Subject : DWM   **Year/ Sem: TE/ V**

Experiment Number: 5

Aim: Installation & study of R Programming (data mining tool) and introduction about basic R programming syntax and implementation of Charts and Graphs using R Programming

Problem Statement: To draw charts and graphs for the given dataset using R.

Objective:

The objective of this experiment is to learn numerous libraries to create charts and graphs in R Programming.

Theory:

Installation: Install R Base and R Studio on Ubuntu

Step 1: Install R

 sudo apt-get update

sudo apt-get install r-base

Step 2: Open R Type capital R on Terminal to come out from R type q()

Step 3: Install R Studio go to website www.rstudio.com/products/RStudio/#Desktop

download R Studio for Ubuntu 32 bit and install it by using Ubuntu software centre to open IDE type RStudio on Terminal

- R is a programming language and software environment for statistical analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand. R is freely available under the GNU General Public License This programming language was named R, based on the first letter of first name of the two R authors (Robert Gentleman and Ross Ihaka).

Features of R Programming

- It is a simple and effective programming language which has been well developed.
- It is data analysis software.
- It is a well-designed, easy, and effective language which has the concepts of user-defined, looping, conditional, and various I/O facilities.
- It has a consistent and incorporated set of tools which are used for data analysis.

- For different types of calculation on arrays, lists and vectors, R contains a suite of operators.
- It provides effective data handling and storage facility.
- It is an open-source, powerful, and highly extensible software.
- It provides highly extensible graphical techniques.
- It allows us to perform multiple calculations using vectors.
- R is an interpreted language.

R - Data Types

In contrast to other programming languages like C and java in R, the variables are not declared as some data type. The variables are assigned with R-Objects and the data type of the R-object becomes the data type of the variable. There are many types of R-objects.

The frequently used ones are –

- Vectors
- Lists
- Matrices
- Arrays
- Factors
- Data Frames

R - Operators

We have the following types of operators in R programming –

- Arithmetic Operators
- Relational Operators
- Logical Operators

1. Pie-chart

    A pie-chart is a representation of values as slices of a circle with different colors. The slices are labelled and the numbers corresponding to each slice is also represented in the chart.

    In R the pie chart is created using the pie () function which takes positive numbers as a vector input. The additional parameters are used to control labels, colour, title etc.

2. Bar Chart:

A bar chart represents data in rectangular bars with length of the bar proportional to the value of the variable. R uses the function barplot() to create bar charts. R can draw both vertical and horizontal bars in the bar chart. In bar chart each of the bars can be given different colours.

3. Boxplots:

Boxplots are a measure of how well distributed is the data in a data set. It divides the data set into three quartiles. This graph represents the minimum, maximum, median, first quartile and third quartile in the data set. It is also useful in comparing the distribution of data across data sets by drawing boxplots for each of them.Boxplots are created in R by using the boxplot() function.

4. Histogram:

A histogram represents the frequencies of values of a variable bucketed into ranges. Histogram is similar to bar chat but the difference is it groups the values into continuous ranges. Each bar in histogram represents the height of the number of values present in that range. R creates histogram using hist () function. This function takes a vector as an input and uses some more parameters to plot histograms.

5. Line chart:

A line chart is a graph that connects a series of points by drawing line segments between them. These points are ordered in one of their coordinate (usually the x-coordinate) value. Line charts are usually used in identifying the trends in data. The plot() function in R is used to create the line graph.

6. Scatterplots:

Scatterplots show many points plotted in the Cartesian plane. Each point represents the values of two variables. One variable is chosen in the horizontal axis and another in the vertical axis. The simple scatterplot is created using the plot() function.

Conclusion: - Hence studied basics of R programming.