HACETTEPE UNIVERSITY

DEPARTMENT OF GEOMATICS ENGINEERING

COURSE: GMT352

GEOGRAPHICAL INFORMATION SYSTEM

Assignment5 - Spatial Autocorrelation

INSTRUCTOR: Berk ANBAROĞLU, Dr.

NAME-SURNAME: Mahir Selek

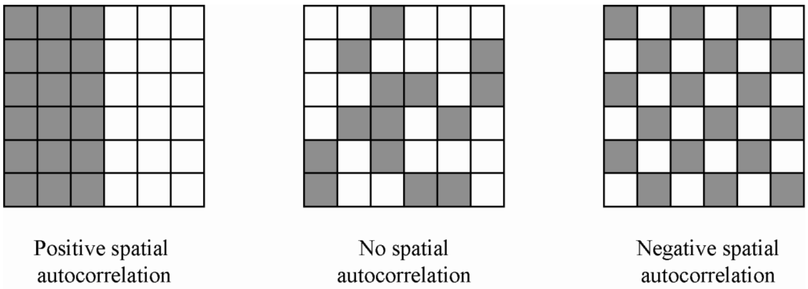
ID: 21532275

SPRING 2020

In Data Science, we tend to explore and investigate data before doing any modelling or processing task. This helps you identify patterns, summarize the main characteristics of the data, or test a hypothesis. The conventional Exploratory data analysis does not investigate the location component of the dataset explicitly but instead deals with the relationship between variables and how they affect each other. Correlation statistical methods are often used to explore the relationship between variables.

In contrast, Exploratory Spatial Data Analysis (ESDA) correlates a specific variable to a location, taking into account the values of the same variable in the neighbourhood. The methods used for this purpose are called Spatial Autocorrelation.

Spatial autocorrelation is describing the presence (or absence) of spatial variations in a given variable. Like, conventional correlation methods, Spatial autocorrelation has positive and negative values. Positive spatial autocorrelation is when areas close to each other have similar values (High-high or Low-low). On the other hand, negative spatial autocorrelation indicates that neighborhood areas to be different (Low values next to high values).



Getting the data

I used the Airbnb dataset (Point dataset) and Layer Super Output Areas — LSOA — neighbourhoods (Polygon dataset) in London for this tutorial. We do spatial join to connect each point of Airbnb listings to neighbourhood areas. listing.csv which I used inside my code

The dataset I use is spatially joined Airbnb properties in London with an average price of properties in each local area (Neighbourhood).

Here are the first 5 rows of the Average prices of Airbnb properties in London :

	LSOA_CODE	LSOA_NAME	price	geometry
0	E01000001	City of London 001A	156.928571	POLYGON ((532050.879 181817.674, 532021.188 18...
1	E01000002	City of London 001B	183.307692	POLYGON ((532267.748 181643.784, 532254.565 18...
2	E01000003	City of London 001C	135.500000	POLYGON ((532071.310 182159.597, 532135.127 18...
3	E01000004	City of London 001D	188.836858	POLYGON ((531172.252 181124.643, 531133.747 18...
4	E01000005	City of London 001E	144.350649	POLYGON ((533378.878 181459.767, 533439.561 18...

Global Spatial Autocorrelation

Global spatial autocorrelation determines the overall pattern in the dataset. Here we can calculate if there is a trend and summarize the variable of interest. Moran's I statistics is typically used to determine the global spatial autocorrelation.

And I get this number for this dataset **0.54**. What does this number mean? This number summarises the statistics of the dataset, just like the mean does for non-spatial data. Moran's I values range from -1 to 1. In my case, this number provides information that there is a positive spatial autocorrelation in this dataset. Remember that I am determining only the global autocorrelation with Moran's I statistics. It does not tell us where this positive spatial autocorrelation exists.

