



Carleton
UNIVERSITY

Canada's Capital University

An Introduction to Stata

Wednesday February 24, 2021

Derek Mikola

Data Services, MacOdrum Library



- **About Stata**
- **Why use Stata? Why use an alternative?**
- **The First Steps**
- **Activities**

What is Stata?

- **Started in mid-1980's as a regression package; extended since!**
- **Written for researchers**
- **Language and Environment originally designed for statistical computing and graphics**
- **Sufficiently flexible**
- **Can be used for multiple tasks beyond just statistical analysis**



- ***“Despite modern user interfaces, the heart of Stata remains the command language. Whatever is done via menus or dialogs is, ideally, echoed as a command. The overwhelming emphasis on a command language follows from a firm belief that statistical analysis cannot be reduced to a small series of standard tasks. In particular, smarter statistical users, especially those near the cutting edge of research in many fields, do not want the statistical equivalent of a burger bar, with choice from a fixed menu, however appealing the individual dishes may seem to some tastes. They do want to go beyond the menu and go inside the kitchen, not to peel the potatoes or fillet the fish, but to order something not on the menu and discuss it with the chef.” (p. 7 – 8)***
- ***Cox, Nick J. 2005. “A brief history of Stata on its 20th anniversary.” *The Stata Journal*, 5 (1) . pp. 2 – 18.***
- **Both positive and negative.**

Why use Stata?

- **Extremely Simple. (Too simple?)**
- **Self-contained & professionally maintained**
- **Stata help manual is extensive**
- **Large online community**
- **Can work with all different types of data**
- **Lots of different *functions* that are used for specific analysis**

Why *not* Stata?

- **Costly**
- **NOT Open Source**
 - | Lags in cutting edge research
 - | Need not have the command you want
- **Issues when communicating with *Mata***
 - | *Mata* is Stata's matrix language.
- **Command names not always intuitive**
- **Only one dataset a time (at least, visible)**
 - | Stata *macros* (variables) aren't visible



Where to get help?

- **Google**
- **Stata Help Manual**
- **<https://www.statalist.org/forums/>**
- **Statistics Consultant in MacOdrum ☺**
- **Stata command sheets**
- **Books!**

- **Data Management**

- | Easy to get data *into* Stata
- | *The data* simplifies many coding issues

- **Statistical Analysis**

- | Basic and Advanced Statistics
- | Stores useful information which can be accessed later
- | Wide range of discipline-specific models
- | Geospatial Analysis

- **Write your own “programs” (functions)**

- **Graphics**

- | Common-place and Fancy (Impulse Response Functions)

Data Editor (Browse) - [census.dta]

