Sardar Patel Institute of Technology

(Autonomous Institute Affiliated to University of Mumbai)
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

Experiment no 4

Name:	Anshuman Dhara
UID:	2021300026
Batch:	A
Exp:	4

Typ

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

- Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot

- Write observations from each chart

Dataset Description:

This data is collated from https://data.gov.in. It has state-wise data on the various crimes committed against women between 2001 to 2021. Some crimes that are included are Rape, Kidnapping and Abduction, Dowry Deaths etc.

Dataset Link:
https://www.kaggle.com/datasets/balajivaraprasad/crimes-against-women-in-india-2001-2021?select=description.csv

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.

Code, Plots and Inference:

```
install.packages("ggplot2")
install.packages("dplyr")
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Warning message in install.packages("ggplot2"):
"installation of package 'ggplot2' had non-zero exit status"
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Warning message in install.packages("dplyr"):
"installation of package 'dplyr' had non-zero exit status"
```

```
[ ] library(ggplot2)
    library(dplyr)
```

```
[ ] data <- read.csv("CrimesOnWomenData.csv")
```

1. Bar chart:

```
colnames(data)
colnames(data) <- trimws(colnames(data))

# Now sum the relevant columns for all years to get the total number of crimes per year
data$total_crimes <- rowSums(data[, c("Rape", "K.A", "DD", "AoW", "AoM", "DV", "WT")], na.rm = TRUE)

# Bar plot: Total crimes per year
ggplot(data, aes(x = Year, y = total_crimes)) +
  geom_bar(stat = "identity", fill = "steelblue") +  # Creates the bar chart
  labs(title = "Total Crimes Per Year", x = "Year", y = "Total Crimes") +  # Adding titles and labels
  theme_minimal()  # A clean theme for the plot
```

'X' · 'State' · 'Year' · 'Rape' · 'K.A' · 'DD' · 'AoW' · 'AoM' · 'DV' · 'WT'
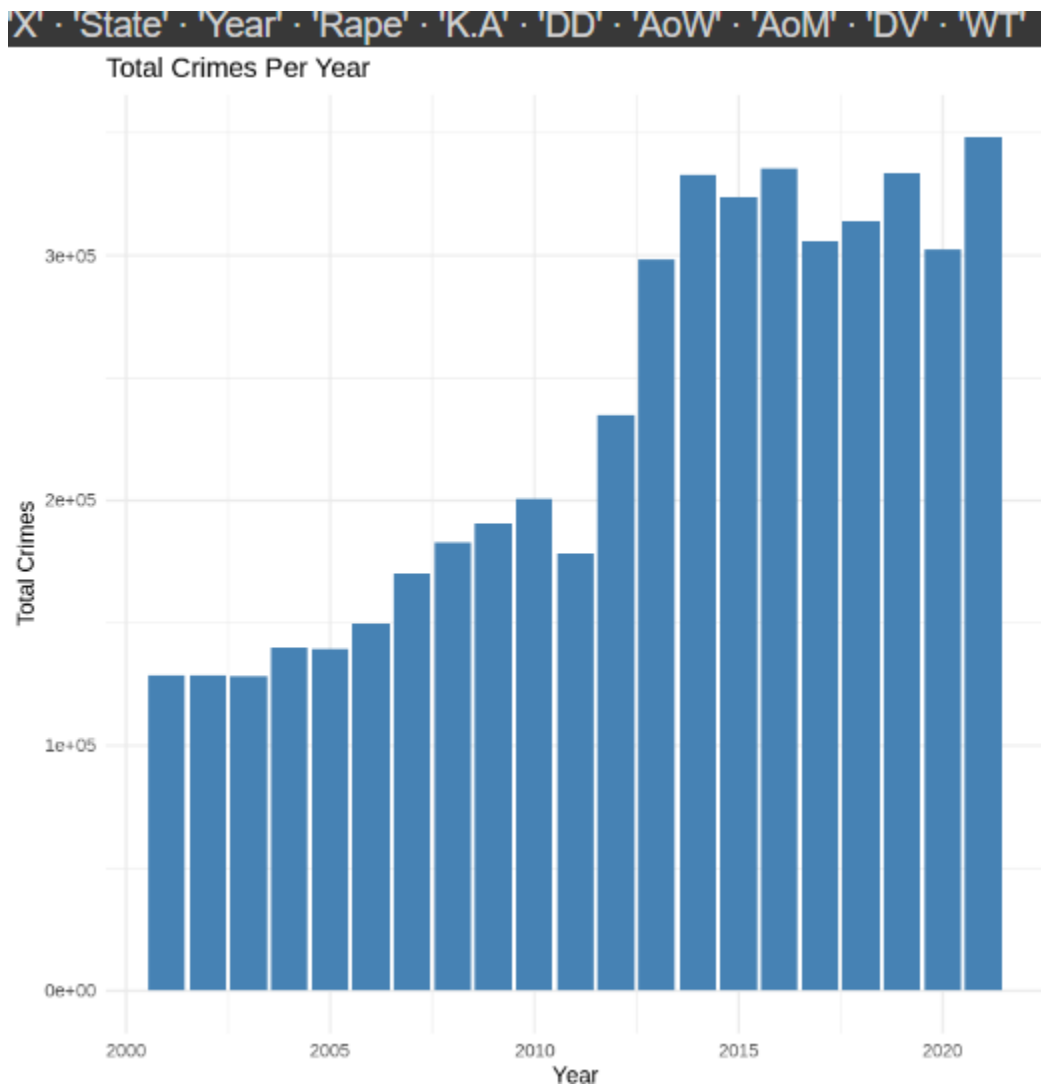
This bar chart shows the total number of crimes of all types occurring from 2001 to 2021.

