

GET DATA

Into Stata

Stata YouTube Channel



Project 1

- Tour of Stata Interface
- Copy/paste data from Excel into Stata
- Import Excel data into Stata
- Importing delimited data in Stata



Cheat Sheets

[Link](#)

Data Transformation with Stata 14.1 Cheat Sheet

For more info see Stata's reference manual (stata.com)

Select Parts of Data (Subsetting)

Select Specific Columns

drop make
remove the 'make' variable

keep make price
opposite of drop, keep only columns 'make' and 'price'

Filter Specific Rows

drop if mpg < 20
drop observations based on a condition (left) or rows 1-4 (right)

drop in 1/4

keep in 1/30
opposite of drop, keep only rows 1-30

keep if inrange(price, 5000, 10000)
keep values of price between \$5,000 - \$10,000 (inclusive)

keep if inlist(make, "Honda Accord", "Honda Civic", "Subaru")
keep the specified values of make

sample 25
sample 25% of the observations in the dataset (use **set seed** # command for reproducible sampling)

Replace Parts of Data

CHANGE COLUMN NAMES

rename (mp28 foreign) (mpairRecord carType)
rename one or multiple variables

CHANGE ROW VALUES

replace price = 5000 if price < 5000
replace all values of price that are less than \$5,000 with 5000

recode price (0 / 5000 = 5000)
change all prices less than 5000 to be \$5,000

recode foreign (0 = 2 "US") (1 = 1 "Not US"), gen(foreign2)
change the values and value labels then store in a new variable, foreign2

REPLACE MISSING VALUES

mvdecode _all, mv(9999)
useful for cleaning survey datasets
replace the number 9999 with missing value in all variables

mvencode _all, mv(9999)
useful for exporting data
replace missing values with the number 9999 for all variables

Label Data

Value labels map string descriptions to numbers. They allow the underlying data to be numeric (making logical tests simpler) while also connecting the values to human-understandable text.

label define mylabel 0 "US" 1 "Not US"
label values foreign mylabel
define a label and apply it to the values in foreign

label list
list all labels within the dataset

Ten Drom (drom@stata.com) • Laura Hughes (laura@stata.com)
Stata.com and Stata are not affiliated with Stata. But we like it.

Reshape Data

webuse intex <http://github.com/ten-drom/stata-helpers/blob/master/try3/Data/webuse/coffeeMaze.do> **load demo dataset**

Next Data (Wide → Long)

reshape long coffeeMaze, i(country) j(year) **new variable**
convert a wide dataset to long

WIDE **LONG** **CAST**

CAST Data (Long → Wide)

reshape wide coffeeMaze, i(country) j(year)
convert a long dataset to wide

xpov, clear varname
transpose rows and columns of data, clearing the data and saving old column names as a new variable called "varname"

Combine Data

Adding (Appending) New Data

webuse coffeeMaze2 **clear** **save coffeeMaze2.dta, replace** **webuse coffeeMaze** **clear** **append using** "coffeeMaze2.dta", gen(_tenum)
add observations from "coffeeMaze2.dta" to current data and create variable "_tenum" to track the origin of each observation

Merging Two Datasets Together

One-to-One

merge 1:1 **using** "int_age.dta", **save int_age.dta, replace** **webuse int_age** **clear** **merge 1:1** **using** "int_age.dta" **into** the loaded dataset and create variable "_merge" to track the origin

Many-to-One

merge m:1 **using** "int2.dta", **save int2.dta, replace** **webuse int2** **clear** **merge m:1** **using** "int2.dta" **into** the loaded dataset and create variable "_merge" to track the origin

Fuzzy Matching: Combining Two Datasets Without a Common ID

findobs **match** records from different data sets using probabilistic matching to find observations

findsim **create** distance measure for similarity between two strings **use** **findsim** **probabilities**

inspired by RStudio awesome Cheat Sheets: <https://github.com/rstudio/cheatsheets>

