US Senate Voting Clusters

1. Introduction

a. Problem Statement

Identify clusters of voting behaviour in the US Senate, beyond the usual Republican/Democrat divide.

b. 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.

The original idea was to create a model where individuals could enter how they would lean on Senate votes, and find where they fit in the groupings identified.

c. Goals

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.

2. Dataset

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.

3. 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.

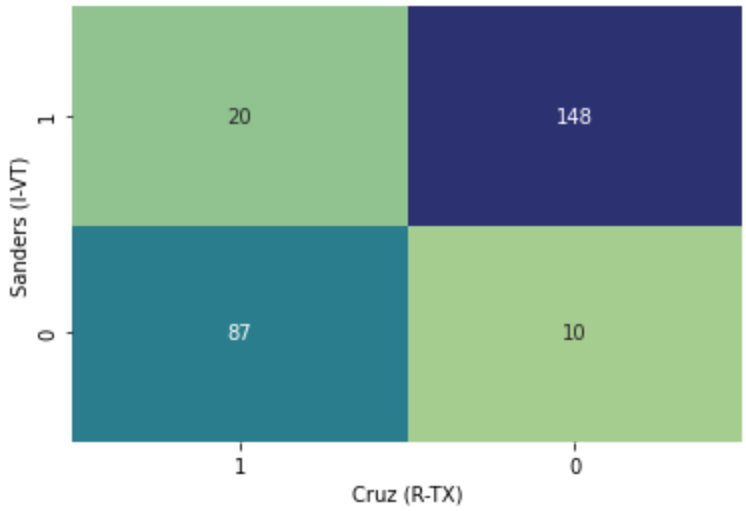
A second issue arose as vote representation. Possible entries included 'Yea', 'Nay', 'Not Present', 'Guilty' and 'Not Guilty'. The 'Yea' and 'Guilty' entries were replaced with 1s, while all other entries were replaced with 0s.

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

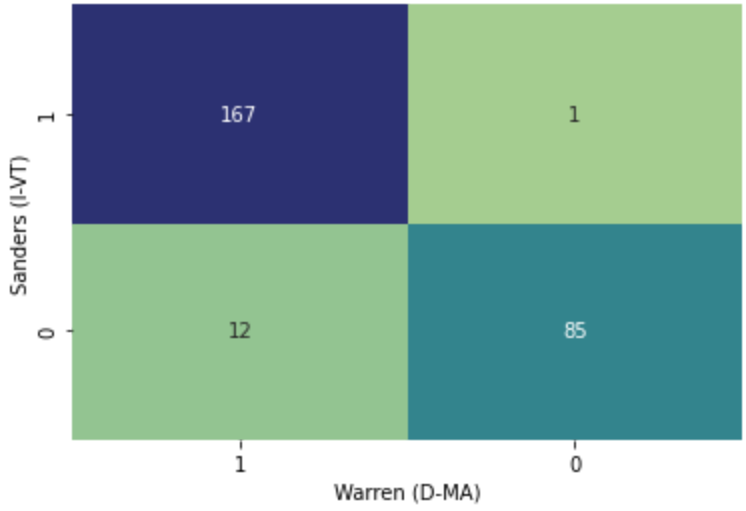
4. Exploratory Data Analysis (EDA)

a. EDA: 2x2 Senator Comparisons

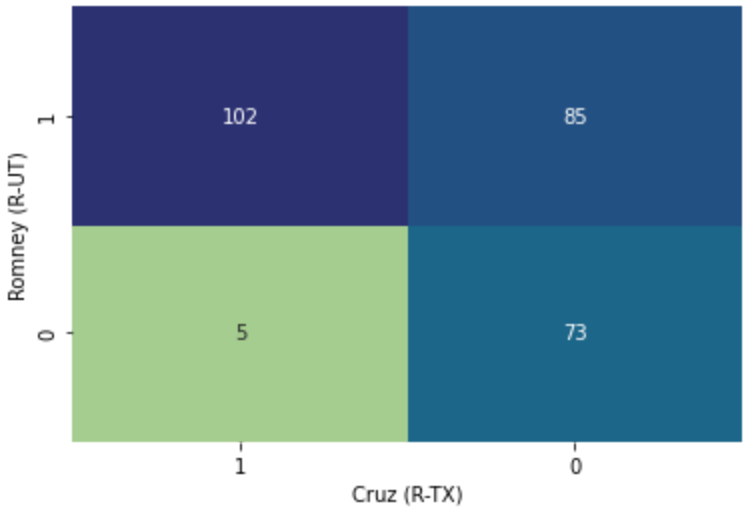
A comparison of the voting records of Bernie Sanders and Ted Cruz show that they voted the same way only 30 times out of 266 votes, indicating next to no alignment in voting patterns.



