



Introduction to Stata

October 19, 2020

Housekeeping

- We are recording
- Please feel free to ask questions!



01

Introduction

02

Stata Interface

03

Loading in a Data File

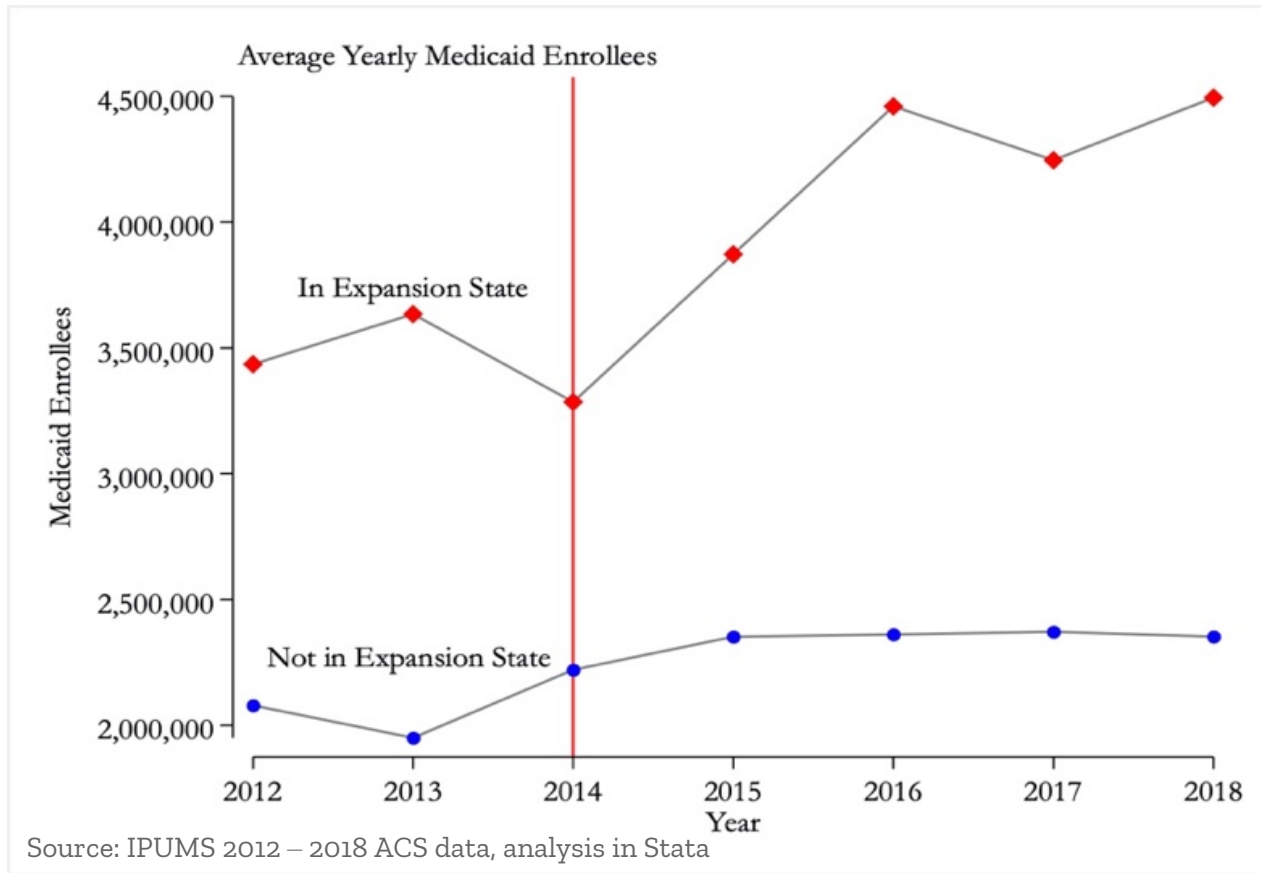
04

Basic Stata Commands

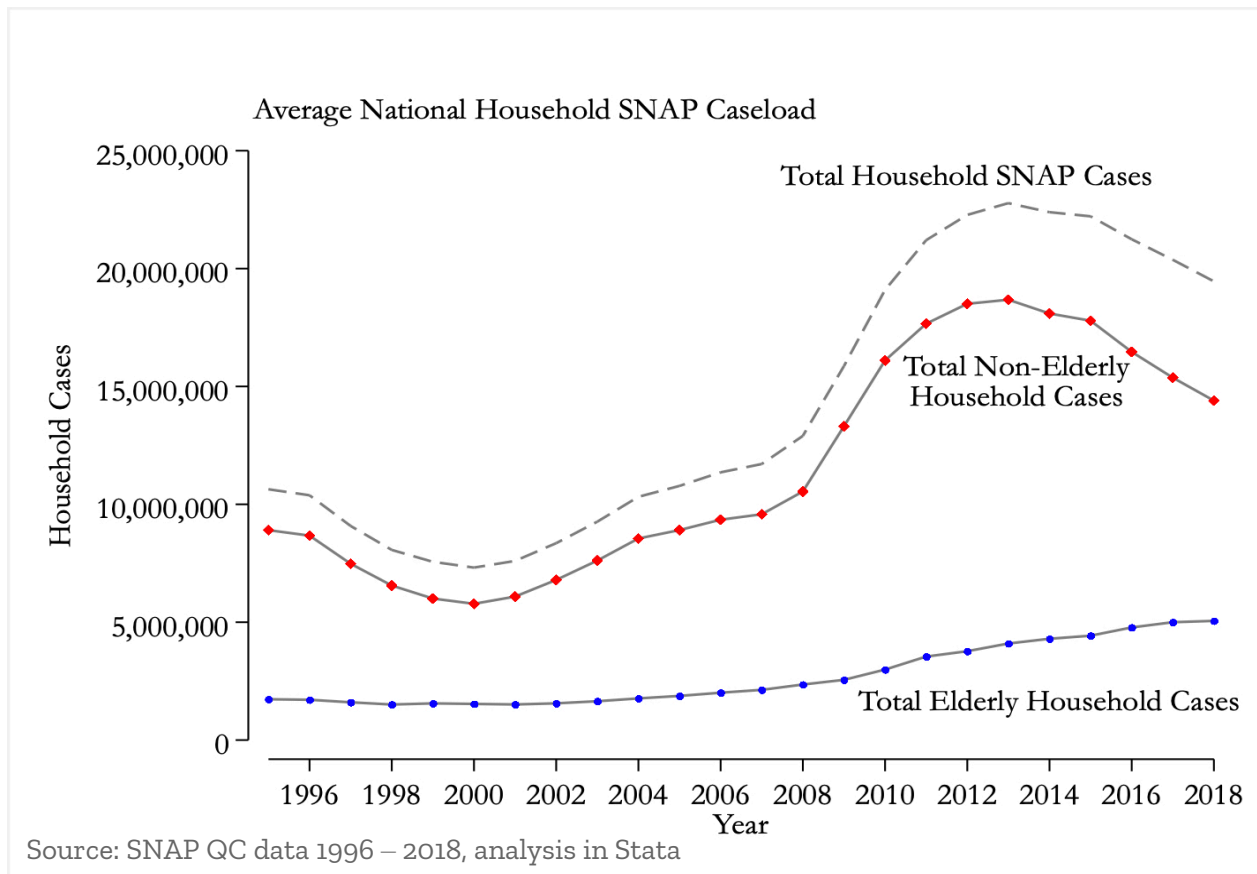
What is Stata?

STATA is a multipurpose statistical package to help perform data analysis, manipulate data, and create graphics.

Average Medicaid Enrollees per Year, Before and After Medicaid Expansion



Average National Household SNAP Caseload



Accessing Stata

1. Download software
2. Strauss
3. Through UD Virtual Lab

Accessing UD Virtual Lab

1. Palo Alto VPN Client
2. VMware Horizon

```
History Results
Command _rc

Review/history:
lists all
commands, can
click on a
previous
command to run

----- (R)
/----- /----- /----- /-----
/----- /----- /----- /-----
Statistics/Data analysis 16.1 Copyright 1985-2019 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC https://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Stata license: Single-user , expiring 14 Jan 2021
Serial number: 301609276424
Licensed to: Ellen Schenk
University of Delaware

Notes:
1. Unicode is supported; see help unicode\_advice.
```

