Problem Set 05 – STATA

Description: This problem set is designed to guide you through collecting information from the Census website's American FactFinder tool, and then estimating and interpreting an OLS regression using that data.

Part 1:
On D2L there is a file called "city_data.csv" that you should download. This file includes observations on 40 major cities in the United States.

a) Pick a city that is not already in the list. Make sure that the city you pick has a 2015 population greater than 150,000 people.
b) Go to the American Factfinder and find the following eight variables:
    1) Population in 2015
    2) Median Income in 2015
    3) Percent of the Population with a BA degree or higher in 2015
    4) Percent of the Population that is Self-Employed in 2015
    5) Population in 2000
    6) Median Income in 2000
    7) Percent of the Population with a BA degree or higher in 2000
    8) Percent of the Population that is Self-Employed in 2000
b) Create the percent change in population from 2000 to 2015.
c) Add you city observations to the csv file and save the csv file.

Part 2:
Go to a computer with STATA (Library and/or Union). Let Population be the "dependent variable" of your analysis. To shorten the notation, we will use the following variable names:

Population YYYY : popYY
Median Income YYYY : incYY
% w/ BA YYYY: pctbaYY
% Self Employed YYYY: selfYY
Change in Population : chgpop

a) What is the mean, max, and min for your data?
b) Print a scatter plot of the dependent variable and independent variable.
c) Generate "logged" versions of pop15, pop00, inc15, inc00
    — call them lnpop15, lnpop00, lninc15, lninc00
d) Regress lnpop15 on lninc15 (using robust standard errors)
    a. Interpret the coefficient.
    b. What is the standard error and t-statistic?
e) Regress lnpop15 on lninc15, pctba15, and self15
    a. Interpret the coefficients on lninc15 and pctba15.
    b. Are the coefficients statistically significant?
f) Regress chgpop on lninc00 pctba00 self00
    a. Interpret the coefficients.
    b. Do these results agree with the regression in (e)?

You do not need to print out all of your results.
You must print out your scatter plot, but other answers can be written by hand.

Problem Set 05 – STATA

Guide to Problem Set

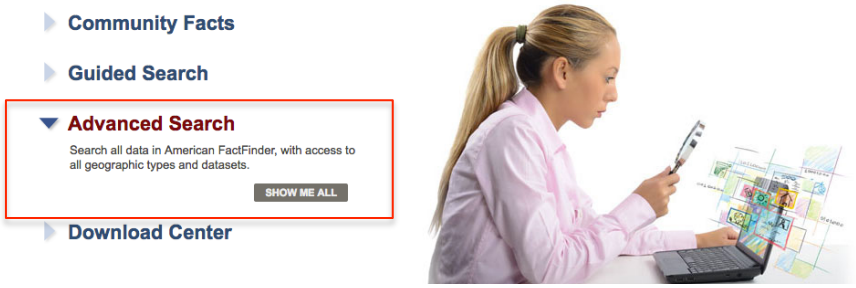
Do Not Include Pages from the Guide when Turning in your problem set!

Part 1:

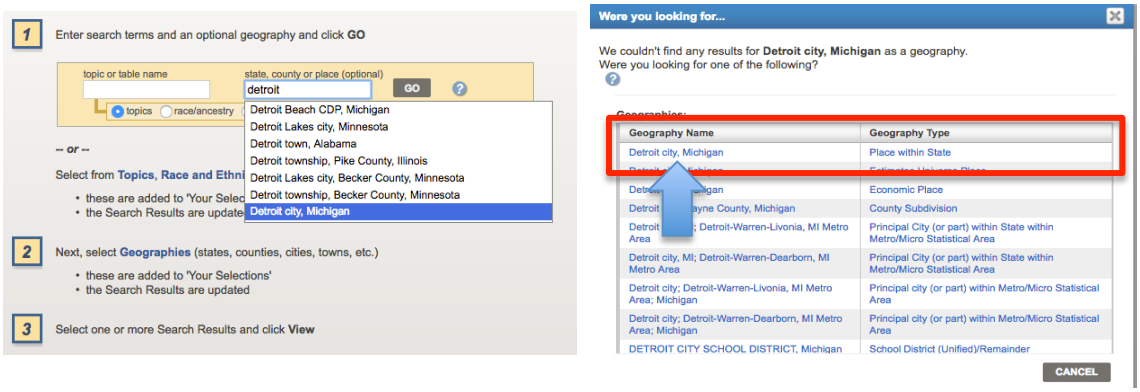
Pick a city with population greater than 150,000 people from the list on Wikipedia that is not already used in the file on D2L:
https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

