Spatial Data Science, Homework 3

Chen Liang

Data - Police expenditures, crime and socio-economic characteristics for Mississippi counties, 1982, projected from EPSG:4326 to EPSG: 3435
https://geodacenter.github.io/data-and-lab/police/

Variables:

1) CNTY_ID: distinct county IDs
2) POLICE: Police expenditures per capita, 1982 (dollars per person)
3) WHITE: Percent county residents that are white, 1980

Geographies and the Characteristics of Spatial Weights

1. Graph 1a and 1b show both the projected original geography with quantile POLICE and the Thiessen polygons with a centroid for each Mississippi county.
2. I used three types of spatial weights:
   1) Queen contiguity with both first- and second- order neighbors included
   2) Rook contiguity with only the first order neighbors and without precision threshold
   3) K-nearest neighbor weights with k = 5
3. Comparisons (See Table 1)
   1) The original and Thiessen polygons give very similar spatial weights but are slightly different in the neighbors identified by queen and rook contiguities. As shown in Graph 1c, counties, particular those at the boundary, are stretched significantly in the Thiessen polygon. Thus, for example, Hancock (circled), has two rook neighbors in the original polygon but three after the Thiessen transformation.
   2) Compared to other contiguity weight matrices, the queen one is much denser. This is because it not only includes the corners (counties with common vertices) which the rook one excludes, but also counts two orders of neighbors.

Table 1: Characteristics of different spatial weights

		Min	Max	Mean	Median	%Non-Zero
Original	Queen	6	24	14.63	14	17.85%
	Rook	2	7	4.9	5	5.98%
	kNN	5	5	5	5	6.10%
Thiessen	Queen	7	22	14.54	15	17.73%
	Rook	3	8	5.17	5	6.31%
	kNN	5	5	5	5	6.10%

Global Spatial Autocorrelation

Graph 2.1-2.6 and Graph 3.1-3.6 show the local univariate Moran's I scatterplots and LISA Cluster Maps for the two variables (POLICE and WHITE) across the six weights.

1. As expected, in all pairs of comparisons, the spatial patterns displayed by original and Thiessen polygons are almost identical.
2. Table 1 shows the pseudo p-values for each variable with given spatial weight matrices. With 999 permutations, the most extreme pseudo p-value possible is 0.001, and each run of simulation gives slightly different values. Nevertheless, WHITE clusters have significantly a spatial pattern, with all of the pseudo p-values regardless of weight matrices are at the minimum possible value of 0.001. In other words, WHITE clusters are not sensitive to the specification of weights.
3. POLICE seems to be not quite spatially clustered, but both rook and kNN suggest that some spatial clusters may exist in the western Mississippi. Yet, the signals are confusing: it seems to be a mix of both high-high and low-low. We can understand this mixed pattern through the quantile map in Graph 1a. The highly police-funded counties (Bolivar-Washington and Warren-Hinds) are separated by Issaquena and Sharkey which had almost no police funding. The latter two counties had very few residents and thus probably received less attention from the police. This alternating pattern clearly confuses our contiguity measures, especially when we include the second-ordered neighbors.
4. As a comparison, WHITE is clearly spatially clustered. Eastern Mississippi has two high-high clusters and western Mississippi has a low-low cluster. This observation is robust with one or two order neighbors and across different contiguity measures. Including the second order queen neighbors makes the clusters slightly larger, probably because mid-Mississippi regions have some spatial-outlier counties that strongly influence the average neighbor values.

Table 2: Global Moran's I Statistics (999 permutations)

		POLICE			White		
		Pseudo P	**z-value**	**Moran's I**	**Pseudo P**	**z-value**	**Moran's I**
Original	**Queen**	0.371	0.1863	-0.006	0.001	11.1697	0.444
	Rook	0.048	2.0949	0.103	0.001	7.8289	0.565
	kNN	0.082	1.6326	0.07	0.001	8.5861	0.567
Thiessen	**Queen**	0.441	-0.2761	-0.022	0.001	10.4016	0.421
	Rook	0.074	1.7	0.086	0.001	8.1078	0.566
	kNN	0.086	1.558	0.07	0.001	8.5861	0.567

Bivariate Local Autocorrelation Test

Bivariate Local Moran's I cluster map

1. Graph 5a shows a bivariate Local Moran's I cluster map, which captures the relationship between the value for WHITE at a location and the average of the neighboring values for POLICE. With 99999 permutations, a rook contiguity weight matrix, and a 0.05 significance filter, it seems that there are three high-high cores in Warren, Rankin, and Stone, and a low-low core in Adams. However, none of these clusters pass a 0.01 significance filter (Graph 5b). I also experiment with a kNN or a queen (order = 2) contiguity weight matrix. As shown in Graph 7a-d, the clustering patterns are highly sensitive to both spatial weights and p-value

cutoffs. Therefore, POLITCE and WHITE seems to be not quite spatially autocorrelated. This conclusion is further confirmed by Graph 6, which is a bivariate local Moran's I scatterplot with the parameters stated above.

