Ben Tankus
Data Viz Roast
01-31-21

<u>**Visual 1:**</u> **A Wordcloud of Every Inaugural Address in the History of the United States (1789-2021) – From Reddit user teddyterminal**

Note: The first half of the speeches are cropped out to make space, here we start around 1890.

Word clouds are hard to glean meaningful information from. I think this visual gets really close to the best a word cloud can be, but I do have some suggestions.

The Good:
I really like how the visual splits the blocks into one inaugural speech per block which keeps the lie factor low and matches the number of dimensions with the dimensions of the data. I also like the color-coded party affiliation which shows the flip-flopping of party victories very well.

The Not-so-Good:
Like most word clouds, it doesn't really tell me anything. One would expect Donald Trump's word cloud to be uniquely full of words like "wall", "America", "division", etc. but the words on his cloud are not much different than Obama's and Biden's, and they were very different presidents. I would recommend dropping the most common words like "America", "citizen", "democracy", "world", and any others past a certain count to see if this is wasted ink or not. If that was removed *maybe* it would be more clear.

Recommendation:
I would recommend adding a visual to clarify modern trends. It's great to see the history all the way back to George Washington, but we aren't really learning anything other than "oh cool, they had inaugurations back then too." It's a good door-buster visual to grab the users attention, and then I would recommend a tree-plot or a simple bar chart with word count.

Visual 2: US Dog & Cat Ownership by State – From Reddit user takeasecond

This visual tries to compare how many households own dog to how many own cats, and I honestly think it's pretty bad.

The Good:

I think the concept and general layout is good, two choropleth maps with clear headings and scale. The metric is also decent and easy to understand, percentage of households that have the pet is easy to visualize. The number of dimensions also match the number of dimensions in the problem.

The Bad:

There is a lot wrong with this visual. First, there is too much black in the visual, the unused ink here puts an unnecessary dark tone to the visual. Second, I don't think the visual is based off correct numbers. The cat map shows most states with 15-25% ownership and then has 65%+ in West Virginia? That doesn't make any sense, and I don't think the representation is accurate. One could argue that the color axis is a lie factor, I can't tell the difference between 30% and 50% as those colors aren't typically considered sequential. Typically, heatmaps are done using one color at different intensities, or at least conventionally sequential colors (ex. green, yellow, red).
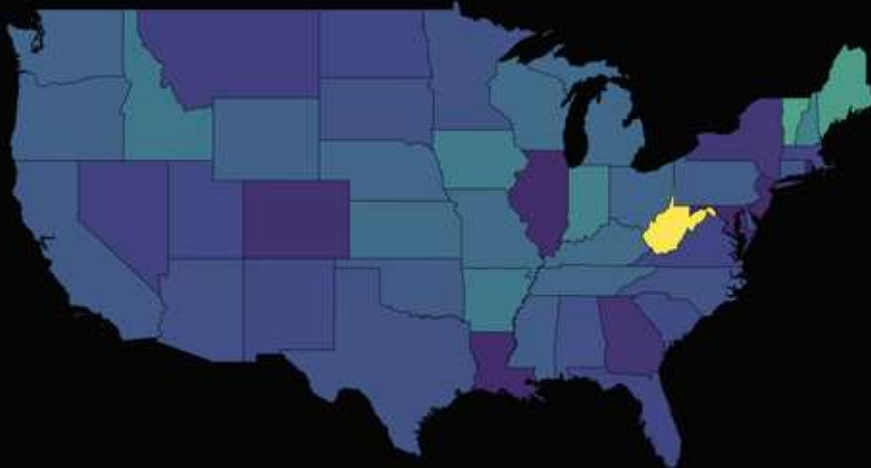
Recommendation:

Obviously number one is to fix the numbers. Data visualizations are useless if you are using bad data. I would also maybe change the metric to consider mixed-species houses as this data may be double counting. Many houses that have one also have the other. I would also recommend improving the background color, and the data color, making the data sequential colors. It's also pretty difficult to compare state-to-state, so maybe a supporting visual of a clustered bar (clustered by state) would help in digesting the data.
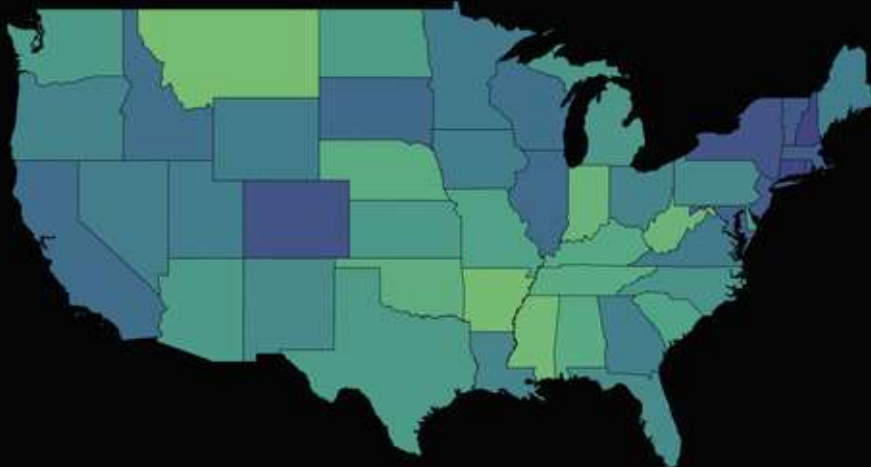
US Dog & Cat Ownership by State
% of Households | 2020

Cat Ownership

Dog Ownership

65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%

Visual 3: Glint job interview visual

This is my visual, and I think it is perfect. I was given a dataset to analyze for a job interview and chose to report out the results as a residual plot. The purpose of this visual is to communicate which locations are "beating" the model's prediction and which locations need help.

The Good:

I don't use ink in poor places, it has very clear indications of where I want the user to look, and the color scales of the residuals clearly indicate "good" and "bad". The representation is clear and does not have a lie factor.

The Not-So-Good:

It's difficult to critique my own visual because if I knew what was wrong, I would fix it. The residual zero-line could be a bit misleading compared to the 90th and 10th percentile lines (dashed). And the text on the side is a bit messy, but that is required based on the PowerPoint slide size constraint.

Recommendation:

I think this is the best visual I can produce in this given category. I could make the visual more engaging by adding more icons and art, but that would change the category from visual to infographic, and risks a messier slide. Really the best thing for me to improve this visual with is gaining more education and experience.