

Analysis of Boston Crime Rates

Joshua Coleman and Ravi Pochiraju

MSDS 430 Final

Summary

Boston is the most populous city in New England. It's population is close to three-quarters of a million and continues to increase every year. Understanding Crime in large cities is an important issue not only to improve safety but to increase public awareness. If these types of studies are published, the citizens would do their part in staying safe and help reduce the burden on law enforcement.

This project explores Boston's Crime statistics from 2015 to 2018 and provides some insights into various types of Crimes that were reported. It also provides details of serious crimes (compared to minor offences) and when these are most likely to occur.

Data Sources

Boston crime data set is available to public on Kaggle.com

Data can be downloaded from : <https://www.kaggle.com/AnalyzeBoston/crimes-in-boston>

Data Clean-up

Studies are only as good as the data itself. If the dataset contains missing pieces of information, has duplicates or contains irrelevant information, we would be chasing rainbows. Clean up was an important exercise for us to perform before we could start any analysis on it.

Our original dataset contained the following fields :

Data Field	Values
Offense Code	Corresponding values found in offense codes CSV
Incident Number	Unique number for each incident
Offense Code Group	Descriptive categorization of offense
Offense Description	Offense details
District	Boston district where incident was reported
Reporting_Area	Report area (0-962)
Shooting	Y if shooting reported
Occured_on_Date	Date crime occurred

Year	Year crime occurred
Month	Month crime occurred
Hour	Hour crime occurred (0-23)
UCR_Part	Seriousness of crime with part 1 being the most serious,e.g.,homicide, and part 3 being the least serious,e.g., motor accident
Street	Street where crime occurred
Lat	Latitude where crime occurred
Long	Longitude where crime occurred
Location	(Lat,Long)

This data set contains records from the Boston Police crime incident report system, showing fields to identify when, where, and what type of crime occurred. There are 319,073 data points with 17 fields to identify the crime. The data set has some missing values. For its relatively large size though it is overall a pretty complete data set with only a couple thousand values missing in some fields. For example the latitude and longitude coordinates are missing or improperly labeled as -1 for about 19,000 records and some missing values in the Reporting Area attribute. We were able to clean up the longitude and latitude section by dropping records with -1, but this clean up was done only before analysis where location was key. Even though there is a column to record shooting it is a column we decided to drop it since it only had 1000 data points of a reported shooting.

Data Insights

- Exploratory data analysis (EDA) is a usual practice of finding patterns in the data set involving visual methods. EDA involves finding correlations among the fields in the data set. It is a preliminary process that helps gather insights before taking a more sophisticated statistical approach. Visuals provided here are the results of our EDA.

Since the data pertains to Boston's crime incident reporting from 2016 to 2017, we intuitively know that **location, time and cause of incident** play an important role in our analysis.

a. Selection of data fields

Since Location, time and cause play a role, the following fields were used :
Street, Lat/Long, District, Area, Year, Month, Day, Hour, Group, Description

