

Atlanta Crime Report Notes:

Intro:

Atlanta is home to some of the largest fortune 500 company headquarters in the nation and is home to approximately 489,000 people (according to the 2019 ACS). The purpose of this study is to evaluate crime data from the city of Atlanta and explore any potential spatial relationships that may exist. The primary objectives of this study are to determine whether there is any correlation between *location* and violent crimes like robbery and rape, to determine whether there are any factors that increase the likelihood of criminality and to roughly determine the safest and most dangerous regions to live within the Atlanta Police Department's jurisdiction.

1. OLS Robbery:

a. First performed in ArcGIS

i. Analysis of Robbery based on the following variables:

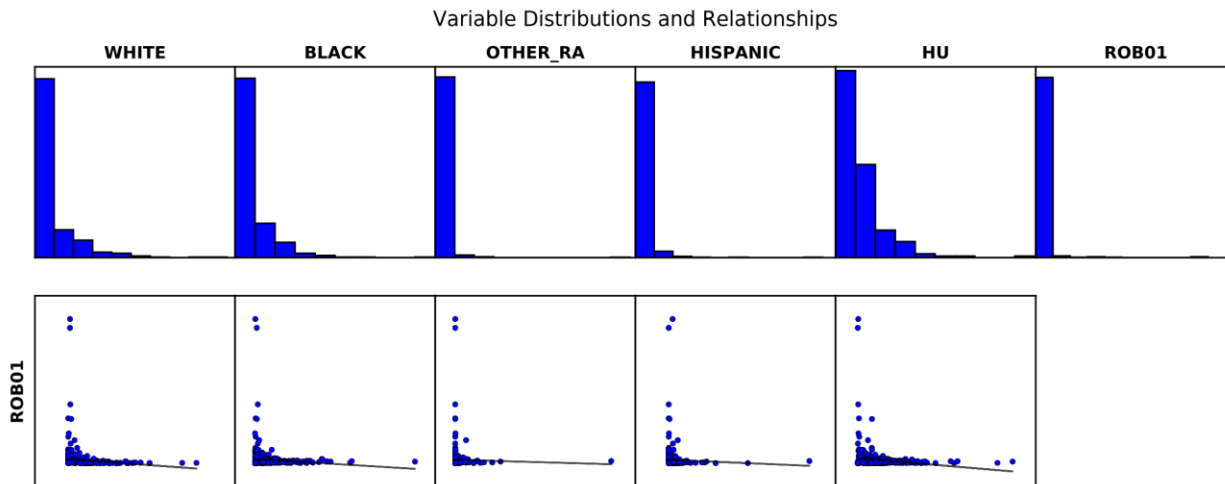
1. White
2. Black
3. Hispanic
4. Other Races
5. Housing Units
6. Vacant Housing Units
7. Median Family Income
8. Median Home Value
9. 1 Adult Households (%)
10. 1 Parent Households (%)
11. Fulltime Employment (%)
12. HS Education (%)
13. Poverty (%)
14. Unemployment (%)
15. Ages 15-24 (%)

ii. Initial Findings (Dep_var: Robbery):

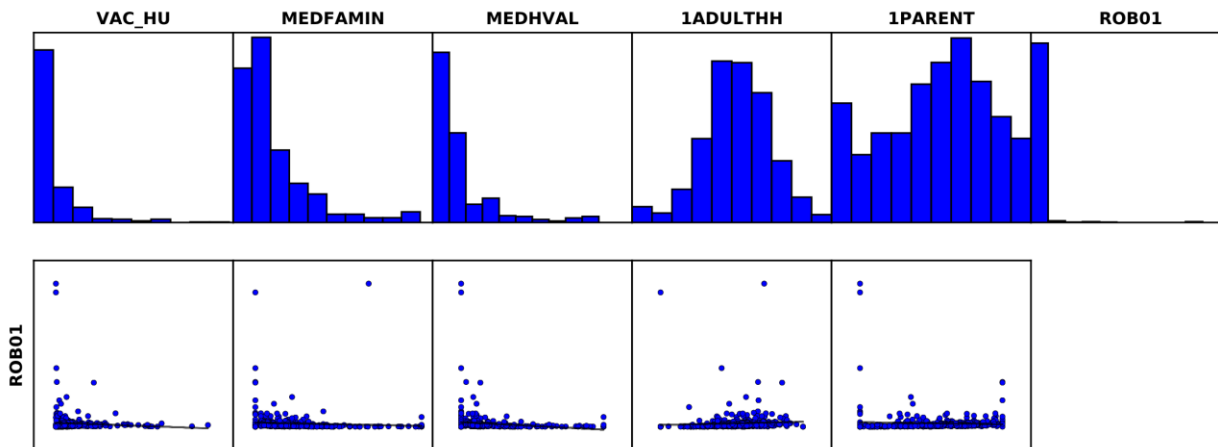
1. Adjusted R-Squared: 0.095
2. Jarque-Bera prob = 0.00 (significant which indicates model bias)
3. Koenker (BP) prob = 0.00 (significant which indicates model bias and heteroskedasticity)
4. AIC: 5102.73
5. Robust P:
 - a. White – Insig.
 - i. Significant P; insignificant robust P
 - b. Black – Insig.
 - i. Significant P; insignificant robust P
 - c. Hispanic – Insig.
 - d. Other Races – Insig.
 - e. Housing Units – Insig.
 - f. Vacant housing units – insig.
 - g. Median Family Income – insig.
 - i. Significant P; insignificant robust P
 - h. Median Home Value – insig.

- i. Significant P; insignificant robust P
- i. 1 Adult Households – insig.
- j. 1 Parent Households – insig.
 - i. Significant P; nearly significant robust P
- k. Fulltime Employment – Insig.
- l. HS Education – insig.
- m. Poverty – Sig.**
- n. Unemployment – insig.
- o. Ages 15-24 – insig.
 - i. Significant P; insignificant robust P

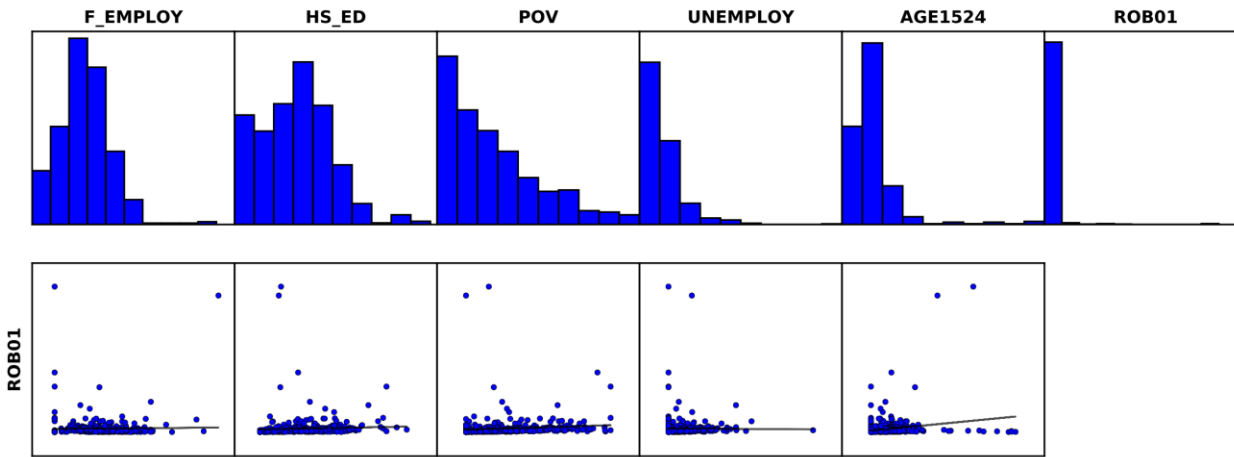
iii. Histograms:



Based on the negative skew for all of the above variables, there seems to be an negative relationship between all races and the crime of robbery.



