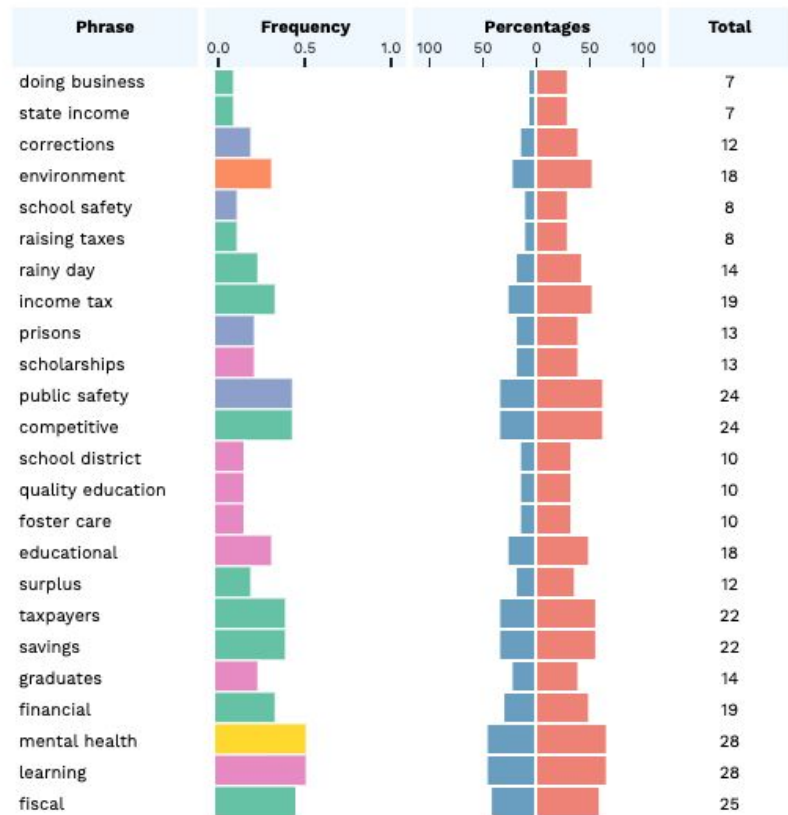


Overview and Motivation

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.

Related Work

We drew some inspiration from one of the assignments we did in class since it catered to the dataset we had found.



Questions

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?

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/>

As mentioned previously, part of the reason we chose this data set was because of how well organized it was. Thus, we have not needed to spend much time doing data processing.

Exploratory Data Analysis

We actually didn't analyze the data beyond determining what data was available and the way it related to each other. This was somewhat deliberate since we didn't want to bias ourselves in any particular direction.

As such, we viewed the data in its raw form as csv files (in Excel).

The next few pages detail the data within our dataset. These details came with the dataset themselves as a PDF.

Social media statistics for members of the 116th Congress

This file contains information on members of Congress' posting activity and engagement on Twitter and Facebook during their time in office in the 116th Congress. Each member is included twice, once for Facebook and once for Twitter.

Variable name	Variable description
Platform	The platform on which the politician was active
Party	The politician's party. Independents are associated with the party they caucused with for the majority of the study period, but identified with parentheses.
Bioguide ID	The Biographical Directory of the United States Congress identifier for the politician.
Number of active accounts	The number of active accounts the politician had on the platform while serving in the 116th Congress (up to May 31, 2020).
Max total followers	The highest observed total number of followers the politician had across all of their accounts on the platform during the time that they served in the 116th Congress (up to May 31, 2020). This value may be empty if researchers did not capture follower data for one or more of their accounts during the period in which it was active on the platform.*
Start of term	The start of the politician's term in the 116th Congress.
End of term	The last day the politician served in the 116th Congress (if applicable).
First post	The date of the first post the politician created on the platform in the 116th Congress.
Last post	The date of the politician's most recent post created on the platform in the 116th Congress (up to May 31, 2020).
Total posts	The total number of posts the politician created on the platform while serving in the 116th Congress (up to May 31, 2020).
Average post favorites/reactions	The number of favorites (for Twitter) or reactions (for Facebook) that the politician's average post receives. On Twitter, this value is only computed for original or quoted tweets; retweets are excluded.
Average post retweets/shares	The number of retweets (for Twitter) or shares (for Facebook) that the politician's average post receives. On Twitter, this value is only computed for original or quoted tweets; retweets are excluded.

*For example, Greg Gianforte has two Twitter accounts but has not posted on either since July 2019. Because researchers did not begin tracking both accounts until September 2019, there is no available follower count information for the period in which Gianforte was active on Twitter.

"Congress Soars to New Heights on Social Media"

Top 10 events associated with increased engagement in each party

This file contains, for each party, the 10 time periods that were associated with the greatest average increase in engagement for the median legislator's average post, across all four metrics of engagement: reactions and shares on Facebook, as well as favorites and retweets on Twitter. Events are defined as beginning on a day in which all four of these measures increased by at least 10% relative to the prior day, and end when one of these measures has a negative day-over-day change. For more information, see the report methodology.

