

# While We Are Waiting...

- If you want to work along with the presentation, all the materials are available on the PRISM website
  - Go to: <http://polisci.osu.edu/prism/luncheons.htm>
  - Download the following zip file onto your desktop
    - *StataIntro\_08.zip*
    - Extract all of the contents of the zip folder to your desktop
      - Double click to open the presentation file: *IntroStata08\_Vfinal.pdf*
      - Double click on Stata to open the program
  - Note: Included in the zip folder
    - Presentation: *IntroToStata08\_Vfinal.pdf*
    - Datasets: *NES04\_VstataIntro08.dta*  
*ICPSR\_08865*
    - Do file: *IntroStata\_V08.do*

PRISM Brownbag:  
An Introduction to 

Dino Christenson & Scott Powell

Ohio State University

January 25<sup>th</sup>, 2008

# Intro to Stata

- I. GUI
- II. Log file
- III. Basic stats
- IV. Data manipulation
- V. Descriptions of variables
- VI. Help files!
- VII. Graphing
- VIII. Do files
- IX. Exporting tables, graphs and data
- X. Importing foreign data
- XI. Closing



# GUI

- Review: Lists commands that have recently been entered
- Results: Show recently obtained results

The screenshot displays the Stata 9.2 interface. The 'Review' window on the left lists the following commands: `use "C:\Users\powel.4\3\D\`, `edit`, and `sum vote`. The 'Results' window on the right shows the Stata logo, version 9.2, copyright information (1984-2007), and contact details for StataCorp. It also displays the license information for a single-user perpetual license, including the serial number 1990543453 and the user 'Department of Political Science, The Ohio State University'. Below this, the command history is shown: `. use "C:\Users\powell.413\Desktop\NE504_V_StataIntro08.dta", clear`, `. edit`, `- preserve`, and `. sum vote`. At the bottom of the Results window, a summary table for the 'vote' variable is displayed:

| variable | Obs | Mean     | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|-----|-----|
| vote     | 811 | .5080148 | .5002443  | 0   | 1   |

# GUI

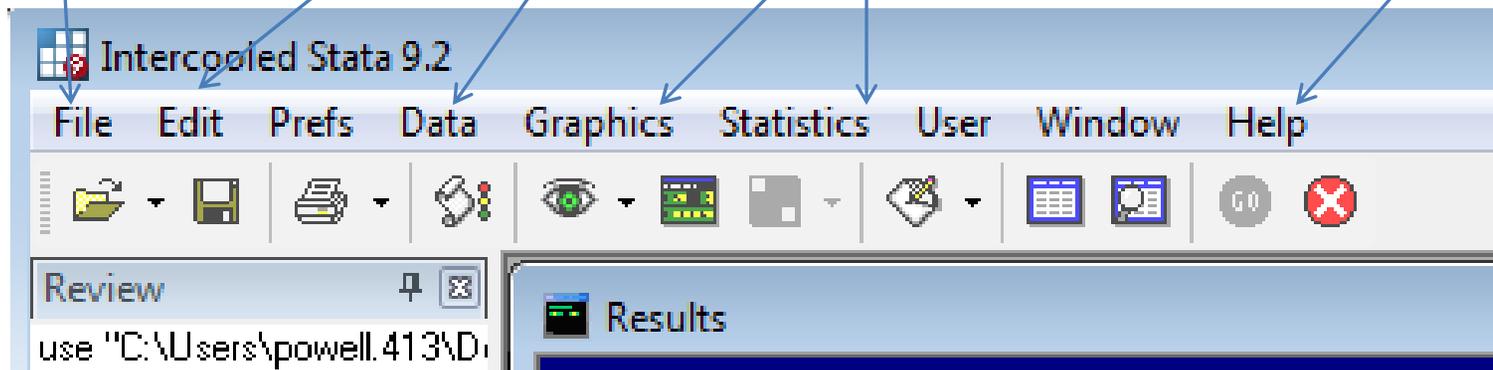
- Variables:  
All the existing variables in your data set
- Command:  
Where commands are entered

The screenshot displays the Stata GUI. On the left, the 'Variables' list shows the following variables: Version, vote, educ, black, hisp, south, marr, age, pray, uspos, styhome, afgwith, abort, ideology, iraq, reptherm, demtherm, female, afghan, pid, and retrosocio. A blue arrow points from the text 'All the existing variables in your data set' to this list. On the right, the Command window shows the following commands: `. use "C:\Users\powell.413\Desktop\NES04_v_StataIntro08.dta", clear`, `. edit`, `-. preserve`, and `. sum vote`. Below the Command window, the Command History shows `sum educ`. A blue arrow points from the text 'Where commands are entered' to the Command window. The bottom status bar shows the current directory as `C:\data`.

| variable | Obs | Mean     | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|-----|-----|
| vote     | 811 | .5080148 | .5002443  | 0   | 1   |

# GUI

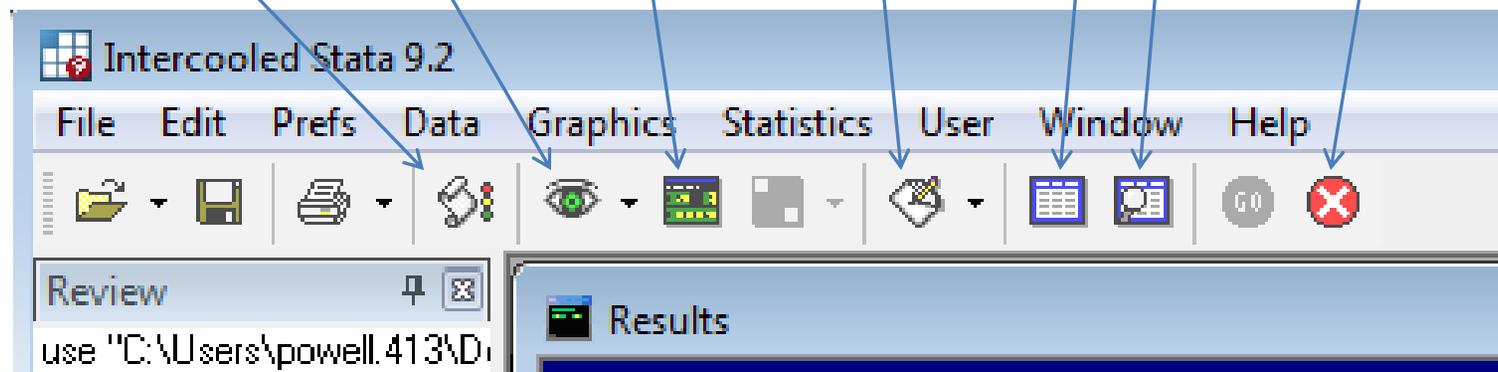
- **File:** More than open and save.
- **Edit:** What you expect
- **Data:** Multiple Avenues for Data Manipulation
- **Graphics:** More to come later
- **Statistics:** Statistical Modeling Options
- **Help:** More to come later



- **Bottom Line:** These menus offer graphical alternatives to directly typing commands into Stata

# GUI

- Begin a new log file
- Bring up the Help Viewer
- Bring up the Results Window
- Begin a new do file
- Edit/View Data
- **STOP!** (The Number Crunching)



# The Log File

- Log or Perish! (or at the very least you might do some crying)
- Log files keep track of everything you do in Stata, both input and output
- However, it does not record when additional windows open up (i.e. graphs, help window, etc.)

```
Viewer (#1) [view C:\Users\powell.413\Desktop\ScottStataIntro.smcl]
Back Refresh Search Help Contents What's New News
Command: view C:\Users\powell.413\Desktop\ScottStataIntro.smcl

. use "C:\Users\powell.413\Desktop\NES04_V_StataIntro08.dta", clear

. *Creating New Variables
. gen left = 1 if ideology <0
(971 missing values generated)

. replace left = 0 if ideology >=0
(971 real changes made)

. recode marr (0=1) (1=0), gen(single)
(1211 differences between marr and single)

. *Matrices
. matrix input mat1 = (1\2\3)
. matrix input mat2 = (1,2,3)
. matrix mat3 = mat1*mat2
. matrix list mat3
symmetric mat3[3,3]
  c1  c2  c3
r1   1
r2   2   4
r3   3   6   9

. *Descriptives: Tabulation
. tab ideology

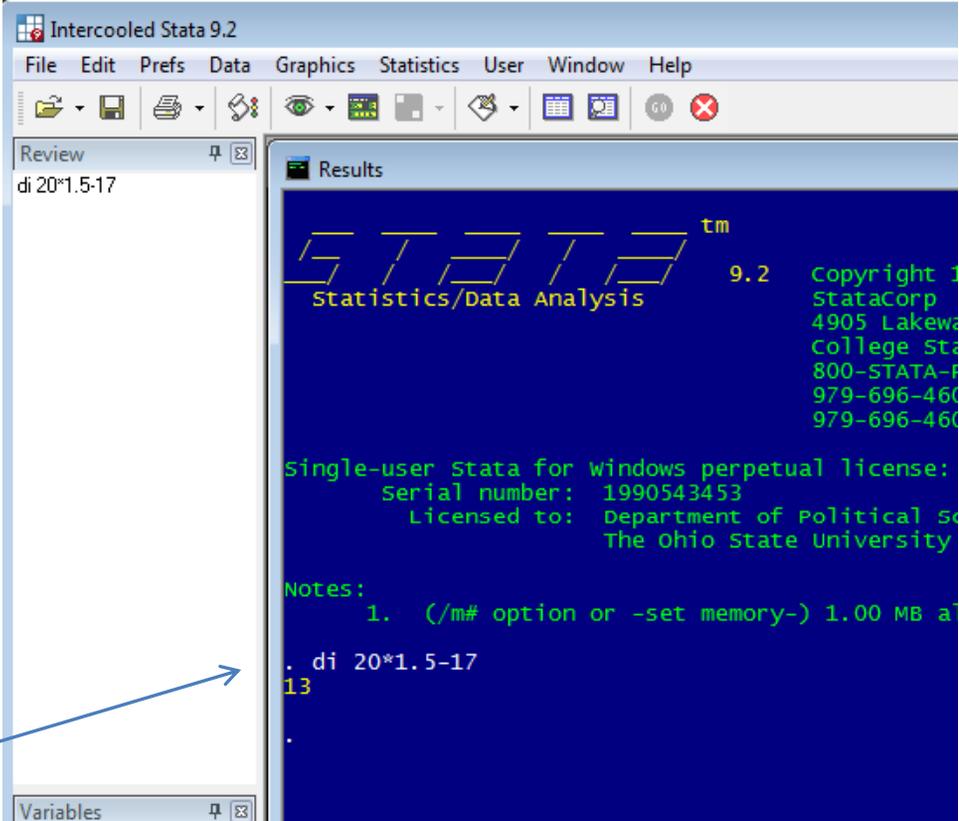
V043085 7pt
-3 to
3=Cons W
76/84 Coded
to midd
zero & 85
mv as .
```

|       | Freq. | Percent | Cum.   |
|-------|-------|---------|--------|
| -3    | 27    | 2.23    | 2.23   |
| -2    | 112   | 9.25    | 11.48  |
| -1    | 102   | 8.42    | 19.90  |
| 0     | 588   | 48.55   | 68.46  |
| 1     | 145   | 11.97   | 80.43  |
| 2     | 201   | 16.60   | 97.03  |
| 3     | 36    | 2.97    | 100.00 |
| Total | 1,211 | 100.00  |        |



# Stata as a Calculator

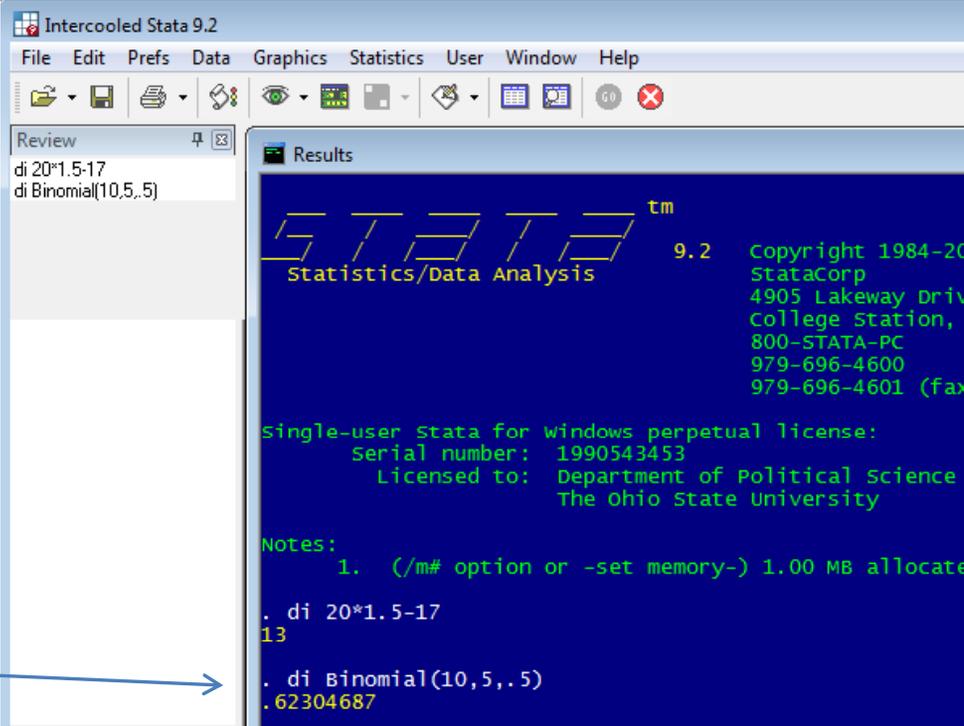
- Stata can be used to compute both basic and advanced mathematical operations
- Use the `display` command, or `di`, followed by the mathematical expression
- `di 20*1.5-17`



The screenshot shows the Stata 9.2 interface. The 'Review' window on the left contains the command `di 20*1.5-17`. The 'Results' window on the right displays the Stata logo and version information, followed by the command `. di 20*1.5-17` and its output `13`. A blue arrow points from the command in the Review window to the output in the Results window.

