INTRODUCTION
to
Stata®

Statistical Software

Program in Statistics and Methodology (PRISM)

Daniel Blake & Benjamin Jones

January 15, 2010

While we are waiting…

- Everyone who wishes to work along with the presentation should log onto a machine
- Launch Stata, there should be an icon on the desktop
- Do not open anything else yet

Outline

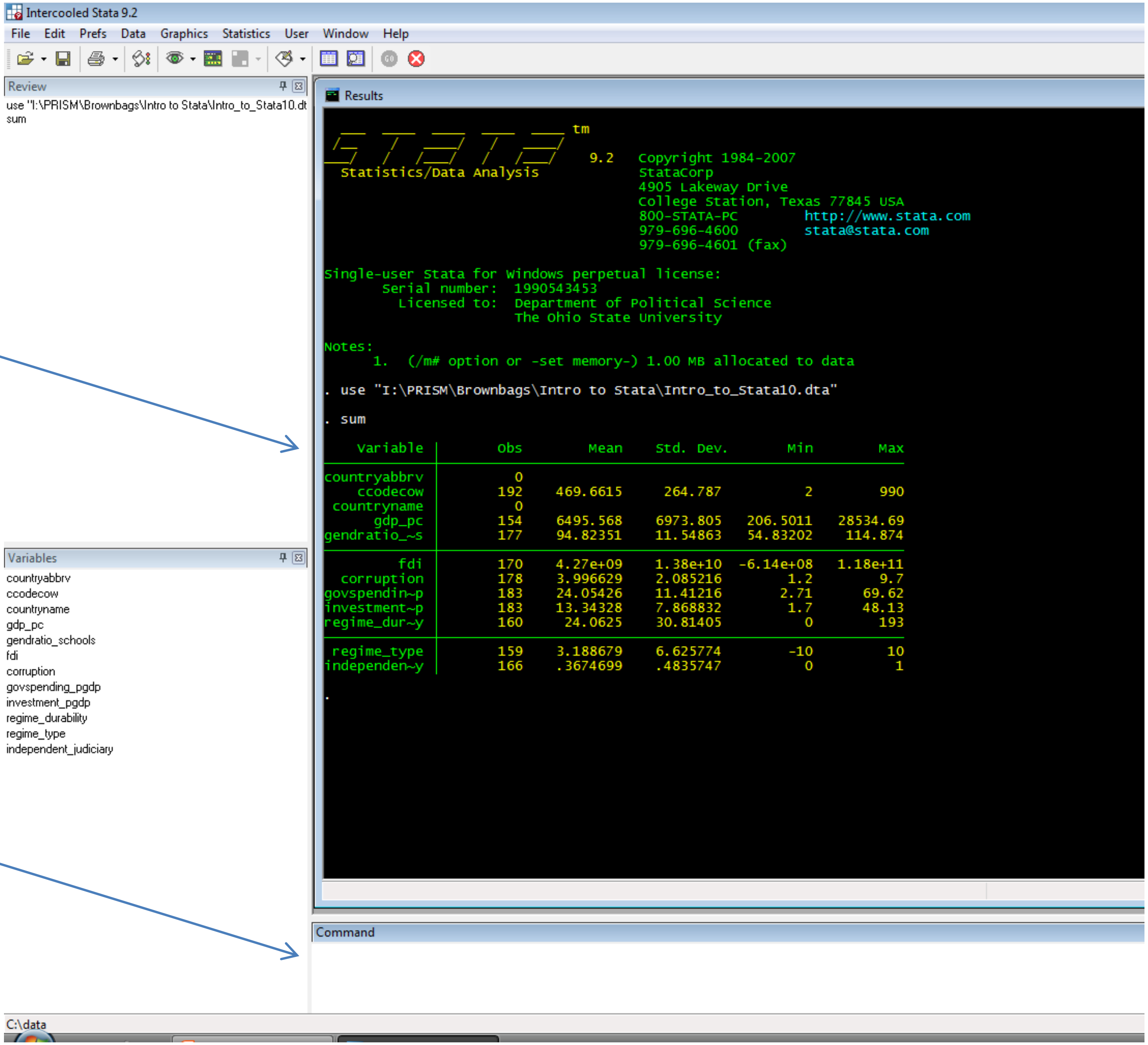
- By the end of this presentation you should be able to:

1. Navigate around the Stata interface
2. Import data into Stata
3. Generate, manipulate and obtain descriptive statistics for variables
4. Create basic graphs and plots
5. Use and appreciate the wonder of Do files
6. Estimate a basic OLS model and use post-estimation commands
7. Keep a log of your work

Stata's Interface: Windows

Four Main Windows:

1. **Review**: Lists past commands
2. **Results**: Shows recently obtained results
3. **Variables**: Names of all variables in data set
4. **Command**: Where commands are entered



```
use "I:\PRISM\Brownbags\Intro to Stata\Intro_to_Stata10.dt
sum
```

```
countryabbrv
ccodecow
countryname
gdp_pc
gendratio_schools
fdi
corruption
govspending_pgdp
investment_pgdp
regime_durability
regime_type
independent_judiciary
```

```
  ___  ____  ____  ____  ____ tm
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   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 "I:\PRISM\Brownbags\Intro to Stata\Intro_to_Stata10.dta"

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
countryabbrv |         0
    ccodecow |       192    469.6615     264.787          2        990
 countryname |         0
      gdp_pc |       154    6495.568    6973.805   206.5011   28534.69
gendratio_~s |       177    94.82351    11.54863   54.83202    114.874
-------------+--------------------------------------------------------
         fdi |       170    4.27e+09    1.38e+10  -6.14e+08   1.18e+11
  corruption |       178    3.996629    2.085216        1.2        9.7
govspendin~p |       183    24.05426    11.41216       2.71      69.62
investment~p |       183    13.34328    7.868832        1.7      48.13
regime_dur~y |       160     24.0625    30.81405          0        193
-------------+--------------------------------------------------------
 regime_type |       159    3.188679    6.625774        -10         10
independen~y |       166    .3674699    .4835747          0          1

.
```

Stata’s Interface: Menus

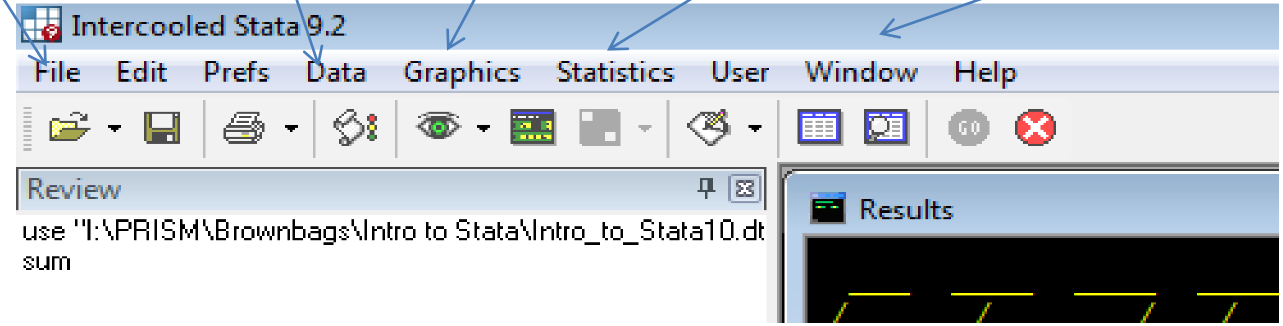
File: Open & save data, log file management.

Data: Data manipulation functions (e.g. sorting, merging)

Graphics: Multiple graphing options

Statistics: Statistical modeling functions

Window: Organizing and bringing up windows



```
use "I:\PRISM\Brownbags\Intro to Stata\Intro_to_Stata10.dt
sum
```

- Menus offer a point-and-click alternative to typing commands directly into Stata’s command window

Stata’s Interface: Toolbar

Begin a new log file

Bring up the Help Viewer

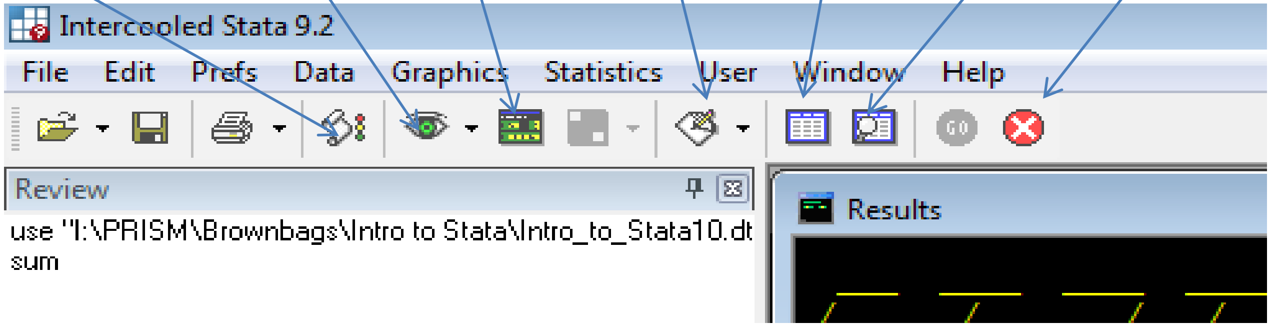
Bring up the Results Window

Begin a new Do file or open an existing one

Opens Data Viewer for Editing

Opens Data Viewer

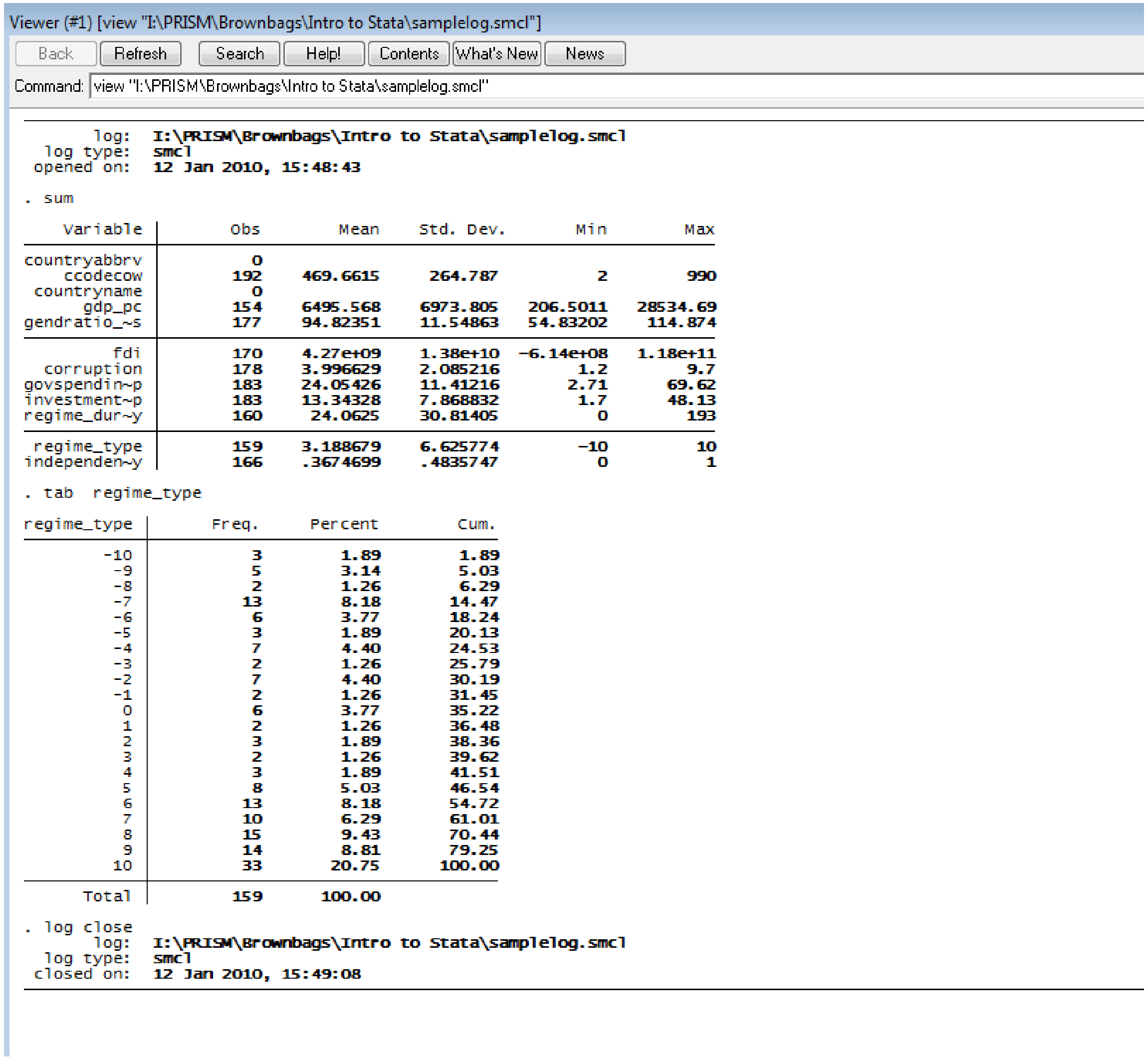
Stop calculations/ processing



```
use "I:\PRISM\Brownbags\Intro to Stata\Intro_to_Stata10.dt
sum
```

Log Files

- Log Files help you keep track of your work in Stata
- After you begin a log, every command, model and all results that appear in the Results window are logged.
- Log files can be saved and are portable.
- Log files can be appended, but their contents cannot be modified.


