué tan fuerte es dicha asociación, independientemente del tamaño de la muestra (Agresti, 2013).

$$V = \sqrt{\frac{\mathcal{X}^2}{N * \min(r-1, c-1)}}$$

donde $\mathcal{X}^2$ es el estadístico chi-cuadrado de independencia, N es el tamaño muestral y $\min(r-1, c-1)$ son las filas y columnas.

El valor de Cramer's V toma valores entre 0 y 1, donde un valor cercano a 0 indica una asociación débil o inexistente, mientras que un valor cercano a 1 indica una asociación fuerte entre las variables categóricas. Este estadístico no proporciona información sobre la dirección de la relación, únicamente sobre su intensidad (Batanero et al., 2011).

Análisis de correspondencias múltiple (ACM)

El análisis de correspondencias múltiples (ACM) es un método de análisis multivariante, extensión del análisis de correspondencias simple, que facilita el estudio conjunto de las relaciones entre múltiples variables categóricas. Su objetivo es reducir la dimensionalidad de los datos y representar, en un espacio de dimensiones reducidas (generalmente un plano factorial), las asociaciones entre categorías y observaciones (Greenacre, M., 2017).

Esta técnica parte de una matriz disyuntiva completa $\mathbf{X}$ de tamaño $n \times m$, donde n corresponde al número de individuos y m al número total de categorías de las variables. Cada elemento x_{ij} de la matriz toma el valor 1 si el individuo i pertenece a la categoría j, y 0 en caso contrario. A partir de esta matriz se construye la matriz de frecuencias relativas $\boldsymbol{P} = \frac{1}{nQ}\boldsymbol{X}$, donde Q es el número de variables categóricas consideradas en el análisis. Posteriormente se definen los vectores de masas de las filas (r) y columnas (c) como:

$$r = \boldsymbol{P} \cdot \mathbf{1}_m \text{ y } c = \boldsymbol{P}^T \cdot \mathbf{1}_n$$

donde $\mathbf{1}_m$ y $\mathbf{1}_n$ son vectores columna de unos. El ACM se basa en el análisis de las disimilitudes entre perfiles utilizando la distancia Chi-cuadrado. Para ello, se construye una matriz centrada y estandarizada respecto a la hipótesis de independencia, dada por:

$$\boldsymbol{S} = \boldsymbol{D}_r^{-1/2}(\boldsymbol{P} - rc^T)\boldsymbol{D}_c^{-1/2}$$

donde $\boldsymbol{D}_r$ y $\boldsymbol{D}_c$ son matrices diagonales que contienen las masas de filas y columnas, respectivamente. Finalmente, se aplica la Descomposición en Valores Singulares (SVD):

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

donde $\boldsymbol{\Sigma}$ es una matriz diagonal cuyos elementos corresponden a los valores singulares, los cuales son no negativos, se encuentran ordenados de manera decreciente y están acotados en el intervalo [0,1]. Por su parte, las matrices $\boldsymbol{U}$ y $\boldsymbol{V}$ contienen los vectores propios asociados a las filas y columnas, respectivamente, los cuales forman sistemas ortonormales. Esta propiedad garantiza la estabilidad de la descomposición y la unicidad de la solución en la representación factorial. A partir de esta descomposición, se calculan las coordenadas factoriales de las filas ($\boldsymbol{F}$) y columnas ($\boldsymbol{G}$) como:

$$\boldsymbol{F} = \boldsymbol{D}_r^{-1/2}\boldsymbol{U}\boldsymbol{\Sigma} \text{ y } \boldsymbol{G} = \boldsymbol{D}_c^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}$$

Estas coordenadas permiten representar tanto a los individuos como a las categorías en un espacio de menor dimensión, facilitando la visualización e interpretación de las relaciones y asociaciones presentes entre las variables categóricas.

Fuzzy C-Means (FCM)

El algoritmo Fuzzy C-Means (FCM) es un método de agrupamiento no supervisado que permite la clasificación de observaciones en múltiples clusters de manera flexible, a diferencia de los métodos de clustering duro como K-means, donde cada observación pertenece exclusivamente a un único grupo. En FCM, cada observación tiene un grado de pertenencia a cada cluster, lo que resulta especialmente útil en fenómenos complejos donde las fronteras entre grupos no son claramente separables (Arango et al., 2016).

Sea un conjunto de datos $\boldsymbol{X} = \{x_1, x_2, \dots, x_n\}$, donde cada $x_i \in \mathbb{R}^p$. El objetivo del algoritmo es particionar los datos en c clusters, minimizando la siguiente función objetivo (Pal y Bezdek, 1995):

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij})^m \parallel x_i - v_j \parallel^2$$

donde u_{ij} representa el grado de pertenencia del individuo i al cluster j, v_j es el centroide del cluster j, $m > 1$ es el parámetro de "fuzzificación", que controla el nivel de difusividad de las asignaciones, y $\parallel x_i - v_j \parallel^2$ es la distancia euclidiana entre la observación y el centroide.

Las restricciones del modelo aseguran que la suma de los grados de pertenencia de cada observación sea igual a 1:

$$\sum_{j=1}^{c} u_{ij} = 1, \forall i$$

El algoritmo se resuelve de manera iterativa actualizando los centroides y los grados de pertenencia hasta alcanzar la convergencia (Suganya y Shanthi, 2012):

$$v_j = \frac{\sum_{i=1}^{n}(u_{ij})^m x_i}{\sum_{i=1}^{n}(u_{ij})^m}$$

$$u_{ij} = \left(\sum_{k=1}^{c}\left(\frac{\parallel x_i - v_j \parallel}{\parallel x_i - v_k \parallel}\right)^{\frac{2}{m-1}}\right)^{-1}$$

Este procedimiento se repite hasta la convergencia, es decir, hasta que los cambios en los centroides o en la matriz de pertenencia sean suficientemente pequeños.

4. Metodología

El presente trabajo se desarrolló bajo un enfoque cuantitativo y un diseño no experimental, ya que no se manipularon las variables. El estudio tuvo un alcance descriptivo, correlacional y analítico-inferencial. En primer lugar, buscó describir la distribución geográfica y demográfica de la violencia intrafamiliar en Colombia. En segundo lugar, se identificaron asociaciones significativas y factores vinculados mediante técnicas estadísticas multivariantes como el análisis de correspondencia múltiple (ACM). Finalmente, la metodología trascendió el análisis tradicional al implementar clustering difuso (Fuzzy C-Means) para la segmentación de perfiles poblacionales.

Para el análisis se utilizó el conjunto de datos oficiales disponibles el día 4 de agosto de 2025 reportados por el Instituto Nacional de Medicina Legal y Ciencias Forenses, correspondientes a los casos de violencia intrafamiliar en Colombia disponibles en la página de datos abiertos

del gobierno nacional, de este conjunto de datos únicamente se trabajó con los registros correspondientes al año 2024. A partir de estos datos, se desarrolló un análisis exploratorio con el propósito de realizar la curaduría y el preprocesamiento de datos, así como la aplicación de análisis descriptivos y geográficos que permitan identificar patrones iniciales y posibles asociaciones entre variables. Posteriormente, se realizaron análisis correlacionales para encontrar asociaciones significativas entre variables e identificar factores asociados a la violencia intrafamiliar. De igual manera, se implementaron técnicas de aprendizaje de máquina que facilitaron la segmentación y caracterización de las víctimas, contribuyendo a una comprensión más profunda del fenómeno.

Población y unidad de análisis

La población objeto de estudio estuvo constituida por el conjunto de personas que reportaron haber sido víctimas de violencia intrafamiliar en Colombia durante el año 2024. Los datos analizados provienen del portal de Datos Abiertos del Estado colombiano, cuya fuente primaria y oficial es el Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF, 2025). Es importante precisar que esta investigación se basa exclusivamente en registros administrativos de carácter forense; en consecuencia, la unidad de análisis corresponde al reporte medicolegal individual y no a la totalidad de los eventos de violencia ocurridos en el territorio nacional.

El conjunto de datos contó con 35 variables y 59.917 registros, donde cada registro corresponde a un reporte individual de violencia intrafamiliar. Cada uno incluye información sobre el género de la víctima, nivel de escolaridad, año y lugar del hecho, así como el tipo y contexto de la violencia, entre otras variables que describen cada caso.

Procesamiento y curaduría de datos

En este apartado se realizó el primer acercamiento a los metadatos, las variables y registros con el fin de identificar valores nulos, categorías como “sin información” o “No aplica”, variables redundantes o que no aportan valor al estudio, con el cual se hicieron las modificaciones que se consideraron pertinentes a cada variable y se tomaron las medidas correspondientes a cada caso.

Análisis exploratorio y descriptivo

Posteriormente, se desarrolló la descripción de las variables con el propósito de comprender las distribuciones, encontrar tendencias y patrones iniciales en los contextos geográficos y

demográficos de los datos. Este análisis se realizó mediante técnicas de estadística descriptiva y análisis geoespacial, lo que permitió identificar concentraciones territoriales y comportamientos diferenciales.

Análisis de asociación

Tras la fase exploratoria, se llevó a cabo un análisis de asociación para identificar vínculos estadísticamente significativos entre las variables categóricas. Se aplicó la Prueba Chi-cuadrado de Independencia y el Chi-cuadrado de tendencia para niveles ordinales. En ambos casos, se utilizó un nivel de significancia de $\alpha = 0{,}05$. Asimismo, se calculó el estadístico V de Cramér para cuantificar la fuerza de dicha asociación independientemente del tamaño de la muestra. Este procedimiento fue un paso previo esencial que permitió filtrar variables redundantes y garantizar la estabilidad de los modelos multivariados posteriores.

Análisis estadístico multivariado

En este apartado se aplicó el Análisis de Correspondencias Múltiples (ACM) con el propósito de identificar asociaciones entre variables categóricas y examinar los factores sociodemográficos, contextuales y territoriales asociados a la violencia intrafamiliar. Esta técnica permitió reducir la dimensionalidad de los datos y representar las relaciones entre categorías y observaciones en un espacio de dimensiones reducidas (plano factorial).

Se construyeron dos configuraciones del ACM. En ambos casos, se definió un conjunto de variables suplementarias (*Pertenencia Grupal*, *Zona del Hecho*, *Escenario del Hecho*, *Actividad Durante el Hecho*, *Mecanismo Causal*, *Diagnóstico Topográfico*, *Días de Incapacidad* y *Tipo de Discapacidad*), las cuales no participan en la construcción de los ejes factoriales, pero pueden proyectarse en el espacio para fines interpretativos. Estas variables fueron clasificadas como suplementarias debido a que presentan bajos niveles de asociación con el resto de variables, según el coeficiente V de Cramér.

La diferencia entre ambas configuraciones radica en el tratamiento de la variable *Contexto del Hecho*. En el ACM I, esta se incluyó como variable activa, mientras que en el ACM II se trató como suplementaria, con el fin de reducir su influencia en la construcción de los ejes y permitir que otras variables emergieran en la estructura del espacio factorial. Adicionalmente, los registros que presentaban más de un 15% de categorías no informativas (como “Sin información”, “Por determinar” u “Otros”) fueron tratados como individuos suplementarios,

con el objetivo de evitar que información incompleta distorsionara la estructura del análisis, manteniendo al mismo tiempo su representación para fines interpretativos.

Segmentación y perfilamiento de individuos

Finalmente, se utilizó la técnica de aprendizaje no supervisado Fuzzy C-Means (FCM) con el propósito de segmentar individuos con características similares. A diferencia de los métodos de clustering tradicionales, el FCM permitió una asignación flexible, en la que cada reporte presenta grados de pertenencia a múltiples clústeres, lo que resulta adecuado para fenómenos complejos donde las fronteras entre grupos no son claramente definidas. Este enfoque facilita el análisis de casos ambiguos o atípicos, permitiendo que una misma observación esté parcialmente asociada a varios grupos. En caso de requerirse una clasificación más definida, cada individuo puede asignarse al clúster con mayor grado de pertenencia (D'Urso et al., 2019).

Para la implementación de la técnica, se emplearon las coordenadas factoriales de los individuos obtenidas a partir del ACM II, dado que estas sintetizan la información más relevante del conjunto de datos y permiten trabajar en un espacio reducido donde se preservan las principales relaciones entre variables. Adicionalmente, se definió el número de clústeres y el parámetro de difusividad m, el cual controla el grado de pertenencia difusa en la asignación. Para este estudio, se utilizó un valor de $m = 2$, comúnmente adoptado en la literatura. Estos elementos influyeron directamente en la formación de los grupos y en la interpretación de los resultados (Bezdek, 1981; D'Urso y Massari, 2019).

Instrumentos y Herramientas

Las etapas de preprocesamiento, curaduría y análisis exploratorio de datos (EDA) se desarrollaron en el lenguaje de programación `Python`, aprovechando la versatilidad de librerías como `pandas` y `numpy` para la manipulación de estructuras de datos, `matplotlib` para la visualización y `geopandas` para el tratamiento de la información geoespacial. Por su parte, las fases de análisis estadístico multivariado (ACM) y la segmentación (Fuzzy C-Means) se ejecutaron en el entorno `R`. Para ello, se emplearon librerías especializadas como `FactoMineR`, `factoextra`, `dplyr`, `ppclust`, entre otras, las cuales garantizaron la precisión en el cálculo de las coordenadas factoriales y la adecuada visualización de los clústeres difusos.

5. Resultados

A continuación, se presentan los resultados, iniciando con la fase de curaduría y preprocesamiento de datos, centrada en la recategorización de variables para garantizar la estabilidad de los modelos multivariados. Posteriormente, se exponen las distribuciones demográficas y el análisis geográfico basado en tasas de incidencia, lo que permite la comparación del riesgo relativo entre territorios. Seguidamente, se evalúan asociaciones mediante las pruebas de chi cuadrado y V de Cramer. Finalmente, se presentan los resultados del Análisis de Correspondencias Múltiples (ACM), utilizado para capturar relaciones estructurales y reducir la dimensionalidad, y de la segmentación mediante Fuzzy C-Means, aplicada sobre las coordenadas factoriales, lo que permite identificar tipologías de violencia intrafamiliar con fronteras flexibles.

Curaduría y preprocesamiento

El proceso de limpieza comenzó con la revisión de los metadatos de cada variable del conjunto de datos, con el fin de identificar los tipos de datos y la incidencia de valores nulos. En este proceso se verificó que el conjunto de datos contó con 59.917 registros y 35 variables, solo una variable temporal completa (año), mientras que las variables mes, día y rango de hora correspondían a componentes temporales desagregados, el resto eran variables categóricas.

Se realizó un análisis de asociación entre variables categóricas. Se aplicó un criterio de exclusión para aquellas con un coeficiente de asociación superior a 0.90, con el fin de mitigar la redundancia de información. Bajo este criterio, y evaluando la relevancia de sus categorías, se descartaron las variables: *Grupo de edad de la víctima, Edad judicial, Identidad de género, Grupo de edad quinquenal, Grupo mayor/menor de edad;* adicionalmente, la variable *Pertenencia étnica* se descartó ya que en realidad está como categoría de la variable *Pertenencia Grupal.*

Asimismo, se descartaron atributos como *ID*, *mes*, *día*, *rango de hora (3 horas)*, códigos DANE y *año del hecho*, debido a que, aunque corresponden a variables de naturaleza temporal o administrativa, no aportan directamente al análisis estructural de las relaciones entre variables categóricas, objetivo central del presente estudio. Se realizó la transformación a escala ordinal en las variables *Ciclo vital* y *Escolaridad*, dado que presentan un orden implícito.

Posteriormente se realizaron recategorizaciones de las variables categóricas: *Presunto Agresor Detallado, Mecanismo Causal de la Lesión no Fatal, Circunstancia del Hecho Detallada, Actividad Durante el Hecho, Escolaridad, Escenario del Hecho y Pertenencia Grupal*. Esto se realizó con el propósito de mejorar la calidad analítica, la estabilidad estadística y la interpretabilidad de los resultados. En particular, se buscó reducir la dispersión categórica y la presencia de niveles con baja frecuencia, los cuales pueden generar inestabilidad en técnicas multivariadas como el Análisis de Correspondencias Múltiples (ACM) y afectar la construcción del espacio factorial.

Estas agrupaciones se realizaron considerando tanto criterios estadísticos basados en la distribución de las variables, como criterios conceptuales, asegurando la coherencia del fenómeno analizado y evitando la fragmentación de la información. Como ejemplo, en la variable *Escolaridad* se unificaron las categorías "Especialización, maestría o equivalente", "Universitario", "Doctorado o equivalente" y "Educación técnica profesional y tecnológica" en una sola categoría denominada "Educación superior", debido a su baja frecuencia. Este proceso permitió una representación más robusta y significativa de los datos, facilitando la identificación de patrones relevantes. Los detalles de las recategorizaciones se presentan en el Anexo A.

Aunque ninguna variable presenta valores nulos, se identificaron categorías como "Sin información", "No aplica" y "Por determinar". Para evaluar la calidad de la información, se construyó un gráfico de barras horizontales en el que las categorías válidas se agruparon bajo la etiqueta "Con información". En el gráfico se incluyeron únicamente las variables que presentaban al menos una de las categorías mencionadas anteriormente diferente a "Con información".

Como se aprecia en la Figura 1, la variable *Transgénero* tiene más del 90% de los registros sin información, por ende, no se tuvo en cuenta en el análisis. Las categorías "Sin información", "Por determinar" y "No aplica" se mantuvieron en todos los análisis.

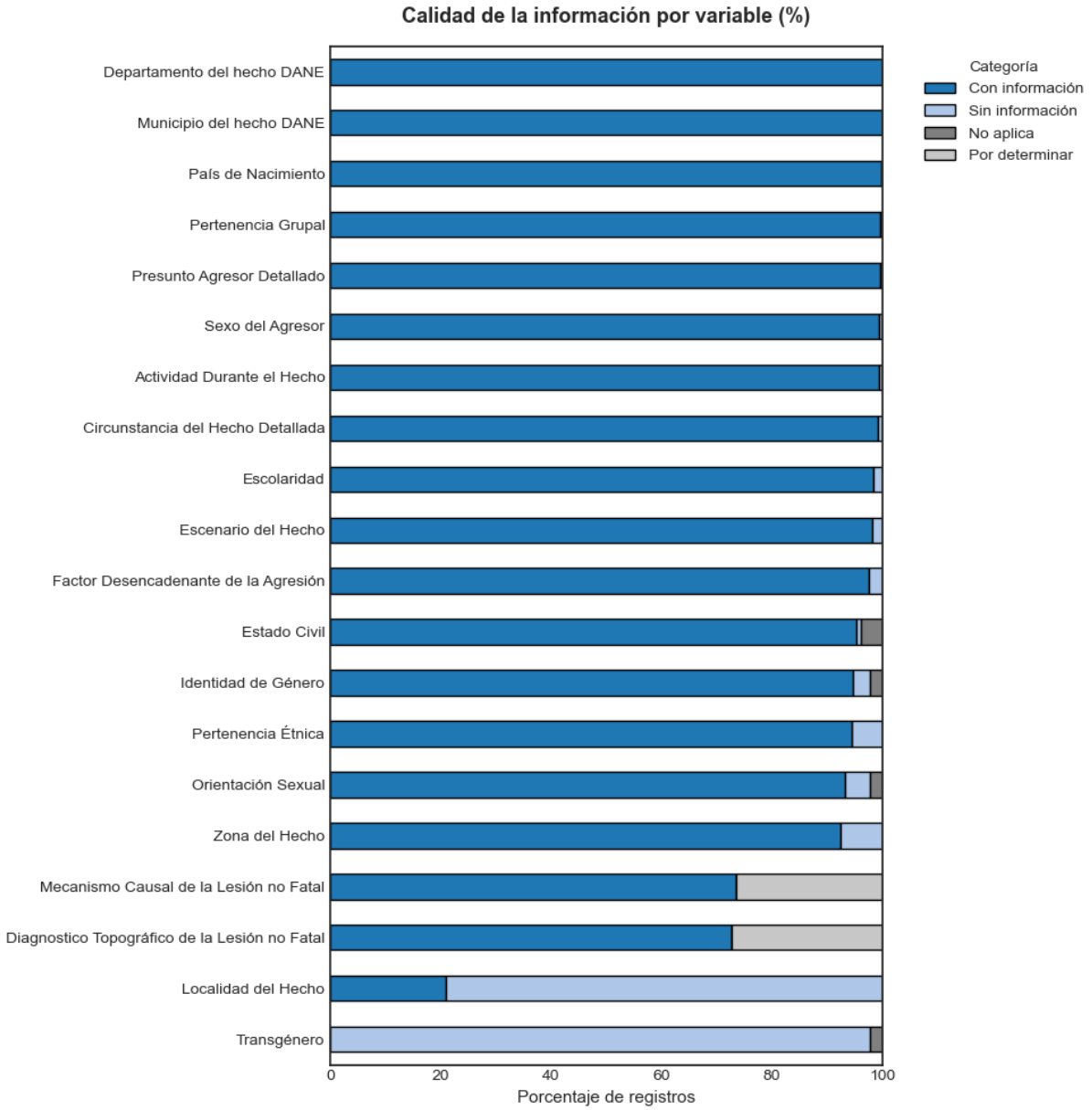


Figura 1*: Calidad de la información por variable*

Para realizar un análisis adecuado de los casos reportados de violencia intrafamiliar a nivel geográfico, se utilizaron las proyecciones oficiales poblacionales publicadas por el Departamento Administrativo Nacional de Estadística (DANE), correspondientes a cada departamento del año 2024, con el fin de construir las tasas y permitir comparaciones consistentes entre departamentos.

Finalmente, para la elaboración de los mapas de Colombia se emplearon archivos en formato GeoJSON, los cuales contienen la geometría espacial de los departamentos, incluyendo sus coordenadas y centroides. Estos insumos permitieron la correcta representación geográfica de la información y su integración con las tasas calculadas previamente. En el caso de los mapas de calor, se utilizó una escala de colores uniforme, con el objetivo de garantizar la

comparabilidad visual entre mapas y evitar interpretaciones erróneas derivadas de cambios en la normalización de la escala. Cuando se consideró necesario se renombró el departamento de "El Archipiélago de San Andrés, Providencia y Santa Catalina" a "San Andrés", con el fin de mejorar la calidad de los gráficos.

Distribución porcentual de variables

Se tienen algunas de las distribuciones más relevantes a continuación.

