

# THE STATA JOURNAL

## Editors

H. JOSEPH NEWTON  
Department of Statistics  
Texas A&M University  
College Station, Texas  
editors@stata-journal.com

NICHOLAS J. COX  
Department of Geography  
Durham University  
Durham, UK  
editors@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, WZB, Germany  
A. COLIN CAMERON, University of California–Davis  
MARIO A. CLEVES, University of Arkansas for  
Medical Sciences  
WILLIAM D. DUPONT, Vanderbilt University  
PHILIP ENDER, University of California–Los Angeles  
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, Germany

FRAUKE KREUTER, Univ. 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, Univ. of California–Berkeley  
J. PATRICK ROYSTON, MRC Clinical Trials Unit,  
London  
PHILIP RYAN, University of Adelaide  
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh  
JEROEN WEESIE, Utrecht University  
NICHOLAS J. G. WINTER, University of Virginia  
JEFFREY WOOLDRIDGE, Michigan State University

## Stata Press Editorial Manager

LISA GILMORE

## Stata Press Copy Editors

DAVID CULWELL and DEIRDRE SKAGGS

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>

**Subscriptions** are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

**Subscription rates** listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
1-year subscription	\$ 79	1-year subscription	\$115
2-year subscription	\$155	2-year subscription	\$225
3-year subscription	\$225	3-year subscription	\$329
3-year subscription (electronic only)	\$210	3-year subscription (electronic only)	\$210
1-year student subscription	\$ 48	1-year student subscription	\$ 79
1-year university library subscription	\$ 99	1-year university library subscription	\$135
2-year university library subscription	\$195	2-year university library subscription	\$265
3-year university library subscription	\$289	3-year university library subscription	\$395
1-year institutional subscription	\$225	1-year institutional subscription	\$259
2-year institutional subscription	\$445	2-year institutional subscription	\$510
3-year institutional subscription	\$650	3-year institutional subscription	\$750

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](mailto:sj@stata.com).



Copyright © 2012 by StataCorp LP

**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, **STATA**, Stata Press, Mata, **mata**, and NetCourse are registered trademarks of StataCorp LP.

# Speaking Stata: Axis practice, or what goes where on a graph

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

**Abstract.** Conventions about what information goes on each axis of a two-way plot are precisely that, conventions. This column discusses—historically, syntactically, and by example—the idea that flouting convention in various ways can lead to small but useful improvements in graph display. Putting  $y$ -axis information on the right or on the top, or putting  $x$ -axis information on the top, often is useful. The most substantial examples are for multiple quantile plots, for which the new command `multqplot` is offered, and table-like graphs, which are made even more table-like by mimicking column headers.

**Keywords:** `gr0053`, `multqplot`, `qplot`, `tabplot`, axes, coordinates, quantile plots, two-way bar charts

## 1 Introduction

What goes where on a rectangular graph (which means most rectangular graphs)? This question is easy to answer, if only with tongue in cheek. Plot what you think of as response or outcome on the vertical or  $y$  axis unless you have good reason to do otherwise. The “otherwise”, for example, can include cases where convention indicates that depth or height should be plotted on the vertical axis. Even if that has been decided, there remains the smaller but still interesting question of where axis titles, labels, and ticks for such variables are best placed.

Such a specific question raises a broader issue: How far is there a binding logic to graphical choices, and how far are we free to follow conventions or to flout them according to taste or circumstance?

The key questions are made concrete by focusing on graphs that are implemented in terms of Stata commands such as `twoway scatter`, `twoway line`, or `twoway connected`. For such graphs, Stata’s defaults will be familiar to you:  $y$ -axis information goes on the left and to the left of the axis line, and  $x$ -axis information goes on the bottom and beneath the axis line. This is a widespread convention in science and mathematics that presumably arises because many graphs show quantities that are 0 or positive and also show axes that are  $y = 0$  and  $x = 0$ . Thus putting extra explanatory information outside the data region and next to the axes seems natural.

In this column, I will discuss situations in which subverting this convention makes sense and how to do that in Stata.

