Chicago's Crime Data Analysis and Visualization

- Jathin Nagesh(6328369), UtrechtUniversity

Introduction

The dataset about the Crimes in Chicago is maintained by the Chicago's police department and can be found in Chicago Data Portal's website. (https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2). Chicago is one of the largest cities in the U.S. The violent crime rate in Chicago was significantly higher compared to other U.S cities. In 2016, it was also responsible for almost half of the increase in homicides in the U.S.

In this analysis, I explore the crimes that took place in Chicago between 2013 to 2017 inclusive.The columns of interest in our data sets are:

1) Primary.Type: The primary description of the type of crime.
2) Description: A secondary description of the type of crime. It is also a subcategory of Primary.Type.
3) Year: Year in which the crime was committed.
4) Latitude: Latitude of the crime committed.
5) Longitude: Longitude of the crime committed.
6) Ward: The ward number suggests the ward in which the particular crime was commiteted. Chicago is divided into 50 wards.
7) Date: Date of the crime.
8) Location.Description: Place where the crime was committed i.e, street, residence etc

Visualizations and Observations:

```
library(ISLR)
library(tidyverse)
library(haven)
library(readxl)
library(plotrix)
library(plyr)
library(dplyr)
library(scales)
library(ggmap)
library(ggplot2)
options(scipen=999)

crimes_2013_orig <- read.csv(file = "C:/Users/jathi/Desktop/Chicago/Crimes_-
_2013.csv",header = TRUE, sep=",", stringsAsFactors = TRUE)
crimes_2014_orig <- read.csv(file = "C:/Users/jathi/Desktop/Chicago/Crimes_-
_2014.csv",header = TRUE, sep=",", stringsAsFactors = TRUE)
crimes_2015_orig <- read.csv(file = "C:/Users/jathi/Desktop/Chicago/Crimes_-
_2015.csv",header = TRUE, sep=",", stringsAsFactors = TRUE)
crimes_2016_orig <- read.csv(file = "C:/Users/jathi/Desktop/Chicago/Crimes_-
_2016.csv",header = TRUE, sep=",", stringsAsFactors = TRUE)
crimes_2017_orig <- read.csv(file = "C:/Users/jathi/Desktop/Chicago/Crimes_-
_2017.csv",header = TRUE, sep=",", stringsAsFactors = TRUE)

crimes_2013 <- select(crimes_2013_orig,
"Primary.Type","Description","Year","Latitude","Longitude","Ward","Date","Loc
ation.Description")
crimes_2014 <- select(crimes_2014_orig,
"Primary.Type","Description","Year","Latitude","Longitude","Ward","Date","Loc
ation.Description")
crimes_2015 <- select(crimes_2015_orig,
"Primary.Type","Description","Year","Latitude","Longitude","Ward","Date","Loc
ation.Description")
crimes_2016 <- select(crimes_2016_orig,
"Primary.Type","Description","Year","Latitude","Longitude","Ward","Date","Loc
ation.Description")
crimes_2017 <- select(crimes_2017_orig,
"Primary.Type","Description","Year","Latitude","Longitude","Ward","Date","Loc
ation.Description")

Crimes_2013_2014 <- join(crimes_2013, crimes_2014, by = "Year", type =
"full", match = "all")
Crimes_2013_2014_2015 <- join(Crimes_2013_2014, crimes_2015, by = "Year",
type = "full", match = "all")
```

```
Crimes_2013_2014_2015_2016 <- join(Crimes_2013_2014_2015, crimes_2016, by =
"Year", type = "full", match = "all")
Crimes_2013_2014_2015_2016_2017 <- join(Crimes_2013_2014_2015_2016,
crimes_2017, by = "Year", type = "full", match = "all")

no_of_crimes_2013 = nrow(crimes_2013)
no_of_crimes_2014 = nrow(crimes_2014)
no_of_crimes_2015 = nrow(crimes_2015)
no_of_crimes_2016 = nrow(crimes_2016)
no_of_crimes_2017 = nrow(crimes_2017)

head(Crimes_2013_2014_2015_2016_2017)
```

```
##                 Primary.Type                          Description Year
## 1                     HOMICIDE                  FIRST DEGREE MURDER 2013
## 2         CRIM SEXUAL ASSAULT                            PREDATORY 2013
## 3         CRIM SEXUAL ASSAULT                            PREDATORY 2013
## 4 OFFENSE INVOLVING CHILDREN      AGG SEX ASSLT OF CHILD FAM MBR 2013
## 5          DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300 2013
## 6          DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300 2013
##   Latitude Longitude Ward                   Date Location.Description
## 1 41.85430 -87.71370   24 05/01/2013 01:26:00 AM                ALLEY
## 2       NA        NA    6 02/10/2013 12:00:00 AM            RESIDENCE
## 3       NA        NA   11 05/10/2013 11:00:00 AM            RESIDENCE
## 4 41.89208 -87.76529   37 01/24/2013 12:00:00 AM            RESIDENCE
## 5       NA        NA   42 03/18/2013 10:25:00 AM
## 6       NA        NA   28 12/05/2013 01:00:00 AM            RESIDENCE
```

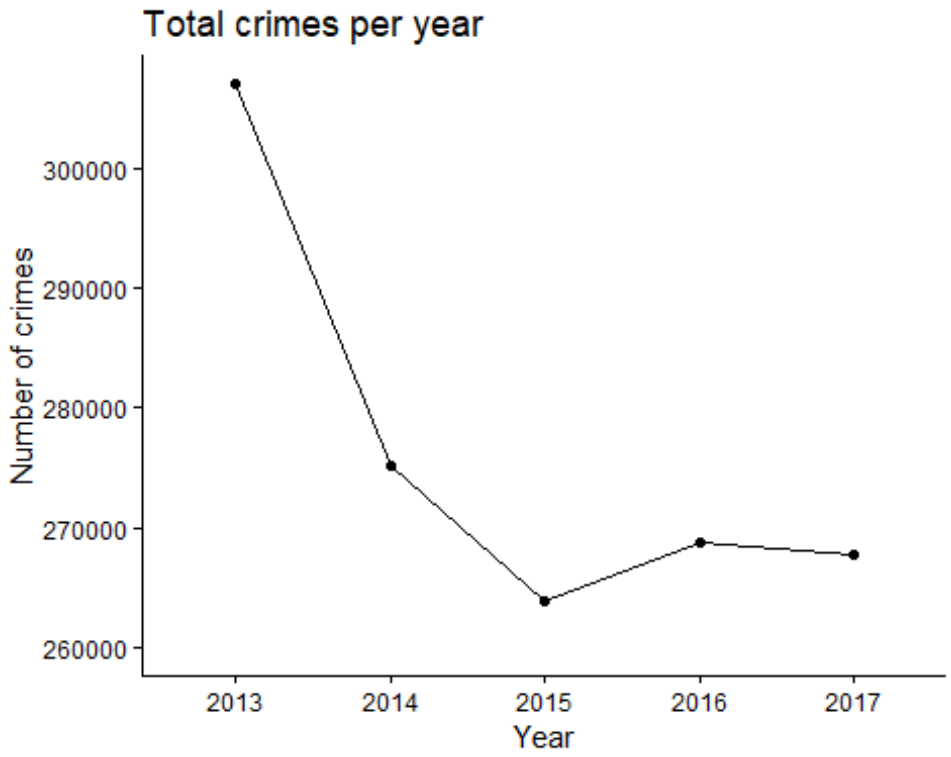
Visualization 1:

Firstly, I analyze the total crime count of each year from 2013 to 2017 to see whether there is any change of crime rate over the past 5 years.

```
Crimes_2013_2014_2015_2016_2017 <- na.omit(Crimes_2013_2014_2015_2016_2017)

year_crimes_df <- data.frame("Year"=c("2013", "2014", "2015", "2016",
"2017"), "No_of_crimes"=c(no_of_crimes_2013,
no_of_crimes_2014,no_of_crimes_2015, no_of_crimes_2016, no_of_crimes_2017))

ggplot(year_crimes_df, aes(x=Year, y=No_of_crimes, group = 1)) + geom_point()
+ geom_line()+labs(x = "Year", y = "Number of crimes", title = "Total crimes
per year")+ylim(260000,307030) + theme(panel.background = element_blank(),
plot.background = element_blank(), strip.background = element_blank(),
panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
axis.line = element_line(color = "black"), axis.text =  element_text(color =
"black"), axis.ticks = element_line(color = "black"))
```



Observation 1:

With this analysis we see that the crime rate in Chicago has decreased significantly over the years. This was mainly due to the significant drop in the crime rate from 2013 to 2014.

The crime rate in chicago dropped by 3.93% in 2017 over the average of the past 4 years.

```
Total_crimes_2013to2016 = no_of_crimes_2013 + no_of_crimes_2014 +
no_of_crimes_2015 + no_of_crimes_2016
Average_crimes_peryear_2013to2016 = Total_crimes_2013to2016/4
percentage_decline = ((Average_crimes_peryear_2013to2016 -
no_of_crimes_2017)/Average_crimes_peryear_2013to2016)*100
percentage_decline
```

```
## [1] 3.931248
```

Visualization 2:

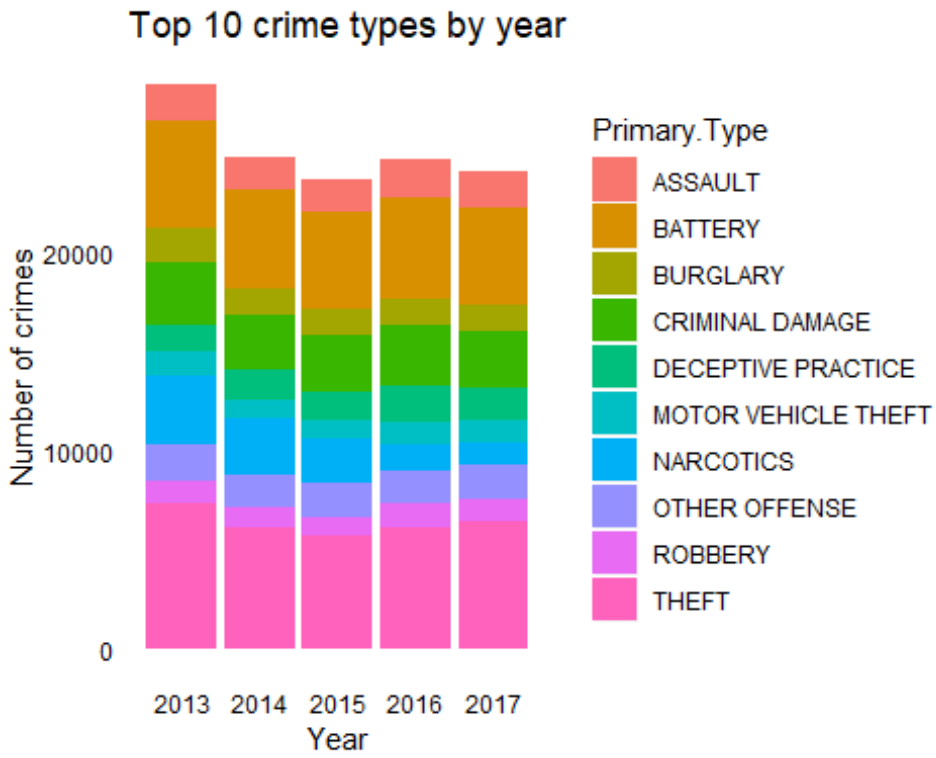
Below bar graph shows the 10 most commonly occuring crime type in each year(2013-2017).

```
crimes_by_type <- count((Crimes_2013_2014_2015_2016_2017))
total_crimes_by_type <- aggregate(crimes_by_type$freq,
by=list(Type=crimes_by_type$Primary.Type), FUN=sum)
```

```r
Top10_crimes <- top_n(total_crimes_by_type, 10)

Crimes_2013_2014_2015_2016_2017_top10 <- Crimes_2013_2014_2015_2016_2017 %>%
select(Primary.Type, Year)%>% filter(Primary.Type == Top10_crimes$Type)

ggplot(Crimes_2013_2014_2015_2016_2017_top10, aes(x=Year, fill =
Primary.Type)) + geom_bar() + labs(x = "Year", y = "Number of crimes", title
= "Top 10 crime types by year") + theme(panel.background = element_blank(),
plot.background = element_blank(), strip.background = element_blank(),
panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
axis.line = element_blank(), axis.text =  element_text(color = "black"),
axis.ticks = element_blank())
```



Observation 2:

We see that theft and battery are the two most commonly committed crime, over the past 5 years, in Chicago.

