A Machine Learning Approach to Investigate Partisanship of Language in Recent Supreme Court Opinions

Maya Jeyendran
Senior Thesis in Statistics and Data Science
Advised by Prof. Zhou Fan

April 30, 2020

Abstract

We study the partisanship of the language of Supreme Court opinions from 1946 to 2018. We find that partisanship remains relatively constant over time with the exception of four peaks. We also study partisanship from a variety of cross-sections including demographic information about Justices as well as topics of opinions themselves.

1 Introduction

Language indicating a high degree of partisanship is common in media, political campaigns, everyday conversations, and more. It is difficult to expect, however, whether the language used in the nation's highest Court would exhibit the same degree of partisanship. One might expect the partisanship appearing in Court opinions to remain relatively low due to, for example, a reasonable belief that Justices impartially evaluate cases exclusively based on their merits and other, hard factors; on the other hand, Justices are people too, some of whom have expressed their political views with little room for ambiguity in interpretation. For example, in 2016, Justice Ginsburg provided the following strong opinion about Donald Trump: "He is a faker. He has no consistency about him. He says whatever comes into his head at the moment. He really has an ego," – a statement she later came under much fire for [1]. Around the same time period, abortion rights activists, politicians, and liberal-leaning Justices raised concerns over "the pace at which the conservative [Court] majority was overruling precedents," and publicly questioned whether landmark verdicts would be overturned. [2]. In 2018, Donald Trump tweeted about necessity of maintaining a Republican-majority Supreme Court to ensure a future of conservative-leaning legislation [3]. Two years earlier, Chief Justice John G. Roberts Jr. stated that "[judges] don't work as Democrats or Republicans," in criticism of the contentious and party-line-driven blockage of Merrick Garland's would-be appointment [4]. Justice Roberts' sentiments were echoed shortly after by Justice Neil M. Gorsuch, who claimed that "[he doesn't] see Republican judges, and [he does] not see Democrat judges. [He] just [sees] judges," [5]. With the motivation these pieces of contrasting information create in mind, we now turn to the original question of whether or not partisanship abounds in the language of recent Supreme Court opinions.

To answer this question, we apply machine learning and text modeling techniques to understand the language that appears in opinions written between 1946 and 2018. We use a discrete-choice model to simulate speech as in Gentzkow et al (2016), and we define the partisanship of an opinion to be a weighted average of the likelihoods of a given phrase to have been spoken by a Justice belonging to a particular party, where higher weights are given to phrases that are much more likely to have been penned by Justices of one political party than the other political party.

Our results indicate that partisanship of the language in these opinions has remained relatively constant over time with occasional peaks that are difficult to contextualize. No discernible, sustained increase or decrease in partisanship was observed, though clear peaks in partisanship that lasted for only a single year were observed in 1954, 1966, 1996, and 2011. We further examine the behavior of this partisanship from a number of standpoints. We study trends in partisanship across the gender of Justices over time, trends in partisanship across the political party of Justices over time, and compute average partisanship scores for each individual Justice. We find that partisanship of opinions written by female Justices is visibly lower than partisanship of opinions written by male Justices at all points in time, partisanship of opinions written by Democratic Justices is considerably higher than partisanship of opinions written by Republican Justices over two distinct periods of time, and while most Justices have average partisanship scores that indicate a neutral writing style, a few Justices exhibit anomalous average partisanship scores. We also briefly study phrases that develop large coefficients for the same features of gender and political party when trained with our model and find that on an intuitive level, these phrases make sense (e.g. "exclusion [of] women" was found to be such a phrase for the gender feature; "partisan bias" was found to be such a phrase for the political party feature). We also find that clearly delineated clusters of opinions corresponding to natural topics such as immigration and the Fourth Amendment clearly emerge; additionally, we find that partisanship between the topics we identify is relatively constant with a couple exceptions we draw attention to.

We provide details about our dataset, models and assumptions, and approaches in the next sections. We present our results on partisanship and attempt to tie the phenomena we observe to major historical and other events as well as demographic information regarding the Justices. We then provide suggestions and opportunities for future work as well as discuss limitations inherent to our approach.

2 Data

We use *this* dataset of Supreme Court opinions from Kaggle; this dataset was originally obtained from the Washington University in St. Louis (WUSTL) Supreme Court database, consists of opinions as well as identifying information such as opinions' authors, dates of publication, and more sourced directly from the Court [6]. We restrict our analysis to opinions penned after 1945, i.e. opinions written post-World War II. To add Justice characteristics such as gender and political party to our dataset of opinions, we supplement the dataset of opinions with another dataset, also from WUSTL's Supreme Court database [6].

Our initial text processing approach has very much in common with the approach used by Gentzkow et al. as well as standard approaches to processing raw text in general (see [7]). We first eliminate opinions whose raw texts are found to contain fewer than 2,000 words. We then remove stop words, punctuation, numbers and words containing numbers, proper nouns, text commonly found in citations (e.g. "U.S.C."), and currency symbols from the texts of the opinions that remain. Following that, we lemmatize words in our texts and group individual lemmas into 2-word phrases, i.e. bigrams. We further restrict our analysis by considering only bigrams that occurred at least four times in a given opinion and at least 15 times across all opinions. This process leaves us with a total of 12,617 opinions and 37,733 bigrams to learn from. We do not distinguish between majority opinions, concurrences, and dissents; we treat them all as opinions.

3 Modeling and Approach

Our work consists of two primary modeling tasks: developing a method to estimate partisanship of individual opinions, and developing a method to classify opinions by topic. We provide details behind the approaches we took to each task here.

3.1 Estimating Partisanship

We employ a two-part approach in which a model without regularization is first fit to identify appropriate parameters to use when re-fitting the model with regularization.

As in Gentzkow et al., we let $\mathbf{c_i}$ be the vector of bigram counts in opinion i as written by the Justice with characteristics $\mathbf{x_i}$; let $\mathbf{m_i}$ be defined as $\sum_\mathbf{j} \mathbf{c_{i,j}}$, i.e. the sum of all phrase counts in opinion i. Let $\mathbf{P(i)} \in \{\mathbf{R}, \mathbf{D}\}$ denote the political party of the Justice who wrote opinion i [7]. Let $\mathbf{q^{P(i)}}(\mathbf{x_i})$ be the probability distribution vector consisting of the probabilities of the Justice with characteristics x_i and political party affiliation $P(i)$ writing the phrases in opinion i [7]. In particular, $x_i \in \mathbb{R}^4$ is a vector of indicators created as follows: its first coordinate is always 1 to represent an intercept term; its second coordinate is 1 if the the Justice who wrote opinion i is female and 0 if male; its third coordinate is 1 if the the Justice who wrote opinion i identifies as a Republican and 0 if they identify as a Democrat; and its fourth and final coordinate is 1 if the Justice who wrote opinion i is a Chief Justice and 0 if the Justice is an Associate Justice. Then, we assume that [7]:

$$\mathbf{c_i} \sim Multinomial(m_i, \mathbf{q^{P(i)}}(\mathbf{x_i})) \tag{1}$$

In other words, we characterize the counts of bigrams in a given opinion solely using the total amount of speech in that same opinion and the probability of a Justice with given characteristics speaking each bigram. In order to determine what $\mathbf{q^{P(i)}}(\mathbf{x_i})$ looks like for a given vector of Justice characteristics and political party affiliation, we employ the following definitions [7]:

$$q_j^{P(i)}(x_i) = e^{u_{i,j}} / \sum_l e^{u_{i,l}} \tag{2}$$

where $u_{i,j}$ is defined for every phrase and opinion as follows [7]:

$$u_{i,j} = \alpha_j + x_i^T \gamma_j + \phi_j \mathbb{I}_{P(i)=R} \tag{3}$$

We may think of α_j as a phrase-specific intercept; γ_j as a set of coefficients specific to each phrase and Justice characteristic; and ϕ_j as a coefficient specific to each phrase and the political affiliation of the Justice. The remainder of our approach deviates from that used in Gentzkow et al. slightly. We estimate the above phrase-Justice specific coefficients in the following two-part approach. First, we compute initial estimates of $\{\alpha_j, \gamma_j, \phi_j\}$ via maximization of the following un-regularized likelihood equation:

$$\sum_{i=1}^{N} \sum_{j=1}^{M} y_{i,j} \beta_{\mathbf{j}}^{\mathbf{T}} \mathbf{x_i} - e^{\beta_{\mathbf{j}}^{\mathbf{T}} \mathbf{x_i}} \tag{4}$$

where N denotes the total number of opinions and M denotes the number of bigrams in opinion i. Encoded in the decision to maximize this equation that we are approximating the likelihood of the multinomial logit model introduced in (1) with the above likelihood of the Poisson log-linear model [7]. We make this approximation to make use of distributed computing to estimate our model parameters for each phrase and in particular to avoid calculating the denominator of the probabilities in (1).

We employ gradient ascent with a learning rate of 1×10^{-6} to learn our matrix of un-regularized coefficients we call β_{NREG}. We found that 25 iterations were sufficient to guarantee an appropriate amount of convergence of the cost function in (4).

Before including a regularization term, we identify the two most prominent clusters of coefficients in β_{NREG} using the standard k-means clustering algorithm. Let $\mu_{\mathbf{1}}$ denote the most prominent cluster of coefficients as identified by this algorithm, and let $\mu_{\mathbf{2}}$ denote the second-most prominent cluster of coefficients as identified by this algorithm. Let $\mathbf{n_j}$ denote number of times phrase j occurs across all opinions. We then re-learn the matrix of coefficients β by maximization of the following *penalized* likelihood equation:

$$\sum_{i=1}^{N} \sum_{j=1}^{M} y_{i,j} \beta_{\mathbf{j}}^{\mathbf{T}} \mathbf{x_i} - e^{\beta_{\mathbf{j}}^{\mathbf{T}} \mathbf{x_i}} + \eta \sum_{j=1}^{M} \frac{1}{n_j} log(e^{-\dfrac{||\beta_{\mathbf{j}} - \mu_{\mathbf{1}}||_2^2}{2}} + e^{-\dfrac{||\beta_{\mathbf{j}} - \mu_{\mathbf{2}}||_2^2}{2}}) \tag{5}$$

We motivate both the necessity of regularization as well as our particular approach here. We use $L1$ regularization to shrink the coefficients of β_{REG} to avoid over-fitting. In the absence of a supervised problem for which we would have been able to use a form of cross-validation or another approach to prevent overfitting, we use this form of regularization to push coefficients away from those that would leave us with an over-fit model. We regularize the coefficients with the Gaussian mixture-model approach in (5) to formulate the belief that clusters in coefficients are likely to emerge, and we wish to pull coefficients away from those clusters.

We also note that (4) is convex in β, implying that the matrix β_{NREG} which resulted from applying gradient ascent to our non-regularized cost function is likely to be close, if not equal, to the true matrix of coefficients that would maximize (4) globally. On the other hand, we observe that (5) is not convex in β; while gradient descent can be applied to non-convex cost functions, convergence to global maxima is not always guaranteed. Therefore, we use our un-regularized matrix β_{NREG} as the initial estimate of coefficients for the gradient descent algorithm we use to maximize (5); we refer to the matrix of coefficients learnt from this second application of gradient ascent as β_{REG}. We again use a learning rate of 1×10^{-6}, run gradient ascent for 25 iterations, and we set η, our regularization coefficient, to be 0.1. In addition to learning this improved β_{REG}, we simultaneously learn improved estimates of μ_1 and μ_2 with gradient ascent to account for the fact that the clusters themselves may shift as better estimates of the true matrix β are computed.

Now that we have shown how we estimate $q_{i,j}^{P(i)}$ using the above approach, we move to defining partisanship. The partisanship of opinion $\mathbf{x}$ is defined in Gentzkow et al. to be [7]:

$$\pi(x) = 1/2 \cdot (\mathbf{q^R}(\mathbf{x}) + \rho(\mathbf{x})) + 1/2 \cdot (\mathbf{q^D}(\mathbf{x}) + (\mathbf{1} - \rho(\mathbf{x}))) \tag{6}$$

where $\mathbf{q^R}(\mathbf{x})$ and $\mathbf{q^D}(\mathbf{x})$ consist of the coordinates $q_{i,j}$ as computed above for each opinion x and political party, and $(\mathbf{1} - \rho(\mathbf{x}))$ is defined to be the coordinate-wise difference between a vector of the appropriate dimension consisting of all 1's and the components of $\rho(x)$. We differ from Gentzkow et al. in how we choose to define ρ. Let $\mathbf{d}(\mathbf{x})$ denote the element-wise absolute difference between $\mathbf{q^R}(\mathbf{x})$ and $\mathbf{q^D}(\mathbf{x})$, i.e. $d_j(x) = |q_j^R(x) - q_j^D(x)|$. We then rank the coordinate-wise differences in order of greatest to least difference, creating a vector called

$\mathbf{rank}(\mathbf{d}(\mathbf{x}))$ in the process. We then let

$$\rho_j(x) = \frac{M - rank_j(d(x)) + 1}{\sum_{i=1}^{M} i} \tag{7}$$

where M denotes the number of phrases in opinion x.

We note that Gentzkow et al. define ρ as below:

$$\rho_j(x) = \frac{q_j^R(x)}{q_j^R(x) + q_j^D(x)} \tag{8}$$

We shy away from this definition for the following main reason. We want to learn more about partisanship from phrases whose probabilities of being spoken by a Republican and Democratic Justice vary significantly, which involves taking into account the absolute difference of the these probabilities for each phrase in each opinion in addition to placing more weight on phrases for which this difference is greater. In other words, we rank the phrases in order of greatest-to-least by this absolute difference, and place the highest proportional weight upon the phrase with the largest such difference, the second highest proportional weight upon the phrase with the second-largest such difference, and so on. We do this in an intuitive way such that the sum of all the weights is 1 (see (7)). We computed the partisanship for each opinion using both the formula in (8) as well as our own formula and found that our formula resulted in larger average partisanship scores on average, and peaks and valleys in partisanship using our definition were exaggerated.

