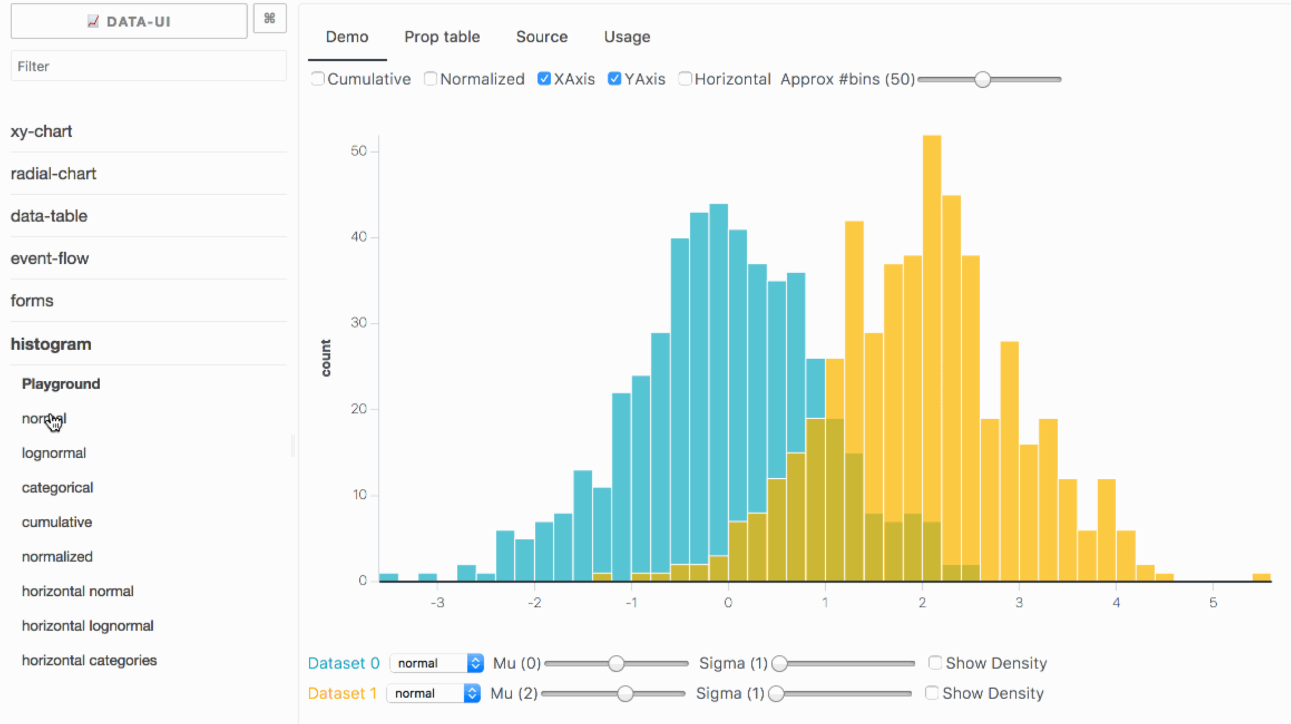


□ Cumulative □ Normalized ☑ XAxis ☑ YAxis □ Horizontal Approx #bins (50)

Dataset 0 normal Mu (0) Sigma (1) □ Show Density

Dataset 1 normal Mu (2) Sigma (1) □ Show Density

EXPLORATORY DATA ANALYSIS

TOPICS OF THIS SESSION

- Exploratory Data Analysis
- Need for graphs/plots
- Various types of graphs and their significance
    - Histogram, Bar plot
    - Pie graph
    - Line chart
    - Scatter Plot
    - Box plot

- Drawing graphs in R

OUTCOMES

After completion of this session you will be able to:

- Understand the significance of Diagrammatic representation of data
- Differentiate between various graphs
- Learn the significance of various graphs
- Learn to choose which graph to use
- Draw graphs in R

EXPLORATORY DATA ANALYSIS

- After the collection and verification of data, it needs to be compiled and displayed in such a way that it highlights the essential features clearly to the users.

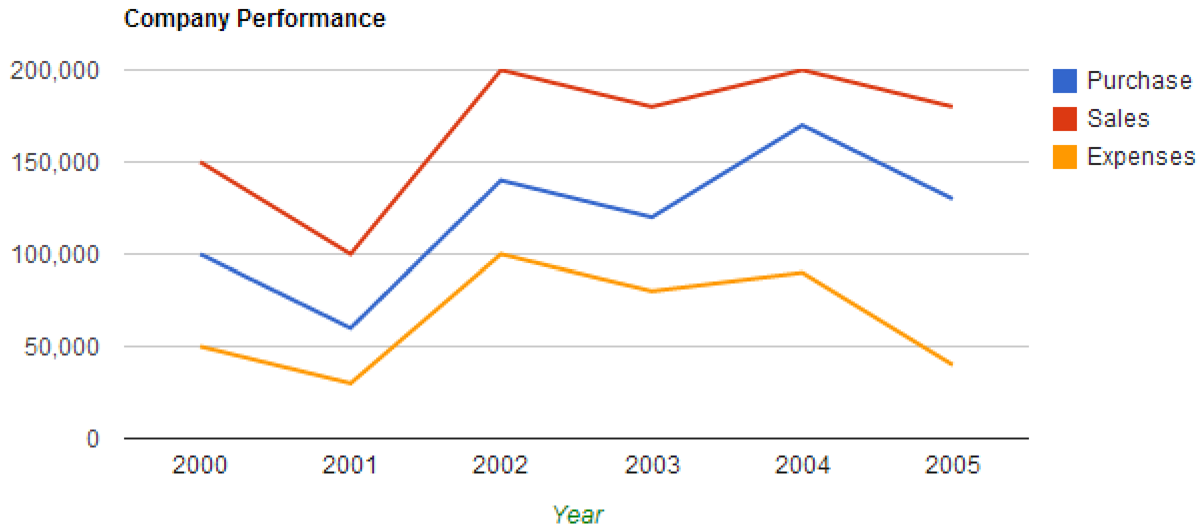
- The statistical analysis can only be performed if it is properly presented.

