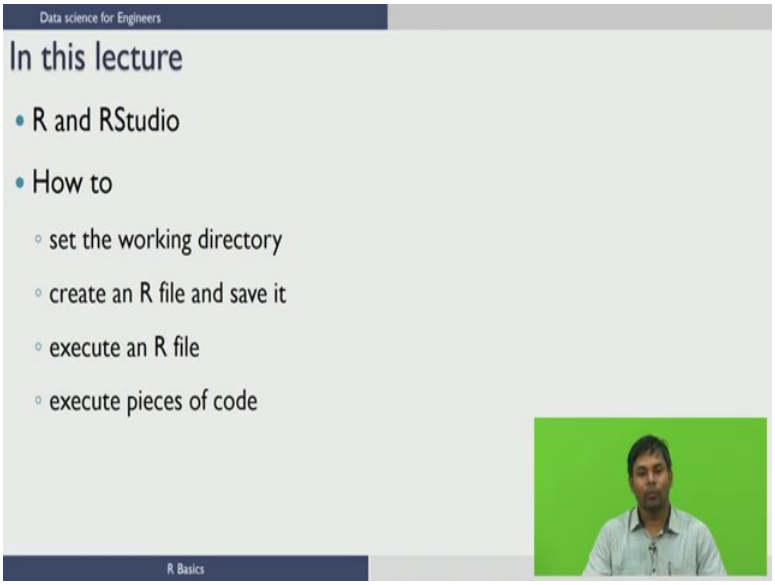
Data Science for Engineers
Prof. Raghunathan Rengaswamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 02
Introduction to R

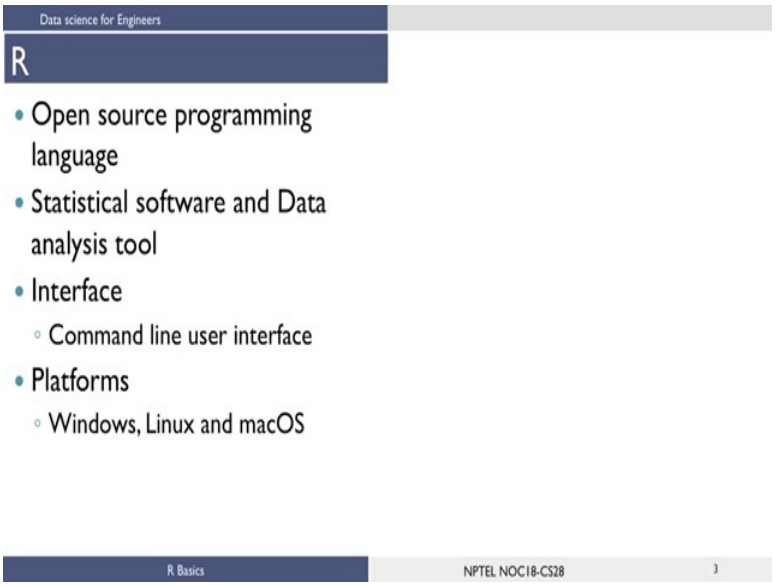
Welcome to the course on data science for engineers. In this module, we are going to introduce R as a programming language to perform data analysis. This lecture, we are going to give a brief introduction about R and Studio.

(Refer Slide Time: 00:25)



