

Congress in the Social Media Space

Grant Keller - u0822035 - u0822035@gcloud.umail.edu

Jairon Terrero - u0879843 - jairon.s.terrero@gmail.com

Repo: <https://github.com/CrownedPhoenix/dataviscourse-project>

Background Information:

Neither of us have a particular background in politics or political science. However, we certainly find information and data in this sphere to be rather interesting. This particular dataset contains information about the usage of social media by members of the 116th Congress between 2015 and 2020. The primary reason we chose this dataset (beyond its intrigue) was because it was relatively organized - we wanted to be able to focus more of our time and effort on the visualization rather than wrangling the data into a workable format.

Project Objective:

Above all, we don't want our presentation of the visualization to carry any explicit bias towards any particular interpretation. As such, we plan to design our visualization in a way that facilitates individual exploration over directed storytelling. We realize that creating a visualization that is completely free of bias is an unrealistic goal, nevertheless, we want to enable observers to come to their own conclusions as much as possible.

All this said, there are a number of questions that we hope to allow users/observers to derive information on (either obviously or via some amount of exploration). These include:

- What are the top terms used most exclusively by members of the two primary political parties?
- On which platform did each political party post the most/least?
- How did the effect on engagement for a particular term change between 2015 and 2020?
- Which terms fostered the most engagement?
 - On which platforms?
 - By which parties?
- How did the usage of social media platforms (and the effect thereof) by various Congress Members differ?
 - By platform?
 - By party?
- How did a particular Congress Member's usage of social media (and the effect thereof) compare to the rest?
- What periods of time corresponded to some of the strongest surges of engagement?

These are only a few of the questions that we hope users will be able to answer, though they are perhaps the most significant. We expect that there may be additional questions for which information can be gleaned via exploration.

Data:

The data we are using was collected by Pew Research Center and was published here:

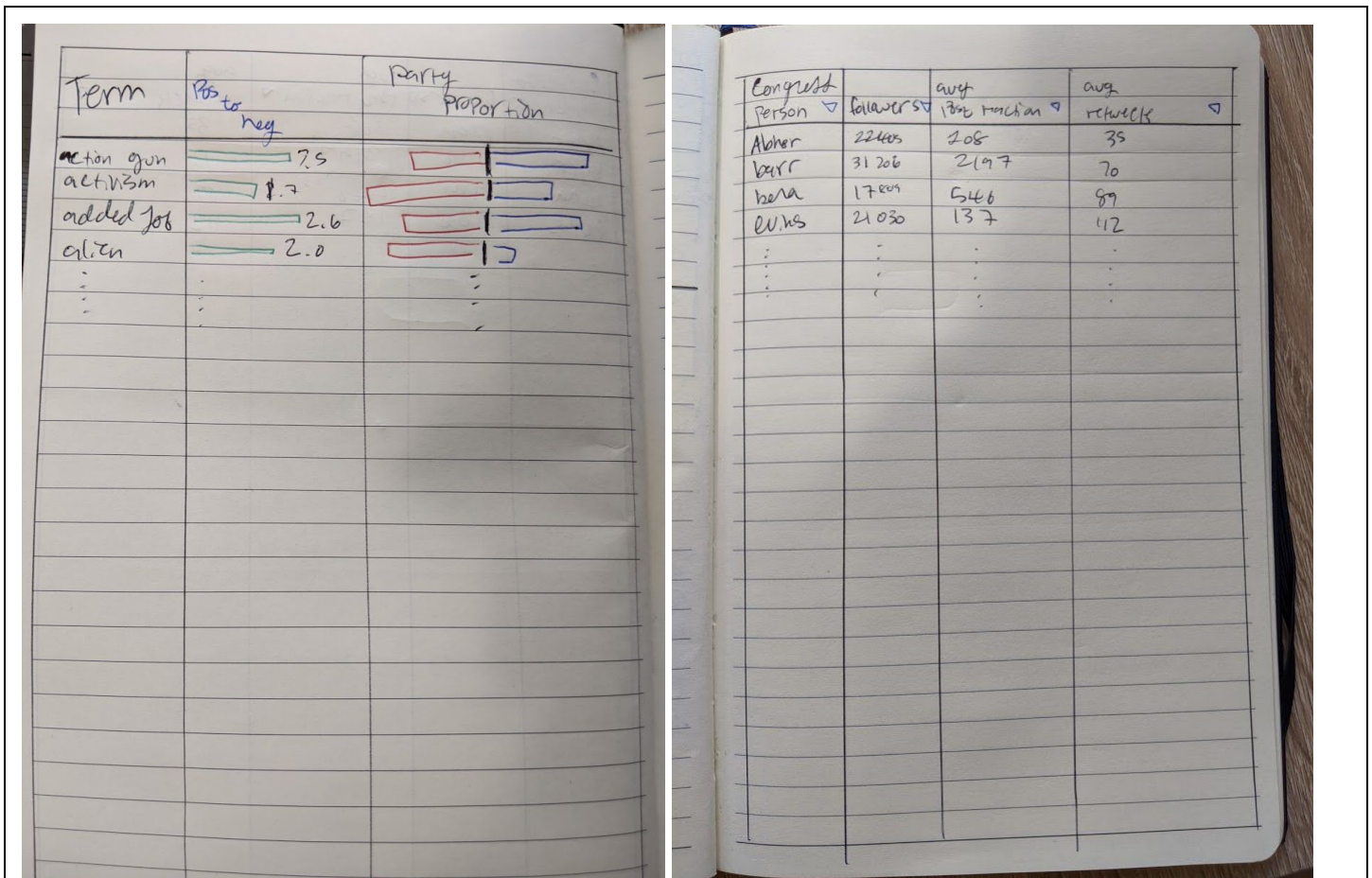
<https://www.pewresearch.org/internet/2020/07/16/congress-soars-to-new-heights-on-social-media/>

