Name: Jayraj Menon

UID: 2021300075

ADV Experiment 4

Aim :

Create basic charts using R programming language on dataset Crime or Police / Law and Order

● Basic - Bar chart, Pie chart, Histogram, Time line chart, Scatter plot, Bubble plot

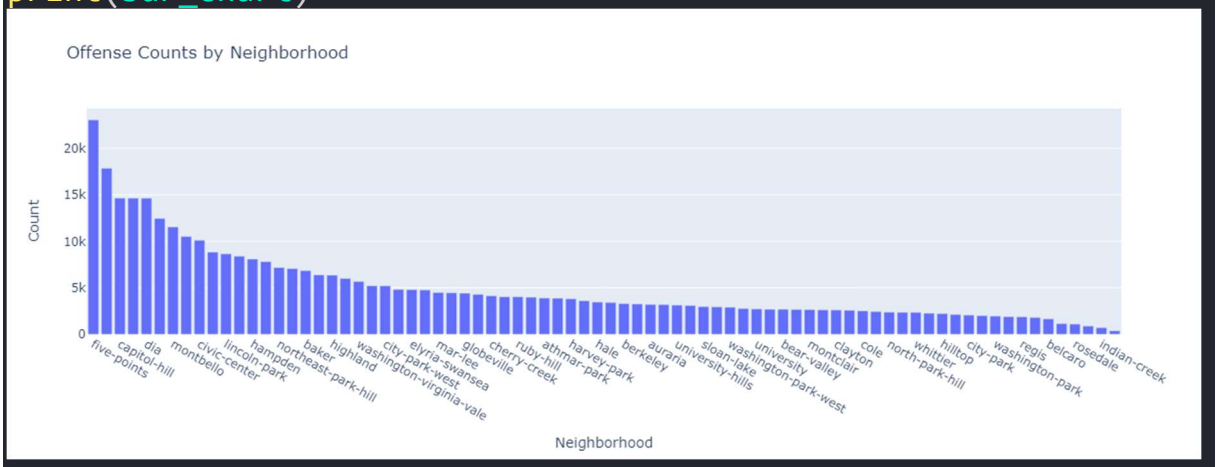
● Write observations from each chart

DATASET:

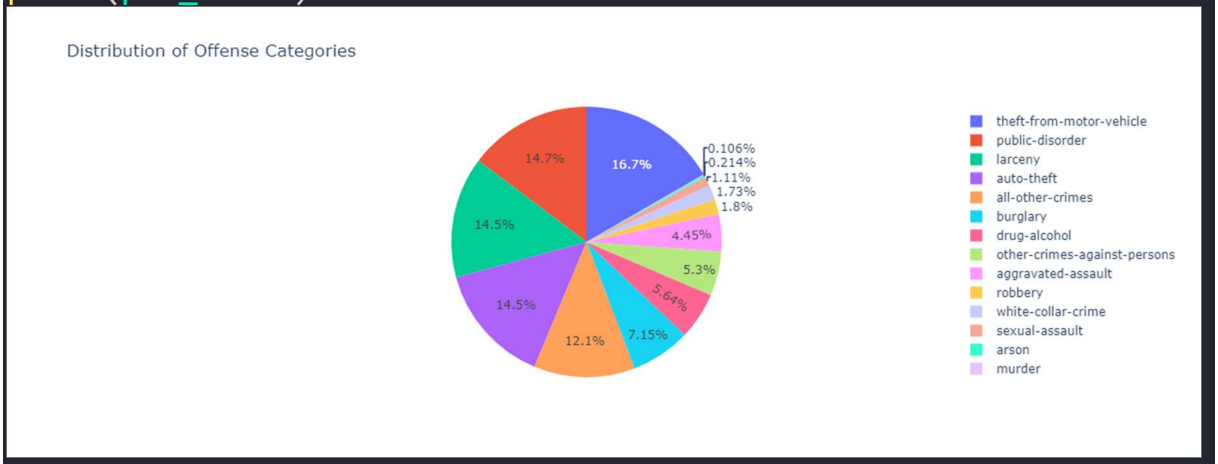
This dataset includes criminal offenses in the City and County of Denver for the previous five calendar years plus the current year to date. The data is based on the National Incident Based Reporting System (NIBRS) which includes all victims of person crimes and all crimes within an incident

```
R.home("bin")
library(tidyverse)
library(lubridate)

data <- read.csv("crime.csv")

data <- data %>%
  mutate(
    first_occurrence_date = mdy_hm(first_occurrence_date),
    reported_date = mdy_hm(reported_date)
  )

# 1. Bar Chart: Offense counts by neighborhood
bar_data <- data %>%
  count(neighborhood_id) %>%
  arrange(desc(n))

bar_chart <- ggplot(bar_data, aes(x = reorder(neighborhood_id,
-n), y = n)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
```