# Some Basic Statistics

- Stata can also perform several probability functions
- Example: What's the probability of tossing a coin ten times and getting five heads?
- `di`  
`Binomial(10,5,.5)`



The screenshot shows the Stata 9.2 interface. The command window on the left contains the command `di 20*1.5-17` and `di Binomial(10,5,.5)`. The results window on the right displays the output of these commands. The first command results in the value 13, and the second command results in the probability 0.62304687. A blue arrow points from the command `Binomial(10,5,.5)` in the list to the corresponding output in the results window.

```
Intercooled Stata 9.2
File Edit Prefs Data Graphics Statistics User Window Help
Review
di 20*1.5-17
di Binomial(10,5,.5)
Results
STATA 9.2 Copyright 1984-2005
Statistics/Data Analysis StataCorp
4905 Lakeway Drive
College Station, TX 77845-1627
800-STATA-PC
979-696-4600
979-696-4601 (fax)

single-user stata for windows perpetual license:
Serial number: 1990543453
Licensed to: Department of Political Science
The Ohio State University

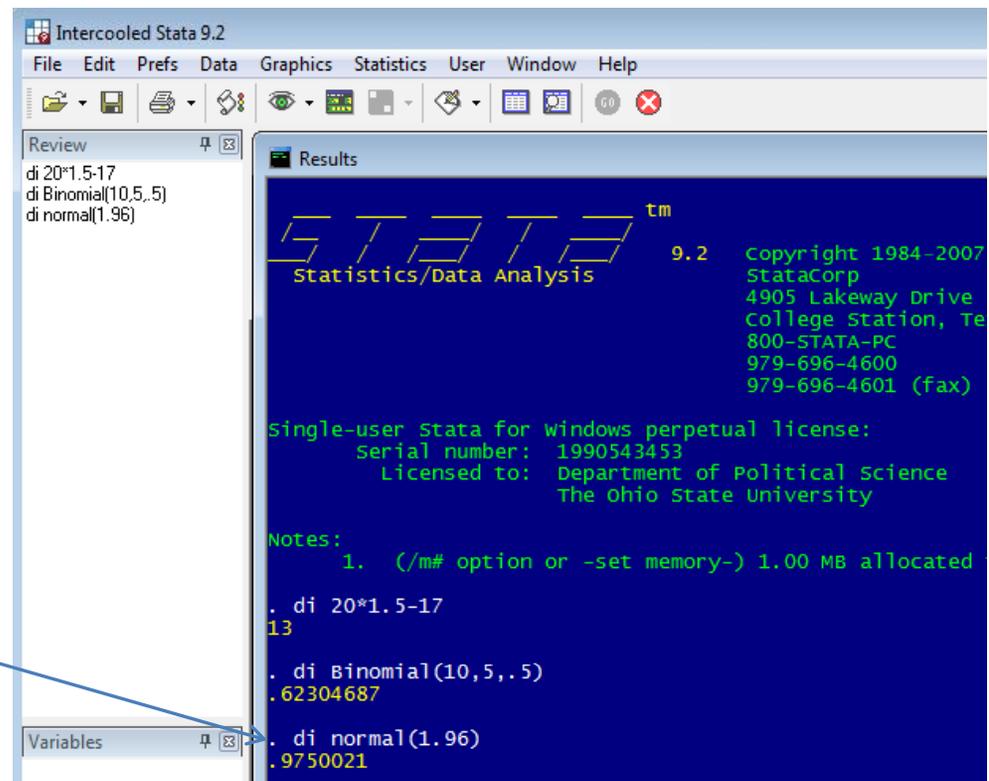
Notes:
1. (/m# option or -set memory-) 1.00 MB allocated

. di 20*1.5-17
13

. di Binomial(10,5,.5)
.62304687
```

# Some Basic Statistics

- Example: CDF for the normal distribution,  $z = 1.96$
- `di normal(1.96)`



The screenshot shows the Stata 9.2 interface with the following content:

```
Intercooled Stata 9.2
File Edit Prefs Data Graphics Statistics User Window Help
Review
di 20*1.5-17
di Binomial(10,5,.5)
di normal(1.96)
Results
STATA 9.2 Copyright 1984-2007
Statistics/Data Analysis StataCorp
4905 Lakeway Drive
College Station, Te
800-STATA-PC
979-696-4600
979-696-4601 (fax)
Single-user stata for windows perpetual license:
Serial number: 1990543453
Licensed to: Department of Political Science
The Ohio State University
Notes:
1. (/m# option or -set memory-) 1.00 MB allocated
. di 20*1.5-17
13
. di Binomial(10,5,.5)
.62304687
. di normal(1.96)
.9750021
```

# Some Basic Statistics

- Stata has many more distribution functions that can be implemented
- For a summary of these, use the following command:
- `help density functions`

The screenshot shows the Stata 9.2 interface. The command window on the left contains the following text:

```
Review
di 20*1.5-17
di Binomial(10,5,.5)
di normal(1.96)
help density functions

Variables

Command
```

The Results window on the right displays the output of the `help density functions` command. It shows the Stata logo and the following text:

```
Single-user stata for wind
serial number: 199
Licensed to: Dep
The

Notes:
1. (/m# option or -

. di 20*1.5-17
13

. di Binomial(10,5,.5)
.62304687

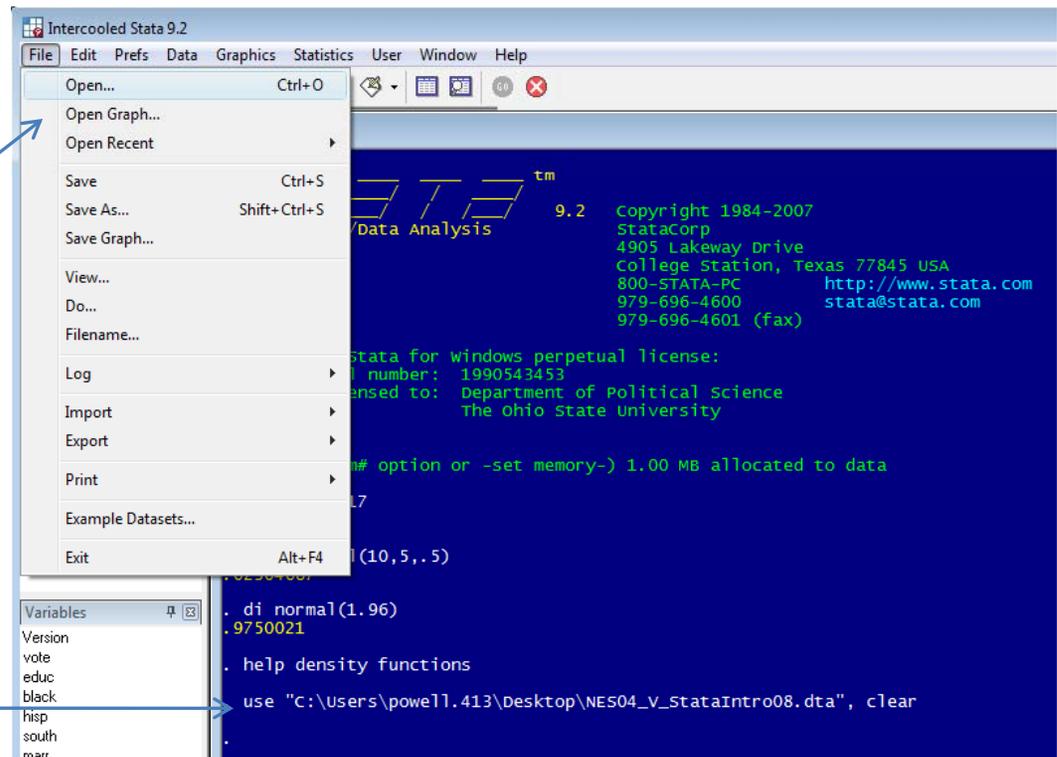
. di normal(1.96)
.9750021

. help density functions
.
```

The help viewer on the right shows the title "[D] functions — Functions" and a description: "This is a quick reference for the probability distribution and density functions." It lists various functions such as `betaden`, `binomial`, `binormal`, `chiz`, `chiztail`, `dgamma`, `Fden`, `Ftail`, and `gammaden`.

# In The Beginning... (Opening a Data Set)

- Several options exist for opening data sets
- Using the GUI allows you to browse or access recent data sets
- It is also possible to type in the use command



# The Data Editor

- Sort data by selected variable
- Move variable to first or last position
- Hide selected variable

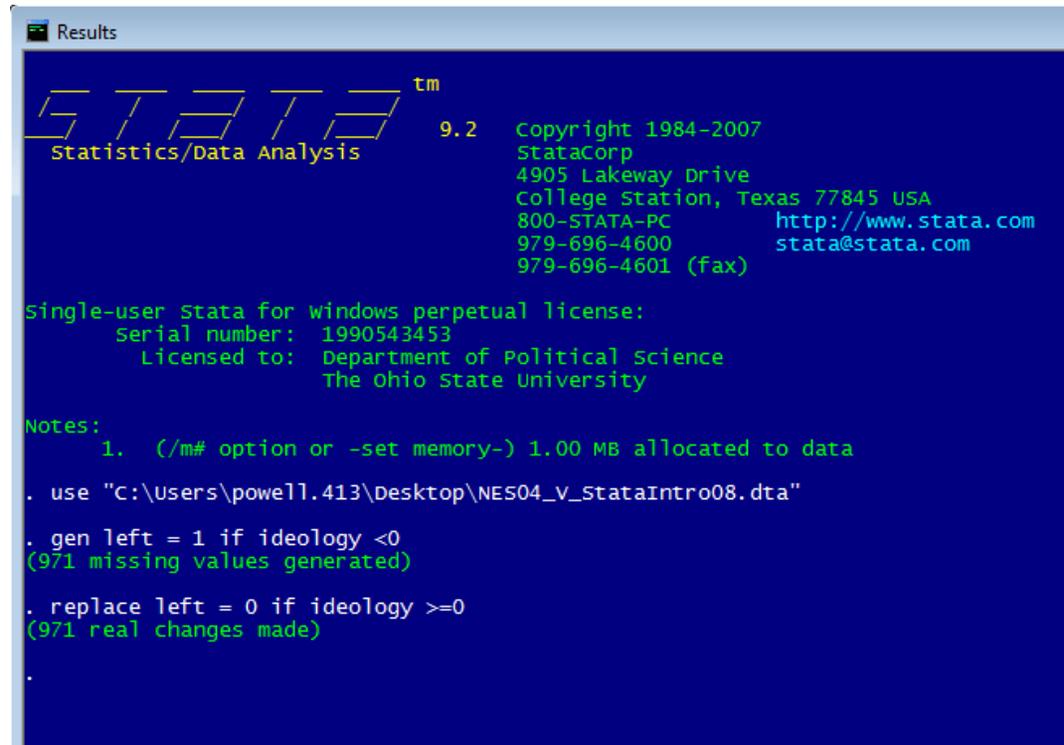
- Preserve changes that you've made
- Restore data to the state of the last "Preserve"

|    | vote | educ | black | hisp | south | marr | age | pray |
|----|------|------|-------|------|-------|------|-----|------|
| 1  | 1    | 4    | 0     | 0    | 0     | 0    | 57  | 1    |
| 2  | 1    | 3    | 0     | 0    | 1     | 1    | 63  | 1    |
| 3  | 1    | 4    | 0     | 0    | 0     | 1    | 49  | 1    |
| 4  | 1    | 4    | 0     | 0    | 0     | 1    | 56  | 0    |
| 5  | 1    | 6    | 0     | 0    | 0     | 1    | 75  | 1    |
| 6  | 1    | 4    | 0     | 0    | 0     | 1    | 51  | 1    |
| 7  | 1    | 4    | 0     | 0    | 0     | 0    | 22  | 1    |
| 8  | 1    | 4    | 0     | 0    | 0     | 0    | 41  | 1    |
| 9  | 1    | 4    | 0     | 0    | 0     | 1    | 66  | 1    |
| 10 | 1    | 4    | 0     | 0    | 0     | 1    | 72  | 1    |
| 11 | 1    | 4    | 0     | 0    | 1     | 0    | 19  | 1    |
| 12 | 1    | 3    | 0     | 0    | 1     | 1    | 59  | 1    |
| 13 | 1    | 3    | 0     | 0    | 1     | 1    | 64  | 1    |
| 14 | 1    | 3    | 0     | 0    | 0     | 1    | 55  | 0    |
| 15 | 1    | 4    | 0     | 0    | 1     | 1    | 84  | 1    |
| 16 | 1    | 4    | 0     | 0    | 1     | 0    | 27  | 1    |
| 17 | 1    | 6    | 0     | 0    | 1     | 1    | 55  | 1    |
| 18 | 1    | 6    | 0     | 0    | 0     | 0    | 46  | 1    |
| 19 | 1    | 3    | 0     | 0    | 1     | 1    | 68  | 1    |
| 20 | 1    | 7    | 0     | 0    | 0     | 1    | 41  | 1    |
| 21 | 1    | 6    | 0     | 0    | 0     | 1    | 56  | 0    |
| 22 | 1    | 6    | 0     | 0    | 0     | 1    | 58  | 1    |
| 23 | 1    | 4    | 0     | 0    | 0     | 1    | 56  | 1    |
| 24 | 1    | 4    | 0     | 0    | 0     | 1    | 25  | 0    |
| 25 | 1    | 4    | 0     | 0    | 0     | 1    | 64  | 1    |
| 26 | 1    | 4    | 0     | 0    | 0     | 1    | 31  | 1    |

- Delete selected variable or observation
- **And**, of course, you can edit each cell

# Manipulating the Data

- Stata can generate new variables and edit existing ones
- Let's create a new variable called "left" using generate and replace
- `gen left = 1 if ideology < 0`
- `replace left = 0 if ideology >= 0`



```
Results
-----
STATAtm 9.2 Copyright 1984-2007
Statistics/Data Analysis StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user stata for windows perpetual license:
  Serial number: 1990543453
  Licensed to: Department of Political Science
              The Ohio State University

Notes:
  1. (/m# option or -set memory-) 1.00 MB allocated to data

. use "C:\Users\powell.413\Desktop\NES04_V_StataIntro08.dta"

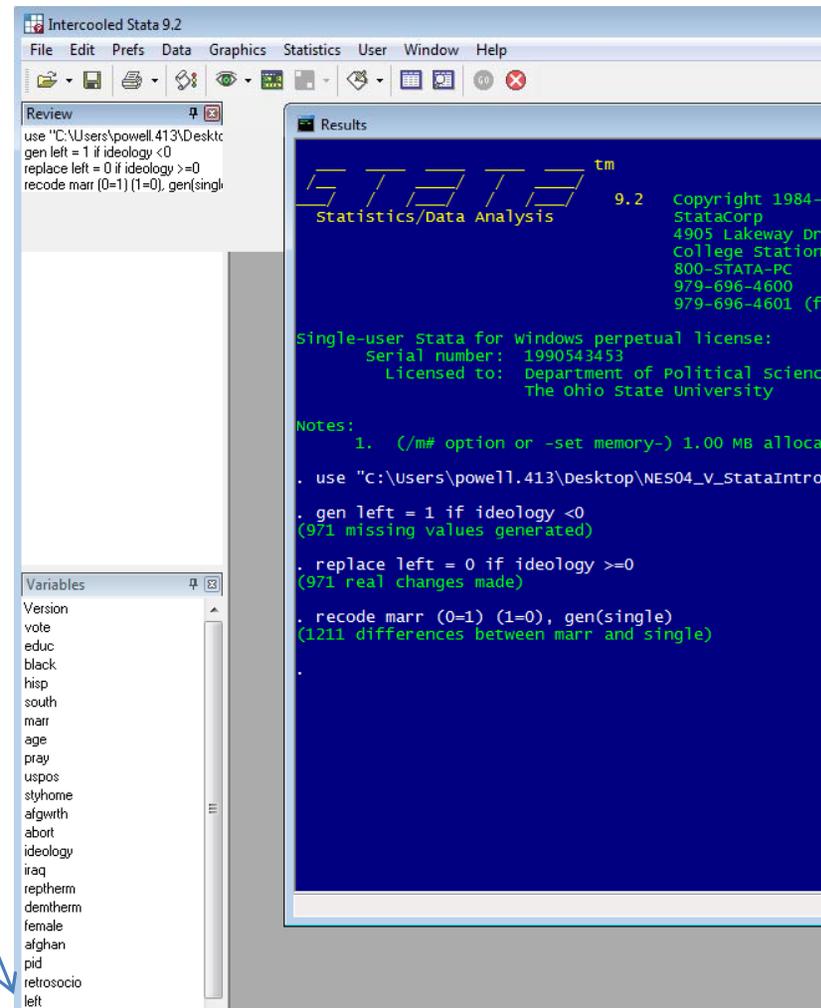
. gen left = 1 if ideology < 0
(971 missing values generated)

. replace left = 0 if ideology >= 0
(971 real changes made)

.
```

