

Vivienda

Ferran Garcia

2024-01-29

Contents

1	Resumen	2
2	Introducción	2
3	Objetivos	2
4	Metodología	3
5	Análisis descriptivo	4
6	Análisis de la autocorrelación espacial	5
6.1	I de Moran global	5
6.2	I de Moran local	7
6.3	Getis-Ord local	8
6.4	Geary local	9
7	Conclusiones	10
8	Annexo	11
8.1	Gráficos: Local Moran	11
8.2	Gráficos: Getis-Ord	14
8.3	Gráfico: Geary	17
9	Bibliografía	18

1 Resumen

Un gran número de factores influyen en el precio de una vivienda: Tamaño, situación del mercado, localización... entre otros. ¿En qué medida está relacionado el precio de una vivienda con la cantidad de transacciones que ocurren y el uso que se les da a las mismas viviendas? Investigamos las relaciones que existen entre el parque de viviendas en términos de viviendas principales y no principales, las transacciones realizadas y el precio medio según la provincia española. Por otra parte tratamos de agrupar las provincias de características similares.

2 Introducción

Desde hace siglos en España han tenido lugar múltiples éxodos rurales en dirección a otras regiones. Dichos movimientos demográficos internos implican una mayor necesidad de alojar personas en núcleos ya de por sí muy poblados. Nos planteamos cuál es el estado actual de las provincias en cuanto a las viviendas.

Ponemos en práctica técnicas estadísticas con un enfoque espacial para identificar relaciones entre provincias y agruparlas en clústers homogéneos. El estudio de la situación del mercado de la vivienda a nivel provincial servirá para conocer mejor los fenómenos económicos y demográficos relacionados con las viviendas y proporcionará información valiosa para los agentes de este mercado.

Para llevar a cabo el análisis usaremos R y algunas de sus librerías que implementan algoritmos usados en *Ciencia de Datos Espacial*. A grandes rasgos, primero buscamos identificar autocorrelación con índices globales y locales para luego agrupar en clústers las provincias con características similares. En el apartado metodológico proporcionamos una explicación más detallada, así como la presentación de las variables que usamos.

3 Objetivos

Este proyecto tiene dos objetivos principales: Por un lado el autor pretende familiarizarse con el área de estudio geográfico orientada a los datos y por otro conocer mejor la situación del mercado de viviendas a nivel provincial. A éstos objetivos del proyecto se le adhieren otros secundarios como servir de ejemplo o guión para consultas futuras sobre el tema o identificar qué provincias las viviendas cambian de mano con más frecuencia, a mayor precio o más pobladas en relación a la cantidad de viviendas. Los detallamos en la lista que sigue:

- Familiarizarse con la *Ciencia de Datos Espacial*:
 - Guión para el tratamiento de datos espaciales
 - Guión para la representación de datos espaciales
 - Guión para el uso de librerías de datos espaciales en R
 - Guión para el análisis espacial exploratorio
- Conocer la coyuntura del mercado de viviendas a nivel provincial a través del análisis de autocorrelación espacial y la formación de clústers, responder a:
 - ¿Qué provincias están más pobladas en relación al parque de viviendas disponibles?
 - ¿Qué diferencias y similitudes hay en el uso¹ de viviendas?
 - ¿En qué provincias hay más transacciones de viviendas?
 - ¿En qué provincias hay más ejecuciones hipotecarias?
 - ¿Qué relaciones de cada variable existen entre provincias?
 - ¿Se pueden clasificar o agrupar las viviendas? ¿Según qué criterio?

¹Uso de vivienda como residencia principal o secundaria.

Por otra parte comprobamos la veracidad de algunas hipótesis sobre el conjunto de datos:

1. Las provincias de los alrededores de Madrid sin costa tienen menores precios y menos habitantes
2. Las provincias de costa o turísticas tienen precios más elevados
3. Las provincias más pobladas en relación su parque de viviendas transaccionan viviendas con mayor frecuencia

4 Metodología

Nuestro objetivo es estudiar la situación de las provincias españolas en relación al mercado de la vivienda. El proyecto se estructura en tres partes análisis: Descriptiva estándar, autocorrelación espacial y clústers con restricciones espaciales. Para la implementación de las diferentes técnicas usaremos R y sus librerías. A continuación detallamos cada parte del proceso y luego presentamos el conjunto de datos. Trabajaremos con un nivel de significación del 5%.

