The Stata Journal (2012)
12, Number 2, pp. 332–341

Speaking Stata: Transforming the time axis

Nicholas J. Cox
Department of Geography
Durham University
Durham City, UK
n.j.cox@durham.ac.uk

Abstract. The time variable is most commonly plotted precisely as recorded in graphs showing change over time. However, if the most interesting part of the graph is very crowded, then transforming the time axis to give that part more space is worth consideration. In this column, I discuss logarithmic scales, square root scales, and scale breaks as possible solutions.

Keywords: gr0052, axis scale, graphics, logarithm, scale break, square root, time, transformation

1 Introduction

Using transformed scales on graphs is a technique familiar to data analysts. Most commonly in Stata, axis scales on a `twoway` graph may be logarithmic. The options `yscale(log)` and `xscale(log)` provide wired-in support. In principle, other transformed scales are hardly more difficult: just calculate the transformation before graphing and, ideally, ensure that the axis labels remain easy to interpret. For example, a logit scale can be useful for proportions within $(0, 1)$ or percentages within $(0, 100)$. For that example and more general comments, see Cox (2008).

In the case of graphs plotting change over time, the time axis usually shows the time variable unchanged. In most sciences, time is plotted on the horizontal or x axis. In the earth sciences, time is sometimes plotted on the vertical or y axis, essentially because older materials are often found at greater depths.

In either case, however, it is common to use a transformed scale for whatever response is plotted versus time. For example, with a logarithmic response scale, periods of exponential growth or decline would be shown by straight line segments.

In this column, I focus on something more unusual: transforming the time axis. The argument is in the first instance pragmatic. Sometimes data span an extended time range, and the amount of information or the pace of change is far greater at one end of the time range. If so, expanding the crowded part of the graph and shrinking the other part of the graph may help readers visualize the detail of change.

If data come from cosmology or particle physics, blowing up the early part of a time range might be a good idea. If data come from the social sciences, especially economic or demographic history, blowing up the later part might be called for. In many examples, very recent growth has been explosive, but the precise details remain of interest. The same recommendation applies, but for different reasons, to many datasets from the earth sciences. Here information from earlier periods tends to be much sparser and indeed in most instances has long since been buried, blasted, or eroded away.

2 Logarithmic scale?

A logarithmic scale for time would show earlier times—all later than some origin, itself conventionally zero—more expansively. `xscale(log)` should suffice, assuming again that time is shown on the x axis. Conversely, a logarithmic scale for times, all those shown being *before* some origin, would do the same for later times. In the second case, you would need to calculate the logarithm of (origin − time) and devise intelligible labels.

However, a major limitation in both cases is that the origin is not plottable because the logarithm of zero is indeterminate. A fix for this is to ensure that the origin is not a data point but is beyond the range of the data. Otherwise put, if time 0 is a data point, the origin must be offset. Karsten (1923, chap. 41) called plotting in terms of logarithm of (offset + latest − time) "retrospective logarithmic projection" and gave economic examples. This common need for an offset is at best awkward and at worst leads to arbitrary choices that are difficult to justify or explain.

It remains true that thinking logarithmically is highly desirable for thinking broadly about the ranges of time scales (and sizes, speeds, and numerosity) to be found in science. See, for example, the engaging essay "Life on log time" in Glashow (1991) or the more extended discussions in Schneider (2009).

Although using a logarithmic scale for either variable or both variables is likely to strike readers as an elementary and obvious technique, its history is in several ways surprisingly short (Funkhouser 1937, 359–361). Lalanne (1846) introduced double logarithmic scales for calculation problems: Hankins (1999, 71) reproduces his graph. In economics, Jevons (1863, 1884) plotted data for responses on a logarithmic scale versus time; for discussion, see Morgan (1990), Klein (1997), and Stigler (1999). Nevertheless, adoption of the idea was slow, and it was even resisted as obscuring the data. Expository articles urging the merits of the technique, often called semilogarithmic or ratio charts, were still appearing many years after its introduction (Fisher 1917; Field 1917; Griffin and Bowden 1963; Burke 1964).

3 Square root scale?

In practice in data analysis, however, logarithmic transformations are likely to be too strong unless change on very different time scales is being considered. A weaker transformation that can work well is the square root, as recently suggested by Wills (2012, 171–172). The square root also has the simple advantage that the root of 0 is also 0 and is thus quite plottable.

