Political Polarization: Real or Illusion?

Junseok Yang
Department of Statistics
University of Illinois Urbana-Champaign
jyang247@illinois.edu

Abstract

This final project investigates whether users' self-reported reasons for joining a non-profit organization "Braver Angels" exhibit signs of political polarization. As a continuation of Project 2, which employed unsupervised learning techniques such as clustering and topic modeling on user responses, this final project extends the analysis by incorporating contextual embeddings derived from language models (e.g., Sentence-BERT) to segment responses more semantically. In addition to comparing clustering results with those from the previous project, this study introduces a new latent construct, extremity level, intended to capture rhetorical or ideological intensity. This variable is defined through a combination of linguistic features—including sentiment, complexity, and semantic distance from cluster centroids—and reduced via Principal Component Analysis (PCA). Finally, regression analysis is used to test whether political lean (Red, Blue, Other) is associated with this extremity score, with additional inferential tests (ANOVA F-test and Tukey HSD) providing insight into group differences.

1 Introduction

Political polarization in the United States has deepened over the past several decades, reshaping not only partisan politics but also social attitudes and individual political behavior. This growing divide encompasses both ideological polarization—divergence in core policy positions—and affective polarization, where individuals express increasingly negative sentiment toward opposing political groups. According to the Pew Research Center, the proportion of Americans with consistently liberal or conservative views more than doubled between 1994 and 2014, rising from 10% to 21% (Center, 2014). Meanwhile, American National Election Studies (ANES) data and Gallup polling have documented steep declines in cross-party favorability, with presidential approval ratings now exhibiting historic partisan gaps, such as the 83-point split between Republican and Democratic evaluations of President Trump (Studies, 2020; Gallup, 2021).

In addition to ideological and emotional divides, structural and technological factors exacerbate polarization. A study by the National Bureau of Economic Research (NBER) found that economic disparities between Democratic- and Republican-leaning regions have grown since the 1970s, reinforcing partisan identity along geographic lines (Dorn et al., 2020). Digital environments also amplify division; online platforms often facilitate ideological echo chambers that heighten political isolation and hostility, as shown by research in the Proceedings of the National Academy of Sciences (Bail et al., 2021). Recent computational linguistics research has further demonstrated how polarization can be detected and quantified through patterns in natural language, including sentiment and lexical framing (Iyyer et al., 2018).

Building on this context, the present project seeks to identify signs of political polarization within open-ended user responses to the question, "Why I joined," collected by the organization. This work extends a prior analysis (Project 2), which applied unsupervised learning and topic modeling to extract thematic clusters from user responses. In this final project, we employ contextual embeddings derived from a language model to improve clustering performance and interpretability. Additionally, we introduce a latent variable, "Extremity level", which combines multiple linguistic features (e.g., sentiment, complexity) as a proxy for ideological or rhetorical intensity. Using statistical techniques such as Principal Component Analysis (PCA), ANOVA, and post-hoc tests, we examine whether political lean is associated with this latent extremity level, offering insight into how polarization may manifest in political self-expression.

2 Data & Pre-processing

2.1 Data

The data set used was provided by a nonprofit organization "Braver Angels" for the purpose of academic usage in this project. It contains 18765 observations with features such as "Member Start Date" (When user joined the organization), "Profession" (User's occupation/job), 'Leans' (Political partisanship) and "Why I joined" (User response). In this project, "Leans" and "Why I joined" are the two main features that are going to be analyzed. For "Leans", only 3 groups were included: "Blue" (10697 users), "Red" (2111 users), and "Other" (1782 users) after dropping duplicate observations.

2.2 Text Cleaning

The primary text variable used in this project, users' open-ended responses to the question "Why I joined", was converted into sentence-level vector representations using contextual embeddings. Specifically, we used the "all-MiniLM-L6-v2" model from the Sentence-BERT framework, implemented via the SentenceTransformer class in Python. This model generates dense semantic representations that preserve contextual meaning, enabling nuanced comparisons between sentences based on meaning rather than simple word overlap.

