

# Assignment No

Title :- Iris (flower data) data analytics

## Problem Statement:

Download Iris dataset or any into dataframe

- 1) How many features are there & types
- 2) display summary statistics of each feature (min, max, mean, range, SD, variance, percentile)
- 3) Create histogram
- 4) Create boxplot.

## Objectives :-

- 1) Student can understand the data descriptive statistics and visualization
- 2) ~~plot~~ also they can understand the boxplot histogram & various attributes.

## Outcome :-

- ① understanding use of R studio/jupyter
- ② able to understand different attributes of data (mean, mode, median, mode, SD, etc.)

## Theory:-

The dataset iris is made up by Edgar Anderson collected the data. It dataset gives measurements in centimeters of the variable sepal length & width & petal length & width resp for 50 flowers each of 3 species of iris.

## format:

iris is a data frame with 150 cases (rows) and 5 variables (column) named sepal.length, sepal.width, petal.length, petal.width and species. here a little summary of what you can basically using → Summary (iris)

The statement gives the some analytical data about the iris dataset i.e. min, 1st qu, median, mean, 3rd Qu, max, Species.

→ head (iris)

It simply returns first six records



→ `typeof(iris[,1])`

It gets the iris dataset first column & gives their general type in that case it is double.

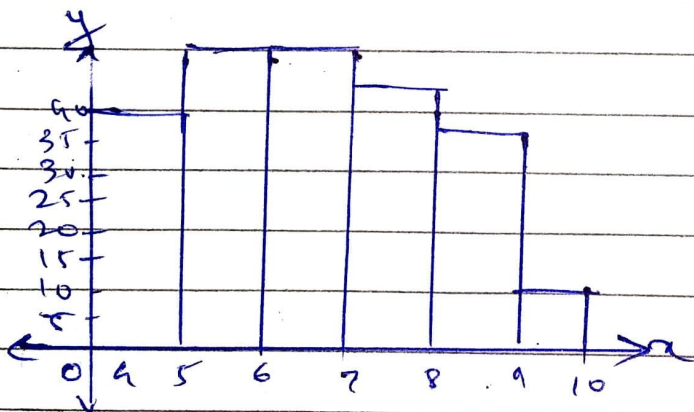
→ Instead of using direct summary statement we given `max()`, `mean()`, `min()`, `median()`, `sd()`, `range()` & `var()`

→ `quantile(col)`

It gives the quantile mode of data in range

→ `hist(iris$sepal.length)`

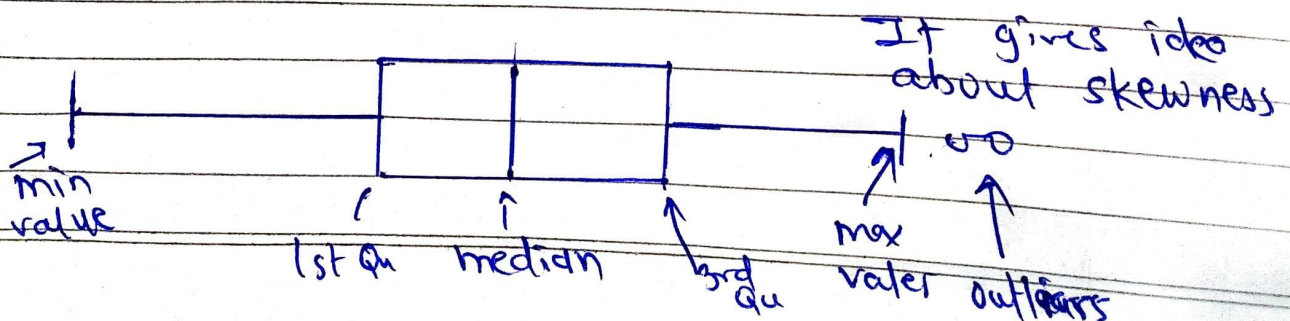
It prints histogram of given column



It gives frequencies w.r.t values.

→ `Boxplotr :- box plot(iris[,1])`

It prints boxplot on 1st column



### Conclusion:-

Hence we understood basic dataset analysis on different parameters of dataset iris.