

PLSC 497 Final Project

Sentiment and Executive-Justice Ideological Distance in Supreme Court Opinions

Taran Samarth

April 29, 2021

Research Question + Background

In an age of extreme polarization and politicization of the courts, Supreme Court opinions have become increasingly partisan and biting in their criticisms of other justices and parties to a case. For example, in a dissent in *Higgs v. United States* this year, Sonia Sotomayor dug into the Trump administration's rush to execute several people on death row and the Court's opinion-less rulings in favor of the government. While researchers have studied how ideological distance affects vote outcomes for or against the executive branch, opinion language offers a more expansive set of information than a mere up-or-down vote from which we can understand justice attitudes towards the executive. Here, I attempt to examine if ideological distance affects justice attitudes towards the executive branch in Supreme Court cases.

So far, little research on the opinions of high courts using quantitative text analysis has been completed (although there certainly has been more in the last two years than ever before). There are many prohibitive reasons: words take on very different meanings and valences in legal writing than in our everyday discourse, the topics addressed by courts vary wildly from transportation to criminal procedure, and hand-coding long texts in order to conduct model diagnostics is difficult and time-consuming. Many of these problems still complicate this analysis, although I attempt to ameliorate some of these issues by implementing a latent semantic scaling (LSS) model.

Although no research has considered the textual treatment of parties like the executive branch within Supreme Court opinions, researchers have examined the effect of ideological distance between the executive branch and justices on votes in Supreme Court cases. Justice ideologies are normally scored using Segal-Cover scores (a static metric scored using newspaper characterizations of each justice as liberal or conservative during their nomination/confirmation period) and Martin-Quinn scores (dynamic metric that is updated after each decision in a case). While these scores can estimate polarization and the mean/median ideology of the Court, they can also be used to estimate ideological distance between each justice and other political actors whose ideologies are similarly scored. Epstein and Posner (2016), for example, use the absolute value of the difference between Presidential DW-NOMINATE and justice Martin-Quinn scores to measure ideological distance. They find that ideological distance between a justice and the sitting President has a significant negative effect on the likelihood that the justice votes for the executive branch. After fitting the sentiment analysis models, I attempt a post-hoc regression in the vein of Epstein and Posner (2016), using a similar set of covariates (including ideological distance), to see if justices talk about the executive branch more negatively when they are in different ideological camps (by mere party membership and other ideological measures).

Data Collection and Preprocessing

In order to identify the relevant set of votes and justice opinions for analysis, I join the vote-level Supreme Court Database with the biographical Supreme Court Justice Database, alongside measures for case salience. This merged dataset ranging from 1946 to 2009 is filtered to cases where justices authored an opinion and

the federal government was either a petitioner or respondent in the case. Here, “federal government” is defined as all cases where the federal government, a specific executive actor, or a Cabinet-level department is a party. I exclude all cases where two different federal agencies or actors are pitted against each other—this rare situation occurred mainly in the 1940s—as neither side straightforwardly represents “the Government” alone. I also add variables representing the party who held the executive and the DW-NOMINATE score for the President in office on the date that the case was argued before the Court. These variables are then used to identify if justices voted in favor of the federal government and if justices voted in favor of their own party.

After identifying all relevant votes with complete data availability ($n = 5003$), I use the case.law API to pull opinion texts for each vote. All votes where the SCDB has identified an opinion that case.law cannot retrieve are dropped—this represents roughly 11.93% of the dataset (597 cases). The texts are then converted into a document-feature matrix where punctuation and numbers are removed. Multi-word pronouns are preserved by compounding collocations of two or three capitalized words that occur at least thirty times in the corpus. Single-letter words and all terms that appear fewer than ten times in the corpus or in fewer than five documents are discarded.

Table 1: Corpus Metadata (N = 4396)

Statistic	Mean	St. Dev.	Min	Max
SCOTUS Term	1973	17.283	1946	2009
Segal-Cover Ideology	-0.554	0.335	-1.000	0.000
NYT Case Salience	0.205	0.403	0	1
Executive as Petitioning Party	0.485	0.500	0	1
Executive DW-NOMINATE	0.100	0.485	-0.504	0.693
Vote Cast for Executive	0.527	0.499	0	1
Justice-Executive in Opposing Parties	0.451	0.498	0	1

