Bharatiya Vidya Bhavan's
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

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

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.

Dataset:
Link: https://www.kaggle.com/datasets/rajanand/crime-in-india

This dataset contains complete information about various aspects of crimes happened in India from 2001.

Property Stolen & Recovered (Crime Head)

Area_Name

Year

Group_Name

Sub_Group_Name

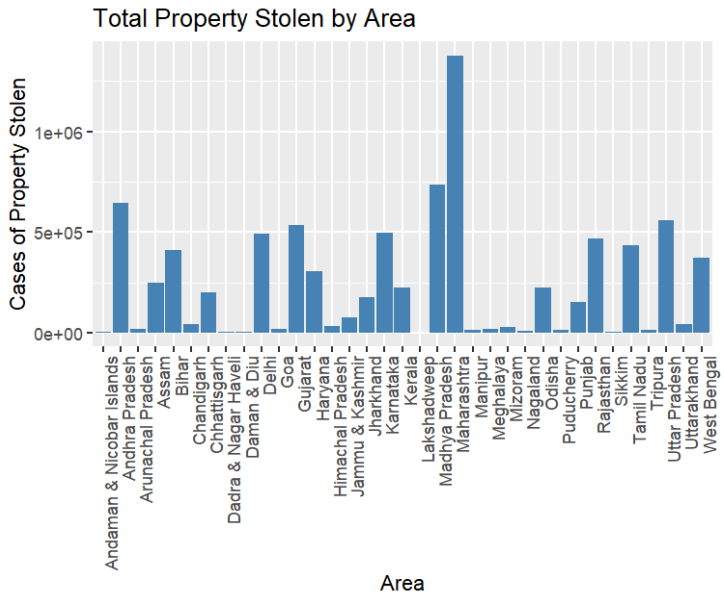
Cases_Property_Recovered

Cases_Property_Stolen

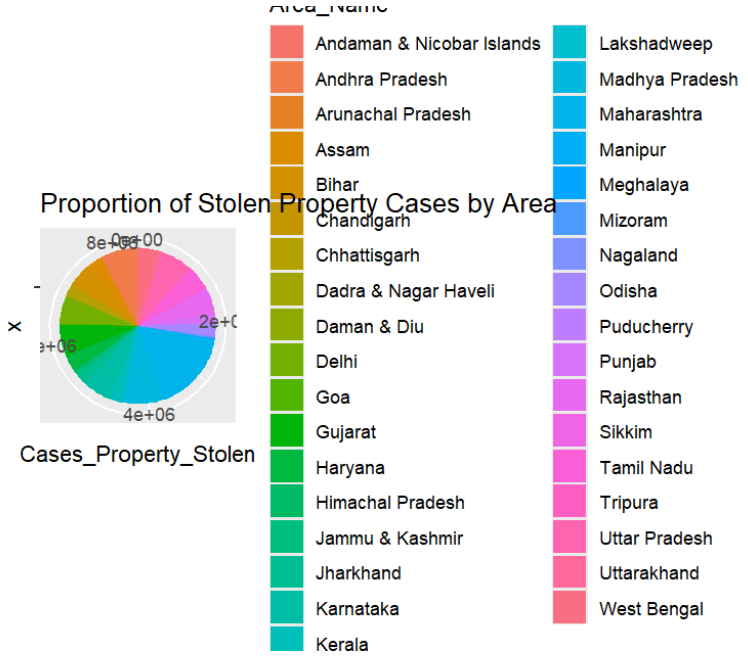
Value_of_Property_Recovered

Value_of_Property_Stolen

Graphs:



1. Wide Variation Across States: There's a significant difference in property theft rates among Indian states. Some states, such as Maharashtra, have significantly higher numbers of cases compared to others.
2. Maharashtra Leads: Maharashtra stands out as having the highest number of reported property theft cases, surpassing all other states by a considerable margin.
3. Southern States: States in the southern region of India, like Kerala and Tamil Nadu, generally have lower reported property theft rates compared to northern and western states.
4. Northeastern States: The northeastern states, including Assam, Arunachal Pradesh, and Manipur, also exhibit relatively lower numbers of property theft cases.
5. Small Territories: The smaller union territories like Andaman & Nicobar Islands, Lakshadweep, and Puducherry have significantly lower reported cases.



