Stata Crib Sheet

Below are some Stata functions that you may find useful. If you cannot remember the syntax for what you want to do, use the drop-down menus to find the command.

**pwd** shows the current (present) working directory.

**cd [directory name]**

*cd* changes the working directory to the drive and directory specified. On a Windows machine, you may need to enclose the directory path in quotation marks [“”].

**use [file]** opens the .dta file.

**browse** Browse data in a spreadsheet format.

**codebook [varlist]**

This will give you some information about the variable(-s) you specify. *codebook* with no variables specified gives you information about all variables in the dataset.

**describe [varlist]**

This will give you descriptive information about the variable you specify: the type of the variable (text/string or numeric) and the variable label.

**list [varlist]**

This will list the values for the variables specified.

If you need to see the first 5 rows of the data use:

**list [varlist] in 1/5**

**summarize [varlist]**

*summarize* calculates and displays a variety of univariate summary statistics. If no varlist is specified, summary statistics are calculated for all the variables in the dataset.

**summarize [varlist], detail** provides additional details, including various percentiles.

**generate [type] newvar[:lblname] = exp [if exp] [in range]**

*generate* creates a new variable. The values of the variable are specified by = *exp*.

command *if* [exp]

*if* at the end of a command means the command is to use only the data specified. *if* is allowed with most Stata commands.
tabulate [varname1] [varname2]
tabulate produces one- and two-way tables of frequency counts.

sort [varlist]
sort arranges the observations of the current data in ascending order of the values of the variables in varlist.

gsort [varlist]
Similar to sort, but gsort allows to transform variables in varlist.

histogram [varname], options
histogram draws histograms of the variable you specify in varname. Options include:
- percent (to display the percent of observations in each bin)
- frequency (to display the number of observations in each bin)

The default option is density.

spmap Visualization of spatial data, to draw a map for a given variable. To use this function, you need to install an external package (if it is not already installed):
ssc install spmap, replace
Here is how to use spmap:
spmap [varname] using basemap, id(idvar) fcolor(colorlist)

scatter [varlist] draws scatter plots.
scatter yvar xvar

regress [depvar] [varlist]
regress fits a model of depvar on varlist using linear regression.

tabstat [varlist]
tabstat provides various useful summary statistics for numeric variables in a single table.

tabstat [varlist], stat([statname])
provides specific statistics as specified (for instance, mean, median, sum, min, max, etc.).

tabstat displays summary statistics for a series of numeric variables in a single table, possibly broken down on (conditioned by) another variable. For instance:

tabstat [varlist] [weight] [if exp] [in range] [, statistics(statname [...])]