L1: INTRODUCTION

Getting started with Stata

Angela Ambroz

May 2015

Today

- Admin
- What is Stata?
- Basics of the basics

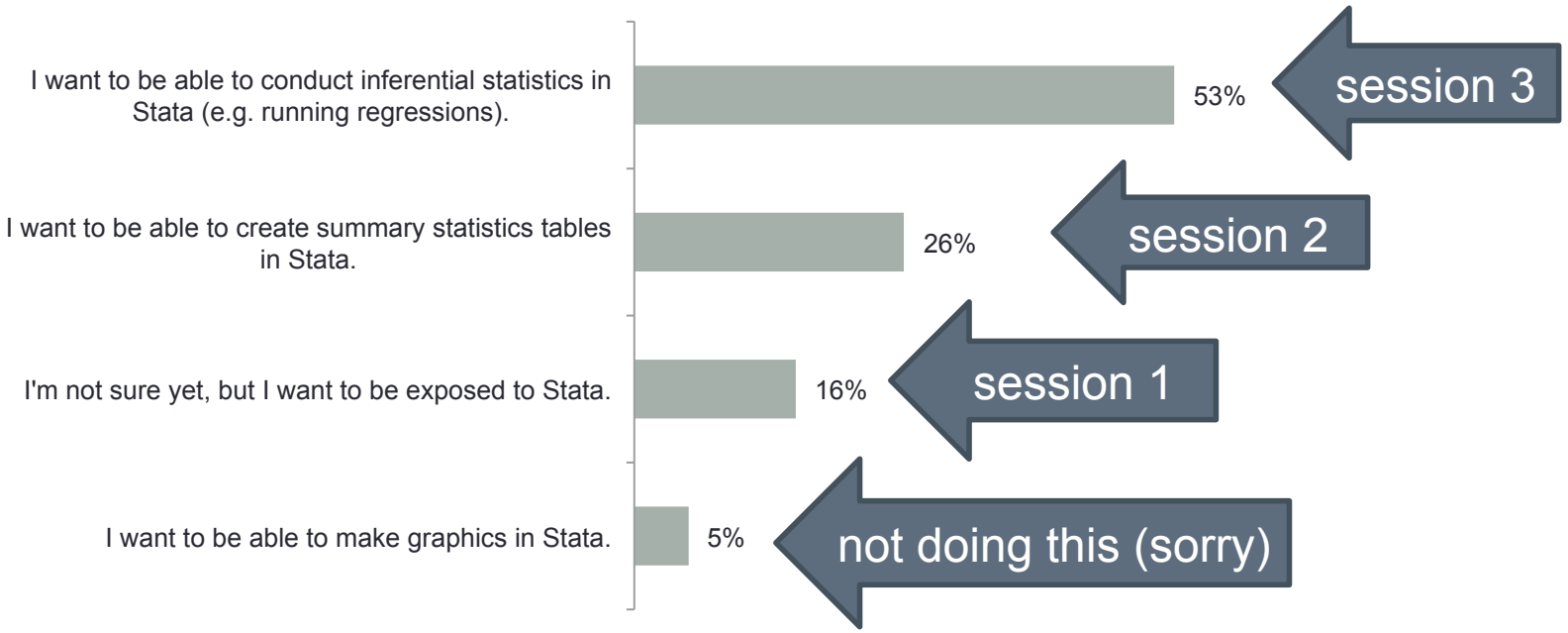
Admin

- 3 lectures
- Homework assignment for each lecture
- Course website: www.angelaambroz.com/stata.html
- 24/7 Support: Skype group chat – CUPCAKE PROMISE
- Make sure you have Stata 10 on your computer
- **Confused? Interrupt me!**

GOAL:
Not being scared/intimidated by Stata. Understanding the basics of how Stata operates. Being able to write and understand .do files.

