MA304-Report:2214122

Data visualization on policing dataset from Dallas,Texas

Introduction

The Center for Policing Equity is a gathering place for experts from various fields, including those dealing with race and equity issues, data analysis specialists, and research scientists. A significant amount of data has been made accessible for the examination of crime and police behavioral data as a result of the participation of police departments from throughout the USA in the database. Their goal is to expose any racial discrepancies in policing and offer suggestions for mending and reestablishing public trust as well as for enhancing public safety.[1]

The Center of Policing Equity (CPE) contributed the policing data set used in our project, which was collected from kaggle and contains the policing data from Dallas, Texas in 2016. The data collection has 2383 records and 47 variables. Our goal is to study various variables and the relationships between them, such as subject race and crime rate, police race and the cases handled by them, and see if any patterns or racial disparities can be found in the data.

Methods

We will perform data analysis,visualize and interpret the data.Firstly the date fields in the data-set have been changed to date format and from the INCIDENT_DATE field we have extracted the Incident month name,Incident month number and the Incident day using functions from lubridate library.Then we have used bar plots,density plots,scatter plots,box plots,time series plot,correlation analysis and leaflet.

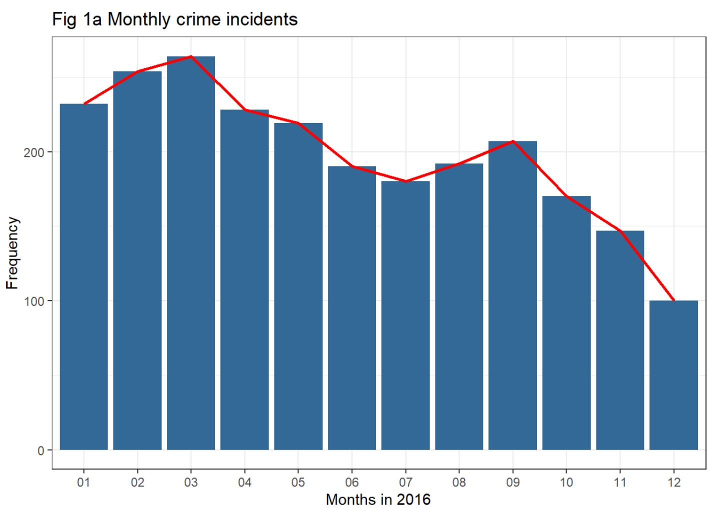
Results

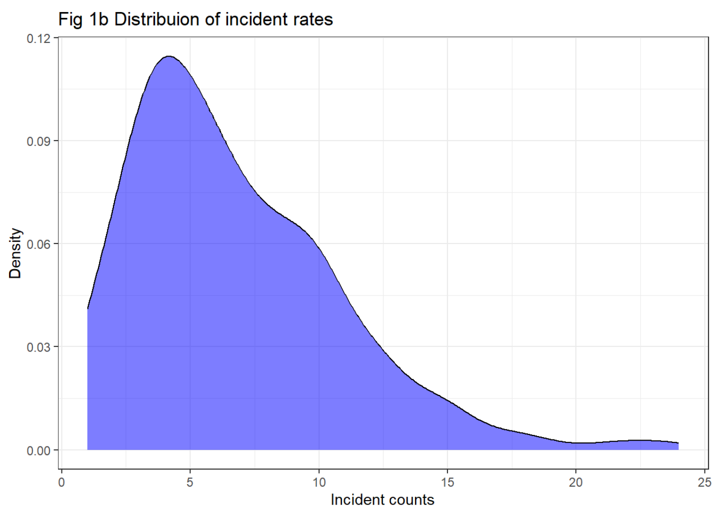
```
#check the data set
str(crimedata)
```

```
## 'data.frame':    2383 obs. of  47 variables:
##  $ INCIDENT_DATE                             : chr  "9/3/2016" "3/22/2016" "5/22/2016" "1/10/2016" ...
##  $ INCIDENT_TIME                             : chr  "4:14:00 AM" "11:00:00 PM" "1:29:00 PM" "8:55:00 PM" ...
##  $ UOF_NUMBER                                : chr  "37702" "33413" "34567" "31460" ...
##  $ OFFICER_ID                                : chr  "10810" "7706" "11014" "6692" ...
##  $ OFFICER_GENDER                            : chr  "Male" "Male" "Male" "Male" ...
##  $ OFFICER_RACE                              : chr  "Black" "White" "Black" "Black" ...
##  $ OFFICER_HIRE_DATE                         : chr  "5/7/2014" "1/8/1999" "5/20/2015" "7/29/1991" ...
##  $ OFFICER_YEARS_ON_FORCE                    : chr  "2" "17" "1" "24" ...
##  $ OFFICER_INJURY                            : chr  "No" "Yes" "No" "No" ...
##  $ OFFICER_INJURY_TYPE                       : chr  "No injuries noted or visible" "Sprain/Strain" "No injuries
noted or visible" "No injuries noted or visible" ...
##  $ OFFICER_HOSPITALIZATION                   : chr  "No" "Yes" "No" "No" ...
##  $ SUBJECT_ID                                : chr  "46424" "44324" "45126" "43150" ...
##  $ SUBJECT_RACE                              : chr  "Black" "Hispanic" "Hispanic" "Hispanic" ...
##  $ SUBJECT_GENDER                            : chr  "Female" "Male" "Male" "Male" ...
##  $ SUBJECT_INJURY                            : chr  "Yes" "No" "No" "Yes" ...
##  $ SUBJECT_INJURY_TYPE                       : chr  "Non-Visible Injury/Pain" "No injuries noted or visible" "N
o injuries noted or visible" "Laceration/Cut" ...
##  $ SUBJECT_WAS_ARRESTED                      : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ SUBJECT_DESCRIPTION                       : chr  "Mentally unstable" "Mentally unstable" "Unknown" "FD-Unkno
wn if Armed" ...
##  $ SUBJECT_OFFENSE                           : chr  "APOWW" "APOWW" "APOWW" "Evading Arrest" ...
##  $ REPORTING_AREA                            : chr  "2062" "1197" "4153" "4523" ...
##  $ BEAT                                      : chr  "134" "237" "432" "641" ...
##  $ SECTOR                                    : chr  "130" "230" "430" "640" ...
##  $ DIVISION                                  : chr  "CENTRAL" "NORTHEAST" "SOUTHWEST" "NORTH CENTRAL" ...
##  $ LOCATION_DISTRICT                         : chr  "D14" "D9" "D6" "D11" ...
##  $ STREET_NUMBER                             : chr  "211" "7647" "716" "5600" ...
##  $ STREET_NAME                               : chr  "Ervay" "Ferguson" "bimebella dr" "LBJ" ...
##  $ STREET_DIRECTION                          : chr  "N" "NULL" "NULL" "NULL" ...
##  $ STREET_TYPE                               : chr  "St." "Rd." "Ln." "Frwy." ...
##  $ LOCATION_FULL_STREET_ADDRESS_OR_INTERSECTION: chr  "211 N ERVAY ST" "7647 FERGUSON RD" "716 BIMEBELLA LN" "560
0 L B J FWY" ...
##  $ LOCATION_CITY                             : chr  "Dallas" "Dallas" "Dallas" "Dallas" ...
##  $ LOCATION_STATE                            : chr  "TX" "TX" "TX" "TX" ...
##  $ LOCATION_LATITUDE                         : chr  "32.782205" "32.798978" "32.73971" "" ...
##  $ LOCATION_LONGITUDE                        : chr  "-96.797461" "-96.717493" "-96.92519" "" ...
##  $ INCIDENT_REASON                           : chr  "Arrest" "Arrest" "Arrest" "Arrest" ...
##  $ REASON_FOR_FORCE                          : chr  "Arrest" "Arrest" "Arrest" "Arrest" ...
##  $ TYPE_OF_FORCE_USED1                       : chr  "Hand/Arm/Elbow Strike" "Joint Locks" "Take Down - Group"
"K-9 Deployment" ...
##  $ TYPE_OF_FORCE_USED2                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED3                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED4                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED5                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED6                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED7                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED8                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED9                       : chr  "" "" "" "" ...
##  $ TYPE_OF_FORCE_USED10                      : chr  "" "" "" "" ...
##  $ NUMBER_EC_CYCLES                          : chr  "NULL" "NULL" "NULL" "NULL" ...
##  $ FORCE_EFFECTIVE                           : chr  " Yes" " Yes" " Yes" " Yes" ...
```