Viewer (#1) [view "I:\PRISM\Brownbags\Intro to Stata\samplelog.smcl"]
Back
Refresh
Search
Help!
Contents
What's New
News

Command: view "I:\PRISM\Brownbags\Intro to Stata\samplelog.smcl"

```
------------------------------------------------------------------------------
      log:  I:\PRISM\Brownbags\Intro to Stata\samplelog.smcl
 log type:  smcl
opened on:  12 Jan 2010, 15:48:43

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
countryabbrv |         0
    ccodecow |       192    469.6615     264.787          2        990
 countryname |         0
      gdp_pc |       154    6495.568    6973.805   206.5011   28534.69
gendratio_~s |       177    94.82351    11.54863   54.83202    114.874
-------------+--------------------------------------------------------
         fdi |       170    4.27e+09    1.38e+10  -6.14e+08   1.18e+11
  corruption |       178    3.996629    2.085216        1.2        9.7
govspendin~p |       183    24.05426    11.41216       2.71      69.62
investment~p |       183    13.34328    7.868832        1.7      48.13
regime_dur~y |       160     24.0625    30.81405          0        193
-------------+--------------------------------------------------------
 regime_type |       159    3.188679    6.625774        -10         10
independen~y |       166    .3674699    .4835747          0          1

. tab  regime_type

regime_type |      Freq.     Percent        Cum.
------------+-----------------------------------
        -10 |          3        1.89        1.89
         -9 |          5        3.14        5.03
         -8 |          2        1.26        6.29
         -7 |         13        8.18       14.47
         -6 |          6        3.77       18.24
         -5 |          3        1.89       20.13
         -4 |          7        4.40       24.53
         -3 |          2        1.26       25.79
         -2 |          7        4.40       30.19
         -1 |          2        1.26       31.45
          0 |          6        3.77       35.22
          1 |          2        1.26       36.48
          2 |          3        1.89       38.36
          3 |          2        1.26       39.62
          4 |          3        1.89       41.51
          5 |          8        5.03       46.54
          6 |         13        8.18       54.72
          7 |         10        6.29       61.01
          8 |         15        9.43       70.44
          9 |         14        8.81       79.25
         10 |         33       20.75      100.00
------------+-----------------------------------
      Total |        159      100.00

. log close
      log:  I:\PRISM\Brownbags\Intro to Stata\samplelog.smcl
 log type:  smcl
closed on:  12 Jan 2010, 15:49:08
------------------------------------------------------------------------------
```

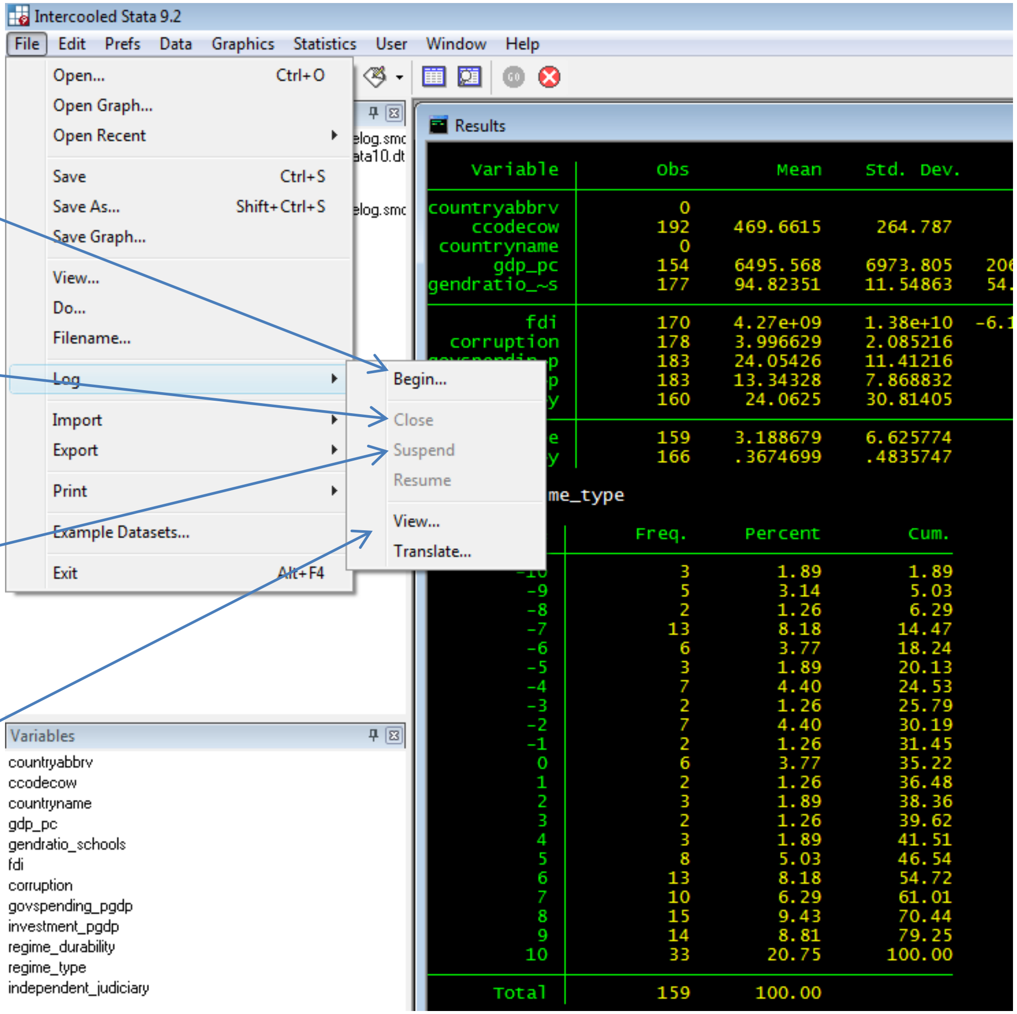
Log Files

To start a log file: File > Log > Begin

To end an ongoing log: File > Log > Close

To pause an ongoing log: File > Log > Suspend

To view a log file: File > Log > View



```
    Variable |       Obs        Mean    Std. Dev.
-------------+--------------------------------------
countryabbrv |         0
    ccodecow |       192    469.6615     264.787
 countryname |         0
      gdp_pc |       154    6495.568    6973.805    206
gendratio_~s |       177    94.82351    11.54863    54.
-------------+--------------------------------------
         fdi |       170    4.27e+09    1.38e+10   -6.1
  corruption |       178    3.996629    2.085216
govspendin~p |       183    24.05426    11.41216
           p |       183    13.34328    7.868832
           y |       160     24.0625    30.81405
-------------+--------------------------------------
           e |       159    3.188679    6.625774
           y |       166    .3674699    .4835747

me_type

             |      Freq.     Percent        Cum.
-------------+-----------------------------------
         -10 |          3        1.89        1.89
          -9 |          5        3.14        5.03
          -8 |          2        1.26        6.29
          -7 |         13        8.18       14.47
          -6 |          6        3.77       18.24
          -5 |          3        1.89       20.13
          -4 |          7        4.40       24.53
          -3 |          2        1.26       25.79
          -2 |          7        4.40       30.19
          -1 |          2        1.26       31.45
           0 |          6        3.77       35.22
           1 |          2        1.26       36.48
           2 |          3        1.89       38.36
           3 |          2        1.26       39.62
           4 |          3        1.89       41.51
           5 |          8        5.03       46.54
           6 |         13        8.18       54.72
           7 |         10        6.29       61.01
           8 |         15        9.43       70.44
           9 |         14        8.81       79.25
          10 |         33       20.75      100.00
-------------+-----------------------------------
       Total |        159      100.00
```

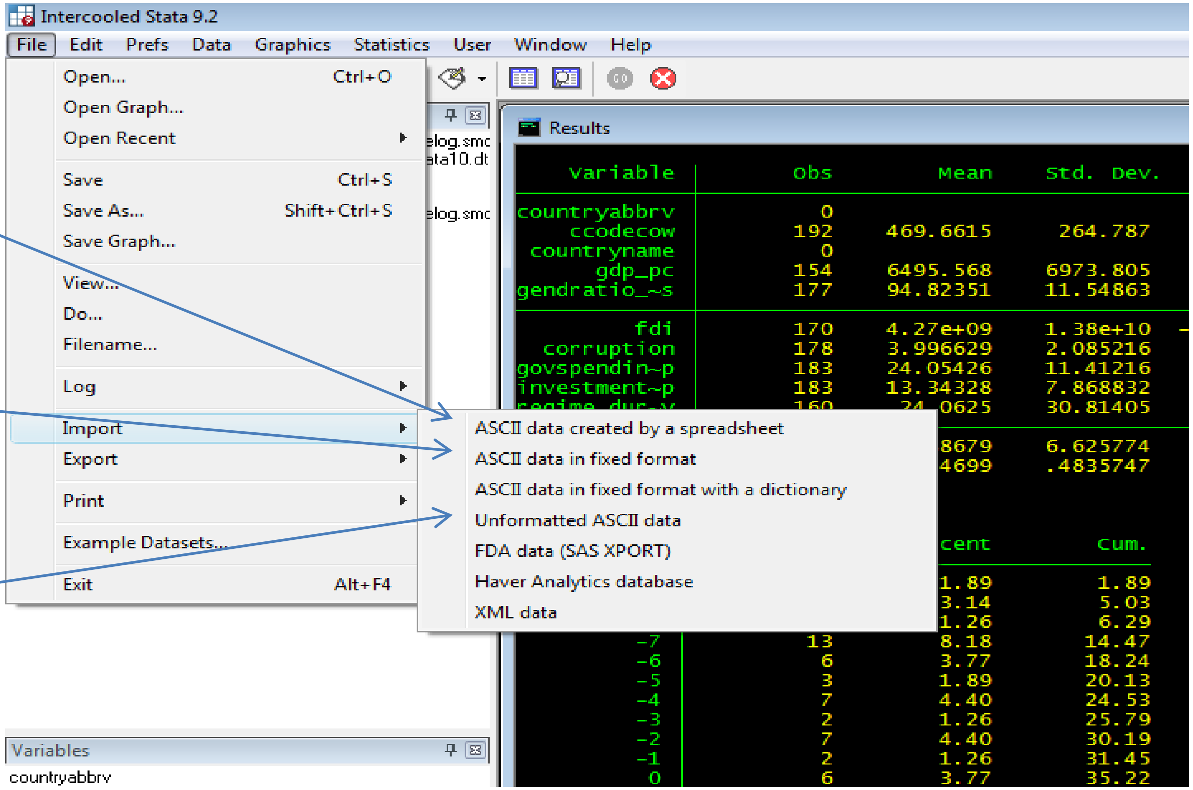
Getting Data into Stata

- Three options:
  1. Entering it by hand using Stata’s Data Editor
  2. Opening existing data files formated specifically for Stata. These files end in <.dta>.
     - File > Open
  3. Import data that is not in Stata format but that Stata can understand (e.g. comma-separated files <.csv>; tab-delimited data; space delimited data).

N.B. If none of the above work → STAT Transfer

Importing Data (non .dta files)

- You must begin with an empty data set before importing data (To do this type `clear`. Note: All unsaved data and changes will be lost.)



```
    Variable |       Obs        Mean    Std. Dev.
-------------+----------------------------------
countryabbrv |         0
    ccodecow |       192    469.6615     264.787
 countryname |         0
      gdp_pc |       154    6495.568    6973.805
gendratio_~s |       177    94.82351    11.54863
-------------+----------------------------------
         fdi |       170    4.27e+09    1.38e+10
  corruption |       178    3.996629    2.085216
govspendin~p |       183    24.05426    11.41216
investment~p |       183    13.34328    7.868832
regime_dur~v |       160     24.0625    30.81405
-------------+----------------------------------
                              8679    6.625774
                              4699    .4835747

                              cent        Cum.
---------------------------------------------
                              1.89        1.89
                              3.14        5.03
                              1.26        6.29
          -7 |        13      8.18       14.47
          -6 |         6      3.77       18.24
          -5 |         3      1.89       20.13
          -4 |         7      4.40       24.53
          -3 |         2      1.26       25.79
          -2 |         7      4.40       30.19
          -1 |         2      1.26       31.45
           0 |         6      3.77       35.22
```

Importing data created by a spreadsheet*

Importing formatted data

Importing formatted data

*You cannot import Excel (.xls) files directly. To get data from Excel to Stata, save your Excel spreadsheet as a comma separated file (.csv). Then import it using this option.

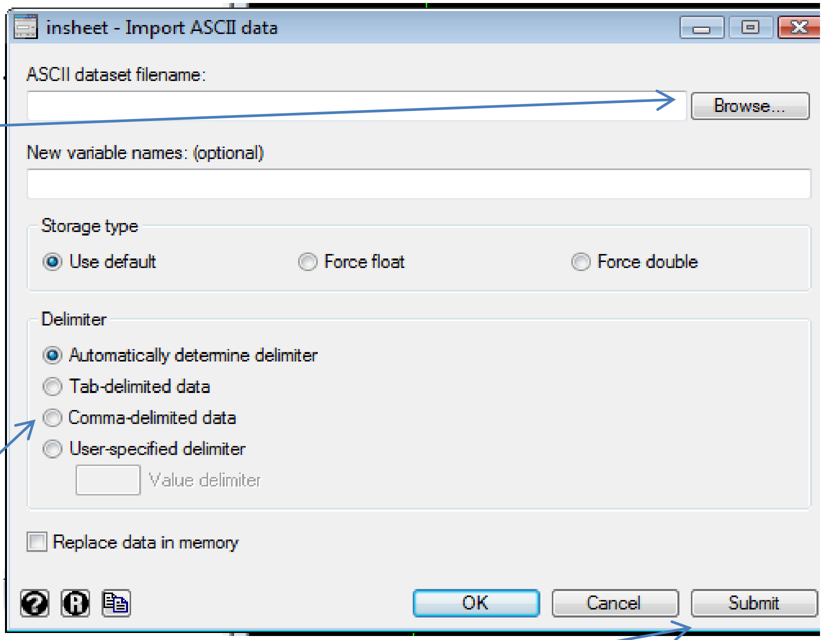
Importing a Comma Separated File

1. Browse and find the file you wish to import.

2. Change file type to .csv in the “Open File” window.

3.Select comma-delimited data

4. Click Submit



insheet - Import ASCII data

ASCII dataset filename:

New variable names: (optional)

Storage type

● Use default ○ Force float ○ Force double

Delimiter

● Automatically determine delimiter

○ Tab-delimited data

○ Comma-delimited data

○ User-specified delimiter

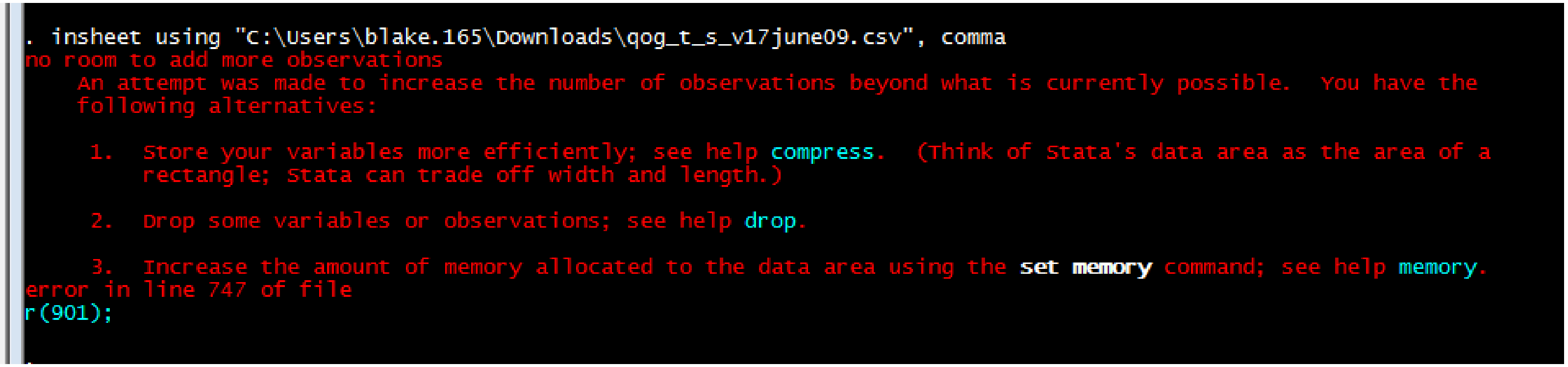
Value delimiter

□ Replace data in memory

- When you import data into Stata and save, it is saved as a Stata data file.

