

CPE Analysis

2023-03-07

The Center for Policing Equity (CPE)

The Center for Policing Equity (CPE) is a research and policy organization that aims to reduce bias and improve policing practices through data-driven methods. The organization was founded in 2007 by Dr. Phillip Atiba Goff, who is a recognized expert in the field of racial bias in law enforcement.

CPE works with police departments across the United States to collect and analyze data on police interactions with communities of color. The organization's goal is to use this data to identify and address racial disparities in policing, as well as to develop evidence-based solutions for improving police-community relations.

Import policing data set of Dallas, Texas in 2016

```
data <- read.csv("37-00049_UOF-P_2016_prepped.csv")
data = data[-1,]
dim(data)
```

```
## [1] 2383  47
```

The data has 2383 rows and 47 columns about police work in Dallas, Texas in 2016.

There are four types of data found in data set.

Incident Information

- INCIDENT_DATE
- INCIDENT_TIME
- UOF_NUMBER
- INCIDENT_REASON
- REASON_FOR_FORCE
- TYPE_OF_FORCE_USED1
- TYPE_OF_FORCE_USED2
- TYPE_OF_FORCE_USED3
- TYPE_OF_FORCE_USED4
- TYPE_OF_FORCE_USED5
- TYPE_OF_FORCE_USED6
- TYPE_OF_FORCE_USED7
- TYPE_OF_FORCE_USED8
- TYPE_OF_FORCE_USED9
- TYPE_OF_FORCE_USED10
- NUMBER_EC_CYCLES
- FORCE_EFFECTIVE

Officer Information

- OFFICER_ID
- OFFICER_GENDER
- OFFICER_RACE
- OFFICER_HIRE_DATE
- OFFICER_YEARS_ON_FORCE
- OFFICER_INJURY
- OFFICER_INJURY_TYPE
- OFFICER_HOSPITALIZATION

Subject Information

- SUBJECT_ID
- SUBJECT_RACE
- SUBJECT_GENDER
- SUBJECT_INJURY
- SUBJECT_INJURY_TYPE
- SUBJECT_WAS_ARRESTED
- SUBJECT_DESCRIPTION
- SUBJECT_OFFENSE

Location Information

- REPORTING_AREA
- BEAT
- SECTOR
- DIVISION
- LOCATION_DISTRICT
- STREET_NUMBER
- STREET_NAME
- STREET_DIRECTION
- STREET_TYPE
- LOCATION_FULL_STREET_ADDRESS_OR_INTERSECTION
- LOCATION_CITY
- LOCATION_STATE
- LOCATION_LATITUDE
- LOCATION_LONGITUDE
- STREET_TYPE

pre-processing

```
date = as.Date(data$INCIDENT_DATE, format = "%m/%d/%y")
data$weekdays = weekdays(date, abbreviate = TRUE)
time_24h <- format(strptime(data$INCIDENT_TIME, "%I:%M:%S %p"), "%H:%M:%S")
data$hours <- as.numeric(substr(time_24h, 1, 2))
data$month_name <- format(date, "%b")
```

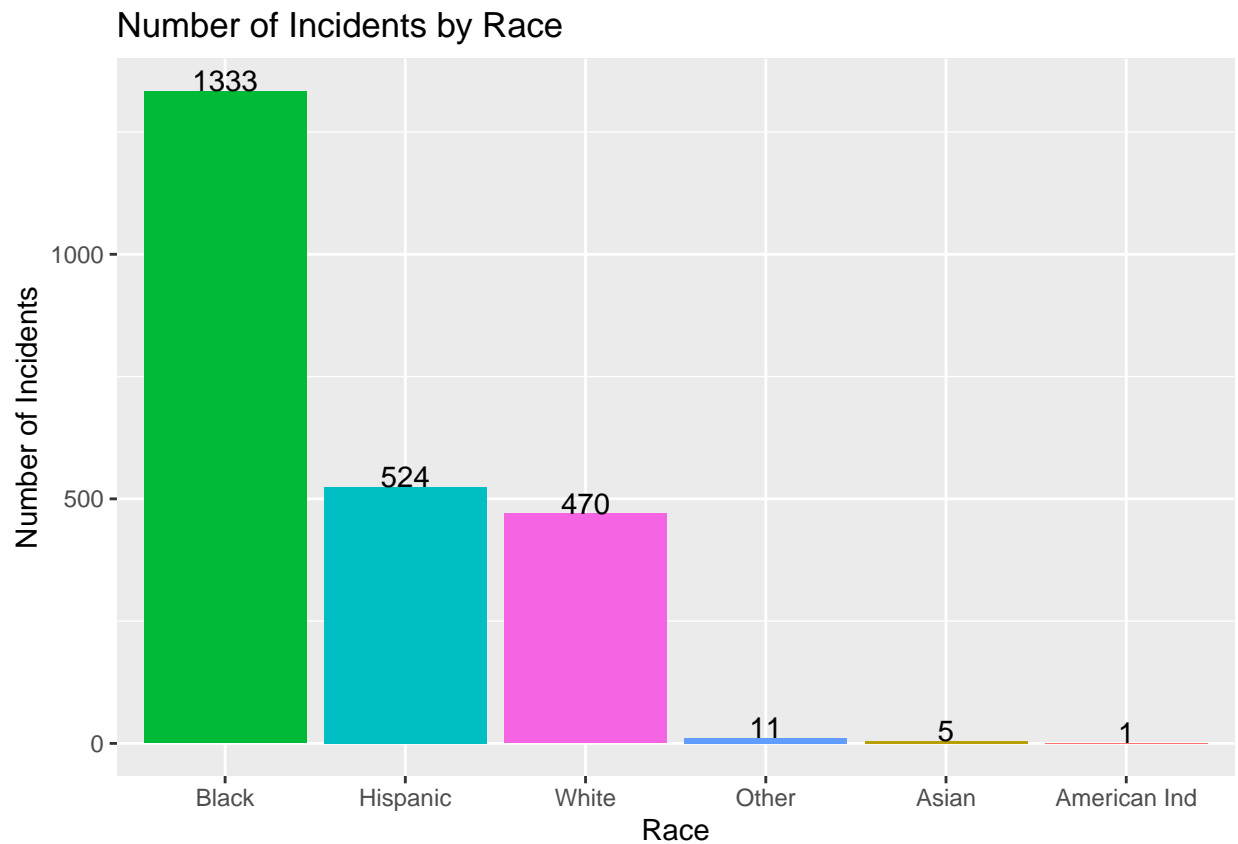
dataframe contains columns INCIDENT_DATE and INCIDENT_TIME. It converts the values in these columns to Date and POSIXct objects respectively and extracts the weekdays, hours and month names from

