Chicago Crime Analysis

wh951

Wei Hong

Background:

Crime is always a serious problem for big cities. Residences want to live in a safe block, city officials want to control the crime rate, the police want to stop crime effectively. Common sense tells us that the crime rate maybe relative to time, location any other indicators. It would be of immense value to everyone if we can create a crime model and predict the crime in the future.

Goals:

There are two aim of the project. First of all, is to analyze the crime statically and draw a city crime map by different categories. Secondly is to find a model for crime and try to guess a future criminal rate in different block at different time. The crime data is get from city of Chicago Data portal. The analyze program will be written in spark and use r or google map JavaScript to create a map. The fitting of the data will be created by using MATLAB

Raw Data Analysis





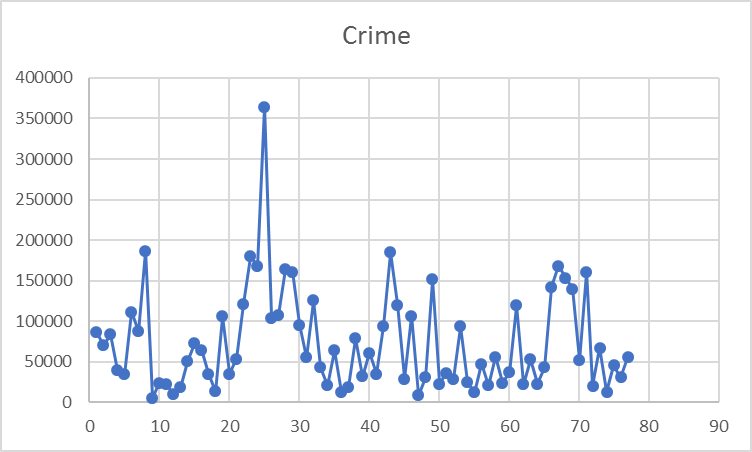
There are 17 fields in the raw data, and 6 of them would be used in the project. Case Number is a unique number for each case; Date is the time when the crime is committed; Block is the community; Primary Type is primary description of the crime report. Arrest tells if the criminal is arrested; Location has the latitude and longitude information of the accident.



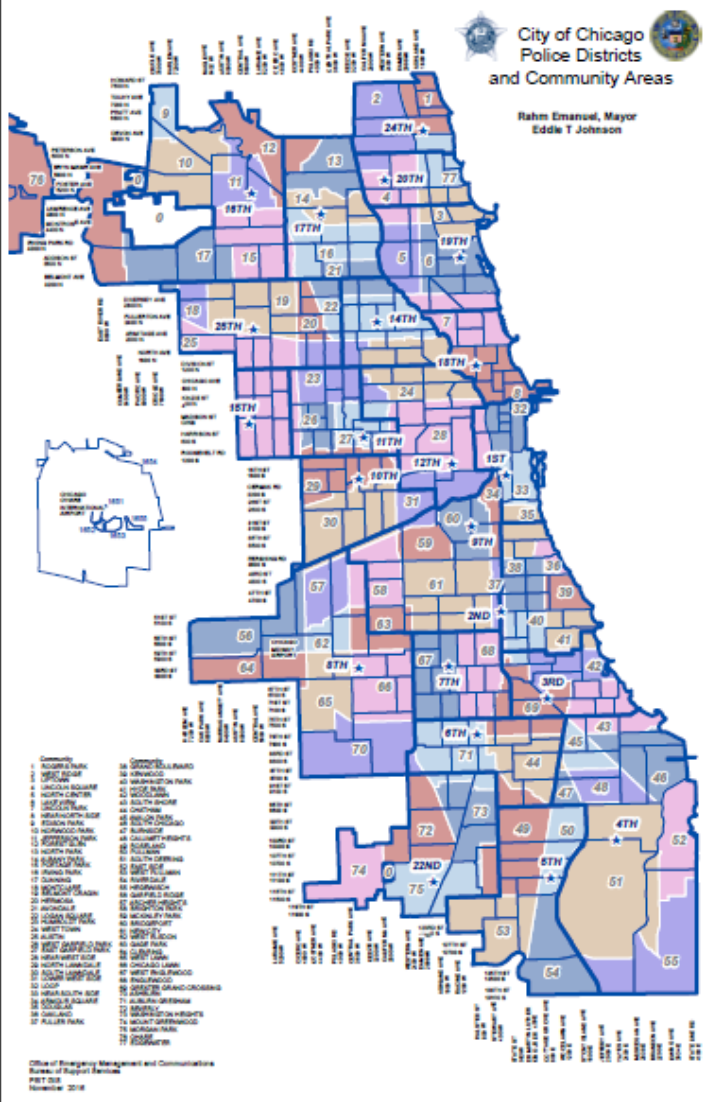
Analysis

There are four results of this project. A crime map generated by using google heat map API; The relationship of crimes and times; The relationship of crimes and community area and the correlation of different type crimes