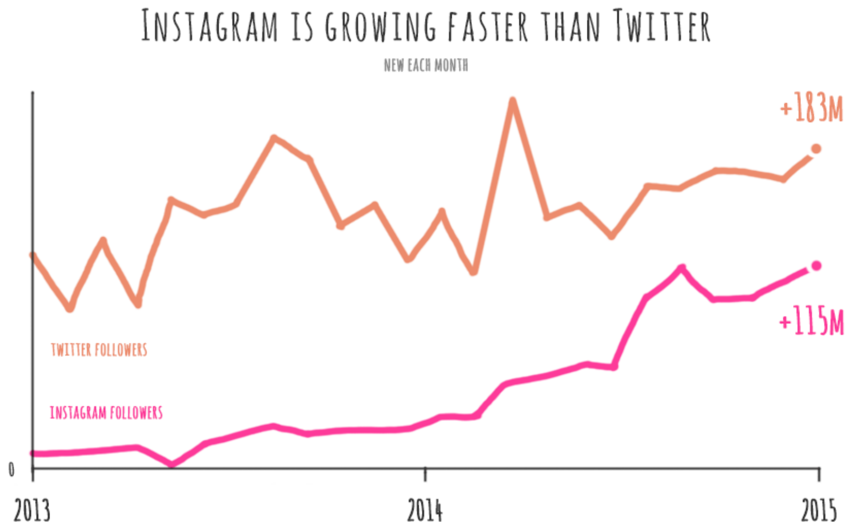
- There are three modes of presentation of data
    - textual presentation
    - tabular presentation
    - diagrammatic presentation

DIAGRAMMATIC REPRESENTATION OF DATA

- The diagrammatic representation of data is one of the best and attractive way of presenting data

- It caters both educated and uneducated section of the society.





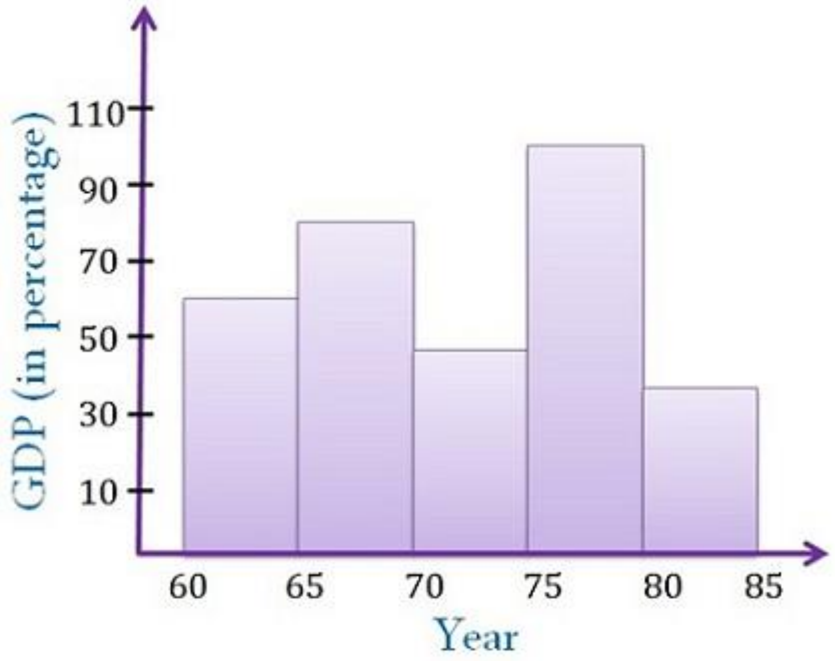
VARIOUS TYPES OF GRAPHS

Graphs in our syllabus:

- Histogram, Bar plot
- Pie graph
- Line chart
- Scatter Plot
- Box plot

HISTOGRAMS

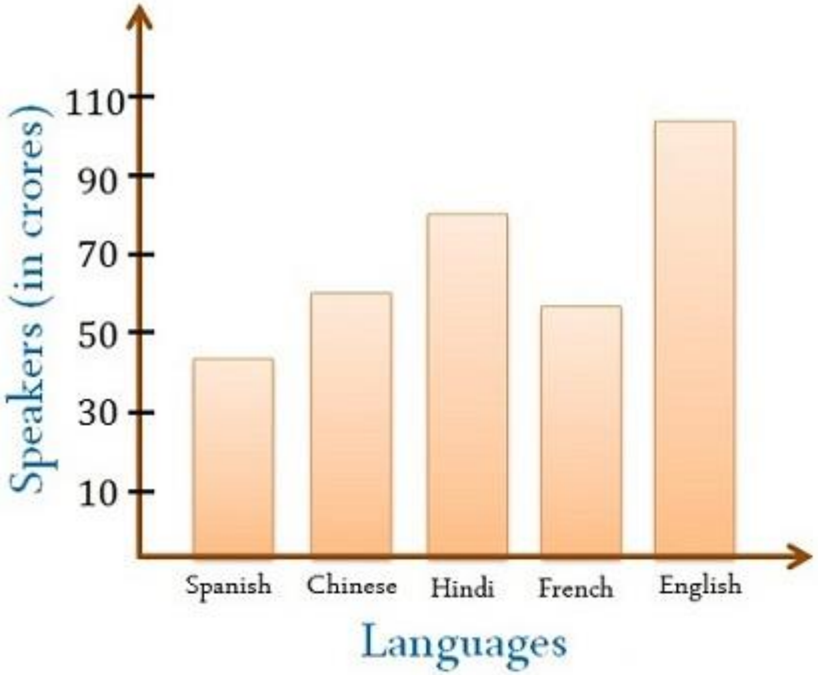
- It a type of bar chart
- used to represent statistical information by way of bars
- shows the frequency distribution of continuous data.
- It indicates the number of observations which lie in-between the range of values, known as class or bin.



