

How Gentrified Degree Affect Crime Rates in Communities of Chicago

Final Project Report

Yutao Chen

1. Introduction

The city of Chicago has been accused of terrible social security for a long time. During many years, the crime rates in Chicago was much higher than the nation's average level (see "Crime in Chicago" 2019). Especially in 2016, The city was responsible for nearly half of the homicides' increase in the U.S. On the other hand, Chicago was also characterized by its history of gentrification, a process that gradually set up segregation between white and non-white residents among different communities (see "Gentrification of Chicago" 2019). Communities that are dominant by the white people tend to be superior in multiple facets like education, health care, economy and also social security, to those where non-whites are majorities. For many of the residents, moving to highly gentrified neighborhoods ensures them to live in places where the crime rates are lower. However, from some previous studies, the correctness of such a hypothesis is put into doubt. For example, McDonald (1986) claimed in his research that "Crime may act as a feedback mechanism to deter neighborhood stability resulting from gentrification". Therefore, in this project, the research question is how the crime rates (per capita) in communities of Chicago relate to their gentrified degrees, and specifically, whether there are any spatial patterns or differences among different communities.

2. Data description

In this research, two parts of the data are required. The first part is the data to measure the gentrified degree of each community in Chicago. According to previous studies (Martin (2017)), quantitative measurements of gentrified degree always incorporate important socioeconomic indicators like income per capita and education level. From my perspective, race composition, especially the ratio of whites or non-whites is also an inevitable indicator to evaluate the gentrified degree of a certain area, since highly gentrified areas are always characterized by a high percentage of white residents. Thus, the entire data set to measure the gentrified to degree contains multiple socioeconomic indicators exported from *Chicago Data Portal* and also population and ratios of white residents in each neighborhood from 2017 census data. The second part is the records of crimes happened in Chicago during 2017. This dataset is responsible for calculating the crime counts in each neighborhood(community) and also other information like the type of crimes. Table 1 illustrates all relevant original variables used in this project, and all variables are standardized before being put into use.

Table 1: Variables Annotation

Variables	Description	Type
Community Number	a unique index for each community	Integer
Community Name	the name for each community	String
Crime Frequency	the frequency of total crimes for each community	Integer
Population	the population for each community	Integer
House Crowding Ratio	the ratio of crowded housing for each community	Real
Poverty Ratio	the ratio of households in poverty	Real
Unemployment Ratio	the ratio of unemployed over 16 for each community	Real

Variables	Description	Type
Uneducated Ratio	the ratio of people without high school diplomas over 25 for each community	Real
Dependency Ratio	the ratio of people under 18 or over 64 for each community	Real
Income	the average income per capita for each community	Real
White Ratio	the ratio of white residents for each community	Real

3. The measurement of the gentrified degree

Before measuring the gentrified degree, I need to mention the difference between gentrified degree and gentrification. If the object is to measure gentrification, then it is better to use panel data, where both sectional and temporal variations are presented in the data sets. But for gentrified degree, it is acceptable to only use the cross_section data because they can reveal that to what degree an area has been gentrified by those socioeconomic and demographical variables.

In this research, the explanatory variable is the gentrified degree for each community, which is derived from the socioeconomic indicators and the white ratio of the original dataset. I did a pretest for these indicators via correlation test and the result shows that they are highly correlated (see Figure 1), which is consistent with Chicago's trait of being highly segregated by socioeconomic conditions at the community level. This ensures that these variables are all suitable for gentrified degree measurement. The exact statistics to represent the gentrified degree is designed to be the principal component score of the original variables. Compared to the simple average, principal component analysis can make sure how much variance among observations is explained, so I can avoid choosing statistics that are inefficient to describe the data. For the selection of principal components, Table 2 suggests that PC1 can explain over sixty percent of the variance in the observations, and all the indicators take a non-negligible proportion in PC1, Their values suggest exactly the opposite with our concept of gentrified areas: lower income per capita, higher unemployment rate, lower white residents ratio, etc. Therefore, the additive inverse of PC1 is chosen to be the statistics to measure gentrified degree, a higher score of which represents a relatively higher gentrified degree.

Table 2: PCA Variable Loadings and variance explained proportion

Indicators	PC1	PC2	PC3	PC4	PC5	PC6	PC7
House Crowding Ratio	+0.2925	+0.6477	-0.1310	+0.1415	+0.2974	-0.2153	-0.5683
Poverty Ratio	+0.3691	-0.2461	-0.6133	+0.4316	-0.1857	-0.4010	+0.2131
Unemployment Ratio	+0.3905	-0.4221	-0.1605	+0.1012	+0.2844	+0.6696	-0.3227
Uneducated Ratio	+0.3671	+0.5128	-0.0080	+0.0336	+0.0576	+0.3974	+0.6633
Dependency Ratio	+0.3597	-0.2414	+0.6760	+0.2949	+0.3705	-0.3309	+0.1469
Income	-0.4330	-0.0486	-0.3268	+0.0249	+0.7964	-0.0781	+0.2497
White Ratio	-0.4173	+0.1349	+0.1303	+0.8335	-0.1461	+0.2663	-0.0609
Proportion of variance	+0.6260	+0.1909	+0.0923	+0.0390	+0.0225	+0.0167	+0.0127
Cumulative proportion	+0.6260	+0.8169	+0.9092	+0.9482	+0.9707	+0.9874	+1.0000

4. Spatial weight

In this project, a highlight is to discover spatial patterns involving the relationship between gentrified degree and crime rates, and the choice of spatial weights is decisive to discover the patterns. Therefore, before deciding the weights used in the following analysis, three candidates are tested to see whether they are suitable to match the gentrified degree distribution in Chicago:

- Contiguity Based Queen Weights (first-order)
- Contiguity Based Queen Weights (first and second-order)
- Distance-Based Weights (distance threshold is the minimal distance that no community is isolated)

Comparisons are made between the box map of gentrified degrees and the connectivity map of different weights. From the box map, we can know that Chicago has a quite segregated structure in the aspect of gentrification, where the northeastern part (near Lake Michigan) is highly gentrified while the west and middle part tends to be the least gentrified. More importantly, these two areas are spatially close or even adjacent to each other. Generally, we hope to see the neighbors of different communities be different, otherwise, they might yield unrealistically similar spatial patterns. Setting the community of North Side as an anchor, the first order queen weights suggest that the neighbors are those similarly highly gentrified communities around (see Figure 2). But when the second order is included, some least gentrified communities are counted as second-order neighbors of Near North Side (see Figure 3). Such a tricky situation also happens when I tested the distance-based weights, where even using the minimal distance threshold that no community is isolated can still cause very unlikely neighbors to be combined (see Figure 4). Therefore, to avoid the lack of differentiation between highly and lowly gentrified areas, only the contiguity based first-order queen weights are used in following spatial pattern analysis.

