The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, Stata, Stata Press, Mata, Mata, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2014)
14, Number 1, pp. 221–225

Stata tip 117: graph combine—Combining graphs

Lars Ängquist
Institute of Preventive Medicine
Bispebjerg and Frederiksberg Hospitals—The Capital Region
Copenhagen, Denmark
lars.henrik.angquist@regionh.dk

1 Introduction

There are many different reasons for wanting to create multipanel graphs, presented in $r \geq 1$ rows and $c \geq 1$ columns: these reasons include making efficient use of restricted display space and enhancing the presentation of results. In basic Stata, the flexible approach to confidently handle these tasks is given by using the `graph combine` functionality (see `help graph combine`). For related discussions and examples, see the stimulating books *An Introduction to Stata for Health Researchers* (Juul and Frydenberg 2010) and *A Visual Guide to Stata Graphics* (Mitchell 2012).

2 Basic usage

First, we start with setting up seven simple, but quite artificial, linear relations disturbed by normally distributed noise based on simulated x and y variables (interpreted in the standard sense).

```
set obs 100
generate xvar=10*runiform()
forvalues i=1/7 {
   generate y`i´=xvar*`i´+runiform()*(`i´*3)
   label variable y`i´ "Outcome variable `i´"
}
```

Second, we simply fit linear regressions that correspond to these relations and then save the seven corresponding graphs in memory.

```
foreach yvar of varlist y* {
   local lbl: variable label `yvar´
   sort xvar
   reg `yvar´ xvar
   local b : display %3.2f _b[xvar]
   predict p, xb
   twoway (scatter `yvar´ xvar) (line p xvar),              ///
      ytitle("`lbl´") xtitle("Explanatory covariate")       ///
      yscale(range(0 80))                                   ///
      legend(off) note("{&beta}=`b´", position(4) ring(0))  ///
      name("graph_`yvar´", replace)
   drop p
}
```


(Here we use the `name(`*string*`)` option—unless we want to actually save the separate graphs to disk. In that case, we would replace this option with `saving(`*string*`)`.)

Finally, we intend to combine the graphs into a multipanel setup. Assuming that the graphs belong to two distinct groups—graphs 1–3 and 4–7, respectively—they are mirrored in the construction. This is achieved by the following:

1. Combine graphs 1–3 into panel 1.
2. Combine graphs 4–7 into panel 2.
3. Combine the resulting 1-row panels, panel 1 ($r \times c = 1 \times 3$) and panel 2 ($r \times c = 1 \times 4$), into a final 2-row panel ($r = 2$; see figure 1).

```
graph combine graph_y1 graph_y2 graph_y3,                                    ///
   name("firstset", replace) ycommon cols(3) title("First set of graphs")
graph combine graph_y4 graph_y5 graph_y6 graph_y7,                           ///
   name("secondset", replace) ycommon cols(4) title("Second set of graphs")
graph combine firstset secondset,                                            ///
   saving("sevenpanelgraph.gph", replace) ycommon cols(1)
graph export sevenpanelgraph.eps, replace
```

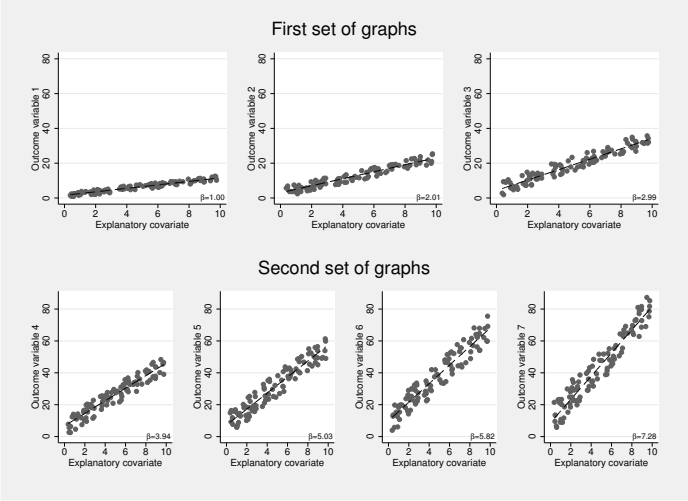

Figure 1. Multipanel graph—a combination of combined graphs

3 Some notes on options

The basic functionality facilitates an easy-to-use combination of graphs. A well-suited set of selected options might improve the display.

3.1 Axes

In many cases, keeping scales constant over panels might enhance the interpretability of the jointly graphed relations. Generally, this might prove to be a valid argument; however, it is imperative for the x axis and y axis when comparing vertically (the `xcommon` option) and horizontally (the `ycommon` option), respectively.

3.2 Margins

To keep the panels as tightly linked as possible—to increase overall comparability—it might be suitable to reduce margins through `imargin(zero)`; for other margin choices, see `help marginstyle`.

3.3 Panel pattern

The final number of panels to use is implicitly given by the stated list of panels in the actual program call. (Remember that each panel might in itself be a previously constructed multipanel. In the above example, a single column, $c = 1$, was used at the combination stage.) To define which $r \times c$ panel-matrix shape will be used, one may choose any of the following options (one is enough): `rows(`*integer*`)` or `cols(`*integer*`)`. To make the graph (distribution of panels) unique, select the `colfirst` option (or not). If the required number of panels is less than the available number $r{\cdot}c$, it may be useful to explicitly—given the unique order—tell Stata which panels should be left empty (instead of the default) by using `holes(`*numlist*`)`.