Variable name	Variable description
Party	The party for which the event was associated with higher levels of engagement.
Start date	The day in which the median party member's average post began experiencing at least 10% higher engagement (across all four metrics) relative to the prior day.
End date	The last day in which all four engagement metrics continued to increase day-over-day.
Average percentage change	The average of the four engagement change metrics.
Percentage change in Facebook reactions	The overall percentage change in the median party member's average Facebook post reactions, as measured by comparing the last day of the event with the day prior to its beginning.
Percentage change in Facebook shares	The overall percentage change in the median party member's average Facebook post shares, as measured by comparing the last day of the event with the day prior to its beginning.
Percentage change in Twitter favorites	The overall percentage change in the median party member's average tweet favorites, as measured by comparing the last day of the event with the day prior to its beginning.
Percentage change in Twitter retweets	The overall percentage change in the median party member's average tweet retweets, as measured by comparing the last day of the event with the day prior to its beginning.

"Congress Soars to New Heights on Social Media"

Top 100 terms associated with engagement boosts in each party

This file contains a list of the 100 terms associated with the highest average predicted increases in engagement for posts from members of a specific party in a specific year (2015-2020). The effects were determined using statistical models that estimated the engagement that posts from members of a particular party received in each year, based on the terms mentioned in the posts. Separate models were trained for each party, year and engagement metric (reactions and shares on Facebook, and favorites and retweets on Twitter). The effects are expressed as a percentage difference between a model's prediction for the engagement that would be received by a post from the median party member that used a specific term, relative to the model's prediction for a post from the median party member that did not use any of the terms included in the model. For more information, see the report methodology.

Variable name	Variable description
Term	The word or phrase (ngram) associated with increased engagement for a given year and party. The terms shown represent the cleaned/stemmed version that was used in the analysis.
Year	The year in which the term was associated with higher engagement for posts created by members of the specified party.
Party	The party for which the term was associated with higher engagement in the specified year.
Average percentage effect	The average of the four different engagement effects.
Number of Facebook posts	The total number of Facebook posts the term appeared in (among posts created in the specified year by members of the specified party).
Percentage effect on Facebook reactions	The effect of the term on a Facebook post's reactions, expressed as a percentage difference between the predicted reactions for a post from the median party member that used the term relative to one that did not mention any terms.
Percentage effect on Facebook shares	The effect of the term on a Facebook post's shares, expressed as a percentage difference between the predicted shares for a post from the median party member that used the term relative to one that did not mention any terms.
Number of tweets	The total number of tweets the term appeared in (among tweets created in the specified year by members of the specified party).
Percentage effect on Twitter favorites	The effect of the term on a tweet's favorites, expressed as a percentage difference between the predicted favorites for a tweet from the median party member that used the term relative to one that did not mention any terms.
Percentage effect on Twitter retweets	The effect of the term on a tweet's retweets, expressed as a percentage difference between the predicted retweets for a tweet from the median party member that used the term relative to one that did not mention any terms.

Top 100 terms used most exclusively by members of each party

This file contains a list of the 100 terms associated with the highest average predicted increases in engagement for posts from members of a specific party in a specific year (2015-2020). The effects were determined using statistical models that estimated the engagement that posts from members of a particular party received in each year, based on the terms mentioned in the posts. Separate models were trained for each party, year and engagement metric (reactions and shares on Facebook, and favorites and retweets on Twitter). The effects are expressed as a percentage difference between a model's prediction for the engagement that would be received by a post from the median party member that used a specific term, relative to the model's prediction for a post from the median party member that did not use any of the terms included in the model. For more information, see the report methodology.

Variable name	Variable description
Term	The word or phrase (ngram) associated with a large party difference. The terms shown represent the cleaned/stemmed version that was used in the analysis.
Party	The party that used the term more often.
Positive to negative ratio	The ratio of between "Proportion of party" and "Proportion of other party" (used to select the top terms).
Proportion of party	The proportion of party members that ever mentioned the term on Facebook or Twitter while serving in office between Jan. 1, 2015, and May 31, 2020.
Proportion of other party	The proportion of members in the other party that ever mentioned the term on Facebook or Twitter while serving in office between Jan. 1, 2015, and May 31, 2020.

