

## **Project 06. Spatial Autocorrelation of Census Data**

### **Abstract**

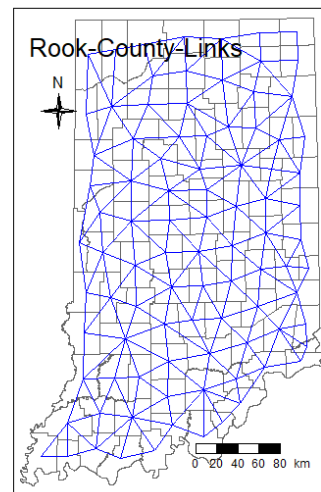
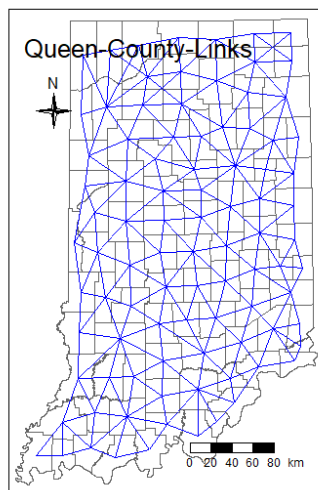
This project conducted spatial autocorrelation analysis on Indiana Census Data at County level. Three attributes were selected: POP2000, MED\_AGE, and HISPANIC, the study showed that most of the distributions are random. The Tippecanoe County level were then selected for the tract level analysis, while most of the results are similar to the State level analysis.

### **1. Data Selection of your interest attributes**

The first step of the project is to input the Indiana Census Data into the program and then select three attributes for subsequent sections. After reviewing all the attributes, I have chosen POP2000 (which is mandatory), MED-AGE (indicating median age), and HISPANIC (representing the number of Hispanic individuals). A screenshot of the column names is provided in the reference.

### **2. Spatial Connectivity evaluation**

The next step of this project is to plot the connection of counties under “Queen” and “Rook” connections. There are a total of 92 regions (counties) in the state of Indiana. For the Queen's Case, there are 480 links. For the Rook's Case, there are 468 links. The links represent the total number of relationships or connections under two methods. For “Queen” cases, on average, each region in this case has approximately 5.217391 connections with other regions. The distribution indicates variances in connectivity; for instance, while 1 region has only 2 links, another region connects with as many as 9 others. Region 90 is identified as the least connected, having only 2 links, while region 28 emerges as the most connected with 9 links. The following map shows the connection of counties under both methods.



For Rook's case (the right map above), with a slightly lesser number of nonzero links (total 468). This means each region, on average, is connected to about 5.086957 others. The connectivity varies across regions. Region 90 remains the least connected with 2 links, and region 28 stands out as the most connected, having links with 9 other regions.

The primary difference between the Queen's and Rook's contiguities lies in their definitions. In a matrix, the Queen's connection is that a region can be connected to another through edge and vertex contiguities, much like the movement of a queen in chess. In contrast, the Rook's connection are the connections to edge contiguities alone, more like a rook's movement in chess. Due to this slightly different connectivity distributions, as we can see in the provided data, the Queen's connection tends to have a slightly higher number of links than the Rook's Case. (The number of links screenshots are listed in the reference)

### 3. Weight matrix evaluation

The first step of the part of the project is to form three matrices chosen from these styles. Three styles, W, B and C were selected.

For the W style, each region, on average, forms approximately 5.086957 links with its counterparts. In terms of distinct connectivity, Region 90 has the fewest connections, with 2 links, while Region 28 stands as the most connected, featuring 9 links. Similarly, the B-style weight matrix average connectivity is consistent at 5.086957 links per region. Regions 90 and 28 continue to represent the least and most connected entities, respectively. The C-style weight matrix has an average of 5.086957 links and the same regions, 90 and 28, as the least and most connected. All three weight matrix styles, W, B, and C, present an identical neighborhood structure across the 92 regions. The key distinctions lie in their weight constants. While  $n$  and  $nn$  remain consistent across the styles, there's variation in  $S_0$  and  $S_2$ . The W-style and C-style share the same  $S_0$  value, but their  $S_2$  values differ. B-style's  $S_0$  and  $S_2$  are distinct from both W and C (screenshot in reference).

### 4. Moran's I evaluation

After choosing W as weight style, the Moran's I statistic for POP2000 is positive, indicating some presence of positive spatial autocorrelation, where similar values tend to be located near each other. However, the p-values for both tests are above the significance of  $p = 0.05$ , suggesting that the observed spatial pattern could arise from random spatial processes. For the HISPANIC attribute, the positive Moran's I statistic implies a mild positive spatial autocorrelation. Similar to the POP2000 attribute, the p-values are above 0.05 for both tests. This suggests that the spatial distribution of the HISPANIC attribute might also be due to random spatial processes. The MED-AGE has a positive Moran's I statistic, indicating potential positive spatial autocorrelation. However, the p-values are

again above 0.05, meaning the spatial pattern observed might not be statistically significant and could be random. All three attributes - POP2000, HISPANIC, and MED-AGE - exhibit some degree of positive spatial autocorrelation as evidenced by their positive Moran's I statistics. However, based on the provided p-values, none of the attributes show statistically significant spatial autocorrelation patterns. This implies that the spatial distribution of these attributes might be largely influenced by random processes. The Table 01 shows the example of Moran test. The rest of the screenshot shows in the reference.

Table 01

*Moran's I Test Three Attributes*

Attributes	Moran's I	P-Value (Normalization)	P-Value (Randomization)
<b>POP2000</b>	0.0640016	0.1265	0.07834
<b>HISPANIC</b>	0.0403126	0.2171	0.1441
<b>MED_AGE</b>	0.0611555	0.1357	0.1267

