Essentials of Data Science with R software - 1

Professor Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur

Lecture No. 02

Installation and Working with R

Hello friends. Welcome to the course, Essentials of Data Science with R software 1 in which we are going to talk about the topics of probability theory and statistical inference. So, from this lecture, first I would like to introduce you with some basic topics of the R software. Why? You see, as soon as I say that I am going to use the R software in this course, then there is a big confusion among the minds of the students that what are the things that you will require from R to understand the topics in this course.

Well, R is a huge software that is a big software ,that the vast, there are so many topics. So, people get some time confused that out of that long list of the topics what exactly do you want. So, in this lecture and in the next couple of lectures, I will very quickly try to give you the important concepts and topics which are needed from R to understand the topics which are covered in this entire course. So, these are very simple things. These are not the new things and I expect that you might be knowing most of them, but I personally feel that if I do not tell you those things then you may have a sort of confusion in your mind.

Well, after taking these lectures, you may feel that it may not be required but if I ask you at this moment that what are the topics which are needed, then you do not have the answer. So, we are going to find out the answer of this question that after a couple of lectures, you must not complain that okay, these are the topic that we knew and why I have covered in this course, but they will help you in giving you a quick refreshing revision of the topics of R software. So well, I will not be going into much detail. I will not be explaining you in much detail, but my objective is only to tell you the list of the topics, but I will be explaining you very briefly.

Well, in case if you have any problem in learning the R software means on this MOOC, on this NPTEL website, I have one more course on introduction to R software whose lectures are available and if you wish, you can have a quick look or quick revision or a quick learning from those video lectures and I am sure that they will give you a sufficient background to initiate a learning of any statistical topic using the R software. So, in this lecture, my basic objective is that suppose that there is a student who does not know even how to download the software and the student does not have any information what is called as an R software.

So, I am just trying to give you here a very quick and brief introduction to the R software and then in the forthcoming lectures, I will be trying to take up some elementary operations which are needed. Well, in case if you are trying to learn the topics in statistics or data sciences, you will always be handling with the data. What is data? Data is only a numerical value. Now using those numerical values, you are trying to do different types of mathematical operations on it and then you are trying to learn how to take the correct statistical inferences out of that.

Why I am calling it mathematics? You see, means if you say arithmetic mean or median, they are the statistical tool but what are they trying to do? They are trying to do some mathematical operations where when you are trying to find out the arithmetic mean that is simply summation xi upon n which that means you have to first sum all the observation and then you have to divide it by the total number of observations. That is all. So, this is the mathematical operation.

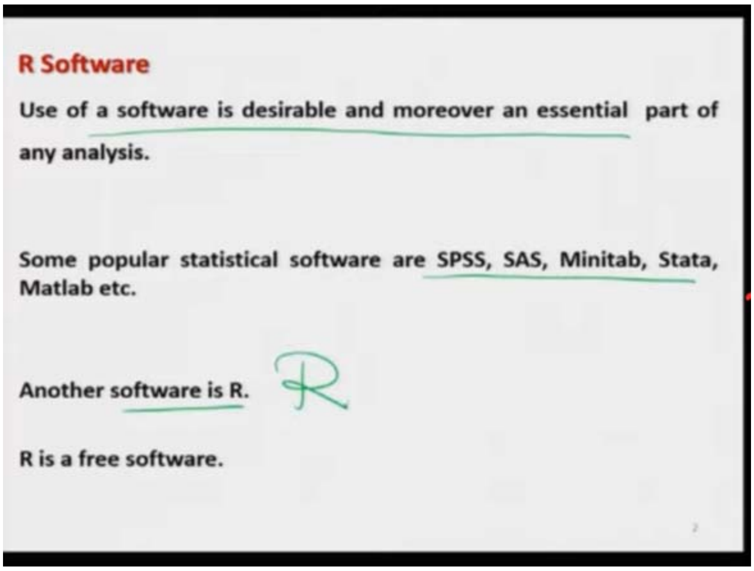
So, whenever you are talking of the statistical tools, that is essentially tool which is based on the concept of mathematics and in order to compute that mathematical quantity, you require the or you need to follow the rules of mathematics. You need to input the numerical values into that mathematical tool so that you get an outcome of the tool and for that we need the help from a software.

Now there are various software. Some are paid, some are unpaid, some are free and so on but I personally believe that as a student in case if there is a software, which is freely available, whose cost is nil and you can always update it according to the latest update available, what can be better than this? So, from this point of view, I have chosen R. R is a software which is a completely free software and beside those things, it has enormous capabilities. R has enormous capabilities means at least if you are talking from the statistic point of view, I do not think if there is any other free software which is handling so many statistical topics.

You want to do time series analysis, you want to do analysis of variance, you want to do financial analysis, you want to do design of experiment, you want to do testing of hypothesis, you want to do econometrics, you want to do Bayesian computation, you want to do simulation, you want to do or you want to apply the Monte Carlo methods-R is there and even for the mathematical operations, if you want to do numerical integration and other things, R is available. So, that is the precise reason that I personally choose this R for this course.

Well, in case if you want to choose any other software, I have no problem, but the only thing is this you have to be sufficiently good enough in that software so that you can handle it. The role of the software comes only to help you out that in case, if you want to find the value of a mathematical function, you should know how to use it. So, with this point of view means I start or I begin this lecture which is just going to give you a basic information about the R software. How R was developed and what are the basic operations? So, let us begin our lecture. So, now in this lecture we are essentially going to talk about that how to install and how to work with the R software.

(Refer Slide Time: 7:46)



R Software

Use of a software is desirable and moreover an essential part of any analysis.

Some popular statistical software are SPSS, SAS, Minitab, Stata, Matlab etc.

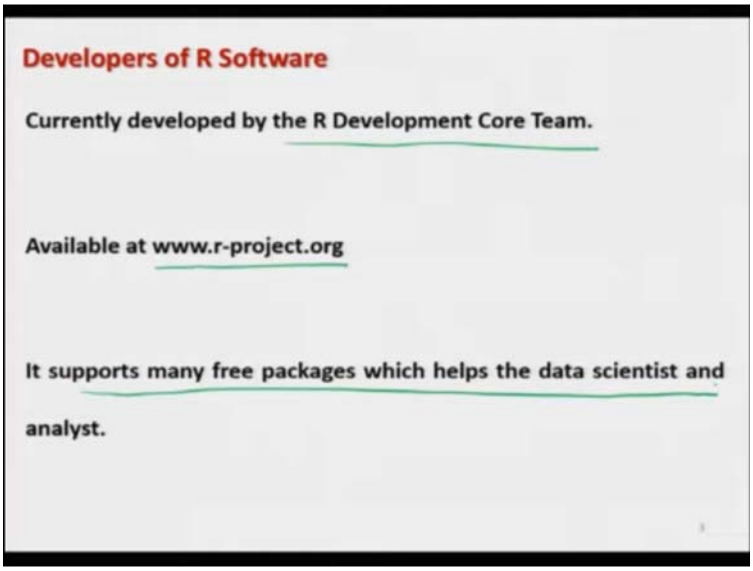
Another software is R. R

R is a free software.

