

Surviving Stata Workshop

infoguides.gmu.edu/software/stata

dsc.gmu.edu/software

Please open Stata

Statistical Software



Easy to start, limited capability

Best for those with basic needs or who analyze experiments



Easy flexible syntax, extensible

Best for academics, especially in economics & public policy



Hard to learn, highly capable

Best for managing huge and/or complex data (govt, financial)

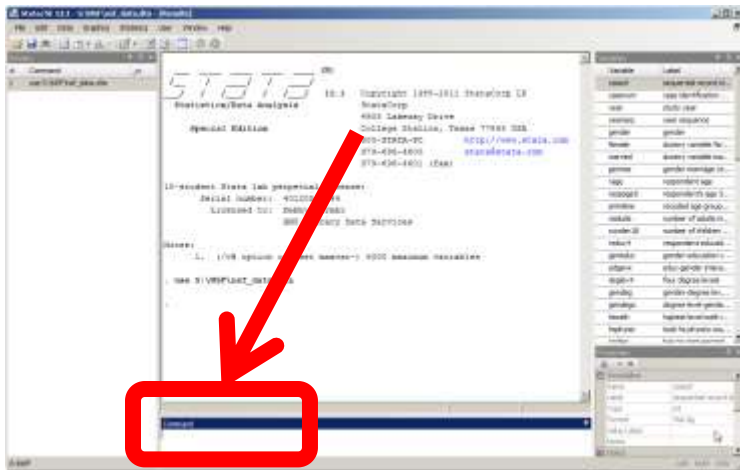


Improving, highly extensible

Best for those who program and are comfortable with data

Submitting Stata Commands

1. Click in the
Command Window



2. Type:
`log using workshop.log`

```
Command  
log using workshop.log
```

3. Press **Enter**

```
. log using workshop.log
```

```
name: <unnamed>  
log: C:\tools\stata\workshop.log  
log type: text
```

Stata Window

The screenshot shows the Stata/SE 15.1 window with several panels and annotations:

- Review Panel:** Located on the left, it contains a search box "Filter commands here" and a table with columns "#", "Command", and "_rc". The command ". log using workshop.log" is listed.
- Results Panel:** The central area displays the output of the command, including the Stata logo, version 15.1, copyright information, and license details for Wendy Mann at George Mason University.
- Variables Panel:** On the right, it shows "Filter variables here" and a message "There are no items to show."
- Properties Panel:** Below the Variables panel, it shows a tree view with "Variables" and "Data" sections, and a table with properties like Name, Label, Type, Format, Value label, Notes, Filename, Label, Notes, Variables (0), Observations (0), Size (0), and Memory (64M).
- Command Panel:** At the bottom, it shows the command ". log using workshop.log" and its execution details: name: <unnamed>, log: C:\tools\stata\workshop.log, log type: text, opened on: 26 Sep 2019, 17:02:56.

Annotations in blue text are placed over the interface:

- "Review" is placed over the Review panel.
- "Variables" is placed over the Variables panel.
- "Results" is placed over the central output area.
- "Command" is placed over the Command panel.

Additional text on the left side of the image reads: "Single Click = Put it back in Command box".

At the bottom left, the path "C:\tools\stata" is visible. At the bottom right, the status bar shows "CAP NUM OVR" and the number "4".

Opening Files

**Those participating in the workshop will be provided the data file.
The following instructions enable others to get and open the file.**

Getting the Data

1. Register for an account at

<https://www.pewsocialtrends.org/profile/registration/>

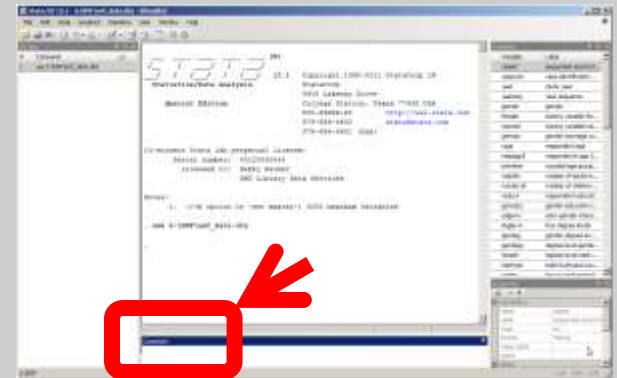
2. Download the datafile from:

<https://www.pewsocialtrends.org/2007/11/09/family-bonds-data/>

3. Unzip the folder to a location you can find again

4. Open Stata

5. Paste or type the following in the Command box; press Enter



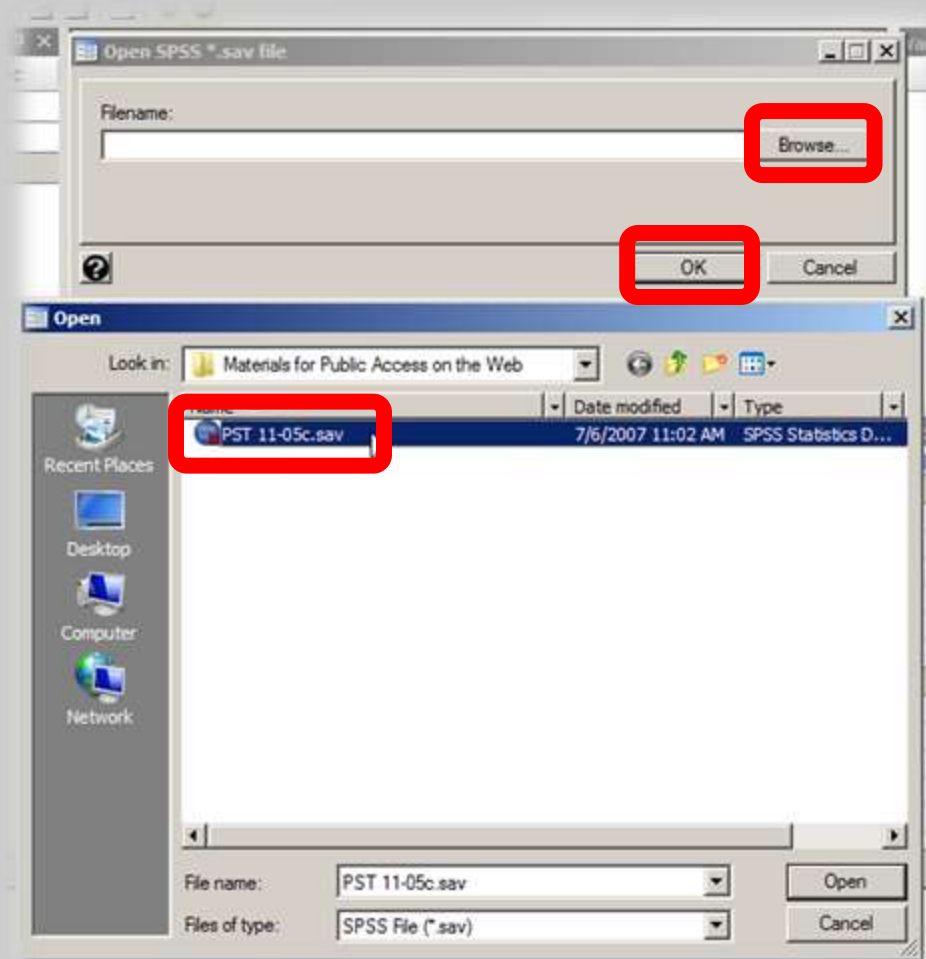
```
net from http://radyakin.org/transfer/usespss/beta
net install usespss
usespss
```

Select the File

Press "Browse..."
in the dialog box


Find and select
the data file (.sav)

Press OK



Stata Vocabulary

`log using workshop.log`

`use https://dsc.gmu.edu/files/` 

use `https://dsc.gmu.edu/files/pew.dta`

The screenshot shows the Stata/SE 15.1 interface. The main window displays the output of the command `use https://dsc.gmu.edu/files/pew.dta`. The output includes the Stata logo, version 15.1, copyright information (1985-2017 StataCorp LLC), and contact details for StataCorp. It also shows the license information for a 20-student Stata lab perpetual license, licensed to Wendy Mann at George Mason University. The command window shows the following commands and their output:

```
. log using workshop.log
. use https://dsc.gmu.edu/files/pew.dta
```

The command window also shows the following output:

```
name: <unnamed>
log: C:\tools\stata\workshop.log
log type: text
opened on: 26 Sep 2019, 17:02:56
```

The variables list on the right side of the window is circled in red. It contains the following variables:

Name	Label
psraid	Unique ID
respid	Record
weight	
int_date	Interview date
lang	Language of Inter...
area	Area code
state	State from FIPS
region	Census region
msa	MSA code
fips	FIPS code
usr	Community type
usr1	Community ty...

The command window also shows the following output:

```
. log using workshop.log
```

The command window also shows the following output:

```
name: <unnamed>
log: C:\tools\stata\workshop.log
log type: text
opened on: 26 Sep 2019, 17:02:56
```

The command window also shows the following output:

```
. use https://dsc.gmu.edu/files/pew.dta
```

The command window also shows the following output:

```
.
```

Command

`use https://dsc.gmu.edu/files`

Looking at Data

browse

Pew Research Center. (2007). Pew Social Trends [Data file]. Retrieved from <http://www.pewsocialtrends.org/category/datasets/?download=5753>

browse

The screenshot shows the SPSS Data Editor window for a dataset named 'pew.dta'. The main window displays a grid of data with columns: psraid, respid, weight, int_date, lang, area, state, and cregion. A purple box highlights the 'lang' column, with the text 'Responses to one question (variable)' overlaid. A red box highlights the 'lang' cell for the 11th row, with the text 'A single response (value)' overlaid. A green box highlights the entire 11th row, with the text 'Responses from one person (observation)' overlaid. The right-hand side of the window shows the 'Variables' and 'Properties' panels. The 'Variables' panel lists the dataset's variables, and the 'Properties' panel shows details for the selected 'psraid' variable.

	psraid	respid	weight	int_date	lang	area	state	cregion
1	1	1	3.0208333	101705	Spanish	727	Florida	south
2	2	1	2.7516667	100505	English	386	Florida	south
3	3	5	2.7516667	100505	Spanish	770	Georgia	south
4	4	5	2.7516667	100505	Spanish	718	New York	north...
5	5	5	2.7516667	100505	English	215	Penns...	north...
6	6	5	1.0208333	101705	Spanish	727	Florida	south
7	7	5	1.2916667	100505	English	727	Florida	south
8	8	5	3.1458333	101705	Spanish	727	Florida	north...
9	9	6	2.3333333	100905	English	727	Florida	south
10	10	6	2.2291667	101705	Spanish	727	Florida	north...
11	11	7	1.0208333	101705	Spanish	816	Missouri	midwest
12	12	7	1.125	100505	English	850	Florida	south
13	13	7	3.0208333	100705	English	813	Florida	south
14	14	7	3.0208333	100705	English	276	Virginia	south
15	15	7	3.0208333	100705	Spanish	408	Calif...	west
16	16	7	3.0208333	100705	Spanish	626	Calif...	west
17	17	7	3.0208333	100905	English	518	New York	north...
18	18	11	3.71	101805	Spanish	908	New J...	north...

