

Exercise 1: Introduction to Stata

New Stata Commands	New Stata Commands
use	
describe	
summarize	
stem	
graph box	
histogram	
log on, off	
exit	

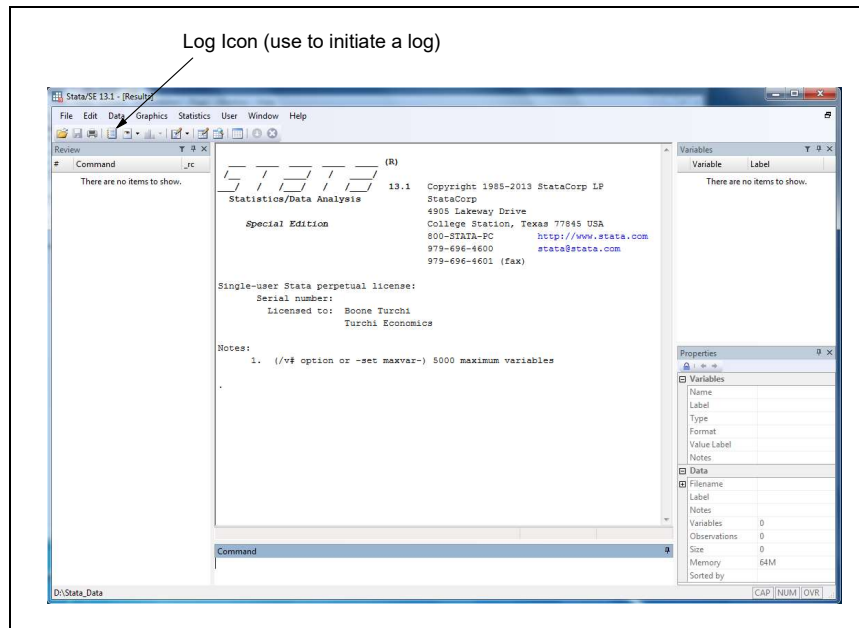
Downloading Data from the Web

You should be able to use Chrome, Firefox, Safari or Edge to download data from the course website.

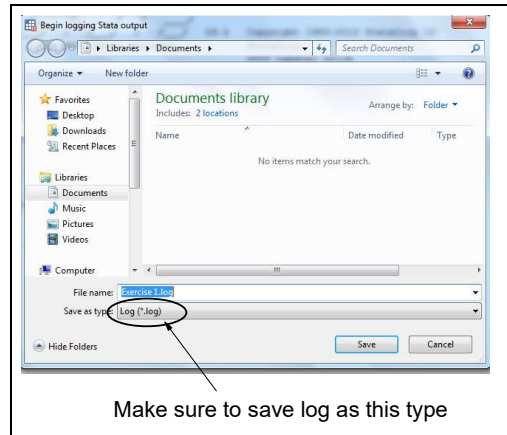
This exercise requires that you use the Stata dataset “cardata.dta” and the *Ec400.dotx* template which will be found on the course web site in the section containing data sets. Download the data set to your hard drive according to the instructions and place the data set in a folder where you can find it easily.

Reading data into Stata:

1. Start Stata by clicking on the Stata icon. The main Stata window will open as shown here:



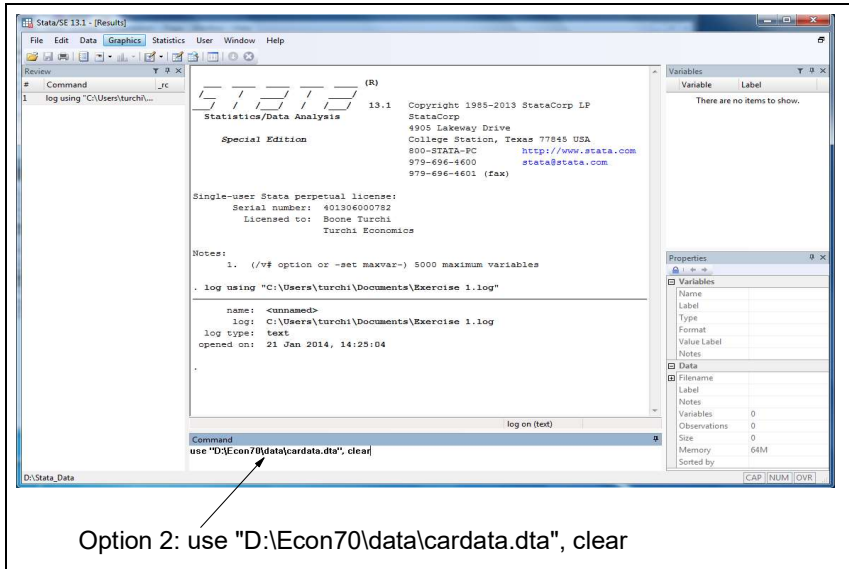
- Next, you want to open a “Log File.” The log file records all the commands that you execute and your numerical results during a given Stata session. To turn on the recording, click on the icon which looks like a notebook page (this icon is located immediately to the right of the Print button on the toolbars {see previous graph}). Now save the log file with the name “Exercise1.log”, in “My Documents” folder (or a folder of your choosing where you can find it again):



