The Stata Journal

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press, and Stata is a registered trademark of StataCorp LP.

The Stata Journal (2005)
5, Number 2, pp. 282–284

Stata tip 21: The arrows of outrageous fortune

Nicholas J. Cox
Durham University
n.j.cox@durham.ac.uk

Stata 9 introduces a clutch of new plottypes for `graph twoway` for paired-coordinate data. These are defined by four variables, two specifying starting coordinates and the other two specifying ending coordinates. Here we look at some of the possibilities opened up by [G] **graph twoway pcarrow** for graphing changes over time. Arrows are readily understood by novices as well as experts as indicating, in this case, the flow from the past towards the present.

Let us begin with one of the classic time-series datasets. The number of lynx trapped in an area of Canada provides an excellent example of cyclic boom-and-bust population dynamics. Trappings are optimistically assumed to be proportional to the unknown population size.

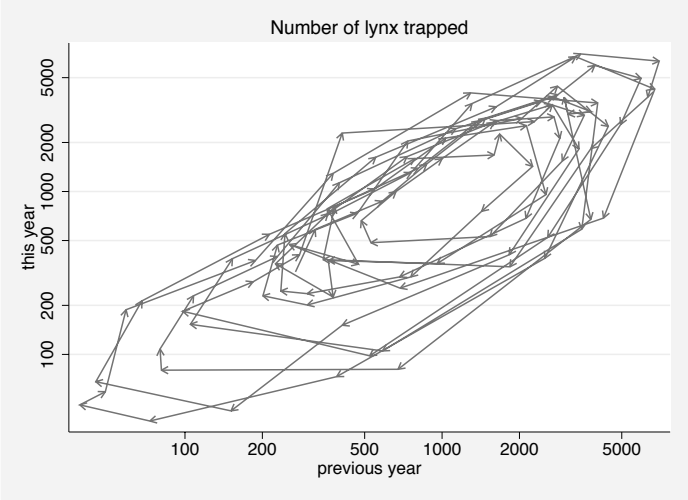
The dataset has already been `tsset` (see [TS] **tsset**).

```
. use http://www.stata-press.com/data/r9/lynx2.dta, clear
(TIMESLAB: Canadian lynx)
. tsline lynx
  (output omitted)
```

Ecologists and other statistically minded people find it natural to think about populations on a logarithmic scale: population growth is after all inherently multiplicative. Logarithms to base 10 are convenient for graphing.

(*Continued on next page*)


```
. gen loglynx = log10(lynx)
. twoway pcarrow loglynx L.loglynx F.loglynx loglynx,
> xla(2 "100" `=log10(200)' "200" `=log10(500)' "500" 3 "1000" `=log10(2000)'
> "2000" `=log10(5000)' "5000")
> yla(2 "100" `=log10(200)' "200" `=log10(500)' "500" 3 "1000" `=log10(2000)'
> "2000" `=log10(5000)' "5000")
> ytitle(this year) xtitle(previous year) subtitle(Number of lynx trapped)
```



Thinking as it were autoregressively, we can plot this year's population versus the previous year's and join data points with arrows end to end. Each data point other than the first and last is the end of one arrow pointing from `L.loglynx` to `loglynx` and the beginning of another pointing from `loglynx` to `F.loglynx`. This dual role and the ability to use time-series operators such as `L.` and `F.` on the fly in graphics commands yield the command syntax just given. Plotted as a trajectory in this space, population cycles are revealed clearly as asymmetric. Depending on your background, you may see this as an example of hysteresis, or whatever else it is called in your tribal jargon.

Another basic comparison compares values for some outcome of interest at two dates. For this next example, we use life expectancy data for 1970 and 2003 from the UNICEF report, *The State of the World's Children 2005*, taken from the web site *http://www.unicef.org* accessed on May 12, 2005. A manageable graph focuses on those countries for which life expectancy was under 50 years in 2003. A `count` on the dataset thus entered shows that there are 33 such countries.

We borrow some ideas from displays possible with `graph dot` (see [G] **graph dot**). Arrows connecting pairs of variables are not supported by `graph dot`. However, as is common with Stata's graphics, whatever is difficult with `graph dot`, `graph bar`, or `graph hbar` is often straightforward with `graph twoway`, modulo some persistence.

A natural sort order for the graph is that of life expectancy in 2003. A nuance to make the graph tidier is to break ties according to life expectancy in 1970. Life expectancy is customarily, and sensibly, reported in integer years, so ties are common. One axis for the graph is then just the observation number given the sort order, except that we will want to name the countries concerned on the graph. For names that might be fairly long, we prefer horizontal alignment and thus a vertical axis. The names are best assigned to value labels. Looping over observations is one way to define those. The online help on `forvalues` and `macros` explains any trickery with the loop that may be unfamiliar to you; also see [P] **forvalues** and [P] **macro**.

```
. gsort lifeexp03 - lifeexp70
. gen order = _n
. forval i = 1/33 {
  2.         label def order `i' "`=country[`i']'", modify
  3. }
. label val order order
```

The main part of the graph is then obtained by a call to `twoway pcarrow`. The arrowhead denotes life expectancy in 2003. Optionally, although not essentially, we overlay a scatter plot showing the 1970 values.

```
. twoway pcarrow order lifeexp70 order lifeexp03 if lifeexp03 < 50
> || scatter order lifeexp70 if lifeexp03 < 50, ms(oh)
> yla(1/33, ang(h) notick valuelabel labsize(*0.75)) yti("") legend(off)
> barbsize(2) xtitle("Life expectancy in years, 1970 and 2003") aspect(1)
```

Apart from Afghanistan, all the countries shown are in Africa. Some show considerable improvements over this period, but in about as many, life expectancy has fallen dramatically. Readers can add their own somber commentary in terms of war, political instability, famine, and disease, particularly AIDS.