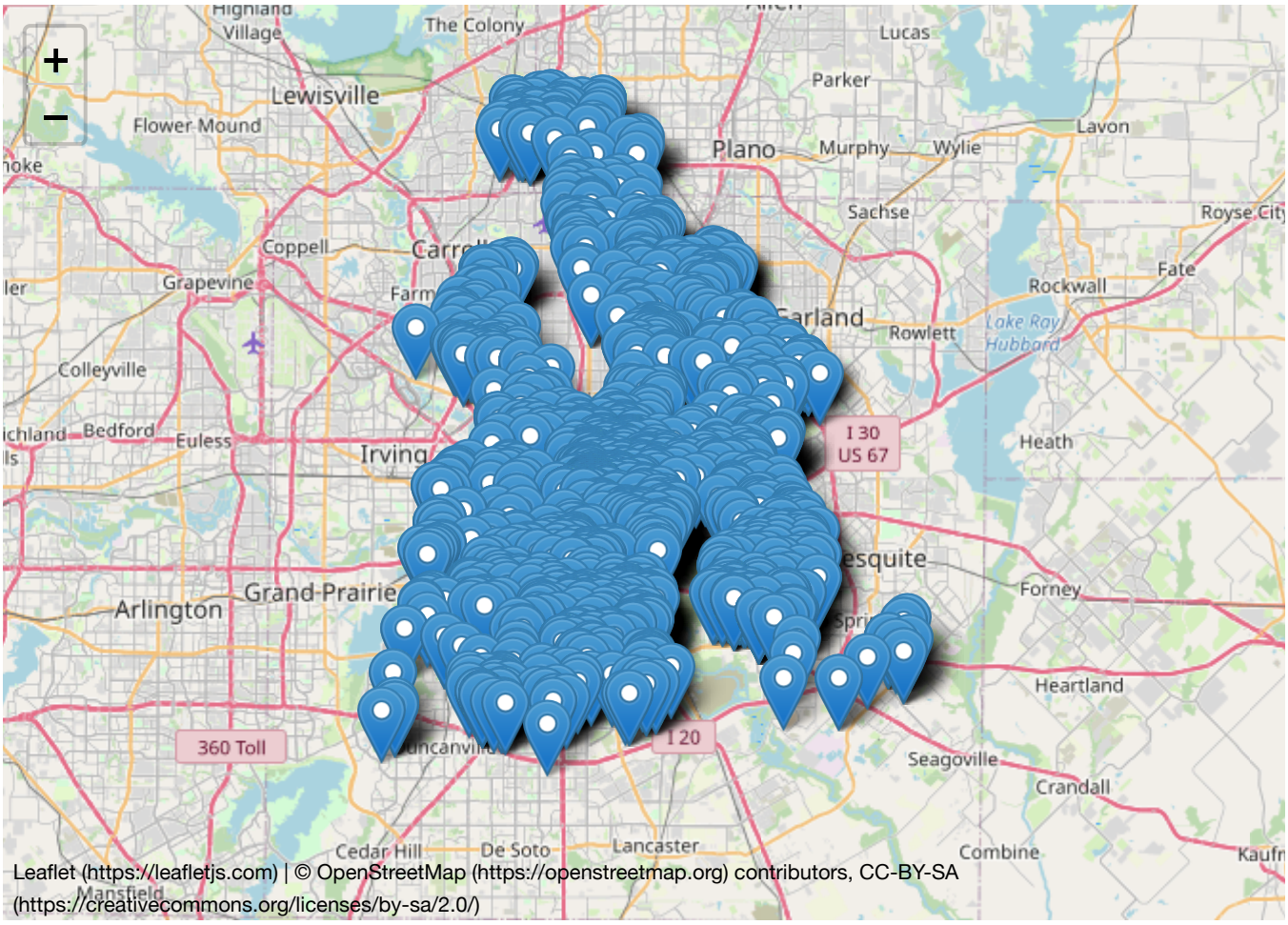
Crime Statistics In Dallas, Texas (Circa 2016 AD)

Crime Summary

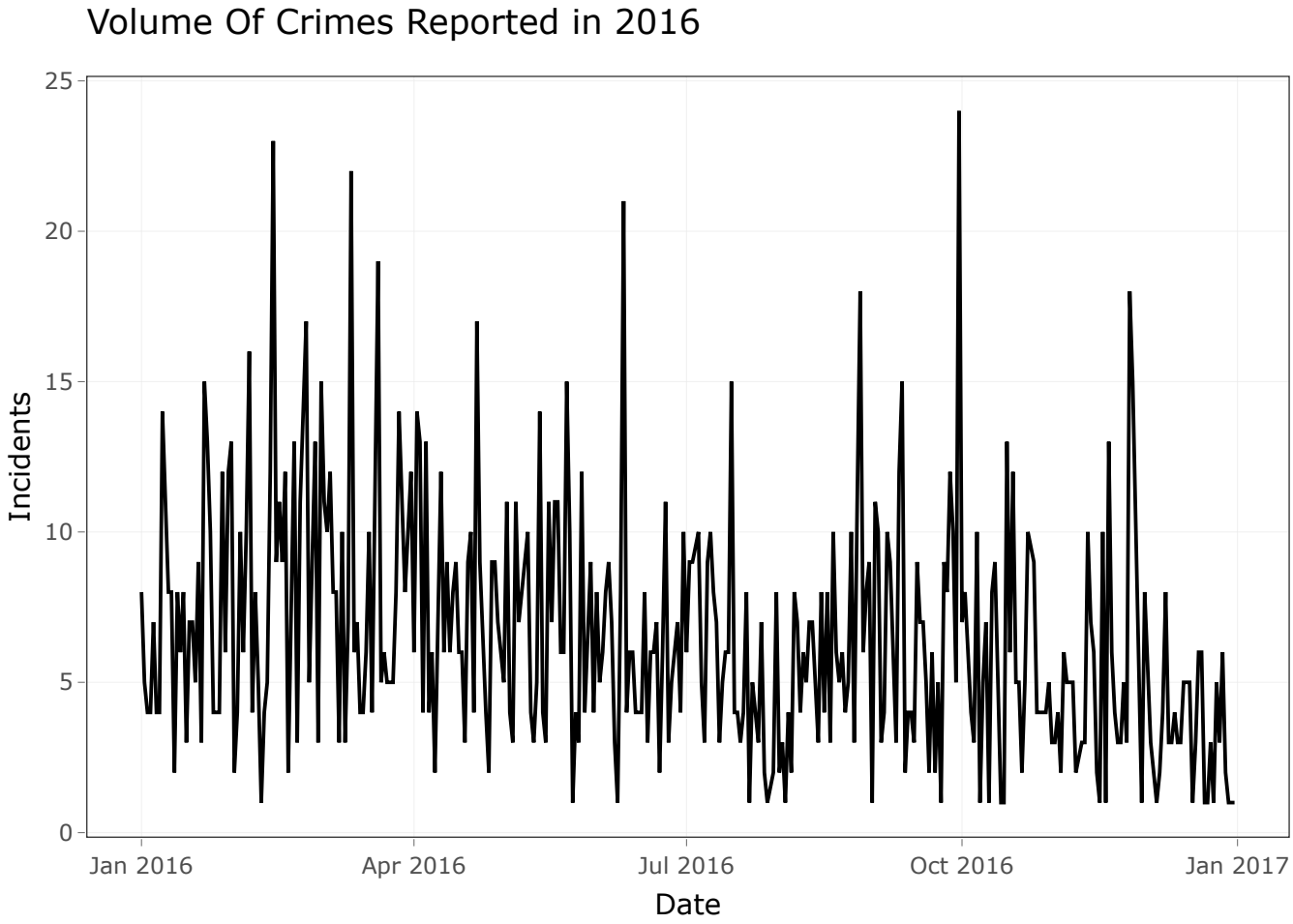
```
# load data
policing <- read.delim("/Users/praseedpai/ACADEMIC/DataVisualisation/Assignment/37-00
049_UOF-P_2016_prepped.csv", sep=",", header = TRUE)
# remove the header
policing = policing[-1,]
# filtering the data based on arrested
crimeData <- policing
# remove null values from logitude and latitude
crimeData <- crimeData[!(is.na(crimeData$LOCATION_LATITUDE) | crimeData$LOCATION_LATI
TUDE==""), ]
# rename the LOCATION_LATITUDE and LOCATION_LONGITUDE columns
names(crimeData)[names(crimeData) == "LOCATION_LATITUDE"] <- "lat"
names(crimeData)[names(crimeData) == "LOCATION_LONGITUDE"] <- "long"
# creating leaflet
crimeData %>%
  leaflet() %>%
  addTiles() %>%
  addMarkers(lng= ~as.numeric(long), lat= ~as.numeric(lat),popup = ~SUBJECT_OFFENSE)
```



Dallas in the state of Texas in the United States, has the highest crime rate when compared to the most other cities across the country. Above map shows the overview of crime incidents reported in Dallas in the year 2016.

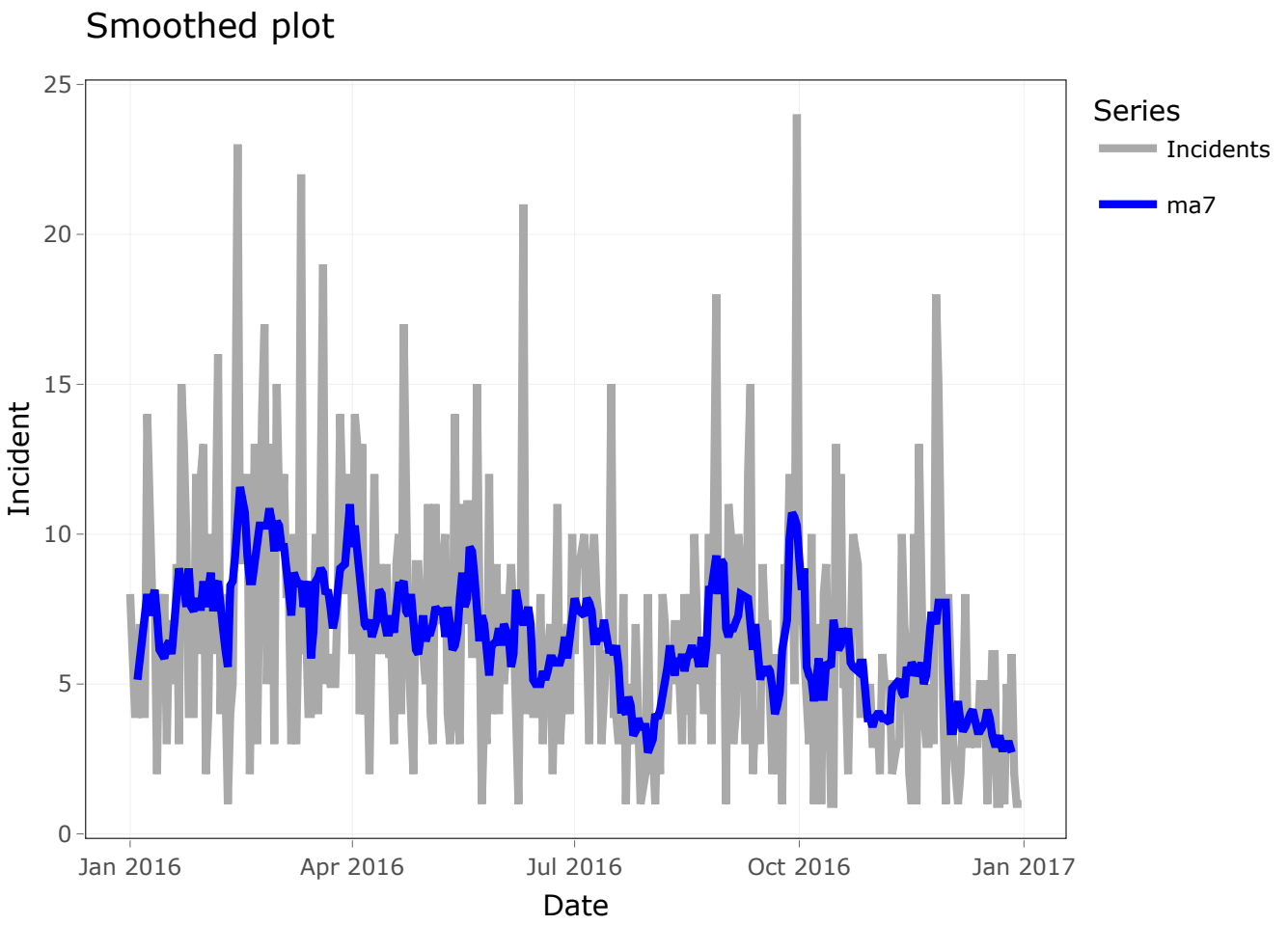
Annual Crime Trend

```
crimeDateData <- crimeData
tdates <- mdy(crimeDateData$INCIDENT_DATE)
incidentsPerday <- as.data.frame(table(tdates))
tscases <- xts(data.frame(Incidents=incidentsPerday$Freq),ymd(incidentsPerday$tdate
s))
p <- autoplot(tscases,facet=NULL)+theme_bw()+xlab("Date")+labs(title="Volume Of Crime
s Reported in 2016")
ggplotly(p)
```



The determining factors in the crime rate include time period, location and type of crime. The above figure shows the crime trend per day in Dallas in the year 2016. Number of incident reported is high during the month of October and lowest during the month of December.

```
## Weekly Moving Average smoothing
ma7 <- forecast::ma(incidentsPerday$Freq, 7)
## Makes a 2-variable time series
tscases <- xts(data.frame(Incidents=incidentsPerday$Freq,ma7=ma7),ymd(incidentsPerday
$tdates))
## and plots them
## plot(tscases,main = "Daily New COVID-19 Cases",col=c("darkgrey","red"))
p <- autoplot(tscases,facet=NULL)+
geom_line(linewidth=1.1) +
scale_color_manual( values = c("darkgrey","blue"))+theme_bw()+labs(x="Date", y="Incid
ent", title="Smoothed plot")
ggplotly(p)
```



