

# Spatial Statistics: Spatial Autocorrelation

---

## Introduction

You will conduct tests for spatial autocorrelation in both Geoda and ArcMap. You will use median housing values for each census tract in Middlesex County, MA from the 2006-2010 American Community Survey.

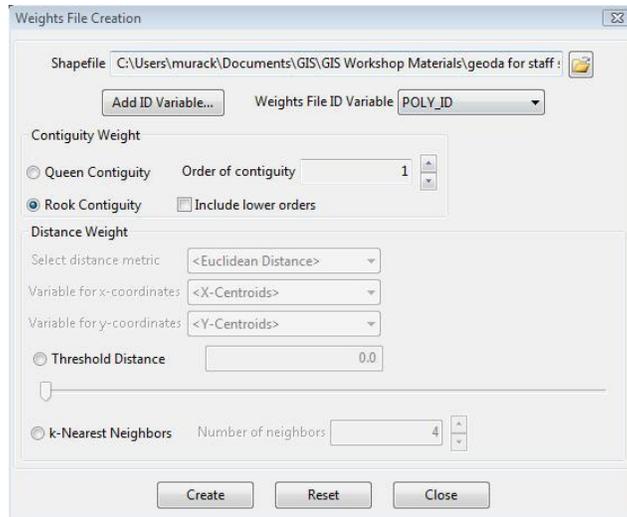
## Geoda

You will first examine the data generally to determine if median household income is clustered or dispersed in Middlesex County, then you will run local statistical tests to see where clusters exist. You will use the Moran's I tests.

## Creating a Spatial Weights Matrix

A spatial weights matrix is a way to numerically represent neighboring relationships and can be based on contiguity, distance, or the number of neighbors. You can create a spatial weights matrix while running a tool or before you begin an analysis. We'll create a weights matrix now to prepare for our spatial autocorrelation analysis.

1. Open Geoda.
2. Open the medianhousing.shp file by clicking File > New Project From > ESRI Shapefile
3. Navigate to the folder on your desktop and open medianhousing.shp.
4. Open the table by clicking on the table button .
5. The column Median\_val contains the median housing values and the column Median\_v\_1 contains the margin of error for each estimate. Close the table.
6. Click on the Weights File Creation button (.
7. Select POLY\_ID as the ID variable. If you did not have an ID variable (or if Geoda did not recognize it), you could create one using the Add ID Variable... button.
8. Select Rook Contiguity and then click Create.



9. Save it to your folder on the desktop and call the file medianhousing\_rook.
10. Close the weights creation box.

**Q:** What is rook contiguity? What neighbors are included in this matrix?

**A:** Any census tracts that share borders (but not vertices) with each other are neighbors.

To learn more about the neighbors, create a weights neighbor histogram.

1. Click on the Weights Neighbor Histogram button (  ). The weights file you just created will be opened by default. If you want to view another weights file, you can use the Open weights file button (  ).
2. Right click in this window and select Display Statistics to see the exact number of census tracts with each number of neighbors.
3. Click on a census tract in the map and the corresponding bar on the histogram will be highlighted in yellow at the bottom.
4. Similarly, select a bar from the histogram and the corresponding polygons will be highlighted on the map. This is called brushing and linking. You can use this tool to identify polygons with large or small numbers of neighbors or to compare the number of neighbors using different spatial weights matrices.

**Q:** What is the most frequently occurring number of neighbors?

**A:** 5

**Q:** What is the smallest number of neighbors? How many tracts have that number of neighbors?

**A:** 1 is the lowest number of neighbors and 1 tract has only 1 neighbor

**Q:** Where do tracts with a small number of neighbors tend to be located?

**A:** Around the edges of the map. In future analyses you may want to analyze an area slightly larger than your area of interest so that your polygons are not cut off from their neighbors.

5. Close the histogram window.

## Univariate Moran's I

Univariate Moran's I is a global statistic that tells you whether there is clustering or dispersion, but it does not inform you of the location of a cluster.