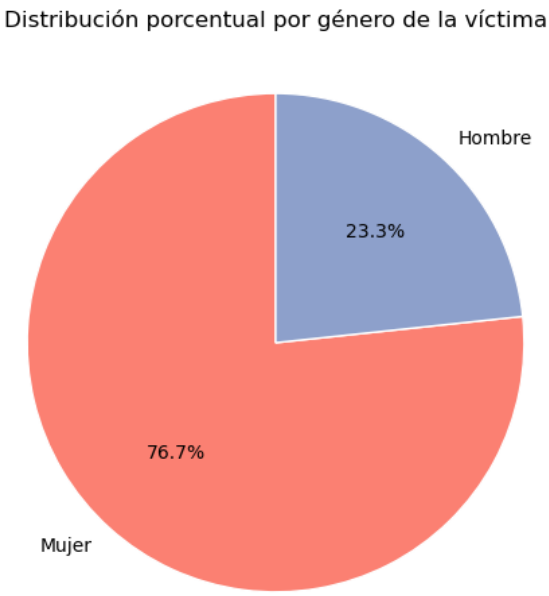


Figura 2: *Distribución porcentual por género de la víctima.*

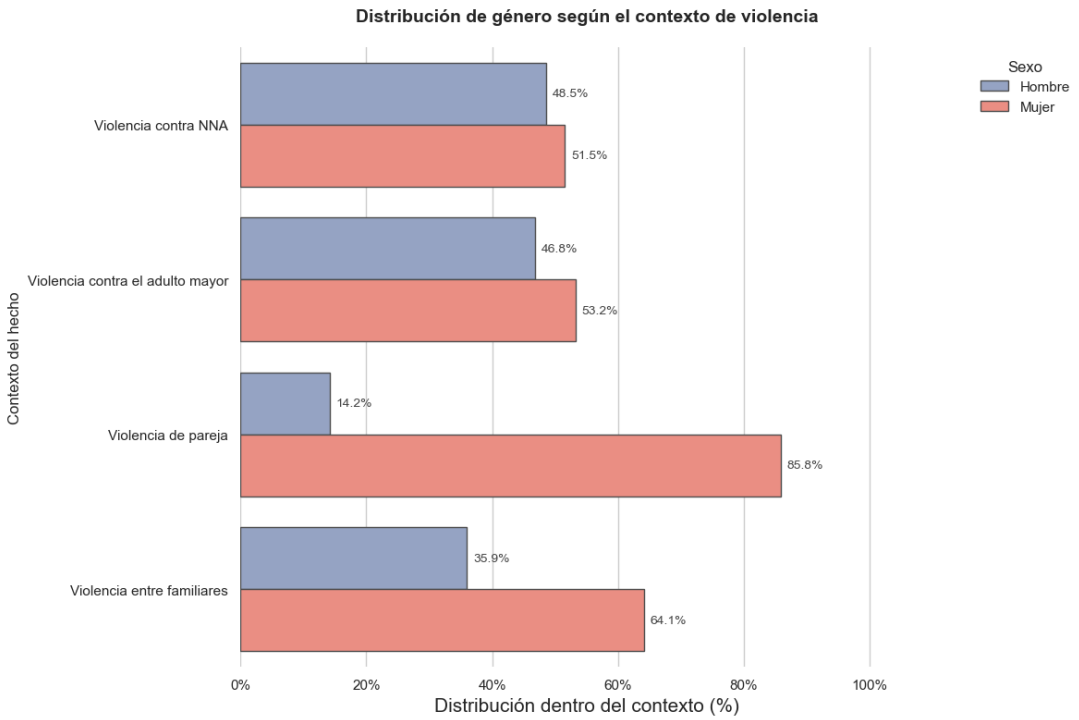


Figura 3: *Distribución porcentual del género según el contexto de violencia.*

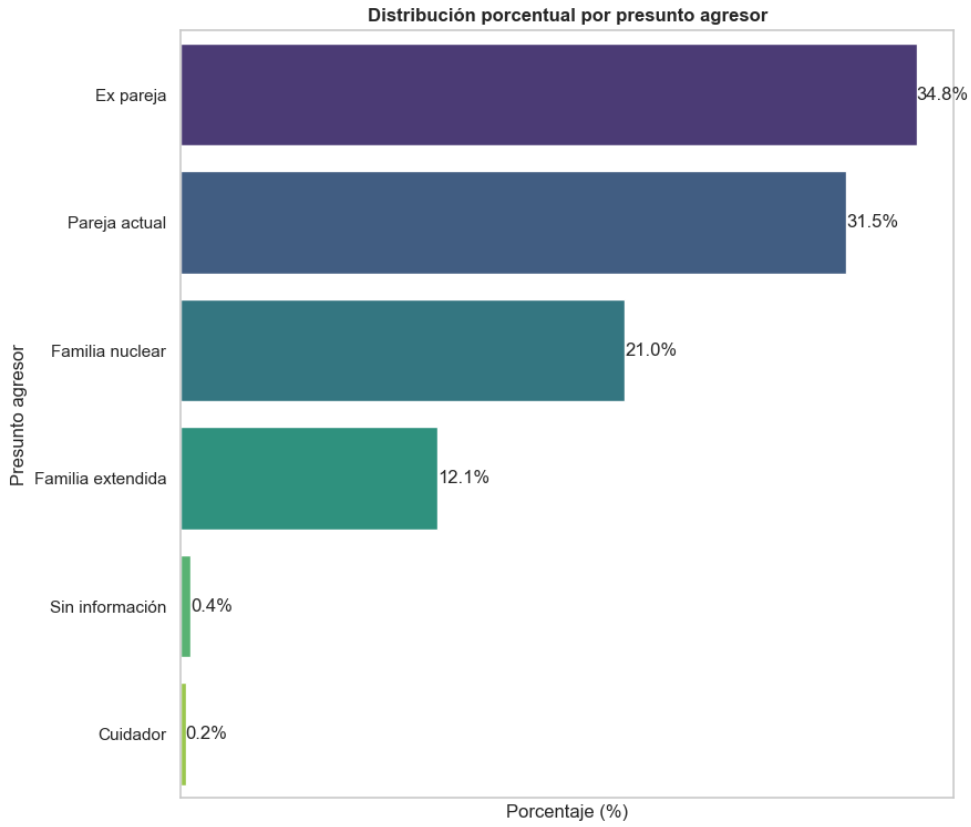


Figura 4: *Distribución porcentual por presunto agresor.*

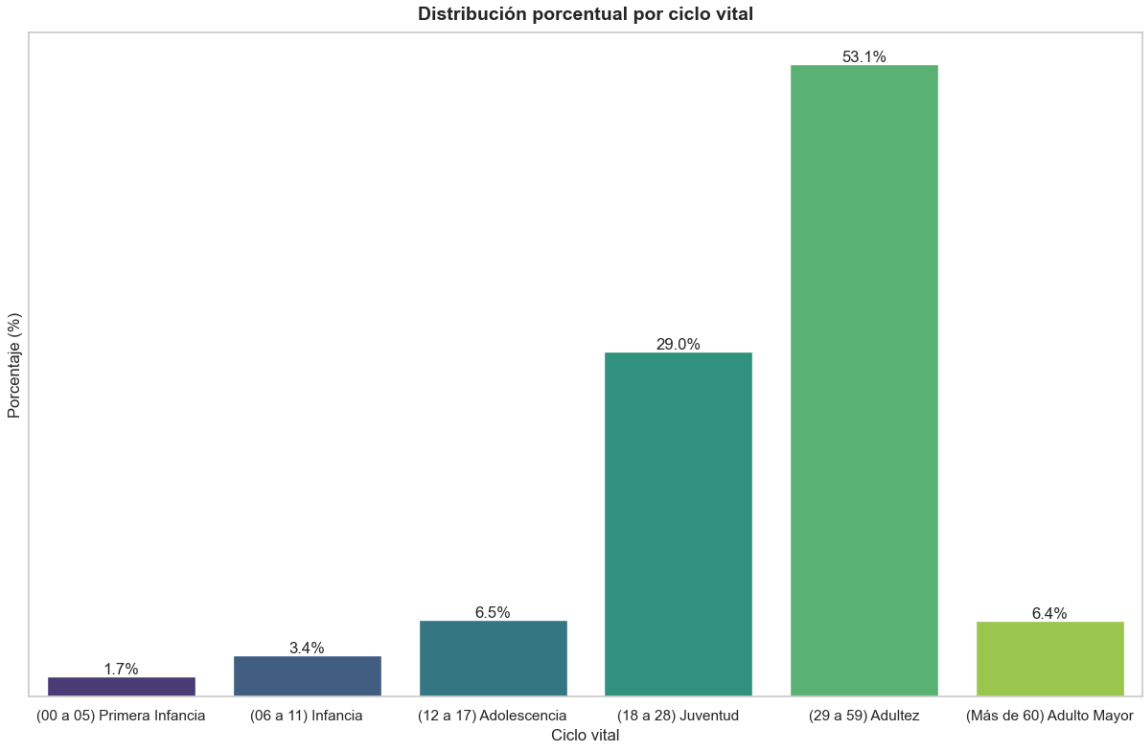


Figura 5: *Distribución porcentual por ciclo vital.*

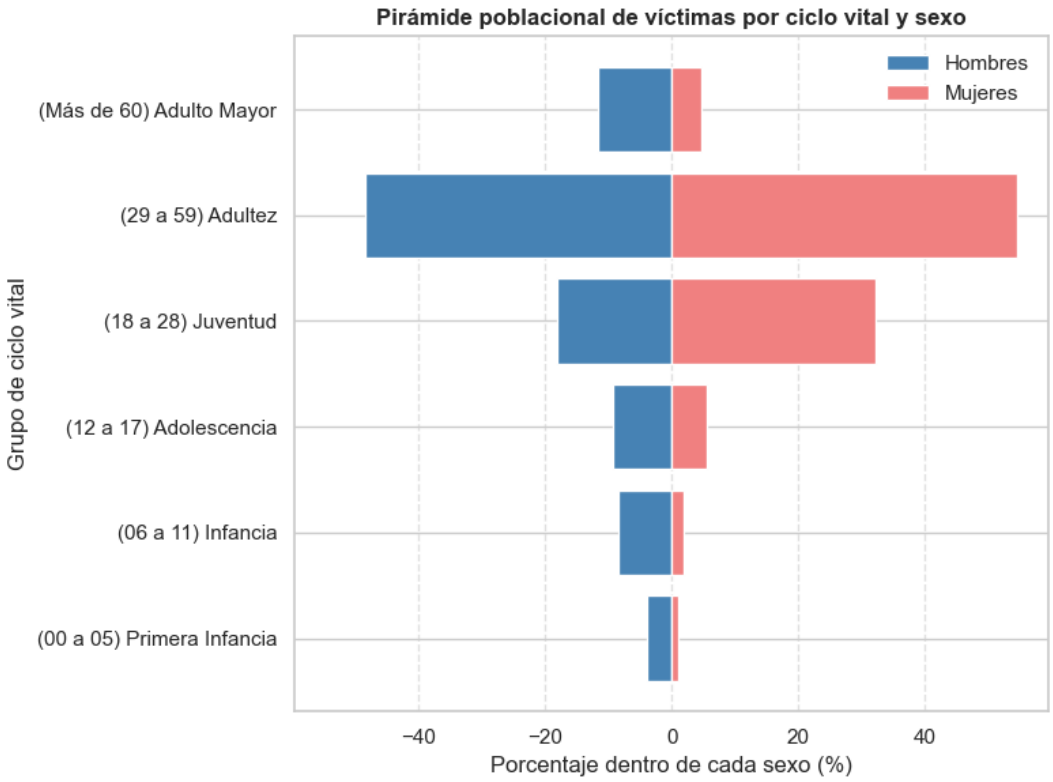


Figura 6*: Pirámide poblacional por grupo de ciclo vital y sexo de la víctima.*

De acuerdo con la Figura 2, la distribución porcentual por sexo de la víctima revela una marcada diferencia de género: el 76,7% de los reportes corresponden a mujeres, confirmando que son el grupo poblacional más afectado por la violencia intrafamiliar en Colombia. Al desagregar esta información por contexto de violencia (Figura 3), se evidencia que la victimización femenina predomina en todas las dimensiones analizadas; destaca especialmente que, en la categoría de violencia entre familiares, las mujeres representan el 64,1% de los registros, mientras que en la violencia de pareja, su participación supera los tres cuartos del total de la categoría.

En cuanto al perfil del presunto agresor detallado (Figura 4), se identifica que el 66% de los casos son perpetrados por la ex pareja o la pareja actual, lo que guarda una estrecha relación con las dinámicas de control y conflicto relacional descritas en el marco conceptual. Respecto al ciclo vital, el análisis muestra que la etapa de la adultez (29 a 59 años) concentra la mayor frecuencia de casos reportados. Por otro lado, los grupos que abarcan la primera infancia, infancia y adolescencia (0 a 17 años) representan de forma conjunta el 11,6% de la población afectada (Figura 5), reflejando un nicho de vulnerabilidad crítica en las etapas tempranas del desarrollo.

Finalmente, el análisis de la distribución porcentual por ciclo vital, representado en la pirámide poblacional (Figura 6), permite identificar diferencias por sexo. Se observa que los hombres presentan una mayor proporción en los ciclos de primera infancia, infancia, adolescencia y adulto mayor, en comparación con las mujeres. Por su parte, las mujeres presentan una proporción ligeramente mayor en los ciclos de juventud y adultez. Estas proporciones corresponden a distribuciones relativas dentro de cada grupo, por lo que la comparación refleja diferencias en la estructura interna por sexo y no en la magnitud absoluta de los casos.

5.1 Distribución espacial y demográfica de las tasas de violencia intrafamiliar en Colombia

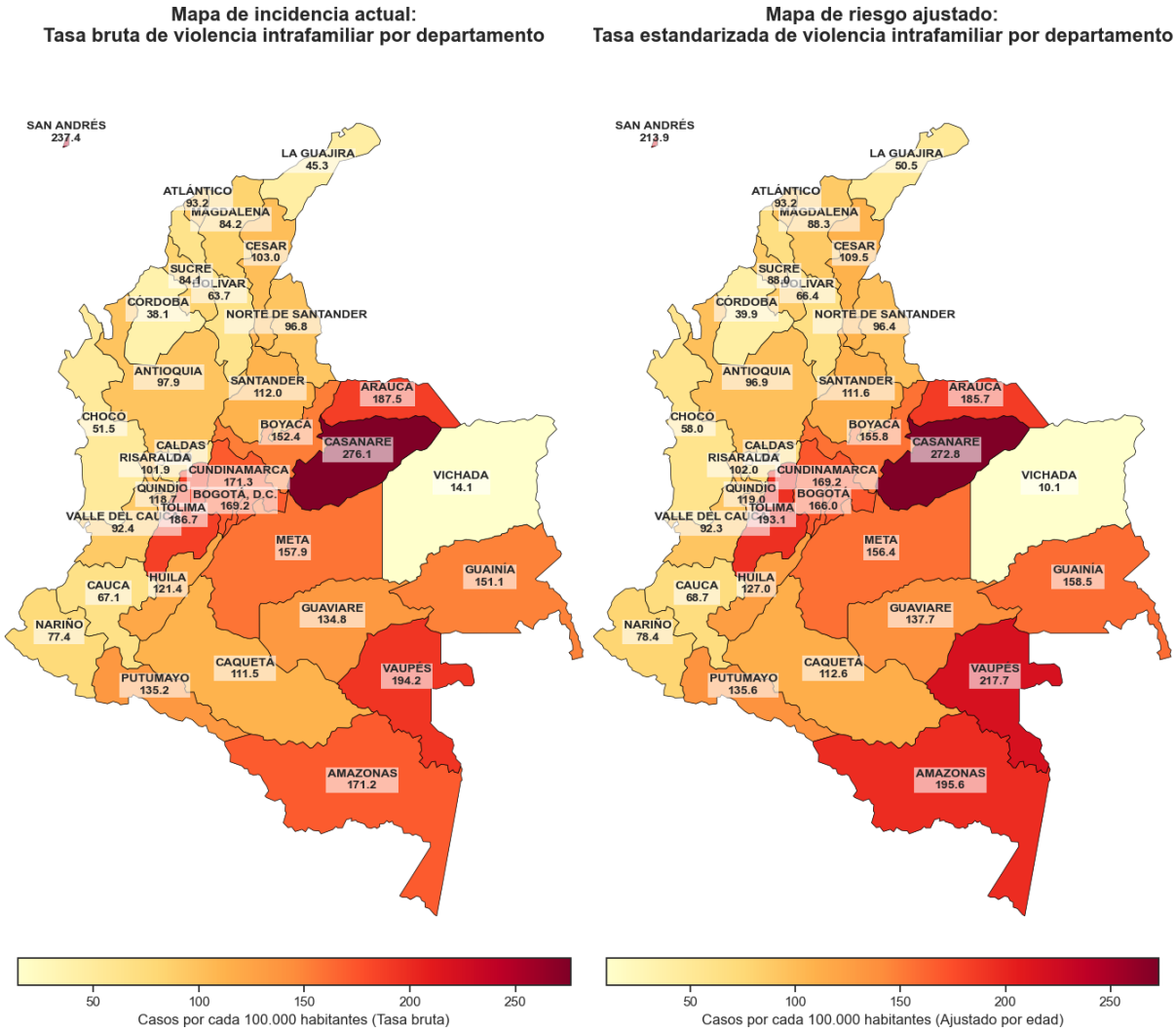


Figura 7: *Mapas de calor de incidencia actual y riesgo ajustado de violencia intrafamiliar por departamento: (a) tasa bruta y (b) tasa estandarizada por ciclo vital.*

Al analizar la Figura 7, se identifica a Casanare como el departamento con mayor incidencia de violencia intrafamiliar en el país, manteniendo valores elevados tanto en la tasa bruta como en la estandarizada. En términos generales, el contraste entre ambos indicadores muestra que la distribución espacial y la intensidad del color del mapa no presentan cambios sustanciales en la mayoría del territorio, lo que sugiere que la estructura demográfica por ciclos vitales no explica en gran medida las diferencias regionales observadas.

No obstante, el proceso de ajuste por edad evidencia excepciones en Amazonas, San Andrés, Vaupés y Tolima. En estos departamentos, la tasa estandarizada supera significativamente a la tasa bruta: mientras que en Amazonas y Vaupés el incremento es superior a 20 unidades, en Tolima el aumento es de aproximadamente 13 unidades. Este ajuste ubica a estos territorios entre los de mayor incidencia en el país, inmediatamente después de Casanare. Estos resultados sugieren que, al controlar por la estructura demográfica, la magnitud de la violencia en estas regiones es mayor de lo que reflejan las tasas brutas, lo que podría estar asociado a factores territoriales específicos.

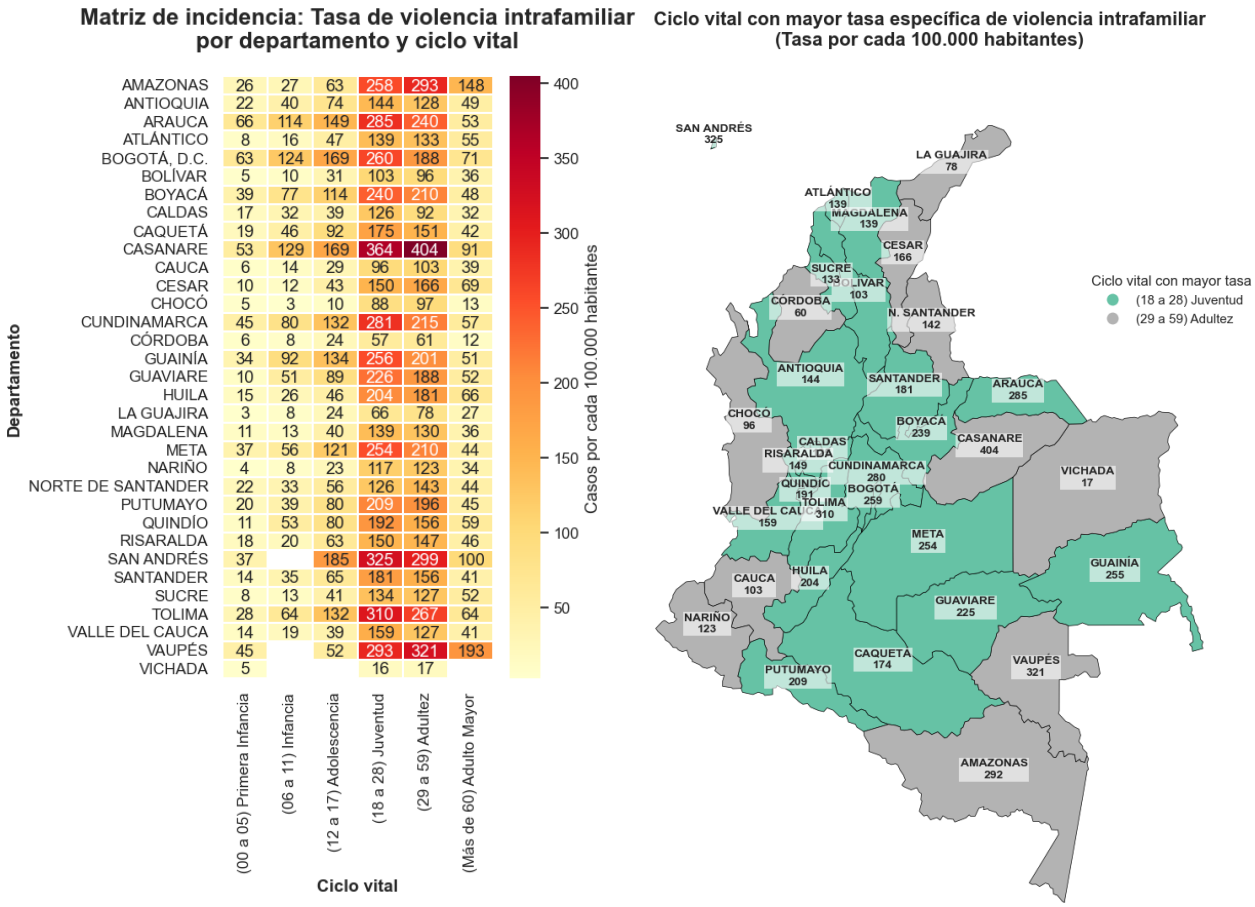


Figura 8: *Análisis territorial de la tasa específica de incidencia de violencia intrafamiliar según ciclo vital: (a) Matriz de incidencia y (b) Mapa de predominancia.*

Respecto a las tasas específicas por ciclo vital, se observa que la mayor concentración de casos ocurre en las etapas de juventud y adultez, tal como lo ilustra la matriz de incidencia de la Figura 8(a). Este análisis reafirma que los departamentos identificados previamente como los más críticos mantienen su tendencia de alta peligrosidad intrafamiliar.

Es relevante contrastar que, en el extremo opuesto, el departamento de Vichada presenta el índice de violencia intrafamiliar más bajo del país, con una tasa específica de apenas 17 unidades y valores similares en el resto de los indicadores.

Un hallazgo relevante de este estudio, detallado en la Figura 8(b), es que, dentro del grupo de departamentos con mayores tasas específicas, en San Andrés, Tolima, Meta, Bogotá, Cundinamarca, Guainía, Boyacá y Arauca, el ciclo vital de la juventud presenta la mayor tasa específica de violencia. En contraste, en el resto de los departamentos con alta incidencia, los niveles más altos se concentran en el ciclo de la adultez, lo que sugiere patrones diferenciales en la intensidad de la violencia intrafamiliar según el ciclo vital entre territorios con alta incidencia.

5.2 Identificación de factores asociados a la violencia intrafamiliar

Prueba Chi Cuadrado de Independencia y Cramer's V

En primer lugar, se aplicó la prueba Chi-cuadrado de independencia junto con la medida de asociación Cramér's V, las cuales permiten identificar qué variables categóricas presentan asociación (Moore et al., 2015; Agresti, 2013) y cuantificar la intensidad de dicha relación. Con el fin de visualizar de manera más clara la fuerza de asociación entre las variables, se construyó un mapa de calor. En este, los valores de Cramér's V cercanos a 1 se representan con tonalidades más intensas, lo que indica asociaciones más fuertes, mientras que los valores cercanos a 0 se muestran con tonalidades más suaves, reflejando asociaciones débiles o inexistentes (Siegel, 1995).

