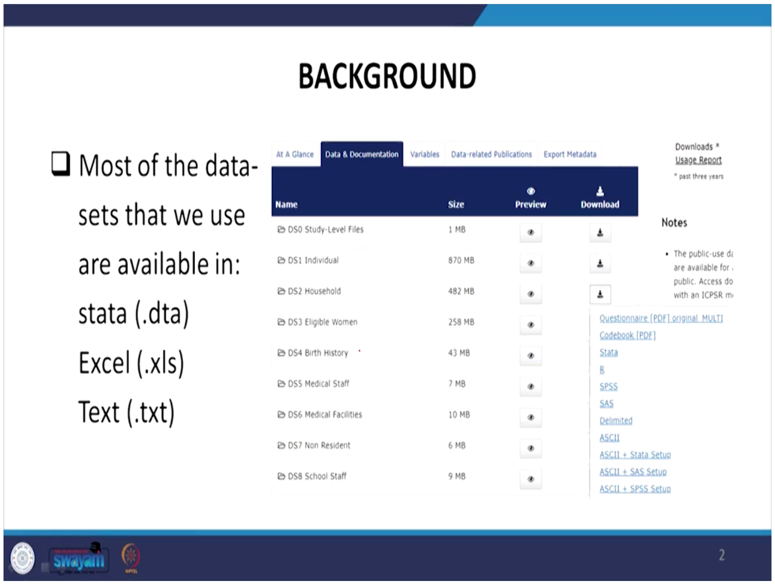
Handling Large-Scale Unit Level Data Using STATA
Professor Pratap C. Mohanty
Department of Humanities and Social Sciences,
Indian Institute of Technology Roorkee
Lecture 12
Exploring Data in Stata

Welcome once again friends to the NPTEL module on Handling Large-Scale Data Using Stata. We are in the 12th lecture of this particular module, where we are trying to explain the use of Stata with some demo data. And in the last class we have already explained the very basic background or understanding of Stata. Just by saying understanding is not enough, we need to explore some operation with the help of Stata.

(Refer Slide Time: 01:27)



BACKGROUND

- ❑ Most of the data-sets that we use are available in:
  stata (.dta)
  Excel (.xls)
  Text (.txt)

At A Glance | Data & Documentation | Variables | Data-related Publications | Export Metadata

Name	Size	Preview	Download
DS0 Study-Level Files	1 MB	👁	⬇
DS1 Individual	870 MB	👁	⬇
DS2 Household	482 MB	👁	⬇
DS3 Eligible Women	258 MB	👁	
DS4 Birth History	43 MB	👁	
DS5 Medical Staff	7 MB	👁	
DS6 Medical Facilities	10 MB	👁	
DS7 Non Resident	6 MB	👁	
DS8 School Staff	9 MB	👁	

Downloads *
Usage Report
* past three years

Notes

- The public-use da
  are available for
  public. Access do
  with an ICPSR m

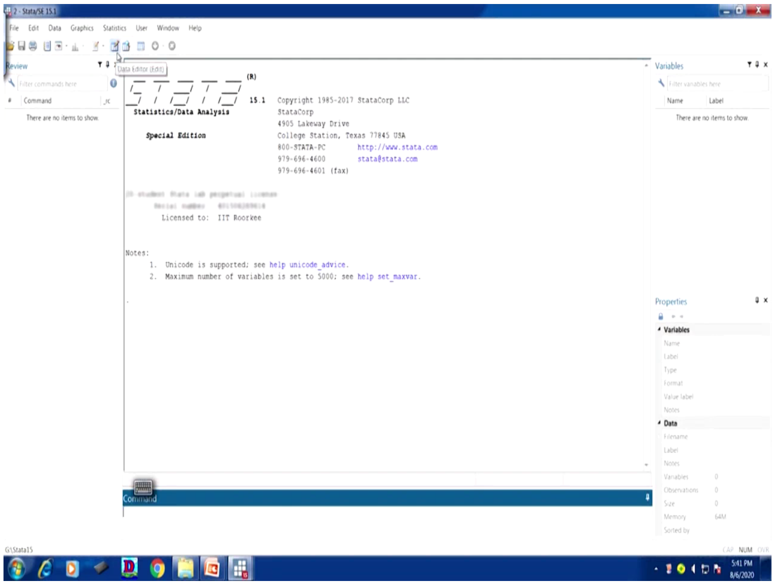
Questionnaire [PDF] original MULTI
Codebook [PDF]
Stata
R
SPSS
SAS
Delimited
ASCII
ASCII + Stata Setup
ASCII + SAS Setup
ASCII + SPSS Setup

2

Let me quickly proceed to exploring the Stata operations. Let us have a database like this. We have already shown you this kind of database earlier, in our earlier lectures while we explain different datasets. The datasets we are referring is from IHDS, India's Human Development Survey which considers different format of dataset. Like you can extract the data or you can operate through data, you can operate through excel, you can operate through SPSS, you can operate through SAS version. There are different versions given at the right hand side.

Since we require Stata version, we can download the Stata version. And generally Stata version data comes with .dta file. And excel version are also there. You can also convert into Stata. But it requires little more procedure. If there are text versions, there are also procedure for extraction. Since I am sticking to exploring Stata with the .dta file, I am not spending much time here.

(Refer Slide Time: 02:31)



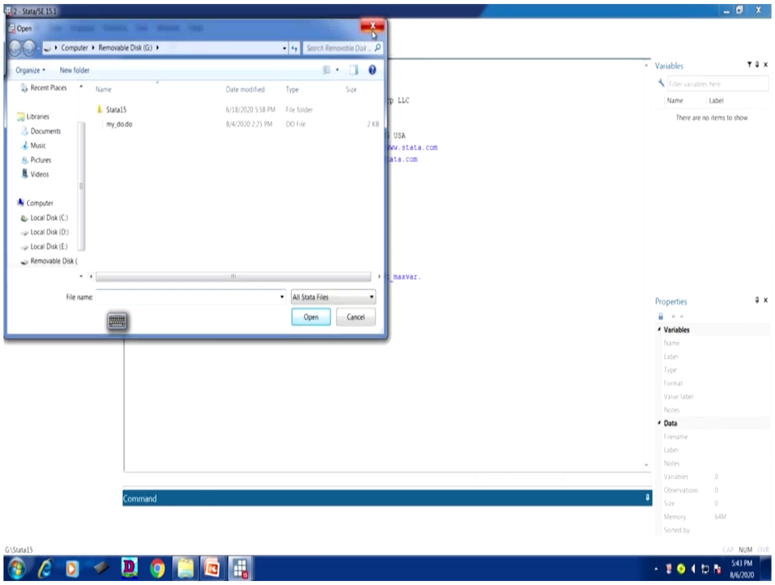
```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

[redacted]
       Licensed to:  IIT Roorkee

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

.
```

Let us open a data file in Stata. How to do it? As you remember that I suggested in the last class to open the dataset through browse window, not through edit window. I will just show you. Here there are two options. One is I am just keeping my cursor here. It shows data editor browse.

So it is always suggested to go for data browse, because browse once you have opened by any chance, it does not give change options. If you change it, then it will be very problematic for you to operate later what kind of change you have done it may be very difficult to track. So better to go by data browse if you just want to see the data. There are many other options also in the data editor we will explore later.

Once again I am going to open a dataset for you. How to do that? If you have a dataset already loaded, which I have already just shown to you IHDS data. We now operate from IHDS at this moment. Because it is a big data, we need to again take a sample data of it for operation. If you are continuing with a big data, generally it consumes byte space. Also while browsing the data, it consumes time. So it is always better for your learning, to go by a sample data or a demo data.

(Refer Slide Time: 04:10)

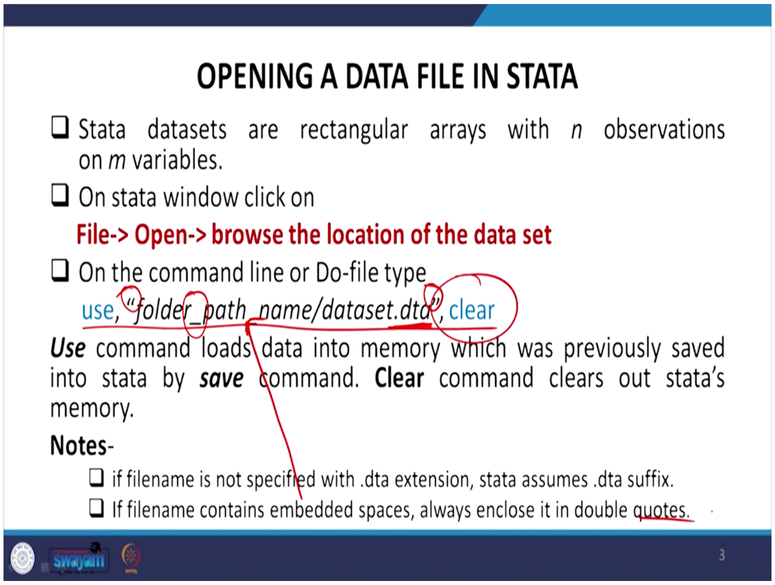


Name	Date modified	Type	Size
Stata15	6/18/2020 5:58 PM	File folder	
my_do.do	8/4/2020 2:25 PM	DO File	2 KB

So to open a data, the simplest way is just click this folder. this is coming, otherwise you go to file and you can get it, otherwise this folder option is there. If you have a loaded data already, you saved it from IHDS or from NFHS or from NSS it will come up with another entry in this particular folder with .dta extension file. And if you double click on it or just click and open, you can open the data. So I am not going to open that data because it is a lengthy one and it will be very difficult to understand the operation of Stata.

So, I will start with the same dataset we have already used from the Stata directory that is through system use, sysuse space directory. Otherwise, if you know the exact directory data that is called lifeexp or auto, we have already used we can open that.

(Refer Slide Time: 05:29)



OPENING A DATA FILE IN STATA

- ❑ Stata datasets are rectangular arrays with *n* observations on *m* variables.
- ❑ On stata window click on

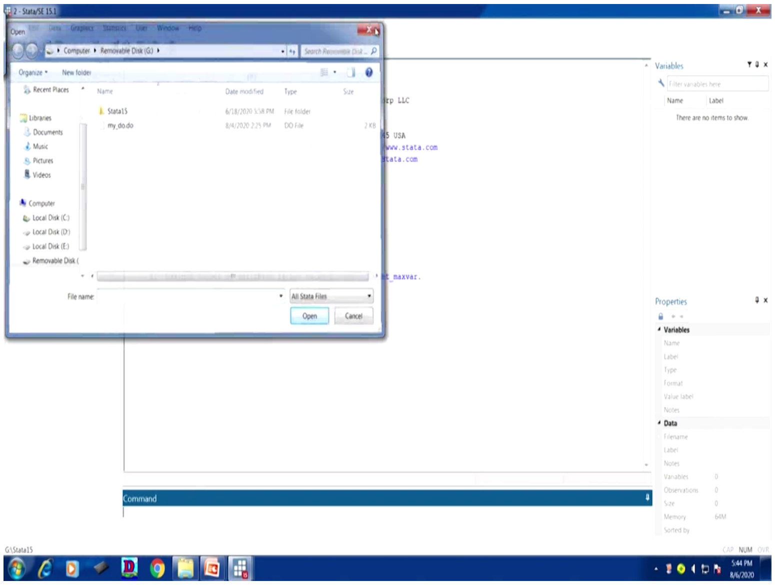
  File-> Open-> browse the location of the data set
- ❑ On the command line or Do-file type

  use, *"folder_path_name/dataset.dta"*, clear

Use command loads data into memory which was previously saved into stata by ***save*** command. **Clear** command clears out stata's memory.

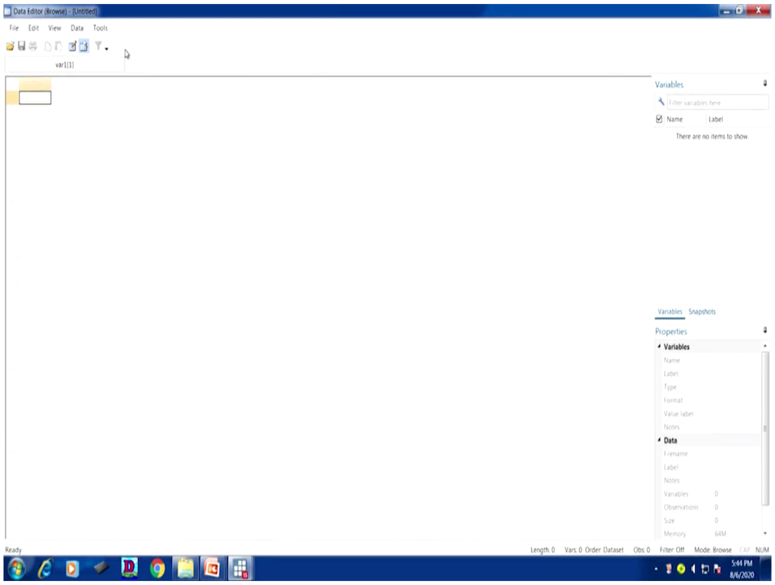
Notes-

- ❑ if filename is not specified with .dta extension, stata assumes .dta suffix.
- ❑ If filename contains embedded spaces, always enclose it in double quotes.



