

TRABAJO DE FIN DE GRADO

GRADO EN ESTADÍSTICA EMPRESARIAL
FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS
UNIVERSIDAD MIGUEL HERNÁNDEZ



EVOLUCION ESTADÍSTICA DE SEPARACIONES, NULIDADES Y DIVORCIOS EN LOS AÑOS 2013-2020

Autor: Jorge Pliego Marín

Tutora: M^a Asunción Martínez Mayoral

Curso: 2022/2023

ÍNDICE

1. CONTEXTO.....	6
2. OBJETIVOS	7
2.1. Información disponible.....	8
3. METODOLOGÍA.....	9
3.1. Procesamiento de los datos.....	9
3.2. Análisis geoespacial.....	13
3.2.1. Tablas geográficas.....	13
3.2.2. Superficies	13
3.3. Agrupación espacial.....	14
3.3.1. Spatial Lag.....	15
3.3.2. Caso Binario	15
3.3.3. Análisis de recuento de uniones	15
3.3.4. Gráficos de densidad del Índice de Moran	17
3.3.5. Moran Scatterplot y puntos fríos y calientes	18
3.4. Software y hardware	19
3.4.1. Librerías	20
4. RESULTADOS	25
4.1. Análisis descriptivo.....	25
4.1.1. Gráficos del total de la tasa de nulidades desde 2013-2020	25
4.1.2. Gráficos de la tasa de divorcios desde 2013-2020.....	27
4.1.3. Gráficos de la tasa de separaciones desde 2013-2020.....	29
4.1.4. Gráficos del total de la tasa de la suma de nulidades, separaciones y divorcios desde 2013-2020.....	31
4.2. Agrupación espacial.....	33

4.2.1. Spatial Lag	34
4.2.2. Caso Binario.....	35
4.2.3. Análisis join counts.....	37
4.2.4. Gráficos de densidad del Índice de Moran.....	39
4.2.5. Moran Scatterplot y puntos calientes y fríos	41
5. CONCLUSIONES.....	48
6. BIBLIOGRAFÍA.....	50
7. ANEXOS.....	52



RESUMEN

En este Trabajo de Fin de Grado se aborda el estudio de la evolución de las separaciones, nulidades y divorcios en España desde 2013 hasta 2020, con el propósito de analizar posibles cambios ocurridos durante este período. Para llevar a cabo este análisis, se emplea una combinación de técnicas descriptivas y análisis de agrupación espacial.

En primer lugar, resulta fundamental realizar una depuración exhaustiva de los datos. Para ello, se accede a la base de datos del Instituto Nacional de Estadística (INE) a través de una API, lo que permite obtener los datos de manera directa y actualizada. Posteriormente, se inicia el proceso de depuración, el cual implica revisar y limpiar los datos para asegurar su calidad y consistencia, este proceso es bastante complejo y largo debido a que se necesitan conocer bastantes aspectos de Python e ir modificando los datos según se necesiten para ir cuadrándolos para su posterior estudio. Además, se presta especial atención a la variable de geometría. En este sentido, se realiza un estudio previo de los conceptos relacionados con la representación espacial, ya que se graficará el mapa de España con los datos relevantes. Este análisis incluye considerar aspectos como la proyección cartográfica adecuada y la comprensión de los sistemas de coordenadas geográficas.

En cuanto al análisis descriptivo, se emplea el entorno de Google Colaboratory, utilizando el lenguaje de programación Python. Dado que Python es ampliamente utilizado en la carrera, se cuenta con experiencia previa en su manejo, lo cual facilita la implementación de los análisis requeridos. En este análisis se realiza un gráfico de la evolución de cada año de las nulidades, separaciones y divorcios por provincias por separado y la suma de los tres aspectos juntos para poder sacar las conclusiones oportunas y ver la distinción entre años y aspectos importantes que destacar. Dentro del análisis descriptivo, se lleva a cabo la creación de un gráfico interactivo. Sin embargo, debido a la incompatibilidad de su interactividad con los formatos tradicionales como Word o PDF, se requiere enlazarlo a una página llamada Netlify mediante un repositorio

en GitHub. Esto permite mostrar el gráfico en su totalidad y aprovechar su interactividad para explorar y visualizar los datos de manera dinámica.

En el análisis de agrupación espacial, se utilizan medidas de vecindad y uniones para identificar patrones espaciales de agrupación. La medida de vecindad se refiere a la relación entre las unidades espaciales, como provincias o municipios, y se basa en la proximidad geográfica. Se establece un criterio para determinar si dos unidades son vecinas, como compartir una frontera o estar a una distancia determinada. Las uniones se refieren a las conexiones entre unidades vecinas, lo que permite identificar grupos de unidades que están más relacionadas entre sí que con el resto del territorio. Estas medidas de vecindad y uniones son útiles para detectar agrupaciones espaciales y evaluar la influencia de la ubicación geográfica en los eventos de separación, nulidad y divorcio.

Para evaluar la autocorrelación espacial, se utiliza el índice de Moran. Este índice mide la similitud espacial entre los valores de una variable en diferentes unidades espaciales. Permite determinar si los valores similares tienden a agruparse en el espacio (autocorrelación positiva) o si los valores disímiles están dispersos (autocorrelación negativa). El índice de Moran calcula la correlación espacial al considerar tanto los valores de la variable en cada unidad espacial como la relación espacial entre las unidades. Proporciona un valor que varía entre -1 y 1, donde valores cercanos a 1 indican una alta autocorrelación espacial positiva y valores cercanos a -1 indican una alta autocorrelación espacial negativa. Este análisis es útil para comprender la estructura espacial de los eventos de separación, nulidad y divorcio, y permite identificar patrones espaciales significativos.

Además, se realiza una regresión lineal espacial para examinar las relaciones espaciales entre las variables de interés. La regresión lineal espacial incorpora la influencia de la ubicación geográfica en el modelo de regresión. Esto implica considerar los efectos espaciales y la dependencia entre las unidades espaciales. Al incluir la información espacial en el análisis, se pueden obtener estimaciones más precisas y se pueden identificar patrones espaciales específicos que no se capturarían en un modelo de regresión lineal tradicional. Esta técnica es valiosa para comprender las relaciones espaciales y la influencia geográfica en los eventos de separación, nulidad y divorcio.

Finalmente, se identifican puntos calientes y puntos fríos mediante el análisis de agrupación espacial local. Esto permite detectar áreas con alta concentración (hotspots) o baja concentración (coldspots) de eventos de separación, nulidad o divorcio.



1. CONTEXTO

El estudio de las nulidades, separaciones y divorcios en España ha adquirido una gran relevancia en los últimos años debido a los cambios y desafíos que ha experimentado la institución matrimonial. De hecho, permite comprender las dinámicas sociales y legales que influyen en las relaciones de pareja y en las decisiones de ruptura.

Una de las tendencias que se han observado en el período analizado es el impacto de la pandemia de COVID-19 en las tasas de separaciones y divorcios. Durante el año 2020, se produjo una notable disminución en los casos de divorcio, separación y nulidad, siendo esta la mayor caída registrada en las últimas dos décadas. Según se informa en un artículo del periódico El País, la crisis sanitaria y las medidas de confinamiento fueron factores determinantes en esta disminución, ya que generaron una convivencia forzada y tensiones adicionales en las parejas, pero también dificultaron el acceso a los trámites y servicios legales relacionados con las rupturas conyugales.



Figura 1. Gráfico sobre la evolución de los divorcios desde 2013 a 2021. (1) INE - Instituto Nacional de Estadística. (s. f.). Nulidades, separaciones y divorcios por provincias. Serie desde 2013(20173). INE.

Sin embargo, es importante señalar que este descenso en las tasas de separaciones y divorcios no debe ser interpretado como una disminución en los

problemas y desafíos que enfrentan las parejas. Por el contrario, la pandemia ha puesto de manifiesto nuevas problemáticas y tensiones que pueden afectar a la estabilidad de las relaciones matrimoniales, como el estrés económico, las dificultades en la conciliación entre el trabajo y la vida personal, y los cambios en los roles y responsabilidades dentro del hogar.

Además de la influencia de la pandemia, otro aspecto relevante a considerar es el impacto de los avances tecnológicos en las relaciones de pareja. El auge de las redes sociales y las aplicaciones de citas en línea ha modificado la forma en que las personas establecen vínculos y se conocen, lo que puede tener implicaciones en la duración y estabilidad de las relaciones matrimoniales. En particular, la popularidad de plataformas como Tinder, fundada en 2012, podría haber influido en la forma en que las parejas se forman y, potencialmente, en las tasas de separaciones y divorcios en los últimos años.

En este contexto, el análisis de los datos de nulidades, separaciones y divorcios obtenidos del Instituto Nacional de Estadística (INE) se vuelve fundamental para comprender la evolución de estas tendencias, identificar posibles factores desencadenantes y plantear reflexiones sobre los desafíos que enfrentan las parejas en la sociedad actual. A través de este estudio, se espera obtener una visión más completa de las dinámicas matrimoniales en España, con el objetivo de brindar información relevante para la toma de decisiones legales, sociales y de políticas públicas en relación con las relaciones de pareja y las rupturas conyugales.

2. OBJETIVOS

El objetivo básico de este trabajo de investigación se centra en el estudio de la distribución territorial y la evolución de nulidades, separaciones y divorcios en España durante el período comprendido entre 2013 y 2020.

Además, se plantea como objetivo secundario el acceso y procesamiento de los datos desde su origen en la web del [INE](#), y la conjugación con otras bases de datos relativas al tamaño de la población y la representación territorial de los

mismos por provincias y comunidades autónomas. El análisis se resuelve utilizando técnicas de estadística espacial en el entorno de programación de Python, y en concreto con los cuadernos Jupyter de Google Colab. Como objetivos específicos, destacamos los siguientes:

- Estudiar la distribución territorial y evolución de las nulidades
- Estudiar la distribución territorial y evolución de los divorcios
- Estudiar la distribución territorial y evolución de las separaciones
- Estudiar la distribución territorial y evolución del total de nulidades, divorcios y separaciones.

2.1. Información disponible

La base de datos usada en este trabajo trata de las nulidades, separaciones y divorcios por provincias desde 2013. Se empieza definiendo los 3 conceptos de esta base y después su distribución.

- Por **nulidad** de matrimonio se entiende la invalidez del matrimonio, por concurrir en su celebración alguno de los vicios de nulidad que se establecen en el [Código Civil](#) (principalmente en relación a los vicios del consentimiento o la capacidad de los contrayentes).
- La **separación** matrimonial se entiende como aquella situación jurídica en el que se mantiene el vínculo conyugal, pero se produce un cese de la convivencia.
- El **divorcio** supone la disolución por sentencia del matrimonio y con ellos el cese definitivo de la convivencia conyugal y del régimen económico matrimonial. Se encuentra regulado en los artículos [85-88](#) del [Código Civil](#).

Tras haber visto el significado de los conceptos, explicamos la distribución de la tabla.

La tabla consta de 53 filas y 36 columnas, en las filas se encuentran las provincias y el total nacional y en las columnas cada año desde el 2021 al 2013 del total, de las separaciones, de los divorcios y de las nulidades

	2021	2020	2019	2018	Total	2016	2015	2014	2013
Total Nacional	90.582 ¹	80.015 ¹	95.320 ¹	99.444 ¹	102.341 ¹	101.294 ¹	101.357 ¹	105.893 ¹	100.437 ¹
02 Albacete	722 ¹	398 ¹	605 ¹	664 ¹	760 ¹	754 ¹	860 ¹	681 ¹	685 ¹
03 Alicante/Alacant	3.595 ¹	3.346 ¹	3.707 ¹	3.973 ¹	4.046 ¹	4.262 ¹	3.913 ¹	4.242 ¹	3.834 ¹
04 Almería	1.716 ¹	1.322 ¹	1.721 ¹	1.485 ¹	1.491 ¹	1.470 ¹	1.458 ¹	1.373 ¹	1.331 ¹
01 Araba/Álava	546 ¹	465 ¹	520 ¹	547 ¹	587 ¹	678 ¹	565 ¹	607 ¹	585 ¹
33 Asturias	1.956 ¹	1.784 ¹	2.182 ¹	2.186 ¹	2.363 ¹	2.408 ¹	2.389 ¹	2.607 ¹	2.361 ¹
05 Ávila	200 ¹	154 ¹	227 ¹	243 ¹	241 ¹	222 ¹	182 ¹	218 ¹	217 ¹
06 Badajoz	1.042 ¹	1.022 ¹	1.230 ¹	1.216 ¹	1.221 ¹	1.208 ¹	1.271 ¹	1.304 ¹	1.050 ¹
07 Balears, Illes	2.401 ¹	2.248 ¹	2.523 ¹	2.596 ¹	2.694 ¹	2.539 ¹	2.503 ¹	2.510 ¹	2.500 ¹
08 Barcelona	11.784 ¹	10.388 ¹	13.016 ¹	13.423 ¹	13.597 ¹	14.016 ¹	14.236 ¹	14.497 ¹	13.847 ¹
48 Bizkaia	1.934 ¹	2.003 ¹	2.542 ¹	2.245 ¹	2.226 ¹	2.133 ¹	2.164 ¹	2.133 ¹	1.971 ¹
09 Burgos	411 ¹	437 ¹	470 ¹	568 ¹	580 ¹	564 ¹	628 ¹	626 ¹	555 ¹
10 Cáceres	635 ¹	515 ¹	626 ¹	716 ¹	719 ¹	683 ¹	707 ¹	694 ¹	523 ¹
11 Cádiz	2.451 ¹	2.048 ¹	2.474 ¹	2.631 ¹	2.569 ¹	2.415 ¹	2.494 ¹	2.657 ¹	2.523 ¹
39 Cantabria	1.114 ¹	1.027 ¹	1.230 ¹	1.236 ¹	1.341 ¹	1.231 ¹	1.375 ¹	1.303 ¹	1.193 ¹

Figura 2. Captura de la tabla de nulidades, separaciones y divorcios por provincias visualizada desde la página web del INE. (2) INE - Instituto Nacional de Estadística. (s. f.-b). Nulidades, separaciones y divorcios por provincias. Serie desde 2013(20173).