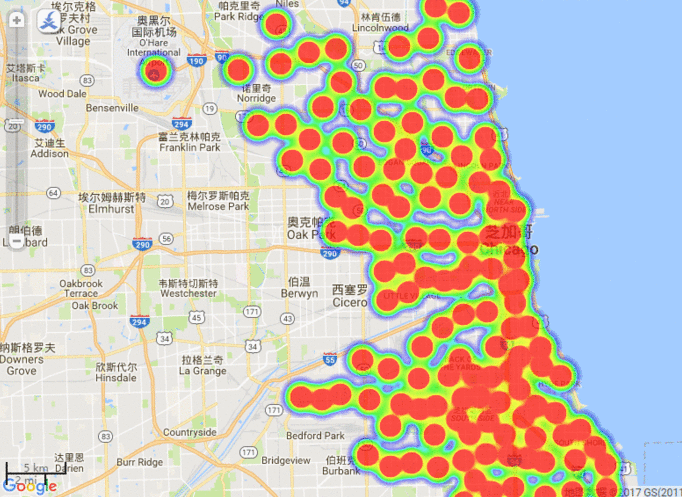
Crime map



The chart above is the crime number of different community area. The picture above is the community area map. For example we get community area 9-13 is a safer area to live than others according to charts.



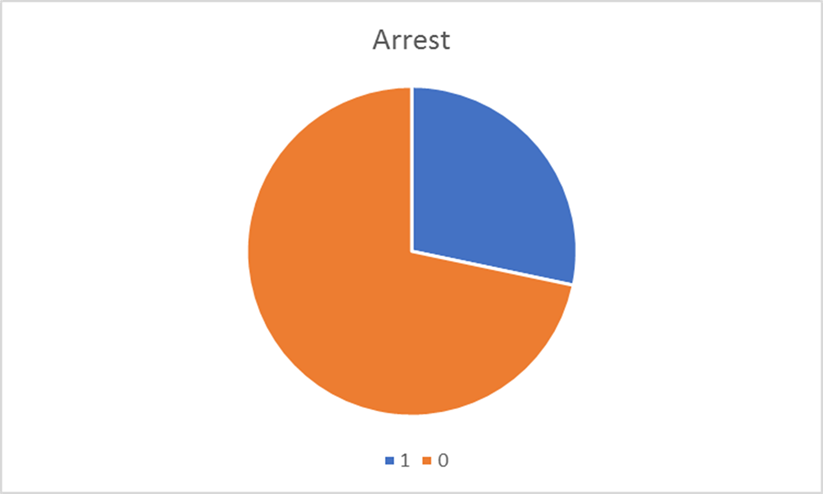
And I have used google heatmap api(<https://developers.google.com/maps/documentation/javascript/examples/layer-heatmap>) to generate a crime heat map for a better view



Clearly, we can see there are lots of crimes in Chicago and not many safe areas, which is a problem for most big cities. And that is also why we need to stop crime efficiently with limited police manpower that we have.

However, for city officials the crime map also is not enough. They want to know if the crime number is relative to something, so that they help to reduce it.