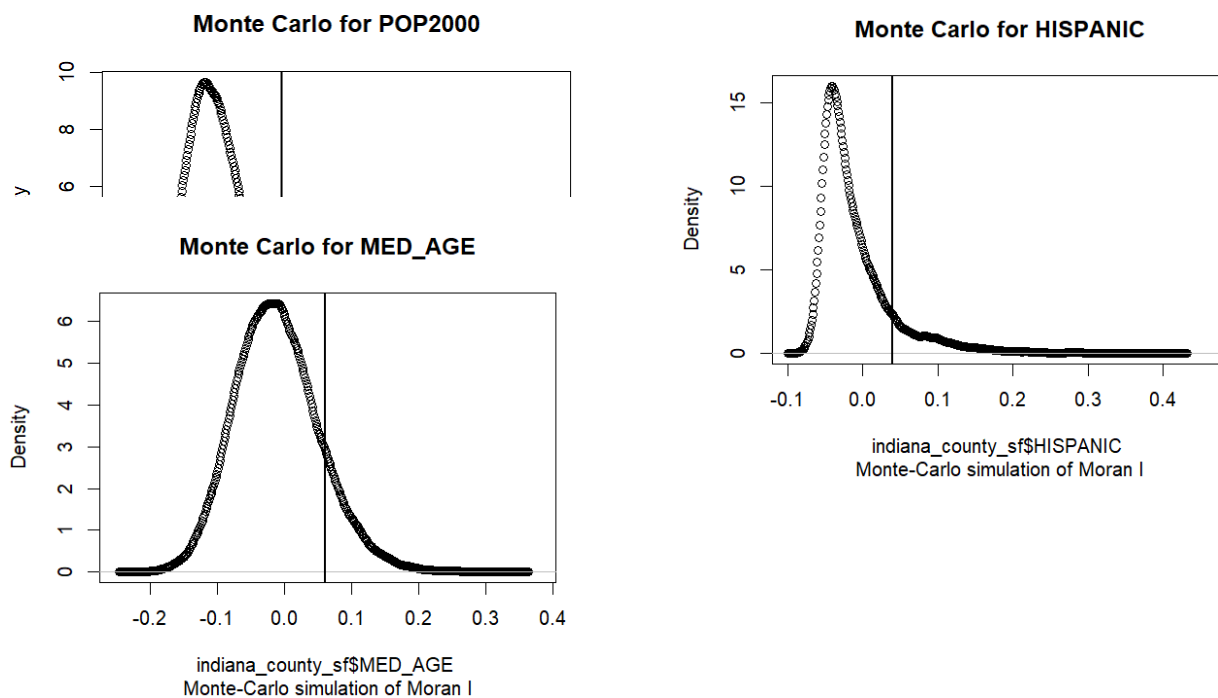
After the Monte Carlo test, The Moran I statistic value is slightly positive, which indicates a weak positive spatial autocorrelation. However, the p-value is slightly above 0.05, indicating that this observation is not statistically significant at the 95% confidence level. Hence, there isn't enough evidence to conclude a strong spatial autocorrelation in the POP2000 data. The Moran I statistic value for the HISPANIC attribute is even closer to 0 compared to POP2000, suggesting a very weak spatial autocorrelation. The p-value again is above 0.05, indicating that this result is not statistically significant at the 95% confidence level. The Moran I statistic for MED-AGE is slightly positive, suggesting a weak positive spatial autocorrelation. But, as with the previous two attributes, the p-value is above 0.05. Therefore, there isn't enough statistical evidence to conclude significant spatial autocorrelation for MED-AGE. The table below shows the result of all three attributes (Table 02).

Table 02

*Monte Carlo Test Moran's I*

Attribute	MORAN's I	Observed Rank	P-Value
<b>POP2000</b>	0.064002	45475	0.09052
<b>HISPANIC</b>	0.040313	44098	0.1181
<b>MED_AGE</b>	0.061156	43673	0.1266

In addition, the density plot was generated for the Monte Carlo test. The observed Moran's I (dotted line) is very close to the peak of the distribution. This suggests that the spatial pattern observed in the POP2000 data closely matches what might be expected from random spatial patterns. The p-value for this attribute is above the significance level of 0.05. For Hispanic, the observed Moran's I is far to the left of the peak of the distribution. This suggests that the spatial pattern in the Hispanic is more dispersed than would be expected from random patterns. For the MED\_AGE, the observed Moran's

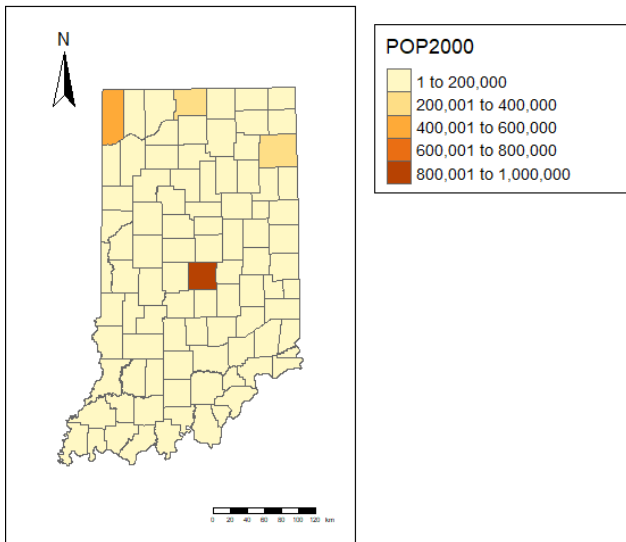


It is again close to the peak of the distribution, similar to POP2000. This suggests that the spatial pattern in the MED\_AGE data is close to what might be expected under spatial randomness.

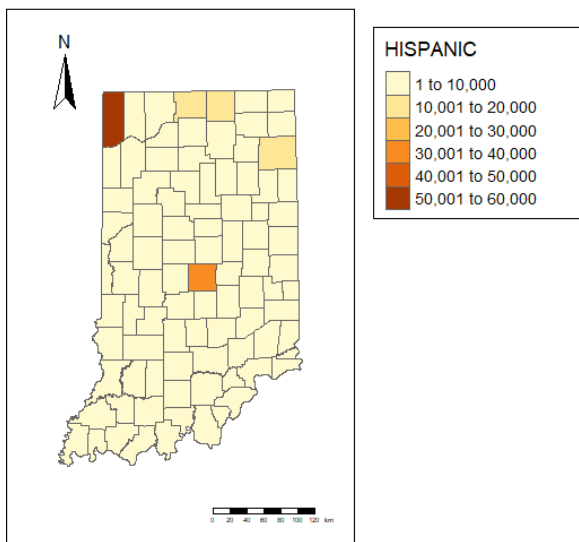
## 5. Local Moran's I evaluation

For the Local Moran's I evaluation section, the first step is to conduct thematic map analysis. The first map below shows the population distribution within Indiana's counties for the year 2000. The

