Data Visualization of Chicago Crime

by

Remya Kannan

DePaul University

TABLE OF CONTENTS

Non-Technical Summary:

Over the past two decades, the United States has experienced an unpredicted drop in crime. Chicago, while often portrayed as a violent city, has seen sustained drops in violent crime and homicide rates during this time, particularly recently. Therefore, police departments in big cities want to understand, predict and if possible prevent or mitigate potential damage from crimes. Through this project, we aim to analyze and visualize crime in Chicago City.

The dataset has been obtained from Kaggle. It has crime records across Chicago over the past decade. The final data set used is a subset with even distribution across the time-period. Data about offenses includes fields depicting the nature of crime, time of crime, block-level location information, legal treatment of crime and resulting punishment. This data set is then transformed for further analysis.

There is general interest as to why crime in Chicago is the most talked about even though there is decrease in the overall crime rates in Chicago. What are the top crimes that occur in the city? What area of the city is most likely to be hit? What time of the day or year do crimes mostly happen? Do relationships between crimes exist?

The following report will highlight the answers to the above questions for analysis. In the first section, we get an overview of the data and its structure. The technical summary includes details on data exploration and the various visualization techniques used to understand and learn various trends in the data. Finally, we analyze the results and draw conclusions based on the trends observed through visualization.

Data Description and Research Objective:

The original Chicago Crime data from Kaggle(https://www.kaggle.com/currie32/crimes-in-chicago) contains 22 attributes, mostly categorical, and more than 6,000,000 records logged daily from beginning of 2012 to end of 2016. We removed the rows with missing values and created few more columns (Year, Day Month, and Weekday) in order perform our visualization on the cleaned data based on the timeline selected.

```
> str(crime_cleaned)
'data.frame':   1419561 obs. of  26 variables:
 $ X                   : int  3 89 197 673 911 1108 1130 1801 1868 1891 ...
 $ ID                  : int  10508693 10508695 10508697 10508698 10508699 10508702 10508703 10508704 10508709 10508982 ...
 $ Case.Number         : Factor w/ 1456599 levels "","161884","223432",..: 1259932 1259876 1259938 1259885 1259906 1259900 1259926 1259948 1259955 1260048 ...
 $ Date                : Date, format: "2016-05-03" "2016-05-03" "2016-05-03" "2016-05-03" ...
 $ Block               : Factor w/ 32774 levels "0000X  I94/EXIT 12",..: 6522 24532 22016 20633 1934 28843 12722 1297 7137 26523 ...
 $ IUCR                : Factor w/ 365 levels "0110","0141",..: 52 52 43 40 85 31 40 40 40 52 ...
 $ Primary.Type        : Factor w/ 33 levels "ARSON","ASSAULT",..: 3 3 28 3 32 3 3 3 3 3 ...
 $ Description         : Factor w/ 342 levels "$500 AND UNDER",..: 121 121 266 284 1 33 284 284 284 121 ...
 $ Location.Description: Factor w/ 143 levels "","ABANDONED BUILDING",..: 19 113 129 125 113 129 35 114 125 129 ...
 $ Arrest              : Factor w/ 2 levels "False","True": 2 1 1 1 1 1 1 1 1 1 ...
 $ Domestic            : Factor w/ 2 levels "False","True": 2 2 1 1 2 1 1 1 1 2 ...
 $ Beat                : int  1022 313 1524 1532 1523 631 133 215 2432 735 ...
 $ District            : num  10 3 15 15 15 6 1 2 24 7 ...
 $ Ward                : num  24 20 37 28 28 8 3 3 40 17 ...
 $ Community.Area      : num  29 42 25 25 25 44 35 38 1 67 ...
 $ FBI.Code            : Factor w/ 26 levels "01A","01B","02",..: 11 11 25 11 8 6 11 11 11 11 ...
 $ X.Coordinate        : num  1154907 1183066 1140789 1143223 1139890 ...
 $ Y.Coordinate        : num  1893681 1864330 1904819 1901475 1901675 ...
 $ Year                : Factor w/ 6 levels "2012","2013",..: 5 5 5 5 5 5 5 5 5 5 ...
 $ Updated.On          : Factor w/ 959 levels "01/01/2016 03:52:56 PM",..: 306 306 306 306 306 306 306 306 306 306 ...
 $ Latitude            : num  41.9 41.8 41.9 41.9 41.9 ...
 $ Longitude           : num  -87.7 -87.6 -87.8 -87.7 -87.8 ...
 $ Location            : Factor w/ 368287 levels "","(36.619446395, -91.686565684)",..: 192927 118612 235207 222691 223552 60912 174917 153900 358859 95475 ...
 $ Day                 : Factor w/ 31 levels "1","2","3","4",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Month               : Ord.factor w/ 12 levels "Jan"<"Feb"<"Mar"<..: 5 5 5 5 5 5 5 5 5 5 ...
 $ Weekday             : Ord.factor w/ 7 levels "Sun"<"Mon"<"Tues"<..: 3 3 3 3 3 3 3 3 3 3 ...
 - attr(*, "na.action")=Class 'omit'  Named int [1:37123] 155 156 157 158 159 160 161 162 163 165 ...
  .. ..- attr(*, "names")= chr [1:37123] "155" "156" "157" "158" ...
```

Technical Summary:

1. Exploratory Visualization of the data:

Bar charts shown in Fig.1, are created to observe the general distribution of data based on various categorical variables (Crime Type, Crime Location, Arrest and Domestic). Overall, we have a deeper understanding of the features to be considered in the analysis and the important features/categorical levels to be included for further analysis. First, we focus on the top 5 or top 10 most frequent crime types and locations for which will be included for analysis further in this study (Time series plot and geography map). Additionally, visualization based on arrest rate by crime type or location is also graphed.

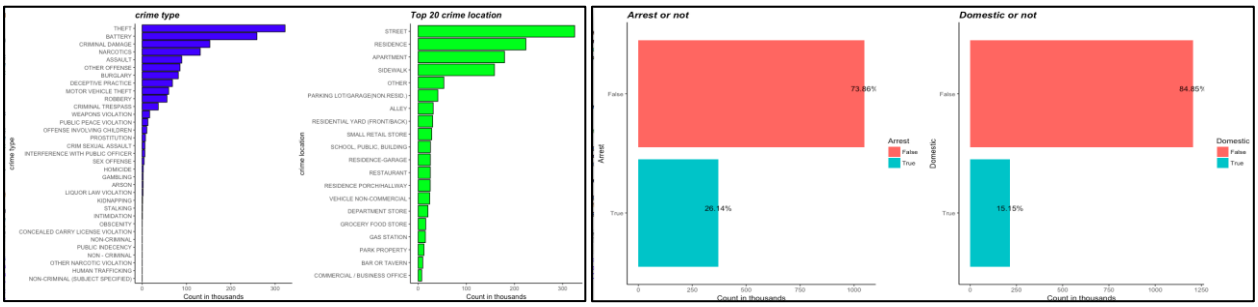


Fig.1. Data exploration is visualized using ggplot in R

2. Selection of Visualization Techniques:

Technique 1: Dot plots, Bar Plots and Tree map

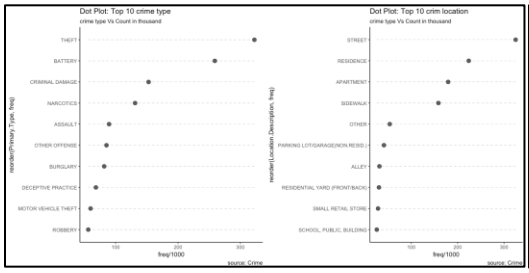


Fig.2. Dot plot of Crime data from 2012-2017 by Top 10 crime type and location

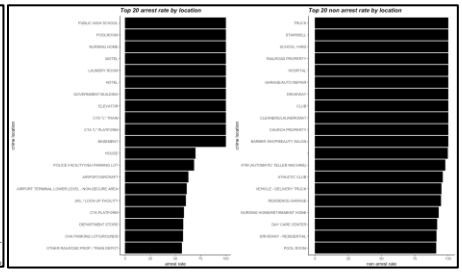


Fig.3. Top20 Arrest and non-Arrest crime location

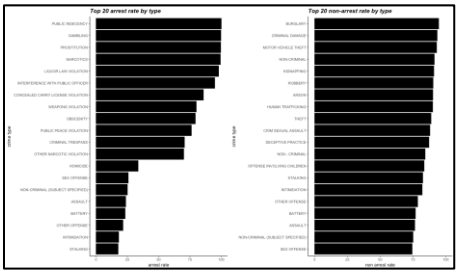


Fig.4. Top20 Arrest and non-Arrest crime type

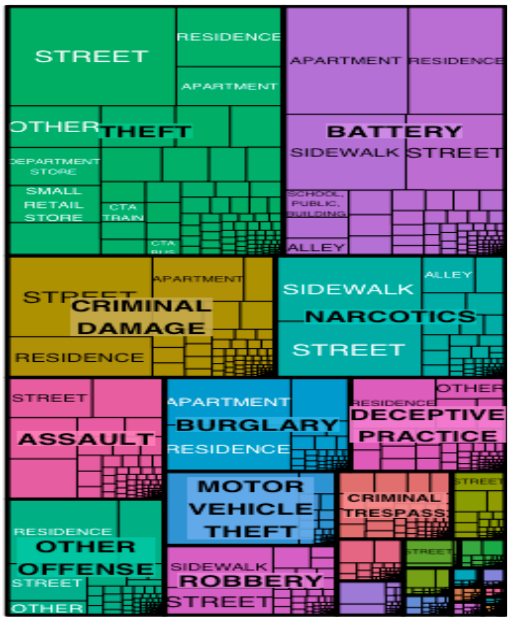


Fig.5. Tree map of crime type and location

For Technique 1, dot charts are created as a revised version of bar charts from previous section on the exploration part. From the dot charts, we observe that the top five crime types in Chicago from beginning of 2012 to end of 2016 are Theft, Battery, Criminal Damage, Narcotic, and Assault. And, the top five crime locations are Street, Resident, Apartment, sidewalk, etc.