Name	Date modified	Type	Size
Stata15	6/18/2020 5:58 PM	File folder	
my_do.do	8/4/2020 2:25 PM	DO File	2 KB

```
rp LLC
5 USA
www.stata.com
tata.com
t_maxvar.
```



Let me first explain you what is all about in the slide and how you should remember those sequencing of entries in the slides for better operation. you go to the file. I have shown you the easiest option of clicking folder. Sometimes you may not get the folder option. You simply go to the file and click open. And open is there. There is a folder just open it, will give you the window for opening the data.

Let me move into the Stata browsing. Like as I told you, once again, I am operating, if I just browse the data. Since we have not yet loaded any dataset here, it is not giving you the dataset. If it is loaded, I can load it and explain you. let us look at, it is like an excel page. It is the data page of Stata. It has a rectangular n by m, rows and columns dataset. So, this is what we are trying to explain in the slide that the datasets are rectangular arrays with n observation. Observations are on the rows and on the column entries are defined as variables. So, we are writing it n into m variables.

Coming to the opening of data, we have already mentioned, otherwise you simply type use. When you know the path name, folder name, where to save or where to start the data and you know the location of the data you have already downloaded and that too it is in .dta format. When you know that, you simply type use, this is the command, use, then folder name with double inverted commas and path space is there, usually Stata does not read the space. You need to use something in between entries like you, it is better to enter underscore. It will make the path of that file name continuous.

So, once you do that, at the end better to always suggested to enter this as clear, because earlier if by any chance any data has already been opened and you are again opening these there might be some confusion with the variables, so better to write clear. It will clear earlier additions. Now you will work with fresh data. There will be no question of repetition.

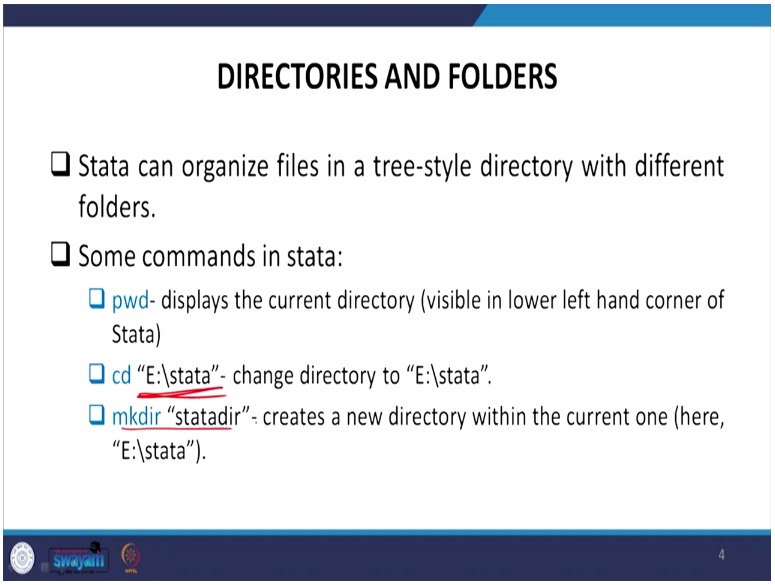
So, I can use the same one, we can do use comma file path name. Again, file path name you can go just by clicking here I can tell you, suppose your file is here by any chance if it is there, if it is in the D file, you just right click here copy address. Click on the copy address, it will give you the path name. use comma you, it will be like use comma double inverted with paste on the file path name you have already copied then you close the path name like since the file is already in .dta name we need not type .dta again, .dta is already there.

If it is a Stata version, you already carefully downloaded the Stata version. So it will come with, automatically come up with .dta file. Do remember that you start with a double inverted comma, you ended with the double inverted comma, then comma clear. It will open the data. The easiest way to click on that folder and double click on that particular file it will show you the data.

If the file name is not specified with .dta extension, Stata assumes .dta suffix. But subject to, like for example if it is an excel file, so obviously it will not be in .dta. Stata will automatically assume is .dta file, because these files can be converted to Stata easily. Excel file is generally convertible to Stata. But it will automatically assumes that you have a .dta file if it is not there.

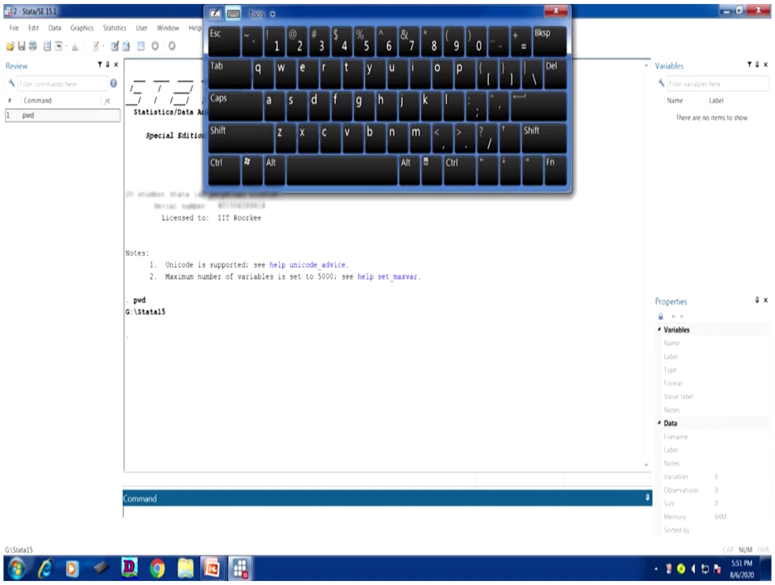
Now issue comes when you have embedded spaces. We have already said embedded space like here. If there are embedded spaces, so it is always better to enclose with double inverted commas. So, double quote is written for that only. Only then it reads the exact path of that data.

(Refer Slide Time: 10:54)



DIRECTORIES AND FOLDERS

- ❑ Stata can organize files in a tree-style directory with different folders.
- ❑ Some commands in stata:
  - ❑ pwd- displays the current directory (visible in lower left hand corner of Stata)
  - ❑ cd “E:\stata”- change directory to “E:\stata”.
  - ❑ mkdir “statadir”- creates a new directory within the current one (here, “E:\stata”).



```
Statistics/Data A

  Special Editio

[redacted]
        Licensed to:  IIT Roorkee

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

. pwd
G:\Stata15

.
```

So, Stata recognize files in a tree-style directory with different folders. Because tree-style basically has a clear sequencing of the folder then sub folders. we will talk about that in a short while. here some commands in Stata are important to read the directories and their folders. If I open the Stata, the directory since we do not have, the directory by default at left hand corner you can see it G, like here pwd since we do not have data, if I just type pwd here, pwd enter, it will show you where your directory is defined. Directory of the Stata is defined.

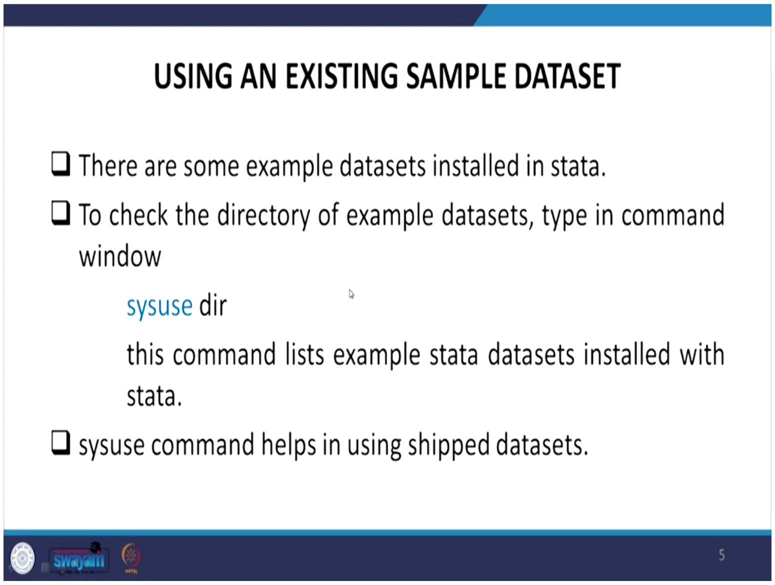
Maybe it is in C drive, but for us it is in G drive. Since our Stata is in G drive, so it is showing G drive. But by default when it is installed, it takes to the C drive. If you do not have enough space in C drive, you can change it to other drive. What you will do, in that case, you have to type cd change directory. If it is in C drive, then cd double quote C colon slash Stata is already there, then end your double quote. Basically since it is already there. We have already checked that it is there in G, it will always show that it is in G file.

You wanted to change it to, suppose you want to change to D, you have to enter D instead of, I was telling you for you as C, if already there in C, you wanted to change it to D, you have to type D. cd double quote D slash double quote D colon slash Stata and double quote ends, enter, it will automatically save it to D file. it depends upon how you are doing it. And we have already written the command here for your use. You can use this particular command.

Let me move, if you have any sub-folder. For example, you are working with some file called wealth, the total national account estimates, you have some data. Within that, you have some particular state GDP files. So you want to make another sub-folder. You do not want to operate from that GDP or the entire GDP file, you want it to operate in a small folder. You have certain small explanation required.

So you need to change the directory, like you wanted to create a new directory. Change directory already said, sub-directory, you can say a sub-directory within the bigger folder, the main folder called E here. For us, E slash Stata E colon. We wanted to get within E another file. For that you need to enter the command mkdir. Then with the same type of format you know the Stata directory, if I add a new name to it, it will create a sub-folder with the name statadir.

(Refer Slide Time: 14:52)

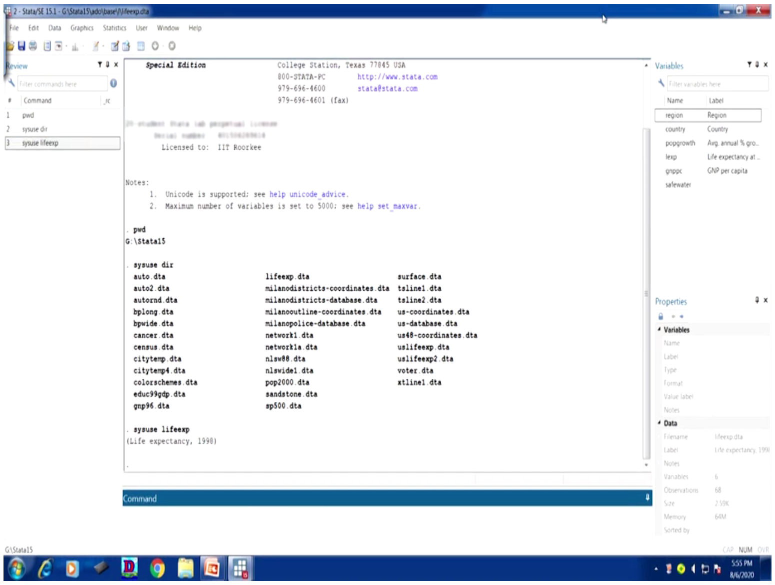


USING AN EXISTING SAMPLE DATASET

- ❑ There are some example datasets installed in stata.
- ❑ To check the directory of example datasets, type in command window

  sysuse dir

  this command lists example stata datasets installed with stata.
- ❑ sysuse command helps in using shipped datasets.

