



TRABAJO FIN DE GRADO
GRADO EN ECONOMÍA
CURSO ACADÉMICO 2021/2022
CONVOCATORIA: JUNIO

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID: UN ANÁLISIS CON R

AUTOR: GRIAZNOV, VASILY

TUTORES: CUESTA SANTA TERESA, MARINA; MARTIN DE DIEGO, ISAAC

En Madrid, a 9 de junio de 2022

RESUMEN

La desigualdad de la renta ha aumentado en la mayoría de los países desarrollados desde la década de 1980. La dimensión espacial de la desigualdad permite analizar las disparidades desde un enfoque geográfico, pero las medidas de desigualdad tradicionales no captan este aspecto. En este trabajo se combinan las disciplinas de economía, geografía y ciencia de datos para comprobar cómo ha variado espacialmente la desigualdad de la renta en el municipio de Madrid a nivel de distritos en el período 2015-2019. Para ello, se realiza un análisis exploratorio de datos y la descomposición espacial del coeficiente de Gini en componentes inter e intragrupal con el lenguaje de programación R. La descomposición demuestra que la renta de los hogares está correlacionada espacialmente y que la mayoría de la desigualdad se explica por el componente de los distritos no vecinos. Esto es, los distritos que pertenecen a un grupo son similares entre sí, pero hay una importante diferencia entre grupos. Estas diferencias se han mantenido estables en el tiempo.

Palabras clave: desigualdad de la renta, desigualdad espacial, Gini, Madrid, R

Códigos JEL: C88, D31, D63, R12

ABSTRACT

Income inequality has increased in the majority of developed countries since the 1980s. The spatial dimension of inequality allows to analyse disparities from a geographic point of view, but the traditional inequality measures fail to capture this aspect. In this paper, the disciplines of economics, geography and data science are used conjointly to assess how income inequality has changed in the city of Madrid at district level over the 2015–2019 period. For this purpose, an exploratory data analysis is carried out alongside a spatial decomposition of the Gini coefficient within and between the groups using the R programming language. The decomposition reveals that the household's income is spatially correlated and that the inequality is mainly explained by the non-neighbour districts component. That is, the districts belonging to the same group share similar characteristic while there are significant differences between the groups. These differences have remained stable across the studied period.

Keywords: income inequality, spatial inequality, Gini, Madrid, R

JEL classification: C88, D31, D63, R12



© 2022 Vasily Griaznov
Algunos derechos reservados

Este documento se distribuye bajo la licencia Atribución 4.0 Internacional (CC BY 4.0) de Creative Commons, disponible en

<https://creativecommons.org/licenses/by/4.0/deed.es>

ÍNDICE

Índice de tablas.....	4
Índice de figuras.....	5
Introducción.....	6
Medidas de desigualdad.....	8
Datos.....	9
I. Organización territorial y administrativa de España.....	9
II. Iteración.....	10
1. Fondo de Reequilibrio Territorial.....	10
2. Indicadores Urbanos.....	10
3. Atlas de Distribución de Renta de los Hogares.....	11
4. Clasificación Nacional de Educación.....	12
III. Análisis exploratorio de datos.....	12
IV. Cartografía.....	13
1. Modificación del tipo de geometría.....	15
Metodología.....	16
I. Descomposición espacial del coeficiente de Gini.....	16
II. Inferencia.....	16
III. Algoritmo <i>k-nearest neighbours</i> (kNN).....	17
Resultados.....	20
I. Renta de los hogares y tamaño medio del hogar.....	20
II. Tasa de paro y población con estudios superiores.....	24
III. Coeficiente de Gini espacial.....	27
Discusión.....	28
Conclusiones.....	30
Bibliografía.....	31
Anexo. Librerías <i>tidyxl</i> y <i>unpivotr</i>	35

ÍNDICE DE TABLAS

Tabla 1. Datos y variables utilizadas.....	9
Tabla 2. Vecinos más cercanos	19
Tabla 3. Renta neta media por hogar (2015-2019).....	20
Tabla 4. Tamaño medio del hogar (2015-2019).....	21
Tabla 5. Tasa absoluta de paro registrado (2015-2019).....	25
Tabla 6. Población de ≥ 25 años con estudios superiores (2015-2019)	25
Tabla 7. Coeficiente de Gini espacial (2015-2019).....	27

ÍNDICE DE FIGURAS

Figura 1. Cartografía de las secciones censales	13
Figura 2. Cartografía de los distritos	14
Figura 3. Cartografía de los distritos nombrados	14
Figura 4. Cartografía de los centroides geográficos.....	15
Figura 5. Distancia euclidiana.....	18
Figura 6. Cartografía de los $k = 5$ vecinos más cercanos	19
Figura 7. $Q-Q$ plot de la renta neta media de los hogares (2015-2019).....	22
Figura 8. Diagrama de caja de la renta neta media de los hogares por distrito (2015-2019)...	22
Figura 9. Cartografía de la renta neta media de los hogares por distrito (2015-2019).....	23
Figura 10. Cartografía del tamaño medio del hogar por distrito (2015-2019).....	23
Figura 11. Matriz de correlación	24
Figura 12. Cartografía de la tasa absoluta de paro registrado por distrito (2015-2019)	26
Figura 13. Cartografía de la población de ≥ 25 años con estudios superiores por distrito (2015-2019).....	26

INTRODUCCIÓN

La desigualdad de la renta ha aumentado en la mayoría de los países desarrollados desde la década de 1980, aunque con ciertas variaciones en el tiempo y la magnitud. En la eurozona, las políticas de austeridad implementadas durante la Gran Recesión (2008-2013) contribuyeron a acelerar este proceso. La desigualdad presenta muchas dimensiones, entre las que destacan la desigualdad de oportunidades, de riqueza, de educación y de salud. La desigualdad de la renta es, por tanto, una manifestación de desigualdades socioeconómicas más amplias (Heshmati, 2006; Salverda et al., 2014). La teoría moderna de la justicia social reconoce que la desigualdad es una medida compuesta entre factores sobre los que los individuos no poseen ningún control (sexo, lugar de nacimiento, origen socioeconómico y composición genética) y factores que son de su plena responsabilidad, como es el esfuerzo (Cabrera et al., 2021; Roemer, 1998).

La dimensión espacial de la desigualdad permite analizar las disparidades desde un enfoque geográfico, se trata una práctica consolidada en el estudio de la segregación. La segregación socioeconómica también aumentó en las capitales europeas en la década de los 2000, especialmente en lo que se refiere a la separación residencial entre los grupos de mayor y menor renta (Tammaru et al., 2015, p. 22).

El objetivo de este trabajo es comprobar cómo varía espacialmente la desigualdad de la renta en el municipio de Madrid a nivel de distritos en el período 2015-2019. En investigaciones anteriores se analizó la segregación socioeconómica en la Comunidad de Madrid (Leal & Sorando, 2015) y la desigualdad espacial en Sudáfrica (McLennan et al., 2016).

El proyecto se realiza con el lenguaje de programación R (R Core Team, 2022), usando los principios de datos limpios (*tidy data*) de las librerías *tidyverse* (Wickham, 2014; Wickham et al., 2019) y el entorno de desarrollo integrado (*integrated development environment*, IDE) RStudio. Se combinan, por tanto, tres disciplinas diferentes: economía, geografía y ciencia de datos. Los datos y el código fuente están disponibles en

<https://github.com/vasilygrz/income-inequality-madrid>

Siguiendo las recomendaciones del Grupo de Canberra sobre Estadísticas de los Ingresos de los Hogares (Canberra Group on Household Income Statistics)¹, que elaboraron las directrices internacionales para mejorar la comparabilidad de las estadísticas nacionales sobre la renta, la unidad de análisis será el hogar (Deininger & Squire, 1996; Lo Prete, 2018). Para ello, se utilizará la descomposición espacial del coeficiente de Gini propuesta por Rey & Smith (2013) con la librería *lctools* (Kalogirou, 2020). Giorgi (2011) proporciona una revisión de la literatura sobre su descomposición en componentes inter e intragrupal.

Asimismo, se llevará a cabo un análisis exploratorio de datos sobre esta y otras variables, a saber: el tamaño medio del hogar, la tasa absoluta de paro registrado y el nivel de estudios de la población de ≥ 25 años. Llop (2018) realizó una visualización de datos para el municipio de Madrid usando variables similares.

¹ <https://bit.ly/3Mvr05J>

Los resultados obtenidos muestran una brecha socioeconómica que divide el municipio de Madrid en una diagonal que va del suroeste al noreste. En el norte, se ubican los distritos con mayor renta, mayor proporción de la población con estudios superiores y menor tasa de paro. Mientras que en el sur sucede justamente lo contrario. El tamaño medio del hogar ha ido aumentando en los distritos del sureste y suroeste.

La descomposición espacial del coeficiente de Gini demuestra que la renta de los hogares está correlacionada espacialmente. El componente asociado a los distritos vecinos explica el 16,85% de la desigualdad global, mientras que el componente de los distritos no vecinos el 83,15%. Esto es, los distritos que pertenecen a un grupo son similares entre sí –aunque ha habido un ligero aumento de la desigualdad entre ellos– pero hay una importante diferencia entre grupos (83,15%). Esta diferencia se ha mantenido estable en el tiempo a pesar de que desde 2015 la renta neta de los hogares ha aumentado un 12,60% y la tasa de paro ha descendido un 29,06% de media.

El trabajo se estructura de la siguiente forma. En el capítulo 2 se hace un repaso del estado del arte de las medidas de desigualdad, en el capítulo 3 se exponen los datos utilizados y el proceso iterativo hasta llegar a ellos, introduciendo la cartografía de los distritos, en el capítulo 4 se describe la descomposición espacial del coeficiente de Gini y el cálculo del algoritmo *k-nearest neighbours* (kNN), en el capítulo 5 se realiza un análisis exploratorio de los datos, en el capítulo 6 se discuten los resultados y el capítulo 7 cierra con las conclusiones.

MEDIDAS DE DESIGUALDAD

Las medidas de desigualdad y segregación son medidas de dispersión que resumen los valores de una variable. El coeficiente de Gini (Gini, 1912), basado en la curva de Lorenz (Lorenz, 1905), se desarrolló para medir la desigualdad de ingresos. Oscila entre cero –igualdad perfecta– y uno –desigualdad perfecta–. Cero significa que todas las personas tienen los mismos ingresos y uno implica que una sola persona tiene todos los ingresos.

Allison (1978) señala que las buenas medidas de desigualdad satisfacen dos criterios:

1. deben tener invariabilidad de escala, de modo que multiplicar a cada miembro de una población por una constante no debería cambiar el nivel de desigualdad; y
2. deben ser sensibles a las transferencias, de modo que una transferencia de ingresos de una persona con bajos ingresos a otra con más ingresos debería aumentar la desigualdad.

El coeficiente de Gini, así como el coeficiente de variación y el índice de Theil, satisface estos criterios.

El coeficiente de Gini evalúa si los valores están concentrados (o dispersos) en las observaciones, independientemente de la ubicación de estas áreas de concentración en relación con las demás. Es decir, el coeficiente puede decirnos si hay desigualdad, pero no dónde se produce dentro de una región. Las medidas de autocorrelación espacial analizan la distribución espacial de los valores de las variables en relación con los lugares vecinos. En un contexto espacial, el coeficiente de Gini puede reinterpretarse como una medida de la cantidad de concentración de los valores en el mapa (Rey & Smith, 2013).