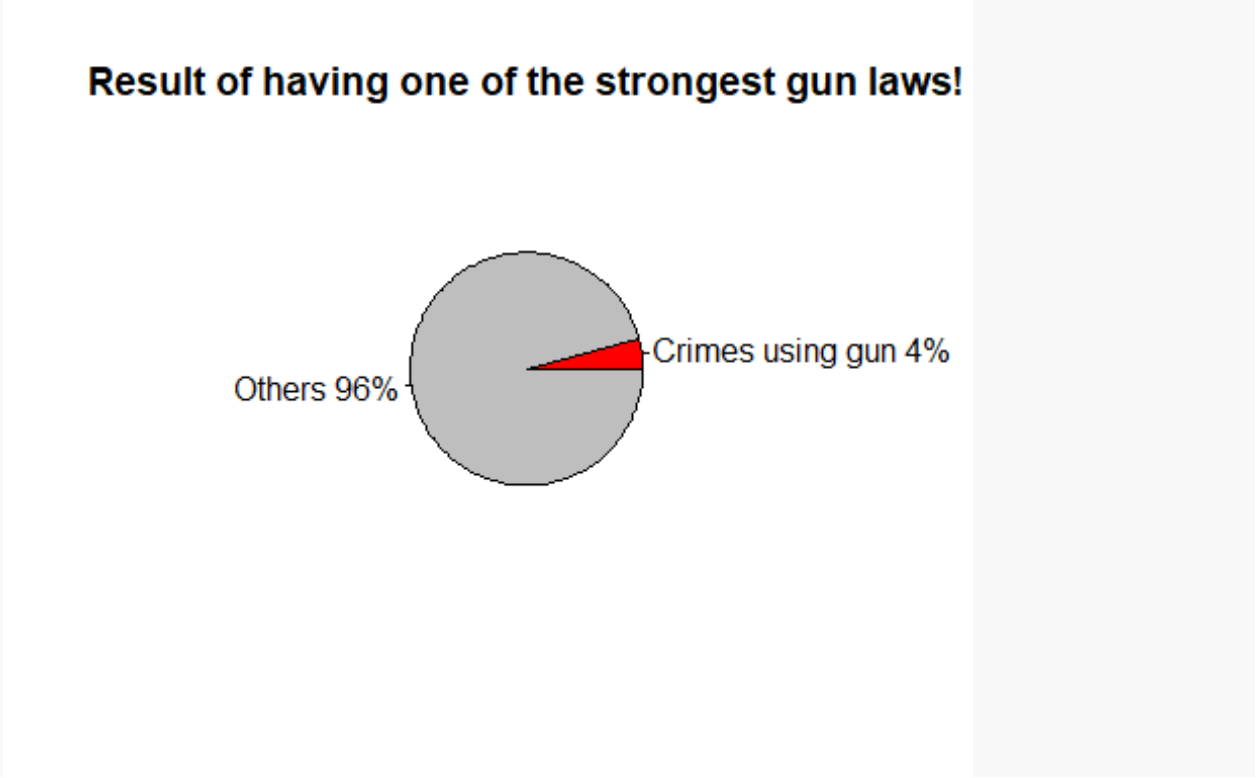
Visualization 3:

Chicago has one of the strongest gun laws in the U.S. But, how does it impact on the number of crimes committed using gun?

```r
crime_using_gun <- sum(grepl("HANDGUN| ARMED",
Crimes_2013_2014_2015_2016_2017$Description))
```

```
Total_crimes_without_gun <- Total_crimes_2013to2016 + no_of_crimes_2017 -
crime_using_gun
Total_crimes <- Total_crimes_2013to2016 + no_of_crimes_2017
gun_vs_nogun <- c(Gun = crime_using_gun, No_Gun  = Total_crimes_without_gun)
gun_vs_nogun <- as.data.frame(gun_vs_nogun)

pie(gun_vs_nogun$gun_vs_nogun, labels = c("Crimes using gun 4%","Others
96%"), main="Result of having one of the strongest gun laws!", col = c("red",
"grey"), radius = 0.6)
```



Observation 3:

It looks as though the strong gun laws do have an impact on the number of crimes committed using gun.

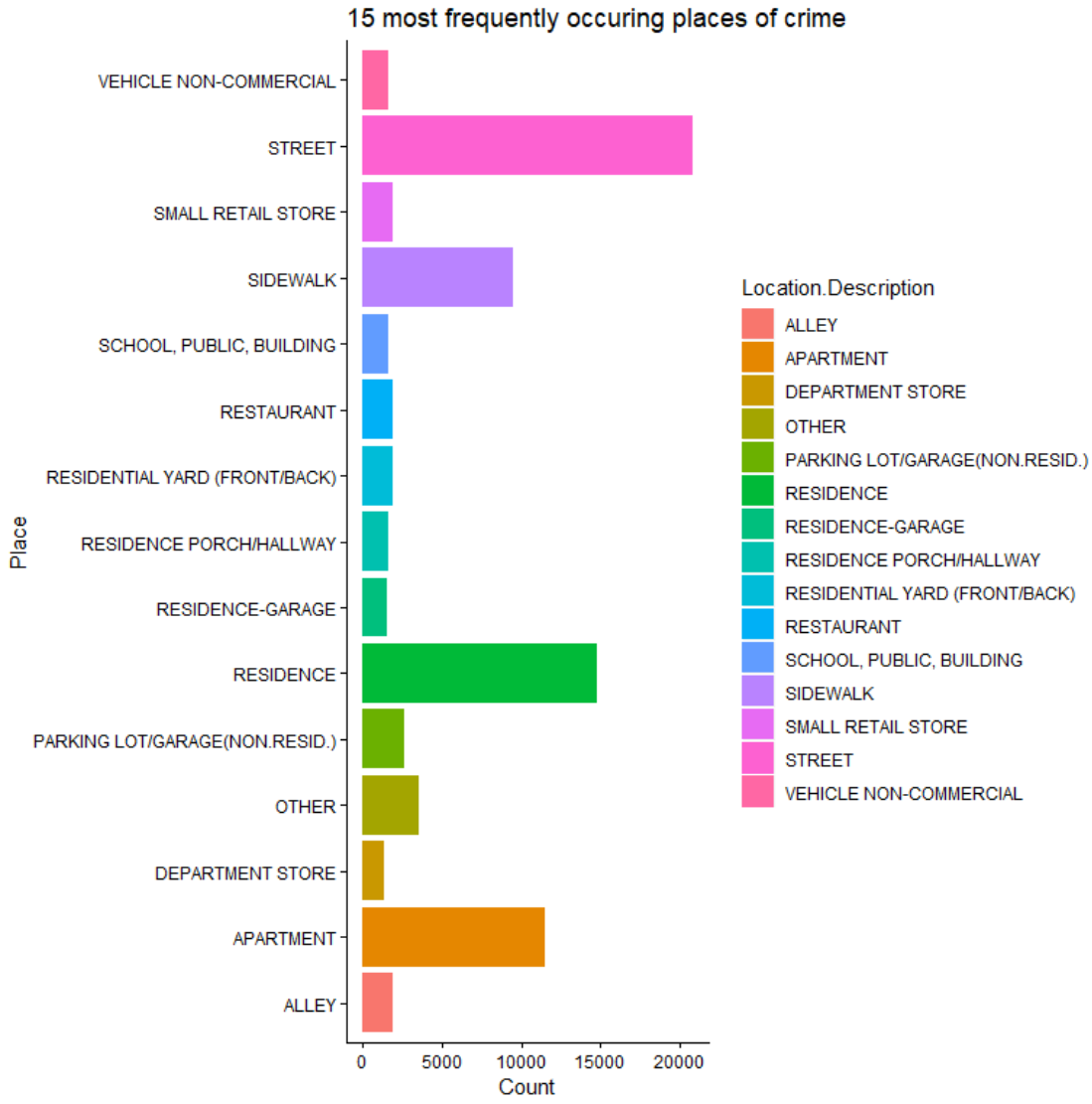
Visualization 4:

Crimes can happen in an N number of places. Let us now look at the most frequent place in which the crime is likely to take place.

```
total_crimes_by_place <- aggregate(crimes_by_type$freq,
by=list(Type=crimes_by_type$Location.Description), FUN=sum)
Top10_crimes_place <- top_n(total_crimes_by_place, 15)
```

```
Crime_Place <- Crimes_2013_2014_2015_2016_2017 %>%
select(Location.Description, Primary.Type) %>% filter(Location.Description ==
Top10_crimes_place$Type)

ggplot(Crime_Place,aes(x=Location.Description, fill=Location.Description)) +
geom_bar() + theme(panel.background = element_blank(), plot.background =
element_blank(), strip.background = element_blank(), panel.grid.major =
element_blank(), panel.grid.minor = element_blank(), axis.line =
element_line(color = "black"), axis.text =  element_text(color = "black"),
axis.ticks = element_line(color = "black")) + (labs(x = "Place", y = "Count",
title = "15 most frequently occuring places of crime"))+ coord_flip()
```



Observation 4:

From my analysis, it looks like the crime happens mostly in 4 places, i.e, streets, residence, sidewalk and apartment. Increasing the police patrol teams, providing security to the apartments can greatly help in the reduction of the crimes that take place in such places.

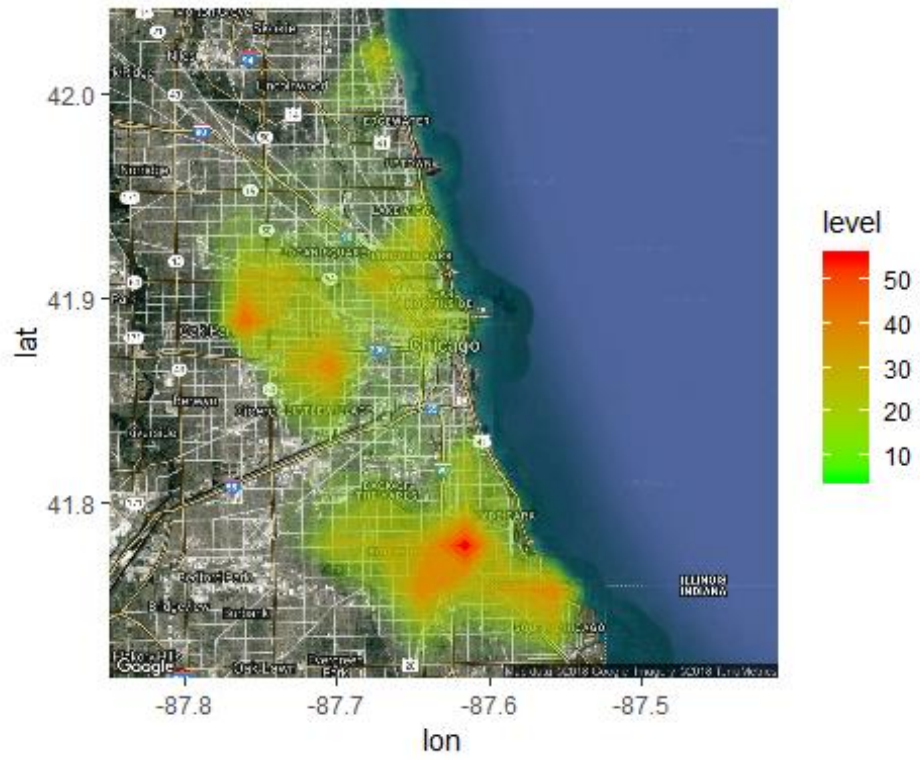
Visualization 5:

Next, with the help of the heat map I will analyze whether the number of crimes took place evenly across Chicago or are there any wards that are more prone to crime than the rest.

```
register_google(key = "YOUR KEY")
crime_prone_zone <- get_map(location = "Chicago", color = "color", zoom = 11,
maptype = "hybrid", source = "google")

latitude_longitude <- data.frame(long = c(
Crimes_2013_2014_2015_2016_2017$Longitude), lat = c(
Crimes_2013_2014_2015_2016_2017$Latitude))

ggmap(crime_prone_zone) + stat_density2d(data = latitude_longitude, na.rm =
TRUE, aes(x = long, y = lat, fill = ..level.., alpha = ..level..), size =
0.01, n=16,  geom = "polygon") + scale_fill_gradient(low = "green", high =
"red") + scale_alpha(range = c(0, 1), guide = FALSE)
```