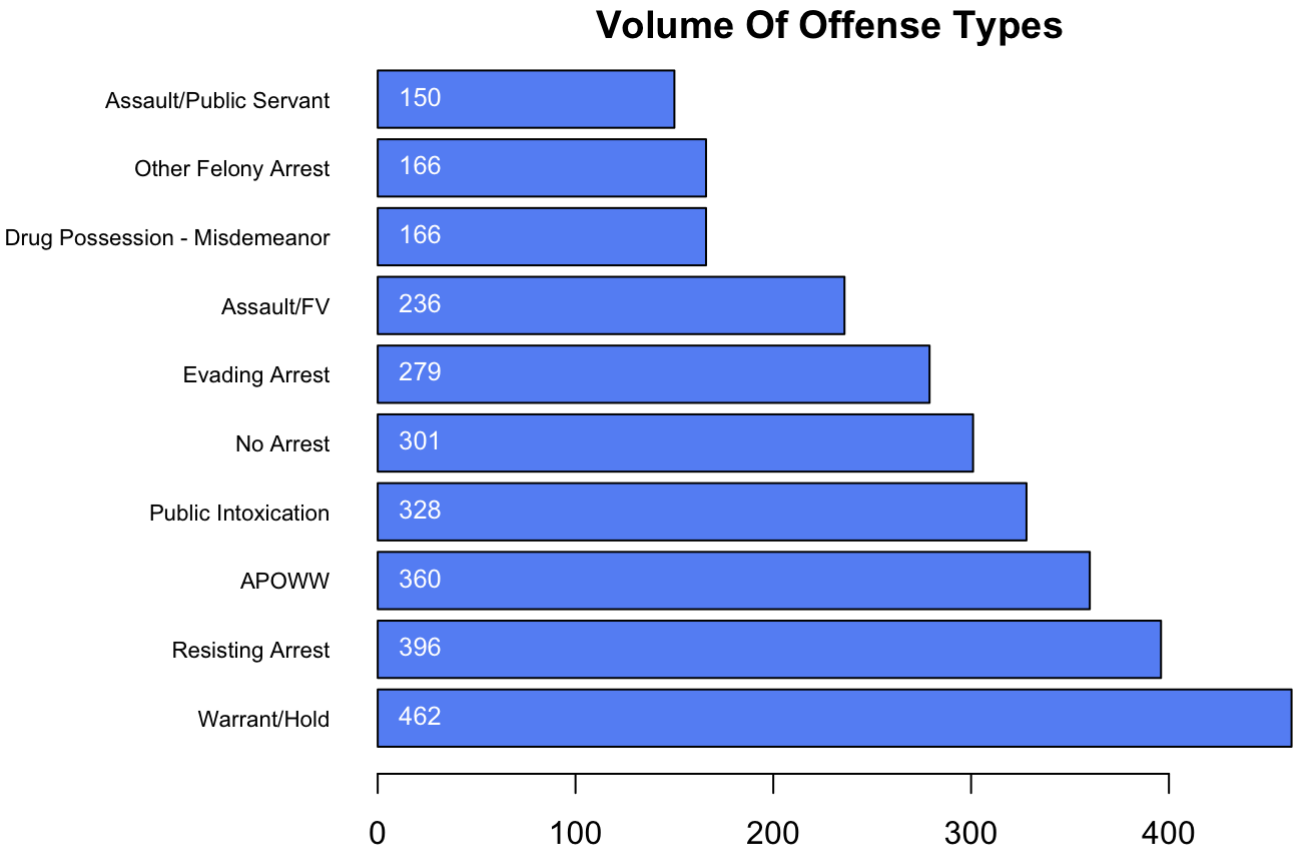
Weekly Moving Average smoothing (above figure) is used to remove the noise or random fluctuations in the data.

Criminal Offenses

```
offenseData <- crimeData
# separate the csv string to multiple columns
separatedDataDF <- separate(data = offenseData, col =SUBJECT_OFFENSE, into = c("var
1", "var2","var3","var4","var5","var6","var7"), sep = "\\,")
```

```
## Warning: Expected 7 pieces. Missing pieces filled with `NA` in 2327 rows [1, 2, 3,
4, 5,
## 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

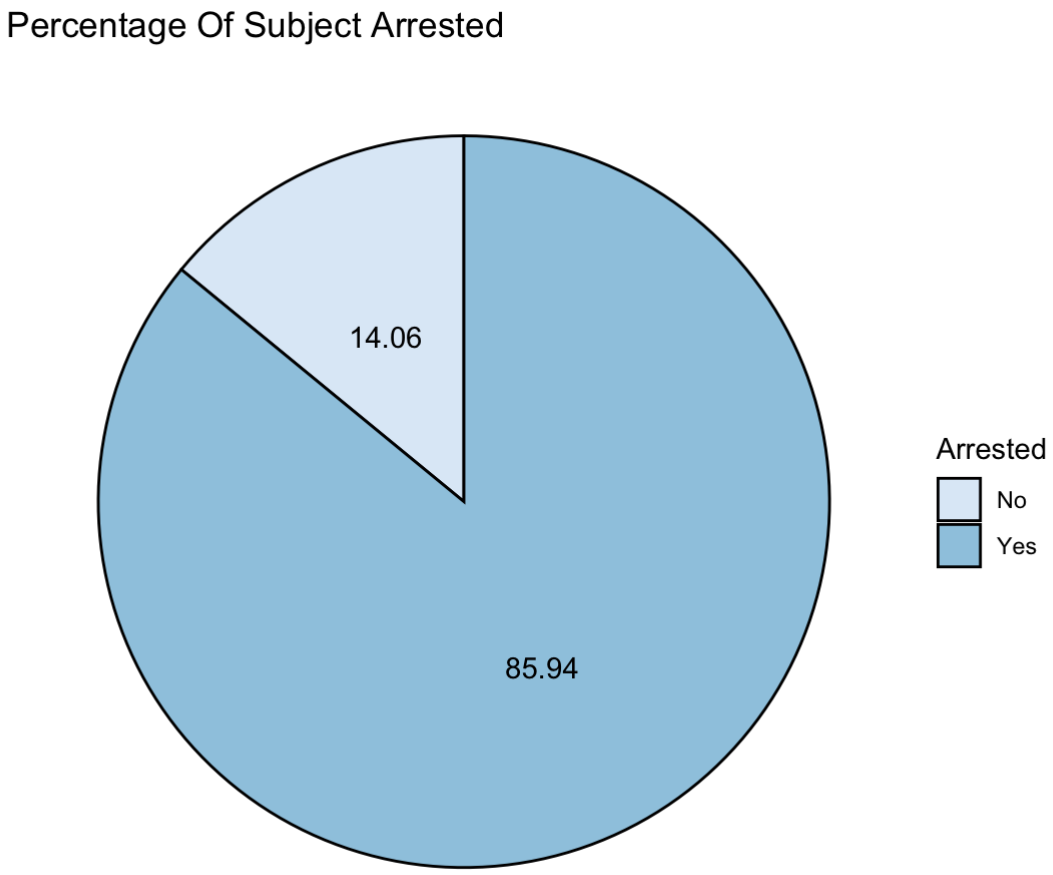
```
# merging column values
separatedCrimeOffenseData <- data.frame(c(separatedDataDF$var1,separatedDataDF$var2,
separatedDataDF$var3,separatedDataDF$var4,separatedDataDF$var5,separatedDataDF$var6,s
eparatedDataDF$var7), separatedDataDF$lat, separatedDataDF$long, separatedDataDF$DIVI
SION, separatedDataDF$INCIDENT_DATE)
# rename the column
names(separatedCrimeOffenseData)[1] <- "data"
names(separatedCrimeOffenseData)[2] <- "lat"
names(separatedCrimeOffenseData)[3] <- "long"
names(separatedCrimeOffenseData)[4] <- "DIVISION"
names(separatedCrimeOffenseData)[5] <- "INCIDENT_DATE"

counts = table(str_trim(separatedCrimeOffenseData$data))
counts = counts[order(counts, decreasing=T)]

par(mar = c(5,10,1,1))
y = barplot(counts[1:10], horiz=T, las=1, cex.names=0.7, col="#6693F5",main="Volume O
f Offense Types")
text(1, y, counts[1:10], pos=4, cex=0.8, col="white")
```



There are different types of criminal offenses reported and most common ones are listed in the above plot. Warrant / Hold rate is the highest about 462 followed by Resisting Arrest 396. A good data set should have mutuallay exclusive attributes and it should be collectively exhaustive. But these offenses are mutually inclusive (some criminal offenses may fall into multiple categories).

Crime Incidents Arrested

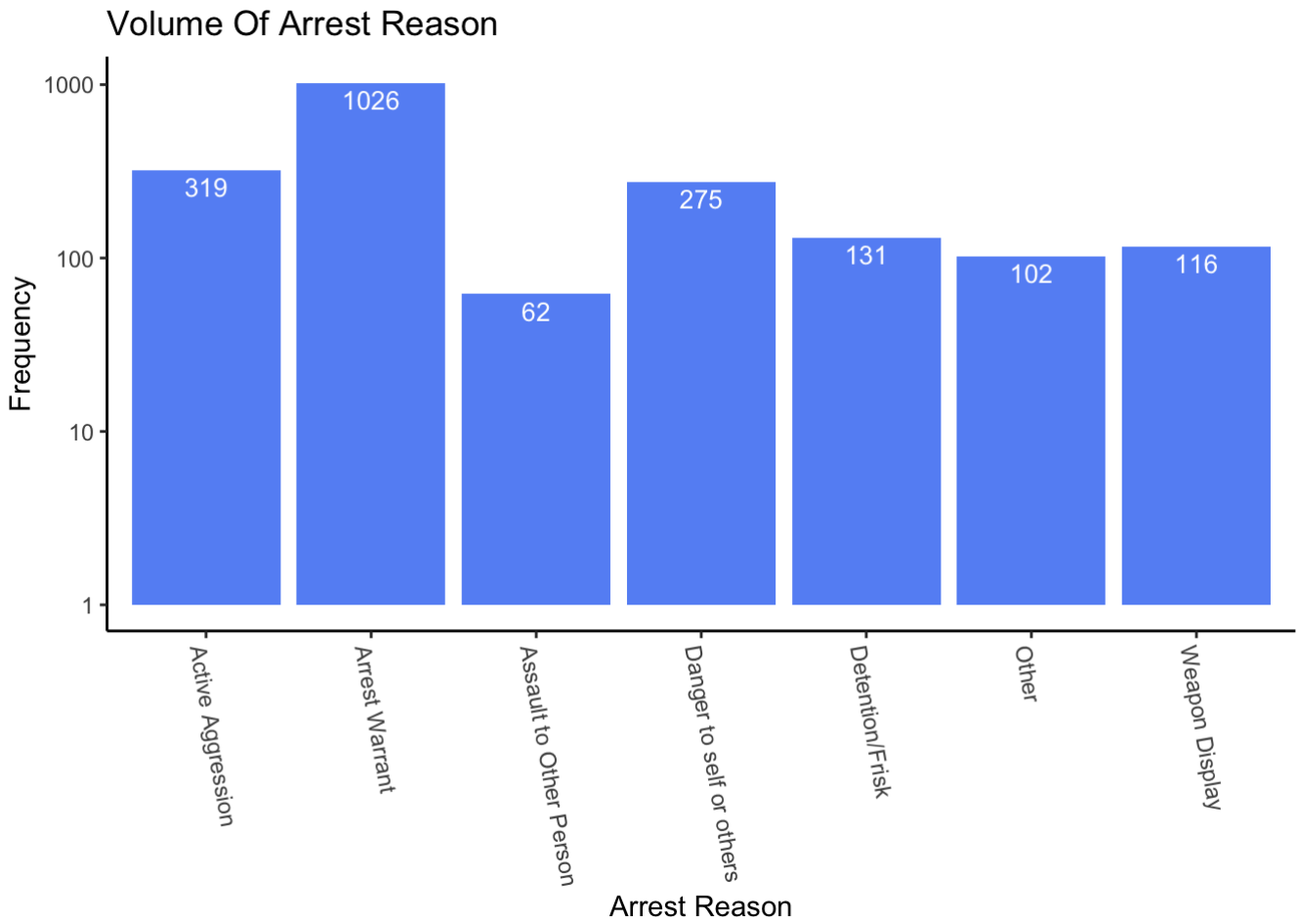
```
crimeStatusDF <- as.data.frame(round( 100*prop.table(table(policing$SUBJECT_WAS_ARRES
TED)),2))
names(crimeStatusDF)[1]<-"Arrested"
ggplot(crimeStatusDF, aes(x = "", y = Freq, fill = Arrested)) +
  geom_col(color = "black") +
  geom_text(aes(label = Freq),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_brewer() +
  theme_void()+labs(title="Percentage Of Subject Arrested")
```



The number of crimes that are reported to the police or other law enforcement agencies are refered as crime reported. Crime arrested, on the other hand, refers to the number of individuals who have been arrested for committing a crime. The above Pie Chart shows the percentage of crime incident reported and arrested. About 85.94% crime subjects are arrested while 14.06% are not arrested.

Arrest Reason

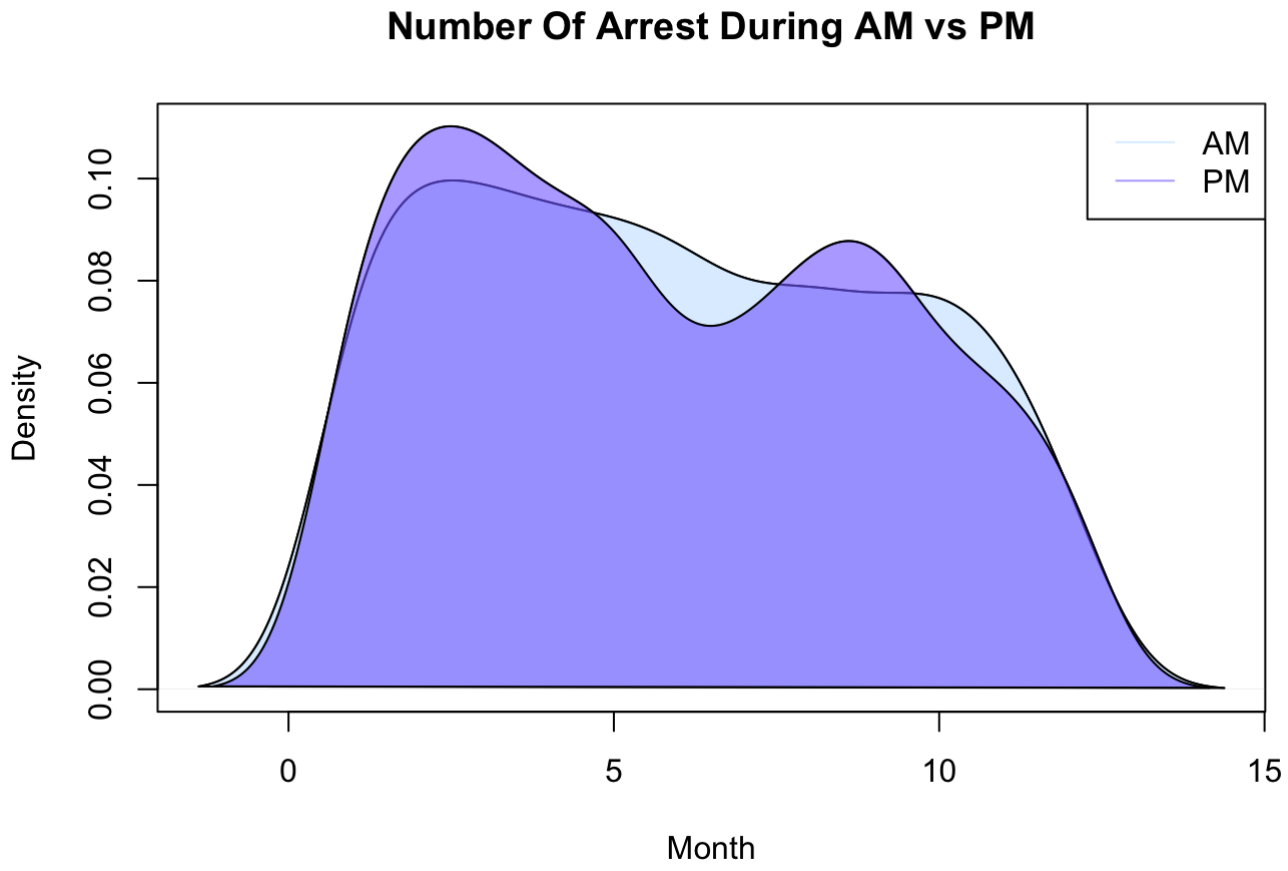
```
crimeData <- policing%>%filter(policing$SUBJECT_WAS_ARRESTED=="Yes")
crimeData$REASON_FOR_FORCE[crimeData$REASON_FOR_FORCE=="Arrest"] <- "Arrest Warrant"
reason <- crimeData$REASON_FOR_FORCE
reasonDF <- as.data.frame(table(reason))
reasonDF <- reasonDF%>%filter(Freq > 7)
reasonDF <- arrange(reasonDF, Freq)
ggplot(data=reasonDF, aes(x=reason, y=Freq)) +
  geom_bar(stat="identity", fill="#6693F5")+
  geom_text(aes(label=Freq), vjust=1.4, color="white", size=3.5)+
  theme_classic()+ theme(axis.text.x = element_text(angle = -80, vjust = 1, hjust =
0))+labs(x = "Arrest Reason", y="Frequency", title="Volume Of Arrest Reason")+  scale
_y_continuous(trans='log10')
```



A crime arrest can occur for a variety of reasons. Some of them are shown in the above figure. Arrest Warrant count is the highest followed by Active Aggression. Just because somebody is arrested, does not mean he is proven gulity. It just shows that matter has to be further investigated by the law enforcement agencies and further by the Judiciary.

Crime Pattern

