Thesis – MSc. Applied Data Science

Utrecht University

Department of Information and Computer Science

Comparing Political Ideology Scaling Methods: A Study of Topic Polarization in the U.S. House using W-NOMINATE and Correspondence Analysis

Candidate

Dion Custers (0029106)

First Examiner

Dr. Javier Garcia-Bernardo

Second Examiner

Elena Candellone

In Cooperation With:

Esther M.A.B. Stehouwer

Lars Teunissen

June 29, 2024

Acknowledgements

I want to extend my thanks to everyone who supported me during the completion of this research. A special thanks to Javier Garcia-Bernardo for his constant assistance and optimism during the project and for helping me establish my final concept. I also want to thank Elena Candellone for the all the time and effort that she spent on assisting me through the process, and for the valuable feedback that greatly improved this final product. Lastly, my sincere thanks to Esther Stehouwer and Lars Teunissen for their great cooperation in the early stages and for continuing to offer valuable insights during the individual phase of our work.

Abstract

Growing concern among scholars about increasing polarization in liberal democracies has led to a surge in research focused on the presence and measurements of political polarization in the United States. Due to party polarization varying significantly across political issues, and with the complexity of measurement techniques seeing an escalation, the subsequent skill threshold for studying this phenomenon has been increasing. This research explores the use of Correspondence Analysis as a potentially less complex alternative for measuring the ideology of legislators by answering two research questions. First, this study identifies trends of party polarization across broad policy domains using the standard NOMINATE method on the voting behavior of U.S. House Representatives. Next, it evaluates whether Correspondence Analysis on the positive votes of these legislators can replicate the identified trends. My results demonstrate equal polarization trends across most policy domains, with only one domain deviating significantly. However, we find that the principal dimension of the Correspondence Analysis does not capture these polarization trends effectively. Although Correspondence Analysis does show clustering in its ideological spatial dimensions, these clusters do not align with the expected party identifications. In conclusion, this research indicates that a less complex model based on the positive voting behavior of U.S. House Representatives does not accurately reflect their ideological positions within specific policy domains. The study proposes further investigation into alternative, less complex methodologies, and smaller-scale research topics, as we argue that reducing the research barrier for scholars engaged in this political phenomenon represents a necessary advancement.

Keywords: Political polarization; U.S. Congress; Topic modelling; NOMINATE; Correspondence Analysis.

Table of Contents

1. Introduction

The rise of far-right populism in liberal democracies such as the Netherlands, France, Argentina, and the United States (Campani et al., 2022; Economist, 2024; Vohra, 2024) has led scholars to become increasingly concerned about growing polarization within democratic countries worldwide (McCoy, 2018). Political polarization can have numerous negative repercussions for both society and democracy, including diminishing trust in governments and politicians (Citrin & Stoker, 2018; Lee, 2022), eroding social cohesiveness (Capshaw, 2005), and undermining the state's democratic system. Theories about the causes of polarization are common in the literature of political science, with branches of this debate questioning its existence or the approaches through which it should be measured (Barber & McCarty, 2013). Recent methods for determining polarization have become increasingly complex, accounting for a broader range of factors, such as multidimensional ideological characteristics (Bateman et al., 2017) and networks of co-sponsorship and co-committee memberships (Neal, 2020). Given the increasing complexity of these recent methodologies, this research seeks to determine if a simpler approach could provide equally strong insights into the trends of political polarization.

1.1 Literature Review

A review of previous research on political polarization reveals a variety of interpretations of this phenomenon. Adapting Neal's (2020) definition, polarization refers to the existence of groups differentiated by one or more characteristics. For political polarization, these characteristics may include political ideology or party affiliation (ibid.). Based on the actors involved, political polarization can be divided into elite polarization, among politicians or legislators, and mass polarization, occurring among the electorate (Ibid.). Regarding the causes of political polarization, two distinct forms are further identified (Iyengar et al., 2019). Affective polarization refers to the phenomenon where individuals' emotional responses to those with similar ideologies become more positive while their feelings toward those with opposing ideologies become more negative. This leads to both a hostile attitude toward others and a decrease in willingness to compromise (ibid.). Ideological polarization omits emotions and refers solely to the extent to which the electorate or legislators hold divergent views on ideological issues and beliefs across a spectrum of policy positions (ibid.).

The polarization of political elites has long been a topic of discussion in political science, with polarisation within the U.S. Congress showing an upward trend since the 1970s (Barber and McCarty, 2013; Jochim & Jones, 2012; Poole & Rosenthal, 1997). Multiple methods for calculating polarization are discussed in the literature. One branch of studies focuses on measuring polarization through the co-behaviour of legislators, ranging from analyses of co-bill sponsorship (Fowler, 2006; Neal, 2020; Zhang et al., 2008), co-voting on bills (Andris et al., 2015), co-membership on committees (Porter et al., 2005), and co-attendance at events (Desmarais et al., 2015), interpreting cooperation in all cases as indicative of a positive relationship (Kirkland & Gross, 2014; Neal, 2020). Although co-sponsorship is frequently used in recent studies, a more prevalent measure of polarization involves analysing the voting history of legislators to assess their political ideology (Barber and McCarty, 2013; Hare & Poole, 2014). The measurement of ideology sees a wide range of methods. In their study, Tausanovitch and Warshaw (2017) reviewed six recent approaches: social media followings (Barberá et al., 2015), campaign donor networks (Bonica, 2014), perceptions of survey respondents (Aldrich & McKelvey, 1977; Hare et al., 2015; Ramey, 2016), expert assessments (Joesten & Stone, 2014; Maestas et al., 2014; Stone & Simas, 2010), and analyses of political texts (Laver et al., 2003; Monroe et al., 2017; Slapin & Proksch, 2008). Among these, however, analysing voting records remains one of the primary methods (Bonica, 2014; Neal, 2020).

While various models allow us to estimate legislators' ideological location based on their roll-call votes, such as the IDEAL method (Carroll et al., 2009; Clinton et al., 2004), the standard remains the NOMINATE method, short for Nominal Three-Step Estimation. Introduced by Poole and Rosenthal in 1985, this method has further developed into the Weighted NOMINATE (W-NOMINATE) and Dynamic Weighted NOMINATE (DW-NOMINATE) (McCarty et al., 1997; Poole & Rosenthal, 1985). NOMINATE places legislators on a two-dimensional spatial map based on their roll call voting behaviour, thereby positioning ideologically similar legislators closer together (Boche et al., 2018; Neal, 2020). The first dimension typically represents the liberal (-1) to conservative (1) spectrum, while the second dimension has varied over time to encompass issues such as race and civil rights (Poole & Rosenthal, 2011). This method has illustrated a rising trend in polarization within the U.S. Congress from the 1970s to the present (Poole & Rosenthal, 2001; Barber & McCarty, 2013; Hare & Poole, 2014; McCarty et al., 2016).

Despite its widespread use and wide recognition as a reliable method of measuring legislator ideology (Bonica, 2014; Carroll et al., 2009), NOMINATE has faced some criticism. Some argue that it overestimates polarization levels (Carson et al., 2010; Egar, 2016) and oversimplifies ideology by focusing predominantly on the left-right policy dimension (Aldrich et al., 2017; Bateman et al., 2017; Crespin & Rohde, 2010; Neal, 2020). Moreover, categorizing members into two blocks may give the impression that both parties are equally polarized on all issues. For instance, Bateman et al. (2017) observed a decrease in polarization on civil rights issues. Additionally, an analysis of NOMINATE scores for specific topics reveals significant variations in polarization across different congressional sessions (Marchi et al., 2021). Another study highlighted that different topics exhibit considerable variability in their dimensional structure (Jochim & Jones, 2012).

While the primary criticism of the NOMINATE method centres on its tendency to oversimplify legislators' ideological values and the measurement of polarization, the literature contains even less complex alternatives. For example, Barberá et al. (2015) utilize Correspondence Analysis (CA) to determine the ideological positions of social media users, as this method reduces computational costs for large-scale networks while still providing reliable results. As a statistical technique, Correspondence Analysis (CA) can visualize relationships between categorical variables by using a contingency table and transforming this data into a low-dimensional space. This makes CA one of the simpler methods to execute and interpret, theoretically requiring only information on a single variable (Abdi & Béra, 2017). However, in the literature on ideology and polarization measurement, CA and its relative Multiple Correspondence Analysis (MCA) are still only primarily applied to texts, videos, or ordinary citizens (Guenduez et al., 2016; Lai et al., 2024; McLay & Ramos, 2021; Sønvisen, 2014; Wiesehomeier & Doyle, 2012).

1.2 Research Aim

Based on the aim to explore whether simplifying polarization measurement can provide equally strong insights into polarization trends, this paper employs the Correspondence Analysis method and compares its results with those received from the NOMINATE method. While NOMINATE considers the entire voting history of a legislator, Correspondence Analysis simplifies this process by focusing solely on positive votes to determine its values. Taking into account that research indicates that polarization can vary by topic, this study investigates whether CA values can adequately represent polarization trends within different policy domains. This paper will thus first address the question of what trends of polarization are found within different broad policy topics, followed by whether Correspondence Analysis, using only positive voting behaviour, can sufficiently replicate these trends. The analysis will focus on the U.S. House of Representatives from 1973 to 2024, i.e., the members of the 93rd to the 118th Congress. In this paper, I answer the research question by categorizing the votes into various policy domains using a Latent Dirichlet Allocation (LDA) topic model. Following, I measure the ideology of House Representatives using the W-NOMINATE and Correspondence Analysis methods, after which the results will be compared using the Pearson Correlation Coefficient. Before presenting the results, a brief overview of the data and methods are given, including data exploration and the necessary preparation for analysis. The methodology section further provides a detailed explanation of the chosen approaches and their rationale. The following results section presents details on the seven selected topics and their subsets, followed by the NOMINATE results, demonstrating a consistent upward trend of polarization across all seven topics. Afterward, the CA results are presented, revealing vastly different polarization trends. Lastly, the legislator positions of both methods are checked on correlation, thereby only showing moderate correlation for the topic of 'Legislation and Policy.' In the final discussion and conclusion section, I address the research question and give a summary of all findings. Additionally, I discuss the encountered limitations and propose some suggestions for future research on CA and topic-based polarization.

2. Data

2.1 Description

This paper uses a combination of data on U.S. legislators and legislative bills from Voteview and ProPublica (Lewis et al., 2021; ProPublica, 2024) to investigate the phenomenon of topic polarization within the U.S. Congress. The study focuses exclusively on the 93rd to 118th House of Representatives and their voting behaviour on legislative bills, spanning the period of 1973 to 2024. As each House member represents a district within a state, this results in the representation of Democratic representatives in Republican states and vice versa. This leads to a wider variety of Democratic and Republican members, with the inclusion of conservative Democrats and progressive Republicans, which in turn can make indications of party polarization more robust. The time frame for study spans from the 93rd Congress to the 118th Congress, selected due to the availability of more extensive information on legislative bills, with especially summaries and sponsors not being digitally accessible for congresses before the 93rd. Voteview is a project designed to facilitate public access to historical and contemporary data on roll call voting within the United States Congress, accompanied by ideological scores calculated using the DW-NOMINATE method (Lewis et al., 2021). The project was initially developed by Poole and Rosenthal in 1989 and subsequently enhanced in 2018 by Boche et al. to create a more accessible and interactive website for its users. From the Voteview project, three datasets were utilized to conduct this research. The first set contains information about the U.S. House members, the second set about the roll calls in the U.S. House, and the third set contains information on every vote taken by every House member. As the original roll call dataset lacks sufficient information on every legislative bill for adequate topic modelling, supplementary data was sourced from a ProPublica dataset containing comprehensive data on U.S. congress bills (ProPublica, 2024). It is important to note that this data was extracted from the official Congress.gov website (Congress.gov, 2024a), where every bill from the 93rd Congress has been digitized and can be accessed freely.

2.2 Exploration

The original three datasets contain 11,063 unique House Representatives from 52 political parties. Of these, 4,778 are Democrats and 4,057 are Republicans. Additionally, it can be observed that 878 members have changed political parties during their tenure in office, with Table 1 illustrating the top 12 parties that have gained the highest number of new members from other parties over the years. The data reveals that the Democratic Party received its highest number of new members from the old Jackson Party, with the Republican Party's biggest donor being the Opposition Party, which traces back to the old Whig and Adams Parties. Furthermore, Figure 1 illustrates the composition of previous Congresses, which used to consist of a variety of different parties. This composition appears to have changed around the 40th Congress, with the Republican (red) and Democratic (blue) parties remaining the top two in the subsequent Congresses. Further information on the other party lines is found in Appendix A.

Total	New Party	Old Party	Members
180	*Democratic Party*	Jackson Party	94
		Republican Party	20
161	*Republican Party*	Opposition Party	46
		Democratic Party	32
104	*Whig Party*	Anti-Jacksonians	60
		Anti-Masonic Party	12
100	*Jackson Party*	Jackson Republican	43
		Crawford Republican Party	26
79	*Adams Party*	Adams-Clay Republican Party	41
		Adams-Clay Federalist Party	10
79	*Anti-Jacksonians*	Adams Party	52
		Jackson Party	16
59	*Democratic-Republican Party*	Anti-Administration Party	45
		Federalist Party	12
38	*Adams-Clay Republican Party*	Democratic-Republican Party	38
37	*Federalist Party*	Pro-Administration Party	34
		Democratic-Republican Party	2
37	*Jackson Republican*	Democratic-Republican Party	36
		Federalist Party	1
37	*Crawford Republican Party*	Democratic-Republican Party	36
		Federalist Party	1
37	*Opposition Party*	Whig Party	33
		Democratic Party	2

Table 1. Top 12 Changes in Party Affiliation Among U.S. House Representatives, 1789-2024.

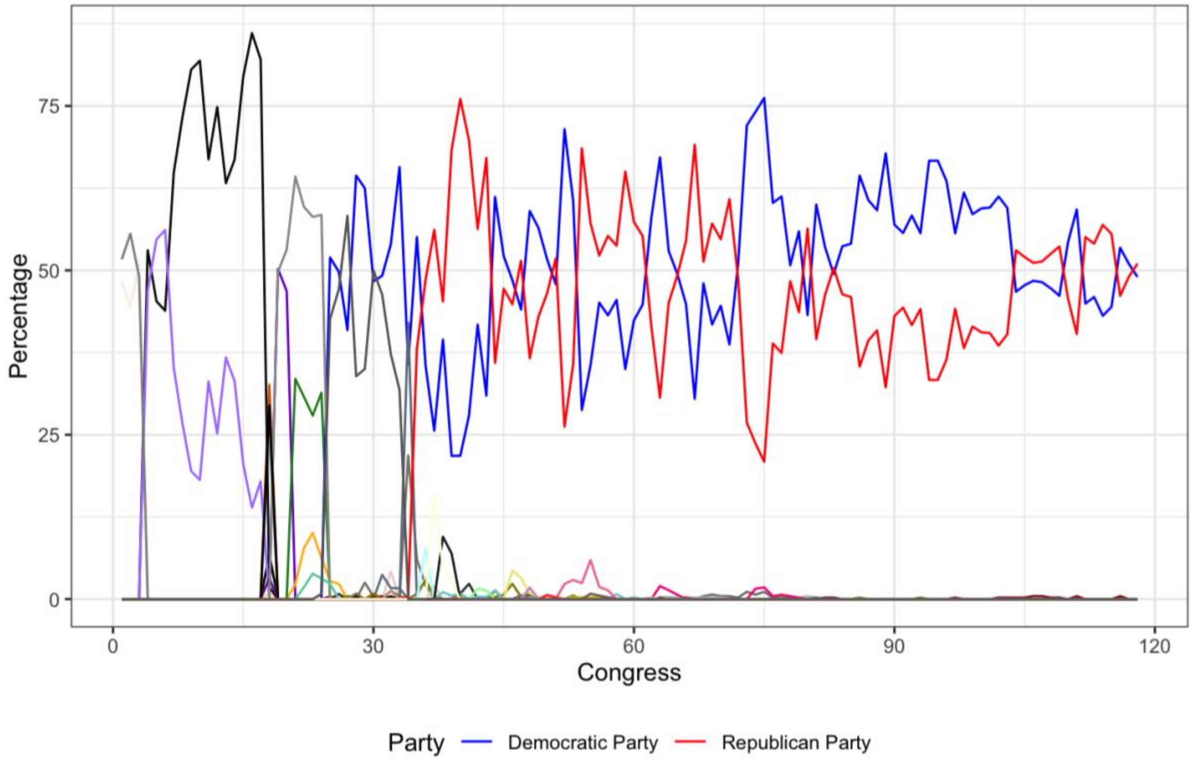


Figure 1. Evolution of Party Representation in the U.S. House of Representatives, 1789-2024.