Memory & Large Data Files

- When importing or opening large data files, you may get the following error:



```
. insheet using "C:\Users\blake.165\Downloads\qog_t_s_v17june09.csv", comma
no room to add more observations
    An attempt was made to increase the number of observations beyond what is currently possible.  You have the
    following alternatives:

     1.  Store your variables more efficiently; see help compress.  (Think of Stata's data area as the area of a
         rectangle; Stata can trade off width and length.)

     2.  Drop some variables or observations; see help drop.

     3.  Increase the amount of memory allocated to the data area using the set memory command; see help memory.
error in line 747 of file
r(901);
```

- To get around this, you need to increase the size of memory using the `set memory` command. For example, to increase the memory available to 500 megabytes:

```
clear
set memory 500m
```

Data Editor

Sort: Sort data by selected variable

<</ >>: Move selected variable to the first or last column

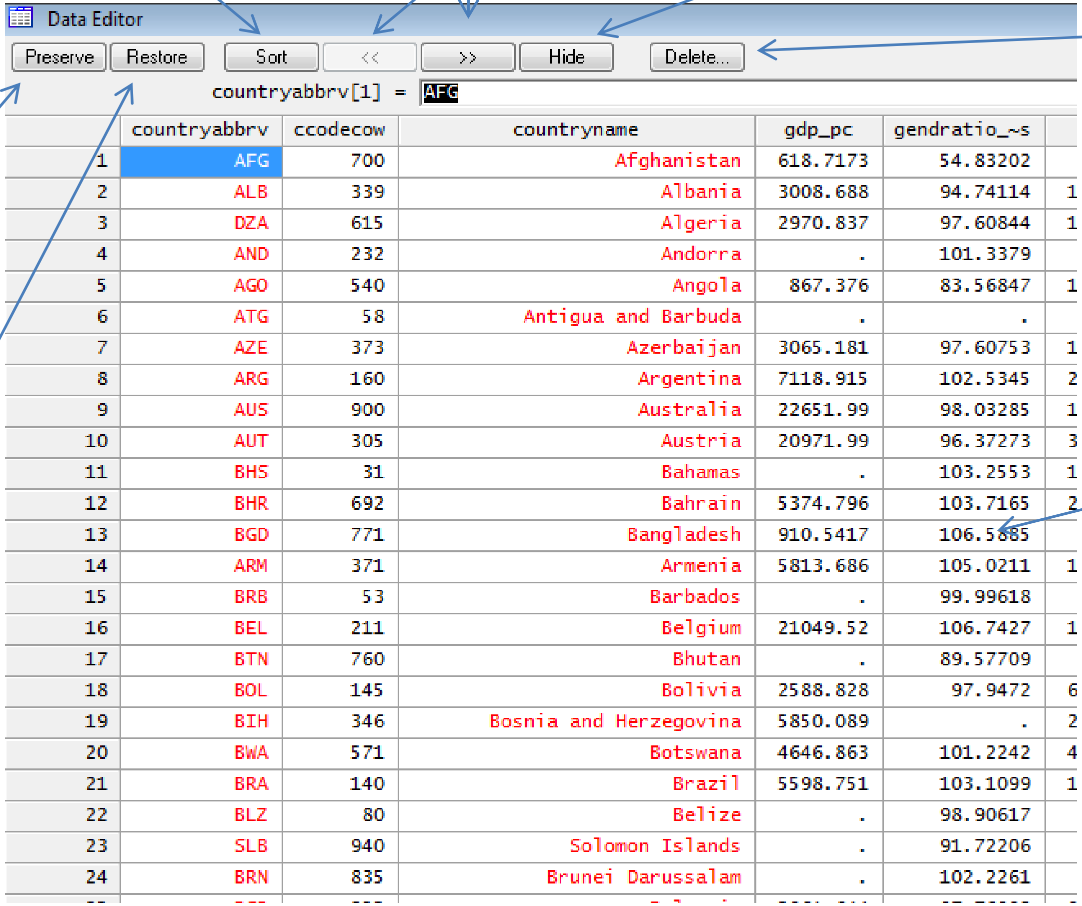
Hide: Hide selected variable

Delete: Delete selected variable or observation - **Choose Wisely or else!**)

Preserve: Save changes made in the data editor

Restore: Undo all changes made in the data editor since last save/preserve

Select cells to edit their contents directly



countryabbrv[1] = AFG

	countryabbrv	ccodecow	countryname	gdp_pc	gendratio_~s	
1	AFG	700	Afghanistan	618.7173	54.83202	
2	ALB	339	Albania	3008.688	94.74114	1
3	DZA	615	Algeria	2970.837	97.60844	1
4	AND	232	Andorra	.	101.3379	
5	AGO	540	Angola	867.376	83.56847	1
6	ATG	58	Antigua and Barbuda	.	.	
7	AZE	373	Azerbaijan	3065.181	97.60753	1
8	ARG	160	Argentina	7118.915	102.5345	2
9	AUS	900	Australia	22651.99	98.03285	1
10	AUT	305	Austria	20971.99	96.37273	3
11	BHS	31	Bahamas	.	103.2553	1
12	BHR	692	Bahrain	5374.796	103.7165	2
13	BGD	771	Bangladesh	910.5417	106.5885	
14	ARM	371	Armenia	5813.686	105.0211	1
15	BRB	53	Barbados	.	99.99618	
16	BEL	211	Belgium	21049.52	106.7427	1
17	BTN	760	Bhutan	.	89.57709	
18	BOL	145	Bolivia	2588.828	97.9472	6
19	BIH	346	Bosnia and Herzegovina	5850.089	.	2
20	BWA	571	Botswana	4646.863	101.2242	4
21	BRA	140	Brazil	5598.751	103.1099	1
22	BLZ	80	Belize	.	98.90617	
23	SLB	940	Solomon Islands	.	91.72206	
24	BRN	835	Brunei Darussalam	.	102.2261	

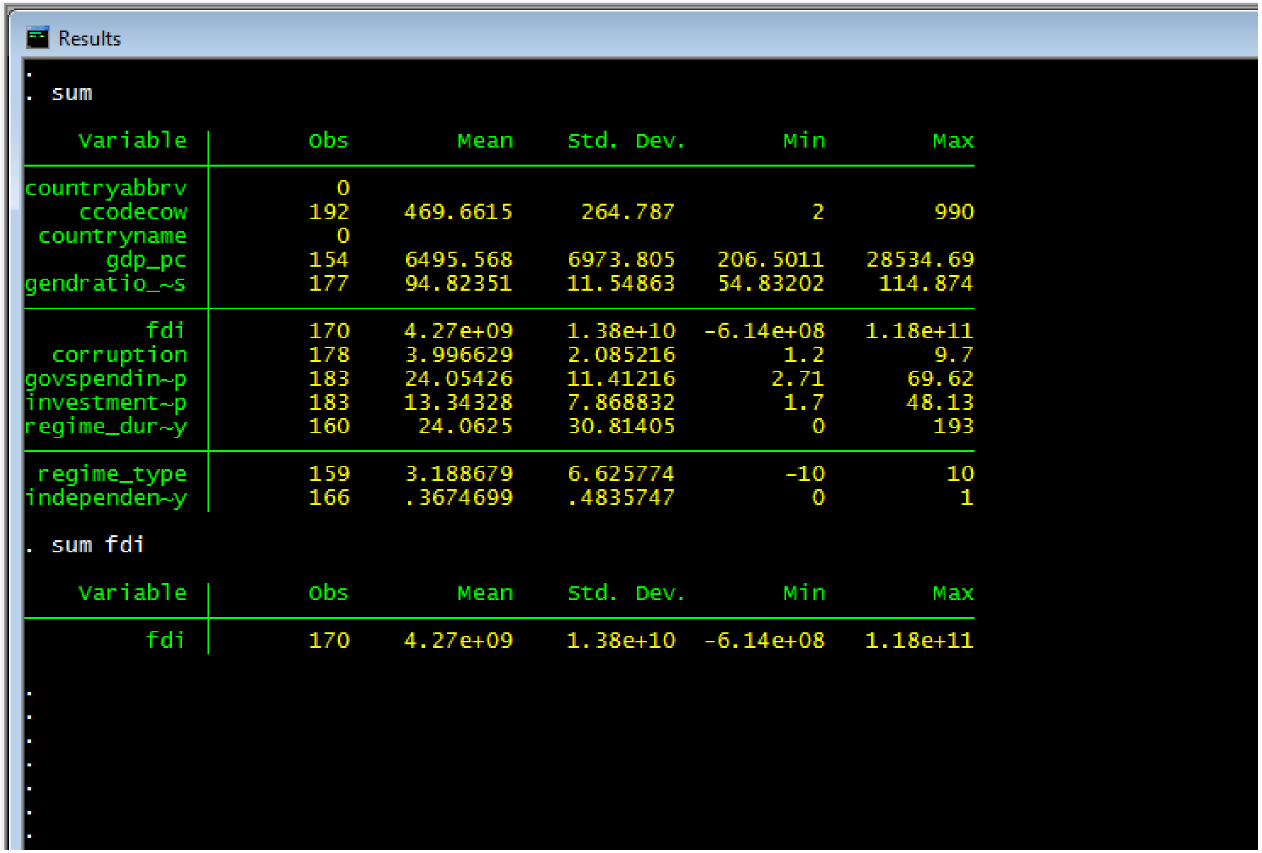
Describing the Data

- Important Commands:
  - `sum` *variables*
    - Provides summary statistics
  - `tab` *variable*
    - Provides table showing distribution of values
  - `tab` *variable1 variable2*
    - Provides a cross-tab of the two variables
  - `describe`
    - Provides a summary of data types in your data

Summary Statistics

To get a summary of all variables, simply type: `sum`.

To get a summary for just one variable (e.g. FDI), type: `sum fdi`.



```
.
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
countryabbrv |         0
    ccodecow |       192    469.6615     264.787          2        990
 countryname |         0
      gdp_pc |       154    6495.568    6973.805   206.5011   28534.69
gendratio_~s |       177    94.82351    11.54863   54.83202    114.874
-------------+--------------------------------------------------------
         fdi |       170    4.27e+09    1.38e+10  -6.14e+08   1.18e+11
  corruption |       178    3.996629    2.085216        1.2        9.7
govspendin~p |       183    24.05426    11.41216       2.71      69.62
investment~p |       183    13.34328    7.868832        1.7      48.13
regime_dur~y |       160     24.0625    30.81405          0        193
-------------+--------------------------------------------------------
 regime_type |       159    3.188679    6.625774        -10         10
independen~y |       166    .3674699    .4835747          0          1

. sum fdi

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         fdi |       170    4.27e+09    1.38e+10  -6.14e+08   1.18e+11

.
.
.
.
.
.
.
```

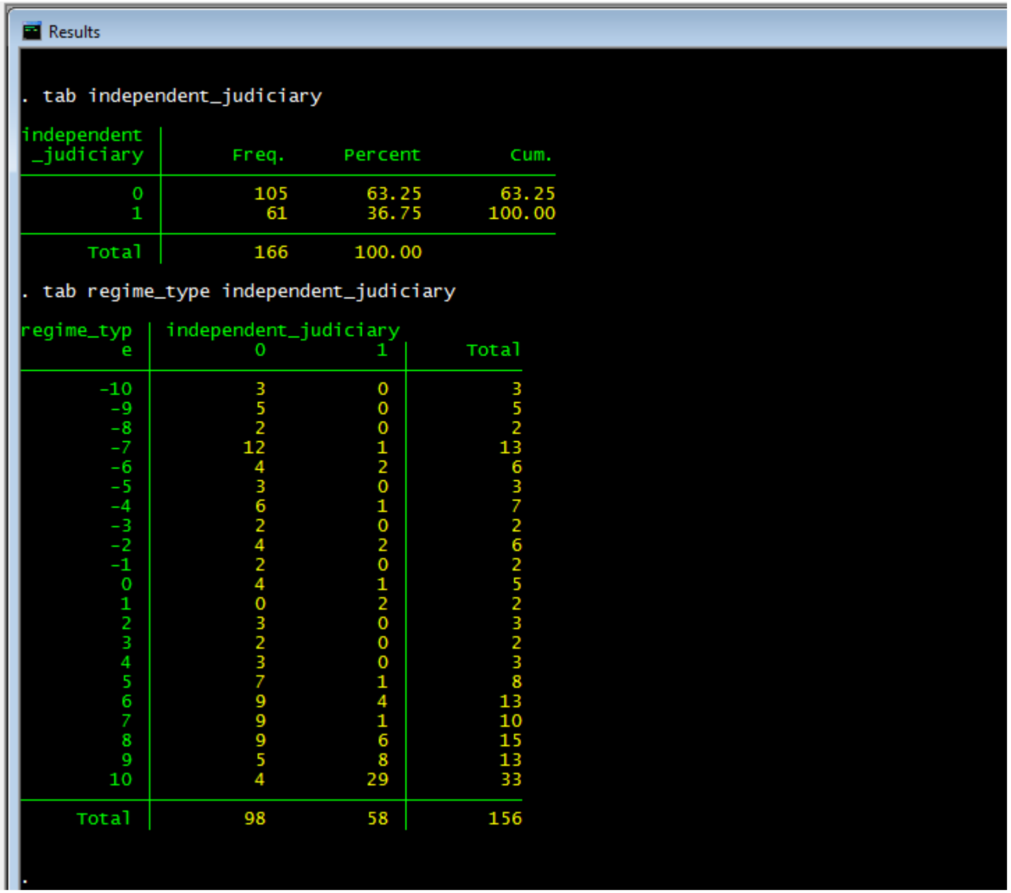
Tables and Cross-Tabs

To look at the distribution of a variable across its values, simply use the tab command e.g.

```
tab independent_judiciary
```

For a cross-tab of independent judiciary and regime type

```
tab regime_type
independent_judiciary
```



```
. tab independent_judiciary

independent |
 _judiciary |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        105       63.25       63.25
          1 |         61       36.75      100.00
------------+-----------------------------------
      Total |        166      100.00

. tab regime_type independent_judiciary

regime_typ | independent_judiciary
         e |         0          1 |     Total
-----------+----------------------+----------
       -10 |         3          0 |         3
        -9 |         5          0 |         5
        -8 |         2          0 |         2
        -7 |        12          1 |        13
        -6 |         4          2 |         6
        -5 |         3          0 |         3
        -4 |         6          1 |         7
        -3 |         2          0 |         2
        -2 |         4          2 |         6
        -1 |         2          0 |         2
         0 |         4          1 |         5
         1 |         0          2 |         2
         2 |         3          0 |         3
         3 |         2          0 |         2
         4 |         3          0 |         3
         5 |         7          1 |         8
         6 |         9          4 |        13
         7 |         9          1 |        10
         8 |         9          6 |        15
         9 |         5          8 |        13
        10 |         4         29 |        33
-----------+----------------------+----------
     Total |        98         58 |       156

.
```