2. However, with more careful brushing in Graph 8a and 8b, it seems that the lack of a global pattern might be attributed to spatial heterogeneity. For example, in the richer region with more white population (selection A), more white residents correspond to more police expenditure in neighboring counties. In the poorer region with more non-white population (selection B), less white residents correspond to more police expenditure in neighboring counties. With a limited sample size, it is hard to conclude with a significance level. But this heterogenous pattern suggests that police expenditure per capita might be higher either when the non-white population is larger (for crime rates or due to racial discrimination), or the county is rich enough to raise extra tax revenue for police force.
3. Also, as shown in Graph 9, Jackson is a significant spatial outlier with much higher number of CRIME (serious crimes per 100,000 residents) compared to any other county. As the capital city of Mississippi, it has been considered a one of the least-safe cities. In the 1980s, Jackson was also haunted by illegal drug smuggling, with its convenient transport infrastructure. Therefore, the reason that Jackson has more police funding could be quite different from the reason that certain other counties have more police funding.

Table 3. Local Moran's I (WHITE vs Lagged POLICE, 999 permutations)				
		Pseudo P	**z-value**	**Moran's I**
Original	**Queen**	0.248	-0.6719	-0.02
	Rook	0.321	-0.4602	-0.026
	kNN	0.441	0.1631	0.008
Thiessen	**Queen**	0.11	-1.2952	-0.037
	Rook	0.351	-0.3577	-0.019
	kNN	0.441	0.1631	0.008

Bivariate Local Geary Cluster Map

1. The spatial heterogeneity problem appears to be better addressed by a bivariate local Geary test, which measures the extent to which neighbors in multi-attribute space are also neighbors in geographical space. Graph 10a-c show the bivariate local Geary cluster maps with 99999 permutations and a rook contiguity weight matrix.
2. With a reference to the quantile maps, the positive cluster in western Mississippi represents a local cluster of vectors with high POLICE and low WHITE, and the two positive clusters in eastern Mississippi represent two local clusters of vectors with high POLICE and high WHITE.
3. With a 0.05 significance filter, we see that the POLICE-WHITE vectors cluster in the three regions we discussed before. That is, the average distance in attribute space between the values at these regions and the values at their neighbors are significantly smaller. With the decrease in significance cutoffs, the clusters get smaller and fewer.
4. As shown in Graph 11a-d, the clusters are relatively not sensitive to different spatial weights, such as kNN and Queen (order = 2) weights. But a key concern I have is that it seems to

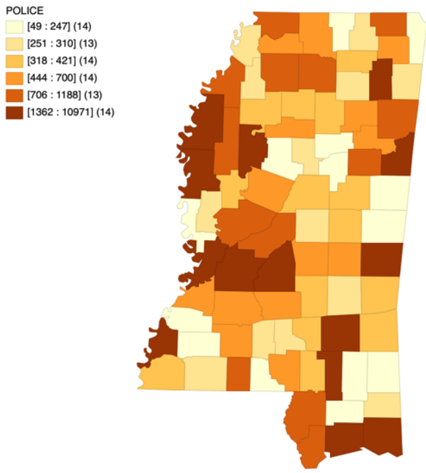
almost replicate the spatial cluster of WHITE. As discussed before, POLICE as a variable itself does not have a strong spatial pattern. Thus, the local bivariate clusters might just represent relatively similar POLICE but strongly different WHITE, such that POLICE-WHITE appears to have a strong spatial implication.

Conclusion

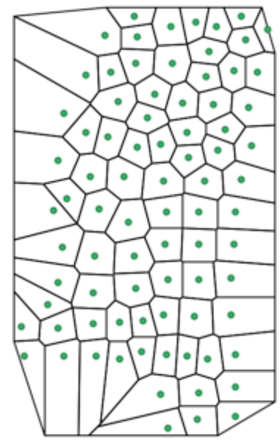
1. It is not clear whether POLICE is spatially clustered. On the contrary, WHITE is significantly spatially clustered, and this pattern is robust to different weights, methods, geography, and p-value cutoffs.
2. In general, POLICE and WHITE seem to be not quite spatially autocorrelated. But it is possible that the two variables have opposite correlations in different regions: negative coefficients in the west and positive coefficients in the east, for example.

