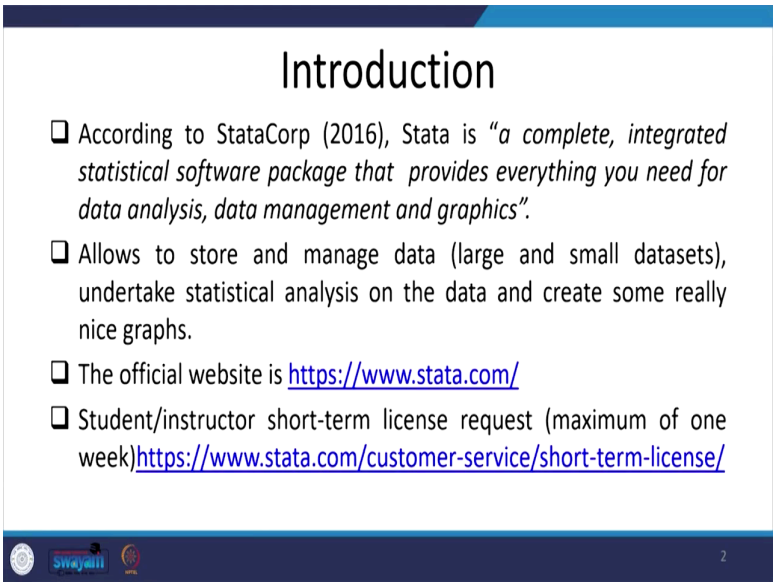
Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee

Lecture - 16
Basic Understanding of STATA

Welcome friends to this NPTEL MOOC module on Handling Healthcare Survey Data, we are on the 3rd week of our module. Myself Dr. Pratap Mohanty, I am a faculty member in the Department of Humanities and Social Sciences. This module has been assisted by Mr. Kamal and Mr. Milind and they are preparing some of the contents of the lecture and we are together developing it for your readings.

In this particular week, we have thought of including one lecture purely on understanding Stata, though in the previous week, we have applied certain aspects of Stata. I did feel that you guys require a certain understanding of the basic statistics or basic understanding of Stata. Now, without delay let us understand the genesis of Stata.

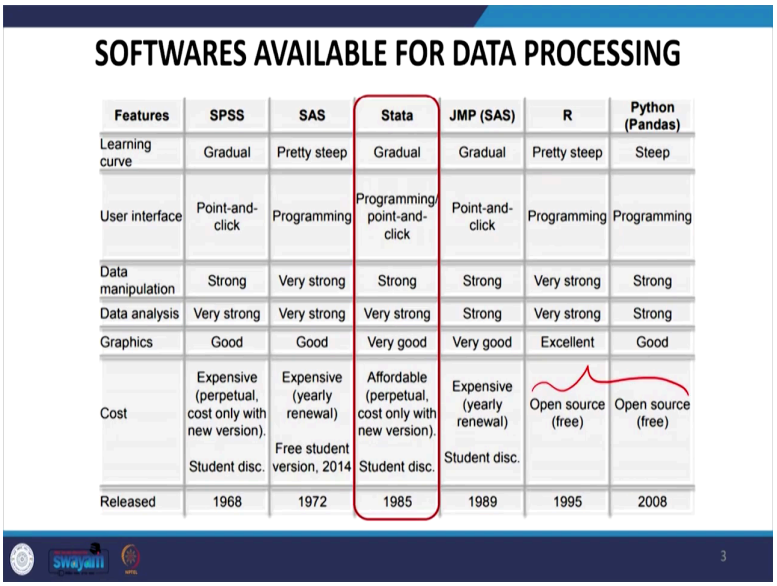
(Refer Slide Time: 01:46)



So, as per the Stata corporation (2016), Stata is a complete, integrated statistical software package that provides everything you need for data analysis, data management and graphics. Stata allows for storage and managing data of both large and small data sets. This helps in undertaking statistical analysis of the data and creating really nice analytical graphs.

The official website is mentioned here as https://www.stata.com/. Students' or instructors' short term license requests can get it through this website. The instructor license for a maximum of one week is actually provided by Stata for your operation. We are also doing the same in front of you all for your exercise. So, you guys can also hold it as early. Now as a comparison with other packages for data processing and analysis.

(Refer Slide Time: 03:04)

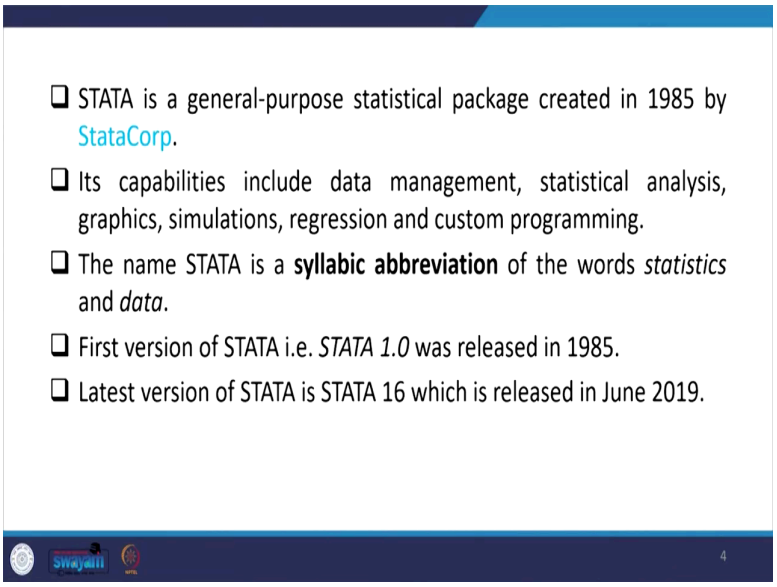


We have listed software like SPSS, SAS, Stata, JMP or SAS, R python etc. Now among all those comparisons one that is Stata has almost all indicators and its analysis is considered to be very strong.

This is also affordable though open-source software is also suggested like R or python, and these are free software for analysis. We all know that SPSS is widely used because it was initiated quite early, and people might have been using it for more time. Though it is more expensive than Stata.

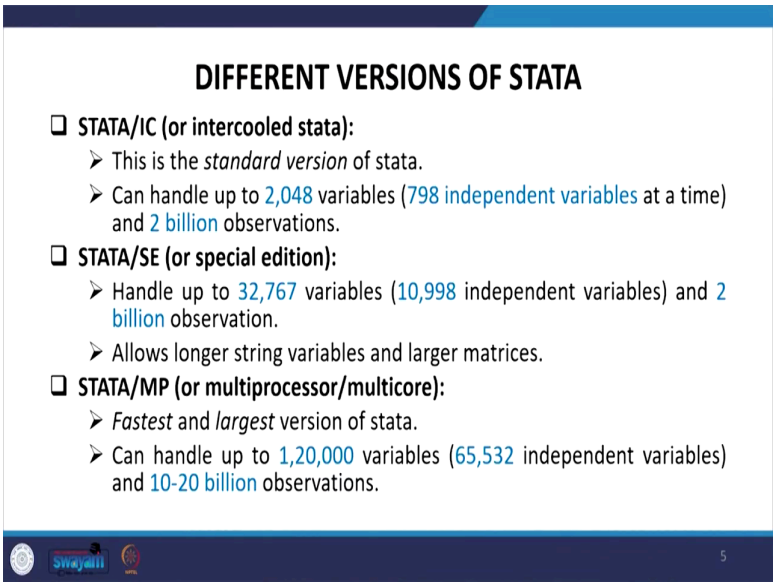
But regarding graphics and user interface there are certain differences between SPSS and Stata. Graphics is quite good in Stata and SPSS usually points and click-based though there is syntax also available.

(Refer Slide Time: 04:45)



Stata is a general-purpose statistical package created in 1985 by Stata Corporation. Its capabilities include data management, statistical analysis, graphics, simulations, regression, and custom programming. The name Stata is a syllabic abbreviation of the words statistics and data. The first version of Stata is Stata 1.0 was released in 1985, and the latest version I referred to in my analysis is Stata 16. Stata 17 has also been released.

(Refer Slide Time: 05:59)



Those are very new to Stata, I am explaining it to them and it is going to be very useful. Those who have already used Stata do not follow all the between the lines and you may skip

some of my contents. So, for first-time users of Stata, I think this is quite useful. Stata IC version, Stata SE version and Stata MP version, there are majorly 3 versions we used.

IC version is also called intercooled stata, this is the standard version of Stata, and this can handle up to 2048 variables, 798 independent variables at a time and 2 billion observations. Whereas the special edition of Stata handles up to 32767 variables, 10998 independent variables and 2 billion observations. This allows for longer string variables and larger matrices, which is another advantage of the SE version.

MP version of the Stata is the best one; it has the fastest and largest version of Stata. This can handle up to 120000 variables with 65532 independent variables and 10 to 20 billion observations.

