BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

Name: Vinit Madhyan
UID: 2021300070
Batch: I

Aim:

Create basic charts using Power BI / Tableau / R / Python / D3.js 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

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
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

Dataset:
https://www.kaggle.com/datasets/adwaitpurao/indian-crimes-dataset

Dataset 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

Column Descriptions:

1. **Report Number**: Unique identifier for each crime report.
2. **Date Reported**: The date the crime was reported to authorities.
3. **Date of Occurrence**: The actual date when the crime occurred.
4. **Time of Occurrence**: Time when the crime took place.
5. **City**: The city where the crime occurred.
6. **Crime Code**: Code categorizing the type of crime.
7. **Crime Description**: A brief description of the crime.
8. **Victim Age**: Age of the crime victim.
9. **Victim Gender**: Gender of the crime victim.
10. **Weapon Used**: Weapon involved in the crime.
11. **Crime Domain**: Area or domain where the crime occurred.
12. **Police Deployed**: Number or level of police personnel deployed.
13. **Case Closed**: Indicates whether the case was resolved.

Charts:

1. Bar Chart

A bar chart represents categorical data using rectangular bars with lengths proportional to the values they represent. It's commonly used to compare different categories.

- **Advantages**:
    - Easy to read and interpret for comparing categorical data.
    - Can display both positive and negative values.
- **Limitations**:
    - Becomes cluttered when there are too many categories.
    - May not be ideal for showing trends over time.
- **Chart**:

Top 10 Cities with Highest Crime Count
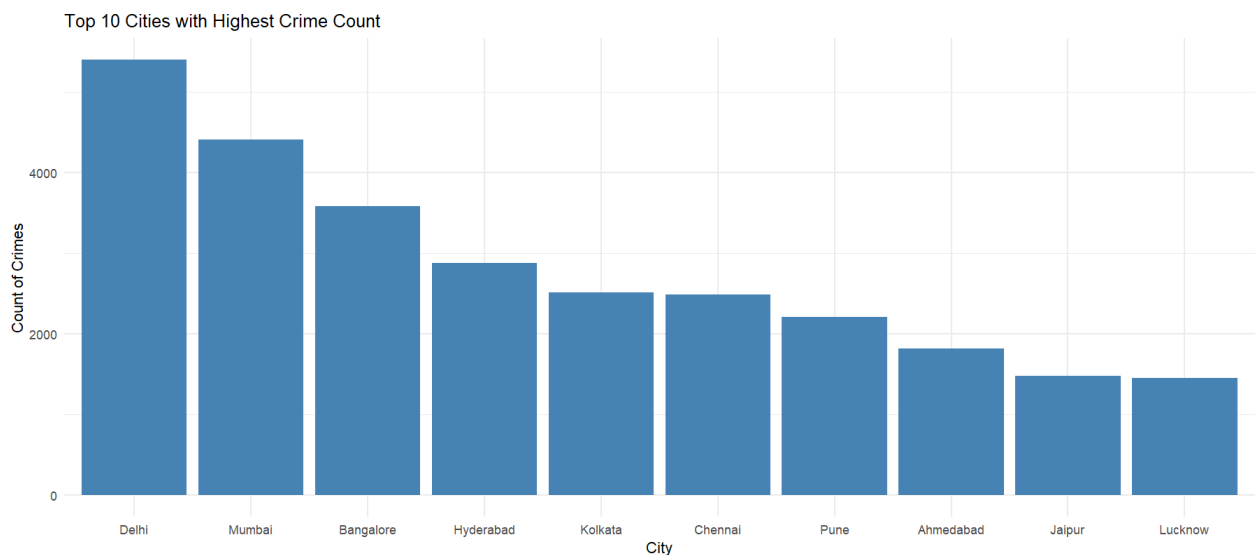
- **Observations**:
    - **Delhi's Crime Dominance**: Delhi has the highest crime count by a notable margin. This could indicate either a larger population, better crime reporting, or potentially higher crime rates compared to other cities.
    - **Tiered Crime Distribution**: There is a clear gap between the top five cities (Delhi, Mumbai, Bangalore, Hyderabad, Kolkata) and the bottom five (Chennai, Pune, Ahmedabad, Jaipur, Lucknow). This suggests that either urbanization or population density might influence the concentration of crimes in larger metropolitan areas.
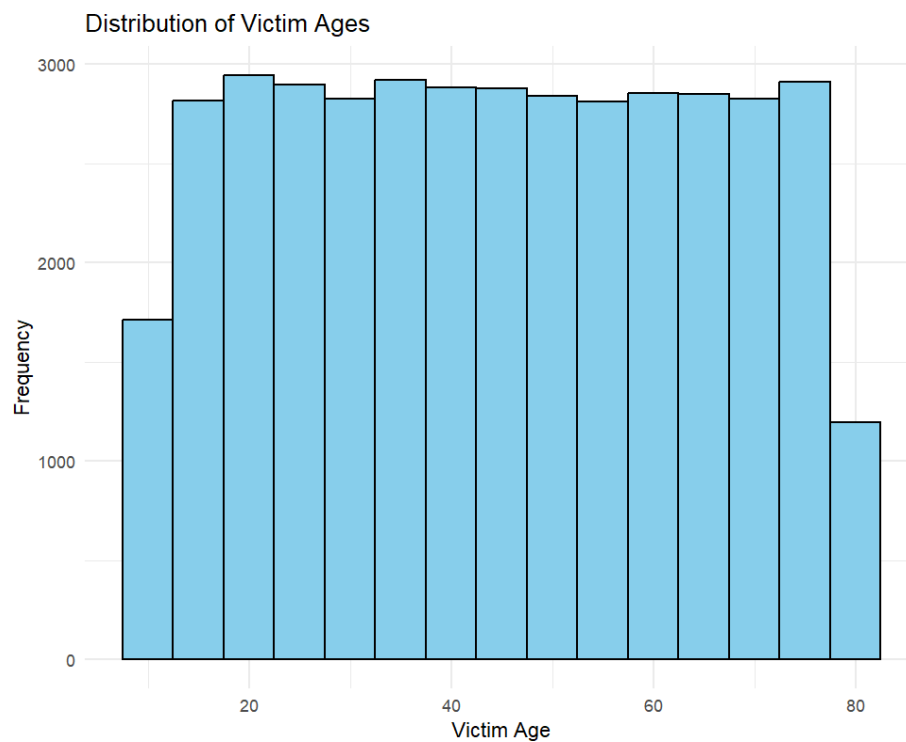
- ○ **Potential for Deeper Insights**: The relatively lower crime counts in cities like Lucknow and Jaipur could suggest either more effective policing, less urban congestion, or underreporting of crimes. It may be worth exploring the reasons behind these differences for a more nuanced understanding of crime patterns.

---

2. Histogram

A histogram displays the distribution of numerical data by grouping data into bins and showing the frequency of data within each bin.

- **Advantages**:
  - ○ Effectively shows the distribution and spread of data.
  - ○ Useful for identifying patterns such as skewness or outliers.
- **Limitations**:
  - ○ Choice of bin size can heavily influence interpretation.
  - ○ Doesn't show exact values, only range distributions.
- **Chart**:

Distribution of Victim Ages

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

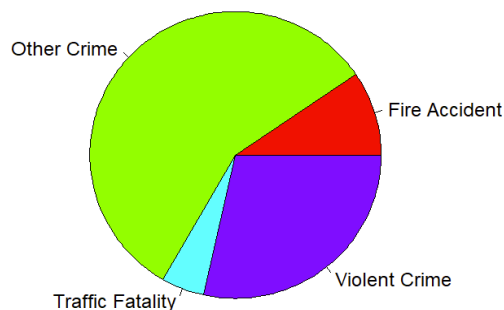**Department of Computer Engineering**

- **Observations**:
  - **Even Distribution Across Most Age Groups**: Victim ages between 20 and 80 show a relatively consistent frequency of incidents, suggesting that crimes do not disproportionately affect any specific age group within this range.
  - **Lower Frequency in Younger and Older Age Groups**: Victims below 20 and above 80 have notably lower frequencies, indicating that these age groups are either less targeted or have fewer opportunities for exposure to situations that lead to victimization.
  - **Peak Crime Ages Around 20 to 40**: The slightly higher frequency of victims in the 20-40 age group could suggest that individuals in this range are more exposed to environments or activities where crimes are more likely, possibly due to greater mobility, work involvement, or social activities.