Primeramente realizamos un análisis descriptivo de las variables. Este contiene una breve discusión sobre las relaciones entre ellas dejando de lado la dimensión geográfica. Representamos la matriz de correlaciones y presentamos los estadísticos resumen habituales para cada variable.

Seguidamente calculamos los índices de autocorrelación espacial habituales. Usamos la I de Moran para describir la autocorrelación espacial global y varios índices locales: Moran local, Getis-Ord y Geary local multiariante. Los representamos gráficamente en un mapa de las provincias.

Utilizamos cinco variables asociadas al año 2022² y las características geográficas de las provincias, las enunciamos juntamente con su fuente:

- Datos vectoriales de las provincias, ArcGis
- Número de transacciones de viviendas, Ministerio de Interior
- Precio por metro cuadrado del suelo urbano, Ministerio de Interior
- Número de ejecuciones hipotecarias, INE
- Población, INE

²El año 2022 es el más reciente en el que todos los datos están disponibles.

5 Análisis descriptivo

Primeramente veamos como es la distribución de las variables. Recogemos los estadísticos resumen habituales en las siguientes tablas:

Table 1: Estadísticos resumen de las variables

	Media	Mediana	Desviación típica	Rango
Transacciones de viviendas	526.32	304.00	540.38	2214.00
Precio del metro cuadrado urbano	132.94	119.93	65.44	294.48
Viviendas principales	376.53	256.24	467.43	2551.36
Viviendas secundarias	158.54	132.26	111.86	556.36
Población	946.37	644.35	1214.13	6654.92
Viviendas en proceso de ejecución hipotecaria	342.42	176.50	451.02	1919.00

Unidades de medida:

Parque e viviendas en miles, población en miles, precio en euros y el resto en unidades.

Parece que la todas las variables presentan asimetría positiva, lo que significa que existen algunas observaciones extremas de valores altos que arrastran la media hacia arriba. Por otro lado la dispersión del número de transacciones, ejecuciones hipotecarias y viviendas principales es notoriamente alto teniendo en cuenta que la población es la variable con mayor desviación típica: En todos estos casos la desviación típica supera a la media y la mediana.

Veamos ahora la correlación pura entre las variables, sin tener en cuenta la dimensión espacial de las mismas. Es aparente que existe una correlación relativamente fuerte entre todas las variables, sin embargo el precio parece estar menos relacionado que el resto.