3. METODOLOGÍA

En este apartado se va a explicar el procedimiento que se ha seguido para la elaboración de este proyecto. Se va a usar un lenguaje de programación llamado Python para realizar este estudio de estadística espacial, ya que un gran beneficio es la abstracción: complejidades, particularidades y peculiaridades asociadas con cada formato de archivo se eliminan. Python representa todos los datos de forma estándar.

3.1. Procesamiento de los datos

El procesamiento de datos se inicia mediante la obtención de una tabla denominada 'datos', la cual almacena la base de datos del Instituto Nacional de Estadística (INE). Utilizando una interfaz de programación de aplicaciones (API), se accede a los archivos directamente desde la página web del INE y se transforman y estructuran de manera adecuada. Además, se dispone de otra tabla llamada 'datosmunicipios', la cual contiene la información sobre la

población total de cada provincia en cada año que se va a estudiar. Asimismo, se cuenta con una tabla adicional llamada 'countries', que contiene la información geométrica necesaria para representar el mapa de España.

Con el objetivo de representar los datos del INE en el mapa de España y obtener resultados significativos, se llevan a cabo los siguientes pasos:

1. Se descarga la base de datos del INE a través de la [API del INE](#) , para esta descarga se usa cuatro tablas distintas llamadas 'new_df1'(contiene los datos anuales del total de separaciones, nulidades y divorcios desde 2013-2020),'new_df2'(contiene los datos anuales del total de nulidades desde 2013-2020),'new_df3'(contiene los datos anuales del total de separaciones desde 2013-2020) y 'new_df5'(contiene los datos anuales del total de divorcios desde 2013-2020), tras obtener las 4 tablas, se añade un índice a cada una que es una secuencia de números empezando por 1 hasta el nº de filas que haya en cada tabla, que en todos los casos es 52, también se nombran las columnas con un nombre más claro para después trabajar con ellas, finalmente se cogen las 4 tablas y se juntan en una tabla única llamada 'datos' con el comando merge a través de la columna índice que se ha creado en cada dataframe. Es importante destacar que para poder unir dos dataframes en Python, es necesario que posean una columna en común. Se asignan nombres más sencillos a las columnas correspondientes a cada año en el dataframe 'datos', utilizando la notación abreviada de los años (por ejemplo, 2020T, 2020N, 2020S, 2020D). Este proceso se repite para cada año de estudio, de manera que se disponga de un dataframe con columnas fácilmente identificables para su posterior uso.

2. Se lee el archivo que contiene la información sobre la población de las provincias y se asigna el nombre 'datosmunicipios' a dicho dataframe.

3. Se procede a crear nuevas variables dentro de la tabla 'datos' que representan las frecuencias relativas. Esto es necesario para realizar un estudio significativo, ya que es común encontrar más separaciones en provincias con mayor población. Para ello, se tiene que crear un orden de las provincias e incluirlo en la tabla, para que siempre mantenga ese mismo orden y a la hora de hacer las operaciones que necesitemos en un futuro, coincidan las provincias de la tabla

'datosmunicipios' con las provincias de la tabla 'datos', seguidamente se realiza una pivoteación de la tabla 'datosmunicipios', de manera que los años se disponen en columnas y las provincias en filas. Luego, se realiza una limpieza del dataframe para conservar únicamente la columna de población, que se utilizará posteriormente en el cálculo de las frecuencias relativas. Para cada año estudiado, se divide el número total de nulidades, separaciones y divorcios en una provincia por la población total de esa provincia en ese mismo año.

A continuación, se multiplica el resultado por 10000, lo cual permite obtener datos más significativos y comparables entre diferentes años, cada columna tendrá su nombre(2020M,2020NM,2020DM,2020SM), así sucesivamente con todos años.

La M es la tasa de la suma total de los 3 casos. La NM, DM y SM es la tasa de cada una de las 3 por separado, sin tener en cuenta las demás, la que empieza por N son nulidades, la que empieza por D son divorcios y por S separaciones.

4. Para unir dos tablas debemos tener una columna de igual nombre y con el mismo significado en cada una de las tablas, entonces creamos una variable llamada números con la misma información que la columna que se encuentra en 'datosmunicipios', esta variable se pone en la columna llamada CC_2 de la tabla 'datos, que tras estos cambios ya es común con la que se encuentra en 'datosmunicipios' y se pueden unir las dos tablas.

5. Se realiza una conversión del tipo de datos de la columna llamada 'CC_2' a int64 en el dataframe 'datos'.

6. Del mismo modo, se realiza una conversión del tipo de datos de la columna 'CC_2' en el dataframe 'countries' a int64.

7. Se fusionan ambos dataframes utilizando el comando 'merge', por lo que se ha realizado un cambio previo de nombres en las columnas y se les ha asignado el nombre 'CC_2'. Asimismo, se ha llevado a cabo la conversión de los tipos de datos para facilitar la fusión.

8. Para el estudio posterior de los gráficos interactivos, se necesitan obtener unas tablas con un orden distinto al que se consigue en 'merge' o en 'datos' que cada año es una columna, entonces se coge cada grupo que se va a estudiar,

2013M-2020M, 2013NM-2020NM, 2013SM-2020SM, 2013DM-2020DM y se crea una nueva tabla para cada grupo en la que se encuentre los años en una misma columna llamada year, otra columna con su valor, una con el nombre de la provincia que se va repitiendo para cada año y la geometría, obtenemos una tabla como esta, que se van a llamar prueba_melted, prueba_melted1, prueba_melted2 y prueba_melted3 respectivamente.

9. Para la visualización de estos gráficos interactivos en Google Colaboratory, debido a la limitación de no mostrarlos directamente, se ha optado por descargar los gráficos en la base de datos del portátil. Posteriormente, se ha creado un repositorio en GitHub utilizando GitHub Desktop, una herramienta que facilita la carga de archivos de gran tamaño. Finalmente, se ha enlazado el repositorio con Netlify, una plataforma que permite la visualización de este tipo de gráficos y proporciona un enlace a una página web para que cualquier persona pueda apreciarlos. Este enfoque garantiza la accesibilidad de los gráficos a través de un entorno en línea, permitiendo una experiencia interactiva y compartible.

10. Se grafica la nueva tabla creada llamada 'merged' para ver si está todo correcto y se puede comprobar como las Islas Canarias tienen mucha separación con la Península Ibérica. Las Islas Canarias, al estar ubicadas en el océano Atlántico frente a la costa noroeste de África, pueden aparecer distorsionadas o alejadas del territorio peninsular en una representación geográfica estándar. Para solucionar este problema y lograr una visión más precisa y equilibrada en el gráfico, se realiza un ajuste específico de las coordenadas de las Islas Canarias.

Este ajuste implica una corrección en la posición geográfica de las Islas Canarias dentro del mapa de España. Se aplican transformaciones a las coordenadas geográficas de las Islas Canarias para acercarlas al territorio peninsular y lograr una representación más coherente y equilibrada.

El objetivo de este ajuste es garantizar que las Islas Canarias se muestran de manera adecuada y no aparezcan excesivamente separadas del resto del territorio español en el gráfico. Esto mejora la interpretación visual y la comprensión espacial de los datos en relación con las Islas Canarias.

En resumen, el procesamiento de los datos se lleva a cabo mediante la obtención y transformación de la base de datos del INE, la creación de variables con frecuencias relativas, la asignación de nombres sencillos a las columnas, la conversión de tipos de datos necesaria para la fusión de los dataframes y la unión de la tabla de datos con la información geométrica del mapa de España. Estos pasos son fundamentales para el posterior análisis y representación.

3.2. Análisis geoespacial

Se continúa con la explicación de algunas estructuras de datos principales que se usan en la ciencia de datos geográficos:

3.2.1. Tablas geográficas

Las tablas geográficas se reconocen debido a que una columna registra información geométrica. Cada objeto geográfico es una fila de la tabla y cada columna es una característica o atributo del objeto.

En el lenguaje usado en este trabajo una característica es una columna y una muestra una fila, que según en qué libros se mire, puede cambiar.

El Open Geospatial Consortium(OGC) ha definido un conjunto de tipos “abstractos” que se pueden usar para definir cualquier tipo de geometría.

Point: ubicación de dimensión cero con una coordenada “x” e “y”.

LineString: Camino compuesto por un conjunto de más de un Point.

Polygon: Superficie que tiene al menos un LineString que comienza y termina en la misma coordenada.

Estos 3 tipos tienen Multi: Son variantes que indican una colección de múltiples geometrías del mismo tipo.

3.2.2. Superficies

Se utilizan para registrar datos de un modelo de datos de campo.

Se representan como cuadrículas donde cada celda contiene una muestra.

Una cuadrícula también se puede considerar como una tabla con filas y columnas, pero esto está vinculado directamente a la ubicación geográfica.

3.3. Agrupación espacial

La agrupación espacial se refiere a la tendencia de las observaciones geográficas a agruparse en el espacio, lo que implica que las observaciones cercanas entre sí son más similares que las observaciones que están más alejadas. Es un concepto clave en el análisis espacial y se utiliza para estudiar patrones y procesos espaciales en diversas disciplinas, como la geografía, la ecología, la economía y la planificación urbana.

La agrupación espacial puede manifestarse de diferentes maneras:

1. **Agrupación espacial positiva:** Las observaciones similares se agrupan en áreas específicas, lo que resulta en la formación de cúmulos o clústeres. Esto puede indicar la presencia de patrones espaciales homogéneos o la influencia de factores locales que promueven la similitud entre las observaciones.
2. **Agrupación espacial negativa:** Las observaciones similares tienden a estar dispersas o separadas en el espacio, lo que indica la presencia de patrones espaciales heterogéneos. Puede ser causada por factores que generan diferencias o barreras entre las observaciones.
3. **Aleatoriedad espacial:** Las observaciones están distribuidas de manera aleatoria en el espacio sin una tendencia a agruparse o dispersarse. No hay patrón discernible ni influencia espacial detectable.

El análisis de agrupación espacial se basa en la construcción de matrices de pesos espaciales que cuantifican las relaciones de proximidad entre las ubicaciones geográficas. Estas matrices se utilizan para evaluar la autocorrelación espacial, que mide el grado de similitud espacial entre las observaciones y puede revelar patrones de agrupación espacial.

3.3.1. Spatial Lag

El "lag espacial" es un concepto utilizado en el análisis espacial para medir la influencia espacial de una variable sobre sí misma o sobre otras variables en un contexto geográfico. También se conoce como "autocorrelación espacial" o "efecto de vecindad espacial".

En el contexto de un análisis de regresión espacial, el "lag espacial" se refiere a la inclusión de una variable que representa la media o la suma ponderada de los valores de una variable en las ubicaciones vecinas. Este valor agregado se utiliza como una covariable en el modelo para capturar la influencia espacial de la variable sobre sí misma o sobre otras variables dependientes.

El cálculo del lag espacial implica la utilización de una matriz de pesos espaciales, que describe la relación de vecindad entre las ubicaciones geográficas. Esta matriz asigna pesos a las conexiones entre ubicaciones vecinas, indicando la intensidad de la relación espacial. Comúnmente se utilizan matrices de pesos espaciales basadas en criterios de vecindad contigua, como la matriz de pesos espaciales Queen o la matriz de pesos espaciales Rook.

Este método se utiliza para explorar la autocorrelación espacial y entender cómo las observaciones en ubicaciones cercanas se relacionan entre sí.

3.3.2. Caso Binario

En este proceso llamado caso binario, vamos a representar la agrupación espacial respecto a una variable con dos resultados posibles, 0 ó 1. El color de cada área en el mapa indica si el valor de esa área para ese año es alto o bajo en comparación con la mediana. La leyenda muestra la correspondencia entre los colores y las etiquetas "0 Low" y "1 High". La figura resultante permite comparar visualmente los patrones espaciales de la variable a lo largo de los años representados.

3.3.3. Análisis de recuento de uniones

El análisis de recuento de uniones (join count analysis) es una técnica utilizada en el análisis de autocorrelación espacial para evaluar la presencia de agrupaciones o dispersión de valores similares en una distribución espacial.

En el contexto de datos espaciales, el recuento de uniones se refiere al número de vecinos que tienen valores similares para una determinada variable. Esta técnica se basa en la idea de que, si los valores de una variable están distribuidos de manera aleatoria en el espacio, entonces el recuento de uniones esperado sería similar al recuento de uniones observado. Sin embargo, si los valores de la variable están agrupados o dispersos de manera no aleatoria, el recuento de uniones observado será significativamente diferente del esperado bajo la aleatoriedad espacial.

