Numeracy
Advancing Education in Quantitative Literacy

Volume 12 | Issue 2 Article 18

2019

Graphicacy for Numeracy: Review of *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* by Claus O. Wilke (2019)

Christy M. Bebeau
University of South Florida, cmbebeau@mail.usf.edu

Graphicacy for Numeracy: Review of *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* by Claus O. Wilke (2019)

Abstract

Wilke, Claus O. 2019. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. (Sebastopol, CA: O'Reilly Media, Inc.). 390 pp. ISBN 978-1-492-03108-6. First edition. First release: 03-15-2019.

Claus O. Wilke has authored an excellent reference about producing and understanding static figures, figures used online, in print, and for presentations. His book is neither a statistics nor programming text, but familiarity with basic statistical concepts is helpful. Written in three parts, the book presents both the math and artistic design aspects of telling a story through figures. Wilke makes extensive use of examples, labels them good, bad, wrong or ugly, and explains why he deems them so. He includes chapters that serve as a directory of visualizations, and the reader will encounter familiar figure types as well as more esoteric ones. He discusses aesthetic considerations, including color, that make figures more compelling, and provides tips to connect with the audience to tell a story. Wilke intends the book for those producing figures, but consumers of graphical content will also find it helpful as his clear and concise style – and his effective use of examples – will help them evaluate the worth of figures they encounter.

Keywords

Data Analysis, Data Science, Graphicacy, Graphical comprehension, Numeracy, Data visualization, Map literacy

Cover Page Footnote

Christy Bebeau is a doctoral student in Geology at University of South Florida. She earned a BSE with a concentration in industrial engineering before working in the utility industry and owning her own information technology consulting firm. She became interested in education, changed careers, cofounded a not-for-profit school, and taught K-12 STEM courses for more than 15 years in the Tampa Bay area. Her interest in data visualization stems from her engineering and information technology backgrounds and was fostered through her studies in geoscience education.

Data visualization skills are important. Understanding graphs is central to the scientific process (Glazer 2011), and it was prominent in one of the founding documents of "numeracy" (Cockcroft 1982, paragraph 39; see also Karaali et al. 2016):

> We would wish the word 'numerate' to imply the possession of two attributes. The first of these is an 'at-homeness' with numbers and an ability to make use of mathematical skills which enables an individual to cope with the practical mathematical demands of … everyday life. The second is an ability to have some appreciation and understanding of information which is presented in mathematical terms, for instance in graphs, charts or tables …

Comprehending information on graphs and maps is being explicitly addressed in recent papers in *Numeracy* (e.g., Pfaff 2011; Juncgk 2012; Perez et al. 2015; Best 2016; Antonacci et al. 2017; Nuhfer et al. 2017; Xie et al. 2018; Bolch and Jacobbe 2019). Yet, there are few resources about designing compelling graphs that convey the story behind the data through clear, attractive, informative styles. Now *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* by Claus O. Wilke, a biology professor and researcher at the University of Texas at Austin, fills the gap. Published just a couple months ago, the book presents key principles for developing static figures for print, online, and presentations.

Before I review the book, I should convey what it is not. The book is not a statistics textbook. The only statistical equation included is the definition of the correlation coefficient in Chapter 12. While some familiarity with basic statistics is helpful, you do not need extensive statistics to benefit from Wilke's focus on figures, explained with concise language and appropriate examples. Furthermore, this is not a computer programming book. Wilke deliberately excludes any computer language recommendations and code from the text. He writes (p. xii):

> I want this book to be useful to you regardless of which software you use, and I want it to remain valuable even once everybody has moved on from ggplot2 and is using the next new thing.

The decision is a prescient one: the next new thing, Julia, is in version 1.1 but already has more than 4 million downloads.[1] Still, for readers who would like access to Wilke's code, he provides the R Markdown source at Github.[2] He also includes suggested references for statistics and programming in an annotated bibliography.

While it is neither a statistics nor a programming book, Wilke's text effectively lives up to its subtitle. It is an excellent primer about making, using, and understanding data through figures. Wilke shows the reader examples of well-

[1] https://julialang.org/. Bezanson, Jeff, Stefan Karpinski, Viral Shah, and Alan Edelman. "The Julia Programming Language." The Julia Language. (accessed June 8, 2019)
[2] https://github.com/clauswilke/dataviz (accessed June 8, 2019)

designed figures as well as those that are incorrect or aesthetically ugly, and he explains the reasons he judges them good, **bad**, **wrong**, or **ugly**. The reader learns what to do and what not to do. The book has three parts, and the reader might choose to jump between them or their chapters, but everyone should read the Preface and Introduction as they act as a key to the symbols and definitions Wilke uses throughout the text. I prefer printed books for my reference library, and this one is visually engaging. For those who like online resources, Wilke provides the complete author manuscript (before final copy-editing) online and open access at his serialmentor.com site.[3]

