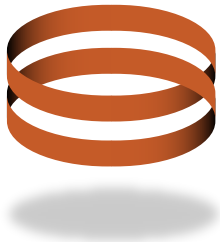
# Regression in Stata

Ista Zahn

Harvard MIT Data Center

April 12 2013



The Institute
*for* Quantitative Social Science
at Harvard University

# Outline

# Topic

# Documents for today

USERNAME: dataclass PASSWORD: dataclass

- Find class materials at: Scratch > StataStatistics
- FIRST THING: copy this folder to your desktop!

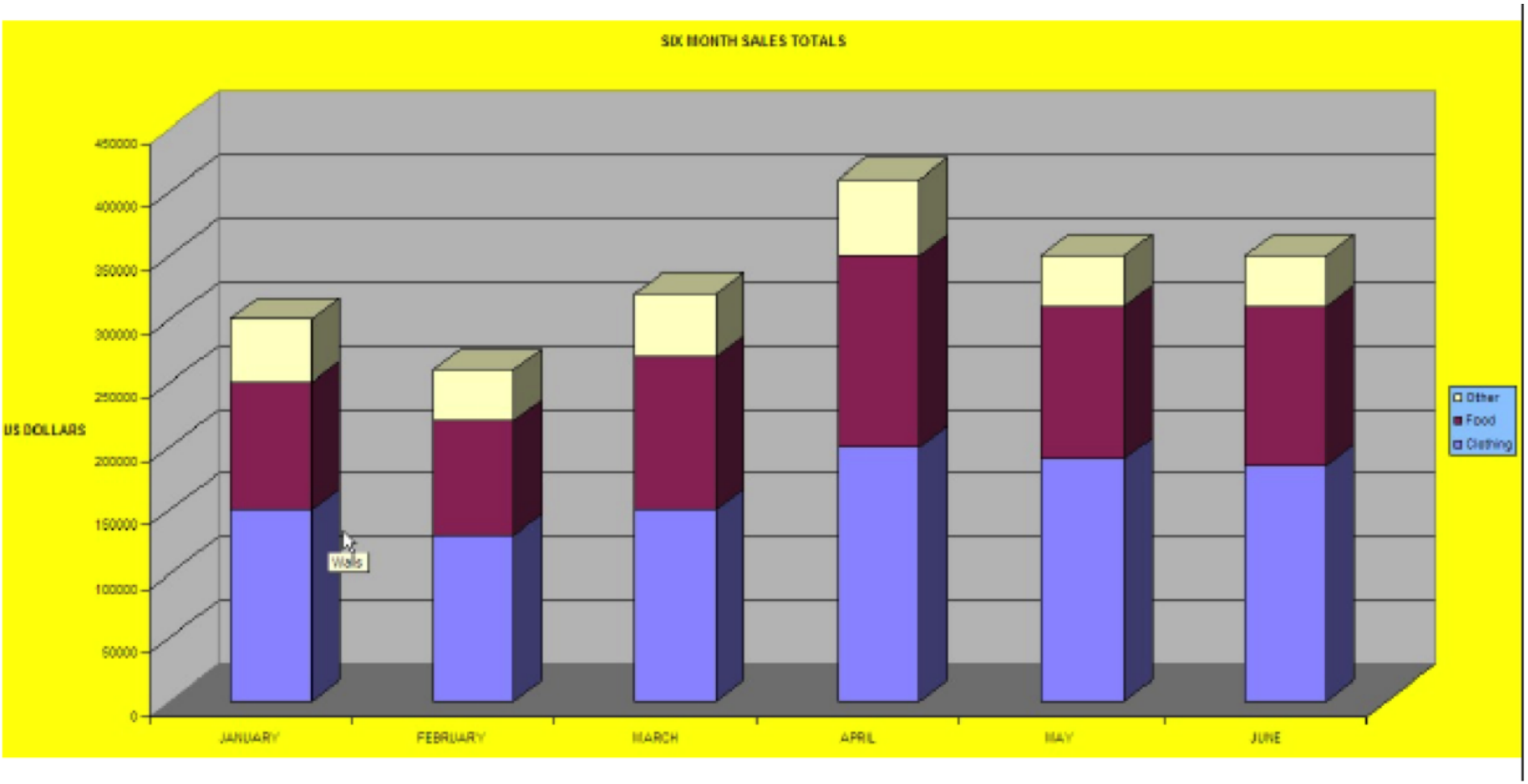
# Organization

- Please feel free to ask questions at any point if they are relevant to the current topic (or if you are lost!)
- There will be a Q&A after class for more specific, personalized questions
- Collaboration with your neighbors is encouraged
- If you are using a laptop, you will need to adjust paths accordingly
- Make comments in your Do-file rather than on hand-outs
- Save on flash drive or email to yourself

# Graphing Strategies

- Keep it simple
- Labels, labels, labels!!
- Avoid cluttered graphs
- Every part of the graph should be meaningful
- Avoid:
  - Shading
  - Distracting colors
  - Decoration