Variables

Name	Label
psraid	Unique ID
respid	Record
weight	
int_date	Interview date
lang	Language of Inter...
area	Area code
state	State from FIPS

Properties

Name	psraid
Label	Unique ID
Type	int
Format	%8.0g
Value label	
Notes	

Navigating Variables

The screenshot shows the SPSS interface. At the top, the title bar reads "seekid3[1]". Below it is a menu bar with "File" and a toolbar with icons for file operations and filters. The main area is a data grid with columns: seemom2, q4041, seemom3, seekid2, q6768, seekid3, and o1. The "seekid3" column is highlighted in yellow. To the right, the "Variables" list is open, showing a search bar containing "see" (highlighted with a red box) and a list of variables. The "seekid3" variable is highlighted with a dotted border (highlighted with a black box). Below the list, the "Properties" window shows details for "seekid3": Name: seekid3, Label: Combined Q65 ar, Type: byte, Format: %8.0g, Value label: LABY. The "Data" section is also visible.

	seemom2	q4041	seemom3	seekid2	q6768	seekid3	o1
1	Every...	Less ...	Every...
2	.	.	.	Once ...	Once ...	Once
3	Once ...	Less ...	Once
4	Once ...	Less ...	Once
5	Once ...	Less ...	Once
6	Once ...	Less ...	Once
7	Once ...	Less ...	Once ...	Once ...	Once ...	Once
8	Every...	Once ...	Every...
9	Once ...	Less ...	Once ...	Once ...	Less ...	Once
10	Every...	Less ...	Every...	Every...	Less ...	Every...	.
11	.	.	.	Every...	Less ...	Every...	.
12	.	.	.	Once ...	Once ...	Once
13	DK/Ref	DK/Ref	DK/Ref
14	Every...	Less ...	Every...
15	Once ...	Less ...	Once
16	Once ...	Less ...	Once
17	.	.	.	Every...	Less ...	Every...	.
18	.	.	.	Every...	Less ...	Every...	.
19	.	.	.	Once ...	Less ...	Once

Type "see"

Double-Click

Values

seekid3[417]				
	seemom3	seekid2	q6768	seekid3
403	Once ...	Once ...	Less ...	Once ...
408	Every...	.	.	.
409	Every...	.	.	.
417	Once ...	Every...	Once ...	Every...
419	Less ...	Once ...	Less ...	Once ...
421	.	Once ...	Less ...	Once ...
423	.	Once ...	Less ...	Once ...
427	Every...	Every...	Once ...	Every...
432	.	Once ...	Less ...	Once ...
435	Once ...	Once
439	Once ...	Once
440	.	Ever...	.	.
450	Every...	Once ...	Less ...	Once ...
453	Once
456	Once ...	Once ...	Once ...	Once ...

LABY:

- 1 Every day
- 2 Once a week or more
- 3 Once a month or more
- 4 Several times a year
- 5 Once a year
- 6 Less than once a year
- 9 DK/Ref

1

Numbers
with Labels

Numbers

Strings



Variables

The screenshot displays the SPSS interface with a data view on the left and the Variables dialog box on the right. A central box lists variable types: Numbers, byte, int, long, float, double. The Variables dialog shows 'seekid3' selected with 'Type' set to 'byte' and 'Value label' set to 'LABY'.

	seemom3	seekid2	q6768	seekid3	oldmom
1	Every...	.	.	.	41
2	.	Once ...	Once ...	Once
3	Once	60
4	Once	42
5	Once	52
6	Once	66
7	Once ...	Once ...	Once ...	Once ...	76
8	Every...
9	Once ...	Once ...	Less ...	Once
10	Every...	Every...	Less ...	Every...	.
11	.	Every...	Less ...	Every...	.
12	.	Once ...	Once ...	Once
13	DK/Ref
14	Every...
15	Once
16	Once
17	.	Every...	Less ...	Every...	.
18	.	Every...	Less ...	Every...	.
19	.	Once ...	Less ...	Once

Variables

Filter variables here

<input checked="" type="checkbox"/>	Name	Label
<input checked="" type="checkbox"/>	q4041	Combined Q40 an...
<input checked="" type="checkbox"/>	seemom3	Combined Q38 an...
<input checked="" type="checkbox"/>	seekid2	Combined Q65 an...
<input checked="" type="checkbox"/>	q6768	Combined Q67 an...
<input checked="" type="checkbox"/>	seekid3	Combined Q65 an...
<input checked="" type="checkbox"/>	oldmom	Age of Oldest Livi...

Properties

Variables

Name	seekid3
Label	Combined O65 ar
Type	byte
Format	%8.0g
Value label	LABY
Notes	

Data

Filename	.
Label	
Notes	

Interface & Windows

Close the browse window

Main Window

The screenshot shows the Stata/SE 15.1 main window. The title bar reads "Stata/SE 15.1 - http://dsc.gmu.edu/files/". The menu bar includes File, Edit, Data, Graphics, Statistics, User, Window, and Help. The toolbar contains various icons for file operations and data management.

Review Panel: Contains a "Filter commands here" search box and a list of commands:

#	Command	_rc
1	use http://dsc.gmu.edu/fil...	
2	br	

Main Command Window: Displays the Stata logo (STATA (R) 15.1) and the following text:

Copyright 1985-2017 StataCorp LLC
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)

Special Edition

20-student Stata lab perpetual license:
Serial number: 401506288390
Licensed to: Wendy Mann
George Mason University

Notes:
1. Unicode is supported; see [help unicode_advice](#).
2. Maximum number of variables is set to 5000; see [help set_maxvar](#).

. use http://dsc.gmu.edu/files/pew.dta
. br
.

Variables Panel: Contains a "Filter variables here" search box and a list of variables:

Name	Label
psraid	Unique ID
respaid	Record
weight	
int	
lang	Language of Inter...
area	Area code
state	State from FIPS
cregion	Census region
msa	MSA code

Properties Panel: Contains a "Filter variables here" search box and a table of properties:

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	

Data	
Variables	228
Observations	3,014
Size	1.48M
Memory	64M
Sorted by	

Command Window: Shows the command `codebook` entered in the Command field, which is circled in red.

Bottom Bar: Shows the file path `C:\tools\stata` on the left and the status `CAP NUM OVR` on the right.