Pre-course survey results

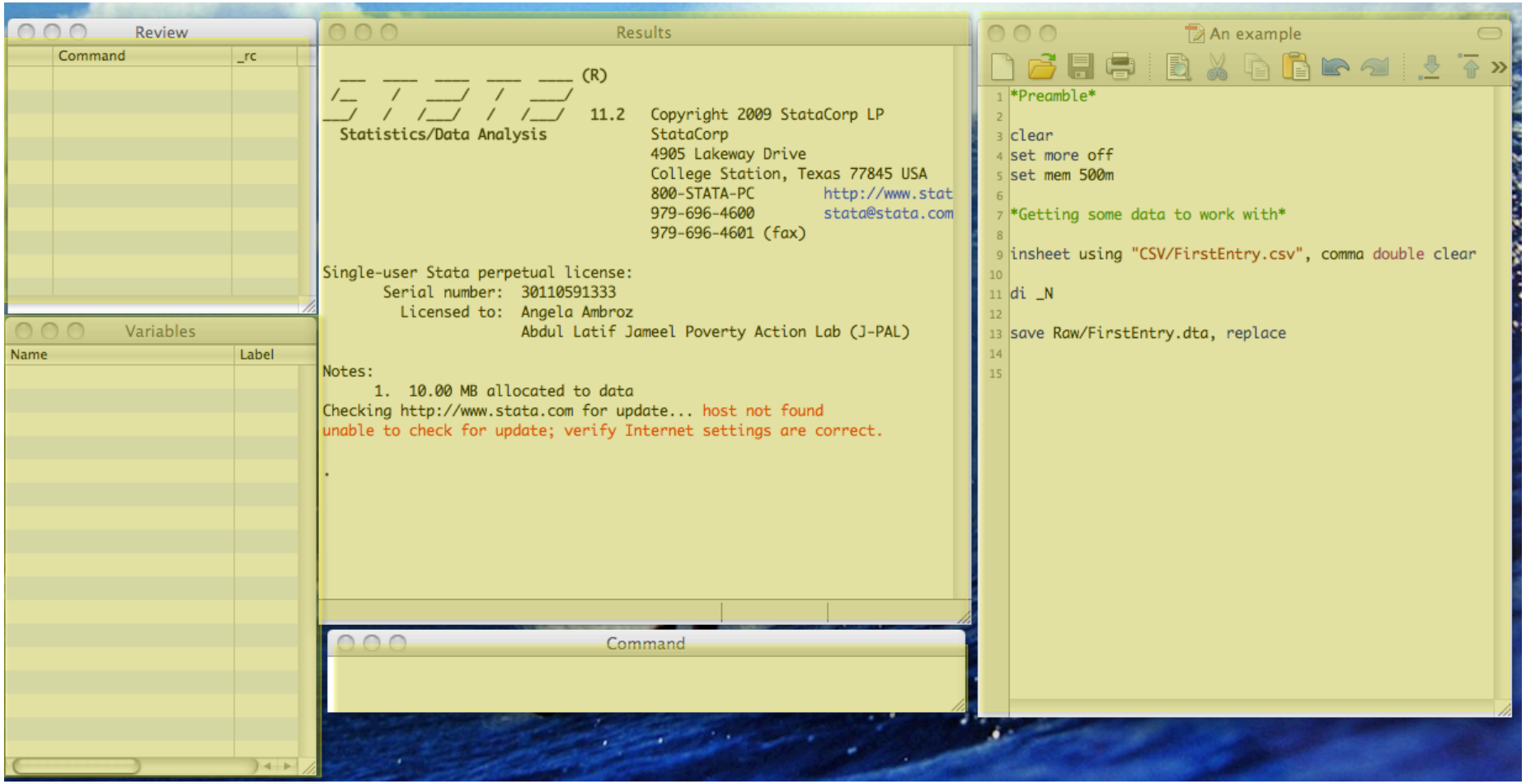
- 74% of you have used Stata before
- 53% of you knew what `browse` does
- 95% of you have used Excel
- 74% of you have used SPSS



Why do we use Stata?

