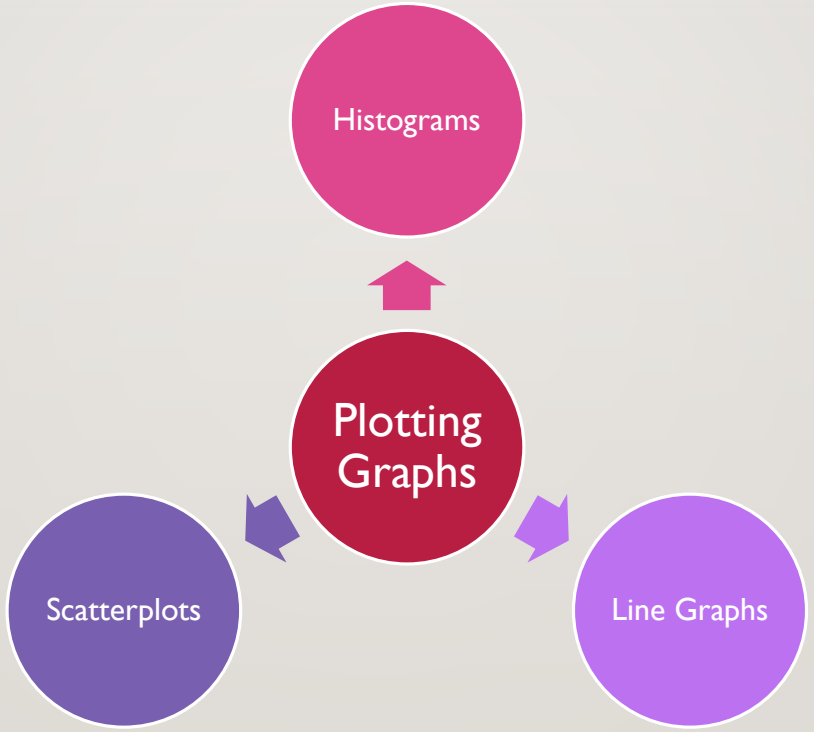
GRAPH PLOTS

KUSHAL SHARMA – DATA SCIENCE PRACTITIONER

AGENDA

- A Quick Revision
- Histograms
- Line Graphs
- Scatterplots
- Demonstration based on dataset in R for statistical concepts

PLOTTING GRAPHS



HISTOGRAM

A histogram represents the frequencies of values of a variable bucketed into ranges. Histogram is similar to bar chart 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.

HISTOGRAM

R creates histogram using hist() function. This function takes a vector as an input and uses some more parameters to plot histograms.

HISTOGRAM – SYNTAX

The basic syntax for creating a histogram using R is:

- hist(v,main,xlab,xlim,ylim,breaks,col,border)

HISTORAM – PARAMETERS USED

v	•**v** is a vector containing numeric values used in histogram.
main	•**main** indicates title of the chart.
col	•**col** is used to set color of the bars.
border	•**border** is used to set border color of each bar.
xlab	•**xlab** is used to give description of x-axis.
xlim	•**xlim** is used to specify the range of values on the x-axis.
ylim	•**ylim** is used to specify the range of values on the y-axis. breaks is used to mention the width of each bar.

