Making Graphs Using Stata

Hsueh-Sheng Wu

CFDR Workshop Series

November 15, 2021

BGSU
Center for Family and Demographic Research

Outline

- What are graphs?
- A sample graph
- Steps of using Stata to create graphs
- How to generate initial graphs?
- How to modify graphs?
- Stata examples
- Additional resources
- Conclusions

What Are Graphs?

- Graphs are pictorial representations or diagrams that display data or values in an organized manner.
- In a manuscript, graphs are used when it is difficult to use texts to concisely describe the relations between variables either because there are too many values in one variable or because more than two variables are examined.
- Different types of graphs highlight different features of the relations between variables. Thus, it is critical that researchers choose graphs that can best represent the relations between variables.
- Graphs need to contain a data field, a caption, axes, scales, and symbols. Sometimes, graphs may also have additional data fields, a title, a subtitle, and a note.

A Sample Graph

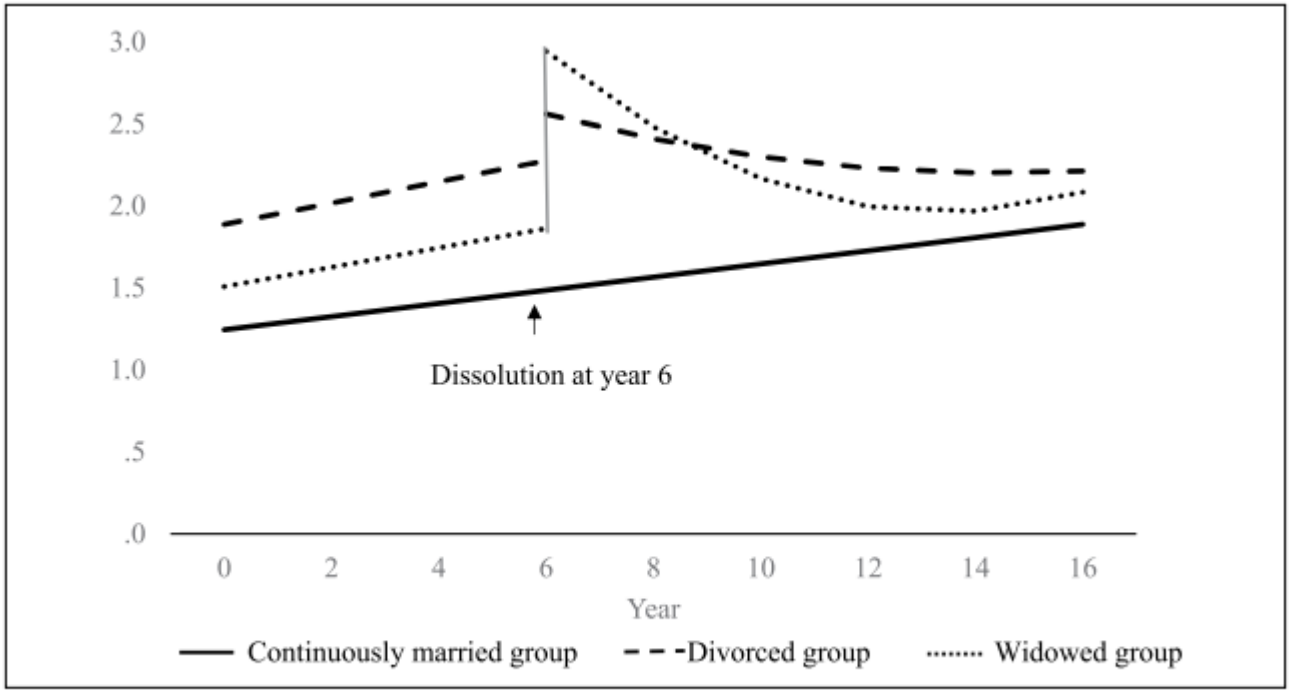
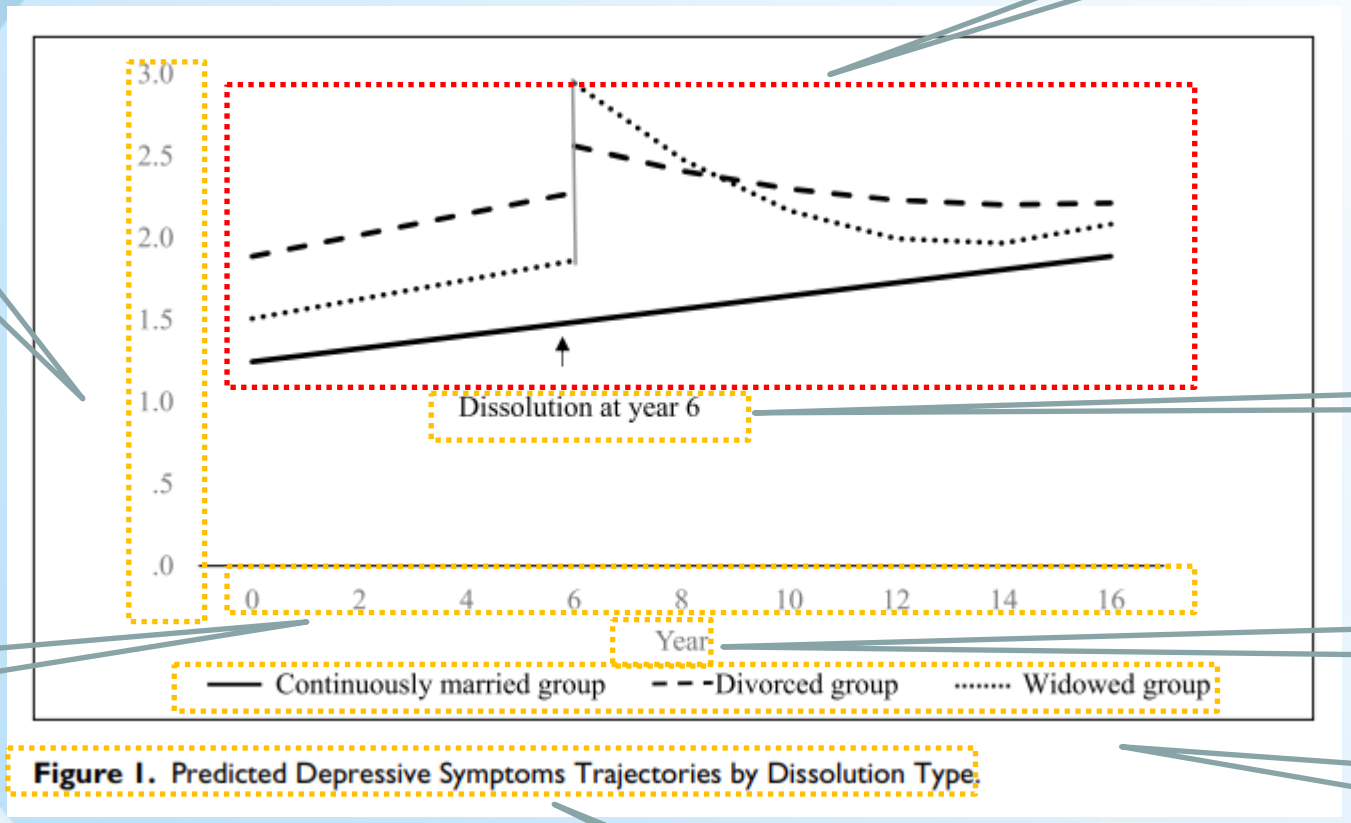


