



# Voting Clusters in the US Senate

Rob Palinic



## Project Goal

To identify voting clusters within the US Senate and identify sub-groups below the usual Democrat/Republican divide.

This clustering model can then be used in a future app where individuals can identify how they would have voted on Senate votes and identify where they fit in relation to the identified groups.



# Background

The American public knows little about the voting patterns of their elected senators, and even less about the actual votes and decisions. It is possible that many die-hard Democrats actually agree with Republican voting patterns, and vice versa.

To further understand voting, it is necessary to understand the variance in groups and to identify diversity of behavior. It is simple to understand how similar two senators are by their voting record, but much more difficult to cluster all one hundred senators together.

This project will graphically show the voting blocs within the US Senate.



# Data Acquisition

Senate voting records for 2021 were scraped from the US Senate website.

Initially, the HTML was scraped but had some data quality issues, as well as some reliability issues (the program did not always return data). Luckily, the Senate offers an XML version which was easily accessed and provided clean and reliable data.

The data set contained 266 votes for all Senators in 2021.



## Data Cleansing

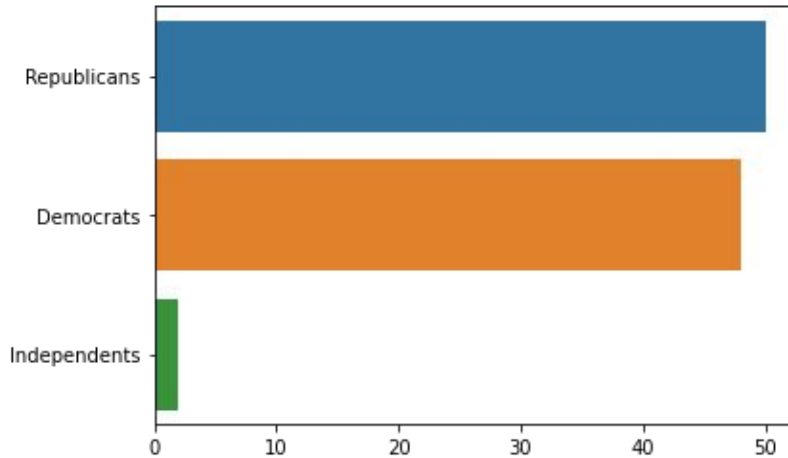
An initial decision needed to be made regarding two replacement Senators, Kamala Harris being replaced by Alex Padilla and Kelly Loeffler being replaced by Raphael Warnock. Since Harris and Loeffler participated in only a handful of votes, I elected to remove their records altogether.

A second issue arose as vote representation. Possible entries included 'Yea', 'Nay', 'Not Present', 'Guilty' and 'Not Guilty'. The 'Yea' and 'Guilty' entries were replaced with 1s, while all other entries were replaced with 0s.

Any missing values (specifically for Padilla and Warnock) were treated as 'Not Present', and given 0s.



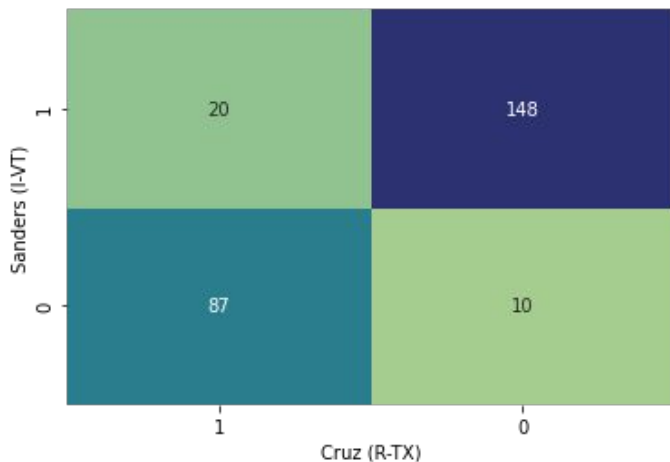
## EDA - Senators by Party Affiliation



Half of the 100 Senators are Republicans, with another 48 being Democrats and two Independents. The Independents caucus with the Democrats, and can be considered Democrats in name, if not in fact.