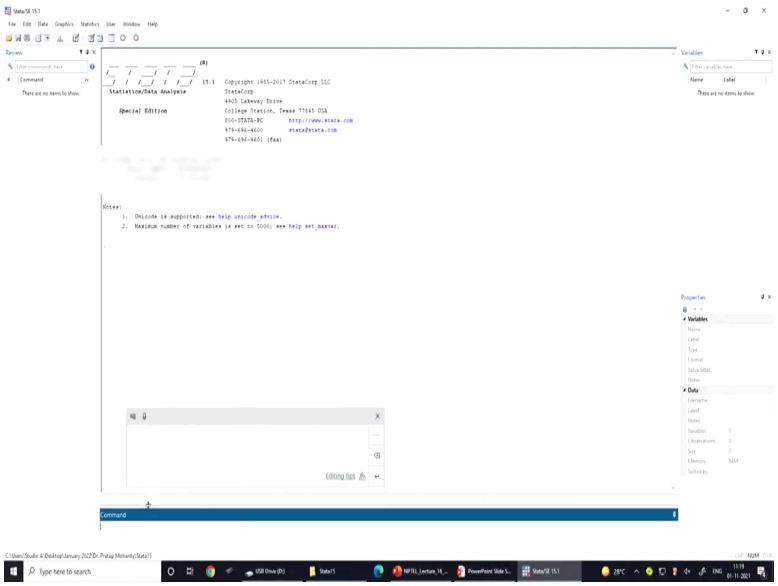
(Refer Slide Time: 07:51)

Now, how you guys can able to launch data on your computer, there are multiple ways to launch it. The first option is to go to the start > all programs > Stata then double click on the Stata icon on the desktop.

Then double click on a previously created Stata do-file. If you want to go with the explanation of the do-file, I will explain all those things. The do-file is called with the extension with .do. In this do file, we can save syntax and all our operations. By double click on a previously created stata data sets one data set can be opened it has the extension of dot dta.

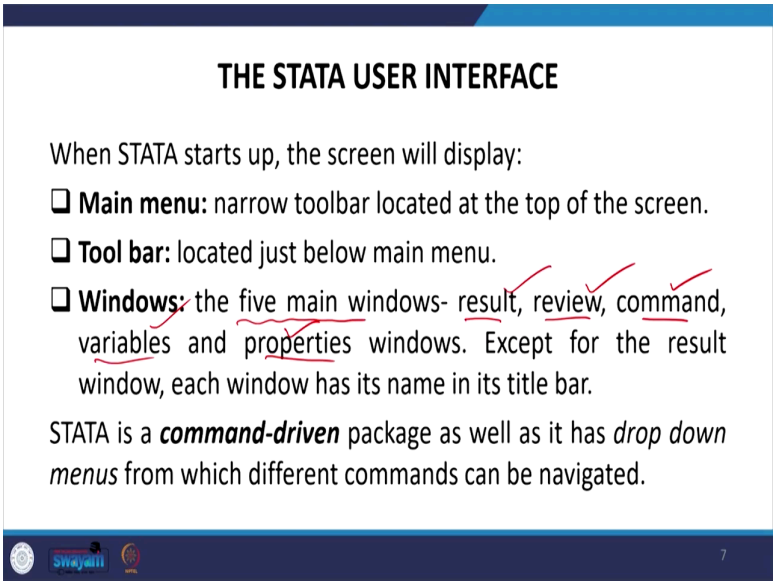
Now, this data you just interface how you should see if I can just open on your screen once then you can go to the respective field.

(Refer Slide Time: 09:37)



This is the variables section here on the right, then it is properties, then the command over here. Then you can check all those review of commands on the left file menu bar. I will explain all those things with my PPT.

(Refer Slide Time: 10:02)



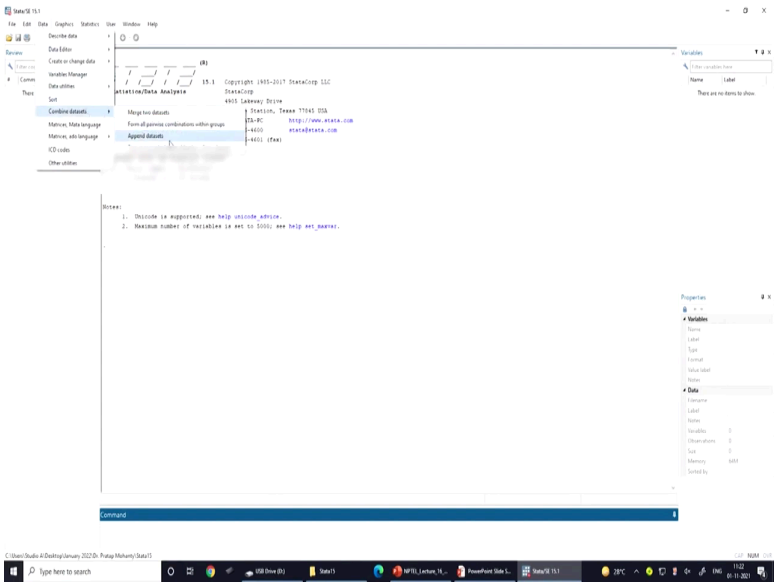
So, PPT is going to be very useful in clarifying all other details. There are three important aspects we deal with in Stata one is called the main menu, toolbar, and windows. Main menu what is that? Basically, this is the narrow toolbar located at the top of the screen. So, all those toolbars which are located on the top of the screen are called the main menu.

So, all narrow toolbar is defined on the main menu. Whereas the toolbar is in fact located just below the main menu. The file, edit, data, graphics, and statistics user window all are called the main menu. Within that there are some toolbars, toolbars are called open files, then save it.

Then what do you mean by windows? Windows display all the 5 main windows these are called the result window, review window, command window and properties window. So, variable windows 1st 2nd 3rd 4th and 5th windows are displayed on the main windows screen of Stata.

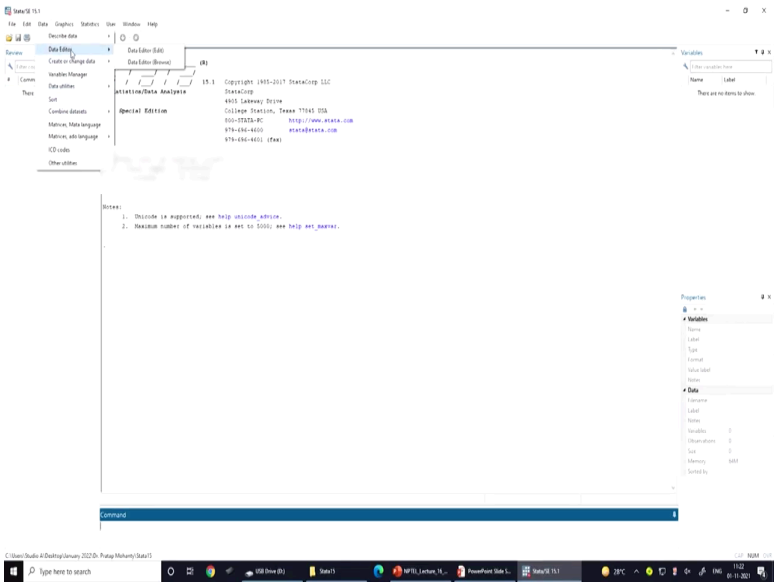
Then variable is here, properties here, the review is here, and the command is here. So, you can able to get all those things through the operation. Except for the result window each window has its name in its title bar. So, from the drop-down menu, we can also able to find out results as well.

(Refer Slide Time: 12:48)



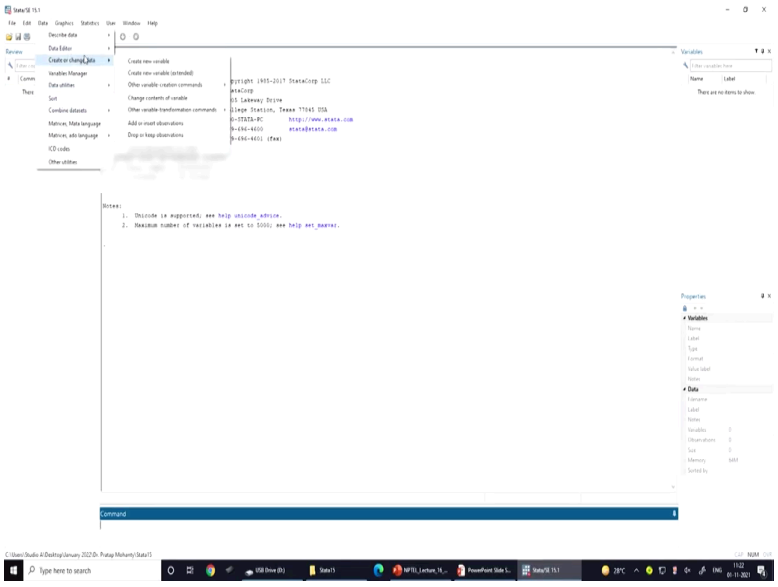
So, a drop-down menu like if you want to go to data then you want to combine data set it has options. If you want to go to data, then you want to sort the data, you can do it.

