Regression Analysis of Portland Violent Crime

Jacob Argueta and Dylan Eglin | GIS II | Professor Duh

Introduction

Geographic Information Systems (GIS) applications and geographic databases have become an important component to local and federal police departments. GIS involvement in police departments has only become prevalent in the last decade and is now one of the primary tools in crime prevention. These applications are widely accessible and fairly affordable compared to other new age technology. Analytic mapping has helped better understand local and national trends. Since the incorporation of GIS tactics in the police force the crime database has grown exponentially. Now researchers are able to relate point data with attributes about the crime. This has made it easier than ever to analyze contributing factors to a certain crime or area.

In effort to replicate what GIS analysts may conduct for the police department or a security company we have experimented with ESRI ArcMap spatial statistic tools using crime data from Portland, Oregon.

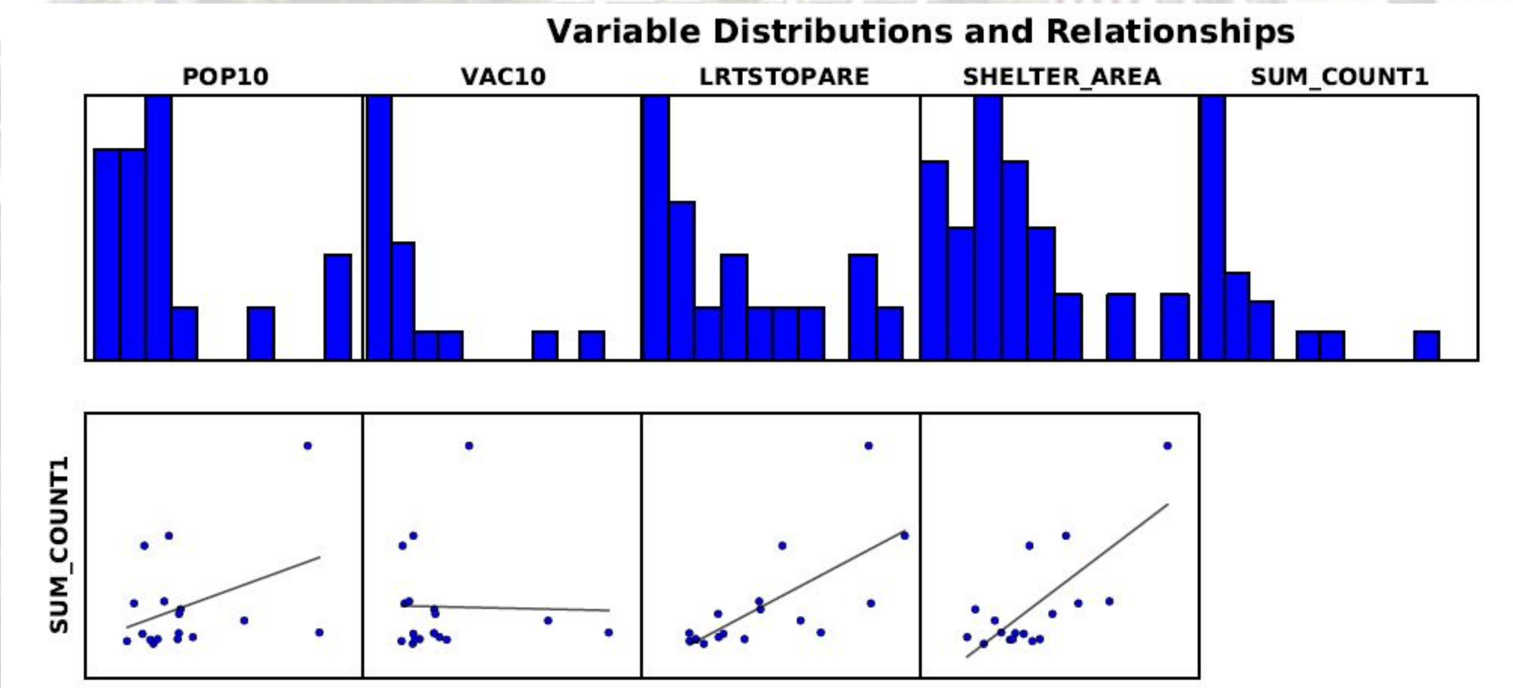
Methodology

- Selected by attribute on crime data and selected all crimes considered "violent" these included: Assault, Robbery, Rape, Homicide and Kidnap
- Focused on data only in Portland Metro Area
- To examine if crime clusters were influenced by variables. Applied buffer to parks with a 1/8 mile radius, applied buffer to transit stations with 300 ft radius, and applied buffer to homeless shelters with 300ft radius