There is a negative relationship between the number of vacant housing units in a census block and the number of robberies that occur. This indicates that the more vacant houses there are, the more robberies occur. There is a negative relationship between median family income and robbery as well as a negative relationship between median home value and robbery. These two findings suggest that robberies take place with higher frequency in lower-income areas. The relationship between robbery and 1 adult household suggests that association between the variables is random for the most part. The relationship between 1 parent households tells a slightly different story. The histogram indicates that at the far end, areas that have many single parents have higher rates of robbery. The middle part of the distribution is more or less normal which indicates that while this factor may be worth considering, it is not a concrete determining factor. This is a preliminary marker for analyzing a demographic with an elevated risk for turning to crime.



The negatively skewed distribution of those with fulltime employment compared to the number of robberies committed in a block indicates a slight positive correlation but is not technically significant. This indicates that there is a slightly elevated chance for crime in an area if there are low numbers of people with fulltime employment. This finding, however, is not significant to consider. The relationship between high school education attainment and the robbery rate is not significant enough to consider. The distribution has a negative skew which indicates slightly elevated risk of being robbed in an area with high numbers of high school dropouts. While this finding is not statistically significant, it should be taken under consideration as a risk factor. There is a statistically significant relationship between poverty and robbery. The highly negative skew of the histogram indicates that the higher the poverty rate in an area, the more likely it is to be robbed. This should be the primary indicator for propensity towards criminality. The highly negative skew towards unemployment tells a similar story, however, the finding was not technically significant so it should not be considered as a factor directly responsible for criminality in an area so much as it should be considered a risk factor – areas with high unemployment are subject to a potentially higher robbery rate. The relationship between age and robbery indicates that the majority of robberies are perpetrated by subjects in their late teens. This makes sense because hormones and development have a great deal to do with a person’s propensity towards violence. Young men are most aggressive up till their early 20’s. If they manage to avoid criminal behavior until their 20’s, the likelihood of turning towards crime drops off very sharply. While there is no directly significant finding here, the interpretation can be made that men in their teens (15-19) have a higher risk of criminal behavior – chiefly, robbery.

2. OLS Rape:

a. First performed in ArcGIS

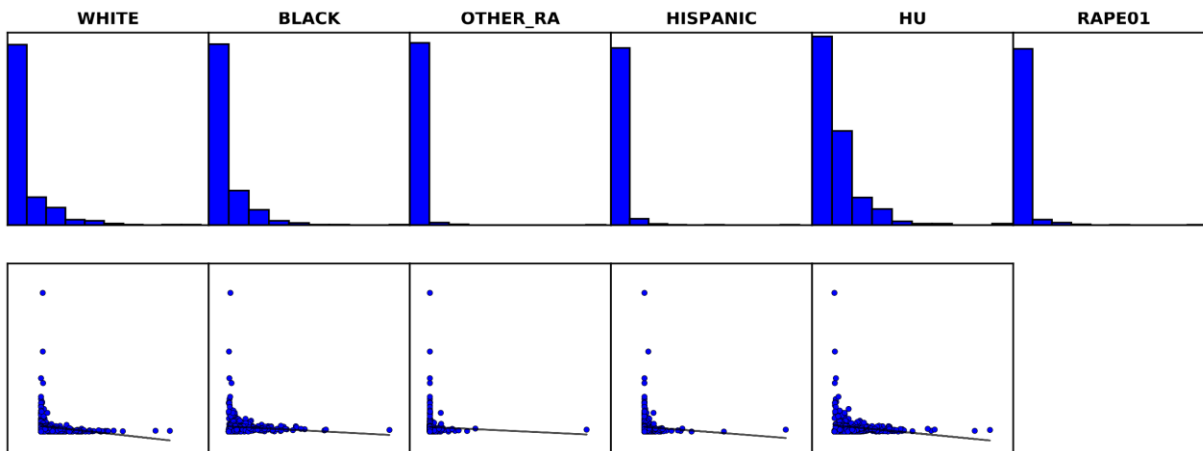
i. Analysis of Rape based on the following variables:

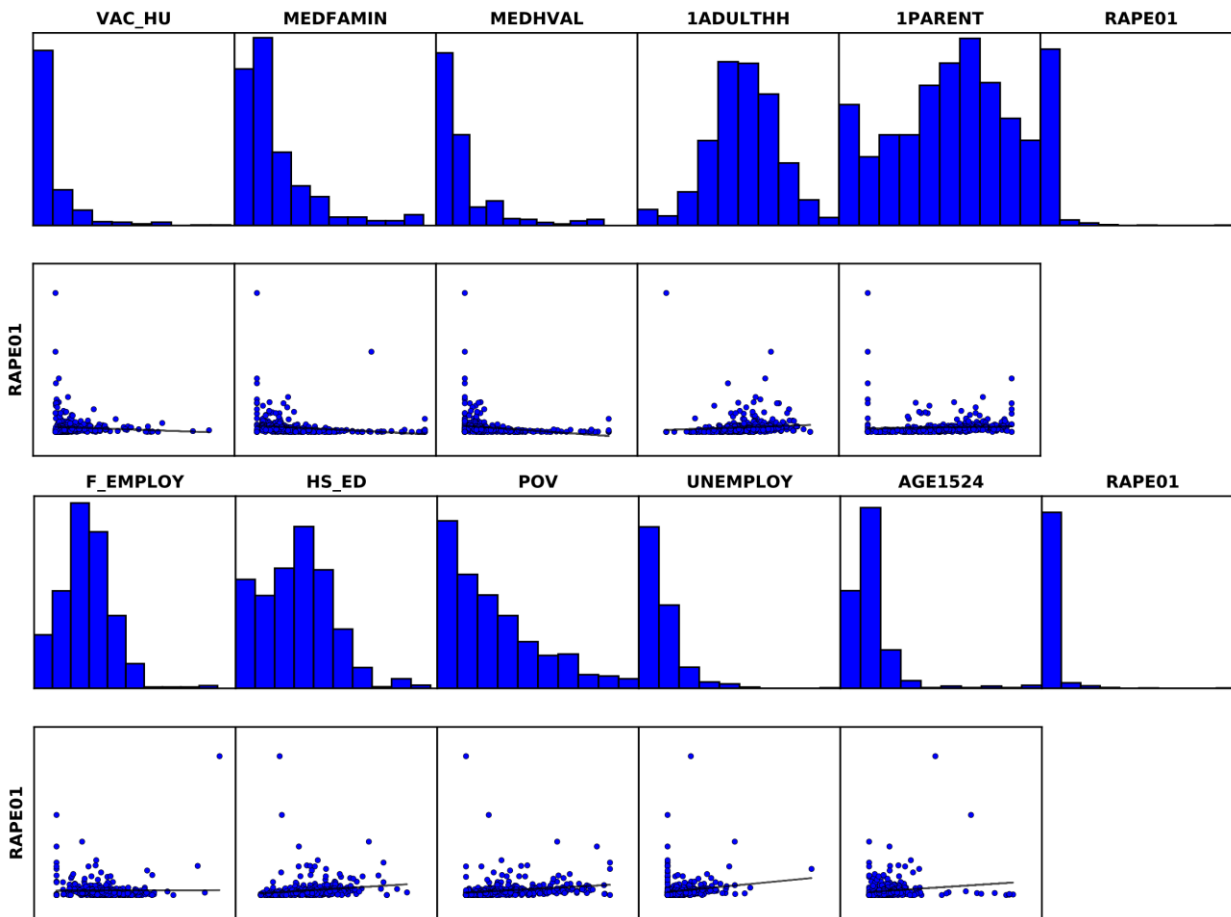
1. White
2. Black
3. Hispanic
4. Other Races
5. Housing Units
6. Vacant Housing Units
7. Median Family Income
8. Median Home Value
9. 1 Adult Households (%)
10. 1 Parent Households (%)
11. Fulltime Employment (%)
12. HS Education (%)
13. Poverty (%)
14. Unemployment (%)
15. Ages 15-24 (%)

ii. Initial Findings (Dep_var: Rape):