Manipulate Strings

Get String Properties

display length("this string has 29 characters")
return the length of the string

charlist make
display the set of unique characters within a string

display strpos("Stata", "a")
return the position in Stata where a is first found

Find Matching Strings

display strpos("Stata", 3, 5)
return the string located between characters 3-5

list make if regexm(make, "D-8")
list observations where make matches the regular expression (here, records that contain a number)

list if regexm(make, "^(Cad|Chev|Datsun)")
return all observations where make contains "Cad", "Chev" or "Datsun"

list if inlist(word1, "Cad", "Chev", "Datsun")
return all observations where the first word of the make variable contains the listed words

TRANSFORM STRINGS

display regex("My string", "My", "four")
replace string1 ("My") with string2 ("four")

replace make = subinstr(make, "Cad", "Cadillac", 1)
replace first occurrence of "Cad" with Cadillac in the make variable

display strtrim(" too much " "Space")
replace consecutive spaces with a single space

display trim(" leading / trailing spaces ")
remove extra spaces before and after a string

display strlower("STATA should not be ALL-CAPS")
change string case, see also **strupper**, **strproper**

display strtoname("LVar name")
convert string to Stata-compatible variable name

display real("100")
convert string to a numeric or missing value

Save & Export Data

save "myData.dta", replace
saveold "myData.dta", replace version(12)
save data in Stata format, replacing the data if a file with same name exists

export excel "myData.xls", if=
firstrow(variables) **replace**
export data as an Excel file (XLS) with the variable names in the first row

export delimited "myData.csv", delimiter(",") **replace**
export data as a comma-delimited file (CSV)

github.com/ten-drom/stata-helpers updated March 2016

Data Analysis with Stata 14.1 Cheat Sheet

Summarize Data

display **sum** **varname**
display summary statistics for varname

display **sum** **varname** **if** **condition**
display summary statistics for varname if condition is true

display **sum** **varname** **by** **groupvar**
display summary statistics for varname by groupvar

display **sum** **varname** **if** **condition** **by** **groupvar**
display summary statistics for varname if condition is true by groupvar

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname if condition is true by groupvar, collapse varname

display **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname** **display** **sum** **varname** **if** **condition** **by** **groupvar** **collapse** **varname**
display summary statistics for varname if condition is true by groupvar, collapse varname, display summary statistics for varname

Lab 2 & Project 1

Intro **Stata**

Learning Goals

- Types of Variables & Control Structures
- Import data
- Save dataset
- Create new variables
- Use basic operators

The 5 Basic Concepts

Found in all Programming Languages, [Link](#).

1. Variables
2. Control Structures
3. Data Structures
4. Syntax
5. Tools

Variables

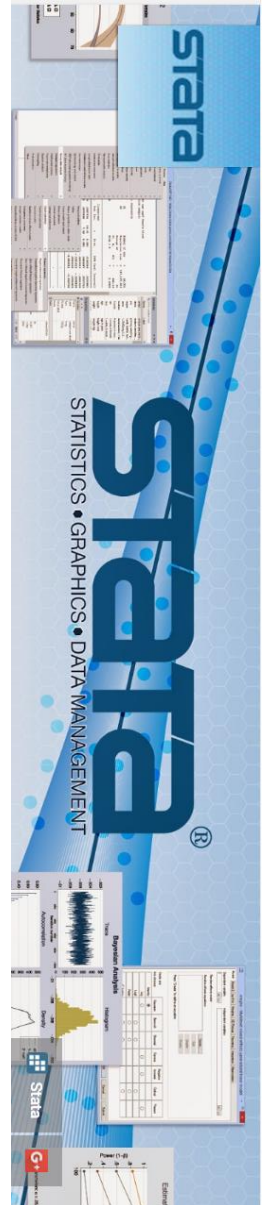
"Any way to store & retrieve information at will"

Stata Datatypes (small -> large), see [Link](#)

- Numeric: byte, int, long, float, double
- String: str1, str2, ... , str2045, strL

Does Var Size Matter?

- Increasing dataset size may slow **Stata** down
- Converting types incorrectly will truncate data



Control Structures

Code Flow: Left-to-Right, Top-to-Bottom

Decisions: Repeat, Skip, or Jump To
Will often be conditional on Variable values

In **Stata**...

- If { ... } Else { ... }
- Foreach x in '*varlist*' { ... }
- While *exp* { ... }
- Function(...)

Statements,
Loops,
Loops,
Calls,

[Link](#)

[Link](#)

[Link](#)

[Link](#)



The 5 Basic Concepts

Found in all Programming Languages, [Link](#).