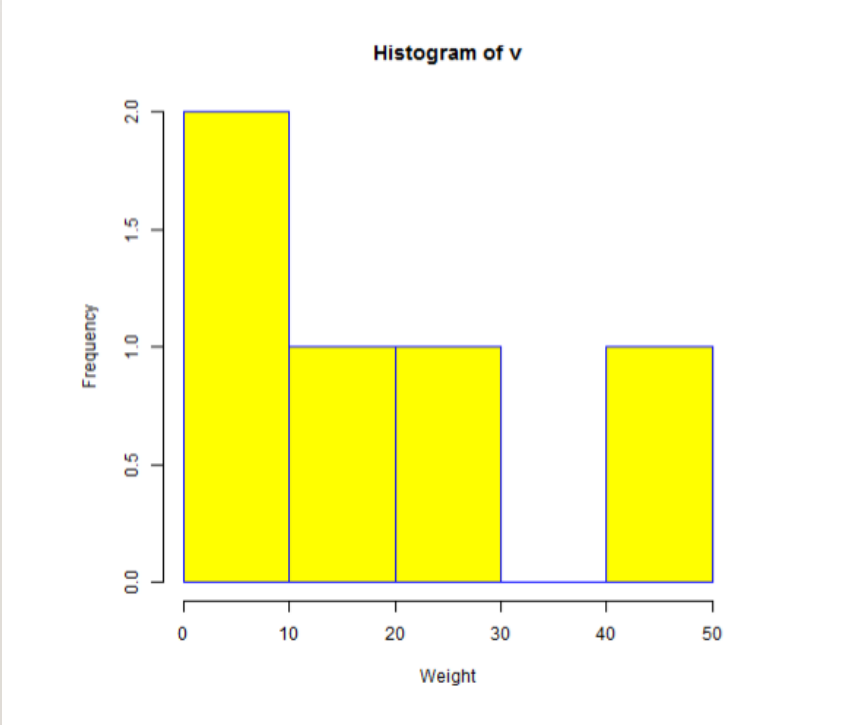
HISTOGRAM – EXAMPLE

A simple histogram is created using input vector, label, col and border parameters.

The script given below will create and save the histogram in the current R working directory.

```
# Create data for the graph.
v <-  c(9,13,21,8,36,22,12,41,31,33,19)
# Give the chart file a name.
png(file = "histogram.png")
# Create the histogram.
hist(v,xlab="Weight",col="yellow",border="blue")
# Save the file.
dev.off()
```

OUTPUT



LINE GRAPHS

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.

LINE GRAPHS

The plot() function in R is used to create the line graph.

LINE GRAPH – SYNTAX

The basic syntax to create a line chart in R is:

- plot(v,type,col,xlab,ylab)

LINE GRAPH - PARAMETERS

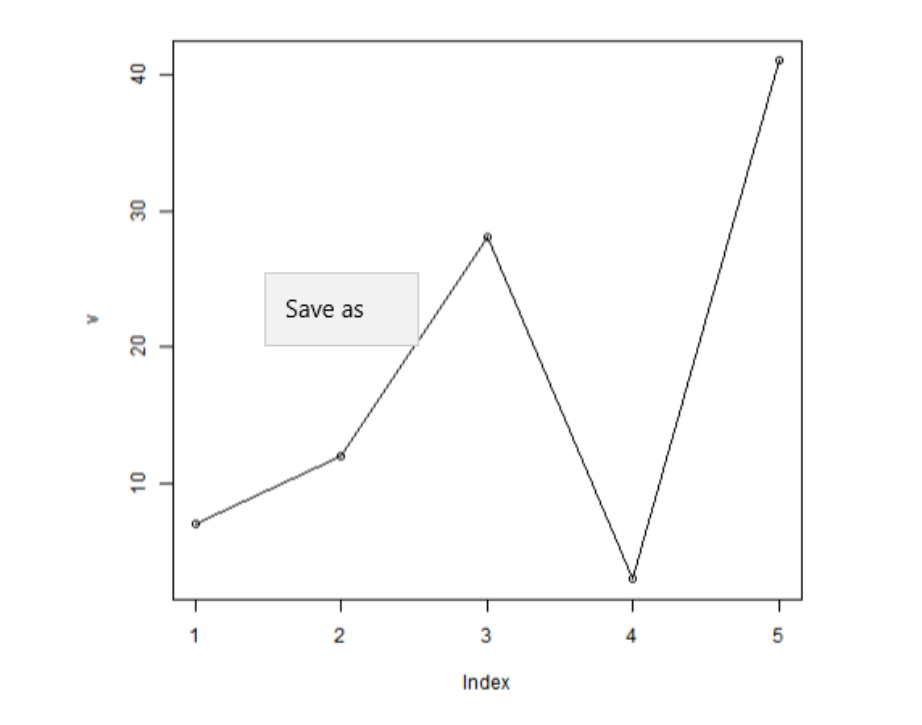
v	• **v** is a vector containing the numeric values.
type	• **type** takes the value "p" to draw only the points, "i" to draw only the lines and "o" to draw both points and lines.
xlab	• **xlab** is the label for x axis.
ylab	• **ylab** is the label for y axis.
main	• **main** is the Title of the chart.
col	• **col** is used to give colors to both the points and lines.

