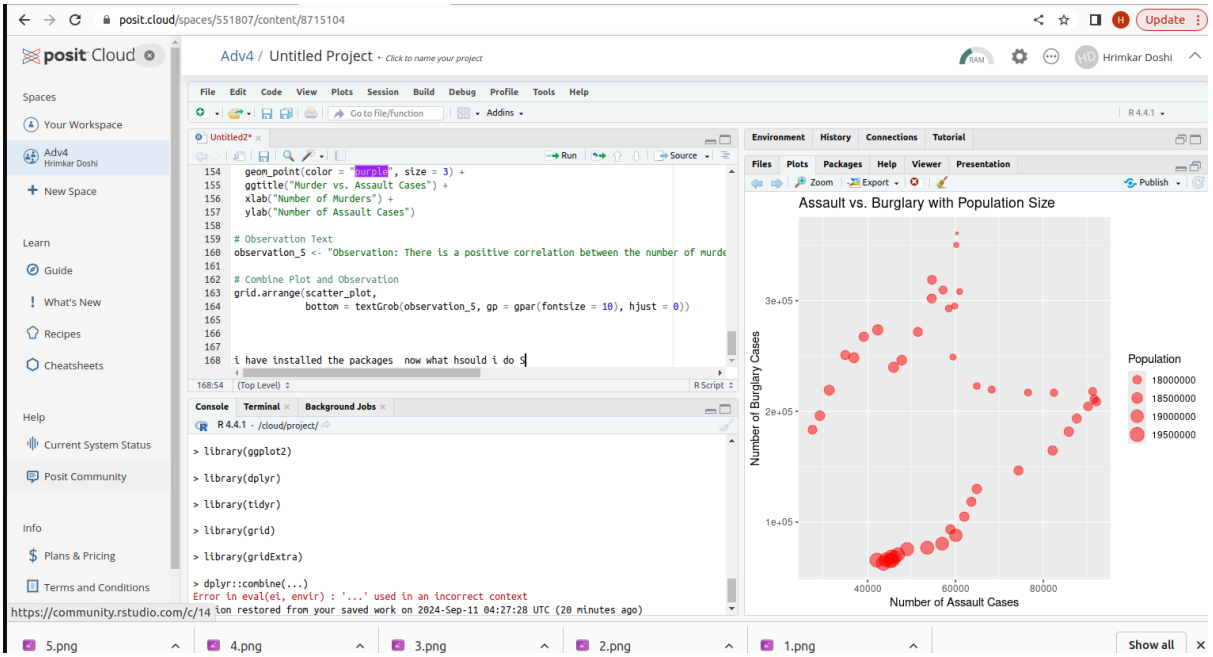
Name	Hrimkar Doshi
UID	202170021
Exp	4
Date	11/09/2024

