Experiment No 03 **Roll no : B1**

Aim : To implement Data Visualisation technique using R.

Theory :

Data visualisation :

Data visualization uses charts, graphs and maps to present information clearly and simply. It turns complex data into visuals that are easy to understand.
With large amounts of data in every industry, visualization helps spot patterns and trends quickly, leading to faster and smarter decisions.

Common Types of Data Visualization

There are various types of visualizations where each has a unique purpose in data representation. Here are the most common types:

1. **Charts and Graphs:** They are used to visualize data, with charts comparing data points across categories or showing trends over time, and graphs analyzing relationships between variables to identify correlations, trends, and outliers. Examples: Bar Charts, Line Charts, Pie Charts, Scatter Plots, Histograms, Box Plots.
2. **Maps**: They are used to display geographical data which provides spatial context to trends and patterns. Examples: Geographic Maps, Heat Maps
3. **Dashboards:** They combine multiple visualizations into a single interface which provides real-time insights and interactive features for users to explore data.

Importance of Data Visualization

Data visualization is essential for understanding and communicating information effectively. Here are some key reasons why it's important:

1. **Simplifies Complex Data:** It turns large and complicated data into visual formats like charts and graphs, making the information easier to understand.
2. **Reveals Patterns and Trends**: It helps identify trends, relationships, and patterns that are not easily seen in raw data or tables.
3. **Saves Time:** Visuals allow quicker interpretation of data, helping users spot key information at a glance instead of manually scanning through numbers.
4. **Improves Communication:** It makes it easier to explain data insights to others, especially those who may not be familiar with the technical details.
5. **Tells a Clear Story**: Data visuals guide the audience through the information step-by-step, making it easier to reach conclusions and make informed decisions.

Real-World Use Cases for Data Visualization

Data visualization is used across various industries to improve decision-making and drive results. Here are a few examples:

1. **Business Analytics:** Used to monitor company performance, track KPIs, and make data-driven decisions by visualizing trends, sales, and customer metrics.
2. **Healthcare:** Helps in analyzing patient records, tracking disease outbreaks, and managing hospital operations through easy-to-read charts and dashboards.
3. **Sports:** Used to visualize player statistics, team performance, and match outcomes, helping coaches and analysts improve strategies and training plans.
4. **Retail and E-commerce:** Enables tracking of sales, customer preferences, and inventory levels, helping businesses adjust stock and marketing efforts effectively.
4. overwhelm viewers. It's important to focus on key data points and avoid clutter.

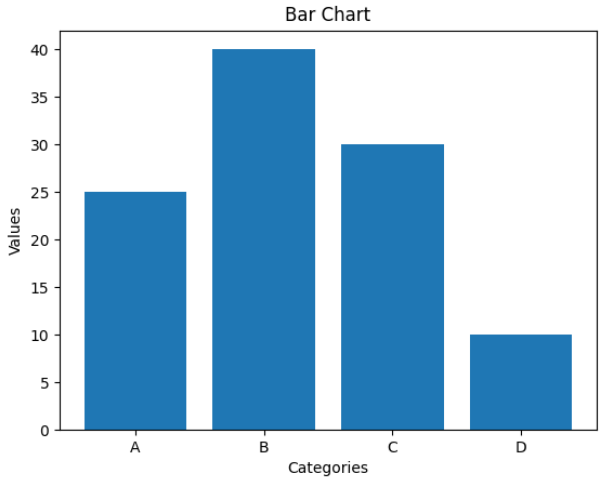
Basic Charts for Data Visualization

Basic charts are best suited for displaying simple comparisons, trends over time and basic relationships within the data. These charts are easy to understand and ideal for communicating insights to a broad audience.

1. Bar Charts

Bar charts are used to compare values across different categories using rectangular bars. X-axis shows categories while Y-axis represents values. Common types include horizontal, stacked and grouped bar charts.

Below is the Example of Bar Chart:



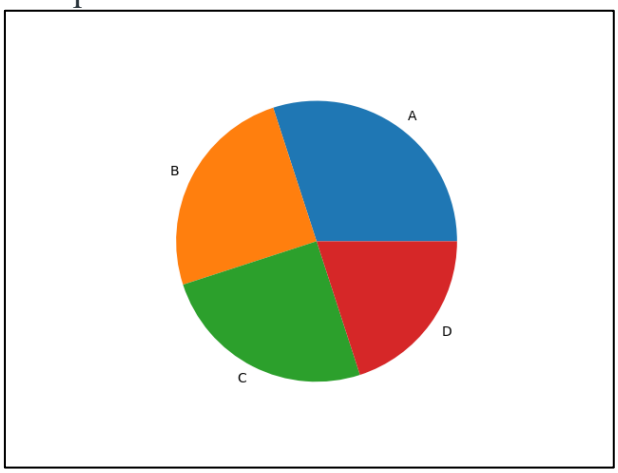
Representation of Bar Chart

When to Use:

- To compare different categories
- To rank values from highest to lowest
- To show relationships between multiple variables

2. Pie charts

Pie chart are round charts divided into slices, where each slice shows a part of the whole. The size of each slice represents its percentage.
Below is the example of pie chart:



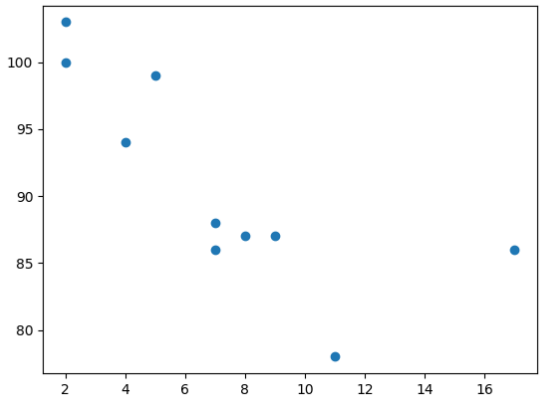
Representation of pie chart

When to Use:

- To show how different parts contribute to a whole
- To highlight a dominant category

3. Scatter Chart (Plots)

Scatter charts use dots to show relationship between two numerical variables. X-axis shows the independent variable and Y-axis shows the dependent variable.
Below is the example of scatter chart:



Representation of Scatter Chart

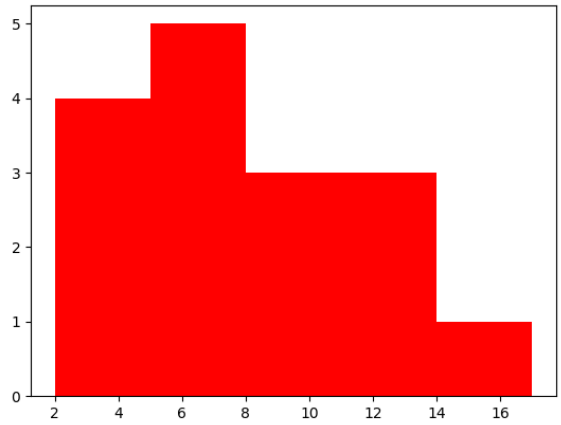
When to Use:

- To observe relationships between two variables
- To detect patterns, clusters or outliers in data

4. Histogram

A histogram displays the distribution of numerical data by grouping values into intervals (bins) and showing their frequency as bars. It helps reveal the shape, spread and patterns in the data.

Below is the example of histogram:



Representation of Histogram

When to Use:

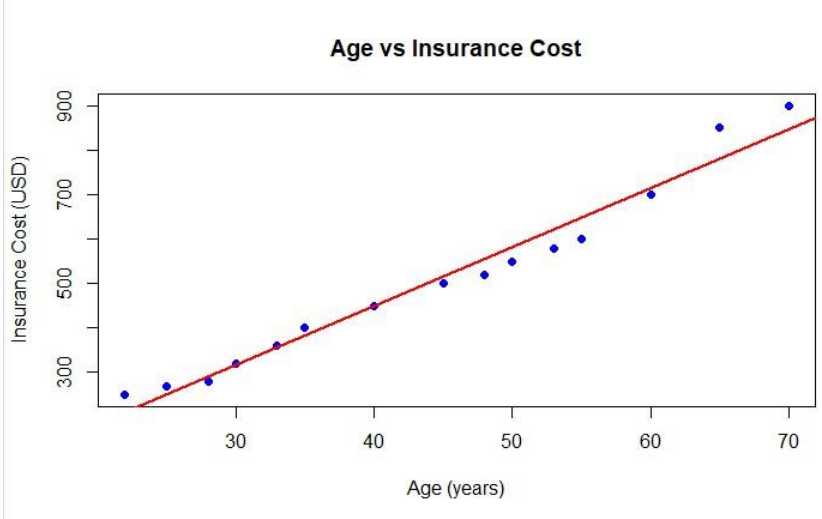
- To visualize the distribution of numerical data
- To explore patterns, trends and outliers

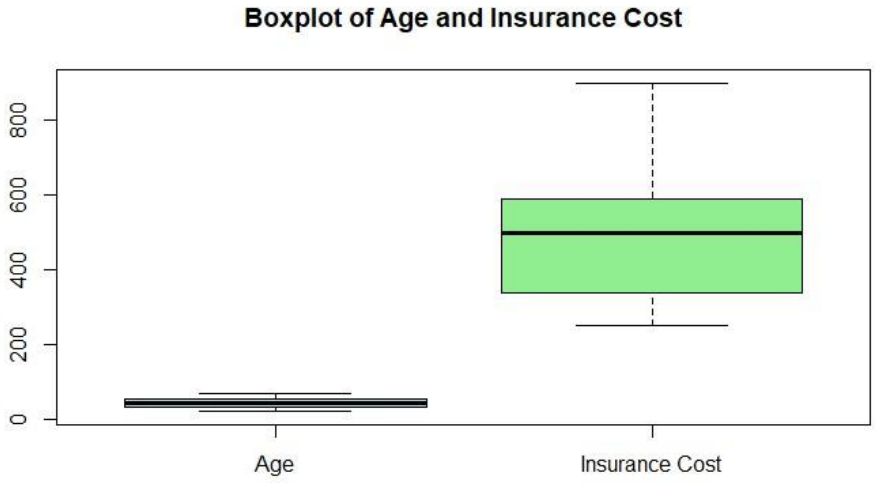
Source :

