Foundations of R Software
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 03
Packages and Libraries in R

Hello friends, welcome to the core Foundations of R software. And, in this lecture, we are going to talk about some Packages and Libraries concept in the R software. Now, before we try to understand what is the concept of library or package, let us try to discuss a very normal phenomena that you follow in your life. Suppose, you want to learn something about some topic, what is your first step? The first step is to get a book.

Now, when you want to get a book then again you have two options. First option is that you go to the market or an online store and buy it. Second option is you go to the library that is the physical library means you know at every college, university, school, etc., they have a library where they keep all types of books. Now, buying all the books is practically very difficult nearly impossible. So, what we try to do that we try to go to our library.

Now, try to recall, what happens after that? You will go to the library you will look for the book, then you will bring the book to a counter where somebody will issue that book to you for some time; then, you will bring that book to your home or your hostel or your college. After that, you will study from that book; and once you are done then you will go back to the library and return it. That's a very usual process which we follow.

Now, suppose if any of the condition is violated here, then what will happen? Suppose, you have to buy all the books that is an expensive affair, that is practically impossible. So, you cannot buy, number one. Is it possible for you that without going to the library you can have the book? Not possible, because unless and until you go to library yourself and sign against your roll number or name, you will not be getting the book.

Once you bring the book to the from the library to your home, would you like to keep it for always? No, because if you start doing it you want to learn many topics and then your home will be full of books and it will be very difficult for you to manage them. So, you

would like to bring the book, but after some time you would like to give it back also. So, all these things are necessary. Now, the same thing happens in the R software.

In case if you try to recall when we downloaded the R software that had some built in functions built in, facilities, but with the growing popularity of the R software, there are many things which have been developed by different academicians, programmers, but all of them are not going to be useful for everyone. So, what will happen? People want to use only a couple of topics as you do not want all the books from the library, but you want to use only couple of books.

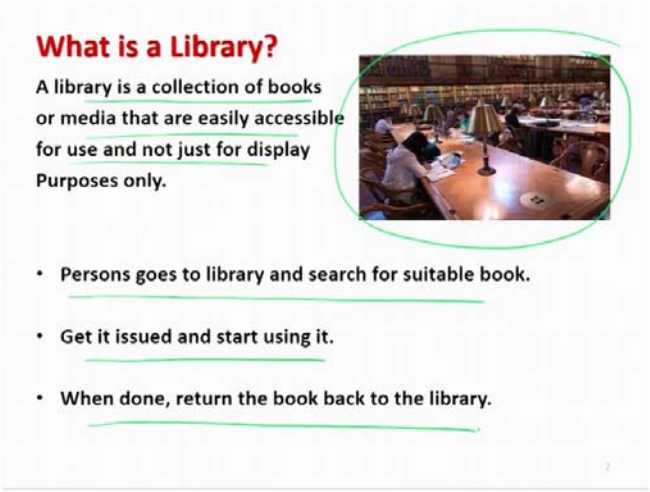
So, similarly in the R package also, you would not like to use all the functionalities. And now, suppose if you try to put all the functionalities in a single package, the size of the package will become too big and it would not be convenient for downloading for uploading; and it will take, say large amount of space on your computer also. So, what people have done? For doing different types of job, they have created different types of packages.

And those packages have been uploaded on the website of the R software. Whenever you want, you can go to the website and you can download them, you can install them, you can work as much as you want and after that you would try to remove them, not removing from the computer, but removing from the from your library. So, that is the same process we follow in the R software. But, how to get it done? That is the objective of this course which I want to discuss today.

So, let us try to begin our lecture and we try to understand these basic fundamental concepts. And there is going to be one difference between my slides and what you will be doing on your computer or what I try to do on this computer on which I am trying to record my video. I have prepared these slides on a different laptop. So, whatever is the outcome that will be related to that laptop. Now I am recording on a different computer so that the contents may be different.

And, similarly when you will be doing it on your computer your contents may be different. So, do not get confused when you try to match the contents and if they are not matching, right. So, with this objective and keeping this point in mind let us begin our lecture today.

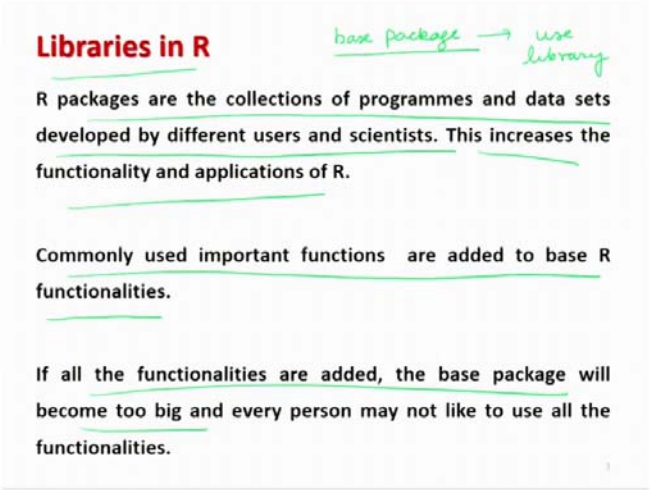
(Refer Slide Time: 06:03)



So, ok so, now we are going to talk about packages and libraries in the R software. So, if you try to recall- how our library looks like, can you see this type of picture? There are so many books, people sitting over there. So, what is happening inside the library? Library is essentially a collection of books or media that are easily accessible for use and not just for display purpose only, right.

So, what happened that if somebody wants a book? The person goes to library and search for the suitable book get it issued and start using it, and when done the book is written to the library, right.