Go to the American FactFinder website:
https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

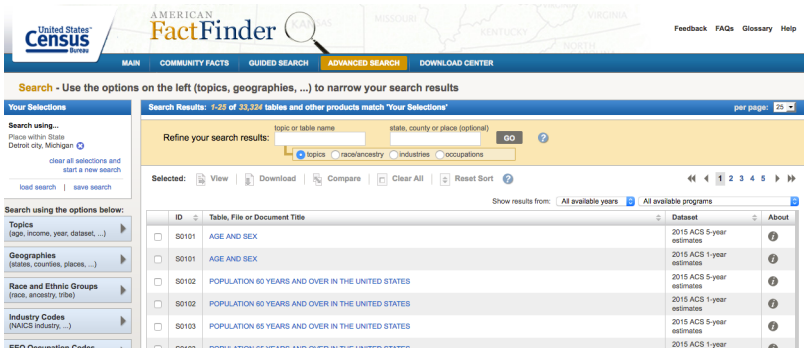
Click on Advanced Search, click "Show Me All"



Suppose the city you choose is Detroit (note: Detroit is already in the data!); type the name of the city in the search bar. Make sure that you choose "*Place within a State*" if you are asked by clicking on the city name in the left column.

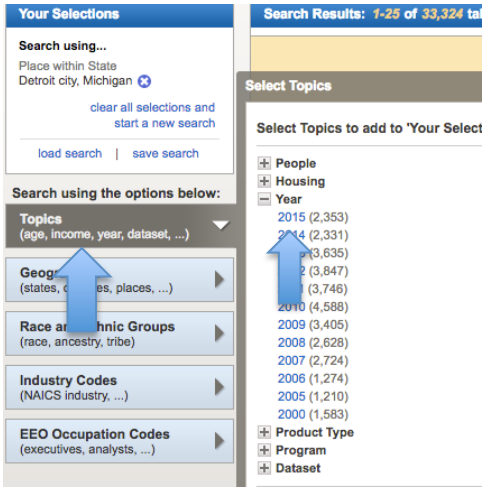


Now you are on the Advanced Search page:



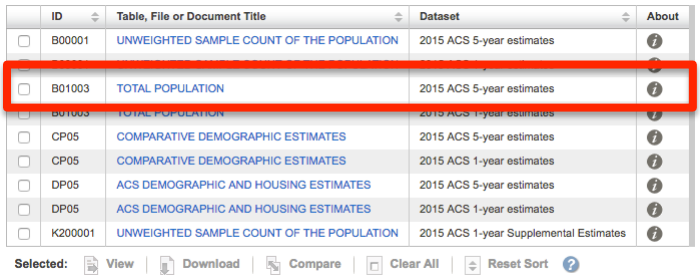
Problem Set 05 – STATA

Now click on the Topic tab to look at our options.
Click on Years; choose 2015.

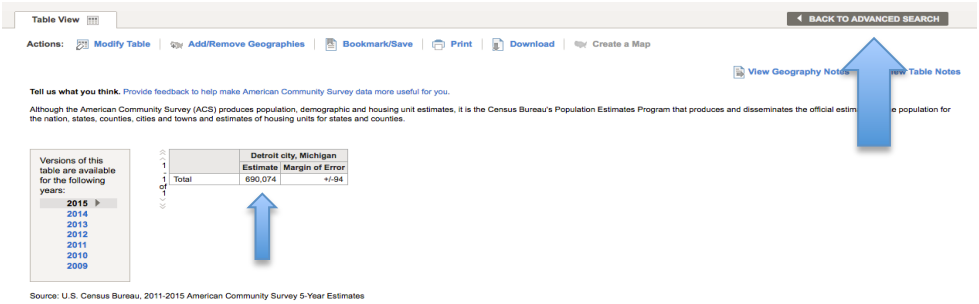


Now, switch to "People," click "Basic Counts" and choose "Population Total."
After choosing this, close the box.