La descomposición de Rey & Smith (2013) revela que, en lugar de tener que desarrollar una nueva medida que combine una medida de desigualdad junto con una medida de agrupación espacial, el coeficiente de Gini estándar anida en realidad una medida de autocorrelación espacial.

DATOS

La [Tabla 1](#) muestra los datos y variables utilizadas. La búsqueda, selección y manipulación de los datos ha requerido un proceso iterativo de más de cuatro meses de duración. Para poder exponerlo, primero hay que describir la organización territorial y administrativa de España.

Variables		Fuentes
1	Cartografía de las secciones censales de España ²	Instituto Nacional de Estadística
2	Cartografía de los distritos ³	Geoportal, Ayuntamiento de Madrid
3	Renta neta media anual de los hogares ⁴	Atlas de Distribución de Renta de los Hogares, Instituto Nacional de Estadística; Agencia Estatal de Administración Tributaria
4	Tamaño medio del hogar	Atlas de Distribución de Renta de los Hogares, Instituto Nacional de Estadística
5	Tasa absoluta de paro registrado ^{5, 6, 7}	Servicio Público de Empleo Estatal, Ministerio de Trabajo y Economía Social; Subdirección General de Estadística, Ayuntamiento de Madrid
6	Nivel de estudios de la población de ≥ 25 años ^{8, 9}	Subdirección General de Estadística, Ayuntamiento de Madrid

Tabla 1. Datos y variables utilizadas

Fuente: Elaboración propia

I. Organización territorial y administrativa de España

El artículo 137 CE establece que:

El Estado se organiza territorialmente en municipios, en provincias y en las Comunidades Autónomas que se constituyan. Todas estas entidades gozan de autonomía para la gestión de sus respectivos intereses.

España se divide territorial y administrativamente en (de mayor a menor):

1. Comunidades Autónomas (incluidas las ciudades autónomas de Ceuta y Melilla)
2. Provincias
3. Municipios
4. Distritos
5. Barrios
6. Secciones censales

Las secciones censales son unidades de carácter administrativo y con finalidad electoral que surgieron a partir de la Ley Orgánica 5/1985 del Régimen Electoral General. Se trata de una unidad que puede variar en el tiempo en función del tamaño de la población, ya que cada

² <https://bit.ly/3O4IzdU>

³ <https://bit.ly/392cUuT>

⁴ <https://bit.ly/3zppEXj>

⁵ <https://bit.ly/3NAvsBs>

⁶ <https://bit.ly/3tnlbAO>

⁷ <https://bit.ly/3zkr8SK>

⁸ <https://bit.ly/3NACvdA>

⁹ <https://bit.ly/3toL82M>

sección está comprendida entre un mínimo de 500 y un máximo de 2.000 electores (Vono de Vilhena & Bayona i Carrasco, 2010, p. 143).

Actualmente el municipio de Madrid lo conforman 21 distritos y 131 barrios. A continuación, se detallan las últimas modificaciones territoriales:

- Creación del barrio “Ensanche de Vallecas” en el distrito de Villa de Vallecas (BOAM nº 7927, de 15 de junio de 2017).
- Creación de los barrios “Casco Histórico de Vicálvaro” (unión entre “Ambroz” y “Casco Histórico”), “Valdebernardo”, “Valderrivas” y “El Cañaverál” en el distrito de Vicálvaro, así como el cambio de denominación del barrio de “San Andrés” por el de “Villaverde Alto, Casco Histórico de Villaverde” en el distrito de Villaverde (BOAM nº 8034, de 17 de noviembre de 2017).
- Modificación de los límites territoriales de los barrios de Las Palomas y Valdefuentes en el distrito de Hortaleza, así como de los barrios de Alameda de Osuna, Aeropuerto y Timón en el distrito de Barajas (BOAM nº 8928, de 12 de julio de 2021).

II. Iteración

1. Fondo de Reequilibrio Territorial

Originalmente se pensó en emplear los datos del estudio del Fondo de Reequilibrio Territorial (FRT)¹⁰ del Ayuntamiento de Madrid tanto para barrios como para distritos, pero se encontraron las siguientes limitaciones:

- El archivo CSV no contiene lo mismo que los archivos Microsoft Excel (XLS).
- No todos los XLS contienen las mismas variables.
- Los valores de la variable “renta neta media anual de los hogares” no se corresponden a los de su fuente (Indicadores Urbanos, INE).

Por ello, se decidió recolectar los datos de sus respectivas fuentes.

2. Indicadores Urbanos

En 1999 se inició el proyecto europeo Urban Audit^{11, 12} –auspiciado por la Dirección General de Política Regional y Urbana de la Comisión Europea (DG REGIO) en colaboración con Eurostat y los institutos nacionales de estadística– con el fin de recopilar información estadística que permita comparar la calidad de vida en las principales ciudades europeas. Desde entonces se ha venido desarrollando en diversas fases de recogida de datos de aproximadamente tres años cada una.

En 2011 la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y la Comisión Europea (CE) desarrollaron una nueva definición de ciudad y su zona de desplazamiento. Los niveles espaciales utilizados son (de mayor a menor):

¹⁰ <https://bit.ly/3NCJin8>

¹¹ <https://bit.ly/3NyK0Sn>

¹² <https://bit.ly/3MNR0TT>

1. Áreas urbanas funcionales (*functional urban areas*, FUA). Se componen de las ciudades y sus zonas de desplazamiento, son zonas supramunicipales definidas por su influencia laboral.
2. Conurbanizaciones (*greater city*). Son áreas donde el entorno urbano de la ciudad se expande más allá de sus límites administrativos, se trata de una aproximación creada para mejorar la comparabilidad internacional.
3. Ciudades (*local administrative units*, LAU). Son unidades administrativas donde la mayoría de la población vive en un centro urbano de, como mínimo, 50.000 habitantes.
4. Áreas submunicipales (*Sub-city districts*, SCD). Constituyen una subdivisión de la ciudad que, en el caso de España, se construyen por agregación de secciones censales, con la condición de que cada SCD tenga entre 5.000 y 40.000 habitantes. Se trata de una división similar a los barrios, pero no está armonizada para toda la Unión Europea (UE).

Estos datos presentan las siguientes limitaciones:

- Los SCD usan la organización administrativa previa a la modificación de 2017, por lo que hay menos SCD que barrios actualmente existentes.
- La división geográfica no necesariamente coincide con la división administrativa, por lo que no será posible asociar ambas.

A través del Área de Atención a Usuarios del INE me transmitieron que podía replicar los SCD agrupando las correspondientes secciones censales.

Por tanto, se decidió usar la información geográfica a nivel de secciones censales y la renta neta media anual de los hogares del Atlas de Distribución de Renta de los Hogares (ADRH), cuya fuente también es la AEAT.

3. Atlas de Distribución de Renta de los Hogares

El ADRH es un proyecto experimental que se inició en 2019. De forma similar a los Indicadores Urbanos, plantea la construcción de indicadores estadísticos de nivel y distribución de renta de los hogares a nivel municipal e inframunicipal, a partir del enlace de información del INE y con datos tributarios, fundamentalmente de la AEAT. La principal diferencia respecto a los Indicadores Urbanos es que la división geográfica se hace a nivel de secciones censales.

A finales de 2022 el Atlas dejará de figurar como estadística experimental y los datos se integrarán en el Censo de Población y Viviendas.

Estos datos presentan la siguiente limitación:

- Acotación de valores extremos de los indicadores de renta media. Para la cota inferior se emplea el percentil 0,1 y para la cota superior el percentil 99,5.

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID

El Anexo I del proyecto técnico¹³ expone cómo se calcula la renta neta:

$$\text{renta neta} = \text{renta bruta} - \text{impuestos y cotizaciones}$$

donde la renta bruta es:

$$\begin{aligned} \text{renta bruta} &= \text{rentas del trabajo} \\ &+ \text{rentas del capital mobiliario} \\ &+ \text{rentas por arrendamiento de inmuebles} \\ &+ \text{rendimientos de actividades económicas} \\ &+ \text{otras rentas} \end{aligned}$$

4. Clasificación Nacional de Educación

Usamos la clasificación de programas, titulaciones y certificaciones en niveles de formación alcanzados (CNED-A), de la Clasificación Nacional de Educación (CNED-2014)¹⁴, para reagrupar los datos del Ayuntamiento de Madrid. La categoría “educación superior” se compone de cuatro niveles:

- CNED 5: Enseñanzas de formación profesional, artes plásticas y diseño y deportivas de grado superior y equivalentes; títulos propios universitarios que precisan del título de bachiller, de duración igual o superior a 2 años.
- CNED 6: Grados universitarios de hasta 240 créditos ECTS, diplomados universitarios, títulos propios universitarios de experto o especialista, y similares.
- CNED 7: Grados universitarios de más de 240 créditos ECTS, licenciados, másteres y especialidades en Ciencias de la Salud por el sistema de residencia, y similares.
- CNED 8: Enseñanzas de doctorado.

III. Análisis exploratorio de datos

El análisis exploratorio de datos (*exploratory data analysis*, EDA) (Tukey, 1977) es un conjunto de técnicas numéricas y gráficas que sirven para:

- resumir las propiedades de los datos;
- identificar características inusuales o interesantes en los datos;
- detectar errores en los datos; y
- sugerir relaciones que podrían incorporarse a un modelo inferencial (Haining & Li, 2020, p. 159).

Cuando analizamos datos espaciales y espacio-temporales debemos tener en cuenta que los valores que están próximos en el espacio no son independientes. La dependencia es una propiedad fundamental de los atributos en el espacio y el tiempo. Asimismo, la presencia de autocorrelación positiva en los datos espaciales y espacio-temporales implica que los valores cercanos en el espacio y/o en el tiempo tienden a ser similares.

¹³ <https://bit.ly/3tknTXp>

¹⁴ <https://bit.ly/3H4Wqi7>

El análisis de los datos se ha realizado con el lenguaje de programación R, usando los principios de datos limpios de las librerías *tidyverse* y el entorno de desarrollo integrado (*integrated development environment*, IDE) RStudio. Las variables fueron inicialmente procesadas con Microsoft Excel e importadas en R con las librerías *tidyxl* y *unpivotr* (Garmonsway, 2020, 2021). En el [Anexo](#) explico cómo importar datos no tabulares.

Los datos y el código fuente se pueden consultar en

<https://github.com/vasilygrz/income-inequality-madrid>

IV. Cartografía

En España hay 36.382 secciones censales, en la Comunidad de Madrid 4.432 y en el municipio de Madrid 2.443. Visualizamos los datos con las librerías *sf* (Pebesma, 2018) y *ggplot2* (Wickham, 2016) ([Figura 1](#)).