Data Processing:

As stated above, part of the reason we chose this data set was because of how well organized it was. Thus, we do not plan on spending much time doing data processing.

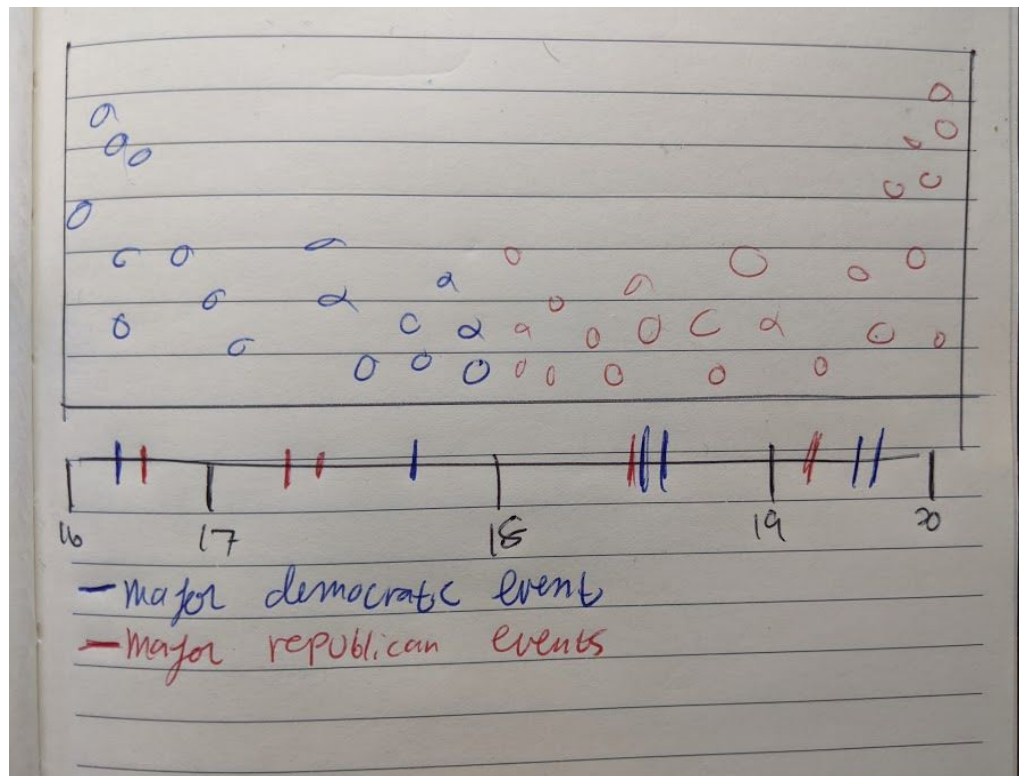
Visualization Design:

Each of our various designs are detailed below in order of their drafting. The later designs tend towards making amendments to and improvements on previous designs - though there are some exceptions. For the most part, the final design is what we expect to follow as our guide.