## 2 A little history

Graphical choices should be defended as matters of logic when there is an underlying logic; otherwise, they may just be explained as matters of taste, convenience, or convention. Even so, a look at the history of axes and coordinates is instructive, and perhaps even surprising. Sources that are especially informative or entertaining include [Cajori \(1929\)](#), [Boyer \(1956\)](#), [Bochner \(1966\)](#), [Kline \(1972\)](#), and [Stillwell \(2010\)](#).

The idea of coordinates was latent in Greek mathematics, in the work of Nicole Oresme (c. 1320–1382), and indeed throughout the practices of surveying, cartography, navigation, and astronomy. The key breakthrough of seeing that geometry and algebra could be united as coordinate or analytic geometry to the benefit of both is generally attributed to Pierre de Fermat (1601–1665) and René Descartes (1596–1650).

Descartes was responsible for the present convention of using letters such as  $x$ ,  $y$ , and  $z$  for coordinates and receives enduring credit through the still widespread term *Cartesian coordinates*. Nevertheless, the work of Fermat and Descartes was still some distance from what is now standard. According to [Bochner \(1966, 40\)](#), “coordinate systems in the full sense of the notion only *begin* to occur in the work of Descartes himself”. The explicit use of both  $x$  and  $y$  axes, or a  $z$  axis when needed, and the idea that either  $x$  or  $y$  might be negative became more evident in the later 17th and 18th centuries. Histories of mathematics differ on who introduced precisely what and when, but outstanding mathematicians such as John Wallis (1616–1703), Isaac Newton (1643–1727), Gottfried Wilhelm Leibniz (1646–1716), who introduced the term *coordinates*, and Leonhard Euler (1707–1783) all contributed.

Over the last two hundred years, practices within what we might now call statistical graphics have been far from consistent. Strikingly, William Playfair (1759–1823) in his *Commercial and Political Atlas (1801a)* shows all his time series with  $y$ -axis labels on the right, corresponding to the most recent data, and with an explanation of what is being shown in a boxed title within the data region. He also uses  $y$ -axis labels on the right in many of his other graphs, for example, in his *Statistical Breviary (1801b)*. Both 1801 works were reprinted together in [Playfair \(2005\)](#).

A century or so later, Brinton’s *Graphic Methods for Presenting Facts (1914)* gave a cross-section of current practice and much advice on good and bad technique. (His admonitions against what are now usually called pie charts and against a third dimension introduced in graphs for artistic effect are a foretaste of often repeated, and often ignored, criticisms from most more recent writers on statistical graphics.) On page 361, Brinton proposed 2 of 25 rules offered “simply as suggestions”:

Figures for the horizontal scale should always be placed at the bottom of the chart. If needed, a scale may be placed at the top also.

Figures for the vertical scale should always be placed at the left of the chart. If needed, a scale may be placed at the right also.

However, such principles are unlikely to persuade those otherwise inclined. Specifically, several examples in Brinton's eclectic collection violate both rules. More generally, what force does a rule have if it is merely a suggestion and there is no reason given for why it is to be the standard?

Nevertheless, most statistical graphics over the last century has followed Brinton's advice, so much so that many texts do not even spell out that these are the conventions. Court (1987, 1642) does precisely that: "On rectangular diagrams, primary scales at bottom and left should increase consistently upward and toward the right, away from the axis intersection; secondary scales at top and right may give equivalent units (metric on primary and English on secondary, or frequency at left and wavelength at right) or may apply to a second set of data." This is clear enough, but Court nowhere explains why it would be illogical to do otherwise.

### 3 Flouting conventions

To cut to the chase, I will give examples of good reasons to flout convention.

1. When you plot time series, both theory and practice often imply most interest in most recent values. In addition, if time is a calendar date, 0 is rarely shown in any case (and if it is shown, it is usually in conjunction with dates BCE, as in my previous column [Cox 2012b]). So putting  $y$ -axis labels on the right puts them where they will be easiest to use. This was presumably Playfair's logic, and it has appealed to others since. Klass (2012) is a friendly recent text with several social science examples. In fact, time-series graphs with variable information at the top and axis labels on the right are commonplaces of financial journalism; for example, have a look at issues of the *Economist*, *Financial Times*, or the *Wall Street Journal*.
2. When you plot several graphs together, pressure on space is much increased. Whenever the  $x$  axis shows a common variable, the top of the graph can be a good place for titles explaining the  $y$  variable, which traditionally are placed on the  $y$  axis. This also satisfies the simple but often ignored idea that text is easiest to read if words run from left to right.
3. Table-like graphs will not suffer from more use of table-like conventions. Two-way tables generally give row and column headers, as standard terminology implies, so why not label columns on the top?