(Refer Slide Time: 12:59)



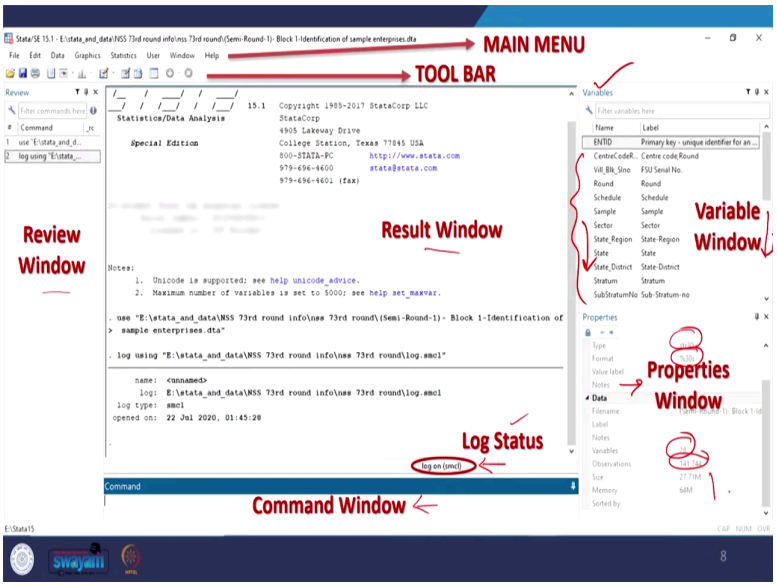
Like you can edit the data.

(Refer Slide Time: 13:00)



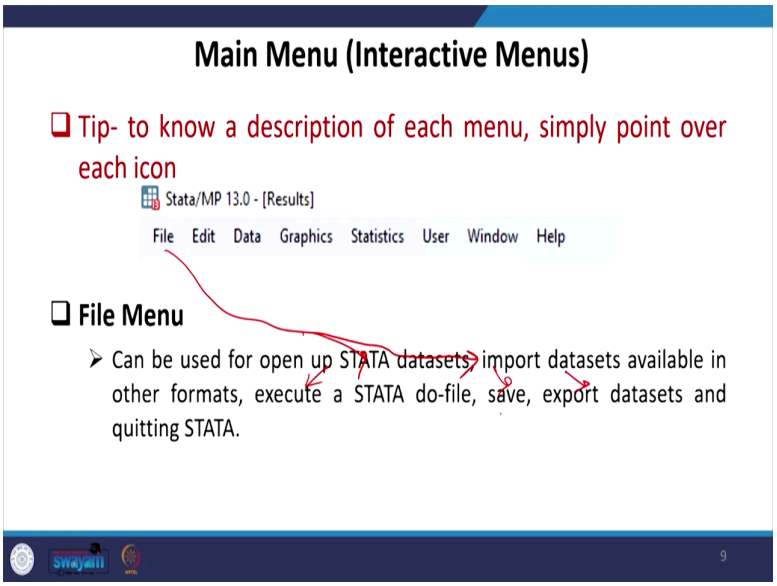
So, these are various options possible through Stata. So, there are 2 options one is called command-based another is called drop-down menu based.

(Refer Slide Time: 13:11)



These are all explained here in the main menu, as I already mentioned that the toolbar is mentioned here, the review is mentioned here, the result window is here, and the log status is here. There is some clarification also required on the log on with SMCL. The properties window explains the type of your variable. Regarding the variables, I have already said that it shows all such variables in this list and its labelling is also defined.

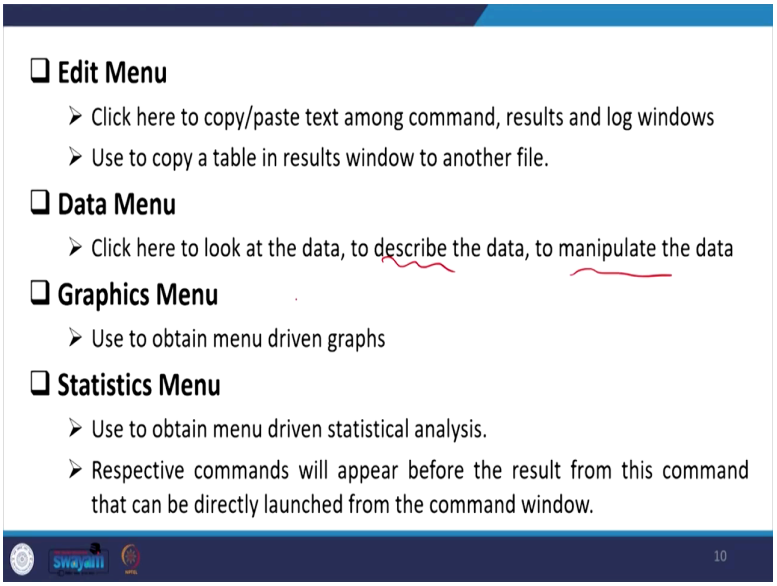
(Refer Slide Time: 14:15)



So, on the Main Menu, we have Interactive Menus as well. Tip to know a description of each menu, simply point over each icon it will show whether it is a file or edit or data graphics etc.

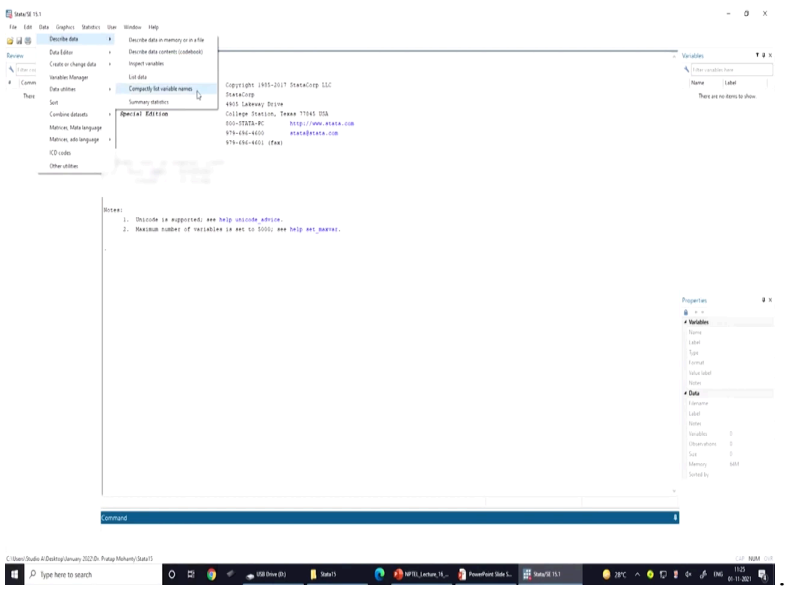
The file menu can be actually used for opening the Stata data sets import data sets available in all the formats. So, on the file, we can get so many things like import, execute, save, export etc.

(Refer Slide Time: 15:04)



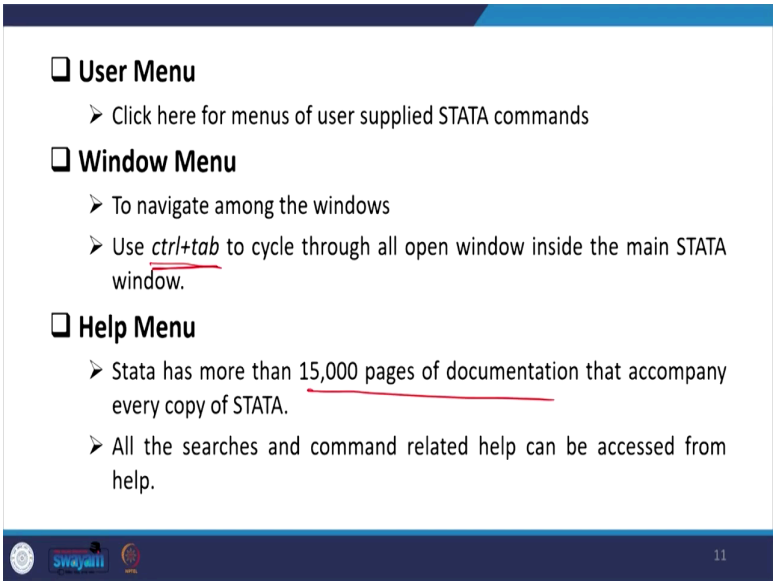
So, for as the edit menu is concerned, we can able to copy or paste text among command results and log windows. This is useful to copy a table in the results window to another file. In the data menu, we can able to get information about the data like describing data or manipulating data.

