



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

Estudio comparativo de la tasa de criminalidad en el territorio español

Autor: Elisa Pernía Villafruela

Tutor: Carlos G. Matrán Bea

2024 - 2025

Agradecimientos.

A mis padres.

RESUMEN

Este proyecto tiene como finalidad aplicar diversas técnicas estadísticas al ámbito criminológico con el objetivo de analizar la criminalidad en el territorio español desde diferentes perspectivas. A partir de datos oficiales, se estudia la distribución y evolución de los delitos atendiendo a factores como la tipología delictiva, la edad, el sexo y la localización demográfica del infractor. Además, se incorporan varias variables socioeconómicas con el fin de explorar posibles condicionantes de la conducta delictiva. El estudio se complementa con herramientas de análisis multivariante, como el análisis de correspondencias y el análisis de componentes principales, que permiten identificar patrones comunes entre comunidades autónomas y perfiles poblacionales.

Palabras clave: Tasa de criminalidad, tipología delictiva, perfil delictivo, regresión logística, delito, clustering, análisis de correspondencias, análisis de componentes principales.

ABSTRACT

This project aims to apply various statistical techniques to the criminological field with the objective of analyzing criminality in the Spanish territory from different perspectives. Using official data, the study examines the distribution and evolution of crimes based on factors such as crime typology, age, sex and the demographic location of the offender. Additionally, several socioeconomic variables are incorporated to explore potential determinants of criminal behavior. The study is complemented by multivariate analysis tools, such as correspondence analysis and principal component analysis, which allow the identification of common patterns among autonomous communities and population profiles.

Keywords: Crime rate, crime typology, offender profile, logistic regression, crime, clustering, correspondance analysis, principal component analysis.

Índice

1. Introducción	6
1.1. Objetivos y motivación del estudio	6
2. Herramientas utilizadas	7
3. Fuentes de datos	9
3.1. Variables utilizadas	9
4. Situación de España frente al resto de Europa	11
5. Análisis descriptivo de la tipología del delito	13
5.1. Distinción por edad	18
5.2. Distinción por sexo	19
6. Factores socioeconómicos	21
6.1. Efecto del turismo en las dinámicas delictivas	27
7. Análisis multivariante	31
7.1. Análisis de correspondencias	31
7.1.1. Efecto Guttman	35
7.2. Análisis de componentes principales	35
7.2.1. Clasificación por clustering	37
8. Comparativa de criminalidad: Castilla y León y la Comunidad de Madrid	40
9. Conclusiones y trabajo futuro	45

1. Introducción

El estudio de la criminalidad es un fenómeno complejo no sólo en el ámbito de las ciencias sociales, sino también en el estadístico. Su estudio no permite únicamente comprender las posibles dinámicas latentes, sino también diseñar políticas más eficaces adaptadas al contexto de cada región con el fin de prevenir y controlar. Es por ello que la estadística se toma como una herramienta esencial para la detección de patrones delictivos o la identificación de factores asociados. En España, la amplia disponibilidad de registros detallados sobre el número de condenas y la tasa de criminalidad desglosadas por distintas características como la edad, el sexo el tipo de delito o la región del infractor permite realizar un análisis exhaustivo de la criminalidad utilizando datos numéricos y herramientas estadísticas. En el marco legal español, según la Real Academia Española (*RAE*) [2] la definición de delito es una acción u omisión voluntaria o imprudente castigada por las leyes penales. Los datos utilizados en este proyecto hacen referencia a personas condenadas en sentencia firme, clasificadas según el Código Penal español. Este trabajo por tanto se centra dentro de esa línea de análisis, con el principal objetivo de identificar algunas tendencias generales del perfil delictivo en el territorio español.

1.1. Objetivos y motivación del estudio

Se persigue con este proyecto ofrecer una aproximación general al fenómeno de la criminalidad en España desde una perspectiva estadística, para lo cual se describe y analiza el comportamiento delictivo en España desde distintas dimensiones; territorial, demográfica y socioeconómica. Para ello, se plantean los siguientes objetivos específicos:

- **Análisis descriptivo:** reconocer la distribución de los delitos según su tipología, diferenciando además entre el sexo y la edad del delincuente, identificando posibles patrones y tendencias.
- **Análisis territorial:** detectar diferencias significativas en la tasa de criminalidad entre las distintas comunidades autónomas del territorio español, considerando factores demográficos y económicos.
- **Análisis multivariante:** aplicar técnicas estadísticas como el Análisis de Correspondencias (*AC*) o el Análisis de Componentes Principales (*ACP*) para identificar relaciones entre variables, reduciendo la dimensionalidad de los datos disponibles.

2. Herramientas utilizadas

Todo el análisis estadístico se ha realizado exclusivamente con el lenguaje de programación *R*, a través del entorno de desarrollo *RStudio*. Además, para la redacción y maquetación del trabajo se ha utilizado la plataforma *Overleaf* (L^AT_EX). A lo largo del desarrollo del estudio se han empleado múltiples librerías especializadas de *R* para distintas funciones, que se describen a continuación:

1. Importación de datos

- **readxl**: se ha utilizado para la lectura de los datos en formato Excel a partir de la función `read_excel()` que permite importar las bases de datos necesarias para el análisis.

2. Manipulación y transformación

- **tidyverse**: este es un conjunto utilizado fundamentalmente para la manipulación y limpieza de los datos. Se utilizaron funciones como:
 - `left_join()` para la unión de bases de datos.
 - `filter()` para la filtración de filas según ciertas condiciones.
 - `group_by()` y `summarise()` para agrupar los datos según el interés del análisis y resumir la información.
 - `mutate()` para la creación de nuevas variables.
- **tidyr**: utilizada para reorganizar la estructura de ciertos datos mediante la función `pivot_wider()`, que transforma los datos de formato largo a ancho.

3. Visualización y representación

- **ggplot2**: utilizada para la creación de gráficos. Se usan para ello múltiples funciones como `ggplot()`, `geom_sf()`, `scale_fill_gradient()` o `labs()`, `theme_minimal()`, entre otras.
- **scales**: utilizada para ajustar escalas y etiquetas de los ejes en los gráficos mediante la función `label_comma()`.
- **ggrepel**: empleada para mejorar la visualización de etiquetas en gráficos con el fin de evitar solapamientos de estas con la función `geom_text_repel()`.
- **patchwork**: usada para la combinación de múltiples gráficos en un mismo panel, con la única finalidad de mejorar la estética.
- **Rnaturalearth** y **rnaturalearthdata**: permiten acceder a datos geográficos para la realización de mapas internacionales. Se utiliza la función `ne_countries()` para obtener los mapas vectoriales para países.
- **mapSpain**: proporciona los datos geográficos oficiales de España. Se empleó `esp_get_ccaa()` para la obtención del mapa dividido por comunidades autónomas.

4. Análisis estadístico multivariante

- **corrplot**: utilizada para representar gráficamente la matriz de correlaciones mediante la función `corrplot()`
- **FactoMineR**: utilizada para la realización de análisis exploratorios multivariantes. Se empleó la función `CA()` para aplicar el análisis de correspondencias.

- **factoextra**: utilizada para visualizar los resultados, tanto del análisis de correspondencias como del análisis de componentes principales. Se utilizan las funciones `fviz_ca_biplot()`, `fviz_contrib()`, `fviz_screepplot()` y `fviz_pca_biplot()` para la representación.
- **e1071**: utilizada para calcular la asimetría de variables.

3. Fuentes de datos

Todas las bases de datos utilizadas en el estudio han sido extraídas de la página oficial del Instituto Nacional de Estadística (INE). Esta elección además de garantizar la fiabilidad al tratarse de una fuente oficial y estandarizada a nivel nacional, permite comparar la variable de interés principal atendiendo a múltiples factores como el sexo, la comunidad autónoma, el grupo de edad o la nacionalidad del infractor, gracias a la amplia disponibilidad de bases de datos detalladas y segmentadas por distintas características poblacionales.

En concreto, la base principal a partir de la cual gira el estudio es la tasa de criminalidad; el número de infracciones penales obtenidos del Ministerio del Interior por cada 1000 habitantes. Esta base de datos ha sido desglosada de distintas maneras, en función a las necesidades y/o intereses del estudio, lo que permite realizar un análisis del comportamiento delictivo en el territorio español, tanto desde un punto de vista descriptivo como multivariante. El periodo de estudio abarca desde 2013 hasta 2023 lo que proporciona una perspectiva temporal amplia. No obstante, en determinadas variables o análisis, se ha optado por utilizar únicamente el dato de un año concreto por considerarse lo suficientemente representativo o en algunos casos se ha tenido que reducir el periodo de tiempo analizado debido a la falta de datos.

Además, con el objetivo de contextualizar y establecer comparaciones con el resto de Europa, se han utilizado también datos proporcionados por Eurostat, concretamente los relativos a la percepción ciudadana de la inseguridad en los respectivos países europeos [6] o la tasa de robos [4] por país. Esta fuente complementa el análisis nacional aportando una visión más amplia.

3.1. Variables utilizadas

El análisis de la criminalidad requiere el uso de múltiples variables que permitan estudiar este fenómeno en distintas dimensiones. A continuación se detallan las principales variables utilizadas en el estudio.

Cabe destacar que las variables de carácter socioeconómico que también forman parte del análisis se desarrollan y explican más adelante en su apartado correspondiente (apartado número 6) dado que tienen un enfoque específico y complementario al resto de variables.

Variables delictivas

- **Tasa de criminalidad:** número de infracciones penales por cada 1.000 habitantes. Es, como se mencionó anteriormente, la variable central del estudio. Ha sido analizada tanto globalmente como desglosada.
- **Tipología del delito:** según la clasificación del Código Penal se ha considerado el tipo de infracción cometida. Estos a su vez se desglosan en subtipologías, que han sido utilizadas únicamente para los más comunes.

() La Ley Orgánica 10/2022, de 6 de septiembre, de garantía integral de la libertad sexual, eliminó del Código Penal el delito de abuso sexual y la referencia a la indemnidad sexual, fusionando, bajo la denominación de agresión sexual, los anteriores delitos de abuso sexual y de agresión sexual. Sin embargo, dado que en un año se juzgan casos acaecidos tanto ese año como en anteriores (que pueden*

ser juzgados según la legislación anterior), durante unos años pueden seguir apareciendo casos de abuso sexual. [11]

Variables demográficas

- **Sexo:** variable cualitativa binaria para analizar diferencias en el comportamiento delictivo entre hombres y mujeres.
- **Edad:** grupo de edad al que pertenece el condenado. Los grupos de edad utilizados han sido: 18–20, 21–25, 26–30, 31–35, 36–40, 41–50, 51–60, 61–70 y mayores de 70.
- **Nacionalidad:** distingue entre españoles y extranjeros, diferenciando además entre el continente del delincuente extranjero.

Variables territoriales

- **Comunidad Autónoma:** variable categórica que indica el territorio del país del que provienen los datos.

4. Situación de España frente al resto de Europa

Antes de comenzar con el análisis detallado, es necesario contextualizar la situación del territorio español dentro del marco europeo. Con el fin de evaluar la criminalidad desde una perspectiva social, se recoge la percepción de inseguridad periódicamente por Eurostat a través de la Encuesta Europea de Condiciones de Vida (EU-SILC) [5]. Uno de los indicadores recogidos por esta encuesta es el porcentaje de población que afirma percibir problemas de delincuencia, violencia o vandalismo [6]. En la figura 1 se muestra por tanto la visión subjetiva de los ciudadanos europeos del nivel de seguridad percibido en 2023. Según los datos disponibles, en el año 2023, España presenta una percepción de inseguridad superior a la media europea; mientras que el promedio de Europa se sitúa por encima del 8 %, en el territorio español más del 13 % de la población considera que percibe inseguridad con respecto a la situación de la criminalidad en su entorno.

Sin embargo, esta percepción es subjetiva y no siempre se alinea con la realidad. Se muestra a continuación el mapa (figura 2) en el que se representa esta vez la tasa de robos registrados en Europa por cada 100.000 habitantes, recogida también por Eurostat [4]. Se observa que España presenta una tasa de robos por debajo de la media europea, reflejando un desajuste entre la percepción de los ciudadanos y la realidad, donde el nivel de inseguridad percibido por la población parece ser mayor que el que se observa gracias a los datos objetivos, en este caso, de robos.

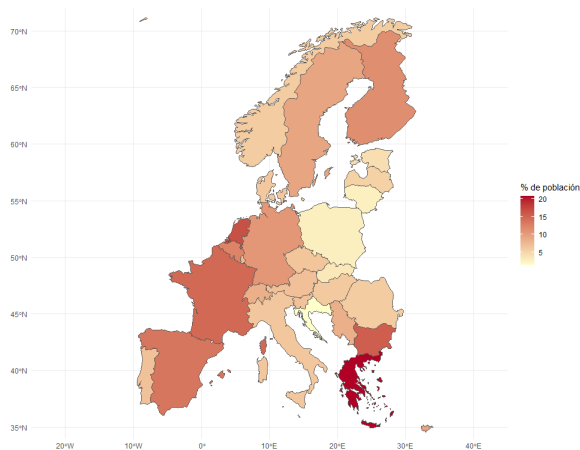


Figura 1: Porcentaje de población que percibe delincuencia, violencia o vandalismo como un problema

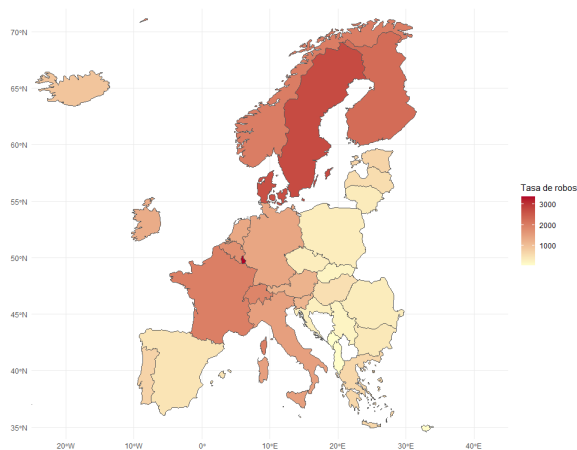


Figura 2: Tasa de robos en Europa

La elección de estos indicadores responde precisamente a este contraste: mientras que uno capta la dimensión subjetiva y emocional del fenómeno delictivo, el otro proporciona una medida concreta y cuantificable. A pesar de que no se traten de variables directamente comparables en términos técnicos, analizarlas conjuntamente permite ilustrar la brecha que puede existir entre la percepción de la ciudadanía y lo que realmente ocurre. Esto es especialmente relevante en el ámbito de las políticas públicas, ya que las percepciones influyen en la demanda social de seguridad, incluso cuando los indicadores objetivos no lo justifican. No obstante, se debe tener en cuenta que la comparación entre países puede llevar a error, pues las diferencias entre estos en cuanto a los códigos penales además de los procedimientos legales pueden dar pie a análisis sesgados.