Also, the bar charts with the top 20 arrests and non-arrest crime by location and type (percentage) are created to detect which location and type has the highest and lowest arrest rates.

Finally, the tree map indicates the type of crime to occur frequently in Chicago based on size and within each type of crime, we also observe the proportion of crime location.

Technique 2: Time Series (Time plot and Heat Map)

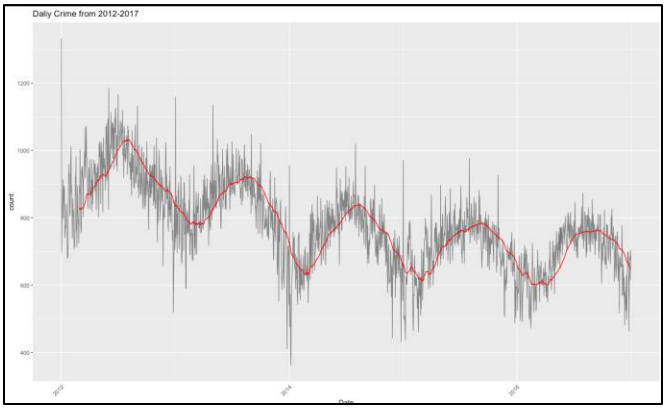


Fig.6.Time Plot of Crime count from beginning of 2012 to end of 2016

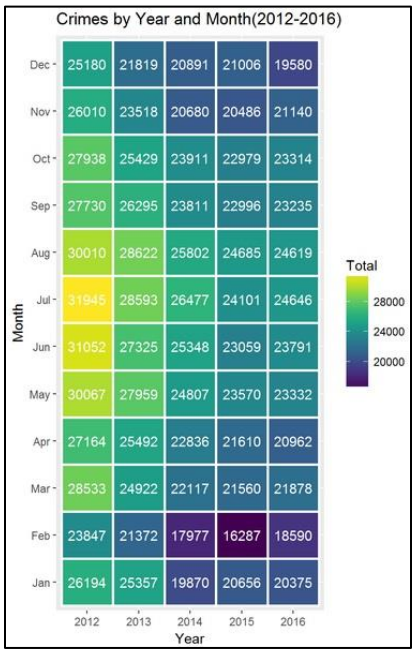


Fig.7. Heat map of Crime (Year vs month)

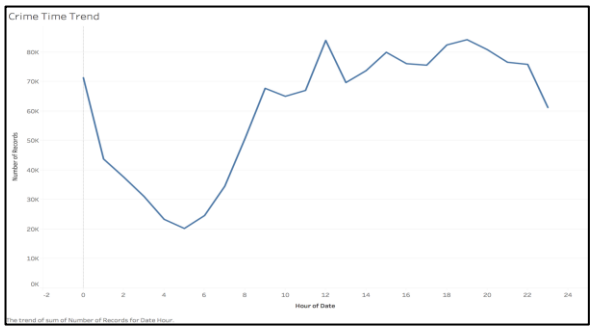


Fig.9. Crime trend by time

Fig.8. Time plot by type

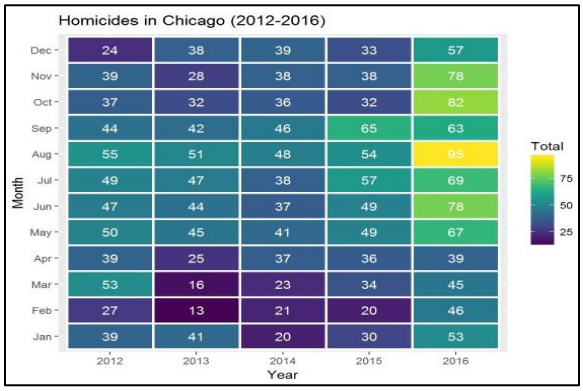


Fig.10. Heatmap for homicide(2012-2016)

From the time plot in Fig.6, with the smoothing line using moving average (set window size =60), we observe that crime in Chicago has a downward trend. Overall, the crime frequency has declined from over 800 to about 650 in average. Also, there seems to be higher crime frequency every June and July than other months of the year, which indicate seasonality. The declining trend in more detail and seasonality can also be detected from the heat map (Fig 7). The time plot in Fig.8 show that, in more detail, top criminal type, narcotics, has decreased dramatically, however, others dont to change much. Fig.9 gives the time of the day the overall crimes occur and it is observed that most of the crime occurs in broad daylight around noon time and after 10pm. Another interesting aspect was to study the homicide rates. Fig.10 show that there is a significant increase in the number of homicides in Chicago in 2016 compared to the previous years.

Technique 3: Geographic Plots (geographical false color heat map)

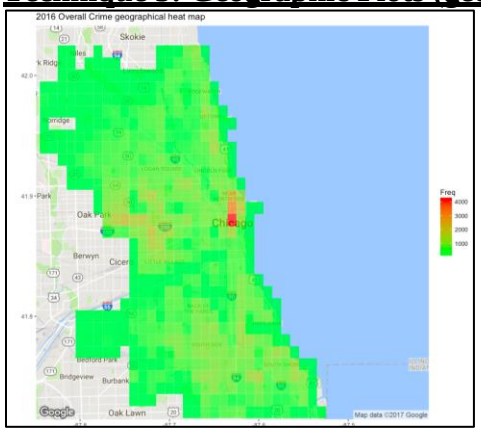


Fig.11 Geographical false color heat map on overall crime on 2016

Fig.12. Density map of the overall crime distribution

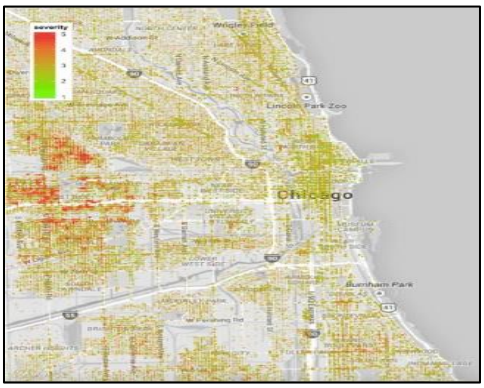


Fig.13. Severity of the crime by location

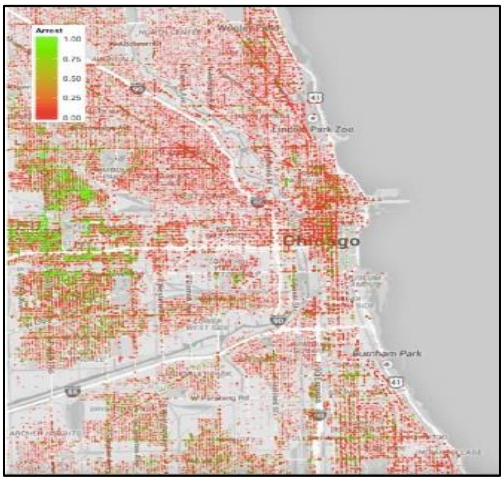


Fig.14. Arrest rate in Chicago

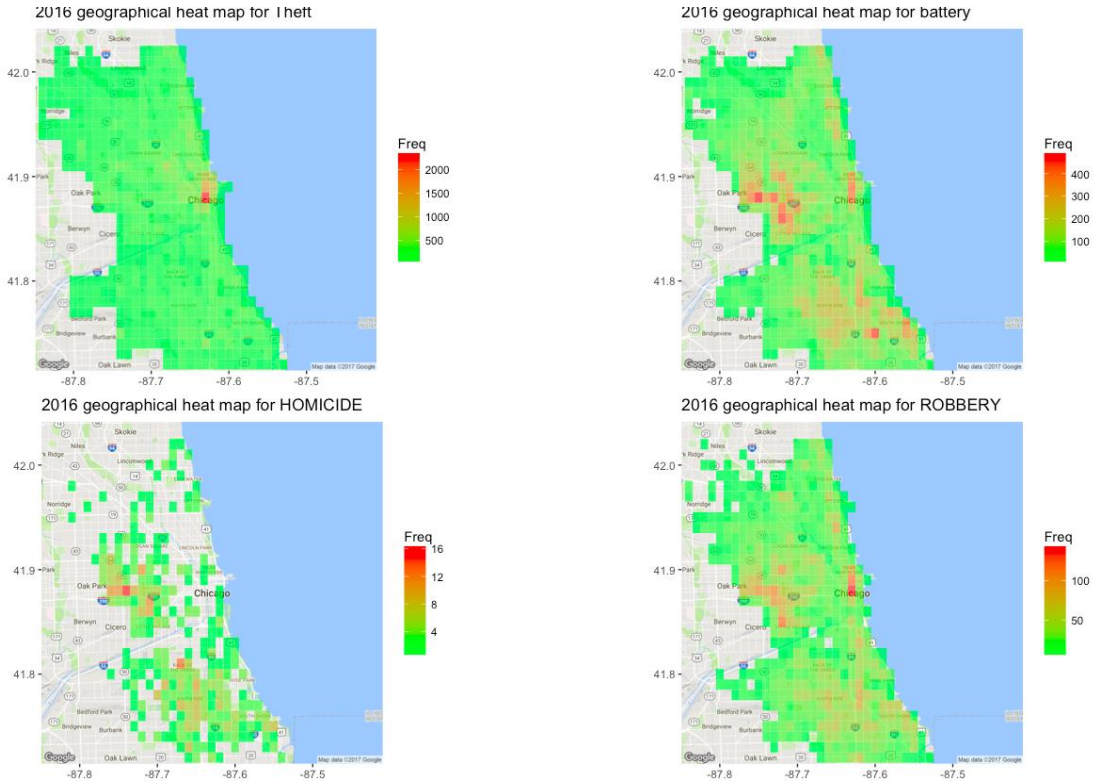


Fig.15. Geographical false color heat map by crime type on 2016

From the geographical heat map on 2016, we can see that overall, downtown Chicago has most frequent occurrence of crime. Also, by crime type, we can see that most of the theft cases happen in downtown Chicago. As for battery, it happens frequently at the area between downtown Chicago and Oak Park, and Southside. We also create the heat map for homicide and Robbery out of curiosity. From the map, we can see that homicide happen most frequently at the area between downtown Chicago and there are 3 hotspots for robbery, downtown Chicago, the area between uptown and Lincoln park, and the area between downtown and oak park.