A similar comparison of Elizabeth Warren (a Democrat) and Bernie Sanders (an independent who caucuses with the Democrats) indicates agreement on 252 out of 266 votes, a high level of agreement.

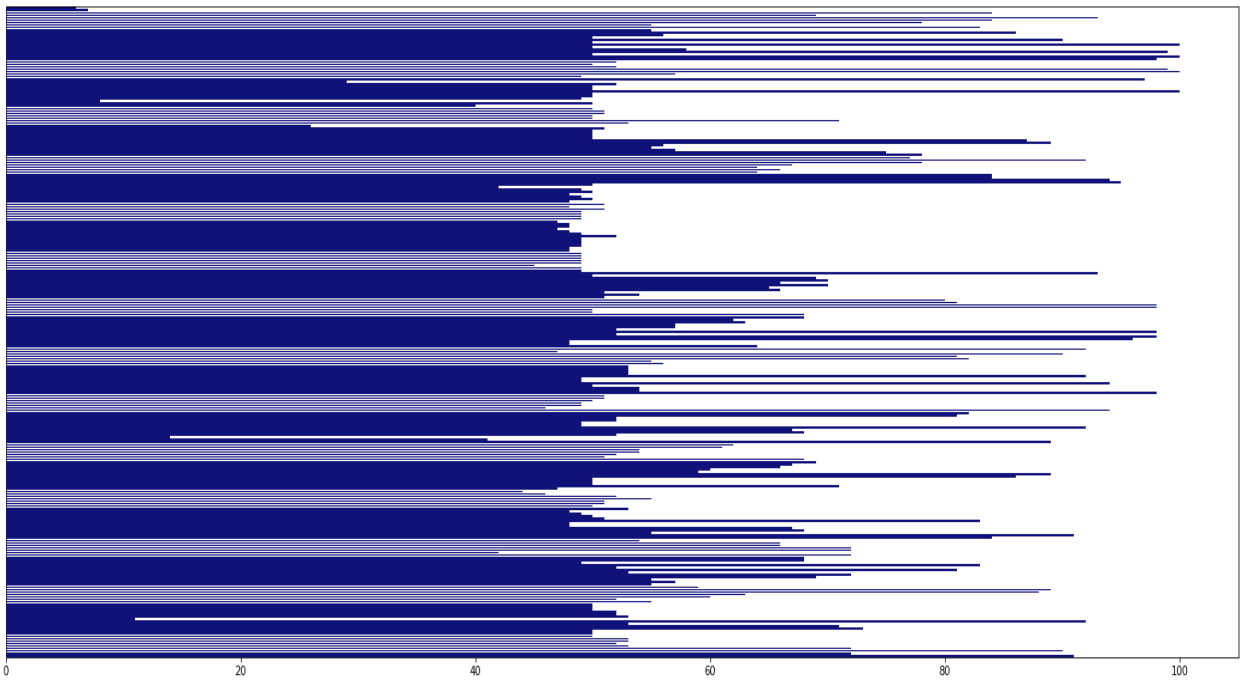


Finally, Mitt Romney and Ted Cruz, both Republicans, voted together 175 out of 266 times. Not as much agreement as between Sanders and Warren.