Results

Review

#	Command	_rc
1	use http://dsc.gmu.edu/fil...	
2		
3	codebook	

Command

```
.
```

Results

```
tabulation:  Freq.   Numeric   Label
              761     2   been married, div/sep/wid/lwp
              528     3   never married
              10     9   dk

marrkid4                                     (unlabeled)

      type: numeric (byte)
      label: MARRKID4
      range: [1,9]
      unique values: 7
      units: 1
      missing .: 0/3,014

      tabulation:  Freq.   Numeric   Label
                    728     1   Married and kids under 18
                    794     2   Married and kids 18+ only
                    270     3   LWP/Div/Sep/Wid/Never married
                        and kids under 18
                    514     4   LWP/Div/Sep/Wid/Never married
                        and kids 18+ only
                    188     5   Married,No kids
                    504     6   LWP/Div/Sep/Wid/Never married,
                        no kids
                    16     9   DK/DK kids
```

Variables

Name	Label
psraid	Unique ID
respid	Record
weight	
int_date	Interview date
lang	Language of Inter...
area	Area code
state	State from FIPS
cregion	Census region
mca	MSA code

Properties

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Filename	
Label	
Notes	
Variables	228
Observations	3,014
Size	1.48M
Memory	64M
Sorted by	

Review

Single Click =
Put in Command
box

Double Click =
Put in box &
Press Enter

Using the Windows

Stata/SE 15.1 - http://dsc.gmu.edu/files/

File Edit Data Graphics Statistics User Window Help

Review

Filter commands here

#	Command	_rc
1	use http://dsc.gmu.edu/fil...	
2	br	
3	codebook	

unique values: 4 missing .: 0/3,014

tabulation:	Freq.	Numeric	Label
	1,715	1	now married
	761	2	been married, div/sep/wid/lwp
	528	3	never married
	10	9	dk

marrkid4 (unlabeled)

type: numeric (byte)
label: MARRKID4

range: [1,9] units: 1
unique values: 7 missing .: 0/3,014

tabulation:	Freq.	Numeric	Label
	728	1	Married and kids under 18
	794	2	Married and kids 18+ only
	270	3	LWP/Div/Sep/Wid/Never married and kids under 18
	514	4	LWP/Div/Sep/Wid/Never married and kids 18+ only
	188	5	Married, No kids
	504	6	LWP/Div/Sep/Wid/Never married, no kids
	16	9	DK/DK kids

Command

codebook pvote04a

codebook pvote04a

Variables

Name	Label
regicert	REGICERT Are you...
pvote04a	PVOTE04A In last ...

type "vote" to find pvote04a & double click

Properties

Variables

Name	pvote04a
Label	PVOTE04A In last yea
Type	byte
Format	%8.0g
Value label	PVOTE04A
Notes	

Data

Filename	
Label	
Notes	
Variables	228
Observations	3,014
Size	1.48M
Memory	64M
Sorted by	

C:\tools\stata

CAP NUM OVR

Interpreting the Codebook

#1: Black, Red, or Blue?

Storage Type

Label Assigned

#1

```
type: numeric (byte)
label: PVOTE04A
```

```
range: [1,9]
```

```
unique values: 4
```

```
units: 1
```

```
missing .: 0/3,014
```

```
tabulation: Freq.   Numeric  Label
              2,314      1  Voted
              652       2  Did not vote (includes too young
              12        8  Don't remember if voted (VOL.)
              36        9  Refused (VOL.)
```

#3

#2: Category or Measure?

of Unique Values

of Values Labeled

AGE What is your

#3: Missing Values?

True Missing

Researcher

```
type: numeric (byte)
```

```
label: AGE, but 75 nonmissing values are not labeled
```

```
range: [18,99]
```

```
unique values: 77
```

#2

```
units: 1
```

```
missing .: 0/3,014
```

```
examples: 33
```

```
44
```

```
54
```

```
66
```

Other Variable Types

```
. codebook fips usr
```

fips

FIPS

Black:
Numeric,
no label

type: numeric (long)

```
range: [1001,56041]          units: 1
unique values: 1164          missing .: 0/3014

mean: 28406.9
std. dev: 15725.7

percentiles:    10%    25%    50%    75%    90%
                6059   13139  29095  42003  48439
```

usr

Community type -

Red:
String

type: string (str1)

```
unique values: 3          missing "": 0/3014

tabulation:  Freq.  Value
              640   "R"
              1509  "S"
```

Syntax Structure

Syntax

command varlist qualifiers, options

histogram age, normal width(1)

tabulate age **if** (age < 99)

label list AGE

generate yrborn = 2005 - age

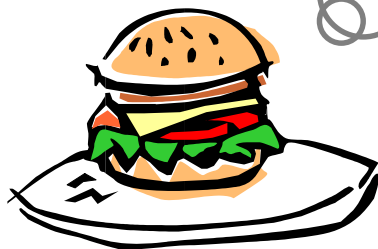
histogram yrborn **if** (age < 99), normal

Structure

order burger if (price < 3), plain mustard

order burger, meat(2) cheese(American)

order burger if (place != "Carls Jr"), nomayo



Abbreviations

histogram age, **normal width(1)**

hist age, **norm w(1)**

tabulate yrborn

tab yrborn

tab yrb

ta y

Wildcards

? = One

* = Multiple

Keeping Track

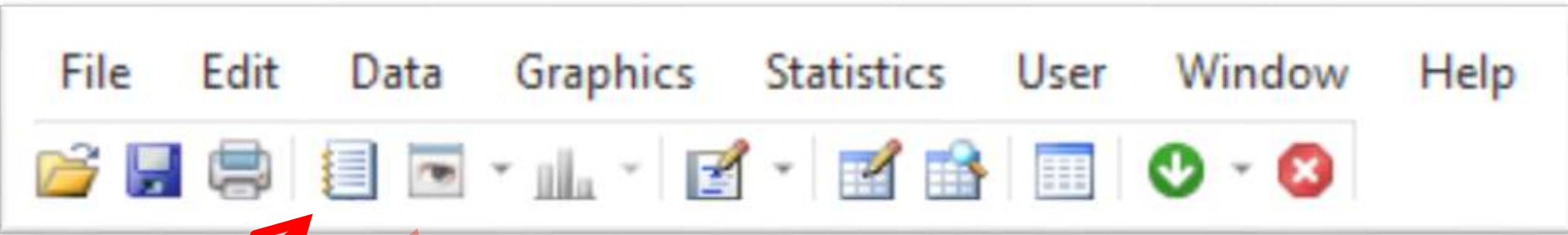
Do Files & Logs

The screenshot displays the Stata software interface with three main windows:

- Do File:** A list of commands including `use`, `describe`, `browse`, `codebook`, `histogram`, `tabulate`, `label list`, `generate`, `histogram`, `tabulate`, `replace`, and `tab`.
- Results:** A table showing the output of the commands, with a large blue text overlay reading "Results ≈ Log File".
- Log File:** A window at the bottom showing the command being executed.
- Variables:** A panel on the right showing the list of variables in the dataset, including `psraid`, `respid`, `weight`, `int_date`, `lang`, `area`, `state`, `cregion`, `msa`, `fips`, `usr`, and `usr1`.
- Properties:** A panel on the right showing the properties of the selected variable `psraid`, including Name, Label, Type, Format, Value Label, and Notes.

Command	70	32	0	0	32
1 use "http://dataservices.gmu..."	71	29	0	0	29
2 describe	72	29	0	0	29
3 browse	73	31	0	0	31
4 codebook pvote04a	74	25	0	0	25
5 codebook pvote04a age	75	39	0	0	39
6 histogram age	76	29	0	0	29
7 tabulate age	77	25	0	0	25
8 label list AGE	78				
9 generate yrborn = 2014 - age	79				
10 histogram age, normal width(1)	80	24	0	0	24
11 histogram age, normal width(1)	81	16	0	0	16
12 histogram yrborn if (age < 9...	82	16	0	0	16
13 tabulate yrborn	83	20	0	0	20
14 tabulate yrborn	84	15	0	0	15
15 generate youth = (age < 30)	85				
16 tab age youth	86	6	0	0	6
17 replace youth = . if (age == ...	87	6	0	0	6
18 tab age youth, missing	88	6	0	0	6
	89	8	0	0	8
	90	4	0	0	4
	92	3	0	0	3
	93	2	0	0	2
	97 or older	1	0	0	1
	Don't know/Refused [V	0	0	51	51
Total		2,516	447	51	3,014

Menu Bar



Logs
*begin/
close*

Viewer
*help &
smcl logs*

Graphs
*use names to
have multiple
windows*

Do Files

**Browse
& Edit**

**Variables
Manager**
*edit labels &
properties*

**More/
Break**

Log File Types

The screenshot shows a Windows File Explorer window titled "Save" with the address bar set to "Windows7_O... > Repository". The left sidebar shows a folder tree with "Repository" selected. The main pane displays two columns of information:

File Type	Description
.log	.smcl
Unformatted Log	Formatted Log
Open in Notepad	Requires Stata

At the bottom, the "Save" dialog is open, showing:

- File name: Surviving_Stata.log
- Save as type: Formatted Log (*.smcl)
- File type list: Formatted Log (*.smcl), Log (*.log), All Files (*.*)

The "Log (*.log)" option is currently selected in the file type list. A mouse cursor is visible at the bottom right of the dialog.

Viewer - view "C:\Repository\Workshops\Surviving_Stata.smcl"

File Edit History Help

view "C:\Repository\Workshops\Surviving_Stata.smcl"

view "C:\Repository\Workshops..."

Dialog | Also See | Jump To

```

name: <unnamed>
log: C:\Repository\Works
log type: smcl
opened on: 3 Oct 2014, 11:46:
. use "http://dataservices.gmu.edu/
. describe
Contains data from http://datas
obs: 3,014
vars: 228
size: 1,555,224

```

Surviving_Stata.log - Notepad

File Edit Format View Help

```

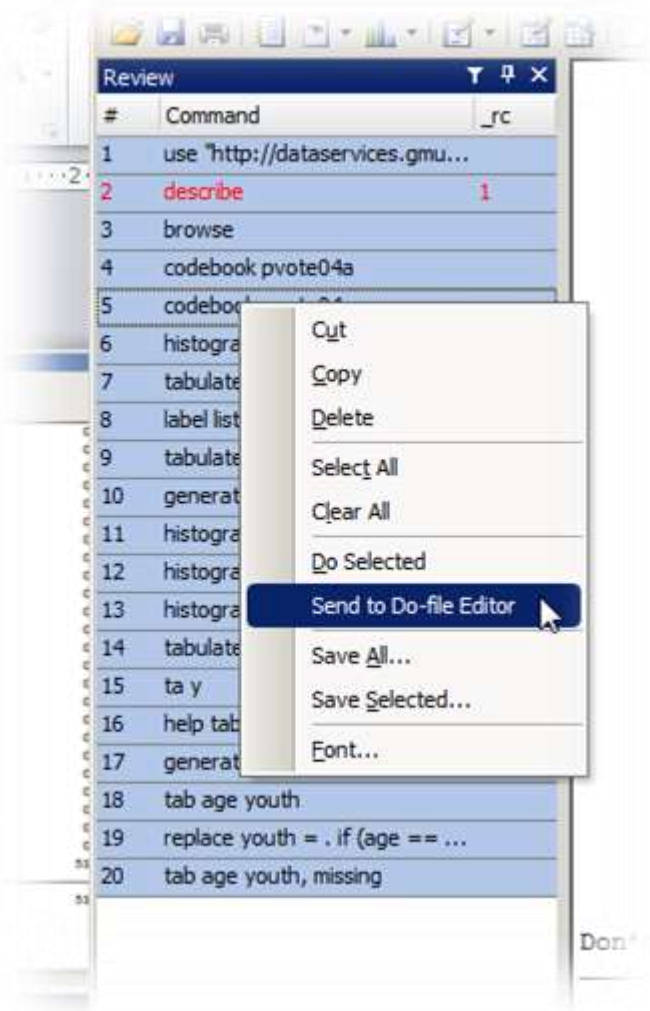
name: <unnamed>
log: C:\Repository\workshops\Surviving_Stata.log
log type: text
opened on: 3 Oct 2014, 11:49:24
. use "http://dataservices.gmu.edu/files,
. describe
Contains data from http://dataservices.gmu.edu/files/
obs: 3,014
vars: 228
size: 1,555,224
24 Sep 2012 16:30