Unlike traditional topic modeling approaches such as Latent Dirichlet Allocation (LDA), contextual embedding models do not require extensive text pre-processing. Rather, practices such as removing punctuations and stopwords, converting to lowercase, or lemmatizing may degrade model performance, since these models are pretrained on naturalistic, unfiltered language. Therefore, we retained the original structure and formatting of the text, including punctuation and casing, when generating embeddings. This allows the model to make use of syntactic and stylistic cues that contribute to meaning in context (Reimers and Gurevych, 2019).

3 Unsupervised Learning: Revisiting Clustering Analysis with K-Means

To explore latent thematic structure in the user response data, this project revisits clustering analysis—now enhanced through the use of contextual embeddings. Rather than relying on traditional vectorization techniques such as TF-IDF or bag-of-words, this approach utilizes sentence-level embeddings generated from the all-MiniLM-L6-v2 model in the Sentence-BERT framework (Reimers and Gurevych, 2019). These embeddings capture the semantic meaning of each response in context, enabling more nuanced similarity comparisons between observations.

With these embeddings, the K-Means clustering algorithm was applied to segment the responses into groups that represent underlying themes. K-Means is a centroid-based algorithm that partitions data into "K" clusters by minimizing within-cluster distances from the centroids. While the algorithm is commonly used in numerical data contexts, its use with contextual embeddings has become increasingly common in natural language processing due to their fixed-length vector structure and dense semantic representation (Wang et al., 2020; Jawahar et al., 2019). Unlike earlier representations (e.g., TF-IDF), contextual embeddings allow K-Means to group text not just by shared vocabulary, but by deeper semantic similarity.

In contrast to the previous project, which compared K-Means with Non-Negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA) on sparse document-term matrices, this final project focuses solely on K-Means clustering using contextual embeddings. This shift reflects a methodological pivot toward meaning-based representations, where the goal is not only to group responses by topic, but also to lay the groundwork for downstream quantitative analysis. In particular, the resulting clusters were used to generate semantic distances between each user and their assigned cluster centroid, which later served as input features to define a latent extremity score. Full parameter settings for such as cluster validation metrics (e.g., silhouette scores), and UMAP visualization are included in the appendix for reproducibility.

3.1 Clustering Performance Evaluation

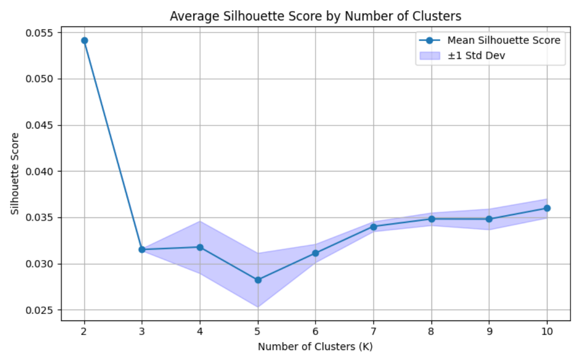


Figure 1: Average Silhouette Score of K-Means by Number of Clusters (Contextual Embeddings)

To evaluate the effectiveness of clustering using contextual embeddings, K-Means was applied across a range of cluster counts (K=2 to 10) using three different random seeds (207, 413, 437). Each model was assessed based on the average silhouette score across these random initializations. This metric measures how well-separated the resulting clusters are, with higher values indicating more cohesive and distinct clusters.

Across all conditions, the average silhouette scores remained relatively low, ranging from approximately 0.03 to 0.035. These values are consistent with the results from the previous project using TF-IDF vectorization with stopword removal via the Gensim stopword list, though slightly higher in general. However, they remain lower than the scores obtained when clustering with TF-IDF vectors without customized stopword filtering with CountVectorizer. This suggests that while contextual embeddings provide semantic richness, the clustering structure remains modest in separation—likely due to the nuanced and overlapping nature of open-ended text data.

The model with K=10 yielded the highest silhouette score with low variability, showing the best consistent performance among models with K>2, specifically with random state of 437 with a score of 0.0374. While the highest silhouette overall was observed at K=2, we selected K=10 as our focus for interpretability and thematic granularity.