these objects. These extracted values are stored in new columns in the data frame called weekdays, hours and month_name

Analysis

Number of Incidents by Race

```
library(ggplot2)
incidents_by_race <- data.frame( table(data$SUBJECT_RACE))
names(incidents_by_race) <- c("Race", "Number of Incidents")
# Removing NULL Race from incidents_by_race data frame
incidents_by_race = incidents_by_race[-5,]
ggplot(data=incidents_by_race, aes(x=reorder( Race, -`Number of Incidents`), y=`Number of Incidents`, fill=`Number of Incidents`)) +
  geom_bar(stat="identity") +
  geom_text(aes(label=`Number of Incidents`), vjust=0) +
  labs(x="Race", y="Number of Incidents", title="Number of Incidents by Race") +
  theme(legend.position = "none")
```



This indicates that Black individuals were involved in the highest number of incidents of police use of force, which is concerning and indicative of potential racial bias in policing practices. Further analysis is needed to understand why this disparity exists.

Top 5 Frequency of subject Offences/crimes

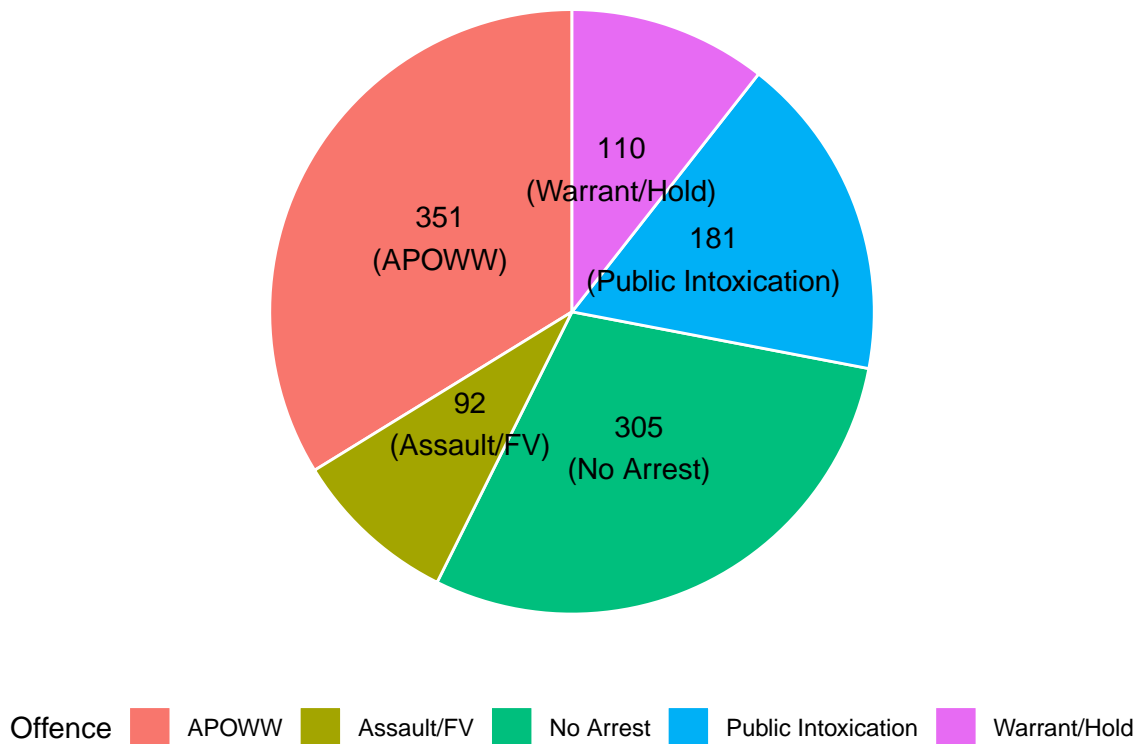
```
subject_offence_frequency = data.frame (table(data$SUBJECT_OFFENSE))
# Sort the data frame by the 'SUBJECT_OFFENSE' column in descending order
pie_data = subject_offence_frequency[order(-subject_offence_frequency$Freq), ]
names(pie_data) <- c("Offence", "Frequency")

# get top 5
pie_data = pie_data[1:5,]

# Calculate the total frequency
total <- sum(pie_data$Frequency )

# Create the pie chart
ggplot(pie_data, aes(x="", y=Frequency, fill=Offence)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) +
  theme_void() +
  theme(legend.position = "bottom") +
  geom_text(aes(label = paste0( Frequency,"\n(" ,Offence, ")" )), position = position_stack(vjust = 0.5))
labs(title = "Top 5 Frequency of subject Offences/crimes", fill = "Offence", x = NULL, y = NULL)
```