2. Pie chart:

```
# howing distribution of crimes in a Andhra Pradesh
crime_counts <- c(data$Rape[1], data$K_A[1], data$DD[1], data$AoW[1], data$AoM[1], data$DV[1], data$WT[1])
pie(crime_counts, labels=c("Rape", "Kidnap & Assault", "Dowry Deaths", "Assault Women", "Assault Modesty", "Domestic Violence", "Trafficking"),
    main="Crime Distribution in Andhra Pradesh")
```


Crime Distribution in Andhra Pradesh
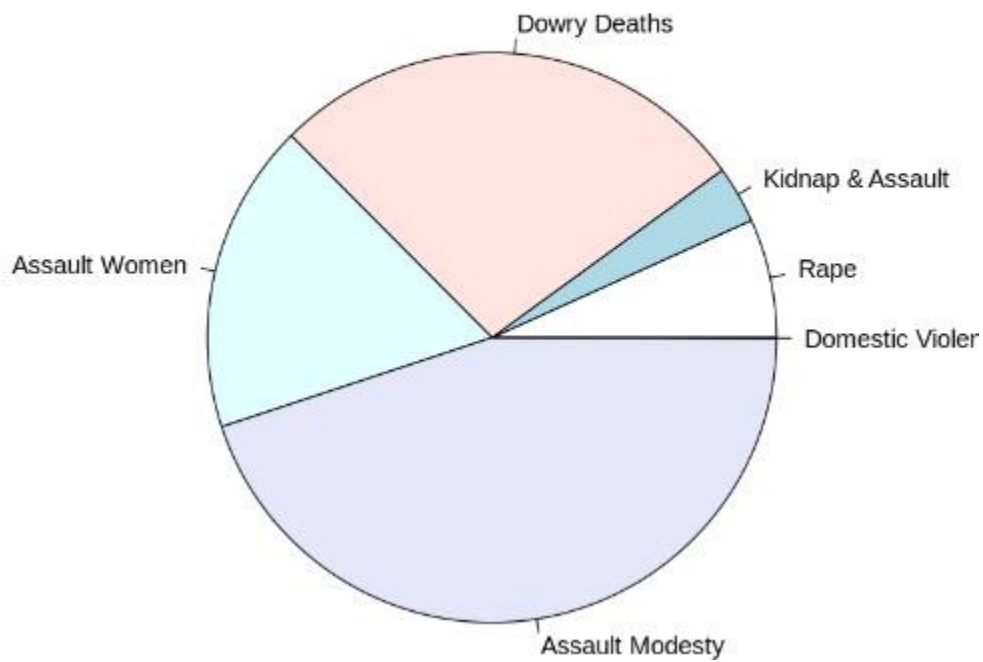
Crime Distribution in Arunachal Pradesh