File Edit View Data Tools

pop65p[18] 404279

	state	state2	region	pop	pop15	pop5_17	pop18p	pop65p	popurban	medage	death	marriage	divorce							
1	Alabama	AL	South	3,893,888	296,412	865,836	2,731,640	448,015	2,337,713	29.30	35,305	49,018	26,745							
2	Alaska	AK	West	401,851	38,949	91,796	271,106	11,547	258,567	26.10	1,604	5,361	3,517							
3	Arizona	AZ	West	2,718,215	213,883	577,604	1,926,728	307,362	2,278,728	29.20	21,226	30,223	19,988							
4	Arkansas	AR	South	2,286,435	175,592	495,782	1,615,061	312,477	1,179,556	30.60	22,676	26,513	15,882							
5	California	CA	West	23,667,902	1,708,400	4,680,558	17,278,944	2,414,250	21,607,606	29.90	186,428	210,864	133,541							
6	Colorado	CO	West	2,889,964	216,495	592,318	2,081,151	247,325	2,329,869	28.60	18,925	34,917	18,571							
7	Connecticut	CT	NE	3,107,576	185,188	637,731	2,284,657	364,864	2,449,774	32.00	26,095	26,048	13,488							
8	Delaware	DE	South	594,338	41,151	125,444	427,743	59,179	419,819	29.80	5,123	4,437	2,313							
9	Florida	FL	South	9,746,324	570,224	1,789,412	7,386,688	1,687,573	8,212,385	34.70	104,190	108,344	71,579							
10	Georgia	GA	South	5,463,105	414,935	1,231,195	3,816,975	516,731	3,409,081	28.70	44,230	70,638	34,743							
11	Hawaii	HI	West	964,691	77,048	197,735	689,108	76,150	834,592	28.40	4,849	11,856	4,438							
12	Idaho	ID	West	943,935	93,531	213,134	637,270	93,680	509,702	27.60	6,753	13,428	6,596							
13	Illinois	IL	N Cntrl	11,426,518	842,241	2,400,796	8,183,481	1,261,885	9,518,039	29.90	102,230	109,823	50,997							
14	Indiana	IN	N Cntrl	5,490,224	418,764	1,199,554	3,871,906	585,384	3,525,298	29.20	47,300	57,853	40,006							
15	Iowa	IA	N Cntrl	2,913,888	221,628	604,245	2,087,935	387,584	1,708,232	30.00	26,348	27,474	11,854							
16	Kansas	KS	N Cntrl	2,363,679	180,877	468,158	1,714,644	306,263	1,575,899	30.10	21,910	24,847	13,410							
17	Kentucky	KY	South	3,660,777	282,731	799,999	2,570,047	409,828	1,862,183	29.10	33,765	32,727	16,731							
18	Louisiana	LA	South	4,205,900	361,533	968,935	2,875,432	404,279	2,887,309	27.40	35,518	43,460	18,108							
19	Maine	ME	NE	1,124,660	78,514	242,873	803,273	140,918	534,072	30.40	10,768	12,040	6,205							
20	Maryland	MD	South	4,216,975	272,274	895,256	3,049,445	395,609	3,386,555	30.30	34,825	46,278	17,494							
21	Massachusetts	MA	NE	5,737,037	337,215	1,153,174	4,246,648	726,531	4,808,339	31.20	54,919	46,273	17,873							
22	Michigan	MI	N Cntrl	9,262,078	685,113	2,066,873	6,510,092	912,258	6,551,551	28.80	75,102	86,898	45,047							
23	Minnesota	MN	N Cntrl	4,075,970	307,249	864,559	2,904,162	479,564	2,725,202	29.20	33,412	37,641	15,371							
24	Mississippi	MS	South	2,520,638	215,279	598,918	1,706,441	289,357	1,192,805	27.70	23,570	27,908	13,846							
25	Missouri	MO	N Cntrl	4,916,686	354,144	1,008,339	3,554,203	648,126	3,349,508	30.90	49,329	54,625	27,595							
26	Montana	MT	West	786,690	64,455	167,440	554,795	84,559	416,402	29.00	6,664	8,336	4,940							
27	Nebraska	NE	N Cntrl	1,569,825	122,946	324,224	1,122,655	205,684	987,859	29.70	14,465	14,239	6,442							
28	Nevada	NV	West	800,493	56,132	159,667	584,694	65,756	682,947	30.20	5,852	114,333	13,842							
29	New Hampshire	NH	NE	920,610	62,512	195,570	662,528	102,967	480,325	30.10	7,594	9,251	5,254							
30	New Jersey	NJ	NE	7,364,823	463,289	1,527,572	5,373,962	859,771	6,557,377	32.20	68,762	55,794	27,796							
31	New Mexico	NM	West	1,302,894	114,731	303,176	884,987	115,906	939,963	27.40	9,816	16,641	10,426							
32	New York	NY	NE	17,558,072	1,135,925	3,551,938	12,870,209	2,160,767	14,858,068	31.90	171,769	144,518	61,972							
33	N. Carolina	NC	South	5,881,766	404,076	1,253,659	4,224,031	603,181	2,822,852	29.60	40,426	46,718	28,050							
34	N. Dakota	ND	N Cntrl	652,717	54,752	136,239	461,726	80,445	318,310	28.30	5,596	6,094	2,142							
35	Ohio	OH	N Cntrl	10,797,630	787,150	2,307,170	7,703,310	1,169,460	7,918,259	29.90	98,268	99,832	58,809							
36	Oklahoma	OK	South	3,025,290	233,307	621,577	2,170,406	376,126	2,035,082	30.10	28,227	46,509	24,226							
37	Oregon	OR	West	2,633,105	198,046	525,011	1,910,048	303,336	1,788,354	30.20	21,756	23,004	17,762							
38	Pennsylvania	PA	NE	11,863,895	747,458	2,375,838	8,740,599	1,530,933	8,220,851	32.10	123,261	93,673	34,922							
39	Rhode Island	RI	NE	947,154	56,692	186,159	704,303	126,922	824,004	31.80	9,300	7,490	3,606							
40	S. Carolina	SC	South	3,121,820	238,516	703,450	2,179,854	287,328	1,689,253	28.10	25,138	53,915	13,595							
41	S. Dakota	SD	N Cntrl	690,768	58,446	147,160	485,162	91,019	320,777	28.90	6,523	8,000	2,811							
42	Tennessee	TN	South	4,591,120	326,888	972,472	3,292,560	517,588	2,773,573	30.10	40,713	59,175	30,206							
43	Texas	TX	South	14,229,191	1,169,061	3,137,045	9,923,085	1,371,161	11,333,017	28.20	108,019	181,762	96,809							

Variables

Filter variables here

<input checked="" type="checkbox"/>	Name	Label	Type	Format	Value label
<input checked="" type="checkbox"/>	state	State	str14	%-14s	
<input checked="" type="checkbox"/>	state2	Two-letter state abbrevi...	str2	%-2s	
<input checked="" type="checkbox"/>	region	Census region	int	%-8.0g	cenreg
<input checked="" type="checkbox"/>	pop	Population	long	%12.0gc	
<input checked="" type="checkbox"/>	pop15	Pop, < 5 year	long	%12.0gc	
<input checked="" type="checkbox"/>	pop5_17	Pop, 5 to 17 years	long	%12.0gc	
<input checked="" type="checkbox"/>	pop18p	Pop, 18 and older	long	%12.0gc	
<input checked="" type="checkbox"/>	pop65p	Pop, 65 and older	long	%12.0gc	
<input checked="" type="checkbox"/>	popurban	Urban population	long	%12.0gc	
<input checked="" type="checkbox"/>	medage	Median age	float	%9.2f	
<input checked="" type="checkbox"/>	death	Number of deaths	long	%12.0gc	
<input checked="" type="checkbox"/>	marriage	Number of marriages	long	%12.0gc	
<input checked="" type="checkbox"/>	divorce	Number of divorces	long	%12.0gc	