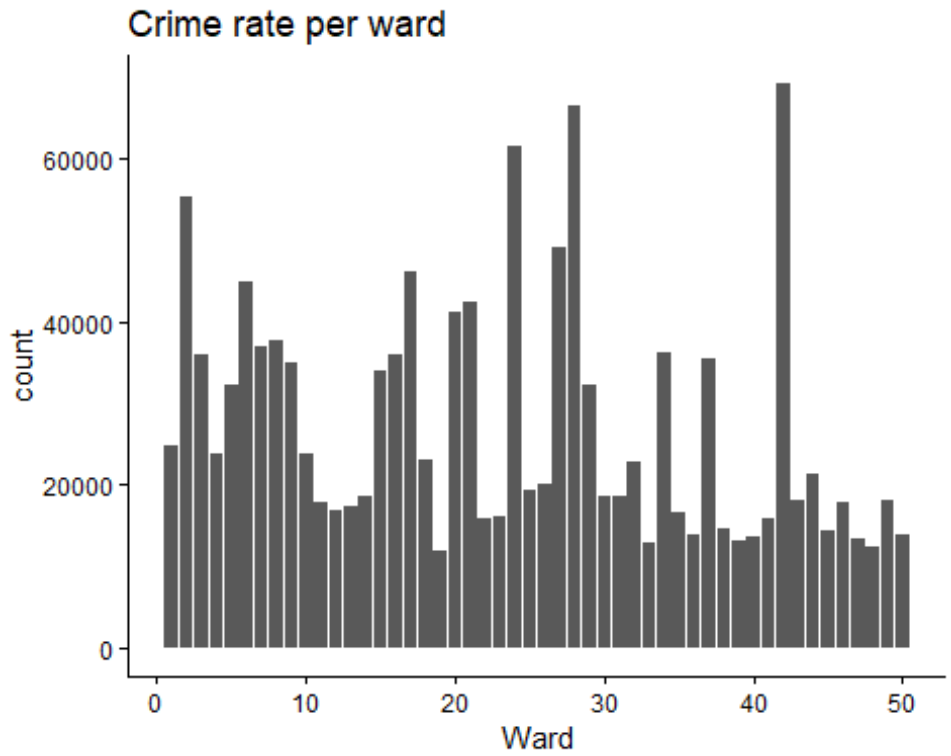
Observation 5:

. As you can see from the heat map, there are a few areas that are more vulnerable to crime than the rest. From such analysis, more police force can be deployed to such vulnerable areas to negate the crimes.

Visualization 6:

We will now look at the wards and the crime rate of each ward for further analysis.

```
ggplot(Crimes_2013_2014_2015_2016_2017,aes(x=Ward)) + geom_bar() +
theme(panel.background = element_blank(), plot.background = element_blank(),
strip.background = element_blank(), panel.grid.major = element_blank(),
panel.grid.minor = element_blank(), axis.line = element_line(color =
"black"), axis.text =  element_text(color = "black"), axis.ticks =
element_line(color = "black")) + (labs(x = "Ward", y = "count", title =
"Crime rate per ward"))
```



Observation 6:

From the above graph it is evident that ward 24, 28 and 42 are the wards in which most of crime takes place. We can further analyse the population, income, poverty rate and other factors to determine the cause of such high crime rate in these wards which is out of scope for this paper.

Visualization 7:

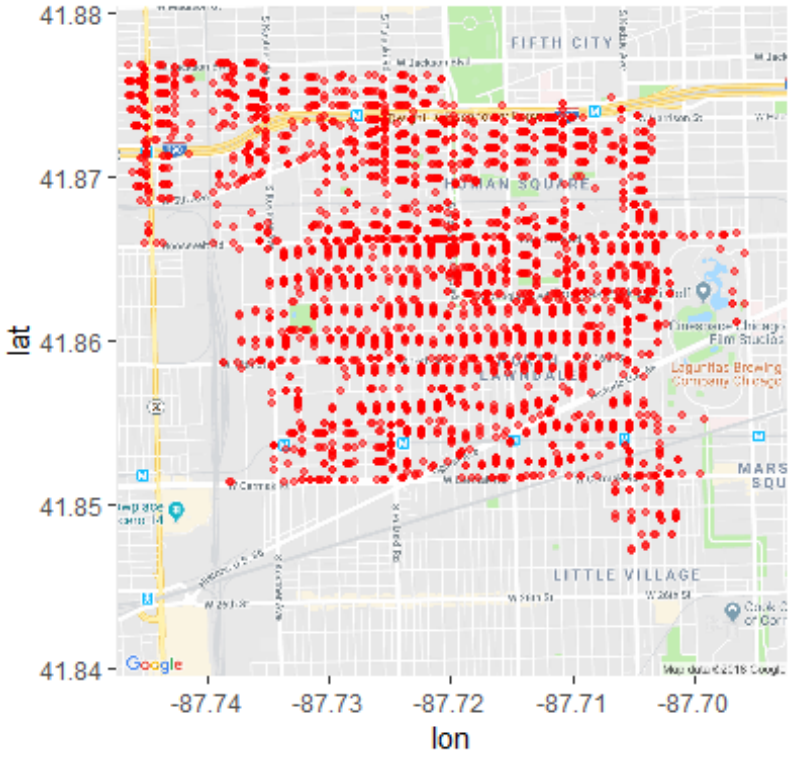
Ward 24, when compared to other wards, has one of the highest crime rates in Chicago. Let us now visualize the crimes committed using a gun in the ward 24 and find out the most vulnerable streets this ward.

```
Crimes_2013_2014_2015_2016_2017_gun <-
mutate(Crimes_2013_2014_2015_2016_2017, Gun = grepl("HANDGUN|
ARMED",Crimes_2013_2014_2015_2016_2017$Description))

Crimes_2013_2014_2015_2016_2017_gun <- Crimes_2013_2014_2015_2016_2017_gun
%>% select(Latitude,Longitude,Ward, Gun)  %>%  filter(Gun == TRUE & Ward ==
24)

crime_prone_zone <- get_map(location = c(lon =-87.72, lat = 41.86), color =
"color",zoom = 14, maptype = "roadmap", source = "google", messaging=FALSE)

ggmap(crime_prone_zone) + geom_point(data =
Crimes_2013_2014_2015_2016_2017_gun, na.rm = TRUE, aes(x = Longitude, y =
Latitude), color="red", size=1, alpha=0.5)
```



Observation 7:

From the above map of ward 24, we can visualize that most of the gun related crimes happen in and around the “HOMAN SQUARE”.

