**Crime Analysis of Chicago and Los Angeles**

**1. Introduction**

Crime in Chicago and Los Angeles has been tracked by the Police Department’s Bureau of Records since the beginning of the 20th century. The city’s violent crime rate, is substantially higher than the US average. Although national crime rates in the United States have stayed near historic lows, Chicago was supposedly responsible for nearly half of 2016’s increase in homicides in the United States. .As of 2017, Chicago’s homicide rate is significantly higher when compared to the larger American cities of New York and Los Angeles, but lower when compared to smaller American cities. The reason for the violence which is localized to some areas of the city, including change in police tactics or increase in gang rivalry, remain unclear.

In this analysis, we explore the Crimes Data Set between the years 2013 - 2017, with the hope that we would be able to provide atleast some answers to the following questions :  
1. How has the number of various crimes changed over time ?  
2. How have the number arrests corresponding to the crimes changed over time?  
3. Which crimes are most frequently committed?  
4. Which locations are these frequent crimes being committed in?  
5. How has the number of Homicides changed over the years in Chicago?

**2. Target Audience**

This project is intended towards Police Department of both the cities. By being able to identify high crime neighbourhoods, we can work on arriving at a solution by using a combination of external demographic, socio-economic, cultural and ethinic data to figure out if the violence is being perpetrated by violent gangs, or are these gangs forming on account of a negligent and an incapable government, or if poverty and poor education is indirectly leading to all these crimes, or if social media such as facebook youtube and mainstream cinema/tv shows are to blame for culturally influencing individuals to take up arms, because the recurring themes among tv shows, movies and rap songs these days seems to be about guns gangs gold and girls , all the while gloryifing an unrealistic “hustling” life style.

**3. Description of Data Sources**

**3.1 Data Sources Overview**

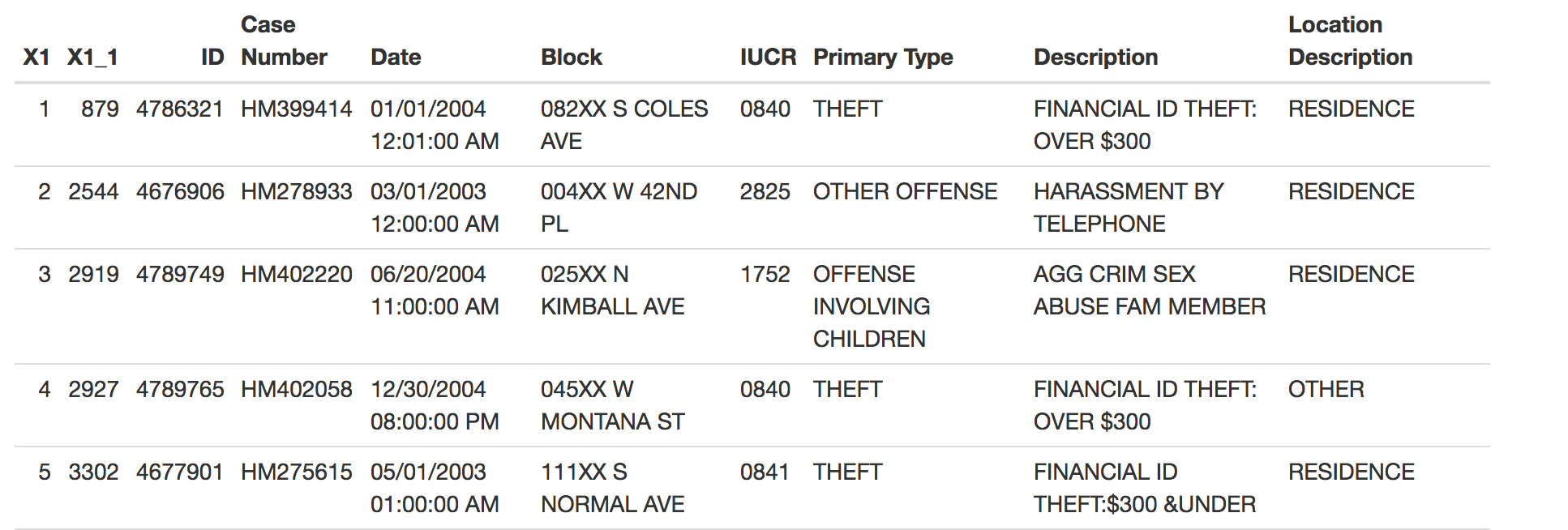
For this project, two data sources were used. Both the data sources came from data.gov