3.2 Result Comparison

Cluster	Top Words
1	Desire, Motivation
2	Believe, Learn, Communication
3	Braver Angels
4	Help, Reduce, Heal
5	Sick, Struggle, Trouble
6	Presentation, Bill Doherty (Co-founder of Braver Angels)
7	Workshop
8	Division concern
9	Polarization concern
10	Recommendation

Table 1: Top topics for K-Means (Contextual Embeddings)

To interpret the clusters, the five responses closest to each cluster centroid (in embedding space) were extracted and manually reviewed. Each cluster was then assigned a descriptive topic label based on the shared themes of these representative sentences (see Table A2 to A11 in appendix). The resulting topics showed substantial overlap with those identified in the previous project, particularly in the model using K-Means with TF-IDF vectors and Gensim's predefined stopword list.

Cluster	Top Words
1	politics, people, polarization, want, talk
2	want, help, learn, people, country
3	work, civil, discourse, need, believe
4	country, division, need, divide, concern
5	angels, braver, work, attended, better
6	family, friends, members, political, want
7	political, divide, bridge, people, current
8	friend, recommended, told, recommendation, member
9	people, need, like, mission, believe
10	polarization, country, political, concerned, concern

Table 2: Top topics for K-Means (Gensim stopwords)

Visual inspection of the cluster structure was conducted using UMAP projections (see Appendix A1) with different combination of number of neighbors (5, 10 , 25, 50, 100) and random state (207, 413, 437). These plots revealed a consistent pattern: one dominant central cluster, with most others forming subgroups around it, and a notably separate Cluster 3. This structural insight aligns with the silhouette analysis, where K=2 produced the highest separation—suggesting a natural division between a core cluster and a distinct outlier group.

Interestingly, Cluster 3 was found to consist predominantly of responses that explicitly mentioned the organization's name, indicating a unique subgroup of respondents focused on identifying or endorsing the organization itself. This thematic separation provides additional support for the observed structural divergence in the embedding space.

4 Defining Latent Variable: "Extremity Level"

4.1 Linguistic Features

To investigate potential signals of political polarization in open-ended textual responses, we introduce a latent variable termed extremity level. This construct is intended to capture the rhetorical or emotional intensity embedded in users' responses to the prompt "Why I joined". The underlying assumption is that individuals with more politically extreme views may express themselves in more emotionally charged, assertive, or atypical ways compared to those with moderate or neutral views. Quantifying this expression enables us to test for associations between political lean and linguistic behavior—an indirect yet interpretable approach to evaluating the presence of polarization in self-reported motivations.

Sentiment reflects emotional tone, which may signal polarization intensity. We computed sentiment scores using the VADER sentiment analysis tool (Hutto and Gilbert, 2014), which produces a compound score representing the overall valence (positive or negative) of the sentence ranging from -1 to 1. Higher magnitude values (positive or negative) indicate stronger emotional expression. This feature captures whether highly emotional responses are more common among politically extreme respondents.

We quantified the proportion of absolutist and modal expressions in each response. These include words that express necessity, obligation, or certainty, such as must, should, need to, and always. Prior research has linked increased use of absolutist terms with mental rigidity, affective intensity, and polarized worldviews (Al-Mosaiwi and Johnstone, 2018). Moreover, political communication studies have shown that conservatives and liberals differ in their use of modal and absolutist language, suggesting it may reflect ideological extremity (Sylwester and Purver, 2015).

Rhetorical extremity may also manifest in linguistic complexity. We constructed a composite measure of complexity combining sentence length and type-token ratio (TTR), a common indicator of lexical diversity. TTR is calculated by dividing the number of unique words (types) by the total number of words (tokens) in the response. To ensure equal weighting, both features were standardized before combination:

$$\begin{aligned}&\text{Complexity}\\&= 0.5 \cdot \text{Sentence Length}_{Normalized}\\&+ 0.5 \cdot \text{Type Token Ratio}_{Normalized}\end{aligned}$$

Figure 2: Equation of Complexity

This formulation balances verbosity with vocabulary diversity. Prior work has suggested that linguistic complexity is associated with both cognitive style and ideological positioning, with simpler language often associated with more affect-driven or absolutist communication (Pennebaker et al., 2003; Graham et al., 2009).