El análisis de recuento de uniones se realiza en tres pasos principales:

1. Construcción de una matriz de pesos espaciales: Se crea una matriz que define la relación de vecindad entre las unidades espaciales, como vecinos contiguos o conectados. Esta matriz se utiliza para determinar qué unidades espaciales están cerca unas de otras.
2. Cálculo del recuento de uniones observado: Se calcula el recuento de uniones para cada unidad espacial en base a los valores de la variable de interés y la matriz de pesos espaciales. Esto implica contar cuántos vecinos tienen valores similares para cada unidad espacial.
3. Generación de distribuciones de recuento de uniones esperadas: Se generan distribuciones de recuento de uniones esperadas bajo la hipótesis nula de aleatoriedad espacial. Esto se logra mediante simulaciones que mantienen la distribución de valores, pero reordenan aleatoriamente las ubicaciones espaciales de las unidades.

El gráfico que se va a realizar va a mostrar la distribución del recuento de uniones observadas y simuladas para cada año procesado. El recuento de uniones observadas se representa mediante una línea vertical roja, mientras que la línea vertical amarilla indica el valor promedio de recuento de uniones simuladas.

En el gráfico, el eje x representa el recuento de uniones, mientras que el eje y muestra la frecuencia o la densidad de los valores. La distribución de recuento de uniones simuladas se genera mediante simulaciones que mantienen la distribución de valores, pero reordenan aleatoriamente las ubicaciones espaciales de las unidades.

3.3.4. Gráficos de densidad del Índice de Moran

El gráfico de densidad del Índice de Moran es una representación visual que muestra la distribución de los valores del Índice de Moran en un conjunto de datos espaciales. El Índice de Moran es una medida de autocorrelación espacial que evalúa si las observaciones en una región están espacialmente agrupadas o dispersas.

En el gráfico de densidad del Índice de Moran, el eje x representa los valores del Índice de Moran, mientras que el eje y representa la densidad de ocurrencia de esos valores. La curva de densidad muestra la forma de la distribución de los valores del Índice de Moran, lo que permite identificar si hay una tendencia hacia la agrupación espacial (valores altos del Índice de Moran) o hacia la dispersión espacial (valores bajos del Índice de Moran).

Una distribución de densidad del Índice de Moran que se sesga hacia la derecha indica una fuerte autocorrelación espacial positiva, lo que significa que las observaciones similares tienden a agruparse en el espacio. Por otro lado, una distribución de densidad sesgada hacia la izquierda indica una fuerte autocorrelación espacial negativa, lo que significa que las observaciones similares tienden a dispersarse en el espacio. Si la distribución de densidad es más simétrica, con un pico centrado, indica una autocorrelación espacial débil o nula.

El gráfico de densidad del Índice de Moran también puede incluir la representación del valor esperado y el valor observado del Índice de Moran.

En este gráfico, se pueden agregar líneas verticales para representar el valor observado y el valor esperado del Índice de Moran. Esto permite visualizar cómo se compara el Índice de Moran observado con el valor esperado y proporciona información sobre la existencia de patrones espaciales significativos.

Si el Índice de Moran observado está por encima del valor esperado, indica una autocorrelación espacial positiva significativa, lo que implica que las observaciones similares están más agrupadas de lo que se esperaría al azar. Si el Índice de Moran observado está por debajo del valor esperado, indica una

autocorrelación espacial negativa significativa, lo que implica que las observaciones similares están más dispersas de lo que se esperaría al azar.

La inclusión de las líneas verticales para el valor observado y el valor esperado en el gráfico de densidad del Índice de Moran ayuda a interpretar de manera más precisa la presencia y la magnitud de la autocorrelación espacial en los datos, permitiendo identificar patrones espaciales significativos y evaluar su relevancia estadística.

3.3.5. Moran Scatterplot y puntos fríos y calientes

El Moran Scatterplot (gráfico de dispersión de Moran) es una representación visual utilizada en el análisis espacial para examinar la autocorrelación espacial de una variable en relación con su lag espacial. El gráfico muestra la relación entre los valores de la variable en el eje X y los valores del lag espacial correspondiente en el eje Y.

En el Moran Scatterplot, los puntos dispersos alrededor de una línea de ajuste indican una autocorrelación espacial positiva. Esto significa que las observaciones vecinas tienden a tener valores similares, formando agrupaciones o patrones espaciales. Por otro lado, si los puntos están dispersos sin seguir una tendencia clara, indica una autocorrelación espacial débil o inexistente.

El Moran Scatterplot es una herramienta útil para identificar patrones espaciales y analizar la autocorrelación espacial de una variable en un contexto geográfico. Ayuda a visualizar las relaciones entre las observaciones y su ubicación espacial.

Los términos "puntos calientes" (hotspots) y "puntos fríos" (coldspots) se utilizan en el contexto del análisis de patrones y la autocorrelación espacial para describir características de agrupamiento o dispersión de valores en un conjunto de datos geográficos.

Un punto "caliente" (hotspot) se refiere a una ubicación geográfica donde los valores de una variable específica son significativamente altos en comparación con sus vecinos. Indica una agrupación espacial de valores altos, lo que implica una autocorrelación espacial positiva. En otras palabras, los puntos calientes representan áreas donde ocurren eventos o condiciones excepcionalmente altas.

Por otro lado, un punto "frío" (coldspot) se refiere a una ubicación geográfica donde los valores de una variable específica son significativamente bajos en comparación con sus vecinos. Indica una agrupación espacial de valores bajos, lo que implica una autocorrelación espacial negativa. Los puntos fríos representan áreas donde ocurren eventos o condiciones excepcionalmente bajas.

En resumen, los puntos calientes y fríos son utilizados para identificar las áreas con valores extremos en relación con sus vecinos cercanos. Estos patrones pueden ser importantes para comprender la distribución espacial de los fenómenos estudiados y pueden proporcionar información sobre posibles procesos subyacentes que generan esos patrones.

3.4. Software y hardware

Python es un lenguaje de programación de alto nivel y de propósito general que se caracteriza por su legibilidad y simplicidad. Fue creado por Guido van Rossum en la década de 1990 y se ha vuelto extremadamente popular en la comunidad de programadores debido a su facilidad de uso y a la amplia gama de bibliotecas y herramientas disponibles.

Python se destaca por su sintaxis clara y legible, lo que facilita la lectura y comprensión del código. Además, cuenta con una amplia variedad de módulos estándar que cubren una amplia gama de funcionalidades, desde el manejo de archivos y el procesamiento de texto hasta la programación en red y el desarrollo web.

Una de las características más destacadas de Python es su enfoque en la legibilidad del código, fomentando el uso de espacios en blanco y una estructura clara y ordenada. Esto permite que los programas escritos en Python sean más fáciles de entender y mantener.

En cuanto a Google Colab, es un entorno de desarrollo en línea que proporciona una plataforma de codificación basada en la nube y está especialmente diseñada para trabajar con el lenguaje de programación Python. Google Colab permite a los usuarios escribir y ejecutar código Python directamente en el navegador web,

lo que elimina la necesidad de configurar y mantener un entorno de desarrollo local.

Google Colab ofrece una serie de características útiles, como la capacidad de ejecutar celdas de código individualmente, lo que facilita la depuración y la experimentación iterativa. También permite el uso de bibliotecas populares de Python, como NumPy, Pandas y Matplotlib, sin necesidad de instalarlas manualmente.

Además, Google Colab proporciona acceso gratuito a recursos de hardware, como unidades de procesamiento gráfico (GPU) y unidades de procesamiento tensorial (TPU), lo que permite realizar cálculos intensivos y entrenar modelos de aprendizaje automático de manera más eficiente.

Para concluir, Python es un lenguaje de programación de alto nivel y fácil de usar, mientras que Google Colab es un entorno de desarrollo en línea que ofrece una plataforma conveniente para escribir, ejecutar y colaborar en proyectos de Python.

3.4.1. Librerías

Las librerías de Python usadas en este trabajo son:

1. “numpy”: “numpy” es una librería de Python para computación científica que proporciona soporte para arreglos y matrices multidimensionales, así como para una amplia variedad de funciones matemáticas para operar sobre estos arreglos, es una de las librerías más utilizadas en el ecosistema de Python para análisis de datos y cálculos numéricos.
2. “pandas”: Define nuevas estructuras de datos basadas en los arrays de la librería “numpy”, permite leer y escribir fácilmente ficheros en formatos CSV, Excel Y SQL, se puede acceder a los datos mediante índices o nombres para filas y columnas, permite trabajar con series temporales, nos permite agrupar, ordenar, filtrar o describir un conjunto de datos y es muy eficiente.
3. “geopandas”: Esta librería permite leer y escribir datos espaciales de múltiples formatos, medir áreas y distancias, crear nuevas geometrías como polígonos

envolventes, buffer u obtener centroides, analizar relaciones espaciales, hacer mapas y realizar geocodificaciones.

4. “contextily”: Es una biblioteca de Python que permite recuperar y agregar mapas de mosaicos de Internet como mapa base a las figuras de “matplotlib”. Además, ofrece la posibilidad de guardar los mapas de mosaicos como archivos ráster geoespaciales en el disco. Con "Contextily", es posible enriquecer visualmente las visualizaciones geoespaciales con mapas de fondo detallados y contextualmente relevantes, mejorando así la comprensión y el análisis de los datos espaciales.

5. “shapely”: "shapely" es una biblioteca de Python con licencia BSD que proporciona herramientas para la manipulación y análisis de objetos geométricos planos. Permite realizar operaciones geométricas como unión, intersección, diferencia y buffer en puntos, líneas y polígonos. Con "shapely", es posible realizar cálculos espaciales avanzados, como determinar la distancia entre objetos, comprobar la validez de geometrías y realizar consultas espaciales. Esta biblioteca es ampliamente utilizada en aplicaciones de análisis geoespacial, geografía computacional y visualización de datos espaciales, facilitando el procesamiento y la manipulación de datos geoespaciales de forma eficiente y precisa.

6. “matplotlib.pyplot”: “matplotlib” es una biblioteca de python que se utiliza para crear gráficos y diagramas 2D mediante scripts de python. Tiene un módulo llamado pyplot que facilita el trazado al proporcionar una función para controlar los estilos de línea, las propiedades de la fuente, los ejes de formato.

7. “rasterio”: La librería “rasterio” es una librería de Python para trabajar con datos geoespaciales en formato raster. Con ella, se pueden leer y escribir datos en formatos como GeoTIFF, NetCDF, entre otros.

8. “mapclassify”: “Mapclassify” es una librería de Python para la clasificación de datos geoespaciales. Esta librería se utiliza para clasificar los datos en diferentes clases, y para crear mapas temáticos. Incluye diferentes métodos de clasificación, como la clasificación k-means, la clasificación de cuantiles, la clasificación de Jenks, entre otros. Estos métodos se utilizan para clasificar los

datos en diferentes clases, según sus valores, también permite visualizar los datos clasificados utilizando diferentes esquemas de colores. La librería incluye diferentes esquemas de colores predefinidos, y también permite crear esquemas de colores personalizados.

9. “libpysal”: Esta librería se utiliza para realizar análisis espaciales avanzados, como el análisis de clústeres espaciales, la regresión y la interpolación espaciales. Incluye una variedad de métodos y algoritmos para el análisis espacial, como la construcción de pesos espaciales, la estimación de modelos espaciales, la simulación espacial, entre otros. Además, también incluye herramientas para la visualización de datos espaciales. Forma parte del proyecto PySAL (Python Spatial Analysis Library), que es una biblioteca de software de código abierto para el análisis espacial en Python. PySAL proporciona una amplia gama de herramientas para el análisis espacial, desde el análisis exploratorio de datos hasta el modelado espacial avanzado.

10. “os”: Las funciones más comunes que ofrece la librería “os” se incluyen:

10.1. Acceso al sistema de archivos: Proporciona funciones para crear, eliminar, mover y renombrar archivos y directorios en el sistema de archivos.

10.2. Variables de entorno: Permite obtener y establecer variables de entorno del sistema operativo.

10.3. Información del sistema operativo: Proporciona información sobre el sistema operativo y el entorno de ejecución, como el nombre de usuario, la ruta del directorio de trabajo actual y el nombre del sistema operativo.

10.4 Funciones para rutas: Funciones para manejar rutas de archivos y directorios, como la concatenación de rutas y la separación de nombres de archivo y directorio.

11. “esda”: “esda” proporciona una serie de métodos y funciones para el análisis exploratorio de datos espaciales, incluyendo:

- Mapas de Moran: para detectar patrones de autocorrelación espacial en datos espaciales.
- Mapas de LISA: para identificar clústeres de áreas similares en datos espaciales. -Análisis de regresión espacial: para modelar la dependencia espacial en datos de regresión.
- Análisis de clúster espacial: para identificar clústeres espaciales en datos agrupados.

12. “seaborn”: Es una librería de visualización de datos en Python que se utiliza para crear gráficos estadísticos atractivos y efectivos. Esta librería se basa en la librería “matplotlib” y proporciona una interfaz fácil de usar para la creación de gráficos complejos con menos líneas de código.

13. “plotly”: “plotly” es una biblioteca de visualización interactiva en Python que permite crear gráficos y visualizaciones de datos de manera flexible y dinámica. Proporciona una amplia gama de tipos de gráficos, desde gráficos de dispersión hasta diagramas de caja, y ofrece capacidades interactivas, como zoom, desplazamiento y herramientas de resaltado. Con “plotly”, puedes crear visualizaciones atractivas y personalizables para explorar y comunicar tus datos de manera efectiva.