The Senate is therefore divided 50/50 between the two parties.

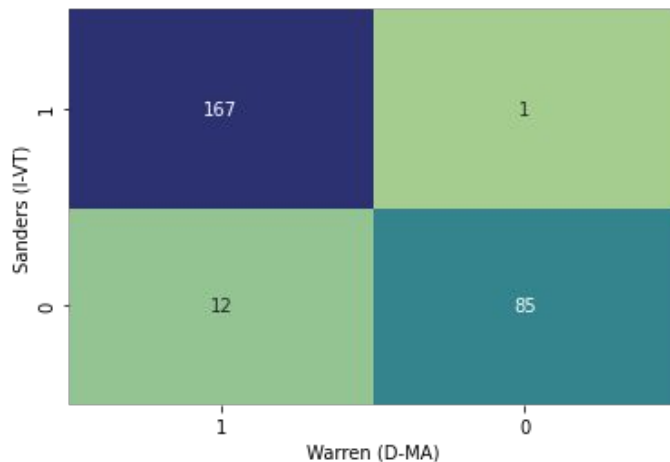
## EDA - Comparing Bernie Sanders and Ted Cruz



An initial exploration was between two well known senators, Bernie Sanders (considered on the far left) and Ted Cruz (considered on the far right).

The 2x2 voting matrix showed that they voted the same way only 30 times, out of 266 votes. These two senators are obviously quite different.

## EDA - Comparing Sanders and Elizabeth Warren

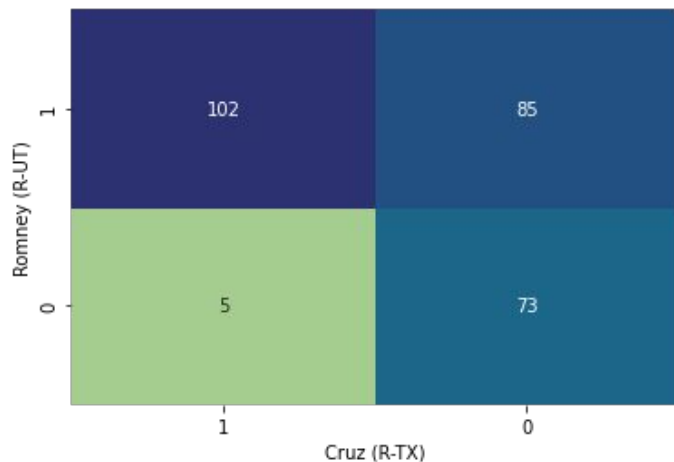


Taking the same approach, we can compare Bernie Sanders with a well known Democrat, Elizabeth Warren.

The two agreed on 252 out of 266 votes. They clearly are very similar.



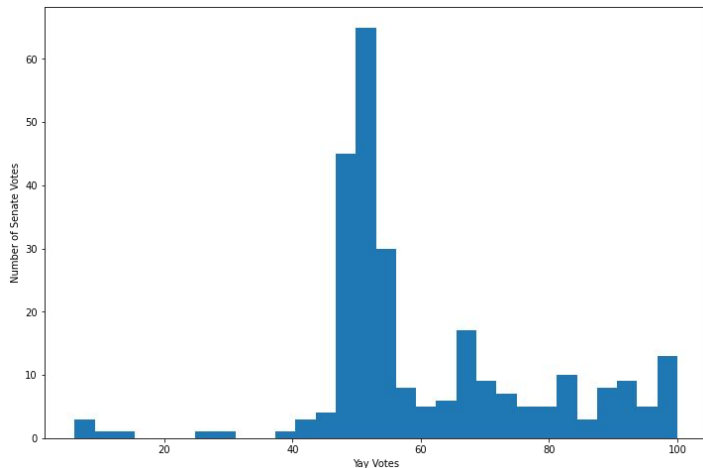
## EDA - Comparing Mitt Romney and Ted Cruz



A similar comparison between Mitt Romney and Ted Cruz shows agreement on 175 out of 266 votes.

They agreed on many votes, but clearly had many divergent opinions.