(Refer Slide Time: 15:34)



Describing data or manipulating data you can easily do that. Once your data is opened, we can describe the data. The graphics menu is useful to obtain different quality graphics as we already mentioned that Stata gives better graphics than the SPSS. From the command window, we can get all statistics.

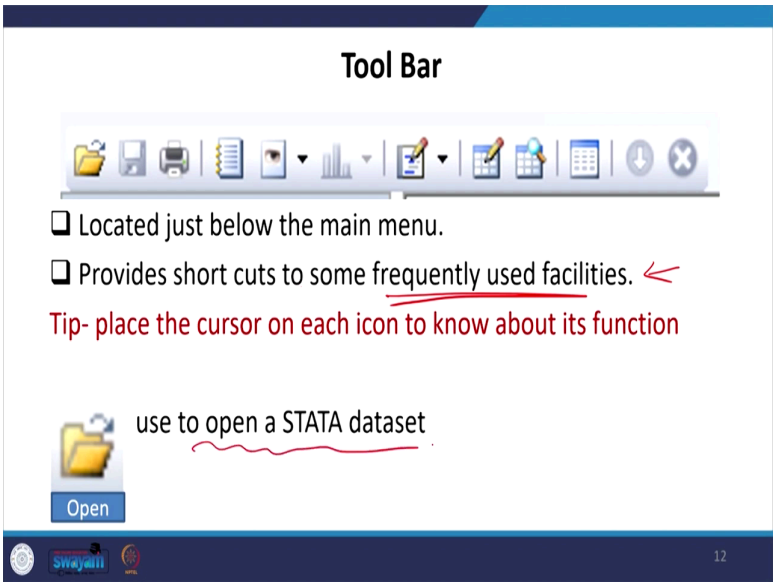
(Refer Slide Time: 16:35)



Once you click on the user menu you will get various possible Stata commands. The window menu helps in navigating among the windows from one to another one, we can use the control plus tab to cycle through all open windows inside the main Stata.

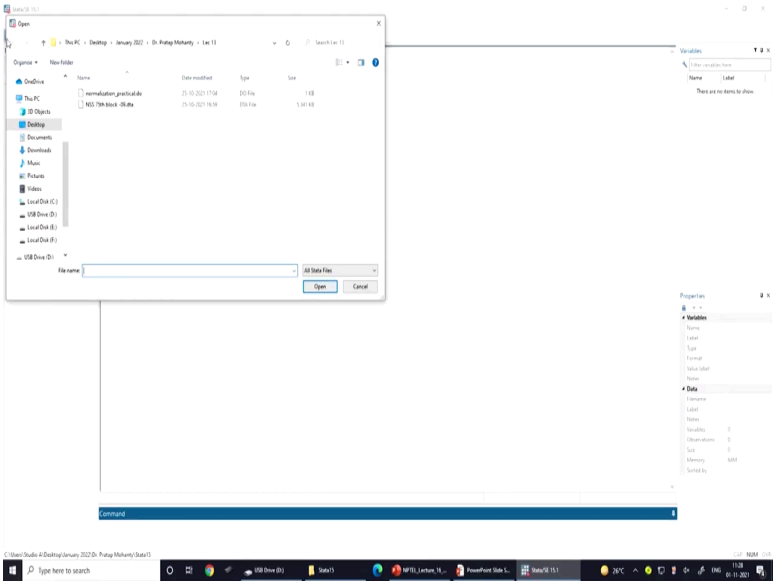
The last one we are also explaining is called the help menu. The help menu is most important. If you are stuck somewhere, and you are not getting the right direction simply type a keyword whichever is striking to you on the help menu. So, this gives searches Stata has more than 15000 pages of documentation that accompany every copy of Stata. All the searches and commands related to help can be accessed from this help menu.

(Refer Slide Time: 17:54)



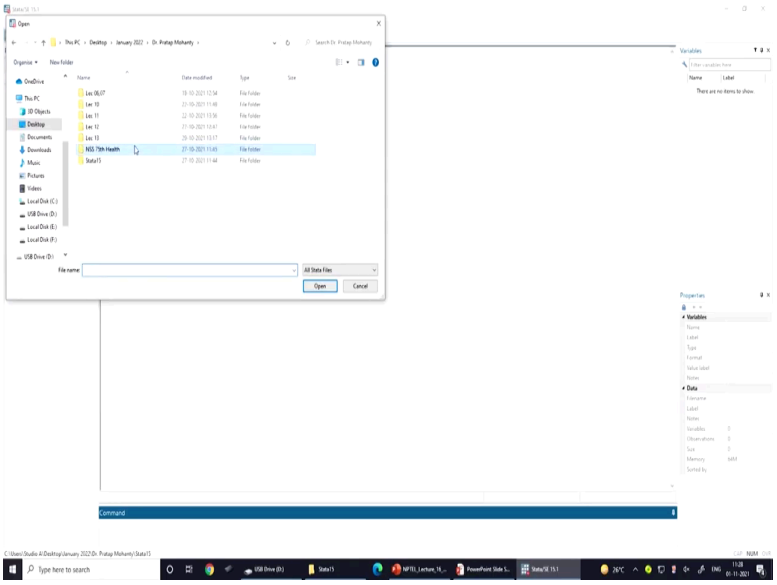
So, this is the toolbar located just below the main menu. This provides shortcuts to some frequently used facilities. So, frequently used facilities are actually explained or given in the toolbar. Once you keep your cursor over that particular icon, it will display its name. How to use Stata data set? Once you click here you can go to file or go to this folder open.

(Refer Slide Time: 18:41)

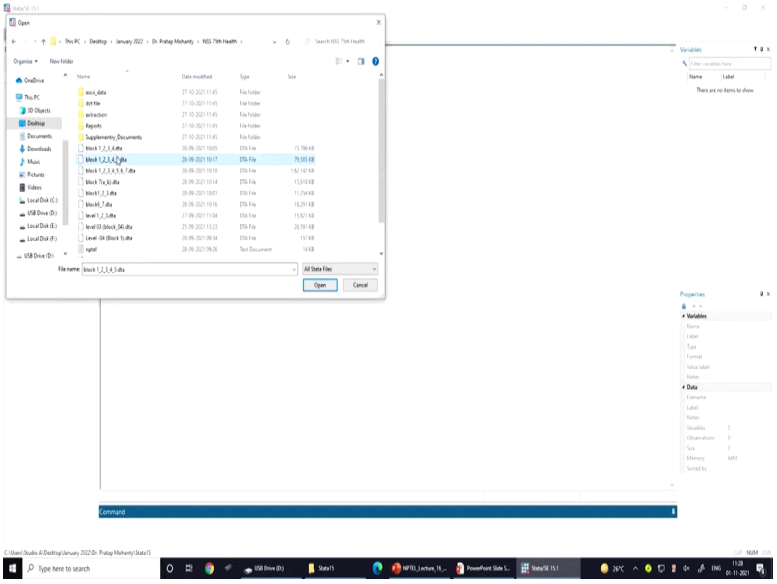


From here wherever you are kept the data you can simply click on that and it will be displayed with the data set.

(Refer Slide Time: 18:54)

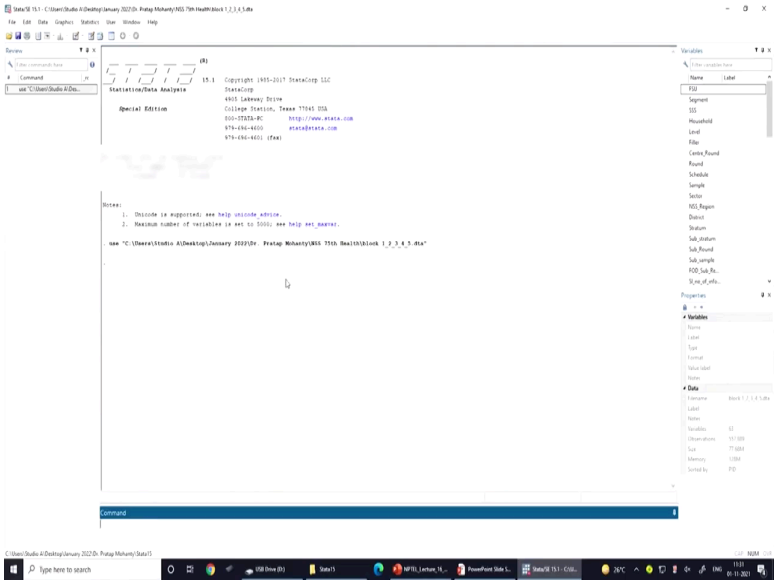


(Refer Slide Time: 18:55)