Top 5 Frequency of subject Offences/crimes



As we can see, APOWW has more offenses than any other. APOWW stands for Apprehension by Peace Officer Without Warrant. It refers to a situation where a peace officer apprehends someone without a warrant. These top 5 offenses cover 1039 out of a total of 2384 offenses.

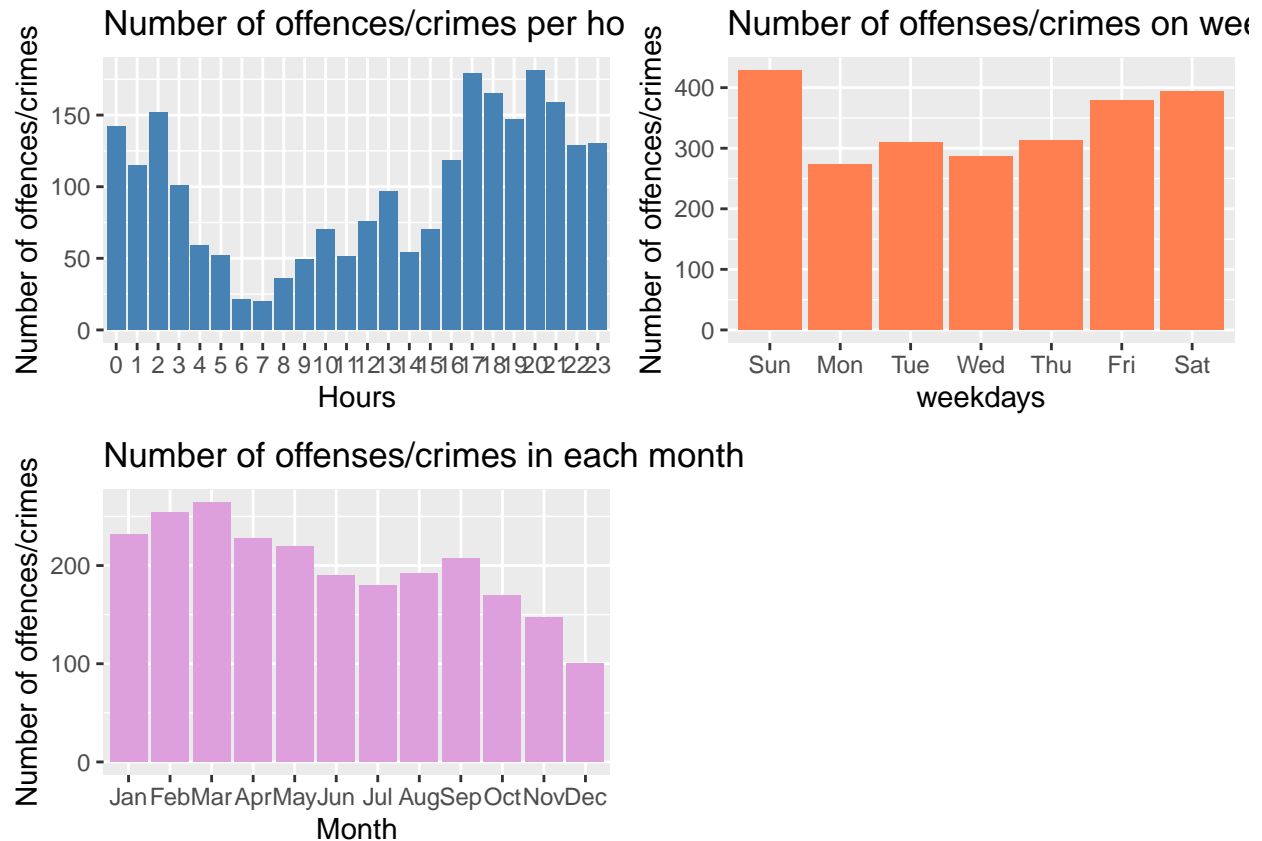
Number of Offences/crimes by Hours, weekdays and months

```
offences_by_hour_frequency = data.frame( table(data$hours))
colnames(offences_by_hour_frequency) = c("hours", "Freq")
hour_plot <- ggplot(data=offences_by_hour_frequency, aes(x=hours, y=Freq)) + geom_bar(stat="identity", fill="plum") +
  xlab("Hours") +
  ylab("Number of offences/crimes") +
  ggtitle("Number of offences/crimes per hour")

offences_by_weekdays_frequency = data.frame(table(data$weekdays))
colnames(offences_by_weekdays_frequency) = c("weekdays", "Freq")
offences_by_weekdays_frequency$weekdays <- factor(offences_by_weekdays_frequency$weekdays,
  levels = c("Sun", "Mon", "Tue", "Wed", "Thu", "Fri", "Sat"))
weekday_plot <- ggplot(data=offences_by_weekdays_frequency, aes(x=weekdays, y=Freq)) + geom_bar(stat="identity", fill="plum") +
  xlab("weekdays") +
  ylab("Number of offences/crimes") +
  ggtitle("Number of offenses/crimes on weekdays")

offences_on_each_month = data.frame(table(data$month_name))
colnames(offences_on_each_month) = c("month", "Freq")
offences_on_each_month$month <- factor(offences_on_each_month$month,
  levels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))
each_month_plot <- ggplot(data=offences_on_each_month, aes(x=month, y=Freq)) +
  geom_bar(stat="identity", fill="plum") +
  xlab("Month") +
  ylab("Number of offences/crimes") +
  ggtitle("Number of offences/crimes in each month")
library(gridExtra)

grid.arrange(hour_plot, weekday_plot, each_month_plot, ncol = 2, nrow = 2)
```