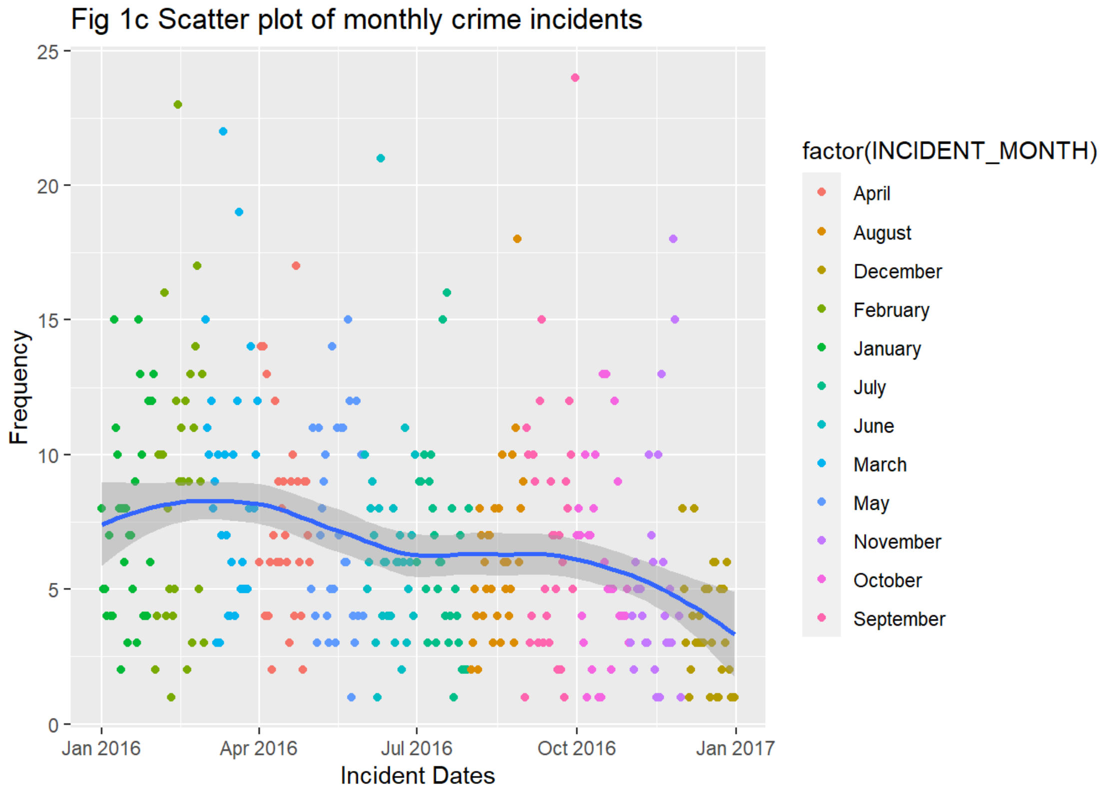
Visualization of crime rates

```
#create a plot for showing crime rates per month in 2016
# remove the na values
  crimedata_month <-  crimedata %>%
  group_by(INC_MONTH) %>%
  filter(INC_MONTH!="NA") %>%
  summarize(count = n())


#plot a ggplot with columns using geom_col() and geom_line() to show crime distribution over 12 months
  crimedata_month %>%
  ggplot(aes(INC_MONTH,count,fill=1)) +
  geom_col(show.legend=FALSE) +
  geom_line(crimedata_month, group=1,mapping=aes(x=INC_MONTH, y =count),
           size = 1,colour ="red") +
  xlab("Months in 2016") +
  ylab("Frequency") +
  ggtitle("Fig 1a Monthly crime incidents")+
  theme_bw()

  #group the crime data by date,month and day
  cr_year <-  crimedata %>%
  group_by(INCIDENT_DATE,INCIDENT_MONTH,INCIDENT_DAY) %>%
  filter(INCIDENT_MONTH!="NA") %>%
  summarize(count = n())

  #density plot for overall distribution of crimes
  ggplot(cr_year, aes(count)) +
  geom_density(alpha = 0.5, colour = "black", fill ="blue")+
  labs(x="Incident counts", y= "Density",
  title="Fig 1b Distribuion of incident rates")  +
  theme_bw()
```






```
 #scatter plot with smoothing to illustrate pattern
   cr_year %>%
   ggplot(aes(x=INCIDENT_DATE,y=count)) +
   geom_point(aes(colour=factor(INCIDENT_MONTH))) +
   xlab("Incident Dates")+
   ylab("Frequency") +
  ggtitle("Fig 1c Scatter plot of monthly crime incidents")+
   stat_smooth()
```



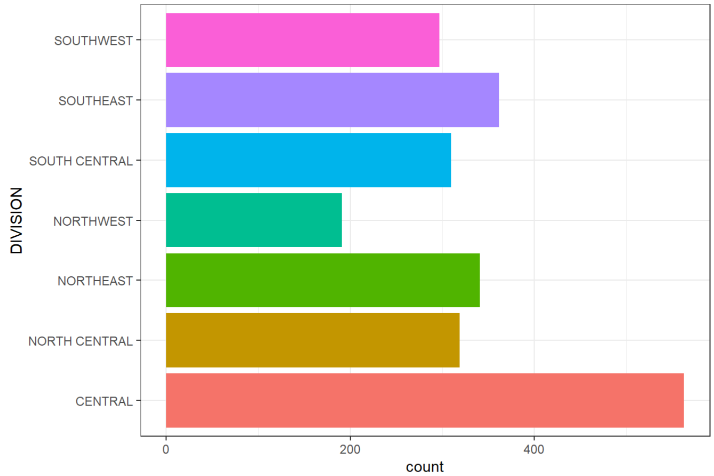
It can be observed from figure 1a that the crime rate is high in January and increases in February,with the highest peak in march.Then the incidents start decreasing gradually with another peak in September and it is lowest in December. Figure 1b shows the overall distribution of crime and the density curve is right skewed.The most number of incidents is 2 to 4 per day with the number of incidents from 15 to 25 very less.The scatter plot(Figure 1c) also shows the same trend throughout the whole year where the smoothing line shows us the upward trend in crime from January to March,with the curve going down for the middle months and then again goign upward September,October and lowest in December.

Crimes across the different divisions in Dallas,Texas

```
#crimes in each division spread over 12 months
  cr_division <-  crimedata %>%
  group_by(INC_MONTH,DIVISION) %>%
  filter(INC_MONTH!="NA") %>%
  summarize(count = n())

#ggplot showing crimes in each division of dallas
  cr_division %>%
  ggplot(aes(INC_MONTH,count,fill = DIVISION)) +
  geom_col(show.legend=FALSE) +
  xlab("Months")+
  ylab("Frequency")  +
  ggtitle("Monthly crime rates across divisions") +
  facet_wrap(~DIVISION,scales="free")   +
  ggtitle("Fig 2a Monthly crime incidents per division")+
  theme_bw()

# Create bar chart with ggplot2
cr_division %>%
ggplot(aes(x = DIVISION,y=count,fill=DIVISION)) +
geom_bar(stat = "identity",show.legend = FALSE) +
ggtitle("Fig 2b Frequency of crime incidents per division")+
coord_flip() +
theme_bw()

cat("\n\n")
```

Fig 2a Monthly crime incidents per division

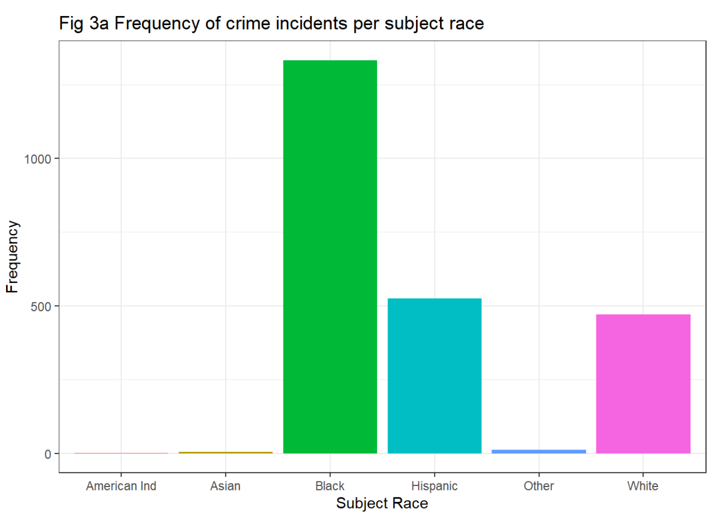
Fig 2b Frequency of crime incidents per division

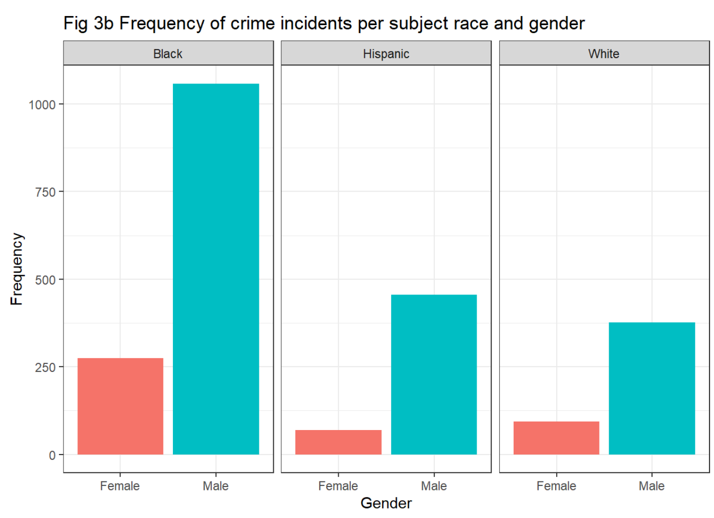


From figure 2a we can see the crimes per division spread over 12 months in 2016.And figure 2b shows the frequency of crimes in each division.From both the figures we can observe that central division has the highest crime rate which is over 50 in some months, followed by southeast and northeast divisions.Southeast has the highest count of over 40 in the month of February and northeast peaks in September with crime rates nearly 40.Northwest has the lowest crime rates with the highest being little over 20 in January.Most of the divisions show a trend of the crime rate going down over the last three months of the year.