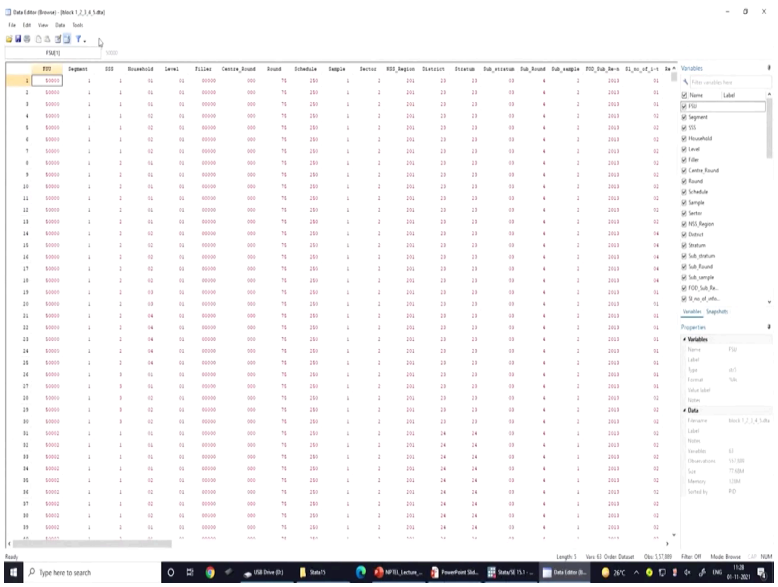
That we can also do it is here. There are some of the data files and any data set we can open.

(Refer Slide Time: 19:02)



So, we have already opened a data set here.

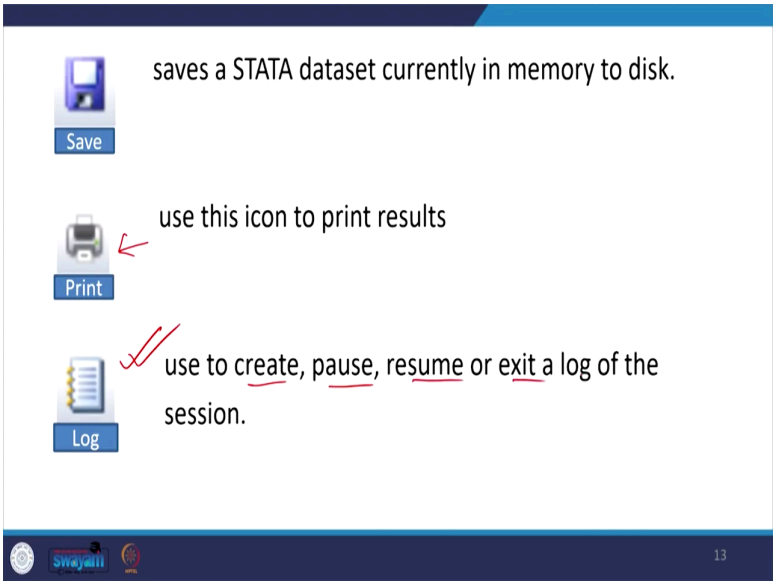
(Refer Slide Time: 19:09)



(Refer Slide Time: 19:13)

This is the browsing data. Why are these in different colours? Some of them are in the red colour, I will also explain them to you later. I have already explained how to open the data set in Stata.

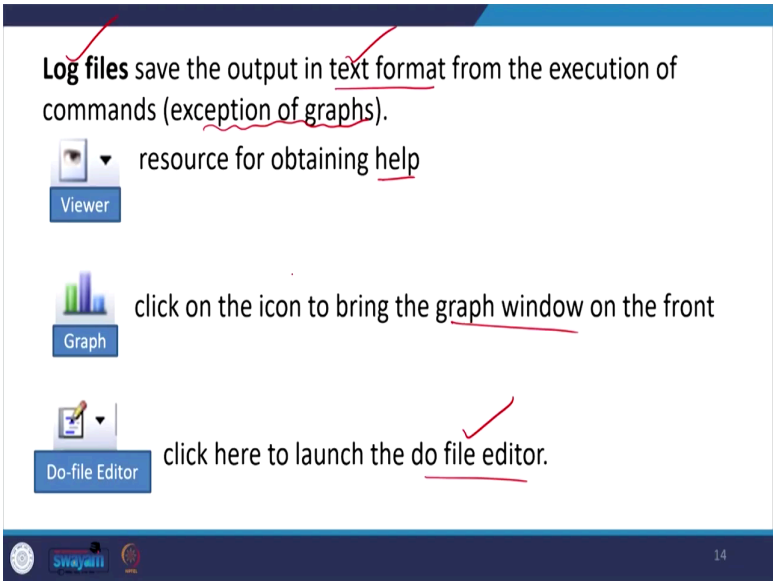
(Refer Slide Time: 19:29)



You need to save the data set currently in memory to disk. Then use this print icon to print the results. Once your result is derived you can print it. The logbook is very important. Now, this is very helpful to create, pause, resume or exit a log of the session.

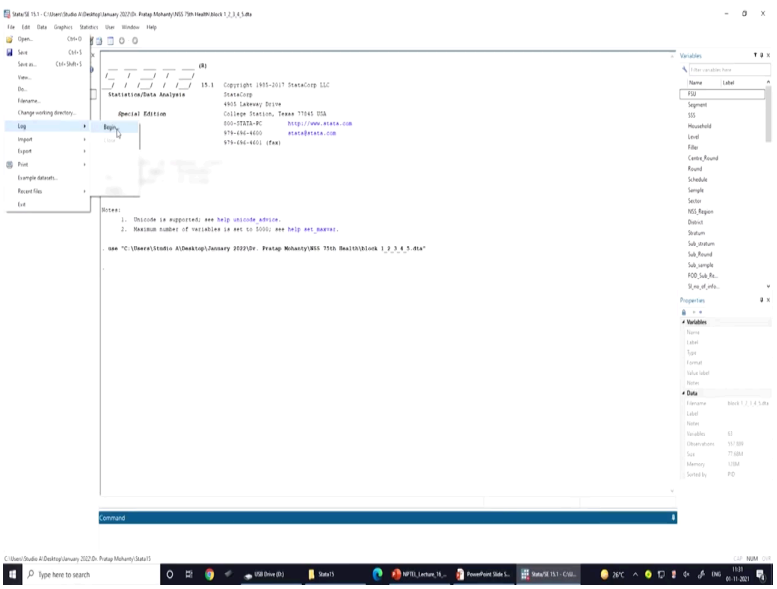
So, there is the various function you can able to do through the log file. How to do it we will also tell you right now. Logfile saves the output in text (txt) format.

(Refer Slide Time: 20:38)



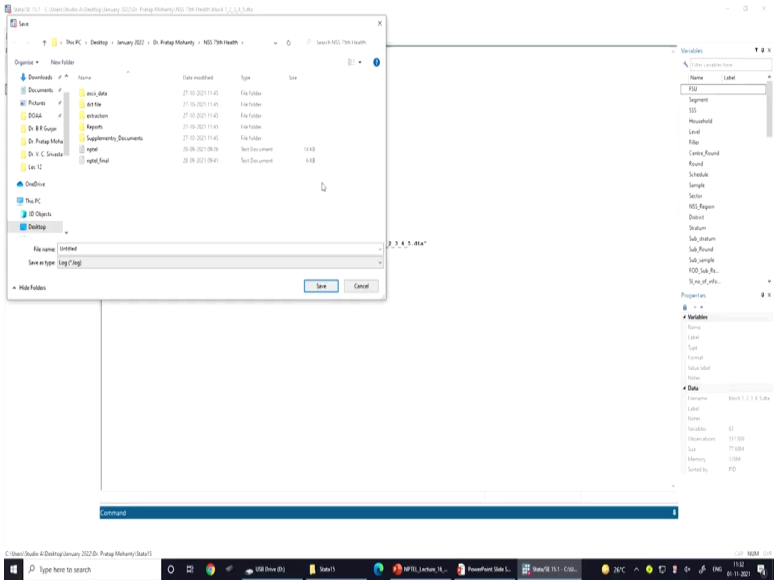
Log files save the output in text format from the execution of commands (exception of graphs). So, graphs are not taken to the log file, but the rest of the commands could be taken to the log file. Like the first icon is for help, the second one is actually on a graph window, and then the third one is on the do-file editor. Click here to launch the do-file editor. I will also guide you on what you mean by the do-file editor, and what you mean by log files.

(Refer Slide Time: 21:31)



So, one check just in a second, I will guide you to go to file and click on the log then click on begin.