In Part 1, "From Data to Visualization," Wilke takes the reader on a tour of the key concepts behind accurately presenting data through figures. In Chapters 2 through 5, he discusses the aesthetics common to all figures: scale, axes, color, and types of data. He includes information likely familiar to most readers (for example, Cartesian coordinates) and builds on that foundation to explore the less familiar (such as non-linear scales and curved axes). He uses easy-to-understand data to demonstrate comparative strengths and weaknesses in how each figure tells the data's story. The chapter about the use of color scales was relevant to me as a geoscientist because Wilke adeptly demonstrates the usefulness of color to represent data values across geographic regions.

Beginning in Chapter 5, Wilke provides samples of visualizations used for different kinds of data. Chapter 5 is an introduction to different types of visualizations and serves as an index to other chapters in which he provides details behind their productions. For example, in Chapter 5 he introduces geospatial data in a half page, including (p. 42),

> The primary mode of showing geospatial data is in the form of a map

and four thumbnail images of different types of maps. He then directs the reader to Chapter 15 where he covers map projections and distortions, map layers, choropleth mapping, and cartograms.

Part 2, "Principles of Figure Design," discusses nuances about the design and use of the figures Wilke introduces in Part 1. For example, in his explanation of visualizations along linear axes (pp. 208–212), Wilke highlights a tip (p. 209), which I am told is a favorite of the QL community:

> Bars on a linear scale should always start at zero.

Then (p. 210), he shows two different line graphs of Facebook stock prices – one starting at $0 and the other at $110 – to demonstrate how the variability of the stock price appears greater on the graph starting at $110. To drive the point home, he labels the errant graph **bad**.

[3] https://serialmentor.com/dataviz/ (accessed June 8, 2019)

https://scholarcommons.usf.edu/numeracy/vol12/iss2/art18
DOI: https://doi.org/10.5038/1936-4660.12.2.18

The chapters in Part 2 cover topics not generally presented when we first learn how to make graphs. They go beyond how to title your figure and label the axes. Wilke discusses how to handle overlapping points with partial transparency, jittering, 2-dimensional histograms, and contour plots, and he provides the pros and cons of each choice. He also describes the misuses of color as well as considerations for color-vision deficiency. I appreciated the chapter on multi-panel figures as I enjoyed exploring how I might combine histograms, scatter plots, and box-and-whisker plots into a mosaic that tells a complete story. The reader can consider Part 2 a checklist of what to do (and not do) when creating figures. Wise data scientists should compare their figures against the topics and tips in Part 2 before delivering final versions.

Part 3, "Miscellaneous Topics," contains three chapters. In Chapter 27, Wilke discusses the common image file formats: bitmap and vector graphics, lossless and lossy comprehension, and converting between image formats. I highlighted his rule of thumb (p. 324),

> always store the original image in the format that maintains maximum resolution, accuracy, and flexibility,

and now I better understand the implications of not doing so. In Chapter 28, Wilke discusses how to choose visualization software, and his consideration about reproducibility and repeatability is fundamental to more than the software product. He takes the concepts, slightly modifies them, adds considerations for data exploration versus data presentation, and leads the reader to contemplate how the types of figures needed drive the selection of software. After he walks the reader though decisions that drive the software selection, he offers them the option of using hand-drawn figures, writing (p. 329),

> such figures represent a unique and personalized take on what might otherwise be a somewhat sterile and routine presentation of data,

and reminding the reader of the importance of connecting to their audiences.

Chapter 29 is perhaps the most salient of all the chapters. Here, Wilke argues the importance of storytelling for communication and describes how figures can complete the the story. He begins by telling a story through two figures about the growth of preprints in the biological sciences on the arXiv repository of electronic preprints. In the first figure, he sets up the story, showing how the trend grew linearly from 50 preprints per month in 2007 to 200 preprints per month in 2014 and then flattened out. Through the figure, he creates the conflict: what happened in 2014 to curtail the growth? Then Wilke shows the second figure as a response. This figure shows a steep linear growth of preprints per month beginning in 2014 and climbing through 2018 but on a new server, bioRxiv, that went live at the end of 2013. In the rest of this closing chapter, Wilke provides tips about communicating with figures. He writes (p. 338):

Never assume your audience can rapidly process complex visual displays.

The headings of the remaining sections, each containing examples, complete his instructions: 1) "Build up toward complex figures," 2) "Make your figures memorable," and 3) "Be consistent but don't be repetitive."

