

Data Science 1 - HW 3

CSYS 5870, Fall 2025

September 26, 2025

Question 1: Revisiting a Previous Visualization

Part 1: Constructing a hand drawn image

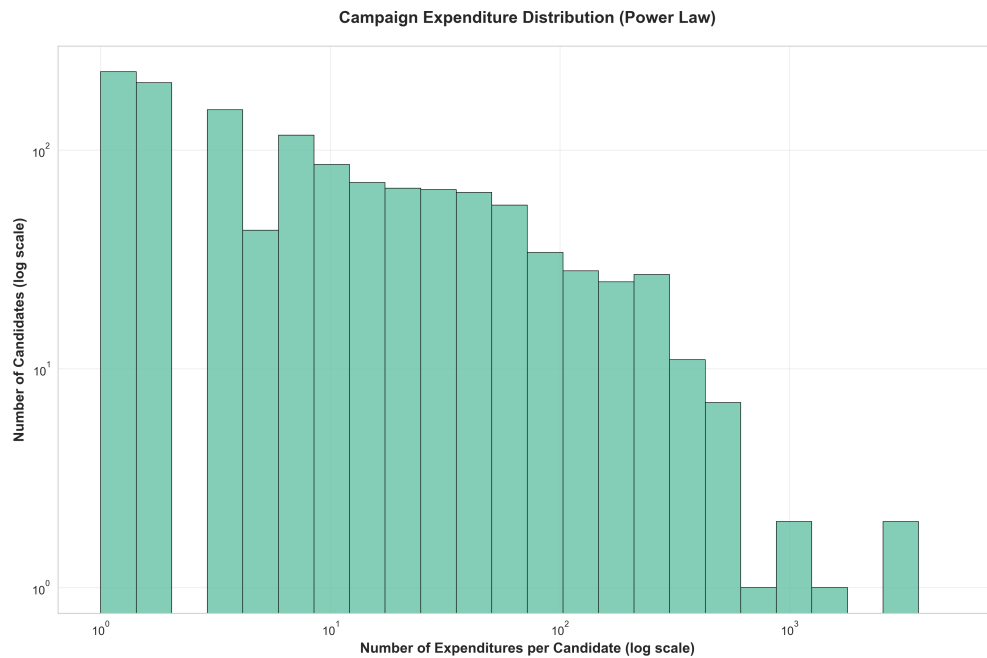


Figure 1: Image selected to revisit

This plot was an attempt at conveying the idea that campaign expenditures are dominated by 2 individuals. The feedback I received is that the image is confusing. "Perhaps you should flip the axis." The idea was that the tail would be more defined. The bigger question was around the "so what?".

- Why is this distribution important?
- What is an expenditure?
- How do these expenditures breakdown across candidate/election type?

- How do expenditures translate to dollars spent?
- How can we visualize how dollars are spent on different types of expenditures?

In trying to answer these questions with a hand drawn picture, I tried to focus on how a visualization might flow from panel to panel. My inspiration came from Claus Wilke's "compound figures" - specifically I tried to "employ a consistent visual language" in telling my story.

