Identifying Voting Blocs for the House of Representatives in the 117th Session of Congress

Brandon Winder

12-9-2021

# Description of Data

This project looks at identifying voting blocs within the House of Representatives (HOR) in the 117th session of Congress. Both member ideology and members’ votes were collected through the website “<https://voteview.com/data>”. The member ideology dataset contained basic demographics for the members in the HOR such as their name, state, district, party, birth, and death year. The members’ votes dataset contained votes taken by every member. For each of these datasets, every member is assigned an ICPSR ID number. This ID number was used to combine the two datasets so that each members demographic data could be matched with their vote for every bill voted on in the 117th session of Congress.

Some data cleaning was necessary, primarly due to missing values in the ideology dataset. First, the votes that were cast by the President and Vice President were removed because the members in the House of Representatives are those of interest for this analysis. It was found that Ron Wright died in 2021 and he only voted on 30 out of the 371 bills in the dataset. So, each of Wright’s votes were removed from the dataset. Then, August Pfluger did not have a reported birth year. So, 371 missing values in the Born variable were replaced by 1978, the year Pfluger was born. Table 1 shows a subset of the data for the first bill voted on in the House of Representatives.

Table 1. Data on the first bill voted on for three members in the HOR.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | icpsr | state\_abbrev | party\_code | bioname | born | cast\_code |
| 23 | 20301 | AL | R | ROGERS, Mike Dennis | 1958 | 6 |
| 28633 | 21948 | CO | D | NEGUSE, Joseph | 1984 | 1 |
| 91573 | 21942 | NJ | D | MALINOWSKI, Tomaz | 1965 | 1 |

# Methods

To determine voting blocs within the House of Representatives, a multidimensional scaling (MDS) technique was used. MDS was used because it is a nice way to visually represent the distances or dissimilarities between individual cases. For this project, the dissimilarities of interest are the difference in votes between each of the members in the HOR. But, the data collected did not contain the number of times each member voted differently than the other members. So, a dissimilarity matrix needed to be created in order to utilize MDS. To do this, each member needed to be compared to every other member in the House of Representatives for each vote cast; then the results were entered into a 437 by 437 matrix to create the dissimilarity matrix used for MDS. The first three members in the matrix can be seen in Table 2. This table shows that Rogers, a Republican Representative from Alabama, voted differently 249 and 74 times from two other Representatives from Alabama, Sewell (Dem.) and Brooks (Rep.), respectively. It also shows that Sewell and Brooks voted differently 292 times.

Table 2. The number of times members voted differently for the first three members.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ROGERS\_AL\_R | SEWELL\_AL\_D | BROOKS\_AL\_R |
| ROGERS\_AL\_R | 0 | 249 | 74 |
| SEWELL\_AL\_D | 249 | 0 | 292 |
| BROOKS\_AL\_R | 74 | 292 | 0 |

The dissimilarity matrix was then used in Kruskal’s non-metric multidimensional scaling with two dimensions. When using two dimensions, the stress reported was 5.51%, which represents a good fit. When increasing the dimensions to three, the stress only decreased to 5.45%. So, the two dimensional MDS was used for the next step in the analysis. The next step taken in this analysis was using K-means clustering to identify voting blocs within the HOR. To do this, the fitted configuration for the two dimensions from the MDS were used within a K-means clustering algorithm two times to generate two different sets of clusters. For the first set, two clusters were created to identify the two main political parties in the HOR. Then, the second time the K-means clustering was used with three clusters to try and identify members that are not voting similarly to their affiliated party.

# Results

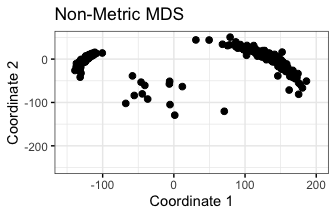


Figure 1. Basic scatter plot of the non-metric multidimensional scaling.

The results from the basic two-dimensional non-metric MDS are plotted in a scatter plot where the x-axis is the first dimensions fitted configuration and the y-axis is the second dimensions fitted configurations (Figure 1). This graph visually represents the similarities and difference for all 437 members in the HOR for the 117th session of Congress. The points in this graph were then modified to represent demographic attributes about the members to obtain a better understanding of the clusters. The points were also color coded and convex ellipses were drawn to visualize the two different clustering sets that were completed (Figure 2). In plot (A) from Figure 2, the red cluster was categorized as “GOP” because all of the members in the cluster are Republicans and the blue cluster is “Dem.” because the majority of the members are Democrats. The K-means clustering algorithm had a classification rate of 99.1% because only 4 Republicans were mis-classified as Democrats (Figure 3).

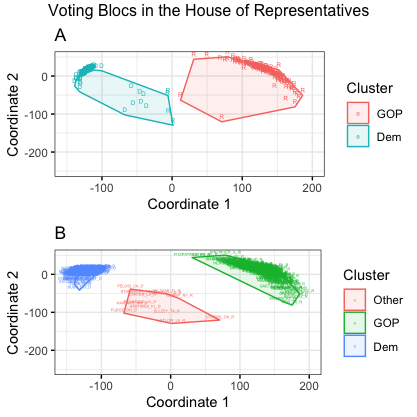


Figure 2. Scatter plots showing the clustering created by the K-means clustering. Plot A shows the clusters when using two centers for K-means. Plot B shows the clusters when using three centers for K-means.

In plot (B) from Figure 2, the green cluster was categorized as “GOP” because all the members in that cluster are Republicans, the blue cluster is “Dem.” because all the members are Democrats, and the red cluster is “Other” because there is a similar number of members from each party.

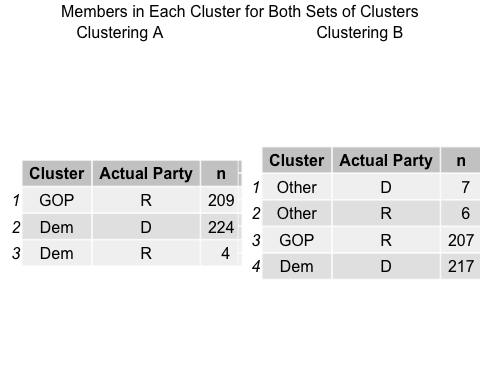


Figure 3. Displays the number of members in each cluster for both sets and compare that to the actual party for the members.

There is seven Democrats and six Republicans in the “Other” cluster which shows that these members are often not voting the same as the rest of their party (Figure 3). Each member that is included in the “Other” cluster can be seen in Table 3. This table shows the state the members represent and the party they are affiliated with. There are 53, 27, 27, 16, 37 total Representatives from CA, FL, NY, OH, and TX, respectively. So, the proportion of Representatives for each of these five States is not very large. But for the other two States in this cluster, the proportion of representatives in the “Other” cluster is much larger (43% for LA and 50% for NM). The only two Democratic Representatives from LA are clustered in the “Other” category. This shows that the Democratic Representatives from LA do not vote similarly to the other Democratic Representatives.

*Table 3. Members included in the “Other” cluster from plot (B) in Figure 2.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CA | FL | LA | NM | NY | OH | TX |
| Pelosi(D) | Salazar(R) | Letlow(R) | Haaland(D) | Tenney(R) | Fudge(D) | Ellzey(R) |
| Valadao(R) | Hastings(D) | Richmond(D) | Stansbury(D) |  | Stivers(R) |  |
|  |  | Carter(D) |  |  |  |  |

# Conclusion

To summarize the analysis done on the votes cast by members in the US House of Representatives of the 117th session of Congress, Kruskal’s non-metric multidimensional scaling was used within K-means clustering algorithms to gain an understanding of voting blocs. The results of this analysis do show what would be expected. There are two large voting blocs that consist of Representatives of the same party and there are a few members that have differing voting patterns than the rest of their party (Figure 2. B). Additionally, when only identifying two clusters there is a classification rate of 99.1% (Figure 3.), showing that generally Representatives vote similarly to others within their political party. Finally, we can identify 13 representatives from seven different States that have differing voting patterns than their party typically has (Table 3). Of these 13 representatives, three of them (1 Rep. and 2 Dem.) are from Louisiana (LA) and LA only has 7 total members in the HOR (5 Rep. and 2 Dem.). Both of the Democratic Representatives from LA do not have similar voting patterns to the rest of the Democrats in the HOR. This is an interesting finding and may be something to look into in further studies to determine if Democrats from Louisiana have differing beliefs than Democrats from other states.