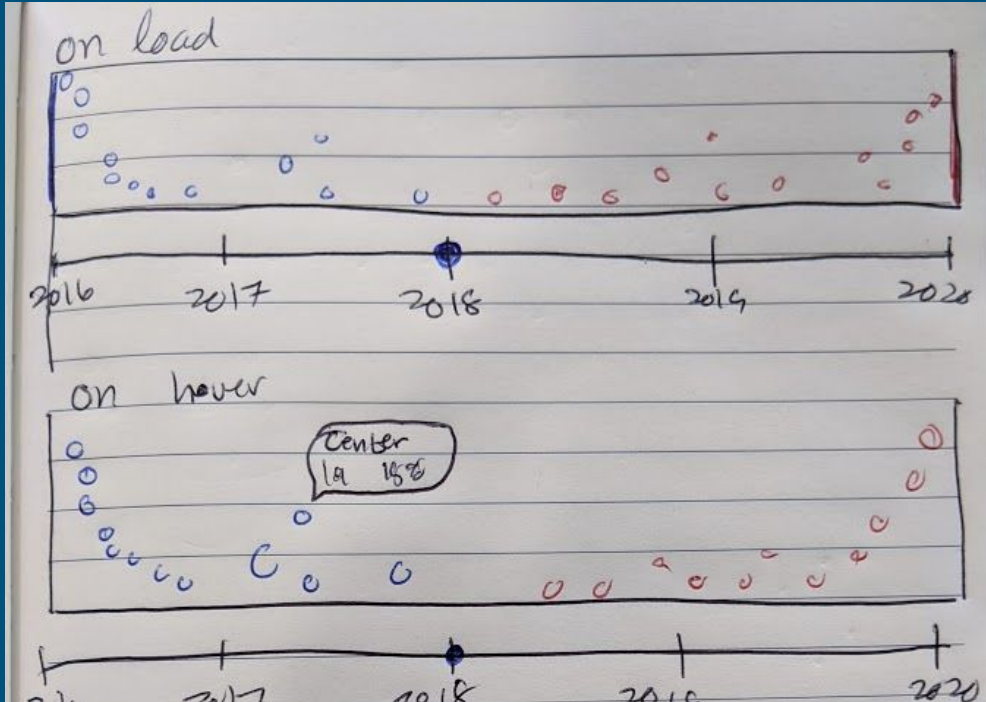
"Congress Soars to New Heights on Social Media"

PEW RESEARCH CENTER

Design Evolution

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.

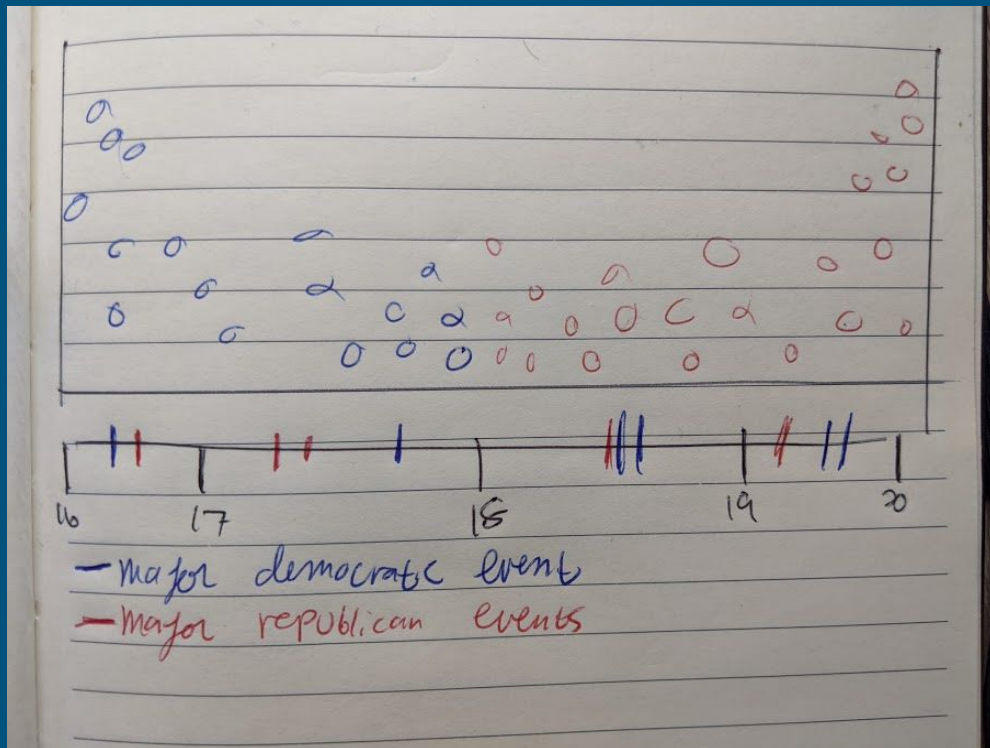
Plot Chart



We considered this plot chart to visualize words and the response they got against which side used what words.

Using a plot chart is great because it is clear and very familiar to everyone. We chose to color the dots based on which party used the word in the highest frequency. We also decided to add a slider that would enable the user to explore the changing of the words over time. The last design element we decided on was to use a tooltip to further explore the data.

