

Preface

The *Stata Technical Bulletin* was started in 1991 as a journal by and for users of the statistical software Stata. As a Stata enthusiast, I subscribed from issue 1, started contributing in 1994, and became an associate editor in 1998. In 2001, a decision was made to transform the *Bulletin* into the *Stata Journal*, and I became one of the editors. The initial meeting setting up the journal was a working breakfast in a Boston hotel, otherwise made memorable by my being offered cottage cheese as an alternative to yogurt with my cereal. Among many other details, two suggestions followed in quick succession: that there should be a column in the *Journal* and that I should write it. I remain gratefully aware of the compliment implied.

I settled on a title, *Speaking Stata*, to match both the privilege of sounding off on my own Stata-linked preoccupations and prejudices and the purpose of explaining the use of Stata's language in a way that grew from, but went beyond, the excellent Stata documentation. As a columnist, I claim no more than certain interests and points of view; I take comfort in the simple thought that others are free to write on quite different topics and from quite different perspectives.

This book collects those columns published between 2004 and 2013 with a major theme of statistical graphics. My strong interest in graphics has grown steadily out of my education and experience as an academic geographer and a long-standing fascination with geometry and spatial patterns. Despite greatly renewed attention given to graphics over the last few decades within statistical science, I think it remains true that they are generally undervalued within several disciplines, but this is not the place to argue that view in any detail. The columns also include coverage of several little-known and even some apparently new ways of graphing data. A keen personal interest in the history of ideas, which surfaces in several columns, underlines the extent to which good ideas are persistently reinvented, so one can never be sure.

I have received strong and consistent counsel not to attempt any serious rewriting. The columns were certainly written one at a time and are intended to be read, mostly, one at a time. Unlike Dickens, Dumas, or Dostoyevsky, I have not been writing a novel in serial installments. Nevertheless, my own interests should give some coherence to the collection.

Learning statistics software, like learning statistics itself, can be a series of small struggles until enlightenment dawns and you wonder why on earth the idea appeared difficult in the first place. While writing these columns, I have tried to be clear rather than clever because I know from my own experience that several ideas in Stata, although beautifully simple when considered in the right way, are far from transparent at first

sight. Similarly, some repetitions remain in the hope that they provide reinforcement or serve the purposes of those with neither time nor inclination to read every column.

However, what is here is a little more than a straight reprinting. References to the Stata manuals have been updated. A few misprints have been corrected. (“Misprint” is British for “typo”.) Notes following this *Preface* identify software changes since the columns were first produced. In any case, readers interested in identifying the most recent versions of any user-written programs should first use `search`, `all` within Stata.

Any writer accumulates many debts that cannot all be discharged or even acknowledged. William (Bill) Gould is the person who conceived of Stata and set up the company that maintains and develops it. StataCorp is a company that aspires to excellence in everything. As I discovered long ago, it is grateful even to be told of all bugs and misfeatures, down to misprints in its manuals. I remain perpetually aware that I follow in Bill’s footsteps and stand on his shoulders.

I have also received much patient and detailed support from many other people at StataCorp. Vince Wiggins has borne the brunt of most of my specific graphics questions over the last decade, alerted me to some subtle tricks, and made encouraging noises throughout. Pat Branton, Lisa Gilmore, and their past and present colleagues have accommodated most of the quirks and dilatory submissions of an author with some strong prejudices about the English (N.B. not American) language and writing generally. Many in the Stata user community have asked good questions and provided most of the answers through encounters at users’ meetings, postings on Statalist, and publications in the *Stata Technical Bulletin* and the *Stata Journal*. My fellow Editor Joe Newton has proved a master of the timely quiet word of approbation. Various specific acknowledgments are recorded in individual columns, but they form a sadly incomplete list. Although it is, as always, invidious to single out a few names, those who know anything of their work would want to echo further special thanks to Kit Baum, Marcello Pagano, and Patrick Royston for their long-standing endeavors for the Stata user community.

At Durham University, several geographical colleagues have kept me honest by providing a series of statistical challenges within a fascinating variety of datasets. Ian Evans in particular has also been an appreciative Stata user for many years. The division of labor in our collaborations that he did most of the more challenging science and I did most of the more challenging Stata programming has proved mutually congenial. It has been a special pleasure to remain in collaboration with my own PhD supervisor for over 40 years.

Most personally, I dedicate this book to my wife, Irene, in recognition of all the encouragement and support she has provided over our time together.