Data Types

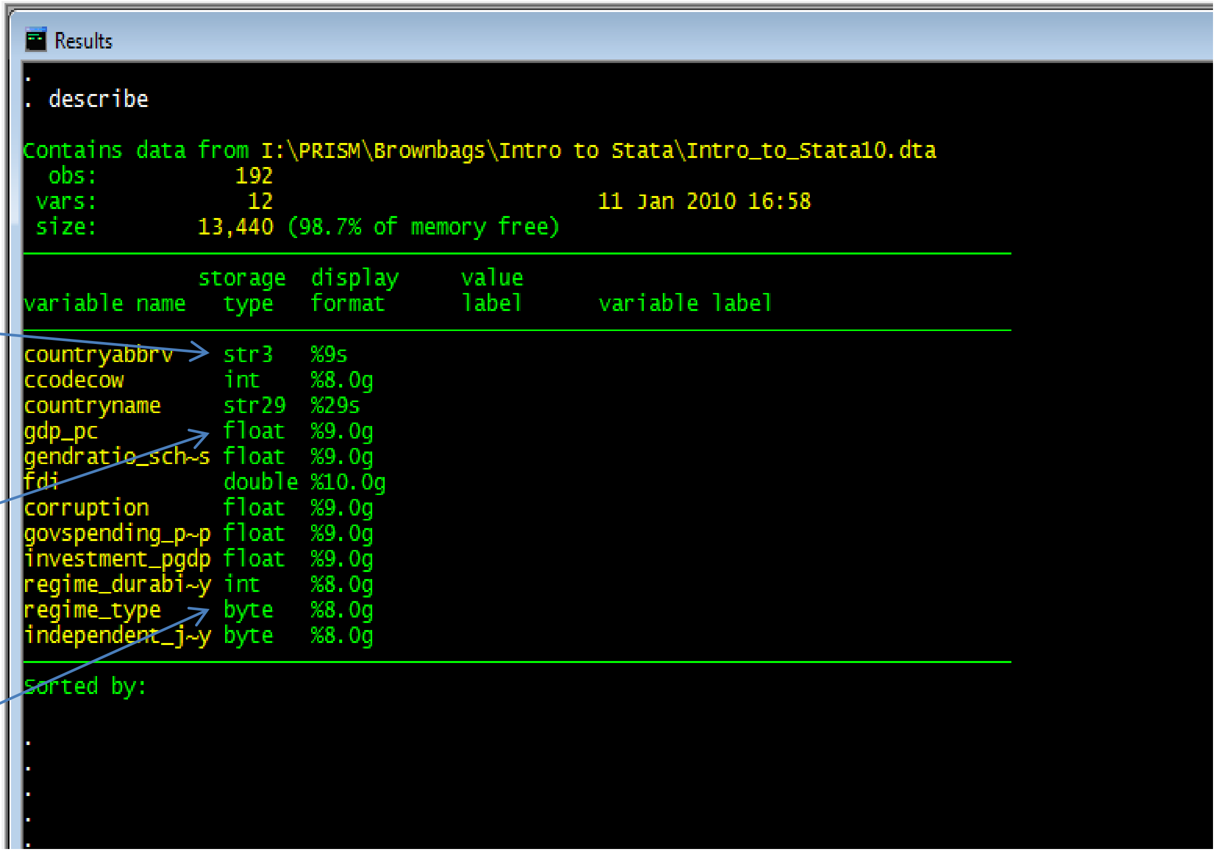
- `describe`

Data Types:

<u>Strings</u>: non-numeric variables

<u>Floats</u>: numeric data with up to 7 digits of accuracy

Byte, int, double and long are other types of numerical data.



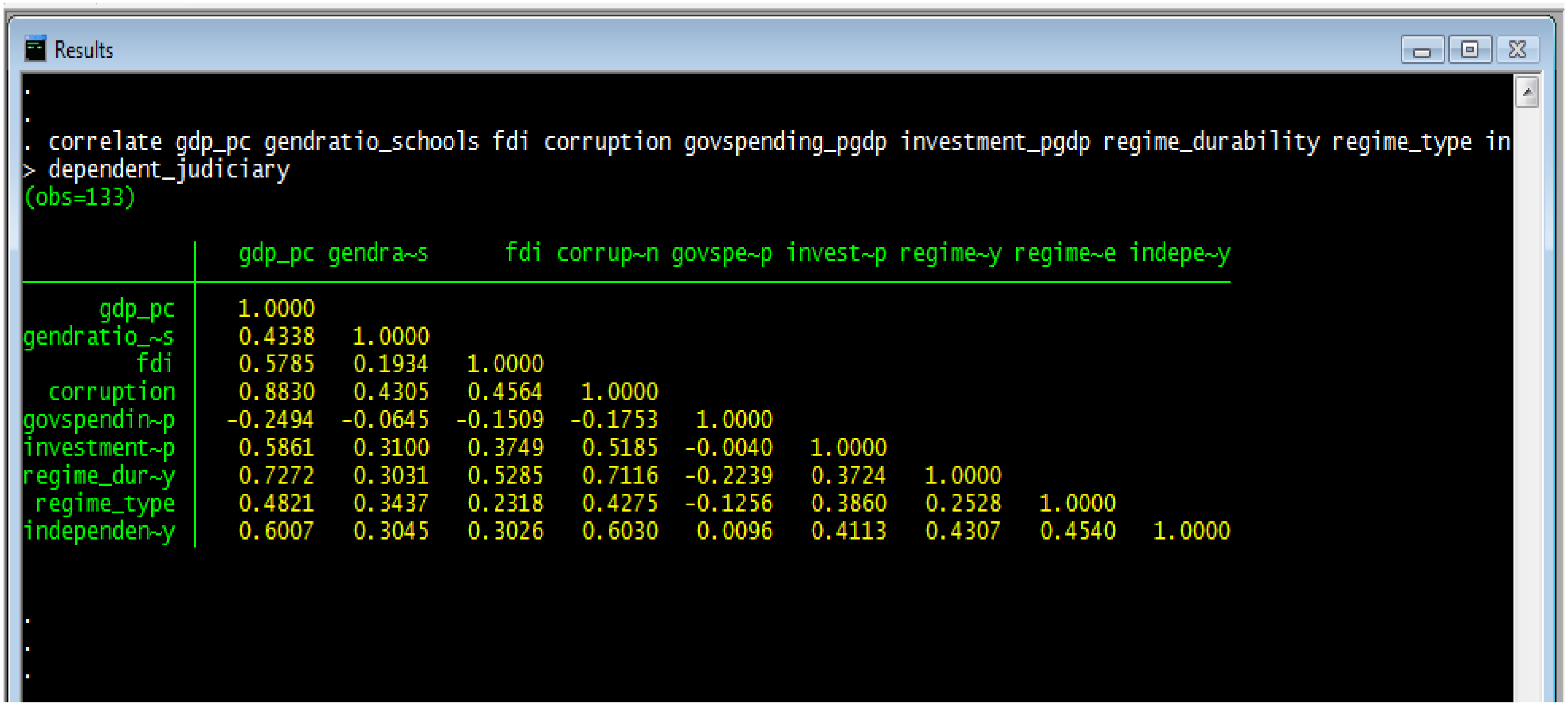
```
.
. describe

Contains data from I:\PRISM\Brownbags\Intro to Stata\Intro_to_Stata10.dta
  obs:           192
 vars:            12                          11 Jan 2010 16:58
 size:        13,440 (98.7% of memory free)
------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------
countryabbrv    str3   %9s
ccodecow        int    %8.0g
countryname     str29  %29s
gdp_pc          float  %9.0g
gendratio_sch~s float  %9.0g
fdi             double %10.0g
corruption      float  %9.0g
govspending_p~p float  %9.0g
investment_pgdp float  %9.0g
regime_durabi~y int    %8.0g
regime_type     byte   %8.0g
independent_j~y byte   %8.0g
------------------------------------------------------------------------
Sorted by:

.
.
.
.
.
```

Useful commands for changing data types: `format`, `destring`, `encode`.

Correlation Matrices

To view the correlation coefficient between one more variables type:

```
correlate variable1 variable2 …
```



```
.
.
. correlate gdp_pc gendratio_schools fdi corruption govspending_pgdp investment_pgdp regime_durability regime_type in
> dependent_judiciary
(obs=133)

             |   gdp_pc gendra~s      fdi corrup~n govspe~p invest~p regime~y regime~e indepe~y
-------------+---------------------------------------------------------------------------------
      gdp_pc |   1.0000
gendratio_~s |   0.4338   1.0000
         fdi |   0.5785   0.1934   1.0000
  corruption |   0.8830   0.4305   0.4564   1.0000
govspendin~p |  -0.2494  -0.0645  -0.1509  -0.1753   1.0000
investment~p |   0.5861   0.3100   0.3749   0.5185  -0.0040   1.0000
regime_dur~y |   0.7272   0.3031   0.5285   0.7116  -0.2239   0.3724   1.0000
 regime_type |   0.4821   0.3437   0.2318   0.4275  -0.1256   0.3860   0.2528   1.0000
independen~y |   0.6007   0.3045   0.3026   0.6030   0.0096   0.4113   0.4307   0.4540   1.0000


.
.
.
```

Generating New Variables

At times it will be useful for you to generate new variables in your data set. Let’s consider regime type, which we measure using Polity data (regime_type). This data ranges from -10 to 10, with -10 being full autocracies and 10 being full democracies. However, it might be useful to simplify this scale, and create a new dummy variable that simply reflects whether a state is a democracy or not.

To do this, let’s begin by generating a new variable called democracy:

```
generate democracy = regime_type
```

This command will generate a new variable called democracy that will be identical to the existing polity variable called regime_type.

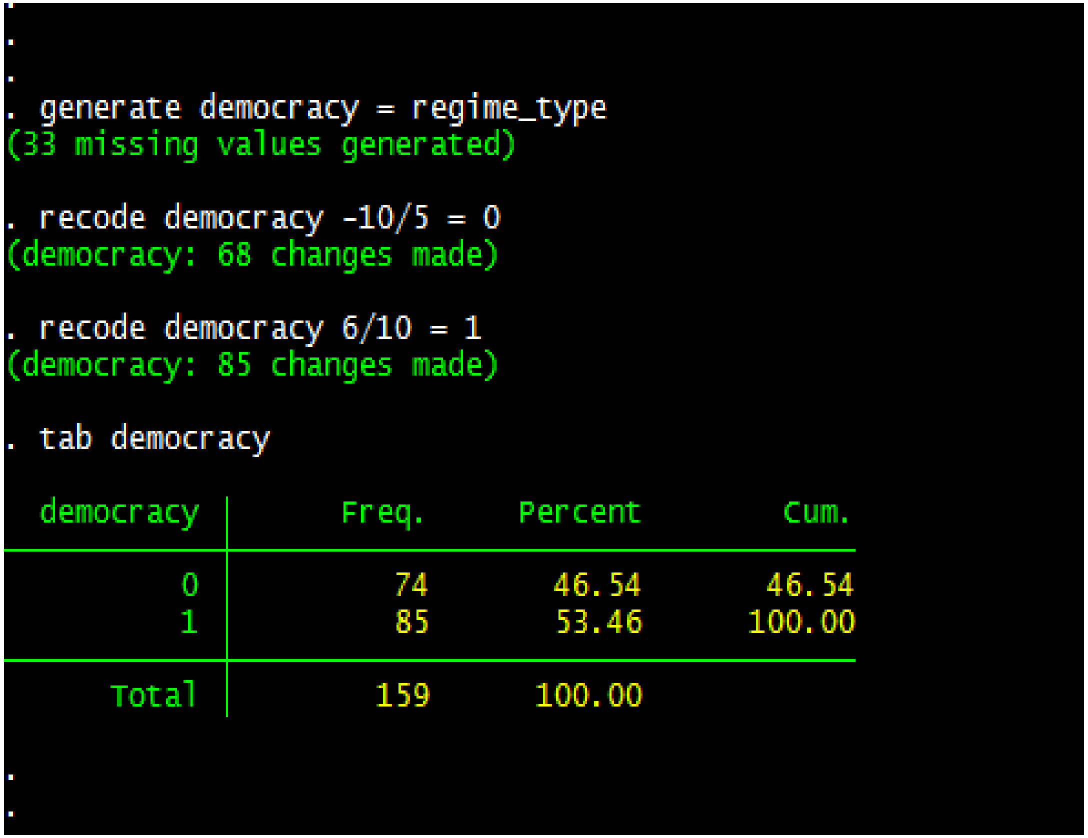
Recoding Variables

Now, we can work to recode our newly generated democracy variable by determining (arbitrarily) that states with a polity score of 6 or higher are democracies, and states with a score of 5 or lower are autocracies.