Pros	Cons
- **Logistics:** Easy to learn, cross-platform, industry standard!	- Working with very large datasets (multiple GBs)
- **Analysis**: Many of our most useful tools are built-in and ready to use (e.g. power calcs)	- Making graphs/tables is time-consuming (faster in Excel)
- **Reproducible research**: Stata is command-driven, saved in .do files	- Pricey
- **Flexibility**	

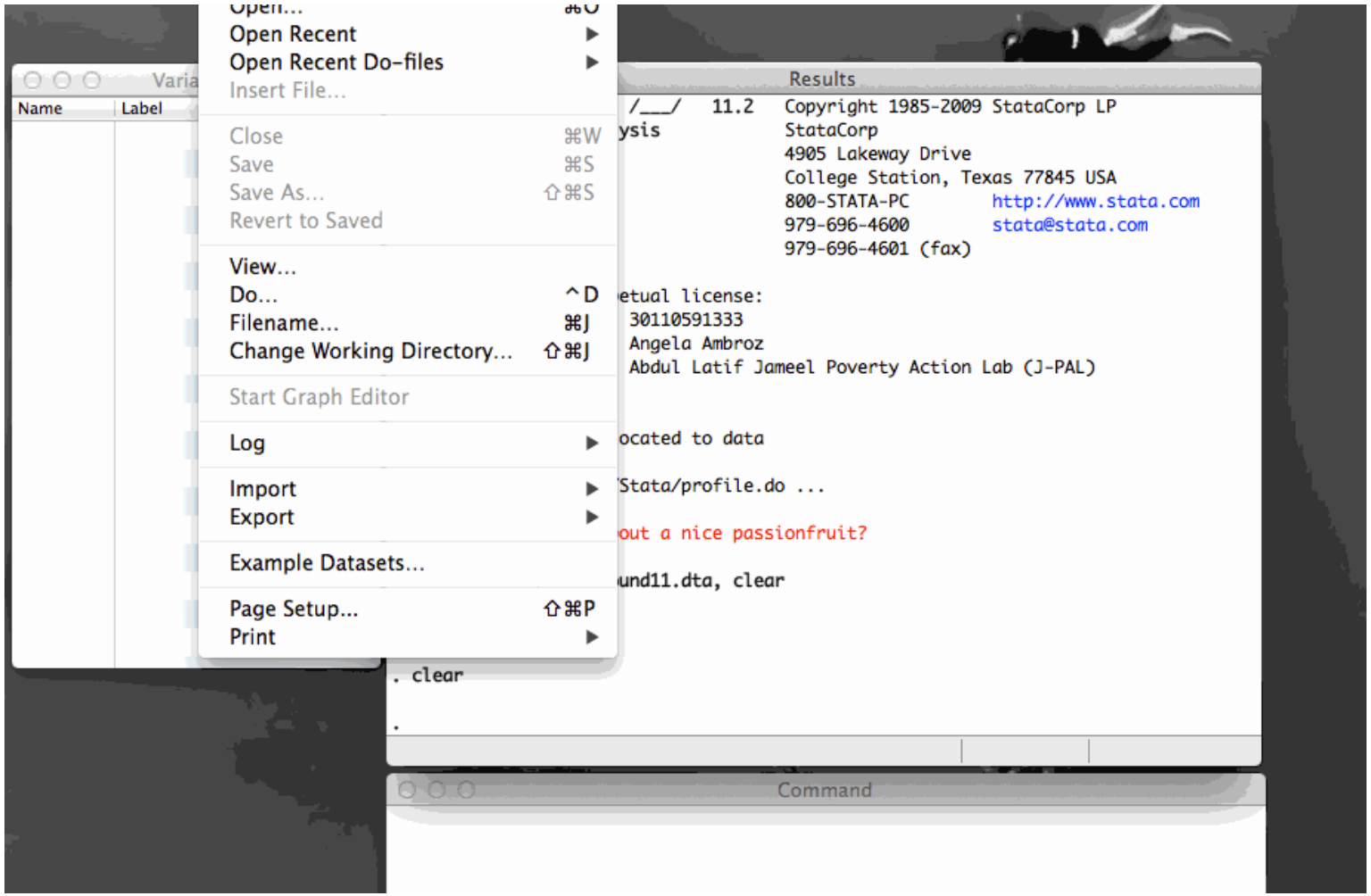
The basics


Review
Command
_rc
Variables
Name
Label
Results
(R)
11.2
Statistics/Data Analysis
Copyright 2009 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC
http://www.stat
979-696-4600
stata@stata.com
979-696-4601 (fax)
Single-user Stata perpetual license:
Serial number: 30110591333
Licensed to: Angela Ambroz
Abdul Latif Jameel Poverty Action Lab (J-PAL)
Notes:
1. 10.00 MB allocated to data
Checking http://www.stata.com for update... host not found
unable to check for update; verify Internet settings are correct.
.
Command
An example
Preamble
clear
set more off
set mem 500m
Getting some data to work with
insheet using "CSV/FirstEntry.csv", comma double clear
di _N
save Raw/FirstEntry.dta, replace