The earliest example of a square root scale known to me is by Moseley (1914) in a classic physics problem that led to predictions of as yet undiscovered elements. Heilbron (1974, 100) reproduces his graph. Karsten (1923, 485) plotted the square root of response versus time in an example in which a logarithmic scale was clearly too severe. Fisher and Mather (1943) plotted frequency counts on a square root scale. Their idea was taken up in much more detail by Mosteller and Tukey (1949), and Fisher (1944, 1950) added corresponding material to the ninth and eleventh editions of his text *Statistical Methods for Research Workers*.

There are contexts ranging from physics to finance in which the square root of time appears naturally, or at least mathematically, in discussions, but we leave that to one side. In the example to follow, the justification for using square roots is, once more, just pragmatic.

Consider some data from Haywood (2011) on world population over the last 8,000 years, available with the media for this journal issue. As is customary, the cautions and caveats associated with such estimates should be flagged. Comprehensive censuses are relatively recent historically, and even long national experience in census-taking is no guarantee of population counts that all experts regard as accurate. Over most of human history, we have only a variety of wild guesses, and competent workers in the field often only agree if they are using each other's estimates. Let us look first at some standard plots.

```
. use haywood
. twoway connected pop year, msymbol(Oh)
. twoway connected pop year, msymbol(Oh) yscale(log)
> ylabel(20 50 200 500 2000 5000)
```

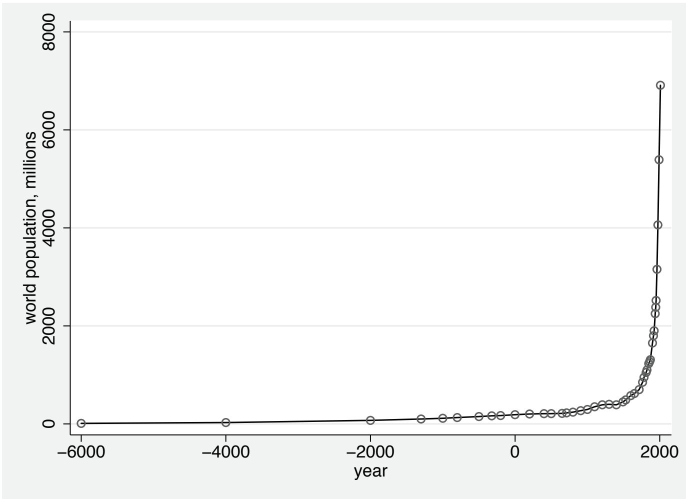

Figure 1. World population over the last 8,000 years (data from Haywood [2011])

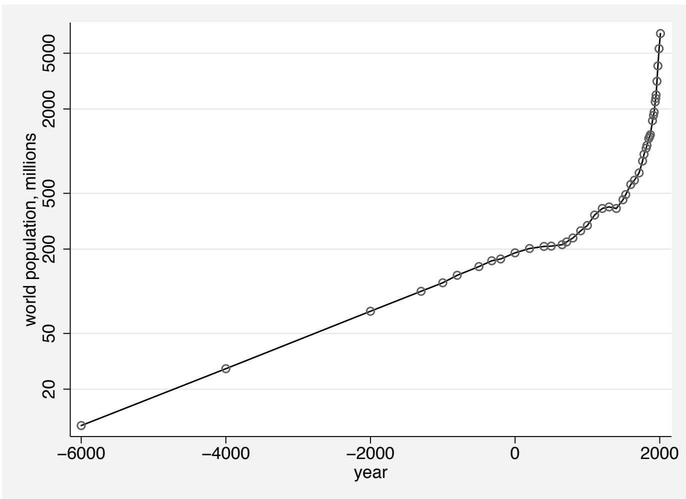

Figure 2. World population over the last 8,000 years; the vertical scale is logarithmic, so periods showing exponential change (constant growth rate) would plot as linear segments (data from Haywood [2011])

Figure 1 draws attention to the explosive nature of population growth over the last few centuries. It has a simple and powerful message, but detail is hard to discern. Figure 2 adds a logarithmic scale and shows a main pattern of accelerating growth rates with strong hints of other fluctuations, reflecting extended periods of higher death rates. For both graphs, using `twoway connected`, a hybrid of `scatter` and `line`, has a natural purpose of emphasizing the precise dates of data points. With more data or data regularly spaced in time, using `line` is likely to be a better choice. That detail aside, figure 1 and figure 2 are standard graphs that are both useful, but neither deals with the problem of crowding of data for later times.