3.2 Topic Modeling

We now classify opinions by topic, where the topics are learnt in the following unsupervised manner.

We clean and lemmatize opinions using the exact same approach described above. Instead of relying upon bigrams to featurize our text, we compute term frequency-inverse document frequency (tf-idf) vectors for each document and aggregate the collection of vectors into a matrix. To avoid dealing with computational issues and to better prepare our features for clustering, we perform Latent Semantic Analysis (LSA) to reduce the dimensionality of our tf-idf matrix. We choose to reduce our dataset to 100 combined features in this way as 100 is the recommended number of combined features to be used in conjunction with LSA [8].

To better understand the output of LSA in this scenario, we examine the top six combined features produced by LSA more closely in *Figure 6*. *Figure 6* consists of "word clouds" for each of these six aggregated features. Each aggregated feature consists of weights to be given to our original features, which, we recall are individual, lemmatized words; in our word clouds, for each aggregated feature, we plot the 20 individual words with the greatest weights proportional to their text size. The output of this process is in highly interpretable and supports the notion that LSA was an appropriate algorithm to apply here (see *Section 4.7.1* further for discussion).

To cluster opinions by topic and display the resulting clusters in a clear manner, we first further reduce the dimensionality of our the matrix of weights produced by LSA and feed this further-reduced data to the *t-sne* algorithm, which aids us in visualizing the eventual clusters of opinions. We further reduce the dimensionality of our data before applying *t-sne* again to avoid running into computational issues. We determined the number of dimensions to further reduce our data to by plotting the amount of explained variance introduced for each successive feature that resulted from applying SVD to the original data earlier. We found that reducing our data to 75 features created an acceptable trade-off between the number of features and amount of additional variance explained by each successive feature (see *Figure 9*).

We then use the k-means clustering algorithm to matrix of 100 features produced by LSA. We determined the number of clusters to use in the k-means algorithm largely through trial and error. We found that the interpretability and coherence of the aggregated features and their constituent words discovered through LSA dropped significantly after the 15th aggregated feature, so we choose 15 to be the number of clusters identified through k-means here. We label each cluster with the three words that have the highest weights for that cluster's identified words.

4 Results and Conclusions

4.1 Summary

We find that distinct, single-year-long peaks in average partisanship occur in 1954, 1966, 1996, and 2011; that a majority of Justices have average partisanship scores around 0.515 with a few Justices exhibiting anomalous scores; that the average partisanship of opinions written by female Justices is lower than the average partisanship of opinions written by male Justices at all points in time; and that partisanship of opinions written by Democrat-identifying Justices is higher than that of opinions written by Republican-identifying Justices over two distinct periods of time. We find that clear clusters of topics emerge and that partisanship of opinions across topics remains relatively constant with the exception of two topics whose corresponding opinions exhibit slightly higher amounts of partisanship.

4.2 Peaks in Partisanship

It is not possible to state with certainty what was responsible for the surges in partisanship in 1954, 1966, and 1996; nevertheless, we aim to contextualize these peaks in terms of historical events and relevant demographic information. It appears that these peaks are unlikely to be the result of noise and nothing else; we computed a 99% confidence interval for the average partisanship of each year under the assumption that the distribution of partisanship of opinions for each year is approximately normal (see *Figure 1.1*) – an assumption we found to hold – and we also computed a 99% bootstrap confidence interval with 1000 samples (see *Figure 1.2*) to better understand each year's average partisanship. We found that the distribution of partisanship of opinions for each year did indeed appear to look Normal for a majority of the years, motivating our first approach to constructing confidence intervals. We found that both approaches to constructing confidence intervals yielded nearly identical results and mirrored the behavior of the average partisanship in shape very closely.

1954 saw the largest average partisanship of any year we studied; the Court in 1954 also consisted of six Democrats, one of who – Justice Robert H. Jackson – died prior to the start of 1955, when a very steep decline in partisanship was observed. 1954 also saw the unanimous verdict to drastically expand civil rights afforded to African American students in the landmark case *Brown v. Board of Education*, a case whose opinion itself exhibited a well-above average individual partisanship score of 0.538; additionally, 1954 saw a sharp peak in partisanship amongst just Democratic-authored opinions (see *Figure 2*). The sole Republican Justices on the Court in 1954 were Chief Justice Earl Warren, who was known for his leadership skills and belief that the Court served to uphold the Constitutional rights afforded to people; and Justice Harold Hitz Burton, who was known for being "dispassionate" and "pragmatic," [9], [10], [11]. It may have been at least partly the case that the peak in partisanship amongst Democratic-authored opinions in this year was due to there only being a two Republican Justices concurrently serving, both of which were well-known for characteristics that would suggest neutral writing styles.

1965 saw the resignation of Arthur Goldberg (D), who was replaced by another Democrat, Abe Fortas (we note that the effects of changes in composition made in one year were likely to have been fully realized in the year after) [11]. We note though, that the peak increase in partisanship from 1965 to 1966 was not as great as the increase in partisanship from 1953 to 1954, nor did partisanship drop in 1967 the way it did in 1955. 1966 also saw one of the most well-known and frequently utilized Supreme Court cases in history: *Miranda v. Arizona*, which established what we now know to be our Miranda rights. Partisanship amongst Republican-authored opinions and Democratic-authored opinions was nearly identical in 1966, creating a sharp contrast to what we observed in 1954 (see *Figure 2*). It is possible that the increase in partisanship in 1966 may have been due to the introduction of a new Justice, and the decrease in partisanship the following year may be attributable to this Justice settling in.

2010 saw the resignation of John Paul Stevens (R), who was replaced by Elena Kagan (D), which, again, the effect of which may have been fully realized the following year [11]. Partisanship in 2010 reached its lowest value before peaking in 2011; partisanship in 2012 reached its second-lowest value after peaking in 2012, though no changes at all in Court composition were observed from 2011 to 2012. It is even more difficult to even attempt to attribute the remaining peak in partisanship to changes in Court composition as Court composition remained exactly the same from 1995 to 1997, though we note that in this case, the values of average partisanship in 1995 and 1997 do not differ from the value of average partisanship in 1996 as dramatically as the values of average partisanship in 2010 and 2012 differ from the value of average partisanship in 2011 [11].

4.3 Partisanship by Individual Justice

We find that the average partisanship does not vary much, if at all, between Justices (see *Figure 4*). To arrive at this conclusion, we simply computed the average partisanship for each Justice who had written a minimum of 30 opinions during the time period we studied (median: 317 opinions; mean: 350 opinions). In particular, the average partisanship for each Justice ranged between 0.5109 and 0.5179, and the distribution of the average partisanship for each Justice doesn't appear to take on any particular shape. 90% of Justices possess a partisanship score between 0.5121 and 0.5172; the Justices whose partisanship scores do not lie inside that interval are Justices Jackson, Rutledge, and Rehnquist, with the former two exhibiting partisanship scores lower than 0.5121 and the latter exhibiting a partisanship score above 0.5172.

