

# Getting Started with Stata

## Session Three: Data Visualization

---

Yiming Cao

Department of Economics, Boston University  
Office: SSW B02  
Email: [yiming@bu.edu](mailto:yiming@bu.edu)  
Website: <http://ymcao.weebly.com>

September 24, 2019

## Last Week

- The command syntax:
  - *[prefix:] command [varlist] [if] [in] [weight] [, options]*
- Basic data management
  - Open and save data set: **use**, **save**
  - Change the names and labels of the variables: **rename**
  - Change the labels of variables: **label variable**
  - Change the labels of values: **label define** and **label values**
  - Create new variables: **generate**, **replace**
  - Delete variables or observations: **keep**, **drop**
- Advanced data management
  - A list of the most useful commands to explore

# Today: Data Visualization in Stata

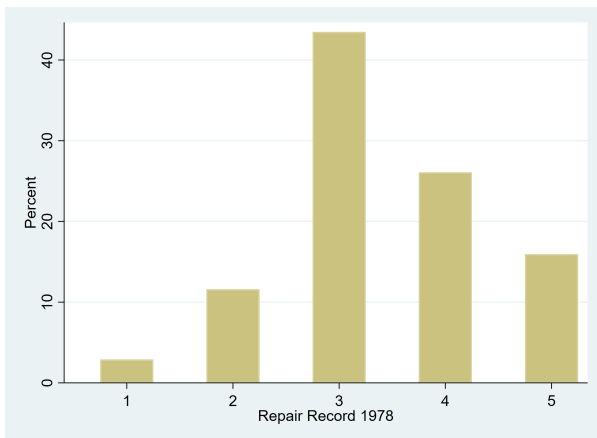
- Pros:
  - Many types provided, and multiple plot types can be overlaid
  - Very flexible
- Cons:
  - Sometimes slow
  - Large syntax (731 pages manual), code can go crazily complicated
- Getting Started
  - Plot one single variable
  - Plot the relation between two variables
  - Save and Export
  - How to become an expertise

## Overview

- Syntax: **graph plottype varlist [if] [in] [weight] [,options]**
- *plottype* can be:  
bar, dot, box, pie, histogram ...
- The following types of graphs can be used without specifying **graph**:  
histogram, kdensity, ...
- Use *options* to customize: grouping of variables, the look of the graph, legend, axis, title, ...

## Histogram graph

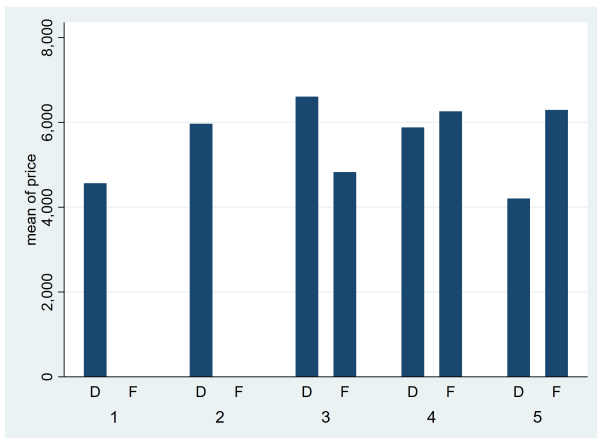
Syntax: **histogram** [varlist] [if] [in] [,options]



Follow my examples posted in the associated dofile.

## Bar Graph

Syntax: `graph bar [varname] [,options]`



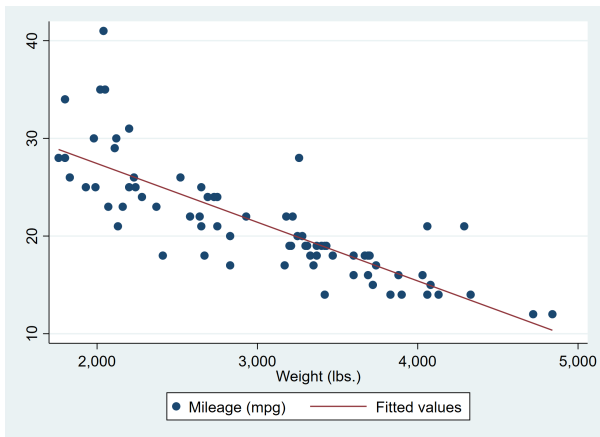
Follow my examples posted in the associated dofile.

## Overview

- Syntax:
  - Basic:  
**[graph] twoway plotype yvar1 [yvar2 [...]] xvar [if] [in] [weight] [,options]**
  - Overlay graphs:  
**[graph] twoway (plotype varlist [,options]) (plotype varlist [, options]) [, options]**  
 Or:  
**[graph] twoway plotype varlist[,options] || plotype varlist[, options] [, options]**
- *plotype* can be:  
 scatter line connected bar dot lfit qfit ...
- Use *options* to customize: *marker\_options connect\_options title\_options legend\_options axis\_options other\_options*

## Twoway Scatter

Syntax: `graph twoway scatter y_1 [y_2 [...]] x [,options]`



Follow my examples posted in the associated dofile.



## Save and Export

- To save the graph in Stata format .gph (e.g.: for future edits): use the **saving** option
- To combine a few .gph graphs: **graph combine**
- To export graphs to .jpg or .png: **graph export**

## Try Reproducing the Graphs

```

1 set more off
2 capture log close
3 log using week3.txt,replace
4 sysuse auto,clear
5 histogram mpg
6 histogram mpg, frequency
7 histogram mpg, fraction
8 histogram mpg, percent
9 histogram mpg, normal
10 histogram mpg, kdensity
11 histogram rep78, percent discrete width(.5)
12 histogram foreign, percent discrete
13 graph bar price
14 graph bar price.by(rep78)
15 graph bar price.over(rep78)
16 graph bar price.over(foreign) over(rep78)
17 label define flb 0 "D" 1 "F"
18 label values foreign flb
19 graph bar price.over(foreign) over(rep78)
20 graph bar price, over(foreign) over(rep78) bar(1,color(black))
21 twoway scatter mpg weight
22 twoway (scatter mpg weight) (lfit mpg weight)
23 twoway (scatter mpg weight,mlabel(make)) (lfit mpg weight)
24 twoway (scatter mpg weight if foreign==0) (scatter mpg weight if foreign==1)
25 twoway (scatter mpg weight if foreign==0) (scatter mpg weight if foreign==1), legend(label(1 "Domestic") label(2 "Foreign"))
26 twoway (scatter mpg weight if foreign==0,msymbol(O) mcolor(black)) (scatter mpg weight if foreign==1,msymbol(X) mcolor(black)), legend(label(1 "Domestic") label(2 "Foreign"))
27 twoway (scatter mpg weight),title("Mileage and Weight")
28 twoway scatter mpg weight,saving(g1,replace)
29 twoway scatter price weight,saving(g2,replace)
30 graph combine g1.gph g2.gph
31 graph export graph.png,width(2048) replace
32 log close

```

Or follow the dofile posted [HERE](#)

## Other types of graphs

- Type **help graph\_other** for a list of statistical graphs available
- Type **help graph\_twoway** for a list of twoway graph types available
- Get a sense of what each type of graphs looks like, and refer to the help document for the detailed syntax and options when using it

## Commonly used options

- Creating and controlling titles: `title()`, `subtitle()`, `note()` and `caption()`  
Type **help title\_options** for more
- Changing the look of markers: `msymbol()`, `msize()`, `mcolor()`  
Type **help marker\_options** for more
- Changing the look of lines: `lpattern()`, `lcolor`, `lwidth()`  
Type **help line\_options** for more
- Setting and controlling axis titles: `ytitle()`, `xtitle()`  
Type **help axis\_title\_options** for more
- Controlling legends: `legend(label(1 "TEXT") [label(2 "TEXT") [...]])`  
Type **help legend\_option** for more

# Mastering Data Visualization

- Walk through the PDF documentation *[G] Graphics* pp. 6–22 (“A quick tour”)
- Refer to the help document or *[G] Graphics* as a dictionary to look up
- **Read this article:**  
Schwabish, Jonathan A. 2014. “An Economist’s Guide to Visualizing Data.” *Journal of Economic Perspectives*, 28 (1): 209-34. [link](#)

## Next Week

- Working with dofiles
- Basic programming
- Mastering Stata