We can see that most crimes are committed at night and in the evening. There is a peak in crime during the weekend, particularly on Sundays and Saturdays. This could be because more people go out for outings during the weekend, which may cause a rise in the number of crimes.

Number of crimes/offences per month in each division

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##   combine

## The following objects are masked from 'package:stats':
##
##   filter, lag

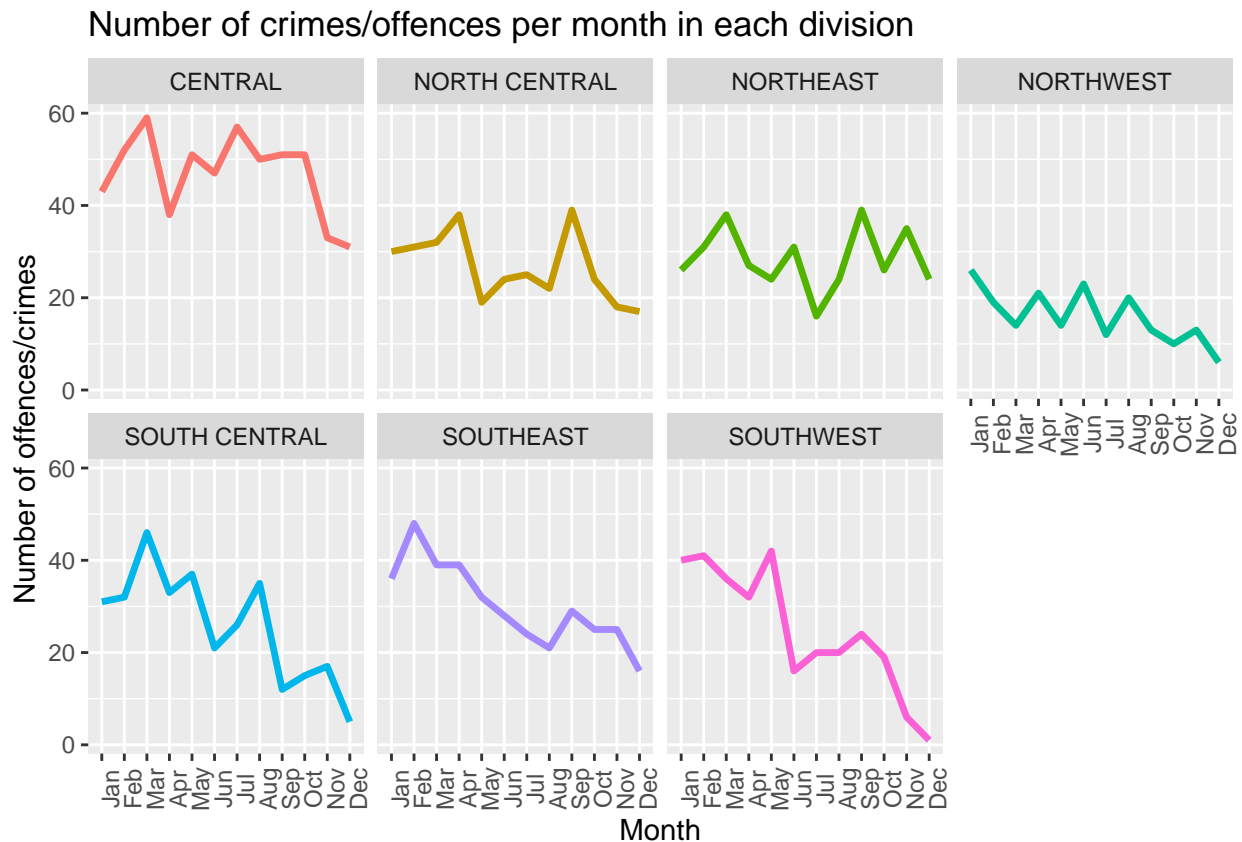
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
df_data <- data %>%
  count(month_name, DIVISION, name = "freq")