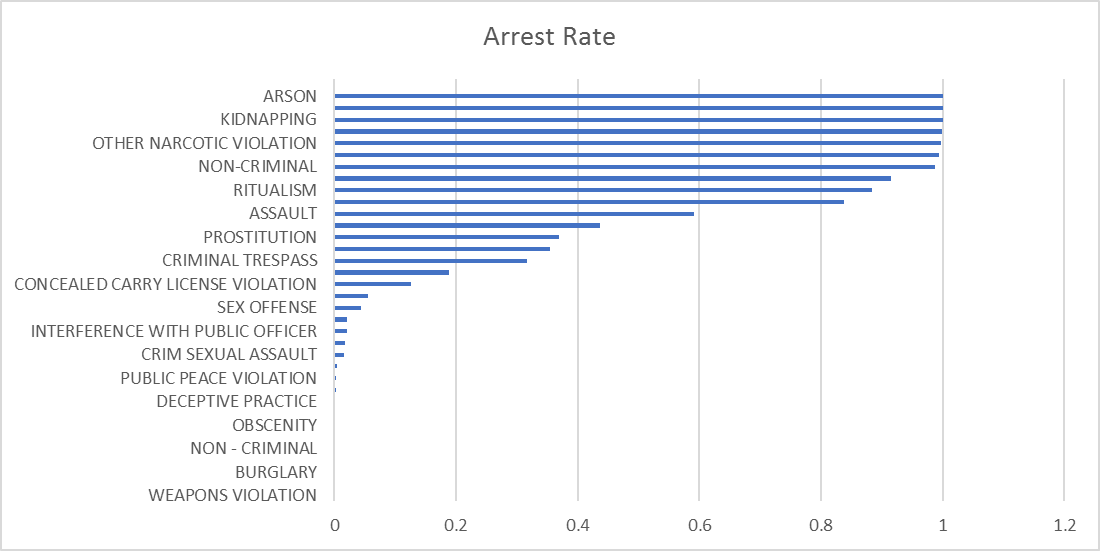
**Crime relationship analysis**



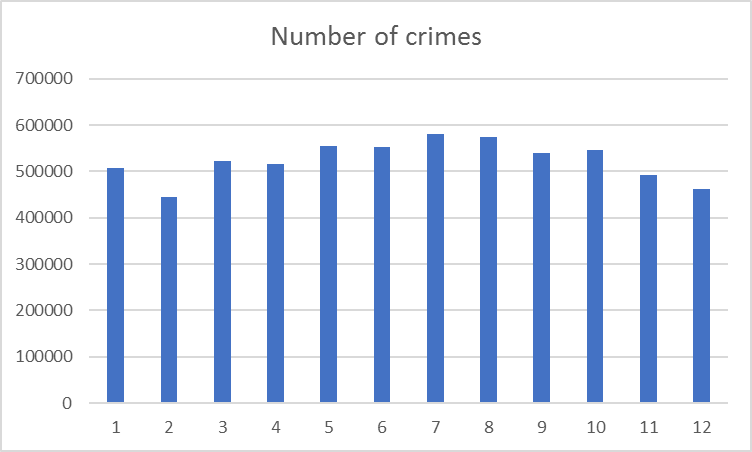
1. My first assumption of crime numbers is related to the arrested rate: if more criminals are arrested, the cost to commit the criminal also increase, which leads to less crime number.



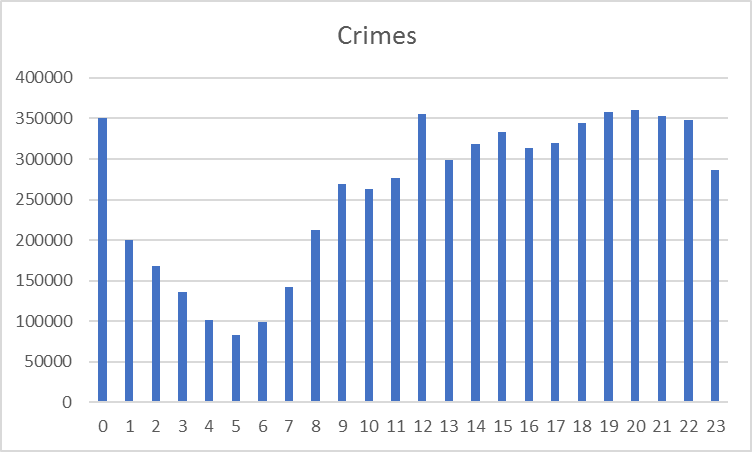
However, the result is disappointing. No any obvious relationship between arrested rate and crimes. The reason maybe that different type crimes have different arrest rate, unless a severe crime, the criminal would be arrested.



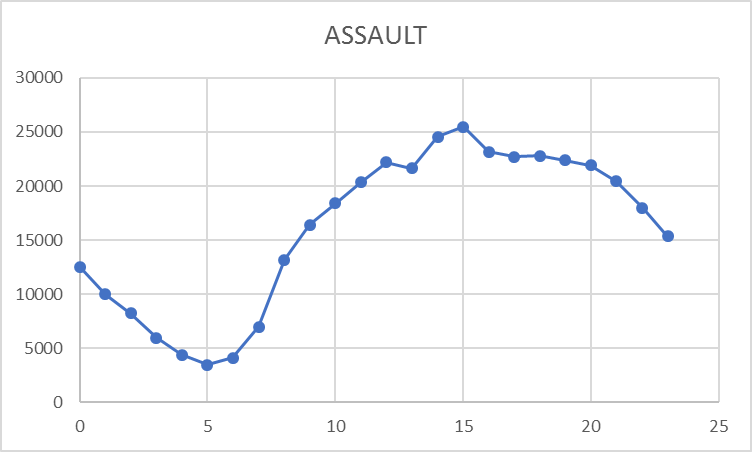
1. The second assumption is that number of crimes is related to the time.