Justice Jackson was a Democratic Truman appointee. Notably, Justice Jackson engaged in a well-documented, ongoing series of arguments with fellow Justice Hugo Black; the reason for their disagreements was that Justice Jackson believed Justice Black had a habit of relying upon personal and political preferences to form conclusions about cases [12]. Justice Jackson went so far as to even claim that "with few exceptions, we all knew which side of a case Black would vote on when he read the names of the parties," [12]. This bit of background information alone establishes that Justice Jackson was an outspoken advocate of judging cases solely based on their merits; his much lower-than-average partisanship score, correspondingly, makes sense.

Justice Rutledge was a Democratic Roosevelt appointee. He was also a documented member of the Court at the time's "liberal wing," and was remembered as "a devoted champion of civil rights,", [13], [14]. It is more difficult to motivate the reasoning behind Justice Rutledge's lower partisanship score; we are left to assume that, while his positions themselves tended to align with liberal positions, the way through which he expressed those positions did not see the usage of much partisan language. We note that, however, that the tenures of these non-partisan Justices overlapped, so it is more than possible that the writing style of Justice Rutledge influenced that of Justice Jackson, and vice versa.

Justice Rehnquist was a Republican Reagan appointee who served as Chief Justice of the Court from 1986 to 2005 and was considered to be part of the Court at the time's more conservative wing. We provide the following quotation regarding Justice Rehnquist that appears in a biography of him written by John A. Jenkins: "As a private citizen, Rehnquist had protested the Court's decision in Brown v. Board of Education, and as a Justice, consistently ruled against racial minorities in affirmative action cases. Only when white males began to make reverse discrimination claims, did Rehnquist become sympathetic to equal protection arguments," [16]. Additionally, according to Professor Geoffrey Stone of the University of Chicago Law School, "[the only areas] Rehnquist showed any interest in enforcing the constitutional guarantee of free expression: in cases involving advertising, religious expression, and campaign finance regulation," but "Rehnquist voted against freedom of advertising if an advertisement involved birth control or abortion," [15]. With this information in mind, it is not unfathomable to suggest that partisan language would have found its way into opinions penned by Justice Rehnquist.

4.4 Partisanship by Gender of Justices

We also analyze average partisanship over time by gender. We find that while peaks in partisanship between genders roughly coincide in time, partisanship of opinions written by female Justice is consistently and clearly lower than partisanship of opinions written by male Justices at all points in time (see *Figure 3*).

Research has indicated that writing styles may indeed differ between men and women, though perhaps in a manner contradicting the conclusions we have drawn here. Past studies have found that, in general, writings (in academic settings and otherwise) by women tend feature more discussion about "relationships" whereas writings by men tend to feature more discussion about "objects";

and "female writing exhibits greater usage of features identified by previous researchers as "involved" while male writing exhibits greater usage of features which have been identified as "informational"" [17], [18]. Most of the conclusions drawn in the studies we found suggested that women's writings having the tendency to appear more emotionally driven than men's writings in general; intuitively, if these conclusions are to be believed and reasonably extended to our analysis, we would expect opinions authored by female Justices to exhibit a greater degree of partisanship than opinions authored by male Justices. Yet, we find the opposite to be true. It may be the case that the texts in these studies may be too unlike Court opinions for their conclusions to be applied (though we note that one study cited here analyzed academic writing, and there is reason to believe academic writing has much in common with legal writing in general). Even with this in mind, though, we must note that there have only been four female Justices in the entire history of the Supreme Court; the introduction of even a single female Justice in the future could potentially contradict the conclusion we draw here. The confidence intervals for partisanship by gender we display in *Figure 3.1* further emphasize that the lack of adequate sample size of female Court Justices can lead to somewhat unstable results; we note the difference in the length of the intervals computed for female Justices and the length of the intervals computed for male Justices as well the overlap between these intervals.

Future work aimed at resolving this particular issue, which we discuss in more detail in the next section, could involve applying the methods outlined here to a much larger corpus of opinions from other Courts (for example, opinions of the U.S. Court of Appeals) for comparative purposes. We nevertheless look forward to the inclusion of more female Justices in the Supreme Court so we can more substantively analyze differences in partisanship across gender in that Court (amongst other reasons).

4.5 Partisanship by Political Party of Justices

Lastly, we assess differences in partisanship across political parties of the Justices whose political affiliations we were able to identify. We find that, overall, partisanship exhibited by Republican-identified Justices tended to be slightly lower than that exhibited by Democratic-identified Justices overall, and much lower than that exhibited by Democratic-identified Justices from 1953 to 1956 as well as from 2001 to 2008. The confidence intervals displayed in *Figure 2.1* support that these observed differences are unlikely to have come about purely as a result of noise.

The Court's composition from 1953 to mid-1956 was heavily Democratic and saw many changes over the period ranging from late 1956 to mid 1957. [11]. The analysis of the peak in partisanship amongst Democrats from 2001 to 2008 differs with regards to Court composition; during this time period, at least six Republicans were serving the Court, and this number was only reduced in 2009 when Justice Sotomayor (D) replaced Justice Souter (R) [11]. It may be possible that partisan language was employed by Democratic Justices in opinions they wrote to "correct" for the conservative majority ruling the Court as well as put forward liberal interpretations of the corresponding rulings when referenced in future cases and establish liberal-leaning precedent that, like most precedents established by the Supreme Court, would be difficult to overturn. Nevertheless, the fact that clear peaks in partisanship of Democratic Justices were observed during a period when the Court was mostly Democratic *and* during a period when the Court was mostly Republican is certainly interesting.

4.6 Polarizing Phrases

In part to evaluate the reasonable-ness of our model, we highlight eight phrases per feature whose learned coefficients deviated significantly from their initial estimates in *Figure 5*; the eight phrases we highlight were chosen from the 15 phrases whose learned coefficients deviated the most from their initial estimates for each feature to represent clear themes observed in those 15 phrases.

We find that phrases that connote procedural issues and more abstract concepts tend to be most polarizing for the *Gender* feature (e.g. "impartial, trial"; "clear, present"; "testimonial, assertion," etc.); the phrase "exclusion, woman" is also a notable polarizing phrase for the *Gender* feature. On the other hand, phrases corresponding to very specific topics that loosely correspond to topics identified in the next section tend to be most polarizing for the *Political Party* feature (e.g. "finance, docket"; "church, judicatory"; "states, jurisdiction," etc.). Procedural phrases are largely absent from the list of most polarizing phrases for the *Political Party* feature, and we also note that the phrase "partisan, bias" is also identified as a highly polarizing phrase for this feature. Polarizing phrases for the *Seniority* feature consist of a relatively evenly-split mix of procedural phrases (e.g. "direction, answer"; "select, nominee") and phrases corresponding to specific topics (e.g. "race, discrimination"; "california, statute"). One phrase in particular, "neutral, principle," appeared to be polarizing across all three features, possibly suggesting that the notion of neutrality itself is polarizing across a variety of cross-sections.