1. Adjusted R-Squared: 0.095
2. Jarque-Bera prob = 0.00 (significant which indicates model bias)
3. Koenker (BP) prob = 0.00 (significant which indicates model bias and heteroskedasticity)
4. AIC: 5102.73
5. Robust P:
 - a. White – Insig.
 - i. Significant P; insignificant robust P
 - b. Black – Sig.**
 - i. Significant P; significant robust P
 - c. Hispanic – Insig.
 - d. Other Races – Insig.
 - e. Housing Units – Insig.
 - f. Vacant housing units – insig.
 - g. Median Family Income – insig.
 - i. Insignificant P; insignificant robust P
 - h. Median Home Value – insig.
 - i. Significant P; insignificant robust P
 - i. 1 Adult Households – insig.
 - j. 1 Parent Households – insig.
 - i. Significant P; nearly significant robust P
 - k. Fulltime Employment – *Nearly sig.*
 - l. HS Education – Sig.**
 - m. Poverty – Sig.**
 - n. Unemployment – insig.
 - o. Ages 15-24 – insig.
 - i. Significant P; insignificant robust P

iii. Histograms:





b. Analysis using GeoDa

i. Initial Findings OLS Robbery:

1. **Median Family Income (\$)**
2. **Median Home Value (\$)**
3. 1 Adult Households (%)
4. **1 Parent Households (%)**
5. *Fulltime Employment (%)*
6. HS Education (%)
7. **Poverty (%)**
8. *Unemployment (%)*
9. **Ages 15-24 (%)**
10. **Population 1990 (cnt)**

* Bold indicates a significant finding. Italics indicates a nearly significant result.

The OLS in GeoDa returned a similar finding to the OLS test conducted in ArcGIS. The reported [adjusted] $R^2=0.114$ with an AIC=5096, and a significant level of heteroskedasticity. This finding essentially corroborates what was found with the initial OLS test. Upon conducting the OLS in GeoDa, I determined that the spatial lag model would be a better fit. Below are the relevant findings:

3. Spatial Lag Model:

c. Initial Findings (Dep_var: Robbery):

- i. Adjusted R-Squared: 0.3707
- ii. Breusch-Pagan test prob = 0.00 (significant which indicates model bias and heteroskedasticity)
- iii. AIC: 4995
- iv. P-value:
 1. **Median Family Income (\$)**
 2. **Median Home Value (\$)**
 3. 1 Adult Households (%)
 4. **1 Parent Households (%)**
 5. **Fulltime Employment (%)**
 6. **HS Education (%)**
 7. **Poverty (%)**
 8. Unemployment (%)
 9. **Ages 15-24 (%)**
 10. **Population 1990 (cnt)**

Similarly, to the OLS tests, the spatial lag model returned findings that indicate the highest risk factors are low MFI, low MHV, high numbers of single parent households, low numbers of fulltime employment, high poverty rate, low rates of high school completion, and high numbers of teens in an area – these are all risk factors that increase the likelihood of crime in a given census block. This model ended up being the best fitting and determined that certain variables that were nearly significant in the OLS to be significant using this test method.

v. Initial Findings OLS Rape:

1. Adjusted R-Squared: 0.129
2. BP test prob = 0.00 (significant which indicates model bias and heteroskedasticity)
3. AIC: 2990
4. Variables with high VIF subsequently removed.
5. P-value:
 - a. *White – Insig.*
 - b. **Black – Sig.**
 - c. *Hispanic – Insig.*
 - d. *Other Races – Insig.*
 - e. *Housing Units – Insig.*
 - f. *Vacant housing units – insig.*
 - g. *Median Family Income – insig.*
 - h. *Median Home Value – insig.*
 - i. *1 Adult Households – insig.*
 - j. *1 Parent Households – insig.*
 - k. *Fulltime Employment – Nearly sig.*
 - l. **HS Education – Sig.**
 - m. **Poverty – Sig.**
 - n. *Unemployment – insig.*
 - o. *Ages 15-24 – insig.*

Variables with high VIF associated with them were removed from subsequent models. Upon conducting the OLS in GeoDa – similarly to what resulted for test with the robbery variable, it was found that the spatial lag model would be a better fit.

- d. Findings Lag Model (Dep_var: Rape):
 - i. Adjusted R-Squared: 0.318
 - ii. Breusch-Pagan test prob = 0.00 (significant which indicates model bias and heteroskedasticity)
 - iii. AIC: 2920
 - iv. P-value:
 - 1. **Median Family Income (\$)**
 - 2. **Median Home Value (\$)**
 - 3. 1 Adult Households (%)
 - 4. **1 Parent Households (%)**
 - 5. **Fulltime Employment (%)**
 - 6. **HS Education (%)**
 - 7. **Poverty (%)**
 - 8. Unemployment (%)
 - 9. **Ages 15-24 (%)**
 - 10. **Population 1990 (cnt)**

Much like the spatial lag model run for the robbery variable, this was the best fitting model to evaluate rape. The risk factors that increase the danger level of an area seem to apply to multiple different types of crimes based on the analysis I conducted.

4. GWR:

- a. Robbery:
 - i. $R^2=0.243$
 - ii. AIC=5044
 - iii. *Significant heteroskedasticity*

The model was run to account for any non-stationarity that might occur but ultimately, the spatial lag model was a better fit.

- b. Rape:
 - i. $R^2=0.153$
 - ii. AIC=2989
 - iii. *Significant heteroskedasticity*

The GWR for the rape variable had a similar result to the test run for the robbery variable in that it was not as good a fit as the spatial lag model for the same variable.

Gettis-Ord was run to identify initial crime hot spots as well as to identify the most impoverished and areas with the lowest employment rates. Below are the associated figures.