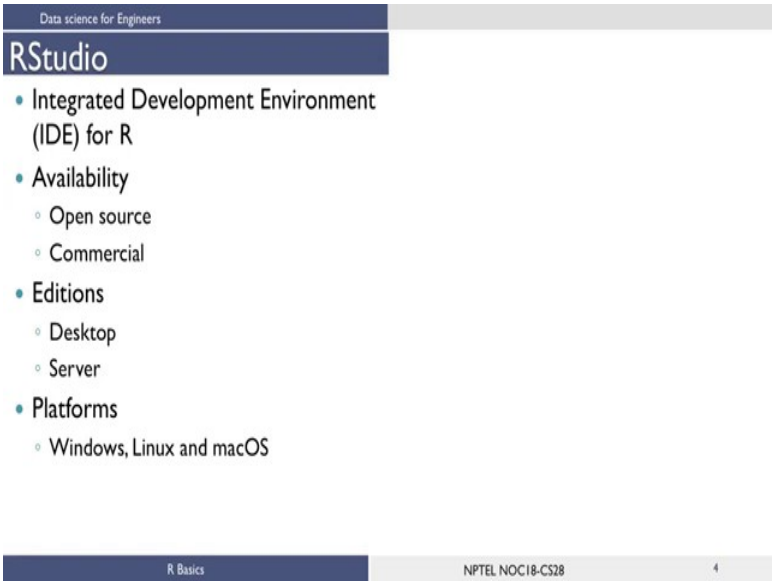
In R studio, we are going to look how to set the working directory, how to create an R file and save it, how to execute an R file and how to execute pieces of R code.

(Refer Slide Time: 00:41)



Let us first see what is R. R is an open source programming language that is widely used as a statistical software and data analysis tool. R generally comes with the Command line interface. R is available across widely used platforms, windows, line x and macOS Now, let us see, what is R Studio.

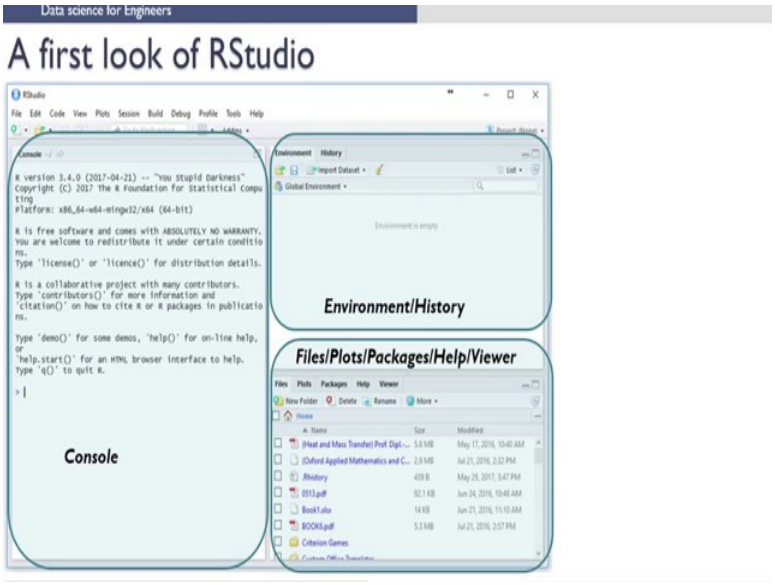
(Refer Slide Time: 01:05)



R Studio is an integrated development environment for R. Integrated development environment, is a GUI, where you can write your quotes, see the results and also see the variables that are generated during the course of programming. R Studio is available as

both Open source and Commercial software. R Studio is also available as both Desktop version and Server version. For this course, we are going to use Open Source Desktop Edition so, that you can solve your assignments using this R Studio. R Studio is also available for various platforms, such as windows, line x and macOS.

(Refer Slide Time: 01:50)