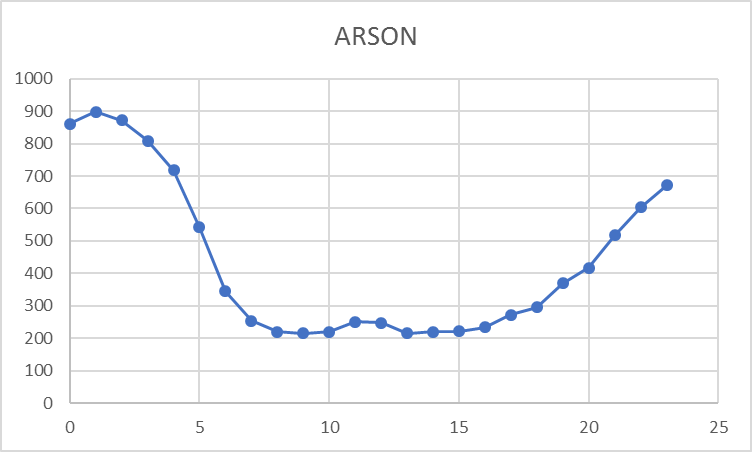
The chart above is number of crimes on different months. As we see, in winter there are a slightly less crimes and a little more in summer. The reasons I think is that people are less likely to go out in winter because it is cold outside.



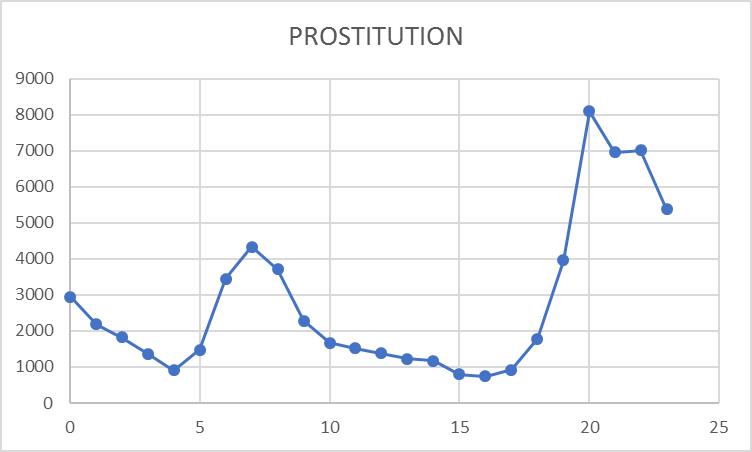
The next chart is the crime number vs the 24 hours in a day. As we can see 1 am to 8 am, there are much less crimes, which are usually sleep time. Criminals are also people. They also need to sleep.