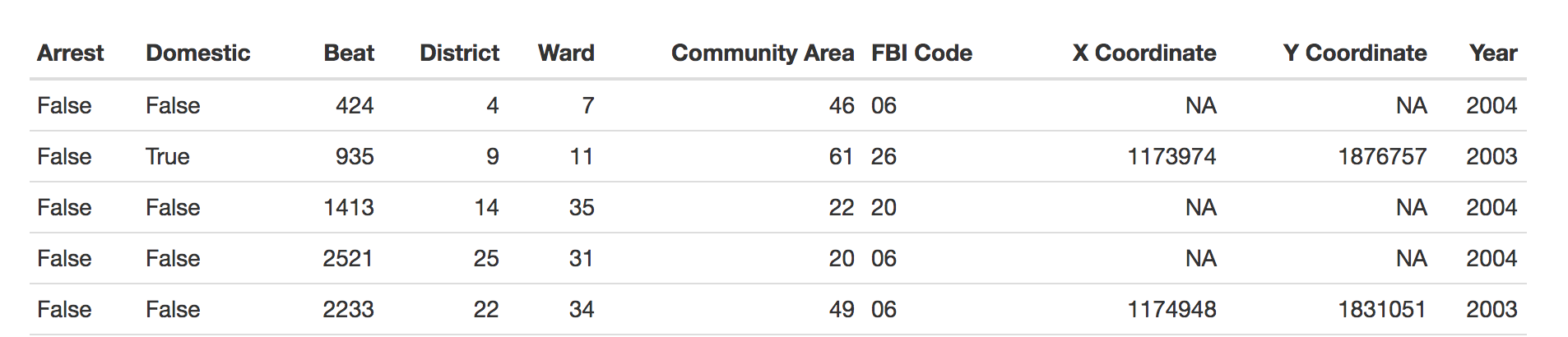
1. Crime Dataset of Chicago from 2001 to Present
2. Los Angeles Crime Dataset 2010 to Present

**3.2 Crime Dataset of Chicago from 2001 to Present**

This dataset contains information on various kinds of crimes committed in Chicago. The dataset size was in millions but we removed some of the unnecessary fields otherwise processing the data would be very difficult.

Below is the screenshot of the original dataset with the field name and few sample data in it:

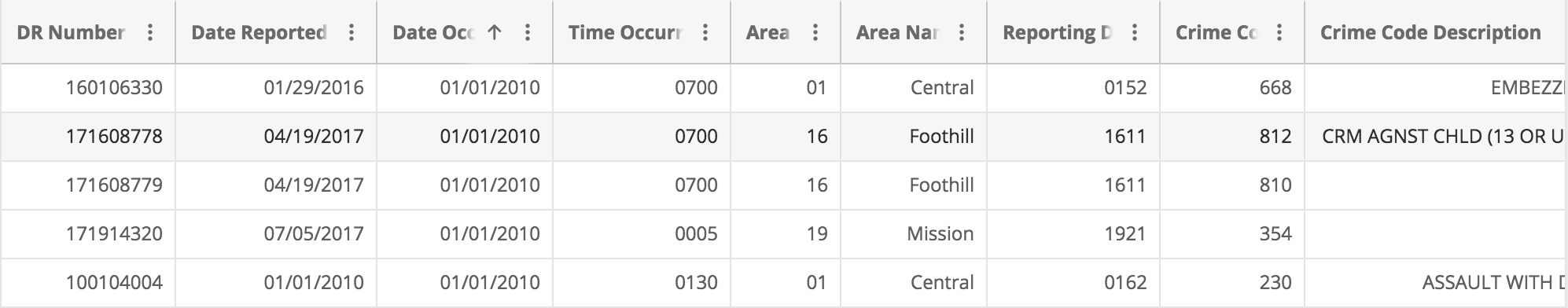


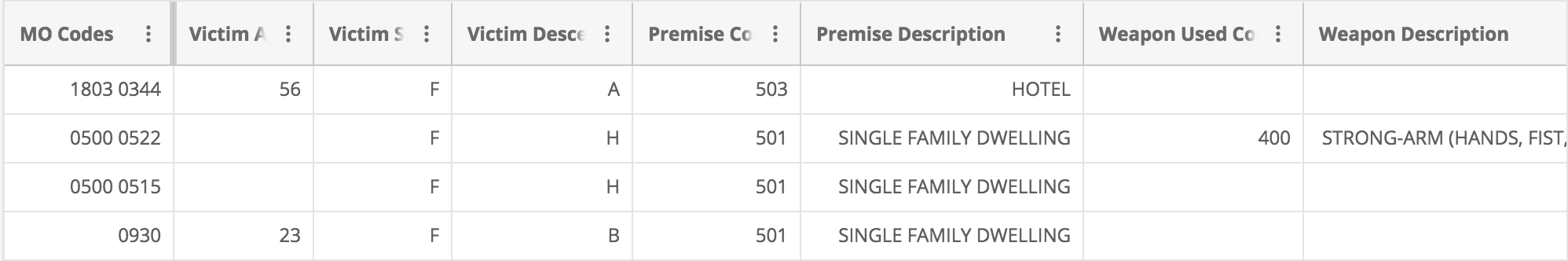


**3.3 Los Angeles Crime Dataset 2010 to Present**

This dataset reflects incidents of crime in the City of Los Angeles dating back to 2010. There were approximately 1.72 Million record initially.