```
seggragated_time <- function(time)  {
  return (str_extract(time, "(AM|PM)"))
}

peroid <- function(time, meridiem) {
  ifelse (time == meridiem, 1,0)
}

arrestedMonth <- month(mdy(crimeData$INCIDENT_DATE))
arrestedTime <- seggragated_time(crimeData$INCIDENT_TIME)
AMperArretedRecords <- peroid(arrestedTime, 'AM')
PMperArrestedRecords <- peroid(arrestedTime, 'PM')
arrestRelatedData <- data.frame(arrestedMonth, AMperArretedRecords, PMperArrestedReco
rds)
clearedArrestedData <- na.omit(arrestRelatedData)

denx <- density(clearedArrestedData$arrestedMonth[clearedArrestedData$PMperArrestedRe
cords ==0])
deny <- density(clearedArrestedData$arrestedMonth[clearedArrestedData$PMperArrestedRe
cords ==1])
# Plot
plot(denx,
     ylim = c(0, max(c(denx$y, deny$y))),
     xlim = c(min(c(denx$x, deny$x)),
              max(c(denx$x, deny$x))), main = "Number Of Arrest During AM vs PM",xlab
= "Month")
lines(deny)
polygon(denx, col = rgb(0.78, 0.89, 1, alpha = 0.6))
polygon(deny, col = rgb(0.51, 0.44, 1, alpha = 0.6))
legend("topright",c("AM","PM"),lty=1,col=c(rgb(0.78, 0.89, 1, alpha = 0.6),rgb(0.51,
0.44, 1, alpha = 0.6)))
```



Crime incidents are reported more commonly in evening than morning. The above figure shows that apart from month June, crime incidents are reported more during the day. This may be may be due to many reasons like timing of activity, witnesses, safety etc.

Division Details

```
divisionData <- crimeData
divTable <- table(divisionData$DIVISION)
kable(formattable(divTable), col.names= c("Division", "Frequency"), align="l", captio
n = "Division Wide Crime Count")
```

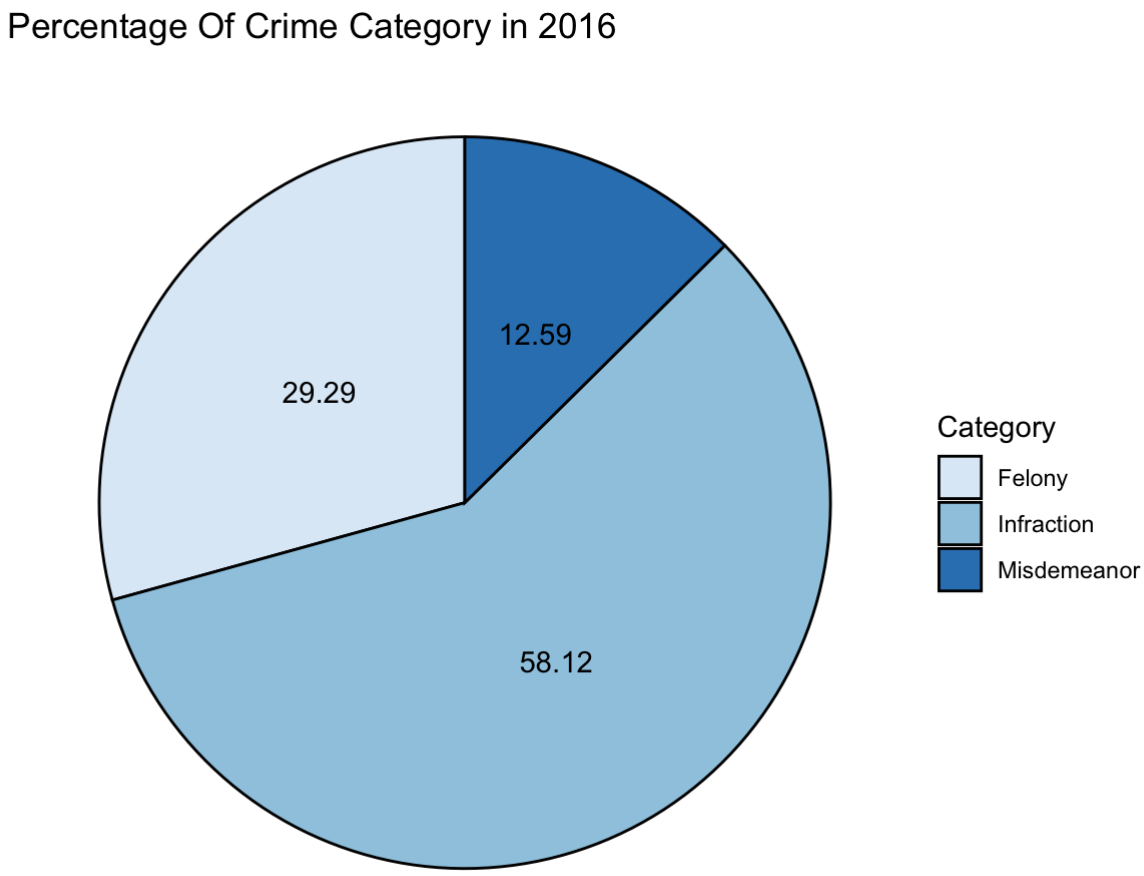
Division Wide Crime Count

Division	Frequency
CENTRAL	499
NORTH CENTRAL	260
NORTHEAST	289
NORTHWEST	171
SOUTH CENTRAL	267
SOUTHEAST	314
SOUTHWEST	248

The above table shows Divisions and crimes frequency (number of crimes) reported. Crime rate is higher in Central division and lowest in North West.

Crime Category

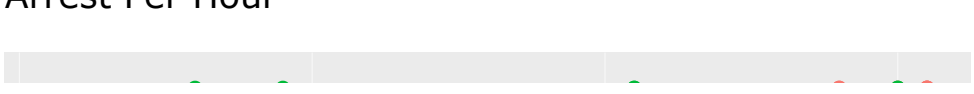
```
# Categorizing the subject offense
crimeCategory <- policing
crimeCategory <- crimeCategory %>%
  mutate(CATEGORY = ifelse(
    grepl('Assault', crimeCategory$SUBJECT_OFFENSE, ignore.case = TRUE),  "Felony",
    ifelse (grepl('Robbery', crimeCategory$SUBJECT_OFFENSE, ignore.case = TRUE) , "Fe
lony",ifelse (grepl('Criminal Mischief', crimeCategory$SUBJECT_OFFENSE, ignore.case =
TRUE) , "Felony",ifelse (grepl('Prostitution', crimeCategory$SUBJECT_OFFENSE, ignore.
case = TRUE) , "Felony",
    ifelse (grepl('Theft', crimeCategory$SUBJECT_OFFENSE, ignore.case = TRUE) , "Misd
emeanor",ifelse (grepl('Burglary', crimeCategory$SUBJECT_OFFENSE, ignore.case = TRUE)
, "Misdemeanor",ifelse (grepl('Felony', crimeCategory$SUBJECT_OFFENSE, ignore.case =
TRUE) , "Felony",ifelse (grepl('Misdemeanor', crimeCategory$SUBJECT_OFFENSE, ignore.c
ase = TRUE) , "Misdemeanor",ifelse (grepl('Crim Trespass', crimeCategory$SUBJECT_OFFE
NSE, ignore.case = TRUE) , "Infraction",ifelse (grepl('Violation', crimeCategory$SUBJ
ECT_OFFENSE, ignore.case = TRUE) , "Infraction", ifelse (grepl('APOWW', crimeCategory
$SUBJECT_OFFENSE, ignore.case = TRUE) , "Infraction",
    'Infraction')))
    )))))))))

categoryDF <- as.data.frame(round( 100*prop.table(table(crimeCategory$CATEGORY)),2))
names(categoryDF)[1]<-"Category"
ggplot(categoryDF, aes(x = "", y = Freq, fill = Category)) +
  geom_col(color = "black") +
  geom_text(aes(label = Freq),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_brewer() +
  theme_void()+labs(title="Percentage Of Crime Category in 2016")
```

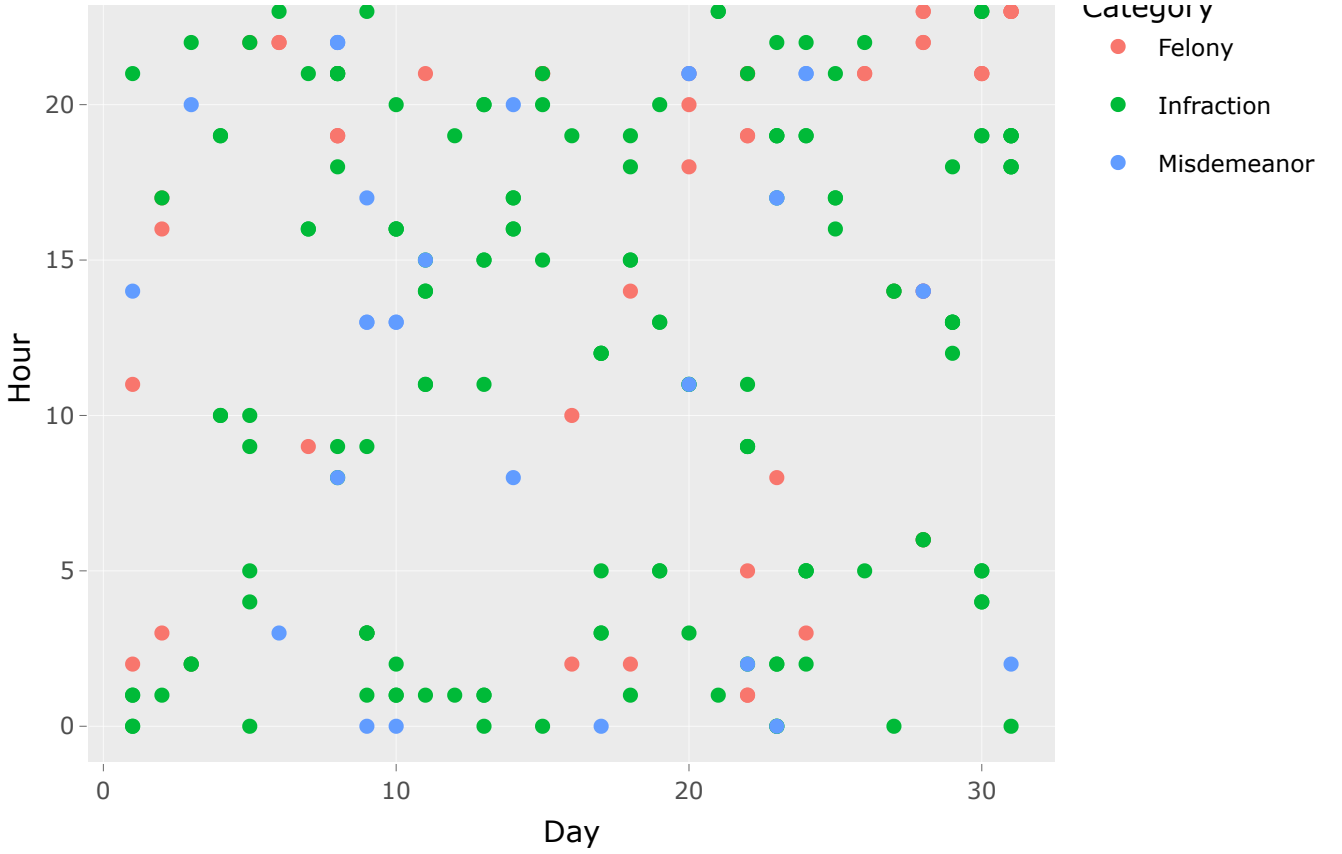


Crimes are broadly divided into felonies, misdemeanors and infractions. Felonies are criminal offenses which are serious in nature and are typically punishable by imprisonment for more than one year or by death in certain states. Examples of felonies include Assualt, Robbery, Criminal Mischief etc. Misdemeanors are criminal offenses are not as intense as felonies and are typically punishable by a fine or imprisonment for up to one year. Examples of misdemeanors include Burglary, theft, and some minor drug offenses. An infraction is a minor offense. It is a violation of a rule, such as a traffic violation or a minor code violation. As per the above Pie Chart,Felony form 29.29%, infraction forms 58.12% and misdemeanors form the 58.12% of crimes.

Crime Analysis