5. Results

5.1 Gentrified degree and crime rates

In this section, I would examine the basic relationship between the gentrified degrees of different communities and their overall crime rates. My general assumption is that **in highly gentrified communities, crime rates tend to be lower, and vice versa**. From the correlation plot (see Figure 5), the crime rate shows a strong negative correlation with gentrified degrees as expected, which means the assumption is established in the level of the whole city.

However, when I added spatial elements into our analysis, the relationships become different to some extent. Based on the Local Moran's I test, there are both high-high and low-low spatial clusters of gentrified degrees. In the high-high clusters ($p = 0.1$, a slightly higher p-value to obtain more observations in this cluster, but the trend is the same when p is smaller), the trend is the opposite of the overall, a higher gentrified degree implying a higher crime rate in the neighborhood (see Figure 6). While in the low-low clusters, the general assumption is still established (see Figure 7). Though the significance level is weaker than that of the overall relationship, these local differences at least suggest that the relationship between gentrified degrees and crime rates is not distributed evenly among the communities in Chicago.

To understand the reason for this local difference, especially the divergence occurred in the high-high clusters, I compared the Local Moran's I test map of gentrified degree with the quantile map of crime rate (see Figure 8), it is clear to see that in the high-high clusters, some neighborhoods have very high crime rates with greater gentrified degrees, such as Loop (the community that has the highest crime rate) and Near North Side (see Figure 9), which could contribute to a positive relationship between the two variables.

5.2 Gentrified degree and crime types

Based on the analysis in 5.1, it is reasonable to speculate that the crimes happened in heavily and slightly gentrified neighborhoods have different characteristics to cause the different patterns. For example, the major types of crimes might be different in these two areas, and the motivation of different types of crime might have different relationships with gentrified degrees. Though I could not precisely obtain the motivations for each type of crime and measure their relationships with gentrified degrees in this study, it is possible to make some deductions according to the relationship between the distribution of different crime types and gentrified degrees.

To categorize different types of crimes, I used the variable *FBI Code* from the criminal record, which contains 29 types of crimes, and calculated per capita rates of each type of crime for every community. The result is suggested in Table 3, where I chose the top 6 crime types with their average rate values within clusters. Besides some common crimes like theft and arson, crimes related to illegal property acquisition like fraud and burglary ranks at higher positions for highly gentrified clusters, while vandalism and aggravated assault happened frequently in lowly gentrified clusters. Thus, it seems that in highly gentrified spatially auto-correlated clusters, wealth is an important factor incurring crimes, which indicates a more gentrified neighborhood is more vulnerable. While in lowly gentrified spatially auto-correlated clusters, crimes tend to be more socially destructive instead of economically, so a more gentrified neighborhood that is associated with a higher educational level and lower unemployment rate is likely to have a lower crime rate.

Table 3: Top 6 Crime Types in Different Auto-correlated Clusters

Cluster Type	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
high-high	Theft (0.0307)	Arson (0.0111)	Fraud (0.0082)	Vandalism (0.0062)	Other Offenses (0.0054)	Burglary (0.0033)
low-low	Arson (0.0352)	Theft (0.0253)	Vandalism (0.0172)	Other Offenses (0.0152)	Drug Abuse Violations (0.0106)	Aggravated Assault (0.0103)

From the cases of high-high and low-low clusters, it is reasonable to check whether this pattern exists among all the communities, which is that *highly gentrified areas are more associated with economically destructive crimes and lowly gentrified areas are dominated by socially destructive crimes*. Before testing this assumption, I firstly executed a PCA on rates of different crime types in different communities, and the result (see Table 4) suggests that PC1 and PC2 are the most explanatory principle components. For PC1, it suggests that serious crimes like homicides, forcible rapes, Robbery, Aggravated Assault and Burglary (FBI 1-5) are always associated with each other. Other illegal behaviors like sex offenses, drug abuse and gambling (FBI 17-19) which are likely to be tied to social instability also take a certain portion in this principle component. As a comparison, the dominant crime types in PC2 like forgery and counterfeiting, fraud and embezzlement (FBI 10-12) are more relevant to business (see Table 5-6). This result is consistent with the two major types of crimes mentioned in the assumption, making it effective to do the following tests with PC1 and PC2.

Table 4: Proportion of Variance by Principle Components

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
0.5279	0.1495	0.0721	0.0468	0.0327	0.0279	0.0259	0.0214	0.0179	0.0172	0.0125	0.0116

Table 5: Variable Loadings for PC1 and PC2 (1)

Index	FBI1	FBI2	FBI3	FBI4	FBI5	FBI6	FBI7	FBI8	FBI9	FBI10	FBI11
PC1	0.250	0.246	0.248	0.278	0.214	0.105	0.259	0.278	0.207	0.078	0.102
PC2	-0.130	0.042	0.137	-0.071	-0.100	0.486	0.012	0.011	-0.137	0.462	0.451

Table 6: Variable Loadings for PC1 and PC2 (2)

FBI12	FBI13	FBI14	FBI15	FBI16	FBI17	FBI18	FBI19	FBI20	FBI22	FBI24	FBI26
-0.004	0.078	0.269	0.251	0.146	0.187	0.210	0.195	0.231	0.051	0.267	0.265

FBI12	FBI13	FBI14	FBI15	FBI16	FBI17	FBI18	FBI19	FBI20	FBI22	FBI24	FBI26
0.379	0.040	-0.050	-0.129	-0.020	0.113	-0.070	-0.111	-0.112	0.256	-0.051	0.027

The examination is performed clustering analysis with scores of two crime type principle components and gentrified degrees, where K-means shows a better performance than hierarchical methods regarding the ratio of between to total sum of squares (k_means: 0.77, hierarchical: 0.73). The objective number of clusters is tested among 2,4 and 6. It is found that when the number of clusters is too small, the clustering results can hardly distinguish well the difference between observations, as the ratios of between to total sum of squares are not satisfactory (less than 50%). However, when the number of clusters becomes too big, though the explanatory strength has been increased, some cluster centers might become too similar to another. Therefore, the final benchmark is set to be the clustering result of 4 clusters (see Figure 10). From the clustering result (Table 7), we see that for clusters with a higher gentrified degree (C2 and C4), the assumption established well since the corresponding cluster centers all suggest that the PC2 dominates PC1. However, for clusters whose gentrified degrees are lower (C1 and C3), the assumption only established partially. In C1, the PC2 score is higher than PC1, while C3 shows consistency with the assumption. Despite this anomaly, we can also find that when the Gentrified degree tends to be neither too high or too low (close to 0), the difference between the scores of two crime type principle components tends to be smaller.