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

General overview of the setup

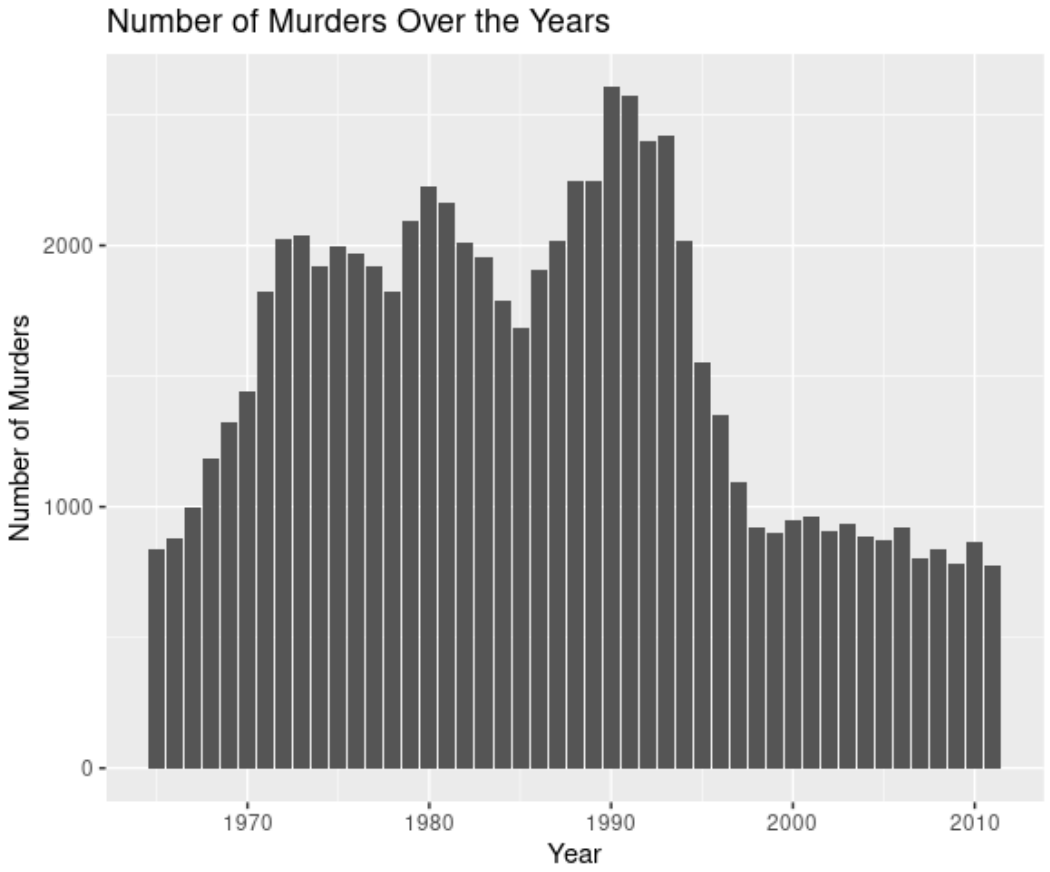


Plotting Graphs

Bar Chart

```
bar_chart <- ggplot(crime_data, aes(x = factor(Year), y = Murder)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  ggtitle("Number of Murders by Year") +
  xlab("Year") +
  ylab("Number of Murders") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Observation Text
```

```
observation_1 <- "Observation: The bar chart shows an increasing trend in the number of murders over the years from 1965 to 1973, with a noticeable spike around the early 1970s."

# Combine Plot and Observation
grid.arrange(bar_chart,
             bottom = textGrob(observation_1, gp = gpar(fontsize = 10), hjust = 0))
```



