

# Crime Report for 3 Major Cities of the United States

CIS 5200-1

Group 5

Era Singh Kajal, Neha Gupta, Tanmai Aurangabadkar, Ying Ying Lai

**Abstract:** This paper aims to research the crime situation of the main cities in United States, for example, Los Angeles, Chicago, and New York City. The datasets that will be used in this paper is from an open data source of the city government. The total size of our combined datasets is approximately 3.1GB and offers a complete overview of crime for the cities. Our main objectives in this paper is to study the relationship among cities between crime rate and several factors such as time, day, year, and location. We also want to find out which city is a relatively safer place to reside in. From this study, we can extract insights about what time is best to stay home due to safety issue and what location to avoid within the city to reduce risk. To process our data, our group will use Hive command to create table, query and calculate the needed data. Then, we will further analyze our data using appropriate visualizations. Eventually, we will present valuable insights that is extracted from the dataset. Based on the above analysis and findings, we can help United States residents to stay safe.

## 1. Introduction

The crime situation in the United States has always been an issue. This country is infamous for the crime types and frequent crime commitment that happens on a daily basis. This not only puts the country's citizen in danger, it puts it's police officer in even more danger because they are the one who stand in the front line and deal with all the crime committers. Among all the factors that cause the high crime rate in the United States, the freedom to own weapons, and the huge gap in income equality are undeniably the two most important factors that contributed the most to this situation. And among all big cities in the United States, New York City, Los Angeles, and Chicago are the top three cities that tops the entire nation in terms of high crime rate. Just to throw in a few facts for each cities: New York City's overall crime rate is 28% lower than national average; for every 100K people, there are 5.58 daily crimes; you have a 1/50 chance of becoming a victim of any crime in New York City. [1] Los Angeles' overall crime rate is 13% higher than national average; for every 100K people, there are 8.75 daily crimes; you have a 1/32 chance of becoming a victim of any crime in Los Angeles.[2] Chicago's overall crime rate is 51% higher than national average; for every 100K people, there are 11.77 daily crimes; you have 1/24 chance of becoming a victim of any crime in Chicago.[3]

## 2. Architecture Workflow

We extracted our crime data from each corresponding website in .csv format: Los Angeles data from LA Open Data, New York data from NYC Open Data, and Chicago

data from Kaggle. We uploaded our data to HDFS and used Hive queries to create external tables on the .csv data that was uploaded in HDFS. Then, we used Hive queries to select the desired data from the external table and filtered out unimportant data. Finally we used Tableau software to reproduce the selected data in the form of information by generating the appropriate graphs, charts and maps.

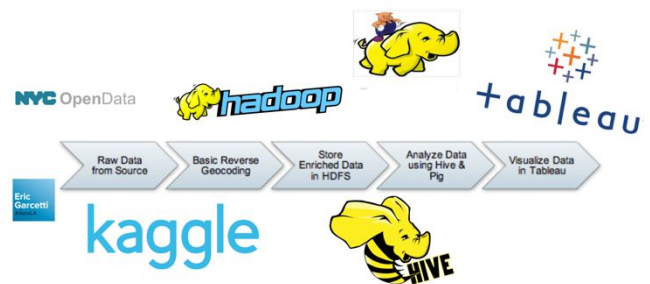


Figure 1: Flow diagram

## 3. Data Analysis

After performing the work that is mentioned in the workflow above, we got the following results below for each city. We present our findings in an appropriate chart using Tableau.

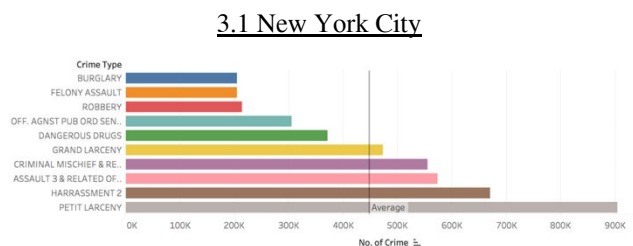


Figure 2: Top 10 Crime Types for New York City

The horizontal bar chart above presents the top 10 crime types for New York City, sorted from the least to the most. We can see that petit larceny being the most and burglary being the least. Other than that, we have 5 crime types that surpass the average crime rate of the top 10 crime types. Petit larceny is considered to be quite serious for New York City because it's crime count surpass the average crime rate a lot.

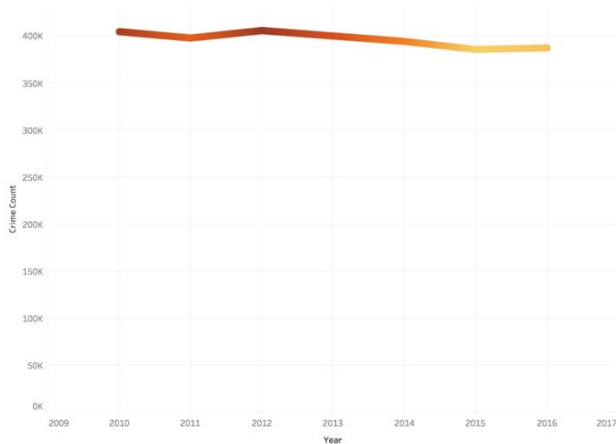


Figure 3: New York City Crime Overview, by Year

The colored line chart above displays the trend of crime rate in New York City from year 2010 to 2017. Crime rate in New York, although decreased a little, but still remain fairly high at the low 400K and high 300K throughout these years.

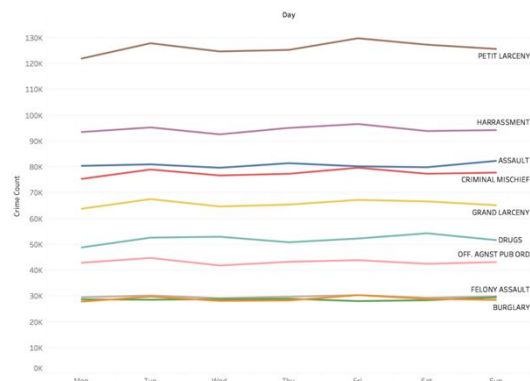


Figure 4: New York City Crime Overview, by Day

The line chart above shows the crime count of each crime type for every day of the week. In overall, there are no big increase or decrease in the crime rate in terms of day. Crime can happen at any day of the week in New York City.

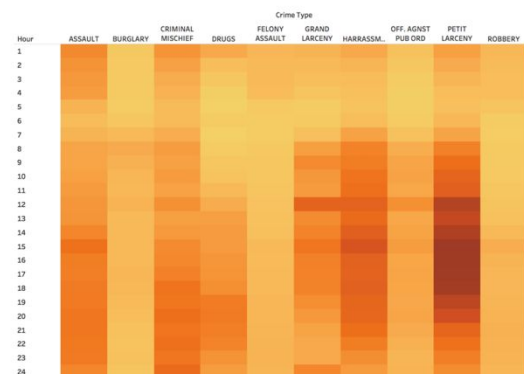


Figure 5: New York City Crime Overview, by Hour

The above heat map presents the detailed crime count of each crime type that happens in terms of time. For New York City, we can see that petit larceny, which is also the

top 1 crime in New York, happens a lot between 3pm to 8pm. Harassment also happens more during this time period.

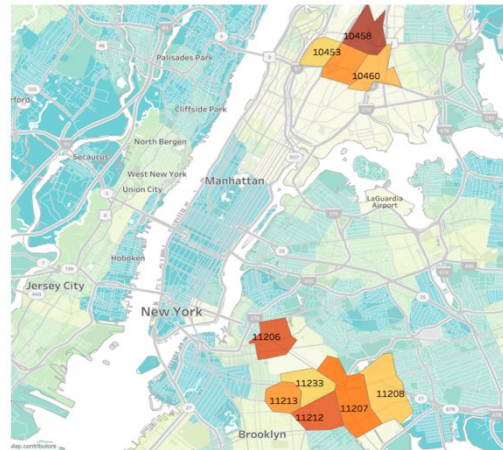


Figure 6: New York City Crime Overview, by ZipCode

The above geo map is presenting the crime rate according to zipcode, the darker the color, the higher the crime count. And this information is presented in adjunct to the corresponding median household income sorted by zipcode. Expectedly, crime rate is high at areas with lower median household income, which means economic status is a main factor when it comes to crime issue.

### 3.2 Los Angeles

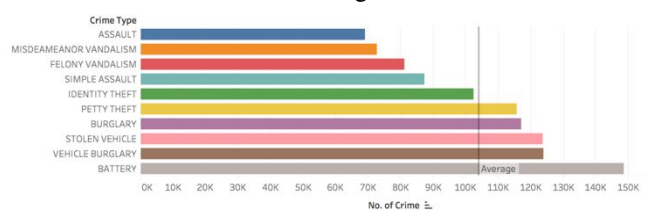


Figure 7: Top 10 Crime Types for Los Angeles

The horizontal bar chart above presents the top 10 crime types for Los Angeles, sorted from the least to the most. We can see that battery being the most and assault being the least. Other than that, we have 5 crime types that surpass the average crime rate of the top 10 crime types. Among the 5 crime types that surpasses the average line, petty theft, burglary, stolen vehicle, and vehicle burglary are pretty equal in terms of seriousness for Los Angeles.

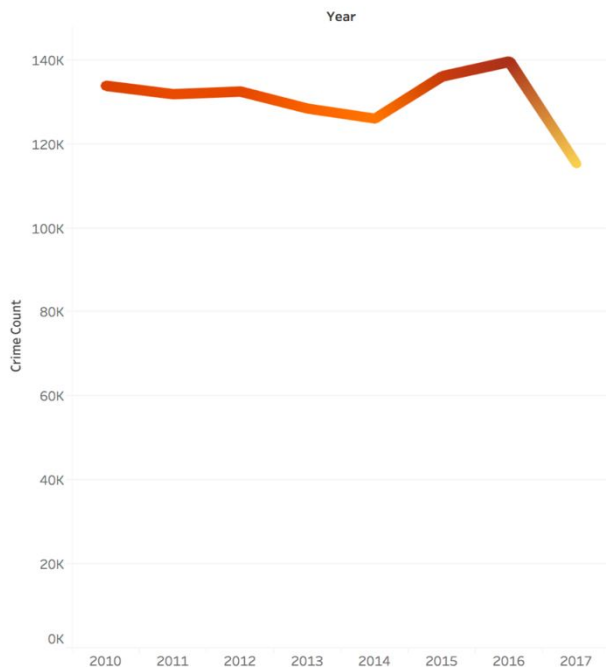


Figure 8: Los Angeles Crime Overview, by Year

The colored line chart above displays the trend of crime rate in Los Angeles from year 2010 to 2017. Crime rate in Los Angeles decreased in overall, but there is a sudden peak in year 2016 and sudden dip in year 2017. We are not sure what caused this huge up and down in such a short two years time.

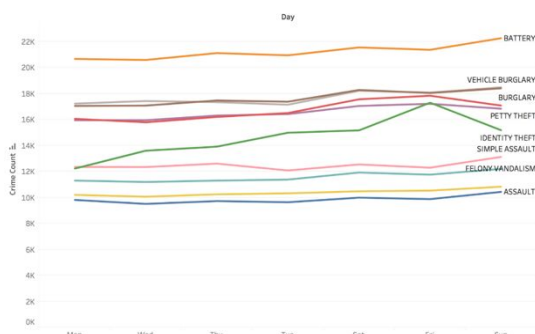


Figure 9: Los Angeles Crime Overview, by Day

The line chart above shows the crime count of each crime type for every day of the week. It is worth noting that identity theft is especially high on Friday, and we are estimating that is due to everyone's dining out on Friday night and using their credit cards, making identity theft much easier to be commit. There is also an upward trend in crime count on Sunday for every crime type in Los Angeles.

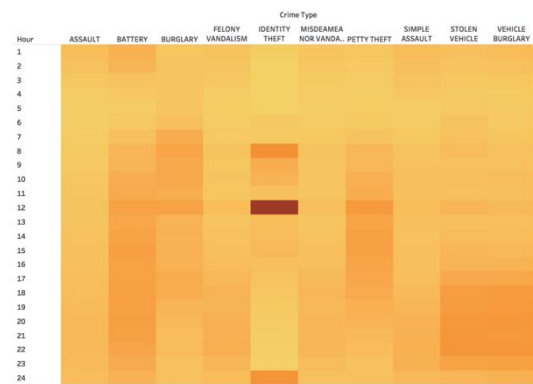


Figure 10: Los Angeles Crime Overview, by Hour

The above heat map presents the detailed crime count of each crime type that happens in terms of time. For Los Angeles, we can see that no matter the crime type, crime can happen at any time throughout the day. But one thing worth noting is that identity theft is unusually high on 12pm during the day. We think this is due to identity theft is being committed using some automated system, and the system is generated to steal personal information every 12 hours.

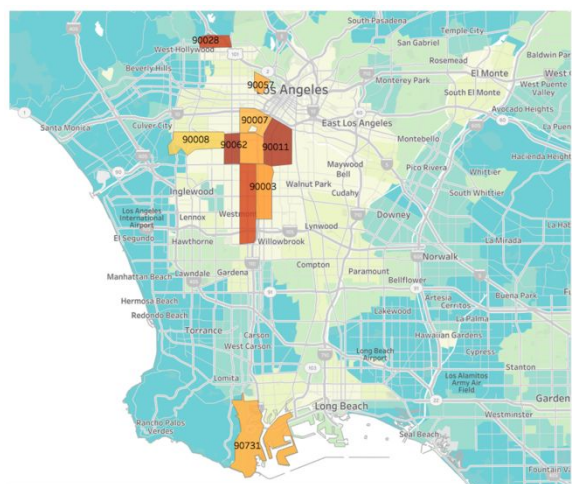


Figure 11: Los Angeles Crime Overview, by ZipCode

The above geo map is presenting the crime rate according to zipcode, the darker the color, the higher the crime count. And this information is presented in adjunct to the corresponding median household income sorted by zipcode. Expectedly, crime rate is high at areas with lower median household income, which means economic status is a main factor when it comes to crime issue.

### 3.3 Chicago

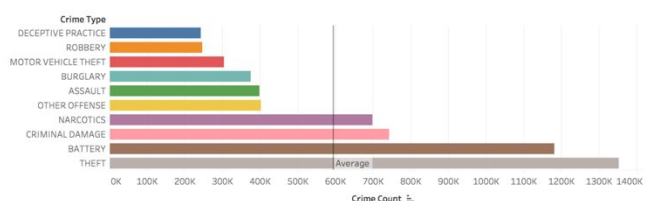


Figure 12: Top 10 Crime Types for Chicago

The horizontal bar chart above presents the top 10 crime types for Chicago, sorted from the least to the most. We can see that deceptive practice being the most and theft being the least. Other than that, we have 4 crime types that surpass the average crime rate of the top 10 crime types. However, theft and battery are considered to be quite serious because it's crime count surpasses the average crime rate by a lot.

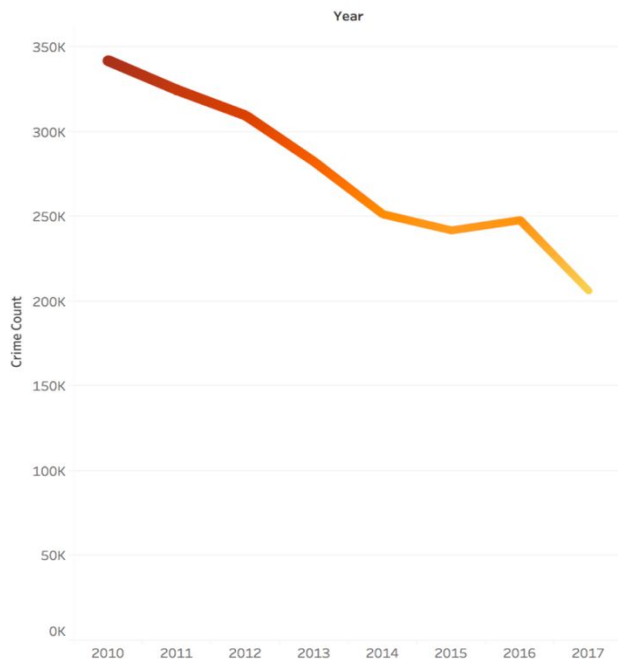


Figure 13: Chicago Crime Overview, by Year

The colored line chart above displays the trend of crime rate in Chicago from year 2010 to 2017. Crime rate in Chicago has surprisingly been steadily decreasing over these years.

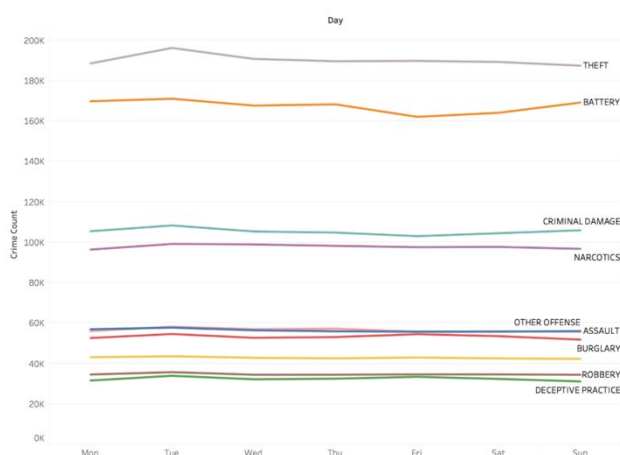


Figure 14: Chicago Crime Overview, by Day

The line chart above shows the crime count of each crime type for every day of the week. In overall, there are no big increase or decrease in the crime rate in terms of day. Crime can pretty much happen at any day of the week in Chicago.

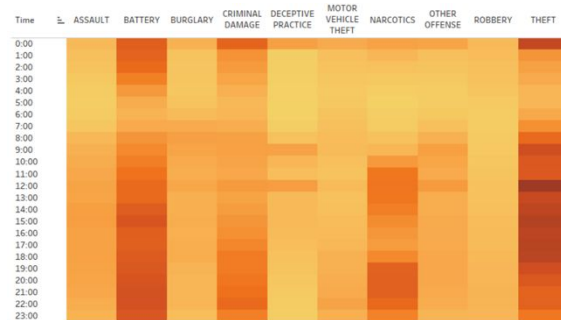


Figure 15: Chicago Crime Overview, by Hour

The above heat map presents the detailed crime count of each crime type that happens in terms of time. For Chicago, we can see that theft, which is the top 1 crime in Chicago, happens a lot during the daytime, which is quite surprising, because we always have the notion that theft happens when the sun is down. Now that this information is presented, residents of Chicago should be aware even if it's day time.

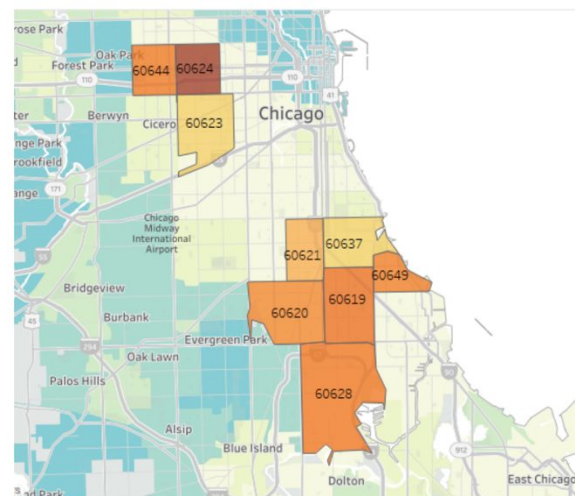


Figure 16: Chicago Crime Overview, by ZipCode

The above geo map is presenting the crime rate according to zipcode, the darker the color, the higher the crime count. And this information is presented in adjunct to the corresponding median household income sorted by zipcode. Expectedly, crime rate is high at areas with lower median household income, which means economic status is a main factor when it comes to crime issue.



#### 4. Conclusion: Comparison in between Cities

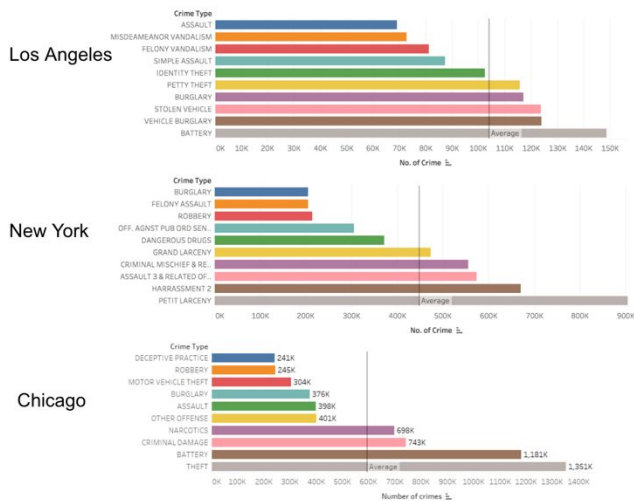


Figure 17: Comparison Chart for 3 Cities

As a conclusion, crime rate will always be high in places where population density is high. In comparison between these three cities, we conclude that in terms of crime severity, Chicago is the most dangerous, then followed by New York City, and Los Angeles is the safest among the three. However, theft-related crime is the common crime type among the three cities and pretty much any crime can be committed anytime throughout the day. Last but not least, economic situation is undeniably the most important factor that caused people to commit crime.

#### 5. Group Submission GitHub Link

<https://github.com/ngupta8/5200-Project>

#### 6. References

- [1] Facts for New York City  
<http://www.areavibes.com/new+york-ny/crime/>
- [2] Facts for Los Angeles  
<http://www.areavibes.com/los+angeles-ca/crime/>
- [3] Facts for Chicago  
<http://www.areavibes.com/chicago-il/crime/>