Ambas figuras han sido elaboradas con el uso de los paquetes `rnaturalearth` y `rnaturalearthdata`, que permiten acceder a datos geográficos vectoriales para la representación de mapas internacionales. A través de la función `ne_countries()`, se obtuvieron las geometrías necesarias para construir los mapas a nivel país mientras que, a partir de los datos obtenidos de Eurostat, se realizó una limpieza eliminando los valores perdidos para más tarde enlazarlos con los datos geográficos mediante el nombre del país.

5. Análisis descriptivo de la tipología del delito

Con el fin de encontrar posibles patrones se analiza a continuación distinguiendo por la tipología de los delitos. Según los datos obtenidos del Instituto Nacional de Estadística a partir de los cuales se está realizando el estudio se categorizan los distintos delitos de la siguiente manera:

1. Homicidio y sus formas
2. Aborto
3. Lesiones
6. Contra la libertad
7. Torturas e integridad moral
8. Contra la libertad e indemnidad sexuales
9. Omisión del deber de socorro
10. Contra la intimidad, derecho a la propia imagen
11. Contra el honor
12. Contra las relaciones familiares
13. Contra el patrimonio y orden socioeconómico
14. Contra la Hacienda Pública y Seguridad Social
15. Contra los derechos de los trabajadores
16. Relativos a la ordenación del territorio
17. Contra la seguridad colectiva
18. De las falsedades
19. Contra la Administración Pública
20. Contra la Administración de Justicia
21. Contra la Constitución
22. Contra el orden público
23. Traición, contra la paz y defensa nacional
24. Contra la Comunidad Internacional

Mediante el siguiente gráfico (figura 3), realizado tras obtener el total de delitos que han sido recogidos desde 2013 hasta 2023, se puede observar como el tipo de delito número 13; 'Contra el patrimonio y orden socioeconómico' es el que mayor número acumula, seguido del tipo 17; 'Contra la seguridad colectiva'. Con menos de la mitad sigue el delito número 3; 'Lesiones'. El resto de tipologías presentan cifras notablemente inferiores. Dado el elevado volumen que concentran los dos primeros tipos mencionados, se profundizará en su análisis con el objetivo de identificar posibles patrones.

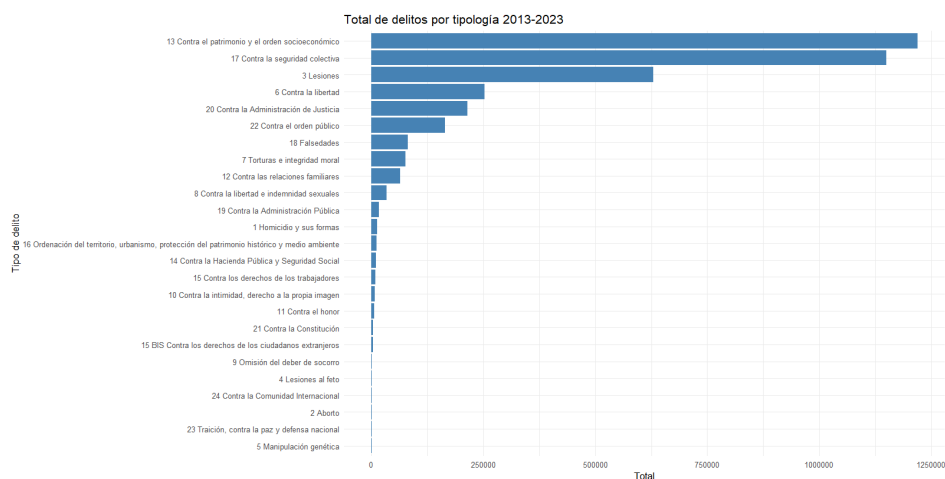


Figura 3: Distribución del total de los delitos según su tipología

El delito número 13: ‘Contra el patrimonio y orden socioeconómico’ es dividido según el INE en los siguientes subtipos:

13.1 Hurtos

13.2 Robos

13.2.1 Robo con violencia

13.2.2 Robo con fuerza

13.3 Extorsión

13.4 Robo y hurto de uso de vehículo

13.5 Usurpación

13.6 Defraudaciones

13.7 Frustración de la ejecución

13.7 BIS Insolvencias punibles

13.8 Alteración de precios en concursos y subastas públicas

13.9 Daños

13.11 Propiedad intelectual e industrial

13.12 De la sustracción de cosa propia a su utilidad social o cultural

13.13 Delitos societarios

13.14 Receptación y blanqueo de capitales

A continuación, se presentan en dos gráficos los subtipos del delito categorizado como número 13. El primer gráfico en el que se presenta el total acumulado de número de caso distinguiendo por el subtipo de delito desde 2013 hasta 2023 permite una visión grafica de la tendencia de las tipologías. Se aprecia que el hurto es con diferencia el tipo de delito más común, superando los 400.000 casos en los 10 años

representados. Le siguen los robos, con más de la mitad de casos, y defraudaciones con unos 200.000. El número de casos del resto de delitos es considerablemente menor a los mencionados anteriormente, alguno como la extorsión cuenta únicamente con unos miles de casos en los 10 años considerados en el estudio. La falta de casos de según que tipos de delitos puede deberse principalmente a dos causas: o bien ocurren puntualmente y hay pocos casos, o bien son más difíciles de detectar y por tanto más difícil de llevar a cabo las denuncias pertinentes.

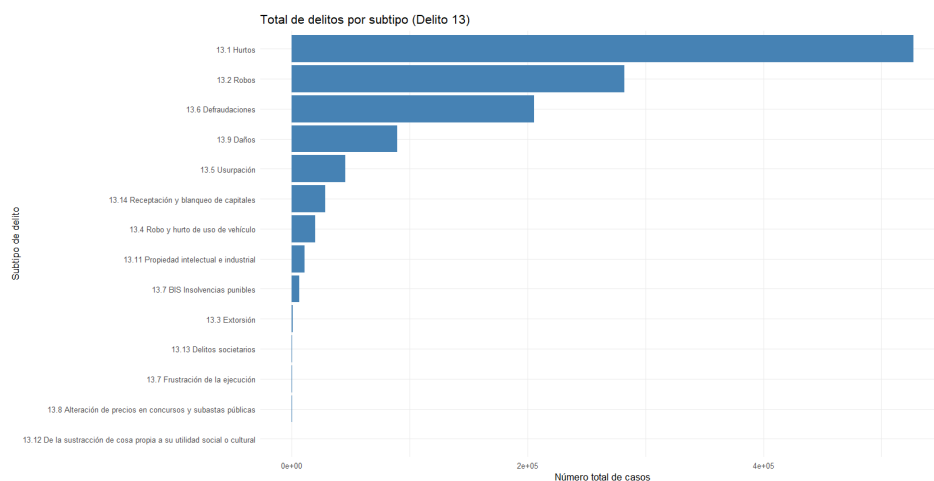


Figura 4: Distribución de los subtipos del delito 13; 'Contra el patrimonio y orden socioeconómico'

La figura 5 en la que se ve representada la evolución anual efectivamente concuerda con lo observado en el anterior; los hurtos es el subtipo de delito más frecuente en el país. Se aprecia como desde 2014 hasta 2019 hay un incremento continuo seguido de un decrecimiento notable en 2020. Esto encaja con el momento histórico de la pandemia de COVID-19, es lógico pensar que las restricciones impuestas afectasen notablemente. A partir del año 2021 se aprecia un nuevo aumento de los casos, aunque sin alcanzar los niveles pre-pandémicos. Cabe destacar que otros delitos como robos o defraudaciones cuentan con un número de casos significativamente menor a los hurtos, pero mantienen unos valores más constantes a lo largo de los años. También es notable como a principio de la década los robos presentaban una mayor frecuencia de casos que las usurpaciones y, sin embargo, a partir de 2016 esta diferencia comenzó a reducirse hasta que en los últimos años los valores de ambos tipos son muy similares, llegando incluso el número de casos de defraudaciones a superar al número de robos en el país. Esto nos lleva a pensar que puede deberse al aumento del uso de la tecnología, que supone un aumento de los fraudes digitales, así como a la mejora en las detecciones y en las denuncias de los delitos cibernéticos.

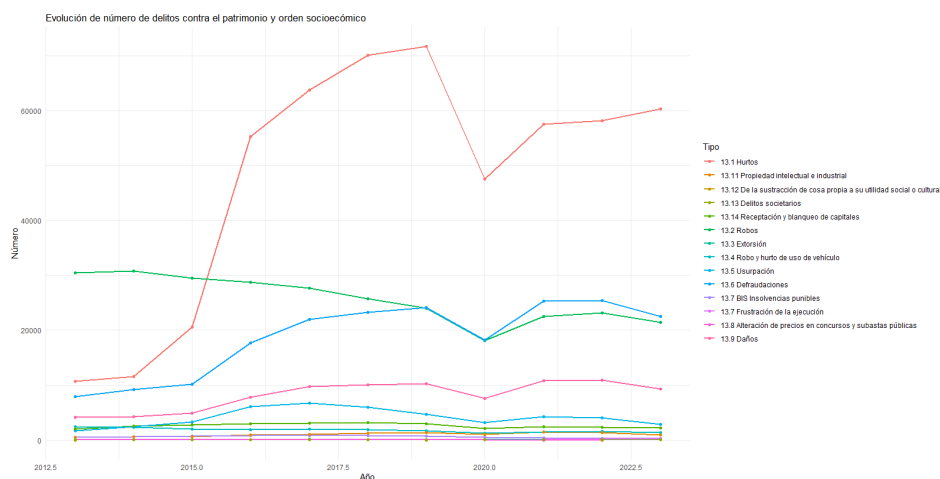


Figura 5: Evolución temporal del delito 13; 'Contra el patrimonio y orden socioeconómico'

El siguiente tipo de delito con mayor número de casos desde que se tienen datos es, como se ha visto anteriormente, el delito número 17: 'Contra la seguridad colectiva'. Este recoge los siguientes subtipos de delitos:

17.1 Delitos de riesgo catastrófico

17.2 Incendios

17.3 Contra la salud pública

17.4 Contra la seguridad vial

En la figura número 6 se muestra la distribución del total de casos ocurridos de los subdelitos del tipo 17 mencionados anteriormente. Se concluye con claridad que el subtipo 17.4: 'Contra la seguridad vial' es, sin lugar a duda, el más frecuente en el país con más de 1 millón de casos en los 10 años estudiados. Le sigue el subtipo 'Contra la salud pública', delito el cual, aunque el número de casos sigue siendo considerablemente alto, pues supera los 100.000, no deja de ocupar muy poca parte de los delitos registrados, siendo menos de una cuarta parte de los delitos contra la seguridad vial. Los otros dos subtipos, 'Incendios' y 'Delitos de riesgo catastrófico' no tienen apenas incidencia comparado con los mencionados anteriormente, lo cual es previsible, pues suelen ser hechos que ocurren de manera puntual.

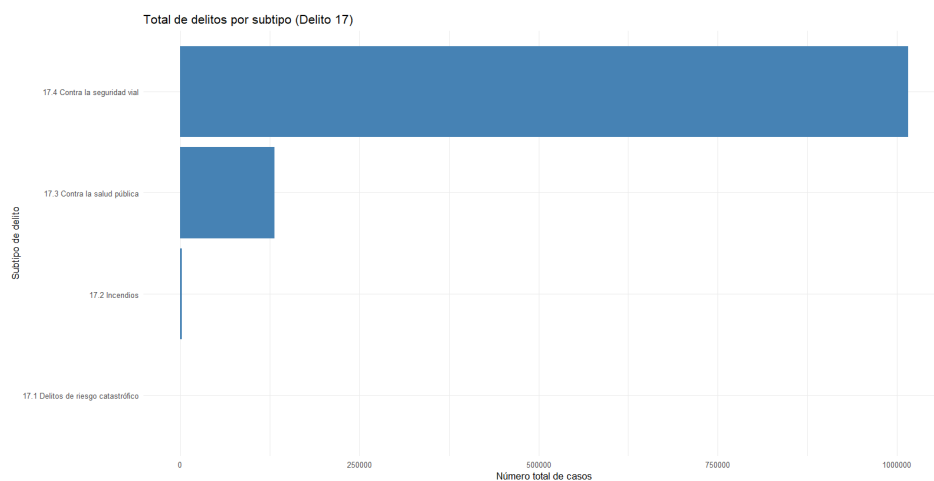


Figura 6: Distribución de los subtipos del delito 17; 'Contra la seguridad colectiva'

En cuanto a la figura número 7 en la que se muestra la evolución temporal se observa como el delito contra la seguridad vial ha mantenido una tendencia estable y ligeramente decreciente durante los primeros años del estudio, con la esperable caída en 2020 y su posterior aumento de casos una vez las restricciones de la pandemia terminaron. En el caso de los delitos contra la salud pública, se aprecia una evolución temporal estable, con un pequeño descenso, de nuevo, en el año del COVID-19. Por otro lado, los incendios y los delitos de riesgo catastrófico mantienen un volumen prácticamente constante a lo largo de todo el periodo, con cifras muy reducidas. En definitiva, llegamos a la conclusión de que los delitos contra la seguridad vial además de ser los más frecuentes del tipo de delito número 17 ('Contra la seguridad colectiva') representan una enorme parte del total de delitos analizados.

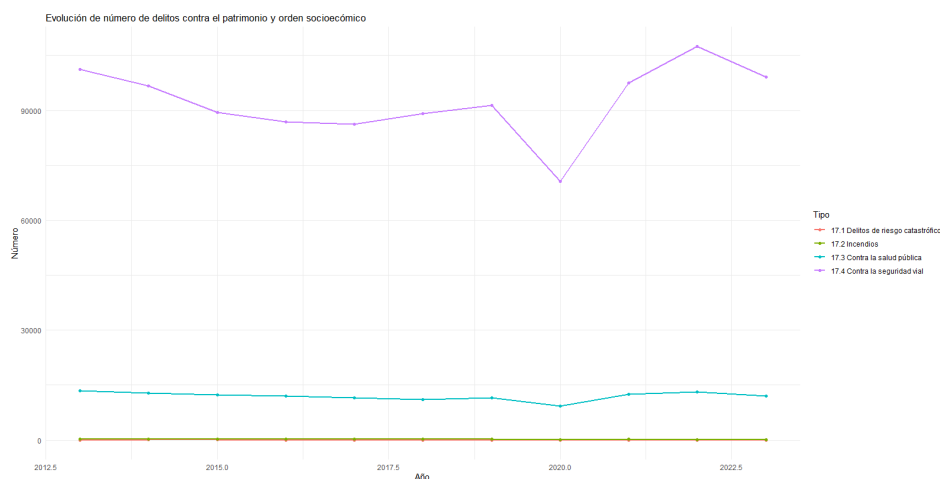


Figura 7: Evolución temporal del 17; 'Contra la seguridad colectiva'

5.1. Distinción por edad

Con el fin de detectar posibles patrones que definan el comportamiento del infractor, se va a estudiar a continuación la frecuencia de cada uno de los tipos de delitos según la edad del delincuente. Para ello se llevan a cabo a continuación una serie de gráficas descriptivas. La figura número 8 representa la proporción del total de delitos cometidos, diferenciando por grupo de edad y tipología delictiva (del nivel 2 según lo clasifica el INE). Esto nos permite observar que tipología delictiva es más frecuente en cada uno de los grupos de edad estudiados. Cabe destacar que se han representado únicamente los delitos más representativos tanto a nivel de volumen de casos como a nivel de relevancia.

