Name	**Harditya Shah**
UID No.	**2021300114**
Batch	**L**

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

Dataset used:
Crime stats data -
https://www.kaggle.com/datasets/paultimothymooney/denver-crime-data

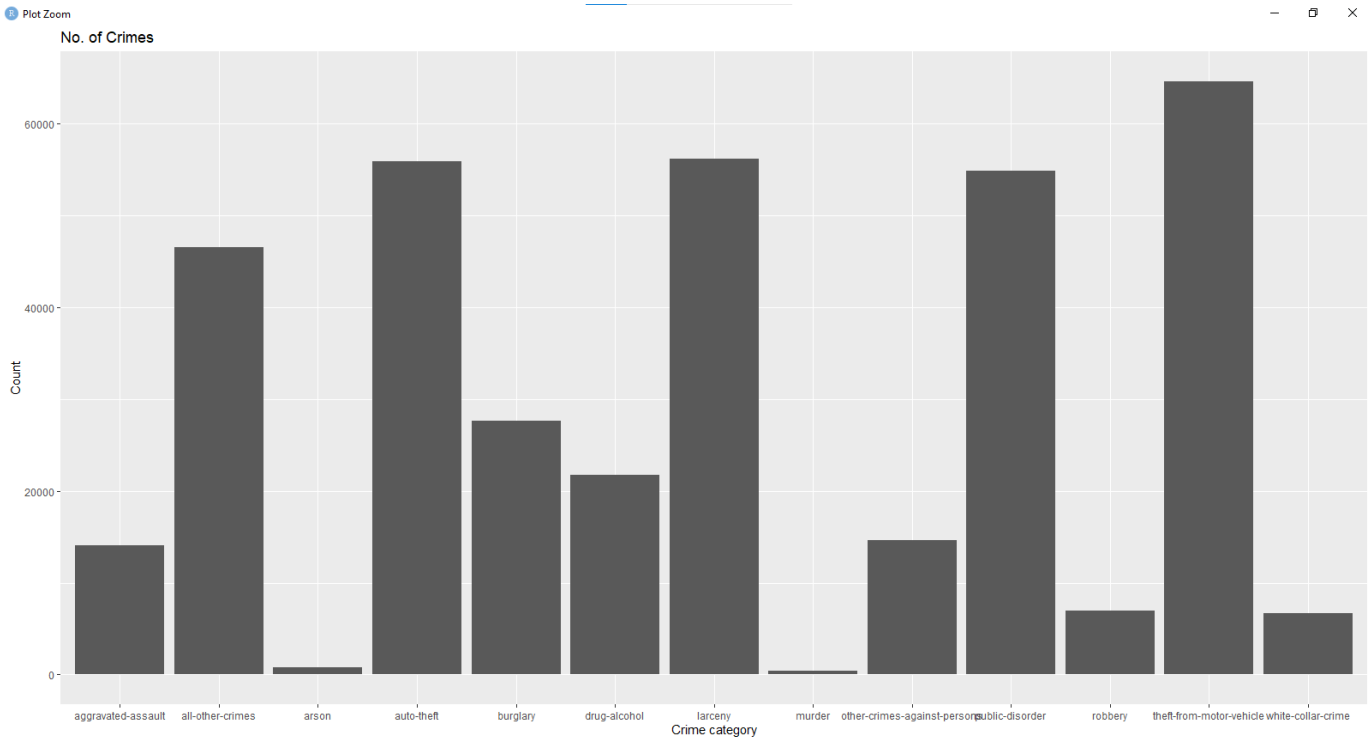
Objectives:
● To understand and apply basic data visualization techniques in R.
● To create various types of charts (Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot) using a crime-related dataset.
● To interpret and analyze the data through visual representations.

