Project 6: Spatial Autocorrelation

1. Objective:

The objective of spatial autocorrelation is to conduct analysis of spatial data and measure the relationship among values of a variable according to their spatial relationship. (Chang)

2. Introduction:

Moran’s I is one of the most popular measures of spatial autocorrelation. It computes a multi-directional value of the spatial correlation by examining the expectation of the value I. If the value of I is close to E(I) the spatial pattern is random. If it is greater than E(I) adjacent points (locations) have similar values (spatially correlated) and it is less than E(I) if the adjacent points have different values.

The methods of computing the multi-dimensional link values are the ‘Rook’, ‘Queen’, and ‘Bishop’. The ‘Rook’ version uses the four cardinal directions as its connection points. The ‘Queen’ uses the four cardinal directions and the diagonal connections, the ‘Bishop’ only uses the diagonal connections. There are other methods of spatial autocorrelation such as Geary’s C, however, only Moran’s I is used in this report.

3. Method

The data used is U.S. Census data from 2000 for the state of Indiana. To begin, the spatial connectivity of the counties was calculated using the ‘Queen’ and ‘Rook’ methods of computing the multi-dimensional link values. The connections of each county were computed using the centroids of the county polygon. The ‘Queen’ method produced 480 nonzero links while the ‘Rook’ method produced 468. They both produced 28 counties with 9 links, but the ‘Rook’ method produced two counties with only two connections while the ‘Queen’ method produced only one county with two connections. This difference

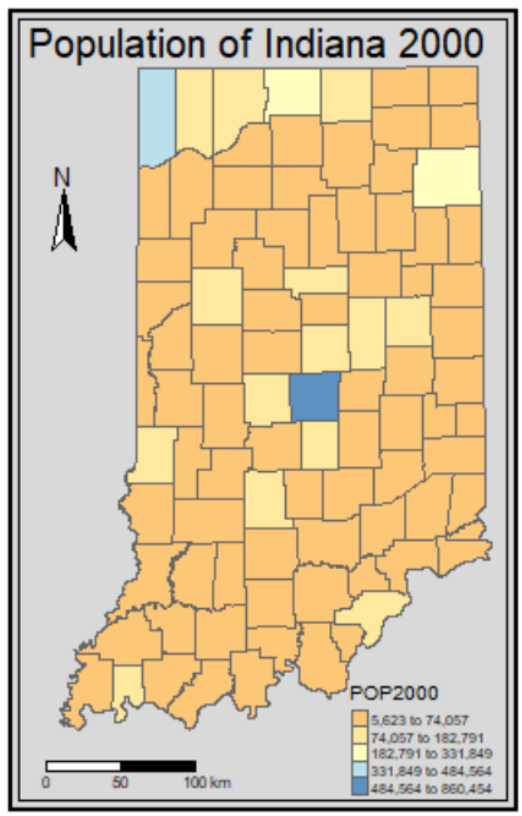

Figure 1: Population distribution of Indiana using Jenks Breaks

can be seen in the adjacent plots in the far northeastern corner of the state. Evaluating the connections of the counties in this manner allows one to familiarize themselves with the overall region and neighborhoods that will be explored spatially using Moran's I.

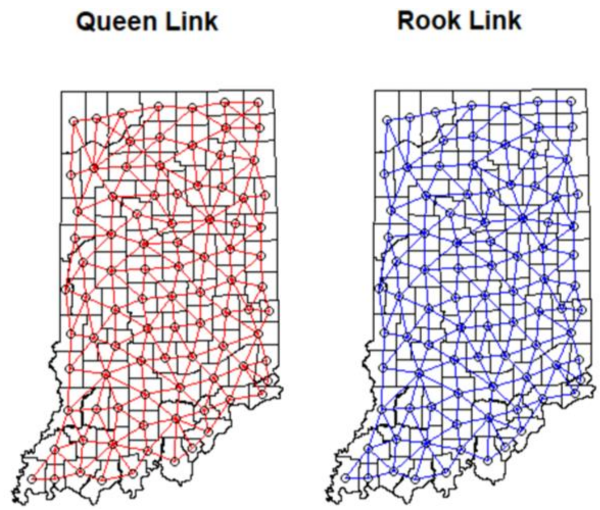

Figure 2: Links of Indiana Counties using the "Queen" and "Rook" methods.