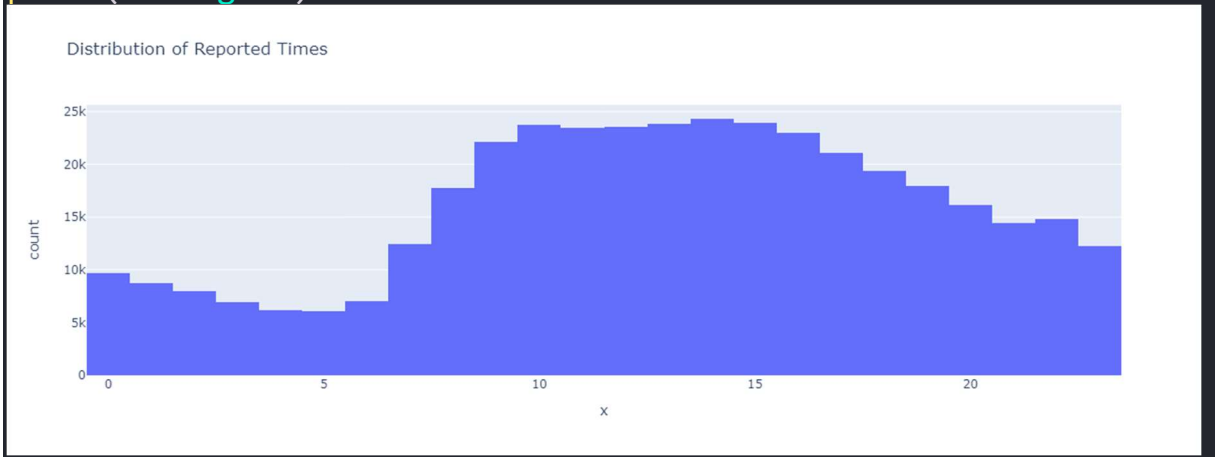
```
  labs(title = "Offense Counts by Neighborhood",
       x = "Neighborhood",
       y = "Count")

print(bar_chart)
```



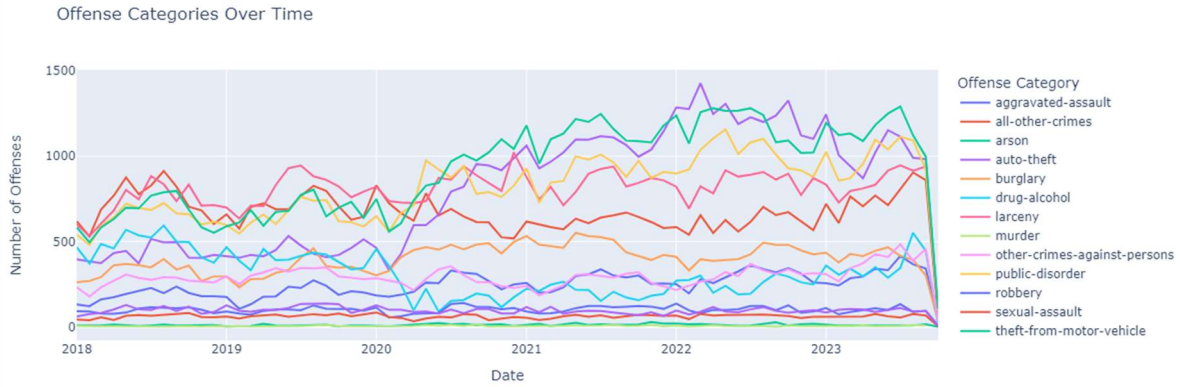

```
# 2. Pie Chart: Distribution of offense categories
pie_data <- data %>%
  count(offense_category_id)

pie_chart <- ggplot(pie_data, aes(x = "", y = n, fill =
offense_category_id)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  labs(title = "Distribution of Offense Categories",
       fill = "Offense Category")

print(pie_chart)
```