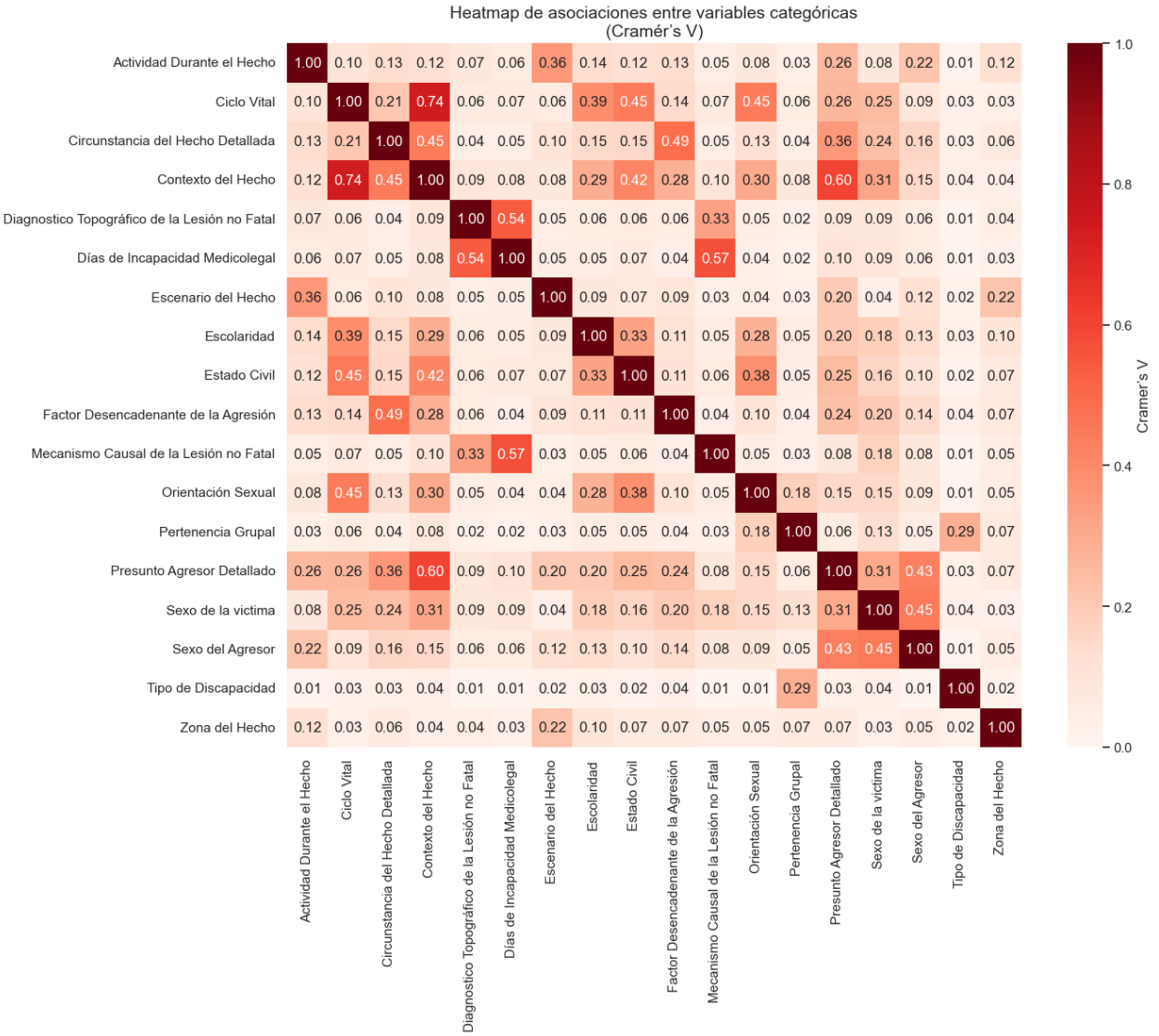


Figura 9*: Mapa de calor de asociaciones entre variables categóricas (Cramer's V).*

Como se puede observar en la Figura 9, existen asociaciones de distinta intensidad entre las variables analizadas. Se identifican asociaciones fuertes, como la existente entre *Presunto agresor* y *Contexto del hecho*. Asimismo, se evidencian asociaciones moderadas, por ejemplo entre *Sexo de la víctima y Circunstancia del hecho*, así como entre *Estado civil* y *Presunto agresor*. Finalmente, también se presentan asociaciones cuya magnitud es débil, lo que indica relaciones poco relevantes desde el punto de vista estadístico.

Si bien se identificaron asociaciones fuertes o moderadas entre algunas variables, es necesario analizar de manera minuciosa las tablas de contingencia correspondientes. Este análisis permite verificar la existencia de celdas con frecuencias nulas o muy bajas, situación que vulnera los supuestos de la prueba Chi Cuadrado y reduce su potencia estadística.

Prueba Chi cuadrado de tendencia

Con el fin de identificar posibles tendencias lineales entre el *sexo de la víctima* y variables de carácter ordinal, se seleccionaron aquellas variables con un orden natural en sus categorías y se sometieron a la prueba Chi Cuadrado de tendencia. Los resultados obtenidos se presentan a continuación:

Tabla 1*: Valores del estadístico de la prueba Chi Cuadrado de tendencia.*

Variable	**χ^2 de tendencia**	**p-valor**
Escolaridad	131.753817	< 0.001
Ciclo vital	15.975207	< 0.001

El análisis estadístico arroja valores $p < 0.05$ para los dos casos, permitiendo rechazar la hipótesis de independencia y confirmar una tendencia lineal. El signo positivo del estadístico revela que dicha tendencia es creciente, demostrando una relación sistemática entre el sexo de la víctima y el comportamiento de las variables ordinales analizadas. Por ende, se entiende que a medida que se avanza de categoría ordinal, los casos reportados van creciendo en ambos géneros en las dos variables. Cabe aclarar que este patrón estadístico, si bien no establece un vínculo causal, identifica una distribución diferencial significativa que resulta relevante para la caracterización del fenómeno objeto de estudio.

Análisis de Correspondencia Múltiple (ACM)

Se analizaron dos configuraciones de Análisis de Correspondencias Múltiples (ACM), diferenciadas por el tratamiento de la variable Contexto del Hecho. En el ACM I, esta variable actúa como eje estructural dominante, explicando el 86% de la inercia en la Dimensión 1 (Anexo C, Tabla C2), lo que permite capturar patrones generales del fenómeno. En contraste, en el ACM II el Contexto del Hecho se incorpora como variable suplementaria, reduciendo su influencia en la construcción de los ejes y permitiendo que otras variables contribuyan de manera más equilibrada. Esto facilita la identificación de patrones menos evidentes dentro del fenómeno.

ACM I

Con el ACM I se valida la alta capacidad explicativa, ya que confirma que el contexto de violencia y el vínculo entre la víctima y el agresor son los elementos primarios que organizan la distribución de los datos, recogiendo además diferencias fundamentales según la etapa de vida. En cuanto a la interpretación de las dimensiones factoriales, la primera dimensión captura una separación entre la *violencia contra niños, niñas y adolescentes (NNA)* , ubicada en el extremo positivo con altas contribuciones de la infancia y la violencia de pareja asociada a conflictos relacionales en el extremo negativo. Por su parte, la Dimensión 2 permite diferenciar los extremos del ciclo de vida, contraponiendo la victimización en la primera infancia frente a la del adulto mayor, bajo la influencia adicional de variables como el estado civil. En la Dimensión 3 se observa una oposición significativa entre la adolescencia y los grupos de primera infancia y adulto mayor, lo que sugiere dinámicas de victimización particulares para estos rangos de edad. Finalmente, la Dimensión 4 comienza a capturar dinámicas más específicas vinculadas a la intolerancia social y al consumo de alcohol o sustancias como factores desencadenantes de las agresiones en contextos familiares distintos a los de NNA. En el Anexo C se encuentran las tablas de varianza explicada, contribuciones, representaciones y las figuras del ACM I.

ACM II

Ahora se realiza el mismo proceso de análisis e interpretación para el ACM II, con la diferencia de que la variable *Contexto del Hecho* se incluye como suplementaria. Esto implica que, aunque puede proyectarse en el espacio factorial para facilitar la interpretación, no participa en la construcción de las dimensiones.

Tabla 2*: Valores propios y varianzas explicadas.*

Dimensión	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	Dim.10
Valores propios	0,37	0,23	0,19	0,19	0,18	0,18	0,17	0,16	0,15	0,14
Varianza explicada (%)	7,10	4,34	3,70	3,56	3,48	3,37	3,30	3,09	2,90	2,66
Varianza acumulada (%)	7,10	11,44	15,14	18,70	22,18	25,55	28,85	31,94	34,83	37,50

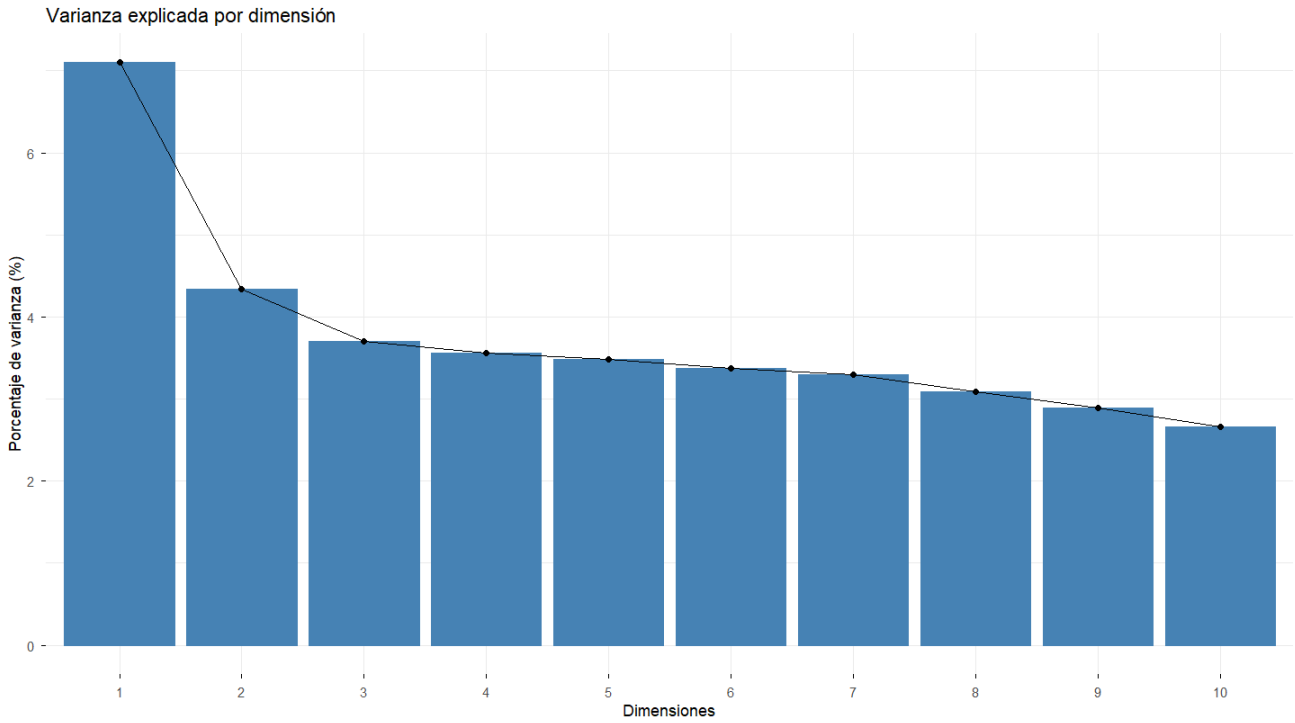


Figura 10*: Porcentaje de varianza acumulada.*

Con base en el análisis de los valores propios y la distribución de la inercia por dimensión, se seleccionaron las primeras cuatro dimensiones para el desarrollo del ACM II. Como se observa en la Figura 10, a partir de la cuarta dimensión el aporte marginal de variabilidad disminuye de forma gradual y homogénea, lo que indica una pérdida en la capacidad discriminante de los ejes subsiguientes. En conjunto, estas dimensiones logran capturar el 18,7% de la variabilidad total (Tabla 2), garantizando la retención de los patrones estructurales más significativos del fenómeno.

A nivel de variables, la varianza explicada por las dimensiones 1 y 2 se consolida como los principales ejes de caracterización sociodemográfica y relacional, mientras que las dimensiones 3 y 4 incorporan un enfoque situacional asociado a las circunstancias y factores detonantes de la agresión (Anexo D, Tabla D1). En particular, la dimensión 1 está fuertemente asociada al *Ciclo Vital* (0,66) y al *Presunto Agresor* (0,61), lo que indica que la estructura primaria de los datos se organiza en torno a la relación víctima–victimario. Por su parte, la dimensión 4 se explica principalmente por la *Circunstancia del Hecho* (0,67) y el *Factor Desencadenante* (0,67), evidenciando un desplazamiento hacia la comprensión de los factores de la agresión.

En cuanto al análisis de categorías, la Dimensión 1 separa con alta precisión la violencia contra menores de las dinámicas de pareja, lo cual se sustenta en las contribuciones de la *Familia Nuclear* (10,00%) y el *Conflicto Relacional* (5,84%) (Anexo D, Tabla D2). La calidad de representación (Anexo D, Tabla D4) de 0,42 para estas categorías confirma que su posición en este eje es altamente confiable y estructural para el fenómeno. La Dimensión 2 introduce una diferenciación crítica en los extremos del ciclo vital, destacando a la Primera Infancia con una contribución de 14,59% (Anexo D, Tabla D2) y una representación de de 0,30 (Anexo D, Tabla D4), consolidándola como el grupo con mayor especificidad en este plano. En los ejes superiores, el modelo logra capturar factores de riesgo específicos: la Dimensión 3 se especializa en la violencia vinculada al Consumo de alcohol y sustancias (Anexo D, Tabla D3 y D5), mientras que la Dimensión 4 queda definida casi exclusivamente por el *Conflicto Familiar*, el cual aporta una contribución del 30,97% (Anexo D, Tabla D3) y una representación en el eje de 0,52 (Anexo D, Tabla D5). Este último hallazgo es vital para el perfilamiento posterior, ya que identifica agresiones motivadas por dinámicas de convivencia complejas que trascienden las etapas del ciclo vital.

En conjunto, al integrar η^2, contribuciones y representaciones, se evidencia que el ACM II está menos concentrado en una sola variable dominante y más distribuido entre varias dimensiones relevantes. El *Ciclo Vital* sigue siendo un eje clave, pero ahora se combina con variables como *Presunto agresor*, *Circunstancia del hecho* y *Factor desencadenante*, lo que permite una lectura más profunda y detallada de las dinámicas de violencia.

Adicionalmente, se observa que, aun cuando el *Contexto del Hecho* se excluye de la construcción de las dimensiones, la estructura general del espacio factorial se mantiene en gran medida. Esto se refleja, por ejemplo, en la Dimensión 2, donde categorías como *primera infancia* continúan teniendo altas contribuciones y buena representación, reproduciendo patrones similares a los observados en el ACM I. En este sentido, el hecho de que el *Contexto del Hecho* sea tratado como suplementario permite confirmar que, aunque sigue estando fuertemente asociado al espacio factorial (η^2= 0.731 en Dim.1) (Anexo D, Tabla D6), no es el único elemento que define la organización de los datos.

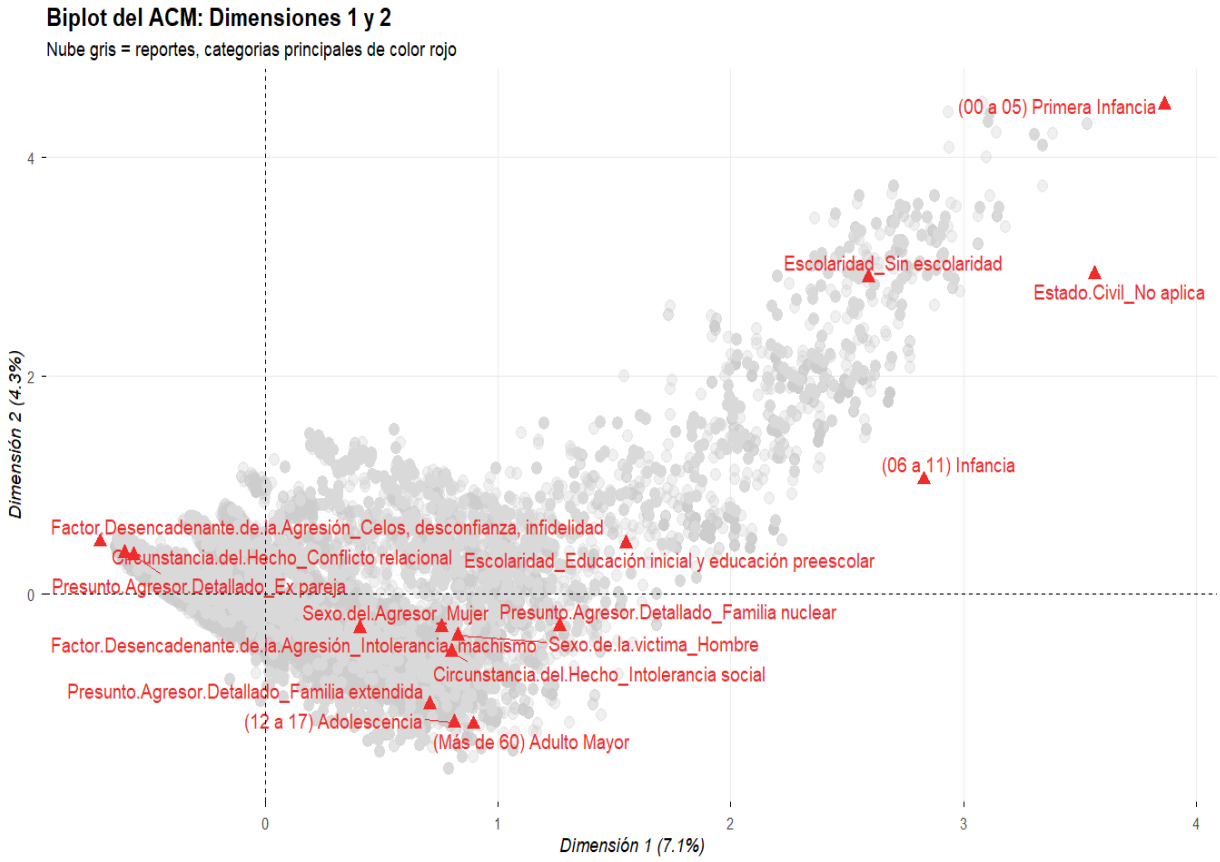


Figura 11: *Representación factorial del ACM II en dimensiones 1 y 2.*

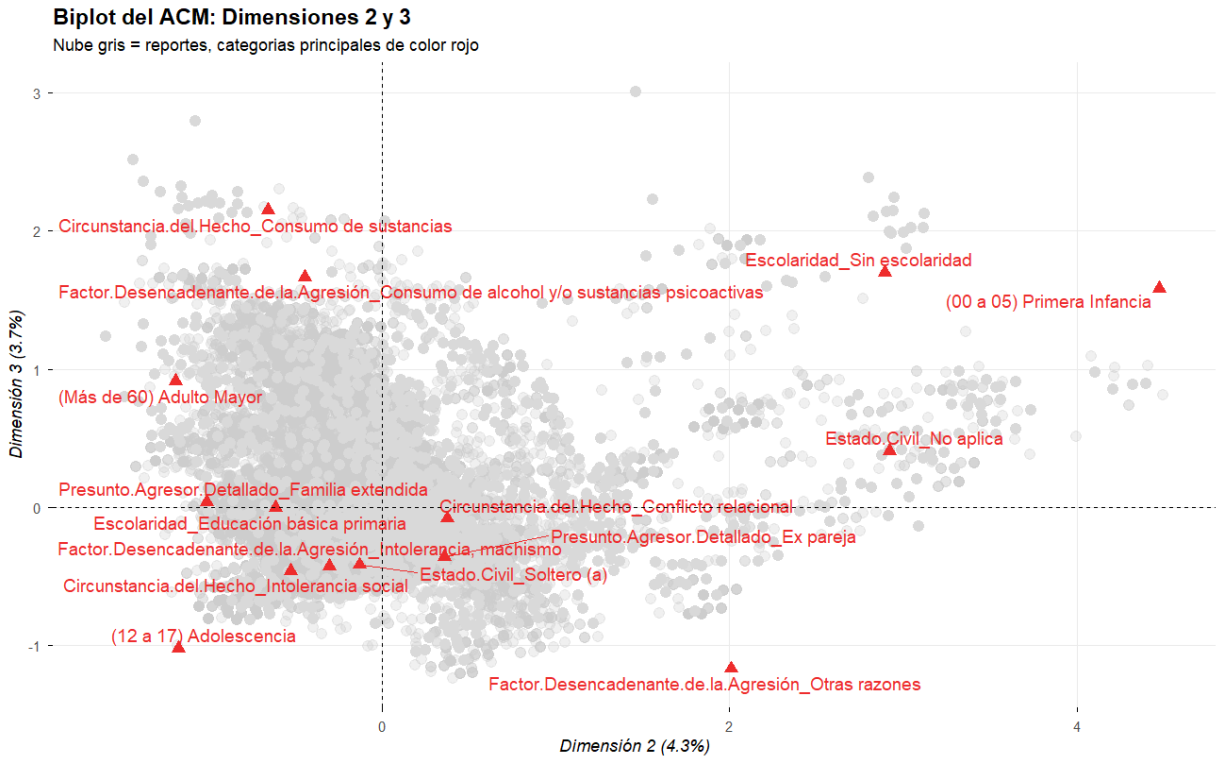


Figura 12: *Representación factorial del ACM II en dimensiones 2 y 3.*

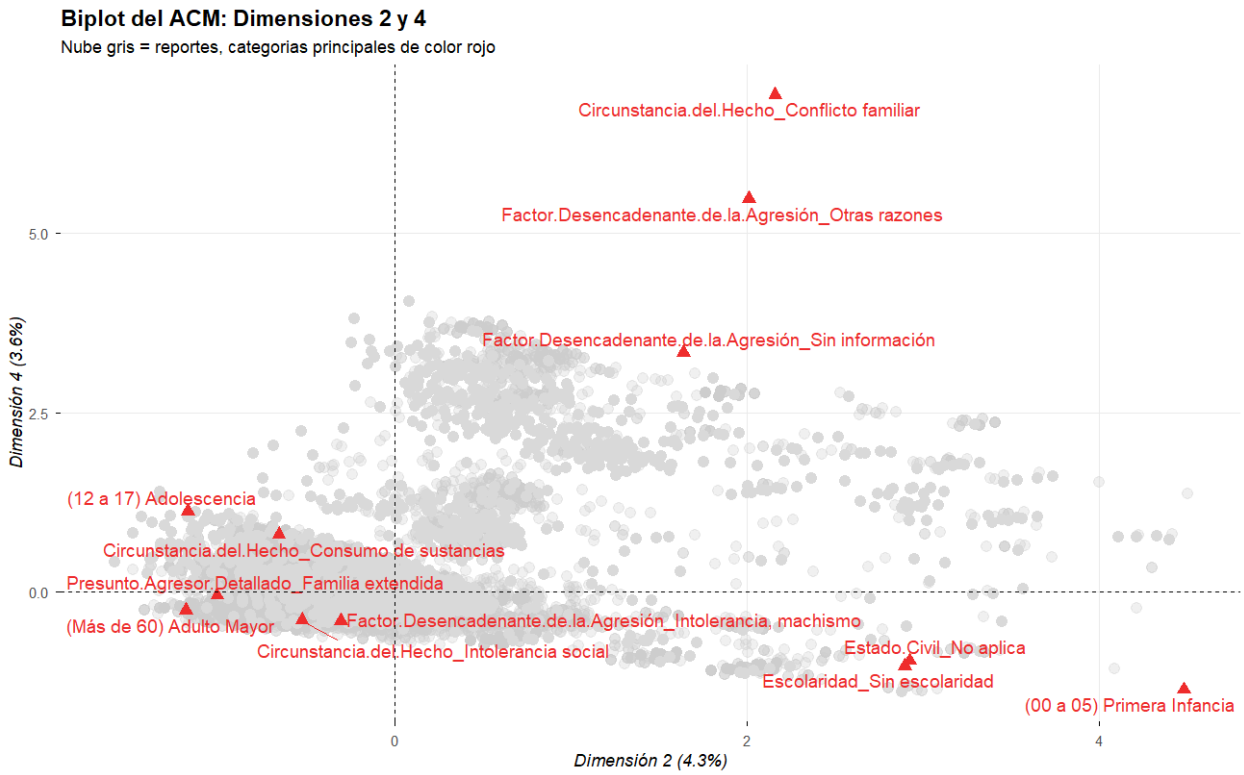


Figura 13: *Representación factorial del ACM II en dimensiones 2 y 4.*

La Figura 11 (Dimensiones 1 y 2) permite visualizar la estructura descrita previamente. En los valores positivos del eje horizontal se agrupan principalmente las categorías asociadas a la violencia contra niños, niñas y adolescentes (NNA), mientras que en los valores negativos se concentran las relacionadas con violencia de pareja. Al incorporar la Dimensión 2 (eje vertical), zonas con valores ligeramente negativos en el eje vertical y cercanos al centro en el eje horizontal aparecen categorías asociadas a adolescencia, junto con características como sexo del agresor mujer y presencia de familia extendida. A medida que aumentan los valores negativos en el eje vertical y se avanza hacia valores positivos en el eje horizontal, se ubican categorías relacionadas con adulto mayor, acompañadas de categorías *como intolerancia social, víctima hombre y agresor de familia nuclear*. Esto sugiere que la Dimensión 2 introduce diferencias dentro de los grupos etarios, asociadas a las condiciones y relaciones en que ocurre la agresión.