- While saving a log file, you can either save it with a new name, or with an existing log file name. If you use an existing name, the program will ask you whether you want to append to or overwrite the existing file. If you wish to continue the previous recording of your activities, then choose “Append”. If you want to start fresh, select the option “Overwrite”. The blank log file will show up in front. Simply click on another part of the Stata window to bring the “Stata Results” window back to the front. **Important Note:** In the "Files of Type: window select "Log(*.log)". It will give you a plain text log. *Do not use the **Formatted Log (*.scml)** setting as it will give you a formatted result which is not easy to paste into Word.*
- Now it’s time to read in the data set “cardata” into Stata. There are a number of ways to do this. The first option is to click the “open file” icon at the far left of the tool bar The second option is to type

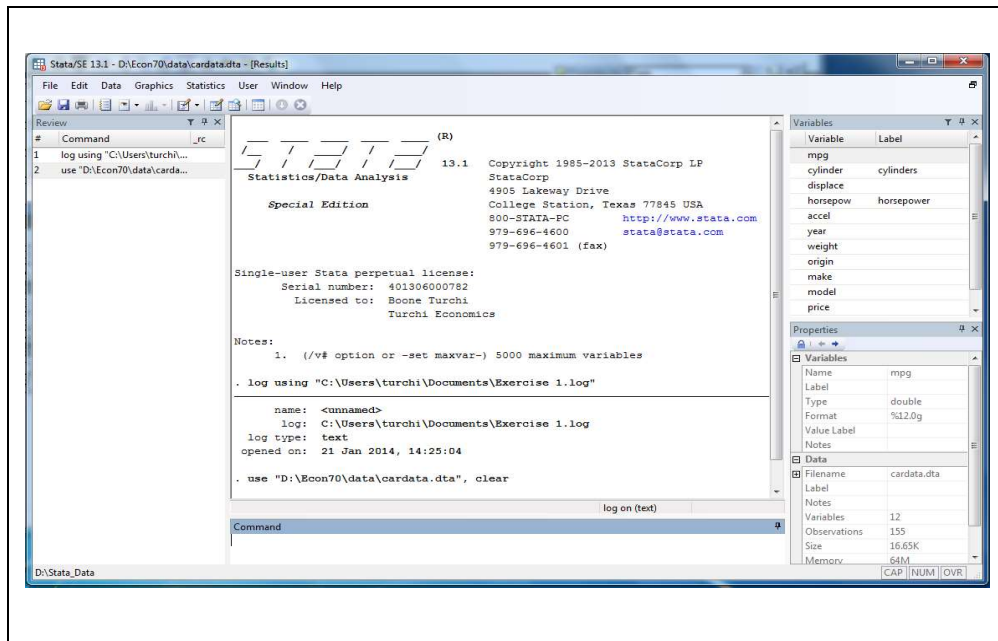
`use D:\...your directory path...\cardata.dta,clear`

Use only one of these options, not both!

