# Lecture 01
Introduction to R software

Welcome to the course on descriptive statistics with R software this is our first lecture; although I believe that all the candidates who are attending this course they have a basic knowledge of R software. And in case if you do not have then there is another course on introduction to R software I would request you that you please go through that lecture. But in order to have a quick review of the mathematical tool statistical tools and the commands, which we are going to use in this course I will try to demonstrate it here in the next couple of lectures. So, next couple of lectures will be devoted, to

the introduction to R software and the tools which we are going to use in the statistical part they will be demonstrated here quickly. Okay?

Refer slide time :(01:10)
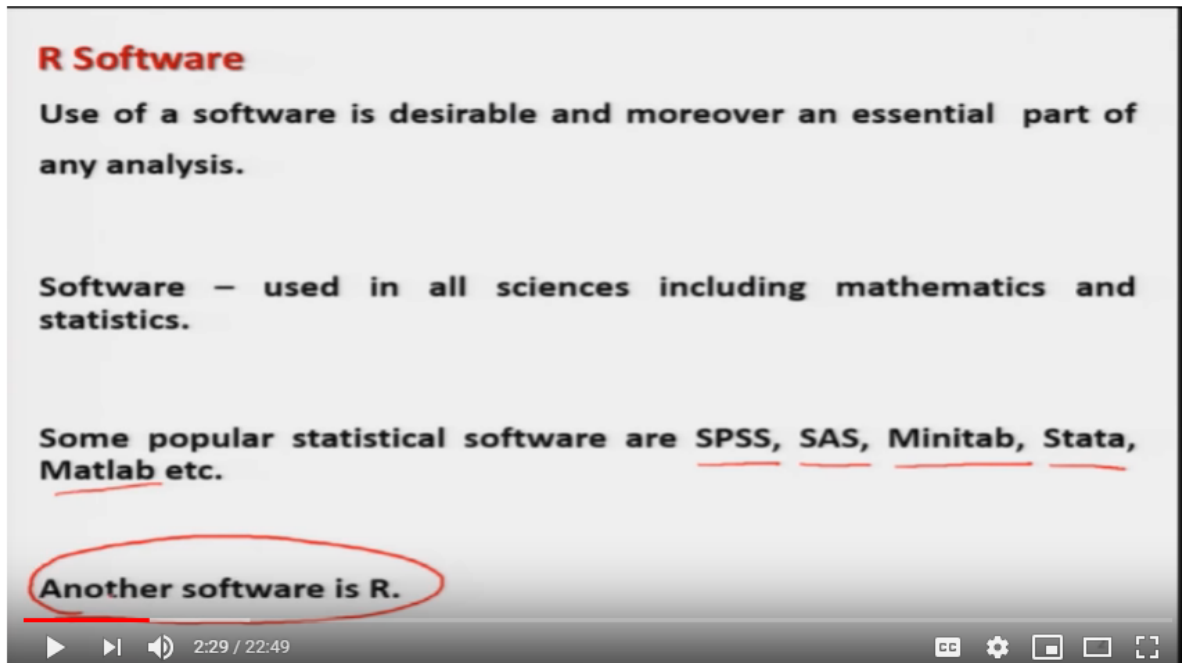
So, first question comes what is this R software and why this is so important for us to use and what is the role of R software in mathematics and statistics and other basic sciences. Right? So, we all know that use of software is desirable and moreover it is an essential part of any statistical analysis or any mathematical analysis or rather I would say that any analysis is incomplete, if this is without the use of mathematics or statistics. So, definitely you need to learn the software how to use it in all basic sciences rather this is physics, chemistry, mathematical statistics, or any discipline in sciences arts Medical Sciences, engineering sciences or anything else Right? Now I would like to concentrate here basically on the statistical software. So, once we go through with the esthetical software there are different types of software which are available in the literature for example there is an SPSS there is a SAS there is a Minitab there Stata there is Matlab and similarly on the same lines there another software here which in are now the question comes.

Refer slide time :(02:23)

Developers of R Software

Currently developed by the R Development Core Team.

Available at www.r-project.org

It is a free software.

It supports many free packages which helps the data scientist and analyst.

What is the difference between the two? And how this are came into picture? This are is basically developed by a team and this is called as our development code team and this software is freely available on the website WWW.r-projefct.org and it is a free software. This is the biggest advantage of this software all of this software they are mostly paid software and sometime it is difficult for a common person to afford them. So, R software gives us a solution and it gives us a very strong edge over the other software that this is a free software and it's not only a software but this also supports many free packages free packages means in the are there is a base package which includes most of the common features which are usable but there are some specialized tasks, they can also be completed using the additional package available with the R software. Right?