Plot Chart

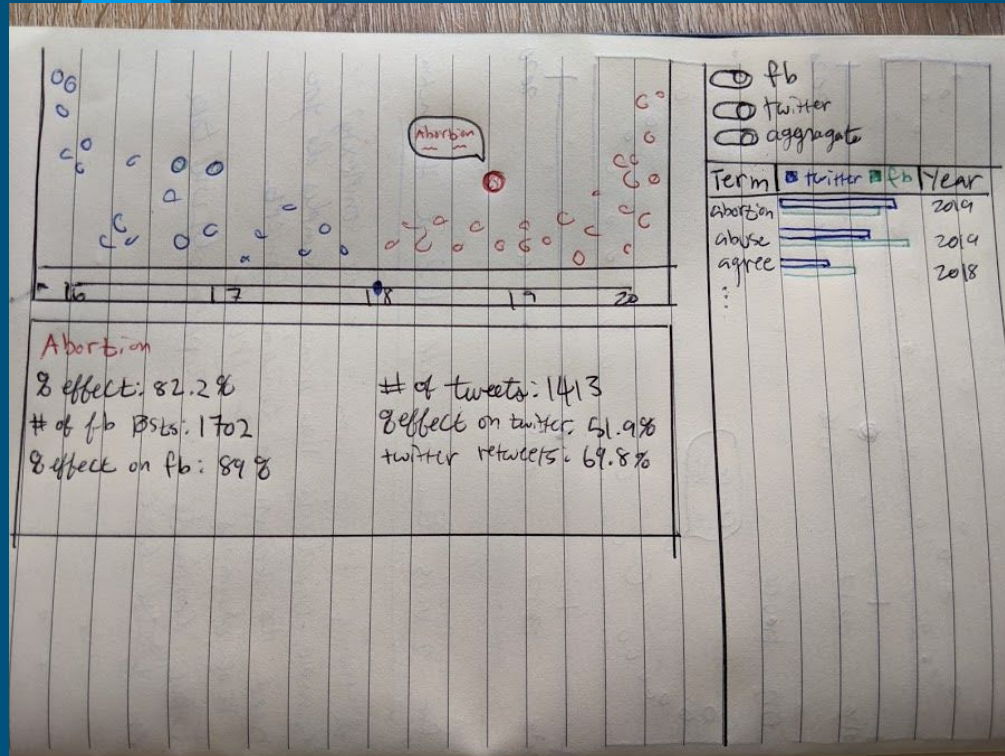


A take on overlaying a display of time regions/windows onto existing data. Each of the red/blue tick marks indicate the beginning or ending of a time region. Hovering over any of the tick marks would shade (semi-transparently) a region of the data chart above. This allows the user to brush their mouse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

Plot Chart Suite



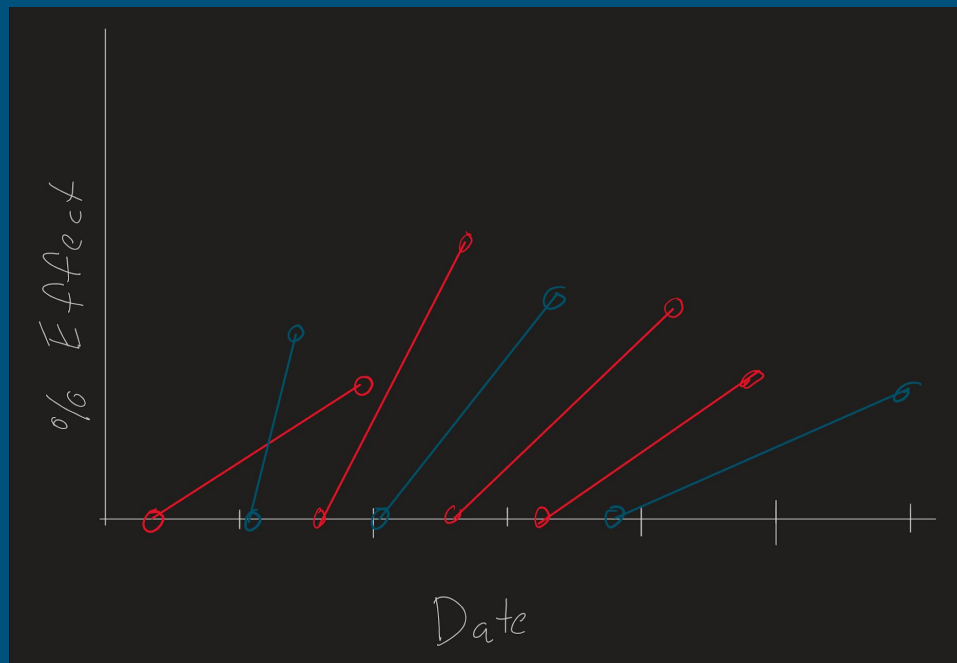
This is where the dot-plot really started to take shape. The dot plot is split into two halves: the left half displays Democrat datapoints while the right displays Republican datapoints.

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).