We are going to look at examples of the second and third kinds after some brief comments on Stata syntax for changing what goes on what axis.

## 4 Stata syntax

Given an interest in experimenting, how is this to be done? With `twoway` graphs, the easiest way to start is with `yscale(alt)` or `xscale(alt)`, which flips axes so that the other nondefault axis is used for title, labels, and ticks. With a Stata standard, consider

```
. sysuse auto
(1978 Automobile Data)
. scatter mpg weight, yscale(alt)
(graph omitted)
. scatter mpg weight, xscale(alt)
```

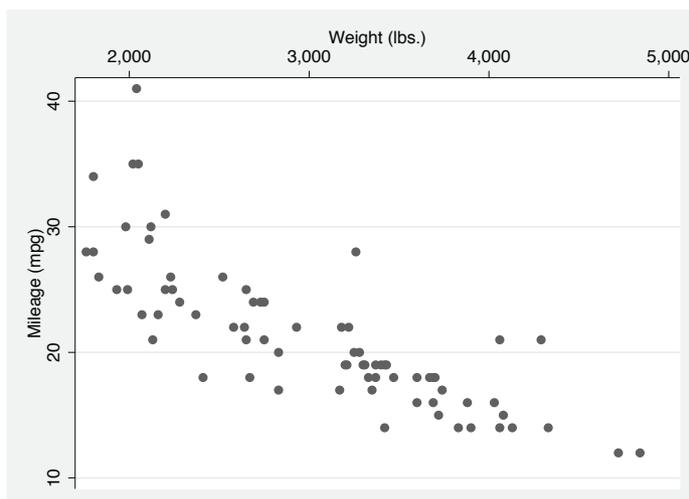


Figure 1. A standard scatterplot, except that  $x$ -axis labels are shown at the top

The second `scatter` command is more interesting and the result is shown as figure 1. It may strike most people as odd, but is this more than a reaction to a conventional choice being set aside? Let me spell it out: the origin (0,0) is not explicit on this graph; that is, neither zero `mpg` nor zero `weight` are near observed values, and no line on the graph is a true axis for which  $y$  or  $x$  is 0. To be sure, in this and many other graphs, there is a case that forcing Stata to show the origin might help many readers to think about the data, and that can be done by spelling out `yscale(r(0 .))` and `xscale(r(0 .))`.

The `alt` suboption here evidently stands for `alternate` and suffices for simple flipping of axes. The more elaborate machinery documented at [G-3] `axis_choice_options` is prominently documented, but not always needed. For example, using the `subtitle()` option is often the most direct method to put a title on the top. The subtitle can be moved around the graph by using its suboptions. It is easy to overlook the extra title options `l1title()`, `l2title()`, `r1title()`, `r2title()`, `t1title()`, `t2title()`, `b1title()`, and `b2title()`. Here `l`, `r`, `t`, and `b` stand for left, right, top, and bottom.

These title options come in pairs: 1 and 2 indicate outer and inner titles. Long-time Stata users may recollect the very similar options in the old graphics of Stata 7 and earlier.

## 5 Multiple quantile plots

Quantile plots have long been a favorite of mine. They plot the ordered values (order statistics or quantiles) from a distribution of one variable against the so-called plotting positions, in effect, estimates of the associated cumulative probability. For discussions and references, see Cox (1999, 2005). Stata implementations include the official `quantile` command (see [R] **diagnostic plots**) and my more general `qplot` command (`quantil2` was an earlier name of the latter). For the most recent update of `qplot`, download from the files for Cox (2012a).

