**Exploring Spatial Patterns of Airbnb Rents and Crime Rates in Chicago**

Final Project for Spatial Data Science

LI LIU

12/7/2018

## Introduction

The research question of this project is to find the spatial patterns of Airbnb rents and crime rates in Chicago. Many economics and sociology theories suggest that there is a negative correlation between a neighborhood's social-economic status and attractions to tourists. When travelers visit a new place, they have abundant choices on Airbnb but limited information on differences between communities. Under this uncertainties, they might filter out some notorious areas and choose safer areas. This pushes up the demand for houses at safer communities. Thus, the market equilibrium rents would increase. The hypothesis to test is whether the neighborhoods of high Airbnb rents are the ones with low crime rates.

The data set is "AirBnB Chicago 2015" available at GeoDa Data and Lab. It contains 20 covariates on Airbnb spots, socioeconomic indicators, and crimes of 77 community areas in Chicago, IL. In particular, the Airbnb part of the data includes response rate, acceptance rate, review rating, price per included guest, room type, number of Airbnb spots. "CRIMEPP" is the crime numbers per capita within each neighborhood.

**Exploratory Data Analysis**

The three continuous variables selected for this exploratory data analysis are "price_pp" (price per person),  "income_pc" (per capita income), and "CRIMEPP" (crime rate).

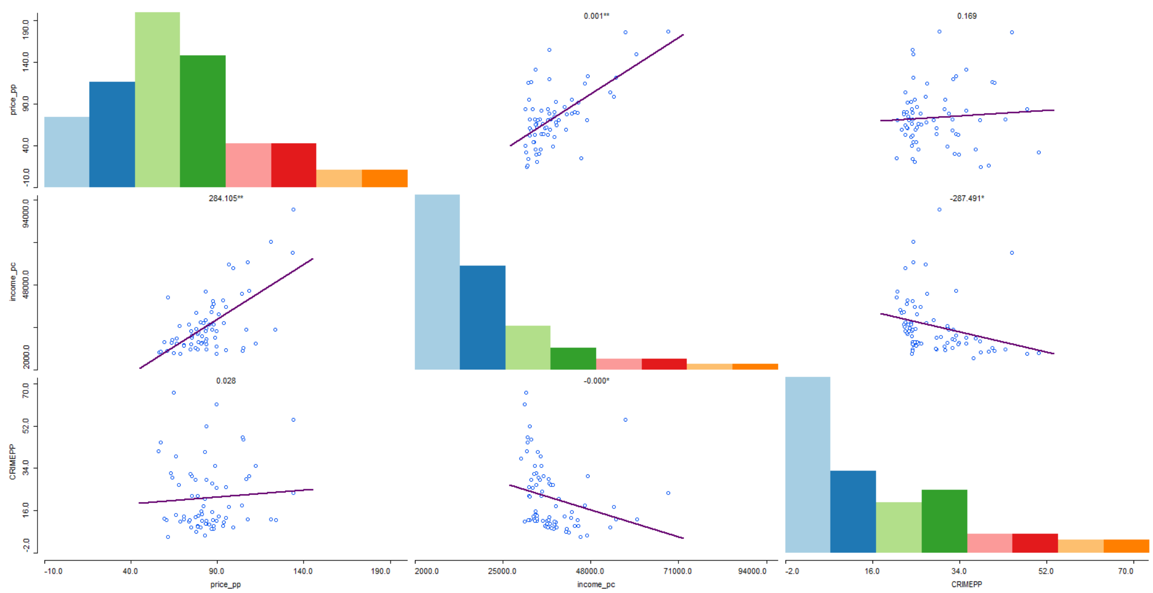


Figure 1. Scatter plot Matrix of "price_pp", "income_pc", and "CRIMEPP"

The scatter plot matrix reveals a strong significant and positive relationship between per capita income and Airbnb rents. This means wealthier neighborhoods are on average more costly to stay. The relationship between crimes rate and Airbnb rents is insignificant. This is surprising and will be explored spatially in this project.

