Essentials of Data Science with R software-1
Professor Shalabh
Department of Mathematics & Statistics,
Indian Institute of Technology Kanpur
Lecture No. 03
Installation and Working with R Studio

Hello friends. Welcome to the course Essentials of Data Science with R Software 1 in which we are trying to deal with the topics of probability theory and statistical inference. So you can recall that in the last lecture, I have given you a very simple brief introduction to the R software that how can you install it and how you can start working on it. When we are working on R software, then we are working directly on the R console. In statistics or data sciences, whenever you are trying to deal with a bigger data sets, you may require lots of programming, you may require different types of functions and it will be very convenient that if you can see everything in a single screen.

In R, you can work in the R console and if you want to look into the program, you have to type the name of the program and then you will see the content. But many times for example, we are interested in looking at different programs side by side because R has a facility that you can call a function inside a function. So, to take a help or to deal with such situation, one option is that we can use some software which is a link between us and R software, but possibly it might be more helpful to work in that software. So, there are some free software which have been developed and which are available which helps us in doing different types of programming in R.

For example, there is a software R studio. There is another software, Tinn R and these packages are trying to help us. Well, there are some other packages also. So, now my idea and my objective in this lecture is that I want to introduce you with these type of supplementary software which are trying to help us while we are working in the R software and I have chosen here R studio, but surely it should not give you a feeling that I am trying to make an advertisement of R studio. Well, I can choose anything actually either R studio or Tinn R or there are some other software. So, please do not think from that point of view, but that means I have to choose one. So, I have chosen here this R studio.

So, in this lecture, I will try to give you a very quick review of this R Studio software that if you are trying to work on this software then how you are going to work and how it is related to the R

console and how you are going to manage the four Windows of R studio. That is the modest objective of this lecture. So, although I believe that if you are working in R, you must be using some similar software, so I am not asking you to shift from that software but here if you want to work in the R studio this may help you.

So, let us begin our lecture. So, in this lecture we are trying to we are going to learn about that how are we going to install the R Studio software and how are we going to work on that but very quickly. So, now you see.

(Refer Slide Time: 04:08)
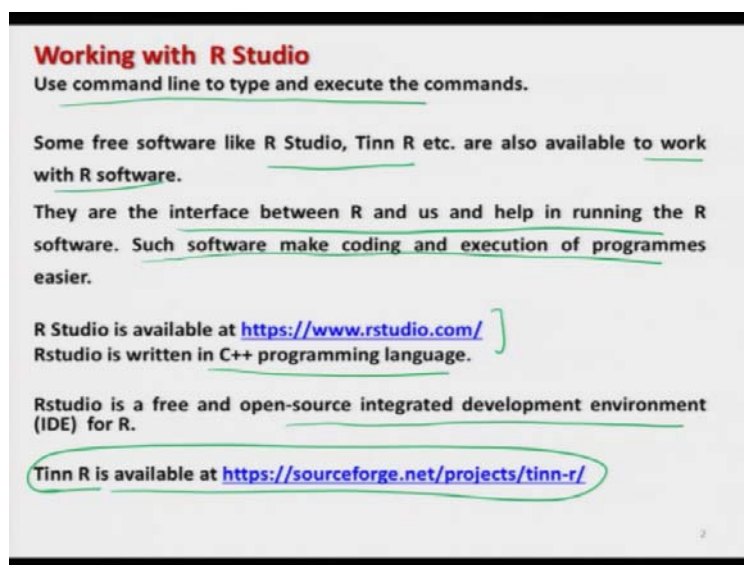
So, now when we are trying to use the R software then we are basically working on the command line to type and execute the basic commands. There are some free software like R Studio, Tinn R, etcetera. They are available and they are actually working along with the R software. So, this R software is working behind this software, behind the screen. So, they are essentially the interface between R and us and they help us in running the R software and the different types of command and in case, if we try to use the software then the software to make the coding and education of the programs easier.

So, this R Studio is available at this site www.rstudio.com and actually this is also a free software and this is written in the C++ programming language and so this is a open source integrated development environment for R and similarly, there is another package Tinn. This is

available from this website. You can simply go to some search engine and just type over there Tinn and you will get this link very easily.

(Refer Slide Time: 05:25)
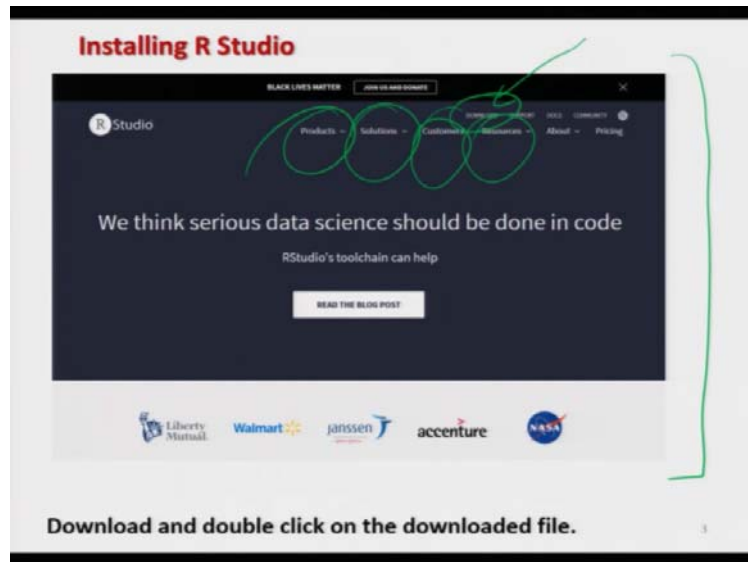
Now, once you come to the website of rstudio.com, it will look like this means, I am trying to show you on the website also, and then you will see here there are different types of things. It is here product, solution, customer, resources and, they will be here download and so on. So, R studio is actually available for different types of things but we are interested only here in the R studio part that is working for the R software.