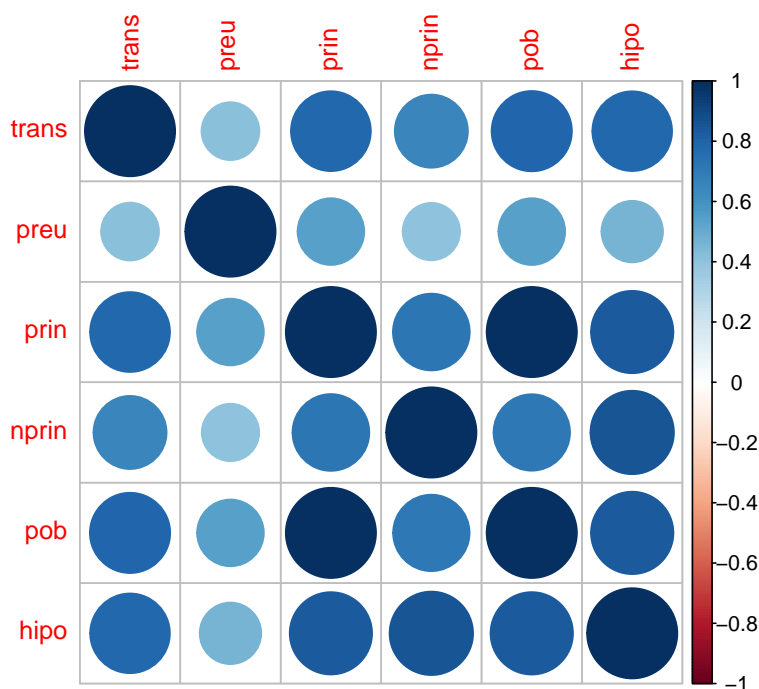


Figure 1: Los nombres de las variables se refieren a transacciones de viviendas, precio por metro cuadrado urbano, parque de viviendas principales, parque de viviendas no principales, población y ejecuciones hipotecarias en curso.

6 Análisis de la autocorrelación espacial

6.1 I de Moran global

Calculamos la I de Moran usando el paquete `ape` con las distancias implementadas por `sf`. Reescalamos el resultado entre -1 y 1 y realizamos la prueba de hipótesis: H_0 : La variable se distribuye aleatoriamente en el espacio contra H_1 : La variable no se distribuye aleatoriamente en el espacio

Lo hacemos para todas las variables del conjunto de datos:

Para la variable población no podemos rechazar H_0 así que el valor ligeramente negativo no es significativo, lo que sugiere una distribución estándar aleatoria.

```
ape::Moran.I(shp$pob, d, scaled = TRUE)[c(1, 4)] # población
```

```
## $observed
## [1] -0.01756222
##
## $p.value
## [1] 0.8563559
```

En este caso podemos rechazar la distribución aleatoria de las transacciones de vivienda y el valor negativo sugiere cierta tendencia hacia la dispersión.

```
ape::Moran.I(shp$trans, d, scaled = TRUE)[c(1, 4)] # transacciones
```

```
## $observed
## [1] -0.05829364
##
## $p.value
## [1] 0.01269392
```

La variable precio ajusta un *p-value* aún menor y también es negativo, algo más que el anterior. De modo que el precio tiende a la dispersión espacial ligeramente más que las transacciones.

```
ape::Moran.I(shp$preu, d, scaled = TRUE)[c(1, 4)] # precio
```

```
## $observed
## [1] -0.1008098
##
## $p.value
## [1] 1.020782e-07
```

De nuevo obtenemos un resultado parecido para el número de ejecuciones hipotecarias. Parece que están dispersas en una medida menor al precio y mayor a las transacciones.

```
ape::Moran.I(shp$hipo, d, scaled = TRUE)[c(1, 4)] # ejecuciones hipotecarias
```

```
## $observed
## [1] -0.08449281
##
## $p.value
## [1] 2.561257e-05
```

La autocorrelación global del número de viviendas principales no es significativa. Es consistente con el resultado obtenido para la población puesto que presumiblemente son dos variables relacionadas estrechamente.

```
ape::Moran.I(shp$prin, d, scaled = TRUE)[c(1, 4)] # viviendas principales
```

```
## $observed
## [1] -0.01591663
##
## $p.value
## [1] 0.7749252
```

En cambio las viviendas no principales sí son significativas y como el resto de variables apunta hacia una tendencia dispersa.

```
ape::Moran.I(shp$nprin/shp$prin, d, scaled = TRUE)[c(1, 4)] # viviendas no principales
```

```
## $observed
## [1] -0.05220966
##
## $p.value
## [1] 0.03486071
```

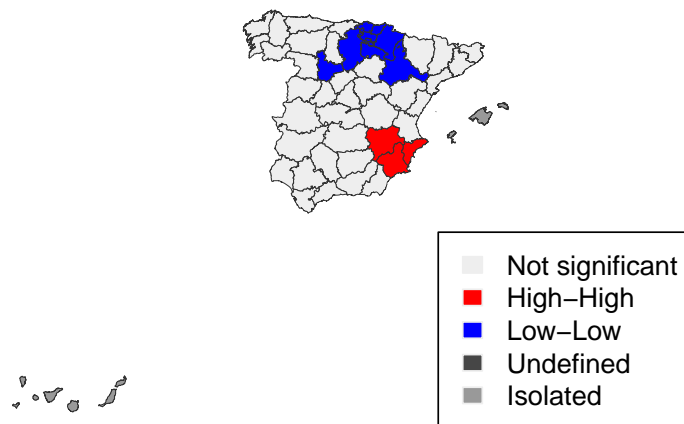
En general, la mayoría de las variables presentan un índice de Moran significativo con tendencia a la dispersión. Las variables con mayor índice son aquellas relacionadas más estrechamente con el mercado inmobiliario: Transacciones, precio y viviendas secundarias. Por otro lado las variables más estables como población y viviendas principales no son significativas. Una posible interpretación es que el mercado de la vivienda es selectivo de una forma no fácilmente agrupable. En la sección de análisis clúster con restricciones espaciales estudiaremos esta cuestión en más profundidad.