Properties

Variables

Name	pop65p
Label	Pop, 65 and older
Type	long
Format	%12.0gc
Value label	
Notes	

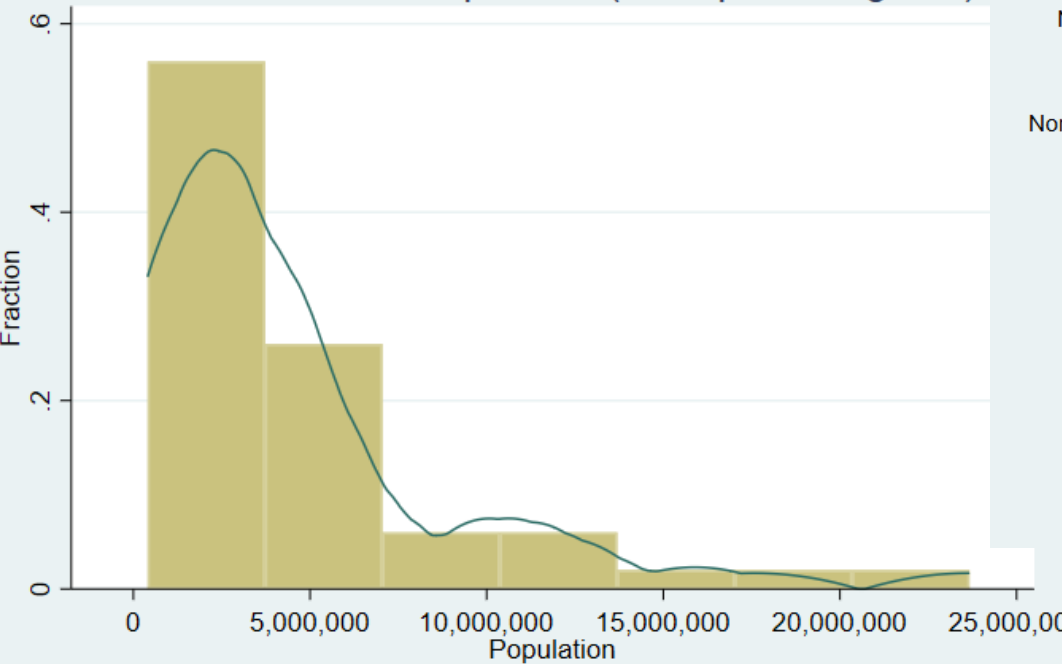
Data

Frame	default
Filename	census.dta
Label	1980 Census data by state
Notes	
Variables	13
Observations	50
Size	2.83K
Memory	64M
Sorted by	

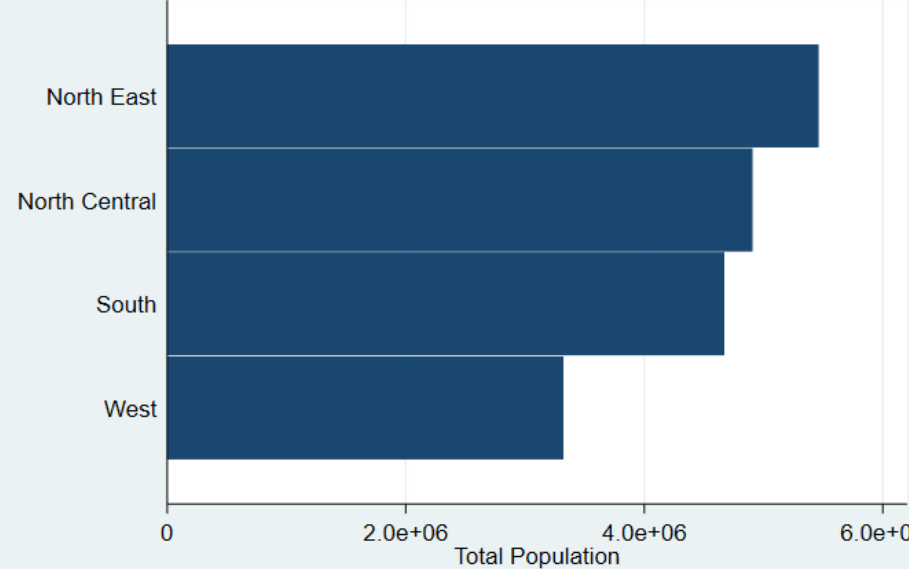


- ***order, sort, reshape, and collapse***, columns (variables) of a dataset (array) easily.
- ***merge, joinby, and append*** new datasets to the one existing in memory.
- ***summarize*** and ***tabulate*** to create summary statistics for continuous and discrete variables, respectively.
- Visualize with ***histogram*** or ***graphs***

U.S. State-Level Population (Example Histogram!)



U.S. Population by Region (Example Bar Graph!)



```
. summarize pop marriage death
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pop	50	4518149	4715038	401851	2.37e+07
marriage	50	47701.4	45130.42	4437	210864
death	50	39474.26	41742.35	1604	186428

```
. tabulate region
```

Census region	Freq.	Percent	Cum.
NE	9	18.00	18.00
N Cntrl	12	24.00	42.00
South	16	32.00	74.00
West	13	26.00	100.00
Total	50	100.00	

And Pretty Tables!

OLS Regression Model

	Monthly~s	Monthly~s	Monthly~s	Monthly~s
getsTreatment	43.60*** (1.484)	8.511*** (0.751)	7.762*** (0.764)	7.499** (0.755)
Observations	5040	5040	5040	5040
Neighbourhood FE		X	X	X
Year FE			X	X
Month FE				X

Standard errors in parentheses

Standard errors are clustered at the neighbourhood level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

- ***recode*** and ***rename*** categories of discrete variables in one fell swoop
- ***label*** variables to provide description of variables. Updates graphics in Stata.
- ***generate*** and ***egenenerate*** new variables from old.



**SO MUCH
MORE**

- **Installation**
- **Interface**
- **Getting Data into Stata**
- **Variables**
- **Do-files**
- **Basic Commands**
- **Conditional Statements**
- **Loops**

- <https://carleton.ca/its/all-services/computers/site-licensed-software/stata/>

You must login with your MCI credentials to gain access to this website. This media is not available at **Hardware Services**. If you prefer to have the software pushed via SCCM to your computer or you would like to access it through VDI (Teaching/Learning) or through RCDC (Research), then please email the **ITS Service Desk** with computer name (pcab123-xx) to request the service you prefer.

Name:	STATA SE 15 & 16
Description:	Stata is a powerful statistical software that enables users to analyze, manage, and produce graphical visualizations of data. It is primarily used by researchers in the fields of economics, biomedicine, and political science to examine data patterns.
OS Supported:	Win, Mac, Linux, Solaris
Available to:	Faculty, Staff and Students on/off campus use
Cost:	Free
How to Get:	To download STATA 15, please go to http://carleton.ca/its/stata/ To download STATA 16, please go to https://carleton.ca/its/stata/download-stata-16/ .





Stata/SE 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command	_rc
	There are no items to show.	

Statistics/Data Analysis 16.0

Copyright 1985-2019 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)

Unlimited-user Stata network license expires 12 Feb 2021:
 Serial number: 401609205665
 Licensed to: Derek Mikola
 Carleton University

Notes:

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

Variables

Filter variables here

Name	Label
There are no items to show.	

Properties

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

Command

Stata/SE 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command	_rc

There are no items to show.