Nicholas J. Cox

Note: What follows is a complete list of my publications on Stata that are mostly graphical. As of this writing, all material published more than three years ago is freely available at <http://www.stata-journal.com> in the case of the *Stata Journal* and at <http://www.stata.com/bookstore/individual-stata-technical-bulletin-issues/> in the case of the *Stata Technical Bulletin*. Graphical tips by me have been collected within Cox and Newton (2014).

SSG indicates reprinting in this book, and 119 indicates reprinting in Cox and Newton (2014).

- Cox, N. J. 1997. gr16.1: Convex hull plots. *Stata Technical Bulletin* 36: 2–3. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 25–27. College Station, TX: Stata Press.
- . 1997. gr22: Binomial smoothing plot. *Stata Technical Bulletin* 35: 7–9. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 36–38. College Station, TX: Stata Press.
- . 1997. gr24: Easier bar charts. *Stata Technical Bulletin* 36: 4–8. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 44–50. College Station, TX: Stata Press.
- . 1997. gr24.1. Easier bar charts: correction. *Stata Technical Bulletin* 40: 12. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, p. 58. College Station, TX: Stata Press.
- . 1997. gr26: Bin smoothing and summary on scatter plots. *Stata Technical Bulletin* 37: 9–12. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 59–63. College Station, TX: Stata Press.
- . 1998. sg85: Moving summaries. *Stata Technical Bulletin* 44: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 145–149. College Station, TX: Stata Press.
- . 1999. gr35: Diagnostic plots for assessing Singh-Maddala and Dagum distributions fitted by MLE. *Stata Technical Bulletin* 48: 2–4. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 72–74. College Station, TX: Stata Press.
- . 1999. gr41: Distribution function plots. *Stata Technical Bulletin* 51: 12–16. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 108–112. College Station, TX: Stata Press.
- . 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.
- . 2001. gr42.1: Quantile plots, generalized: update to Stata 7.0. *Stata Technical Bulletin* 61: 10–11. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 55–56. College Station, TX: Stata Press.
- . 2003. Software update: gr41_1: Distribution function plots. *Stata Journal* 3: 211.

- . 2003. Software update: gr41_2: Distribution function plots. *Stata Journal* 3: 449.
- . 2004. Software update: gr22_1: Binomial smoothing plot. *Stata Journal* 4: 490–491.
- . 2004. Software update: gr42_2: Quantile plots, generalized. *Stata Journal* 4: 97.
- . 2004. Speaking Stata: Graphing agreement and disagreement. *Stata Journal* 4: 329–349 (SSG).
- . 2004. Speaking Stata: Graphing categorical and compositional data. *Stata Journal* 4: 190–215 (SSG).
- . 2004. Speaking Stata: Graphing distributions. *Stata Journal* 4: 66–88 (SSG).
- . 2004. Speaking Stata: Graphing model diagnostics. *Stata Journal* 4: 449–475 (SSG).
- . 2004. Stata tip 6: Inserting awkward characters in the plot. *Stata Journal* 4: 95–96 (119).
- . 2004. Stata tip 12: Tuning the plot region aspect ratio. *Stata Journal* 4: 357–358 (119).
- . 2004. Stata tip 15: Function graphs on the fly. *Stata Journal* 4: 488–489 (119).
- . 2005. Software update: gr41_3: Distribution function plots. *Stata Journal* 5: 470–471.
- . 2005. Software update: gr42_3: Quantile plots, generalized. *Stata Journal* 5: 470–471.
- . 2005. Speaking Stata: Density probability plots. *Stata Journal* 5: 259–273 (SSG).
- . 2005. Speaking Stata: The protean quantile plot. *Stata Journal* 5: 442–460 (SSG).
- . 2005. Speaking Stata: Smoothing in various directions. *Stata Journal* 5: 574–593 (SSG).
- . 2005. Stata tip 21: The arrows of outrageous fortune. *Stata Journal* 5: 282–284 (119).
- . 2005. Stata tip 24: Axis labels on two or more levels. *Stata Journal* 5: 469 (119).
- . 2005. Stata tip 27: Classifying data points on scatter plots. *Stata Journal* 5: 604–606 (119).
- . 2006. Software update: gr26_1: Bin smoothing and summary on scatter plots. *Stata Journal* 6: 151.
- . 2006. Software update: gr42_4: Quantile plots, generalized. *Stata Journal* 6: 597.
- . 2006. Speaking Stata: Graphs for all seasons. *Stata Journal* 6: 397–419 (SSG).
- . 2007. Software update: gr0012_1: Density probability plots. *Stata Journal* 7: 593.

