# DATA 22700

### Homework Assignment 4: Persuasive or deceptive visualization?

*Friday, Week 8*

It is tempting to think of data and visualization as neutral actors, with a single "correct" set of design choices that truthfully report the data. However, outside of egregious errors (e.g., when dates are sorted incorrectly or the y-axis is not scaled uniformly), we see that "ground truth" in data is much more contextual and situated.

Design choices we make give visualization a rhetorical power that influences what a reader concludes and remembers about the data, and blurs the line between persuasion and deception. For instance, contrast Simon Scarrs Iraqs Bloody Toll with a more conventional representation of the same data, and consider why Scarrs visualization won an award while another visualization that made similar design choices– Gun deaths in Florida by Christine Chan–was widely considered to be misleading.

In this assignment, we will grapple with these ethical concerns by using visualization to offer two different, opposing perspectives to the same question posed about a single dataset.

Students will *work alone*.

**Students should submit their assignment as a PDF document.**

---

### Technical specification

**Select one dataset from the list provided below.** These datasets are intentionally chosen to cover politically charged topics, as ethical analysis and visualization is especially critical is such situations.

Spend time exploring and becoming familiarizing with the dataset through exploratory visual analysis. Once students feel like they have developed an understanding of the data, **students must devise a proposition** about it: a statement that asserts a judgement or opinion about the trends the student might have uncovered. An example of a proposition might be, "Gun deaths spiked after Florida enacted its 'Stand Your Ground' law."

When devising propositions, please be respectful and inclusive towards other people. In particular, while students are welcome to formulate controversial propositions, **propositions that demean or dehumanize people** (e.g., based on their race, gender expression, sexual orientation, disability, or other aspects of their identity) **are unacceptable**. If students are unsure whether or not their proposition is suitable, please contact the course staff via email.

**Students must design two visualizations** to persuade a reader about each side of the proposition (i.e., one visualization should persuade the reader that the proposition is true while the other should persuade the reader that the proposition is not true).

We encourage students to use whatever design choices produce the most persuasive visualization for each side of their proposition. Students are welcome to not only use techniques we would usually consider to be earnest (e.g., effective and expressive encodings, transparently communicating data transformations, citing sources, etc.) but also those we might some-

times consider to be deceptive (e.g., violating conventions, skewed or slanted titles and labels, truncated scales and axes, filtering outliers, etc.). In doing so, students are likely to discover that there is not always a clear distinction between persuasive rhetoric and deception.

Note, however, that this goal of persuasion also means that, **if students use any deception techniques, they should not be immediately obvious** (including to the course staff!) as they might otherwise backfire and dissuade your reader.

Students are free to use any visualization tools they like and should carefully consider data transformation, visual encoding, textual content (i.e., titles, axes, labels), and annotations. In this assignment, we construe "visualization" broadly (i.e., a single visualization can include several concatenated or inset charts).

**Students must document they design decisions and rationale**. For each visualization, *enumerate 3-5 design decisions* which are central to making the visualization persuasive (note, a "design" decision can also refer to decisions you made about data transformation or textual content). Then, for each decision:

1. Score, on a diverging scale from -2 to 2, how deceptive or earnest the decision is (where -2 is fully deceptive, 0 is neutral, and 2 is fully earnest). Students should only use a score of 0 if they are really unable to decide one way or another. Scores should be in increments of 1 unit (i.e., a 5-point Likert-style scale).

2. Write a short paragraph documenting the 3-5 decisions, with corresponding rationale/justifications. For each decision, briefly explain how it helps make the visualization persuasive? What other alternatives did the student consider, and why did they select this one?

After documenting these decisions for both visualizations, **students must write a final paragraph reflecting on their overall design process**. What was straightforward or difficult? What (if anything) was surprising? How (if at all) did this assignment change the student's views on ethical analysis and visualization? What bounds (if any) can students draw to distinguish "acceptable" persuasive choices vs. "misleading" ones (and if none, why not)?

**Technical Note:** although you are encouraged to create a visualization as close to your final visualization using Altair or matplotlib, it is allowed to post process your chart manually, e.g., using a vector graphics editor or some other graphical software.

---

### Datasets

Please select one dataset from the options below. **Students must use the same dataset for both visualizations**, but they are free to transform the data differently for each design.

- Illinois COVID-19 Data. Illinois has collected detailed data about the spread of COVID-19 in the state including cases, testing, hospitalizations, and vaccination broken out across different demographic characteristics. Multiple datasets are linked from the site, and some of them are broken down by different grouping factors such as spans of time

(e.g., day, month) or geography (e.g., municipality, county). Students may need to join multiple data sources from the site in order to make a compelling submission.

- Federal Campaign Finance Data. All candidates and political action committees that participate in federal elections (i.e., for the Presidency, Senate, or House of Representatives) are required to disclose how much money they raised and spent on a quarterly basis to the Federal Election Commission (FEC). As this data is split across many different datasets (including money raised, money spent as well as candidates and committees), we recommend downloading the data in bulk and joining it together as needed.

- Civilian Complaints Against New York City Police Officers. This is a dataset compiled by ProPublica, an independent, nonprofit investigative journalism newsroom. It contains more than 12,000 civilian complaints filed against the NYPD, with demographic information about the complainant and officer, the category of the alleged misconduct, and the result of the complaint.

- Gentrification and Demographic Analysis. This is a dataset compiled by BuzzFeed News to understand gentrification, or how the character and demographics of neighborhoods change as more affluent people and business move in and potentially displace existing residents. The process of data collection, cleaning, and analysis is well-documented by the BuzzFeed News team, so be sure to read the accompanying article which contains important context and details.

- Social mobility in the U.S. Raj Chettys group at Harvard studies the factors that contribute to (or hinder) upward mobility in the United States (i.e., will our children earn more than we will). Their work has been extensively featured in The New York Times. This page lists data from all of their papers, broken down by geographic level or by topic. We recommend downloading data in the CSV/Excel format, and encourage you to consider joining multiple datasets from the same paper (under the same heading on the page).

- Human Development Indicators, 19602020. The World Bank has tracked global human development by indicators such as climate change, economy, education, environment, gender equality, health, and science and technology since 1960. The linked repository contains indicators that have been cleaned and formatted to simplify visual analysis and visualization design. However, youre also welcome to browse and use the original data by indicator or by country. Click on a category to download the CSV file on the right-hand sidebar.