Theory:
Data visualization is an essential skill in data analysis that helps in understanding trends, patterns, and relationships within a dataset. R, a powerful statistical programming language, provides a wide range of tools for creating visually appealing and informative charts. In this experiment, we will use basic chart types to analyze crime data and derive insights.
Chart Types:
1. Bar Chart: A bar chart is used to display categorical data with rectangular bars representing the frequency or count of each category.
2. Pie Chart: A pie chart shows the proportion of categories as slices of a pie, useful for comparing parts of a whole.
3. Histogram: A histogram is used to represent the distribution of numerical data by grouping it into bins.
4. Timeline Chart: A timeline chart visualizes data points in chronological order, often used to show trends over time.
5. Scatter Plot: A scatter plot displays the relationship between two numerical variables using points in a Cartesian plane.
6. Bubble Plot: A bubble plot is an extension of a scatter plot where the size of the points (bubbles) represents an additional variable.

```
crime <- read.csv("C:/Users/students/Downloads/crime.csv")
>  View(crime)
> crime <- crime %>% na.omit()
```

Bar-

```
> ggplot(crime, aes(x = offense_category_id)) +
+     geom_bar() +
+     ggtitle("No. of Crimes") +
+     xlab("Crime category") +
+     ylab("Count")
```

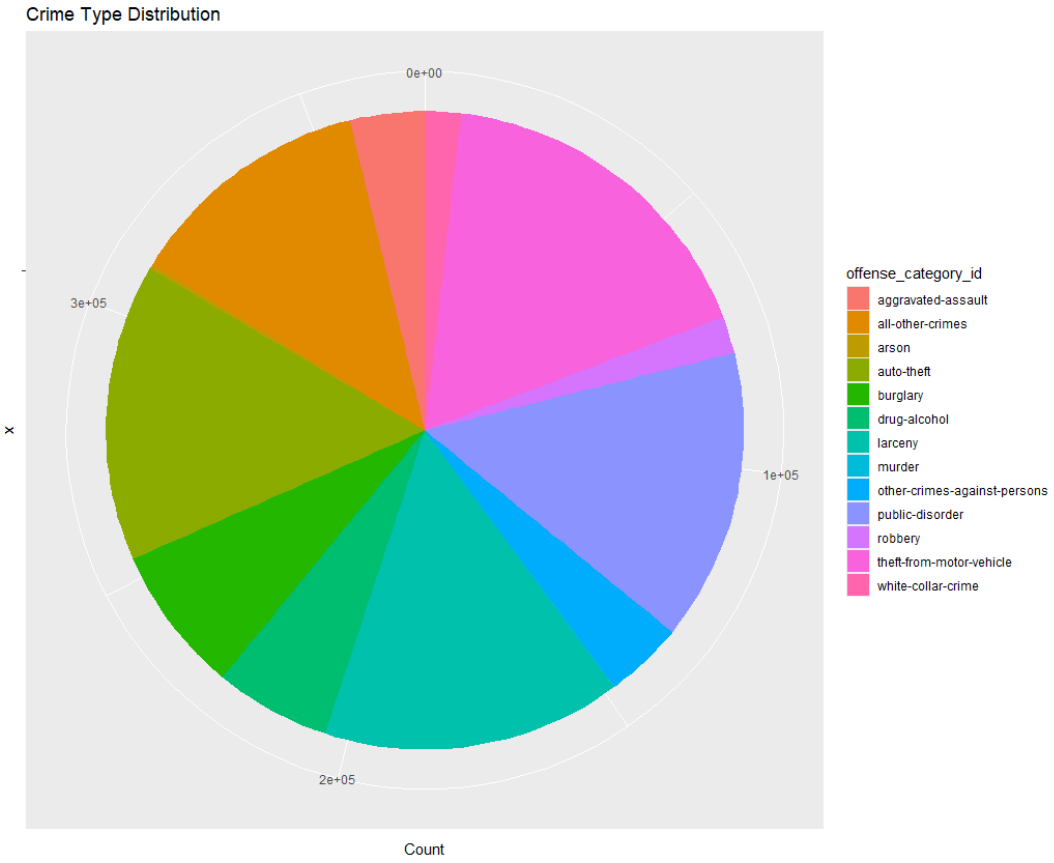


Observations:

1. Larceny and robbery are the most common crime categories, with larceny showing the highest count of incidents.
2. Categories like murder and aggravated assault have significantly lower counts, indicating they are less frequent compared to property crimes.