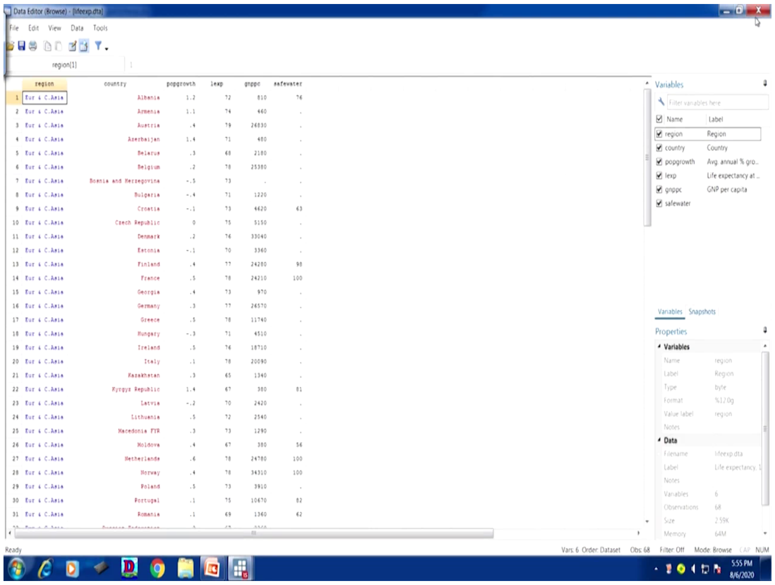


#	Command	_rc
1	pwd	
2	sysuse dir	
3	sysuse lifeexp	

```
     Special Edition                College Station, Texas 77845 USA
                                    800-STATA-PC        http://www.stata.com
                                    979-696-4600        stata@stata.com
                                    979-696-4601 (fax)

[redacted]
         Licensed to:  IIT Roorkee


Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.

. pwd
G:\Stata15

. sysuse dir
  auto.dta                    lifeexp.dta                     surface.dta
  auto2.dta                   milanodistricts-coordinates.dta tsline1.dta
  autornd.dta                 milanodistricts-database.dta    tsline2.dta
  bplong.dta                  milanooutline-coordinates.dta   us-coordinates.dta
  bpwide.dta                  milanopolice-database.dta       us-database.dta
  cancer.dta                  network1.dta                    us48-coordinates.dta
  census.dta                  network1a.dta                   uslifeexp.dta
  citytemp.dta                nlsw88.dta                      uslifeexp2.dta
  citytemp4.dta               nlswide1.dta                    voter.dta
  colorschemes.dta            pop2000.dta                     xtline1.dta
  educ99gdp.dta               sandstone.dta
  gnp96.dta                   sp500.dta

. sysuse lifeexp
(Life expectancy, 1998)

.
```

Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Filename	lifeexp.dta
Label	Life expectancy, 1998
Notes	
Variables	6
Observations	68
Size	2.59K
Memory	64M
Sorted by	



1

	region	country	popgrowth	lexp	gnppc	safewater
1	Eur & C.Asia	Albania	1.2	72	810	76
2	Eur & C.Asia	Armenia	1.1	74	460	.
3	Eur & C.Asia	Austria	.4	79	26830	.
4	Eur & C.Asia	Azerbaijan	1.4	71	480	.
5	Eur & C.Asia	Belarus	.3	68	2180	.
6	Eur & C.Asia	Belgium	.2	78	25380	.
7	Eur & C.Asia	Bosnia and Herzegovina	-.5	73	.	.
8	Eur & C.Asia	Bulgaria	-.4	71	1220	.
9	Eur & C.Asia	Croatia	-.1	73	4620	63
10	Eur & C.Asia	Czech Republic	0	75	5150	.
11	Eur & C.Asia	Denmark	.2	76	33040	.
12	Eur & C.Asia	Estonia	-.1	70	3360	.
13	Eur & C.Asia	Finland	.4	77	24280	98
14	Eur & C.Asia	France	.5	78	24210	100
15	Eur & C.Asia	Georgia	.4	73	970	.
16	Eur & C.Asia	Germany	.3	77	26570	.
17	Eur & C.Asia	Greece	.5	78	11740	.
18	Eur & C.Asia	Hungary	-.3	71	4510	.
19	Eur & C.Asia	Ireland	.5	76	18710	.
20	Eur & C.Asia	Italy	.1	78	20090	.
21	Eur & C.Asia	Kazakhstan	.3	65	1340	.
22	Eur & C.Asia	Kyrgyz Republic	1.4	67	380	81
23	Eur & C.Asia	Latvia	-.2	70	2420	.
24	Eur & C.Asia	Lithuania	.5	72	2540	.
25	Eur & C.Asia	Macedonia FYR	.3	73	1290	.
26	Eur & C.Asia	Moldova	.4	67	380	56
27	Eur & C.Asia	Netherlands	.6	78	24780	100
28	Eur & C.Asia	Norway	.4	78	34310	100
29	Eur & C.Asia	Poland	.5	73	3910	.
30	Eur & C.Asia	Portugal	.1	75	10670	82
31	Eur & C.Asia	Romania	.1	69	1360	62

☑ Name	Label
☑ region	Region
☑ country	Country
☑ popgrowth	Avg. annual % gro...
☑ lexp	Life expectancy at ...
☑ gnppc	GNP per capita
☑ safewater	

Variables	
Name	region
Label	Region
Type	byte
Format	%12.0g
Value label	region
Notes	
Data	
Filename	lifeexp.dta
Label	Life expectancy, 1
Notes	
Variables	6
Observations	68
Size	2.59K
Memory	64M

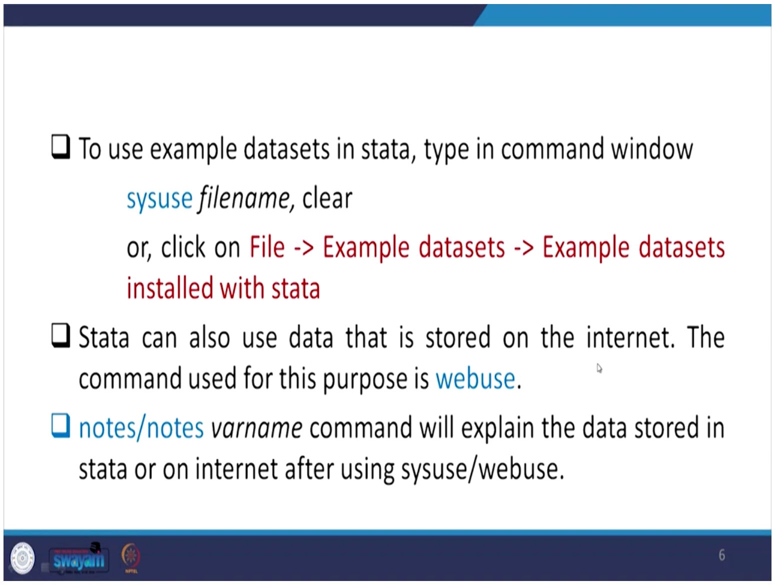
Vars: 6 Order: Dataset Obs: 68 Filter: Off Mode: Browse CAP NUM

This is going to be very useful, where we are going to discuss that how to operate some existing data. I have already shown to you last time. So, what I will do, I will open once again this and start opening a new database with the sysuse command. Let me just tell you, if I just type dir, it will give you the directory, all the dataset that I told you in the last lecture.

When you know those name, lifeexp we are using. You can use auto also; it will take another data file also. network one if you use, it will open the network data. I am interested because some of the important files, some of the important variations are there in lifeexp data. So I will be using lifeexp. So I use again sysuse space lifeexp enter. It has already opened. You can find out from our right hand side with the variable names.

You just wanted to see that what kind of data and how it looks, you can also open the data browser. Data browser will give you the entries of those data. I will tell you in a short while why different colors are there I will explain you. Let me move into the section we are trying to explain. So, we already discussed sysuse command and this is going to be very useful.

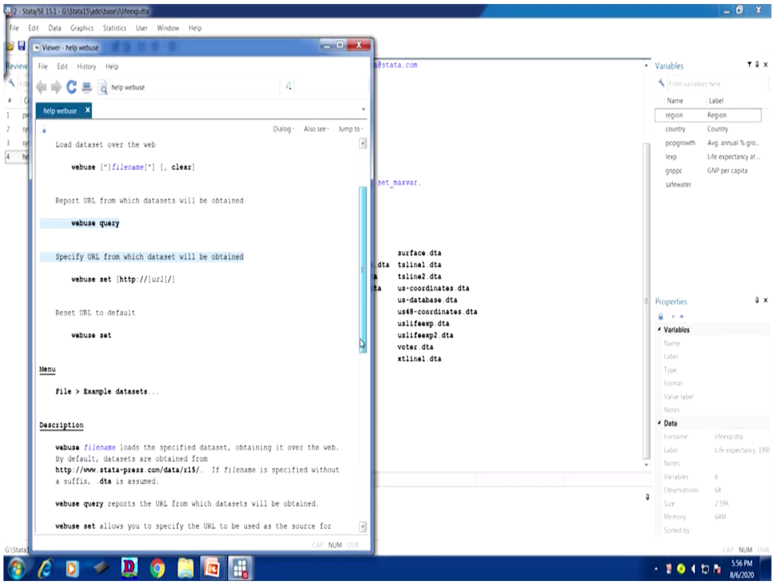
(Refer Slide Time: 16:55)



- ❑ To use example datasets in stata, type in command window

  sysuse *filename*, clear

  or, click on File -> Example datasets -> Example datasets installed with stata
- ❑ Stata can also use data that is stored on the internet. The command used for this purpose is webuse.
- ❑ notes/notes *varname* command will explain the data stored in stata or on internet after using sysuse/webuse.



Stata/SE 15.1 - G:\Stata15\ado\base\l\lifeexp.dta
File Edit Data Graphics Statistics User Window Help
Viewer - help webuse
File Edit History Help
help webuse
Dialog
Also see
Jump to

```
    Load dataset over the web

        webuse ["]filename["] [, clear]


    Report URL from which datasets will be obtained

        webuse query


    Specify URL from which dataset will be obtained

        webuse set [http://]url[/]


    Reset URL to default

        webuse set


Menu

    File > Example datasets...


Description

    webuse filename loads the specified dataset, obtaining it over the web.
    By default, datasets are obtained from
    http://www.stata-press.com/data/r15/.  If filename is specified without
    a suffix, .dta is assumed.

    webuse query reports the URL from which datasets will be obtained.

    webuse set allows you to specify the URL to be used as the source for
```

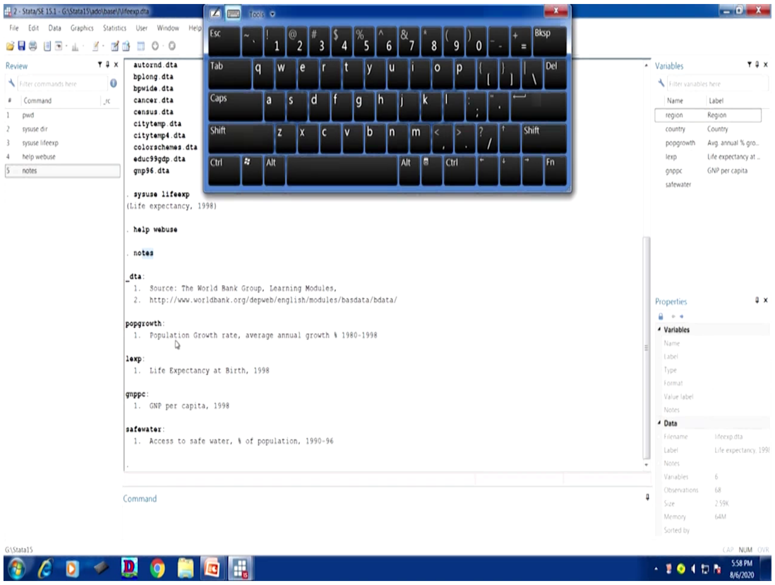
```
surface.dta
tsline1.dta
tsline2.dta
us-coordinates.dta
us-database.dta
us48-coordinates.dta
uslifeexp.dta
uslifeexp2.dta
voter.dta
xtline1.dta
```

Variables

Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

Properties

Filename	lifeexp.dta
Label	Life expectancy, 1998
Variables	6
Observations	68
Size	2.59K
Memory	64M



```
autornd.dta
bplong.dta
bpwide.dta
cancer.dta
census.dta
citytemp.dta
citytemp4.dta
colorschemes.dta
educ99gdp.dta
gnp96.dta

. sysuse lifeexp
(Life expectancy, 1998)

. help webuse

. notes

_dta:
  1.  Source: The World Bank Group, Learning Modules,
  2.  http://www.worldbank.org/depweb/english/modules/basdata/bdata/

popgrowth:
  1.  Population Growth rate, average annual growth % 1980-1998

lexp:
  1.  Life Expectancy at Birth, 1998

gnppc:
  1.  GNP per capita, 1998

safewater:
  1.  Access to safe water, % of population, 1990-96
```

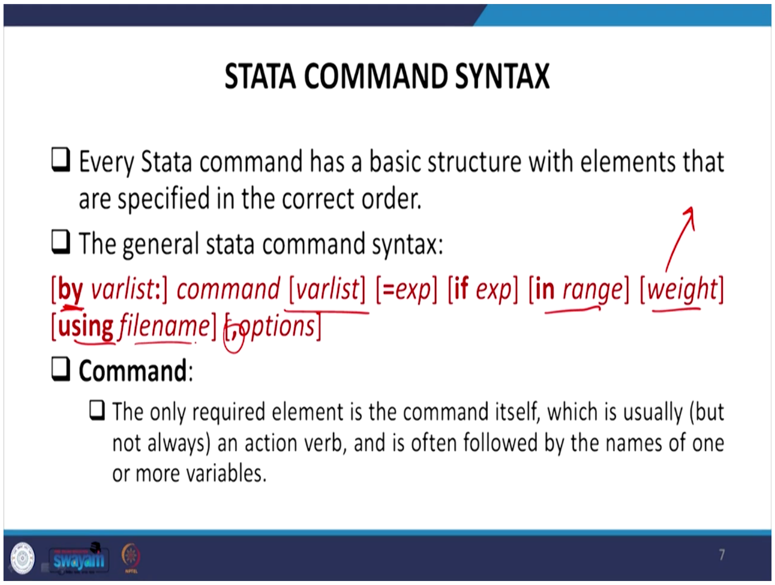
Another aspect of importance for user, for example, these are Stata installed data. We have shown you Stata installed data. Suppose you do not know where is the directory and you want to operate from other data source, there are other options also. Stata gives many other options. So, you have to go for help and webuse. There are a number of datasets available in the help command. So, simply type help webuse, help space webuse. It will redirect you to many links there are so many, but view complete video.

