

Análisis de los Suicidios Registrados en España en el Periodo 2014-2023

Fabián Calvo Castillo, Mario Martínez Guillen, Axel Valton Juan

¹ Máster Universitario en Ciencia de Datos - Universitat de València - ETSE;
facalcas@alumni.uv.es, avaljuan@alumni.uv.es

* Correspondence: marcelino.martinez@uv.es;

Simple Summary: Proyecto Análisis Exploratorio de los Datos.

Abstract: Este estudio realiza un análisis exploratorio de los datos oficiales del Instituto Nacional de Estadística (INE) sobre los suicidios en España entre 2014 y 2023. Las bases de datos incluyen información desglosada por sexo, grupo de edad, mes, nacionalidad y comunidad autónoma, así como datos poblacionales. Tras la importación y limpieza de los datos, se llevaron a cabo tareas de visualización con el objetivo de identificar patrones o posibles factores asociados a la variación en las tasas de suicidio. Finalmente, se desarrolló un análisis para determinar la relación entre el número de suicidios y los factores disponibles. Los resultados aportan una visión global sobre la evolución reciente del fenómeno, proporcionando una base para estudios posteriores.

Keywords: Suicidios; Mortalidad por Suicidio; España; Estadísticas de Mortalidad; Salud Pública; Análisis Exploratorio de Datos; Visualización de Datos

1. Introducción

El suicidio representa un grave problema de salud pública en España y el mundo, cuya comprensión es esencial para desarrollar estrategias de prevención efectivas. Este estudio realiza un análisis exploratorio de los datos registrados de la mortalidad por suicidio en España entre 2014 y 2023, con el objetivo de identificar patrones, tendencias, disparidades demográficas, geográficas y etarias a partir de los datos del INE. El análisis se estructura en torno a tres conjuntos de datos diferenciados: distribución anual por comunidad autónoma, sexo y edad, distribución mensual de los casos en los años estudiados y la distribución de los casos por nacionalidad. A través de la exploración y visualización de dichas variables, se busca describir la magnitud de la problemática que se enfrenta, e identificar posibles relaciones entre las variables analizadas.

2. Método

2.1. Preparación y Adquisición de los Datos

El análisis se fundamenta en los datos de mortalidad por suicidio en España para el periodo de 2014 a 2023, proporcionados por el INE. El punto de partida consistió en múltiples archivos .csv (uno para cada año de estudio), los cuales presentaban un formato estructurado, pero con trabajo por hacer en cuanto al formato de columnas y su distinción por año.

El primer paso fue verificar la codificación de los conjuntos de datos, con el resultado que los conjuntos de Nacionalidad y Sexo y el conjunto de Comunidad Autónoma, Sexo y Edad comparten codificación, siendo esta la ISO-8859-1. Mientras que el conjunto de datos de Mes, Sexo y Edad presenta una codificación de UTF-8.

Por esta razón, la importación de los conjuntos de datos se realiza en la codificación respectiva, y el análisis y graficación de los datos parte de la codificación asignada.

Citation: Calvo, F.; Martínez, M.; Valton, A. Análisis de los Suicidios Registrados en España en el Periodo 2014-2023. *Journal Not Specified* **2024**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2025 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

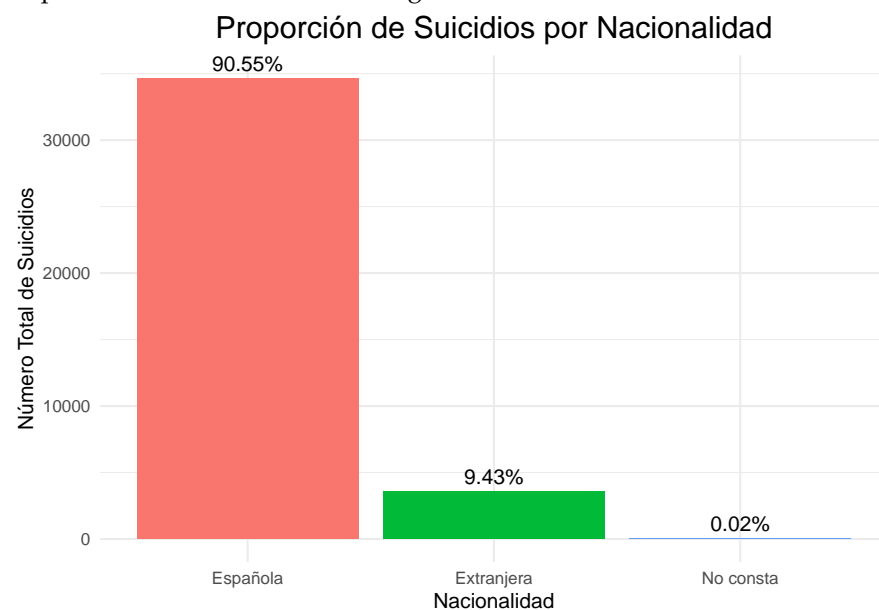
Se destacan tareas de transformación y limpieza en los conjuntos de datos, como la creación de una variable Año para cada uno de los conjuntos, debido a su ausencia en los ficheros originales proporcionados por el INE. Mediante el uso de expresiones regulares se extrae el año (proveniente del nombre del fichero .csv) y se crea esta variable. Asimismo, se ajustó la codificación de cada conjunto de datos y los tipos de datos numéricos y categóricos para las respectivas variables de los conjuntos. Una vez los ficheros se han tratado, se procede a su concatenación para la creación de los tres respectivos dataframes que serán utilizados para el análisis en el proyecto.