Model

I implement a latent semantic scaling (LSS) model in this analysis. LSS, developed by Watanabe (2020), is a semisupervised semantic analysis model based on word embeddings. Rather than merely matching terms from a dictionary with assigned valences to a corpus, LSS computes word polarities by estimating proximities between terms and researcher-assigned seedwords that are chosen for their unequivocal positive or negative meanings. Notably, LSS avoids the context-insensitivity problem from dictionary-based sentiment analysis—term polarities are derived solely within the corpus and are consequently context-specific. LSS also allows us to find sentiment around a certain term or dimension. Here, I pass in a list of tokens (unigrams, bigrams, and trigrams) that appear with statistically significant frequency around the bigram “the_government,” such that polarities are only fitted for terms that modify or address references to the federal government as a party before the Court. After estimating word polarities, document polarities are calculated by reducing each document to the model terms and then computing the document’s average weighted word polarity.

Additionally, unlike most unsupervised machine learning models, LSS models can be fit across an entire corpus, rather than a training sample, such that the model does not “suffer the differences in language usage between training and test sets.” I take this exact approach because, simply put, coding the sentiment of over 4000 opinions, such that I could complete some model diagnostics, is way too time-consuming. If I had the capacity, I would have three or four people code sentiment towards the federal government in these opinions and check the correlation over time between the LSS fits and the manual coding to complete some diagnostic analysis.

The primary researcher-controlled model parameter is the set of seedwords. Initially, I implement the model with a basic set of positive/negative words listed in Table 1, using wildcard endings to capture minor changes

(e.g. “bad*” to capture “bad” and “badly”). However, wildcards ended up including many irrelevant terms—“great*” turned “great_britain” into a positive seedword, for example—so I reduced the over-included set by hand to a set of 21 positive and 24 negative words, shown in Table 2.

Table 2: Seedwords for LSS Model

Sentiment	Terms
Positive	accurate, accurately, brilliant, brilliantly, wonderful, wonderfully, excellent, excellently, good, great, greatly, positive, positively, correct, correctly, terrific, fortunate, fortunately, superior, nice, nicely
Negative	erred, err, negatively, fail, poorly, negative, terrible, failure, bad, badly, poor, failing, errors, negatives, error, incorrect, wrong, erroneous, wrongly, errs, failures, fails, failed, erroneously

Since LSS is based on word embeddings, researchers must also select the size of word vectors (k). Watanabe (2020) proposes using the cohesion statistic which measures the marginal improvement of each additional vector component in capturing relevant semantics for the model (the math here deals with linear algebra which I can’t really explain well). He compares the k -cohesion plot with a k -correlation plot (where the correlation coefficient compares the fitted model values against the manually coded sentiment values). The two plots are, in his analysis of economic and political texts, similar—optimal points on the k -cohesion plots match with peak values on the k -correlation plots. In the absence of hand-coded sentiment values to replicate the same correlation analysis, I implement Watanabe’s suggestion that the highest peak on the k -cohesion plot is the near-optimal choice. Per Figure 1, I set k to 430.