# Manipulating the Data

- Notice that we now have a new variable in our list
- Let's create a new variable by recoding an existing one
- `recode marr (0=1) (1=0), gen(single)`
- Other Expressions to know: `>=`, `<=`, `&`, `|`, `~`, `^`, `-`, `/`, `*`, `+`, `~=`



The screenshot shows the Stata 9.2 interface. The 'Review' window displays the following commands:

```
use "C:\Users\powell.413\Desktop\NES04_v_stataIntro.dta"
gen left = 1 if ideology <0
replace left = 0 if ideology >=0
recode marr (0=1) (1=0), gen(single)
```

The 'Results' window shows the output of these commands:

```
STATA 9.2 Copyright 1984-2005 StataCorp. 4905 Lakeway Drive College Station, TX 77845-1222
Single-user Stata for windows perpetual license:
Serial number: 1990543453
Licensed to: Department of Political Science, The Ohio State University

Notes:
1. (/m# option or -set memory-) 1.00 MB allocated

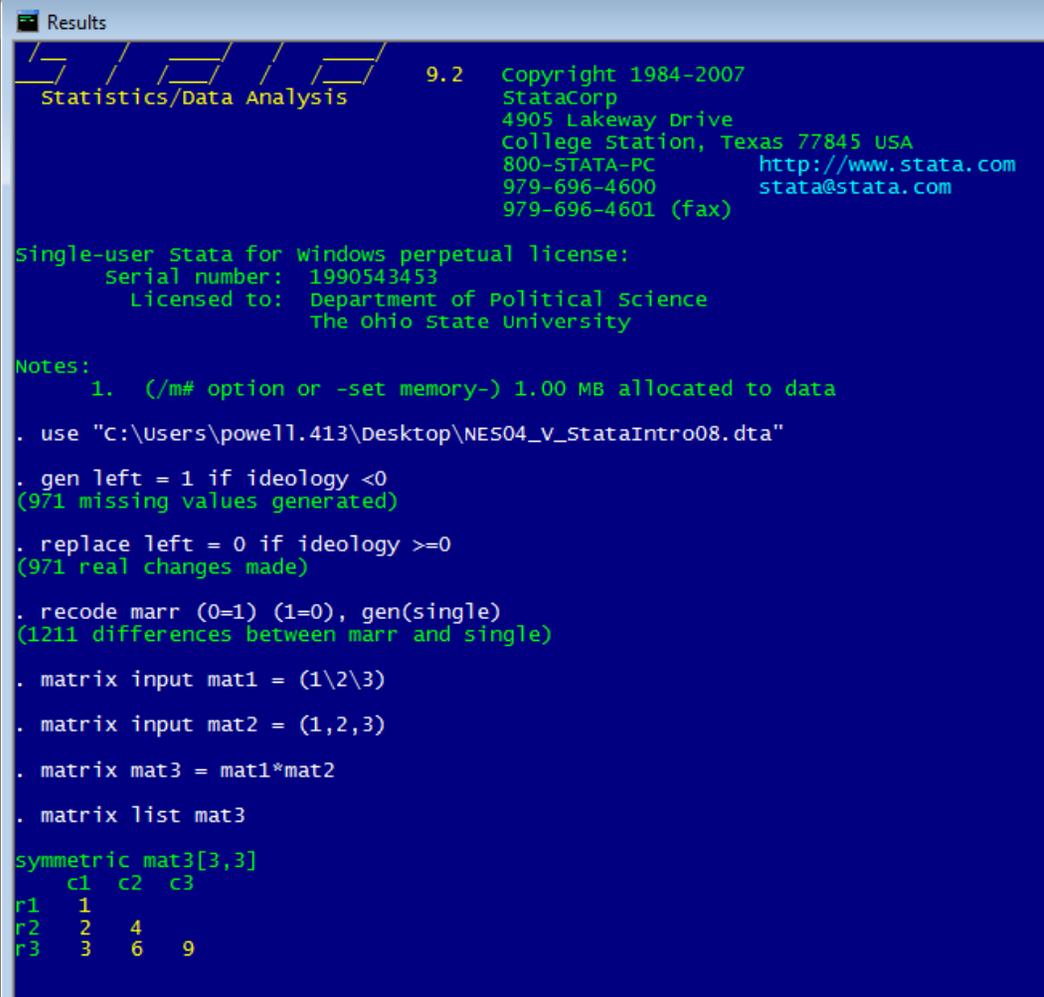
. use "C:\Users\powell.413\Desktop\NES04_v_stataIntro.dta"
. gen left = 1 if ideology <0
(971 missing values generated)
. replace left = 0 if ideology >=0
(971 real changes made)
. recode marr (0=1) (1=0), gen(single)
(1211 differences between marr and single)
.
```

The 'Variables' window lists the following variables:

```
Version
vote
educ
black
hisp
south
marr
age
pray
uspos
styhome
afgwrth
abort
ideology
iraq
replthem
demthem
female
afghan
pid
retrosocio
left
```

# Manipulating the Data

- Stata also has the ability to generate vectors and matrices
- matrix input  
mat1 =  
(1\2\3)
- matrix input  
mat2 =  
(1,2,3)
- matrix mat3 =  
mat1\*mat2
- matrix list  
mat3



```
Results
-----
Statistics/Data Analysis 9.2 Copyright 1984-2007
                          StataCorp
                          4905 Lakeway Drive
                          College Station, Texas 77845 USA
                          800-STATA-PC http://www.stata.com
                          979-696-4600 stata@stata.com
                          979-696-4601 (fax)

Single-user stata for windows perpetual license:
  Serial number: 1990543453
  Licensed to: Department of Political Science
              The Ohio State University

Notes:
  1. (/m# option or -set memory-) 1.00 MB allocated to data

. use "C:\Users\powell.413\Desktop\NES04_v_StataIntro08.dta"

. gen left = 1 if ideology <0
(971 missing values generated)

. replace left = 0 if ideology >=0
(971 real changes made)

. recode marr (0=1) (1=0), gen(single)
(1211 differences between marr and single)

. matrix input mat1 = (1\2\3)

. matrix input mat2 = (1,2,3)

. matrix mat3 = mat1*mat2

. matrix list mat3

symmetric mat3[3,3]
   c1  c2  c3
r1   1
r2   2   4
r3   3   6   9
```

# Describing the Data

- Now let's have a look at what we created
- tab ideology
- tab left
- tab marr
- tab single

The screenshot shows the Stata 9.2 interface. The Command window contains the following commands:

```
use "C:\Users\powell.413\Desktop"
gen left = 1 if ideology < 0
replace left = 0 if ideology >= 0
recode marr (0=1) (1=0), gen(single)
matrix input mat1 = (1\2\3)
matrix input mat2 = (1,2,3)
matrix mat3 = mat1*mat2
matrix list mat3
tab ideology
tab left
tab marr
tab single
```

The Results window displays the following tables:

**Summary Statistics for ideology**

| ideology | Freq. | Percent | Cum.   |
|----------|-------|---------|--------|
| -3       | 27    | 2.23    | 2.23   |
| -2       | 112   | 9.25    | 11.48  |
| -1       | 102   | 8.42    | 19.90  |
| 0        | 588   | 48.55   | 68.46  |
| 1        | 145   | 11.97   | 80.43  |
| 2        | 201   | 16.60   | 97.03  |
| 3        | 36    | 2.97    | 100.00 |
| Total    | 1,211 | 100.00  |        |

**Summary Statistics for left**

| left  | Freq. | Percent | Cum.   |
|-------|-------|---------|--------|
| 0     | 971   | 80.12   | 80.12  |
| 1     | 241   | 19.88   | 100.00 |
| Total | 1,212 | 100.00  |        |

**Summary Statistics for marr**

| marr  | Freq. | Percent | Cum.   |
|-------|-------|---------|--------|
| 0     | 563   | 46.49   | 46.49  |
| 1     | 648   | 53.51   | 100.00 |
| Total | 1,211 | 100.00  |        |

**Summary Statistics for single**

| single | Freq. | Percent | Cum.   |
|--------|-------|---------|--------|
| 0      | 648   | 53.51   | 53.51  |
| 1      | 563   | 46.49   | 100.00 |
| Total  | 1,211 | 100.00  |        |

# Describing the Data

- To produce a list and summary of all variables, use the `sum` command

• `sum`

- You can also use this command to summarize individual variables

• `sum ideology`

```
Results
and
marr=1&6=1,
2/5=0, 8=-.)

```

|       | Freq. | Percent | Cum.   |
|-------|-------|---------|--------|
| 0     | 648   | 53.51   | 53.51  |
| 1     | 563   | 46.49   | 100.00 |
| Total | 1,211 | 100.00  |        |

```
. sum
```

| Variable   | Obs  | Mean      | Std. Dev. | Min | Max |
|------------|------|-----------|-----------|-----|-----|
| version    | 0    |           |           |     |     |
| vote       | 811  | .5080148  | .5002443  | 0   | 1   |
| educ       | 1212 | 4.302805  | 1.612265  | 0   | 7   |
| black      | 1204 | .1528239  | .3599672  | 0   | 1   |
| hisp       | 1204 | .0722591  | .2590241  | 0   | 1   |
| south      | 1212 | .2970297  | .4571384  | 0   | 1   |
| marr       | 1211 | .535095   | .4989729  | 0   | 1   |
| age        | 1212 | 47.27228  | 17.14157  | 18  | 90  |
| pray       | 1192 | .7323826  | .4429028  | 0   | 1   |
| uspos      | 1212 | 2.531353  | 1.622969  | 1   | 5   |
| styhome    | 1199 | .1901585  | .3925898  | 0   | 1   |
| afgwrth    | 1187 | .7034541  | .456927   | 0   | 1   |
| abort      | 1047 | 2.209169  | 1.085799  | 1   | 4   |
| ideology   | 1211 | .2047894  | 1.290529  | -3  | 3   |
| iraq       | 1209 | 2.588089  | 1.932398  | 1   | 5   |
| reptherm   | 1176 | 53.23384  | 26.95131  | 0   | 100 |
| demtherm   | 1178 | 58.36927  | 24.16135  | 0   | 100 |
| female     | 1212 | .5330033  | .4991155  | 0   | 1   |
| afghan     | 1210 | 3.798347  | 1.81365   | 1   | 5   |
| pid        | 1195 | -.1271967 | 2.091833  | -3  | 3   |
| retrosocio | 1212 | -.2178218 | .7934991  | -1  | 1   |
| left       | 1212 | .1988449  | .3992958  | 0   | 1   |
| single     | 1211 | .464905   | .4989729  | 0   | 1   |

```
. sum ideology
```

| variable | Obs  | Mean     | Std. Dev. | Min | Max |
|----------|------|----------|-----------|-----|-----|
| ideology | 1211 | .2047894 | 1.290529  | -3  | 3   |

# Describing the Data

- The tab command can also be used to create cross-tabs when implemented with two variables
- `tab left marr`
- Summary statistics can be separated using the `by` command, but you have to sort first
- `sort left`
- `by left: sum educ`

```
Results
retrosocio      1212  -.2178218  .7934991  -1  1
      left      1212  .1988449  .3992958   0  1
      single    1211  .464905  .4989729   0  1

. sum ideology
      variable |      obs      Mean      Std. Dev.      Min      Max
-----+-----+-----+-----+-----+-----
      ideology |    1211  .2047894  1.290529      -3      3

. tab left marr
           |      V043251 dummy part
           |      and marr=1&6=1,
           |      2/5=0, 8=.
      left |      0      1      Total
-----+-----+-----+-----
           |    432    538    970
           |    131    110    241
      Total |    563    648   1,211

. by left: sum educ
not sorted
r(5);
. sort left
. by left: sum educ

-> left = 0
      variable |      obs      Mean      Std. Dev.      Min      Max
-----+-----+-----+-----+-----
           educ |    971  4.162719  1.551949         0         7

-> left = 1
      variable |      obs      Mean      Std. Dev.      Min      Max
-----+-----+-----+-----+-----
           educ |    241  4.86722  1.726933         1         7
```

# Describing the Data

- In Stata, data exists in several formats
- For a summary of data types in your data set, use the describe command
- describe

```
Results
. describe
Contains data from C:\Users\powell.413\Desktop\NES04_V_StataIntro08.dta
obs:      1,212
vars:     23
size:     141,804 (86.5% of memory free)
16 Jan 2008 17:15
```

| variable name | storage type | display format | value label | variable label   |
|---------------|--------------|----------------|-------------|--|
| Version       | str25        | %25s           |             | Study.1. Dataset version   |
| vote          | float        | %9.0g          |             | 0 challenger and 1 incumbent   |
| educ          | float        | %9.0g          |             | V043254 8pt scale unchanged  |
| black         | float        | %9.0g          |             | V043299 dummy B,BH and<br>BW=10/15=1, 10/70=0, 88/89=.                                 |
| hisp          | float        | %9.0g          |             | V043299 dummy 15/35=0 70=0<br>BH=14=1, H and HW=40/45=1,<br>88/89=.                    |
| south         | float        | %9.0g          |             | V041203 9 states south res<br>40/49=1, 1/59=0, 50/73=0                                 |
| marr          | float        | %9.0g          |             | V043251 dummy part and<br>marr=1&6=1, 2/5=0, 8=.                                       |
| age           | float        | %9.0g          |             | V043250 age 18 to 90 unchanged   |
| pray          | float        | %9.0g          |             | V043221 dummy prays=more than<br>once a week, dk and ref as mvs                        |
| uspos         | float        | %9.0g          |             | V043112 3pt 8/9=3 mv recoded to<br>center  |
| styhome       | float        | %9.0g          |             | V043113 dummy disagr=5=0, 8/9=.  |
| afgwrth       | float        | %9.0g          |             | V043131 dummy worth it=1 and dk<br>and ref 8/9 coded as mvs                            |
| abort         | float        | %9.0g          |             | V045132 4pt 4=high=Cons=never<br>permit abort, 1=low=always<br>permit, 2 and 3 qualifi |
| ideology      | float        | %9.0g          |             | V043085 7pt -3 to 3=Cons w<br>76/84 coded to midd zero & 85<br>mv as .                 |
| iraq          | float        | %9.0g          |             | V043134 3pt scale w 1=not worth<br>it and 5=high=worth it and<br>3=mid; same as afgwa  |
| reptherm      | float        | %9.0g          |             | V043050 100pt therm scale w mvs<br>to mvs and none recoded to<br>center                |
| demtherm      | float        | %9.0g          |             | V043049 100pt therm scale w mvs<br>to mvs and none recoded to<br>center                |
| female        | float        | %9.0g          |             | V041109a dummy with male as 0<br>and female as 1                                       |

# Describing the Data

- Strings are non-numeric variables
- Floats are numeric data types that store up to 7 digits of accuracy, rounding thereafter
- byte, int, long, and double are other numeric types
- Useful commands for changing data types: format, destring, encode

```
Results
. describe
Contains data from C:\Users\powell.413\Desktop\NES04_V_StataIntro08.dta
obs:      1,212
vars:      23
size:     141,804 (86.5% of memory free)
16 Jan 2008 17:15
```