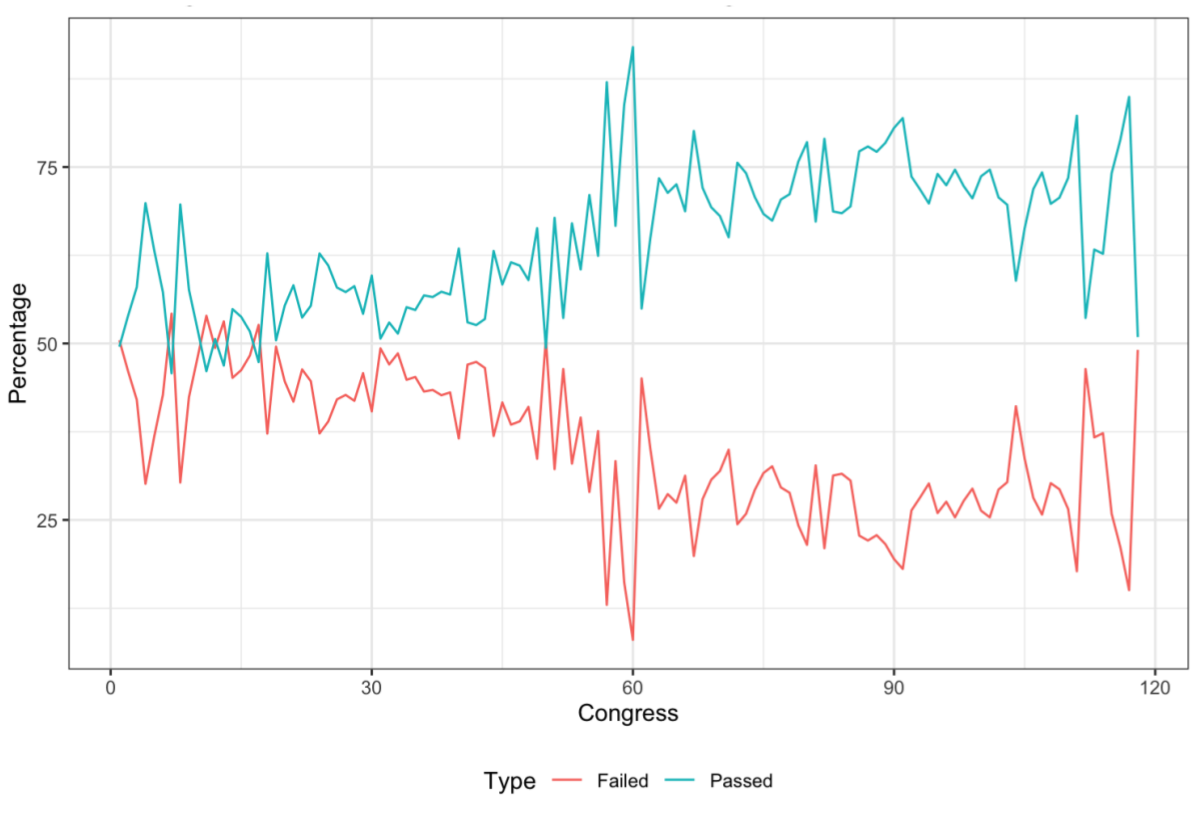


Figure 2. Percentage of Passed and Failed votes in U.S. House Congresses, 1789-2024.

Looking at the roll calls, of which there are 58,779, these are split into multiple different kinds of votes, with the main types being on the agreements to amendments or resolutions and the passage of legislation. However motions for suspension of the rules, adjourning, recommitting and quorum calls also make up a large part of the roll calls, which on their own are more about processes within congress and do not say much or anything about the actual content of a bill. Further, Figure 2 shows an increase in the number of passed roll calls after the 50th congress, which appears to align with the ascendance of the Democratic and Republican Parties.

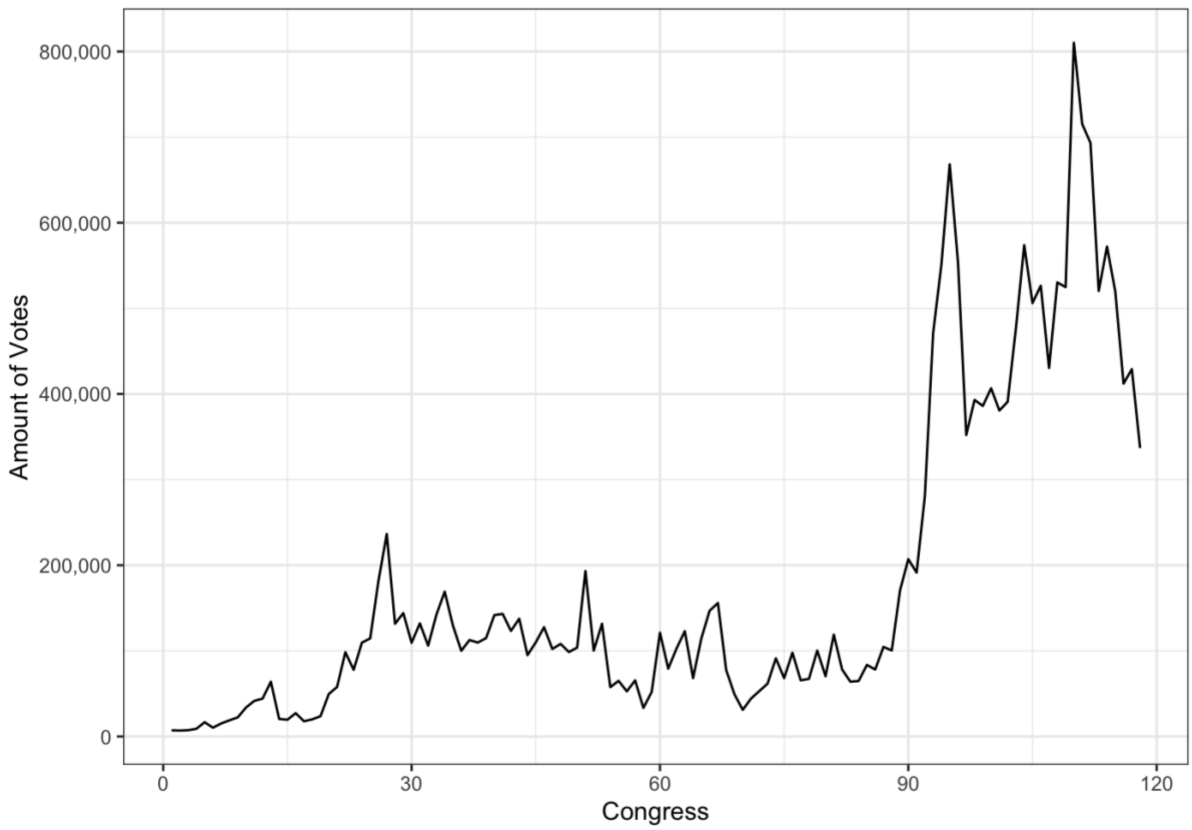


Figure 3. Number of Votes per Congress (1st to 118th)

Lastly, an examination of the roll call votes per Congress in Figure 3 reveals a notable increase in the number of votes cast per Congress starting around the 93rd. This top Congress received over 800,000 votes, aligning with the introduction of electronic voting in 1973 (Roberts, 2007). Figure 3 further illustrates that this quantity has increased over the past few congresses. However, as the most recent Congress has not yet concluded, this amount may still fluctuate. With a total of 21.36 million votes cast, the votes from the 93rd-118th Congresses represent 61.5% of all votes, with 13.1 million votes taken by the House Representatives within this period.

2.3 Preparation

As previously mentioned, each original dataset was first filtered to only include members of the House of Representatives from the 93rd to the present 118th Congress. The original roll call dataset contained minimal information about each legislative bill, such as the title or the roll call question, leading to poor performance in topic modelling. To address this, we included additional information for each roll call based on the ProPublica dataset (ProPublica, 2024).

Specifically, we added a detailed summary of each roll call, thereby expanding the lexicon of each bill and enhancing the performance. To evaluate the coherence of the derived topics, we also incorporated policy information for each roll call. This policy information consists of one of 32 policy area terms assigned to every bill by the Congressional Research Service, which describes the predominant subject matter of the measure (Congress.gov, 2024b). Given the importance of the summary for topic modelling, we excluded 1,818 roll calls that lacked a summary in the bulk dataset. Consequently, we removed quorum calls, motions on adjourning, and motions to suspend the rules, resulting in a dataset comprising only legislative bills. The final roll call dataset includes 28,449 bills from the 93rd Congress in 1973 to the 118th Congress in 2024. Following the initial filtering of the original dataset of House of Representative members, each member was assigned an ID based on their name and party affiliation. This step was necessary due to the ID provided by Voteview, called ICPSR, which is supposed to be based on a unique name and party combination, containing errors that affected the calculation of members' ideological values. The final member dataset comprises 11,580 unique rows, representing 2,217 unique individuals and 2,246 unique person-party combinations, accounting for members who changed parties during their tenure. Lastly, to facilitate for the Correspondence Analyses, we added a binary variable to the original member votes dataset, assigning 1 for a yea vote and 0 otherwise. The final votes dataset ultimately contains 12,343,235 member votes.

3 Methods

The aim of this research is to identify polarization trends within various policy domains and to determine whether Correspondence Analysis, using only positive voting behaviour, can accurately replicate these patterns. To achieve this, the research is structured into three main sections. The first section focuses on identifying polarization trends using the NOMINATE method. Initially, we employ topic modelling to analyse the bills and categorize them into different topics. Subsequently, we use the W-NOMINATE scaling method to calculate the ideological values for each topic based on a subset of related bills. We then determine the party polarization within each topic by averaging the ideological values of representatives per party per Congress. These values are plotted to visualize the party polarization trends for each topic. The second section addresses identifying the same party polarization trends per topic, this time using Correspondence Analysis. We begin by applying Correspondence Analysis to the positive votes of House Representatives, followed by determining the party polarization by averaging the representatives' ideological values per party per Congress and plotting the resulting averages. Additionally, we measure the bimodality coefficient of the NOMINATE and CA data points to check for the presence of clusters within the data should the plots appear uninformative. Finally, the third section involves comparing the results of both methods. We use the Pearson Correlation Coefficient to compare the data coordinates obtained from the W-NOMINATE method and Correspondence Analysis.

3.1 Topic Modelling

The first section of this research involves classifying the content of bills in our dataset and ideologically scaling the members of the House of Representatives based on their voting behaviour. Given the large number of bills and their substantial content, manual classification is not feasible. Although new bills are originally labelled into 32 different topics by the Congressional Research Service (Congress.gov, 2024b), using these labels is not an option due to significant variations in topic distribution, with smaller topics presenting challenges for the NOMINATE ideological scaling method. Additionally, these 32 topics are used only in the latter part of our research period, complicating the classification of older bills. For this reason, we employ topic modelling in this study. This unsupervised machine-learning approach is capable of extracting topics from large volumes of text, with this research in particular utilizing a Latent Dirichlet Allocation (LDA) model (Blei et al., 2003). While previous research has employed various other topic models for legislative and political texts, including Structural Topic Models (STM) and BERTopic, with some yielding positive results (Ebeling et al., 2023; Gerrish & Blei, 2011; Marchi et al., 2021; Wang et al., 2013), attempts to utilize STM and BERTopic models for this research produced only disappointing results. BERTopic, in particular, generated a multitude of small topics, with the reduction of the number of topics resulting in over half of the bills being labelled as outliers without a designated topic. Attempts to enhance these results by incorporating sentence embedding methods, such as LegalBert (Chalkidis et al., 2020), and implementing outlier reduction techniques, were unsuccessful. In contrast, the LDA model produced topics of similar size that covered most of the bills, with clearer distinctions between topics based on their top words.

The general concept of the LDA model is that each document is comprised of a series of latent topics, which are described by a specific probability distribution, with each document being a random mixture of these latent topics (Jelodar et al., 2019). Each topic is further represented by a probability distribution over certain words. The model then infers the latent topics by iteratively assigning words to topics based on their co-occurrence patterns and updates the topic distribution to best fit the observed data (Chen et al., 2016). The document probability distribution over each topic indicates the degree to which a topic is present in a document. This allows for the identification of the primary topics discussed in a document and the inferred topic distributions (Tong & Zhang, 2016). In this paper, the number of topics to be identified by LDA is determined based on the average cosine similarity between the inferred topic distributions of the bills and a ground truth overall topic distribution (Blei et al., 2003). The ground truth topic distribution was calculated using the original 32 policy domains, thereby measuring the word probability distributions for each document and calculating the average probability distribution across all documents, representing the overall topic distribution in the corpus (ibid.). The average cosine similarity was then calculated by taking the mean of the cosine similarities between the ground truth, A, and the inferred topic distributions, B, using the following formula:

$$Cosine\ Similarity = \frac{A \cdot B}{\|A\| \|B\|}$$

The LDA topic model in this research is trained and performed on the full summaries of 28,499 legislative bills. Before executing the LDA model, the summaries undergo pre-processing to ensure optimal results. The pre-processing steps include tokenizing the documents and converting all tokens to lowercase. Subsequently, numbers, punctuation, and stop words are removed, with the latter category including high-influence words added based on their excess presence across all bills. Next, the final corpus is created using Term Frequency-Inverse Document Frequency (TF-IDF), which highlights the importance of words in documents by adjusting for the fact that some words appear more frequently in general (Simha, 2021). After extracting and labelling the topics, their quality is assessed using the Cv Coherence score (Röder et al., 2015; Syed & Spruit, 2017). This measure consists of four parts: first is the segmentation of the data into word pairs using a sliding window approach. In this paper, a Boolean sliding window of 110 words is used, meaning that for each target word, the context of 110 words before and 110 words after is considered (Řehůřek & Sojka, 2010; Röder et al., 2015). After segmenting the data, the probabilities of words or word pairs are calculated based on a given reference corpus. Following this, a confirmation measure is calculated that quantifies the degree to which the presence of one word set supports the presence of another. The final step is the aggregation of these confirmation measures into an overall single coherence score (Röder et al., 2015; Syed & Spruit, 2017). The Cv coherence score indicates the extent to which the top words within a topic co-occur frequently and consistently across the documents in the corpus. Lower scores suggest that many documents within a topic do not share a similar policy domain. Cv coherence scores above 0.5 are considered good, scores between 0.4 and 0.5 are adequate, and scores below 0.4 are deemed poor (Röder et al., 2015; Stevens et al., 2012). The final step is the division of the complete bill and vote datasets into their respective topical subsets.

