

June 12th, 2019

Shanglun Li

INVESTIGATE CHICAGO COMMUNITIES CRIME ON A SPATIAL ASPECT

ABSTRACT. This paper used spatial techniques to analyze the 2014 Chicago communities crime data. The paper focused on the relationship between crime rate and several variables representing income level, education level, and unemployment rate.

CONTENTS

1. Introduction	1
1.1. Background	1
1.2. Data Description	2
1.3. Research Questions	3
1.4. Variables of Interest	3
2. Exploratory Data Analysis (EDA)	4
3. Statistical and Spatial Analysis	8
4. Conclusion	14
Acknowledgments	14
References	15
Appendix A. Histograms of the Variables	17
Appendix B. Quantile Maps of the Variables	20
Appendix C. Scatterplot Matrix	24
Appendix D. Rook Weight	26
Appendix E. Moran's I Scatterplot	27
Appendix F. Cluster and Significant Map	29
Appendix G. Principle Component Analysis	31

1. INTRODUCTION

1.1. Background.

It is well-known that Chicago is not a safe city. The city's overall crime rate, especially the violent crime rate, is higher than the US average. Chicago was

responsible for nearly half of 2016's increase in homicides in the US, though the nation's crime rates remain near historic lows. This paper aims to analyze the crime rate of Chicago communities from a spatial aspect. It draws attention to some types of social tie that is particularly conducive to criminological behaviors: educational level, employment rate, and income rate. Neighborhood scholars have recently begun to note that "neighborhood effects" are in fact largely approximating the impact of spatially-bounded, local social networks (Hipp et al., 2012; Sampson, 2012). Specifically, social networks are thought to be a key intervening mechanism explaining the observed relationship between neighborhood features and local crime rates (Hipp and Boessen, 2013; Kubrin and Wo, 2016; Soller and Browning, 2014). Importantly, while social ties are typically construed in terms of inhibiting crime, such ties may also facilitate problematic behaviors (Browning et al., 2004). At a basic level, crime and deviance require information, skills, and logistical support, as well as deviant social norms that support such behavior. For example, social learning theory suggests that social ties—especially among peers—facilitate the acquisition of such resources (McGloin and Nguyen, 2013). Likewise, research on street gangs suggests that internal processes directly related to social ties among members—such as cohesion or loyalty—facilitate a range of group and individual deviance and criminal behavior (Hughes, 2013).

1.2. Data Description.

Three datasets were chosen to include in the project:

First of all, the crime data from Chicago Data Portal is selected (link: <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>). It reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

Secondly, the Public health and socio-economic indicators dataset is selected, which is imported from the Geoda Center, which is the Public health and socio-economic indicators for the 77 community areas of Chicago, IL, 2014. (link: https://geodacenter.github.io/data-and-lab//comarea_vars/).

Finally, we selected the Chicago community areas data set from Chicago Data

Portal. The dataset was imported from Chicago Data Portal. It is a boundary file for the community areas. (link: <https://data.cityofchicago.org/Facilities-GeographicBoundaries/Boundaries-Community-Areas-current-/cauq-8yn6>).

1.3. Research Questions.

We are targeting the following research question(s):

Q1: Is there a statistical relationship between crime rate versus education level, unemployment rate, and income level?

Q2: Is there a spatial correlation with respect to the variables of interest?

1.4. Variables of Interest.

After the datasets have been imported, they have been modified by aggregating and rearranging the variables; community ID, community name, number of incidents of crime per thousand capita in 2014 (calculated by the count of crime divided by population of community, then times one thousand), number of people living in poverty per thousand capita in 2014, number of people unemployed per thousand capita in 2014, number of adults without a high school diploma in 2014, and per capita income in 2014, which we named ComAreaID, community, CRIMP1000, POVP1000, UNEMPP1000, NOHSP1000, PerCIInc14, respectively (For more detailed description of variables, refer to the table below). According to our interest, the explanatory variables used in our analysis were no high school diploma rate, unemployment rate, income per capita, and poverty rate, and crime rate were used as the response variables. Table 1 provides the name and a brief description of each variable that were created, including the associated levels and the type and usage of the variables in the future analysis.

Variable Attributes

Variable	Description	Levels	Comments
ComAreaID	The ID number of the community in Chicago	No.1-77	ID variable
community	The community name in Chicago	text name	ID variable
CRIMP1000	Number of incidents of crime per thousand capita in 2014	[22, 366]	Response variable
POVP1000	Number of people living in poverty per thousand capita in 2014	[24, 628]	Explanatory variable

UNEMPP1000	Number of people unemployed per thousand capita in 2014	[28, 153]	Explanatory variable
NOHSP1000	Number of adults without a high school diploma in 2014	[13, 273]	Explanatory variable
PerCInc14	Per capita income in 2014	[14905, 2075859]	Explanatory variable

TABLE 1. The table includes the name, description, levels, comments for each variable. Explanatory variable, response variable, and variable type are noted in the comments.

2. EXPLORATORY DATA ANALYSIS (EDA)

The data was reviewed and modified prior to the spatial and statistical analysis followed by the procedure stated in the former section. There were no notable errors or missing values. We looked at the descriptive statistics of some variables including their histograms, shown in Table 2 (refer Appendix A for the histograms).

Descriptive Statistics for Variables with respect to the Whole Sample

Variable	N	Mean	Standard Deviation	Minimum	Maximum
CRIMP1000	77	109.811	75.4334	22.014	365.317
POVP1000	77	233.46	122.773	24.470	627.677
UMEMPP1000	77	74.5495	31.1279	28.6728	152.69
NOHSP1000	77	120.104	63.3869	13.2498	272.405
PerCInc14	77	289557	360033	14905	2075860

TABLE 2. The table displays some variables with respect to the whole sample that were meaningful to include. There were a total of 77 observations across the study. The table also displays additional information pertaining to the mean, standard deviation, maximum and minimum for each variable.

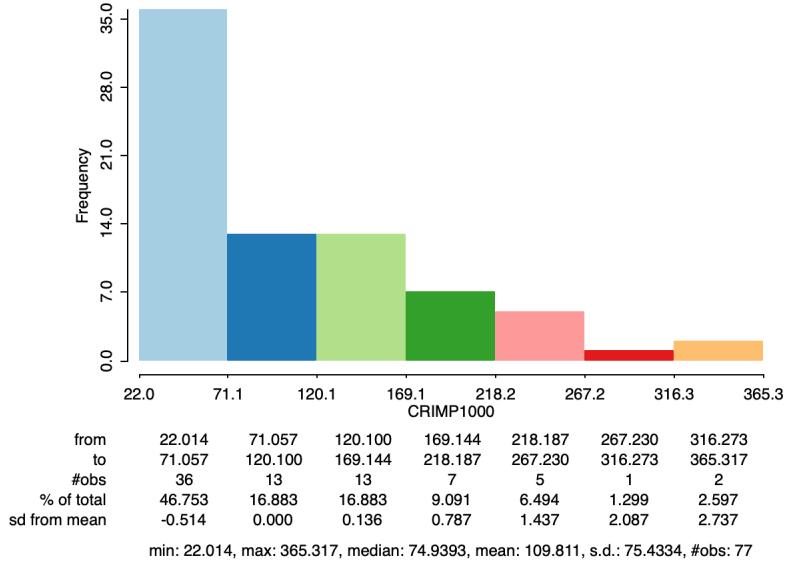


FIGURE 1. Histogram of Number of Crime happed per thousand capita in 2014.

As we can see from the table 2 and the histograms of crime rate in Figure 1 (refer to the other histograms in Appendix A), the distribution of all the five variables are right skewed, since there is a lower bound for all the variables, which is zero.

