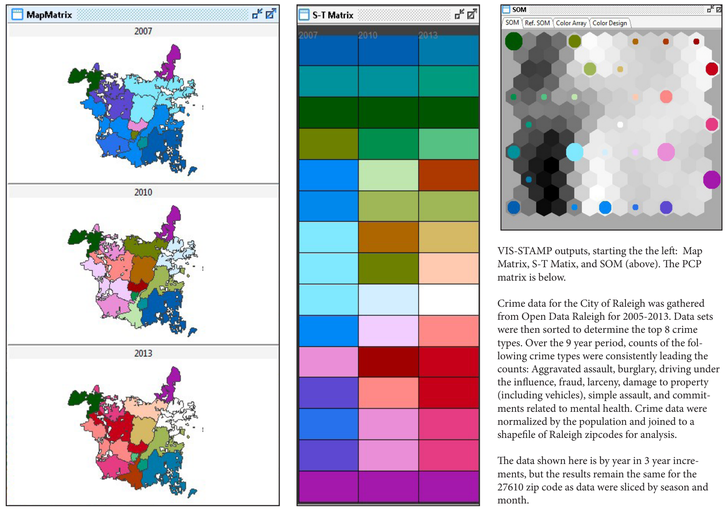
Examining Spatio-temporal Crime Patterns and High Crime Areas in Raleigh, North Carolina

Emily Antoszyk, University of Pennsylvania, Spring 2014

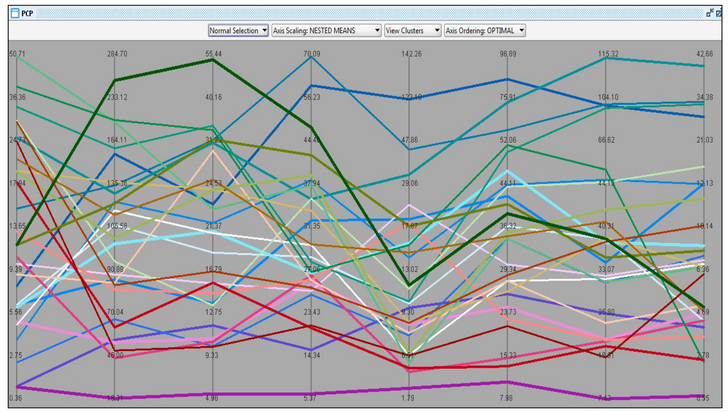
Data Mining with VIS-STAMP



VIS-STAMP outputs, starting the the left: Map Matrix, S-T Matix, and SOM (above). The PCP matrix is below.

Crime data for the City of Raleigh was gathered from Open Data Raleigh for 2005-2013. Data sets were then sorted to determine the top 8 crime types. Over the 9 year period, counts of the following crime types were consistently leading the counts: Aggravated assault, burglary, driving under the influence, fraud, larceny, damage to property (including vehicles), simple assault, and commitments related to mental health. Crime data were normalized by the population and joined to a shapefile of Raleigh zipcodes for analysis.

The data shown here is by year in 3 year increments, but the results remain the same for the 27610 zip code as data were sliced by season and month.

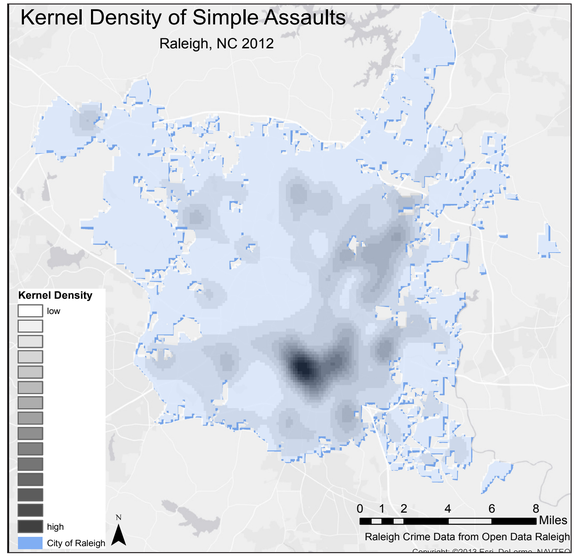


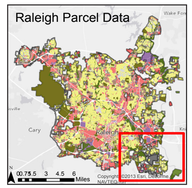
VIS-STAMP outputs four data visualization tools: the Map Matrix, Self-Organizing Map (SOM), Parallel Coordinate Plot (PCP) and Spatial Temporal Matrix (S-T Matrix).