Results window:
Non-graphic
output

```
Command window:
Where commands go


/Volumes/GoogleDrive/Shared drives/SNAP_Elderly_Caseloads/Current Project Programs and Files/Data/Collapsed Data
```

Variables window: right now empty because we do not have any loaded data. When you have a data set open, the variable names will show in the window

Name	Label
------	-------

Properties

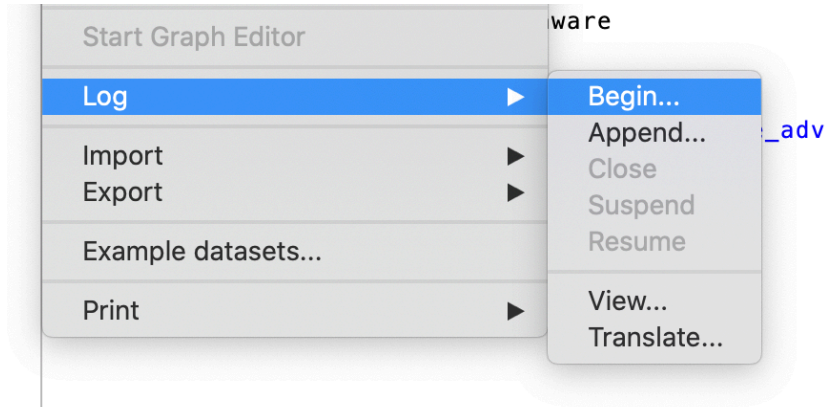
- Variables
 - Name
 - Label
 - Type
 - Format
 - Value label
 - Notes
- Data
 - Frame default
 - Filename
 - Label
 - Notes
 - Variables 0



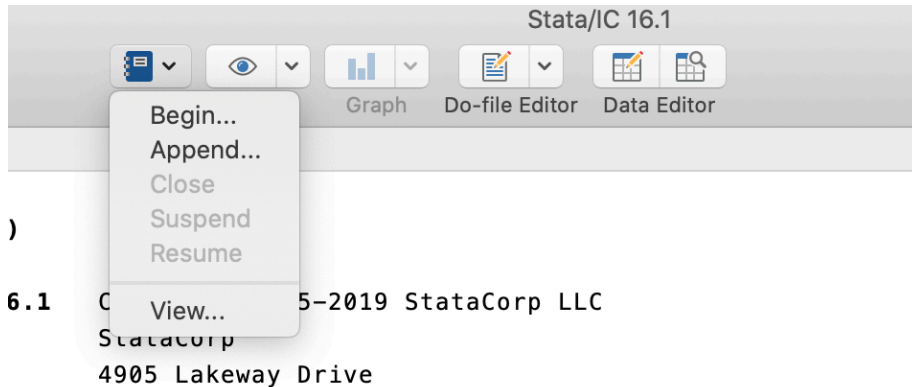
If you want to record anything that you do in a STATA session so that you can look at results or commands later, you should open a log-file. A log-file is simply a record of all the commands you enter into STATA and the output from those commands (kind of like a diary). The key is to make sure you have a log file open at the beginning of a STATA session, and to close it once you have finished, and before you close STATA. The following two slides offer 3 different ways to start a log file.

Log Files

1. File -> Log -> Begin



2. At top of Stata Interface



Opening a Log File

Opening a Log File (continued)

3. Type into Stata command window

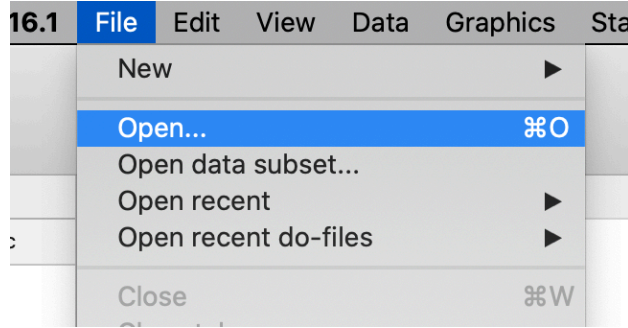
```
Command
log using "/Users/ellen_schenk/Documents/IPA Stata/Workshop_1.log", replace
```

Working directory

Name of log file

Replaces file if it already exists

1. File -> Open -> navigate to .dta file



Variables	
Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu. in.)
gear_ratio	Gear Ratio
foreign	Car type

Opening a Data File

After opening the data set, we can see that the variable window now shows the list of variables, with the variable name and label.