So, we have to click here to download and then it will give us the file and then we have to save it and then we have to simply double-click on the sites and then you have to click on that file and then you have to just read the section and then you have to continue by clicking it and then it will install the software on your computer.
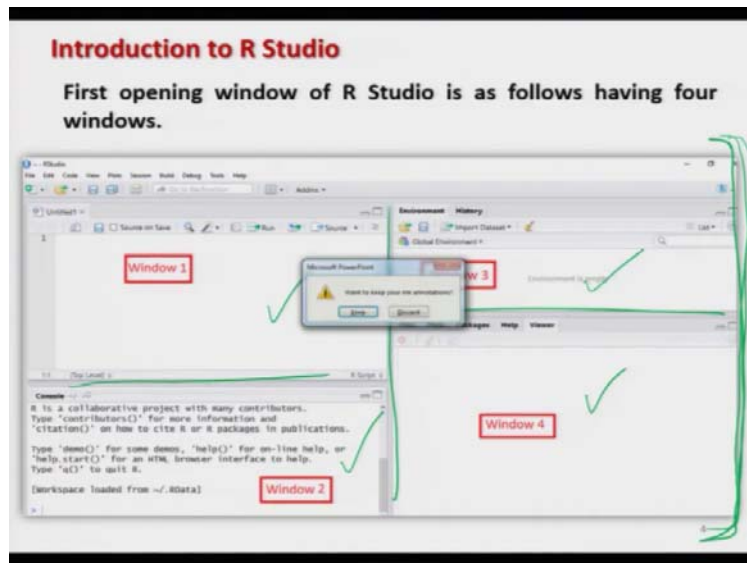
Now, in case if you try to click on the icon of R studio and open it, it will look like this, what it is given here. There are four windows – 1, 2, 3 and here 4. These four windows are trying to help us in executing different types of things and the important part is this all those executions can be seen in a single screen. The size of these windows can be adjusted by clicking on these border lines which are here like this, which I will try to show you and these windows will give us different type of information. So, first let us try to see how are you going to get this software and then we will try to see that how it will look like.