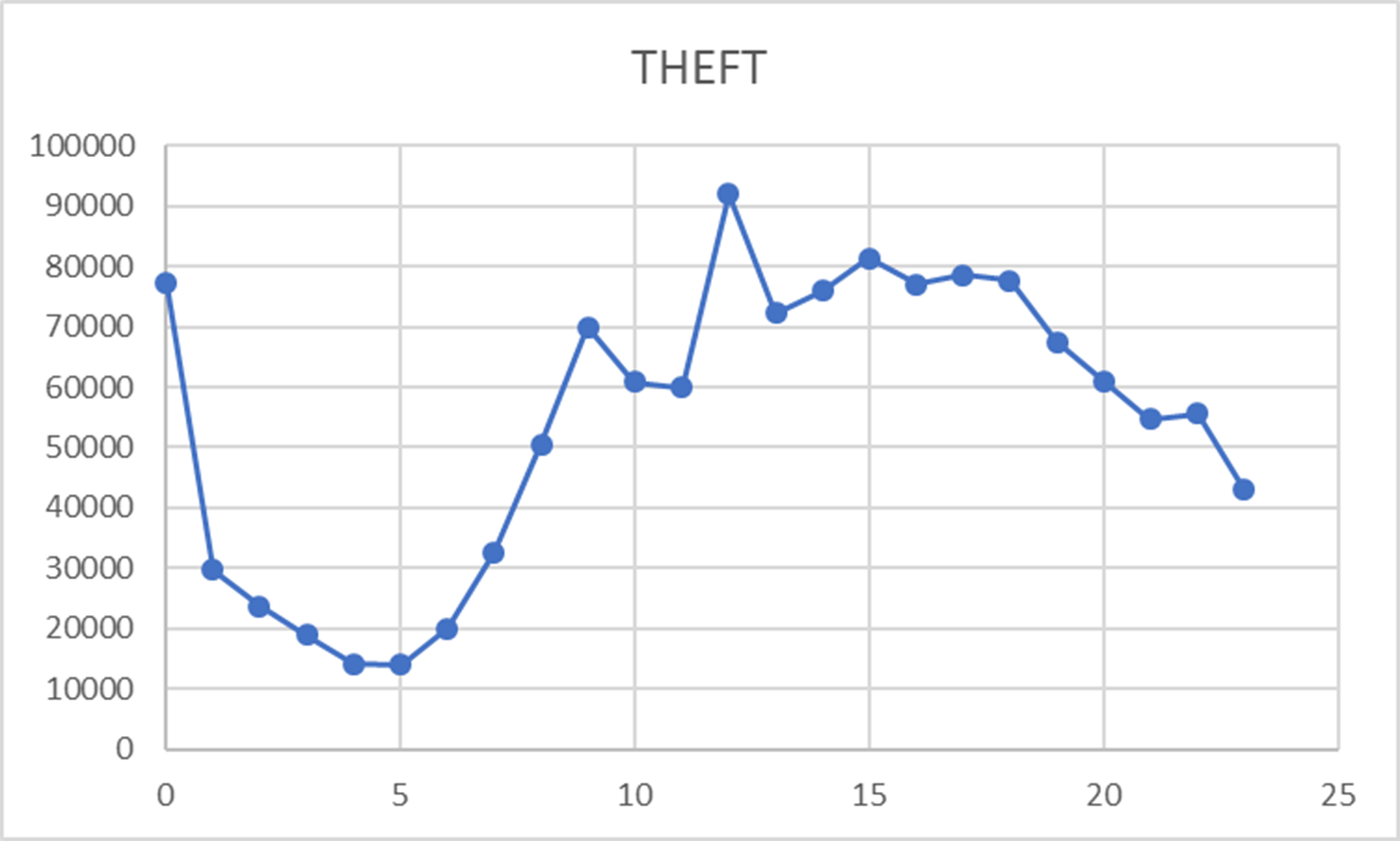
If we draw the chart of different type of crimes, we can find out assault happens most in day time, when there are more people on the street than any other time.

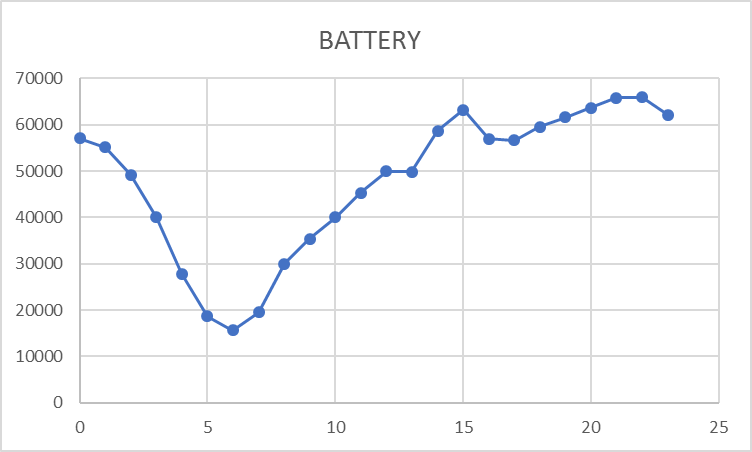


Arson happens most during the middle night, since fire is more visible at night.