Visualization 8:

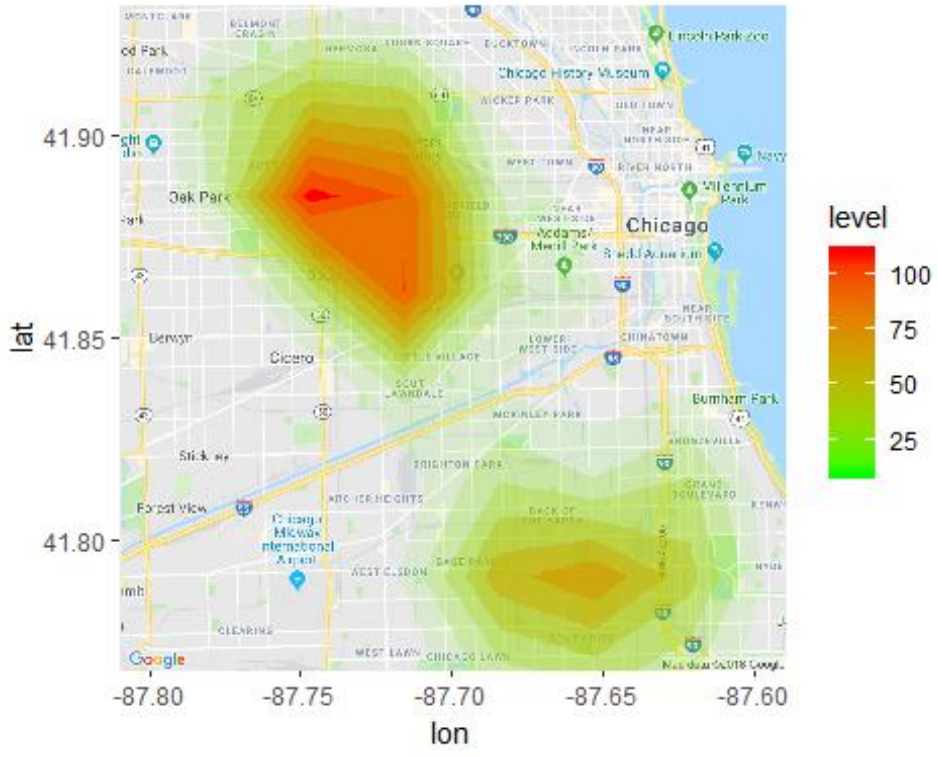
As explained earler, Homicide rate is significantly higher in the Chicago when compared to the rest of the U.S. We will now look at the homicide prone wards in Chicago.

```
Crimes_2013_2014_2015_2016_2017_homicide <-
mutate(Crimes_2013_2014_2015_2016_2017, Homicide =
grepl("HOMICIDE",Crimes_2013_2014_2015_2016_2017$Primary.Type))

Crimes_2013_2014_2015_2016_2017_homicide <-
Crimes_2013_2014_2015_2016_2017_homicide %>%
select(Latitude,Longitude,Primary.Type, Homicide)  %>%  filter(Homicide ==
TRUE)

Homicide_prone_zone <- get_map(location = c(lon =-87.7, lat = 41.85), color =
"color",
                 zoom = 12, maptype = "roadmap", source = "google",
messaging=FALSE)

ggmap(Homicide_prone_zone) + stat_density2d(data =
Crimes_2013_2014_2015_2016_2017_homicide, na.rm = TRUE, aes(x = Longitude, y
= Latitude, fill = ..level.., alpha = ..level..), size = 0.01,n=8,  geom =
"polygon") + scale_fill_gradient(low = "green", high = "red") +
scale_alpha(range = c(0, 1), guide = FALSE)
```



Observation 8:

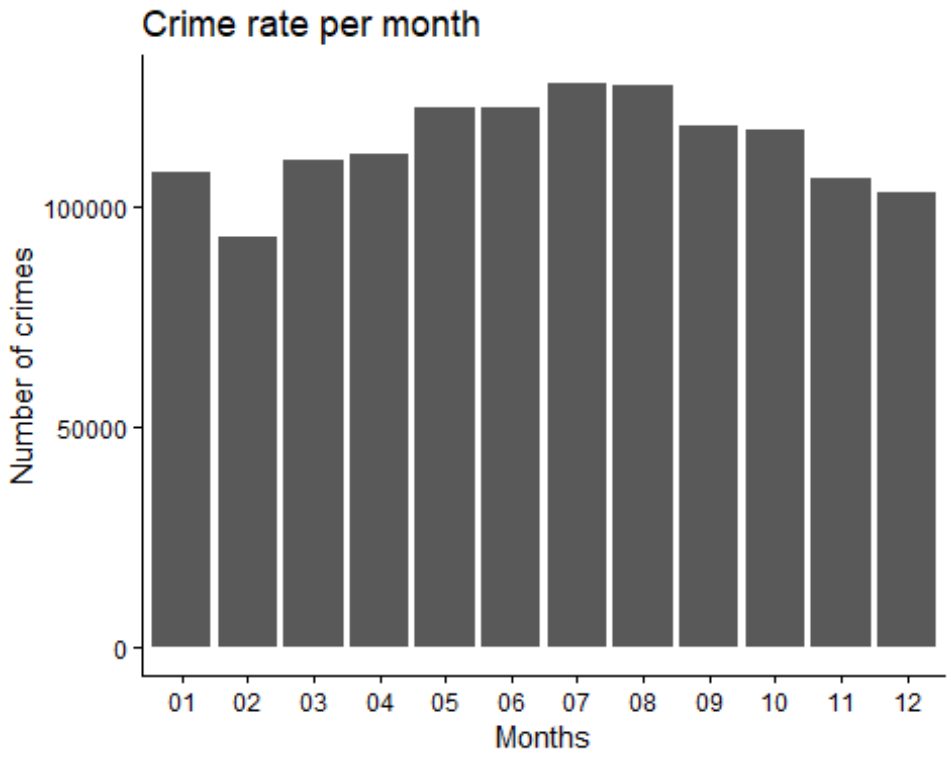
The below heat map suggests that there are very few wards in which most of the homicides take place every year. The color red on the map indicates a high homicide rate whereas color green indicates a low homicide rate.