Pie-

```
> crime_summary <- crime %>% group_by(offense_category_id) %>%
summarise(Count = n())
> ggplot(crime_summary, aes(x = "", y = Count, fill = offense_category_id)) +
geom_bar(width = 1, stat = "identity") +
+    coord_polar("y") +
+    ggtitle("Crime Type Distribution")
> ggplot(crime, aes(x = offense_type_id)) +
+    geom_bar() +
+    ggtitle("No. of Crimes") +
+    xlab("Crime category") +
+    ylab("Count")
```

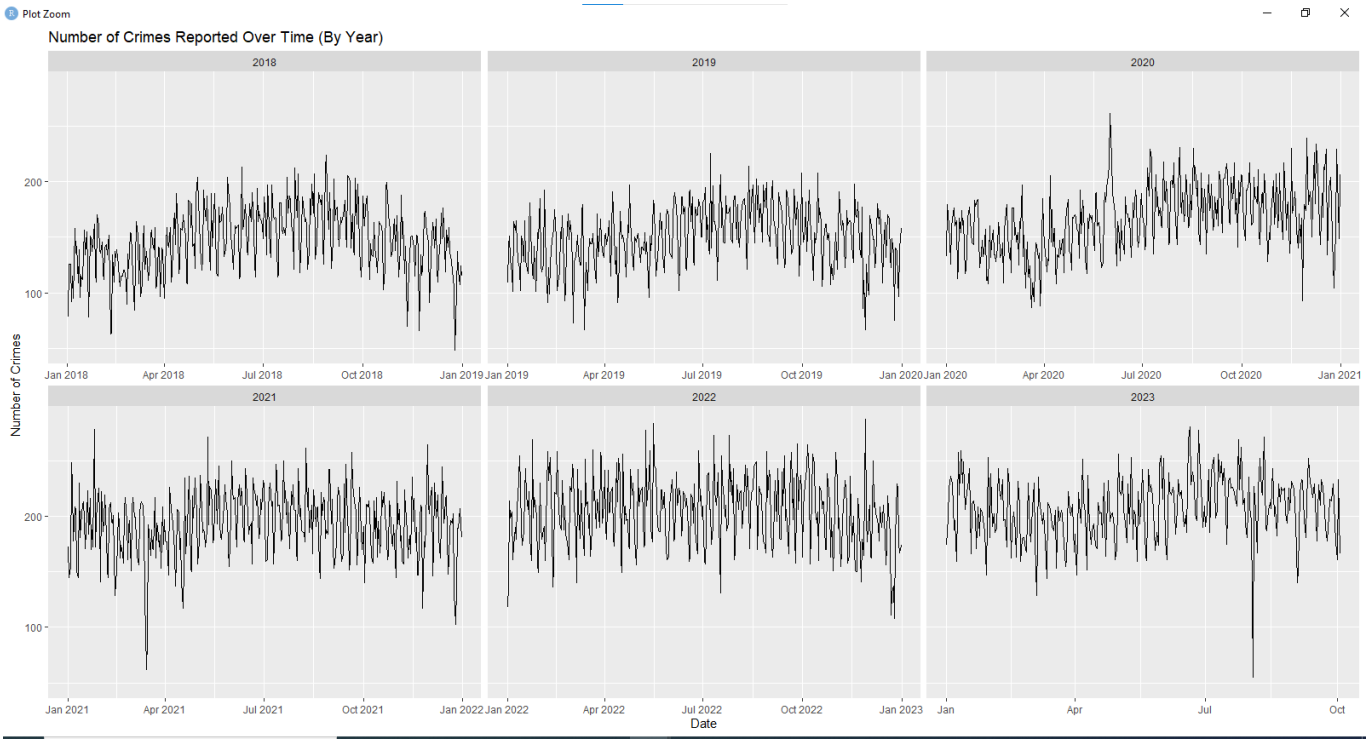


Observation:

1. Property crimes, particularly larceny and burglary, dominate the crime statistics, suggesting a need for targeted prevention strategies in these areas.