Histogram

BAR GRAPHS

- graphically represents the comparison between categories of data.
- displays grouped data by way of parallel rectangular bars of equal width but varying the length.

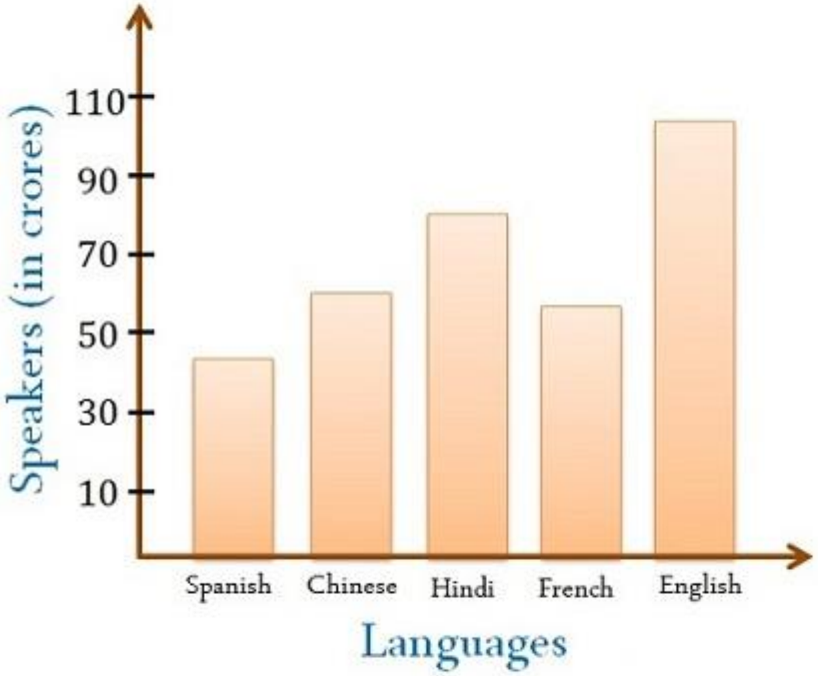


Bar Graph

BAR GRAPHS

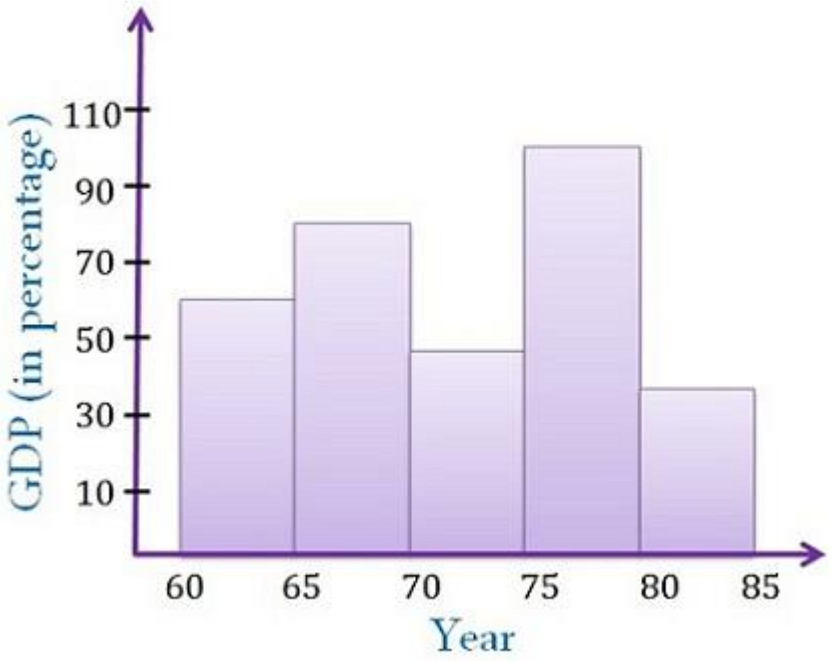
- Each rectangular block indicates specific category and the length of the bars depends on the values they hold.

- The bars in a bar graph are presented in such a way that they do not touch each other, to indicate elements as separate entities.