df_data = df_data[!is.na(df_data$month_name), ]
df_data$month_name <- factor(df_data$month_name, levels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

ggplot(df_data, aes(x = month_name, y = freq, color = DIVISION, group = DIVISION)) +
  geom_line(size = 1.2) +
  facet_wrap(~ DIVISION, ncol = 4) +
  labs(x = "Month") +
  labs(y = "Number of offences/crimes") +
  labs(title = "Number of crimes/offences per month in each division") +
  theme(legend.position = "none", axis.text.x = element_text(angle = 90))

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```

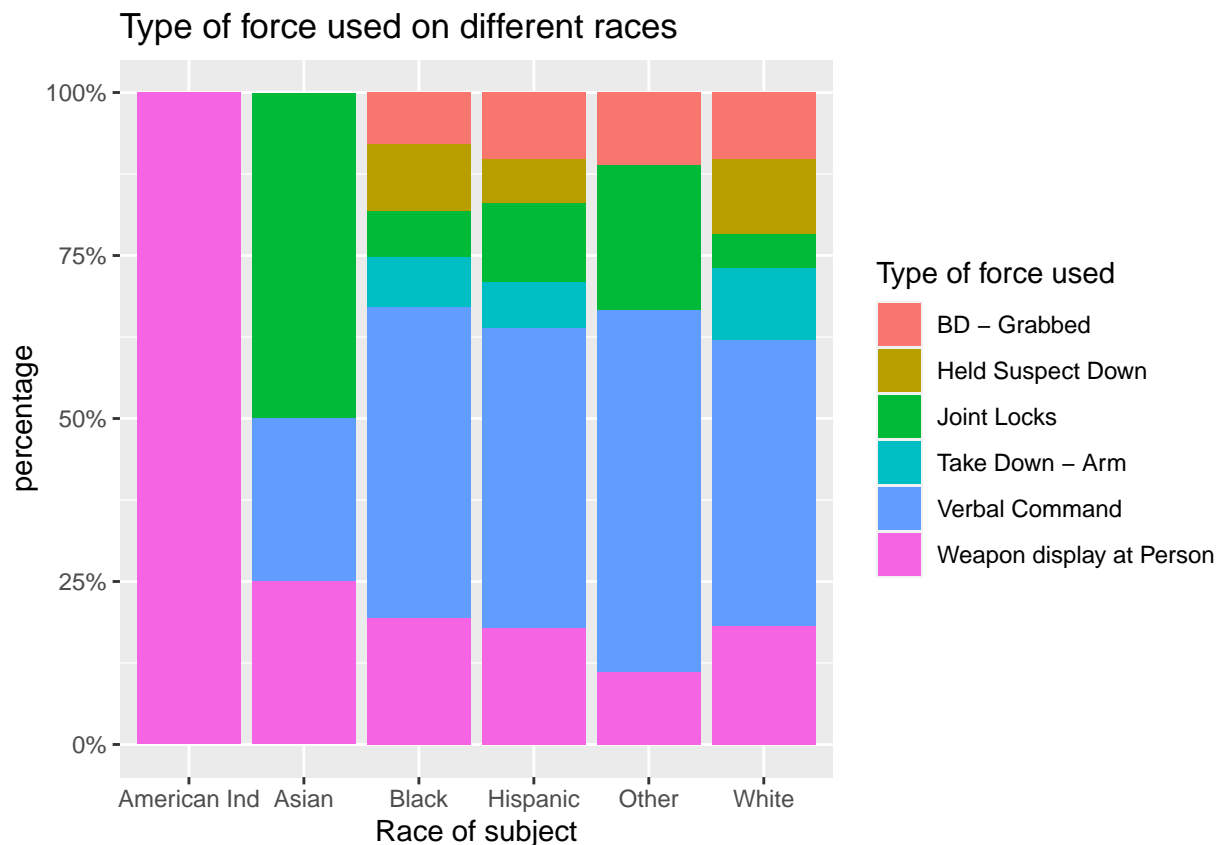


The North Central division has consistently high numbers of crimes throughout the year, while the South Central and Southeast divisions have the lowest numbers. The Central division shows a spike in crime in February and again in October, while the other divisions show more sporadic patterns. Overall, the plot suggests that crime rates vary widely across divisions and over time.

