Visualizing Crime in Philadelphia; A Multivariate Spatio-Temporal Analysis

Intent

The intent of this project is to explore the use and validity of VIS-STAMP as a methodology for exploring patterns of change in social phenomena. Challenges arise when examining and analyzing data sets which contain attributes with space, time, and multiple variables. Analysis is routinely performed in a single dimension, and these limited analytical approaches can result in significant patterns being missed or overlooked. A more thorough analysis requires software that can process, examine and analyze complex, multi-dimensional data.

Crime data exhibits these complex spatio-temporal and multivariate characteristics. Examine Figure 1, which shows crime type rates indexed to 2006 levels. Visualizing data at this level can help us see the overall, general trends. However, this analysis fails to explore these trends spatially. VIS-STAMP offers a method to develop a holistic view of how crime patterns have changed in Philadelphia from 2006 to 2013.

One approach to overcome the challenge of analyzing complex data sets is VIS-STAMP, a free software package developed by University of South Carolina's Department of Geography. VIS-STAMP uses visual, cartographic, and computational methods to display data in an integrated system. VIS-STAMP performs multivariate clustering and visualizes the clusters spatially and temporally. The software is comprised of a Self-Organizing Map (Figure 2), a scalable Parallel Coordinate Plot (Figures 3 & 4), and a re-orderable map matrix (Figure 5). Each of these components acts as a window to view the data, and allows for user interaction to generate visual graphics reflecting clusters, their relationships and potential patterns. The Self-Organizing Map and the Parallel Coordinate Plot are organized by color, and act as the map legends.

Methodology

This analysis is performed at the neighborhood level. The crime data is released by the Philadelphia Police Department and contains 15 types of crime. These were aggregated and reclassified as indicated in Figure 6. Both the crime data and the neighborhood boundary files were acquired from OpenDataPhilly.org.

Within VIS-STAMP, the values of crime type are normalized to Z-scores, and then processed by a Self-Organizing Map (SOM). The SOM clusters multivariate data it into a 2D layout and groups similar clusters by color. The SOM allows users to select the number of clusters to be generated in order to moderate and control the total number of clusters. Euclidean distance is used to assess the similarity between spatial objects. Each colored circle represents a cluster, where the circle size denotes the proportional size of the cluster. The hexagons are shaded to denote the dissimilarity of clusters, where darker shades represent more dissimilarity. The colors assigned to clusters are then mimicked throughout the rest of the analysis

The Parallel Coordinate Plot (PCP) acts as a legend, showing the meaning of the colors assigned by the SOM. The PCP visualizes the composition of the variables that make up a cluster. The axis of the PCP can be manipulated and shown several ways (Figures 3 & 4).

Findings

In this analysis, 16 crime clusters were generated. In Figure 7 each of the clusters is colored and named to simplify discussion. An in-depth analysis of findings is not feasible within the constraints of this poster. Instead general trends and examples will be discussed. Figure 4 demonstrates the overall, global cluster trends in Philadelphia.

Both the Orchid and Purple clusters exhibit an interesting pattern of crime distribution. Both have generally low crime rates, but conversely have high theft rates. Generally these neighborhoods are in the Southwest and Center City.

Blue is the largest cluster, and the cluster most similar to adjacent clusters. The PCP indicates that Blue contains the lowest crime rates for all crime types. Looking to the Map Matrix, this cluster appears to be predominately located in parks and industrial areas, i.e. low population areas. In contrast, Red contains the highest overall rates of crime, except for theft. The number of Red neighborhoods generally decreases over time, and are predominately located in North, North Central, and West Philadelphia.

Fuchsia is the smallest cluster and the cluster least similar to adjacent clusters. Fuchsia represents relatively high crime rates across all crime types. The crime type distribution is unique; Fuchsia has the second highest theft rate but with a lower homicide rate than other high crime areas. Fuchsia neighborhoods are geographically clustered, predominately in the Near Northeast.

Figure 5 geographically illustrates the pattern of neighborhood cluster change over time. Generally, neighborhood areas appear to shift to clusters with lower crime distributions. For example, the decrease in the number of Red neighborhoods (the highest crime cluster) is readily apparent.