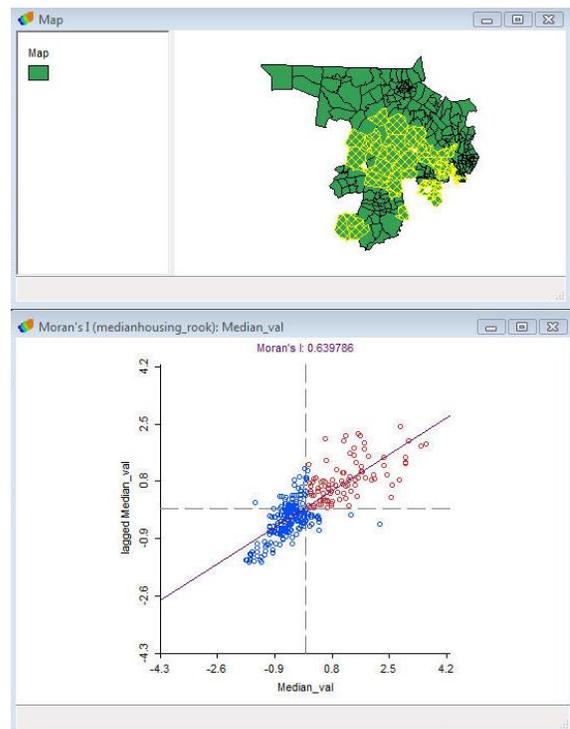
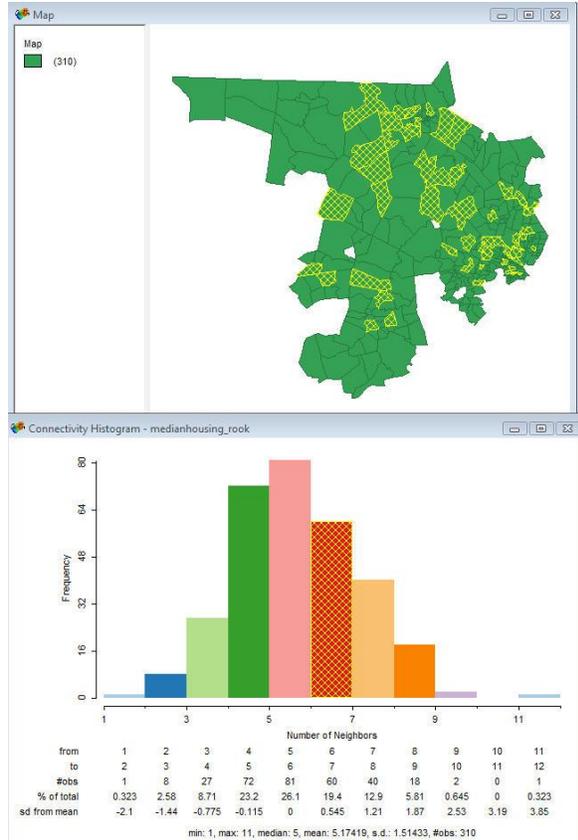
1. Click Space > Univariate Moran's I
2. Select Median\_val as the variable and click Ok.

The result is a Moran's scatter plot with the I value displayed at the top. The X-axis is the value of I and the Y-axis is the spatial lag, which is the weighted average of neighboring values. The slope of the line is Moran's I.

**Q:** What is the I value? What does this mean?

**A:**  $I = .639786$ . Since it is positive, there is an overall pattern of clustering of median housing values. It is also quite close to 1, indicating the values are highly clustered.

The upper right and lower left quadrants represent positive spatial autocorrelation, which means clustering of like values while the lower right and upper left quadrants represent negative spatial autocorrelation or spatial outliers.



3. Select the points in the different quadrants by drawing a box around them with your cursor to see the general areas where clustering and dispersion are occurring.

**Q:** Do there appear to be more clusters or outliers?

**A:** Clusters since most points fall into the upper right or lower left quadrants.

4. Right click on the scatter plot and select Randomization > 999 Permutations

This displays the I value (yellow line) in relation to the reference distribution. You can click Run over and over again to generate new reference distributions. If the yellow line overlaps the reference distribution, your distribution may be due to chance rather than autocorrelation.