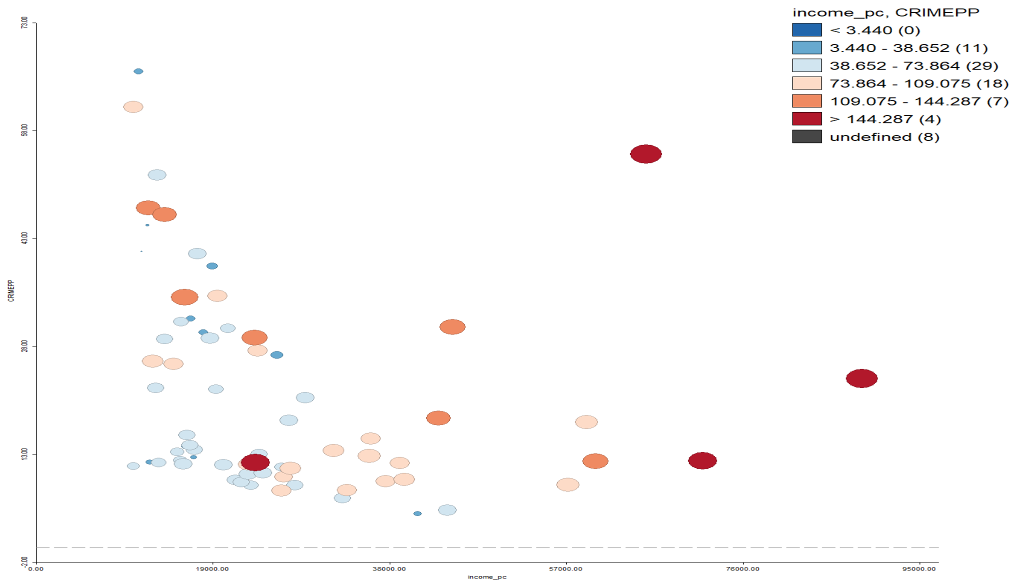



Figure 2. Bubble Chart among “price_pp”, “income_pc”, and “CRIMEPP”

From the bubble chart, we observe that the higher prices(larger bubbles) are situated in the bottom right corner of the scatter plot. Communities with high Airbnb rents are associated with low crime rates and high per capita income. One outlier at the upright corner is the Loop area, as the high crime rate is probably caused by the relatively small amount of residents.

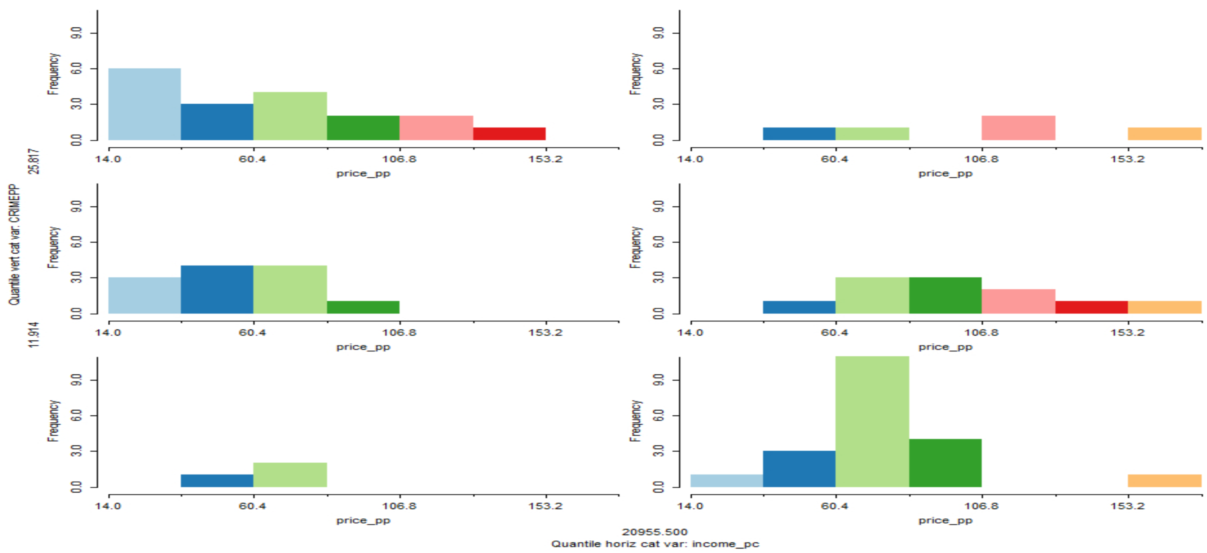



Figure 3. Conditional Histogram of “price_pp” on “income_pc” and “CRIMEPP”