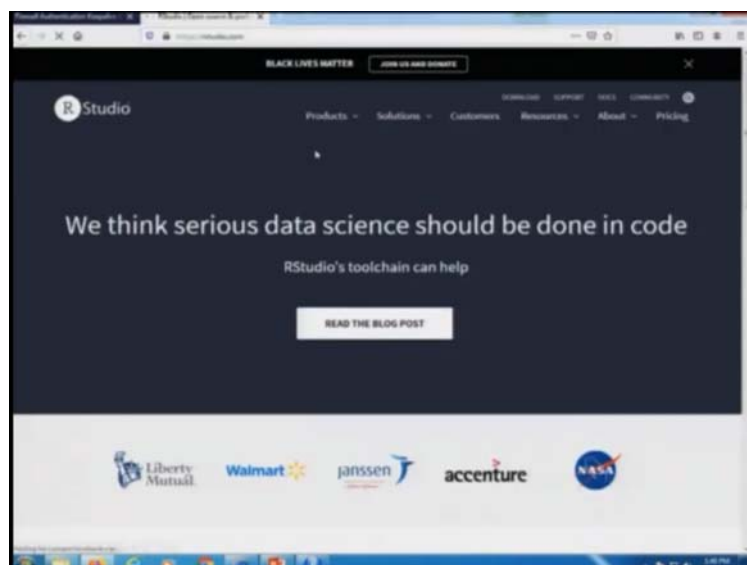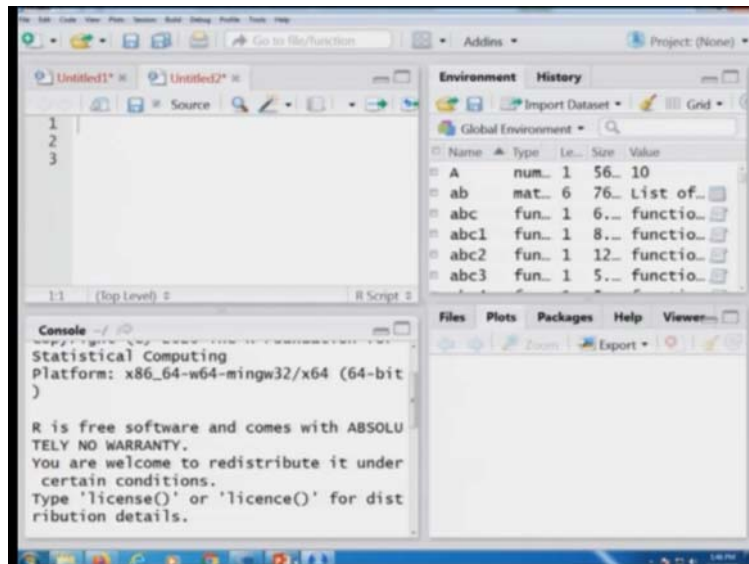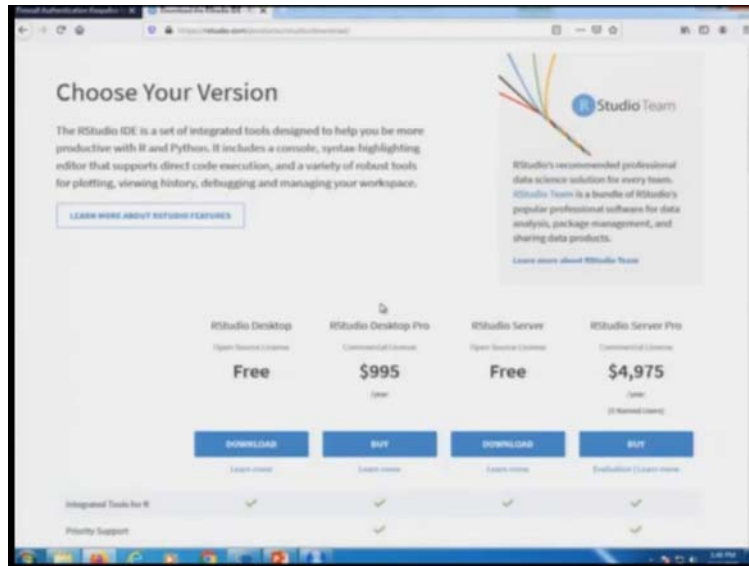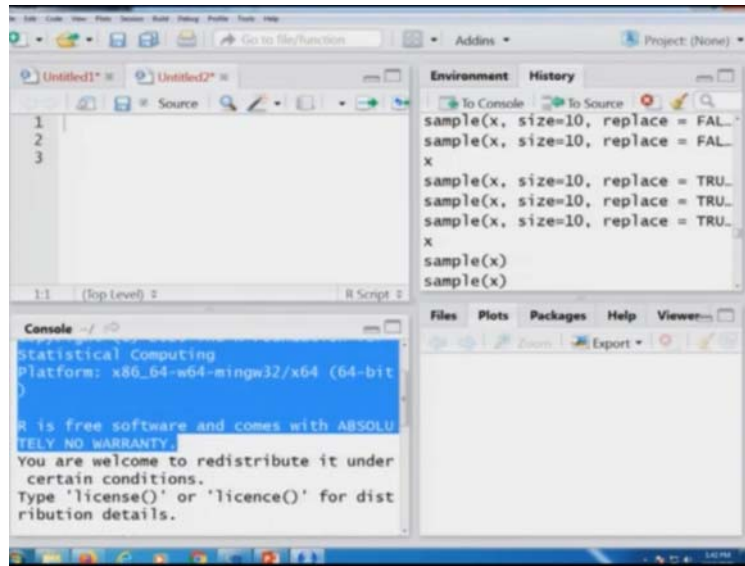
So, if you come to the Internet site and simply try to press here rstudio.com and you try to press it here, you will get here, this type of screen, this type of website of R studio and if you go ahead download and then you will see here that you have to choose here your version. There is a free software, there is a paid software also and there is R Studio for desktop, RStudio for server. So, I will simply say that in case, if you are working on that desktop just try to click it here on the free part, which is here download and in case if you are working on a server then you try to click over here under the free part and just download it from here.