Three weight matrices were used to evaluate the initial Moran's I of the population of the state of Indiana. These matrices were the built-in R weights W, S, and U. The weights were generated using the links computed with the 'Queen' method and the function 'nb2listw'. The style was set to either "W", "S", or "U" to calculate the different weights. All three of these weights returned similar Moran's I values and had E(I) values less than I which was initially interpreted to mean there is spatial correlation or clustering among the population of Indiana. However, when inspecting the p-values for the 'moran.test' function when randomization equals both true and false it returns a p-value greater than 0.05 in all instances, meaning there is no spatial autocorrelation. The null hypothesis that the population is randomly distributed is accepted.

Moran's I Statistic for POP2000 with W, S, & U Weights						
Weight	Moran's I	E(I)	Variance	Std. Dev.	P-Value Random=T	P-Value Random=F
W	0.060866814	-0.010989011	0.002700251	1.3828	0.08336	0.1322
S	0.063051853	-0.010989011	0.002581705	1.4572	0.07253	0.1196
U	0.064521027	-0.010989011	0.002545685	1.4966	0.06725	0.1131

Next the two variables, housing units (HSE_UNITS) and vacant housing units, (VACANT) where selected from the Census data and evaluated. W is the weight used throughout the evaluation of these variables. Evaluating these two variables also produced results that showed no spatial clustering. The Monte Carlo test on the variable HSE_UNITS returns a rank of 43743 and a p-value of 0.1252. On the variable VACANT it returns a rank of 28620 and a p-value of 0.4276.

Moran's I Statistic for HSE_UNITS & VACANT with W Weights						
Variable	Moran's I	E(I)	Variance	Std. Dev.	P-Value Random=T	P-Value Random=F
HSE_UNITS	0.043970705	-0.010989011	0.002463345	1.1073	0.1341	0.1966
VACANT	-0.011939037	-0.010989011	0.001808446	-0.02234	0.5089	0.5059