NOTE: Anything highlighted in yellow you should type into the command window.

describe

```
. describe
```

```
Contains data from /Applications/Stata/ado/base/a/auto.dta
  obs:                74          1978 Automobile Data
  vars:                12          13 Apr 2018 17:45
                                   (_dta has notes)
```

variable name	storage type	display format	value label	variable label
make	str18	%-18s		Make and Model
price	int	%8.0gc		Price
mpg	int	%8.0g		Mileage (mpg)
rep78	int	%8.0g		Repair Record 1978
headroom	float	%6.1f		Headroom (in.)
trunk	int	%8.0g		Trunk space (cu. ft.)
weight	int	%8.0gc		Weight (lbs.)
length	int	%8.0g		Length (in.)
turn	int	%8.0g		Turn Circle (ft.)
displacement	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type

```
Sorted by: foreign
```

Basic
Commands:
Describe

At the top of the output you will see some overall features of the file, including the number of variables. Below that you will see a list of every variable, including the variable name, the “storage type” (byte, float, int, etc.) and the variable label. If you see – more- at the bottom of your screen, you need to press the space bar to continue scrolling.

Most Stata commands can be abbreviated. You can execute the describe command by typing `des` or `d`

*Basic
Commands:
Describe*

Stata is a language; like any language it has rules. For each command, there are various options that you can add on to modify the command (or just the output, what is printed to the screen). Commands always come before options and are separated by commas.

describe, short

```
. d, s
```

```
Contains data from /Applications/Stata/ado/base/a/auto.dta
```

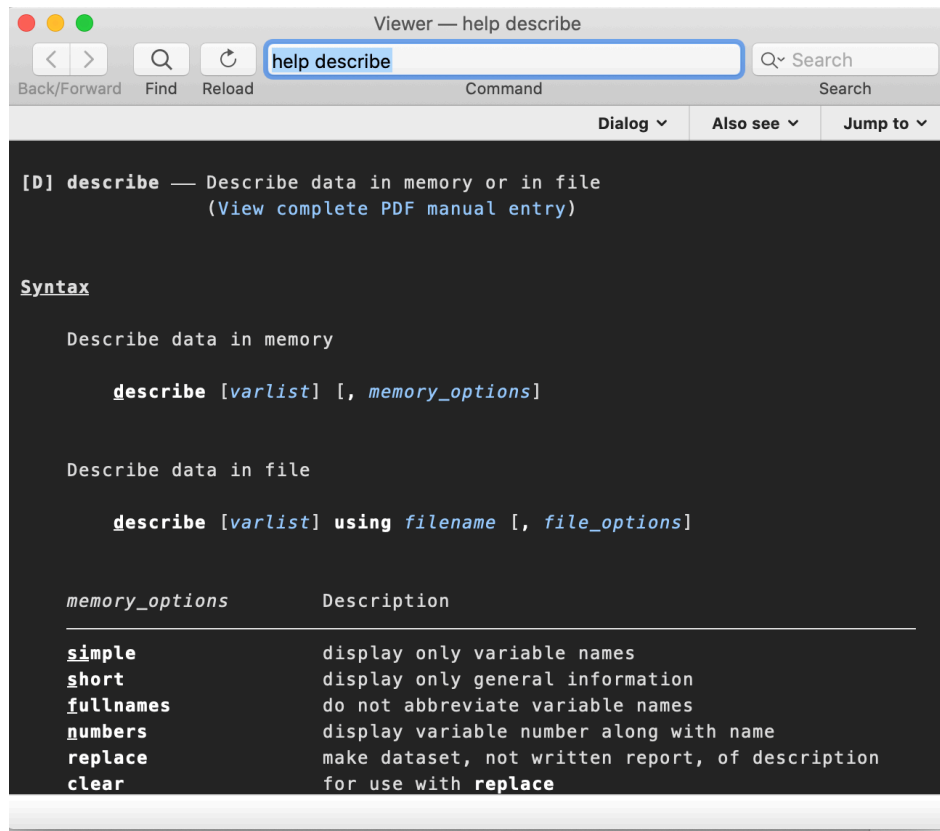
```
  obs:                74                1978 Automobile Data
```

```
  vars:                12                13 Apr 2018 17:45
```

```
Sorted by: foreign
```

