Name: Omkar Shirwadkar
UID: 2021600062
Batch: D
Class: CSE(AI & ML)

EXPERIMENT NO: 4
Create basic charts using R programming language on dataset Crime or Police / Law and Order

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

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

Description:
This dataset captures a comprehensive snapshot of criminal activities across multiple Indian cities from 2020 to 2024. It includes detailed information about the type of crime, the date and time of occurrence, the weapon used, victim demographics, and the level of police deployment. With crimes ranging from identity theft to homicide, this dataset offers valuable insights for researchers, policymakers, and law enforcement agencies aiming to understand crime patterns and improve public safety. The data also includes information on whether the case was closed, providing a clear view of crime resolution rates across cities

Attributes/Columns:
The dataset contains the following columns:

incident_id, offense_id, offense_code offense_code_extension, offense_type_id, offense_category_id, first_occurrence_date, last_occurrence_date, reported_date, incident_address, geo_x, geo_y, geo_lon, geo_lat, district_id, precinct_id, neighborhood_id, is_crime, is_traffic, victim_count.

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.

Code:

```r
install.packages("ggplot2") # Install ggplot2 if you haven't already library(ggplot2)

# Load the ggplot2 library setwd("/home/mca/Downloads/denver") # Replace
with your directory path


# Load the dataset
crime_data <- read.csv("crime.csv", stringsAsFactors = FALSE) # Use stringsAsFactors =
FALSE to avoid factor conversion

# Check if the data loaded correctly
head(crime_data)

str(crime_data) # Structure of the dataset summary(crime_data)
# Summary statistics

# 1 Create a bar chart of offense categories ggplot(data =
crime_data, aes(x = offense_category_id)) +
  geom_bar(fill = "blue") + labs(title = "Number of
  Crimes by Offense Category", x = "Offense
  Category", y = "Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# 2 Convert reported_date to Date type
crime_data$reported_date <- as.Date(crime_data$reported_date)

# Create a time series plot crime_counts <-
as.data.frame(table(format(crime_data$reported_date, "%Y-%m")))

ggplot(crime_counts, aes(x = Var1, y = Freq)) +
  geom_line(group = 1, color = "red")
  + labs(title = "Crimes Over Time", x
  = "Month-Year", y = "Number of
  Crimes") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# 3 Create a heatmap of crimes by neighborhood and offense category heatmap_data

<- table(crime_data$neighborhood_id, crime_data$offense_category_id)

ggplot(as.data.frame(heatmap_data), aes(Var1, Var2)) +
```

```
geom_tile(aes(fill = Freq), color = "white") + scale_fill_gradient(low =
"white", high = "blue") + labs(title = "Heatmap of Crimes by
Neighborhood and Offense Category", x = "Neighborhood ID", y =
"Offense Category ID") +
theme_minimal()
```

```
# 4 Create a box plot of victim count by offense category
ggplot(crime_data, aes(x = offense_category_id, y = victim_count)) +
geom_boxplot(fill = "lightblue") + labs(title = "Victim Count
Distribution by Offense Category", x = "Offense Category",
y = "Number of Victims") +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
theme_minimal()
```

```
# 5 Pie chart of offense categories crime_category_counts <-
table(crime_data$offense_category_id) crime_category_df <-
as.data.frame(crime_category_counts)
```

```
ggplot(crime_category_df, aes(x = "", y = Freq, fill = Var1)) +
geom_bar(stat = "identity", width = 1) + coord_polar("y")
+ labs(title = "Proportion of Crimes by Offense
Category") + theme_void() + theme(legend.position =
"right")
```

Plots:

1. Bar Chart (Offense Categories)

This bar chart illustrates the number of crimes associated with each offense category. The length of each bar represents the frequency of each offense category in the dataset, allowing for easy comparison of crime types and identification of the most prevalent offenses.