Table 7: Kmeans-cluster centers (n = 4)

	Gentrified	Crime_PC1	Crime_PC2
C1	-1.0757	-1.1652	-0.2958
C2	2.4285	-2.5625	0.2568
C3	-2.1144	5.0836	-0.6197
C4	4.4103	2.7020	13.4881

To find the reason for this anomaly, I used local Moran’s I test to reveal how the communities in C1 distribute on the scores of PC1 and PC2. Figure 11 suggests that this anomaly is mostly contributed by several spatial outliers. Firstly, the community of Albany Park is a low-high spatial outlier among surrounding highly gentrified clusters but has the trait of a low PC1 score, which is more like gentrified communities. Secondly, the community of South Deering is a high-low spatial outlier regarding the distribution of PC2, and it is also a member of low-low gentrified clusters when the p level is 0.05. Also, the majority of C1 are some communities locating in the middle of Chicago, which are primarily labeled with low gentrified degrees. However, these communities do not reveal a significant trait as other lowly gentrified neighborhoods to balance the impact of the above outliers, otherwise, some high-high/low-low spatial clusters should be formed on the Local Moran’s I map of PC1/PC2.

6. Discussions

In this section, I would discuss those unusual cases discovered in the above analysis process. These special cases are typically impactful in forming some unexpected patterns among all the communities, and most of them are outliers in different aspects. I would primarily explain these special cases based on their attribute values available in the dataset and try to build the connection between them and our intuitive recognitions.

1. Loop: The community of Loop is the downtown area of Chicago, which is one of the most gentrified neighborhoods. However, Figure 12 suggests it serves as an outlier in crime rates and especially, with the highest PC2 crime type scores. The positive relationship discovered in the high-high gentrified clusters is largely contributed by this area. The excessive amounts of economical or commercial crimes that happened in Loop lead us to cast doubt on the claim that gentrification can help to reduce the crimes in places that are gentrified. From my view, it is more of a transformation than a reduction concerning crimes of gentrified areas. The transformation starts with socially destructive behaviors

to more and more economically destructive. For those communities which are on the process, they might experience a period when the crime rates are decreasing, but when their gentrified degrees reach a certain level, they are likely to start suffering from the increase of economically destructive crimes. It is hard to decide whether socially destructive or economically destructive crimes cause more harm to our life since our ultimate goal is to restrain the happening of both, but gentrification is unlikely to serve as a perfect solution.

2. Albany Park: The community of Albany Park is mentioned as a low-high spatial outlier in the last section. It locates in the north of Chicago, surrounded by other highly gentrified neighborhoods. Compared to its neighbors, this community shows a milder value on most of the socio-economic indicators like poverty ratio and unemployment ratio, and also on white ratio (see Figure 13). But for House Crowding Ratio and Uneducated Ratio, it presents excessively high values which are more close to communities with low gentrified degrees. For crime types, however, the neighborhood shows a similarity to its highly gentrified neighbors. These unique traits it displays might be related to its history of immigrants since Albany Park is continuously labeled as one of the most diverse areas in the country. The attraction of people from different races might prevent or at least, slow down the gentrification of this area. Meanwhile, the gentrification happened around Albany Park protects it from being disturbed by most socially destructive crimes.
3. South Deering: The community of South Deering locates at the south of Chicago, where social security is thought to be terrible. As a lowly gentrified neighborhood, it presents unexpected high values on Larceny-theft (FBI 6) and forgery and counterfeiting (FBI 10). This might be related to its remains of heavy industry, offering more possibilities for petty crimes in this area, which has become an issue since the 1990s (“South Deering, Chicago” 2019)).

7. Conclusions

In this study, the relationship between the gentrified degrees of communities in Chicago and their crime rates per capita is explored. The gentrified degrees are measured with the first principle component of multiple socioeconomic indicators as well as the ratio of white residents in different communities. Generally, the crime rate shows a strong negative correlation with gentrified degrees, which means that neighborhoods with better conditions in socioeconomic aspects like average income and education level are likely to be safer. This relationship is found to be violated in highly gentrified spatially auto-correlated clusters because of the existence of outliers. To further understand this anomaly, we examined the relationships between different crime types and gentrified degrees and found out my assumption that highly gentrified areas are more associated with economically destructive crimes and lowly gentrified areas are dominated by socially destructive crimes exists for highly gentrified communities and part of lowly gentrified communities. For some slightly lowly gentrified neighborhoods, this assumption is inapplicable, and the reason for this difference is also discussed.

8. Future work

This project only explores the relationships between gentrified degrees and crime rates and crime types, where the functions of other intermediate variables like police power are not considered. These variables might also make a difference in the relationships. Also, as we mentioned, only the influence of gentrified degrees are measured in this study, instead of the gentrification process. The latter is a more complex but interesting work for further exploration.

Reference

“Crime in Chicago.” 2019. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Crime_in_Chicago&oldid=922238685.