Top 6 Forces used on different races

```
type_of_force_df = data.frame(table(data$TYPE_OF_FORCE_USED1))
type_of_force_df = type_of_force_df[order(-type_of_force_df$Freq), ]
colnames(type_of_force_df) = c("Force_type", "Freq")
type_of_force_df = type_of_force_df[1:6,]

forced_on_races_dataframe = subset(data, SUBJECT_RACE!="NULL" & (TYPE_OF_FORCE_USED1 %in% type_of_force_df$Force_type))
ggplot(forced_on_races_dataframe, aes(x = SUBJECT_RACE, y=100, fill = TYPE_OF_FORCE_USED1)) +
  geom_bar(stat = "identity", position = "fill") +
  scale_y_continuous(labels = scales::percent) +
  xlab("Race of subject") +
  ylab("percentage") +
  guides(fill=guide_legend(title="Type of force used")) +
  ggtitle("Type of force used on different races")
```

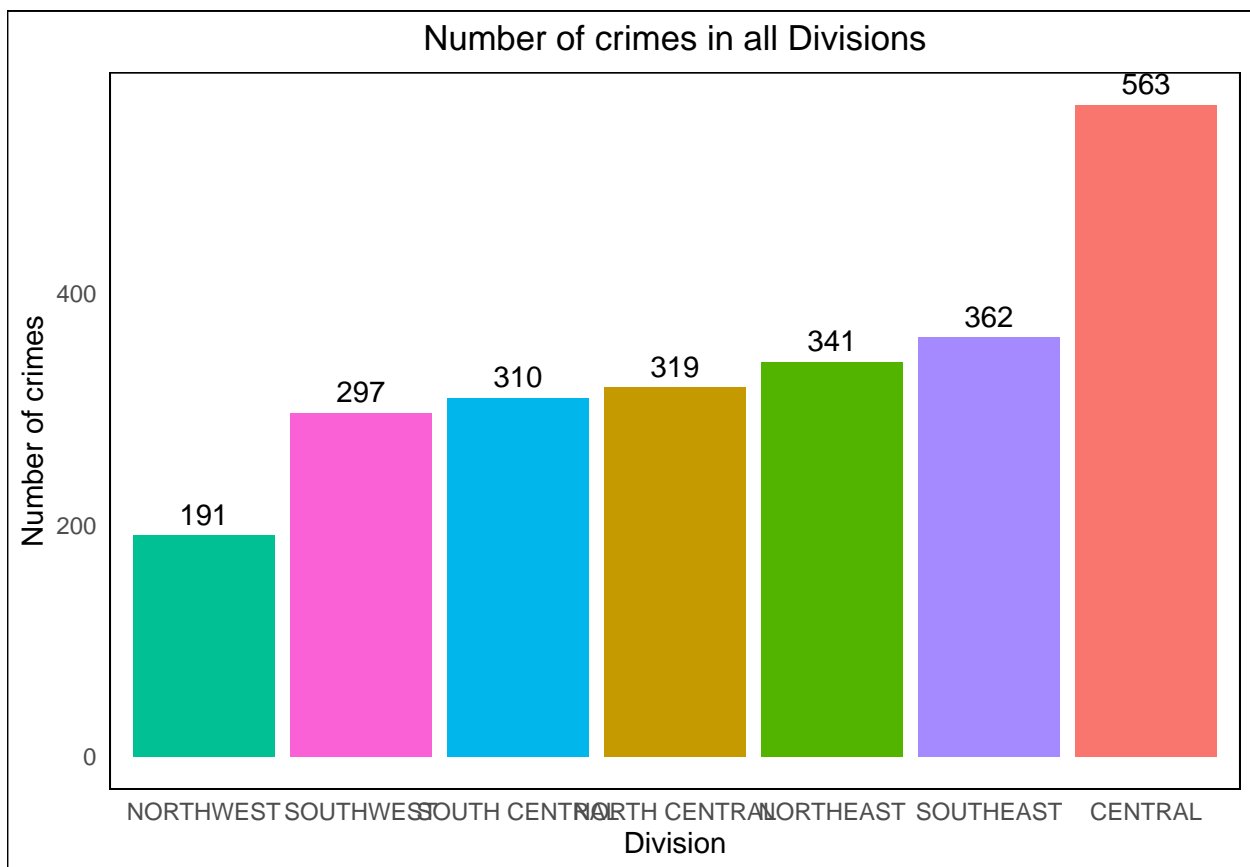


According to records, weapon display is the most commonly used force against individuals of American Indian race because it is the only offense recorded for that race. Verbal commands are more frequently used on Black, Hispanic, and other races as well as white individuals. The second most common use of force is weapon display at person.

Number of crimes in Divisions

```
division_freq = data.frame( table(data$DIVISION))
colnames(division_freq) = c("Division", "Freq")

ggplot(data = division_freq, aes(x =reorder( Division , Freq), y = Freq, fill = Division)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = Freq), vjust = -0.5, position = position_dodge(width = 0.9)) +
  labs(title = "Number of crimes in all Divisions", x = "Division", y = "Number of crimes") +
  theme_minimal() +
  theme(
    legend.position = "none",
    panel.border = element_rect(color = "black", fill = NA),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    plot.title = element_text(hjust = 0.5),
    plot.background = element_rect(fill = "white")
  )
```