Kidnap & Assault

Rape

Domestic Violen

Assault Modesty

Assault Women

Dowry Deaths
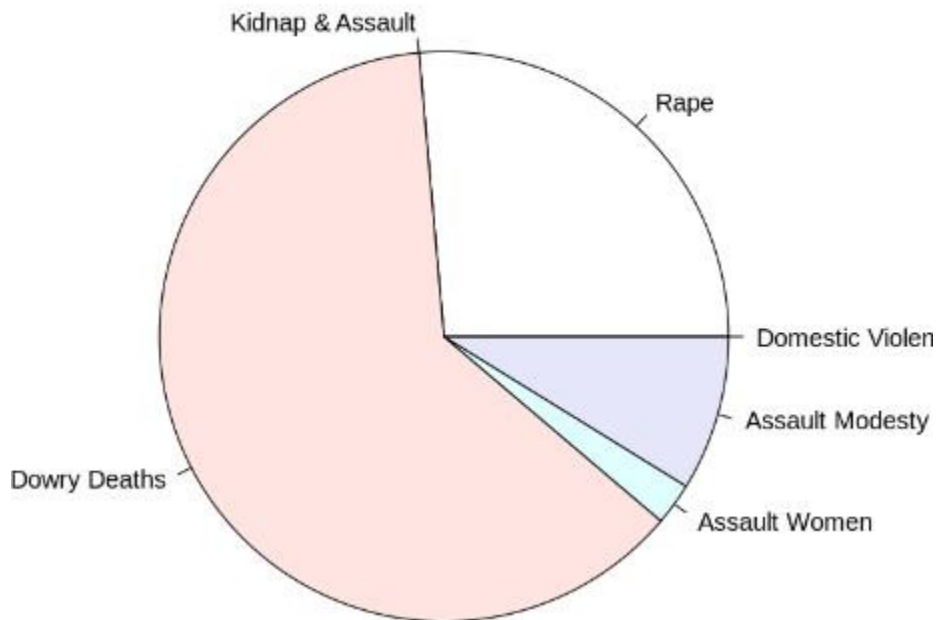
The pie chart shows the relative proportion of different crimes in a particular state, indicating which type of crime is most or least common.

3. **Histogram:**

```
# Histogram for Rape cases across all states
hist(data$Rape, col='pink', main="Distribution of Rape Cases", xlab="Number of Rape Cases")
```