Visualization 9:

Are there any seasons in which the rate of crime increases or decreases?

```
Crimes_2013_2014_2015_2016_2017$Date <-
as.Date(Crimes_2013_2014_2015_2016_2017$Date, format = "%m/%d/%Y")
Crimes_2013_2014_2015_2016_2017_month <-
mutate(Crimes_2013_2014_2015_2016_2017, Month =
format(Crimes_2013_2014_2015_2016_2017$Date,"%m"))
Crimes_2013_2014_2015_2016_2017_day <-
mutate(Crimes_2013_2014_2015_2016_2017, Day =
weekdays(Crimes_2013_2014_2015_2016_2017$Date))

ggplot(Crimes_2013_2014_2015_2016_2017_month, aes(x=Month))+geom_bar() +
theme(panel.background = element_blank(), plot.background = element_blank(),
strip.background = element_blank(), panel.grid.major = element_blank(),
panel.grid.minor = element_blank(), axis.line = element_line(color =
"black"), axis.text =  element_text(color = "black"), axis.ticks =
element_line(color = "black")) + labs(x = "Months", y = "Number of crimes",
title = "Crime rate per month")
```



Observation 9:

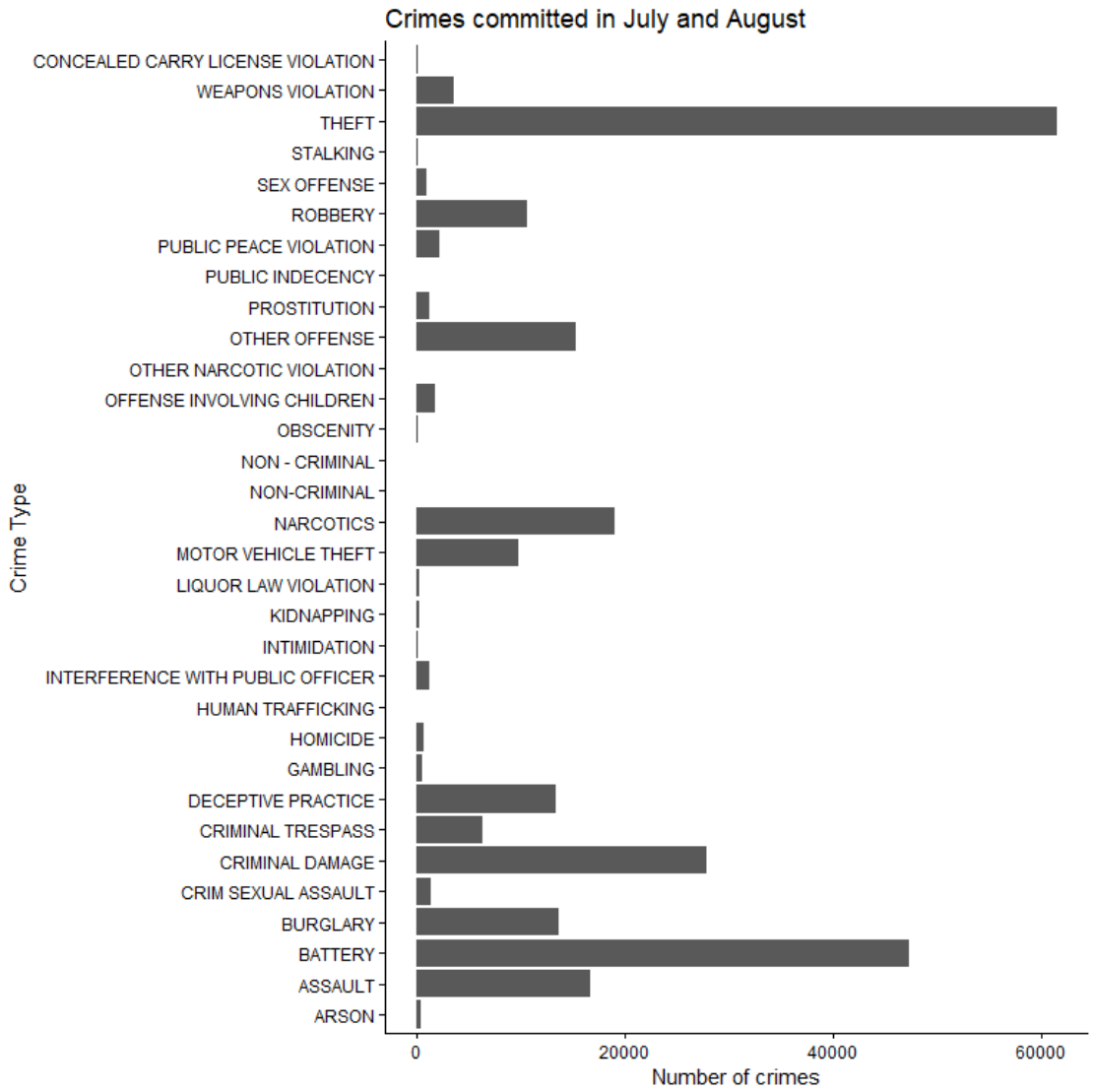
From the above graph we can conclude that the rate of crime increases during the summer, i.e, July and August; Someone who is planning to visit Chicago and is worried about being a victim of a crime, can plan his/her travel in the month of February as it is considered to be relatively safe.

Visualization 10:

Now, let us look at the type of crimes mostly committed in the month of July and August.

```
test <- Crimes_2013_2014_2015_2016_2017_month %>% select(Month, Primary.Type)
%>% filter(Month == "07" | Month == "08")

ggplot(test, aes(x=test$Primary.Type))+geom_bar()+coord_flip() +
theme(panel.background = element_blank(), plot.background = element_blank(),
strip.background = element_blank(), panel.grid.major = element_blank(),
panel.grid.minor = element_blank(), axis.line = element_line(color =
"black"), axis.text =  element_text(color = "black"), axis.ticks =
element_line(color = "black")) + labs(x = "Crime Type", y = "Number of
crimes", title = "Crimes committed in July and August")
```



Observation 10:

Theft is the most common crime committed in this period. We can hypothesize that the theft rate has partially to do with number of people going on vacations during the summer season.

Visualization 11:

Does the chance of a crime happening equal across all 7 days in a week or are there any specific days in which a crime is more likely to happen?