Getis-Ord Gi Hot Spot Analysis - Robbery

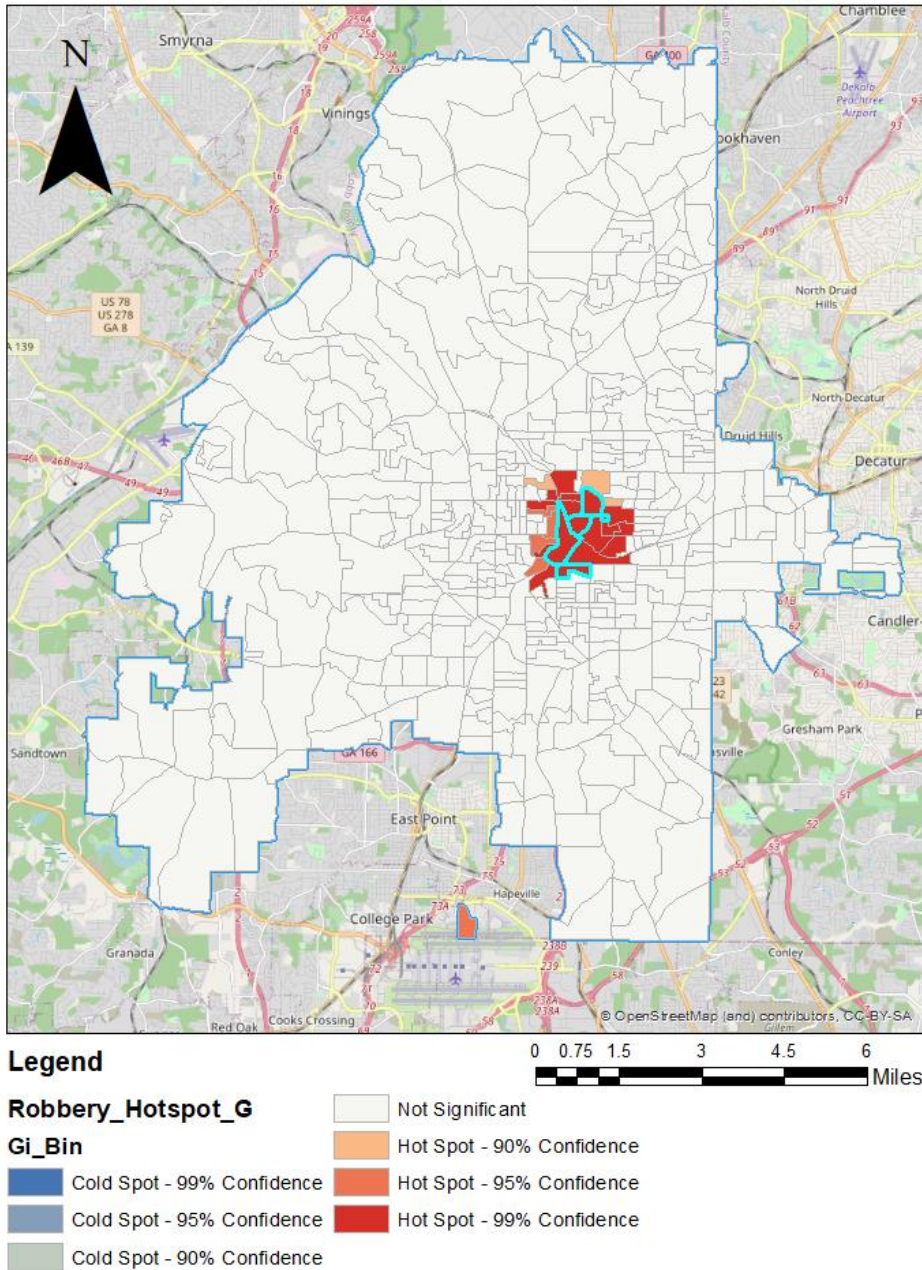


Figure 1. Hotspot analysis: Robbery

The same area ended up being a hotspot for both rape and robbery. The highest concentration of both robberies and rapes appears to occur in the heart of the city – roughly spanning the area from spaghetti Junction to the South, Georgia State to the East and Centennial Olympic Park to the Northwest.

