

Introduction to Data Visualization for Epidemiologists: Self-Study

Last updated: March 2023

BACKGROUND

This document provides a high-level introduction to data visualization and shares some of our favourite resources ahead of your upcoming “Data Visualization in R” course. The best way to get better at visualizing your data is practicing and getting feedback! For 2023, we are providing this self-study document to provide you with resources and information on data visualization in public health. If you’ve already taken the TDU’s Data Visualization module in 2022, the first two pages of this document provide the same information, but we have added an improved and expanded list of “Going Further” resources in this document that we hope you find useful. Happy visualizing!

INTRODUCTION TO DATA VISUALIZATION

WHAT IS DATA VISUALIZATION?

Data visualization is a way of sharing the story of your data with an audience. Transforming abstract data into a graphical or visual representation facilitates communication, amplifies cognition, and reveals patterns in complex information.

Effective data visualizations can help us reveal patterns and relationships in data – along with statistics! Data is like the box of Legos dumped on the floor. It is:

- Not structured
- Without inherent communicative purpose
- Too granular to make sense of (*i.e.*, many tiny, individual pieces)
- Impossible to interpret without the necessary context, and
- Painful when accidentally stumbled upon 🤯

Visualizations are like the [Lego blocks stacked](#). Visualizations have:

- Structure
- A communicative purpose
- A message that accounts for the whole picture (the individual pieces do not matter as much as the pattern)
- A clear, easily interpretable meaning

Swedish physician and statistician, Hans Rosling, efficiently summarized the importance of data visualization when he said, “the purpose of a visualization is to go from numbers, to information, to understanding” and ultimately wisdom. Moreover,

Data visualization is the graphical display of abstract information for two purposes: *sense-making and communication*”

- Stephen Few, Author of “Show Me the Numbers: Designing Tables and Graphs to Enlighten”

Data visualization has been an effective tool for public health communication for centuries - starting with John Snow’s Cholera map in 1854. Today, data visualization is more important than ever, as we live in a world that is abundant with data. For example, 90% of the world’s data have been collected in the last two years alone, and with increasing data literacy within the population and increased access to data overall, it is crucial to make sense of this abundant information. Without effective methods to interpret and communicate the story that underlies the data, the meaning of the data cannot be fully realized. Data visualization thus allows public health practitioners to more effectively communicate important information to their target audience in order to educate and inspire change within the population.

When you are trying to deliver specific information or get a certain message across, wrapping it in story makes it more accessible to the people you are trying to reach. Data visualization, and any other form of storytelling with data, relies on picking the appropriate:

- story/message
- numbers/indicators
- medium for your audience.

Effective data visualizations cannot and should not negate the need for thorough methods and sound data analysis.

One of the most important and useful tools for visualization is the infographic, which combines data and visuals to allow users to comprehend and retain information more effectively. While data visualizations help to generate hypotheses, infographics tell a premeditated story to point out relationships and guide the audience to conclusions. They are often used to summarize a longer report or article, display key research findings, help raise awareness about an issue, cause or event, support communications of a strategy, or to share information with communities.

WHY DO WE NEED DATA VISUALIZATION?

The ability to showcase data and epidemiological findings in visual formats, such as graphs, infographics, and dashboards, has increasingly become an important skill for epidemiologists to have. Data visualization allows for statistics or scientific findings to be better understood by both decision makers and members of the public as they work towards public health action. Visual outputs help individuals, especially those who are visual learners, conceptualize complex

information quickly and help epidemiologists effectively communicate findings. Ensuring that visual outputs are accessible to a wide variety of people is one of the cornerstones of effective communications and visual product development, as a data visualization is only useful if it provides a way for its audience to understand the message as intended.

GOING FURTHER

Are you interested in learning more about data visualization? Check out some of the resources below. Have you discovered resources that you think your colleagues would also enjoy? Please let us know why you love it, and we'll add it to the list for future learners!