```
divisionData <- crimeCategory
#table(divisionData$DIVISION)
divisionData <- crimeCategory
divisionData <- divisionData %>% filter(month(mdy(divisionData$INCIDENT_DATE)) == 1)
divisionData$datetime <- paste(divisionData$INCIDENT_DATE
                  , divisionData$INCIDENT_TIME)
divisionData <- divisionData %>%
  mutate(day =day(mdy(INCIDENT_DATE)), month =month(mdy(INCIDENT_DATE)) , hour=hour(m
dy_hms(datetime)))
Category <- factor(divisionData$CATEGORY)
p <- ggplot(divisionData, aes(day, hour, color=Category)) + geom_point() +
  labs(x="Day", y="Hour", legend="Crime CATEGORY", title="Arrest Per Hour")
ggplotly(p)
```

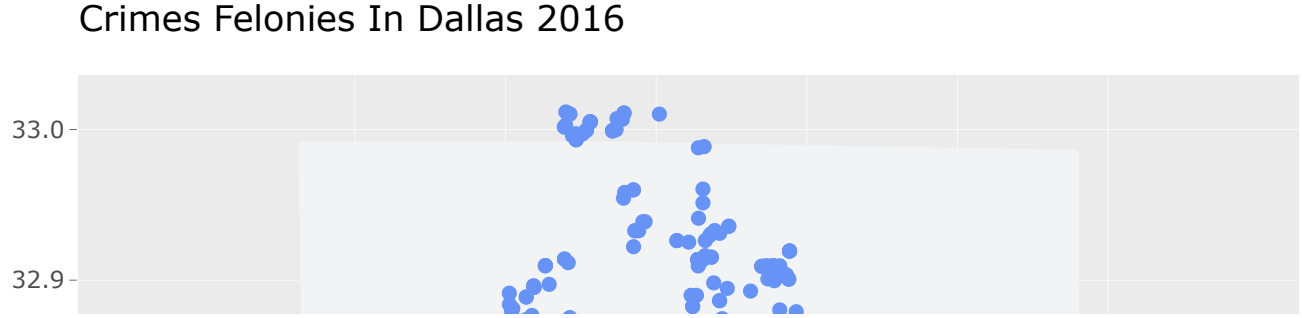


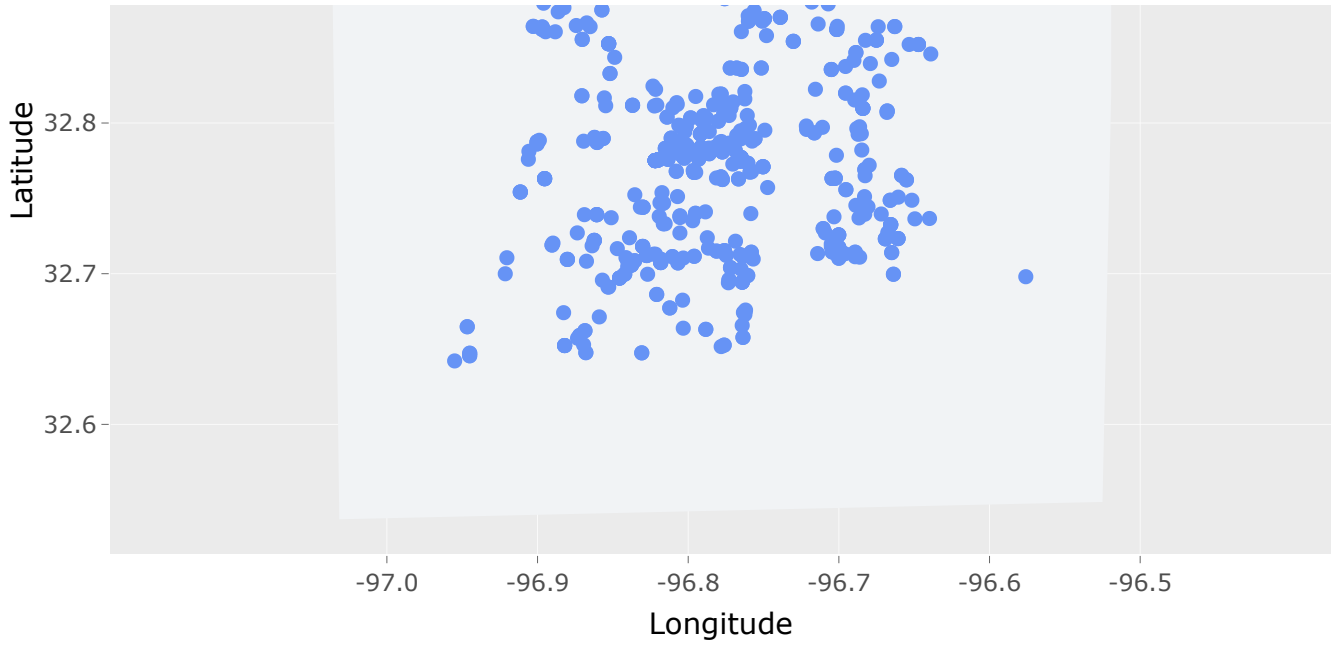


The above figure shows the crime category reported per hour in a day during the month of January (choice of month was arbitrary). Infractions are reported more and felonies the less.

Felonies

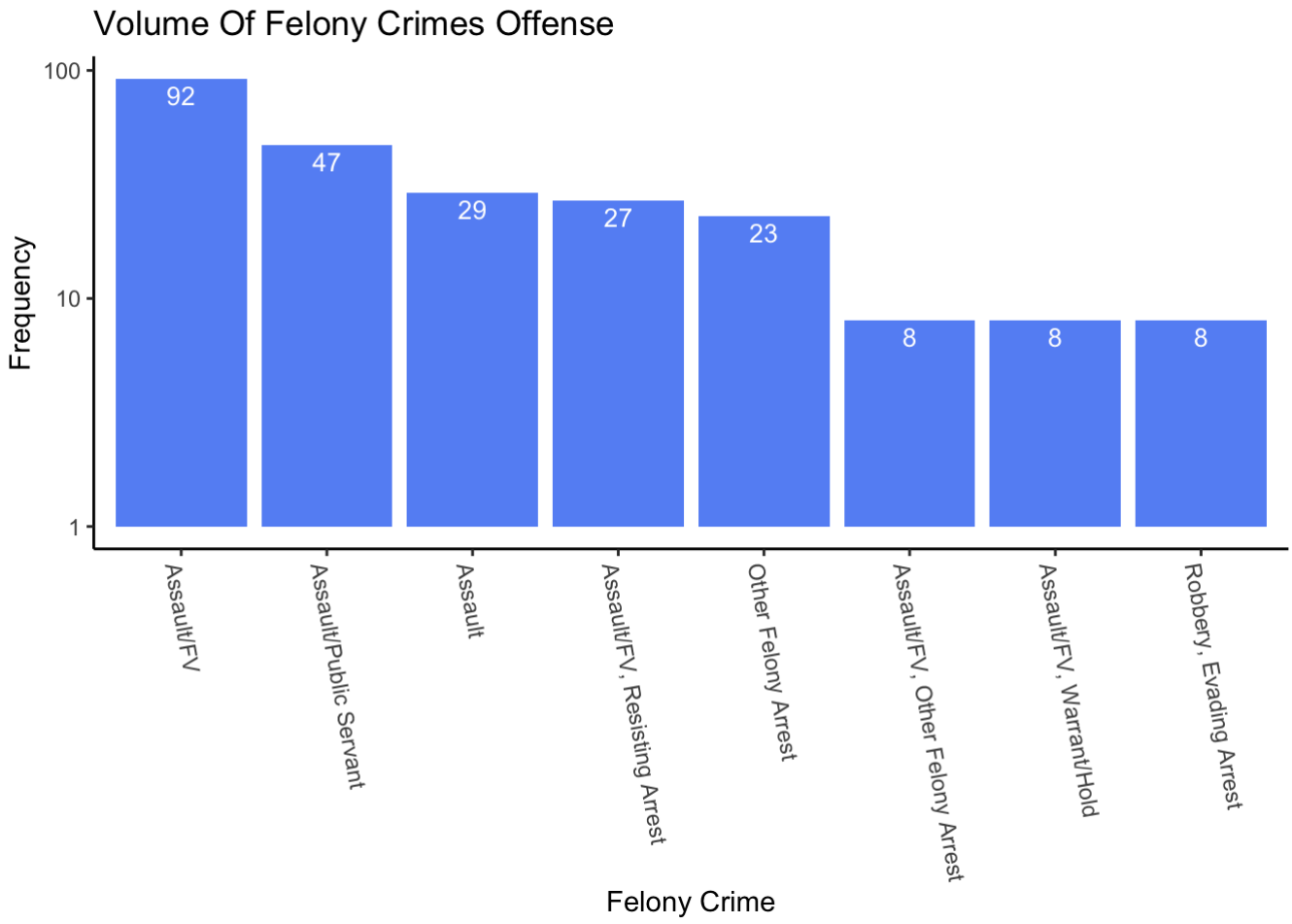
```
maplist<-map("county", namesonly = TRUE, plot=FALSE)
m = map_data('county', region = 'texas,dallas')
felonies <- crimeCategory%>%filter(CATEGORY == "Felony")
felonies <- na.omit(felonies)
felonies <- felonies%>%mutate(long = as.numeric(LOCATION_LONGITUDE), lat=as.numeric(L
OCATION_LATITUDE))
p <- ggplot() +
 geom_polygon( data=m, aes(x=as.numeric(long), y=as.numeric(lat),group=group),colour
="#f1f3f5", fill="#f1f3f5" )+
 geom_point(data=felonies,aes(x=long,y=lat),colour="#6693F5")+
 geom_text(data=felonies, aes(x=long, y=lat, label=""), size=2, hjust=-0.1) +
 ggtitle("Crimes Felonies In Dallas 2016")+
 xlab('Longitude')+
 ylab('Latitude')+
 coord_fixed()
ggplotly(p)
```





The above figure shows the felonies reported in year 2016 in Dallas Texas. The longitude and latitude of the the place is being used to generate the plot.

```
crimeRateData <- crimeCategory
felonies <- crimeCategory%>%filter(CATEGORY == "Felony")
feloniesTB <- table(felonies$SUBJECT_OFFENSE)
feloniesCount = as.data.frame(feloniesTB[order(feloniesTB, decreasing=T)])
feloniesCount <- feloniesCount[1:8, ]
ggplot(data=feloniesCount, aes(x=Var1, y=Freq)) +
  geom_bar(stat="identity", fill="#6693F5")+
  geom_text(aes(label=Freq), vjust=1.4, color="white", size=3.5)+
  theme_classic()+ theme(axis.text.x = element_text(angle = -80, vjust = 1, hjust =
0))+labs(x = "Felony Crime", y="Frequency", title="Volume Of Felony Crimes Offense")+
scale_y_continuous(trans='log10')
```



The above figure shows volume of the felonies crime offense. Assault / FV has the highest count of 92 followed by Assault Public Servant the count of 47.

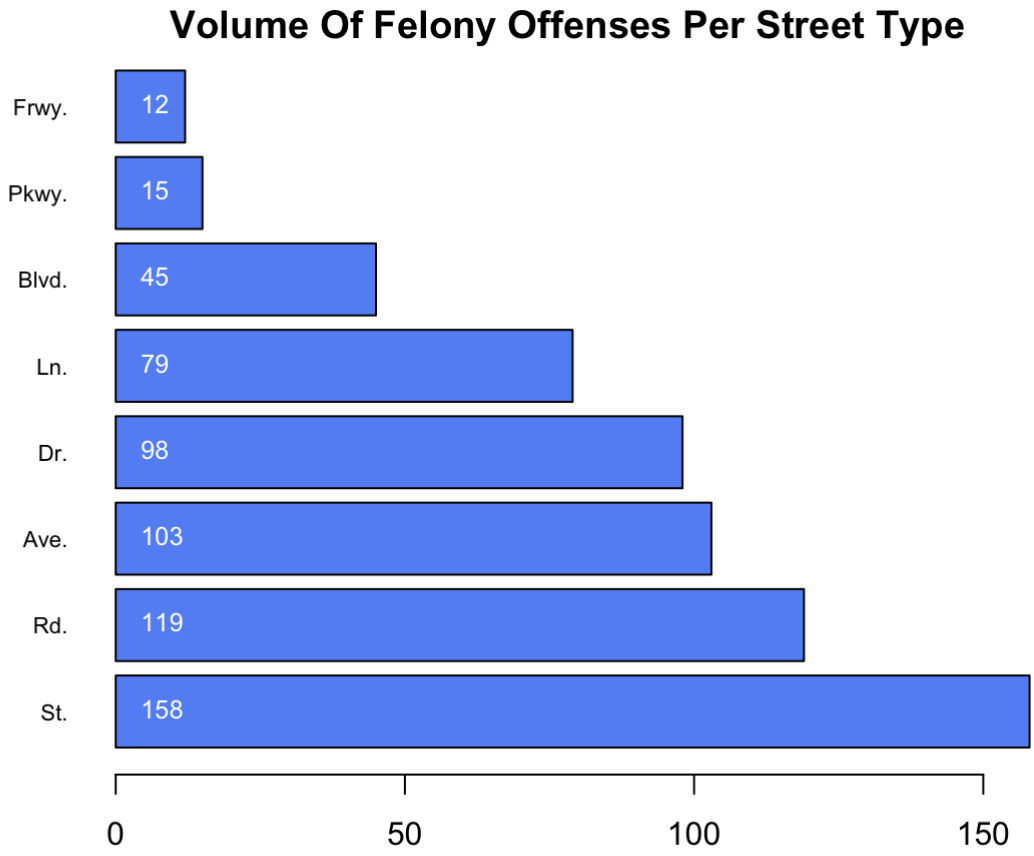
```
crimeRateData <- crimeCategory
crimeRateData$SUBJECT_RACE[crimeRateData$SUBJECT_RACE=="NULL"] <- "Other"
crimeRateData$SUBJECT_GENDER[crimeRateData$SUBJECT_GENDER=="NULL"] <- "Unknown"
felonies <- crimeRateData%>%filter(CATEGORY == "Felony")
feloniesTable <- table(felonies$SUBJECT_RACE, felonies$SUBJECT_GENDER)
kable(formattable(feloniesTable), align="l", caption = "Table shows the count of Gend
er and Race")
```

Table shows the count of Gender and Race

	Female	Male	Unknown
Asian	0	2	0
Black	92	309	0
Hispanic	18	152	0
Other	0	5	6
White	11	103	0

As per the above table, from a gender point of view men commit more felonies crimes than women and from a racial perspective, the people of Black origin commit more felonies than others followed by the Whites.