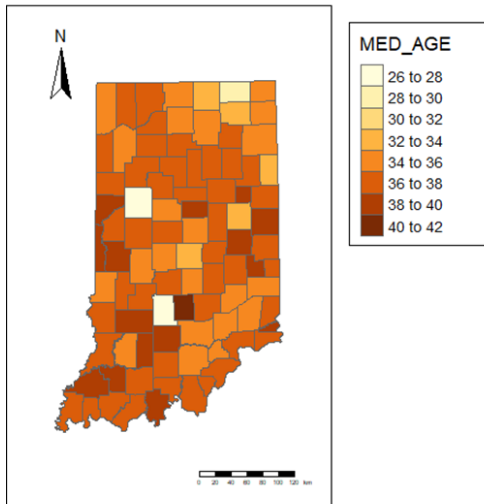
majority of the counties are shaded pale, indicating populations ranging from 1 to 200,000. However, there are two areas that stand out: one county in the northwestern part of Indiana is colored light orange, signifying a population between 200,001 to 400,000. Similarly, a county in the central region (Indianapolis Area) is depicted in medium orange, indicating its population falls within the 400,001 to 600,000 range.



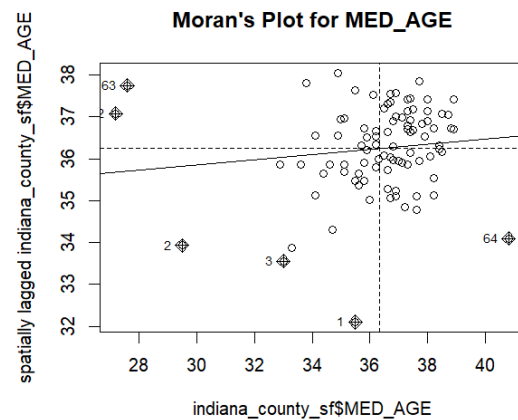
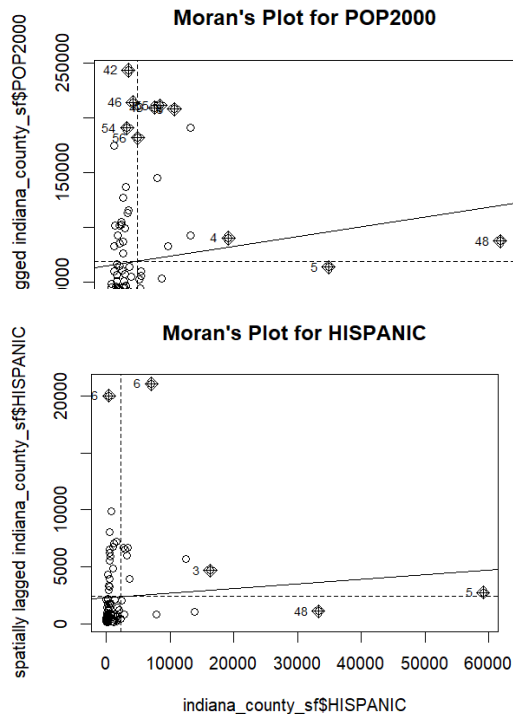
For the Hispanic distribution, most of the counties are colored pale, suggesting a Hispanic population between 1 to 10,000. There's a noticeable county in the northern region shaded in light orange, which indicates a Hispanic population ranging from 10,001 to 20,000. Similarly, in the Marion County (Indianapolis) is highlighted in medium orange, another thing to notice is that the Lake county (northwestern) has Hispanics ranging from 50,000 to 60,000, which is more than Indianapolis area.



The third map illustrates the median age distribution across Indiana's counties. The majority of counties are shaded in various shades of orange, indicating a generally older median age. Another iconic thing to notice is that there are two counties that the med age ranges from 26 to 28. Where these two counties are where Purdue University and Indiana University is located at (Tippecanoe and Monroe). A significant amount of college students might be the reason of cause med age younger than other counties.



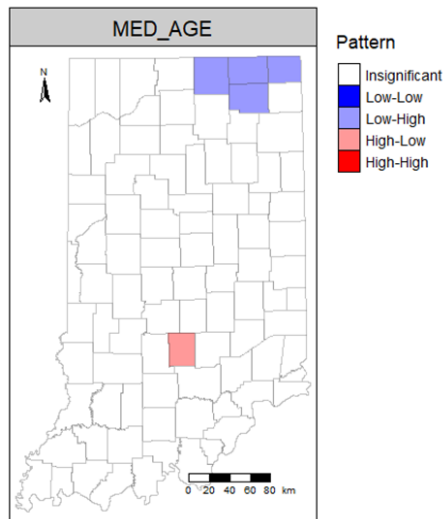
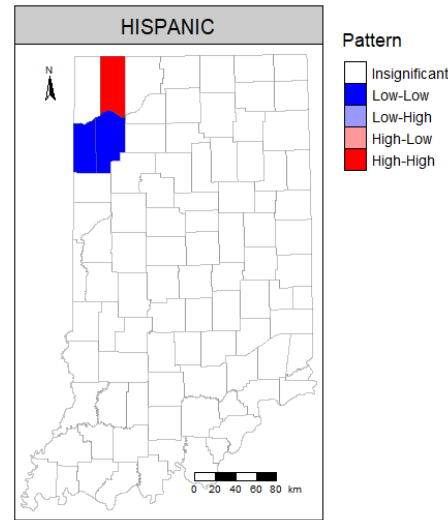
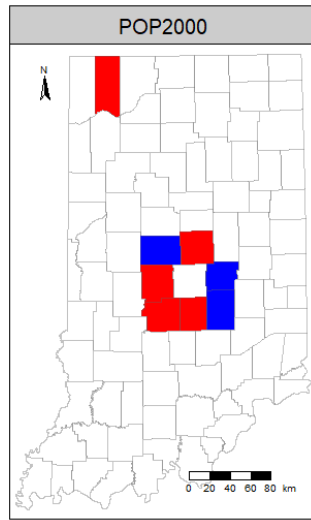
The moran.plot function was used to find the distribution of lagged averages with respect to the total average. Three plots below show the distribution of three attributes.



The plot for POP2000 suggesting that counties with higher populations tend to be surrounded by counties with higher populations as well. This implies positive spatial autocorrelation. For the MED\_AGE plot, the distribution of points around the horizontal line (the overall median age) seems random, which may suggest that the median

age of a county does not necessarily correlate with the median age of its neighboring counties. For the Hispanic plot, similar to the POP2000 plot, the points tend to form a positive slope. Counties with higher Hispanic populations tend to be near counties with higher Hispanic populations as well. Again, this indicates positive spatial autocorrelation.

The following map shows the Local Moran's I's clusters of all three attributes based on High-High, Low-Low High-Low and Low-High.



For POP2000, the counties around the Indianapolis area are a unique mix of both High-High (HH) and Low-Low (LL) clusters, forming a ring. This suggests that while there are pockets of high population density adjacent to other high-density areas, there are also counties of low population density neighboring other low-density areas. This is common to see at the urban-rural area's divide, where the city's core and certain suburbs exhibit high population densities, and other areas or the more rural areas show a lower

density. In addition, another clear High-High pattern can be seen at the northwestern corner. This suggests another significant concentration of population area besides the central Indianapolis (Lake County as mentioned before).

The spatial distribution for HISPANIC reveals multiple significant patterns. In the northwest, there's a distinct Low-Low (LL) cluster, suggesting that areas with low Hispanic populations are neighboring

other areas with similarly low Hispanic concentrations. Just beside this LL cluster, a High-High (HH) pattern emerges, indicating regions with a high Hispanic presence are adjacent to similar areas. This could be indicative of areas where there is a strong Hispanic community.

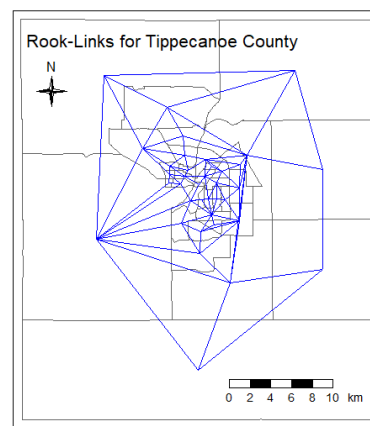
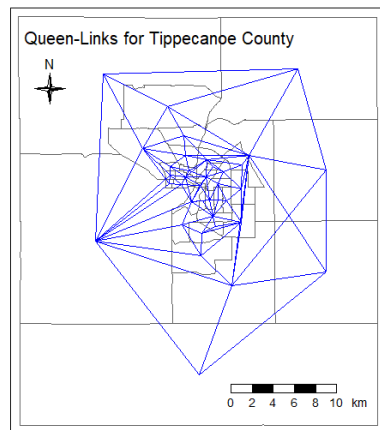
For MED\_AGE, the clustering appears to be more dispersed compared to the other two attributes. A distinct High-Low (HL) cluster is present in the middle, suggesting areas with a higher median age are adjacent to areas with a younger median age. To the northeast, a Low-High (LH) pattern is visible, which indicates areas with a lower median age being adjacent to areas with an older median age.

## 6. County level study

For the county level analysis, I have selected the county where Purdue University is located at. After sorting through the attributes, the Tippecanoe County shows the county ID of 157.

994959658	128724.61	JAY	Jay	38	075	18075	2
428882449	83330.95	BLACKFORD	Blackford	5	009	18009	1
760587830	128462.38	HOWARD	Howard	34	067	18067	8
1302322797	145012.51	TIPPECANOE	Tippecanoe	79	157	18157	14
948193947	144068.62	WARREN	Warren	86	171	18171	
1048797387	132964.60	CLINTON	Clinton	12	023	18023	3

The following map shows the connection of tracts in Tippecanoe County.



For the Moran's I statistics, the result of three attributes (same as the county level) in Tippecanoe County is shown in Table 03. Based on these results, the spatial distribution of Tippecanoe MED\_AGE shows significant clustering, while the spatial distribution of Tippecanoe POP2000 and Tippecanoe Hispanic seems to have a weaker, yet still evident, positive spatial correlation.

Table 03

*Moran's I Statistics Tippecanoe County*

Attributes	Moran's I	P-value (randomization)	P-value (normality)
<b>POP2000</b>	0.06939	0.1574	0.1573
<b>HISPANIC</b>	0.16311	0.0166	0.0241
<b>MED_AGE</b>	0.4253	1.438e-06	1.374e-06



Table 04 shows the results of the Monte Carlo simulation results for Moran's I statistics.

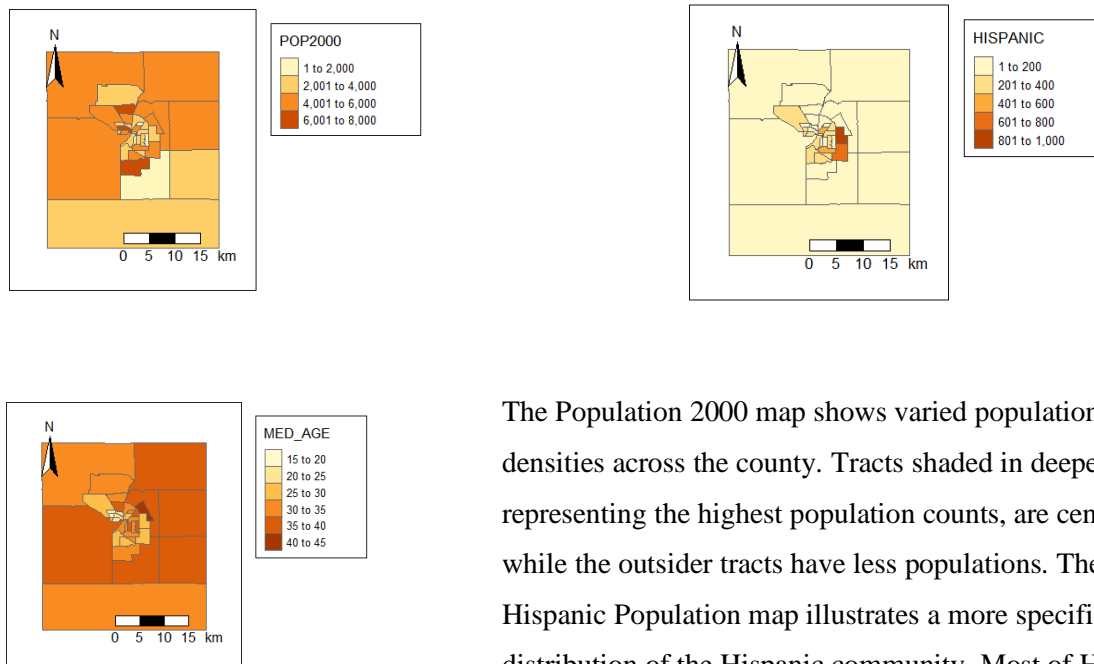
Table 04

*Monte Carlo Moran's I Tippecanoe County*

Attributes	Moran's I	Observed Rank	P-Value
<b>POP2000</b>	0.069393	42060	0.1588
<b>HISPANIC</b>	0.16311	48547	0.02908
<b>MED_AGE</b>	0.4253	49999	4e-05