Below is the screenshot of the original dataset







**4. Data Integration**

Two datasets used for this project. Both the dataset included description, location ,date, time etc of Chicago and Los Angeles.. We also integrated the datasets together using the Google visualization API. With this API we were able to plot points and show results from various datasets on one chart. Altogether we performed five queries. Four of these queries relied on data from single datasets. Integrating these datasets together allowed us to make comparisons that were not possible using one dataset alone.

More information about the individual queries can be seen below in section 5.

**5. Data Product Results**

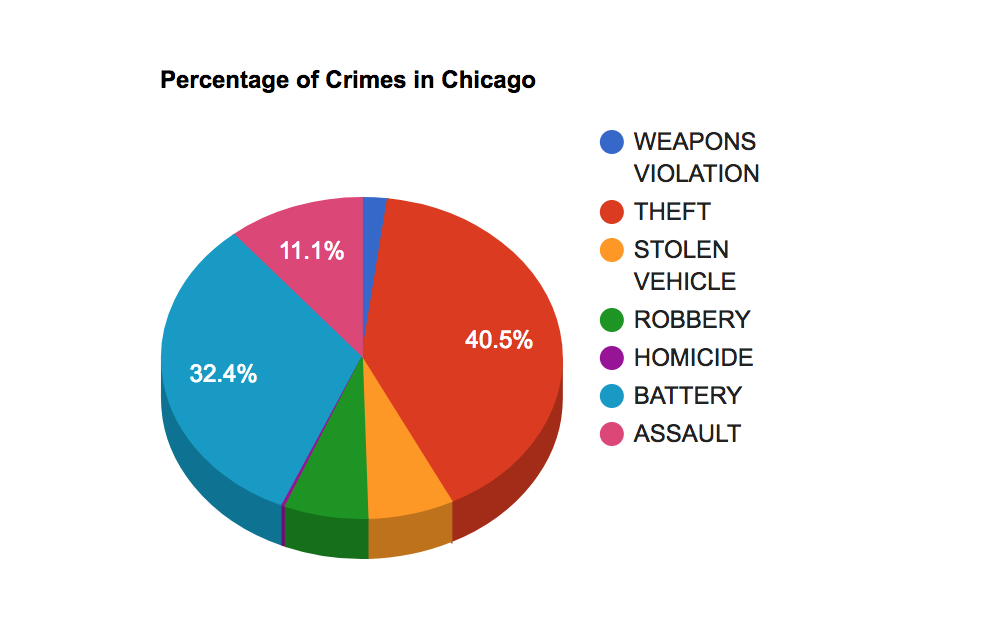
We created six query visualizations for our data sets.

1. Frequency of various types of Crimes in Chicago
2. Total number of Crimes vs total number of arrests in Chicago
3. Frequency of various Crimes in common areas in Los Angeles
4. Frequency of various types of Crimes in Los Angeles
5. Comparison of various types of crimes in all the years between Chicago and Los Angeles