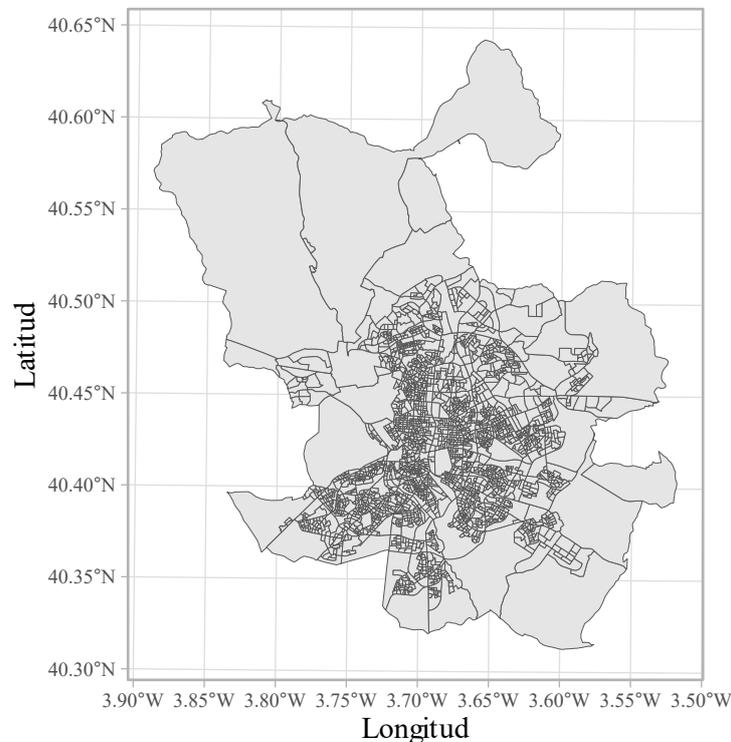


Figura 1. Cartografía de las secciones censales

Fuente: Elaboración propia

El sistema de referencia de coordenadas (*coordinate reference system*, CRS) es el European Terrestrial Reference System 1989 (ETRS89), la unidad de medida son grados y el tipo de geometría es polígono.

Agregamos las secciones censales en distritos ([Figura 2](#)). La agregación da como resultado un área total de 604,4653 km², mientras que la superficie administrativa oficial es de 604,4551 km². Hay una diferencia de 10,2 m².

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID

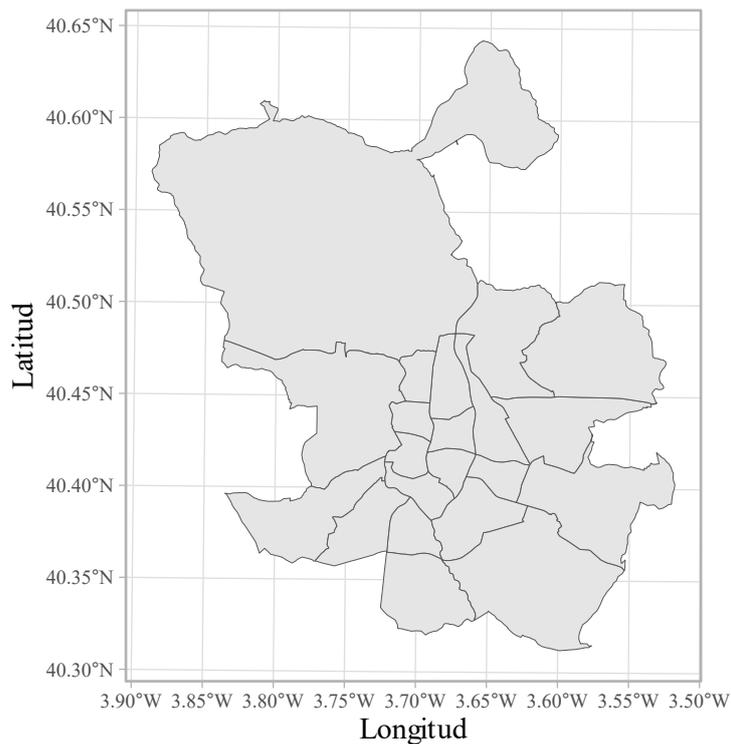


Figura 2. Cartografía de los distritos

Fuente: Elaboración propia

Nombramos los distritos con la librería *ggrepel* (Slowikowski, 2021). Cada distrito tiene asociado un código (Figura 3).

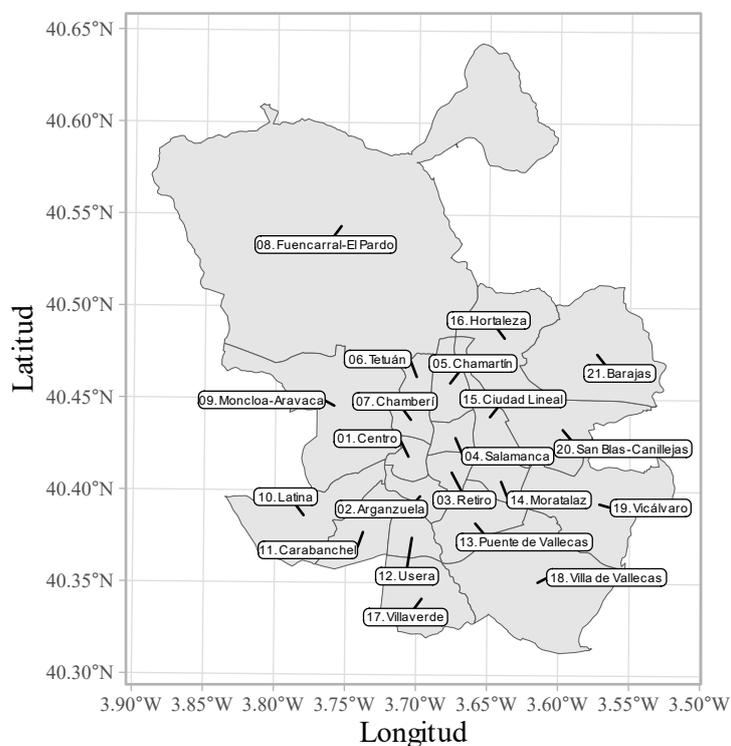


Figura 3. Cartografía de los distritos nombrados

Fuente: Elaboración propia

1. Modificación del tipo de geometría

Cuando trabajamos con datos espaciales es habitual tener que modificar su tipo de geometría para realizar diferentes tipos de análisis. Para convertir polígonos en puntos debemos calcular su centroide geográfico, un elemento que representa el centro de masa de un objeto espacial (Lovelace et al., 2019, p. 94).

La [Figura 4](#) muestra los centroides geográficos de los distritos.

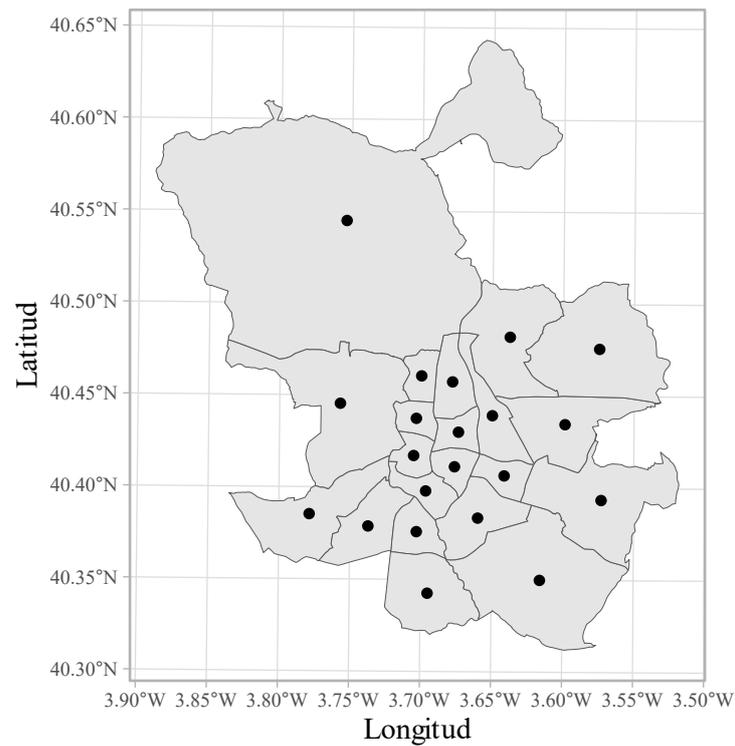


Figura 4. Cartografía de los centroides geográficos

Fuente: Elaboración propia

La [Figura 3](#) también se creó utilizando esta operación geométrica.

METODOLOGÍA

I. Descomposición espacial del coeficiente de Gini

El coeficiente de Gini estándar se define como la diferencia media relativa de los valores de una variable:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2\mu} \quad (1)$$

donde x_i es el valor de la variable x observada en el lugar $i = [1, 2, \dots, n]$, x_j es el valor de la variable observada en un lugar $j = [1, 2, \dots, n]$ y $\mu = \frac{1}{n} \sum_{i=1}^n x_i$.

Rey & Smith (2013) muestran que la suma de las diferencias por pares se puede descomponer como:

$$\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| = \sum_{i=1}^n \sum_{j=1}^n w_{ij} |x_i - x_j| + (1 - w_{ij}) |x_i - x_j| \quad (2)$$

donde w_{ij} es una matriz de pesos espaciales binarios que expresa la relación de vecindad entre los lugares i y j .

Reescribiendo la ecuación (1) obtenemos el coeficiente de Gini espacial:

$$G^s = \underbrace{\frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} |x_i - x_j|}{2n^2\mu}}_{\text{desigualdad intragrupal}} + \underbrace{\frac{\sum_{i=1}^n \sum_{j=1}^n (1 - w_{ij}) |x_i - x_j|}{2n^2\mu}}_{\text{desigualdad intergrupala}} \quad (3)$$

$$= G_n^s + G_{ns}^s$$

El primer término, G_n^s , es el par de observaciones vecinas correspondientes a la desigualdad intragrupal, mientras que el segundo, G_{ns}^s , es el par de observaciones no vecinas correspondientes a la desigualdad intergrupala. A medida que aumenta la autocorrelación espacial positiva, el valor del par de observaciones vecinas se acercará a cero.

Nótese que $G = G^s$ cuando $w_{ij} = 1 \quad \forall i, j$, ya que el par de observaciones no vecinas será cero. Los pesos desempeñan el papel de determinar a qué grupo pertenecen los pares de observaciones. En el caso de las ponderaciones binarias, la pertenencia es mutuamente excluyente.

II. Inferencia

Para comparar las propiedades del coeficiente de Gini espacial como prueba de autocorrelación espacial frente a las de la I de Moran, la prueba más comúnmente empleada para encontrar autocorrelación espacial, se realiza una simulación de Monte Carlo.

La I de Moran (Moran, 1950) se define como:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - \mu)(x_j - \mu)}{\sum_{i=1}^n (x_i - \mu)^2} \quad (4)$$

La prueba de autocorrelación basada en el coeficiente de Gini espacial se define como:

$$SG = \frac{\sum_{i=1}^n \sum_{j=1}^n (1 - w_{ij}) |x_i - x_j|}{2n^2 \mu G^s} \quad (5)$$

donde G^s es el coeficiente de Gini espacial.