(Refer Slide Time: 21:37)

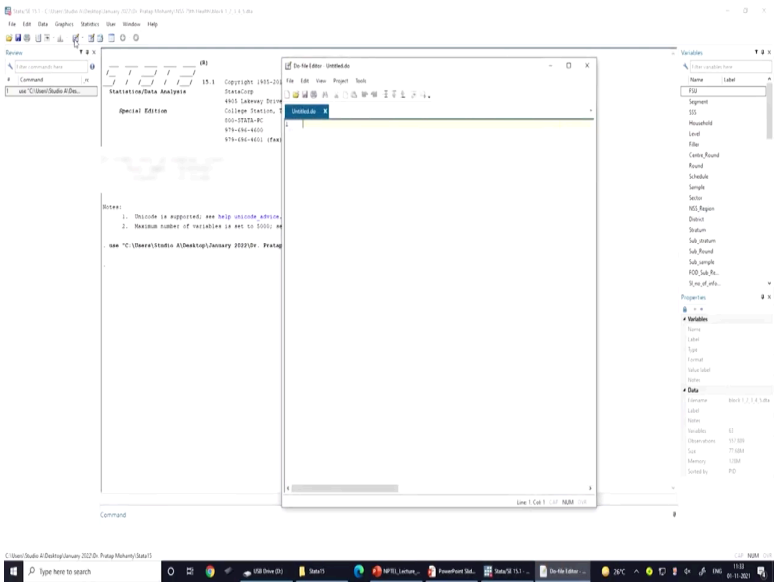


Now once you click on begin it directs you to save it with a name. you can take the name as log 75th round. By default, it has taken the dot SMCL format. This indicates that you are going to save the Stata result as it is like Stata. If you are doing it dot log, then by default save the results in txt format except for the graphs.

The dot log format is going to create a text file and it consumes significantly less space, these are in a notepad file. So, this will be in some kb but if it is in SMCL then it takes like a picture. If you save it on your computer this will be consuming huge space.

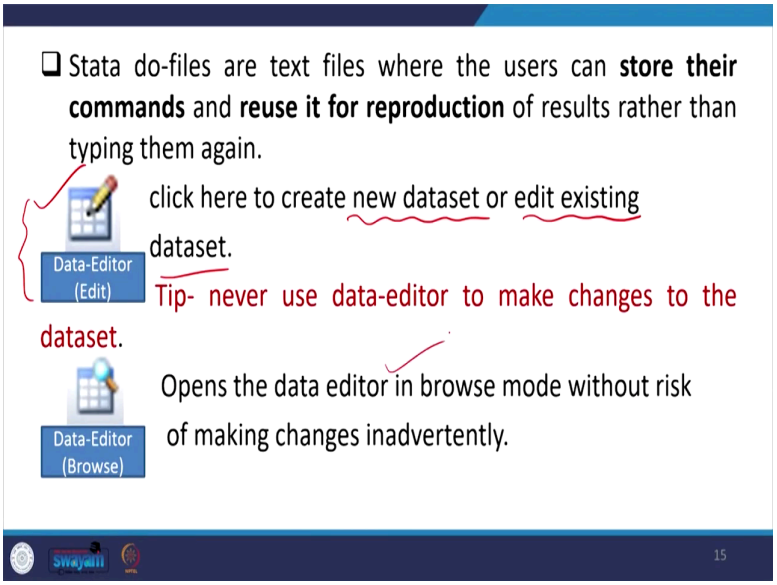
Then the next one is the do-file editor, I will also tell you it is located here. I am keeping my cursor over it.

(Refer Slide Time: 23:23)



Once I click here then the do-file editor is open. Now, what does it stands for? It stands for writing all your commands in this particular place. It is like syntax, and you can save this for later purposes. That is one of the biggest advantages of Stata.

(Refer Slide Time: 23:49)

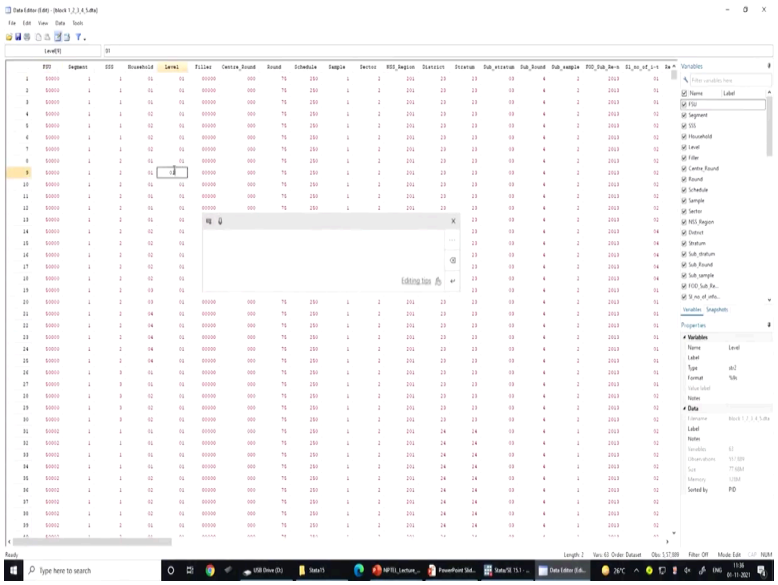


Stata do-files are text files where the users can store their commands and reuse them for the reproduction of results rather than typing them again. There are huge command pages, and you may not even remember them.

If you have forgotten to save commands, then you have to retry all those commands and there is a chance of entering an error. So, your do file is actually correcting those things that are in fact one of the biggest advantages of Stata. So, this is the icon which I have already told you and this helps in creating new data sets or editing existing data sets.

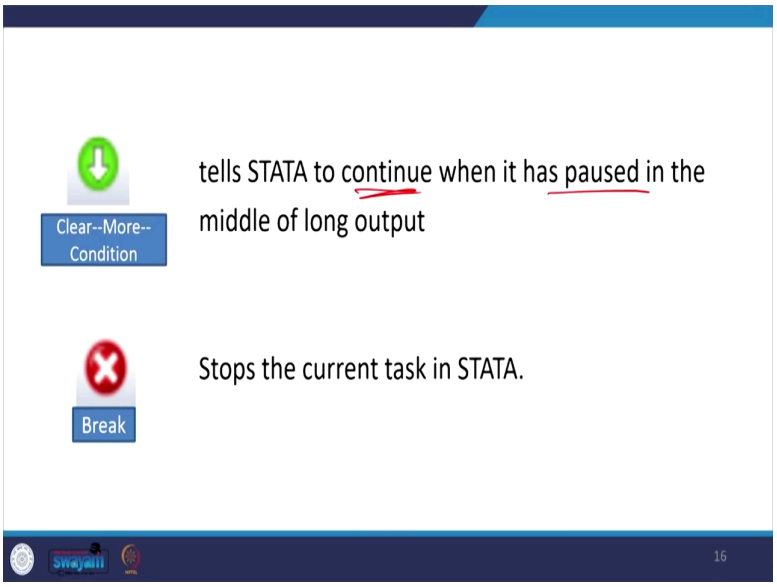
A data editor is useful for creating new data set or edit the existing data set. So, one tip is here you should not use the data editor to make changes to the data set. If you open the data editor there might be some possibility of committing mistakes which cannot be checked later on. A browse data set is useful to see all your data it is often used by the researcher in the Stata window.

(Refer Slide Time: 25:52)



So, we suggest that you please open data browse instead of data edit. This is the one screen opened in front of you is data browse. Anything you cannot be able to change on your own. Whereas in the case of data edit you can type and change it. So, once you are changing then it is very difficult to come back to the original data. You are supposed to correct your data set by the data editor.

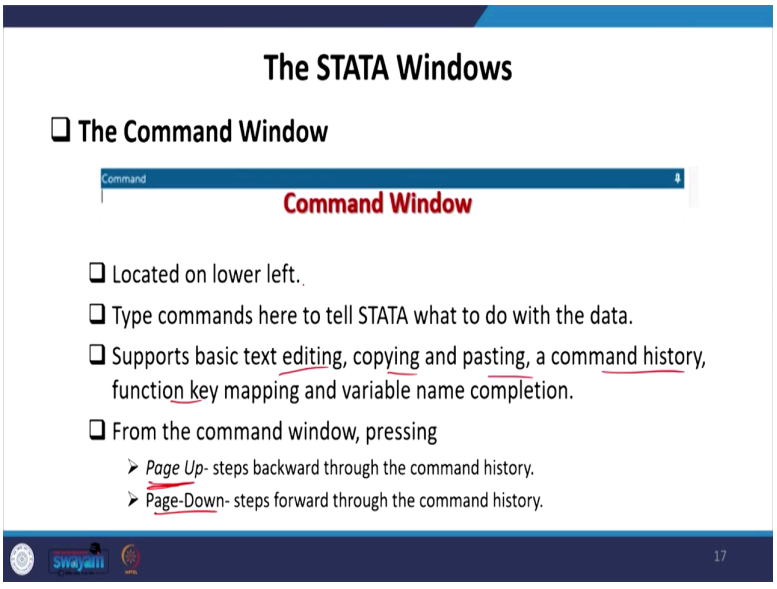
(Refer Slide Time: 26:35)



Then next is clear more condition, this option basically tells Stata to continue when it has paused in the middle of long output. You have given a command; it is kept on rolling and the final result is not coming then in that case you can click on this particular icon.