3.2 NOMINATE

The next step involves the ideological scaling of the members of the House of Representatives and the calculation of party polarization. In this research, this is initially executed using the NOMINATE method developed by Poole and Rosenthal (1985), specifically the W-NOMINATE method (Poole & Rosenthal, 1997). As discussed in the literature review, there are three alterations of the NOMINATE procedure, all designed to analyse the roll call voting records of politicians and create an ideological spatial map of these legislators. The original NOMINATE method uses a probabilistic model to analyse roll call votes, calculating the positions of legislators by evaluating the probability of a legislator voting *yea* or *nay* on a bill based on their and the bill's ideological location (Carroll et al., 2009; Poole & Rosenthal, 1985; Lewis & Poole, 2004). The underlying assumption is that the likelihood of a *yea* vote increases as the distance between the legislator's ideal point and the bill's position decreases. Specifically, legislator i's utility for outcome y, which represents a *yea* vote, on bill j, is given by:

$$U_{ijx} = \beta exp\left[-\frac{\sum_{k=1}^{s} w_k^2\, d_{ijyk}^2}{2}\right] + \epsilon_{ijy}$$

Here, d_{ijyk}^2 represents the Euclidean distance between the ideal point of a legislator x_i and the location of the *yea* bill z_{jyk}, with k indicating the dimension (Poole & Rosenthal, 1997; Lewis & Poole, 2004). Additionally, β and w are positive parameters/weights that, along with the locations of the legislator and bill, are estimated using a constrained nonlinear maximum likelihood procedure. This procedure iteratively adjusts the ideological positions of the legislator and the bill to maximize the likelihood of the observed voting patterns, thereby minimizing the discrepancy between the predicted and actual votes (ibid.).

While previous research has utilized DW-NOMINATE scores published by Voteview (Barber & McCarty, 2013; Cameron & Park, 2009; Lindgren & Southwell, 2014), this study requires a different approach as it aims to compute the ideological values of legislators for specific topics rather than for all bills collectively. For this reason, this research cannot use the published DW-NOMINATE scores, as these are calculated based on the entire voting history of a legislator (Lewis et al, 2021). Due to W-NOMINATE being the only NOMINATE method with existing implementations in R or Python (Poole et al., 2024), this method is chosen for this research, as the manual implementation of this measure is beyond the scope of this paper. Unlike the original NOMINATE method, W-NOMINATE assigns weights to votes based on their distribution, incorporating the ideological significance of important votes (Poole, 2005). Additionally, due to it being a static algorithm, it constrains legislator and bill ideal points to lie between -1 and 1 in every dimension (Everson et al., 2016; Poole & Rosenthal, 1997). Although DW-NOMINATE further accounts for the ideological shifts of legislators over time, research suggests that legislators' ideological positions are relatively stable during their tenure, making W-NOMINATE a sufficient fit for this study's objectives (Jochim & Jones, 2012).

The NOMINATE algorithm is designed to calculate the ideological position of a legislator. However, to orient the results such that 1 represents the right and -1 the left, it requires a conservative legislator per dimension as input (Poole et al., 2007). For this study, which focuses solely on the first dimension, Rep. Sensenbrenner is selected based on his participation in 21 of our 26 relevant congresses and his conservative DW-NOMINATE score of 0.638 (Lewis et al., 2021). Additionally, NOMINATE applies two minimum thresholds for calculating a legislator's ideological position. The first threshold excludes bills from the analysis if their vote distributions are excessively skewed, where the minority side constitutes less than 2.5% of the total vote (Poole et al., 2007). This criterion removes 2,462,241 votes from the final analysis. The second threshold excludes legislators with fewer than 20 votes, as insufficient voting data would impair adequate positioning (ibid.). This threshold results in the exclusion of 129 legislators across seven topics.

Furthermore, the NOMINATE scores for House members are calculated based on their voting history within specific policy domains. Party polarization is then assessed by calculating the absolute difference between the mean spatial positions of Republican and Democratic members. In order to identify polarization trends, the mean values and differences are calculated for each Congress, with the resulting trends plotted from the 93rd to the 118th Congress. While there are different methods of calculating party polarization, using the mean spatial positions of party members is a well-established method in polarization studies, as seen in Voteview and other research (Barber & McCarty, 2013; Lewis et al., 2021).

3.3 Correspondence Analysis

The next step involves ideologically scaling members of the House of Representatives and calculating party polarization using Correspondence Analysis (CA) (Greenacre, 2016). CA analyses contingency matrices with categorical data, extending Principal Component Analysis (PCA) by reducing dimensionality in sparse data and identifying dimensions that explain the highest variance and total inertia (Abdi & Williams, 2010; Abdi & Béra, 2017). Unlike PCA or Multiple Correspondence Analysis (MCA), which are suited for continuous and nominal categorical data respectively, CA is particularly effective for binary data, as used in this research, where a *yea* vote is coded as 1 and any other vote as 0 (Halford, 2023).

In Correspondence Analysis, the first step of the process involves standardizing the contingency matrix by its column and row sums, ensuring that all values sum to 1 (Greenacre, 2010). This is followed by centering the matrix relative to its weighted column averages. A weighted singular value decomposition (SVD) is then performed on this matrix to identify the plane closest to the data in terms of weighted least squares (Barberá et al., 2015; Greenacre, 2010). This procedure reveals the principal dimensions and coordinates that capture the most variance (inertia) within the data (Abdi & Béra, 2017; Barberá et al., 2015; Greenacre, 2010). The eigenvalues of each dimension, which quantify the amount of variance explained, are often presented in a scree plot to determine the minimal number of dimensions that provide significant cumulative variance and are thus crucial for data representation (Cattell, 1966). However, as the primary objective of this research is to examine the ideological positioning of legislators on the left-right spectrum, we decided to focus on the dimension with the highest inertia, as this is the most likely to represent the ideological scale in question (Barberá et al., 2015; Bonica et al., 2014).

While both NOMINATE and CA differ in the voting behavior they analyze, it's essential to recognize that NOMINATE is a probabilistic model estimating the ideological ideal points of legislators, whereas CA analyses the relationships within the data, focusing on the co-occurrences between legislators and spatially locating them based on these relationships.

Once the CA scores are calculated for each legislator on a given topic, the degrees of polarization are assessed by measuring the absolute difference between the mean spatial positions of Republican and Democratic members. This analysis is conducted for each congress and plotted across the 93rd-118th Congress period. It is important to note that CA assigns values to legislators randomly, which can vary across topic analyses. For interpretability purposes, the values are thus first normalized to a scale between 1 and -1, and, if necessary, inverted so that Republicans typically have positive values. This normalization and inversion further aligns CA values with NOMINATE values. Furthermore, to conform with the NOMINATE method, the same thresholds are also applied here: votes on bills where the minority side constitutes less than 2.5% of the total are removed, with the threshold for a minimum of 20 votes per legislator also being maintained.

3.4 Bimodality Coefficient and Pearson Correlation Coefficient

After the ideological locations of all legislators for each topic are calculated using the NOMINATE and CA methods, the next step consists of calculating Sarle's Bimodality Coefficient (BC) for all data points per topic (SAS Institute Inc, 2004, p. 984). This metric helps determine whether the data exhibits a bimodal or multimodal distribution, providing additional insight should the polarization trends be inconclusive. The BC is chosen for its simplicity, interpretability, and robustness, as it incorporates both the skewness and kurtosis of the data (Pfister et al., 2013). The original formula for BC with a sample size of n is:

$$BC = \frac{m_3^2 + 1}{m_4 + \frac{3(n-1)^2}{(n-2)(n-3)}}$$

where m_3 represents skewness, indicating the asymmetry of the distribution, and m_4 denotes kurtosis, measuring the tailedness or peakedness of the distribution (SAS Institute Inc., 2004). A BC value greater than the benchmark of $5/9 \approx 0.555$ suggests a bimodal or multimodal distribution, whereas lower values indicate unimodality (Pfister et al., 2013; SAS Institute Inc., 2004).

The final step of this research involves comparing the W-NOMINATE values with the Correspondence Analysis (CA) values for each policy domain to identify any correlation between the ideological spatial locations of legislators based on their entire voting behavior and positive voting behavior. This comparison is conducted using the Pearson Correlation Coefficient (PCC) (Benesty et al., 2009), which measures the linear correlation between two sets of data points. The PCC provides an intuitive interpretation, with 1 indicating a perfect positive relationship and -1 indicating a perfect negative relationship (Adler & Parmryd, 2010). A value near 0 suggests no dependency between the data sets, implying that the spatial ideological locations derived from the two methods are unrelated (ibid.). The formula for the Pearson Correlation Coefficient of two vectors x and y is:

$$PCC(x,y) = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

where m_x is the mean of vector x and m_y is the mean of vector y (Benesty et al., 2009; SciPy, 2024).

4 Results

4.1 Topic Modelling and NOMINATE

After data preparation, a total of 28,449 legislative bills were used within the LDA topic model. Despite initial analysis (Figure 4) indicating that five topics exhibit a higher average cosine similarity score, we ultimately decided on seven topics to encompass a broader range of policy domains. Table 2 presents the labels, top words, size, and Cv coherence scores for the final seven identified topics. Six of these topics are relatively balanced in size, with only the "Social Services" topic being substantially larger. However, it's lower Cv coherence score suggests that this topic includes bills that do not fall within this policy domain, indicating that this topic is partly filled with residual bills.

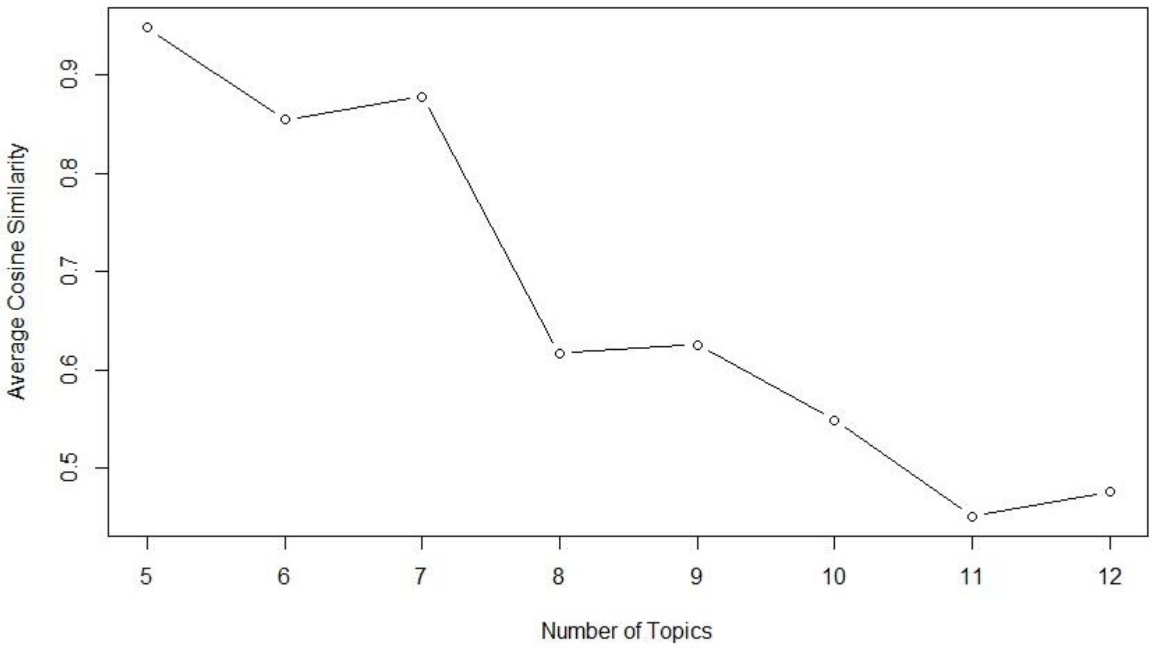


Figure 4. Topic Model Performance Measured by Average Cosine Similarity per Number of Topics.

Further analysis of the Cv coherence scores revealed that the topics of ‘Defense’, ‘International Relations’, and ‘Legislation’ showed a high coherence quality. Additionally, the topics of ‘Environment’, ‘Infrastructure’, and ‘Government Budget’ demonstrated adequate coherence quality, with the topic of ‘Social Services’ being the only topic that showed poor coherence. This poor performance indicates that the topic of ‘Social Services’ does not accurately represent its actual policy area, and will thus not be taken into consideration within the concluding analyses.

Number	Topic	Top Terms	Size	Cv Coherence
Topic 0	*Environment and Natural Resources*	lands; energy; land; interior; oil; forest; gas; water; conservation; environmental	3172	0.4478
Topic 1	*Infrastructure and Development*	housing; funds; transportation; development; energy; highway; projects; assistance; water; administration	2434	0.4243
Topic 2	*Government Budget and Administration*	funds; administration; fy; agencies; health; makes; expenses; related; commission; fund	3432	0.4259
Topic 3	*Defense and Military*	defense; military; dod; funds; security; forces; fy; personnel; air; army	3021	0.5620
Topic 4	*International Relations and Government*	president; assistances; us; international; budget; foreign; countries; expresses; development; government	3564	0.5332
Topic 5	*Legislation and Policy*	hr; consideration; rule; forth; sets; bill; house; committee; resolution; res	3880	0.5325
Topic 6	*Social Services and Public Welfare*	health; education; taks; provisions; shall; public; bill; report; provides; grants	8946	0.2867

Table 2. Seven Topics from the LDA Model with Top Words, Cv Coherence Score, and Topic Size.

Using these seven topics, we computed W-NOMINATE scores for 2,246 unique legislators per policy domain, excluding 129 legislators over the seven topics, as they did not meet the threshold of a minimum of 20 votes. The mean spatial locations of the Republican and Democratic members were then calculated per Congress for each topic, with Figure 5 illustrating the ideological trend lines of both parties and their absolute ideological differences. The grey difference line indicates a rise in polarization in all policy domains from 1973 to

2024, with "International Relations" being the only exception, showing stagnation after the 104th and 112th Congress.