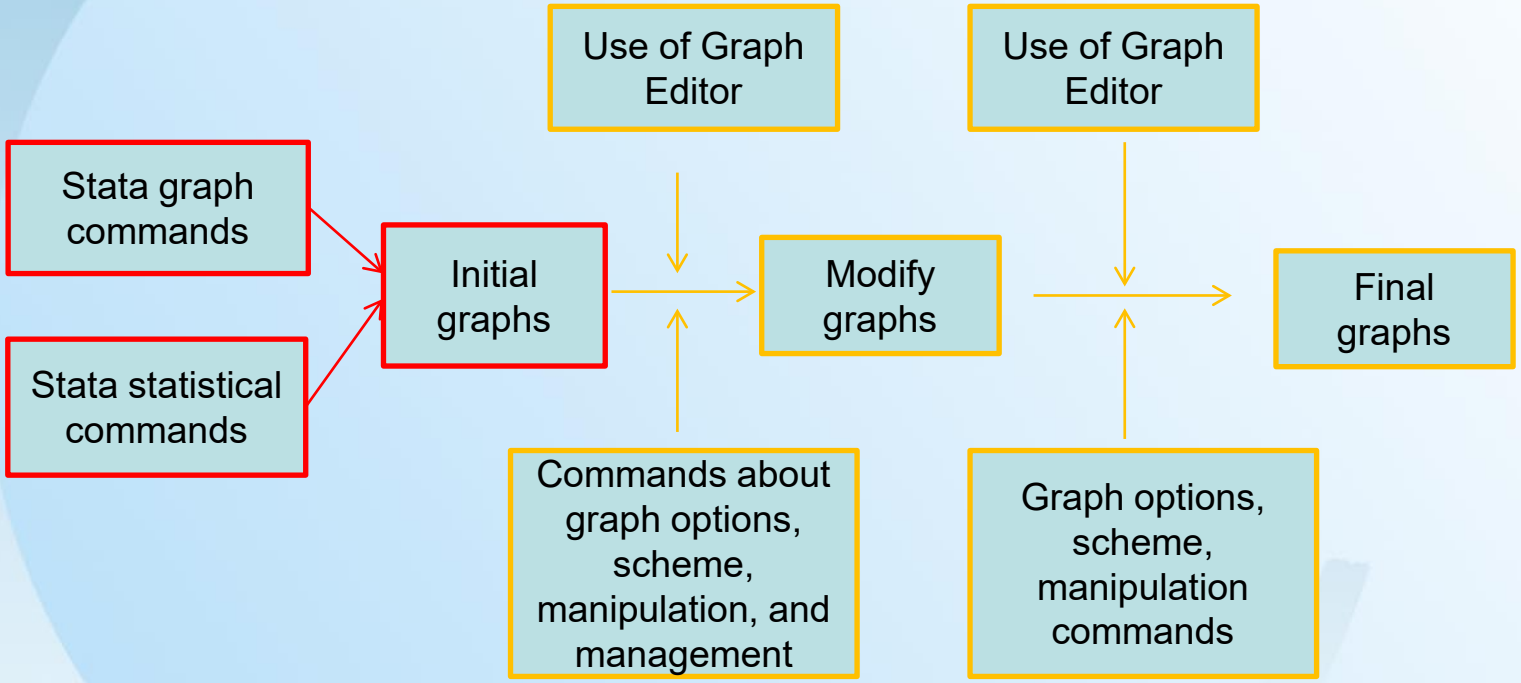
Figure 1. Predicted Depressive Symptoms Trajectories by Dissolution Type.

Lin, I-Fen, Susan L. Brown, Matthews R. Wright, and Anna M. Hammersmith. 2019. "Depressive Symptoms Following Later-Life Marital Dissolution and Subsequent Repartnering." *Journal of Health and Social Behavior* 60(2):153-168. doi:10.1177/0022146519839683

A Sample Graph (Cont.)



Steps of Using Stata to Create Graphs



BGSU Center for Family and Demographic Research

Steps of Using Stata to Create Graphs (Cont.)

- Researchers need to decide what message they want to deliver using graphs. Thus, everytime graphs are modified, they are getting closer to accurately delivering the message.
- The initial graphs can be generated via two methods:
  - Stata graph commands can generate bar charts, scatter plots, and many other different charts and plots.
  - Some Stata statistical commands also generate graphs, for example, the life table or marginplot commands.
- The initial graphs can be modified via two methods:
  - Graph editor allows researchers to modify graphs interactively.
  - Commands lines can be added to the original Stata comands to modify initial graphs.