**5.1 Frequencies of various types of Crimes in Chicago**

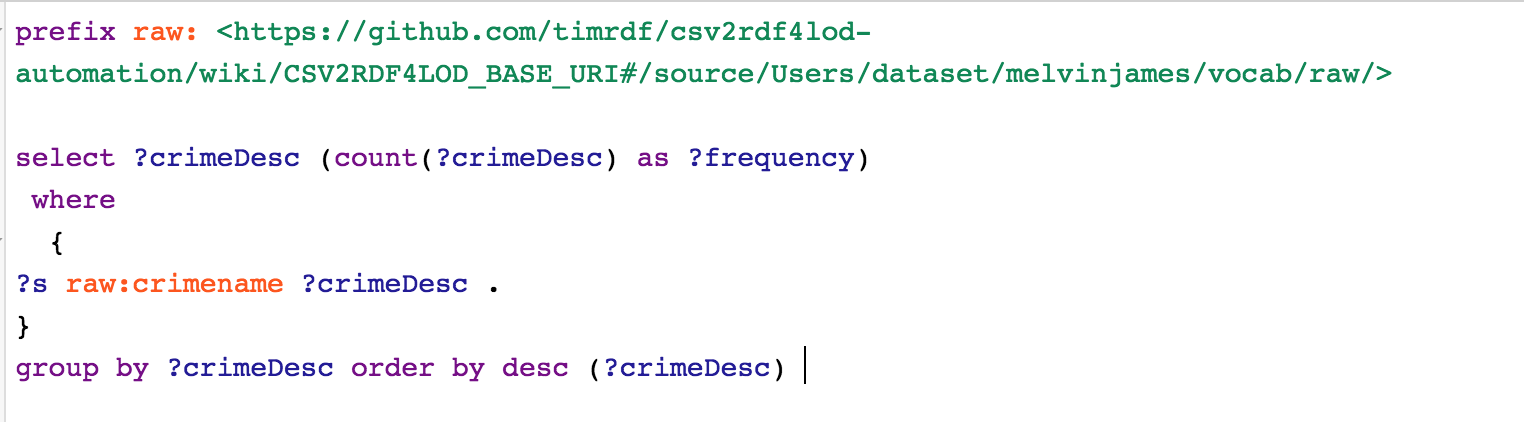
The first query was a simple display of the frequency of each of the 8 crimes in the city of Chicago between 2013 through 2015. We used a single dataset of Chicago in this query. Below is the result of the query

The results of this query can be seen in figure 5.1.1.



**Figure 5.1.1**

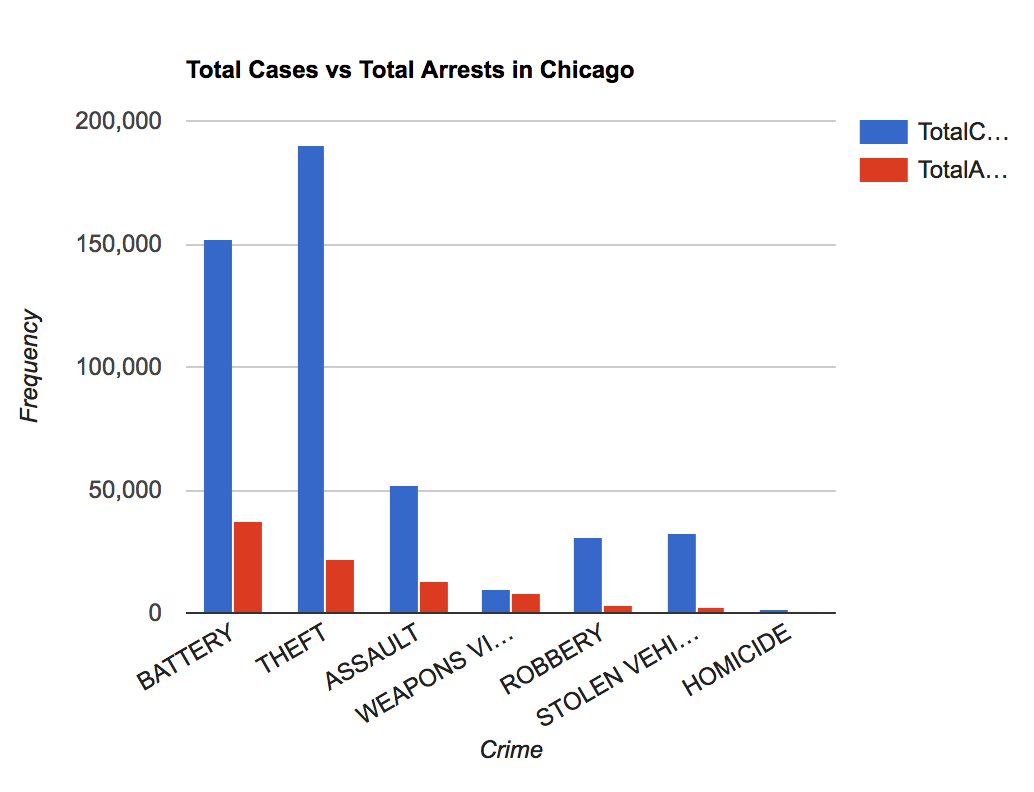
In figure 5.1.2 you can see the text of the SPARQL query that was used to return the results. Here we do a group by of the crime name and then count the frequency in each of the categories.

  
**Figure 5.1.2**

For the visualization in figure 5.1.1. we simply displayed the result geomap using the Google Visualization API. From this map, we can clearly see which crime is most frequent in Chicago in those years.