Getis-Ord Hot Spot Analysis - Robbery

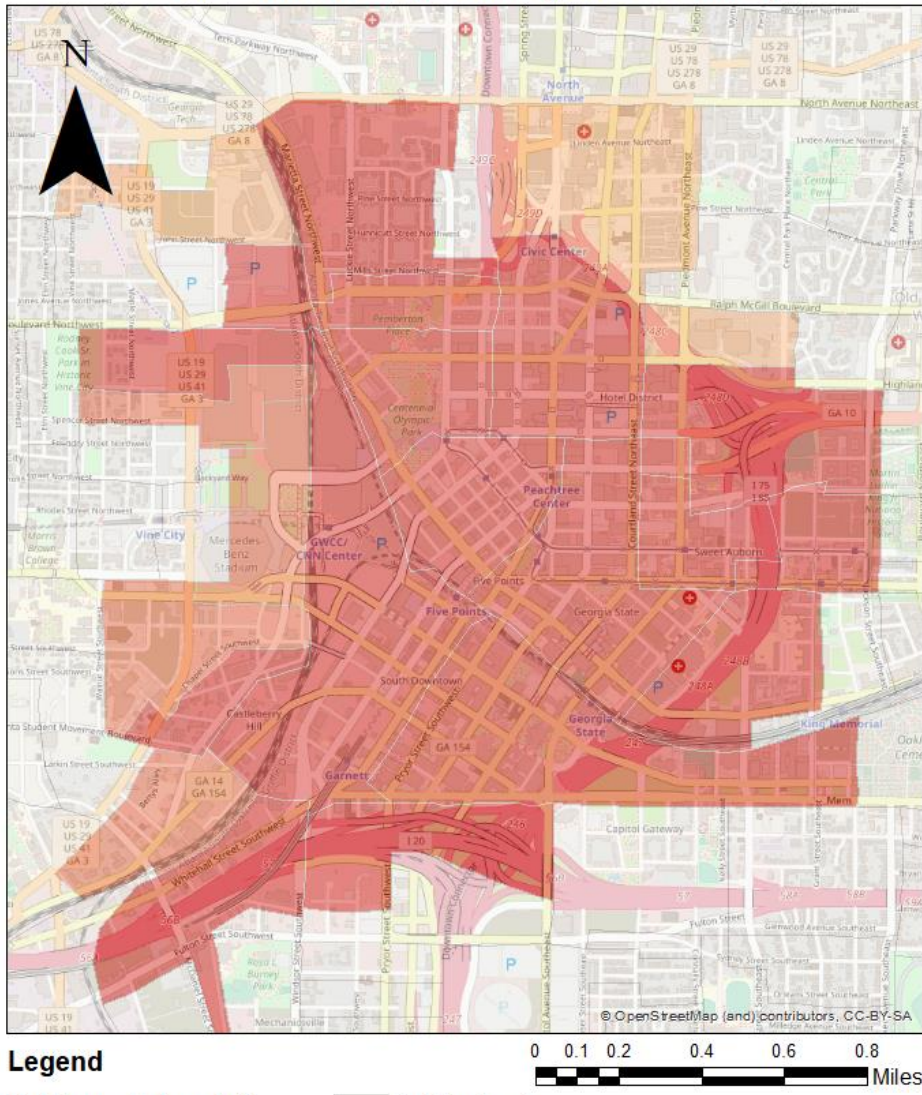


Figure 2. Hotspot analysis: Robbery

OLS Robbery

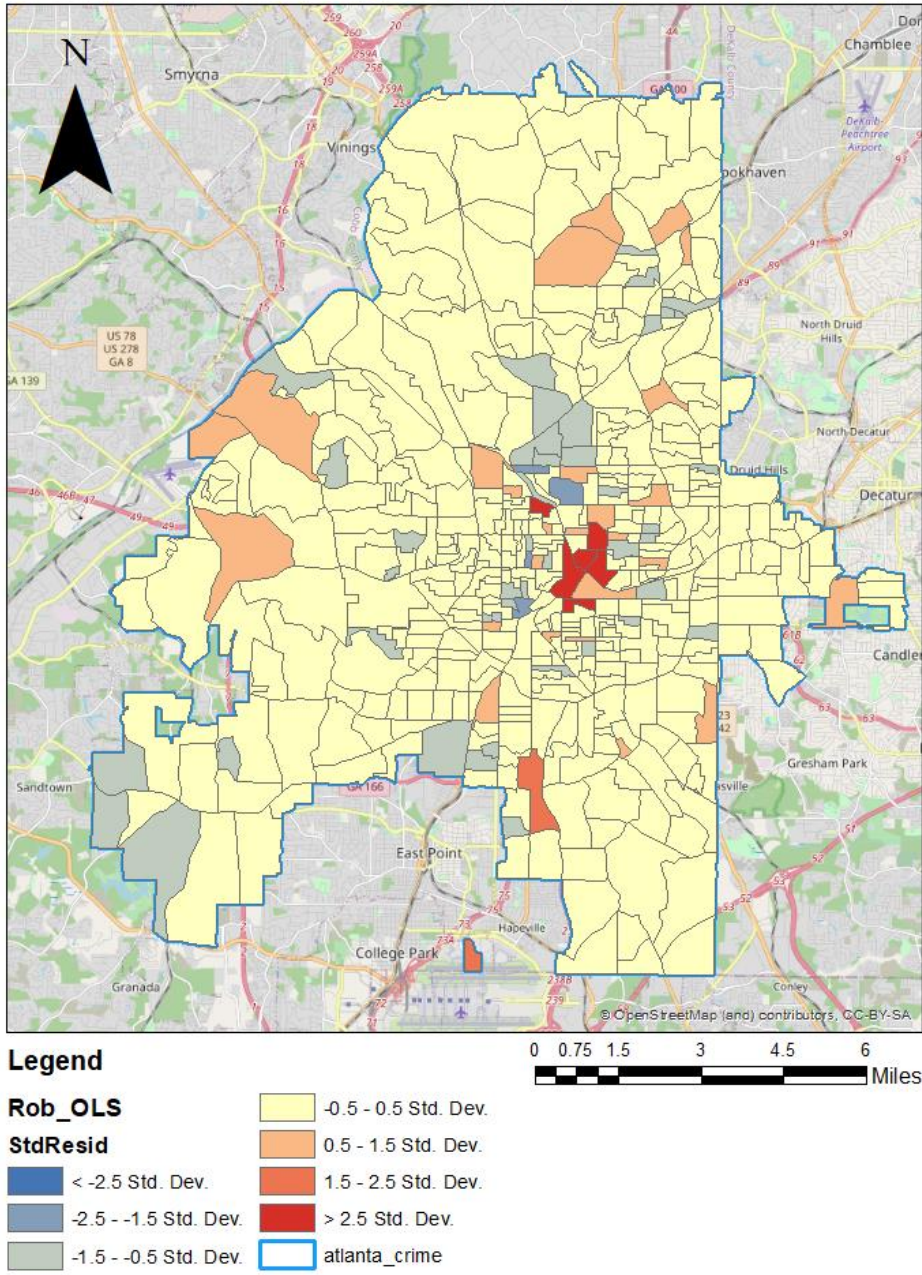
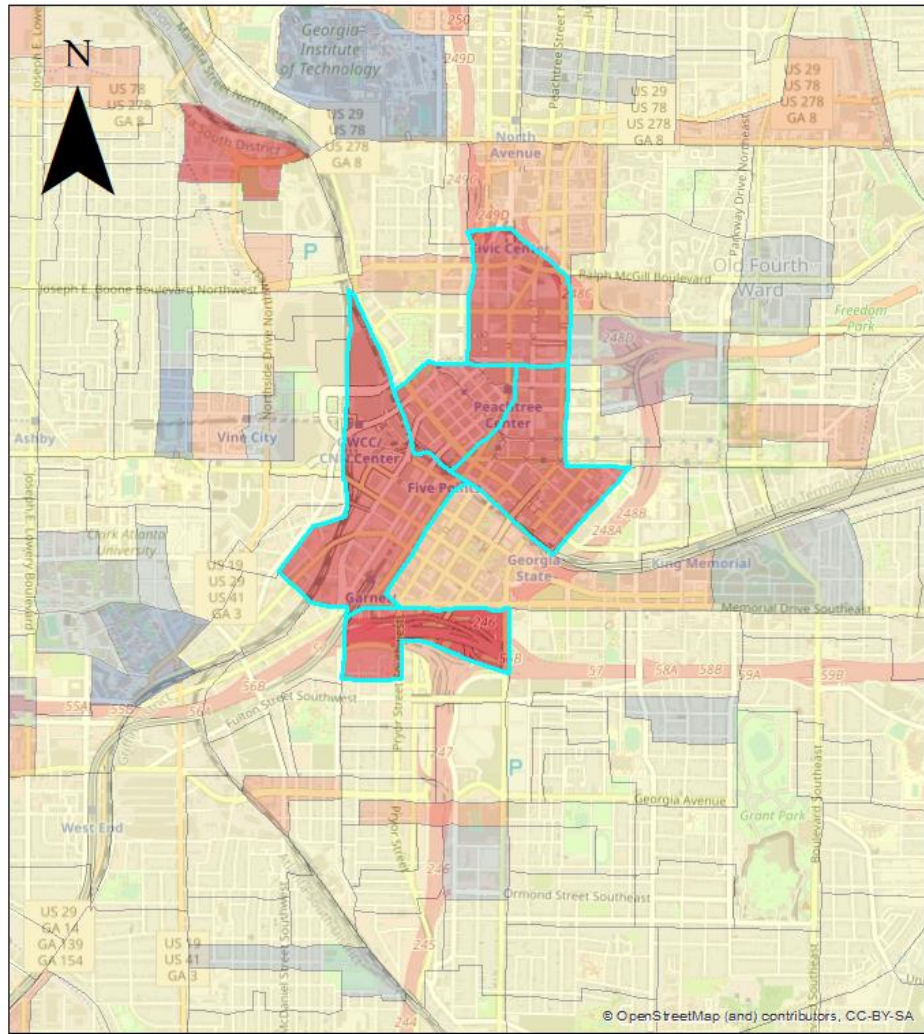


Figure 3. OLS: Robbery

OLS - Robbery



Legend

Rob_OLS	 -0.5 - 0.5 Std. Dev.
StdResid	 0.5 - 1.5 Std. Dev.
 < -2.5 Std. Dev.	 1.5 - 2.5 Std. Dev.
 -2.5 - -1.5 Std. Dev.	 > 2.5 Std. Dev.
 -1.5 - -0.5 Std. Dev.	

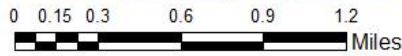


Figure 4. OLS: Robbery zoomed in

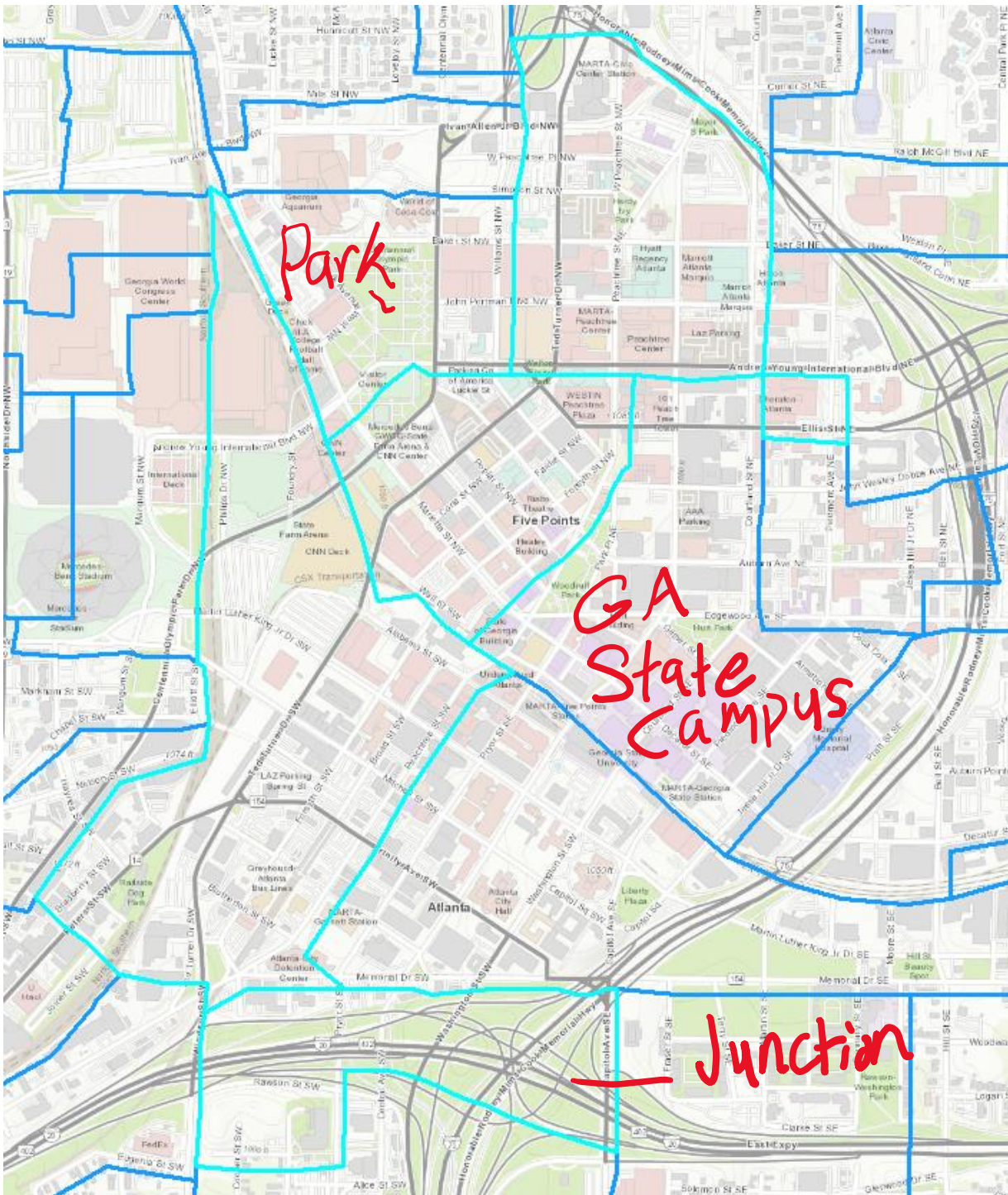


Figure 5. Annotated problem area.