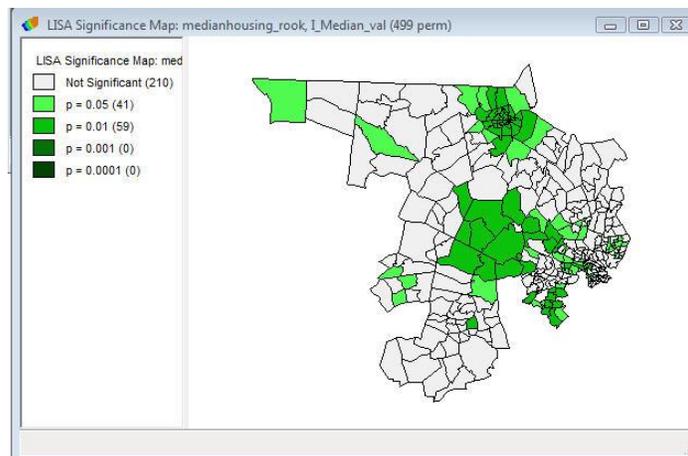
5. Click Done and close the Moran's scatter plot.

## Univariate Local Moran's I

Univariate Local Moran's I calculates local Moran statistics to demonstrate local spatial autocorrelation.

1. Click Space > Univariate Local Moran's I
2. Select Median\_val as the variable and click OK.
3. Check Significance Map and Cluster Map and click OK.

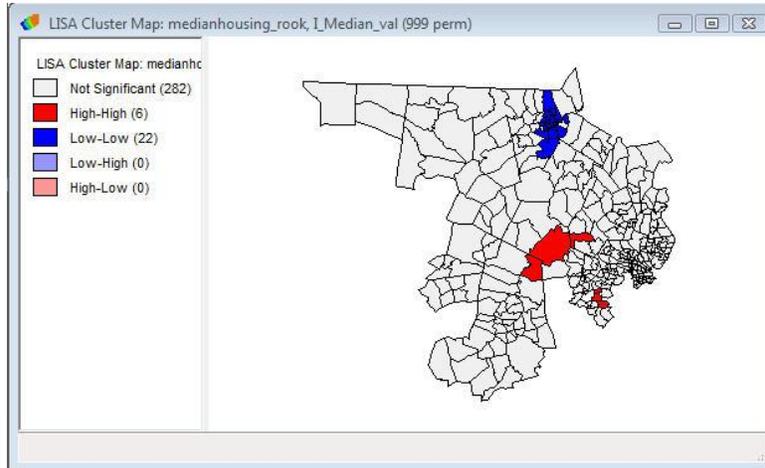
The significance map shows the locations with significant local Moran statistics for various p values. When conducting spatial statistics, you'll often find significant results at  $p=.05$ , so it is good to look at the results for more restrictive p-values. You can adjust the significances displayed and the permutations by right clicking and selecting Randomization and/or Significance filter. The map below shows 499 permutations. Your map will look different depending on the number of permutations that you run and the significance level you choose.



The LISA Cluster Map shows the significant locations color coded by type of spatial autocorrelation. Again, your map will look different depending on the number of permutations and significance level.

The above map shows  $p < .001$  and 999 permutations.

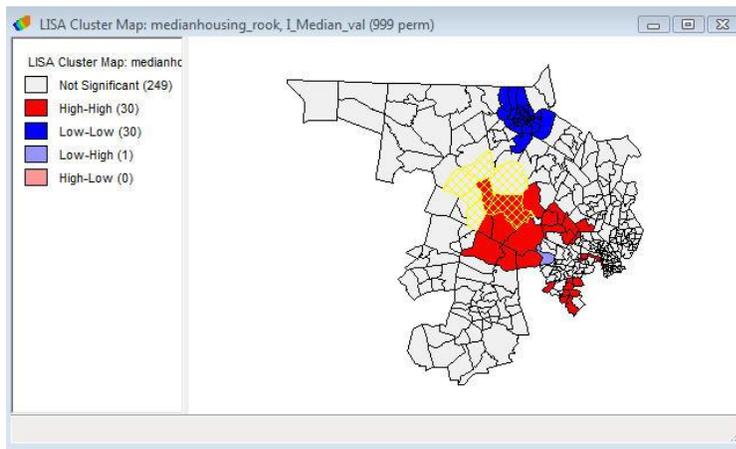
- High-high and low-low = spatial clusters
- High-low and low-high = spatial outliers



The spatial clusters on the map refer to the core of the cluster. The cluster is classified as such when the value at a location (either high or low) is more similar to its neighbors (as summarized by the weighted average of the neighboring values, the spatial lag) than would be the case under spatial randomness. The cluster itself extends to its neighbors.