Using a square root scale for time before the present requires two steps before we can create a graph. First, calculate the new time scale.

```
. generate time = sqrt(2010 - year)
```

The latest year for data is 2010. Clearly, other datasets would lead to different constants.

Second, produce better axis labels. The key here is that any text you like can be shown as text in axis labels. In practice, we need to work out what labels we want to see and then where they would be put in terms of the new variable, here `time`. This can take some experimentation, but it is best done by looping over the labels we want and working out the option call that is needed.

```
. foreach y of num -6000(2000)2000 500 1000 1500 {
.       local call `call´ `=sqrt(2010 - `y´)´ "`y´"
. }
```

Here the result is put into a local macro. Every time around the loop, we calculate the square root of 2010 minus some year and add that value and the text to be shown on the time axis—namely, the year itself—as extra content in the local macro. So the first time around, the year is −6000. We use Stata to calculate `sqrt(2000 - (-6000))` on the fly and insert the result in the macro. The special syntax `‘=`*expression*`’` instructs Stata to calculate the result of the *expression* and use that result immediately.

If you do the calculation for −6000 yourself with `display`, you will find that the result is given as 89.442719. On-the-fly calculation gives more decimal places, far more than are really needed for graphical purposes, but no harm is done thereby. The calculation

```
89.44271909999159 "-6000"
```

is added to the macro, which is created empty. The next time around, the macro will contain

```
89.44271909999159 "-6000" 77.45966692414834 "-4000"
```

and so on.

If you do this yourself, make sure to blank out the local macro by typing

```
. local call
```

before revising any earlier guess. Otherwise, earlier insertions in the macro will just be carried forward.

Now the graph shown as figure 3 is possible. Reversing the scale is needed because the time variable increases going backward, at least if you want to show the direction of time in the usual manner.

```
. twoway connected pop time, msymbol(Oh) xscale(reverse) xlabel(`call´)
> xtitle(year)
```

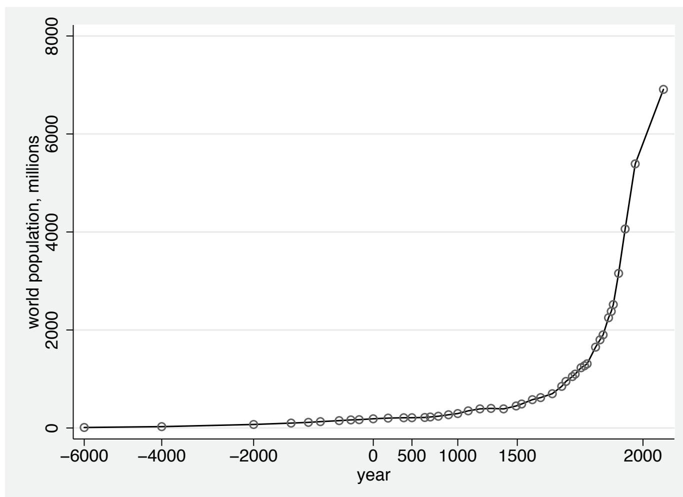

Figure 3. World population over the last 8,000 years; the horizontal axis represents time before 2010 on a square root scale (data from Haywood [2011])

You can see the advantage of automating the production of the option call: you would not prefer to do all the little calculations one by one with a flurry of copy-and-paste to transfer results to the graph command.

Nothing stops the use of `yscale(log)` too, although the standard result that exponential change shows as linear segments no longer holds. My own preference is to stop at transforming the time scale.

As usual, smaller graphical details remain open to choice. The axis title of “year” might seem redundant. A method other than negative dates might be preferred for showing years BCE. The y-axis labels could be rotated to horizontal.

4 Scale break?

Another popular solution to the problem of crowding for early or later times is to use a scale break so that the time series is divided into two or more panels with different scales. Scale breaks themselves divide statistically minded people. Some argue for them as a simple practical solution, so long as it is made explicit, and some argue against them as awkward, artificial, and usually unnecessary, because transformation is typically a better solution. Good discussions are given by Cleveland (1994) and Wilkinson (2005). For more discussion of scale breaks in Stata graphics, see Cox and Merryman (2006).

