

THE STATA JOURNAL

Editor

H. Joseph Newton
Department of Statistics
Texas A&M University
College Station, Texas 77843
979-845-8817; fax 979-845-6077
jnewton@stata-journal.com

Editor

Nicholas J. Cox
Department of Geography
Durham University
South Road
Durham DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors

Christopher F. Baum
Boston College

Nathaniel Beck
New York University

Rino Bellocco
Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy

Maarten L. Buis
Tübingen University, Germany

A. Colin Cameron
University of California–Davis

Mario A. Cleves
Univ. of Arkansas for Medical Sciences

William D. Dupont
Vanderbilt University

David Epstein
Columbia University

Allan Gregory
Queen's University

James Hardin
University of South Carolina

Ben Jann
University of Bern, Switzerland

Stephen Jenkins
London School of Economics and
Political Science

Ulrich Kohler
WZB, Berlin

Frauke Kreuter
University of Maryland–College Park

Peter A. Lachenbruch
Oregon State University

Jens Lauritsen
Odense University Hospital

Stanley Lemeshow
Ohio State University

J. Scott Long
Indiana University

Roger Newson
Imperial College, London

Austin Nichols
Urban Institute, Washington DC

Marcello Pagano
Harvard School of Public Health

Sophia Rabe-Hesketh
University of California–Berkeley

J. Patrick Royston
MRC Clinical Trials Unit, London

Philip Ryan
University of Adelaide

Mark E. Schaffer
Heriot-Watt University, Edinburgh

Jeroen Weesie
Utrecht University

Nicholas J. G. Winter
University of Virginia

Jeffrey Wooldridge
Michigan State University

Stata Press Editorial Manager
Stata Press Copy Editor

Lisa Gilmore
Deirdre McClellan

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.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index[®]
- Current Contents/Social and Behavioral Sciences[®]
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch[®])
- Scopus[™]
- Social Sciences Citation Index[®]

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

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

Stata tip 102: Highlighting specific bars

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

A frequent need when drawing a bar or dot chart is to highlight a subset of observations while keeping the overall sort order. The stipulation of keeping the overall sort order is what provides the challenge here, because otherwise we could just add subdivision by another variable to the command, as when distinguishing foreign cars among those with the best repair record:

```
. sysuse auto  
. graph hbar (asis) mpg if rep78 == 5, over(make, sort(1) descending)  
. graph hbar (asis) mpg if rep78 == 5, over(make, sort(1) descending)  
> over(foreign) nofill
```

Figure 1 shows the graphs for these two commands.

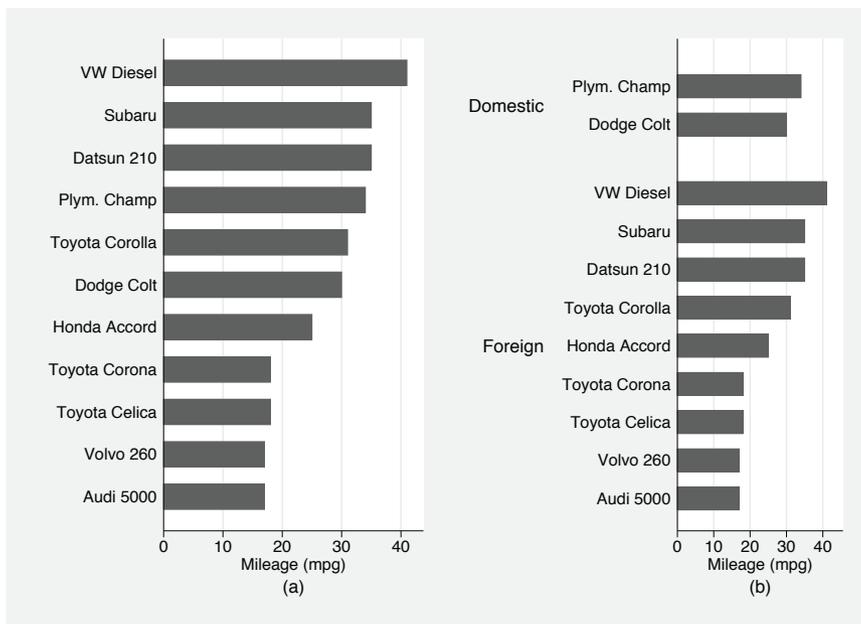


Figure 1

In figure 1(a), the ordering is within all the observations specified. In figure 1(b), the extra option `over(foreign)` subdivides observations according to the further variable

`foreign`. Note also the crucial detail of `nofill`. This can be a useful kind of graph, but it is not what we want here.