# Terrible Graph



# Much Better Graph


Source: Bureau of Labor Statistics, http://www.bls.gov/data/

# Opening Stata

- Once you have Stata open, let's call up the datafile for today

```
// Step 1: tell Stata where to find data:
cd /Users/dataclass/Desktop/StataGraphics/dataSets
// Step 2: call up our dataset:
use TimePollPubSchools.dta
```

# Basic Graphing

- Always know what you're working with before you get started
  - Recognize scale of data
  - If you're using multiple variables – how do their scales align?
- Before any graphing procedure review variables with `codebook`, `sum`, `tab`, etc.
- HELPFUL STATA HINT: If you want your command to go on multiple lines use /// at end of each line

# Basic Graphing: Single Continuous Variables

Example: Histograms

- Stata assumes you're working with continuous data
- Very simple syntax:
  - `hist varname`
- Put a comma after your varname and start adding options
  - `bin(#)` : change the number of bars that the graph displays
  - `normal` : overlay normal curve
  - `addlabels` : add actual values to bars

# Our First Dataset

- Time Magazine Public School Poll
- Based on survey of 1,000 adults in U.S.
- Conducted in August 2010
- Questions regarding feelings about parental involvement, teachers union, current potential for reform

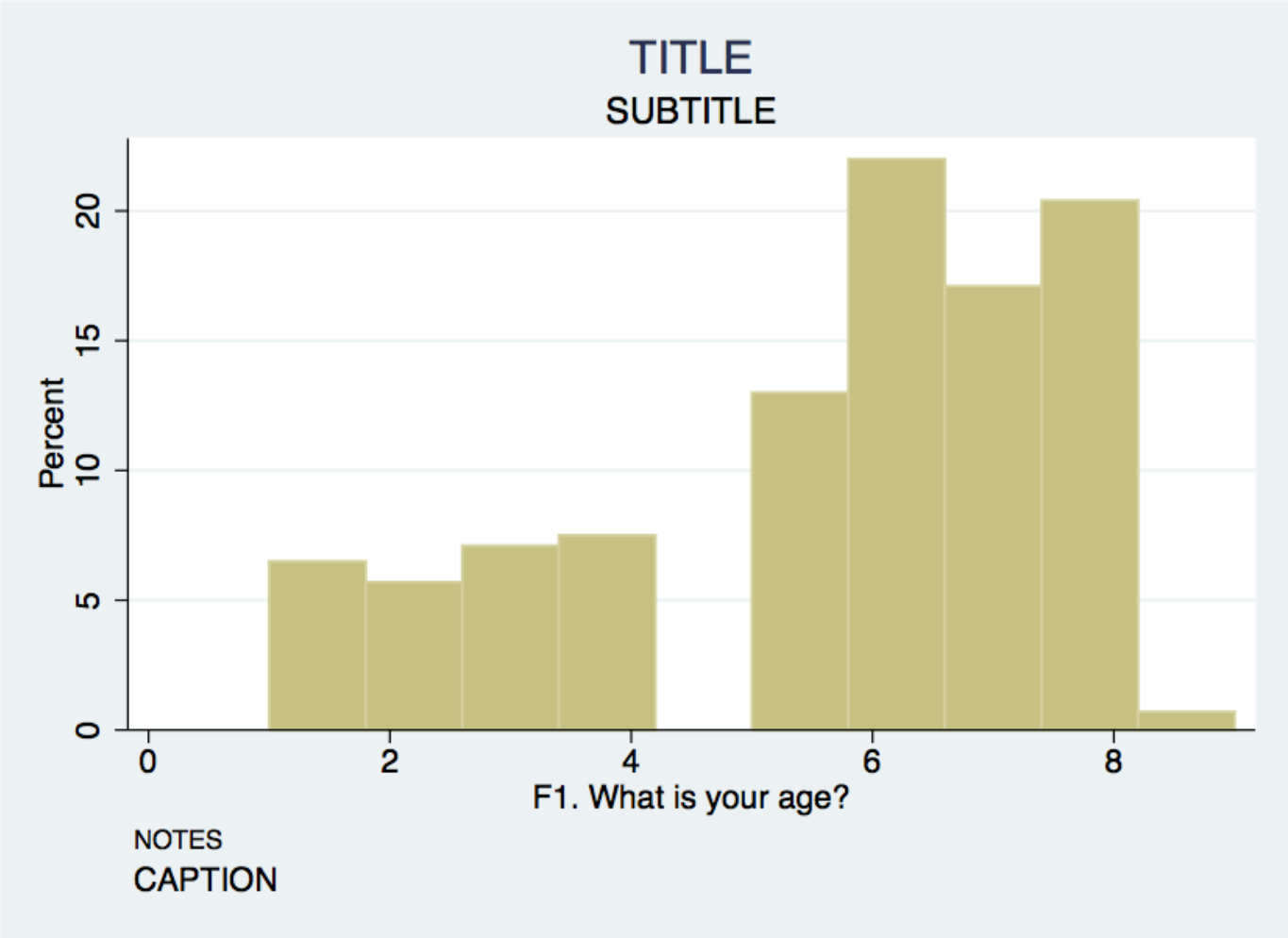
# Basic Graphing: Single Continuous Variables

Example: Histograms

- Change the numeric depiction of your data
- Add these options after the comma
  - Choose one: density fraction frequency percent
- hist varname, percent
- Be sure to properly describe your histogram:
  - `title(insert name of graph)`
  - `subtitle(insert subtitle of graph)`
  - `note(insert note to appear at bottom of graph)`
  - `caption(insert caption to appear below notes)`

# Basic Graphing: Single Continuous Variables

```
hist F1, bin(10) percent title(TITLE) ///
  subtitle(SUBTITLE) caption(CAPTION) note(NOTES)
```



# Basic Graphing: Single Continuous Variables

Example: Histograms

- Axis title options (default is variable label):
  - `xtitle(insert x axis name)`
  - `ytitle(insert y axis name)`
- Don't want axis titles?
  - `xtitle("")`
  - `ytitle("")`

# Basic Graphing: Single Continuous Variables

Example: Histograms

- Add labels to X or Y axis:
  - xlabel(insert x axis label)
  - ylabel(insert y axis label)
- Tell Stata how to scale each axis
  - xlabel(start#(increment)end#)
  - xlabel(0(5)100)
- This would label x-axis from 0-100 in increments of 5

# Basic Graphing: Single Continuous Variables

# Basic Graphing: Single Categorical Variables

Example: Histograms

- What if my variable is not continuous?
  - Simply specify "discrete" with options
- Stata will produce one bar for each level (i.e. category) of variable
- Use xlabel command to insert names of individual categories
  - . . . , xlabel(1 "White" 2 "Black" 3 "Asian" 4 "Hispanic" 5 "Other")

# Basic Graphing: Single Categorical Variables

# Exercise 1: Histograms Bar Graphs

# The Twoway Family

- Next Dataset:
  - National Neighborhood Crime Study (NNCS)
  - N=9,593 census tracts in 2000
  - Explore sources of variation in crime for communities in the United States
- Tract-level data: crime, social disorganization, disadvantage, socioeconomic inequality
- City-level data: labor market, socioeconomic inequality, population change
  - The Twoway Family
- twoway is basic Stata command for all twoway graphs
- Use twoway anytime you want to make comparisons among variables
- Can be used to combine graphs (i.e., overlay one graph with another
  - e.g., insert line of best fit over a scatter plot

# The Twoway Family

- Most basic:
  - tw scatter TPERCAP TVIOLNT
  - tw dropline TPERCAP TVIOLNT
  - tw lfitci TPERCAP TVIOLNT
  - Twoways and the By Statement

# Twoway Title Options

- Same title options as with histogram
  - title(insert name of graph)
  - subtitle(insert subtitle of graph)
  - note(insert note to appear at bottom of graph)
  - caption(insert caption to appear below notes)

# Twoway Title Options

# Twoway Title Options

# Twoway Symbol Options

# Twoway Symbol Options

# Twoway Symbol Options

# Overlaying Twoway Graphs

- Very simple to combine multiple graphs. . . just put each graph command in parentheses
  - twoway (scatter var1 var2) (lfit var1 var2)
- Add individual options to each graph within the parentheses
- Add overall graph options as usual following the comma
  - twoway (scatter var1 var2) (lfit var1 var2), options

# Overlaying Twoway Graphs

# Overlaying Twoway Graphs

# Exercise 2: The TwoWay Family

# Line Graphs

- Line graphs helpful for a variety of data
  - Especially any type of time series data
- We'll use data on US life expectancy from 1900-1999
  - webuse uslifeexp, clear

# Line Graphs

# Line Graphs

# Line Graphs: Adding Options

- As usual. . . just keep adding options after the comma!
- Same rules apply for titles that we've already seen for histograms and the twoway graphs
- Let's review how we can play with the appearance of our lines
- Full listing of options type "help line$_{\text{options}}$"

# Line Graphs: Changing Options

# Stata Graphing Lines

# Line Graphs: Changing Options

# Exporting Graphs

- From Stata, right click on image and select "save as" or try syntax:
  - graph export myfig.esp, replace
- In Microsoft Word: insert > picture > from file
  - Or, right click on graph in Stata and copy and paste into Word

# Topic

# Help Us Make This Workshop Better

- Please take a moment to fill out a very short feedback form
- These workshops exist for you–tell us what you need!
- `http://tinyurl.com/StataGraphicsFeedback`

# Additional resources

- training and consulting
  - IQSS workshops:
    `http://projects.iq.harvard.edu/rtc/filter_by/workshops`
  - IQSS statistical consulting: `http://rtc.iq.harvard.edu`
- Stata resources
  - UCLA website: `http://www.ats.ucla.edu/stat/Stata/`
  - Great for self-study
  - Links to resources
- Stata website: `http://www.stata.com/help.cgi?contents`
- Email list: `http://www.stata.com/statalist/`