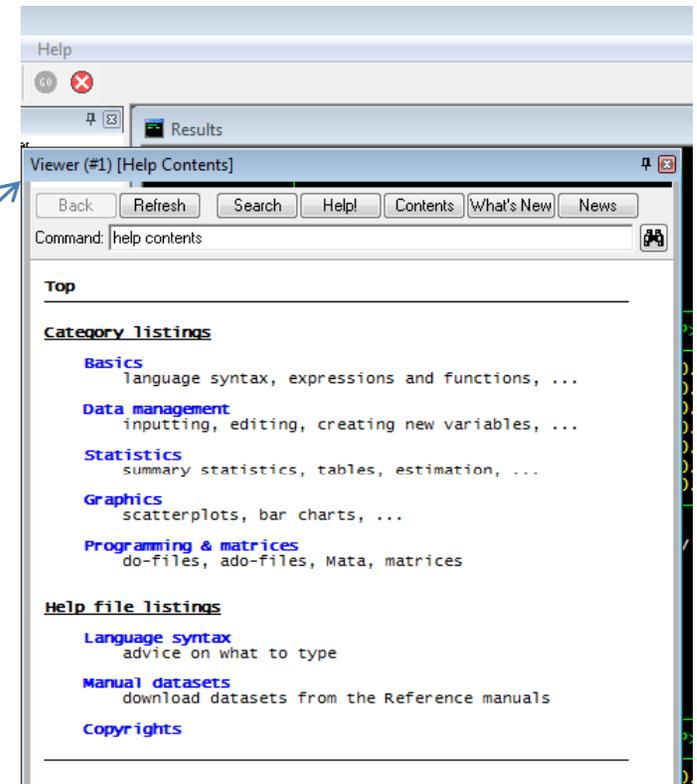
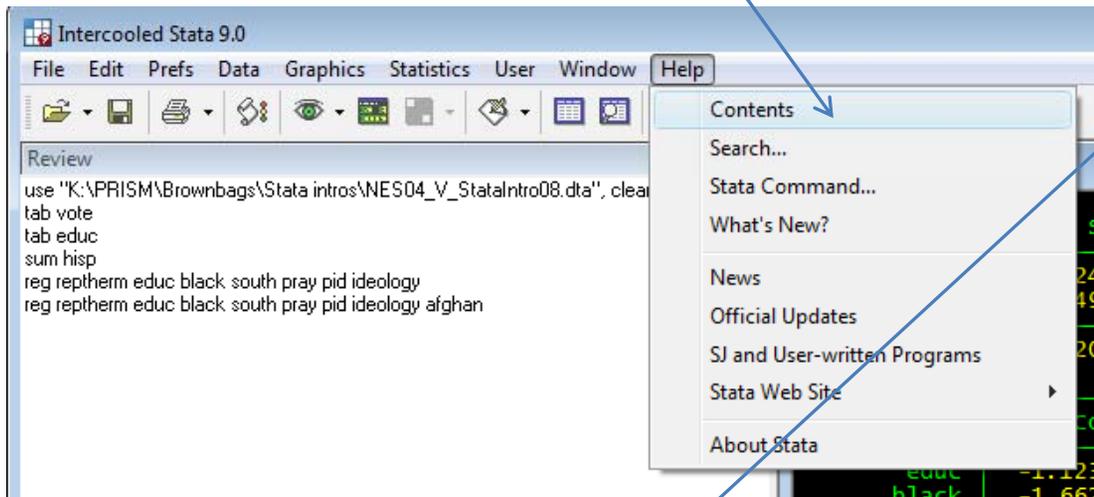
| variable name | storage type | display format | value label | variable label   |
|---------------|--------------|----------------|-------------|--|
| Version       | str25        | %25s           |             | Study.1. Dataset version   |
| vote          | float        | %9.0g          |             | 0 challenger and 1 incumbent   |
| educ          | float        | %9.0g          |             | V043254 8pt scale unchanged  |
| black         | float        | %9.0g          |             | V043299 dummy B,BH and<br>BW=10/15=1, 10/70=0, 88/89=.                                 |
| hisp          | float        | %9.0g          |             | V043299 dummy 15/35=0 70=0<br>BH=14=1, H and HW=40/45=1,<br>88/89=.                    |
| south         | float        | %9.0g          |             | V041203 9 states south res<br>40/49=1, 1/59=0, 50/73=0                                 |
| marr          | float        | %9.0g          |             | V043251 dummy part and<br>marr=1&6=1, 2/5=0, 8=.                                       |
| age           | float        | %9.0g          |             | V043250 age 18 to 90 unchanged   |
| pray          | float        | %9.0g          |             | V043221 dummy prays=more than<br>once a week, dk and ref as mvs                        |
| uspos         | float        | %9.0g          |             | V043112 3pt 8/9=3 mv recoded to<br>center  |
| styhome       | float        | %9.0g          |             | V043113 dummy disagr=5=0, 8/9=.  |
| afgwrth       | float        | %9.0g          |             | V043131 dummy worth it=1 and dk<br>and ref 8/9 coded as mvs                            |
| abort         | float        | %9.0g          |             | V045132 4pt 4=high=Cons=never<br>permit abort, 1=low=always<br>permit, 2 and 3 qualifi |
| ideology      | float        | %9.0g          |             | V043085 7pt -3 to 3=Cons w<br>76/84 coded to midd zero & 85<br>mv as .                 |
| iraq          | float        | %9.0g          |             | V043134 3pt scale w 1=not worth<br>it and 5=high=worth it and<br>3=mid; same as afgha  |
| reptherm      | float        | %9.0g          |             | V043050 100pt therm scale w mvs<br>to mvs and none recoded to<br>center                |
| demtherm      | float        | %9.0g          |             | V043049 100pt therm scale w mvs<br>to mvs and none recoded to<br>center                |
| female        | float        | %9.0g          |             | V041109a dummy with male as 0<br>and female as 1                                       |

# Help Viewer

- The capabilities of Stata are vast
- What you can do with Stata depends on your knowledge of the commands
- Fortunately Stata comes with user friendly help
- Stata's greatest selling point
  - All commands are easily referenced
  - All commands come with helpful descriptions and examples
  - All commands have been peer reviewed

# Help Viewer

- To open the Help Viewer click on Help → Contents



- The Help Viewer opens and allows you to browse the entire Stata database and online resources
- It acts like an internet browser...

# Help Viewer

- Take the now familiar `tab` command
- In command prompt or in help viewer prompt:  
`help tabulate`
- Provides information on:
  - Command title
  - Command syntax
- Note: blue font is linked; click on it to get more info on the given word

```
Viewer (#1) [help tabulate oneway]
Back Refresh Search Help Contents What's New News
Command: help tabulate oneway

help tabulate, help tab1          dialogs: tabulate
                                         tabulate ..., generate(
                                         tab1

Title
[R] tabulate oneway — One-way tables of frequencies

Syntax
One-way tables of frequencies
    tabulate varname [if] [in] [weight] [, tabulate_options]
One-way tables for each variable — a convenience tool
    tab1 varlist [if] [in] [weight] [, tab1_options]

tabulate_options    description
-----
Main
subpop(varname)    exclude observations for which varname = 0
missing            treat missing values like other values
nofreq            do not display frequencies
nolabel           display numeric codes rather than value labels
plot              produce a bar chart of the relative frequencies
sort              display the table in descending order of frequency

Advanced
generate(varname) create indicator variables for varname
matcell(matname)  save frequencies in matname; programmer's option
matrow(matname)   save unique values of varname in matname

tab1_options        description
-----
Main
missing            treat missing values like other values
nolabel           display numeric codes rather than value labels
plot              produce a bar chart of the relative frequencies
sort              display the table in descending order of frequency

by may be used with tabulate and tab1; see by.
fweights, awweights, and iweights are allowed by tabulate. fweights are
allowed by tab1. See weight.
```

# Help Viewer

- Also provides information on:
  - Command Description
  - Command Options
  - Command Examples
  - Related commands

```
Examples
. tabulate mpgcat
. tabulate mpgcat, generate(mpg)
. tabulate region [freq=pop]
. tab1 sex race age

Also see
Manual: [R] tabulate oneway
Online: collapse, encode, epitab, svy: tabulate oneway, svy: tabulate
twoway, table, tabstat, tabulate summarize, tabulate twoway,
xtides, xttab
```

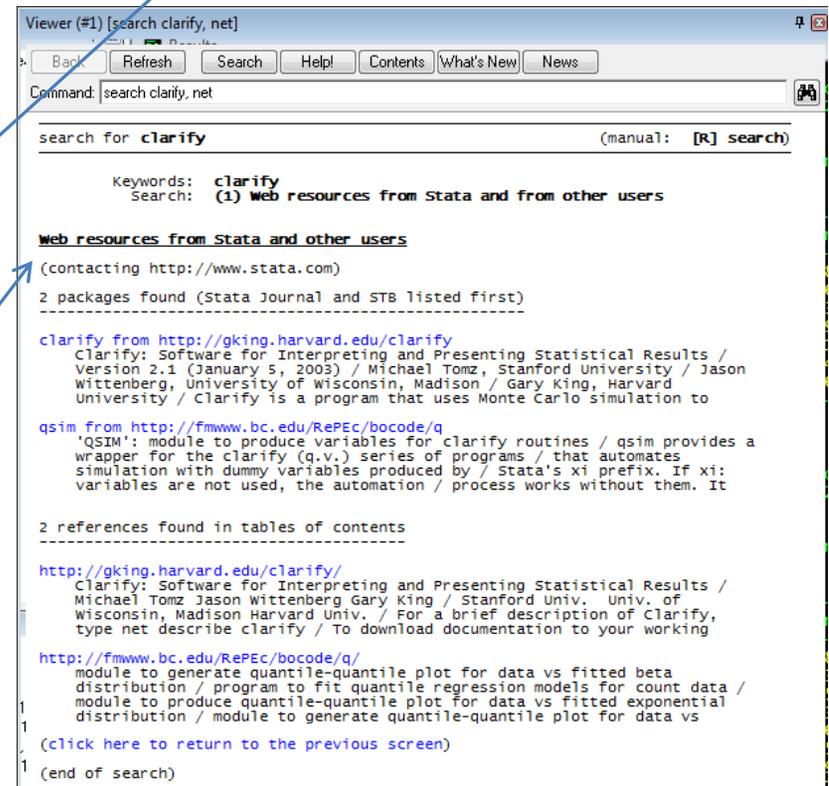
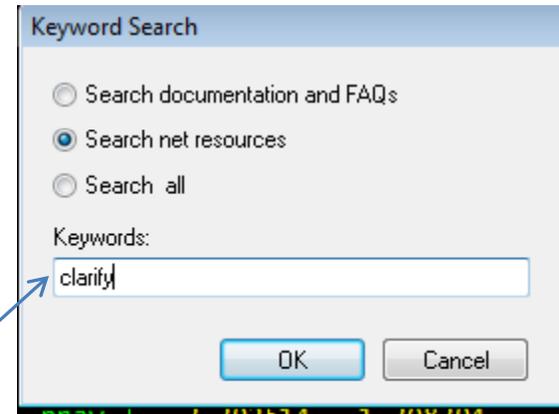
```
Viewer (#1) [help tabulate oneway]
Back Refresh Search Help Contents What's New News
Command: help tabulate oneway

Description
tabulate produces one-way tables of frequency counts.
For information on two-way tables of frequency counts along with measures
of association, including the common Pearson chi-squared, the likelihood
ratio chi-squared, Cramér's V, Fisher's exact test, Goodman and Kruskal's
gamma, and Kendall's tau-b, see tabulate twoway.
tab1 produces a one-way tabulation for each variable specified in
varlist.
Also see table and tabstat if you want one-, two-, or n-way tables of
frequencies and a wide variety of statistics. See tabulate summarize for
a description of tabulate with the summarize() option; it produces tables
(breakdowns) of means and standard deviations. table is better than
tabulate, summarize(), but tabulate, summarize() is faster. See epitab
for 2 x 2 tables with statistics of interest to epidemiologists.

Options
Main
subpop(varname) excludes observations for which varname = 0 in tabulating
frequencies. The mathematical results of tabulate ..., subpop(myvar)
are the same as tabulate ... if myvar != 0, but the table may be
presented differently. The identities of the rows and columns will
be determined from all the data, including the myvar = 0 group, so
there may be entries in the table with frequency 0.
Consider tabulating answer, a variable that takes on values 1, 2, and
3, but consider tabulating it just for male=1 subpopulation. Assume
that answer is never 2 in this group. tabulate answer if male=1
produces a table with two rows: one for answer 1 and one for answer
3. There will be no row for answer 2 because answer 2 was never
observed. tabulate answer, subpop(male) produces a table with three
rows. The row for answer 2 will be shown as having 0 frequency.
missing requests that missing values be treated like other values in
calculations of counts, percentages, and other statistics.
nofreq suppresses the printing of the frequencies.
nolabel causes the numeric codes to be displayed rather than the value
labels.
plot produces a bar chart of the relative frequencies in a one-way table.
(Also see histogram.)
sort puts the table in descending order of frequency (and ascending order
of the variable within equal values of frequency).
Advanced
generate(varname) creates a set of indicator variables reflecting the
observed values of the tabulated variable. The generate() option may
not be used with the by prefix.
matcell(matname) saves the reported frequencies in matname. This option
```

# Help Viewer

- Add-on packages also easy to find with the help viewer
  - For e.g., “Clarify” by G. King
  - Search clarify: Help → Search...
  - Type: clarify
  - Help finds the add-on package site and provides links for its description and download

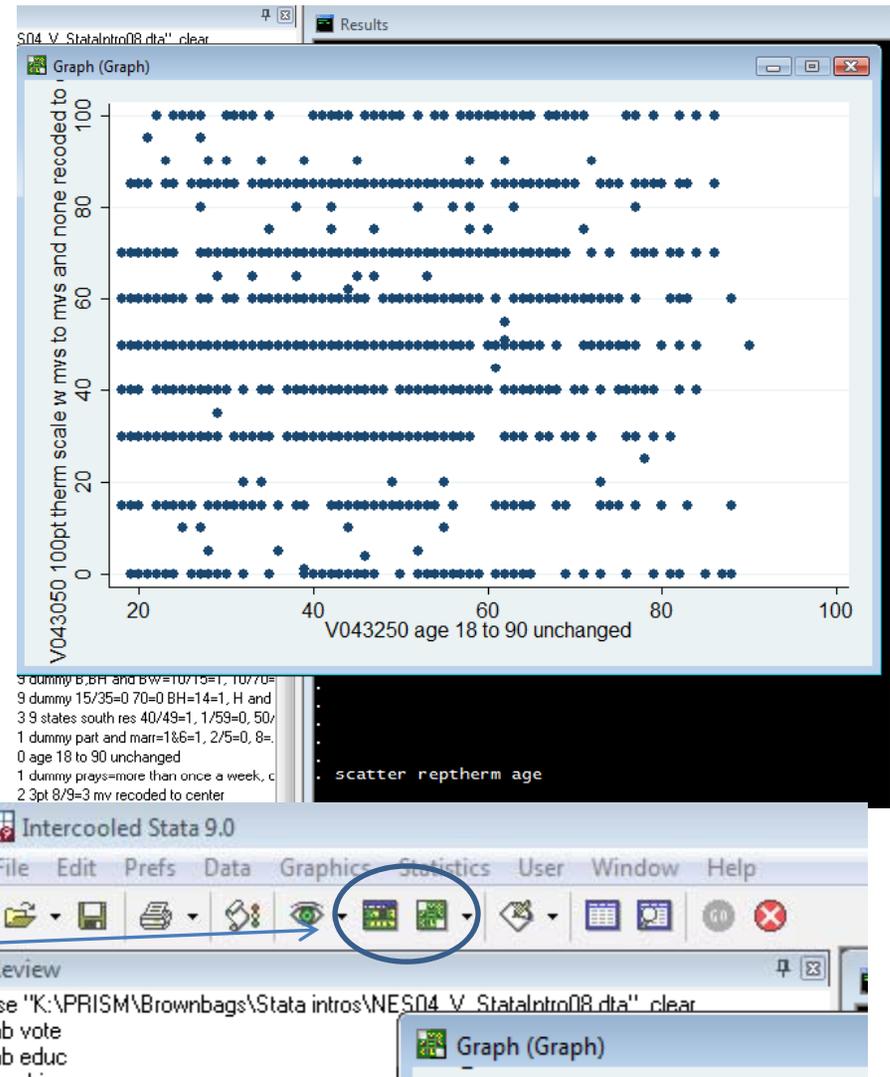


# Graphing

- Stata has numerous graphing capabilities
  - ANOVA and post-estimation OLS
  - Time Series: ARCH, ARIMA, VAR...
  - Duration Analysis: exponential, weibull, cox...
  - Event Count: negative binomial, poisson, Hurdle...
  - Limited Dependent Variables: logit, probit, multinomial logit and probit, ordered logit and probit...
  - Selection Models: heckman, censored probit, tobit,...
  - And, if it is not canned, we can program it – but that is for another brownbag
- Furthermore Stata 10 is supposed to be a drastic improvement in the flexibility of graphing functions
  - Competition with R?
- Let's quickly look at some of the basic graphs you can create