3.4 Scaling

Each panel is downscaled when using multipanels, text and markers, etc. It is possible to rescale the downscaling through the `iscale(`*scale*`)` option, where *scale* is either an absolute (positive) or a relative value. For example, the absolute value `1` means the original size, and the relative value `*1` implies the same size as the default selection; `0.75` and `*0.75` will adjust the size to the three-quarter size counterparts.

4 A second example

For our second example, we will play around with the individual panel sizes. For this, we will use one of the seven graphs (the sixth) from figure 1, which is inspired by the informative help file (see the end of the `help graph combine` post), to complement it with the two corresponding underlying histograms (see result in figure 2).

```
histogram xvar,                                     ///
    percent start(0) width(1)                       ///
    xscale(range(0 10) off)                         ///
    fxsize(100) fysize(25)                          ///
    yscale(range(0 15)) ytitle("")                  ///
    ylabel(0(5)15, angle(horizontal))               ///
    kdensity kdenopts(lpattern(dash))               ///
    plotregion(margin(zero))                        ///
    note("N (%)", ring(0) position(10))             ///
    name("hist_xvar", replace)
histogram y6,                                       ///
    percent start(0) width(10) horizontal           ///
    xtitle("") xlabel(0(10)20) xscale(rev)          ///
    fxsize(25) fysize(100)                          ///
    yscale(range(0 80) off)                         ///
    ylabel(10(20)70, angle(horizontal))             ///
    kdensity kdenopts(lpattern(dash))               ///
    plotregion(margin(zero))                        ///
    note("N (%)", ring(0) position(4))              ///
    name("hist_y6", replace)
```

In the next step, these three panels are combined (note that we use some of the options just discussed). The main point here is that the options `fxsize(`*number*`)` and `fysize(`*number*`)` govern the widths and heights of the panels; that is, in the example above, the thin sides are left at 25% of the original sizes.

```
graph combine hist_y6 graph_y6 hist_xvar,               ///
    holes(3) rows(2)                                    ///
    imargin(0 2 0 0)                                    ///
    title("  Twoway graph with histograms", ring(0))    ///
    saving(graphwithhistograms.gph, replace)
graph export graphwithhistograms.eps, replace
```

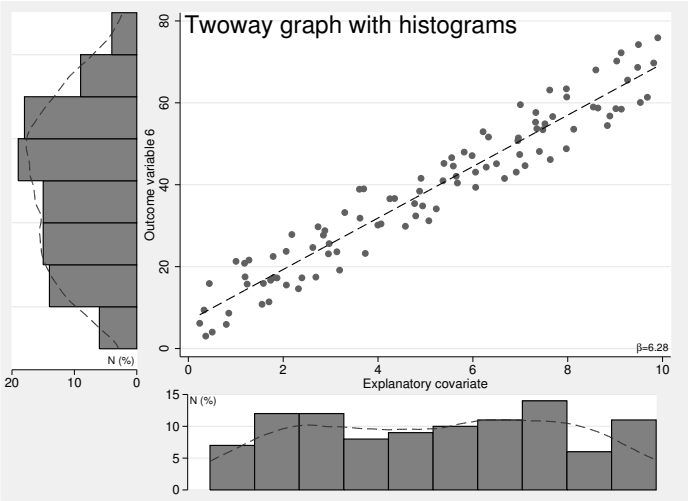

Figure 2. Multipanel graph—a scatterplot with a prediction line and two complementary histograms

5 Discussion and alternatives

In many situations where the subgraphs combine corresponding true data subsets of the present loaded data, a similarly performing alternative would be to use the `by()` option (see `help by_option`). Here the syntax `by(`*varlist*`[ , `*options*`])` allows combined graphing of the corresponding defined graph with respect to all present categories specified by the categorical variables given in *varlist*. In this setting, the options `total` and `missing` add panels based on the total dataset (over nonmissing groups) and missing data for individuals, respectively.

5.1 by() options

As noted above, the option `by()` allows for suboptions. Some suboptions are similar to the ones available for `graph combine`—for example, `colfirst`, `cols()`, `rows()`, `holes()`, `iscale()`, and `imargin()`. Similar functionality, but with different names and adapted settings, is given by `compact` (reduces margins between panels), `norescale` (uses the same scales over panels), and `noedgelabel` (restricts the number of displayed labels). Note that an option with `no`, such as `norescale`, generally has a counterpart, such as `rescale`, with a quite obvious implication.

Usually, this type of solution might be convenient in different cases; however, in most situations, this solution is less flexible and more restrictive by nature. Furthermore, graphing several subgroups within a single panel (together but separately marked) is an alternative solution that allows the smaller number of subgroups to be totally displayed while applying distinct colors and markers. For other cases, the multipanel design may be the best choice because one (or several) background groups can be added to each panel to enhance overall comparability. For example, see the discussion of overlaid graphs in Cox (2010), where subgroups are plotted against completely complementary data while using discrete gray-scaled backdrop markers for the background group. This is referred to as adopting a substrate, or subset, graphing design.

References

Cox, N. J. 2010. Speaking Stata: Graphing subsets. *Stata Journal* 10: 670–681.

Juul, S., and M. Frydenberg. 2010. *An Introduction to Stata for Health Researchers*. 3rd ed. College Station, TX: Stata Press.

Mitchell, M. N. 2012. *A Visual Guide to Stata Graphics*. 3rd ed. College Station, TX: Stata Press.