The Politics of Negativity: Sentiment, Bipartisanship, and Text Mining in the California Legislature

Group 4: Riley Berman, Min Kim (Authors), Sasha Mok (Secondary), Toby Liu (RA)

Credits: (Presentation) Riley Berman

Overview

Bipartisanship

Data: LegiScan

- California bill data from public tracking service LegiScan.
- Interacted with its API through its documentation.
- Many different type of bills. Restrict to Senate and Assembly bills (~90%) before 2025.

Data: LegiScan

- California bill data from public tracking service LegiScan.
- Interacted with its API through its documentation.
- Many different type of bills. Restrict to Senate and Assembly bills (~90%) before 2025.
  - ~35K bills.
  - 2009-2010 to 2025-2026 Regular Sessions.
  - 2009-2010 Regular Session = December 2008-November 2010.
  - Exclude Chad Mayes: the only Independent (2019-Present) in our dataset.

The legend himself

What is Bipartisanship?

- Lugar Center-McCourt School Bipartisan Index: bit complicated.
- We devised three metrics:

Bipartisanship

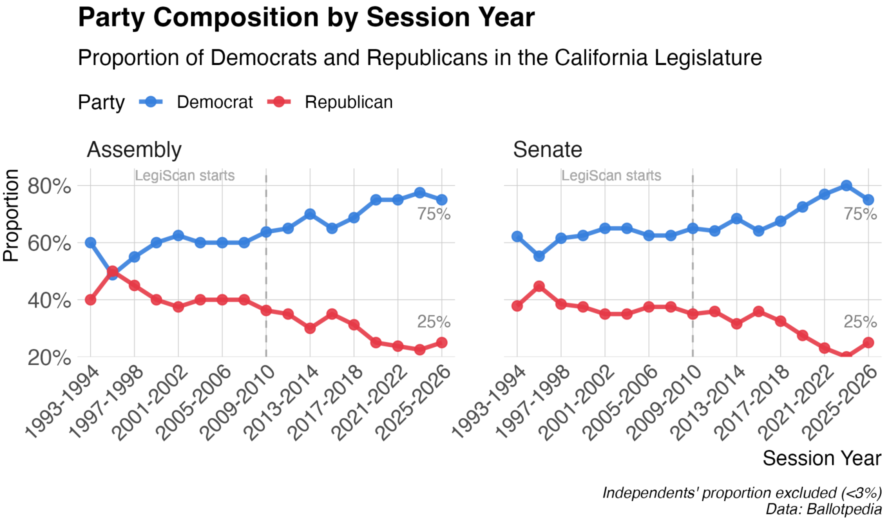
A bill is *bipartisan1* if at least one member of each party is a sponsor.

A bill is *bipartisan2* if the minority party makes up at least 25% of its sponsors.

A bill has a *bipartisan3* score given by $\frac{\text{minority sponsors}}{\text{majority sponsors}} \in [0, 1]$.

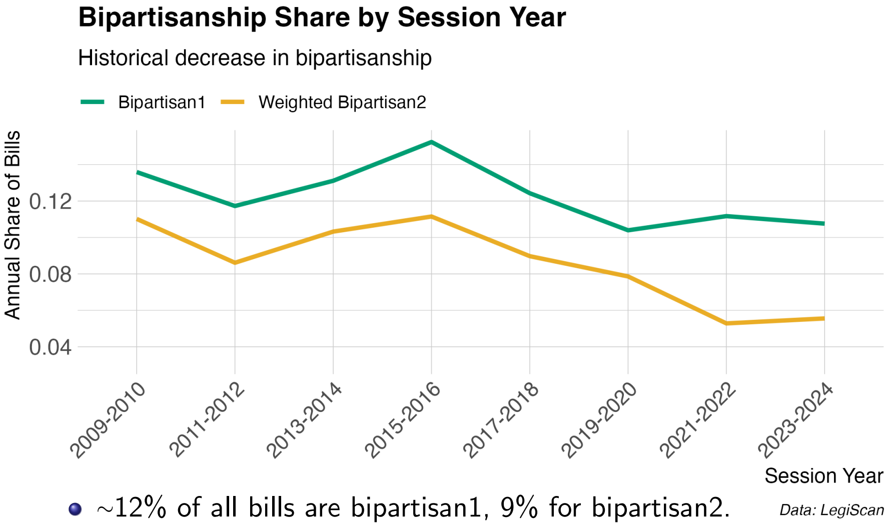
For fairness, we weight bipartisan2 and bipartisan3 by their annual party proportions (Ballotpedia).

Graphs: Party Percentages



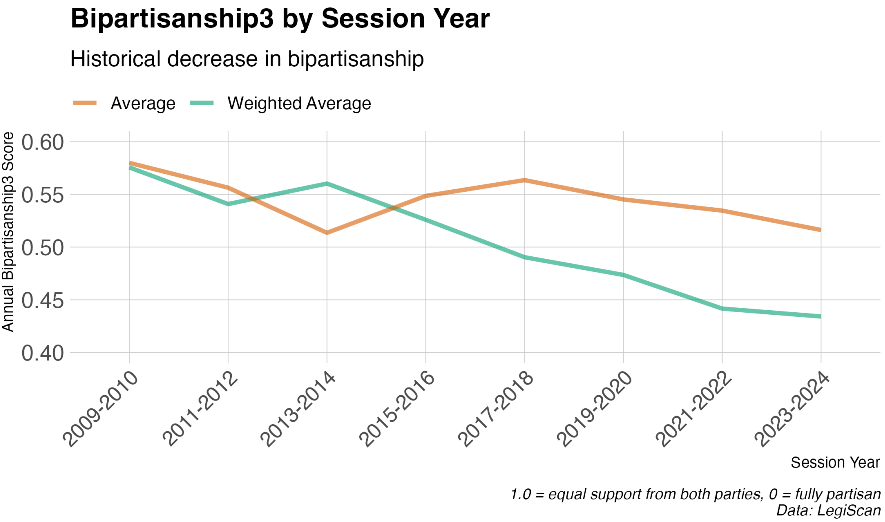
Weight by dividing bill's $\frac{\text{Republican sponsors}}{\text{Republican proportion}}$ and $\frac{\text{Democrat sponsors}}{\text{Democrat proportion}}$.

Graphs: bipartisan1 and bipartisan2

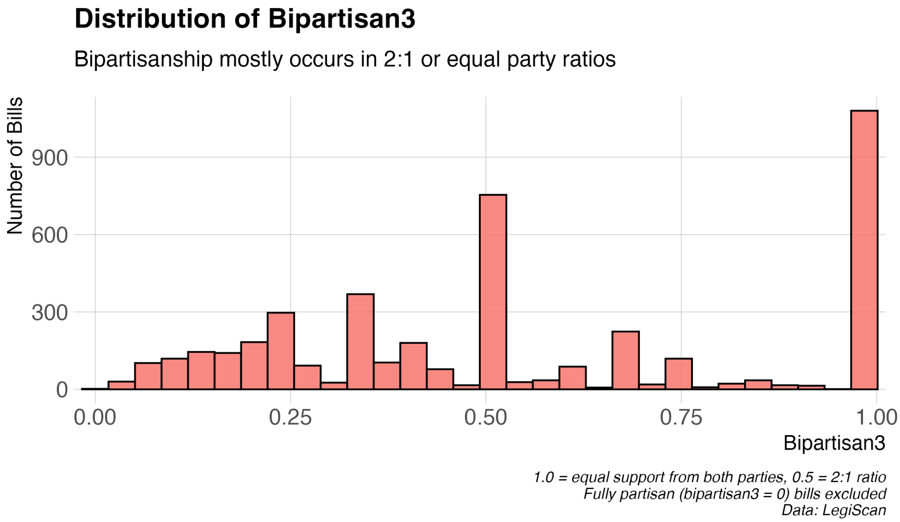


- ~12% of all bills are bipartisan1, 9% for bipartisan2.
- ~2.9% and 5.5% decrease respectively (= 150 and 260 less bipartisan bills).

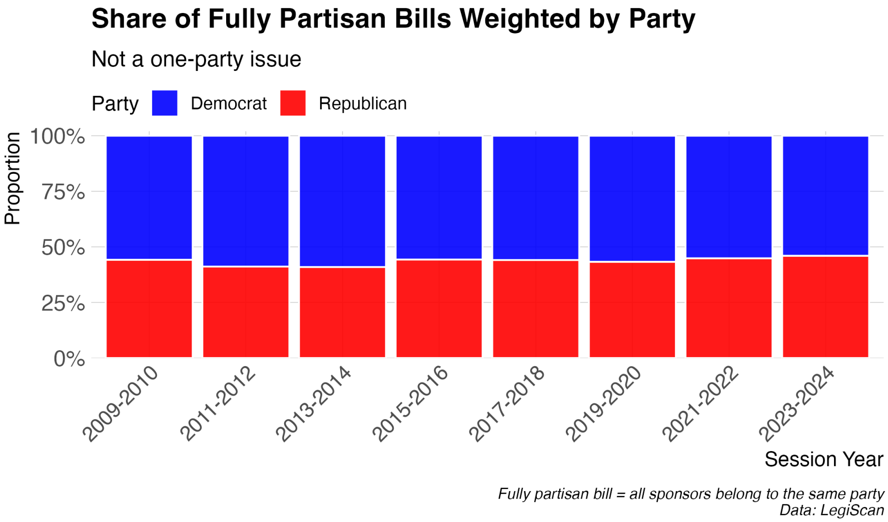
Graphs: bipartisan3



Graphs: bipartisan3 Insights



Graphs: Not a Single-Party Issue



Conclusion

1. Bipartisanship seems to be declining.
2. bipartisan3 > 0.5 (and its weighted variable) has decreased significantly over time, suggesting that politicians are increasingly working in more lopsided, partisan ratios.
3. Not a one-party issue.
4. Modern research: *bipartisanship pays*.
    - "The 'secret sauce' for effective lawmaking."
    - Election incentives.
5. Potential "solutions:"
    - Encourage cross-party collaboration and support bipartisan politicians across party lines.
    - Show politicians that bipartisanship *matters* through the vote.
    - Reform legislative rules/committees to encourage minority participation.

Twitter

Data

- Tweets from the official Twitter handles of all California State Senators.
- Tried running own Python scraping code.
- Problem: Elon Musk.
- Scraper service through Appify, expensive.
- $\sim$10.6K tweets ($\sim$14%) from 2013-2025.
- 36 senators.

Data

- Tweets from the official Twitter handles of all California State Senators.
- Tried running own Python scraping code.
- Problem: Elon Musk.
- Scraper service through Appify, expensive.
- $\sim$10.6K tweets ($\sim$14%) from 2013-2025.
- 36 senators.
- Date of tweet, likes, retweets, quotes, current follower count, and more (originally 155 variables).
  - Popularity = tweet + likes + retweets + quotes.
  - Popularity distribution is right-skewed $\Rightarrow$ use median.
- Apply sentiment analysis model VADER.

Sentiment Analysis: VADER

"No, I am your father" has a compound score of -0.296.

VADER

- Valence Aware Dictionary and sEntiment Reasoner.[1]
- Originally developed in Python, 2014.
- Lexicon (dictionary) designed for social media.
- Popular rule-based model, *not* machine learning $\Rightarrow$ not a "black box," predictable.
- Accounts for intensifiers, capitalization, punctuation, emoticons, slang, among others.
- R port or original package.
- Can score more positively then Bing, AFINN, and NRC.

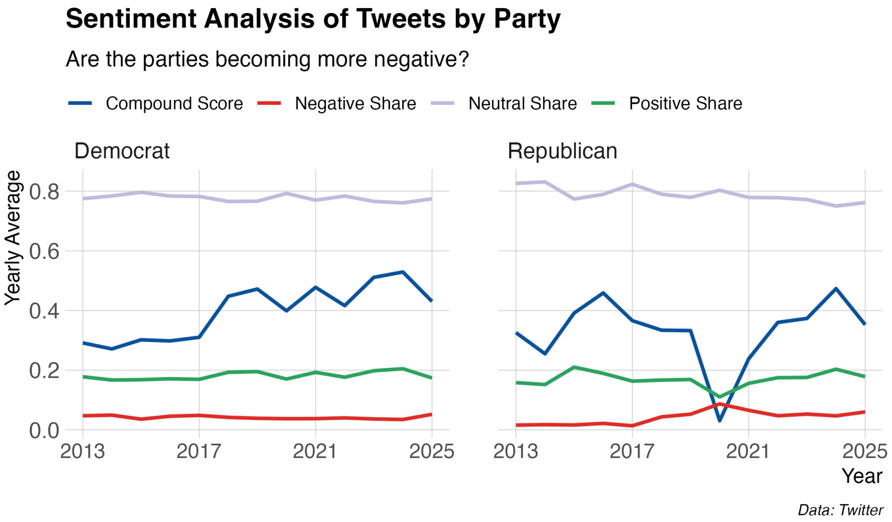
VADER: An Example is Worth a Thousand Words...

- Compound = sentiment score $\in [-1, 1]$ = [very negative, very positive].
- Compound $> 0.05 \Rightarrow$ positive, $< -0.05 \Rightarrow$ negative.
- Positive = percentage of text that is positive.

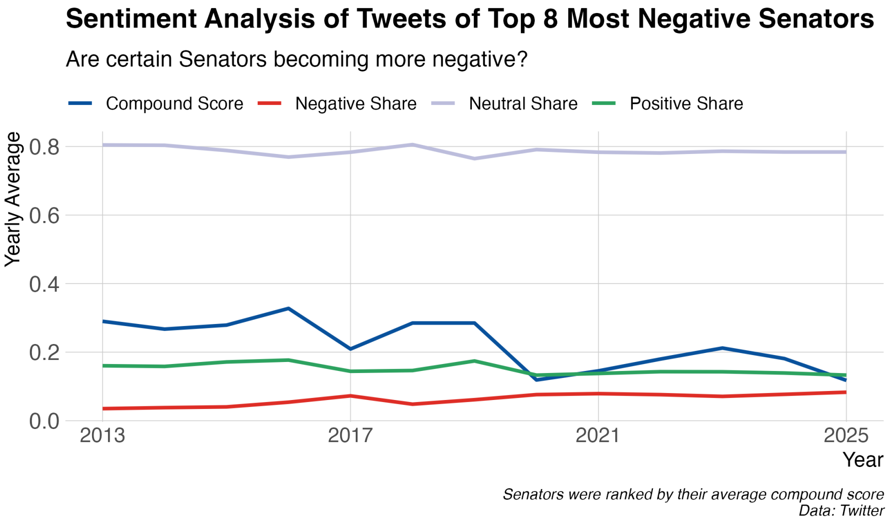
Text	Compound	Positive	Neutral	Negative
I love today and I hate tomorrow	0.1280	0.326	0.388	0.287
I love today 😊 and I hate tomorrow	0.7579	0.466	0.365	0.169
I love today, but I hate tomorrow!	-0.5778	0.201	0.386	0.413
I love today, but I really HATE tomorrow!	-0.7416	0.168	0.388	0.444
I do not love today	-0.5216	0.000	0.543	0.457
Today is not bad	0.4310	0.487	0.513	0.000

Sentiment Analysis: VADER

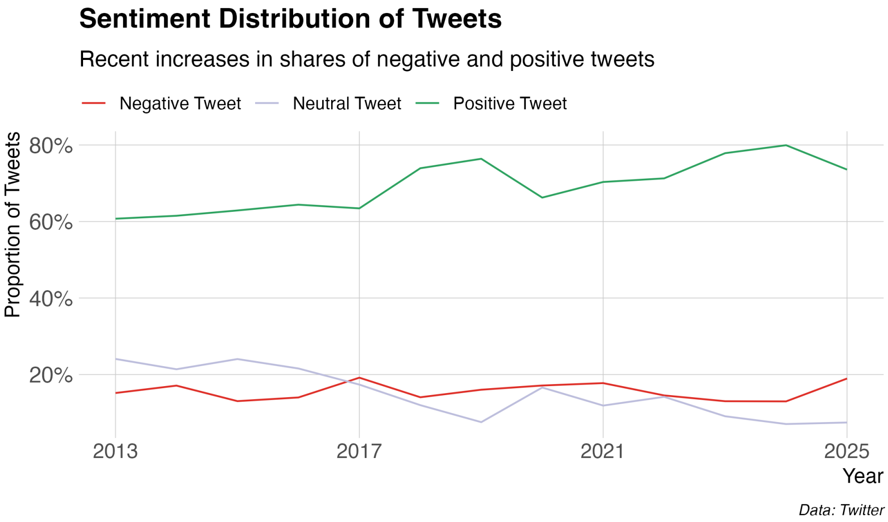
Graphs: Party Sentiment



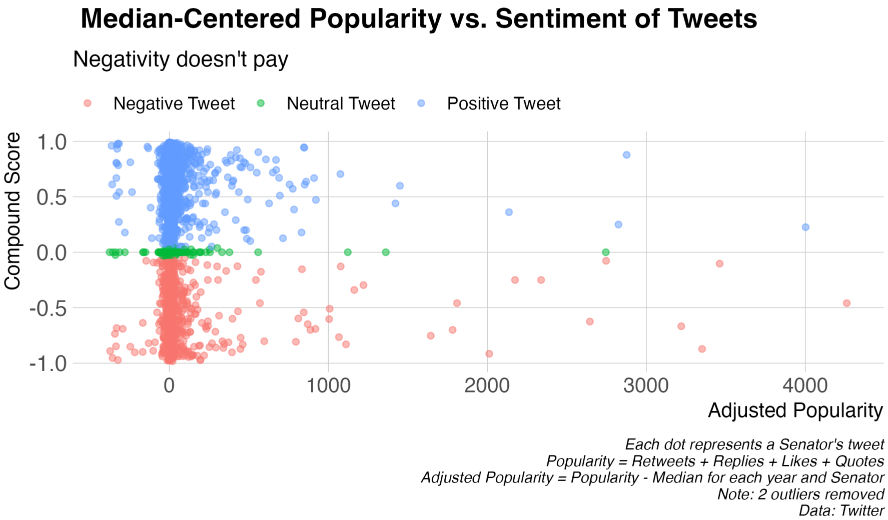
Graphs: Certain Senators' Sentiment



Graphs: Positive, Negative, and Neutral Tweets



Graphs: Negativity Doesn't Pay



Political Polarization

- **Idea:** Does cross-party hostility translate into real differences in bipartisan behavior? Do more politicians who publicly berate the other party *actually cooperate less with them politically*?
  - If yes $\Rightarrow$ problem.
  - If no $\Rightarrow$ problem ("weaponization of political polarization").
- Filter tweets for a string of key words (ex: "democrat, woke, trump, far-left, far-right, etc.") and for only negative tweets.
- Very rare ($< 1\%$ of all tweets).

Political Polarization: Example

Example:

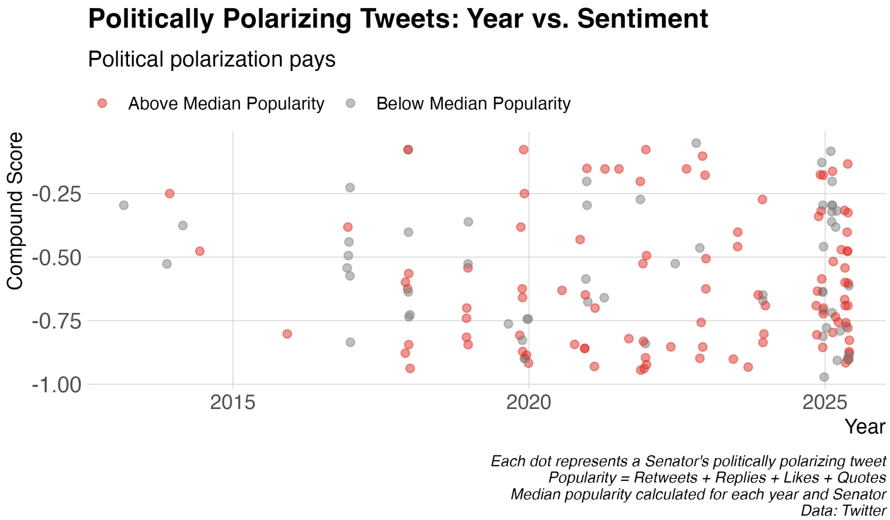
@AlvaradoGilSD4 (Senator Marie Alvarado-Gil, Republican)

"Once again, CA Dems are prioritizing criminals over victims.

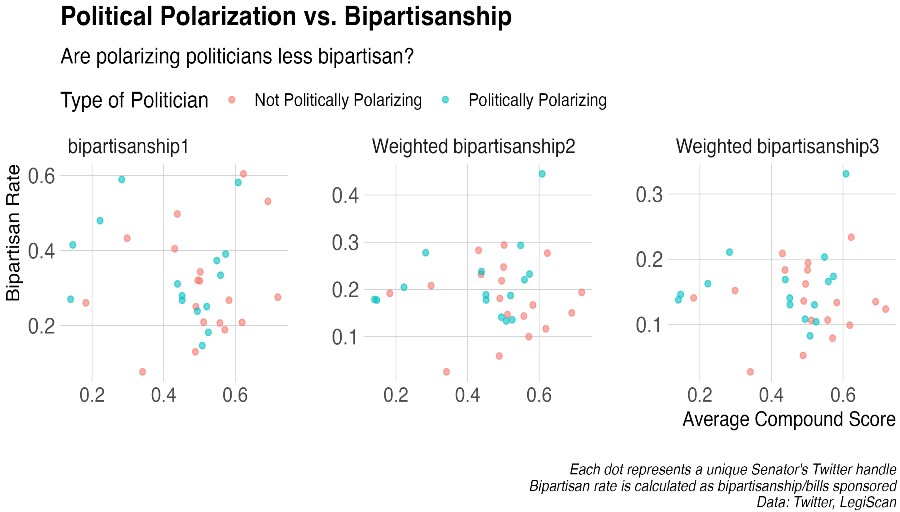
SB 672 would offer parole to certain killers sentenced to life without parole, just because they were "youthful" when they committed their crime/s & have served 25 years.

I'll be voting NO when this hits the Floor."

Graphs: Political Polarization



Graphs: Polarization Pays



Conclusion

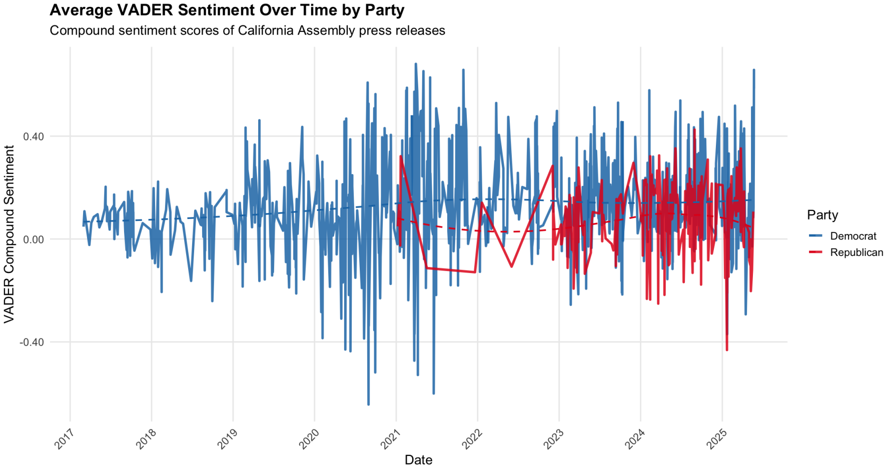
1. Increasing tweet negativity for *some* Senators.
2. More positive and negative tweets, less neutral tweets.
3. Negativity language does not appear to pay significantly, but politically polarizing language does.
4. Politically polarizing politicians don't seem to actually act on that cross-party resentment in the legislature.
5. Potential "solutions:"
    - Be aware of performance tweeting. Make others and politicians aware of the behavior.
    - Value of cross-party collaboration.
    - If the incentives are there, the behavior *will follow.*

Speeches

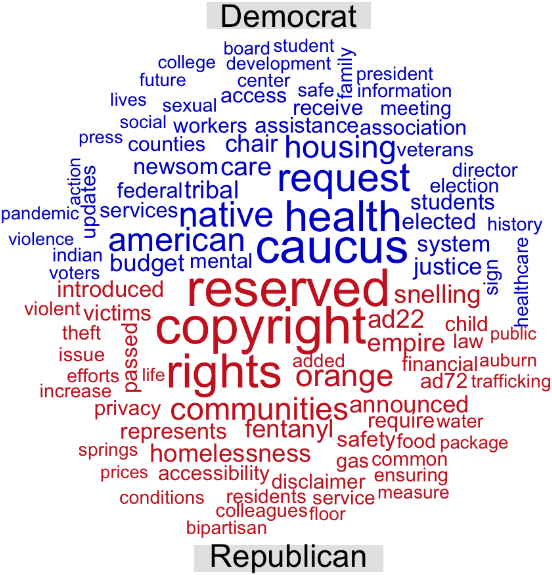
Data

- Web scraped from California Assembly website.
- Press releases of Democrats and Republicans.
- 2017-2024.
- Apply VADER.
- Text Mining.

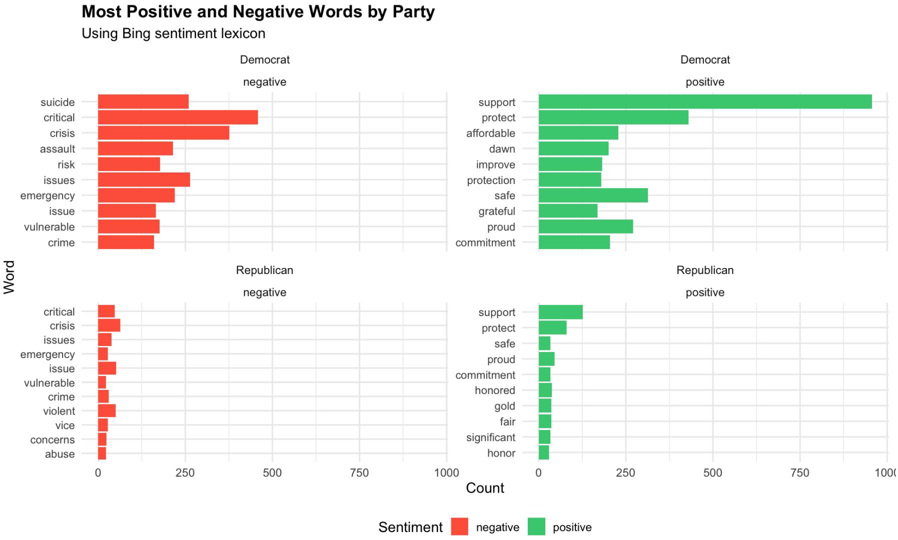
Graphs: Does Sentiment Change?



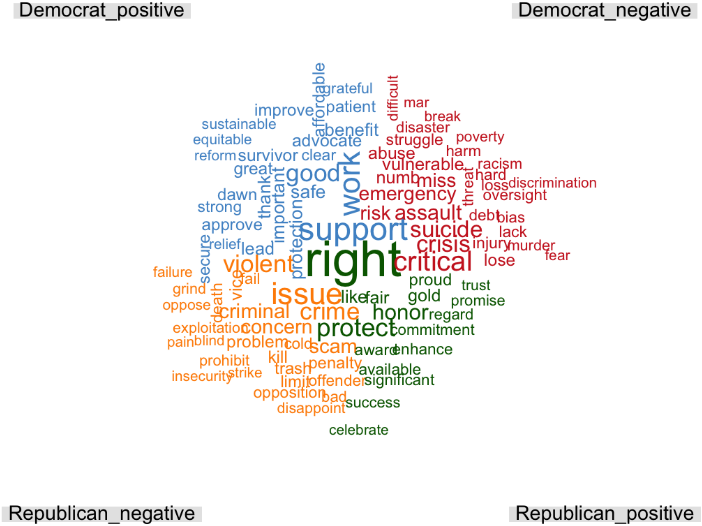
Graphs: Party topics



Graphs: Charged Vocabulary by Party



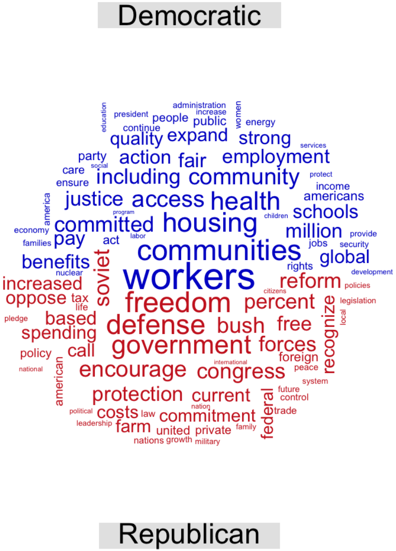
Graphs: Deeper Look



Party Platforms

- Web scraped national party platform statements from 1840-2024.
- Text mining.

Party Platforms: Party Topics



Conclusion

1. Average sentiment in press releases has stayed consistent.
2. Party topics do differ.
3. Charged (negative/positive) vocabulary is similar.
4. *Same game, different goals.*

Skills Learned and Used

- R packages:
  - **Web Scraping:** `RSelenium`, `rvest`, `netstat`, `wdman`.
  - **Data Wrangling:** `tidyverse`, `janitor`, `lubridate`, `fuzzjoin`, `data.table`, `tidylog`, `jsonlite`.
  - **Graphing:** `ggplot2`, `hrbrthemes`, `scales`, `knitr`, `kableExtra`, `webshot2`.
  - **Sentiment Analysis:** `reticulate`, `vader`.
  - **Text Mining:** `tidytext`, `wordcloud`, `reshape2`, `SnowballC`, `textdata`, `textstem`.
- Interacting with an API through an API key to retrieve data.
- Multiple R projects (5 subprojects).
- Managing a GitHub repository, locally cloning, and using Git Large File Service (LFS).
  - https://github.com/RileyBerman/DataHack_Group4_Project
- LaTeX+ Beamer.
- ChatGPT + Copilot (especially for graphing, commenting code).

Limitations and Further Research

- **Limitations:**
  - Broke college kids, Twitter is hard and expensive to scrape.
  - Limited time frame (∼6 weeks).
  - Learning curve.
  - VADER is not infallible.
- Our datasets have a wealth of information on bill text and titles. These could be interesting to apply sentiment analysis to.
- Deeper analysis to see exactly why our bipartisan metrics show a recent decline.
- Our datasets contain granular information on roll call/voting behavior of all California legislature politicians. We could explore these.
- A larger batch of tweets could enhance or contradict our findings.

Extra Time: Interesting Problems and Insights

- **LegiScan:** had to resort to Ballotpedia for session party counts. 120+ politicians could be recorded for a given session (vacancies, special appointments, deaths).
- *The Nature of Politics:* Ted Gaines and Brian Dahle both represented SD-001 in the 2019-2020 session. Gaines actually left early (January 7, 2019) to assume a new office as a Member of the California State Board of Equalization.
- *What are Even Term Limits?* Susan Eggman ended her tenure in the Assembly by replacing Cathleen Galgiani in the Senate in 2020, whom she had previously replaced in the Assembly in 2012. Same thing with Ted Gaines and Brian Dahle.

Extra Time: Interesting Problems and Insights

- This "double succession" where a California politician succeeds the same individual in both the Assembly and the Senate, happens 3 other times!

Incumbent	Replacement	Role Progression	Session Year
Jean Fuller	Shannon Grove	Assembly to Senate	2011-2012, 2019-2020
Marty Block	Toni Atkins	Assembly to Senate	2011-2012, 2017-2018
Cathleen Galgiani	Susan Eggman	Assembly to Senate	2013-2014, 2021-2022
Joel Anderson	Brian Jones	Assembly to Senate	2013-2014, 2019-2020
Ted Gaines	Brian Dahle	Assembly to Senate	2013-2014, 2019-2020

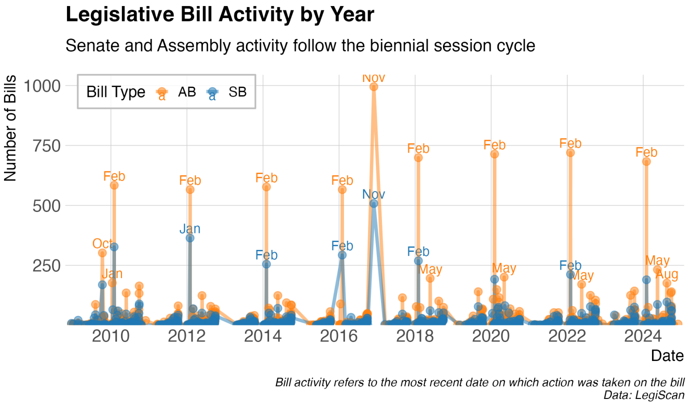
Data: LegiScan

Extra Time: Interesting Problems and Insights

- **LegiScan:** The failure rate of bills is understated. Many bills die out in committees, chambers, and politicians' desks, never to be heard of again. Why we created a new variable to account for this.
  - Spike in the proportion of failed bills in 2015-2016 session.
- **LegiScan:** The "X" of a bill's number (ex: "ABX11") indicates that the bill occurred in a special (extraordinary) session.
  - Highest number of extraordinary sessions occur in... 2009-2010 session (post-recession)!

Extra Time: Interesting Problems and Insights

- Why do we analyze in session years not years?



- Legislature activity is highly cyclical: lawmakers introduce bills in their first (odd-numbered) year, then work on them in the following (even-numbered) year.

Thanks for Listening!

References

[1] C.J. Hutto and Eric Gilbert. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text". In: *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI: Association for the Advancement of Artificial Intelligence (AAAI), 2014.