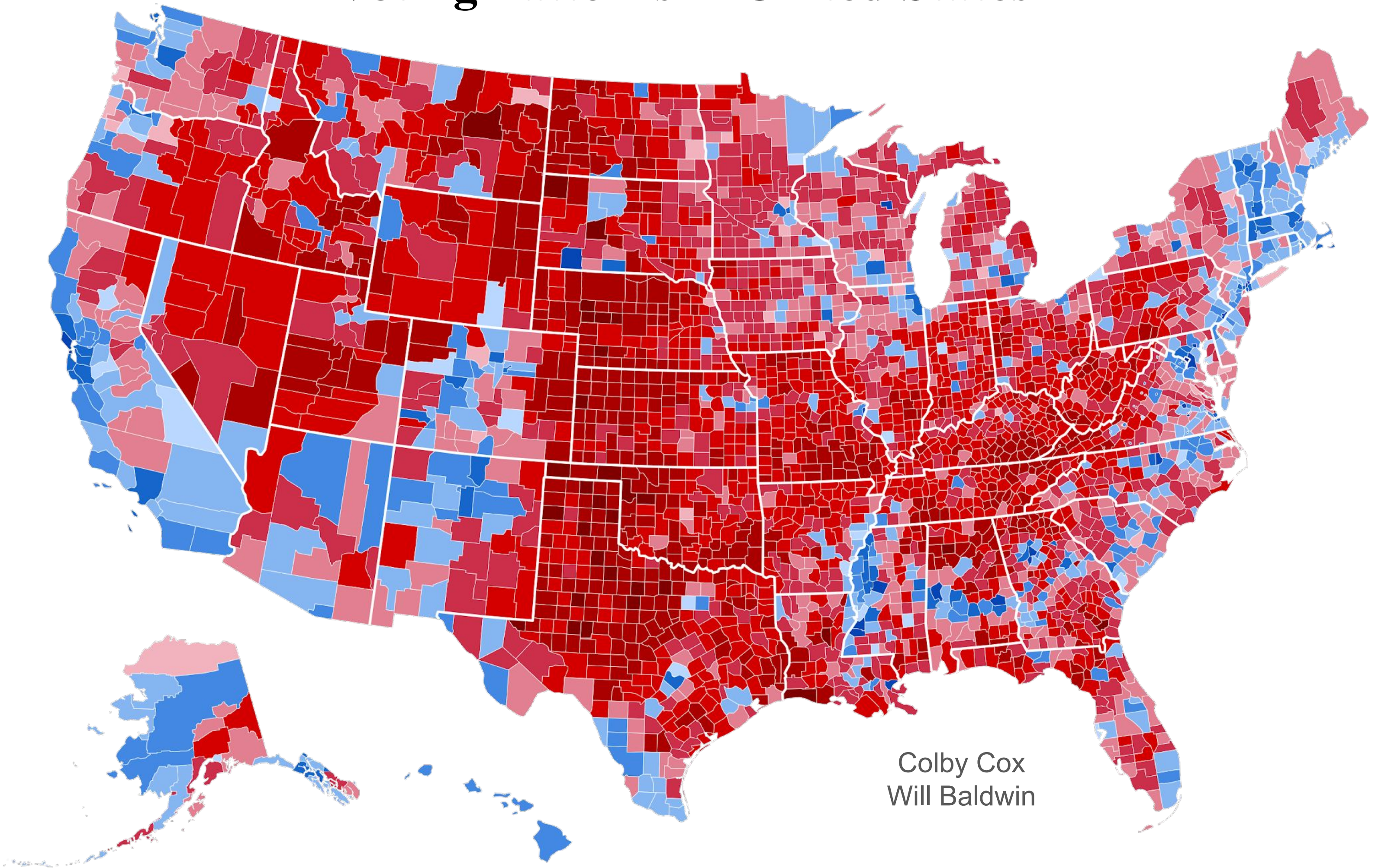
Voting Patterns in United States



Colby Cox
Will Baldwin

Abstract

- Analyze how United States citizens vote with the help of multiple statistical analysis methods.
- Various datasets were collected to have a better understanding of the factors leading to how Americans cast their vote.
- Find subtle influences that cause United States citizens to make their decision when they cast their vote.