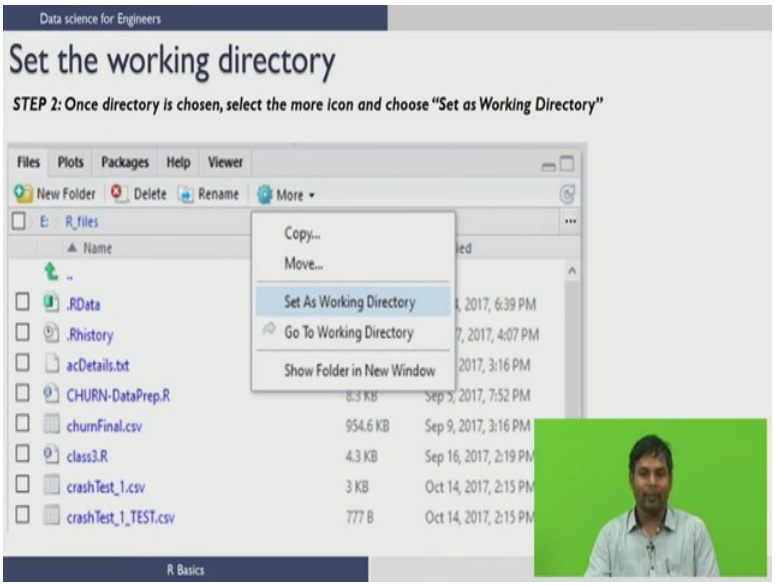


Now, let us see, how an R Studio looks, when you first run application. This is how an R Studio Interface looks. When you first run the application, to the left, we see Console panel, where you can type in the comments and see the results that are generated when you type in the commands. To the top right, you have Environmental History pane. It contains 2 types: the Environment type, where, it shows the variables that are generated during the course of programming, in a workspace, which is temporary and in the History tab, you will see all the commands that are used till now from the beginning of usage of R Studio. The right bottom, you have another panel, which contains multiple tab, such as files, plots, packages and help.

The Files tab shows the files and directories that are available in the default workspace of R. The Plots tab shows the plots that are generated during the course of programming. And the Packages tab helps you to look, what are the packages that are already installed in the R Studio and it also gives an user interface, to install new packages. The Help tab is a most important one, where you can get help from the R Documentation on the functions that are in built in R. The final and last tab is the Viewer tab, which can be

used to see the local web content that is generated using R, are some other application. For this course, you are not going to use this tab from much. So, we limit ourself not discuss more about that, viewer tab. So, we have got an idea about how R Studio looks. Let us see, how to set the working directory in R Studio.

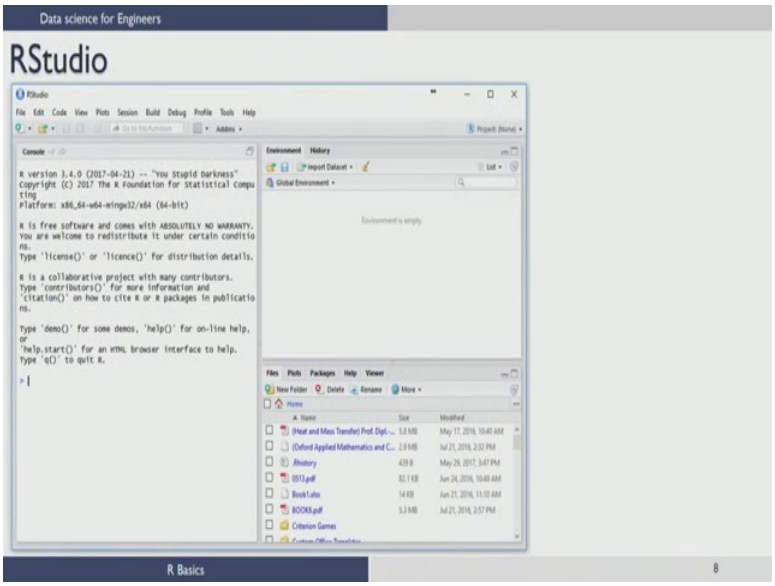
(Refer Slide Time: 03:52)



The working directory in R Studio can be set in 2 ways. The first, way is to use the console and using the command Set working directory. You can use this function Set working directory and give the path of the directory which u want to be the working directory for r studio, in the double codes.