Event Chart

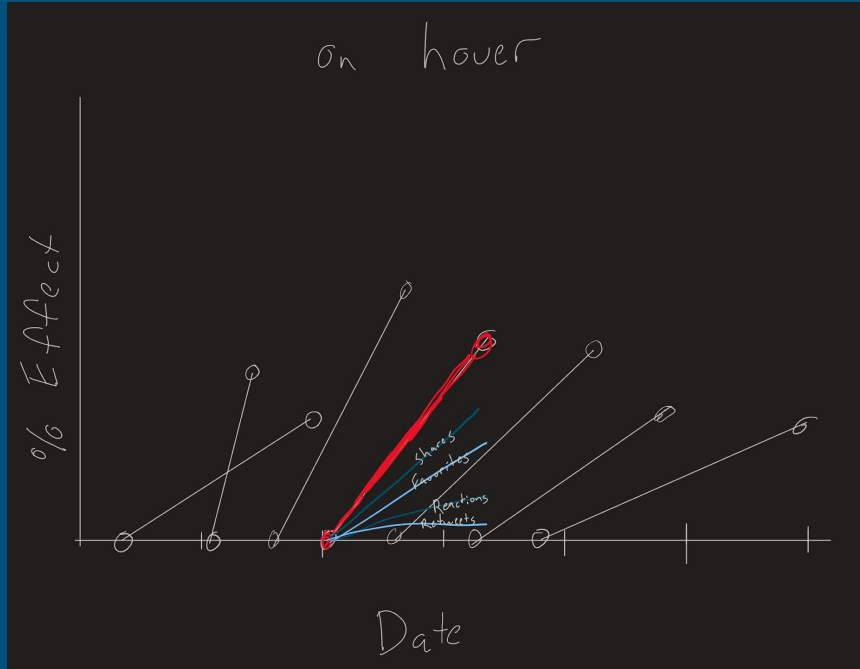


Here we have an updated attempt on displaying important regions of time associated with significant increases in engagement. The bottom axis here represents the various dates across the temporal landscape. The vertical axis represents the percent effect on engagement at a given time index.

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.

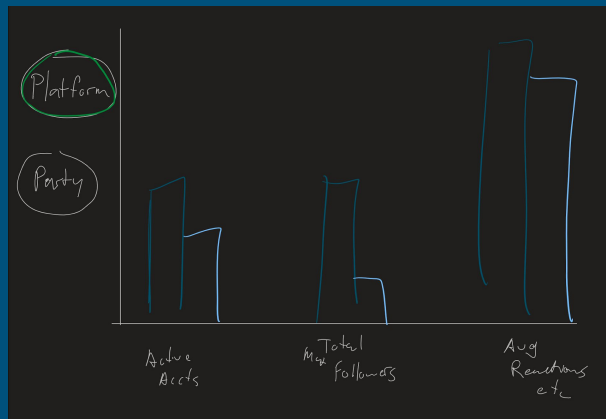
This particular part of the dataset has only 100 data points - but it seems that this could become crowded with too many datapoints.

Event Chart



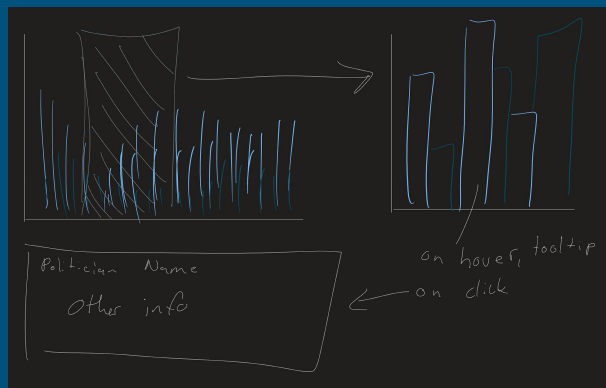
This second pane shows that additional details about the composition of a particular line can be rendered on click. The other lines will grey-out and the selected line will appear with its constituents.

Bar City



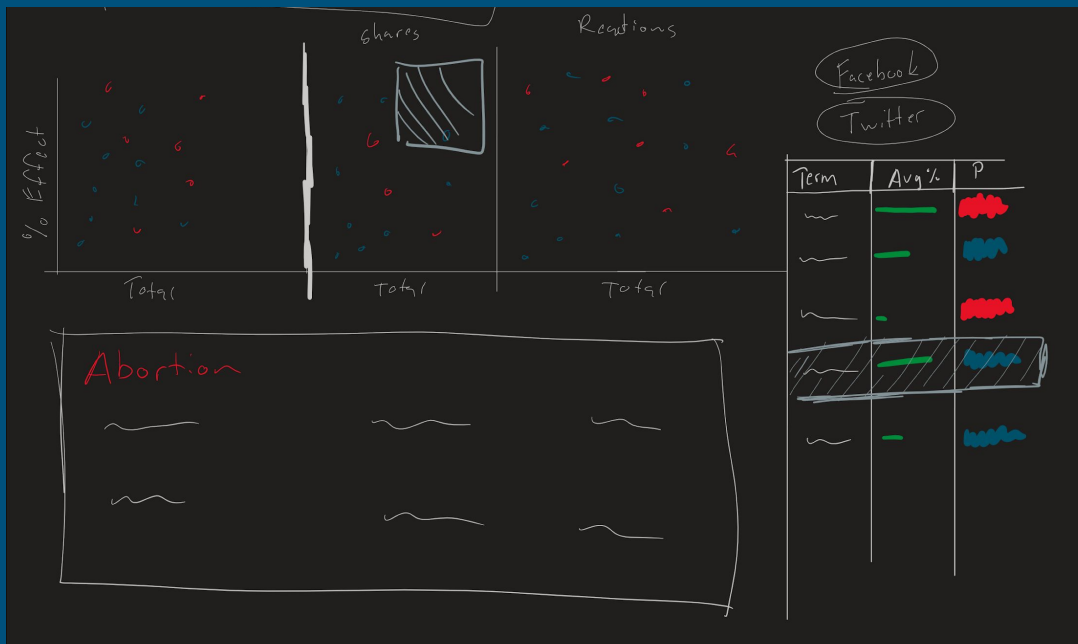
The top diagram is deliberately simple. It shows the relationship between the total/aggregation of a particular feature between either party or platform.