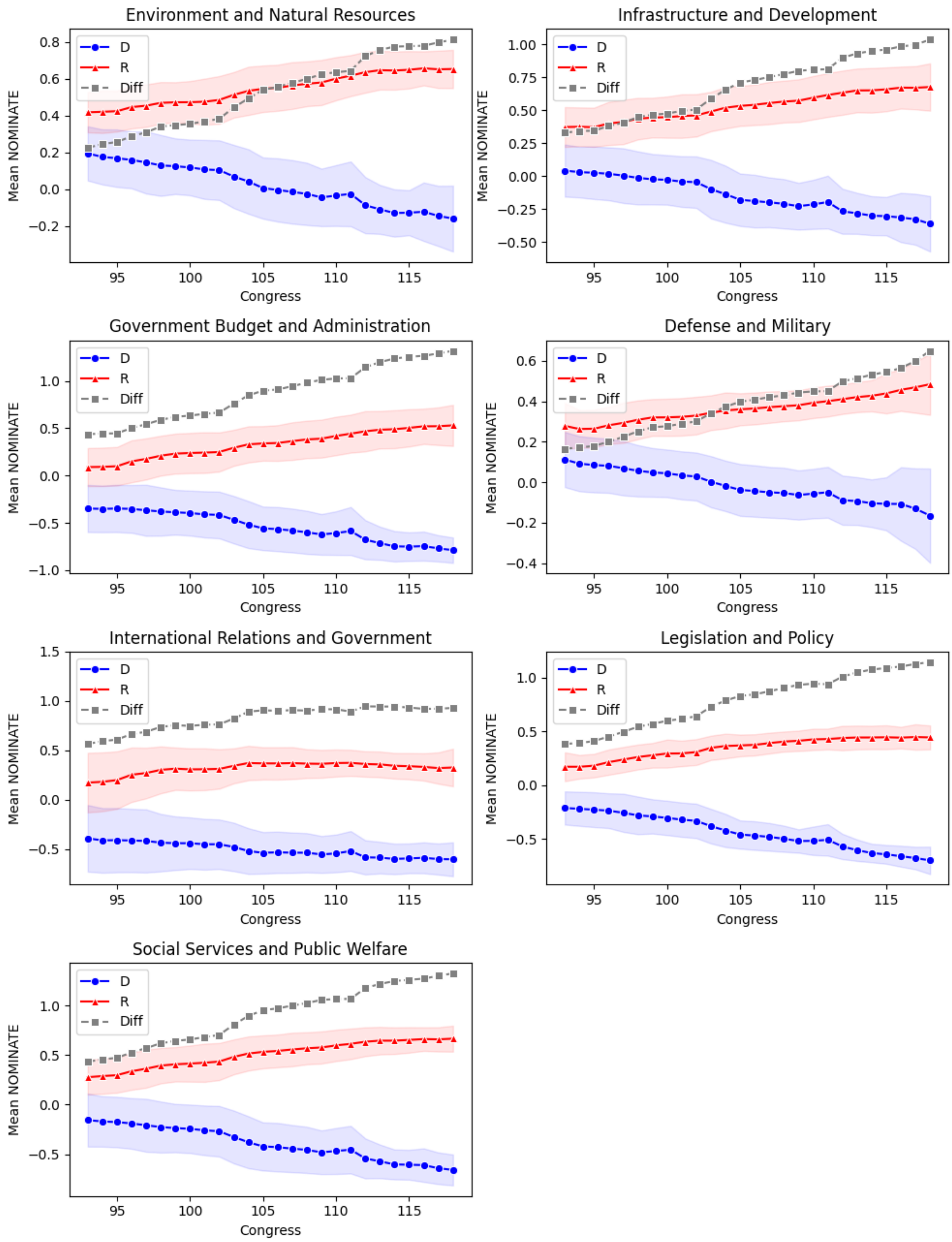


Figure 5. Ideological Trend Lines of Mean NOMINATE Scores for Republican and Democratic Parties and Their Difference, Congress 93-118.

Figure 5 shows that both the Republican and Democratic parties have become more ideologically polarized, with the Democrats trending ideologically left-wing and Republicans trending ideologically right-wing with each new Congress. Notably, 'International Relations' is the exception to this trend, with the ideological position of the Republicans seeing a stagnation from the 104th Congress, with even showing decline during the later Congresses. Another interesting outlier is the 111th Congress, where the Democratic Party experienced a peak of conservatism before recovering in the following Congress. Table 3 shows that on average, the Democratic Party has shifted more to the left. This is indicated by their average ideological change of 0.3826 in comparison to the Republican Party's average of 0.2856. The Democrats show shifts greater than 0.3 for five of the seven topics, with the topics of 'Social Services' even exceeding a change of 0.5 on the ideological scale. Figure 5 additionally shows that the early Congresses also showed a remarkable conservative position of the Democratic Party on 'Environment', 'Infrastructure', and 'Defense'. While their position does become more progressive in later Congresses, this is an intriguing phenomenon as being conservative on these topics seems to be in contradiction to the traditional ideological stances of the Democratic party.

Topic	Democratic Party	Republican Party	Difference
Environment and Natural Resources	0.3547	0.2355	0.5901
Infrastructure and Development	0.4037	0.3037	0.7074
Government Budget and Administration	0.4402	0.4406	0.8808
Defense and Military	0.2779	0.2061	0.4840
International Relations and Government	0.2118	0.1529	0.3647
Legislation and Policy	0.4876	0.2728	0.7603
Social Services and Public Welfare	0.5025	0.3879	0.8904
Average	0.3826	0.2856	0.6682

Table 3. Difference in Mean Spatial Location between the 93rd and 118th Congress.

According to Figure 6, the topics showing the largest ideological shifts are 'Government Budget', 'Legislation', and 'Social Services', with these also demonstrating the highest increase in polarization in Table 3. However, due to the low coherence score of the 'Social Services"' topic, this topic will be excluded from further analysis. As for 'Legislation', this topic primarily consists of bills related to committees and congressional oversight, with 'Government Budget' pertaining to government operations and budgetary issues. Figure 6 further shows that the topics with the smallest differences include 'Defense', 'Environment', and 'International Relations', which all have final ideological differences of less than 1. Notable is the topic of 'Defense', with a difference of under 0.7, given its usual association as an Republican policy issue. Figure 6 demonstrates that, despite the differences in values, all topics exhibit a similar pattern of polarization, with a sudden and sharp increase observed between the 102nd and 104th Congresses for all seven topics. After this increase, the

'International Relations' topic shows a divergent trend, whereas the remaining topics demonstrate a pattern of gradual increase followed by a peak in polarization and subsequent return to moderate growth.

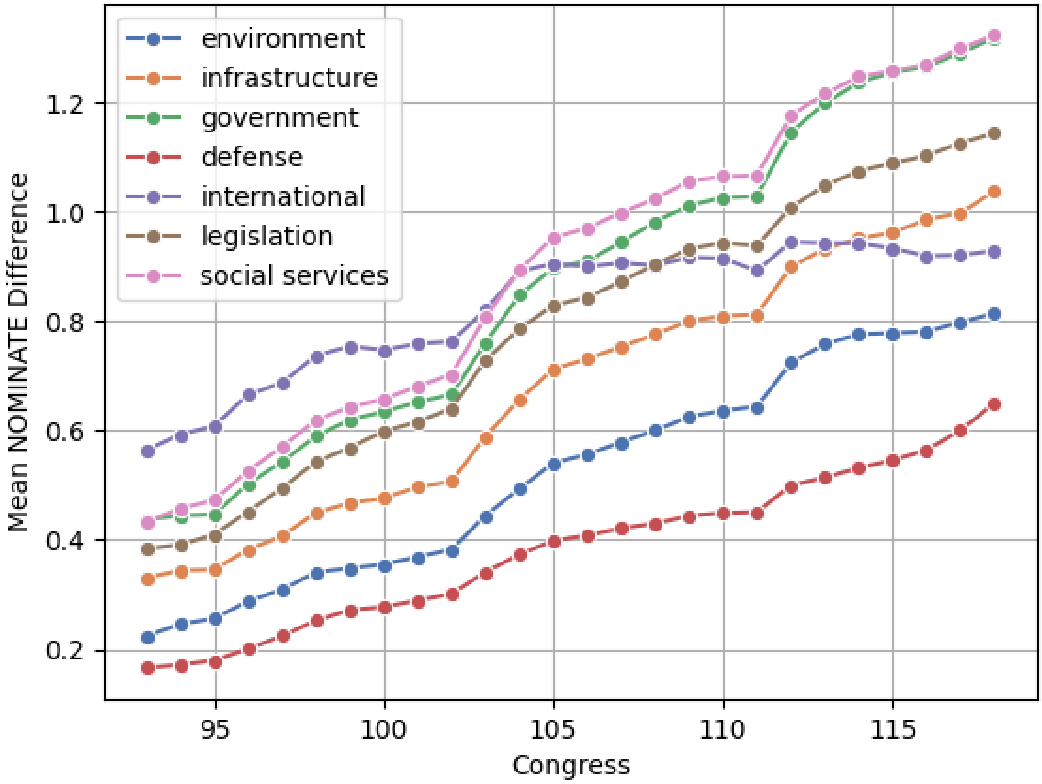


Figure 6. Ideological Differences Between Parties by Policy Using W-NOMINATE.

4.2 Part Two

Performing Correspondence Analyses on the seven topics overall resulted in relatively low inertia across the board. As depicted in Figure 7, each topic, except for 'Legislation', required ten to twelve principal components to explain 50% of the cumulative variance of the data. Given that the first principal component exhibits the highest inertia and, consequently, explains the most variance, the spatial locations from this dimension were selected for the subsequent sections of the analysis. As we expected that the left-right ideology would be the primary ideological difference between the legislators, this choice maximized the potential to capture this division within the CA data points. The degree of variance explained through the first dimension for each topic is presented in Table 4, with the complete eigenvalue scree plots for each topic further found in Appendix B.

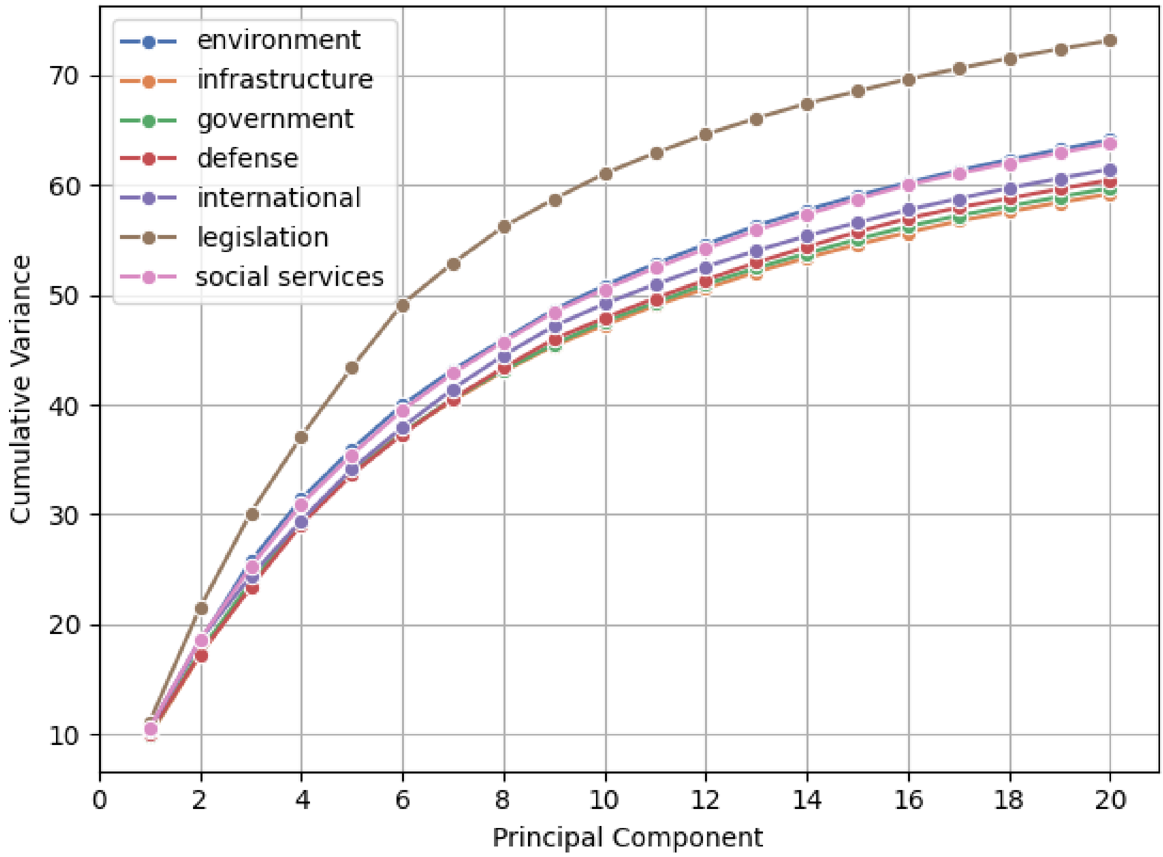


Figure 7. Cumulative Variance per Principal Component by Topic in CA.

Topic	Variance 1st Dimension	NOMINATE Bimodality Coefficient	CA Bimodality Coefficient	Pearson Correlation
Environment and Natural Resources	10.48%	0.4296	0.7234	0.0618**
Infrastructure and Development	9.83%	0.4328	0.6916	0.0963***
Government Budget and Administration	9.93%	0.5172	0.6653	0.1362***
Defense and Military	10.06%	0.2535	0.7081	0.0635**
International Relations and Government	10.39%	0.5517	0.6385	0.1123***
Legislation and Policy	11.11%	0.5574	0.5818	0.3705***
Social Services and Public Welfare	10.53%	0.5663	0.6835	0.0804***

Table 4. Analysis of Variance (1st CA Dimension), Bimodality Coefficient (NOMINATE and CA), and Pearson Correlation Coefficient per Topic. Note: Significant results indicated by: *= $p < 0.05$, **= $p < 0.01$, ***= $p < 0.001$.

Figure 9 illustrates the trend lines of both parties and their absolute difference based on the mean spatial location of their members per congress. In contrast to the polarization trends observed using the NOMINATE method, these are not replicated by the Correspondence Analysis, with only the topic of 'Legislation and Policy' showing a similar trend when examining the difference line. It is further notable that similar waves can be observed in the remaining topics when examining the Republican and Democratic parties, with both parties showing a nearly identical average spatial location for all 26 congresses. Additionally, we can observe that polarization levels peaked in the 104th and 112th Congresses for all topics, followed by a sharp drop in the subsequent congresses. Figure 8 provides an even clearer image of 'Legislation and Policy' as an outlier, with the rest of the policy domains showing very similar polarization trends and all seeing a small increase of polarization during the 93rd-118th congresses.