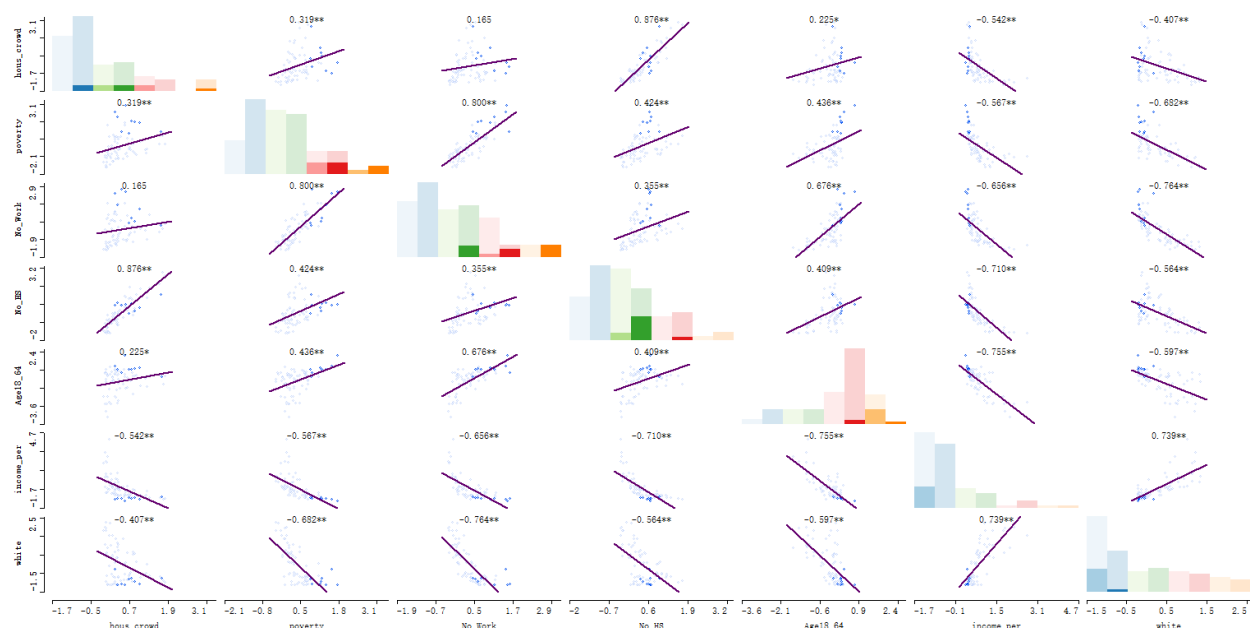


Figure 1: Correlation Matrix Plot of gentrified degree indicators

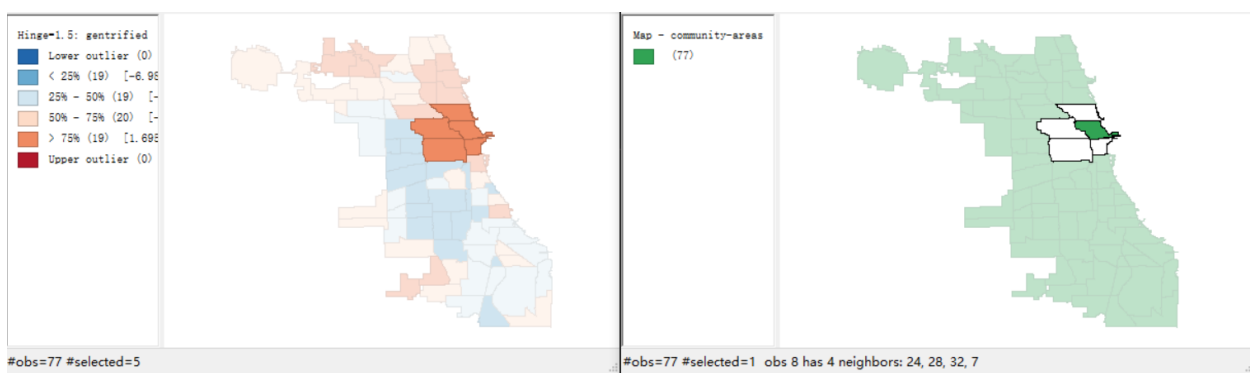


Figure 2: Neighbors: Contiguity Based Queen Weights (first-order)

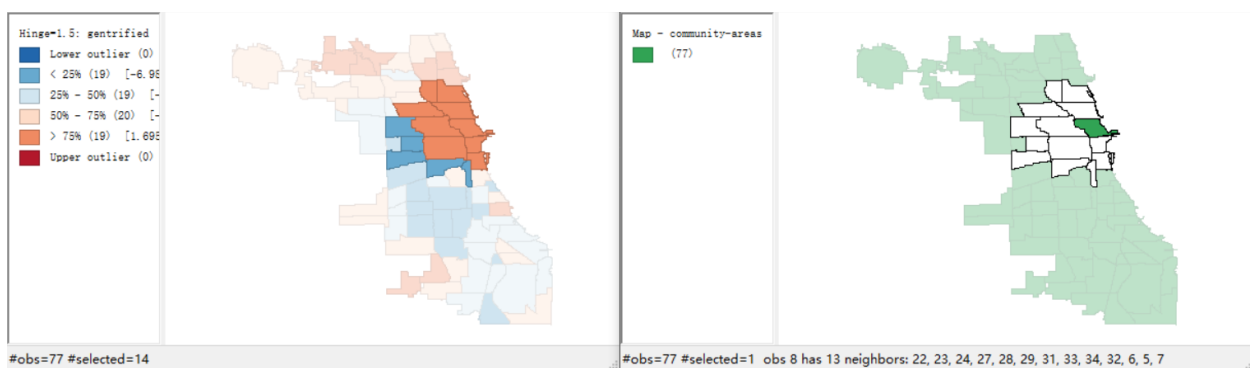


Figure 3: Neighbors: Contiguity Based Queen Weights (first and second-order)

