**Multivariate Procedures with R**

**Prof. Shalabh**

**Department of Mathematics and Statistics**

**IIT Kanpur**

**Week – 01**

**Lecture – 04**

**Introduction to R Studio**

Hello friends, welcome to the course Multivariate Procedures with R. So, you can recall that in the last lecture we had talked about the data editors and we had used the data editor inside the R software as well as inside the R studio and we had some idea that how this R studio will look like. And as promised in the last lecture, in this lecture also we are going to continue with the R studio software. Well, R studio has many features. But in this lecture my objective is to make you familiar with this software. I will not be in a position to explain you all the functionalities of the R studio software for which I depend on you that you will learn them yourself.

My idea here and my objective in this lecture here is that I try to give you some basic concept that how are you going to work in the R software using this R studio. And remember one thing, this R studio is a software which is visible only on the face. The background material is R software. Whatever are the executions, whatever is happening, R software is the soul.

The only advantage is that those various types of option, commands, they have been organized in such a way which makes the life of a programmer easier. And more you practice on this R studio, better you will learn. There are many options once again I would say which you have to learn yourself. Where you are going to learn, how to learn the R studio and move forward. So let us begin our lecture. Right?

So now in this lecture when we are talking about the R studio software, so as I said that this is only an interface between R software and us. And it is actually more useful for the beginners that who are starting with the R software because it makes coding easier. But definitely you know that when you are trying to write your own program, then you want to write a function in which you want to write many many things which are to be executed in an automated way. Now this R studio when you try to use the built-in

commands by clicking on a button, then you may not be able to understand that what is the command which is working behind the scene.

 And the same command has to be written when you are trying to write down the function. So, it is important and this is my advice that better you work or at least learn those commands also, you can use whatever you want depending on your choice and convenience. Right? So, as I told you in the earlier lecture that R studio is a software which is written in the C++ programming language and it is also a free and open-source integrated development environment for R. So, it is free, you can download it and you do not have to pay any cost.

 Yes, there is a paid version also but for us it is sufficient to have the free version. So now this software is available at this website. Right? This is posit.co and if you try to just navigate in that website, you will reach to the R studio software. There are some other software also but then you have to see what you want on that site.

 So as soon as you go to this website, you will see here this type of screen, where there is an option for download R studio, download R studio server etc. etc. And so you come to this here download R studio and this file will be downloaded. You can double click on the file and then this R studio will be installed on your laptop or computer. And as I shown you in the last lecture, there will be 4 windows here.

 Let me call here window 1, window 2, window 3 and window 4. So, window 1 is used for writing the syntax and commands. Window 2 if you try to see here, this is nothing but your R GUI window. This is R software. Window number 3, there will be some environment history where you can see the variables which are stored in your laptop or computer.

 And then in the fourth window there are several options, file, plots, packages, health, viewer etc. So, this is how it will look like. And I can show you here that if you want to change the size of these windows, you can work, actually this partition here is like this. I can mark this thing and if you try to move your mouse and try to drag these lines, you can change the size of these windows here in this direction, this direction, this direction or this direction. Right? So, this is how it will actually work.

 So now I give you here a quick background about what are you going to do in these different windows. So, in the window number 1, you can see here, this is basically the script section and then you are going to write your commands etc. that you want to execute here in this case. Right? And basically, if you try to see here, there are different options. First if you try to see these menus are usually the similar menus which you see in other software, these icons you are also familiar with those things, so I need not to explain them.

And here also you can see here, this is here save button, search. The main button which I want to show you here is the run and here source. Right? So, source is used to connect this R GUI window or and between the function and R-U-N here run, this is used to execute your command. Right? And similarly, if you try to look here that if you want to open a new file here to write your command, then you simply have to come here at this place and then you will see here, there is a point here R script. R script means it will open a new script file where you can type your command. Right?