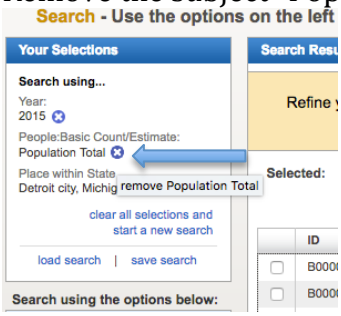
Now you have datasets available to look at.
Choose "Total Population" using "2015 ACS 5-Year"



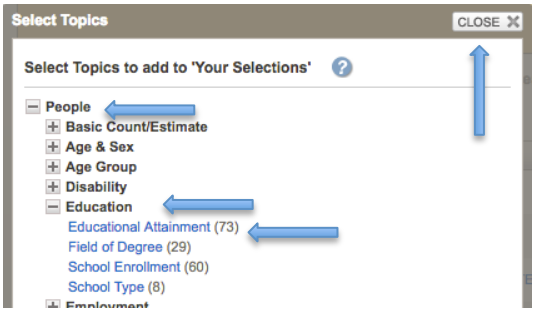
Once you choose this, you should get a table with an estimate of the population. Copy the number down somewhere (you'll need it to put in the "city_data.csv" file. Then click "Back to Adv Search"

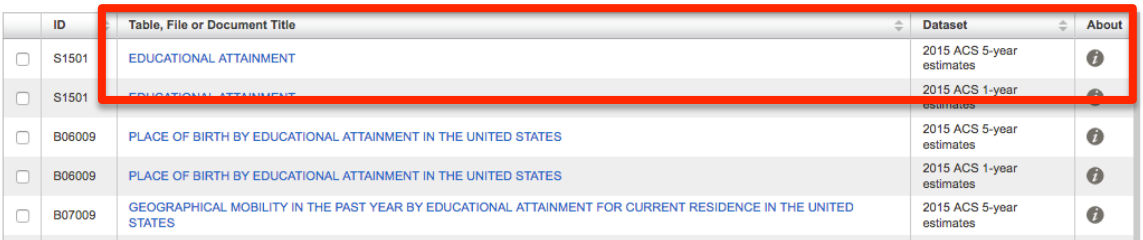


Remove the subject "Population Total" to pick a new topic.



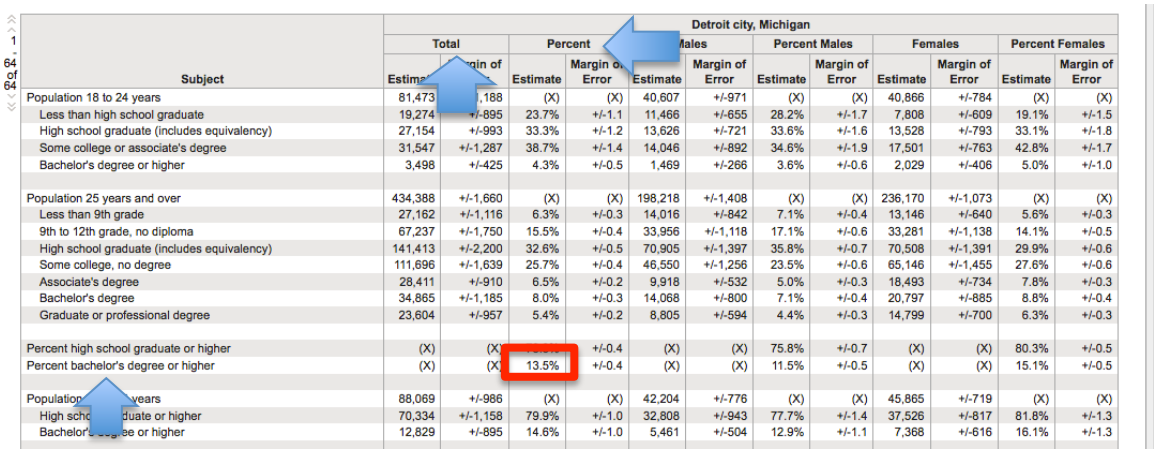
Click on "Topic" → "People" → "Education" then choose "Education Attainment" then close. Then choose the "2015 ACS 5 Year" option.