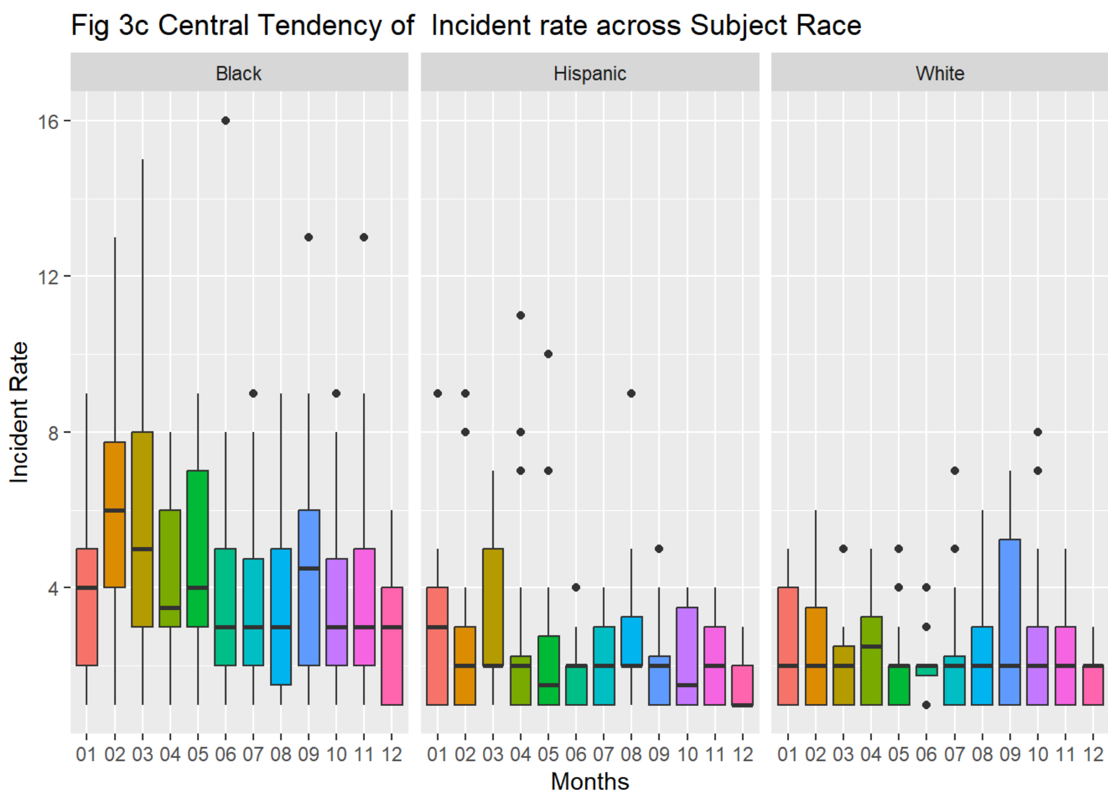
Crime incidents across different races

```
#removing the null values
crsubjectrace = crimedata[!crimedata$SUBJECT_RACE == "NULL", ]
crsubjectracebox = crimedata[!crimedata$SUBJECT_RACE == "NULL", ]

#grouping by crimes committed by subject race per month
 crsubjectrace = crsubjectrace %>%
 group_by(INC_MONTH,SUBJECT_RACE) %>%
 filter(INC_MONTH!="NA") %>%
 summarise(count = n() )

 #box plot of crimes committed by each subject
 crsubjectracebox = crsubjectracebox   %>%
  filter(SUBJECT_RACE == "Black"   | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic" ) %>%
  group_by(INCIDENT_DATE,INC_MONTH,SUBJECT_RACE) %>%
  summarize(avg =n()) #reference [2]

 #Gender count of subjects
 crgender = crimedata[!crimedata$SUBJECT_GENDER == "NULL", ]
 crgender = crgender %>%
 group_by(SUBJECT_RACE,SUBJECT_GENDER) %>%
  filter(SUBJECT_RACE == "Black"   | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic" ) %>%
filter (SUBJECT_GENDER !='Unknown') %>%
 summarise(count = n() )

# Create bar chart with ggplot2
crsubjectrace %>%
ggplot(aes(x = SUBJECT_RACE,y=count,fill=SUBJECT_RACE)) +
geom_bar(stat = "identity",show.legend = FALSE) +
xlab("Subject Race")+ylab("Frequency")+
ggtitle("Fig 3a Frequency of crime incidents per subject race")+
theme_bw()

#Frequency of crime incidents per subject race and gender
crgender %>%
ggplot(aes(x = SUBJECT_GENDER,y=count,fill=SUBJECT_GENDER)) +
geom_col(show.legend=FALSE)  +
ggtitle("Fig 3b Frequency of crime incidents per subject race and gender")+
facet_wrap(~SUBJECT_RACE) +
xlab("Gender") + ylab("Frequency") +
theme_bw()
```





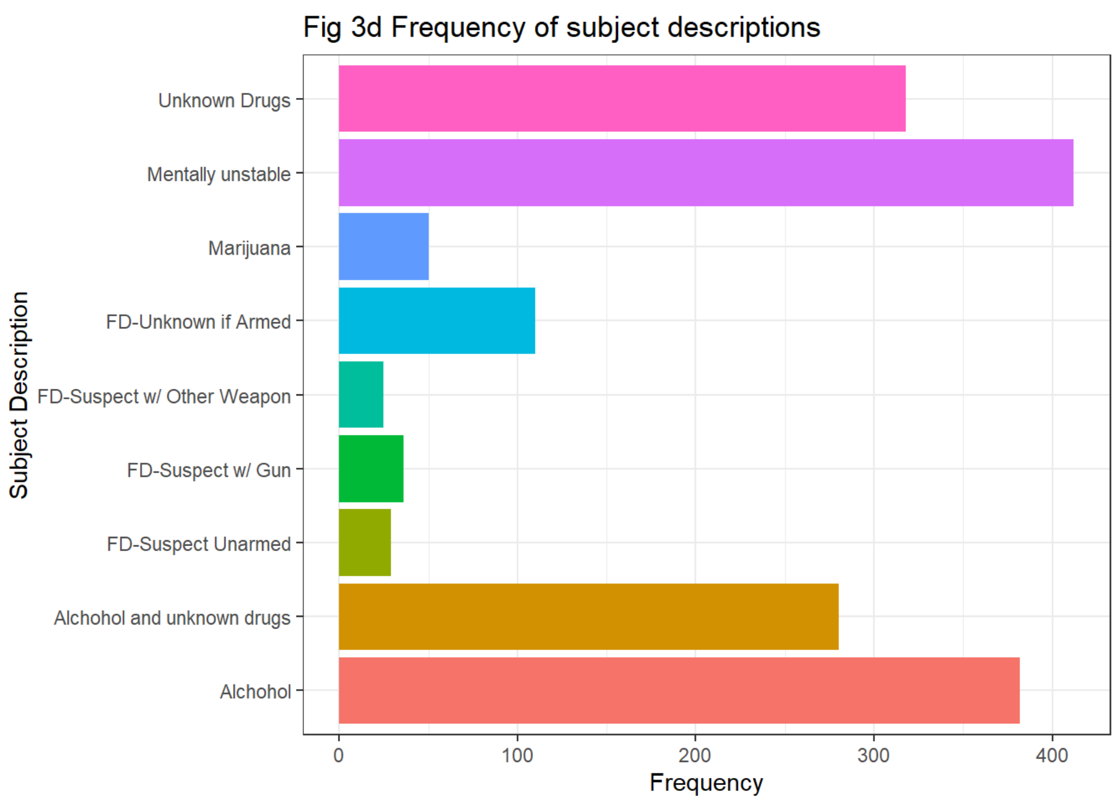
```
#box plot of incident rate
crsubjectracebox %>%
ggplot(aes(x = INC_MONTH, y= avg, fill = INC_MONTH)) +
geom_boxplot() +
labs(x= 'Months',y = 'Incident Rate',
title = paste("Fig 3c Central Tendency of",
              ' Incident rate across Subject Race ')) +
theme(legend.position="none") +
facet_wrap(~SUBJECT_RACE)
```



Firstly a bar graph was plotted in figure 3a to see the frequency of crimes in each race and it was observed that the maximum crimes are committed by black followed by Hispanic and white races.For the rest of the races the numbers are almost negligible.Figure 3b shows that for all three races males committed more crimes than females.
Next a box-plot is done in figure 3c for the top three races involved in crimes.For Black people the highest crime rates is in March followed by February and May and the lowest is in December.
Hispanic people committed maximum crimes in March followed by January and October.December had the lowest crime rates.
September has the highest crime incidents for white people.January and February are the second and third highest months in crime rates respectively.The lowest number of incidents reported was in June.