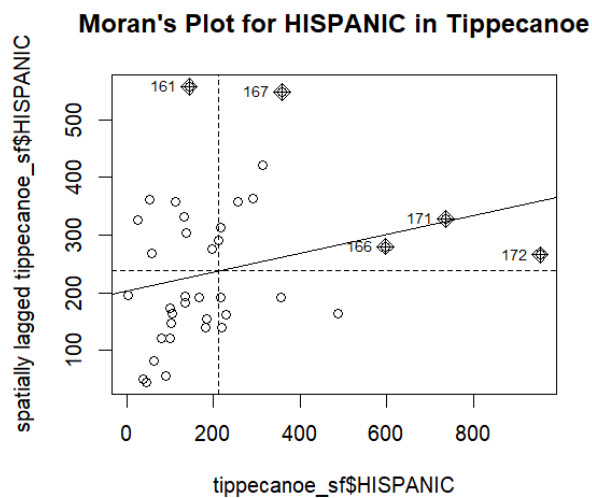
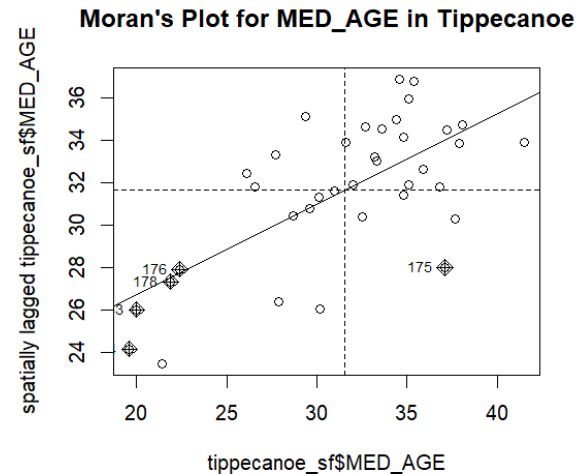
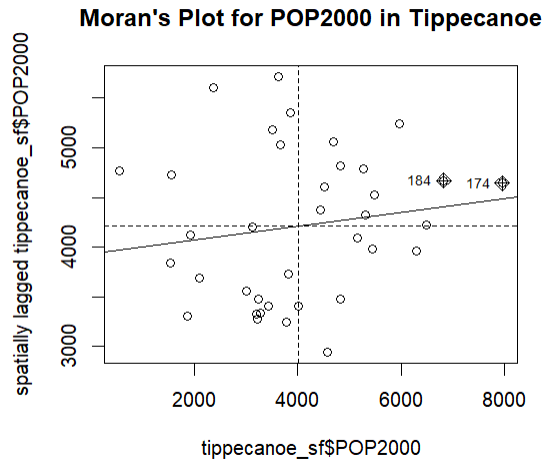
The spatial distribution of the population in 2000 in Tippecanoe County is not significantly different from a random distribution. In other words, there's no strong evidence of spatial autocorrelation for the population data of that year. With a p-value below 0.05, there's evidence to suggest that the spatial distribution of the Hispanic population in Tippecanoe County is not random. The positive Moran's I value indicates a positive spatial autocorrelation, meaning similar values (either high or low) of Hispanic population counts are more likely to be found near each other. The p-value for the median age is less than 0.05, indicating a strong spatial autocorrelation in the data. The positive Moran's I value suggests that areas with similar median ages are clustered together. Given the extremely low p-value, this pattern is very unlikely to be due to random chance. The density plot is shown in reference.

The thematic map of all three attributes for Tippecanoe County is shown below.



The Population 2000 map shows varied population densities across the county. Tracts shaded in deeper colors, representing the highest population counts, are centralized, while the outsider tracts have less populations. The Hispanic Population map illustrates a more specific distribution of the Hispanic community. Most of Hispanic communities are located in central areas, while outer tracts

have mostly less than 200 Hispanics. For the median age map, the central region is characterized by younger age groups, that is where Purdue University is located at. Probably the younger median age is caused by the college students.