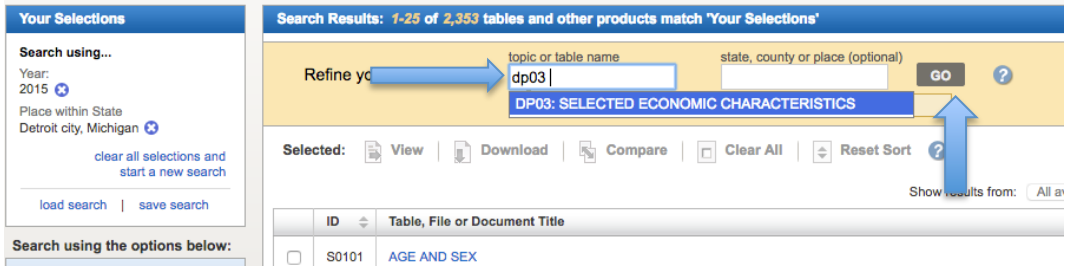
Problem Set 05 – STATA

Find the percent of the population with a BA or higher degree from the table.



Subject	Detroit city, Michigan											
	Total		Percent		Males		Percent Males		Females		Percent Females	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error
Population 18 to 24 years	81,473	,188	(X)	(X)	40,607	+/-971	(X)	(X)	40,866	+/-784	(X)	(X)
Less than high school graduate	19,274	+/-895	23.7%	+/-1.1	11,466	+/-655	28.2%	+/-1.7	7,808	+/-609	19.1%	+/-1.5
High school graduate (includes equivalency)	27,154	+/-993	33.3%	+/-1.2	13,626	+/-721	33.6%	+/-1.6	13,528	+/-793	33.1%	+/-1.8
Some college or associate's degree	31,547	+/-1,287	38.7%	+/-1.4	14,046	+/-892	34.6%	+/-1.9	17,501	+/-763	42.8%	+/-1.7
Bachelor's degree or higher	3,498	+/-425	4.3%	+/-0.5	1,469	+/-266	3.6%	+/-0.6	2,029	+/-406	5.0%	+/-1.0
Population 25 years and over	434,388	+/-1,660	(X)	(X)	198,218	+/-1,408	(X)	(X)	236,170	+/-1,073	(X)	(X)
Less than 9th grade	27,162	+/-1,116	6.3%	+/-0.3	14,016	+/-842	7.1%	+/-0.4	13,146	+/-640	5.6%	+/-0.3
9th to 12th grade, no diploma	67,237	+/-1,750	15.5%	+/-0.4	33,956	+/-1,118	17.1%	+/-0.6	33,281	+/-1,138	14.1%	+/-0.5
High school graduate (includes equivalency)	141,413	+/-2,200	32.6%	+/-0.5	70,905	+/-1,397	35.8%	+/-0.7	70,508	+/-1,391	29.9%	+/-0.6
Some college, no degree	111,696	+/-1,639	25.7%	+/-0.4	46,550	+/-1,256	23.5%	+/-0.6	65,146	+/-1,455	27.6%	+/-0.6
Associate's degree	28,411	+/-910	6.5%	+/-0.2	9,918	+/-532	5.0%	+/-0.3	18,493	+/-734	7.8%	+/-0.3
Bachelor's degree	34,865	+/-1,185	8.0%	+/-0.3	14,068	+/-800	7.1%	+/-0.4	20,797	+/-885	8.8%	+/-0.4
Graduate or professional degree	23,604	+/-957	5.4%	+/-0.2	8,805	+/-594	4.4%	+/-0.3	14,799	+/-700	6.3%	+/-0.3
Percent high school graduate or higher	(X)	(X)		+/-0.4	(X)	(X)	75.8%	+/-0.7	(X)	(X)	80.3%	+/-0.5
Percent bachelor's degree or higher	(X)	(X)	13.5%	+/-0.4	(X)	(X)	11.5%	+/-0.5	(X)	(X)	15.1%	+/-0.5
Population years	88,069	+/-986	(X)	(X)	42,204	+/-776	(X)	(X)	45,865	+/-719	(X)	(X)
High sch duate or higher	70,334	+/-1,158	79.9%	+/-1.0	32,808	+/-943	77.7%	+/-1.4	37,526	+/-817	81.8%	+/-1.3
Bachelor' ee or higher	12,829	+/-895	14.6%	+/-1.0	5,461	+/-504	12.9%	+/-1.1	7,368	+/-616	16.1%	+/-1.3

Again, return to the "Adv Search" page and clear the topic.
Instead of going to the topic, type "DP03" in the table search bar, choose the suggestion, and then click "Go".