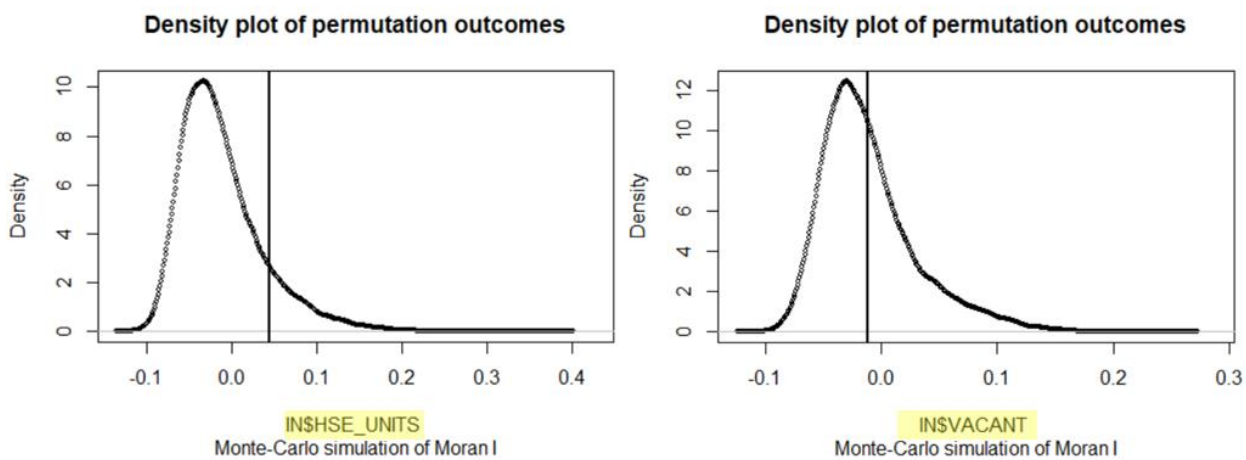

Figure 3: Monte Carlo plots of the variables HSE_UNITS and VACANT

The local Moran's I values where then calculated for each of the variables HSE_UNITS and VACANT. Local Moran's I is also called LISA (Local Indicators of Spatial Association) and it calculates an index value for each feature, point or polygon. (Chang) In the map below, it appears there is some spatial clustering of both housing units and vacant units near the urban centers of the state. These are Indianapolis in the central region and Gary in the northwest. The Moran's I analysis did not reveal any spatial clustering; however, I believe if we evaluated these variables at a smaller level such as census block or tract there would be clustering. The data is dispersed throughout the entire county polygon which dilutes it to some effect. The fact that these variables are clustered around the population centers also makes sense. More people require more housing.

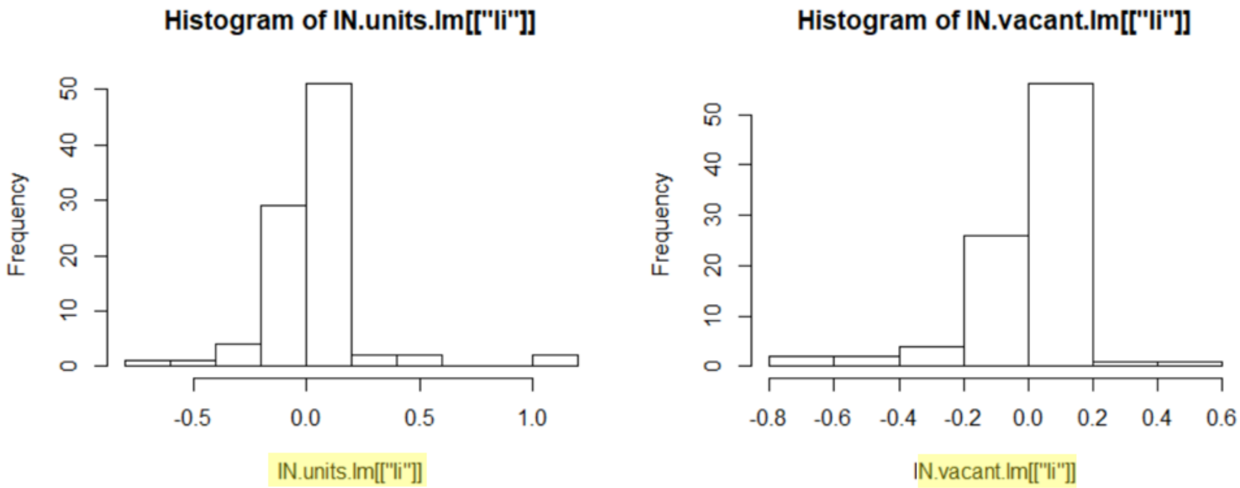

Figure 4: Histograms of the local Moran's I values for HSE_UNITS and VACANT.

