

PDF file containing a written description of your project. There are no specific page or word limits. This document should contain:

A. A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection. (10 pts)

The data we selected reports the results of monthly national surveys by Cards Against Humanity from September 2017 - March 2018 (excluding December 2017 and January 2018, when the poll was limited exclusively to Republican-leaning respondents and Democratic-leaning respondents, respectively). We used data from <https://thepulseofthenation.com/#future>.

Each file contained a set of questions and 800-1000 rows of respondents' answers. Questions were the headers for the columns. For example, a question could be "Who would you prefer as president of the United States, Darth Vader or Donald Trump?" Respondents would answer "Donald Trump" or "Darth Vader".

Our visualization represents a selected subset of all questions from multiple months. In addition to the responses to each question, we used the respondents' reported demographic data - including gender, race, education, political affiliation, and approval of Trump - to provide more dimensionality to the visualization.

In processing the files, we discovered inconsistencies among column headers: for example, in some files, the column for gender data was labeled "Gender" while in others it was "What gender do you identify with?". To circumvent these discrepancies, when reading in data we matched columns not by exact strings but instead checked for membership, e.g.

```
if (s.toLowerCase().includes("gender"))
```

in our mutateRow function.

Sometimes the answers to the demographic-oriented questions presented at a finer level of granularity than we preferred. Our solution involved using the mutate row pattern to append columns with more generalized versions of the responses. For example, political affiliation initially read as one of 'Independent', 'Strong Democrat', 'Strong Republican', 'Democrat', 'Republican', 'DK/REF', 'Not very Strong Republican',

'Not very strong Democrat', 'Not Strong Republican', 'Not Strong Democrat'; in our visualization we further generalize to 'Independent', 'Democrat', and 'Republican'. This occurs via one of the `classify_{category}` functions, following the first `<script>` tag, and as with the column headers, these functions check for membership rather than exact matches, e.g.

```
if (s.toLowerCase().includes("democrat")){ return "democrat"; }
```

to filter 'Strong Democrat', 'Democrat', 'Not very strong Democrat', and 'Not Strong Democrat' down to "democrat". We grouped answers such as 'DK/REF' into a "residual" category, which we filter out from the final visual by the pattern of filtering by desired answers, e.g.

```
var node1 = node.filter(function (d){ return d[question] == answer1 });
```

As a result, we are able to keep only the responses that match certain answers.

To manage integration of data files, we match the question the user selects in the drop down menu to a string, e.g.

```
if( question == "Do you consider rap a form of music?"){  
    c = "201711-CAH_PulseOfTheNation.csv";  
}
```

We would then pass `c` into the `d3.csv` call, and in this way we are able to load the file that corresponds to the correct question.

B. A description of the mapping from data to visual elements. Describe the scales you used, such as position, color, or shape. Mention any transformations you performed, such as log scales. (10 pts)

We created individual circles to represent respondents. Circles represent a one to one mapping with a respondent. Circles were grouped into clusters of circles by category.

Users can look at the data in two dimensions or three dimensions.

We have three scales that breakdown groups of circles ('respondents'). First, for the horizontal x scale, we separated respondents based on their answer into two sides of the visualization using a line. Users can breakdown groups of respondents into gender and presidential approval.

For the vertical y scale, users can further break down these groups into age, political affiliation, race, and education. We position a cluster of circles vertically and separate these groups into separate sections in the page.

Lastly, each cluster can then be further broken down into groups by age, political affiliation, race, and education. Colors were used to distinguish circle or clusters of circles from each other.

We provided a legend that associates a color to a specific group such as 18-24 age or independent, etc.

In a hover feature, we break down the groups in comparison to its row, column, and other clusters who answered similarly by percentages. For example, for the question “Who would you prefer as president of the United States, Darth Vader or Donald Trump?” we show that if one were to hover on the cluster for female and republican, 63% of republicans respondents who answered Darth vader were Female and other analysis.

C. The story. What does your visualization tell us? What was surprising about it? (5 pts)

Our goal was to map each respondent of our data to a circle and group them in clusters by demographics. This visualization present survey data on a variety of topics to users giving them the ability to group and breakdown of respondents by a myriad of demographics in three dimensions.

Our data and how we integrated it into our visualization lends itself to telling a story in the details. When you first open the page you will see a seemingly random question broken down spatially by respondents answers to the question, their gender and their political affiliation. But this is just the tip of the iceberg, you can find out so much information about how people of different demographics feel about a ton of different issues. Question ranged from their feelings on rap music to who they are voting for in the next election to how they feel about peeing in the shower.

For example looking at the question, who would you rather have as President, Donald Trump or Darth Vader, 61% of respondents would rather have Darth Vader as president than Donald Trump. It was surprising to see that this question had a similar response/breakdown to would you vote for Donald Trump in 2020 where 66% would not vote for him in 2020. With our visualization, If one were to dig deeper, we can break down these groups further by age range, political affiliation education, and race. For example, if one were to choose political affiliation on the vertical axis, we show an overwhelming majority of republicans would vote for donald trump over darth vader. Next, on a third dimension, we were able to breakdown these groups further into age, education, race and political affiliation.

Thus, our visualization allows the user to select many combinations in terms of demographics and three dimensions to breakdown respondents into multiple groups.