4.7 Topic Modeling

4.7.1 Emergence of Topics

We will first briefly discuss the six aggregated features and their constituent words as discovered by LSA mentioned in the *Modeling and Approach* section; the weighted word clouds can be seen in *Figure 6.* The the larger the word appears, the greater the weight it was given within its corresponding aggregated feature. Our first LSA-aggregated feature concerns the theme of unions; words like "collective," "bargaining," "labor," "employer," and "union" itself emerge with high weights. Our second aggregated feature concerns the theme of search, seizure, and the Fourth Amendment; words like "fourth," "search," "arrest," and "seizure," appear with high weights. The third feature concerns religion in schools; the fourth feature concerns interstate commerce; the fifth feature concerns deportation and the notion of "illegal aliens"; and the sixth feature concerns Native American reservations and tribal land. Each of these six features presents a clear, human-interpretable theme that we have good reason to expect to feature prominently in Supreme Court cases.

Accordingly, there is much overlap between the aggregated features created by LSA and the clusters identified by k-means. We note that in addition to having understandable topic labels for clusters and, by extension, opinions, we are able to identify separations between clusters and note that the clusters themselves emerge very clearly. For example, the cluster pertaining to Native American land and water sits far away from the cluster pertaining to patents and patent infringement, and the cluster pertaining to sentencing and death sits far away from the cluster pertaining to arbitration. On an intuitive level, these clusters being separated makes sense; it's unlikely that the death penalty has much to do with arbitration, and it's unlikely that Native American land and water have much to do with patent infringement. There is some, but not much, overlap between clusters. We find that the cluster pertaining to lawsuits themselves overlaps with the greatest number of clusters; this is relatively unsurprising because the concepts of procedure and lawsuits themselves are relevant to cases of all kinds, and many cases that one might think would pertain to other topics are resolved on matters of procedure. In general, we find that the topics identified by these clusters intuitively align with what we might expect the Supreme Court to rule about.

4.7.2 Partisanship Across Topics

We find that partisanship across topics remains relatively constant, with the exception that opinions about witnesses, grand juries, and testimony (one topic) as well as opinions about elections and voting exhibit slightly higher average partisanship scores (see *Figure 8*). It is difficult to speculate as to why these opinions about these topics have higher partisanship scores.

Opinions about deportation of aliens and immigration exhibit the lowest average partisanship score by topic. Given that in recent years, language outside the Court surrounding the topic of immigration in particular frequently conveys the political affiliation of the person(s) producing the language, and that immigration has always been a relatively contentious issue, we find this to be surprising. Yet, we are relieved that we do not observe unusually partisan language accompanying cases pertaining to this topic in the language of the Supreme Court, at least for the time being.

5 Limitations and Future Work

There are some limitations to our data and approach that we must discuss here. First, we note that our dataset is much less rich than the one used in Gentzkow et. al — their dataset consists of some 529,980 unique bigrams, 7,254 unique speakers, and 33,373 speaker sessions (each of which consists of multiple speeches) and includes speeches from 1873 to 2016 [7]. For contrast, our dataset consists of 37,733 bigrams, 39 unique Justices, and 12,617 individual opinions, yet we apply much of the same approach. Our analysis of trends in partisanship based on gender is similarly limited based on size constraints; we note that there have only been four female Justices (Justice O'Connor, Justice Ginsburg, Justice Kagan, and Justice Sotomayor) on the Supreme Court, and the first opinion authored by a female Justice came to be in 1981. Furthermore, the number of female Justices appointed to the Court has always been less than the number of male Justices, often by a significant amount, which makes it difficult to substantively assess differences in partisanship across gender, at least for the time being [11]. As mentioned earlier, the addition of even a single female Justice in the future can have the potential to alter our results.

Again due a lack of sufficient data, we were unable to apply very rigorous techniques to reduce the amount of noise in our texts (i.e. we would have liked to have employed higher thresholds for the number of times phrases occurred in each opinion and the number of times phrases occurred across all opinions), the implications of this being that our overall partisanship may have disproportionately been affected by phrases our model was unable to learn very much about. We make an effort to mitigate this undesirable effect by ranking phrases based on the difference between their likelihoods of being written by a Democrat versus a Republican in our partisanship metric with the hope that doing so weights phrases our model was unable to learn much about less.

There are multiple avenues for future work. One such avenue could involve experimenting with the way the text of opinions is featurized — "high-signal" unigrams could be included in addition to the existing bigrams, or a different approach in which pre-trained word embeddings are utilized to featurize the text could be explored. More information about judges, such as their educational background, the President who appointed them, their geographic origin, and other demographic attributes could be encoded and utilized to further enhance the model used here.

To significantly increase the amount of text available for analysis, we propose the following three approaches. The first approach could involve not just considering the opinion section of each Supreme Court case but considering the entire case itself (which we note would require treating concurrences and dissents appearing in the same case but written by Justices who did not write the majority opinion differently); doing so would multiply the amount of text available for analysis by several factors. We also may see the signals extracted from phrases appearing in this larger corpus increase in quality. The second approach could involve extending the analysis to a larger corpus of opinions, such as, for example, opinions from the highly influential U.S. Courts of Appeals, which sit directly under the Supreme Court. There are a total of thirteen U.S. Courts of Appeals defined based on geography and 179 judges presiding over them; the total volume of opinions originating from these Courts will be massive. Additionally, more robust conclusions about the ways through which demographic information of judges factors into partisanship could be formed with this larger sample size of judges. The third approach to increasing the amount of text available for analysis could involve including opinions written by Supreme Court Justices over their entire judicial career as many Justices served on other Courts before ascending to the Supreme Court. Taking this approach would also help us to understand how Justices' writing styles changed (if at all) after they began serving to the Supreme Court both at the individual Justice level as well as across all Justices.

6 Acknowledgements

I would like to thank my advisor, Prof. Zhou Fan, for his guidance, patience, and motivation as I worked on this project. I would also like to thank the Statistics and Data Science department for giving me the opportunity to write this thesis to partially fulfill my degree requirements.