Furthermore, the conditional histogram illustrates that the communities with below-median income per capita are associated with lower Airbnb rents as well. The “nice” communities (high income and low crime rate) don’t necessarily have the most expensive rates. The most costly rents are in the neighborhoods with above-median income per capita. The rents of neighborhoods with high crime rates seem to be distributed evenly at different ranges.

**Geo-visualization**

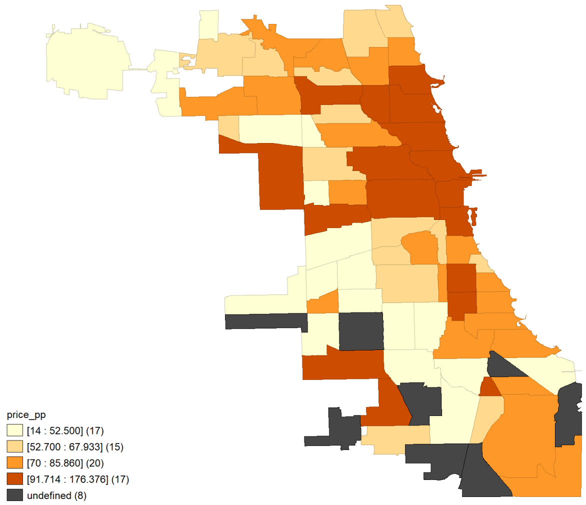



Figure 4. Quantile Map for Price

The quantile map shows the distribution of Airbnb rents across the city is not spatially random. Northern sides seem to be more expensive than the southern. All the 8 missing values are southern neighborhoods, probably because there are no Airbnb spots there.

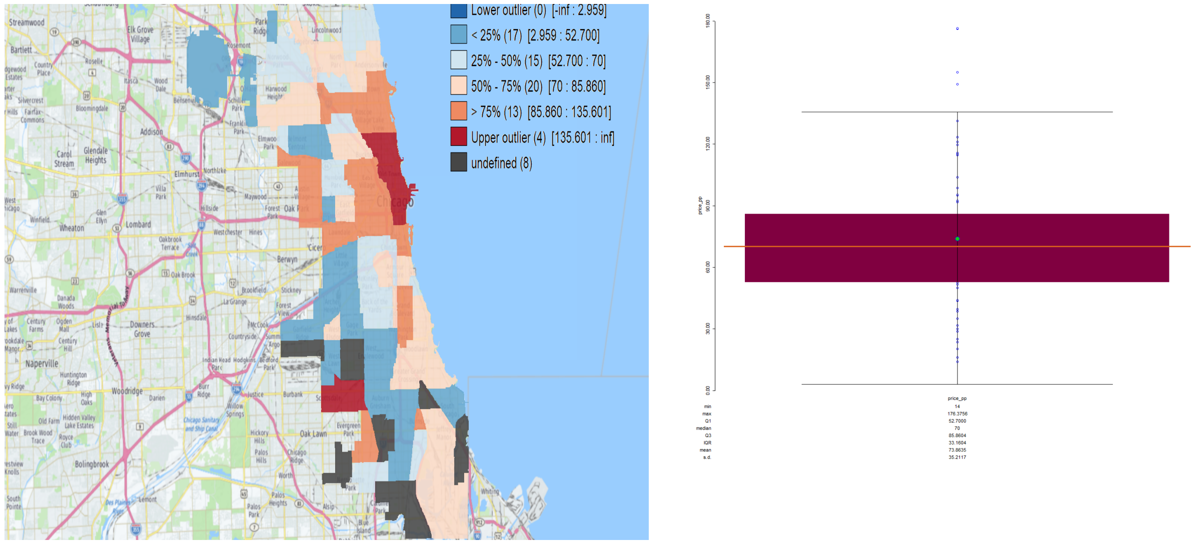



Figure 5. Box Map and Boxplot for Price

Box map and boxplot show clearly that the Loop has the most expensive rate and its surrounding neighborhoods have high rates as well. The lower outliers (cheapest) are the areas close to Chicago boundary. There is one upper outlier in the bottom left part which seems to be strange as the rents are much larger than the surroundings. This neighborhood is Ashburn where the price is 155 with only 2 Airbnb spots.