Relationship between crimes:

Now that we have studied and explored the various crime types and the features associated with it, another interesting analysis would be to study the relationships between the crime types. We know that most of the time criminals are charged with multiple crime offence and it would be helpful if we understood the correlation between certain crime types. For instance, based on our analysis, we observe that if a weapons violation and arson crime is called in, most likely the same crime could involve kidnapping and homicide. Similarly, various

relationships between different correlated crimes and studied and the top relations are shown in Fig. 19. Fig 20 brings forward the visualization of these relations.

Confidence	Support	Lift	Rule
75.56%	24.82%	1.25	Weapons Violation & Arson ==> Kidnapping & Homicide
76.40%	24.82%	1.24	Weapons Violation & Kidnapping & Homicide ==> Arson
75.56%	24.82%	1.23	Kidnapping & Homicide ==> Crim Sexual Assault & Arson
74.73%	24.82%	1.23	Kidnapping & Arson ==> Weapons Violation & Homicide
52.31%	24.82%	1.22	Kidnapping & Crim Sexual Assault ==> Homicide & Arson
58.12%	24.82%	1.22	Homicide & Arson ==> Kidnapping & Crim Sexual Assault
75.56%	24.82%	1.22	Kidnapping & Homicide ==> Arson
58.12%	24.82%	1.73	sex offense & obscenity & homicide ==> gambling

Fig.19. Correlations between crime types

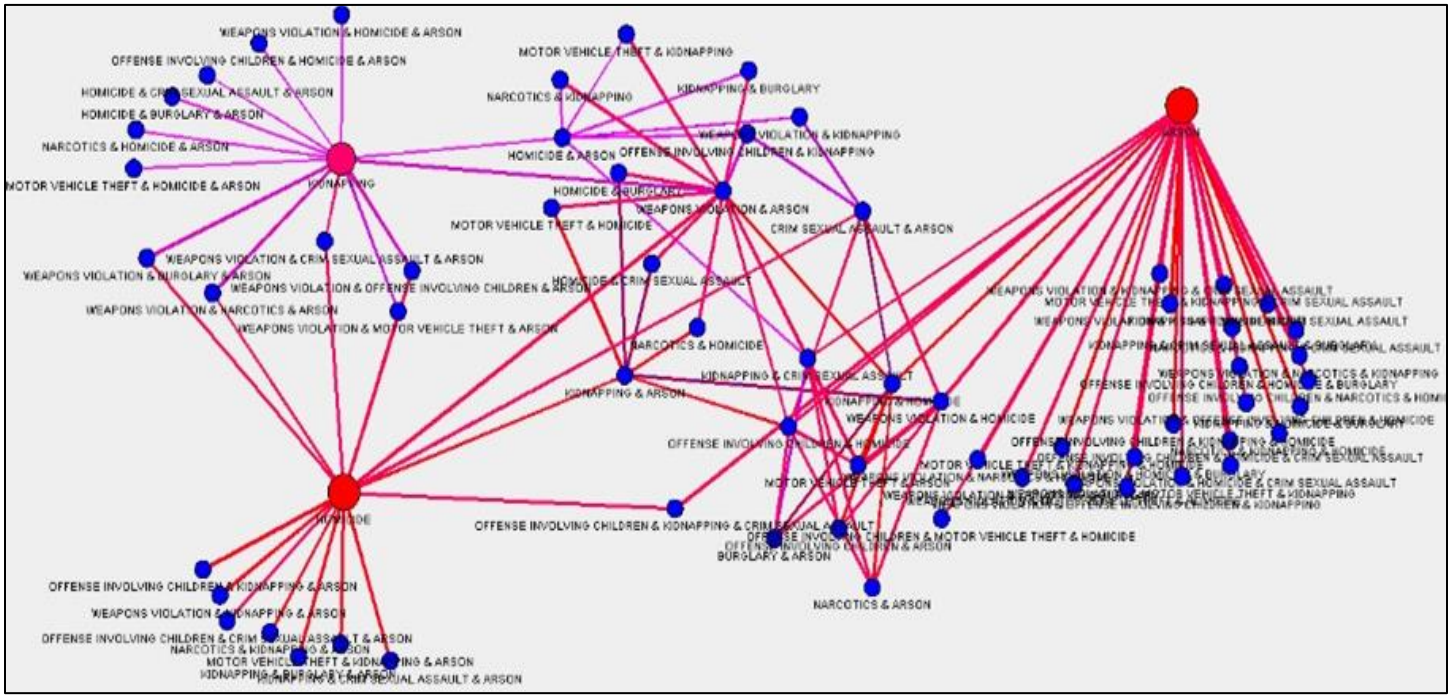


Fig.20. Visualization of the various relationships between correlated crime types

Results and Analysis:

From the above analysis, we observe that crime in Chicago has decreased over time. Yet, the crime rates in Chicago is the most discussed topic. Being one of the biggest cities it does have an impact on the overall crime numbers of the country. Thus, even though a lot of other smaller cities have higher crime rates, Chicago's crime rates are still talked about. This calls for increased police training and research to specific types of crime targeting certain locations.

We also observe that crime is most likely to happen in the months of June-August every year and is most likely to occur in downtown, south side and between oak park and downtown. From the analysis on the relationships between crime types, extensive training and research must also include studying the correlations between crime types to better prepare and handle situations.

The techniques we selected have been refined many times to help us bring out the patterns of data. To sum up, dot plot (better data ink ratio than box plot) of top 10 crime time location and type are used with reversing axis to make the graph less clutter and easier to read. The time plot handles the overlapping issues. Heat map enables us to choose the right color that bring out the high crime hotspot. Lastly, for geometrical heat map, false color method is used (green indicate low crime, red indicate crime hot spot) to help reader to see the crime hotspot easily.

As part of future studies, we would like to study on the narcotics a little further and the various relationships associated with it. Model building using techniques such as decision trees, neural networks, etc. may be explored to predict the percentage increase of the occurrence of crime in Chicago in the future years. From the visualization perspective, we would like to study interactive maps and graphing techniques and implement it towards our analysis.

Appendix:

A. Additional plots and graphs

A.1. Exploratory Analysis Visualization:

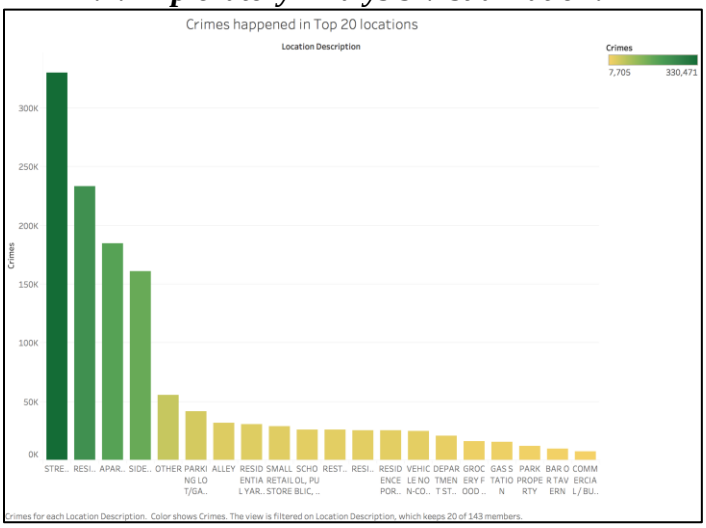


Fig.21. Top crimes in the top 20 locations of Chicago

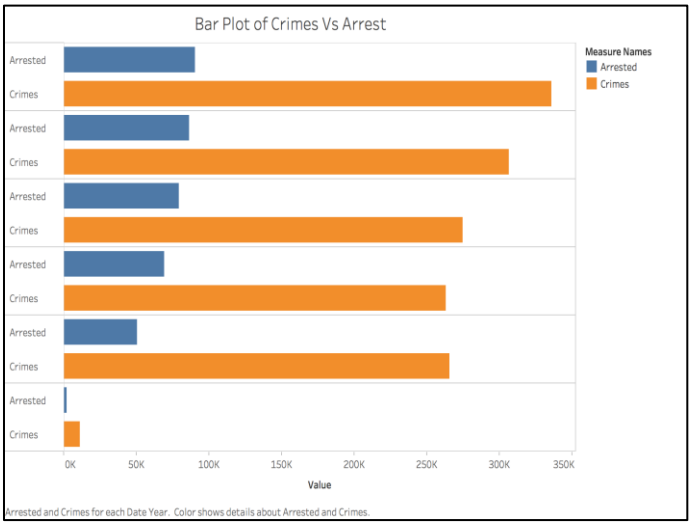


Fig.22. Bar plot of Arrests vs crimes

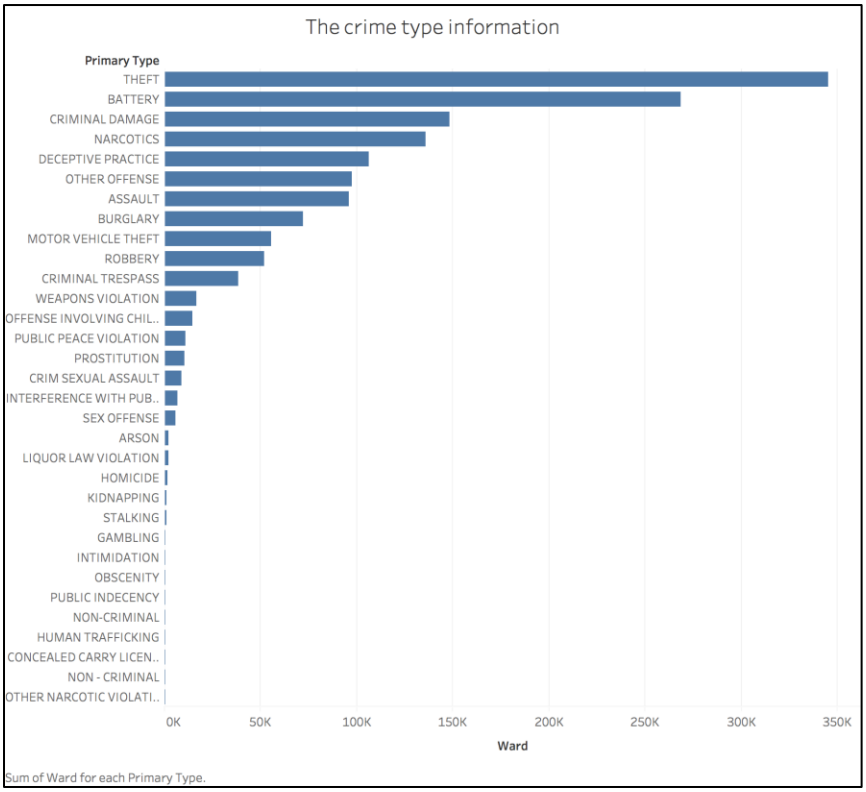


Fig.23. Crime type information

A.2.. Hexplot:

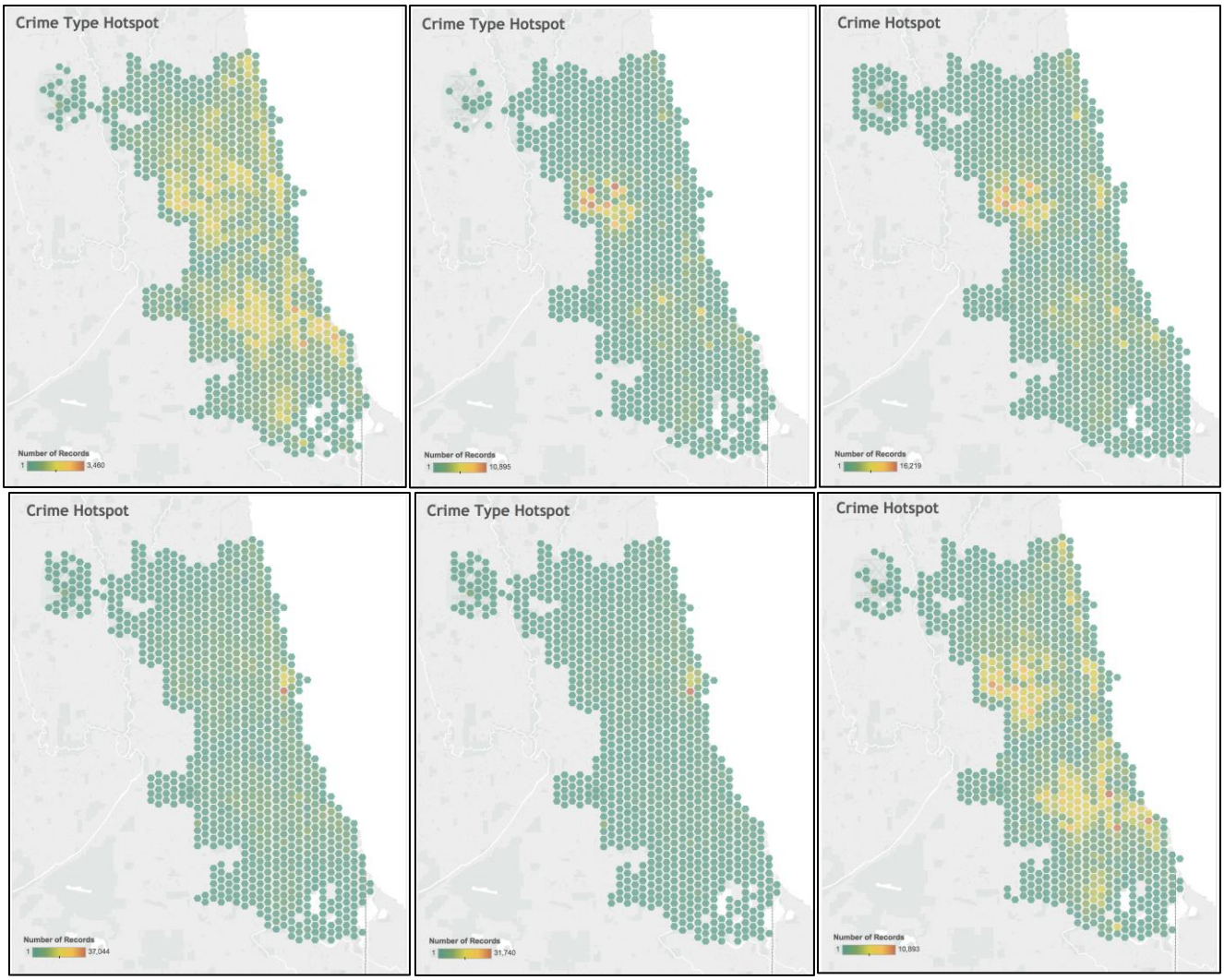


Fig.24. Criminal hotspot for various crime types

B. Reports and Notes

B.1. Exploration stage:

Summary of Work

Overall, I did help on first and second part of basic exploration and time series. During the first step, I helped cleaning data, and graphed couple charts to see little bit of the crime rate, type and location. Furthermore, I did analysis of the data from time series perspective, did analysis deeply in theft section. Such that what time at what place occur what type of theft the most? In addition, I did the presentation and helped organize the power point.

First, I did initial exploration of the data by making some dot chats, histograms, and bar charts on types of crime and frequency of crime occurs. This step initially gave me an idea about what I can do with this data (Even most them are not used in the final report). This clearly get us to understand the data better, such that which variable is saying what. Also, this step also gives me some idea such as "what technics we can use with this data".

Secondly, this is the step that I helped the most. We worked on time series data analysis, which gives us the best idea of how the time effect the data. In this part, I figured which type of crime occur the most by using both histogram and line graph. Under the most frequently occurred crime--theft, which type of thefts is the most popular one during each time of the day. In addition, by looking at yearly scale, I figured that most crime occur during summer June – August. Another observation we figured is that during February (may because of the cold) crime rate are less. Also, I dig into the data and found the afternoon to evening is when the thefts are out.

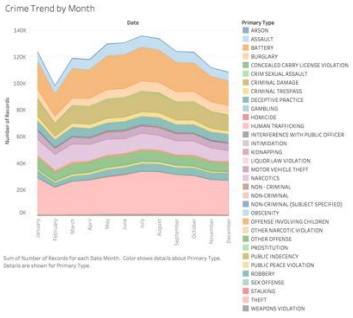
Summary of takeaways

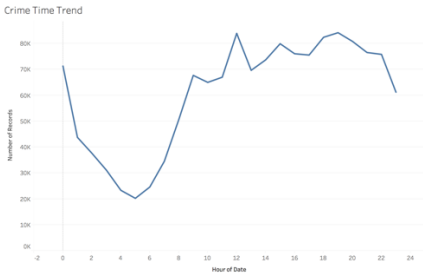
Throughout the project, I learned about the crime rate in Chicago:

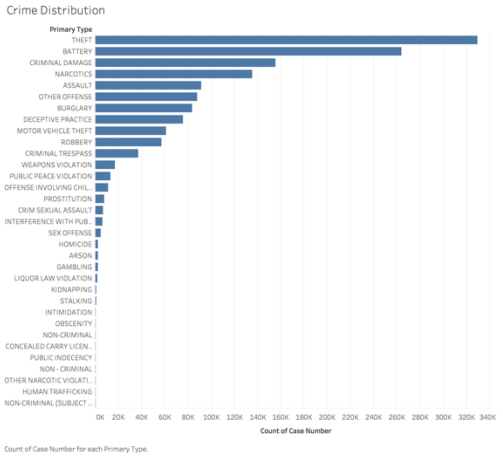
1. Geographically, the crime has higher frequency in west part of Chicago.

2. The Chicago area crime are more domestic type.
3. In addition, by digging into the data, I found out that the most “popular “crime is theft (under $500) and mostly spread in downtown area.
4. This also separated in two types of most occurred theft, during afternoon, retail thefts are more active than street thefts.
5. Also, we found the severity and rest rate are related, as well as some of the crimes are related as well.
6. Around 22% of the crime are theft! However, the police are not good at caching them.
7. Crime has declined, but the arrest rate has not increased, then there should be an interesting problem.

This are some graphs that are used in the final graph below on Time Series:







Crime by Year and Month

Month of D..	Date 2012	2013	2014	2015	2016
January	26,194	25,357	19,870	20,656	20,375
February	23,847	21,372	17,977	16,287	18,590
March	28,533	24,922	22,117	21,560	21,878
April	27,164	25,492	22,836	21,610	20,962
May	30,067	27,959	24,807	23,570	23,332
June	31,052	27,325	25,348	23,059	23,791
July	31,945	28,593	26,477	24,101	24,646
August	30,010	28,622	25,802	24,685	24,619
September	27,730	26,295	23,811	22,996	23,235
October	27,938	25,429	23,911	22,979	23,314
November	26,010	23,518	20,680	20,486	21,140
December	25,180	21,819	20,891	21,006	19,580

Number of Records: 16,287 – 31,945

Sum of Number of Records (color) broken down by Date Year vs. Date Month. The view is filtered on Date Year, which has multiple members selected.

B.2. Exploratory visualization:

【 What I did 】

1. I created exploratory visualization like bar chart of the crime type and location, which can help us decide which attribute and which level we need to go deeper. For the locations, we can clearly know what are the top ten locations that had more crimes in Chicago, and the top ten crime types that occurred frequently in Chicago.
2. Based on this situation, I also created the tree map by using two subdivisions--crime types with the location and the location with crime types. The division cells are sized by the number of crimes.

3. For our data, it is obvious that we can use time series technique to get the changing trend. We created the line chart for every year to show the difference of the arrest and non-arrest crime each year from 2012 to 2016. It can help us see more clearly the trend of crime over these years. I also created line chart by using daily crime number and monthly crime number.
4. For the technique 2, I draw the heat map to compare the number of crimes and the number of arrest. Using the heat map, we can easily conclude that the number of arrest is decreasing but the total number of crimes is still increasing. That is an interesting discovery for us.
5. By using technique 3, I draw hex plot for different crime types. Using color schema can show clearly which part of Chicago get more diverse types of crime.
6. I also use the location attribute to draw the world map to show the locations which had crime occurred frequently. Even bar chart can show the same result, but a world map is another way to show the information.
7. Review and revise the presentation power point, and give some suggestions to make our powerpoint

looks cleaner and simpler.

【 What I Learnt 】

1. Before I took this course, I did not know that R had many strong functions and could draw diverse graphs. Thanks to this course, I now surely have more knowledge about R.
2. During visualization project, I know how to choose variables and color scheme to make my graph look better and easier to understand. After creating a graph, I learned how to check the graph critically and objectively, I asked myself to try to get the information from the graph, and asked myself whether I can get information effectively and efficiently.
3. After this course and the group project, I know there is so much more to learn about R coding on my own end. Professor can only provide us some inspiration and instructions, the rest depends on our own exploration. Especially when I was doing my homework, I reviewed the class video and Googled the answer to finish my homework. This is an effective learning process for me to expand my knowledge.
4. Group project is really a great way to learn about how to communicate with classmates effectively and learn from each other. When I had problem with the group projects, my team member helped me by patiently telling me how to fix my problems.

B.3. Techniques and notes:

Contributions:

I cleaned the data, took care of missing values, and performed exploratory visualization - including the initial bar chart that helps us to decide what variables and levels we want to explore in more detail. After exploring the data (distribution) overall, for technique 1, I then performed the dot charts which have better data ink ratios (less clutter) than bar charts on top 10 crime types and top 20 crime locations. This helps the reader to decode and clearly see which crime types and locations happen most frequently. I also created a tree map on crime types and locations so it presents the categorical data with 2 levels.

As for technique 2, I wanted to focus on crime over the time (time series) so first I created a line plot on a overall crime frequency from 2012-2017 with moving average smoothing line that helps readers to detect the overall decreasing trend, and then I created a heat map by year and month. This also tells us that crime has decreased for the past 5 year, and there are recurring high crime rates in June-August every year. Lastly, before exploring technique 3, I designed a yearly line plot based on top 3 crime types on the same graph that provided the information in more detail to the reader.

For technique 3, I wanted to explore the data geographically, so I performed a Geographical false color heat map on crime frequency overall and on the selected 4 types of crime on 2016. I plot the heat map based on the area (count the crime frequency by the group of latitude and longitude rounded in 2 decimal points) so we can analysis and see the crime by area which provide more useful information visually rather than looking at the crime frequency by each longitude and latitude data point which will not be helpful for reader to decode since it will be very cluttered with too much overlapping. Also, I use “false color” criteria to show the hot spots

clearly. The color scheme I chose is green for low and red for high. I sent alpha to 0.6 so we can see the name of area on the map.

Summary/Reflection (What I learnt):

Throughout the project, I learnt how to take care of cluttering and overlapping. Also, there are many criteria when creating data visualization. I learnt that the best way to create the best visualization is to experiment and chose the one that best brings out the patterns. For example, the color ("false color" helped to bring out the hotspot), alpha (prevented overlapping), and the moving average window (helped readers to detect the pattern), and axis scale are all things we needed to try before deciding on the graph we want to use. In the future, I think it will be good to create an interactive geographical map and time series plot.

B.4. Time Series Analysis and report:

Contributions:

The inspiration for working with the Chicago Crime Data was aimed at challenging ourselves to working with a dataset relevant to the city we currently reside it and to understand the impact of various factors towards a crime occurring in Chicago. The motivation to work for this study was my idea. I obtained the dataset from Kaggle in the raw format and cleaned it with the help of my team mates for further analysis.

On understanding the structure of the data further based on the exploration analysis by my team mates, my focus of study was in the time series analysis of the dataset. Firstly, I determined the time series plot of the Chicago crimes and arrests to understand the increase or decrease in crime numbers over the years and the arrest rates for the corresponding years. This further led to the analysis of the top crime types by location to determine the time series plot of the number of crimes by the locations. The next field of analysis was to determine the time series plot of the crime series number by the top crime type. Another interesting aspect of the analysis was to determine the homicide number in Chicago over the years.

Once the time series analysis brought to light the initial picture of the Chicago crime analysis, I was further interested to study if there exists a relationship between crimes. The analysis had some interesting results, for instance, if there is a criminal charge for weapons violations and arson, the charge is most likely to be linked with kidnapping and homicide. This information helps the police train specifically not only by the top crime types and locations, but also conduct research on the top possible relationships between crimes. A visualization of this relationship best helped determine what are the possible combinations and how they come to be.

I was then interested to see how the crime data would look like over the geographic map of Chicago. I first determined the crime density and then proceeded to visualize the severity of the crime by location. From the previous time series analysis, I had come to the conclusion that the arrest rate was much lower than the crime rate and was interested to see the results on the map of Chicago with interesting visualization.

Summary/Reflection:

This visualization study helped to understand the Chicago Crime data better. Not only did the exploratory analysis plots help prioritize the areas to focus, but also helped understand the data structure better. The time series analysis plots gave a clear overview of the Crime data over the years by crime type, location and numbers. Plotting the relationship between different crime types helped understand how there exist correlation between different crime type that cant be avoided. Geographical mapping helped understand the data over the exact location on the map. Overall, the key concepts learnt throughout the course has been applied for visualization of the analysis. The takeaway from this study is the importance of not only plotting the various graphs in analysis but it is also equally important to determine the aesthetics of the visualization. Many obstacles encountered in this study in the form of cluttering, overlapping, labelling, etc. were avoided due to the concepts learnt through the course. Furthermore, I had a hands-on experience to understanding how color scheme brought out some of the most important patterns in the data and thus helped in the analysis. Overall, it brought to light the importance of visualization as part of analysis as it helps convey the information to the viewer better inorder to aid in further decision making.