b. EDA: Yea's By Vote

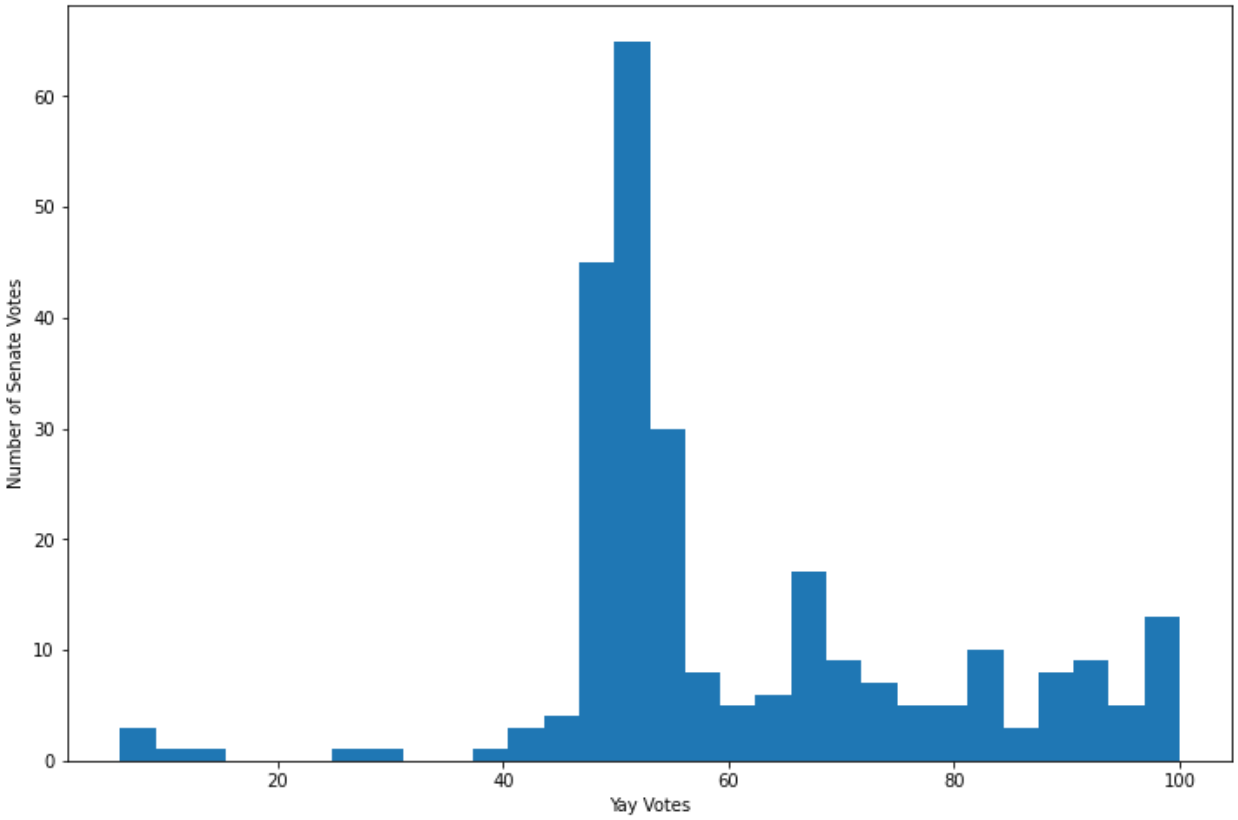
Each of the 266 Senate votes has been graphically shown below by the number of Yea votes received. It is clear that many votes are split down the middle, but that there are also many votes with either little support or near unanimous support. There is a lot of variability to be seen, which will support model creation.