Mississippi County Map

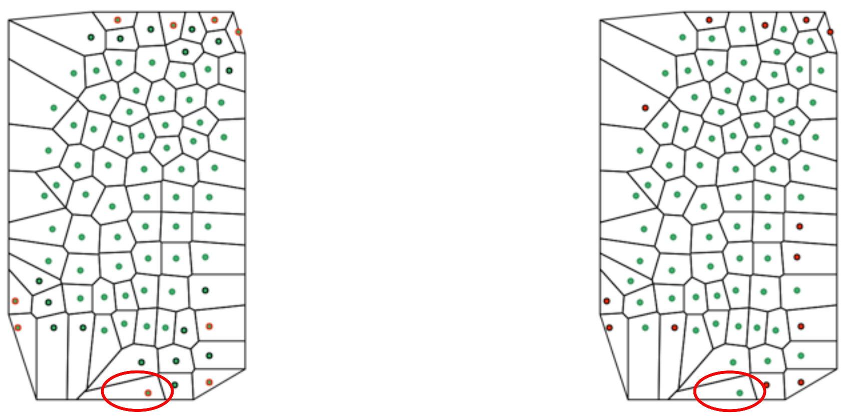
Graph 1a: Original Geography (Quantile POLICE).



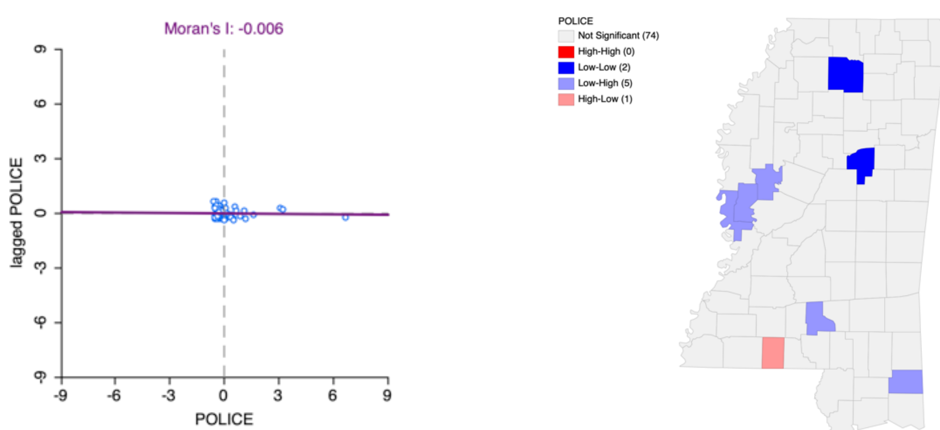
Graph 1b: Centroids and Thiessen polygons



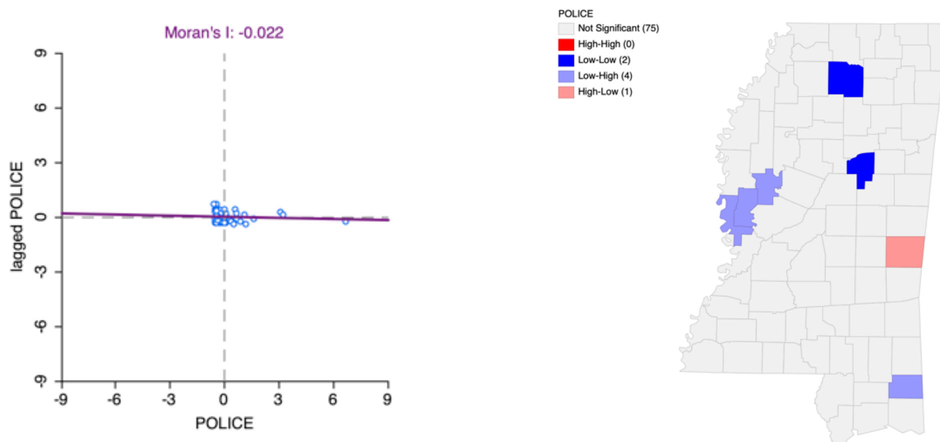
Graph 1c. Selected (red) counties with three Rook neighbors, Thiessen vs Original



