

# Segmenting Neighborhoods in Chicago, Illinois

Wendy Hu

## Introduction/Business Problem

Chicago, on Lake Michigan in Illinois, is the third largest city in the U.S and one of famous international financial centers in the world. The city is also renowned for its museums, including the Art Institute of Chicago with its noted Impressionist and Post-Impressionist works. According to 2018 Time Out City Life Index, Chicago was ranked first in the global quality of life survey of 15,000 people in 32 cities.

So, despite Chicago is famous for crime capital, there are many people considering to settle in this city. When they choose the community and apartment they will live in, safety and neighborhood facility can be top determinant factors. For people who consider to moving in this city, they are willing to read reports in answering the following questions:

1. What community is the safest in Chicago?
2. Is crime rate connected with average income and education?
3. Using Foursquare data, how different are venues distributed in neighborhood of different crime rate?

## Data

Chicago Socioeconomic Indicators:

<https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv>

This dataset contains a selection of six socioeconomic indicators of public health significance and a “hardship index,” by Chicago community area, for the years 2008 – 2012. The indicators are the percent of occupied housing units with more than one person per room (i.e., crowded housing); the percent of households living below the federal poverty level; the percent of persons in the labor force over the age of 16 years that are unemployed; the percent of persons over the age of 25 years without a high school diploma; the percent of the population under 18 or over 64 years of age (i.e., dependency); and per capita income.

Chicago Public Schools:

<https://ibm.box.com/shared/static/f9gijv1gimxxzycdhplzt01qtz0s7ew7.csv>

This dataset shows all school level performance data used to create CPS School Report Cards for the 2011-2012 school year.

Chicago Crime Data:

<https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv>

This dataset 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

Venue Data:

<https://foursquare.com/>

Venues data can be extracted from this website. We can utilize this website to analysis what venues distributed around different neighborhoods.

## Methodology

1. Load all the datasets using pandas package
2. Explore datasets using bar plots
3. Test if crime rate correlate with income and education level using scatter plot and regression analysis
4. Check what is the most common venue in the communities with highest and lowest crime rate
5. Analyze if crime rate has impact on facility distribution, such as coffee shop and restaurants.

## Visualization of Data

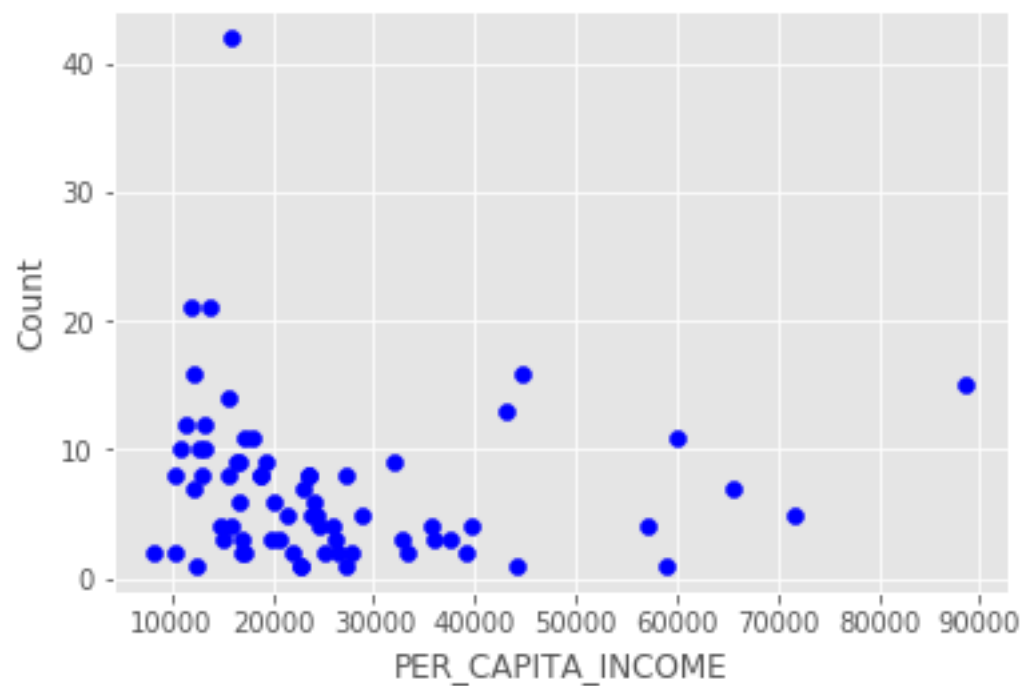
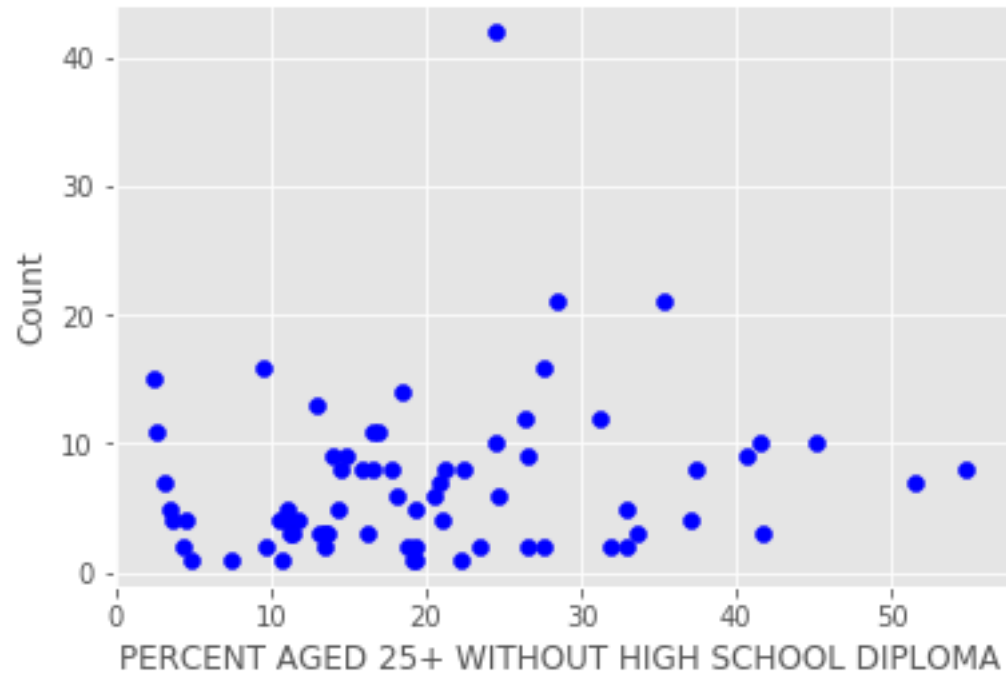
After loading data, bar plot and scatter plot are created to view the crime count and also its relationship with education level and income. The region of greatest crime count is found to be Austin which is on the far west side of Chicago. It's a really big area, stretching all the way from North Avenue to Roosevelt Road. It has a pretty high crime rate. The safest regions include burnside and Bridgeport.