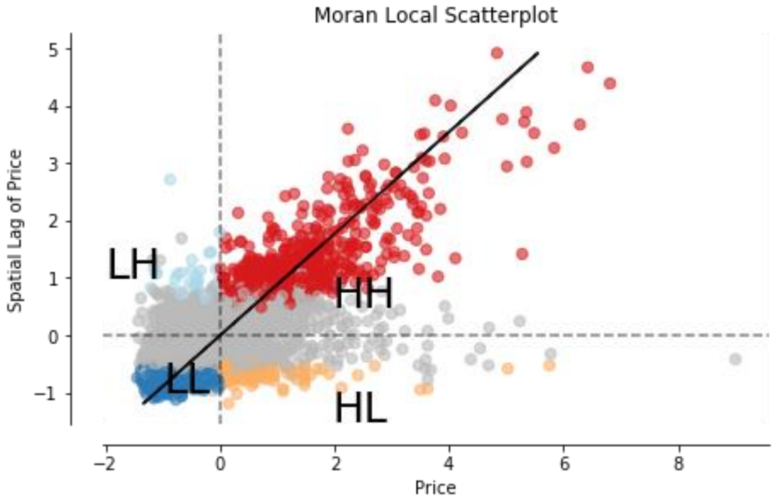
I used Moran's I plot to visualize the global spatial autocorrelation, which is identical to other scatter plots, with a linear fit that shows the relationship between the two variables.

Both Moran's I and Moran's I Scatter plot show positively correlated observations by location in the dataset.

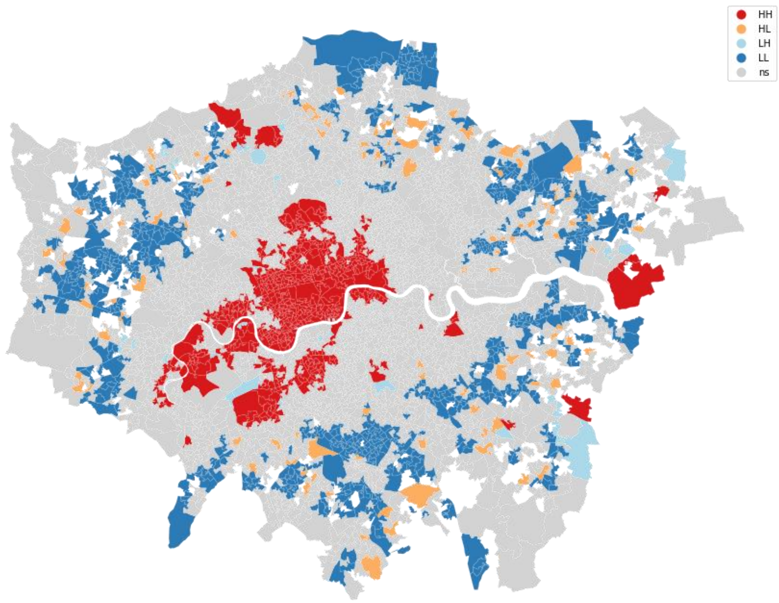
Local Spatial Autocorrelation

So far, I have only determined that there is a positive spatial autocorrelation between the price of properties in neighbourhoods and their locations. But I have not detected where clusters are. Local Indicators of Spatial Association (LISA) is used to do that. LISA classifies areas into four groups: high values near to high values (HH), Low values with nearby low values (LL), Low values with high values in its neighborhood, and vice-versa.

The scatter plot divides the areas into the four groups, as we mentioned.



Now, this is cool, and I can see all values classified into four groups, but the exciting part is to see where these values cluster together in a map. Again, there is a function in Pysal (splot) to plot a map of the LISA results.



The map above shows the variation in the average price of Airbnb Properties. The red colors indicate neighbourhoods clustered together, which have high prices surrounded by high prices as well (mostly the center of the city). The blue areas indicate where prices are low, also surrounded by areas with low-value prices (Mostly peripheries). Equally interesting is also Low-high and High-low area concentration.