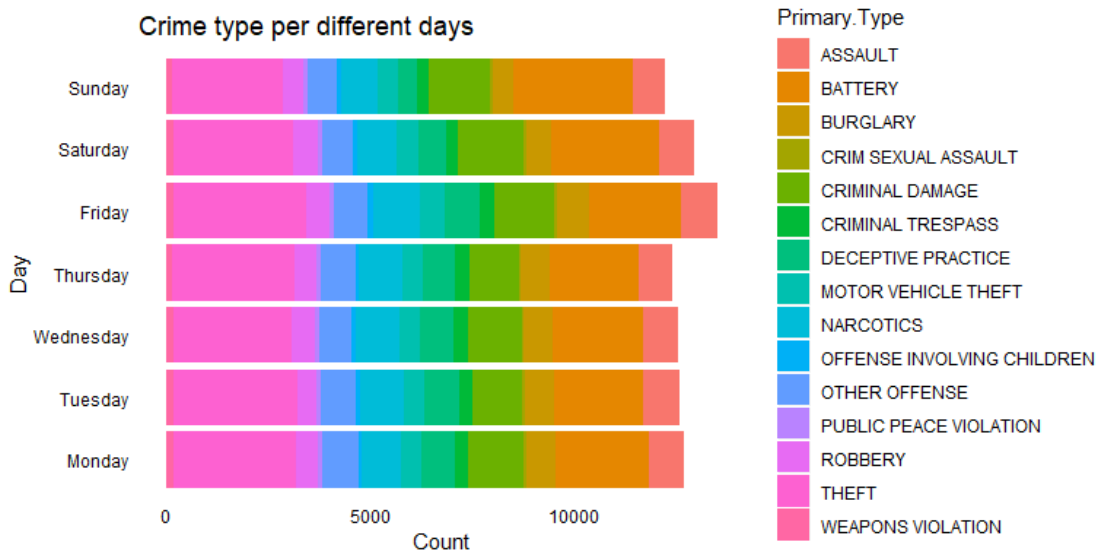
```
Crimes_2013_2014_2015_2016_2017_day$Day  =
factor(Crimes_2013_2014_2015_2016_2017_day$Day, levels=c("Monday", "Tuesday",
```

```
"Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))

Top15_crimes <- top_n(total_crimes_by_type, 15)

crime_count_per_day <- Crimes_2013_2014_2015_2016_2017_day %>% select(Day,
Primary.Type) %>% filter(Primary.Type == Top15_crimes$Type)

ggplot(crime_count_per_day, aes(x=Day, fill = Primary.Type)) + geom_bar() +
labs(x = "Day", y = "Count", title = "Crime type per different days") +
theme(panel.background = element_blank(), plot.background = element_blank(),
strip.background = element_blank(), panel.grid.major = element_blank(),
panel.grid.minor = element_blank(), axis.line = element_blank(), axis.text =
element_text(color = "black"), axis.ticks = element_blank()) + coord_flip()
```



Observation 11:

From my analysis, the crime is most likely to happen on a Friday.

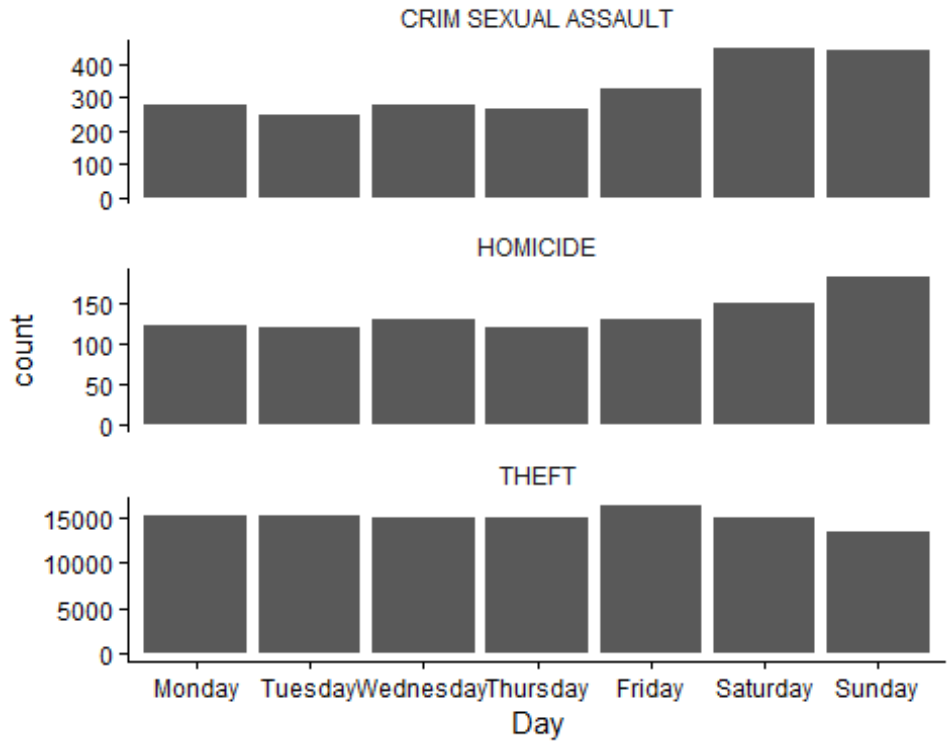
Visualization 12:

Let us now look at randomly selected 3 crime types and the most likely days in which they can occur.

```
crime_day <- Crimes_2013_2014_2015_2016_2017_day %>% select(Day,
Primary.Type) %>% filter(Primary.Type == c("HOMICIDE", "CRIM SEXUAL ASSAULT",
"THEFT"))

ggplot(crime_day, aes(x=Day))+geom_bar()+facet_wrap(~Primary.Type, 3, scales
= "free_y")+theme(panel.background = element_blank(), plot.background =
element_blank(), strip.background = element_blank(), panel.grid.major =
element_blank(), panel.grid.minor = element_blank(), axis.line =
```

```
element_line(color = "black"), axis.text =  element_text(color = "black"),
axis.ticks = element_line(color = "black"))
```



Observation 12:

Sexual assaults and homicides are most likely to happen over the weekend, whereas theft is more like to happen on a Friday.

Conclusion

With such data analysis and visualization, we can learn a lot about the crimes that happen in one's city and guide/help the law enforcements agencies to significantly reduce the number of crimes.