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.



Sorting functionality is a given as well.

Plot Chart



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

Final Draft



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.

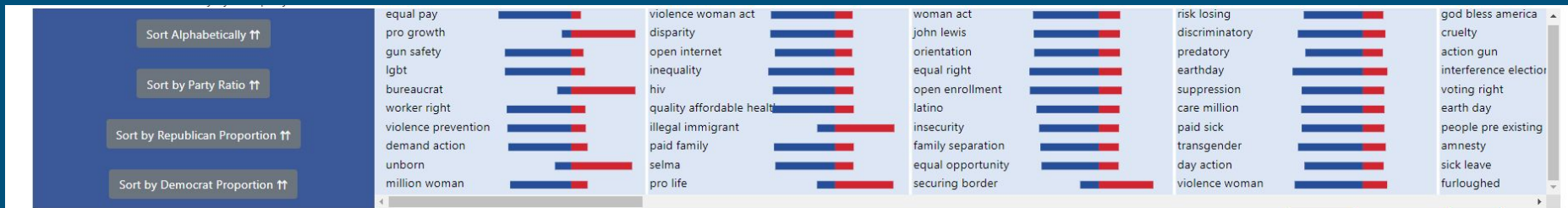
We may not be able to complete every section of this visualization as there are a number of relatively complex pieces, but we hope to do most (if not all) of them.

Implementation

Our goal was to design a visualization that was interactive and exploratory.

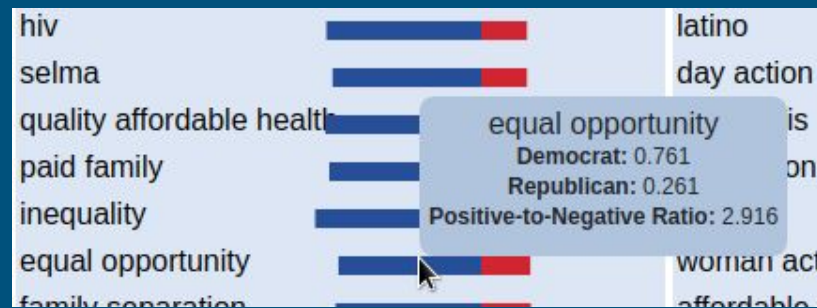
Our visualization consists of 3 tiers, each representing a different aspect of the data set.

First Tier



This part of the visualization highlights the various terms used most exclusively by members of each party. It allows the user to sort the data to highlight the terms most frequently used on social media by the respective parties.

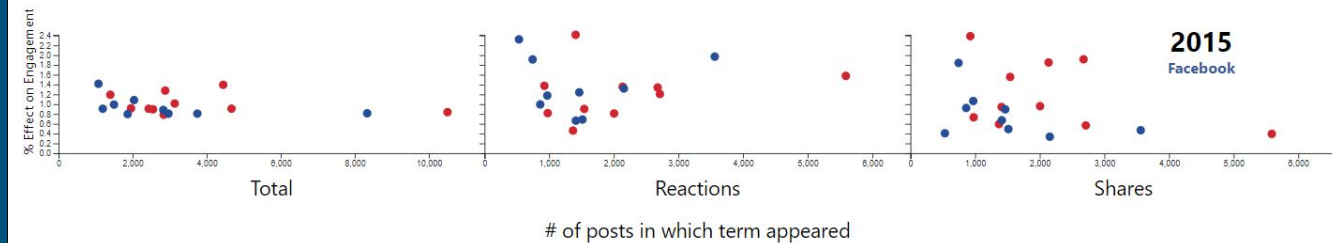
When the user hovers over a bar or term they are presented with the data in a tooltip to further allow the user to explore the usage of the terms.



Second Tier

How did different terms affect engagement?

Consider: Some terms appeared in relatively many posts but had relatively less impact compared to other terms that appeared less frequently.