1. Variables
2. Control Structures
3. Data Structures
4. **Syntax**
5. Tools

help summarize

Title

[R] `summarize` — Summary statistics

Syntax

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

`options`

Description

Main

<u>detail</u>	display additional statistics
<u>meanonly</u>	suppress the display; calculate only the mean; programmer's option
<u>format</u>	use variable's display format
<u>separator</u> (#)	draw separator line after every # variables; default is <code>separator(5)</code>
<u>display_options</u>	control spacing and base and empty cells

[Link](#)

Stata Syntax

summarize [[varlist](#)] [[if](#) *exp*] [[in](#) *range*] [[weight](#)] [, *options*]

Example—sysuse auto

- summarize mpg price
- summarize in 1/10
- summarize if foreign == 1
- summarize if foreign == 1 & mpg > 30
- summarize , detail

Stata Shortcuts

[Link](#)

Prefix Commands: Repeat Stata command on subsets of the data

[by varlist:] summarize [*varlist*] [*if exp*] [*in range*] [*weight*] [, *options*]

Example—sysuse auto

- `by foreign: summarize rep78`
- `sort foreign make`
- `by foreign (make): summarize rep78`
- `by foreign rep78, sort: summarize price`

cc	year	life					
Afghanistan	1800	28.21					
Afghanistan	1801	28.2					
Afghanistan	1802	28.19					
Afghanistan	1803	28.18					
Afghanistan	1804	28.17					
Afghanistan	1805	28.16					
Afghanistan	1806	28.15					
Afghanistan	1807	28.14					
Afghanistan	1808	28.13					
Afghanistan	1809	28.12					
Afghanistan	1810	28.11					
Afghanistan	1811	28.1					
Afghanistan	1812	28.09					
Afghanistan	1813	28.08					
Afghanistan	1814	28.07					
Afghanistan	1815	28.06					
Afghanistan	1816	28.05					
Afghanistan	1817	28.04					
Afghanistan	1818	28.03					

Excel files

import excel

Import & export Excel files

- **import excel** [[using](#)] **filename** [, *import_excel_options*]

Options	Descriptions
sheet("Sheetname")	worksheet to load
cellrange([start][:end])	cell range to load
firstrow	treat first row of Excel data as variable names
allstring	import all Excel data as strings
clear	replace data in memory

Example

[UCLA](#)

References

[Princeton](#)

[UNC](#)

[Stata](#)

p2003, tp2004, tp2005, tp2006, tp2007, tp2008, tp2009, tp2010, tp2011, tp2012, 14, tp2015

"Afghanistan", 3280000, 3280000, 3323519, 3448982, 3625022, 3810047, 3973968, 19695, 4710171, 5021241, 5351413, 5813814, 6394908, 7034081, 7752118, 7839426, 38312, 8150037, 8270024, 8398309, 8534913, 8679848, 8833127, 8994793, 9164945, 31555, 9728645, 9935358, 10148841, 10368600, 10599790, 10849510, 11121097, 116896, 12022514, 12315553, 12582954, 12831361, 13056499, 13222547, 13283279, 196923, 12667001, 12279095, 11912510, 11630498, 11438949, 11337932, 11375768, 067570, 12789374, 13745630, 14824371, 15869967, 16772522, 17481800, 18034130, 9038420, 19701940, 20531160, 21487079, 22507368, 23499850, 24399948, 2518361, 26528741, 27207291, 27962207, 28809167, 29726803, 30682500, 31627506, 325265

"Albania", 410445, 423591, 438671, 457234, 478227, 506889, 552800, 610036, 672819950, 901122, 963956, 1015991, 1123210, 1263171, 1287499, 1316086, 1348097, 9969, 1459089, 1500152, 1543224, 1588478, 1636054, 1685201, 1735645, 1791532, 6125, 1947786, 1998695, 2049147, 2099657, 2150602, 2202040, 2253842, 2305999, 1229, 2464338, 2517869, 2571845, 2626290, 2681245, 2735329, 2788315, 2842620, 6799, 3041003, 3121336, 3197064, 3253659, 3281453, 3275438, 3240613, 3189623, 6727, 3092034, 3092471, 3102898, 3114851, 3121965, 3124093, 3123112, 3117045, 2172, 3050741, 3010849, 2968026, 2929886, 2901883, 2886010, 2880667, 2883281, 6679