14. “requests”: “requests” es una biblioteca de Python que facilita la realización de solicitudes HTTP de manera simple y eficiente. Con esta biblioteca, los desarrolladores pueden enviar solicitudes a servidores web y recibir respuestas de forma sencilla. Proporciona una interfaz amigable que abstrae la complejidad del protocolo HTTP, permitiendo realizar operaciones como obtener contenido de una URL, enviar datos a través de formularios, establecer encabezados personalizados y gestionar cookies de manera intuitiva. “requests” es ampliamente utilizada en aplicaciones web, servicios API y tareas de raspado web, brindando una forma conveniente de interactuar con recursos en la web.

15. “datetime”: La biblioteca “datetime” es una biblioteca incorporada en Python que proporciona clases y funciones para trabajar con fechas y tiempos. Permite

crear, manipular y formatear objetos relacionados con fechas y horas de manera conveniente.

16. “cufflinks”: La biblioteca “cufflinks” es una extensión de la biblioteca plotly que permite generar fácilmente gráficos interactivos a partir de los objetos de datos de Pandas en Python. Proporciona una interfaz simplificada para crear visualizaciones utilizando la sintaxis familiar de Pandas.

17.” from IPython.display import display, HTML”: Se utiliza en entornos de Jupyter Notebook o Jupyter Lab para importar las funciones display y HTML del módulo IPython.display. Estas funciones se utilizan para mostrar contenido enriquecido y HTML dentro del entorno de la notebook.

La función display se utiliza para mostrar diversos tipos de contenido, como texto, imágenes, tablas y gráficos, de manera formateada y visualmente agradable dentro de la notebook. Puedes pasar objetos o variables a la función display y la notebook mostrará el contenido de manera adecuada.

La función HTML permite mostrar código HTML directamente en la notebook. Puedes utilizar esta función para incrustar contenido HTML personalizado en tu notebook, como tablas HTML, fragmentos de código o elementos interactivos.

Tras esta introducción explicando cada librería, se debe tener en cuenta que muchas veces hay que instalarlas y después importarlas.

4. RESULTADOS

En esta sección se presentan los resultados derivados el análisis descriptivo y del análisis de agrupación espacial.

4.1. Análisis descriptivo

Se presentan a continuación los mapas con distribución provincial y la evolución temporal durante el periodo 2013-2020 de las tasas de divorcios, separaciones, nulidades y el total de todas ellas.

4.1.1. Gráficos del total de la tasa de nulidades desde 2013-2020

En la Figura 3 se muestran las tasas provinciales de nulidades durante el periodo estudiado. Como se aprecia, estas tasas son extremadamente bajas, algunas de ellas con valor 0, y aportan poco valor en el estudio global. No obstante, se ha considerado interesante mostrarlas.

Con el fin de diferenciar provincias, se ha considerado la categorización (en tasas por 100.000 habitantes), dada por {0, 0.01, 0.04 y 0.35}. Se observa una progresiva disminución de las nulidades desde el año 2013, destacada por la suavización progresiva del color en los mapas, por lo que parece ser una práctica que ha ido perdiendo popularidad entre la población (es notorio especialmente en 2018/2019 respecto de 2013). No obstante, la escasa incidencia de nulidades durante el año 2020 podría estar justificada por el parón administrativo ocasionado por el confinamiento de ese año.

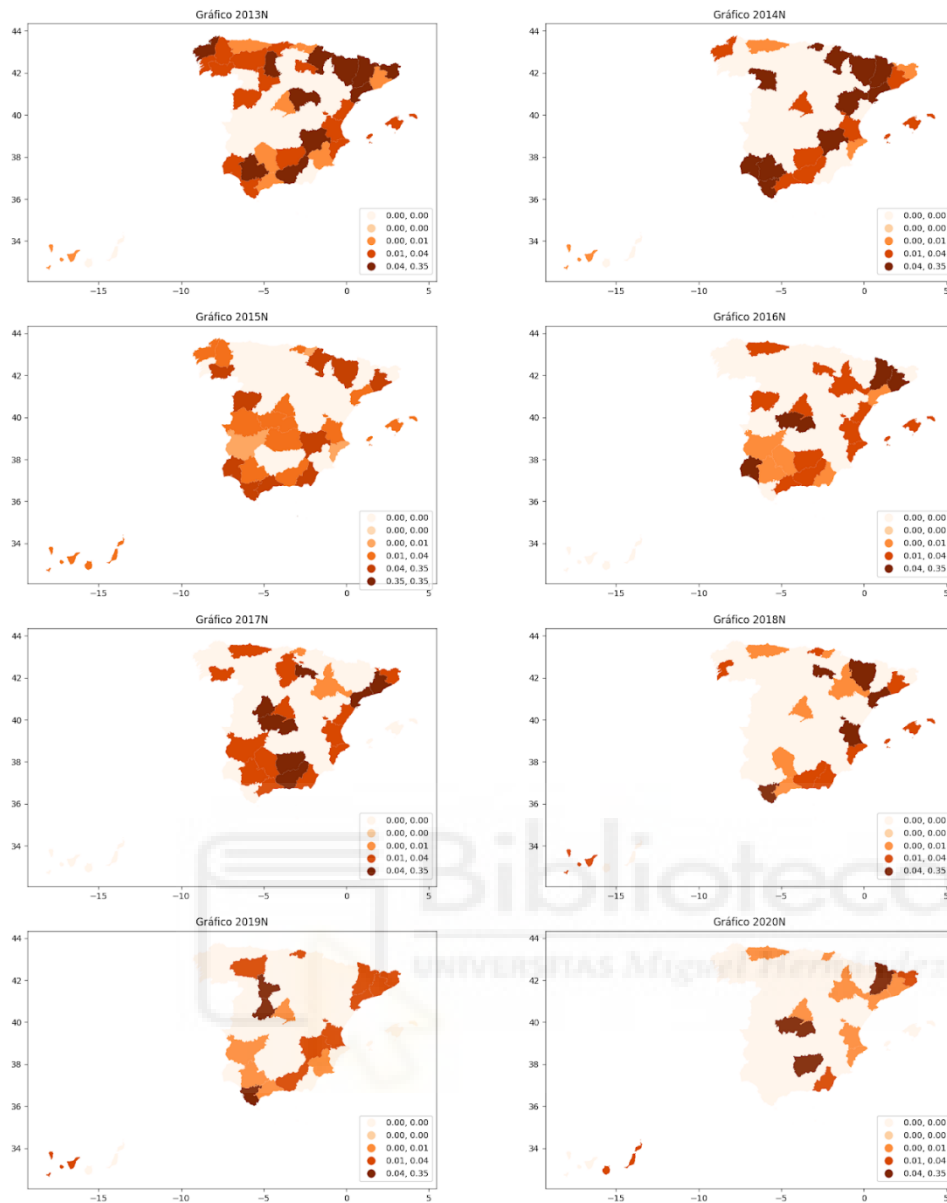


Figura 3. Tasa de divorcios de 2013-2020. "Fuente: elaboración propia"

4.1.2. Gráficos de la tasa de divorcios desde 2013-2020

Con relación al análisis de los divorcios, se puede observar claramente en la Figura 4, que es el tipo de disolución matrimonial más frecuente, representando más del 90% del total. Se ha considerado la categorización (en tasas por 100.000 habitantes), dada por {8-15,15-21,21-27,27-34}. Existe una clara tendencia en la distribución geográfica de los divorcios, donde las regiones del este y las zonas costeras presentan las tasas más altas. Por el contrario, las provincias situadas en el centro de España muestran tasas más bajas.

Es destacable el comportamiento de las Islas Canarias, específicamente Las Palmas y Santa Cruz de Tenerife, que experimentan un descenso significativo en sus tasas a lo largo del periodo estudiado. Inicialmente, se encontraban entre las provincias con las tasas más elevadas, pero finalizan siendo de las provincias con tasas más bajas.

Por otro lado, la provincia de Guadalajara se destaca como la única con una tasa excepcionalmente alta en el año 2020. Si analizamos su evolución a lo largo de los primeros años, se observa que su tasa era relativamente baja y fue aumentando progresivamente hasta alcanzar el máximo en 2020, convirtiéndose en la única provincia con una tasa tan elevada.

También se ve cómo se van elevando las tasas conforme van avanzando los años, teniendo siempre en cuenta la excepción del año 2020 por el COVID-19, ya que en este año todas las tasas bajan debido en gran parte al confinamiento.

En general, la mayoría de las provincias mantienen una tasa relativamente constante a lo largo de los años, sin presentar cambios bruscos. Se puede afirmar que existe una cierta estabilidad en el comportamiento de las tasas de divorcio a nivel provincial en España.

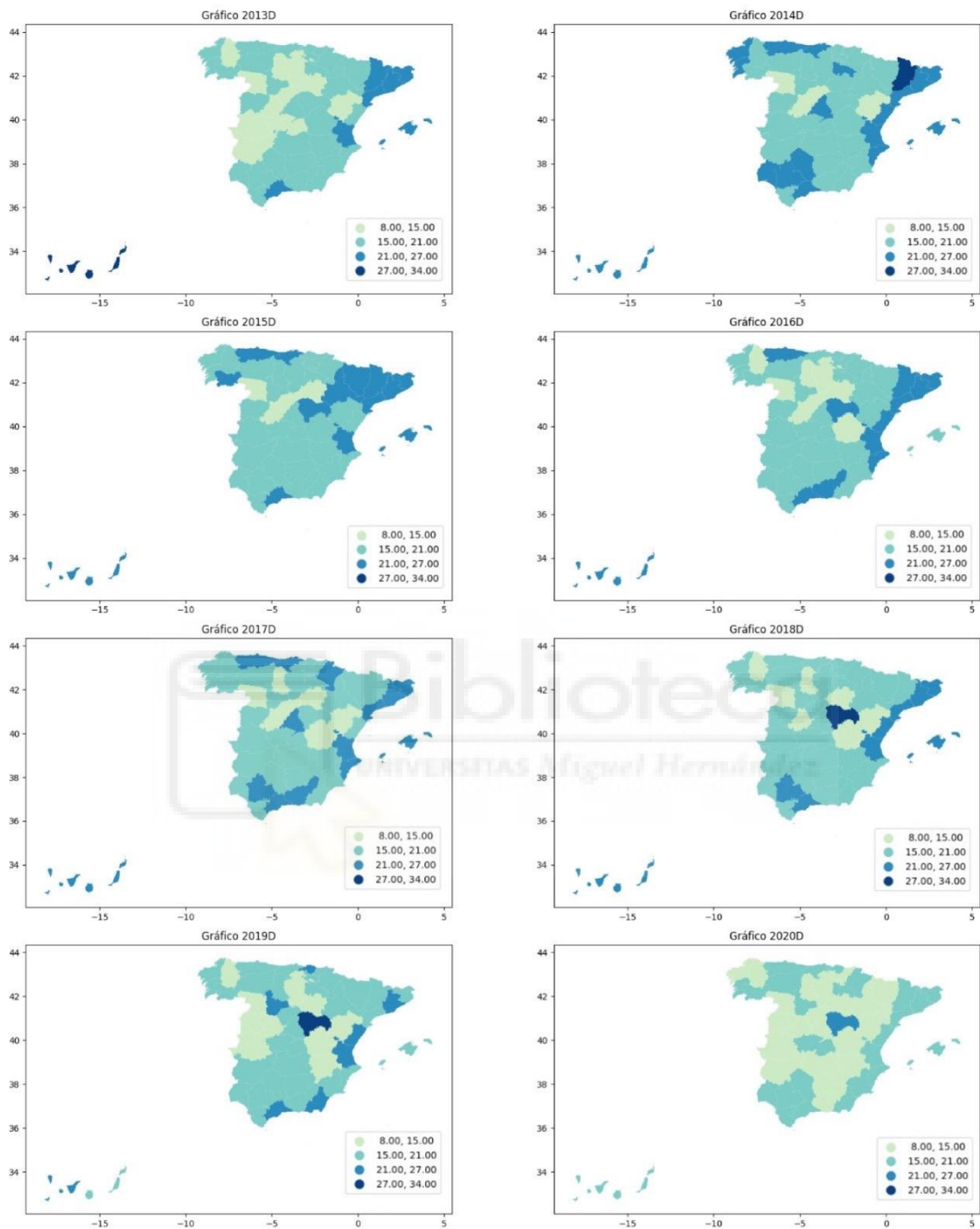
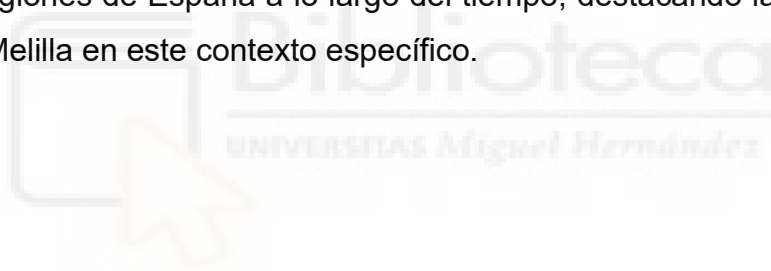


Figura 4. Tasa de divorcios de 2013-2020. "Fuente: elaboración propia"

4.1.3. Gráficos de la tasa de separaciones desde 2013-2020

En el contexto de las separaciones en España, se ha considerado la categorización (en tasas por 100.000 habitantes), dada por $\{0-0'75, 0'75-1'5, 1'5-2'25, 2'25-3'10\}$, se puede observar en la Figura 5 una tendencia acentuada en las provincias de Ceuta y Melilla. Estas provincias se destacan debido a sus altas tasas de separaciones en los primeros años analizados. Para resaltar su importancia, se han representado visualmente en los mapas como triángulos grandes. A medida que avanzan los años, se observa una disminución en las tasas, con la mitad de las provincias ubicadas aproximadamente entre $1'5-2'25$ y la otra mitad entre $0'75-1'5$, exceptuando el año 2020. Este año, afectado por la pandemia del COVID-19, todas las provincias experimentaron una disminución en sus registros, lo que se refleja en el mapa de España, aunque algunas zonas cercanas a Portugal y algunas provincias del norte se mantuvieron relativamente estables. Estos hallazgos evidencian la evolución de las separaciones en diferentes regiones de España a lo largo del tiempo, destacando la importancia de Ceuta y Melilla en este contexto específico.