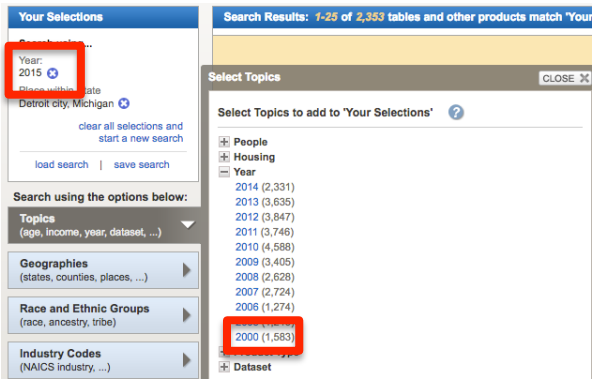
There will be only two options; choose the "2015 ACS 5 Year" option as before.
The table you will open is longer than the previous tables, so you must scroll down to find the variables you are interested in.
First find self`15:

Subject	Detroit city, Michigan			
	Estimate	Margin of Error	Percent	Percent Margin of Error
CLASS OF WORKER				
Civilian employed population 16 years and over	213,591	+/-3,130	21	(X)
Private wage and salary workers	180,304	+/-3,032	84.4%	+/-0.6
Government workers	24,406	+/-1,026		+/-0.5
Self-employed in own not incorporated business workers	8,643	+/-673	4.0%	+/-0.3
Unpaid family workers	238	+/-90		+/-0.1

Then find inc15 — Be careful that you find "Median Earnings for Workers" — there are other 'median income' variables but we are not using them!

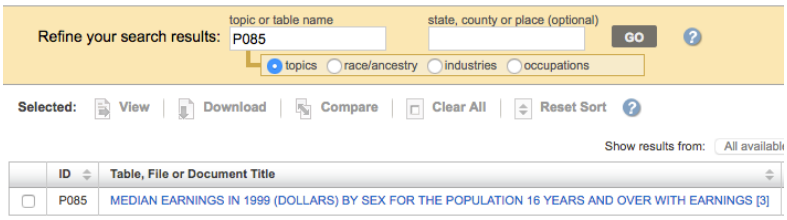
Subject	Detroit city, Michigan			
	Estimate	Margin of Error	Percent	Percen
Per capita income (dollars)	15,038	+/-239	(X)	
Nonfamily households	111,474	+/-1,664	111,474	
Median nonfamily income (dollars)	18,513	+/-446	(X)	
Mean nonfamily income (dollars)	28,255	+/-773	(X)	
Median earnings for workers (dollars)	19,867	+/-389	(X)	
		+/-1,185	(X)	
Median earnings for female full-time, year-round workers (dollars)	30,856	+/-407	(X)	
HEALTH INSURANCE COVERAGE				

Now, **remove** the year 2015 and then **select** the year 2000.



You will see the following tables; you'll need to scroll to find the variables you want.
"Profile of General Demographic Characteristics" — Population
"Selected Social Characteristics" — Percent with BA Degree
"Selected Economic Characteristics" — Percent Self Employed

For some reason, "Median Income for Workers" is not in the table above, so search for the following table like before: P085