Comma delimited

1990>gdp1991>gdp1992>gdp1993>gdp1994>gdp1995>gdp1996>gdp1997>
1999>gdp2000>gdp2001>gdp2002>gdp2003>gdp2004>gdp2005>gdp2006>
2008>gdp2009>gdp2010>gdp2011>gdp2012>gdp2013>gdp2014>gdp2015**CRLF**
n" —>603>603>603>603>603>603>603>603>603>603>604>604>604>604>604>
>604>604>604>607>609>611>613>615>617>619>621>623>625>627>630>632>
>640>643>645>647>649>651>654>656>658>660>663>665>667>669>672>674>
>683>685>688>690>692>695>697>699>702>704>707>709>711>714>716>719>
>729>731>734>736>739>741>744>746>749>751>754>756>759>761>764>767>
>777>780>782>785>788>790>793>796>798>801>804>807>809>812>815>818>
>829>833>837>841>845>849>853>857>862>868>874>879>885>890>896>902>
>925>931>937>943>949>955>961>967>973>979>985>991>997>1004 —>
6 —>1022 —>1029 —>1035 —>1050 —>1069 —>1116 —>1122 —>
5 —>1134 —>1176 —>1187 —>1206 —>1192 —>1189 —>1185 —>
2 —>1168 —>1173 —>1187 —>1178 —>1174 —>1092 —>1046 —>
0 —>1201 —>1231 —>1119 —>1179 —>1155 —>1158 —>1284 —>
4 —>1429 —>1384 —>1486 —>1230 —>1113 —>1087 —>1028 —>
>810>725>872>895>921>947>972>962>862>1053 —>1097 —>1067 —>
3 —>1298 —>1311 —>1548 —>1637 —>1695 —>1893 —>1884 —>
5**CRLF**

Tab delimited

Insheet

Read data created by a spreadsheet

- **insheet** [[varlist](#)] using filename [, options]

Options	Descriptions
tab	tab-delimited data, .txt
comma	comma-delimited data, .csv
delimiter("char")	use char as delimiter
clear	replace data in memory
[no]names	variable names included on first line of the file

Examples

[UCLA](#)

[UCLA](#)

References

[Princeton](#)

[Stata](#)

"Benin"		"Sub-Saharan Africa"
"BJ"	"Africa"	
"Botswana"		"Sub-Saharan Africa"
"BW"	"Africa"	
"Burkina Faso"		"Sub-Saharan Africa"
"BF"	"Africa"	
"Burundi"		"Sub-Saharan Africa"
"BI"	"Africa"	
"Cameroon"		"Sub-Saharan Africa"
"CM"	"Africa"	
"Cape Verde"		"Sub-Saharan Africa"
"CV"	"Africa"	
"Central African Rep."		"Sub-Saharan Africa"
"CF"	"Africa"	
"Chad"		"Sub-Saharan Africa"
"TD"	"Africa"	
"Comoros"		"Sub-Saharan Africa"
"KM"	"Africa"	
"Congo, Dem. Rep."		"Sub-Saharan Africa"
"CD"	"Africa"	
"Congo, Rep."		"Sub-Saharan Africa"
"CG"	"Africa"	
"Cote d'Ivoire"		"Sub-Saharan Africa"
"CI"	"Africa"	
"Equatorial Guinea"		"Sub-Saharan Africa"
"GQ"	"Africa"	
"Eritrea"		"Sub-Saharan Africa"
"ER"	"Africa"	
"Eritrea and Ethiopia"		"Sub-Saharan Africa"
"Eritrea_Ethiopia"	"Africa"	
"Ethiopia"		"Sub-Saharan Africa"
"ET"	"Africa"	
"Gabon"		"Sub-Saharan Africa"
"GA"	"Africa"	
"Gambia"		"Sub-Saharan Africa"
"GM"	"Africa"	

