

The Investigation of Crime Spread in Chicago City

Principal Investigator: Zhe Wang (zwang752@wisc.edu)

Introduction:

Chicago is one of the major cities with the highest number of crime cases every year in the United States. While the tourism business in Chicago enlarges quickly, it is useful to realize the crime ratio reality in different areas of Chicago. The ultimate crime ratio statistics can inform those tourists where the dangerous zones are, so largely increase the safety index of those people who are not familiar with this city. My target of this report is to demonstrate the spread of crime ratio in Chicago city in 2016 and clarify the frequency of crime behaviors in different period of time, to inform tourists when and where to go. In addition, I endeavored to discover the relationship between crime ratios and poverty level of each areas in Chicago city.

Description of the data and methodology:

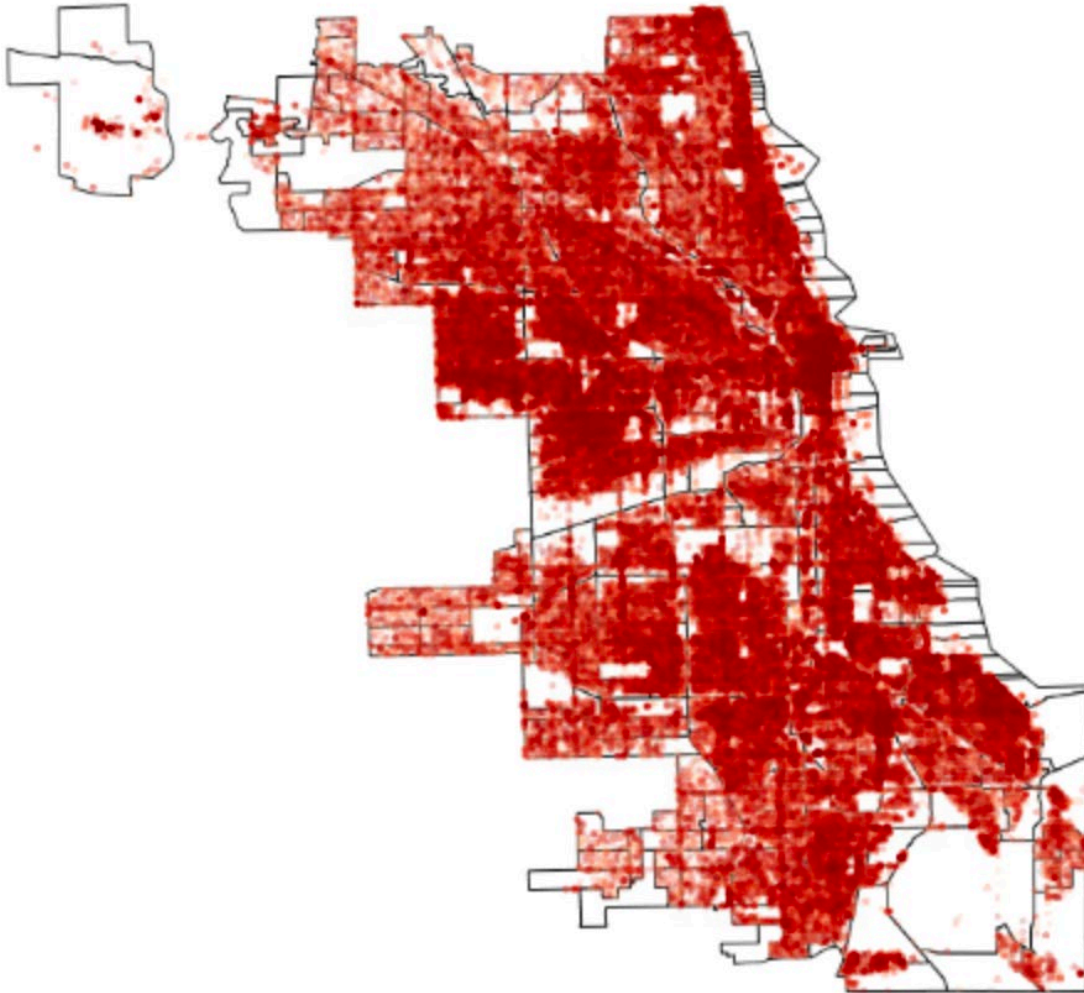
In this report I have used the census tracts and community areas profile to plot the boundaries of Chicago city. In census tracts there are geometric information coordinates of different tracts in it. However, in community area profile, the geometric information is separated by 77 different areas in Chicago city. In addition, I loaded the data of crime behaviors from year 2012 to 2017. In this report, I will mainly focus on 2016. Main information such as crime types (battery, weapon, etc.), crime time, crime locations and crime description are included. I also loaded the data of household information to bring the poverty level as a factor into my analysis. Some main variables are community area numbers, percentage households below poverty level and percentage crowd ratio.