Distribution of Rape Cases
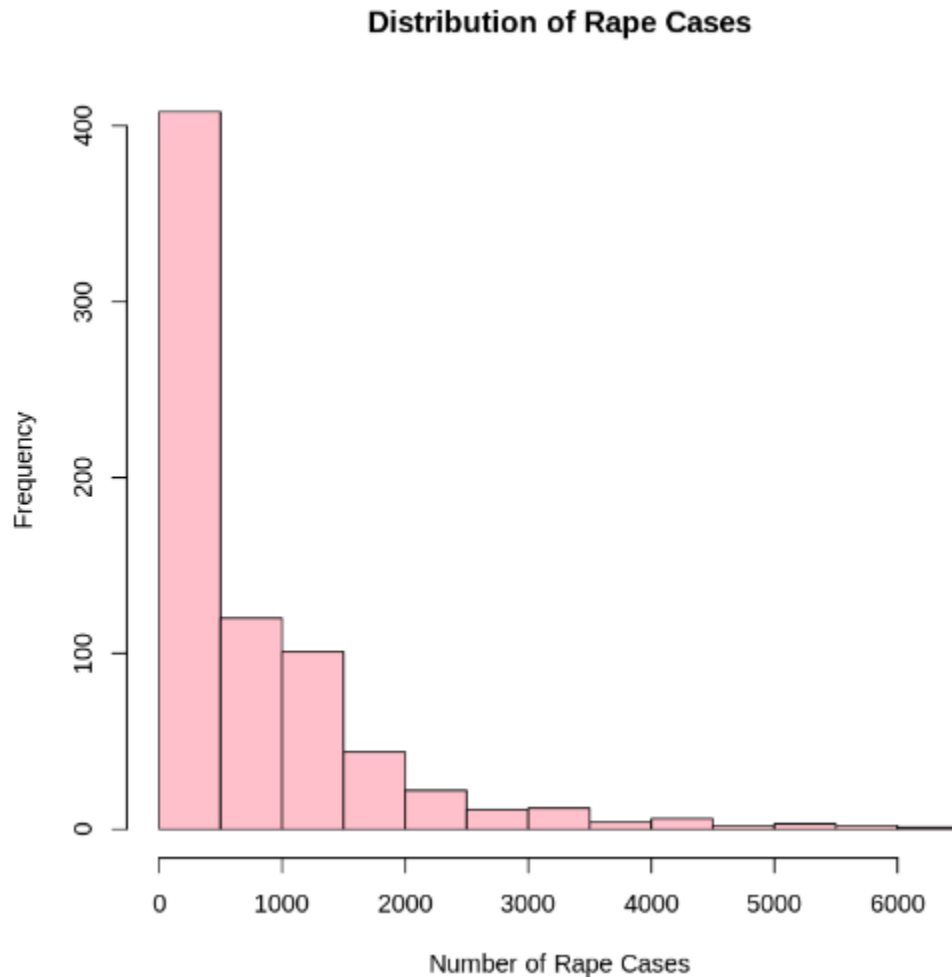
This plot shows the distribution of the number of rape cases across the dataset, identifying whether most states report low, medium, or high numbers of cases.
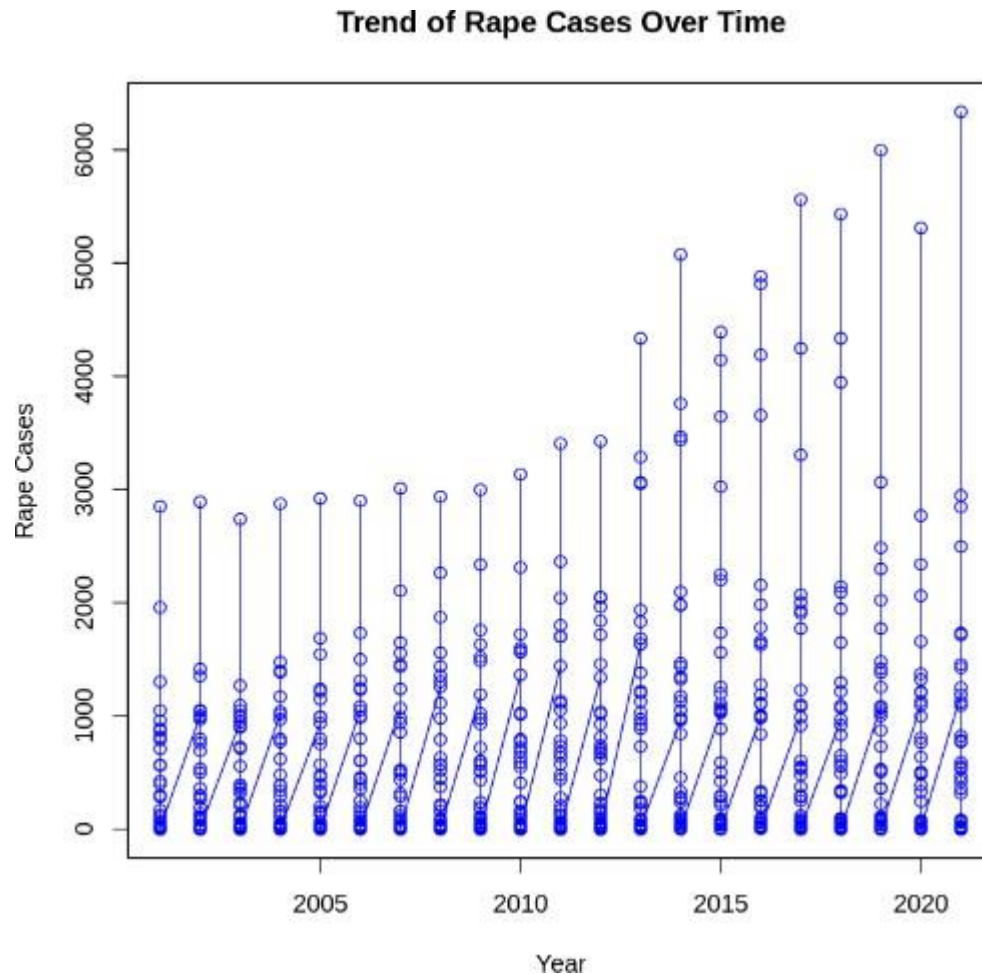
```
# Histogram for Rape cases across all states
hist(data$DD, col='pink', main="Distribution of Dowry Deaths", xlab="Number of Dowry Deaths")
```