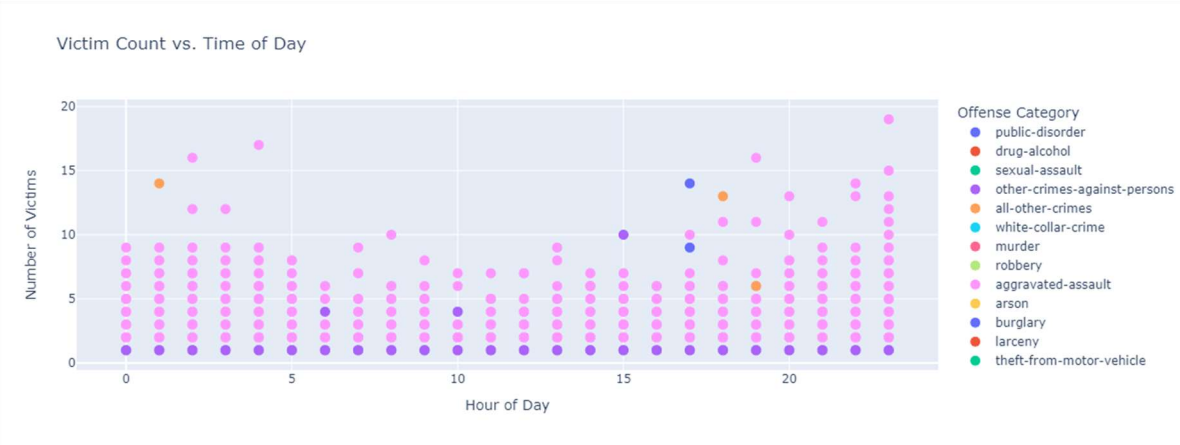
```r
# 3. Histogram: Distribution of reported times
histogram <- ggplot(data, aes(x = hour(reported_date))) +
  geom_histogram(binwidth = 1) +
  labs(title = "Distribution of Reported Times",
       x = "Hour of Day",
       y = "Count")

print(histogram)
```




```r
# 4. Timeline: Comparing different offense categories over time
timeline_data <- data %>%
  mutate(month = floor_date(reported_date, "month")) %>%
  count(month, offense_category_id)

timeline_chart <- ggplot(timeline_data, aes(x = month, y = n,
color = offense_category_id)) +
  geom_line() +
  labs(title = "Offense Categories Over Time",
       x = "Date",
       y = "Number of Offenses",
       color = "Offense Category")

print(timeline_chart)
```