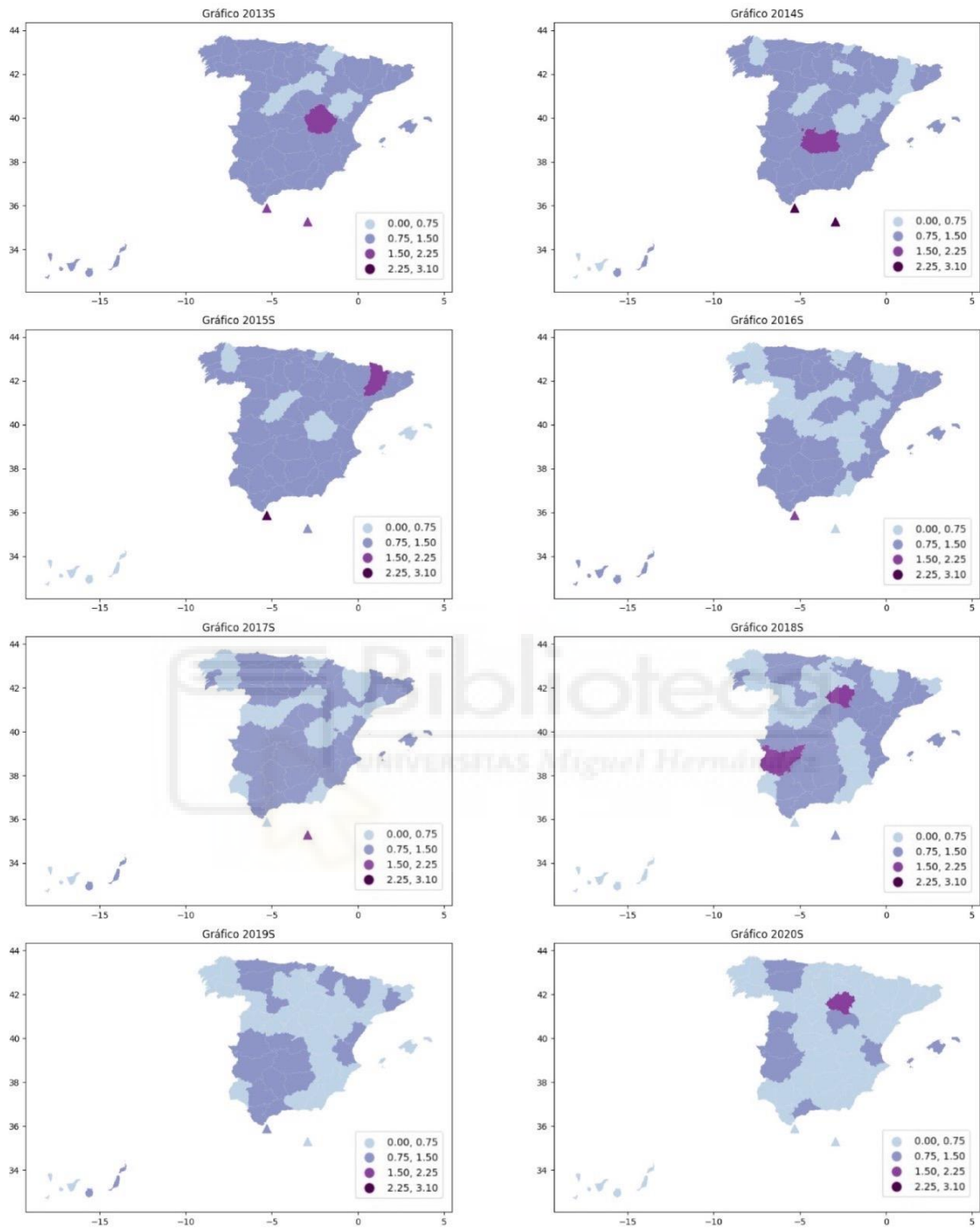


Figura 5. Tasa de separaciones de 2013-2020. "Fuente: elaboración propia"

4.1.4. Gráficos del total de la tasa de la suma de nulidades, separaciones y divorcios desde 2013-2020.

En esta Figura 6 se ha considerado la categorización (en tasas por 100.000 habitantes), dada por {9-15,15-21,21-27,27-34}, una muy similar a la de separaciones, ya que como se puede ver ocupan más del 90% de los registros totales.

Se observa de manera evidente una escasa presencia de provincias con una tasa considerablemente alta de frecuencia. De hecho, únicamente en cuatro de los ocho años estudiados encontramos una provincia destacada: Valencia en 2013, Lérica en 2014, Guadalajara en 2018 y Guadalajara nuevamente en 2019.

Asimismo, resulta notable la tendencia de las zonas costeras a presentar tasas más elevadas en comparación con el centro de España a lo largo de todos los años analizados, tal como se aprecia en la mayoría de los gráficos mediante tonalidades más intensas y proporcionadas en los colores.

Un aspecto que se puede destacar es la notable neutralidad de las tasas a lo largo de la mayoría de los años. Estas no se sitúan ni en niveles muy bajos ni en niveles muy altos, sino que se mantienen en un rango central.

Un dato de gran relevancia es la disminución de la tasa en el año 2020, como se había mencionado al inicio de este proyecto. La irrupción del COVID-19, junto con el confinamiento prolongado en los hogares y la excepcionalidad de dicho año, ocasionaron una considerable caída en el número total de nulidades, separaciones y divorcios.

Además, se observa un descenso gradual en Santa Cruz y Las Palmas, las cuales comenzaron en 2013 siendo de las provincias con tasas más elevadas y finalizaron en 2019 ubicándose entre las de menor incidencia.

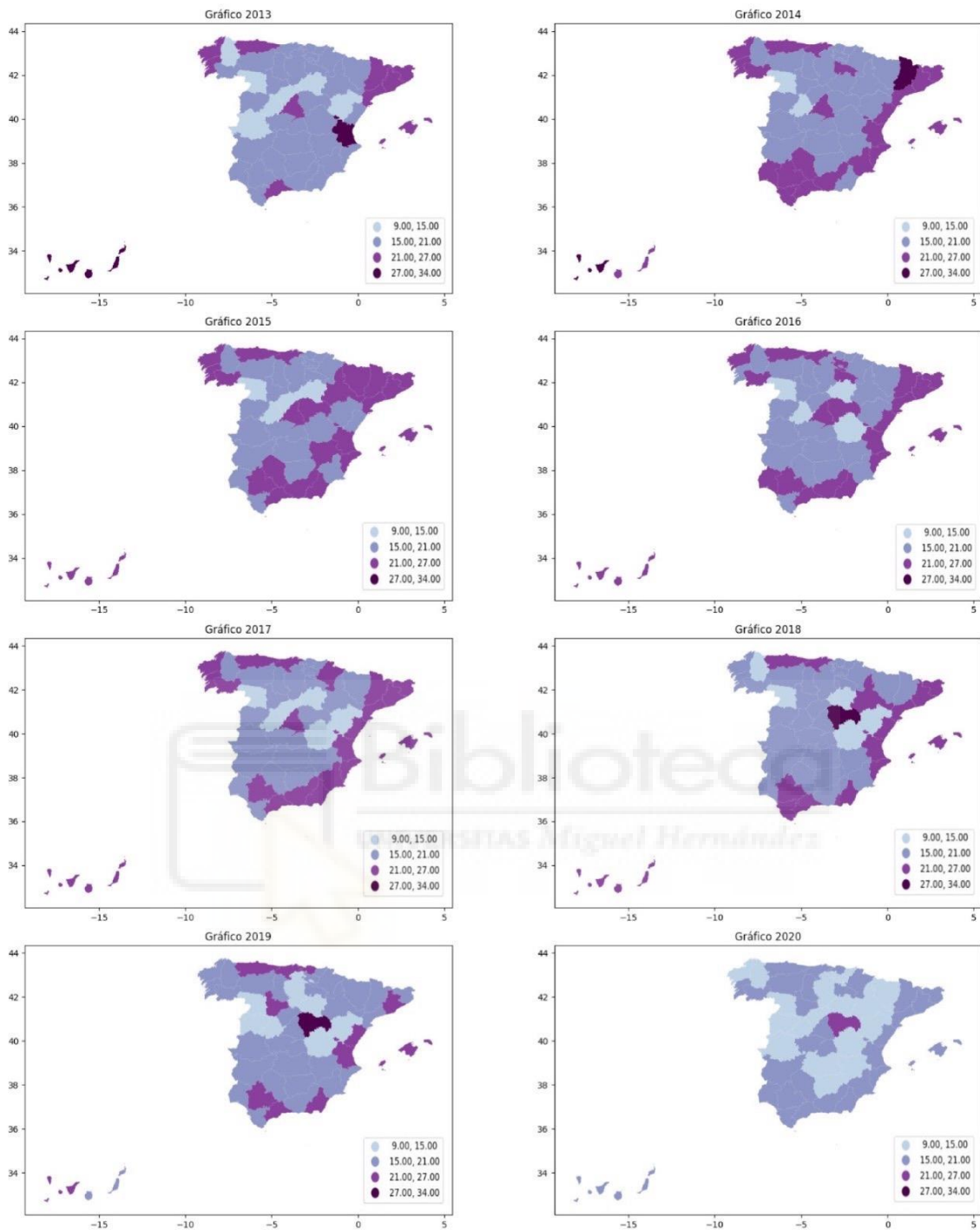


Figura 6. Tasa del total de nulidades, separaciones y divorcios de 2013-2020. "Fuente: elaboración propia"

A continuación, se proporciona el enlace del gráfico interactivo en el que podemos reproducir mediante animación, la evolución temporal de las tasas globales sobre el mapa de España:

[Graph \(flourishing-axolotl-1cf241.netlify.app\)](http://Graph(flourishing-axolotl-1cf241.netlify.app))



Figura 7. Gráfico interactivo de la tasa total de la suma de separaciones, nulaciones y divorcios.
"Fuente: elaboración propia"

4.2. Agrupación espacial

Se presentan a continuación los distintos tipos de agrupaciones espaciales que se han estudiado en este trabajo, en todos los estudios se han eliminado tanto las Islas, como Ceuta y Melilla debido a que no tienen vecinos y pueden influenciar en el estudio de una manera incorrecta.

4.2.1. Spatial Lag

En esta Figura 6, se puede ver como se dividen los datos en 4 grupos, esta elección es para continuar con la categorización previa en otros gráficos, explicando ya los resultados, se puede ver como en los primeros años, el centro de España muestra valores bajos en comparación con otras provincias. Sin embargo, a medida que avanzan los años, estos valores aumentan gradualmente, lo que indica un cambio en la distribución espacial. Es interesante notar que las provincias de Andalucía presentan una agrupación espacial constante, ubicándose en los dos grupos más altos en la mayoría de los años.

Otro aspecto destacable es la presencia de valores altos en la zona este de España durante la mayoría de los años analizados.

Es evidente que los primeros 4 años muestran gráficos similares entre sí. Sin embargo, a partir de 2017, se observan cambios en las agrupaciones espaciales, hasta llegar a 2019 o 2020, donde los patrones difieren significativamente de los observados en 2013. En estos años posteriores, los valores altos y bajos se encuentran más separados, en contraste con la cercanía que muestran en 2013.

En resumen, el análisis espacial de las tasas de nulidades, separaciones y divorcios revela una coherencia notable en los años 2013, 2014, 2015, 2016 y 2017 los cuales exhiben similitudes significativas en su distribución geográfica. Estos años se agrupan de manera consistente, lo que sugiere la existencia de un patrón espacial establecido. Sin embargo, al examinar los años posteriores, se observa un cambio en la distribución, lo que dificulta la obtención de conclusiones claras. Estos años presentan una distribución geográfica que parece aleatoria, sin un patrón discernible. Es posible percibir una evolución en la distribución geográfica entre los primeros cinco años, que pueden agruparse en una categoría, y los años restantes.

