

Preface

The art of data visualization has advanced significantly since the turn of the millennium. In particular, accessibility and ease of use have taken center stage. Beautiful, complex, and sophisticated data visualizations are no longer limited to those with highly technical and specialized coding skills. Many software platforms now make it easy for anyone to start creating graphs and visualizing data. Now a much broader user base can find creative ways to reveal patterns and connections in data and, in turn, weave them into their stories and narratives.

Stata exemplifies both ease of use and powerful capabilities in data visualization. Continuous updates to Stata's graphics features, combined with a large and innovative user base, have transformed it from statistical software with basic graphic functions into a true graphics powerhouse. This evolution allows users to create an extensive array of sophisticated and visually compelling graphs with relative ease, leveraging both built-in functionalities and community-contributed commands.

This book aims to unleash that powerhouse and show users the wide range of statistical graphs that can be created with Stata. It is written in an informal style and contains no equations. The aim of the book is not to explain how each graph and statistical command works “under the bonnet”, and there is an assumption that users have a basic understanding of statistics. This book is not intended to be linear and read from beginning to end. It is designed to be flipped through casually, during which readers can identify a particular type of graph or a series of graphs they are interested in. They then focus on that content, either reading the text around the graph before applying the relevant code and idea to their own data or reading the entire chapter for a more comprehensive introduction and discussion of that type of graph.

Each graph is directly preceded by the command that generates it. This will help users quickly adapt the code to their own data and create similar-looking graphs. As much as possible, only Stata datasets are used. These come preinstalled with Stata and will already reside on local hard drives. In other instances, examples use online datasets from Stata's official website that are fast to download. Custom or bespoke data sources are avoided because they can be difficult to access or work with. Users will, therefore, notice repetition of certain datasets such as `auto.dta`, `citytemp.dta`, `nlsw88.dta`, or `uslifeexp.dta`. These datasets often have particular variables or data setups that make them useful for teaching purposes, and their repetition throughout the book will help users better understand how different graph types represent the same data.

The book is split into four major parts, of which three are related to statistical concepts around data visualization. The first part looks at graphs that focus on univariate (one variable) data. Here the emphasis is often on presenting distributions of single variables. The second part examines graphs for bivariate (two variables) data. These graphs often examine how two variables are related and how such data are jointly distributed. The third part presents graphs for multivariate (three or more variables) data, which aim to examine how many variables are jointly related. These are often highly specialized graphs that use some trick to insert additional data dimensions. The final part looks at some useful graphic capabilities that users might encounter in data analysis, including how regression results can be easily visualized or how maps can be created.

Each chapter first demonstrates a basic graph, after which a series of important options is highlighted. This book does not aim to present and explain the many options available for each graph command. However, some options matter more than others, and this book tries to highlight how selected options can significantly change a graph. Also note that many graphs have access to generic graph options that change marker colors, line shapes, aspect ratios, axis titles, etc. These options are not explained in detail, because it is assumed that users have some basic knowledge of the graphic capabilities in Stata. However, these options are highlighted and explained when used.

A vital feature of this book is that it goes beyond the standard, official Stata installation. This book aims to represent the wide range of graph commands written by other Stata users and available at the Statistical Software Components Archive. Each chapter will clarify what additional commands or packages need to be installed to be able to follow along with that chapter. All chapters are standalone, and the code will work from top to bottom. The chapters are designed so that readers can execute all the relevant code without worrying about previous or later chapters.

I hope you enjoy this book and keep it close when working with data. Stata learning and teaching have come a long way since I started learning many years ago, and I am glad to provide you with a friendly new resource to augment Stata's extensive documentation. Of course, you are always encouraged to read the relevant help file and examine Stata's manuals in depth for all the nitty-gritty details. However, to rapidly accumulate high-level knowledge of many different processes and abilities, you cannot beat the plethora of Stata Press books now available for users.

Best wishes,
Franz Buscha