Here these links are not visible because you should have the exact name of that particular file. It is better that you search in the Google, other dataset available Stata. It will give you many other datasets. Another important aspect to be noted, once you know the data you wanted to, since we have already opened a dataset called lifeexp of the Stata data, we wanted to understand the data.

Simply type, you want it to, what kind of data it is, simply type notes, it will give you all the variables like from the notes, what kind of variables are there. Here region, country, population growth, then life expectancy, then GNP per capita, safe water all those information are there.

What kind of information it gives, population growth rate, annual, under population growth variable, population growth rate, annual growth rate by percentage from the year 1980 to 1998. Similarly, other information you can find out on your own comfortable approach.

(Refer Slide Time: 19:36)



STATA COMMAND SYNTAX

- ❑ Every Stata command has a basic structure with elements that are specified in the correct order.
- ❑ The general stata command syntax:

[**by** *varlist*:] *command* [*varlist*] [=*exp*] [**if** *exp*] [**in** *range*] [*weight*] [**using** *filename*] [,*options*]

- ❑ **Command**:
  - ❑ The only required element is the command itself, which is usually (but not always) an action verb, and is often followed by the names of one or more variables.

7

So, let me proceed for other set of information which is very important for the users. Let us come to Stata command syntax. Usually, Stata has a clear defined order of syntax. A clear defined command syntax is there with Stata. If you deviate that order Stata is not going to read your command. You have to follow this order. If you have so many indicators to search or to be operated with your data, then you have to follow this ordering. This the basic structure of the syntax.

What is there? for example, there are so many syntaxes. It is highlighted with the square bracket. So, square bracket only indicates what command is here, what kind of different indicators are there in the command. First is by command. By is there if you have a variable, you need to sort with that particular variable, for example, you wanted to locate from data you have life expectancy and you have GDP and you have already derived a cross tabulation or a descriptive statistic between these two variables. But GDP you have already categorized and life expectancy you have already categorized and it gives, suppose by three or four whichever indicator you have already categorized.

You wanted to sort it by another variable. For example, maybe, by region, sort by region. Your interest is to sort all the variables through region. Region is the indicator. you start with writing sort by region, then you enter the command, like tab, tab is the command I am going to explain you later, tab then variable list, two variables tab means basically a cross tabulation we are deriving, two-way table if you are trying to derive, then the two variables to be entered here. We will explain all those things later. till this we have already done it.

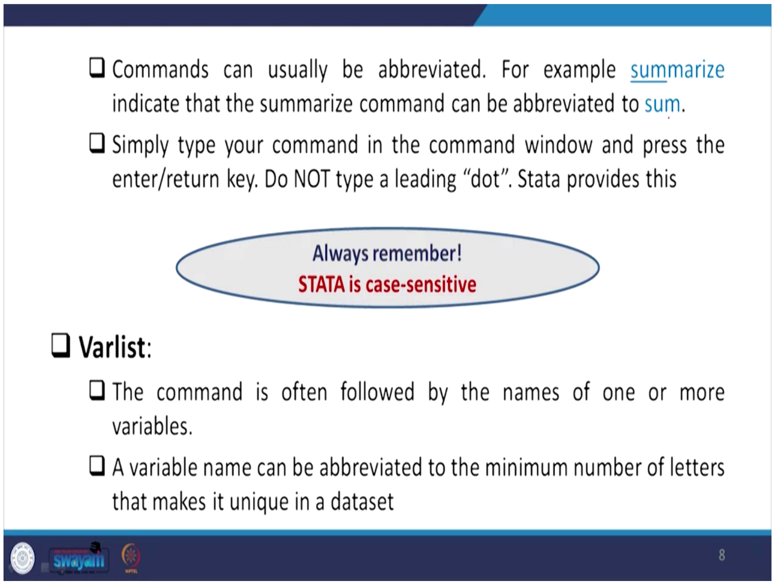
What is important, if you have certain variables which require some other manipulation, some extent of manipulation, like what manipulation you want to add certain variables and derive the total effect of it or if you want that your variables should only show you till certain limit, not exceeding that limit. For example, if you are working with the age category, you wanted to highlight the youth age and their indicators in any databases maybe in IHDS databases. So, you can use the less than expression, less than or equal to with that particular age. If you type, it will only show you that particular result not with entire age category.

Otherwise that you have used less than equal to command. If you have a range defined from this age to that age required. Weight, we will explain all those things while we will operate with range, operate with weight. Weight is technical at this moment I am not explaining in detail. Certain weightage is given to certain variables and weights basically are attached to a variable to represent that particular variable in the analysis. And some of the explanation we already made earlier, in our earlier lecture more clarity can be derived. How weight variables differentiate your result.

Similarly, after that using from the filename, from whose filename you wanted to operate. you are giving a command, but you do not have a data. Whose filename you wanted to use to derive the result? So filename, next comes with filename. Then comes option comma is there, comma must be entered, bracket should be avoided, bracket is not required except weight. Bracket, while entering the command brackets are not required or Stata does not read brackets. We are explaining this to highlight which sequences are there.

In the options generally we wanted to generate a new variable. After your expression like you want less than of this age. this should be defined as a new variable. So generate, gen a new variable. We will tell you later, but at this moment, I am just giving you a sample of those explanation.

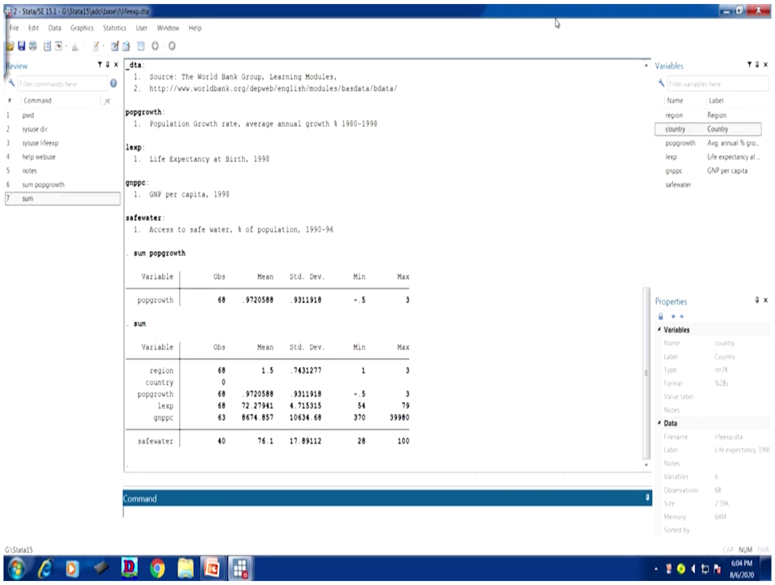
(Refer Slide Time: 24:32)



- ❑ Commands can usually be abbreviated. For example summarize indicate that the summarize command can be abbreviated to sum.
- ❑ Simply type your command in the command window and press the enter/return key. Do NOT type a leading "dot". Stata provides this

Always remember!
STATA is case-sensitive

- ❑ **Varlist:**
  - ❑ The command is often followed by the names of one or more variables.
  - ❑ A variable name can be abbreviated to the minimum number of letters that makes it unique in a dataset



#	Command	_rc
1	pwd	
2	sysuse dir	
3	sysuse lifeexp	
4	help webuse	
5	notes	
6	sum popgrowth	
7	sum	

```
_dta:
  1.  Source: The World Bank Group, Learning Modules,
  2.  http://www.worldbank.org/depweb/english/modules/basdata/bdata/

popgrowth:
  1.  Population Growth rate, average annual growth % 1980-1998

lexp:
  1.  Life Expectancy at Birth, 1998

gnppc:
  1.  GNP per capita, 1998

safewater:
  1.  Access to safe water, % of population, 1990-96

. sum popgrowth

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   popgrowth |         68    .9720588    .9311918        -.5          3

. sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      region |         68         1.5    .7431277          1          3
     country |          0
   popgrowth |         68    .9720588    .9311918        -.5          3
        lexp |         68    72.27941    4.715315         54         79
       gnppc |         63    8674.857    10634.68        370      39980
-------------+---------------------------------------------------------
   safewater |         40        76.1    17.89112         28        100

.
```

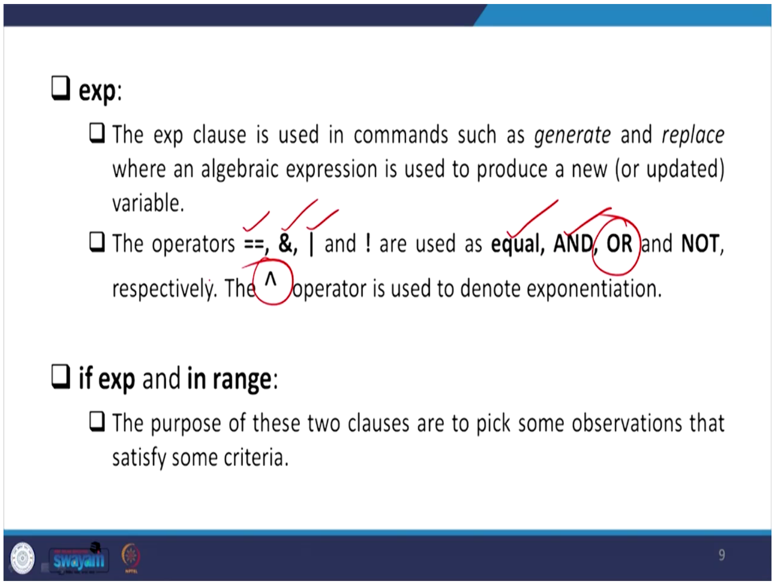
Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

Commands can usually be abbreviated. Suppose you want to summarize the data. Summarize give you sum of everything. For example, I will show you here like you want summarize, you can write summarize, but instead you can write sum. Sum, if you just click on the variable, what is going to give you, it will give you only one variable information. It gives how many observations are there, what is the mean of it, what is the standard deviation, what is the min, maximum, a standard summarize information. You can get detailed summarize also. There are other options, other commands. We will tell you during our operation later.

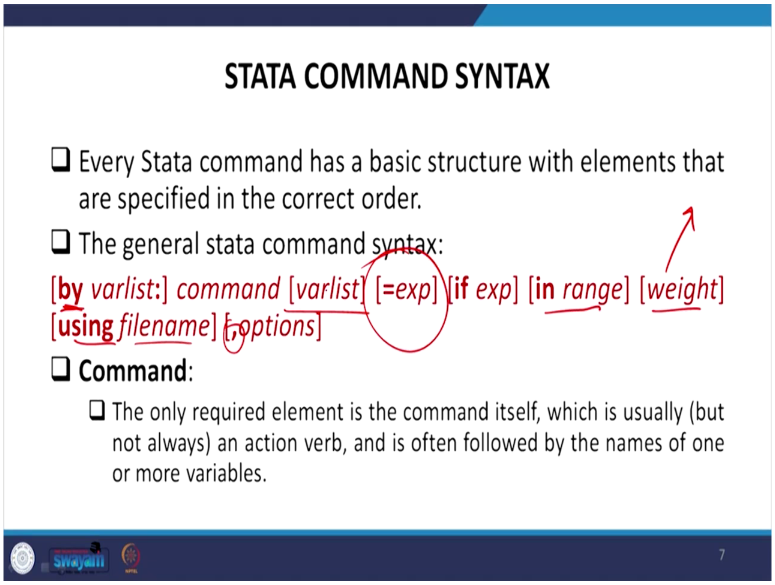
If you want to summarize all, then you have to click sum all the variables. It will give you only sum. It gives automatically sum, simply type sum, it will generate the detailed result. It is I think highlighted here. So, no need to type all the variables, only sum is going to give you the detailed result. But dot is very important. Sum ended with dot, dot should not be entered. So, since Stata is very case sensitive, do not type a leading dot after the command. And simply you type the command and enter. Press enter that will give you the result.