- Joined area of buffer with block group data to use as explanatory variable in regression analysis
- Tabulate intersection to calculate area of the block groups that is within the buffer radius
- In order to look at spatial trends and clustering we ran multiple geostatistical tools including Cluster and Outlier(Anselin Local Moran's I) and Hot Spot Analysis (Getis-Ord Gi*)
- Ran Ordinary Least Squares (OLS) to observe the influence of the explanatory variables on the dependent variable (violent crimes per block group)

Results

Ordinary Least Square Diagnostics

Dependent Variable	Violent Crimes per Block group
Akaike's Information Criterion (AICc) [d]	187.12
Adjusted R-Squared [d]	0.758



Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Also note that the dependent variable, sum_count1 represents the count of violent crimes per block group.

Summary of OLS Results - Model Variables

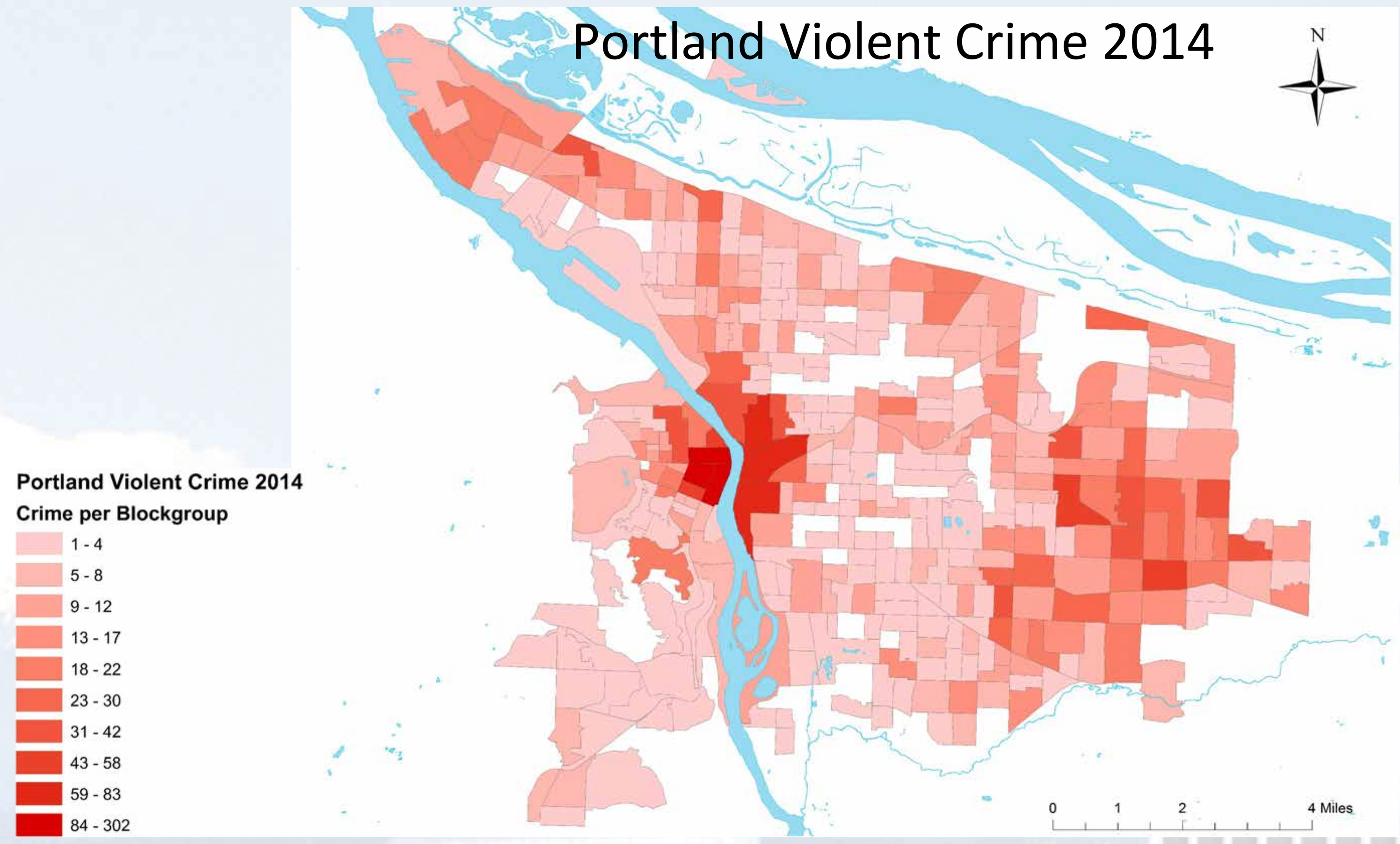
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-68.116360	23.578944	-2.888864	0.013605*	19.759731	-3.447231	0.004835*	--------
POP10	0.088146	0.029994	2.938790	0.012401*	0.023371	3.771649	0.002670*	6.143188
VAC10	-0.432618	0.152805	-2.831178	0.015142*	0.108830	-3.975173	0.001848*	6.313846
LRTSTOPARE	0.000049	0.000017	2.880671	0.013814*	0.000013	3.847609	0.002326*	1.700159
SHELTER_AREA	0.000061	0.000068	0.891275	0.390296	0.000040	1.529366	0.152105	2.377684