SG se puede interpretar como la parte de la desigualdad global que se asocia a los pares de observaciones no vecinas. La inferencia de este estadístico se basa en permutaciones espaciales aleatorias de los datos. Primero, se calcula el valor de la ecuación (5) a partir de los datos originales. A continuación, los valores de los datos originales se permutan espacialmente para simular aleatoriedad espacial y se calcula el estadístico de prueba para este nuevo patrón cartográfico. Se realiza un número de M permutaciones adicionales y se compara el valor original del estadístico con la distribución de valores obtenida a partir de los datos permutados aleatoriamente. El pseudo valor p observado del estadístico de prueba (Davison & Hinkley, 1997, pp. 140-141) se define entonces como:

$$p(SG) = \frac{C + 1}{M + 1} \quad (6)$$

donde C es el número de M muestras de permutación que generaron valores de SG mayores o iguales al valor de G^s observado para los datos originales. Para una prueba con un nivel de significación de $\alpha = 0,05$ se recomienda usar $M = 99$, y para una prueba con un nivel de significación de $\alpha = 0,01$ se recomienda $M = 999$ (Schabenberger & Gotway, 2005, p. 87).

Para la simulación de datos a partir de permutaciones de los originales, se genera en cada realización una variante espacialmente autocorrelacionada:

$$x = \rho Wx + \epsilon \quad (7)$$

donde x es un vector $n \times 1$, ρ es el nivel de autocorrelación espacial, W es una matriz de pesos espaciales binarios $n \times n$ y $\epsilon \sim N(0, \sigma^2 I)$.

La hipótesis a contrastar es:

$$\begin{aligned} H_0: & \text{ La renta no está correlacionada espacialmente} \\ H_1: & \text{ La renta está correlacionada espacialmente} \end{aligned} \quad (8)$$

III. Algoritmo *k-nearest neighbours* (kNN)

El algoritmo *k-nearest neighbours* (kNN) (Fix & Hodges, 1989), k vecinos más cercanos, es una técnica de aprendizaje supervisado (*supervised learning*) que identifica las k observaciones que son similares o más cercanas al valor que se está prediciendo. Se puede utilizar para problemas de clasificación, donde se usa la clase más común, o regresión, donde

se usa el valor medio. La similitud entre observaciones se determina mediante medidas de distancia que calculan las diferencias por pares entre ellas. La distancia euclidiana es la más común y mide la distancia en línea recta entre dos puntos (Boehmke & Greenwell, 2019).

La distancia entre dos puntos con coordenadas (x_1, y_1) y (x_2, y_2) es d :

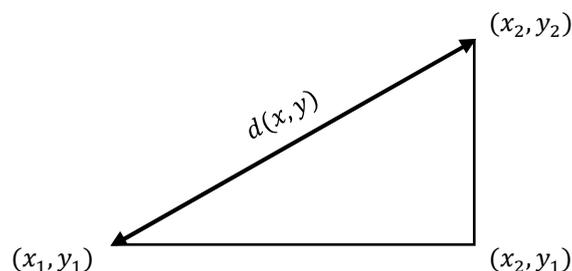


Figura 5. Distancia euclidiana

Fuente: Elaboración propia basada en Crawley (2012, p. 830)

Según el teorema de Pitágoras, el cuadrado de la hipotenusa es la suma de los cuadrados de los dos lados adyacentes:

$$d(x, y)^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 \quad (9)$$

Por tanto:

$$d(x, y) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (10)$$

La distancia del vecino más cercano es el valor mínimo de d (Crawley, 2012, pp. 830-831). No hay una hoja de ruta sobre cómo escoger el valor óptimo de k , que toma valores $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$, pero la regla general es tomar la raíz cuadrada del tamaño muestral¹⁵.

Usamos el algoritmo kNN para calcular los k vecinos más cercanos de cada distrito. Para ello, se usan los centroides geográficos de la Figura 4 para calcular la distancia entre ellos. Los pesos espaciales binarios, w_{ij} , de la ecuación (3) determinan a qué grupo pertenece cada distrito. Cuando se ordena la renta neta de los hogares por distritos a través de un diagrama de caja (*box plot*) se identifican grupos de 4-5 componentes, como se muestra en la Figura 8. Usando una medida de cinco niveles, similar a la escala Likert, podemos dividir estos grupos en:

1. Renta muy baja
2. Renta baja
3. Renta media
4. Renta alta
5. Renta muy alta

De esta forma, para 21 distritos tendremos $\sqrt{21} \cong 5$ vecinos más cercanos. Los calculamos con las funciones *col.knn* y *knn2nb* de la librería *spdep* (Bivand & Wong, 2018), los resultados se muestran en la Tabla 2 y la Figura 6.

¹⁵ Existen técnicas de remuestreo (*resampling*) más avanzadas, entre las que destaca *k-fold cross-validation*.

Distrito	Vecinos más cercanos				
	1	2	3	4	5
01. Centro	07	02	03	04	12
02. Arganzuela	01	03	12	13	11
03. Retiro	04	02	01	14	13
04. Salamanca	03	15	07	01	05
05. Chamartín	06	04	07	15	16
06. Tetuán	05	07	04	01	15
07. Chamberí	01	06	04	05	03
08. Fuencarral-El Pardo	06	09	05	16	07
09. Moncloa-Aravaca	07	06	01	05	10
10. Latina	11	12	09	02	01
11. Carabanchel	12	10	02	01	17
12. Usera	02	11	13	17	03
13. Puente de Vallecas	14	03	02	12	18
14. Moratalaz	13	03	15	04	20
15. Ciudad Lineal	04	05	14	03	20
16. Hortaleza	05	15	21	06	20
17. Villaverde	12	11	13	02	18
18. Villa de Vallecas	13	19	14	17	12
19. Vicálvaro	20	14	18	13	15
20. San Blas-Canillejas	15	14	21	19	16
21. Barajas	20	16	15	05	19

Tabla 2. Vecinos más cercanos

Fuente: Elaboración propia

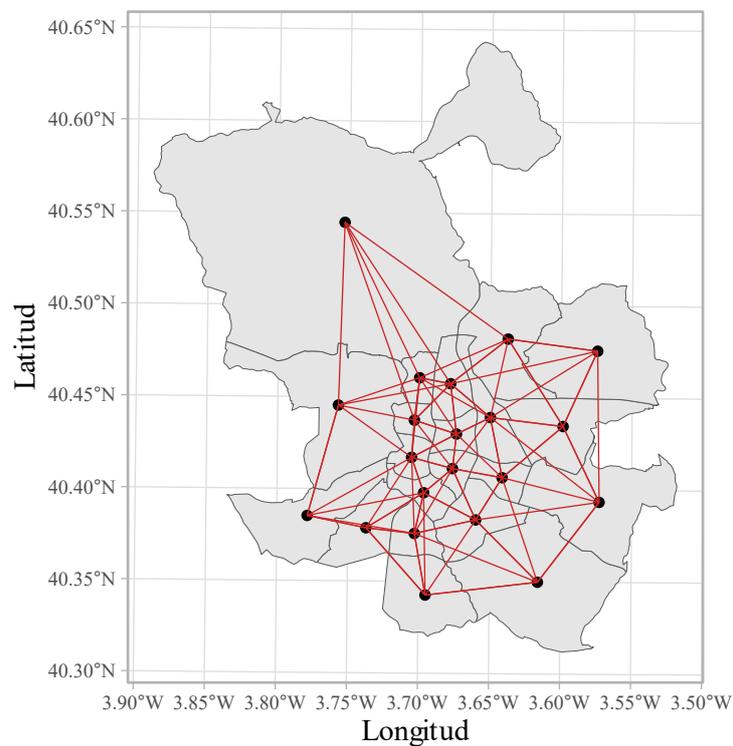


Figura 6. Cartografía de los $k = 5$ vecinos más cercanos

Fuente: Elaboración propia

RESULTADOS

I. Renta de los hogares y tamaño medio del hogar

La renta neta de los hogares en el período 2015-2019 varía entre un mínimo de 24.690 € en Puente de Vallecas (2015), con un tamaño medio del hogar de 2,67 personas, y un máximo de 70.735 € en Chamartín (2018), con 2,50 personas. La renta media es de 41.508,84 €, con una desviación típica de 12.058,20 €, y el tamaño medio del hogar es de 2,51 personas, con una desviación típica de 0,19 personas. El distrito Centro es una excepción pues posee el menor tamaño medio del hogar del municipio, con 1,97 personas.

En el municipio de Madrid la renta neta media ha crecido un 12,60%. Los mayores incrementos han ocurrido en Centro (21,59%), Chamberí (14,44%) y Hortaleza (14,27%). Por su parte, el tamaño medio del hogar ha crecido un 0,80%. Los mayores incrementos se han producido en Carabanchel (3,13%), Villaverde (2,94%) y Usera (2,94%), mientras que los mayores decrementos se han dado en Retiro (-1,22%), Chamberí (-0,89%) y Arganzuela (-0,84%) (Tabla 3, Tabla 4). Visualizamos los resultados con la librería *tmap* (Tennekes, 2018) (Figura 9, Figura 10).

Distrito	Renta neta media por hogar (€)					Variación (%)
	2015	2016	2017	2018	2019	2015-2019
01. Centro	31.392	32.458	33.473	36.072	38.171	21,59
02. Arganzuela	40.203	41.122	42.088	43.789	45.262	12,58
03. Retiro	51.795	52.963	53.111	55.775	57.027	10,10
04. Salamanca	53.978	57.147	57.720	61.466	60.435	11,96
05. Chamartín	60.962	65.260	65.995	70.735	69.558	14,10
06. Tetuán	34.702	35.609	36.142	37.557	39.332	13,34
07. Chamberí	48.997	50.882	51.644	54.327	56.073	14,44
08. Fuencarral-El Pardo	48.859	50.087	50.379	52.858	54.493	11,53
09. Moncloa-Aravaca	56.458	57.423	58.965	64.066	62.482	10,67
10. Latina	30.157	30.490	30.847	31.872	33.237	10,21
11. Carabanchel	27.940	28.301	28.721	29.842	31.190	11,63
12. Usera	25.573	25.954	26.501	27.674	29.061	13,64
13. Puente de Vallecas	24.690	25.029	25.527	26.652	27.975	13,30
14. Moratalaz	34.753	35.059	35.514	36.755	37.951	9,20
15. Ciudad Lineal	36.833	37.421	37.621	39.425	40.948	11,17
16. Hortaleza	47.113	48.688	49.695	52.944	53.836	14,27
17. Villaverde	26.599	26.915	27.392	28.407	29.873	12,31
18. Villa de Vallecas	30.175	30.861	31.447	32.765	34.168	13,23
19. Vicálvaro	31.466	32.152	32.713	33.870	35.510	12,85
20. San Blas-Canillejas	34.498	35.077	35.546	37.112	38.560	11,77
21. Barajas	46.347	47.389	47.782	49.615	51.108	10,27
Madrid	38.539	39.613	40.195	42.283	43.393	12,60

Tabla 3. Renta neta media por hogar (2015-2019)