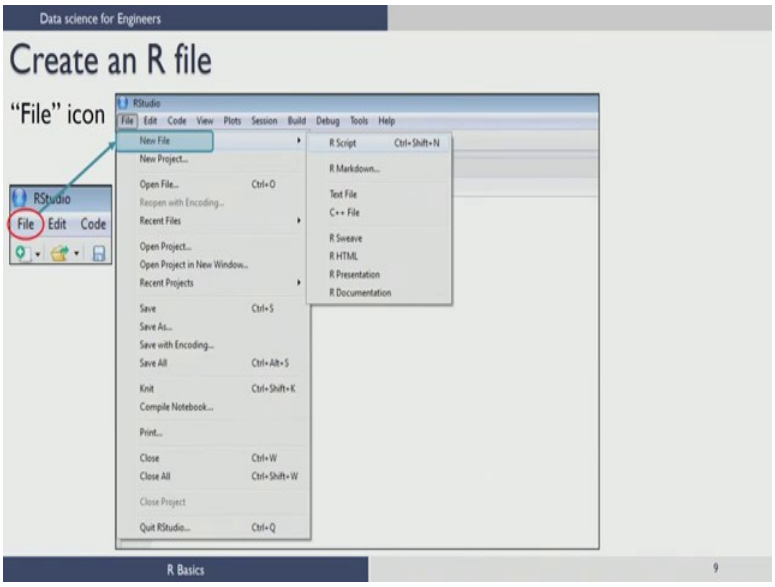
R, to set the working directory from the GUI, you need to click on this 3 dots button. When you click this, this will open up a file browser, which will help you to choose your working directory. Once you choose your working directory, you need to use this setting button in the more tab and click it and then you get a popup menu, where you need to select Set as working directory. This will select the current directory, which you have chosen using this file browser as your working directory.

(Refer Slide Time: 04:50)



Once you set the working directory, you are ready to program in R Studio.

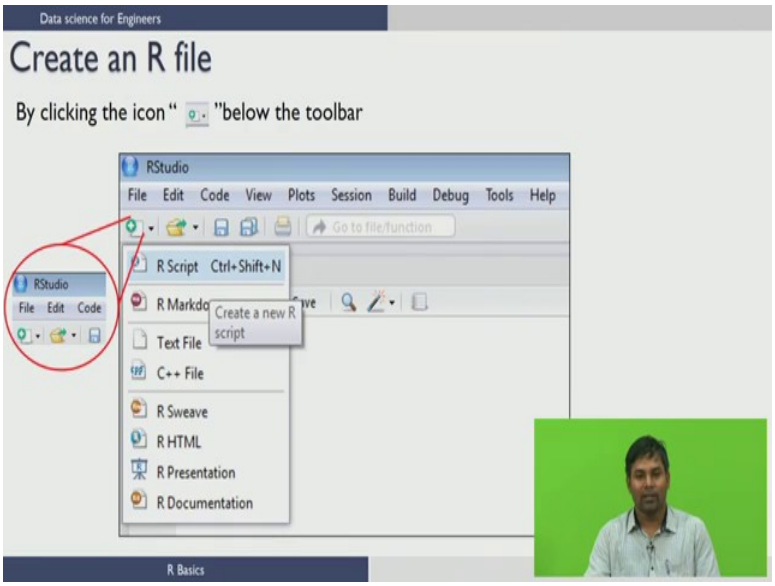
(Refer Slide Time: 04:56).



Let us illustrate how to create an R file and write some code. To create an R file, there are 2 ways: The first way is: you can click on the file tab, from there when you when you click it will give a drop down menu, where you can select new file and then R script, so that, you will get a new file open.

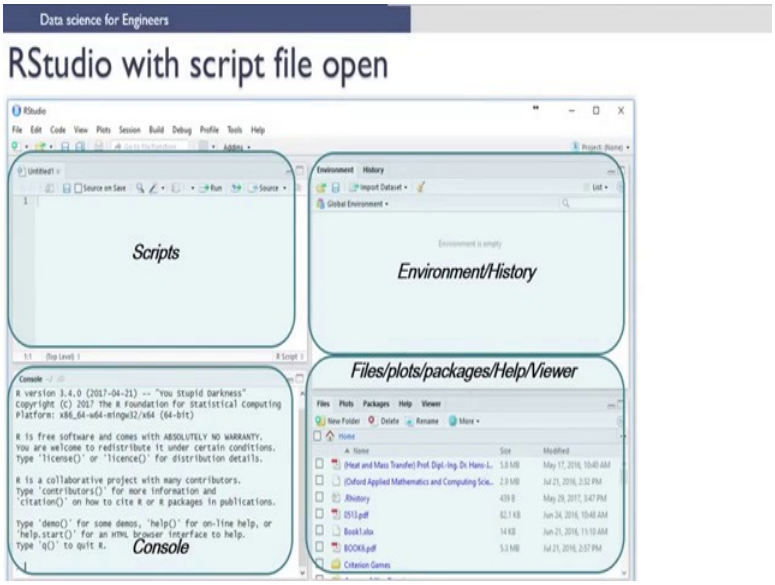
.

