

## Project 6. Moran's I: Spatial Autocorrelation

### Abstract

This project will detail the use of the Moran's I test statistic on county census data for Indiana. Moran's I details the overall spatial autocorrelation of a dataset and is available in the 'spdep' package in R.

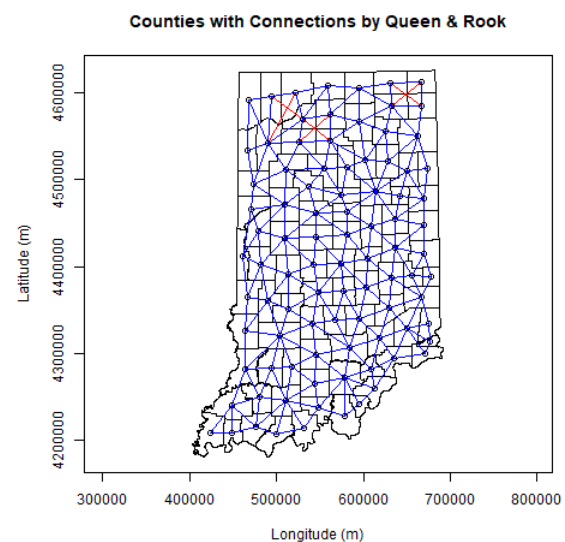
### 1. Spatial Connectivity Evaluation

To begin this analysis of spatial autocorrelation, it was necessary to generate a set of links for contiguous counties. Fortunately, this is quite easy with the aid of the `poly2nb` function. This function establishes neighbors from polygons under two different boundary conditions, queen and rook. The queen neighborhood is defined as at least one shared boundary while the rook neighborhood is defined as having more than one shared boundary. This indicates that the queen boundaries will always have more connections. These neighborhoods can be seen in Figure 1 with queen connections in red and rook in blue.

### 2. Weight Matrix Evaluation

Forming weight matrices in R is trivial with the `nb2listw` function. This function takes the results from the spatial connectivity analysis and a style argument. From this point forward, the rook connectivity definition will be used. Three styles were chosen for evaluation of the counties, these styles are W, B, and C. W refers to a row standardized weighting method that sums over all the links to the county in question. W assigns its weights such that they

sum to one. Online reading points out the usefulness of row standardization when the boundaries are arbitrarily defined. B refers to binary, that is all connections to the county in question are given a weight of 1. C refers to a global standardization, assigning the same weights to all connections in the entire dataset. After establishing the weights, `moran.test` was run with the assumption of normality. The results have been summarized in Table 1. As can be seen by the P-Values, the null hypothesis under each weighting scheme fails to be



**Figure 1:** Neighborhood connections for queen (red) and rook (blue) for Indiana counties.

rejected, that is to say, we fail to reject the notion that the data is randomly distributed. Interestingly the B and C methods are identical. This is due to their weighting being simple scaled versions of one another.

Table 1: Summary of Moran's I test statistic information output by R.

Weighting Method	Moran's I Statistic	Expectation	Variance	P-Value
W	0.064	-0.011	0.004	0.127
B	0.066	-0.011	0.004	0.111
C	0.066	-0.011	0.004	0.111

A quick look at the population in Figure 2 does seem to support the notion of random distribution but not enough to sway Moran's I to reject the null hypothesis.

Unfortunately, the Moran's I statistic wasn't particularly informative. Only so much information can be gleaned from a single value for 92 counties. Fortunately though, the moran.test function provides finer details by way of plotting the data. These plots can be seen in Figure 3a, 3b, and 3c.

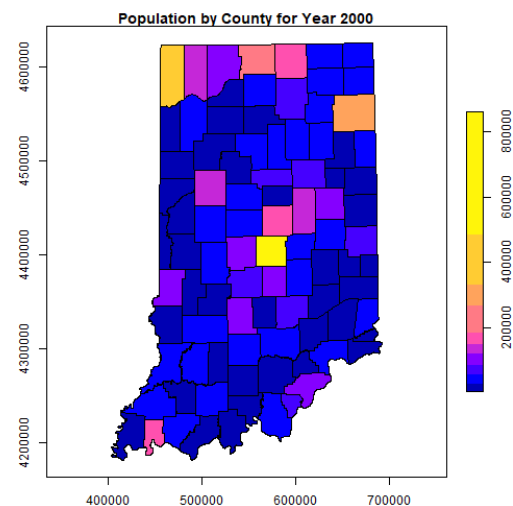


Figure 2: Population by county with Jenks breaks.