---

3. Pie Chart

A pie chart displays parts of a whole as slices of a circle, with each slice proportional to its contribution.

- **Advantages**:
  - Simple and effective for representing proportional data.
  - Useful for visualizing percentage breakdowns.
- **Limitations**:
  - Difficult to compare slices of similar sizes.
  - Not effective with too many categories or small data differences.
- **Chart**:

Crime Distribution by Domain

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

- **Observations**:
  - **Dominance of 'Other Crimes'**: The majority of crimes fall under the "Other Crime" category, which suggests a broad spectrum of criminal activity that doesn't neatly fit into more specific categories. This could indicate the need for more refined classification or a focus on crimes that might be less reported or understood.
  - **Significant Presence of Violent Crimes**: Violent crimes make up a considerable portion of the chart, highlighting the serious nature of offenses in this category and the potential need for targeted interventions or law enforcement efforts.
  - **Smaller Proportion for Fire Accidents and Traffic Fatalities**: Both fire accidents and traffic fatalities make up relatively small sections of the chart, which may suggest either fewer incidents in these domains or possibly more effective preventive measures in place compared to other categories.

---

4. Scatter Plot

A scatter plot represents the relationship between two numerical variables using dots, each dot corresponding to a pair of values.

- **Advantages**:
  - Shows relationships and correlations between two variables.
  - Helps in identifying trends, clusters, or outliers.
- **Limitations**:
  - May not show exact trends if the data points overlap.
  - Hard to interpret with large datasets or multiple categories.

- **Chart**:

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

Filtered Scatter Plot: Homicide and Identity Theft (30-50, Male, Mumbai)
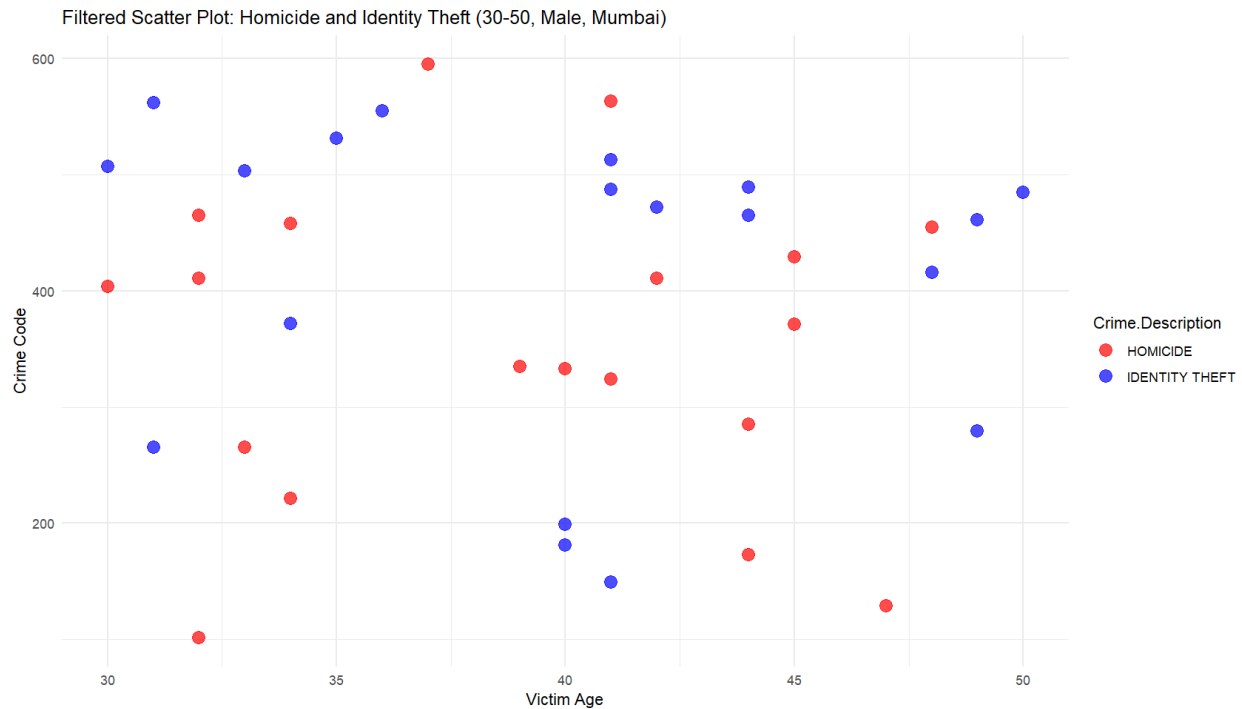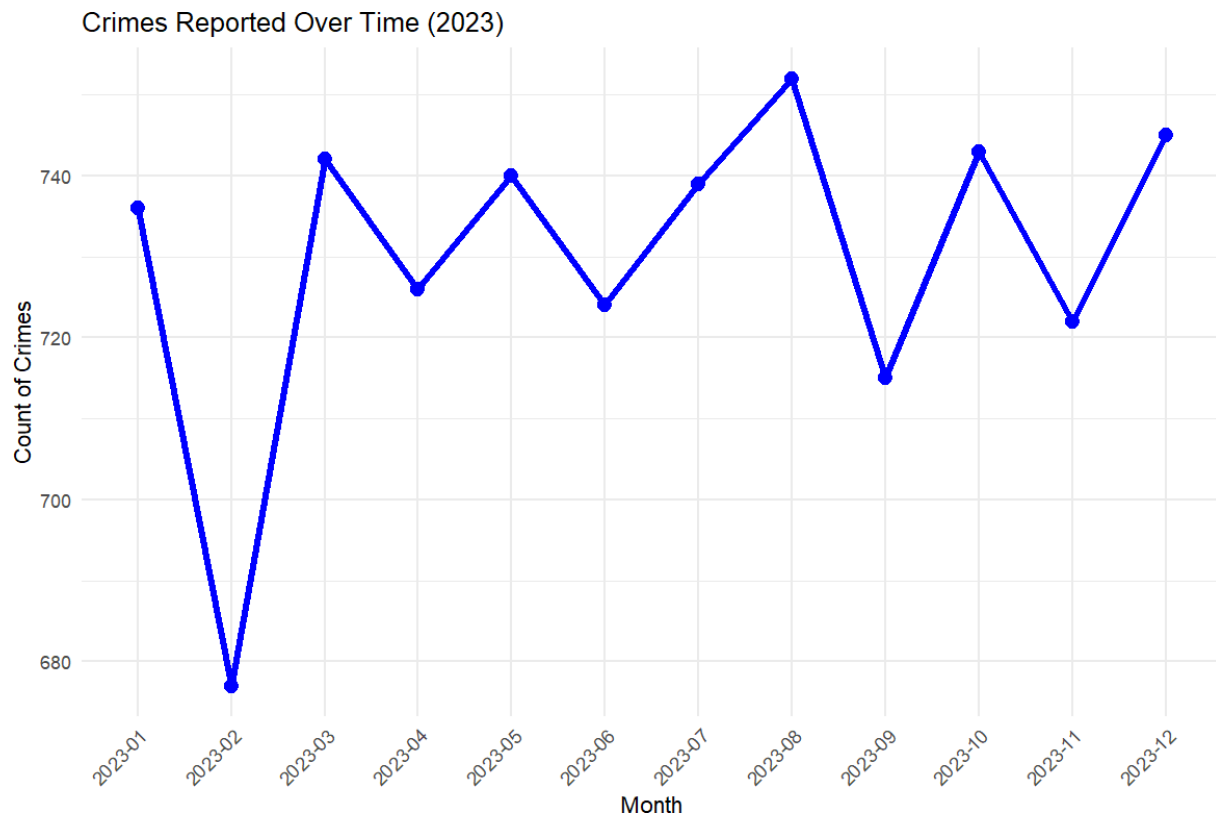
- **Observations**:
  - **Homicide (Red) shows a broader age distribution**: Incidents of homicide occur across a wide range of victim ages (30–50 years), with no clear clustering around specific ages. This suggests that homicide in this demographic group (males in Mumbai) is not significantly age-dependent.
  - **Identity Theft (Blue) tends to cluster around certain crime codes**: Most identity theft incidents are associated with crime codes in the 600 range, indicating a possible concentration of these crimes in higher-numbered crime categories. The victim ages also appear to be somewhat evenly spread across the 30–50 range but are not as frequent as homicide cases.
  - **Homicide incidents are more frequent than Identity Theft**: There are visibly more red dots (homicide) on the plot compared to blue dots (identity theft), indicating that, within the specified data range, homicide is the more frequently reported crime for males in Mumbai aged 30–50.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

5. Time Series Chart

A time series chart displays data points over time, typically using a line to connect the points.

- **Advantages**:
  - Ideal for showing trends, patterns, and cycles over time.
  - Helps in forecasting and predictive analysis.
- **Limitations**:
  - Can become complex with multiple data series.
  - Limited in analyzing non-temporal variables.
- **Chart**:

Crimes Reported Over Time (2023)

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

- **Observations**:
  - **Significant Drop in February**: The count of crimes dropped sharply in February 2023, reaching the lowest point of the year, with fewer than 680 crimes reported. This is a significant outlier compared to the other months, which generally maintain higher crime levels.
  - **Fluctuating Crime Rates Throughout the Year**: The number of crimes fluctuates regularly from month to month. After the February dip, crime rates increase in March, peaking multiple times (around May and October) before falling again. This suggests there is no consistent trend of increasing or decreasing crime but rather cyclical spikes and dips.
  - **Highest Crime Peaks in May and October**: Crime reports hit their highest points in May and October, with both months recording around 750 crimes. This indicates that certain months, possibly due to seasonal or societal factors, see more criminal activity than others.

---

6. Bubble Chart

A bubble chart is a variation of a scatter plot, where the size of the bubble represents a third variable.

- **Advantages**:
  - Displays three variables simultaneously in a clear format.
  - Effective for highlighting correlations and size-based comparisons.
- **Limitations**:
  - Hard to compare bubbles accurately, especially if sizes are similar.
  - May be challenging to interpret for large datasets.
- **Chart**:

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

Crime Distribution by City and Domain

- **Observations**:
  - **Violent Crime**: All four cities (Bangalore, Delhi, Hyderabad, and Mumbai) show a significant number of violent crimes, with the largest bubbles indicating a relatively high crime count across this category in every city.
  - **Other Crime**: Delhi and Mumbai have a notably high count of "Other Crimes" compared to Bangalore and Hyderabad, as indicated by the larger green bubbles. Hyderabad also has a moderate number of other crimes.
  - **Traffic Fatalities and Fire Accidents**: Traffic fatalities and fire accidents are relatively low in all cities, with traffic fatalities being the smallest category overall. Fire accidents occur slightly more frequently than traffic fatalities but are still relatively lower in comparison to other crime categories.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]

Department of Computer Engineering

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:

Through this experiment, I gained significant insights into using R for data visualization. I created bar charts, pie charts, timelines, scatter plots, bubble charts, and histograms to effectively represent various crime categories. I identified trends over time, visualized correlations between variables such as victim demographics and crime types, and understood crime distribution across cities. This experience enhanced my ability to interpret crime data, improved my proficiency in R, and highlighted the importance of clear visual communication in crime analysis.