JOUR(4|7)79V Project Proposal

For our final project, we propose investigating the unconscious biases involved in our politics. To do this, we will look at the Google autocomplete suggestions for various demographics and groups within the United States Congress. Within each of these groups, we will compute a "word cloud" of the most suggested search terms, and compare them, in order to find search terms disproportionately represented within certain groups. We expect that demographic groups, such as female or minority lawmakers, could potentially generate greater interest in specific areas, such as their appearance or life story. In addition, partisan groups, such as Republicans and Democrats, should have different associated search terms, since they should tend to tackle different policy areas. Examining these could give insight into what aspects of these perspectives are potentially misleading voters in making their political judgments.

In order to do so, we will use data from GovTrack using API provided by the site (https://www.govtrack.us/developers/api). To see if there is any difference in autosuggestion results, we ran two sets of exploratory analysis to compare gender (male and female) and party (democrat and republican). We found each keyword suggested as an autocompletion for the name of a member of Congress in each category.

First we ran word cloud analysis on each category that resulted in four figures as shown below.

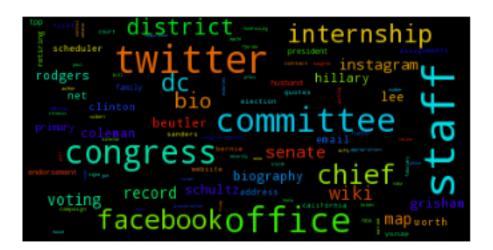


Figure 1 Autosuggestion results for female representatives

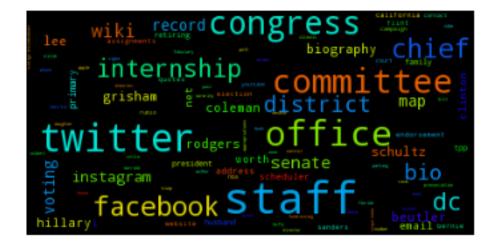


Figure 2 Autosuggestion results for male representatives

```
actor flint florida twitter in netimic ration top actor flint florida twitter in wiki hillary ballotpedia voting assignments book beutler wiki hillary ballotpedia voting assignments ted wiki hillary phone to senate biography house number scheduler paulic Committee texas gndorsement office and president california email record sanders linkedin senator district scalia internship sensentrener biography court congressment office senator district scalia internship sensentrener colemna are colemna a
```

Figure 3 Autosuggestion results for republicans

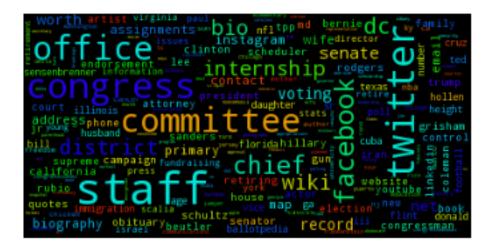


Figure 4 Autosuggestion results for Democrats

To better capture what are the most disproportionately represented terms in each category, we then compared the percentage of members of Congress in each category who had each keyword suggested. For example, for 121 Democrats and 139 Republicans, 'committees' was a suggested search term. Since there are 238 Democratic and 301 Republican members of Congress, 51% of Democrats and 46% of Republicans have 'committees' suggested. This means that the term skews 4.2% Democrat, making it the 5th most proportionally Democratic term in our preliminary results.

	Democrat	Republican	Female	Male
1	Internship	Primary	Internship	Wife
2	Office	Facebook	Staff	Congress
3	TPP	Voting record	DC office	Wiki
4	Hillary Clinton	Trump	Twitter	Congressman
5	Committees	For Congress	Chief of staff	LinkedIn

Our preliminary results definitely seem to have some associated stories. For example, every social network is primarily female except LinkedIn (which is suggested for only male politicians), but they are also primarily Republican except Instagram, which is noteworthy considering that the majority of female members of Congress are Democrat. This might also be related to the fact that most campaign-related queries, such as "for Congress," are primarily Republican, as well as the current Republican control over Congress. There is some conventional wisdom that Democrats are more focused on presidential politics, which is also potentially related.

For further analysis to discover interesting finding and construct news worthy stories, we plan to make comparisons between other possible data points provided through API by govtrack.us. For example, we can compare representatives who own social media accounts versus those who don't, or representatives who co-sponsored different types of bills with a focus on controversial ones. Comparing across different time stamps can be another option. Also we can filter out the representatives who are recently associated with some political scandals and see how autosuggestion takes different patterns across those individuals.

Through this lens, we should be able to find a number of newsworthy results related to how the public perceives Congress. Given the disparity between its historically low approval rating and the relatively high approval ratings of the individual members, this kind of aggregated analysis of the individual members might be able to provide some insight.