7 Figures

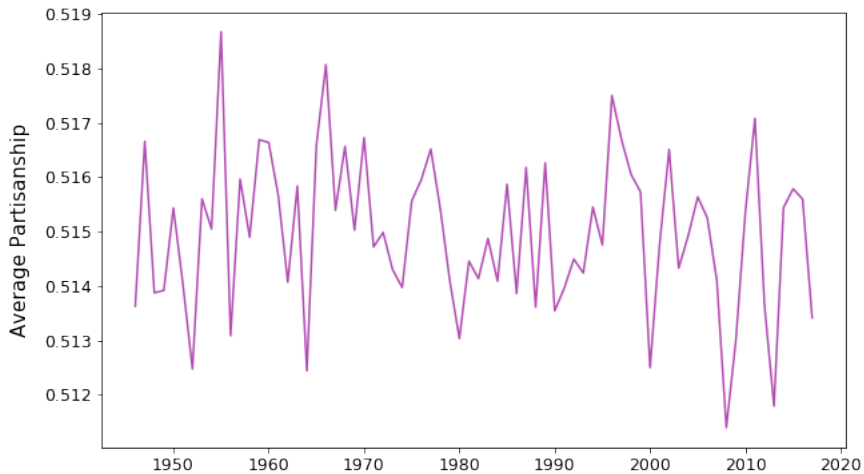

Figure 1: Average Partisanship of Opinions Over Time

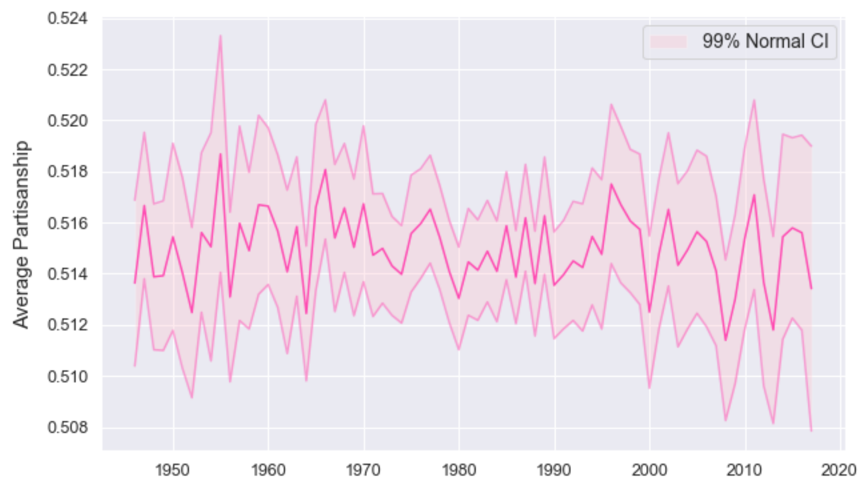

Figure 1.1: Average Partisanship of Opinions Over Time with Normal CI

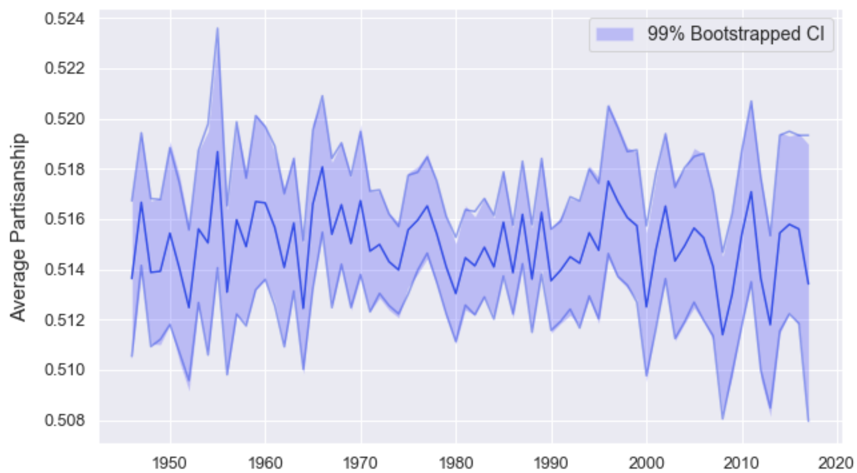

Figure 1.2: Average Partisanship of Opinions Over Time with Bootstrapped CI

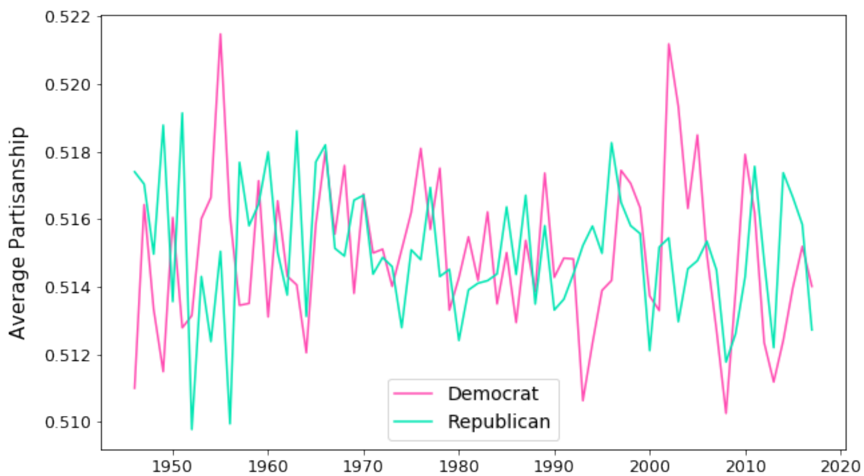

Figure 2: Average Partisanship of Opinions Over Time by Political Affiliation of Justices

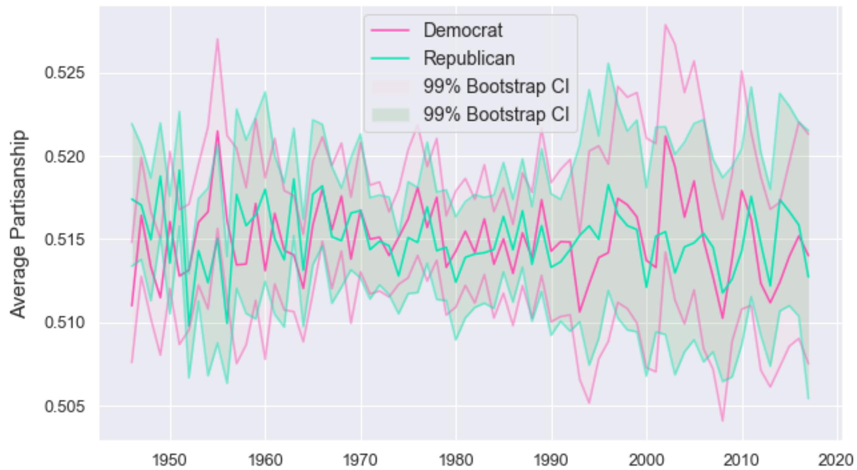

Figure 2.1: Average Partisanship of Opinions Over Time by Political Affiliation of Justices with Bootstrapped CI