Getis-Ord Hot Spot Analysis - Rape

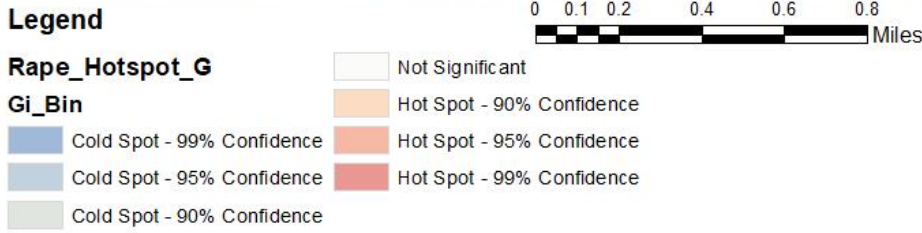
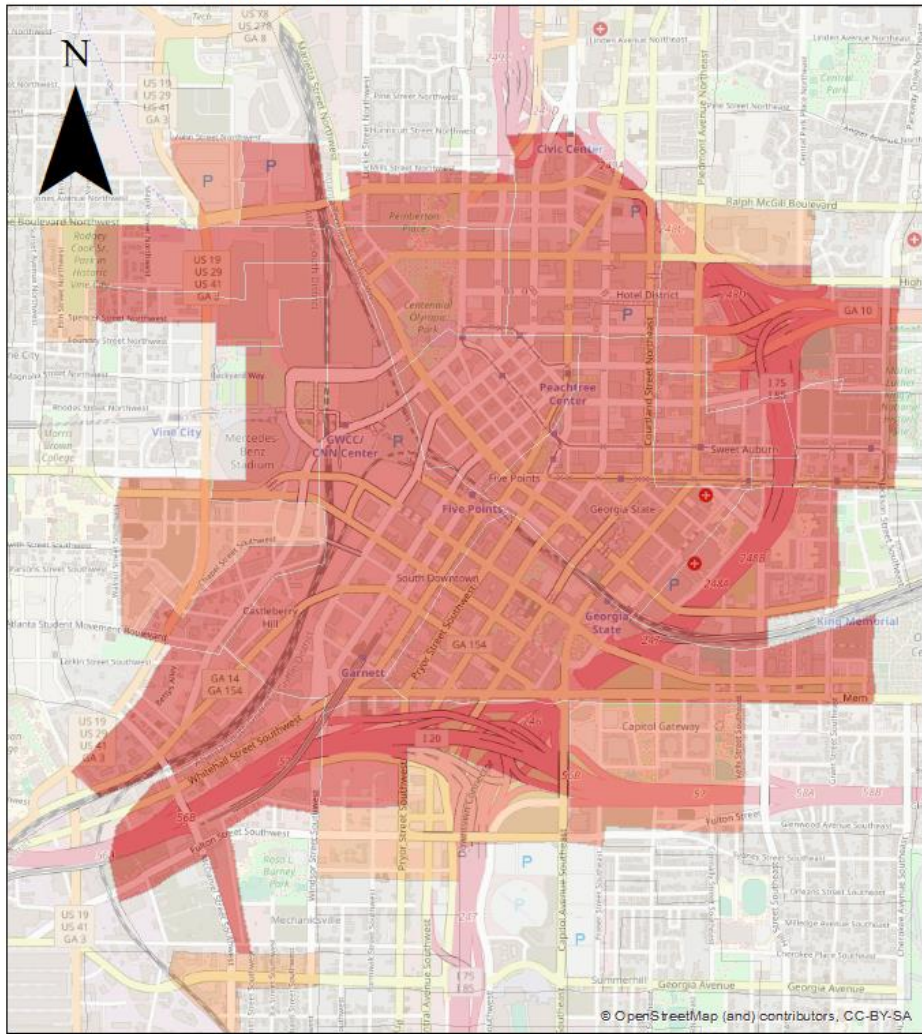