Stata does not support scale breaks directly, perhaps mostly because a graph with a scale break is not easily made consistent with the general philosophy of allowing graphs to be superimposed or combined with a common axis. But you can devise your own scale breaks by producing separate graphs before combining them again.

```
. twoway connected pop y if year < 0, msymbol(Oh) yscale(r(10 7000) log)
> ylabel(10 20 50 100 200 500 1000 2000 5000, angle(h))
> saving(part1, replace) xlabel(-6000(1000)0) xtitle("")
. twoway connected pop y if year > 0, ms(Oh) ysc(r(10 7000) log off)
> yla(10 20 50 100 200 500 1000 2000 5000, ang(h))
> saving(part2, replace) xla(250(250)2000) xtitle("")
. graph combine "part1" "part2", imargin(small) b1title(year, size(small))
```

The result is shown in figure 4. It is likely to qualify as many users' best bet for this kind of problem, if only because it is easier to explain than other solutions.

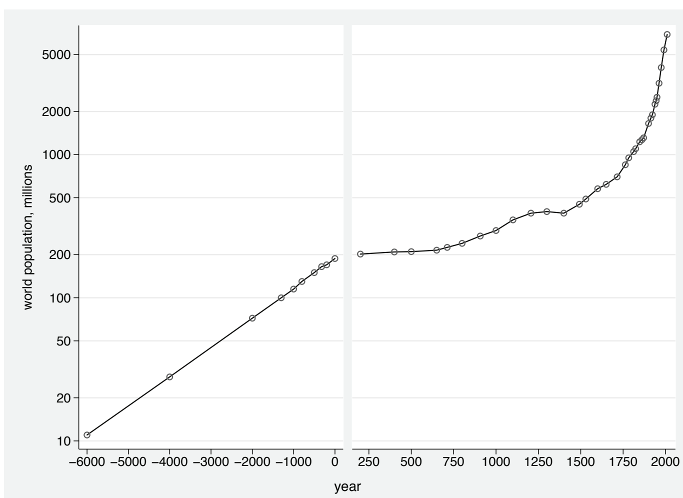

Figure 4. World population over the last 8,000 years; scale break at change from BCE to AD (data from Haywood [2011])

Let us focus on some of the small details in these commands that might otherwise seem puzzling.

1. I tend to include the `replace` suboption of `saving()` even the first time around, because in practice I need several iterations before all my choices seem right. Nevertheless, do check that you are not losing a valuable graph from some previous project.

2. `graph combine` has a `ycommon` option. But with `yscale(log)` too, I get better results by being explicit in `yscale()` in both graph commands that the y axes have the same limits.

3. Why spell out the `ylabel()` call for the second graph if the labels are omitted by `yscale(off)`? The reason is that the label locations determine horizontal grid lines. (Naturally, if you did not want those grid lines, you should omit them.)

4. The title "year" should certainly not be given twice, so we blank it out in the individual graph commands and put it back (once) in the `graph combine` command.

5. The first panel ends up a bit smaller in this example as a side-effect of omitting the y axis and its labels from the second panel. That seems about right, but the relative sizes of the panels could be further controlled by using the `fxsize()` option with the initial graph commands. Cox and Merryman (2006) give an example.

In principle, two or more scale breaks are possible, giving three or more panels. You just need to produce the individual graph panels as separate graphs and then combine them. But the trade-offs are clear. Each panel must be big enough to make clear what the different scales are, but without overloading the time axis with many different labels or ticks. For that and other reasons, even a graph with two scale breaks and three panels is likely to be too complicated for most purposes.

5 Conclusions

For showing change over time graphically, transformation almost always means transforming the response. Logarithmic scales are by far the most widely used example and need little publicity. In this column, I focused on a less common but intriguing possibility: transforming time. Although logarithmic scales have been used on occasion for time, they can be awkward at best. In contrast, a square root scale for time is easy to implement and can work well. Using a scale break to divide a graph into separate panels is also a possibility.