Four main methods I have used to visualize data are maps, bar charts, scatter plot and line plot. Each method has their uniqueness in interpreting the valid information to readers, and I will make them as precise and accurate as possible to avoid misleading information.

Analysis:

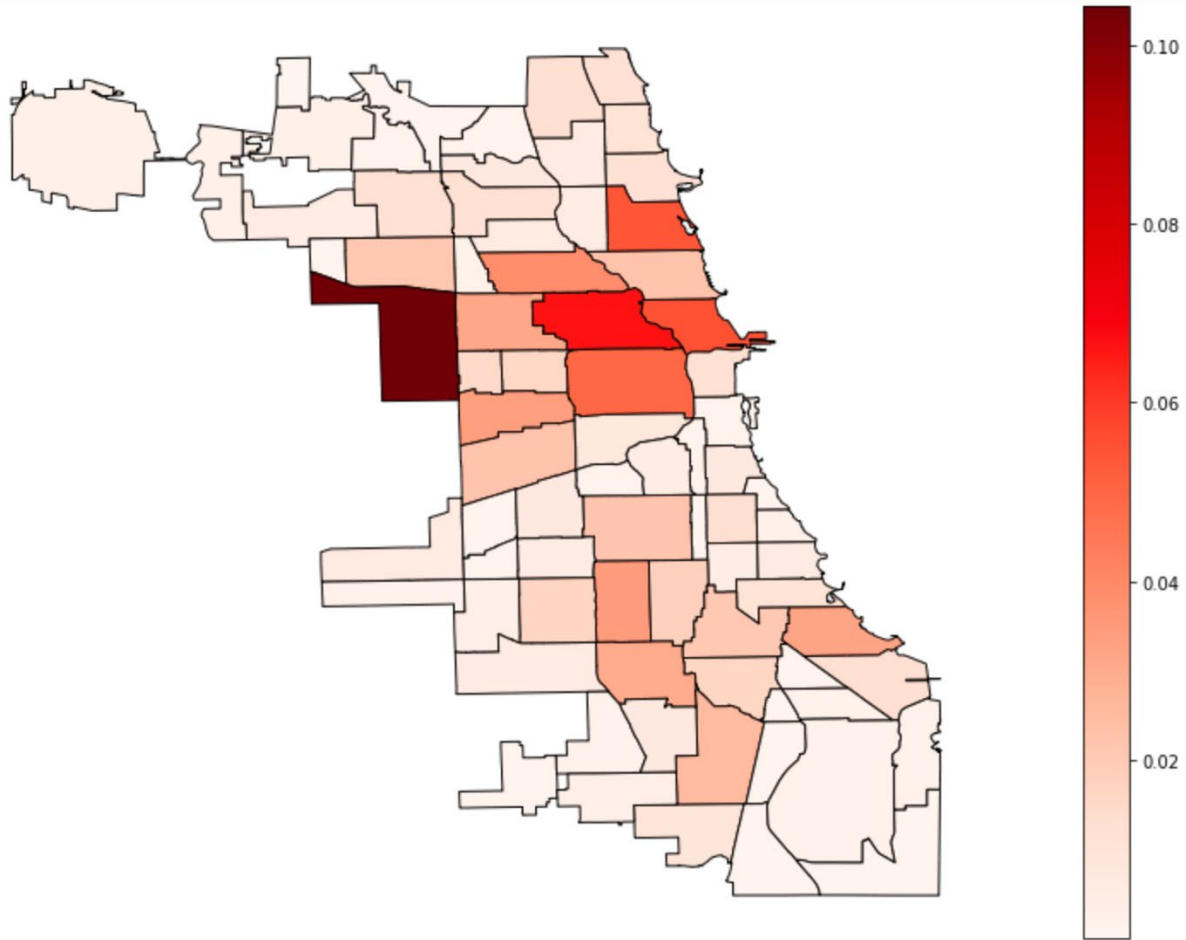
1. The spread of crimes in Chicago

At the beginning, I tried to merge the crime data and census tracts to plot Chicago maps and crime behaviors as points on the map. I used the longitude and latitude variables in crime data to create point objects and plot those points on the map plotted by geometric census tracts. Here is the result:



This plot does not look good. Too many information and point objects are squeezed together, so the graph looks messy. We can only observe a weak relationship even if I have ameliorated this graph several times. Therefore, the using point object to pinpoint the crime location in Chicago map is not a valid method to inform readers crime information.