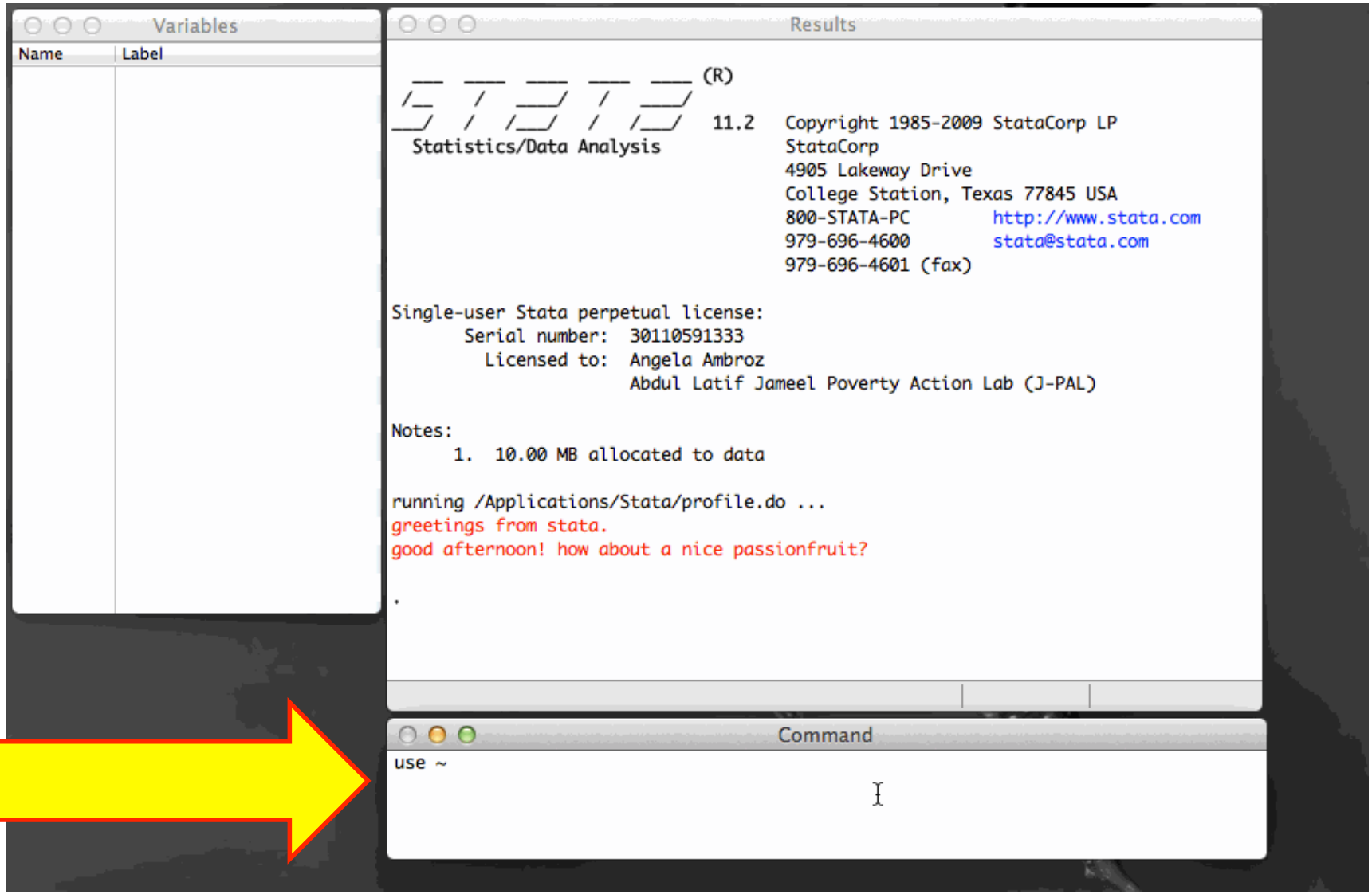
How to talk to Stata

- Stata has both a command-line and graphical user interface
- It uses a specific syntax, similar to programming languages
- Most people write commands in a .do file, save it, and then run it