6.3 Getis-Ord local

En cuanto al índice de Getis-Ord la idea e interpretación son similares, la diferencia es que este índice solo señala categorías *Low-Low* o *High-High* así que la información que se puede extraer del mismo es más limitada.

Veamos un ejemplo con la variable viviendas no principales, los gráficos para el resto de variables se encuentran en el anexo. En este caso el índice identifica que hay más viviendas no principales, en relación con el resto de provincias, en Murcia, Alicante o Albacete. Por contra indica que hay menos viviendas principales hacia el norte, en el País Vasco, parte de Aragón y parte de Castilla y León.

Mapa de Getis-Ord de viviendas no principales



6.4 Geary local

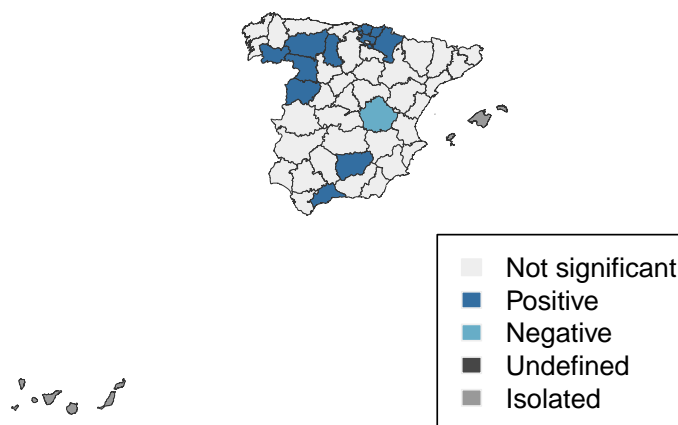
El propósito de éste índice es el mismo que los dos anteriores. Basándose en la medida para cada provincia y el *spatial lag*³ asociado a ésta. De nuevo hay distintas clasificaciones posibles para las observaciones significativas: *Positive* corresponde a provincias con una medida y un *spatial lag* superiores a la media. El concepto es análogo para la categoría *Negative*. Por otra parte tenemos las provincias que no cuya medida no va de la mano con sus vecinos, éstas van al grupo *Undefined*. Finalmente se separan las *Isolated*, aquellas que no tienen vecinos. La definición de vecindad es de suma importancia en casos como el nuestro, con islas. Sin embargo queda fuera del ámbito de este proyecto el estudio de definiciones de vecindad. Veamos un ejemplo, en el anexo hay otras combinaciones de variables:

Transacciones, precio y población:

Intuimos que existe cierta relación entre la frecuencia de transacciones y el precio, así como con el número de habitantes de una provincia. Veamos si algunas zonas comparten medidas en cuanto a estas tres variables.

Cuenca es la única que tiene un perfil negativo, es decir está poco poblada y tiene tanto precios como frecuencia de transacciones de viviendas bajas. Este es el caso para Cuenca y sus provincias vecinas. Por otro lado identificamos por la zona del País Vasco y Navarra clasificaciones positivas, también en Málaga. Hasta ahora los resultados tienen cierto sentido en el contexto español: Las provincias en se clasifican como altas y las que no lo están mucho como bajas. Sin embargo notamos que la parte oeste de Castilla y León también ha sido clasificada como alta, algo que a primera vista no esperaríamos. Será necesario indagar más en esta cuestión en investigaciones posteriores. Quizás cambiando la definición de contiguidad, aumentando el número de iteraciones o transformando las variables obtengamos otros resultados.