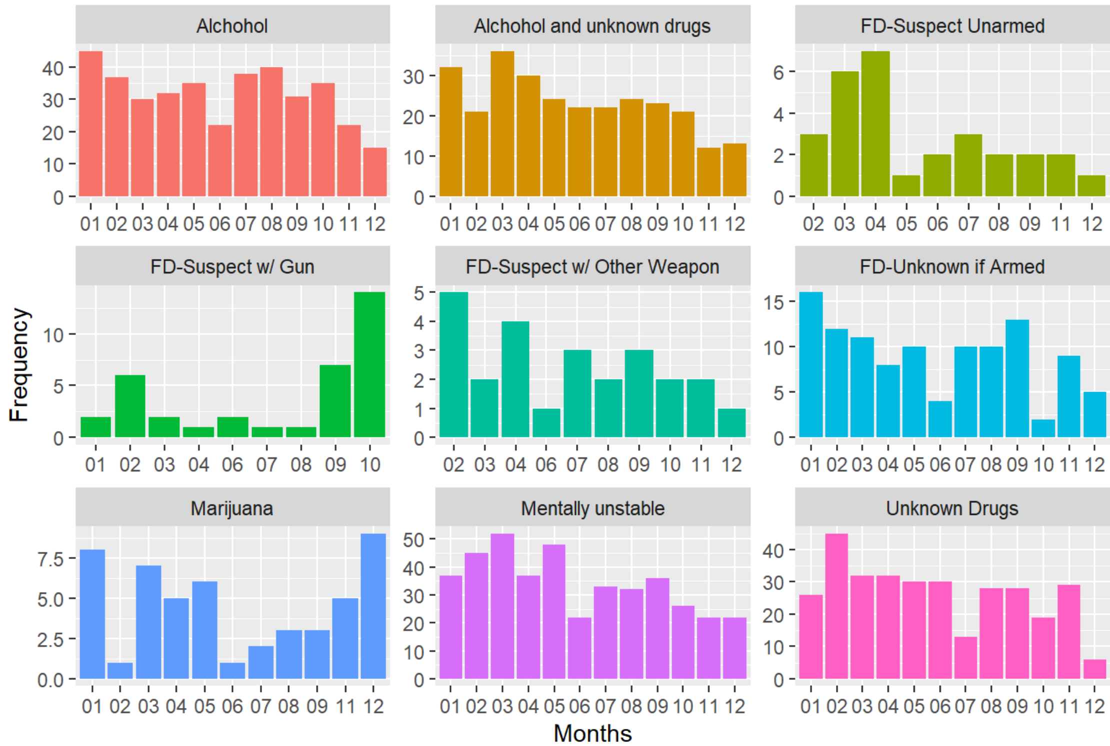
Subject Descriptions

```
crSubDesc = crimedata[!crimedata$SUBJECT_DESCRIPTION == "NULL", ]
crSubDesc=crSubDesc %>%
  group_by(INCIDENT_DATE,INC_MONTH,INCIDENT_DAY,SUBJECT_DESCRIPTION) %>%
  filter(SUBJECT_DESCRIPTION != "FD-Motor Vehicle" & SUBJECT_DESCRIPTION !=     "NULL" & SUBJECT_DESCRIPTION != "FD-
Animal" & SUBJECT_DESCRIPTION != "Animal"
         & SUBJECT_DESCRIPTION !="CIT_INFL_A" &
         SUBJECT_DESCRIPTION !="None detected" & SUBJECT_DESCRIPTION !="Unknown")    %>%
  summarize(count = n())

# Create bar chart with ggplot2
crSubDesc %>%
ggplot(aes(x = SUBJECT_DESCRIPTION,y=count,fill=SUBJECT_DESCRIPTION)) +
geom_bar(stat = "identity",show.legend = FALSE) +
xlab("Subject Description")  +
ylab("Frequency")+
ggtitle("Fig 3d Frequency of subject descriptions")+
coord_flip() +
theme_bw()
```




```
  crSubDesc %>%
  ggplot(aes(INC_MONTH,count,fill=SUBJECT_DESCRIPTION)) +
  geom_col(show.legend=FALSE) +
  xlab("Months")+
  ylab("Frequency")  +
  facet_wrap(~SUBJECT_DESCRIPTION,scales="free")   +
  ggtitle("Fig 3e Monthly distribution of subject description")+
  theme(legend.position="none")
```

Fig 3e Monthly distribution of subject description



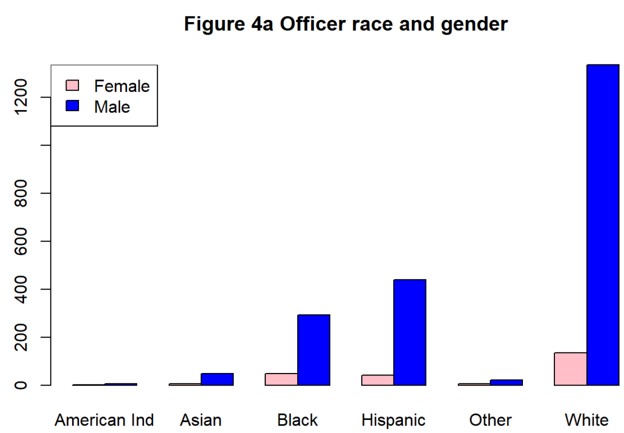
We can observe from Figure 3d that for the people committing crimes the maximum number of people are described as mentally unstable,followed by people under the influence of alcohol and/or unknown drugs.If we see the distribution of the subject descriptions against a timeline (Figure 3e) it shows Mentally Unstable and Unknown Drugs and Alcohol are constantly stable across the year. Suspect with Gun and other weapon has recurring peaks. Marijuana is also mostly low but peaks during some months.

Officer race and gender description

```
#table for officer race and gender
datatable = table(crimedata$OFFICER_GENDER,crimedata$OFFICER_RACE)
knitr::kable(datatable,format = "html",longtable = TRUE)
```

	American Ind	Asian	Black	Hispanic	Other	White
Female	2	7	49	42	6	134
Male	6	48	292	440	21	1336

```
crimedata_offrace <-  crimedata %>%
  group_by(INC_MONTH,INCIDENT_MONTH,OFFICER_RACE) %>%
  summarize(count = n())

#bar plot showing officer race and gender
barplot(datatable, beside = TRUE, main="Figure 4a Officer race and gender",col=c("pink","blue"))
legend("topleft",c("Female","Male"),fill = c("pink","blue"))

#officer race and incidents handled over 12 months
 ggplot() +
  geom_line(data=subset(crimedata_offrace, OFFICER_RACE=="Black" ) ,aes(y=count,x= INC_MONTH,colour="green",group=
1),size=1 ) +
  geom_line(data=subset(crimedata_offrace, OFFICER_RACE=="Hispanic" ) ,aes(y=count,x= INC_MONTH,colour="yellow",grou
p=1),size=1 ) +
  geom_line(data=subset(crimedata_offrace, OFFICER_RACE=="White" ) ,aes(y=count,x= INC_MONTH,colour="red",group=1),s
ize=1 ) +
  scale_color_discrete(name = "Y series", labels = c("BLACKS", "WHITES","HISPANICS")) + labs(x="Months of 2016", tit
le="Figure 4b Officer Race vs Incident Handled Rates")  +
  scale_color_discrete(name = "Legend", labels = c("BLACKS", "WHITES","HISPANICS"))  + theme(axis.text.x=element_tex
t( vjust=0.5),legend.position="bottom")+guides(colour=guide_legend(nrow=2))
 #reference [2]

cat("\n\n")
```