```

----- (R)
STATA 16.0 Copyright 1985-2019 StataCorp LLC
Statistics/Data Analysis StataCorp
Special Edition 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)

Unlimited-user Stata network license expires 12 Feb 2021:
Serial number: 401609205665
Licensed to: Derek Mikola
Carleton University

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

```

Command

Variables

Filter variables here

Name	Label

There are no items to show.

Properties

Variables

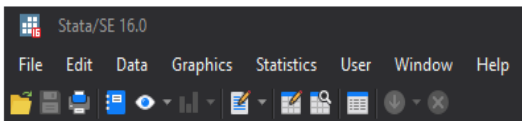
Name	Label	Type	Format	Value label	Notes

Data

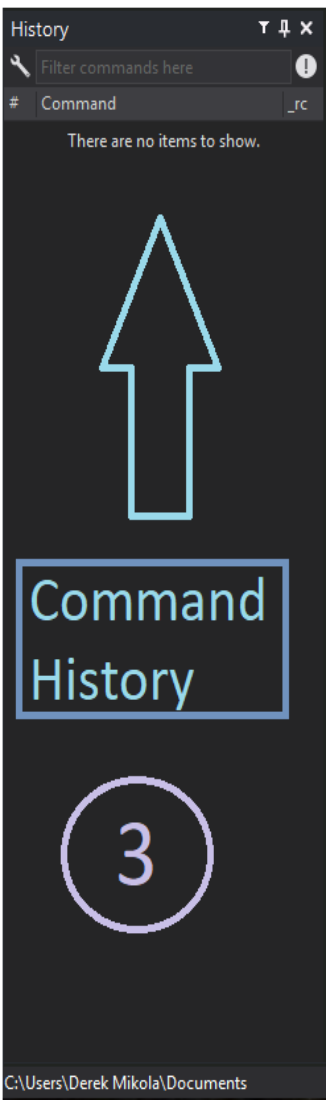
Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

C:\Users\Derek Mikola\Documents

CAP NUM OVR



Toolbar



History

Filter commands here

Command _rc

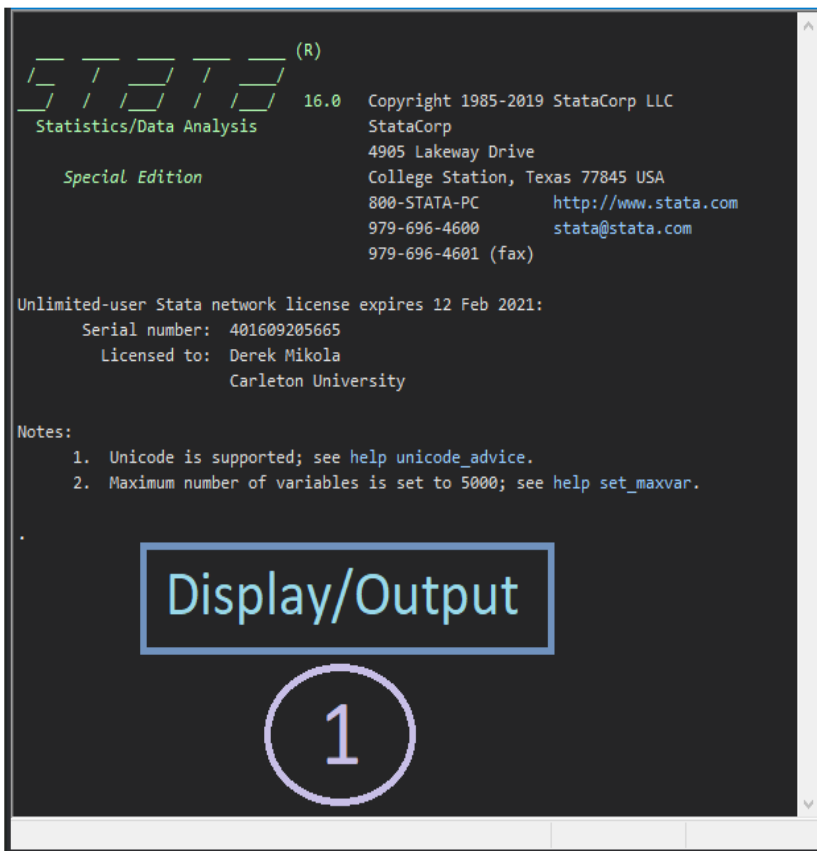
There are no items to show.



Command History



C:\Users\Derek Mikola\Documents



(R)

Statistics/Data Analysis 16.0 Copyright 1985-2019 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STAT-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

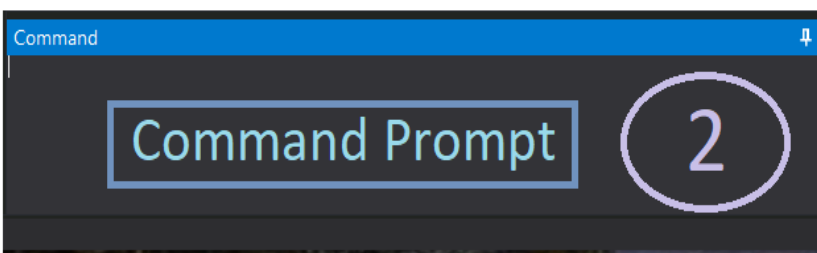

Special Edition

Unlimited-user Stata network license expires 12 Feb 2021:
Serial number: 401609205665
Licensed to: Derek Mikola
Carleton University

Notes:

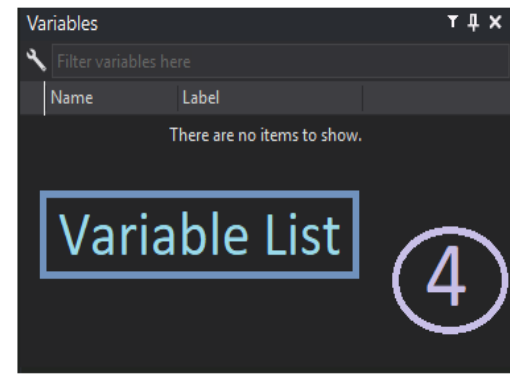

1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set_maxvar.

Display/Output



Command

Command Prompt

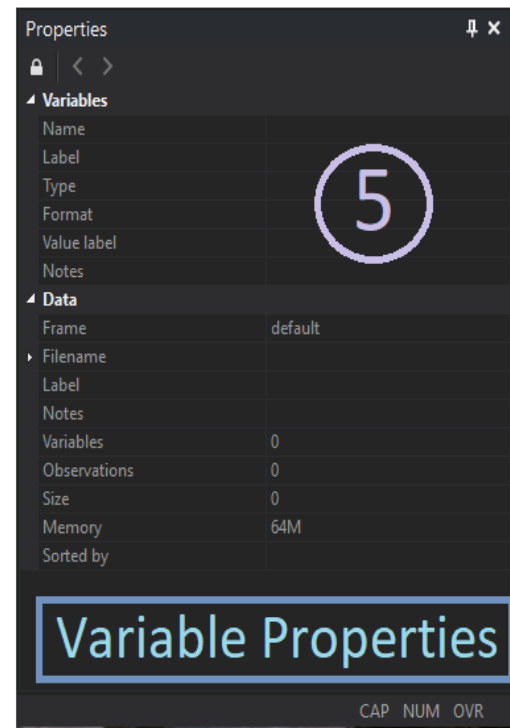



Variables

Filter variables here

Name	Label
There are no items to show.	

Variable List



Properties

Variables


Name	Label
Type	
Format	
Value label	
Notes	

Data

Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

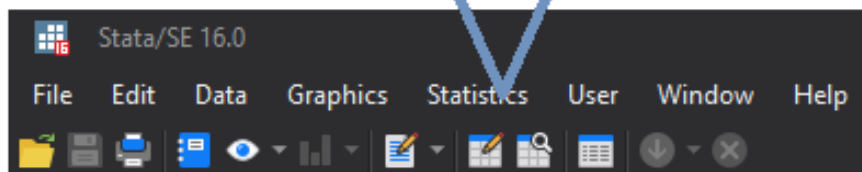
Variable Properties

CAP NUM OVR



- 1. Display/Output:** after running a command, output or error will display.
- 2. Command Prompt:** where you can manually input commands.
- 3. Command History:** what commands have been run since opening *this* interface.
- 4. Variable List:** all variables currently in Stata's memory.
- 5. Variable Properties:** after selecting a variable, what properties it has.
- 6. Toolbar**