```
crimeRateData <- crimeCategory
crimeRateData <- crimeRateData[!(is.na(crimeRateData$STREET_TYPE) | crimeRateData$STR
EET_TYPE=="NULL"), ]
crimeRateData <- crimeRateData%>%mutate(street = str_remove(crimeRateData$STREET_TYP
E, "[.]"))
felonies <- crimeRateData%>%filter(CATEGORY == "Felony")
location <- felonies$STREET_TYPE
tb <- table(location)
counts = tb[order(tb, decreasing=T)]
par(mar = c(5,10,1,1))
y = barplot(counts[1:8], horiz=T, las=1, cex.names=0.7, col="#6693F5",main="Volume Of
Felony Offenses Per Street Type" )
text(1, y, counts[1:8], pos=4, cex=0.8, col="white")
```

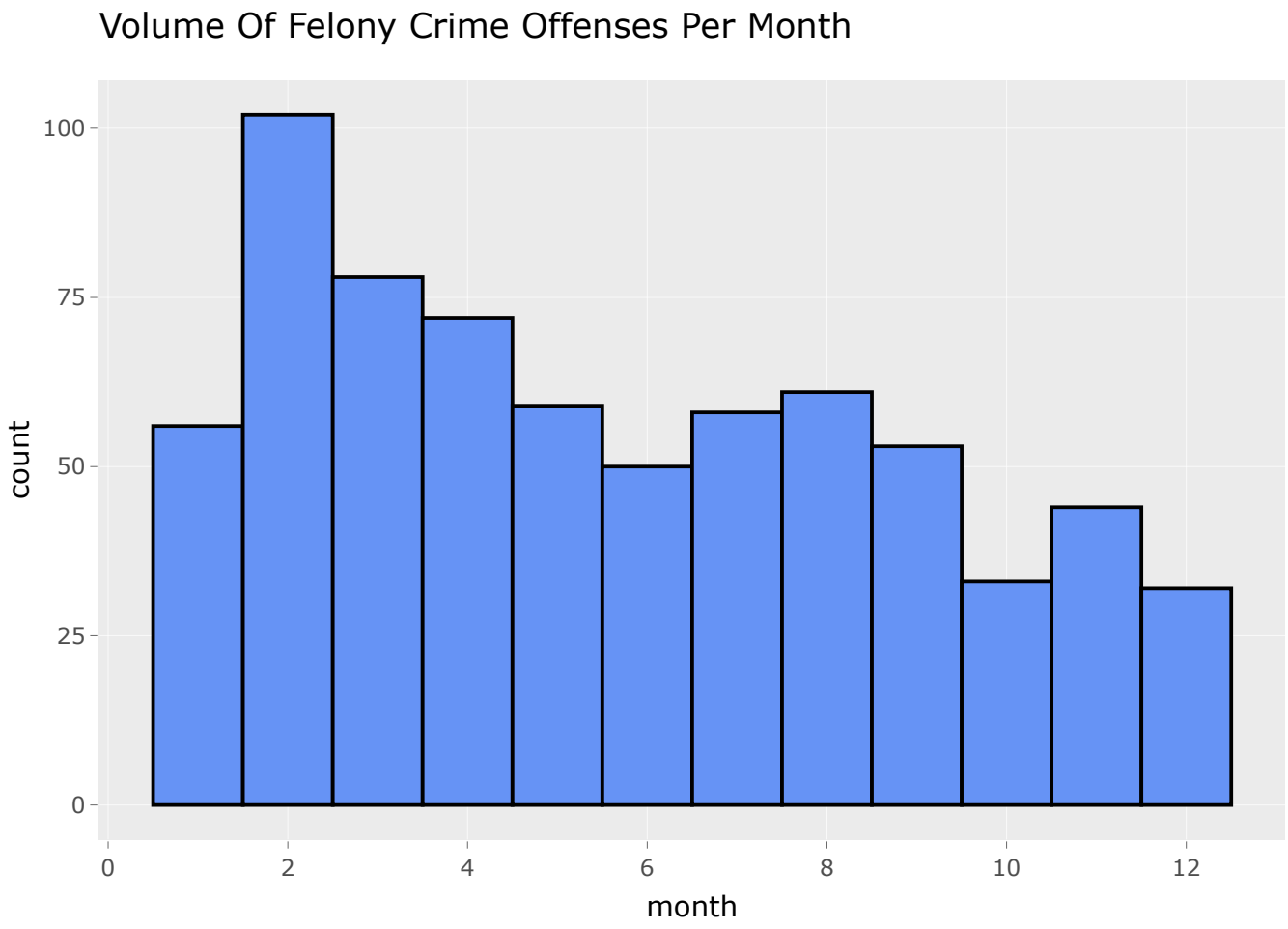


The location of a crime offense can also vary and can have a significant impact on the type of crime that is committed. Common street types for crime offenses include Freeway, Parkway, Boulevard, Lane, Drive, Avenue, Road, Street. Highest count of felonies crime are reported in Street (158) and lowest in Freeway (12).

```
crimeMonthData <- crimeCategory
crimeMonthData$datetime <- paste(crimeMonthData$INCIDENT_DATE
                   , crimeMonthData$INCIDENT_TIME)
crimeMonthData <- crimeMonthData%>%drop_na(datetime)
crimeMonthData$SUBJECT_DESCRIPTION[crimeMonthData$SUBJECT_DESCRIPTION=="NULL"] <- "Un
known"
crimeMonthData <- crimeMonthData %>%
  mutate(day =day(mdy(INCIDENT_DATE)), month =month(mdy(INCIDENT_DATE)) ,year =year(m
dy(INCIDENT_DATE)) , hour=hour(mdy_hms(datetime)))
```

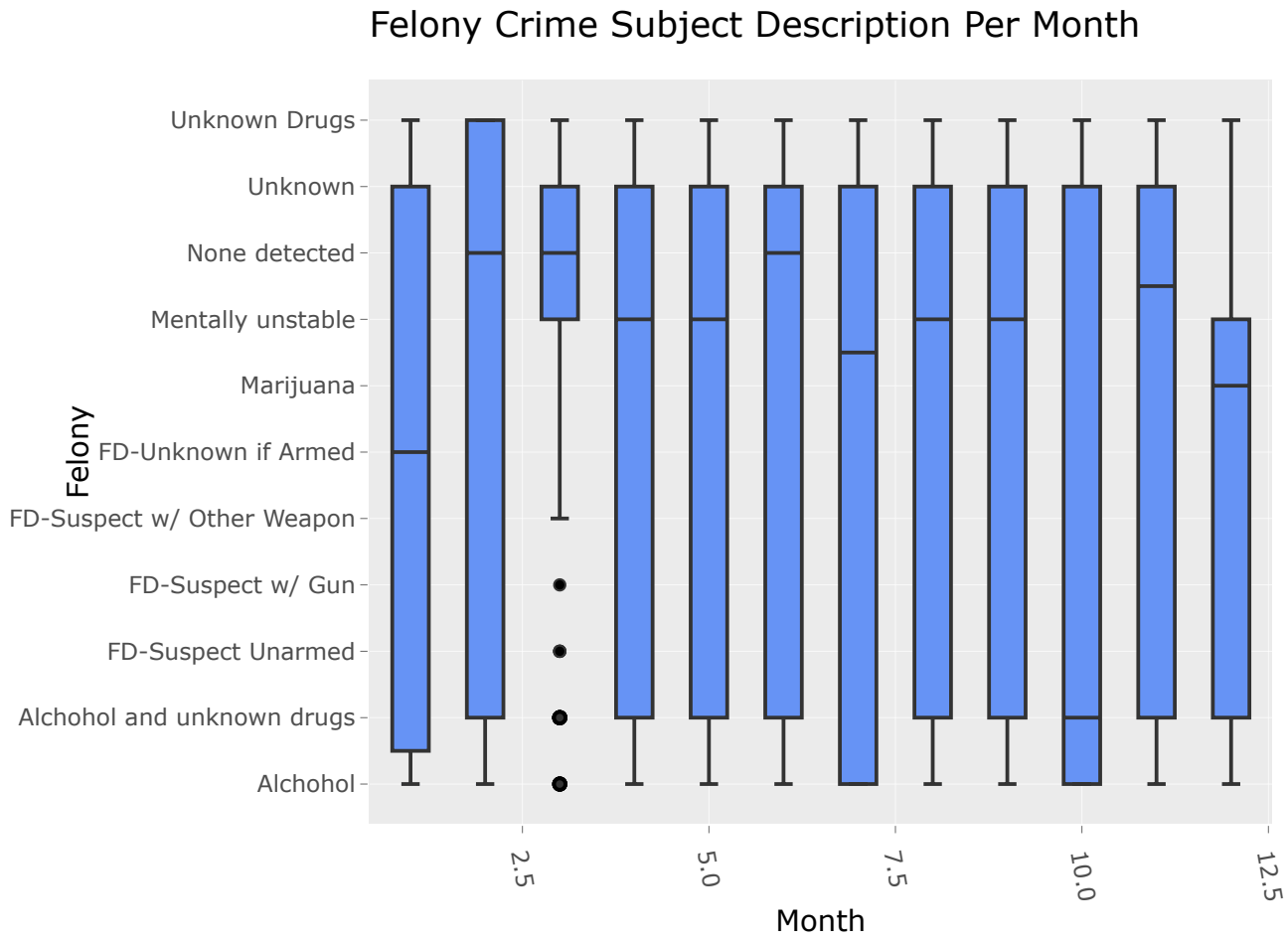
```
## Warning: There was 1 warning in `mutate()`.
## ℹ In argument: `hour = hour(mdy_hms(datetime))`.
## Caused by warning:
## !  10 failed to parse.
```

```
felonies <- crimeMonthData%>%filter(CATEGORY == "Felony")
G = felonies %>% ggplot(aes(month)) + geom_histogram(bins=12, fill="#6693F5", color =
"black")+ scale_x_continuous(breaks =c(0, 2, 4,6,8,10,12))+labs(title="Volume Of Felo
ny Crime Offenses Per Month")
ggplotly(G)
```



The above figure shows the count of Felonies crimes per month. Highest reported in the month of February and lowest in the month of December.

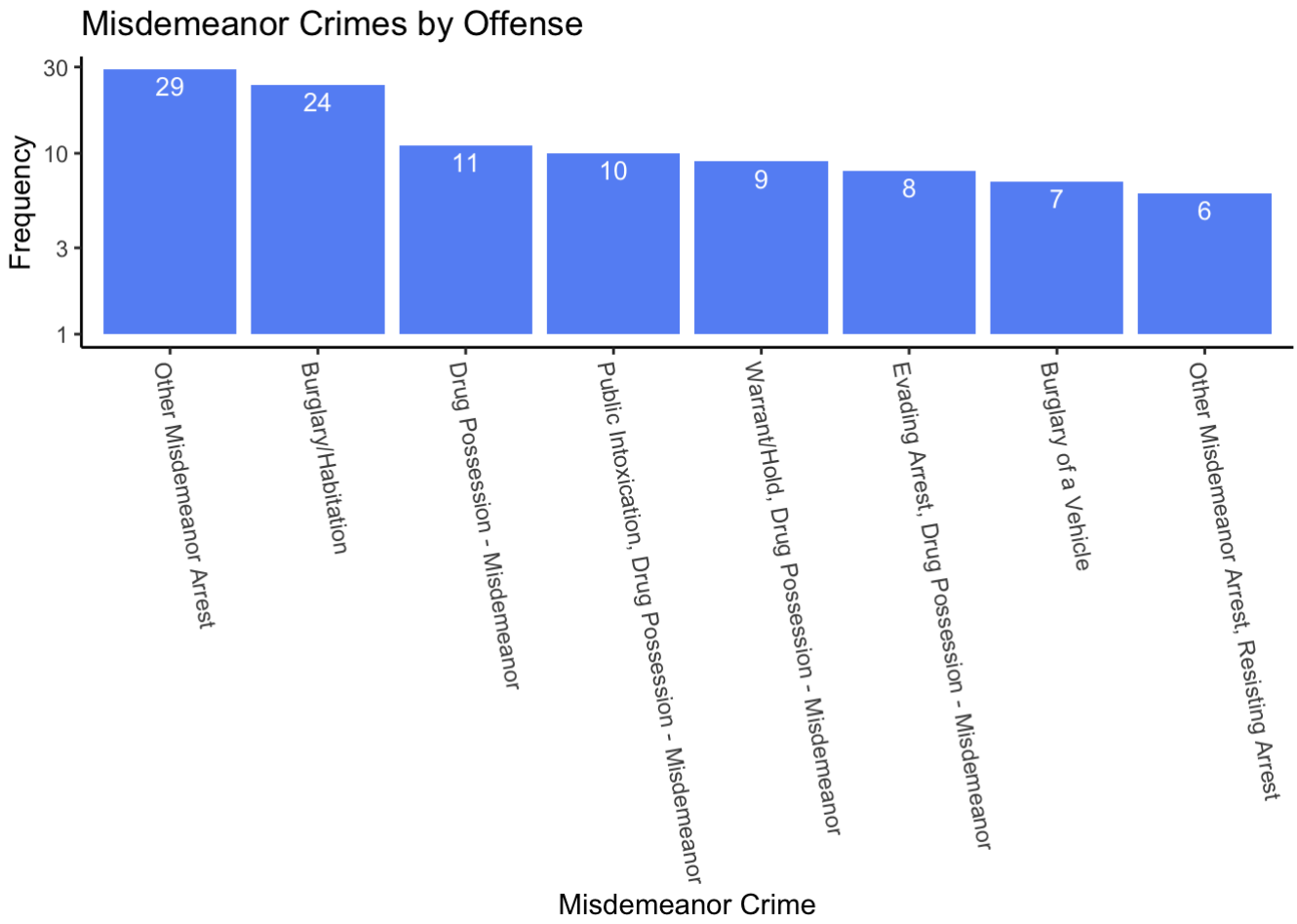
```
felonies <- crimeMonthData%>%filter(CATEGORY == "Felony")
p <- ggplot(felonies, aes(x = month, y = factor(SUBJECT_DESCRIPTION))) +
geom_boxplot(fill="#6693F5")+ theme( axis.text.x = element_text(angle = -80, vjust =
1, hjust = 0))+ labs(x = "Month", y="Felony")+labs(title="Felony Crime Subject Descri
ption Per Month")
ggplotly(p)
```



The above figure shows the subject description at the time of arrest for each month. The maximum count is shown as Unknown Alcohol and minimum Alcohol. Outliers are found in the month of March.

Misdemeanor Crime