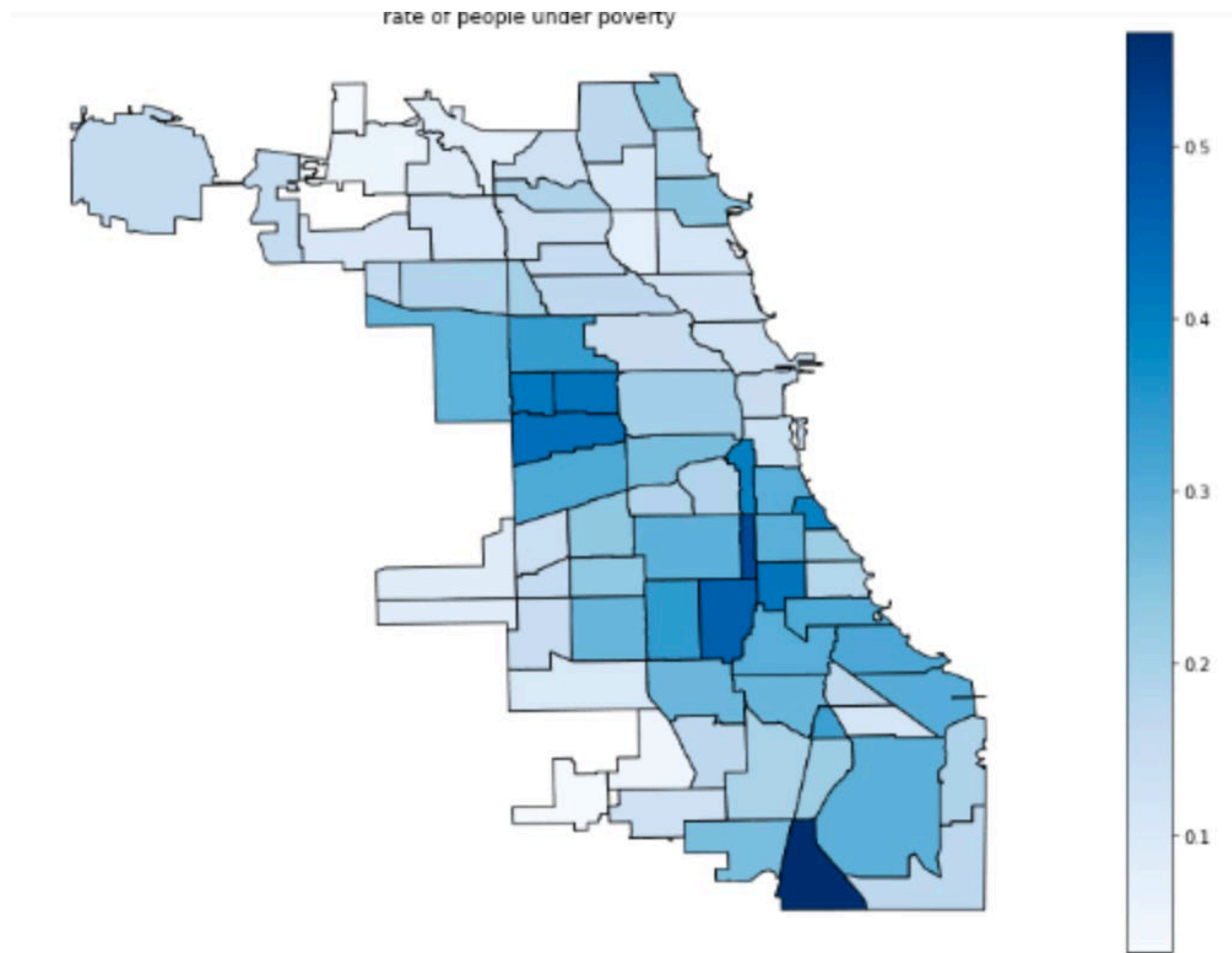
Hence, I tried another method: merge the community area profile with the crime data. This time I used community areas to separate Chicago city into 77 areas and plot the crime ratio in each of them to the overall crimes in Chicago. And I choose to use different dark level of red to demonstrate the visual difference. Here is the result:



In the map we can clearly observe the crime patterns. In some areas in the high middle and south, high crime ratios exist. The darker the red color in one community area tells a higher number of crime cases happened in this area in 2016 compared to other areas in Chicago.

