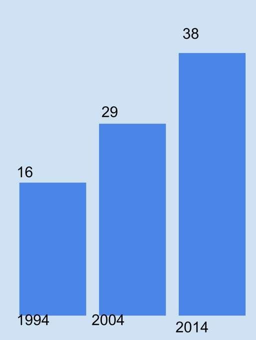
The Effect of Race, Gender, State, and Party on Partisanship in U.S. Politics

Nandita Gupta[1], Sophie Alexis[1], and John Hanay[2]
[1]Henry M. Gunn High School, [2]Ruckus Wireless

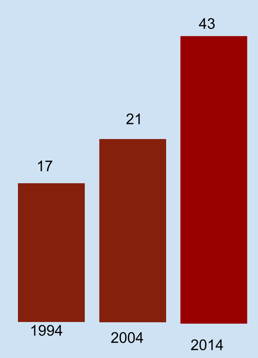
INTRODUCTION

The media presents a picture of an America that has grown increasingly polarized; both catalyzed by and reflected by the speeches today's politicians make. We wanted to see whether our nation is as divided as it appears on our computer screens. Matthew Gentzkow's paper on behavioral economics described a computer program which determines the party of the speaker. We realized that we could measure polarization by the accuracy of the program; the higher the accuracy, the greater the polarization.

DATA AND FINDINGS

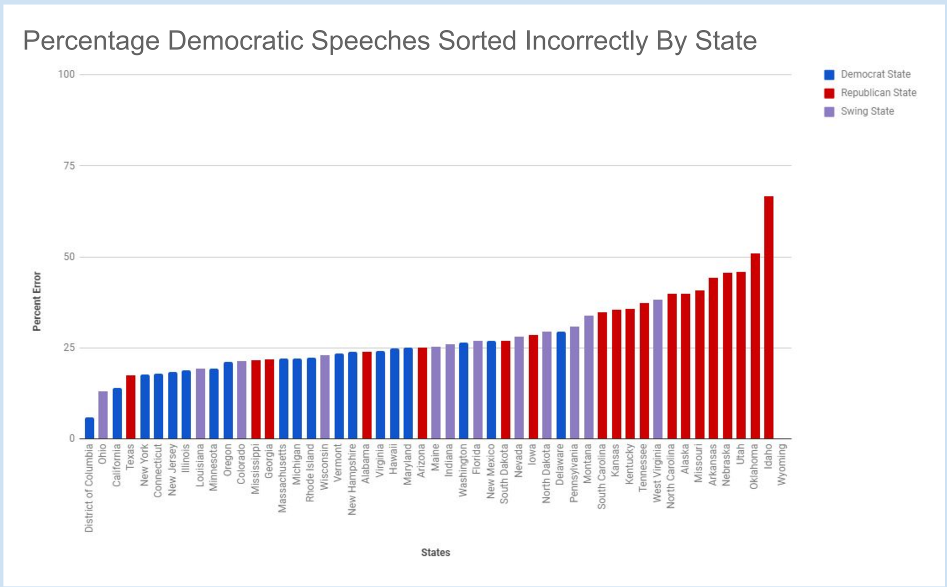


% Democrats who viewed the GOP as "very unfavorable"



% Republicans who viewed the Democratic Party as "very unfavorable"

Source: Suh, M. (2014, June 11). Political Polarization in the American Public. Retrieved October 30, 2017, from http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/



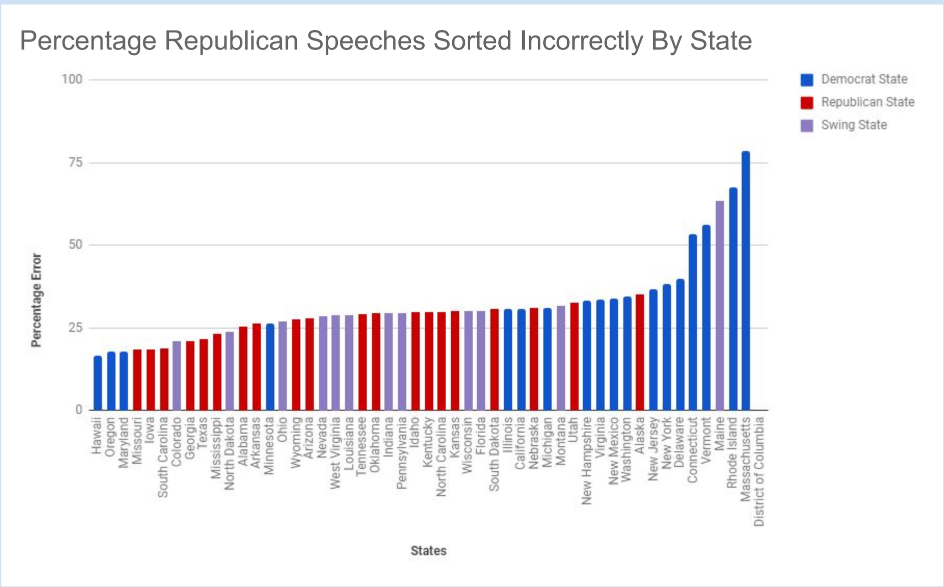


Figure 1 and 2: These graphs represent the percentage of speeches in each state that were sorted incorrectly by the classifier. The bars are colored to represent blue, red, and purple states based on senator affiliations.

All Data	Dem. Actual	Rep. Actual
Dem. Predicted	63,803	19,529
Rep. Predicted	17,836	49,847

1990's	Dem. Actual	Rep. Actual
Dem. Predicted	9,663	3,115
Rep. Predicted	2,442	8,943

2000's	Dem. Actual	Rep. Actual
Dem. Predicted	30,857	10,322
Rep. Predicted	7,604	20,763

2010's	Dem. Actual	Rep. Actual
Dem. Predicted	25,936	6,932
Rep. Predicted	5,923	19,141

Tables 1-4: These tables are commonly called "confusion matrices," referring to when the classifier is "confused." The data show how many speeches in each category the classifier sorted correctly and incorrectly.

Top 50 words the classifier used on speeches:

accordingly
routine
solute
defend
aisle
mobile
morristown
loudoun
commonwealth
descript
entitle
liability
syracuse
occur
rodger
hoosier
pensacola
takeover
riverside
enemy
twenty-second
therefore
thus
wilkin
government-run
radical
congratulate
encourage
bless
proudly
denton
govern
immediate
unborn
kentuckian
freedom
speaker
bureaucracy
spend
cincinnati
illegal
liberal
earmark
revise
obama-care
bureaucrat
brownwait
diaz-balart
ros-lehtinen
recognize

CONCLUSIONS, NEXT STEPS

Polarization through Time:
We expected that the speeches from the current decade would indicate greater polarization than the speeches of the previous decades. We were unable to draw any such conclusions based on our decades classifiers.

Regional Biases:
Our hypothesis on the effect of regional biases on speech was proven correct. In more Republican-leaning states, the Democratic politicians' speeches were classified more as Republican. Democrats from redder states tend to behave more like Republicans in response to their constituency, and vice versa. Region was an accurate metric of a politician's views.

Future Work:
We would like to build a regional classifier to see if the region to which a politician belongs is easier to predict than the party.

RESEARCH METHODOLOGIES

Databases:

- A collection of approximately 570,000 speeches by different congressmen, downloaded from the Congressional Record
  - Used this as our training and test data for the classifier
- A database of speakers and their gender and race
  - Was used for reading into the program so that we could get polarization data based off of race and gender

Vocabulary:

- TFIDF Matrix:
  - Term Frequency Inverse Document Frequency
  - Takes the percentage frequency of each term in each document and puts it as the inverse of the percentage frequency in all the documents
    - Eliminates words like "the" and "congress" which are used frequently across the aisle from becoming determining factors
- Logistic Regression Model: Plots the TFIDF along a logistic regression curve and classifies the speeches.

DISCUSSION, ANALYSIS, AND EVALUATION

Time Period Analysis:
Based on the picture the media has given us, we expected there to be more polarization as we approach the present. However, when fine-tuning the classifier and removing state names and Congressmen's names, which would skew the data, there appeared to be no real increase in Congressional polarization. Our data turned out to be inconclusive.

Geographical Analysis:
We did, however, find that the classifier behaves as expected in terms of regional biases. In more Republican-leaning states, the Democratic politicians' speeches were classified more as Republican. We can infer that this is because Democrats from redder states tend to behave more like Republicans in response to their constituency. The same occurred with Republicans in bluer states, where they were classified incorrectly as more Democratic.

ACKNOWLEDGEMENTS / REFERENCES

We'd like to express our deepest gratitude to Vineet Gupta for help in data collection, Mudita Jain for programming consultation, and our mentor John Hanay for overall guidance.

Works Cited:

Cho, W. K., Gimpel, J. G., & Hui, I. S. (2013). Voter Migration and the Geographic Sorting of the American Electorate. Annals of the Association of American Geographers, 103(4), 856-870. doi:10.1080/00045608.2012.720229
Congressional Record. (n.d.). Retrieved December 10, 2017, from https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=CREC
Gentzkow, M., Shapiro, J. M., & Taddy, M. (2016). *Measuring polarization in high-dimensional data: Method and application to congressional speech* (No. w22423). National Bureau of Economic Research.
Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences*, *37*(3), 297-307.
Jensen, J., Naidu, S., Kaplan, E., Wilse-Samson, L., Gergen, D., Zuckerman, M., & Spirling, A. (2012). Political polarization and the dynamics of political language: Evidence from 130 years of partisan speech [with comments and discussion]. *Brookings Papers on Economic Activity*, 1-81.
Lauderdale, B. E., & Herzog, A. (2016). Measuring political positions from legislative speech. *Political Analysis*, *24*(3), 374-394.
Mccarty, N. (n.d.). Reducing Polarization by Making Parties Stronger. Solutions to Political Polarization in America, 136-145. doi:10.1017/cbo9781316091906.009
Members of the U.S. Congress. (n.d.). Retrieved Nov. & dec., 2017, from https://www.congress.gov/members
Nivola, P.S. (2005, January 1). Thinking About Political Polarization. Retrieved from https://www.brookings.edu/research/thinking-about-political-polarization/
Suh, M. (2014, June 11). Political Polarization in the American Public. Retrieved October 30, 2017, from http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/
Lee, Michelle Ye Hee. "Analysis | Everything You Wanted to Know about Bernie Sanders's Record on Guns." *The Washington Post*, WP Company, 26 Jan. 2016, www.washingtonpost.com/news/fact-checker/wp/2016/01/26/everything-you-wanted-to-know-about-berniesanderss-record-on-guns/?utm_term=.786789c37bc0.
File:114th United States Congress Senators.svg. (n.d.). Retrieved April 10, 2018, from https://commons.wikimedia.org/wiki/File:114th_United_States_Congress_Senators.svg#filelinks