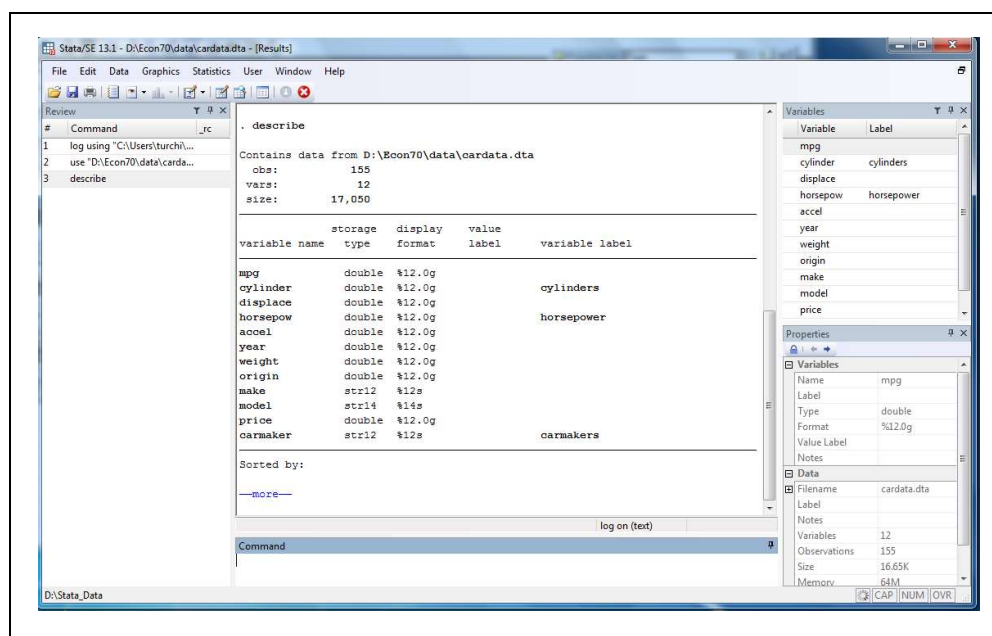


(Actually, there’s a third way: find the data set [on your desktop or in the folder where you saved it] and double-click it. Stata will open with the file already in memory and ready to use.)

5. Stata will read in the data. Names of variables will appear in the “variables window”. Previous commands will appear in the “Review” window where they can be selected and used again in the “Stata comand” window:



6. Now, we’re ready to do some data analysis. Enter “describe” in the “Stata Command” window and press “enter.” You’ll see the following screen:



In the “Stata Results Window” you’ll see a description of the data set including number of observations, number of variables, size of the file, followed by a listing of the variables in the data set. Under “Storage type” you’ll see “double” and “str12” “str14”. Variables that are “double” are numeric variables stored in double precision for extra accuracy. “Str12 and Str14” variables are character variables that represent things like model name, etc.

If “-more” appears at the bottom of the “Results” screen simply tap the space bar to show more of the results.

Performing descriptive analysis :

By now you should be familiar with the procedure of typing in commands at the “Stata command” window, and pressing “enter” to get them executed. In the same manner, type in the syntaxes/commands, which are similar to those written below in courier letters. Pressing “enter” at the end of each command will show you the output for that command in the Stata Results window.

Describing a variable

Type the following command in the “Stata command” window:

```
summarize mpg
```

In the results window you’ll get:

Economics 400: Professor Turchi

```
. summarize mpg
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
mpg	154	28.79351	7.37721	15.5	46.6

The results show for variable “mpg” the number of nonmissing observations, the mean (or average) of the variable, its standard deviation and the minimum and maximum values in the data set.

Now, go to the Review window and click on “summarize mpg” and you’ll find it in the command window again. At the end of the command line type “, detail” so your command line looks like:

```
. summarize mpg, detail {enter}
```

Now you’ll get additional detail for mpg:

```
. summarize mpg, detail
```

```
mpg
-----
Percentiles      Smallest
1%              16.2        15.5
5%              17.6        16.2
10%             19.1        16.5      Obs           154
25%             22.4        16.9      Sum of Wgt.   154

50%             28.9
75%             34.3        44
90%             38          44.3      Variance      54.42323
95%             40.9        44.6      Skewness      .1115568
99%             44.6        46.6      Kurtosis      2.155094
```

showing additional information about “mpg” including distribution of values, the median (28.9 mpg) and other summary measures of the distribution.

Now, why don’t you try doing the same thing for two other variables: “price” “horsepower” and “weight”. The command would be:

```
summarize price horsepower weight {enter}
```

Now, try this:

```
summarize
```

What do you get?

Stem and Leaf plot

You can also describe the data graphically. Type:

```
stem mpg
. stem mpg

Stem-and-leaf plot for mpg

mpg rounded to nearest multiple of .1
plot in units of .1

 15* | 5
 16* | 259
 17* | 005667
 18* | 11256
 19* | 12224489
 20* | 2222355668
 21* | 156
 22* | 034
 23* | 002567899
 24* | 023
 25* | 01448
 26* | 04668
 27* | 0000222459
 28* | 000148
 29* | 05889
 30* | 00479
 31* | 00035689
 32* | 00012344789
 33* | 0578
 34* | 0011234557
 35* | 017
 36* | 00000114
 37* | 000237
 38* | 00001
 39* | 014
 40* | 89
 41* | 5
 42* |
 43* | 14
 44* | 036
 45* |
 46* | 6
```

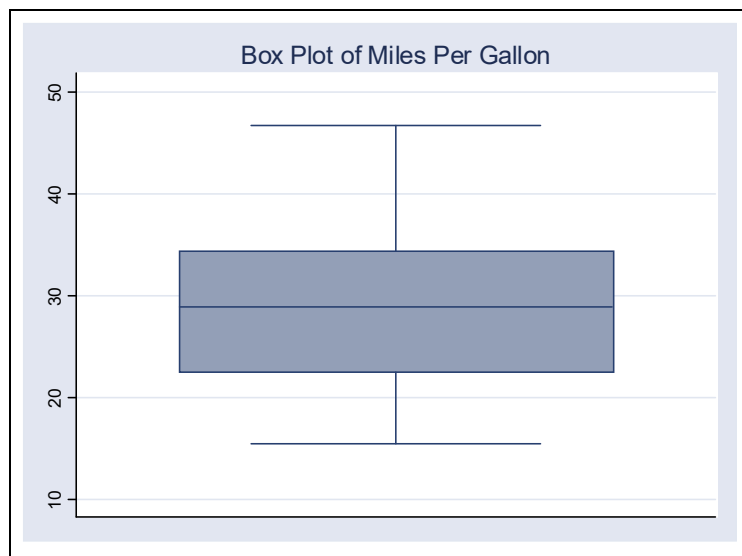
Here's the diagram shows the distribution of mpg.

Box-and-whisker plot

Another way to describe a variable is with a boxplot. Here we're making extensive use of Stata's graphics facilities:

Enter the command:

```
graph box mpg, title(" Box Plot of Miles per Gallon")
```



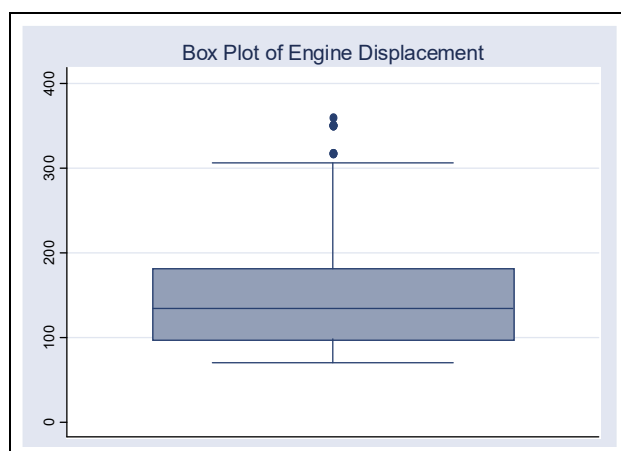
This plot appears in a new “Stata Graph” window and shows the median, the two middle quartiles of the data and “whiskers”. If you want to save this graph, you could go to the Edit menu and select “copy graph”. Then, open a Microsoft Word document and “paste” the graph onto a page. Or right-click on the graph itself and you’ll get some options.

Open a Word document, copy the graph in Stata, and then go to the Word document and go to Edit/paste (or type Cntrl-V) and see if you were able to copy the graph to a Word document successfully.

Now enter