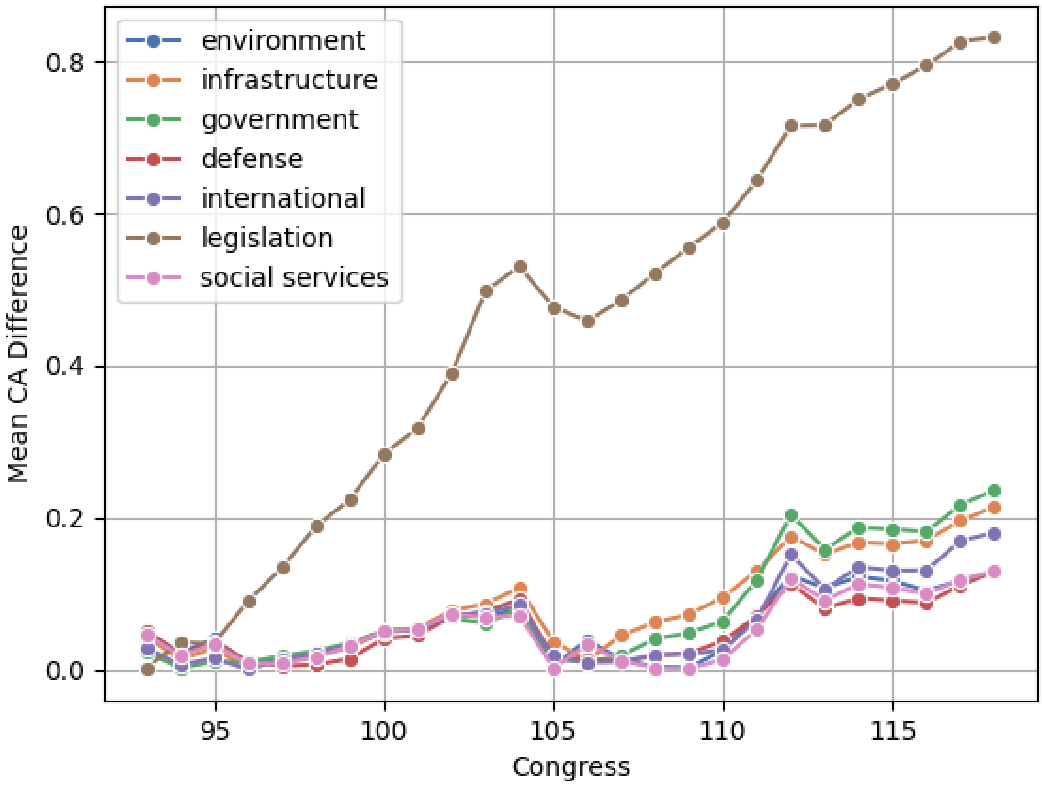


Figure 8. Ideological Differences Between Parties by Policy Using CA.

A further examination of the bimodality coefficients of both the values of the CA and W-NOMINATE methods, as presented in Table 3, reveals surprisingly that the CA data exhibits greater bimodality across all topics. Notably, the NOMINATE data even scores below the benchmark of 0.555 for four out of the seven topics. Most striking is the particularly low bimodality score of 0.2535 for the 'Defense and Military' topic. However, an examination of Figure 10, which depicts the NOMINATE values plotted against the CA values in a joint scatter-and-density plot, reveals the existence of two distinct peaks within the NOMINATE values. This observation provides evidence of bimodality in the data despite the low score received through the Bimodality Coefficient. Besides this, however, there is considerable overlap between the two parties in the ideological center, explaining why its trend of polarization in Figure 6 scores the lowest in comparison to the other topics. As for the bimodality coefficients of the CA values per topic, the values in Table 4 make it evident that these distributions are bimodal or multimodal. However, with the ideological trend lines of the Democratic and Republican parties being so similar, this indicates that these clusters represent a cross-party ideological scale that differs from the left-right party ideology that we expected.

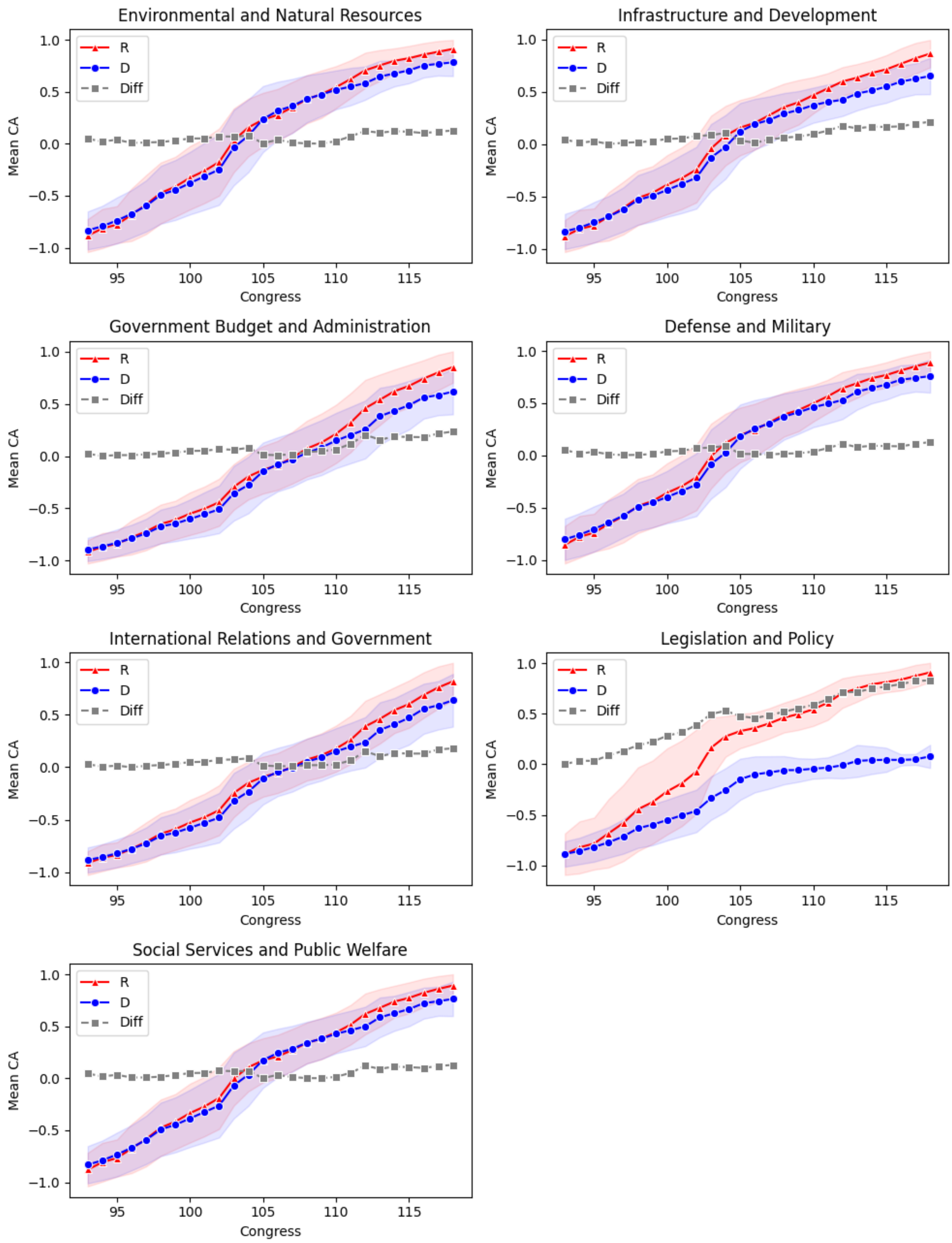


Figure 9. Ideological Trend Lines of Mean CA Scores for Republican and Democratic Parties and Their Difference, Congress 93-118.

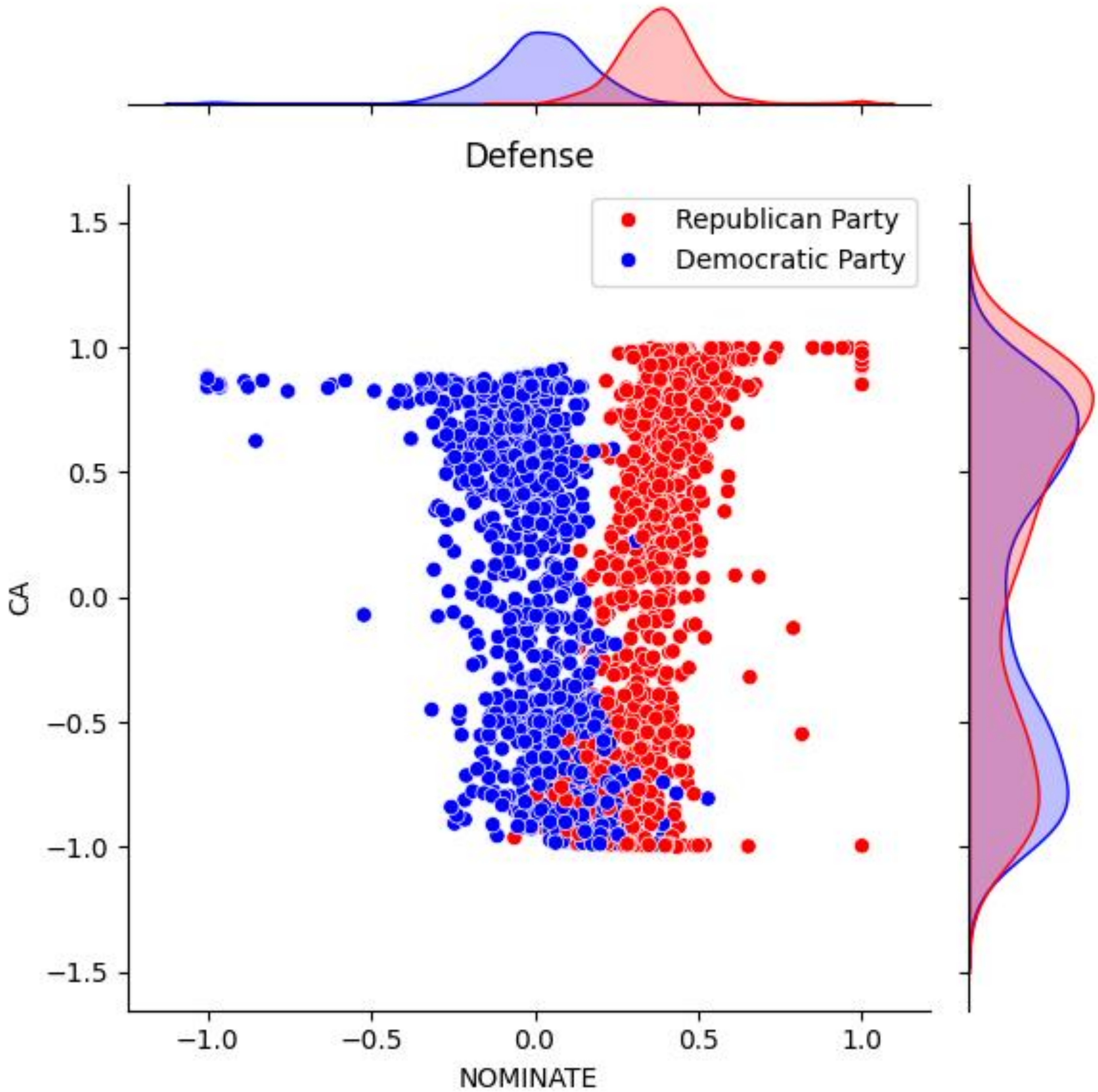


Figure 10. Scatter-and-density-plot of Ideological Spatial Location of U.S. House Representatives for Topic *Defense and Military* using W-NOMINATE and Correspondence Analysis.

The topic of 'Legislation and Policy' demonstrates bimodality according to its bimodality coefficient of 0.5818 and exhibits the clearest form of polarization when analysed using the average CA coordinates, thereby suggesting that the CA values for this topic show division between the political parties. Looking at Figure 11, this scatterplot indeed shows less overlap between the Republican and Democratic legislators, with a significant number of Republicans scoring closer to 1 in comparison to the Democrats. However, a substantial number of both Republicans and Democrats are positioned near the -1 value, demonstrating that there is also no clear distinction between the two parties regarding this topic. Scatterplots for the other policy domains are provided in Appendix C, displaying a similar structure to the 'Defense' topic in Figure 10, with Appendix D presenting violin boxplots showing the density of all CA data points for each topic.

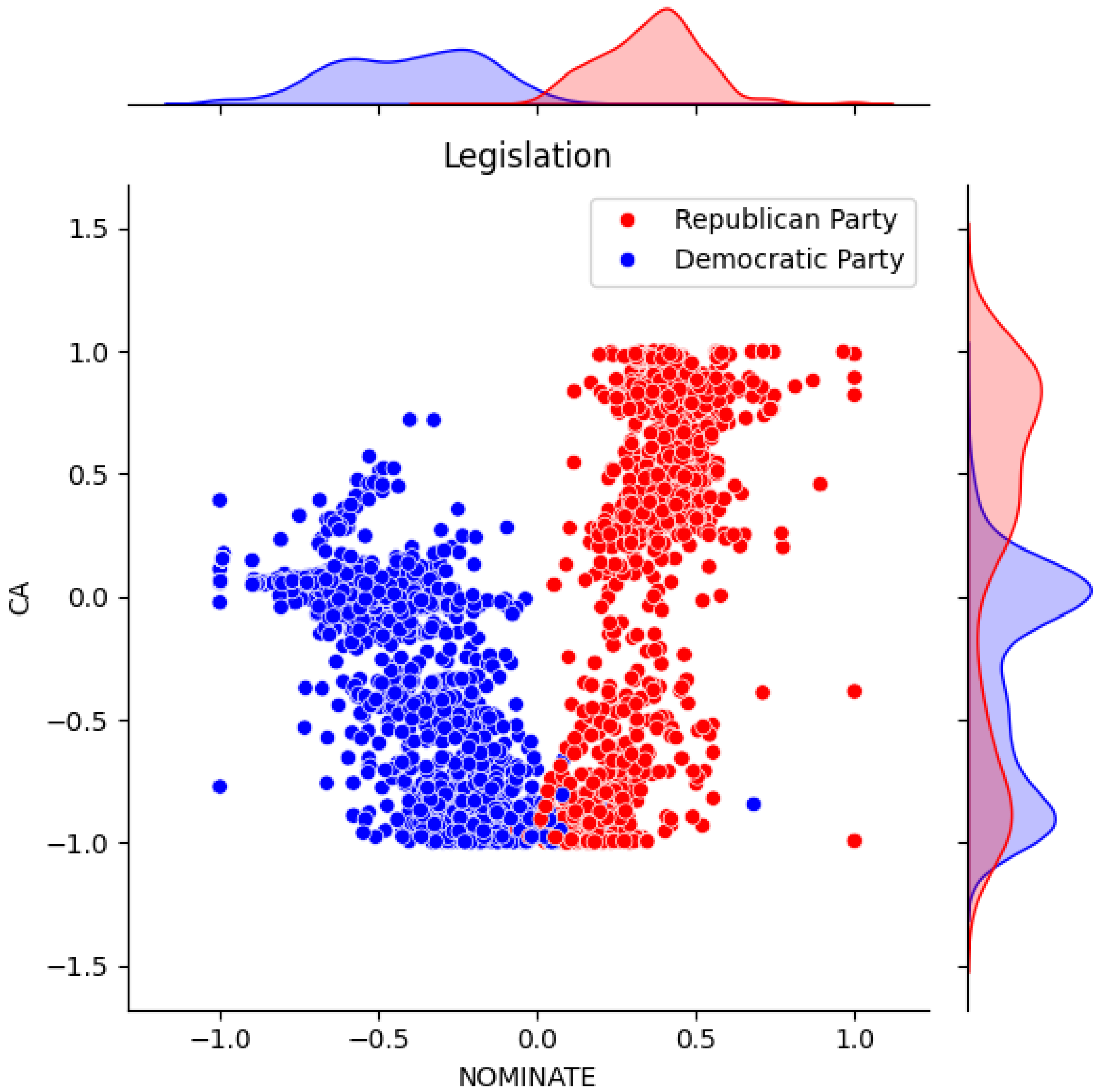


Figure 11. Scatter-and-density-plot of Ideological Spatial Location of U.S. House Representatives for Topic *Legislation and Policy* using W-NOMINATE and Correspondence Analysis.