Los datos de mortalidad por suicidio almacenados en estos dataframes son absolutos, y pueden no tener una interpretación correcta si se comparan grupos de distinto tamaño. Es por ello que también se han importado datos de población en España durante el periodo de análisis, esto con el fin de contar con las herramientas para responder a preguntas donde la proporcionalidad del tamaño de los grupos que se analizan es fundamental para entender los resultados. En específico, se importan dos datasets, uno con el desglose de poblaciones por comunidades autónomas en el periodo de estudio, y otro con el desglose por Nacionalidad y Sexo. Estos datasets se han obtenido en archivos tipo Excel, descargados desde la página del INE, que no presentaban un formato tidy, por lo que se les realizan las transformaciones respectivas para que adopten un formato adecuado para la futura graficación de resultados. De esta forma, cuando sea necesario se puede disponer de un dato relativo al tamaño del grupo, tratándose de la tasa de suicidios.

2.2. Inspección de los Datos

Con los datos consolidados, se realizó una inspección inicial de las variables disponibles. El dataframe que cuenta con la información de nacionalidad y sexo, se compone de 4 variables siendo estas Año, Nacionalidad, Sexo y Total.

La variable Año es un factor con niveles de 2014 a 2023, mientras que Sexo incluye los niveles "Hombres" y "Mujeres". Por su parte la variable Nacionalidad presenta tres niveles siendo estos: "Española", "Extranjera" y "No consta". Siendo el nivel No consta casos en los que no fue posible determinar la nacionalidad de la persona. Como parte de las decisiones tomadas con este conjunto, se analiza la proporción que esta categoría No consta presenta en los datos, con los siguientes resultados.



Como se observa en el gráfico, la categoría No consta representa una proporción insignificante de los casos con apenas un 0.02%, por lo que se excluye del análisis segmentado por Nacionalidad que se realizará. En cuanto a la variable Total, esta contiene 60 registros que reflejan el número anual de suicidios, desglosado según la nacionalidad de las personas.

A continuación se presenta un resumen de los datos presentes en esta variable, desglosada como hemos dicho por Año, Nacionalidad y Sexo.

Mínimo	Q1	Mediana	Media	Q3	Maximo	Valores_NA
63	185.75	571	956.2	1324	2796	0

Como para nuestro análisis posterior precisaremos de la tasa de suicidios por nacionalidad (agrupando el sexo y dividiendo entre la población por nacionalidad), inspeccionamos si se presentan outliers. Usaremos la regla boxplot para esta detección.

Como podemos observar no hay outliers.

El conjunto de datos que presentan los datos de suicidio separados por Mes, Sexo y Edad se compone de cinco variables siendo estas: Año, Mes de defunción, Sexo, Edad y Total suicidios.

Para las variables Mes de defunción, Sexo y Edad se hizo la limpieza de categorías no necesarias como “Todos los meses”, “Ambos sexos” y “Todas las edades”; esto para que en la visualización no se agregue ruido innecesario debido a que se tiene la capacidad de presentar esta información mediante agrupamientos de los datos. Por último, se presentan los datos de la variable Total suicidios.

Mínimo	Q1	Mediana	Media	Q3	Maximo	Valores_NA
0	4	8	10.63	16	45	0

Como se observa, existen meses sin suicidios registrados, y el máximo registrado en un mes es de 45 personas (ocurrido en agosto de 2020). Finalmente se muestra que la media de suicidios es de 10.63. Se destaca que este conjunto no presenta valores NA.

Finalmente, el conjunto de datos con datos separados por comunidad autónoma presenta las variables Año, Territorio, Sexo, Edad y Total.