La Figura 12 (Dimensiones 2 y 3) permite identificar patrones diferenciados asociados a las condiciones y factores de la agresión. En el primer cuadrante se agrupan categorías vinculadas a niños, niñas y adolescentes (NNA), mientras que cerca del origen se ubican situaciones de transición relacionadas con conflicto relacional con expareja. En el cuadrante con valores

negativos en el eje horizontal y positivos en el eje vertical se agrupan categorías asociadas a circunstancias del hecho como consumo de sustancias y factores desencadenantes como el alcohol, junto con el ciclo vital del adulto mayor. Por otro lado, en los valores negativos del eje vertical se ubican categorías relacionadas con adolescencia, acompañadas de factores como intolerancia y machismo, escolaridad básica primaria y agresores pertenecientes a la familia extendida, evidenciando que esta dimensión permite diferenciar dinámicas de violencia según sus factores desencadenantes y contextuales.

En la Figura 13 (Dimensiones 2 y 4) se observa un patrón adicional. En los valores altos tanto del eje horizontal como del eje vertical se agrupan categorías asociadas a conflicto familiar, junto con factores desencadenantes como "otras razones" y casos sin información. Esta configuración sugiere que la Dimensión 4 captura escenarios menos estructurados, aportando una capa adicional de análisis relacionada con dinámicas de convivencia complejas que no se explican únicamente por el ciclo vital. Finalmente, los gráficos del ACM II muestran una estructura similar a la del ACM I en términos de la separación principal por tipo de violencia, pero incorporan mayor detalle en las dinámicas internas, especialmente en relación con los factores desencadenantes, las circunstancias del hecho y las relaciones entre los involucrados.

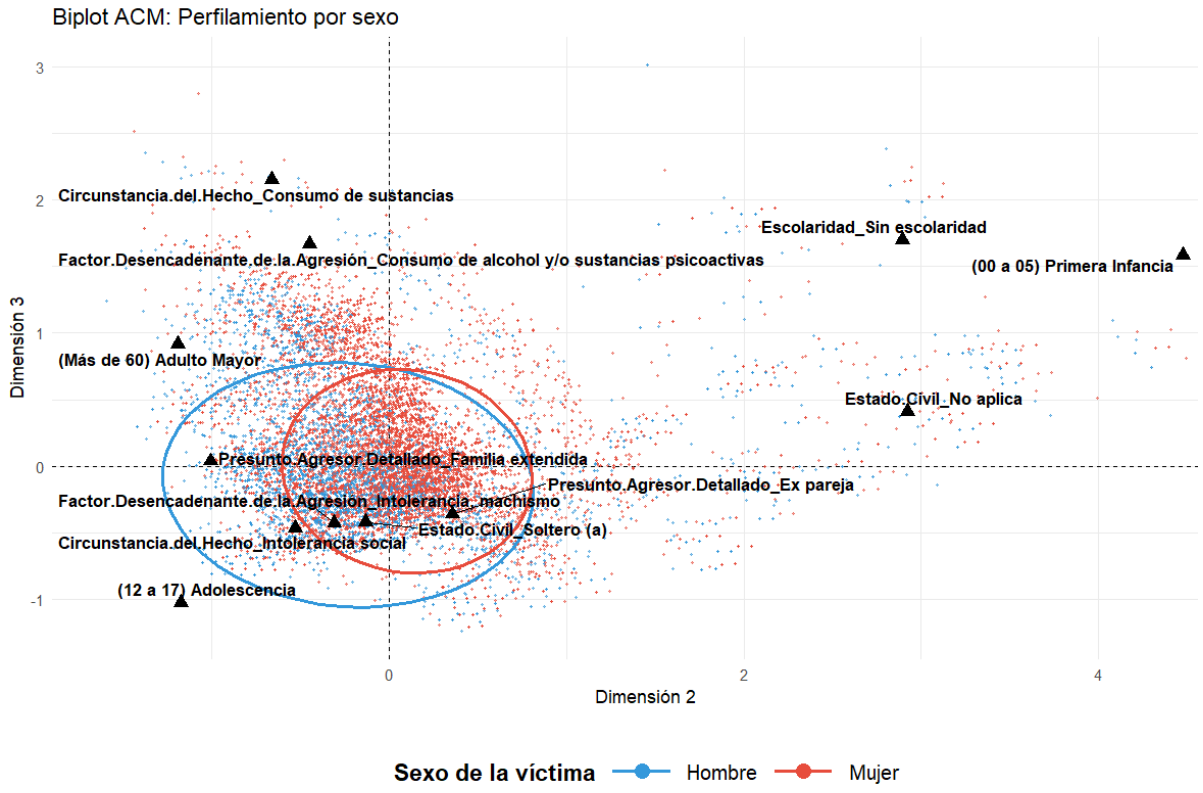


Figura 14: *Mapa factorial del ACM II por género (Dimensiones 2 y 3).*

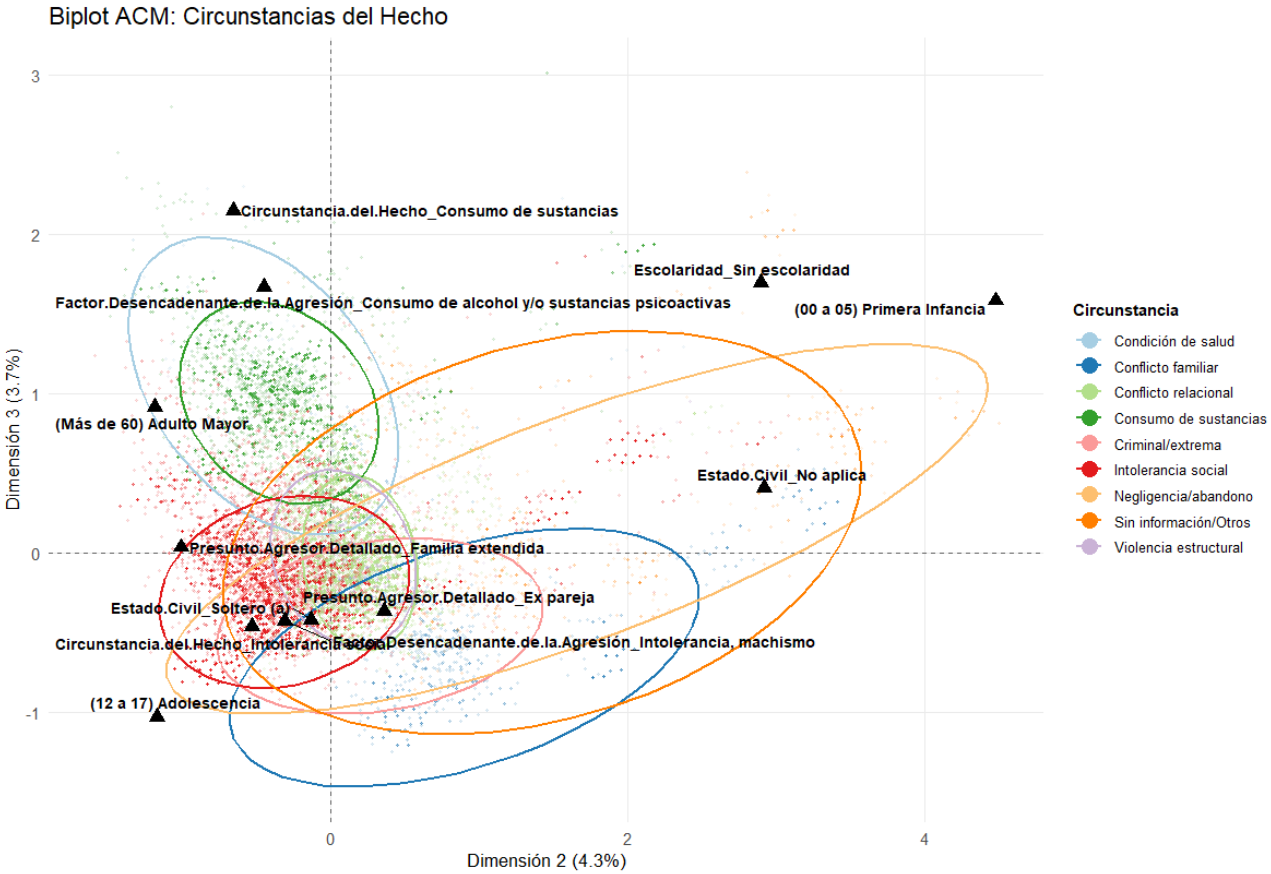


Figura 15: *Mapa factorial del ACM II por circunstancias del hecho (Dimensiones 2 y 3).*

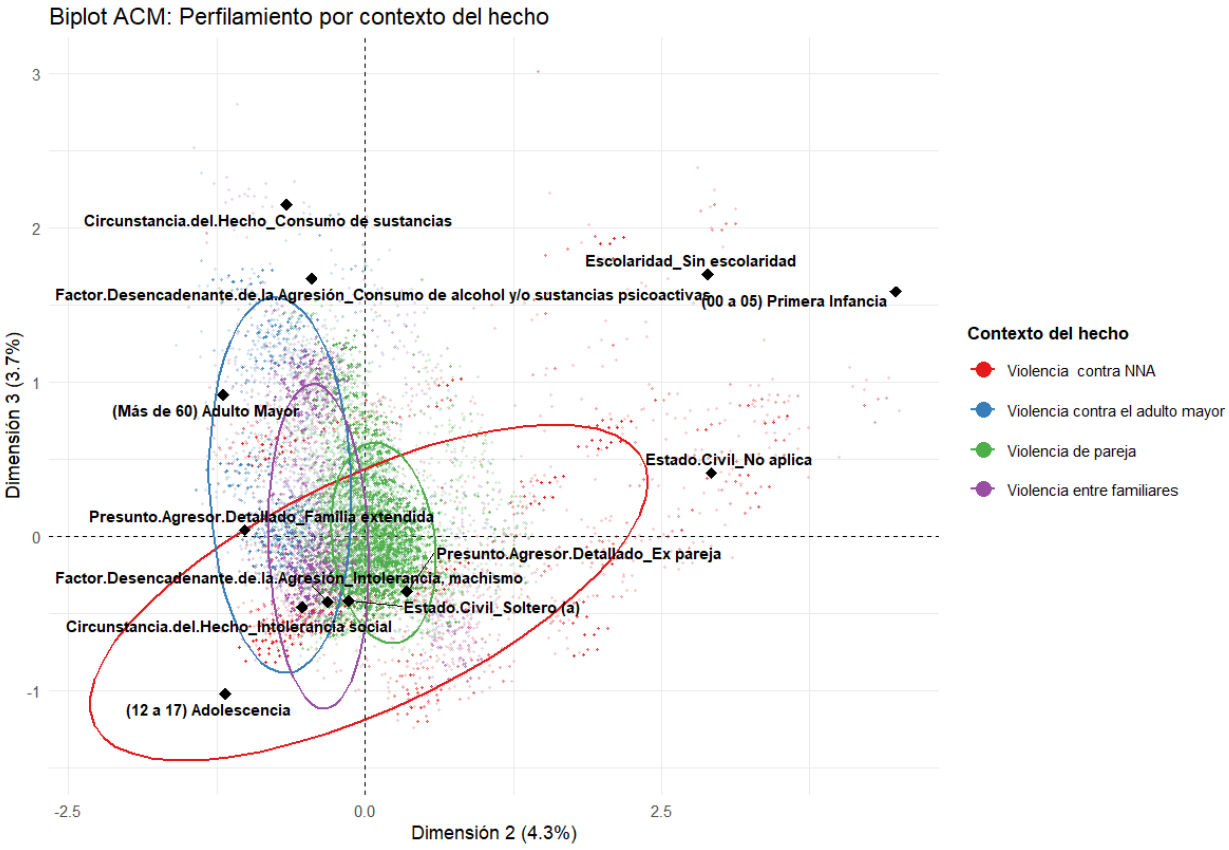


Figura 16: *Mapa factorial del ACM II por contexto del hecho (Dimensiones 2 y 3).*

En los gráficos de perfilamiento se trabaja principalmente con las dimensiones 2 y 3, ya que en este plano emergen con mayor claridad categorías adicionales y se identifican patrones complementarios a los observados en el plano tradicional (Dimensiones 1 y 2).

En el perfilamiento por sexo de la víctima (Figura 14), aunque existe cierta mezcla entre los grupos, se observa una ligera diferenciación espacial. El grupo de hombres (círculo azul) tiende a ubicarse más hacia el lado izquierdo del plano, mientras que el de mujeres se concentra más hacia la zona central. En términos de asociación, las mujeres se relacionan más con categorías como “agresor expareja”, lo que refuerza su vínculo con dinámicas de violencia de pareja. Por su parte, los hombres se asocian más con categorías como “adolescencia”, “intolerancia social, estado civil soltero” y “familia extendida”, lo que sugiere una mayor presencia en contextos de violencia entre familiares.

En el perfilamiento por circunstancia del hecho (Figura 15) se identifican agrupaciones más definidas. Las categorías asociadas al adulto mayor se ubican próximas a condiciones como “consumo de sustancias” y “condición de salud”, lo que indica una mayor asociación con estos factores. En el caso de la adolescencia, se observa cercanía con el “conflicto familiar”. Por otro lado, en zonas centrales del plano se localizan categorías como “intolerancia social” y “machismo”, lo que sugiere que estos factores actúan como elementos transversales presentes en distintos contextos de violencia. Finalmente, el perfilamiento por contexto del hecho (Figura 16) confirma la coherencia de los resultados, ya que las categorías se ubican en concordancia con las que más contribuyen a la construcción de las dimensiones. Esto refuerza la consistencia del modelo y valida las relaciones identificadas previamente en el análisis del ACM II.

5.3 Segmentación y caracterización de perfiles mediante clustering difuso

Fuzzy C-Means

Con base en la estructura factorial obtenida en el ACM II, se aplicó el algoritmo Fuzzy C-Means con el objetivo de segmentar los individuos en perfiles de victimización, permitiendo capturar posibles solapamientos entre grupos.

Para determinar el número óptimo de clústeres (k), se contrastaron los resultados del método del codo, el coeficiente de silueta y métricas propias del clustering difuso, como el coeficiente de partición (Partition Coefficient, PC) y la entropía de partición (Partition Entropy, PE). Al comparar las métricas (Anexo E, Tabla E1), se evidenció una degradación en la calidad del agrupamiento a partir de k = 5, reflejada en una disminución del PC y un aumento significativo del PE, lo que indica una mayor ambigüedad en la asignación de los registros. Para k = 3 y k = 4, los valores de PC y PE resultaron competitivos, sugiriendo soluciones estables desde el punto de vista difuso. Aunque el método del codo y el coeficiente de silueta mostraron indicios de mayor cohesión en k = 5, esta configuración fue descartada debido a la pérdida de calidad en las métricas difusas. En consecuencia, se seleccionó k = 4, ya que representa el mejor equilibrio entre la estabilidad estadística del modelo y el nivel de desagregación necesario para capturar la diversidad de perfiles de victimización.

Tabla 3*: Cantidad y proporción de reportes asignados por cluster.*

Cluster	1	2	3	4	NA	Total
Reportes asignados	32723	8217	2424	16553	0	59917
%	54,61	13,714	4,05	27,63	0	100

La partición en cuatro clusters evidencia una estructura claramente asimétrica, lo que refleja la distribución del fenómeno en el conjunto de datos analizado. El cluster 1 se configura como el perfil dominante al concentrar el 54,61% de los reportes, seguido en relevancia por el cluster 4 (27,63%) y el cluster 2 (13,71%). En contraste, el cluster 3 agrupa únicamente el 4,05% de los casos, por lo que existen nichos de victimización muy desiguales con respecto a la cantidad de reportes asignados y con características muy específicas entre clusters.

En cuanto a la calidad de la asignación, los resultados de la tabla de pertenencia (Anexo E, Tabla E2) muestran un comportamiento consistente que respalda la robustez de la técnica. La mayoría de los individuos presentan una asociación clara a sus respectivos grupos: el 67,20% de los casos se clasifican como “Perfiles Claros” ($u > 0.7$), lo que indica una baja difusividad en la asignación y sugiere la existencia de tipologías bien definidas. Por otro lado, el 28,96% de los registros se ubica en un nivel intermedio de pertenencia, lo cual constituye un hallazgo relevante, ya que refleja la naturaleza compleja y transicional de la violencia intrafamiliar, donde ciertos casos comparten características de múltiples perfiles. Solo el 3,82% presenta una alta ambigüedad ($u < 0.4$), lo que confirma que la técnica de clustering logra una adecuada

diferenciación entre los individuos, manteniendo un bajo nivel de incertidumbre en la asignación.

Finalmente, la certeza promedio por clúster (Anexo E, Tabla E3), tiene valores entre 0,72 y 0,76, lo que ratifica la estabilidad de la segmentación en todos los grupos. En conjunto, estos indicadores reflejan una certeza global cercana al 0,75, lo que demuestra que la técnica posee una adecuada capacidad discriminatoria para separar las dinámicas de violencia.

Tabla 4*: Tabla de proporciones de categorías asignadas por cluster.*

Clúster	Nombre asignado	Ciclo vital dominante	Contexto dominante	Sexo predominante	Factor principal	Agresor principal	Segundo agresor
1	Violencia de pareja (actual y expareja)	Adultez (58.94%) / Juventud (38.80%)	Pareja (99.25%)	Mujer (92.12%)	Conflicto relacional (85.30%)	Expareja (55.88%)	Pareja actual (43.37%)
2	Violencia relacional con consumo de sustancias	Adultez (57.83%) / Juventud (19.35%)	Pareja (45.73%) / Familiar (35.46%)	Mujer (72.21%)	Consumo de sustancias (62.44%)	Pareja actual (36.06%)	Familia nuclear (33.41%)
3	Violencia contra NNA	Infancia (56.56%) / Primera infancia (42.08%)	NNA (99.01%)	Hombre (56.27%)	Intolerancia social (66.96%)	Familia nuclear (83.46%)	Familia extendida (10.64%)
4	Violencia familiar mixta / transicional	Adultez (46.83%) / Juventud (18.50%) / Adolescencia (16.80%)	Familiar (48.31%) / Pareja (21.11%) / NNA (19.01%)	Mujer (53.20%)	Intolerancia social (85.11%)	Familia nuclear (46.43%)	Familia extendida (32.26%)

El análisis de los clusters permite identificar cuatro perfiles diferenciados de violencia, los cuales se distinguen no solo por el contexto del hecho, sino también por el ciclo vital de las víctimas, los factores desencadenantes y el tipo de agresor involucrado. En conjunto, se observan tanto estructuras altamente especializadas como configuraciones más híbridas que reflejan la complejidad del fenómeno.

El cluster 1 (Violencia de pareja (Actual y expareja)) es el más definido, con una concentración casi total en este tipo de violencia (99.25%), afectando principalmente a mujeres adultas

(92.12%). Se asocia de forma predominante a conflictos relacionales (85.30%) y los agresores corresponden en su mayoría a la expareja (55.88%) seguido de la pareja actual (43.37%). Además, este perfil se concentra en etapas de adultez y juventud, lo que refuerza la idea de dinámicas de ruptura o relaciones sentimentales conflictivas estructuradas (Tabla 4).

El cluster 3 (Violencia contra NNA) también presenta una alta especialización, concentrando el 99.01% de los casos en niños, niñas y adolescentes, principalmente en primera infancia e infancia. En este perfil predominan los agresores de la familia nuclear (83.46%) y los factores principales son la intolerancia social (66.96%) y el conflicto familiar (10.85%), lo que refleja dinámicas de violencia intradomiciliaria donde la convivencia y las relaciones de autoridad juegan un papel central (Tabla 4).

El cluster 2 (Violencia relacional con consumo de sustancias) muestra un comportamiento más mixto, con presencia tanto de violencia de pareja (45.73%) como entre familiares (35.46%). Su principal característica diferenciadora es el alto peso del consumo de sustancias (62.44%), lo que sugiere contextos de convivencia donde el consumo de alcohol o drogas actúa como detonante clave de la agresión. Este perfil afecta mayoritariamente a mujeres adultas y se asocia principalmente con la pareja actual y, en menor medida, con miembros de la familia nuclear (Tabla 4). Por último, el cluster 4 (Violencia mixta o transicional) es el más difuso, combinando distintos contextos de violencia (familiar, de pareja y contra NNA) y una distribución más amplia del ciclo vital. Se caracteriza por una fuerte presencia de intolerancia social (85.11%) y por la participación de agresores tanto de la familia nuclear (46.43%) como extendida (32.26%), lo que indica que agrupa casos intermedios o de transición entre diferentes dinámicas de violencia (Tabla 4).