Timeline-

```
# Convert 'reported_date' to Date format
> crime$reported_date <- as.Date(crime$reported_date, format="%m/%d/%Y
%I:%M:%S %p")
>
> # Extract the year from 'reported_date'
> crime$year <- format(crime$reported_date, "%Y")
>
> # Aggregate the data by date and year to count the number of incidents per
day
> daily_crimes <- crime %>%
+    group_by(reported_date, year) %>%
+    summarise(crime_count = n())
>
> # Plot the timeline chart with separate plots for each year
> ggplot(daily_crimes, aes(x = reported_date, y = crime_count)) +
+    geom_line() +
+    facet_wrap(~ year, scales = "free_x") +
+    ggtitle("Number of Crimes Reported Over Time (By Year)") +
+    xlab("Date") +
+    ylab("Number of Crimes")
```

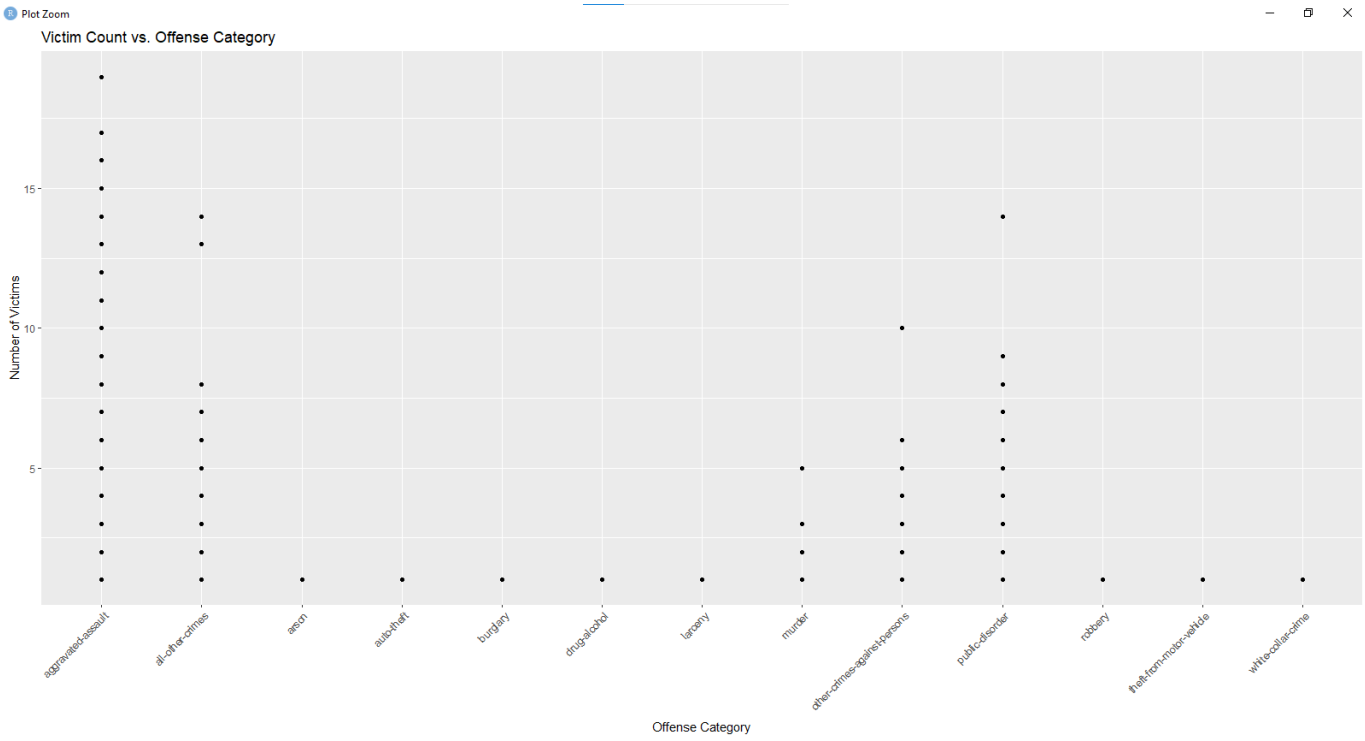


Observation:

1. Each year shows similar patterns, but slight variations in peak crime months could inform targeted interventions during high-crime periods.
2. The data exhibits seasonal fluctuations, with noticeable peaks and troughs throughout each year, suggesting that crime rates may vary by season.