## Results

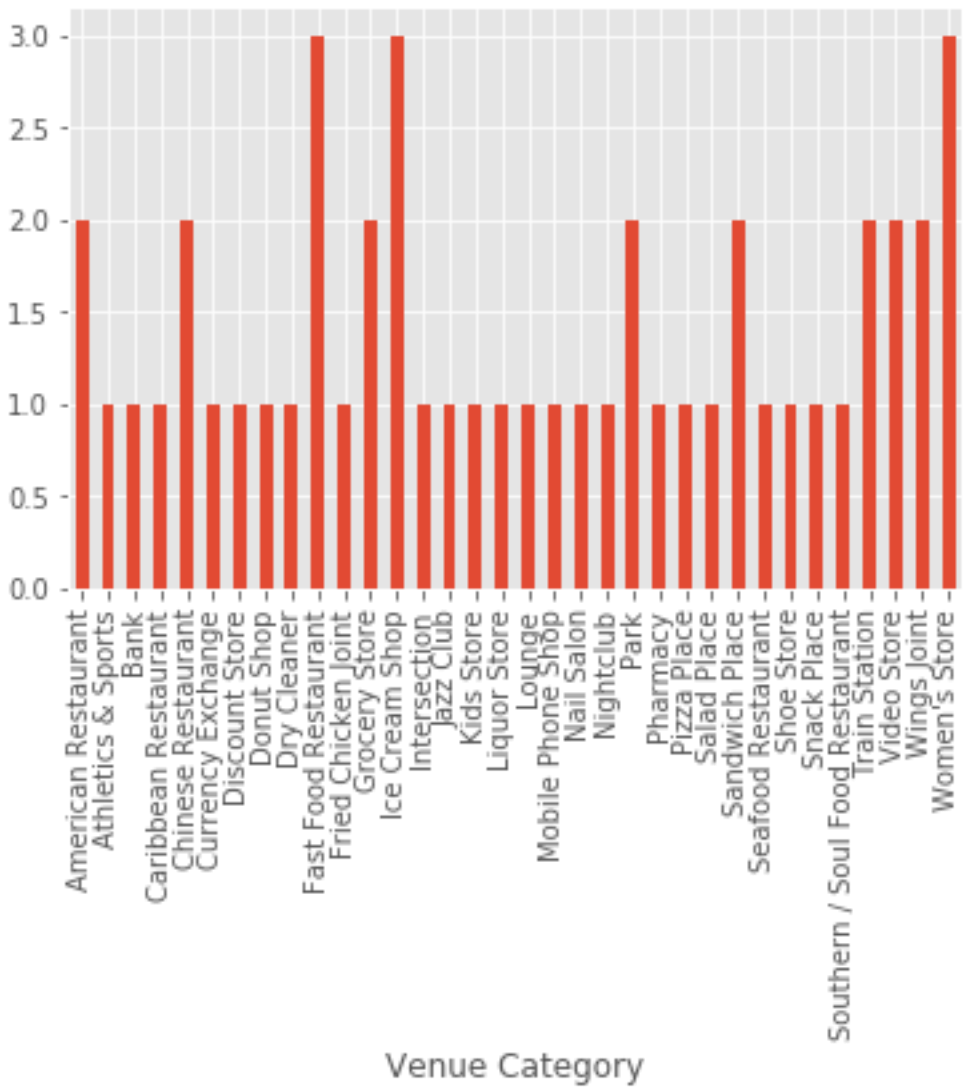
The analysis enabled us to discover and describe visually and quantitatively. According to summary data, we can know that the most dangerous region is Austin and the safest regions include burnside and Bridgeport.