Las variables Año se define con un factor con el correspondiente periodo de estudio del proyecto, la variable Territorio incluye 19 niveles que representan las comunidades y ciudades autónomas españolas, De forma coherente con los otros datasets, las variables Sexo y Edad también se establecen como factores según los niveles analizados. Finalmente la variable Total muestra el número de suicidios presentados en bajo las condiciones anteriormente explicadas, con los siguientes resultados.

Mínimo	Q1	Mediana	Media	Q3	Maximo	Valores_NA
0	0	3	6.64	9	85	0

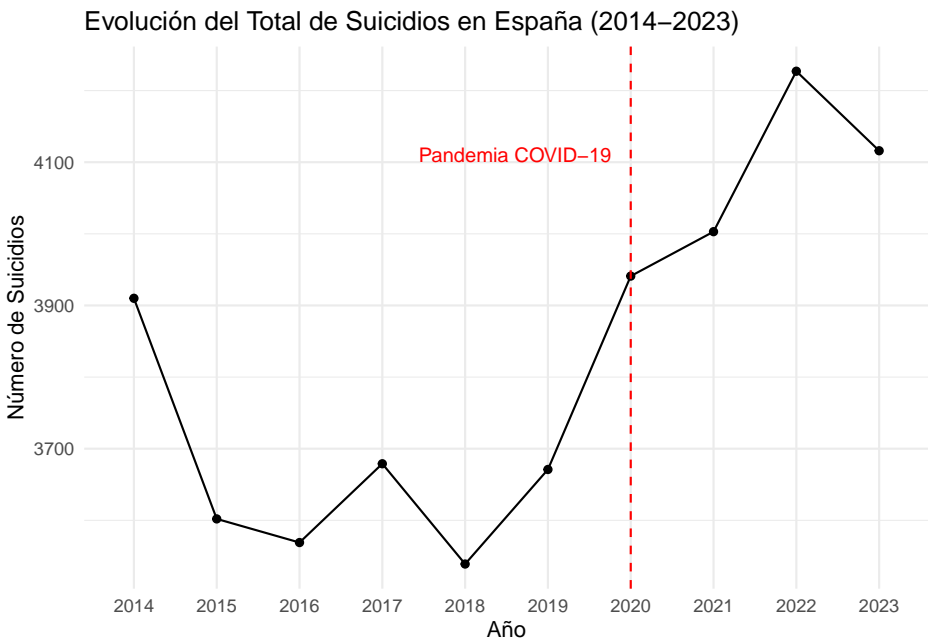
Como más adelante precisaremos también de las tasas de suicidio agrupadas por comunidad autónoma, hacemos un análisis de sus outliers con la regla boxplot.

Outliers Tasa Suicidios por Comunidad autónoma
15

Tras analizar dichos outliers, hemos visto que se corresponde a tasas de suicidios superiores a 12 personas o inferiores a 3 personas por cada 100000 habitantes, no siendo valores demasiado extremos respecto del resto de tasas. Por tanto, las mantendremos para nuestro análisis posterior.

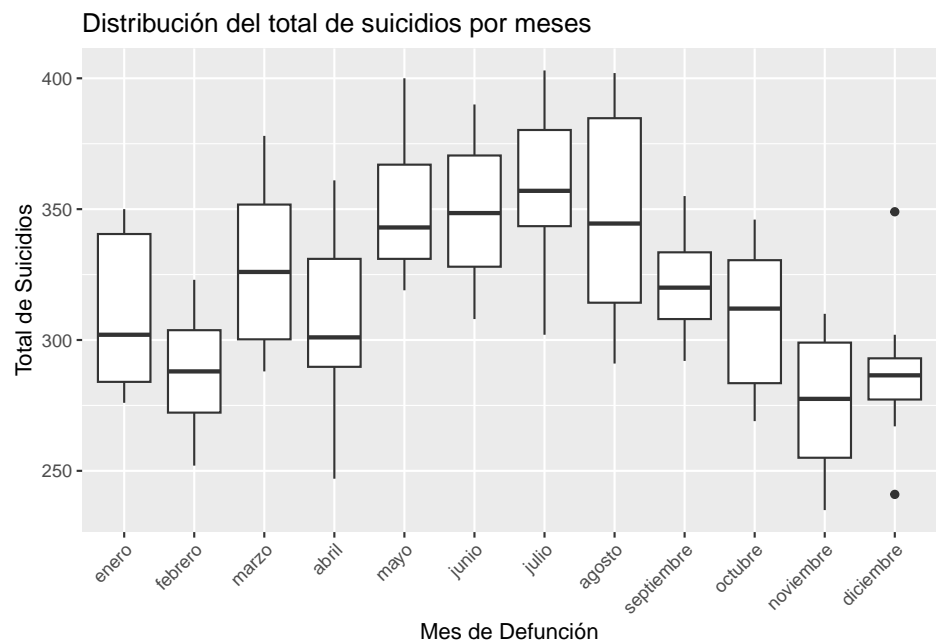
3. Resultados

Una de las primeras preguntas del análisis, es la evolución del número de suicidios en España a lo largo del tiempo. Examinar esta tendencia permite identificar patrones y posibles efectos de acontecimientos específicos en la variación del número de casos registrados. Dado el interés por el comportamiento reciente de los datos, se destaca la pandemia del COVID-19 como un posible factor asociado al aumento de casos que se han registrado, señalándose este hecho con una línea en el gráfico, se ilustra esta evolución con el siguiente gráfico.



Se visualiza un aumento notable en el número de suicidios en España, concentrado principalmente en los últimos años del periodo analizado. Se distinguen dos fases: una etapa prepandémica (2014-2018) con una tendencia variable y el mínimo en el 2018, y una etapa posterior (a partir de 2019) caracterizada por un incremento pronunciado en el número de suicidios que culmina en 2022, año en el que se registró el máximo de suicidios. En 2023 se muestra un leve descenso respecto al año anterior, aunque con cifras superiores a las del periodo prepandémico. Estos hallazgos sugieren un posible impacto de la pandemia y sus consecuencias sociosanitarias en la mortalidad por suicidio en España.

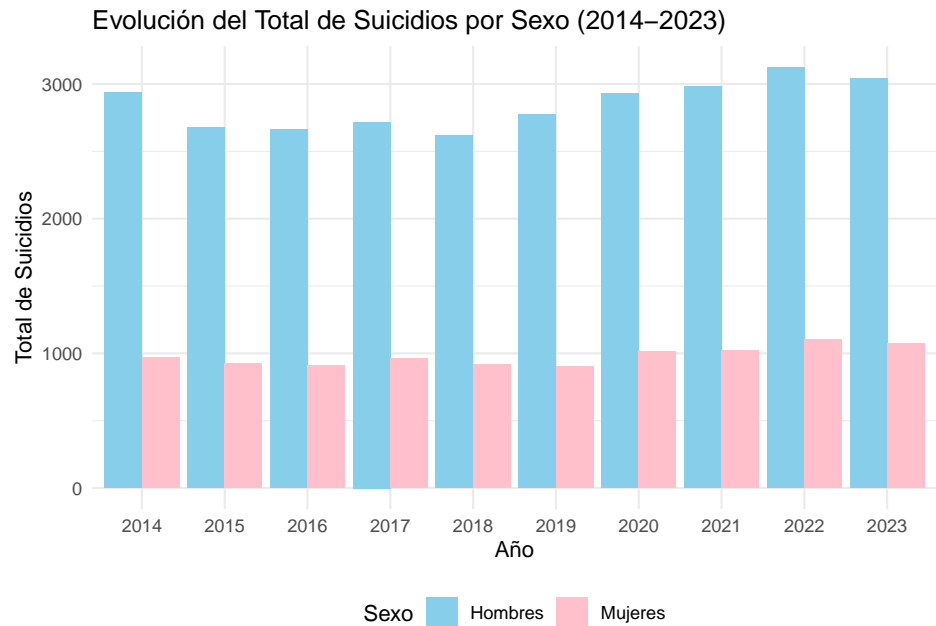
Como continuación del análisis del periodo 2014-2023, se examina la posible estacionalidad en la ocurrencia de los suicidios. Para ello, los casos se agrupan según el mes en el que tuvieron lugar, y se estudia para todos los años. El siguiente gráfico muestra un conjunto de boxplots que resumen la distribución mensual, mostrando la mediana, la variabilidad y la posible presencia de valores extremos en determinados meses.



A partir de este gráfico, se puede refutar una posible idea de que los suicidios ocurren aleatoriamente durante el año, debido a que se observa un patrón estacional. Contrario a la expectativa de mayor incidencia en invierno, los datos muestran que los meses de mayor frecuencia de suicidio se presentan en los meses de verano. Las medianas más altas se sitúan entre mayo y agosto, y progresivamente se presenta una disminución con el paso del tiempo. Además, las cajas más compactas en invierno (diciembre y febrero en particular) indican no solo una incidencia menor, sino también una variabilidad baja en estos meses a lo largo de los años.

Se destaca el mes de diciembre por presentar dos valores atípicos: el valor extremo superior corresponde a diciembre de 2021 (durante la pandemia) con 349 suicidios, mientras que el valor extremo inferior corresponde a 2017 (año prepandémico), con 241 casos registrados.

Es relevante analizar los datos de suicidios por sexo, para identificar posibles diferencias en su comportamiento y derivar a conclusiones específicas de esta problemática. El siguiente gráfico presenta la evolución anual del número de suicidios desglosado por sexo.



124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140

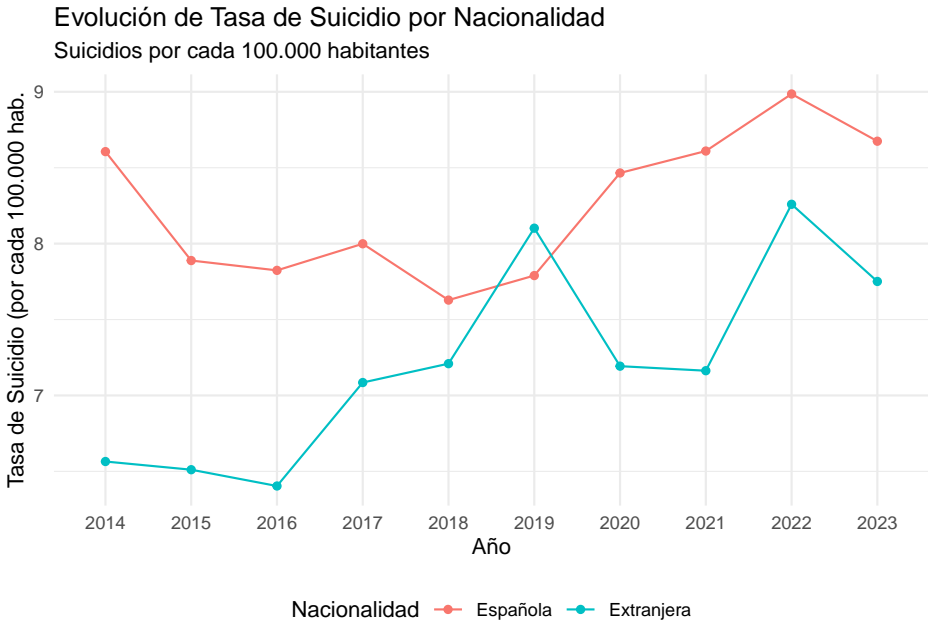
El gráfico evidencia una profunda brecha de género en la mortalidad por suicidio, con un número consistentemente mayor de casos en hombres que en mujeres a lo largo del periodo estudiado. También se observa que tanto en hombres como en mujeres, post pandemia, los registros de suicidios se incrementaron.

Para respaldar esta observación, se realiza el análisis de estas variables para identificar que no es cuestión de aleatoriedad los resultados obtenidos, sino que realmente existe una relación entre el sexo y el número de suicidios. Se realiza la prueba no paramétrica de Wilcoxon, debido a que no se comparan grupos independientes, sino que se comparan datos pareados de hombres y mujeres para cada año, relativizados a la población total en España dividida en estos 2 grupos.

```
##
## Wilcoxon signed rank exact test
##
## data: total_hombres and total_mujeres
## V = 55, p-value = 0.0009766
## alternative hypothesis: true location shift is greater than 0
```

Con el valor de p obtenido en esta prueba, se rechaza la hipótesis nula, que indica que no existe diferencia estadística entre el número de suicidios entre hombres y mujeres mediante el análisis de la mediana de las diferencias. Este análisis confirma que la diferencia observada no es causal, la disparidad entre la mortalidad por suicidio en España entre hombres y mujeres es un fenómeno estadísticamente significativo y de gran magnitud. Mediante este tipo de hallazgos, se destaca que las estrategias de prevención de estos sucesos se deben direccionar y adquirir un enfoque de género diferenciado.

Como continuación del análisis, se estudian los datos de suicidio registrados bajo la segmentación de Nacionalidad. El número de suicidios ocurridos en españoles es ampliamente superior a los extranjeros, pero este análisis es trivial y poco informativo debido a que la población española residente en España es mucho mayor que la extranjera. Para obtener información de estos datos, se analiza el número de suicidios ocurridos por cada cien mil habitantes, haciendo uso de los datos poblaciones de españoles y extranjeros registrados territorio español durante el 2014 al 2023. Obteniendo lo siguiente.