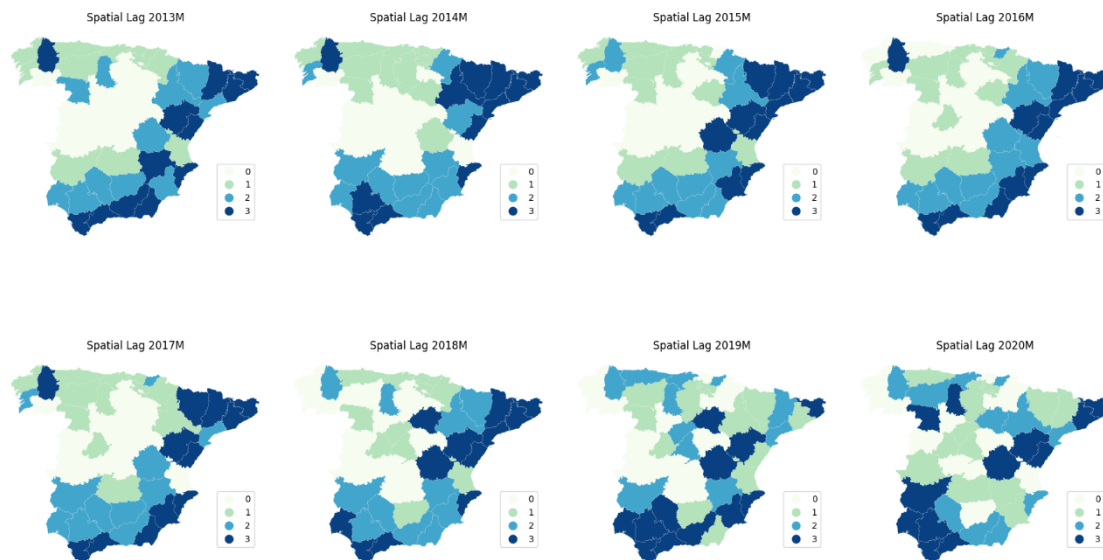


Figura 6. Spatial Lag de la tasa de nulidades, separaciones y divorcios de 2013-2020. "Fuente: elaboración propia"

4.2.2. Caso Binario

Estos gráficos representan la división de las provincias en dos categorías: "Low" (baja tasa) o "High" (alta tasa), en relación con la mediana de la variable en ese año.

Durante los años 2013 y 2014, se observa en la Figura 7, una coincidencia en la agrupación espacial de la tasa de nulidades, separaciones y divorcios en la mayoría de las provincias, excepto dos. En 2013, las provincias de León y Almería se ubican por encima de la mediana, mientras que, en 2014, estas provincias se sitúan por debajo y Navarra y Guadalajara experimentan un aumento. Además, se puede notar que las zonas de Andalucía, el Mediterráneo y Galicia presentan las tasas más altas en estos primeros años.

En 2015, algunas provincias muestran una dispersión respecto a los años anteriores, y se observan cambios significativos en dos o tres de ellas, como Murcia y Castellón, en este año la zona de Barcelona se agrupa en tasas altas.

En los años 2016 y 2017, se vuelve a evidenciar una dinámica similar a la observada en 2013 y 2014.

Con relación a los tres últimos años analizados, la zona del Mediterráneo sigue manteniendo las tasas más altas. Sin embargo, se presentan datos curiosos, en 2019, todas las provincias de Andalucía se encuentran por encima de la mediana, lo cual se asemeja a lo observado anteriormente, donde solían estar en esa posición. En cambio, en 2020, cuatro provincias andaluzas experimentan una disminución en sus tasas. Dado que el año 2020 fue atípico, pueden influir múltiples factores que expliquen este cambio brusco. Durante estos tres últimos años, también se observa un aumento en la tasa en algunas provincias del centro de España, así como la inclusión de otras provincias que no mostraban tasas tan altas en años anteriores.

En resumen, según la agrupación espacial en el caso de una variable binaria, se ve como la zona mediterránea suele agruparse en las tasas más altas en casi todos los años, y la zona céntrica en las tasas más bajas, hay que tener en cuenta que en este estudio se pueden obtener patrones más claros debido a que se divide solamente en 2 grupos, entonces es lógico que se vean más similitudes respecto a otros estudios.

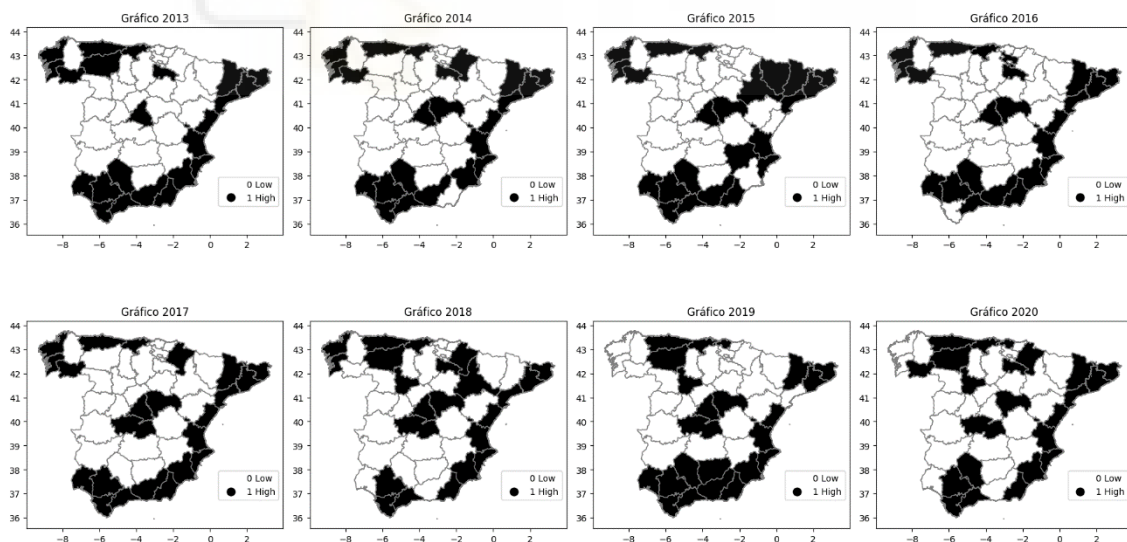


Figura 7. Caso binario de agrupaciones espaciales de la tasa de nulidades, separaciones y divorcios de 2013-2020. "Fuente: elaboración propia"

4.2.3. Análisis join counts

En este análisis de agrupación espacial, viendo la Figura 8, se pueden extraer varias conclusiones significativas. Cuando la línea vertical amarilla, que representa el promedio de los recuentos de uniones simuladas, se encuentra cercana a la línea roja, que representa los recuentos de uniones observadas, indica que existe una autocorrelación espacial que se aproxima a la media.

Una interpretación pertinente de esta situación sería que, en ese año en particular, las separaciones entre provincias no presentan un patrón espacial significativamente agrupado ni disperso. En otras palabras, las provincias con valores altos de separación no tienden a agruparse espacialmente cerca de otras provincias con valores altos, y lo mismo ocurre para las provincias con valores bajos. Esto se evidencia en los años 2013, 2015 y 2019, los cuales son especialmente notables. Al examinar los gráficos descriptivos de las nulidades, separaciones y divorcios en general, se observa que en estos tres años los resultados están dispersos, y las tasas altas se encuentran distribuidas a lo largo de todo el mapa.

Por otro lado, cuando la línea vertical amarilla está más separada de la línea vertical roja, se puede interpretar que existe una tendencia a la formación de grupos de provincias con valores similares. Las provincias con valores altos de separación tienden a agruparse espacialmente cerca de otras provincias con valores altos, y lo mismo ocurre para las provincias con valores bajos. Estos casos se presentan en los años restantes, como 2014, 2016, 2017 y 2018.

Cabe destacar que el año 2020 resulta especialmente interesante para esta segunda conclusión, a pesar de ser un año atípico debido a la pandemia de COVID-19. En este año, la línea vertical amarilla y la línea vertical roja están notablemente separadas en comparación con los demás años. Al observar el gráfico descriptivo de este año, se evidencia claramente cómo las provincias se agrupan tanto en tasas altas como bajas.

En resumen, este análisis de agrupación espacial proporciona información valiosa sobre los patrones de separaciones entre provincias en cada año. Permite identificar años en los que las provincias no presentan una agrupación espacial significativa, así como años en los que se forman grupos de provincias

con valores similares. Estas conclusiones enriquecen el estudio al ofrecer perspectivas sobre la distribución espacial de las separaciones.

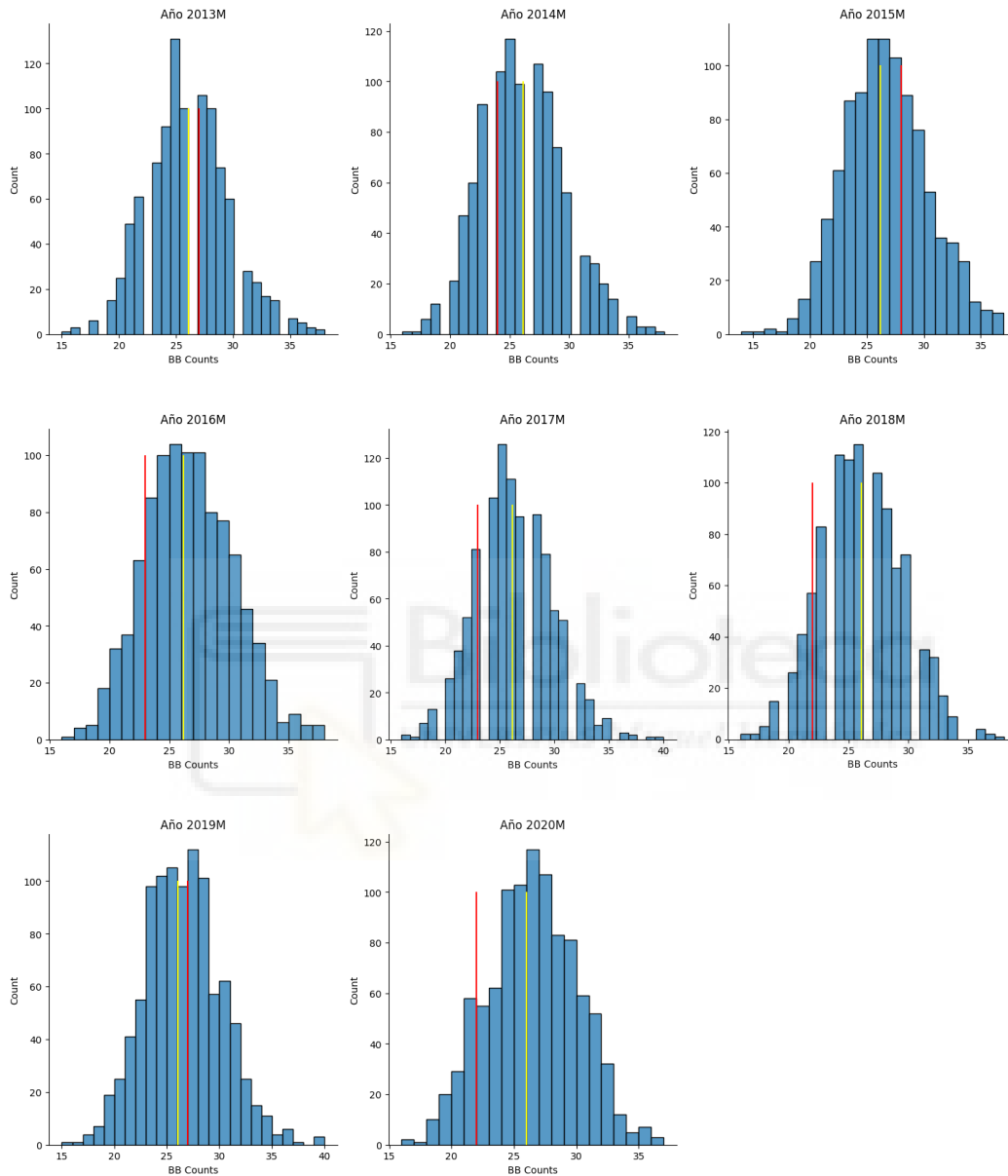


Figura 8. Análisis joint counts de la tasa de nulidades, separaciones y divorcios de 2013-2020
"Fuente: elaboración propia"

4.2.4. Gráficos de densidad del Índice de Moran

Tras observar los gráficos generados en la Figura 8, se aprecia que todos presentan una distribución bastante simétrica, lo que indica que la autocorrelación espacial es débil. Esto significa que la tendencia de agrupamiento o dispersión de valores similares en el espacio es menos pronunciada.

En cada gráfico se pueden distinguir dos líneas verticales: una línea roja que representa el valor observado del índice de Moran, y una línea azul que representa el valor esperado del índice de Moran. En los años 2013 a 2017, se observa que la línea azul (valor esperado) es inferior a la línea roja (valor observado), lo cual indica una autocorrelación espacial positiva. Por otro lado, en los años 2018 a 2020, ocurre lo contrario, donde la línea azul es superior a la línea roja, lo que indica una autocorrelación espacial negativa.

También se estudian y se realizan todos los p-valor simulados, todos son mayores que 0.05, estos resultados hacen llegar a la conclusión de que los resultados observados podrían ser atribuibles al azar o al azar espacial. No hay suficiente evidencia para concluir que existe una asociación espacial significativa entre los valores analizados.

En resumen, cuando la autocorrelación espacial es débil y la distribución del Índice de Moran es casi simétrica, esto implica que las separaciones, nulidades y divorcios entre provincias muestran una agrupación o dispersión espacial poco significativa. Los valores observados y esperados son cercanos, lo que sugiere que los patrones espaciales en los datos son consistentes con una distribución aleatoria o al azar.

