Stata: Simple Introduction

- Stata is well documented online (including the official documentation in pdf format)

- I use the tutorial: http://data.princeton.edu/stata/ as a starter for the one here

- Many others available

- Also see Stata Corporation youtube page: https://www.youtube.com/user/statacorp

Outline:

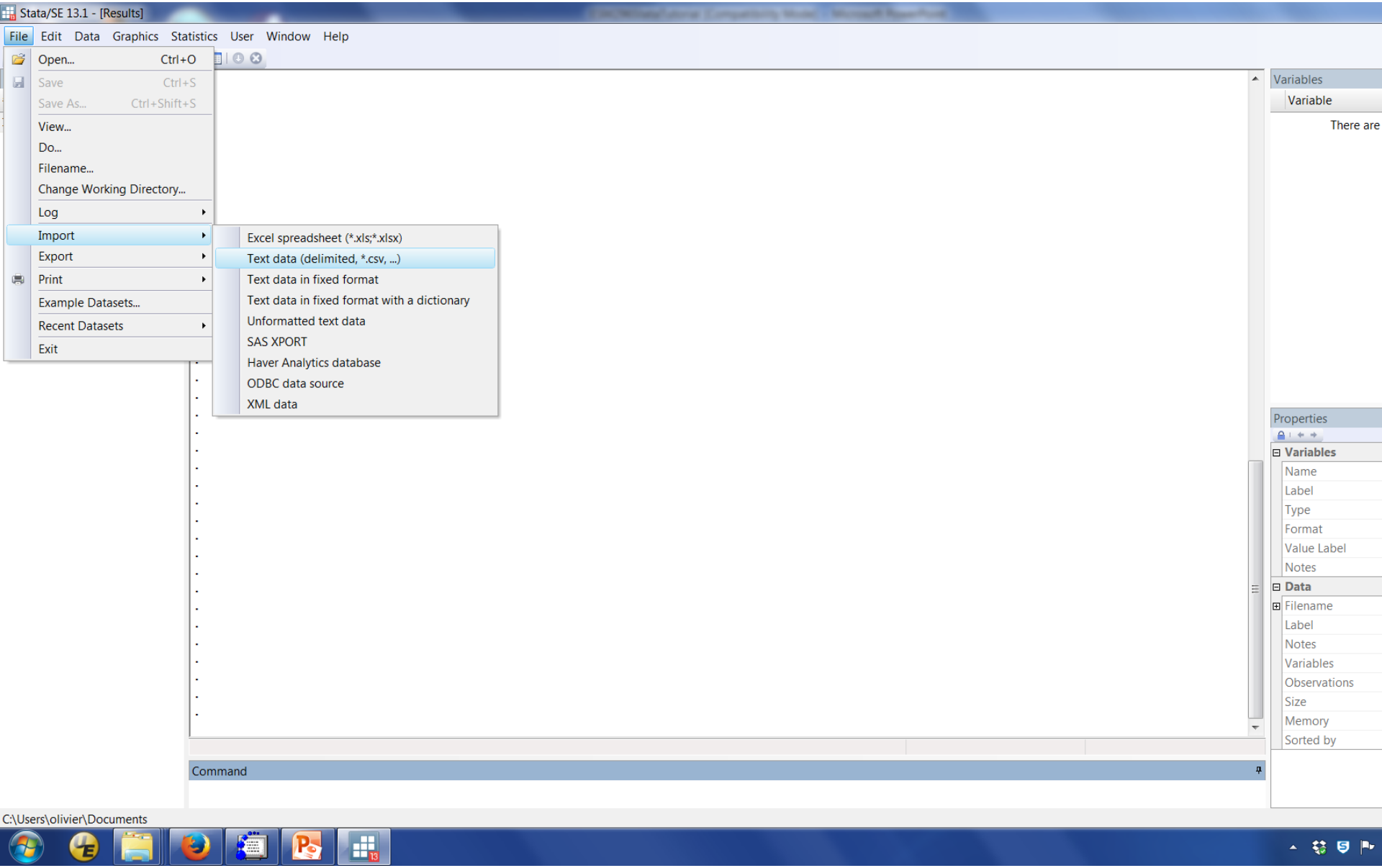
- Creating a “log” file to document your work
- Importing .csv (or .xls) files
- Examining the data for issues
- Saving dataset in Stata format
- Merging 2 Stata datasets
- Summary statistics
- Scatterplot of data
- Generating new variables
- Linear regression
- Post estimation commands (predicted values, hypothesis tests, linear combination of coefficients)
- Plotting scatterplot and fitted regression line