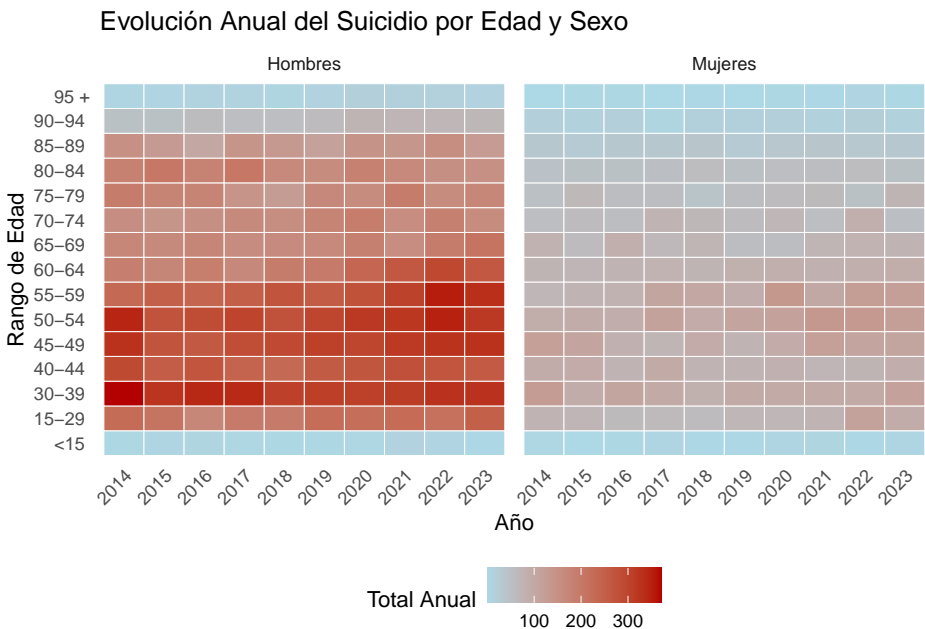
Este gráfico contradice el análisis si toma únicamente el número absoluto de suicidios registrados, evidenciando una mayor paridad en las tasas de suicidio por cada 100.000 habitantes entre españoles y extranjeros. Se evidencia que la población española tenía

una tasa de suicidio superior a la extranjera al inicio del periodo de estudio (2014 a 2016), sin embargo la tasa en extranjeros fue incrementando con el tiempo, alcanzando un pico en 2019 y sobrepasando a la tasa española. La tasa de suicidio en españoles presentó un comportamiento general más suave, con incrementos tras la pandemia, mientras que la extranjera presentó picos y valles más pronunciados. Se concluye con este gráfico que la población extranjera está lejos de ser un grupo de bajo riesgo ante la problemática de los suicidios, con una alta volatilidad, y que su pico máximo se alcanzara en 2019 (pre-pandemia) rompe con la narrativa de un impacto pandémico unificado para todos los grupos, y sugiere que existen otros factores que afectaron a esta población antes de la crisis sanitaria.

Aún con la mayor paridad observada, si realizamos en test de Wilcoxon con las tasas calculadas diferenciadas por grupos observamos que existen diferencias significativas entre las medianas de ambos grupos debido al bajo p-valor, luego la nacionalidad también es influyente en los suicidios.

```
##
## Wilcoxon signed rank exact test
##
## data: total_españoles and total_extranjeros
## V = 54, p-value = 0.001953
## alternative hypothesis: true location shift is greater than 0
```

Con los resultados de los análisis anteriores se han supuesto conclusiones como el aumento en el número de suicidios post pandemia, también que la gran mayoría de estos suicidios se presentan en hombres y que la población extranjera en España presenta variaciones a lo largo del tiempo en la tasa de suicidio por cada cien mil habitantes. Sin embargo no se considera si existen grupos etarios que presentan mayor frecuencia en el número de suicidios. Mediante el siguiente mapa de calor facetado, se visualiza si existen rangos de edad con una mayor incidencia y si los resultados entre hombres y mujeres se asemejan.