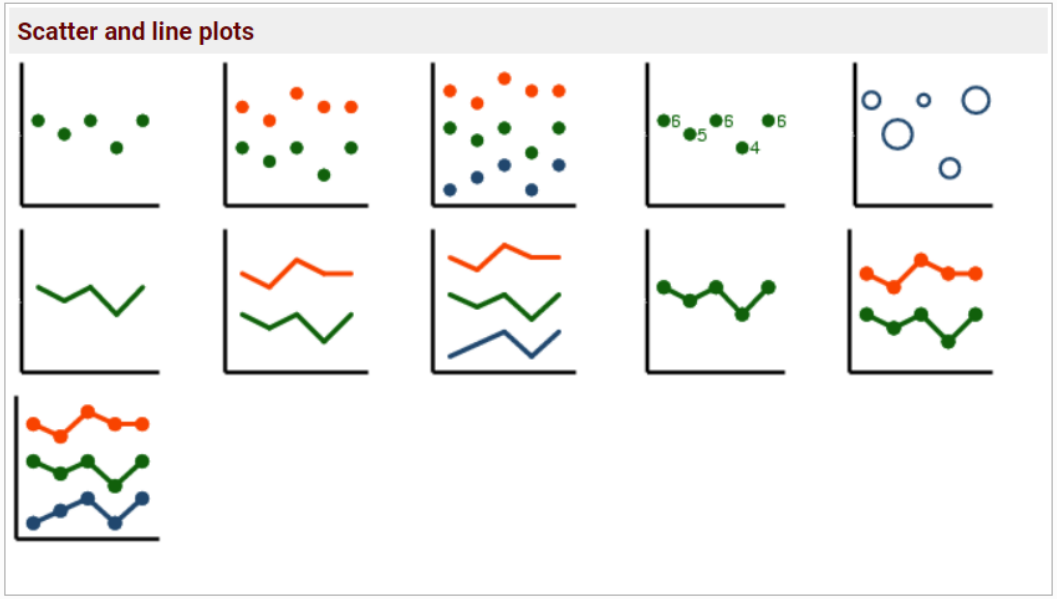
How to Generate Initial Graphs?

https://www.stata.com/support/faqs/graphics/gph/stata-graphs/

Visual overview for creating graphs

To view examples, scroll over the categories below and select the desired thumbnail on the menu at the right.

- Scatter and line plots
- Range and area plots
- Bar graphs
- Pie charts
- Dot charts
- Distribution plots
- ROC analysis
- Regression fit plots
- Survival graphs
- Time-series plots
- VAR and VEC



BGSU Center for Family and Demographic Research

How to Generate Initial Graphs (Cont.)

https://www.stata.com/features/publication-quality-graphics/

Many graph styles
- Bar charts (video tutorial)
- Box plots (video tutorial)
- Histograms (video tutorial)
- Spike plots
- Pie charts (video tutorial)
- Scatterplot matrices
- Dot charts
- Line charts
- Area charts
- Two-way scatterplots (video tutorial)
- Filled and outlined contour plots (video tutorial)
- **grmap**: Visualization of spatial data

Graphic features
- Combine graphs
- Various plotting symbols
- Various connecting line options
- Axis scaling and labeling
- Multiple graph windows
- Control color and transparency
- Control sizes of all graph elements

Watch Transparency in Stata graphs.
Watch Modifying sizes of elements in graphs.

Regression fit graphs
- Added-variable plots (partial-regression leverage plots)
- Component-plus-residual plots (partial residual plots)
- Augmented component-plus-residual plots (augmented partial-residual plots)
- Leverage-versus-squared residual plots
- Residual-versus-fitted plots
- Residual-versus-predictor plots (independent variable plots)

Distributional diagnostic plots
- Symmetry plots
- Quantile plots
- Quantile–normal plots
- Normal probability plots

Symbols and multiple fonts
- Unicode characters
- Bold and italic
- Serif and sans serif
- Monospace and proportional
- Greek letters
- Mathematical symbols
- Superscripts and subscripts

Watch Unicode in Stata.

Graph Editor
- Add
- Remove
- Move
- Modify
- Record and apply edits to other graphs

Watch Modifying graphs using the Graph Editor.

Meta-analysis
- Forest plots
- Funnel plots
- L'Abbé plots
- Bubble plots
- Galbraith plots New

Lasso
- Cross-validation function
- Path of coefficients
- BIC function New

Bayesian graphs
- Bayesian IRFs, dynamic-multiplier functions, and FEVDs New
  - IRFs, dynamic-multiplier functions, and FEVDs
  - Combine graphs
  - Overlay graphs
- Trace plots
- Autocorrelation plots
- Distributional plots

BGSU Center for Family and Demographic Research

How to Generate Initial Graphs (Cont.)

- Quantile–chi-squared plots
- Probability plots
- Quantile–quantile plots
- Cumulative-distribution plots
- Ladder-of-powers plots
- Spike plots and rootograms
- Dotplots
- Density-distribution sunflower plots

Margins plots, profile plots, and interaction plots

Smoothing
- Local polynomial
- LOWESS
- Robust nonlinear
- Kernel density

Item response theory
- Item characteristic curve plots
- Test characteristic curve plots
- Item information function plots
- Test information function plots

Multivariate graphs
- Biplots
- Dendrograms
- Shepard diagrams
- Configuration plots
- Correspondence analysis projection plots
- Scree plots
- Score and loading plots
- Procrustes overlay plot

Nonparametric regression conditional mean plots

Quality control
- c charts
- p charts
- R charts
- X-bar charts
- Shewhart charts
- Standard-error bar charts

Output formats
- TIFF
- PNG
- SVG
- PDF
- PostScript
- Encapsulated PostScript (EPS)
- Encapsulated PostScript (EPS) with TIFF preview
- Windows Metafile
- Windows Enhanced Metafile
- Export graphs to Excel files

Survival plots
- Kaplan–Meier survival curves
- Nelson–Aalen cumulative hazard curves
- Parametric fitted survival curves
- Proportional hazard diagnostic curves

Time-series graphs
- Correlograms
- Periodograms
- Cumulative sample spectral density
- Line plots
- Range plots with lines
- IRFS, dynamic-multiplier functions, and FEVDs

Panel-data line plots
- Graphs by panel
- Overlaid panels

Power, precision, and sample size
- Automated and customizable
- Comparative — multiple lines, subgraphs, or graphs

Treatment effects
- Overlap plots
- Box plots for covariate balance
- Density plots for covariate balance

ROC analysis
- ROC curve of classifier with binormal fit
- Marginal and covariate-adjusted ROC curves
- Sensitivity and specificity versus probability cutoff

Additional resources
- *Graphics Reference Manual*
- *A Visual Guide to Stata Graphics, 3rd Edition* by Michael N. Mitchell
- *Speaking Stata Graphics* by Nicholas J. Cox
- In the spotlight: marginsplot
- The Stata Blog: Scheming your way to your favorite graph style
- NetCourse 120: Statistical Graphics Using Stata
- Using Stata Effectively: Data Management, Analysis, and Graphics Fundamentals training course

BGSU Center for Family and Demographic Research

How to Modify Graphs

- Graphs can be modfied using command lines.
- This method keeps a record of how graphs are created.
- Works best if researchers already know what modifications to the graph should be made

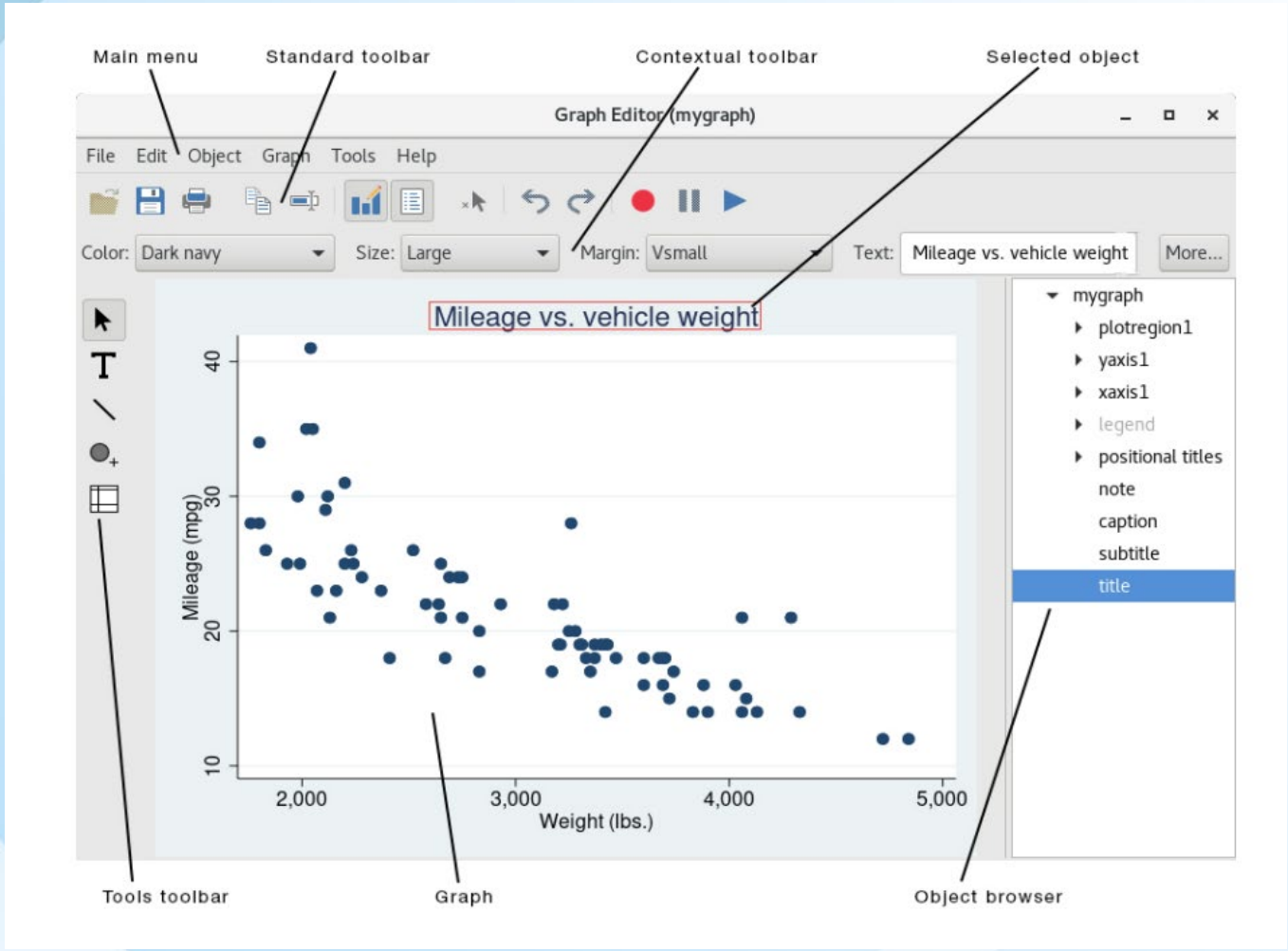
- Modify graphs generated by Stata graph commands
  - Stata graphics reference manual release 17 (https://www.stata.com/bookstore/graphics-reference-manual/)
  - The Cheatsheet of Data Visualization with Stata, created by Laura Hughes and Tim Essam and available at: (https://www.stata.com/bookstore/statacheatsheets.pdf)

- Modify graphs generated by other Stata commands
  - Some options in Stata graph commands can be applied to graphs created by some statistical commands, but not others.

```
ltable timevar [deadvar] [if] [in] [weight] [, options]
webuse selvin
ltable t died [freq=pop], graph
ltable t died [freq=pop], graph title("Life Table") subtitle("Using Stata data")
```

How to Modify Graphs (Cont.)

Graphs can be modified using Graph Editor



Stata Examples

```
*********************************
* Bar Chart
*********************************


use https://www.stata-press.com/data/r17/nlsw88, clear

sum
tab2 wage race
sort race
by race: sum wage, detail

* You can plot mean, median, percentile, sum, count, percent, minimum value, and maximum value

graph bar (mean) wage, over(race)
graph bar (median) wage, over(race)

* The difference between the -by- and -over-option

graph bar wage , by(race)
graph bar wage, over(race)
graph bar wage, over(race) over(collgrad)
graph bar wage, over(race) over(collgrad) over(union)


* rotate the variables using the Graph Editor


graph bar wage, over(race) over(collgrad) over(union)
* save the recording and play it back on two graphs
graph bar wage, over(race) over(collgrad) over(union)
graph bar wage, over(collgrad) over(race) over(union)
```

Stata Examples (Cont.)

```
********************
* Combine plots
********************

use https://www.stata-press.com/data/r17/uslifeexp, clear

line le_male year, saving("D:\jason\workshop\Stata graph\male.gph", replace)
line le_female year if inrange(year, 1900,1950), saving("D:\jason\workshop\Stata graph\female.gph", replace)

* Each graph uses its own x-axis

graph combine "D:\jason\workshop\Stata graph\male.gph"  "D:\jason\workshop\Stata graph\female.gph", cols(1)  ///
saving("D:\jason\workshop\Stata graph\combine100.gph", replace)

* Make both graphs use the same x-axis

graph combine "D:\jason\workshop\Stata graph\male.gph"  "D:\jason\workshop\Stata graph\female.gph", cols(1)  ///
xcommon iscale(.5) saving("D:\jason\workshop\Stata graph\combine50.gph", replace)
```

Stata Examples (Cont.)

```
********************
* Scatter plots
********************

use https://www.stata-press.com/data/r17/sp500, clear

replace volume = volume/1000

twoway rspike hi low date ||                                        ///
line close date ||                                                  ///
bar volume date, barw(.25) yaxis(2) ||                              ///
in 1/57                                                             ///
, yscale(axis(1) r(900 1400))                                       ///
yscale(axis(2) r( 9 45))                                            ///
ytitle(" Price -- High, Low, Close")                                ///
ytitle(" Volume (millions)", axis(2) bexpand just(left)) ///
legend(off)                                                         ///
subtitle("S&P 500", margin(b+2.5))                                  ///
note("Source: Yahoo!Finance and Commodity Systems, Inc.")
```

Additional Resources

- Video turorials
  - https://www.stata.com/features/publication-quality-graphics/
- Visual overview for creating graphs
  - https://www.stata.com/support/faqs/graphics/gph/stata-graphs/
- Stata cheatsheet
  -https://www.stata.com/bookstore/stata-cheat-sheets/
- Stata graphics reference manual release 17
  - https://www.stata.com/bookstore/graphics-reference-manual/
- Michael N. Mitchell (2012) A Visual Guide to Stata Graphics, Third Edition, College Station, TX : Stata Press

Conclusions

- Graphs are visual presentationof data or the relations between variables.
- Graphs have two parts: data field(s) and non-data objects. The data field contains plots or charts, which are created by Stata -graph- command or the graph option. The non-data obects can be modified via command lines or graph editor.
- Different graphs draw attention to different aspects of relations between variables. Thus, it is important to browser through different graphs and then choose graphs that can best represent ideas manifested by the relations between variables.
- Given different types of graphs Stata can generate, it is not realistic for researchers to master all of them. Instead, researchers should focus on the graphs they will use and then learn every details of these graphs.
- The command line and graph editor have different strengths and weaknesses. The use of command lines keeps the record of how graphs are crated but requires users to know exactly what codes are needed. In contrast, the use of graph editor modifies graphs by moving objects without typing codes but does not keep tract of how graphs are created.
- For further question, feel free to contact me at wuh@bgsu.edu or stop by my office (5D Williams Hall).