So, we know of that whenever you are trying to do any type of computation or simulation, the use of software is desirable and moreover, it is an essential part of any analysis and in case, if you try to draw from the statistics and mathematics point of view, then there are some popular statistical software like as SPSS, SAS, Minitab, Stata, Matlab, etcetera. Well, please do not think that I am trying to write down these names here so I am trying to do is sort of an advertisement.

Well, I am just informing you and just like them, there is another software which is called as R which is capital R. The main difference between these software and R is that R is a free software. Here, you do not have to pay any cost, you do not have to pay any amount of money to anyone.

(Refer Slide Time: 08:37)



Developers of R Software

Currently developed by the R Development Core Team.

Available at www.r-project.org

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

This R software is being developed by R Development Core Team. So, that is a group of people which is called as R development Core Team and this team there are various people from all over the world from different academic institution, industry, etc. and they are all trying to work together to develop this software and the software is freely available at www dot r-project dot org and one can download the package directly from here and this R software supports many free packages and which helped the data scientist and an analysist.

So, what do you mean by free package? The meaning of free package is that you see, in case if you have a software and suppose you want to do each and everything from a single software, you want to do the financial time series, you want to do the time series, you want to do design of experiments means everything. Well, they are the part of any statistical analysis. Somebody may like to do an econometric analysis also. So, now the question is that for each of the topic you need a program.

Now, in case if you try to create a program for all the topics in the statistics, the size of the program will become very large and it is also not necessary that every user will need all the topics. It may be possible that somebody who is working in the design of experiment he or she may like to do the or conduct the analysis of variance, but then another person who is working in econometrics, that person may not be interested in the design of experiment but then in case, if all those components are immediate into a single software, then the size of the software will become big and then it will have its own complications. So, that is why this R software has been divided into two parts.

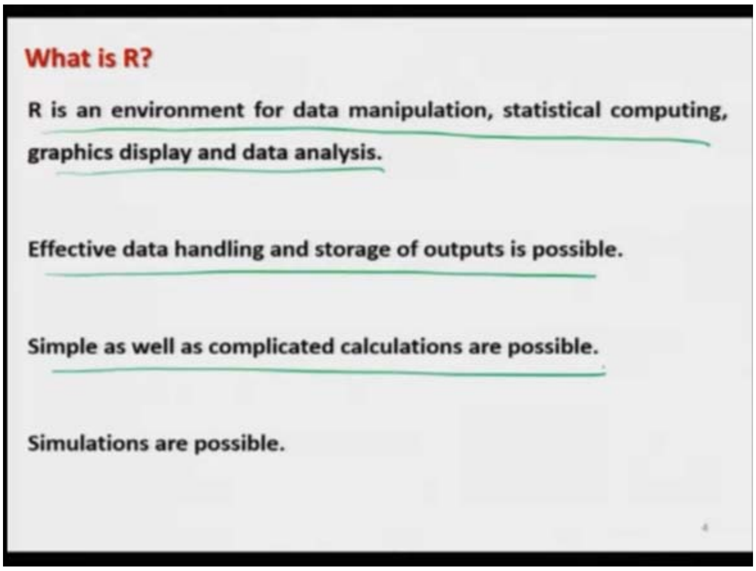
The one part, which contains the basic essential parts like a which is the part of the base package for example, arithmetic mean. That is such a tool which people expect that everybody will need it or most of the people will need it but there can be a tool on design of experiment which everybody may not need it or they can be a time series process which everybody may not need it. So, that is why they have created the package in which all the basic tools are there and that is called as a base package and second part is the separate packages.

People have developed different types of programs for conducting different types of analysis. Somebody has developed the program for the time series analysis, somebody has developed a program for say spatial analysis, etc. So, now these different tasks are embedded into different packages. Packages means you can say in very simple words different programs. So, whenever you want any package or any program, you can simply go to the website of the R and you can download it from there.

Now when we are talking about these packages so, there are some packages which have been developed by the R Development Core Team, but there is a very good advantage of R that even if you are trying to develop a new tool, new statistical tool then you can also create your own program. You can submit it to the R Development Core Team. I am sure that they will try to take it. They will try to discuss it with the experts, they will try to scrutiny it and if everything is fine, they will upload it on their website so that if anybody wants to use your tool which you have developed, the person can simply download that package from the website and can use it directly. Means everybody will not be interested in doing the programming herself or himself. So that is called the package.

So all these packages are freely available. Some of the packages are the part of the base package of R and some of the packages are contributed, some of the packages have been developed by the R Development Core Team and some of the packages are developed by individual academicians from all over the world. So, this is what I meant.

(Refer Slide Time: 13:30)



What is R?

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

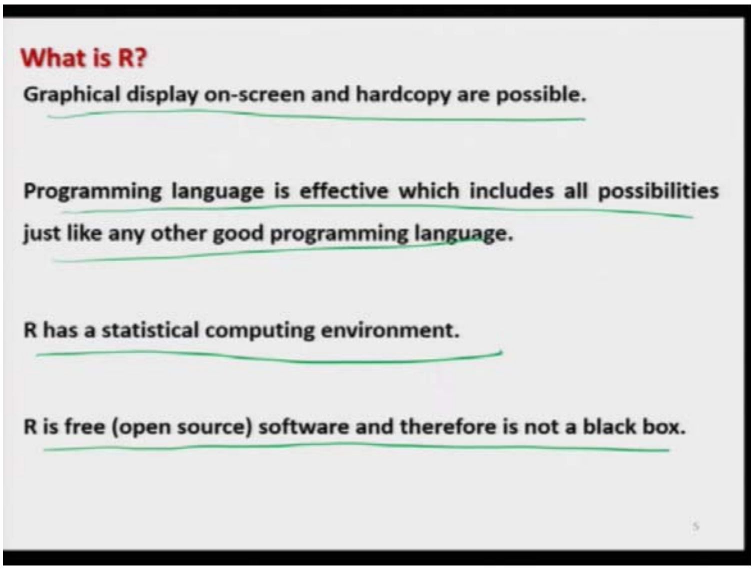
Effective data handling and storage of outputs is possible.

Simple as well as complicated calculations are possible.

Simulations are possible.

So now, in case if you try to think about the R software. R is an environment for data manipulation, statistical computing, graphical display and data analysis just like any other software and it provides you are an effective data handling and storage of output is possible and it is just like any other software in any software, you can control the input, you can control the output and similarly the same thing is possible to do in R software also. Simple as well as complicated calculations are possible, Monte Carlo simulations are possible and various types of simulation, whatever you want to do, you can program it and they can be executed without any problem.

(Refer Slide Time: 14: 07)



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.

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

Any software, whenever you are trying to see the outcome, the outcome can be seen on the computer screen as well as some time you want to have the file of the outcome in the soft format it or you also want to have a hard copy of the output. For example, you want to print a graph. So, in R also, graphical display on a screen as well as hard copies are possible. You can save do graphics in different formats like a jpeg file, png file, pdf file, etc.

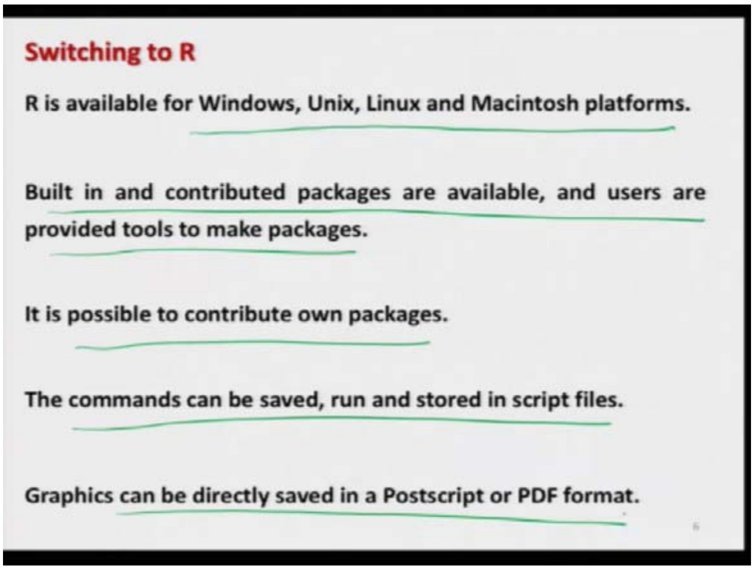
And this programming language, R also has a programming language which is effective, which include all possibilities just like any other good programming language and actually this R has this very strong advantage that R has some built-in packages and R has a programming language also, and that was the reason that R became very popular. There are many statistical software which are working like click click click means you go to the that button, try to click it and then try to input the data and they will give you the outcome but if you want to change the pattern or change something you cannot do it, but in R, you can actually do it.

Suppose if you suppose somewhere you want to find an outcome of the square of the arithmetic mean, I am sure that no statistical software, which is based on click click click type of job, that cannot compute the mean square. But in R, there is a command to compute the arithmetic mean which can be obtained just like by m e a n and if you try to write mean square, you can simply get the square of the mean. So, the arithmetic mean is going to be computed by the R software and now through the programming you can compute the square of the mean. That is what I meant and that was advantage of this R software that you have both the facilities in your hand.

You can use a built-in packages, you can use a contributed packages as well as you can do the programming and also you can do the programming which is from the output of this packages. All these features, they have made this R to be one of the popular software. So, that is why R has the biggest advantage that you can program whatever you want as well as you can use the packages and you can combine them also.

So, this R has a statistical computing environment and as I said that R is a free software, that is an open source software and it is not like a black box. That means if you really want to know what is happening inside the programming of this R software, you can easily know it and you will actually have complete information of any program of any package of whatever you are going to use.

(Refer Slide Time: 17:03)



Switching to R

R is available for Windows, Unix, Linux and Macintosh platforms.

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

It is possible to contribute own packages.

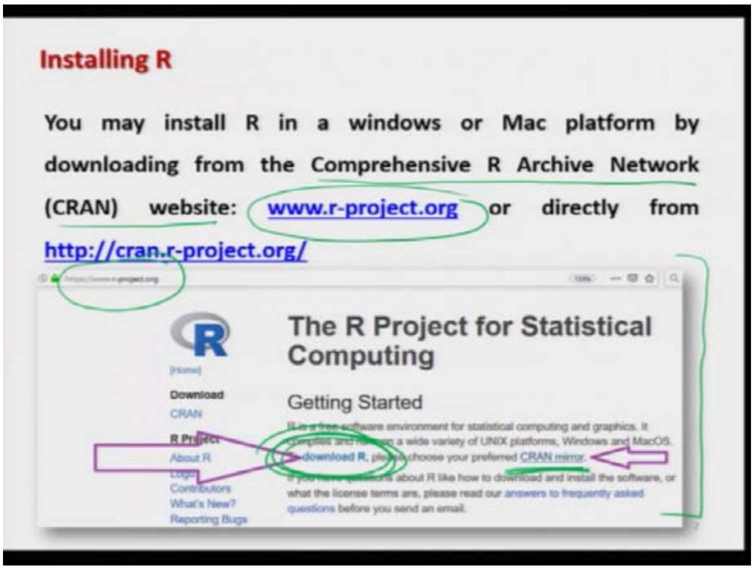
The commands can be saved, run and stored in script files.

Graphics can be directly saved in a Postscript or PDF format.

And this R is available for the Windows platform, Unix platform, Linux platform, Macintosh platform, so whatever you want to do, you can just download it freely from the website r-project.org, and can just install it and can use it and I explained you that is R has some built-in packages, some contributed packages, both are available as well as the users can contribute their own package. They can create their own packages and they can contribute it and those packages can be used by the people all over the world.

So, it is possible to contribute your own packages here and yeah, just like in any software you always try to write a program, you try to run the command, you try to save the command, you want to store the commands in a program file that is called as a script file. So, these things are also possible in R software and as I said, the graphics can also be directly saved in a Postscript, PDF, jpeg, etc. popular formats.

(Refer Slide Time: 18:04)



Installing R

You may install R in a windows or Mac platform by downloading from the Comprehensive R Archive Network (CRAN) website: www.r-project.org or directly from http://cran.r-project.org/

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

So now, after a brief introduction to this R Software, after illustrating the advantages of using the R software, the question comes, how are you going to install it? How are you going to get this software? So far, so as I said this R software is available for different types of platform and this is available from the website www dot r-project dot org and this website has something like what you call as CRAN websites. So, CRAN means Comprehensive R Archive Network. What does this mean? CRAN means actually what happened that this R became a popular software and different people all over the world, they were trying to access it.

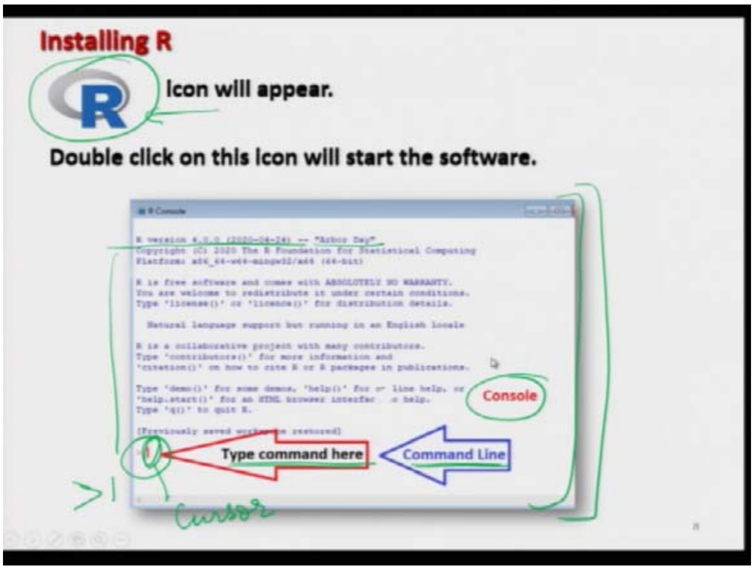
So, now obviously if you try to upload the R software and this package is only at one place, only on 1 server and everybody from all over the world, they are going to download it from there within the load on that server will become too heavy and there is a possibility that the server may crash because it is crossing its capability. So, what people thought that they decided to host the R software in different servers in their academic institution in different countries. So, you will see that every when you want to install a package or install a software, it will ask you from which of the websites you want to download it.

Those names of the websites, they may be located in Australia, they may be located in USA. They may be located in Germany and so on and in different countries. So, you can just click on that any of the link and you can download it. One doubt, which I would like to clarify here that sometimes people think that if they try to say that, you try to download this software from this website from this University, from this institute, this is better. This is only a myth because the software is the same which is uploaded at different places.

Sometimes people say that you try to download it from the country, which is close to your country, right, but do you think that with this internet does it make any difference that you are downloading a file from say America, Germany, Australia or any neighbouring country of India. It does not make any difference practically means I do not know if it has some very deep technical interpretation but in general as a common sense, common user, it will not make practically any difference whether you are downloading the R software from Australia, from an African country, from any European country or from any of the American countries.

So, you can download it from any of the website and the collection of these website that is called as CRAN – Comprehensive R Archive Network. So, what do you have to do? You simply have to go to the website www.r-project.org on any this say, this Mozilla, internet or say, Safari or any of the program that supports the internet, and then from there, for example, if you go to this r-project.org, you will get here a page like this. I have given you are a screenshot, and then you will see here, there is something like here download R. You just click over there or it will also give you that if you have a choice for a particular CRAN mirror, so then you click over here and after that it will go over to that side and then you can just download it.

(Refer Slide Time: 22:03)



Installing R

Icon will appear.

Double click on this icon will start the software.

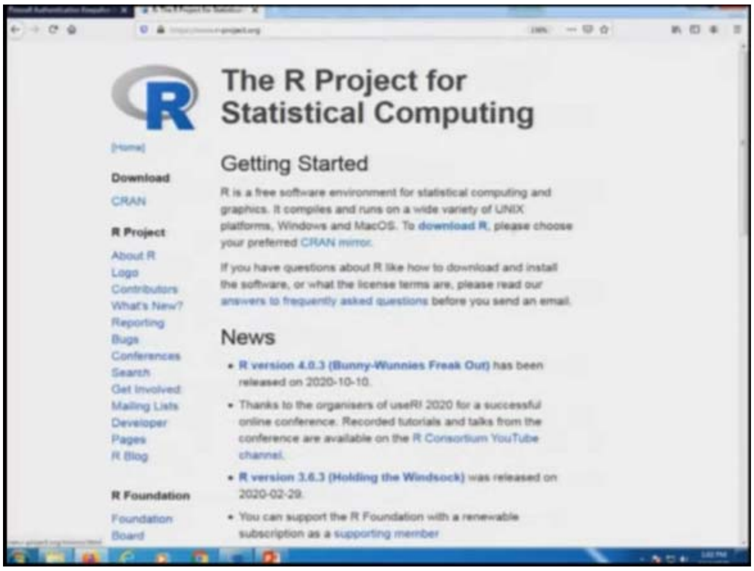
R Console

```
R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

>|
```

So, now once you get that software from there, you just download it save it on your computer and then just try to double click on the package and then try to read the instruction and just keep on double-clicking it or clicking it and it will be installed on your computer and once it is installed, you will see an icon like this one. This is the icon for R software and if you try to double click on this software, it will give you this type of window, it will give you hear all sorts of information.

This is the R version 4.0.0 and its name is R body so on different types of things and you will see that here there will be a sign here like greater than or like this. Actually, this is the place where we type our commands and this is called as command line. This vertical line means this is the cursor and here you have to type your command and this entire window where you are going to type your command, this is called as console. So, before going forward, before moving forward let me try to show you this on the internet that how are you going to do.

(Refer Slide Time: 23:31)



The R Project for Statistical Computing

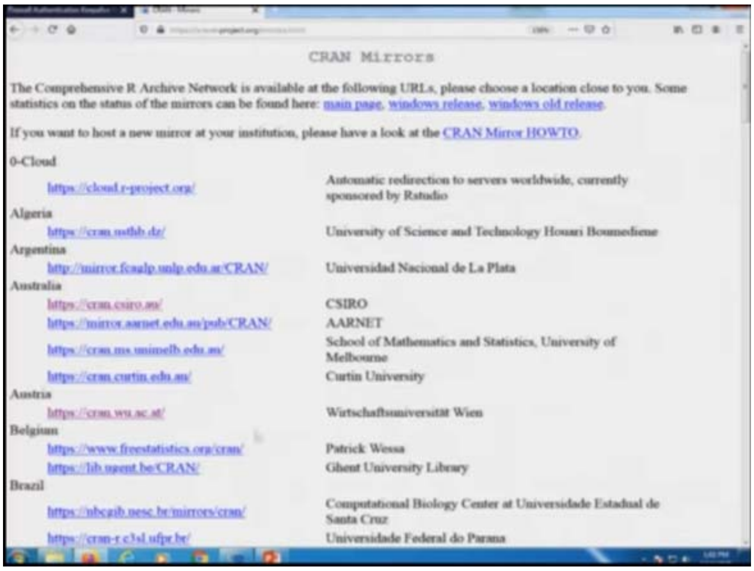
Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- R version 4.0.3 (Bunny-Wunnies Freak Out) has been released on 2020-10-10.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the R Consortium YouTube channel.
- R version 3.6.3 (Holding the Windsock) was released on 2020-02-29.
- You can support the R Foundation with a renewable subscription as a supporting member

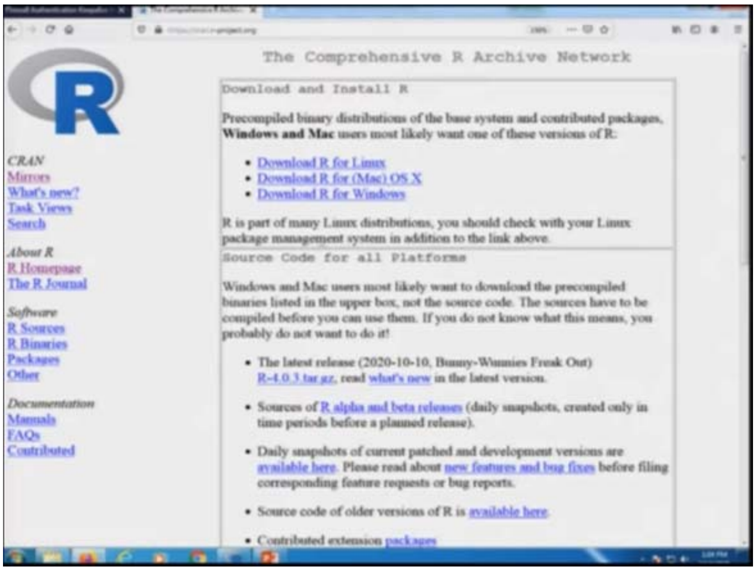


CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: main page, windows release, windows old release.

If you want to host a new mirror at your institution, please have a look at the CRAN Mirror HOWTO.

0-Cloud	
https://cloud.r-project.org/	Automatic redirection to servers worldwide, currently sponsored by Rstudio
Algeria	
https://cran.usthb.dz/	University of Science and Technology Houari Boumediene
Argentina	
http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
Australia	
https://cran.csiro.au/	CSIRO
https://mirror.aarnet.edu.au/pub/CRAN/	AARNET
https://cran.ms.unimelb.edu.au/	School of Mathematics and Statistics, University of Melbourne
https://cran.curtin.edu.au/	Curtin University
Austria	
https://cran.wu.ac.at/	Wirtschaftsuniversität Wien
Belgium	
https://www.freestatistics.org/cran/	Patrick Wessa
https://lib.ugent.be/CRAN/	Ghent University Library
Brazil	
https://nbcgib.uesc.br/mirrors/cran/	Computational Biology Center at Universidade Estadual de Santa Cruz
https://cran-r.c3sl.ufpr.br/	Universidade Federal do Parana



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

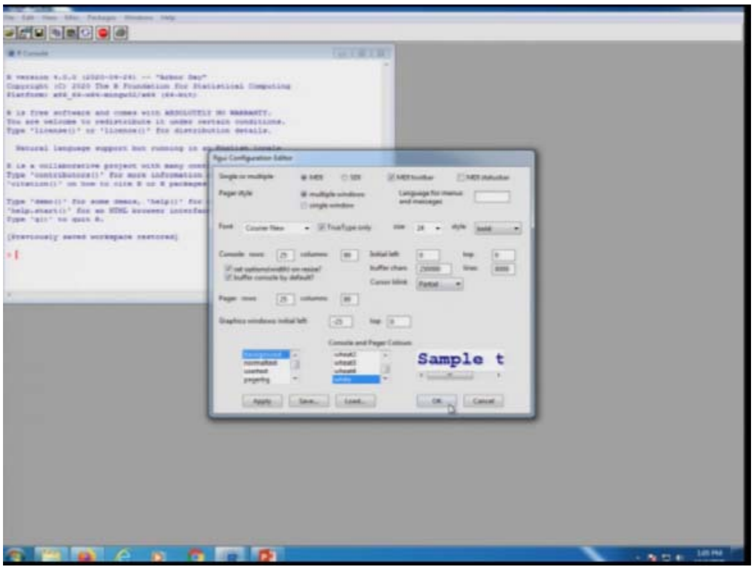
- The latest release (2020-10-10, Bunny-Wunnies Freak Out) R-4.0.3.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

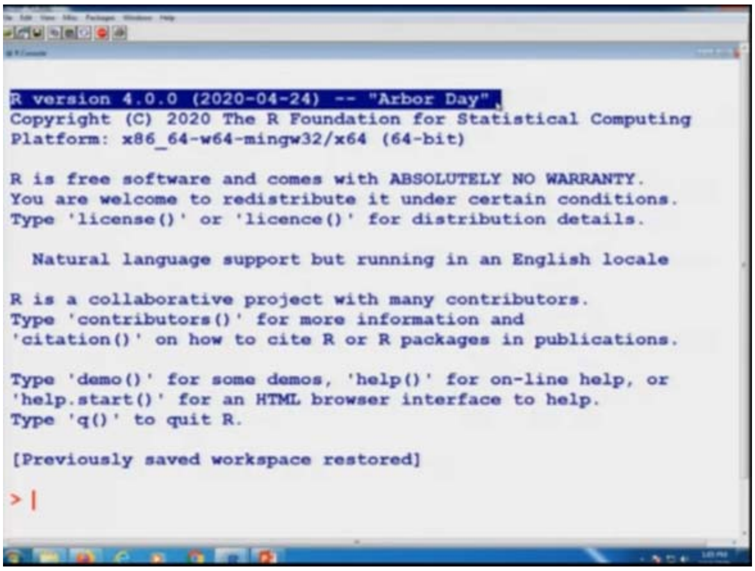
So if you simply try to click here this r-project.org, you will get here this type of website you can see here and you can see here, this is the see here the button- download R or here you can say CRAN mirror and if I try to press here at download our it will give you hear different types of this website as I told you because this is from Algeria, this is from Argentina, Australia, Austria, and you can see here there is a long list over here and even if you can see here that in India also, this R software is hosted at National Institute of Science Education and Research, that is NISER in Bhubaneshwar in the Orissa state.

So, well these says CRAN mirrors will always be getting changed and so at this moment when I am trying to record this video, these are the available CRAN mirror. So, if you try to just click on anything, for example, if you want to go to this this homepage in Austria now in case if you try to download the R software from CRAN mirror in Austria, then you can just click over here and then you will see here that you come to here, this CRAN mirror and here you can see here download R for Linux download R for Mac, download, R for Windows and if you try to click here you will get the software on your computer and you can see here that other older these versions of R are also available.

So anyway, so those things you can do here and once you try to install this thing over here, then you will get here this R console.

(Refer Slide Time: 25:12)

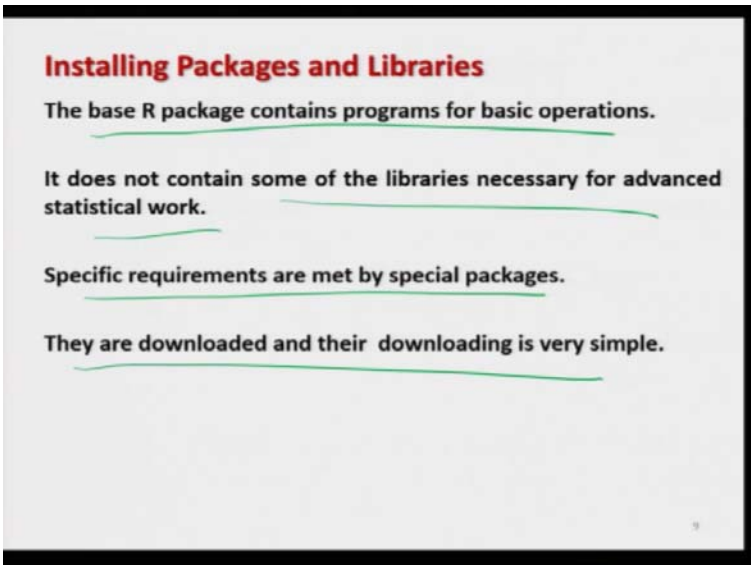




```
R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> |
```

You can see here, this is here the R console over here and before that you can see here on your computer there is an icon here. Here I can show you here like this one, click it here, this will give you the same thing. If you want to increase the font size, etc. here, you can do it here. I can make it either font size to be 24, I can make it here bold. I can do different types of things and you can see here, this is the same thing which I have shown you here that R version 4.0.0 from Arbor Day, etc., etc. So, now let us come back to our slides and try to continue with the lecture.

(Refer Slide Time: 25:54)



Installing Packages and Libraries

The base R package contains programs for basic operations.

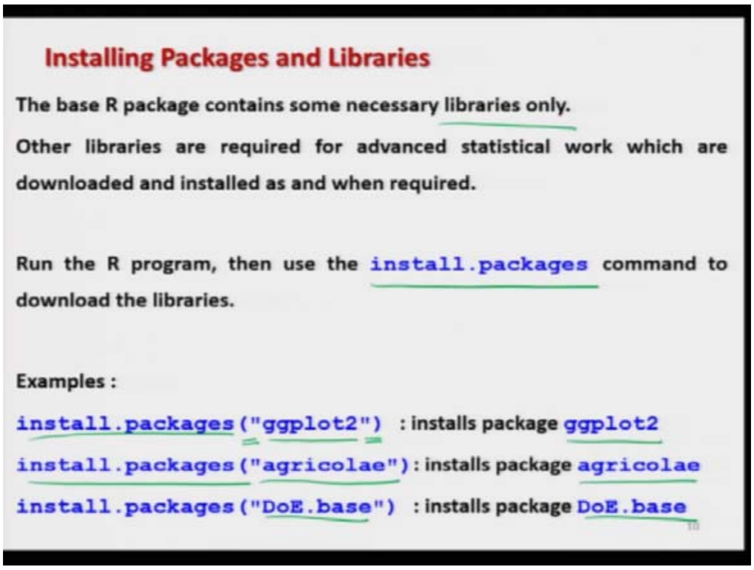
It does not contain some of the libraries necessary for advanced statistical work.

Specific requirements are met by special packages.

They are downloaded and their downloading is very simple.

So, now I believe that you can install the R software on your computer and now you are, I mean, there should not be any problem now in downloading the package also. So, as I said you, the package which you are trying to download here, this is the base package and it has all the features and program for the basic operation and it does not contain some of the libraries which are needed for some advanced statistical work or your need what you really want to do and those specific requirements are met by special packages and they are actually such packages are downloaded from the website of the R software and the downloading is very very simple.

(Refer Slide Time: 26:36)



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 as and when required.

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

Examples :

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

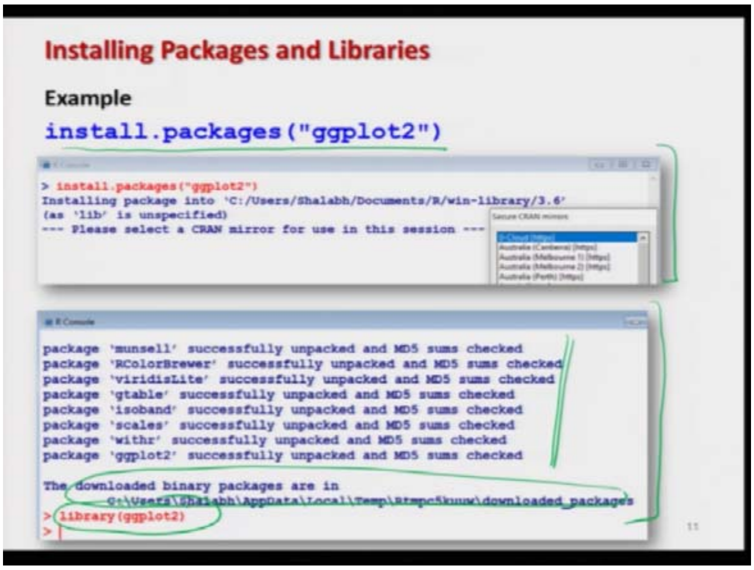
`install.packages("agricolae")`: installs package `agricolae`

`install.packages("DoE.base")` : installs package `DoE.base`

I will simply try to give you here an idea that in case if you want to download a package, the package will contain some libraries. So, these libraries are going to help you in executing a particular type of task. So, if you want to install a particular package then the command that you know is install dot packages – Install dot packages and after this, you have to just execute this command on the R console on the command line, you simply have to type here installed dot packages and if you want to install the package say ggplot2, this is the package for creating different types of graphics so what you have to write inside the parentheses, you have to just write inside the double quotes ggplot2.

After that if you simply try to just say click click click and so on, it will install the package ggplot2 and similarly if you want to install the package say agricolae or say DoE dot base that means they are the packages which are used in the design of experiment and analysis of variance then you simply have to use here the command install dot packages and then within parentheses within double quotes, you have to write down the name of the package like as agricolae, or DoE.base and then it will install that a package on your computer.

(Refer Slide Time: 28:06)



Installing Packages and Libraries

Example

```
install.packages("ggplot2")
```

```
> install.packages("ggplot2")
Installing package into 'C:/Users/Shalabh/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
```

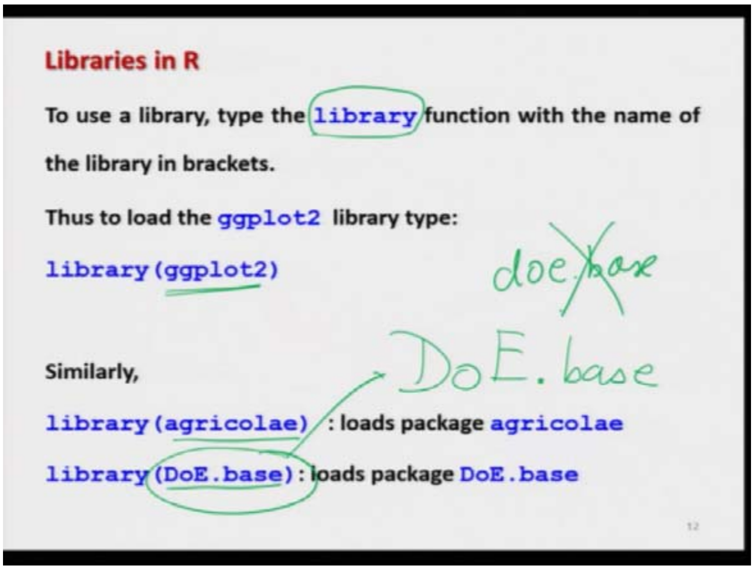
Secure CRAN mirrors
0-Cloud [https]
Australia (Canberra) [https]
Australia (Melbourne 1) [https]
Australia (Melbourne 2) [https]
Australia (Perth) [https]
R Console
11

```
package 'munsell' successfully unpacked and MD5 sums checked
package 'RColorBrewer' successfully unpacked and MD5 sums checked
package 'viridisLite' successfully unpacked and MD5 sums checked
package 'gtable' successfully unpacked and MD5 sums checked
package 'isoband' successfully unpacked and MD5 sums checked
package 'scales' successfully unpacked and MD5 sums checked
package 'withr' successfully unpacked and MD5 sums checked
package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Shalabh\AppData\Local\Temp\Rtmpc5kuuw\downloaded_packages
> library(ggplot2)
> 
```

It means if you want to suppose try to install the ggplot2, then you try to type here install dot packages and ggplot2, you have to write it in double quotes, within the parentheses and first screenshot you will get here like this. It will give you the choice for the CRAN mirrors. Once you choose the CRAN mirror that from where you want to download it, it will just start downloading it and installing it and you will get here a window like this one. You will get these types of message.

You do not have to bother about these messages because they are simply indicating that something is happening to install the package on your computer and finally you will see here this type of message that there is no error message and all the packages have been downloaded. Now after this if you want to use that package, you have to use the command here, library and inside the parentheses, you have to write down the name of the package that you want to use.

(Refer Slide Time: 29:08)



Libraries in R

To use a library, type the `library` function with the name of the library in brackets.

Thus to load the `ggplot2` library type:

`library(ggplot2)`

Similarly,

`library(agricolae)` : loads package `agricolae`

`library(DoE.base)` : loads package `DoE.base`

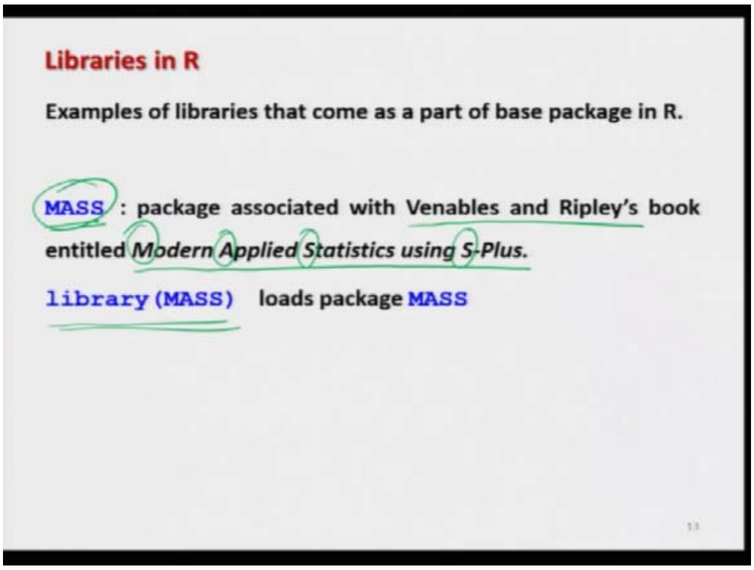
doe.~~base~~

DoE.base

So, one thing what you have to just recall, if you want to use a library, you have to use the function library, all in small letters and then you have to write down the name of the package inside the parentheses but remember here you are not going to use the double quotes, you simply have to write it here library and the package name inside the parentheses. Similarly, if you want to use the package agricolae or DoE dot base, you have to simply write down the name of the package inside the parentheses and you have to write out or you have to use the function library and remember one thing, that this R is the case sensitive.

So for example, if you can see here that in this package DoE base, the D is capital, o is small letter and capital E is used with a dot and after this base is written all in small lowercase alphabets. So, this is what you have to keep in mind means if you try to make it here, only here small d, small o, small e dot base, it will not work. That is what you have to keep in mind.

(Refer Slide Time: 30:15)



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.*

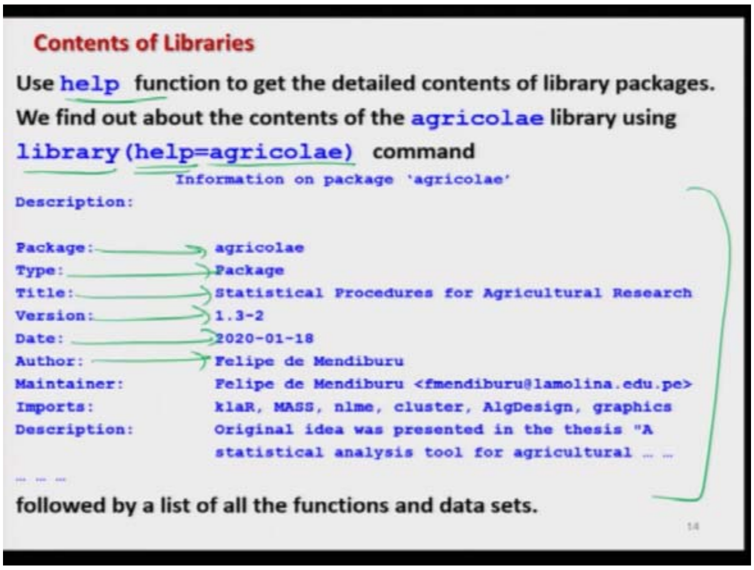
`library(MASS)` loads package MASS

Well, these are the examples, ggplot, agricolae, etc, they are the packages which you are trying to download because they are not the part of the base package of R whereas there are some packages which are the part of the base package and among those packages, there is package here MASS, all in uppercase alphabets. This is a package which is available inside the base package of R and actually this package is coming from a book whose name is Modern Applied Statistics using S-Plus.

So, this MASS means the first letter of modern, first one letter of applied, first letter of statistics and first letter of S-plus. S Plus was an earlier package, which was very popular and R was developed on the same lines of the S-Plus software. So in that book, which was written by Professor Venebles and Ripley's book that was published from Springer. They had used different types of datasets so the same data sets are compiled and they are uploaded inside the R software, the base package of R software and they are available in the package name MASS.

So, if you want to use this package, you simply have to simply write down here library and inside parentheses MASS in uppercase alphabets but the main thing here is that you need not to install it. You need not to download it.

(Refer Slide Time: 31:50)



Contents of Libraries

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

We find out about the contents of the agricolae library using

library(help=agricolae) command

```
                 Information on package 'agricolae'
Description:

Package:         agricolae
Type:            Package
Title:           Statistical Procedures for Agricultural Research
Version:         1.3-2
Date:            2020-01-18
Author:          Felipe de Mendiburu
Maintainer:      Felipe de Mendiburu <fmendiburu@lamolina.edu.pe>
Imports:         klaR, MASS, nlme, cluster, AlgDesign, graphics
Description:     Original idea was presented in the thesis "A
                 statistical analysis tool for agricultural … …

… … …
```

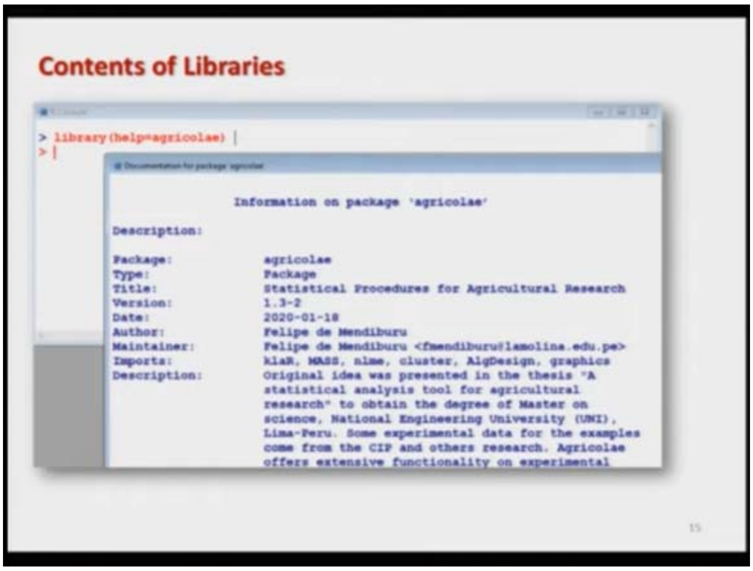
followed by a list of all the functions and data sets.

14

When you install a package or if you want to use a library, as I said that R is not a black box. You can have the complete information about that package or that library and for example, if you want to know something about the package agricolae that that is used for the analysis of variance, you can use here the help function to get the detailed content of the library package.

For example, if you want to know about agricolae, you simply have to write library and inside the parentheses, you have to write help equal to agricolae and once you try to do it, you will get here this type of information that what is the name of the package, what is the type, what is the title, what is the version, what is the date when it was incorporated and who is the author and so on. All those details are there and after this, there is a long list of the of the functions and data set which are available in this package.

(Refer Slide Time: 32:58)



Contents of Libraries

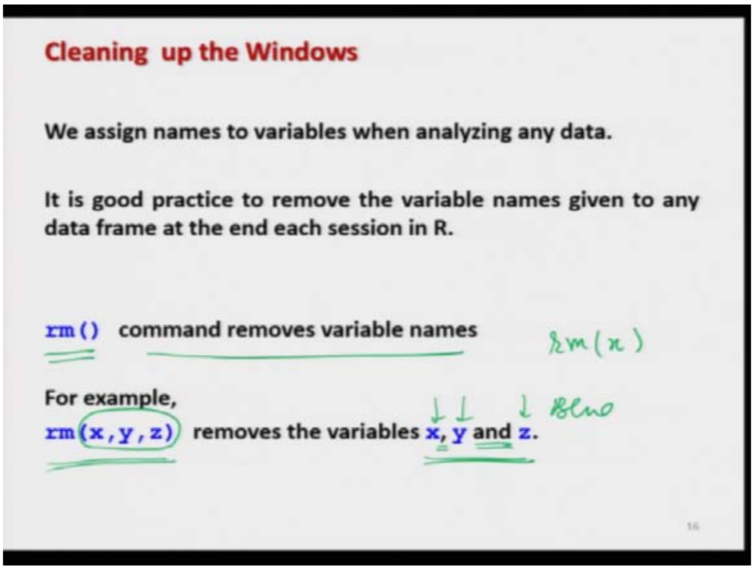
```
> library(help=agricolae) |
> |
```

Information on package 'agricolae'

```
Description:

Package:            agricolae
Type:               Package
Title:              Statistical Procedures for Agricultural Research
Version:            1.3-2
Date:               2020-01-18
Author:             Felipe de Mendiburu
Maintainer:         Felipe de Mendiburu <fmendiburu@lamolina.edu.pe>
Imports:            klaR, MASS, nlme, cluster, AlgDesign, graphics
Description:        Original idea was presented in the thesis "A
                    statistical analysis tool for agricultural
                    research" to obtain the degree of Master on
                    science, National Engineering University (UNI),
                    Lima-Peru. Some experimental data for the examples
                    come from the CIP and others research. Agricolae
                    offers extensive functionality on experimental
```

Beside those things, means if you go to the website of R project and if you try to look the package agricolae, you will get a complete documentation of this package and if you try to do it on the R software, you will get here this type of detail. So, I am just trying to show you here. I am not going into the details, but definitely just to make you convince and confident, I am trying to give you hear their screenshot.

(Refer Slide Time: 33: 13)



Cleaning up the Windows

We assign names to variables when analyzing any data.

It is good practice to remove the variable names given to any data frame at the end each session in R.

rm() command removes variable names

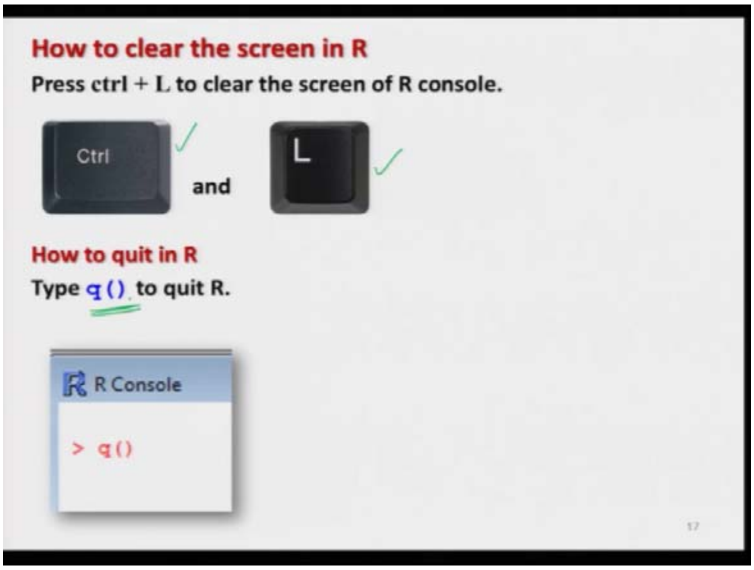
For example,
rm(x,y,z) removes the variables x, y and z.

Now once you have done the programming and then you would like to assign some names to the different types of variable. So, now once you are done with the programming, then you would definitely like to remove those variable from your R console so that we can free the space and they will not create any confusion if you are trying to do a new programming on the same computer.

For example, if you are trying to give a variable name as height and today you are trying to store the heights of students of say, class 7 in your data set and you are trying to do this analysis and suppose after a week somebody else comes to work on the same computer and the person also has the data on the heights of the say, class 10 and suppose that person gives here the variable name as height. Then the earlier name will be vanished or there will be some confusion. So, this is a better practise that when you are done with your programming always try to remove those variables from the console.

So, in order to remove those variables, we have a command here rm. rm command removes the variable name. For example, if you have some variable say x, you simply have to write down here rm x and in case if you have more than 1 variables say, x, y and z then you simply have to write down here rm x, y, z. So, now you can see here that in the beginning, I have told you that how you have to read it. You can see here this x, y and z, they are in blue colour and comma they are in black colour. So, that means you have to simply understand that these are the part of the R command whereas, if you try to see here, this x, y, z as well as comma, they are also in the blue colour. That means you have to type this x, y, z inside the R console.

(Refer Slide Time: 35:09)



How to clear the screen in R

Press ctrl + L to clear the screen of R console.

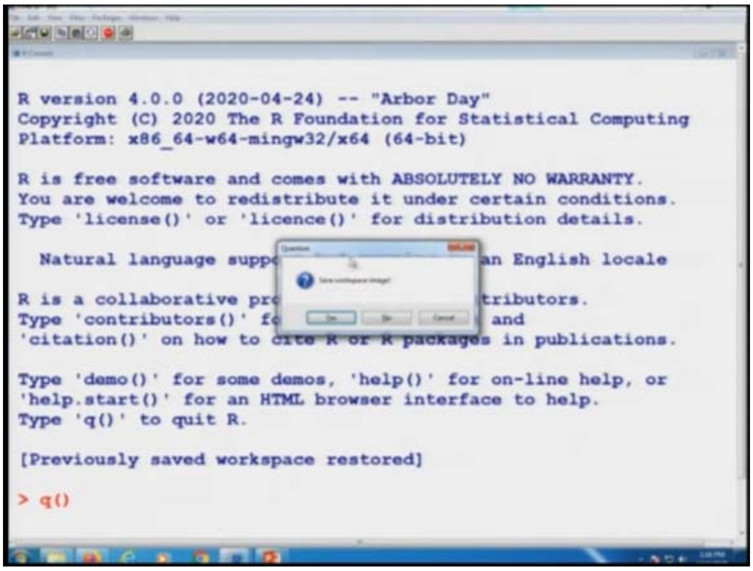
How to quit in R

Type q() to quit R.

```
> q()
```

So, now once you are trying to do something and suppose you have to type something and if you want to clear the screen then the command here is Ctrl plus L. So, that means you try to use this control key here and then while pressing the Ctrl key you try to click on the key L and then this will clear the screen. So, control L will clear the screen and if you want to quit from the R software simply try to use here the command q and write down the parentheses. So, if you try to write down here q parentheses, it will quit.

(Refer Slide Time: 35:49)



```
R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language supp                    an English locale

R is a collaborative pr                    tributors.
Type 'contributors()' f                    and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> q()
```

For example, if you try to see here, if you try to come on the same program over here R, where you have come and suppose you want to hear quit. I simply have to write down here q and this parenthesis and it will say, do you want to save the workspace image. That means whatever you have worked in this question, is it going to do you want to really save it or you want to ignore it. Suppose I say no, I do not want to save it. I will select no and then it will quit from the R software.

So now, we stop in this lecture, I have given you are very basic lecture and some basic information about the R software so that even if you have not done the R software earlier, this will give you some time and some idea that how to begin before we come to the main topics of the statistics. So, I would say that why do not you open your our software on your computer and try to have a quick look, try to revise these commands and try to see what you can do and in the next lecture, I will try to give you some more basic commands which are related to R software that will help you in the further lectures. So, you try to practice it and I will see you in the next lecture. Till then, goodbye.