Similarly, the command is often followed by names of one or more variables. You can, if you have one variable, it will give you the information of one. If it is more variable than it will certainly give you more information. But for simplicity some commands automatically take entire variables like summary. You need not type everything. Simply sum, it will by default take all. If you want two, three information, then two, three information can be collected individually. A variable name can be abbreviated to a minimum number of letters, which I have already said instead of typing everything.

(Refer Slide Time: 27:10)



- ❑ **exp**:
  - ❑ The exp clause is used in commands such as *generate* and *replace* where an algebraic expression is used to produce a new (or updated) variable.
  - ❑ The operators **==**, **&**, | and **!** are used as **equal**, **AND**, **OR** and **NOT**, respectively. The ^ operator is used to denote exponentiation.
- ❑ **if exp** and **in range**:
  - ❑ The purpose of these two clauses are to pick some observations that satisfy some criteria.



STATA COMMAND SYNTAX

□ Every Stata command has a basic structure with elements that are specified in the correct order.

□ The general stata command syntax:

[**by** *varlist*:] *command* [*varlist*] [=*exp*] [**if** *exp*] [**in** *range*] [*weight*] [**using** *filename*] [,*options*]

□ **Command**:

□ The only required element is the command itself, which is usually (but not always) an action verb, and is often followed by the names of one or more variables.

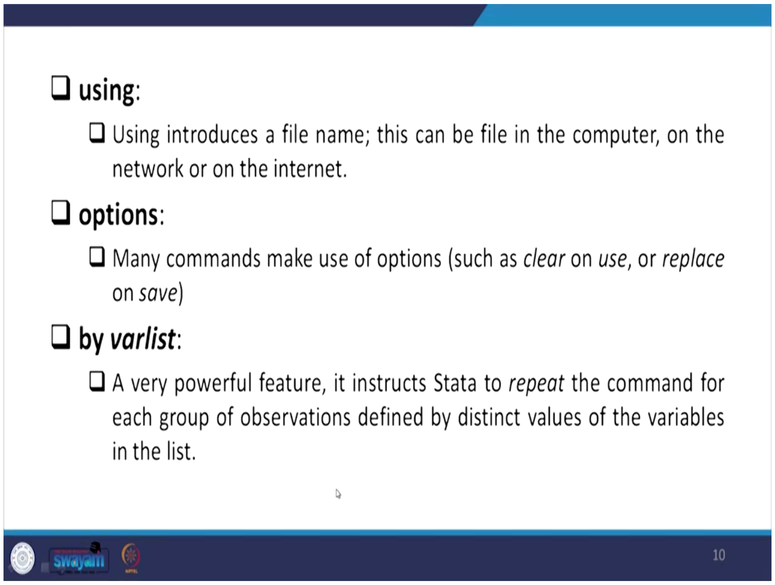
7

Similarly, there are some expression command here in the sequencing of our syntax, command syntax I have already said expression is here. So, as I already told you, exp stands for expression. I told you it might be within a range, it might be less than or equal to or maybe addition, may be power, there are different ways of operating that particular command.

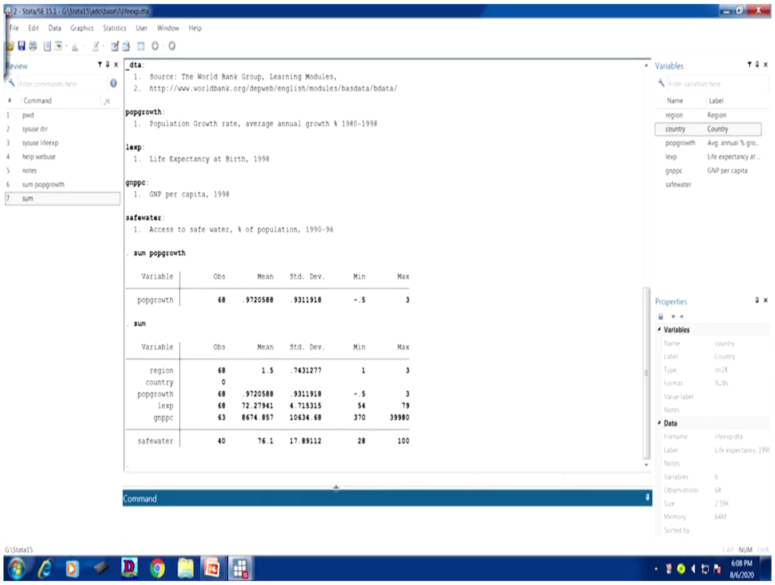
So, here it is written. You can use these. You can use double equal to & stands for if there are joint variables and their information you want so that can be taken here, these stands for, or these, use these variable or another variable or maybe also you or this range or another range if you wanted to enter or is important for you.

This exclamation mark is used for not. If that is not the variable you want, that is a particular figure you do not want, you just simply add exclamation mark. And if you have a power, like you want two to the power q, so you have to enter the power command. There is no such, in the keyboard there is no cube format entry, only power is there in the keyboard, so power command can be entered. So I have already explained this. I am not spending much time like expression with range I have already explained.

(Refer Slide Time: 28:39)



- ❑ **using**:
  - ❑ Using introduces a file name; this can be file in the computer, on the network or on the internet.
- ❑ **options**:
  - ❑ Many commands make use of options (such as *clear* on *use*, or *replace* on *save*)
- ❑ **by *varlist***:
  - ❑ A very powerful feature, it instructs Stata to *repeat* the command for each group of observations defined by distinct values of the variables in the list.



2 - Stata/SE 15.1 - G:\Stata15\ado\base\l\lifeexp.dta
File Edit Data Graphics Statistics User Window Help
Review
Filter commands here
Variables
Filter variables here
Properties
Command

#	Command
1	pwd
2	sysuse dir
3	sysuse lifeexp
4	help webuse
5	notes
6	sum popgrowth
7	sum

```
_dta:
  1.  Source: The World Bank Group, Learning Modules,
  2.  http://www.worldbank.org/depweb/english/modules/basdata/bdata/

popgrowth:
  1.  Population Growth rate, average annual growth % 1980-1998

lexp:
  1.  Life Expectancy at Birth, 1998

gnppc:
  1.  GNP per capita, 1998

safewater:
  1.  Access to safe water, % of population, 1990-96

. sum popgrowth

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   popgrowth |         68    .9720588    .9311918        -.5          3

. sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      region |         68         1.5    .7431277          1          3
     country |          0
   popgrowth |         68    .9720588    .9311918        -.5          3
        lexp |         68    72.27941    4.715315         54         79
       gnppc |         63    8674.857    10634.68        370      39980
-------------+---------------------------------------------------------
   safewater |         40        76.1    17.89112         28        100

.
```

Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

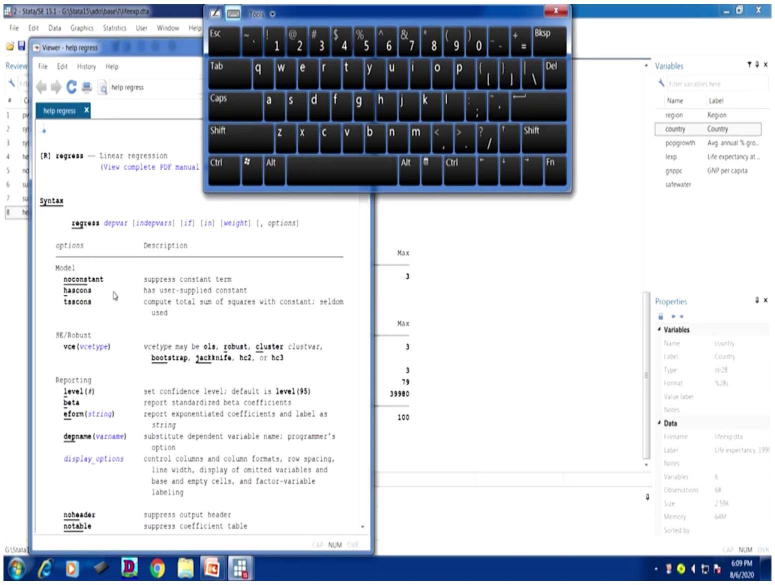
Similarly, Using a file, you, either click on the window here. If you do not know the path name in your command, you simply click on the window and it will automatically attach the path name to that particular place. Otherwise, you better type, if you know the path name and even click, get the path name, which I told you in the last class, you get the path name and type, it will open the file. But while we have already started typing the command, you cannot click the window to get the path name. You have to get it from that particular folder of the D file. If it is in D file, click on the D file, it will automatically come. So I think these are already done a bit. I am not spending much time on it.

(Refer Slide Time: 29:34)

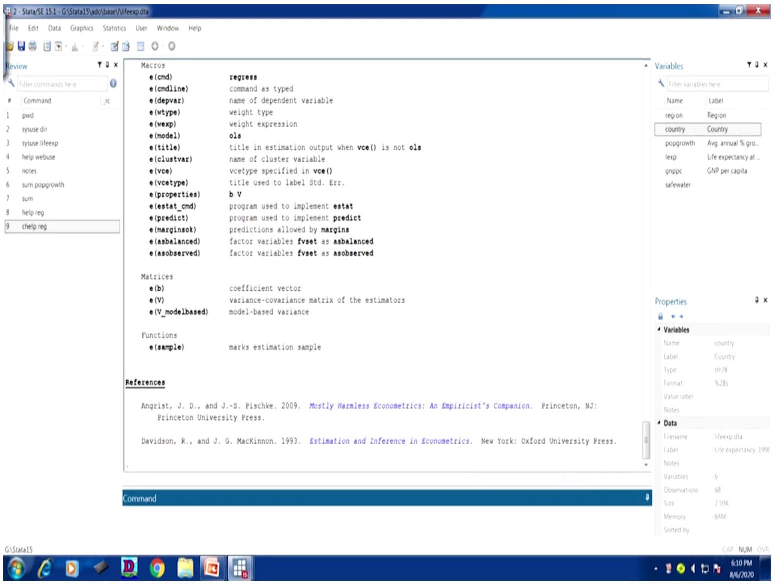
GETTING HELP ON STATA

- To obtain help on a command (or function) type help *command_name*, which displays the help on a separate window called the *Viewer*.

help *command_name*- display help on viewer window
chelp command_name- display help on result window
On menu bar- **Help -> stata command...**



```
[R] regress — Linear regression
              (View complete PDF manual

Syntax

    regress depvar [indepvars] [if] [in] [weight] [, options]

  options                  Description
  ─────────────────────────────────────────────────────────────
  Model
    noconstant             suppress constant term
    hascons                has user-supplied constant
    tsscons                compute total sum of squares with constant; seldom
                             used

  SE/Robust
    vce(vcetype)           vcetype may be ols, robust, cluster clustvar,
                             bootstrap, jackknife, hc2, or hc3

  Reporting
    level(#)               set confidence level; default is level(95)
    beta                   report standardized beta coefficients
    eform(string)          report exponentiated coefficients and label as
                             string
    depname(varname)       substitute dependent variable name; programmer's
                             option
    display_options        control columns and column formats, row spacing,
                             line width, display of omitted variables and
                             base and empty cells, and factor-variable
                             labeling

    noheader               suppress output header
    notable                suppress coefficient table
```


2 - Stata/SE 15.1 - G:\Stata15\ado\base\l\lifeexp.dta
File Edit Data Graphics Statistics User Window Help
Review
Variables
Properties
Command

```
1  pwd
2  sysuse dir
3  sysuse lifeexp
4  help webuse
5  notes
6  sum popgrowth
7  sum
8  help reg
9  chelp reg
```

