Name :- Manvi Pandya
Class:- TE10 (L-10)
Roll No. 33235

Assignment-8

- **Aim:**

  Visualize the data using R/Python by plotting the graphs for Air quality dataset and Iris Dataset.

- **Objective:**
  1) To understand and apply the analytical concept of Big data using R/Python.
  2) To study detailed data visualization techniques in R programming.

- **SOFTWARE REQUIREMENTS:**

  1. Ubuntu 16.04 / 18.04
  2. R Studio

- **Theory :-**

  1) R – Pie Charts

  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, color, title etc.

  Syntax :
  The basic syntax for creating a pie-chart using the R is –
  pie(x, labels, radius, main, col, clockwise)
  Following is the description of the parameters used –

  1. x is a vector containing the numeric values used in the pie chart.
  2. labels is used to give description to the slices.
  3. radius indicates the radius of the circle of the pie chart.(value

between −1 and +1).
4. main indicates the title of the chart.
5. col indicates the color palette.
6. clockwise is a logical value indicating if the slices are drawn clockwise or anti clockwise.

Example

A very simple pie-chart is created using just the input vector and labels.

The below script will create and save the pie chart in the current R working directory.

```
# Create data for the graph.
X <- c(21, 62, 10, 53)
labels <- c(“London”, “New York”, “Singapore”, “Mumbai”)

# Give the chart file a name.
png(file = “city.jpg”)
# Plot the chart.
Pie(x,labels)
# Save the file.
Dev.off()
```

When we execute the above code, it produces the following result –

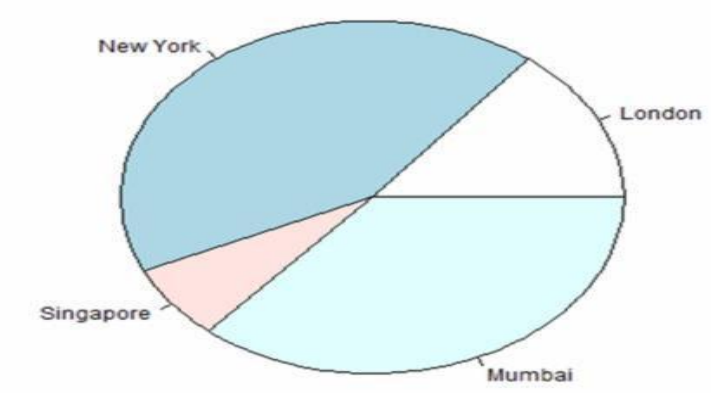

Fig 1. Pie Chart

2) **R – Bar Charts :-**

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 colors.

Syntax

The basic syntax to create a bar-chart in R is –

barplot(H, xlab, ylab, main, names.arg, col)

Following is the description of the parameters used –

1. H is a vector or matrix containing numeric values used in bar chart.
2. xlab is the label for x axis.
3. ylab is the label for y axis.

Example :-

A simple bar chart is created using just the input vector and the name of each bar.

The below script will create and save the bar chart in the current R working directory.

```
# Create the data for the chart.
H <- c(7,12,28,3,41)

# Give the chart file a name.
png(file = “barchart.png”)
# Plot the bar chart.
Barplot(H)
# Save the file.
Dev.off()
```

When we execute the above code, it produces the following result –

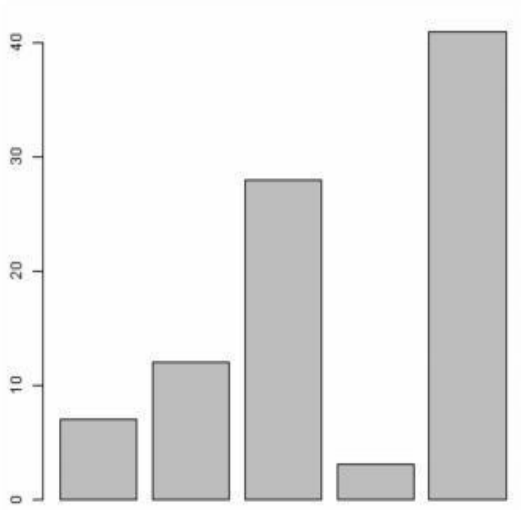

Fig. 2. Bar Chart

3) R – 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.

Syntax

The basic syntax to create a boxplot in R is –

boxplot(x, data, notch, varwidth, names, main)

Following is the description of the parameters used –

1. x is a vector or a formula.
2. data is the data frame.
3. notch is a logical value. Set as TRUE to draw a notch.
4. varwidth is a logical value. Set as true to draw width of the box proportionate to the sample size.

5.names are the group labels which will be printed under each boxplot.
6.main is used to give a title to the graph.

Example
We use the data set “mtcars” available in the R environment to create a basic boxplot. Let’s look
at the columns “mpg” and “cyl” in mtcars.

```
Input <- mtcars[,c(‘mpg’,’cyl’)]
print(head(input))
```

When we execute above code, it produces following result –

	mpg	cyl
Mazda RX4	21.0	6
Mazda RX4 Wag	21.0	6
Datsun 710	22.8	4
Hornet 4 Drive	21.4	6
Hornet Sportabout	18.7	8
Valiant	18.1	6

Creating the Boxplot :-

The below script will create a boxplot graph for the relation between mpg (miles per gallon) and cyl (number of cylinders).

```
# Give the chart file a name.
png(file = “boxplot.png”)

# Plot the chart.
Boxplot(mpg ~ cyl, data = mtcars, xlab = “Number of Cylinders”,ylab =
“Miles Per Gallon”, main = “Mileage Data”)

# Save the file.
Dev.off()
```