Figure 6. Hotspot Rape

OLS Rape

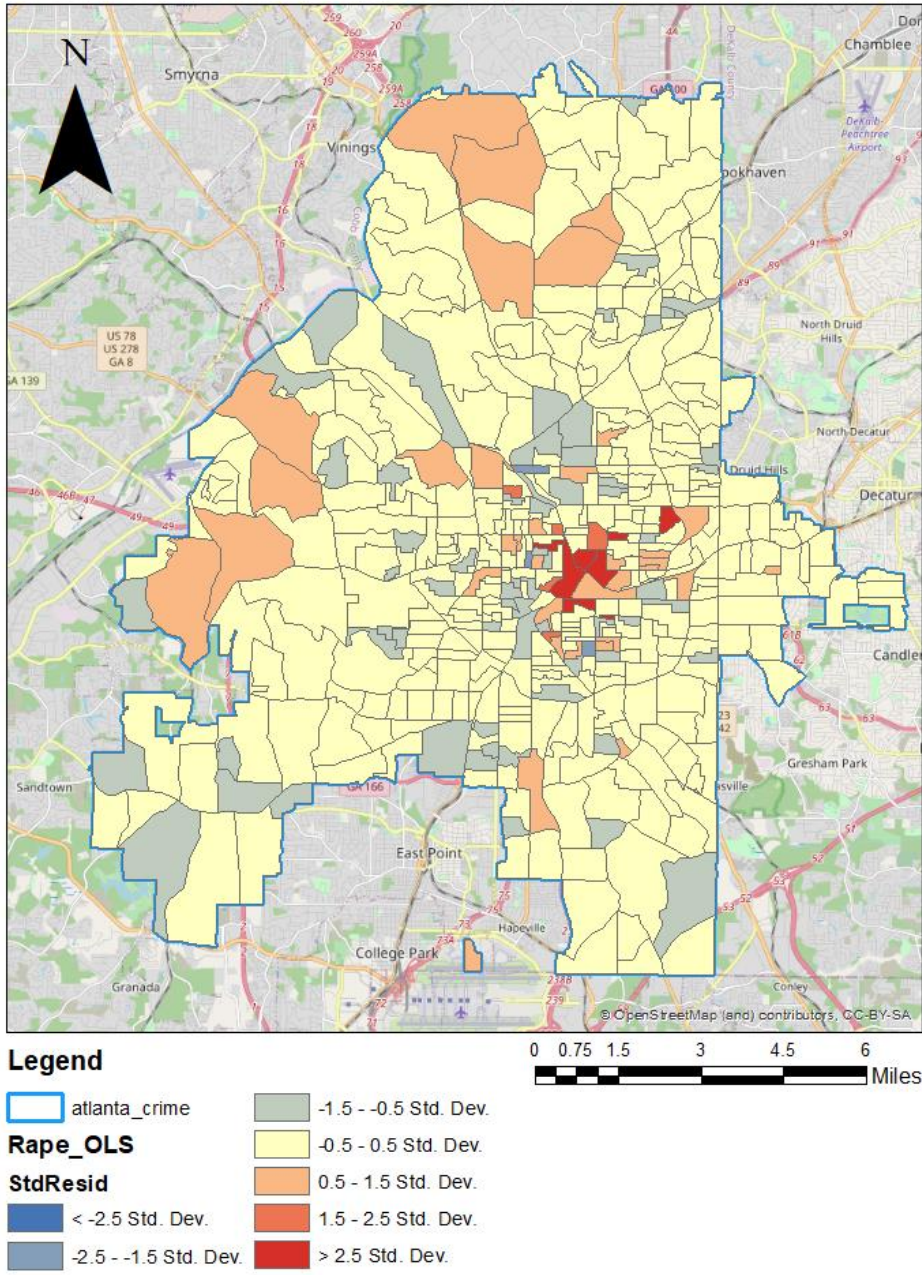
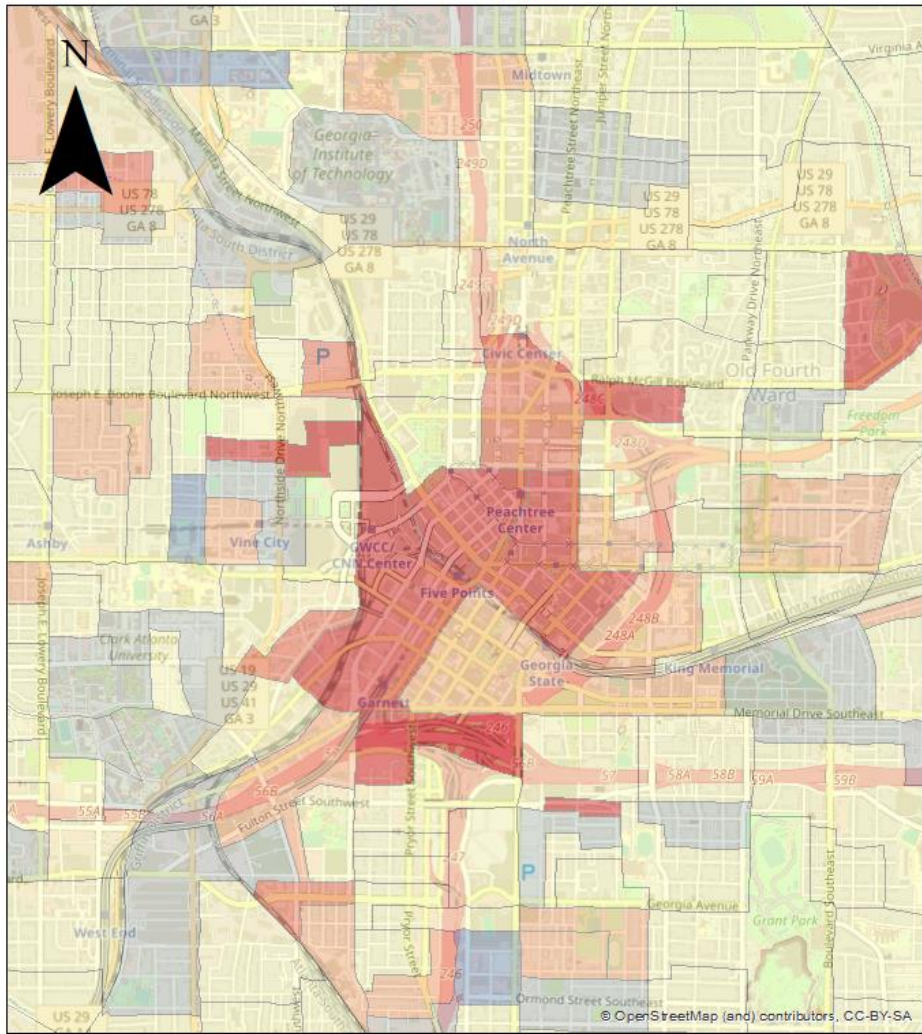


Figure 7. OLS: Rape

OLS - Rape



Legend

Rape_OLS	StdResid
Light yellow	-0.5 - 0.5 Std. Dev.
Orange	0.5 - 1.5 Std. Dev.
Dark orange	1.5 - 2.5 Std. Dev.
Red	> 2.5 Std. Dev.
Blue	< -2.5 Std. Dev.
Light blue	-2.5 - -1.5 Std. Dev.
Dark blue	-1.5 - -0.5 Std. Dev.

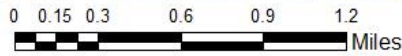
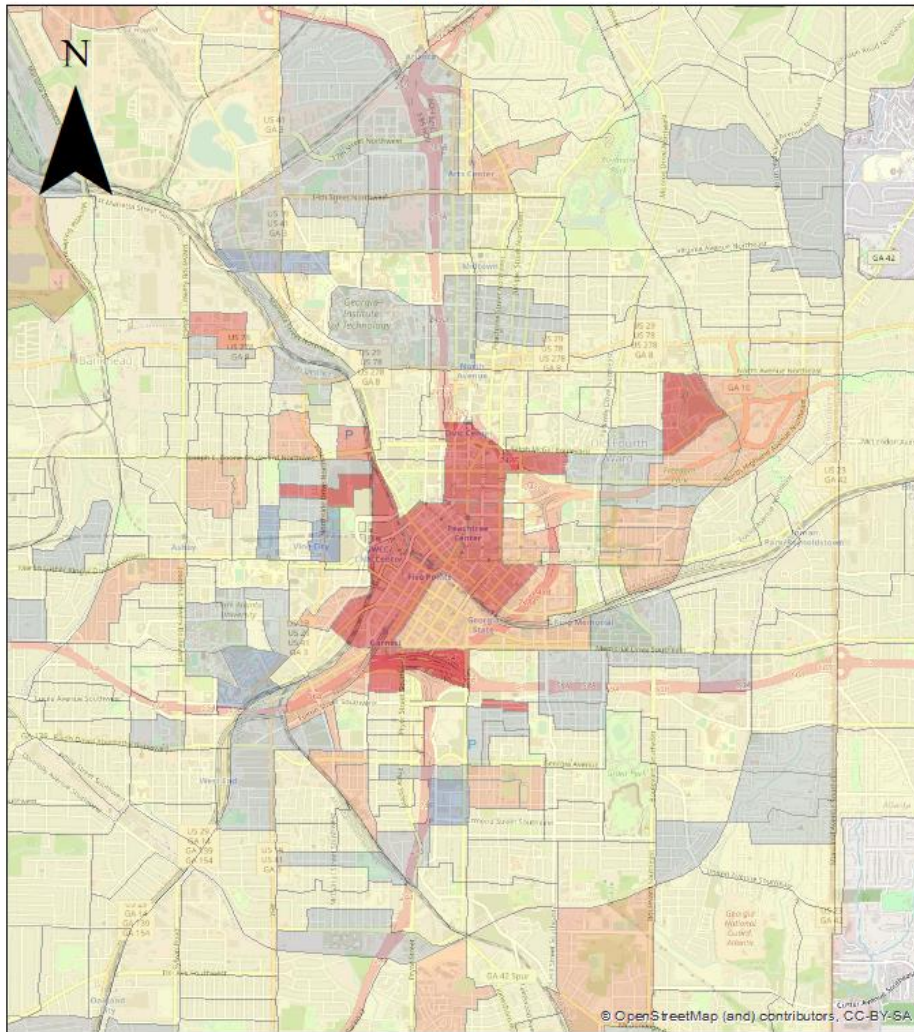


Figure 8. OLS: Rape zoomed in.

GWR - Rape



Legend

Rape_GWR	StdResid
Light Yellow	-0.5 - 0.5 Std. Dev.
Light Orange	0.5 - 1.5 Std. Dev.
Blue	< -2.5 Std. Dev.
Dark Blue	-2.5 - -1.5 Std. Dev.
Light Grey	-1.5 - -0.5 Std. Dev.
Dark Orange	1.5 - 2.5 Std. Dev.
Red	> 2.5 Std. Dev.

