

Stata Overview, Interface, and File Types

Stata

- Stata advertises itself as a fast, easy-to-use, and powerful statistical software package
- True! In my humble opinion, Stata is very fast, one of the easiest package to learn, and has support for a wide (and growing) array of analyses
- Stata also prides itself on being both a point-and-click **and** command-line based package
 - *This workshop is all about the command line*

Stata Pros / Cons

PROS

- Fast, Strong, Efficient
- Intuitive (for the most part)
- Well-organized and understandable documentation
- Proprietary (Unified Vision, Overarching Plan)
- Reproducible
- Publication-level graphics

CONS

- Proprietary (Licenses, Top-down control)
- Idiosyncratic Syntax
- Less flexible for bespoke analyses
- Proprietary != External Reproducibility
- Smaller User Ecosystem
- Less flexible data management
- Too many windows!

Stata = Great!

- Take my opinions with a grain of salt — Stata is a **GREAT** option for organizations trying to improve internal reproducibility
- Let's all take a look at the Stata Interface together:
 - Main Window
 - Command (*where you type commands*)
 - Results (*where you view results*)
 - Review (*a history of commands you have run*)
 - Variables (*a quick look at the variables in your dataset*)

Datasets

- In Stata, a dataset is a group of observations with values for a group of variables, along with metadata about the data, the variables, and what their levels mean (in the form of labels and notes)
- A Stata dataset file is called a **dta-file** and has the file extension **.dta**
- Stata can read many other types of data files, but works best with **.dta** files because of its rich metadata tracking system

Do File Editor

- Stata's power gets unlocked when you combine a series of commands together into what's called a **do-file**.
- In other languages this might be called a *script*, but in Stata it's a do file.
- These files are Stata-specific and have the extension `.do` (e.g. `clean_data.do`)
- Let's run the `setup.do` file to download some files we're going to use!

Data Browser

- Stata has a built-in Data Browser which allows you to peruse your data set
 - This shows your data in a traditional excel-file type spreadsheet format, but presents categorical variable labels in a useful way
- There is also a Data Editor for manually changing values in the spreadsheet
 - **I strongly advise against using this!**

Viewer Window

- Whenever you ask for **help** with a command or concept, the Viewer Window will pop up, showing you Stata's built-in documentation
- At the top of these help files there are often links to even more detailed Stata manuals (in PDF form).

Graph Editor

- The Graph Editor is a point and click tool for editing graphs
- It allows the user to customize almost every part of a graph that has already been created (but the graph must be created first to manipulate)
- Stata graphs can be saved with the **.gph** extension. They are not flat files (like a jpeg or a tif), they can be loaded and re-manipulated
- **Point-and-click solutions work counter to automation and reproducibility**
 - However, sometimes you want to just change that one little thing...

Windows Galore!

- Stata has a lot of distinct windows — my advice is to mainly write scripts and run commands from the Do File Editor
- I keep the Main Window open on one side of my screen, with the Do File Editor on the other
- I tend not to use the Data Browser often, but your preferences will evolve as you work more with Stata