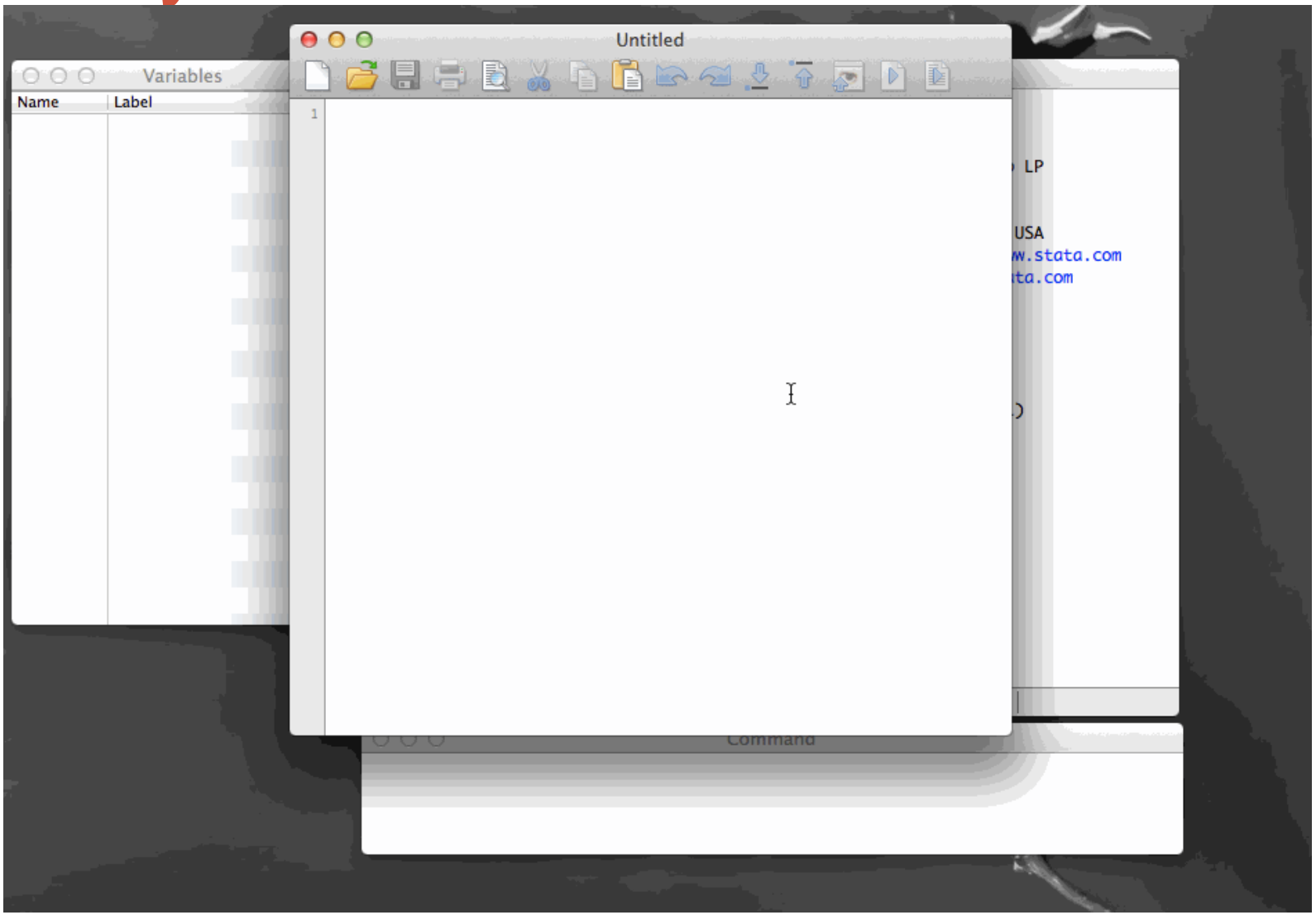
Graphical user interface



Command line interface



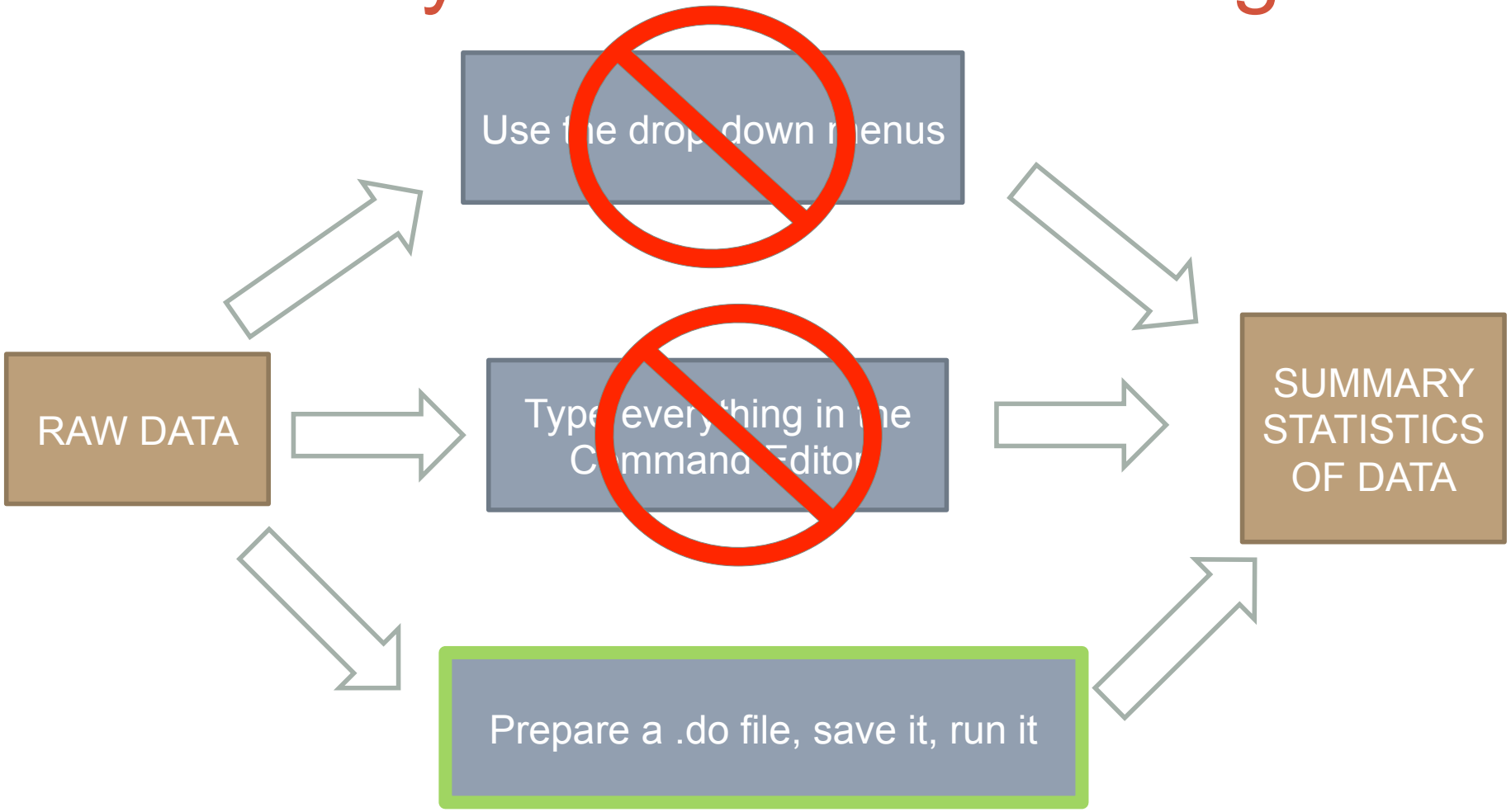
Save your commands: .do files



What is a .do file?

- A .do file is a **plain text file** where you save your Stata commands.
- It's kind of like a **mini program**.
- Saving commands in a .do file makes sure that your work is **reproducible and understandable.**
- Added bonus: you can do and save your analysis in a .do file **without altering the dataset**! (unlike Excel)

Three ways to do the same thing



Word to the wise

Best practice

Test ideas in the Command Editor window

Do your analysis (and save it!) in a .do file!

How does Stata see data?

- Stata won't show you the data ***unless you explicitly ask it*** (via `browse`)
- Data can be either:
  - `numeric: -1, 0.31495, 30000`
  - `string:"Angela", " Dar es Salaam", "1,000", "1/4", "  "`
  - Encoded: `100 ("Dar es Salaam")`
- Missing values show up as `.` (numeric) and " " (string)

How does Stata see data?

variable name	storage type	display format	value label	variable label
uhn	int	%10.0g		**unique household number**
EA	int	%10.0g	ea	**Enumeration Area**
hh_hasphone_bl	byte	%9.0g		**hh had phone at baseline**
hhsize	byte	%9.0g		**nr of hh members from roster**
resp_urban	byte	%9.0g	urb_lab	**Respondent urban**
resp_farmer	byte	%9.0g	Farm_lab	**Respondent is farmer**
resp_male	byte	%9.0g	male_lab	**Respondent is male**
comp_std7	byte	%30.0g	st7lab	**Completed Standard 7 or higher**
comp_form4	byte	%26.0g	fr4lab	**Completed Form 4 or higher**
quin_asset_we~h	byte	%10.0g	assetq_lab	**wealth quintiles**

Variables
The questions in your questionnaire

Values
The answers in your questionnaire

Labels
Meta-data on your questions and answers

How to talk to Stata (via commands)

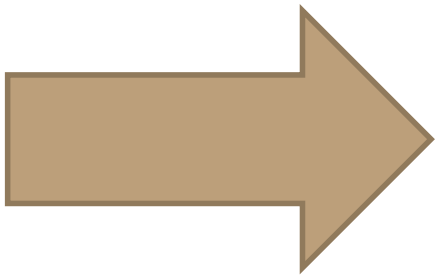
- There are **thousands*** of official Stata commands
- There are **hundreds*** of unofficial, user-written Stata commands (`.ado`)
- There is **no point** in trying to memorize a lot of commands – this is a giant waste of time
- Instead, memorize only one:

```
help
```

* I think, I'm not actually sure.

How to use “help”

You type `help browse`, and Stata will open this:



A key Stata skill: Learning how to read and understand “`help`” files

Title

[D] edit — Browse or edit data with Data Editor

Syntax

Edit using Data Editor

edit [*varlist*] [*if*] [*in*] [, **nolabel**]

Browse using Data Editor

browse [*varlist*] [*if*] [*in*] [, **nolabel**]

Menu

edit

Data > Data Editor > Data Editor (Edit)

browse

Data > Data Editor > Data Editor (Browse)

General syntax structure

- Stata's commands following **programming conventions**.
- As such, they use **logic symbols (`&`, `<`, `>`, etc.)**
- Variable names never have spaces
- Variable names never start with numbers
- Strings always have `"quotes around them"`
- All commands can be sh<u>ort</u>ened:

```
g
  b
ta
```

General syntax structure: Logic symbols

This…	means this…	Example
+ - * /	Add, subtract, multiply, divide	
=	Assignment	`generate weekly_amount = daily_average * 7`
==	Equals	`browse if region==“Dar es Salaam”`
!=	Does not equal	`browse if region!=“Dar es Salaam”`
>=, <=, >, <	Greater/Less than	`browse if age<=5`
&	And	`browse if age<=5 & gender==“Female”`
|	Or	`browse if party==“CHADEMA” | party==“CUF”`

In-class exercise

Let's now take 5-10 minutes, and translate this into English:

```
gen adult_respondent=1 if
age>=18 & name!=""
```

hint: `help...`

Homework

- `homework1-1.do` - Read and explain a simple .do file (written by someone else)
- `homework1-2` (online Google form) – Practice writing and running some simple commands
- **Goal**:
  - **To be able to open a .do file**
  - **To get familiar with the different windows in Stata**
  - **To be able to use and understand the "help" command**
  - **To start writing your first commands**
- Questions? E-mail/Skype.
- Finished? E-mail me your .do file. Fill out the Google form.
- **DEADLINE: COB Monday, May 18, 2015**