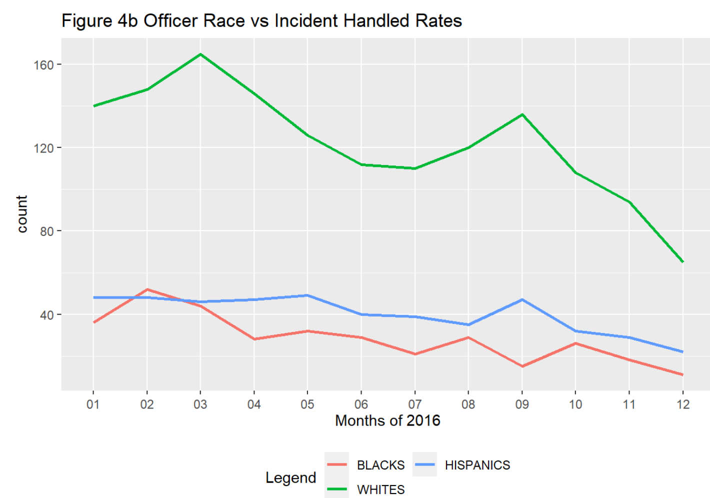
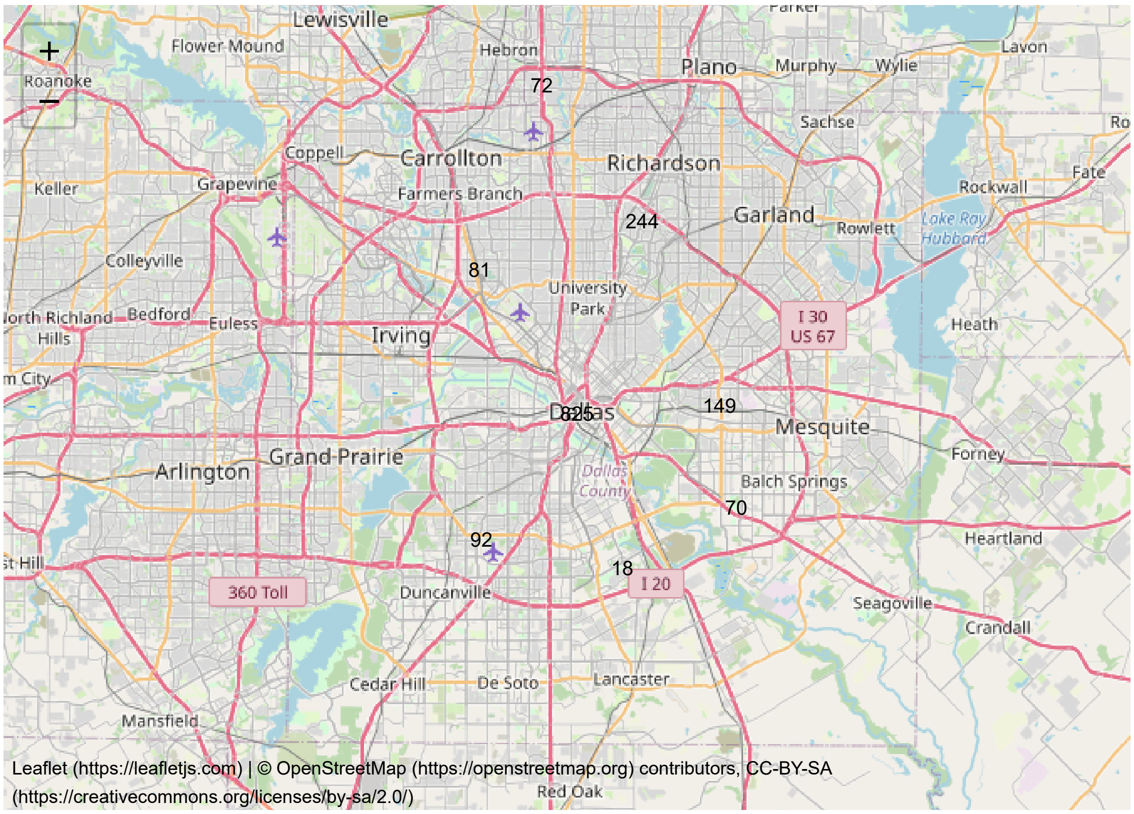


Figure 4a shows that the maximum number of officers are White followed by Hispanics and Blacks.Also the number of males are more than females in the police force.The plot in Figure 4b shows that the maximum cases handled over the months are by the white officers followed by Hispanic and Black.So it is totally opposite compared to the data we saw previously for crimes committed where the Black people were committing highest number of crimes followed by Hispanics and Whites.This indicates that most of the crimes being committed by Black people are being handled by the White police officers.

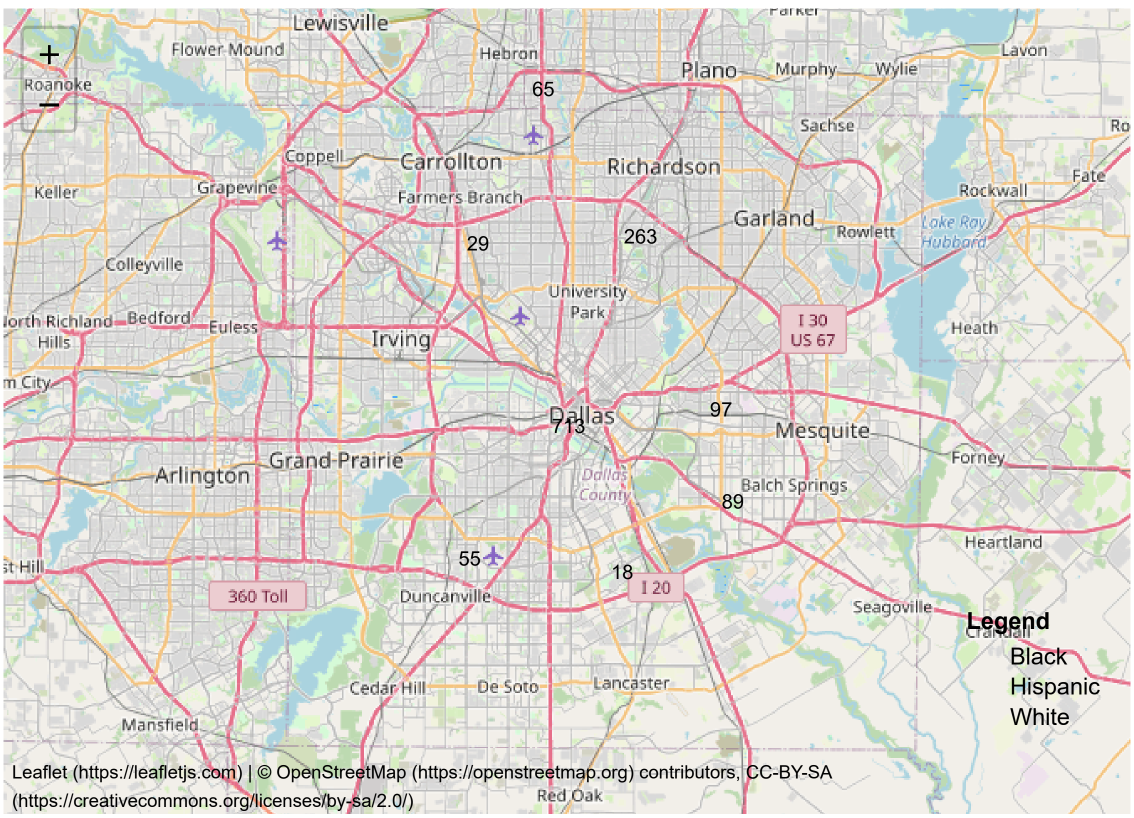
Interactive maps

```r
crimes <- crimedata %>%
  filter(LOCATION_LATITUDE!=''& LOCATION_LONGITUDE!='')%>%
  group_by(INCIDENT_DATE,INCIDENT_MONTH,INCIDENT_DAY,
           LOCATION_LONGITUDE,LOCATION_LATITUDE) %>%
  count() %>%
  arrange(desc(n))

#rename columns
names(crimes) <- c("IncDate","IncMonth","IncDay","longitude","latitude","n")

#transform the longitude and latitude columns to numeric
crimes$longitude = as.numeric(crimes$longitude)
crimes$latitude = as.numeric(crimes$latitude)

#shows the dates of the incidents
 crimes %>% leaflet() %>% addTiles() %>%
setView(lng = -96.808891, lat = 32.779167, zoom = 10) %>%
addCircleMarkers(popup=~n,clusterOptions = markerClusterOptions(),label=crimes$IncDate)
```