1. Data Editor/Viewer



2. Open Do-File Editor

1. Data Editor/Viewer

Data Editor (Browse) - [census.dta]

File Edit View Data Tools

pop65p[18] 404279

	state	state2	region	pop	pop15	pop5_17	pop18p	pop65p	popurban	medage	death	marriage	divorce						
1	Alabama	AL	South	3,893,888	296,412	865,836	2,731,640	448,015	2,337,713	29.30	35,305	49,018	26,745						
2	Alaska	AK	West	401,851	38,949	91,796	271,106	11,547	258,567	26.10	1,604	5,361	3,517						
3	Arizona	AZ	West	2,718,215	213,883	577,604	1,926,728	307,362	2,278,728	29.20	21,226	30,223	19,988						
4	Arkansas	AR	South	2,286,435	175,592	495,782	1,615,061	312,477	1,179,556	30.60	22,676	26,513	15,882						
5	California	CA	West	23,667,902	1,708,400	4,680,558	17,278,944	2,414,250	21,607,606	29.90	186,428	210,864	133,541						
6	Colorado	CO	West	2,889,964	216,495	592,318	2,081,151	247,325	2,329,869	28.60	18,925	34,917	18,571						
7	Connecticut	CT	NE	3,107,576	185,188	637,731	2,284,657	364,864	2,449,774	32.00	26,095	26,048	13,488						
8	Delaware	DE	South	594,338	41,151	125,444	427,743	59,179	419,819	29.80	5,123	4,437	2,313						
9	Florida	FL	South	9,746,324	570,224	1,789,412	7,386,688	1,687,573	8,212,385	34.70	104,190	108,344	71,579						
10	Georgia	GA	South	5,463,105	414,935	1,231,195	3,816,975	516,731	3,409,081	28.70	44,230	70,638	34,743						
11	Hawaii	HI	West	964,691	77,048	197,735	689,108	76,150	834,592	28.40	4,849	11,856	4,438						
12	Idaho	ID	West	943,935	93,531	213,134	637,270	93,680	509,702	27.60	6,753	13,428	6,596						
13	Illinois	IL	N Cntrl	11,426,518	842,241	2,400,796	8,183,481	1,261,885	9,518,039	29.90	102,230	109,823	50,997						
14	Indiana	IN	N Cntrl	5,490,224	418,764	1,199,554	3,871,906	585,384	3,525,298	29.20	47,300	57,853	40,006						
15	Iowa	IA	N Cntrl	2,913,888	221,628	604,245	2,087,935	387,584	1,708,232	30.00	26,348	27,474	11,854						
16	Kansas	KS	N Cntrl	2,363,679	180,877	468,158	1,714,644	306,263	1,575,899	30.10	21,910	24,847	13,410						
17	Kentucky	KY	South	3,660,777	282,731	799,999	2,570,047	409,828	1,862,183	29.10	33,765	32,727	16,731						
18	Louisiana	LA	South	4,205,900	361,533	968,935	2,875,432	404,279	2,087,309	27.40	35,518	43,460	18,108						