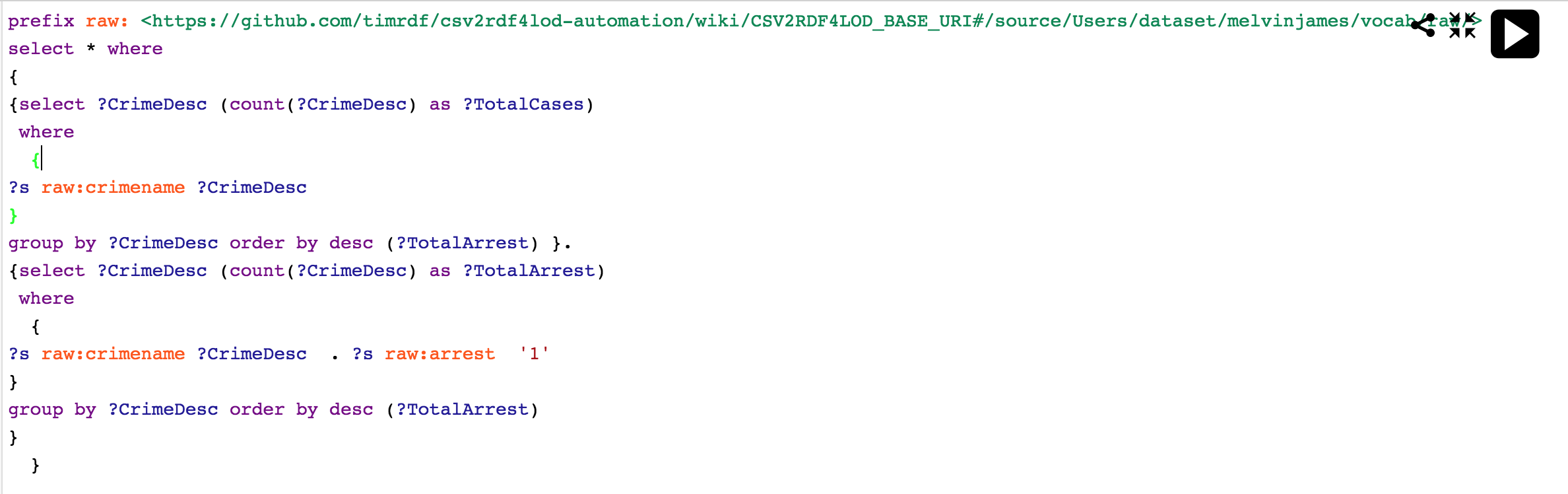
**5.2 Total number of crime vs total number of arrests in Chicago**

The second query was an extension of the first query. For query two we compared the total

  
**Figure 5.2.1**

number of cases and the respective arrest in those cases from 2013 to 2015. The results of this query can be seen in figure 5.2.1.

In figure 5.2.2 you can see the text of the SPARQL query that was used to return the results.

  
**Figure 5.2.2**

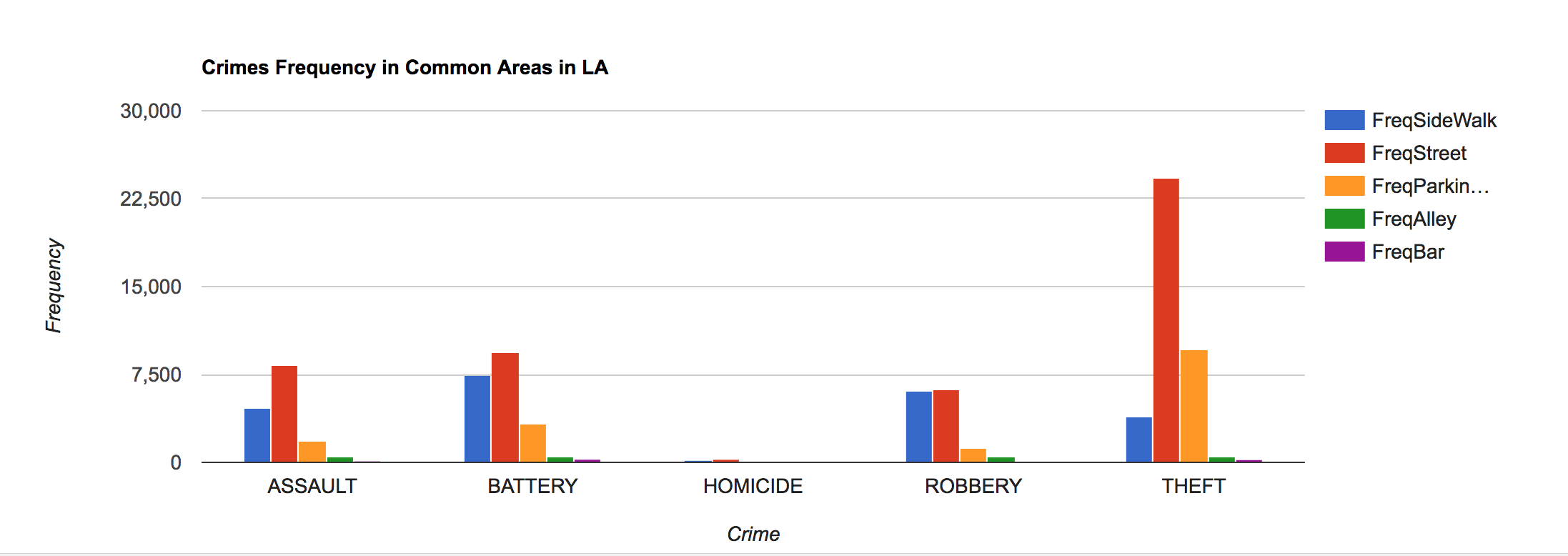
For the visualization in figure 5.2.1. We simply displayed the total cases of a crime and total arrest of that crime in the combochart using the Google Visualization API.