Once you download it from here and then you simply double click it and read the condition and can try to accept it. Click it again. Finally you will be able to install the software on your computer. Once you have installed it and if you try to double click on this R studio icon, this type of window will come. So this is the same window which I had shown you but in this window something is written because I was working earlier and my earlier commands are visible here, somewhere here where I am trying to move my cursor.

So you can see here, this is my window number here one. Please try to see the movement of my cursor where this is window number 1. This is here window number 2, which is where I am trying to move my cursor. This is here window number 3 and this is here window number 4. Try to have a very clear look so that I can explain you from the screenshot of this window. One thing. I will try to show you here that these are here the borderlines where you are getting this 4

dimensional arrow and if you try to just press the mouse of your computer and try to move it here, you can increase or decrease the size of this windows. You can see here like this.

So, these are the things which are in your control and if you try to see here before I try to go to my slides you can see here, what is this where I am trying to highlight, Can you identify it what is this? Do not you think that this is the same as what you had obtained in the R software like here and you started it? So, now you can understand that what is the advantage and now let me come back to the slides and try to explain you the utility of all these 4 windows very quickly.

(Refer Slide Time: 9:51)
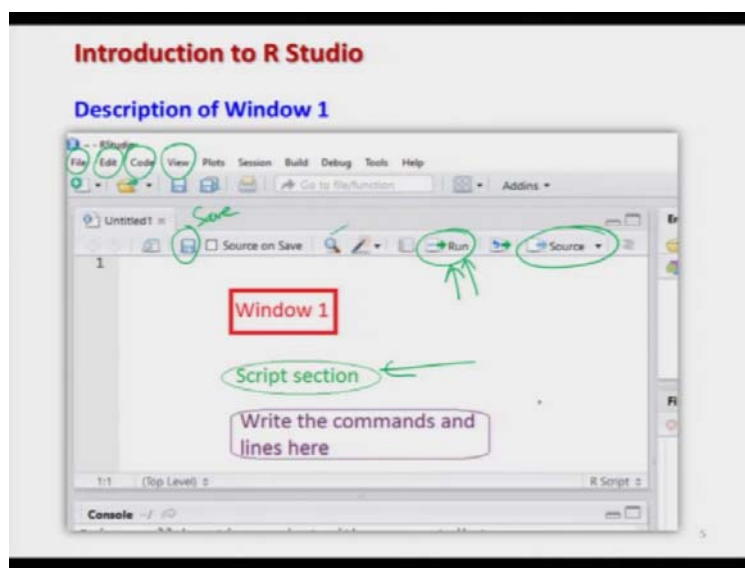
So, let us try to understand the utility of all the windows one by one. So, first I try to concentrate on the window number 1. So, window number 1 is the place where you try to write down all the programs, where you try to write down all the syntax and commands for your program and this is called as a script section, and then you know that there are different option. This is the option for save and then there is the option for search, there is option for your run. So, if you want to execute a particular syntax command or program you simply have to highlight it and then simply try to click here on the run.

It will execute the program and if you want to link the source of the program, you simply have to click here on the icon of source and simply you can see here, there are different options here file, edit, code, view, etcetera. So, these are very standard thing and I do not think the type of student

to whom I am now a teaching this course, it is a very difficult thing for them. So, means you can just experiment with this window and try to see what is there.

You can make are different types of script file. Well, one thing is that whenever we are trying to save the script file, script file means the programming file and that is the common language, the program for a software. So, script is the programming file in which we are trying to write down the commands of the R software. So here in this window number 1, we are trying to write those commands.

(Refer Slide Time: 11:23)
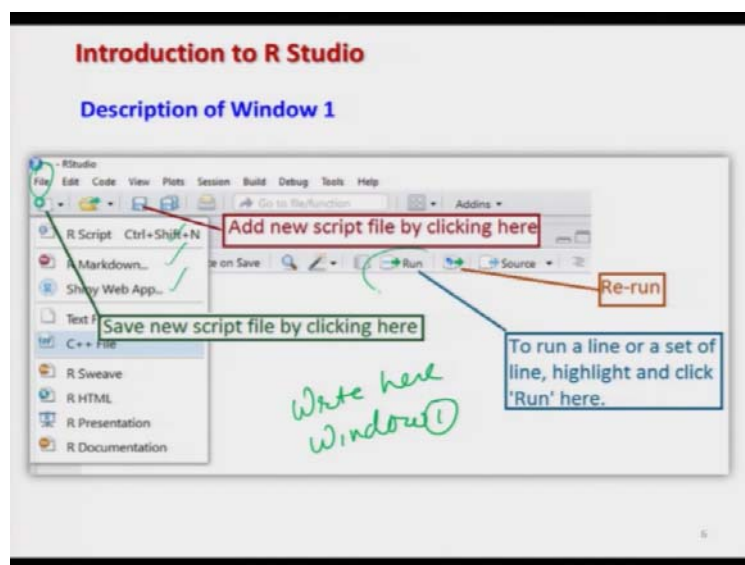
Now we come to our this thing this a window number here one means if you try to open here this file, you will get here R Script, R Markdown, Shiny Web page, etc., etc.. There are many many things. So, you can see here now this are R is developed and this R markdown Shiny webpage and they are the new features. Well, I am not dealing with those things here. So, I am not discussing it here, but I just want to show you here. So, now I would say this, simply you try to execute it yourself.

(Refer Slide Time: 11:58)
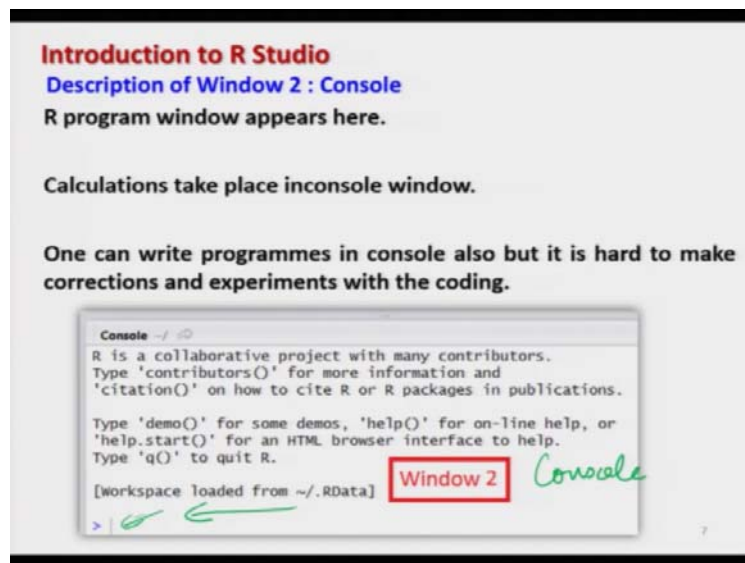
Now in case if you come to the window number here 2, this is called your console. If you try to see here, this is the console that we had discussed earlier also and this is the place where in R software you try to write down the commands and programs for execution. So, this is the same window which is appearing here and but what is happening, you are not going to write anything over here at the command line but what are you going to do, you are going to write here in window 1 and as soon as you see here run means execute, whatever is the execution that is happening in this window number 2 inside the R console.