Distribution of Dowry Deaths
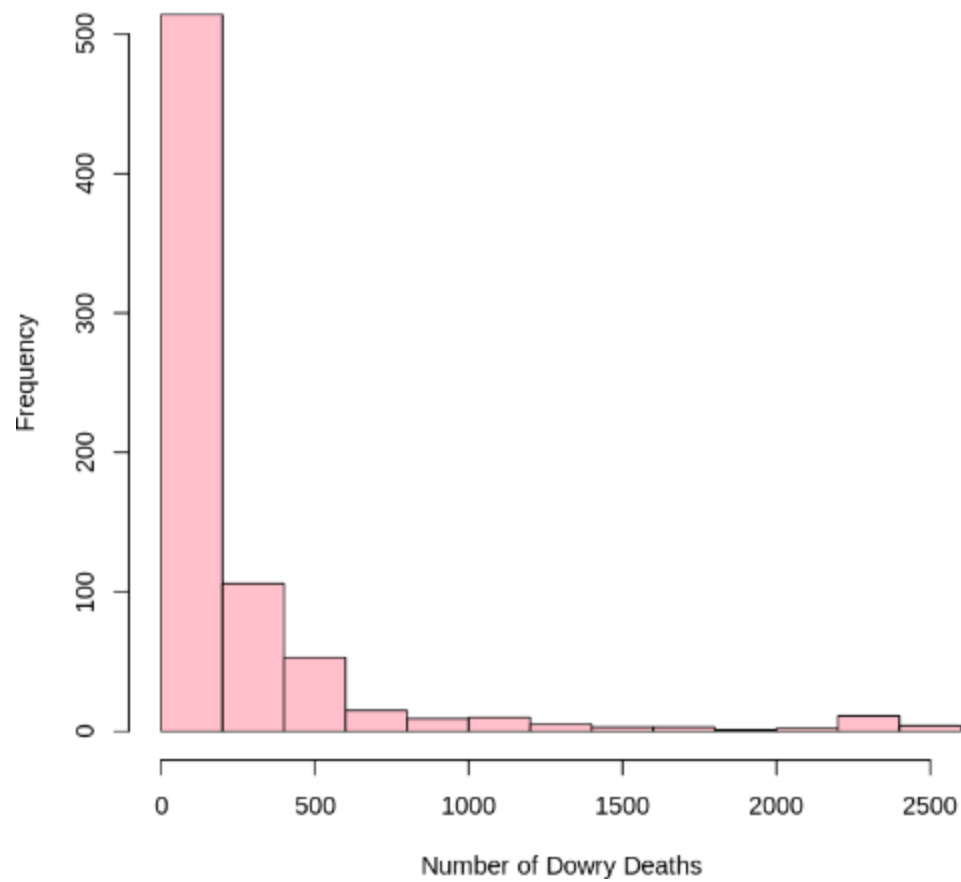
4. Line chart:

```
# Line plot for rape cases over the years in a state
plot(data$Year, data$Rape, type='o', col='blue', main="Trend of Rape Cases Over Time", xlab="Year", ylab="Rape Cases")
```