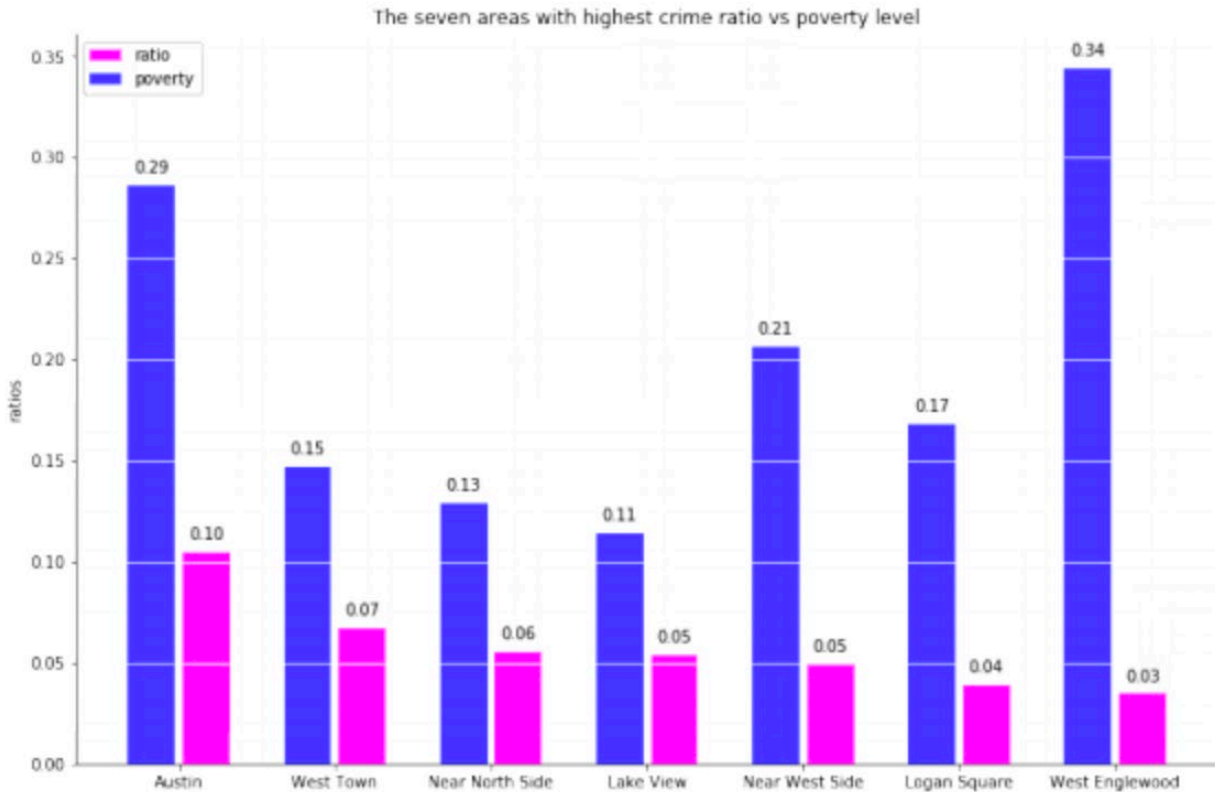
2. The relationship between crime ratio and poverty level in Chicago

My intuitive guess of the reason of this crime spread is the different poverty levels of these areas. It is reasonable to believe that poor people have more incentive to become criminals due to their life pressure. Therefore, I merged the household information with the community area profile, and plotted a map demonstrate poverty level in different areas in Chicago city. Again, I used different level of darkness in color blue to demonstrate different level of poverty.