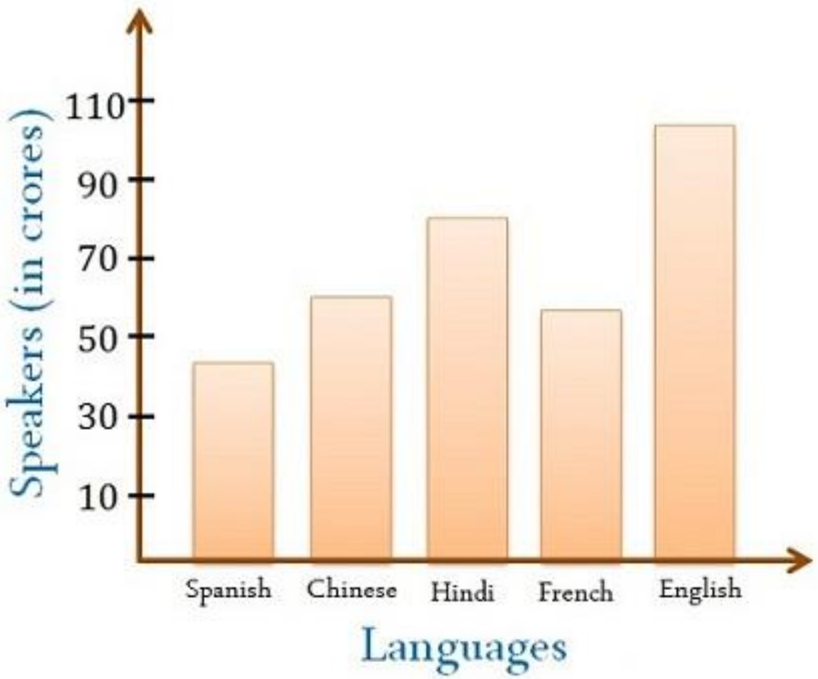


Bar Graph

HISTOGRAMS VS BAR GRAPHS



Histogram



Bar Graph

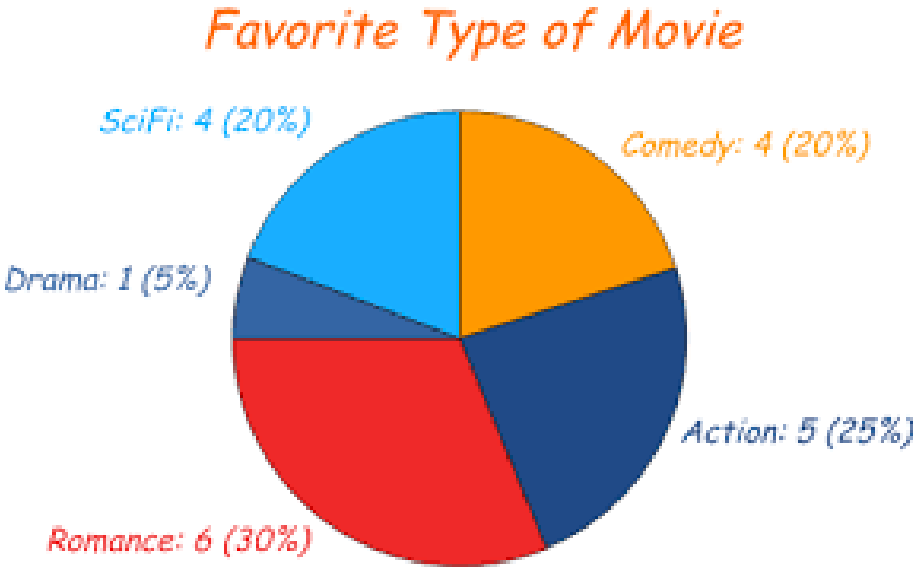
HISTOGRAMS VS BAR GRAPHS

BASIS FOR COMPARISON	HISTOGRAM	BAR GRAPH
Indicates	Distribution of non-discrete variables	Comparison of discrete variables
Presents	Quantitative data	Categorical data
Spaces	Bars touch each other, hence there are no spaces between bars	Bars do not touch each other, hence there are spaces between bars.
Elements	Elements are grouped together, so that they are considered as ranges.	Elements are taken as individual entities.
Can bars be reordered?	No	Yes
Width of bars	Need not to be same	Same

PIE CHARTS

- It is a circular statistical graphic, which is divided into slices to illustrate numerical proportion.

- the arc length of each slice is proportional to the quantity it represents.

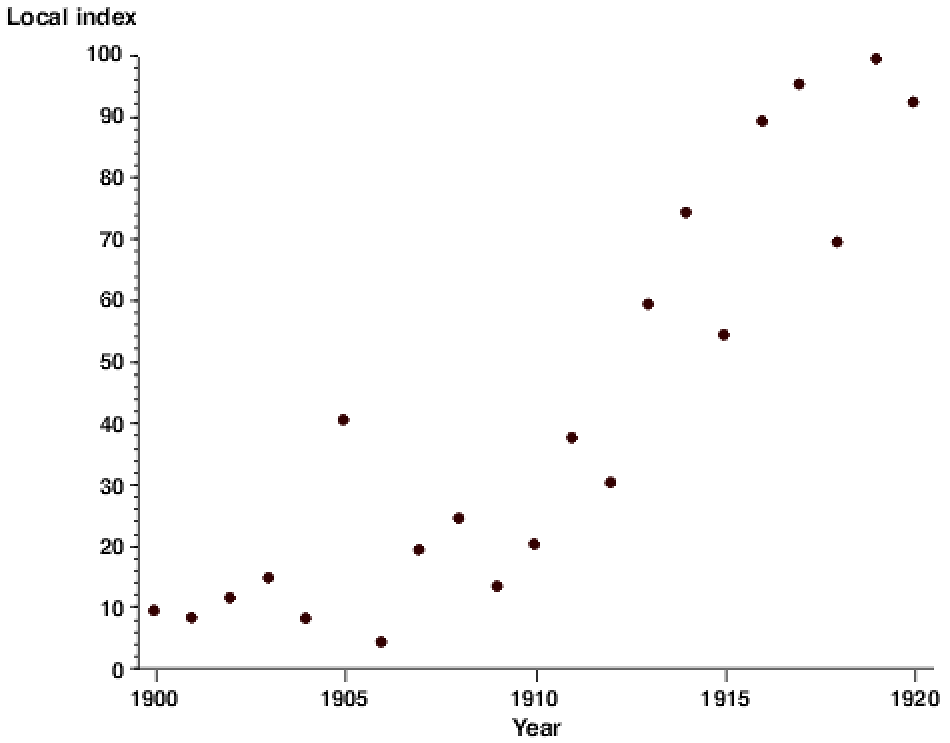


SCATTER PLOTS

- also called a scatterplot, scatter graph, scatter chart, scattergram, or scatter diagram