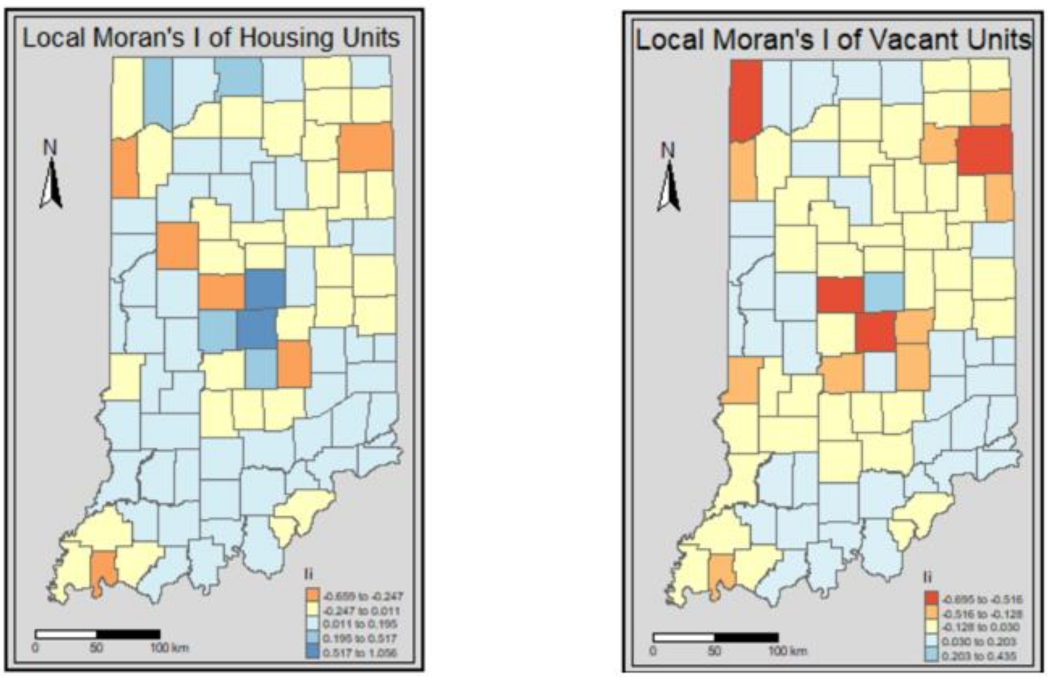

Figure 5: Thematic maps of local Moran's I values for HSE_UNITS and VACANT using Jenks breaks.

4. Conclusion

Moran's I and spatial autocorrelation is a very effective statistic for data analytics and geographic analysis. Often people will look at a thematic and make an inference on clustering and correlation purely on the apparent distribution they see in the map. Moran's I is a rigorous statistical method that can be used to definitively evaluate the spatial relationship of variables or objects.

References:

Chang, K.-Tsung. (2016). *Introduction to geographic information systems*. New York, NY: McGraw-Hill Education.

Appendix:

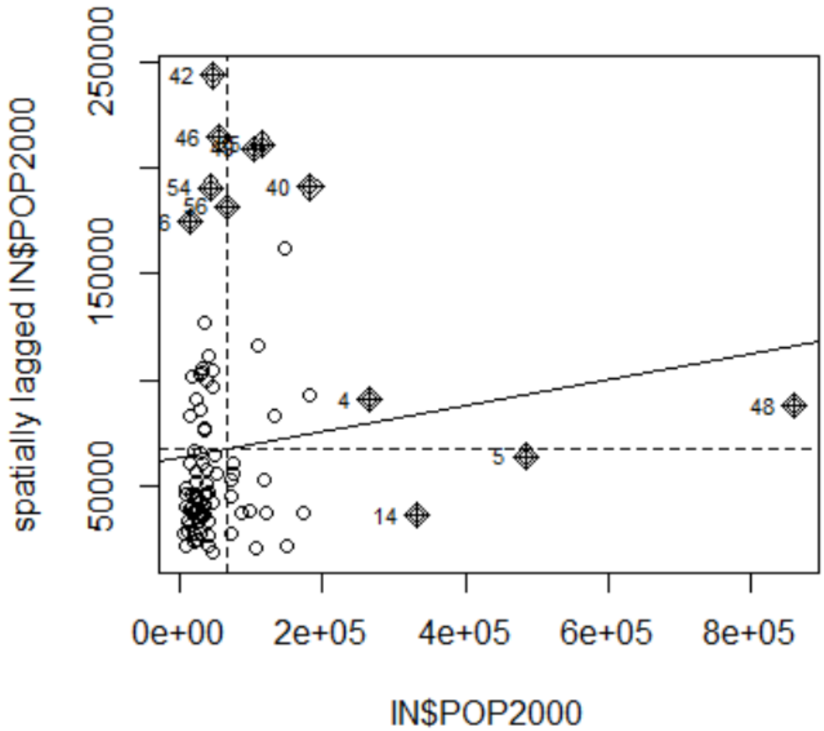

Figure 6: Moran's I plot of the Indiana Population using the W weight.

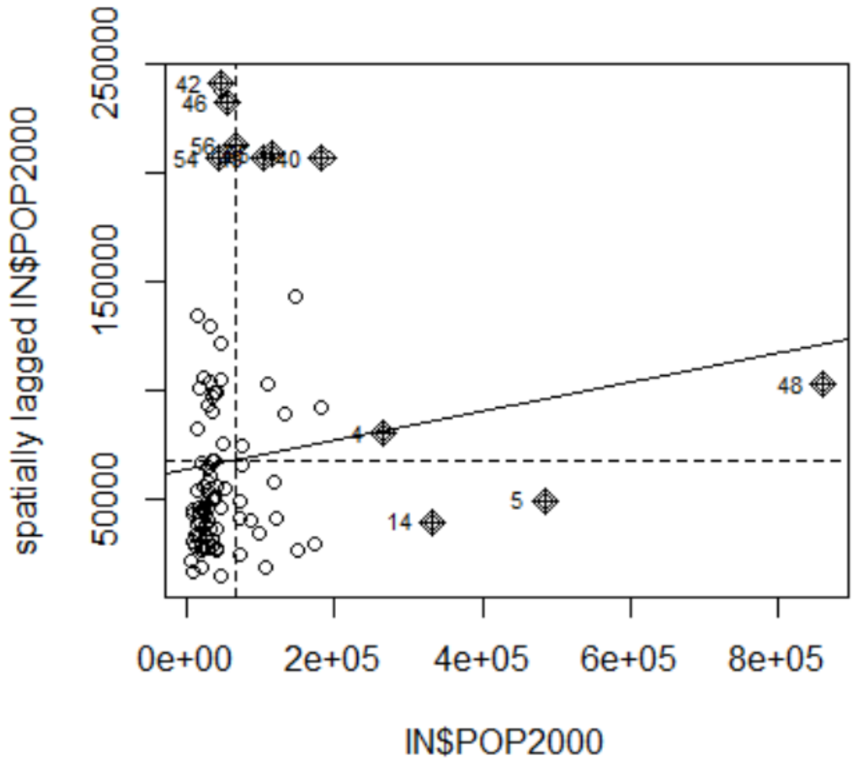

Figure 7: Moran's I plot of the Indiana Population using the S weight.

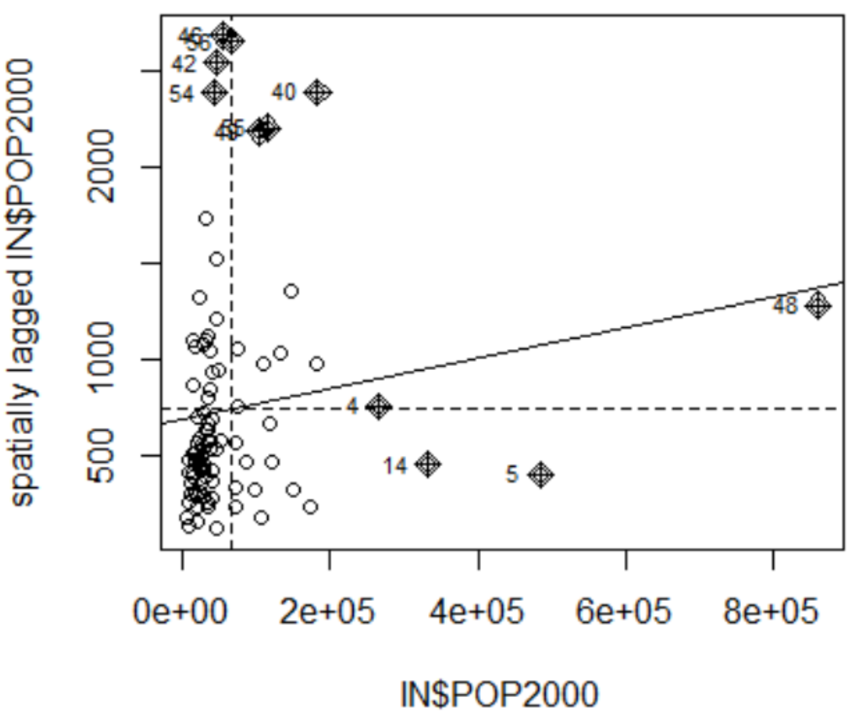

Figure 8: Moran's I plot of the Indiana Population using the U weight.