1. Select one of the census tracts that has significant clustering by clicking it on the map.
2. Right click in the white space on the map and select Add Neighbors to Selection

The map now shows all the neighbors of the target census tract.



ok to exit without saving your project.

**Q:** Where are the clusters on your map? Where are the outliers?

**A:** This will depend on the number of permutations and the significance level. In the map above, there is a cluster of low values in the upper right corner and a cluster of high values in the center of Middlesex County. There is one outlier near the cluster of high values.

3. Close the maps and Geoda. It is

## ArcMap

You will first examine the data generally to determine if median household income is clustered or dispersed in Middlesex County, then you will run local statistical tests to see where clusters exist. You used the Moran's I tests in Geoda and will use the Getis-Ord tests in ArcMap. Only the general Getis-Ord test is available in Geoda.

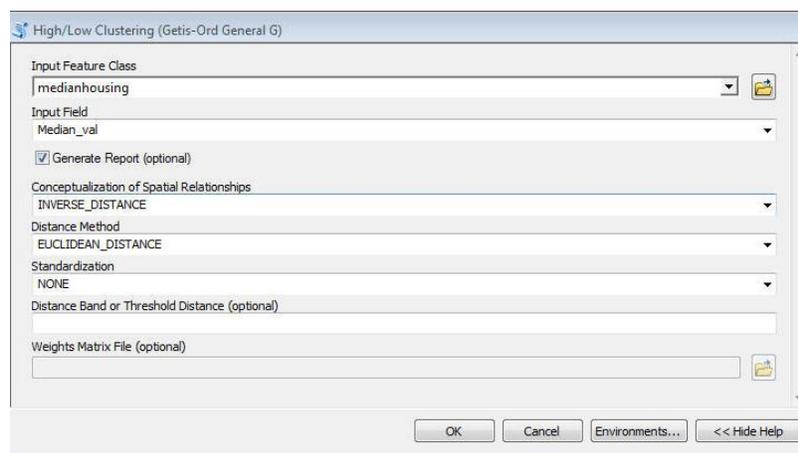
1. Open ArcMap and a blank map.
2. Add the data layer, "medianhousing" to the map from your folder on the desktop.

In the following exercises you will use the default setting for spatial weights in each tool, however you also have the option of creating a spatial weights matrix using this tool: Spatial Statistics > Modeling Spatial Relationships > Generate Spatial Weights Matrix. If you are familiar with network analysis you can construct a spatial weights matrix based on your road network using the Generate Network Spatial Weights tool.

## Getis-Ord General G

You will first run the Getis-Ord General G statistic to see if areas with a high or low median income are clustered.

1. Open the toolbox if it is not already open by clicking Geoprocessing > ArcToolbox.
2. Expand Spatial Statistics Tools.
3. From the *Analyzing Patterns* tools, select *High/Low Clustering (Getis Ord General G)*.
4. For the input feature class, select "medianhousing" from the drop down menu. For the input field, select "Median\_val".
5. Check the box next to *Generate Report*.
6. Keep the Conceptualization and Distance methods as the default. Feel free to experiment with other conceptualizations after you run this analysis.
7. Click *Ok*.