So, now you can see whatever you are viewing from this window number 1 that is being executed behind the screen and whatever is being executed behind the screen that is visible on window number 2. So, that is the advantage that you can see the script as well as the execution of the program simultaneously when you are working in the software like R studio.
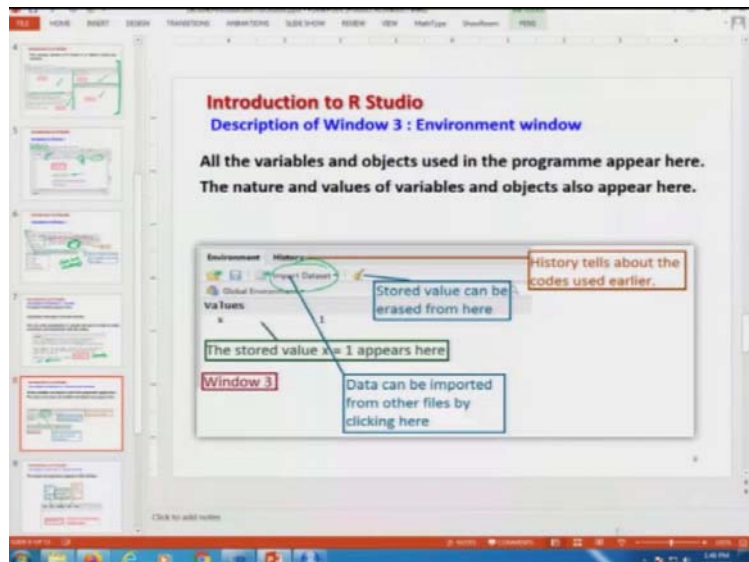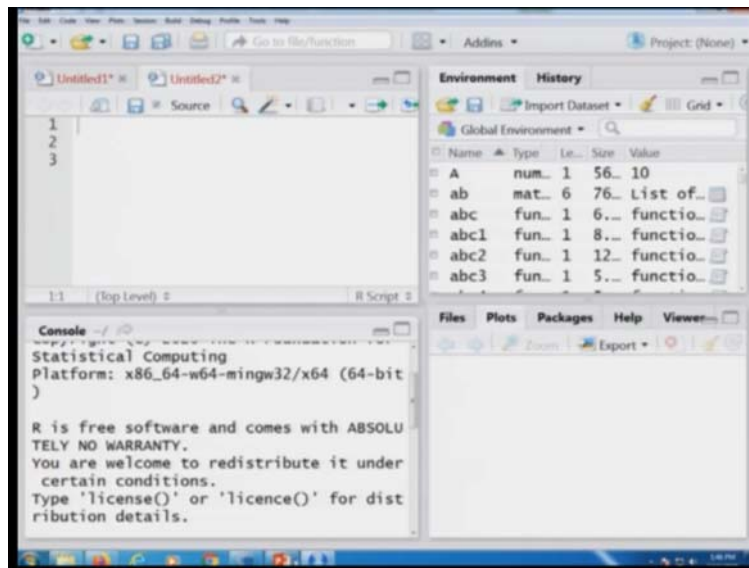
And similarly, if you come here, in your number 3, well I can just show you here before you forget that what was your window number 3. I can show you here like this. This is you're here window number 3 and so on, you can see here this is environment, history and so on and then you have import data sets and then global environment, etcetera, etcetera. So, I am now talking about this window over here.

So, this window number 3 gives you an information on different things. For example, whatever variables you are going to use or define a new program, they are available there. Although they are also available in the R console also, but here you can view them very clearly directly on your

screen and in case, if you do not want them, you can just hide them by minimizing the size of the screen of the window 3. That is up to you and then here you will try to see here that there are here different types of option like an imported data set.

For example, if you want to import the data set from this MS Excel or different types of software possibly this can be done. For example, if you try to see here if I try to click here import data set, import data set from CSV, from Excel from SPSS data and so on. These are different types of things are there and you can see here that it is trying to give, if you try to look at the variable number here A, it is trying to give you here in that case over here a number and it is trying to give you different types of information and simply, if you try to come over history, it is trying to show you that what I was exactly doing on this software in the past.

So, this is trying to give you the details about the history. So, all those things are available to you when you are trying to work with our studio and they are available very conveniently actually. So, this is the use of window number 3.