Number of Crimes by Offense Category
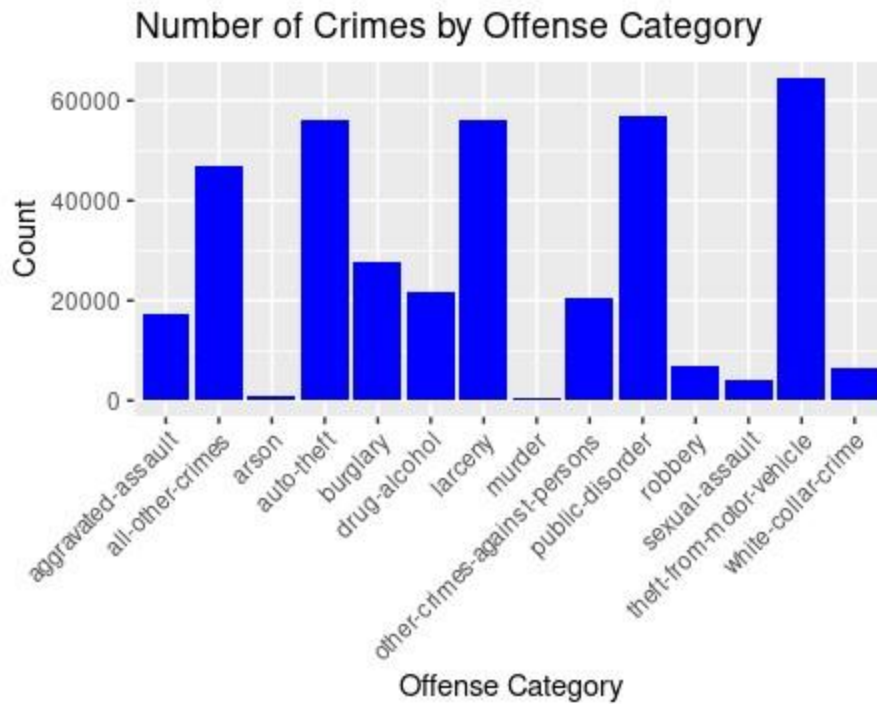
2. Time Series Chart

This timeline chart displays how the number of reported crimes changes over time. Each point represents the count of crimes reported in a specific month, and the line connects these points to illustrate trends over time. This visualization helps in identifying patterns, such as periods of increased or decreased crime activity.