Let us suppose we have data on basin (catchment or watershed) areas for various large rivers in the world, and we want to show where the Mississippi falls in the rank order for the very largest basins. Some example data from Allen (1997) are included with the media for this issue.

```
. use rivers
```

Figure 2 as a first graph shows that the Mississippi ranks third on area of basin in this dataset, after the Amazon and Zaire.

```
. graph hbar (asis) area if area >= 1000, over(name, sort(1) descending)
```

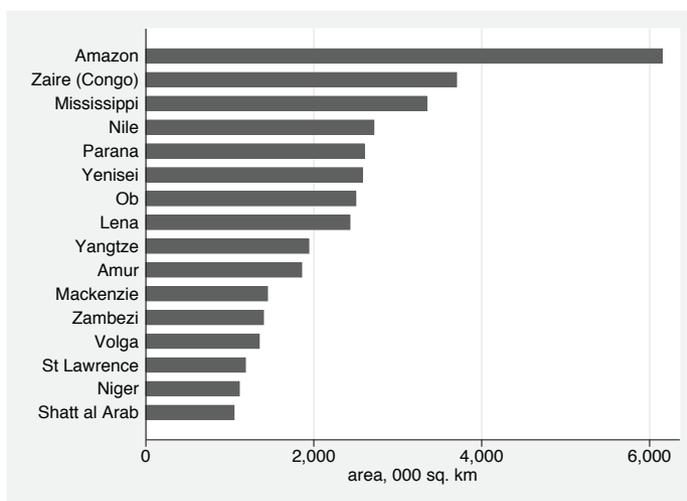


Figure 2

Highlighting a particular bar means giving it a different color. Some acquaintance with the bar chart commands shows that they are willing to combine bars for different variables, which will be assigned different colors. Knowing this, we simply need to put data for two subsets, the Mississippi and the others, into two different variables. `separate` (see [D] `separate`) is a command designed for precisely this purpose. For other graphical applications of `separate`, see Cox (2005). It is naturally also possible to use `generate` directly.

```
. separate area, by(name == "Mississippi")
```

In this example, the equality supplied to `by()` is either false or true, numerically 0 or 1, and so `separate` creates two new variables, `area0` and `area1`.

```
. graph hbar (asis) area0 area1 if area >= 1000, nofill
> over(name, sort(area) descending) legend(off) ytitle("`': var label area`'")
```

We are plotting bars for values that are nonmissing on `area0` and missing on `area1`, or vice versa. But `graph` plots no bars when values are missing. This is easy to fix: `nofill` gets us the intended effect. In this case, we suppressed the legend, imagining that, depending on the purpose, we could add a title for a presentation—as, say

```
title(Mississippi ranks third in catchment area)
```

or underline the message of the graph in informative text supplied in a text or word processor. Because two response variables are being shown on the same graph, we have to step in to provide an informative *y*-axis title, in this case by automating use of the variable label for `area`. Nothing stops us from just providing an axis title explicitly, as when no such variable label has been defined.

In principle, using `stack` should have the same effect as using `nofill`. In practice, small complications can exist if there are other missing values in the data; these complications are fixable with an appropriate `if` exclusion.

The main problem now being solved, we could clearly heighten the contrast by adding `bar(1, bfcOLOR(none))`. Figure 3 shows the graph after that tweak.

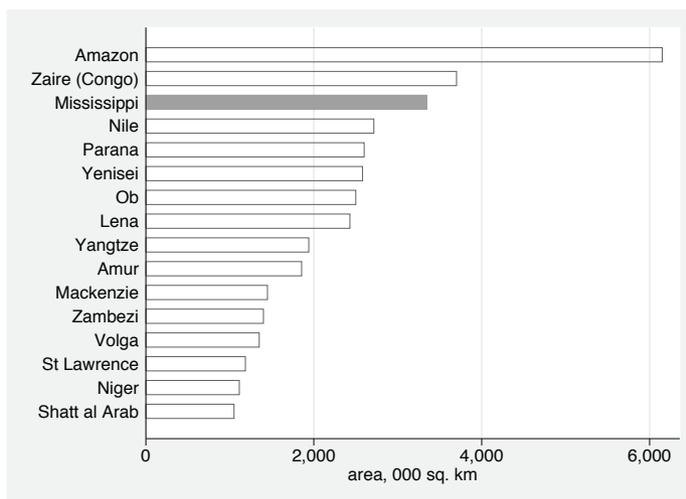


Figure 3

Similar needs are met by variations on this theme.

In our example, the subset to be highlighted is a single observation, but nothing depends on that being true.

Equally, three or more subsets could be distinguished. For a more elaborate subdivision, we might want a legend, although there is a trade-off: the more complicated the design, for which a legend becomes necessary, the less the impact of the graph is likely to be.

The examples all are based on showing values `asis`. If graphs of this kind are needed, but for means or other summary statistics, it is often easiest to `collapse` or `contract` the dataset first and then use `separate` and `graph hbar (asis)`.

The same device can be used with `graph bar`, `graph dot`, or various subcommands of `twoway`, such as `twoway bar`. In practice, when we want this, the individual observations include names that are informative, so horizontal alignment makes those names more readable. If `graph dot` were to be used, we should consider heightening the contrast by adding, for example, `marker(2, msize(*3))`.

References

- Allen, P. A. 1997. *Earth Surface Processes*. Oxford: Blackwell Science.
- Cox, N. J. 2005. Stata tip 27: Classifying data points on scatter plots. *Stata Journal* 5: 604–606.