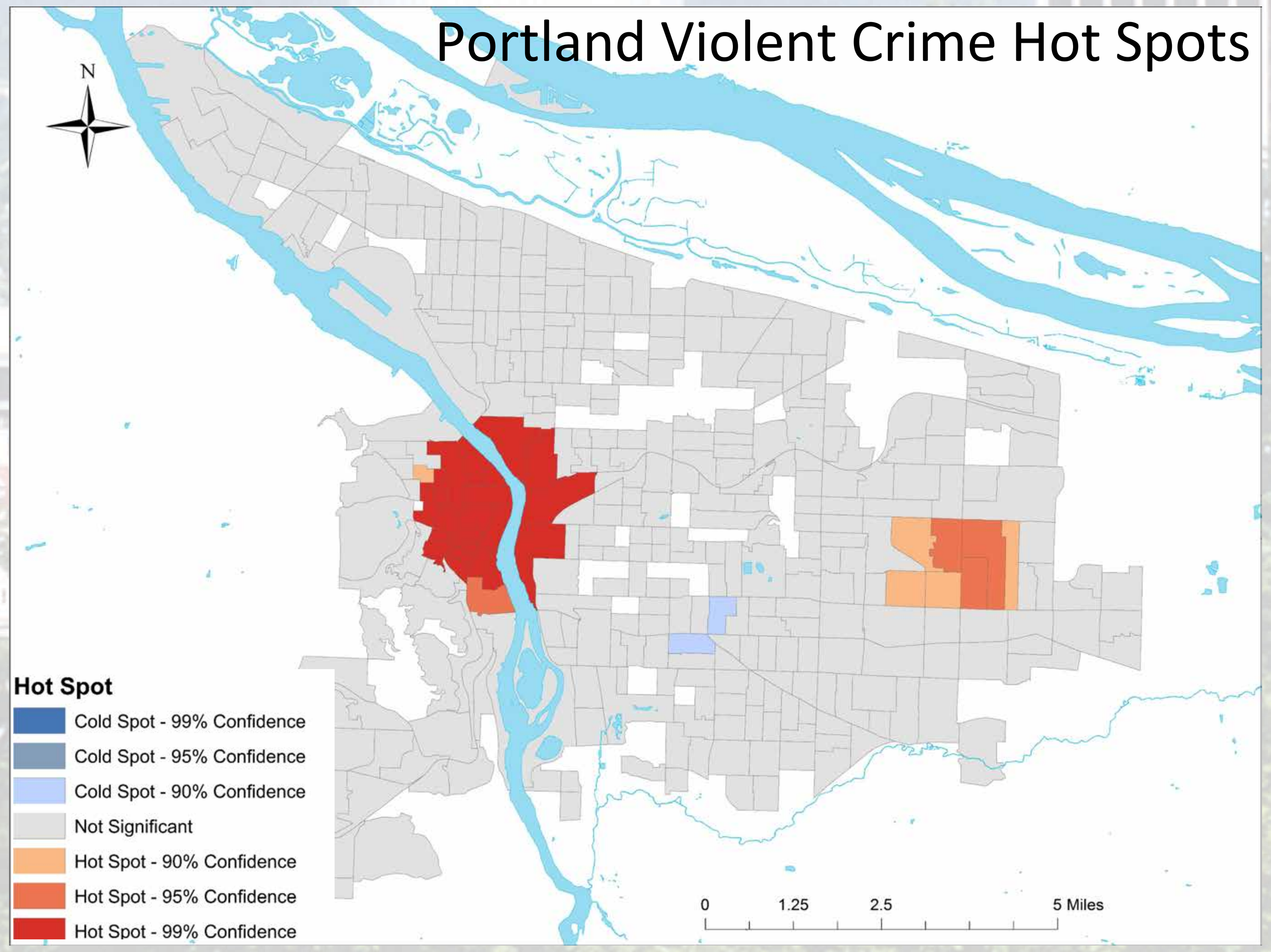
Data sources

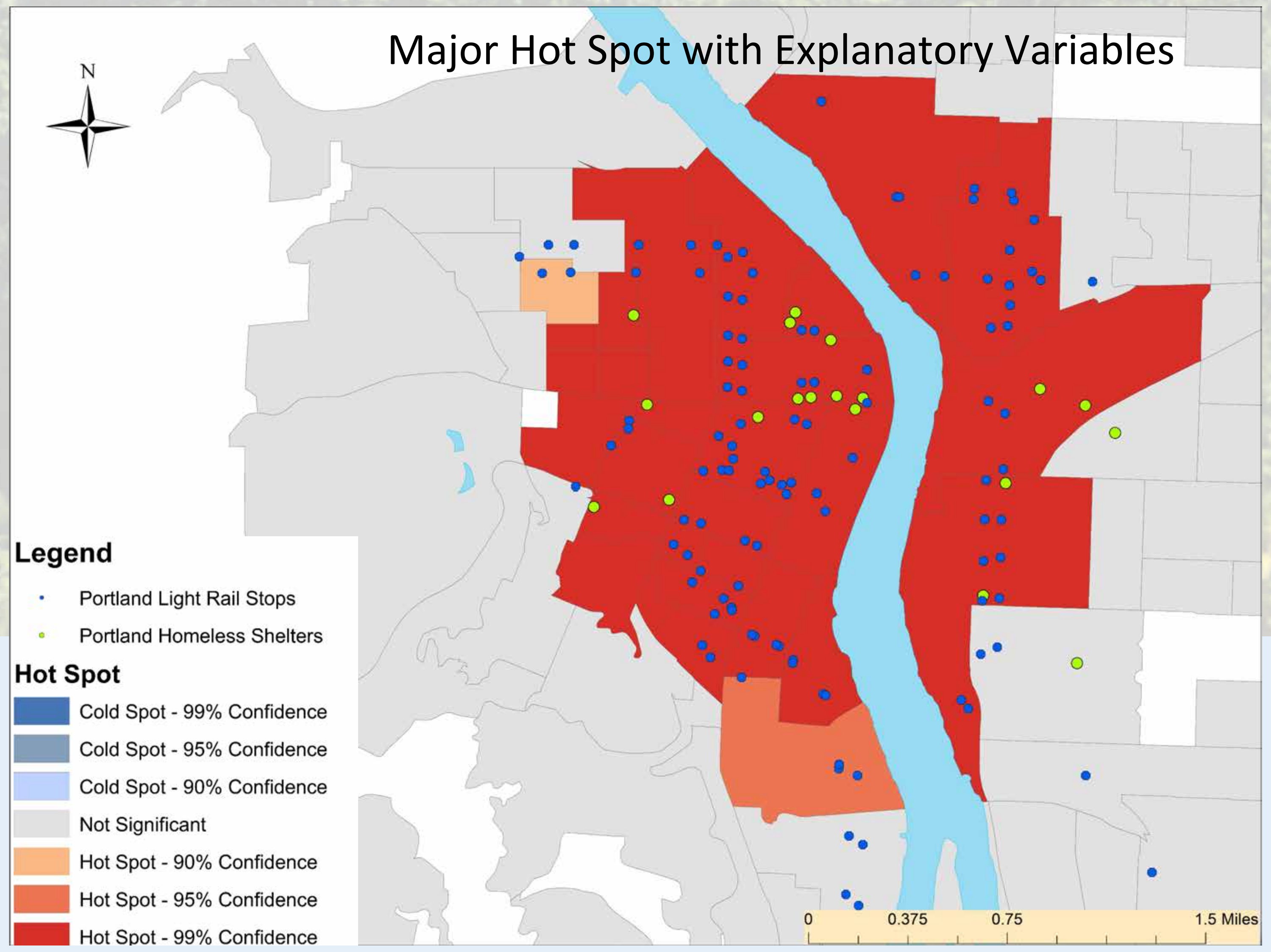
Oregon Metro (RLIS Discovery)
- 2010 Census Block Groups
- 2010 Tax lots
- Transit Centers
- Portland Parks
- River fill