```

.log

variable name	storage type	display format	value label	variable label
psraid	int	%8.0g		Unique ID
respid	int	%8.0g		Record
weight	double	%10.0g		
int_date	long	%12.0g		Interview date
lang	byte	%8.0g	LANG	Language of Interview
area	int	%8.0g		Area code
state	byte	%8.0g	STATE	State from FIPS
cregion	byte	%8.0g	CREGION	Census region
msa	long	%12.0g		MSA code
fips	long	%12.0g		FIPS code
usr	str1	%1s		Community type - alpha
usr1	byte	%8.0g	USR1	Community type - numeric
tz	str1	%1s		Time zone
form	byte	%8.0g	FORM	Sample split
sex	byte	%8.0g	SEX	Respondent's sex
q2	byte	%8.0g	Q2	* Q.2 Generally, how would you say things are these days in your life -- would you
q3a	byte	%8.0g	LABA	* Q.3a Your standard of living -- of what you can buy or do...Please tell me wheth
q3b	byte	%8.0g	LABA	* Q.3b Your household income...Please tell me whether you are satisfied or dissat

Do- Files



The Do-file Editor window shows a menu bar with "File", "Edit", "View", "Project", and "Tools". A red box highlights the "File" menu, and a red arrow points to the "Execute selection (do)" button in the toolbar. The main text area contains the following Stata code:

```
1 log using mylog, replace
2 use "http://dataservices.gmu.edu/files,
3 describe
4 browse
5 codebook pvote04a
6 codebook pvote04a
7 histogram age
8 tabulate age
9 label list AGE
10 tabulate age if (age < 99)
11 generate yrborn = 2014 - age
12 histogram age, normal width(1)
13 histogram age, normal width(10)
14 histogram yrborn if (age < 99), normal
15 tabulate
16 ta y
17 help t
18 genera
19 tab ag
20 replac
21 tab ag
22 log cl
23
```

1. Select syntax
2. Press **Ctrl-D**
(D for Do)

* Make Comments
* with asterisks

Execute selection (do) Line: 12, Col: 1 CAP NUM OVR

```
1 /*****
2 Syntax for Surviving Stata
3 Workshop by Data Services
4 George Mason University
5 DAK 2014-10-01
6 *****/
7
8 * Open data
9 use "http://dataservices.gmu.edu/files"
10
11 * Examine data
12 describe
13 browse
14 codebook pvote04a
15 codebook pvote04a age
16
17 * Examine age
18 histogram age
19 tabulate age
20 label list AGE
21 tabulate age if (age < 99)
22 generate yrborn = 2014 - age
23
24 * Use options
25 histogram age, normal width(1)
26 histogram age, normal width(10)
27 histogram yrborn if (age < 99), normal
28
29 * Try abbreviations
30 tabulate yrborn
31 ta y
32
33 * Get help
34 help tab
35
36 * Create an indicator variable
37 generate youth = (age < 30)
38 tab age youth
39 replace youth = . if (age == 99)
40 tab age youth, missing
41
```

**.do file
in
Do File Editor**

```
Surviving_Stata.do - Notepad
File Edit Format View Help

/*****
Syntax for Surviving Stata
Workshop by Data Services
George Mason University
DAK 2014-10-01
*****/

* Open data
use "http://dataservices.gmu.edu/files"

* Examine data
describe
browse
codebook pvote04a
codebook pvote04a age

* Examine age
histogram age
tabulate age
label list AGE
tabulate age if (age < 99)
generate yrborn = 2014 - age

* Use options
histogram age, normal width(1)
histogram age, normal width(10)
histogram yrborn if (age < 99), normal

* Try abbreviations
tabulate yrborn
ta y

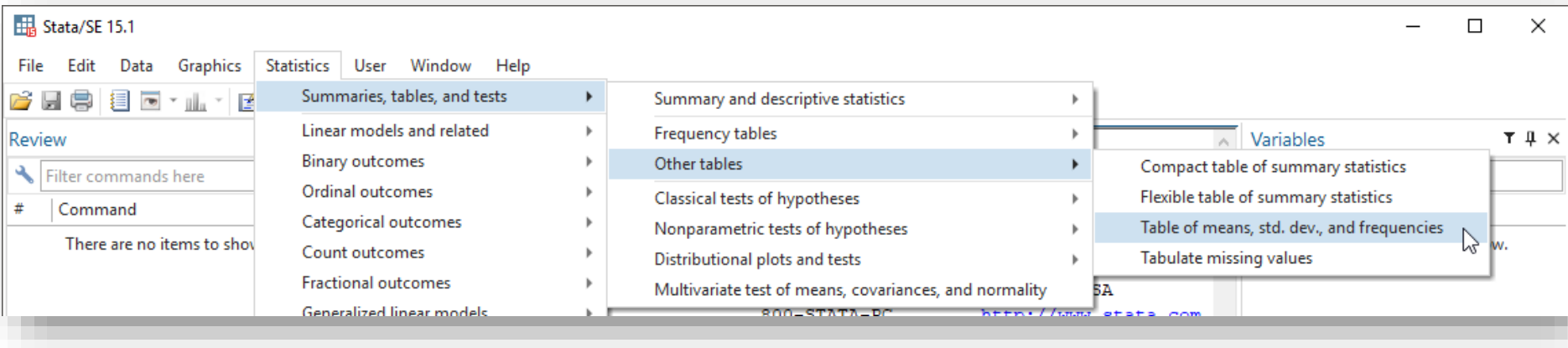
* Get help
help tab

* Create an indicator variable
generate youth = (age < 30)
tab age youth
replace youth = . if (age == 99)
tab age youth, missing
```

**.do file
in
Notepad**

Using help

Why Use Syntax?