```
crimeRateData <- crimeCategory
misdemeanor <- crimeCategory%>%filter(CATEGORY == "Misdemeanor")
misdemeanorTB <- table(misdemeanor$SUBJECT_OFFENSE)
misdemeanorCount = as.data.frame(misdemeanorTB[order(misdemeanorTB, decreasing=T)])
misdemeanorCount <- misdemeanorCount[1:8, ]
ggplot(data=misdemeanorCount, aes(x=Var1, y=Freq)) +
  geom_bar(stat="identity", fill="#6693F5")+
  geom_text(aes(label=Freq), vjust=1.4, color="white", size=3.5)+
  theme_classic()+ theme(axis.text.x = element_text(angle = -80, vjust = 1, hjust =
0))+labs(x = "Misdemeanor Crime", y="Frequency", title="Misdemeanor Crimes by Offens
e")+  scale_y_continuous(trans='log10')
```



The above figure shows volume of the Misdemeanor crime. Other Misdemeanor Arrest has the highest rate of 29 followed by Burglary / Habitation of 24.

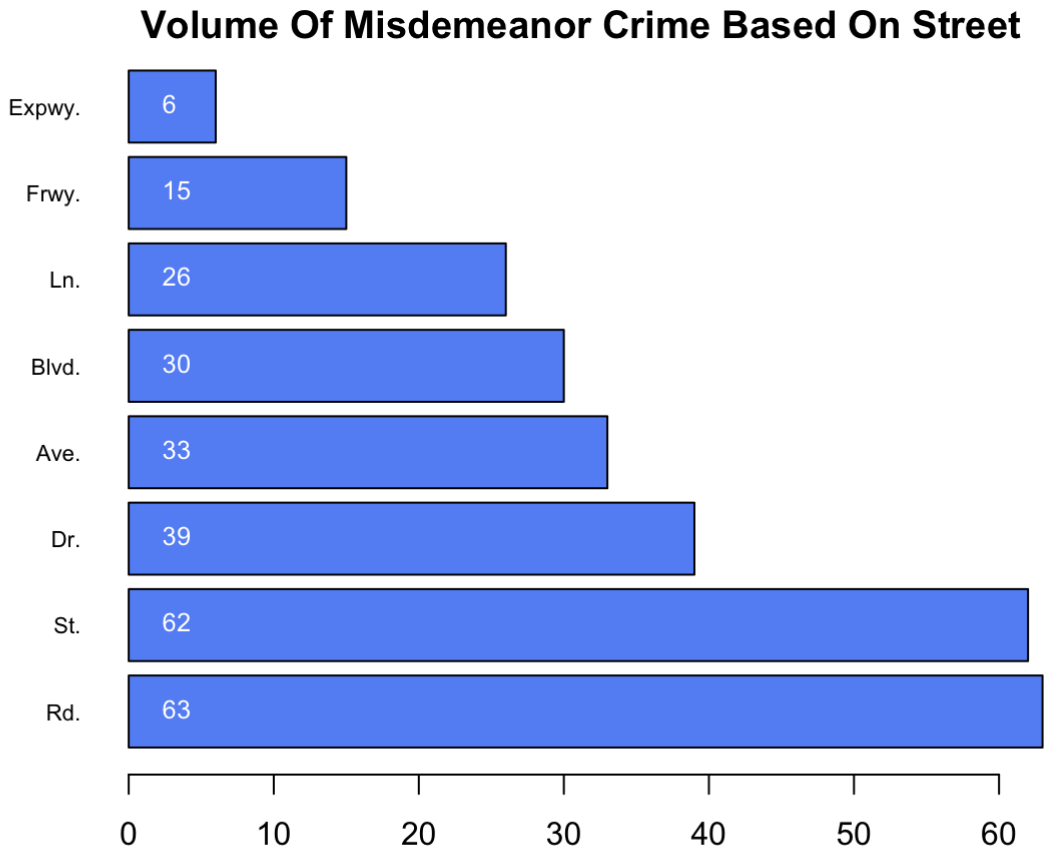
```
crimeRateData$SUBJECT_RACE[crimeRateData$SUBJECT_RACE=="NULL"] <- "Other"
crimeRateData$SUBJECT_GENDER[crimeRateData$SUBJECT_GENDER=="NULL"] <- "Unknown"
misdemeanor <- crimeRateData%>%filter(CATEGORY == "Misdemeanor")
misdemeanorTable <- table(misdemeanor$SUBJECT_RACE, misdemeanor$SUBJECT_GENDER)
kable(formattable(misdemeanorTable), align="l", caption = "Table shows the count of G
ender and Race")
```

Table shows the count of Gender and Race

	Female	Male	Unknown
Asian	0	1	0
Black	19	153	0
Hispanic	7	58	0
Other	0	3	2
White	7	50	0

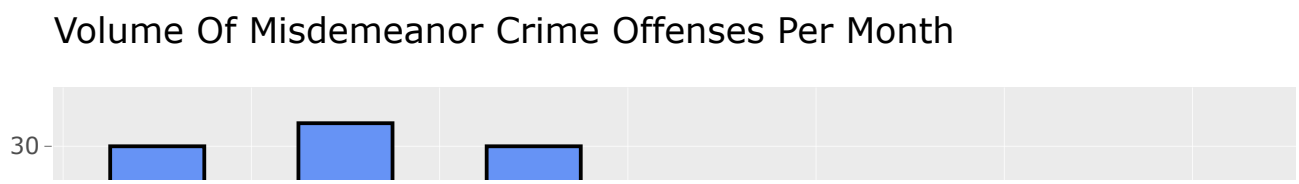
As per the above table, from a gender point of view men commit more misdemeanor crimes than women and from a racial perspective, the people of Black origin commit more misdemeanor than others followed by the Hispanic.

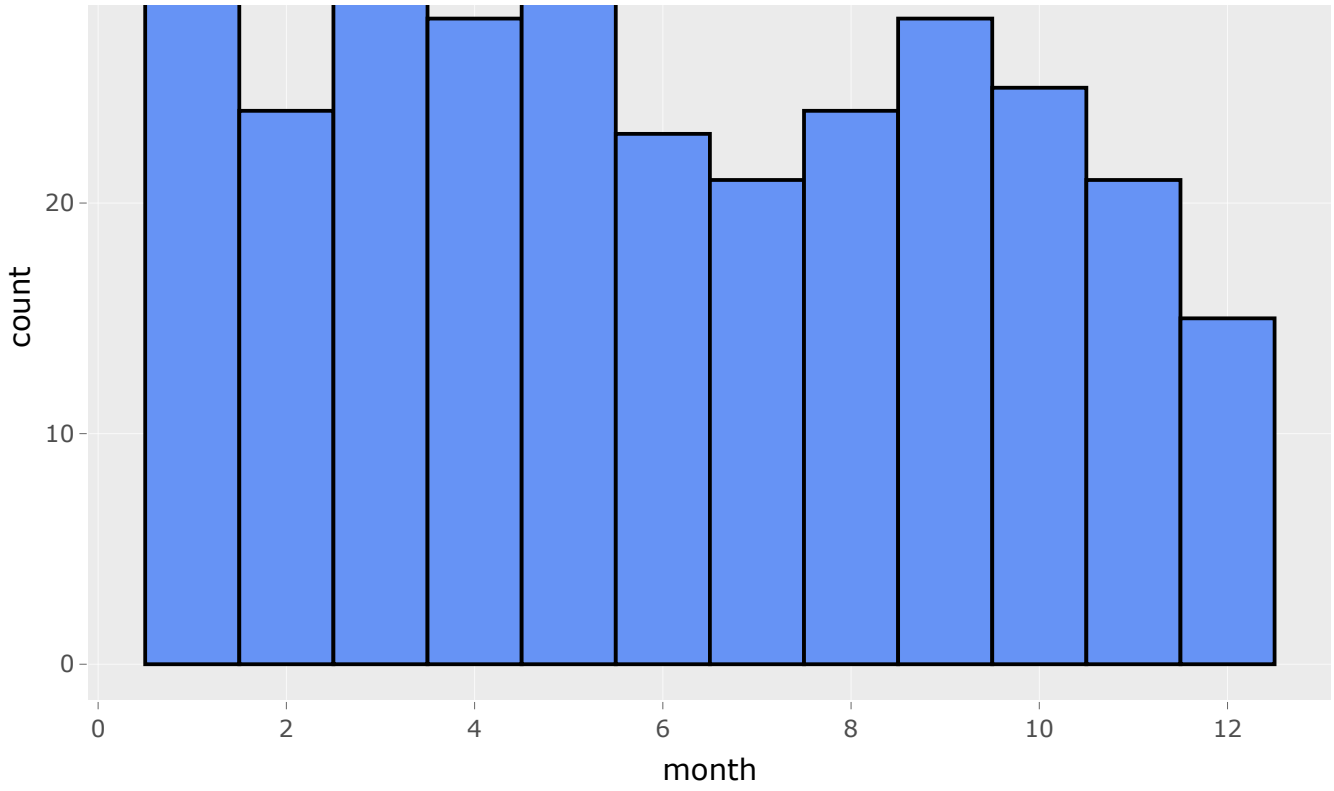
```
crimeRateData <- crimeRateData[!(is.na(crimeRateData$STREET_TYPE) | crimeRateData$STR
EET_TYPE=="NULL"), ]
crimeRateData <- crimeRateData%>%mutate(street = str_remove(crimeRateData$STREET_TYP
E, "[.]"))
misdemeanor <- crimeRateData%>%filter(CATEGORY == "Misdemeanor")
mislocation <- misdemeanor$STREET_TYPE
tb <- table(mislocation)
counts = tb[order(tb, decreasing=T)]
par(mar = c(5,10,1,1))
y = barplot(counts[1:8], horiz=T, las=1, cex.names=0.7, col="#6693F5", main="Volume O
f Misdemeanor Crime Based On Street")
text(1, y, counts[1:8], pos=4, cex=0.8, col="white")
```



The above bar plot shows the highest rate of Misdemeanor crime reported in Road (63) and lowest in Expressway (6).

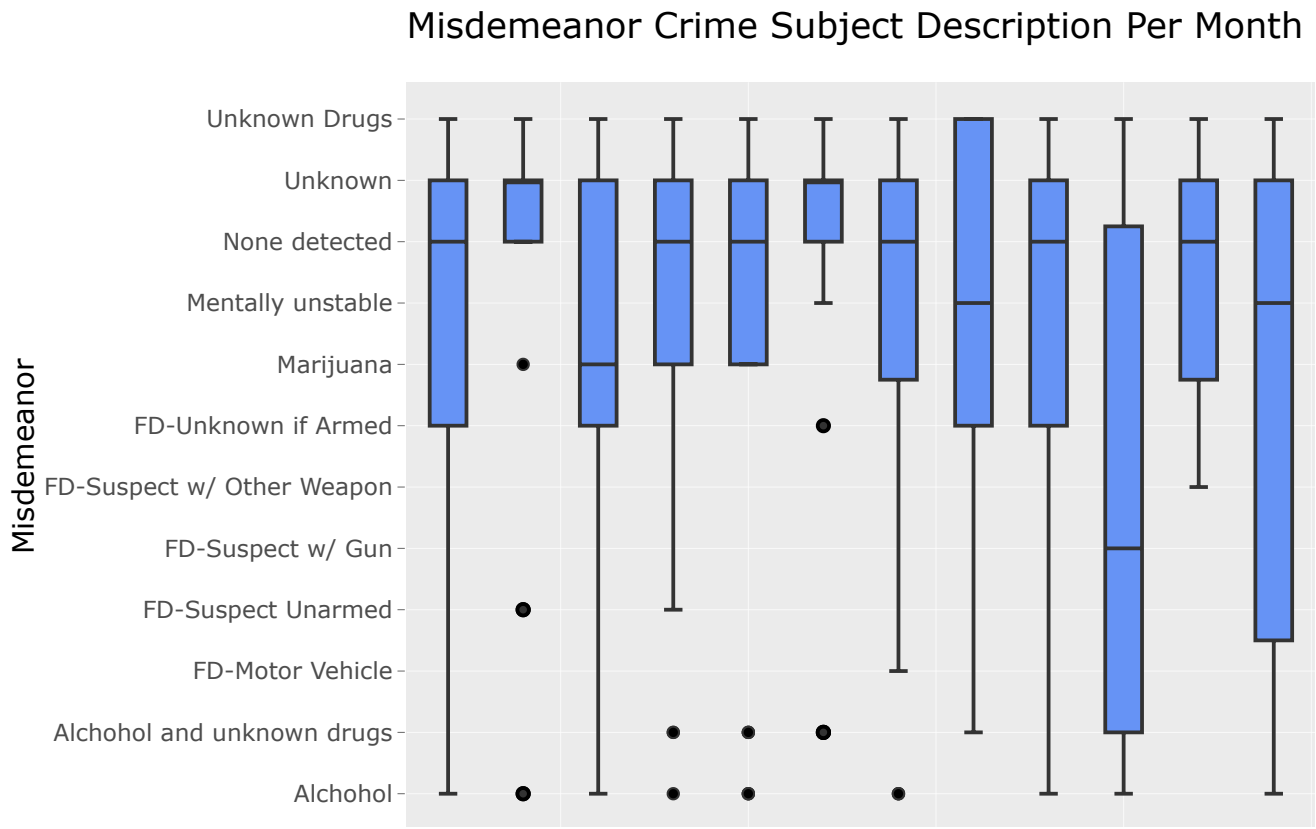
```
misdemeanor <- crimeMonthData%>%filter(CATEGORY == "Misdemeanor")
G = misdemeanor %>% ggplot(aes(month)) + geom_histogram(bins=12, fill="#6693F5", colo
r = "black")+ scale_x_continuous(breaks =c(0, 2, 4,6,8,10,12))+labs(title="Volume Of
Misdemeanor Crime Offenses Per Month")
ggplotly(G)
```

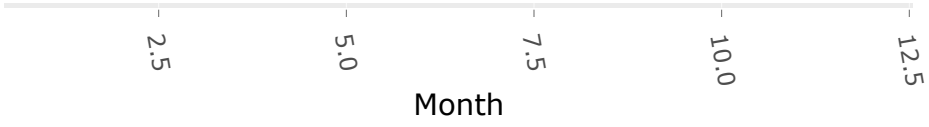




The above plot depicts that number of Misdemeanor crimes offenses are highest in the month of March and lowest in the month of December.