b. Evaluating relationships between data fields

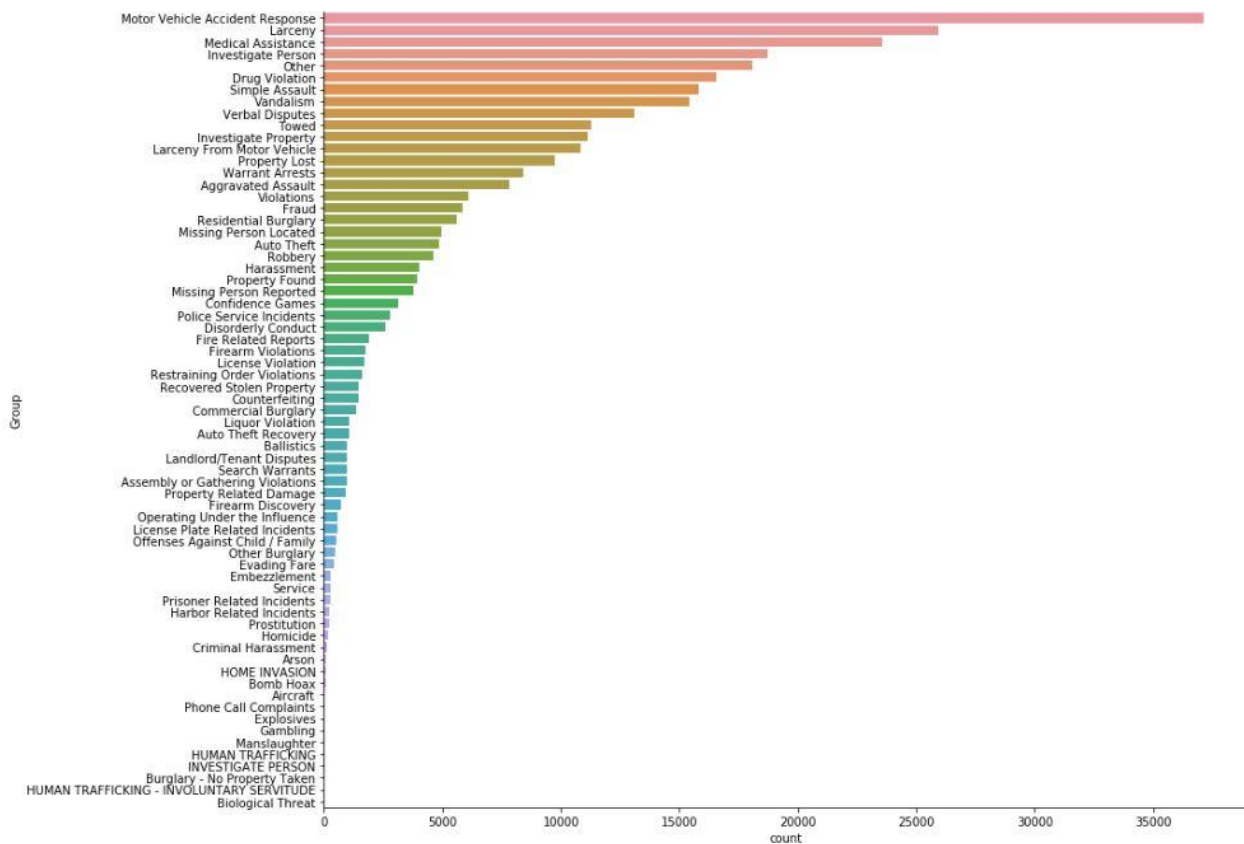
After isolating the most important fields we wanted to analyze, some of the questions we asked were :

1. What are the different types of incidents that are reported ?
2. How many incidents are serious vs minor ?
3. Do the Crimes occur at a certain time ? Is it possible that more incidents occur during the weekdays than weekends (as people may not be in their homes during the day)
4. Where do the Crimes occur frequently - in what zipcodes or districts ?
5. Do Crimes increase during the holiday period ? (as people may travel and homes might be vacant)
6. Have the number of Crimes increased in the last 2 years ?

In order to answer some of the questions we asked, we used one of Python's visualization modules (Seaborn) that provides intuitive graphics.