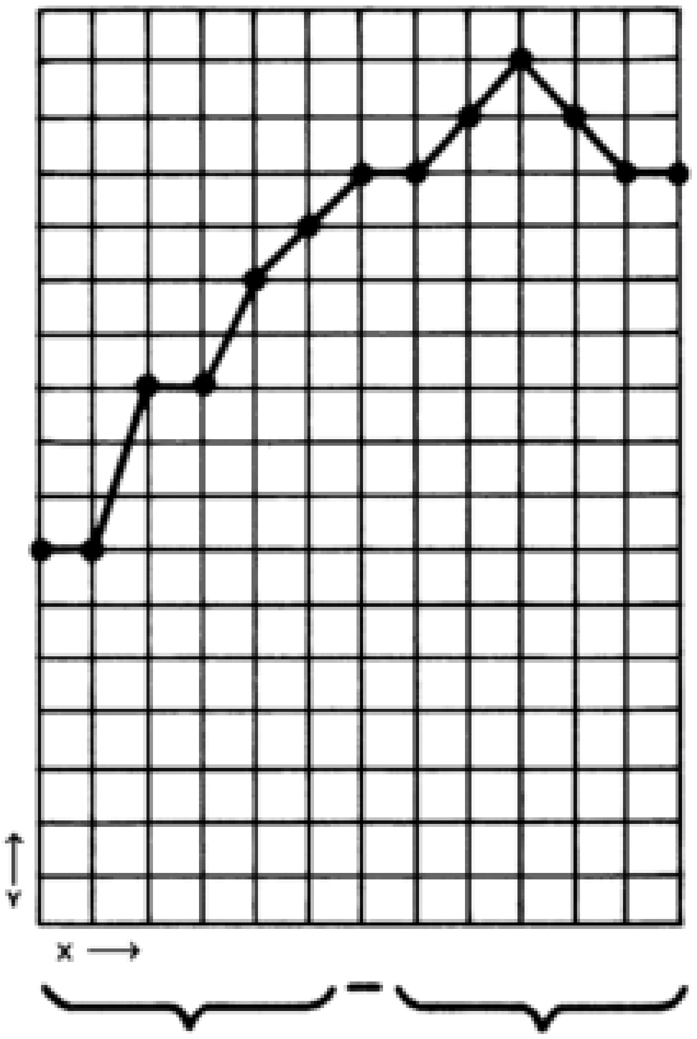
- display values for typically two variables for a set of data.

- If the points are color-coded, one additional variable can be displayed.