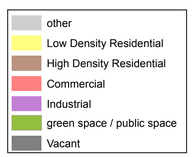
- Map Matrix: gives a visual depiction of clusters by areal designation over time, here it shows changes in crime clustering all over Raleigh, but one zip code - 27610 (blue cluster to the bottom right in each map) - is consistently high crime over the 9-year period.
- SOM: shows the relationship clusters by color, with larger color circles indicating larger clusters. The background coloring in the SOM is also significant; white and lighter backgrounds show more similarity between adjacent clusters while dark grey or black backgrounds indicate more polarized clusters. This SOM reveals that blue clusters are relatively large and mostly set apart from adjacent neighbors.
- PCP: shows the nested means of each crime type for each cluster, in this case, the blue strands are associated consistently with higher crime crime areas across crime types.
- S-T Matrix: inspects each zip code and its accompanying clusters, enabling users to see how individual zip codes have changed over time (high crime to low, steady increase/decrease, stasis). While some change / remain the same over time, the blue cluster in the 27610 zip code is still of interest.

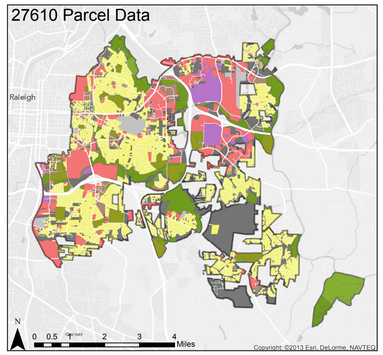
Running analyses in VIS-STAMP helped to inform subsequent regressional analyses.

Visualizing Crime in Raleigh









	27610	Raleigh
Variables		
Male Population*	46.2%	48.3%
Caucasian*	19.9%	57.5%
Black / African American*	65.5%	29.4%
Asian*	1.5%	4.3%
Hispanic*	17.3%	11.4%
Median household income**	$44,085	$53,610
Persons with income below the poverty level**	20.5%	14.6%
Unemployment rate**	14%	6%
High school dropout rate**	9.6%	17.1%
Percent vacant**	8%	9.4%
Family households**	67%	75.7%
Renter occupied**	38.7%	44.2%
Population Density	1536.8 persons / sq. mi.	2827.2 persons / sq. mi.

*Data from the 2010 decennial census
**Estimates based on 5-year estimates 2008-2012 from the US Census

Maps and charts give context to the study area. The first map shows crime density of the eight major crime types from the regressional analyses across Raleigh. Parcel maps in the center show the layout of the city in terms of current land use, and a close up of the 27610 zip code. The final table shows statistics for the city of Raleigh gathered from the US Census as compared with statistics from the 27610 zip code.

Regressional Analyses

Block Groups Raleigh	Adjusted R^2
Aggravated Assault	0.45
Burglary*	0.43
DUI*	0.33
Fraud*	0.33
Larceny*	0.40
Mental Health Commitments*	0.23
Damage to government & private property, or cars*	0.45
Simple Assault*	0.57

*Indicates variables that were transformed using a log transformation.

Table 1 (left): Adjusted R-squared values for each OLS regression model.

Table 2 (below): Reports significant predictors for each dependent crime type and the direction of the relationship resulting from each OLS regression model for all block groups within the city of Raleigh. For instance, an increase in the total population is associated with an average increase in the instance of aggravated assault whereas an increase in distance to the nearest bus stop is associated with an average decrease in the instance of aggravated assault.