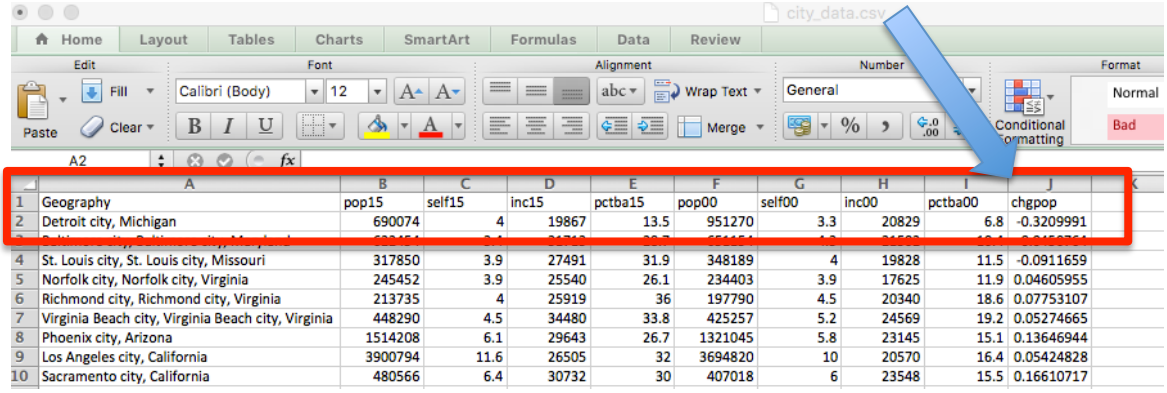


The variable is in this table; you won't need to scroll.

Problem Set 05 – STATA

Now that you have all the data for your city, add it to the csv file.
You can open a csv file in MS Excel.

Note: for "chgpop," find the percentage change in population from 2000 to 2015.
You can do this by the following calculation: ln[pop15] – ln[pop00]



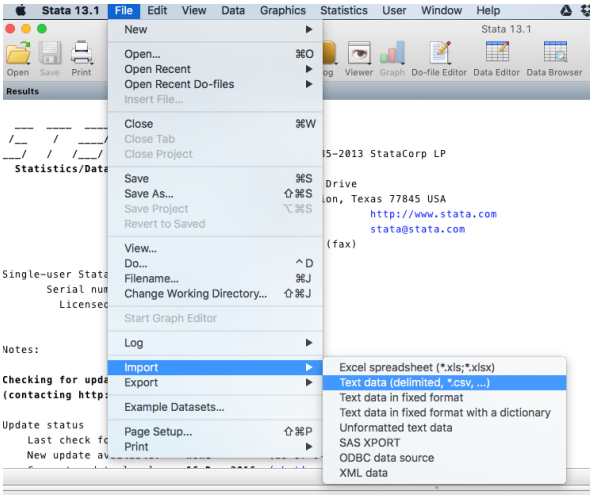
	A	B	C	D	E	F	G	H	I	J
1	Geography	pop15	self15	inc15	pctba15	pop00	self00	inc00	pctba00	chgpop
2	Detroit city, Michigan	690074	4	19867	13.5	951270	3.3	20829	6.8	-0.3209991
4	St. Louis city, St. Louis city, Missouri	317850	3.9	27491	31.9	348189	4	19828	11.5	-0.0911659
5	Norfolk city, Norfolk city, Virginia	245452	3.9	25540	26.1	234403	3.9	17625	11.9	0.04605955
6	Richmond city, Richmond city, Virginia	213735	4	25919	36	197790	4.5	20340	18.6	0.07753107
7	Virginia Beach city, Virginia Beach city, Virginia	448290	4.5	34480	33.8	425257	5.2	24569	19.2	0.05274665
8	Phoenix city, Arizona	1514208	6.1	29643	26.7	1321045	5.8	23145	15.1	0.13646944
9	Los Angeles city, California	3900794	11.6	26505	32	3694820	10	20570	16.4	0.05424828
10	Sacramento city, California	480566	6.4	30732	30	407018	6	23548	15.5	0.16610717

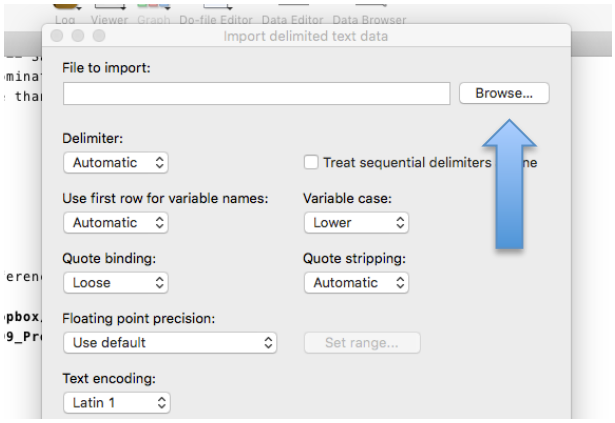
Note: when you go to save, Excel will give you two pop-up warnings.
* For the first one, click "Continue"
* For the second, click "Don't Save." — Excel wants you to change the file format, but you can ignore this and stick with csv.