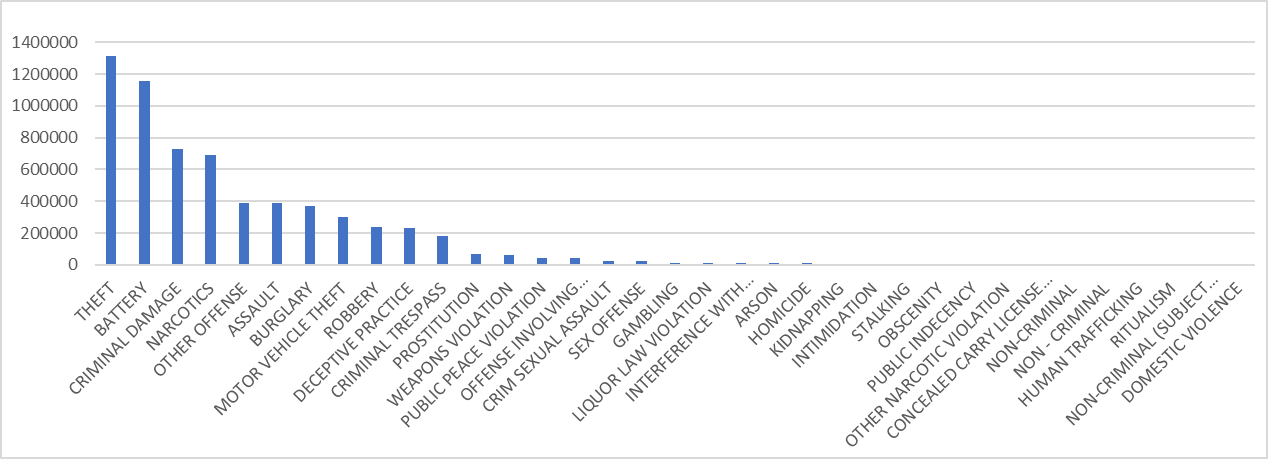


Prostitution also happens most at night when people are more likely be free and catches less attention.

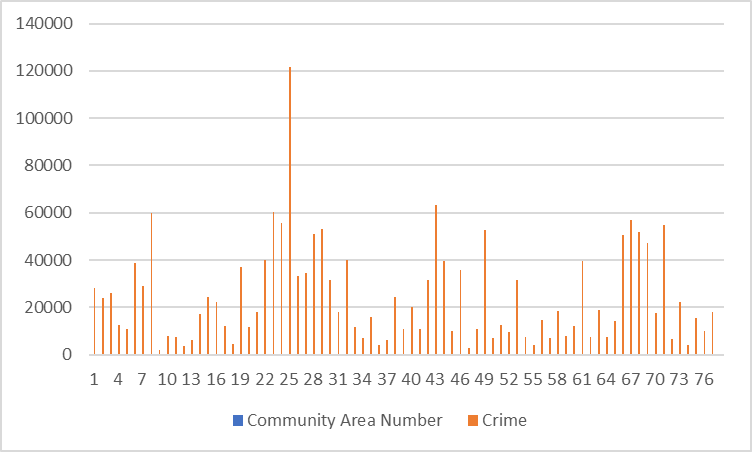




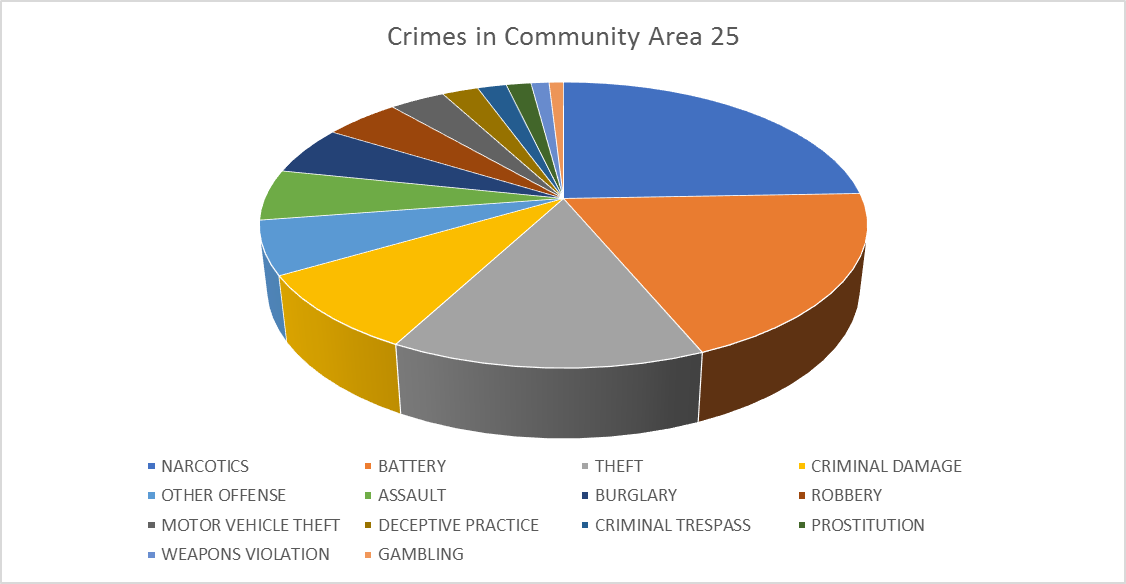
And theft and battery fits the overall crime curve well since they share the most percent of all crimes



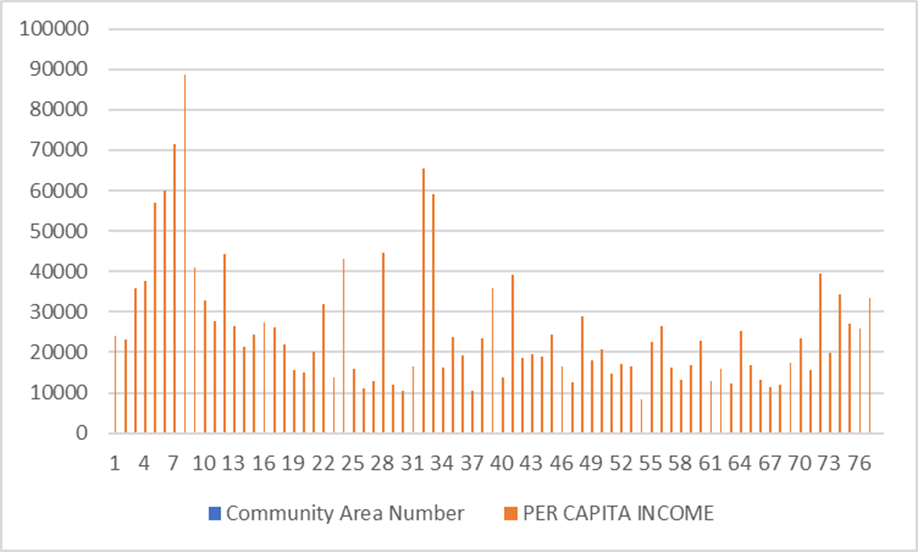
1. The relationship between crimes and different community area



As we can see, different areas have every distinct crime number. However, the community 25 has a very significant high crimes number than others.



After analyzing the number of different type of crimes, I find out the narcotics has a significant high weight. Since the people who are using narcotic are more likely to lose control of themselves, they are be the easy target of criminals or commit the crimes under the influence of drugs.

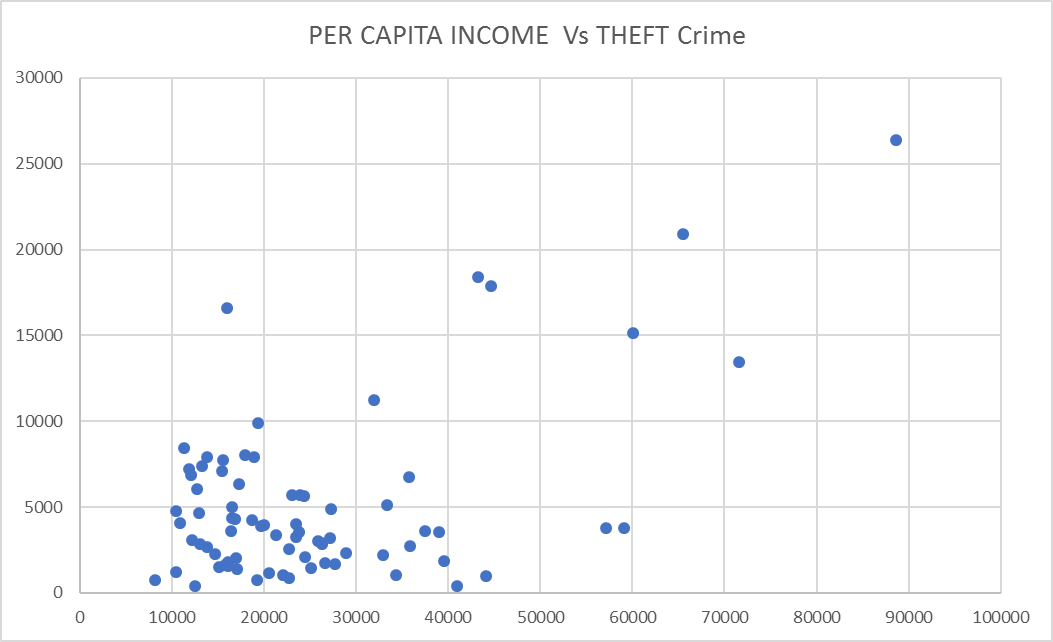


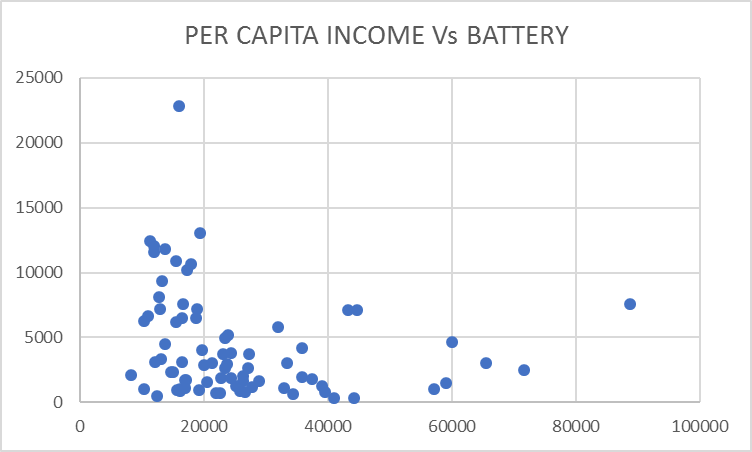
Another assumption I think is that rich community areas are safer and less crimes than the poor ones. Since people have a better life in rich community, they are less likely to commit the crimes.



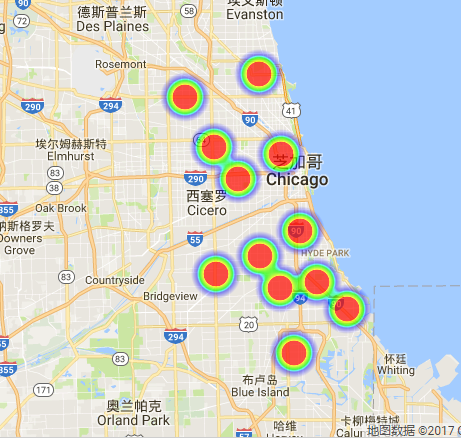


I used MATLAB analyzing the result and draw a fit for the picture of income vs crimes. And it turns out more likely to be a polynomial fit: poor and rich community areas has more crimes, while the middle class has the less crime number. The result surprised me at first. However, after I have draw the income vs Theft and battery crime, I got some clues.

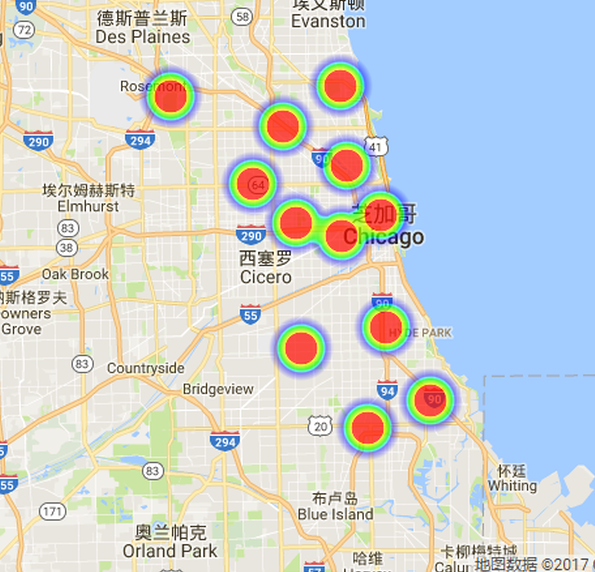




In rich community, less battery crimes are committed, but more theft happens. Meanwhile, in poor area, thief has less things valuable to steal, but more battery happens. So the middle class areas are the safest area.

1. **Correlation between different crimes**

Battery heatmap



Theft heatmap

Another interesting thing is that theft and battery don’t only have a similar time curve but also have a similar heatmap of location. So people can deploy more forces at certain time and certain area to stop crimes.

Conclusion

This project used scala to analyze the crime data of Chicago city, draw a city crime map by using google map and find out the relationship of crimes and other indicators: Arrested rate is not related to crime numbers, Season has a weak while the time in day has a strong influence on the number crime, Middle class community areas are safest areas and crimes are more likely to happen at some certain time and place.

The crime data analysis is a very interesting project. If I have more time, maybe we can find out more helpful relationships to analyze and even build a model to predict the crime.