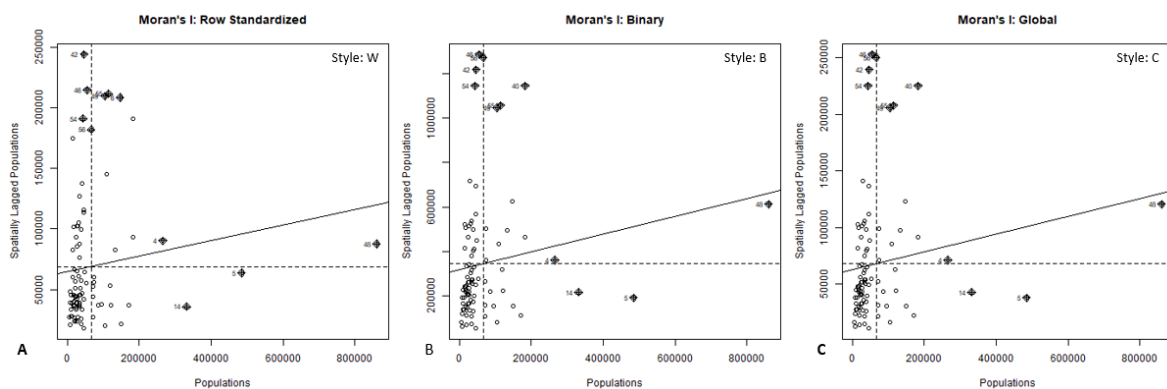


Figure 3a, 3b, 3c: Diagnostic plots for the Moran's I test for W, B, and C styles. The individual county populations are on the x-axis and the sum of the weighted neighborhood is on the y-axis for each.

Upon inspection of Figure 3, it becomes apparent that the Binary and Global weighting methods are identical aside from the spatially lagged populations. This is intuitive given their weighting methods. The row normalized (style W) plot in 3a shows the greatest difference from the other two in the distribution of its points. All three plots do show a quite a few low

population counties in the presence of low population counties. Despite this, the Moran's I still wasn't able to reject the notion that the county populations are randomly distributed.

### 3. Moran's I Evaluation

Further exploration of the counties dataset was carried out using the Moran's I statistic but with two other attributes from the county census data. I opted to choose two attributes that were closely related, population of males and females, to see if some form of relationship or unexpected distribution could be found. Intuition would suggest that these will be randomly distributed, but Moran's I will help provide this insight quantitatively. The raw population distributions for both male and female populations can be seen in Figure 4a and 4b.

Interestingly there appears to be a clustering of similar by male population counties near Indianapolis.

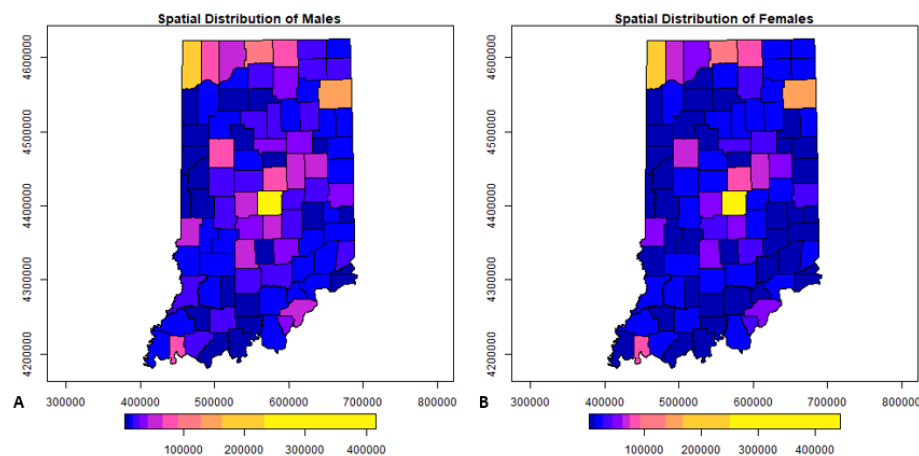


Figure 4a, 4b: Spatial distribution of male (A) and female (B) populations according to Jenks breaks.

To test the spatial auto-correlation for these attributes both a typical Moran test as well as a monte-carlo simulated Moran test were carried out and can be seen in Figure 5. A review of the corresponding P-Values of the Moran statistic for male and female populations by county were 0.119 and 0.134 respectively. Under the monte-carlo simulation framework, the P-Values dropped quite significantly for male and female populations to 0.084 and 0.104 respectively. This seems to be due to the randomization of the population distributions and shows a stronger spatial autocorrelation than the moran.test alone but still not enough to reject the null hypothesis of these populations being randomly distributed.