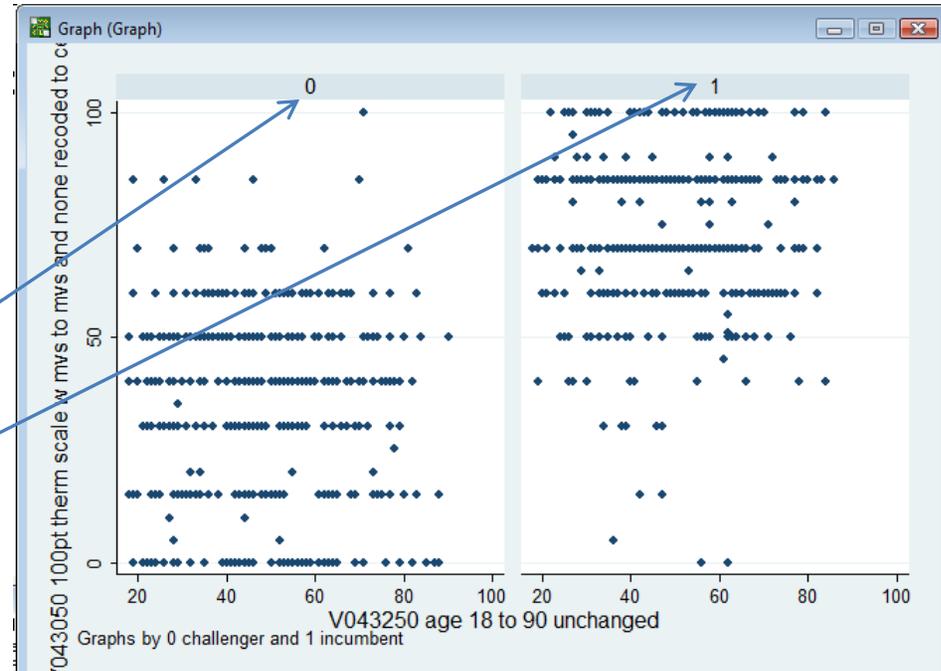
# Scatterplots

- Perhaps we want to check if our data hints that people become more favorable to conservative values as they age
- We can graph the variables with respect to one another
  - scatter rephtherm age
- Graph viewer appears above the results viewer
- Toggle to and fro with graph viewer buttons on toolbar



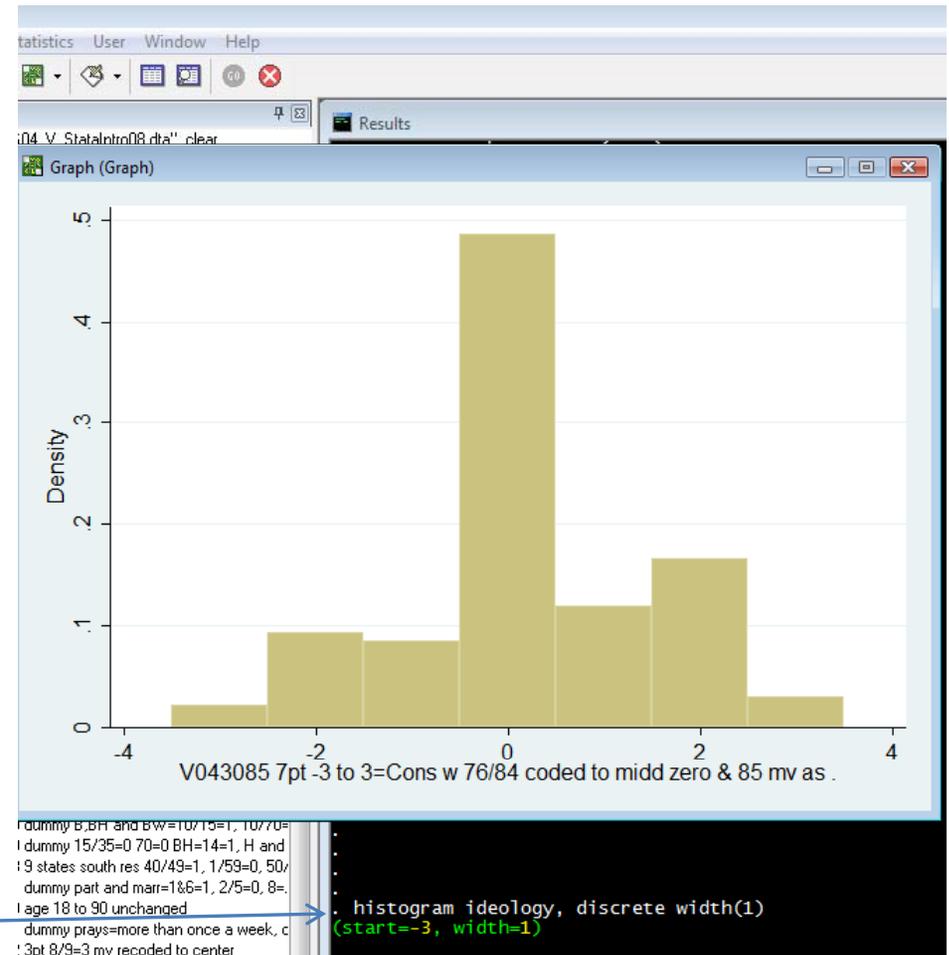
# Scatterplots

- We can also look at the same relationship by a particular sample of our data
- Perhaps there is a difference between those that voted for Bush (1) and Kerry (0)
- Let's sort by vote
- Try scatter rephtherm age, by(vote)



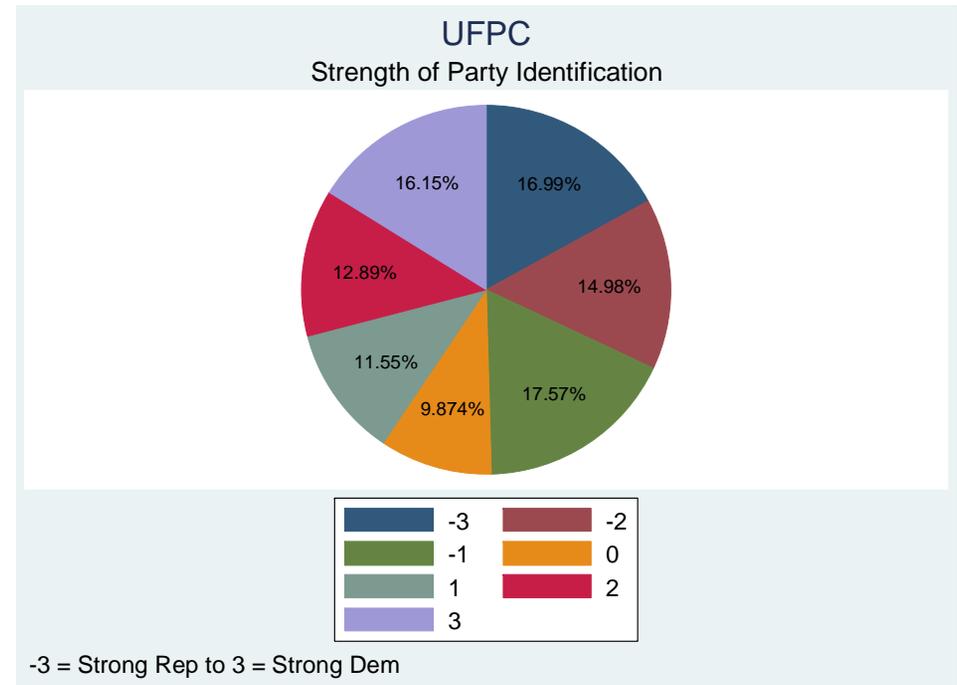
# Bar Charts & Histograms

- Say we are interested in the distribution of a categorical variable
- Try creating a bar chart for our measure of political ideology
- Type
- `hist ideology, discrete width(1)`



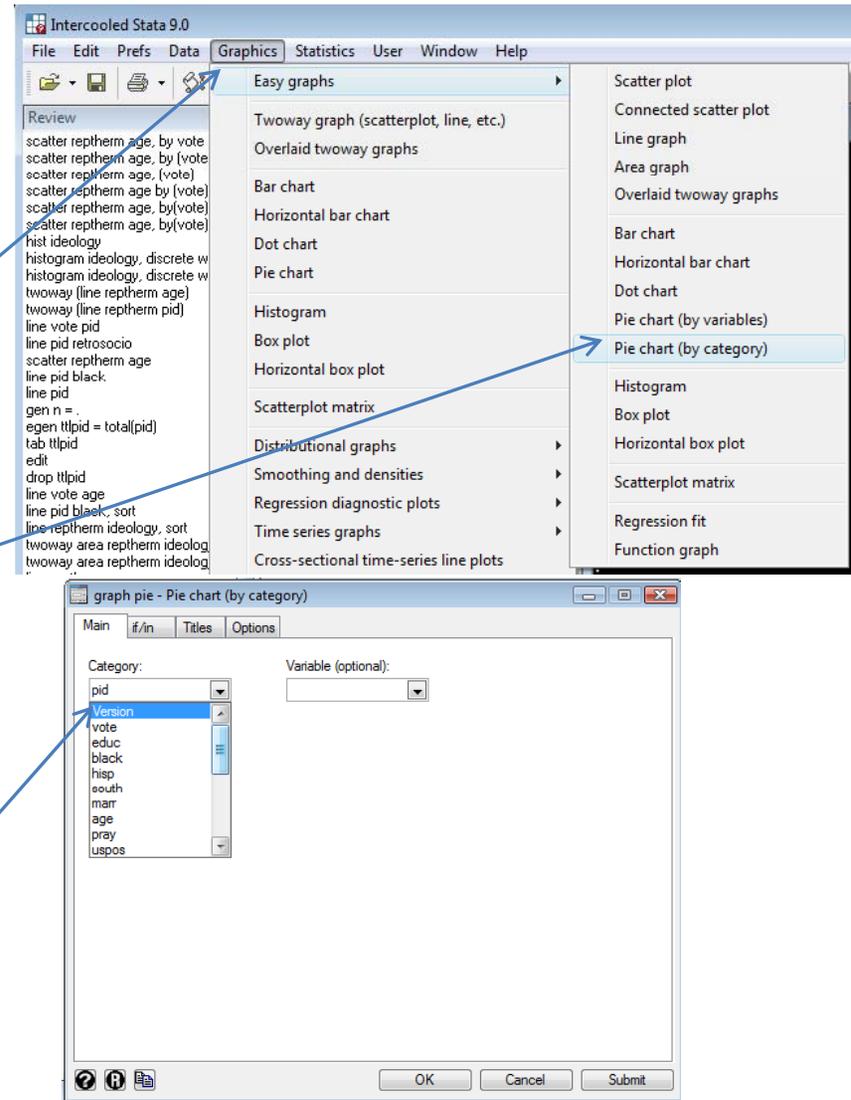
# UFPC

- Say you need to paint a really basic picture of party id strength for your coworkers
- Try a pie chart
  - `graph pie, over(pid)`
  - Use options for presentation:
  - `title(UFPC)`
  - `subtitle(Strength of Party Identification)`
  - `caption(-3 = Strong Rep to 3 = Strong Dem)`
  - `plabel(_all percent) cw`
- Then quit your job; you're working with imbeciles



# Graphing with GUI

- Of course, we did not need the exact commands to create the graphs above
- We could have used the GUI toolbar to create any of those graphs
- Just go to Graphics and select the appropriate graph
- A new viewer will appear
- Select from the drop-down menu to fill in the necessary variables and options

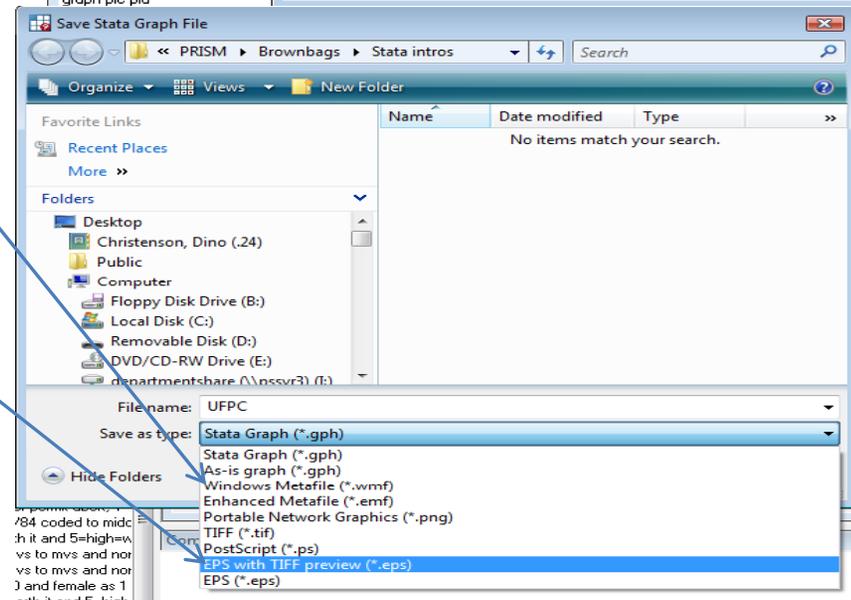
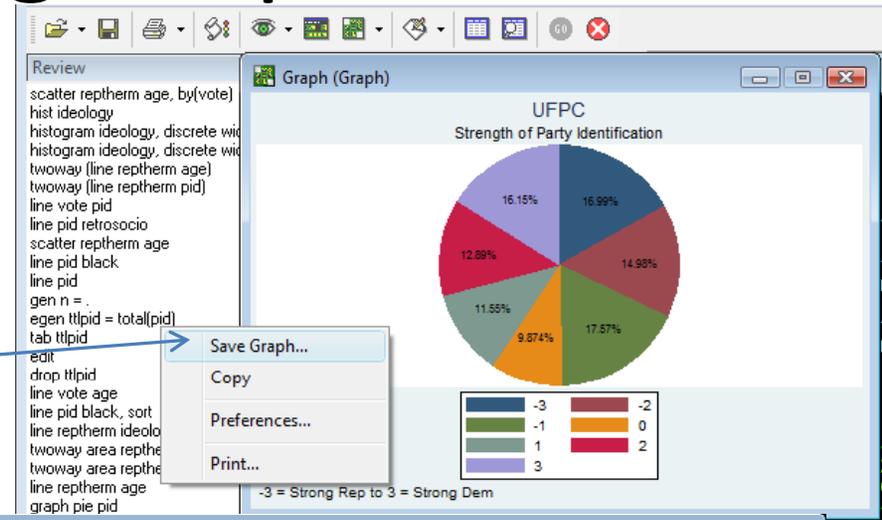