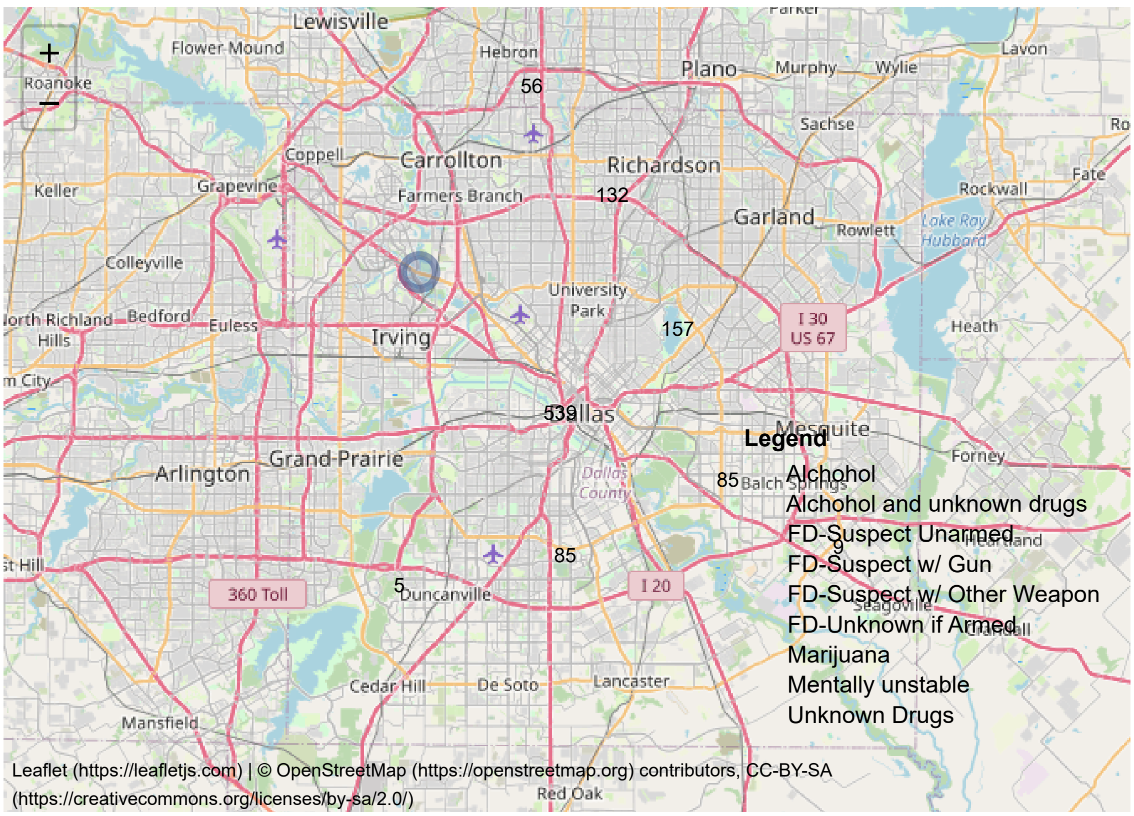


Leaflet (https://leafletjs.com) | © OpenStreetMap (https://openstreetmap.org) contributors, CC-BY-SA (https://creativecommons.org/licenses/by-sa/2.0/)

```
##subject race
crimessubjectrace <- crimedata %>%
  filter(LOCATION_LATITUDE!=''& LOCATION_LONGITUDE!='')%>%
   filter(SUBJECT_RACE == "Black"  | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic") %>%
  group_by(LOCATION_LONGITUDE,LOCATION_LATITUDE,SUBJECT_RACE) %>%
  count() %>%
  arrange(desc(n))

names(crimessubjectrace) <- c("longitude","latitude","race","n")
crimessubjectrace$longitude = as.numeric(crimessubjectrace$longitude)
crimessubjectrace$latitude = as.numeric(crimessubjectrace$latitude)


pal <- colorFactor("viridis", levels = crimessubjectrace$race)

##show subject race against incidents
crimessubjectrace %>% leaflet() %>% addTiles() %>%
setView(lng = -96.808891, lat = 32.779167, zoom = 10) %>%
addCircleMarkers(popup=~n,clusterOptions = markerClusterOptions(),label=crimessubjectrace$race,color = ~pal(race))
%>%
  addLegend(data = crimessubjectrace,
            position = "bottomright",
            pal = pal, values = ~race,
            title = "Legend",
            opacity = 1)
```



```
##show subject description against incident

 crimessubjectdesc <- crimedata %>%
  filter(LOCATION_LATITUDE!=''& LOCATION_LONGITUDE!='')%>%
   filter(SUBJECT_DESCRIPTION != "FD-Motor Vehicle" & SUBJECT_DESCRIPTION !=     "NULL" & SUBJECT_DESCRIPTION != "FD
-Animal" & SUBJECT_DESCRIPTION != "Animal"
         & SUBJECT_DESCRIPTION !="CIT_INFL_A" &
         SUBJECT_DESCRIPTION !="None detected" & SUBJECT_DESCRIPTION !="Unknown")    %>%
  group_by(LOCATION_LONGITUDE,LOCATION_LATITUDE,SUBJECT_DESCRIPTION) %>%
  count() %>%
  arrange(desc(n))

  names(crimessubjectdesc) <- c("longitude","latitude","subjectdesc","n")
crimessubjectdesc$longitude = as.numeric(crimessubjectdesc$longitude)
crimessubjectdesc$latitude = as.numeric(crimessubjectdesc$latitude)

 pal1 <- colorFactor("viridis", levels = crimessubjectdesc$subjectdesc)

 crimessubjectdesc %>% leaflet() %>% addTiles() %>%
setView(lng = -96.808891, lat = 32.779167, zoom = 10) %>%
addCircleMarkers(popup=~n,clusterOptions = markerClusterOptions(),label=crimessubjectdesc$subjectdesc,color = ~pal1
(subjectdesc))  %>%
  addLegend(data = crimessubjectdesc,
           position = "bottomright",
           pal = pal1, values = ~subjectdesc,
           title = "Legend",
           opacity = 1)
```



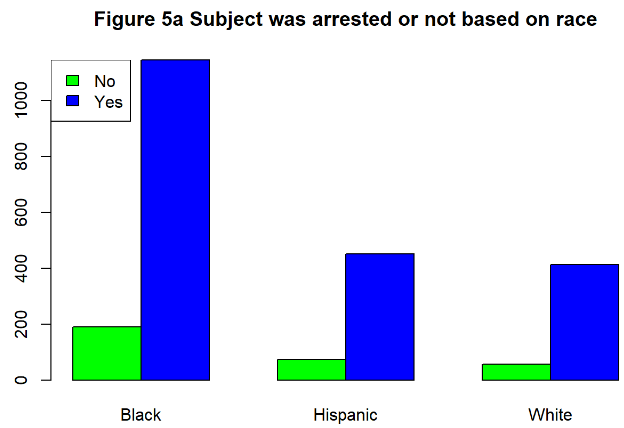
We have introduced some interactive maps showing the crimes that occurred in different locations all over Dallas,Texas.The first map shows the incident date against each location and the second and third map shows the race and description of the subject involved in the crime respectively.

Subject Arrested or not

```
#subject arrested or not for three races black,white and hispanic
  crimedata_subjectarrest <-  crimedata %>%
  filter(SUBJECT_RACE == "Black"  | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic")

datatable <- table(crimedata_subjectarrest$SUBJECT_WAS_ARRESTED,crimedata_subjectarrest$SUBJECT_RACE)
knitr::kable(datatable,format = "html",longtable = TRUE)
```

	Black	Hispanic	White
No	189	73	57
Yes	1144	451	413

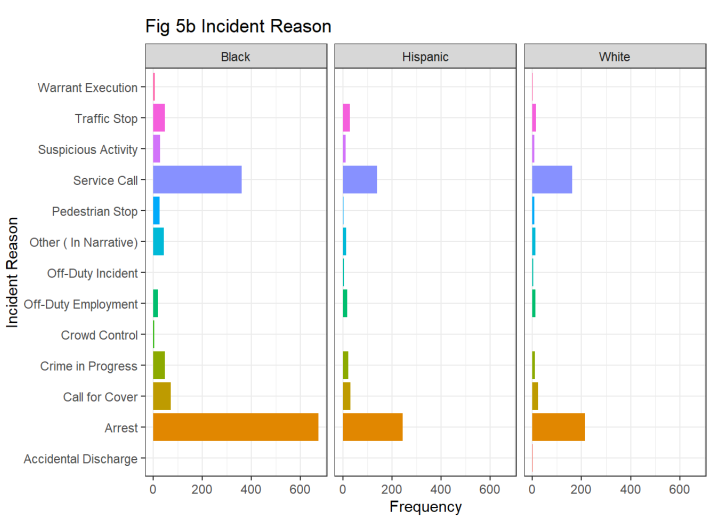
```
#bar plot for arrests made classified on race basis
barplot(datatable, beside = TRUE, main="Figure 5a Subject was arrested or not based on race",col=c("green","blue"))
legend("topleft",c("No","Yes"),fill = c("green","blue"))
```

