DATA 23700 Autumn 2025

Assignment 5: Data Storytelling

Due November 17, 2025

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. Instead we should think of *visualizations as rhetorical devices* that can provide support for arguements and storytelling.

In this assigment, students will practice using visualization for *storytelling*. In doing so, they will demonstrate the ability to (1) structure an argument around data, (2) use text and visualization together to situate data in context, and (3) navigate design choices that give visualization a rhetorical power and influence what a reader concludes and remembers about the data.

Students will *work alone*.

Students should submit their assignment as a PDF document on Gradescope. Depending on the chosen format/genre of the student's data story (see below), the PDF should either contain the submission itself (e.g., comic) or a link to a webpage hosting the submission (e.g., slide show, video). Regardless, of the submission's genre, the PDF should include the write-up.

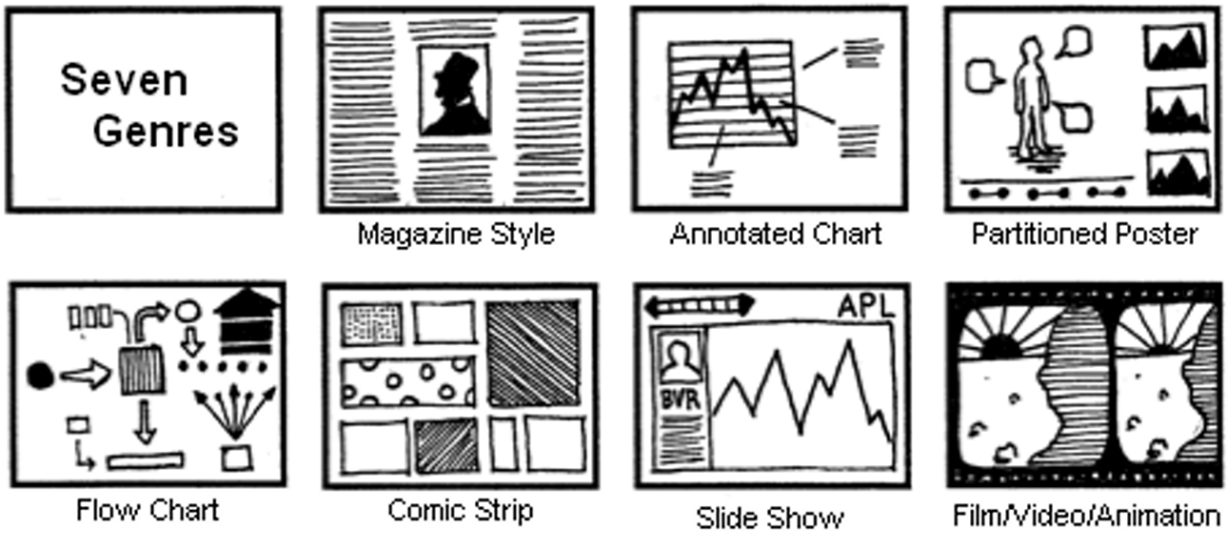
Technical specification

Select one dataset from the list provided below. These datasets are intentionally chosen to cover politically charged topics, such that students will be able to stake out a data-driven argument.

Spend time exploring and becoming familiar 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 a short data story to persuade the reader of their proposition. This story should be in the format of one of the *seven geners of narrative visualization* identified by Segel and Heer (see Figure below).



For a static submission, we suggest an annotated chart, a partitioned poster, or a comic strip. These should fit on one page. For a dynamic submission, we suggest either a slide show or a video.

Your data story should consist of:

- A sequence of at least four visualizations. They can build on each other (e.g., adding encodings to give a more nuanced view of the same pattern, changing annotations to highlight different details).
- Some sort of text accompanying each visualization. These can be captions or annotations as appropriate for the student's chosen narrative genre. For videos, some spoken narrative exposition would be good.
- The proposition the student is arguing for should be explicit upfront (e.g., in a title).
- The layout or flow of information should facilitate a particular reading order.
- The views composed together should culminate into a whole story, narrative, or argument based on the chosen data. This story should be self-contained and require no additional explanation to interpret.

Students are free to use any visualization and presentation tools they like. They should carefully consider data transformation, visual encoding, and textual content (i.e., titles, axes, labels, annotations, captions). The resulting submission should be polished and professional enough to hypothetically share publicly (e.g, in a publication). Students will benefit from taking this opportunity to practice using Figma, Powerpoint, video editing software, or similar.

For the write-up portion of the assignment, **students must write two paragraphs 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 storytelling with visualization? Did you make any choices to persuade your reader that

felt like they could be considered misleading? Whether or not you did, how did you think about what distinguishes acceptable persuasive choices from misleading ones?

Datasets

Please select one dataset from the options below. If you want to use a different dataset, please check with your instructor or TAs.

- 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 Chetty's 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, 1960–2020. 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, you're 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.