`qplot` can be used to show the distributions of two or more variables, but that makes most sense when variables are measured in the same units and are broadly comparable in magnitude. However, a helpful, broad exploratory view of a dataset may be obtained by looking at several quantile plots together. In essence, that just implies cycling over a `qplot` call for each of several variables, letting each `qplot` find its own  $y$ -axis scale.

There are many ways to do this. The `multqplot` command distributed with this column is certainly indicative rather than definitive. It loops over a numeric *varlist* and uses `graph combine` to show the resulting graphs in one image. A formal statement of its syntax is given in a later section.

Let's see an example first. We return to the `auto.dta` example. `make` is a string variable, but the other 11 variables are numeric, and so typing

```
. multqplot price-foreign
```

shows all the univariate distributions (figure 2).

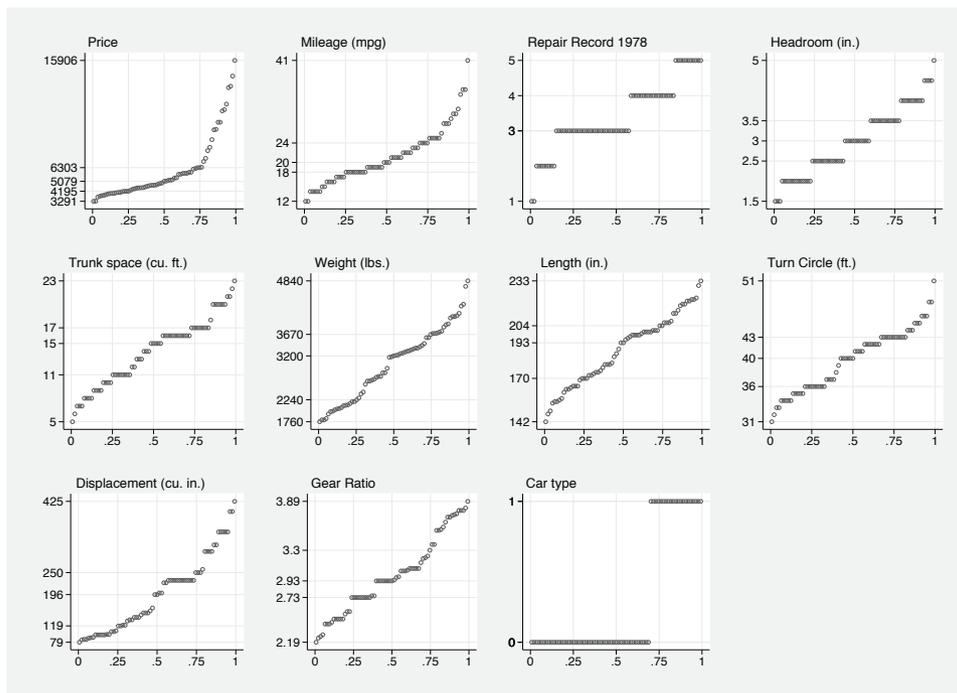


Figure 2. Multiple quantile plot of the numeric variables in `auto.dta`; the quantiles of each variable are plotted against the equivalent fractions of the data, calculated as  $(\text{rank} - 0.5) / \text{sample size}$

Coincidentally, this example with 11 graphs is about the limit of what is comfortably seen on a modestly sized monitor. It is also true that even when reduced to about  $1/12$ th of the usual graph size, a quantile plot still shows much interesting and intelligible detail. Those faced with hundreds or thousands of variables would need a different strategy. For example, arranging to see numerous quantile plots in a slide show is likely to be more practical.

Several choices in the design of the individual quantile plots shown by `multqplot` reflect a desire to make the most of a small space:

1. The variable information (a variable label or, in its absence, a variable name) goes on the top. Left-to-right orientation makes it more readable.
2. That frees up the  $y$  axis to some extent, and the default is to show the values of five summary values: minimum, lower quartile, median, upper quartile, and maximum. As grid lines are also shown by default for probabilities 0, 0.25, 0.5, 0.75, and 1, the combined effect is to hybridize quantile and box plots (compare especially [Parzen \[1979\]](#)). The box corresponding to 0.25, 0.5, and 0.75 quantiles is not, however, emphasized visually; but it would be possible to program having

two sets of grid lines with different thicknesses. Conversely, this plot shows much more information than a conventional box plot in so far as all quantiles are shown. If the same value is also reported as two or more summary quantiles, then fewer than five values are shown on the  $y$  axis. This will always happen with binary variables—for example, `foreign` in this dataset—and may easily happen with other variables with relatively few distinct values.