This is here. This is here means another important thing for us that if you press here at this place, this is for saving the script that whatever you are trying to write down as a different types of command function in a particular sequence, they are going to be saved here. Then another important point where I would like to have your attention is over this run. So as soon as you highlight the or a set of commands which you have written in the script section, after highlighting them you can click here run. Right? And then this syntax will be executed over the R console. Right? And then here this next button after this run is the rerun that if you want to execute it again, then you can see here rerun.

Beside those things if you try to see here, there are different options here R notebook, R markdown, shiny, plumber, text file, etc. means as I can tell you that this R has developed in an enormous way in the last decade. And many options been added and so those options also have been added in the R studio software which people in executing those programs. Right? Now if you come to window number here too, you can very easily identify how it looks like. This is your here window number 2.

This is nothing but your R GUI window. So, this is actually the soul. Here your R software is being actually executed. Right? So, whatever you try to highlight in the script section and execute it by pressing run, they are executed here. Right? So here I do not need to explain you more because we already have worked a lot in the R GUI window and this is the same thing.

So, this is called as console. Now the third window is about the environment window. So if you try to see here, it looks like this. So, you can see here there are different options. That first option let me try to take you here.

This is import dataset. Whenever you are working in statistics and if you want to analyze any data, then there are two options. Either you can enter the data directly inside the software or the data is coming from some external source. From some external source means the data is stored in some TXT file, CSV file, data spreadsheet, etc. etc. and you would like that data to be read inside the R software so that it is converted into a framework which R can understand it.

In order to do it, there are different types of commands and sometimes we use different types of special package so that R can understand. For example, if you want to read file

which is created into this MS Excel software, then you need a special software or a special package which has to be installed and once you have to load that package, after that only you can read this dataset. So anyway, so we have got different commands to import the dataset and this command helps you that just by clicking on the file; by choosing the suitable file and just clicking on it, it will directly import the dataset into the R software. So, you need not to remember the command. Right? So, this is for this thing that the data can be imported from other files by clicking here.

Now there is a command here, history. So, history will tell about the different types of variables and details which have been used in this laptop in the past. For example, if you are doing something and tomorrow you will be doing something, so all those things whatever you have done in this laptop or computer that is available in the history. Now if you try to see what is the main role of this window. This window gives you all the details about the dataset and variables which are going to be stored here, right.

Suppose if you try to store up some variable here x equal to 1, y equal to 2 and x is a matrix, y is a matrix, etc. etc. So, all such information is readily available in this window. Because you will agree with me that whenever we are trying to use any R command, it depends that whether the data is in which format, whether it is scalar, matrix or something else. So, you need to know the nature of the data and this is available directly from this window.

So, this is how it is going to be useful. And now this icon here, this icon is if you try to click here, then the stored value can be erased. So, this is how we try to work in this window number 3. Now about window number 4, this is very detailed and there are many options which are available here. Right? So, I will try to give you some quick idea and just you will try to see when you try to execute subcommand yourself on the R studio.

So, here is about the files that means all the files which are stored in this laptop, they will be visible here. And then the next option here is plot. So, whenever you are trying to create any graphics, then this is the window or this is the part where your graphics will be visible and you can make different types of operations for those graphics. Then you have here packages. So, as we had discussed earlier that there are two types of packages in R software.

One are built-in and another are the contributed. Contributed means different people are trying to do different type of work. They try to create the package so that other people can use their work and after reviewing, once they are approved, they are available on the website of R software and whenever we want to use them, we have to install them here. So, all those packages which have been installed on the laptop on which you are working, they are visible here, number 1. And then you will see, in the tutorial I will show you when I come to the R console that in this if you try to click over this package, then there

will be different option that if you want to install a new package or if you want to update any new package, etc., etc. Right?

Then after that you can see here the second last option here is help. So, if you remember in the beginning of the course, we had talked about different ways in which the help can be seen in the R software. And here if you try to click, all those options will be visible here and you can find out that how would you like to take the help. Then it is here last here is viewer. That means all your graphics which you try to create, they can be viewed in this part.