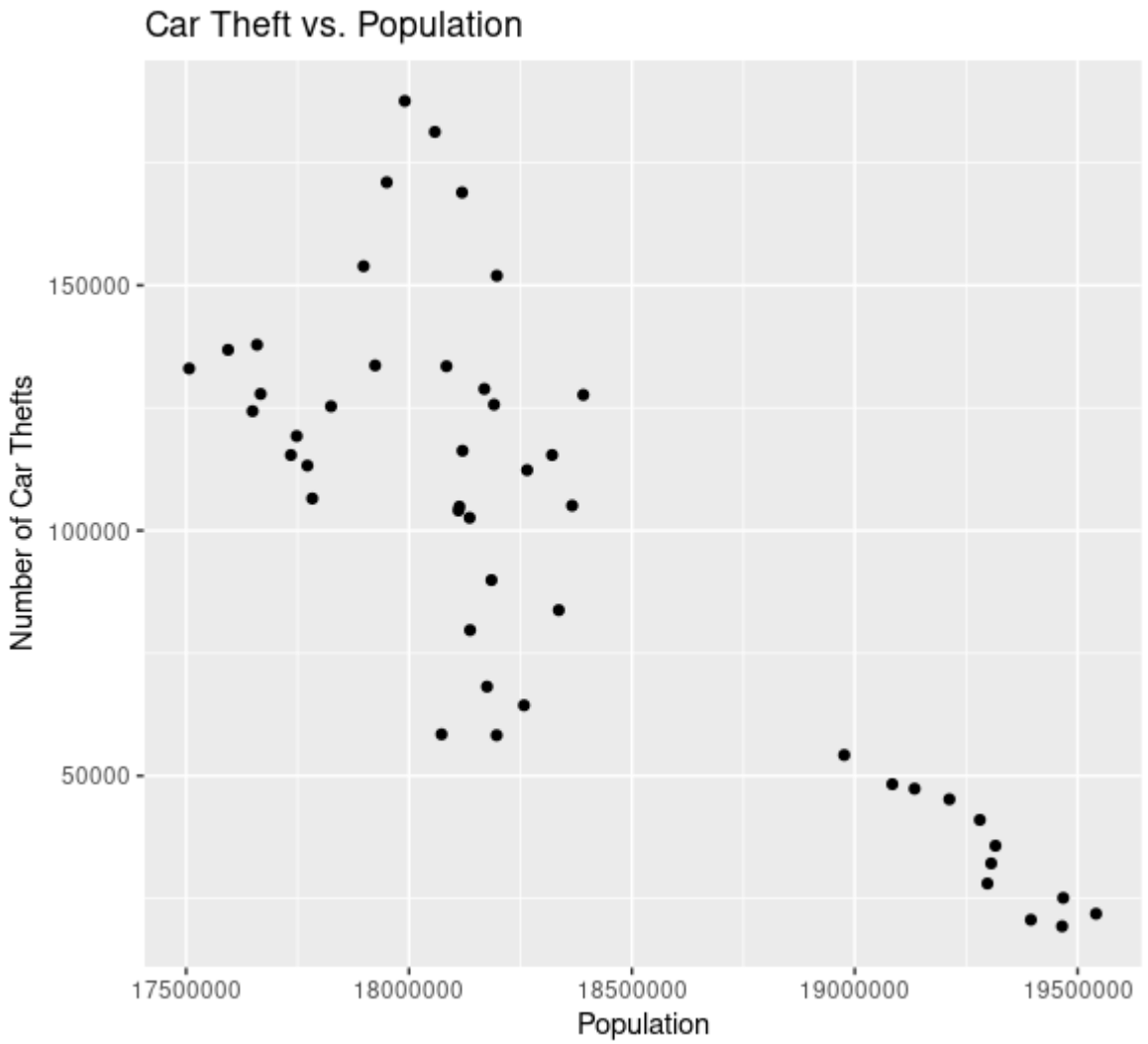
Number of Murders Over the Years (Bar Chart):

- The number of murders increased from the early 1970s, peaked around 1990, and then gradually decreased through the 2000s.
- The peak period for murders appears to be the late 1980s to early 1990s, after which a significant decline can be observed.
- By the 2000s, the number of murders stabilized at a lower rate compared to the peak years.

Scatter Chart

```
ggplot(crime_data, aes(x = population, y = car theft)) +

 geom_point(color = "purple", size = 3) +

 ggtitle("population vs. car theft") +
```

```
xlab("population") +

ylab("car theft")-
```