3. Although the default display of quantiles has just been discussed in detail, any set of quantiles may be selected to show numerically on the  $y$  axis. My advice is not to try to show too many, because they will be difficult to read.
4. The default  $x$ -axis title, “fraction of the data”, would be the same for all individual `qplots` and is suppressed partly to avoid repetition and partly because of lack of space. Putting an explanation in the text caption for the graph, as here, is thus natural.

`multqplot` makes some discreet use of grid lines. Fashion has moved away from grid lines recently in statistical graphics, but they can be useful; see Cox (2009b) for more discussion.

Further advocacy of quantile plots is not the main purpose of this column, and my previous papers in that vein are easily accessible. But those to whom quantile plots are unfamiliar might note some of their advantages:

1. They show well any outliers, gaps, or granularity, which is useful in data checking or assessment of data quality.
2. They scale well over a large range of possible sample sizes.
3. They entail a minimum of arbitrary choices compared with, say, bin origin and bin width for histograms or kernel type and width for kernel density estimation.
4. They signal behavior that might be awkward in later modeling of the data, whether variables are responses or predictors.
5. They behave reasonably with ordinal or binary variables. (They cannot do anything other than reflect the numerical coding of nominal variables.)

That said, showing small multiples for any univariate distribution plot is easy to copy for other displays, and `multqplot` could serve as a starting point for programmers. When you code alternatives, it may prove useful to rethink the style of the standard display.

## 6 Table-like graphs

Another favorite theme of mine in these columns has been the idea that the conventional distinction between tables and graphs may usefully be subverted. For much more

discussion and some literature references, see [Cox \(2008\)](#). In Stata, it is much easier to make graphs more table-like than to make tables more graph-like. Although it is clearly a cosmetic detail, moving  $x$ -axis information to the top can make graphs even more table-like by mimicking the convention of column headers.

To see this in practice, we will revisit an example used in the documentation; see [\[R\] kappa](#). The data concern the structure of agreement and disagreement between two radiologists in the diagnosis of a sample of tissues.

```
. webuse rate2
(Altman p. 403)
. kap rada radb, tab
```

Radiologist t A's assessment	Radiologist B's assessment				Total
	Normal	benign	suspect	cancer	
Normal	21	12	0	0	33
benign	4	17	1	0	22
suspect	3	9	15	2	29
cancer	0	0	0	1	1
Total	28	38	16	3	85

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
63.53%	30.82%	0.4728	0.0694	6.81	0.0000

We can make the table graphical without losing any vital information. `tabplot` ([Cox 2004](#)) uses the simple idea of a two-way bar chart to represent cell frequencies from a table like that above. The most up-to-date version of `tabplot` may be installed on your system by typing `ssc install tabplot`; that version is needed to replicate the results here. The command

```
. tabplot rada radb, showval
```

gives a close replica of the core of the table (figure 3).

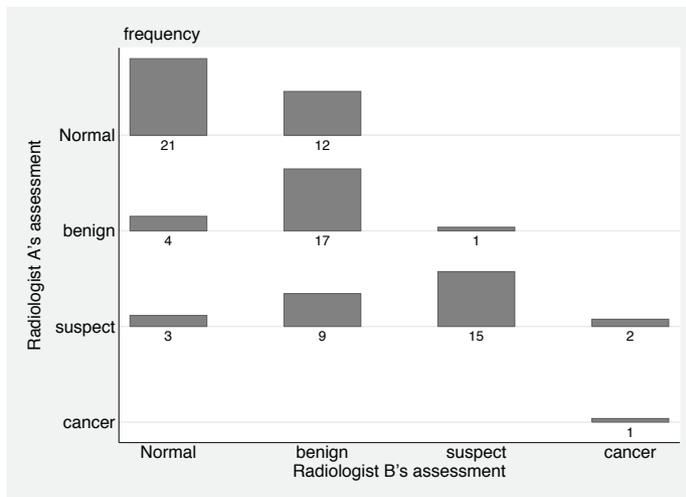


Figure 3. Two-way bar chart showing agreement and disagreement in diagnosis of two radiologists: bars are proportional in height to frequencies, and frequencies (cell counts) are shown below each bar