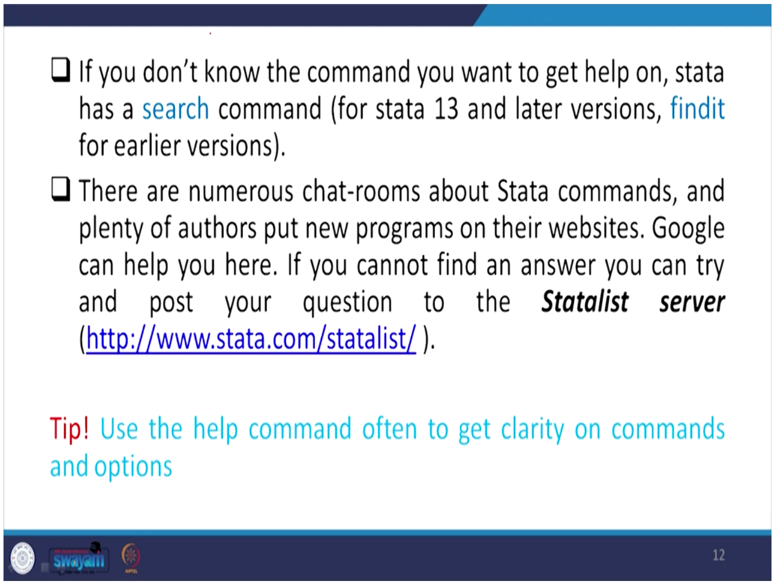
```
  Macros
    e(cmd)              regress
    e(cmdline)          command as typed
    e(depvar)           name of dependent variable
    e(wtype)            weight type
    e(wexp)             weight expression
    e(model)            ols
    e(title)            title in estimation output when vce() is not ols
    e(clustvar)         name of cluster variable
    e(vce)              vcetype specified in vce()
    e(vcetype)          title used to label Std. Err.
    e(properties)       b V
    e(estat_cmd)        program used to implement estat
    e(predict)          program used to implement predict
    e(marginsok)        predictions allowed by margins
    e(asbalanced)       factor variables fvset as asbalanced
    e(asobserved)       factor variables fvset as asobserved

  Matrices
    e(b)                coefficient vector
    e(V)                variance-covariance matrix of the estimators
    e(V_modelbased)     model-based variance

  Functions
    e(sample)           marks estimation sample


References

  Angrist, J. D., and J.-S. Pischke. 2009.  Mostly Harmless Econometrics: An Empiricist's Companion.  Princeton, NJ:
      Princeton University Press.

  Davidson, R., and J. G. MacKinnon. 1993.  Estimation and Inference in Econometrics.  New York: Oxford University Press.
```

Let us move to help command and chelp command. It is very simple to explain. Like you want help, suppose I will simply type help here, help I want regression, I simply reg, instead you can write regression reg, regression, or you can type reg in short. To enter, it will open in another window with its information of various type of regression and there are different entries here. You will get that information from this new window.

But chelp, instead of that, if I simply write chelp, as I told you already, the page up command in the keyboard, will give you the earlier command. So, here instead of help, you simply write chelp. So, write chelp. It will not give you another new window. It will give you on the same window. The information will pop-up in the same window. Sometimes it is easy to operate from the same window. So this option helps a lot.

Otherwise in the top down, click based approach you just go to the main menu bar and from the main menu bar there is a help option. Here is a help option, I have already highlighted you. So click on the help option, other information will come for your clarity.

(Refer Slide Time: 31:25)



- ❑ If you don't know the command you want to get help on, stata has a search command (for stata 13 and later versions, findit for earlier versions).
- ❑ There are numerous chat-rooms about Stata commands, and plenty of authors put new programs on their websites. Google can help you here. If you cannot find an answer you can try and post your question to the ***Statalist server*** (http://www.stata.com/statalist/).

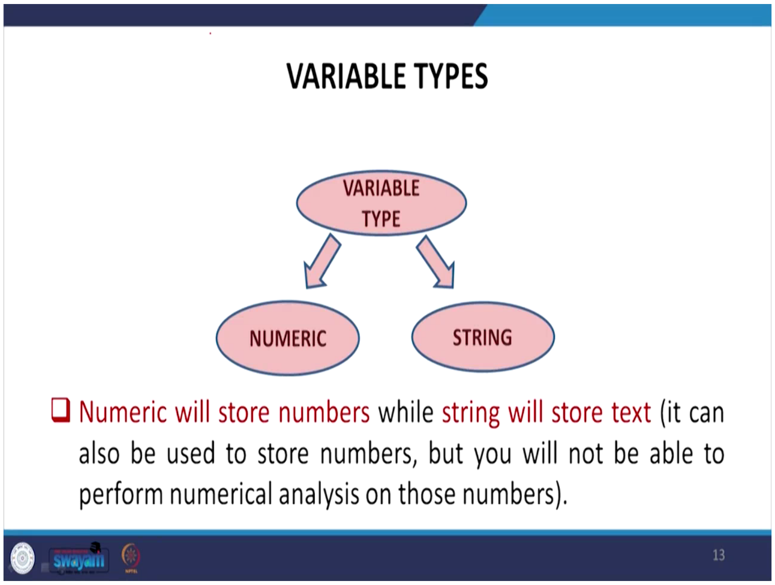
Tip! Use the help command often to get clarity on commands and options

12

Similarly, if you do not know the command, like we know the command reg, so we typed reg for help. If you do not know the command, simply type search and any command. You want regression, write regression. You want analysis, cross-tab analysis writes it, will give you another window for options, for help options.

Find it is usually used in the earlier version of Stata. From 13th version onwards, search command is given in the Stata, not find it command. So, there are numerous information you can get it from the chat room of Stata version or of the StataCorp. And otherwise, you can get all that information from Stata server.

(Refer Slide Time: 32:22)



VARIABLE TYPES

VARIABLE TYPE

NUMERIC

STRING

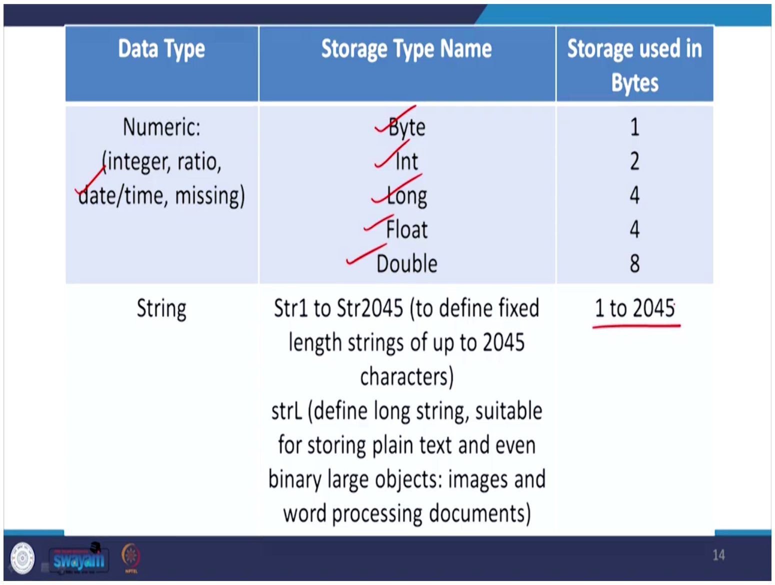
❑ Numeric will store numbers while string will store text (it can also be used to store numbers, but you will not be able to perform numerical analysis on those numbers).

13

And let me move into some other information like variable types, what kind of variables are there and how it is important for our analysis and how this is being entered. Broadly, variables are of numeric type and string type. So, some of the explanation we made earlier, lecture on numeric and string, but further I will explain you here how it is stored in Stata interface or in Stata data.

So, numeric will store numbers, while the string generally stored in text file. And it can also be used to store numbers, but you will not be able to perform numerical analysis if it is a string entry. So string entry is, if you simply operate on the string since it is in text entry, some characters are there. maybe character representing a picture, a graph.

(Refer Slide Time: 33:45)



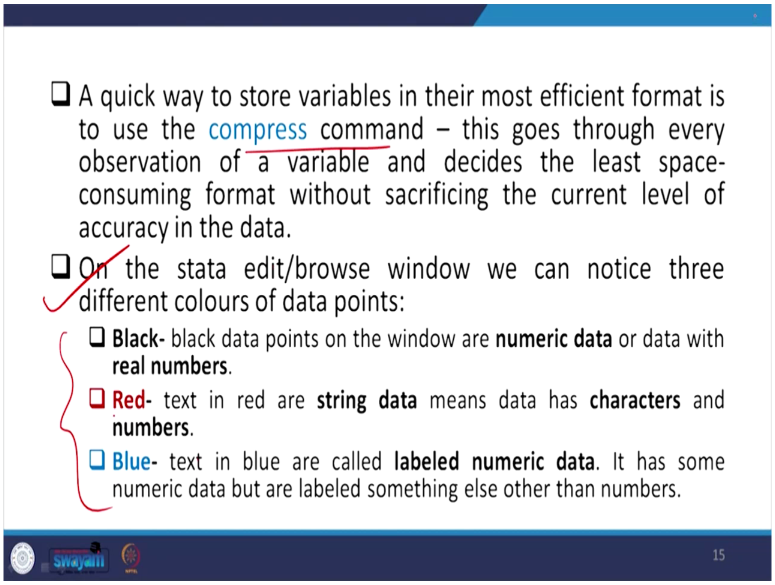
Data Type	Storage Type Name	Storage used in Bytes
Numeric: (integer, ratio, date/time, missing)	Byte Int Long Float Double	1 2 4 4 8
String	Str1 to Str2045 (to define fixed length strings of up to 2045 characters) strL (define long string, suitable for storing plain text and even binary large objects: images and word processing documents)	1 to 2045

14

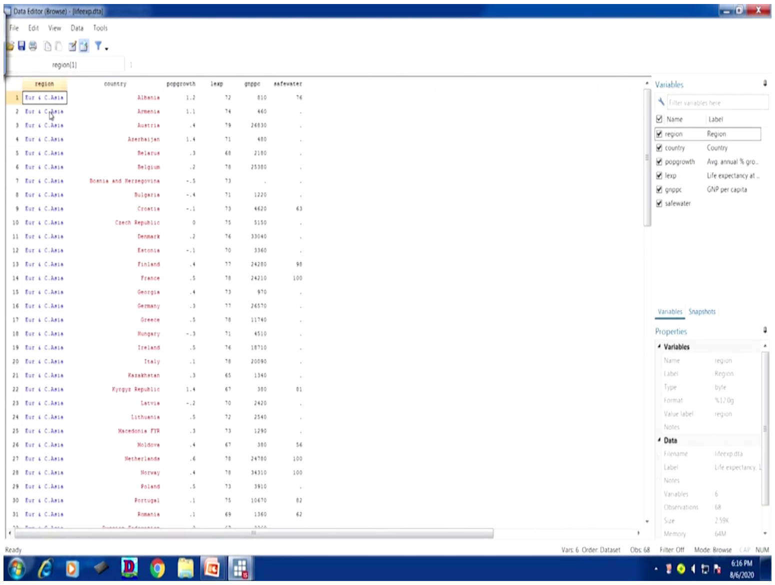
But if you simply just go by the command, we have already shown you some basic command, for the string variable it will not operate. I will clarify further. Like numeric data, it is either in integer, in a full number or in real number or in ratio format, in date or missing format. Missing data are also there. They generally given in a numeric number. So storage type is either byte, int, long, I will show you in the variable window, float or double, they consumed the byte space with one, for byte, if it is in byte space, integer two similarly, eight for double.

Whereas the string data, varies from 1 to 2,045 to define fixed length string up to 2,045 characters. So it consumes till 2,045 characters. If it is a long string, then it comes with strL and this is generally suitable for storing plain text, even the binary large objects as well images and word processing documents.