Mapa Geary: Transacciones, precio y población



³El lag espacial es la suavización de la medida atendiendo a los vecinos, en un caso simple es la media de la medida sobre la entidad y sus colindantes.

7 Conclusiones

En este trabajo hemos plantado la cuestión de identificar similitudes entre provincias respecto a medidas de autocorrelación espacial. Para ello hemos utilizado datos de *ArcGis*, el *Ministerio de Interior* y el *INE*. Los hemos procesado utilizando *R* y nos hemos servido de los numerosos recursos recogidos en la bibliografía.

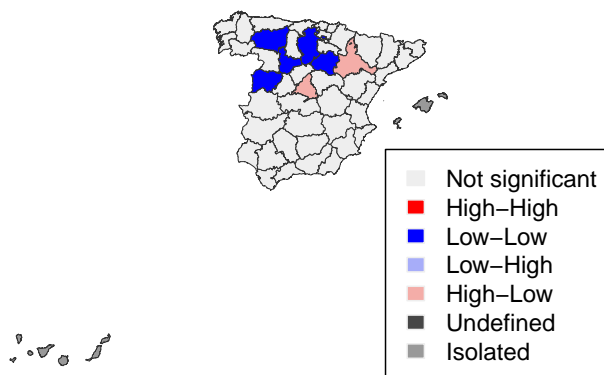
Comprobamos que las variables que hemos usado son asimétricas hacia la derecha y observamos una correlación importante entre todas ellas, aunque con el precio notablemente más baja. Hemos constatado que la autocorrelación espacial es significativa a un nivel del 5% para algunas variables como las transacciones de viviendas o el precio de éstas en sentido de dispersión espacial. Por contra lo hemos descartado para otras como la población o el número de viviendas principales. Finalmente calculamos y representamos los índices de autocorrelación local. Con ello hemos sido capaces de identificar potenciales cabezas de clúster y outliers en diversos sentidos usando índices univariantes y en sentidos complejos empleando un índice multivariante.

En resumen, nos hemos familiarizado con los conceptos fundamentales del análisis espacial y hemos aprendido a manejar los datos y librerías más relevantes en *R*. Las próximas direcciones de estudio deberán trabajar sobre estas bases indagando en las definiciones y detalles capaces de mejorar la capacidad del análisis en el ámbito de la *Ciencia de Datos Espacial*.

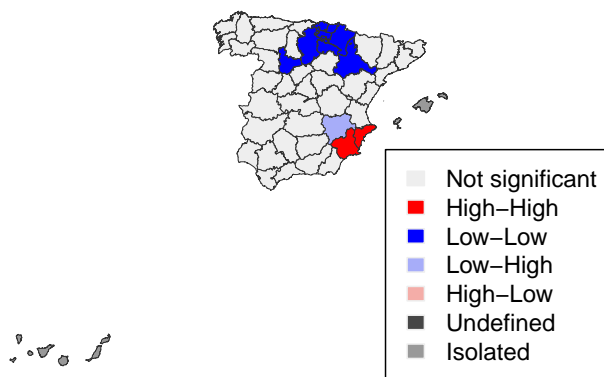
8 Anexo

8.1 Gráficos: Local Moran

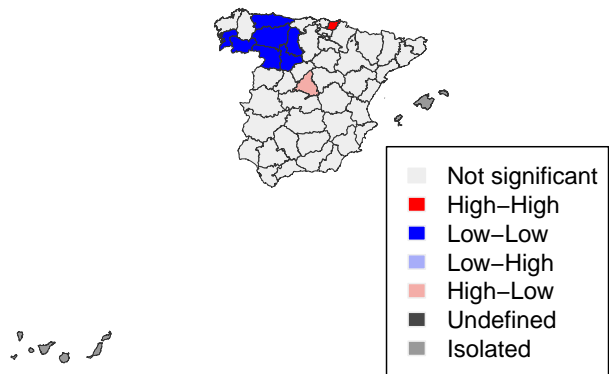
Mapa de Local Moran de viviendas principales