Then, some quantile maps were created in order to see if there exists some potential spatial patterns in the dataset (refer the maps in Appendix B). As we can see from the maps, people in South part of Chicago tend to have higher crime rates, higher unemployment rate, lower income, and lower education level, while people in North part of Chicago including Loop area tend to have lower crime rates, lower unemployment rate, higher income, and higher education level. In addition, as we can see from the crime histogram, most of the communities have a value less than 90.7. Then, we selected the bins that are higher than 90.7. As we can see from the equal intervals map, the distribution of the communities are in the South part of Chicago, downtown and the communities on its West as shown in Figure 2. Thus, there is potentially some spatial correlations to these variables that we can investigate later on.

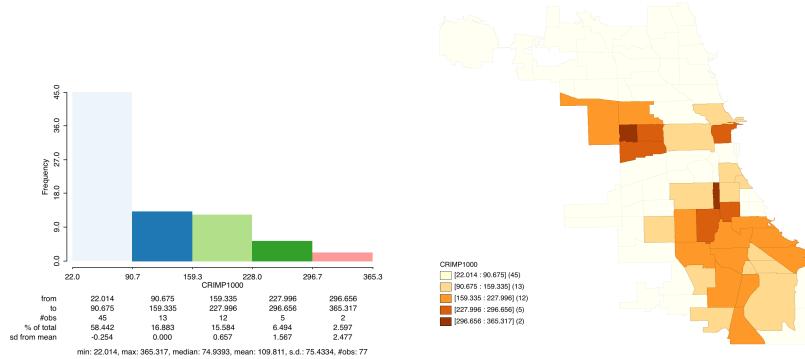


FIGURE 2. Histogram of Incidents of crime per thousand capita and the Corresponding Equal Intervals Map.

Then, we drew the scatterplot matrix among unemployment rate, high school education rate, income per capita, poverty rate, and crime rate (refer the scatterplot matrix in Appendix C). As we can see from the scatterplot matrix, crime rate has a positive relationship with unemployment rate and poverty rate, and a negative relationship with income level. And the LOWESS smoother does not imply any structural break with respect to crime versus other variables except crime versus income level. In addition, we can see that after we selected the higher income per capita bins in the income histogram, the communities are all located in the North part of Chicago. They are all with relatively low crime rate (lower than 154.6). In addition, the points shown in the scatterplot can be potential outliers.

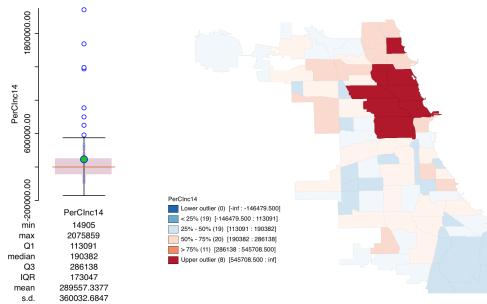


FIGURE 3. Boxplot of Income per capita and Income per capita Box Map.

To check the high income per capita communities, we draw the box plot of Income per capita and the box map of it with hinge = 1.5 as shown in Figure 3. As we can see from the box plot and box map, there is no lower outlier, but there are some upper outliers, which are exactly the communities to the North of downtown Chicago.

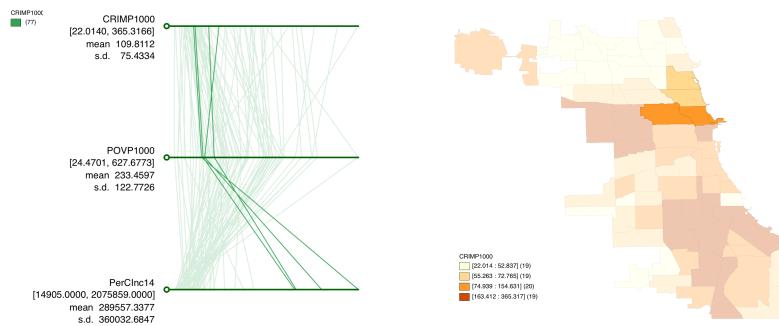


FIGURE 4. Parallel Coordinate Plot and Crime Choropleth Map.

Then, we plot the parallel coordinate plot as shown in Figure 4. We select the observations with highest income per capita. We find that it is the four communities located to the North of downtown Chicago, and it makes sense in the plot that high income per capita implies a low poverty rate and a low crime rate.

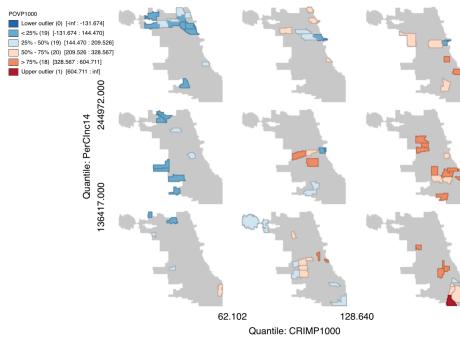


FIGURE 5. Conditional Map with Crime Rate x-Axis, Income per capita y-Axis, and Poverty Rate Base.

In addition, we draw the conditional map as shown in Figure 5, since income level and poverty rate are variables with same idea of representation of the population. The horizontal axis is the crime per thousand capita; the vertical axis is the income per capita; and the base box map is based on the people living in poverty per thousand capita. We can see from the map on the first row and first column that communities with higher income per capita and lower poverty rate tend to have a low crime rate. Similarly, as we can see from the maps on the second row and third column and on the third row and third column, communities with lower income per capita and higher poverty rate tend to have a higher crime rate.

Finally, we created the rook contiguity weight for the community polygon map (as shown in Appendix D). As we can see from the weight table and the histogram, the distribution of the number of neighbors of community is approximately normal.

3. STATISTICAL AND SPATIAL ANALYSIS

First, we specifically look at the scatterplot of crime per thousand capita versus the other four explanatory variables as shown in Figure 6. As we can see from the scatterplots, the p-value for the linear fit of the 77 data points is less than 0.05 for scatterplots of crime rate versus unemployment rate and crime rate versus poverty rate, which are significant. On the other hand, the scatterplots of crime rate versus no high school diploma rate and crime rate versus income level are not significant, which means crime rate is not significantly correlated with education level and income level. In addition, we found that if we select all the north communities, we can see that the Chow test is significant for all the scatterplots, which means the slopes we select and the slopes of the rest of the map are significantly different to each other, suggesting spatial heterogeneity between North part of Chicago and the rest. For the two significant scatterplots, we can see that the slopes are both positive, which implies that crime rate and unemployment rate are positively correlated, as well as crime rate and poverty rate. Thus, the higher poverty rate and unemployment rate will have a higher crime rate.

