

What is Data Visualization?

And how do I avoid being *bad* at it?

Sarah Moore

Week 2, Day 2

What kinds of data have you worked with thus far?

Would you consider it “political”?

Which of these sounds like a problematic source of data?

- 1) Protest data with phone tower geo-location data in a repressive country
- 2) De-identified tweets associated with a BLM hashtag
- 3) Survey data obtained without informed consent
- 4) Data that has been obtained by a pro-gun lobby group
- 5) Interview data that redacts names, locations, and other identifying characteristics

A Rundown of This Week's Readings

- ▶ Cairo (2014) "Ethical Infographics" Meta

A Rundown of This Week's Readings

- ▶ Cairo (2014) “Ethical Infographics” Meta
- ▶ Beers (2022) “Trust In The Digital Age Requires Data Ethics And A Data Literate Workforce” Applied and Job Readiness

A Rundown of This Week's Readings

- ▶ Cairo (2014) “Ethical Infographics” Meta
- ▶ Beers (2022) “Trust In The Digital Age Requires Data Ethics And A Data Literate Workforce” Applied and Job Readiness
- ▶ Correll (2018) “Ethical Dimensions of Visualization Research” Conceptual, Theoretical, Agenda-Setting

Keywords and key takeaways from the readings

Motivating Questions:

- ▶ What types of ethics exist around data and data visualization?
- ▶ What does good data literacy provide society? What are the consequences of poor data literacy?
- ▶ What are our responsibilities around data management and data diffusion?

Data visualization:

Exploration OR Dissemination

What is data visualization?

- ▶ Creation of some graphical shape or form to allow other people to comprehend data and ideas.
- ▶ Visualization as a *task* to be performed by the viewer. Therefore, we have to ask not only what we want to convey, but what *task* are we expecting the reader to perform.
- ▶ According to Cairo (2014), then we should contemplate the way that our evidence confers with a moral choice. **What does this mean?**

“Increase *understanding* while *minimizing* harm.”

Does the purpose of data visualization change across contexts?

- ▶ This is to say, do we have different aims of data visualization given what our task at hand is?
- ▶ Is data viz for the purpose of dissemination always a *rhetorical* device?
- ▶ Does this mean our *ethical obligation* is uniform across these activities?

What is politics?

What does it have to do with power? With knowledge?

What does it have to do with data visualization?

How is data political?

- ▶ Multiple, potentially overlapping layers to this:
 - 1) Data collected *about* politics
 - 2) Data collected for a *political purpose*.
 - 3) Data portrayed for *political gain*.
 - 4) Data whose existence, analysis, or portrayal has *political consequences*.
- ▶ There are probably a lot of data out there that *are* mundane or trivial. But there are also a lot of data out there wherein the misrepresentation or even publication has potential effects on politics.

Here are a few examples:

- ▶ Cairo's discussion of this fivethirtyeight piece.
- ▶ This clarification by Mother Jones on anti-abortion propaganda.
- ▶ This interesting Tweet by former presidential hopeful, Andrew Yang.
- ▶ Follow @GraphCrimes on Twitter for more.
- ▶ What are some consequences if these are taken seriously?

Examples of social science data?

What are sources for data in politics?

Quantitative :

- ▶ Event Data (e.g. ACLED)
- ▶ Surveys and Experiments (e.g. General Social Survey or LAPOP)
- ▶ Administrative Data (e.g. Tax Data)
- ▶ Expert-Coded Information (e.g. V-Dem or Quality of Governance)

What are sources for data in politics?

Qualitative :

- ▶ Interviews and Focus Groups
- ▶ Participant Observation

Mixed or Multi-Method:

- ▶ Text Data (e.g. tweets with a specific hashtag, constitutions)

But, what should we be aware of? When is data “bad”?

- ▶ Data sources! Data can be its own source of bias. Need to assess the quality, scope, and reliability of our data.
- ▶ Who funded it? Who collected it? What was the purpose of the data's collection?
- ▶ When people say they have measured difficult concepts in their data, like *political stress*, or even somewhat more straightforward things, like *violence levels*, what are they really measuring?

What are some concerning trends you all have encountered with regard to data viz or data diffusion?

So, what can we do?

Correll talks about some main points to facilitate what *our* obligations are as researchers.

- 1) Make invisible *visible*
- 2) Collect data with *empathy*

▶ What's this concept of *anthropomorphizing* data?

- 3) Challenge structures of power

▶ With caveat that sometimes exposing and promoting truth requires acting with existing vessels of power.

Data Literacy is Reciprocal

- ▶ Audiences can only be so data literate insofar as the person performing the analysis is literate in what they are portraying.
- ▶ Different types of data require that we are attuned to what is good/bad practice both ethically and procedurally in displaying it.
- ▶ As I mentioned before, visualization and analysis of data does not exist in some vacuum. What is the minimum context necessary to appropriately inform an audience?

Let's wrap this up...

All quarter we will have plenty of time to discover the dos and don'ts of actually *making* visualizations.

- ▶ Let's think through what the dos and don'ts are of dealing with data before you even get through specifying a visualization.

Some Main Points

- ▶ We will be pretty minimal with exploratory data visualization in and of itself. Once we want to use data viz as a communication device need to be more critical.
- ▶ Lots of choices in data viz, but also a context (and a subtext).
- ▶ Be aware of what you are working with and to whom you want to speak.
- ▶ Be able to justify your decisions about measurement, data inclusion or exclusion, and how you got your data.

Data Sources

- ▶ Harvard Dataverse
- ▶ Project and institute sites
- ▶ NU Libraries Social Science Data Guide
- ▶ Other assorted places along the way
- ▶ Make your own dataset!