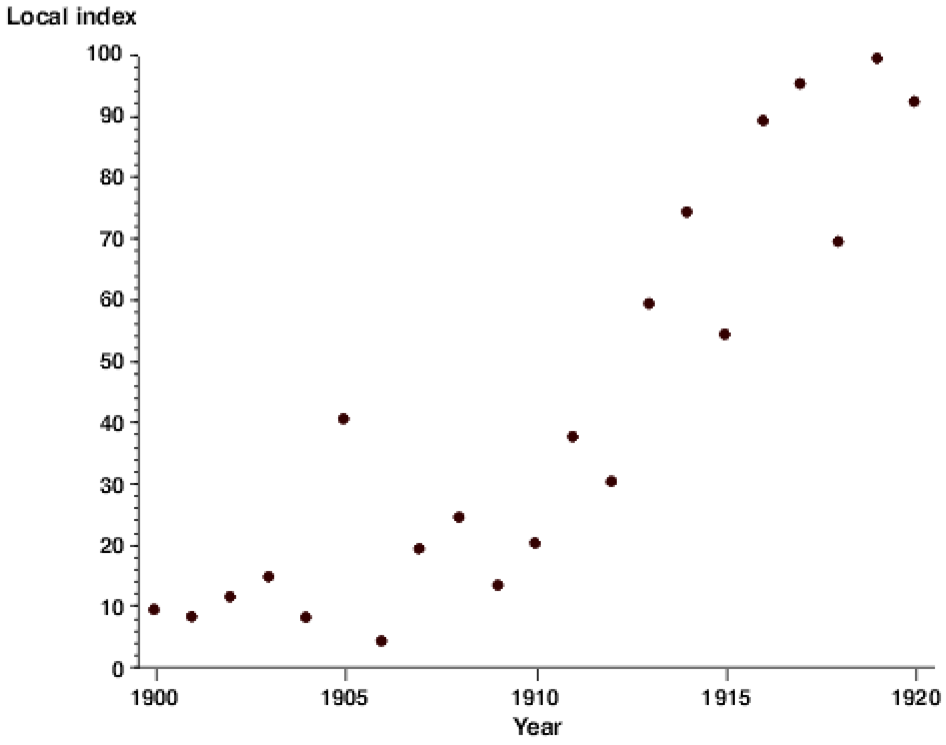


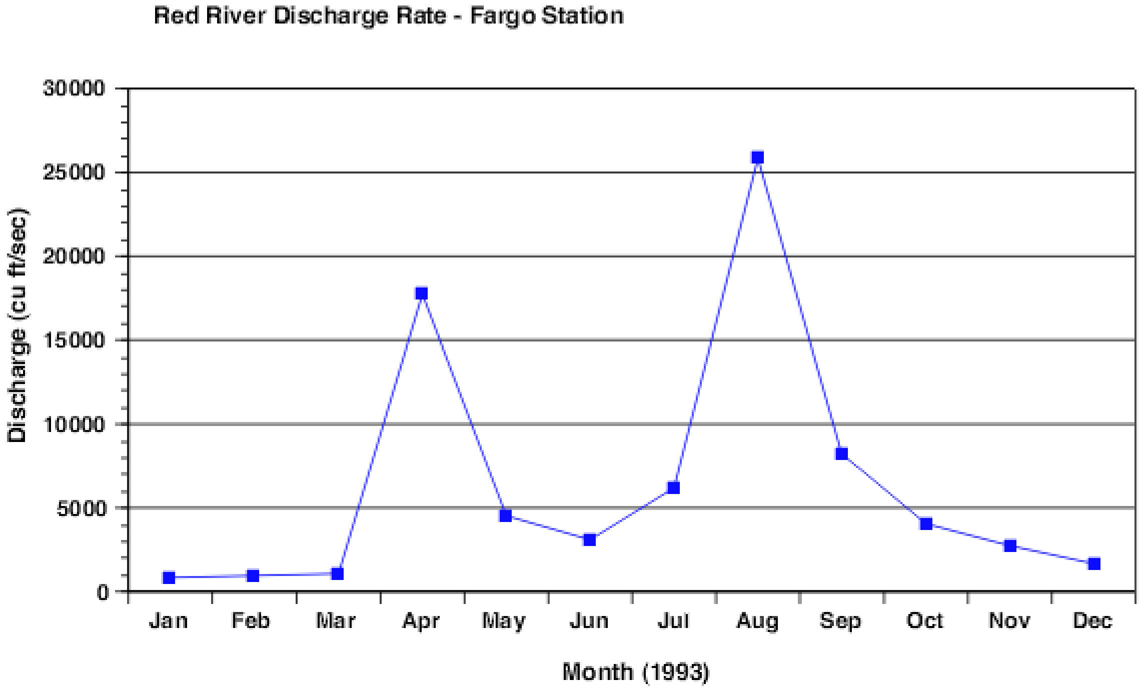
LINE CHARTS

- displays information as a series of data points called 'markers' connected by straight line segments.
- It is a basic type of chart common in many fields.
- It is similar to a scatter plot except that the measurement points are ordered (typically by their x-axis value) and joined with straight line segments.



SCATTER PLOTS VS LINE CHARTS

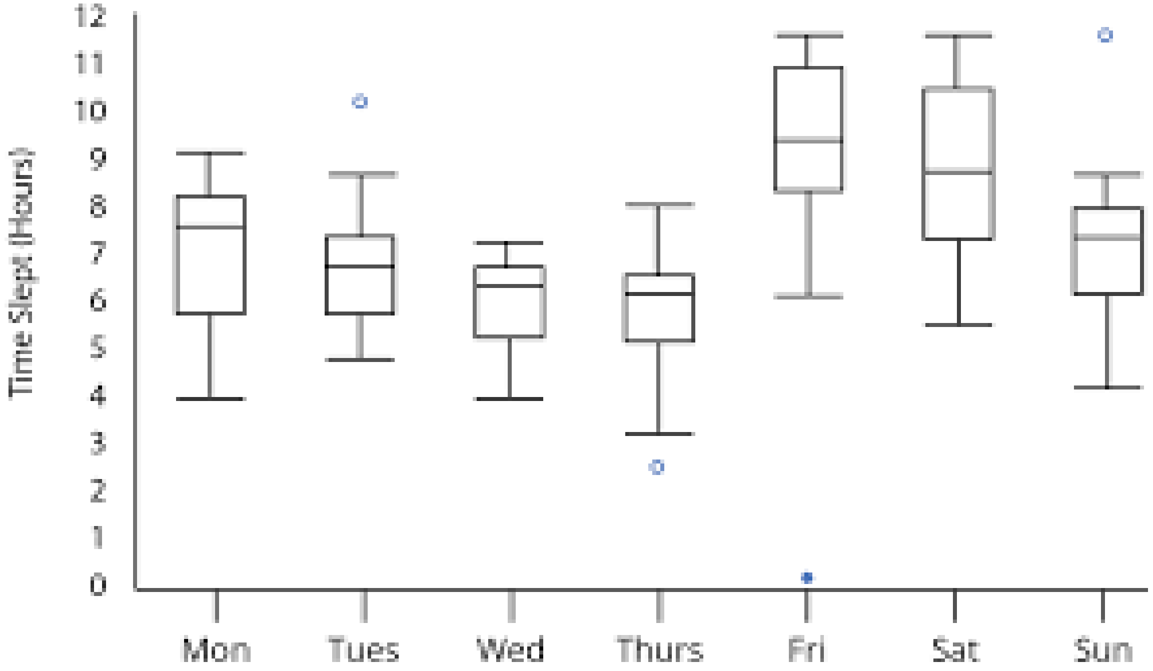




BOX PLOTS

Boxplots are a measure of how well distributed is the data in a data set.