1. What are the types of incidents that are reported ?

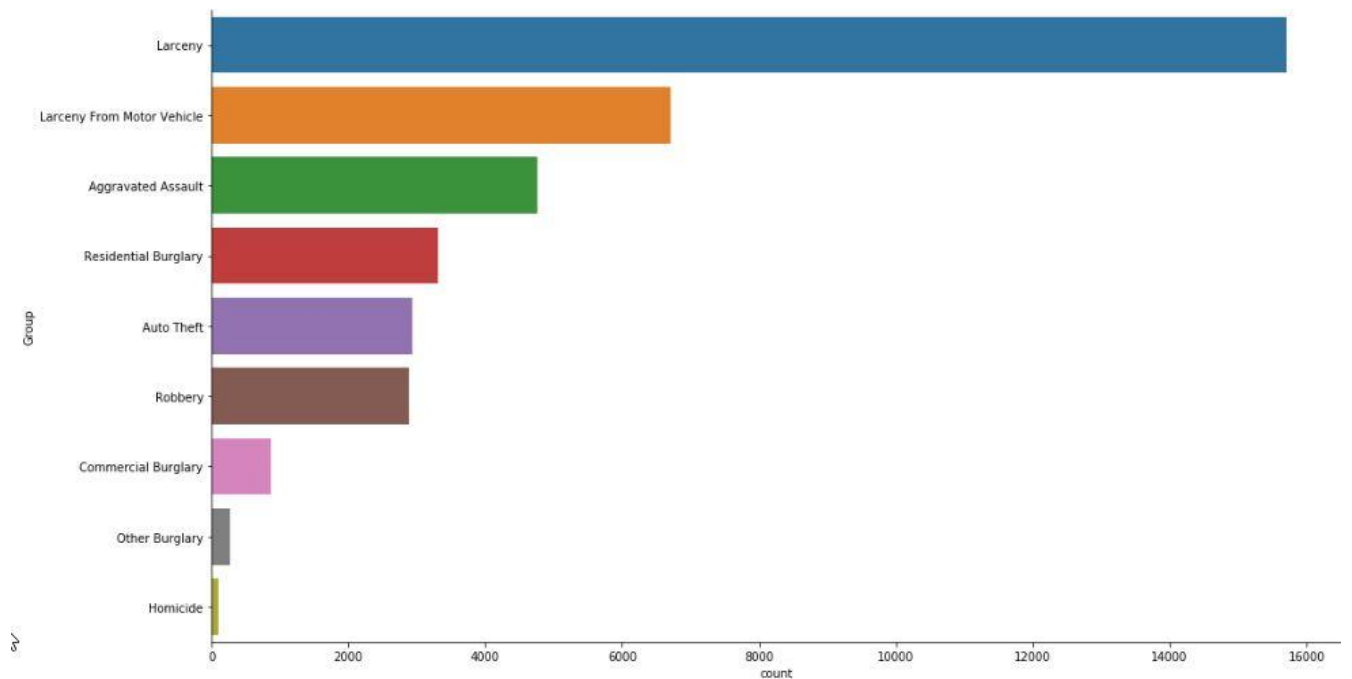
Our initial question was to evaluate the different types of incidents that Boston police logged into the system. This helps us differentiate the most frequent from the rarest.



Motor vehicle accident responses and Larceny seem to be the top 2 frequently reported incidents (in 2016 and 2017) and some of the more serious incidents appear at the bottom (Human Trafficking, Manslaughter etc.). This chart provides the Police department a way to appropriately allocate their workload. If the number of motor vehicle related incidents continue to trend up, there might be more patrolling officers than personnel in the office.

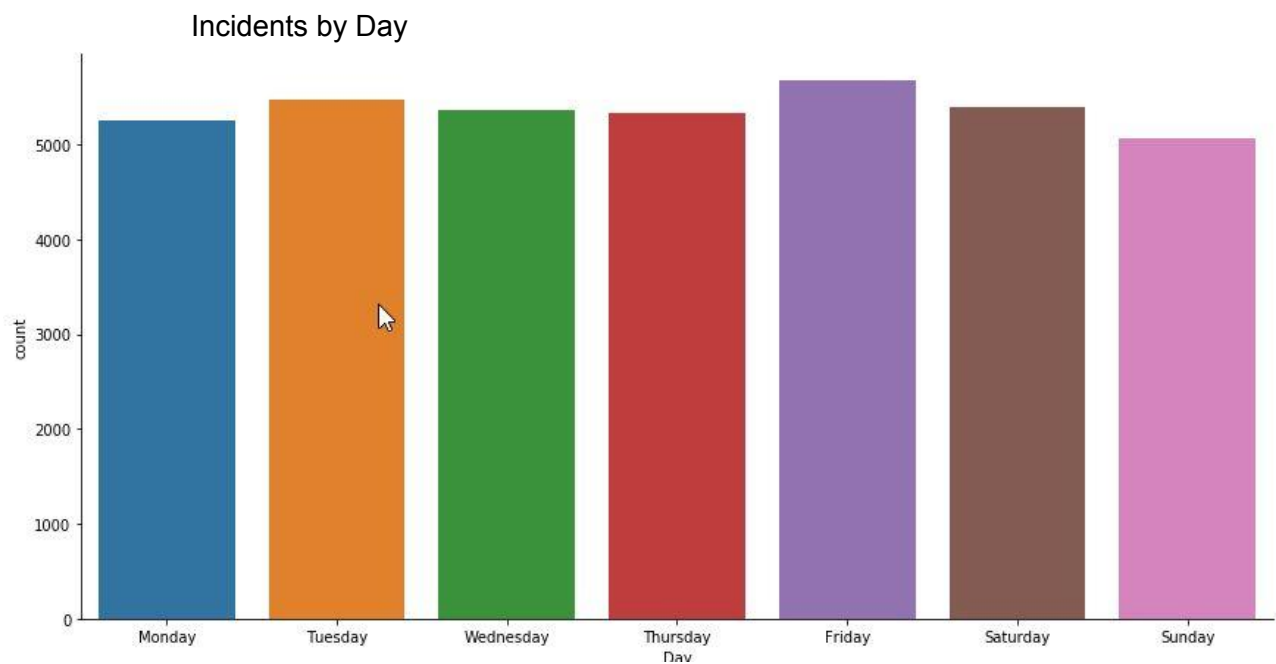
2. Out of the reported Crimes, how many are serious vs minor offences ?