The next icon is break like you might have given a wrong command and you want to stop this command. So, this break is going to stop your command.

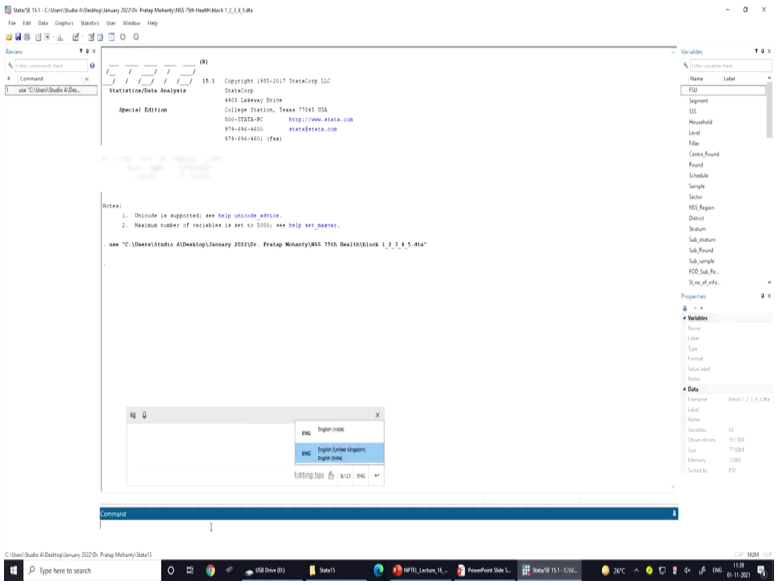
(Refer Slide Time: 27:31)



We are supposed to give the command or write commands like regression or tabulation. We can do it through the command window which is located usually at the bottom but drag it to your convenient position. This is located on the lower left. Type commands here to tell Stata what to do with the data. This supports basic text editing, copying, pasting, command history, functions key mapping and variable name completion.

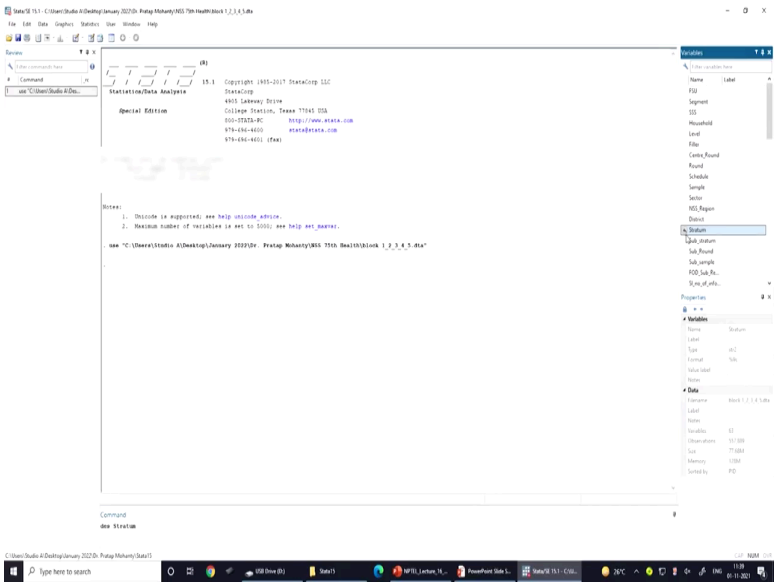
From the command window pressing page up steps backward through the command history. Page down steps forward through the command history.

(Refer Slide Time: 29:00)



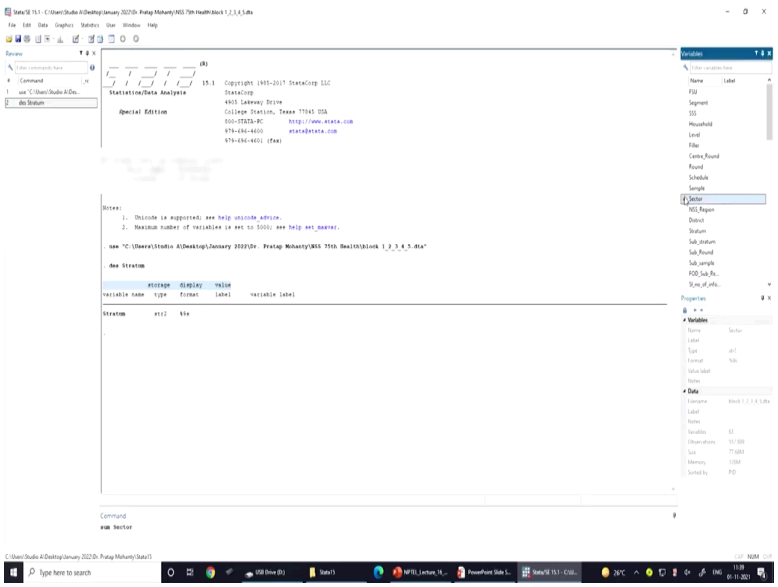
Suppose I will type tab.

(Refer Slide Time: 29:27)



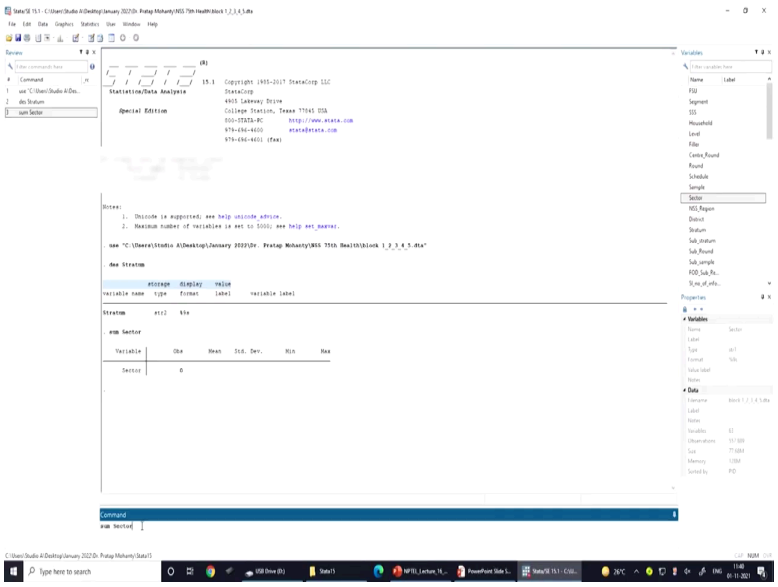
Or describe one variable here.

(Refer Slide Time: 29:38)



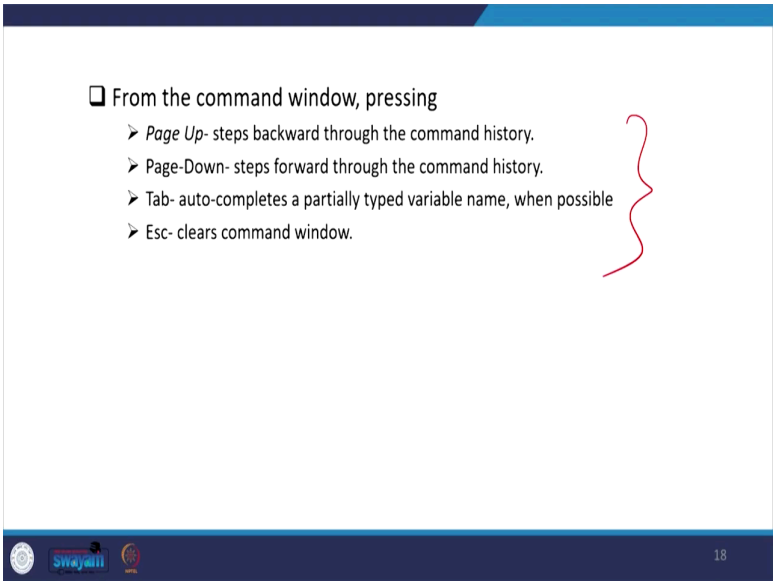
So, this describes the variable. Again, suppose I type ‘sum variable_name’. The variable name let it be a sector.

