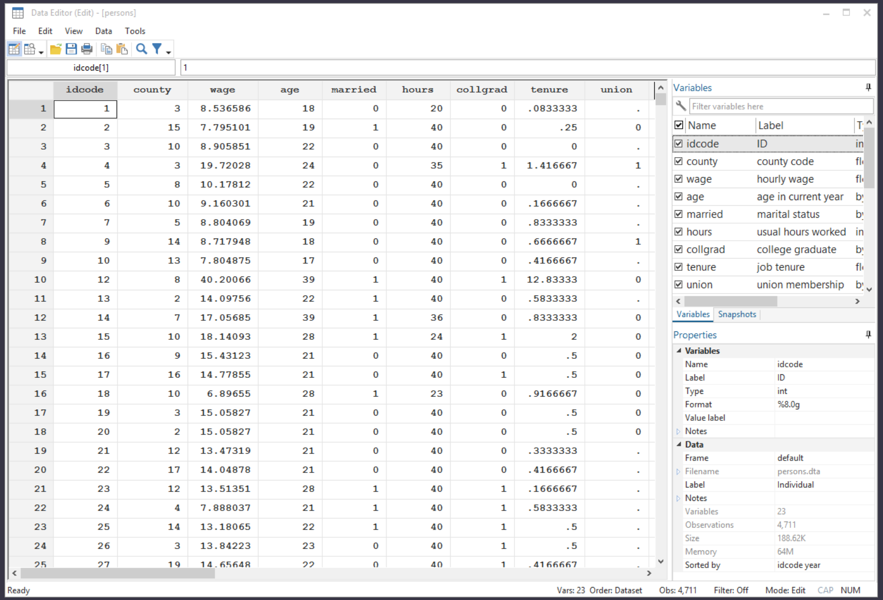
Stata Level 1: Fundamentals of Data Analysis

JUAN SEBASTIAN CUERVO

Goal 1 - Why use Stata?

JUAN SEBASTIAN CUERVO

Goal 1: Why use Stata? - Level 1



idcode[1] 1

	idcode	county	wage	age	married	hours	collgrad	tenure	union
1	1	3	8.536586	18	0	20	0	.0833333	.
2	2	15	7.795101	19	1	40	0	.25	0
3	3	10	8.905851	22	0	40	0	0	.
4	4	3	19.72028	24	0	35	1	1.416667	1
5	5	8	10.17812	22	0	40	0	0	.
6	6	10	9.160301	21	0	40	0	.1666667	.
7	7	5	8.804069	19	0	40	0	.8333333	.
8	9	14	8.717948	18	0	40	0	.6666667	1
9	10	13	7.804875	17	0	40	0	.4166667	.
10	12	8	40.20066	39	1	40	1	12.83333	0
11	13	2	14.09756	22	1	40	0	.5833333	.
12	14	7	17.05685	39	1	36	0	.8333333	0
13	15	10	18.14093	28	1	24	1	2	0
14	16	9	15.43123	21	0	40	0	.5	0
15	17	16	14.77855	21	0	40	1	.5	0
16	18	10	6.89655	28	1	23	0	.9166667	0
17	19	3	15.05827	21	0	40	0	.5	0
18	20	2	15.05827	21	0	40	0	.5	0
19	21	12	13.47319	21	0	40	0	.3333333	.
20	22	17	14.04878	21	0	40	0	.4166667	.
21	23	12	13.51351	28	1	40	1	.1666667	.
22	24	4	7.888037	21	1	40	1	.5833333	.
23	25	14	13.18065	22	1	40	1	.5	.
24	26	3	13.84223	23	0	40	1	.5	.
25	27	19	14.65648	22	0	40	1	.4166667	.

Variables

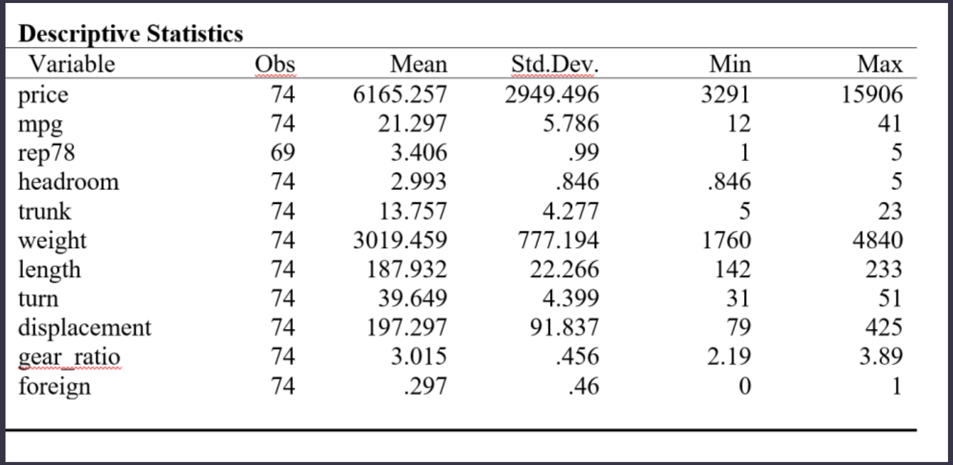
Filter variables here

☑ Name	Label	T
☑ idcode	ID	in
☑ county	county code	fl
☑ wage	hourly wage	fl
☑ age	age in current year	by
☑ married	marital status	by
☑ hours	usual hours worked	in
☑ collgrad	college graduate	by
☑ tenure	job tenure	fl
☑ union	union membership	by

Properties

Variables	
Name	idcode
Label	ID
Type	int
Format	%8.0g
Value label	
Notes	
Data	
Frame	default
Filename	persons.dta
Label	Individual
Notes	
Variables	23
Observations	4,711
Size	188.62K
Memory	64M
Sorted by	idcode year

Ready Vars: 23 Order: Dataset Obs: 4,711 Filter: Off Mode: Edit CAP NUM

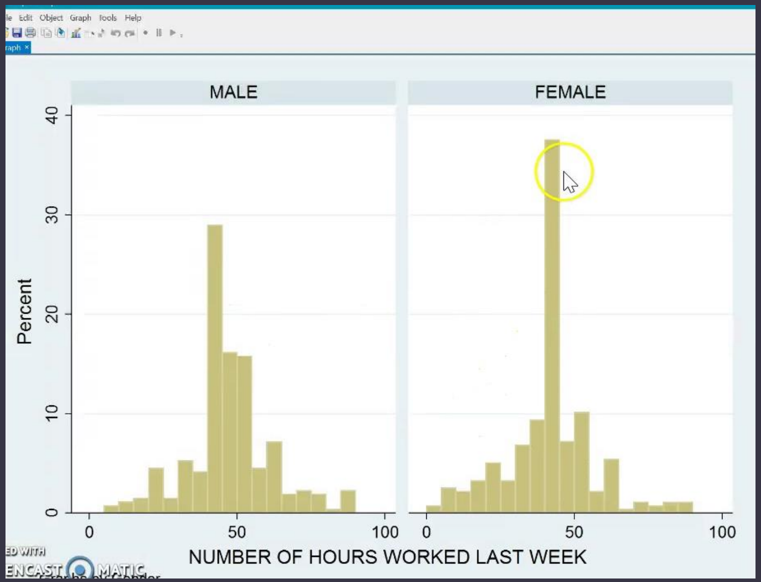


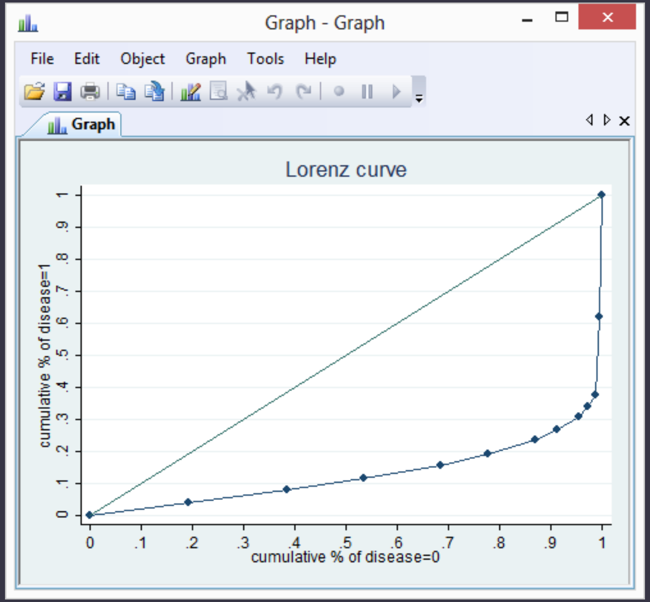
Descriptive Statistics

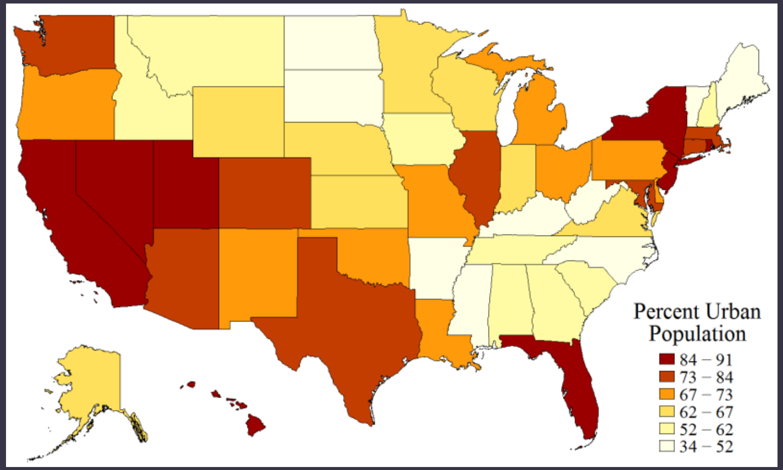
Variable	Obs	Mean	Std.Dev.	Min	Max
price	74	6165.257	2949.496	3291	15906
mpg	74	21.297	5.786	12	41
rep78	69	3.406	.99	1	5
headroom	74	2.993	.846	.846	5
trunk	74	13.757	4.277	5	23
weight	74	3019.459	777.194	1760	4840
length	74	187.932	22.266	142	233
turn	74	39.649	4.399	31	51
displacement	74	197.297	91.837	79	425
gear_ratio	74	3.015	.456	2.19	3.89
foreign	74	.297	.46	0	1

Organize large quantities of information

Goal 1: Why use Stata? -Level 2

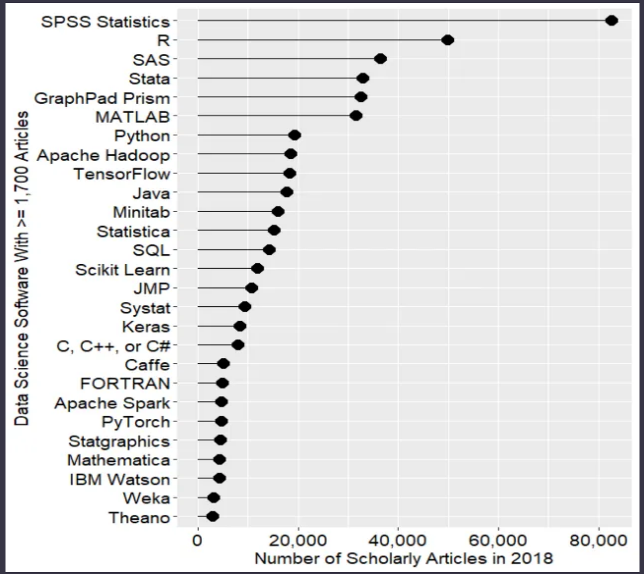

le Edit Object Graph Tools Help
aph
MALE
FEMALE
Percent
0
10
20
30
40
0
50
100
0
50
100
NUMBER OF HOURS WORKED LAST WEEK



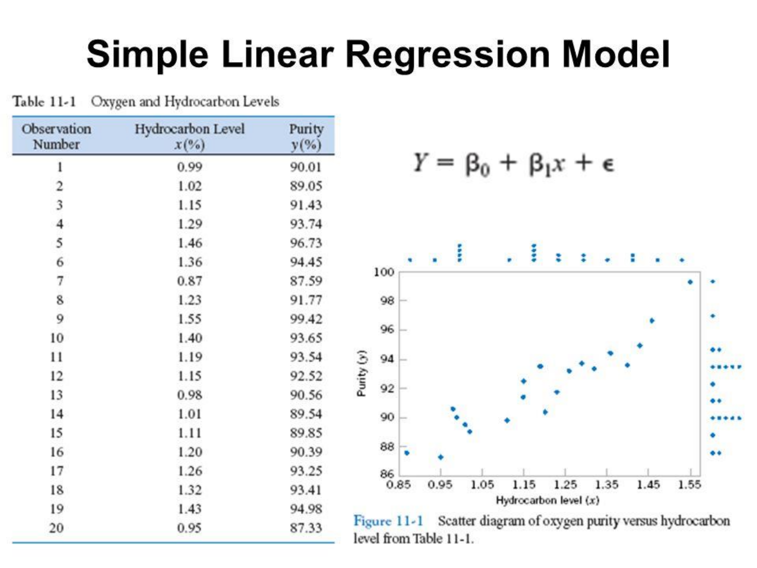


Visualize data more easily

Goal 1: Why use Stata? - Level 3



Stata in Data Science



Simple Linear Regression Model

Table 11-1 Oxygen and Hydrocarbon Levels

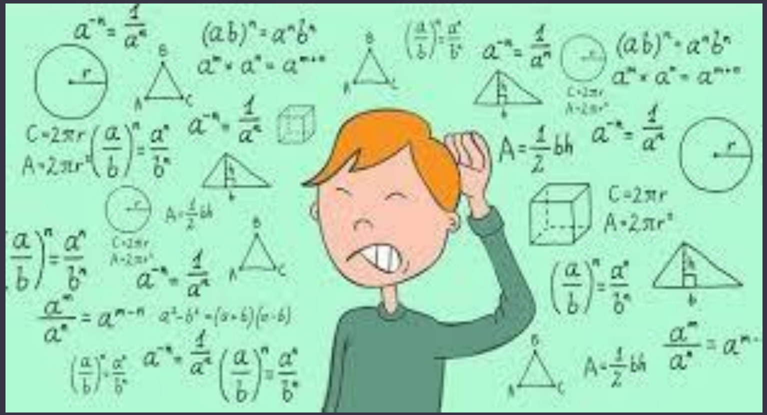
Observation Number	Hydrocarbon Level $x(\%)$	Purity $y(\%)$
1	0.99	90.01
2	1.02	89.05
3	1.15	91.43
4	1.29	93.74
5	1.46	96.73
6	1.36	94.45
7	0.87	87.59
8	1.23	91.77
9	1.55	99.42
10	1.40	93.65
11	1.19	93.54
12	1.15	92.52
13	0.98	90.56
14	1.01	89.54
15	1.11	89.85
16	1.20	90.39
17	1.26	93.25
18	1.32	93.41
19	1.43	94.98
20	0.95	87.33

$$Y = \beta_0 + \beta_1 x + \epsilon$$

100
98
96
94
92
90
88
86
Purity (y)
0.85
0.95
1.05
1.15
1.25
1.35
1.45
1.55
Hydrocarbon level (x)

Figure 11-1 Scatter diagram of oxygen purity versus hydrocarbon level from Table 11-1.

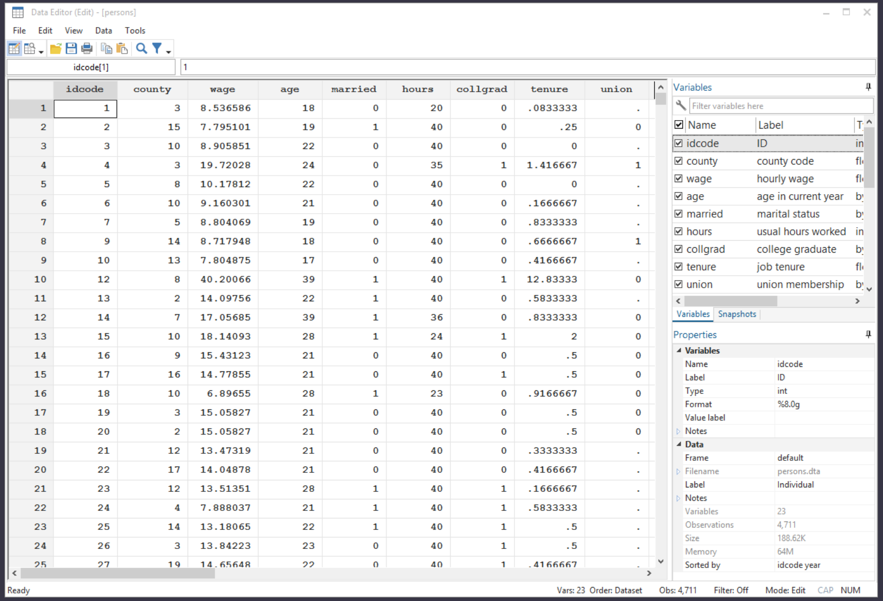
Statistical inference



Less fear to quantitative

Generate insights in a data-oriented world - without overwhelm

Scope of course - Level 1

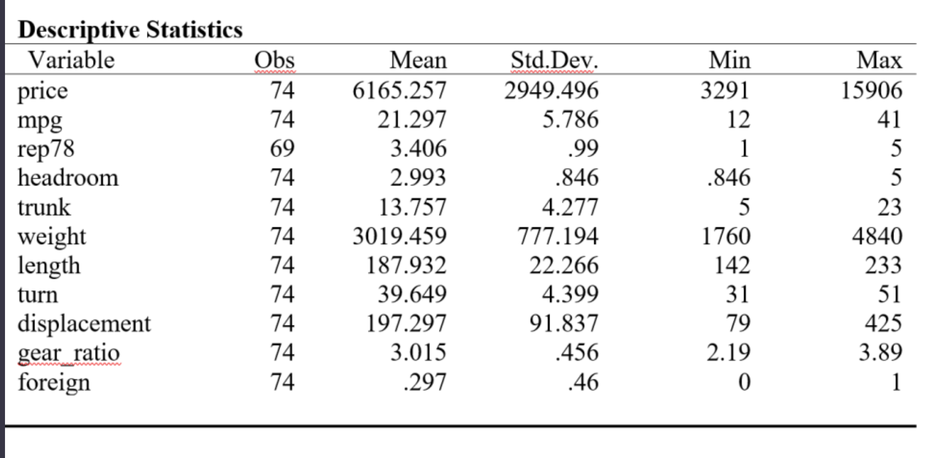


	idcode	county	wage	age	married	hours	collgrad	tenure	union
1	1	3	8.536586	18	0	20	0	.0833333	.
2	2	15	7.795101	19	1	40	0	.25	0
3	3	10	8.905851	22	0	40	0	0	.
4	4	3	19.72028	24	0	35	1	1.416667	1
5	5	8	10.17812	22	0	40	0	0	.
6	6	10	9.160301	21	0	40	0	.1666667	.
7	7	5	8.804069	19	0	40	0	.8333333	.
8	9	14	8.717948	18	0	40	0	.6666667	1
9	10	13	7.804875	17	0	40	0	.4166667	.
10	12	8	40.20066	39	1	40	1	12.83333	0
11	13	2	14.09756	22	1	40	0	.5833333	.
12	14	7	17.05685	39	1	36	0	.8333333	0
13	15	10	18.14093	28	1	24	1	2	0
14	16	9	15.43123	21	0	40	0	.5	0
15	17	16	14.77855	21	0	40	1	.5	0
16	18	10	6.89655	28	1	23	0	.9166667	0
17	19	3	15.05827	21	0	40	0	.5	0
18	20	2	15.05827	21	0	40	0	.5	0
19	21	12	13.47319	21	0	40	0	.3333333	.
20	22	17	14.04878	21	0	40	0	.4166667	.
21	23	12	13.51351	28	1	40	1	.1666667	.
22	24	4	7.888037	21	1	40	1	.5833333	.
23	25	14	13.18065	22	1	40	1	.5	.
24	26	3	13.84223	23	0	40	1	.5	.
25	27	19	14.65648	22	0	40	1	.4166667	.

☑	Name	Label	T
☑	idcode	ID	in
☑	county	county code	fl
☑	wage	hourly wage	fl
☑	age	age in current year	by
☑	married	marital status	by
☑	hours	usual hours worked	in
☑	collgrad	college graduate	by
☑	tenure	job tenure	fl
☑	union	union membership	by

Variables	
Name	idcode
Label	ID
Type	int
Format	%8.0g
Value label	
Notes	
Data	
Frame	default
Filename	persons.dta
Label	Individual
Notes	
Variables	23
Observations	4,711
Size	188.62K
Memory	64M
Sorted by	idcode year

Ready   Vars: 23 Order: Dataset   Obs: 4,711   Filter: Off   Mode: Edit   CAP   NUM



Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
price	74	6165.257	2949.496	3291	15906
mpg	74	21.297	5.786	12	41
rep78	69	3.406	.99	1	5
headroom	74	2.993	.846	.846	5
trunk	74	13.757	4.277	5	23
weight	74	3019.459	777.194	1760	4840
length	74	187.932	22.266	142	233
turn	74	39.649	4.399	31	51
displacement	74	197.297	91.837	79	425
gear_ratio	74	3.015	.456	2.19	3.89
foreign	74	.297	.46	0	1

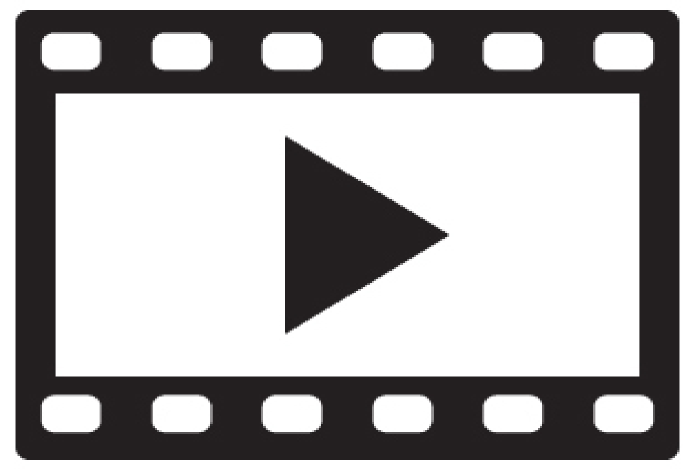
Organize large quantities of information

Course Methodology

Why?

1. Section goals

How?



2. Short video lessons

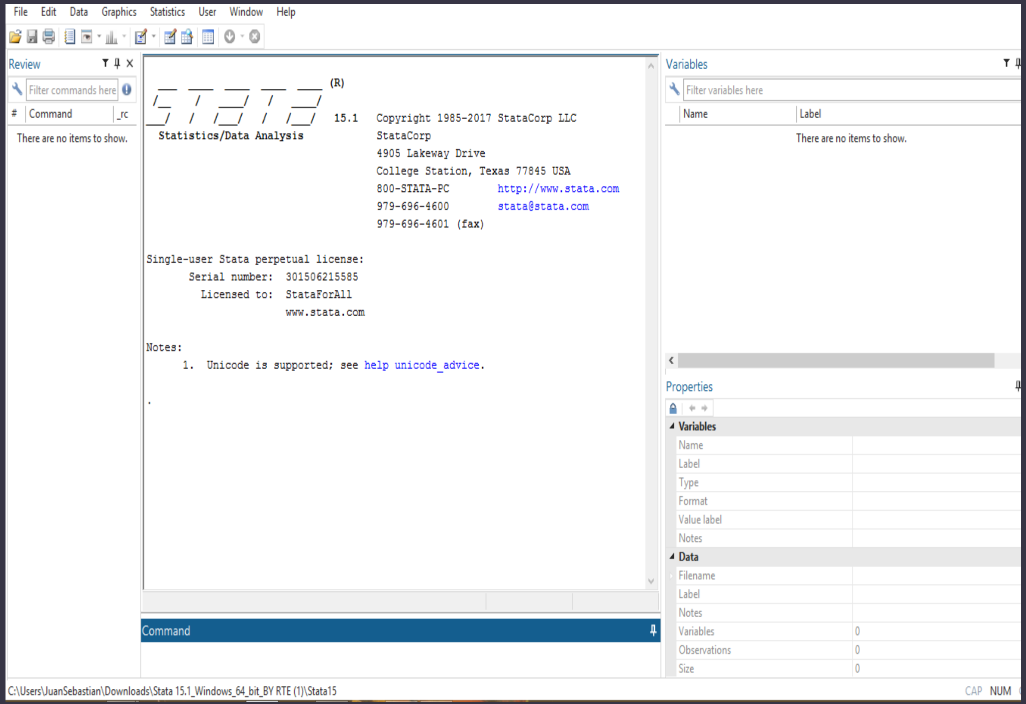
Sure?

3. Comprehension Check

Goal 2 - Where do I Start?

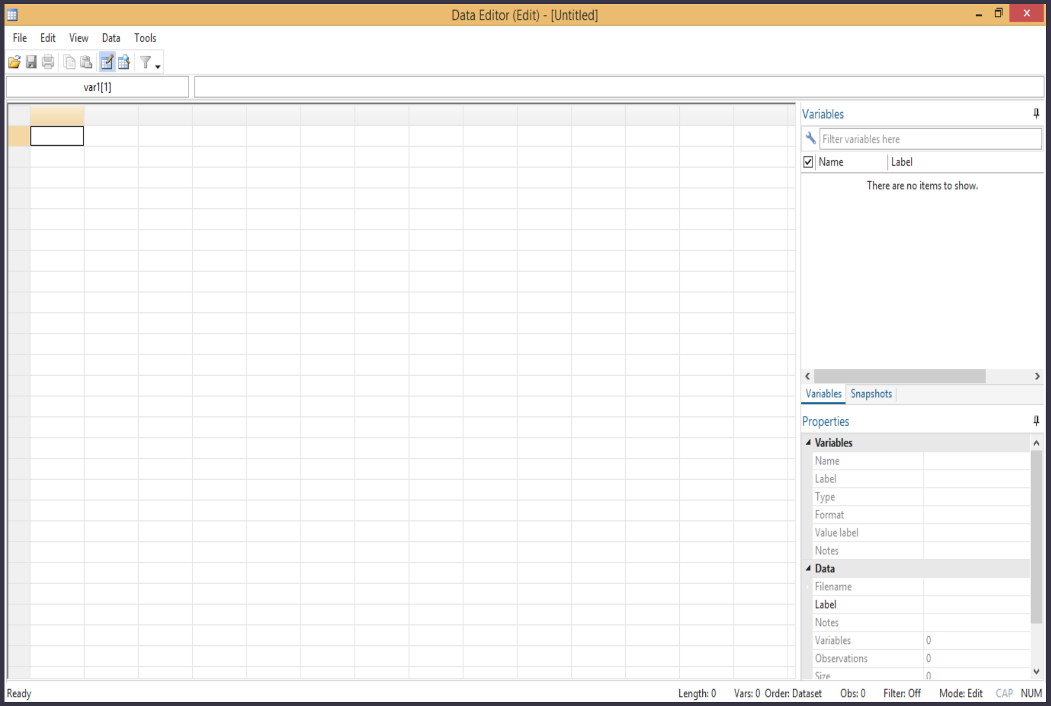
JUAN SEBASTIAN CUERVO

Goal 2 - Where do I Start ?



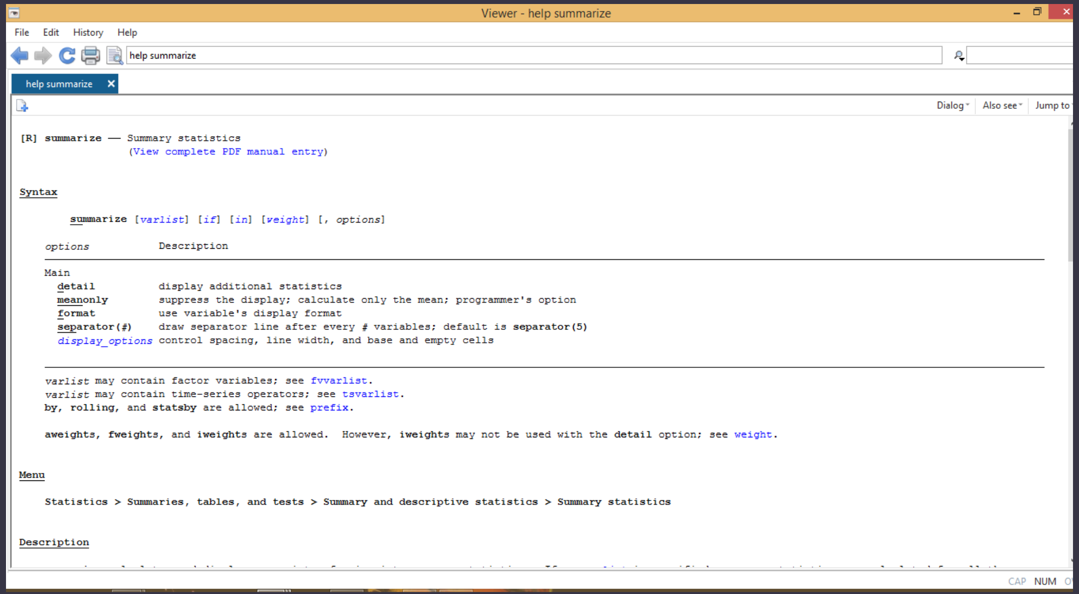
```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Single-user Stata perpetual license:
       Serial number:  301506215585
         Licensed to:  StataForAll
                       www.stata.com

Notes:
      1.  Unicode is supported; see help unicode_advice.

.
```

Stata interface and windows



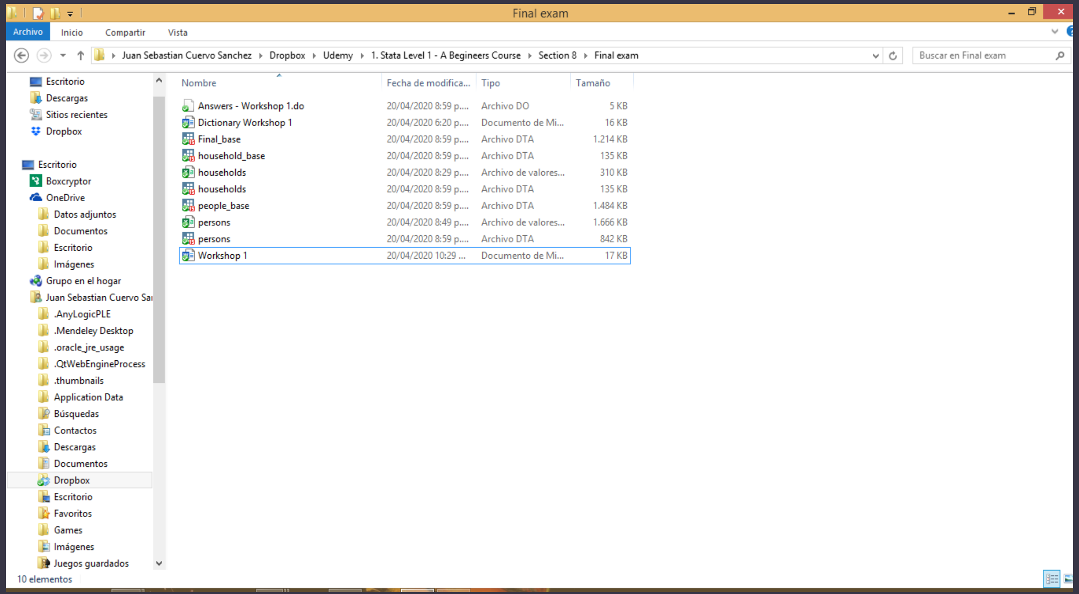
Data editor and browse

Goal 2 - Where do I Start ?



```
[R] summarize — Summary statistics
              (View complete PDF manual entry)


Syntax

        summarize [varlist] [if] [in] [weight] [, options]

    options          Description
    ------------------------------------------------------------------------
    Main
      detail         display additional statistics
      meanonly       suppress the display; calculate only the mean; programmer's option
      format         use variable's display format
      separator(#)   draw separator line after every # variables; default is separator(5)
      display_options control spacing, line width, and base and empty cells
    ------------------------------------------------------------------------
    varlist may contain factor variables; see fvvarlist.
    varlist may contain time-series operators; see tsvarlist.
    by, rolling, and statsby are allowed; see prefix.

    aweights, fweights, and iweights are allowed.  However, iweights may not be used with the detail option; see weight.


Menu

    Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics


Description
```

Help and search in Stata



Nombre	Fecha de modifica...	Tipo	Tamaño
Answers - Workshop 1.do	20/04/2020 8:59 p....	Archivo DO	5 KB
Dictionary Workshop 1	20/04/2020 6:20 p....	Documento de Mi...	16 KB
Final_base	20/04/2020 8:59 p....	Archivo DTA	1.214 KB
household_base	20/04/2020 8:59 p....	Archivo DTA	135 KB
households	20/04/2020 8:29 p....	Archivo de valores...	310 KB
households	20/04/2020 8:59 p....	Archivo DTA	135 KB
people_base	20/04/2020 8:59 p....	Archivo DTA	1.484 KB
persons	20/04/2020 8:49 p....	Archivo de valores...	1.666 KB
persons	20/04/2020 8:59 p....	Archivo DTA	842 KB
Workshop 1	20/04/2020 10:29 ...	Documento de Mi...	17 KB

Establish a working directory

Goal 3 - How to import data in Stata?

JUAN SEBASTIAN CUERVO

Goal 3: How to import data in Stata?

Stata (.dta)

Text (.txt)

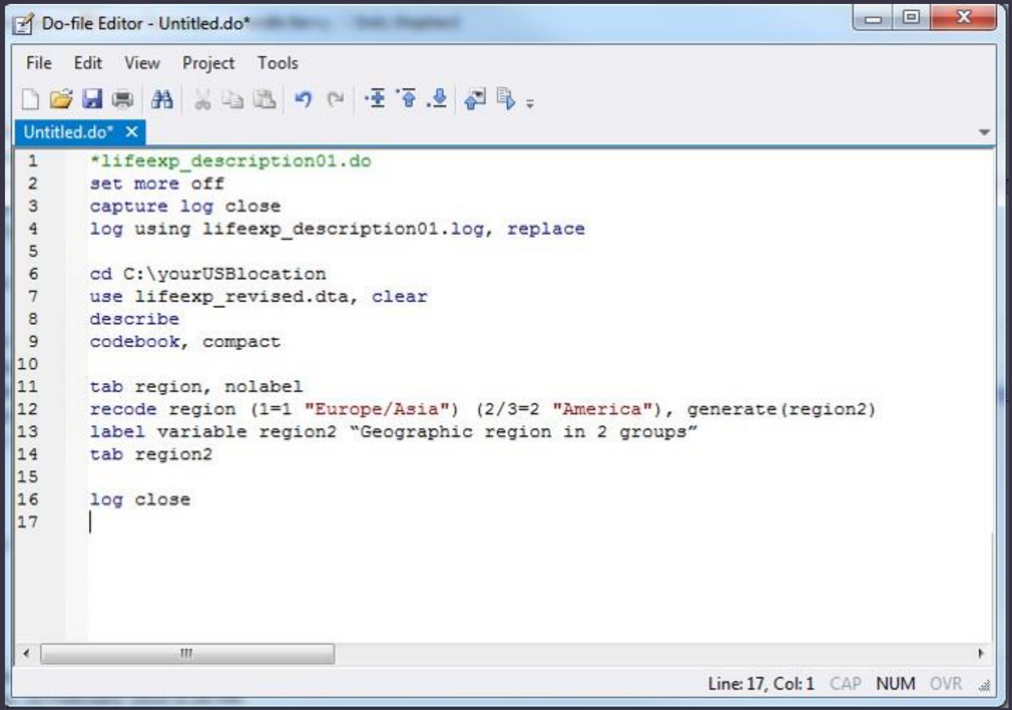
Excel (.dta)

Save files

Goal 4 - Why keep track of your work in Stata?

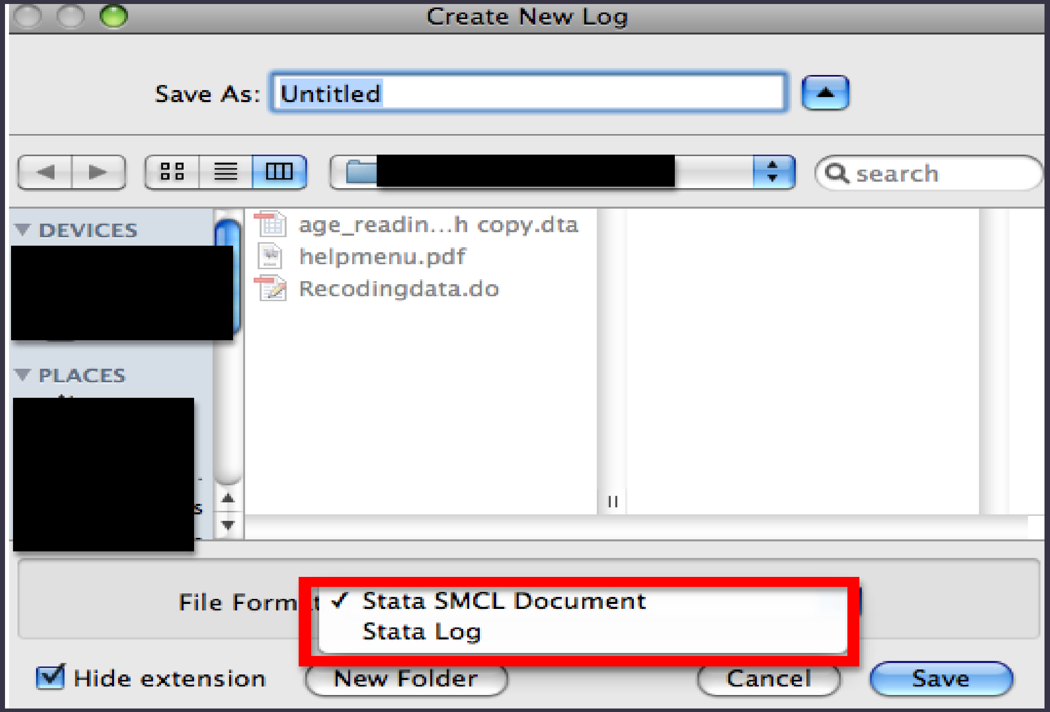
JUAN SEBASTIAN CUERVO

Goal 4: Why keep track of your work in Stata?



```
*lifeexp_description01.do
set more off
capture log close
log using lifeexp_description01.log, replace

cd C:\yourUSBlocation
use lifeexp_revised.dta, clear
describe
codebook, compact

tab region, nolabel
recode region (1=1 "Europe/Asia") (2/3=2 "America"), generate(region2)
label variable region2 “Geographic region in 2 groups”
tab region2

log close
```

Data reproducibility - Do file



Save As: Untitled

[redacted]

[redacted]

[redacted]

age_readin...h copy.dta
helpmenu.pdf
Recodingdata.do

File Format: ✓ Stata SMCL Document
Stata Log

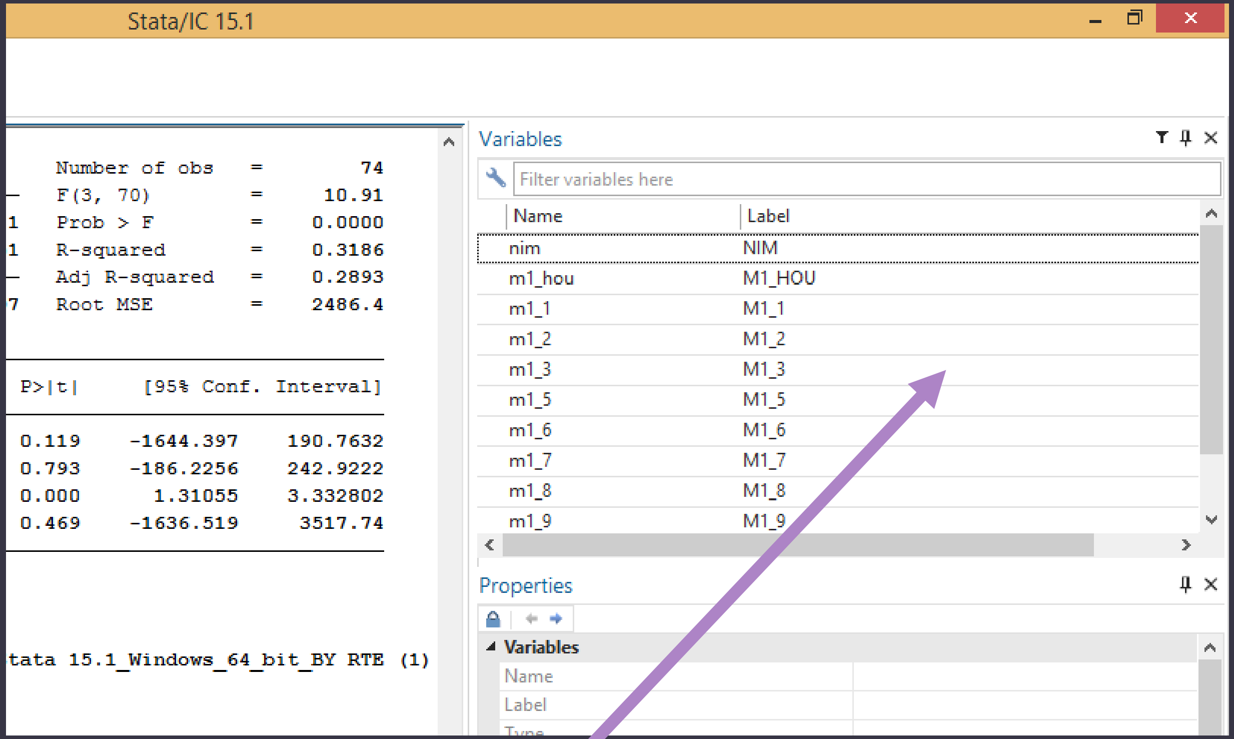
☑ Hide extension

Log files

Goal 5 - Why do I want to alter my dataset?

JUAN SEBASTIAN CUERVO

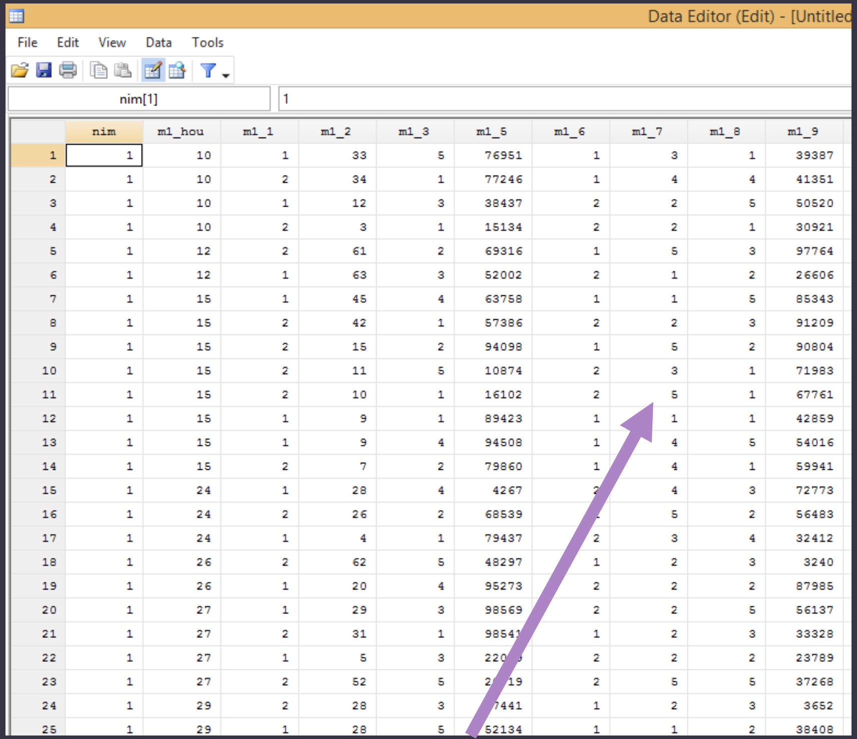
Goal 5: Why do I want to alter my dataset?



```
   Number of obs    =        74
—  F(3, 70)         =     10.91
1  Prob > F         =    0.0000
1  R-squared        =    0.3186
—  Adj R-squared    =    0.2893
7  Root MSE         =    2486.4

 P>|t|      [95% Conf. Interval]

 0.119    -1644.397    190.7632
 0.793    -186.2256    242.9222
 0.000      1.31055    3.332802
 0.469    -1636.519     3517.74

tata 15.1_Windows_64_bit_BY RTE (1)
```

Name	Label
nim	NIM
m1_hou	M1_HOU
m1_1	M1_1
m1_2	M1_2
m1_3	M1_3
m1_5	M1_5
m1_6	M1_6
m1_7	M1_7
m1_8	M1_8
m1_9	M1_9

What are these variables?



nim[1] 1

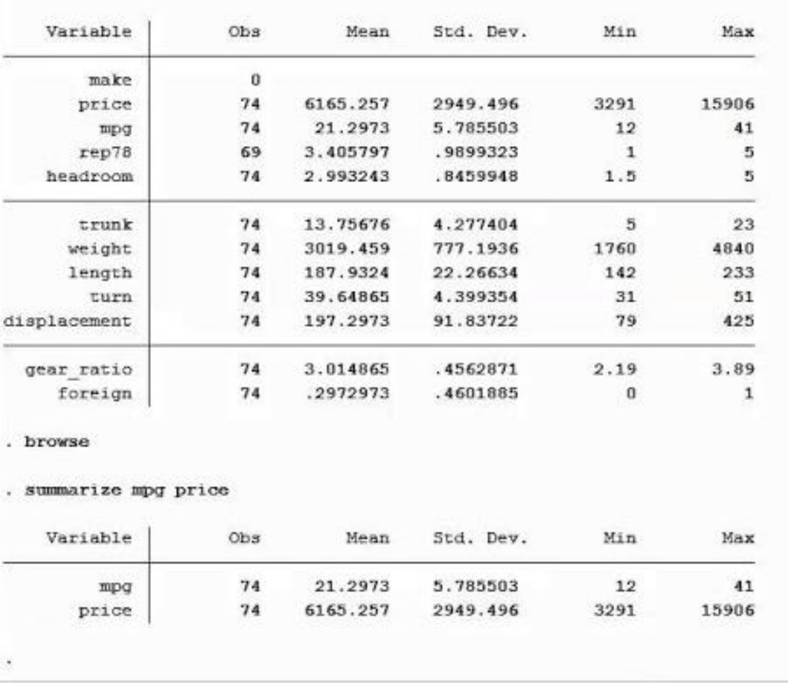
	nim	m1_hou	m1_1	m1_2	m1_3	m1_5	m1_6	m1_7	m1_8	m1_9
1	1	10	1	33	5	76951	1	3	1	39387
2	1	10	2	34	1	77246	1	4	4	41351
3	1	10	1	12	3	38437	2	2	5	50520
4	1	10	2	3	1	15134	2	2	1	30921
5	1	12	2	61	2	69316	1	5	3	97764
6	1	12	1	63	3	52002	2	1	2	26606
7	1	15	1	45	4	63758	1	1	5	85343
8	1	15	2	42	1	57386	2	2	3	91209
9	1	15	2	15	2	94098	1	5	2	90804
10	1	15	2	11	5	10874	2	3	1	71983
11	1	15	2	10	1	16102	2	5	1	67761
12	1	15	1	9	1	89423	1	1	1	42859
13	1	15	1	9	4	94508	1	4	5	54016
14	1	15	2	7	2	79860	1	4	1	59941
15	1	24	1	28	4	4267	2	4	3	72773
16	1	24	2	26	2	68539	1	5	2	56483
17	1	24	1	4	1	79437	2	3	4	32412
18	1	26	2	62	5	48297	1	2	3	3240
19	1	26	1	20	4	95273	2	2	2	87985
20	1	27	1	29	3	98569	2	2	5	56137
21	1	27	2	31	1	98541	1	2	3	33328
22	1	27	1	5	3	220[illegible]	2	2	2	23789
23	1	27	2	52	5	2[illegible]19	2	5	5	37268
24	1	29	2	28	3	[illegible]7441	1	2	3	3652
25	1	29	1	28	5	52134	1	1	2	38408

What does these values mean?

Goal 6 - How to analyze my data ?

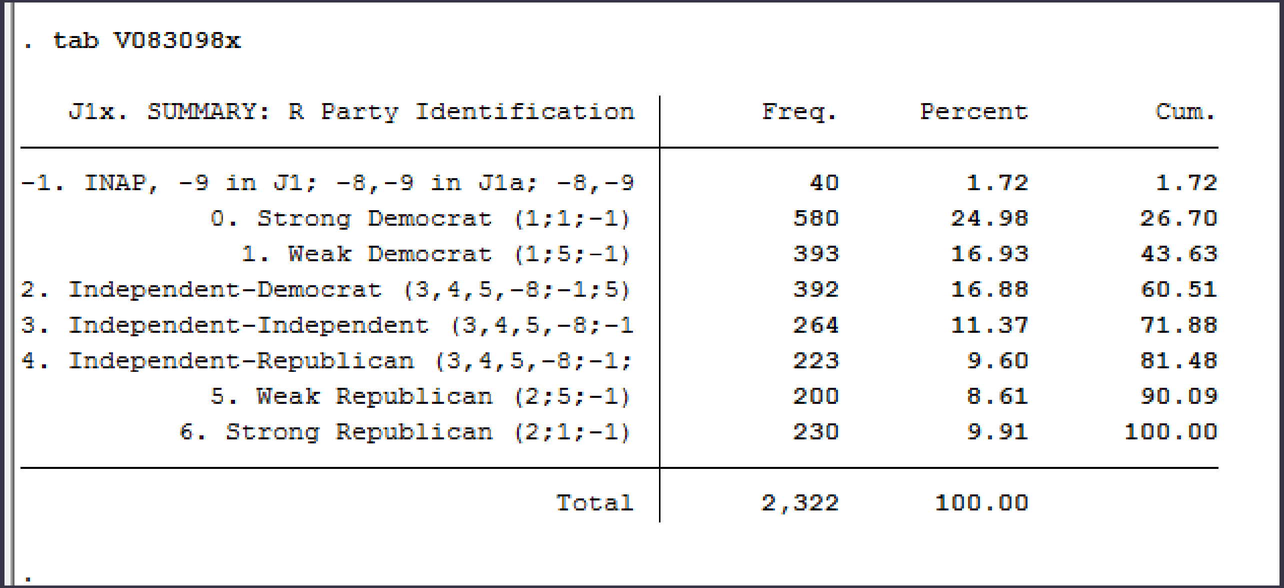
JUAN SEBASTIAN CUERVO

Goal 6: How to analyze my data?



```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        make |         0
       price |        74    6165.257    2949.496       3291      15906
         mpg |        74     21.2973    5.785503         12         41
       rep78 |        69    3.405797    .9899323          1          5
    headroom |        74    2.993243    .8459948        1.5          5
-------------+---------------------------------------------------------
       trunk |        74    13.75676    4.277404          5         23
      weight |        74    3019.459    777.1936       1760       4840
      length |        74    187.9324    22.26634        142        233
        turn |        74    39.64865    4.399354         31         51
displacement |        74    197.2973    91.83722         79        425
-------------+---------------------------------------------------------
  gear_ratio |        74    3.014865    .4562871       2.19       3.89
     foreign |        74    .2972973    .4601885          0          1

. browse

. summarize mpg price

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         mpg |        74     21.2973    5.785503         12         41
       price |        74    6165.257    2949.496       3291      15906

.
```

Obtain summary statistics



```
. tab V083098x

   J1x. SUMMARY: R Party Identification |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
-1. INAP, -9 in J1; -8,-9 in J1a; -8,-9 |         40        1.72        1.72
             0. Strong Democrat (1;1;-1) |        580       24.98       26.70
               1. Weak Democrat (1;5;-1) |        393       16.93       43.63
2. Independent-Democrat (3,4,5,-8;-1;5) |        392       16.88       60.51
3. Independent-Independent (3,4,5,-8;-1 |        264       11.37       71.88
4. Independent-Republican (3,4,5,-8;-1; |        223        9.60       81.48
             5. Weak Republican (2;5;-1) |        200        8.61       90.09
           6. Strong Republican (2;1;-1) |        230        9.91      100.00
----------------------------------------+-----------------------------------
                                  Total |      2,322      100.00

.
```

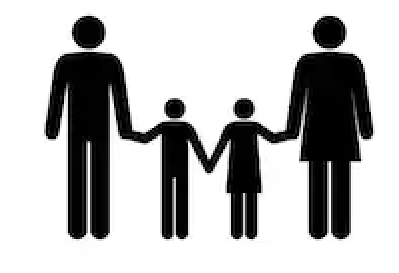
Obtain frequencies, percents

Goal 7 - Why do I need to join databases?

JUAN SEBASTIAN CUERVO

Goal 7: Why do I need to join databases?

Household database



People database

Cross information

Household	People
1	1
1	2
1	3
1	4
2	1
2	2

Goal 7: Why do I need to join databases?

Database 1

Key	Var 1	Var 2
1		
2		
3		
4		
5		

Database 2

Key	Var 3	Var 4
1		
2		
3		
4		
5		

Merge

Key
variable

Resulting database

Key	Var 1	Var 2	Var 3	Var 4
1				
2				
3				
4				
5				

Goal 7: Why do I need to join databases?

Database 1

Key	Var 1	Var 2
1		
2		
3		
4		
5		

Database 2

Key	Var 1	Var 2
6		
7		
8		

Append

Resulting database

Key	Var 1	Var 2
1		
2		
3		
4		
5		
6		
7		
8		

Goal 7: Why do I need to join databases?

Database 1

Year	Production
1992	10.000
1992	16.000
1993	5.000
1993	7.000
1993	4.000
1993	4.000
1994	3.000
1994	3.000

Collapse
Statistic
(Mean)

Results database

Year	Mean Production
1992	13.000
1993	5.000
1994	3.000

Goal 7: Why do I need to join databases?

Reshape

Database 1

Var 1	Var 2	Income
1	1	10.000
1	2	30.000
1	3	40.000
2	1	15.000
2	2	8.000

Reshape

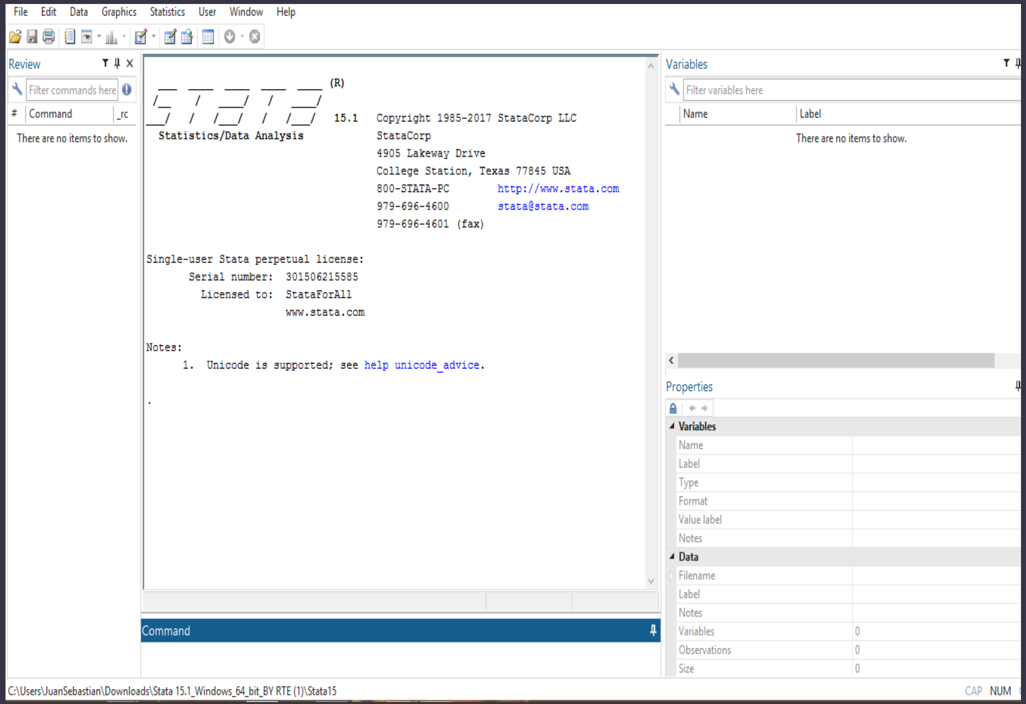
Results database

Var 1	Income 1	Income 2	Income 3
1	10.000	30.000	40.000
2	15.000	8.000	-

Recap Course

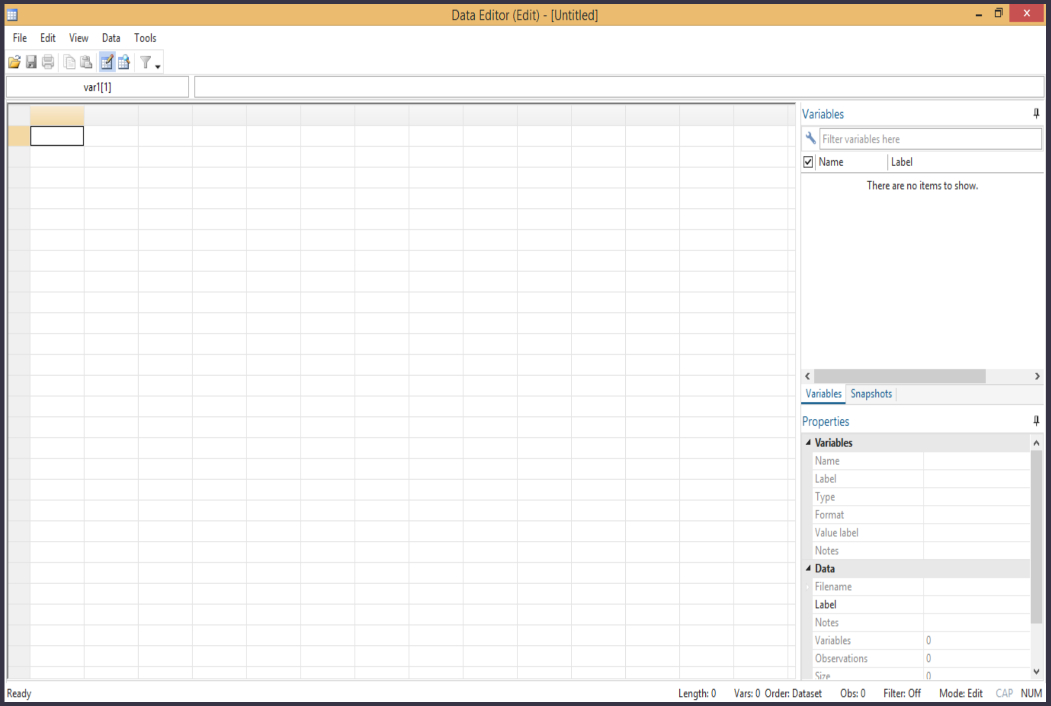
JUAN SEBASTIAN CUERVO

Goal 1: Why use Stata? - Level 1



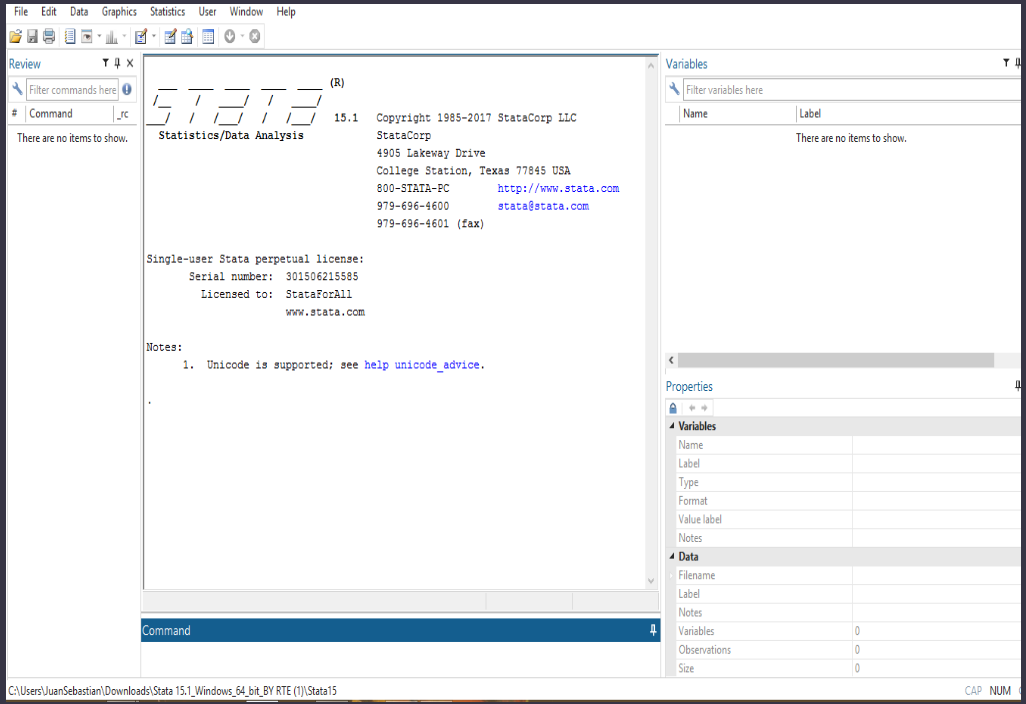
```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Single-user Stata perpetual license:
       Serial number:  301506215585
         Licensed to:  StataForAll
                       www.stata.com

Notes:
      1.  Unicode is supported; see help unicode_advice.

.
```

Stata interface and windows



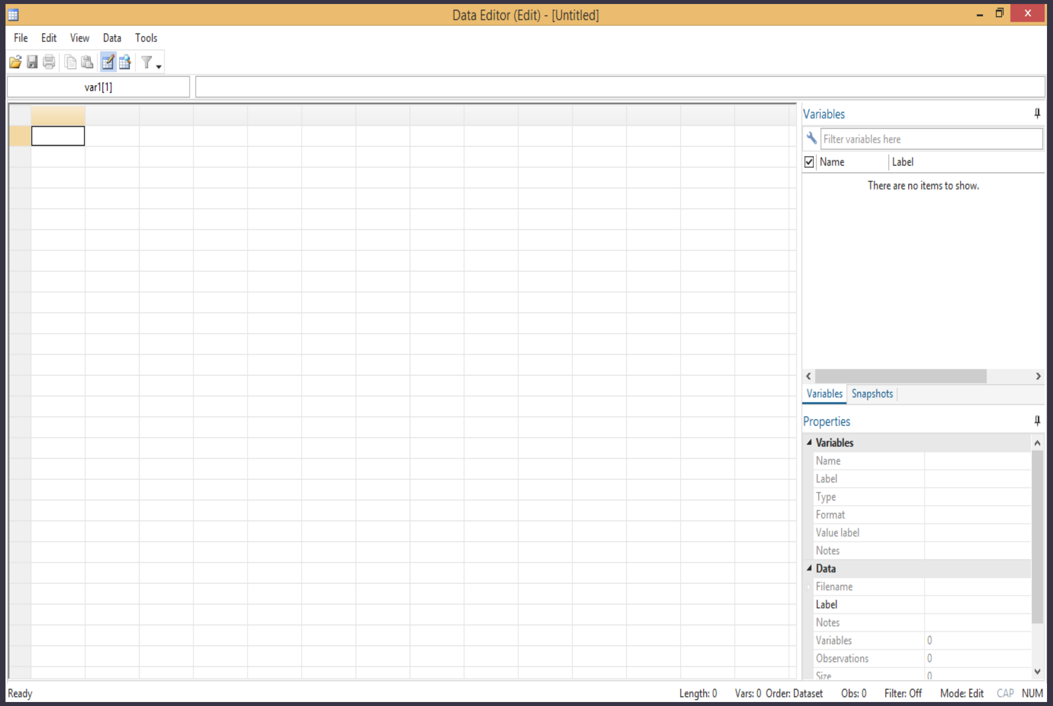
Data editor and browse

Goal 2 - Where do I Start ?



```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.1   Copyright 1985-2017 StataCorp LLC
  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 perpetual license:
       Serial number:  301506215585
         Licensed to:  StataForAll
                       www.stata.com

Notes:
      1.  Unicode is supported; see help unicode_advice.

.
```

Stata interface and windows



Data editor and browse

Goal 3: How to import data in Stata?

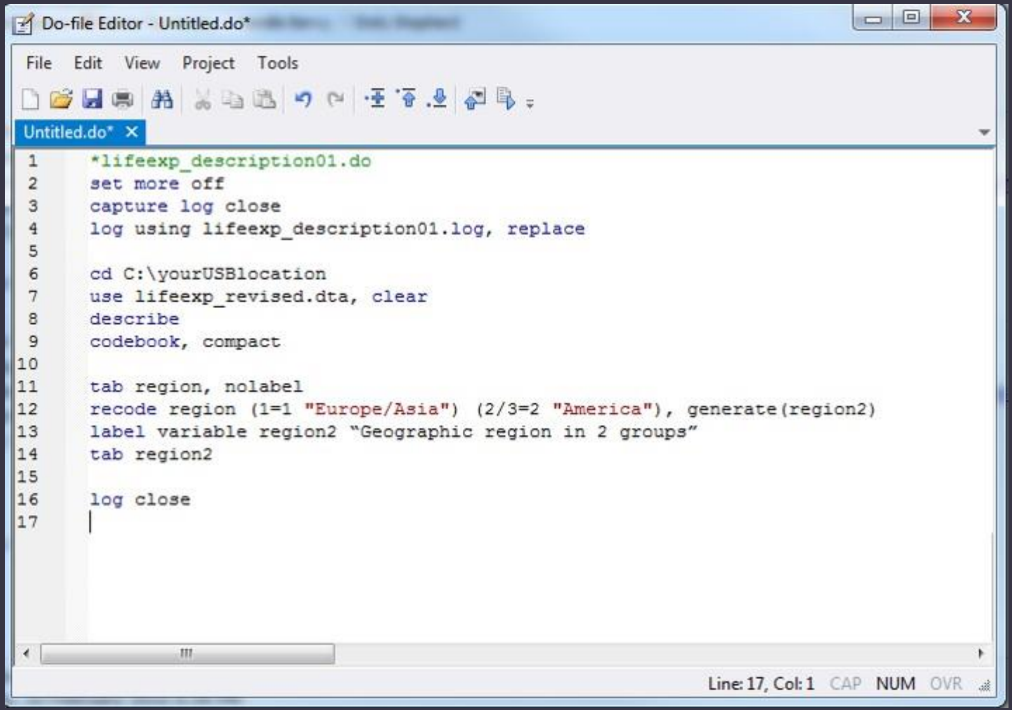
Stata (.dta)

Text (.txt)

Excel (.dta)

Save files

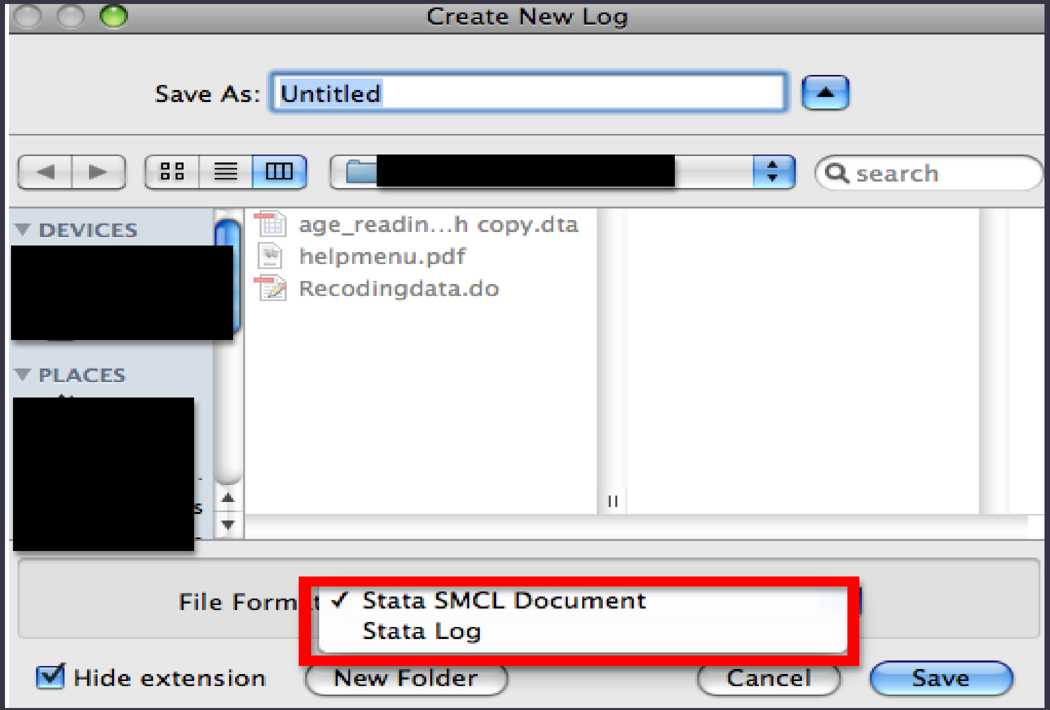
Goal 4: Why keep track of your work in Stata?



Do-file Editor - Untitled.do*
File Edit View Project Tools
Untitled.do*
Line: 17, Col: 1 CAP NUM OVR

```
*lifeexp_description01.do
set more off
capture log close
log using lifeexp_description01.log, replace

cd C:\yourUSBlocation
use lifeexp_revised.dta, clear
describe
codebook, compact

tab region, nolabel
recode region (1=1 "Europe/Asia") (2/3=2 "America"), generate(region2)
label variable region2 “Geographic region in 2 groups”
tab region2

log close
```

Data reproducibility - Do file



Save As: Untitled

[redacted]

[redacted]

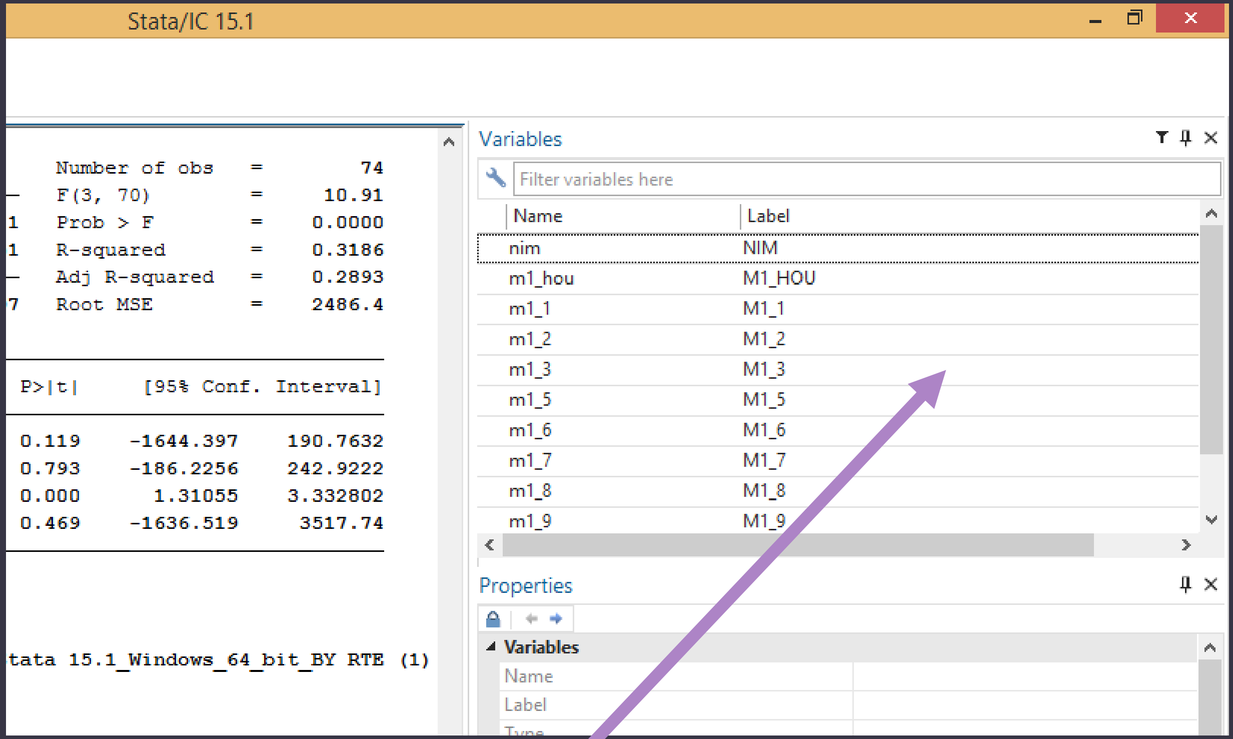
[redacted]

File Format: ✓ Stata SMCL Document
Stata Log

☑ Hide extension

Log files

Goal 5: Why do I want to alter my dataset?



```
    Number of obs    =        74
—   F(3, 70)         =     10.91
1   Prob > F         =    0.0000
1   R-squared        =    0.3186
—   Adj R-squared    =    0.2893
7   Root MSE         =    2486.4

 P>|t|     [95% Conf. Interval]

 0.119    -1644.397    190.7632
 0.793    -186.2256    242.9222
 0.000      1.31055    3.332802
 0.469    -1636.519     3517.74


tata 15.1_Windows_64_bit_BY RTE (1)
```

Variables

Filter variables here

Name	Label
nim	NIM
m1_hou	M1_HOU
m1_1	M1_1
m1_2	M1_2
m1_3	M1_3
m1_5	M1_5
m1_6	M1_6
m1_7	M1_7
m1_8	M1_8
m1_9	M1_9

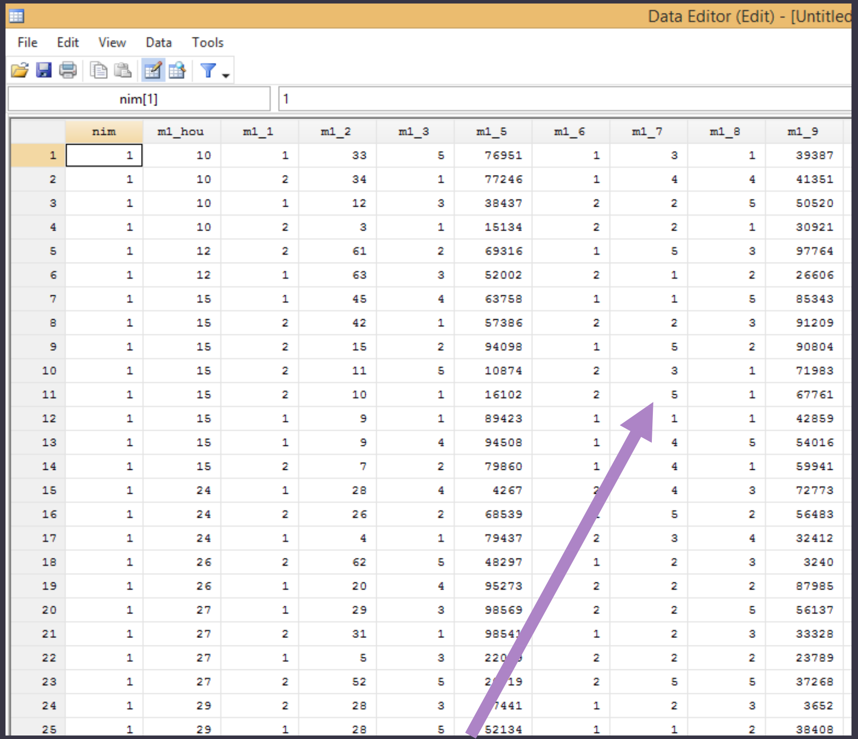
Properties

Variables

Name

Label

Type



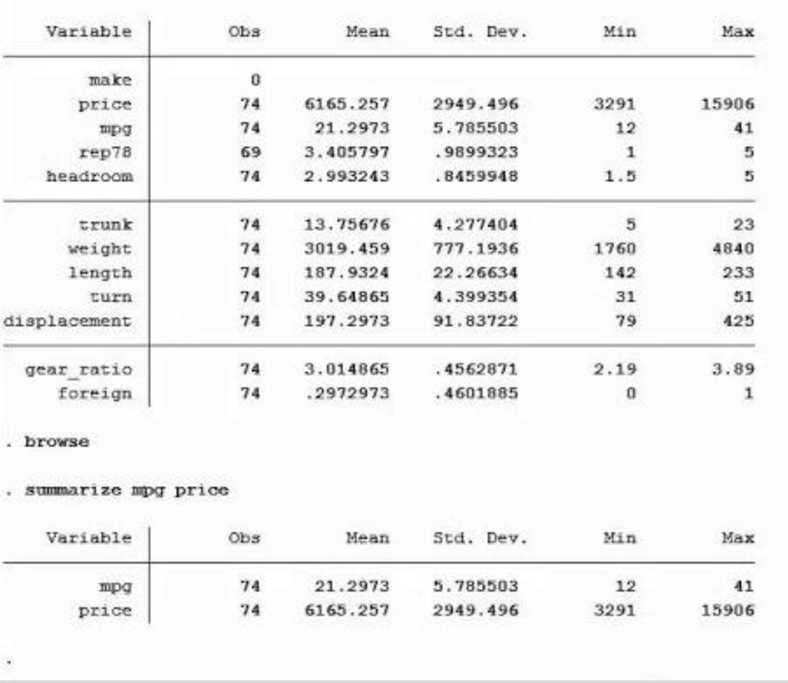
nim[1] 1

	nim	m1_hou	m1_1	m1_2	m1_3	m1_5	m1_6	m1_7	m1_8	m1_9
1	1	10	1	33	5	76951	1	3	1	39387
2	1	10	2	34	1	77246	1	4	4	41351
3	1	10	1	12	3	38437	2	2	5	50520
4	1	10	2	3	1	15134	2	2	1	30921
5	1	12	2	61	2	69316	1	5	3	97764
6	1	12	1	63	3	52002	2	1	2	26606
7	1	15	1	45	4	63758	1	1	5	85343
8	1	15	2	42	1	57386	2	2	3	91209
9	1	15	2	15	2	94098	1	5	2	90804
10	1	15	2	11	5	10874	2	3	1	71983
11	1	15	2	10	1	16102	2	5	1	67761
12	1	15	1	9	1	89423	1	1	1	42859
13	1	15	1	9	4	94508	1	4	5	54016
14	1	15	2	7	2	79860	1	4	1	59941
15	1	24	1	28	4	4267	2	4	3	72773
16	1	24	2	26	2	68539	[illegible]	5	2	56483
17	1	24	1	4	1	79437	2	3	4	32412
18	1	26	2	62	5	48297	1	2	3	3240
19	1	26	1	20	4	95273	2	2	2	87985
20	1	27	1	29	3	98569	2	2	5	56137
21	1	27	2	31	1	98541	1	2	3	33328
22	1	27	1	5	3	[illegible]	2	2	2	23789
23	1	27	2	52	5	[illegible]	2	5	5	37268
24	1	29	2	28	3	[illegible]	1	2	3	3652
25	1	29	1	28	5	52134	1	1	2	38408

What are these variables?

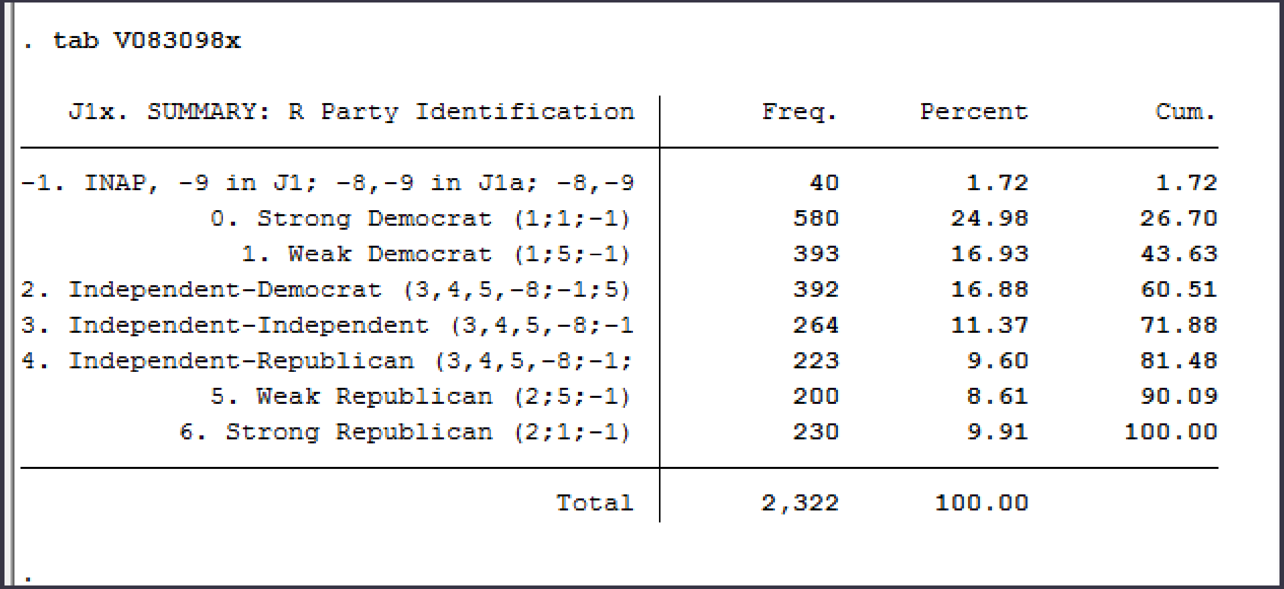
What does these values mean?

Goal 6: How to analyze my data?



```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        make |          0
       price |         74    6165.257    2949.496       3291      15906
         mpg |         74     21.2973    5.785503         12         41
       rep78 |         69    3.405797    .9899323          1          5
    headroom |         74    2.993243    .8459948        1.5          5
-------------+---------------------------------------------------------
       trunk |         74    13.75676    4.277404          5         23
      weight |         74    3019.459    777.1936       1760       4840
      length |         74    187.9324    22.26634        142        233
        turn |         74    39.64865    4.399354         31         51
displacement |         74    197.2973    91.83722         79        425
-------------+---------------------------------------------------------
  gear_ratio |         74    3.014865    .4562871       2.19       3.89
     foreign |         74    .2972973    .4601885          0          1

. browse

. summarize mpg price

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         mpg |         74     21.2973    5.785503         12         41
       price |         74    6165.257    2949.496       3291      15906

.
```

Obtain summary statistics



```
. tab V083098x

   J1x. SUMMARY: R Party Identification |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
-1. INAP, -9 in J1; -8,-9 in J1a; -8,-9 |         40        1.72        1.72
            0. Strong Democrat (1;1;-1) |        580       24.98       26.70
              1. Weak Democrat (1;5;-1) |        393       16.93       43.63
2. Independent-Democrat (3,4,5,-8;-1;5) |        392       16.88       60.51
3. Independent-Independent (3,4,5,-8;-1 |        264       11.37       71.88
4. Independent-Republican (3,4,5,-8;-1; |        223        9.60       81.48
            5. Weak Republican (2;5;-1) |        200        8.61       90.09
          6. Strong Republican (2;1;-1) |        230        9.91      100.00
----------------------------------------+-----------------------------------
                                  Total |      2,322      100.00

.
```

Obtain frequencies, percents

Goal 7: Why do I need to join databases?

Database 1

Key	Var 1	Var 2
1		
2		
3		
4		
5		

Database 2

Key	Var 3	Var 4
1		
2		
3		
4		
5		

Merge

Key variable

Resulting database

Key	Var 1	Var 2	Var 3	Var 4
1				
2				
3				
4				
5				