CREATING MORE EFFECTIVE GRAPHS

CREATING MORE EFFECTIVE GRAPHS

A succinct and highly readable guide to creating effective graphs

NAOMI B. ROBBINS
NBR
Wayne, New Jersey

CHART HOUSE

Published by Chart House, Wayne, New Jersey.

Publisher's Cataloging in Publication

Robbins, Naomi B.
Creating more effective graphs / Naomi B. Robbins.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-9859111-2-6 (pbk.)
1. Statistics–Graphic methods. I. Title.

HA31.R535
519.5′02′1–dc22

Cover Designer: Uhlein Design
Publication Assistant: Sarah Klein

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Ed, Joyce, and Rich

CONTENTS

PREFACE xiii

PREFACE TO CHART HOUSE PRINTING xix

1 INTRODUCTION 1

1.1 What We Mean by an Effective Graph 3
1.2 General Comments 10
1.2.1 Captions 10
1.2.2 The Data We Plot 10
Summary 11

2 LIMITATIONS OF SOME COMMON CHARTS AND GRAPHS 12

2.1 Pie Charts 13
2.2 Charts with a Three-Dimensional Effect 19
2.3 Bar Charts: Stacked and Grouped 29
2.4 Difference between Curves 35
2.5 Bubble Plots 43
Summary 45

3 HUMAN PERCEPTION AND OUR ABILITY TO DECODE GRAPHS 46

3.1 Elementary Graphical Perception Tasks 47

3.2 Ordered Elementary Tasks 61
3.3 Role of Distance and Detection 63
Summary 63

4 SOME MORE EFFECTIVE GRAPHS IN ONE OR TWO DIMENSIONS 64

4.1 Distribution of One Variable 67
4.1.1 Strip Plots 67
4.1.2 Dot Plots 69
4.1.3 Histograms 77
4.1.4 Jittering 81
4.2 Comparing Distributions: Box Plots 86
4.3 Relationship of Two Variables: Scatterplots 95
4.4 Time Series 101
4.5 Line Graphs 109
4.6 Comments 112
Summary 113

5 TRELLIS GRAPHICS AND OTHER WAYS TO DISPLAY MORE THAN TWO VARIABLES 114

5.1 Alternative Presentations of Three Variables 115
5.1.1 Stacked Bar Chart 115
5.1.2 Labeled Scatterplot 117
5.1.3 Trellis Display 119
5.2 More than Three Variables 121
5.2.1 Superposed Data Sets 121
5.2.2 Trellis Multipanel Displays 125
5.2.3 Scatterplot Matrices 127
5.2.4 Mosaic Plots 131
5.2.5 Linked Micromaps 137
5.2.6 Parallel Coordinate Plots 141

5.2.7 Nightingale Rose 145
5.2.8 Financial Plot 151
5.3 Comments 152
Summary 153

6 GENERAL PRINCIPLES FOR CREATING EFFECTIVE GRAPHS 154

6.1 Terminology 157
6.2 Visual Clarity 159
6.2.1 Clarity of Data 159
6.2.2 Clarity of Other Elements 179
6.3 Clear Understanding 197
6.4 General Strategy 224
Summary 225

7 SCALES 226

7.1 Aspect Ratio: Banking to 45° 229
7.2 Scales: Must Zero be Included? 232
7.3 When to Use Logarithmic Scales 243
7.4 Scale Breaks 257
7.5 Using Two *Y* Scales 263
7.6 Data Hidden in the Scales 269
7.7 Other Principles Involving Scales 277
Summary 291

8 APPLYING WHAT WE'VE LEARNED: *BEFORE* AND *AFTER* EXAMPLES 292

8.1 Grouped Bar Chart 295
8.2 Ten Small Graphs 299
8.3 Radar Chart 308

8.4 Multiple Pie Charts 317
8.5 Tables 322
Summary 327

9 SOME COMMENTS ON SOFTWARE 328

9.1 Statistical Software: S Language 328
9.2 Drawing Programs: Illustrator 330
9.3 Spreadsheets: Excel 331
9.3.1 Moving an Axis in Excel 333
9.3.2 Line Charts with Uneven Time Intervals 335
9.3.3 Dot Charts from Excel 337
9.3.4 Data Labels with Excel 339
Summary 341

10 QUESTIONS AND ANSWERS 342

1 When should I use a table, and when should I use a graph?
2 Should I use different graphs for presentations and for written reports?
3 How do graphs for data analysis and graphs for communication differ?
4 What should I use instead of pie charts?
5 What if I just want an impression of the direction of the data? Then may I use three-dimensional charts?
6 I use three-dimensional charts but I include data labels. That's OK, isn't it?
7 I want my graphs to attract the reader's attention. How should I decorate them?
8 Why do you think we see so many bad graphs?
9 When should I use each type of graph?