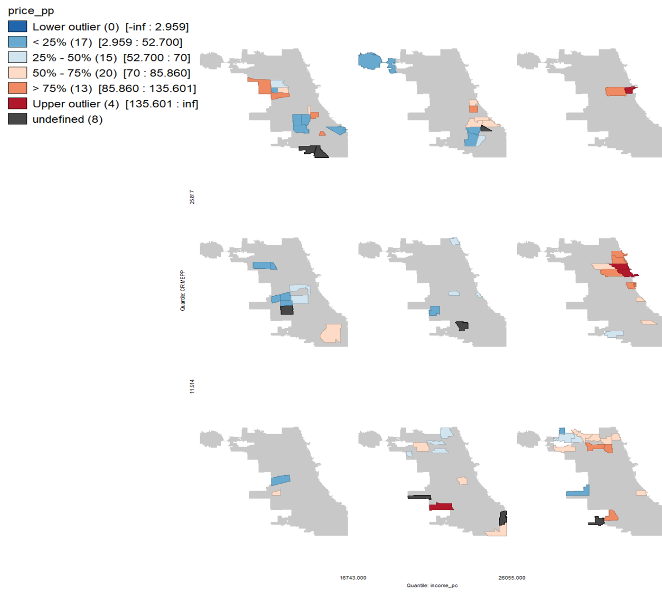
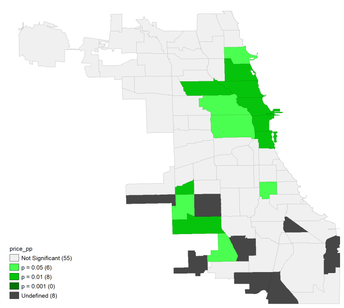
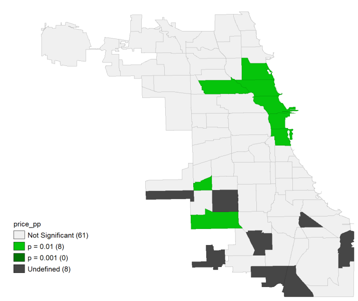
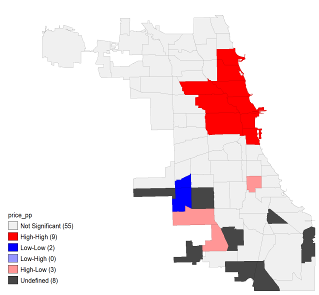
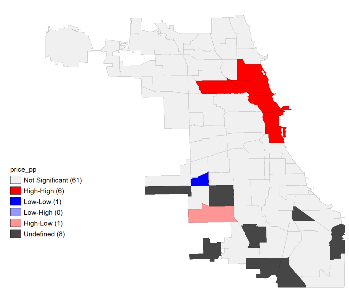



Figure 6. Conditional Map for Airbnb rents using “Income_pc” and “CRIMEPP”

The conditional map suggests that the “nice” neighborhoods (low crime rate and high average income) are mostly in the northern part (Hyde Park is included as well) and the “bad” ones (high crime rate and low average income) are either in the upper left or the south.

**Local Spatial Autocorrelation**

Local spatial autocorrelation enables us to identify the location of spatial clusters and outliers. Such clusters for Airbnb rents and crime rates are locations with significant positive local spatial autocorrelations. Only the crime rates variable is selected as the representative of social-economic status because it is the major factor that tourists in Chicago would concern about. The two spatial weights used to impose structure are “Distance Band Weight” and “K-Nearest Neighbor Weight”. “Distance Band Weight” using Euclidean distance implies distance decay between centroids and “K-Nearest Neighbor Weight” keeps the same number of neighbors for all observations. Two test statistics are “Local Moran’s I” and “Local Geary”. The two columns refer to the maps associated with significant levels at p=0.05 and p=0.01.