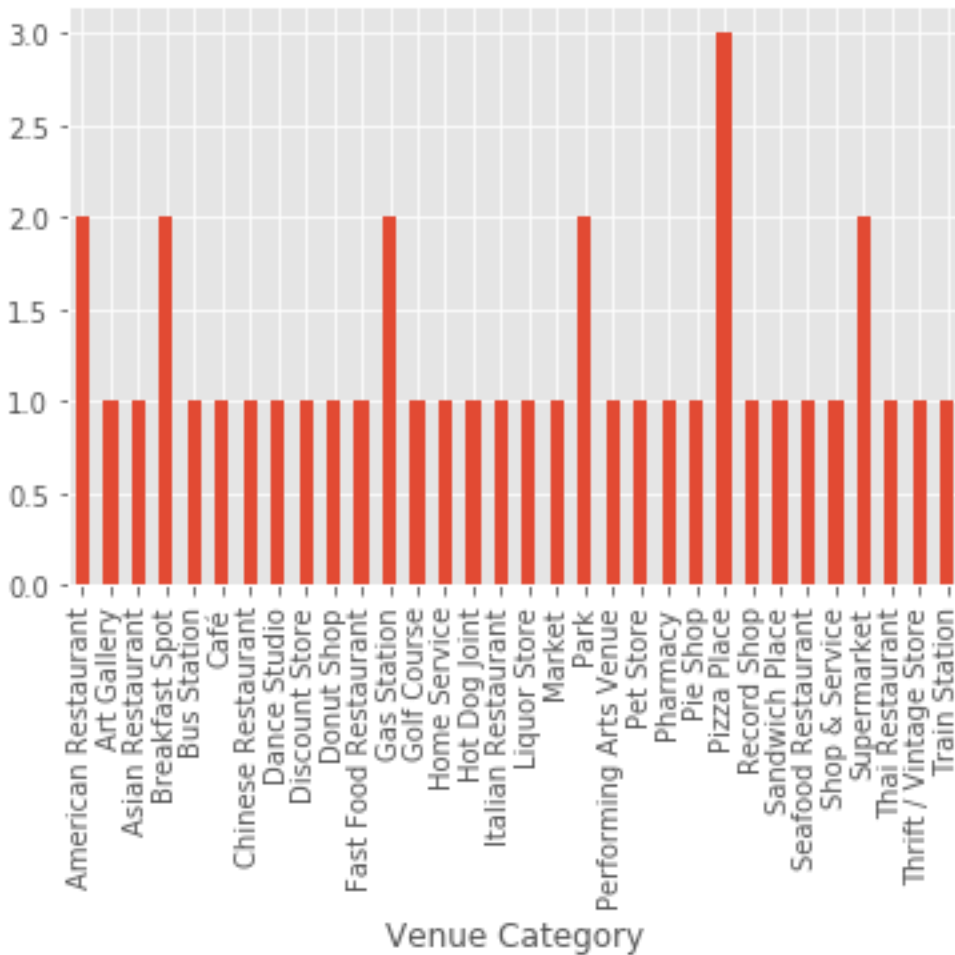
There is no significant relationship between crime rate with education level, which is against my original assumption. However, the independent variable used in the scatter plot is the percentage of PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA. It may be needed to check other variables like the percentage with bachelor.

There is significant decreasing relationship between crime rate with income. From scatter plot, we can see crime rate decreased along with increasing incomes.



we were able to determine the top common venues within a 1000m radius of the centroid of the highest crime community. The most common venues in the highest crime community are pizza place. The most common venues in the lowest crime community are fast food restaurant and ice cream shop.





## Discussion

These analysis enables us to gain an understanding of the crime count by communities. When consider moving to Chicago, it is obvious to avoid Austin which has the highest crime rate. It also has much higher crime count than other communities. However, there are many factors to consider when moving, such as working distance, house price. It could be interesting to analysis these factors with crime rate together.

The most common venues in Austin and Burnside are different. They are pizza place and fast food restaurant respectively. It is convenient to live in both communities. It is surprised that ice cream shop is one of common venues in Austin. It may beneficial to open an ice cream shop in Burnside.

## Conclusion

Overall, Chicago is a great city to move in since all communities have good life facility. Venues are quite similar in different communities of crime rate. And it is necessary to consider crime rate. Communities differ significantly in crime rate.