c. EDA: Yea Counts by Votes

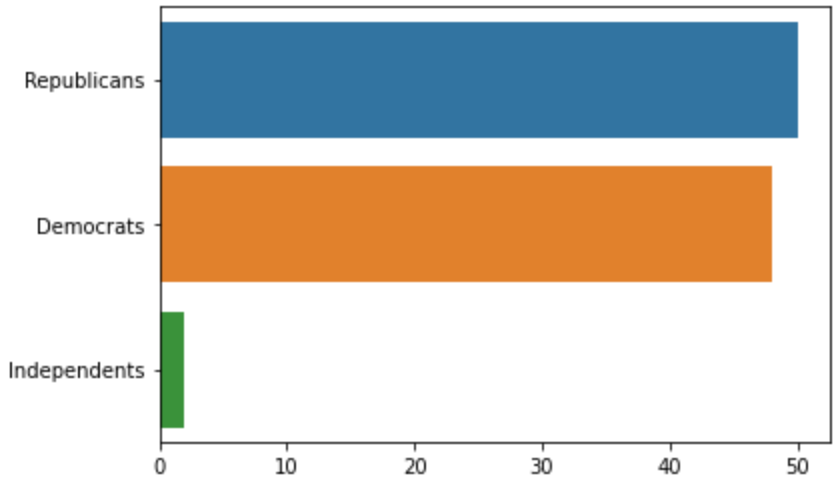
A histogram of voting patterns indicates that the highest frequency is for votes to have around 50 Yea's. This indicates votes supported by one or another party. At the same time, there are many votes that show clear bi-partisan support, some with only a few senators, some unanimous. This is indicated by the bars on the right side of the histogram.

At the same time, very few votes do not have party support, as indicated by the bars on the left of the histogram. These appear to be votes supported by only a few individual senators.



d. EDA: Senator Count by Party

The US Senate has 100 senators. In 2021, this is broken down by 50 Republican Senators, 48 Democrats, and 2 Independents, who caucus with the Democrats. In effect, this means a 50/50 split between the two parties in the Senate.

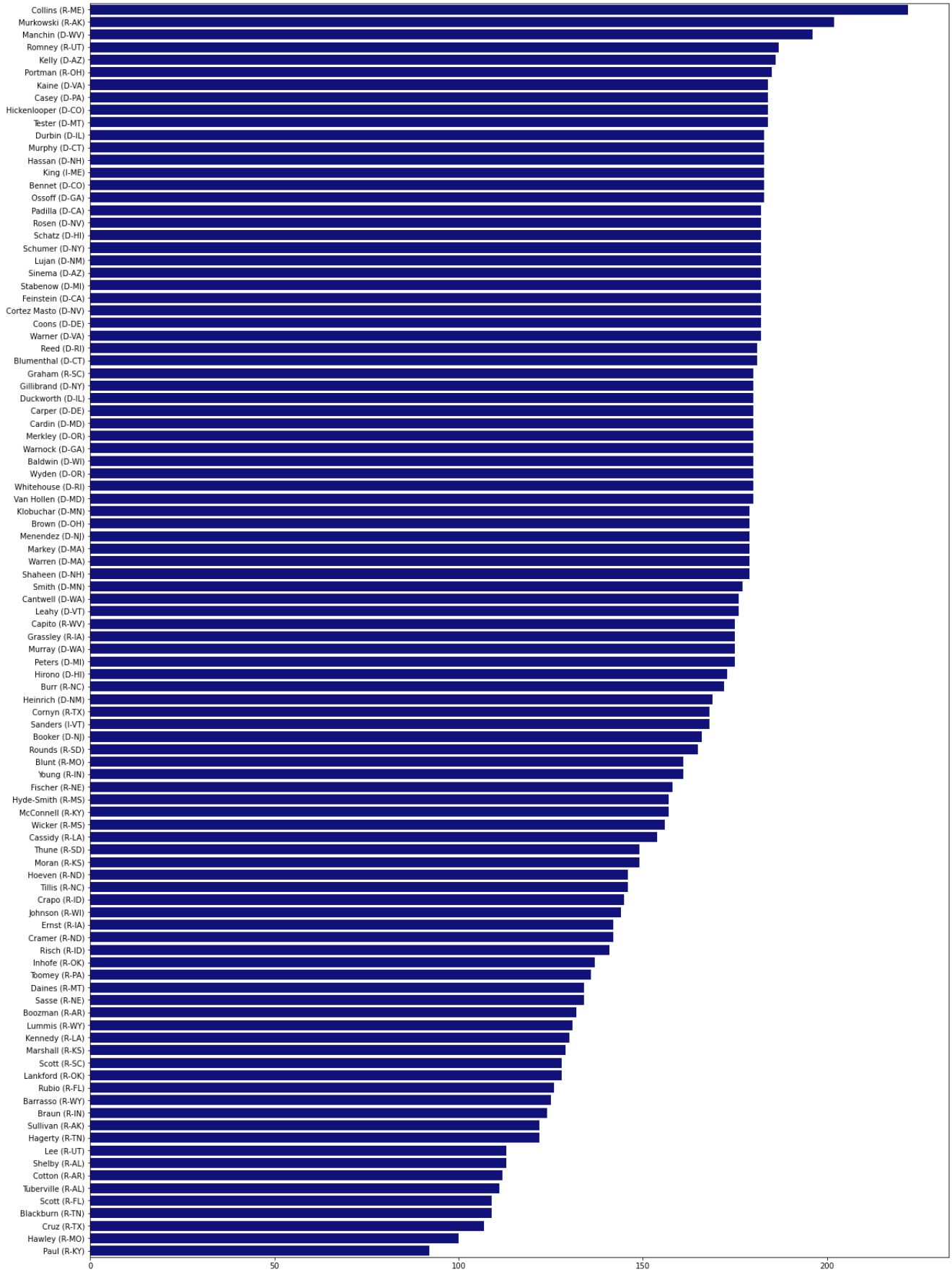


e. EDA: Yea Votes by Senator

The final piece of exploration is to understand the number of 'Yea' votes by Senator. Surprisingly, at the top is a mix of Republicans and Democrats. At the bottom appears to be mainly Republican Senators.

This piece of information already indicates that for the Republicans at least, there will be a divergence of opinion and opportunities for clustering.

US Senate Voting Clusters

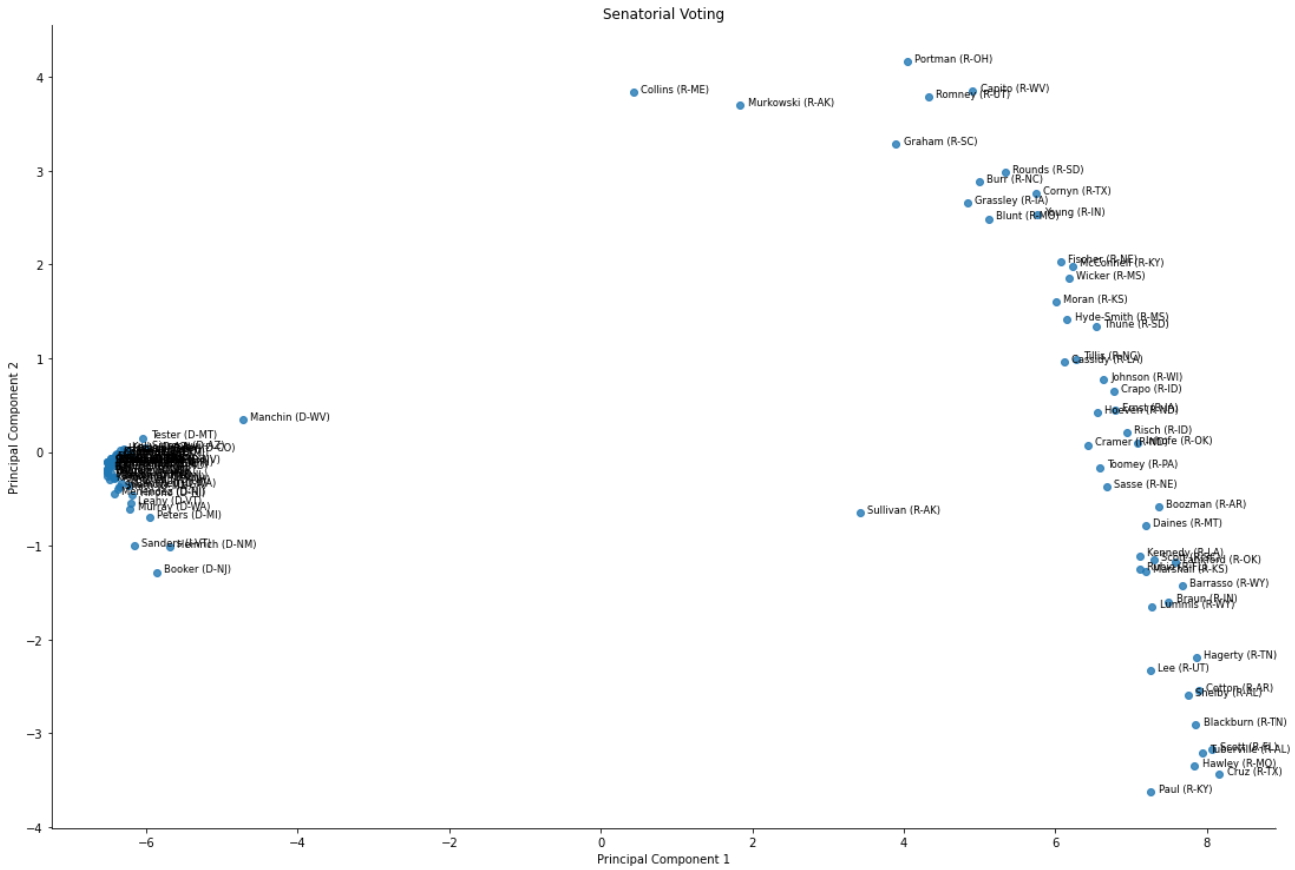


5. Modeling

The cleaned data will have 100 Records (Senators) and 266 Fields (Votes). The goal is to graphically show voting clusters in a two by two scatter plot. For this reason, dimensionality reduction is necessary, reducing 266 fields to 2. The tool for this will be Principal Component Analysis (PCA). Following the modelling, KMeans clustering will be applied to identify individual clusters.

a. Principal Component Analysis (PCA)

The PCA model indicated that there are two clearly defined groups within the Senate, with Democrats in the left hand, tight cluster, and the Republicans to the right.



US Senate Voting Clusters

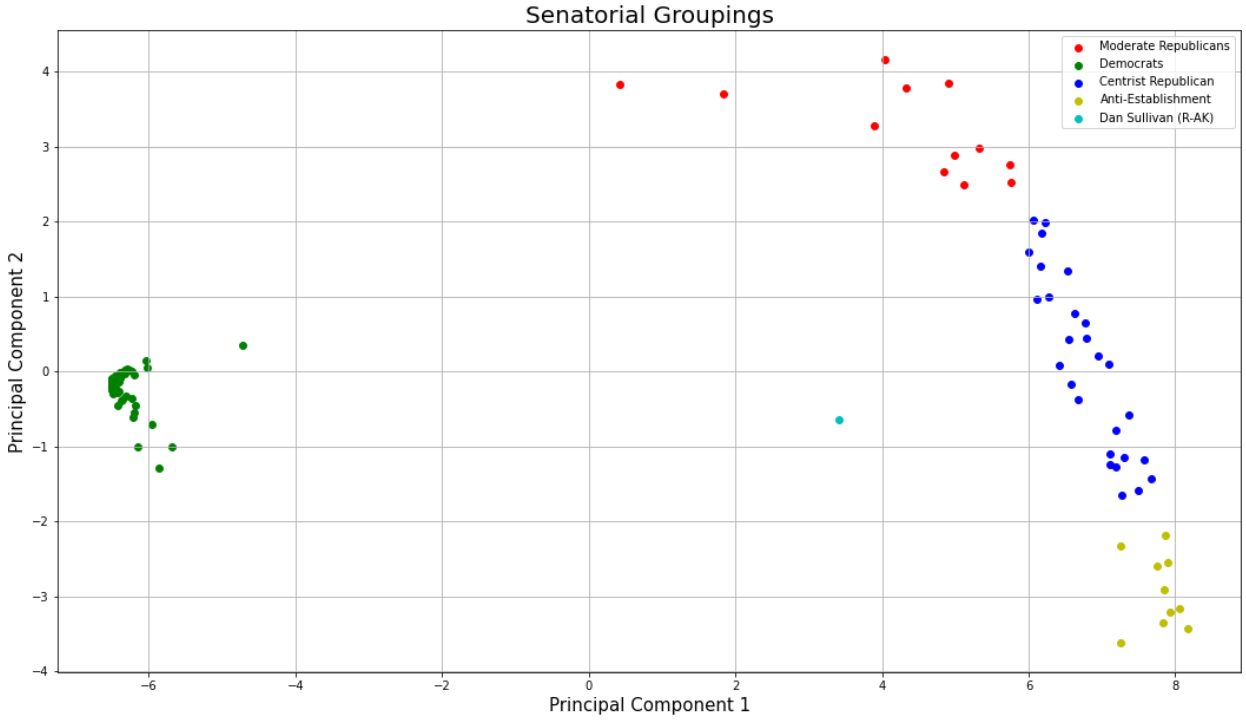
While the Democrats are a monolithic block, either through shared political visions or strong party discipline, the Republicans appear to be more of a continuum, with one outlier Senator in the middle. Also, while the Republicans show variance along the second Principal Component, that variance seems to be correlated with the first component. The Republicans at the top are those that vote closest to the Democrats.

In the 2x2 Senator comparison shown in the EDA section, Ted Cruz was identified as being quite different from both Bernie Sanders and Mitt Romney. Ted Cruz is located at the bottom right, Bernie Sanders is in the Democrat camp, and Mitt Romney is at the opposite end of the Republican party.

The two components have explained variances of 0.75 and 0.05, indicating a high confidence in the component decomposition.

b. KMeans Clustering

The KMeans clustering algorithm was applied and easily identified five clusters. A smaller number of clusters would aggregate disparate groups, while a larger number of clusters would unnecessarily split the Republican party into smaller groups.



The voting blocks were broken up into five clusters, one for the Democrats and four for the Republicans.

As mentioned before, the Democrats are a monolithic block. The outlier to the right (although still within the Democrat orbit) is Manchin (D-WV). West Virginia is a Republican State, and Manchin prides himself on bi-partisanship, which accounts for his position slightly outside of the cluster.

For the Republicans, three groupings can be identified. The 'Moderate Republicans' (to use Capital Hill terminology) include individuals such as Mitt

Romney. Fact checking indicated that these Senators are all unknown for their willingness to work with the Democrats.

At the other end is the aforementioned Ted Cruz, as well as Rand Paul. This grouping at the other end of the spectrum can be called the 'Anti-Establishment' Republicans. While they may not personally agree, or even like one other, they do vote in a similar manner and are well known for opposing or having a skeptical view towards government.

Between these two groups are what I call the 'Centrist Republicans', centrist in the sense that they are in the middle of the two Republican wings. The Republican Senate leader is a member of this group.

Dan Sullivan (R-AK) was given his own cluster as he is quite different from the Republicans along both Principal Components. While he may vote more 'Democrat' than most Republicans, a quick read through Wikipedia shows that he holds many views that are opposed by the Democrats.

6. Conclusion

The clustering model worked very well identifying voting groups within the US Senate in 2021. The most surprising result was the party discipline of the Democrats, and the range of viewpoints within the Republican party.

Online research of individual senators indicated that these clusters match political knowledge as shown in the press and by commentators. The model works.

Finally, a quick check of 'Moderate Republicans' brought up the article below, which described a meeting between the moderates and President Biden. The Senators named in the article are the exact same senators shown in the 'Moderate Republican' grouping.
https://thehill.com/homenews/senate/547063-moderate-gop-senators-and-biden-clash-at-start-of-infrastructure-debate

7. Next Steps

a. Perform the same analysis for the last 30 years of data

The US Senate has 30 years of data on its website. It would be interesting to understand how the groupings have changed over the years, and whether the groupings are influenced by the party of the President.

b. Analyze the US Congress and/or Canadian Parliament

The US Congress has over 300 members, and would make a very interesting study along the same lines to determine if there is more variability, especially among the Democrats.

Also, analysis of the Canadian parliament, with five federal parties, would also bring interesting results.

c. Build a website to compare individuals to senators

Since the model works, building a website where people can enter their own votes is feasible, and may show surprises as to where individuals' beliefs reside.