*Basic
Commands:
Describe*

If you want more details about this Stata command, you can get help by typing `help describe`



The screenshot shows a Stata help viewer window titled "Viewer — help describe". The search bar at the top contains "help describe". The main content area displays the following text:

```
[D] describe — Describe data in memory or in file
      (View complete PDF manual entry)

Syntax

Describe data in memory

      describe [varlist] [, memory_options]

Describe data in file

      describe [varlist] using filename [, file_options]

memory_options      Description
-----
simple               display only variable names
short               display only general information
fullnames           do not abbreviate variable names
numbers             display variable number along with name
replace             make dataset, not written report, of description
clear               for use with replace
```


Basic
Commands:
Help

To learn more about the variables and the organization of the data, use the **browse** command.

The screenshot shows the Stata Data Editor (Browse) window for the file 'auto.dta'. The main window displays a table of car data with columns for 'make', 'price', 'mpg', 'rep78', 'headroom', 'trunk', 'weight', 'length', 'turn', 'displacement', 'gear_ratio', and 'foreign'. The 'make' column is highlighted, and the first row is selected. The sidebar on the right shows the 'Variables' list with checkboxes for 'make', 'price', 'mpg', 'rep78', 'headroom', and 'trunk'. Below the variables list, the 'Properties' section shows details for the selected variable 'make', including its name, label, type, format, value label, and notes. The 'Data' section shows the frame, filename, label, and notes for the dataset.

	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9	Buick Riviera	10,372	16	3	3.5	17	3,880	207	43	231	2.93	Domestic
10	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic
11	Cad. Deville	11,385	14	3	4.0	20	4,330	221	44	425	2.28	Domestic
12	Cad. Eldorado	14,500	14	2	3.5	16	3,900	204	43	350	2.19	Domestic
13	Cad. Seville	15,906	21	3	3.0	13	4,290	204	45	350	2.24	Domestic
14	Chev. Chevette	3,299	29	3	2.5	9	2,110	163	34	231	2.93	Domestic
15	Chev. Impala	5,705	16	4	4.0	20	3,690	212	43	250	2.56	Domestic
16	Chev. Malibu	4,504	22	3	3.5	17	3,180	193	31	200	2.73	Domestic
17	Chev. Monte Carlo	5,104	22	2	2.0	16	3,220	200	41	200	2.73	Domestic
18	Chev. Monza	3,667	24	2	2.0	7	2,750	179	40	151	2.73	Domestic
19	Chev. Nova	3,955	19	3	3.5	13	3,430	197	43	250	2.56	Domestic
20	Dodge Colt	3,984	30	5	2.0	8	2,120	163	35	98	3.54	Domestic
21	Dodge Diplomat	4,010	18	2	4.0	17	3,600	206	46	318	2.47	Domestic

Vars: 12 Order: Dataset Obs: 74 Length: 18 Filter: Off



This command directs you to the data viewer inside Stata where the data appears; it looks like an Excel spreadsheet.

You should note the following:

- Each observation (in this case each car for which there is recorded information) appears on a separate row of the spreadsheet.
- Each variable appears in a separate column.

`count`

Count tells you how many observations there are in your data set. In our data set, we have 74 observations.

```
. count  
74
```

*Basic
Commands:
Count*

20

This is a good time to introduce the very important “if” command. You will often want to limit your analysis to some subsample of the population: women, children, or people with a smartphone. Let’s browse the data listing only domestic cars: `browse if foreign==0`. Note that you need two equal signs for this kind of logic statement.