<table>
<tr><td>Airbnb Rents<br><br>Weight:<br><br>Distance band</td><td>Significance level: 0.05</td><td>Significance level: 0.01</td></tr>
<tr><td>Significance Map</td><td>
</td><td>
</td></tr>
<tr><td>Cluster Map</td><td>
</td><td>
</td></tr>
</table>

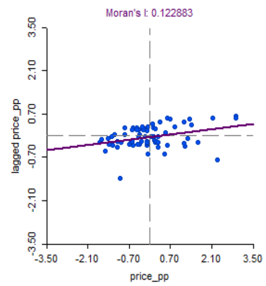
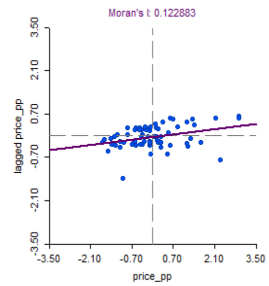
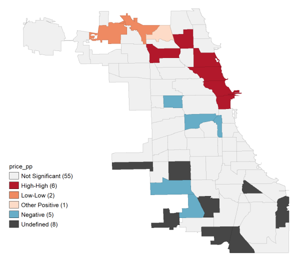
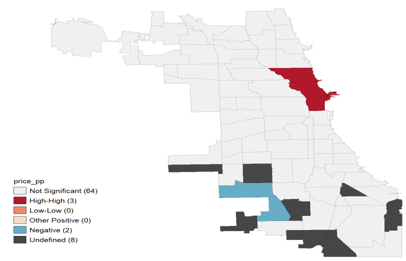
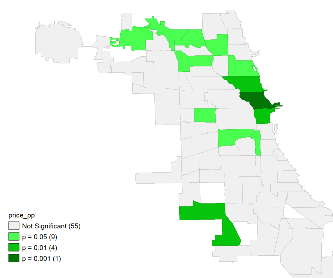
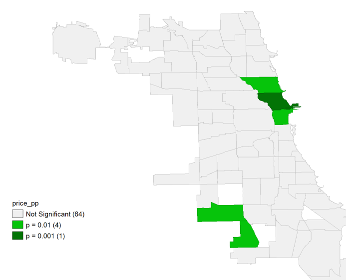
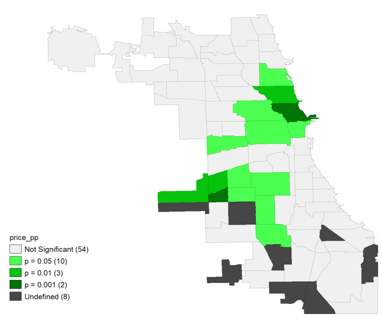
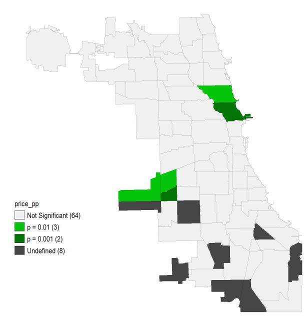
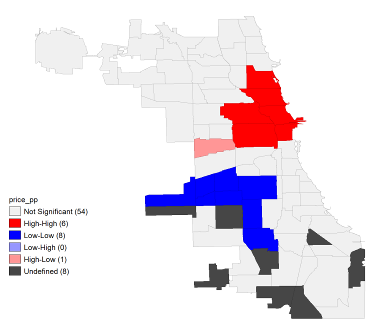
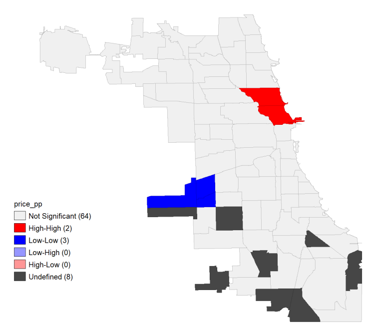
<table>
<tr>
<td>Moran’s I</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Cluster Map with Local Geary</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Significance Map with Local Geary</td>
<td>
</td>
<td>
</td>
</tr>
</table>

