# **Data Processing** with Stata 14.1 Cheat Sheet

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

# Useful Shortcuts

- **F2** keyboard buttons Ctrl + 9 describe data open a new .do fi Ctrl + 8 Ctrl + D highlight text in .d open the data editor clear then ctrl + d exec in the command l delete data in memory AT COMMAND PROMPT PgUp PgDn scroll through previous com Tab autocompletes variable name after typi
- cls clear the console (where results are dis

### Set up

#### pwd

- print current (working) directory
- cd "C:\Program Files (x86)\Stata13" change working directory

#### dir

display filenames in working directory

#### fs \*.dta

List all Stata data in working directory	underlined parts are shortcuts –
	use "conture"

# log using "myDoFile.txt", replace

create a new log file to record your work and search mdesc find the package mdesc to install

## ssc install mdesc

install the package mdesc; needs to be done of

### **Import Data**

sysuse auto, clear load system data (Auto data) for many examples, we use the auto dataset.
<b>use</b> "yourStataFile.dta", <b>clear</b> load a dataset from the current directory <sub>frequently used</sub>
import excel "yourSpreadsheet.xlsx", /* commands are highlighted in yellow
*/ sheet("Sheet1") cellrange(A2:H11) firstrow import an Excel spreadsheet
<pre>import delimited"yourFile.csv", /*</pre>
webuse set "https://github.com/GeoCenter/StataTraining/raw/master/Day2/Data webuse "wb_indicators_long" set web-based directory and load data from the web

	<i>command</i> across you going to <b>do</b> apply a new variable apply the function each unique to <i>varlists</i> ? <i>command</i> to <i>if</i> something is true combination of	
file	variables in varlist1 bysort rep78 : summarize price if foreign == 0 & p	price <= 9000, detail
.do file,	To find out more about any command – like what opt	tions it takes – type <b>help</b> command
ecutes it I line mmands	Arithmetic       Logic       == tests if something is equal         +       add (numbers) combine (strings)       & and       == equal       < less than         !       or ~ not       != not       <= less than or equal to	Change Data Types           Stata has 6 data types, and data can also be missing:           no data         true/false         words         numbers           missing         byte         string         int long float double           To convert between numbers & strings:
ping part lisplayed)	<ul> <li>subtract</li> <li>multiply</li> <li>divide</li> <li>foreign != 1 &amp; price &gt;= 10000</li> <li>if foreign price</li> <li>foreign != 1 &amp; price &gt;= 10000</li> <li>if foreign price</li> <li>buck Rivera 0 10,372</li> <li>raise to a power</li> </ul>	gen foreignString = string(foreign)       "1"         1       tostring foreign, gen(foreignString)       "1"         1       decode foreign , gen(foreignString)       "foreign"         2       gen foreignNumeric = real(foreignString)       "1"         1       destring foreignString, gen(foreignNumeric)       "1"         1       destring foreignString, gen(foreignNumeric)       "1"         1       foreignString, gen(foreignNumeric)       "1"         1       foreignString, gen(foreignNumeric)       "1"
	Explore Data	recast double mpg generic way to convert between types
	VIEW DATA ORGANIZATION describe make price display variable type, format, and any value/variable labels SEE DATA DISTRIBUTION codebook make price overview of variable type, stats, number of missing/unique values	Summarize Data include missing values create binary variable for every rep78 rvalue in a new variable, repairRecord tabulate rep78, mi gen(repairRecord)
ined parts prtcuts – apture" "	count count if price > 5000 number of rows (observations) Can be combined with logicsummarize make price mpg print summary statistics (mean, stdev, min, max) for variables	one-way table: number of rows with each value of rep78 tabulate rep78 foreign, mi two-way table: cross-tabulate number of observations for each combination of rep78 and foreign
d results ontain ands that a's toolkit	ds, has(type string)inspect mpglookfor "in."show histogram of data,search for variable types,number of missing or zerovariable name, or variable labelobservations	bysort rep78: tabulate foreign for each value of rep78, apply the command tabulate foreign tabstat price weight mpg, by(foreign) stat(mean sd n) create compact table of summary statistics displays stats formats numbers for all data
once	isid mpg check if mpg uniquely identifies the data bit of a variable BROWSE OBSERVATIONS WITHIN THE DATA	table foreign, contents(mean price sd price) f(%9.2fc) row create a flexible table of summary statistics collapse (mean) price (max) mpg, by(foreign) - replaces data
nples, we dataset.	browse or Ctrl + 8 Missing values are treated as the largest positive number. To exclude missing values, use the !missing(varname) syntax	calculate mean price & max mpg by car type (foreign) Create New Variables
uently used mands are lighted in yellow ) <b>firstrow</b>	<b>ist</b> make price if price > 10000 & !missing(price)       clist (compact form)         list the make and price for observations with price > \$10,000         display price[4]         display the 4th observation in price; only works on single values	<pre>generate mpgSq = mpg^2 gen byte lowPr = price &lt; 4000 create a new variable. Useful also for creating binary variables based on a condition (generate byte) generate id = _n bysort rep78: gen repairIdx = _n _n creates a running index of observations in a group</pre>
names(2)	gsort price mpg (ascending)       gsort -price -mpg (descending)         sort in order, first by price then miles per gallon         duplicates report finds all duplicate values in each variable       assert price!=.	<pre>generate totRows = _N bysort rep78: gen repairTot = _N _N creates a running count of the total observations per group pctile mpgQuartile = mpg, nq = 4</pre>
aster/Day2/Data" the web	livelsof rep78 display the unique values for rep78	create quartiles of the mpg data <b>egen</b> meanPrice = mean(price), by(foreign) calculate mean price for each group in foreign
ghes@usaid.gov) inspired by RStudio's awesome Cheat Sheets ( <u>rstudio.com/resources/cheatsheets</u> ) <u>geocenter.github.io/StataTraining</u> updated January 2016 Disclaimer: we are not affiliated with Stata. But we like it. <u>CC BY 4.0</u>		

**Basic Syntax** 

[if exp]

[in range]

[weight]

[using filename]

[,options]

All Stata functions have the same format (syntax):

command

[varlist2]

[=exp]

[**by** varlist1:]

Tim Essam (tessam@usaid.gov) • Laura Hughes (lhug follow us @StataRGIS and @flaneuseks