While we could recode each value individually, it is much easier to recode the entire range at once, which is easily accomplished using the /.

```
recode democracy -10/5  = 0
recode democracy 6/10  = 1
```

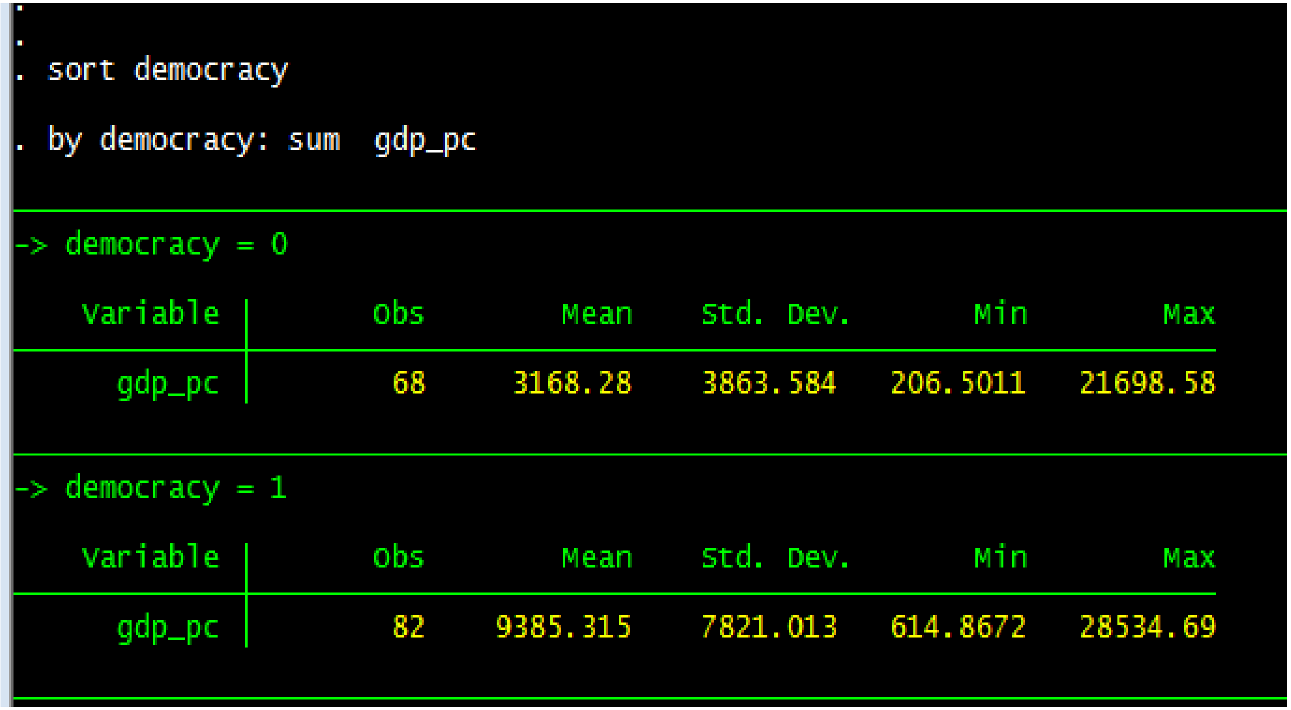
As the results window indicates, we have now successfully recoded the democracy variable so that it is now dichotomous.



```
.
.
.
. generate democracy = regime_type
(33 missing values generated)

. recode democracy -10/5 = 0
(democracy: 68 changes made)

. recode democracy 6/10 = 1
(democracy: 85 changes made)

. tab democracy

  democracy |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         74       46.54       46.54
          1 |         85       53.46      100.00
------------+-----------------------------------
      Total |        159      100.00

.
.
```

“Sort” and “By”

- `Sort variable1 variable2…` allows you to sort your data according to one or more variables in ascending order (place a “-” before a variable if you wish to sort in descending order).



```
.
.
. sort democracy

. by democracy: sum  gdp_pc

--------------------------------------------------------------------------
-> democracy = 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+------------------------------------------------------------
      gdp_pc |        68     3168.28    3863.584   206.5011   21698.58

--------------------------------------------------------------------------
-> democracy = 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+------------------------------------------------------------
      gdp_pc |        82    9385.315    7821.013   614.8672   28534.69

--------------------------------------------------------------------------
```

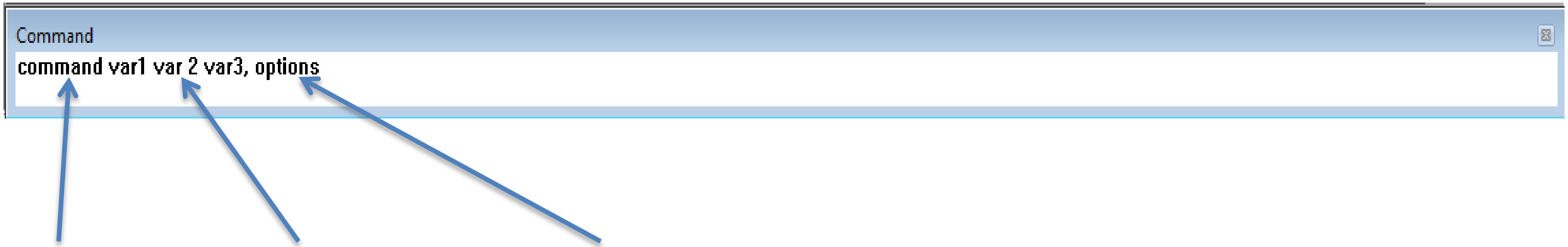
- `Sort` and `by` together allow you to sort data according to a categorical variable, and then run various commands for each category separately.

- Using our newly created democracy variable, let’s investigate the average GDP level for democracies and non democracies.

```
sort democracy

by democracy: sum gpd_pc
```

Command Structure: The Basics

The command structure in Stata is usually straightforward, although like any language there are some exceptions to the basic rules. The vast majority of commands follow this basic structure:



```
command var1 var 2 var3, options
```

The first word is usually just the name of the given command you want to perform, for example `regress` in order to estimate a regression.

Following the command name is a list of variables. When estimating models, the dependent variable is listed first.

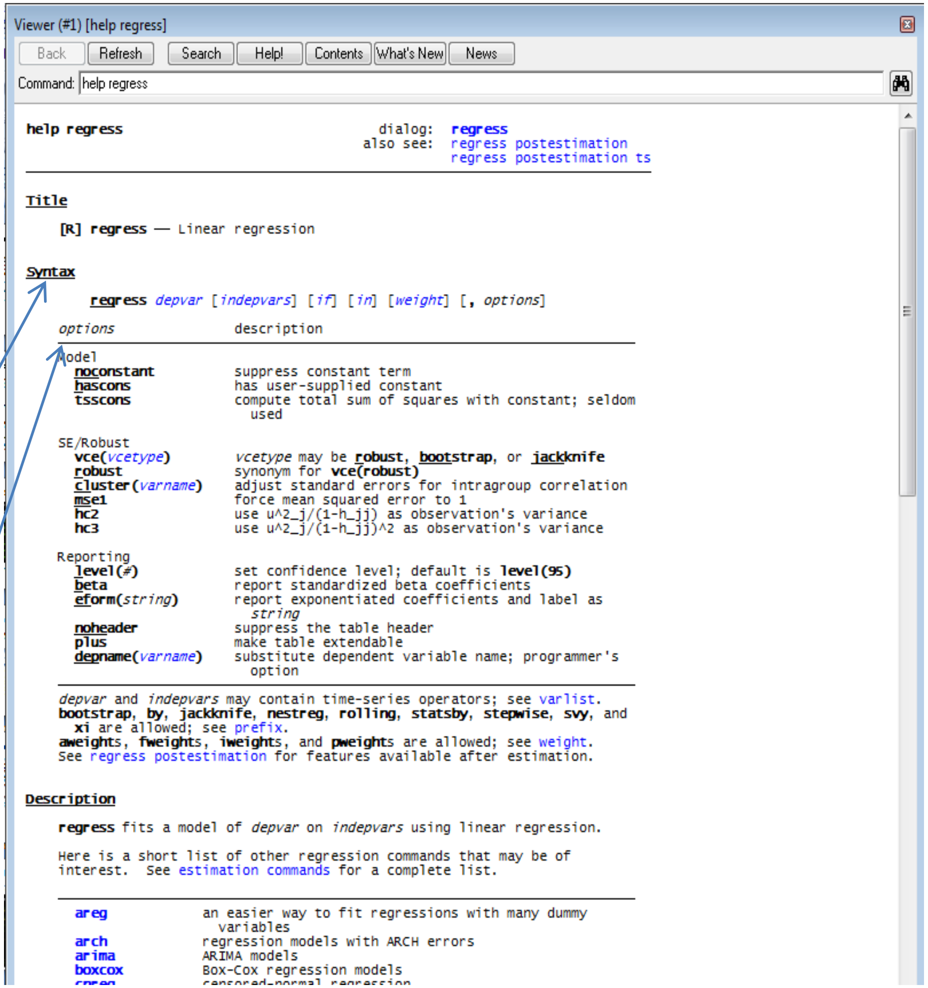
Following the variable list, you may enter a comma, which tells Stata that you now want to employ one or more available options for the given command. For example, you may want to estimate your regression model with robust standard errors. This can easily be accomplished by typing <`, robust`> after the last variable.

Command Structure: Options

Every command in Stata has numerous options that can be applied to it to provide very useful additional functions. The easiest way to determine what these options are is through the help menu. To get help for a specific command, simply type:

```
help command
```

The help file provides the basic syntax for the given command, as well as information on all of the possible options that can be added.



```
help regress

help regress                                dialog:  regress
                                          also see:  regress postestimation
                                                     regress postestimation ts

Title

    [R] regress — Linear regression

Syntax

        regress depvar [indepvars] [if] [in] [weight] [, options]

    options                description
    ---------------------------------------------------------------
    Model
      noconstant           suppress constant term
      hascons              has user-supplied constant
      tsscons              compute total sum of squares with constant; seldom
                             used

    SE/Robust
      vce(vcetype)         vcetype may be robust, bootstrap, or jackknife
      robust               synonym for vce(robust)
      cluster(varname)     adjust standard errors for intragroup correlation
      mse1                 force mean squared error to 1
      hc2                  use u^2_j/(1-h_jj) as observation's variance
      hc3                  use u^2_j/(1-h_jj)^2 as observation's variance

    Reporting
      level(#)             set confidence level; default is level(95)
      beta                 report standardized beta coefficients
      eform(string)        report exponentiated coefficients and label as
                             string
      noheader             suppress the table header
      plus                 make table extendable
      depname(varname)     substitute dependent variable name; programmer's
                             option
    ---------------------------------------------------------------
    depvar and indepvars may contain time-series operators; see varlist.
    bootstrap, by, jackknife, nestreg, rolling, statsby, stepwise, svy, and
      xi are allowed; see prefix.
    aweights, fweights, iweights, and pweights are allowed; see weight.
    See regress postestimation for features available after estimation.

Description

    regress fits a model of depvar on indepvars using linear regression.

    Here is a short list of other regression commands that may be of
    interest.  See estimation commands for a complete list.

    ---------------------------------------------------------------
      areg               an easier way to fit regressions with many dummy
                           variables
      arch               regression models with ARCH errors
      arima              ARIMA models
      boxcox             Box-Cox regression models
      cnreg              censored-normal regression
```

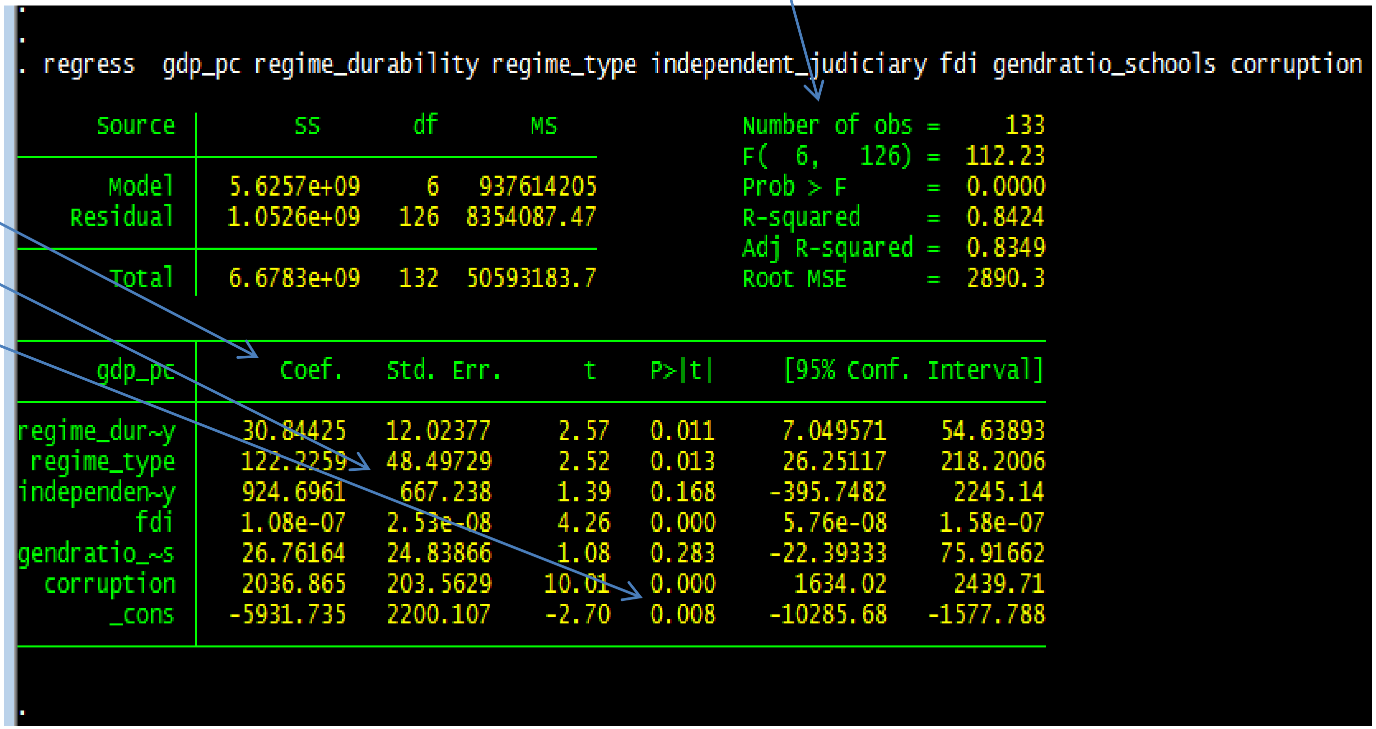
Model Estimation: OLS

Let’s say that we want to model GDP per capita using Ordinary Least Squares regression, and our theory tells us that GDP varies as a function of regime durability, regime type, and several other covariates. We type:

```
regress gdp_pc independent variables
```

Above the main results, Stata also provides a range of goodness of fit measures including R^2.

By default, Stata provides coefficient estimates, standard errors, and p values in the main table.

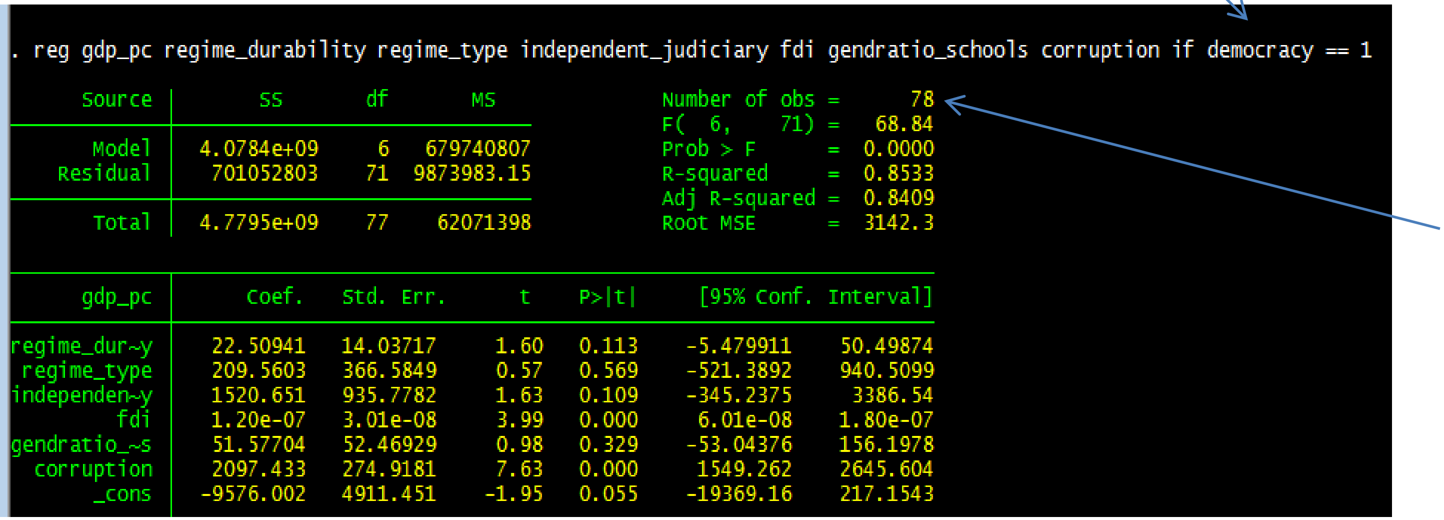


```
.
.
. regress  gdp_pc regime_durability regime_type independent_judiciary fdi gendratio_schools corruption

      Source |       SS       df       MS              Number of obs =     133
-------------+------------------------------           F(  6,   126) =  112.23
       Model |  5.6257e+09     6   937614205           Prob > F      =  0.0000
    Residual |  1.0526e+09   126  8354087.47           R-squared     =  0.8424
-------------+------------------------------           Adj R-squared =  0.8349
       Total |  6.6783e+09   132  50593183.7           Root MSE      =  2890.3

------------------------------------------------------------------------------
      gdp_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
regime_dur~y |   30.84425   12.02377     2.57   0.011     7.049571    54.63893
 regime_type |   122.2259   48.49729     2.52   0.013     26.25117    218.2006
independen~y |   924.6961    667.238     1.39   0.168    -395.7482     2245.14
         fdi |   1.08e-07   2.53e-08     4.26   0.000     5.76e-08    1.58e-07
gendratio_~s |   26.76164   24.83866     1.08   0.283    -22.39333    75.91662
  corruption |   2036.865   203.5629    10.01   0.000      1634.02     2439.71
       _cons |  -5931.735   2200.107    -2.70   0.008    -10285.68   -1577.788
------------------------------------------------------------------------------

.
```

If

- If statements can be used to limit operations to a subset of the data fulfilling the conditions set out by the if statement.