Serious crimes play a key role in keeping the cities a lot safer. Gathering data on different types of serious crimes would not only be insightful but also in taking adequate steps in reducing it. Residential burglaries can be minimized if public takes precaution by not leaving their doors unlocked or by installing security systems.

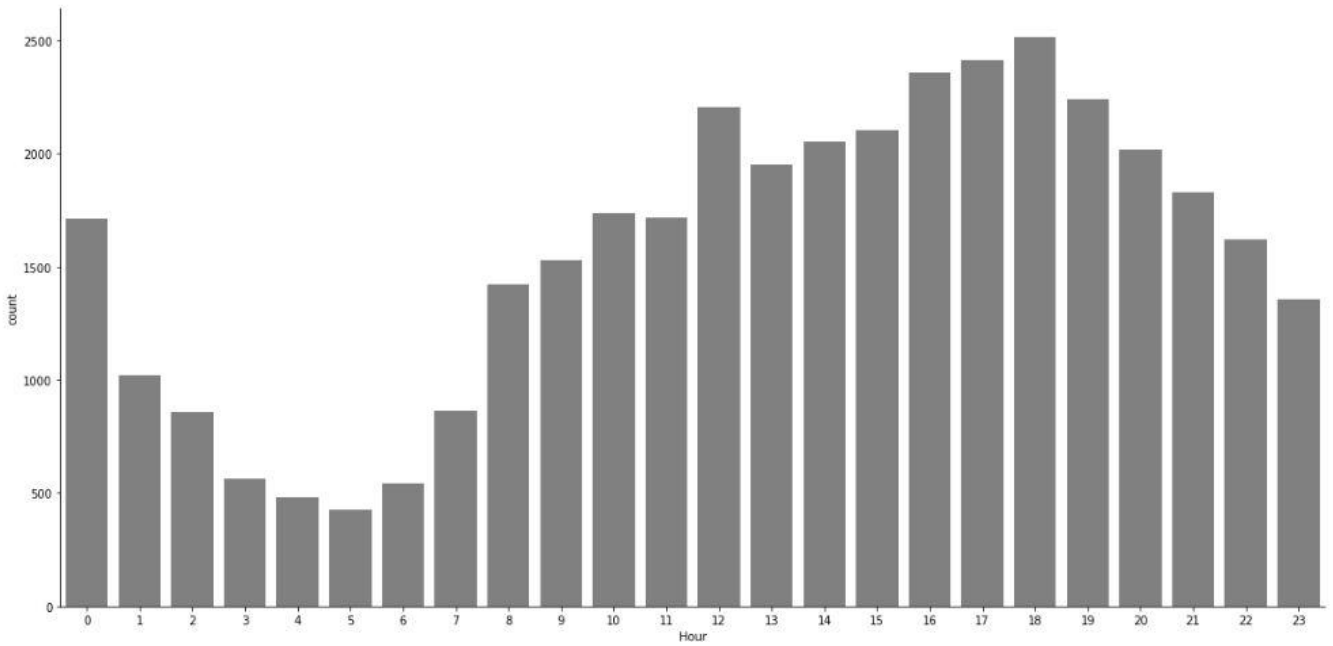


About 45% of reported incidents (16K out of 37K) happen to be Serious crimes (Larceny) reported followed by motor vehicle related thefts.

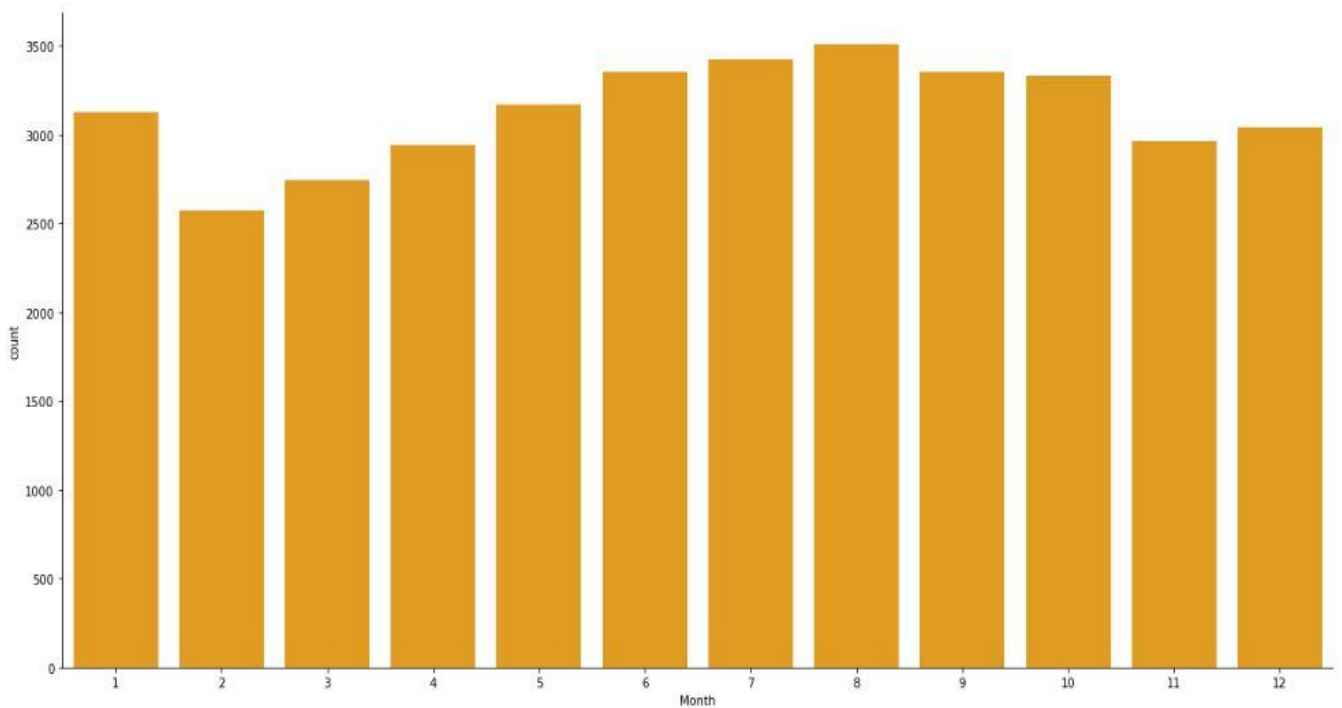
- When do these Serious Crimes occur - are there specific times of the day or days of the week or months when these reported incidents are higher ?



Incidents by Hour



Incidents by Month

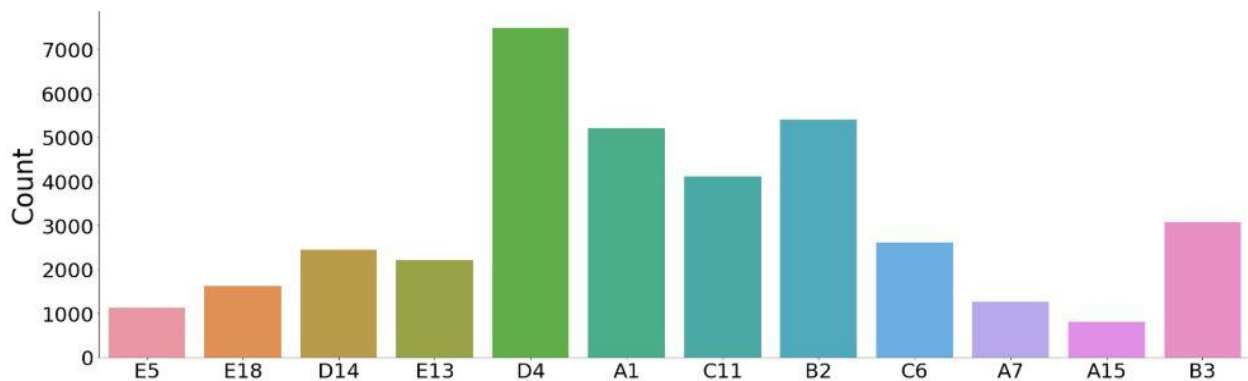


Crime rates are low past midnight till early morning. The rate gradually increases towards the evening and appears to peak at 6 pm. Variation across days of the week doesn't seem much,

though Friday has the highest crime rate. During the year, summer months (June through August) have a high rate compared to the rest of the year. January is an outlier that has a high crime rate. If we had more information in the data set, we would have explored why January stands out.