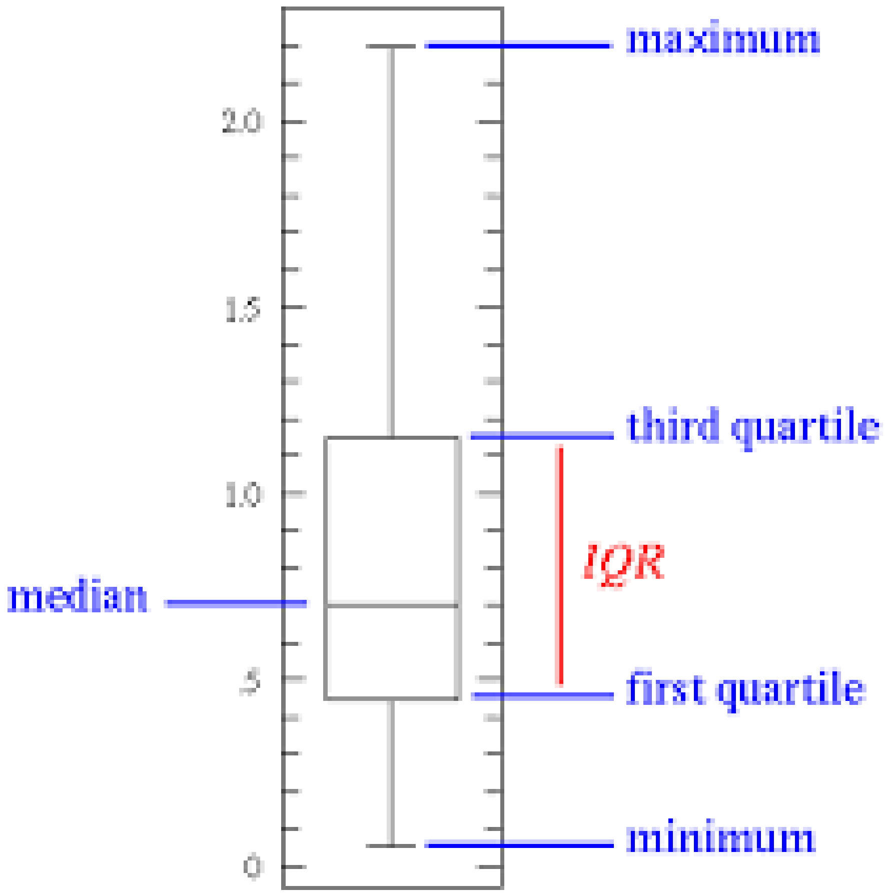
It divides the data set into three quartiles.



BOX PLOTS

- This graph represents the minimum, maximum, median, first quartile and third quartile in the data set.

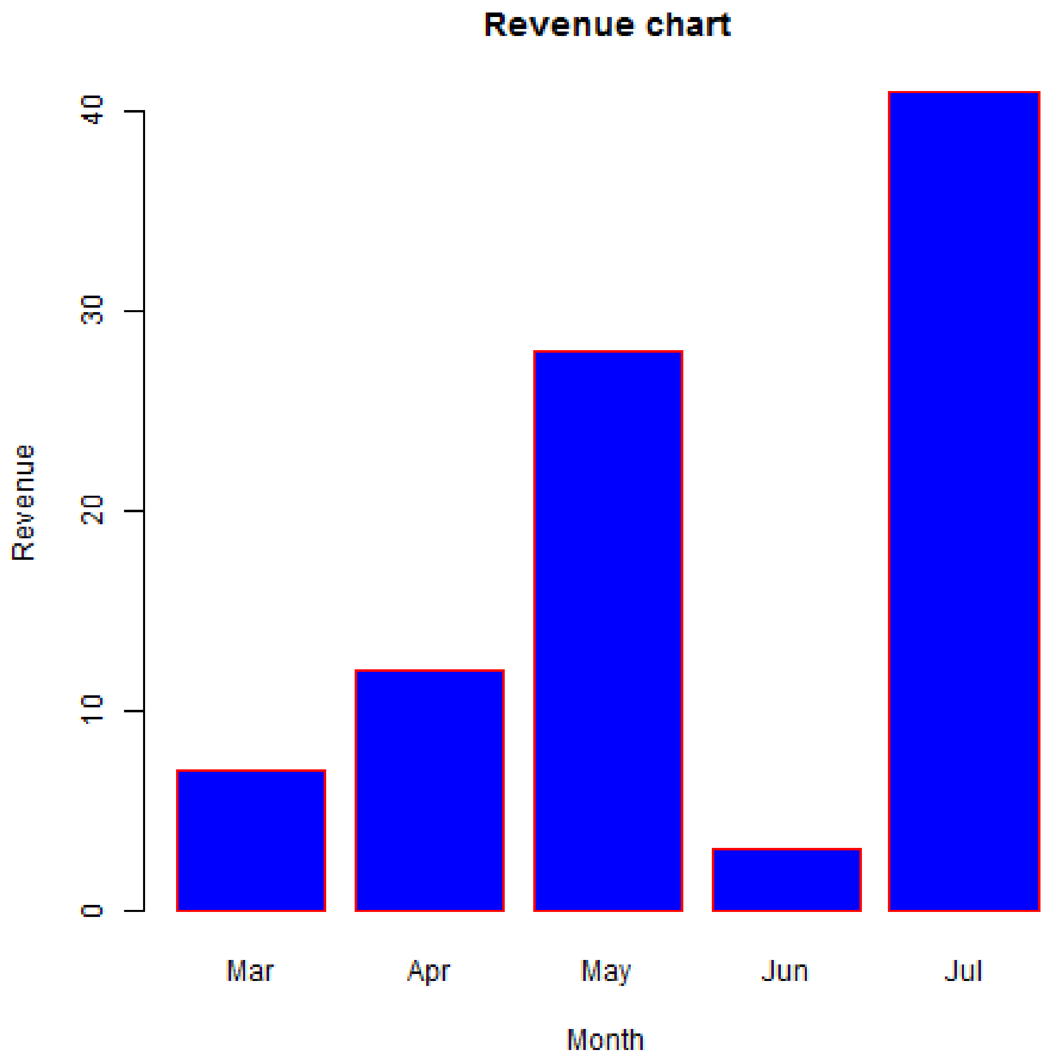
- It is also useful in comparing the distribution of data across data sets by drawing boxplots for each of them



