

Computational Bootcamp 4: Stata

Henry Watson

Georgetown University

8/18/22

What We'll Be Covering Overall

- ① Software installation, file management
- ② Basics of R: writing code, creating objects, thinking in matrices
- ③ More R: working with datasets
- ④ Stata: pros & cons vs. R, working with datasets
- ⑤ LaTeX: producing documents with Markdown and Overleaf

What We'll Be Covering Today

- 1 Why learn Stata when we have R?

What We'll Be Covering Today

- ① Why learn Stata when we have R?
- ② The Stata Interface

What We'll Be Covering Today

- ① Why learn Stata when we have R?
- ② The Stata Interface
- ③ Loading in datasets

What We'll Be Covering Today

- ① Why learn Stata when we have R?
- ② The Stata Interface
- ③ Loading in datasets
- ④ Manipulating datasets

Why learn Stata?

Why learn Stata?

- Stata blends point-and-click menus with reproducible code

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual
 - Generally speaking, anything you can do in R you can do in Stata, and vice versa

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual
 - Generally speaking, anything you can do in R you can do in Stata, and vice versa
 - Still fairly common in political science and adjacent fields — good to communicate with other scholars
- Another Tool in the Toolbox

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual
 - Generally speaking, anything you can do in R you can do in Stata, and vice versa
 - Still fairly common in political science and adjacent fields — good to communicate with other scholars
- Another Tool in the Toolbox
 - Stronger at dataset cleaning and manipulation

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual
 - Generally speaking, anything you can do in R you can do in Stata, and vice versa
 - Still fairly common in political science and adjacent fields — good to communicate with other scholars
- Another Tool in the Toolbox
 - Stronger at dataset cleaning and manipulation
 - Variable labeling and other "quality of life"

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual
 - Generally speaking, anything you can do in R you can do in Stata, and vice versa
 - Still fairly common in political science and adjacent fields — good to communicate with other scholars
- Another Tool in the Toolbox
 - Stronger at dataset cleaning and manipulation
 - Variable labeling and other "quality of life"
 - Better and more efficient at some statistical applications e.g. time series and panel data

Why learn Stata?

- Stata blends point-and-click menus with reproducible code
- Being Bilingual
 - Generally speaking, anything you can do in R you can do in Stata, and vice versa
 - Still fairly common in political science and adjacent fields — good to communicate with other scholars
- Another Tool in the Toolbox
 - Stronger at dataset cleaning and manipulation
 - Variable labeling and other "quality of life"
 - Better and more efficient at some statistical applications e.g. time series and panel data
 - Can install user-made packages, but isn't as reliant on them as R

Downsides of Stata

- Stata presumes you're working with a dataset

Downsides of Stata

- Stata presumes you're working with a dataset
 - Weak for simulating data

Downsides of Stata

- Stata presumes you're working with a dataset
 - Weak for simulating data
- More difficult to work with multiple datasets simultaneously

Downsides of Stata

- Stata presumes you're working with a dataset
 - Weak for simulating data
- More difficult to work with multiple datasets simultaneously
 - *This isn't as true as it used to be thanks to Frames*

Downsides of Stata

- Stata presumes you're working with a dataset
 - Weak for simulating data
- More difficult to work with multiple datasets simultaneously
 - *This isn't as true as it used to be thanks to Frames*
- Poor tool for data visualization

Downsides of Stata

- Stata presumes you're working with a dataset
 - Weak for simulating data
- More difficult to work with multiple datasets simultaneously
 - *This isn't as true as it used to be thanks to Frames*
- Poor tool for data visualization
- Not free (but...)

The Stata Interface

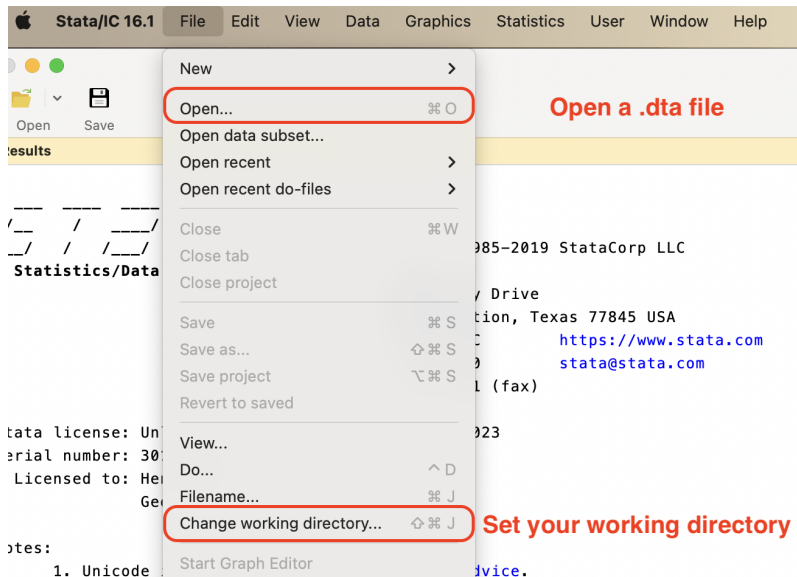
The screenshot shows the Stata 16.1 interface with several key components highlighted by red arrows and labels:

- Open/Save Stata Datasets (.dta)**: Points to the 'Open' and 'Save' icons in the top-left toolbar.
- Do-file Editor (Scripts)**: Points to the 'Do File Editor' icon in the top toolbar.
- Data Viewer and Editor**: Points to the 'Data Editor' icon in the top toolbar.
- Results (Output)**: Points to the main command window area displaying Stata's startup information, including the license, serial number, and update status.
- Variables Pane (Similar to R's Environment)**: Points to the 'Variables' pane on the right side of the interface, which lists the variables in the current dataset.
- Variable Properties**: Points to the 'Properties' pane on the right side of the interface, which shows the properties of the selected variable.
- Command Line (Similar to R's Console)**: Points to the 'Command' window at the bottom of the interface, where users can enter Stata commands.

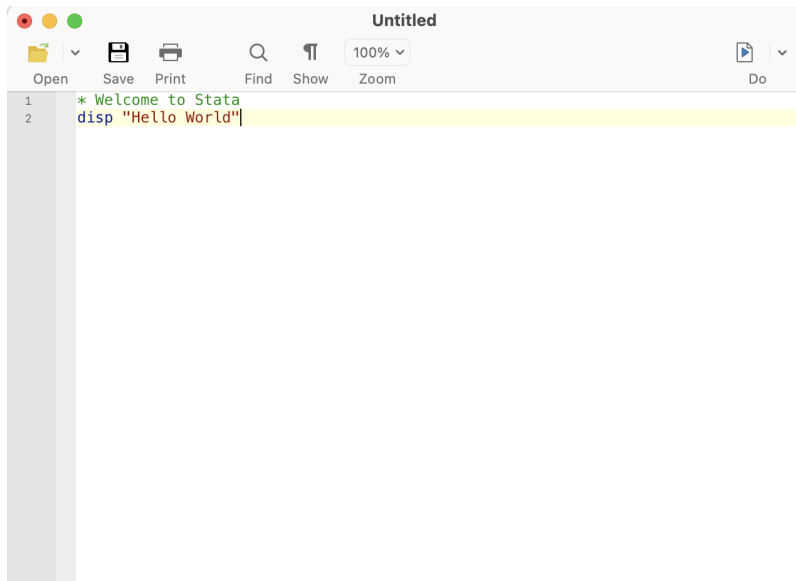
The main command window displays the following text:

```
Stata 16.1 Copyright 1989-2019 StateCorp LLC  
StateCorp  
4965 Lakeway Drive  
College Station, Texas 77845 USA  
888-STATA-PC https://www.stata.com  
979-696-4688 stata@stata.com  
979-696-4683 (fax)  
  
Stata license: Unlimited-user, expiring 7 Feb 2023  
Serial number: 381780311370  
Licensed to: Henry Watson  
Georgetown University  
  
Notes:  
1. Unicode is supported; see help unicode_advice.  
  
Checking for updates...  
(contacting http://www.stata.com)  
  
Update status  
Last check for updates: 15 Aug 2022  
New update available: none (as of 15 Aug 2022)  
Current update level: 14 Jun 2022 (what's new)  
  
Possible actions  
Do nothing; all files are up to date.  
Click to edit automatic update checking preferences
```

The Stata Interface



The Stata Interface



A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)
- Can comment out multiple lines of code by enclosing multiple lines between /* and */

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)
- Can comment out multiple lines of code by enclosing multiple lines between /* and */
- Very common to abbreviate Stata commands

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)
- Can comment out multiple lines of code by enclosing multiple lines between /* and */
- Very common to abbreviate Stata commands
 - Can abbreviate to as few letters as is still unambiguous to Stata

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)
- Can comment out multiple lines of code by enclosing multiple lines between /* and */
- Very common to abbreviate Stata commands
 - Can abbreviate to as few letters as is still unambiguous to Stata
 - Example: "generate" is almost always written as "gen"

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)
- Can comment out multiple lines of code by enclosing multiple lines between /* and */
- Very common to abbreviate Stata commands
 - Can abbreviate to as few letters as is still unambiguous to Stata
 - Example: "generate" is almost always written as "gen"
- Syntax is different from R; think of it as reading a logical sentence where commands are verbs
- Example: `generate tall = 1 if height > 72`

A Few Quick Notes on Stata Code

- Your script of Stata code is called a "Do File"
- Running code is Shift+Command+D (for "Do")
- Commenting lines of code is asterisk (*)
- Can comment out multiple lines of code by enclosing multiple lines between /* and */
- Very common to abbreviate Stata commands
 - Can abbreviate to as few letters as is still unambiguous to Stata
 - Example: "generate" is almost always written as "gen"
- Syntax is different from R; think of it as reading a logical sentence where commands are verbs
- Example: `generate tall = 1 if height > 72`
- Missings are . for numeric and "" for strings