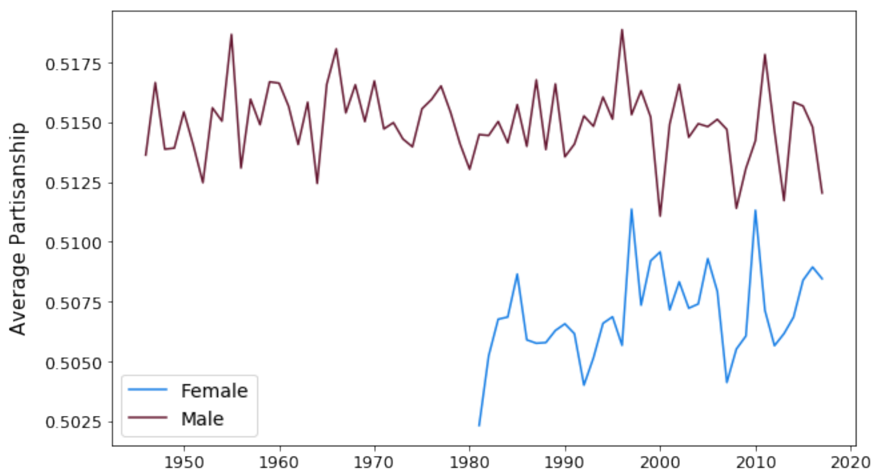

Figure 3: Average Partisanship of Opinions Over Time by Gender of Justices

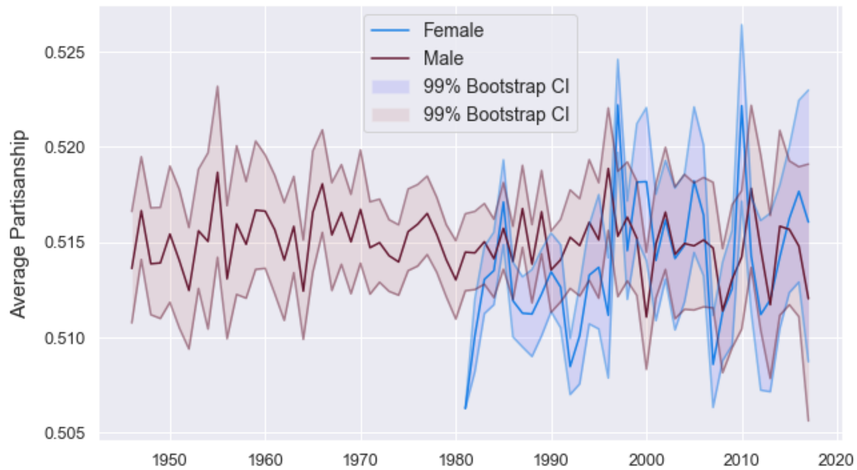

Figure 3.1: Average Partisanship of Opinions Over Time by Gender of Justices with Bootstrapped CI

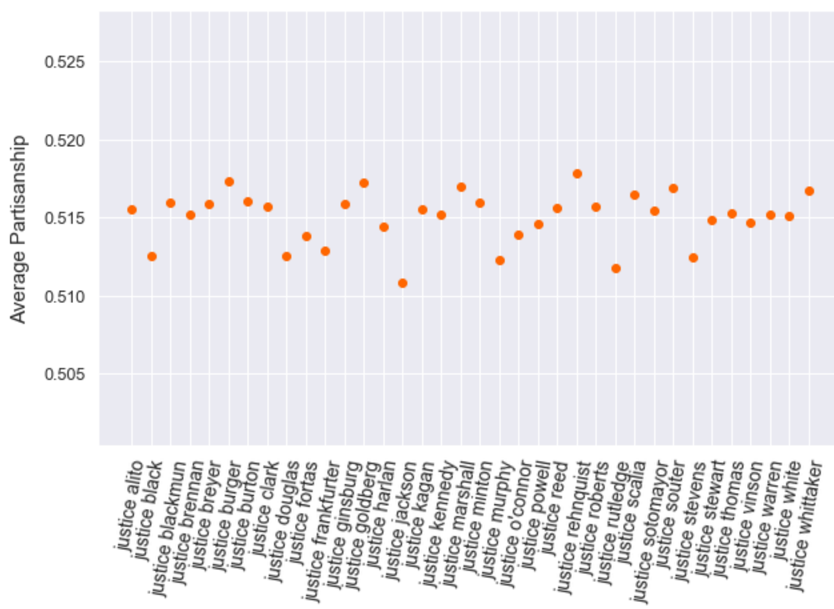

Figure 4: Average Partisanship of Opinions for Individual Justices

Gender	Political Party	Seniority
exclusion, woman	*partisan, bias*	*race, discrimination*
clear, present	*finance, docket*	*human, relation*
impartial, trial	*federal, expenditure*	*federal, expenditure*
support, constitution	*church, judicatory*	*select, nominee*
special, education	*states, jurisdiction*	*california, statute*
testimonial, assertion	*pre, merger*	*direction, answer*
neutral, principle	*neutral, principle*	*neutral, principle*