Trend of Rape Cases Over Time

This plot will display the trend of rape cases over time for a particular state, showing whether the number of cases has increased, decreased, or remained stable over the years.

We can observe that the rape cases are constantly increasing.

5. Scatter plot:

```
# Scatter plot for Rape vs Domestic Violence
plot(data$Rape, data$DV, main="Rape vs Domestic Violence", xlab="Rape Cases", ylab="Domestic Violence Cases", col="red")
```

Rape vs Domestic Violence
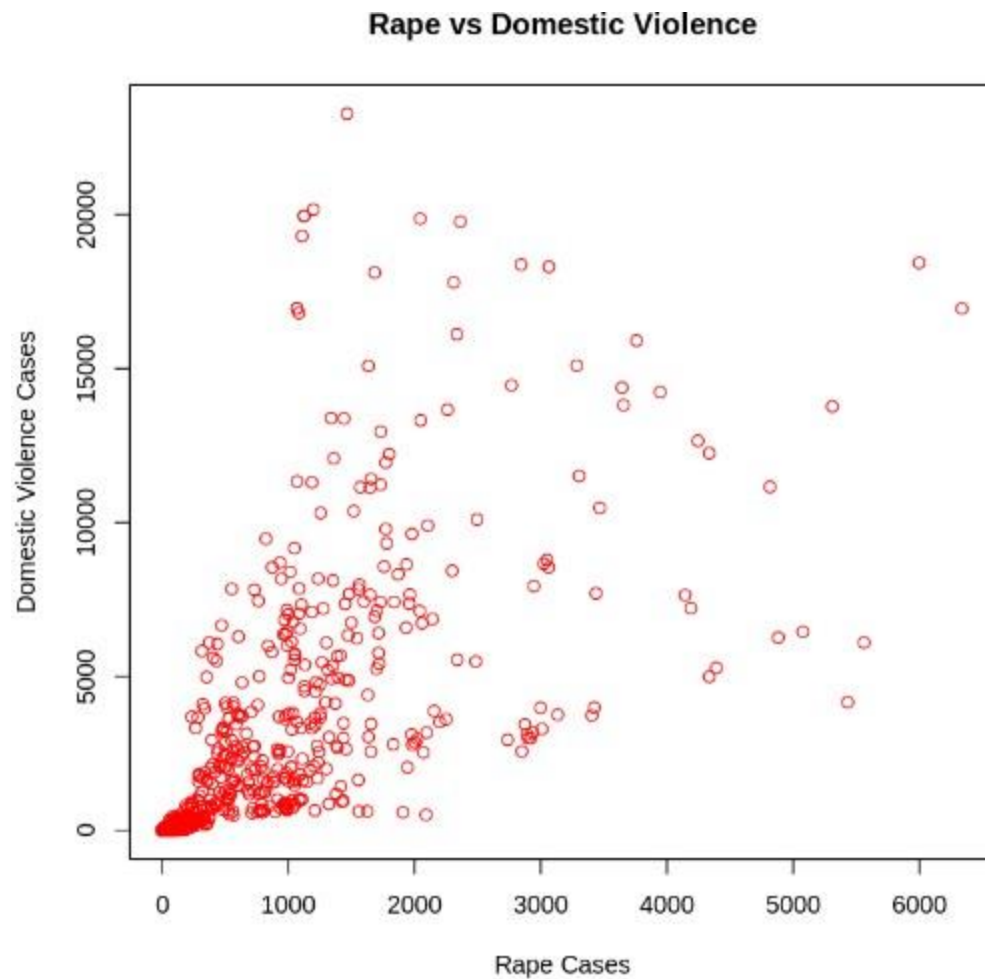
This plot will help you understand if there is a correlation between rape and domestic violence cases across states.

