The Stata Journal (2008)
8, Number 2, pp. 295–298

Stata tip 62: Plotting on reversed scales

Nicholas J. Cox and Natasha L. M. Barlow
Durham University
Durham City, UK
n.j.cox@durham.ac.uk and n.l.m.barlow@durham.ac.uk

Stata has long had options allowing a reversed scale on either the y or the x axis of many of its graph types. Many graph users perhaps never even consider specifying such options. Those who do need to reach for them may wish to see detailed examples of how reversed scales may be exploited to good effect.

The usual Cartesian conventions are that vertical or y scales increase upward, from bottom to top, and horizontal or x scales increase from left to right. Vertical scales increasing downward are needed for graphs with vertical categorical axes following a table-like convention, in which the first (lowest) category is at the top of a graph, just as it would be at the top of a table. Commands such as `graph bar` and `graph dot` follow this convention. Indeed, it is likely to be so familiar that you may have to reflect briefly to see that is how such graphs are drawn. Other examples of this principle are discussed elsewhere in this issue (Cox 2008).

Reversed scales are also common in the Earth and environmental sciences. Here, in fields such as pedology, sedimentology, geomorphology, limnology, and oceanography, it is common to take measurements at varying depths within soils, sediments, rocks, and water bodies. Even though depth is not a response variable, it is conventional and convenient to plot depth below surface on the vertical axis; hence, the need for a reversed scale. An extra twist that gives spin to the graph problem is that frequently detailed analyses of materials at each level in a core, bore, or vertical profile yield several response variables, which are all to be plotted on the horizontal axis. Such multiple plotting is easiest when overlay is possible.

Let us look at some specific syntax for an example and then add comments. The data shown here come from work in progress by the second author. Sediment samples at 2 cm intervals down a core from Girdwood, Alaska, were examined for several elements associated with placer mining pollution (LaPerriere, Wagener, and Bjerklie 1985). In this example, the concentrations of gold, cadmium, arsenic, lead, copper, and zinc, measured in parts per million (ppm), are then plotted as a function of depth (cm). See figure 1.

(*Continued on next page*)

 gr0035

```
. local spec clwidth(medium) msize(*0.8)
. twoway
> connect depth Au, `spec´ ms(Oh) cmissing(n) ||
> connect depth Cd, `spec´ ms(Th)  ||
> connect depth As, `spec´ ms(Sh)  ||
> connect depth Pb, `spec´ ms(O)   ||
> connect depth Cu, `spec´ ms(T)   ||
> connect depth Zn, `spec´ ms(S)
> yscale(reverse)  xscale(log) xscale(titlegap(*10))
> ylabel(50(10)90, angle(h)) xlabel(0.1 0.3 1 3 10 30 100)
> ytitle(Depth (cm)) xtitle(Concentration (ppm))
> legend(order(1 "Au" 2 "Cd" 3 "As" 4 "Pb" 5 "Cu" 6 "Zn") position(3) column(1))
```

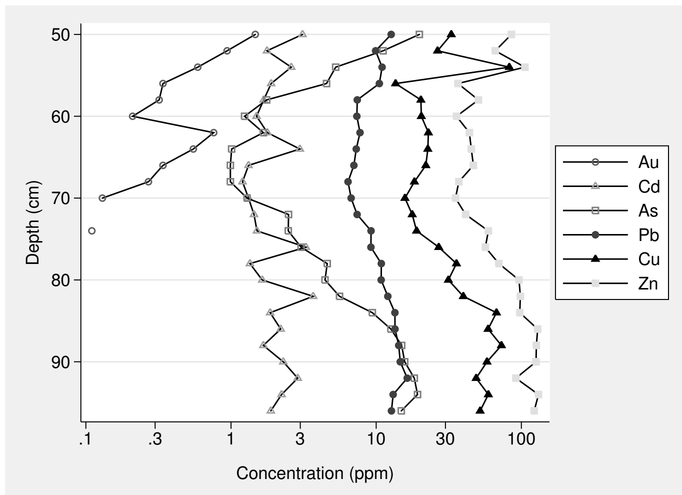

Figure 1. Variation in concentrations of various elements with depth, measured at a site in Girdwood, Alaska.

A key detail here is that Stata's model for scatter and similar plots is asymmetric. One or more y variables are allowed, but only one x variable, in each individual plot. Thus, if you wish to have several x variables, you must either superimpose several variables, as here, or juxtapose several plots horizontally.

Further, it is necessary to spell out what is desired as text in the legend. By default, `twoway` would use information for the y axis variables, which is not what is wanted for this kind of graph.