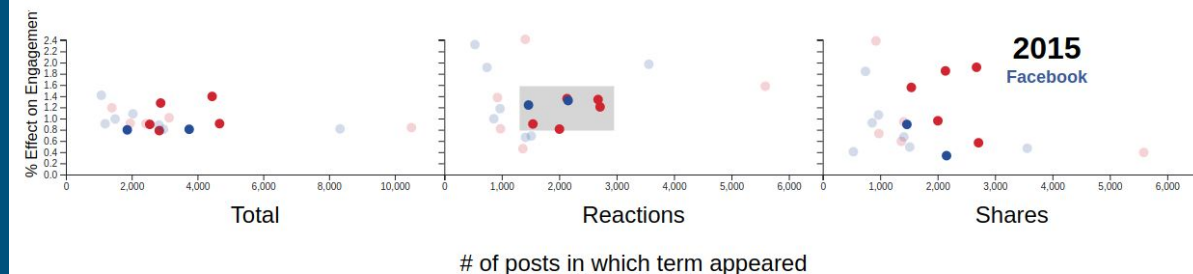
Facebook			Twitter	
Term	Avg %	Party	Reaction %	Share %
agree	1.395	R	1.34	1.916
planned parenthood	1.277	R	1.356	1.851
syrian refugee	1.192	R	1.375	2.387
happy birthday	1.014	R	2.415	0.943
netanyahu	0.915	R	0.817	0.732
prayer	0.909	R	1.208	0.567
clinton	0.907	R	0.462	0.593
refugee	0.896	R	0.903	1.556
happy	0.838	R	1.577	0.394
year ago	0.784	R	0.811	0.96
planned parenthood	1.416	D	1.913	1.842
iran	1.084	D	1.176	1.065
gun violence	0.992	D	0.996	0.923

This part of the visualization explores the effectiveness of terms in how the followers of congress people reacted over time. The slider allows the user to scroll through time to view how terms may have had stronger or weaker reactions during a given year.

Second Tier cont.

How did different terms affect engagement?

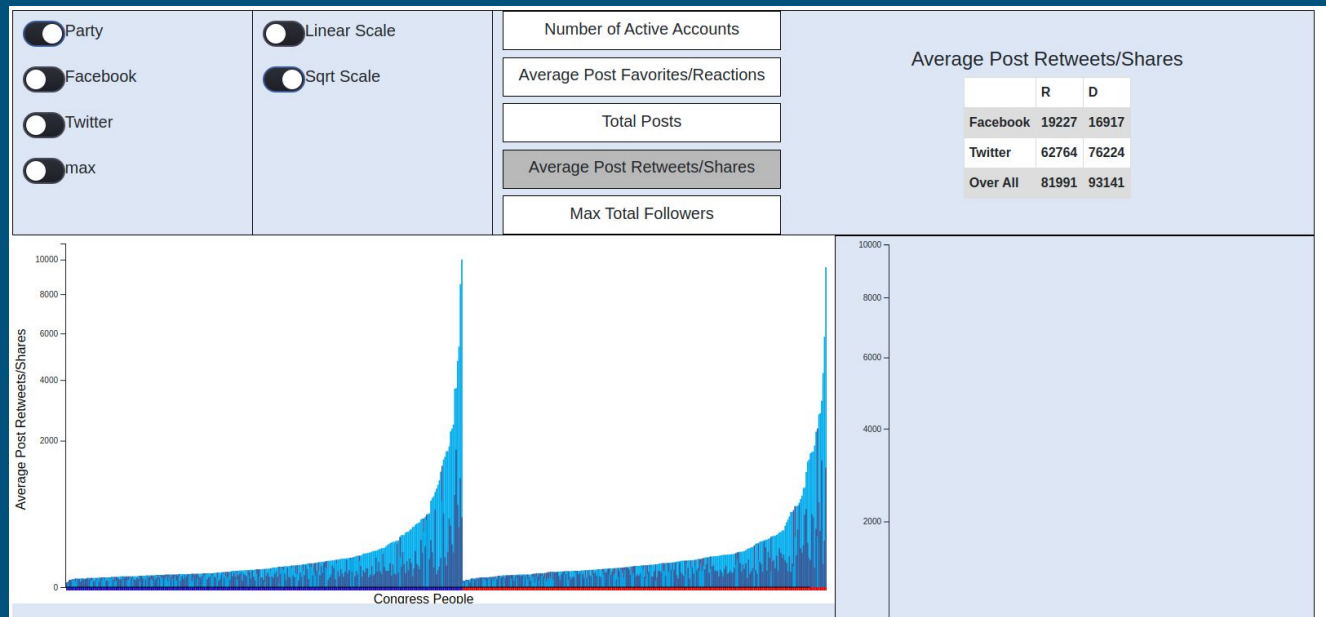
Consider: Some terms appeared in relatively many posts but had relatively less impact compared to other terms that appeared less frequently.