We can go further. First, there is enough space to show more detail. Let's show not only the frequencies but also those frequencies as a percentage of the total. In this case, the calculation is simple, although in many other datasets, you would need to watch for missing values.

```
. by rada radb, sort: generate freq = _N
. generate percent = 100 * freq/_N
```

`tabplot`'s `showval` option has a related option with the same name that takes arguments, including the name of a variable whose values are to be shown. (Rather, that is how it appears to the user: the simple programming of this subterfuge is explained in Cox [2009a].) The variable constructed for that purpose is a concatenation of values, spaces, and a percent sign. A `format()` suboption is used to round the percentages. (Showing percentages is itself another convention that usually works well, but there are alternatives. Showing numbers expressed per thousand [per mille] can even save a character [12.3 becomes 123], at the small cost of a little explanation.)

```
. generate show = string(freq) + "    " + string(percent, "%2.1f") + "%"
```

The final flourish is to flip the  $x$ -axis information to the top. At the same time, the default subtitle—in this case, **Frequency**—is just suppressed. Other choices are possible, but in this case, it seems simplest to transfer such explanation to the text caption as given below the graph. The result is figure 4.

```
. tabplot rada radb, showval(show) xscale(alt) subtitle("")
```

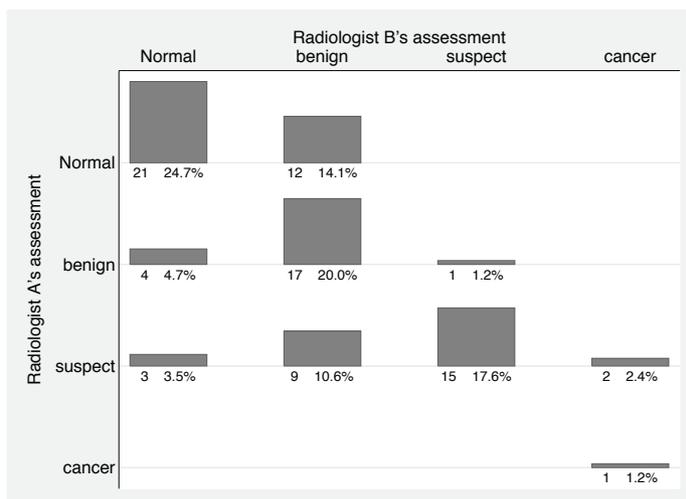


Figure 4. Two-way bar chart showing agreement and disagreement in diagnosis of two radiologists: Bars are proportional in height to frequencies, and frequencies are shown as both cell counts and percentage of the table total. *x*-axis labels have been moved to the top.

This is a relatively simple table. A conservative view that the graph does not add much to what is evident from the table is defensible, but the graph might prove as interesting or more so to some readers. Often tables are much more complicated, however, and then a graph may really help a reader to see structure more easily. But to return to the main idea of this column: Is the second graph preferable to the first?

## 7 Syntax of multqplot

```
multqplot varlist [if] [in] [, allobs quantile(numlist) varnames
      qplot_options combine(combine_options) ]
```

### 7.1 Description

`multqplot` combines and shows quantile plots for each of the numeric variables in *varlist* by using `qplot`, which must also be installed.

By default,

- plots are shown only for those observations for which every variable in *varlist* is nonmissing;
- each variable is explained by a `subtitle(, place(w))` at the top of each quantile plot, which shows the variable label or the variable name if no variable label is defined;
- quantiles for 0(25)100% of the data are shown by horizontal labels on the *y* axis;
- fractions of the data corresponding to those quantiles are shown by horizontal labels on the *x* axis;
- grid lines are shown matching *y*- and *x*-axis labels;
- titles are blank on both *y* and *x* axes.

Variations from the basic design may be obtained by particular option choices.