19	Maine	ME	NE	1,124,660	78,514	242,873	803,273	140,918	534,072	30.40	10,768	12,040	6,205						
20	Maryland	MD	South	4,216,975	272,274	895,256	3,049,445	395,609	3,386,555	30.30	34,825	46,278	17,494						
21	Massachusetts	MA	NE	5,737,037	337,215	1,153,174	4,246,648	726,531	4,808,339	31.20	54,919	46,273	17,873						
22	Michigan	MI	N Cntrl	9,262,078	685,113	2,066,873	6,510,092	912,258	6,551,551	28.80	75,102	86,898	45,047						
23	Minnesota	MN	N Cntrl	4,075,970	307,249	864,559	2,904,162	479,564	2,725,202	29.20	33,412	37,641	15,371						
24	Mississippi	MS	South	2,520,638	215,279	598,918	1,706,441	289,357	1,192,805	27.70	23,570	27,908	13,846						
25	Missouri	MO	N Cntrl	4,916,686	354,144	1,008,339	3,554,203	648,126	3,349,508	30.90	49,329	54,625	27,595						
26	Montana	MT	West	786,690	64,455	167,440	554,795	84,559	416,402	29.00	6,664	8,336	4,940						
27	Nebraska	NE	N Cntrl	1,569,825	122,946	324,224	1,122,655	205,684	987,859	29.70	14,465	14,239	6,442						
28	Nevada	NV	West	800,493	56,132	159,667	584,694	65,756	682,947	30.20	5,852	114,333	13,842						
29	New Hampshire	NH	NE	920,610	62,512	195,570	662,528	102,967	480,325	30.10	7,594	9,251	5,254						
30	New Jersey	NJ	NE	7,364,823	463,289	1,527,572	5,373,962	859,771	6,557,377	32.20	68,762	55,794	27,796						
31	New Mexico	NM	West	1,302,894	114,731	303,176	884,987	115,906	939,963	27.40	9,816	16,641	10,426						
32	New York	NY	NE	17,558,072	1,135,925	3,551,938	12,870,209	2,160,767	14,858,068	31.90	171,769	144,518	61,972						
33	N. Carolina	NC	South	5,881,766	404,076	1,253,659	4,224,031	603,181	2,822,852	29.60	40,426	46,718	28,050						
34	N. Dakota	ND	N Cntrl	652,717	54,752	136,239	461,726	80,445	318,310	28.30	5,596	6,094	2,142						
35	Ohio	OH	N Cntrl	10,797,630	787,150	2,307,170	7,703,310	1,169,460	7,918,259	29.90	98,268	99,832	58,809						
36	Oklahoma	OK	South	3,025,290	233,307	621,577	2,170,406	376,126	2,035,082	30.10	28,227	46,509	24,226						
37	Oregon	OR	West	2,633,105	198,046	525,011	1,910,048	303,336	1,788,354	30.20	21,756	23,004	17,762						