Se resaltan en color verde los valores del percentil 80

Fuente: Elaboración propia

Distrito	Tamaño medio del hogar (personas)					Variación (%)
	2015	2016	2017	2018	2019	2015-2019
01. Centro	1,98	1,96	1,96	1,96	1,98	0,00
02. Arganzuela	2,37	2,36	2,35	2,35	2,35	-0,84
03. Retiro	2,46	2,45	2,44	2,44	2,43	-1,22
04. Salamanca	2,31	2,30	2,30	2,30	2,30	-0,43
05. Chamartín	2,49	2,48	2,48	2,48	2,48	-0,40
06. Tetuán	2,34	2,34	2,34	2,35	2,36	0,85
07. Chamberí	2,24	2,23	2,22	2,22	2,22	-0,89
08. Fuencarral-El Pardo	2,68	2,67	2,68	2,68	2,69	0,37
09. Moncloa-Aravaca	2,57	2,55	2,56	2,56	2,57	0,00
10. Latina	2,48	2,47	2,47	2,48	2,49	0,40
11. Carabanchel	2,56	2,57	2,58	2,61	2,64	3,13
12. Usera	2,72	2,72	2,73	2,76	2,80	2,94
13. Puente de Vallecas	2,60	2,59	2,60	2,62	2,65	1,92
14. Moratalaz	2,52	2,50	2,50	2,50	2,50	-0,79
15. Ciudad Lineal	2,47	2,46	2,47	2,47	2,48	0,40
16. Hortaleza	2,65	2,65	2,65	2,66	2,67	0,75
17. Villaverde	2,72	2,72	2,74	2,76	2,80	2,94
18. Villa de Vallecas	2,56	2,57	2,57	2,59	2,61	1,95
19. Vicálvaro	2,74	2,73	2,72	2,73	2,73	-0,36
20. San Blas-Canillejas	2,60	2,60	2,60	2,61	2,63	1,15
21. Barajas	2,67	2,67	2,67	2,67	2,67	0,00
Madrid	2,50	2,49	2,49	2,50	2,52	0,80

Tabla 4. Tamaño medio del hogar (2015-2019)

Se resaltan en color rojo los valores del percentil 5 y en verde los del percentil 95

Fuente: Elaboración propia

La renta neta de los hogares presenta una distribución de colas ligeras respecto a una distribución normal (Figura 7). Cuando se ordena la renta a través de un diagrama de caja (*box plot*) se identifican grupos de 4-5 componentes. Tenemos así grupos de renta muy baja (Puente de Vallecas, Usera, Villaverde, Carabanchel), renta baja (Latina, Villa de Vallecas, Vicálvaro, Centro), renta media (Moratalaz, San Blas-Canillejas, Tetuán, Ciudad Lineal, Arganzuela), renta alta (Barajas, Hortaleza, Fuencarral-El Pardo, Chamberí, Retiro) y renta muy alta (Salamanca, Moncloa-Aravaca, Chamartín) (Figura 8).

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID

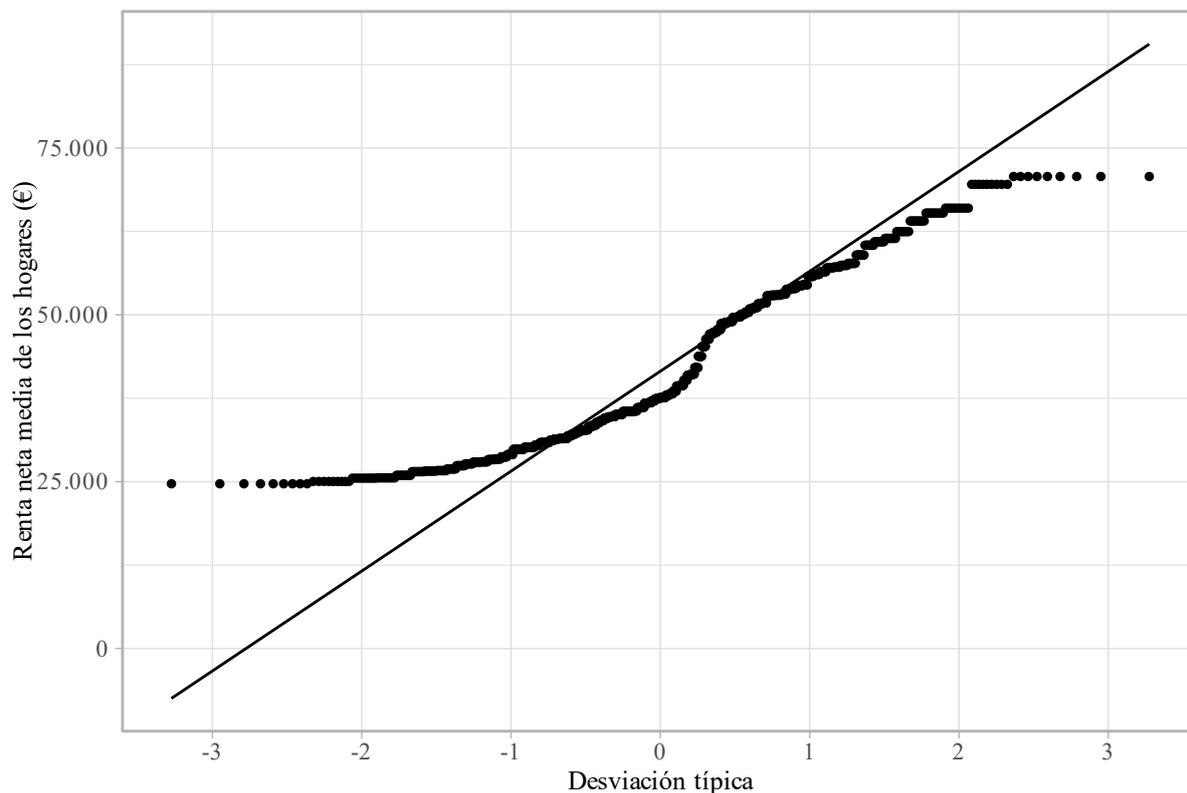


Figura 7. Q-Q plot de la renta neta media de los hogares (2015-2019)

Fuente: Elaboración propia

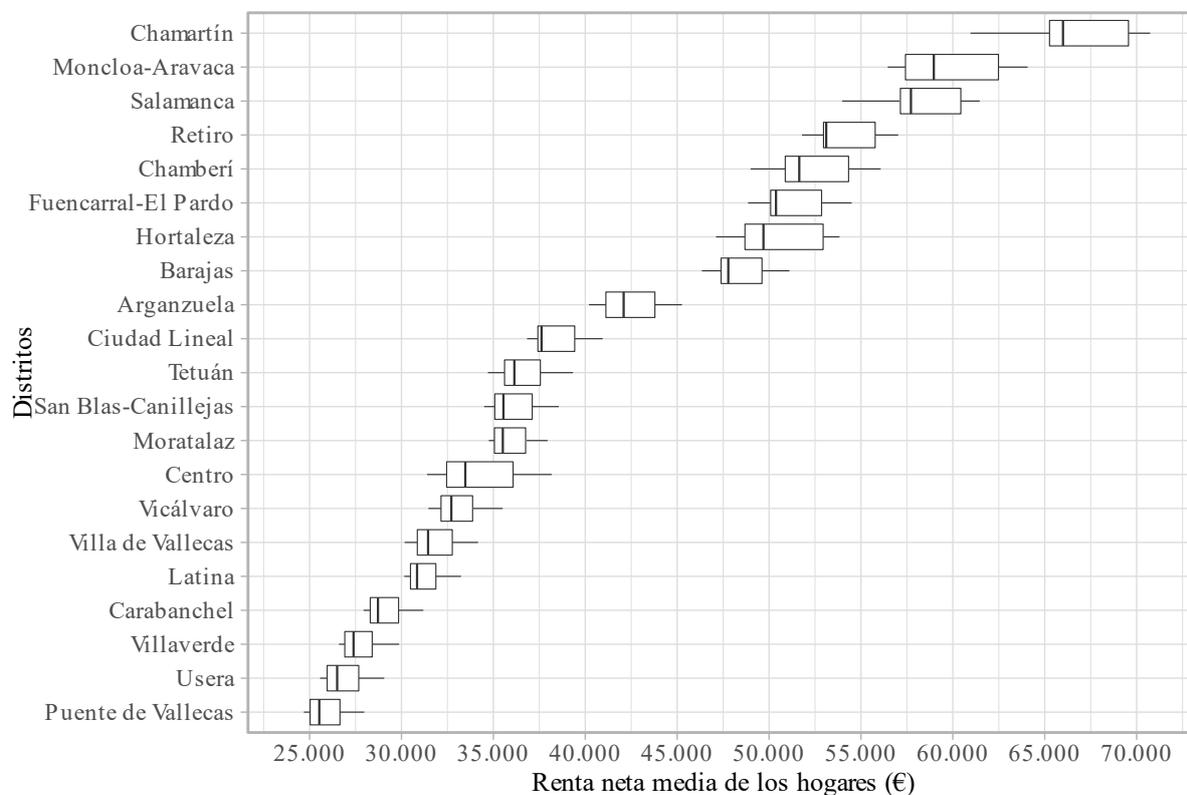


Figura 8. Diagrama de caja de la renta neta media de los hogares por distrito (2015-2019)

Fuente: Elaboración propia

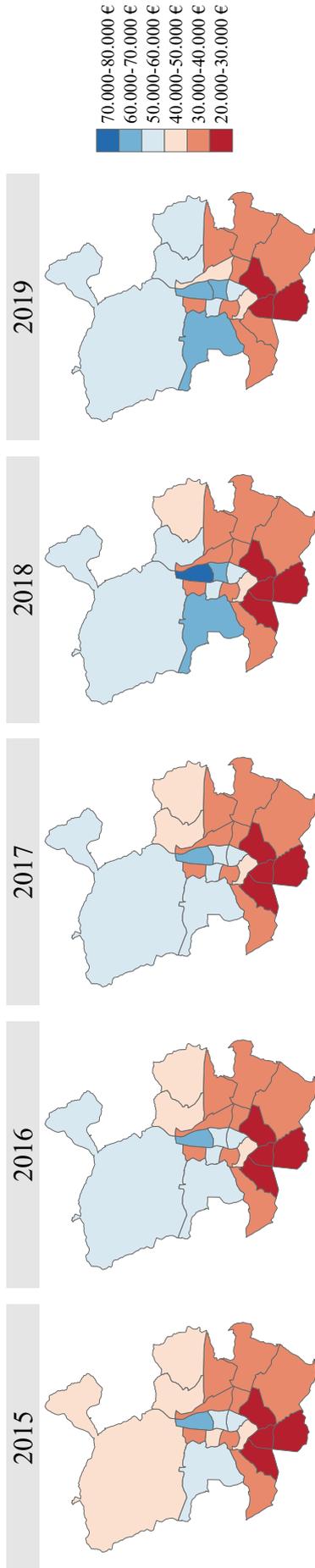


Figura 9. Cartografía de la renta neta media de los hogares por distrito (2015-2019)
Fuente: Elaboración propia

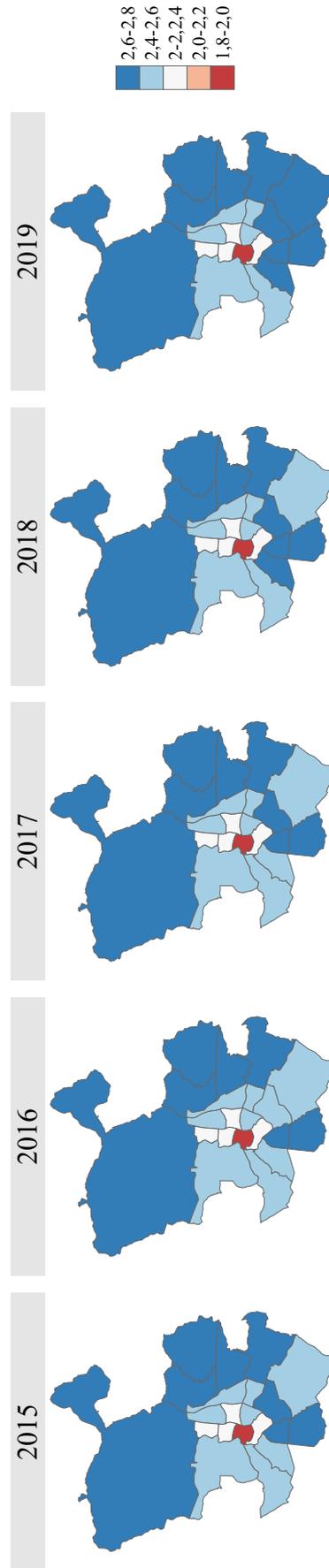


Figura 10. Cartografía del tamaño medio del hogar por distrito (2015-2019)
Fuente: Elaboración propia

II. Tasa de paro y población con estudios superiores

La tasa de paro en el período 2015-2019 varía entre un máximo de 15,34% en Puente de Vallecas (2015) y un mínimo de 4,85% en Chamartín (2019). Mientras que la proporción de personas de ≥ 25 años con estudios superiores varía entre un mínimo de 15,51% en Puente de Vallecas (2015) y un máximo de 61,55% en Chamartín (2019).

