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
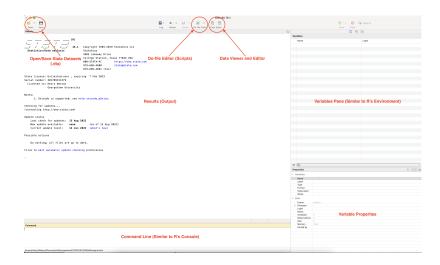 - Weak for simulating data

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
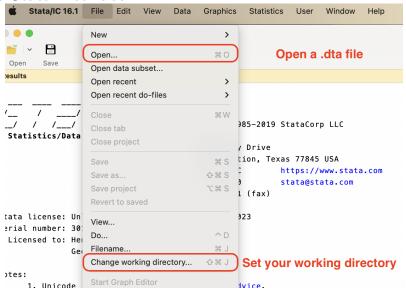   - *This isn't as true as it used to be thanks to Frames*

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*
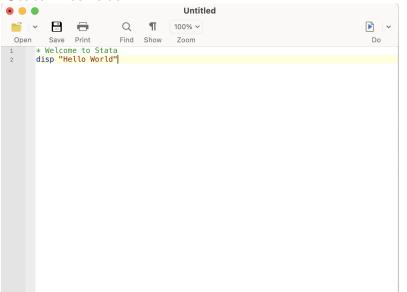- Poor tool for data visualization
- Not free (but...)

The Stata Interface

The 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