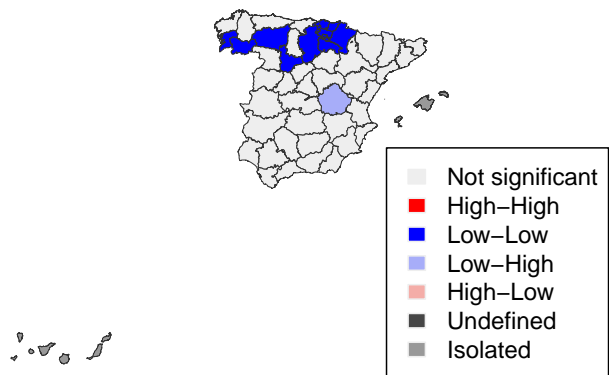
Mapa de Local Moran de viviendas no principales



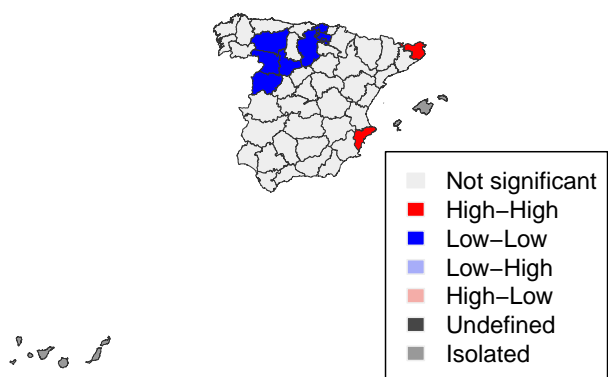
Mapa de Local Moran de precio



Mapa de Local Moran de transacciones

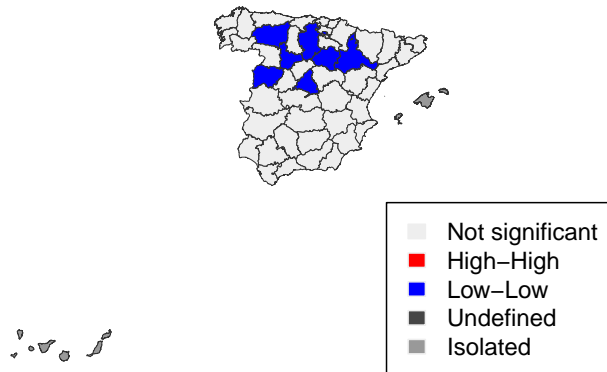


Mapa de Local Moran de ejecuciones hipotecarias

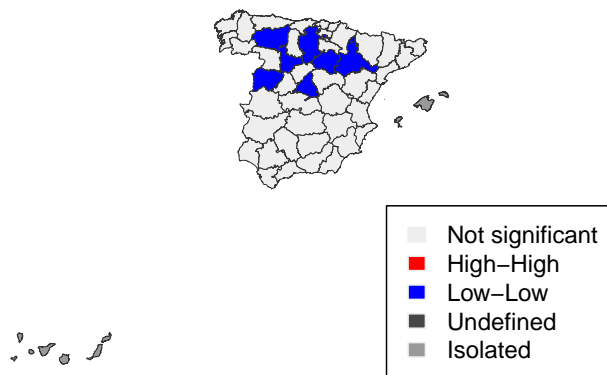


8.2 Gráficos: Getis-Ord

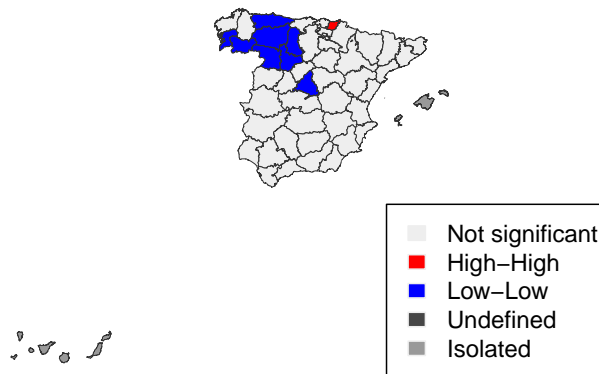
Mapa de Getis-Ord de viviendas principales