Figure 5: Selection of Highly Partisan Phrases for Features

Figure 6: Aggregate Topics and their Features as Created by LSA

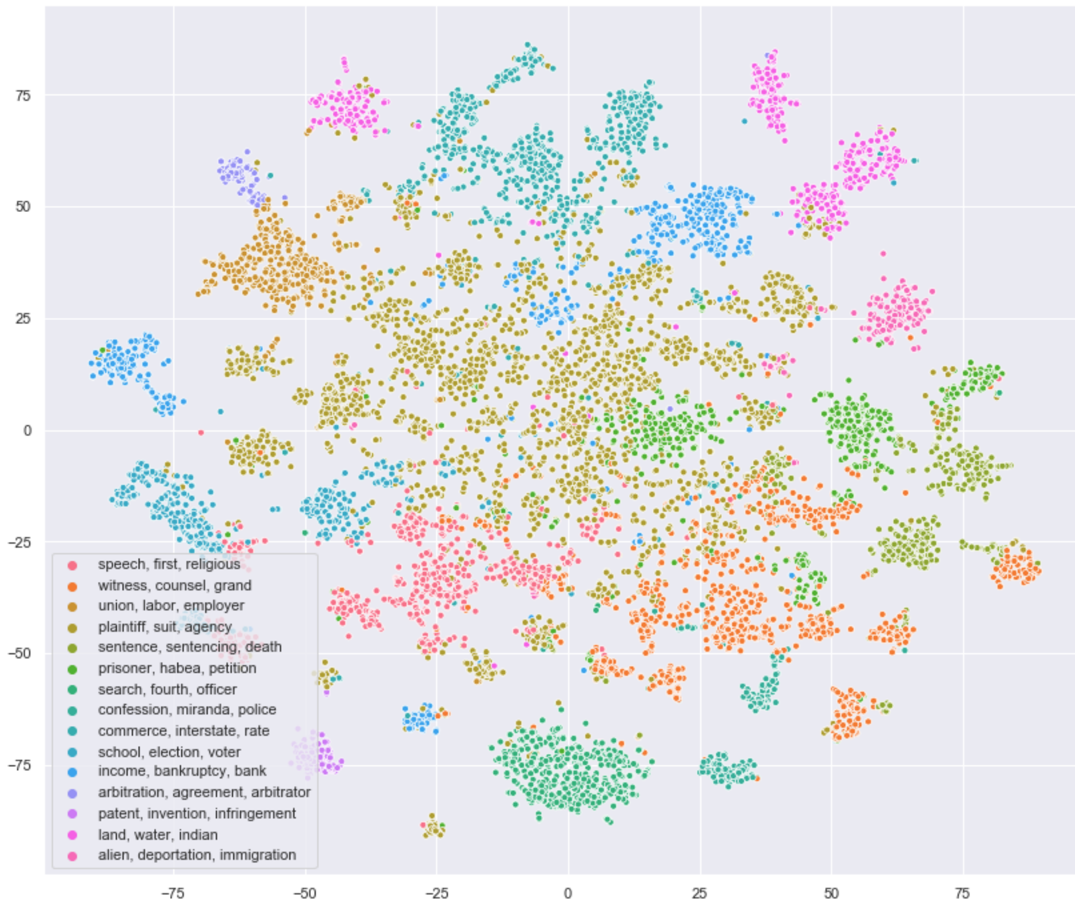

Figure 7: Computed Clusters of Topics of Opinions

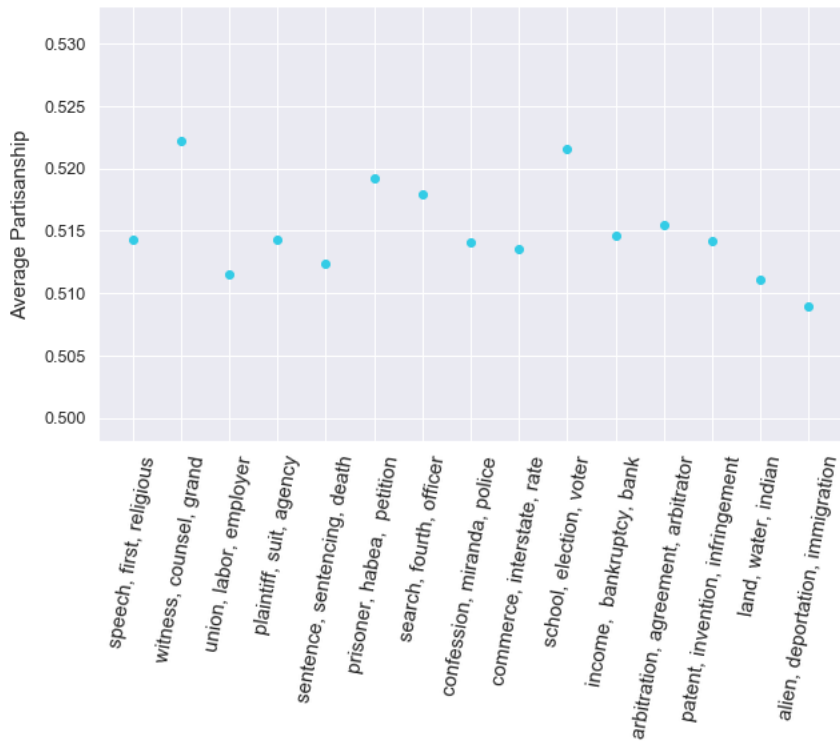

Figure 8: Partisanship of Opinions by Topic

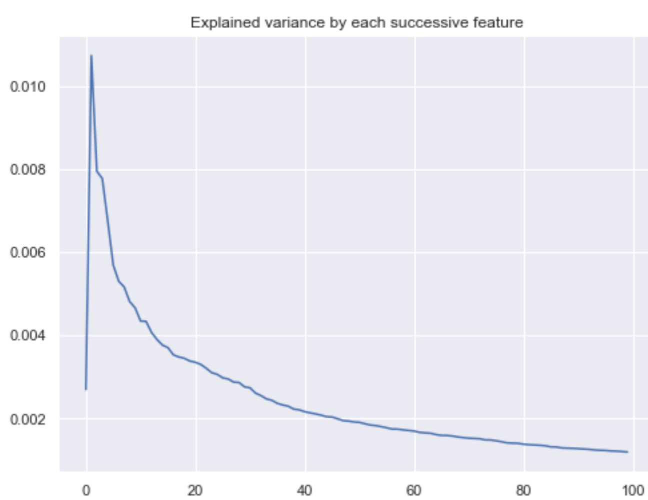

Figure 9: Amount of Variance Explained by Addition of Successive Features in SVD

References

[1] Tom Kertscher. *What Ruth Bader Ginsburg said about Donald Trump.* https://www.politifact.com/article/2016/jul/13/what-ruth-bader-ginsburg-said-about-donald-trump/. September 24, 2019.

[2] Andrew Chung *U.S. Supreme Court not politicized, says Chief Justice Roberts.* https://www.reuters.com/article/us-usa-court-chiefjustice/u-s-supreme-court-not-politicized-says-chief-justice-roberts-idUSKBN1WA08F. July 13, 2016.

[3] Donald Trump. *Twitter.* https://twitter.com/realdonaldtrump/status/978932860307505153?lang=en. March 28, 2018.

[4] Adam Liptak. *John Roberts Criticized Supreme Court Confirmation Process, Before There Was a Vacancy.* https://www.nytimes.com/2016/03/22/us/politics/john-roberts-criticized-supreme-court-confirmation-process-before-there-was-a-vacancy.html. March 21, 2016.

[5] United States Congressional Records. *Confirmation Hearing on the Nomination of Hon. Neil M. Gorsuch to be an Associate Justice of the Supreme Court of the United States.* https://www.congress.gov/115/chrg/shrg28638/CHRG-115shrg28638.htm. March 2017.

[6] Washington University in St. Louis. *The Supreme Court Database.* http://scdb.wustl.edu/data.php?s=1.

[7] Matthew Gentzkow, Jesse M. Shapiro, and Matt Taddy. *Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech.* https://siepr.stanford.edu/sites/default/files/publications/16-028.pdf. July 2016.

[8] scikit-learn. *sklearn.decomposition.TruncatedSVD*.
https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html.
2019.

[9] Melvin I. Urofsky. *The Warren Court: Justices, Rulings, and Legacy*. 2001.

[10] United States Supreme Court. *Proceedings of the Bar and Officers of the Supreme Court of the United States: May 24, 1965*. May 24, 1965.

[11] United States Supreme Court. *Justices 1789 to Present*.
https://www.supremecourt.gov/about/members_text.aspx.

[12] Dennis J. Hutchison. *The Black-Jackson Feud*. 1988.

[13] Ball, Howard. *Hugo L. Black: Cold Steel Warrior*. 1988.

[14] The New York Times. *Obituary of Justice Rutledge*.
https://timesmachine.nytimes.com/timesmachine/1949/09/12/84220509.html.
September 12, 1949.

[15] University of Chicago Law School. *Stone Says Rehnquist's Legacy Doesn't Measure Up*.
https://www.law.uchicago.edu/news/stone-says-rehnquists-legacy-doesnt-measure.
September 12, 1949.

[16] John A. Jenkins. *The Partisan: The Life of William Rehnquist*. 2012.

[17] Shlomo Argamon, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni. *Gender, Genre, and Writing Style in Formal Written Texts*.
http://writerunboxed.com/wp-content/uploads/2007/10/male-female-text-final.pdf.
2007.

[18] Deborah Tannen. *Gender differences in topical coherence: Creating involvement in best friends' talk*. 1990.