C. Visualization Codes

C.1. Section 1:

```
library(ggplot2)

#load data
crime1617<- read.csv("Chicago_Crimes_2012_to_2017.csv")
head(crime1617)

#######################Preproccessing#####################################

#seperate the date colume to day, month, year
library(lubridate)
crime1617$Date <- as.Date(crime1617$Date, "%m/%d/%Y %I:%M:%S %p")
crime1617$Day <- factor(day(as.POSIXlt(crime1617$Date, format="%m/%d/%Y %I:%M:%S %p")))
crime1617$Month <- factor(month(as.POSIXlt(crime1617$Date, format="%m/%d/%Y %I:%M:%S
%p"), label = TRUE))
crime1617$Year <- factor(year(as.POSIXlt(crime1617$Date, format="%m/%d/%Y %I:%M:%S %p")))
crime1617$Weekday <- factor(wday(as.POSIXlt(crime1617$Date, format="%m/%d/%Y %I:%M:%S
%p"), label = TRUE))

#missing values?
colSums(is.na(crime1617))
#remove rows containg missing values
crime_cleaned<- na.omit(crime1617)
colSums(is.na(crime_cleaned))

#remove year 2017 due to lack of observation (preventing distoration)
crime_cleaned <- crime_cleaned[crime_cleaned$Year!='2017',]

#overview of columns
str(crime_cleaned)
head(crime_cleaned)



#write.table(crime_cleaned,"crime_cleaned.txt",sep="\t",row.names=FALSE)



###################### Preproccessing done ############################

#exploratory
library(gridExtra)
library(plyr)

# barplot by type
type_crime = arrange(count(crime_cleaned,c("Primary.Type")),(desc(freq)))

bar_type=ggplot(data=type_crime, aes(x=reorder(Primary.Type,freq),y=freq/1000)) +
  geom_bar(colour="black", fill="blue",stat = "identity") +
  ylab('Count in thousands')+coord_flip()+xlab("crime type")+ggtitle("crime
type")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))



# barplot by Location
location_crime = arrange(count(crime_cleaned,c("Location.Description")),(desc(freq)))

bar_loc=ggplot(data=location_crime[1:20,],
aes(x=reorder(Location.Description,freq),y=freq/1000)) +
```

```
  geom_bar(colour="black", fill="green",stat = "identity") +
  ylab('Count in thousands')+coord_flip()+xlab("crime location")+ggtitle("Top 20 crime
location")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))


grid.arrange(bar_type,bar_loc, ncol = 2)

# barplot by arrest
arrest_crime = arrange(count(crime_cleaned,c("Arrest")),(desc(freq)))

bar_arrest=ggplot(data=arrest_crime, aes(x=reorder(Arrest,freq),y=freq/1000,fill=Arrest))
+
  geom_bar(stat = "identity") +
  ylab('Count in thousands')+coord_flip()+xlab("Arrest")+ggtitle("Arrest or
not")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))+
  geom_text(aes(label = sprintf("%.2f%%", freq/sum(freq) * 100)),
            vjust = -.5)


# barplot by Domestic
dos_crime = arrange(count(crime_cleaned,c("Domestic")),(desc(freq)))

bar_Domestic=ggplot(data=dos_crime,
aes(x=reorder(Domestic,freq),y=freq/1000,fill=Domestic)) +
  geom_bar(stat = "identity") +
  ylab('Count in thousands')+coord_flip()+xlab("Domestic")+ggtitle("Domestic or
not")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))+
  geom_text(aes(label = sprintf("%.2f%%", freq/sum(freq) * 100)),
            vjust = -.5)



grid.arrange(bar_arrest,bar_Domestic, ncol = 2)


##technique 1

# dot Plot by type

library(ggplot2)
library(scales)
theme_set(theme_classic())

type_crime = arrange(count(crime_cleaned,c("Primary.Type")),(desc(freq)))

dot_type=ggplot(data=type_crime[1:10,], aes(x=reorder(Primary.Type,freq),y=freq/1000))  +
  geom_point(col="black", size=3, alpha= 0.7) +   # Draw points
  geom_segment(aes(x=Primary.Type,
                   xend=Primary.Type,
                   y=min(freq/1000),
                   yend=max(freq/1000)),
               linetype="dashed",
               size=0.1) +   # Draw dashed lines
  labs(title="Dot Plot: Top 10 crime type",
       subtitle="crime type Vs Count in thousand",
       caption="source: Crime") +
  coord_flip()


# dot Plot by location

location_crime = arrange(count(crime_cleaned,c("Location.Description")),(desc(freq)))
dot_loc=ggplot(data=location_crime[1:10,],
aes(x=reorder(Location.Description,freq),y=freq/1000))  +
```

```
  geom_point(col="black", size=3, alpha= 0.7) +   # Draw points
  geom_segment(aes(x=Location.Description,
                   xend=Location.Description,
                   y=min(freq/1000),
                   yend=max(freq/1000)),
               linetype="dashed",
               size=0.1) +   # Draw dashed lines
  labs(title="Dot Plot: Top 10 crim location",
       subtitle="crime type Vs Count in thousand",
       caption="source: Crime") +
  coord_flip()


grid.arrange(dot_type, dot_loc, ncol = 2)

# top 20 arrest rate location
ar_loc = arrange(count(crime_cleaned,c("Location.Description","Arrest")),(desc(freq)))
library(dplyr)

ar_locG=ar_loc%>%group_by(Location.Description)%>%
  mutate(per=paste0(round(freq/sum(freq)*100, 2), "%")) %>%
  ungroup
ar_locGT =ar_locG[ar_locG$Arrest == "True",]


ar_locGT$per=as.numeric(sub("%", "", ar_locGT$per))
ar_locGT = arrange(ar_locGT,(desc(per)))
head(ar_locGT)
bar_ar_loc=ggplot(data=ar_locGT[1:20,], aes(x=reorder(Location.Description,per),y=per)) +
  geom_bar(colour="black", fill="black",stat = "identity") +
  ylab('arrest rate')+coord_flip()+xlab("crime location")+ggtitle("Top 20 arrest rate by
location")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))

bar_ar_loc=bar_ar_loc+ scale_y_continuous(limits = c(0, 100))



# top 20 nonarrest rate location


ar_locGF =ar_locG[ar_locG$Arrest == "False",]


ar_locGF$per=as.numeric(sub("%", "", ar_locGF$per))
ar_locGF = arrange(ar_locGF,(desc(per)))
head(ar_locGF)
bar_F_Loc=ggplot(data=ar_locGF[1:20,], aes(x=reorder(Location.Description,per),y=per)) +
  geom_bar(colour="black", fill="black",stat = "identity") +
  ylab('non arrest rate')+coord_flip()+xlab("crime location")+ggtitle("Top 20 non arrest
rate by location")+theme(plot.title = element_text(color="Black", size=14,
face="bold.italic"))

bar_F_loc=bar_F_Loc+ scale_y_continuous(limits = c(0, 100))




grid.arrange(bar_ar_loc,bar_F_loc, ncol = 2)




#arrest top 10 type
```

```
ar_type = count_(crime_cleaned,c("Primary.Type","Arrest"))

ar_typeG=ar_type%>%group_by(Primary.Type)%>%
  mutate(per=paste0(round(n/sum(n)*100, 2), "%")) %>%
  ungroup
ar_typeGT =ar_typeG[ar_typeG$Arrest == "True",]
head(ar_typeGT)

ar_typeGT$per=as.numeric(sub("%", "", ar_typeGT$per))
ar_typeGT = arrange(ar_typeGT,(desc(per)))
head(ar_typeGT)
bar_ar_typeGT=ggplot(data=ar_typeGT[1:20,], aes(x=reorder(Primary.Type,per),y=per)) +
  geom_bar(colour="black", fill="black",stat = "identity") +
  ylab('arrest rate')+coord_flip()+xlab("crime type")+ggtitle("Top 20 arrest rate by
type")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))

bar_ar_typeGT=bar_ar_typeGT+ scale_y_continuous(limits = c(0, 100))
bar_ar_typeGT



# top 20 nonarrest rate type

ar_typeGF =ar_typeG[ar_typeG$Arrest == "False",]
head(ar_typeGF)

ar_typeGF$per=as.numeric(sub("%", "", ar_typeGF$per))
ar_typeGF = arrange(ar_typeGF,(desc(per)))
head(ar_typeGF)
bar_ar_typeGF=ggplot(data=ar_typeGF[1:20,], aes(x=reorder(Primary.Type,per),y=per)) +
  geom_bar(colour="black", fill="black",stat = "identity") +
  ylab('non arrest rate')+coord_flip()+xlab("crime type")+ggtitle("Top 20 non-arrest rate
by type")+theme(plot.title = element_text(color="Black", size=14, face="bold.italic"))

bar_ar_typeGF=bar_ar_typeGF+ scale_y_continuous(limits = c(0, 100))
bar_ar_typeGF



grid.arrange(bar_ar_typeGT,bar_ar_typeGF, ncol = 2)



##technique 2
#line plot by Date
library(plyr)
library(scales)
daily_crime = count_(crime_cleaned,"Date")
dCrime_lineplot = ggplot(data = daily_crime, aes(x = Date, y = n)) + ylab('count') +
xlab('Date')  + geom_line(col='blue')
dCrime_lineplot = dCrime_lineplot+scale_x_date(date_breaks="1 years",labels =
date_format("%m/%d/%Y"),limits = c(as.Date("2012-01-01"), as.Date("2016-12-31")))
dCrime_lineplot = dCrime_lineplot+theme(axis.ticks.x=element_blank(),axis.text.x =
element_text(angle = 45, hjust = 1)) +ggtitle("Daliy Crime from 2012-2017")
dCrime_lineplot

#with mv smotthing line
f = rep(1/60, 60)
f
daily_crime$mAve = filter(daily_crime$n, f, sides=1)

dCrime_lineplot = ggplot(data = daily_crime, aes(x = Date, y = n)) + ylab('count') +
xlab('Date')  + geom_line(color="black",alpha=0.5)
```

```
dCrime_lineplot = dCrime_lineplot+scale_x_date(date_breaks="1 years",labels =
date_format("%m/%d/%Y"),limits = c(as.Date("2012-01-01"), as.Date("2016-12-31")))
dCrime_lineplot = dCrime_lineplot+theme(axis.ticks.x=element_blank(),axis.text.x =
element_text(angle = 45, hjust = 1)) +ggtitle("Daliy Crime from 2012-2017")
dCrime_lineplot + geom_line(aes(x= Date, y=mAve), color="red")



#time plot by top 3 crime type yearly
ggplot(df,aes(x=x,y=y,group=col,colour=factor(col))) + geom_line()

dailytype_crime = count_(crime_cleaned,c("Year","Primary.Type"))
top3crimtype = subset(dailytype_crime, Primary.Type=="THEFT" | Primary.Type=="BATTERY"|
Primary.Type=="CRIMINAL DAMAGE")
ggplot(top3crimtype,aes(x = Year, y = n,group=Primary.Type,colour=Primary.Type)) +
geom_line()







#heat map

Monthly_crime =count_(crime_cleaned,c("Month","Year"))

crimes <- ggplot(Monthly_crime, aes(Year, Month, fill = n)) +geom_tile(size = 1, color =
"white")  +ggtitle("Crimes by Year and Month(2012-2016)")
crimes


##tree map
library(treemap)

crimemap =count_(crime_cleaned,c("Primary.Type","Location.Description"))


treemap(crimemap,
        index=c("Primary.Type", "Location.Description"),
        vSize="n",
        type="index")

##technique 3 :geographical crime at 2016-2017



library(ggmap)
install.packages("ggmap", type = "source")
chicago <- get_map(location = 'chicago', zoom = 11)
ggmap(chicago)

#subset for only 2016
crime_cleaned=subset(crime_cleaned,crime_cleaned$Year==2016)

#geographical heat map overall
LatLonCounts <- as.data.frame(table(round(crime_cleaned$Longitude,2),
round(crime_cleaned$Latitude,2)))
head(LatLonCounts)
str(LatLonCounts)

LatLonCounts$Long <- as.numeric(as.character(LatLonCounts$Var1))
LatLonCounts$Lat <- as.numeric(as.character(LatLonCounts$Var2))
```

```
LatLonCounts_cleaned <- subset(LatLonCounts, Freq > 0)
ggmap(chicago) + geom_tile(data = LatLonCounts_cleaned, aes(x = Long, y = Lat,fill =
Freq),alpha=0.7)+ theme(axis.title.y = element_blank(), axis.title.x =
element_blank())+labs(title = "2016 Overall Crime geographical heat map") +
scale_fill_gradient(low = "green", high = "red") +scale_alpha(range = c(0, 0.3), guide =
FALSE)

#take a look at top 2 crime type
levels(crime_cleaned$Primary.Type)
#geographical heatmap for theft and BATTERY
#theft
theft=subset(crime_cleaned,crime_cleaned$Primary.Type=="THEFT")

LatLonCounts_theft <- as.data.frame(table(round(theft$Longitude,2),
round(theft$Latitude,2)))


LatLonCounts_theft$Long <- as.numeric(as.character(LatLonCounts_theft$Var1))
LatLonCounts_theft$Lat <- as.numeric(as.character(LatLonCounts_theft$Var2))


LatLonCounts_theft_cleaned <- subset(LatLonCounts_theft, Freq > 0)
a=ggmap(chicago) + geom_tile(data = LatLonCounts_theft_cleaned, aes(x = Long, y = Lat,
fill = Freq), alpha=0.6)+ theme(axis.title.y = element_blank(), axis.title.x =
element_blank())+labs(title = "2016 geographical heat map for Theft")
a=a+ scale_fill_gradient(low = "green", high = "red")
a

#battery
battery=subset(crime_cleaned,crime_cleaned$Primary.Type=="BATTERY")

LatLonCounts_tbattery <- as.data.frame(table(round(battery$Longitude,2),
round(battery$Latitude,2)))


LatLonCounts_tbattery$Long <- as.numeric(as.character(LatLonCounts_tbattery$Var1))
LatLonCounts_tbattery$Lat <- as.numeric(as.character(LatLonCounts_tbattery$Var2))


LatLonCounts_tbattery_cleaned <- subset(LatLonCounts_tbattery, Freq > 0)
b=ggmap(chicago) + geom_tile(data = LatLonCounts_tbattery_cleaned, aes(x = Long, y = Lat,
fill = Freq), alpha = 0.6)+ theme(axis.title.y = element_blank(), axis.title.x =
element_blank())+labs(title = "2016 geographical heat map for battery")
b=b+ scale_fill_gradient(low = "green", high = "red")
b







#take look at some crime type we are interested
#geographical heatmap for homicide
HOMICID=subset(crime_cleaned,crime_cleaned$Primary.Type=="HOMICIDE")

LatLonCounts_HOMICID <- as.data.frame(table(round(HOMICID$Longitude,2),
round(HOMICID$Latitude,2)))


LatLonCounts_HOMICID$Long <- as.numeric(as.character(LatLonCounts_HOMICID$Var1))
LatLonCounts_HOMICID$Lat <- as.numeric(as.character(LatLonCounts_HOMICID$Var2))
```

```
LatLonCounts_HOMICID_cleaned <- subset(LatLonCounts_HOMICID, Freq > 0)
c=ggmap(chicago) + geom_tile(data = LatLonCounts_HOMICID_cleaned, aes(x = Long, y = Lat,
fill = Freq), alpha = 0.6)+ theme(axis.title.y = element_blank(), axis.title.x =
element_blank())+labs(title = "2016 geographical heat map for HOMICIDE")
c=c+ scale_fill_gradient(low = "green", high = "red")
c


#geographical heatmap for ROBBERY

ROBBERY=subset(crime_cleaned,crime_cleaned$Primary.Type=="ROBBERY")

LatLonCounts_ROBBERY <- as.data.frame(table(round(ROBBERY$Longitude,2),
round(ROBBERY$Latitude,2)))


LatLonCounts_ROBBERY$Long <- as.numeric(as.character(LatLonCounts_ROBBERY$Var1))
LatLonCounts_ROBBERY$Lat <- as.numeric(as.character(LatLonCounts_ROBBERY$Var2))


LatLonCounts_ROBBERY_cleaned <- subset(LatLonCounts_ROBBERY, Freq > 0)
d=ggmap(chicago) + geom_tile(data = LatLonCounts_ROBBERY_cleaned, aes(x = Long, y = Lat,
fill = Freq), alpha = 0.6)+ theme(axis.title.y = element_blank(), axis.title.x =
element_blank())+labs(title = "2016 geographical heat map for ROBBERY")
d=d+ scale_fill_gradient(low = "green", high = "red")
d


grid.arrange(a, b,c,d, ncol = 2, nrow=2)
```