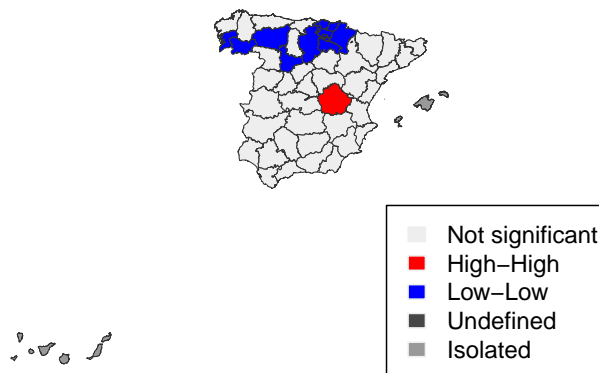
Mapa de Getis-Ord de población



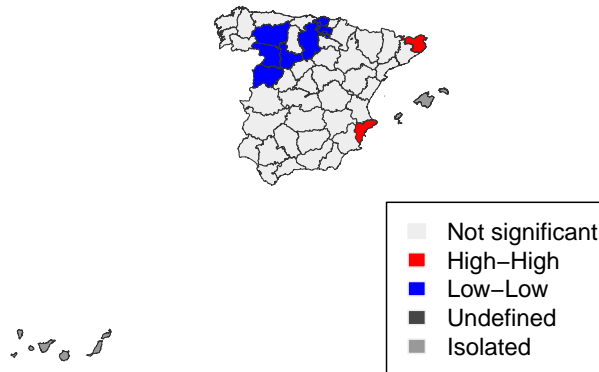
Mapa de Getis-Ord de precio



Mapa de Getis-Ord de transacciones



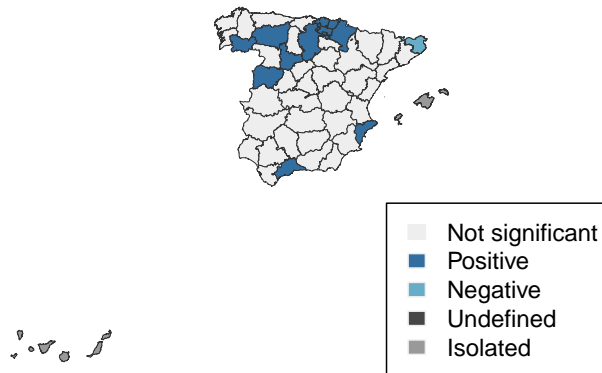
Mapa de Getis-Ord de ejecuciones hipotecarias



8.3 Gráfico: Geary

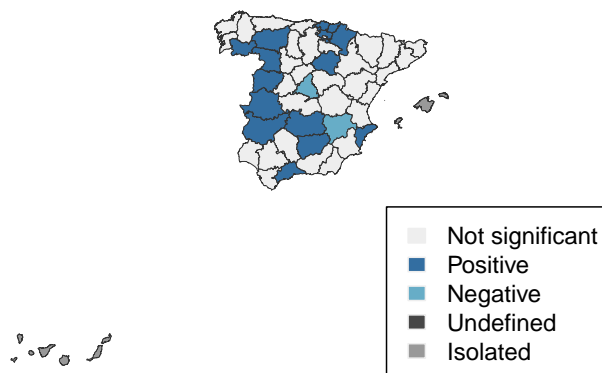
Ejecuciones hipotecarias y transacciones:

Mapa Geary: Ejecuciones hipotecarias y transacciones



Viviendas principales y precio:

Mapa Geary: Viviendas principales y precio



9 Bibliografía

- Estimación del parque de viviendas
- Precios del suelo urbano y número de transacciones
- Shapefile de las provincias
- Autocorrelación espacial global I
- Autocorrelación espacial II
- Autocorrelación espacial local
- rgeoda
- I de Moran