- . 2007. Speaking Stata: Turning over a new leaf. *Stata Journal* 7: 413–433 (SSG).
- . 2007. Stata tip 47: Quantile–quantile plots without programming. *Stata Journal* 7: 275–279 (119).
- . 2007. Stata tip 55: Better axis labeling for time points and time intervals. *Stata Journal* 7: 590–592 (119).
- . 2008. Speaking Stata: Between tables and graphs. *Stata Journal* 8: 269–289 (SSG).
- . 2008. Speaking Stata: Spineplots and their kin. *Stata Journal* 8: 105–121 (SSG).
- . 2008. Stata tip 59: Plotting on any transformed scale. *Stata Journal* 8: 142–145 (119).
- . 2009. Speaking Stata: Creating and varying box plots. *Stata Journal* 9: 478–496 (SSG).
- . 2009. Speaking Stata: Paired, parallel, or profile plots for changes, correlations, and other comparisons. *Stata Journal* 9: 621–639 (SSG).
- . 2009. Stata tip 76: Separating seasonal time series. *Stata Journal* 9: 321–326 (119).
- . 2009. Stata tip 78: Going gray gracefully: Highlighting subsets and downplaying substrates. *Stata Journal* 9: 499–503 (119).
- . 2009. Stata tip 82: Grounds for grids on graphs. *Stata Journal* 9: 648–651 (119).
- . 2010. Software update: gr0009_1: Speaking Stata: Graphing model diagnostics. *Stata Journal* 10: 164 (SSG).
- . 2010. Software update: gr0021_1: Speaking Stata: Smoothing in various directions. *Stata Journal* 10: 164 (SSG).
- . 2010. Software update: gr41_4: Distribution function plots. *Stata Journal* 10: 164.
- . 2010. Software update: gr42_5: Quantile plots, generalized. *Stata Journal* 10: 691–692.
- . 2010. Speaking Stata: Graphing subsets. *Stata Journal* 10: 670–681 (SSG).
- . 2010. Speaking Stata: The statsby strategy. *Stata Journal* 10: 143–151 (SSG).
- . 2011. Stata tip 102: Highlighting specific bars. *Stata Journal* 11: 474–477 (119).
- . 2011. Stata tip 104: Added text and title options. *Stata Journal* 11: 632–633 (119).
- . 2012. Software update: gr42_6: Quantile plots, generalized. *Stata Journal* 12: 167.

- . 2012. Speaking Stata: Axis practice, or what goes where on a graph. *Stata Journal* 12: 549–561 (SSG).
- . 2012. Speaking Stata: Transforming the time axis. *Stata Journal* 12: 332–41 (SSG).
- . 2013. Speaking Stata: Creating and varying box plots: Correction. *Stata Journal* 13: 398–400 (SSG).
- . 2013. Speaking Stata: Trimming to taste. *Stata Journal* 13: 640–666 (SSG).
- . 2014. Stata tip 119: Expanding datasets for graphical ends. *Stata Journal* 14: 230–235 (119).
- Cox, N. J., and N. L. M. Barlow. 2008. Stata tip 62: Plotting on reversed scales. *Stata Journal* 8: 295–298 (119).
- Cox, N. J., and A. R. Brady. 1997. gr25: Spike plots for histograms, rootograms, and time-series plots. *Stata Technical Bulletin* 36: 8–11. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 50–54. College Station, TX: Stata Press.
- . 1997. gr25.1: Spike plots for histograms, rootograms, and time-series plots: update. *Stata Technical Bulletin* 40: 12. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, p. 58. College Station, TX: Stata Press.
- Cox, N. J., and H. J. Newton, eds. 2014. *One Hundred Nineteen Stata Tips*. Stata Press, College Station, TX.
- Gray, J. P., and N. J. Cox. 1998. gr16.2: Corrections to condraw.ado. *Stata Technical Bulletin* 41: 4. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, p. 58. College Station, TX: Stata Press.
- Royston, P., and N. J. Cox. 2005. A multivariable scatterplot smoother. *Stata Journal* 5: 405–412.
- Sasieni, P., P. Royston, and N. J. Cox. 2005. Software update: sed9.2: Symmetric nearest neighbor linear smoothers. *Stata Journal* 5: 285.