"Benin"		"Sub-Saharan Africa"
"BJ"	"Africa"	
"Botswana"		"Sub-Saharan Africa"
"BW"	"Africa"	
"Burkina Faso"		"Sub-Saharan Africa"
"BF"	"Africa"	
"Burundi"		"Sub-Saharan Africa"
"BI"	"Africa"	
"Cameroon"		"Sub-Saharan Africa"
"CM"	"Africa"	
"Cape Verde"		"Sub-Saharan Africa"
"CV"	"Africa"	
"Central African Rep."		"Sub-Saharan Africa"
"CF"	"Africa"	
"Chad"		"Sub-Saharan Africa"
"TD"	"Africa"	
"Comoros"		"Sub-Saharan Africa"
"KM"	"Africa"	
"Congo, Dem. Rep."		"Sub-Saharan Africa"
"CD"	"Africa"	
"Congo, Rep."		"Sub-Saharan Africa"
"CG"	"Africa"	
"Cote d'Ivoire"		"Sub-Saharan Africa"
"CI"	"Africa"	
"Equatorial Guinea"		"Sub-Saharan Africa"
"GQ"	"Africa"	
"Eritrea"		"Sub-Saharan Africa"
"ER"	"Africa"	
"Eritrea and Ethiopia"		"Sub-Saharan Africa"
"Eritrea_Ethiopia"	"Africa"	
"Ethiopia"		"Sub-Saharan Africa"
"ET"	"Africa"	
"Gabon"		"Sub-Saharan Africa"
"GA"	"Africa"	
"Gambia"		"Sub-Saharan Africa"
"GM"	"Africa"	

Unformatted data

Infile

Read unformatted text data

- **infile varlist** [*varlist* [*_skip*(#)] ...]] **using filename** [*if*] [*in*] [, *options*]

Options	Descriptions
clear	replace data in memory
automatic	create value labels from nonnumeric data

Example

[UCLA](#)

References

[Princeton](#)

[Stata](#)

The 5 Basic Concepts

Found in all Programming Languages, [Link](#).

1. Variables
2. Control Structures
3. Data Structures
4. Syntax
5. **Tools**

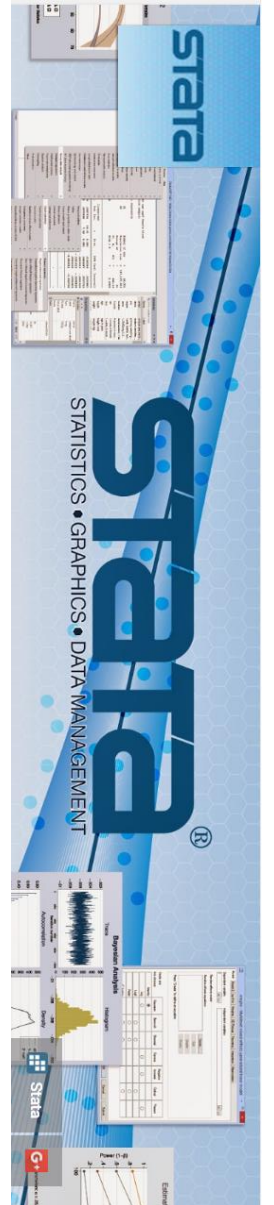
Tools

Anything that helps gets your job done faster!

IDE—Limited options when using **Stata**, Link.

- Integrated Development Environment
- Check syntax, organize & view files, Code completion & navigation

User-written procedure files—see findit carryforward



Findit

Uncertain source—**findit** estout

- Search **Stata** docs for user package, select, install.
- Not always trivial to find correct package 😞
- See [Link](#) or [Link](#).

Known source—**ssc install** estout

List of UCLA developed packages, [Link](#).

Carryforward

- **carryforward** varlist [*if exp*] [*in range*], {gen(*newvarlist*) | replace}

Options

by sort :

Descriptions

Used with extended macro function

Limits flow of values to within by groups

x	y
12	12
4	4
.	4
.	4
.	4
3	3
.	3
7	7
.	7
.	7

Example

[UCLA](#)

References

[Statalist](#)

[Replace](#) [_n-1]

Stata Blog

Project 1

- Using import excel with real world data—Import Raw *GapMinder* Data
- Example, [Link](#).
- Data; [Link1](#), [Link2](#).

Advanced Topic—Extended Macro Function

- See **help extended_fcn**
- More Info, [Link](#).