- For example, we may wish to re-estimate the OLS model using only observations that are democracies. To do this, we put the following statement at the end of the command: `if democracy == 1`.



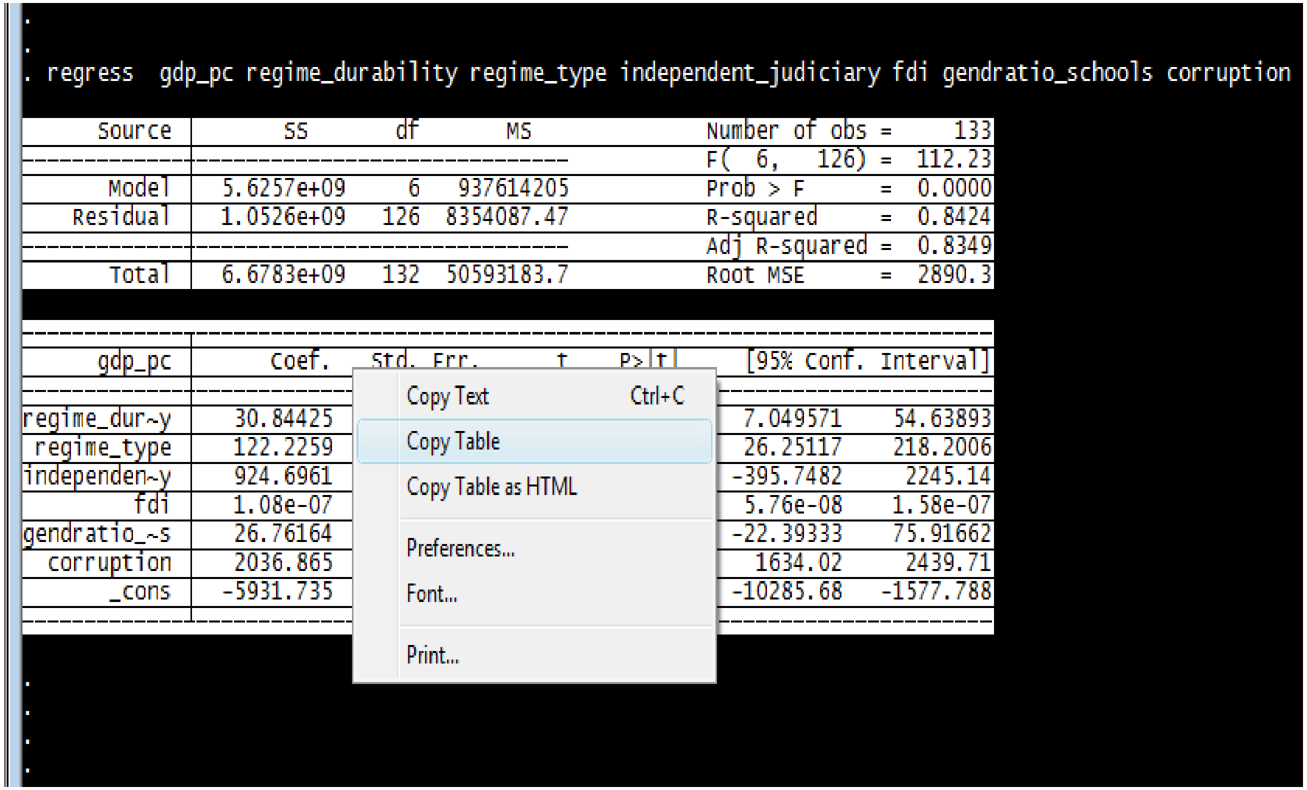
```
. reg gdp_pc regime_durability regime_type independent_judiciary fdi gendratio_schools corruption if democracy == 1

      Source |       SS       df       MS              Number of obs =      78
-------------+------------------------------           F(  6,    71) =   68.84
       Model |  4.0784e+09     6   679740807           Prob > F      =  0.0000
    Residual |   701052803    71  9873983.15           R-squared     =  0.8533
-------------+------------------------------           Adj R-squared =  0.8409
       Total |  4.7795e+09    77    62071398           Root MSE      =  3142.3

------------------------------------------------------------------------------
      gdp_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
regime_dur~y |   22.50941   14.03717     1.60   0.113    -5.479911    50.49874
 regime_type |   209.5603   366.5849     0.57   0.569    -521.3892    940.5099
independen~y |   1520.651   935.7782     1.63   0.109    -345.2375     3386.54
         fdi |   1.20e-07   3.01e-08     3.99   0.000     6.01e-08    1.80e-07
gendratio_~s |   51.57704   52.46929     0.98   0.329    -53.04376    156.1978
  corruption |   2097.433   274.9181     7.63   0.000     1549.262    2645.604
       _cons |  -9576.002   4911.451    -1.95   0.055    -19369.16    217.1543
```

Note the lower number of observations

Exporting a Table

- Exporting a table is easy. Simply highlight the output in Stata, right-click and select Copy Table. From there, you can paste your results into Word or Excel.

CAUTION: This output alone is not sufficient to turn in for your homework, let alone a paper. This is just an easy way to transfer the information. You will need to make this look more professional.



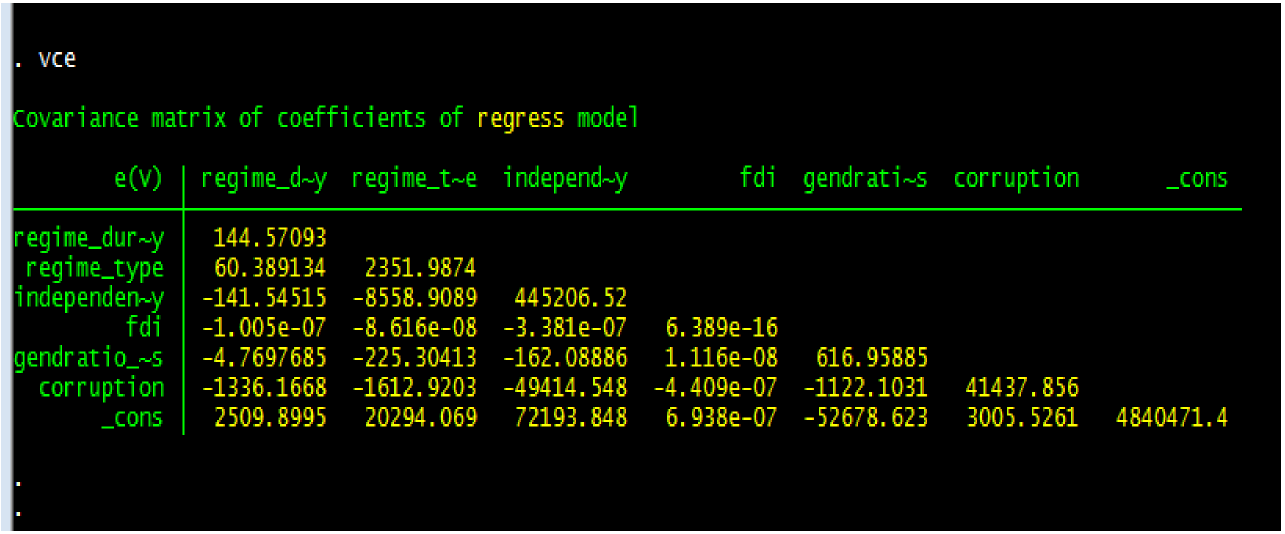
```
.
.
. regress  gdp_pc regime_durability regime_type independent_judiciary fdi gendratio_schools corruption

      Source |       SS       df       MS              Number of obs =     133
-------------+------------------------------           F(  6,   126) =  112.23
       Model |  5.6257e+09     6   937614205           Prob > F      =  0.0000
    Residual |  1.0526e+09   126  8354087.47           R-squared     =  0.8424
-------------+------------------------------           Adj R-squared =  0.8349
       Total |  6.6783e+09   132  50593183.7           Root MSE      =  2890.3

------------------------------------------------------------------------------
      gdp_pc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
regime_dur~y |   30.84425                                 7.049571    54.63893
 regime_type |   122.2259                                 26.25117    218.2006
independen~y |   924.6961                                -395.7482     2245.14
         fdi |   1.08e-07                                 5.76e-08    1.58e-07
gendratio_~s |   26.76164                                -22.39333    75.91662
  corruption |   2036.865                                  1634.02     2439.71
       _cons |  -5931.735                                -10285.68   -1577.788
------------------------------------------------------------------------------
.
.
.
.
```

Useful stata commands that can help you do this are `estout` and `outreg2`.

Post-Estimation

Stata offers a range of post-estimation options. After estimating a model, all of these commands will apply to the most recently estimated model. Among others, these commands include: `predict, adjust, level, test, vce` and others.

For example, after estimating our regression model, by typing `vce`, we can obtain the variance covariance matrix for the model.



```
. vce

Covariance matrix of coefficients of regress model

        e(V) | regime_d~y  regime_t~e  independ~y         fdi  gendrati~s  corruption       _cons
-------------+------------------------------------------------------------------------------------
regime_dur~y |  144.57093
 regime_type |  60.389134   2351.9874
independen~y | -141.54515  -8558.9089   445206.52
         fdi | -1.005e-07  -8.616e-08  -3.381e-07   6.389e-16
gendratio_~s | -4.7697685  -225.30413  -162.08886   1.116e-08   616.95885
  corruption | -1336.1668  -1612.9203  -49414.548  -4.409e-07  -1122.1031   41437.856
       _cons |  2509.8995   20294.069   72193.848   6.938e-07  -52678.623   3005.5261   4840471.4

.
.
```

Similarly, we could obtain the residuals for the model using the predict command, and the option resid:

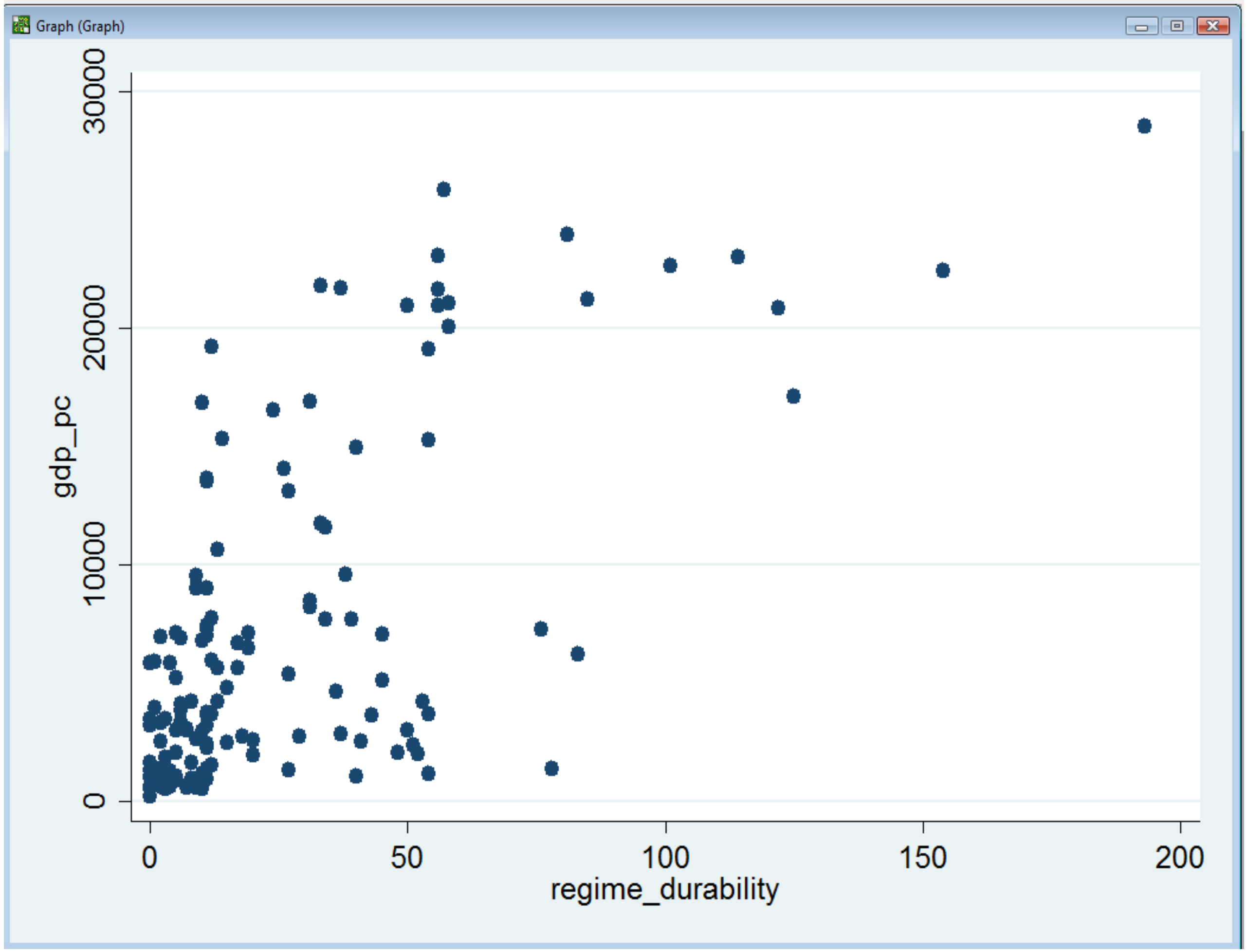
```
predict newvar, resid
```

Graphing

Stata is capable of producing a range of different graphs, and the command structure for all of these is similar.