38	Pennsylvania	PA	NE	11,863,895	747,458	2,375,838	8,740,599	1,530,933	8,220,851	32.10	123,261	93,673	34,922						
39	Rhode Island	RI	NE	947,154	56,692	186,159	704,303	126,922	824,004	31.80	9,300	7,490	3,606						
40	S. Carolina	SC	South	3,121,820	238,516	703,450	2,179,854	287,328	1,689,253	28.10	25,138	53,915	13,595						
41	S. Dakota	SD	N Cntrl	690,768	58,446	147,160	485,162	91,019	320,777	28.90	6,523	8,000	2,811						
42	Tennessee	TN	South	4,591,120	326,888	972,472	3,292,560	517,588	2,773,573	30.10	40,713	59,175	30,206						
43	Texas	TX	South	14,229,191	1,169,061	3,137,045	9,923,085	1,371,161	11,333,017	28.20	108,019	181,762	96,809						

Variables

Filter variables here

<input checked="" type="checkbox"/>	Name	Label	Type	Format	Value label
<input checked="" type="checkbox"/>	state	State	str14	%-14s	
<input checked="" type="checkbox"/>	state2	Two-letter state abbrevi...	str2	%-2s	
<input checked="" type="checkbox"/>	region	Census region	int	%-8.0g	cenreg
<input checked="" type="checkbox"/>	pop	Population	long	%12.0gc	
<input checked="" type="checkbox"/>	pop15	Pop, < 5 year	long	%12.0gc	
<input checked="" type="checkbox"/>	pop5_17	Pop, 5 to 17 years	long	%12.0gc	
<input checked="" type="checkbox"/>	pop18p	Pop, 18 and older	long	%12.0gc	
<input checked="" type="checkbox"/>	pop65p	Pop, 65 and older	long	%12.0gc	
<input checked="" type="checkbox"/>	popurban	Urban population	long	%12.0gc	
<input checked="" type="checkbox"/>	medage	Median age	float	%9.2f	
<input checked="" type="checkbox"/>	death	Number of deaths	long	%12.0gc	
<input checked="" type="checkbox"/>	marriage	Number of marriages	long	%12.0gc	
<input checked="" type="checkbox"/>	divorce	Number of divorces	long	%12.0gc	

Properties

Variables

Name	pop65p
Label	Pop, 65 and older
Type	long
Format	%12.0gc
Value label	
Notes	

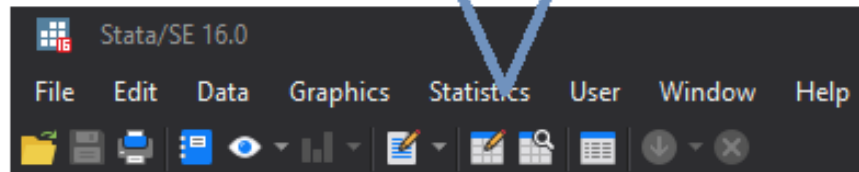
Data

Frame	default
Filename	census.dta
Label	1980 Census data by state
Notes	
Variables	13
Observations	50
Size	2.83K
Memory	64M
Sorted by	

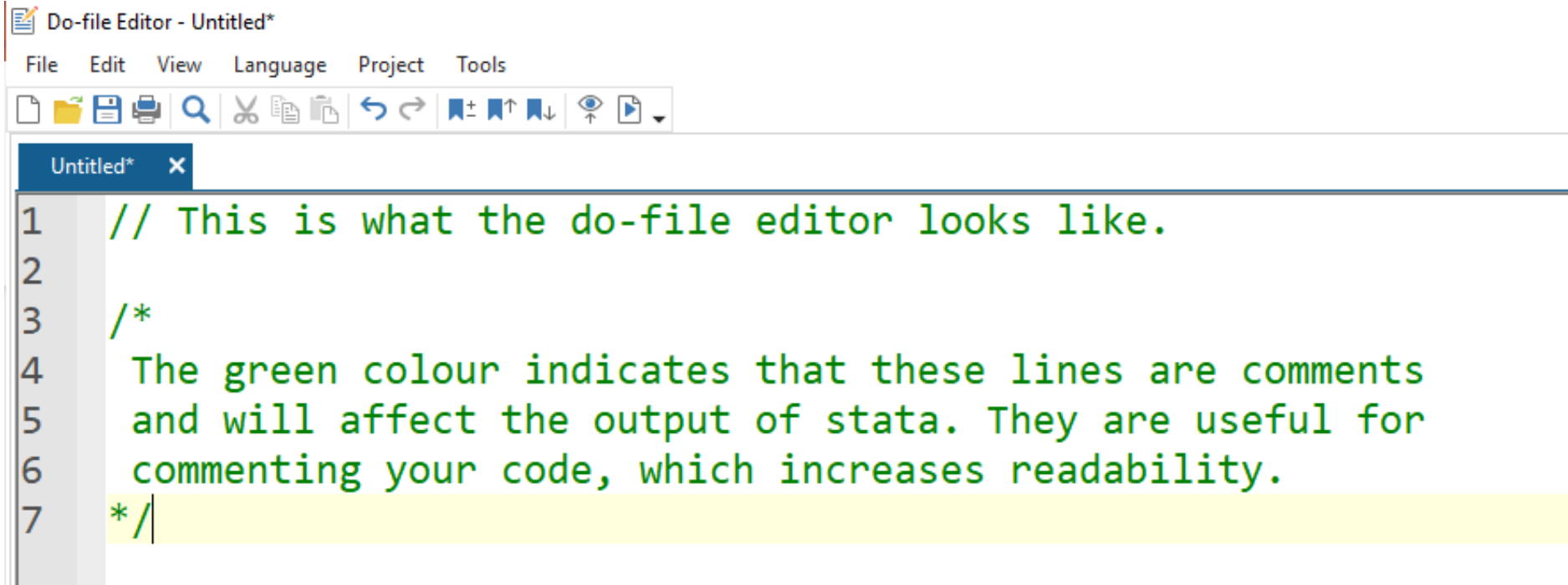
22



1. Data Editor/Viewer



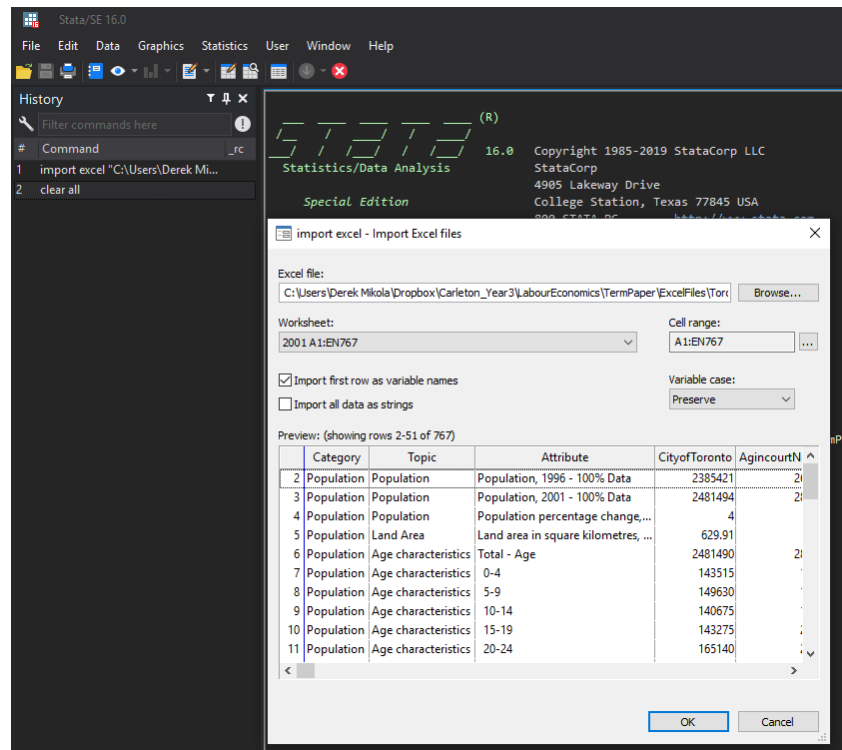
2. Open Do-File Editor

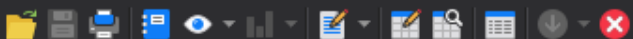