(Refer Slide Time: 15:10)
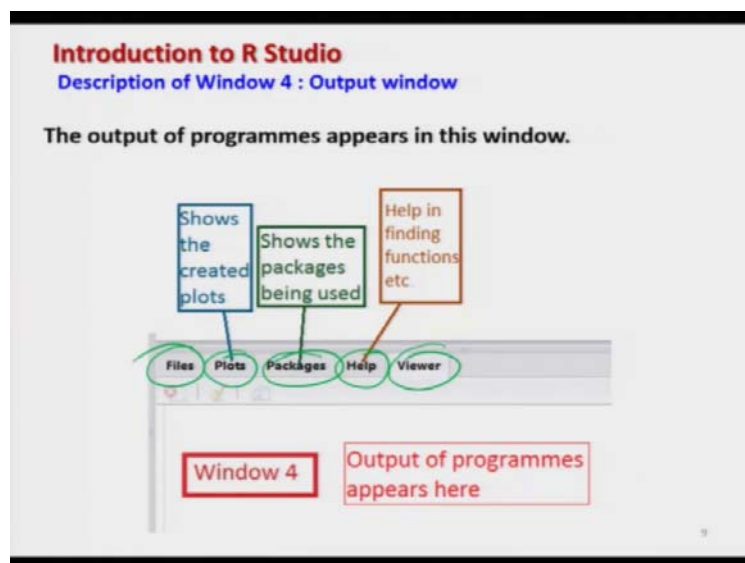
Now in case if you come to here window number 4, the window number 4 is essentially the output window. Here different types of outcomes of your graphics, they will be available, different type of information the packages that you want to install or that are already installed if you want to load a package load a library or if you want to take different types of help that is also available over there.

For example, you can see here the first icon here is file, second here it is plots, plot is trying to show you the created plot that mean the plot which you have constructed through your programming and then it is information about the packages. Packages means what are the packages that are available on your computer under this R software which can be used and if some packages not available, you have to see that where is this and how to install it and so on. All sorts of information about the packages available here.

Then there is next option here help. For example, if you want to know some information command, syntax or how to use, this help is going to do for example, if I want to know that how to find out the mean and suppose I will simply try to type here mean and it will give me detailed information about the mean function and then you have the finally the viewer. That means here you can view here different types of graphics and different formats. So, this is essentially the fourth window which is mainly used for being the output of the program.

(Refer Slide Time: 16:51)
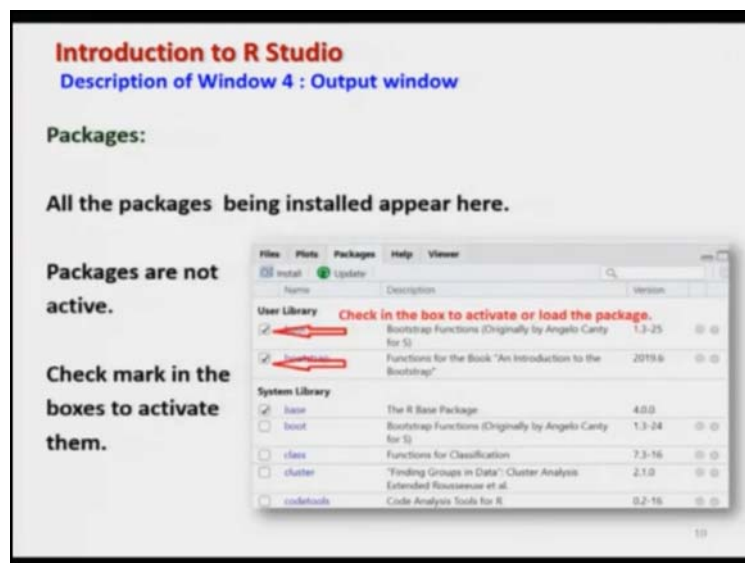
So, before I try to show you, let me try to just show you that how you can execute it. For example, in case if you want to install a package you have learnt in the last lecture that you have a command install dot and then you to give the name of the package within the double quotes inside the parentheses and then it will install it, but here in case if you want to install the package inside the R studio, you simply have to click here at install. It will open a new window and you

simply have to type the name of the package and those packages will be appearing in the drop down menu also and then you can choose from them directly.

And in case, if you want to use a package or use a library, you have learnt in the last lecture that you have to use the command library where you have to give the name of the packages inside the parentheses. So, here you do not have to do these things but what you have to do, you simply have to make a check mark this one in this box against the name of the package.

Suppose if I want to use a package here boot. The boot is the package which is used for bootstrapping so means I will simply type here is somewhere here boot and then it will give me here the package here boot or bootstrap, these two packages are available, I simply come here and I try play here a check mark like this. As soon as I say, place this check mark inside the box, the packages are uploaded. So, in case if you try to see what is really happening. Do you really think that just by making a check mark in the box or if you type the name of the package, the package can be uploaded or installed? No.

Our studio cannot do this job because this because these packages are available on the website of the R software but whatever you were doing inside the R console by typing the installed dot packages or library, the same thing is executed at the front of this R studio very easily. You simply have to type the name of the package and it will give you the possible name and from here you can choose. You do not have to type install dot packages and similarly, if you want to load a package earlier you had to write library and then package, but now you do not have to do those things. You simply have to make a check mark.