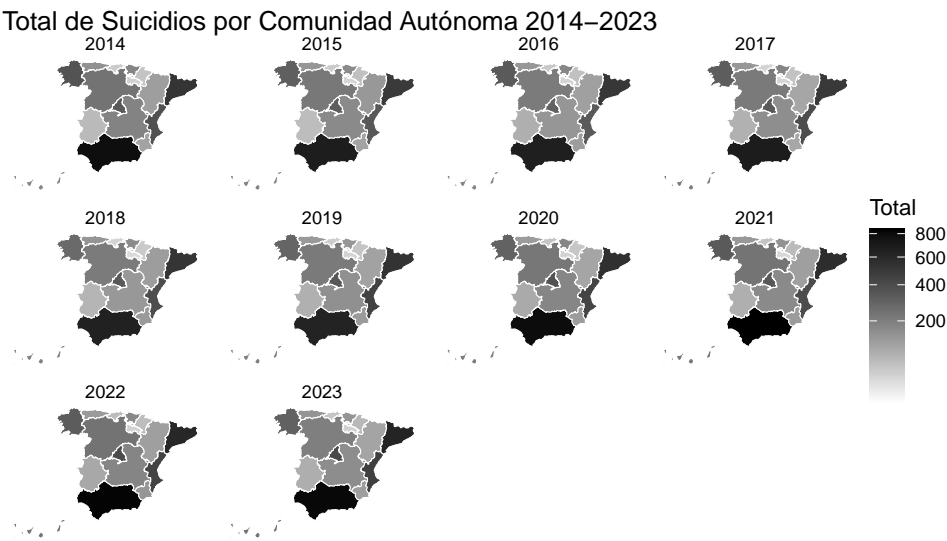
La primera observación que se rescata de este gráfico, es la drástica diferencia entre los dos paneles, siendo el de los hombres el que presenta una mayor incidencia (con más rojos, y más intensos) mientras que el de las mujeres presenta valores mucho menores. También se identifica que la incidencia de los suicidios no se distribuye uniformemente, destacando “puntos calientes” en la mediana edad. Esto se visualiza ya que a partir de los 30 años

hasta los 50, es donde se pinta una mayor incidencia de suicidios, principalmente en hombres. Aunque este patrón también ocurre en mujeres pero con una menor magnitud. Visualizando el gráfico de izquierda a derecha, se vuelve a encontrar el impacto del COVID-19 en la incidencia de suicidios, ya que se visualizan los puntos de mayor intensidad, en ambos paneles, al lado derecho tras el 2020.

Con este mapa de calor se sintetiza, que la narrativa de suicidio en España durante el periodo de 2014 a 2023 se define principalmente en hombres de la mediana edad, y que el impacto de la pandemia no creó un problema nuevo, sino que intensificó el riesgo en estos grupos.

Una pregunta natural que surge con los datos es ¿cómo se distribuyen los suicidios a lo largo de las distintas comunidades autónomas en España?

Para visualizarlo, se presentan unos gráficos en forma de mapa donde cuanto más intenso es el tono de gris, mayor cantidad de suicidios hay en dicha comunidad autónoma.

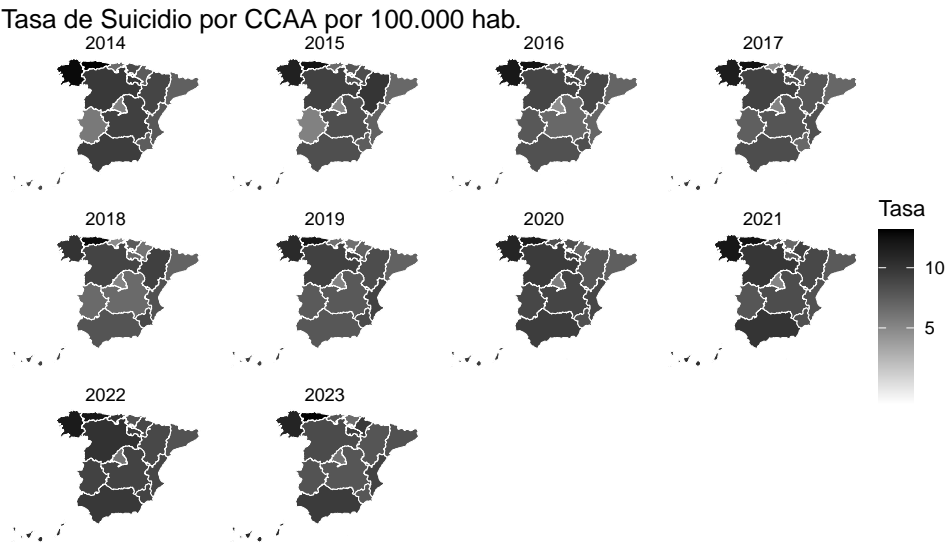


Como se desprende de los mapas, el patrón de total de suicidios es relativamente estable a lo largo de todos los años.

Andalucía aparece de forma constante como la comunidad con las cifras más altas, mientras que regiones como Cataluña, Comunidad Valenciana y Madrid muestran valores intermedios. Por otro lado, comunidades con poblaciones más pequeñas como La Rioja, Cantabria o Navarra mantienen tonos significativamente menores.