C.2. Section 2:

```
---
output: html_document
---

<h1><center> Chicago Crime Report by Remya Kannan </center></h1>

<br>


<h4> Crime in Chicago</h4>
<p> Crime in Chicago has been tracked by the Chicago Police Department's Bureau of
Records since the beginning of the 20th century.Chicago was responsible for nearly half
of 2016's increase in homicides in the US.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.</p>



<br>
<h4>The importance of the dataset</h4>

1. Chicago Crime Numbers have been in the media for all the wrong reasons.
2. President widely used these numbers during his presidential campaign
3. He still talks about it.
4. These numbers are impacting the overall numbers of the country.

```{r setup, include=FALSE}

knitr::opts_chunk$set(echo = FALSE, warning = FALSE, fig.width = 6, fig.height = 5,
fig.align = "center")

library(highcharter)
library(ggplot2)
```
```

```
library(dplyr)
library(tidyr)
library(viridis)
library(plotly)
library(lubridate)
library(xts)
library(maps)
library(ggmap)
library(gridExtra)

setwd("/Users/remya/Desktop/Depaul_5th_quarter/CSC_465/Project")

chicagocrimes20122016 <- read.csv("Chicago_Crimes_2012_to_2017.csv")

## Reading the data
chicagocrimes20122016 <- chicagocrimes20122016[chicagocrimes20122016$Year!='2017',]


### Working with Date and Time
chicagocrimes20122016$Day <- factor(day(as.POSIXlt(chicagocrimes20122016$Date,
format="%m/%d/%Y %I:%M:%S %p")))
chicagocrimes20122016$Month <- factor(month(as.POSIXlt(chicagocrimes20122016$Date,
format="%m/%d/%Y %I:%M:%S %p"), label = TRUE))
chicagocrimes20122016$Year <- factor(year(as.POSIXlt(chicagocrimes20122016$Date,
format="%m/%d/%Y %I:%M:%S %p")))
chicagocrimes20122016$Weekday <- factor(wday(as.POSIXlt(chicagocrimes20122016$Date,
format="%m/%d/%Y %I:%M:%S %p"), label = TRUE))

chicagocrimes20122016$Date <- as.Date(chicagocrimes20122016$Date, "%m/%d/%Y %I:%M:%S %p")

## Creating timeseries
by_Date <- na.omit(chicagocrimes20122016) %>% group_by(Date) %>% summarise(Total = n())
tseries <- xts(by_Date$Total, order.by=as.POSIXct(by_Date$Date))


## Creating timeseries of arrests made
Arrests_by_Date <- na.omit(chicagocrimes20122016[chicagocrimes20122016$Arrest ==
'True',]) %>% group_by(Date) %>% summarise(Total = n())
arrests_tseries <- xts(Arrests_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

## BY Location
by_location <- chicagocrimes20122016 %>% group_by(Location.Description) %>%
summarise(Total = n()) %>% arrange(desc(Total))

## By Primary Type
by_type <- chicagocrimes20122016 %>% group_by(Primary.Type) %>% summarise(Total = n())
%>% arrange(desc(Total))


## By District
by_district <- chicagocrimes20122016 %>% group_by(District) %>% summarise(Total = n())
%>% arrange(desc(Total))


## By Primary Type
by_ward <- chicagocrimes20122016 %>% group_by(Ward) %>% summarise(Total = n()) %>%
arrange(desc(Total))


## By FBI Code
by_fbi <- chicagocrimes20122016 %>% group_by(FBI.Code) %>% summarise(Total = n()) %>%
arrange(desc(Total))

## By Arrest
```

```
by_arrest <- chicagocrimes20122016 %>% group_by(Arrest) %>% summarise(Total = n()) %>%
arrange(desc(Total))

## By Domestic
by_domestic <- chicagocrimes20122016 %>% group_by(Domestic) %>% summarise(Total = n())
%>% arrange(desc(Total))

## By Year
by_year <- chicagocrimes20122016 %>% group_by(Year) %>% summarise(Total = n()) %>%
arrange(desc(Total))

## Lat and Long of Crimes
LatLonCounts <- as.data.frame(table(round(chicagocrimes20122016$Longitude,2),
round(chicagocrimes20122016$Latitude,2)))
LatLonCounts$Long <- as.numeric(as.character(LatLonCounts$Var1))
LatLonCounts$Lat <- as.numeric(as.character(LatLonCounts$Var2))
LatLonCounts2 <- subset(LatLonCounts, Freq > 0)

## Lat and Long of Arrests
arrests_data <- na.omit(chicagocrimes20122016[chicagocrimes20122016$Arrest == 'True',])
LatLonArrestsCounts <- as.data.frame(table(round(arrests_data$Longitude,2),
round(arrests_data$Latitude,2)))
LatLonArrestsCounts$Long <- as.numeric(as.character(LatLonArrestsCounts$Var1))
LatLonArrestsCounts$Lat <- as.numeric(as.character(LatLonArrestsCounts$Var2))
LatLonArrestsCounts2 <- subset(LatLonArrestsCounts, Freq > 0)
```

```{r}

hchart(tseries, name = "Crimes") %>%
  hc_add_series(arrests_tseries, name = "Arrests") %>%
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the
Chicago Police Department", style = list(fontSize = "12px")) %>%
  hc_title(text = "Times Series plot of Chicago Crimes and Arrests") %>%
  hc_legend(enabled = TRUE)
```