```
misdemeanor <- crimeMonthData%>%filter(CATEGORY == "Misdemeanor")
p <- ggplot(misdemeanor, aes(x = month, y = factor(SUBJECT_DESCRIPTION))) +
geom_boxplot(fill="#6693F5")+ theme( axis.text.x = element_text(angle = -80, vjust =
1, hjust = 0))+ labs(x = "Month", y="Misdemeanor",title="Misdemeanor Crime Subject De
scription Per Month")
ggplotly(p)
```

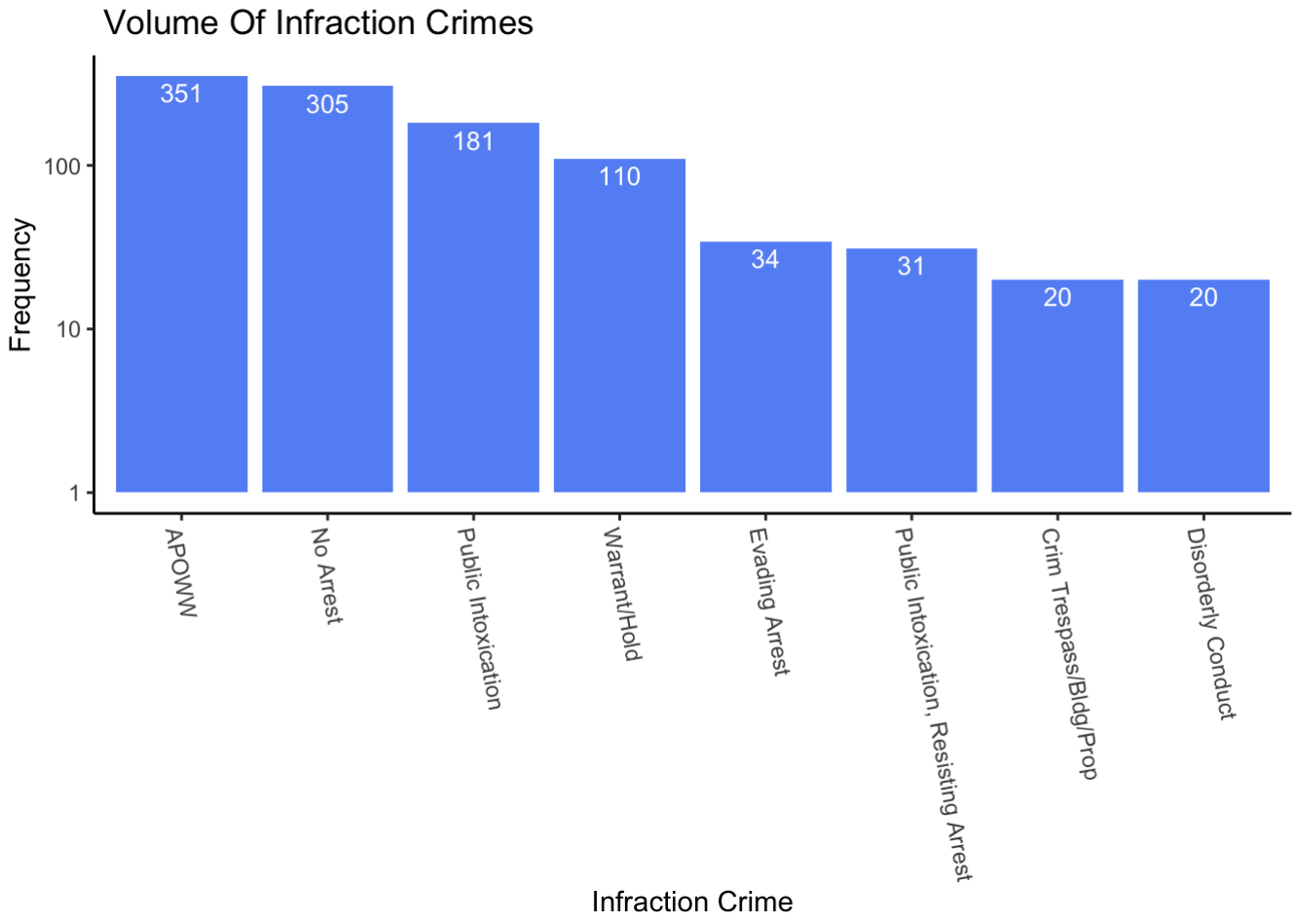




The above figure shows the subject description at the time of arrest for each month. The maximum value is shown for Unknown Drugs and minimum Alcohol. Outliers are found in the month of February, April, May, June, July.

Infraction Crime

```
crimeRateData <- crimeCategory
infraction <- crimeCategory%>%filter(CATEGORY == "Infraction")
infractionTB <- table(infraction$SUBJECT_OFFENSE)
infractionCount = as.data.frame(infractionTB[order(infractionTB, decreasing=T)])
infractionCount <- infractionCount[1:8, ]
ggplot(data=infractionCount, aes(x=Var1, y=Freq)) +
  geom_bar(stat="identity", fill="#6693F5")+
  geom_text(aes(label=Freq), vjust=1.4, color="white", size=3.5)+
  theme_classic()+ theme(axis.text.x = element_text(angle = -80, vjust = 1, hjust =
0))+labs(x = "Infraction Crime", y="Frequency", title=" Volume Of Infraction Crimes")
+  scale_y_continuous(trans='log10')
```



From the above figure it is evident that APPOW has the highest frequency of 351 followed by No Arrest of 305.

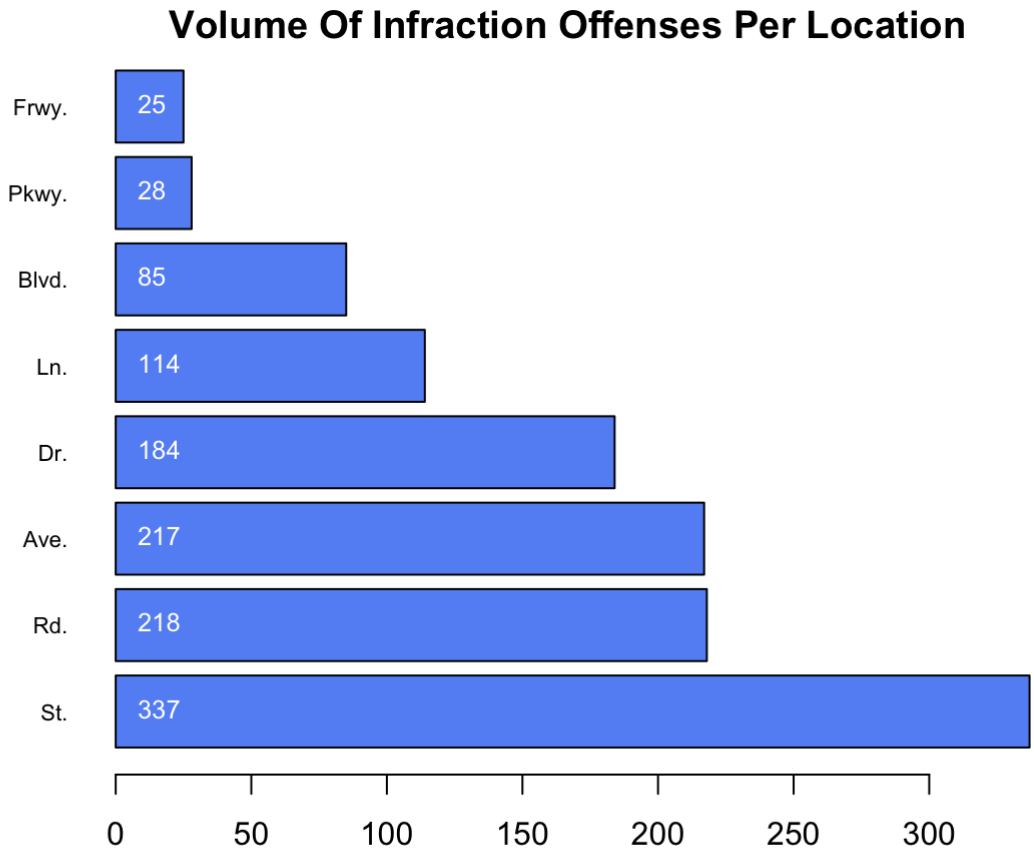
```
crimeRateData$SUBJECT_RACE[crimeRateData$SUBJECT_RACE=="NULL"] <- "Other"
crimeRateData$SUBJECT_GENDER[crimeRateData$SUBJECT_GENDER=="NULL"] <- "Unknown"
infraction <- crimeRateData%>%filter(CATEGORY == "Infraction")
infractionTable <- table(infraction$SUBJECT_RACE, infraction$SUBJECT_GENDER)
kable(formattable(infractionTable), align="l", caption = "Table shows the count of Ge
nder and Race")
```

Table shows the count of Gender and Race

	Female	Male	Unknown
American Ind	0	1	0
Asian	0	2	0
Black	163	596	1
Hispanic	44	245	0
Other	4	28	2
White	75	224	0

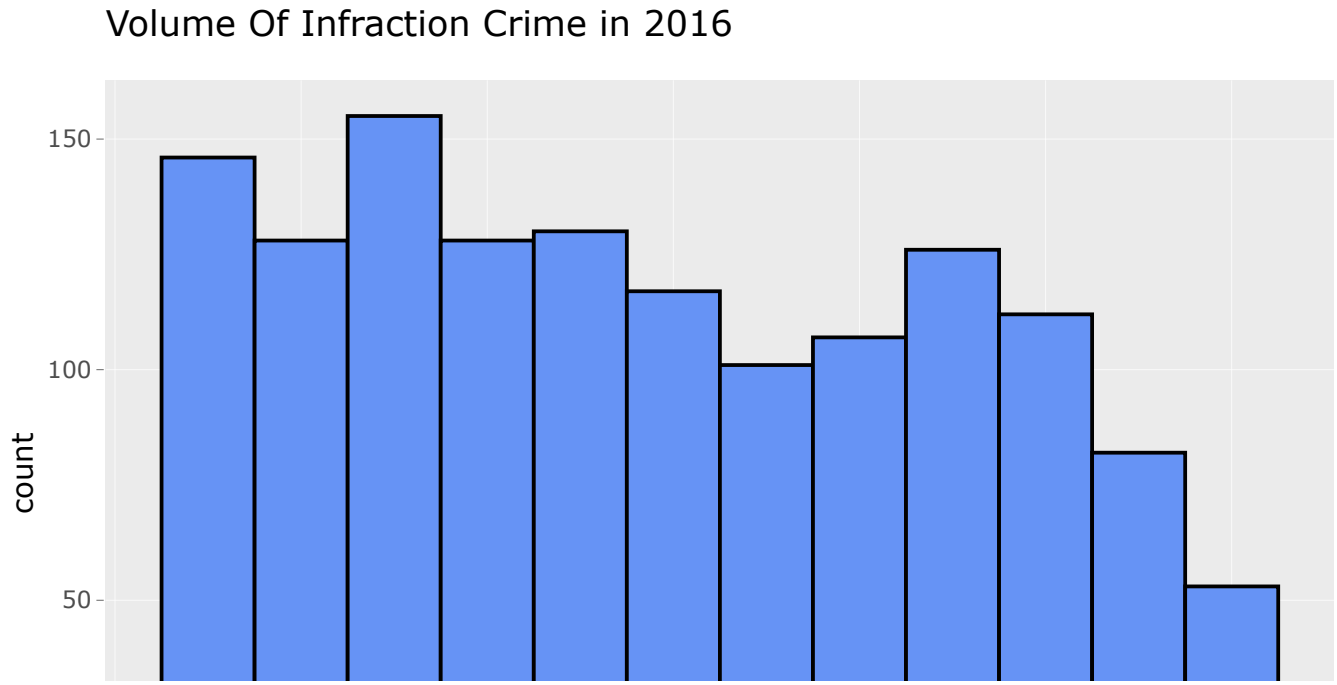
As per the above table, from a gender point of view men commit more infraction crimes than women and from a racial perspective, the people of Black origin commit more infraction than others followed by the White and Hispanic.

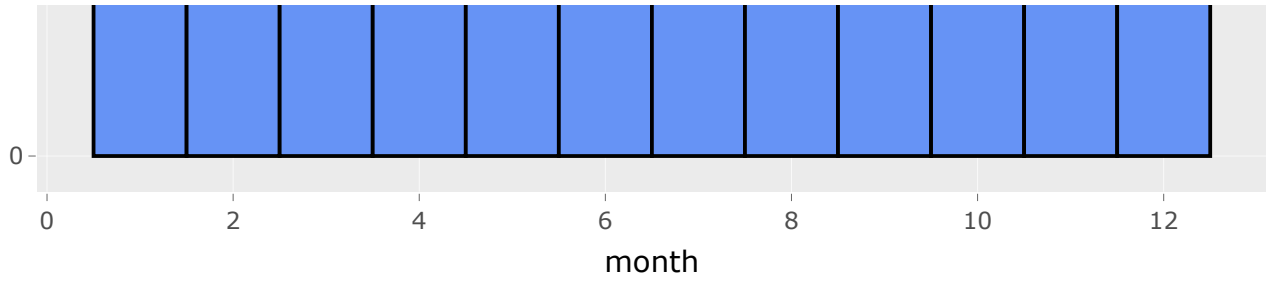
```
crimeRateData <- crimeRateData[!(is.na(crimeRateData$STREET_TYPE) | crimeRateData$STR
EET_TYPE=="NULL"), ]
crimeRateData <- crimeRateData%>%mutate(street = str_remove(crimeRateData$STREET_TYP
E, "[.]"))
infraction <- crimeRateData%>%filter(CATEGORY == "Infraction")
inflocation <- infraction$STREET_TYPE
tb <- table(inflocation)
counts = tb[order(tb, decreasing=T)]
par(mar = c(5,10,1,1))
y = barplot(counts[1:8], horiz=T, las=1, cex.names=0.7, col="#6693F5", main="Volume O
f Infraction Offenses Per Location")
text(1, y, counts[1:8], pos=4, cex=0.8, col="white")
```



The above table shows the highest frequency of Infraction crime reported in Street (337) and lowest in Freeway (28).

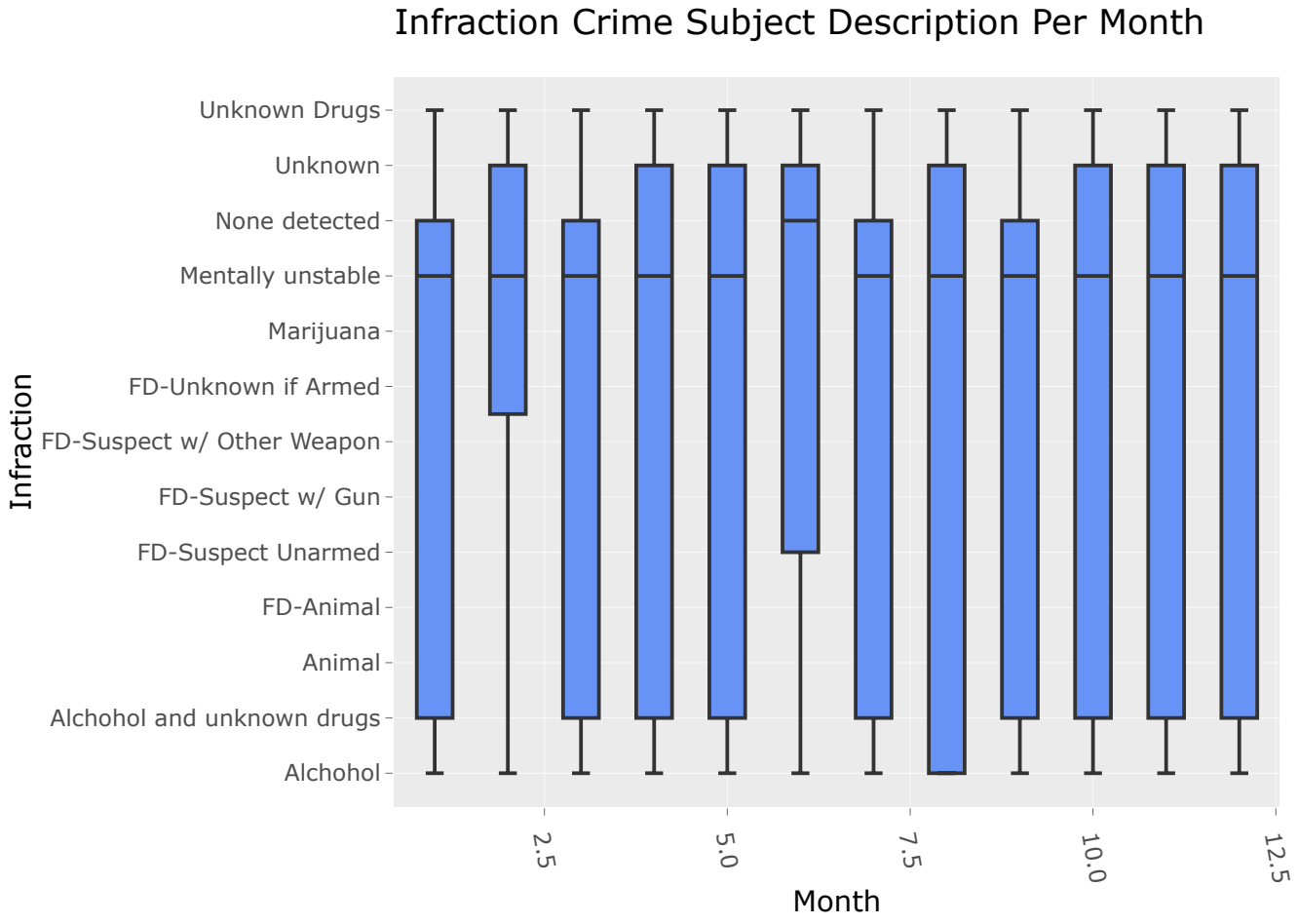
```
infraction <- crimeMonthData%>%filter(CATEGORY == "Infraction")
G = infraction %>% ggplot(aes(month)) + geom_histogram(bins=12, fill="#6693F5", color
= "black")+ scale_x_continuous(breaks =c(0, 2, 4,6,8,10,12))+labs(title="Volume Of In
fraction Crime in 2016")
ggplotly(G)
```





The above plot shows the count of Infraction crimes per month. Highest reported in the month of March and lowest in the month of December.