Data Editor (Browse) — auto.dta

	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9	Buick Riviera	10,372	16	3	3.5	17	3,880	207	43	231	2.93	Domestic
10	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic
11	Cad. Deville	11,385	14	3	4.0	20	4,330	221	44	425	2.28	Domestic
12	Cad. Eldorado	14,500	14	2	3.5	16	3,900	204	43	350	2.19	Domestic
13	Cad. Seville	15,906	21	3	3.0	13	4,290	204	45	350	2.24	Domestic
14	Chev. Chevette	3,299	29	3	2.5	9	2,110	163	34	231	2.93	Domestic
15	Chev. Impala	5,705	16	4	4.0	20	3,690	212	43	250	2.56	Domestic
16	Chev. Malibu	4,504	22	3	3.5	17	3,180	193	31	200	2.73	Domestic
17	Chev. Monte Carlo	5,104	22	2	2.0	16	3,220	200	41	200	2.73	Domestic
18	Chev. Monza	3,667	24	2	2.0	7	2,750	179	40	151	2.73	Domestic
19	Chev. Nova	3,955	19	3	3.5	13	3,430	197	43	250	2.56	Domestic
20	Dodge Colt	3,984	30	5	2.0	8	2,120	163	35	98	3.54	Domestic
21	Dodge Diplomat	4,010	18	2	4.0	17	3,600	206	46	318	2.47	Domestic

Vars: 12 Order: Dataset Obs: 52 of 74

Basic Commands: “If” statements

Here, we are only looking at domestic cars.

Tabulate: tabulate produces one-way and two-way tables of frequency counts and percentage distributions. We can see what tab does easily by example. Suppose we want to know the percentage domestic cars: `tab foreign`

```
. tab foreign
```

Car type	Freq.	Percent	Cum.
Domestic	52	70.27	70.27
Foreign	22	29.73	100.00
Total	74	100.00	

*Basic
Commands:
Tabulate*

Here, we can see that about 30% of cars are foreign and 70% are domestic.

We can get the distribution of cars with prices greater than \$5,000 (remember, 1978 data!): `tab foreign if price>5000`

```
. tab foreign if price>5000
```

Car type	Freq.	Percent	Cum.
Domestic	23	62.16	62.16
Foreign	14	37.84	100.00
Total	37	100.00	

```
.
```

*Basic
Commands:
Tabulate*

We can also look at foreign cars with prices greater than \$5000 and miles per gallon less than or equal to 30: `tab foreign if price>5000 & mpg<=30`

```
tab foreign if price>5000 & mpg<=30
```

Car type	Freq.	Percent	Cum.
Domestic	23	63.89	63.89
Foreign	13	36.11	100.00
Total	36	100.00	

*Basic
Commands:
Tabulate*

Summarize: provides you with more information about your data

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

```
.
```

*Basic
Commands:
Summarize*

25

Tells us number of observations (some may be missing, that's why we have less than 74), mean, standard deviation, min, max. If 0, string variable (i.e. make is a string variable)

Remember options that can follow commands? Try **summarize, detail**

Price				
Percentiles		Smallest		
1%	3291	3291		
5%	3748	3299		
10%	3895	3667	Obs	74
25%	4195	3748	Sum of Wgt.	74
50%	5006.5		Mean	6165.257
		Largest	Std. Dev.	2949.496
75%	6342	13466		
90%	11385	13594	Variance	8699526
95%	13466	14500	Skewness	1.653434
99%	15906	15906	Kurtosis	4.819188

This tells us more detailed statistics about each variable.

*Basic
Commands:
Summarize*

Codebook gives us more information about our data. Try `codebook foreign`

```
foreign                                     Car type

      type: numeric (byte)
      label: origin

      range: [0,1]                          units: 1
unique values: 2                            missing .: 0/74

      tabulation: Freq.  Numeric  Label
                   52      0      Domestic
                   22      1      Foreign
```

*Basic
Commands:
Codebook*

*Remember to
close your log
file!*

log close

To exit Stata:

`clear` -> closes open
data file

`exit` -> exit Stata

Logic Command

Symbol

And

&

Or

| (vertical bar, on same key
as "\")

Not equal to

!=

Less than/less than or
equal to

<, <=

Greater than, greater than
or equal to

>, >=

***Stata Logical
Comments***

- STATA help command
- Stata website – www.stata.com
- Do a Google search
- Email me!
- YouTube tutorials for common Stata commands

*Get Help
with Stata*

Thanks!

Does anyone have any questions?

ellends@udel.edu

Next Stata Workshop:
Monday, November 2- Data
Management