|  |  |
| --- | --- |
| Subject: Data Analytics and Visualization Lab | Course ID: CSL-601 |
| Semester: VI | Course: AI & DS |
| Laboratory: 407 | Name of teacher: Prof. Gitanjali Korgaonkar |
| Name of Student: Ayush Gupta | Roll No: VU2S2324001 |

**EXPERIMENT NO. 3**

**Aim:**

To understand and implement visualization of data.

**Theory:**

Data visualization is the graphical representation of information and data. It allows us to identify patterns, trends, and insights that may not be immediately apparent from raw data. Visualization plays a crucial role in data analysis, enabling better decision-making, storytelling, and communication of findings.

R provides several powerful libraries for data visualization, among which ggplot2 is one of the most widely used. It is based on The Grammar of Graphics, which allows users to create complex visualizations by layering components such as data, aesthetics, geometric objects, and statistical transformations.

**Types of Visualizations in Data Analysis**

1. **Histogram:**

A histogram is used to visualize the distribution of a single numerical variable. It divides data into intervals (or bins) and shows how frequently values occur within each bin. This helps in understanding the shape of the data distribution, detecting skewness, and identifying potential outliers.

1. **Scatter Plot:**

A scatter plot is used to examine the relationship between two numerical variables. Each point in the plot represents an observation, with one variable on the x-axis and the other on the y-axis. Scatter plots help in identifying correlations, trends, and patterns in the data.

1. **Box Plot:**

A box plot (also known as a whisker plot) is a method for visualizing the distribution of numerical data and detecting outliers. It displays the median, quartiles, and extreme values of a dataset, making it useful for comparing distributions across different groups.

1. **Bar Chart:**

A bar chart is used to visualize categorical data by displaying the frequency or count of each category. It helps in understanding the distribution of categorical variables and making comparisons between different categories.

1. **Line Plot:**

A line plot is used to visualize trends in numerical data over time. It connects data points using a continuous line, making it ideal for showing patterns, fluctuations, and changes over a period.

**Importance of Data Visualization**

* **Enhances Data Understanding**: Helps in grasping complex patterns in data more intuitively.
* **Identifies Trends and Outliers**: Detects anomalies and variations that may not be visible in raw data.
* **Supports Decision-Making**: Provides insights that aid in making informed business and research decisions.
* **Improves Data Communication**: Makes data more accessible to a broader audience, including non-technical stakeholders.

By utilizing different types of visualization techniques, we can effectively analyze data and extract meaningful insights. This experiment focuses on implementing various visualizations in R using ggplot2.

**Learning Objectives:**

* Understood the importance of data visualization in data analysis.
* Learned how to create various types of visualizations in R using ggplot2.
* Explored the significance of histograms, scatter plots, box plots, bar charts, and line plots.
* Gained practical experience in interpreting visualized data for better decision-making.

**Conclusion:**

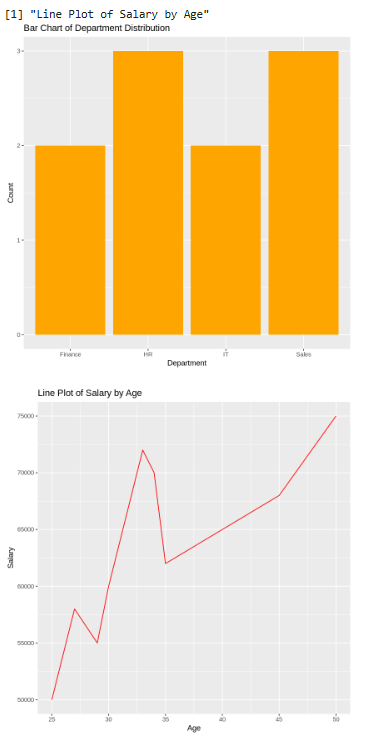
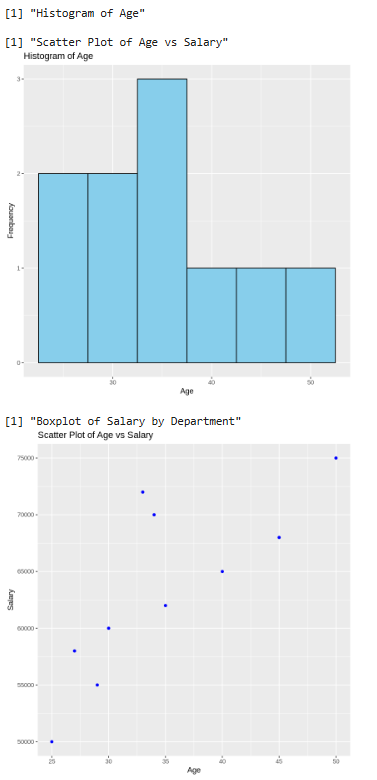
Different data visualization techniques were explored and implemented using ggplot2 in R. Various types of plots, such as histograms, scatter plots, box plots, bar charts, and line plots, were generated to analyze data effectively. The experiment demonstrated how visualization enhances data interpretation, helps in identifying patterns, and simplifies decision-making.

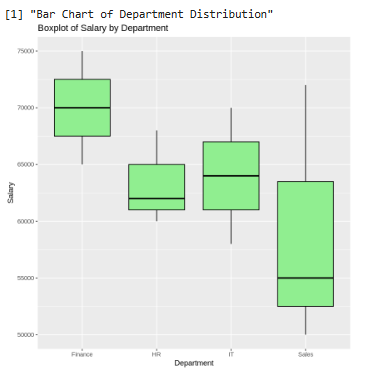


**Program:**

****

**Output:**

****

****