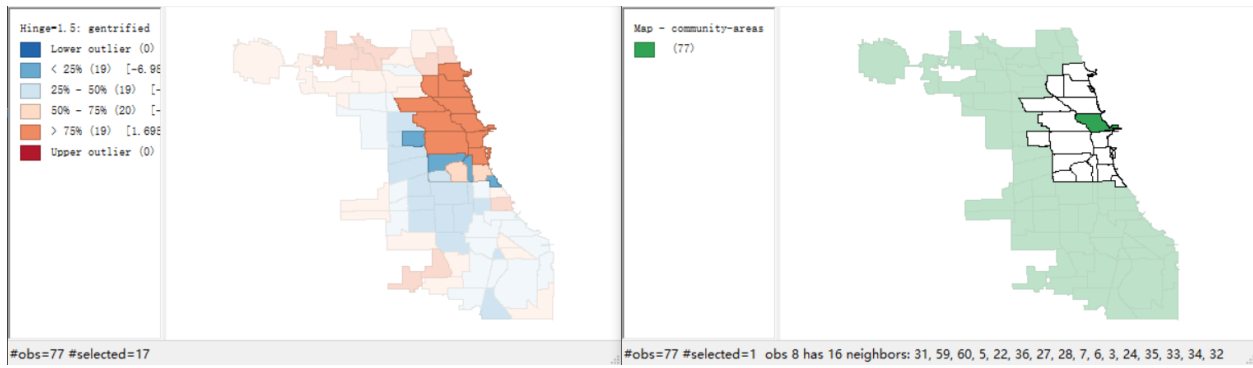


Figure 4: Neighbors: Distance-Based Weights

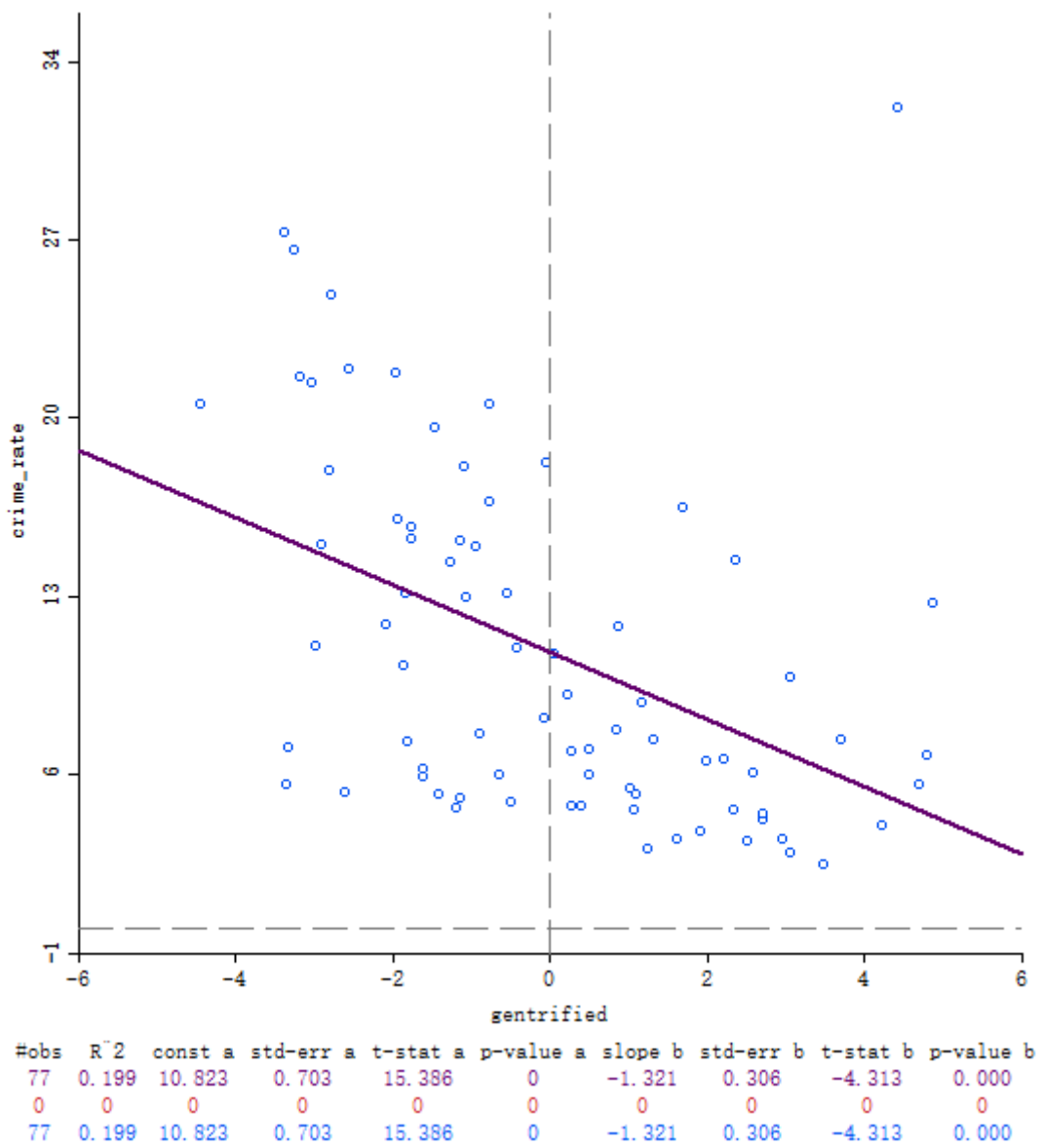


Figure 5: Relationship between Gentrified Degree and Crime Rate Per Capita

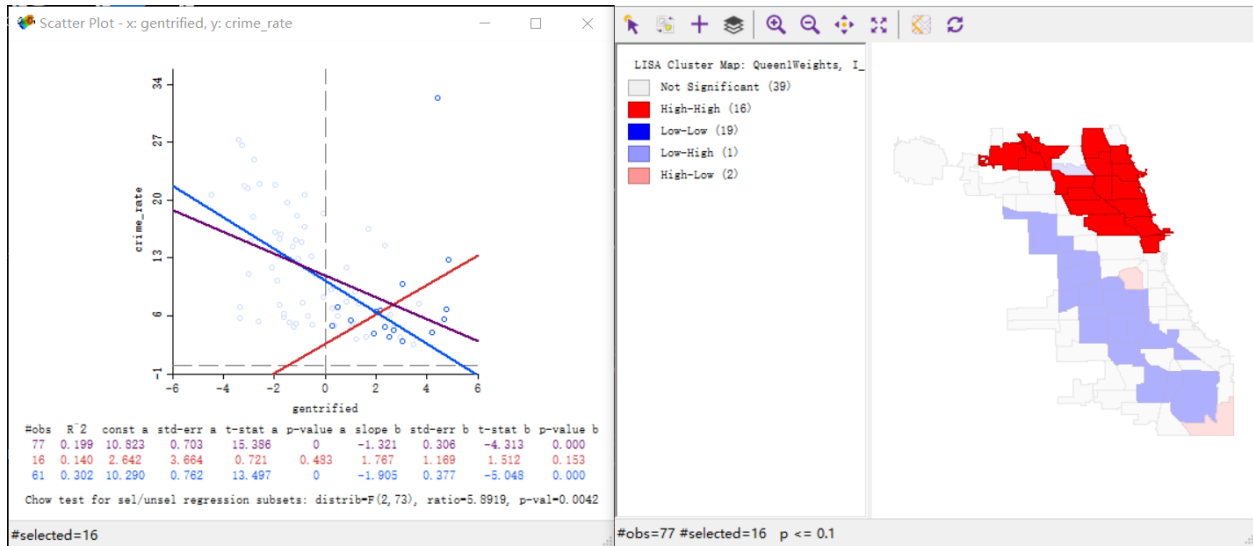


Figure 6: Local Relationship between Gentrified Degree and Crime Rate Per Capita-High-High Clusters

