

The Practice of Data Visualization (Course Notes)

Concepts and Tools for Your Visual Data Endeavours

Patrick Boily, Stephen Davies, Jen Schellinck

June 2023

Quadrangle | Data Action Lab

This work is licensed under a [Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International License](#) [↗](#) .

Below is a human-readable summary of (and not a substitute for) the license. Please see [this page](#) [↗](#) for the full legal text.

You are free to:

Share – copy and redistribute the material in any medium or format

Remix – remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution – You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike – If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions – You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

This one goes out to the “Welsh” contingent: Elwyn, Llewellyn, and Gwynneth. Your world is going to be a whole lot different than Bronwyn’s and mine was; maybe data can even help make some of it better. But one thing’s for sure: data is not going away any time soon – best be prepared.
(Patrick)

As W. Edwards Deming once said, "In God we trust, all others bring data"; Douglas Adams remarked, "but which god and why?". This book could be dedicated to all the gods, goblins, faeries, bogeymen, and fantastical beasts that sneak into data repositories at night to twist and break things, ensuring that I have a job helping to fix it all the next day! Transitioning from fantastical stories to real ones, I could also quote Dan Heath, "Data are just summaries of thousands of stories – tell a few of those stories to help make the data meaningful". This book is actually dedicated to the people in my life that have helped me make my own life story meaningful – Kristi, Piper, Findlay, Tilly, Tristan, Tony, Margaret, and Carol. (Stephen)

To my students, both young and old. You are the next generation of data interpreters. I look forward to seeing what you do. (Jen)

Happy reading!

Preface

The Practice of Data Visualization [PDV] arose once we realized, after a multitude of data-focused training sessions offered over the years, that we were always trying to cram more (and more!) material in our courses. In a very pragmatic way, it provides us with a way to be complete without drowning our audience; a dutiful companion to our various workshops, it also stands on its own, readable from cover to cover in and of itself.

We think that PDV will give its readers a flavour for the data visualization ecosystem, from some of its history and challenges, for the types of tools that analysts use to display data insights and to tell data stories. Of course, the book cannot be exhaustive, but it is comprehensive, enough so that what we have not covered will not come as a shock whenever it is encountered way over yonder on computer screens and newspapers.

It also fits within the context of a larger series of data science training books (*Data Understanding, Data Analysis, and Data Science* [1, DUDADS]), whose overall goal is to survey the data science landscape while leaving no glaring hole – from programming, mathematical, and statistical pre-requisites to data streams, natural language processing, and data streams, and through machine learning, big data, feature engineering, and the likes.

But PDV is not an encyclopaedia or a reference book; the self-contained chapters have an underlying flow:

1. in *A Data Visualization Primer* (part I), we introduce the fundamental notions of data visualization, what we would share with an audience that no prior knowledge of the discipline and only a small amount of time to devote to its study, namely: a bit of historical perspective, Tufte’s fundamental principles of analytical design, and an introduction to dashboards (*Data Visualization Essentials*, chapter 1); a first look at data exploration and at multivariate charts (*Data Visualization and Exploration*, chapter 2), and examples of misleading charts and of (subjectively) great and awful visualizations (*Hall-of-Fame/Hall-of-Shame*, chapter 3);
2. in *The Foundations of Visual Design* (part II), we present guidelines that will help analysts (and actors with data-adjacent roles) design and recognize effective visualizations, including: the “celebrated” Gestalt theory (*The Mechanics of Visual Perception*, chapter 4); chart constraints dictated by memory limitations, colouring and other aesthetic considerations, decluttering, and the organizing principle of the grammar of graphics (*Visual Design and Data Charts*, chapter 5), and alternatives to visualization and practical accessibility suggestions (*Universal Design and Accessibility*, chapter 6);
3. which naturally segue into *Visualization and Storytelling* (part III), where we describe: the constituents of this most human (and ancient) of activities [to use visualizations as support for stories, we must first understand stories] (*Stories and Storytelling*, chapter 7), and what we pragmatically elect to omit from visualizations and dashboards in order to produce impactful charts (*Effective Storytelling Visuals*, chapter 8);
4. finally, in *Some Practical Aspects of Data Visualization* (part IV), we throw off the shackles of visualization agnosticism on which the rest of the book has relied and focus on specific techniques (*Visualization Toolbox*, chapter 9), tools, and practical aspects, illustrated through the vagaries of R, ggplot2, and Power BI (*Visualization Software*, chapter 10; *Basic Visualizations in R*, chapter 11; *ggplot2 Visualizations in R*, chapter 12; *Power BI for Beginners*, chapter 13, with thanks to Tristan Shaeen).

May your visualizations always display in your favour!

Patrick Boily, Wakefield
Stephen Davies, Manotick
Jen Schellinck, Ottawa
June 2023

Contents

- I A Data Visualization Primer 1**
 - 1 Data Visualization Essentials 3**
 - 1.1 Historical Perspectives 3
 - 1.2 Infographics & Visualizations 6
 - 1.3 Analytical Design 7
 - 1.4 Minard’s March to Moscow 20
 - 1.5 Dashboards 24
 - 2 Data Visualization and Exploration 31**
 - 2.1 Exploratory Data Analysis 32
 - 2.2 Workhorse Visualizations 38
 - 2.3 Multivariate Observations 42
 - 2.4 Communicating Results 47
 - 3 Hall-of-Fame / Hall-of-Shame 49**
 - 3.1 Misleading Charts 49
 - 3.2 Home Runs and Strike Outs 52

- II The Foundations of Visual Design 61**
 - 4 The Mechanics of Visual Perception 63**
 - 4.1 Background 64
 - 4.2 Gestalt Principles 68
 - 4.3 Gestalt Guidelines 76
 - 4.4 Examples 77
 - 5 Visual Design and Data Charts 81**
 - 5.1 Visualization and Memory 81
 - 5.2 Aesthetic Considerations 84
 - 5.3 Decluttering 90
 - 5.4 The Grammar of Graphics 91
 - 5.5 Additional Examples 94
 - 6 Universal Design and Accessibility 101**
 - 6.1 Perception Alternatives 102
 - 6.2 Colourblindness 105
 - 6.3 General Resources 106
 - 6.4 Practical Suggestions 107

III Visualization and Storytelling	109
7 Stories and Storytelling	111
7.1 What are Stories?	111
7.2 Elements of Storytelling	127
7.3 Form and Structure	133
7.4 How to Tell a Story	139
8 Effective Storytelling Visuals	149
8.1 Stories and Illustrations	149
8.2 Data and Stories	154
8.3 Evolving a Story Chart	160
8.4 Anatomy of a Dashboard	167
IV Some Practical Aspects of Data Visualization	175
9 Visualization Toolbox	177
9.1 Exploration Visualizations	179
9.2 Presentation Visualizations	183
9.3 The Rest of The Landscape	188
9.4 Misc. & Charts to Avoid	197
10 Visualization Software	199
10.1 GUI Tools	199
10.2 Command Line Tools	201
10.3 Future Tools	202
11 Basic Visualizations in R	211
11.1 Scatterplots	211
11.2 Barplots	213
11.3 Histograms	216
11.4 Curves	218
11.5 Boxplots	219
11.6 Examples	220
12 ggplot2 Visualizations in R	235
12.1 ggplot2's Grammar	235
12.2 ggplot2 Miscellenea	244
12.3 Examples	252
13 Power BI for Beginners	309
13.1 Importing & Modeling Data	312
13.2 Calculations and Charts	316
13.3 Organizing Data	328
13.4 Data Wrangling	337
Exercises	347
Index	363
References	367