Este resultado parece lógico, ya que comunidades autónomas con mayor población tienen mayor número de suicidios. Es por ello que para hacer una comparación entre las comunidades autónomas, se necesita representar una tasa de suicidios que tenga en cuenta la población.

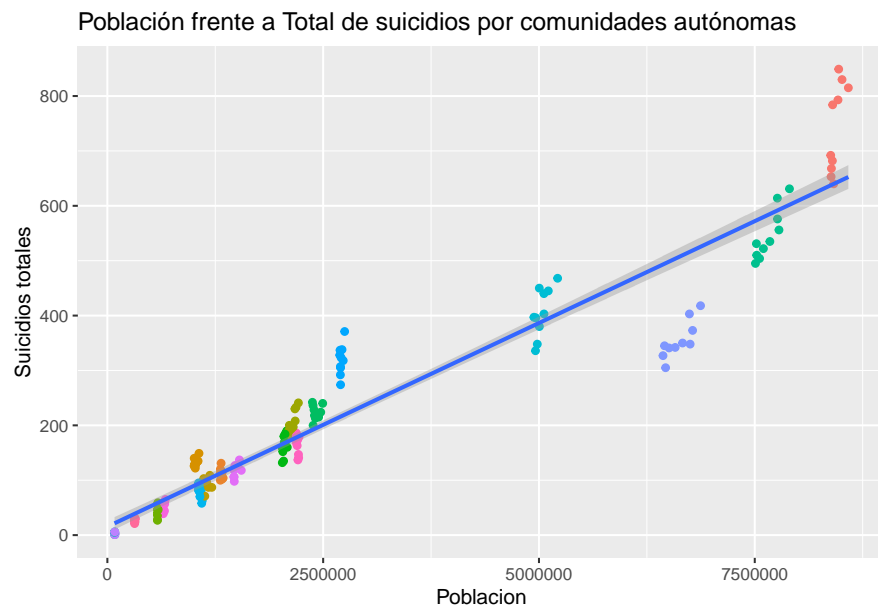
En los siguientes gráficos se representan las tasas de suicidios por cada 100000 habitantes en cada comunidad autónoma, habiendo extraído los datos de población del dataset comentado al principio del documento.



Habiendo considerado la población de cada comunidad, se puede observar que Asturias aparece repetidamente como la comunidad con tasas más elevadas a lo largo de los años, destacando también Galicia y Canarias sobre todo en los años más próximos. Andalucía, Cataluña, Comunidad Valenciana y Madrid ya no destacan sobre el resto ya que comparado con su población no tienen tantos suicidios. En definitiva, los mapas muestran que las tasas más elevadas se concentran en la franja noroeste del país. Con el primer conjunto de mapas que se mostró ya se ha podido vislumbrar que la población influye en el número total de suicidios. ¿Se puede decir entonces que están correlacionados? Obteniendo la correlación de Pearson entre el número de suicidios y el total de población para la comunidades autónomas, resulta que este es de 0.95, es decir, existe una fuerte relación lineal entre los suicidios y el total de población y, consecuentemente, entre los suicidios y la comunidad autónoma.

correlacion
0.9558

Si se visualiza la población frente a los suicidios (con el correspondiente ajuste lineal) se obtiene lo siguiente:



Efectivamente se observa que existe una relación fuertemente positiva entre población y total de suicidios. Esto confirma que las cifras absolutas están fuertemente condicionada por el tamaño poblacional lo que refuerza aún más las idea del uso de tasas poblacionales para comparar las regiones.

4. Conclusiones

El análisis sobre los datos de suicidios en España nos ha permitido identificar patrones relevantes que nos ayudan a comprender mejor este fenómeno. La realización de gráficos y tests estadísticos, así como el análisis de correlaciones, nos ha permitido establecer una realción entre las tasas de suicidios y variables sociodemográficas como el sexo, la población o la nacionalidad.

Los datos muestran que ciertos grupos presentan tasas notablemente superiores, destacando la gran diferencia en el sexo. Esto subraya la necesidad de diseñar estrategias de prevención más específicas y adaptadas a cada grupo.

Así mismo, el estudio también revela una mayor concentración de suicidios en ciertas épocas del año, correspondiéndose estas con los meses de verano, así como en personas de mediana edad.

El análisis realizado desvela la importancia de abordar el tema tratándolo desde una perspectiva integral, desarrollando iniciativas preventivas eficaces y mejorando los sistemas de apoyo a la población de riesgo.

Data Availability Statement: Los datos analizados en este estudio están disponibles públicamente y fueron obtenidos del Instituto Nacional de Estadística (INE). https://ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736176780&menu=resultados&idp=1254735573175#_tabs-1254736194710

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.