4.3 Part Three

Given the limited polarization shown in party trend lines for most topics, and with significant overlap existing between party members' CA spatial locations, no clear parallel seems to exist between the values of the NOMINATE and CA methods. Nevertheless, comparing the W-NOMINATE and CA spatial locations of legislators shows a significant positive Pearson correlation coefficient for all topics, as seen in Table 4. Although this would suggest a correlation between high (conservative) NOMINATE and CA values, the low scores of around 0.1 imply virtually negligible correlation for the topics 'Environment', 'Infrastructure', 'Government Budget', 'Defense', 'International Relations', and 'Social Services'. Only the policy domain of "Legislation and Policy" shows a low to moderate positive correlation, aligning with the previous observations made regarding this topic. Taking the trend lines, Bimodality Coefficient, and Pearson Correlation Coefficient all into account, this indicates that only the CA values for the topic of ‘Legislation and Policy’ can adequately simulate the ideological scaling of the NOMINATE method, with the rest not accurately reflecting polarization trends within the topics.

5 Discussion & Conclusion

Over the years, methods of measuring political polarization have increasingly become more complex, taking into consideration networks between legislators and the existence of multidimensional ideological positions. In this paper, I have laid the focus on the comparison of the methods of NOMINATE and Correspondence Analysis, thereby investigating whether the less complex method of CA on the positive voting behavior of legislators can sufficiently locate their ideological position. Recent literature reported that various policy domains can see different levels of political polarization (Bateman et al., 2017); this paper performs this analysis on the voting records of U.S. House Representatives from the 93rd to the 118th Congress, separated into different policy domains. The topics were first acquired through LDA topic modelling, with the final seven topics being: Government Budget and Administration, Environment and Natural Resources, International Relations, Legislation and Policy, Infrastructure and Development, Defense and Military, and Social Services and Public Welfare. To find the actual trends of polarization per topic, I used the ideological scaling method W-NOMINATE, which spatially locates legislators based on their voting behavior and ideologically positions similar legislators closer together. Afterward, Correspondence Analyses were performed using only the positive voting behavior of representatives, again using the subsequent spatial values of members to find the trends of polarization within Congress, as well as using a Bimodality Coefficient to see if clear division could be found within data. Following, the W-NOMINATE and the CA spatial locations of the legislators were compared to each other using the Pearson Correlation Coefficient, thereby checking if the results of the CA method could be considered similar to those of the NOMINATE method.

Analysing the results of the W-NOMINATE method provides insight into the first part of the research question. It reveals similar trends of polarization between the Republican and Democratic parties across all seven topics, with the exception of later congresses within the topic of ‘International Relations and Government.' Initially, this may seem to contradict earlier studies suggesting that specific topics, such as ‘civil rights’ and ‘veteran affairs,’ exhibit less polarization (Marchi et al., 2021; Bateman et al., 2017). However, the low coherence score for ‘Social Services and Public Welfare’ suggests that both findings might still be valid. The second part of the research question was divided into two sections: first, we wanted to determine whether the same polarization trends per topic could be reflected using the CA method; second, we wanted to compare the spatial locations of both methods through correlation analysis. Notably, we found that only the topic of ‘Legislation and Policy’ displayed a significant polarization trend over the period from 1973 to 2024, with the CA values of this topic also being the only ones that showed a moderate positive correlation with the NOMINATE values. A negative conclusion can thus be reached surrounding the use of a Correspondence Analysis on positive voting behavior for the ideological locating of legislators.

This research thus confirms that trends of political polarization persist across broad policy domains, as shown through the W-NOMINATE method, challenging earlier findings that suggested smaller-scale analyses might obscure this phenomenon. Furthermore, through the comparison with CA, the effectiveness of the CA method was tested multiple times, thereby demonstrating its unsuitability for the ideological scaling of legislators. While CA has been proven to work adequately when analysing social media data or political texts, this research thus illustrates its shortcomings when performed on positive voting behavior, with only one out of seven topics receiving sufficient spatial values.

However, these findings are subject to several limitations, starting with the topic modelling approach. While aiming for an equal distribution of bills across topics does lead to more robust ideological scaling within the W-NOMINATE method, it also led to the inclusion of irrelevant bills in some categories. For instance, the 'Social Services' topic included residual bills, ultimately affecting the reliability of its polarization trends. Further research should refine topic modelling techniques for legislative bills, as there is no established state-of-the-art model in this field. Second, this study used W-NOMINATE instead of the more recent DW-NOMINATE, which accounts for changes in legislators' ideological positions over time. Although Poole and Rosenthal (1997) suggested that legislators' ideological spatial locations are relatively stable over their tenure, future studies should consider DW-NOMINATE, especially as legislators' ideological shifts might become more pronounced over time. The reliance on W-NOMINATE was primarily due to its availability in R, highlighting a need for more accessible tools and software for the more advanced methods. Finally, this study focused solely on CA applied to positive voting behavior, thereby severely limiting its scope. Future research should explore alternative methods to simplify the measurement of political polarization, such as PCA combined with CA, Multiple Correspondence Analysis on voting behavior, or analyses focusing on negative voting behavior. In conclusion, this paper offers insights into polarization trends within broad policy domains and evaluates CA for the ideological positioning of legislators, identifying its limitations. As political polarization remains a critical issue for the future of Western democracies, ongoing research into its patterns and manifestations is crucial. Simplifying methods for measuring polarization should thus be a key objective for the advancement of social data science, as it will lower the bar for future scholars interested in researching this phenomenon.

References

Abdi, H., & Béra, M. (2017). Correspondence Analysis. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining* (pp. 1–12). Springer. https://doi.org/10.1007/978-1-4614-7163-9_140-2

Abdi, H., & Williams, L.J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*, 433-459. https://doi.org/10.1002/wics.101

Adler, J., & Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry*, *77A*(8), 733-742. https://doi.org/10.1002/cyto.a.20896

Aldrich, J. H., & McKelvey, R. D. (1977). A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections. *American Political Science Review*, *71*(1), 111–130. https://doi.org/10.2307/1956957

Aldrich, J. H., Montgomery, J. M., & Sparks, D. B. (2017). Polarization and Ideology: Partisan Sources of Low Dimensionality in Scaled Roll Call Analyses. *Political Analysis*, *22*(4), 435–456. https://doi.org/10.1093/pan/mpt048

Andris, C., Lee, D., Hamilton, M. J., Martino, M., Gunning, C. E., & Selden, J. A. (2015). The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives. *PLOS ONE*, *10(*4), e0123507. https://doi.org/10.1371/journal.pone.0123507

Arbatli, E., & Rosenberg, D. (2021). United we stand, divided we rule: How political polarization erodes democracy. *Democratization*, *28*(2), 285–307. https://doi.org/10.1080/13510347.2020.1818068

Barber, M. & McCarty, N. (2013). Causes and Consequences of Polarization. In J. Mansbridge, & C. J. Martin (Eds.), *Negotiation Agreement in Politics* (pp. 19-53). American Political Science Association.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Bateman, D. A., Clinton, J. D., & Lapinski, J. S. (2017). A House Divided? Roll Calls, Polarization, and Policy Differences in the U.S. House, 1877–2011. *American Journal of Political Science*, *61*(3), 698–714. https://doi.org/10.1111/ajps.12281

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson Correlation Coefficient. In J. Benesty, J. Chen, Y. Huang, & I. Cohen, (Eds.), *Noise Reduction in Speech Processing* (pp. 1-4). Springer, https://doi.org/10.1007/978-3-642-00296-0_5

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*, 993-1022.

Boche, A., Lewis, J. B., Rudkin, A., & Sonnet, L. (2018). The new Voteview.com: Preserving and continuing Keith Poole's infrastructure for scholars, students and observers of Congress. *Public Choice*, *176*(1), 17–32. https://doi.org/10.1007/s11127-018-0546-0

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science*, *58*(2), 367–386. https://doi.org/10.1111/ajps.12062

Cameron, C. M., & Park, J. K. (2009). How will they vote? Predicting the future behavior of Supreme Court nominees, 1937–2006. *Journal of Empirical Legal Studies*, *6*(3), 485-511. https://doi.org/10.1111/j.1740-1461.2009.01151.x

Campani, G., Fabelo Concepción, S., Rodriguez Soler, A., & Sánchez Savín, C. (2022). The rise of Donald Trump right-wing Populism in the United States: Middle American radicalism and anti-immigration discourse. *Societies*, *12*(6), 154. https://doi.org/10.3390/soc12060154

Capshaw, N. C. (2005). The Social Cohesion Role of the Public Sector. *Peabody Journal of Education*, *80*(4), 53–77. https://doi.org/10.1207/S15327930pje8004_5

Carroll, R., Lewis, J. B., Lo, J., Poole, K. T., & Rosenthal, H. (2009). Comparing NOMINATE and IDEAL: Points of Difference and Monte Carlo Tests. *Legislative Studies Quarterly*, *34*(4), 555–591. https://doi.org/10.3162/036298009789869727

Carson, J. L., Finocchiaro, C. J., & Rohde, D. W. (2010). Consensus, Conflict, and Partisanship in House Decision Making: A Bill-Level Examination of Committee and Floor Behavior. *Congress & the Presidency*, *37*(3), 231–253. https://doi.org/10.1080/07343469.2010.486393

Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. http://dx.doi.org/10.1207/s15327906mbr0102_10

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The Muppets straight out of Law School (arXiv:2010.02559). *arXiv*. https://doi.org/10.48550/arXiv.2010.02559

Chen, Q., Yao, L., & Yang, J. (2016). Short text classification based on LDA topic model. *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, 749-753. https://doi.org/10.1109/ICALIP.2016.7846525

Citrin, J., & Stoker, L. (2018). Political Trust in a Cynical Age. *Annual Review of Political Science 21*(21), 49–70, https://doi.org/10.1146/annurev-polisci-050316-092550

Clinton, J., Jackman, S., & Rivers, D. (2004). The Statistical Analysis of Roll Call Data. *American Political Science Review*, *98*(2), 355–370. https://doi.org/10.1017/S0003055404001194

Congress.gov. (2024a, June 24). *About Congress.gov*. https://www.congress.gov/about

Congress.gov. (2024b, June 24). *Field Values for Legislative Policy Area Term Vocabulary*. https://www.congress.gov/help/field-values/policy-area

Crespin, M. H., & Rohde, D. W. (2010). Dimensions, Issues, and Bills: Appropriations Voting on the House Floor. *The Journal of Politics, 72*(4), 976–989. https://doi.org/10.1017/S0022381610000472

Desmarais, B. A., Moscardelli, V. G., Schaffner, B. F., & Kowal, M. S. (2015). Measuring legislative collaboration: The Senate press events network. *Social Networks*, *40*, 43–54. https://doi.org/10.1016/j.socnet.2014.07.006

Ebeling, R., Nobre, J., & Becker, K. (2023). A multi-dimensional framework to analyze group behavior based on political polarization. *Expert Systems with Applications, 233*. https://doi.org/10.1016/j.eswa.2023.120768

Economist. (2024, April 1). *Latin America's new hard right: Bukele, Milei, Kast and Bolsonaro.* The Economist. https://www.economist.com/the-americas/2024/04/01/latin-americas-new-hard-right-bukele-milei-kast-and-bolsonaro

Egar, W. T. (2016). Tarnishing Opponents, Polarizing Congress: The House Minority Party and the Construction of the Roll-Call Record. *Legislative Studies Quarterly*, *41*(4), 935–964. https://doi.org/10.1111/lsq.12135

Everson, P., Valelly, R., Vishwanath, A., & Wiseman, J. (2016). NOMINATE and American political development: a primer. *Studies in American Political Development*, *30*(2), 97-115. https://doi.org/10.1017/S0898588X16000067

Fowler, J. H. (2006). Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis, 14*(4), 456–487. https://doi.org/10.1093/pan/mpl002

Gerrish, S. M., & Blei, D. M. (2011). Predicting legislative roll calls from text. *Proceedings of the 28th International Conference on Machine Learning,* 489 – 496, https://oar.princeton.edu/handle/88435/pr1p50h

Greenacre, M.J. (2010), Correspondence analysis. *WIREs Computational Statistics, 2*, 613-619. https://doi.org/10.1002/wics.114

Greenacre, M. (2016). *Correspondence Analysis in Practice* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315369983

Guenduez, A. A., Schedler, K., & Ciocan, D. (2016). Generic frames and tonality: Mapping a polarizing issue in a multifaceted context. *European Journal of Communication, 31*(5), 584–599. https://doi.org/10.1177/0267323116666478

Halford, M. (2023). Prince [Computer Software]. MIT. https://maxhalford.github.io/prince/

Hare, C., Armstrong II, D. A., Bakker, R., Carroll, R., & Poole, K. T. (2015). Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions. *American Journal of Political Science, 59*(3), 759–774. https://doi.org/10.1111/ajps.12151

Hare, C., & Poole, K. T. (2014). The Polarization of Contemporary American Politics. *Polity, 46*(3), 411–429. https://doi.org/10.1057/pol.2014.10

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science, 22*, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhou, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*, 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jochim, A. E., & Jones, B. D. (2012). Issue Politics in a Polarized Congress. *Political Research Quarterly, 66*(2), 352–369. https://doi.org/10.1177/1065912912448930

Joesten, D. A., & Stone, W. J. (2014). Reassessing Proximity Voting: Expertise, Party, and Choice in Congressional Elections. *The Journal of Politics*, *76*(3), 740–753. https://doi.org/10.1017/S0022381614000140

Kirkland, J. H., & Gross, J. H. (2014). Measurement and theory in legislative networks: The evolving topology of Congressional collaboration. *Social Networks, 36,* 97–109. https://doi.org/10.1016/j.socnet.2012.11.001

Körösényi, A. (2013). Political polarization and its consequences on democratic accountability. *Corvinus Journal of Sociology and Social Policy, 4*(2). https://doi.org/10.14267/cjssp.2013.02.01

Lai, A., Brown, M. A., Bisbee, J., Tucker, J. A., Nagler, J., & Bonneau, R. (2024). Estimating the Ideology of Political YouTube Videos. *Political Analysis*, 1–16. https://doi.org/10.1017/pan.2023.42

Laver, M., Benoit, K., & Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review*, *97*(2), 311–331. https://doi.org/10.1017/S0003055403000698

Lee, A. H.-Y. (2022). Social Trust in Polarized Times: How Perceptions of Political Polarization Affect Americans' Trust in Each Other. *Political Behavior*, *44*(3), 1533–1554. https://doi.org/10.1007/s11109-022-09787-1

Lewis, J. B., & Poole, K.T. (2004). Measuring Bias and Uncertainty in Ideal Point Estimates via the Parametric Bootstrap. *Political Analysis*, *12*(2),105-127. https://doi.org/10.1093/pan/mph015

Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A. & Sonnet, L. (2021). Voteview: Congressional Roll-Call Votes Database. https://voteview.com/

Lindgren, E., & Southwell, P. (2014). The effect of electoral competitiveness on ideological voting patterns in the US House, 2002-10. *Politics & Policy*, *42*(6), 905-918. https://doi.org/10.1111/polp.12097

Maestas, C. D., Buttice, M. K., & Stone, W. J. (2014). Extracting Wisdom from Experts and Small Crowds: Strategies for Improving Informant-based Measures of Political Concepts. *Political Analysis*, *22*(3), 354–373. https://doi.org/10.1093/pan/mpt050

Marchi, S. de, Dorsey, S., & Ensley, M. J. (2021). Policy and the structure of roll call voting in the US house. *Journal of Public Policy*, *41*(2), 384–408. https://doi.org/10.1017/S0143814X20000069

McCarty, N. M., Poole, K. T., & Rosenthal, H. (1997). *Income Redistribution and the Realignment of American Politics*. SEI Press

McCarty, N., Poole, K. T., & Rosenthal, H. (2016). *Polarized America, second edition: The Dance of Ideology and Unequal Riches.* MIT Press.