Facebook		Twitter		
Term	Avg %	Party	Reaction %	Share %
agree	1.395	R	1.34	1.916
planned parenthood	1.277	R	1.356	1.851
prayer	0.909	R	1.208	0.567
refugee	0.896	R	0.903	1.556
year ago	0.784	R	0.811	0.96
congratulation	0.808	D	1.321	0.337
supreme court	0.798	D	1.241	0.895

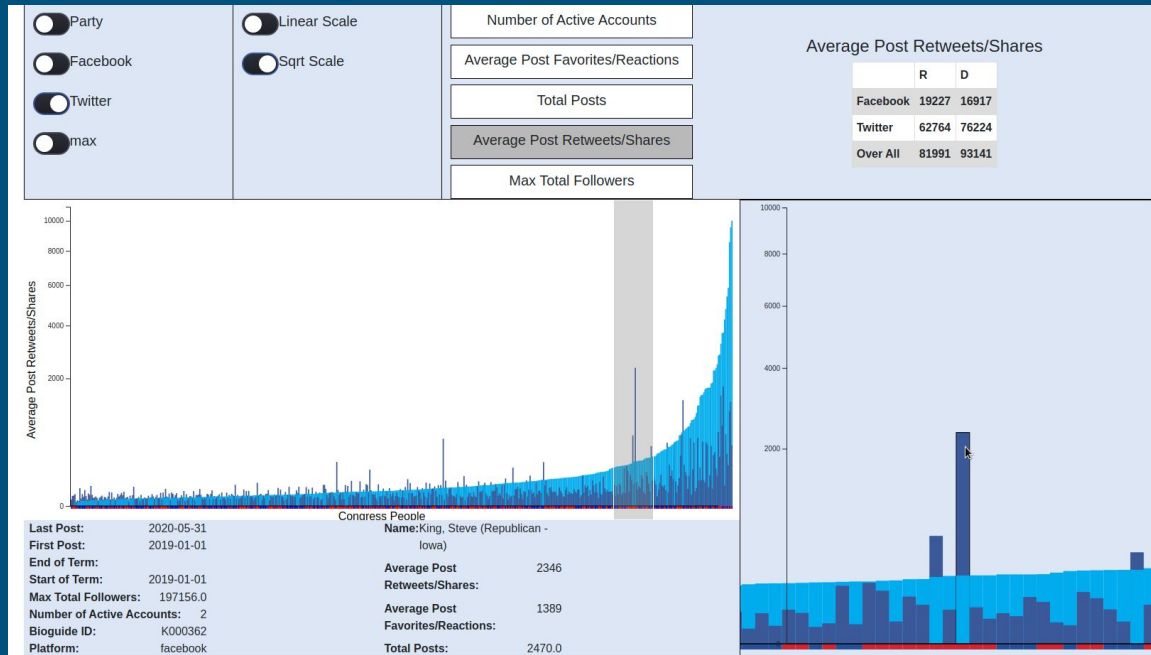
The user can also select a subgroup of points to have the group highlighted across the three charts as well as in the data table. This facilitates the ease of group selection and enhances the users understanding of how some terms might be used a lot but receive relatively low amounts of Reactions or Shares.

Third Tier



The third visualization allows the user a birds eye view of all congresspeople's platform usage. The user can change the term the graph is sorted by, the scale of the Y-axis and which feature to sort on. On selecting a new feature such as 'Total Posts' presents a new data table on the right that gives an aggregate breakdown of that feature.

Third Tier cont.



The panel on the right provides a zoom-in functionality. This is used when the brush is selected on the left hand graph. When a bar from the zoomed panel is selected the data for that bar is displayed giving the user access to the data behind the peaks and valleys and the congresspeople social platform usage.

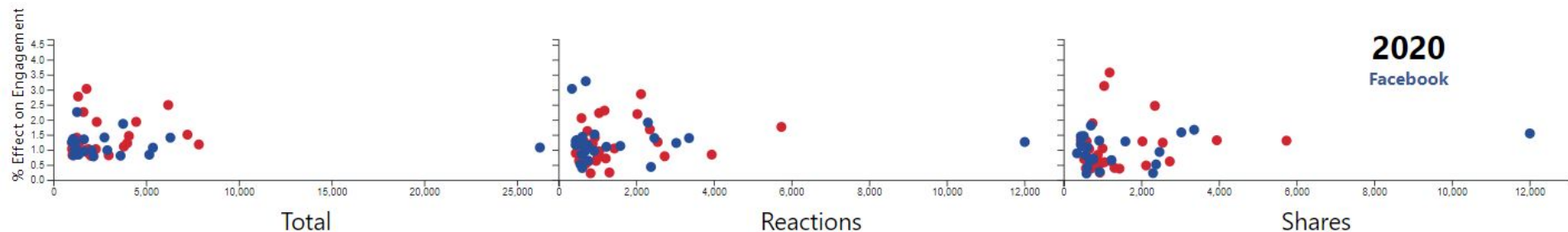
Evaluation: What we learned about the data

It was interesting to see that certain terms were very polarized between the parties. For example, 'equal pay' was used a great deal more by Democrat congress members; whereas 'pro growth' was used a great deal more by Republican congress members.



Evaluation: What we learned about the data

Another interesting thing to see was how great an outlier 'trump' was in 2020 (the blue dot on the right) by Democrat congress members compared to Republican congress members.



Evaluation: Improvements

As with all things, there are plenty of ways we could improve upon this visualization:

- Styling. The styling is a little rough. It could be improved to be more smooth.
- On the 2nd tier: It would be neat to show the history of how a point/term changed from year to year by using a tail that passes through its previous coordinates.
- We have a lot of information being displayed and in some cases it might be helpful to isolate some of it a bit more so it's easier for the user to parse.