```
infraction <- crimeMonthData%>%filter(CATEGORY == "Infraction")
p <- ggplot(infraction, aes(x = month, y = factor(SUBJECT_DESCRIPTION))) +
geom_boxplot(fill="#6693F5")+ theme( axis.text.x = element_text(angle = -80, vjust =
1, hjust = 0))+ labs(x = "Month", y="Infraction",title="Infraction Crime Subject Desc
ription Per Month")
ggplotly(p)
```



The above plot shows the subject description at the time of arrest for each month. The maximum value is found to be Unknown Drugs and minimum Alcohol.

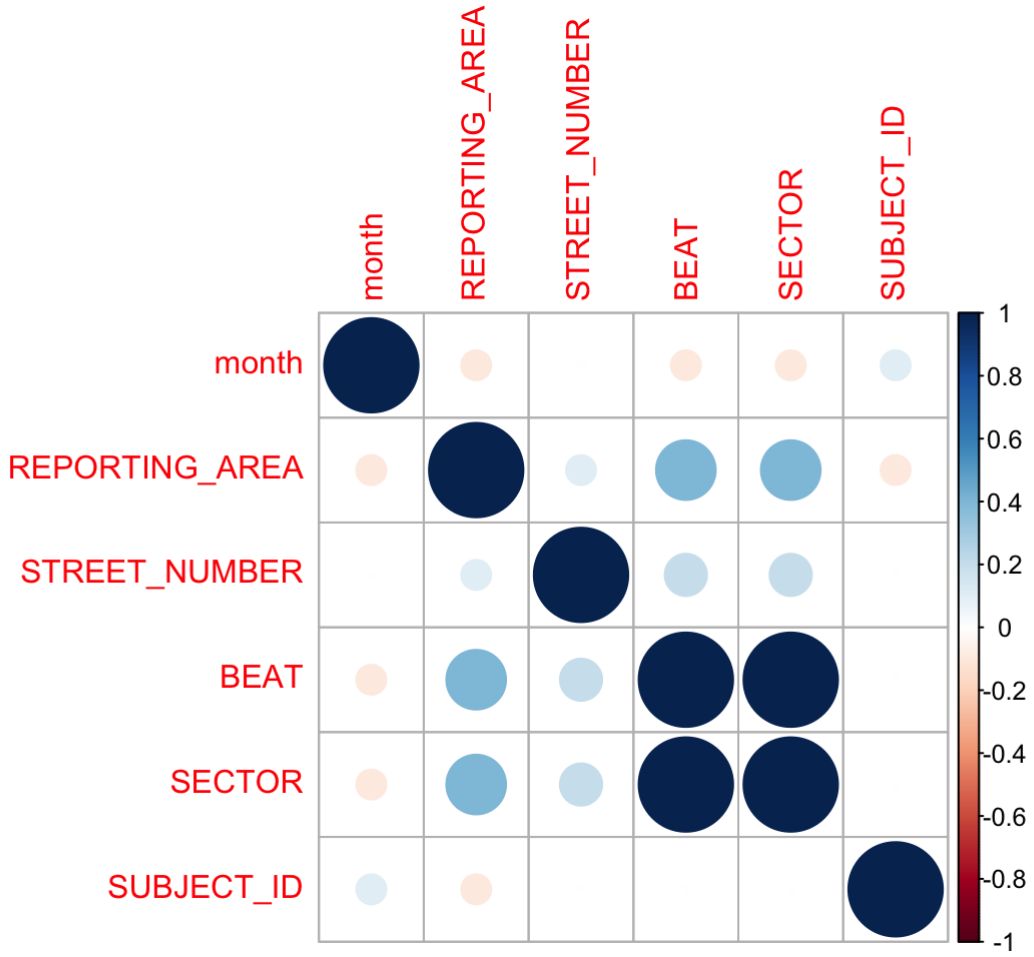
Correlation Between Month,REPORTING AREA, STREET

NUMBER, BEAT, SECTOR,SUBJECT ID

```
CorData <- crimeMonthData
attach(CorData)
CorData$OFFICER_ID <- as.numeric(OFFICER_ID)
CorData$REPORTING_AREA <- as.numeric(REPORTING_AREA)
CorData$STREET_NUMBER <- as.numeric(STREET_NUMBER)
CorData$SUBJECT_ID <- as.numeric(SUBJECT_ID)
CorData$UOF_NUMBER <- as.numeric(UOF_NUMBER)
```

```
## Warning: NAs introduced by coercion
```

```
CorData$STREET_NUMBER <- as.numeric(STREET_NUMBER)
CorData$BEAT <- as.numeric(BEAT)
CorData$SECTOR <- as.numeric(SECTOR)
df <- CorData %>%select(month,REPORTING_AREA, STREET_NUMBER, BEAT, SECTOR,SUBJECT_ID)
corr.mat <- round(cor(df), 1)
# To show the diagonl of the data
pval.cor <- cor_pmat(df)
corrplot(corr.mat, method="circle")
```

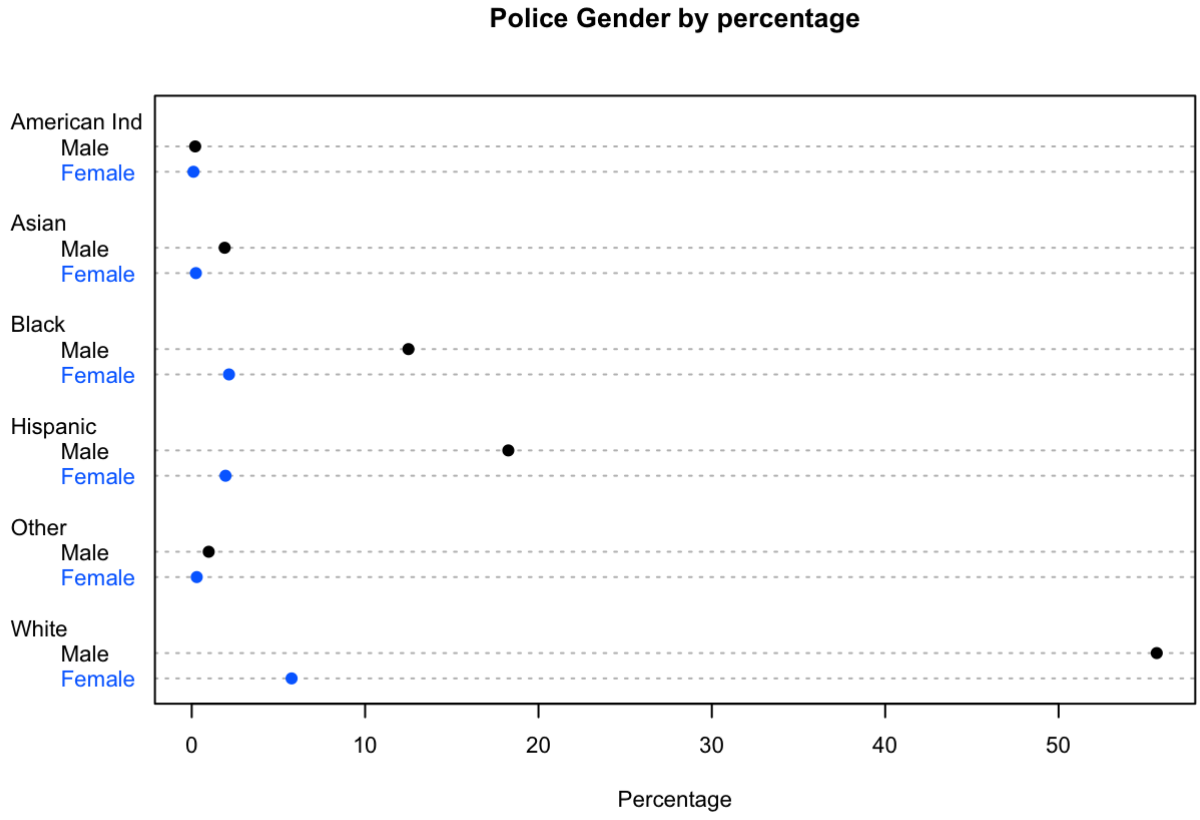



```
detach(CorData)
```

The above figure shows a correlation between the street number, Beat and Sector. Also a correlation is visible with Subject Id and month. Thers is a high correlation between Beat and Sector. A correlation is visible between the Reporting area, Beat, Sector and small correlation with Reporting area and Street Number.

Police Details

```
policeDetails <- crimeData
attach(policeDetails)
gender <- (c("Male","Female"))
policeDetailsDF<-as.data.frame(round( 100*prop.table(table(OFFICER_GENDER, OFFICER_RACE)), 2))
policeDetailsDF<-policeDetailsDF[order(policeDetailsDF$Freq),]
policeDetailsDF$color[policeDetailsDF$OFFICER_GENDER=="Male"] <- "black"
policeDetailsDF$color[policeDetailsDF$OFFICER_GENDER=="Female"] <- "#006FFF"
dotchart(policeDetailsDF$Freq,labels=policeDetailsDF$OFFICER_GENDER,groups=factor(policeDetailsDF$OFFICER_RACE),color=policeDetailsDF$color,cex=.7,pch=19,main="Police Gender by percentage",xlab="Percentage")
```




```
detach(policeDetails)
```

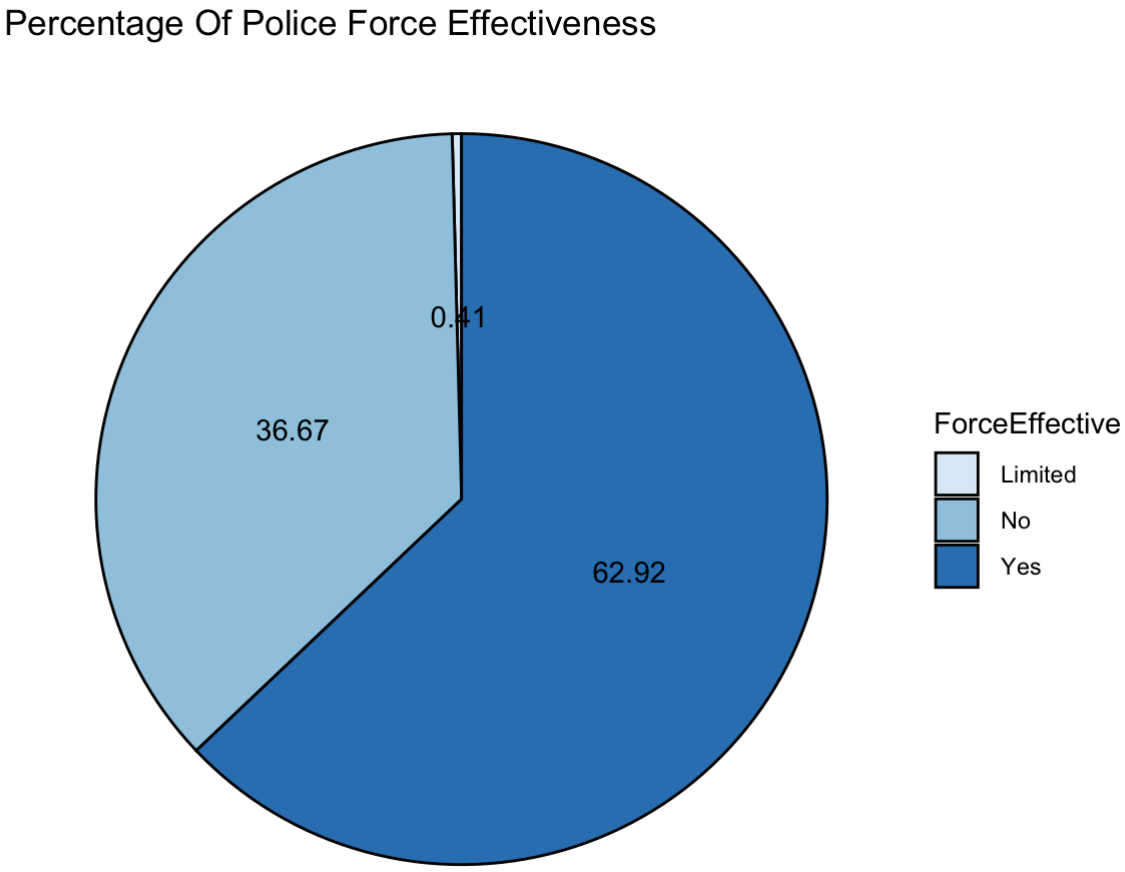
The above table shows male-domination in the police force and from a racial perspective, White race are dominant followed by Hispanic and Black.

Police Efficiency

```
forceEffective <- policing
# separate the csv string to multiple columns
forceEffectiveDF <- separate(data = forceEffective, col =FORCE_EFFECTIVE, into = c("v
ar1", "var2","var3"), sep = "\\,", convert = TRUE, remove = FALSE, fill="right")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 387 rows [38, 40, 48, 4
9, 53,
## 58, 69, 70, 93, 95, 110, 114, 115, 117, 121, 140, 141, 142, 147, 149, ...].
```

```
# merging column values
forceEffectiveData <- data.frame(c(forceEffectiveDF$var1,forceEffectiveDF$var2, force
EffectiveDF$var3,forceEffectiveDF$var4))
forceEffectiveDataDF <- as.data.frame(round( 100*prop.table(table(forceEffectiveDat
a)),2))
names(forceEffectiveDataDF)[1]<-"ForceEffective"
ggplot(forceEffectiveDataDF, aes(x = "", y = Freq, fill = ForceEffective)) +
  geom_col(color = "black") +
  geom_text(aes(label = Freq),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y") +
  scale_fill_brewer() +
  theme_void()+labs(title="Percentage Of Police Force Effectiveness")
```



Despite the high density of crimes, the data shows that law enforcement agencies are effective in controlling the situation. The above Pie Chart shows the effectiveness of the ability of police departments to investigate and solve the crimes.