Tabla 5*: Perfilamiento de los clusters.*

Variable	**Clúster 1**	**Clúster 2**	**Clúster 3**	**Clúster 4**
Sexo de la víctima	Mujer	Mujer	Hombre	Mujer
Ciclo vital	Adultez (29–59)	Adultez (29–59)	Infancia (6–11)	Adultez (29–59)
Escolaridad	Media / secundaria	Media / secundaria	Inicial / preescolar	Media / secundaria

Estado civil	Soltero(a)	Unión libre	No aplica	Soltero(a)
Pertenencia grupal	Ninguno	Ninguno	Ninguno	Ninguno
Zona del hecho	Cabecera municipal	Cabecera municipal	Cabecera municipal	Cabecera municipal
Escenario del hecho	Espacio privado	Espacio privado	Espacio privado	Espacio privado
Actividad durante el hecho	Desplazamiento	Cuidado personal	Cuidado personal	Cuidado personal
Circunstancia del hecho	Conflicto relacional	Consumo de sustancias	Intolerancia social	Intolerancia social
Contexto del hecho	Violencia de pareja	Violencia de pareja	Violencia contra NNA	Violencia entre familiares
Sexo del agresor	Hombre	Hombre	Hombre	Hombre
Tipo de agresor	Ex pareja	Pareja actual	Familia nuclear	Familia nuclear
Mecanismo de lesión	Contundente	Contundente	Por determinar	Contundente
Diagnóstico topográfico	Politraumatismo	Politraumatismo	Por determinar	Politraumatismo
Factor desencadenante	Celos / desconfianza / infidelidad	Consumo de sustancias	Intolerancia / machismo	Intolerancia / machismo
Días de incapacidad	1 a 30	1 a 30	1 a 30	1 a 30
Discapacidad	Ninguna	Ninguna	Ninguna	Ninguna
Orientación sexual	Heterosexual	Heterosexual	No aplica	Heterosexual

La caracterización del individuo prototípico por clúster permite evidenciar cuatro dinámicas de violencia intrafamiliar. En primer lugar, el cluster 1 describe un perfil dominante de mujeres

adultas, solteras, residentes en cabecera municipal, que son agredidas por su ex pareja en espacios privados, generalmente durante desplazamientos, bajo contextos de conflicto relacional motivados por celos o infidelidad. En contraste, el cluster 2 también corresponde a mujeres adultas, pero en unión libre, donde la agresión proviene de la pareja actual y está fuertemente asociada al consumo de alcohol o sustancias psicoactivas, reflejando escenarios de violencia en convivencia. Por su parte, el cluster 3 representa un nicho específico de victimización infantil, principalmente niños entre 6 y 11 años, agredidos dentro del núcleo familiar bajo contextos de intolerancia social, con una alta presencia de mecanismos y diagnósticos no determinados, lo que puede indicar dificultades en la identificación o tipificación de la agresión. Finalmente, el cluster 4 agrupa situaciones de violencia entre familiares en adultos, con mayor equilibrio entre sexos, donde predominan factores como la intolerancia y el machismo, evidenciando conflictos de convivencia más amplios dentro del entorno familiar.

En la Figura 17 (Dimensiones 1 y 2), donde los individuos se encuentran etiquetados según su clúster, se evidencia una separación clara entre el clúster 1 (violencia de pareja: actual y expareja) y el clúster 3 (violencia contra NNA), sin indicios de solapamiento significativo. Este comportamiento confirma que ambos perfiles presentan estructuras altamente diferenciadas y consistentes con los hallazgos previos, en los que se identificaron como tipologías bien definidas. Este patrón se mantiene en la Figura 18 (Dimensiones 2 y 3), lo que refuerza la estabilidad de estos clusters y su coherencia con las categorías que más contribuyen a la construcción de las dimensiones.

Por otro lado, los clusters 4 (violencia mixta o transicional) y 2 (violencia relacional con consumo de sustancias) muestran un comportamiento distinto. En la Figura 17 se evidencia un solapamiento importante, reflejado en la mezcla de puntos azules (cluster 4) y rosados (cluster 2), lo que es coherente con su naturaleza más híbrida identificada en el análisis difuso. Sin embargo, en la Figura 18, este solapamiento disminuye, lo que permite una mejor diferenciación entre ambos perfiles. En particular, el cluster 2 se posiciona más cercano a categorías como adulto mayor, consumo de alcohol y sustancias, mientras que el cluster 4 se agrupa alrededor de etiquetas como intolerancia social, machismo y agresor de familia extendida. Este resultado sugiere que, aunque ambos clústeres comparten ciertas características en dimensiones principales, logran diferenciarse al incorporar variables relacionadas con las condiciones y factores desencadenantes de la agresión.

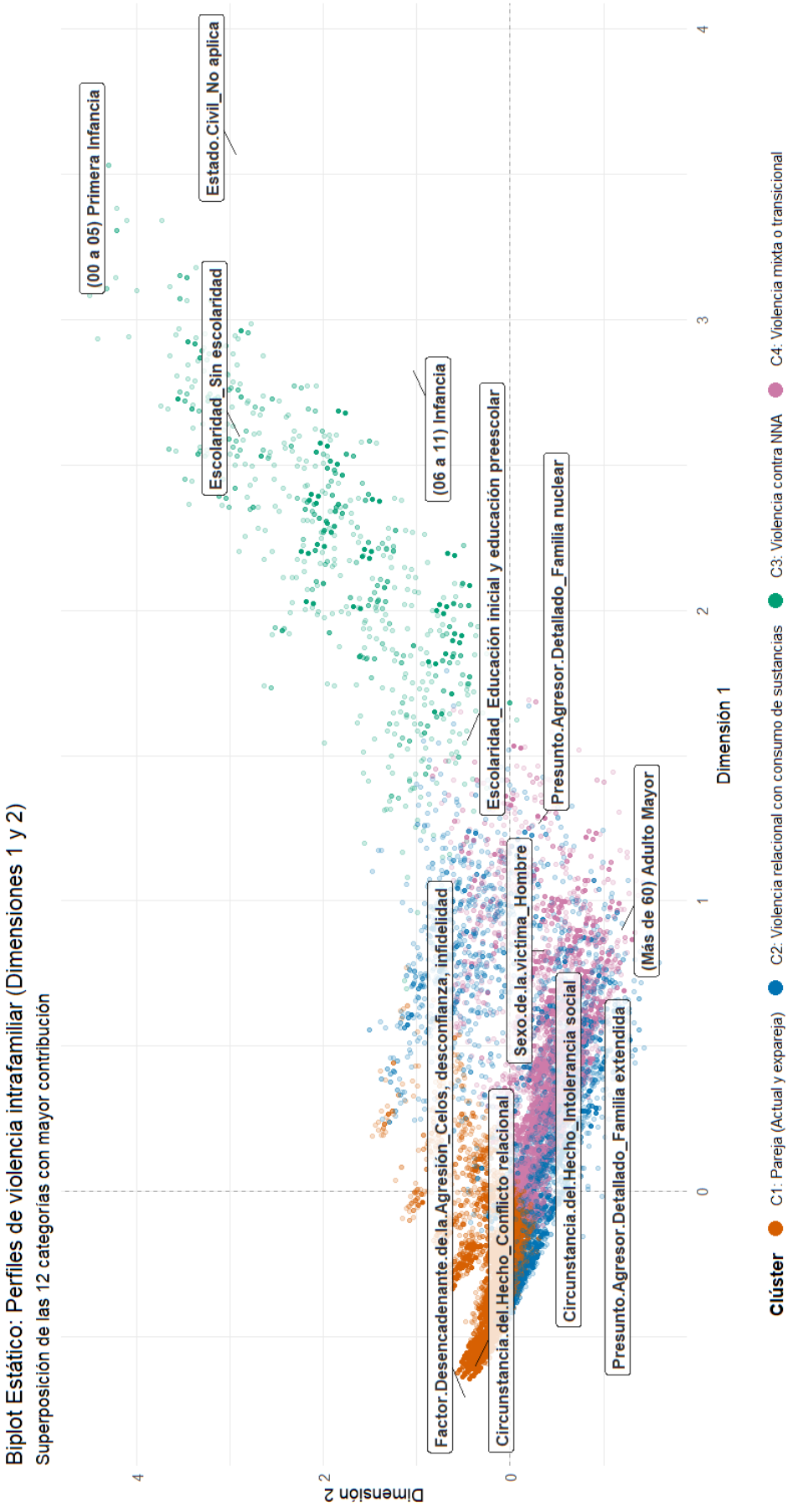


Figura 17*: Mapa factorial dimensiones 1 y 2 del ACM II por clusters.*

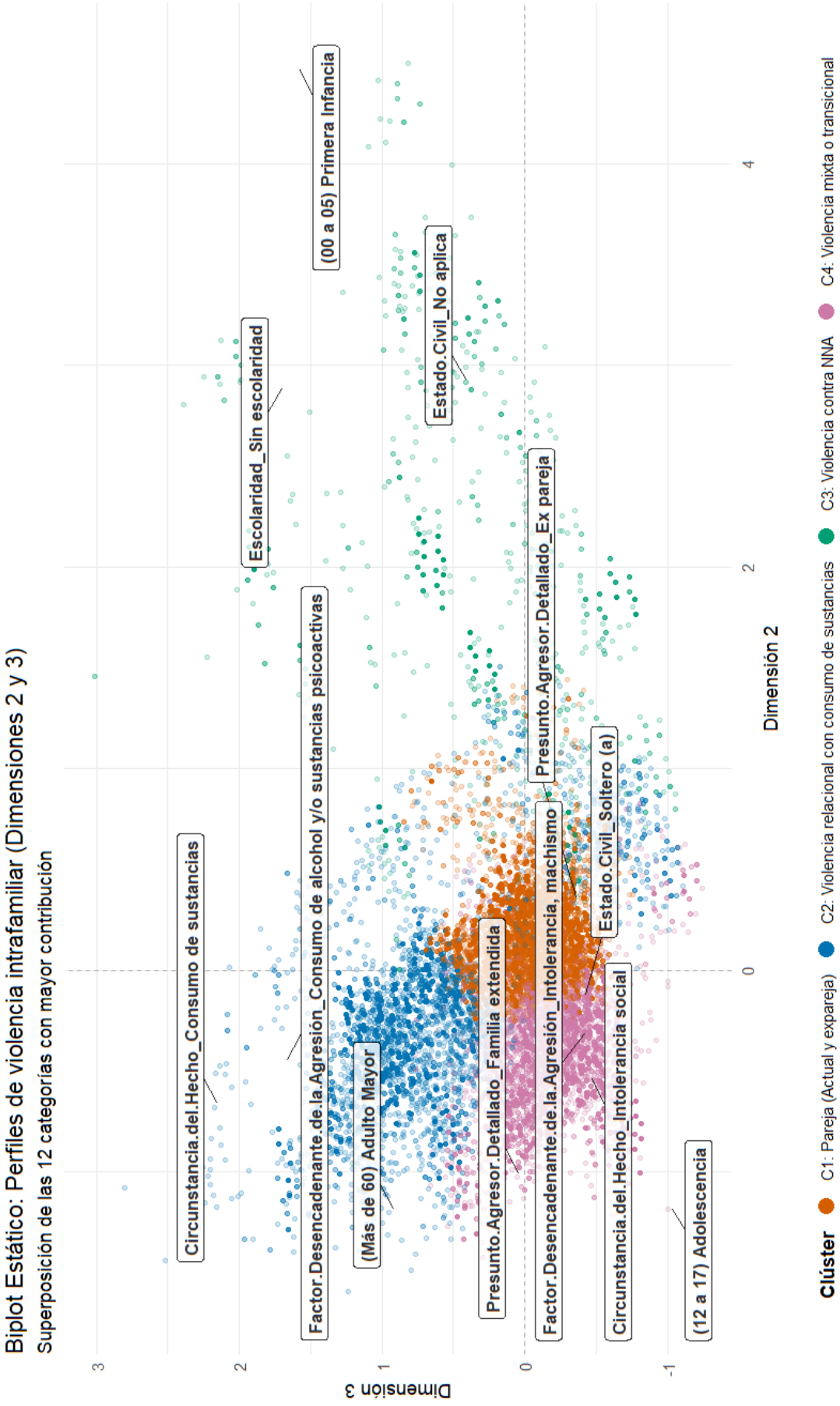


Figura 18: *Mapa factorial dimensiones 2 y 3 del ACM II por clusters.*

6. Conclusiones

El objetivo general del trabajo fue analizar las dimensiones geográficas, sociodemográficas y contextuales de la violencia intrafamiliar en Colombia mediante técnicas avanzadas que permitan generar insumos para la toma de decisiones orientadas a su prevención y atención. En este sentido, los resultados obtenidos permitieron dar cumplimiento a los objetivos específicos planteados.

En relación con el primer objetivo específico, orientado a describir la distribución geográfica y demográfica del fenómeno, el análisis evidenció que las variaciones regionales no se explican únicamente por la estructura demográfica. Departamentos como Casanare, Amazonas, San Andrés, Vaupés y Tolima presentan mayores niveles de criticidad tras ajustar por edad, lo que sugiere la presencia de factores territoriales específicos que influyen en la ocurrencia de la violencia.

Respecto al segundo objetivo, enfocado en identificar factores sociodemográficos, contextuales y territoriales, el análisis de correspondencias múltiples (ACM II) sugirió que la violencia intrafamiliar no depende únicamente del contexto declarado, sino que se organiza en torno a una estructura relativamente estable, principalmente definida por el ciclo vital de la víctima y el vínculo con el agresor. Las dimensiones 1 y 2 separaron de forma consistente la violencia contra menores de las dinámicas de pareja, mientras que las dimensiones 3 y 4 capturan elementos situacionales, destacando factores como el consumo de sustancias y el conflicto familiar.

En cuanto al tercer objetivo, relacionado con la segmentación y caracterización de la población mediante técnicas de aprendizaje de máquina, la aplicación de Fuzzy C-Means permitió profundizar el análisis realizado con el ACM II, al facilitar la identificación de perfiles de victimización más flexibles, capturando transiciones y solapamientos entre grupos. Se identificaron cuatro tipologías diferenciadas. El cluster 1 (54,61%) corresponde al perfil dominante de violencia de pareja postruptura y actual, concentrado en mujeres adultas y asociado principalmente a exparejas en contextos de conflicto relacional. El cluster 2 (13,71%) representa violencia de pareja y entre familiares en convivencia, donde el consumo de sustancias tiene un peso relevante como factor detonante de la agresión por parte de la pareja actual o miembros de la familia nuclear. El cluster 3 (4,05%) se configura como un perfil

especializado de victimización infantil, con alta concentración en primera infancia e infancia dentro del entorno de familia nuclear. Finalmente, el cluster 4 (27,63%) actúa como un perfil mixto o transicional, caracterizado por la presencia de intolerancia social y la interacción entre familia nuclear y extendida en distintos rangos etarios.

La implementación de lógica difusa evidenció que la violencia intrafamiliar no presenta fronteras rígidas. Aunque la mayoría de los casos muestra pertenencia clara a un clúster, una proporción relevante se ubica en niveles intermedios, reflejando la coexistencia de características propias de múltiples tipologías, especialmente entre violencia de pareja y violencia familiar extendida. Esta borrosidad es coherente con la complejidad estructural de las relaciones familiares.

Finalmente, se concluye que, aunque las mujeres representan la mayoría de las víctimas en términos absolutos, su mayor concentración se presenta en contextos de violencia de pareja (cluster 1). En contraste, en la violencia contra menores (cluster 3) y en dinámicas de intolerancia familiar (cluster 4) aumenta la proporción relativa de víctimas hombres, lo que sugiere patrones diferenciados según el tipo de relación y contexto de la agresión. Asimismo, se identifica un perfil específico asociado al consumo de sustancias (clúster 2), en el que este factor actúa como detonante relevante dentro de las dinámicas de violencia.

El uso conjunto del ACM y el clustering difuso permitió articular una lectura complementaria del fenómeno, donde el primero captura la estructura global de las relaciones entre variables, mientras que el segundo facilita la identificación y caracterización de perfiles específicos, incluyendo sus zonas de transición y sus grados de pertenencia.

En conjunto, estos hallazgos permiten una comprensión articulada del fenómeno, integrando sus dimensiones geográficas, sociodemográficas y contextuales, y aportan evidencia empírica que puede orientar estrategias más focalizadas de prevención e intervención.

7. Limitaciones y trabajos futuros

El modelo Fuzzy C-Means presenta una alta sensibilidad a la inicialización y un costo computacional elevado, lo que implica la necesidad de realizar ajustes cuidadosos en parámetros como la semilla, el número de clústeres y los procedimientos de validación. En este

contexto, el uso de muestreo puede ser necesario para reducir la complejidad computacional; sin embargo, si la muestra no es representativa, puede afectar la calidad de la segmentación, generando asignaciones inadecuadas de los individuos a los clústeres. En particular, la presencia de grupos pequeños o subgrupos puede no ser capturada adecuadamente, lo que limita su identificación en los resultados.

Adicionalmente, el análisis se basa en registros oficiales, lo que introduce un posible sesgo de subregistro, dado que los casos no reportados no son considerados dentro del estudio. Esta limitación puede afectar la representatividad del fenómeno analizado.

Por otra parte, debido a limitaciones de tiempo y a la necesidad de realizar una selección más rigurosa de variables, no fue posible desarrollar en su totalidad modelos de carácter explicativo, como la regresión logística multinomial. En particular, la inclusión de covariables exige un análisis cuidadoso para evitar problemas como la multicolinealidad, dado que la incorporación o exclusión de variables puede afectar significativamente la estabilidad e interpretación de los resultados. Asimismo, se identificó que el conjunto de variables disponible podría no ser suficiente para una identificación robusta de factores de riesgo.

Como línea de trabajo futuro, se recomienda profundizar en este enfoque con un énfasis explicativo más que predictivo, orientado a la identificación de factores de riesgo. En este sentido, en futuros trabajos se pueden desarrollar modelos como la regresión logística multinomial, así como explorar extensiones metodológicas, entre ellas la regresión logística difusa y modelos multinivel que incorporen información geográfica. Adicionalmente, se sugiere ampliar el conjunto de variables y realizar análisis temporales más exhaustivos que permitan capturar la dinámica del fenómeno a lo largo del tiempo.

8. Referencias

[1] Agresti, A. (2013). *Categorical Data Analysis* (3.ª ed.). John Wiley & Sons.

[2] Arango, M. A., Jaramillo, J. D. y Jaramillo, L. (2016). Técnicas de clustering para detectar patrones espaciales de criminalidad en jóvenes y adultos en Medellín. *Revista Criminalidad*, *58*(1), 25-45.

[3] Batanero, C., Díaz, C. y Gea, M. M. (2011). *Estadística con Proyectos*. Universidad de Granada.

[4] Bernal-Monroy, E. R., García-Viana, I., Martínez-Casas, P. y López-Sánchez, R. (2025). Detection of victimization patterns and risk of gender violence through machine learning algorithms. *Systems*, 12(1), 21. https://doi.org/10.3390/informatics12010021

[5] Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Springer.

[6] Cervantes, G. (s.f.). *Factores psicosociales asociados a la violencia intrafamiliar en los sectores Calzada y Bosque Centro del barrio El Bosque (Barranquilla)* [Trabajo de grado, Universidad Nacional Abierta y a Distancia UNAD]. Repositorio Institucional. https://repository.unad.edu.co/jspui/bitstream/10596/19834/1/gcervantezm.pdf

[7] Cochran, W. G. y Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, 10(4), 417-451.

[8] Departamento Administrativo de la Función Pública. (s.f.). *Ley 294 de 1996 (Norma 5387)*. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=5387

[9] D'Urso, P. y Massari, R. (2019). Fuzzy clustering of mixed data. *Information Sciences*, *505*, 513-534.

[10] Greenacre, M. (2017). *Correspondence Analysis in Practice* (3.ª ed.). CRC Press.

[11] Instituto Colombiano de Bienestar Familiar (ICBF). (2017). *Conceptos generales sobre la violencia intrafamiliar*. https://www.icbf.gov.co/cargues/avance/compilacion/docs/11001-03-06-000-2017-00122-00%28C%29.htm

[12] Instituto Colombiano de Bienestar Familiar (ICBF). (2023). *Boletín Datos para el cambio: Situación de violencias contra niñas, niños y adolescentes*.

[13] Instituto Nacional de Medicina Legal y Ciencias Forenses. (2006). *Violencia intrafamiliar*. https://www.medicinalegal.gov.co/documents/20143/49496/Violencia+Intrafamiliar.pdf

[14] Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF). (2025). *Violencia intrafamiliar Colombia años 2015 a 2024* [Conjunto de datos]. Portal de Datos Abiertos del Estado Colombiano. https://www.datos.gov.co/Justicia-y-Derecho/Violencia-intrafamiliar-Colombia-a-os-2015-a-2024-/ers2-kerr/about_data

[15] Losada, I. L., Ortiz, H. A., Recalde, J. A. y Rodríguez, A. M. (2020). *Factores de riesgo que limitan la denuncia oportuna de la violencia intrafamiliar en mujeres víctimas* [Trabajo de grado, Politécnico Grancolombiano]. Repositorio Institucional Alejandría. https://alejandria.poligran.edu.co/handle/10823/1971

[16] Ministerio de Justicia y del Derecho. (2024). *Informe sobre el fenómeno de la violencia intrafamiliar en Colombia (2016-2023)*. https://www.minjusticia.gov.co/Sala-de-prensa/Paginas/MinJusticia-presenta-relevante-informe-sobre-fenomeno-de-violencia-intrafamiliar-en-Colombia-(2016-2023).aspx

[17] Ministerio del Interior y Ministerio de Justicia y del Derecho. (s.f.). *Violencia intrafamiliar: definición y modalidades*. https://comunal.mininterior.gov.co/Documentos/Bot%C3%B3n%20Oferta%20Institucional/VIF_Minjustia-Violencia%20Intrafamiliar.pdf

[18] Moore, D. S., McCabe, G. P. y Craig, B. A. (2015). *The Basic Practice of Statistics* (7.ª ed.). W. H. Freeman & Company.

[19] Organización de las Naciones Unidas. (2020). *La violencia contra las mujeres: una pandemia en la sombra*. Noticias ONU. https://news.un.org/es/story/2020/04/1473082

[20] Organización Mundial de la Salud. (2022). *Prevención de la violencia*. Organización Panamericana de la Salud. https://www.paho.org/es/temas/prevencion-violencia

[21] Pal, N. R. y Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, *3*(3), 370-379. https://doi.org/10.1109/91.413225

[22] Siegel, S. (1995). *Estadística no paramétrica aplicada a las ciencias de la conducta* (4.ª ed.). Trillas.

[23] Suganya, R. y Shanthi, R. (2012). Fuzzy C-Means Algorithm- A Review. *International Journal of Scientific and Research Publications*, *2*(11), 440-442.