Once you make a check inside the box, the package is uploaded but behind this thing inside the R console, what is happening. It is time to execute the command library. So, you know that is a difference between learning means if you do not know that by putting a check mark or a check inside the box, what is happening in the R console, you may not understand the basic fundamental.

That what is really happening and if there is a problem you cannot solve it. So, that is why my objective in this brief lecture or in a short lecture is simply to make you understand that when you are working in the R studio, do not try to take it for granted that what R studio is doing you do not know. You must know it. You must not know that what is happening at the backend and

this is how you have to learn it and that is why you need to learn the commands like installed dot packages or this library command.

(Refer Slide Time: 21:00)
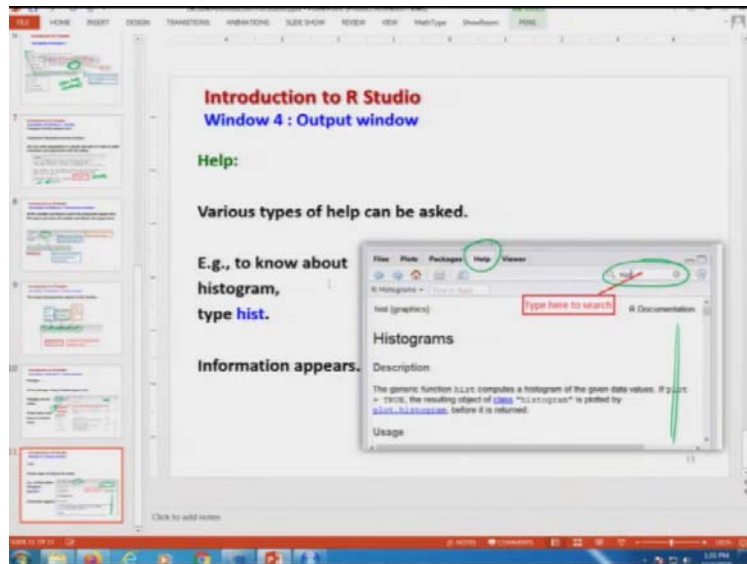
So let us now come back to our slides and now similarly if you try to go to the help part here which is here and if you try to for example, if I want to have some idea about the histogram so I can type here in this window hist and as soon as I type it, I will get here this type of detail where all the information about the command about histogram is available and one can use it here and similarly if you want to use any other command, all this information is available at your fingertips. You do not have to remember, you do not have to recall everything but the only thing is this you must have some idea about these things. So, before I go to anything else let me try to come back to R studio and try to show you here what is really happening.

(Refer Slide Time: 21:48)
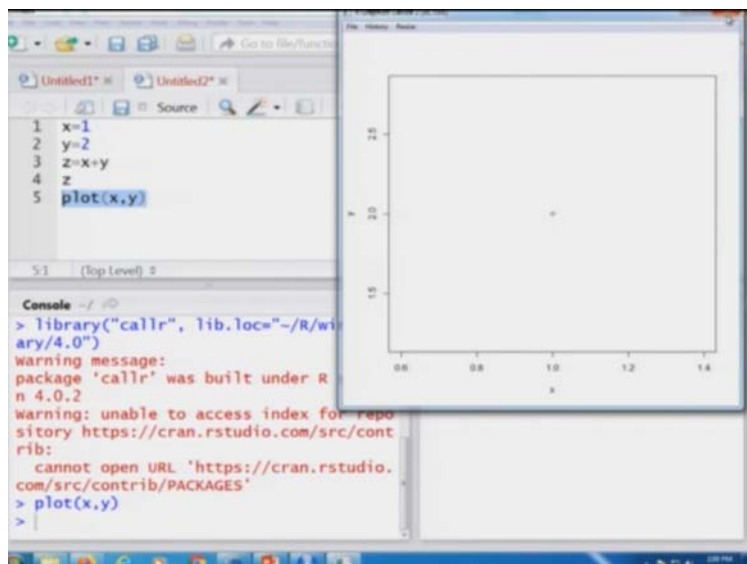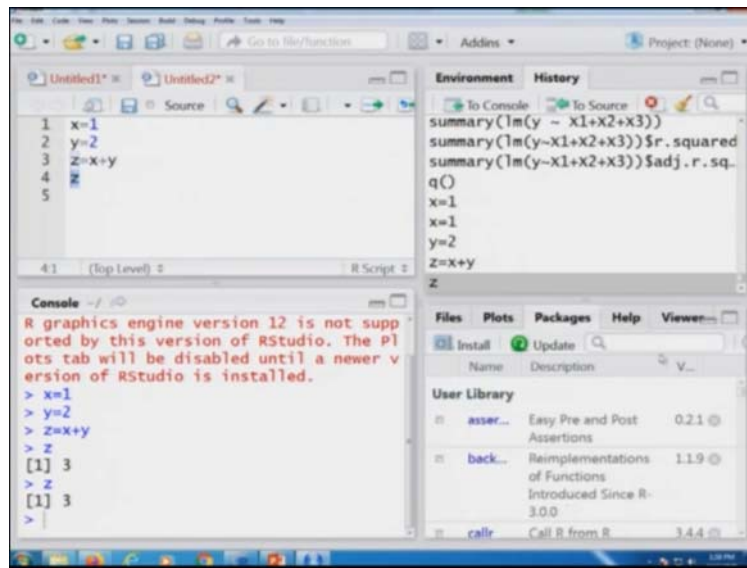
So, now before I come to anything, let me try to first clear this window. So, for cleaning the window, you know, the command was here Ctrl L. So I can simply use here Ctrl L otherwise, you can clear this window by the given commands over here. Now, I try to write down here. Suppose I try to write down here x equal to 1. Suppose, you can see here means I am writing here x equal to here 1 and I will bring the third window at the bottom. So, now what you have to see. As soon as I try to highlight it and I try to execute it here, you have to observe what is happening here in the window number 2 and here window number 3.