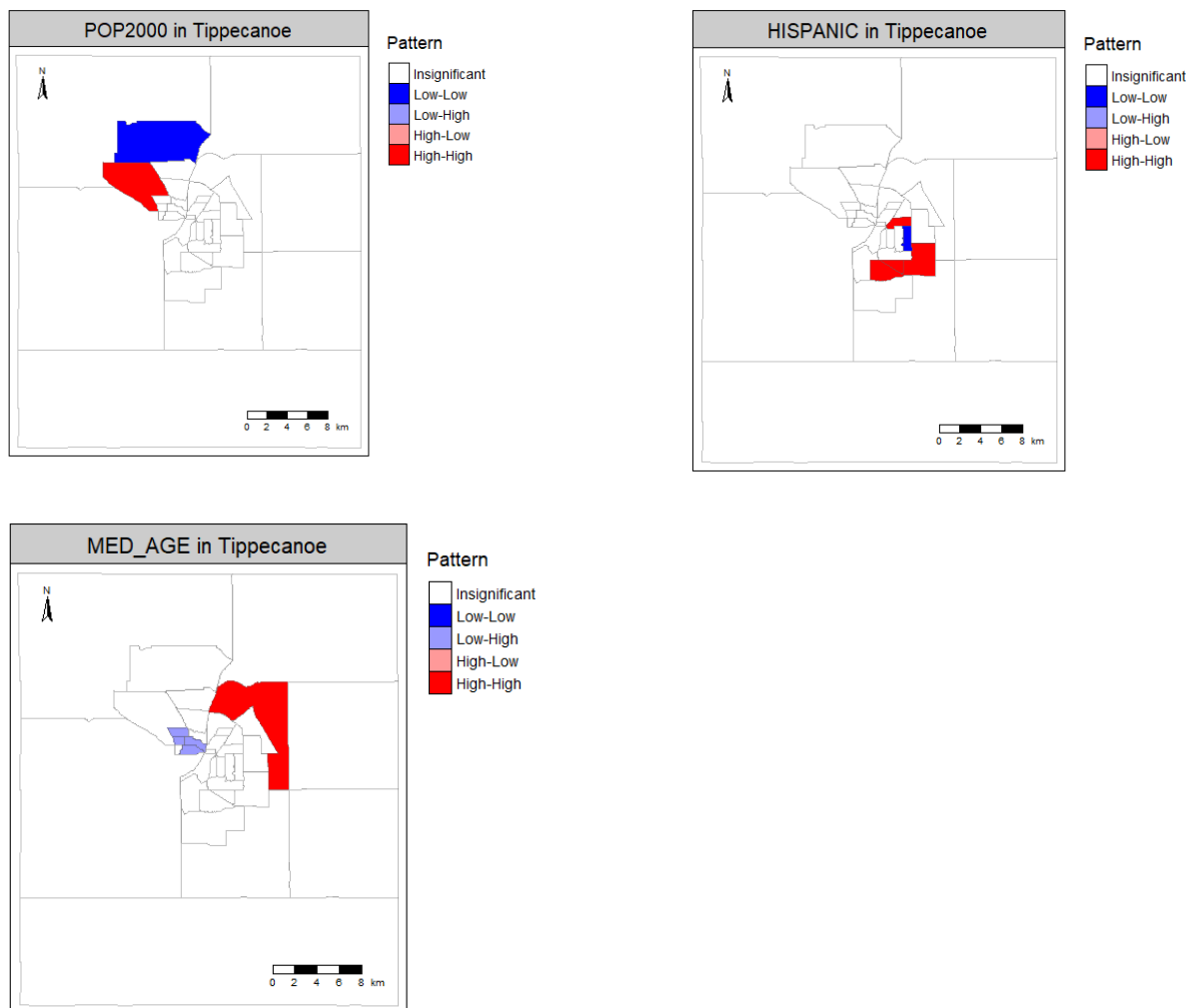


For the Moran' I' lagged average plot, the population 2000 scatterplot appears to be more random in its distribution, with the line of best fit relatively flat. This suggests a weaker or no significant spatial autocorrelation for population values in 2000. Data points are spread across the plot without a distinct pattern around the line. For the med age plot, the upward trend of the scatterplot suggests a positive spatial autocorrelation. Tracts with higher median ages tend to be surrounded by regions with similar higher median ages, and vice versa for lower values. For

the Hispanic, although not as upward as the median age plot, indicating a slight positive spatial autocorrelation.

The following map shows the Tippecanoe County's Local Moran's I's clusters of all three attributes based on High-High, Low-Low High-Low and Low-High. The POP2000 analysis reveals that the northwestern-central region, marked in red, showcases a High-High pattern, indicating areas with high populations surrounded by similarly populous areas. In contrast, the northern tract right next to the red tract, colored in blue, depicts a Low-Low pattern, signifying areas with lower populations neighbored by regions with similar population counts. In the analysis of the Hispanic population in

Tippecanoe, the central southern region, represented in red, displays a High-High pattern, highlighting zones with a dense Hispanic population surrounded by areas with similarly high Hispanic populations. There is a blue tract in the middle of two High-High regions, shows it is surrounded by areas with low Hispanic populations. For the median age distribution in Tippecanoe, the eastern region, colored in red, represents a High-High clustering, indicating areas with older median ages in proximity to others with similarly higher median ages. Notably, a smaller section in lighter blue, located centrally, signifies a Low-High trend, indicating an area with a younger median age surrounded by regions that possess older median ages.



For the MAUP effect, the state-level map simplifies the median age data into broader categories. When we zoom into Tippecanoe County, we see a much more detailed breakdown of median age, with tracts. For someone only looking at the Indiana state-level map, they might assume Tippecanoe

County has a relatively homogenous median age falling within the 26 to 28 range. However, the county-level map reveals internal variations ranging from 15 to 45, which is a significant difference. And once when we look at the thematic map of population in Tippecanoe County, we can see that most of the tracts with higher population are located in the central region, where most of them are at lower med age.

## **7. Summary/Conclusion/Concluding Remarks**

The project primarily focused on spatial analysis of Indiana's counties using Census data, three attributes were selected for the analysis. The spatial connectivity highlighted differences between Queen's and Rook's connections, revealing variances in regional connectivity. Moran's I evaluation identified weak positive spatial autocorrelation in all three attributes, but these were not statistically significant based on p-values. The Monte Carlo test mirrored these findings. Local Moran's I evaluation presented spatial patterns using thematic maps. POP2000 and HISPANIC exhibited spatial clusters in regions, such as Indianapolis for high population density and northwestern areas for Hispanic communities. MED-AGE patterns were more dispersed, with notable younger median ages in counties hosting major universities. The Tippecanoe County were also selected to conduct the analysis, despite some of the results reveal the same as compared to State level, It is clearly to see that the majority of the population at Purdue University (college students with lower med age) affect the distribution of population and age at Tippecanoe County.