9. Anexos

Anexo A. Información, recategorización, metadatos y distribuciones porcentuales

Tabla A1*: Metadatos de las variables después de modificaciones.*

N	Variable	Descripción	Tipo de dato	Con información valida (%)
1	Sexo de la víctima	Sexo biológico reportado de la víctima (Hombre o Mujer).	Categórico	100
2	Ciclo Vital	Etapa del ciclo vital de la víctima (niñez, adolescencia, adultez, vejez).	Categórico	100
3	Escolaridad	Nivel educativo alcanzado por la víctima.	Categórico	98.38
4	Estado civil	Estado civil de la víctima al momento del hecho.	Categórico	95.23
5	Tipo de discapacidad	Tipo de discapacidad reportada, si existe.	Categórico	100
6	Pertenencia grupal	Afiliación a grupos sociales o comunitarios.	Categórico	99.69
7	Departamento del hecho DANE	Nombre del departamento correspondiente al hecho.	Categórico	99.98

8	Localidad del hecho	Localidad o zona específica dentro del municipio.	Categórico	20.96
9	Zona del hecho	Clasificación urbana o rural del lugar.	Categórico	92.36
10	Escenario del hecho	Lugar físico donde ocurrió el hecho (vivienda, vía pública, etc.).	Categórico	98.25
11	Actividad durante el hecho	Actividad que realizaba la víctima al momento del hecho.	Categórico	99.43
12	Circunstancia del hecho	Descripción de las condiciones o contexto del hecho.	Categórico	99.15
13	Contexto de violencia	Tipo o modalidad de violencia (de pareja, contra Niños y niñas, contra el adulto mayor, entre otros familiares)	Categórico	100
14	Mecanismo causal	Medio o mecanismo utilizado para causar la lesión.	Categórico	73.54
15	Diagnóstico topográfico de la lesión	Localización anatómica de la lesión reportada.	Categórico	72.79
16	Sexo del presunto agresor	Sexo reportado del agresor.	Categórico	99.51
17	Presunto agresor	Relación o rol del agresor con la víctima (pareja, familiar, desconocido, etc.).	Categórico	99.56
18	Días de incapacidad medicolegal	Número de días de incapacidad dictaminados por Medicina Legal.	Categórico	100
19	Factor Desencadenante de la Agresión	Causa que desencadeno el hecho(intolerancia, machismo,etc)	Categórico	97.52
20	Orientación sexual	Orientación sexual de la victima	Categórico	93.73

Tabla A2: *Recategorización de las categorías de la variable Pertenencia Grupal.*

Categoría original	**Categoría recategorizada**
Ninguno	Ninguno
Persona en condición de desplazamiento	Vulnerabilidad social
Persona privada de la libertad	Vulnerabilidad social
Persona en situación de calle	Vulnerabilidad social
Persona en situación de prostitución	Vulnerabilidad social
Persona adicta a una droga natural o sintética	Vulnerabilidad social

Desmovilizados o reinsertados	Vulnerabilidad social
Niños, niñas, adolescentes bajo la protección del ICBF	Población protegida
Niños, niñas, adolescentes en condición de abandono	Población protegida
Adolescentes en conflicto con la ley	Población protegida
Persona mayores en hogares de cuidado	Población protegida
Persona con capacidades diferentes	Población protegida
Grupos étnicos	Grupos poblacionales
Sector social LGBTI (OSIGD)	Grupos poblacionales
Campesinos (as) y/o trabajadores (as) del campo	Grupos poblacionales
Servidor público	Ocupación/rol
Funcionarios judiciales	Ocupación/rol
Maestro/Educador	Ocupación/rol
Conductores de vehículos de servicio público	Ocupación/rol
Persona que ejerce actividades de periodismo	Ocupación/rol
Persona que ejerce actividades gremiales o sindicales	Ocupación/rol
Religioso	Ocupación/rol
Líder	Ocupación/rol
Mujer cabeza de hogar o de familia	Condición familiar
Pertenencia múltiple	Sin información/Otros
Tribus urbanas	Sin información/Otros
Otros	Sin información/Otros
Sin información	Sin información/Otros

Tabla A3: *Recategorización de las categorías de la variable Escenario del Hecho.*

Categoría original	**Categoría recategorizada**
Vivienda	Espacio privado
Lugares de hospedaje (hoteles, campamentos y otros tipos de hospedaje no permanente, moteles, etc.)	Espacio privado
Piscina y Jacuzzi (vivienda)	Espacio privado

Calle (autopista, avenida, dentro de la ciudad)	Espacio público urbano
Vía pública	Espacio público urbano
Parqueaderos, estacionamientos	Espacio público urbano
Terminales de pasajeros	Espacio público urbano
Transporte masivo	Espacio público urbano
Carretera (fuera de la ciudad)	Espacio público rural
Zonas de actividades agropecuarias	Espacio público rural
Espacios terrestres al aire libre (bosque, potrero, montaña, playa, etc.)	Espacio público rural
Espacios acuáticos al aire libre (mar, río, arroyo, humedal, lago, etc.)	Espacio público rural
Terreno baldío	Espacio público rural
Lugar de explotación de minas y canteras	Espacio público rural
Establecimiento comercial (tienda, centro comercial, almacén)	Comercio y servicios
Establecimiento comercial (plaza de mercado, galería)	Comercio y servicios
Establecimientos de expendio de comidas (restaurantes, asaderos, salsamentarias, etc.)	Comercio y servicios
Estaciones de servicio (bombas de gasolina)	Comercio y servicios
Establecimientos financieros y relacionados (bancos, fiduciarias, etc.)	Comercio y servicios
Lugares de esparcimiento con expendio de alcohol	Ocio y recreación
Áreas deportivas y/o recreativas	Ocio y recreación
Piscina y jacuzzi (establecimientos turísticos, recreativos, deportivos)	Ocio y recreación
Lugares de actividades culturales (cines, teatros, museos, bibliotecas, etc.)	Ocio y recreación
Centros educativos	Institucional
Centro de atención médica (hospital, clínica, consultorio, etc.)	Institucional
Oficinas y/o edificios de oficinas	Institucional
Establecimientos dedicados a la administración pública (cortes, juzgados, ministerios, etc.)	Institucional
Guarniciones militares y/o de policía	Institucional
Centros de reclusión	Institucional
Lugares de cuidado de personas (hospicios, orfelinatos, hogares geriátricos, etc.)	Institucional

Vehículo de servicio particular	Transporte privado
Establecimiento industrial (fábrica, planta) y/o obras en construcción	Industrial
Taller	Industrial
Sitio de culto (capilla, iglesia, templo, etc.)	Religioso
Otros	Sin información/Otros
Sin información	Sin información/Otros

Tabla A4: *Recategorización de las categorías de la variable Actividad Durante el Hecho.*

Categoría original	**Categoría recategorizada**
Actividades vitales o relacionadas con el cuidado personal	Cuidado personal
Actividades de desplazamiento de un lugar a otro	Desplazamiento
Actividades de trabajo doméstico no pagado para el uso del propio hogar	Hogar y cuidado
Actividades relacionadas con el cuidado no pagado de miembros del hogar	Hogar y cuidado
Actividades relacionadas con la asistencia a eventos culturales, de entretenimiento y/o deportivos	Recreación
Actividades de tiempo libre	Recreación
Actividades relacionadas con los deportes y el ejercicio físico	Recreación
Actividades relacionadas con el trabajo remunerado	Trabajo
Actividades relacionadas con el estudio y el aprendizaje	Educación
Actividades ilícitas o delictivas	Actividad delictiva
Actividades relacionadas con manifestaciones públicas (marchas, protestas, etc.)	Sin información/Otros
Otros	Sin información/Otros
Sin información	Sin información/Otros

Tabla A5: *Recategorización de las categorías de la variable Circunstancia del Hecho.*

Categoría original	**Categoría recategorizada**
Conflicto de pareja	Conflicto relacional

Celos	Conflicto relacional
Infidelidad	Conflicto relacional
Conflicto familiar	Conflicto familiar
Intolerancia	Intolerancia social
Amenaza o intimidación	Intolerancia social
Embriaguez (alcohólica y no alcohólica)	Consumo de sustancias
Consumo de alcohol y/o sustancias psicoactivas	Consumo de sustancias
Machismo	Violencia estructural
Violencia económica	Violencia estructural
Económicas	Violencia estructural
Enfermedad física o mental	Condición de salud
Abandono	Negligencia/abandono
Negligencia	Negligencia/abandono
Ajuste de cuentas	Criminal/extrema
Reclutamiento de niños, niñas y adolescentes	Criminal/extrema
Presunta esclavitud sexual o prostitución forzada	Criminal/extrema
Esclavitud o trabajo forzoso	Criminal/extrema
Otros	Sin información/Otros
Sin información	Sin información/Otros

Tabla A6: *Recategorización de las categorías de la variable Mecanismo Causal de la Lesión.*

Categoría original	**Categoría recategorizada**
Contundente	Contundente
Corto contundente	Contundente
Cortante	Cortante/punzante
Punzante	Cortante/punzante
Corto punzante	Cortante/punzante
Abrasivo	Abrasivo
Proyectil de arma de fuego	Arma de fuego
Térmico	Térmico/eléctrico

Eléctrico	Térmico/eléctrico
Agente químico corrosivo	Químico/tóxico
Agente químico irritante	Químico/tóxico
Agente químico tóxico	Químico/tóxico
Tóxico	Químico/tóxico
Agentes y mecanismos biológicos	Biológico/asfixia
Generadores de asfixia	Biológico/asfixia
Agentes y mecanismo explosivo	Por determinar
Biodinámico	Biodinámico
Mecanismo múltiple	Múltiple
Por determinar	Por determinar
Otros	Sin información/Otros
Sin información	Sin información/Otros

Tabla A7: *Recategorización de las categorías de la variable Presunto Agresor.*

Categoría original	**Categoría recategorizada**
Compañero(a) permanente	Pareja actual
Esposo(a)	Pareja actual
Novio(a)	Pareja actual
Amante	Pareja actual
Ex compañero(a) permanente	Ex pareja
Ex esposo(a)	Ex pareja
Ex novio(a)	Ex pareja
Ex amante	Ex pareja
Padre	Familia nuclear
Madre	Familia nuclear
Hijo(a)	Familia nuclear
Hermano(a)	Familia nuclear
Padrastro	Familia nuclear
Madrastra	Familia nuclear

Tío(a)	Familia extendida
Primo(a)	Familia extendida
Sobrino(a)	Familia extendida
Cuñado(a)	Familia extendida
Suegro(a)	Familia extendida
Yerno	Familia extendida
Nuera	Familia extendida
Abuelo(a)	Familia extendida
Nieto(a)	Familia extendida
Otros familiares civiles o consanguíneos	Familia extendida
Encargado del cuidado	Cuidador
Otros	Sin información/Otros
Sin información	Sin información/Otros

Se presentan figuras complementarias, las cuales contienen las distribuciones porcentuales de algunas variables.

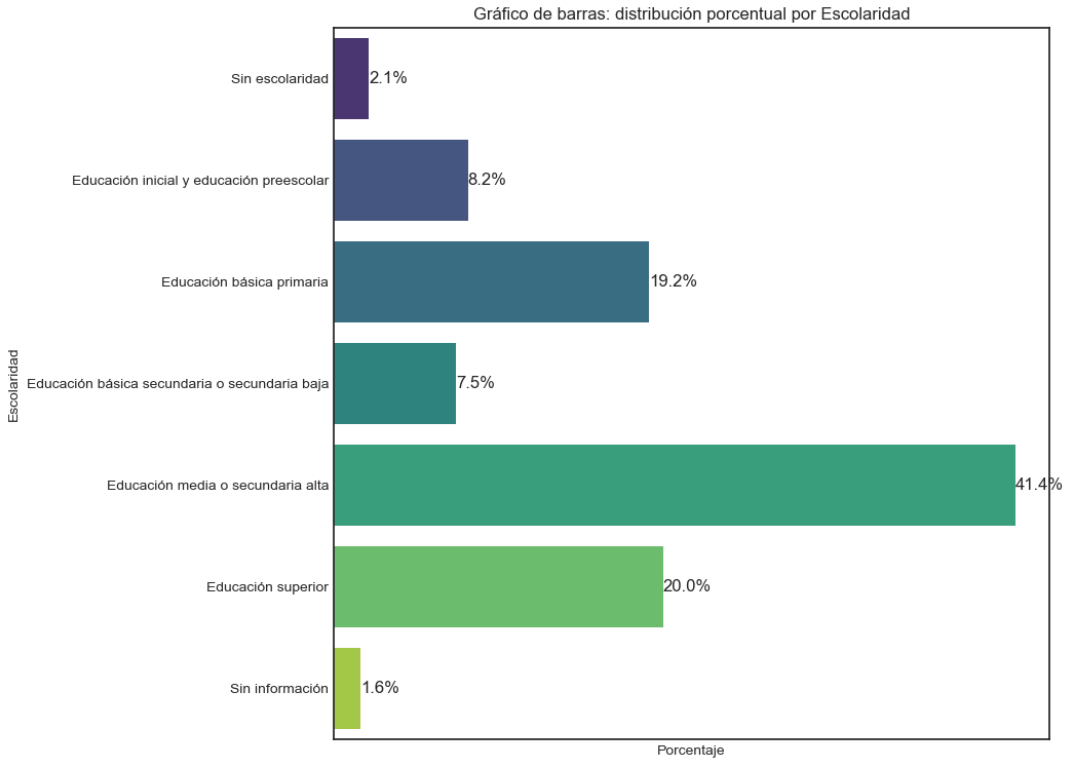


Figura A1: *Distribución porcentual por escolaridad.*

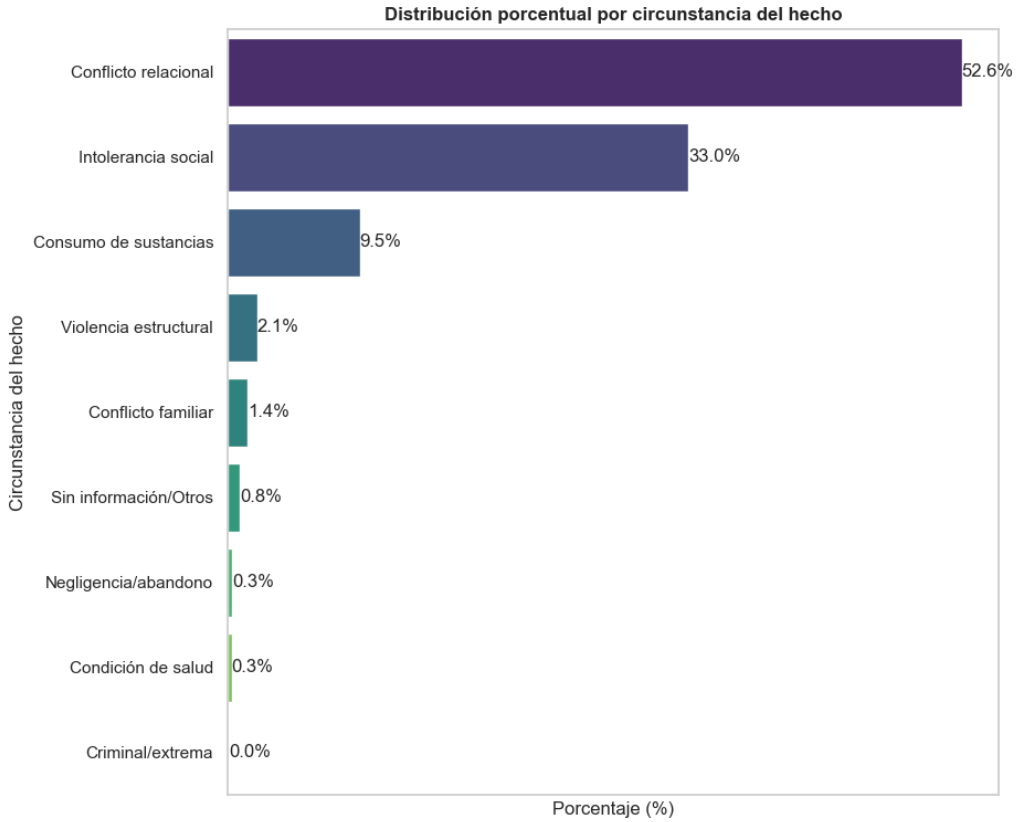


Figura A2: *Distribución porcentual de Circunstancia del hecho detallada.*

Anexo B. Información complementaria de las tasas construidas

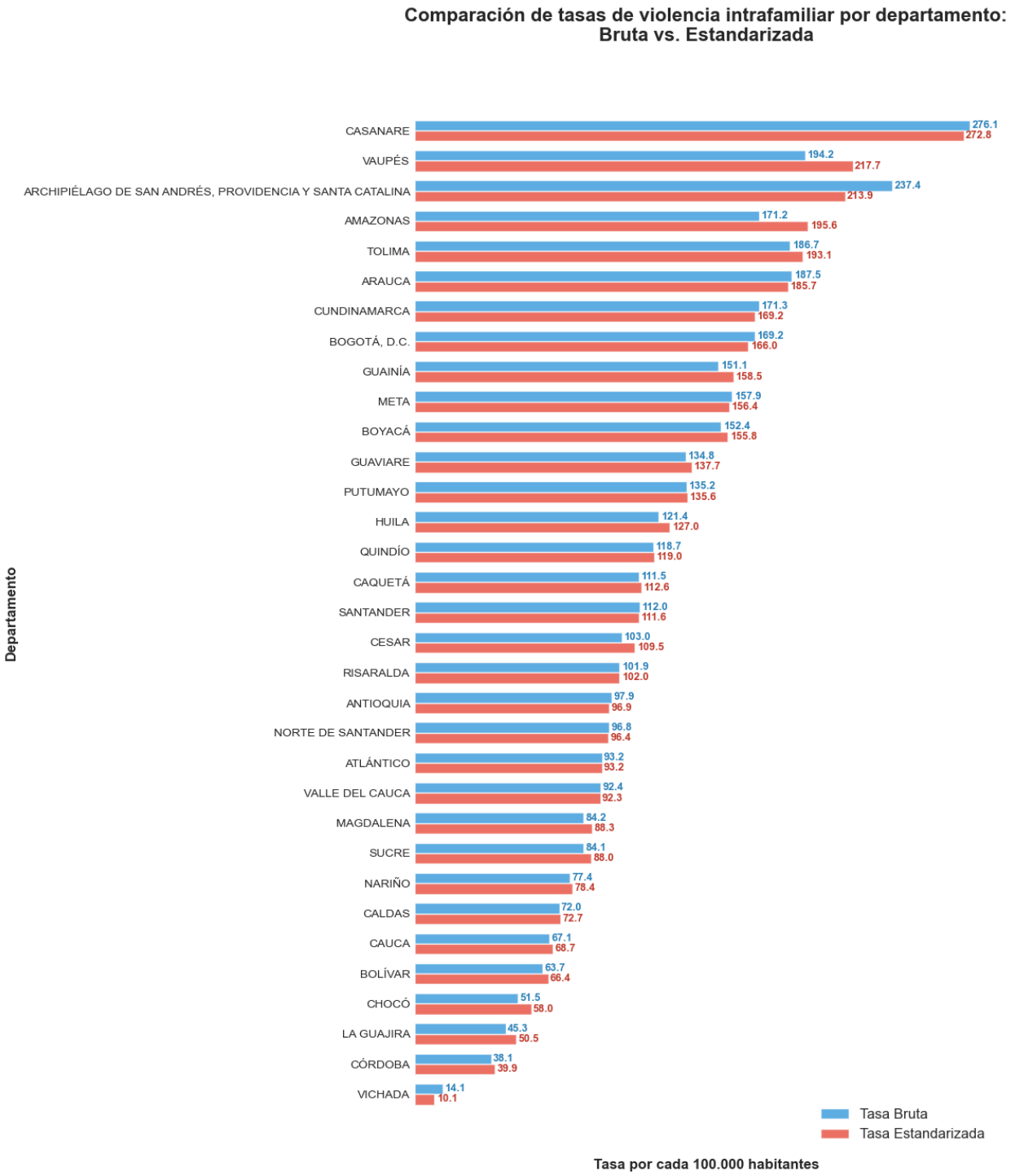


Figura B1: *Comparación de Tasa Bruta y Tasa Estandarizada por departamento.*

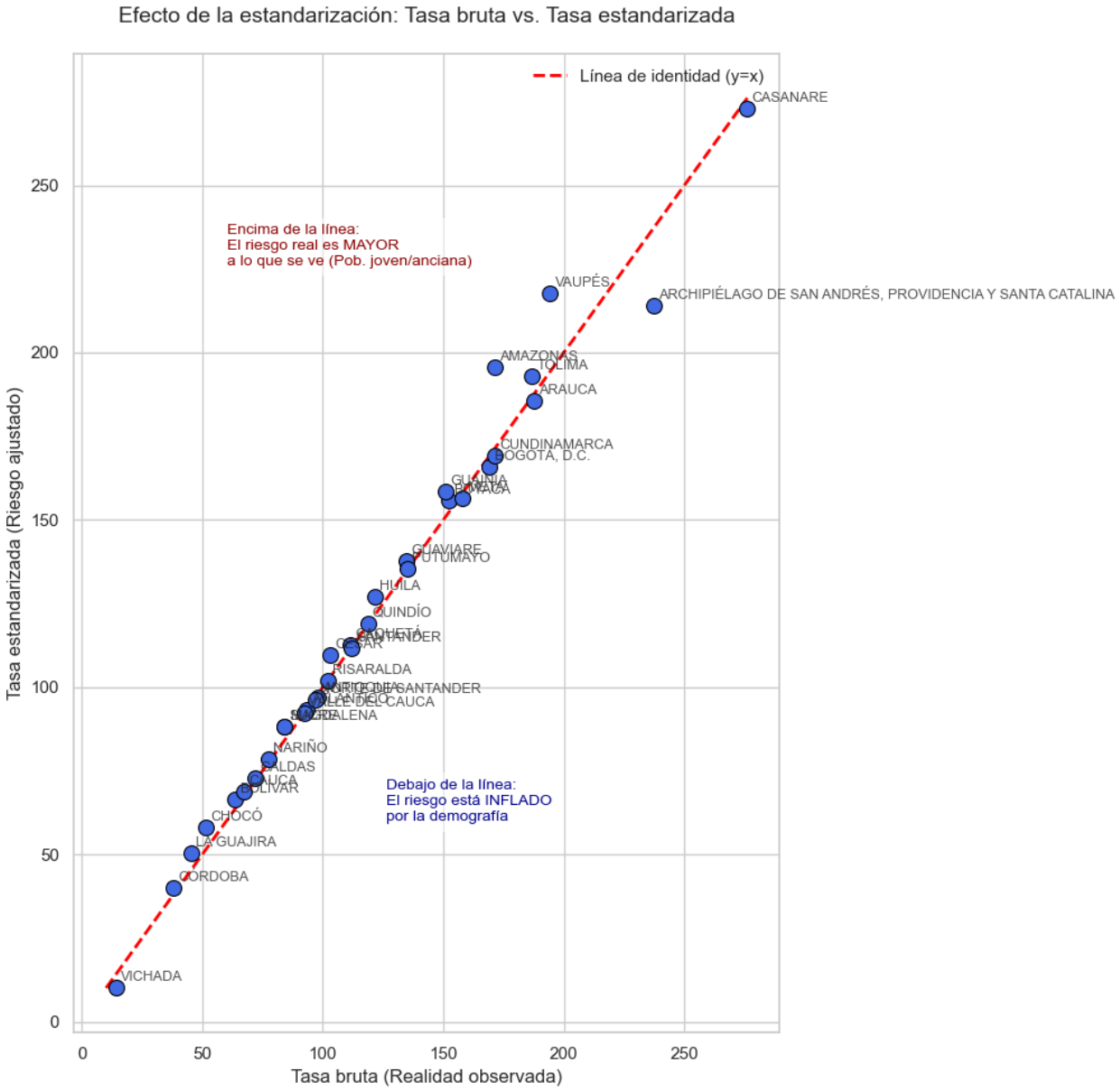


Figura B2: *Relación entre Tasa Bruta y Tasa Estandarizada (Análisis Comparativo).*

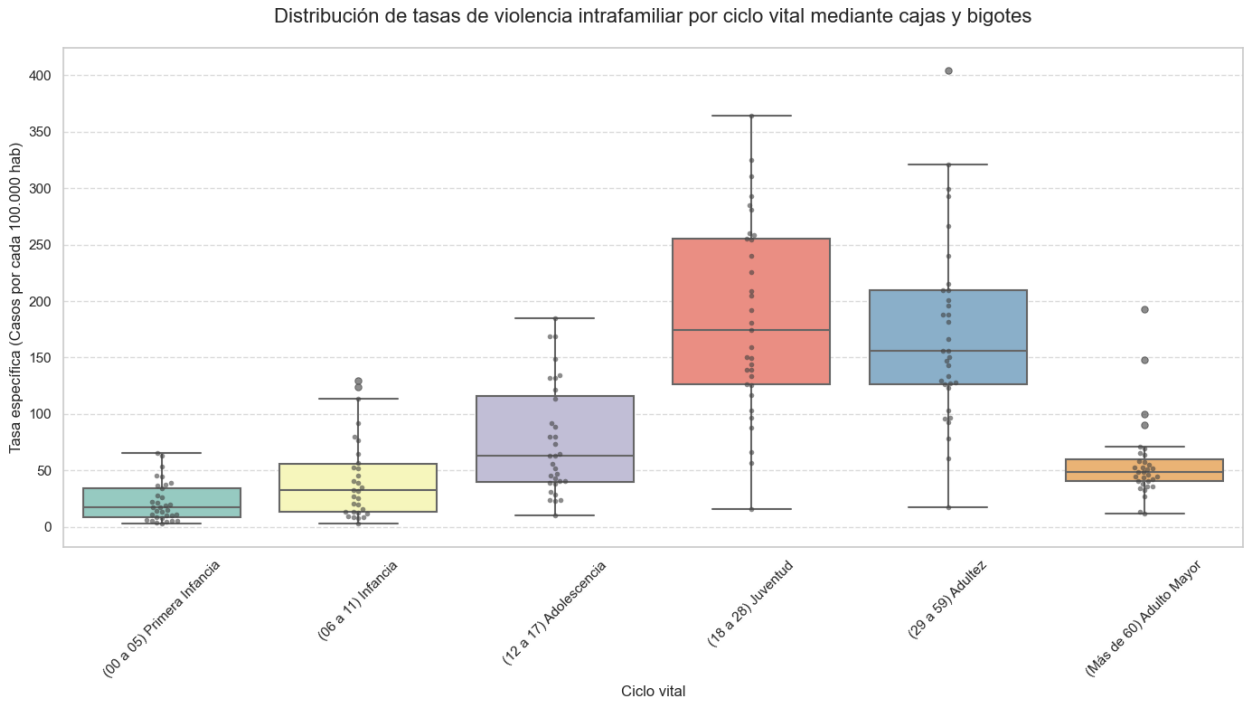


Figura B3: *Cajas y bigotes de la tasa específica por ciclo vital.*

Anexo C. Tablas y Figuras del ACM I

A continuación, se presentan los valores propios (eigenvalues) del ACM I, los cuales indican la proporción de variabilidad explicada por cada dimensión, junto con una representación gráfica que muestra el porcentaje de inercia explicado por cada una de ellas. Este análisis permite contar con un soporte tanto cuantitativo como visual para determinar el número de dimensiones más relevantes a considerar en la interpretación y el análisis detallado del ACM I.