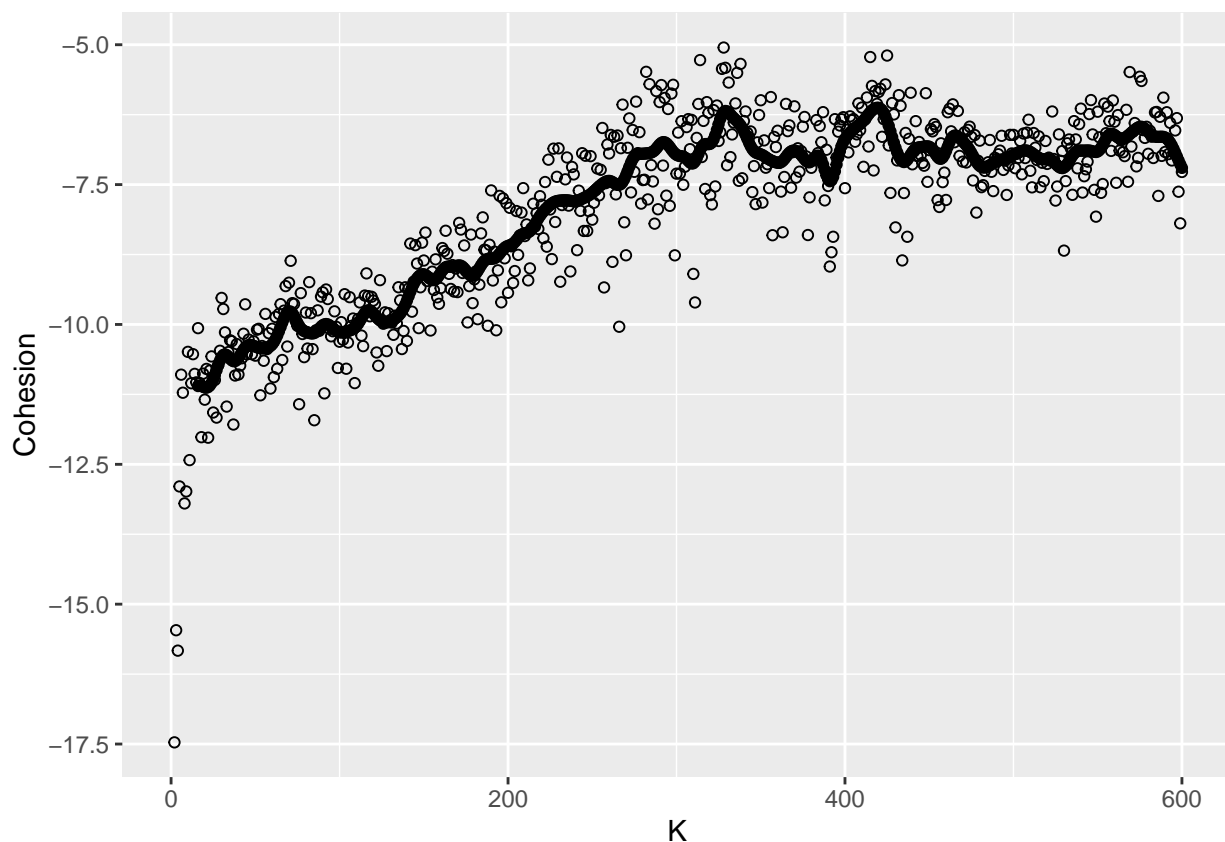


Figure 1: k -Cohesion Plot for LSS Model

The resulting fitted sentiment values from the LSS model is used as the dependent variable in the post-

hoc regression analysis. Covariates for the regression analysis include ideological distance, executive branch experience, federal government as respondent, and case salience. Ideological distance is measured by taking the absolute value of the difference between justice Segal-Cover and presidential DW-NOMINATE scores. Executive branch experience is measured using the Supreme Court Justice Database’s variables for previous employment in the executive branch, and federal government as respondent is a self-explanatory binary indicator. Case salience is a binary indicator, developed by Epstein and Segal (2000) and updated in 2011, which measures if the case appeared as the lead case within a New York Times front-page story the day after the Court issued its decision.

In general, I expect that Justices with executive branch experience will speak more positively about the federal government, but speak more negatively as ideological distance between the executive and the justice increases. Given the Court’s tendency to reverse, I expect that they will speak more positively about the federal government when the executive is a petitioning party. I also expect that Justices will speak more positively about the federal government in salient cases—important cases draw more media attention, so Justices may be more averse to using harsh language to avoid outcry which could undermine Court legitimacy.

Findings

The fitted LSS model develops polarity estimates for 1,208 terms that appeared with statistically significant frequency around “the government” across the corpus. As is evident in Figure 2, most of these polarities are close to zero, and those which stray significantly away from zero tend to have clear meanings (e.g. the most negative terms are negative seedwords “failing,” “fails,” and “failed”). There are still terms, such as the positive seedword “correct,” that take on sentiment scores in the opposite polarity than expected. Some of these are evaluative of arguments like “assert” and “sustain” which take negative sentiments but are coded positively in the Lexicoder Sentiment Dictionary (LSD); other legal terms that are negative in everyday speech but are merely terms of art in law like “indictment” or “charge” take on positive or neutral sentiments when they are coded as negative in the LSD. In general, however, there are more terms with large negative polarity values than large positive polarity values, indicating that the negative seedwords were likely more effective in estimating polarities for terms than the positive seedwords.

