Review of A Visual Guide to Stata Graphics by Mitchell

Ulrich Kohler
Wissenschaftszentrum Berlin
kohler@wz-berlin.de

Abstract. The new book by Mitchell (2004) is reviewed.

Keywords: gn0014, graphics, Stata texts

1 Introduction

With the arrival of Stata 8, users have been confronted by totally new and far more flexible graphics facilities. The flexibility of the new graphics engine, however, is accompanied by a more complicated syntax. Where previously there was one command with some options, there is now a command with subcommands and sub-subcommands, with options and suboptions holding arguments or lists of arguments. Correspondingly, the *Graphics Reference Manual* has now grown in length from 148 pages to 518 pages.

StataCorp has made several attempts to explain the new graphics language. They wrote an introduction for the manual ([G] **graph intro**), added clickable example graphs in the online documentation, and supplied pull-down menus in the new GUI. What is common to these attempts is that the user is guided from the command to the graph. In Mitchell's book, it is the other way around.

Mitchell lets readers pick out a concrete graph and then presents the command to produce it, along with explanations. In a sense, the book works like the graph palettes of some other software, in which users can point and click to select the graph they want. However, Mitchell's book goes one step further, as he also teaches the general concepts of the Stata graphics language.

2 Content and structure

A Visual Guide to Stata Graphics is an exceptional book. Not only is it printed in color—and around 1000 graphs are included—but the book is also exceptional in how its elements—text, commands, graphs, table of contents—are put together and in the amount, accessibility, and documentation of the online supplements distributed with it.

The chapters of the book are mainly ordered according to Stata's graph types, with sections on `graph twoway`, `graph matrix`, `graph bar`, `graph dot`, and `graph pie`. In addition, there are two chapters on graph options, which explain options to be used with most graphs—such as *axis_options* or *marker_options*—and explain options to be used with all graphs—such as *title_options*. General topics of the graphics language are

demonstrated in the introduction. Finally, the appendix contains material that is more advanced or probably less often used. Here are examples of specialized statistical graphs, notes on how to combine graphs, and a brief introduction to customized schemes.

While the outline of the book is fairly conventional, its realization is not. The book has three graphs on almost every page. Above each graph is the command that produced it, and alongside is a short description explaining the peculiarities of the command. Within each topic, many successive graphs are the same, except that each graph changes one specific aspect of its predecessor. In the command and the descriptive text for the succeeding graph, the part that differs is printed in color. Hence, looking at two successive graphs and the colored part of the command indicates directly the meaning of the respective language element.

The book layout restricts the explanations alongside each graph to a few sentences. The author therefore often gives cross-references to sections in which a specific language element is discussed in more detail. Naturally, a printed book does not allow you to click on cross-references, but as a substitute, the book prints a "visual table of contents" on the right margin of every odd-numbered page. This visual table of contents lists the chapters of the book and the sections of the chapter one is in. A colored marker indicates the reader's current position. This makes it extremely easy to follow cross-references and find a specific topic of interest.

All the graphs printed in the book are readily reproducible with the commands above them. The names of the dataset and the scheme file used are printed beside each graph. Both datasets and scheme files are distributed via the Internet, and they can be easily installed with Stata's `net` utilities. Moreover, the command `vguse` can be used to load the datasets from the Internet.

3 Goals and evaluation

A book review should determine whether an author achieved their goals, which requires being clear on what those goals were and what they were not. Mitchell's book, while it is about Stata graphics, is not about graphs. It does not explain what can be learned from graphs nor why one graph type should be preferred to another (for example, dot charts over bar charts). Moreover, while it is about designing Stata graphs, it is not about graphics design. There is complete silence about why you might want to design a graph in one way rather than another (for example, using hollow circles rather than filled circles in scatterplots). Naturally, such decisions can hardly be criticized, but one comment might be allowed: given the large number of weird graphs in publications and given that it is easy to misuse the capacities of Stata's new graphics engine, it would have been worthwhile to supply references to some books that do raise such questions.

What are the goals the author wanted to reach? I think there are three:

- Provide an overview of Stata's graphics capabilities.

- Let Stata users quickly find the `graph` command that produces the graph they have in mind.

- Teach the reader the general ideas of Stata's graphics language.

You can flip first of all through Mitchell's book, which in itself will give you an overview of Stata's graphics capabilities. You will see lots of bars and pies and lots of scatterplots and line plots, some with overlaid curves and some without. You will see plots in fancy colors and plots in black and white, and much more. Flipping through the book a second time, you will almost certainly see many surprising graphs and unknown features. The book thus definitely achieves the goal of providing an overview of Stata's graphics capacities, although the section on combining graphs deserves more prominence than being half-hidden in the appendix.

People picking up the book to look for a specific `graph` command can follow two strategies. First, they can flip through the book until they find a graph very similar to what they have in mind. After typing the printed `graph` command, they can look for a graph with the features not yet present in the graph, and so on. They will often produce the desired graph after just a few such iterations.

A second strategy is useful to those who already know which elements of the graphics language are needed but do not know the right syntax. Such users can again flip through the book until the colored marker of the visual table of contents appears at the desired place. A syntactical example of the language element in question will be nearby. In preparing this review, I used these strategies to find solutions for questions about graphs appearing on Statalist, and they worked pretty well. When looking up the syntax of odd graph options, it seems faster to use Mitchell's book than to use the online help or the manual.

The strength of Mitchell's book for looking up the right syntax is heavily based on the visual table of contents. Taking advantage of this tool, however, requires a certain amount of familiarity with the new graphics language. It is therefore an important question whether the book can also provide a more abstract understanding of the graphics language. The answer to this question is a slightly doubtful yes. A beginner working through the introduction should end up with a rough idea of the language. Playing with examples in the other chapters and reading the explanations should be enough to gain a deeper understanding. The slight doubt arises because it might be faster to understand the language structure by following the suggested reading in [G] **graph intro**.

What is not so evident in Mitchell's book is the interaction between producing graphs and other parts of data analysis. For many graphs, it is necessary to `generate` specific variables or to `collapse`, `xpose`, or `reshape` the data. Mitchell's book is impressive in covering what can be done with `graph`. The drawback of this is that it promotes the impression that what cannot be done with `graph` alone cannot be done at all. Such

a conclusion, however, would be a complete mistake. In the appendix, Mitchell shows several examples of graphs requiring some data management beforehand. Unfortunately, appendix material tends to get overlooked.

4   Conclusion

The main criticism of Mitchell's book therefore is that some of the material in the appendix is too interesting and too important to belong there. This comment should, however, not belie the strengths of the book. Let me end with a brief assessment. In his preface, Mitchell expresses the hope that readers like the book. I do. It is excellent.

5   Reference

Mitchell, M. 2004. *A Visual Guide to Stata Graphics*. College Station, TX: Stata Press.

About the Author

Ulrich Kohler is a sociologist at the Wissenschaftszentrum Berlin (Social Science Research Center), who has used Stata for several years. His research interests include social inequality and political sociology. With Frauke Kreuter, he is author of the forthcoming textbook *Data Analysis Using Stata*.