

Stata tip: Custom legends for graphs that use translucency

Tim P Morris
MRC Clinical Trials Unit at UCL
London, UK
tim.morris@ucl.ac.uk

As of version 15, Stata graphs permit translucent elements, invoked by “%#” following a *colorstyle* (where # is a percentage of opaque). Many `twoway` commands will include a legend. This gives control over the labels, but not over the symbols.

I will now demonstrate a trick to produce legend symbols ‘by hand’ for plot types `scatter` and then `line`, and why the arrival of translucency has prompted the need for this. Note that this tip is not well showcased by the `sj` graph scheme. Hence I advise readers to try running the examples using a scheme including colors.

First, we will simulate a dataset with two groups each containing 1,000 observations per group on `y` and `x` (note that `drawnorm` sets the dataset size from empty).

```
. clear
. matrix means = (4,1)
. matrix corr = (1, .5 \ .5, 1)
. drawnorm y1 x1 , n(1000) seed(1)
. drawnorm y2 x2 , corr(corr) means(means)
```

I begin with a scatter plot:

```
twoway
    (scatter y1 x1, mc(%30) mlc(white%1))
    (scatter y2 x2, mc(%30) mlc(white%1))
    ,
    legend(order(1 "Group 1 (default symbol)" 2 "Group 2 (default
symbol)"))
```

The scatter plot above includes translucent symbols with (30% of opaque). The translucency inherited by the symbols in the legend is not necessarily desirable as it matches the palest possible shade seen within the plot, which may make it harder to match the legend to the plotted data. To produce a legend symbol of our own choosing, we can add what I term a ‘ghost’ plot for each symbol we wish to control. This is achieved as follows:

```
. twoway (scatteri . . , msize(medlarge))
> (scatteri . . , msize(medlarge))
> (scatter y0 x0, pstyle(p1) mc(%30) mlc(white%1))
> (scatter y1 x1, pstyle(p2) mc(%30) mlc(white%1))
> , legend(order(1 " Group 1 (custom symbol)" 2 " Group 2 (custom
symbol)"))
```

Figure 1 shows the plot with the first and second versions of the legend, placed in one graph here to emphasize the difference: the top row uses the default translucent symbols and the bottom row uses the opaque custom symbols used by the ghost plots.

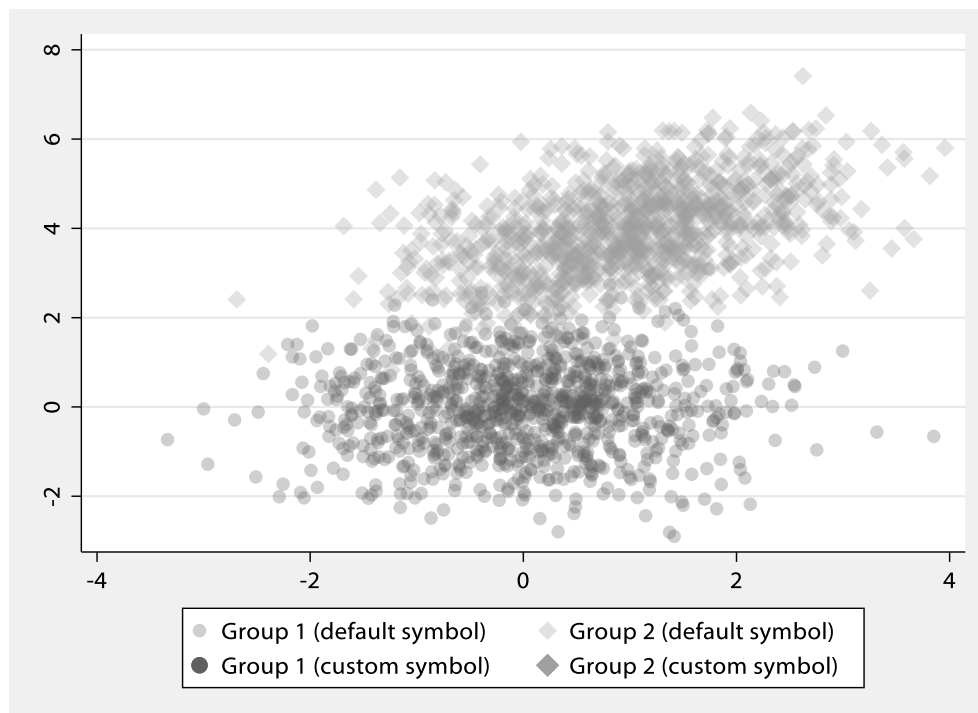


Figure 1. Scatter plot of simulated data with the default (upper row) and custom (lower row) legend symbols.

The ghost plots here were invoked using `scatteri`, and the points `. .` ensured nothing was plotted – but the symbol attributes nonetheless appeared in the legend. Notice that the option `legend(order(1 "... " 2 "..."))` now refers to our ghost plots (since they appear first and second in the `twoway` command). Our original plots are now in positions 3 and 4, meaning that the symbols they use would have changed. This is remedied by the option `pstyle(p1)` for plot #3 and `pstyle(p2)` for plot #4.

To demonstrate this method using a line plot type, we first (twice) `reshape` the simulated data:

```
. generate int id = _n
. reshape long y x, i(id) j(group)
. rename (y y1) (x y0)
. reshape long y, i(id group) j(time)
```

The following `twoway` command then produces a legend with user-manipulated symbols:

```
. twoway
> (function . , lc(gs10) lw(medthick))
> (function . , lc(gs6) lw(medthick))
> (line y time if group==1, pstyle(p1line) lc(gs10%5))
> (line y time if group==2, pstyle(p2line) lc(gs6%5))
> , legend(order(1 "Group 1 (custom symbol)" 2 "Group 2 (custom
symbol)") cols(1))
```

Most of this works in the same way as with `scatter`. The 'ghost' plot type is `function` and again we use it to plot nothing (`.` denotes missing). The *pstyles* used by plots 3 and 4 are now `p1line` and `p2line` respectively.

Using this approach, any options required for the legend can be achieved by manipulating options within the ghost plots, and this manipulation does not have to affect the 'living' plots.

The trick outlined in this tip has uses beyond translucency. In presentations it is sometimes desirable to add elements gradually to build up a graph. For example, the legend that comes with `msymbol(p)` is almost never readable even when the color of a swarm of points is. Another use is in presentations. One may begin with the graph region with no data but a legend, then introduce the data for group 1 and subsequently group 2; using this trick, the legend could be produced to be present from the beginning.