Introduction

- Investigate the influence of various factors on voting behavior.
- Our data consists of state, age, income, percentage of male/female, and percentage of race/ethnicity in each state
- Our methods of testing were data clustering and multidimensional scaling.
- The data clustering process involved clustering the factors and comparing them to see if there was a match for the Democratic and Republican voting groups.
- The multidimensional scaling process included running the MDS functions and examining how data grouped compared to voting history.
- The overall goal is to determine if these factors truly have significant impact on voting or if they are based on assumptions.

Data

	Vote <chr>	Age <dbl>	HouseholdIncome <dbl>	Male <dbl>
Alabama	Republican	39.5	51734	0.483
Alaska	Republican	35.3	75463	0.517
Arizona	Democratic	38.5	62055	0.495
Arkansas	Republican	38.6	48952	0.491
California	Democratic	37.3	80440	0.498
Colorado	Democratic	37.3	77127	0.503
Delaware	Democratic	41.4	70176	0.482
Florida	Republican	42.7	59227	0.489
Georgia	Democratic	37.3	61980	0.482
Hawaii	Democratic	40.0	83102	0.488

1-10 of 49 rows | 1-5 of 12

Female <dbl>	WhiteTotalPerc <dbl>	BlackTotalPerc <dbl>	IndianTotalPerc <dbl>
0.517	0.674986	0.265945	0.005061
0.483	0.633606	0.032421	0.145590
0.505	0.737727	0.045317	0.043353
0.509	0.753670	0.152012	0.006357
0.502	0.560507	0.057209	0.007920
0.497	0.815192	0.041459	0.009441
0.518	0.674384	0.219902	0.003679
0.511	0.716390	0.159357	0.002623
0.518	0.572455	0.315677	0.003325
0.512	0.241482	0.018814	0.002470

Previous 1 2 3 4 5 Next

AsianTotalPerc <dbl>	HawaiianTotalPerc <dbl>	OtherTotalPerc <dbl>	TwoOrMoreTotalPerc <dbl>
0.013878	0.000417	0.039712	0.024385
0.064165	0.014227	0.109992	0.093396
0.033341	0.002040	0.138223	0.069639
0.015316	0.003471	0.069174	0.039836
0.148282	0.003803	0.222278	0.079348
0.031968	0.001563	0.100378	0.059320
0.039815	0.000729	0.061491	0.038822
0.027839	0.000629	0.093162	0.060189
0.041326	0.000678	0.066540	0.037386
0.376402	0.104007	0.256824	0.242887

1-10 of 49 rows | 10-12 of 12 columns Previous 1 2 3 4 5 Next 2 3 4 5 Next

Analysis

- K-means clustering was applied to the numeric data with the number of clusters set to two
- Data could be categorized into different groups based on their similarities
- A silhouette plot was also created to help visualize the form of the clustering result
- Variation of information and corrected rand index were both used to help compute the quality of the clustering results

```r
```{r}
data_num <- data %>% select_if(is.numeric)
data_dist <- dist(data_num)
data_eclust <- eclust(data_num, "kmeans", 2)
data_silhouette <- fviz_silhouette(data_eclust)
data_clus <- data %>%
  mutate(cluster = data_eclust$cluster,
         vote_numeric = (data$Vote) %>% as_factor() %>% as.numeric())
data_stats <- cluster.stats(data_dist, data_clus$vote_numeric, data_clus$cluster)
data_vi <- data_stats$vi
data_rand <- data_stats$corrected.rand
#table(data$State, data_eclust$cluster)
```
```