Crimes Over Time
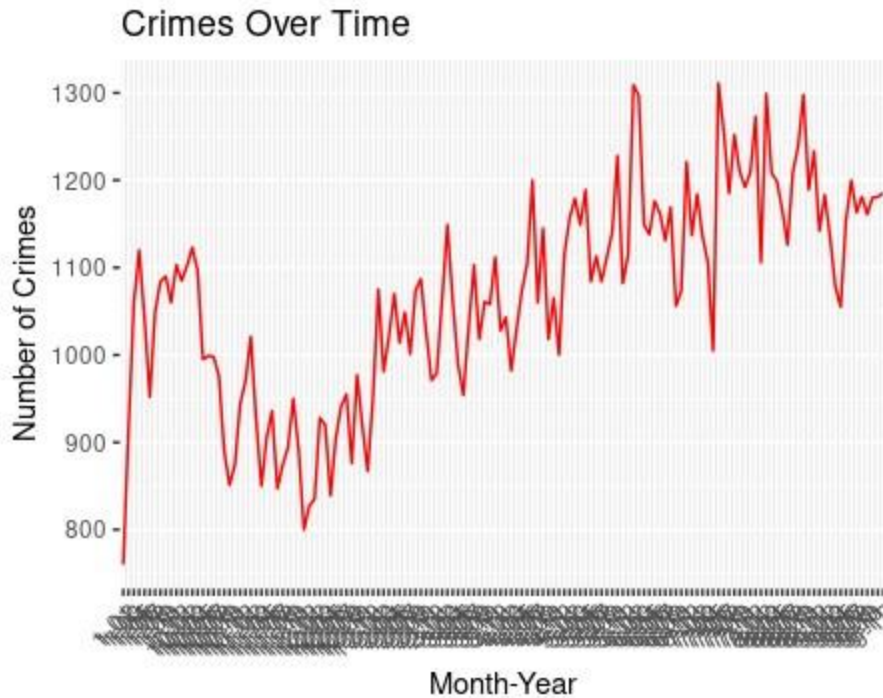
3. Heatmap

The heatmap visualizes the relationship between neighborhoods and offense categories. Each cell represents the frequency of crimes occurring in a specific neighborhood for each offense category. The color intensity indicates the number of incidents, allowing for quick identification of areas with high crime rates and the types of crimes prevalent in those neighborhoods.
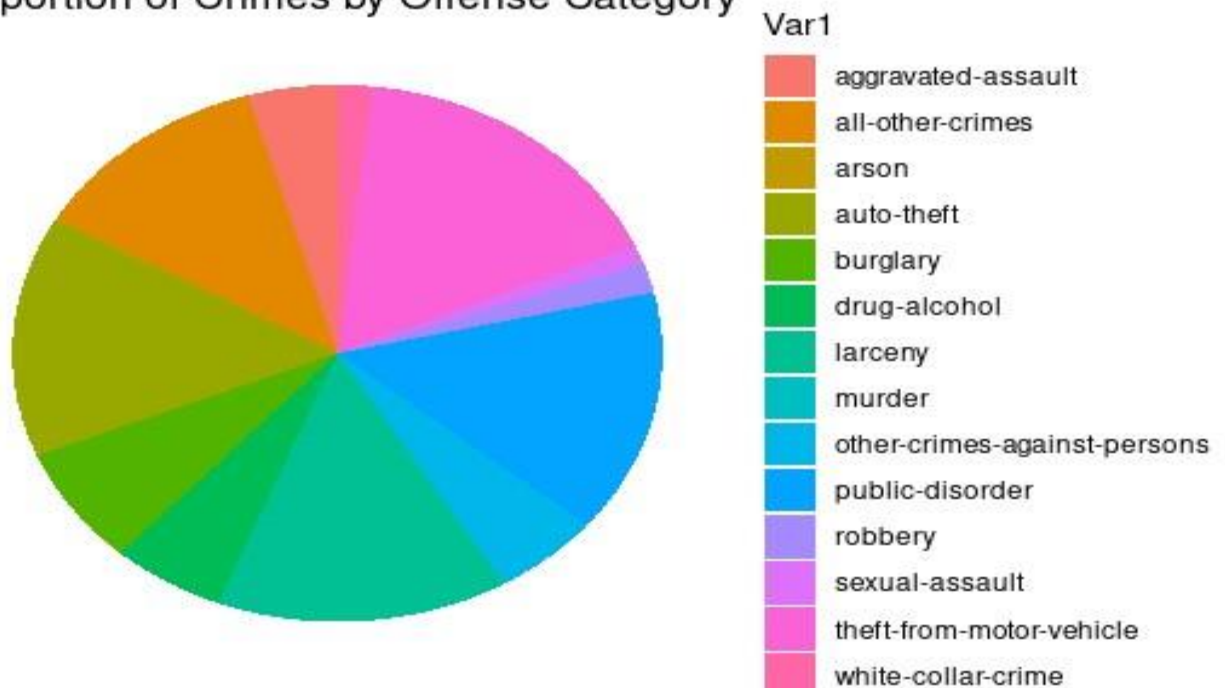

4. Box Plot

This box plot shows the distribution of the number of victims for each offense category. It displays the median, quartiles, and potential outliers, helping to understand how victim counts vary across different crime types.

Victim Count Distribution by Offense Category

Number of Victims (y-axis)
Offense Category (x-axis): aggravated-assault... larceny... auto-theft... burglary... drug-alcohol... robbery... other-crimes-against-persons... disorderly... white-collar... all-other-crimes... theft-from-motor-vehicle

5. Pie Chart

The pie chart visualizes the proportion of different offense categories in the Denver crime dataset. Each slice represents a crime category, providing an easy-to-understand breakdown of how various offenses contribute to the overall crime rate. This visualization helps identify which categories are most prevalent, highlighting trends such as whether certain types of crimes, like property crimes or drug offenses, dominate the dataset. The chart offers a clear, high-level overview of crime distribution across categories.

Proportion of Crimes by Offense Category

Var1
- aggravated-assault
- all-other-crimes
- arson
- auto-theft
- burglary
- drug-alcohol
- larceny
- murder
- other-crimes-against-persons
- public-disorder
- robbery
- sexual-assault
- theft-from-motor-vehicle
- white-collar-crime

Conclusion:

Successfully created multiple types of charts using R to visualize crime data. Gained insights into the distribution, frequency, and relationships within the crime Dataset. Developed an understanding of how different chart types can be used to analyze and present data effectively. This experiment demonstrated the power of data visualization in uncovering patterns and trends in a crime dataset. By using R, we efficiently created visual representations that allowed us to explore the data from different perspectives, leading to better-informed conclusions.