En el municipio de Madrid, la tasa de paro ha disminuido de 10,62% a 7,53% (-29,06%), mientras que a nivel nacional disminuyó de 22,06% a 14,11% (-36,05%)¹⁶. Pese a esta reducción, España sigue siendo el segundo país de la UE y la eurozona, por detrás de Grecia, con la mayor tasa de paro¹⁷. Los distritos donde más se ha reducido la tasa de paro son Tetuán (-32,02%), Villaverde (-31,83%) y Ciudad Lineal (-31,34%). En 2019 Puente de Vallecas (10,84%) y Villaverde (10,43%) seguían teniendo tasas superiores al 10%, mientras que Salamanca (4,87%) y Chamartín (4,85%) tenían tasas inferiores al 5%.

La población de ≥ 25 años con estudios superiores ha crecido un 9,83% en el municipio de Madrid, los mayores incrementos han ocurrido en los distritos de Puente de Vallecas (14,31%), Centro (13,10%) y Tetuán (12,84%). En 2019 las mayores tasas se dan en Chamartín (61,55%), Chamberí (61,23%) y Salamanca (60,63%), y las menores en Puente de Vallecas (17,73%), Usera (18,68%) y Villaverde (18,90%) (Tabla 5, Tabla 6). Visualizamos los resultados (Figura 12, Figura 13).

La matriz de correlación (Figura 11), creada con la librería *ggcorrplot* (Kassambara, 2019), muestra que existe una correlación positiva casi perfecta entre la renta neta y los estudios superiores. La tasa de paro presenta una fuerte correlación negativa entre la renta neta y los estudios superiores.

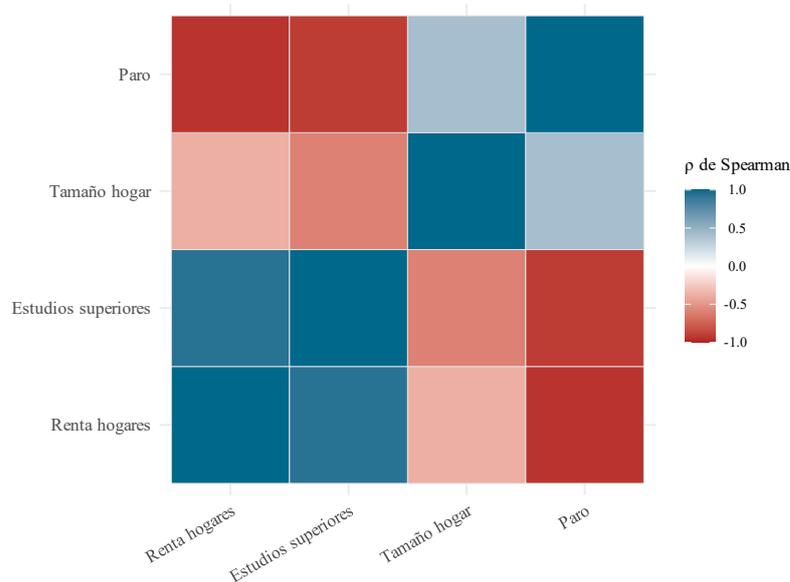


Figura 11. Matriz de correlación
Fuente: Elaboración propia

¹⁶ Cálculo hecho a partir de la Encuesta de Población Activa (INE): <https://bit.ly/3H8vTk3>

¹⁷ <https://bit.ly/3NXcyEA>

Distrito	Tasa absoluta de paro registrado (%)					Variación (%)
	2015	2016	2017	2018	2019	2015-2019
01. Centro	10,09	9,29	8,48	7,77	7,09	29,75
02. Arganzuela	9,03	8,18	7,38	6,64	6,29	30,39
03. Retiro	7,29	6,65	6,17	5,76	5,46	25,04
04. Salamanca	6,85	6,27	5,76	5,23	4,87	28,86
05. Chamartín	6,64	6,18	5,62	5,11	4,85	26,99
06. Tetuán	10,54	9,55	8,66	7,65	7,17	32,02
07. Chamberí	7,21	6,56	5,93	5,47	5,07	29,62
08. Fuencarral-El Pardo	8,05	7,35	6,67	6,06	5,80	27,91
09. Moncloa-Aravaca	7,16	6,58	6,01	5,59	5,24	26,76
10. Latina	12,13	11,15	10,01	9,12	8,49	30,01
11. Carabanchel	13,14	12,06	10,79	9,90	9,30	29,20
12. Usera	13,87	12,48	11,23	10,30	9,78	29,50
13. Puente de Vallecas	15,34	14,07	12,74	11,67	10,84	29,33
14. Moratalaz	11,47	10,37	9,64	8,97	8,28	27,82
15. Ciudad Lineal	10,61	9,63	8,62	7,76	7,28	31,34
16. Hortaleza	8,83	8,09	7,40	6,72	6,29	28,72
17. Villaverde	15,30	13,82	12,46	11,32	10,43	31,83
18. Villa de Vallecas	12,45	11,35	10,47	9,68	9,24	25,77
19. Vicálvaro	12,44	11,29	10,53	9,60	8,95	28,01
20. San Blas-Canillejas	11,50	10,58	9,56	8,54	7,97	30,70
21. Barajas	8,22	7,45	6,73	5,95	5,73	30,29
Madrid	10,62	9,70	8,81	8,03	7,53	-29,06

Tabla 5. Tasa absoluta de paro registrado (2015-2019)

Se resaltan en color verde los valores del percentil 80

Fuente: Elaboración propia

Distrito	Población de ≥25 años con estudios superiores (%)					Variación (%)
	2015	2016	2017	2018	2019	2015-2019
01. Centro	46,82	48,58	49,92	51,18	52,96	13,10
02. Arganzuela	46,49	47,79	49,05	50,47	51,81	11,46
03. Retiro	53,18	54,27	55,31	56,50	57,67	8,45
04. Salamanca	55,34	56,69	57,98	59,40	60,63	9,56
05. Chamartín	56,64	57,81	58,98	60,47	61,55	8,65
06. Tetuán	37,21	38,25	39,36	40,66	41,98	12,84
07. Chamberí	55,61	57,14	58,52	59,92	61,23	10,10
08. Fuencarral-El Pardo	46,00	46,93	47,86	48,92	49,89	8,46
09. Moncloa-Aravaca	51,80	52,77	53,78	54,96	55,91	7,95
10. Latina	25,19	25,73	26,33	26,98	27,79	10,33
11. Carabanchel	22,49	23,04	23,54	24,23	24,94	10,91
12. Usera	17,06	17,31	17,52	17,96	18,68	9,48
13. Puente de Vallecas	15,51	15,89	16,32	16,94	17,73	14,31
14. Moratalaz	30,74	31,27	31,94	32,68	33,64	9,44
15. Ciudad Lineal	34,44	35,01	35,73	36,70	37,82	9,81
16. Hortaleza	41,58	42,79	43,90	45,19	46,46	11,74
17. Villaverde	17,55	17,76	18,14	18,41	18,90	7,71
18. Villa de Vallecas	28,89	29,56	30,09	30,67	31,30	8,35
19. Vicálvaro	27,91	28,31	28,63	29,19	29,96	7,35
20. San Blas-Canillejas	29,20	29,86	30,61	31,33	32,26	10,48
21. Barajas	43,99	44,92	45,60	46,23	47,05	6,94
Madrid	36,38	37,24	38,07	39,00	39,96	9,83

Tabla 6. Población de ≥25 años con estudios superiores (2015-2019)

Se resaltan en color verde los valores del percentil 80

Fuente: Elaboración propia

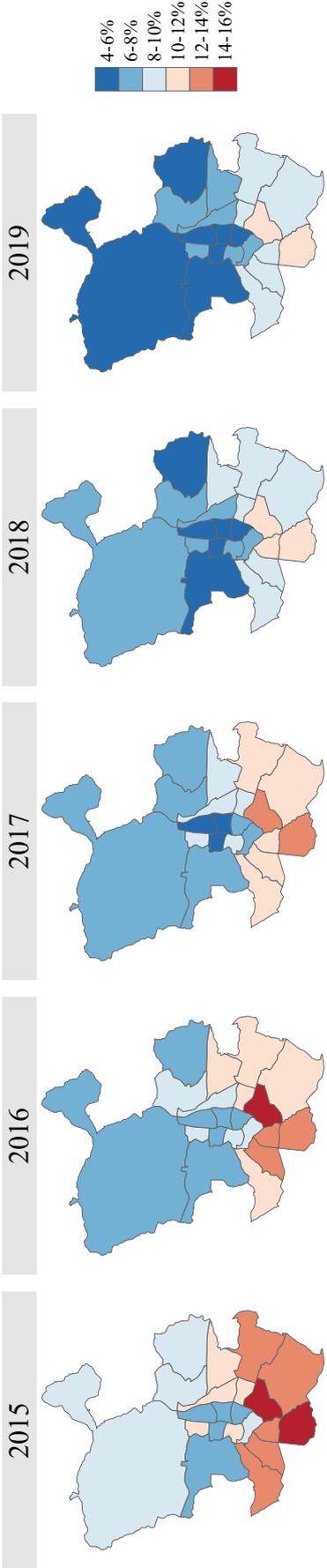


Figura 12. Cartografía de la tasa absoluta de paro registrado por distrito (2015-2019)
Fuente: Elaboración propia

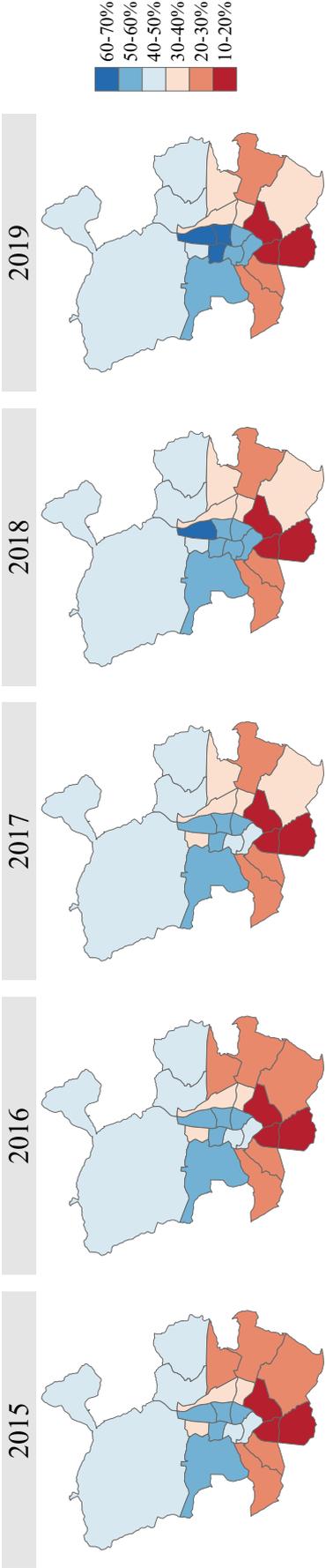


Figura 13. Cartografía de la población de ≥ 25 años con estudios superiores por distrito (2015-2019)
Fuente: Elaboración propia

III. Coeficiente de Gini espacial

La [Tabla 7](#) muestra los resultados del cálculo de la descomposición espacial del coeficiente de Gini.