(Refer Slide Time: 05:18)



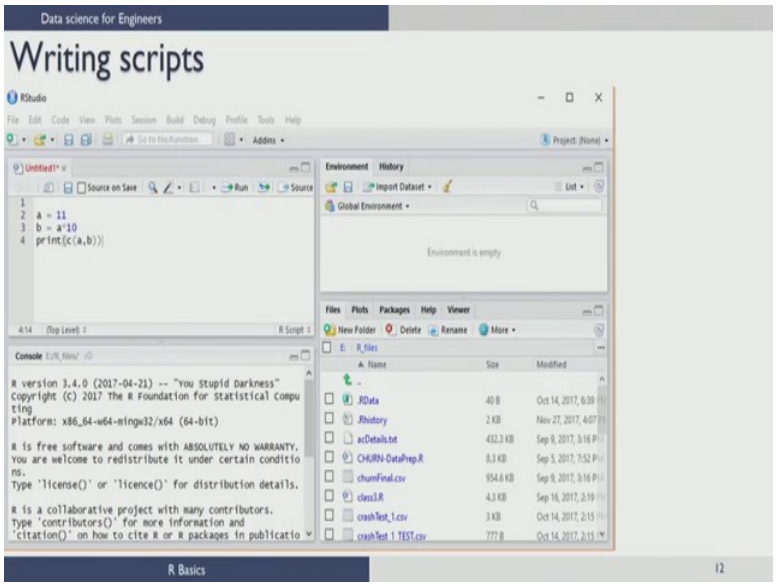
The other way is to use the + button, that is just below the file tab and you can choose R script, from there, to open a new R script file.

(Refer Slide Time: 05:30)



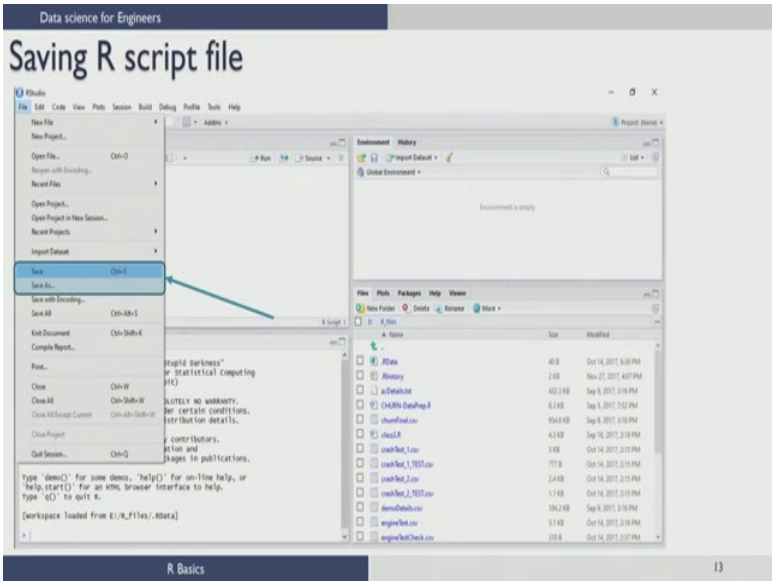
Once you open an R script file, this is how an R Studio with the script file open looks like. So, 3 panels console environmental history and files and plots panels are there. On top of that, you have a new window, which is now being opened as a script file. Now you are ready to write a script file or some program in R Studio.