Data visualization checklist

Check out this Checklist! The [Data Visualization Checklist](#) is meant to be used as a guide for the development of high impact data visualizations. Comprised of 24 guidelines, it shows how graphs should be formatted to best show the story in your data. The 24 guidelines are broken down in to 5 sections: Text, Arrangement, Color, Lines and Overall. What's handy about this checklist is that you can rate each aspect of your data viz to see how it scores. Well-formatted data visualizations that score between 90-100% of available points indicate a product where viewers are better able to read, interpret, and retain content.

Want to upload your data viz directly online? The Evergreen Data website allows you to do just that; as well as offering an example of each of the 24 guidelines/checkpoints.

Accessibility considerations in data visualization design

Are you “Keen” to learn more about **accessibility considerations in data visualization**? This [blog](#) by Keen software (builder of tools and infrastructure to help developers deliver analytics) covers some key concepts to think about when reporting data to support the diverse capabilities of intended users in order to offer the richest experience possible. Topics covered include:

- Recognizing that all disabilities are not one and the same
- Tips on how to keep visuals clear and contrasted
- Leveraging white space
- Not relying on colour to be the only visual cue
- Using technical markers to convey metrics
- Simplicity is key

Accessibility matters – accessible data viz is better data viz

Inclusive design principles and accessibility are important to take into consideration when designing data visualization; they help a broader audience understand your graphic. Designing with accessibility in mind can even help make your visualizations easier to understand for people without disabilities.

This Storytelling With Data resource offers 5 easy ways to make your data visualization more accessible and includes free tools to check colour contrasts, as well as examples and links to other resources if you want to take a deeper dive.

Accessible data visualizations

Want to make complex data visualizations accessible? Making data viz accessible means that users will not miss out on important information and that your viz will be understandable for everyone. This guide by the University of Wisconsin outlines how to communicate data, how to use color and pattern, as well as how to work with interaction and animation. It also includes numerous resources to support further learning, while not being too overwhelming.

Ten simple rules for better figures

FREE! Small list of open-source tools! Now that we have your attention... this short article is a great resource in providing a basic set of rules to improve figure design and to explain some of the common pitfalls. Scientific visualization is that graphical interface between people and data; following some of the simple rules outlined in this article will help you to bridge that connection.

- Rule 1: Know Your Audience
- Rule 2: Identify Your Message
- Rule 3: Adapt the Figure to the Support Medium
- Rule 4: Captions Are Not Optional
- Rule 5: Do Not Trust the Defaults
- Rule 6: Use Color Effectively
- Rule 7: Do Not Mislead the Reader
- Rule 8: Avoid “Chart junk”
- Rule 9: Message Trumps Beauty
- Rule 10: Get the Right Tool

And what’s even better, is that it lists FREE open-source tools to use.

Choose an effective visual with the SWD Chart Guide

Tell a story with your data! This interactive and engaging Storytelling With Data Chart Guide helps you to choose just that right effective visual to communicate your data. This guide shares the ‘good and the bad’ of commonly used charts and graphs to know when and how to effectively apply them. It also identifies common pitfalls, including some things to avoid. This resource is based on a “back-to-basics” blog series offering an interactive way to select a visual that then sends you to a succinct article outlining the basic concepts, common questions, and ways to use that chart or graph.

Data viz cheatsheet

This PolicyViz cheatsheet is a high-level summary of some of the core lessons taught in the author's classes and workshops. It is not meant to be comprehensive, but rather serve as a good reminder of best practices and core things to keep in mind when visually presenting data.

Topics to remember include:

- Showing the data
 - Reducing the clutter
 - Integrating the text and the graph
 - Pre-attentive processing
 - Remembering your audience
 - Including annotation
 - Using pie charts with care
 - Avoiding 3D
 - Starting bar and column charts at zero
 - Making labels easy to read
 - Trying small multiples
 - Using maps carefully
 - Colour and font considerations
- Visualization mapping: Form and Function