(Refer Slide Time: 35:10)



- ❑ A quick way to store variables in their most efficient format is to use the compress command – this goes through every observation of a variable and decides the least space-consuming format without sacrificing the current level of accuracy in the data.
- ❑ On the stata edit/browse window we can notice three different colours of data points:
  - ❑ **Black-** black data points on the window are **numeric data** or data with **real numbers**.
  - ❑ **Red-** text in red are **string data** means data has **characters** and **numbers**.
  - ❑ **Blue-** text in blue are called **labeled numeric data**. It has some numeric data but are labeled something else other than numbers.



	region	country	popgrowth	lexp	gnppc	safewater
1	Eur & C.Asia	Albania	1.2	72	810	76
2	Eur & C.Asia	Armenia	1.1	74	460	.
3	Eur & C.Asia	Austria	.4	79	26830	.
4	Eur & C.Asia	Azerbaijan	1.4	71	480	.
5	Eur & C.Asia	Belarus	.3	68	2180	.
6	Eur & C.Asia	Belgium	.2	78	25380	.
7	Eur & C.Asia	Bosnia and Herzegovina	-.5	73	.	.
8	Eur & C.Asia	Bulgaria	-.4	71	1220	.
9	Eur & C.Asia	Croatia	-.1	73	4620	63
10	Eur & C.Asia	Czech Republic	0	75	5150	.
11	Eur & C.Asia	Denmark	.2	76	33040	.
12	Eur & C.Asia	Estonia	-.1	70	3360	.
13	Eur & C.Asia	Finland	.4	77	24280	98
14	Eur & C.Asia	France	.5	78	24210	100
15	Eur & C.Asia	Georgia	.4	73	970	.
16	Eur & C.Asia	Germany	.3	77	26570	.
17	Eur & C.Asia	Greece	.5	78	11740	.
18	Eur & C.Asia	Hungary	-.3	71	4510	.
19	Eur & C.Asia	Ireland	.5	76	18710	.
20	Eur & C.Asia	Italy	.1	78	20090	.
21	Eur & C.Asia	Kazakhstan	.3	65	1340	.
22	Eur & C.Asia	Kyrgyz Republic	1.4	67	380	81
23	Eur & C.Asia	Latvia	-.2	70	2420	.
24	Eur & C.Asia	Lithuania	.5	72	2540	.
25	Eur & C.Asia	Macedonia FYR	.3	73	1290	.
26	Eur & C.Asia	Moldova	.4	67	380	56
27	Eur & C.Asia	Netherlands	.6	78	24780	100
28	Eur & C.Asia	Norway	.4	78	34310	100
29	Eur & C.Asia	Poland	.5	73	3910	.
30	Eur & C.Asia	Portugal	.1	75	10670	82
31	Eur & C.Asia	Romania	.1	69	1360	62

Name	Label
☑ region	Region
☑ country	Country
☑ popgrowth	Avg. annual % gro...
☑ lexp	Life expectancy at ...
☑ gnppc	GNP per capita
☑ safewater	

Variables

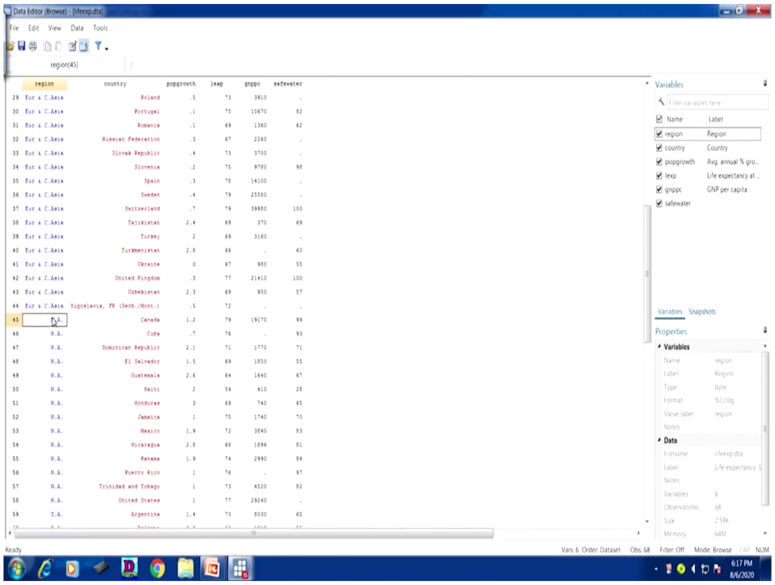
Name	region
Label	Region
Type	byte
Format	%12.0g
Value label	region
Notes	

Data

Filename	lifeexp.dta
Label	Life expectancy, 1
Notes	
Variables	6
Observations	68
Size	2.59K
Memory	64M

Ready

Vars: 6 Order: Dataset Obs: 68 Filter: Off Mode: Browse CAP NUM



region[45] 2

	region	country	popgrowth	lexp	gnppc	safewater
29	Eur & C.Asia	Poland	.5	73	3910	.
30	Eur & C.Asia	Portugal	.1	75	10670	82
31	Eur & C.Asia	Romania	.1	69	1360	62
32	Eur & C.Asia	Russian Federation	.3	67	2260	.
33	Eur & C.Asia	Slovak Republic	.4	73	3700	.
34	Eur & C.Asia	Slovenia	.2	75	9780	98
35	Eur & C.Asia	Spain	.3	78	14100	.
36	Eur & C.Asia	Sweden	.4	79	25580	.
37	Eur & C.Asia	Switzerland	.7	79	39980	100
38	Eur & C.Asia	Tajikistan	2.4	69	370	69
39	Eur & C.Asia	Turkey	2	69	3160	.
40	Eur & C.Asia	Turkmenistan	2.8	66	.	60
41	Eur & C.Asia	Ukraine	0	67	980	55
42	Eur & C.Asia	United Kingdom	.3	77	21410	100
43	Eur & C.Asia	Uzbekistan	2.3	69	950	57
44	Eur & C.Asia	Yugoslavia, FR (Serb./Mont.)	.5	72	.	.
45	N.A.	Canada	1.2	79	19170	99
46	N.A.	Cuba	.7	76	.	93
47	N.A.	Dominican Republic	2.1	71	1770	71
48	N.A.	El Salvador	1.5	69	1850	55
49	N.A.	Guatemala	2.6	64	1640	67
50	N.A.	Haiti	2	54	410	28
51	N.A.	Honduras	3	69	740	65
52	N.A.	Jamaica	1	75	1740	70
53	N.A.	Mexico	1.9	72	3840	83
54	N.A.	Nicaragua	2.8	68	1896	81
55	N.A.	Panama	1.9	74	2990	84
56	N.A.	Puerto Rico	1	76	.	97
57	N.A.	Trinidad and Tobago	1	73	4520	82
58	N.A.	United States	1	77	29240	.
59	S.A.	Argentina	1.4	73	8030	65

Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

Variables	
Name	region
Label	Region
Type	byte
Format	%12.0g
Value label	region
Notes	
Data	
Filename	lifeexp.dta
Label	Life expectancy, 1
Notes	
Variables	6
Observations	68
Size	2.59K
Memory	64M

Vars: 6 Order: Dataset Obs: 68 Filter: Off Mode: Browse CAP NUM

6:17 PM 8/6/2020

Since there are so many different windows and so many different commands and those commands takes huge space, usually consume our much of a space and very difficult to operate, it is always suggested that you start with compress command. You type compress with a data, it will convert your data to a compressed data format in Stata. The data which we have shown you here is already in a compressed format. If you type compress is not going to be compressed further. They have already suitably compressed the data to the best compressed format.

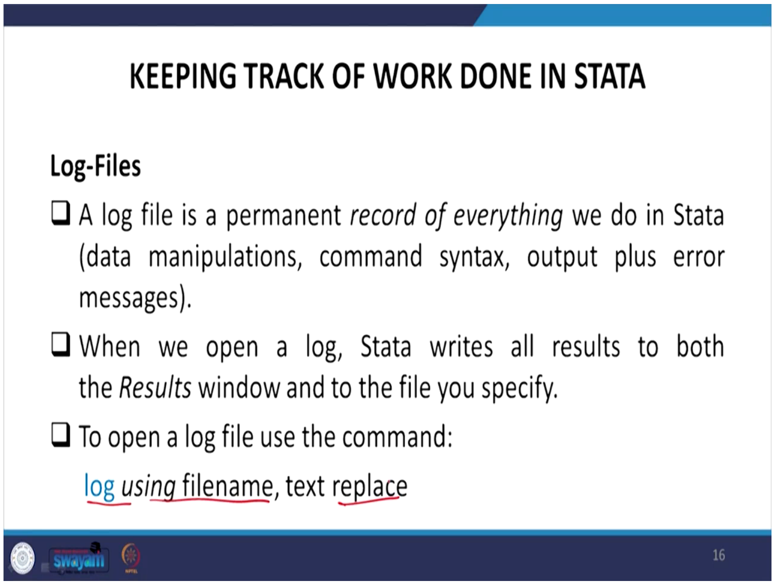
It is interesting to note the important part here is how the data is loaded in Stata, in .dta format. Usually in three colors we told you even in the earlier, but I will show you, black, red and blue. Better you go to the data browser to see it correctly the data which we have already loaded from the Stata directory. So this is the life expectancy data. And in this data, what I will do, I will just show you. I am just going to click the first entry, very first entry. When I enter it, one is written here, but the name is shown to us Europe and Central Asia, but one is entered in the data.

So I will tell you why they have made different colors to different entries. There is a black color, there is a red color and there is a blue color. Black color are simply real numbers or the numeric. Whereas the blue color is not purely numeric, it is a labeled with numeric. Central Asia they have coded as one. If I just click on, I am just scrolling further, I am just putting a click here and going little down till here.

NA stands for North America, but the entry with a label two is there. So whichever the variables which are labeled one, they are highlighted with blue color. Whereas the red color are string.

They are character space. With the same command we cannot operate those one. We have to convert it to other format maybe sometimes we do destringing, we will tell you later. At this moment, I am not spending much time. I have already clarified three colors. Other details we will clarify in our other lectures.

(Refer Slide Time: 38:20)



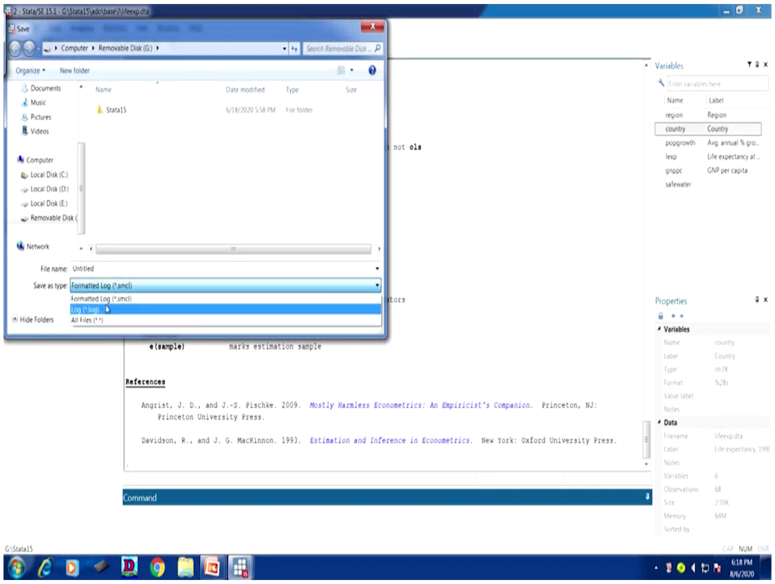
KEEPING TRACK OF WORK DONE IN STATA

Log-Files

- ❑ A log file is a permanent *record of everything* we do in Stata (data manipulations, command syntax, output plus error messages).
- ❑ When we open a log, Stata writes all results to both the *Results* window and to the file you specify.
- ❑ To open a log file use the command:

  log *using* filename, text replace

16



2 - Stata/SE 15.1 - G:\Stata15\ado\base\l\lifeexp.dta

Save

Computer ▸ Removable Disk (G:) ▸

Name	Date modified	Type	Size
Stata15	6/18/2020 5:58 PM	File folder	

File name: Untitled

Save as type: Formatted Log (*.smcl)

Formatted Log (*.smcl)
Log (*.log)
All Files (*.*)

```
not ols

tors

    e(sample)          marks estimation sample

References

    Angrist, J. D., and J.-S. Pischke. 2009.  Mostly Harmless Econometrics: An Empiricist's Companion.  Princeton, NJ:
        Princeton University Press.

    Davidson, R., and J. G. MacKinnon. 1993.  Estimation and Inference in Econometrics.  New York: Oxford University Press.
```

Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

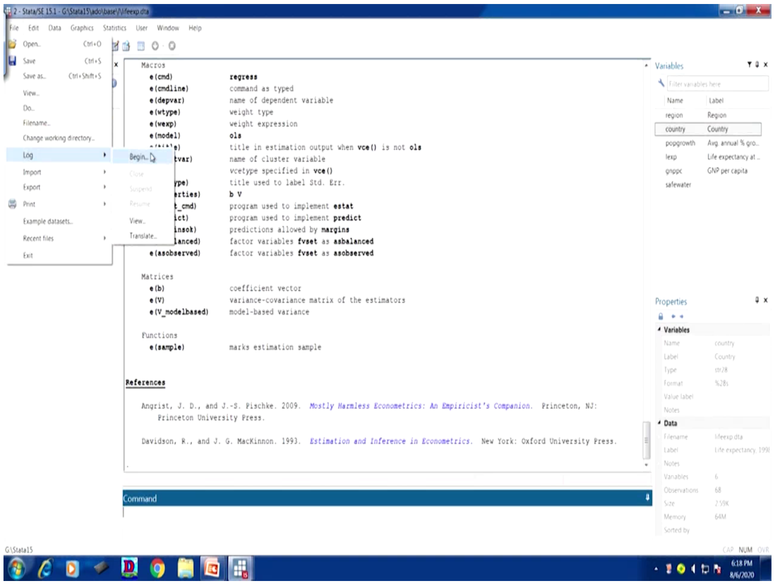
Properties

Variables

Name	country
Label	Country
Type	str28
Format	%28s
Value label	
Notes	

Data

Filename	lifeexp.dta
Label	Life expectancy, 1998
Notes	
Variables	6
Observations	68
Size	2.59K
Memory	64M
Sorted by	