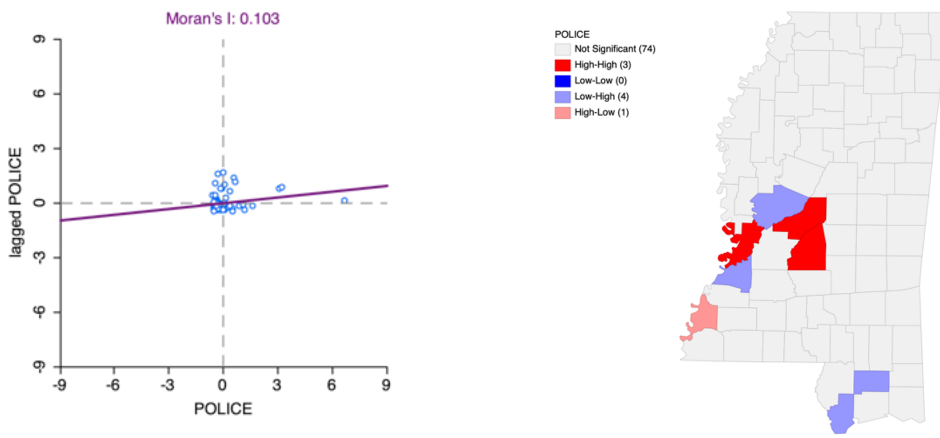
Graph 2.1: Original Graph, Queen (2), POLICE



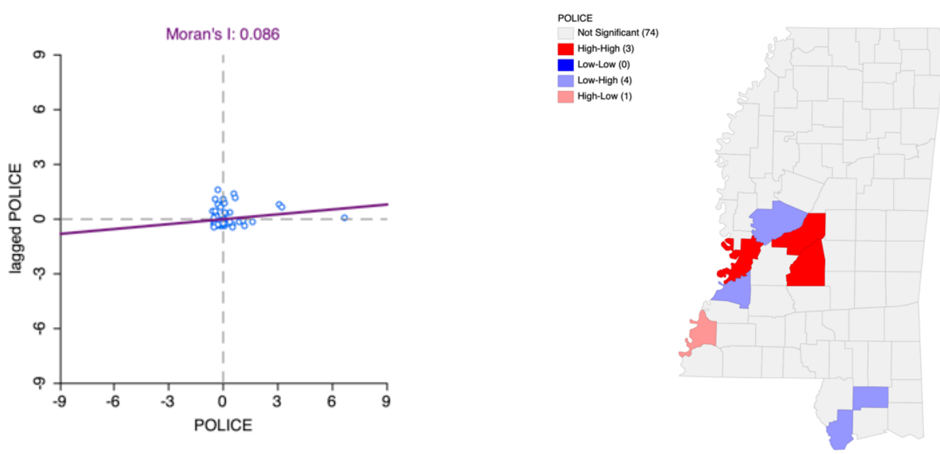
Graph 2.2: Thiessen polygons, Queen (2), POLICE



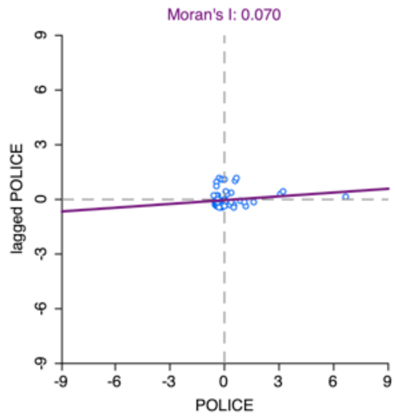
Graph 2.3: Original Graph, Rook (1), POLICE