	2015	2016	2017	2018	2019	2015-2019
G (%)	38,80	38,90	38,60	38,50	38,10	-
G^s	0,1566	0,1625	0,1614	0,1668	0,1553	0,1631
G_n^s	0,0181	0,0190	0,0189	0,0194	0,0184	0,0275
G_{ns}^s	0,1385	0,1435	0,1425	0,1474	0,1369	0,1356
G_n^s (%)	11,56	11,68	11,70	11,64	11,82	16,85
G_{ns}^s (%)	88,44	88,32	88,30	88,36	88,18	83,15
p valor ($M = 99$)	0,01	0,01	0,01	0,01	0,01	0,01
p valor ($M = 999$)	0,001	0,001	0,001	0,001	0,001	0,001

Tabla 7. Coeficiente de Gini espacial (2015-2019)

Fuente: Elaboración propia, ADRH (INE)

G es el coeficiente de Gini estándar, G^s es el coeficiente de Gini espacial, G_n^s es el componente correspondiente a los distritos vecinos y G_{ns}^s es el componente correspondiente a los distritos no vecinos. Se muestran los p valores del test de la simulación de Monte Carlo para $M = 99$ y $M = 999$ simulaciones.

DISCUSIÓN

El municipio de Madrid presenta una brecha socioeconómica que lo divide en una diagonal que va del suroeste al noreste. En el norte, se ubican los distritos con mayor renta, mayor proporción de la población con estudios superiores y menor tasa de paro. Mientras que en el sur sucede justamente lo contrario. Leal & Sorando (2015) obtuvieron resultados similares para el área metropolitana de la Comunidad de Madrid, donde mostraron la relación entre la segregación espacial y el nivel de ocupación de las personas. El informe del Fondo de Reequilibrio Territorial¹⁸ del Ayuntamiento de Madrid igualmente muestra que los distritos más vulnerables son aquellos que tienen una mayor tasa de paro y un menor nivel de renta.

El tamaño medio del hogar ha ido aumentando en los distritos del sureste y suroeste, posiblemente motivado por personas jóvenes e inmigrantes que viven en régimen de alquiler. Que la población con estudios superiores haya aumentado significativamente en los distritos de Puente de Vallecas y Tetuán puede deberse a varios factores. En Puente de Vallecas ha podido incrementarse el número de personas que son la primera generación de su familia que han ido a la universidad, mientras que en Tetuán puede deberse a su ubicación, próxima al centro económico y financiero de la capital, y a los comparativamente menores precios de la vivienda en alquiler.

Recursos valiosos como la información, la influencia y las oportunidades laborales se intercambian a través de relaciones interpersonales. Las redes personales o egocéntricas varían en tamaño y calidad en función de a quién incluyen y qué acceso ofrecen a dichos recursos. Estas redes reproducen la desigualdad socioeconómica cuando las personas que tienen acceso a recursos valiosos impiden que otros accedan a ellos, generándose así redes de exclusividad. Asimismo, una gran parte de la desigualdad de la renta se canaliza a través de factores no elegidos por el individuo, como son el nivel educativo y de ocupación de los padres (Cabrera et al., 2021; van Eijk, 2010).

En las últimas tres décadas ha aumentado considerablemente la proporción de personas con educación superior en los países desarrollados. En Europa, esto ha provocado que la prima salarial por poseer estudios superiores se haya estabilizado o incluso reducido. La estructura del mercado de trabajo en España genera desajustes entre los egresados (oferta) y las empresas (demanda) porque éstas son incapaces de acoger a todos los egresados que las universidades producen. Aquí, el modelo productivo y la estructura empresarial, donde la inmensa mayoría son pymes, juega un papel relevante. España posee la mayor tasa de subempleo de la UE¹⁹ (Crivellaro, 2016; Felgueroso et al., 2010).

Observando la [Tabla 7](#), la desigualdad espacial global en el período 2015-2019 se sitúa en 0,1631. En un contexto espacial, cero implica que los valores están dispersos a lo largo del mapa y uno que están concentrados en un único punto. Por tanto, este valor demuestra que la desigualdad no está concentrada en un único distrito, sino que está repartida entre varios. Concretamente, la descomposición espacial muestra que el componente asociado a los distritos vecinos, G_n^S , representa el 16,85% de la desigualdad global. Entre 2015 y 2019 se ha

¹⁸ <https://bit.ly/3toq3p2>

¹⁹ <https://bit.ly/3mvothq>

incrementado un 2,25%. Su valor es próximo a cero, signo de una elevada autocorrelación espacial positiva. Por su parte, el componente de los distritos no vecinos, G_{ns}^s , representa el 83,15% de la desigualdad global, entre 2015 y 2019 ha disminuido un 0,29%.

La desigualdad se produce entre los miembros de los diferentes grupos. Esto es, los distritos que pertenecen a un grupo son similares entre sí –aunque ha habido un ligero aumento de la desigualdad entre ellos– pero hay una importante diferencia entre grupos (83,15%). Esta diferencia se ha mantenido estable en el tiempo a pesar de que desde 2015 la renta neta de los hogares ha aumentado un 12,60% y la tasa de paro ha descendido un 29,06% de media.

En el municipio de Madrid, el coeficiente de Gini estándar ha disminuido un 1,80%, situándose en 2019 en 38,10%, mientras que a nivel nacional disminuyó de 34,60% a 33% (-4,62%)²⁰. Pese a esta reducción, el valor sigue siendo alto.

Los resultados del test para medir la autocorrelación espacial son estadísticamente significativos tanto para $M = 99$ simulaciones ($\alpha = 0,05$) como para $M = 999$ simulaciones ($\alpha = 0,01$). En ambos casos hay evidencia significativa para rechazar la hipótesis nula, lo que implica que la renta de los hogares está correlacionada espacialmente.

²⁰ <https://bit.ly/3xyTiYE>

CONCLUSIONES

En este novedoso trabajo se han combinado tres disciplinas diferentes –economía, geografía y ciencia de datos– para comprobar cómo ha variado espacialmente la desigualdad de la renta en el municipio de Madrid a nivel de distritos en el período 2015-2019. Esto se ha realizado con el lenguaje de programación R, usando los principios de datos limpios.

Los resultados obtenidos a través del análisis exploratorio de datos muestran que una brecha socioeconómica divide al municipio de Madrid en una diagonal que va del suroeste al noreste. En el norte, se ubican los distritos con mayor renta, mayor proporción de la población con estudios superiores y menor tasa de paro. Mientras que en el sur sucede justamente lo contrario. La descomposición espacial del coeficiente de Gini demuestra que la renta de los hogares está correlacionada espacialmente y que la mayoría de la desigualdad se explica por el componente de los distritos no vecinos. Esto es, los distritos que pertenecen a un grupo son similares entre sí, pero hay una importante diferencia entre grupos. Estas diferencias se han mantenido estables en el tiempo.

Se ha utilizado el algoritmo kNN para calcular los k vecinos más cercanos de cada distrito y crear así grupos. Los pesos espaciales binarios determinan a qué grupo pertenece cada distrito. Usar otro tipo de sistema de ponderación y/o medida de distancia ofrecerá resultados diferentes. Asimismo, podría ser de interés generar los k vecinos más cercanos utilizando una técnica de remuestreo.

Este trabajo se podría extender (1) realizando un análisis de regresión entre la renta de los hogares, la tasa de paro y el nivel educativo, (2) usando un nivel administrativo inferior (barrios) o (3) incluyendo variables sobre el mercado inmobiliario, tanto en compra como en alquiler.

La desigualdad de la renta lleva asociada la desigualdad de riqueza, que es particularmente difícil de medir, la cual impide a un número significativo de la población encontrar vivienda tanto en Madrid como en otras ciudades y desarrollar así su proyecto vital.

BIBLIOGRAFÍA

- Allison, P. D. (1978). Measures of Inequality. *American Sociological Review*, 43(6), 865-880. <https://doi.org/10.2307/2094626>
- Boehmke, B., & Greenwell, B. (2019). *Hands-On Machine Learning with R*. CRC Press.
- Cabrera, L., Marrero, G. A., Rodríguez, J. G., & Salas-Rojo, P. (2021). Inequality of Opportunity in Spain: New Insights from New Data. *Hacienda Pública Española / Review of Public Economics*, 237(2), 153-183. <https://hpe-rpe.org/ief/210/237-2-2021/4793/inequality-of-opportunity-in-spain-new-insights-from-new-data-2.pdf>
- Crawley, M. J. (2012). *The R Book* (2.^a ed.). John Wiley & Sons.
- Crivellaro, E. (2016). The College Wage Premium over Time: Trends in Europe in the Last 15 Years. En L. Cappellari, S. W. Polachek, & K. Tatsiramos (Eds.), *Inequality: Causes and Consequences* (Vol. 43, pp. 287-328). Emerald Group Publishing. <https://doi.org/10.1108/S0147-912120160000043016>
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
- Deininger, K., & Squire, L. (1996). A New Data Set Measuring Income Inequality. *The World Bank Economic Review*, 10(3), 565-591. <https://doi.org/10.1093/wber/10.3.565>
- España, Cortes Generales. (1978). Constitución Española. *Boletín Oficial del Estado*, 311. [https://www.boe.es/eli/es/c/1978/12/27/\(1\)/con](https://www.boe.es/eli/es/c/1978/12/27/(1)/con)
- España, Jefatura del Estado. (1985). Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General. *Boletín Oficial del Estado*, 147. <https://www.boe.es/eli/es/lo/1985/06/19/5/con>
- Felgueroso, F., Hidalgo, M., & Jiménez Martín, S. (2010). Explaining the fall of the skill wage premium in Spain. *Documentos de Trabajo (FEDEA)*, 19-2010. <http://documentos.fedea.net/pubs/dt/2010/dt-2010-19.pdf>
- Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3), 238-247. <https://doi.org/10.2307/1403797>
- Garmonsway, D. (2020). *tidyxl: Read Untidy Excel Files*. <https://CRAN.R-project.org/package=tidyxl>
- Garmonsway, D. (2021). *unpivotr: Unpivot Complex and Irregular Data Layouts*. <https://CRAN.R-project.org/package=unpivotr>