Wilke intends the book for scientists, analysts, consultants and anyone who prepares figures for publication (p. xi). Certainly, these people will benefit from his clearly written explanations of appropriate examples. I would urge them to read the Preface, Introduction, Chapter 5, and Chapter 29 before they delve into the other chapters. Another audience who might benefit from the book are consumers of prepared graphs. They can improve their graphicacy skills by becoming familiar with the types of graphs Wilke presents in the book. I'd suggest they pay attention to the examples Wilke labels **bad** or **wrong**. Through these, they can learn how poorly designed figures distort or misrepresent the data and how to read the different types of figures with a discerning eye.

As an educator, I tend to judge whether a book might be of use in one of my classes. Wilke's style and clarity make this book an excellent candidate. As I read, I found myself wishing for end-of-chapter questions or exercises to check student comprehension or a companion website where students would have opportunities to practice both reading and producing figures. The book could augment a statistics class to improve the figures the students make. It might also be of use in a programming class where students would create similar figures. It could be a useful adjunct to quantitative literacy and quantitative reasoning courses, especially given its online, open-access availability, and for lab sections and off-campus assignments where reading and interpreting graphs are relevant.

In *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*, Professor Wilke has produced a well-written, clearly organized reference to effective visualization. His extensive use of examples and his concise explanations of what makes figures good, **bad**, **wrong** and **ugly** guides the reader to improved figures. The directory of visualizations in Part 2 includes not only familiar figures but also more esoteric styles the reader might consider. Overall, Wilke reminds the reader of the importance of connecting to the audience and using the figures to tell a compelling story.

References

Antonacci, Nathan J., Michael Rogers, Thomas J. Pfaff, and Jason G. Hamilton. 2017. "Figures and First Years: An Analysis of Calculus Students' Use of Figures in Technical Reports." *Numeracy* 10(2): Article 10. https://doi.org/10.5038/1936-4660.10.2.10

Best, Joel. 2016. "Infographics as Eye Candy: Review of *World War II in Numbers: An Infographic Guide to the Conflict, Its Conduct, and Its Casualties* by Peter Doyle (2013)." *Numeracy* 9(1): Article 8. https://doi.org/10.5038/1936-4660.9.1.8

Bolch, Charlotte A., and Tim Jacobbe. 2019. "Investigating Levels of Graphical Comprehension Using the LOCUS Assessments." *Numeracy* 12(1): Article 8. https://doi.org/10.5038/1936-4660.12.1.8

Cockcroft, Sir Wilfred H. 1982. *Mathematics Counts*. Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the Chairmanship of Dr. WH Cockcroft. London: Her Majesty's Stationery Office. http://www.educationengland.org.uk/documents/cockcroft/cockcroft1982.html

Glazer, Nirit. 2011. "Challenges with graph interpretation: a review of the literature." *Studies in Science Education* 47(2):183–210. https://doi.org/10.1080/03057267.2011.605307

Jungck, John R. 2012. "Incorporating Quantitative Reasoning in Common Core Courses: Mathematics for *The Ghost Map*." *Numeracy* 5(1): Article 7. https://doi.org/10.5038/1936-4660.5.1.7

Karaali, Gizem, Edwin H. Villafane Hernandez, Jeremy A. Taylor. 2016. "What's in a Name: A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9(1): Article 2. https://dx.doi.org/10.5038/1936-4660.9.1.2

Nuhfer, Edward, Steven Fleisher, Christopher Cogan, Karl Wirth, and Eric Gaze. 2017. "How Random Noise and a Graphical Convention Subverted Behavioral Scientists' Explanations of Self-Assessment Data: Numeracy Underlies Better Alternatives." *Numeracy* 10(1): Article 4. https://doi.org/10.5038/1936-4660.10.1.4

Perez, Victor W., Joel Best, and Rachel J. Bacon. 2015. "Cancer Clusters in Delaware? How One Newspaper Turned Official Statistics into News." *Numeracy* 8(1): Article 7. https://doi.org/10.5038/1936-4660.8.1.7

Pfaff, Thomas J., Michael Rogers, Ali Erkan, and Jason G. Hamilton. 2011. "Go Figure: Calculus Students' Use of Figures and Graphs in Technical Report Writing." *Numeracy* 4(1): Article 6. https://doi.org/10.5038/1936-4660.4.1.6

Xie, Ming, H. L. Vacher, Steven Reader, and Elizabeth Walton. 2018. "Quantitative Map Literacy: A Cross between Map Literacy and Quantitative Literacy." *Numeracy* 11(1): Article 4. https://doi.org/10.5038/1936-4660.11.1.4