Table 1. Local Spatial Autocorrelation for Airbnb Rents with Distance Band weight

The significance map of Airbnb rents shows that about 70% (55/77) of the observations’ p-values are not significant. Only 18% (14/77) of them are significant at 0.05 level, and 10% (8/77) of them are significant at 0.01 level. Most of the observations with small p-values are high-high clusters in the downtown area (9 urban-area neighborhoods). There are three high-low spatial outliers with two on the border. The clusters identified using Local Geary are mostly matched with the results from local Moran’s I, but with three fewer high-high clusters. When resetting the p-value to 0.01, the number of clusters in cluster map and significance map with Local Geary statistic decreases. This suggests the downtown surrounding neighborhoods are sensitive to the change of significance level, but the downtown neighborhoods is a stable cluster.

| **Airbnb Rents**<br>**Weight:**<br>**K-Nearest neighbor** | Significance level: 0.05 | Significance level: 0.01 |
|---|---|---|
| Significance Map |  |  |
| Cluster Map |  |  |

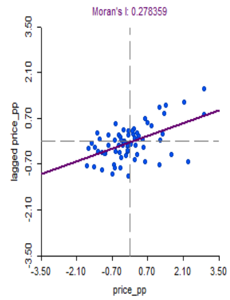
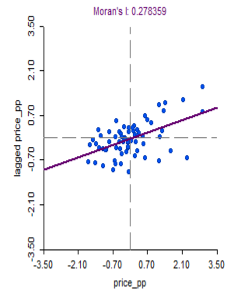
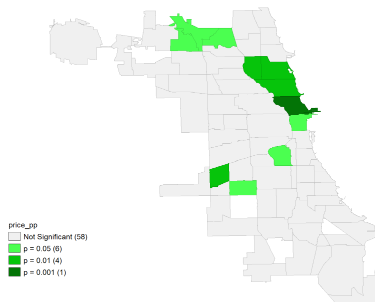
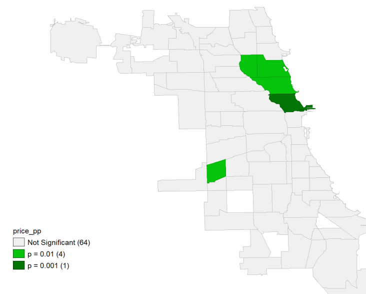
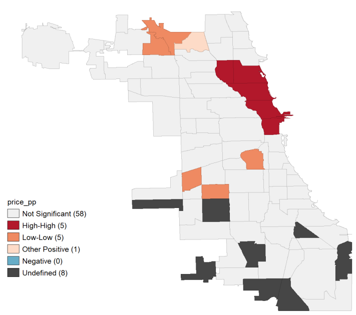
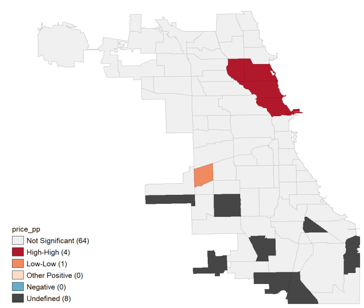
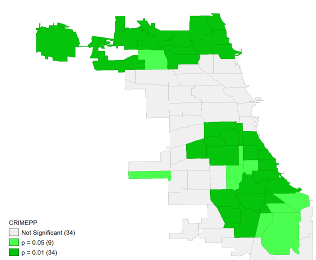
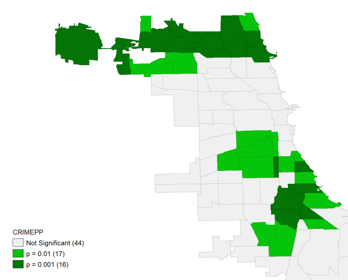
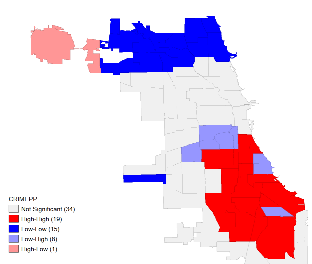
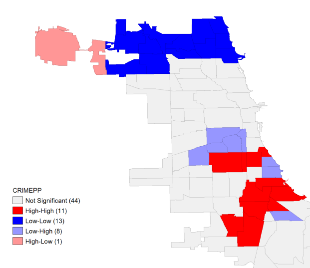
<table>
<tr>
<td>Moran’s I</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Significance Map<br><br>with Local Geary</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Cluster Map with<br><br>Local Geary</td>
<td>
</td>
<td>
</td>
</tr>
</table>

