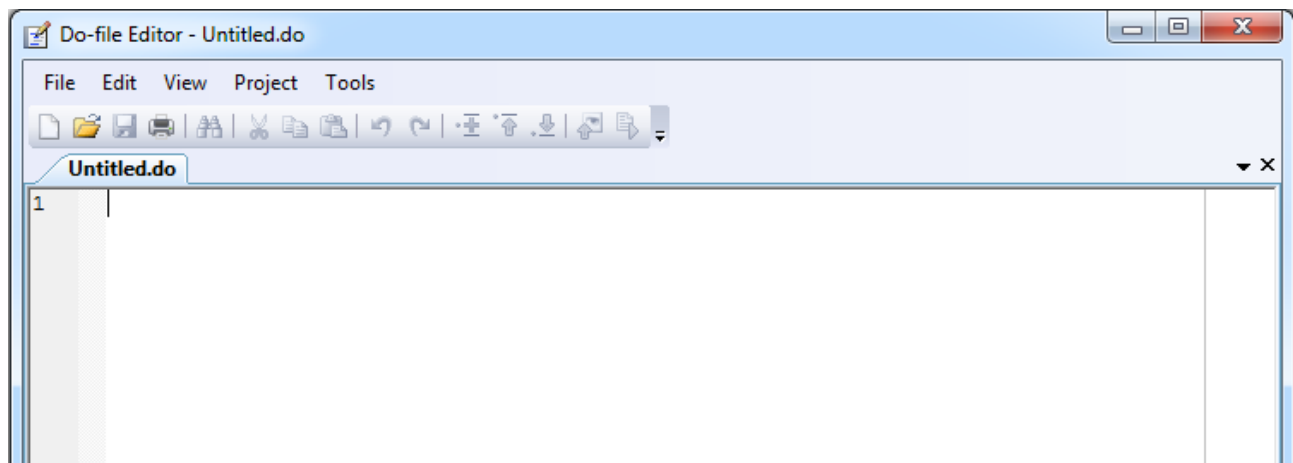
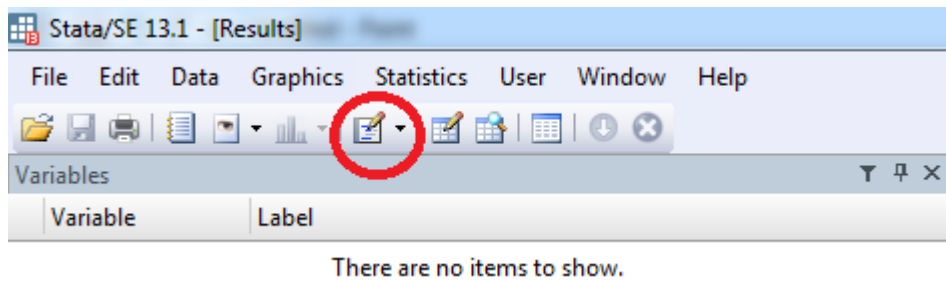


STATA GUIDE TO DATA ANALYSIS FOR LATINO NATIONAL SURVEY

1. Open the software program called Stata
2. If you want the interface to match what we are using in class
 - a. Click on Edit → Preferences → General Preferences
 - b. Under Color scheme choose Classic
 - c. You can set the color scheme to whatever you prefer
3. You always want to record and document all the steps you have taken during data analysis using your .do file editor. This will open a new window that is essentially a basic text document where you can type in, and save, all the commands you use.
4. Click on the icon of the .do file editor and a new window will appear



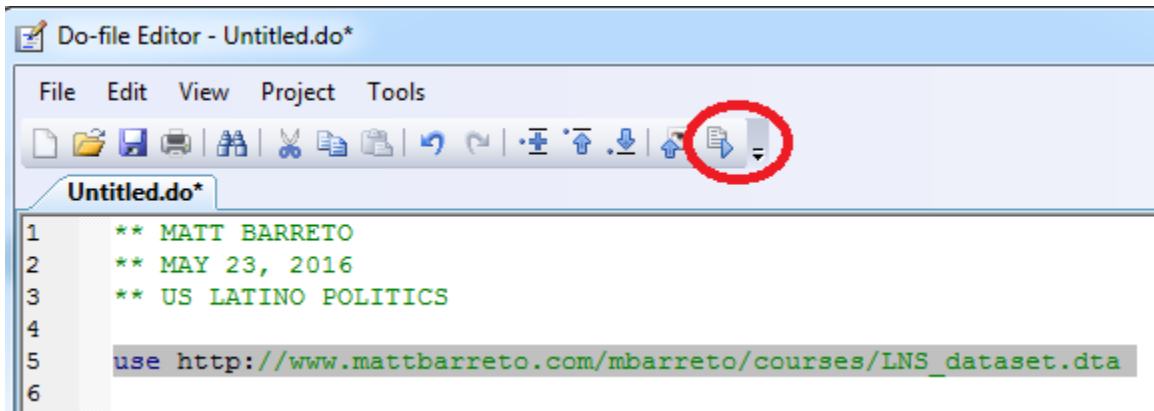
At the top of your .do file, type in your name, date, course title, project title.

NOTE: Start with an asterisk * if you are writing a note to yourself and not an actual command

**** YOUR NAME**
**** MAY 23, 2016**
**** US LATINO POLITICS**

IMPORTANT: Now click File → Save and save your .do file somewhere that you can find it

5. Now we want to actually open the Latino National Survey (LNS) dataset in our Stata window. To do this type the following command into your .do file editor:
 - a. use http://www.mattbarreto.com/mbarreto/courses/LNS_dataset.dta
 - b. Highlight the line that begins with use and then click execute, this will open the dataset



6. In the main Stata window the dataset will open. You will see a list of all the questions asked on the survey in the section titled 'Variable' and 'Label'. The main Stata window should show you that you executed a .do file, then it should show the actual command you executed (use) and if there are no errors, it will then say "end of do-file"

Lesson 1: tabulate basic demographic data

7. The main thing Stata will help you do is examine the percentage results of the survey. The basic way we do this is using the tabulate command, which is abbreviated to 'tab'
 - a. In your .do file, type in the following commands and then highlight ONLY those lines and execute
 - i. tab LANGPREF
 - ii. tab IDPREF
 - iii. tab ANCESTRY
 - iv. tab BORNUS

If you did it correctly – you should get the following output on your screen:

```

. tab LANGPREF

S3. Would |
you prefer |
that I    |
speak in  |
English or |
Spanish?  |
          |
          | Freq.    Percent    Cum.
-----+-----
English   |    3,291    38.12    38.12
Spanish   |    5,343    61.88   100.00
-----+-----
Total     |    8,634   100.00

```

```
. tab IDPREF
```

S4. The most frequently used terms to describe persons of Latin American descent	Freq.	Percent	Cum.
Hispanic	3,030	35.53	35.53
Latino	1,115	13.08	48.61
Either is acceptable	2,817	33.04	81.65
Don't care	1,565	18.35	100.00
Total	8,527	100.00	

```
. tab ANCESTRY
```

B4. Families of ANSWERFROM(AQS4) origin or background in the United States come	Freq.	Percent	Cum.
Argentina	28	0.33	0.33
Bolivia	36	0.42	0.75
Chile	17	0.20	0.94
Colombia	139	1.62	2.56
Costa Rica	32	0.37	2.94
Cuba	420	4.89	7.83
Dominican Republic	335	3.90	11.74
Ecuador	103	1.20	12.94
El Salvador	407	4.74	17.68
Guatemala	149	1.74	19.41
Honduras	87	1.01	20.43
Mexico	5,704	66.47	86.90
Nicaragua	51	0.59	87.50
Panama	14	0.16	87.66
Paraguay	3	0.03	87.69
Peru	65	0.76	88.45
Puerto Rico	822	9.58	98.03
Spain	105	1.22	99.25
Uruguay	7	0.08	99.34
Venezuela	27	0.31	99.65
Don't Know	30	0.35	100.00
Total	8,581	100.00	

```
. tab BORNUS
```

B5. Were you born in the mainland United States, Puerto Rico or some other count	Freq.	Percent	Cum.
Mainland US	2,450	28.38	28.38
Puerto Rico	467	5.41	33.79
Some other country	5,717	66.21	100.00
Total	8,634	100.00	

```
. end of do-file
```

8. Some variables have too many response option categories for making basic tables. For example let's tab age:
 - a. In your .do file editor type: `tab AGE` (now highlight that and click execute)

9. The variable AGE might be easier to work with if we recode it into age categories. To do this we will generate a new variable called age_cat and then recode it. Never recode the original variable in case you need to come back to it later. Always create a new variable and recode that. In you .do file editor type the following:

```

** New commands to generate new variables and recode
** gen = generate new variable
** recode = recode the answer options
gen age_cat = AGE
recode age_cat 18/29=1 30/39=2 40/49=3 50/59=4 60/97=5
tab age_cat

```

10. Highlight and execute the three lines of code and you should get the following:

```
. tab age_cat
```

age_cat	Freq.	Percent	Cum.
1	2,299	28.24	28.24
2	2,061	25.32	53.56
3	1,591	19.54	73.10
4	1,096	13.46	86.56
5	1,094	13.44	100.00
Total	8,141	100.00	

Lesson 2: exploring the data with cross-tabulations

11. Let's look at the relationship between place of birth and strength of different identities
 - a. `tab AMERICAN BORNUS`
 - b. This is a cross-tab of the total number of people in each response category. You always read a cross-tab from top-to-bottom, within each category. Such as "among those born in Mainland US..."

```

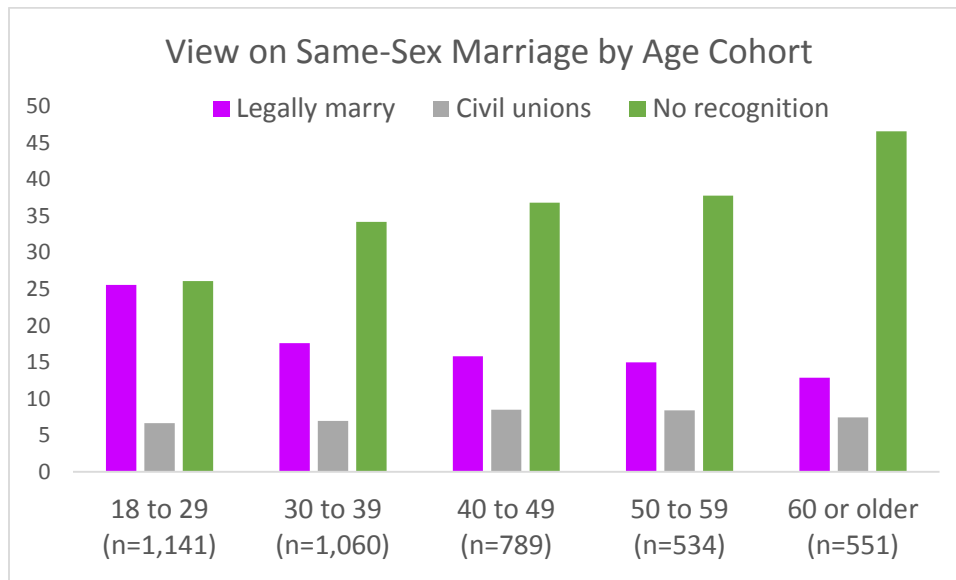
L8. The first one |
    is ...        |
'American.' how  |      B5. Were you born in the
strongly or not  |      mainland United States, Puerto
do you think of  |      Rico or some other count
yourself         |      Mainland  Puerto Ri  Some othe |      Total
-----+-----+-----+-----+-----+
    Not at all   |           76           28       1,137 |      1,241
Not very strongly |          132           52       1,260 |      1,444
Somewhat strongly |           449          127       1,607 |      2,183
    Very strongly |          1,761          249       1,428 |      3,438
-----+-----+-----+-----+
                |          2,418          456       5,432 |      8,306

```


13. Now let's look at the relationship between age and support for same-sex marriage

a. tab SAMESEX age_cat, col nofreq

L21. What is your view about same-sex couples? Should they be permitted to ... (age_cat					Total
	1	2	3	4	5	
Legally marry?	25.59	17.64	15.84	14.98	12.89	18.53
Enter into civil unio	6.66	6.98	8.49	8.43	7.44	7.44
Or should they receiv	26.12	34.25	36.88	37.83	46.64	34.63
No opinion/NA	41.63	41.13	38.78	38.76	33.03	39.41
Total	100.00	100.00	100.00	100.00	100.00	100.00



14. Often times, you want to take a third variable into account, in this case church attendance. To do this we are going to use the same cross-tab as before and add an 'if statement'

a. First let's examine the church attendance variable: tab ATTENDCH

b. tab SAMESEX age_cat if ATTENDCH==1, col nofreq

L21. What is your view about same-sex couples? Should they be permitted to ... (age_cat					Total
	1	2	3	4	5	
Legally marry?	15.15	11.60	9.49	9.57	10.00	11.26
Enter into civil unio	4.55	3.31	5.06	4.26	1.82	3.85
Or should they receiv	40.91	55.25	56.33	56.38	60.91	53.78
No opinion/NA	39.39	29.83	29.11	29.79	27.27	31.11
Total	100.00	100.00	100.00	100.00	100.00	100.00

c. tab SAMESEX age_cat if ATTENDCH==5, col nofreq

L21. What is your view about same-sex couples? Should they be permitted to ... (age_cat					Total
	1	2	3	4	5	
Legally marry?	33.52	26.98	21.59	26.32	25.76	28.07
Enter into civil unio	6.04	11.11	7.95	7.89	12.12	8.55
Or should they receiv	22.53	28.57	31.82	31.58	40.91	29.00
No opinion/NA	37.91	33.33	38.64	34.21	21.21	34.39
Total	100.00	100.00	100.00	100.00	100.00	100.00

What does this tell us about the influence of age?

What does this tell us about the influence of church attendance?

Do both variables have their own independent influence on views towards same sex marriage?

Resources:

UCLA Institute for Digital Research and Education

Stata help, notes, tutorials:

<http://www.ats.ucla.edu/stat/stata/>

Remote access to Stata software via UCLA Social Science computing

<https://computing.sscnet.ucla.edu/labs/remote-access/>

note: need to register remote access account

Remote access to Stata software via UCLA libraries

<http://www.library.ucla.edu/clicc/software>