Aside from a holistic, visual examination of neighborhood change, the number of times each neighborhood underwent a cluster change is calculated. The change is mapped in Figure 8. All cluster changes are counted (not just unique cluster shifts). The number of cluster changes ranges from 0 to 7, with some neighborhoods never shifting clusters while some other neighborhoods shift annually. Interestingly, high change neighborhoods and low change neighborhoods do not appear to be geographically separated.

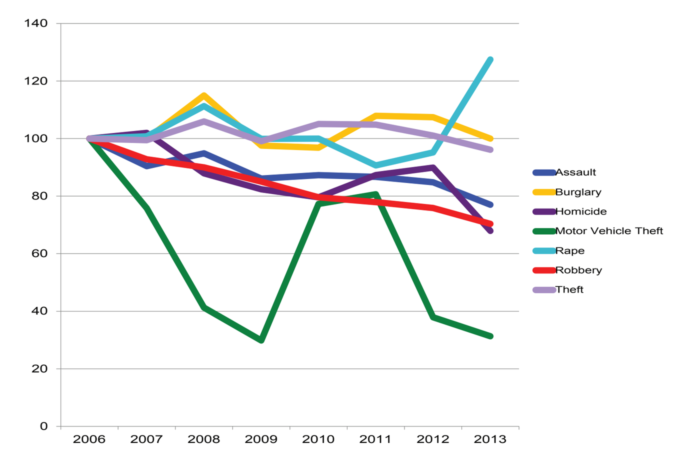

Figure 1 - Indexed Crime

Initial Offense Index	Reclassified Offenses
'Aggravated Assault Firearm'	Assault
'Aggravated Assault No Firearm'	
'Burglary Non-Residential'	Burglary
'Burglary Residential	
Homicide – Criminal'	Homicide
'Homicide - Criminal '	
'Homicide - Gross Negligence'	
'Homicide - Justifiable '	
'Motor Vehicle Theft'	Motor Vehicle Theft
'Rape'	Rape
'Recovered Stolen Motor Vehicle'	N/A
'Robbery Firearm'	Robbery
'Robbery No Firearm'	
'Theft from Vehicle'	Theft
'Thefts'	

Figure 6 - Crime Classifications

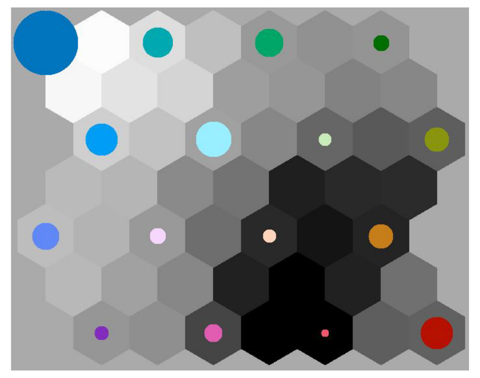

Figure 2 - SOM

Figure 7 - Cluster Names

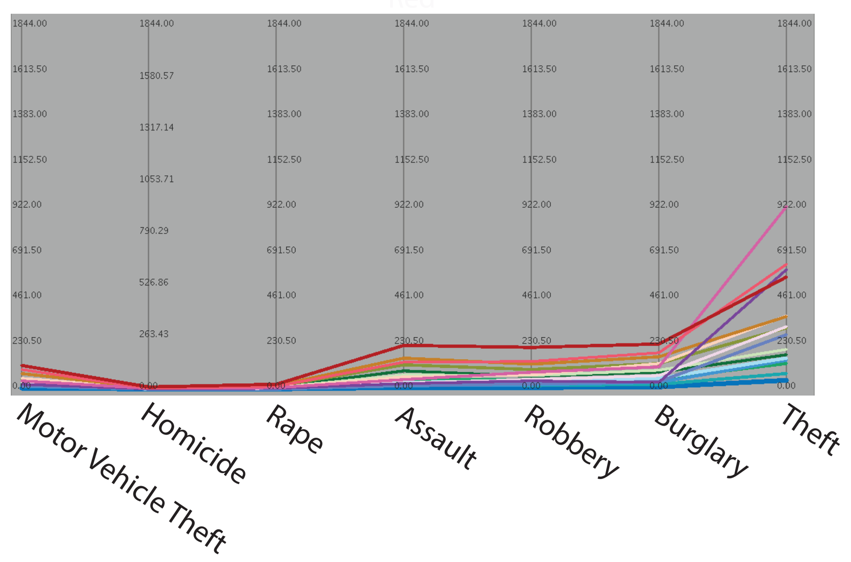

Figure 4 - Global PCP (axis is scaled linearly by the global minimum and maximm values)

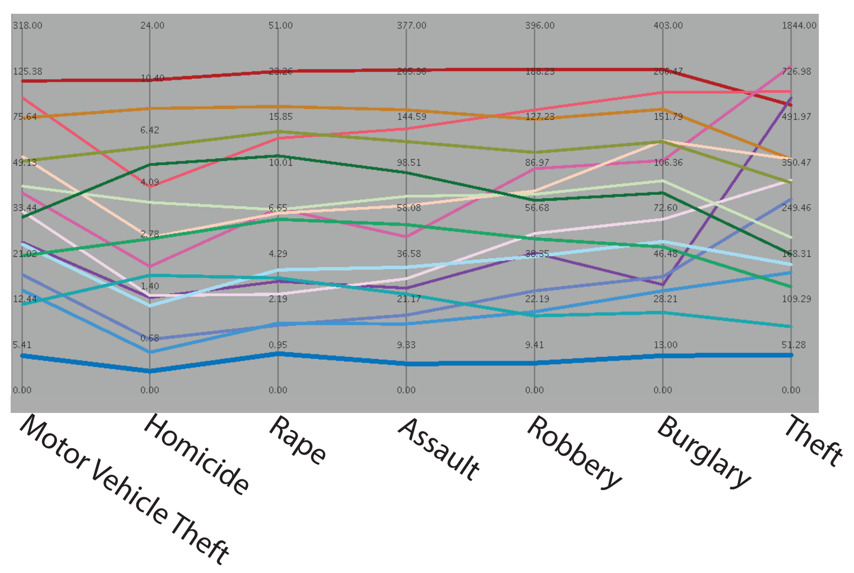

Figure 3 - Nested Means PCP (axis scaled non-linearly by adjusting the spacing of intervals according to data distribution values)