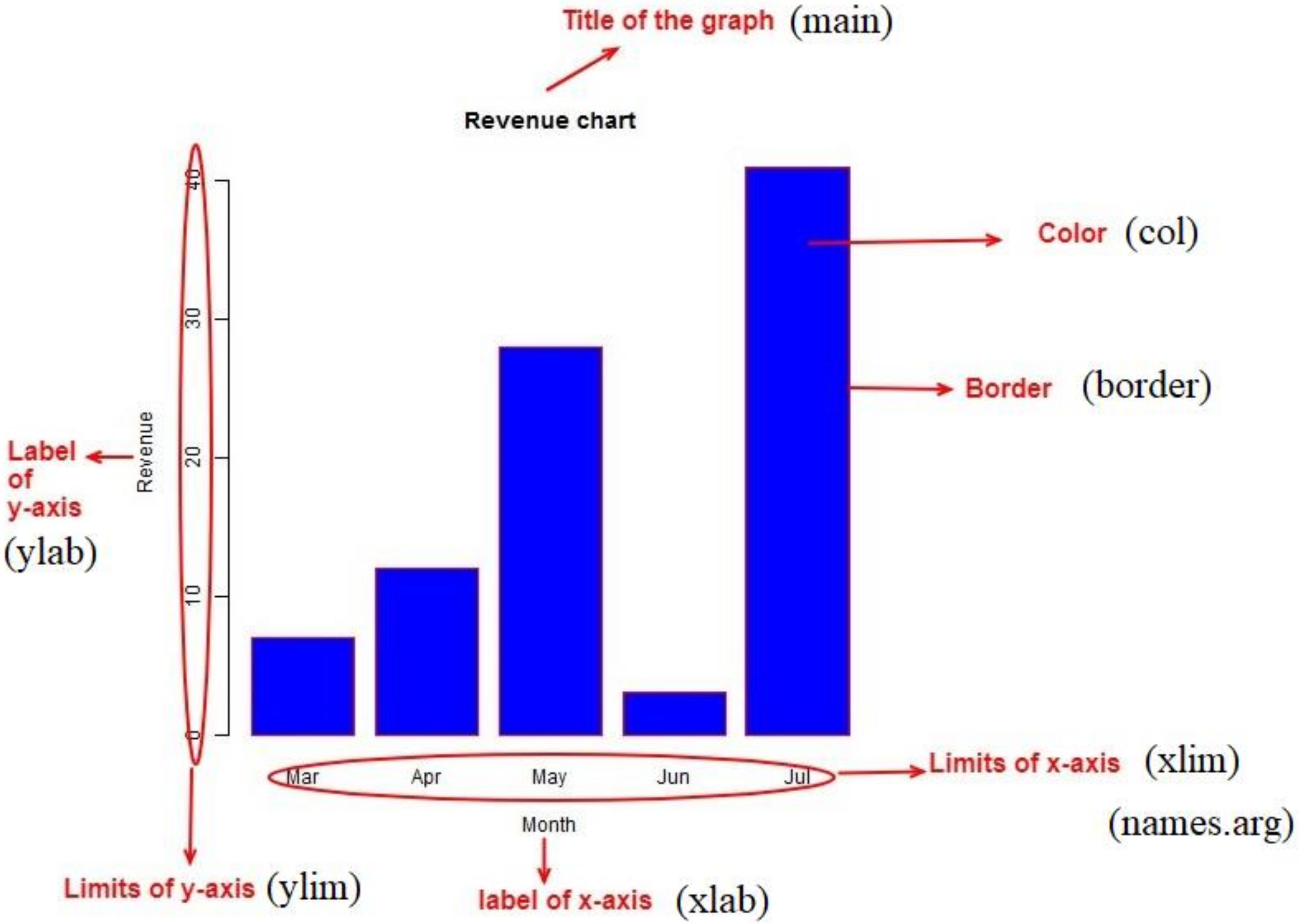
DRAWING PLOTS IN R - FUNCTIONS

Name of the Plot/Graph	Function
Box Plot	boxplot()
Histogram	hist()
Bar Plot	barploat()
Pie Chart	pie()
Line Chart	plot()
Scatter Plot	

PARTS OF A GRAPH



PARTS OF A GRAPH



COMMON ARGUMENTS

x is the data set whose values are the horizontal coordinates.

y is the data set whose values are the vertical coordinates.

main is the tile of the graph.

xlab is the label in the horizontal axis.

ylab is the label in the vertical axis.

COMMON ARGUMENTS

xlim is the limits of the values of x used for plotting.

ylim is the limits of the values of y used for plotting.

axes indicates whether both axes should be drawn on the plot

col is used to give colors to both the points and lines

border is used to set border color of each bar

names.arg is a vector of names appearing under each bar