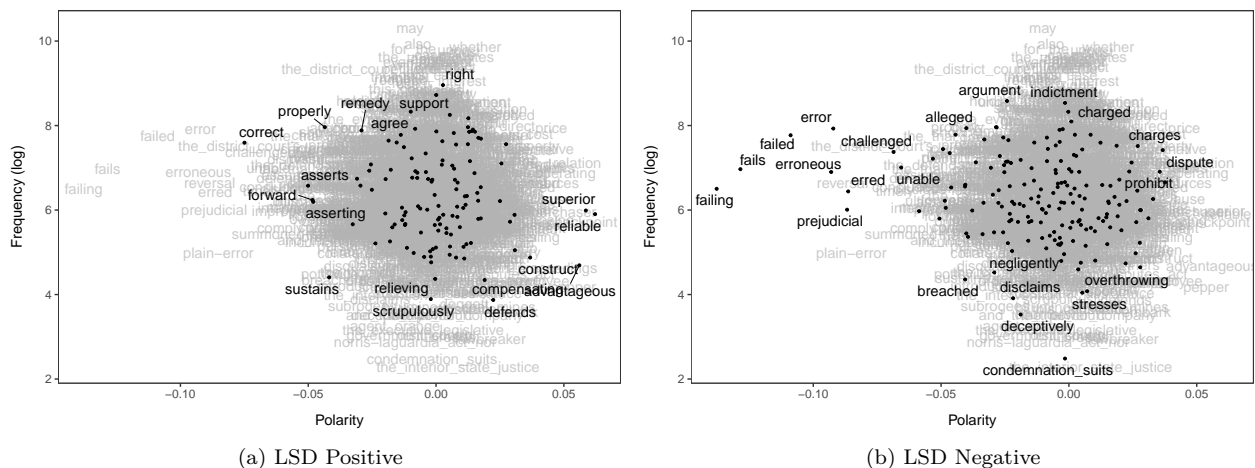


Figure 2: Model Term Polarities with Lexicoder Sentiment Dictionary (LSD) Highlighted Terms

Using the fitted document sentiment scores, I plot the Court’s average sentiment towards the federal government over time in Figure 3. Between 1946 and 2009, the Court has become generally more negative in its textual treatment of the federal government. Truman, for example, nominated four justices to the Court, so receiving a generally positive treatment from a Court full of his own appointees is no surprise (Epstein and Posner, 2009). Other significant changes in sentiment towards the federal government cannot be easily

explained including the dips at the end of the Ford administration and the middle of the Reagan years. Figure 4 in the appendix¹ also shows the average opinion sentiment by justice. As expected, justices like Antonin Scalia and Ruth Bader Ginsburg generally write negatively about the federal government, but so do all Justices who served after 1968. The few justices with positive average sentiments served in the first twenty years of the dataset, and those datapoints are not easily explainable either. Further research that might explain these could provide some meaningful historicity to the model, though, and would definitely be worthwhile outside of a class project!