(Refer Slide Time: 05:51)



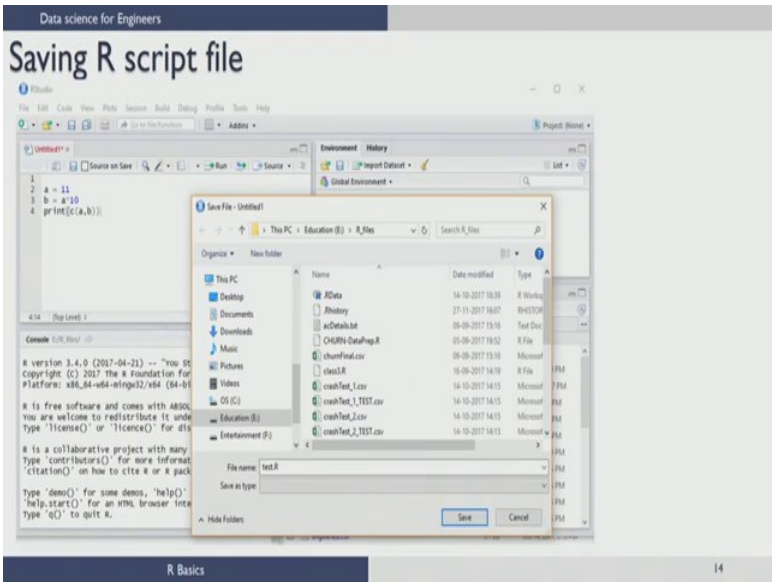
So, let us illustrate this with a small example, where I am assigning a value of 11 to a, in the first line of the code which I have written and you have b which is a times 10, that is the second command, I am evaluating the value of a times 10 and assign the value to the b and the third statement, which is print c of a, b concatenates this a and b and print the result. So, this is how you write a script file in R. Once you write a script file, you have to save this file before you execute it.

(Refer Slide Time: 06:37)



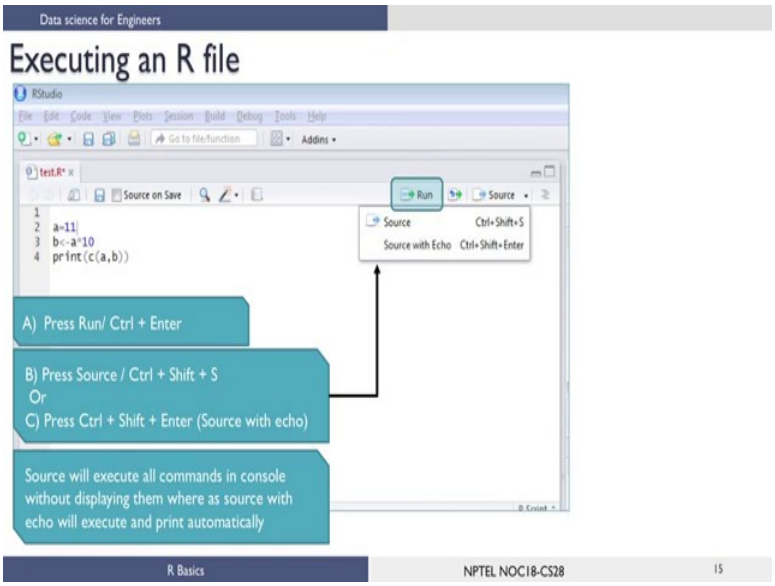
Let us see, how to save the R file. From the file menu, if you click the file tab, you can either save the file, when you want to save the file, if you click the save button, it will automatically save the file has untitled x. So, this x can be 1 or 2 depending upon how many R scripts you have already opened, or it is a nice idea, to use the Save as button, just below the Save one, so that, you can rename the script file according to your wish. Let us suppose we have click the, Save as button.

(Refer Slide Time: 07:14)



This will pop out a window like this, where you can rename the script file as test R, are the one which you are intended to. Once you rename, you can say save, that will save the script file.