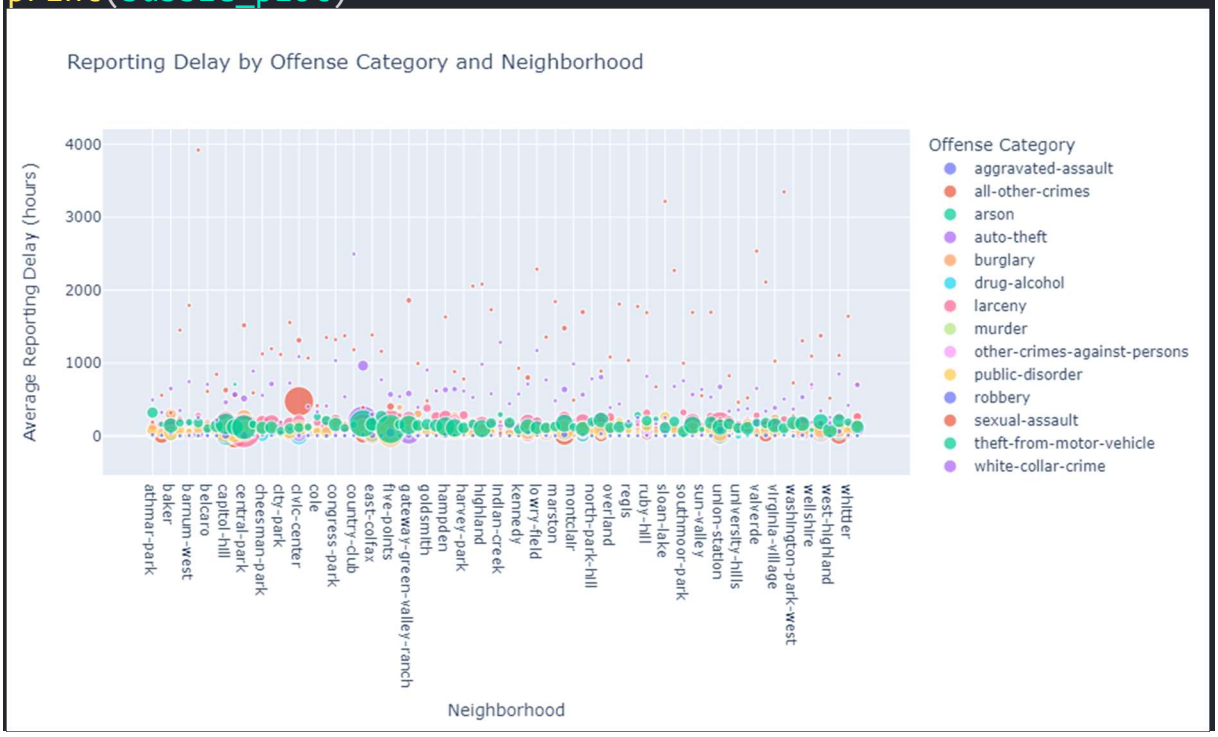



```r
# 5. Scatter Plot: Victim Count vs. Time of Day
data <- data %>%
  mutate(hour_of_day = hour(reported_date))

scatter_plot <- ggplot(data, aes(x = hour_of_day, y =
victim_count, color = offense_category_id)) +
  geom_point(size = 3, alpha = 0.6) +
  labs(title = "Victim Count vs. Time of Day",
       x = "Hour of Day",
       y = "Number of Victims",
       color = "Offense Category")

print(scatter_plot)
```




```r
# 6. Bubble Plot: Reporting Delay by Offense Category and
Neighborhood
data <- data %>%
  mutate(reporting_delay = as.numeric(difftime(reported_date,
first_occurrence_date, units = "hours")))
```

```r
bubble_data <- data %>%
  group_by(neighborhood_id, offense_category_id) %>%
  summarise(
    incident_count = n(),
    avg_reporting_delay = mean(reporting_delay, na.rm = TRUE),
    total_victims = sum(victim_count, na.rm = TRUE)
  ) %>%
  ungroup()

bubble_plot <- ggplot(bubble_data, aes(x = neighborhood_id, y =
avg_reporting_delay, size = incident_count, color =
offense_category_id)) +
  geom_point(alpha = 0.7) +
  scale_size(range = c(1, 20)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Reporting Delay by Offense Category and
Neighborhood",
       x = "Neighborhood",
       y = "Average Reporting Delay (hours)",
       size = "Number of Incidents",
       color = "Offense Category")

print(bubble_plot)
```



Observations

Distribution of Reported Times:

The distribution of reported times is roughly bell-shaped, indicating a tendency towards certain hours of the day.

There are peaks in reported times around mid-day and early evening.

Distribution of Offense Categories:

Larceny and theft-from-motor-vehicle are the most common offense categories.

Murder and arson are among the least frequent offense categories.

Victim Count vs. Time of Day:

Victim counts appear to be highest during daytime and early evening hours.

Certain offense categories, like public-disorder and theft, seem to have higher victim counts compared to others.

Offense Categories Over Time:

Larceny and theft-from-motor-vehicle consistently show high offense counts throughout the period.

There appears to be some seasonal variation in offense counts for certain categories, like aggravated assault.

Offense Counts by Neighborhood:

Five Points, Capitol Hill, and Central Park have the highest offense counts.

Indian Creek and Cheesman Park have relatively low offense counts.

Reporting Delay by Offense Category and Neighborhood:

Reporting delays vary significantly across different offense categories and neighborhoods.

Certain neighborhoods, like Five Points and Capitol Hill, show a wider range of reporting delays across different offense categories.