Most people do the same few things over and over again.

log	cd	use	save	clear	drop	keep
recode	encode	generate	egen	replace	rename	
tabulate	summarize	ttest	logistic	regress	estimates	

Getting to Help

The image shows a screenshot of the SPSS 'tabsum' dialog box, titled 'tabsum - Table of means, std. dev., and frequencies'. The dialog has three tabs: 'Main', 'by/if/in', and 'Weights'. The 'Main' tab is active. It contains several input fields and checkboxes:

- Variable 1:** An empty dropdown menu.
- Variable 2: (optional):** An empty dropdown menu.
- Summarize variable:** An empty dropdown menu.
- Output:** A section containing two sub-sections:
 - Summary table:** Four checkboxes, all unchecked:
 - Suppress means
 - Suppress standard deviations
 - Suppress frequencies
 - Suppress number of obs.
 - Other:** Two checkboxes, both unchecked:
 - Show numeric codes, not labels
 - Do not break wide tables
- Treat missing values as categories

A tooltip is overlaid on the right side of the dialog, containing the following text:

help tabsum
or
help tab then click
tabulate, summarize()

At the bottom left of the dialog, there is a red arrow pointing to a help icon (a question mark in a circle). The bottom of the dialog features three buttons: 'OK' (highlighted with a blue border), 'Cancel', and 'Submit'.

The Help Viewer

Viewer - help tabulate_summarize

File Edit History Help

help tabulate_summarize

help tabulate_summarize

[R] `tabulate, summarize()` — One- and two-way tables of summary statistics
([View complete PDF manual entry](#))

Syntax

`tabulate varname1 [varname2] [if] [in] [weight] [, options]`

options	Description
Main	
<code>summarize(varname3)</code>	report summary statistics for varname3
<code>[no]means</code>	include or suppress means
<code>[no]standard</code>	include or suppress standard deviations
<code>[no]freq</code>	include or suppress frequencies
	include or suppress number of observations
	show numeric codes, not labels
	do not break wide tables
	treat missing values of <code>varname1</code> and <code>varname2</code> as categories
	allowed; see <code>weight</code> .

Dialog Also see Jump to

Syntax

Menu

Description

Links to PDF documentation

Options

Examples

help tabsum
or
help tab then click
`tabulate, summarize()`

CAP NUM OVR

Syntax

italics only



```
tabulate varname1 [varname2] [if] [in] [weight] [, options]
```

Underline *Blue* [Brackets] *Italics*

must type

click here

replace me

[not needed]

Simplifying Syntax

Title

[D] `generate` — Create or change contents of variable

Syntax

Create new variable

```
generate [type] newvar[:lblname] =exp [if] [in]
```

```
generate newvar =exp
```

Practice with Options

help histogram

More Options

[R] histogram — Histograms for continuous and categorical variables
(View complete PDF manual entry)

Syntax

```
histogram varname [if] [in] [weight] [, [continuous_opts | discrete_opts] options]
```

continuous_opts

Main

bin(#)

width(#)

start(#)

discrete_opts

Main

discrete

width(#)

start(#)

options

options

Description

Main

density

draw as density; the default

fraction

draw as fractions

frequency

draw as frequencies

percent

draw as percentages

bar_options

rendition of bars

binrescale

recalculate bin sizes when

addlabels

add height labels to bars

addlabopts (marker_label_options)

affect rendition of labels

Density plots

normal

add a normal density to the

normopts (line_options)

affect rendition of normal

kdensity

add a kernel density estima

kdensity (kdensity_options)

affect rendition of kernel

Quick Start - NEW!

histogram — Histograms for continuous and categorical variables

[Description](#)
[Options](#)

[Quick start](#)
[Remarks and examples](#)

[Menu](#)
[References](#)

[Syntax](#)
[Also see](#)

Description

`histogram` draws histograms of *varname*, which is assumed to be the name of a continuous variable unless the `discrete` option is specified.

`hist` is a synonym for `histogram`.

Quick start

Histogram of `v1`

```
histogram v1
```

Add a normal density curve to the graph

```
histogram v1, normal
```

Add a kernel density estimate to the graph

```
histogram v1, normal kdensity
```

Add "My Title" as the title of the graph

```
histogram v1, normal kdensity title("My Title")
```

Specify the number of bins as 10

```
histogram v1, bin(10)
```

Specify the width of the bins as 2

```
histogram v1, width(2)
```

Specify that `v2` should be treated as discrete

```
histogram v2, discrete
```

More on Syntax

Great Tutorial

The screenshot shows the SSCC (Social Science Computing Cooperative) website. The header includes the SSCC logo and the tagline "Supporting Statistical Analysis for Research". A navigation menu lists "HOME", "WINSTAT", "HELP DESK", "STAT CONSULTING", and "KNOWLEDGE BASE". A search bar is located in the top right. The main content area is titled "Articles on Statistical Computing" and features a list of articles: Stata Basics, Stata Topics, R Basics, R Topics, SAS Basics, SAS Topics, SPSS, Mplus, Python, and NVivo. A detailed view of the "Stata Basics" article is shown, including a description and a numbered list of 11 topics.

SSCC social science computing cooperative

Supporting Statistical Analysis for Research

• HOME • WINSTAT • HELP DESK • STAT CONSULTING • KNOWLEDGE BASE •

Search

Articles on Statistical Computing

KNOWLEDGE BASE

- TRAINING
- ACCOUNTS
- RESOURCES
- ABOUT SSCC
- SSCC NEWS

Stata Basics

Stata Topics

R Basics

R Topics

SAS Basics

SAS Topics

SPSS

Mplus

Python

NVivo

This page lists articles related to various programs for doing statistical analysis, as well as general information on working with statistical data.