Lastly, from the K-Means clustering results, we derived semantic distance features by computing the Euclidean distance between each user's sentence embedding and each of the ten cluster centroids. Each cluster represents a broad theme or motivational category (e.g., division concern, organization recommendation, polarization reduction). These distances serve two purposes: 1) To quantify how atypical a user's response is (i.e., distant from all central themes), and 2) To examine whether proximity to certain clusters is predictive of ideological leanings. This approach is supported by research suggesting that ideologically extreme messages often deviate semantically from mainstream discourse patterns and occupy unique areas in embedding space (Demszky et al., 2019).

By integrating these components, we formed a feature set that captures both emotional and semantic indicators of rhetorical extremity. These features were then combined using Principal Component Analysis (PCA) to derive a latent score, with the first principal component interpreted as the extremity level. This variable is used in downstream regression and hypothesis testing to evaluate whether political lean is associated with systematic variation in rhetorical intensity.

4.2 Principal Component Analysis

Principal Component Analysis (PCA) is a widely used technique in multivariate statistics for reducing dimensionality while preserving as much variance in the data as possible. By transforming the original variables into a new set of orthogonal components, PCA identifies the directions (principal components) along which the data vary the most (James et al., 2013). Each principal component is a linear combination of the original features, ordered by the proportion of variance it explains. The PCA was performed using the linguistic features discussed in 4.1, which are sentiment score, absolutist/modal language usage, complexity, and distance to each cluster centroids.

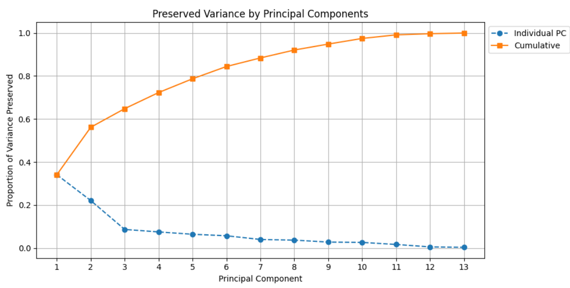


Figure 3: Percentage of Variance Preserved by Number of Principal Components (PC)

From Figure 3 and 4, interpreted as the extremity level, accounted for 34.1% of the total variance

$$
\begin{aligned}
&\text{PC1 (Extremity Level)}\\
&= -0.023 \cdot X_{\text{sentiment\_score}}\\
&- 0.058 \cdot X_{\text{absolutist\_modal\_freq}}\\
&- 0.066 \cdot X_{\text{complexity}}\\
&+ 0.348 \cdot X_{\text{dist\_cluster\_1}} + 0.378 \cdot X_{\text{dist\_cluster\_2}}\\
&+ 0.130 \cdot X_{\text{dist\_cluster\_3}} + 0.433 \cdot X_{\text{dist\_cluster\_4}}\\
&+ 0.433 \cdot X_{\text{dist\_cluster\_5}} - 0.037 \cdot X_{\text{dist\_cluster\_6}}\\
&+ 0.128 \cdot X_{\text{dist\_cluster\_7}} + 0.398 \cdot X_{\text{dist\_cluster\_8}}\\
&+ 0.376 \cdot X_{\text{dist\_cluster\_9}} - 0.136 \cdot X_{\text{dist\_cluster\_10}}
\end{aligned}
$$

Figure 4: Equation of first Principal Component (PC1)

in the original feature set. PC1 was driven primarily by semantic distance to specific cluster themes, particularly the distance to cluster 4 (help / reduce polarization / heal), 5 (sick / struggle / trouble), and 8 (division concern).

Delving into the equation more with Appendix A2 and A3, features with strong positive associations (increasing PC1 values) included distance to the clusters 1, 2, 3, 4, 5, 7, 8, and 9, which they represent themes such as personal motivation, civic dialogue, and deep concern about division.

In contrast, features with negative associations (decreasing PC1 values) included sentiment score, absolutist/modal language usage, complexity, and the distance to cluster 6 (presentation/Bill Doherty) and 10 (recommendation). This suggests that responses that are highly emotionally expressive, lexically intense, and linguistically simple may be situated at one end of the extremity spectrum, whereas responses that are semantically distant from core motivational themes (e.g., expressing concern or struggle) are positioned at the other. This latent dimension offers a continuous and interpretable scale of rhetorical extremity to be used in subsequent regression analysis and hypothesis testing.