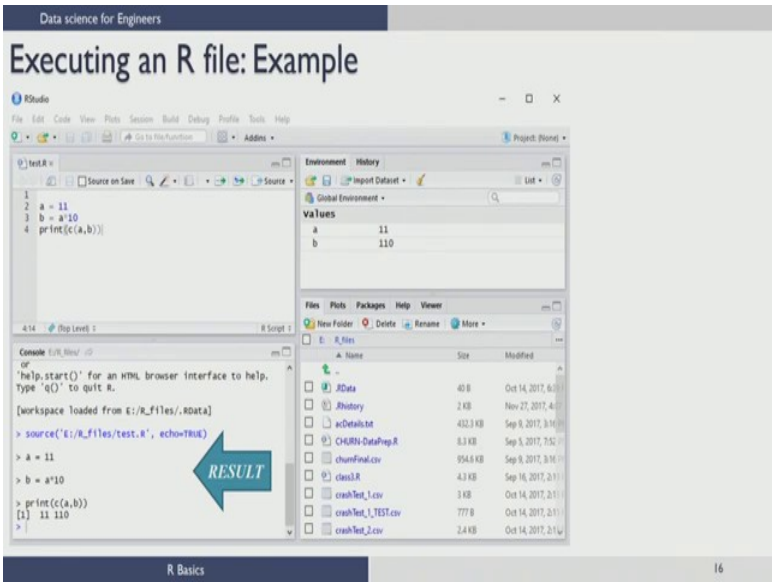
(Refer Slide Time: 07:31).



So now, we have seen how to open an R script and how to write some code in the R script file. The next task is to execute the R file. There are several ways you can execute the commands that are available in the R file. The first way is to use run command.

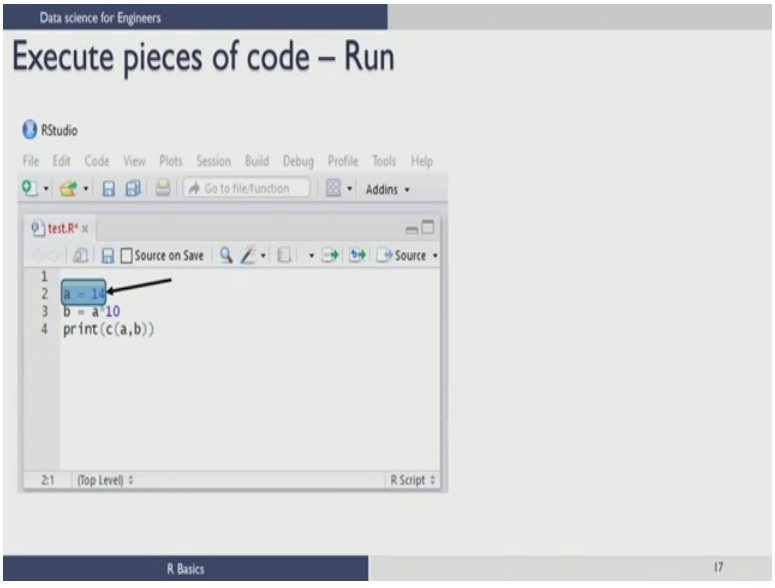
This run command, can be executed using the GUI, by pressing the run button there, or you can use the Shortcut key, this is control + enter, what it does is, it will execute the line in which the cursor is there. The other way is to run the R code 'R' using source R source with echo. The difference between source and source with echo is the following: The Source command executes the whole R file and only prints the output, which you wanted to print. Whereas, source with echo prints the commands also, along with the output you are printing.

(Refer Slide Time: 08:38)



So, this is an example, where I have executed the R file, using the source with echo, you can see, in the console, that it printed a the command a = 11 and the command b = a time 10 and also the output print c of a, b with the values. So, a = 11 and b = 11 times 10, this is 110. So, this is how, the output will be printed in console. So, that is the result.