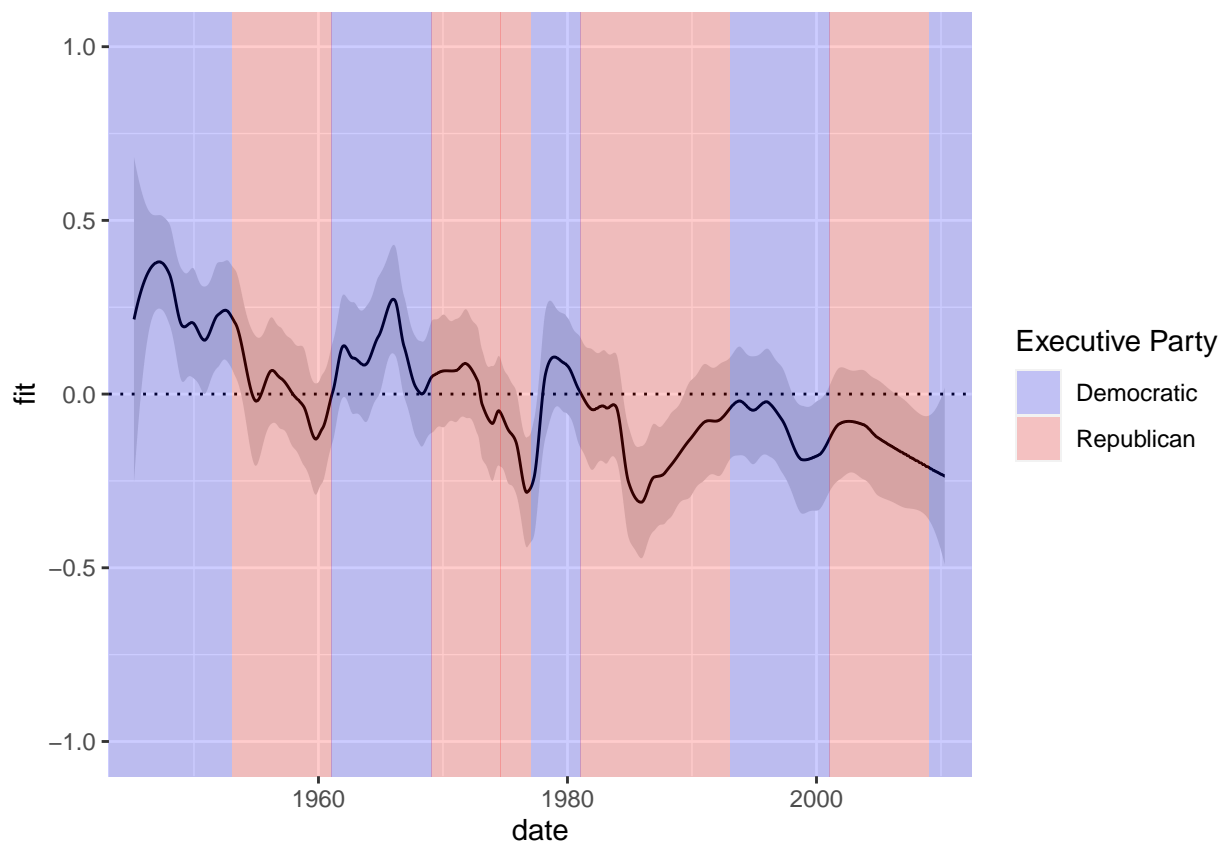


Figure 3: Average Court Sentiment Over Time

The model does show that the average sentiment in cases where the executive and the justice are in opposing parties is significantly more negative than those where the executive and justice are co-partisans—one answer, albeit a simple one, to the original research question. However, a mere party membership analysis is not robust to ideological differences within each party nor changes in party ideology over time. To accommodate for ideological distance and other mediating variables, I conduct a robust linear regression analysis² on document fits with ideology, federal government as petitioner, case salience, and executive experience as covariates.

The two regression models in Table 3 show that ideology—measured either by mere party membership or continuous ideological distance—has a significant effect on the LSS sentiment values for language regarding the federal government. When a justice is not a co-partisan with the executive, the expected sentiment value

¹For these group means, I use the variance-sum property to calculate standard errors and confidence intervals. Given the fit and standard error of the fit for each document, I square the document standard error, sum them across the group, and divide by the square of the number of documents within the group to get group-level variance (the square root of which is the group-level standard error).

²I use robust linear regression because, although fits are rescaled by LSS onto a normal distribution, there are still a handful of extreme outliers. Using the weighted least squares regression function available in the `robustbase` package reduces the effect that outliers interfere with the results.

drops by .087. Similarly, in the ideological distance model, a one-unit increase in the difference between a justice’s Segal-Cover score and the president’s NOMINATE score decreases the expected sentiment value by .120. With the assumption that the initial LSS model accurately estimates opinion sentiment values around the term “the government,” this analysis provides affirmative evidence for the question on whether or not justices speak more negatively about the federal government when the executive is not co-partisan with the justice and is party to the case. The models also show that justices speak more positively about the federal government in salient cases and when the justices themselves have executive experience, concurring with Epstein and Posner’s (2016) findings of the same effects on binary vote outcomes in favor of a Justice’s appointing President.

Table 3: Sentiment Regression Models

	Sentiment	
	(1)	(2)
Justice-Executive in Opposing Parties	−0.084*** (0.025)	
Ideological Distance		−0.114*** (0.029)
Federal Government as Petitioner	0.102*** (0.025)	0.111*** (0.025)
Case Salience	0.237*** (0.027)	0.239*** (0.027)
Executive Experience	0.067** (0.026)	0.061* (0.026)
Constant	−0.045 (0.028)	−0.001 (0.033)
N	4,396	4,396
R ²	0.022	0.023
Adjusted R ²	0.021	0.022
Residual Std. Error (df = 4391)	0.781	0.782

*p < .05; **p < .01; ***p < .001

Conclusion and Limitations

The analysis here appears to show that justices speak more positively about co-partisan executives than out-partisans, but the most glaring limitation is that the LSS model still lacks thorough diagnostics to check whether or not it accurately estimates opinion text sentiments towards the federal government in these cases. While I make as many efforts to refine the model as Watanabe (2020) proposes (e.g. cohesion statistic optimization), LSS still has not been checked against manual coding in a legal context, and it probably should before these results are considered statistical evidence in favor of the research question.