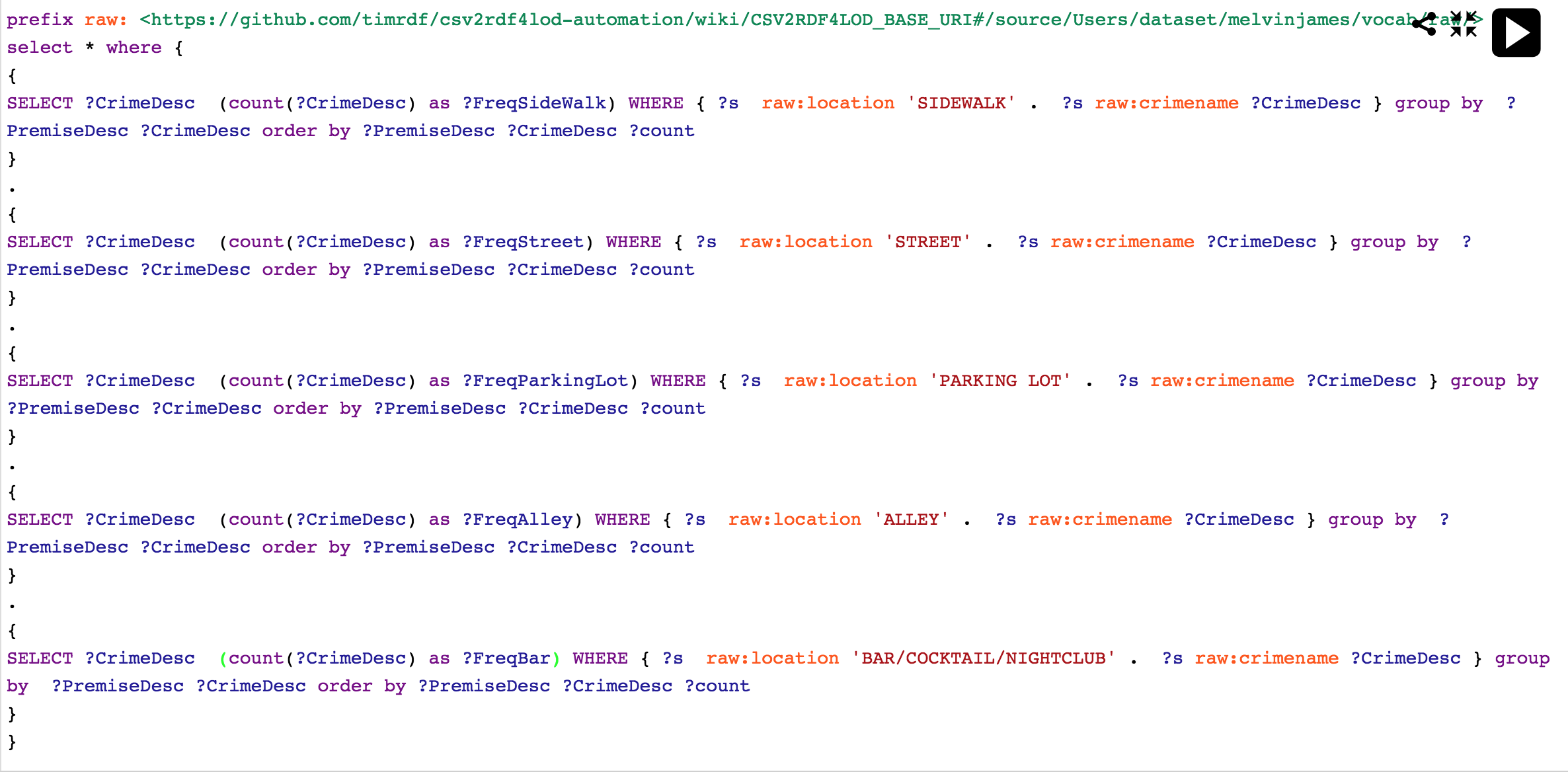
**5.3 Frequency of various crimes in common areas in Los Angeles**

The third query was a simple display of the various crimes committed in common areas of Los Angeles. The results of this query can be seen in figure 5.3.1.