Tabla C1: *Valores propios y varianza explicada del ACM I.*

Dimensión	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9	Dim.10
Valores Propios	0,41	0,25	0,20	0,17	0,17	0,16	0,16	0,16	0,15	0,14
Varianza explicada (%)	8,28	5,04	3,97	3,46	3,37	3,28	3,26	3,14	2,97	2,75
Varianza Acumulada	8,2	13,3	17,2	20,7	24,1	27,4	30,6	33,8	36,7	39,5

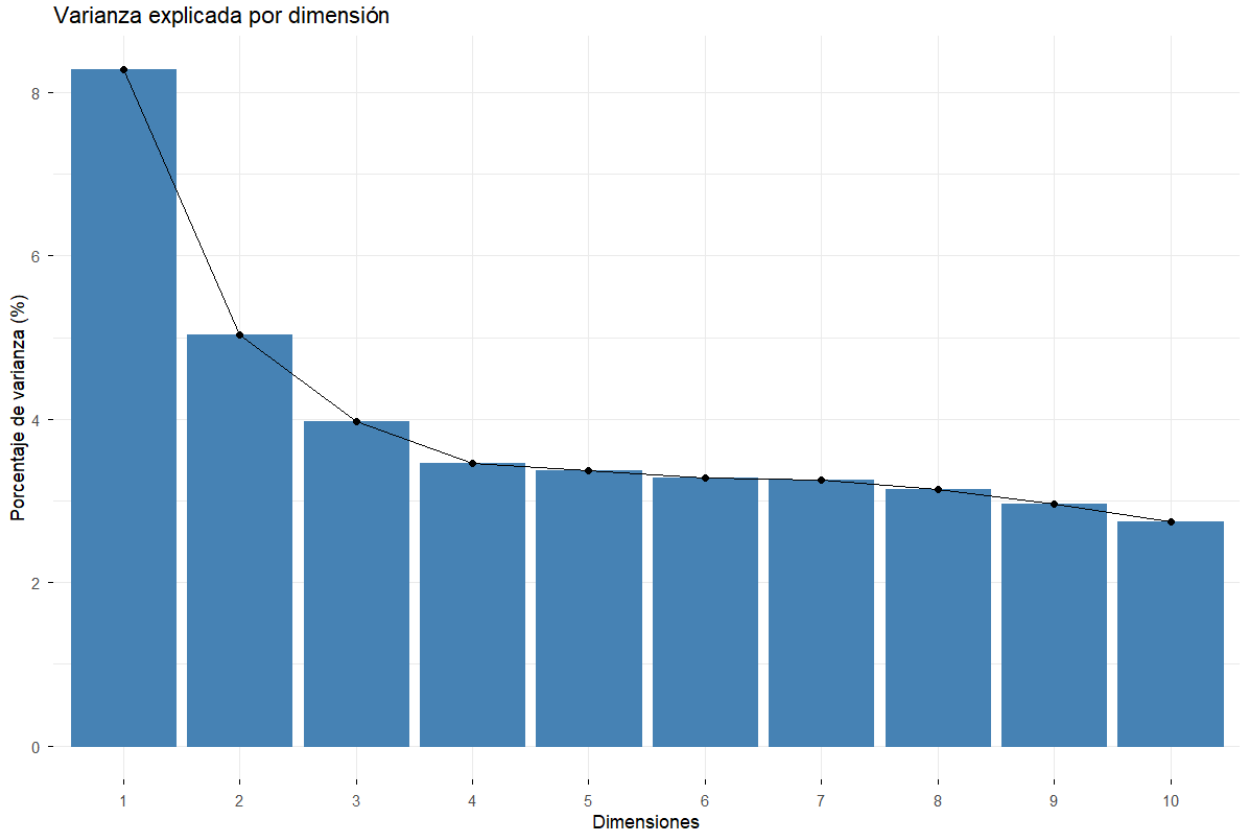


Figura C1: *Varianza explicada por dimensión del ACM I.*

Se presenta el resumen del ACM I, donde se incluyen las tablas con los valores de η^2 de las variables que contribuyen a la construcción del espacio factorial, las mayores contribuciones en las primeras cuatro dimensiones y las categorías mejor representadas (cos^2) en dichas dimensiones.

Tabla C2: *Variables con mayor aporte a la varianza explicada en ACM I.*

Variable	**Dim1**	**Dim2**	**Dim3**	**Dim4**	**Dim5**
Contexto del hecho	0.86	0.71	0.56	0.30	0.02
Presunto agresor detallado	0.72	0.25	0.08	0.15	0.04
Ciclo vital	0.65	0.60	0.60	0.47	0.13
Circunstancia del hecho	0.54	0.22	0.11	0.15	0.61
Estado civil	0.42	0.43	0.31	0.23	0.07
Escolaridad	0.36	0.09	0.20	0.18	0.10
Factor desencadenante de la agresión	0.28	0.17	0.07	0.17	0.61
Sexo de la víctima	0.18	0.02	0.01	0.01	0.03

Sexo del agresor	0.13	0.00	0.02	0.02	0.07
Días de incapacidad medicolegal	0.00	0.02	0.04	0.05	0.00

Tabla C3*: Categorías que más aportan a las dimensiones 1 y 2.*

Dimensión 1		**Dimensión 2**	
Categoría	**%**	**Categoría**	**%**
Violencia contra NNA	11.62	Violencia contra el adulto mayor	12.05
Presunto agresor: Familia nuclear	9.34	Estado civil: No aplica	11.85
Estado civil: No aplica	8.82	Ciclo vital: Adulto mayor (60+)	10.74
Ciclo vital: Infancia (6–11)	5.87	Presunto agresor: Familia extendida	7.49
Circunstancia: Intolerancia social	5.75	Ciclo vital: Primera infancia (0–5)	7.41
Violencia de pareja	5.55	Violencia contra NNA	7.25
Circunstancia: Conflicto relacional	5.29	Violencia entre familiares	7.08
Ciclo vital: Primera infancia (0–5)	4.12	Ciclo vital: Infancia (6–11)	4.80
Escolaridad: Educación inicial/preescolar	3.89	Estado civil: Viudo(a)	3.59

Tabla C4*: Categorías con mayor aporte en las dimensiones 3 y 4.*

Dimensión 3		**Dimensión 4**	
Categoría	**%**	**Categoría**	**%**
Ciclo vital: Adulto mayor (60+)	18.35	Ciclo vital: Adolescencia (12–17)	13.06
Violencia contra el adulto mayor	16.85	Violencia entre familiares	9.34
Violencia entre familiares	9.64	Ciclo vital: Primera infancia (0–5)	5.42
Ciclo vital: Adolescencia (12–17)	6.12	Circunstancia: Consumo de sustancias	5.01

Estado civil: Viudo(a)	5.93	Factor desencadenante: Consumo de alcohol y/o SPA	4.91
Escolaridad: Sin escolaridad	5.83	Estado civil: Unión libre	4.76
Ciclo vital: Primera infancia (0–5)	4.62	Escolaridad: Sin escolaridad	4.48

Tabla C5*: Categorías que mejor se representan en las dimensiones 1 y 2.*

Dimensión 1		**Dimensión 2**	
Categoría	**cos²**	**Categoría**	**cos²**
Violencia de pareja	0.69	Violencia contra el adulto mayor	0.32
Violencia contra NNA	0.53	Estado civil: No aplica	0.31
Presunto agresor: Familia nuclear	0.49	Ciclo vital: Adulto mayor (60+)	0.29
Circunstancia: Conflicto relacional	0.47	Violencia entre familiares	0.22
Estado civil: No aplica	0.38	Presunto agresor: Familia extendida	0.22
Circunstancia: Intolerancia social	0.36	Violencia contra NNA	0.20
Ciclo vital: Infancia (6–11)	0.25	Ciclo vital: Primera infancia (0–5)	0.19
Factor desencadenante: Celos, desconfianza o infidelidad	0.20	Violencia de pareja	0.13
Presunto agresor: Ex pareja	0.19	Ciclo vital: Infancia (6–11)	0.12

Tabla C6*: Categorías que mejor se ven representadas en las dimensiones 3 y 4.*

Dimensión 3		**Dimensión 4**	
Categoría	**cos²**	**Categoría**	**cos²**
Ciclo vital: Adulto mayor (60+)	0.39	Ciclo vital: Adolescencia (12–17)	0.24

Violencia contra el adulto mayor	0.35	Violencia entre familiares	0.20
Violencia entre familiares	0.24	Ciclo vital: Adultez (29–59)	0.14
Estado civil: Soltero(a)	0.14	Estado civil: Soltero(a)	0.14
Ciclo vital: Adolescencia (12–17)	0.13	Estado civil: Unión libre	0.12
Estado civil: Viudo(a)	0.12	Factor desencadenante: Consumo de alcohol y/o SPA	0.10
Escolaridad: Sin escolaridad	0.12	Circunstancia: Consumo de sustancias	0.10
Ciclo vital: Primera infancia (0–5)	0.09	Ciclo vital: Primera infancia (0–5)	0.10
Circunstancia: Intolerancia social	0.09	Escolaridad: Sin escolaridad	0.08

Ahora se presentan los gráficos correspondientes, donde se muestra la distribución de los individuos en las dimensiones, junto con las 12 categorías que más contribuyen a la construcción de estos ejes factoriales.

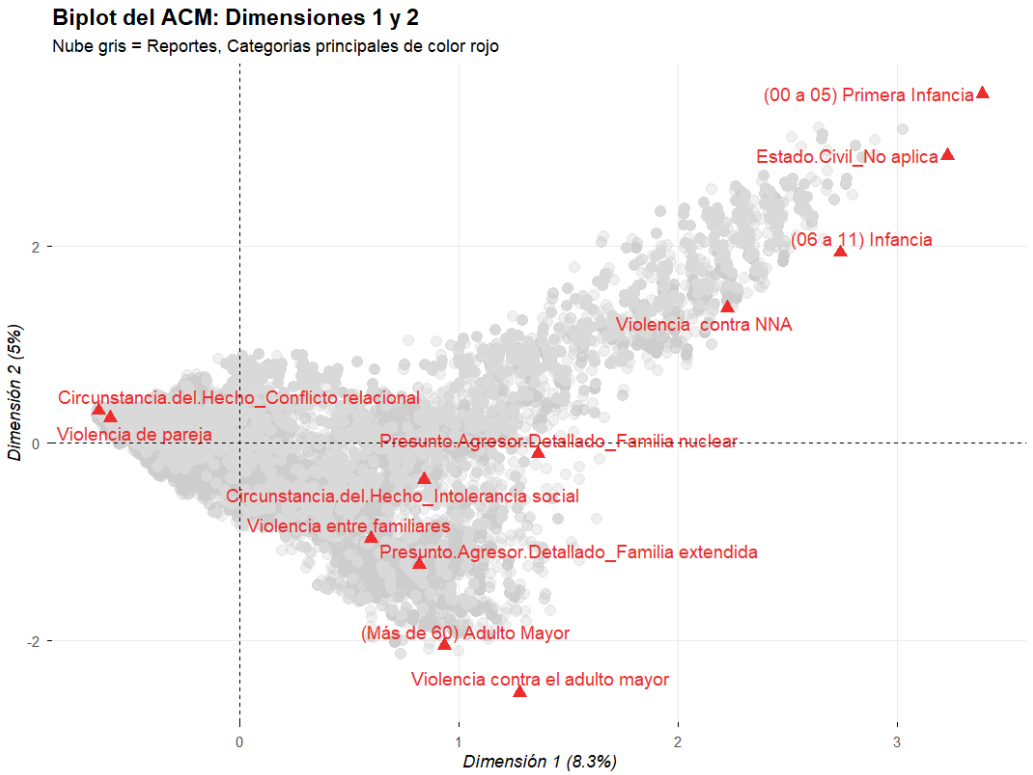


Figura C2: *Representación factorial del ACM I en dimensiones 1 y 2.*

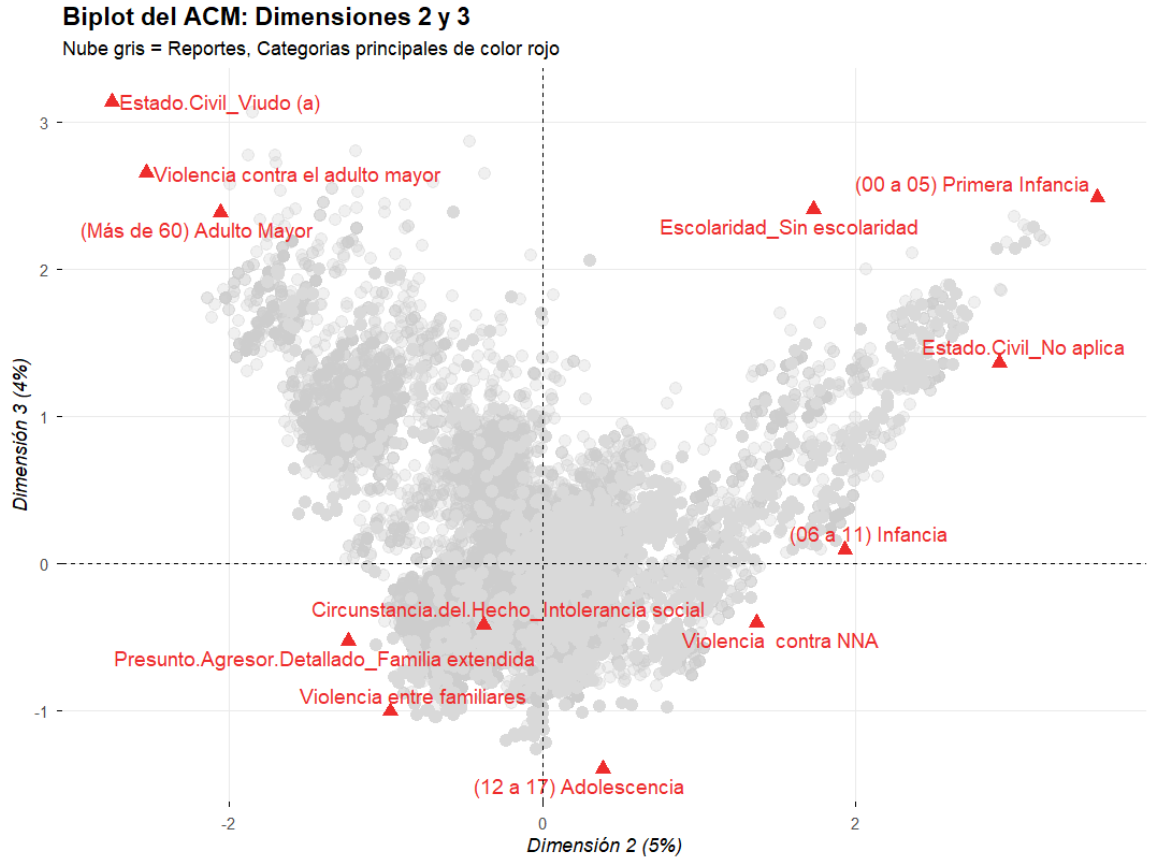


Figura C3: *Representación factorial del ACM I en dimensiones 2 y 3.*

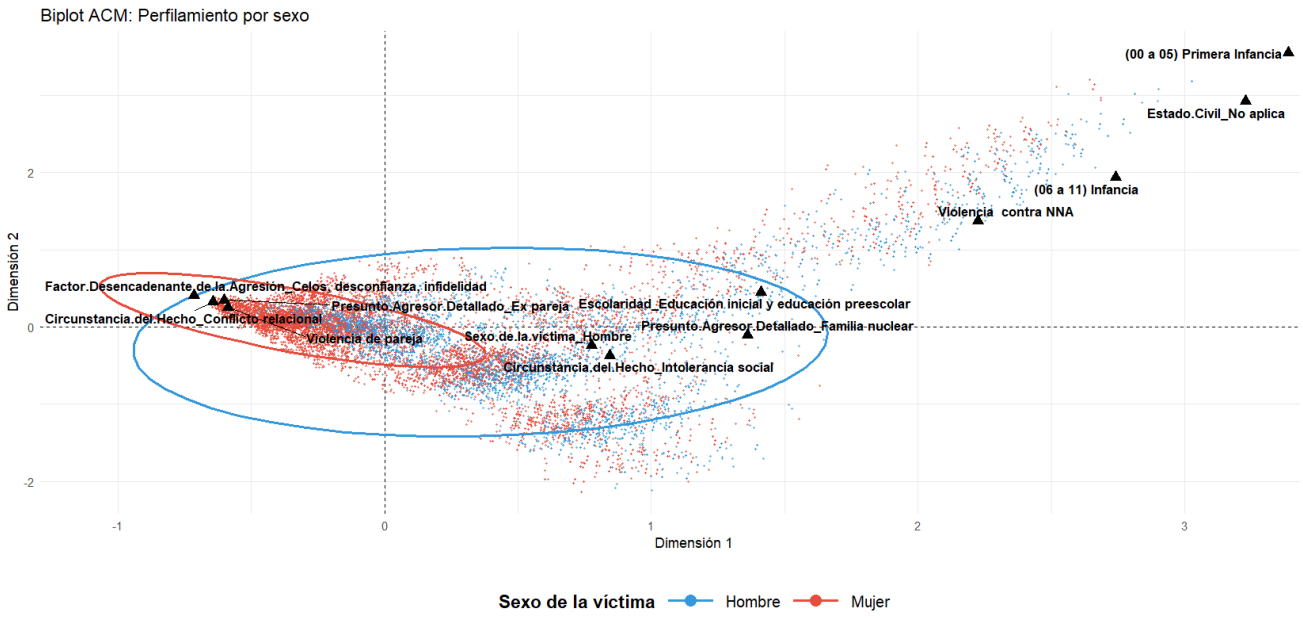


Figura C4: *Mapa factorial del ACM I por género (Dimensiones 1 y 2).*

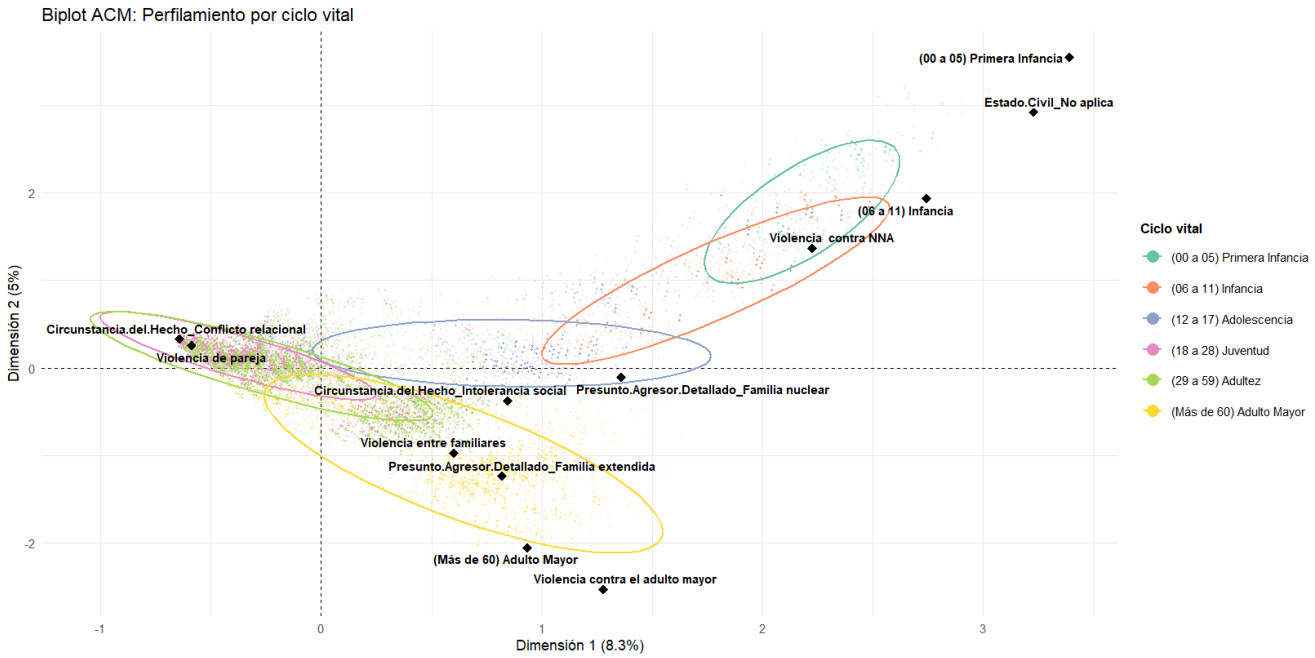


Figura C5*: Mapa factorial del ACM I por ciclo vital (Dimensiones 1 y 2).*

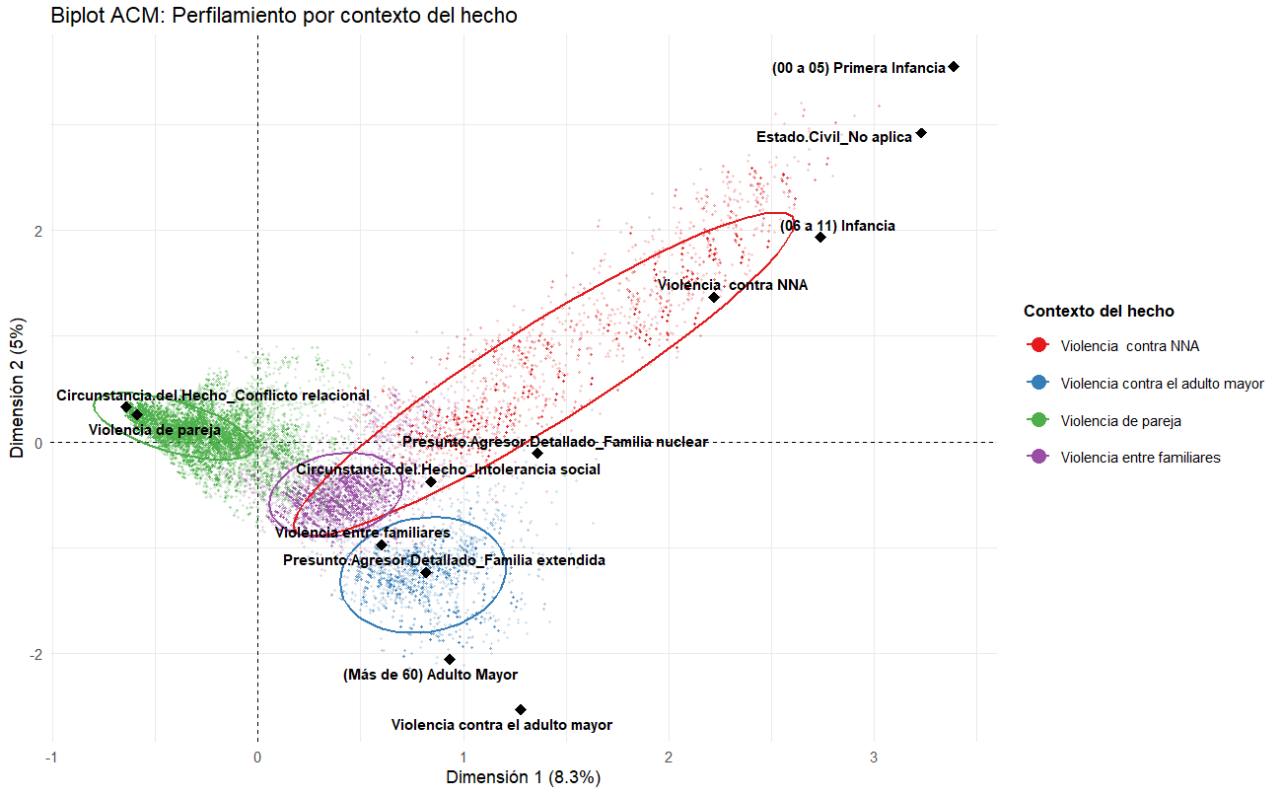


Figura C6*: Mapa factorial del ACM I por contexto del hecho (Dimensiones 1 y 2).*