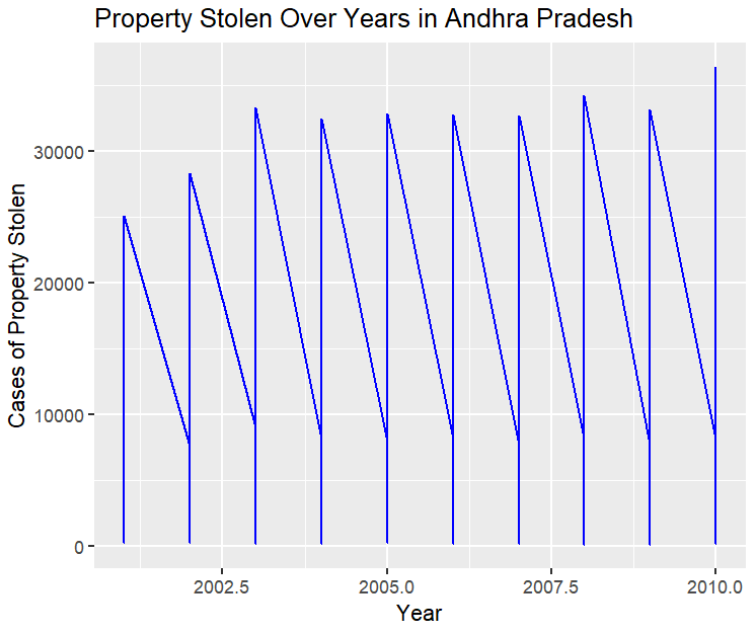
Key Observations:

Maharashtra Dominates: Maharashtra accounts for a significantly larger proportion of stolen property cases compared to other states, indicating a disproportionate concentration of property theft in this region.

Southern States: States in the southern region of India, such as Kerala, Tamil Nadu, and Andhra Pradesh, generally have lower proportions of stolen property cases.

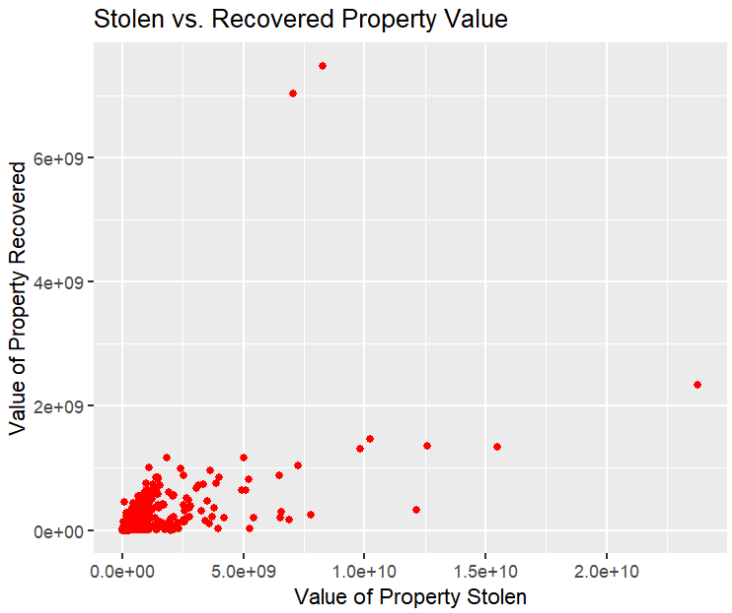
Northeastern States: The northeastern states, including Assam, Arunachal Pradesh, and Manipur, also exhibit relatively lower proportions.

Small Territories: The smaller union territories like Andaman & Nicobar Islands, Lakshadweep, and Puducherry have significantly lower proportions of stolen property cases.



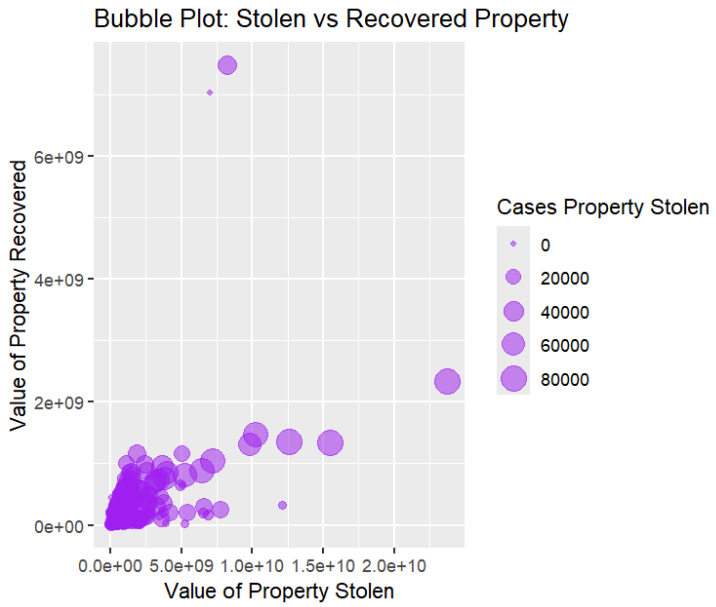
Key Observations:

1. Fluctuating Trends: The graph shows a clear pattern of fluctuations in the number of property theft cases reported in Andhra Pradesh over the years.
2. Periodic Peaks: There appear to be periodic peaks in the number of cases, occurring roughly every two years.
3. Overall Decrease: Despite the fluctuations, there seems to be a general trend of decreasing property theft cases over the decade depicted in the graph.
4. Sharp Drops: The peaks are followed by sharp drops in the number of cases, suggesting potential interventions or campaigns that have been effective in reducing property theft temporarily.



Key Observations:

1. Positive Correlation: There appears to be a positive correlation between the value of stolen property and the value of recovered property. This suggests that, in general, larger thefts tend to result in higher value recoveries.
2. Wide Dispersion: While there is a general trend of higher recovery values for higher theft values, there is also a wide dispersion of data points. This indicates that the relationship between stolen and recovered property values is not perfectly linear.
3. Outliers: A few outliers are visible, particularly in the upper right corner of the graph. These represent cases where the value of recovered property is significantly higher than the value of stolen property. This could be due to factors like additional items recovered, or perhaps errors in the data.
4. Clustering: There seems to be a clustering of data points around the lower end of the x-axis, suggesting that a significant proportion of thefts involve relatively low-value items.

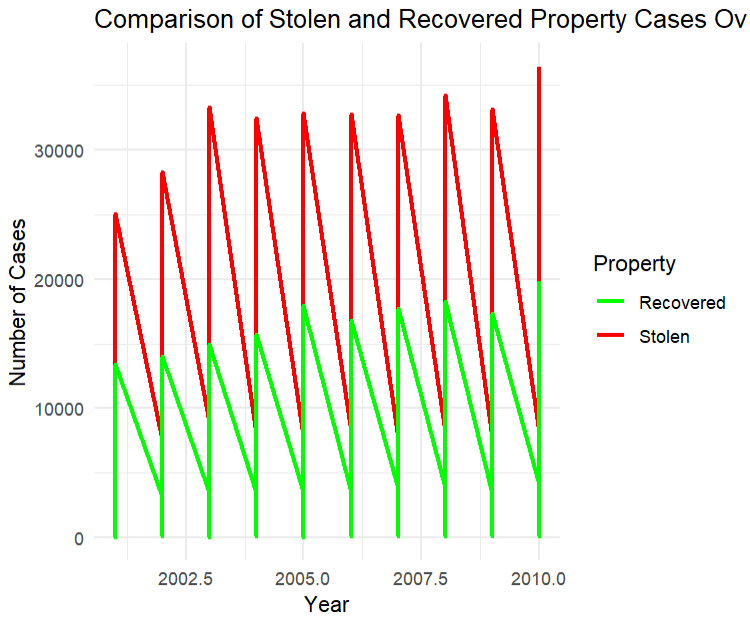


Positive Correlation: There appears to be a positive correlation between the value of stolen property and the value of recovered property. This suggests that, in general, larger thefts tend to result in higher value recoveries.

Bubble Size Indicates Case Count: The size of the bubbles represents the number of cases, providing additional information about the frequency of thefts at different value levels.

Clustering: There is a clustering of data points around the lower end of the x-axis, indicating that a significant proportion of thefts involve relatively low-value items.

Outliers: A few outliers are visible, particularly in the upper right corner of the graph. These represent cases where the value of recovered property is significantly higher than the value of stolen property, possibly due to additional items recovered or data errors.



Key Observations:

1. **Fluctuating Trends:** Both stolen and recovered property cases exhibit a pattern of fluctuations over the years.
2. **Periodic Peaks:** There appear to be periodic peaks in the number of both stolen and recovered cases, occurring roughly every two years.
3. **Similar Patterns:** The patterns of stolen and recovered cases are relatively similar, suggesting a strong correlation between the two.
4. **Increasing Recovery Rates:** While both stolen and recovered cases fluctuate, the number of recovered cases seems to be increasing relative to the number of stolen cases over the years.

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 conclusion.