LINE GRAPHS - EXAMPLE

A simple line chart is created using the input vector and the type parameter as "O". The below script will create and save a line chart in the current R working directory.

```
# Create the data for the chart.

v <- c(7,12,28,3,41)

# Give the chart file a name.

png(file = "line_chart.jpg")

# Plot the bar chart.

plot(v,type="o")

# Save the file.

dev.off()
```

OUTPUT



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.

SCATTERPLOT

The simple scatterplot is created using the plot() function.

SCATTERPLOT – SYNTAX

The basic syntax for creating scatterplot in R is :

- plot(x, y, main, xlab, ylab, xlim, ylim, axes)

SCATTERPLOT – PARAMETERS

x	• **x** is the data set whose values are the horizontal coordinates.
y	• **y** is the data set whose values are the vertical coordinates.
main	• **main** is the tile of the graph.
xlab	• **xlab** is the label in the horizontal axis.
ylab	• **ylab** is the label in the vertical axis.
xlim	• **xlim** is the limits of the values of x used for plotting.
ylim	• **ylim** is the limits of the values of y used for plotting.
axes	• **axes** indicates whether both axes should be drawn on the plot.

SCATTERPLOT – EXAMPLES

We use the data set **"mtcars"** available in the R environment to create a basic scatterplot. Let's use the columns "wt" and "mpg" in mtcars.

```
input <- mtcars[,c('wt','mpg')]
print(head(input))
```

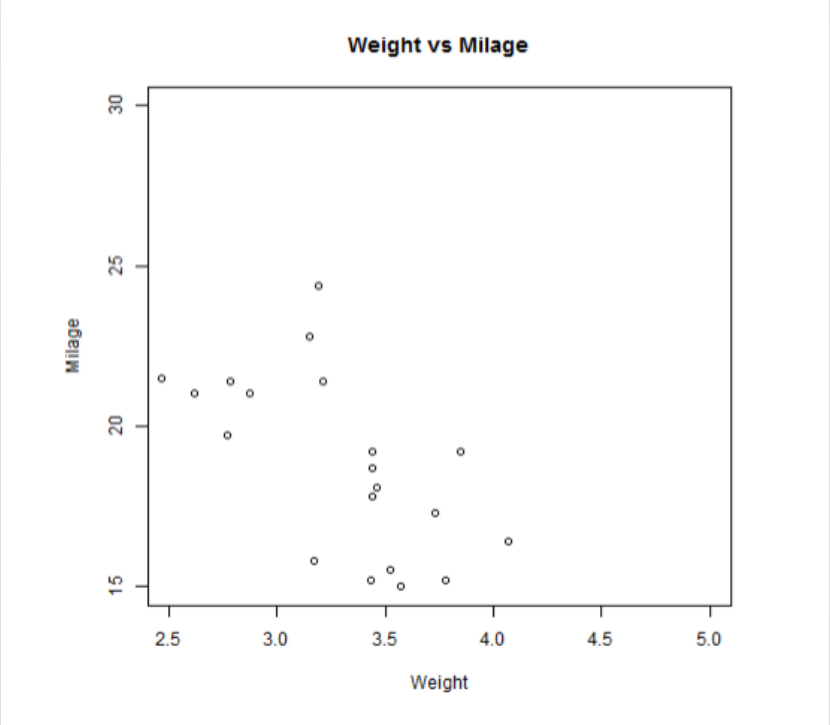
When we execute the above code, it produces the following result:

	wt	mpg
Mazda RX4	2.620	21.0
Mazda RX4 Wag	2.875	21.0
Datsun 710	2.320	22.8
Hornet 4 Drive	3.215	21.4
Hornet Sportabout	3.440	18.7
Valiant	3.460	18.1

CREATING THE SCATTERPLOT

- The below script will create a scatterplot graph for the relation between wt(weight) and mpg(miles per gallon).
- # Get the input values.
- input <- mtcars[,c('wt','mpg')]
- # Give the chart file a name.
- png(file = "scatterplot.png")
- # Plot the chart for cars with weight between 2.5 to 5 and mileage between 15 and 30.
- plot(x=input$wt,y=input$mpg,
- xlab="Weight",
- ylab="Milage",
- xlim=c(2.5,5),
- ylim=c(15,30),
- main="Weight vs Milage"
-)
- # Save the file.
- dev.off()

OUTPUT



DISCUSSION Q&A

THANK YOU!