So basically, if you try to see this whole window 4, this is related to the output of the program and different outputs will appear here in different ways. Right? So, for example in this fourth window suppose if you want to know about these packages, yeah, that is important so that is why I am trying to show you here. Yeah, this screenshot may look different when you try to look into your computer because it varies that due to the installed packages. Means I have installed some packages so it is appearing in my computer. But if you try to install some other packages, then they will be visible here.

So, the first option here is install. As soon as you press here install, a pop-up window come and then where you have to type the name of the package which you want to install. And then it will, and then you say click click and it will install the package on your laptop. Then the next option here is update. That well these packages they also get updated over a period of time with the new development and the package which is installed on your computer that was available on the date when you installed it. Now if there is any change, if the package has been updated, if you press here update, then all the packages will be updated and so on.

So, if you try to see here, let me try to, please try to have a view on this part. Suppose if I try to see here, say agree here it is agreement and BayesM. Right? So, these are two packages which I had installed on my laptop. So this agreement package is used for statistical tools for measuring agreement and similarly BayesM is used for Bayesian inference for marketing or say micro-econometrics and similarly there are other packages composition, deoptium, energy etc.

So, this is indicating that these many packages are available in your laptop in your software. Now suppose if you want to load any packages, then how are you going to do it? So just in this box you have to make a, for example here I can bring my cursor and see here check. As soon as you do it, this package will be uploaded in your RGI window that you can see there. The library function will be executed and now this package is ready to use. So, you can see here it is more convenient to load the package and you need not to remember any command or etc. to do it. Right?

And similarly, if you try to look into other option here help which is here help. So suppose if you want to know something about some function, suppose I want to know about histogram. So, if you come here in this window and as soon as you start typing say H, I, S, T etc. something will start coming in the bottom and then you can choose here what you want to have and as soon as you say here hist, so this will come and all the details about the histogram they will be, they will appear here. And after that you can choose what you want, how much you want to read and whatever the help for histogram you had obtained in the RGI window, the same information will also be available here. Right?

Now let me try to show you one example that how are you going to execute it in the RStudio and then I will try to take another example to illustrate that how are you going to work on it. Suppose I want to create a histogram. Now what you have to do, you have to concentrate on these four, these windows that how the things are going to change. Right? So, if you try to look at the first window here, I have given the data that x is equal to c 1, 2, 1, 1 so on and then I am writing here the command HIST x which is the command to create the histogram.

Now I try to highlight it and then I press here run. That is my step 2. Step 1 is to type the commands here; step 2 click on the run. Then you will see here as soon as you press here run this HIST x will appear here and then yeah, this data will also be appear here which is not visible here but it will be there. So, all those commands which are getting executed in the window number here 1, they will appear in the RGI window, right. And whatever is the value which you have defined here x, this will now appear here and it is trying to indicate you the nature and values which are stored in x.

So, it is a number. There are 13 numbers, 1 colon 13 that is indicating it and the values are 1, 2, 1, 1 so called. And if you try to click over there then it will give you all the values, right. And now as soon as you this command hist(x) is executed a histogram will appear here in the fourth window. Right? Now there are different options available.

There is option here zoom if you want to make it bigger. Yeah, because the window is small so it will look smaller and if you want to have a bigger view you can simply zoom it. Now as I told you that these graphics can be stored in different formats so say well image, JPEG, say well PDF, copy to clipboard so that you can paste it into other applications, other software, this is possible. Right? And then if you want to delete the graphics, clean the graphics then these options are here. So, you can see here well as soon as you try to work in the first window then all the fourth windows become active and the relevant information is stored at different places. Right?

And here you can see here this is here the partition. This is like this. So just by dragging over this red colour line which I am marking here, you can control the size of the window. You can place the window at appropriate place.