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID

- Gini, C. (1912). *Variabilità e mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche*. Tipografia di Paolo Cuppini.
- Giorgi, G. M. (2011). The Gini Inequality Index Decomposition: An Evolutionary Study. En J. Deutsch & J. Silber (Eds.), *The Measurement of Individual Well-Being and Group Inequalities: Essays in memory of Z. M. Berrebi* (pp. 185-218). Routledge.
- Haining, R., & Li, G. (2020). *Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach*. CRC Press.
- Heshmati, A. (2006). The World Distribution of Income and Income Inequality: A Review of the Economics Literature. *Journal of World-Systems Research*, 12(1), 61-107. <https://doi.org/10.5195/jwsr.2006.372>
- Kalogirou, S. (2020). *lctools: Local Correlation, Spatial Inequalities, Geographically Weighted Regression and Other Tools*. <https://CRAN.R-project.org/package=lctools>
- Kassambara, A. (2019). *ggcorrplot: Visualization of a Correlation Matrix using «ggplot2»*. <https://CRAN.R-project.org/package=ggcorrplot>
- Leal, J., & Sorando, D. (2015). Economic crisis, social change and segregation processes in Madrid. En T. Tammaru, S. Marcińczak, M. van Ham, & S. Musterd (Eds.), *Socio-Economic Segregation in European Capital Cities* (pp. 214-237). Routledge.
- Llop, J. (2018, abril). *La línea de la desigualdad*. JotaData. <https://jllopc.github.io/linea-de-la-desigualdad/>
- Lo Prete, A. (2018). Inequality and the finance you know: Does economic literacy matter? *Economia Politica*, 35(1), 183-205. <https://doi.org/10.1007/s40888-018-0097-3>
- Lorenz, M. O. (1905). Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association*, 9(70), 209-219. <https://doi.org/10.2307/2276207>
- Lovelace, R., Nowosad, J., & Muenchow, J. (2019). *Geocomputation with R*. CRC Press.
- Madrid, Ayuntamiento. (2017a). Acuerdo de 31 de mayo de 2017 del Pleno por el que se aprueba la creación del barrio denominado «Ensanche de Vallecas» en el Distrito de Villa de Vallecas y la modificación del Reglamento Orgánico de los Distritos de la Ciudad de Madrid. *Boletín Oficial del Ayuntamiento de Madrid*, 7927.
- Madrid, Ayuntamiento. (2017b). Acuerdo de 31 de octubre de 2017 del Pleno del Ayuntamiento de Madrid por el que se aprueba la creación de los barrios «Casco Histórico de Vicálvaro», «Valdebernardo», «Valderrivas» y «El Cañaveral» y la delimitación territorial interna actualizada del Distrito de Vicálvaro, así como el cambio de denominación del «Barrio de San Andrés» por el de «Villaverde Alto, Casco Histórico de Villaverde» y la delimitación territorial interna actualizada del Distrito de Villaverde,

- junto con la correspondiente modificación del Reglamento Orgánico de los Distritos de la Ciudad de Madrid. *Boletín Oficial del Ayuntamiento de Madrid*, 8034.
- Madrid, Ayuntamiento. (2021). Acuerdo de 29 de junio de 2021 del Pleno del Ayuntamiento de Madrid por el que se aprueba la modificación de los límites territoriales entre los Distritos de Hortaleza y Barajas. *Boletín Oficial del Ayuntamiento de Madrid*, 8928.
- McLennan, D., Noble, M., & Wright, G. (2016). Developing a spatial measure of exposure to socio-economic inequality in South Africa. *South African Geographical Journal*, 98(2), 254-274. <https://doi.org/10.1080/03736245.2015.1028980>
- Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17-23. <https://doi.org/10.2307/2332142>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439. <https://doi.org/10.32614/RJ-2018-009>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rey, S. J., & Smith, R. J. (2013). A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences*, 6(2), 55-70. <https://doi.org/10.1007/s12076-012-0086-z>
- Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press.
- Salverda, W., Nolan, B., Checchi, D., Marx, I., McKnight, A., Tóth, I. G., & van de Werfhorst, H. (Eds.). (2014). *Changing Inequalities in Rich Countries: Analytical and Comparative Perspectives*. Oxford University Press.
- Schabenberger, O., & Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. CRC Press.
- Slowikowski, K. (2021). *ggrepel: Automatically Position Non-Overlapping Text Labels with «ggplot2»*. <https://CRAN.R-project.org/package=ggrepel>
- Tammaru, T., Marcińczak, S., van Ham, M., & Musterd, S. (Eds.). (2015). *Socio-Economic Segregation in European Capital Cities*. Routledge.
- Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software*, 84(6), 1-39. <https://doi.org/10.18637/jss.v084.i06>
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- van Eijk, G. (2010). *Unequal networks: Spatial segregation, relationships and inequality in the city*. IOS Press. <https://repository.tudelft.nl/islandora/object/uuid%3A38a9a18d-5ac1-4f94-9ee4-02e48f364f79>

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID

- Vono de Vilhena, D., & Bayona i Carrasco, J. (2010). El asentamiento residencial de los latinoamericanos en las principales ciudades españolas (2001-2009). *Notas de Población*, 37(91), 129-159. <https://repositorio.cepal.org/handle/11362/12875>
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10). <https://doi.org/10.18637/jss.v059.i10>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2.^a ed.). Springer. <https://ggplot2-book.org/>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

ANEXO. LIBRERÍAS *TIDYXL* Y *UNPIVOTR*

Para importar datos no tabulares de Microsoft Excel en R hay que usar las librerías *tidyxl* y *unpivotr* de Duncan Garmonsway²¹.

La función *behead* de *unpivotr* se emplea para extraer los datos de las filas que sirven como encabezados una por una. Hecho esto, se puede pivotar hacia otra zona para extraer los datos de las columnas. Si dos o más celdas están combinadas, los datos sólo existen en una de ellas; la del extremo izquierdo, las demás están vacías. Para tratar esto, se emplea la dirección *up-left* en *behead*.

Veamos cómo usarlas con el archivo “renta neta media anual de los hogares.xlsx” que contiene doble encabezado y celdas combinadas. Tomando como ejemplo el rango E1:I1, “Renta neta media por hogar (€)”, se ha hecho lo siguiente:

1. El rango está combinado, estando el valor únicamente en E1. Para extraer esta fila, tomamos como referencia la celda I5 y vamos hacia arriba y a la izquierda; *behead(up-left)*.

	A	B	C	D	E	F	G	H	I
1					Renta neta media por hogar (€)				
2	CCAA	MUN	CDIS	NDIS	2015	2016	2017	2018	2019
3	Comunidad de Madrid	Madrid			38.539	39.613	40.195	42.283	43.393
4	Comunidad de Madrid	Madrid	01	Centro	31.392	32.458	33.473	36.072	38.171
5	Comunidad de Madrid	Madrid	02	Arganzuela	40.203	41.122	42.088	43.789	45.262
6	Comunidad de Madrid	Madrid	03	Retiro	51.795	52.963	53.111	55.775	57.027
7	Comunidad de Madrid	Madrid	04	Salamanca	53.978	57.147	57.720	61.466	60.435
8	Comunidad de Madrid	Madrid	05	Chamartín	60.962	65.260	65.995	70.735	69.558
9	Comunidad de Madrid	Madrid	06	Tetuán	34.702	35.609	36.142	37.557	39.332
10	Comunidad de Madrid	Madrid	07	Chamberí	48.997	50.882	51.644	54.327	56.073
11	Comunidad de Madrid	Madrid	08	Fuencarral-El Pardo	48.859	50.087	50.379	52.858	54.493
12	Comunidad de Madrid	Madrid	09	Moncloa-Aravaca	56.458	57.423	58.965	64.066	62.482
13	Comunidad de Madrid	Madrid	10	Latina	30.157	30.490	30.847	31.872	33.237

²¹ Véase el tutorial que ha elaborado el autor: <https://bit.ly/3Q8ox4c>

DESIGUALDAD ESPACIAL DE LA RENTA EN MADRID

2. La fila 2 pasa a ser la fila 1. Para extraer los años vamos hacia arriba; behead(up).

	A	B	C	D	E	F	G	H	I
1	CCAA	MUN	CDIS	NDIS	2015	2016	2017	2018	2019
2	Comunidad de Madrid	Madrid			38.539	39.613	40.195	42.283	43.393
3	Comunidad de Madrid	Madrid	01	Centro	31.392	32.458	33.473	36.072	38.171
4	Comunidad de Madrid	Madrid	02	Arganzuela	40.203	41.122	42.088	43.789	45.262
5	Comunidad de Madrid	Madrid	03	Retiro	51.795	52.963	53.111	55.775	57.027
6	Comunidad de Madrid	Madrid	04	Salamanca	53.978	57.147	57.720	61.466	60.435
7	Comunidad de Madrid	Madrid	05	Chamartín	60.962	65.260	65.995	70.735	69.558
8	Comunidad de Madrid	Madrid	06	Tetuán	34.702	35.609	36.142	37.557	39.332
9	Comunidad de Madrid	Madrid	07	Chamberí	48.997	50.882	51.644	54.327	56.073
10	Comunidad de Madrid	Madrid	08	Fuencarral-El Pardo	48.859	50.087	50.379	52.858	54.493
11	Comunidad de Madrid	Madrid	09	Moncloa-Aravaca	56.458	57.423	58.965	64.066	62.482
12	Comunidad de Madrid	Madrid	10	Latina	30.157	30.490	30.847	31.872	33.237

3. Para extraer las columnas A, B, C y D vamos a la izquierda; behead(left).
Repetimos la operación cuatro veces.

	A	B	C	D	E	F	G	H	I
1	CCAA	MUN	CDIS	NDIS	2015	2016	2017	2018	2019
2	Comunidad de Madrid	Madrid			38.539	39.613	40.195	42.283	43.393
3	Comunidad de Madrid	Madrid	01	Centro	31.392	32.458	33.473	36.072	38.171
4	Comunidad de Madrid	Madrid	02	Arganzuela	40.203	41.122	42.088	43.789	45.262
5	Comunidad de Madrid	Madrid	03	Retiro	51.795	52.963	53.111	55.775	57.027
6	Comunidad de Madrid	Madrid	04	Salamanca	53.978	57.147	57.720	61.466	60.435
7	Comunidad de Madrid	Madrid	05	Chamartín	60.962	65.260	65.995	70.735	69.558
8	Comunidad de Madrid	Madrid	06	Tetuán	34.702	35.609	36.142	37.557	39.332
9	Comunidad de Madrid	Madrid	07	Chamberí	48.997	50.882	51.644	54.327	56.073
10	Comunidad de Madrid	Madrid	08	Fuencarral-El Pardo	48.859	50.087	50.379	52.858	54.493
11	Comunidad de Madrid	Madrid	09	Moncloa-Aravaca	56.458	57.423	58.965	64.066	62.482
12	Comunidad de Madrid	Madrid	10	Latina	30.157	30.490	30.847	31.872	33.237