Refer slide time :(03:46)

What is R?

R is an environment for data manipulation, statistical computing, graphics display and data analysis.

Effective data handling and storage of outputs is possible.

Simple as well as complicated calculations are possible.

Simulations are possible.

So, that is why it is very, very popular and R is an environment for data manipulation for this tentacle computing, graphic display, as well as data analysis, or even who can think of other things also and this gives us a an effective way to handle the data and storage of the outcome is also possible the stories can be done one by one or the stories can be done in a loop or in other ways which will be sort of automated way you run the program and, and the outcome of the analysis, will be stored in some file and that can be exported to different types of software's .Right? And this case simple as well as complicated computations are also possible simulations like a Monte Carlo simulation statistics they are also possible now what is in case if you are involved in research, without doing simulation it is very difficult to survive unless and until you do a research to give the mathematical aspect if this tactical aspect, is not only the sufficient thing but you also need to do the Monte Carlo simulation to demonstrate the utility and application of your tools using the data. So, this R software also works in that thing. And now if you know that in most of the software whatever are your graphics means any software will have rather two types of outcome graphics, as well as some numerical values. So, the numerical value as I already told you that they can be stored in a file similarly the graphic can also be stored they there are two options, the graphics can be displayed on screen.

Refer slide time :(05:35)

What is R?

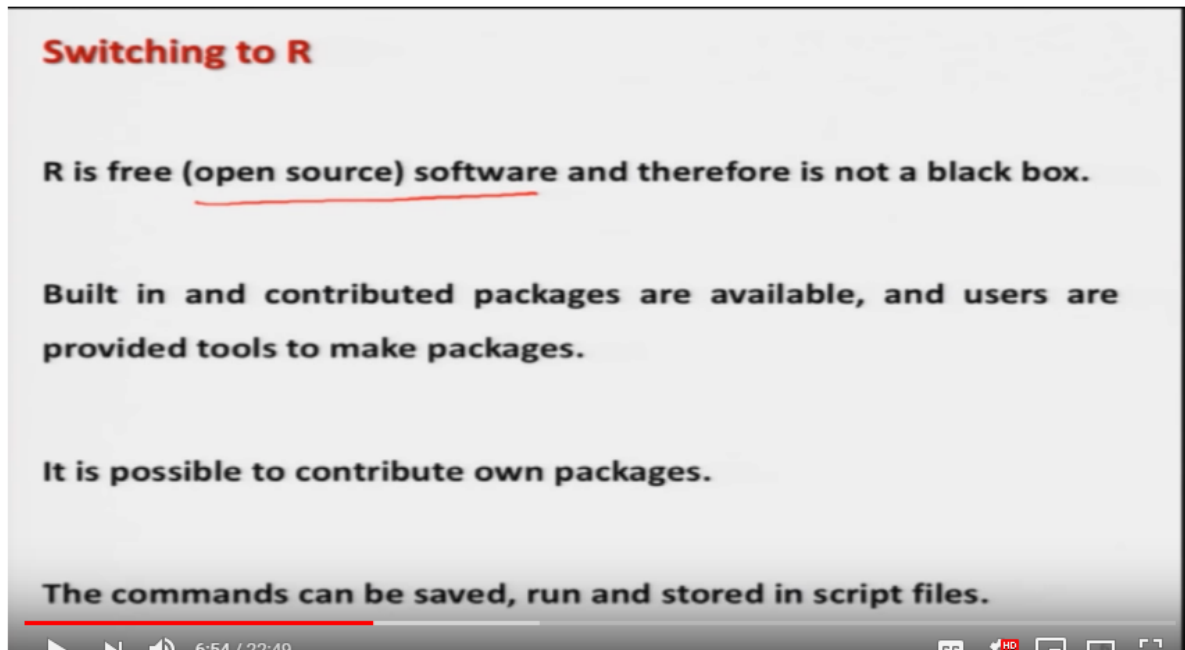Graphical display on-screen and hardcopy are possible.

Programming language is effective which includes all possibilities just like any other good programming language.

R has a statistical computing environment.

It has a computer language which is convenient to use for statistical and graphical applications.

As well as they can be stored and their hard copies are also possible. And in R there is a programming language it's not only that there are some built-in functions, that can be used but you can also create your own programs. And this is the main advantage that you can use the built-in function, as well as you can create your own function using the programming language. And cut is why this R has an environment of using the programming language which is its own programming language, it is different than other languages although it is very, very similar to other languages but it has its own syntax and commands and this R software has a basically a statistical computing environment. There are many, many things in statistics and also in mathematics which are directly supported by using the built-in function or by using the contributed packages. And the environment which are provides that is very suitable for the statistical and mathematical, calculations, computations, and brick educating in graphical output also. Right?

Refer slide time :(06:43)

Switching to R

R is free (open source) software and therefore is not a black box.

Built in and contributed packages are available, and users are provided tools to make packages.

It is possible to contribute own packages.

The commands can be saved, run and stored in script files.

And as I said the biggest advantage of R software that it is a free software and it is an open source. This means what this it's not like a black box if you really want to see what is happening inside the program you can just open it and can and can have an idea that how are is computing something and as I said earlier there are some packages which are built-in they are already embedded in the R software and there are some contributed packages, contributed packages mean somebody is doing his doing their research and he or she develops something new they can write a program and after a scrutiny or after checking or verifying by is by the our core development team this program can be uploaded on the website of R software and then anybody can just download it and can use it in their own research work. Right? So, in R it is also possible to contribute our own packages and this gives us a very strong advantages over the other software this type of facility is not really available in many other software's. Right? And all those commands whatever we are using they can be saved if they can be their output can be can be used and so, on this R software is available for all sorts of operating system like as

Refer slide time :(08:19)

Windows, Unix, Linux Macintosh and whatever are the graphics they can also be stored in different types of format the most common format of graphics is this PostScript file or the PDF format although they can also be stored in JPEG format they can directly be copied from the software and they can be pasted in say the type of software like as this MS Office ms word. And so, on

Refer slide time :(08:46)
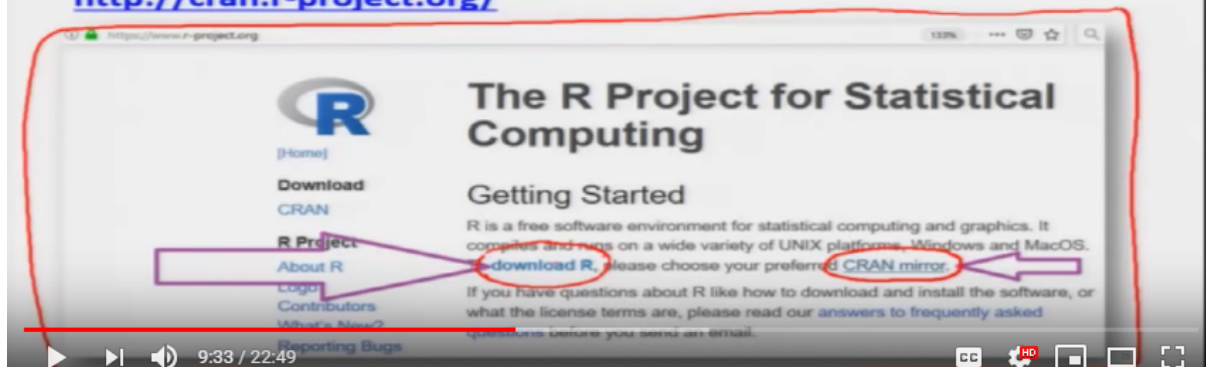
Now after this brief introduction and advantages of R I will just try to show you that how you can install the R on your computer so, in order to install the software what you have to do that you have to go to this website. Now once you go to this website you can this is here a sort of the screenshot of the webpage that you will get. And here if you try to see there are here links this link or you can also come through here this CR e n mirror and you can simply click here just click here and after that the software, will be downloaded and once you download it you simply have to click on the software it

will ask you various option. And you simply have to press different clicks and then the software will be installed and once this is installed on your computer. So, now I will come on the desktop and I will show you that how the things are going to happen so, you can see here there is an R icon and I simply double click on it and you will get this type of window here .Right? Just in order to make it more. Clear I will try to increase the phone size user so, that you can see whatever I am doing here clearly Right? So, you will get here this type of thing and now this is screen what you are seeing here this is called 'console', and by pressing the here control L that means you have to press two key ctrl + L together this will clear the screen and we will use it for executing different types of commands. So, now it was quite mean come here to over slides and I will try to show you here what's really happening. Right?

Refer slide time :(10:51)
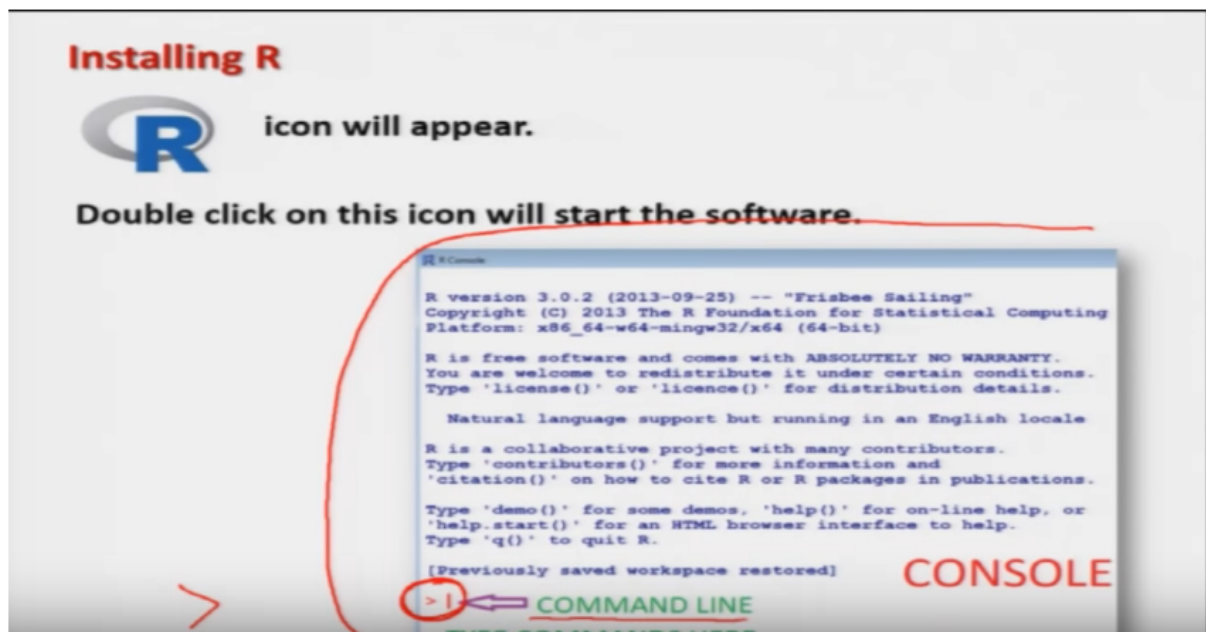
So, if you see here this is screenshot which is here this is the same thing which I just showed you on the R software this is called 'console', and here you have seen that there is a sign here something like greater than, this is the place where you write all that are commands you try to type all the commands and this place is called as command line. So, that will be our common terminology come on console or type the command on the command line and so on.

Refer slide time :(11:27)

Working with R

Use command line to type and execute the commands.

Some free software like R Studio, Tinn R etc. are also available to work with R software. It is an interface between R and us.

Such software are more useful for beginners.

It makes coding and execution of programmes easier.

R Studio is available at https://www.rstudio.com/

Tinn R is available at https://sourceforge.net/projects/tinn-r/

So, you need to be familiar with this thing. So, what we try to do? Whenever we want to execute any command I will go to the command line and I will type it there and I will press ENTER and that will execute the command. When you are trying to work with R software there are two options first option is that you can use directly the R software and you can type your commands, you can store your commands. And see the outcome directly on the R software on the console and second option is that there are some software's which are available they are the free software's for example one software is R studio and another software is Tinn R and similarly there are other software's are also this software they help us they are the interface between the R and us. So, using the R studio or Tinn R or similar type of software will help you more in executing the commands of r and getting the outcome but here my objective is not to teach you here the R software. So, I will be working only on the console and, and I will leave it up to you that this software you want to use. Okay? So, this R studio software this is available at this way website www.rstudio.com// and similarly this Tinn R software is available at this site. So, if you wish you can just download it and install it

Refer slide time :(12:58)

Installing Packages and Libraries

The base R package contains some necessary libraries only.

Other libraries are required for advanced statistical work which are downloaded and installed.

Run the R program, then use the `install.packages` function to download the libraries.

Examples :

`install.packages("ggplot2")` : installs package ggplot2
`install.packages("graphics")` : installs package graphics
`install.packages("cluster")` : installs package cluster

15:15 / 22:49

On the computer and start working on this now I will try to just take a quick commands and quick things which are needed for us to work instead stick using the R software. So, as I told you that in our there is a base package base. Package contains some essential libraries which are usually common among the users and these libraries are required to do an aesthetical work. And some of the libraries are the part of the base package and some of the libraries have to be downloaded from the website and these libraries are needed to execute a particular type of task. Right? So, first I will try to demonstrate here how you will install the package from this celebre and how you are going to bring it to a platform where it is available to use in any data set. Okay? So, first command which I would try to illustrate here is install dot packages, installed dot packages is a function which is used to download the libraries? And in order to use it this is the syntax suppose I want to download here a package ggplot2 this is a this is used for graphics so, I have to type here install dot packages and inside these two brackets here, I have to write double quotes here, and say here. And then I have to write down ggplot2 that is the name of the package which I want to install similarly in case if I want to install the package graphics this can be done by typing install dot packages and inside the brackets because, in the double course you have to write graphics and similarly. If you want to install another package here say cluster you simply have to type your install dot packages and inside the bracket within the double quotes just write CLUSTER and once we enter R software will guide you how to install this software? And then so, you simply have to choose a mirror and then you have to simply click, click and this a package will be installed here. I'm not demonstrating here but I would but I would like to leave it up to you that you practice and after installing the lab this package suppose I want to use it then
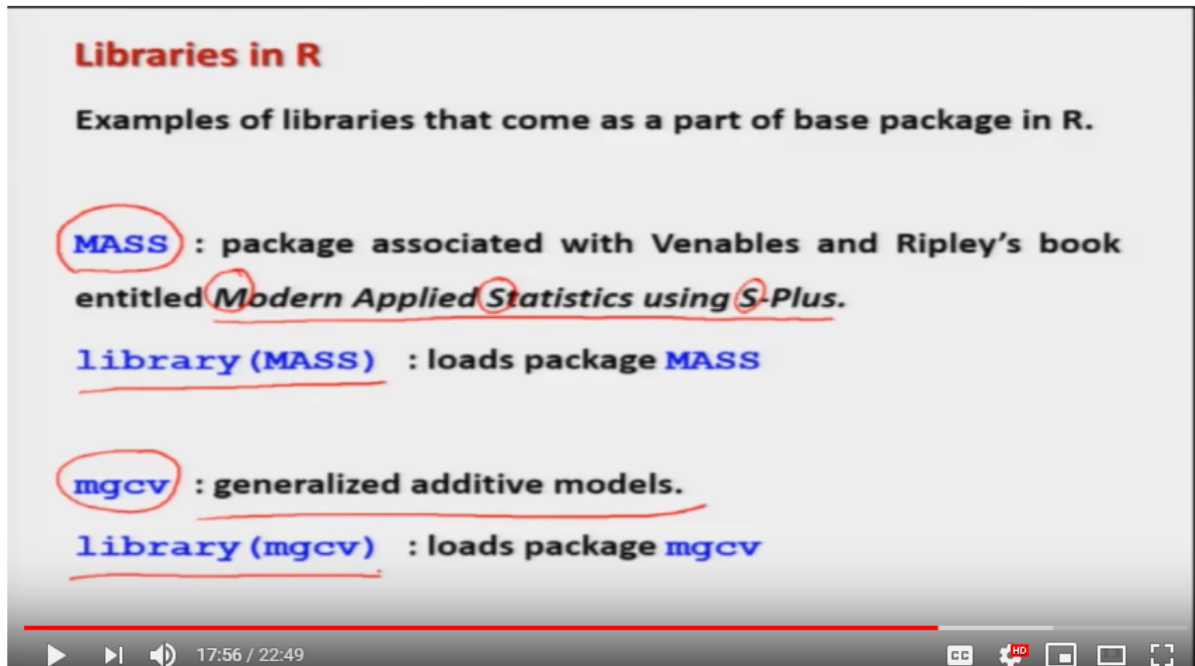
Refer slide time :(15:44)

Libraries in R

Examples of libraries that come as a part of base package in R.

MASS : package associated with Venables and Ripley's book entitled *Modern Applied Statistics using S-Plus.*

`library(MASS)` : loads package MASS

mgcv : generalized additive models.

`library(mgcv)` : loads package mgcv

17:56 / 22:49

I used to hear another command here say library and does syntax is very, very simple it is simply here library and inside the bracket you have to write down the name of the package, which you would like to use for example if I want to use the package cluster then I have to load the library as library. And inside the bracket I have to write here cluster and similarly if I want to use the package ggplot2 then I have to write down here the command library and inside the bracket I have to write down here ggplot2 and similarly if you want to use the package graphics it means again just write library and then you have to use here the graphics inside the brackets. So, this is how you can install package and this is how you have to load the library before you start using it .There are some libraries which come as a part of the package in R for example there is one very popular library which is here mass MASS in capital letters actually MASS means modern applied statistics using s plus this is a book which was written by Rand bliss and wearable's and replay and this mass corresponding to the first letter of M say from modern and a from applied as from statistics and s from s plus s place is in the software so and then s plus n are they are very, very similar .So, so in case if you want to use the data set or the libraries to use in that book you have to use this package mass by writing the command library. And inside the bracket capital m capital it and capital SS and for that you need not to download it because, this package is already available in the base package and similarly there are say the different types of libraries for doing specialized job for example if you want to use the generalized additive models then there is another library say MGCV and in order to use the general I additive generalized additive models you use the command library. And MGCV so, this is how you can install a package and can use the library. And in case if you want to take some help or if you want to know what are the contents of the library what is contained in that thing then you simply have to use a function help

Refer slide time :(18:13)

Contents of Libraries

Use `help` function to get the detailed contents of library packages.

Here is how we find out about the contents of the `cluster` library:
`library(help=cluster)` returns the following:

```
          Information on package 'cluster'
Description:

Package:                cluster
Version:                1.14.4
Date:                   2013-03-26
Priority:               recommended
Author:                 Martin Maechler, based on S
                        original by Peter ... ...
... ... ...
```
followed by a list of all the functions and data sets.

Help and suppose I want to know what are the contents of the libraries see here cluster. Right? So, I simply have to type here library and inside this brackets you simply right here help, help is equal to see her name, name of the library which is her cluster and once you do it this will give you here different types of information here that what is the package and what is the version? What is the date? When it was incorporated? What is the priority well? And who wrote this one? And there will be many that information availability

Refer slide time :(18:53)

Contents of Libraries

```
> library(help=cluster)
> |
```

```
R Documentation for package 'cluster'

                    Information on package 'cluster'

Description:

Package:          cluster
Version:          1.14.4
Date:             2013-03-26
Priority:         recommended
Author:           Martin Maechler, based on S original by Peter
                  Rousseeuw <rousse@uia.ua.ac.be>,
                  Anja.Struyf@uia.ua.ac.be and
                  Mia.Hubert@uia.ua.ac.be, and initial R port by
                  Kurt.Hornik@R-project.org
Maintainer:       Martin Maechler <maechler@stat.math.ethz.ch>
```

And this for example here you can see here a screenshot of the same thing for example if you want to see it here I can show you here that if I simply try to show you here I can just copy this command

here. Right? So, you can see here that this is these are the details about the package cluster. So, I have simply and there is a lot of information complete information about the package. And this is what I meant when I said that R is not a black box means you have to complete details. Right? And if you want to know the you know what is the programming? Of this package you can also know about it. So, let me come back here and here you can see this is a screenshot which I shown you here

And after this yeah this is very important now some final words about using the R software that whenever you are trying to start a new programming you will always be using some variables, which are given by some names like as X Y Z ABC and so on. So, it is possible that today you are trying to write down a program in which you have used the variable name say x and y and say after some time you write another program in which you are using the same name x and y. so, what will happen? That as soon as you define the new variable x and y the, the earlier x and y will be raised or it is also possible that some of your friend comes with the same computer and he defines x and y in a different way so, the way you had defined earlier x and y will be raised. So, it is always a better option that you try to remove the names before you leave or we are before you start a new program. So, in order to remove a variable the command here is RM and inside the bracket you need to write down the name of the variable for example if I have three variables here see here X say Y and say Z then I simply need to write here R M and inside the bracket you have to write X Y Z which is here. And once you do it this will remove the variables X Y Z and then you can use here one of variable at a time or you can use more than one variable at a time.

How to quit in R

Type q() to quit R.

> q()

R Console

> q()
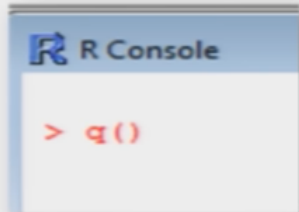
The cut is your choice and simply once you are done and if you want to close the session you want to quit R then the command here is Q and you have to write a two parenthesis. So, once you write Q and opening and closing bracket you will come out of the R program. So, now I will stop here and I would request you that you please try to have a quick look on the basics of are the commands in case if you have done it earlier that will help you and he will see you in the next lecture till then. Good bye