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist*, *62*(1), 16–42. https://doi.org/10.1177/0002764218759576

McLay, R., & Ramos, H. (2021). Do Polarization Narratives Apply to Politics on the Periphery? The Case of Atlantic Canada. *Frontiers in Sociology*, *6*. https://doi.org/10.3389/fsoc.2021.655880

Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2017). Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis*, *16*(4), 372–403. https://doi.org/10.1093/pan/mpn018

Neal, Z. P. (2020). A sign of the times? Weak and strong polarization in the U.S. Congress, 1973–2016. *Social Networks*, *60*, 103–112. https://doi.org/10.1016/j.socnet.2018.07.007

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. (2013). Good things peak in pairs: a note on the bimodality coefficient. *Frontiers in Psychology*, *4*(700). https://doi.org/10.3389/fpsyg.2013.00700

Poole, K. T. (2005). *Spatial models of parliamentary voting*. Cambridge University Press.

Poole, K. T., Lewis, J. B., Lo, J., & Carroll, R. (2008). Scaling Roll Call Votes with W-NOMINATE in R. *Journal of Statistical Software.* http://dx.doi.org/10.2139/ssrn.1276082

Poole, K., Lewis, J., & Carroll, J. L. and R. (2024). wnominate: Roll Call Analysis Software (1.5) [Computer software]. https://cran.r-project.org/web/packages/wnominate/index.html

Poole, K. T., & Rosenthal, H. (1985). A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, *29*(2), 357–384. https://doi.org/10.2307/2111172

Poole, K. T., & Rosenthal, H. (1997). *Congress: A Political-economic History of Roll Call Voting*. Oxford University Press.

Poole, K. T., & Rosenthal, H. (2001). D-Nominate after 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting. *Legislative Studies Quarterly*, *26*(1), 5–30.

Poole, K. T., & Rosenthal, H. (2011). *Ideology and Congress*. (2nd ed.) Transaction Publishers.

Porter, M. A., Mucha, P. J., Newman, M. E. J., & Warmbrand, C. M. (2005). A network analysis of committees in the U.S. House of Representatives. *Proceedings of the National Academy of Sciences*, *102*(20), 7057–7062. https://doi.org/10.1073/pnas.0500191102

ProPublica. (2024). U.S. Congress: Bulk Data on Bills [Data set]. ProPublica. https://www.propublica.org/datastore/dataset/congressional-data-bulk-legislation-bills

Ramey, A. (2016). Vox Populi, Vox Dei? Crowdsourced Ideal Point Estimation. *The Journal of Politics*, *78*(1), 281–295. https://doi.org/10.1086/683395

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceeding of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45-50. http://is.muni.cz/publication/884893/en

Roberts, J. M. (2007). The Statistical Analysis of Roll-Call Data: A Cautionary Tale. *Legislative Studies Quarterly*, *32*(3), 341–360. http://www.jstor.org/stable/40263426

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*. 399–408. https://doi.org/10.1145/2684822.2685324

SAS Institute Inc. (2004). *SAS/STAT® User's Guide.* SAS Publishing.

SciPy. (2024). scipy.stats.pearsonr [Computer Software]. The SciPy Community. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html#r8c6348c62346-1

Simha, A. (2021, October 6). *Understanding TF-IDF for Machine Learning: A gentle introduction to term frequency-inverse document frequency*. Capital One. https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/

Slapin, J. B., & Proksch, S.-O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, *52*(3), 705–722. https://doi.org/10.1111/j.1540-5907.2008.00338.x

Sønvisen, S. A. (2014). Contemporary fisher images: Ideologies, policies and diversity. *Journal of Rural Studies*, *34*, 193–203. https://doi.org/10.1016/j.jrurstud.2014.01.011

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL '12)*. 952–961.

Stone, W. J., & Simas, E. N. (2010). Candidate Valence and Ideological Positions in U.S. House Elections. *American Journal of Political Science*, *54*(2), 371–388. https://doi.org/10.1111/j.1540-5907.2010.00436.x

Svolik, M. W. (2019). Polarization versus Democracy. *Journal of Democracy*, *30*(3), 20–32. https://doi.org/10.1146/annurev-polisci-050316-092550

Syed, S., & Spruit, M. (2017). Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 165-174. https://doi.org/10.1109/DSAA.2017.61

Tausanovitch, C., & Warshaw, C. (2017). Estimating Candidates' Political Orientation in a Polarized Congress. *Political Analysis*, *25*(2), 167–187. https://doi.org/10.1017/pan.2017.5

Tong, Z., & Zhang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. *Computer Science & Information Technology*. *6*. 201-210. http://dx.doi.org/10.5121/csit.2016.60616

Vohra, A. (2024, March 13). *A Far-Right Populist Takeover of the EU is Underway*. Foreign Policy. https://foreignpolicy.com/2024/03/13/eu-parliament-elections-populism-far-right/

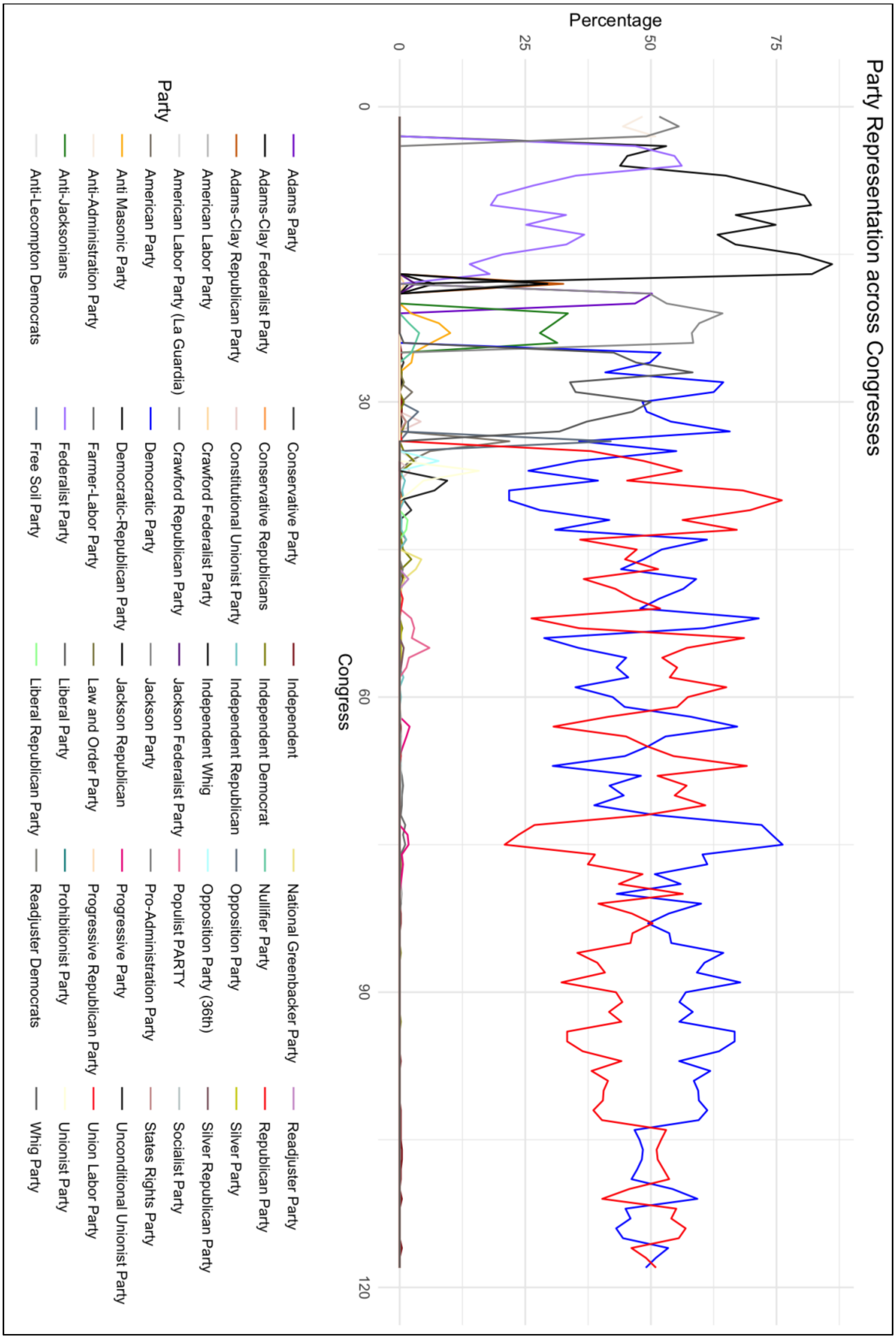
Wang, E., Salazar, E., Dunson, D., & Carin, L. (2013). Spatio-Temporal Modeling of Legislation and Votes. *Bayesian Analysis*, *8*(1), 233–268. https://doi.org/10.1214/13-BA810

Wiesehomeier, N., & Doyle, D. (2012). Attitudes, Ideological Associations and the Left–Right Divide in Latin America. *Journal of Politics in Latin America*, *4*(1), 3–33. https://doi.org/10.1177/1866802X1200400101

Zhang, Y., Friend, A. J., Traud, A. L., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2008). Community structure in Congressional cosponsorship networks. *Physica A: Statistical Mechanics and Its Applications*, *387*(7), 1705–1712. https://doi.org/10.1016/j.physa.2007.11.004

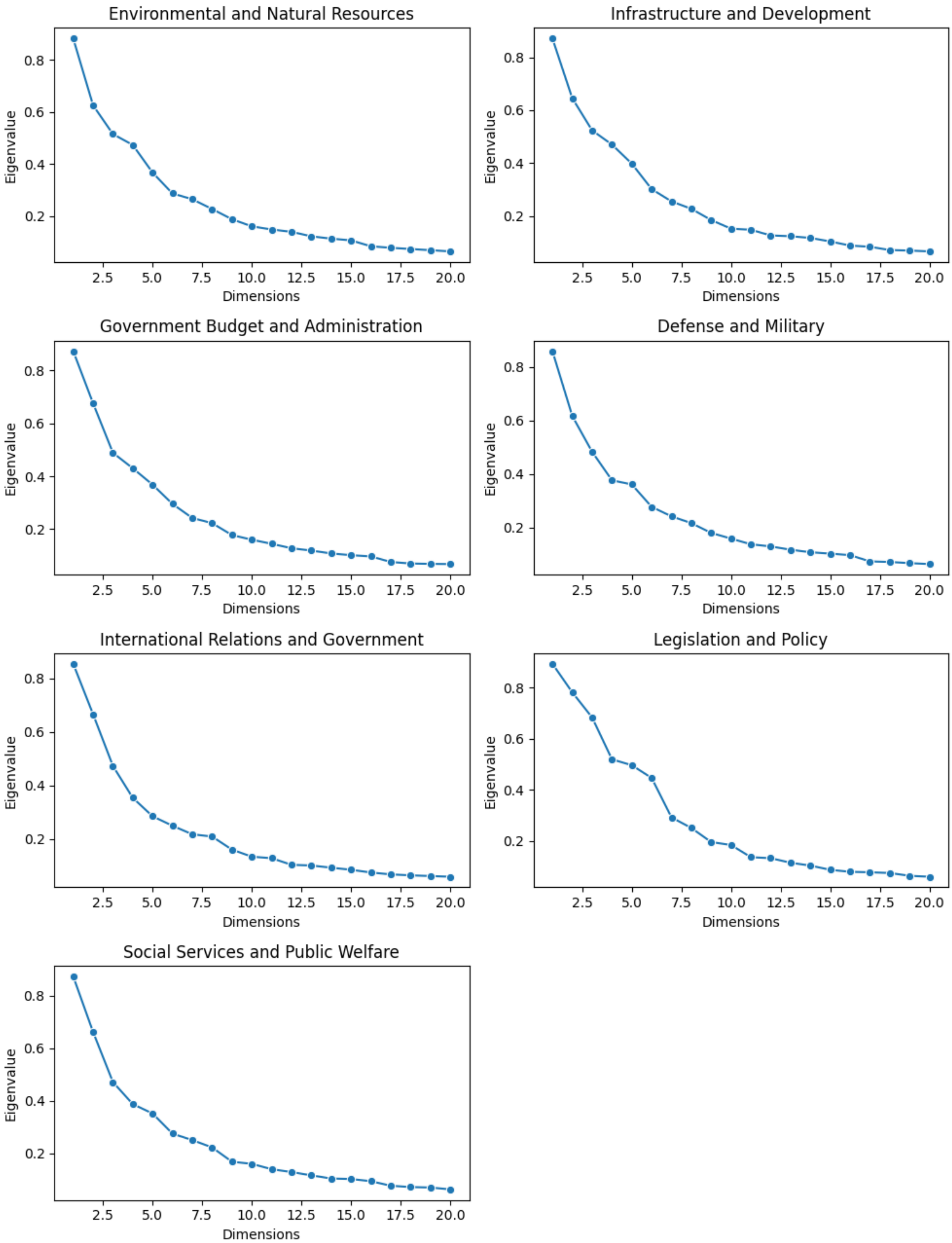
Appendix A:

Evolution of All Party Representation in the U.S. House of Representatives, 1789-2024.