This design is reminiscent of something we did in Homework 6. The diverging bars make a great encoding of predominance on either side of the red/blue (Rep/Dem) spectrum. Sorting functionality will further aid interpretability and (most importantly) facilitate exploration.

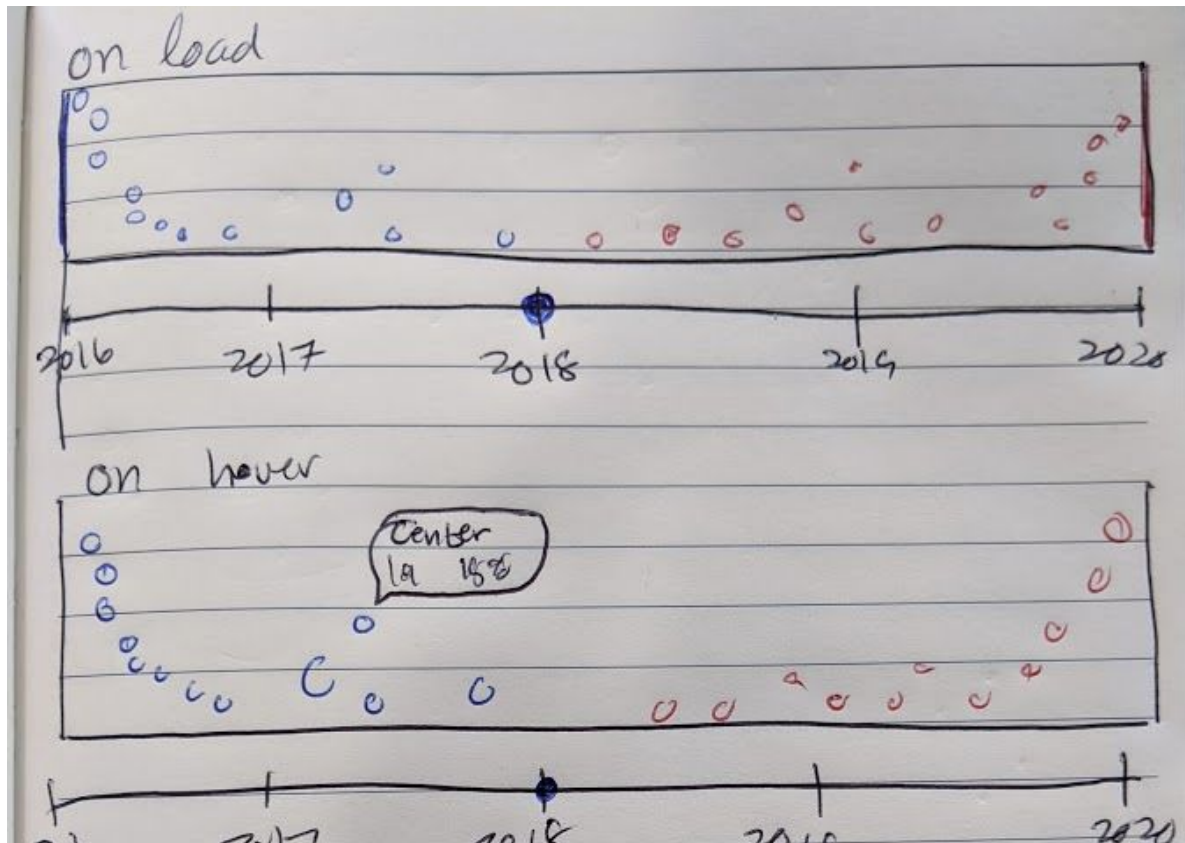
Since this is something we've done before, it seemed appropriate to include it given our experience.



A take on overlaying a display of time regions/windows onto existing data. Each of the red/blue tick marks indicate the beginning or ending of a time region. Hovering over any of the tick marks would shade (semi-transparently) a region of the data chart above. This allows the user to brush their mouse across the bottom axis and see important windows of time pop out.

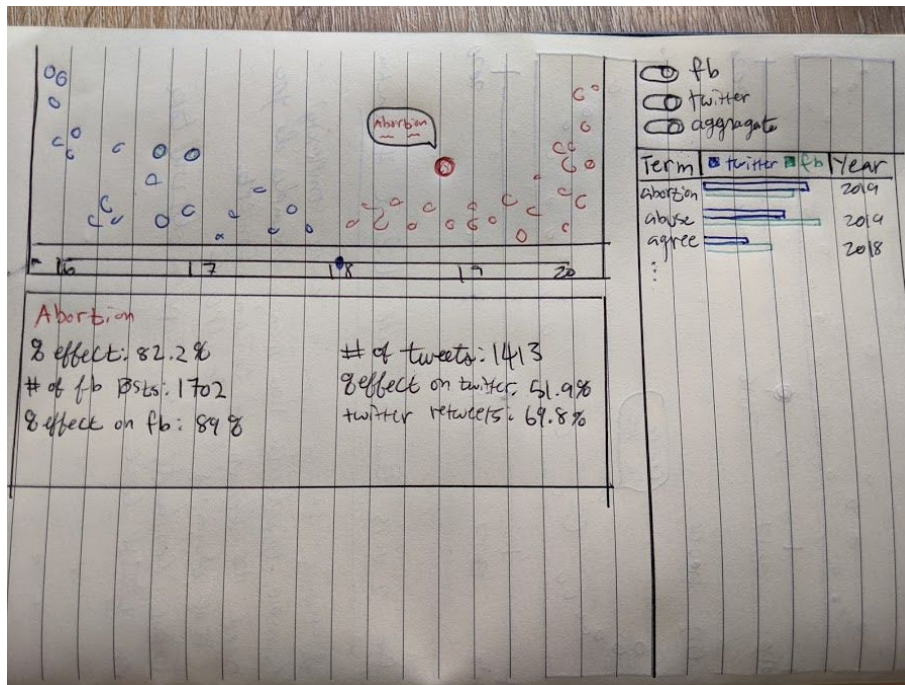
Specifically, this functionality catered to a piece of our data that captured time regions associated with a significant increase in engagement for a particular party.

Ultimately, we decided against this model. However, the format of the dot-plot above did continue to evolve.



This is another viz that has some semblance to a previous homework assignment. The time slider at the bottom changes the position of the dots on the chart and a tooltip displaying additional information about a datapoint will appear on hover.

More details about the specifics of the dot plot below.



This is where the dot-plot really started to take shape. The dot plot is split into two halves: the left half displays Democrat datapoints while the right half displays Republican datapoints. Each half has its own bottom axis which represents the total amount of social media posts including the phrase associated with the datapoint. These axes are diverging (we ultimately decided against this).

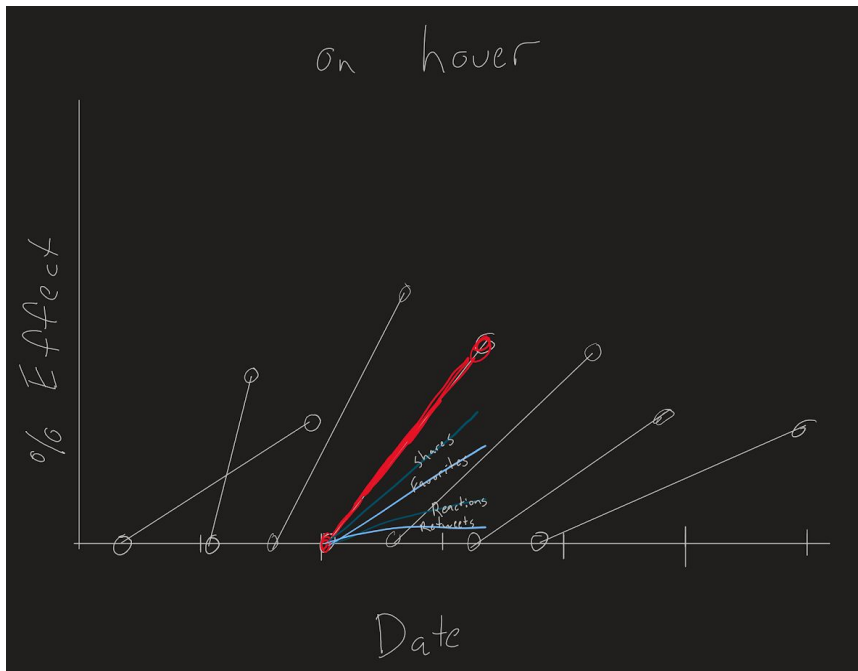
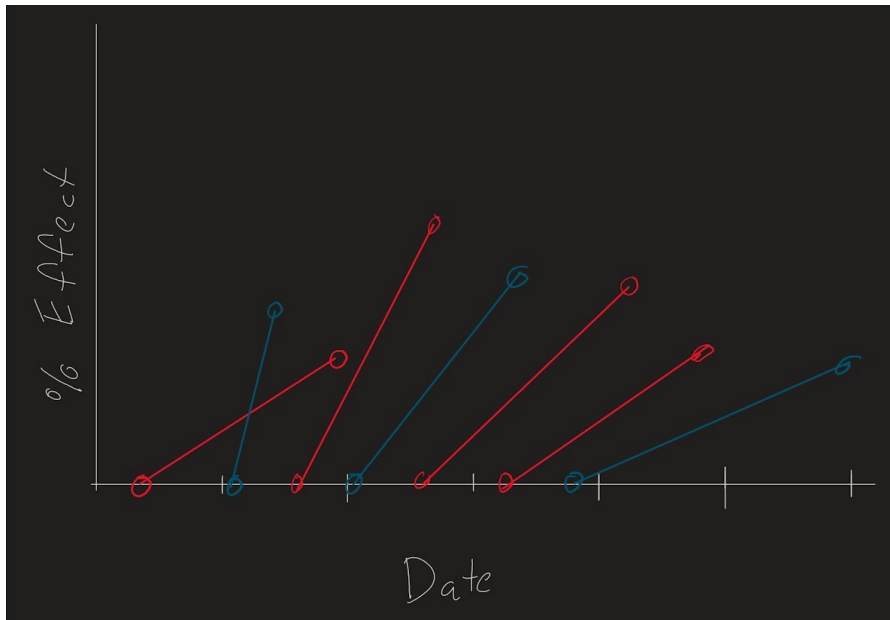
The vertical axis represents the average percent effect that the post containing a given phrase had on engagement. Thus, dots near the top are associated with greater engagement.

As a whole, the dot plot encodes the relationship between the frequency of a term and its effect on engagement. Some terms may be used few times, but have a relatively large effect on engagement, while there may be used a great deal and have a relatively small effect.

On the right hand side of this design are some switches that modify the data shown in the dot plots. This allows the user to select which social platform they want information on, or if they want an aggregate/average of both.

Below these switches is a tabled barchart showing a redundant encoding of the difference in average engagement between two platforms for a particular phrase as well as the year.

Finally, an info box is provided on clicking any data point (another callback to previous assignments).

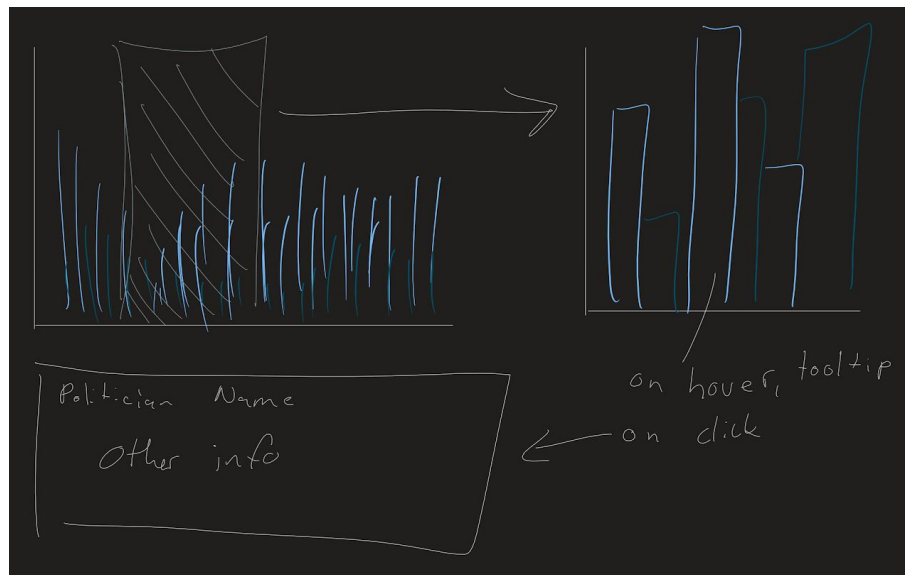
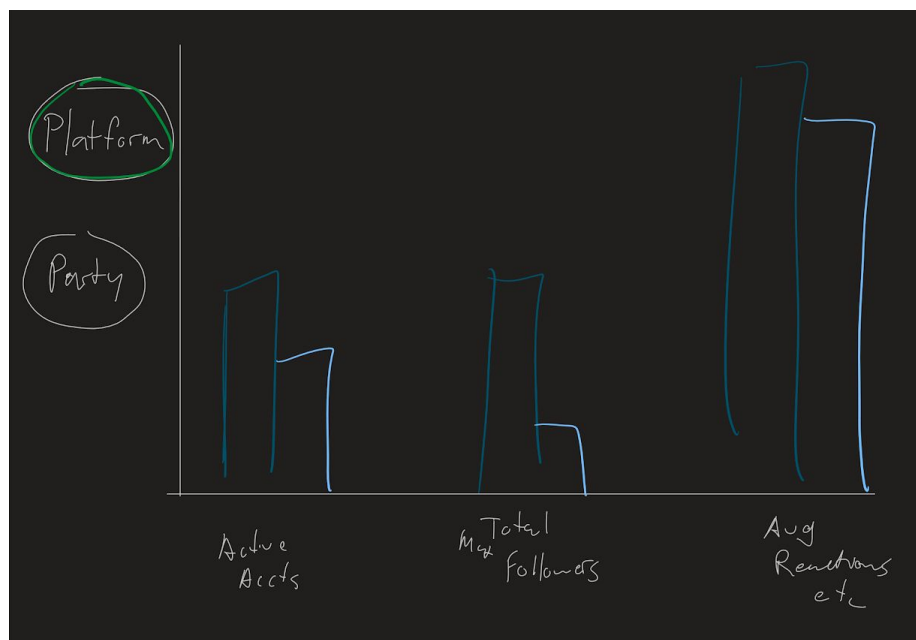


Here we have an updated attempt on displaying important regions of time associated with significant increases in engagement. The bottom axis here represents the various dates across the temporal landscape. The vertical axis represents the percent effect on engagement at a given time index.

The lines between two dots represent related datapoints - more specifically, the beginning and the end of an event. The vertical difference and the line between the first and the second point redundantly encode the difference in engagement. This allows the user to use the slope of the line to see windows with large effects on engagement relative to other regions.

This particular part of the dataset has only 100 data points - but it seems that this could become crowded with too many datapoints.

The second pane shows that additional details about the composition of a particular line can be rendered on click. The other lines will grey-out and the selected line will appear with its constituents.



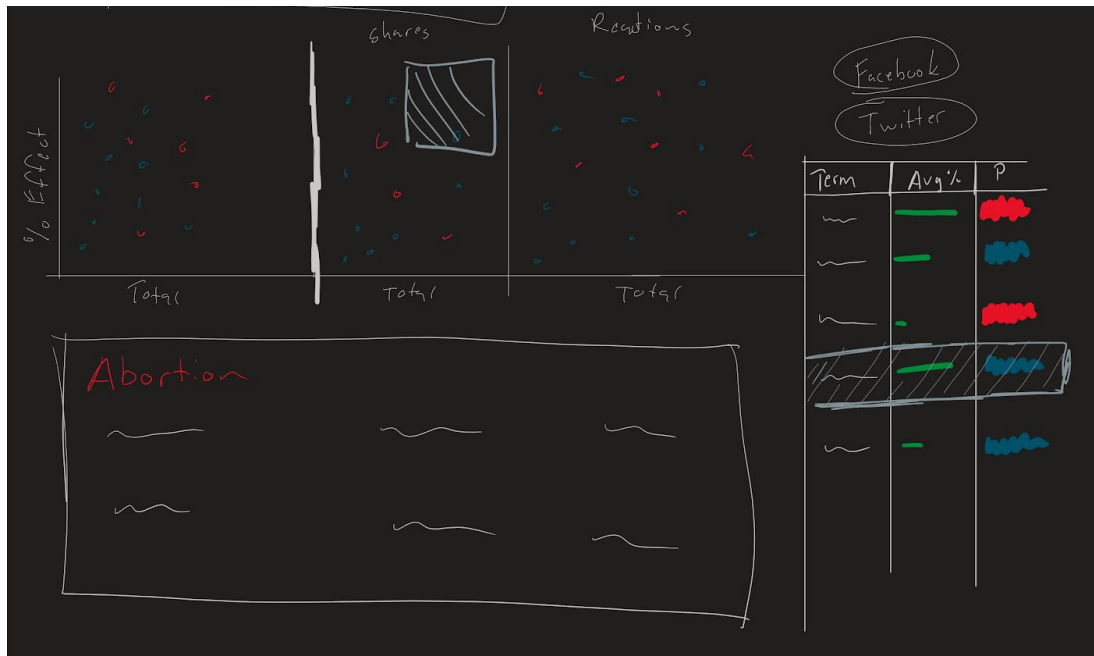
The top diagram is deliberately simple. It shows the relationship between the total/aggregation of a particular feature between either party or platform (depending on what is selected on the left).

The most important piece of this sketch is the second pane. When the user clicks on one of the bars in the top pane, the second pane reloads its dataset to display the constituents of the aggregate. Since there are hundreds of bars, we use a scrubber/brush to allow the user to hone in on a particular portion of the chart they are interested in (presumably high or low points). Then inside the zoomed/brushed section, they can click on a bar to get details about that particular datapoint.

For example, consider the line of thought: It seems that Facebook is the platform that tended to garner the most followers, but which Congress Members were associated with accounts that contributed most to that number? And what other details are there about that Congress Member?