So, now I just highlight it and I see here execute. As soon as I clear execute, it is trying to execute here x equal to 1 and the R studio and you can see here that x equal to 1 has come here in the window number 3. So, now this is giving us an information that as soon as I try to put your x equal to 1, this is being executed in the window number here 2. Well, it is giving us some warning because this package has been updated and so but I am not bothered about these things, but they definitely are not asking you to ignore these messages.

Warning messages are simply like warning. They will not stop your work, but you must know and you try to update your software but anyway at this stage that is not my objective and similarly if you try to see here, suppose if I see here y is equal to 2 and suppose if I try to define here z is equal to x plus y. So, I am trying to write down all the programs over here in this window number here 1 and if I try to highlight all the 3 lines and if I try to execute it here you can you have to see here what is going to happen after this x equal to 1 and what is going to happen here in the window number 2 in the console window.

So, I now press here x equal 1, y equal to 2 and that is equal to x plus y, I press here execute and you can see here this has been executed here the window number 2 and in case if you try to see what has happened after this window, you can see here this is x equal to 1, y equal 2 and z equal to y, x plus y has been executed.

Now in case if you want to see the value of the z, you have here two options, either you try to press here z and try to see here execute and you can see here the value of z is coming out to be here 3 and here you can read the outcome z and if you try to see you here also, it is trying to show you that a new variable has been executed z and whose value if you want to see, you simply have to press enter and it will show you the value of the z is 3.

And similarly, if you now to come to hear this window number 4, for example, if you want to see where packages, so packages is going to give you the different information that what are the packages which are available on my this laptop. Well, remember one thing that when you are trying to install these packages, these packages are installed only on the computer on which you are doing. It means if you have a computer in your office and say another computer at your home, so in case if you are trying to install it on the computer in your office, it does not mean that it is going to automatically install in the computer at your home.

So, similarly, if you want to stay here, suppose if I want to see here whether a package to has been, suppose I want to see about this call R package and I do not know whether this package is there or not. So, I can see here callr and you can see here that as well as I am trying to type this name, this is appearing here and in case if I want to install it, I simply have to make here check mark and if I want to install here package suppose, if I want to type here agricolae. So, you can see as I try to type this here and if I try to install it will install it.

Well, I am not doing it here but anyway, these things can be done without any problem and here we can see here in these files, it will try to show you the information about the files available in the working directory. Similarly, if you come to here plots, it will give you here different types of plots in case you are trying to plot it here. For example, if I see here plot, see here x and y, so you can see here. Although there are not any observations, but it will give me here like this. There is only here one plot here.

So, and then you will means if I want to export it means I can save it here as image or PDF, whatever you want and similarly if you come here on the help and suppose if I want to get a help on the histogram and my command is hist suppose. As soon as I start typing here his, is you can see here in the drop-down these option are coming and if I choose here hist, you can see here all the information about histogram is coming over here and that means you can read it and then you can just do whatever you want.

So, with this brief introduction to the R software, let me come to an end to this lecture. My objective was not really to make you learn about the R Studio software, but my objective was just to give you a brief introduction so that you can brush up your knowledge. You are confident and those students who have not learnt it earlier, this will give an idea that if they try to learn it in the given time as soon as possible that is not a very difficult software and it is going to help them.

So, one thing I would like to tell you here and I would like to clarify here that although I have introduced you here with the R Studio software but definitely in all my lectures, I am going to work only on the R console, directly on the R software. Why because in case if you work directly on the R Studio software only, then it will be R studio dependent and there is a possibility that

different people, different student, different colleagues think they might be using different types of software and over a period of time some new software may also come.

So, my objective here is that I would not like to make you learn so that you are software-based. These intermediate software maybe changing, maybe improving but R studio software will always remain the same, R console will always remain the same. Only the version and the improvement in the R software will be there. So, I will be working directly on the R console and there another advantage will be when I am trying to work on the R console you will also see that how the commands are working and how is the real outcome which is appearing on the R software and how to handle with those things.

So, that is why, now today onwards after this brief introduction to the R studio, I will not come back to the R studio software, but I will be working only on the R console. So, you try to have a look on this R studio. Try to have a quick revision, try to understand its utility and different types of options which are available and I will see you in the next lecture. Till then, goodbye.