From the above time series plot, we observe that the crime rate in Chicago has decreased in 2016 when compared to 2012. There is a clear indication that crime numbers oncrease during the middle of the year mostly during the summer months and drops towards the end/starting of the year (winter months). Another interesting note is the arrest rate is lower compared to the number of crimes. Additionally, crimes distibuted over months would be interesting to look into.


```{r}

hchart(arrests_tseries) %>%
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_title(text = "Times Series plot of Arrests made in Chicago (2012-2016)") %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the
Chicago Police Department", style = list(fontSize = "12px"))
```

````
<br>
<br>

The plot above shows the decrease in the arrest rate from 2012-2016!

<br>
<br>


```{r fig.height=7, fig.width=9}


arrests_count <- arrests_data %>% group_by(Year, Month) %>% summarise(Total = n())

arrests <- ggplot(arrests_count, aes(Year, Month, fill = Total)) +
  geom_tile(size = 1, color = "white") +
  scale_fill_viridis()  +
  geom_text(aes(label=Total), color='white') +
  ggtitle("Arrests by Year and Month(2012-2016)")


crime_count <- chicagocrimes20122016 %>% group_by(Year, Month) %>% summarise(Total = n())

crimes <- ggplot(crime_count, aes(Year, Month, fill = Total)) +
  geom_tile(size = 1, color = "white") +
  scale_fill_viridis()  +
  geom_text(aes(label=Total), color='white') +
  ggtitle("Crimes by Year and Month(2012-2016)")


grid.arrange(crimes, arrests, ncol = 2)

```


<br>

The heatmap indicates the decrease in the number of arrests by more than half between
2012 and 2016 but the crimes have not reduced at the same rate.
<br>

<!-- ```{r} -->

<!-- hchart(by_year, "column", hcaes(x = Year, y = Total, color = Year)) %>% -->
<!--   hc_colorAxis(stops = color_stops(n = 10, colors = c("#440154", "#21908C",
"#FDE725"))) %>% -->
<!--   hc_add_theme(hc_theme_darkunica()) %>% -->
<!--   hc_title(text = "Number of Crimes by Year") %>% -->
<!--   hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the
Chicago Police Department", style = list(fontSize = "12px")) %>% -->
<!--   hc_legend(enabled = FALSE) -->

<!-- ``` -->

<br>
<br>


```{r}

hchart(by_location[1:20,], "column", hcaes(x = Location.Description, y = Total, color =
Total)) %>%
  hc_colorAxis(stops = color_stops(n = 10, colors = c("#440154", "#21908C", "#FDE725")))
%>%
````
````

```
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_title(text = "Locations with most Crimes - Top 20") %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the
Chicago Police Department", style = list(fontSize = "12px")) %>%
  hc_legend(enabled = FALSE)
```


From the above time plot we observe that the streets are the most common location for crimes to occur followed by apartments and residence.

Let us understand how crime numbers vary in some of these top crime locations


```{r}
streets <- chicagocrimes20122016[chicagocrimes20122016$Location.Description=="STREET",]
## Creating timeseries
streets_by_Date <- na.omit(streets) %>% group_by(Date) %>% summarise(Total = n())
streets_tseries <- xts(streets_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

residence <-
chicagocrimes20122016[chicagocrimes20122016$Location.Description=="RESIDENCE",]
## Creating timeseries
residence_by_Date <- na.omit(residence) %>% group_by(Date) %>% summarise(Total = n())
residence_tseries <- xts(residence_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

apartment <-
chicagocrimes20122016[chicagocrimes20122016$Location.Description=="APARTMENT",]
## Creating timeseries
apartment_by_Date <- na.omit(apartment) %>% group_by(Date) %>% summarise(Total = n())
apartment_tseries <- xts(apartment_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

sidewalk <-
chicagocrimes20122016[chicagocrimes20122016$Location.Description=="SIDEWALK",]
## Creating timeseries
sidewalk_by_Date <- na.omit(sidewalk) %>% group_by(Date) %>% summarise(Total = n())
sidewalk_tseries <- xts(sidewalk_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

hchart(streets_tseries, name = "Streets") %>%
  hc_add_series(residence_tseries, name = "Residence") %>%
  hc_add_series(apartment_tseries, name = "Apartment") %>%
  hc_add_series(sidewalk_tseries, name = "Sidewalk") %>%
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the
Chicago Police Department", style = list(fontSize = "12px")) %>%
  hc_title(text = "Crimes in Streets/Residence/Apartment/Sidewalk") %>%
  hc_legend(enabled = TRUE)
```


The number of crimes have significantly reduced in the top crime locations.


```{r}

hchart(by_type, "column", hcaes(Primary.Type, y = Total, color = Total)) %>%
  hc_colorAxis(stops = color_stops(n = 10, colors = c("#440154", "#21908C", "#FDE725"))) %>%
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_title(text = "Crime Types") %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the Chicago Police Department", style = list(fontSize = "12px")) %>%
  hc_legend(enabled = FALSE)

```


Taking a look at the popular crimes, theft is the topmost followed by battery, narcotics and criminal damage.

Let us see how crime numbers vary in top crimes


```{r}

thefts <- chicagocrimes20122016[chicagocrimes20122016$Primary.Type=="THEFT",]
## Creating timeseries
thefts_by_Date <- na.omit(thefts) %>% group_by(Date) %>% summarise(Total = n())
thefts_tseries <- xts(thefts_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

battery <- chicagocrimes20122016[chicagocrimes20122016$Primary.Type=="BATTERY",]
## Creating timeseries
battery_by_Date <- na.omit(battery) %>% group_by(Date) %>% summarise(Total = n())
battery_tseries <- xts(battery_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

criminals <- chicagocrimes20122016[chicagocrimes20122016$Primary.Type=="CRIMINAL DAMAGE",]
## Creating timeseries
criminals_by_Date <- na.omit(criminals) %>% group_by(Date) %>% summarise(Total = n())
criminals_tseries <- xts(criminals_by_Date$Total, order.by=as.POSIXct(by_Date$Date))

narcotics <- chicagocrimes20122016[chicagocrimes20122016$Primary.Type=="NARCOTICS",]
## Creating timeseries
narcotics_by_Date <- na.omit(narcotics) %>% group_by(Date) %>% summarise(Total = n())
narcotics_tseries <- xts(narcotics_by_Date$Total, order.by=as.POSIXct(by_Date$Date))


hchart(thefts_tseries, name = "Thefts") %>%
  hc_add_series(battery_tseries, name = "Battery") %>%
  hc_add_series(criminals_tseries, name = "Criminal Damage") %>%
  hc_add_series(narcotics_tseries, name = "Narcotics") %>%
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the Chicago Police Department", style = list(fontSize = "12px")) %>%
  hc_title(text = "Crimes in Thefts/Battery/Criminal Damage/Narcotics") %>%
  hc_legend(enabled = TRUE)
```

```
```


From the above pllot we observe that the number of narcotics crimes have reduced and the number of thefts and battery crimes have remained the same.

How big is the increase in homicides in Chicago?

```{r}

homicide <- chicagocrimes20122016[chicagocrimes20122016$Primary.Type=="HOMICIDE",]

homicide_year <-  homicide %>% group_by(Year) %>% summarise(Total = n())

hchart(homicide_year, "column", hcaes(Year, Total, color = Year)) %>%
  hc_add_theme(hc_theme_darkunica()) %>%
  hc_title(text = "Homicide 2012-2016")  %>%
  hc_credits(enabled = TRUE, text = "Sources: City of Chicago Administration and the Chicago Police Department", style = list(fontSize = "12px"))
```



```{r}

homicide_count <- homicide %>% group_by(Year, Month) %>% summarise(Total = n())

ggplot(homicide_count, aes(Year, Month, fill = Total)) +
  geom_tile(size = 1, color = "white") +
  scale_fill_viridis()  +
  geom_text(aes(label=Total), color='white') +
  ggtitle("Homicides in Chicago (2012-2016)")
```


There is a significant increase in the number of homicides in Chicago in 2016 compared to previous years. However, there is very little development on the reasons for the increase in the number. Being one of the biggest cities, Chicago does have an impact on the overall crime numbers of the country. Hence the talk about crime even though a lot of other smaller cities have bigger crime rates.

C.3. Section 3:

Please find the Tableau code attached in the report folder.

C.4. Section 4:

Please find the Tableau code attached in the report folder.