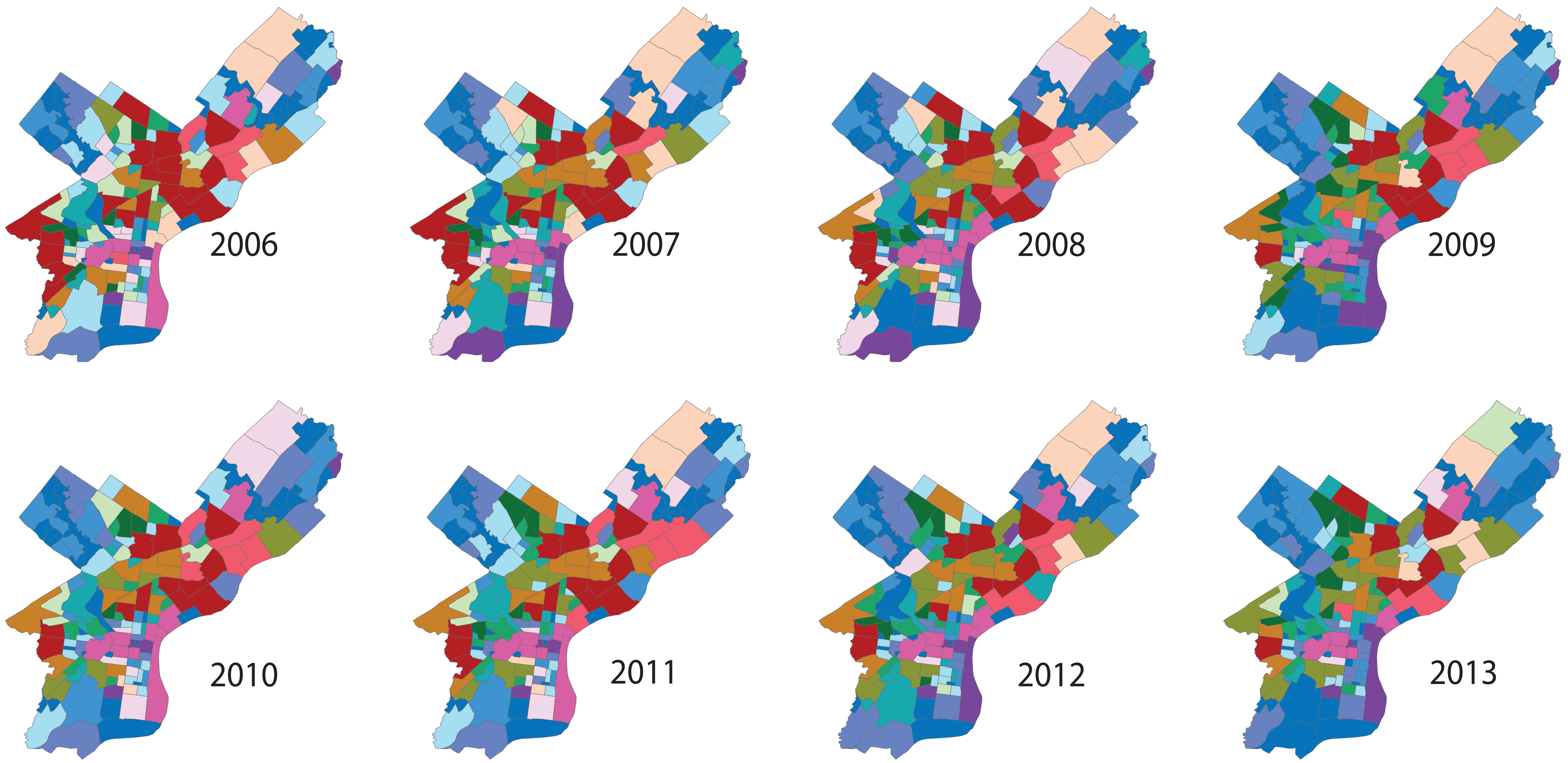

Figure 5 - Re-orderable Map Matrix

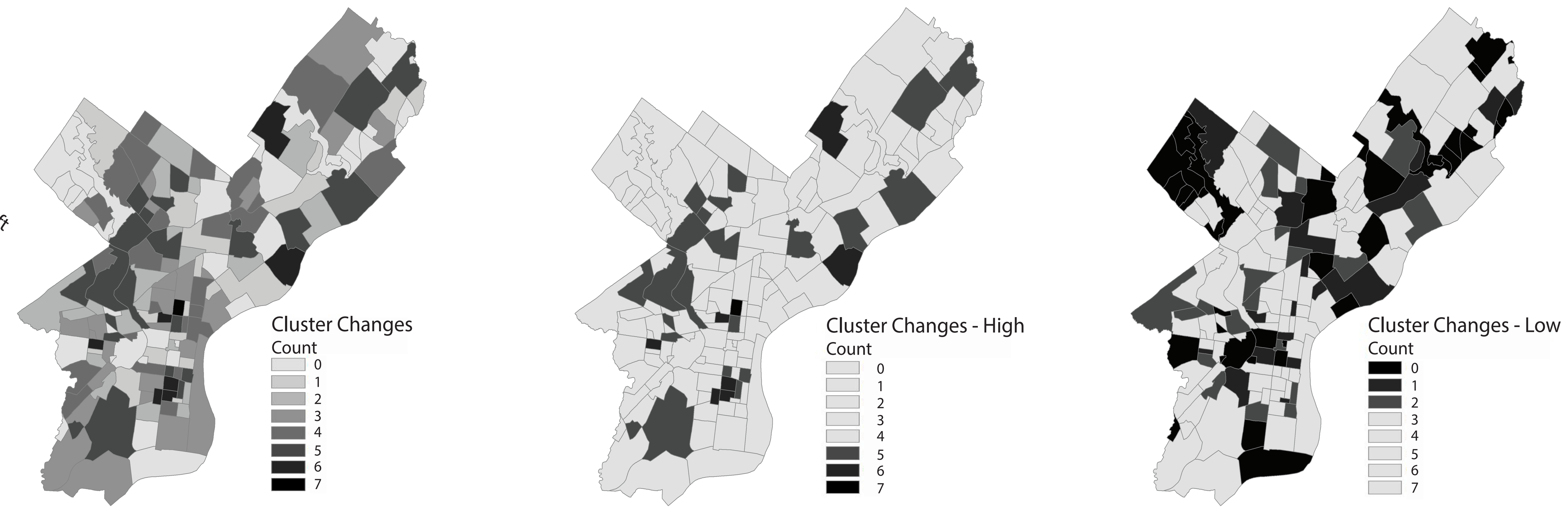

Figure 8- Maps of Cluster Changes

Penn

Morgan Findley
MUSA 2014