(Refer Slide Time: 06:53)



Now, similar is the concept in the R software also that there are libraries in R also. And, while understanding it, if you keep in mind the usual process of a physical library then possibly it will be very easy for you to understand the concept of library and packages in the R software. So, as we already have discussed couple of times that when this R software was built up, then many people around the world they got associated and they distributed their work.

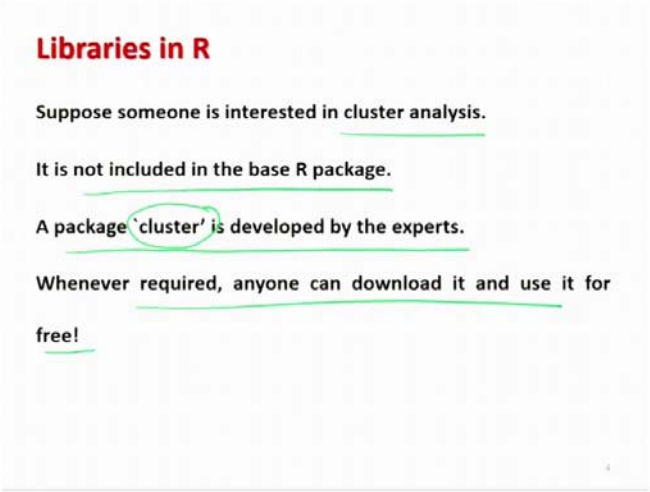
So, some of the basic functions which are commonly used, they were put in together inside the R software and that part is called as base package, right. So, what you download from the website of www r hyphen project dot org that is essentially the base package. And then, if you want to do something more, then you have to use the concept of library, and then you have to get that package and then you have to work for it.

So, these R packages are essentially the collections of program and data sets which have been developed by different users, different scientists, different academicians from different parts of this world. And, this was the biggest advantage of this R software that so many people got involved and they started thinking in different directions that what people want and this increased the functionality and applications of R tremendously, right.

Now, in case if all the functionalities are added to the package then the size of the package will become too huge and it will occupy large amount of space on the individual computers also. So, what was thought? That the commonly used important functions were added to the base package of the R software; and just to avoid that the package it does not become too big, then additional features were given separately, right.

Because, it is obvious suppose, as a statistician I want to use linear models only although I am not interested in the cluster analysis. Similarly, somebody may be interested in the cluster analysis and may not like to use the linear regression analysis. So, why to put both the things together inside the base package?

(Refer Slide Time: 09:46)



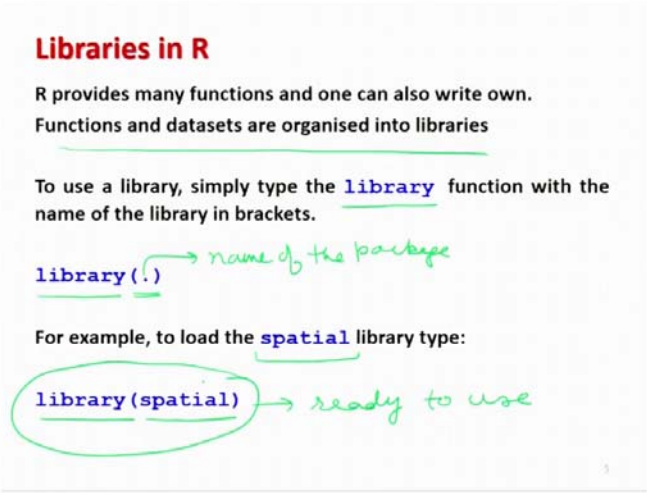
So, that was the concept. And, now for example, if you try to see now as I said that there are some functionalities which are added in the base package of the R software and others are additional. So, for example, let me take an example of the cluster analysis. Cluster analysis is a methodology where we try to group different objects together. Well, that has a statistical background and people use it, but at least I am not interested in telling you what is cluster analysis, but I am just trying to take an example so that I can explain you that how the packages work.

So, this aspect of cluster analysis is not included in the base package of R software. But, in order to conduct the cluster analysis, a package for cluster analysis is developed by the

experts who knows cluster analysis and this package is called as cluster. The name of the package has been given as cluster.

So, now, this is not a part of the base package. So, whenever we want to use it we can download it and use it completely for free. And, similarly you can think of not only cluster analysis, but so many applications of statistics and mathematics which are not needed by everyone. So, whenever they need it they can download it, they can use it just for free.

(Refer Slide Time: 11:17)



So, now the question is how to get it done? And besides those things as we already have discussed that R software has an option that people can write their own functions and this gives very big advantage that different people can write different functions and they can have databases also so that anybody can use it. And, in order to use them, these functions and data sets they have been organized into libraries.

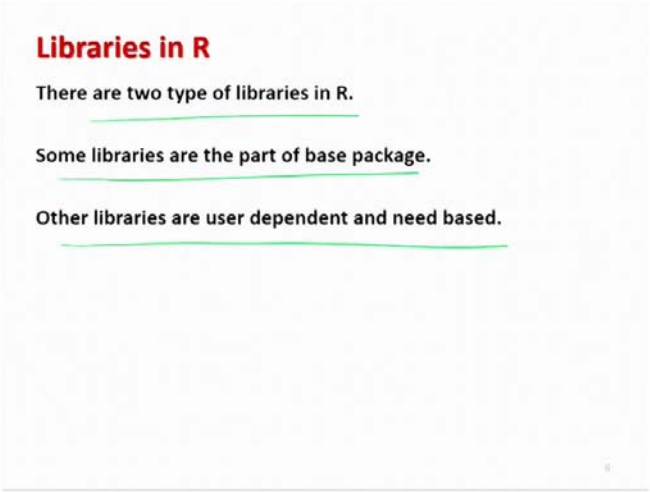
It is just like if you go to a library, there will be a book, there will be a CD, there will be a DVD, there will be now E-books, there will be data sets which are available online, etc. And, you can access them through library. So, similarly in the R software also if you want to use a library, first, you have to inform the software and for that the command is library l i b r a r y all in small letters, lower case alphabets.