Part 2:

Two do the estimation you will need to have completed you data set of 41 cities and have STATA open.

First, import your data. You will need to choose "Browse…" to find the file on your computer. STATA will let you preview the data, but click OK if it looks good.

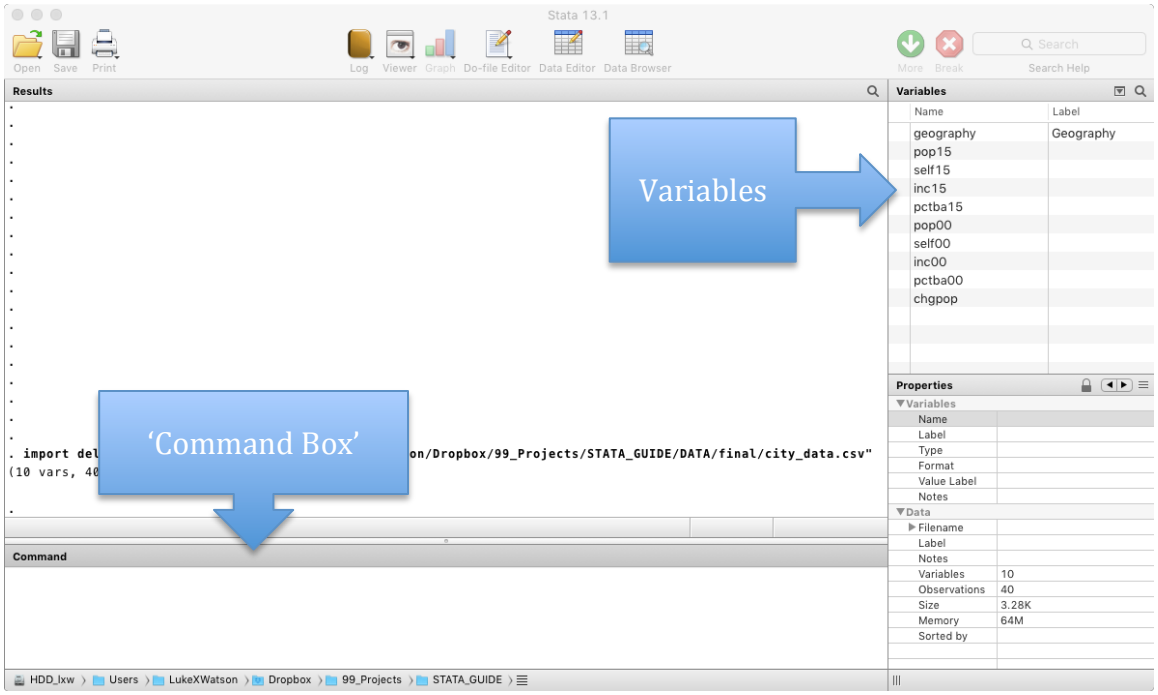




Problem Set 05 – STATA

You are now ready to do the estimation.

There are several ways of getting STATA to do what you want, but I am going to recommend for now that you operate STATA through "interactive commands" by typing commands into the 'command box'



The following commands are all that you need to complete **Part 2:**

```
sum
twoway scatter yaxisvar xaxisvar
gen lnvar = ln(oldvar)
reg depvar indepvar1 indepvar2 indepvar3, robust
```

The following commands you may find interesting:

```
twoway lfitci yaxisvar xaxisvar || scatter yaxisvar xaxisvar
predict fitted
predict uhat, resid
rvpplot
rvfplot indepvar
sum uhat, detail
list geography if abs(uhat)>1
```

Find more commands at the following website:
http://homepages.rpi.edu/~simonk/pdf/UsefulStataCommands.pdf

Problem Set 05 – STATA

How to read the STATA regression output: