



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

Example

<https://app.powerbigov.us/view?r=eyJrJlJoiYmU4MDhiYWItYjEwOS00ZDg5LTk1OTUtNzNIMmU0MDFjNTk5liwidCI6IjI5MzU3MDIILWMxMGMtNDgwOS1hMzAyLTg1MmQzNjlmODcwMCMJ9>

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.

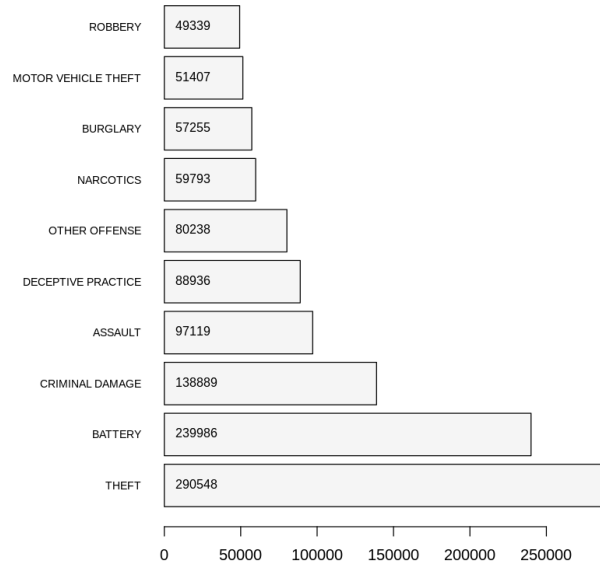


Bharatiya Vidya Bhavan's

Sardar Patel Institute of Technology

(Autonomous Institute Affiliated to University of Mumbai)

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India



Observations

1. Overall Distribution:

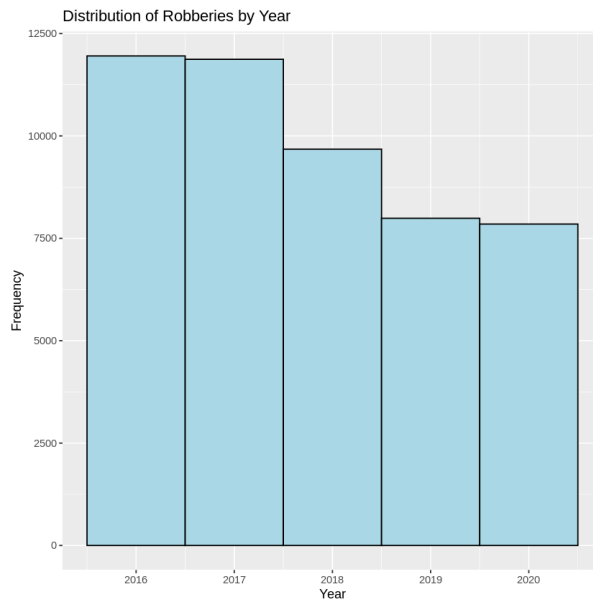
- The bar plot visualizes the top 10 crime types based on their frequency. The `counts` variable represents the number of occurrences for each `Primary.Type`, with the plot showing the most common types.

2. Top Crime Types:

- The bars are sorted in descending order of frequency, allowing for easy identification of the most prevalent crime types. This arrangement helps to highlight which crimes are most frequently reported.

3. Crime Counts:

- The numbers displayed next to each bar indicate the exact count of incidents for the top 10 crime types. This provides a clear view of how many times each crime type was reported.



Observations

1. Overall Distribution:

- The histogram shows the distribution of robbery incidents over the years, specifically for incidents where the latitude is greater than 41. Each bar represents the number of robbery incidents reported in a particular year.

2. Trend Over Time:

- By examining the histogram, you can identify trends in the frequency of robberies over time. Look for increases or decreases in the number of robberies year by year.

3. Robbery Trends:

- Significant peaks in certain years might indicate periods with higher robbery rates, while troughs may represent years with fewer incidents. This can provide insights into trends or patterns in robbery occurrences.

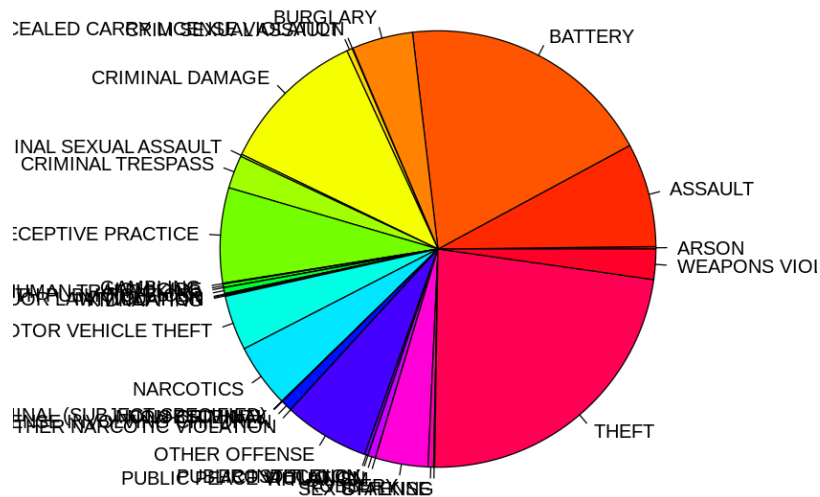
4. Data Range:

- The plot only includes data for years where the latitude is greater than 41, so the results reflect only this subset of the data. Ensure that this latitude constraint aligns with the focus of your analysis.

5. Histogram Details:

- The bin width is set to 1 year, which provides a detailed view of annual changes in robbery frequency. Adjusting the bin width can smooth the data or highlight broader trends if needed.

Proportion of Crime Types



Observations

1. Overall Distribution:

- The pie chart displays the proportion of different crime types within the dataset. Each slice of the pie represents a different crime type, with the size of each slice corresponding to the proportion of that crime type relative to the total number of crimes.

2. Crime Type Proportions:

- By looking at the chart, you can quickly identify which crime types are most and least frequent. Larger slices indicate more prevalent crime types, while smaller slices represent less common ones.

3. Color Coding:

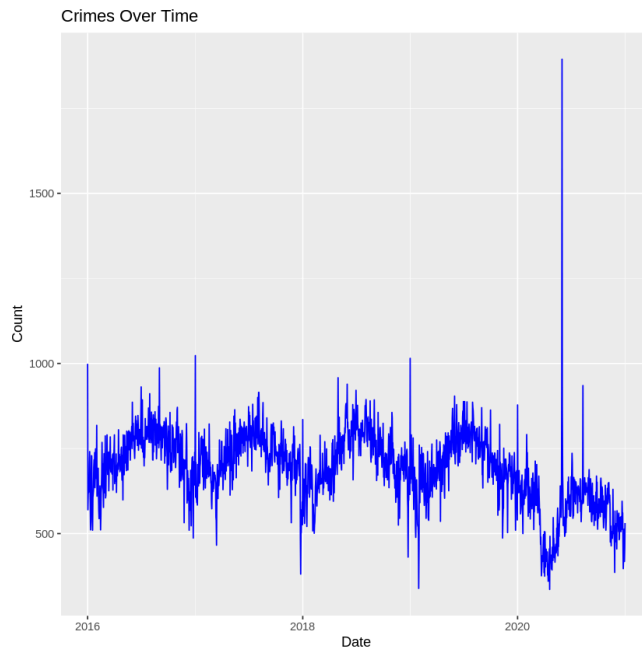
- The use of different colors for each crime type helps in distinguishing between them. The `rainbow()` function provides a range of colors, but be cautious of color-blind accessibility and ensure that the colors are distinct enough to be differentiated easily.

4. Visual Clarity:

- While pie charts are useful for showing proportions, they can become cluttered with too many categories. If there are many crime types, the chart may become difficult to interpret. Grouping smaller categories or using a different type of visualization could improve clarity.

5. Title and Labels:

- The chart includes a title ("Proportion of Crime Types"), which helps contextualize the data being presented. However, for detailed analysis, consider adding labels or a legend to provide specific information about each slice.



Observations

1. Overall Trend:

- The timeline chart illustrates the trend in the number of reported crimes over time. The x-axis represents dates, and the y-axis shows the count of crimes reported on those dates.

2. Temporal Patterns:

- By examining the line plot, you can observe fluctuations in crime frequency over time. Look for trends such as increases or decreases in crime rates, which could indicate seasonal variations, long-term trends, or responses to external factors.

3. Granularity of Data:

- The chart uses `stat = "count"` to aggregate the number of crimes reported on each date. This approach provides a clear view of daily or aggregated crime counts, but be aware that if there are many data points, the chart might become cluttered. Adjusting the aggregation level or plotting data by week or month might provide a clearer trend.