And after writing this command, inside the parenthesis you try to write down the name of the package that you want to use. It is just like means if you go to a library and if you want to have a particular book and if you ask the some employee in the library, ok, I want a book. The person will simply ask you what book, which book, then you have to tell the name and then that employee in the library can help you in finding out that book.

So, similarly suppose I ask that person in the employee that I need the spatial book or in our language I want to use a spatial library. A spatial library for example, is used for the spatial analysis and statistics. So, I simply have to type here library inside the parentheses I have to write s p a t i a l, a spatial which is the name of the library. Now, what will happen?

This library will become ready to use. If that package is there and if you don't load the library then you cannot use it. It is just like if the book is inside the library and if you have not issued it you cannot read it, you cannot understand it. So, first you have to get the book issued. So, this library spatial command works just like as if you are getting the book issued in your name.

(Refer Slide Time: 14:07)

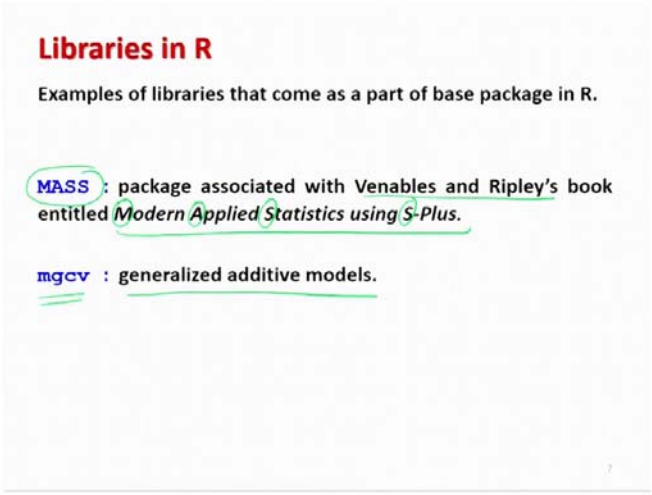


And as I said that there are two types of libraries which are available in the R software. Some libraries which are very common popular and they are commonly used, they have been made the part of the base package of the R software; whereas, other libraries they

are user dependent and they are need based. Wherever you need, you can get it for example, a spatial analysis is a very specialized topic and everybody may not like to use it always; whereas, finding out the arithmetic mean or variance, these are very common things which everybody may like to use it.

So, functions like mean, variance, etc.; they have become they have been actually made the part of the base package of the R software; whereas, spatial cluster analysis etc., they are not the part of the base package, but they have to use to be used separately.

(Refer Slide Time: 15:08)



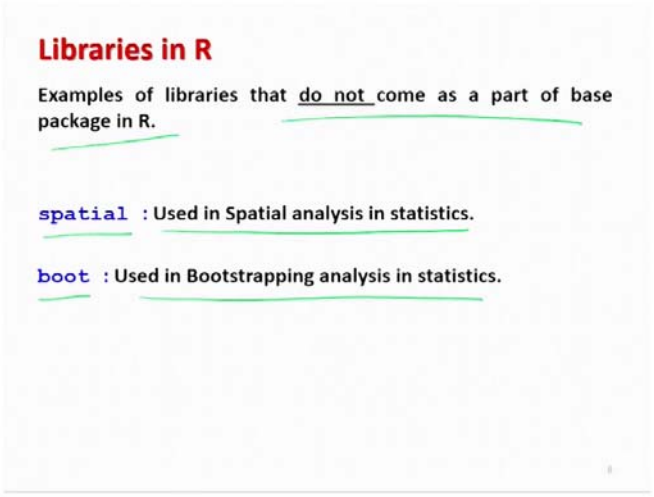
So, some of the libraries they are also the part of base package in R and one of the library which is quite popular is MASS; capital M capital A capital double S. Actually this name is coming from here. Modern Applied Statistics using S-Plus. So, this is here M, this is here A, this is here capital S from statistics and capital as from S plus.

Actually, if you try to recall that earlier we had discussed that this R package was developed on the lines of another software S plus. So, when this S plus was there then Professors Venables and Ripley they wrote a book whose name was modern applied statistics using a S plus. And in that book, they have used some data sets. So, when this R was developed because that was developed on the same lines as the S plus.

So, in order to understand it people thought that ok if they try to use the same data set and because R has got the similar command people will be more comfortable in using

and learning the R software. So, all those data sets commands etc. whatever were available in that book, they have been compiled in this package MASS, right. Similarly, there is another library which is built in which is mgcv, all in small letters lowercase alphabets. This is a library for using the generalized additive models. And then not only these two, but the, but there is a long list which are the part of the base package.

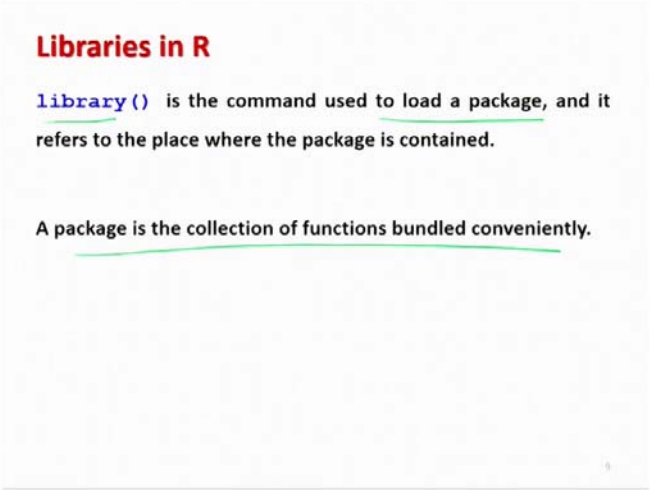
(Refer Slide Time: 16:57)



And two examples of those libraries which are not the part of the base package R say for example, spatial s p a t i a l, all in lower case alphabets. Actually, this is a library which is used for the spatial analysis in statistics. And similarly, there is another library for example, boot b double o t. Actually, this is a library which is used for boot strapping methodology in statistics, right.