Works Cited

- Epstein, Lee, and Eric A. Posner. 2016. “Supreme Court Justices’ Loyalty to the President.” *The Journal of Legal Studies* 45(2): 401–36. <http://dx.doi.org/10.1086/688395>.
- Epstein, Lee, and Jeffrey A. Segal. 2000. “Measuring Issue Salience.” *American Journal of Political Science*. 44(1): 66-83.
- Epstein, Lee, Thomas G. Walker, Nancy Staudt, Scott Hendrickson, and Jason Roberts. (2021). “The U.S. Supreme Court Justices Database.” January 17. <http://epstein.wustl.edu/research/justicesdata.html>.

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2021). Voteview: Congressional Roll-Call Votes Database. <https://voteview.com/>

Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2020 Supreme Court Database, Version 2020 Release 01. <http://supremecourtdatabase.org>

Watanabe, Kohei. 2020. “Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages.” *Communication Methods and Measures*: 1–22. <http://dx.doi.org/10.1080/19312458.2020.1832976>.

Appendix

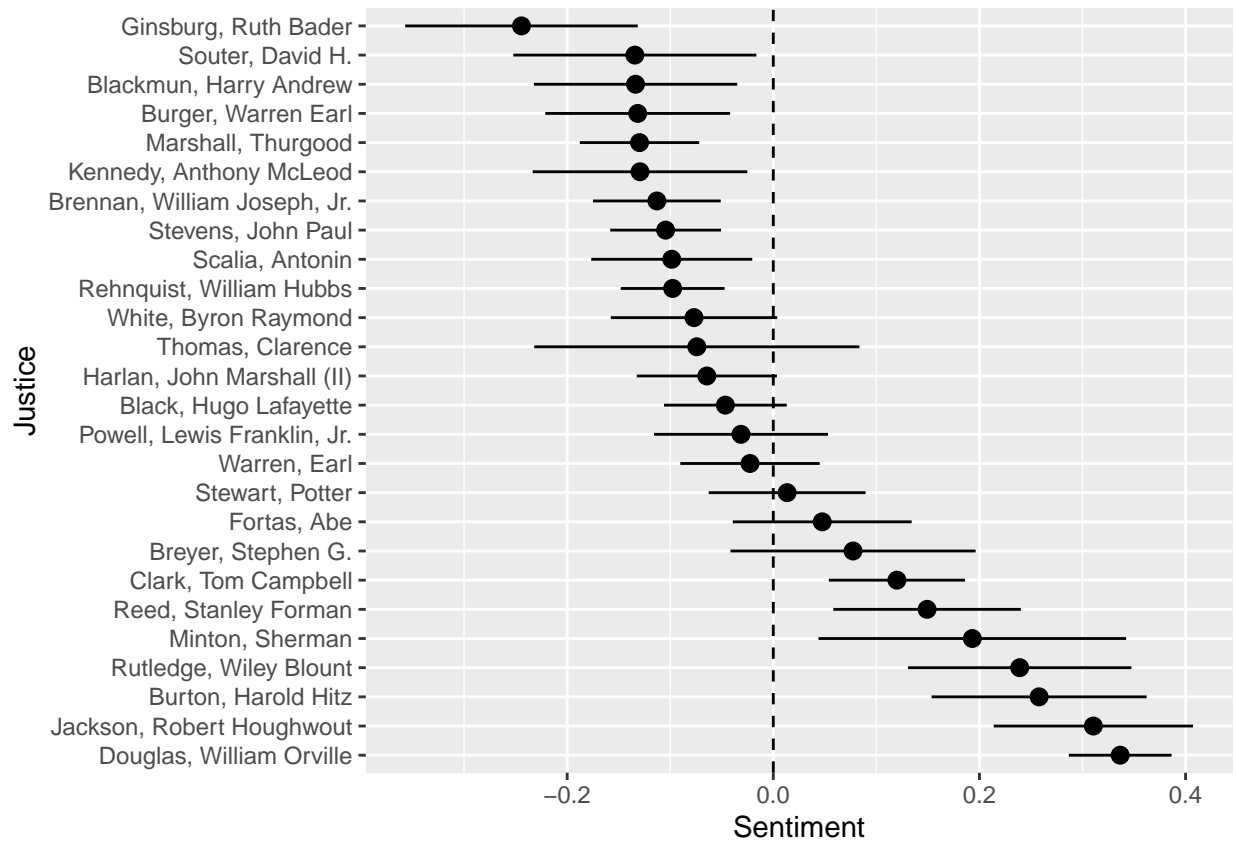


Figure 4: Average Sentiments by Justice