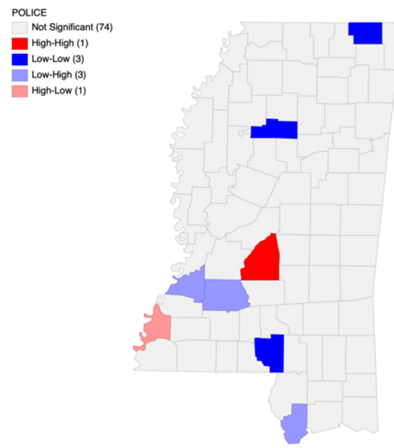


Graph 2.4: Thiessen polygons, Rook (1), POLICE

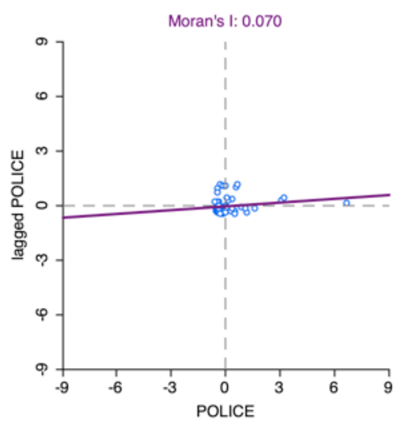


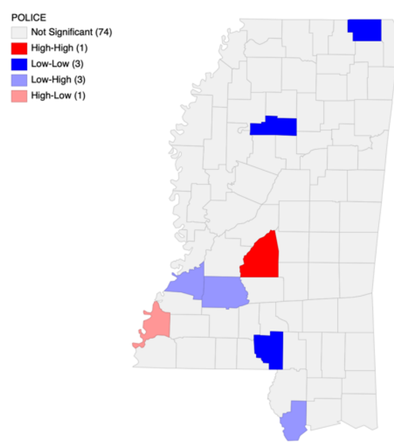
Graph 2.5: Original Graph, kNN (5), POLICE



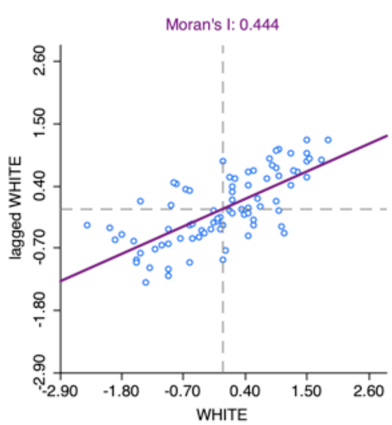


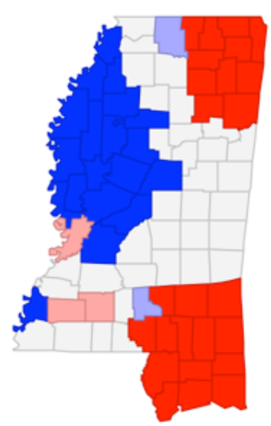
Graph 2.6: Thiessen polygons, kNN (5), POLICE



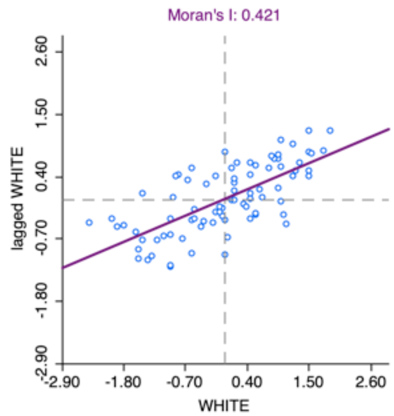


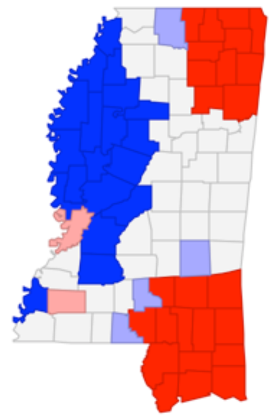
Graph 3.1: Original Graph, Queen (2), WHITE



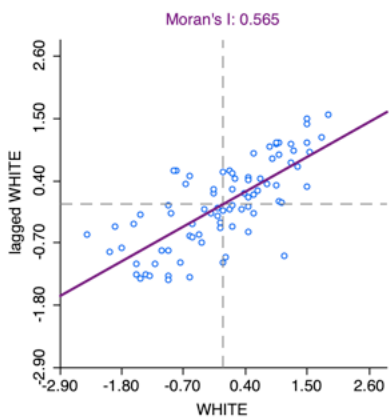


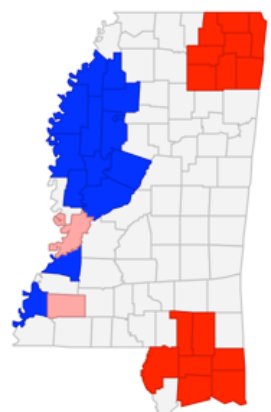
Graph 3.2: Thiessen polygons, Queen (2), WHITE



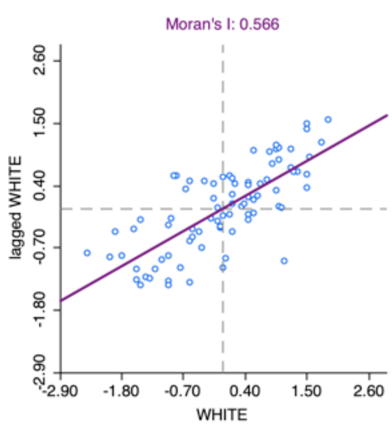


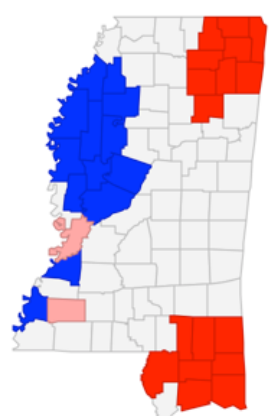
Graph 3.3: Original Graph, Rook (1), WHITE



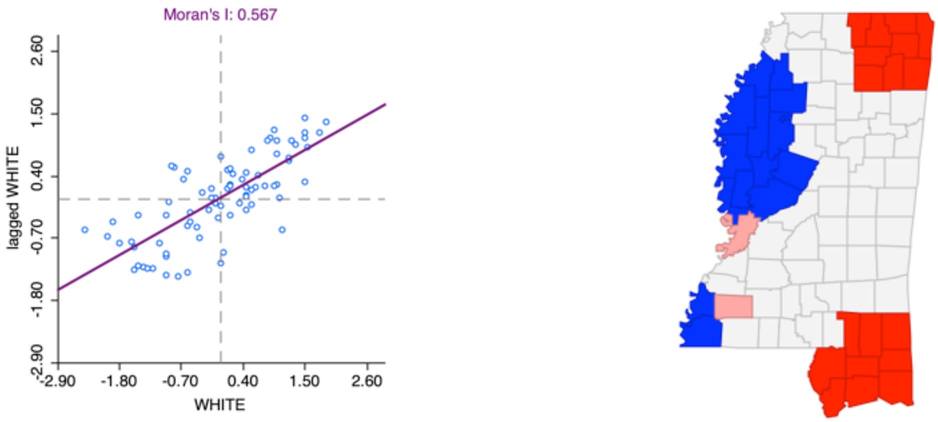


Graph 3.4: Thiessen polygons, Rook (1), WHITE

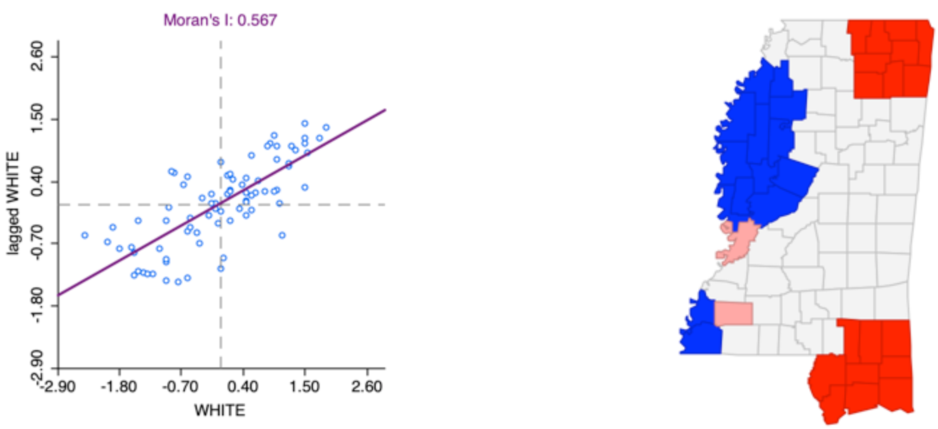




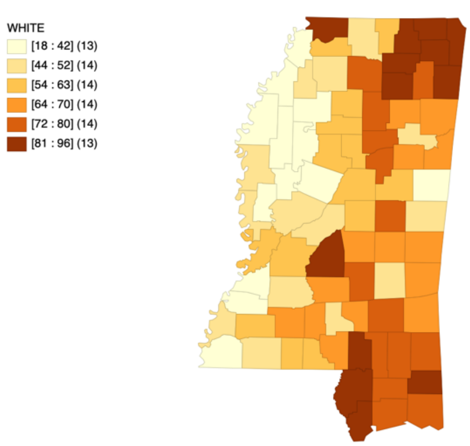
Graph 3.5: Original Graph, kNN (5), WHITE



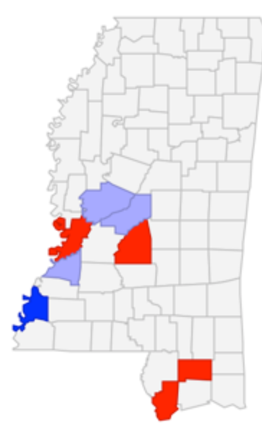
Graph 3.6: Thiessen polygons, kNN (5), WHITE



Graph 4. Quantile WHITE

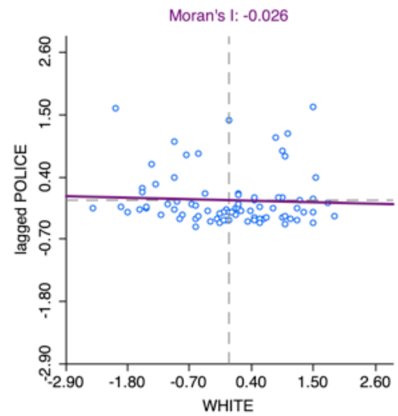


Graph 5a. BiLISA Cluster Map, Rook, alpha = 0.05

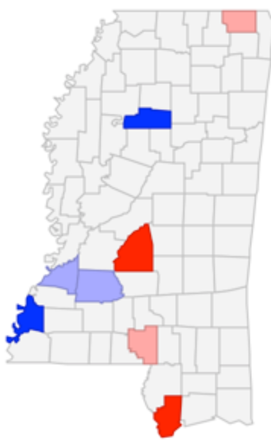


Graph 5b. BiLISA Cluster Map, Rook, alpha = 0.01

Graph 6. Bivariate Local Moran's I (Rook)

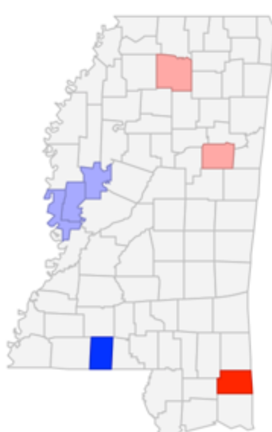


Graph 7a. BiLISA Cluster Map, kNN, 0.05



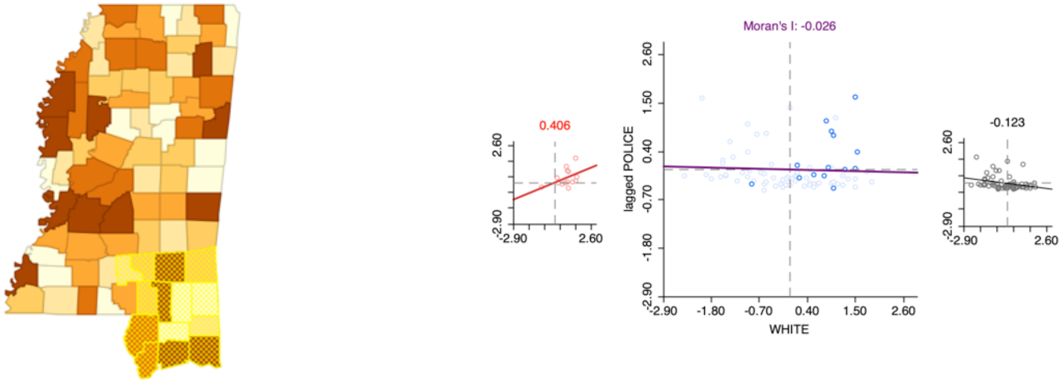
Graph 7b. BiLISA Cluster Map, kNN, 0.01

Graph 7c. BiLISA Cluster Map, Queen, 0.05

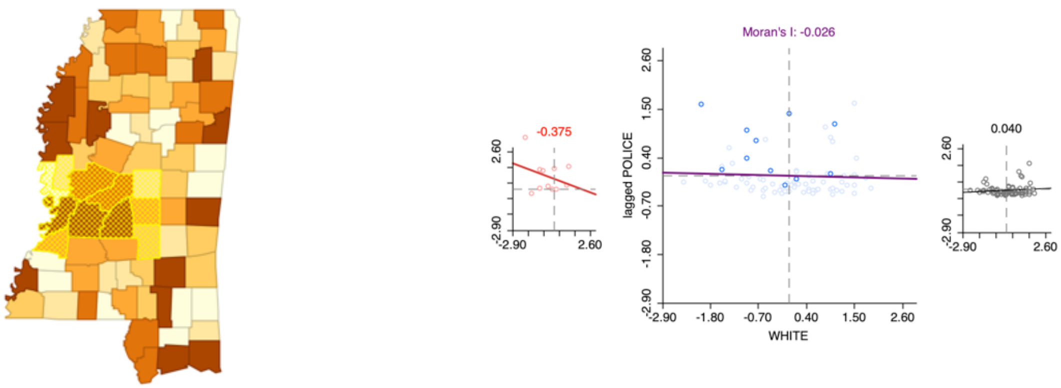


Graph 7d. BiLISA Cluster Map, Queen, 0.01

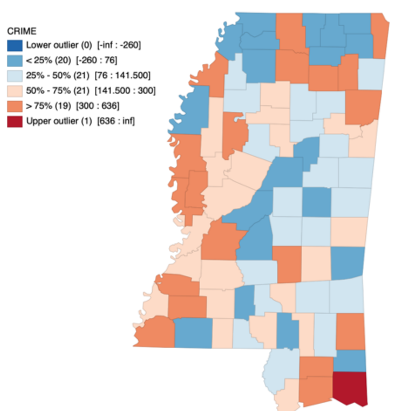
Graph 8a. Selection A (Quantile POLICE)



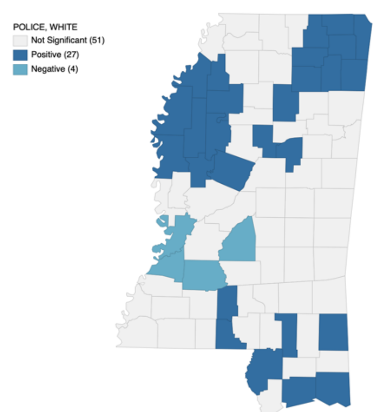
Graph 8b. Selection B (Quantile POLICE)



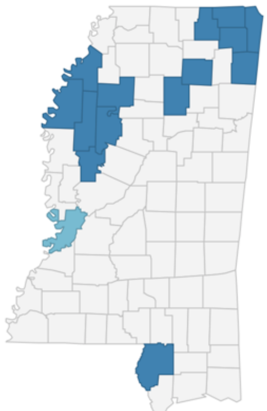
Graph 9. CRIME, Box plot (hinge = 1.5)



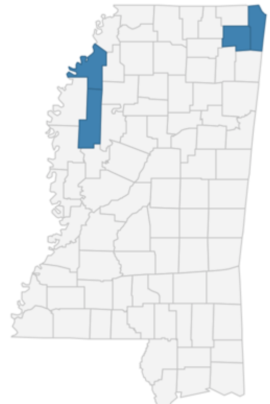
Graph 10a. Local Geary, alpha = 0.05 (Rook)



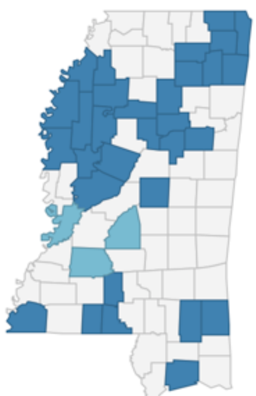
Graph 10b. Local Geary, alpha = 0.01 (Rook)



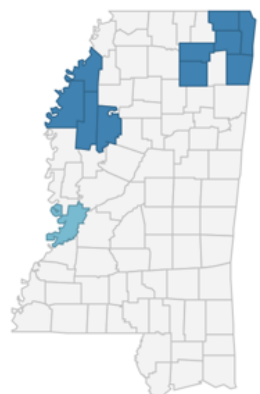
Graph 10c. Local Geary, alpha = 0.001 (Rook)



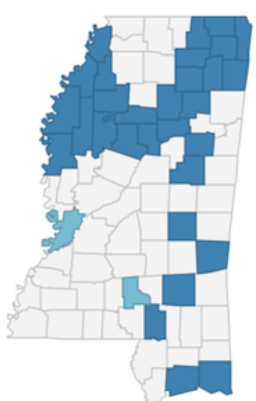
Graph 11a. Local Geary, alpha = 0.05 (kNN)



Graph 11b. Local Geary, alpha = 0.001 (kNN)



Graph 11c. Local Geary, alpha = 0.05 (Queen)



Graph 11d. Local Geary, alpha = 0.001 (Queen)