Creating a log file

```
. log using "C:\Teaching\ESM 296\Logfile.txt", text replace
-------------------------------------------------------------------------------
name:  <unnamed>
log:  C:\Teaching\ESM 296\Logfile.txt
log type:  text
opened on:  23 Jan 2017, 19:24:52
```

- The log file will collect all the commands and output created in a Stata session
- It is formatted in Courier New font

Importing .xls or .csv files



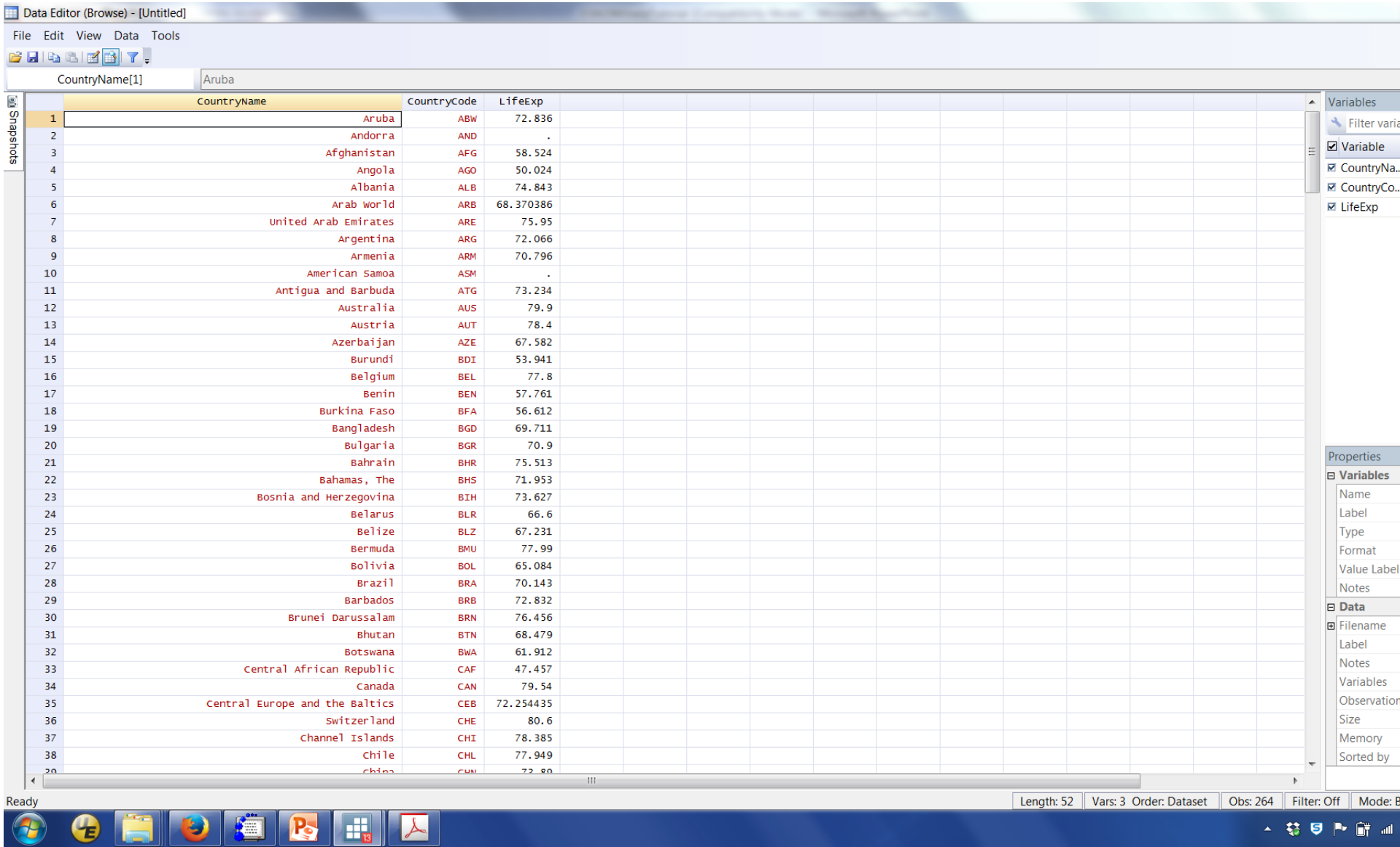
Importing a .xls file

```
. import excel "C:\Teaching\ESM 296\LifeExpectancy_2012.xlsx",
sheet("Data") firstrow
```