Stata Basics

Stata for Researchers

1. Introduction
2. Usage and Syntax
3. Do Files
4. Working With Data
5. Statistics
6. Working with Groups
7. Hierarchical Data
8. Combining Data Sets
9. Graphics
10. Project Management
11. Learning More

Stata Basic Syntax

command *varlist* *qualifiers*, *options*

command 1, 2, or 3 words that indicate what to do
varlist 0, 1, or more variables to work with
qualifiers add labels or limits to the instructions
options specify additions or changes to the output

Legend

Replace any syntax in *italics* with your own variable names or values.

var any variable ("var")
cat a categorical (nominal) var
num a numeric (interval/scale) var
x the predictor or Independent var
y the response or Dependent var

Describe Values

ANY VARIABLE(S)

`codebook` *vars*

CATEGORICAL

`tabulate` *cat*

`graph` *bar*, *over*(*cat*)
also `graph` *pie*, ...

NUMERIC

`summarize` *num*

`mean` *num*

`histogram` *num*, *freq*
if *integer*, add *discrete*

MORE OPTIONS

Install once before use.

`ssc install` *fre*
`fre` *cat*

`ssc install` *univar*
univar *num*

Show Relationships

TWO NUMERIC

`pwcorr` *num num*

`scatter` *y x* || *lfit y x*

TWO CATEGORICAL

`tabulate` *x y*, *row*

`tabulate` *y x*, *col*

NUMERIC + CATEGORICAL

`tabulate` *cat*, *sum*(*num*)

`table` *cat*, *c*(*mean num*)

`graph` *bar num, over*(*cat*)
also `graph` *box* ...

3-WAY RELATIONSHIPS

`table` *cat cat*, *by*(*cat*)

`table` *cat cat*, *c*(*mean num*)

`scatter` *num num*, *by*(*cat*)

`scatter` *num num* [*w=num*]

Data Management

SIMPLIFY THE FILE

`keep` *var1 var2 etc*

`keep` if *var* != 4

CHANGE VALUES

`generate` *newvar = var*
↑ avoid changing original var

`recode` *var* (2 3 = 1)

`recode` *var* (99 = .)
missing value ↑

`replace` *var = 1* if *var* == 9

`replace` *var = "US"*
if *var* == "VA"
strings must be quoted ↑

Significance Tests

TEST

Chi-Square

Correlation

One-Way ANOVA

SYNTAX

`tabulate` *cat cat*, *chi2*

`pwcorr` *num num*, *sig*

`oneway` *num cat*, *tabulate*

P-VALUE

after `Pr =`

under Pearson's *r*

under Prob > F

OPERATORS

== Equal to < Less Than
!= Not <= Less than
Equal to or Equal to
& AND | OR

Save & Close

Stata Windows

Review commands you ran, left or right click to **reuse**

Begin a **log** to start recording **results** that appear here

The **working directory** of files set with `cd "path"`




Main Window

If **data** is loaded/open, a list of the **variables** will be here

See variable **properties**, **unlock** to change

Type **commands** here, then press Enter

Viewer



Log File .smcl / .log

Keeps a record of what you did and the results. Not reusable.

File → Log → **Begin**
File → Log → **Close**
File → Log → **View**



log using *mylog*
log *close*

Begin a log *before* doing any work; close a log when done.

help *command* or **help** *topic*
Shows help docs in Viewer

Unless you purchased Stata, use the extension *.log* so you can view in any text editor.

Data Editor



Data File .dta

Data in Stata format can include labels and metadata.

File → **Open** & File → **Save As**
Window → **Data Editor** or
Data → **Data Editor**

use *mydata*
save *mydata*

No need to specify *.dta*; add the option *replace* as needed.

clear or add *clear* as an option
Closes the data *without* saving

If your data is not in Stata format, choose File → Import or use **import** (GSS Section 8).


Do File Editor



Do File .do

Save, edit, and re-run syntax to replicate analysis or fix errors.

Window → **Do File Editor**
In the Do File Editor:
File → **Open** & File → **Save As**



doedit
do *mydofile.do*

Highlight syntax then **Execute** (mouse or Ctrl-L) (Ctrl-D)

* *describe* what is happening
Ways to comment *// comment*

A do file can be opened in any text editor. Type in it or copy/paste Review and make changes.

command *varlist* *qualifiers* using *file*, *options*

Review

Review

What are the **4 parts** of Stata Syntax, in order?

command varlist qualifiers, options

What do [brackets] mean in Stata Syntax Help?

not needed / optional element

What does *italics* mean in Stata Syntax Help?

replace with your own word(s) or value(s)

Syntax

command ***varlist*** ***qualifiers***, ***options***

command 1, 2, or 3 words specifying the task

varlist 0, 1, 2 or more variables

= exp A math or logical statement to set a value

if exp A logical statement to limit cases

, options A word or selector to alter the command

Qualifiers

=exp expression (math or logical)
generate yrborn = 2005 - age
generate youth = (age < 30)

[if] logical statement
histogram age if (age != 99)
replace youth = . if (age == 99)

[in] range of case numbers
tabulate pvote04a in 1/100

... names or labels
label yrborn "Year of Birth"
label list AGE

Logical Operators

| or

& and

== is equal to

!= not equal to

< less than

<= less than or equal to

End

For these workshop **slides**:

<https://dsc.gmu.edu/workshops/stata>

For more resources on **Stata**:

<https://infoguides.gmu.edu/software/stata>

For help with **511**:

<https://infoguides.gmu.edu/statsclass>

To provide **feedback** about this workshop:

<https://dsc.gmu.edu/workshops/feedback>

Debby Kermer dkermer@gmu.edu