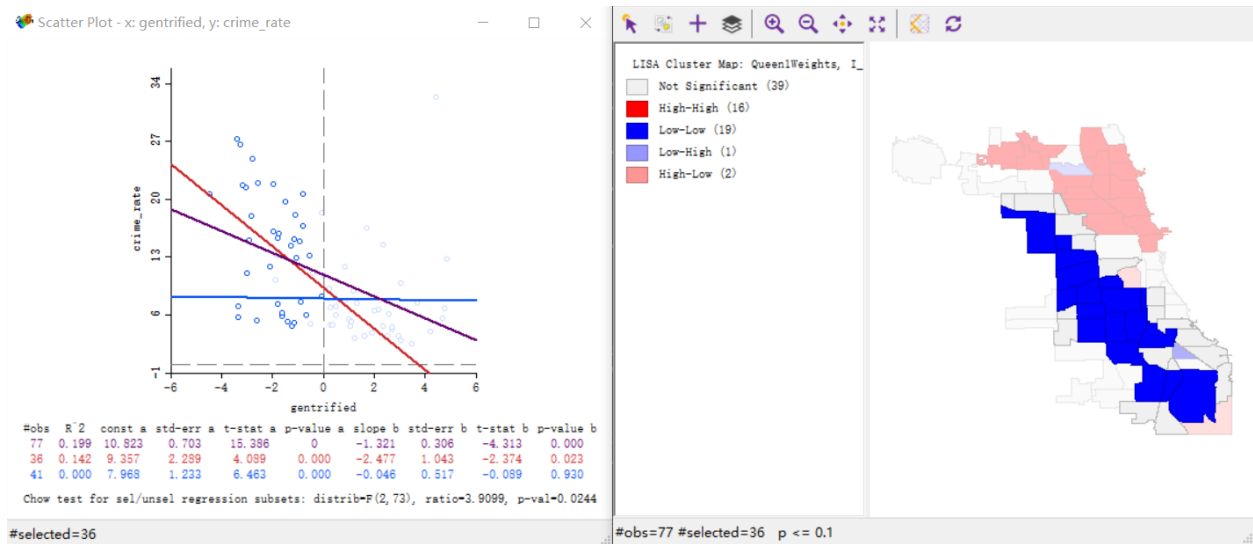


Figure 7: Local Relationship between Gentrified Degree and Crime Rate Per Capita-Low-Low Clusters