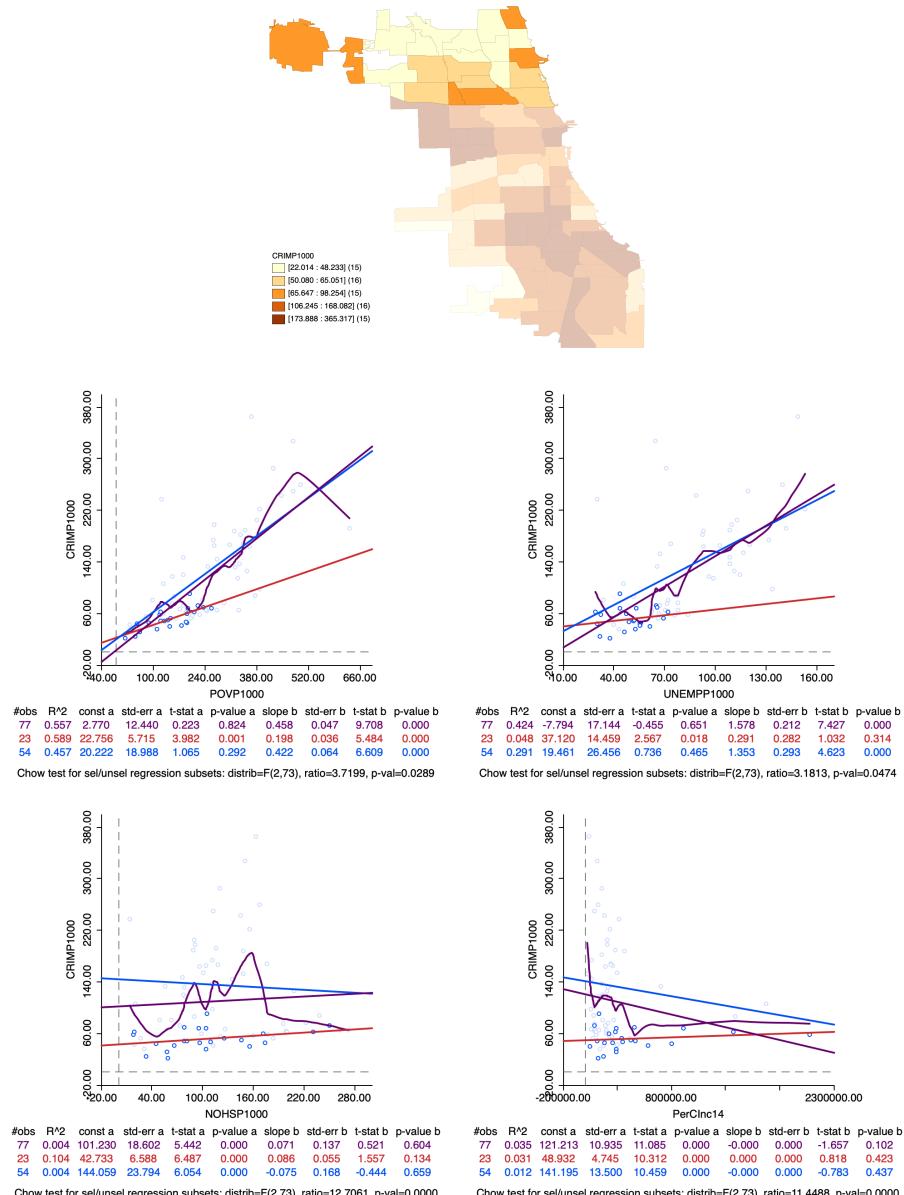


FIGURE 6. Scatterplot of Crime versus Variables of Interest and Crime Choropleth Map.

After fitted the scatterplots and did the summary, Table 3-6 shown below had been generated:

Statistical Summary for Poverty Rate

Parameter	Coefficient	Standard Error
Intercept	2.770	12.440
POVP1000	0.458	0.047

TABLE 3. The table displays the coefficient and the standard error for each variable.

Statistical Summary for Unemployment Rate

Parameter	Coefficient	Standard Error
Intercept	-7.794	17.144
UNEMPP1000	1.578	0.212

TABLE 4. The table displays the coefficient and the standard error for each variable.

Statistical Summary for No High School Diploma Rate

Parameter	Coefficient	Standard Error
Intercept	101.230	18.602
POVP1000	0.071	0.137

TABLE 5. The table displays the coefficient and the standard error for each variable.

Statistical Summary for Income Level

Parameter	Coefficient	Standard Error
Intercept	121.213	10.935
POVP1000	-0.000	0.000

TABLE 6. The table displays the coefficient and the standard error for each variable.

Then, we draw the Moran's I scatterplots of the five variables including crime rate, no high school diploma rate, income level, unemployment rate, and poverty rate with respect to Rook weight as shown in Figure 7 and 8 for crime rate (refer to the rest in Appendix E). Among the five Moran scatterplot, the slopes of all the

scatterplots are significant, since the p-value is less than 0.05, which implies that for all the five variables, there exists significant spatial correlation for the variables, since from the simulation, we can see that the p-values are all 0.001. In addition, we can see from the Moran Scatterplots that there is a potential structural break in the map between north and the rest, since the p-value for Chow test is less than 0.05 for all Moran scatterplots after we select the north part of Chicago.

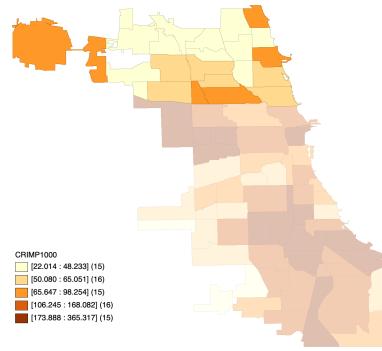


FIGURE 7. Crime Choropleth Map with North Part Selected.

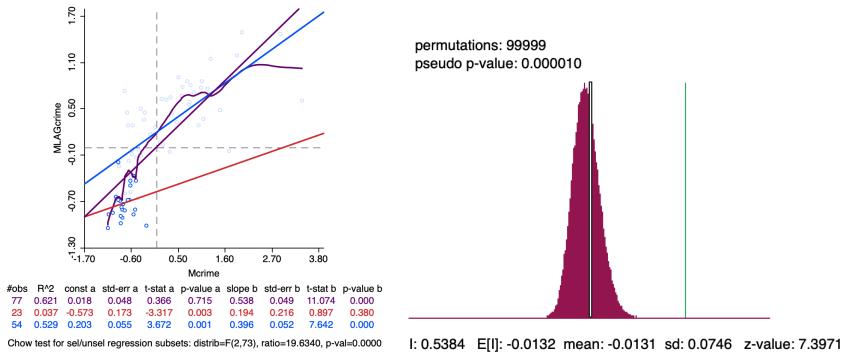


FIGURE 8. The Moran Scatterplot of Crime Rate with respect to Rook Weight.

Next, we drew the cluster map and significance map of the five variables with respect to Rook weight. We chose the p-value cutoff to be 0.05, and the number of permutation to be 999 as shown in Figure 9 for crime rate (refer to the rest in

Appendix F). As we can see from the maps, for crime rate, the communities on the North seem to have a lower crime rate cluster as shown in blue color to be the core of the cluster. In addition, the middle South and West communities seem to have a high crime rate cluster with red colors as the core of the cluster. Moreover, we can see that there is no low-high or high-low communities in the map, which implies that the communities' crime rates are positively correlated. For poverty rate, the communities on the North seem to have a lower poverty rate cluster as shown in blue color to be the core of the cluster. In addition, the middle South communities seem to have a high poverty rate cluster with red colors as the core of the cluster. Moreover, we can see that there is just two low-high or high-low communities in the map, which implies that the communities' poverty rates are mostly positively correlated. For no high school diploma rate, some north communities and two communities on south seem to have a lower no high school diploma rate cluster as shown in blue color to be the core of the cluster. In addition, the West communities seem to have a high no high school diploma rate cluster with red colors as the core of the cluster. Moreover, we can see that there is no low-high or high-low communities in the map, which implies that the communities' no high school diploma rates are positively correlated. For unemployment rate, the communities on the North seem to have a lower unemployment rate cluster as shown in blue color to be the core of the cluster. In addition, the South communities seem to have a high unemployment rate cluster with red colors as the core of the cluster. Moreover, we can see that there are only two low-high or high-low communities in the map, which implies that the communities' unemployment rates are positively correlated. Finally, for income level, the communities on the North seem to have a higher income level cluster as shown in red color to be the core of the cluster. In addition, the South and West Loop communities seem to have a lower income level cluster with blue colors as the core of the cluster. Moreover, we can see that there are only one low-high or high-low communities in the map, which implies that the communities' income level are positively correlated.