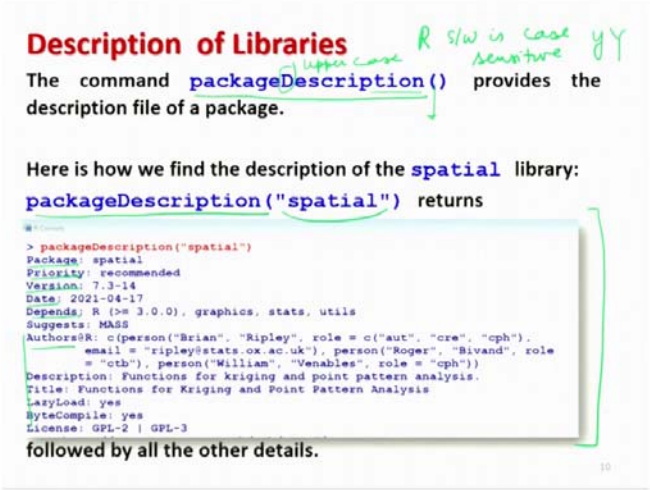
This spatial analysis, bootstrapping, etc. they are different types of methodologies which are used in the statistics. And, now means because I am from statistics background so that is why I am trying to give you example of those libraries and packages which are used in statistics right. So, but as I said there is a long list of such libraries which are available.

(Refer Slide Time: 17:54)



Now, in case if you want to begin, the first step is that you have to use the command library to load a package, right. And, as I said a package is the collection of functions which are bundled conveniently, right. And once you try to use this library command the R will bring the required library or required package from the software to the memory of the computer and it becomes usable right.

(Refer Slide Time: 18:31)



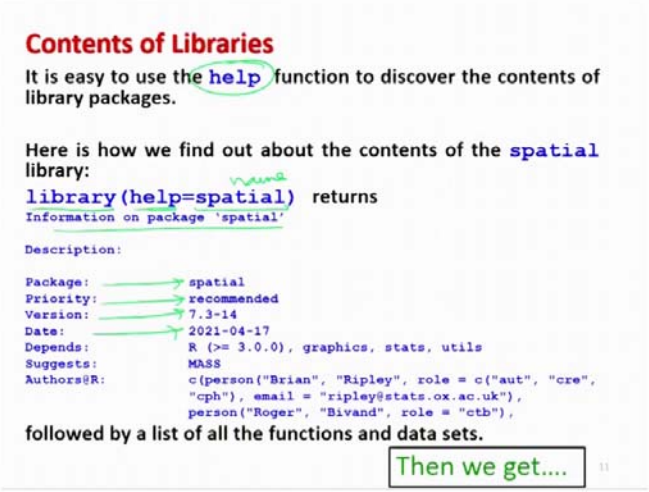
Now, suppose you get a package and you want to see, what is this package, what are the details of this package. So, as we have discussed already that R is not a black box, if you want to have any type of information that is available and one can see it. So, in order to see the description of the packages, we have a command here package description. Now, if you try to see you have to be a little bit watchful here that all the alphabets they are in the lower case, but this capital D this is in the upper case right.

So, yeah I would try to explain you in more detail also that R software is case sensitive. The this is the R programming is case sensitive; that means, small y and capital Y, they are different. So, similarly whenever you are trying to use any name of the package you have to be very careful that whatever is given in the name in the lower case alphabet and whatever is given in the upper case alphabet, that has to be used exactly in the same way, right.

So, this name is package description p a c k a g e, then capital D e s c r i p t i o n, right. And, inside the parenthesis you have to write down the name of the library about which you want to know the details; for example, in case if I want to know the details about the spatial library. So, what I have to do? I have to write down the packageDescription and within the double quotes inside this parenthesis we have to write a spatial and it will give you this type of detail, right.

So, you can see here this is the package name priority version date when it was built and depends suggests and authors etc., etc. And, there will be some other details also, but I have given you here a brief screenshot just for the sake of understanding.

(Refer Slide Time: 20:54)



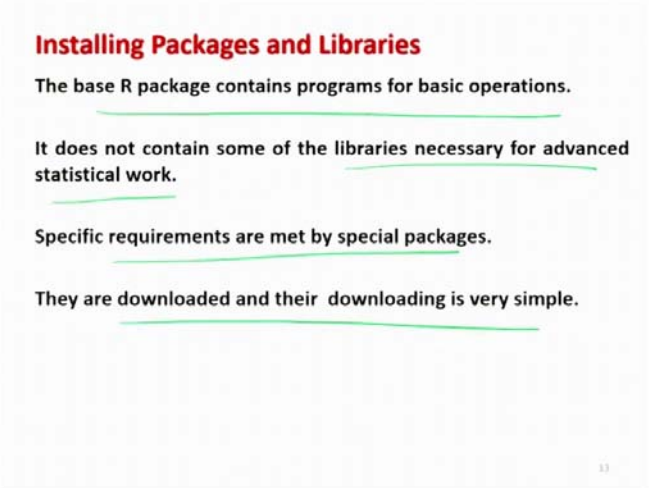
Now, means after that you would also try to seek some help from this library function. So, for example, in case if I want to seek the help for the spatial package, which has to be used in the library function. So, we try to use the function help with the library package in this following way.