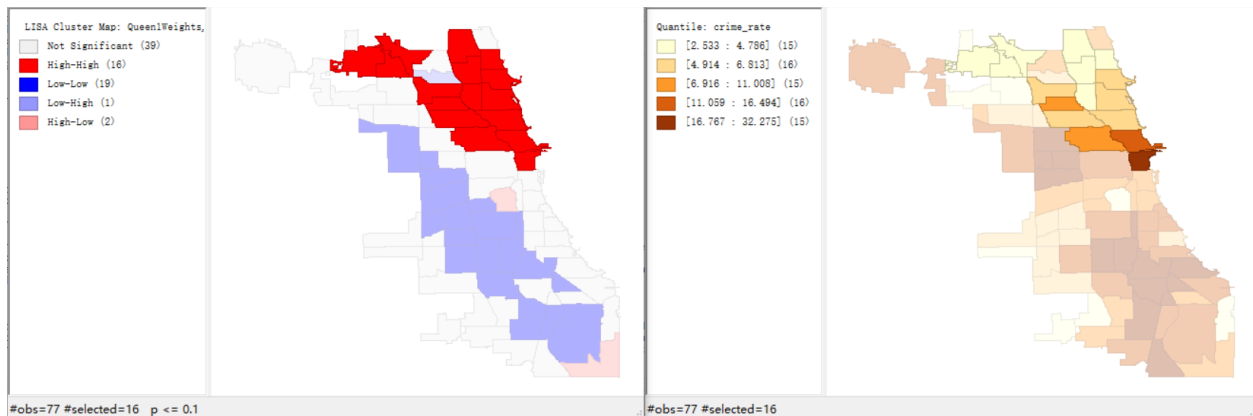


Figure 8: Comparison between Spatial Cluster Map and Quantile Map

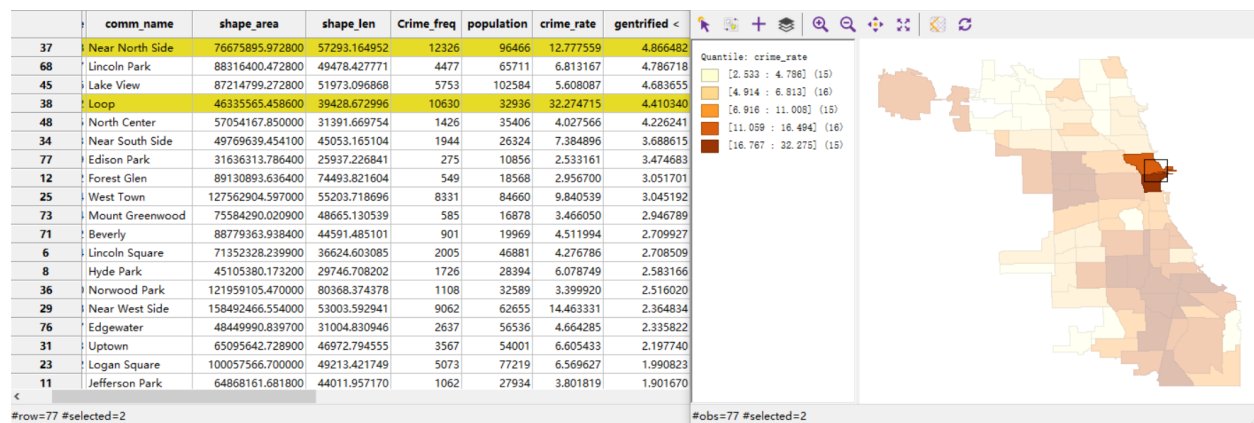


Figure 9: Special Neighborhoods in the High-High Cluster

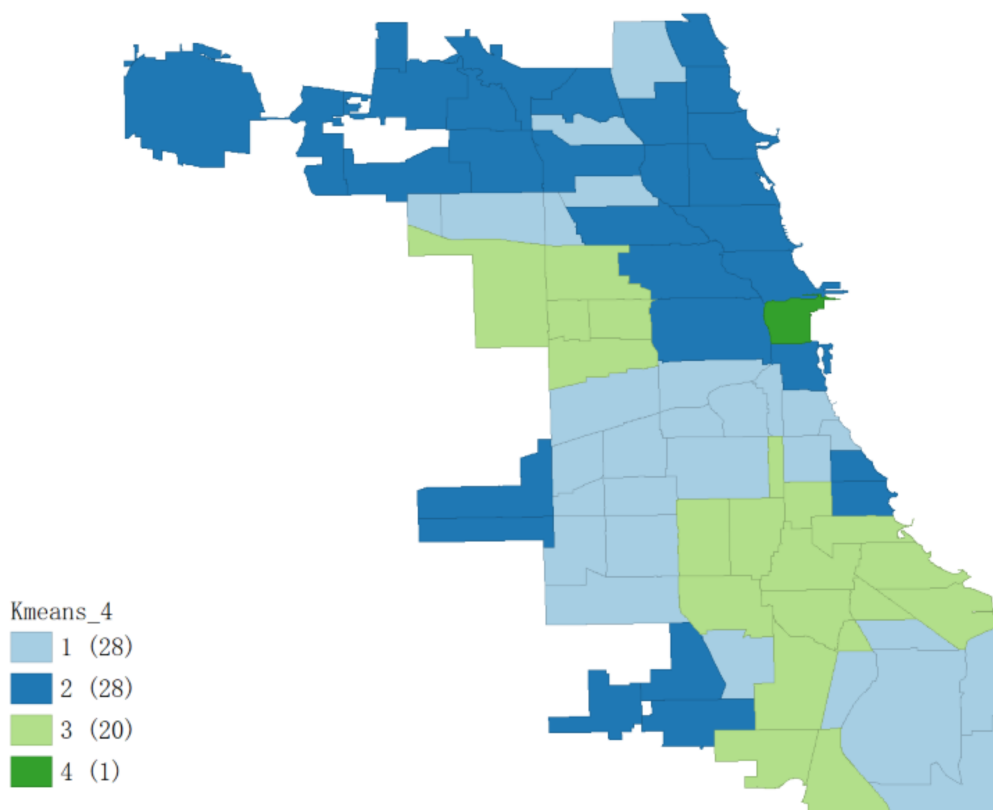


Figure 10: K-means clustering results ($n = 4$)