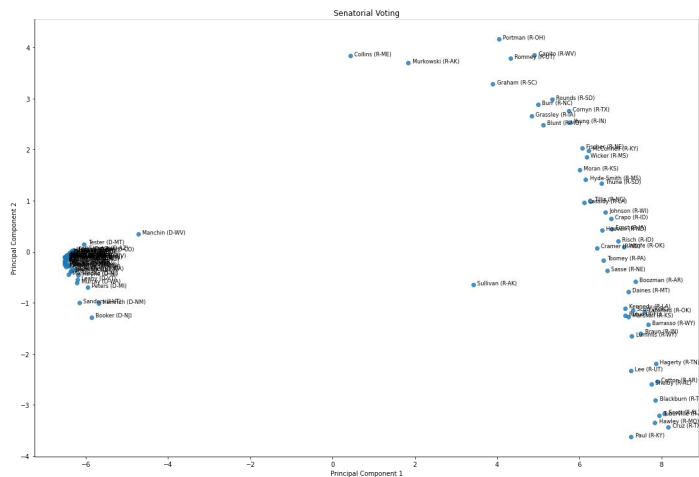
## EDA - Yeas per Vote



Most votes have around 50 votes, with many of the votes having significant majorities. This indicates party support for just about every vote in the Senate.

There are also many votes with very, very few Yea's. These appear to be individual motions without broad support, indicating senators misaligned with the rest of the senate.

# Modeling - PCA Results

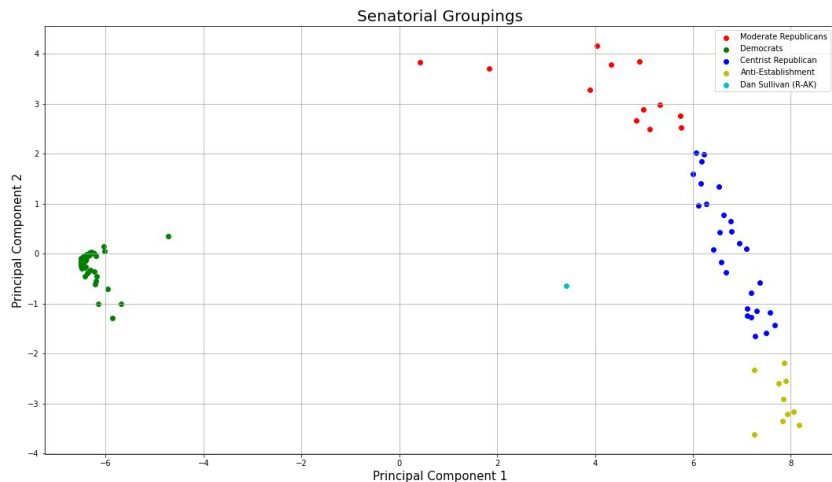


Principal Component Analysis (PCA) was used for dimensionality reduction, reducing 266 fields to 2 principal components.

Graphically, it appears that there are two main groups, the tight clustered Democrats on the left, and the wide range of Republicans on the right.

The explained variance ratios are .75 and .5, indicating a model with excellent results.

# K-Means Clustering



K-Means clustering was then used to define senatorial voting clusters. The Democrats are tightly grouped into one cluster, while three Republican clusters can be identified, plus one lone individual.

The moderate Republicans in red at the top have been identified as willing to cooperate with the Democrats. The anti-establishment group at the bottom include Ted Cruz and Rand Paul.

The lone individual Republican, Dan Sullivan, votes akin to Democrats, but has very Republican political viewpoints.



## Conclusions

The model was a success, identifying that the Democrats are a monolithic voting bloc, while the Republicans can be divided into three groups plus a lone individual.

Verification in the press indicated that the identified moderates are also identified as a bloc by the mainstream press.

As the clustering model has worked so well, creating an online website to show individuals where their voting patterns lie is a feasible project.



## Next Steps - Future Projects

1. Do the same analysis but each year of available senate data (30 years total are available) do understand how clusters have evolved over time
2. Perform the same analysis on the US Congress
3. Perform the same analysis on the Canadian House of Commons
4. Build a website to allow people to enter their own vote on Senate votes, then show where they would fit graphically and which senator is the most similar to them. This was the original goal of the project.