Los gráficos de perfilamiento, basados en las Dimensiones 1 y 2 por su mayor capacidad discriminante, muestran patrones coherentes con el ACM. El sexo de la víctima evidencia que las

mujeres se asocian principalmente con violencia de pareja, mientras que los hombres presentan mayor dispersión hacia infancia, adulto mayor y violencia familiar. El ciclo vital confirma una correspondencia clara entre edades y tipos de violencia, y el contexto del hecho muestra una buena separación de grupos, destacándose como variable clave en la segmentación. En conjunto, estos resultados validan la consistencia entre las métricas del ACM y la distribución visual de los individuos.

Anexo D. Tablas de valores propios, contribuciones y representatividad del ACM II

A continuación, se presenta el resumen del **ACM II**, que incluye las tablas con los valores de η^2 de las variables que participan en la construcción del espacio factorial, junto con las principales contribuciones en las primeras cuatro dimensiones y las categorías mejor representadas (cos^2).

Tabla D1*: Variables con mayor aporte a la varianza explicada.*

Variable	**Dim 1**	**Dim 2**	**Dim 3**	**Dim 4**
Ciclo Vital	0,659507	0,519679	0,203982	0,128848
Presunto Agresor Detallado	0,609641	0,218812	0,090479	0,011661
Circunstancia del Hecho	0,505417	0,321777	0,554556	0,673008
Estado Civil	0,498693	0,353357	0,195111	0,084988
Escolaridad	0,416971	0,27576	0,074839	0,09016
Factor Desencadenante de la Agresión	0,288468	0,258931	0,563512	0,671597
Sexo de la victima	0,208083	0,042989	0,004436	0,000817
Sexo del Agresor	0,150424	0,024604	0,046827	0,001964
Días de Incapacidad Medicolegal	0,001005	0,023638	0,004555	0,008897

Tabla D2*: Categorías que más aportan las dimensiones 1 y 2.*

Dimensión 1 – Categoría	**Contrib. (%)**	**Dimensión 2 – Categoría**	**Contrib. (%)**
Estado civil: No aplica	13.33	Estado civil: No aplica	14.66
Presunto agresor: Familia nuclear	10.00	Ciclo vital: Primera infancia (0–5)	14.59
Ciclo vital: Infancia (6–11)	7.75	Escolaridad: Sin escolaridad	8.19
Ciclo vital: Primera infancia (0–5)	6.64	Presunto agresor: Familia extendida	6.18

Circunstancia: Intolerancia social	6.43	Circunstancia: Intolerancia social	4.58
Circunstancia: Conflicto relacional	5.84	Ciclo vital: Adulto mayor (60+)	4.45
Escolaridad: Educación inicial / preescolar	5.79	Ciclo vital: Adolescencia (12–17)	4.24
Sexo de la víctima: Hombre	4.79	Circunstancia: Conflicto relacional	3.64
Factor desencadenante: Celos, desconfianza o infidelidad	4.19	Escolaridad: Educación básica primaria	3.60
Escolaridad: Sin escolaridad	4.02	Factor desencadenante: Celos, desconfianza o infidelidad	3.09

Tabla D3*: Categorías que más aportan las dimensiones 3 y 4.*

Dimensión 3 – Categoría	Contrib. (%)	Dimensión 4 – Categoría	Contrib. (%)
Circunstancia del Hecho: Consumo de sustancias	25,54	Circunstancia del Hecho: Conflicto familiar	30,97
Factor Desencadenante: Consumo de alcohol y/o sustancias psicoactivas	25,27	Factor Desencadenante: Otras razones	23,89
Factor Desencadenante: Intolerancia, machismo	4,97	Factor Desencadenante: Sin información	8,57
Estado Civil: Soltero(a)	4,78	Adolescencia (12 a 17)	4,77
Circunstancia del Hecho: Intolerancia social	4,09	Factor Desencadenante: Intolerancia, machismo	4,42
Adolescencia (12 a 17)	3,78	Circunstancia del Hecho: Consumo de sustancias	3,72
Escolaridad: Sin escolaridad	3,31	Factor Desencadenante: Consumo de alcohol y/o sustancias psicoactivas	3,08
Adulto mayor (>60)	3,08	Circunstancia del Hecho: Intolerancia social	3,03
Estado Civil: Unión libre	2,72	Circunstancia del Hecho: Sin información/Otros	2,20
Presunto agresor: Ex pareja	2,60	Estado Civil: No aplica	1,91

Tabla D4*: Categorías que mejor están representadas en las dimensiones 1 y 2.*

Dimensión 1 – Categoría	**cos²**	**Dimensión 2 – Categoría**	**cos²**
Estado civil: No aplica	0.46	Estado civil: No aplica	0.31
Presunto agresor: Familia nuclear	0.42	Ciclo vital: Primera infancia (0–5)	0.30
Circunstancia: Conflicto relacional	0.42	Escolaridad: Sin escolaridad	0.17
Circunstancia: Intolerancia social	0.32	Circunstancia: Conflicto relacional	0.16
Ciclo vital: Infancia (6–11)	0.27	Presunto agresor: Familia extendida	0.14
Ciclo vital: Primera infancia (0–5)	0.22	Circunstancia: Intolerancia social	0.14
Escolaridad: Educación inicial / preescolar	0.21	Ciclo vital: Adulto mayor (60+)	0.10
Sexo de la víctima: Hombre	0.21	Ciclo vital: Adolescencia (12–17)	0.09
Sexo de la víctima: Mujer	0.21	Escolaridad: Educación básica primaria	0.09
Factor desencadenante: Celos, desconfianza o infidelidad	0.19	Factor desencadenante: Celos, desconfianza o infidelidad	0.09

Tabla D5*: Categorías que mejor están representadas en las dimensiones 3 y 4.*

Dimensión 3 – Categoría	**cos²**	**Dimensión 4 – Categoría**	**cos²**
Factor desencadenante: Consumo de alcohol y/o sustancias psicoactivas	0.52	Circunstancia: Conflicto familiar	0.52
Circunstancia: Consumo de sustancias	0.49	Factor desencadenante: Otras razones	0.40
Factor desencadenante: Intolerancia, machismo	0.16	Factor desencadenante: Sin información	0.15
Estado civil: Soltero(a)	0.16	Factor desencadenante: Intolerancia, machismo	0.14
Circunstancia: Intolerancia social	0.11	Ciclo vital: Adolescencia (12–17)	0.09
Ciclo vital: Adolescencia (12–17)	0.07	Circunstancia: Intolerancia social	0.08
Presunto agresor: Ex pareja	0.07	Circunstancia: Consumo de sustancias	0.07
Estado civil: Unión libre	0.07	Factor desencadenante: Consumo de alcohol y/o sustancias psicoactivas	0.06
Presunto agresor: Pareja actual	0.07	Estado civil: Soltero(a)	0.05
Escolaridad: Sin escolaridad	0.06	Circunstancia: Sin información / otros	0.04

Tabla D6*: Variables suplementarias que mejor se ven representadas en cada dimensión.*

	Dim.1	Dim.2	Dim.3
Pertenencia.Grupal	0,015	0,001	0,001
Zona.del.Hecho	0,002	0	0,005
Escenario.del.Hecho	0,01	0,003	0,007
Actividad.Durante.el.Hecho	0,026	0,006	0,014
Mecanismo.Causal.de.la.Lesión.no.Fatal	0,007	0,026	0,002
Diagnostico.Topográfico.de.la.Lesión.no.Fatal	0,008	0,02	0
Tipo.de.Discapacidad	0,002	0,001	0,001
Orientación.Sexual	0,361	0,309	0,029
Contexto.del.Hecho	0,731	0,266	0,073

Anexo E. Tablas y Figuras complementarias de la técnica Fuzzy C-Means

Tabla E1*: Métricas de validación para la selección de clústeres.*

Cantidad de clusters(k)	**PC**	**PE**
3	0.6478	0.6368
4	0.6469	0.6845
5	0.5184	0.9291
6	0.5177	0.9539

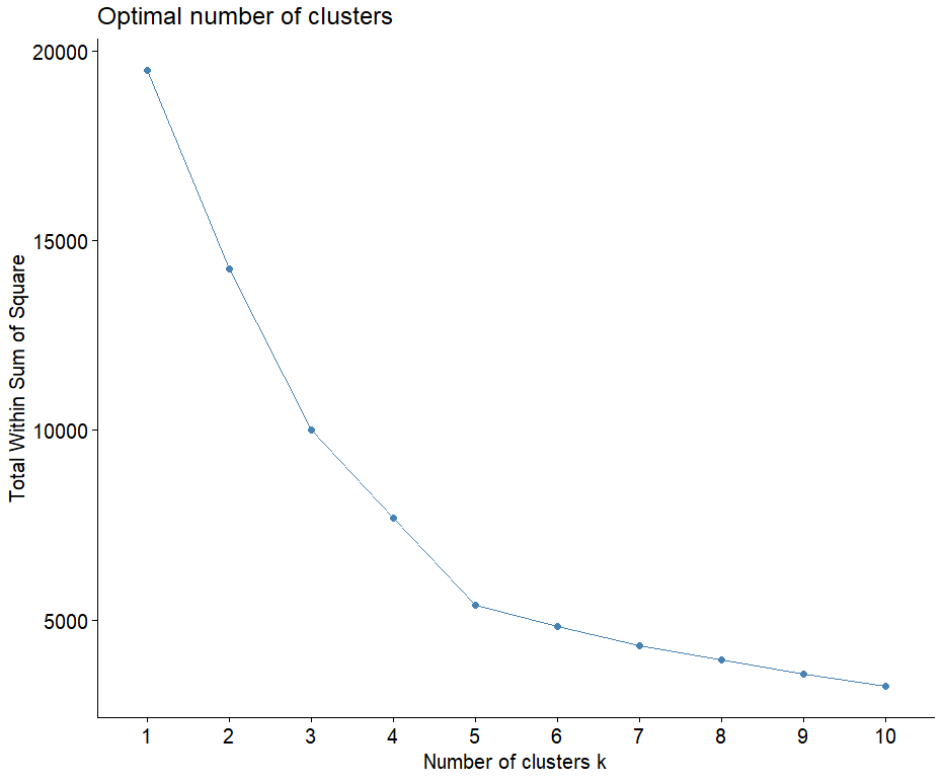


Figura E1: *Método del codo para determinar el número de clusters óptimo.*

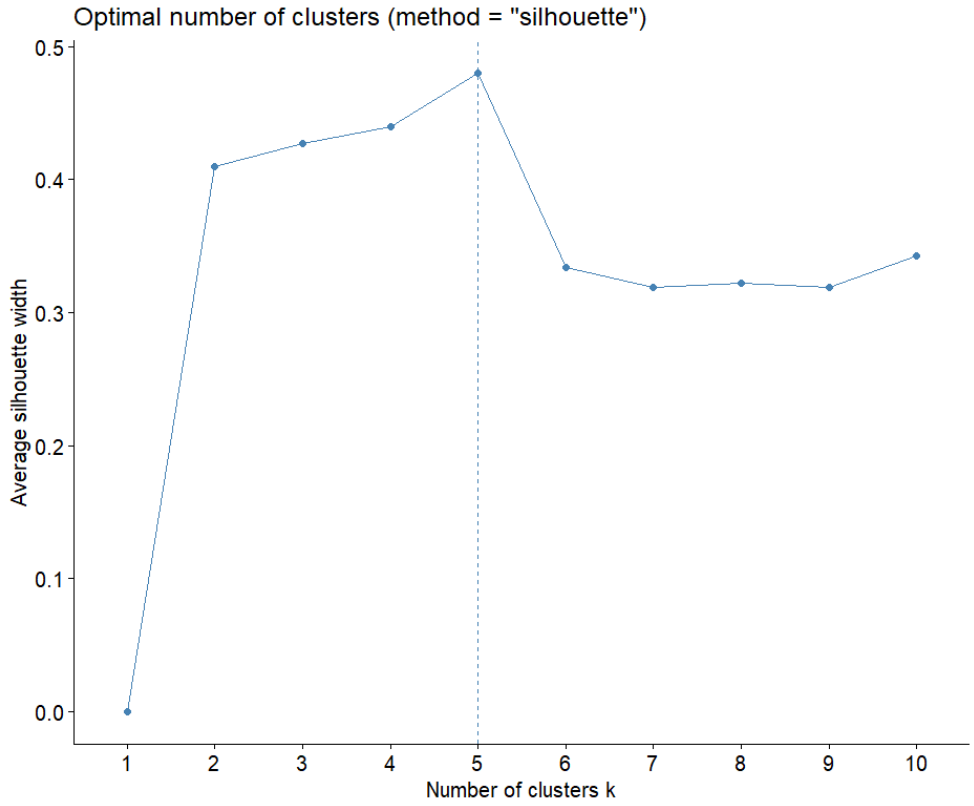


Figura E2: *Método de la silueta para encontrar el número de clusters óptimo.*

Tabla E2*: Frecuencia y porcentaje por calidad de pertenencia.*

Tabla de pertenencia			
	Muy Ambiguo [0 , 0.4]	Intermedio [0.4 , 0.7]	Perfil Claro [0.7 , 1]
Frecuencia	2291	17357	40269
%	3.82	28.96	67.20

Tabla E3*: Media de certeza de pertenencia de cada cluster.*

Tabla certeza	
Cluster_Fuzzy	Certeza_Promedio
1	0,76
2	0,75
3	0,72
4	0,74

Tabla E4*: Distribución porcentual del contexto del hecho por cluster.*

Cluster	**Contexto de violencia**			
	Violencia contra NNA	Violencia contra el adulto mayor	Violencia de pareja	Violencia entre familares
1	0,01	0,01	99,25	0,73
2	7,89	10,92	45,73	35,46
3	99,01	0,66	0,00	0,33
4	19,01	11,57	21,11	48,31

Tabla E5*: Distribución porcentual del sexo de la víctima por cluster.*

Cluster	**Sexo de la victima**	
	Hombre	Mujer
1	7,87	92,12
2	27,78	72,21
3	56,27	43,72
4	46,8	53,19

Tabla E6*: Distribución porcentual de circunstancias del hecho por clúster Parte I.*

Cluster	**Circunstancias del hecho**		
	Negligencia/abandono	Sin información/Otros	Violencia estructural
1	0.06	0.17	2.94
2	0.22	0.06	1.25
3	4.46	8.09	0.37
4	0.06	2.98	0.91

Tabla E7*: Distribución porcentual de circunstancias del hecho por cluster Parte II.*

Cluster	**Circunstancias del hecho**					
	Condición de salud	Conflicto familiar	Conflicto relacional	Consumo de sustancias	Criminal/extrema	Intolerancia social
1	0.01	0.00	85.30	1.05	0.02	10.45
2	0.02	0.60	12.57	0.08	0.08	85.11
3	0.41	10.85	1.49	7.18	0.21	66.96
4	1.74	5.66	18.29	62.44	0.00	7.91

Tabla E8*: Distribución porcentual del presunto agresor por cluster.*

Cluster	**Presunto agresor**					
	Cuidador	Ex pareja	Familia extendida	Familia nuclear	Pareja actual	Sin información
1	0.00	55.88	0.39	0.35	43.37	0.00
2	0.20	10.83	32.26	46.43	10.28	0.01
3	3.63	0.00	10.64	83.46	0.00	2.27
4	0.09	9.31	18.62	33.41	36.06	2.52

Tabla E9*: Valores Gap de pertenencia.*

	`Cluster 1`	`Cluster 2`	`Cluster 3`	`Cluster 4`
`Cluster 1`	0	1069	0	1631
`Cluster 2`	1174	0	18	772
`Cluster 3`	2	55	0	133
`Cluster 4`	966	589	117	0

La métrica de *gap de pertenencia* evalúa la incertidumbre en la asignación de individuos a clústeres, midiendo la diferencia entre su mayor y segunda mayor probabilidad de pertenencia; valores bajos (menores a 0.15) indican ambigüedad. Los resultados muestran un solapamiento entre el Clúster 1 (violencia de pareja) y el Clúster 4 (violencia mixta o transicional), así como entre el Clúster 1 y el Clúster 2 (violencia relacional asociada a sustancias).En contraste, el Clúster 3 (violencia contra NNA) presenta bajo solapamiento con los demás, evidenciando un perfil más independiente y claramente diferenciado.

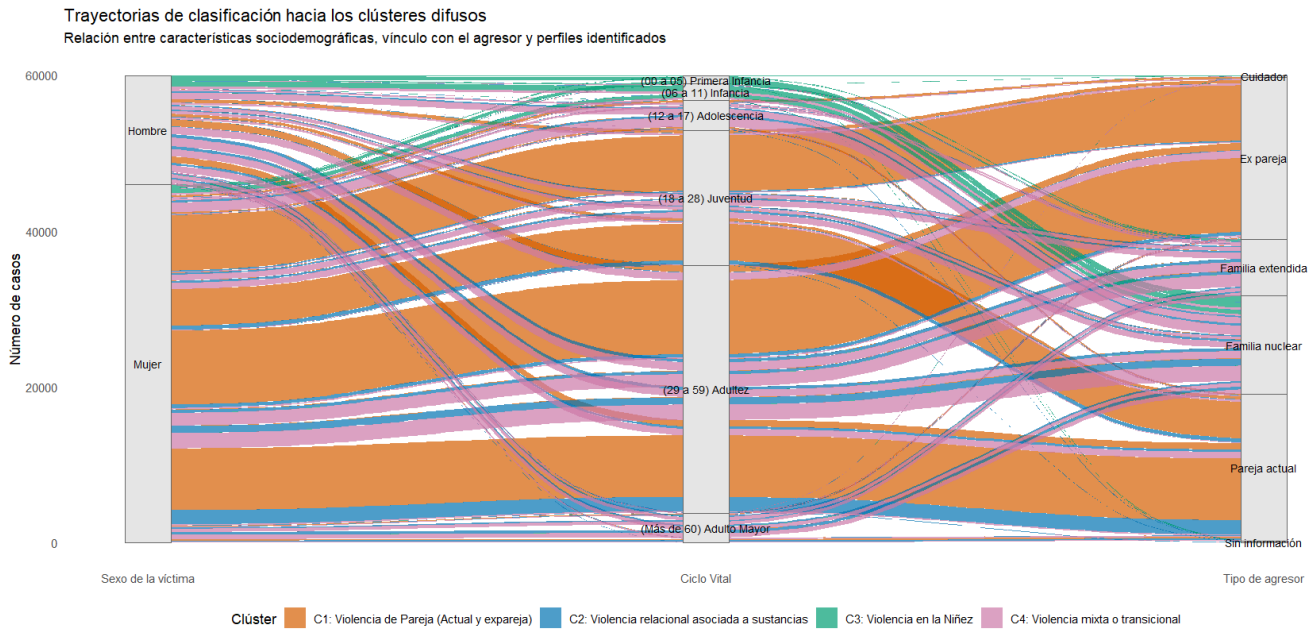


Figura E3*: Relación entre características demográficas, vínculo con el agresor y perfiles identificados.*

Anexo F. Repositorio en GitHub

El código utilizado para el procesamiento y análisis de los datos, así como para la implementación de las técnicas metodológicas empleadas en el presente trabajo, se encuentra disponible en el siguiente repositorio:

https://github.com/AndresUsecheFH/Trabajo-de-grado---Identificacion-de-factores-asociados-a-la-violencia-intrafamiliar-en-Colombia

Con el fin de facilitar la visualización conjunta del código y sus resultados, se incluyeron versiones en formato HTML de los análisis realizados en Python y en R, cuyos enlaces pueden consultarse en el archivo README del repositorio.

El análisis desarrollado en Python incluye el preprocesamiento de los datos, el análisis exploratorio, la identificación de asociaciones y su intensidad, así como la construcción y visualización de tasas de incidencia a nivel departamental y de capital:

https://andresusechefh.github.io/Trabajo-de-grado---Identificacion-de-factores-asociados-a-la-violencia-intrafamiliar-en-Colombia/Violencia_Intrafamiliar_2024.html

Por otra parte, el análisis realizado en R comprende la aplicación de la prueba exacta de Fisher para asociaciones que no cumplen los supuestos del chi-cuadrado, el análisis de Correspondencias Múltiples (ACM) y la segmentación de observaciones mediante el algoritmo Fuzzy C-Means:

https://andresusechefh.github.io/Trabajo-de-grado---Identificacion-de-factores-asociados-a-la-violencia-intrafamiliar-en-Colombia/Violencia_intrafamiliar-ACM-FUZZY.html