To produce a scatter plot with GDP per capita along the y axis and regime durability along the x axis, the command is:

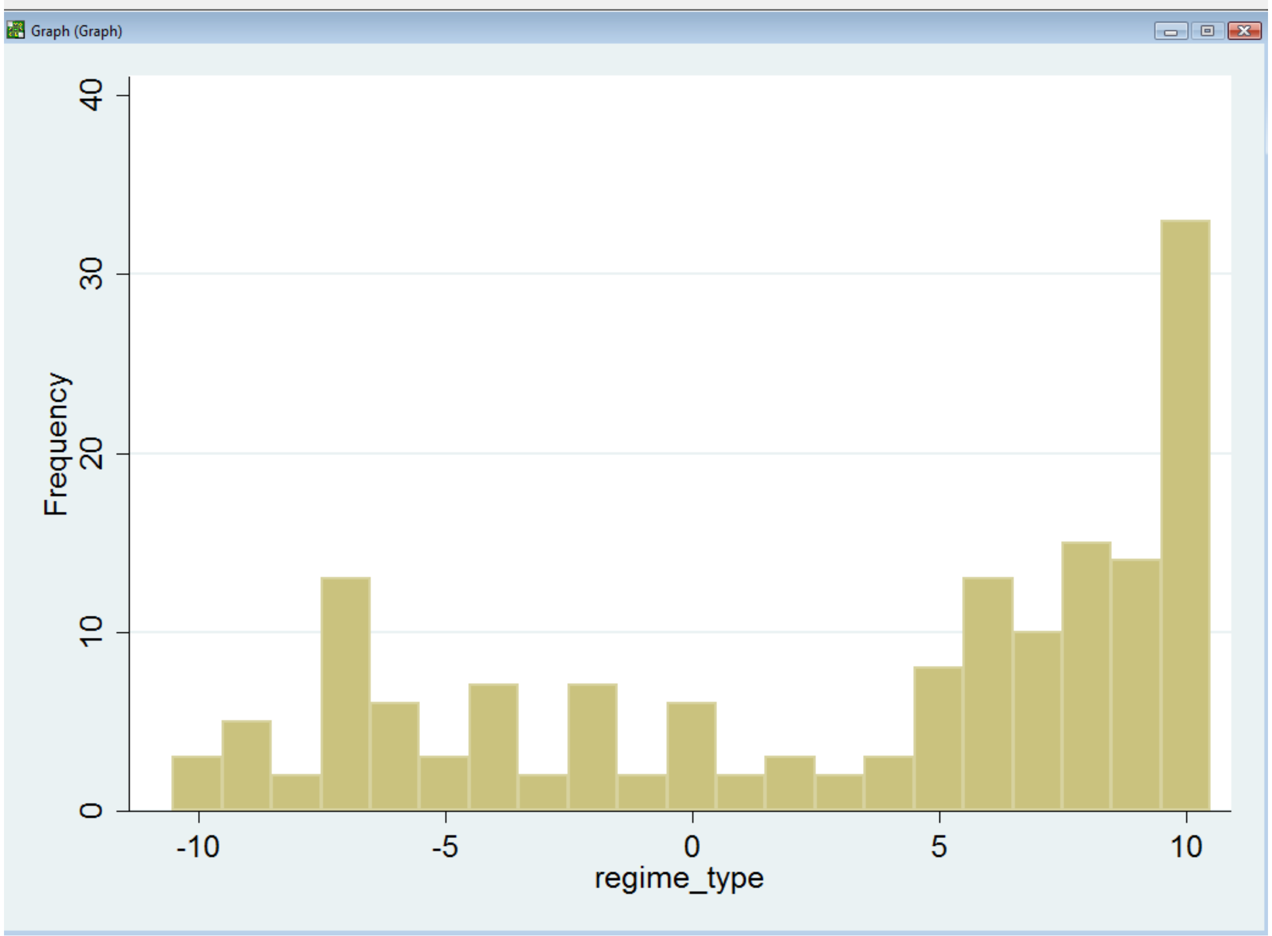
```
scatter gdp_pc
regime_durability
```



Graphing II

Another useful graph that can help you with preliminary data analysis is a histogram.

In order to produce this graph, it is necessary to use several options, which appear behind the comma. In this case, I have indicated that the variable is discrete, and that I want the frequencies to be displayed, as opposed to percentages, etc.

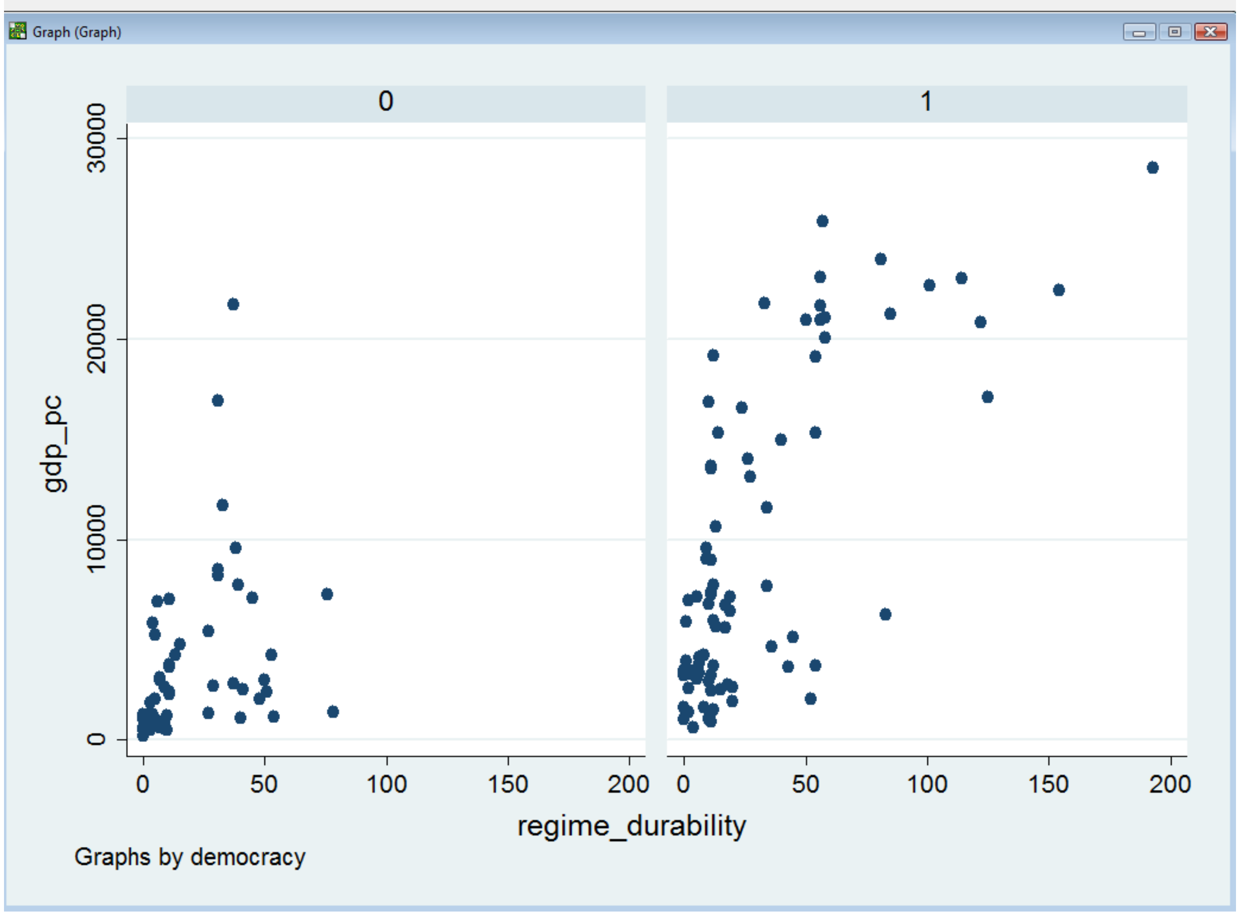


```
histogram regime_type, discrete frequency
```

Graphing III

In addition to different sorts of graphs, it is also possible to produce more sophisticated graphs in Stata. For example, we might want to see if the trend we observed in the previous scatter plot is consistent for both democracies and non-democracies. Using our democracy dummy variable, and the sort/by commands this is possible.

```
sort democracy
```



```
scatter gdp_pc regime_durability, by(democracy)
```

Exporting Graphs

- Now that you have produced a graph in Stata, you will ideally want to do something with it.

- Graphs in Stata appear in their own window. From there, you can either print the graph, or save it elsewhere on your computer. **NOTE**: producing a new graph will overwrite the previous graph, so be sure to save your graph before you move onto the next one. (PNG – Portable Network Graphics is typically a very flexible format for saving graphs.)

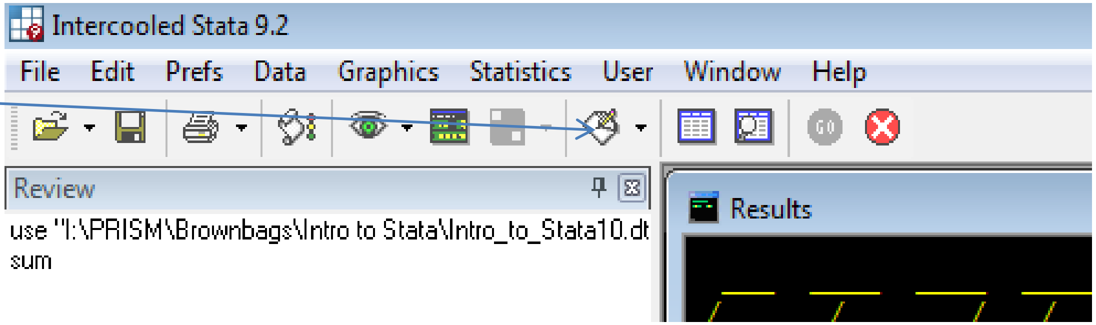
- As an alternative, you can also copy and paste graphs into documents, but it is probably a good idea to keep a saved copy just in case you accidentally crop the graph in the document.

- Finally, if you are lucky enough to be working on a mac ;), you can click and drag a graph and drop it into a document of your choosing.

The Wonder of Do files

- Do files allow you to store multiple commands for future use, manipulation or for batch processing

- Click here to open a new or saved do-file



```
use "I:\PRISM\Brownbags\Intro to Stata\Intro_to_Stata10.dt
sum
```

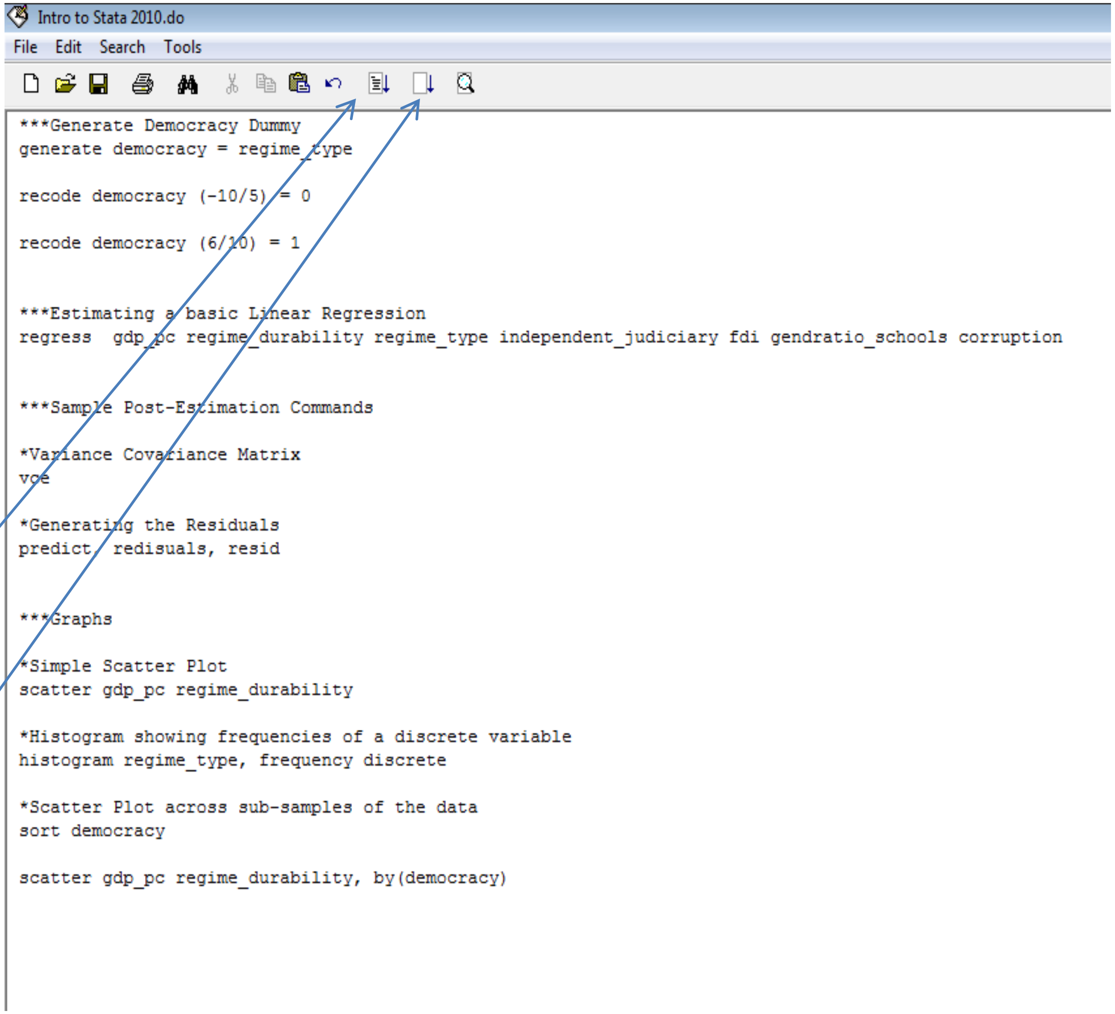
Do File Editor

- The Do file editor works just like a text editor making it easy to write, edit and save your commands.

- To annotate your Do file use an * at the beginning of each line that you do <u>not</u> want Stata to attempt to run as a command.