Table 2. Local Spatial Autocorrelation for Airbnb Rents with K-nearest neighbor weight

When using K-nearest neighbor weight, more observations become significant at level 0.05 in the significant map. When using local Geary statistic, more neighborhoods’ levels become insignificant. There are only two high-high spatial clusters and three observations significant at the p=0.01 level. Maps with Local Geary show two more high-high spatial clusters. This further illustrates the existence of a strong and stable cluster around the downtown neighborhoods for Airbnb rents.

Next, we apply the same methods (significance map, cluster map, Moran's I, significance map and cluster map with Local Geary), two spatial weights (Distance Band and K-nearest neighborhoods) and sensitivity analysis to the crime rate variable.

| **Crime Rates**<br>**Weight:**<br>**Distance band** | Significance level: 0.05 | Significance level: 0.01 |
|---|---|---|
| Significance Map |  |  |
| Cluster Map |  |  |

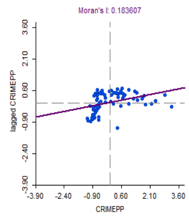
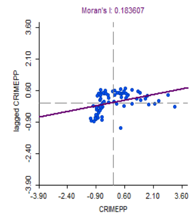
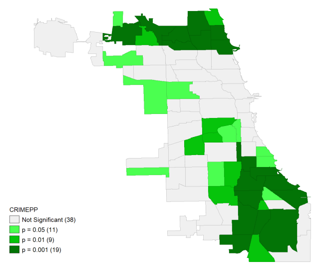
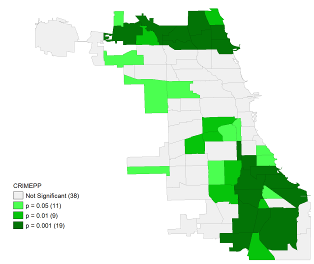
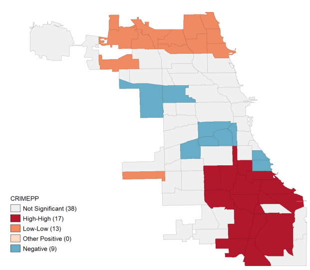
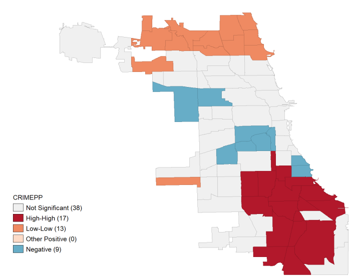
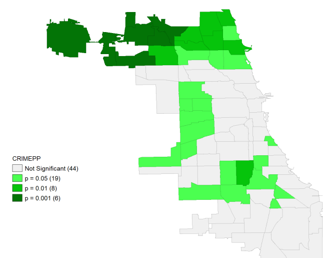
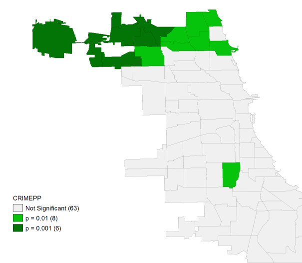
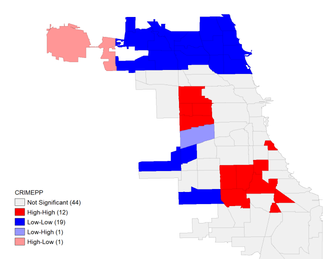
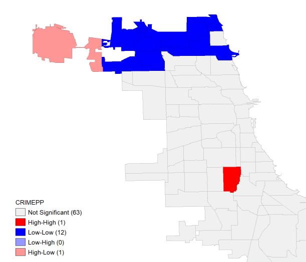
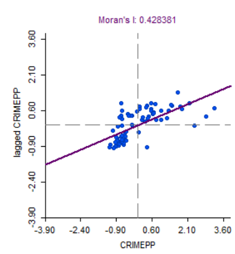
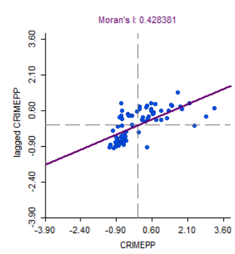
| | | |
|---|---|---|
| Moran's I |  |  |
| Significance Map with Local Geary |  |  |
| Cluster Map with Local Geary |  |  |