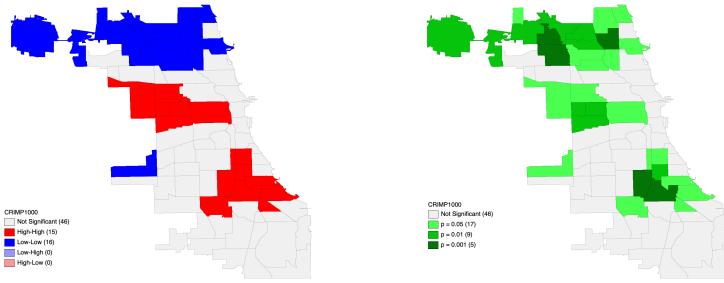


FIGURE 9. The Cluster Map and Significance Map of Crime Rate with respect to Rook Weight.

In order to find spatial clusters, we also tried Principle Component Analysis on the five variables. I set the transformation option to be Standardize. As we can see from the output, the 5 variables are reduced to 3 dimensions based on 95% threshold criterion, i.e. the first three principle explain 90.56% of the total variance of the original variables. According to the squared correlations table (as shown in Appendix G), the first principle component (PC1) is mainly made up of the 3 variables including crime rate, poverty rate, and unemployment rate. The second principle component (PC2) is mainly explained by three variables including no high school diploma rate, crime rate, and income level. The third component (PC3) is explained mainly by the no high school diploma rate and income level.

Then, we use the max-p regionalization with respect rook contiguity weight. In order to have not too many cluster, we set the minimum bound to be 17% of the 2014 population so that the number of clusters is appeared to be 5 as shown in Figure 10. We set the iteration number to 1000. As we can see from the cluster map, south, west, loop, and north have been clustered roughly, which is identical to our prior analysis.

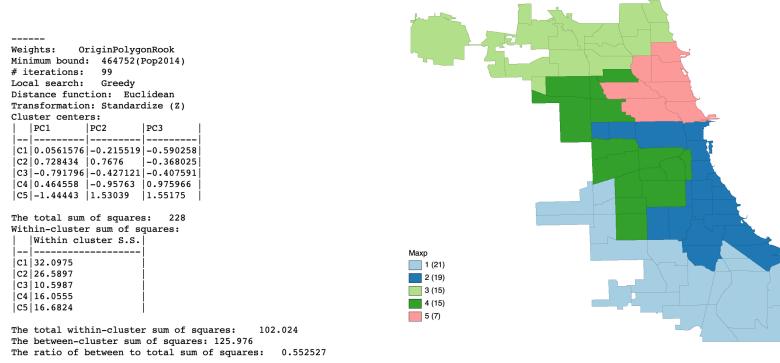


FIGURE 10. Max-p cluster map and characteristics (Minimum population bound = 17%)

4. CONCLUSION

Statistically speaking, poverty rate and unemployment rate are statistically related to crime rate, since from the analysis in section 3, we found that the p-values for the slope of the scatterplot with unemployment rate versus crime rate and that of the scatterplot with poverty rate versus crime rate are both significant. In addition, they are all positively correlated, which means a higher unemployment rate and higher poverty rate will make the crime rate higher.

Spatially, there is a spatial correlation with respect to our variables of interest. Since as we can see from the Moran's I scatterplots, the p-values are all less than 0.05 for all the five variables. In addition, there is a structural break between north part of Chicago and the rest. Moreover, the spatial clusters are clustered roughly by north, south, loop and west. Also, we can see that the neighbors of Chicago are influencing each other positively, which also implies similarity among neighbor communities.

Acknowledgments. It is a pleasure to thank my mentor, Professor Benjamin Soltoff, for providing helpful suggestions about constructing model during the process of finishing this project.

REFERENCES

- [1] Chicago Driving Uptick in Murders; National Crime Rate Stays Near 'Historic Lows'. (n.d.). Retrieved from <https://www.usnews.com/news/articles/2016-09-19/chicago-drives-uptick-in-murders-national-crime-rate-stays-near-historic-lows>
- [2] Friedman, Matthew, et al. Crime Trends, 1990-2016. Brennan Center for Justice at New York University School of Law, 2017.
- [3] Graif, C., Lunguanu, A., & Yetter, A. M. (2017). Neighborhood isolation in Chicago: Violent crime effects on structural isolation and homophily in inter-neighborhood commuting networks. *Social Networks*, 51, 40-59. doi:10.1016/j.socnet.2017.01.007
- [4] Hipp, John R., Butts, Carter T., Acton, Ryan, Nagle, Nicholas N., Boessen, Adam, 2013. Extrapolative simulation of neighborhood networks based on population spatial distribution: do they predict crime? *Social Networks* 35 (4), 614–615.
- [5] Hughes, Lorine A., 2013. Group cohesiveness, gang member prestige, and delinquency and violence in Chicago, 1959–1962. *Criminology* 51 (4), 795–832.
- [6] Kubrin, Charis E., Wo, James, 2016. Social disorganization theory's greatest challenge: linking structural characteristics to crime in socially disorganized neighborhoods. In: Piquero, Alex R. (Ed.), *Handbook of Criminological Theory*. Wiley-Blackwell, Oxford, pp. 121–136.
- [7] Marotta, P. (2016). Assessing Spatial Relationships Between Rates of Crime and Rates of Gonorrhea and Chlamydia in Chicago, 2012. *Journal of Urban Health*, 94(2), 276-288. doi:10.1007/s11524-016-0080-7
- [8] Mears, Daniel P., Bhati, Avinash S., 2006. No community is an island: the effects of resource deprivation on urban violence in spatially and socially proximate communities. *Criminology* 44 (3), 509–548.
- [9] McGloin, Jean M., Nguyen, Holly, 2013. The importance of studying Co-offending for criminological theory and policy. In: Morselli, C. (Ed.), *Crime and Networks*. Routledge, New York, pp. 13–27.
- [10] Moody, James, White, Douglas R., 2003. Structural cohesion and embeddedness: a hierarchical concept of social groups. *Am. Sociol. Rev.* 68 (1), 103–127.
- [11] Papachristos, Andrew V., 2009. Murder by structure: dominance relations and the social structure of gang homicide. *Am. J. Sociol.* 115 (1), 74–128.
- [12] Sampson, Robert J., 2012. *Great American City*. University of Chicago Press, Chicago, IL.

- [13] Sanburn, J. (2016, September 19). Chicago Responsible for Nearly Half of U.S. Homicide Spike. Retrieved from <http://time.com/4497814/chicago-murder-rate-u-s-crime/>
- [14] Soller, Brian, Browning, Christopher R., 2014. Neighborhood effects and social networks. In: Bruinsma, G.J.N., Weisburd, D.L. (Eds.), Encyclopedia of Criminology and Criminal Justice. Springer, New York, pp. 3255–3265.

APPENDIX A. HISTOGRAMS OF THE VARIABLES

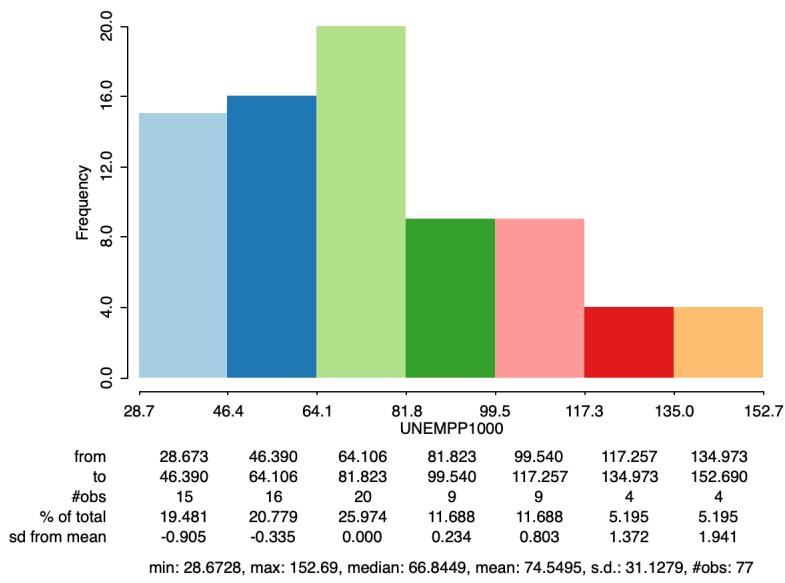


FIGURE 11. Histogram of Number of Unemployment Population per thousand capita in 2014.