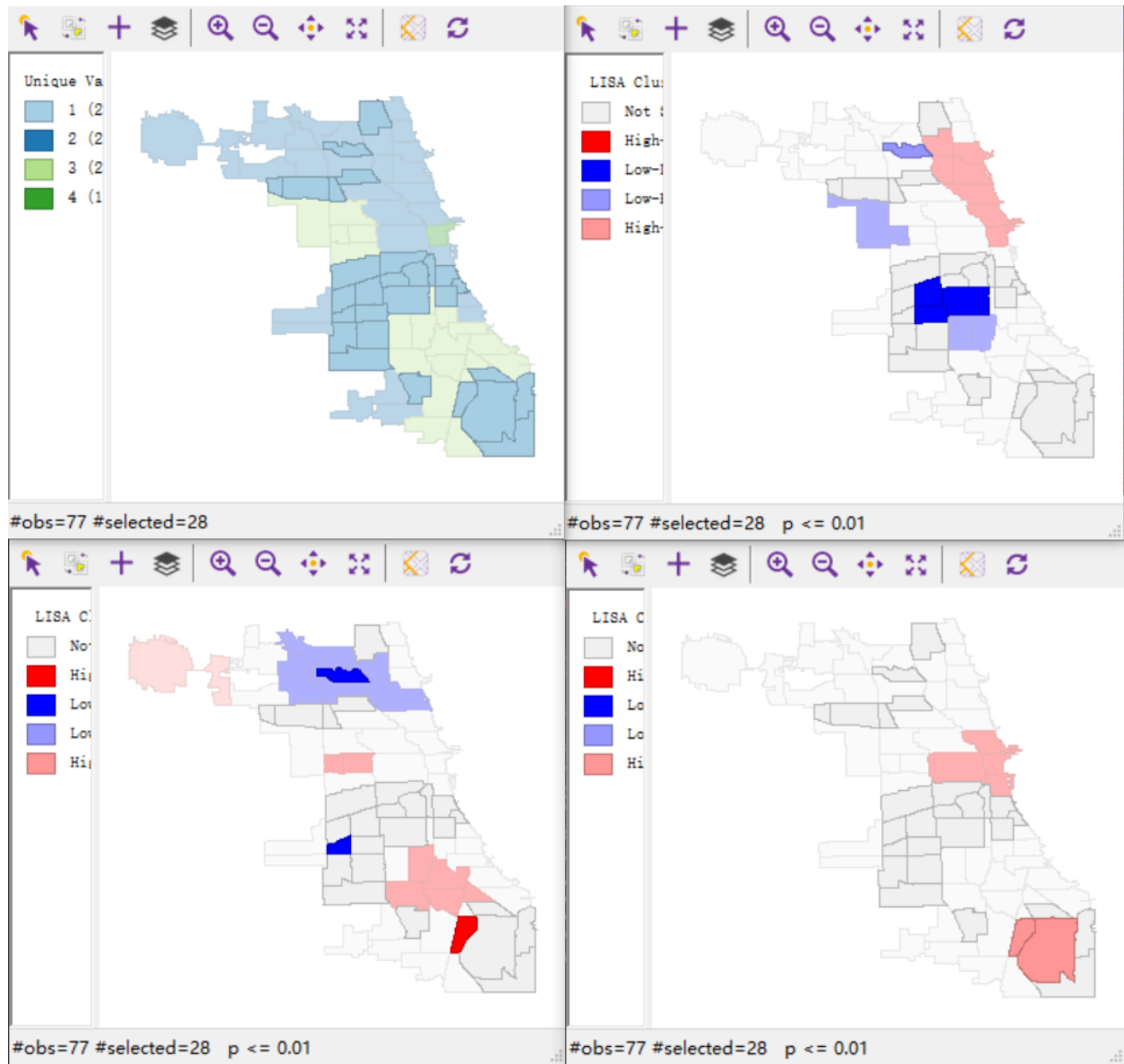


Figure 11: Crime Type Test Anomaly

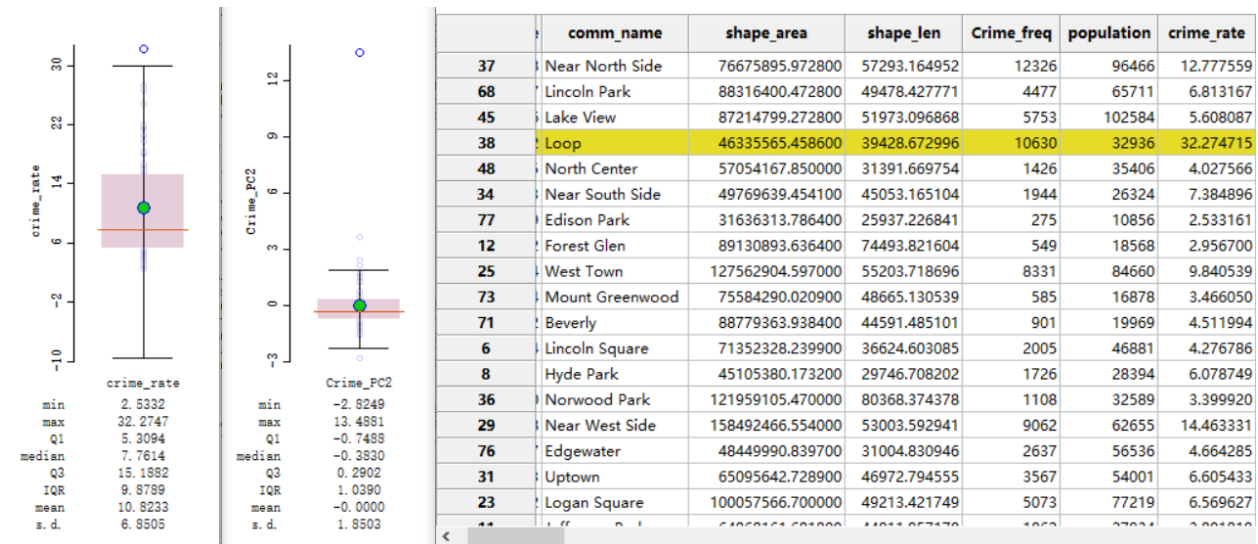


Figure 12: Outlier: Loop

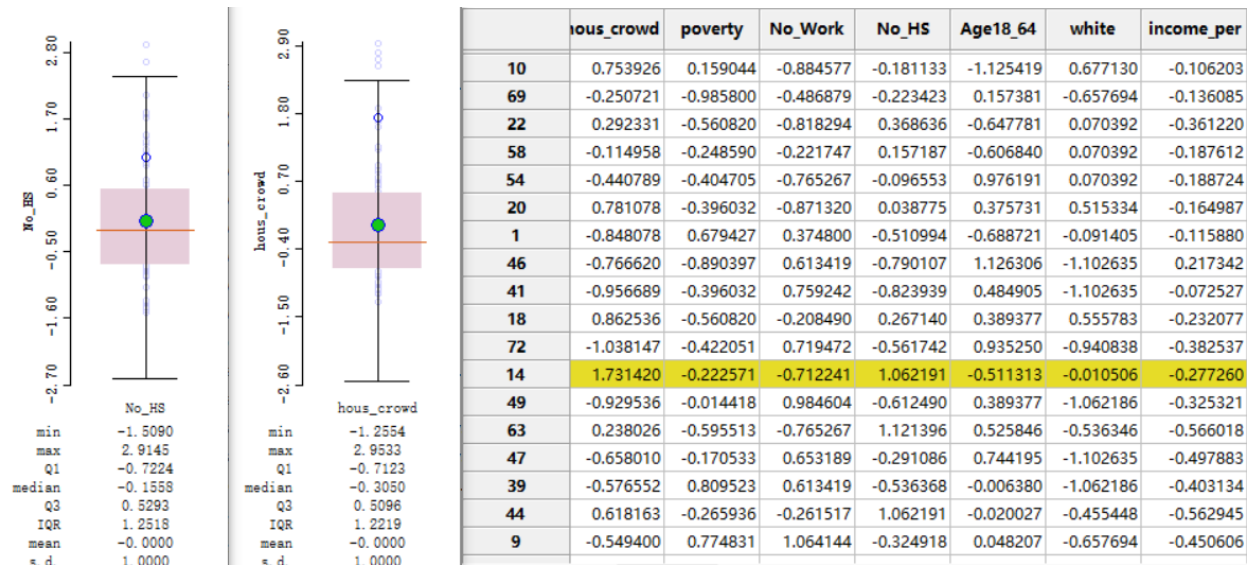


Figure 13: Outlier: Albany Park

“Gentrification of Chicago.” 2019. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Gentrification_of_Chicago&oldid=914124667.

Martin, Richard. 2017. “A Quantitative Approach to Gentrification: Determinants of Gentrification in U.S. Cities, 1970-2010.” PhD thesis, Department of Insurance, Legal, Studies,; Real Estate, Terry College of Business, University of Georgia, 206 Brooks Hall, Athens, GA 30602: University of Georgia.

McDonald, Scott C. 1986. “Does Gentrification Affect Crime Rates?” *Crime and Justice* 8 (January): 163–201. <https://doi.org/10.1086/449122>.

“South Deering, Chicago.” 2019. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=South_Deering,_Chicago&oldid=926344540.