```
1 // This is what the do-file editor looks like.  
2  
3 /*  
4 The green colour indicates that these lines are comments  
5 and will affect the output of stata. They are useful for  
6 commenting your code, which increases readability.  
7 */
```


- **Files > Import > *select.typ* (ex. *gdp.xlsx*)**

This will open a pop-up window and allow you to browse and select the appropriate data.





History



```

Filter commands here
# Command _rc
1 import excel "C:\Users\Derek Mi...
2 clear all

```

(R) 16.0 Copyright 1985-2019 StataCorp LLC
 StataCorp
 4905 Lakeway Drive
 College Station, Texas 77845 USA
 Special Edition

import excel - Import Excel files

Excel file: C:\Users\Derek Mikola\Dropbox\Carleton_Year3\LabourEconomics\TermPaper\ExcelFiles\Tor... Browse...

Worksheet: 2001 A1:EN767 Cell range: A1:EN767 ...

Import first row as variable names Variable case: Preserve

Import all data as strings

Preview: (showing rows 2-51 of 767)

	Category	Topic	Attribute	CityofToronto	AgincourtN
2	Population	Population	Population, 1996 - 100% Data	2385421	21
3	Population	Population	Population, 2001 - 100% Data	2481494	21
4	Population	Population	Population percentage change, ...	4	
5	Population	Land Area	Land area in square kilometres, ...	629.91	
6	Population	Age characteristics	Total - Age	2481490	21
7	Population	Age characteristics	0-4	143515	
8	Population	Age characteristics	5-9	149630	
9	Population	Age characteristics	10-14	140675	
10	Population	Age characteristics	15-19	143275	
11	Population	Age characteristics	20-24	165140	

OK Cancel

TermPaper

■ Variables

- | *Columns* of the dataset within Stata
- | Have *names* and (potentially) *labels*
- | *Rows* of the dataset (should be) the unit of observation (*the person, the year, the firm*)

■ Macros

- | Loosely, variables stored in memory but not displayed
- | *locals* store values temporarily; *globals* store values forever.

■ Vectors/Matrices

- | Arrays, often returned in as *stored values* immediately following the execution of commands

- **Can be executed through the *command prompt* or a *do-file*.**
- **As a principle, always use do-files; never command prompt.**
 - | Command prompt commands will disappear every time you close the window
 - | Do-files are saved forever and can execute a series of commands.

- Command *variable1 ... [variablek] [if] [in] [weight] [,options]*

- generate gdp = consumption + savings
 - | makes a variable called `gdp' from variables in storage
 - | this is a *row-wise summation*

- generate cobbDougGDP = (capital^(1/3))*(labour^(2/3))
 - | transforms `capital' and `labour' into a new variable (column) called `cobbDougGDP'

- **Command *variable1 ... [variablek] [if] [in] [weight] [,options]***

- **summarize gdp consumption**
 - | gives average and std. dev. of the variables gdp and consumption
- **summarize gdp consumption if year > 1985**
 - | As above, but, only for the period after 1985

- **regress gdp consumption**
 - | ols regression of gdp on consumption
- **regress gdp consumption if year > 1985 & year < 2007, robust**
 - | As above, but, only using data between 1985 and 2007 and using White's standard errors.

For help with a known command:

- **help *commandName*** in the command prompt
- **Google!** “how do I use *commandName*?” which will usually direct you to the appropriate page.

For help with an unknown command:

- **Google!**

- | *“How do I refuzzulate the carbonator on rocketship 27 to land on Juniper and get an A in econ?”*
- | *“Want to order variables by country in Stata”*
- | *“How do I fill-in missing times in Stata?”*
- | *“How do I declare my dataset to be panel data in Stata?”*
- | *“How do I run a logistic regression in Stata?”*



- **There are a few ways to grab a specific element or column/row of a data type**
- **We use square brackets [] to grab a specific row of a given variable**
- **gen 1980 = year[10]**

If Statement is TRUE than do this

We use curly brackets {} to open and close our “if statement”

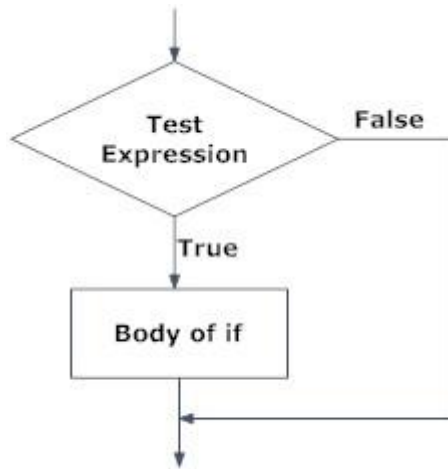


Fig: Operation of if statement.

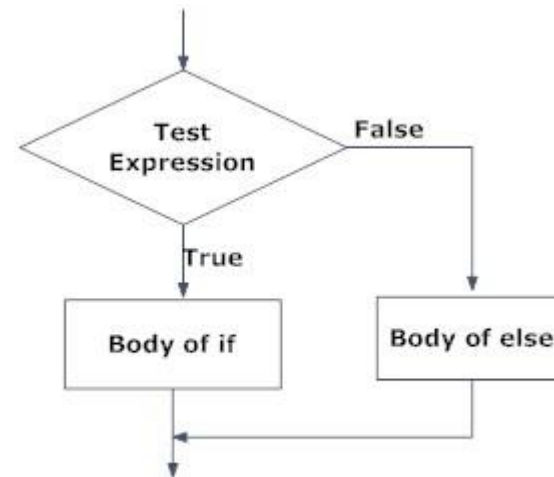


Fig: Operation of if...else statement



```
if Conditional Statement {  
    thing to happen if Condition is TRUE  
}
```

```
if Conditional Statement{  
    thing to happen if Condition is TRUE  
} else {  
    thing to happen if Condition is FALSE  
}
```



- **Loops are used to repeat a specific task over a block of code**
- **For Loops**
 - | Runs for a specific period of time (10 times, 20 times, 1000 times)
- **While Loops**
 - | Runs until a specific condition is met (run until object is greater than 10, run until you encounter a specific object)



Looks similar to our if statements

```
forval i = 1/10{  
    display(`I`)  
}
```

i is an object

1/10 is a range of numbers

print(`I`) will be done for each value of **i**



Thank you!

For more information ...

<https://library.carleton.ca/services/statistical-consulting>

dataservices@carleton.ca