As always, when users have choices, they also have to make decisions on what to do. The question is likely to depend on specific circumstances of dataset, analysis, and audience, as well as researchers' general attitudes. In choosing any of the graphs here for a presentation, what the audience would find easiest to understand—or in some cases, what they would find most stimulating to consider—would be the first question to consider.

6 Acknowledgment

Stephen M. Stigler provided invaluable details for references to successive editions of R. A. Fisher's text.

7 References

Burke, T. 1964. Semi-logarithmic graphs in geography: A pertinent addendum. *Professional Geographer* 16(1): 19–21.

Cleveland, W. S. 1994. *The Elements of Graphing Data*. Rev. ed. Summit, NJ: Hobart.

Cox, N. J. 2008. Stata tip 59: Plotting on any transformed scale. *Stata Journal* 8: 142–145.

Cox, N. J., and S. Merryman. 2006. Stata FAQ: How can I show scale breaks on graphs? http://www.stata.com/support/faqs/graphics/scbreak.html.

Field, J. A. 1917. Some advantages of the logarithmic scale in statistical diagrams. *Journal of Political Economy* 25: 805–841.

Fisher, I. 1917. The "ratio" chart for plotting statistics. *Publications of the American Statistical Association* 15: 577–601.

Fisher, R. A. 1944. *Statistical Methods for Research Workers*. 9th ed. Edinburgh: Oliver & Boyd.

———. 1950. *Statistical Methods for Research Workers*. 11th ed. Edinburgh: Oliver & Boyd.

Fisher, R. A., and K. Mather. 1943. The inheritance of style length in *Lythrum salicaria*. *Annals of Eugenics* 12: 1–23.

Funkhouser, H. G. 1937. Historical development of the graphical representation of statistical data. *Osiris* 3: 269–404.

Glashow, S. L. 1991. *The Charm of Physics*. New York: American Institute of Physics.

Griffin, D. W., and L. W. Bowden. 1963. Semi-logarithmic graphs in geography. *Professional Geographer* 15(5): 19–23.

Hankins, T. L. 1999. Blood, dirt, and nomograms: A particular history of graphs. *Isis* 90: 50–80.

Haywood, J. 2011. *The New Atlas of World History: Global Events at a Glance*. London: Thames & Hudson.

Heilbron, J. L. 1974. *H. G. J. Moseley: The Life and Letters of an English Physicist, 1887–1915*. Berkeley: University of California Press.

Jevons, W. S. 1863. *A Serious Fall in the Value of Gold Ascertained, and Its Social Effects Set Forth.* London: Edward Stanford.

———. 1884. *Investigations in Currency and Finance.* London: Macmillan.

Karsten, K. G. 1923. *Charts and Graphs: An Introduction to Graphic Methods in the Control and Analysis of Statistics.* New York: Prentice Hall.

Klein, J. L. 1997. *Statistical Visions in Time: A History of Time Series Analysis 1662–1938.* Cambridge: Cambridge University Press.

Lalanne, L. 1846. Mémoire sur les tables graphiques et sur la géométrie anamorphique appliquée à diverses questions qui se rattachent à l'art de l'ingénieur. *Annales des Ponts et Chaussées* 11: 1–69.

Morgan, M. S. 1990. *The History of Econometric Ideas.* Cambridge: Cambridge University Press.

Moseley, H. G. J. 1914. The high-frequency spectra of the elements. Part II. *Philosophical Magazine,* 6th ser., 27: 703–713.

Mosteller, F., and J. W. Tukey. 1949. The uses and usefulness of binomial probability paper. *Journal of the American Statistical Association* 44: 174–212.

Schneider, D. C. 2009. *Quantitative Ecology.* 2nd ed. London: Academic Press.

Stigler, S. M. 1999. *Statistics on the Table: The History of Statistical Concepts and Methods.* Cambridge, MA: Harvard University Press.

Wilkinson, L. 2005. *The Grammar of Graphics.* 2nd ed. New York: Springer.

Wills, G. 2012. *Visualizing Time: Designing Graphical Representations for Statistical Data.* New York: Springer.

About the author

Nicholas Cox is a statistically minded geographer at Durham University. He contributes talks, postings, FAQs, and programs to the Stata user community. He has also coauthored 15 commands in official Stata. He wrote several inserts in the *Stata Technical Bulletin* and is an editor of the *Stata Journal*.