(Refer Slide Time: 29:56)



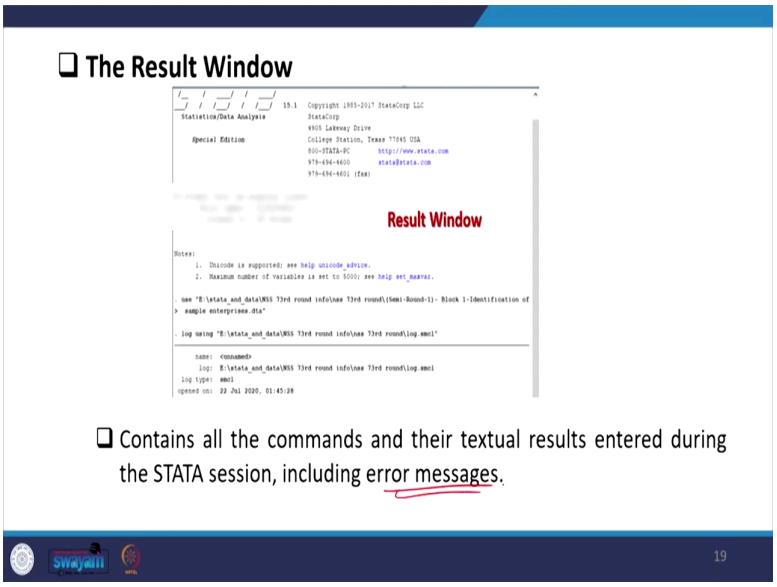
Now I wanted to show you page up and page down. If I click on the page up it will show me the previous command. Now the first click on the page up command gives us the first command which we have entered.

(Refer Slide Time: 30:57)



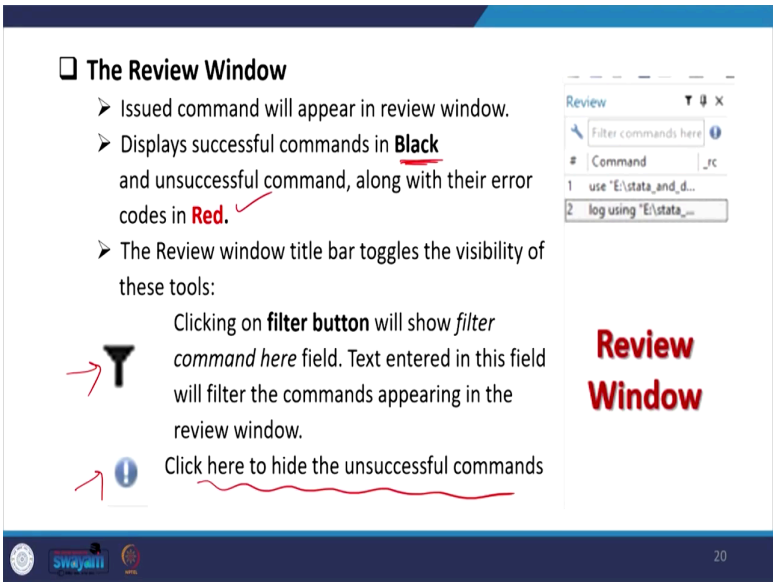
So, these are important you can just experiment on your own, we will be often using these shortcut keys for our Stata operation.

(Refer to Slide Time: 31:20)



The Result window contains all the commands and their textual results entered during the Stata session, including error messages as well.

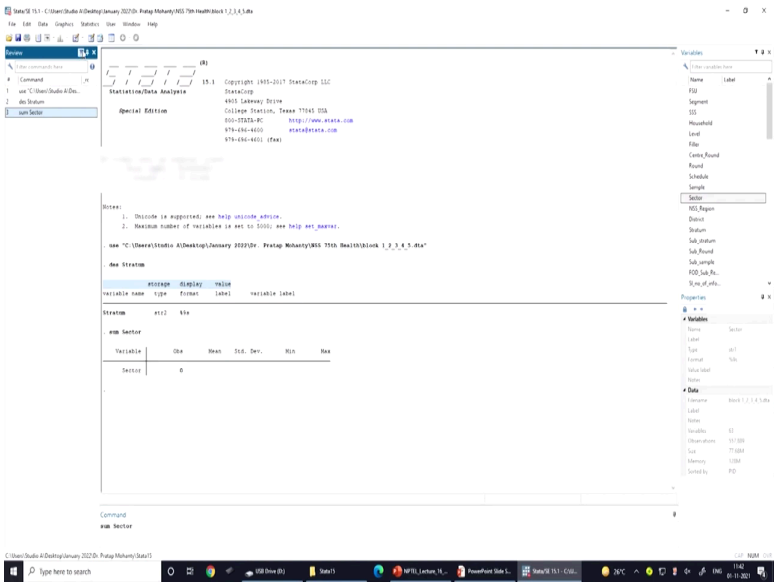
(Refer Slide Time: 31:36)



The review window usually keeps it on the left-hand side. So, this gives all the commands we have already entered, or this will be shown it here. If your command was incorrect it will show in red color that is why we have highlighted it. This displays successful commands in black, if the command was unsuccessful, it was not a correct command, then it will show as red.

The review command title bar toggles the visibility of these tools and clicking on the filter button will show the filter.

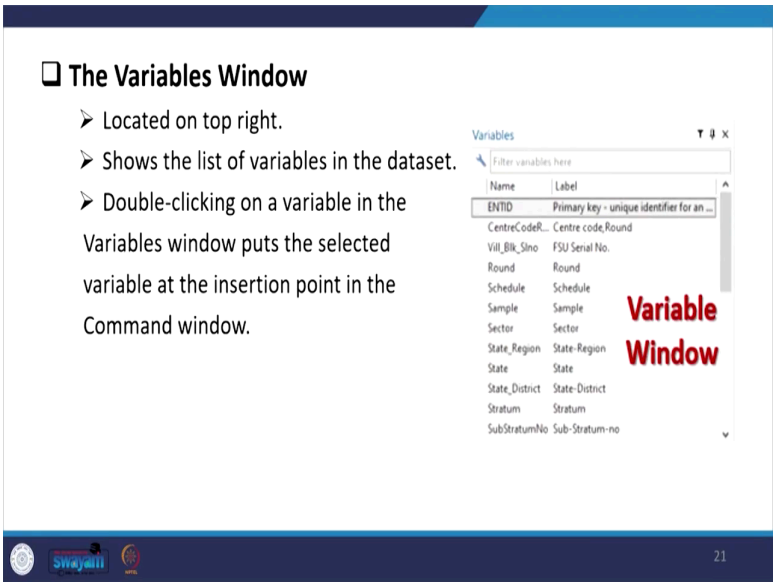
(Refer Slide Time: 32:38)



If you click on the filter will show the filter command here field. If you are just simply copying all those together all your wrong commands are also going to be copied. So, before taking copying all your commands on the review window you are supposed to filter the command.

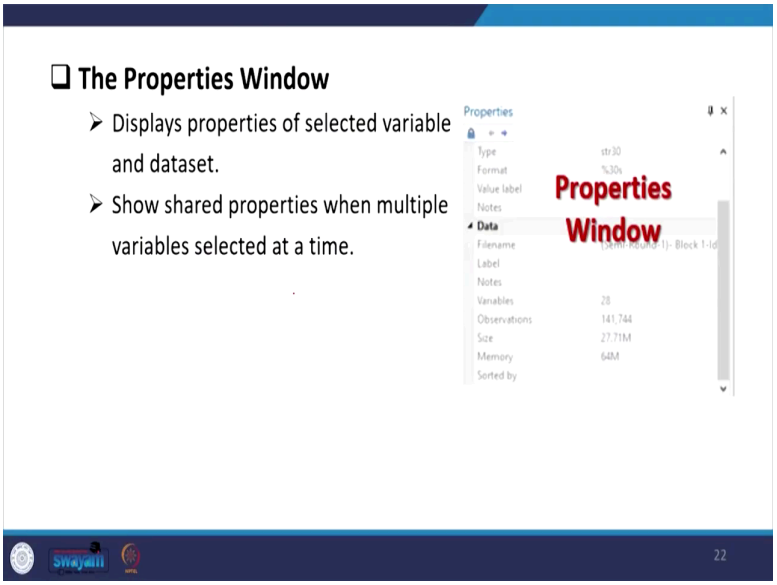
The filter button will show the command correctly. So, click here to hide the unsuccessful command, once you click here all those wrong commands will be closed then only the filter one will be left.

(Refer Slide Time: 33:42)



The variables window is basically located on the top right, this shows the list of variables in the data set. Double-clicking on a variable in the variable window puts the selected variable at the insertion point in the command window.

(Refer Slide Time: 34:12)



The properties window displays the properties of the selected variable and dataset. I have already shown it to you earlier. This shows properties here on the right-hand side and shows observations, size, and variable type with its label. Then this shows shared properties when multiple variables are selected at a time.

So, these are all details I think we need not explain to you much on this. Rest of the clarification, I am sure you will get and understand very clearly once you operate Stata on your own. There might be initial errors, but do not worry at all. You have to take the patience and work out through the Stata. This will make you more handy with Stata operations. This week onwards we will be focusing on Stata operations, and this is going to be quite useful throughout your work. With this, I think I should stop here and look forward to your participation.

Thank you.