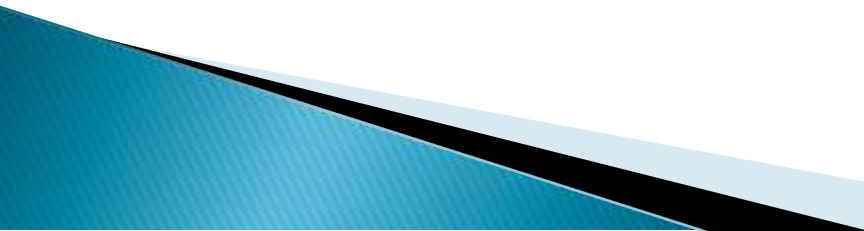# "An Introduction to Statistical Software R"
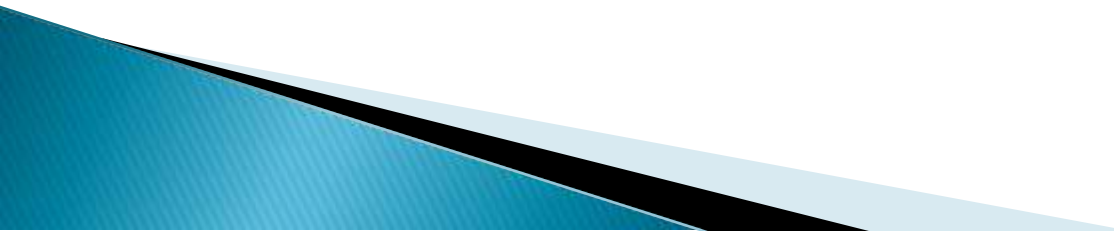
by

Geethu Gopinath
Department of Statistics
Christ College(Autonomous), Irinjalakuda

What is R?

- R is an open source programming language.
- It was developed by **Ross Ihaka** and **Robert Gentleman** in 1993
- It has been developed for statistical computing and graphics supported by **R Foundation**.
- **R** is the most popular language in the world of data science.
- As of September 2020, **R** ranks 9th in TIOBE (The Importance Of Being Earnest) index.

Why it is so popular?

- Free to download
- Less storage space
- Complete source code available
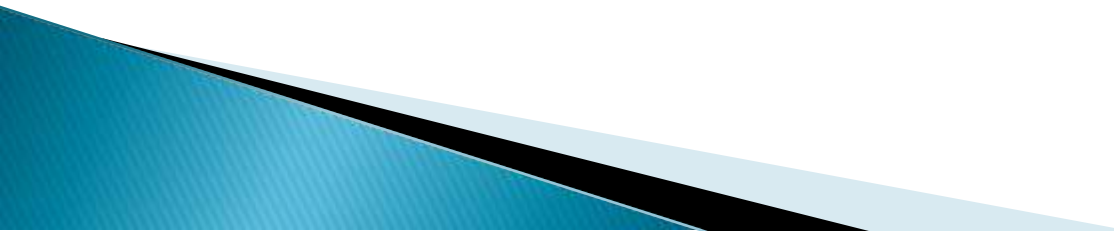- User friendly
- Updation –every 6 months

How to download and Install R?

- R is freely available from the Comprehensive R Archive Network
(CRAN) at http://cran.r-project.org/
- CRAN is the first port of call for everything to do with R. It is from
here to download and install R.
- Find contributed packages to solve particular problems
- Find the answers to frequently asked questions, Read about the latest developments, Get programming tips and much more besides.

R as a calculator

- Addition (+)
- Subtraction(-)
- Multiplication(*)
- Division(/)
- Exponentiation(^)

Methods of data input in R

- c function
- Seq function
- Scan function
- rep function
- class function
- data.frame function

1. c function

- Combines or concatenates terms together.

- Example

a) y <- c(1, 2, 3, 4, 5)

  # Constructs a vector

a) Names <- c("A", "B", "C")

  #Construct a vector of
  character strings

2. Sequence operator and seq function

- Generate regular sequences.

- Example
a)   1:4, seq(5:20)

  # Generates consecutive numbers

a)   seq(2,10,by=2)

  #Specifies interval and increment

3. scan and rep Functions

- **Scan()-** Used to scan and read data. It is usually used to read data into vector or list or from file in R Language.

- **rep() -** replicates numeric values, or text, or the values of a vector for a specific number of times.

a) 1 1 1 1 2 2 3 3 3

X=c(rep(1,4),rep(2,2),rep(3,3))

b) Y=2:10

rep(Y, each=4)

4. class function

▸ Useful in deciding the class of data object.

▸ Example

a)   a<- c(5, 8, 6, 10); class(a)
               # Output of this command is "numeric"

a)   d<- c("x", "y") ; class(d)
               # Output of this command is "character"

5.data.frame function

- Useful for creating table
- Example
  1) Suppose we have two vectors:-
     x=(1,2,3,4), y=(2,3,4,5).Create
     following data frame:(x, y)
     data.frame(x,y) #command used

BUILT-IN FUNCTIONS IN R

- max function: max(x)
- min function: min(x)
- length function: length(x)
- sort function: sort(x)
- Mean(x)
- Median(x)
- sum(x)
- Example
- X<-c(4,5,2,13)

Data accessing or indexing

- Access the elements of an object , which can either be numerical or logical
- Square brackets [ ] are used for data accessing
- Example

    x < - 2:12

    x

    x[3]   # access third element from
            dataset

Steps:-

- ▸ Enter the data in Excel with column names
- ▸ Save the file as *CSV(comma delimited) text file with* name (say data)
- ▸ Close the file
- ▸ To import the file in R console, use **read.csv** function

R command is:

read.csv(file.choose(),header=T)