## 7.2 Options

`allobs` specifies that all observations specified are to be used as far as possible, regardless of the structure of missing values. The default is to use only those observations with nonmissing values on all the variables specified.

`quantile(numlist)` specifies a numlist of percentages between 0 and 100 (inclusive) that select particular quantiles to be calculated and shown as labels on the *y* axis. The numbers 0 and 100 are interpreted as the sample minimum and maximum, respectively. Quantiles are calculated with `summarize` (see [R] `summarize`), if possible, and with `_pctile` (see [D] `pctile`) otherwise.

`varnames` specifies that variable names should always be used to indicate which variable is being plotted. This option is provided for the circumstance in which at least one variable label is so long that labels should not be shown.

`qplot_options` are options of `qplot` used to tune individual quantile plots; see `help qplot`.

`combine(combine_options)` specifies any of the options documented in [G-2] `graph combine` that are used to tune the combination or joint display of the quantile plots. Note that this includes options for saving the graph to a file (see [G-3] `saving_option`).

## 7.3 Remarks

As usual, this program is at best indicative, not definitive. Some might prefer that variables that are categorical by some definition or declaration be shown in some other

way, say, by a bar chart. That is deliberately not an option here, but readers are welcome to clone the code and write their own variant program.

## 8 Conclusion

Appealing and effective graphics arises from a small series of good large decisions and a large series of good small decisions, most of which are to accept program defaults. This column has focused on one of the small decisions, placement of axis information. You are encouraged to be mildly subversive when it helps! Minute tweaks, such as putting  $y$ -axis information on the right or on the top or putting  $x$ -axis information on the top, often are useful. Indeed, they may already be someone else's orthodoxy.

## 9 References

- Bochner, S. 1966. *The Role of Mathematics in the Rise of Science*. Princeton, NJ: Princeton University Press.
- Boyer, C. B. 1956. *History of Analytic Geometry*. New York: Yeshiva University.
- Brinton, W. C. 1914. *Graphic Methods for Presenting Facts*. New York: Engineering Magazine Company.
- Cajori, F. 1929. *A History of Mathematical Notations. Volume II: Notation Mainly in Higher Mathematics*. Chicago: Open Court.
- Court, A. 1987. Ten proposed rules of numerical diagrams. *Eos, Transactions, American Geophysical Union* 68: 1642.
- Cox, N. J. 1999. 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.
- . 2004. Speaking Stata: Graphing categorical and compositional data. *Stata Journal* 4: 190–215.
- . 2005. Speaking Stata: The protean quantile plot. *Stata Journal* 5: 442–460.
- . 2008. Speaking Stata: Between tables and graphs. *Stata Journal* 8: 269–289.
- . 2009a. Stata tip 79: Optional arguments to options. *Stata Journal* 9: 504.
- . 2009b. Stata tip 82: Grounds for grids on graphs. *Stata Journal* 9: 648–651.
- . 2012a. Software update: gr42\_6: Quantile plots, generalized. *Stata Journal* 12: 167.
- . 2012b. Speaking Stata: Transforming the time axis. *Stata Journal* 12: 332–341.

- Klass, G. M. 2012. *Just Plain Data Analysis: Finding, Presenting, and Interpreting Social Science Data*. 2nd ed. Lanham, MD: Rowman & Littlefield.
- Kline, M. 1972. *Mathematical Thought from Ancient to Modern Times*. New York: Oxford University Press.
- Parzen, E. 1979. Nonparametric statistical data modeling. *Journal of the American Statistical Association* 74: 105–131.
- Playfair, W. H. 1801a. *Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure, and Debts of England during the Whole of the Eighteenth Century*. 3rd ed. London: Wallis.
- . 1801b. *The Statistical Breviary: Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe; Illustrated with Stained Copper-Plate Charts, Representing the Physical Powers of Each Distinct Nation with Ease and Perspicuity: to Which is Added, a Similar Exhibition of the Ruling Powers of Hindoostan*. London: Wallis.
- . 2005. *The Commercial and Political Atlas and Statistical Breviary*. Cambridge University Press: Cambridge. Edited by H. Wainer and I. Spence.
- Stillwell, J. 2010. *Mathematics and Its History*. 3rd ed. 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*.