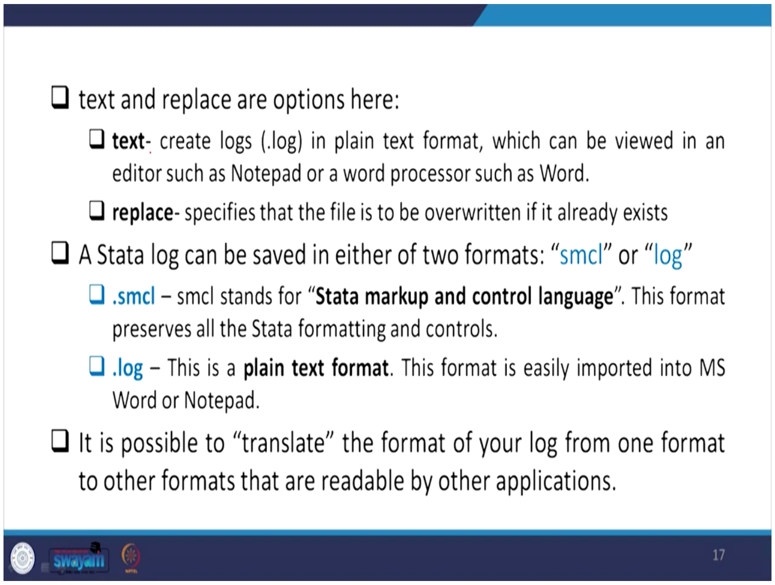
```
  Macros
    e(cmd)              regress
    e(cmdline)          command as typed
    e(depvar)           name of dependent variable
    e(wtype)            weight type
    e(wexp)             weight expression
    e(model)            ols
    e(title)            title in estimation output when vce() is not ols
           tvar)        name of cluster variable
                        vcetype specified in vce()
           ype)         title used to label Std. Err.
           erties)      b V
           t_cmd)       program used to implement estat
           ict)         program used to implement predict
           insok)       predictions allowed by margins
           lanced)      factor variables fvset as asbalanced
    e(asobserved)       factor variables fvset as asobserved

  Matrices
    e(b)                coefficient vector
    e(V)                variance-covariance matrix of the estimators
    e(V_modelbased)     model-based variance

  Functions
    e(sample)           marks estimation sample


References

  Angrist, J. D., and J.-S. Pischke. 2009.  Mostly Harmless Econometrics: An Empiricist's Companion.  Princeton, NJ:
     Princeton University Press.

  Davidson, R., and J. G. MacKinnon. 1993.  Estimation and Inference in Econometrics.  New York: Oxford University Press.
```

Variables

Name	Label
region	Region
country	Country
popgrowth	Avg. annual % gro...
lexp	Life expectancy at ...
gnppc	GNP per capita
safewater	

Properties

Variables	
Name	country
Label	Country
Type	str28
Format	%28s
Value label	
Notes	
Data	
Filename	lifeexp.dta
Label	Life expectancy, 1998
Notes	
Variables	6
Observations	68
Size	2.59K
Memory	64M
Sorted by	



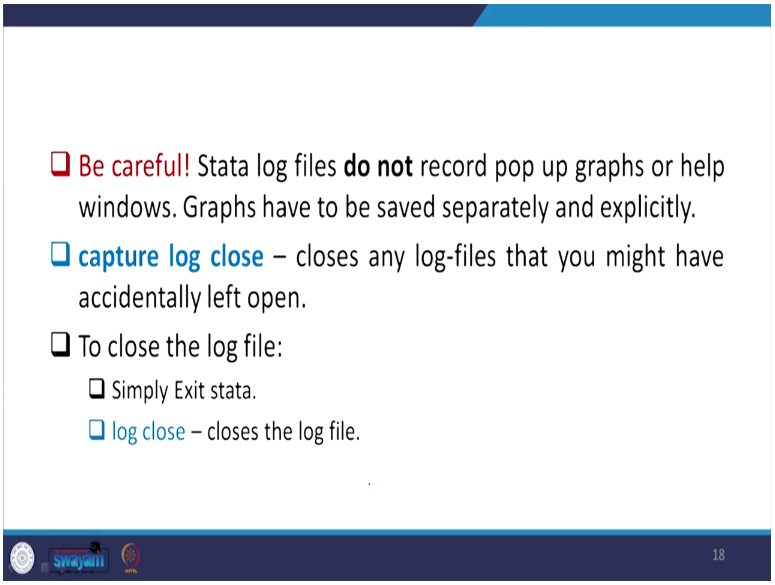
- ❑ text and replace are options here:
  - ❑ **text**- create logs (.log) in plain text format, which can be viewed in an editor such as Notepad or a word processor such as Word.
  - ❑ **replace**- specifies that the file is to be overwritten if it already exists
- ❑ A Stata log can be saved in either of two formats: "smcl" or "log"
  - ❑ .smcl – smcl stands for "**Stata markup and control language**". This format preserves all the Stata formatting and controls.
  - ❑ .log – This is a **plain text format**. This format is easily imported into MS Word or Notepad.
- ❑ It is possible to "translate" the format of your log from one format to other formats that are readable by other applications.

Log file, as I already told you, is also important. How to open log file? Simply you click here log begin. It will start saving with the file log. But log by default it is coming with .smcl. But for our further use, like we want to copy those for our word file, better to convert it into txt log file. Log is there another format, if you save it, it will save in txt, text format. And text format is generally easy for converting the log file to our word file. This saves differently.

You can start from file, then there is a log begin. If you click log begin, you can also do that. So, all those details are there. We may continue from the next class. You can read it. We will continue. I have just given you the command called log. Otherwise, I have already shown you how to operate through click based approach. Otherwise, you can simply type log using filename if you know you can enter a file name, command txt replace. Replace is there if any existing log file are there, it will automatically replace.

Txt, if you do not give txt it will automatically save by default in smcl format, which is not generally suggested if you are copying that to a word file. So this is the way you can do it. Rest of the details we will continue.

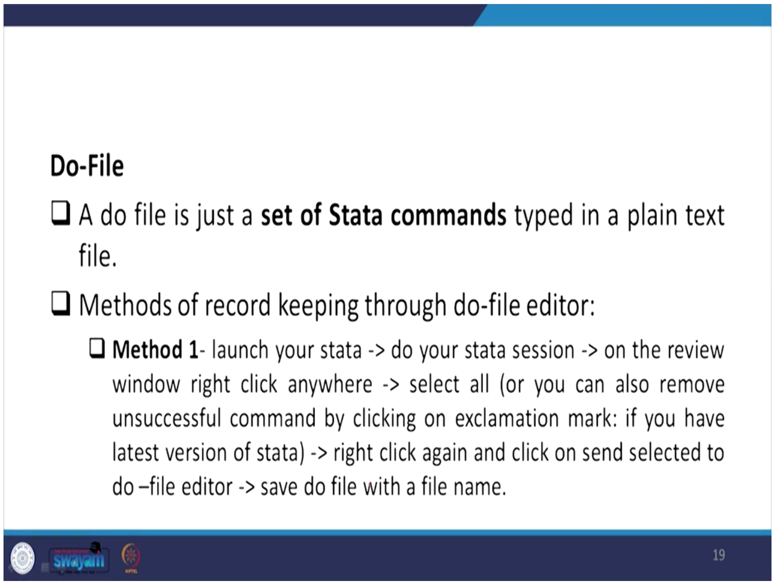
(Refer Slide Time: 40:26)



- □ Be careful! Stata log files **do not** record pop up graphs or help windows. Graphs have to be saved separately and explicitly.
- □ **capture log close** – closes any log-files that you might have accidentally left open.
- □ To close the log file:
  - □ Simply Exit stata.
  - □ log close – closes the log file.

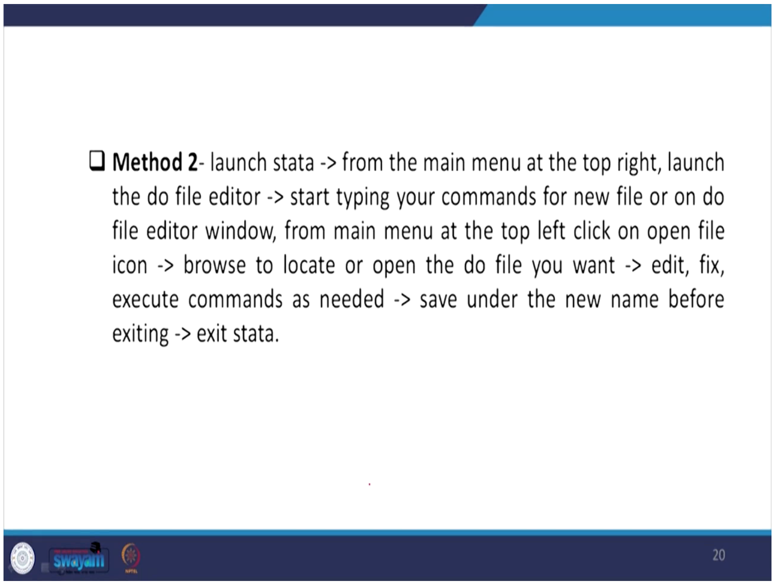
We have other details also for your explanation. We will do it in our next class, like we will talk about how to read log files and where to use log files and how to use log files. We will also clarify capture, close, whether to close the log file or not, what is wrong if you do not close it.

(Refer Slide Time: 40:34)



Do-File

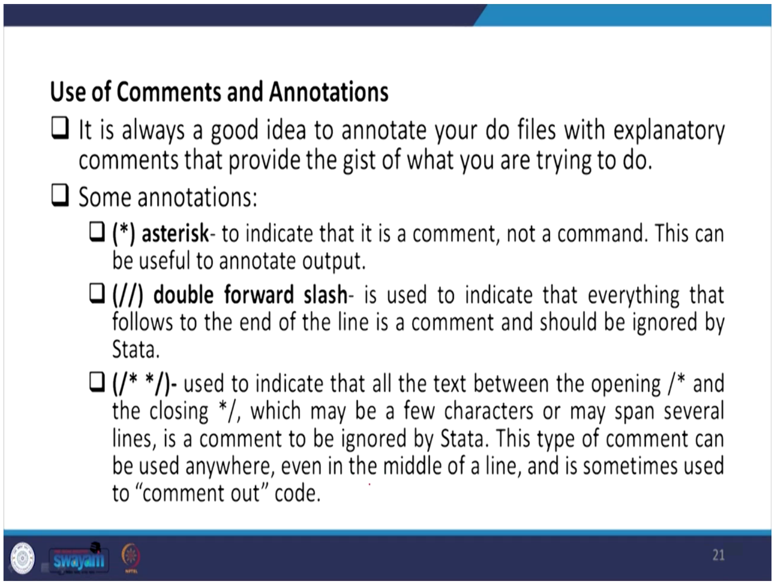
- ❑ A do file is just a **set of Stata commands** typed in a plain text file.
- ❑ Methods of record keeping through do-file editor:
  - ❑ **Method 1**- launch your stata -> do your stata session -> on the review window right click anywhere -> select all (or you can also remove unsuccessful command by clicking on exclamation mark: if you have latest version of stata) -> right click again and click on send selected to do –file editor -> save do file with a file name.



- ❑ **Method 2**- launch stata -> from the main menu at the top right, launch the do file editor -> start typing your commands for new file or on do file editor window, from main menu at the top left click on open file icon -> browse to locate or open the do file you want -> edit, fix, execute commands as needed -> save under the new name before exiting -> exit stata.

We have also information on do-file. How to operate do-file, what kind of commands we generally do, what is the best method of operating do-file. We will continue those things in the next class.

(Refer Slide Time: 40:44)



Use of Comments and Annotations

- ❑ It is always a good idea to annotate your do files with explanatory comments that provide the gist of what you are trying to do.
- ❑ Some annotations:
  - ❑ **(*) asterisk**- to indicate that it is a comment, not a command. This can be useful to annotate output.
  - ❑ **(//) double forward slash**- is used to indicate that everything that follows to the end of the line is a comment and should be ignored by Stata.
  - ❑ **(/* */)-** used to indicate that all the text between the opening /* and the closing */, which may be a few characters or may span several lines, is a comment to be ignored by Stata. This type of comment can be used anywhere, even in the middle of a line, and is sometimes used to "comment out" code.

21

Within the do-file, it is very interesting to note some asterisk entry, double forward slash, triple forward slash, even forward asterisk from the beginning and the end, how it reads. If you execute from there, it operates with those commands. We will continue that in the next class. I think, it is better to close here. There are so many information given for you, for your preparation. You please go through other details, we will continue. With this, let me close here. Thank you.