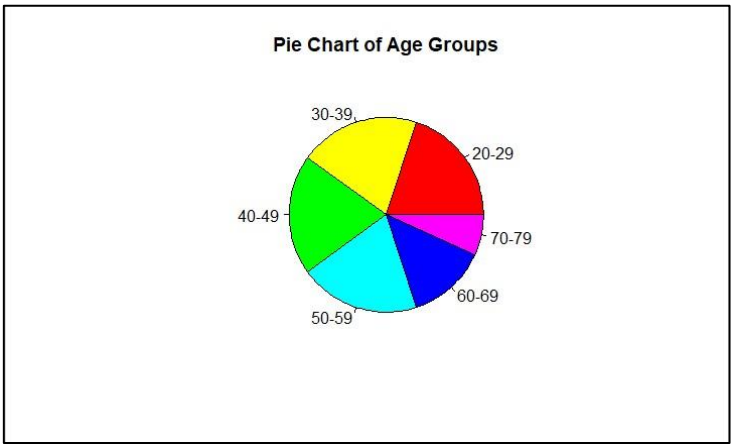
```
# Sample dataset: Age and Insurance Cost for 15 people
> insurance_data <- data.frame(
+     Age = c(22, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 28, 33, 48, 53),
+     Insurance_Cost = c(250, 270, 320, 400, 450, 500, 550, 600, 700, 850, 900, 280, 360, 520,
580)
> print(insurance_data)
   Age Insurance_Cost
1   22            250
2   25            270
3   30            320
4   35            400
5   40            450
6   45            500
7   50            550
8   55            600
9   60            700
10  65            850
11  70            900
12  28            280
13  33            360
14  48            520
15  53            580
> plot(insurance_data$Age, insurance_data$Insurance_Cost,
+      main = "Age vs Insurance Cost",
+      xlab = "Age (years)",
```

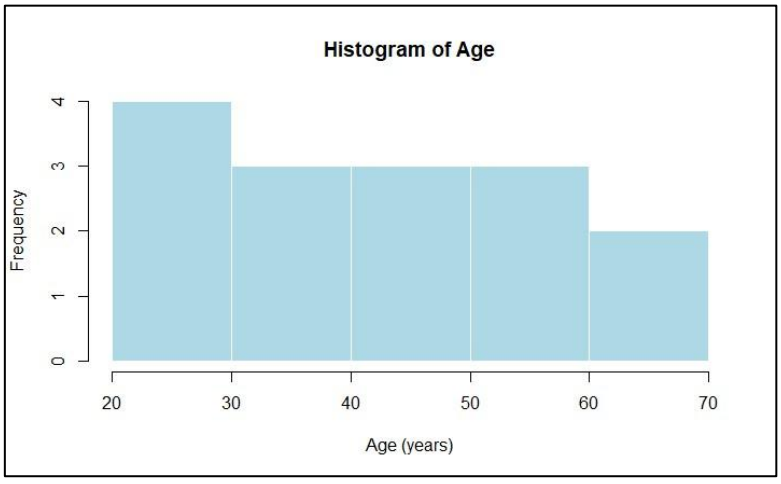
```
+     ylab = "Insurance Cost (USD)",
+     pch = 19, col = "blue")
> abline(lm(Insurance_Cost ~ Age, data = insurance_data), col = "red", lwd = 2)
>
> boxplot(insurance_data$Age, insurance_data$Insurance_Cost,
+         names = c("Age", "Insurance Cost"),
+         main = "Boxplot of Age and Insurance Cost",
+         col = c("lightblue", "lightgreen"))
>
> abline(lm(Insurance_Cost ~ Age, data = insurance_data), col = "red", lwd = 2)
> hist(insurance_data$Age,
+     main = "Histogram of Age",
+     xlab = "Age (years)",
+     col = "lightblue",
+     border = "white")
>
> # Define age groups
> age_groups <- cut(insurance_data$Age,
+                   breaks = c(20, 30, 40, 50, 60, 70, 80),
+                   labels = c("20-29", "30-39", "40-49", "50-59", "60-69", "70-79"),
+                   right = FALSE)
> # Count of each age group
> age_group_table <- table(age_groups)
>
> # Pie chart
> pie(age_group_table,
+     main = "Pie Chart of Age Groups",
+     col = rainbow(length(age_group_table)))
```

Output :









Conclusion : Hence we have successfully implemented data visualization techniques using R.