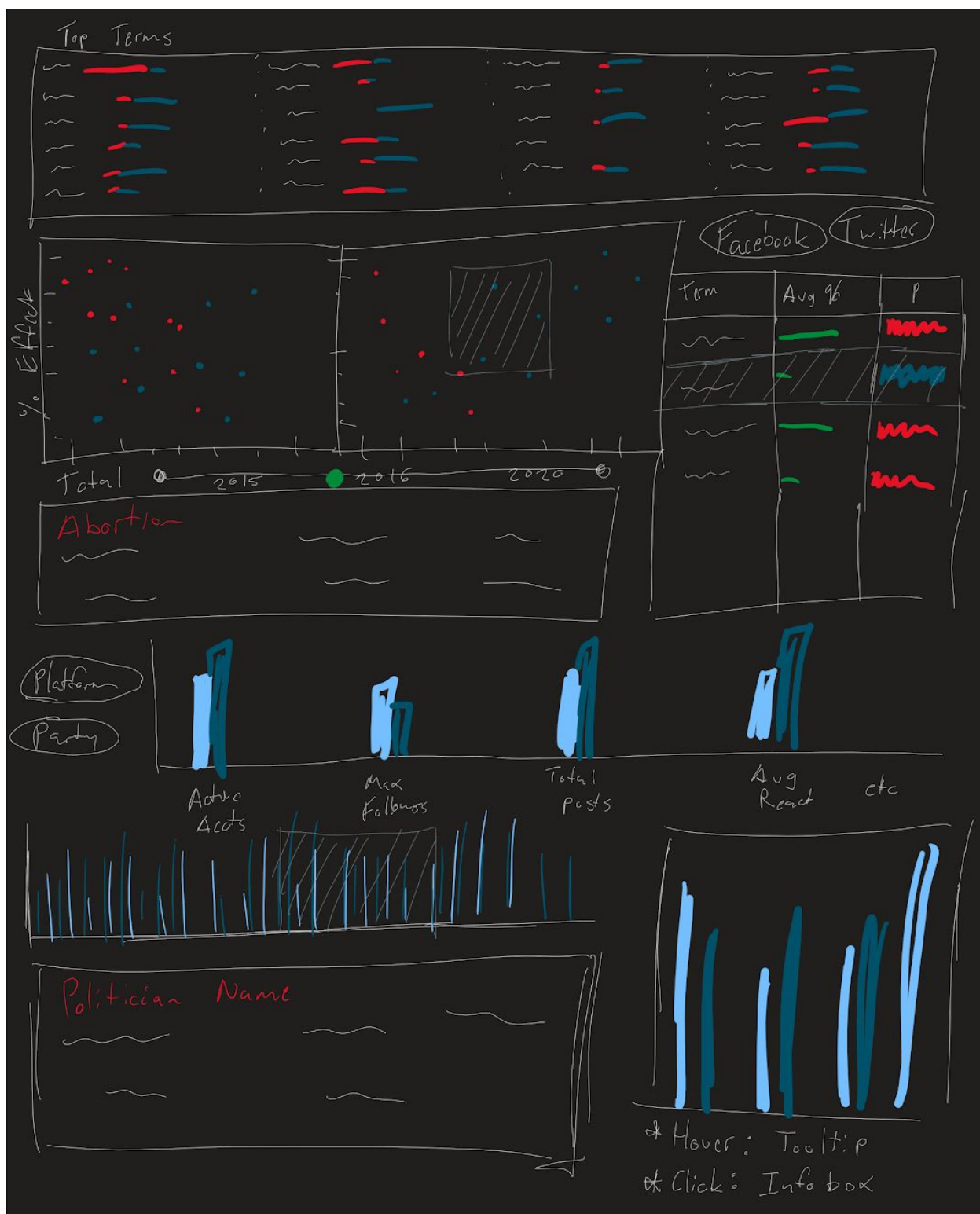
One improvement to this sketch might be to make the bottom pane display a single bar chart instead of a double bar chart and also allow the user to toggle whether the bars are colored by party instead of platform.

Sorting functionality is a given as well.



This is an improvement on the dot-plots from above. Here the axes remain the same but the three separate dot-plots encode: the average % engagement effect across all platforms, the % effect on reactions/favorites, the % effect on shares/retweets. The latter two dot-plots swap between platforms using the buttons to the right. They can be brushed.

The brushing will highlight the dots in all the dot-plots as well as their associated rows in the bar table to the right. The bar table redundantly encodes the average % effect, but with the terms easily displayed. The table also allows the user to sort and make selections on the data table itself that will also be highlighted in the dot-plots



This final sketch incorporates all the previous sketches with some slight variations. This will be our baseline for our development process - though we may make some minor changes as we deem them appropriate.

We may not be able to complete every section of this visualization as there are a number of relatively complex pieces, but we hope to do most (if not all) of them.

Must Have features:

See the final sketch diagram. It is composed of 3 primary pieces: the top terms chart with divergent bars, the dot-plot section about engagement, and the bottom barchart section about Congress Member social media details.

We want to finish at least the first chart and also one of the two following charts. In other words, the two possible combinations we would consider successful are (1,2) and (1,3). There are a lot of components to each of these sections as detailed in the sketches.

Some important features for each of the three components:

1. Top Trends:
 - a. Sort feature
 - b. Horizontal stacking of the table that makes it use horizontal real estate instead of vertical
2. Dot Plots:
 - a. Toggle between platform
 - b. Info box
 - c. Dot-plots
 - d. Brushability of plots
 - e. Link between bar-table and dot-plot brush/highlighting
 - f. Bar-table
 - g. Time slider
 - h. Preservation of highlighted dots while sliding time slider
3. Bar Chart:
 - a. Toggle party/platform
 - b. Click to detail attribute constituents
 - c. Brush over constituents
 - d. Toggle between party platform display on constituents
 - e. Info box on bar click

Optional Features:

Our optional feature will be the third component of the three. We will complete at least two, and if we can, the third.

Project Schedule:

We plan to meet biweekly to discuss our progress on the project and schedule time to work on it and/or make commitments.