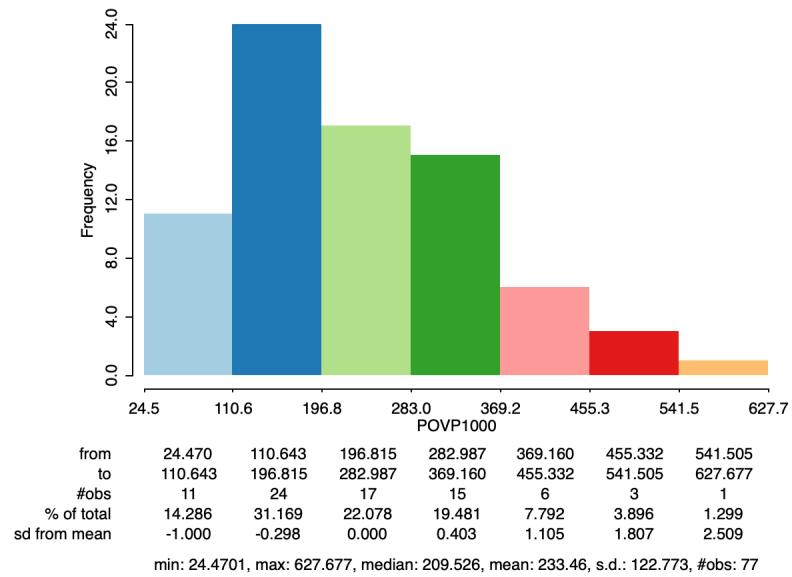


FIGURE 12. Histogram of Number of People Living in Poverty per thousand capita in 2014.

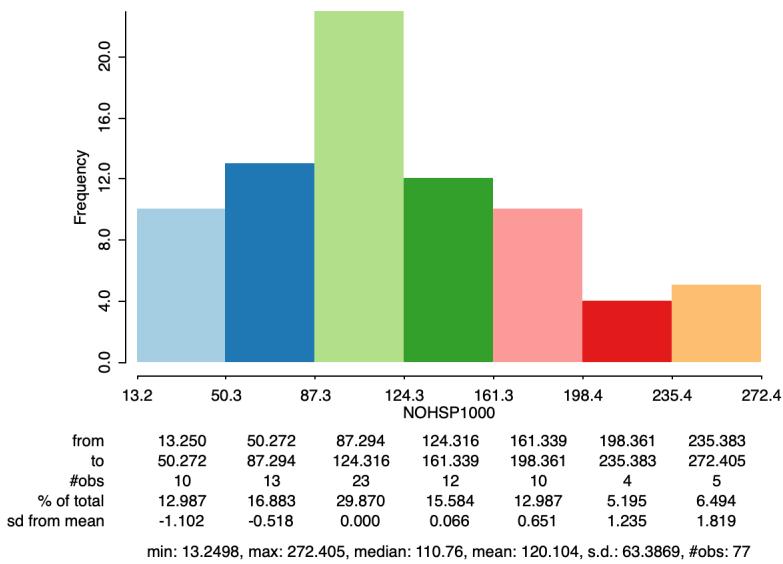


FIGURE 13. Histogram of Number of People without High School Diploma per thousand capita in 2014.

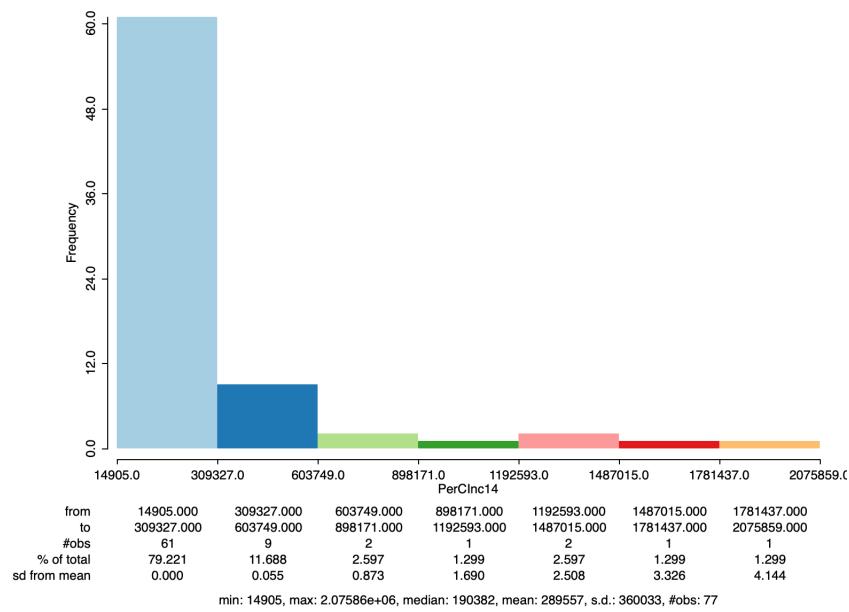


FIGURE 14. Histogram of Income per capita in 2014.

APPENDIX B. QUANTILE MAPS OF THE VARIABLES

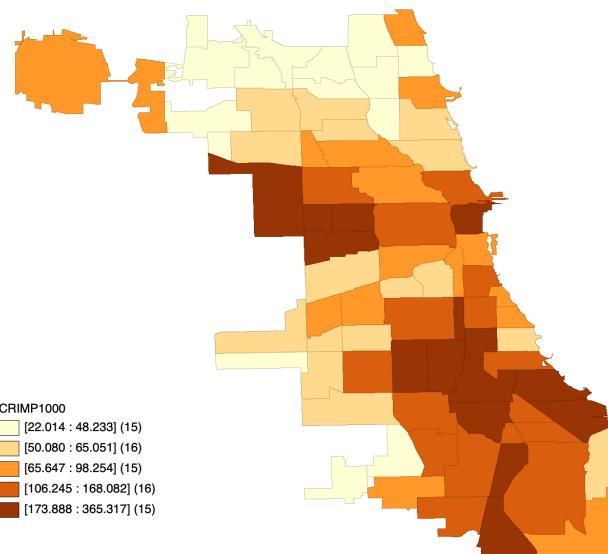


FIGURE 15. Quantile Map of Number of Crime happed per thousand capita in 2014.

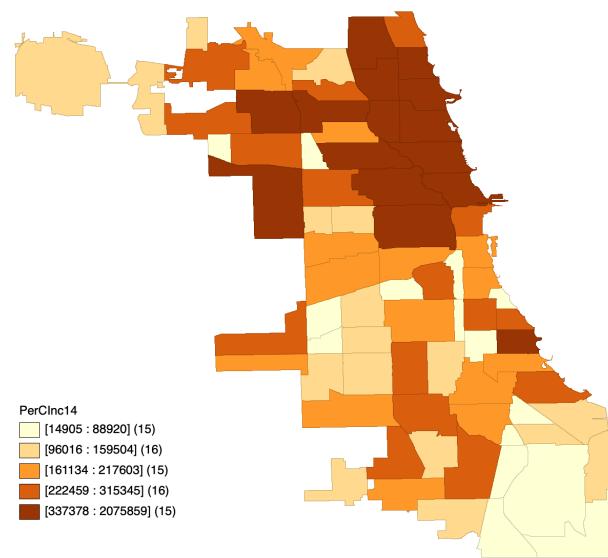


FIGURE 16. Quantile Map of Income per capita in 2014.