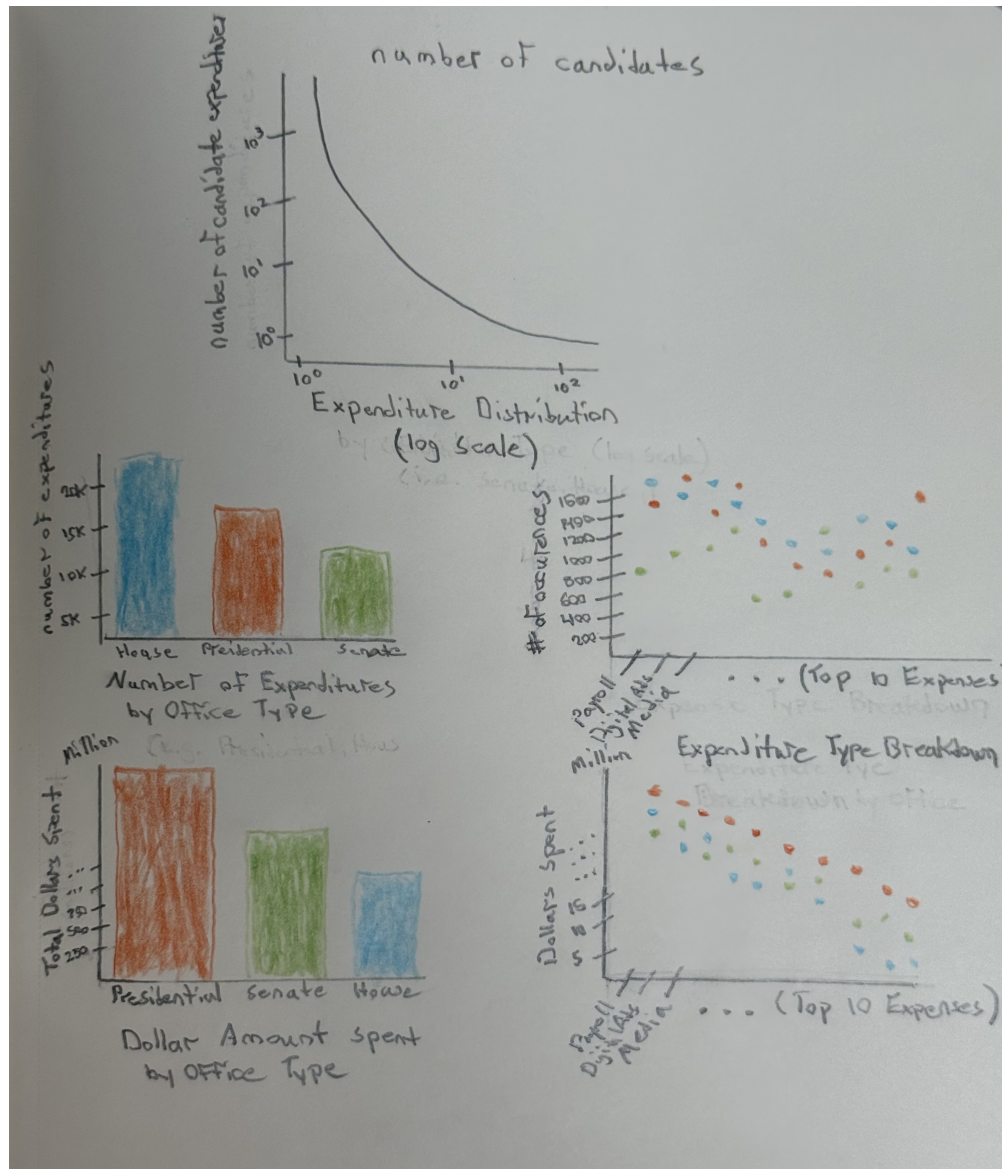


Figure 2: A Visually Consistent Plot

In the image above, I start with a cleaner-looking power law distribution by flipping the axis and smoothing out the tail. In the plots below, presented in a top-to-bottom, left-to-right fashion, I begin to unpack and answer the questions that emerged from the provided feedback.

First, more context is provided on the expenditure distribution by breaking it down by House, Senate, and Presidential race types. Then the viewer can better understand the actual nature of expenditures through frequency analysis. Finally, we translate expenditures into dollars, following the same narrative structure. Each plot provides a key piece of information that helps tell a bigger story.

What the viewer must keep in mind while interpreting this data is that there were primarily 2 presidential candidates in this dataset, compared to hundreds of House and Senate candidates. As one navigates these visualizations with this context in mind, the true magnitude of presidential elections in this country becomes apparent—not only in terms of money, but in human effort and organizational complexity as well.

Part 2: Translate hand drawn image to plot

Below there are 3 iterations. The first plot shows a lack of success in the scatter plots due to a lack of consistency across expenditure types.

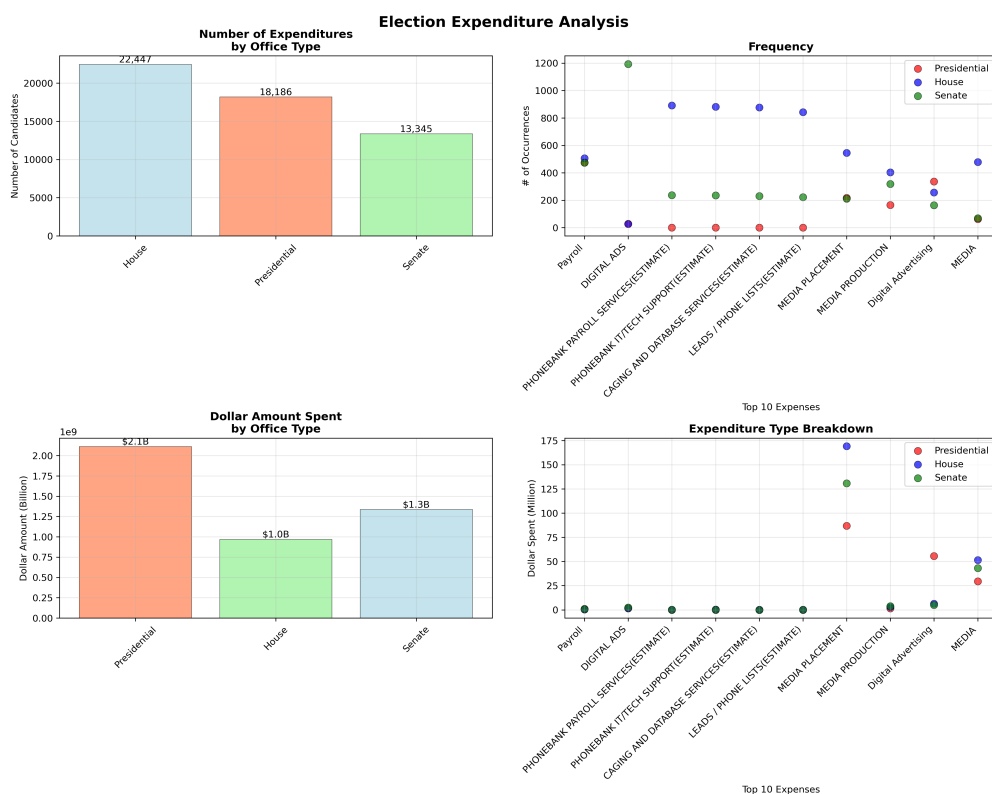


Figure 3: Iteration 1

To address the lack of consistency I decided to create horizontal bar plots that displayed spending for presidential and non-presidential (house and senate) office types. The idea being that the flow would go as follows:

expenditure breakdown count by office type -> translate expenditure counts to money spent -> breakdown the way the money is spent by office types

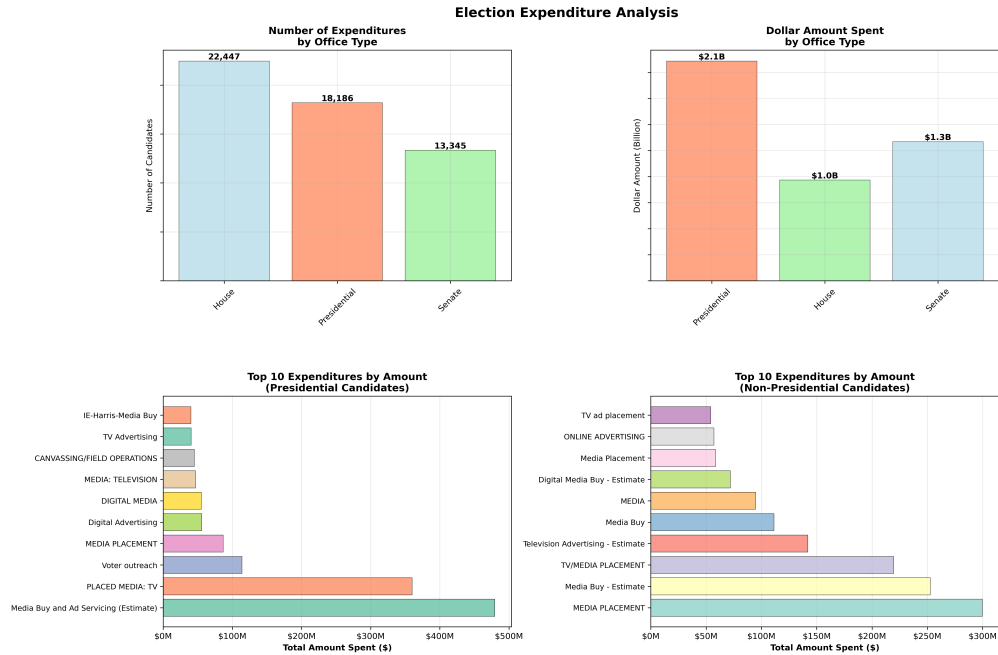


Figure 4: Iteration 2

However, I still found this approach a bit confusing, and I felt it didn't address the questions I set out to answer with these plots. Also, I felt that I totally failed to achieve a visually consistent language. What follows is something much nicer!

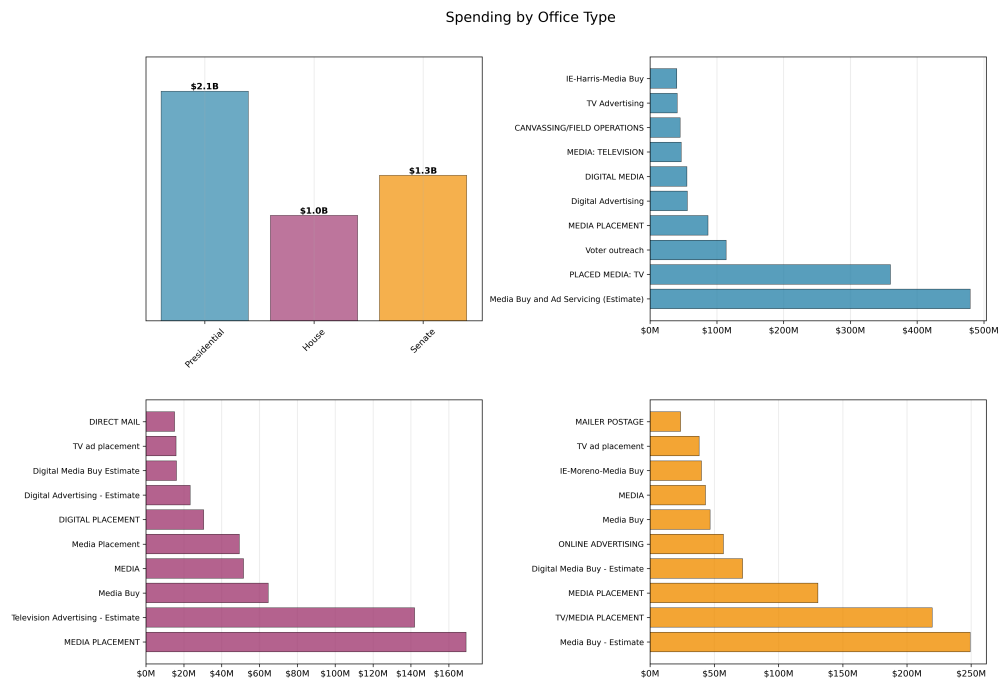


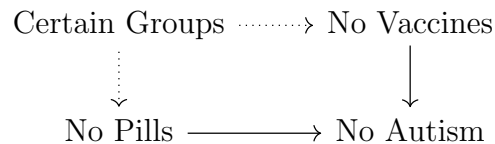
Figure 5: Iteration 3

Above we see much cleaner panel! It focuses entirely on spending and doesn't confuse the viewer by making the transition from count expenditures to how those translate into dollars. Also I removed titles as I felt the plots were self-explanatory. One criticism I have of this panel is that I still haven't gotten a sense of how to manipulate fonts. I would love to be able to select a font that better matches the overall aesthetic.

Question 2: Causal Statements

Statement 1

Causal Statement: "There are certain groups of people that don't take vaccines and don't take any pills and have no autism." - Donald Trump



Analysis Questions:

Counterfactual:

The implicit counterfactual here is that if certain groups (e.g. "The Amish") vaccinated and took pills, they would have autism in their population.

Ideal Experiment:

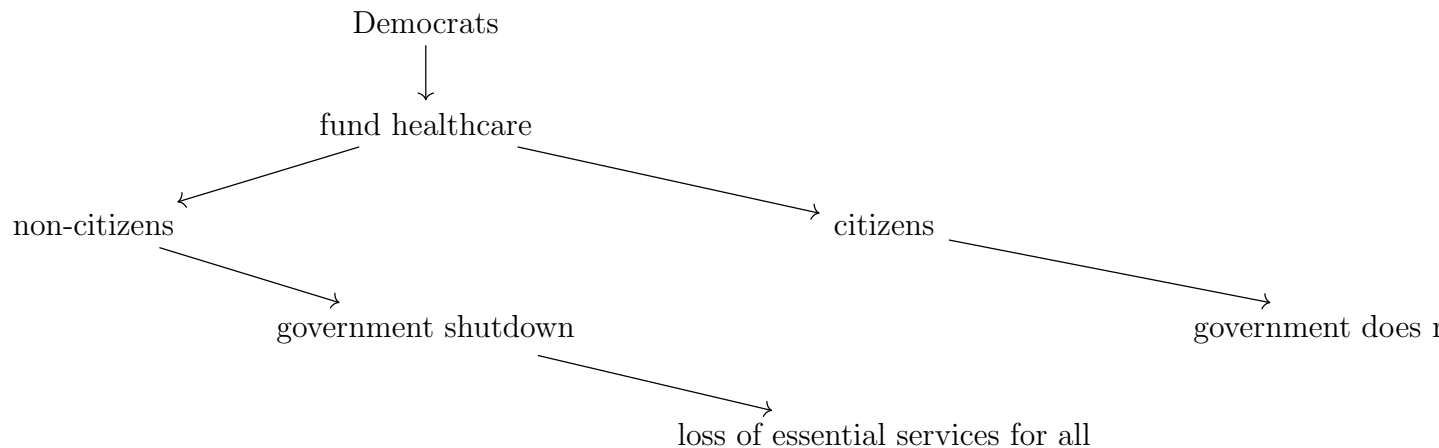
An ideal experiment here would be to research communities where access to Tylenol and vaccines is limited by socioeconomic circumstances rather than a belief system. This kind of experiment would be generational and track groups that engaged with Tylenol, vaccines, both as well as a control group.

Potential Confounds:

- Socioeconomic status: Wealthier families may be more likely to both avoid vaccines AND have better access to early autism diagnosis/services
- Education/beliefs: Parents who avoid vaccines may also avoid seeking autism diagnoses due to stigma or distrust of medical establishment

Statement 2

Causal Statement: Democrats are about to shutdown the government because they demand we fund healthcare for illegal aliens.” - JD Vance



0.0.1 Analysis Questions:

Counterfactual:

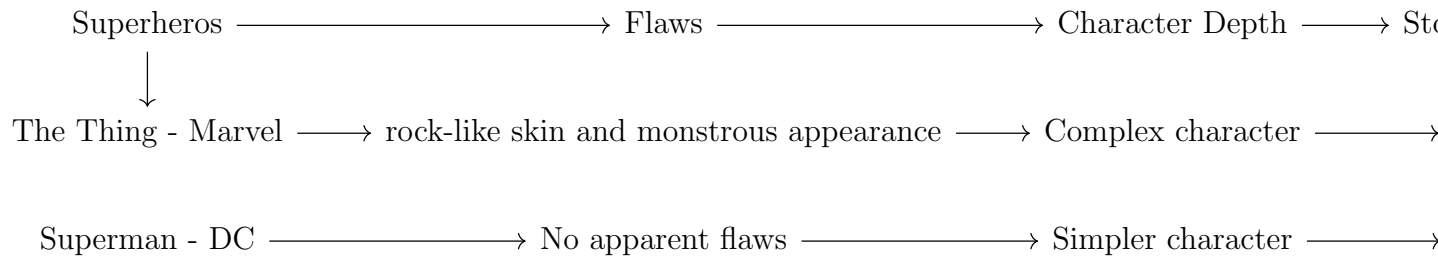
The reverse implication of this statement would be that Democrats do not fund healthcare for non-citizens and the government stays open.

Potential Confounds:

- Budget negotiations complexity: A shutdown might ultimately occur due to a failure to negotiate a multitude of complex issues. Publ
- Public opinion: This may be an effort to pressure Democrats to cave by controlling the narrative.
- Distraction: This could be an effort to distract from actual healthcare funding disputes (Medicare, Medicaid, ACA tax credits).

Statement 3

Causal Statement: "Marvel is better than DC because they have less stereotypical superhero ideas like The Thing versus Superman. The Thing's powers can impede him. They stop him from living a normal life because of his rock-like skin and monstrous appearance, but Superman has all the benefits of being a superhero - speed, superhuman strength, but none of the downsides."



Analysis Questions:

Counterfactual:

If DC characters had superpowers balanced with flaws (like The Thing), their characters would have more depth, and their stories would be better.

Ideal Experiment:

Create an extensive and agreed upon rubric for evaluating the complexity and depth of superheroes and apply it to Marvel and DC.

Potential Confounds:

- Fan community expectations: Fans of superheroes might have developed certain norms and expectations of characters over time.
- Artistic talent: Individual writer/artist quality may matter more than publisher philosophy - Jerry Siegel/Joe Shuster vs. Stan lee/Jack Kirby
- Cherry-picking examples: Selecting specific character pairs (Thing vs Superman) rather than representative samples.
- Publication era effects: Characters created in different time periods may reflect different norms, conventions, expectations, and behaviors.