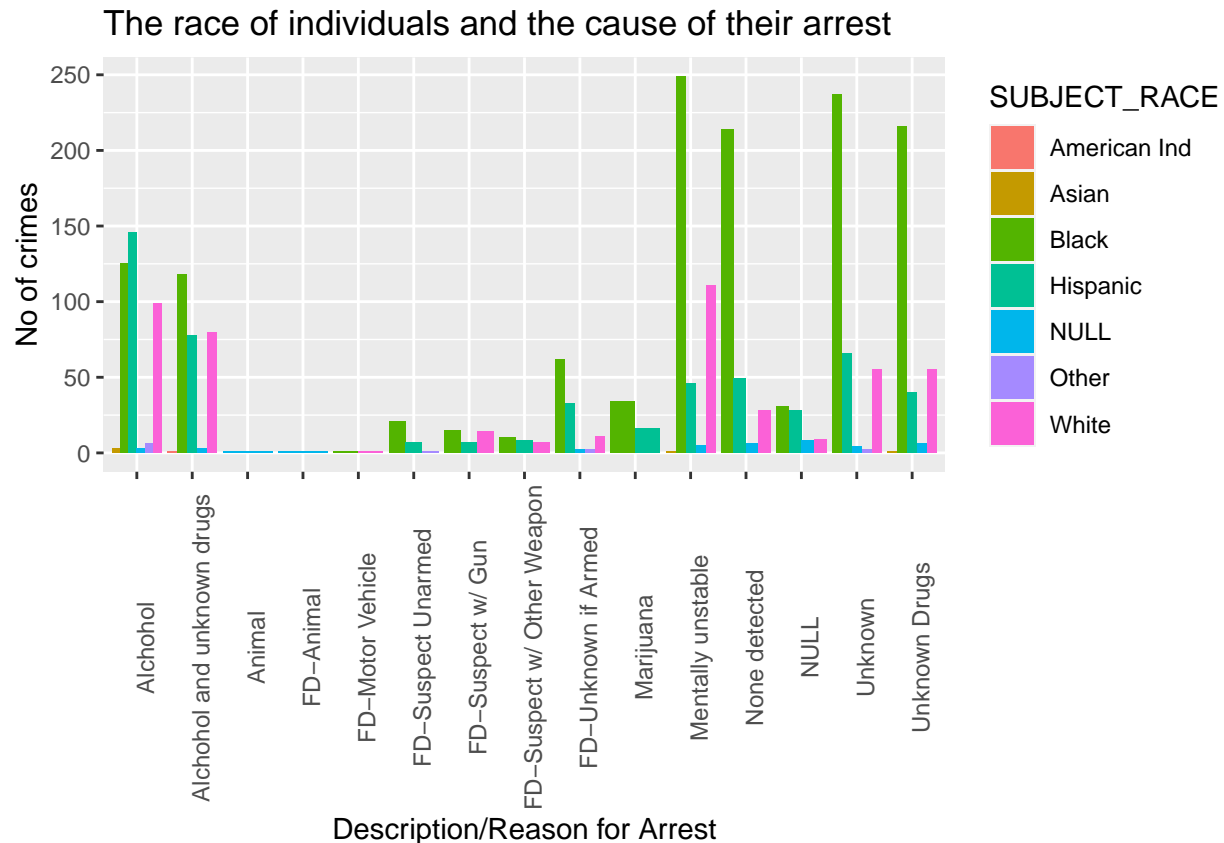


We can see that the highest frequency is in the Central Division with a count of 563, followed by Southeast Division with a count of 362. The lowest frequency is in the Northwest Division with a count of 191. The chart shows that the Central and Southeast Divisions require more attention, as they have a higher frequency compared to other divisions.

The race of individuals and the cause of their arrest

```
library(dplyr)
data_count <- count(data, SUBJECT_DESCRIPTION , SUBJECT_RACE)

ggplot(data_count, aes(x = SUBJECT_DESCRIPTION, y = n, fill = SUBJECT_RACE )) +
  geom_bar(position = "dodge", stat = "identity") +
  labs(title="The race of individuals and the cause of their arrest", x="Description/Reason for Arrest")
  theme(axis.text.x = element_text(angle = 90))
```



We can see that the most common subject description in subject Crimes is “Mentally unstable,” with a total of 411 fatal crimes. Out of these, the majority of the crimes were of Black subjects (249 crimes), followed by White subjects (111 crimes), and Hispanic subjects (46 crimes). The second most common subject description in crimes is “None detected.” The third most common subject description in crimes is “Unknown,” with a total of 364. From this analysis, we can observe that crimes in the United States are disproportionately concentrated among Black subjects. Additionally, subjects described as “Mentally unstable,” “None detected,” and “Unknown” account for a significant proportion of crimes.

```
data$month_numbers <- match(data$month_name, month.abb, nomatch = 0)
df <- data %>%
  filter(SUBJECT_RACE == "Hispanic" | SUBJECT_RACE == "Black" | SUBJECT_RACE == "White") %>%
```