I will try to write down here library, then within the parenthesis I will write down here help h e l p equal to and then I have to give here the name of the library for which we want to know the detail and it will give you some details like information on packages spatial which is packages spatial, priority recommended, version, date etc. etc.; and followed by some list of functions and data sets whatever are available with this one, right.

(Refer Slide Time: 21:57)

So, I have given you here a screenshot. So, means yeah you can do it; that is not a big deal, you simply have to type this library help is equal to spatial on your computer and you will get the same thing. There are many many details, so, I would not tell like to show you it on the R console, but you can do it yourself that is not difficult.

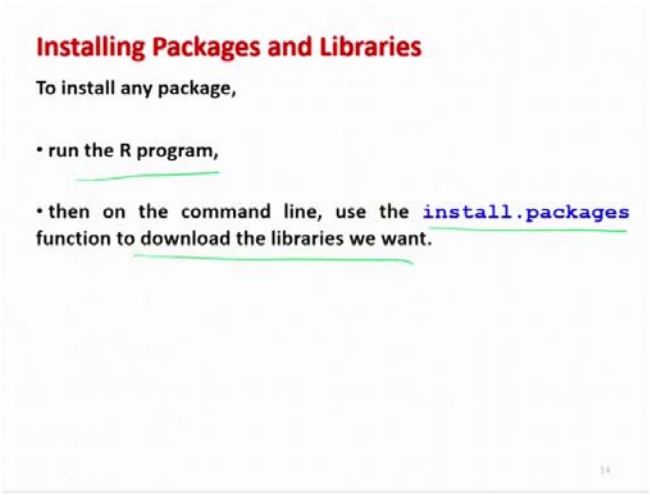
(Refer Slide Time: 22:14)



And yeah means that the time now when I have to make you independent also. So, whatever I am doing here you try to do it yourself on your computer and try to verify, are you getting the same outcome, ok.

So, now I come to another aspect whenever you want to use any library. First of all as I said that you have to bring the package from the website of the R software to your computer, right.

(Refer Slide Time: 23:40)

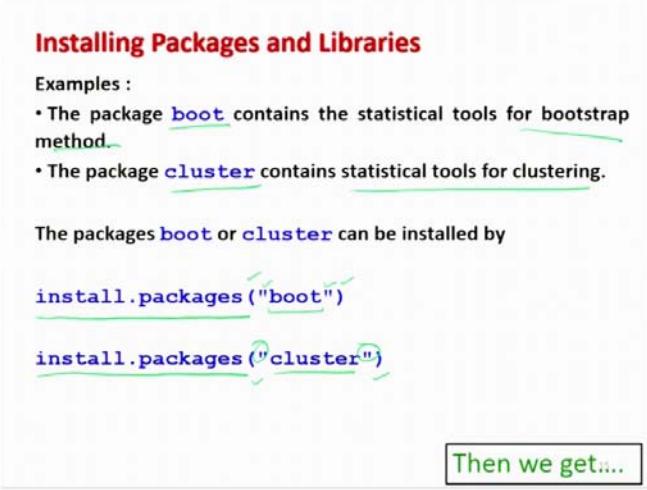


So, as we have discussed that the base package contains programs only for the basic operations and it may not contain the functions and libraries for some advanced work related to for example, statistics. So, these special requirements are met by these special packages.

So, now the question is, how will you bring these special packages to your computer so that you can use them? So, obviously, as we have discussed that these packages are available on the website of the R software. So, you simply have to go to the website, you have to download them and then you have to install them, it is just like as you download any software and after that you have to install it, right. So, similarly you have to download these packages and then you have to install it. And after that, you can use them.

So, now the question is that how to download and how to install these packages on your computer. So, to install any packages first you have to execute the R program. And, on the R console inside, on the command line, you have to simply type install dot packages. So, it is i n s t a double l dot p a c k a g e s, all in lower case alphabets. And, this function will help you in downloading the libraries what you want.

(Refer Slide Time: 24:23)



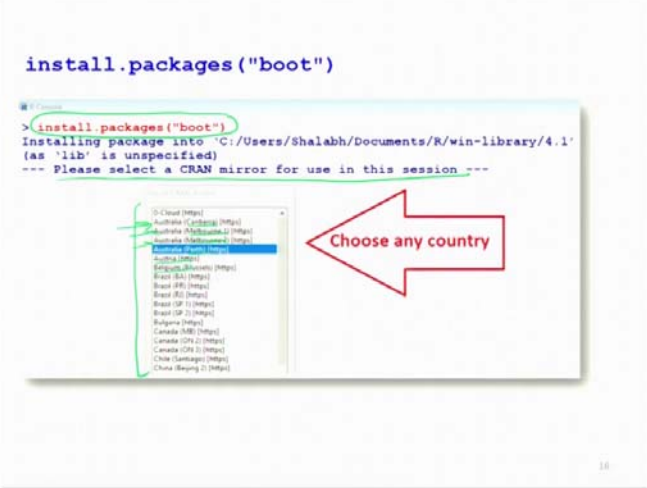
So, now whatever you want you have to give that name inside the parenthesis within the double quotes, right; for example, in case if you want to download the package boot or say cluster. So, as we have just discussed boot is the package which contains the statistical tool for the bootstrap methodology whereas, cluster package contains the statistical tool for the cluster analysis, right.

So, these two packages are not available in the base package of R, but they are available on the website of R software. So, the question is how to install them on your computer. So, I will try to show you here that how to get it then. And, through screenshot I will try to show you here what are you really going to observe. Yeah, sometime it may take a couple of say this seconds minutes etc. depending on the size of the package.

So, yeah so, I would not like to use this time in showing you here, but I will try my best to explain you through the screenshot and believe me I am promising you this is a very simple process. So, suppose I want to install the package boot. So, I have to simply type

here install dot packages, within the parentheses within double quotes I have to write the name of the package boot b double o t or similarly, if I want to install the package cluster. So, I have to type here install dot packages within the parentheses within double quotes, you have to write the name cluster c l u s t e r, then you will see what happens.

(Refer Slide Time: 26:02)



So, suppose if I take the example of a here boot. So, now, you can see here that as soon as I type install dot packages and then boot on the inside the R software on the command line something will happen like this, that something will be executed and it will be like this please select a CRAN mirror for use in this session. So, if you try to recall, we had discussed earlier that this R packages or this R software that has been uploaded in on different website and it has been hosted by different institutions.

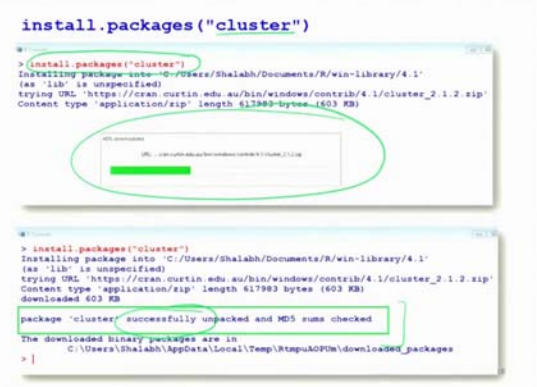
So, similar to what you have observed while downloading the R package because the same thing you will observe when you are trying to download a package. So now, it will ask you from where you want to download the package. And it has a list of countries in the alphabetical order, this is from Australia, at Canberra, Australia in Melbourne, Australia and Melbourne 2, Australia and Perth, Austria Belgium, etc. So, you can choose any country.

(Refer Slide Time: 27:14)

And after that what will happen? This type of a screen will come, right. And it will show you the progress of this downloading, e.g., it is showing here 61 percent downloaded. And, after this the whole package will be downloaded and automatically installed on your computer, right. And, what you can see here after this installation has been done after sometime depending on the size of the package you will get here a message like package boot successfully unpacked and MD 5 sums checked.

So that means, ok as soon as you have to just keep an eye on the word successfully unpacked and sums checked. So, this successfully will indicate that the package has been installed successfully. After this, this boot package has arrived on your computer. It is something like this a book has been bought in the library, but you, but since you have not issued it, you cannot use it that is what you have to keep in mind, right.

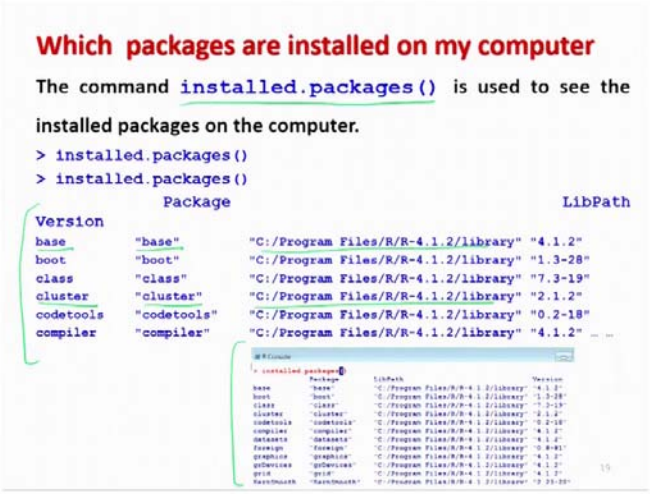
(Refer Slide Time: 28:27)



Similarly, as a as an another example in case if you are interested in the cluster package, if you try to write down here install dot packages within parentheses within double quotes cluster. So, once again you will see this type of progress in the downloading of the software and finally, you will get here a message which I have highlighted in green. And it and you have to simply look into this part whether the downloading has been successful or not.

So, in case if you try to do it so, now you have brought the two packages related to boot and cluster in your library, but in order to use it you will have to upload it first, you have to tell the library please issue it in my name so that I can use it.

(Refer Slide Time: 29:07)



So, I will try to show you how are you going to do it. So, the next question comes here that once you start working in the R software, so from time to time you will be downloading different types of packages and those packages will be available on your computer. And, suppose you want to check that what are the packages which are installed on my computer. So, in order to do that we have a command here installed dot packages and within the parenthesis, you do not have to write anything.

So, it is i n s t a double l e d dot p a c k a g e s and all in lower case alphabets. So, if you just use this package, this command over here you can see here this type of detail. Well, I am not showing you on the R console because these details are going to be different than

what I have on my computer. So, you can see here it will show you that the base package then cluster and then the this is the name of the package and then where it is located etc. etc. And this is here the screenshot you will get here long list of the packages which are installed on your computers, right.

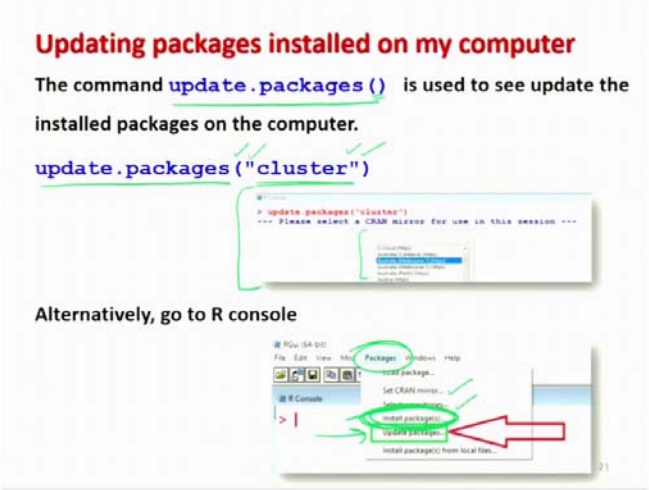
(Refer Slide Time: 30:36)

So, that is how you can get it. And, well, suppose in case if you keep on storing all these packages on your computer. So, definitely they are going to take some space on your computer. So, in case if you are not using them, you would like to free some space and for that you need to remove the package, you have to uninstall the packages from your computer which are inside the R software.

So, for that we have a command here remove dot packages r e m o v e dot p a c k a g e s and then you have to give within the parenthesis within double quotes you have to give the name of the package which you want to remove, right; obviously, only those packages can be removed which are already installed on your computer, right.

Suppose, we already have installed the package cluster. So, suppose I want to remove it. So, I will write here remove dot packages and within the parentheses, within the double quotes, I will write the name of the package cluster; and if you try to do it here, then you it will remove the package, ok.

(Refer Slide Time: 31:36)



Now, next aspect is that as we have discussed that R is a free software and R has a very big advantage that whenever there is some update in the software, all those updated versions can be downloaded for free. So, now, suppose somebody has developed some package and it is uploaded on the website of the R software and people are using it means after some time, some more features are added in that package and a revised version or an updated version of that package is available.

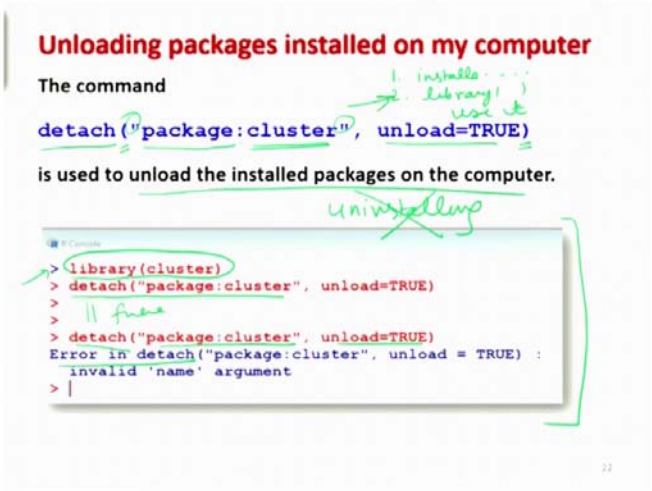
So, that academician will upload it on the website of the R software. And suppose you want to use that updated version. So, you would not like to remove the package, but you simply want to update the package. So, the question is now, how to update it? So, for that the command here is update dot packages u p d a t e dot p a c k a g e s and within the parenthesis within double quotes you have to write down the name of the package which you want to update. So, obviously, the package has to be there on your computer.

Suppose, I want to update the package cluster. So, I will write down here update dot packages with within parentheses, within double quotes, I will write down here c l u s t e r and you will get here this type of screen. And, once again it will ask you that which of the CRAN mirror you would like to use in which countries? Suppose, I use here Australia Melbourne; and then it will just you just click here and it will be done automatically. The other alternative is that means, if you want to use the R console directly here, if you go to the RGui window there is here a package and then there will be

here set CRAN mirror select etc. etc. Then, there will be an option here update packages. So, that also you can do and here I would like to one thing here more that another approach to install these packages is that you can download them means only the package somewhere on your computer.

And then using this one, this option here install packages you can, if you just click here the computer will ask you where is the package located, you just go to the to that location where you have downloaded the package. And if you give it there you can also install it, right.

(Refer Slide Time: 34:24)



So, I will try to show you it on the R console also and after. So, as we have discussed that once you have downloaded the package then in order to use it you have to use the command library. So, the first step will be that you have used the command install dot packages install it, then after that you have to use the command library whatever is the package name. Now, you have now use it now you have done your work and you want to remove it remove it you do not want to uninstall it.

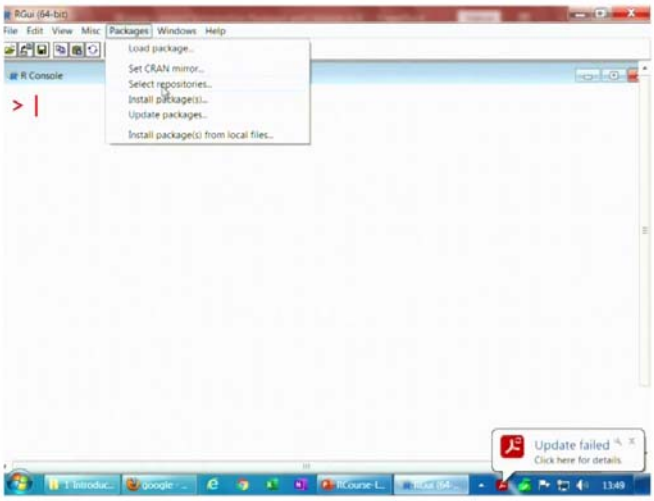
Remember, you just want to tell the R software that ok my job is done. Now, whatever I have done with this command library, I have to do opposite of this. So, for that the command here is it detach d e t a c h. And, then within the parenthesis, within double quotes, you have to write down the name of the package colon say for example, cluster

and then you have to write here unload is equal to true. And, this will unload the install package. Remember, I am not talking of uninstalling, remember.

No, I am not asking you to uninstall the package right. So, in case if you try to do it with the cluster package actually, you can see here this type of screen short will come, right. And means, I can show you here that once I have loaded the cluster package then I am using detach for this cluster package.

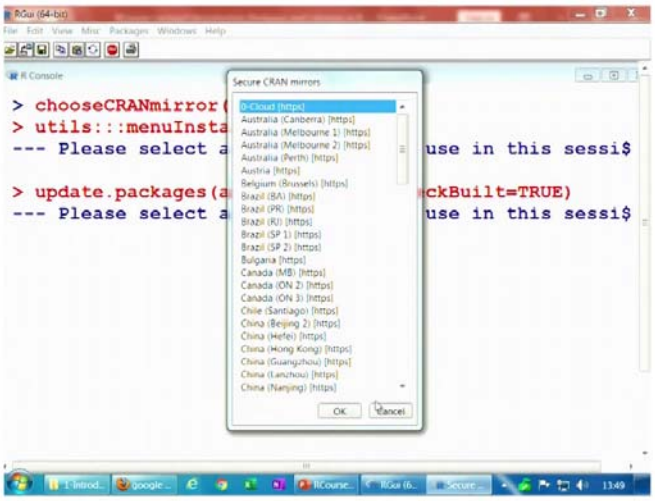
Then, you can see here there is nothing means everything is fine. Now, once I have detached it; that means, this package is not available in the library and if I try to use this command once again detach package cluster unload equal to TRUE this is giving me here error. Because the package can be detach only if it was loaded. And for loading, you have to use the command library. So, this is what you have to be very careful when you are trying to use it here.

(Refer Slide Time: 36:50)



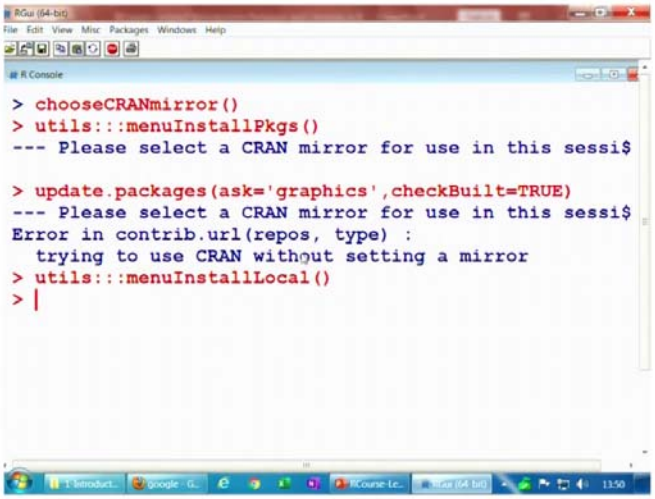
So, let me try to give you here a glimpse in the R software also in the RGui window then how the things are going to do for example, if you try to see here load package, set CRAN mirrors, select repositories etc. So, these things are needed for example, means every time you are trying to select that from where you want to download your package from Australia or Belgium or any other place.

(Refer Slide Time: 37:17)

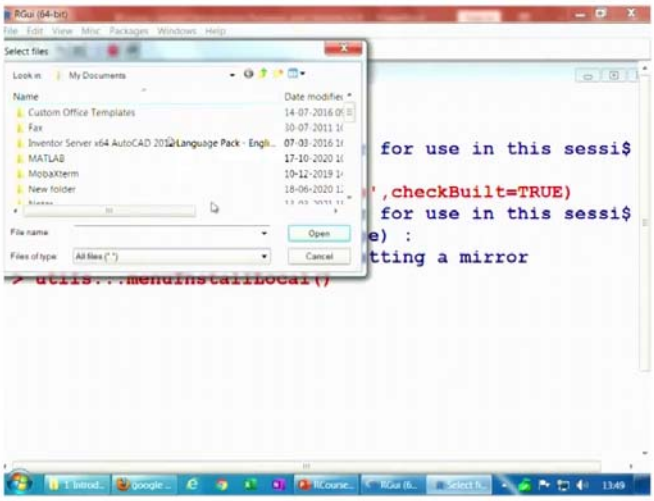


So, you can select here which repositories you want to use; you can get here this say this CRAN mirror also, for example, you can see here you can set here whatever country you want. And similarly, menu install dot packages that is also here if you try to see here it will give you this type of option here whatever you have typed, but I am just simply trying to show you the things from the basic fundamental point of view, then similarly you have here update packages. So, you can see here which package you want to update from which of the country.

(Refer Slide Time: 37:52)



(Refer Slide Time: 37:54)



So, I will just cancel it and then the last option is install package from the local file. So, if you have somewhere downloaded the package then you can also install it from here also right.

So, now, let me come to an end to this lecture. And yeah before leaving, I would like to tell you one thing more that whenever you are trying to download this software or these packages, they are local for example, if you have downloaded the R software in a computer at your office and at your home.

Now, suppose in your office you have installed the package cluster. So, that will remain only in that laptop or computer, that cannot be automatically install in another computer which is in your name. So, that is what you have to be careful. So, if you are working in your office at home and if you are using in the package cluster. So, you have to install at both the places. And, then many things you can see here, they are also available by this click from the RGui window. So, why you have not used that one?

My answer is simply, I am trying to teach you the thing from the basic fundamental means after some time we are going to talk about the R studio software also where whatever I have told you, these things are very easy to use, but then I do not want you to be dependent on any particular software, but I would like to teach you this R software on the R software only. And, I am more interested in developing the things in the basic

fundamental point of view. So, that when you are trying to do some good programming, you are not dependent on the software that is my simple objective. So, now, you try to look for this command and try to practice it; just try to install, load, unload etc. some packages and try to see what happens.

Once you become comfortable with this commands and with this language I promise you the learning of our programming will become very easy. So, you try to revise it, and I will see you in the next lecture till then. Goodbye.