GEOG 4140/6140 – Winter 2021
Lab 6: Cluster Analysis
Due Thursday, March 11th by 11:59 PM

Overview

Large point datasets can be challenging to visualize and analyze, since just displaying the points shows little detail about the actual distribution or any attributes of the points. Crime data from Manhattan illustrates this well, as even a single year of data contains over 100,000 occurrences, making visual interpretation near impossible. There are various techniques to analyze the distribution of points and display areas of relatively high or low density of occurrences. In this lab we will use a variety of methods to help visualize and quantify the clustering of the points representing the occurrence of crime in Manhattan.

The following resource may be of use in completing this assignment:

Create Fishnet
How Create Fishnet works
Fundamentals of field calculations
Calculate Field
Kernel Density
How Kernel Density works
Spatial Join
Add Geometry Attributes
Spatial weights
How Generate Spatial Weights Matrix works
Generate Spatial Weights Matrix
How Spatial Autocorrelation (Global Moran's I) works
Spatial Autocorrelation (Global Moran's I)
How Cluster and Outlier Analysis (Anselin Local Moran's I) works
Cluster and Outlier Analysis (Anselin Local Moran's I)

Required Data (*Sources*)

1. Manhattan_pts feature class (*Lab data on Canvas*)
2. Census_Tracts feature class (*Lab data on Canvas*)

Workflow

1. Open ArcGIS Pro and create a new project titled *Lab06YourLastName* in your working directory for this lab.
    a. Import the *Manhattan_pts* crime points feature class into your Lab 6 geodatabase, and add these data to the map.

2. Our first analysis of this point data creates a regularly spaced grid and calculates the density of crimes within each grid cell.

    a. Use **Create Fishnet** to produce a regularly spaced grid that covers all of Manhattan with cells 0.0625 km^2 in area. Because the projection is using meters as its linear unit, you will want to use a conversion tool to convert from square km to square m, then take the square root of that number to find the desired cell width in meters.

        i. Set the output extent to match the extent of the Manhattan Points layer, specify the cell width, turn off the "Create Label Points" option, and set the geometry type to Polygon.

    b. Then calculate the number of crimes per square kilometer occurring in each cell.

        i. Run the spatial join tool to get a count of crime occurrences per cell.

        ii. Run the ' add geometry attributes' tool, to calculate area in square km for each cell.

        iii. Add a new field named CrimePer_SqKM. Run a field calculator on this field to calculate the number of crime occurrences per square km.

3. Import the *Census_Tracts* crime points feature class into your Lab 6 geodatabase, and add these data to the map.

    a. Calculate the crimes per square kilometer for each census tract. (Hint: use the **Add Geometry Attributes** tool to calculate the area in square kilometers for each census tract. Then, repeat the pattern from Step 2b to determine the crimes committed per square kilometer for each census tract.)

4. Create a layout map with two panels: one showing the number of crimes per square kilometer within each grid cell and the other showing the number of crimes per square kilometer within each census tract for Manhattan. Be sure to include all standard map elements. (Deliverable 1)

5. The next method of analysis we will use is the Kernel Density function. This analysis will create a raster displaying the density of points in a specified radius from each cell. We will apply this method to the Manhattan Points data.

    a. Calculate the **Kernel Density** with a search radius of 500 meters. Then run it a second time with a search radius of 1 kilometer. Observe the differences between the two outputs.

    b. (If you get an error regarding the spatial analyst license) go to the "Project tab -> Licensing -> Configure your licensing options , and make sure Spatial Analyst is enabled.

6. Find the location of the highest crime density by visually comparing all 4 of the analyses created in the previous steps (Kernel Density 500m, Kernel Density 1000m, Fishnet, and the Census Tracts). Note any differences in magnitude and location of these maxima.
(Deliverable 2)

7. While the previous analyses have provided visual information about the density of crime occurrences, we will now focus on statistical measures of clustering. In order to do this, we must first run the Generate Spatial Weights Matrix tool for the census tracts.

a. For the Conceptualization of Spatial Relationships use "Contiguity edges corners"

b. The output from this step will generate a file with a ".swm" extension. Save this file to a folder on your N drive.

8. Next, we will measure spatial autocorrelation in our data using Moran's I Index.
    a. Use the **Spatial Autocorrelation (Global Moran's I)** tool, with the census tracts as the input file and the normalized crimes field as the input field. Click generate report and specify the spatial weights matrix from step 7 as the conceptualization of spatial relationships.
    b. The output from this step will be an ".html" file that can be viewed in a web browser. (Deliverable 3)

9. In order to understand the clustering of the data, we will use the **Cluster and Outlier Analysis** tool.
    a. Again, be sure that you use the spatial weights matrix from Step 7.

10. Create a layout map showing the results of the Cluster and Outlier analysis from Step 9. Be sure to include all required map elements. (**Deliverable 4**)

<u>Deliverables</u>

1. A map of the gridded crime data and the census tract aggregated data for Manhattan in two separate map panes. (10 points)

2. Describe the differences in the types of analysis conducted for this lab, compare the results, and describe some pros and cons of each method. In addition, describe some of the data requirements for each method. (3 points)

3. What percent likelihood is there that the pattern of crime occurrences in Manhattan could occur by chance; how did you reach your conclusion? (1 point)

4. A map of the Cluster and Outlier analysis results. (6 points)