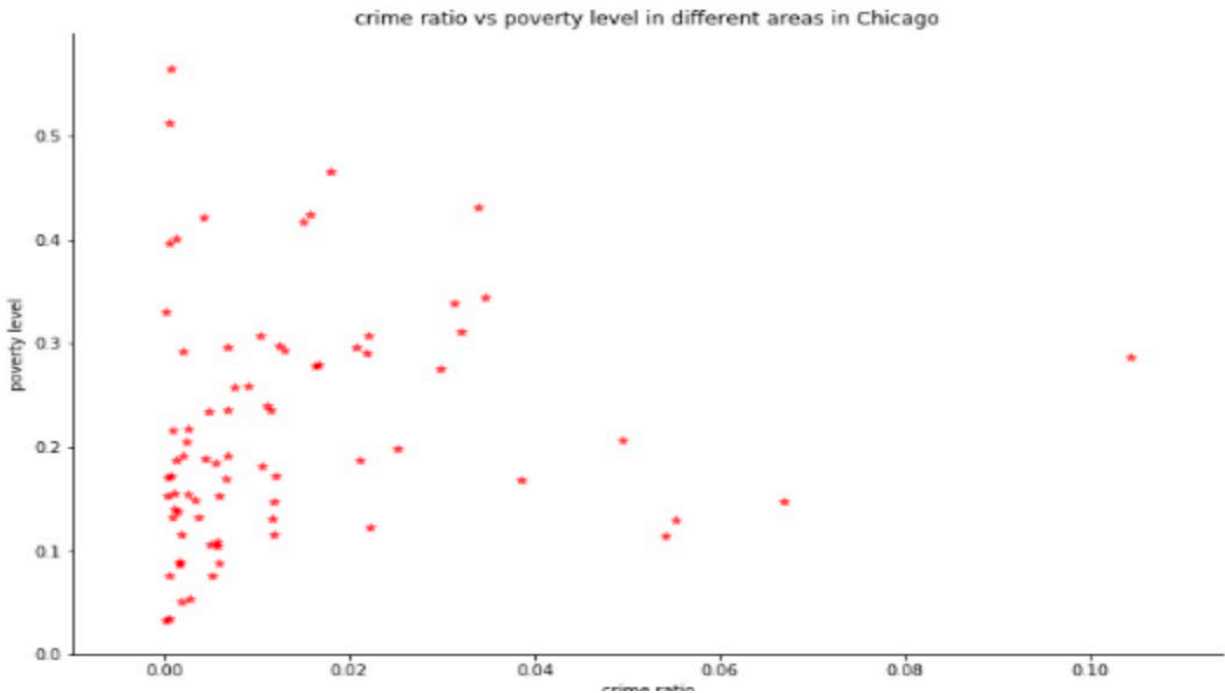


In the poverty level map plot, the darker of the blue color in an area means the poorer the area was. While we compare this plot with the crime spread plot previously by bare eyes, we can observe a weak connection between the color darkness in different areas. A weak connection exists that areas with higher poverty level, which has darker blue colors, generally contain higher crime ratio, represented by darker red. However, the result is not very clear, so we need more specific comparison to reach a conclusion.

So, I plotted a bar charts. I chose seven community areas with the highest crime ratio and make them rectangles on the plot. Moreover, I put the poverty level of each of them as another rectangle to the right of original rectangles. The comparison will be obvious: my target is to observe whether a higher crime ratio city is also poorer than the others. Here is the result:



The result is not expected. There exists a positive linear relationship at the 4 areas with the highest crime ratios, but the other three areas demonstrate the relationship adversely. To clarify the relationship and use all areas as samples, I plotted a scatter plot in which x axis represents the crime ratio, and y axis represents the poverty level. Since scatter plots are good tools to visualize a linear pattern between two factors, my target is to discover whether there exists a linear relationship between crime ratios and poverty level. If a positive relationship exists for these two factors, just like what I expected, then a higher level of ratio crime should also mean a higher level of poverty. Then in the graph, every point is expected to surround the line $Y = X$. Here is the result:



Unexpectedly, we actually cannot see a clear pattern of how the relationship presents between crime ratio and poverty level in this scatter plot.

There are several explanations:

- 1.Data is insufficient or inaccurate.
- 2.Factors which affect crime ratio deeper are not listed, such as population density, etc.
- 3.The observing time of our poverty level report data does not match with our crime ratio data collecting time of 2016.

So to clarify the key factors of high crime ratio in each area in Chicago, more samples are needed, and more factors should be taken into consideration.

3. The frequency spread of crime behaviors in a year

Finally, I plotted a line plots to visualize the frequency spread of crime behaviors in each month in 2016. I summed the number of cases in each month as y axis and use date time object as x axis to plot the result. Here is the result:



We can clearly observe that in July, the crime frequency reaches its peak at 24646 cases, but in February the frequency falls to its neap at 18590 cases. And in most summer time, the crime frequency is higher than its average in a year. So, while the weather is warm, criminals have more incentive to crime; if the weather is getting colder, criminal behaviors decrease quickly.

Conclusion:

1. Crime behavior reaches its peak at summer on August and fall to its neap on March. To be more specific, crime behaviors increases in the spring and decreases while winter comes. The average line demonstrate that the crime frequency is higher than its average in a year while the city enters summer and Autumn season, while the crime frequency is lower than its average when the city enters winter and spring season. My guess is that in winter, the opportunity cost to crime increases because of the weather getting colder, so less criminals are observed.
2. The relationship between poverty level and crime ratio is not clear in my current data. A weak positive relationship is demonstrated in the concentrated areas in the two maps. More data are needed to further analyze how crime ratio is affected by different factors.
3. In general, the best for tourism is winter and spring season, if we consider safety index as the top factor in our decision. And those areas in the crime spread map with darker red are not suggested to go for tourism.