Moran's I, while an interesting concept, is very limiting in that it doesn't give you a sense of the Moran's I values spatially. This is where a local Moran's I comes in handy. The local Moran's I for male and female populations can be seen in Figure 6. The distributions seen in Figure 6a and b are quite similar except for in the tail. There seems to be a slightly greater male population. Spatially though, the Indianapolis male population and female population both seem to be very similar to their surroundings. One particularly surprising result was in the top left portion of Indiana (Porter County), the female population's local Moran's I is higher than that of the male population.

#### 4. Conclusions

Moran's I is an interesting and useful concept for quick analysis of spatial autocorrelation but in many ways, it feels too subjective (likely due to the choice in weighting schemes). The monte-carlo simulations do feel more rigorous and perhaps more useful in forming scientific arguments. Though it didn't seem to make much difference for the attributes reviewed here, boundary conditions are important and can greatly influence Moran's I.

#### References

IndianaMap. (n.d.). Layer Gallery. Retrieved October 22, 2019, from

<https://maps.indiana.edu/layerGallery.html?category=Census>.

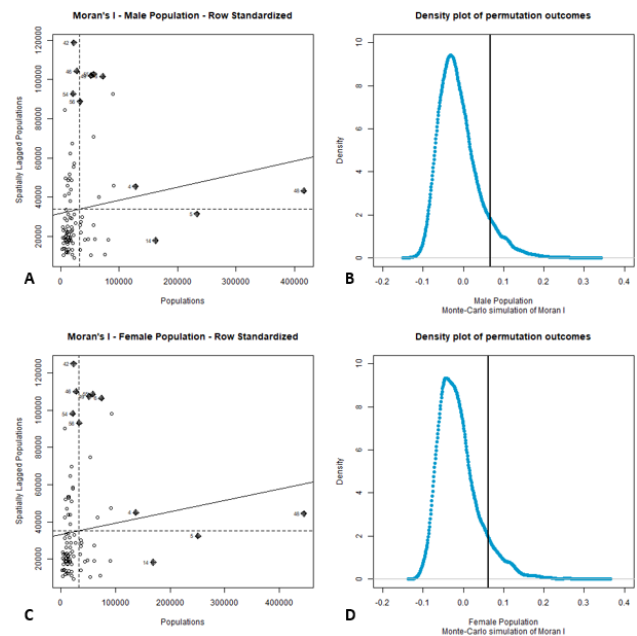


Figure 5a, 5b, 5c, 5d: Row standardized Moran tests for male (A) and female (C) populations and monte-carlo simulations for male (B) and female (D) populations.

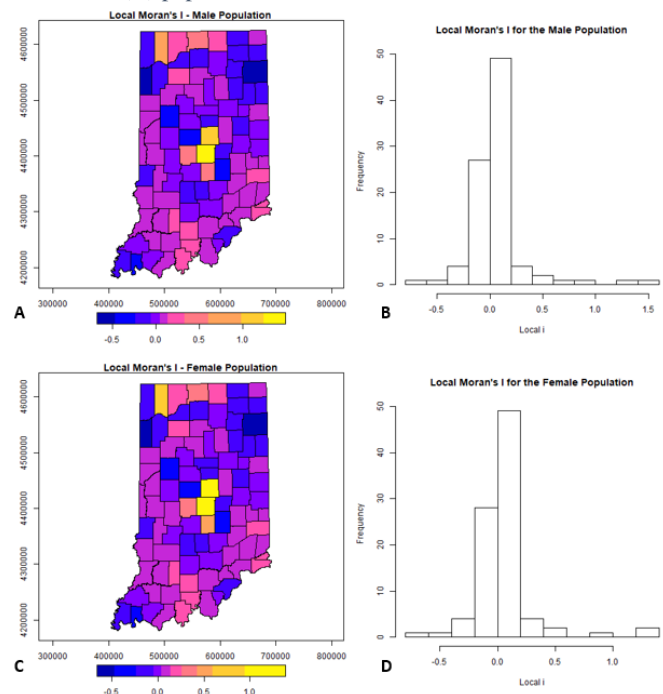


Figure 6a, 6b, 6c, 6d: Moran's local I spatial distribution and numerical distribution for males (A and B) as well as females (C and D).