6. **Bubble plot:**

```
# Bubble plot for Rape vs Kidnap & Assault with bubble size based on Dowry Deaths
symbols(data$Rape, data$K_A, circles=data$DD, inches=0.5, fg="white", bg="lightgreen", xlab="Rape Cases", ylab="Kidnap & Assault Cases",
```

Bubble Plot: Rape vs K&A with Dowry Deaths
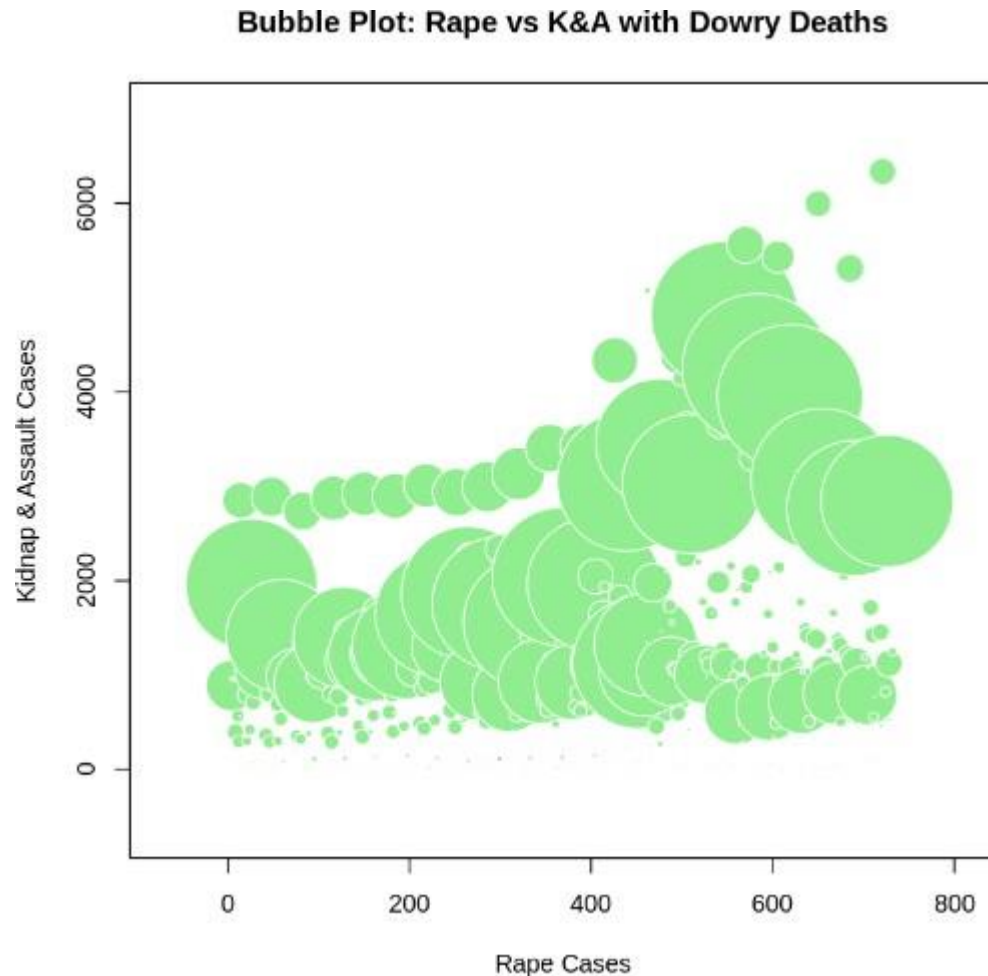
This plot show the relationship between rape cases and kidnap and assault cases, with the size of the bubbles indicating the number of dowry deaths. It helps to visualize multiple variables at once.

Outcomes:

- 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.

Conclusion:

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.