```
graph box displace, title("Box Plot of Engine Displacement")
```

You’ll get the following box and whiskers plot:



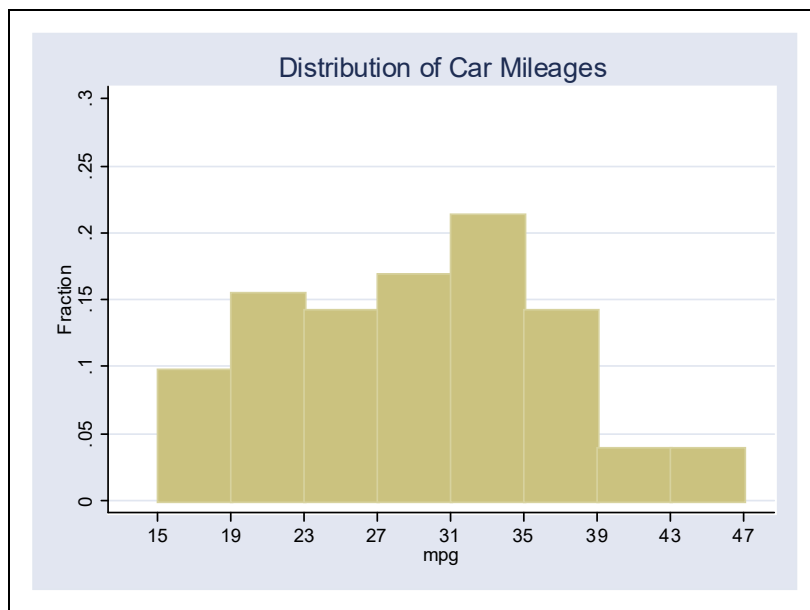
Note the small circles on the top of the whisker. These observations are “outliers”, cars that have really big engines.

Histograms

Finally, let's create a histogram of the mpg data. Enter the following command:

```
histogram mpg, fraction title("Distribution of Car Mileages")  
xlabel(15 (4) 48) ylabel(0 (.05) .3) start(15) width(4)
```

(This more elaborate command specifies lots of information about the axes, how the frequencies should be shown [it's "fraction" but could also be "frequency" or "percent"] and how many bars there should be and how wide each bar should be. Check Hamilton pp. 70-78 for further information. Getting the axes and bars lined up is kind of tricky; experimentation will make you an expert in doing this.)



Save this histogram to the Word document, identically as you did the box plot.

Note: All during the session, the log file keeps recording your commands and the results (except for the graphs) for future reference. You can view all that has been saved in the log file, by going to the File/View menu or, even easier, clicking the *log* icon on Stata's toolbar; there you'll see three options (1) view a snapshot of the log; (2) suspend the log; (3) close the log. Choose (1) and the log window will pop up and you'll see your work.

If at some point in any session, you want to suspend the log's recording of commands and results, you can simply suspend the log by clicking on the *log* icon and choosing "suspend". When you want to start recording again, you can click the *log* icon and choose "resume suspended log."

1. Save the Word file with the name *ex1_graphs*.

2. Close the log file by clicking the log icon and choosing “close log file.”
3. End your Stata session by typing “exit” in the command window.

Be sure to store your Word and log files where you can find them. We’ll need them later.

Turning in the exercise:

Once you’ve completed all your computer work, you need to format the material and submit it for grading. You will need your log file and the two graphs that you saved.

Using the “Instructions for Completing Computer Exercises” create a Microsoft Word document (*using the Ec400.dotx template*) that contains this material and turn in the exercise on the date required.

The exercise should contain the following items in the order specified:

1. Include the results of the “describe” command which describes the “cardata” data set.
2. Include the results of the “summarize mpg, detail” command
3. Include the results of the “stem mpg” command (be sure to put the stem-leaf display in “courier new” font so the results line up properly.
4. Include the graph produced with the “graph box mpg” command and write a few sentences which describe what this box plot is showing the reader (e.g., median mpg, min, max, bottom of top quartile, top of bottom quartile, etc.)
5. Include the graph produced with the “histogram mpg” command and write a few sentences in which you compare and contrast the kind of information the histogram gives you in comparison with the box plot. Which do you prefer in this instance?
6. Finally, print out your log file (again specifying couriers or other fixed pitch type) and attach it to the exercise as an appendix to the exercise.