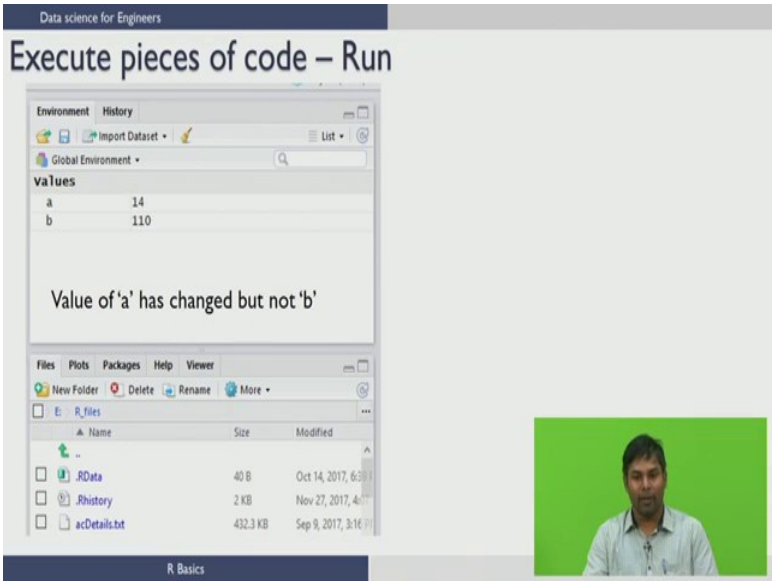
(Refer Slide Time: 09:14)



Now, let us see how to execute the pieces of code in R. As you have seen earlier, you can use run command, to run the single line, right. So now, let us try to assign value 14 for a and then try to run it. So, how do you do this? Take your cursor to the line, which you

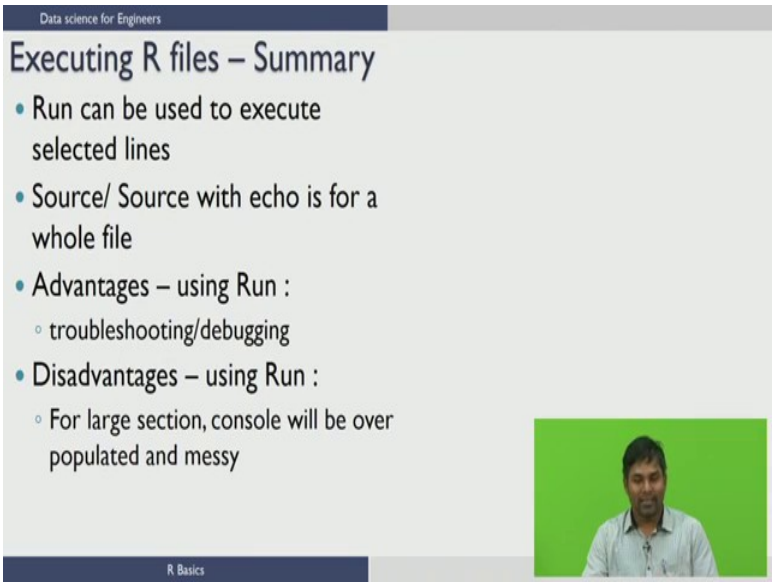
want to edit, replace that 11 by 14 and then use control enter or the run button. This will execute only the line, where the cursor is placed.

(Refer Slide Time: 09:42)



In the Environment pane, you can see that, only value of a, has been changed and the b value remains same. This is because, we have executed only the line 2 of the code, which change the value of a, but we have not executed the code of line 3. So, the b value reminds as is. Value of a, has changed, but not the value of b.

(Refer Slide Time: 10:05)



In summary, we can say that, Run can be used to execute the selected lines of R code. Source and Source with echo can be used to run the whole file. The advantage of using Run is, you can troubleshoot or debug the program when something is not behaving according to your expectations. The disadvantages of using run command is, it populates the console and make it messy unnecessarily.

In the next lecture, we are going to see how to add comments to the R file and how to add comments to the single line and multiple lines etc.

Thank you.