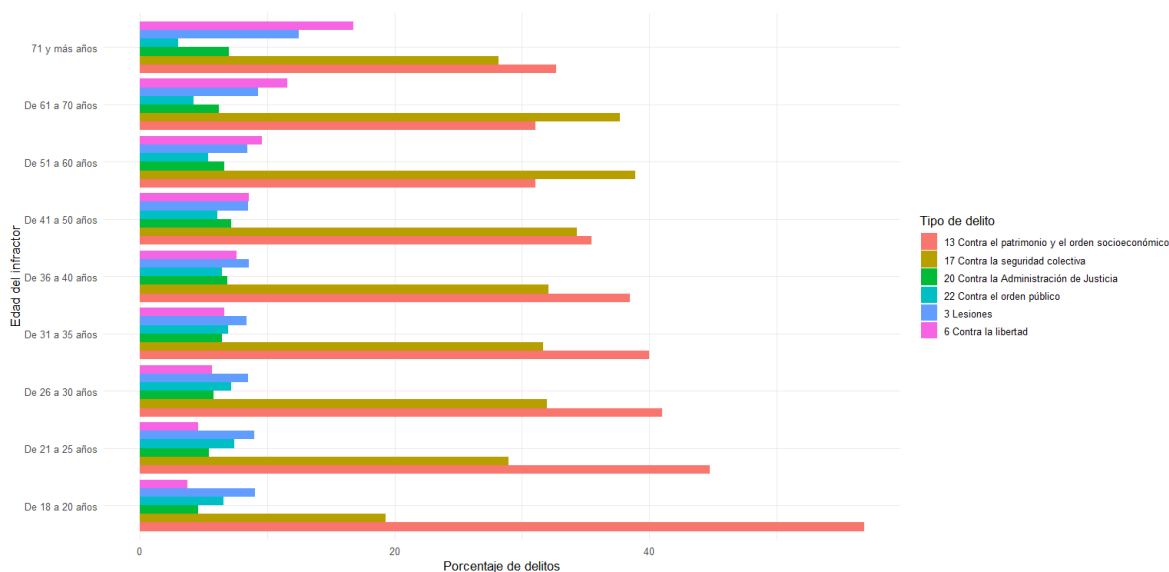


Figura 8: Porcentaje de delitos por tipo y edad

A primera vista destaca que el delito 13; ‘Contra el patrimonio y el orden socioeconómico’ es el más frecuente en el grupo de edad de 18 a 20 años con diferencia, pues representa más del 50 % de todos los delitos cometidos en el rango de edad de la juventud. Este sigue siendo el más frecuente en el resto de los grupos de edad hasta los 40 años y vuelve a ser el más común a partir de los 71, en la etapa de la vejez. Se observa como el delito 17; ‘Contra la seguridad colectiva’ es el más frecuente en el rango de edad que va desde los 50 hasta los 70 años. Recordamos que dentro de este tipo de delito se incluyen los delitos contra la seguridad vial y, aunque de primeras podríamos pensar que los conductores nóveles son los más temerarios al volante, esto tiene coherencia ya que los conductores de edad adulta pueden cometer infracciones derivadas del exceso de confianza, además de que generalmente tienen más tendencia coger el coche y por tanto están más expuestos.

En cuanto al grupo de la vejez, cabe destacar que hay una gran incidencia del tipo de delito número 6; ‘Contra la libertad’ en comparación con el resto de los rangos de edad, pues, mientras que el delito ‘Contra el patrimonio y el orden socioeconómico’ disminuye con la edad del infractor, el delito contra la libertad aumenta con esta. Este incluye los siguientes subdelitos:

6.1 Detenciones ilegales y secuestro

6.2 Amenazas

6.3 Coacciones

5.2. Distinción por sexo

A continuación, se va a profundizar desde otro enfoque, esta vez teniendo en cuenta el sexo del infractor. Para comenzar, en la figura número 9 se representa el porcentaje de delitos cometidos distinguiendo por el sexo del infractor considerando los datos desde el año 2013 hasta el 2023. Es evidente la diferencia abismal entre la frecuencia de delincuentes hombres contra el número de delincuentes de sexo femenino. Por ello vamos a analizar más a fondo, viendo por ejemplo que tipos de delitos son más comunes en cada uno de los sexos.

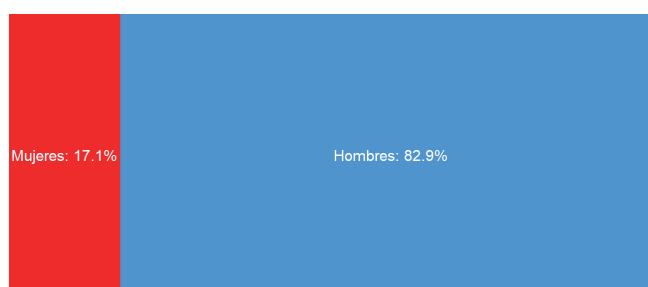


Figura 9: Distribución de delitos por sexo

Para empezar, en la figura número 10 se representa la distribución de los tipos de delitos distinguiendo entre si el infractor era hombre o mujer. Evidentemente se confirma con claridad la desigualdad comentada anteriormente entre sexos; se afirma que los hombres son responsables de la mayor parte de los delitos cometidos en el país. No hay ningún tipo predominado por el sexo femenino, todas las tipologías son más frecuentes en los hombres. En el delito número 8 ‘contra la libertad e indemnidad sexuales’ la participación de las mujeres es prácticamente anecdótica, pues únicamente han sucedido 919 casos en los 10 años que se están estudiando frente a los 32.388 llevados a cabo por el sexo masculino, lo que se traduce en un 97.24 % frente a un 2.76 % (cuadro 1).

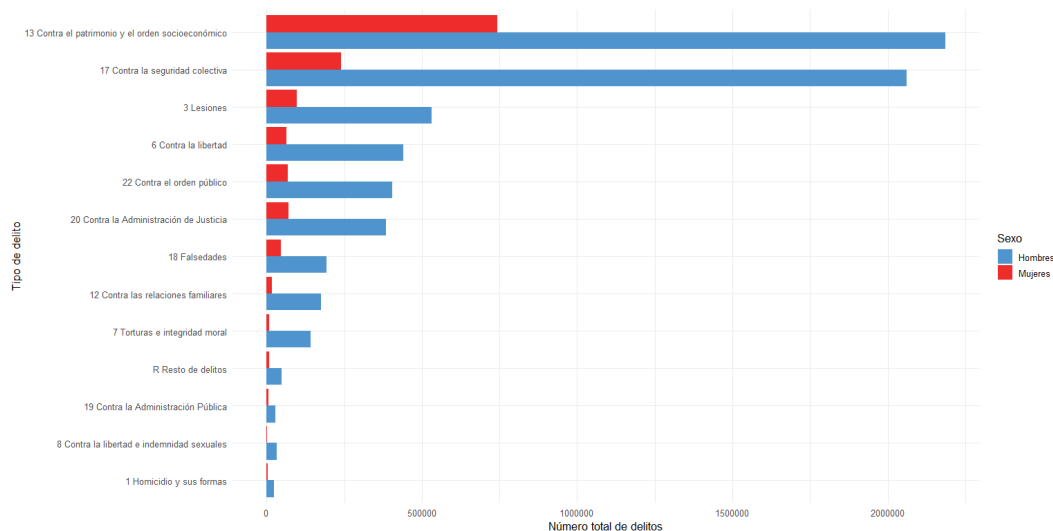


Figura 10: Frecuencia de delitos por sexo del infractor

Sexo	Frecuencia	Porcentaje
Masculino	32.388	97.2
Femenino	919	2.8

Cuadro 1: Distribución de infractores por sexo

Por ello vamos a profundizar más a fondo en este tipo de delito de violencia sexual. Según el INE los delitos en contra de la libertad e indemnidad sexuales se dividen según los siguientes tipos:

8.1 Agresiones sexuales

8.2 Abusos sexuales

8.2 BIS Abusos y agresiones sexuales a menores de 16 años

8.3 Acoso sexual

8.4 Exhibicionismo y provocación sexual

8.5 Prostitución y corrupción de menores

El Instituto Nacional de Estadística sólo proporciona los datos divididos en los anteriores subtipos para los condenados menores de edad, por lo que se ha de tener en cuenta que se estudia ahora únicamente esta parte de la población desde 2017 hasta el año 2023, ya que es toda la disponible.

Delitos sexuales	Hombres	Mujeres	Total
8.1 Agresiones sexuales	864	8	872
8.2 Abusos sexuales	901	11	912
8.2 BIS Abusos y agresiones sexuales a menores de 16 años	1661	19	1680
8.3 Acoso sexual	18	0	18
8.4 Exhibicionismo y provocación sexual	133	2	135
8.5 Prostitución y corrupción menores	294	35	329
Total	3871	75	3946

Cuadro 2: Distribución de delitos sexuales por sexo del infractor

En cuanto al sexo del delincuente, se constata una amplia mayoría de hombres (98.1 %) frente a las mujeres (1.9 %), superando la diferencia de porcentajes que se observó al contar a toda la población en lugar de únicamente a los menores de edad.

El sexo femenino aparece principalmente en el subtipo número 5, ‘Prostitución y corrupción de menores’ aunque con tan sólo 35 casos en los 7 años que se tienen en cuenta. Destacar también que no hubo ningún caso de acoso sexual por parte del sexo femenino.

En cuanto al tipo de delito sexual más común entre los menores de edad, se encuentra el delito de abuso y agresión sexual a menores de 16 años con un total de 1680 casos (más del 42 % de los casos), lo cual es coherente ya que dentro del ámbito de la delincuencia juvenil es habitual que las víctimas pertenezcan a grupos de edad similares y, por tanto, no es extraño que estas sean también menores de edad. Por tanto, tiene sentido que los delitos sexuales cometidos por menores se concentren específicamente en esta tipología, donde la víctima es menor de 16 años.

En definitiva, se pone de manifiesto una brecha de género en la criminalidad que no se expresa únicamente en el tipo de delitos, sino también el tipo de conducta delictiva.

6. Factores socioeconómicos

Con la finalidad de observar qué factores socioeconómicos pueden estar más relacionados con el incremento de la variable de estudio se va a estudiar las siguientes variables con mayor profundidad:

- **Coefficiente de Gini:** proporción acumulada de la población ordenada por los ingresos equivalentes con la proporción acumulada de los ingresos recibidos por los mismos. Es una medida de desigualdad que toma el valor 0 en caso de equidad perfecta y el valor 100 en caso de desigualdad perfecta.
- **Nivel de educación:** porcentaje de la población de más de 16 años por nivel de formación alcanzado, sexo y comunidad autónoma.
- **Riesgo de pobreza:** porcentaje de personas cuyos ingresos están por debajo del umbral de pobreza, definido como el 60 % de la mediana de los ingresos equivalentes por unidad de consumo.
- **Tasa de paro:** porcentaje de personas activas que se encuentran en situación de desempleo. La población activa incluye tanto a las personas con trabajo como a las que están buscando empleo. Esta tasa se estima a partir de la *Encuesta de Población Activa (EPA)*¹.
- **Renta media:** ingreso medio neto por persona, calculado a partir de la *Encuesta de Condiciones de Vida (ECV)*.
- **Número de turistas:** número total de turistas recibidos por comunidad autónoma y su tasa correspondiente.

Con la finalidad de estudiar la posible relación entre estos factores y la variable de estudio, se ha realizado el siguiente gráfico de correlación. En él se muestra la intensidad de la asociación entre la tasa de criminalidad y cada una de las cinco variables seleccionadas: coeficiente de Gini, nivel educativo, riesgo de pobreza, tasa de paro, renta media y número de turistas acogidos. Dado que el interés principal se centra en las relaciones entre estas variables y la tasa de criminalidad, se representa únicamente la parte superior de la matriz de correlaciones, permitiendo visualizar con mayor claridad dichas asociaciones. Para la realización del gráfico se ha calculado la media de las respectivas variables de interés. Cabe destacar que algunas variables presentan valores ausentes para determinadas comunidades autónomas, en especial Ceuta y Melilla, y, por este motivo, dichas observaciones han sido excluidas del análisis.

¹Los datos de 2006 a 2020 están calculados según la base poblacional 2011. Desde 2021 se utiliza la base 2021.

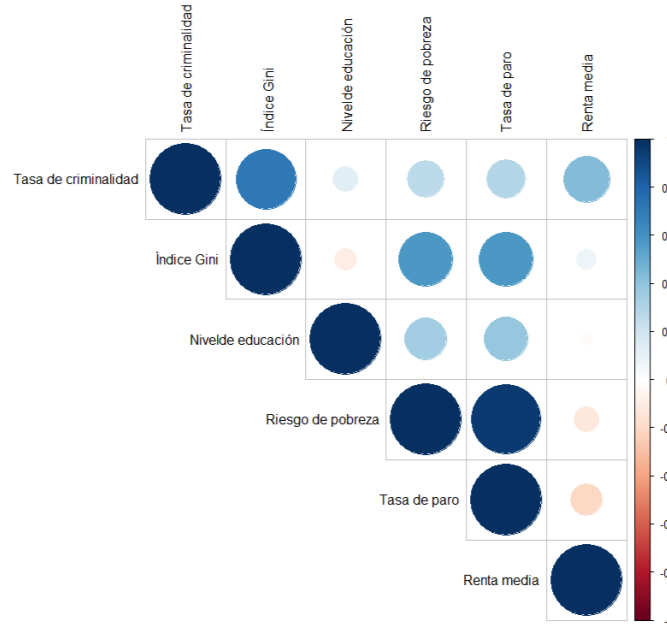


Figura 11: Matriz de correlaciones

En la figura número 11 se aprecia que la tasa de criminalidad presenta una relación especialmente notable con una de las variables analizadas: coeficiente de Gini, seguida de la renta media con una correlación mucho menor. Esta aparente asociación sugiere que la desigualdad en la distribución de la renta podría desempeñar un papel relevante en la explicación de los niveles de criminalidad en el país. A continuación, se muestran los valores de las correlaciones obtenidas, que permiten cuantificar con mayor precisión las relaciones mencionadas:

	Índice Gini	Nivel educación	Riesgo de pobreza	Tasa de paro	Renta media
Tasa de criminalidad	0.717	0.123	0.269	0.293	0.431

Cuadro 3: Coeficientes de correlación

Se confirma lo observado gráficamente anteriormente; las variables que parecen tener más relación con la tasa de criminalidad son el índice de Gini, con una correlación de 0.717 seguido de la renta media con un valor de 0.431. Pese a que este último es un valor bastante bajo, se estudiará más adelante para asegurarnos de la posible repercusión de la renta media por hogar en el índice de criminalidad.