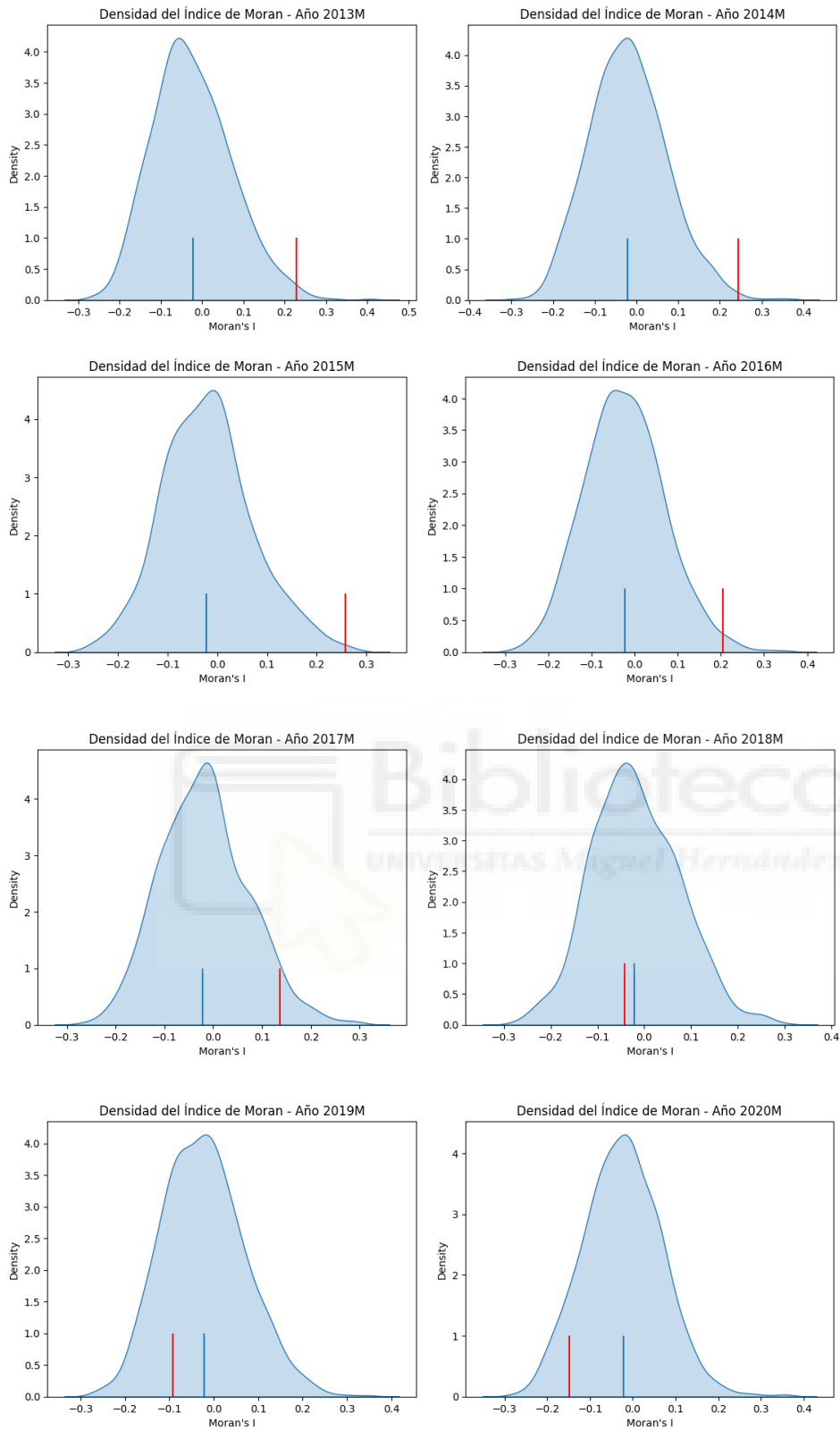


Figura 9. Densidades del Índice de Moran. "Fuente: elaboración propia"

4.2.5. Moran Scatterplot y puntos calientes y fríos

Se va a realizar un estudio individualizado de cada año para sacar unos resultados más específicos en este último apartado.

Es muy importante describir de nuevo el significado de punto caliente y punto frío, porque si no creará muchas dudas al verlo en el gráfico.

Un punto "caliente" (hotspot) se refiere a una ubicación geográfica donde los valores de una variable específica son significativamente altos en comparación con sus vecinos. Indica una agrupación espacial de valores altos, lo que implica una autocorrelación espacial positiva. En otras palabras, los puntos calientes representan áreas donde ocurren eventos o condiciones excepcionalmente altas.

Por otro lado, un punto "frío" (coldspot) se refiere a una ubicación geográfica donde los valores de una variable específica son significativamente bajos en comparación con sus vecinos. Indica una agrupación espacial de valores bajos, lo que implica una autocorrelación espacial negativa. Los puntos fríos representan áreas donde ocurren eventos o condiciones excepcionalmente bajas.

Dicho esto, se vuelve a explicar debido a que los puntos fríos pueden ser las provincias que tienen mayor tasa en los gráficos descriptivos, esto se debe a que aunque tengan mayores tasas, puede ser que su vecino tenga una tasa superior, entonces se esté considerando punto frío, en cambio un punto caliente puede tener menor tasa, pero tener a su alrededor tasas menores y convertirse en punto caliente.

Año 2013

En la Figura 10, se presenta el Moran Scatterplot correspondiente al año 2013, donde se observa que los puntos dispersos están alejados de la línea de regresión, lo cual indica una falta de una tendencia clara de agrupamiento o dispersión espacial de los valores de la variable en estudio. Esto sugiere que existe poca autocorrelación espacial en este año en particular.

Sin embargo, al examinar los puntos calientes y fríos en el mapa de clusters de Moran local, se puede identificar cierto patrón espacial. Los puntos calientes, que representan valores más altos que sus vecinos, tienden a agruparse en la zona del Norte, mientras que los puntos fríos, que representan valores más bajos, se agrupan en la zona del Mediterráneo. Aunque esta agrupación no es extremadamente significativa, sugiere una posible distribución espacial de los valores de la variable.

En resumen, en el año 2013, al examinar el Moran Scatterplot, no se observa una autocorrelación espacial clara. Sin embargo, al analizar el mapa de clústeres de Moran local, se identifica una tendencia hacia la agrupación espacial en ciertas regiones geográficas. Se puede apreciar que los puntos fríos se encuentran más concentrados en relación con la línea de regresión. Esto se refleja en el gráfico de puntos calientes y fríos, donde los puntos están más agrupados, lo que indica la posible existencia de clústeres geográficos con valores similares en la tasa analizada.

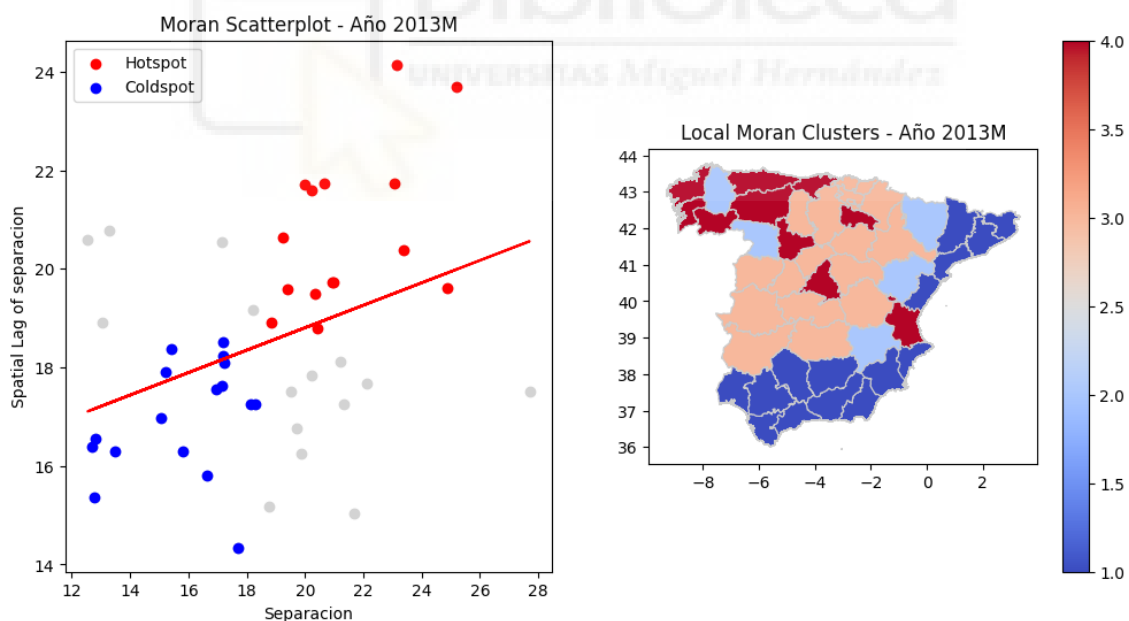


Figura 10. Moran Scatterplot y puntos calientes y fríos de 2013. "Fuente: elaboración propia"

Año 2014

En cuanto a a interpretación del año 2014, en la que se puede apreciar una mayor dispersión con relación a los puntos calientes y fríos respecto del año 2013. Los puntos calientes, que indican valores más altos, se encuentran ampliamente dispersos en el mapa, sin presentar una agrupación espacial clara. Por otro lado, los puntos fríos, que representan valores más bajos, tienden a concentrarse en las regiones de Andalucía y Barcelona. En el caso de los puntos fríos, se puede considerar la existencia de una posible agrupación espacial.

No obstante, es importante señalar que la autocorrelación espacial general en este año sigue siendo débil, ya que los puntos en su conjunto no presentan una tendencia clara con relación a la línea de regresión. Aunque se observa cierta agrupación en los puntos fríos, esta no es suficientemente pronunciada para afirmar una autocorrelación espacial significativa.

En conclusión, en el año analizado se evidencia una mayor dispersión en los puntos calientes, mientras que los puntos fríos muestran una agrupación más notoria en las regiones mencionadas. Sin embargo, es importante tener en cuenta que la autocorrelación espacial global sigue siendo débil, ya que los puntos en su conjunto no siguen una tendencia clara.

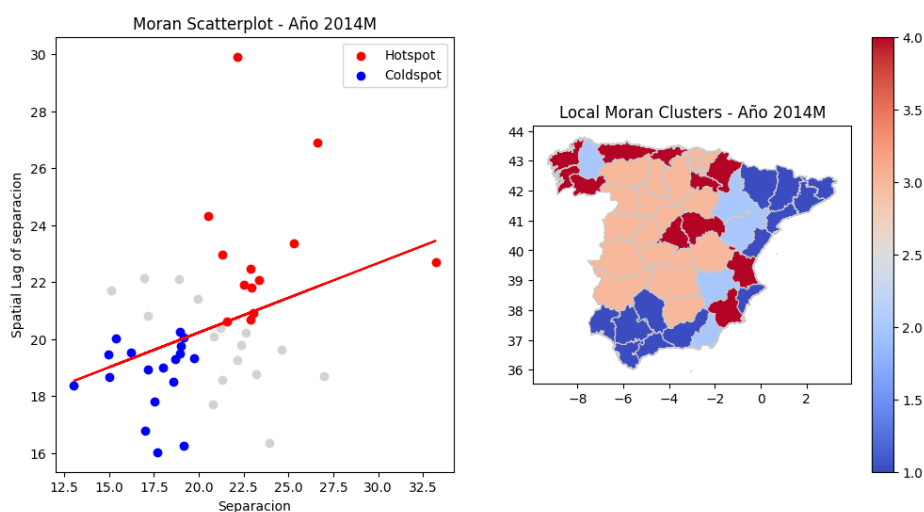


Figura 11. Moran Scatterplot y puntos calientes y fríos de 2014. "Fuente: elaboración propia"

Año 2015

En el análisis correspondiente al año 2015, se observa nuevamente que la región de Andalucía y Barcelona exhibe una concentración de valores bajos en comparación con sus vecinos cercanos, lo cual se clasifica como puntos fríos en términos de autocorrelación espacial. Por otro lado, los puntos calientes se encuentran dispersos y no muestran una agrupación significativa.

Esto indica que en este año no se puede afirmar con suficiente confianza que exista una autocorrelación espacial significativa en la distribución de los valores analizados. Aunque se identifican puntos fríos en las áreas mencionadas, la falta de agrupamiento evidente en los puntos calientes sugiere que la distribución de valores altos no presenta una asociación espacial fuerte.

Estos hallazgos son relevantes para comprender la configuración espacial de los fenómenos estudiados en el año 2015 y proporcionan información para evaluar posibles patrones o procesos subyacentes que puedan influir en la distribución geográfica de los datos.

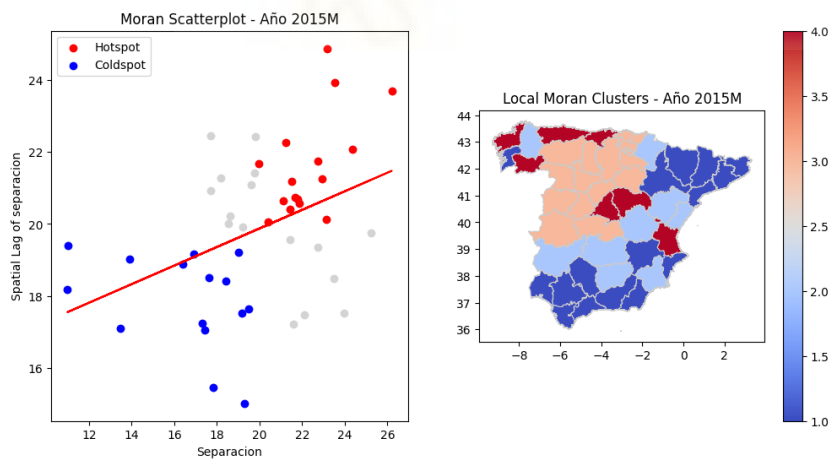


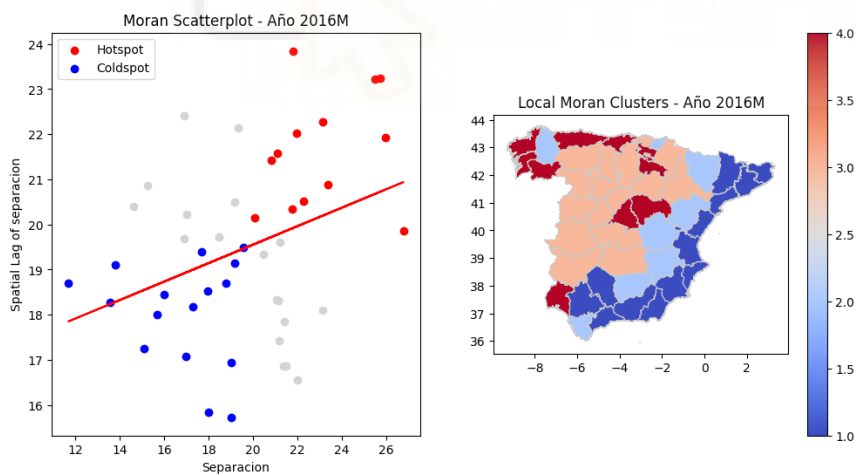
Figura 12. Moran Scatterplot y puntos calientes y fríos de 2015. "Fuente: elaboración propia"

Año 2016 y 2017

Los años 2016 y 2017 se estudian conjuntamente debido a su similitud en cuanto al Moran Scatterplot y la distribución de puntos fríos y calientes. En ambos años, se observa una tendencia positiva en la recta de regresión del Moran Scatterplot, lo que indica una autocorrelación espacial positiva a priori.