ESRI
- Portland Homeless Shelters

Civic Apps
- 2014 Crime Data







Regression Analysis

Regression analysis allows the user to explore and model spatial relationships. It aids in understanding how factors influence observed spatial patterns. Data does not need to be normally distributed in order to run regression analysis.
In this study, the spatial pattern that is being analyzed is the crime hot spots in Portland. The analysis takes place on the block group level. After running the Ordinary Least Squares regression analysis, the tool produces various result pages that identify how well the chosen explanatory variables illustrate the spatial pattern. Important values in this result include the AIC, which should be as low as possible, the R-squared value, which is ideally as close to one as possible, and the probability scores for each of the variables.

Discussion

Four explanatory variables, population, vacant residencies, nearby light rail stops, and nearby homeless shelters, were identified as having significant correlation with the incidence of violent crimes in Portland, Oregon after the conduction of regression analysis. Using these explanatory variables yielded the lowest AIC score and highest R-squared value out of the variables that were tested. Several other explanatory variables were used in this regression analysis before this model was chosen. Previous explanatory variables included nearby parks, land zoned for industrial use, and land zoned for commercial use. These three variables did not aid in improving the model so they were removed.
When examining the data, it is apparent that the regression analysis identified three of the four explanatory variable as being statistically significant. While the variable that is statistically insignificant, in this case homeless shelters, is usually removed, it was found that removal of this variable drastically increased the AIC score as well as lowered the R-squared value. This means that removal of this variable reduced the effectiveness of the model. Further regression analysis of this topic could benefit from increasing the number of explanatory variables that were tested. Examples of useful explanatory variables for further study could include: average income per block group, average education level per block group, and incidence of other crime types per block group.
The continued use of regression analysis when addressing the occurrence of crime will most likely be beneficial to law enforcement and city planners. By identifying variables that explain crime incidence, both law enforcement and city planners will have some idea of the issues that need to be addressed in certain locations. This type of data analysis will be beneficial when allocating resources across the city.