4. Where did major Crime occur ?

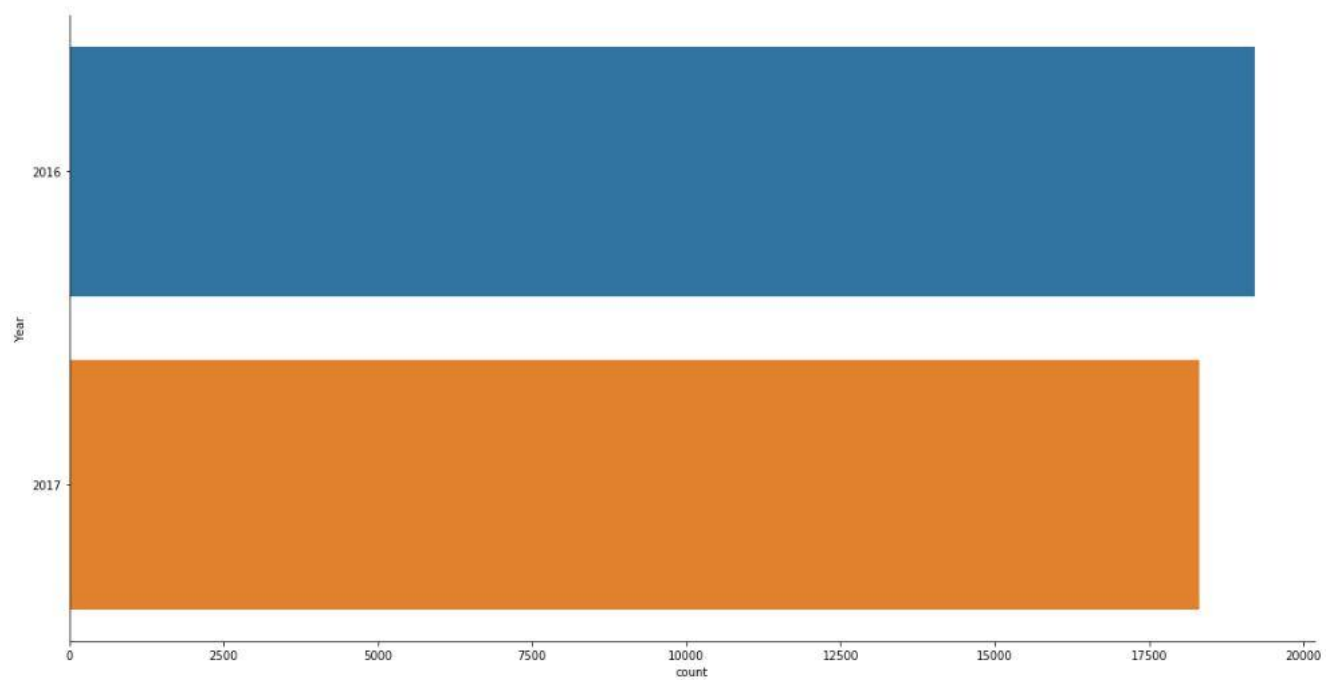
If all the Serious crimes occurred in specific districts or areas, increasing law enforcement in that area might be helpful in reducing Crime rate.



Majority of serious crimes occur in the heart of the city. The districts having the most serious crimes reported are D4, B2, and A1. The amount of serious crimes tapers off in the outskirts of the city.