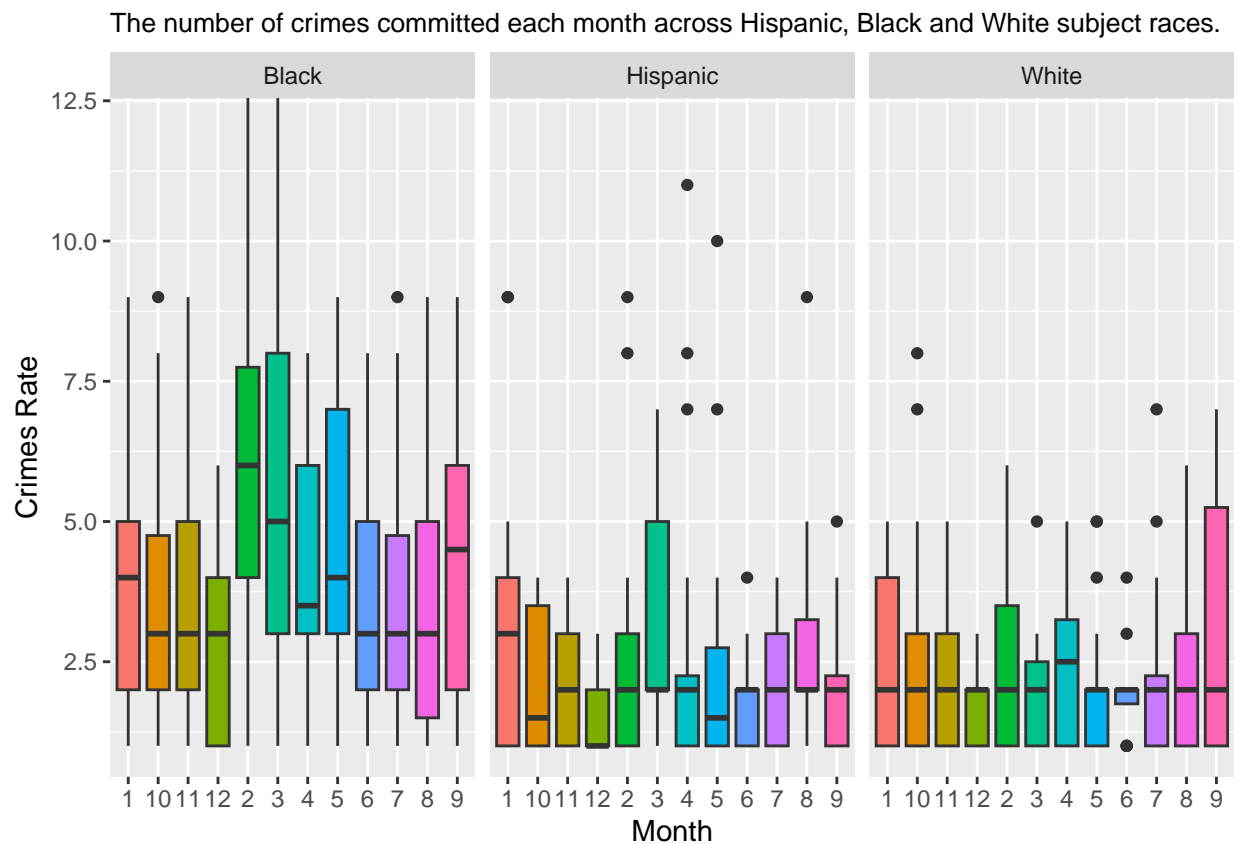
```

group_by(INCIDENT_DATE, month_numbers, SUBJECT_RACE) %>%
  summarize(avg = n(), .groups = "drop")
df <- subset(df, month_numbers != 0 )
df$INCIDENT_DATE = as.Date(df$INCIDENT_DATE)
df$month_numbers = as.character(df$month_numbers)
df$SUBJECT_RACE = as.factor(df$SUBJECT_RACE)
df$avg = as.numeric(df$avg)

ggplot(df , aes(x = (month_numbers), y= avg, fill = month_numbers)) +

  geom_boxplot() +
  labs(x= 'Month') +
  labs(y = 'Crimes Rate') +
  labs(title = paste("The number of crimes committed each month across Hispanic, Black and White su
  theme(plot.title = element_text(size = 10)) +
  theme(legend.position="none") + facet_wrap(~SUBJECT_RACE) +
  coord_cartesian(ylim = c(1, 12))

```



According to the data presented, it appears that the number of crimes committed by Black people is highest in March and May, with lower numbers in July and October. For Hispanics, the highest number of crimes is also in March, followed by January, with the lowest numbers in June and October. For White people, the highest number of crimes is in September, with the lowest number in June.

Conslusion

The dataset shows concerning disparities in police use of force among different races, with Black individuals involved in the highest number of incidents. APOWW is the most frequent offense recorded. Most crimes occur at night and on weekends. The Central and Southeast Divisions require more attention. Crimes are disproportionately concentrated among Black subjects, and those described as “Mentally unstable,” “None detected,” and “Unknown” account for a significant proportion of crimes.

References

R Graph Gallery. (n.d.). R Graph Gallery. <https://r-graph-gallery.com/>

STHDA. (n.d.). ggplot2 barplots: Quick start guide - R software and data visualization. Retrieved March 24, 2023, from <http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization>