

# **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



<b>III Visualization and Storytelling</b>	<b>109</b>
<b>7 Stories and Storytelling</b>	<b>111</b>
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
<b>8 Effective Storytelling Visuals</b>	<b>149</b>
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
<b>IV Some Practical Aspects of Data Visualization</b>	<b>175</b>
<b>9 Visualization Toolbox</b>	<b>177</b>
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
<b>10 Visualization Software</b>	<b>199</b>
10.1 GUI Tools . . . . .	199
10.2 Command Line Tools . . . . .	201
10.3 Future Tools . . . . .	202
<b>11 Basic Visualizations in R</b>	<b>211</b>
11.1 Scatterplots . . . . .	211
11.2 Barplots . . . . .	213
11.3 Histograms . . . . .	216
11.4 Curves . . . . .	218
11.5 Boxplots . . . . .	219
11.6 Examples . . . . .	220
<b>12 ggplot2 Visualizations in R</b>	<b>235</b>
12.1 ggplot2's Grammar . . . . .	235
12.2 ggplot2 Miscellenea . . . . .	244
12.3 Examples . . . . .	252
<b>13 Power BI for Beginners</b>	<b>309</b>
13.1 Importing & Modeling Data . . . . .	312
13.2 Calculations and Charts . . . . .	316
13.3 Organizing Data . . . . .	328
13.4 Data Wrangling . . . . .	337
<b>Exercises</b>	<b>347</b>
<b>Index</b>	<b>363</b>
<b>References</b>	<b>367</b>



# List of Figures

1.1	A few historically significant data visualizations (I)	4
1.2	A few historically significant data visualizations (II)	5
1.3	Data visualizations and infographics	7
1.4	2012 Health and Wealth of Nations (Gapminder)	8
1.5	Comparisons in the Gapminder chart: country-to-country	9
1.6	Comparisons in the Gapminder chart: country-to-country overlap	10
1.7	Comparisons in the Gapminder chart: country-to-average	10
1.8	Approximate line of best fit for the Gapminder chart	11
1.9	Close-up, bottom right quadrant	12
1.10	Potential outliers in the Gapminder chart	12
1.11	Potential clusters in the Gapminder chart	14
1.12	Non-integrated Gapminder chart	15
1.13	Legend inset for the Gapminder chart	18
1.14	2013 Health and Wealth of Nations	19
1.15	Minard's map of Napoleon's Russian campaign	20
1.16	Layering of six variables in Minard's <i>March to Moscow</i>	21
1.17	Change over time in Minard's <i>March to Moscow</i>	22
1.18	Use of colour to emphasize the inflection point in Minard's <i>March to Moscow</i>	22
1.19	Overlay of Minard's <i>March to Moscow</i> on the geographical map of the region	23
1.20	Visual hierarchy in Minard's <i>March to Moscow</i>	23
1.21	Exploratory dashboard showing metrics relating to various cities ranked on the Global Cities Index	26
1.22	Exploratory dashboard for the NHL draft class of 2015	27
1.23	Anonymous 'ugly' dashboards	28
1.24	'Zen' dashboard I	29
1.25	'Zen' dashboard II	30
2.1	The (messy) analytical process	32
2.2	Pre-analysis uses of data visualization	33
2.3	Exploratory data analysis of the algae blooms dataset	37
2.4	Rug plot	38
2.5	Histogram and stacked histograms	38
2.6	Effect of histogram bin size	39
2.7	Line charts	40
2.8	Boxplots	41
2.9	Minimalist NASA CM1 scatterplot	42
2.10	Deluxe NASA CM1 scatterplot	43
2.11	Choropleths of US elevation	44
2.12	Network diagram of European language lexical distance	45
2.13	Physical visualization of people's demographics	46
2.14	Data visualization suggestions, by question type	47
3.1	Misleading chart: cost of gas	49
3.2	Misleading chart: UK voting results	51
3.3	NBA FG% against league average	53
3.4	Hertzprung-Russel diagram	54
3.5	A model of breast cancer causation	55

3.6	World's population "density" map . . . . .	56
3.7	Mapping paid paternity leave . . . . .	57
3.8	Language map of Europe . . . . .	58
3.9	COVID incidence in Georgian (US) counties . . . . .	59
3.10	Gubernatorial recall process in the United States . . . . .	60
4.1	Record of eye movement during free examination . . . . .	63
4.2	Eye movement on a chart . . . . .	64
4.3	How many 6s are there in this image? (I) . . . . .	64
4.4	How many 6s are there in this image? (II) . . . . .	65
4.5	How many 8s are there in this image? . . . . .	65
4.6	Pre-attentive features . . . . .	66
4.7	Combination of pre-attentive attributes . . . . .	67
4.8	How many 6s, 4s, 9s, etc. are found in this image? . . . . .	67
4.9	Ordering the digits to count them, with colours . . . . .	67
4.10	Re-arranging the digits to determine the relative frequencies of the digits . . . . .	68
4.11	Gestalt principle: simplicity . . . . .	69
4.12	Gestalt principle: continuation . . . . .	69
4.13	Gestalt principle: proximity (I) . . . . .	70
4.14	Gestalt principle: proximity (II) . . . . .	70
4.15	Gestalt principle: similarity (I) . . . . .	71
4.16	Gestalt principle: similarity (II) . . . . .	71
4.17	Gestalt principle: focal point (I) . . . . .	71
4.18	Gestalt principle: focal point (II) . . . . .	72
4.19	Gestalt principle: isomorphic correspondence . . . . .	73
4.20	Gestalt principle: figure / ground duality . . . . .	73
4.21	Gestalt principle: common fate . . . . .	74
4.22	Gestalt principle: closure . . . . .	74
4.23	Gestalt principle: uniform connectedness . . . . .	75
4.24	Combinations of Gestalt principles . . . . .	76
4.25	Gestalt examples: simplicity, continuity, isomorphic correspondence, common fate . . . . .	77
4.26	Gestalt example: figure / ground duality, focal point, isomorphic correspondence . . . . .	78
4.27	Gestalt example: figure / ground duality, continuity . . . . .	78
4.28	Gestalt example: simplicity, figure / ground duality, enclosure, isomorphic correspondence . . . . .	79
4.29	Gestalt example: similarity, continuity . . . . .	80
5.1	Using pre-attentive processes to easily find Waldo . . . . .	82
5.2	Failure of pre-attentive processes to easily find Waldo . . . . .	82
5.3	Illustration of the difference between a pre-attentive and an attentive search . . . . .	83
5.4	The main additive colour wheels . . . . .	85
5.5	Zeileis, Hornik, and Murrell's 24-colour palette . . . . .	86
5.6	Tol's 14-colour rainbow scheme . . . . .	87
5.7	Kelly's 22 colours of maximum contrast . . . . .	87
5.8	Carter and Carter's 26 colours of high contrast . . . . .	88
5.9	Competing colour palettes . . . . .	88
5.10	Sizing considerations . . . . .	89
5.11	Standard 4-chart layout . . . . .	89
5.12	Alternative 4-chart layouts . . . . .	90
5.13	A cluttered chart . . . . .	91
5.14	A decluttered chart . . . . .	91
5.15	The layered grammar of graphics . . . . .	92

5.16	Gapminder’s 2012 Health and Wealth of Nations . . . . .	93
5.17	Scatterplot matrices of the iris dataset . . . . .	94
5.18	Decluttering a chart . . . . .	94
5.19	Gapminder: proportion of population by income . . . . .	95
5.20	Gapminder: life expectancy relative to fertility rates . . . . .	96
5.21	Gapminder: infant mortality by continent . . . . .	97
5.22	Gapminder: HIV infection rates relative to income per capita . . . . .	98
5.23	Gapminder: fertility rates by continent . . . . .	99
5.24	Gapminder: infant mortality relative to fertility rates . . . . .	100
6.1	Braille symbols . . . . .	101
6.2	Sonification: TRAPPIST-1 . . . . .	102
6.3	Sonification: LHC . . . . .	103
6.4	Physicalization: khipus . . . . .	104
6.5	Physicalization: thermoform . . . . .	104
6.8	Colourblind-friendly colour scales . . . . .	105
6.6	Gestalt principle: focal point (reprise) . . . . .	105
6.7	Colourblindness effects . . . . .	105
6.9	Short graphs vs wide graphs . . . . .	108
7.1	Distorted projection of the night sky . . . . .	112
7.2	da Varese’s <i>Map of the Heavens</i> . . . . .	113
7.3	Modern IAU constellations (I) . . . . .	113
7.4	Modern IAU constellations (II) . . . . .	114
7.5	Cassiopeia’s night sky neighbourhood . . . . .	114
7.6	Failure to move beyond flattening the curve . . . . .	116
7.7	Schematic diagram of systems thinking . . . . .	117
7.8	Written by a Kid (La Munkya) . . . . .	119
7.9	How do we get to zero greenhouse gas emissions (infographic) . . . . .	120
7.10	Same data, different conclusions . . . . .	122
7.11	Visual Cinnamon data story (?) . . . . .	122
7.12	Visual Cinnamon data story (!) . . . . .	123
7.13	Representation of Axis bullets on returning Allied bombers . . . . .	125
7.14	Anti-vaccine “argument” against COVID-19 vaccines. . . . .	126
7.15	Full data argument in favor of COVID-19 vaccines. . . . .	126
7.16	Newspaper article: Wakefield nurse fires up Freedom Convoy . . . . .	130
7.17	Traditional constellations in 4 cultures . . . . .	131
7.18	Periodic table of storytelling . . . . .	134
7.19	Storytelling molecules . . . . .	134
7.20	Playing with tropes . . . . .	136
7.21	Narrative structures . . . . .	138
7.22	Robinah Babirye’s <i>This is Just the Beginning</i> . . . . .	143
7.23	Hans Rosling’s <i>The Best Stats You’ve Ever Seen</i> . . . . .	144
7.24	Yash Gupta’s Bank dashboard . . . . .	144
7.25	Parenting memes . . . . .	145
7.26	This is That’s <i>Thought Leader</i> . . . . .	145
7.27	Charlie Booker’s <i>How to Report the News</i> . . . . .	146
7.28	Emily Surname’s <i>Immigration Report</i> . . . . .	146
7.29	Gigi D.G.’s <i>Cucumber Quest</i> . . . . .	146
7.30	The top 6 reasons this infographic is just wrong enough to sound convincing . . . . .	147

8.1	Interpretation of the coffee cup image . . . . .	150
8.2	Interpretation of the 3 amigos image . . . . .	151
8.3	Population distribution in the Americas in 2020 . . . . .	151
8.4	Monthly attrition rate in May 2019 . . . . .	151
8.5	Staff and students by ethnicity at the University of Washington . . . . .	152
8.6	Monthly attrition rates, revisited (2019) . . . . .	152
8.7	COVID-19: The Story of the Impossible Train (Illustrated) . . . . .	153
8.8	COVID-19 vaccination rates . . . . .	154
8.9	History of atomic models . . . . .	155
8.10	Why Numbers Matter’s <i>Do Maps Lie?</i> . . . . .	156
8.11	Northern constellations and Hertzsprung-Russell diagram . . . . .	158
8.12	Cosmic history . . . . .	159
8.13	Data storytelling trope: trend lines . . . . .	162
8.14	Data storytelling trope: cluster bar chart . . . . .	163
8.15	Data storytelling trope: intersecting lines . . . . .	163
8.16	Data storytelling trope: circular charts . . . . .	164
8.17	Data storytelling trope: US colour coding . . . . .	164
8.18	Data storytelling trope: Canadian colour coding . . . . .	165
8.19	Data storytelling trope: broken axes . . . . .	165
8.20	Evolving a storytelling chart: example 1 . . . . .	166
8.21	Evolving a storytelling chart: example 2 . . . . .	166
8.22	Evolving a storytelling chart: example 3 . . . . .	166
8.23	Defining the visualization context . . . . .	167
8.24	Exploration dashboard . . . . .	168
8.25	Situational dashboard . . . . .	169
8.26	Storybook dashboard . . . . .	169
8.27	Storyboard narrative structure . . . . .	172
8.28	Simple storyboard . . . . .	172
8.29	Storytelling dashboard for sales numbers . . . . .	174
8.30	Storytelling dashboard for the <i>Health and Wealth of Nations</i> . . . . .	174
9.1	Chart types . . . . .	178
9.2	Chart types, an alternative. . . . .	178
9.3	Scatterplots and bubble charts . . . . .	181
9.4	Histograms . . . . .	182
9.5	Two-way table . . . . .	182
9.6	Text block . . . . .	183
9.7	A simple table . . . . .	184
9.8	Tables and heat map tables . . . . .	184
9.9	Line graphs and sparklines . . . . .	185
9.10	Funnel and gauge charts . . . . .	186
9.11	Bar charts and variants . . . . .	187
9.12	Geographical maps . . . . .	188
9.13	Heat maps and choropleths . . . . .	189
9.14	Small multiples . . . . .	190
9.15	Area charts . . . . .	191
9.16	Treemaps . . . . .	191
9.17	Text visualizations . . . . .	192
9.18	Radar and parallel coordinates charts . . . . .	193
9.19	Trees and network diagrams . . . . .	194
9.20	Animated and interactive charts I . . . . .	195

9.21	Animated and interactive charts II . . . . .	196
9.22	Chernoff faces . . . . .	197
10.1	A plot generated from a ChatGPT suggestion . . . . .	206
12.1	Schematics of the grammar of <code>ggplot2</code> graphics . . . . .	236
13.1	The components of the default Power BI view . . . . .	309
13.2	Conceptual interactions in the Borealis Terraformers LLC data . . . . .	311
13.3	Power BI's Power Query Editor . . . . .	313
13.4	Power Query saved steps . . . . .	314
13.5	Unlinking tables . . . . .	315
13.6	Linking tables . . . . .	315
13.7	Linking tables edit menu . . . . .	316
13.8	Calculation buttons in Modeling ribbon . . . . .	317
13.9	Accessing table tools . . . . .	319
13.10	Accessing column tools . . . . .	319
13.11	Creating a new column with DAX . . . . .	319
13.12	Cleaning up a new column . . . . .	320
13.13	Formatting decimal places . . . . .	320
13.14	Our first Power BI chart! . . . . .	320
13.15	The sum of a numerical value as the default statistic . . . . .	321
13.16	The other statistics for numerical variables . . . . .	321
13.17	A different Power BI chart . . . . .	322
13.18	Changing the aesthetics . . . . .	322
13.20	On the way to a calculated measure . . . . .	322
13.19	Another Power BI chart . . . . .	323
13.21	On the way to a calculated measure . . . . .	323
13.22	On the way to a calculated measure . . . . .	323
13.23	Creating a new measure . . . . .	324
13.24	Trouble in paradise? . . . . .	324
13.25	No trouble in paradise . . . . .	325
13.26	A chart created from a measure . . . . .	325
13.27	A chart using result of logical computation . . . . .	326
13.28	A bar chart built entirely out of measures . . . . .	327
13.29	Re-ordering a bar chart built entirely out of measures . . . . .	328
13.30	Using a slicer on a chart . . . . .	328
13.31	Chart for one specific project . . . . .	329
13.32	Hierarchical structure of dates . . . . .	329
13.33	Building a custom hierarchy . . . . .	330
13.34	Bar chart with custom hierarchical structure. . . . .	330
13.35	Matrix icon in the visualization panel . . . . .	331
13.36	Pivot tables in Power BI . . . . .	331
13.37	Groups dialog box (categorical) . . . . .	332
13.38	New groups: Group 1, Group 2, Group 3 . . . . .	332
13.39	Project Identifier grouped variable . . . . .	333
13.40	New variable in Accounting Transactions table . . . . .	333
13.41	Groups dialog box (numerical) . . . . .	334
13.42	Groups dialog box (numerical: list and bins) . . . . .	334
13.43	Histograms using grouped numerical variable . . . . .	335
13.44	Various ways of sorting a chart . . . . .	335

13.45	Custom sort order (I)	336
13.46	Custom sort order (II)	336
13.47	Removing rows using filters	338
13.48	Removing rows using the menu	339
13.49	Replacing values dialog box	339
13.50	Splitting a column	340
13.51	Trimming a text column	341
13.53	Location of the Append Queries button	343
13.54	The Append dialog box	344



# List of Tables

- 12.1 Common ggplot2 geometries and options . . . . . 238
- 12.2 Common ggplot2 geometries options . . . . . 239
  
- 13.1 Accounting transactions . . . . . 310
- 13.2 Accounting journal voucher codes . . . . . 310
- 13.3 Projects tombstone . . . . . 310
- 13.4 Projects FTE records . . . . . 311