When we execute the above code, it produces the following result –

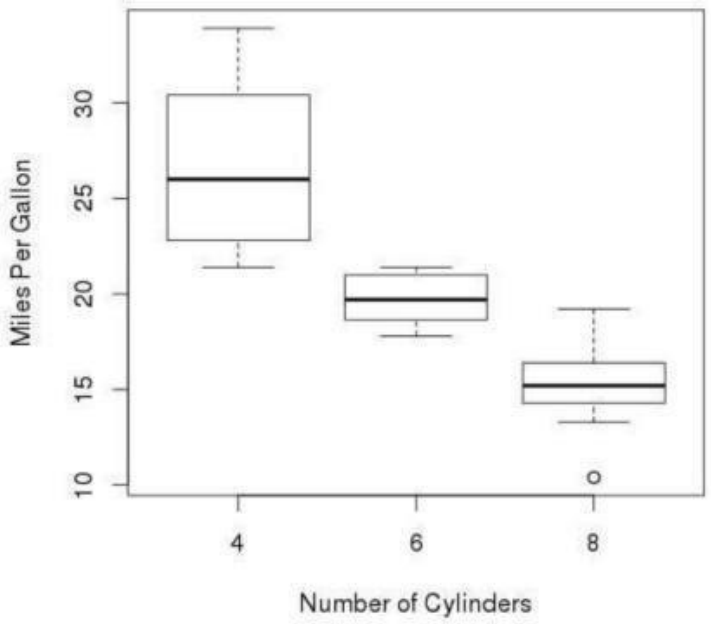

Fig. 3. Box Plot

4) **Scatter plot Matrices:-**

When we have more than two variables and we want to find the correlation between one variable versus the remaining ones we use scatterplot matrix. We use pairs() function to create matrices of scatterplots.

Syntax

The basic syntax for creating scatterplot matrices in R is –

pairs(formula, data)

Following is the description of the parameters used –

1.formula represents the series of variables used in pairs.
2.data represents the data set from which the variables will be taken.

Example

Each variable is paired up with each of the remaining variable. A scatterplot is plotted for each pair.

```
# Give the chart file a name.
png(file = “scatterplot_matrices.png”)
```

Plot the matrices between 4 variables giving 12 plots.

One variable with 3 others and total 4 variables.
Pairs(~wt+mpg+disp+cyl,data = mtcars,
main = “Scatterplot Matrix”)

# Save the file.
Dev.off()

When the above code is executed we get the following output.

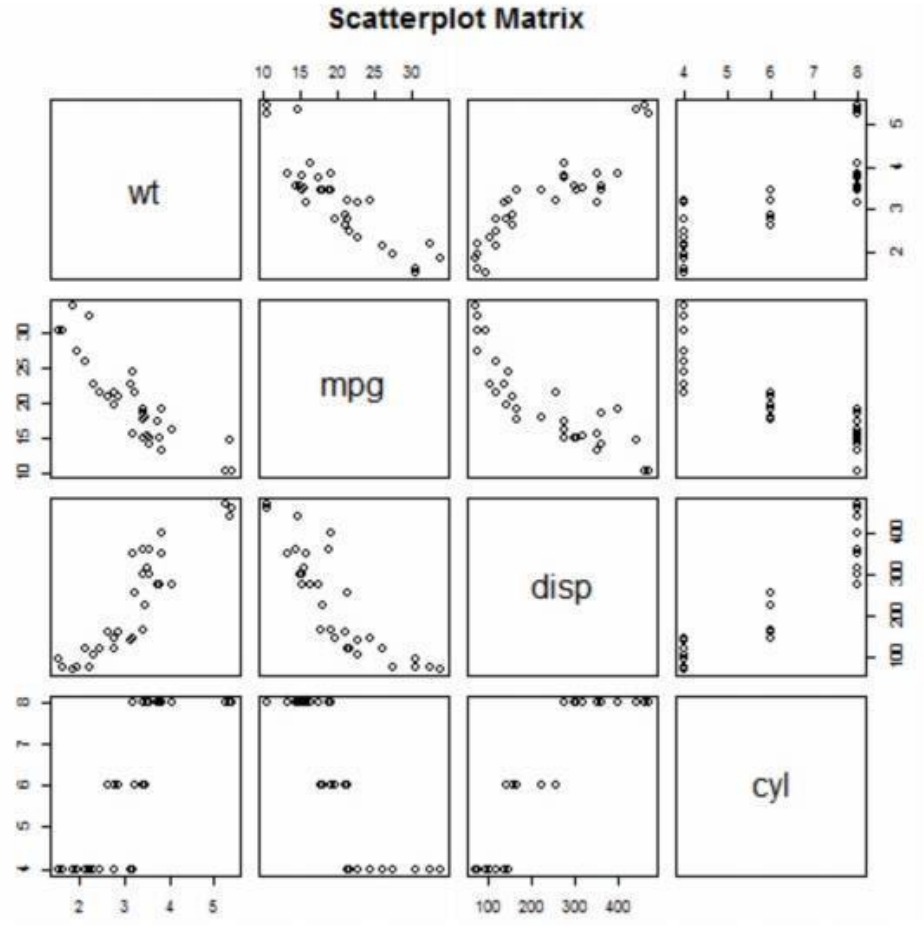

Fig. 7 Scatter Plot Matrix

- **Conclusion:-**

  Thus we have learnt Visualize the data using R/Python by plotting the graphs.