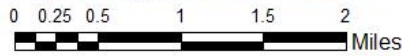


Figure 9. GWR: Rape

Getis-Ord Hot Spot Analysis - Poverty

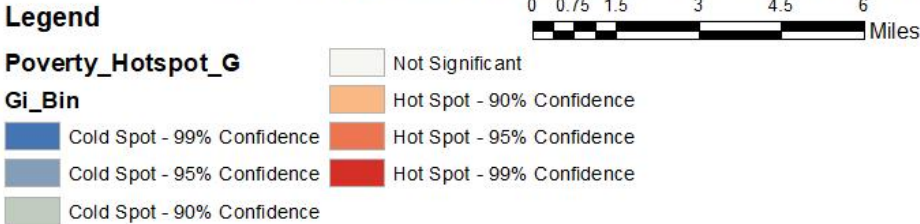
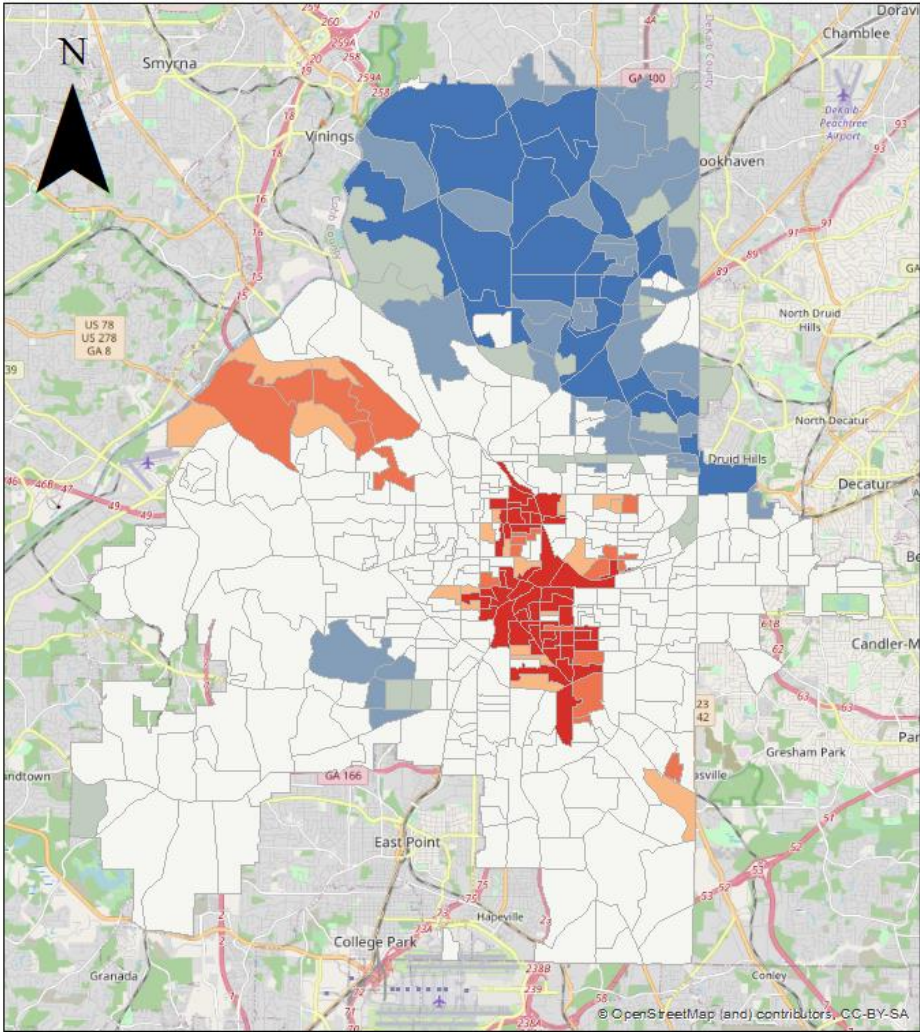


Figure 10. Hotspot: Poverty

Getis-Ord Hot Spot Analysis - Unemployment

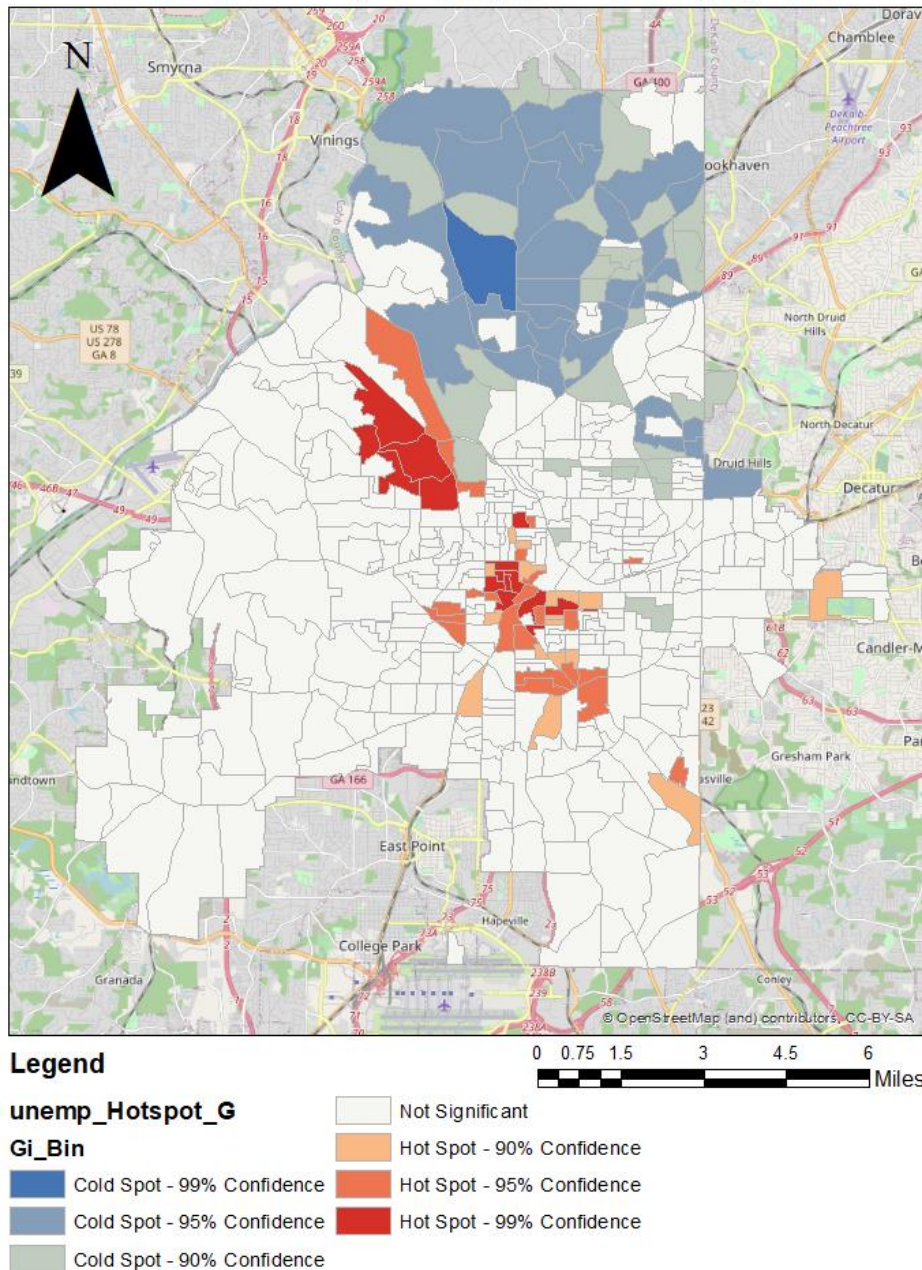


Figure 11. Hotspot: Unemployment

Conclusions:

Poverty is the primary driver for robbery. Low-income communities with high numbers of single parent households are at high risk for violent crime. Lack of education is the best indicator for predicting rape. The northern part of the city is the safest; the central portion is the most dangerous. The Spatial lag models were the best fits for both Robbery & Rape. Model heteroskedasticity indicates further analysis must be conducted.

Resources:

- NHGIS – American Community Survey
- Atlanta Police Department
- City-Data.com
- GeoDa
- ArcGIS