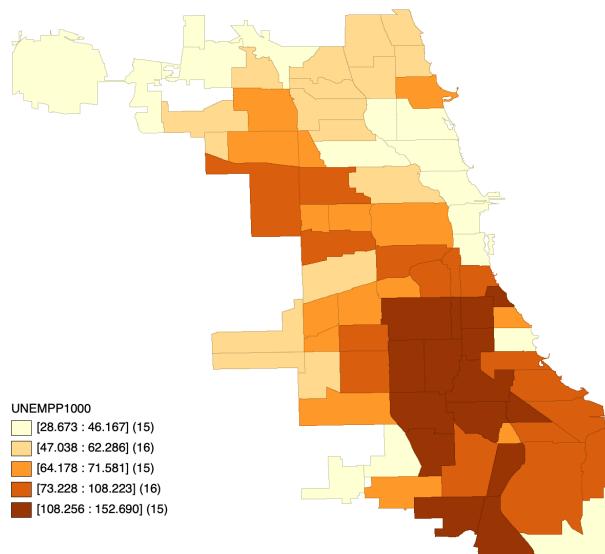


FIGURE 17. Quantile Map of Number of Unemployment Population per thousand capita in 2014.

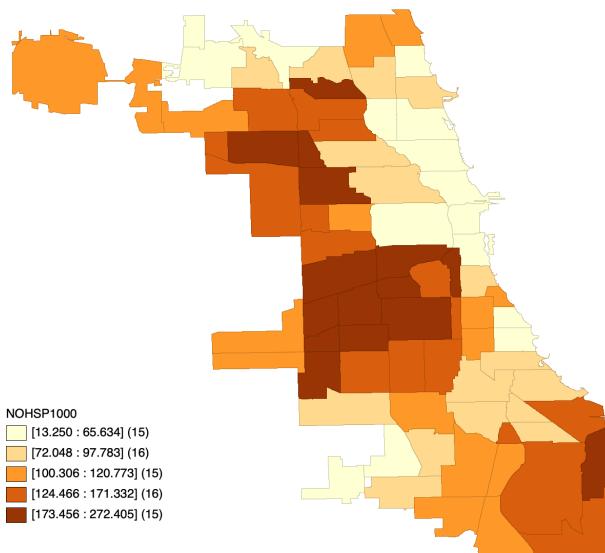


FIGURE 18. Quantile Map of Number of People without High School Diploma per thousand capita in 2014.

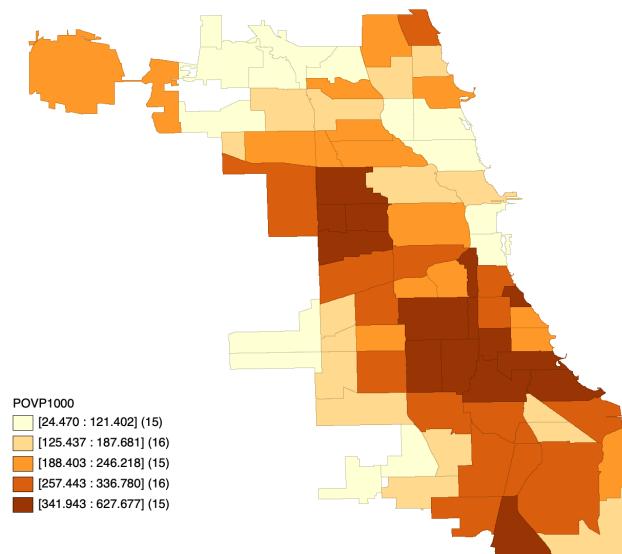


FIGURE 19. Quantile Map of Number of People Living in Poverty per thousand capita in 2014.

APPENDIX C. SCATTERPLOT MATRIX

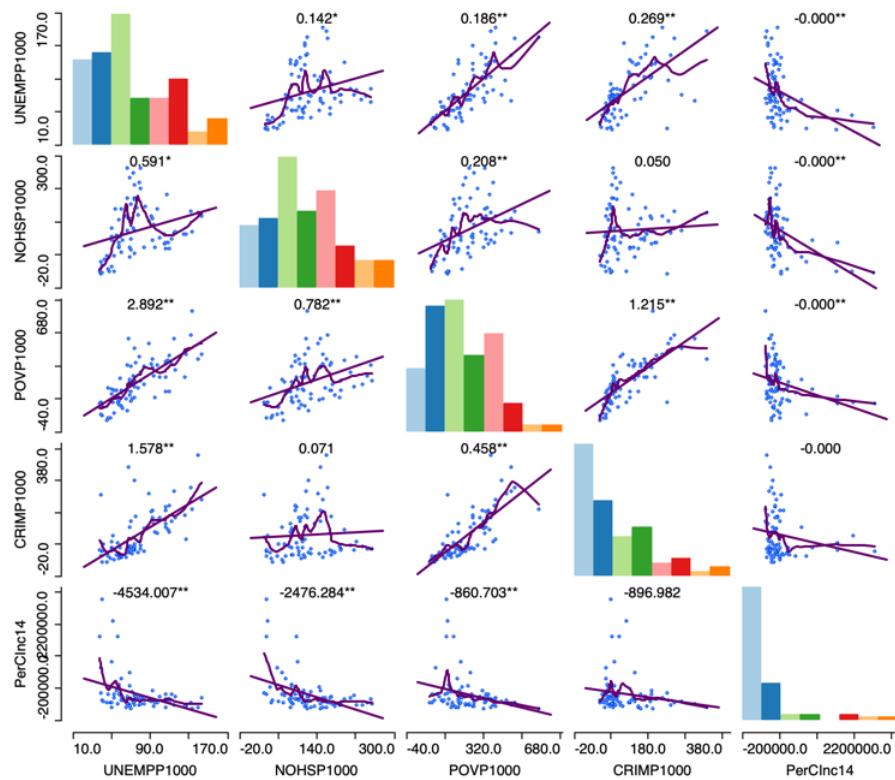


FIGURE 20. Scatterplot Matrix of Variables of Interest.

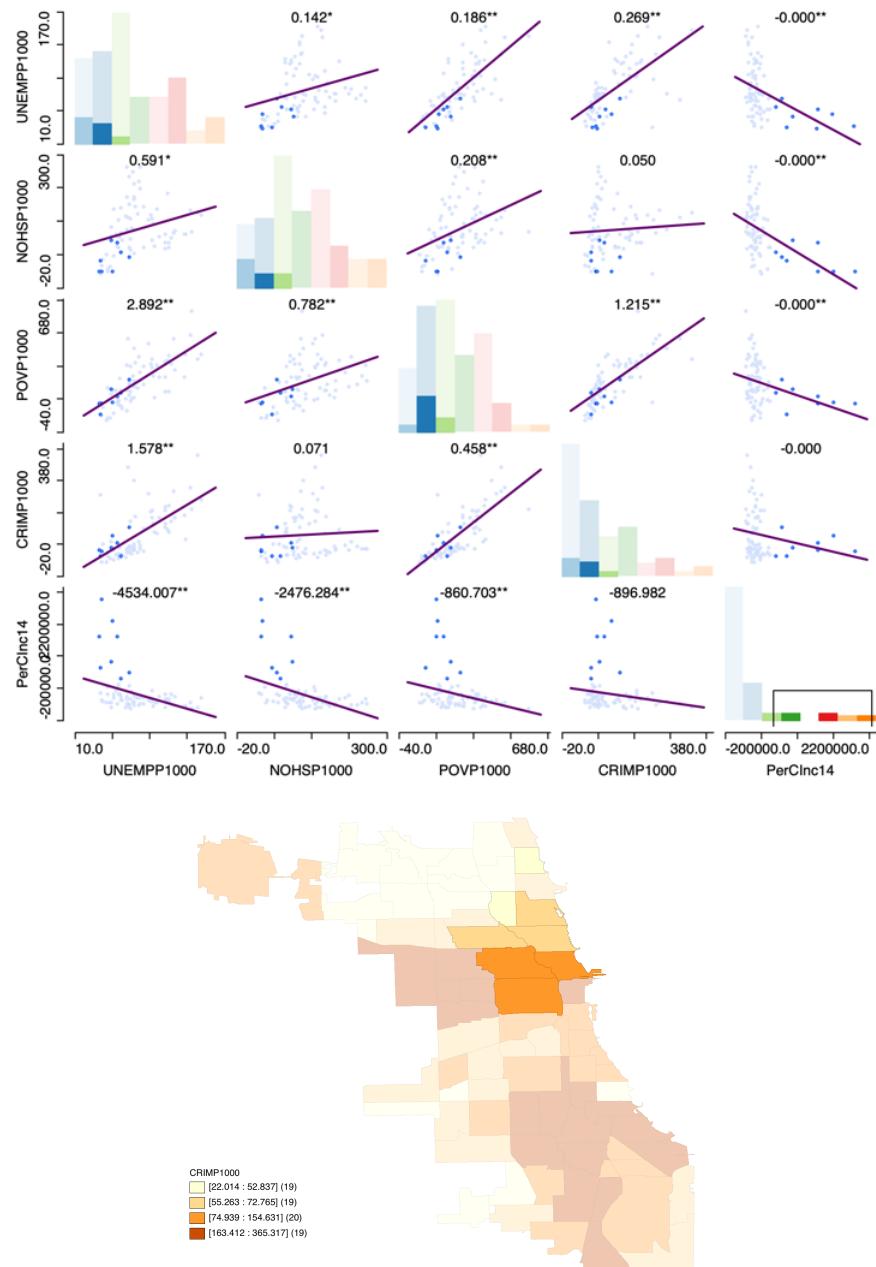


FIGURE 21. Scatterplot Matrix of the Five Variables and Crime Choropleth Map.

APPENDIX D. ROOK WEIGHT

Property	Value
type	rook
symmetry	symmetric
file	OriginPolygonRook.gal
id variable	ComAreaID
order	1
# observations	77
min neighbors	1
max neighbors	9
mean neighbors	4.65
median neighbors	5.00
% non-zero	6.04%

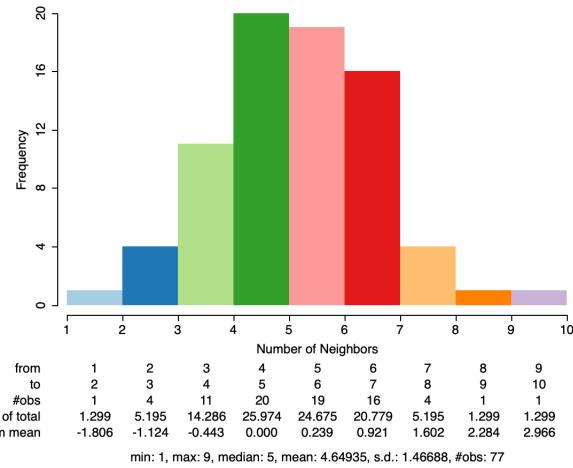
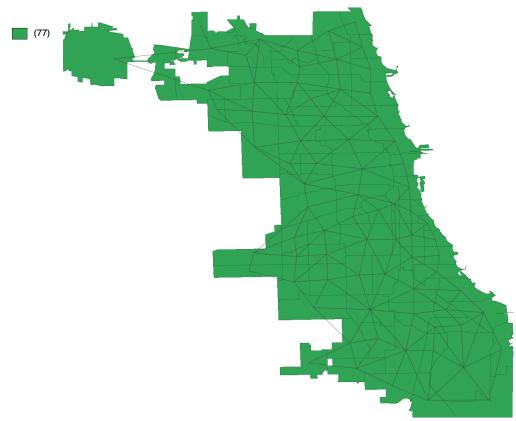


FIGURE 22. Rook Weight with respect to Original Geography.

APPENDIX E. MORAN'S I SCATTERPLOT

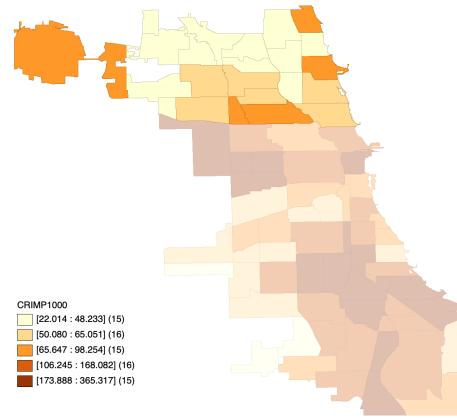


FIGURE 23. Crime Choropleth Map with North Part Selected.

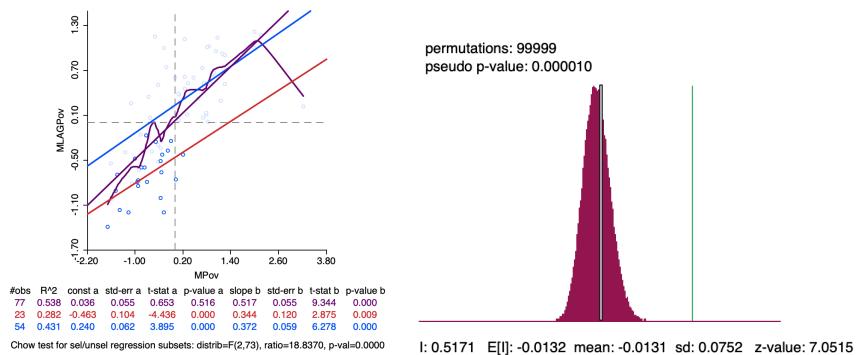


FIGURE 24. The Moran Scatterplot of Poverty Rate with respect to Rook Weight.

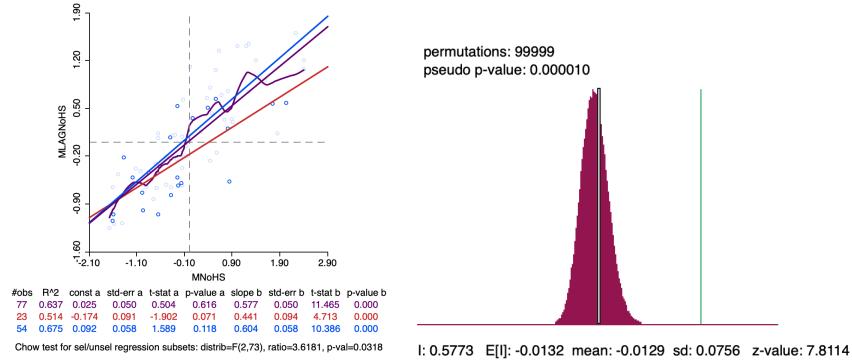


FIGURE 25. The Moran Scatterplot of No High School Diploma Rate with respect to Rook Weight.

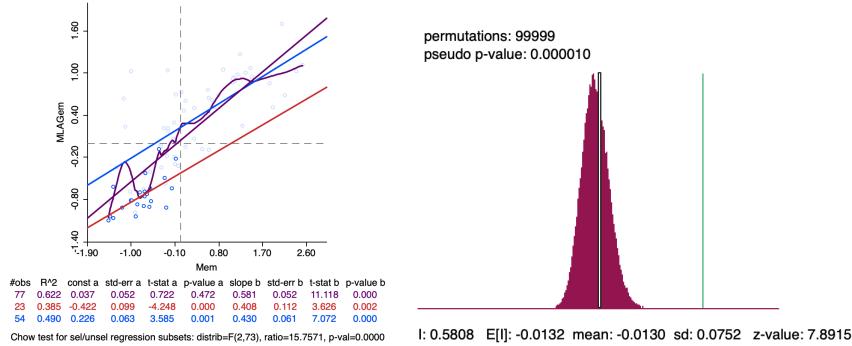


FIGURE 26. The Moran Scatterplot of Unemployment Rate with respect to Rook Weight.

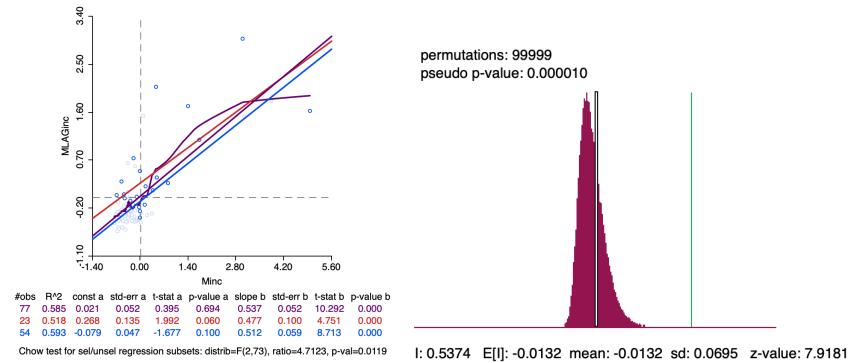


FIGURE 27. The Moran Scatterplot of Income Level with respect to Rook Weight.

APPENDIX F. CLUSTER AND SIGNIFICANT MAP

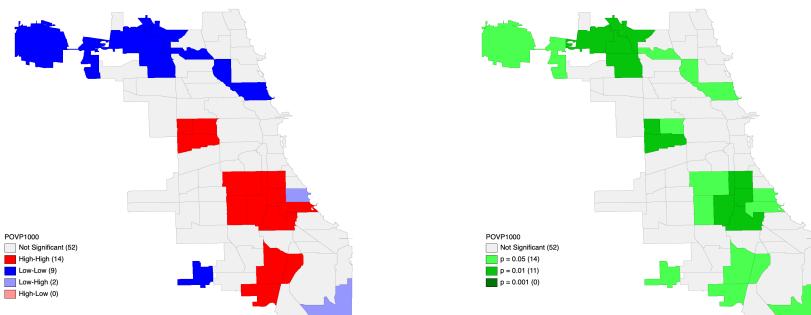


FIGURE 28. The Cluster Map and Significance Map of Poverty Rate with respect to Rook Weight.

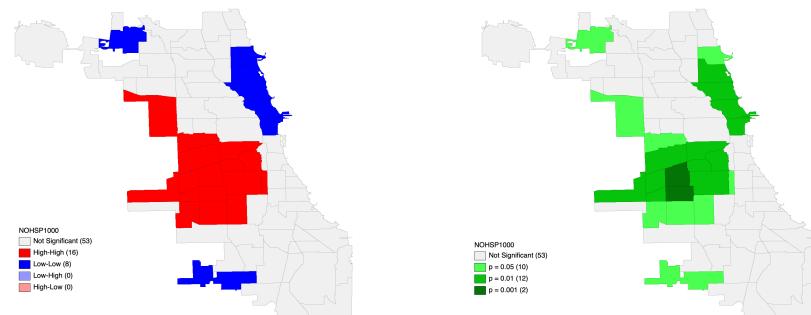


FIGURE 29. The Cluster Map and Significance Map of No High School Diploma Rate with respect to Rook Weight.

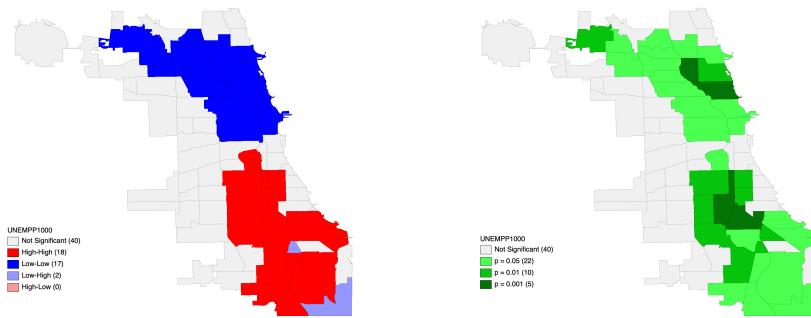


FIGURE 30. The Cluster Map and Significance Map of Unemployment Rate with respect to Rook Weight.

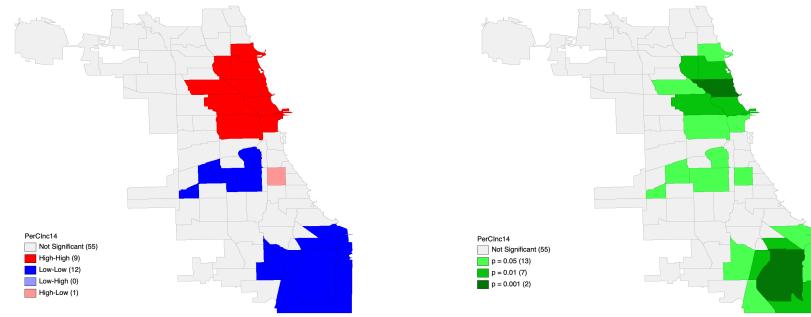


FIGURE 31. The Cluster Map and Significance Map of Income Level with respect to Rook Weight.

APPENDIX G. PRINCIPLE COMPONENT ANALYSIS

PCA method: svd

Standard deviation:

1.662251 1.077649 0.776793 0.554215 0.406241

Proportion of variance:

0.552615 0.232266 0.120682 0.061431 0.033006

Cumulative proportion:

0.552615 0.784881 0.905563 0.966994 1.000000

Kaiser criterion: 2.000000

95% threshold criterion: 3.000000

Eigenvalues:

2.76308

1.16133

0.603408

0.307155

0.165032

Variable Loadings:

	PC1	PC2	PC3	PC4	PC5
UNEMPP1000	0.525054	0.141586	-0.119044	-0.806963	-0.197257
NOHSP1000	0.309563	-0.649624	0.613058	0.0745908	-0.317418
POVP1000	0.543921	0.170144	0.278339	0.157682	0.756879
CRIMP1000	0.470223	0.468109	-0.0772764	0.527739	-0.524676
PerCIInc14	-0.33396	0.556657	0.725634	-0.199687	-0.110386

Squared correlations:

	PC1	PC2	PC3	PC4	PC5
UNEMPP1000	0.761731	0.0232805	0.00855118	0.200016	0.00642154
NOHSP1000	0.264784	0.490095	0.226785	0.00170894	0.0166278
POVP1000	0.817455	0.0336194	0.0467474	0.00763702	0.0945407
CRIMP1000	0.610944	0.254477	0.00360334	0.0855454	0.0454309
PerCIInc14	0.308164	0.359857	0.317721	0.0122479	0.00201093

FIGURE 32. PCA Result Table.