- I used the drop menus to locate my .xls file on my computer
- You need to specify if the first row are actual variables or the variable names (as it is often the case). "firstrow" command
- If you know the location, you can type in the full directory name instead of using

Examining the data as a spreadsheet

. **Just press the “Data Editor” menu…**



	CountryName	CountryCode	LifeExp
1	Aruba	ABW	72.836
2	Andorra	AND	.
3	Afghanistan	AFG	58.524
4	Angola	AGO	50.024
5	Albania	ALB	74.843
6	Arab World	ARB	68.370386
7	United Arab Emirates	ARE	75.95
8	Argentina	ARG	72.066
9	Armenia	ARM	70.796
10	American Samoa	ASM	.
11	Antigua and Barbuda	ATG	73.234
12	Australia	AUS	79.9
13	Austria	AUT	78.4
14	Azerbaijan	AZE	67.582
15	Burundi	BDI	53.941
16	Belgium	BEL	77.8
17	Benin	BEN	57.761
18	Burkina Faso	BFA	56.612
19	Bangladesh	BGD	69.711
20	Bulgaria	BGR	70.9
21	Bahrain	BHR	75.513
22	Bahamas, The	BHS	71.953
23	Bosnia and Herzegovina	BIH	73.627
24	Belarus	BLR	66.6
25	Belize	BLZ	67.231
26	Bermuda	BMU	77.99
27	Bolivia	BOL	65.084
28	Brazil	BRA	70.143
29	Barbados	BRB	72.832
30	Brunei Darussalam	BRN	76.456
31	Bhutan	BTN	68.479
32	Botswana	BWA	61.912
33	Central African Republic	CAF	47.457
34	Canada	CAN	79.54
35	Central Europe and the Baltics	CEB	72.254435
36	Switzerland	CHE	80.6
37	Channel Islands	CHI	78.385
38	Chile	CHL	77.949
39	China	CHN	73.89

Examining the data

```
. list
```

	CountryName	Count~de	LifeExp
1.	Aruba	ABW	72.836
2.	Andorra	AND	.
3.	Afghanistan	AFG	58.524
4.	Angola	AGO	50.024
5.	Albania	ALB	74.843
6.	Arab World	ARB	68.370386
7.	United Arab Emirates	ARE	75.95
8.	Argentina	ARG	72.066
9.	Armenia	ARM	70.796
10.	American Samoa	ASM	.
11.	Antigua and Barbuda	ATG	73.234
12.	Australia	AUS	79.9

Saving in Stata (.dta) format

```
. save "C:\Teaching\ESM 296\LifeExpectancy_2012.dta"
```

If you want to overwrite (be careful) an existing dataset, add the “,replace” to the above command

Merging 2 .dta files:

```
. use "C:\Teaching\ESM 296\LifeExpectancy_2012.dta", clear

. merge 1:1 CountryName using "C:\Teaching\ESM 296\PM25_2012.dta"

    Result                           # of obs.
    -----------------------------------------
    not matched                             0
    matched                               264  (_merge==3)
    -----------------------------------------
```

Also can be done by using the “Data” -> “Combine datasets” -> … menus

Both datasets need to be sorted by the “merging” variable, here “CountryName” -> sort Countryname…