This window can come here. This window can come here. I would say that you try to experiment with the R software and try to do it here. Now let me try to take this example in the R studio so that I can show you that how are you going to work but before that I can show you with the screenshot and then I will try to show you in the R studio. So, suppose I want to create a histogram of some values which I have stored here in some variable here x and if I execute the command here HIST x inside the parenthesis then this type of histogram will come. Right? And if you want to do it in the R software, so you can see here that here I have this type the command like here this and as soon as I highlight it and then I see here run, the same values are coming here and then HIST, the command is executed and yeah in this window somewhere in the x you will find out the value of here x which will be the details about this x data vector and in the fourth window you can see here the same histogram which you have seen here. So let me try to execute these things in the R studio window so that you get more convinced. Right?

So, you can see here I have written here the data x here and now if you try to write down here hist(x). Right? So now if I try to highlight both the things here then you have to observe what is going to happen. Right? Here in this window where I am trying to move the cursor, I will show you that the same values of x will be stored there and these two commands they are going to be executed at this place and somewhere here in the fourth window you will see that the histogram will appear. So now let me click here on the run, so you can see here as soon as I run this commands that x is executed here and then histogram of x is executed here and this histogram is coming here and somewhere you can see here this data. Right?

This is here, this is 1 to 100 values 91.9, 51.7 etc. and you can see here these are the same values which are here 91.9, 51.7. So, you can see here now this histogram is created here. Suppose if you want to export it, save as image, save as PDF, copy to clipboard etc. Right?

If you want to see here zoom you can see here it looks bigger, it is easy to look into this thing and then if I go on here package you can see here that in this laptop the packets like abind, AER, Alabama, Amelia etc. they are present here. So let me try to show you, so let me try to clear this console window so that you can see what is happening. So now if I want to load this package abind, I have to simply come here and say click and it will be checked and you can see here this library A-Bine is loaded. And similarly, if I try to say here install, so install you see this pop-up window is open and then means yeah you type the package here and it will be installed here.

But definitely when you are trying to use this type of command then you should have an active internet connection because this control is going out of the R studio or R software to an internet site from where it is trying to download and it is trying to install it on your computer. Right? And yeah, if you want to now means say you had loaded this package abind and now if you do not want to upload it but you want to remove it just see here check once again and it will use the command here detach and it will unload the package. Right? And similarly, if you try to go on different aspects, suppose if I see a history so and then you can see here whatever commands I had used in the beginning of the lecture they are available here.

If you try to recall we had used the command here example lm demo perspective demo graphics which I used in the earlier lectures they are available here so that I can see what I had done in the past. Right? And yeah, means so and here if you try to see here if you want to get here any help suppose as soon as I see here H I S T, hist you can see here this hist is coming here and as soon as I say this details about the histogram, they are available here. Right? So now this is a brief introduction about the R studio and there are many more functions which are available in this software but my request will be you please try to have a careful look and try to play with these different options. There is no harm.

Your laptop is not going to be damaged even if you press a wrong button. It will give you say that it is an error and then you have to execute it again. And you try to explore that what are the different options are available by which your life as the programmer can be made easy. Right?

How you can install the package how you can load them etc etc. And a good point is that whatever you are doing by clicking on the button the respective command is visible in the R GUI window. So that is a great help that sometime if I want to write some program and if I am not getting an idea what is the right command you can seek the help and then get the command do it here and then whatever is the correct command you can copy and paste into your text editor so that you can write the right function without any error. So, I would stop here but as I said I would repeat once again that after this I will not be talking about the R studio software but now, I have given you both the ideas working inside the R GUI window and working inside the R studio software. My objective was simply to introduce you that there exists some external software which are also helpful. Now whether you want to use it or not that is your decision how you are comfortable that you have to decide.

But I will be continuing from the next lecture for all our computation and demonstration only in the R GUI window. So, you try to have a look try to have a try to take some examples from your past knowledge about the R software and try to execute them and try

to make yourself familiar and comfortable and we will see you in the next lecture till then, goodbye. Thank you.