In figure 5.3.2 you can see the text of the SPARQL query that was used to return the results.

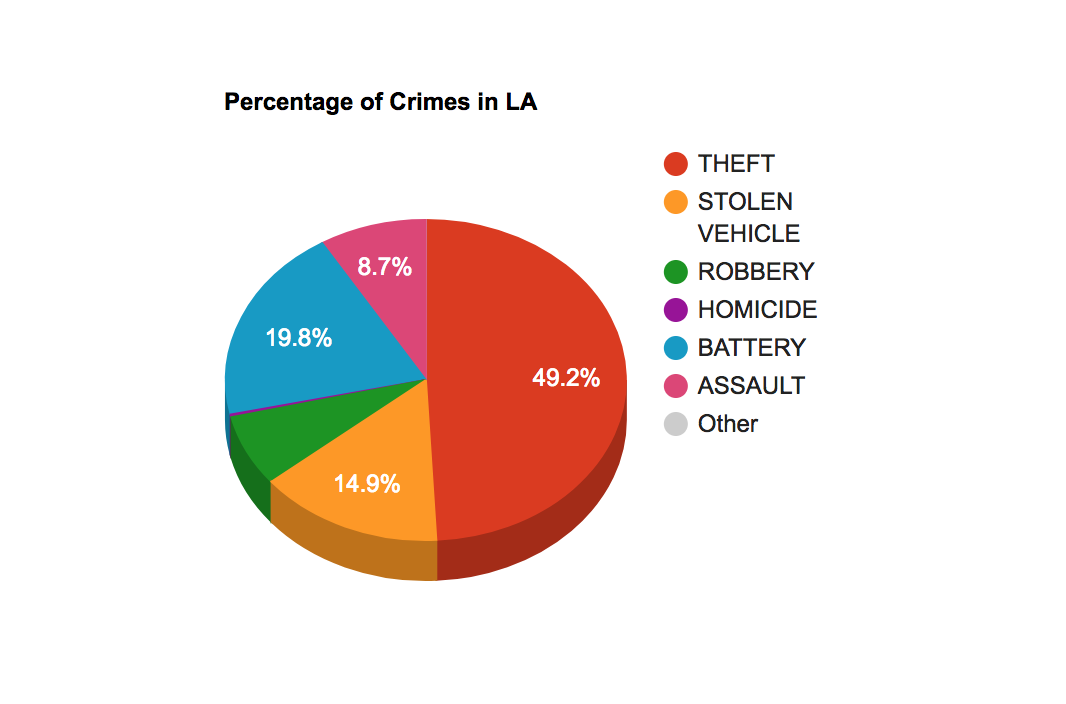
For the visualization in figure 5.3.1.. From this map, we can see which crime is common in what areas.

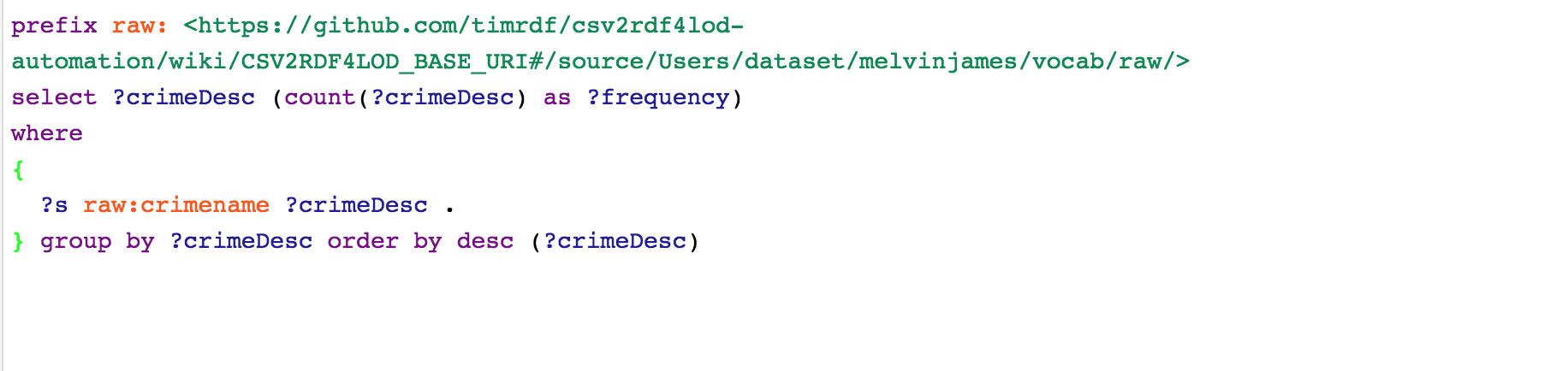
  
**Figure 5.3.1**

**  
Figure 5.3.2**

**5.4 Frequencies of various types of Crimes in Los Angeles**

The fourth query is just like the first query.

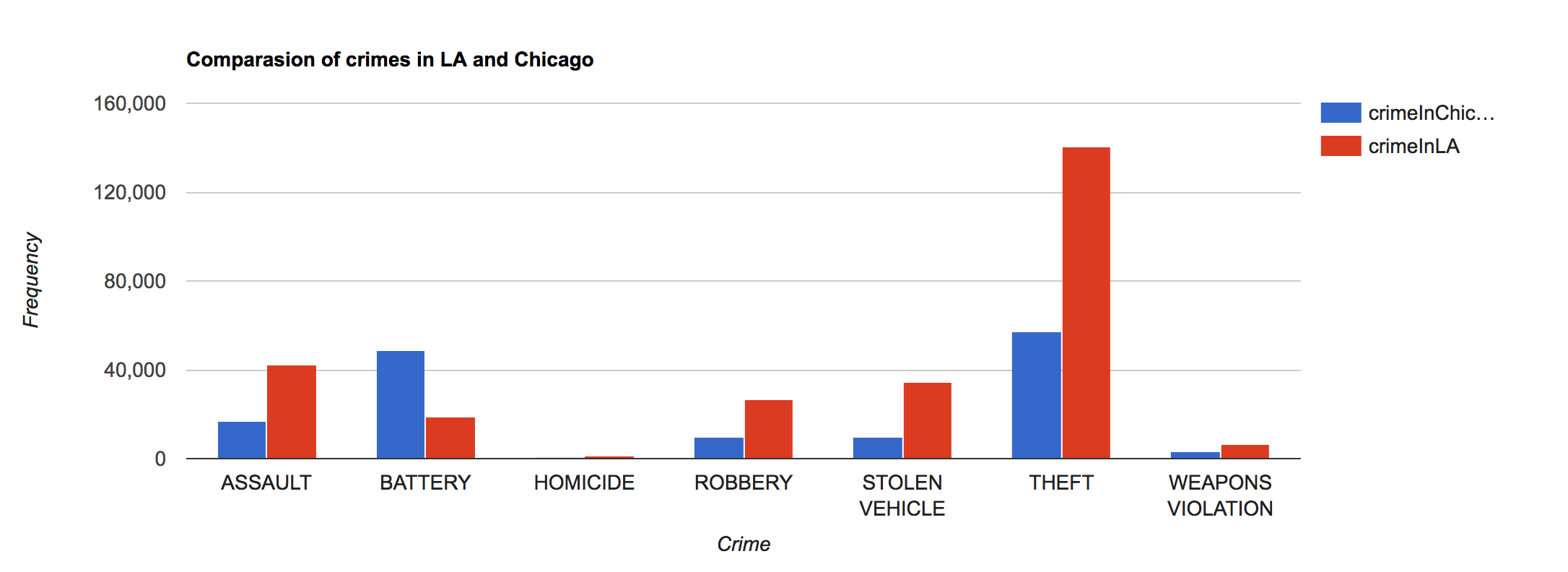
  
**Figure 5.4.1**

  
**Figure 5.4.2**

**5.5 Comparison of various types of crimes between Chicago and Los Angeles**

Query five used both the data sets.

In figure 5.5.2 you can see the text of the SPARQL query that was used.



**Figure 5.5.1**

  
**Figure 5.5.2**

**6. Custom Project Justification**

The project uses Fuseki order to serve the datasets through SPARQL endpoint. We had to move to a custom project because of unavailability of required SPARQL endpoint. Also we used two datasets: LACrimeData and CHCrimeData for deriving meaningful inferences from it. We have compared many features of each dataset to reach a plausible conclusion.. Mashing up different data sets now only helps in identifying new indicators and factors but also incepts a whole new perspective towards a problem as this. It is done so by analyzing different results to find reasons for the mismatch found in the assumption.

**7. Summary**

We have gathered a lot of knowledge regarding crimes occurring in the cities of Los Angeles and Chicago. We were able to understand the frequency of different types of crimes. Further we were able to infer from the results that are generated, if adequate number of arrests are being made for every case. The police can have a better understanding where and when major crimes can occur. As for civilians, one can know which places are safer to visit and be ready knowing about details of crimes happening in these cities.