For these data a logarithmic scale is helpful, indeed essential, for showing several elements that vary greatly in abundance. Gold is reported as less than 0.1 ppm at depth. Such censored values cannot be shown by point symbols.

Even very experienced Stata users will not usually think up an entire graph command like this at the outset. Typically, you start with a fairly simple design and then elaborate it by a series of very small changes (which may well be changes of mind back and forth).

At some point, you are likely to find yourself transferring from the Command window to the Do-file Editor and from an interactive session to a do-file. We also find it helpful, once out of the Command window, to space out a command so that its elements are easier to see and so that it is easier to edit. Spending a few moments doing that saves some fiddly work later on. What is shown above is in fact more compressed than what typically appears within the Do-file Editor in our sessions.

The legend used here is, like all legends, at best a necessary evil, as it obliges careful readers to scan back and forth repeatedly to see what is what. The several vertical traces are fairly distinct, so one possibility is to extend the vertical axis and insert legend text at the top of the graph. `‘=Au[1]’`, for example, instructs Stata to evaluate the first value of `Au` and use its value. The same effect would be achieved by typing in the actual value. Here we are exploiting the shortness of the element names, which remain informative to any reader with minimal chemical knowledge. The legend itself can then be suppressed. The text elements could also be repeated at the bottom of the graph if desired. See figure 2.

```
. twoway
> connect depth Au, `spec´ ms(Oh) cmissing(n) ||
> connect depth Cd, `spec´ ms(Th)  ||
> connect depth As, `spec´ ms(Sh)  ||
> connect depth Pb, `spec´ ms(O)   ||
> connect depth Cu, `spec´ ms(T)   ||
> connect depth Zn, `spec´ ms(S)
> yscale(reverse r(46 .)) xscale(log) xscale(titlegap(*10))
> ylabel(50(10)90, angle(h)) xlabel(0.1 0.3 1 3 10 30 100)
> ytitle(Depth (cm)) xtitle(Concentration (ppm))
> text(48 `=Au[1]´ "Au" 48 `=Cd[1]´ "Cd" 48 `=As[1]´ "As" 48 `=Pb[1]´ "Pb"
>      48 `=Cu[1]´ "Cu" 48 `=Zn[1]´ "Zn")
> legend(off)
```

(*Continued on next page*)

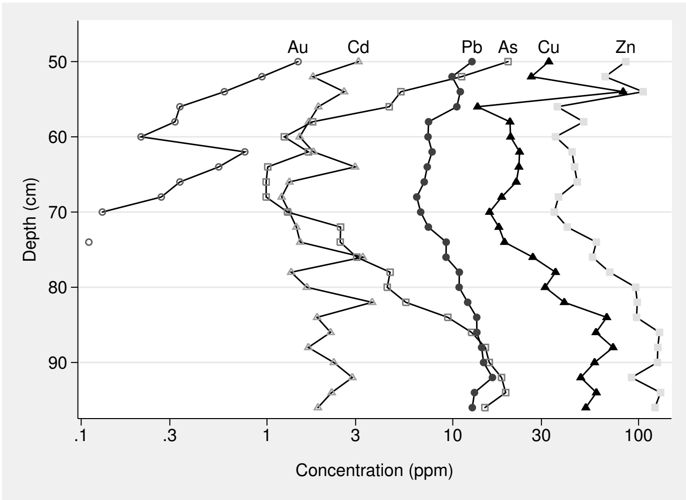

Figure 2. Variation in concentrations of various elements with depth, measured at a site in Girdwood, Alaska. The legend has been suppressed and its text elements placed within the graph.

The structure of overlays may suggest that a program be written that takes one y and several x variables and then works out the overlay code for you. Such a program might then be available particularly to colleagues and students less familiar with Stata. That idea is more tempting in this example than in general. In other cases, there might be a desire to mix different `twoway` types, to use different marker colors, or to make any number of other changes to distinguish the different variables. Any code flexible enough to specify any or all of that would lead to commands no easier to write than what is already possible. In computing as in cookery, sometimes you just keep old recipes once you have worked them out, at least as starting points for some later problem.

In the Earth and environmental sciences, reversed horizontal scales are also common for showing time, whenever the units are not calendar time, measured forward, but time before the present, measured backward. Within Stata, such practice typically requires no more than specifying `xscale(reverse)`.

References

Cox, N. J. 2008. Speaking Stata: Between tables and graphs. *Stata Journal* 8: 269–289.

LaPerriere, J. D., S. M. Wagener, and D. M. Bjerklie. 1985. Gold-mining effects on heavy metals in streams, Circle Quadrangle, Alaska. *Journal of the American Water Resources Association* 21: 245–252.