Table 3. Local Spatial Autocorrelation for Crime Rates with Distance Band weight

The cluster map shows that the high-high spatial clusters of crime rates are all in the southern side. None of the downtown areas is classified as spatial clusters and statistically significant. Local Geary's maps show similar results, with 4 more insignificant observations.

When restricting p-value to 0.01, 6 more observations become insignificant. The significance map and cluster map with Local Geary are the same as the ones at p=0.05. This shows these spatial clusters and outliers are insensitive to the change of significance level. The high-high spatial clusters are located on the southern side and low-low clusters are in the northern side. None of the downtown neighborhoods are classified as spatial clusters.

<table>
<tr><td>Crime Rates<br>Weight:<br>K-Nearest Neighbor</td><td>Significance level: 0.05</td><td>Significance level: 0.01</td></tr>
<tr><td>Significance Map</td><td>
</td><td>
</td></tr>
<tr><td>Cluster Map</td><td>
</td><td>
</td></tr>
<tr><td>Moran's I</td><td>
</td><td>
</td></tr>
</table>

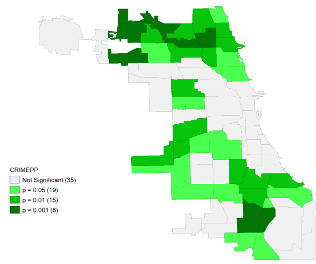
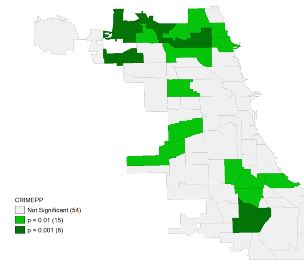
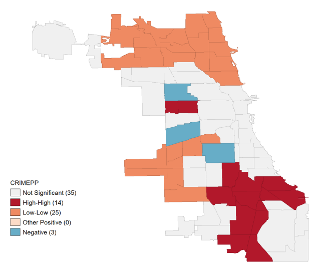
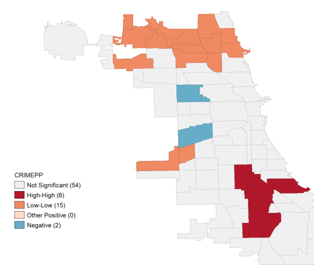
<table>
<tr>
<td>Significance Map with Local Geary</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>Cluster Map with Local Geary</td>
<td>
</td>
<td>
</td>
</tr>
</table>

Table 4. Local Spatial Autocorrelation for Crime Rates with K-nearest neighbor weight

When using K-nearest neighborhoods as spatial weights, the low-low spatial cluster is on the northern border. Maps with Local Geary show slightly greater dissimilarities between neighbors. When setting the p-value to 0.01, there is one high-high spatial cluster in the near south side. Most of the northern-side observations still are low-low spatial clusters and significant at $p<0.01$.

**Conclusion**

This project shows the neighborhoods of high Airbnb rents are spatially different from the ones with low crime rates. The correlation between crimes rate and Airbnb rents is weak. By deploying local spatial autocorrelation analysis and sensitivity analysis, we find that the downtown neighborhoods have the highest Airbnb rents, the northern neighborhoods are the safest and the southern neighborhoods are the most dangerous. This result implies that location convenience to tourist attractions might be the major factor influencing market demand for Airbnb rentals. Tourists might prefer to choose the downtown Airbnb spots as they also have access to more restaurants and entertainments during the short stay.