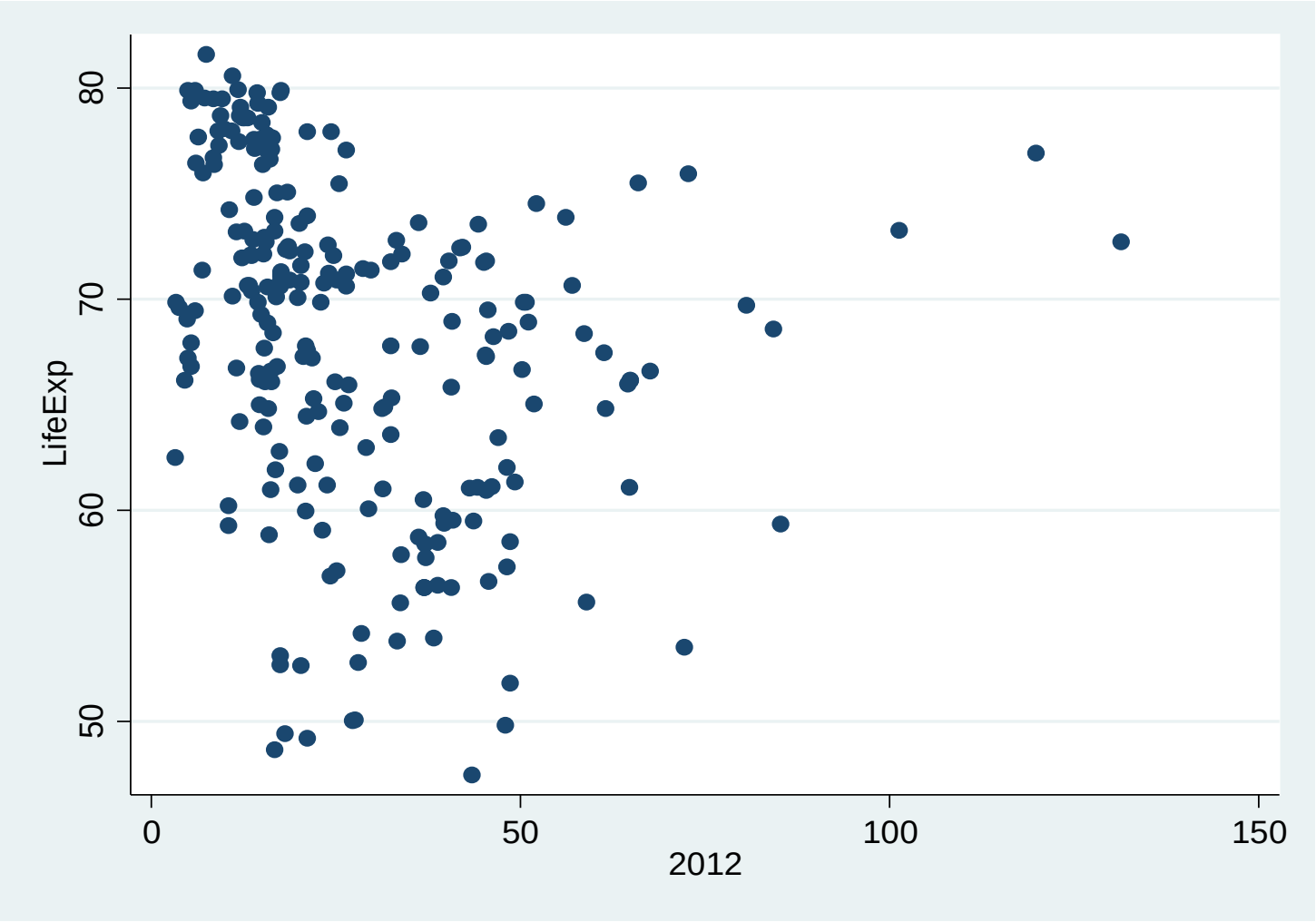
Summary statistics

```
. summ (or summarize)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
CountryName	0				
CountryCode	0				
LifeExp	246	68.39469	7.852758	47.457	81.6
PM25	240	27.54711	19.86892	3.2	131.4
_merge	264	3	0	3	3

Scatterplot of variables

```
. graph twoway scatter LifeExp PM25
```



Generating new variables

```
. gen ln_LifeExp=log(LifeExp)
(18 missing values generated)
```

Simple linear regression

```
. reg ln_LifeExp PM25, robust

Linear regression                               Number of obs =     235
                                                F(  1,   233) =    9.22
                                                Prob > F      =  0.0027
                                                R-squared     =  0.0636
                                                Root MSE      =  .11658