Scatter-

```
crime_data <- crime_data %>% na.omit()
> ggplot(crime_data, aes(x = offense_category_id, y = victim_count)) +
+     geom_point() +
+     ggtitle("Victim Count vs. Offense Category") +
+     xlab("Offense Category") +
+     ylab("Number of Victims") +
+     theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

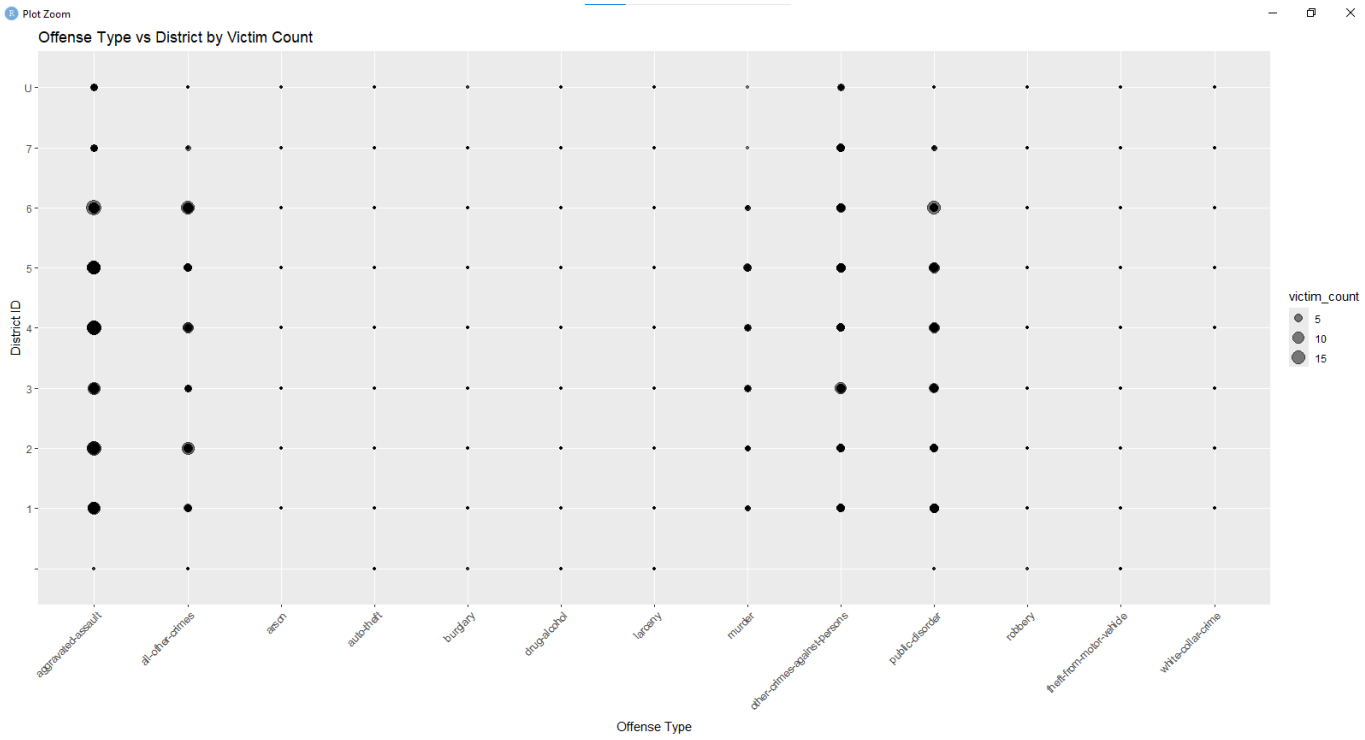


Observation:

1. Certain categories like Aggravated assault and Public Disorder may have outlier points, indicating specific offenses that have attracted more victims than others.
2. Categories with higher victim counts could be prioritized for intervention or further investigation to understand underlying causes.

Bubble-

```
ggplot(crime_data, aes(x = offense_category_id, y = district_id, size =
victim_count)) +
+    geom_point(alpha = 0.5) +
+    ggtitle("Offense Type vs District by Victim Count") +
+    xlab("Offense Type") +
+    ylab("District ID") +
+    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Observation:

1. Offense types like "assault" and "theft" appear to have higher victim counts across multiple districts, suggesting these are common issues.
2. Some districts like 3, 4, 5, 6 have a wider range of offense types with varying victim counts, indicating diverse crime issues.