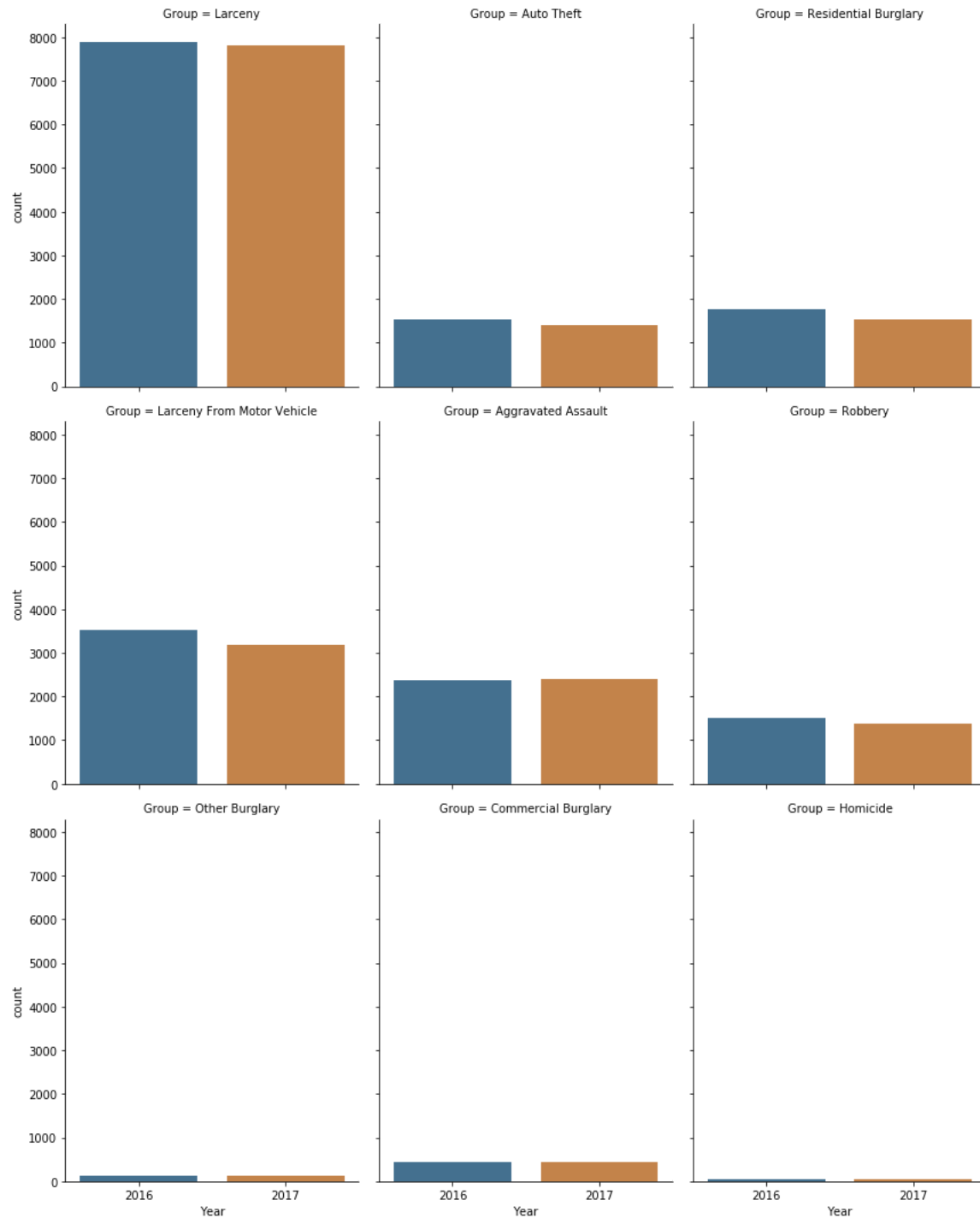
5. Did the Crime rate go up in 2017 from 2016 ?

Useful statistic in exploring further is whether the rates have been decreasing or increasing over time. If the rates are increasing over time, we can take additional measures in keeping the city safer.



There is a slight decrease in serious crime rate in 2017 from 2016.

6. What contributed to the decrease in 2017 from 2016 ?



Even though the overall Crime rate has decreased in 2017, Larceny from motor vehicle (the largest nbr of incidents) hasn't changed much. Residential burglaries, Auto thefts and Homicides have decreased slightly though Commercial burglaries stayed the same. Aggravated assaults have gone up in 2017 compared to the previous year.

Conclusion

Our goal was to glean any insights we could from exploratory data analysis and find answers to the questions we asked with the available data. As evident from the visuals, some meaningful relationships do occur. The Crimes are high in certain districts (D4, B2, A1) . The incident rate also increases towards the evening and peaks around 6 pm. Summer months have a higher Crime rate than the others (with January being an outlier). We believe summer months tend to higher rates as people are out in the open more than in the winter. Larceny incidents are the highest among all the others. In general, Crimes have decreased in 2017 from 2016. There were fewer homicides and crimes in 2017 than in 2016.

A more thorough study could be possible if the dataset was richer than we worked with. If dataset contained income, education level and demographics, this study could have been more inferential and predictive.