Analysis

- Second method used was a non-metric multidimensional scaling
- The stress plot function was used to create a plot to help assess the solution
- NMDS results were transformed into a dataframe with rows representing the states and columns representing MDS1 and MDS2
- 2D scatter plot created, each point represents a state
- Used as visualization tool to help better understand the data

````
data_MDS <- metaMDS(data[,2:11], distance = "gower")

stressplot(data_MDS)

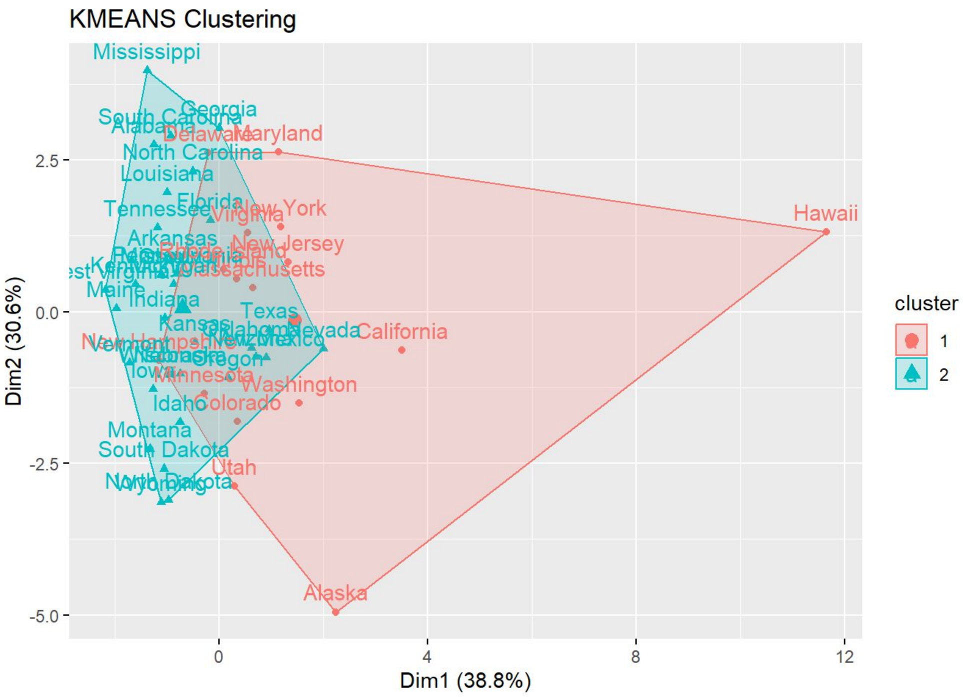
data_MDS$stress

data_2D_df <- data_MDS$points %>%
  as_tibble(rownames = "States")
colorVector <- c("blue", "red")
stateColorVector <- c()
for (i in data$Vote)  {
  if (i == "Democratic") {
    stateColorVector <- rbind(stateColorVector, "blue")
  }
  else {
    stateColorVector <- rbind(stateColorVector, "red")
  }
}
data_2D_df %>%
  ggplot(mapping = aes(x = MDS1, y = MDS2, color = stateColorVector)) +
  scale_colour_manual(values=colorVector) +
  geom_text(label = rownames(data), size = 2) +
  labs(x = "Dimension 1",
       y = "Dimension 2",
       title = "NMDS 2-D representation of Voting Dataset")
```
````
````

Analysis

- The K-means clustering was used because it can identify natural groupings in the data.
- It was a great way to understand if there were any obvious patterns or outliers.
- By clustering the data, K-means can expose the similarities and differences of how different demographics vote.
- Clustering the data helped interpret and analyze the data easier because of how it reduces the dimensions.
- Non-metric multidimensional scaling was also another great way to analyze this data.
- MDS was used to help uncover any underlying patterns within the data that might not have been seen or was overlooked.
- It was also used to provide a visual representation of any similarities or differences of voting patterns, but its primary purpose is to reduce the dimensions to reveal any obvious interpretations within the data.
- They are both statistical methods that are ingrained in the statistical analysis world and are widely used when searching for relationships in data.

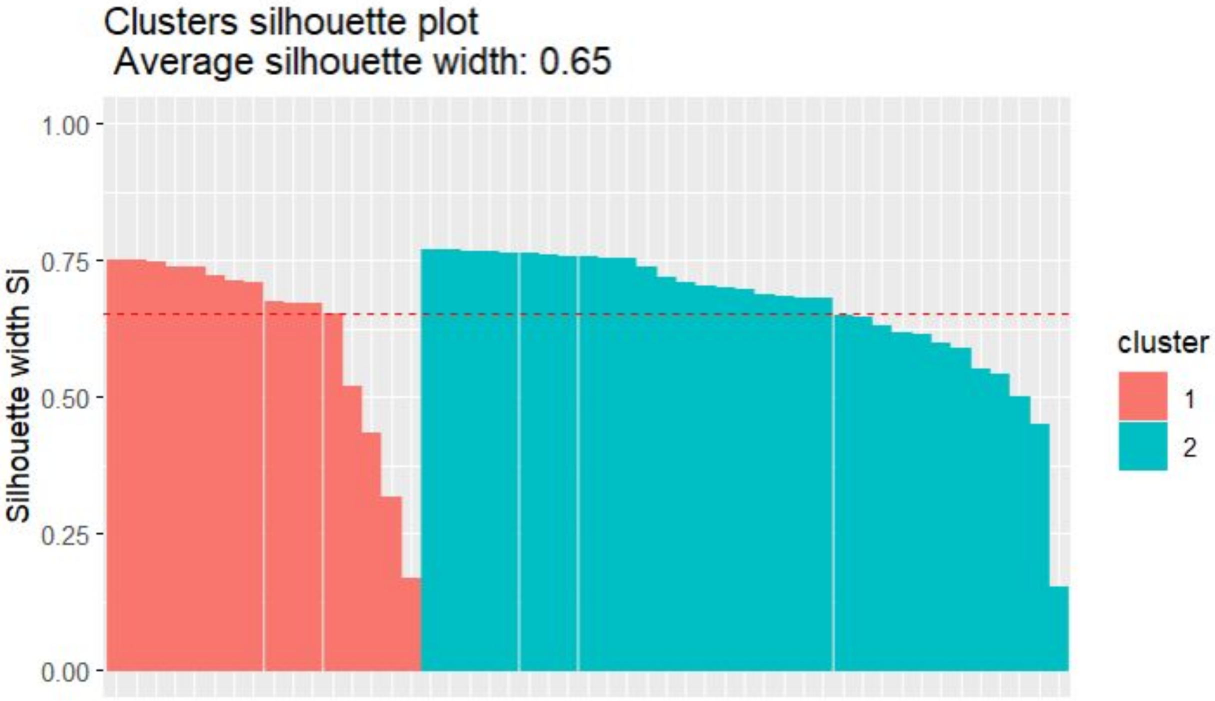
Results



	cluster <fctr>	size <int>	ave.sil.width <dbl>
1	1	16	0.62
2	2	33	0.66

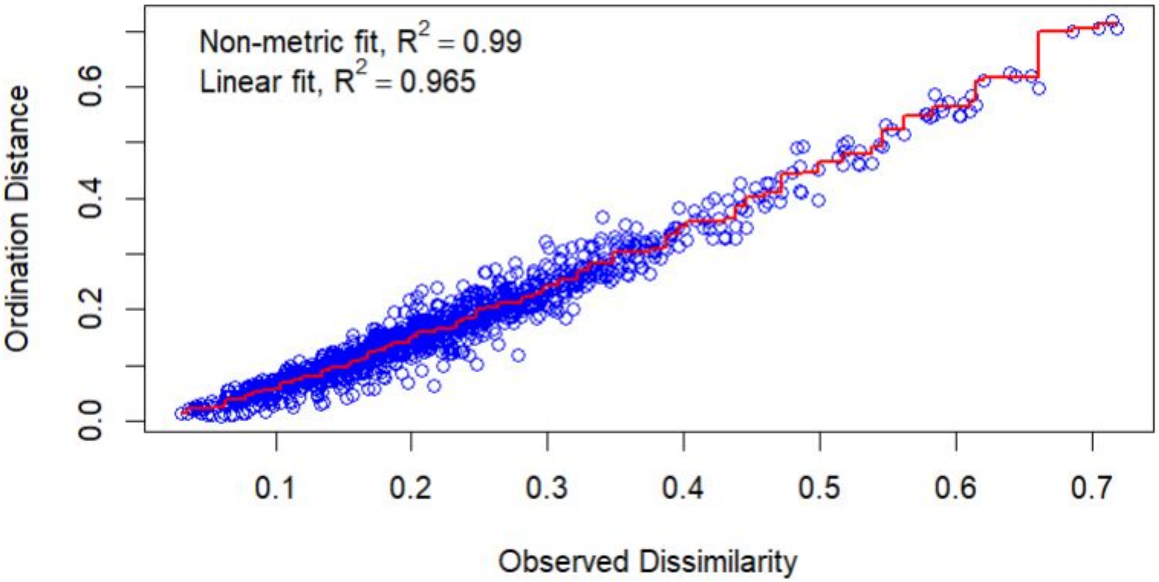
2 rows

Results



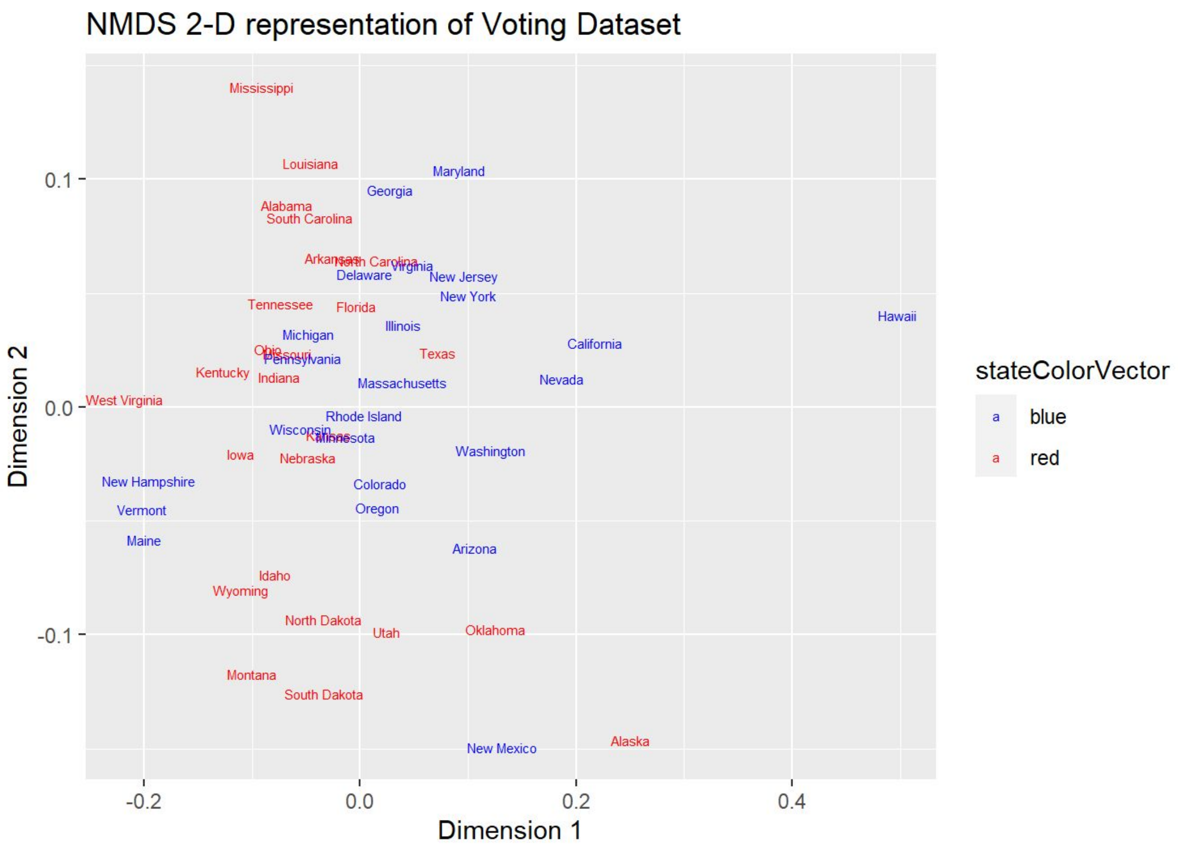
Clusters silhouette plot
Average silhouette width: 0.65
Silhouette width Si
1.00
0.75
0.50
0.25
0.00
cluster
1
2

Results



Non-metric fit, R2 = 0.99
Linear fit, R2 = 0.965
Ordination Distance
0.0
0.2
0.4
0.6
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Observed Dissimilarity

Results



NMDS 2-D representation of Voting Dataset
Mississippi
Louisiana
Maryland
Georgia
Alabama
South Carolina
Arkansas
North Carolina
Virginia
Delaware
New Jersey
New York
Tennessee
Florida
Hawaii
Illinois
Michigan
California
Ohio
Missouri
Pennsylvania
Texas
Kentucky
Indiana
Massachusetts
Nevada
West Virginia
Rhode Island
Wisconsin
Kansas
Minnesota
Washington
Iowa
Nebraska
New Hampshire
Colorado
Vermont
Oregon
Maine
Arizona
Idaho
Wyoming
North Dakota
Utah
Oklahoma
Montana
South Dakota
New Mexico
Alaska
0.1
0.0
-0.1
-0.2
0.0
0.2
0.4
Dimension 2
Dimension 1
stateColorVector
a blue
a red

Discussion

- Generally the two methods show positive results, but there are some outliers.
- Texas, which is completely surrounded by democratic states on MDS, could be republican because of targeted political campaigning in the area.
- MDS splits pretty well, but clustering is a lot less clear.
- Checking the rand index shows a slightly positive result
- Clustering may be a less proficient way to prove this, as it is random and the factors at play here affect the voting very slightly, but enough to matter to a politician.

Further Extension

- Add more factors said to affect voting, such as living in a rural or urban area or various mental factors like statewide depression.
- Testing under factor analysis to see how factors affect each other, like how if one population is higher there will be less of the other populations.
- Looking into the outliers like Texas and New Mexico.

Limitations

- Electoral votes are not a good measure of population or political makeup of a state.
- Data gathered is just a few of the factors that affect voting.

Conclusion

- The factors age, income, sex, and ethnicity do affect voting.
- This is important for politicians campaigning, as they need to guide their campaign towards getting votes.
- Limitations
    - Data gathered - other factors could be used said to direct voting
    - Observations limited by state, take a closer look at counties
    - Electoral votes are not the best measure for population or political makeup of a state

References

- “Arkansas 2020 Election Results.” *CNN*, Cable News Network, www.cnn.com/election/2020/results/state/arkansas. Accessed 6 Dec. 2023.
- “List of U.S. States and Territories by Median Age.” *Wikipedia*, Wikimedia Foundation, 27 Sept. 2023, en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_median_age.
- Team, Good Party Politics. “The Psychology of Voting: What Motivates the Electorate.” *Good Party*, 24 Oct. 2023, goodparty.org/blog/article/psychology-of-voting-what-motivates-electorate.
- *U.S. Census Bureau Quickfacts: Arkansas*, www.census.gov/quickfacts/fact/table/AR/PST045222. Accessed 7 Dec. 2023.
- *US States by Race 2023*, worldpopulationreview.com/states/states-by-race. Accessed 6 Dec. 2023.