------------------------------------------------------------------------------
             |               Robust
  ln_LifeExp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PM25 |   -.001524   .0005018    -3.04   0.003    -.0025127   -.0005352
       _cons |   4.256063   .0147634   288.28   0.000     4.226976    4.285149
------------------------------------------------------------------------------
```

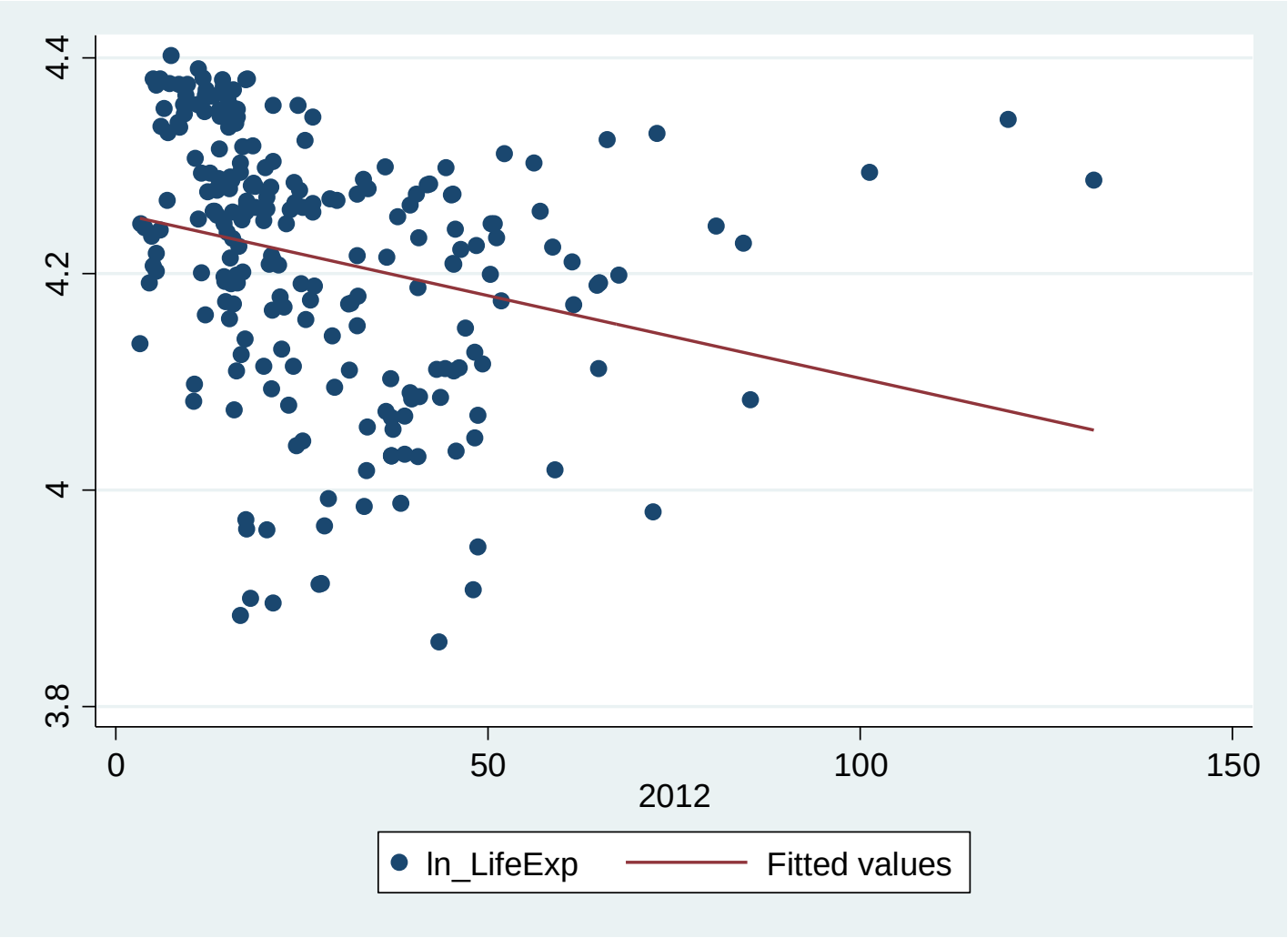
Post estimation commands (1)

```
. predict ln_LE_fit, xb
(24 missing values generated)
```

Creates new Stata variable (ln_LE_fit) with fitted values from regression

Scatterplot of data and fitted regression line

```
. graph twoway (scatter ln_LifeExp PM25) (lfit ln_LifeExp PM25)
```



Post estimation commands (2)

```
. test PM25=-0.003

 ( 1)  PM25 = -.003

       F(  1,   233) =    8.65
            Prob > F =    0.0036
```

Hypothesis tests …

Post estimation commands (3)

```
. lincom 5*PM25

 ( 1)  5*PM25 = 0

------------------------------------------------------------------------------
  ln_LifeExp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.0076198   .0025092    -3.04   0.003    -.0125635   -.0026761
------------------------------------------------------------------------------
```

Linear combination of regression coefficient(s) and associated standard errors

Closing the log file

```
. log close
name:  <unnamed>
log:  C:\Teaching\ESM 296\Logfile.txt
log type:  text
closed on:  23 Jan 2017, 20:14:15
```