Table 3 (bottom): Results of the Pearson Correlations run for each crime predictor in the 27610 zip code.

Predictors tested for significance in all cases include spatial variables - whether or not a block group belongs to the Urban Progress Zone (UPZ), distance from the nearest park and distance from the nearest bus stop - as well as demographic variables including population density, percent of male population, number of white, black, Asian, and Hispanic residents, family households, high school drop out and unemployment counts, median household income figures, renter occupancy, vacant properties, and those with income below the poverty level.

	UPZ	*Pop. Density*	*White*	*Black**	*Asian**	*Hispanic**	*Dropout*	*Dist. to nearest park**	*Dist. to nearest bus stop**	*Vacant**	*Renter occupied**	*Persons with income below the poverty level**	*Family households**	*Unemployed**
Aggravated Assault	inc	dec	dec	inc			inc		dec			inc		inc
Burglary*			dec	inc		inc			dec		inc		inc	
DUI*	inc	dec							dec		inc		dec	
Fraud*		dec		inc	Inc				dec	inc	inc		inc	
Larceny*		dec		inc					dec		inc			
Mental Health Commitments*				inc						inc	inc	inc		
Damage to government & private property, or cars*						inc								
		dec		inc			dec		dec	inc	inc			
Simple Assault*		inc	dec	inc					dec		inc	inc		dec

*Indicates variables that were transformed using a log transformation.
Percent male population, median household income (variable transformed), were removed from this table because of lack of significance in any of the above tests.

	UPZ	*% Male Pop.*	*Pop. Density*	*White*	*Black**	*Asian*	*Hispanic*	*Drop out*	*Median HH income*	*Distance to nearest Park**	*Distance to nearest bus stop**	*Renter occupied**	*Unemployed**	*Vacant**	*Persons with income below poverty level**	*Family HHs**
Aggravated Assault*	+			-					-	-	-	+			+	
Burglary					+	+	+	+	-		-	+	+	+	+	+
DUI				-												
Fraud		+			+							+			+	+
Larceny												+			+	
Mental Health Commitments*		+										+				
Damage to government & private property, or cars			+		+		+		-		-	+			+	
Simple Assault			+						-	-	-	+			+	

*Indicates variables that were transformed using a log transformation.

To further explore the crime data, SAS statistical software was used to run eight separate analyses for each crime type at the block group level for all of the block groups in Raleigh. Each analysis began using 16 variables, which were whittled down to seven or fewer using stepwise regressional tactics. Selecting significant predictors through stepwise regression enabled a leaner analysis through OLS regression, the results of which are in Tables 1 & 2, above. Higher r-squared values (on a scale of 0-1, higher values indicating how much of the variation within the dependent variable, or crime type, is explained by the model) show models that are better explained by the predictive variables; in the batch of models shown to the above, the variables are best associated with the instance of simple assaults, and least likely to predict the instance of mental health commitments.

The 27610 zip code, consisting of 33 block groups, was identified as consistently high crime in the VIS-STAMP analyses. While there are not enough records (i.e. block groups) to conduct a more intensive stepwise / OLS regressional analysis, Pearson Correlation Matrices for each crime type give indication of potential predictive variables in each case. It is acknowledged that correlation does not indicate causation, but it is nonetheless interesting to see similar patterns emerge in the two analyses; in particular, that predictor variables seem to associate more strongly and consistently with more violent crimes: aggravated & simple assault, burglary, and damage to property.

Special thanks to Dana Tomlin for project guidance, Eugene Brusilovsky for aid with VIS-STAMP & Statistical Methods, Tara Jackson & Krista Heinlen from the University of Pennsylvania's Cartographic Modeling Lab for project direction assistance
Crime data: OpenData Raleigh, demographic data: Social Explorer, vacancy data: USPS, Transit data: GoTriangle.org, Urban Progress Zone boundaries: City Planning Department of the City of Raleigh