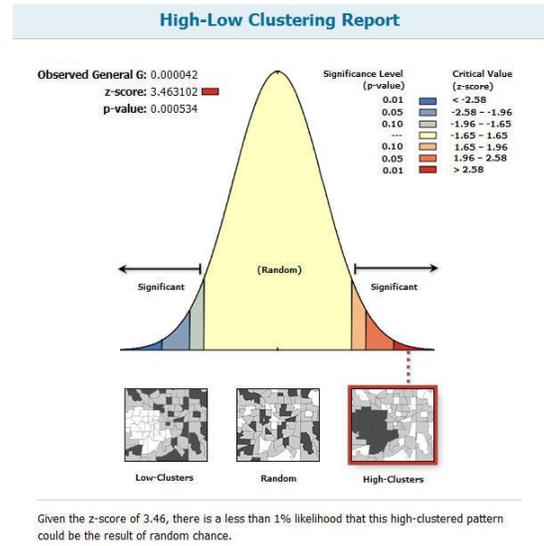


ArcMap warns you that it used the distance of 12597.9 meters for the distance of the search threshold. By default, ArcMap will use the minimum distance to insure all features have at least one neighbor. You can change this distance to make the analysis more appropriate for your project. The larger the value, the more neighbors that will be considered during the calculation.

8. Once the test has finished running, click Geoprocessing > Results (in the top menu bar)
9. Expand Current Session and double click Report File: GeneralG\_Result.html. **Note:** If the graphics do not display properly try using Chrome or Firefox to open the report.

**Q:** What is the G score? What is the Z score? What does this tell you?

**A:** Because the Z score is highly significant, the median household income in each tract is significantly more clustered than a random pattern.



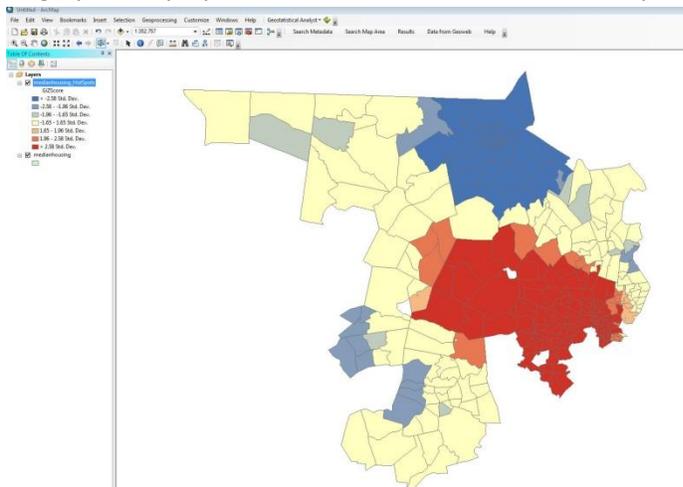
## Hot Spot Analysis (Getis-Ord $G_i^*$ )

While the General G test tell you something about the overall pattern of median housing values in Middlesex County, you will now use a local statistical test to learn more about where clusters are located.

To learn more about where median housing values are clustered, calculate a local version of the G score that you calculated earlier.

1. From the *Mapping Clusters* toolset, select *Hot Spot Analysis (Getis-Ord  $G_i^*$ )*.
2. Use the same inputs as you did for the Local Moran's I test. Save the output feature class to your folder on the desktop. Click *Ok*.

A graphic display of the Z scores will be added as a layer to your map.



**Q:** What do you see?

**A:** You will notice a large statistically significant (Z score < 2.58 Std. Dev or > 2.58 Std. Dev.) “hot spot” (dark red) and a large statistically significant “cold spot.” There are also some sections with moderately high or low Z scores.

To be a statistically significant hot spot, the tract needs to be surrounded by tracts with high, positive values and have a much higher, positive value than its neighbors. The

inverse is true for the cold spots. This is similar to the HH or LL relationships that the Local Moran's I test detects. However, Local Moran's I can also detect HL or LH relationships, whereas Hot Spot Analysis just looks for clusters of similar high or low values.

**Q:** What did you learn about median housing values in Middlesex County? Feel free to add a basemap to ArcMap to give your data context. You can do this by clicking on the drop down arrow next to the Add Data button.

**A:** Low housing values are clustered together in northern Middlesex County around Lowell while high values are clustered in the Western suburbs of Boston.

If you want to examine the median housing values on the map, you can use the Map menu in Geoda to map the Median\_val column. In ArcMap, right click on the medianhousing layer, select Properties > Symbology and select Quantities and then Median\_val from the drop down menu.

MIT OpenCourseWare  
<http://ocw.mit.edu>

RES.STR-001 Geographic Information System (GIS)Tutorial  
January IAP 2016

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.