## **Acknowledgement**

## References

### Column Names

```
> colnames(indiana_county_sf)
[1] "AREA" "PERIMETER" "NAME_U" "NAME_L" "NCAPC" "CNTY_FIPS" "STFID" "POP2000"
[9] "WHITE" "BLACK" "AMERI_ES" "ASIAN" "HAWN_PI" "OTHER" "MULT_RACE" "HISPANIC"
[17] "MALES" "FEMALES" "AGE_UNDE5" "AGE_5_17" "AGE_18_21" "AGE_22_29" "AGE_30_39" "AGE_40_49"
[25] "AGE_50_64" "AGE_65_UP" "MED_AGE" "MED_AGE_M" "MED_AGE_F" "HOUSEHOLDS" "AVE_HH_SZ" "HSEHLD_1_M"
[33] "HSEHLD_1_F" "MARHH_CHD" "MARHH_NO_C" "MHH_CHILD" "FHH_CHILD" "FAMILIES" "AVE_FAM_SZ" "HSE_UNITS"
[41] "VACANT" "OWNER_OCC" "RENTER_OCC" "geometry"

> colnames(indiana_tract_sf)
[1] "ID" "FIPSSTCO" "TRT2000" "STFID" "TRACTID" "COUNTY" "POP2000" "WHITE"
[9] "BLACK" "AMERI_ES" "ASIAN" "HAWN_PI" "OTHER" "MULT_RACE" "HISPANIC" "MALES"
[17] "FEMALES" "AGE_UNDE5" "AGE_5_17" "AGE_18_21" "AGE_22_29" "AGE_30_39" "AGE_40_49" "AGE_50_64"
[25] "AGE_65_UP" "MED_AGE" "MED_AGE_M" "MED_AGE_F" "HOUSEHOLDS" "AVE_HH_SZ" "HSEHLD_1_M" "HSEHLD_1_F"
[33] "MARHH_CHD" "MARHH_NO_C" "MHH_CHILD" "FHH_CHILD" "FAMILIES" "AVE_FAM_SZ" "HSE_UNITS" "VACANT"
[41] "OWNER_OCC" "RENTER_OCC" "geometry"
```

### Spatial Connectivity Links

```
> summary(queen_nb)
Neighbour list object:
Number of regions: 92
Number of nonzero links: 480
Percentage nonzero weights: 5.671078
Average number of links: 5.217391
Link number distribution:

  2  3  4  5  6  7  8  9
1 13 13 25 24 11  4  1
1 least connected region:
90 with 2 links
1 most connected region:
28 with 9 links

> summary(rook_nb)
Neighbour list object:
Number of regions: 92
Number of nonzero links: 468
Percentage nonzero weights: 5.529301
Average number of links: 5.086957
Link number distribution:

  2  3  4  5  6  7  8  9
2 16  9 28 24  9  3  1
2 least connected regions:
1 90 with 2 links
1 most connected region:
28 with 9 links
```

### Weight Matrix

```
> summary(weight_W_county)
Characteristics of weights list object:
Neighbour list object:
Number of regions: 92
Number of nonzero links: 468
Percentage nonzero weights: 5.529301
Average number of links: 5.086957
Link number distribution:

  2  3  4  5  6  7  8  9
2 16  9 28 24  9  3  1
2 least connected regions:
1 90 with 2 links
1 most connected region:
28 with 9 links

Weights style: W
Weights constants summary:
  n  nn S0  S1  S2
W 92 8464 92 38.51639 375.3014

> summary(weight_B_county)
Characteristics of weights list object:
Neighbour list object:
Number of regions: 92
Number of nonzero links: 468
Percentage nonzero weights: 5.529301
Average number of links: 5.086957
Link number distribution:

  2  3  4  5  6  7  8  9
2 16  9 28 24  9  3  1
2 least connected regions:
1 90 with 2 links
1 most connected region:
28 with 9 links

Weights style: B
Weights constants summary:
  n  nn S0  S1  S2
B 92 8464 468 936 10296

> summary(weight_C_county)
Characteristics of weights list object:
Neighbour list object:
Number of regions: 92
Number of nonzero links: 468
Percentage nonzero weights: 5.529301
Average number of links: 5.086957
Link number distribution:

  2  3  4  5  6  7  8  9
2 16  9 28 24  9  3  1
2 least connected regions:
1 90 with 2 links
1 most connected region:
28 with 9 links

Weights style: C
Weights constants summary:
  n  nn S0  S1  S2
C 92 8464 92 36.17094 397.8803
```

## Moran Test

```
> moran.test(x = indiana_county_sf$HISPANIC, listw = weight_W_county, randomisation = TRUE)

Moran I test under randomisation

data: indiana_county_sf$HISPANIC
weights: weight_W_county

Moran I statistic standard deviate = 1.062, p-value = 0.1441
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.04031256      -0.01098901      0.00233342

> moran.test(x = indiana_county_sf$HISPANIC, listw = weight_W_county, randomisation = FALSE)

Moran I test under normality

data: indiana_county_sf$HISPANIC
weights: weight_W_county

Moran I statistic standard deviate = 0.78208, p-value = 0.2171
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.04031256      -0.01098901      0.004302853

> moran.test(x = indiana_county_sf$POP2000, listw = weight_W_county, randomisation = TRUE)

Moran I test under randomisation

data: indiana_county_sf$POP2000
weights: weight_W_county

Moran I statistic standard deviate = 1.4163, p-value = 0.07834
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.064001565      -0.010989011      0.002803318

> moran.test(x = indiana_county_sf$POP2000, listw = weight_W_county, randomisation = FALSE)

Moran I test under normality

data: indiana_county_sf$POP2000
weights: weight_W_county

Moran I statistic standard deviate = 1.1432, p-value = 0.1265
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.064001565      -0.010989011      0.004302853

> moran.test(x = indiana_county_sf$MED_AGE, listw = weight_W_county, randomisation = TRUE)

Moran I test under randomisation

data: indiana_county_sf$MED_AGE
weights: weight_W_county

Moran I statistic standard deviate = 1.1421, p-value = 0.1267
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.061155598      -0.010989011      0.003990295

> moran.test(x = indiana_county_sf$MED_AGE, listw = weight_W_county, randomisation = FALSE)

Moran I test under normality

data: indiana_county_sf$MED_AGE
weights: weight_W_county

Moran I statistic standard deviate = 1.0998, p-value = 0.1357
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.061155598      -0.010989011      0.004302853
```

## Tippecanoe County

## Weight Matrix

```
> summary(weight_W_tippecanoe)
Characteristics of weights list object:
Neighbour list object:
Number of regions: 37
Number of nonzero links: 192
Percentage nonzero weights: 14.02484
Average number of links: 5.189189
Link number distribution:

3 4 5 6 7 8 9
8 6 9 6 4 1 3
8 least connected regions:
156 159 163 167 185 189 190 191 with 3 links
3 most connected regions:
166 186 188 with 9 links

Weights style: W
Weights constants summary:
      n  nn S0      S1      S2
W 37 1369 37 15.00203 154.3961
> summary(weight_B_tippecanoe)
Characteristics of weights list object:
Neighbour list object:
Number of regions: 37
Number of nonzero links: 192
Percentage nonzero weights: 14.02484
Average number of links: 5.189189
Link number distribution:

3 4 5 6 7 8 9
8 6 9 6 4 1 3
8 least connected regions:
156 159 163 167 185 189 190 191 with 3 links
3 most connected regions:
166 186 188 with 9 links

Weights style: B
Weights constants summary:
      n  nn S0 S1      S2
B 37 1369 192 384 4448

> summary(weight_C_tippecanoe)
Characteristics of weights list object:
Neighbour list object:
Number of regions: 37
Number of nonzero links: 192
Percentage nonzero weights: 14.02484
Average number of links: 5.189189
Link number distribution:

3 4 5 6 7 8 9
8 6 9 6 4 1 3
8 least connected regions:
156 159 163 167 185 189 190 191 with 3 links
3 most connected regions:
166 186 188 with 9 links

Weights style: C
Weights constants summary:
      n  nn S0      S1      S2
C 37 1369 37 14.26042 165.1832
```