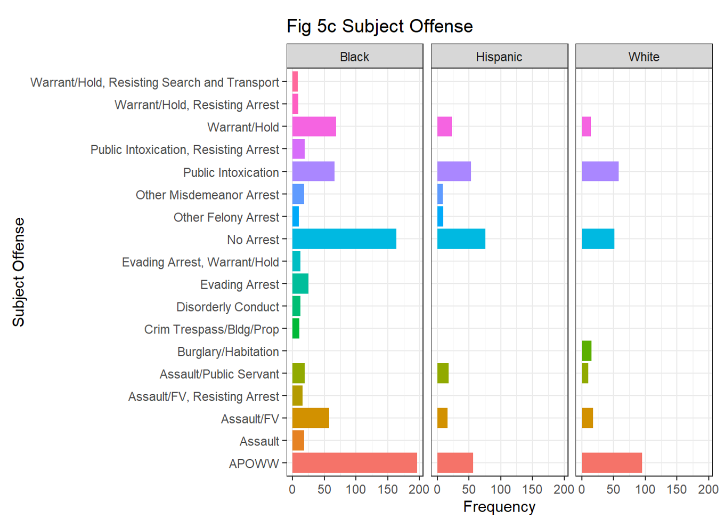


We observe from the bar plot that the black subjects have the highest number of arrests but this is expected given they have the highest crime rates also. The Hispanics come second and the white subjects have the lowest number of arrests.

Incident Reason and subject offense

```
#Incident Reason
crimedata_incidentreason <- crimedata %>%
  filter(INCIDENT_REASON!="NULL") %>%
  filter(SUBJECT_RACE == "Black"  | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic") %>%
  group_by(SUBJECT_RACE,INCIDENT_REASON) %>%
  count() %>%
  arrange(desc(n))

# # Create bar chart with ggplot2 force used on crime subjects
crimedata_incidentreason %>%
ggplot(aes(x = INCIDENT_REASON,y=n,fill=INCIDENT_REASON)) +
geom_bar(stat = "identity",show.legend = FALSE) +
xlab("Incident Reason")  +
ylab("Frequency")+
ggtitle("Fig 5b Incident Reason")+
coord_flip()  +
facet_wrap(~SUBJECT_RACE)+
theme_bw()

#Subject Offense

crimedata_subjectoffense <- crimedata %>%
  filter(SUBJECT_RACE == "Black"  | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic") %>%
  group_by(SUBJECT_RACE,SUBJECT_OFFENSE) %>%
  count() %>%
  arrange(desc(n))

## Create bar chart with ggplot2 force used on crime subjects
crimedata_subjectoffense %>%
  filter(n>7) %>%
ggplot(aes(x = SUBJECT_OFFENSE,y=n,fill=SUBJECT_OFFENSE)) +
geom_bar(stat = "identity",show.legend = FALSE) +
xlab("Subject Offense")  +
ylab("Frequency")+
ggtitle("Fig 5c Subject Offense")+
coord_flip()  +
facet_wrap(~SUBJECT_RACE)+
theme_bw()
```

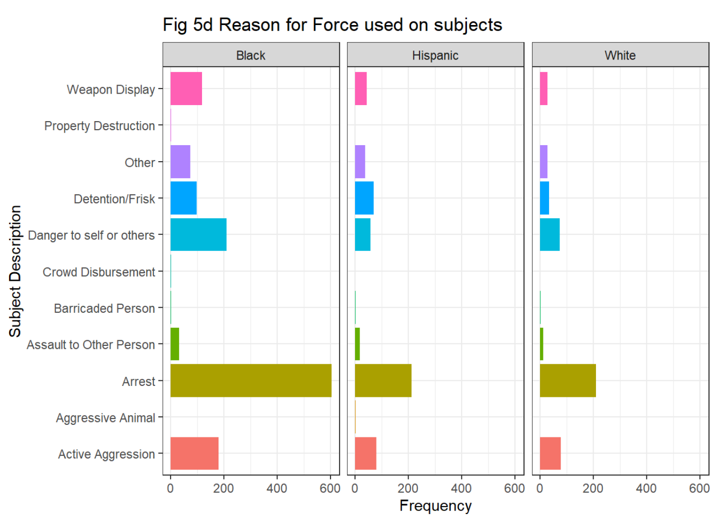


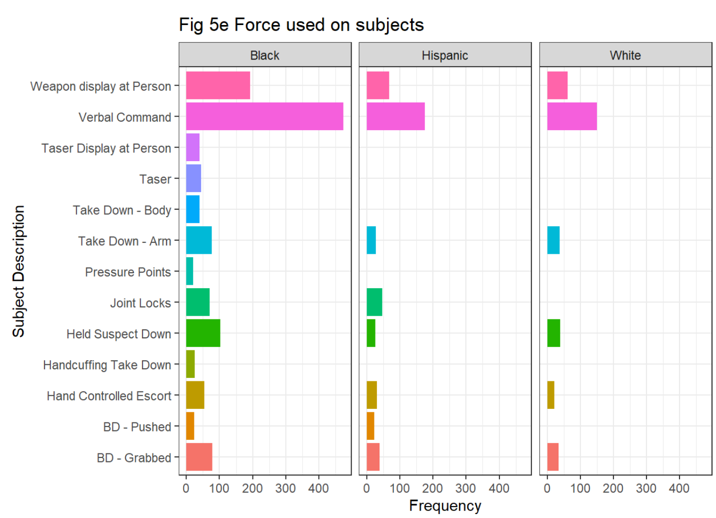


The next step is to look at the subject offences and incident reasons for the various races. For all three races, the most common incident reasons are arrest and service call(Figure 5b). However, compared to Hispanics and Whites, the numbers for black subjects are higher. Black folks are also described as having more calls for cover, crimes in progress, and suspicious activity than the other two races combined. Even traffic stops are more common for people of color.Therefore, even though we see that all three races report the identical incident causes, black people consistently have the largest numbers. In Texas, a warrant exits, known as an APOWW—apprehension by a police officer without a warrant—which can be completed by any law enforcement official. In terms of offenses recorded against each subject race, APOWW has the highest number of incidences. Other common crimes include various assaults, public intoxication, and crimes involving warrants.As expected the numbers are highest for people of color.(Figure 5c)

Reason for force and type of force used on the subjects

```
##Reason for force

crimesubject_reasonforforce <- crimedata %>%
  filter(SUBJECT_RACE == "Black"  | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic") %>%
  filter(REASON_FOR_FORCE!="NULL") %>%
  group_by(SUBJECT_RACE,REASON_FOR_FORCE) %>%
  count() %>%
  arrange(desc(n))

## Create bar chart with ggplot2 reason for force used on crime subjects
crimesubject_reasonforforce %>%
ggplot(aes(x = REASON_FOR_FORCE,y=n,fill=REASON_FOR_FORCE)) +
geom_bar(stat = "identity",show.legend = FALSE) +
xlab("Subject Description")  +
ylab("Frequency")+
ggtitle("Fig 5d Reason for Force used on subjects")+
coord_flip()  +
theme_bw()+
facet_wrap(~SUBJECT_RACE)

crimesubjectforceused <- crimedata %>%
  filter(SUBJECT_RACE == "Black"  | SUBJECT_RACE == "White" | SUBJECT_RACE == "Hispanic") %>%
  group_by(SUBJECT_RACE,TYPE_OF_FORCE_USED1) %>%
  count() %>%
  arrange(desc(n))

## Create bar chart with ggplot2 force used on crime subjects
crimesubjectforceused %>%
  filter(n>20) %>%
ggplot(aes(x = TYPE_OF_FORCE_USED1,y=n,fill=TYPE_OF_FORCE_USED1)) +
geom_bar(stat = "identity",show.legend = FALSE) +
xlab("Subject Description")  +
ylab("Frequency")+
ggtitle("Fig 5e Force used on subjects")+
coord_flip()  +
theme_bw()+
facet_wrap(~SUBJECT_RACE)
```