APPENDIX A: CHECKLIST OF POSSIBLE GRAPH DEFECTS 375

APPENDIX B: LIST OF FIGURES WITH SOURCES 378

REFERENCES 389

INDEX 395

PREFACE

The idea for this book came while reading *The Elements of Graphing Data* by William S. Cleveland. Cleveland addresses his book to people in science and technology. As I read it, I kept thinking that his ideas deserved a wider audience: that people in the business world, the financial world, the world of nonprofits, and many other groups would benefit from his ideas and methods. *Creating More Effective Graphs* addresses this need.

Goals

This book is intended to be quick and easy to read. It is not intended to be a substitute for the classic books in the field: Cleveland's *The Elements of Graphing Data* and Edward Tufte's *The Visual Display of Quantitative Information*. Rather, it is hoped that it will serve as an introduction to the subject, encouraging readers to study the classics. The number of confusing, poor graphs one sees is staggering. However, at a recent visit to a business school library, I was delighted to see how many more clear, accurate, well-designed graphs appear in journals currently than they did a decade ago. The credit for this improvement goes to the influence of these authors.

Audience

This book offers guidance to readers with a wide range of backgrounds. As a result, at times I review some high school math, as I do in Chapter 4 when reviewing logarithms. Some readers will appreciate this review. At other times I include more complex technical details for the more technical readers. Nontechnical readers can feel comfortable skipping these sections without worrying about being able to understand later parts of the book. This book is a collection of independent examples and advice; it is not like a course, where material builds on earlier sections, keeping you lost forever if you do not understand a fundamental concept. So please skip the paragraphs that are too easy or too hard for you and benefit from the rest. (Note that Chapter 3 provides basic information useful when reading later chapters.)

Terminology

Unfortunately, there is little standard terminology in the literature on data graphs. Words such as *chart*, *graph*, and *plot* are often used interchangeably. Some authors have given precise meanings to these words and have used them consistently. The problem is that they then no longer use the term of the creator of the graph. For example, John Tukey (1977) introduced a box and whisker plot in his 1977 book. For over 25 years, statisticians have called them box plots or box and whisker plots. I am unwilling to call them box graphs for the sake of consistent terminology. So my choice has been to continue with the name that the originator of the technique used. The term *graph* in this book always means data graph, as opposed to the nodes and edges of a graph in the branch of mathematics called graph theory.

Examples

You will notice that a disproportionate number of examples come from the world of museums. There are a number of reasons for this. I have found that the museum community is very interested in the accurate display of information and wants to display information effectively. Their data are less likely to be private, so they are more willing to allow me to share their data. We all understand what we mean by their terms, so I don't have to waste space defining the concepts behind the data. Finally, since many of us go to museums, we can relate to the examples. If I used examples from the pharmaceutical industry, people in nonprofits or marketing would feel that the examples do not apply to them.

Wherever possible, the figures are real figures with real data. In some cases, using real data would violate confidentiality, so labels or data have been changed. In other cases, I have reproduced real problems that are frequently seen with simulated data.

How to Read

You can start at the beginning and read through the text or use the book as a reference to look up the topics you need. Either way, I recommend reading Chapter 3 before later chapters, since it helps to understand why some presentations work and some do not. Reading the entire book is not a big investment of time.

Chapter 1 explains what we mean by an effective graph.

Chapter 2 shows problems with many of the common graphs that are ubiquitous today.

Chapter 3 briefly discusses the tasks we perform when we decode information from a graph, and which of these tasks we perform well.

Chapters 4 and 5 present methods. Chapter 4 describes graphs with one or two variables that are more effective than those of Chapter 2. Graphs with more than two variables are described in Chapter 5.

Chapters 6 and 7 contain principles. **Chapter 6** presents general principles for creating effective graphs. Principles for choosing scales appear in **Chapter 7**.

Chapter 8 applies what we have learned by looking at *before* and *after* examples.

Chapter 9 contains some comments about software.

Chapter 10 includes questions and answers.

Appendix A is a checklist of graph defects.

Appendix B lists all figures and their sources.

The book contains examples of both good and bad graphs. It should be clear from reading the book which is which. However, to prevent readers who just skim the book from thinking that I recommend the bad graphs, the icon you see on the margin is placed near some graphs. The absence of an icon does not necessarily mean that the graph form or the editing choices are recommended.

Tools Needed

All of the figures drawn for this book were produced on home equipment and print well on an inkjet printer. The one luxury was powerful software, S-Plus, available from Insightful

Corporation. Equally powerful open-source software, R, is freely available to download. It does not take a large budget to improve your graphs.

Recognizing that many people make graphs with Microsoft Excel, much of the advice applies to Excel users, and I provide some tips for Excel users in Chapter 9.

Acknowledgments

You will soon become aware that every chapter in the book has been influenced by the work of William Cleveland. I cannot thank Bill enough for his generosity in sharing his ideas, encouraging me in this field, and allowing me to quote his materials.

As deadlines approached, Linda Clark helped to produce some of the figures. The list of figures in Appendix B credits her for the ones she drew. Paul Murrell provided some code for Figure 8.9 before the documentation for the features used was available. Kenneth Klein wrote a macro to allow Excel users to draw dot plots. Marc Tracey provided me with information on Illustrator for Chapter 9.

Thanks to all who allowed me to use their unpublished data or figures: Edith Flaster, the Monterey Bay Aquarium, Beverly Serrell, the St. Louis Science Center, Marc Tracey, private clients, and anonymous contributors. Thanks also to the Center for the Study of Philanthropy, the Graduate Center, CUNY, for its contribution.

This project would never have gotten off the ground without the encouragement of Steve Quigley, Executive Editor at Wiley. It benefited throughout from the top-notch skills of the Wiley team, from the copyeditor to the compositor. Special

thanks go to Angioline Loredo, Associate Managing Editor. It was a pleasure to work with her.

The book benefited greatly from reviewers, who gave generously of their time to tell me what they liked and didn't like, and what they understood and didn't understand: Edith Flaster, Anne Freeny, Bert Gunter, Amy Juviler, Nancy Klujber, Melanie Meharchand, Beth Lisberg Najberg, Edward Robbins, Joyce Robbins, Richard Robbins, Ervin Schoenblum, and an anonymous reviewer from Wiley. I appreciate their candid remarks.

PREFACE TO CHART HOUSE PRINTING

It has been an exciting eight years since *Creating More Effective Graphs* was first published at the end of 2004. During this time, there have been dramatic advances in interactive and dynamic graphs, with great interest in the growing field of data visualization. Data journalism is thriving and infographics appear everywhere. Higher processing power has enabled businesses, government and society to derive previously unavailable insights from large data sets. New software vendors have appeared offering choices that fall between the point and click of spreadsheets and the programming necessary for programming languages. And social media platforms have facilitated a thriving online conversation about, and called attention to, data visualization.

Nonetheless, the foundations for visualizing data remain unchanged. The core principles about which I wrote eight years ago — designing effective graphs to communicate data — are just as applicable today, and provide a necessary foundation for any data visualization work. This work has stood the test of time and is more relevant than ever.

Even though it is more economical to use color today than it was when the book was first printed, there are still journals that only publish in black and white and many people produce figures that will be reproduced on black and white copy machines. Therefore, there still is a need to learn how to

create graphs that communicate well without the use of color. That said, some guidelines for using color appear on my Web site, http://www.nbr-graphs.com.

This volume is not a new edition; it is a reprinting of the original edition with a new publisher.

For those who are interested, I post about graphs and other areas of interest to me at www.twitter.com/nbrgraphs. I invite you to read these posts and join the conversation.

1 Introduction

In today's world we are overwhelmed with data. Graphs can be incredibly powerful tools in creating order from the chaos of numbers. A basic knowledge of graphing techniques is needed to ensure that data are presented effectively.

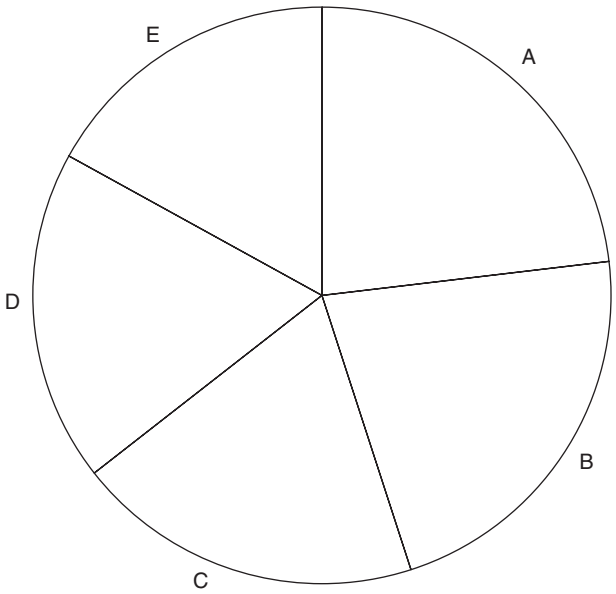
This book will teach you:

- How to make clear and accurate graphs that improve understanding of data
- How to avoid common problems that cause graphs to be ineffective, confusing, or even misleading
- When to use new graphing techniques to simplify complex data presentation
- How to be more critical and analytical when viewing graphs

You will learn guidelines and principles for creating effective graphs. These principles are to numbers what the rules of grammar are to words. Not many people today write without a computer. But we all know that being skilled at using word processors does not make us great writers. Similarly, learning how to use a software package to make graphs is not enough to become an effective communicator of numerical data.

Creating More Effective Graphs, by Naomi B. Robbins
ISBN 978-0-9859111-2-6

Fig. 1.1 Similar Pie Wedges

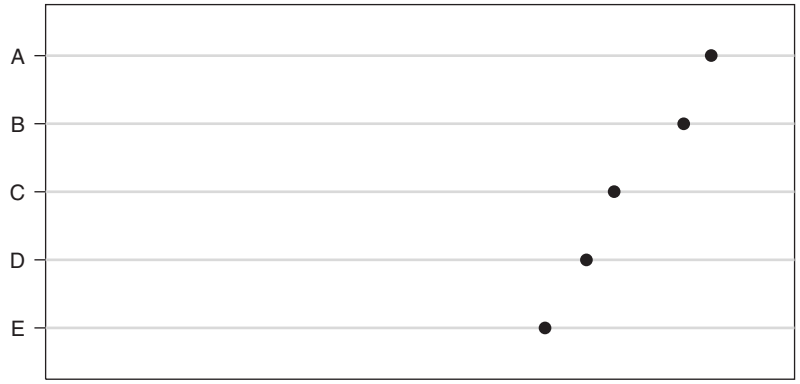


Whether you draw graphs manually or with a computer, and regardless of the software package you use, increasing your knowledge of the principles of effective graphs will greatly improve your work.

1.1 WHAT WE MEAN BY AN EFFECTIVE GRAPH

We begin by discussing what we mean by an *effective graph*. Figure 1.1 is the familiar *pie chart*. This one has five wedges labeled A through E. Study this chart and try to place the wedges in size order from largest to smallest. Use a pencil and paper to write down your results.

Fig. 1.2 Similar Pie Wedges: Dot Plot



The same data are plotted in Figure 1.2, this time using a *dot plot*. Once again take out paper and pencil and order the size of categories A through E.[1] Most of you probably had a lot of trouble placing the wedges of the pie chart in size order. Some of you may have found this to be quite frustrating, even though a few of you probably had no trouble at all. But even those of you who could order the wedges of the pie chart easily must admit that this task is much easier using a dot plot. Cleveland (1984) introduced dot plots to take advantage of the results of experiments on human perception and the decoding of graphical information. In Chapter 3 I discuss this topic briefly.

[1] I have left out the tick marks and labels since a reviewer suggested that it was an unfair comparison to show tick labels on the dot plot and no labels on the pie chart. Tick mark labels help you estimate the values, which is not the task that you were asked to do.

Now we're ready to define what we mean by an *effective graph*. One graph is more effective than another if its quantitative information can be decoded more quickly or more easily by most observers. Here, Figure 1.2, the dot plot is more effective than Figure 1.1, the pie chart.

This definition of effectiveness assumes that the reason we draw graphs is to communicate information—but there are actually many other reasons to draw graphs. Figures help

to keep an audience attentive during presentations. A graph may make a page in a document more attractive and inviting by adding variety and avoiding a totally gray page of text, thereby increasing readership. This book and our definition of effectiveness do not address these other reasons except in the questions and answers in Chapter 10. There you will see that we should communicate clearly and accurately even if our primary purpose for including a graph is one of the other reasons that I have listed.

Fig. 1.3 Similar Pie Wedges: Table

A	23.0
B	22.0
C	19.5
D	18.5
E	17.0
Total	100.0

Sometimes graphs do not provide the best solution for presenting data. A table provides another way to show the data in Figures 1.1 and 1.2. The table in Figure 1.3 not only shows the exact values of the categories, but also shows the total. Tables are preferable to graphs for many small data sets.

1.2 GENERAL COMMENTS

1.2.1 Captions

Cleveland (1994) says that a graph should include a caption that draws attention to important features of the data and describes conclusions that are drawn from the data. Why, then, although we use identifying figure headings in this book, did we not use captions? It is because the style of the book is to make you think about graphs and answer questions before you read the material that normally is contained in a caption.

1.2.2 The Data We Plot

The information to be plotted may be quantitative or categorical. Quantitative variables have numerical values (e.g., heights, measurements, salaries). Categorical variables have labels as values (e.g., gender, country, occupation).

An early step in any statistical analysis, including the presentation of data, is to check and clean the data. Are there typos? Do the numbers make sense? Why are a few data points far from the others? Are there inconsistencies, such as a birth date later than the corresponding death date? The advice in this book applies *after* the data have been checked. We assume that the numbers are correct and discuss how to display them effectively. However, another value of graphing is that graphs are a useful tool in helping to check data. We will see several examples where graphs have helped to identify incorrect data. Best (2001) discusses how numbers get distorted and take on lives of their own.

SUMMARY

One graph is more effective at communicating quantitative information than another if most readers can decode the quantitative information from it more quickly or more easily. Tables are useful for displaying small data sets.