- To get Stata to execute commands in a Do file you can:

1) Run only the portion of commands you highlight
2) Run all commands in the open Do file



```
***Generate Democracy Dummy
generate democracy = regime_type

recode democracy (-10/5) = 0

recode democracy (6/10) = 1


***Estimating a basic Linear Regression
regress  gdp_pc regime_durability regime_type independent_judiciary fdi gendratio_schools corruption


***Sample Post-Estimation Commands

*Variance Covariance Matrix
vce

*Generating the Residuals
predict redisuals, resid


***Graphs

*Simple Scatter Plot
scatter gdp_pc regime_durability

*Histogram showing frequencies of a discrete variable
histogram regime_type, frequency discrete

*Scatter Plot across sub-samples of the data
sort democracy

scatter gdp_pc regime_durability, by(democracy)
```

Do Files

- If you forget to work in the do file, you can capture all your commands from the review window and paste them into a Do file:

1) Right click in the Review editor
2) Select “Copy Review Contents to Clipboard”
3) Open a new or old Do file where you want to record the commands
4) Paste into your Do file

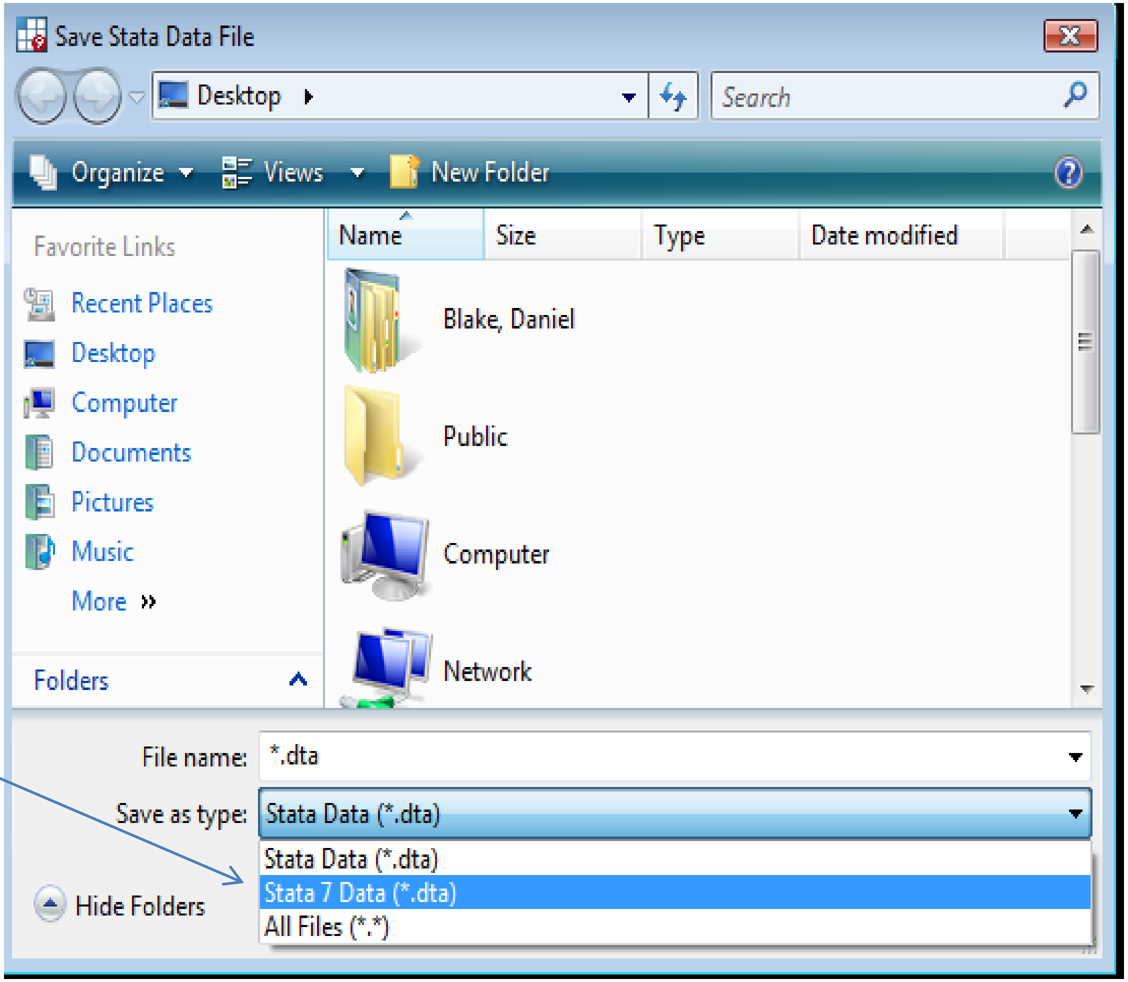
Saving Data

- Caution: Data saved in a newer version of Stata often cannot be opened in an older version of Stata.

- To save data for use in an older version:

  File > Save As >

  Select an older version of Stata in the "Save As Type" drop down menu



Save Stata Data File

File name: *.dta

Save as type: Stata Data (*.dta)

Stata Data (*.dta)
Stata 7 Data (*.dta)
All Files (*.*)

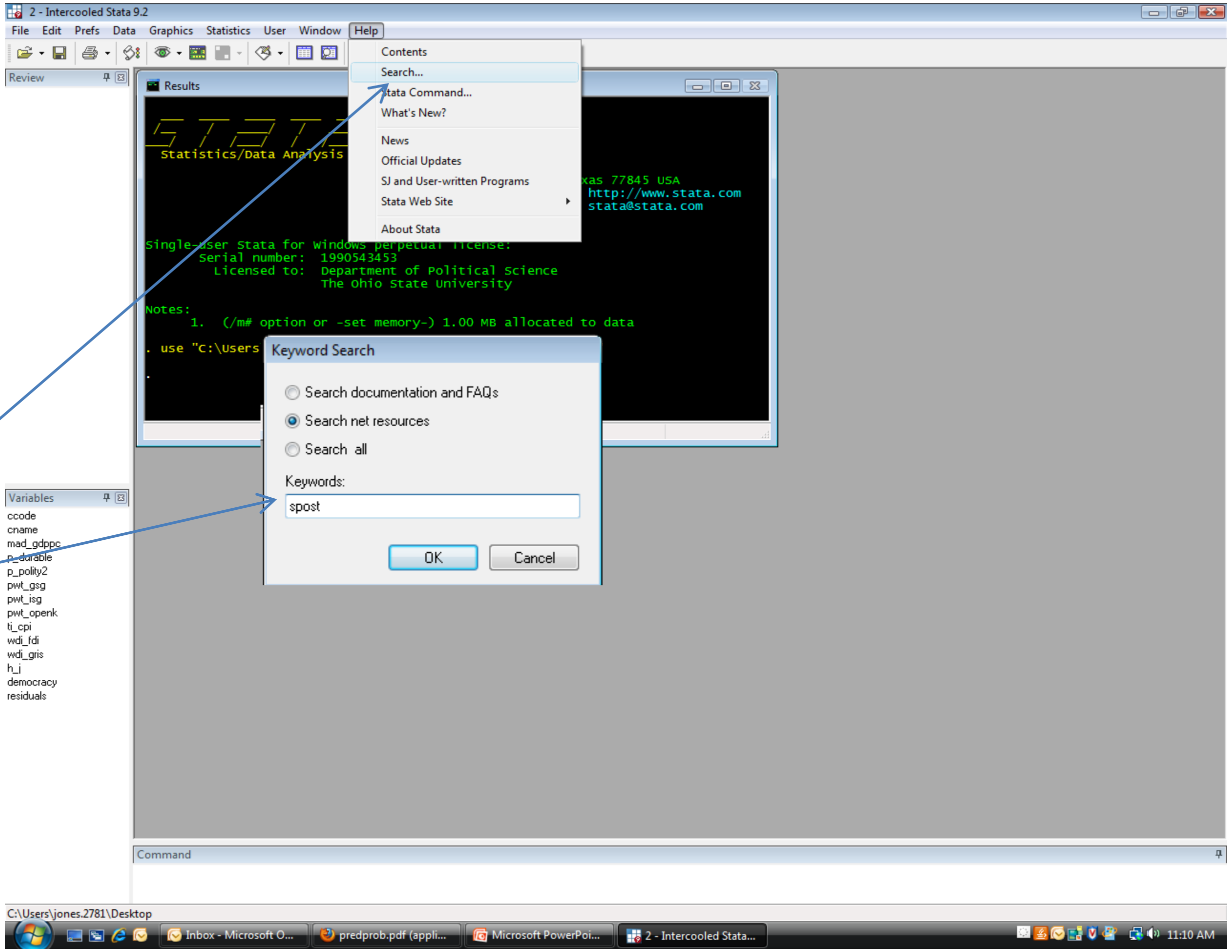
Installing packages: Searching

The first step is to do a net search for the package that you are interested in. Let's say for example that you want to install J. Scott Long's spost package.

- To find packages you can search using using the `findit` command by typing:

```
findit spost
```

- You can also find packages searching the net via the Help menu



```
Statistics/Data Analysis
                                            xas 77845 USA
                                             http://www.stata.com
                                             stata@stata.com

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

.
```

ccode
cname
mad_gdppc
p_durable
p_polity2
pwt_gsg
pwt_isg
pwt_openk
ti_cpi
wdi_fdi
wdi_gris
h_j
democracy
residuals

Keyword Search

○ Search documentation and FAQs
● Search net resources
○ Search all

Keywords: spost

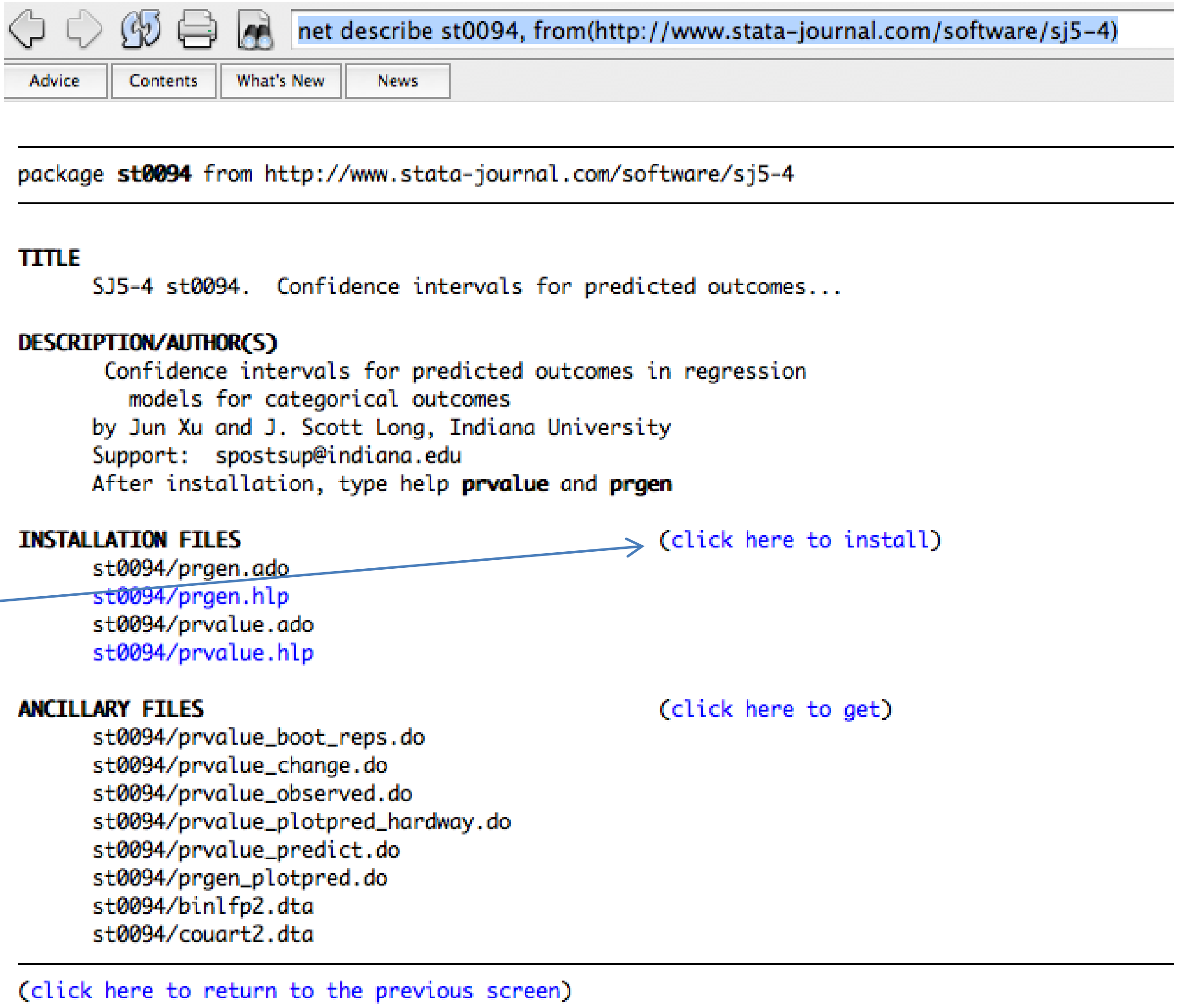
OK　Cancel

C:\Users\jones.2781\Desktop

Installing Packages: Installing

Once you have found the package that you want, you can select it from the list by clicking on the highlighted title.

From there, installing the package is as easy as clicking the install link.



```
net describe st0094, from(http://www.stata-journal.com/software/sj5-4)
```

```
package st0094 from http://www.stata-journal.com/software/sj5-4

TITLE
      SJ5-4 st0094.  Confidence intervals for predicted outcomes...

DESCRIPTION/AUTHOR(S)
       Confidence intervals for predicted outcomes in regression
         models for categorical outcomes
      by Jun Xu and J. Scott Long, Indiana University
      Support:  spostsup@indiana.edu
      After installation, type help prvalue and prgen

INSTALLATION FILES                                    (click here to install)
      st0094/prgen.ado
      st0094/prgen.hlp
      st0094/prvalue.ado
      st0094/prvalue.hlp

ANCILLARY FILES                                       (click here to get)
      st0094/prvalue_boot_reps.do
      st0094/prvalue_change.do
      st0094/prvalue_observed.do
      st0094/prvalue_plotpred_hardway.do
      st0094/prvalue_predict.do
      st0094/prgen_plotpred.do
      st0094/binlfp2.dta
      st0094/couart2.dta

(click here to return to the previous screen)
```

Useful Operators

- There are several important operators and expressions used when manipulating variables and constructing if statements.

Operator	Function	Operator	Function
var1 = var2	Makes var1 equal to var2	&	Boolean operator AND
var1 == var2	Used when comparing values (e.g. if var1 equals var 2)	| (pipe)	Boolean operator OR
var1 != var2	As above, but means does not equal.	<=/>=	Greater than or equal to/ Less than or equal to
x/y	Values in range from x to y (also x divide by y when manipulating variables)		

Upcoming PRISM Events

This upcoming Thursday, January 21st at 11:15 we will be having our first Methods Lunch of the new quarter. We will meet in the East stairwell and walk to the Wexner Center.

Our next brownbag will be held on Friday, February 26th. Any requests?

- Intro to R?
- Advanced Data Management in Stata?
- Intro LaTex/BibTex?
- PRISM fashions in the new millennium?