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

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	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

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.

Stata tip 122: Variable bar widths in two-way graphs

Ben Jann
University of Bern
Bern, Switzerland
jann@soz.unibe.ch

Two-way bar charts in Stata use a fixed bar width as specified by option `barwidth()` (see [G-2] **graph twoway bar**). Some types of plots found in the literature, however, require variable bar widths. One example is equal probability histograms in which the bar widths are adjusted so that each bar covers the same area; see the `eqprhistogram` command by Cox (1999a). Another example is spine plots for two-way categorical data as implemented in `spineplot` (Cox 2008, 2014).

In this tip, I highlight the `bartype(spanning)` option of the `twoway bar` command, an undocumented feature that can be used to produce bars of different widths. In fact, `eqprhistogram` and `spineplot` are based on this functionality. Consider a plot of income or wealth shares by population percentiles, as is sometimes used in inequality research.[1] Such a plot could be produced as follows:

```
. sysuse nlsw88
(NLSW, 1988 extract)
. sort wage
. generate cumwage = sum(wage)
. replace  cumwage = cumwage/cumwage[_N]
(2246 real changes made)
. _pctile  cumwage, percentiles(20 40 60 70 80 90 95 97 99)
. return list
scalars:
                  r(r1) =  .0803691893815994
                  r(r2) =  .2007369846105576
                  r(r3) =  .3634746670722961
                  r(r4) =  .4661617577075958
                  r(r5) =  .5882071852684021
                  r(r6) =  .7350806593894958
                  r(r7) =  .8268628716468811
                  r(r8) =  .8735467791557312
                  r(r9) =  .9496700167655945
```

1. For an amazing example, see http://www.youtube.com/watch?v=slTF_XXoKAQ. The percentile share plot is a binned and rescaled version of the quantile plot (see [R] **diagnostic plots** and Cox [1999b]). In inequality research, the quantile plot is also known as Pen's "Parade of Dwarfs (and a few Giants)" (Pen 1971, 48–59).

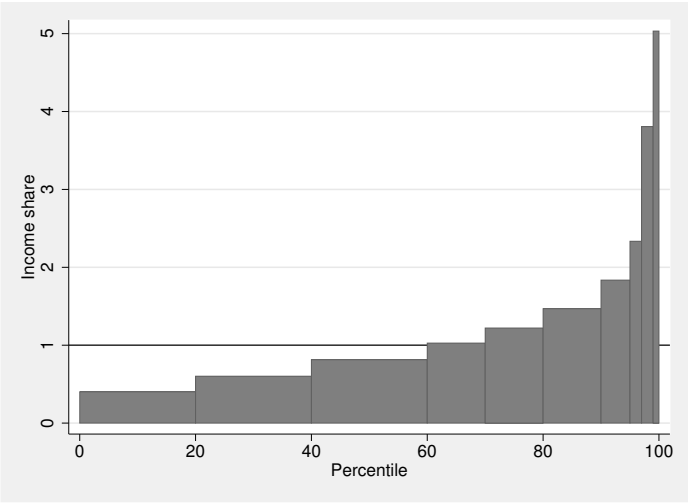

```
. matrix S = ( 0,  (r(r1) - 0    ) / ( 20 - 0 ) * 100)
>          \ ( 20, (r(r2) - r(r1)) / ( 40 - 20) * 100)
>          \ ( 40, (r(r3) - r(r2)) / ( 60 - 40) * 100)
>          \ ( 60, (r(r4) - r(r3)) / ( 70 - 60) * 100)
>          \ ( 70, (r(r5) - r(r4)) / ( 80 - 70) * 100)
>          \ ( 80, (r(r6) - r(r5)) / ( 90 - 80) * 100)
>          \ ( 90, (r(r7) - r(r6)) / ( 95 - 90) * 100)
>          \ ( 95, (r(r8) - r(r7)) / ( 97 - 95) * 100)
>          \ ( 97, (r(r9) - r(r8)) / ( 99 - 97) * 100)
>          \ ( 99, (    1 - r(r9)) / (100 - 99) * 100)
>          \ (100, .)

. matrix list S

S[11,2]
             c1         c2
 r1           0  .40184595
 r2          20  .60183898
 r3          40  .81368841
 r4          60  1.0268709
 r5          70  1.2204543
 r6          80  1.4687347
 r7          90  1.8356442
 r8          95  2.3341954
 r9          97  3.8061619
r10          99  5.0329983
r11         100          .

. svmat S

. twoway bar S2 S1, bartype(spanning)
> yline(1) xtitle(Percentile) ytitle(Income share)
```



First, the running sum of ordered wages is computed and divided by the wage total. Second, `_pctile` is used to compute a series of percentiles from the cumulated wages (see [D] **pctile**). Results are collected in a matrix, where income shares are computed as differences between consecutive percentiles and normalized by the population share.

Third, `svmat` (see [P] **matrix mkmat**) is used to store the matrix columns as variables, and `twoway bar` is applied with the `bartype(spanning)` option to create the graph. Variable `S1` (the first column in the matrix) specifies the lower bounds of the bars on the x axis, and variable `S2` (the second column) specifies the heights of the bars. The width of a bar is determined by `bartype(spanning)` such that it spans the x axis to the lower bound of the next bar. An extra row in the matrix is needed to provide the upper bound for the last (rightmost) bar.

Such a plot provides a very intuitive view on the wage distribution. Suppose you have 100 dollars to distribute among 100 people. The people are lined up along the x axis in ascending order of their shares, which are depicted by the heights of the bars. If the distribution is equal, each person gets 1 dollar (horizontal line). For the data at hand, however, we see that there is inequality. For example, the person with the largest share gets 5 dollars, and the 20 people with the lowest shares get only about 40 cents on average. The area of a bar reflects the fraction of total wages received by the corresponding group. For example, the bottom 20% of people receive $0.4 \times 20 = 8\%$ of the sum of wages.

1 References

Cox, N. J. 1999a. eqprhistogram: Stata module for equal probability histogram. Statistical Software Components S432701, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s432701.html.

———. 1999b. gr42: Quantile plots, generalized. *Stata Technical Bulletin* 51: 16–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 113–116. College Station, TX: Stata Press.

———. 2008. Speaking Stata: Spineplots and their kin. *Stata Journal* 8: 105–121.

———. 2014. *Speaking Stata Graphics*. College Station, TX: Stata Press.

Pen, J. 1971. *Income Distribution*. London: Allen Lane.