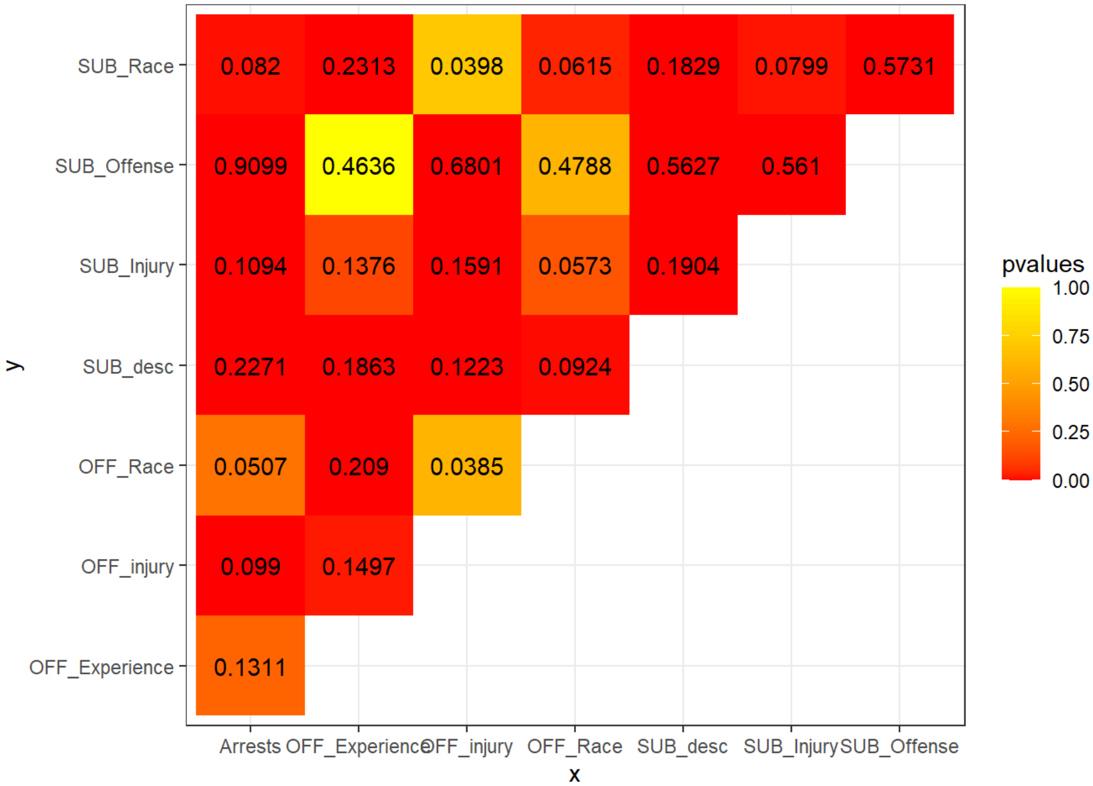




This section focuses on the reasons given for and types of forces employed by the police personnel against the subjects. This has been compared for all three races. The three main justifications for using force are,it was done during an arrest, when the individual was actively aggressive and if they are a threat to both themselves and other people.The suspects being held, being frisked, or showing signs of having a weapon are other typical grounds.(Figure 5d) Next, we investigate the types of force employed by the officers against the suspects. For all three races, verbal command has the highest percentage, showing that cops typically try to address problems or persons verbally without engaging in physical conflict. Displaying weapons comes in second on the list, then utilizing physical force, such as holding the suspect down, taking them down by the arm, locking their joints, etc.(Figure 5e)

Correlation analysis

```
library(tidyverse)
library(lsr)

crimecorrelation = select(crimedata,OFFICER_RACE,OFFICER_YEARS_ON_FORCE,OFFICER_INJURY,SUBJECT_RACE,SUBJECT_INJURY,SUBJECT_WAS_ARRESTED,SUBJECT_DESCRIPTION,SUBJECT_OFFENSE)

names(crimecorrelation) = c("OFF_Race","OFF_Experience","OFF_injury","SUB_Race","SUB_Injury","Arrests","SUB_desc","SUB_Offense")
# function to get chi square p value and Cramers V
f = function(x,y) {
    tbl = crimecorrelation %>% select(x,y) %>% table()
    pvalues = round(chisq.test(tbl)$p.value, 4)
    cramV = round(cramersV(tbl), 4)
    data.frame(x, y, pvalues, cramV) }

# create unique combinations of column names
# sorting will help getting a better plot (upper triangular)
df_comb = data.frame(t(combn(sort(names(crimecorrelation)), 2)), stringsAsFactors = F)

# apply function to each variable combination
df_res = map2_df(df_comb$X1, df_comb$X2, f)

# plot results
df_res %>%
  ggplot(aes(x,y,fill=pvalues))+
  geom_tile()+
  geom_text(aes(x,y,label=cramV))+
  scale_fill_gradient(low="red", high="yellow")+
  # theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))+
  #scale_x_discrete(guide = guide_axis(n.dodge=3))+
  theme_bw() #Reference [3]
```



Finally a correlation analysis has been done for the variables OFFICER_RACE,OFFICER_YEARS_ON_FORCE,OFFICER_INJURY,SUBJECT_RACE,SUBJECT_INJURY,SUBJECT_WAS_ARRESTED, SUBJECT_DESCRIPTION,SUBJECT_OFFENSE.All the variables are put through the chi-square test to see if they are independent of one another, and the p values are gathered. The variables are then subjected to Cramér's V, which gauges the degree to which two category fields are related. Therefore, we can determine whether the variables are related based on the color gradient utilized for the chi square values. We can determine the strength of the association between the two variables by looking at the values for Cramers V.The values ranges from 0 to 1.
The subject was arrested or not and the subject offense have a high cramers v value of 0.9 indicating a strong relationship. Second highest is the relation between Subject Offense and Officer Injury.The lowest values are between Officer injury and officer race and subject race.

Discussions and Conclusion

The results can be summed up as follows: the crime rate was highest in the first three months of the year and lowest in the last month of December. The central division of Dallas had the greatest crime rate, followed by the southeast and northeast divisions, when we looked at the frequency of crimes across the city's seven districts. The majority of the individuals were classified as having unstable minds or being under the influence of drugs and alcohol.
White, Hispanic, and Black people made up the top three racial groups in crime. In contrast, White officers made up the majority of the force, followed by Hispanic and Black cops. Therefore, it appears that the majority of crimes committed by black individuals or incidents reported against people of color are handled by white officers. For black folks, there were a lot more arrests. We looked into the incident cause and the subject offence to see if the reported reasons or offences would differ for the three races, but this was not the case. Arrests and service calls are the most frequent incident causes for all three races. The numbers for black subjects are higher than those for Hispanics and Whites. In comparison to the other two races combined, black people are said to have more calls for cover, crimes in progress, and suspicious activities. People of color are more likely to experience traffic stops overall.Black people continually make up the biggest percentage, despite the fact that all three races describe the same incidence causes.The same pattern of frequent crimes being reported in all three instances, with colored persons once again having the highest numbers, is shown when it comes to offences.We cannot draw a clear inference from the available statistics that this is because of the high number of offences reported against them or because of a racial disparity in the system.
The reason for using force and the type of force used by the officers were found to share a overall similarity between all three races .For instance, if the person displayed hostility, was arrested, or represented a threat to himself or others. Likewise if the police desired to detain or frisk a person,then force was used.For offences they handled, the police are always shown to utilize verbal orders more frequently than displaying weapons, regardless of the race. Then, there are additional physical tackles that happen frequently.

In conclusion, we may argue that even if there is a racial imbalance between the subjects and the officers, there isn't enough concrete evidence in the police and crime data to demonstrate that racism played a role in how the various reported crimes were handled. For all three primary racial groups involved in crimes, the majority of recorded events, subject offences, and force used on subjects are identical , with black individuals having the highest numbers.For instance, we observe that the officers primarily employ verbal commands to restrain the suspect.However, given that it may mean that more crimes are being reported or complaints are being made against black people, the high number of crimes against them that have been recorded may be cause for alarm. For instance, more arrests and convictions for them than, say, white individuals were the consequence of traffic stops, suspicious activity reported more frequently for them, or any other documented events.

Further examination of these concerns and investigation of more data may shed more light and provide conclusive evidence to the issues of racial disparities in Dallas,Texas.

References

1.“Data Science for Good: Center for Policing Equity,” Kaggle, Oct. 29, 2018. https://www.kaggle.com/datasets/center-for-policing-equity/data-science-for-good (https://www.kaggle.com/datasets/center-for-policing-equity/data-science-for-good)

2.”EDA - Time Series Analysis - Policing equity,” Kaggle, Dec. 16, 2018. https://www.kaggle.com/code/vincentkr18/eda-time-series-analysis-policing-equity/notebook (https://www.kaggle.com/code/vincentkr18/eda-time-series-analysis-policing-equity/notebook)

3.”Plot the equivalent of correlation matrix for factors (categorical data)? And mixed types?“,Stack Overflow. https://stackoverflow.com/questions/52554336/plot-the-equivalent-of-correlation-matrix-for-factors-categorical-data-and-mi (https://stackoverflow.com/questions/52554336/plot-the-equivalent-of-correlation-matrix-for-factors-categorical-data-and-mi)