Spatial Exploratory Data Analysis- ArcGIS Supplement

In this supplemental document, the underlining spatial analysis and preliminary exploration of the data is given. This is meant to viewed in conjunction with the STATA output as well as the main project, which uses Spatial Point Patterns and Geostatistics done in R. Since it is important to consider the visualization of the spatial data before modeling, we use ArcGIS to create high quality images and utilize some of the basic functions built into the software to compare and contrast methodologies and implementations. Spatial Analysis is a multi-disciplinary field so it helps to keep an open mind about the different ways to approach the problems of analyzing geographic based data.
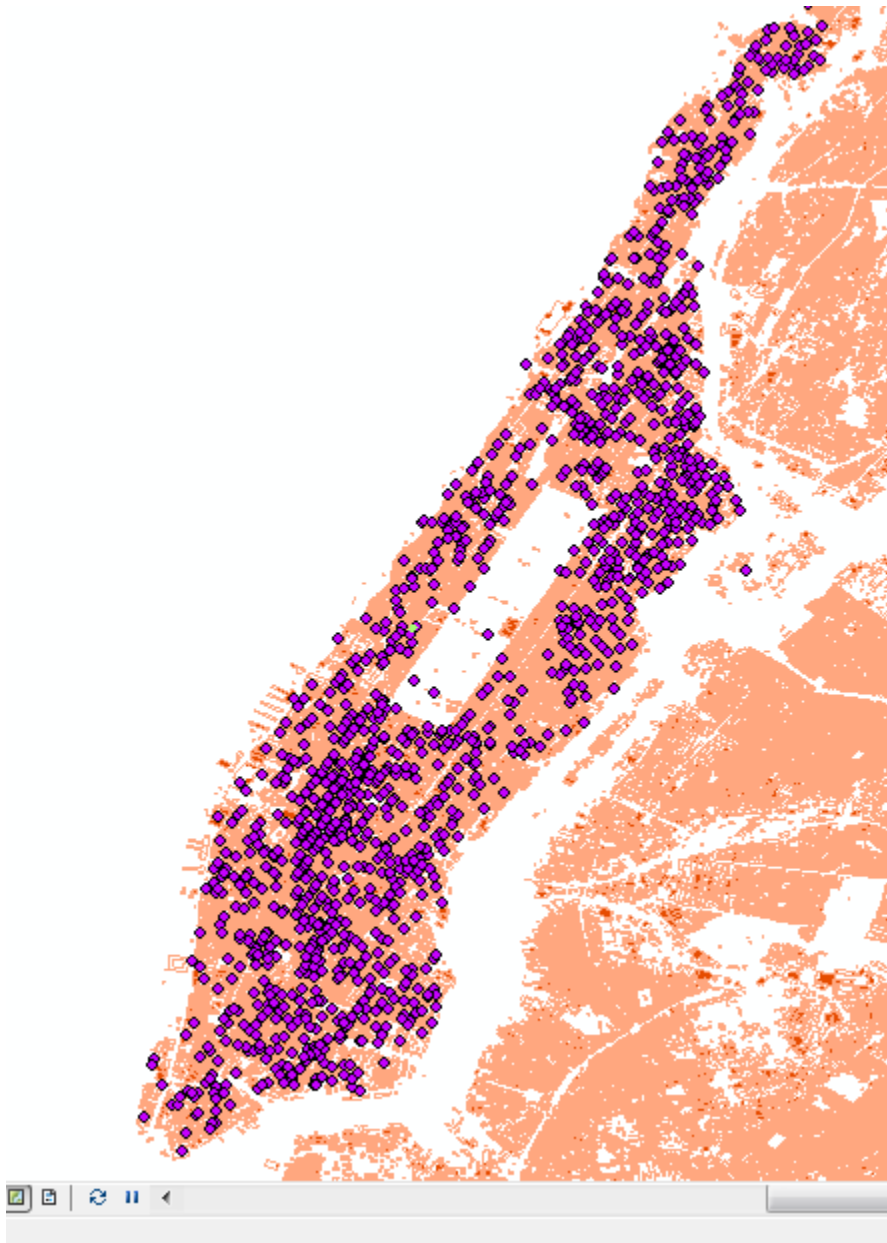
General Views of Points

Figure 1- Plot of (aggregate) individual points where a complaint was logged. The underlying map layer is provided also provided by NYC Open Data. Only occurrences in Manhattan are shown.

From a bird's eye view it seems that only parts of the Upper East Side remain free of incidences. But appearance can be deceiving when looking at this. It's hard to discern which part, if any, have more concentration (and if so, how much?) of incidences than other parts of the borough. Also, be aware that these points are generated using point scale arbitrarily. If we change the font sizes, the pattern might look very different depending on the analyst's choices.

Severity Level of Complaint Considerations

Figure 2-A: Points of Felonies Only.

GIS software allows easy filtering based on attributes from the dataset. Comparing the three major types of offense levels can yield different interpretations of the severity of the crime led by the complaints.

We see that the area of Upper Manhattan on the East Side above Central Park as well as the bulk of Midtown area contain large concentration of Felony level offenses.
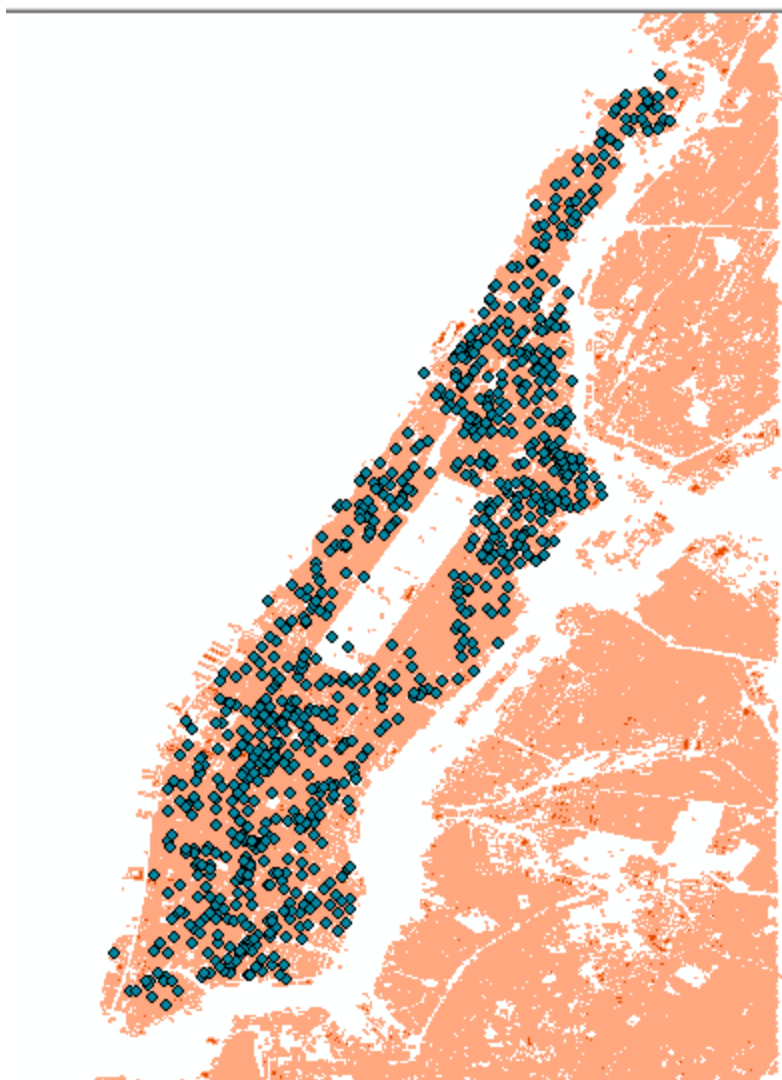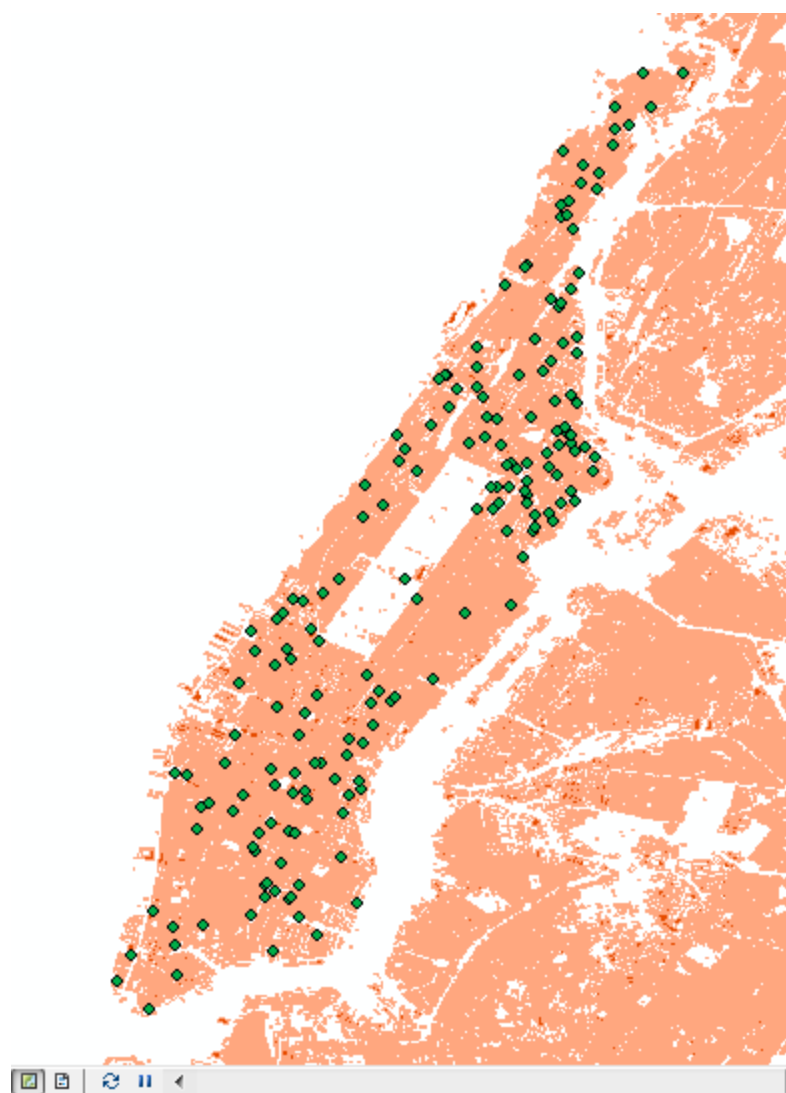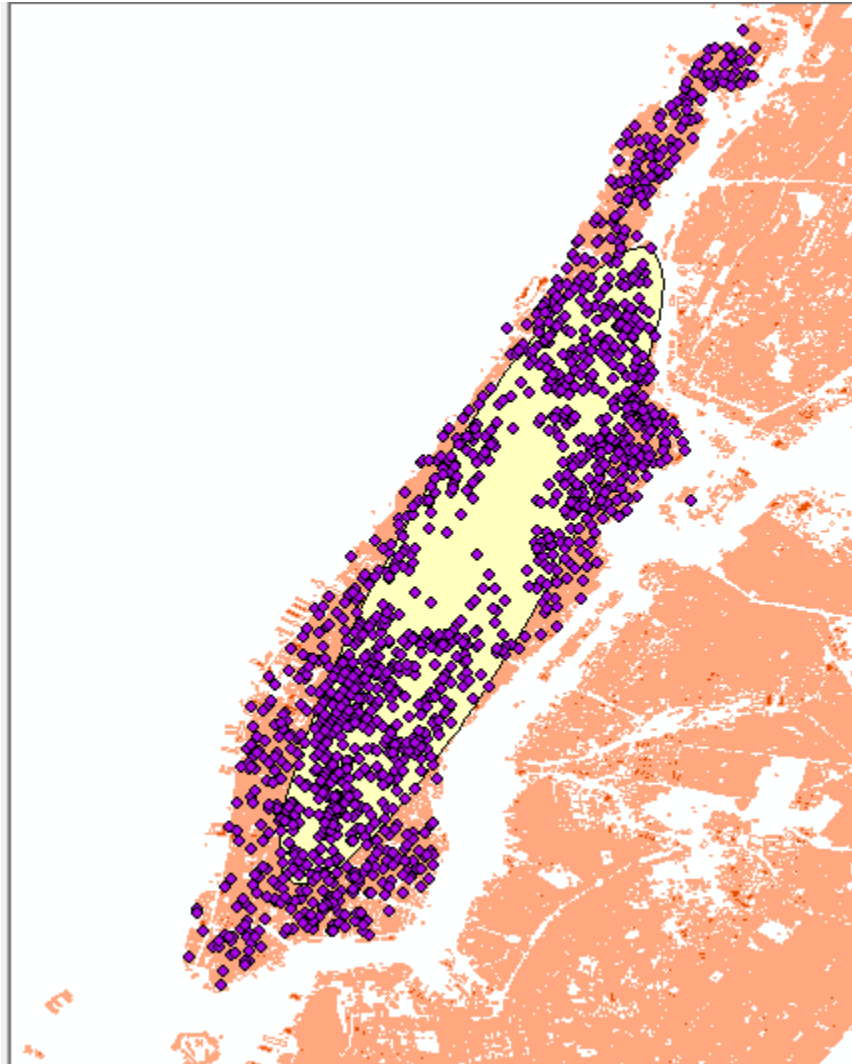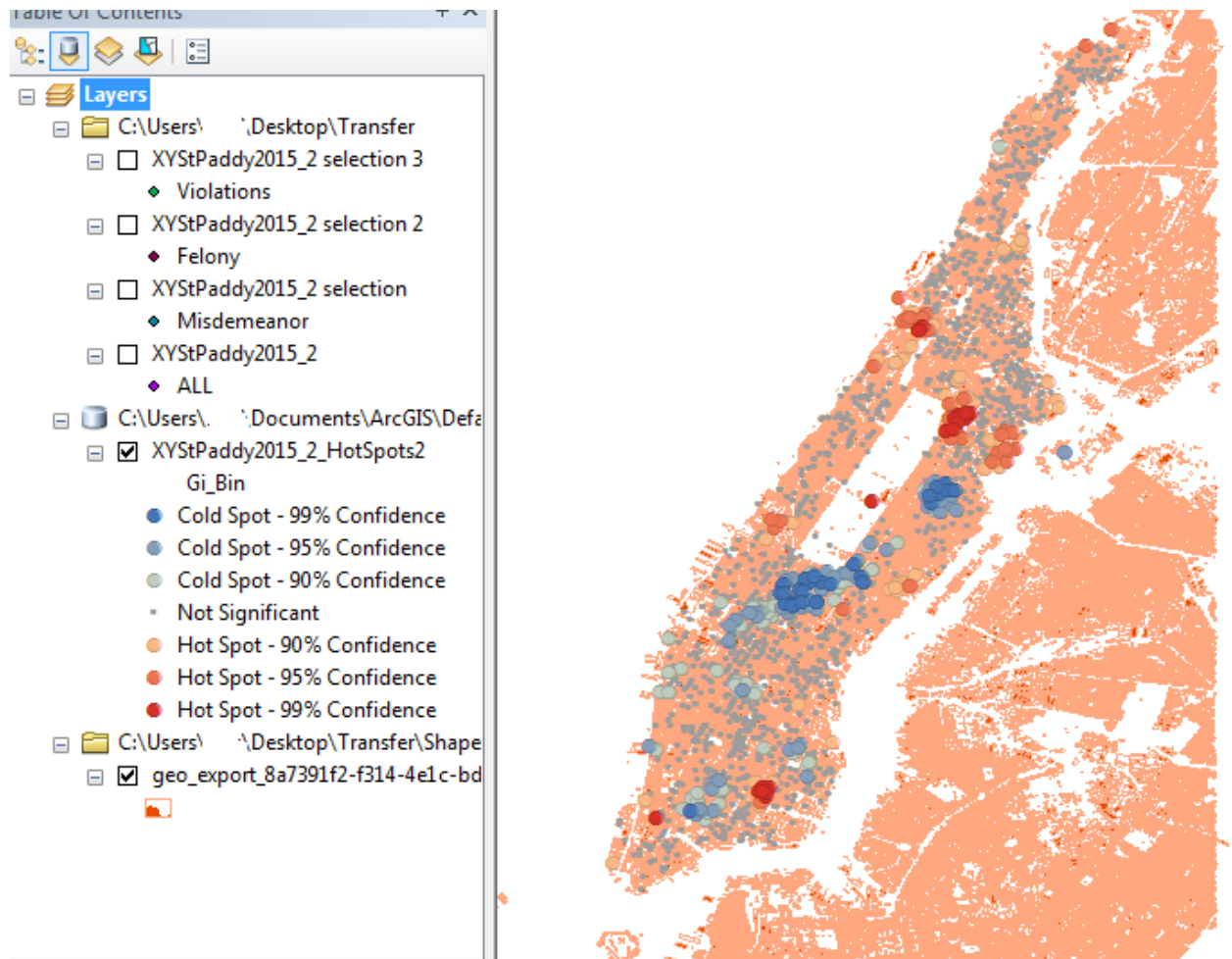
Basic Functions of Geostatistics

GIS software like ArcGIS have basic geostatistical functions built in natively through its Geoprocessing mode and ArcToolbox. While anything done using these functions can also be done in R, it can be useful to consider them due to convenience. Here, we include some of the most commonly used measurements in the, mostly related to clustering and dispersion of points, and their exploratory visualizations.

## Directional Distribution (Standard Deviational Ellipse)

Hot Spot Analysis

Table Of Contents

- Layers
  - C:\Users___\Desktop\Transfer
    - ☐ XYStPaddy2015_2 selection 3
      - ◆ Violations
    - ☐ XYStPaddy2015_2 selection 2
      - ◆ Felony
    - ☐ XYStPaddy2015_2 selection
      - ◆ Misdemeanor
    - ☐ XYStPaddy2015_2
      - ◆ ALL
  - C:\Users___\Documents\ArcGIS\Defa
    - ☑ XYStPaddy2015_2_HotSpots2
      - Gi_Bin
      - ● Cold Spot - 99% Confidence
      - ● Cold Spot - 95% Confidence
      - ● Cold Spot - 90% Confidence
      - · Not Significant
      - ● Hot Spot - 90% Confidence
      - ● Hot Spot - 95% Confidence
      - ● Hot Spot - 99% Confidence
  - C:\Users___\Desktop\Transfer\Shape
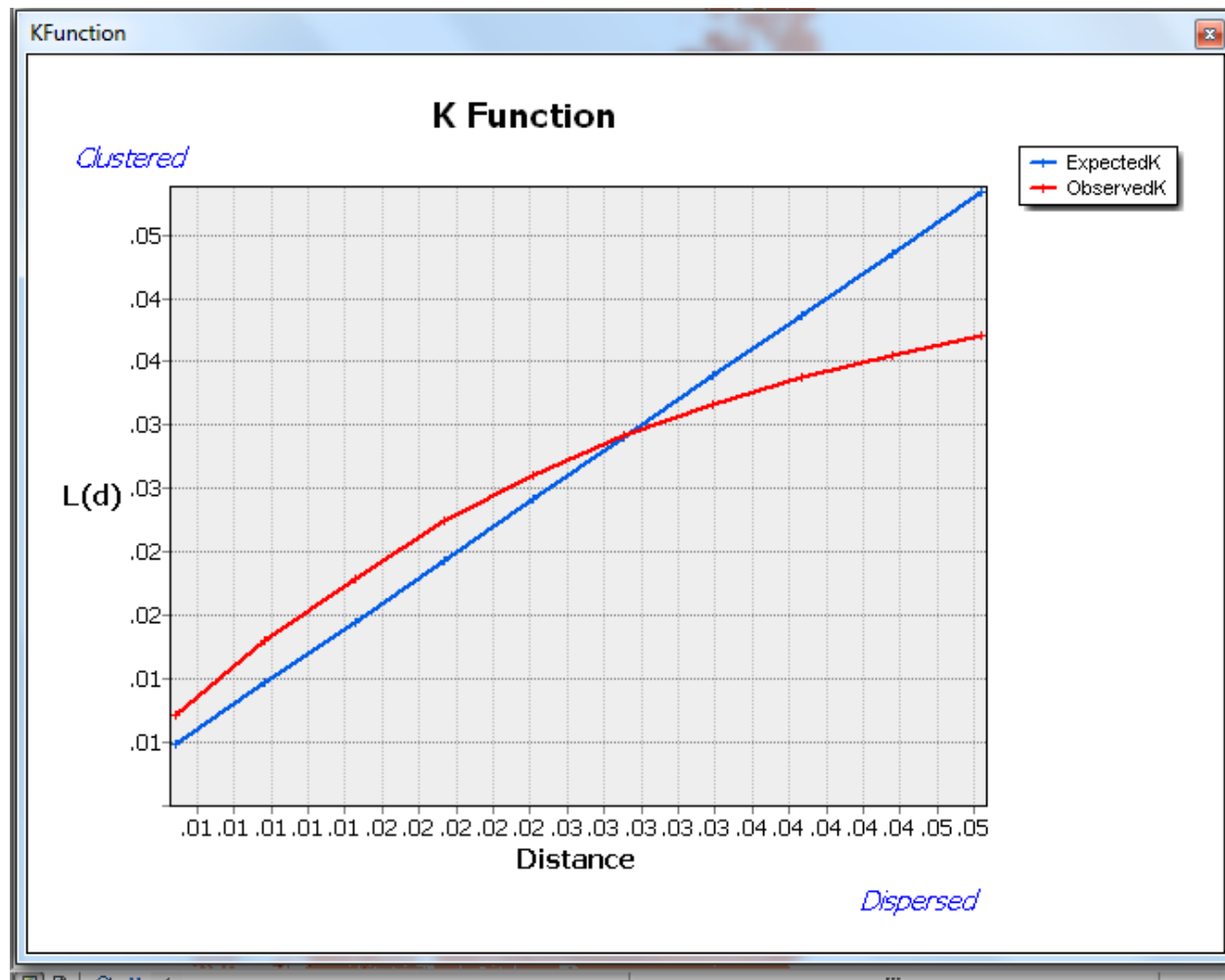   - ☑ geo_export_8a7391f2-f314-4e1c-bd

Spatial Autocorrelation (Moran's I)

Local (Neighborhood) Moran's I For Cluster and Outlier Analysis
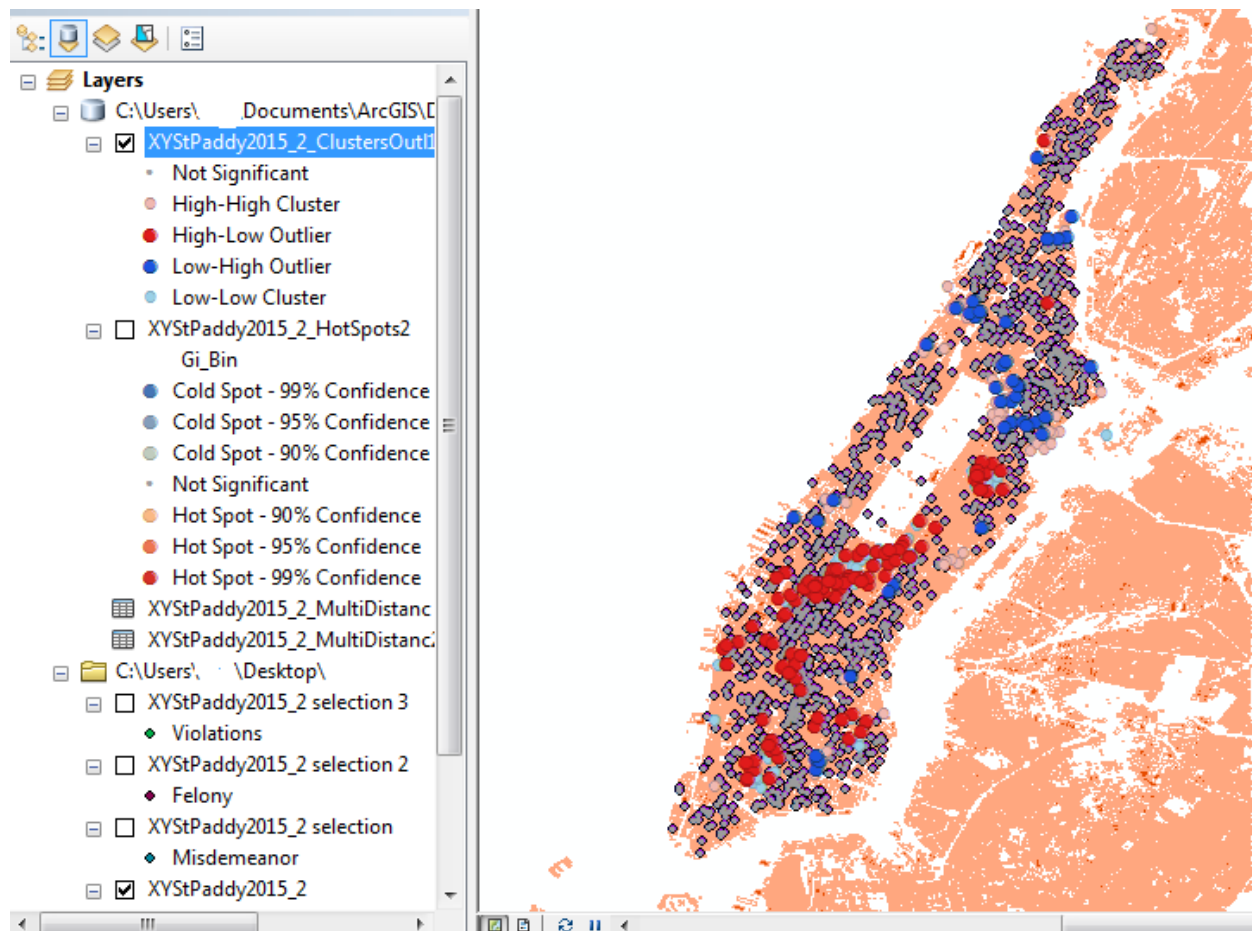
Overall Spatial Clustering (Ripley's K Function)

Ripley's K Function with Offense Code as weighted variable and minimal of 10 distance bands.