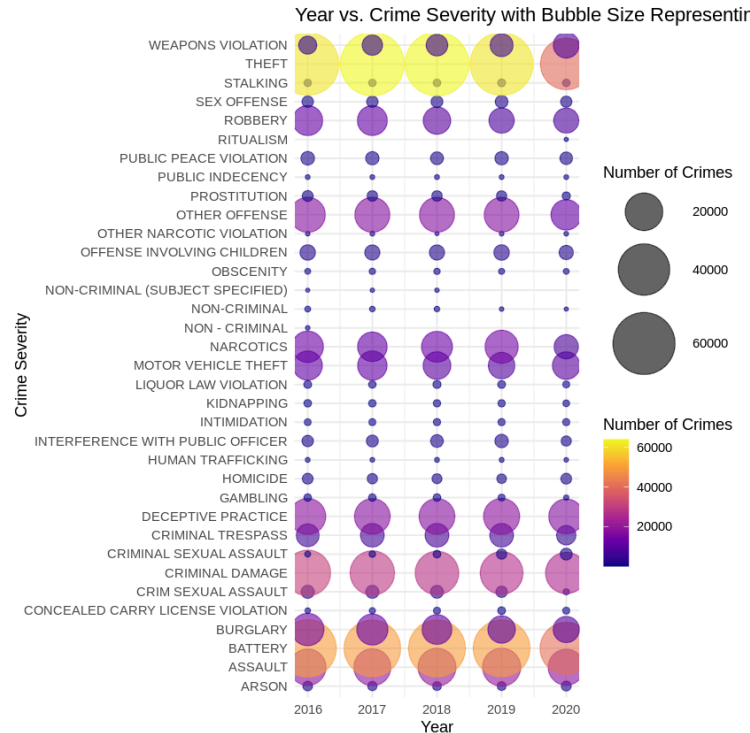
4. Visual Clarity:

- The line is colored blue, which provides a clear and distinguishable visual representation of the trend. Ensure the chart is easy to read by adjusting axis

labels and titles as needed.

5. Date Conversion:

- Converting the **Date** column to the Date type ensures accurate plotting and interpretation of time-series data. The format used ("%m/%d/%Y %H:%M") aligns with the dataset's date-time format.



Observations

1. Overall Distribution:

- The bubble chart displays the relationship between **Year** and **Primary.Type**, with the size of the bubbles representing the number of crimes for each combination. Each bubble represents a specific **Primary.Type** of crime in a given year.

2. Trends Over Time:

- Observe how the bubbles change size and position across different years. An increase in the size of the bubbles over the years for specific crime types indicates a rise in the number of those crimes. Conversely, a decrease suggests a decline.

3. Crime Severity:

- The **Primary.Type** of crime is plotted on the y-axis. Check for any noticeable changes in the size of the bubbles for different crime types. Large bubbles in specific years may indicate higher crime rates for those types.

4. Bubble Size and Color:

- The size of the bubbles reflects the number of crimes reported, with larger bubbles indicating more crimes. The color intensity, ranging from lighter to darker

shades, also represents the number of crimes, with darker shades indicating higher counts.

5. **Patterns and Anomalies:**

- Look for any patterns, such as consistent increases or decreases in specific crime types over the years. Identify any anomalies or years with unusually high or low numbers of crimes for particular **Primary.Type**.

6. **Comparison of Crime Types:**

- Compare the sizes of bubbles across different crime types in the same year. This can help understand which types of crimes are more prevalent and how their prevalence changes over time.

Scatter Plot of Crime Coordinates



Overall Distribution:

- The scatter plot displays the distribution of crime locations based on **X Coordinate** and **Y Coordinate**. Each point represents a crime incident, plotted according to its geographic coordinates.

Spatial Patterns:

- The plot reveals the spatial distribution of crimes across the area covered by the coordinates. Look for any patterns, clusters, or gaps in the distribution of points. Dense areas with many points might indicate high-crime hotspots.

Geographic Clusters:

- Identify clusters where points are concentrated. These clusters could represent areas with higher crime rates or specific locations that might need further investigation.

Distribution Across Coordinates:

- Examine the spread of points along both the **X Coordinate** (longitude) and **Y Coordinate** (latitude) axes. A wide spread suggests a large geographic area covered by the dataset, while a more concentrated spread indicates a smaller, possibly more targeted area.

Outliers:

- Check for any points that are far removed from the main cluster. These outliers might represent isolated incidents or data errors that could warrant further investigation.

Patterns of Crime Locations:

- Look for any visible patterns, such as alignment with streets, neighborhoods, or other geographic features. This can provide insights into crime distribution related to specific locations or areas.

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.