# Exporting Graphs & Tables

- So why did the last chart, the UFPC, look so nice and the others... not so much?
  - 1. Used titles
  - 2. Used a key
    - The graph was understandable on its own
  - 3. Exported the graph as a picture
- Stata allows you to export its output – both tables and graphs – in various formats
  - Depending on your typesetting system you will want to save the output in different manners

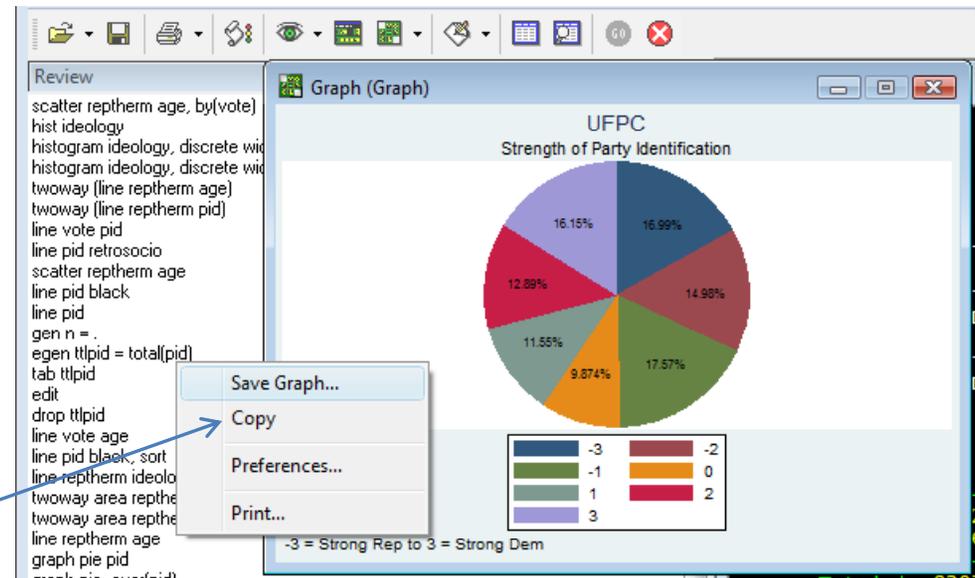
# Exporting Graphs

- To save and export a graph, right click on the graph (control click to my Mac friends)
  - Click Save Graph
  - Save in the appropriate format
    - Word: .wmf or .png
    - Latex: .eps
- Alternatively, go to the main toolbar and click File → Save Graph
  - Follow same procedure



# Exporting Graphs

- Shortcut to word users
- To merely copy a graph, right click on the graph (control click to my Mac friends)
  - Click Copy
  - Paste it in your word processor
  - Note: you do not have a separate saved graph in this case



# Exporting Tables

- The Stata table output is not appropriate for a conference paper or article submission
- Why not?
  - 1. Too much information
  - 2. Vertical lines
  - 3. Variable names
  - 4. No title or explanation
- Therefore, when you write a paper you will need to transform the output
- You've all seen article worthy tables (e.g. Balla & Wright 2001)

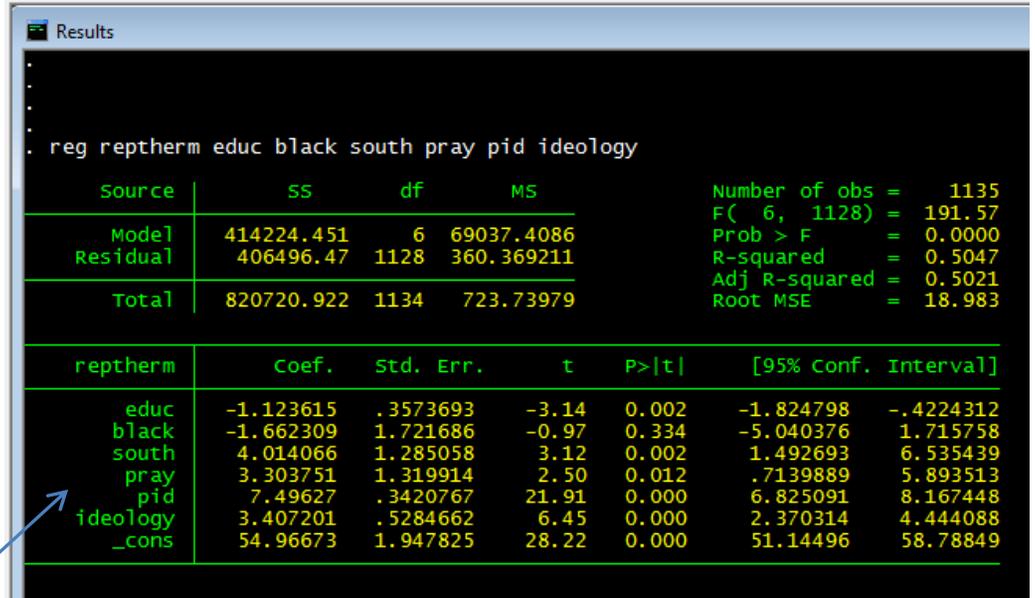
**TABLE 4** The Determinants of EPA Selection of NDWAC Members

| Variable                           | Parameter Estimate<br>(Standard Error) |                   |
|------------------------------------|--|-------------------|
| AWWA Support                       | -1.40<br>(1.81)                        | -1.08<br>(1.13)   |
| NAWC Support                       | 3.54**<br>(1.95)                       | 2.36**<br>(1.31)  |
| NRWA Support                       | 2.65<br>(2.16)                         | 2.07*<br>(1.46)   |
| ASDWA Support                      | 4.41**<br>(1.98)                       | 3.22**<br>(1.56)  |
| NRDC Support                       | 3.07**<br>(1.65)                       | 3.02***<br>(1.30) |
| Congressional Support              | 3.86***<br>(1.56)                      |                   |
| Demographic Balance                | -.28<br>(.86)                          |                   |
| Geographic Balance                 | .53<br>(.42)                           |                   |
| Reappointment                      | 1.89<br>(1.71)                         |                   |
| Science/Engineering                | -.56<br>(.96)                          |                   |
| Water System Experience            | .58<br>(1.76)                          |                   |
| <i>Log Likelihood</i>              | -17.76                                 | -25.34            |
| <i>Likelihood Ratio Chi Square</i> | 33.24***                               | 18.45***          |
| <i>Percent Correctly Predicted</i> | 90.83                                  | 85.79             |
| <i>Percent Reduction in Error</i>  | 44.44                                  | 13.17             |
| <i>Pseudo R<sup>2</sup></i>        | .48                                    | .27               |
| <i>Number of Observations</i>      | 109                                    | 110               |

*Note:* \*\*\* = statistically significant at  $p < .01$ , one tailed. \*\* = statistically significant at  $p < .05$ , one tailed. \* = statistically significant at  $p < .10$ , one tailed. A variety of diagnostic tests indicate that multicollinearity is not a concern.

# Exporting Tables

- Let's run a simple OLS regression of some key political and demographic variables on the republican thermometer measure
  - Explanatory variables: educ black south pray pid ideology
  - Dependent variable: repherm
- Stata output



Results

```
. reg repherm educ black south pray pid ideology
```

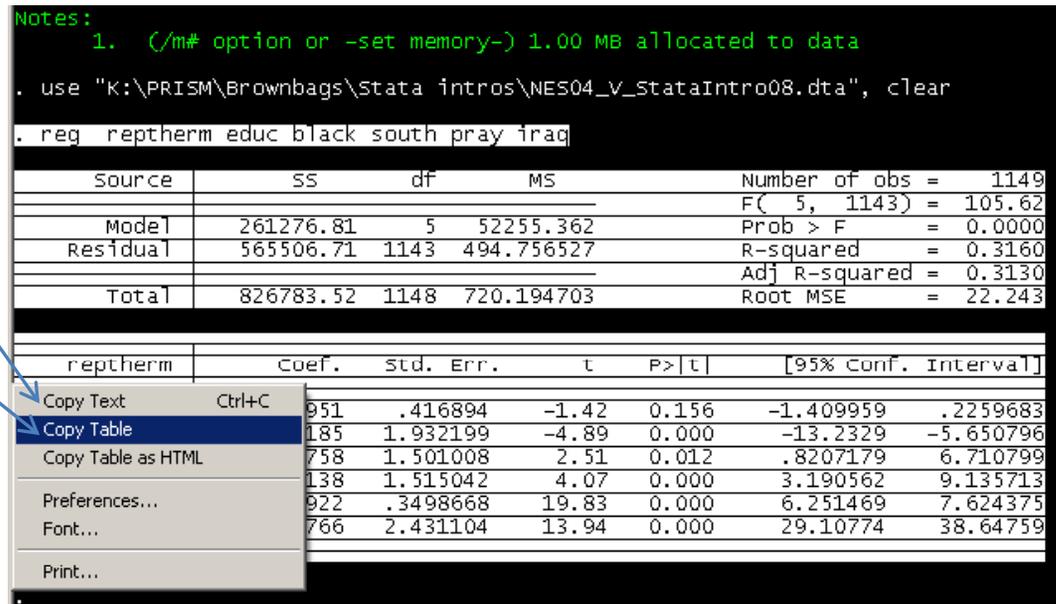
| Source   | SS         | df   | MS         |
|----------|------------|------|------------|
| Model    | 414224.451 | 6    | 69037.4086 |
| Residual | 406496.47  | 1128 | 360.369211 |
| Total    | 820720.922 | 1134 | 723.73979  |

Number of obs = 1135  
F( 6, 1128) = 191.57  
Prob > F = 0.0000  
R-squared = 0.5047  
Adj R-squared = 0.5021  
Root MSE = 18.983

| repherm  | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|----------|-----------|-----------|-------|-------|----------------------|
| educ     | -1.123615 | .3573693  | -3.14 | 0.002 | -1.824798 - .4224312 |
| black    | -1.662309 | 1.721686  | -0.97 | 0.334 | -5.040376 1.715758   |
| south    | 4.014066  | 1.285058  | 3.12  | 0.002 | 1.492693 6.535439    |
| pray     | 3.303751  | 1.319914  | 2.50  | 0.012 | .7139889 5.893513    |
| pid      | 7.49627   | .3420767  | 21.91 | 0.000 | 6.825091 8.167448    |
| ideology | 3.407201  | .5284662  | 6.45  | 0.000 | 2.370314 4.444088    |
| _cons    | 54.96673  | 1.947825  | 28.22 | 0.000 | 51.14496 58.78849    |

# Exporting Tables

- To export, highlight the table with the mouse
- Right click on the highlighted table
  - For Word: Copy Text
  - For Excel: Copy Table
- Edit in your chosen program in accord with journal specifications



Notes:  
1. (/m# option or -set memory-) 1.00 MB allocated to data

```
. use "K:\PRISM\Brownbags\Stata intros\NES04_V_StataIntro08.dta", clear  
. reg rephtherm educ black south pray iraq
```

| Source   | SS        | df   | MS         | Number of obs = 1149 |   |        |
|----------|-----------|------|------------|----------------------|---|--------|
| Model    | 261276.81 | 5    | 52255.362  | F( 5, 1143)          | = | 105.62 |
| Residual | 565506.71 | 1143 | 494.756527 | Prob > F             | = | 0.0000 |
| Total    | 826783.52 | 1148 | 720.194703 | R-squared            | = | 0.3160 |
|          |           |      |            | Adj R-squared        | = | 0.3130 |
|          |           |      |            | Root MSE             | = | 22.243 |

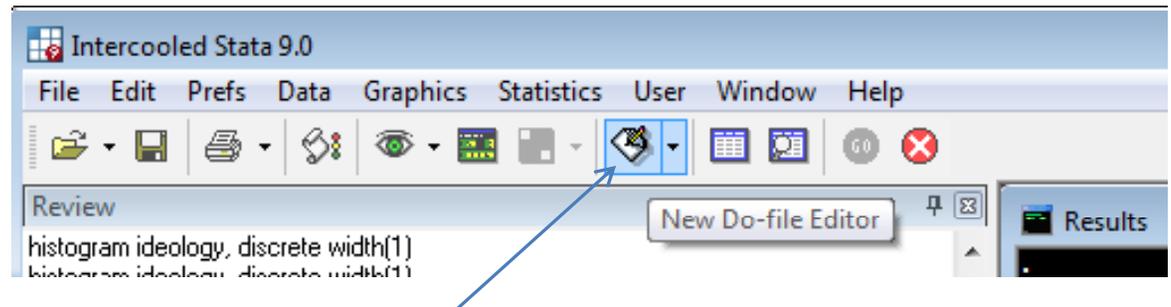
  

| rephtherm | Coef. | Std. Err. | t     | P> t  | [95% Conf. Interval] |           |
|-----------|-------|-----------|-------|-------|----------------------|-----------|
|           | .951  | .416894   | -1.42 | 0.156 | -1.409959            | .2259683  |
|           | 1.185 | 1.932199  | -4.89 | 0.000 | -13.2329             | -5.650796 |
|           | .758  | 1.501008  | 2.51  | 0.012 | .8207179             | 6.710799  |
|           | 1.138 | 1.515042  | 4.07  | 0.000 | 3.190562             | 9.135713  |
|           | .922  | .3498668  | 19.83 | 0.000 | 6.251469             | 7.624375  |
|           | .766  | 2.431104  | 13.94 | 0.000 | 29.10774             | 38.64759  |

Context menu options: Copy Text (Ctrl+C), Copy Table, Copy Table as HTML, Preferences..., Font..., Print...

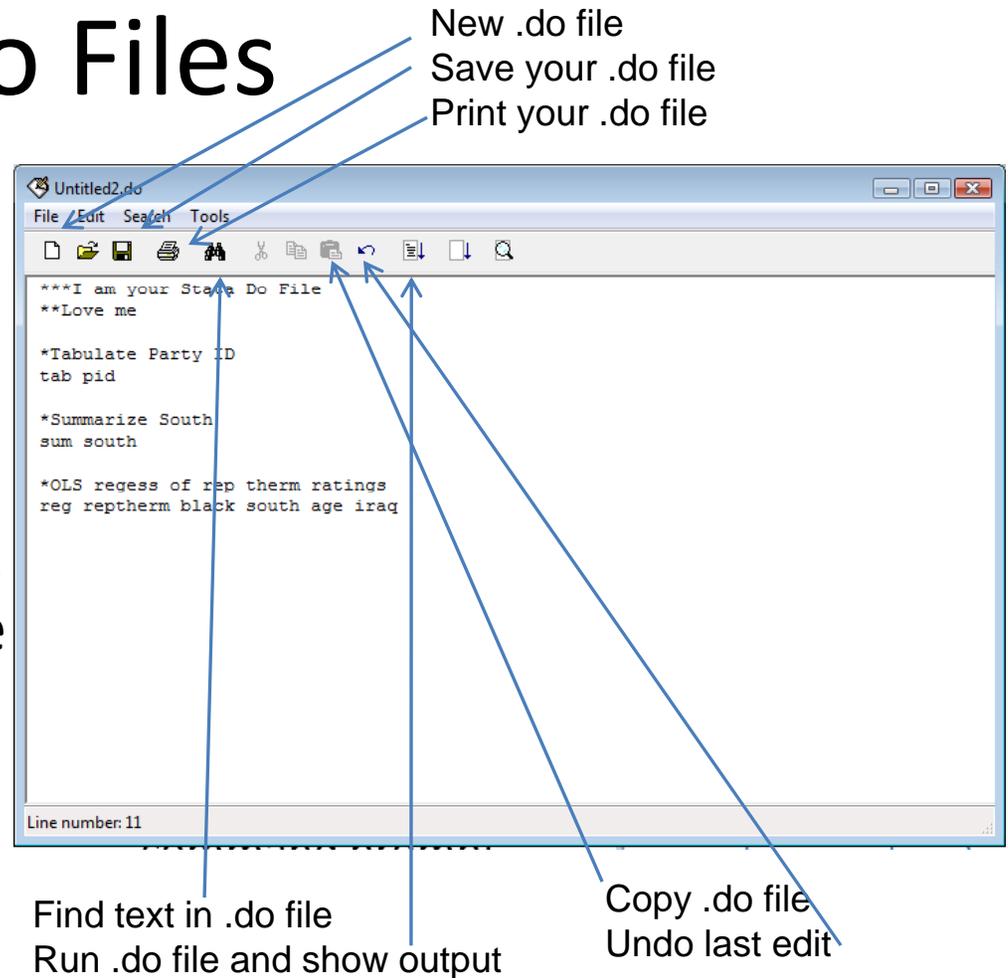
# Do Files

- We've accomplished quite a bit and we have a log file of our work to prove it
- But is there an easy way to rerun all our work?
- What if we wanted to make some small changes to our analyses and largely repeat this work?
- Use a Do file!
- Click here to open a new or saved .do file



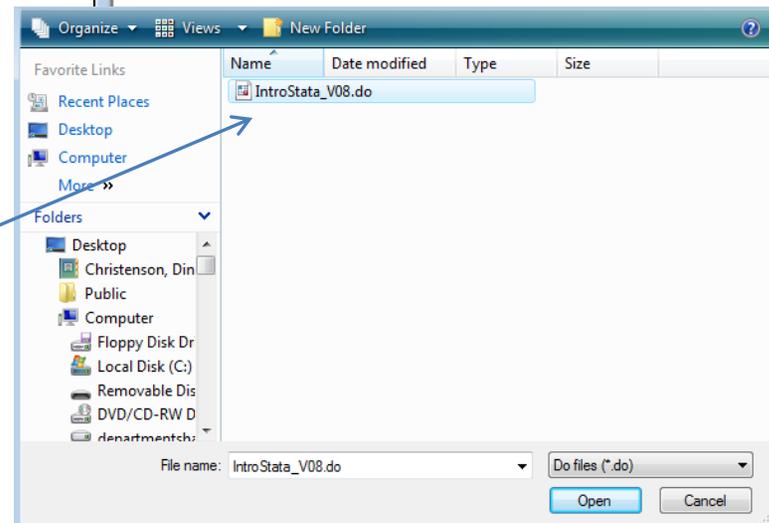
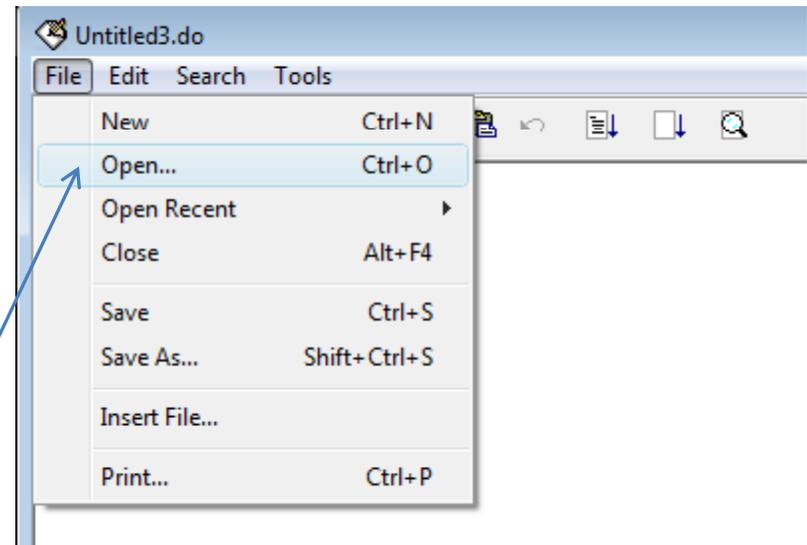
# Do Files

- A Stata do file saves text in a text editor format
  - It is often easier to create your commands in an editor than at the command prompt
  - Also easier to record your commands for future use and manipulation



# Do Files

- Typical text editing functions can be used in here: replace, copy...etc.
- Asterisk \* tells Stata not to run that line
  - Therefore annotate your .do file with titles and explanations beginning with an \*
- Let's look at all the commands used in today's presentation
  - Open a do file
  - Select Open... in do file toolbar
  - Select *IntroStata\_V08.do*
  - Click Open



# Do Files

- The .do file presents all the commands from today in a simple editor
- From here we can edit the commands
- We can run the entire series of commands at one fell swoop
  - Bring cursor to the first line and click on the Run button
- We can also select portions to run by highlighting the appropriate text and clicking the same button

```
IntroStata_V08.do
File Edit Search Tools
***Help in Stata
help contents
help tabulate
help tabulate oneway
search clarify, all

***Graphing in Stata
*Scatterplot
scatter repherm age

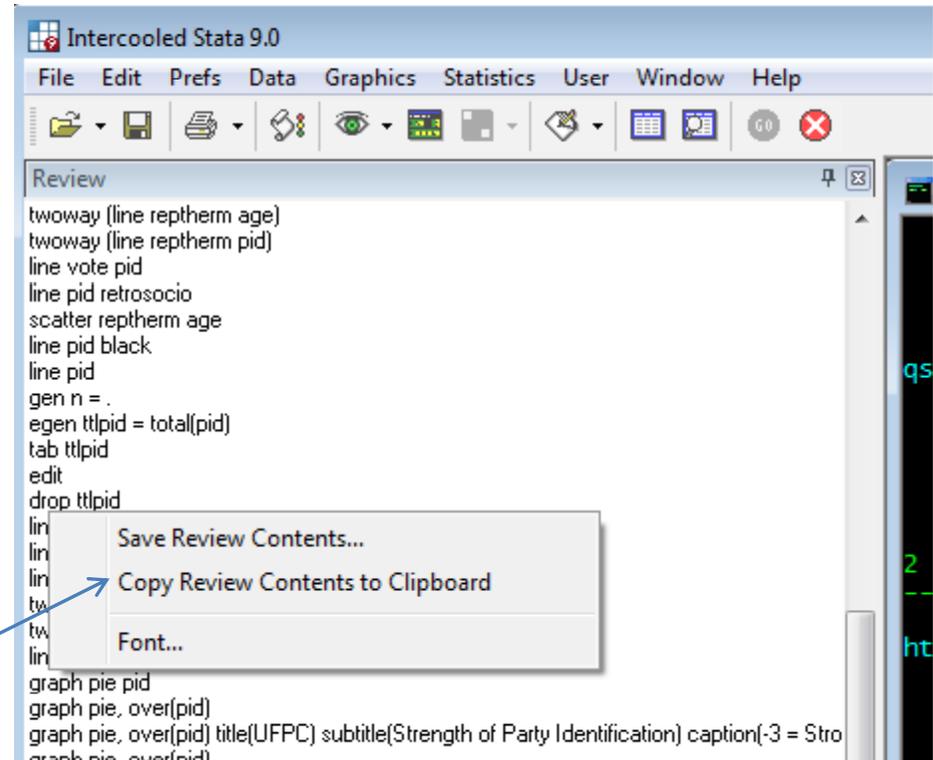
*Scatterplot - sort by another variable
scatter repherm age, by(vote)

*Histogram / Bar Chart
hist ideology, discrete width(1)

Line number: 19
```

# Do Files

- Note: if you forget to work in the do file, you can capture all your commands from the review editor:
  - Right click in the review editor → Copy Review Contents to Clipboard
  - Paste into your do file and edit



# Importing Foreign Data

- Often times we aren't lucky enough to have data in Stata's database format
- Stata's data files are stored as .dta files
  - They are just EZ-form data files
    - Used in various programs
  - Not to be confused with .dat files
    - Which are usually ASCII comma delimited and often viewed in text editors
- Not to worry!
- Beyond working with .dta files, Stata allows you to import data in various formats:
  - ASCII (.txt, .raw, .csv)
  - FDA (SAS export)
  - XML (.xml)

# Importing Foreign Data

- For example, say we wanted to use data stored at ICPSR
- [www.icpsr.umich.edu](http://www.icpsr.umich.edu)
- ICPSR has tons of data on various topics
- Hover on “Data” and select “Browse” to view their many datasets
- You can also search for a particular dataset



The screenshot shows the ICPSR website interface. The top navigation bar includes links for Data, Courses & Learning Tools, Our Research, Membership, About ICPSR, Help, and Home. A dropdown menu is open under the 'Data' link, showing options: Search, Browse, Publications Based on Our Data, Archives, How to Deposit Data, Data Use Tutorial, and Data User Help Center. Below the dropdown is a search box with a 'Search' button and a 'Web Site' checkbox. The 'QUICK LINKS' section features a dropdown menu and a paragraph stating: 'Established in 1962, ICPSR is the world's largest archive of digital social science data. We acquire, preserve, and distribute original'. The 'ANNOUNCEMENTS' section lists several dates and links, including '2008.01.20 New data releases...', '2008.01.04 Email verification added to MyData...', '2007.12.14 ICPSR to launch PK-3 data resource center...', '2007.11.19 International Data Resource Center launched...', and '2007.10.11 ICPSR Summer Internship...'. A 'more >>' button is located at the bottom right of the announcements section. The right sidebar features a 'Feature' section titled 'National Archive of Computerized Data on Aging (NACDA)' with a small image and text describing the archive's focus on gerontological research.

# Importing Foreign Data

- Today I'm interested in American state politics
- I find that ICPSR has 14 relevant datasets
- I simply select to download the dataset I'm interested in: 8655 Survey of City Council Members...
- If you are a returning user, it will request your login and password
- If you are a new user, you will have to register first
  - It's free and easy to register
  - No self-respecting methods student will make it through their first year without registering & downloading a dataset here

|   |            |
|---|------------|
| 7986 Quality of Life in the Detroit Metropolitan Area, 1975<br>Rodgers, Willard L., Robert W. Marans, et al.<br><a href="#">description</a>   <a href="#">download</a>   <a href="#">related literature</a> | 1992-02-16 |
| 7328 San Francisco Bay Region Local Politics, 1966-1967<br>Eulau, Heinz, Prewitt, Kenneth<br><a href="#">description</a>   <a href="#">download</a>   <a href="#">related literature</a>                    | 1992-02-16 |
| 8655 Survey of City Council Members in Large American Cities, 1982<br>Welch, Susan, Bledsoe, Timothy<br><a href="#">description</a>   <a href="#">download</a>   <a href="#">related literature</a>         | 2007-03-05 |
| 8709 United States Fiscal Austerity and Urban Innovation Project, 1983-1984<br>Clark, Terry Nichols, et al.<br><a href="#">description</a>   <a href="#">download</a>   <a href="#">related literature</a>  | 1992-02-16 |
| 3735 Urban Morality Issues Incidents in Ten Cities, 1990-2000: [United States]<br>Chen, Filing P.   | 2005-12-15 |

**ICPSR** | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Data | Courses & Learning Tools | Our Research | Membership | About ICPSR

### Create A New Account

In order to download data, you will need to create an account. Please fill out *all* fields in the form below.

**Email Address:**

**Note:** Before you can log in, you will need to verify your account. When you submit this form, a message will be sent to the email address specified here. It will contain a URL to confirm your email address.

**Re-type Email Address:**

**Password:**

**Re-type Password:**

**First Name:**

**Last Name:**

**Organizational Affiliation:** --select a category--

**Department/Field:** Not applicable

**Downloads:** When you download data from our site, which statistical package do you prefer?

SAS Transport (CPORT/CIMPORT) file

# Importing Foreign Data

- Download will usually allow you to import the data with various set-up files
  - These files make importing to your preferred program easier
- In this case we just want the “Stata Setup” files with the data file
- Add these to the “Data Cart” in Step 3
- Then select “Download” in Step 5 (you can review your cart in Step 4)

Description & Citation Browse Documentation **Download Data** Related Literature

## Download -- Study No. 8655

**Title:** Survey of City Council Members in Large American Cities, 19

**Principal Investigator(s):** Welch, Susan, and Timothy Bledsoe

### Step 1. Select available data formats

Documentation Only  
 ASCII Data File + SAS Setup Files  
 ASCII Data File + SPSS Setup Files  
 ASCII Data File + Stata Setup Files  
 All Files

Documentation files are automatically included with download.  
[more information on data format types](#)

### Step 2. Select datasets

DS1: Survey of City Council Members in Large American Cities

### Step 3. Add to data cart

### Step 4. Review cart (optional)

### Step 5. Download cart contents

**0 file(s); 0KB**

# Importing Foreign Data

- After agreeing to their terms and conditions
  - The data files are compressed in a zip drive
  - You are prompted to open or save the files
- Save the drive in your preferred folder

If you do not agree, you can click on the "I Do Not Agree" button to return to the home page.

The study you have requested is archived and distributed by ICPSR. By downloading the data...

Any intentional identification or disclosure of a person or establishment violates the assurance providers of the information. Therefore, users of data obtained from the ICPSR archive and/o Archives agree:

- To use these datasets solely for statistical analysis and reporting of aggregated information of specific individuals or organizations, except when identification is authorized in writing
- To make no use of the identity of any person or establishment discovered inadvertently, such discovery
- To produce no links among ICPSR datasets or among ICPSR data and other datasets if organizations not already known in all of the datasets before the datasets were linked
- To not redistribute or sell the data, except in the following exceptions: a) if you are assisting users with other research that involves analyzing ICPSR data in approved ways, you must obtain prior approval from ICPSR

ICPSR further asks that any publications based on ICPSR data should be cited as follows: These citations acknowledge the source of the data and the ICPSR's role in providing the data.

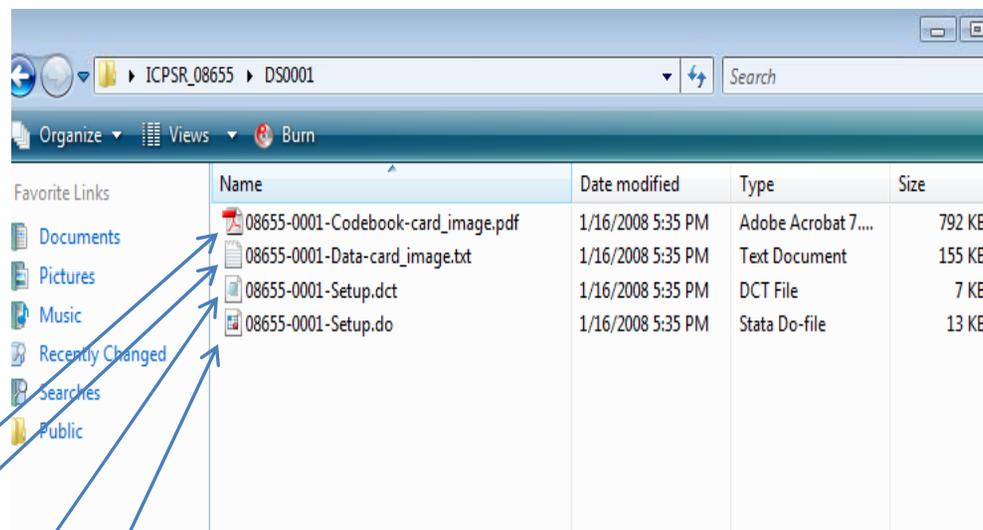
In addition, the user acknowledges that the user assumes full responsibility for use of the data.

By continuing past this point to the next page, you agree to the requirements and give your assurance that the use of statistical data obtained from ICPSR will conform to widely-accepted standards of practice and legal restrictions that are intended to protect research subjects.

When you've finished downloading, you can return to the [front page](#).

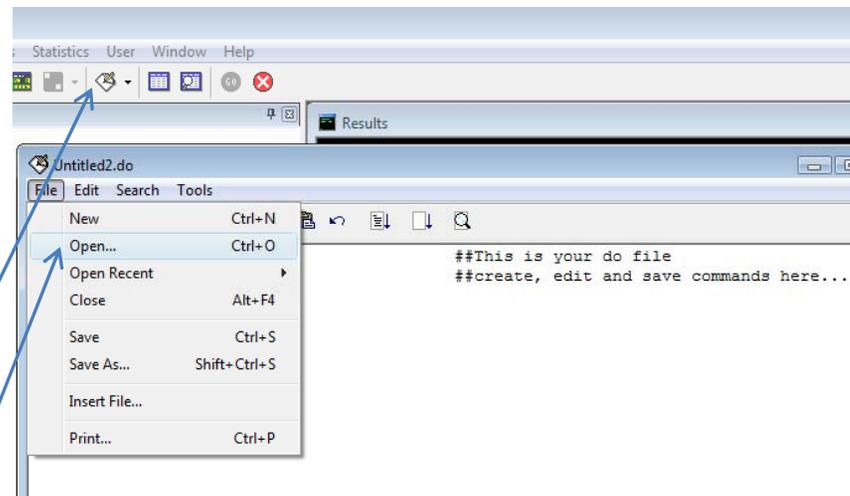
# Importing Foreign Data

- Now we have the data and setup files in a zip drive on our computer
  - Extract the contents from your zip drive
  - View the contents
    - Codebook as .pdf
    - Data as .txt
    - Setup dictionary as .dct
    - Setup do file as .do



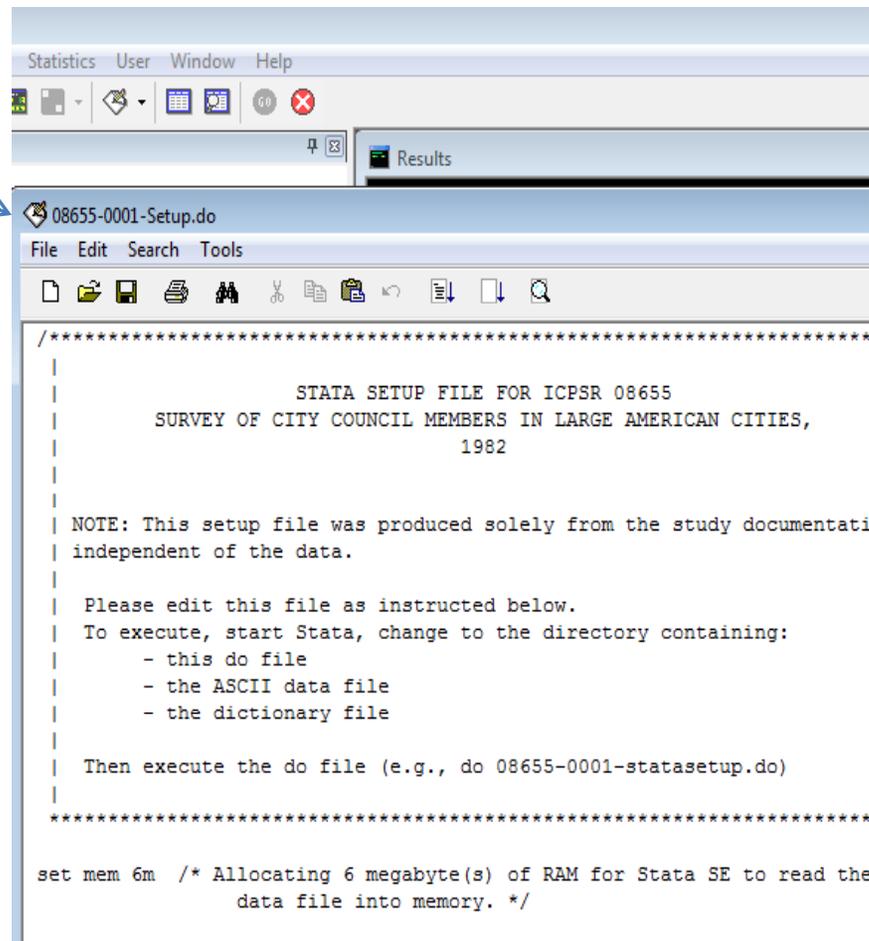
# Importing Foreign Data

- Let's return to your Stata GUI
- Type `clear` to completely reset your data
  - Doing so deletes any variables you have stored or you have created
- Click here to open a “do file”
  - In the do file, select open
  - Browse for the setup do file:
  - *08655-0001-Setup.do*



# Importing Foreign Data

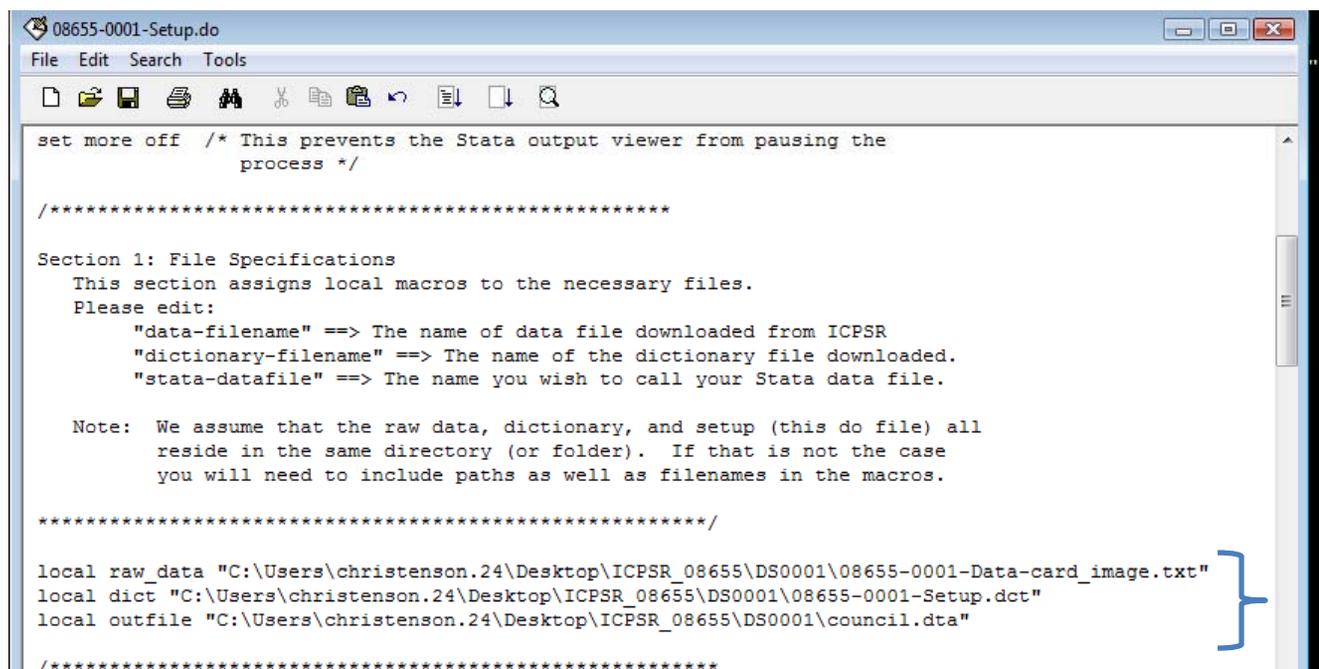
- The setup do file
- This file will define and label your data for the Stata editor by calling
  - The dataset
  - A corresponding dictionary file



```
Statistics User Window Help
08655-0001-Setup.do
File Edit Search Tools
*****
|
|          STATA SETUP FILE FOR ICPSR 08655
|          SURVEY OF CITY COUNCIL MEMBERS IN LARGE AMERICAN CITIES,
|          1982
|
| NOTE: This setup file was produced solely from the study documentati
| independent of the data.
|
| Please edit this file as instructed below.
| To execute, start Stata, change to the directory containing:
|   - this do file
|   - the ASCII data file
|   - the dictionary file
|
| Then execute the do file (e.g., do 08655-0001-statasetup.do)
|
|*****
set mem 6m /* Allocating 6 megabyte(s) of RAM for Stata SE to read the
          data file into memory. */
```

# Importing Foreign Data

- Edit the do file to pull from the appropriate folder
  - You must tell it where to find the raw data (.txt) and the dictionary file (.dct) → we stored it on the desktop
  - And you must specify the name of the output file (.dta)



```
08655-0001-Setup.do
File Edit Search Tools

set more off /* This prevents the Stata output viewer from pausing the
              process */

/*****

Section 1: File Specifications
This section assigns local macros to the necessary files.
Please edit:
    "data-filename" ==> The name of data file downloaded from ICPSR
    "dictionary-filename" ==> The name of the dictionary file downloaded.
    "stata-datafile" ==> The name you wish to call your Stata data file.

Note: We assume that the raw data, dictionary, and setup (this do file) all
      reside in the same directory (or folder). If that is not the case
      you will need to include paths as well as filenames in the macros.

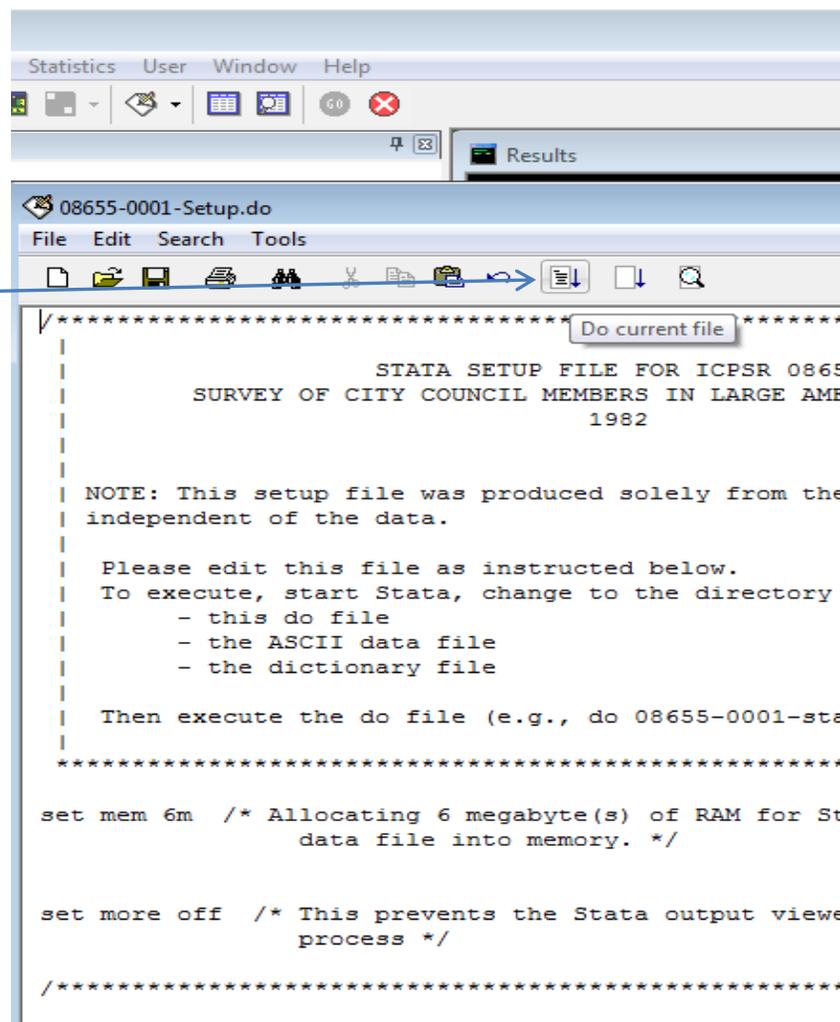
*****/

local raw_data "C:\Users\christenson.24\Desktop\ICPSR_08655\DS0001\08655-0001-Data-card_image.txt"
local dict "C:\Users\christenson.24\Desktop\ICPSR_08655\DS0001\08655-0001-Setup.dct"
local outfile "C:\Users\christenson.24\Desktop\ICPSR_08655\DS0001\council.dta"

/*****
```

# Importing Foreign Data

- Once we've told the do file editor where to find the dictionary and data...
- We run the do file
- In result viewer, Stata returns in green
  - file C:\council.dta saved
  - Or, if you got it wrong, it returns an error code in red
  - If wrong, make sure you specified the right directory



```
***** Do current file *****
                                STATA SETUP FILE FOR ICPSR 0865
                                SURVEY OF CITY COUNCIL MEMBERS IN LARGE AME
                                1982

NOTE: This setup file was produced solely from the
independent of the data.

Please edit this file as instructed below.
To execute, start Stata, change to the directory
- this do file
- the ASCII data file
- the dictionary file

Then execute the do file (e.g., do 08655-0001-sta
*****

set mem 6m /* Allocating 6 megabyte(s) of RAM for St
data file into memory. */

set more off /* This prevents the Stata output viewe
process */

/*****
```



# Other Importing Options

- SPSS data (.sav) can be easily exported to Stata format (.dta) from SPSS
  - In SPSS, just click Save As and select the appropriate Stata version (an export wizard is now available in SPSS as well)
  - FYI: You can also export from SPSS to just about anything else (SAS, Excel, ASCII, dBase & SAS)
- The PRL lab has Stat/Transfer 
  - An easy way to move data between packages and into different databases
  - Especially good with large and labeled databases

# Congratulations

- By now you can move comfortably around Stata
- You can
  - Keep a log of your work
  - Use Stata as a statistics calculator
  - Create variables
  - Load a Stata dataset
  - Examine your data
  - Run some descriptive functions
  - Make basic graphs
  - Search for help on commands and packages
  - Export Stata output into your preferred document
  - Create, edit, run and save commands from a do file
  - And even import foreign datasets

# Remember

1. Begin by opening a log
  - Always keep a log
2. To increase memory for large datasets, type `set mem 100m`
3. Begin all analyses with simple descriptives
  - Know your data
4. Utilize `gen` to generate variables
  - The `egen` command is a helpful extension to `gen`
5. Usefulness of the Review window
  - Don't need to retype the command (just click from the review)
  - Also helpful are the page up/down keys within the command prompt
6. `__n` is Stata programming code for observation number
7. Use `.do` files
  - Annotate your do files utilizing the `*`

# See You Next Time

- PRISM's next brownbag  
*Contemporary Methods of Ideal Point Estimation*  
Presenter: Josh Clinton of [Princeton University](#)  
January 30, 2008  
12:00-1:00pm
- PRISM's Spring brownbag  
*Bayesian Inference with WinBUGS*  
Presenters: Dino Christenson & Scott Powell  
Date & Time TBA (Spring 2008)  
Updates at <http://polisci.osu.edu/prism/luncheons.htm>
- PRISM's next methods lunch  
February 5<sup>th</sup>, 12 noon