El índice de Gini mide el grado de desigualdad en la distribución del ingreso dentro de una población. Un valor cercano a 0 indica una distribución equitativa, mientras que los valores más cercanos a 100 reflejan una concentración extrema de la riqueza en pocas manos. Este vínculo con la criminalidad no resulta sorprendente, ya que los altos niveles de desigualdad económica pueden generar tensiones sociales, faltas de oportunidades y contextos propicios para la realización de delitos.

Por ello, se lleva a cabo a continuación la modelización de un modelo lineal simple con el objetivo de analizar en qué medida la tasa de criminalidad puede explicarse a partir del índice de Gini, como indicador de desigualdad en la distribución de la renta.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Y_i : valor de la tasa de criminalidad para la comunidad autónoma i

β_0 : intercept del modelo

β_1 : coeficiente que mide el efecto de la variable explicativa

X_i : valor del índice de Gini para la comunidad autónoma i

ε_i : término de error aleatorio

El resultado del ajuste mediante regresión lineal simple se resume a continuación. Los coeficientes se han obtenido a partir del resumen estadístico del modelo `summary()`.

Variable	Coeficiente	Error estándar	t-valor	p-valor
Intercept	-35.3477	21.2567	-1.663	0.11708
Índice Gini	2.3098	0.6327	3.651	0.00237

Cuadro 4: Estimación de los coeficientes del modelo lineal simple

$R^2 = 0.4705$, R^2 ajustado=0.4352, error estándar residual=9.589, $F(1,15)=13.33$, $pvalor=0.002365$

A partir de la modelización se confirman las sospechas planteadas tras el análisis de correlación, ya que el coeficiente de regresión asociado al índice de Gini es positivo y estadísticamente significativo ($pvalor=0.00237$). Según el modelo estimado, un incremento de una unidad en el índice de Gini se asocia, en promedio, con un aumento de 2.31 puntos en la tasa de criminalidad.

No obstante, esta estimación debe interpretarse con cautela, ya que el R^2 ajustado del modelo es de 0.435, lo que indica que el índice de Gini explica aproximadamente el 43.5% de la variabilidad en la tasa de criminalidad entre regiones. Por tanto, aunque se detecta una relación significativa, es importante tener en cuenta que existen otros factores relevantes no incluidos en el modelo que también influyen en la criminalidad.

Ahora que sabemos que existe una relación entre ambas variables, vamos a estudiar más a fondo la variable índice de Gini. Para ello se muestra a continuación la tabla resumen (cuadro 5) seguida del gráfico de dispersión entre las medias del índice mencionado anteriormente y la de la tasa de criminalidad por comunidad autónoma. Cabe destacar que no aparecen dos de las comunidades autónomas; Comunidad Foral de Navarra y País Vasco debido a la ausencia de datos del índice de Gini de estas.

Comunidad Autónoma	Tasa media de criminalidad	Índice de Gini medio
Andalucía	42.49	33.68
Aragón	32.46	30.35
Asturias, Principado de	27.24	30.71
Balears, Illes	62.41	32.35
Canarias	43.45	34.13
Cantabria	31.66	30.49
Castilla - La Mancha	34.29	31.89
Castilla y León	31.95	30.64
Cataluña	57.96	32.83
Ceuta	58.61	42.35
Comunitat Valenciana	49.56	33.50
Extremadura	26.53	31.65
Galicia	29.84	30.39
Madrid, Comunidad de	58.04	36.25
Melilla	56.24	42.74
Murcia, Región de	38.46	32.53
Rioja, La	29.27	31.30

Cuadro 5: Tasa media de criminalidad e índice de Gini medio por comunidad autónoma

En el gráfico (figura 12) se aprecia la relación entre el índice y la tasa de criminalidad en cada una de las comunidades autónomas del territorio español. Vemos que hay asociación directa entre ambas variables, existe una relación positiva entre ambas, lo que indica que a mayor desigualdad mayor tasa de criminalidad. Comunidades autónomas con un índice más alto como la Comunidad de Madrid, Ceuta y Melilla tienden a tener una mayor de tasa de criminalidad, mientras que aquellas con menor desigualdad como Aragón, Cantabria, Castilla y León y Galicia presentan una tasa de criminalidad media mucho menor. En cuanto a Extremadura, Principado de Asturias y la Rioja, las tres regiones tienen una criminalidad inferior a la esperada.

Es importante además tener en cuenta que los puntos que representan a la Comunidad de Madrid, Islas Baleares y Cataluña se encuentran fuera de la banda que representa el intervalo de confianza, en este caso del 95 %, para la predicción del modelo, lo que indica que presentan tasas de criminalidad superiores a las que serían estadísticamente esperables. Esto sugiere que pueden existir factores adicionales que pueden explicar el comportamiento delictivo de los individuos de las regiones y que no hemos recogido aún, en estos casos parece que la tasa no se explica únicamente por la desigualdad económica medida con el índice de Gini, por lo que convendría más adelante estudiar más a fondo ambas comunidades.

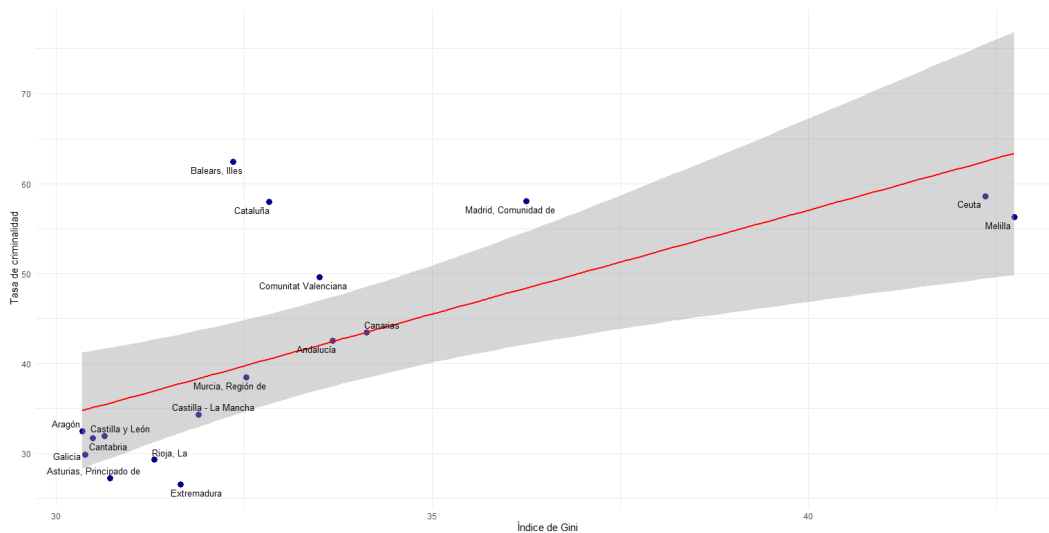


Figura 12: Gráfico de dispersión entre la tasa media de criminalidad y el índice de Gini

Pese a que la correlación obtenida en el anterior análisis entre la criminalidad y la variable que representa la renta no es muy alta, vamos a profundizar en esta con el fin de buscar una posible relación entre las tres comunidades que mencionamos anteriormente de las que obtenemos una tasa estadísticamente superior a la esperada; Comunidad de Madrid, Islas Baleares y Cataluña. Con la intención de buscar posibles patrones en regiones del país con tasas de criminalidad similares, se lleva a continuación un análisis descriptivo de la tasa de interés en las distintas comunidades.

A continuación, se analiza más en detalle la tasa de criminalidad global en el país, distinguiendo por comunidades autónomas. Se presenta, en la figura número 13, un mapa que representa la tasa de criminalidad en 2021. Para ello se ha utilizado el software *R*, usando la librería *mapSpain*. Pese a que la opción de los nombres de los países ha sido la llamada `ine.ccaa.name`, pues los datos se han conseguido a través del INE, cabe destacar que ha sido necesario la realización de cambios en los nombres de las comunidades del territorio, pues algunas no coincidían.

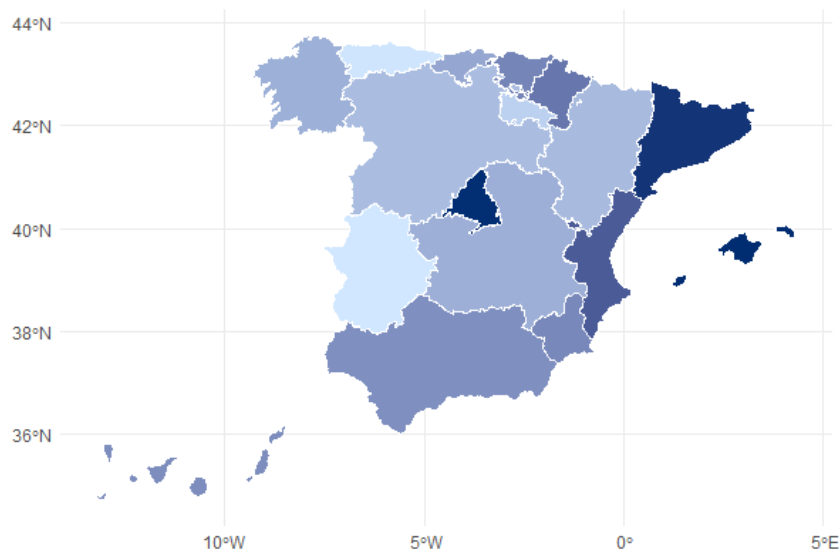


Figura 13: Tasa de criminalidad en España en 2021

Se observa que, en las comunidades autónomas con mayor densidad de población: Andalucía, Cataluña, Comunidad de Madrid y Comunidad Valenciana la tasa de criminalidad es notablemente mayor al resto de comunidades. También se aprecia que las islas: Islas Canarias e Islas Baleares tiene una tasa alta, teniendo en cuenta que la población es mucho menor que en las comunidades mencionadas anteriormente. Esto nos lleva a pensar que, pese a que la población fija es menor, la elevada tasa puede deberse al alto grado de turismo que afrontan estas comunidades en comparación con el resto de las comunidades del país (las islas acogieron más de 11 millones de turistas en 2021, representando esto más del 35 % del total de visitantes recibidos en el territorio español).

A continuación, se muestra el histograma en el que se representa la tasa de criminalidad en cada una de las comunidades autónomas, con la única intención de confirmar lo que se plasmó en la figura anterior.

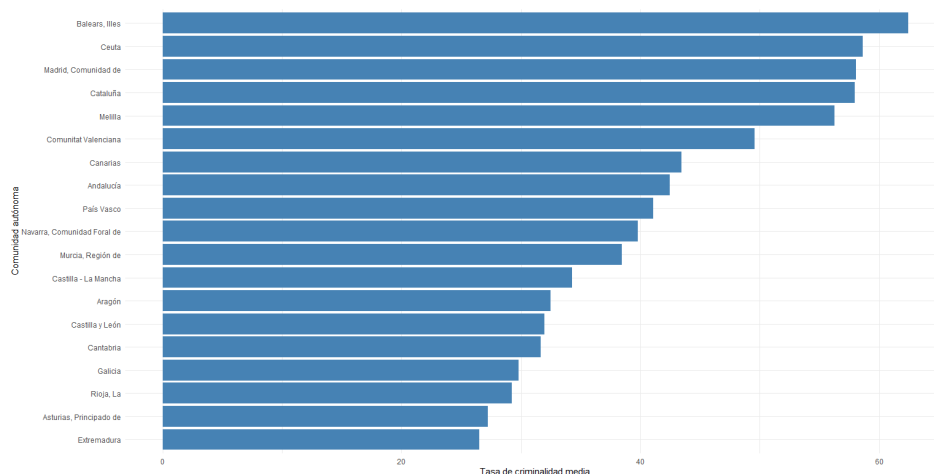


Figura 14: Tasa de criminalidad media por comunidad autónoma

Con el fin de encontrar un posible patrón que explique el por qué, pese a que exista una relación entre la desigualdad de la región y su tasa de crímenes las regiones con mayor tasa de crímenes cuentan con una baja tasa de desigualdad, se estudia a continuación la variable que representa la renta media en el país. A continuación, se muestra el mapa en el que se representa la renta media por comunidades autónomas:

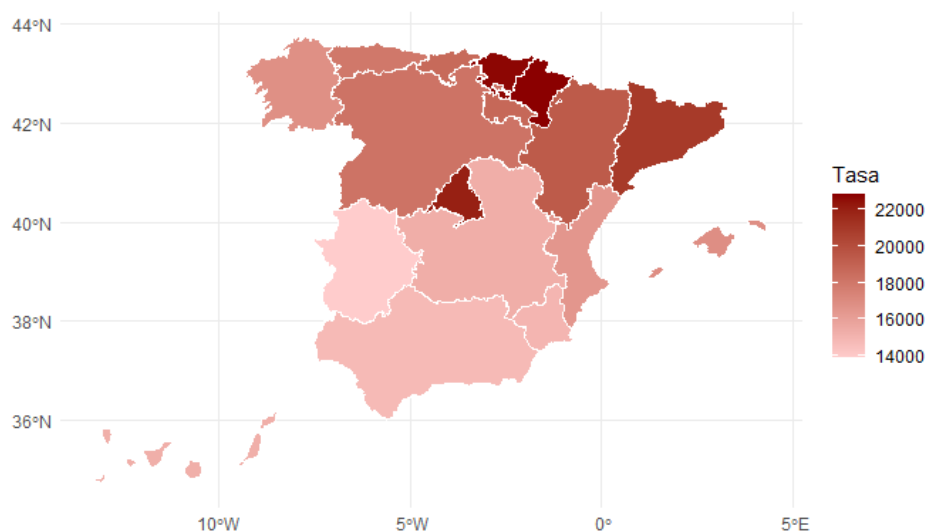


Figura 15: Renta media en España en 2021

Las comunidades autónomas con mayor renta media del territorio español son País Vasco, Comunidad Foral de Navarra, Comunidad de Madrid y Cataluña. Aunque, como se comentó anteriormente, no contamos con los datos de la desigualdad en la distribución de ingresos de País Vasco y la Comunidad Foral de Navarra si podemos analizar lo que sucede con la Comunidad de Madrid y Cataluña. Estas dos últimas además de ser de las regiones con mayor renta media, coincide que son las que mayor número de crímenes por habitantes tienen. Sin embargo, en cuanto a las Islas Baleares, no se cumple esto; es la región con mayor tasa pese a que la renta media no es de las más altas del país. Una posible explicación a este comportamiento podría residir en la influencia del turismo: posiblemente las dinámicas delictivas de esta región sean distintas a la del resto de las comunidades, ya sea por el perfil del delincuente o el tipo de delito, ambas se pueden ver influenciadas por la afluencia masiva de visitantes temporales. También hay que tener en cuenta que estas tres regiones registran una elevada población flotante, no sólo turistas si no también estudiantes o trabajadores temporales.

6.1. Efecto del turismo en las dinámicas delictivas

En muchas regiones del país, el turismo es una importante actividad económica, especialmente en aquellos destinos de costa. Las Islas Baleares y Cataluña son las dos comunidades autónomas que más turistas han recogido desde 2016 hasta la actualidad, como se observa en el siguiente cuadro:

Comunidad Autónoma	Total de turistas
Cataluña	138.462.062
Balears, Illes	105.301.768
Canarias	104.925.171
Andalucía	88.534.963
Comunitat Valenciana	72.968.239
Madrid, Comunidad de	53.878.129
País Vasco	12.952.409
Galicia	11.497.641
Castilla y León	9.848.685
Murcia, Región de	7.693.860
Aragón	3.997.890
Cantabria	3.058.558
Extremadura	2.794.833
Navarra, Comunidad Foral de	2.792.788
Asturias, Principado de	2.427.666
Castilla - La Mancha	1.792.789
Rioja, La	997.900
Ceuta	150.495
Melilla	88.630

Cuadro 6: Total acumulado de turistas por comunidad autónoma

Con el fin de estudiar la posible influencia del turismo en el país en las actividades delictivas de cada región se lleva a cabo a continuación un análisis de correlación entre la tasa de turistas y la tasa de criminalidad. Puesto que los censos de población se realizan cada 10 años en este estudio no se dispone de la población en cada región a partir de 2020, por lo que hay que tener en cuenta que se han utilizado las tasas de turismo y criminalidad únicamente con los datos desde 2016 hasta 2020. Cabe también destacar que tanto Ceuta como Melilla no se han tenido en cuenta en el estudio debido a la falta de datos.

Para calcular la tasa de turismo, se ha utilizado el cociente entre el número total de turistas recibidos y la población residente en cada comunidad autónoma. Posteriormente se ha evaluado la relación entre dicha tasa y la tasa de criminalidad mediante un análisis de correlación de Pearson, a través de la función `cor.test()`. Esta prueba permite cuantificar la intensidad y dirección de la asociación lineal entre ambas variables. A partir del análisis se obtienen los siguientes resultados:

Estadístico	Valor
Coefficiente de correlación (r)	0.636
Grados de libertad (df)	83
Pvalor	$6,409 \times 10^{-11}$
Intervalo de confianza (95 %)	[0.489, 0.747]

Cuadro 7: Resultados del test de correlación de Pearson

Se obtiene un valor positivo: 0.636 , lo que indica que la correlación entre la tasa de turistas y la de criminalidad es moderada y positiva; en un mismo espacio a medida que aumentan los turistas parece aumentar también el número de crímenes. Aunque este valor no implica una relación causal, el *pvalor* obtenido es muy inferior a 0.01 (prácticamente 0) lo que sugiere que la asociación observada entre ambas tasas es estadísticamente significativa según el test de correlación de Pearson. Este resultado, por tanto, es consistente con la hipótesis de que el aumento de turistas podría estar relacionado con un

incremento del número de delitos. Se va analizar a continuación qué sucede en las Islas Baleares, pues, aunque no se sitúa entre las comunidades con mayor renta media o con el índice de desigualdad más alto, presenta la tasa de criminalidad más alta del país. Este comportamiento irregular podría deberse a que es la segunda región con mayor número de turistas al año, lo que puede favorecer al aumento de una serie de delitos como hurtos o robos.

A continuación, se muestra un diagrama de barras en el que se representan los tipos de delitos más frecuentes en las Islas Baleares. En este se observa que los delitos más comunes son con diferencia aquellos contra la seguridad colectiva y aquellos contra el patrimonio y orden socioeconómico.

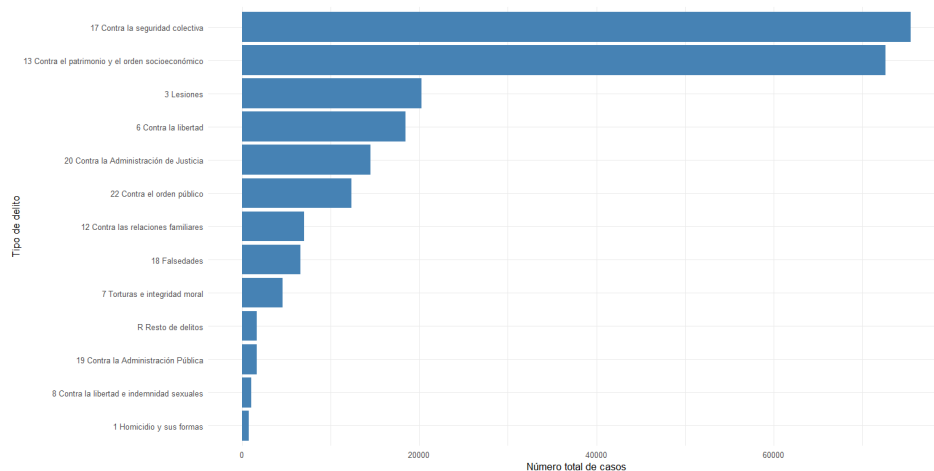


Figura 16: Número total según el tipo de delito en Islas Baleares

Sin embargo, es cierto que estos resultados no cambian mucho respecto al resto de comunidades visto en el previo análisis descriptivo por tipología (véase el punto 5), por lo que a continuación se muestran unos gráficos en los que se representa la nacionalidad del infractor para buscar otra posible relación. Se ha utilizado para ello únicamente los datos recogidos en el año 2019.

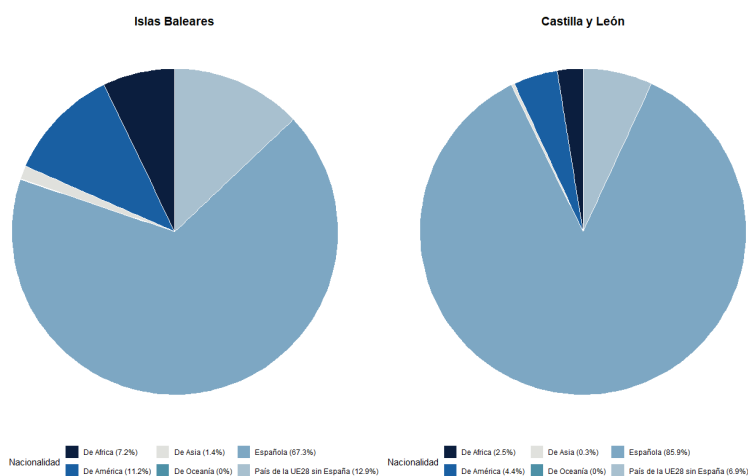


Figura 17: Distribución de los delitos por nacionalidad del infractor

Se compara con otra comunidad autónoma con características distintas a las islas, se ha elegido para este análisis Castilla y León ya que la intención principal es analizar cómo pueden haber influido diferentes factores sociales a los actos delictivos de la región y, mientras que Islas Baleares presenta una de las tasas de criminalidad más elevadas del país a la vez que una alta tasa de turistas, Castilla y León se caracteriza por una baja densidad poblacional y se sitúa significativamente por debajo de esta en ambas tasas. Esta comparación permite contrastar por tanto dos contextos socioeconómicos y demográficos opuestos. En la figura número 17 se muestra la distribución de la nacionalidad de las personas condenadas por delitos en las dos comunidades elegidas. Se observa una diferencia notable en la proporción de infractores de nacionalidad extranjera, lo que podría reforzar la hipótesis de que existe un efecto del turismo en el número de crímenes cometidos. En las Islas Baleares sólo el 67 % de infractores tienen nacionalidad española, mientras que el restante corresponde a extranjeros destacando tanto el resto de los países de la Unión Europea como el continente americano. En cambio, en Castilla y León el porcentaje de delincuentes de nacionalidad española es muy superior; alcanza casi el 86 %.

En definitiva, en la figura número 18 se representa distinguiendo únicamente entre la proporción de criminales españoles y no españoles con la idea de reforzar visualmente lo observado en la figura anterior.

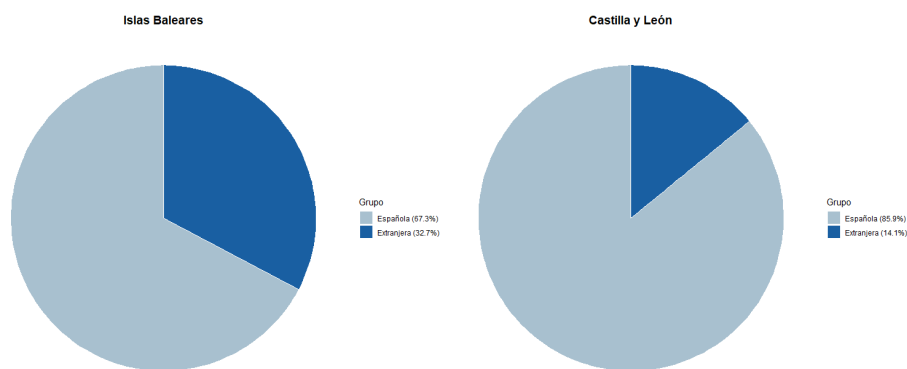


Figura 18: Distribución de los delitos según nacionalidad del infractor

7. Análisis multivariante

7.1. Análisis de correspondencias

El análisis de correspondencias (*AC*) es un procedimiento multivariante de interdependencia; no hay variable respuesta. Se considera para ello variables categóricas con el fin de estudiar posibles relaciones entre las variables de estudio y sus categorías. El objetivo principal es presentar, en un número reducido de dimensiones, las similitudes y disparidades existentes entre las categorías de las variables, en este caso los grupos de edad y las comunidades autónomas del país, con el fin de estudiar posibles asociaciones entre estas. Como únicamente van a ser estudiadas dos variables, se muestra a continuación los resultados tras realizar el Análisis de Correspondencias Simples para estudiar la posible relación entre comunidades y la edad de los condenados en el año 2023. Antes de efectuar el análisis se ha explorado la distribución de los datos, ya que se observa que categorías como Andalucía concentran una gran cantidad de datos. Se evaluó la posible existencia de asimetría en estos y aunque la asimetría positiva está presente en varios perfiles fila y columna no se va a efectuar ninguna transformación, pues el análisis de correspondencias trabaja a partir de perfiles relativos. No obstante, se tendrá en cuenta a la hora de analizar los resultados obtenidos, ya que es esperable que las categorías con frecuencias más elevadas tengan mayor peso en la representación gráfica. Se muestra a continuación la tabla de contingencias de los datos mencionados:

Comunidad autónoma	18-20	21-25	26-30	31-35	36-40	41-50	51-60	61-70	71 y más	Total
01 Andalucía	5075	9281	10250	11158	11078	17175	8672	2766	821	76276
02 Aragón	998	1341	1350	1274	1421	2191	1081	375	117	10148
03 Asturias	435	746	820	897	1013	1802	1005	387	184	7289
04 Illes Balears	1001	1711	1748	2093	1999	3006	1337	414	125	13434
05 Canarias	1445	2372	2617	3079	3090	5035	3133	993	273	22037
06 Cantabria	426	614	593	663	745	1230	666	265	101	5303
07 Castilla y León	1087	1937	1924	2023	1988	3684	2032	732	252	15659
08 Castilla-La Mancha	823	1540	1628	1749	1758	2903	1334	448	143	12326
09 Cataluña	6566	11124	10434	10066	9733	14847	6050	1692	526	71038
10 C. Valenciana	3757	6669	6623	7144	7347	12774	5821	1687	531	52353
11 Extremadura	482	848	958	1087	1070	1790	1076	416	131	7858
12 Galicia	993	2045	2220	2403	2537	4697	2544	948	352	18739
13 Madrid, C. de	3753	6688	6676	6796	6395	10472	5125	1499	499	47903
14 Murcia, R. de	956	1805	1934	1883	1900	3058	1296	333	71	13236
15 Navarra	376	651	600	613	653	907	452	119	55	4426
16 País Vasco	1942	3045	2983	2539	2654	4212	2179	605	195	20354
17 La Rioja	181	313	311	329	322	516	278	92	37	2379
18 Ceuta	145	191	207	184	149	216	144	30	16	1282
19 Melilla	176	241	198	145	139	164	60	28	3	1154
Total	30617	53162	54074	56125	55991	90679	44285	13829	4432	403194

Cuadro 8: Distribución de delitos por edad y comunidad autónoma

De primeras, gracias a la tabla de contingencias se puede ver cómo en las columnas destaca la categoría de 41 a 50 años, mientras que en las filas destaca Andalucía, seguida de la Comunidad de Madrid. Se muestra ahora el *screeplot* elaborado con el fin de observar el porcentaje de varianza explicada en cada una de las dimensiones.

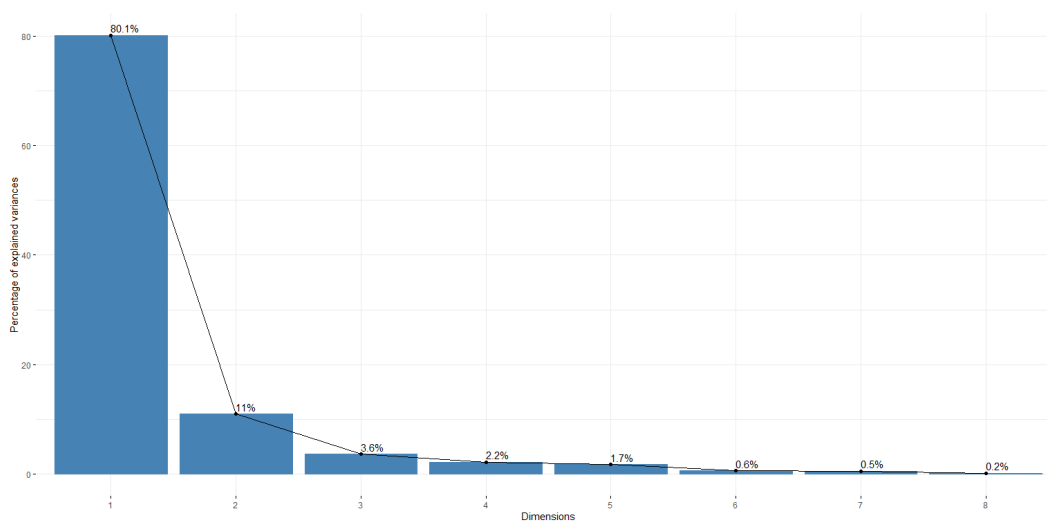


Figura 19: Porcentaje de varianza explicada por cada componente principal

Dimensión	Valor propio	% Varianza	% Acumulada
1	0.00970	80.11	80.11
2	0.00133	10.98	91.09
3	0.00044	3.64	94.73
4	0.00026	2.17	96.90
5	0.00021	1.75	98.65
6	0.00008	0.62	99.27
7	0.00007	0.54	99.81
8	0.00002	0.19	100.00

Cuadro 9: Autovalores y varianza explicada

Tal y como se observa en las anteriores figuras, la extracción de dos dimensiones es suficiente para alcanzar más del 90 % de la variabilidad total del modelo.

Una vez que ha sido seleccionado el número de dimensiones, se analiza la contribución de cada una de las dimensiones por parte de los perfiles fila y columna. Respecto a los perfiles fila, se observa en la figura 20 que la mayor contribución en la primera dimensión viene dada por Cataluña mientras que en la segunda por País Vasco, seguida de Andalucía.

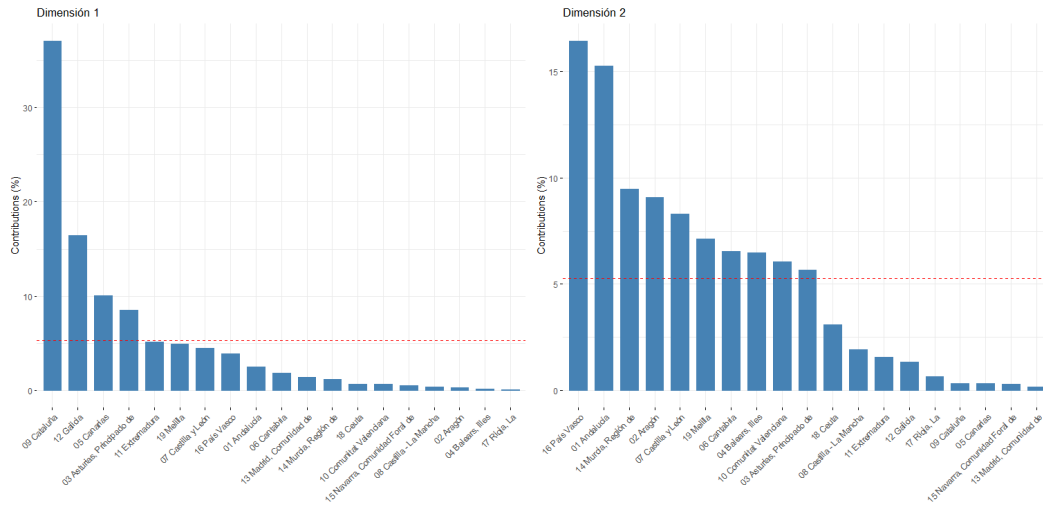


Figura 20: Contribución de las filas

En cuanto a los perfiles columna, se puede apreciar que en la primera dimensión la mayor contribución corresponde con el grupo de edad de 51 a 60 años, seguido de los grupos de 21 a 25 y 61 a 70 años. Por otro lado, en la dimensión 2 el grupo de edad que más contribuye es, con una diferencia considerable, el de 18 a 30 años.

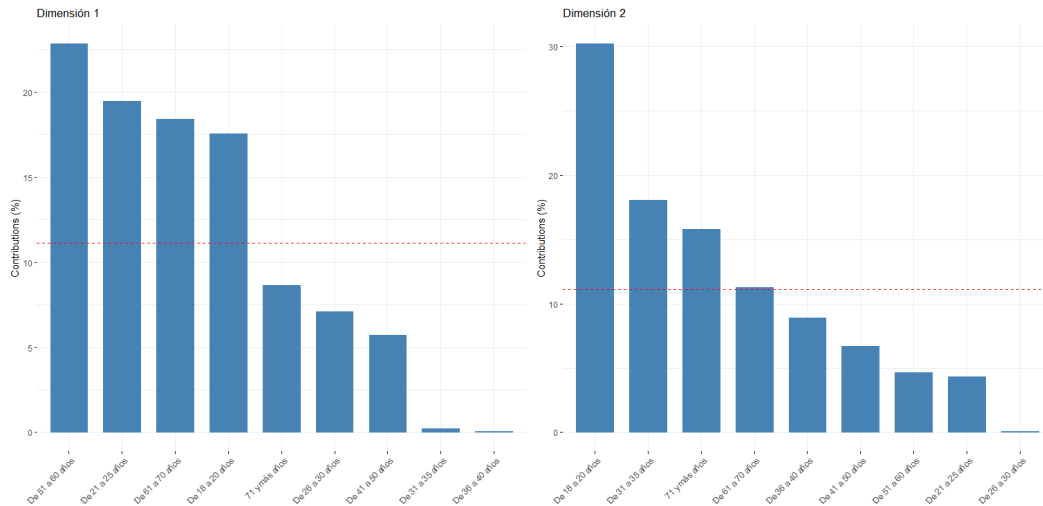


Figura 21: Contribución de las columnas

Una vez se ha seleccionado el número de dimensiones a extraer y se ha estudiado como han sido elaboradas mediante las contribuciones de los dos perfiles se analiza el *biplot*. En este se muestran las posiciones relativas de ambas variables proyectadas sobre las dos dimensiones seleccionadas, que explican entre las dos el 91.1 % de la inercia total.

En el eje horizontal, que representa la primera dimensión, se observa como la comunidad de Melilla y, en menor medida, Ceuta, aparecen alejadas del resto de comunidades. Concretamente Melilla y el grupo de edad de 18 a 20 años se ven representadas en la misma dirección, lo que indica que esta comunidad tiene una proporción alta de delincuentes de este grupo etario en comparación con el resto. Lo mismo pasa con el Castilla y León y los infractores de más de 71 años o entre el Principado de

Asturias y el grupo etario de 61 a 70. Estos últimos además parecen tener una posible asociación, lo que indica que en dicha región la delincuencia debido a personas de ese grupo de edad es mayor que en el resto de las comunidades. Además, tanto la región como el grupo parecen estar bien representados, pues presentan un valor del coseno al cuadrado suficientemente alto.

En cuanto a asociaciones entre comunidades, se observa una posible relación entre Andalucía y Castilla – La Mancha (y algo más lejana Comunidad Valencia) lo que incita a pensar que tienen perfiles de delincuencia similares. Se observa además que la primera dimensión discrimina por edades ancianas y jóvenes, presentándose en el sector positivo los infractores de mayor edad, mientras que el eje vertical, que representa la segunda dimensión, se observa que este distingue los distintos grupos de edad de manera más extrema, situándose las edades menores (de 18 a 20) y mayores (mayores de 71) en el eje positivo.

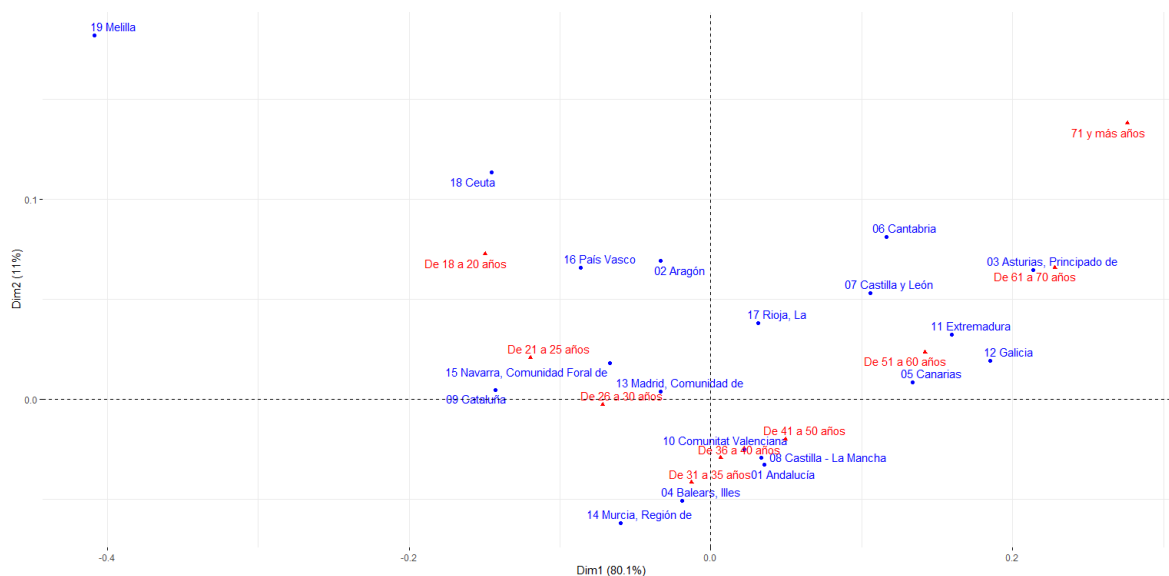


Figura 22: Biplot del análisis de correspondencias

Cabe destacar el gráfico también permite ver cuáles son las comunidades y los grupos de edad que menos aportan a la creación de las dos dimensiones extraídas; estos son los que se encuentran más cercanos al centro del eje. Lo mismo ocurre con los que se encuentran más cercanos al eje de coordenadas, como la Comunidad de Madrid o Cataluña, que no aportan apenas a la definición de la segunda componente o los que se encuentran próximos al eje de abscisas que no aportan a la primera, como las Islas Baleares. No obstante, como se mencionó anteriormente, es posible que debido a la asimetría observada las comunidades con mayor frecuencia pueden tener más peso en la representación gráfica.

7.1.1. Efecto Guttman

Además, se puede intuir la presencia de un efecto Guttman. Este patrón aparece cuando las categorías de alguna de las variables tienen un orden natural, como es en este caso el orden de las edades, y la representación gráfica de sus coordenadas en los ejes forman una parábola. Se observa, en este caso, cómo los grupos etarios más jóvenes (de 18 a 20 años) y los más mayores (de más de 71 años) se sitúan en los extremos del gráfico, mientras que los grupos intermedios se posicionan más cerca del origen. Esto supone una estructura en la que las clases extremas ya mencionadas se oponen en la primera dimensión, mientras que el segundo eje discrimina a las clases medias. La identificación de este efecto ayuda, por tanto, a interpretar que los delitos cometidos por los individuos más jóvenes y más mayores siguen dinámicas distintas.

7.2. Análisis de componentes principales

El análisis en componentes principales (*ACP*) se trata, al igual que el *AC*, de un procedimiento multivariante de interdependencia que tiene como finalidad estudiar las relaciones entre las variables de interés y los individuos del conjunto de datos, que en este caso son los infractores. El objetivo principal es reducir la dimensión del conjunto de datos intentando minimizar lo máximo posible la pérdida de información. Para ello, se construye un conjunto de variables nuevo a partir de combinaciones lineales de las variables originales que recojan la mayor cantidad de información posible.

Previo al análisis en componentes principales, se estudia la posible asimetría de cada uno de los tipos de delitos. Teniendo en cuenta:

- Asimetría > 0 : distribución asimétrica positiva, los datos tienden a concentrarse en los valores más bajos, dejando una cola hacia la derecha.
- Asimetría $= 0$: distribución simétrica, los datos se distribuyen de forma equilibrada a ambos lados de la media.
- Asimetría < 0 : distribución asimétrica negativa, los datos tienden a concentrarse en los valores más altos, dejando una cola hacia la izquierda.

Tipo de delito	Asimetría
Homicidio y sus formas	1.126965
Contra las relaciones familiares	1.670472
Contra el patrimonio y el orden socioeconómico	1.305689
Contra la seguridad colectiva	1.268264
Falsedades	1.197947
Contra la Administración Pública	1.868516
Contra la Administración de Justicia	1.402852
Contra el orden público	1.449863
Lesiones	1.229296
Contra la libertad	1.683932
Torturas e integridad moral	1.698898
Contra la libertad e indemnidad sexuales	1.199739
Resto de delitos	1.721447

Cuadro 10: Coeficiente de asimetría de las variables delictivas

Todas las tipologías de los delitos tienen un valor positivo, lo que muestra que siguen una distribución asimétrica a la derecha. Es por ello por lo que se realiza una transformación logarítmica previa al *ACP*, para ello se utiliza la función `log1p` en lugar de `log`, ya que al tener valores muy próximos a cero se evitan obtener resultados indefinidos.

$$X' = \log(1 + X)$$

Posteriormente, aunque todas las variables tienen las mismas unidades, se escalan mediante la función `scale` con la finalidad de evitar que el posterior análisis de componentes principales de más peso a las variables con mayor rango de valores. Esta hace que cada variable, en este caso el tipo del delito tenga media 0 y varianza 1.

$$Z = \frac{X - \mu}{\sigma}$$

Tras aplicar la función que permite llevar a cabo el análisis (`prcomp`) se debe elegir el número de componentes que se extraerán. Para ello se muestra en la figura 23 el *scree-plot* en el que se aprecia que con una única componente se alcanza una varianza explicada de más del 90 %.

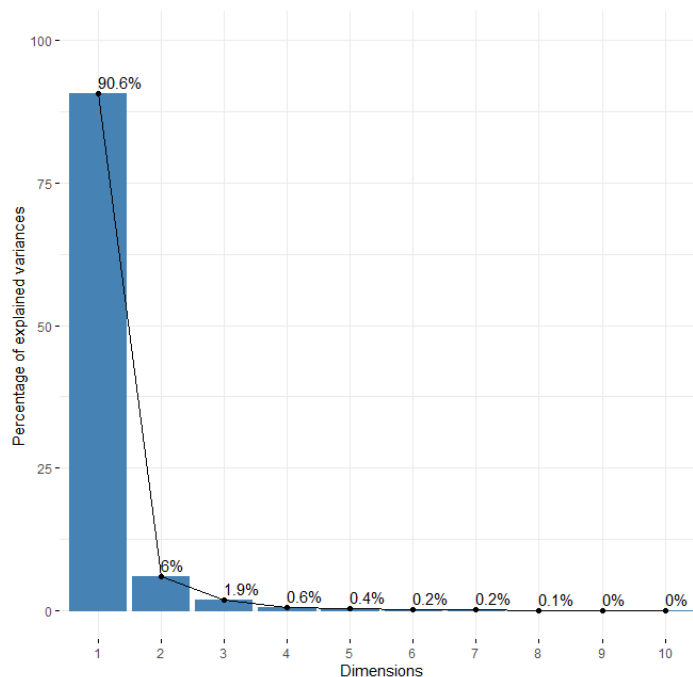


Figura 23: Porcentaje de varianza explicada por cada componente principal

El (*biplot*) que se presenta a continuación permite representar de manera simultánea las distintas regiones y los tipos de delitos en el espacio formado por las dos primeras componentes principales. Pese a que con una componente sería suficiente para explicar más de un 90 % de la varianza total, se representa también la segunda, explicándose así conjuntamente un 92,5 % con el fin de ofrecer una representación lo más fiel posible a la estructura de los datos.

Por una parte, en la figura 24 se aprecia, en el primer componente, como La Rioja, Comunidad Foral de Navarra, Principado de Asturias, Extremadura o Cantabria se sitúan próximas en el gráfico, siendo estas comunidades con poca tasa de criminalidad. Esto demuestra que la distribución de los tipos de delitos en estas regiones se distribuye de manera similar. Por lo contrario, en la esquina inferior

izquierda del gráfico se agrupan Andalucía, Comunidad Valenciana, Cataluña y Comunidad de Madrid, siendo estas las regiones con mayor tasa de criminalidad, que parecen también tener similitudes en la distribución de la tipología de los delitos. Las regiones Ceuta y Melilla, por otra parte, parecen tener perfiles delictivos muy distintos al resto de comunidades.

Además de las agrupaciones de las distintas regiones, es importante analizar la información que nos proporciona la orientación de los vectores correspondientes a los distintos tipos de delitos. La dirección de cada flecha indica hacia dónde aumentan los valores de esa variable en el espacio de las componentes principales, mientras que su longitud nos muestra el grado de contribución de dicha variable a la varianza explicada. Islas Canarias parece estar bastante asociada con los delitos contra la libertad e indemnidad sexuales, mientras que Castilla y León parece estarlo con el delito contra la administración de justicia, entre otros.

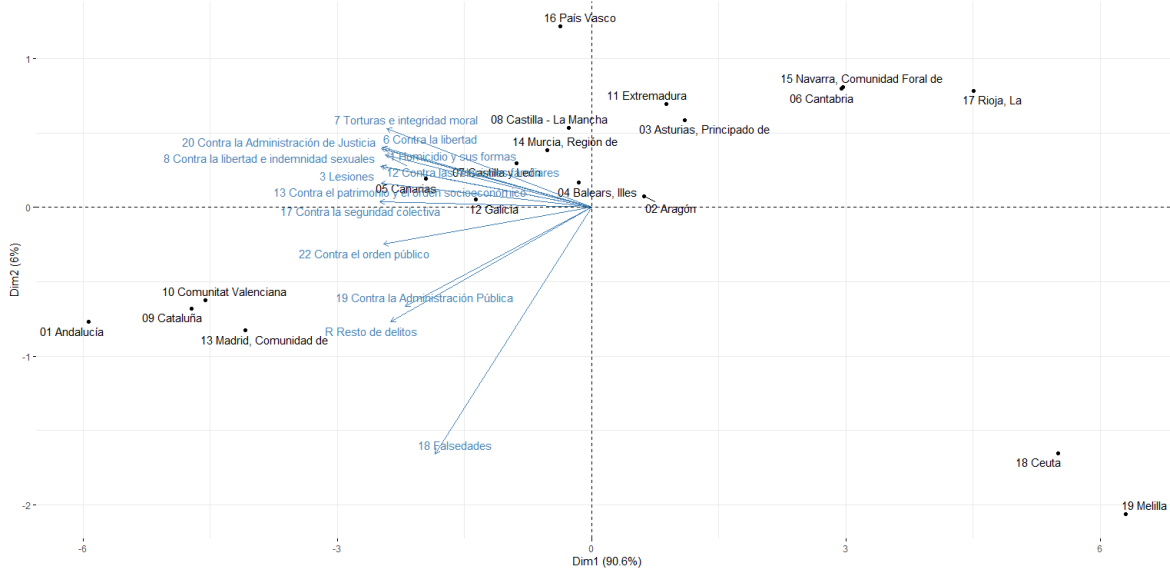


Figura 24: Biplot del análisis de componentes principales

7.2.1. Clasificación por clustering

Se quiere agrupar a continuación las 19 comunidades autónomas en k grupos (*clusters*) de tal forma que los individuos dentro de un mismo grupo sean lo más homogéneos posible, además de lo más diferentes posibles a los individuos en otros grupos. Como se trata de un agrupamiento no supervisado, se debe decidir primero el número de *clusters* a tomar. Para ello, se ha partido de la matriz de datos normalizada y transformada logarítmicamente que recoge el número total de delitos por tipo y comunidad autónoma. Tras aplicar el *ACP*, se han extraído los scores de cada comunidad en el nuevo espacio y sobre estos se ha calculado una matriz de distancias euclídeas:

$$d_E(x_i, x_{i'}) = \|x_i - x_{i'}\| = \sqrt{(x_{i1} - x_{i'1})^2 + \dots + (x_{ip} - x_{i'p})^2}$$

Posteriormente se ha aplicado un algoritmo de agrupamiento jerárquico con el método de *Ward* para el cálculo del índice de agregación $\delta(A, B)$.

$$\delta(A, B) = \frac{n_A \cdot n_B}{n_A + n_B} d(g_A, g_B)^2$$

El proceso anterior permite crear el dendrograma que se muestra a continuación y que permite

identificar los distintos *clusters* atendiendo a los perfiles delictivos en cada región. Las comunidades autónomas fueron asignadas a cada uno de los grupos mediante la función `cutree()`.

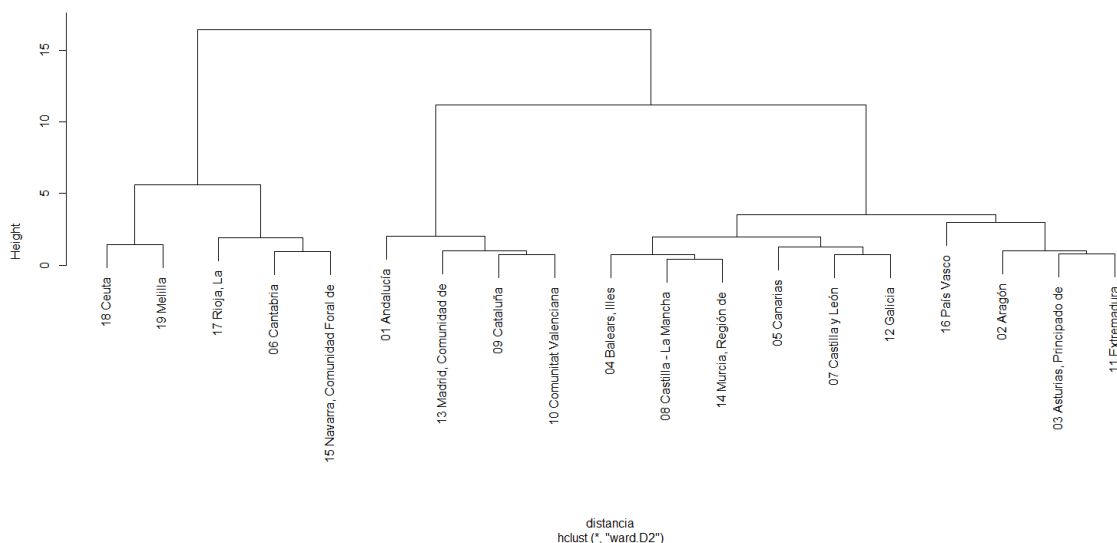


Figura 25: Dendrograma jerárquico

Observando el dendrograma, se ha decidido crear 4 clusters distintos ($k=4$), ya que, a partir de esa división, se aprecian agrupaciones bien definidas y separadas por una altura significativa, obteniéndose la siguiente distribución. El grupo que agrupa un mayor número de comunidades es el grupo 2.

Cluster	Número
1	4
2	10
3	3
4	2

Cuadro 11: Agrupación de los *clusters*

Se representa gráficamente a partir de esta decisión la agrupación de las comunidades autónomas en el plano definido por las dos primeras componentes principales del análisis de componentes principales, distinguiendo por colores los puntos según al grupo al que pertenece cada una de las regiones. Esta representación permite observar cómo las comunidades agrupadas en zonas cercanas en el plano tienden a tener perfiles delictivos similares.

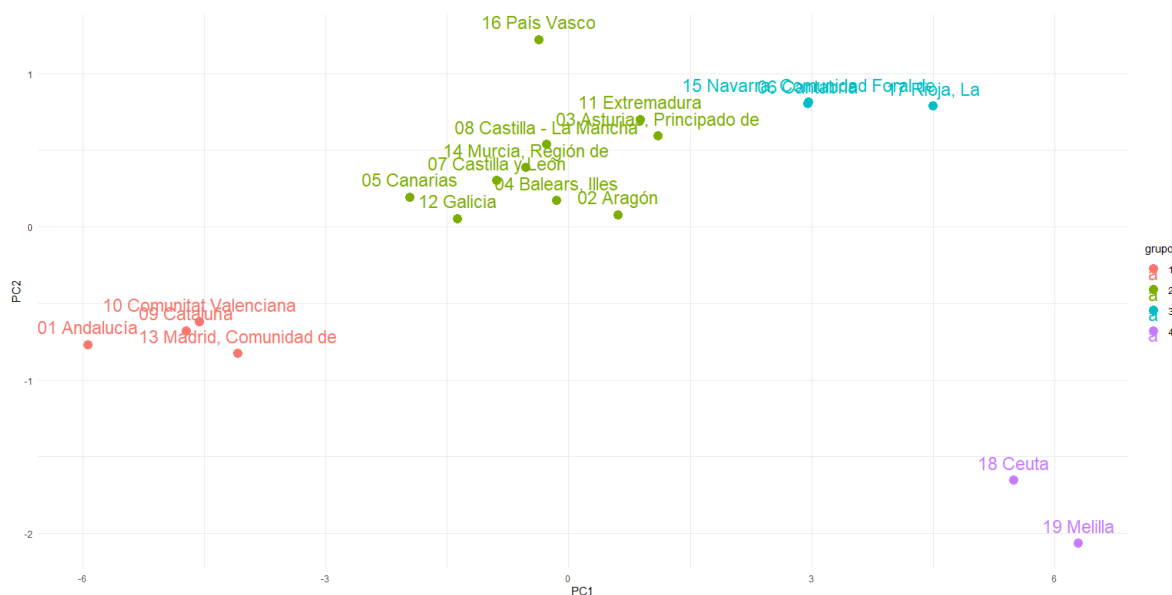


Figura 26: *Clustering* de comunidades autónomas

El *cluster* número 1 formado por Andalucía, Cataluña, Comunidad Valenciana y Comunidad de Madrid reúne aquellas comunidades con mayor población del país. El grupo 4, en cambio, agrupa únicamente dos comunidades; Ceuta y Melilla, siendo estas las únicas regiones situadas geográficamente en el continente africano. En cuanto al grupo 3 recoge tres regiones; Comunidad Foral de Navarra, Cantabria y La Rioja, estas además de tener una baja tasa de criminalidad, se encuentran geográficamente cercanas, por lo que no sorprende la similitud delictiva. Por último, el segundo *cluster* recoge el resto de las comunidades autónomas del territorio peninsular e insular, conformando el grupo más numeroso. Estas regiones comparten un perfil delictivo más equilibrado, sin valores extremos ni en distribución tipología del delito, ni en tasa de criminalidad, ni en volumen de población.

En definitiva, se ha clasificado las comunidades autónomas en función de su comportamiento delictivo, detectando patrones regionales.

8. Comparativa de criminalidad: Castilla y León y la Comunidad de Madrid

Para este análisis se han seleccionado dos comunidades autónomas con características sociodemográficas y territoriales claramente diferenciadas: Castilla y León y la Comunidad de Madrid. La elección de estas regiones responde al interés de contrastar dos modelos poblacionales distintos: por un lado Castilla y León, una comunidad extensa territorialmente, con predominio del medio rural y una población más envejecida; por otro, Madrid, una comunidad urbana, con una alta densidad poblacional y gran diversidad social. Esta diferenciación se refleja además en el análisis de clustering, donde ambas quedaron clasificadas en grupos distintos, lo que refuerza la conveniencia de analizarlas de forma comparativa en relación con la criminalidad.

A continuación, se analiza la evolución temporal del número de personas condenadas en cada una de las regiones en estudio desde 2013 hasta 2023. Este gráfico permite observar las tendencias generales en la criminalidad registrada en ambas comunidades, así como identificar posibles momentos de cambio.

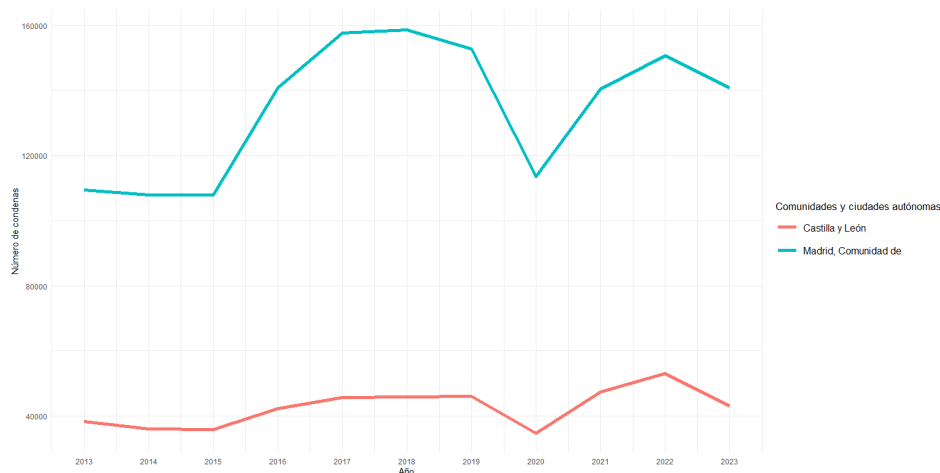


Figura 27: Evolución anual del número de condenas

La Comunidad de Madrid presenta sistemáticamente un número más elevado de personas condenas, lo cual es coherente con su mayor volumen de población. Debido a la variación del tamaño de ambas comunidades resulta más adecuado analizar la tasa de criminalidad, es decir, el número de condenas por cada 100.000 habitantes. Esto permite comparar de manera más precisa y justa la incidencia delictiva relativa en cada territorio. Al estudiar la evolución de esta tasa se puede observar si ambas comunidades siguen una trayectoria paralela o divergente en términos de criminalidad a lo largo de los 10 años de estudio. Este enfoque resulta especialmente útil para identificar patrones estructurales o eventos puntuales que hayan podido influir en la evolución delictiva, tales como reformas legislativas, crisis económicas u otros cambios sociales significativos.

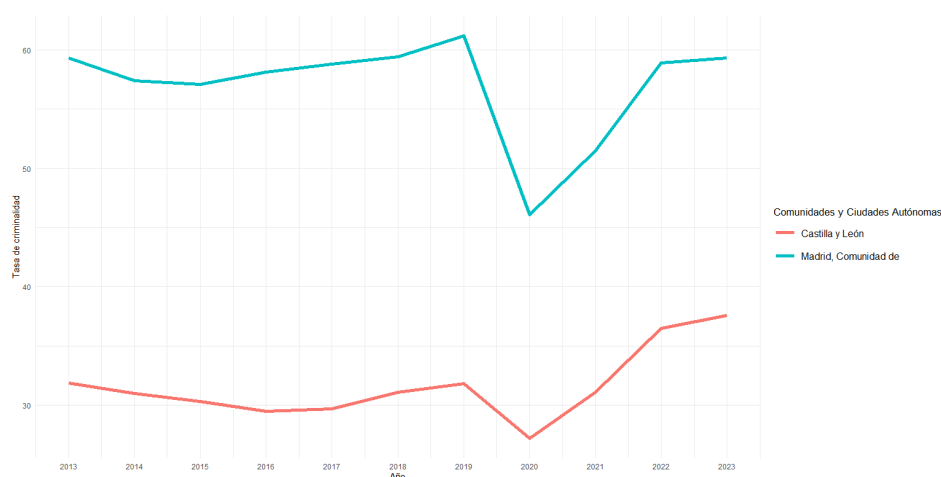


Figura 28: Evolución anual de la tasa de criminalidad

Se observa que la Comunidad de Madrid presenta, de forma sostenida, una tasa de criminalidad significativamente superior a la de Castilla y León a lo largo de los 10 años. Madrid cuenta con una tasa cercana a los 59 delitos por cada 100.000 habitantes al inicio del periodo, manteniéndose relativamente estable hasta 2019, año en el que alcanza su punto máximo. Es en 2020 cuando se produce una caída que puede atribuirse a las restricciones impuestas durante la pandemia por COVID-19. A partir de 2021, la tasa vuelve a incrementarse progresivamente, estabilizándose nuevamente cerca de los niveles prepandémicos en 2023.

Por su parte, Castilla y León muestra una evolución más estable y con valores consistentemente menores. La tasa disminuye ligeramente entre 2013 y 2016, manteniéndose en torno a los 29 delitos por cada 100.000 habitantes. Al igual que en Madrid, en 2020 se detecta la esperable caída debido a la pandemia seguida de un aumento sostenido, alcanzando en 2023 el valor más alto. Este repunte postpandemia se puede asociar a la recuperación de la actividad social y económica.

En resumen, la capital del país presenta una tasa de criminalidad notablemente superior a la de Castilla y León a lo largo del periodo analizado. No obstante, ambas regiones muestran una evolución paralela marcada por una disminución significativa en 2020 y una posterior recuperación. Estas dinámicas reflejan tanto las características estructurales de cada territorio como el impacto de eventos circunstanciales.

Otro eje de análisis fundamental es la distribución de las personas condenadas según su sexo. Esta variable permite observar posibles diferencias estructurales en el comportamiento delictivo entre ambos sexos y su peso en cada comunidad. En línea con lo que reflejan las estadísticas penales a nivel nacional, en ambas comunidades el número de hombres condenados es muy superior al de mujeres, como ya se vio anteriormente. Esta tendencia se mantiene en ambas comunidades aunque con diferentes magnitudes.

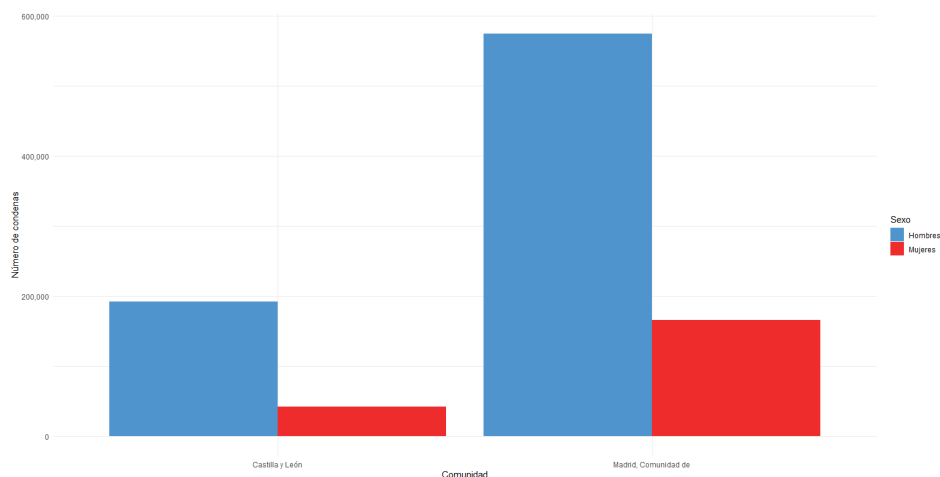


Figura 29: Número de condenas por sexo

En Castilla y León, el número de condenas impuestas a hombres ronda las 190.000, mientras que las correspondientes a mujeres apenas superan las 50.000. Esta diferencia indica que, en esta comunidad, aproximadamente cuatro de cada cinco condenas recaen sobre el sexo masculino.

En la Comunidad de Madrid, la brecha de género en las condenas es aún más pronunciada en términos absolutos. Las condenas a hombres superan las 600.000, mientras que las condenas a mujeres se sitúan por encima de las 150.000. Si bien la proporción sigue siendo similar, el volumen total de condenas es mucho mayor en esta comunidad, lo que puede estar relacionado tanto con factores poblacionales como con dinámicas delictivas propias de un entorno urbano con mayor densidad poblacional.

Otro aspecto relevante en el estudio de la criminalidad es la edad de las personas condenadas. Analizar esta variable permite comprender mejor los perfiles demográficos asociados a la delincuencia e identificar los grupos etarios con mayor incidencia delictiva. A continuación, se presenta la distribución del total de personas condenadas entre 2013 y 2023 distinguiendo por grupo de edad.

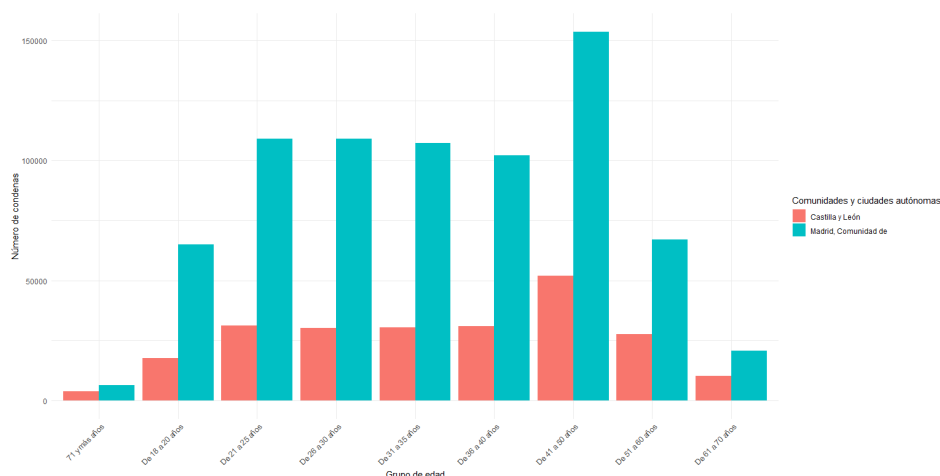


Figura 30: Número de condenas por grupo etario

En ambas comunidades se observa una concentración notable de las condenas en los grupos de edad comprendidos entre los 18 y 44 años, siendo especialmente elevado el número de personas condenadas en el intervalo de 41 a 50 años. Esta tendencia puede estar relacionada con etapas vitales de mayor

actividad social, laboral y económica, en las que también aumenta la exposición a situaciones de riesgo o conflicto.

Castilla y León, a pesar de su menor volumen poblacional, muestra una distribución proporcionalmente similar a la de Madrid, aunque con cifras absolutas más reducidas. En cambio, la Comunidad de Madrid presenta una curva más pronunciada, con una mayor concentración de delitos en las franjas de edad jóvenes y medias, lo que podría estar relacionado tanto con la estructura demográfica de la región como con los patrones de criminalidad propios del entorno urbano.

En ambos territorios, el número de condenas en menores de edad y mayores de 65 son considerablemente más bajas, lo que es coherente con el perfil delictivo más común y con el menor peso de estos grupos en las estadísticas penales. No obstante, es interesante destacar que en Castilla y León el número de condenas en edades avanzadas (a partir de los 51 años) tiene un peso relativamente más significativo, posiblemente relacionado con su población más envejecida.

Para analizar de manera conjunta los factores que influyen a la criminalidad registrada en las dos regiones de estudio, se ha ajustado un modelo de regresión de Poisson. Este tipo de modelo es apropiado para datos de conteo como el número de condenas registradas. Se ha considerado como variable respuesta el número total de personas condenadas incluyendo como variables explicativas la comunidad autónoma, el sexo, el grupo de edad y el año. Las categorías de referencia son Castilla y León, el sexo masculino y el grupo de edad de mayores de 71 años.

$$\log(E[Y_i]) = \beta_0 + \beta_1 \cdot \text{Comunidad}_i + \beta_2 \cdot \text{Sexo}_i + \beta_3 \cdot \text{Edad}_i + \beta_4 \cdot \text{Año}_i$$

donde:

- Y_i es el número de condenas en la observación i .
- β_0 es el intercepto del modelo.
- $\beta_1, \beta_2, \beta_3, \beta_4$ son los coeficientes asociados a las variables explicativas: comunidad autónoma, sexo, grupo de edad y año, respectivamente.

El modelo asume que:

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \text{con} \quad \lambda_i = E[Y_i]$$

y que la relación entre las variables y la media de Y_i se modela mediante un enlace logarítmico.

Los resultados del modelo de regresión de Poisson se resumen en la siguiente tabla.

Variable	Estimación	Error estándar	Valor z	Pvalor
(Intercept)	-44.99	0.458	-98.21	2e-16
Comunidad: Madrid	1.151	0.00168	686.50	2e-16
Sexo: Mujeres	-1.305	0.00247	-527.57	2e-16
Edad: 18 a 20 años	2.096	0.0105	199.48	2e-16
Edad: 21 a 25 años	2.623	0.0103	255.49	2e-16
Edad: 26 a 30 años	2.615	0.0103	254.64	2e-16
Edad: 31 a 35 años	2.606	0.0103	253.61	2e-16
Edad: 36 a 40 años	2.571	0.0103	249.95	2e-16
Edad: 41 a 50 años	3.006	0.0102	295.93	2e-16
Edad: 51 a 60 años	2.232	0.0104	213.90	2e-16
Edad: 61 a 70 años	1.117	0.0114	97.80	2e-16
Año	0.0245	0.000227	108.00	2e-16

Cuadro 12: Resultados del modelo de regresión de Poisson

Los resultados muestran que, manteniendo constantes el resto de variables, el número esperado de condenas en la Comunidad de Madrid es significativamente superior al esperado en Castilla y León. En cuanto al sexo, se observa que el coeficiente asociado a las mujeres es negativo, lo que refleja que el número esperado de condenas a mujeres es considerablemente menor que en hombres. Esta diferencia es estadísticamente significativa y refuerza la desigualdad de distribución de la criminalidad por sexo observada previamente.

Respecto a la edad del infractor, los coeficientes muestran una clara tendencia ascendente respecto al grupo de referencia. Los tramos comprendidos entre los 21 y los 50 años presentan los valores más elevados, lo que indica que estos grupos concentran una proporción mucho mayor de personas condenadas. La diferencia es especialmente marcada en el grupo de 41 a 50 años, que muestra la mayor intensidad delictiva relativa. A partir de los 51 años, los coeficientes comienzan a decrecer, lo que sugiere una disminución de la criminalidad con la edad, lo que resulta coherente con los patrones criminológicos habituales.

Por último, en cuanto al efecto del año, este resulta significativo y positivo, lo que sugiere que con el paso del tiempo, se ha producido un ligero incremento en el número de condenas. Este resultado puede reflejar tanto cambios en las dinámicas sociales como posibles modificaciones legislativas a lo largo de los años.

9. Conclusiones y trabajo futuro

Este trabajo ha permitido relizar una caracterización del fenómeno de la criminalidad en España a través del uso de técnicas estadísticas, tanto descriptivas como multivariantes. A lo largo del análisis se ha observado que los delitos más frecuentes son aquellos que atentan contra el patrimonio y el orden socioeconómico, con especial incidencia en los hurtos, seguidos por los delitos contra la seguridad colectiva, particularmente los relacionados con la seguridad vial. Desde una perspectiva demográfica, el perfil del infractor responde, en general, a un varón joven, constatándose una brecha de género notable, especialmente en delitos de índole sexual, donde la participación femenina es mínima.

El enfoque territorial ha puesto en manifiesto importantes desigualdades entre comunidades autónomas, con tasas de criminalidad más elevadas en regiones densamente pobladas o con fuerte presencia turística, como la Comunidad de Madrid, Cataluña o Islas Baleares, frente a comunidades como Castilla y León, que presentan valores significativamente más bajos. En cuanto a los factores socioeconómicos, el índice de Gini se ha revelado como el indicador con mayor correlación con la tasa de criminalidad, sugiriendo que la desigualdad en la distribución de la renta puede desempeñar un papel determinante en la aparición de conductas delictivas. También se ha encontrado una relación positiva entre el turismo y la criminalidad, lo que podría estar asociado con la mayor proporción de infractores de nacionalidad extranjera en estas zonas.

Desde el punto de vista metodológico, la aplicación de técnicas como el análisis de correspondencias, el análisis de componentes principales y el *clustering* ha permitido identificar patrones comunes y diferencias significativas entre comunidades y grupos poblacionales, enriqueciendo así la comprensión global de la criminalidad.

Como líneas de trabajo futuro, sería interesante ampliar el análisis incorporando nuevas variables, como la reincidencia o el nivel educativo del infractor, así como profundizar en el impacto de políticas públicas concretas en la evolución de la criminalidad. También podría explorarse el uso de modelos predictivos más avanzados, como redes neuronales o árboles de decisión, que permitan estimar con mayor precisión las actividades delictivas según contextos territoriales y/o perfiles sociales. En definitiva, el presente estudio pone en valor el papel de la estadística como herramienta clave para el diseño de estrategias eficaces de prevención y control del crimen, así como para la toma de decisiones informadas por parte de las instituciones al cargo.

Referencias

- [1] Gobierno de España. *El turismo internacional cierra 2021 con un incremento de visitantes del 64,4 % y un aumento del gasto del 76 %*. 2022. URL: <https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/industria/paginas/2022/020222-turimo-frontur-egatur.aspx>.
- [2] Real Academia Española. *Delito*. Diccionario de la lengua española (23.^a ed.) 2024. URL: <https://www.rae.es/drae2001/delito>.
- [3] Instituto Nacional de Estadística. *Menores de edad condenados por delitos sexuales*. 2025. URL: <https://www.ine.es/jaxiT3/Tabla.htm?t=28750>.
- [4] Eurostat. *Criminal Offences by Category*. 2025. URL: https://ec.europa.eu/eurostat/databrowser/view/crim_off_cat/default/table?lang=en.
- [5] Eurostat. *European Union Statistics on Income and Living Conditions*. 2025. URL: <https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions>.
- [6] Eurostat. *European Union Statistics on Income and Living Conditions (ILC_{MDDW03})*. 2025. URL: https://ec.europa.eu/eurostat/databrowser/view/ilc_mddw03/default/table?lang=en.
- [7] Michael J. Greenacre. *La práctica del análisis de correspondencias*. Fundación BBVA, 2007.
- [8] Hosteltur. *España cerró 2021 con 31 millones de turistas, un 63 % menos que en 2019*. Consultado el 10 de julio de 2025. 2022. URL: https://www.hosteltur.com/149602_espana-cerro-2021-con-31-millones-de-turistas-un-63-menos-que-en-2019.html (visitado 10-07-2025).
- [9] Instituto Nacional de Estadística. *Tasa de criminalidad por comunidades autónomas*. 2025. URL: <https://www.ine.es/jaxi/Tabla.htm?path=/t00/ICV/dim6/l0/&file=61101.px&L=0>.
- [10] Instituto Nacional de Estadística. *Tasa de criminalidad por sexo*. 2025. URL: <https://www.ine.es/jaxiT3/Tabla.htm?t=25998>.
- [11] Instituto Nacional de Estadística. *Tasa de criminalidad por tipo de delito*. 2025. URL: <https://www.ine.es/jaxiT3/Tabla.htm?t=25997>.