En el caso de los puntos fríos, se encuentran ubicados principalmente en la zona del Mediterráneo y suelen estar agrupados en estas áreas. Por otro lado, los puntos calientes están dispersos por todo el mapa de España, sin una concentración espacial clara.

Estos patrones espaciales consistentes en ambos años sugieren la existencia de una estructura espacial similar en la distribución de los datos. La presencia de puntos fríos localizados en el Mediterráneo y su agrupación indica una tendencia a la baja en esas regiones, mientras que la dispersión de los puntos calientes indica valores más altos distribuidos de manera más aleatoria en todo el territorio.



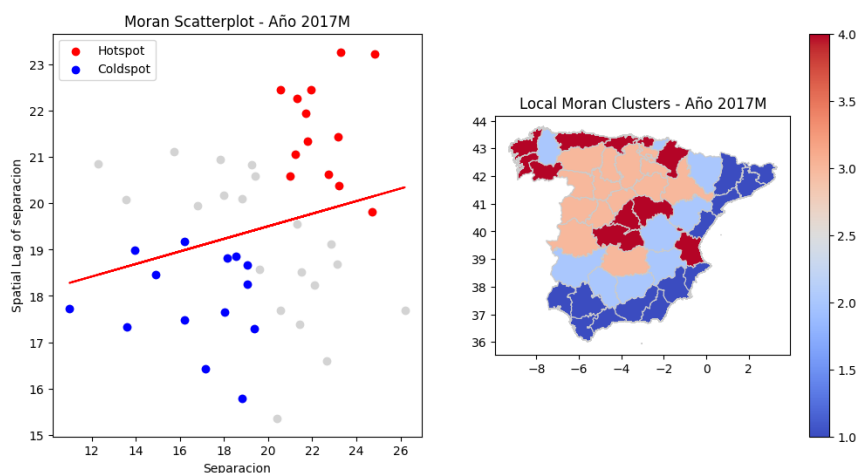


Figura 13. Moran Scatterplot y puntos calientes y fríos de 2016 y 2017. “Fuente: elaboración propia”

Año 2018, 2019 y 2020

En los años 2018, 2019 y 2020 se observa un cambio en la tendencia de los datos, ya que la autocorrelación espacial comienza a ser negativa. Además, los patrones de puntos fríos y calientes se mantienen similares a los años anteriores hasta el año 2018, pero a partir de 2019 se produce una dispersión de estos puntos por todo el mapa de España. En esta etapa, los puntos fríos y calientes se mezclan sin seguir ninguna agrupación o patrón discernible.

Estos resultados indican que, a partir de 2019, la distribución espacial de los valores analizados se vuelve aleatoria y no muestra una estructura clara de agrupación o tendencia espacial. Esto implica que los valores altos y bajos se distribuyen de manera más aleatoria en todo el territorio español, sin concentraciones significativas en áreas específicas.

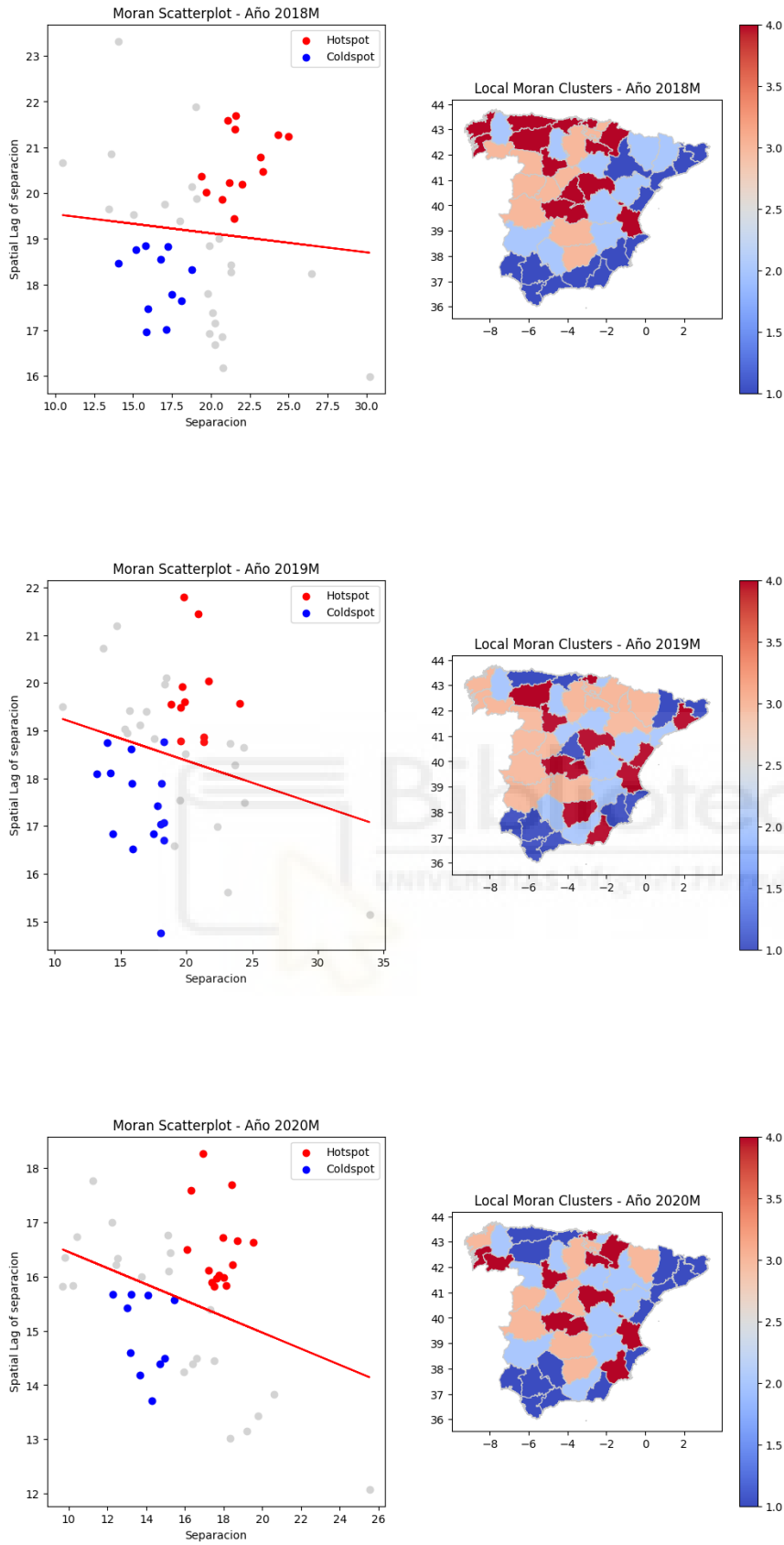


Figura 14. Moran Scatterplot y puntos calientes y fríos de 2018,2019 y 2020. “Fuente: elaboración propia”

5. CONCLUSIONES

Para concluir, este estudio ha analizado la evolución de las separaciones, divorcios y nulidades en España desde 2013 hasta 2020. Se destaca el arduo trabajo de lectura, limpieza y transformación de los datos realizado en la metodología. En cuanto a los resultados del análisis descriptivo, se observa que el año 2020 fue excepcional debido a la pandemia de COVID-19 y las medidas de confinamiento, lo que provocó una disminución significativa en los registros. Excluyendo este año, se observa un descenso gradual en el número de divorcios, separaciones y nulidades desde 2013 hasta 2019, además de unas mayores tasas por la zona del Mediterráneo en la mayoría de años estudiados.

Es importante señalar la escasa presencia de nulidades en España, lo cual hace que este tipo de registro sea menos relevante en el análisis. Sin embargo, se ofrece una breve explicación de esta situación.

Respecto a la agrupación espacial, se llevaron a cabo diversos estudios utilizando herramientas como el Spatial Lag, el caso binario, el índice de Moran, el Moran scatterplot y los puntos fríos y calientes. En general, los resultados indican que los datos se distribuyen de manera aleatoria cada año. Aunque en algunos estudios se observa cierta tendencia en la zona del Mediterráneo, como en el caso binario o en ciertos años de los puntos fríos y calientes, al realizar el índice de Moran se concluye que la autocorrelación espacial no es significativa en ningún año.

En conclusión, los resultados sugieren que no existe una autocorrelación espacial significativa en los datos de separaciones, divorcios y nulidades en España durante el periodo estudiado, tienden a tener una distribución aleatoria por lo general.

En conclusión, este Trabajo Fin de Grado me ha brindado la valiosa oportunidad de adquirir conocimientos en el ámbito de los mapas geográficos y su implementación en Python, así como de aprender a adaptar la geometría para su visualización. Además, he obtenido experiencia en la descarga de datos del

Instituto Nacional de Estadística (INE), en el manejo de repositorios en GitHub y en el uso de aplicaciones como Netlify para la visualización de gráficos interactivos. Por último, he adquirido nociones básicas de estadística espacial relacionadas con el análisis de agrupamiento espacial. Este trabajo ha sido una experiencia enriquecedora que me ha permitido ampliar mi conocimiento y habilidades en diferentes áreas, integrando la geografía, la programación, la estadística y la visualización de datos.



6. BIBLIOGRAFÍA

1. INE - Instituto Nacional de Estadística. (s. f.). *Nulidades, separaciones y divorcios por provincias*. Serie desde 2013(20173). INE. <https://www.ine.es/jaxiT3/Datos.htm?t=20173#!tabs-grafico>
 2. INE - Instituto Nacional de Estadística. (s. f.-b). *Nulidades, separaciones y divorcios por provincias*. Serie desde 2013(20173). INE. <https://www.ine.es/jaxiT3/Datos.htm?t=20173#!tabs-tabla>
 3. *PyPI · The Python Package Index*. (s. f.). PyPI. <https://pypi.org/>
 4. *Spatial Data — Geographic Data Science with Python*. (s. f.). https://geographicdata.science/book/notebooks/03_spatial_data.html
 5. INE - Instituto Nacional de Estadística. (s. f.). *INE. Instituto Nacional de Estadística*. INE. <https://www.ine.es/index.htm>
 6. Iberley. El valor de la confianza. (2022, 14 enero). *Regulación de la nulidad matrimonial, separación y divorcio*. Iberley, Información legal. <https://www.iberley.es/temas/regulacion-nulidad-matrimonial-separacion-divorcio-59552>
 7. Iberley. El valor de la confianza. (2023, 2 marzo). *Real Decreto de 24 de julio de 1889 por el que se publica el Código Civil. - TITULO IV. DEL MATRIMONIO*. Iberley, Información legal. https://www.iberley.es/legislacion/real-decreto-24-julio-1889-publica-codigo-civil-7499513?ancla=118043#ancla_118043
 8. *¿Qué es Python? - Explicación del lenguaje Python - AWS*. (s. f.). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/python/>
 9. *Google Colaboratory*. (s. f.). <https://colab.research.google.com/>
 10. Dani. (2021, 28 enero). *extractor_ine/ine.py at main · dani537/extractor_ine*. GitHub. https://github.com/dani537/extractor_ine/blob/main/ine.py
 11. *How Spatial Autocorrelation (Global Moran's I) works—ArcGIS Pro | Documentation*. (s. f.). <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>
- Agrupación de datos*. (s. f.-b) https://help.highbond.com/helpdocs/analytics/141/user-guide/es/Content/analyzing_data/summarizing_data/grouping_data_overview.htm

12. Amesquita, D. (2022). ¿Qué es el I de Moran? (Definición y ejemplo). Statologos.
<https://statologos.com/morans-i/>
13. *Spatial Joins*. (s. f.-b). LOST. https://lost-stats.github.io/Geo-Spatial/spatial_joins.html
14. Anselin, L. (s. f.). Local Spatial Autocorrelation (1).
https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html#objectives



7. ANEXOS

A continuación, se muestra un enlace de Github, donde se encuentra todo el código usado en Google Collaboratory, con todos los procedimientos llevados a cabo en este trabajo

[tfg_bueno/3ficheros \(1\).ipynb at main · Gancho66/tfg_bueno \(github.com\)](https://github.com/Gancho66/tfg_bueno/blob/main/tfg_bueno/3ficheros%20(1).ipynb)