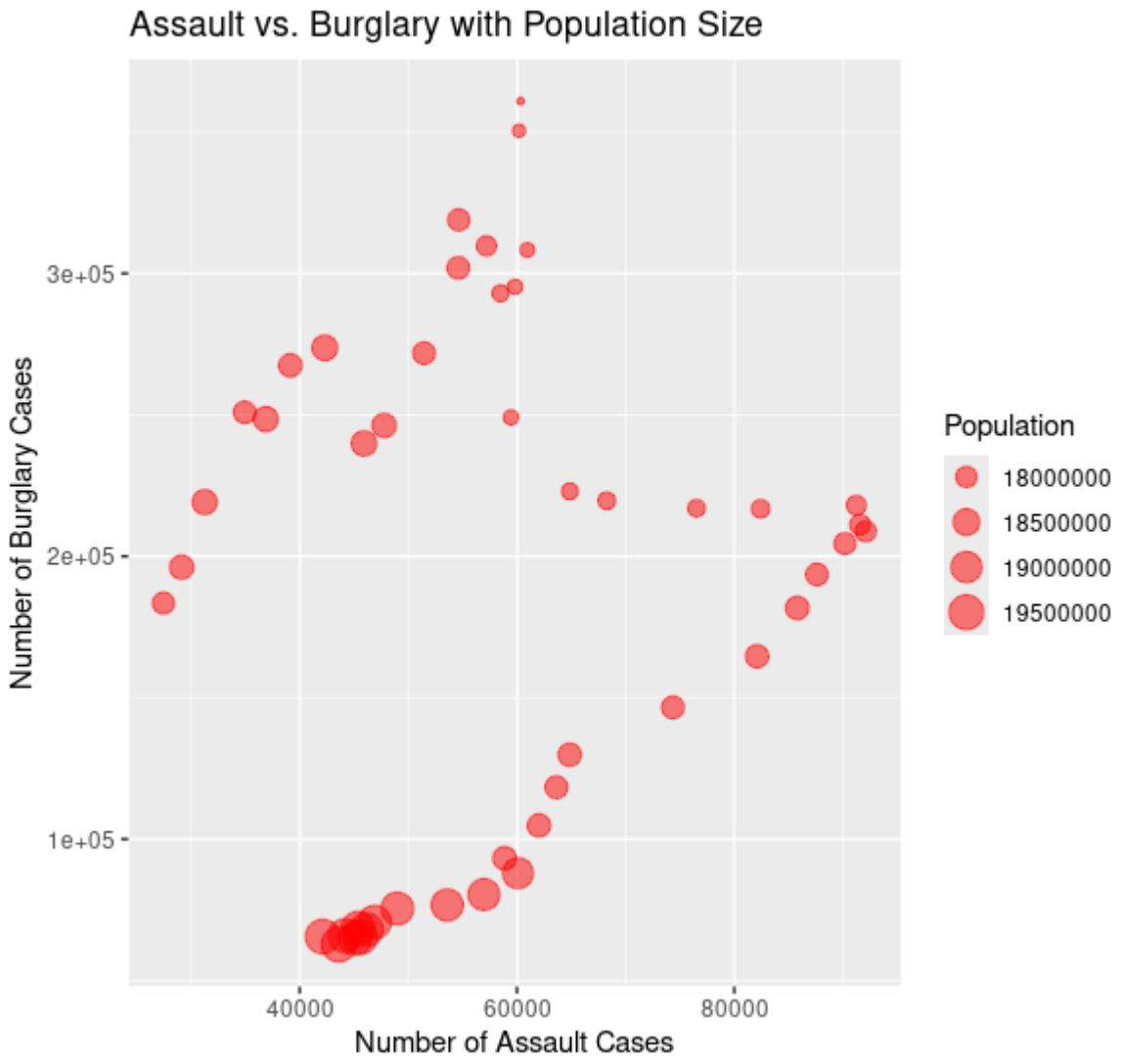
Car Theft vs. Population (Scatter Plot):

1. There is an inverse relationship between the number of car thefts and the population size in this scatter plot.
2. As the population increases, the number of car thefts generally decreases.
3. In the range where the population is below 18.5 million, car thefts seem more frequent, with the highest point of car thefts occurring around a population of 17.8 million.

Scatter plot

```
ggplot(crime_data, aes(x = Assault, y = Burglary, size = Population)) +
  geom_point(alpha = 0.5, color = "red") +
  ggtitle("Assault vs. Burglary with Population Size") +
```

```
xlab("Number of Assault Cases") +
ylab("Number of Burglary Cases")
```



The bubble plot "Assault vs. Burglary with Population Size" visualizes the relationship between the number of **assault cases** and **burglary cases**, with the **population size** represented by the size of the bubbles.

Here are key observations:

1. **Positive Correlation**:
    - There appears to be a positive correlation between the number of assault cases and burglary cases. As the number of assault cases increases, burglary cases also tend to rise.
2. **Population Size Representation**:
    - Larger bubbles indicate cities or regions with a larger population, while smaller bubbles represent areas with smaller populations.