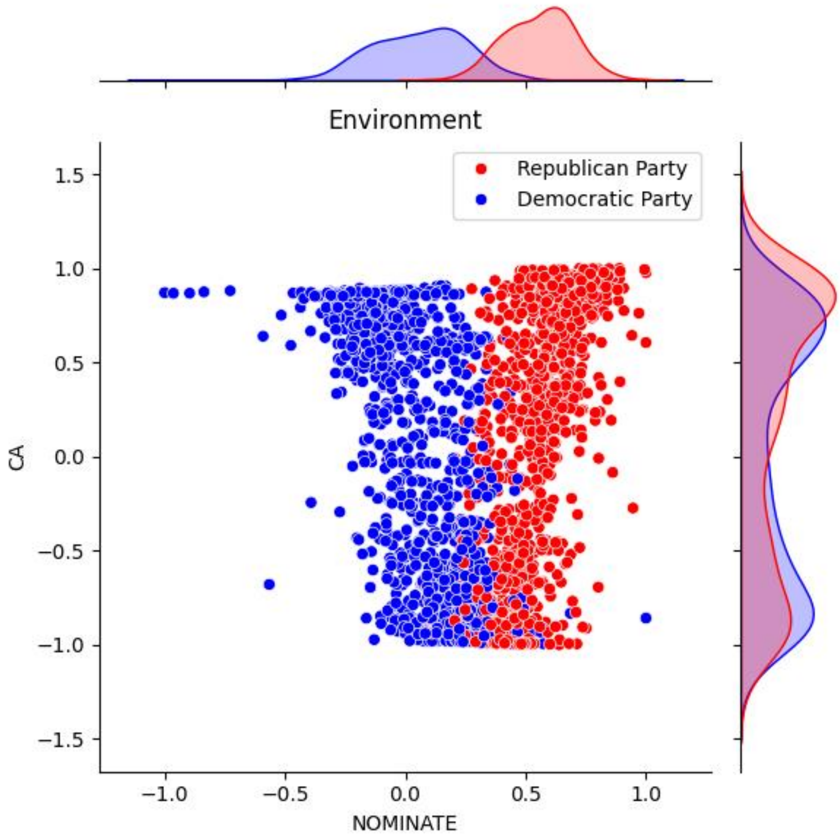
Appendix B:

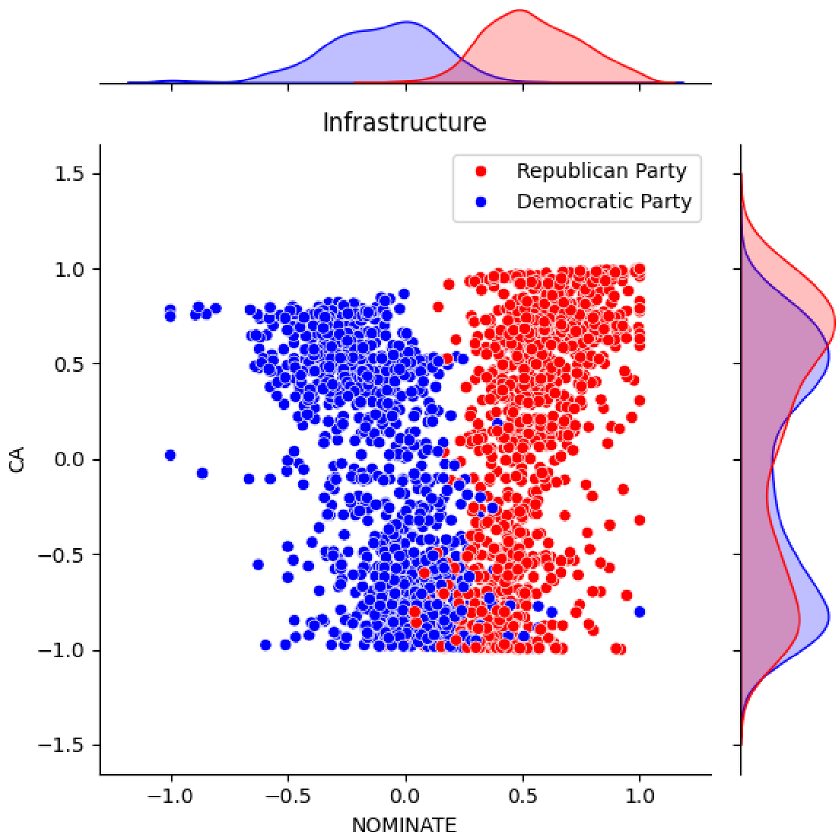
Eigenvalue Screeplots of CA Data for each Policy Domain.

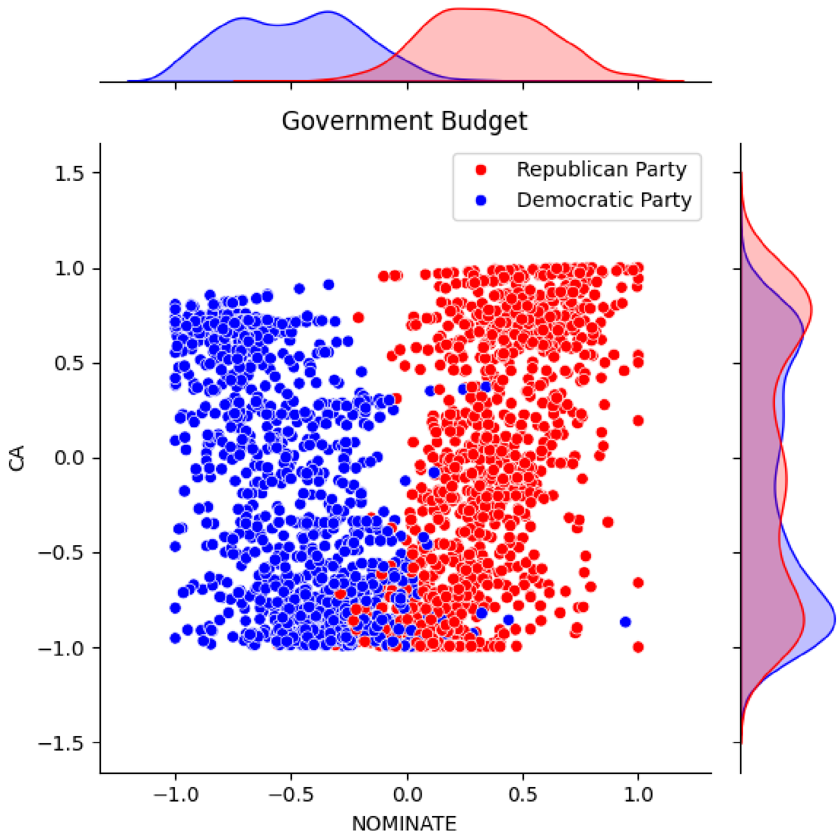


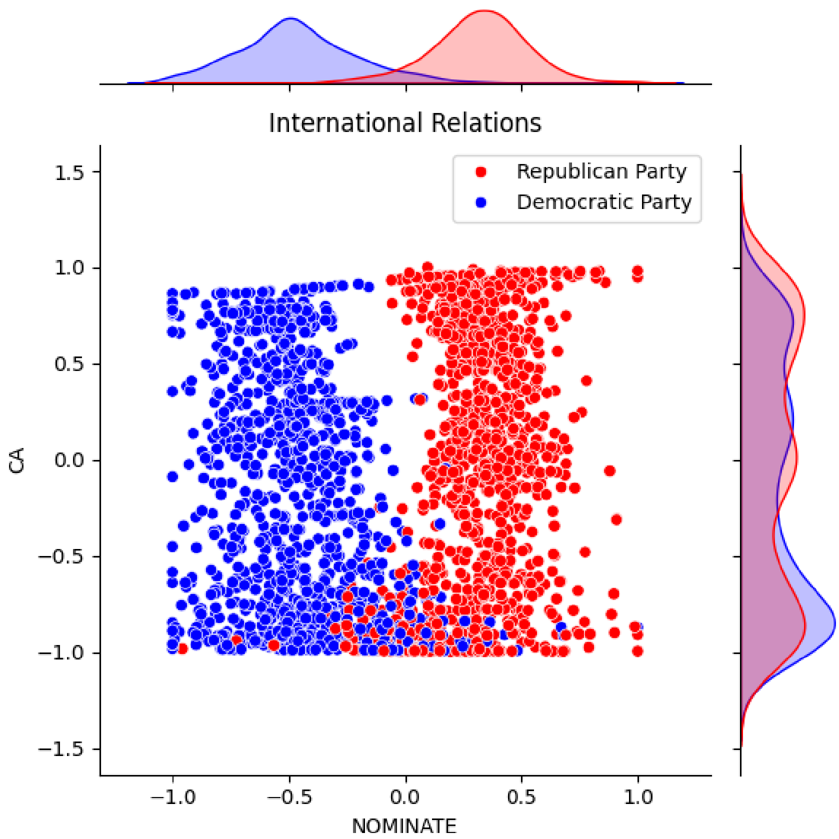
Appendix C:

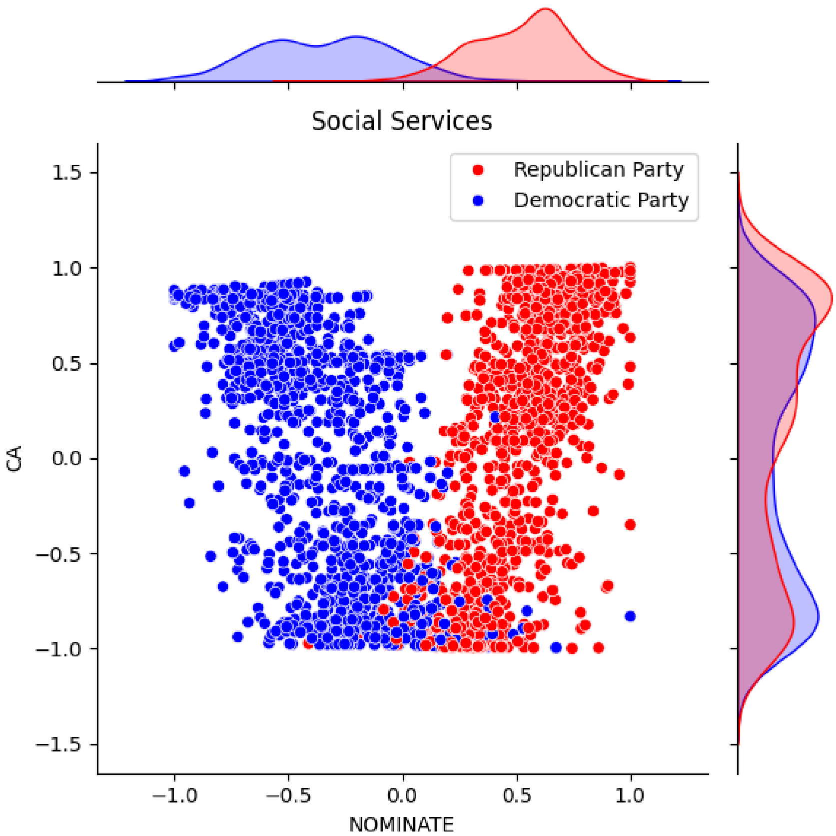
Scatter-and-density-plots of All Topics except *Defense* and *Legislation*.





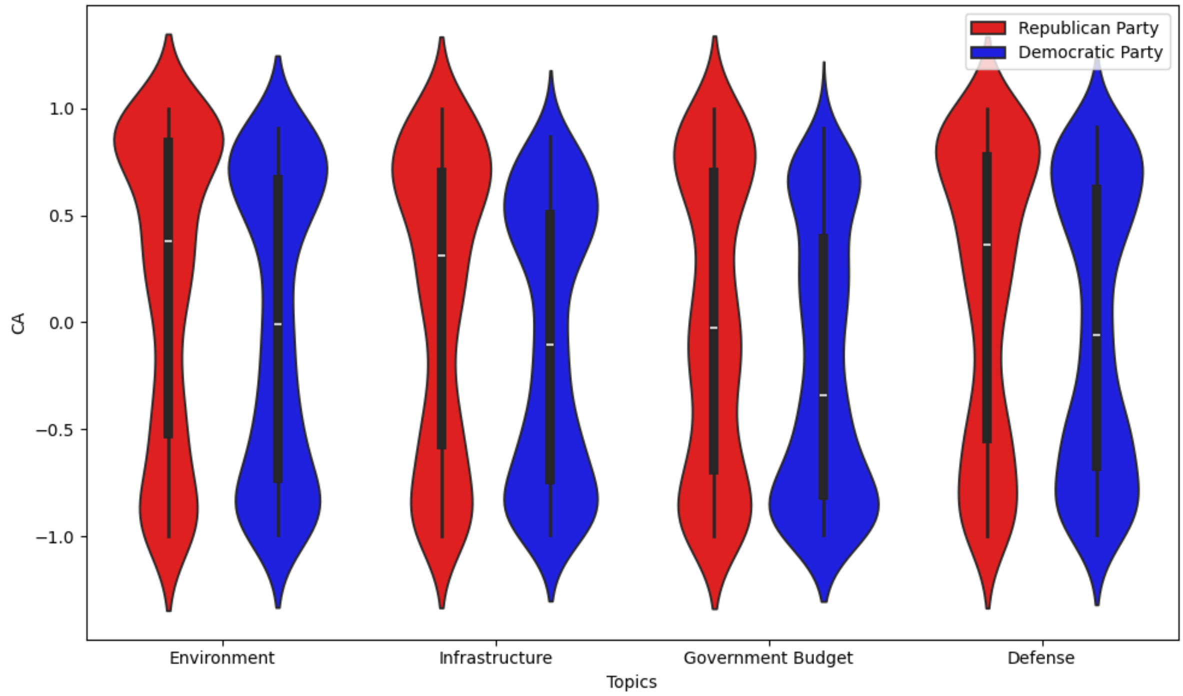

Government Budget
Republican Party
Democratic Party
CA
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
NOMINATE

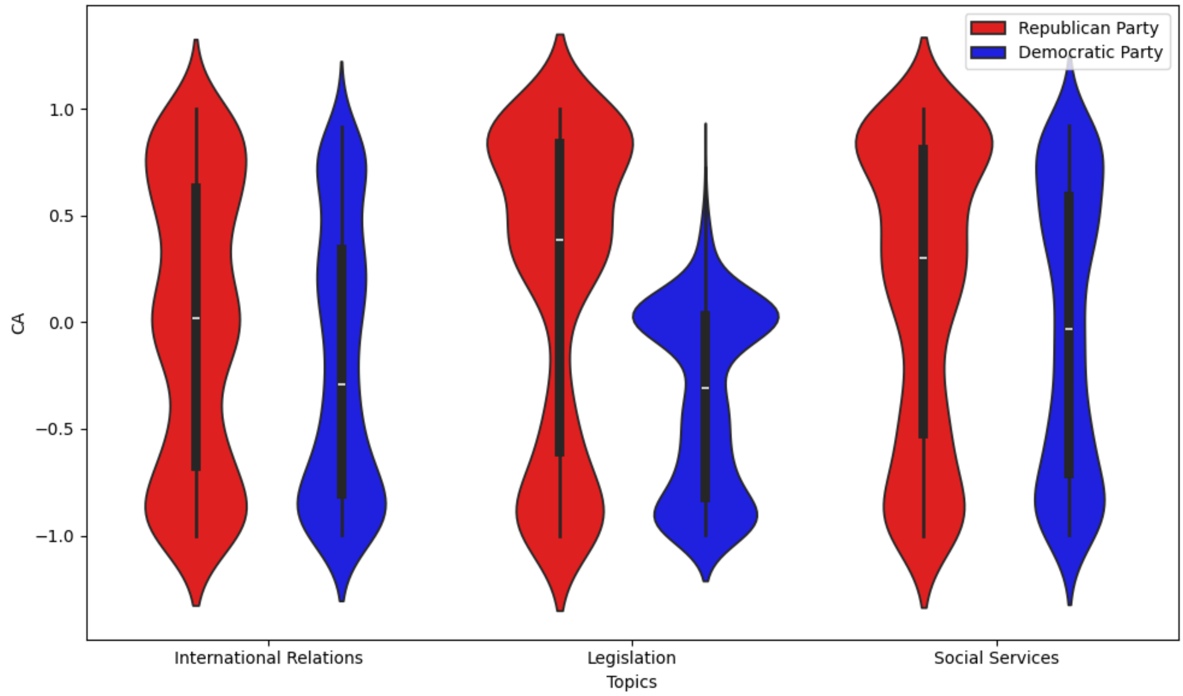

International Relations
Republican Party
Democratic Party
CA
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
NOMINATE


Social Services
Republican Party
Democratic Party
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
CA
−1.0
−0.5
0.0
0.5
1.0
NOMINATE

Appendix D:

Violin Density Boxplots of All Topics





Appendix E:

https://github.com/DionCU