## Moran I

```
> moran.test(x = tippecanoe_sf$POP2000, listw = weight_W_tippecanoe, randomisation = TRUE)

Moran I test under randomisation

data: tippecanoe_sf$POP2000
weights: weight_W_tippecanoe

Moran I statistic standard deviate = 1.0052, p-value = 0.1574
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.069392602          -0.027777778          0.009344821

> moran.test(x = tippecanoe_sf$POP2000, listw = weight_W_tippecanoe, randomisation = FALSE)

Moran I test under normality

data: tippecanoe_sf$POP2000
weights: weight_W_tippecanoe

Moran I statistic standard deviate = 1.0056, p-value = 0.1573
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.069392602          -0.027777778          0.009337434

> moran.test(x = tippecanoe_sf$HISPANIC, listw = weight_W_tippecanoe, randomisation = TRUE)

Moran I test under randomisation

data: tippecanoe_sf$HISPANIC
weights: weight_W_tippecanoe

Moran I statistic standard deviate = 2.1297, p-value = 0.0166
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.163111205          -0.027777778          0.008033735

> moran.test(x = tippecanoe_sf$HISPANIC, listw = weight_W_tippecanoe, randomisation = FALSE)

Moran I test under normality

data: tippecanoe_sf$HISPANIC
weights: weight_W_tippecanoe

Moran I statistic standard deviate = 1.9755, p-value = 0.02411
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.163111205          -0.027777778          0.009337434

> moran.test(x = tippecanoe_sf$MED_AGE, listw = weight_W_tippecanoe, randomisation = TRUE)

Moran I test under randomisation

data: tippecanoe_sf$MED_AGE
weights: weight_W_tippecanoe

Moran I statistic standard deviate = 4.6795, p-value = 1.438e-06
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.425302947          -0.027777778          0.009374394

> moran.test(x = tippecanoe_sf$MED_AGE, listw = weight_W_tippecanoe, randomisation = FALSE)

Moran I test under normality

data: tippecanoe_sf$MED_AGE
weights: weight_W_tippecanoe

Moran I statistic standard deviate = 4.6888, p-value = 1.374e-06
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.425302947          -0.027777778          0.009337434
```



## Monte Carlo Tippecanoe

```
> print(moran_mc_pop2000_tippecanoe)
```

Monte-Carlo simulation of Moran I

```
data: tippecanoe_sf$POP2000
weights: weight_W_tippecanoe
number of simulations + 1: 50001
```

```
statistic = 0.069393, observed rank = 42060, p-value = 0.1588
alternative hypothesis: greater
```

```
> print(moran_mc_hispanic_tippecanoe)
```

Monte-Carlo simulation of Moran I

```
data: tippecanoe_sf$HISPANIC
weights: weight_W_tippecanoe
number of simulations + 1: 50001
```

```
statistic = 0.16311, observed rank = 48547, p-value = 0.02908
alternative hypothesis: greater
```

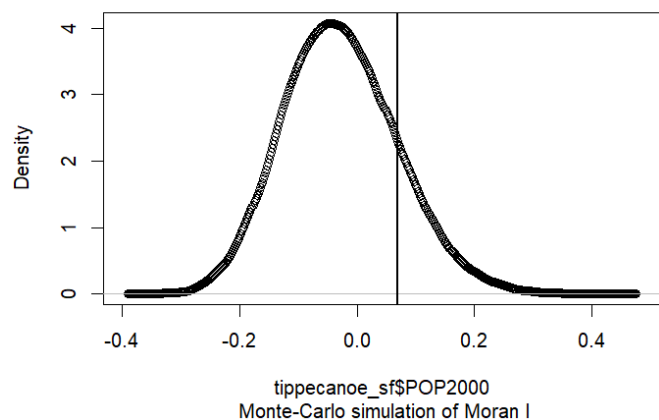
```
> print(moran_mc_med_age_tippecanoe)
```

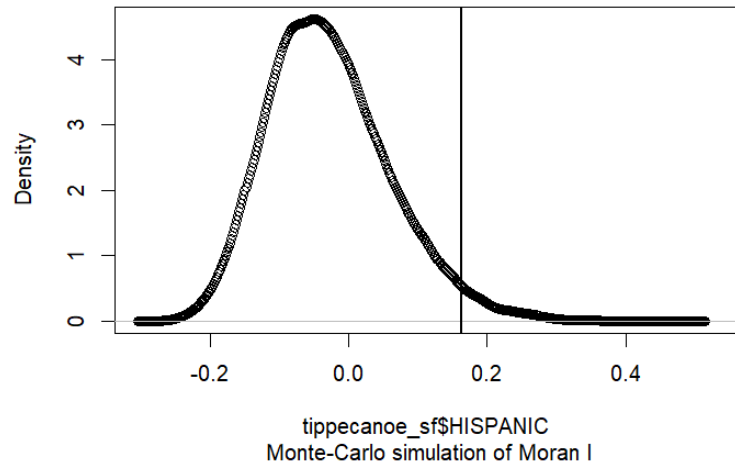
Monte-Carlo simulation of Moran I

```
data: tippecanoe_sf$MED_AGE
weights: weight_W_tippecanoe
number of simulations + 1: 50001
```

```
statistic = 0.4253, observed rank = 49999, p-value = 4e-05
alternative hypothesis: greater
```

## Monte Carlo for POP2000 in Tippecanoe County Tracts



**Monte Carlo for HISPANIC in Tippecanoe County Tracts****Monte Carlo for MED\_AGE in Tippecanoe County Tracts**