- Larger population centers often have a higher number of both assaults and burglaries, as seen from the larger bubbles positioned towards higher values on both axes.

3. **Spread of Data**:
   - The data points (bubbles) are spread across the graph, indicating some variance in the number of assault and burglary cases. However, larger bubbles tend to cluster more towards higher crime values.
4. **Transparency (Alpha)**:
   - The semi-transparent nature of the bubbles allows overlap visualization, making it easier to distinguish areas where data points are dense.

Timeline chart

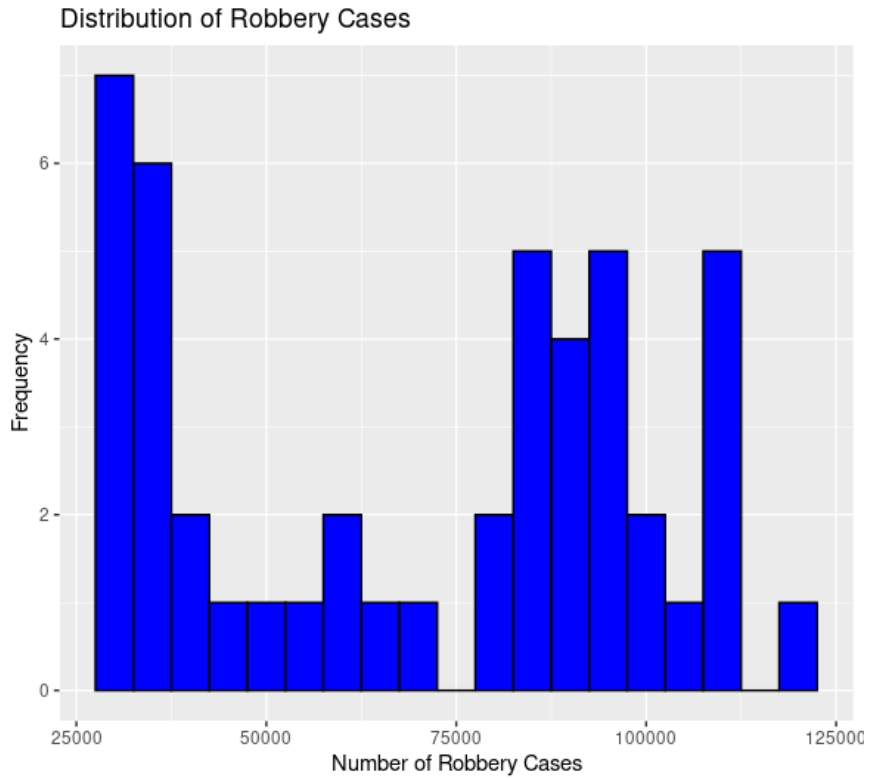
```
crime_data <- crime_data %>%
  mutate(Total_Crime = Murder + Rape + Robbery + Assault + Burglary + CarTheft)


timeline_chart <- ggplot(crime_data, aes(x = Year, y = Total_Crime)) +
  geom_line(color = "darkgreen", size = 1) +
  ggtitle("Total Crimes Over the Years") +
  xlab("Year") +
ylab("Total Number of Crimes")
```

Rape Cases Over the Years (Line Graph):

- There is a rise in rape cases starting in the late 1960s, reaching a peak in the early 1990s.
- After the early 1990s, the number of rape cases shows a gradual decline.
- The most noticeable drop in cases seems to occur between 2000 and 2010.

Histogram Plot

```
histogram <- ggplot(crime_data, aes(x = Robbery)) +
  geom_histogram(binwidth = 5000, fill = "orange", color = "black") +
  ggtitle("Distribution of Robbery Cases") +
  xlab("Number of Robbery Cases") +
  ylab("Frequency")
```



Distribution of Robbery Cases:

- A histogram that depicts the frequency distribution of robbery cases. The distribution is bimodal, showing two distinct peaks, one around 25,000 cases and the other around 100,000 cases.