5 Regression Analysis: Association Between Extremity and Political Lean

5.1 Translating Research Question to Mathematical Expression

To evaluate whether the extremity level, a latent variable derived from linguistic behavior, is associated with political identity, we conduct a multiple linear regression analysis using participants' self-reported political lean and cluster label (from the K-Means analysis) as predictors. The central research question is whether linguistic extremity systematically varies by political affiliation, controlling for the type of motivational theme captured by cluster membership.

Linear regression models the expected value of a continuous outcome variable as a linear function of explanatory variables. In this context, the response variable is the first principal component (PC1), interpreted as the extremity level. The predictor variables include categorical indicators of political lean, which are Blue, Red, or Other, and K-Means cluster labels (from 1 to 10), which represent semantically distinct categories of joining reasons. This approach allows us to quantify whether individuals with different political affiliations express themselves with systematically different rhetorical or emotional intensity in open-ended responses.

Linear regression is an appropriate method here as it provides a flexible framework for estimating conditional mean differences and testing for statistical significance under well-established assumptions. Its interpretability and capacity to control for multiple covariates make it particularly suitable for evaluating relationships between latent traits and observed group memberships (James et al., 2013).

Assume Y = Extremity Level, L = Political Lean (Blue/Red/Other), and C = Cluster (1 to 10. The regression model can be formally expressed as:

$$
\begin{aligned}
&E[Y \mid L, C]\\
&= \beta_0 + \beta_1 \cdot \text{Lean}_{\text{Other}} + \beta_2 \cdot \text{Lean}_{\text{Red}}\\
&+ \sum_{i=2}^{10} \beta_{i+1} \cdot \text{Cluster}_i
\end{aligned}
$$

Figure 5: Equation of Linear Regression Model

The baseline (reference) groups are $\text{Lean}_{\text{Blue}}$ and Cluster_1. and all other levels are encoded as binary indicator variables. The coefficients β_1 and β_2 capture the mean difference in extremity level between Blue and the other lean categories, after controlling for semantic clustering.

5.2 Model Coefficients & Interpretation

The fitted linear regression model quantifies how the latent variable extremity level varies as a function of political lean and semantic cluster membership. The estimated model is:

$$
\begin{aligned}
&E[Y \mid L, C] \\
&= 0.56 - 0.02 \cdot \text{Lean}_{\text{Other}} + 0.07 \cdot \text{Lean}_{\text{Red}} \\
&- 1.68 \cdot \text{Cluster}_2 - 0.30 \cdot \text{Cluster}_3 \\
&- 2.04 \cdot \text{Cluster}_4 - 1.50 \cdot \text{Cluster}_5 \\
&+ 2.46 \cdot \text{Cluster}_6 + 0.65 \cdot \text{Cluster}_7 \\
&- 1.44 \cdot \text{Cluster}_8 - 2.46 \cdot \text{Cluster}_9 \\
&+ 3.17 \cdot \text{Cluster}_{10}
\end{aligned}
$$

Figure 6: Equation of Estimated Linear Regression Model

At the conventional significance level of $\alpha = 0.05$, all coefficients except for $\text{Lean}_{\text{Other}}$ (p-value = 0.553) are statistically significant. This suggests that, holding cluster constant, Red-leaning users exhibit a small but significantly higher predicted extremity level compared to Blue-leaning users, while Other-leaning users do not differ significantly from the Blue group.

For interpreting the coefficients in the model, the intercept $\beta_0 = 0.56$ represents the expected extremity level for a user in Cluster 1 who identifies as Blue after controlling for all other variables. For Red-leaning users, the model predicted to score 0.07 units higher on the extremity level scale than Blue-leaning users after controlling for cluster membership, on average. On the other hand, the expected extremity level for Other-leaning users is 0.02 units lower than Blue-leaning users.

Similarly, each cluster coefficient reflects the expected difference in extremity level between that cluster and the baseline cluster (Cluster 1), after holding political lean fixed. While Cluster 6, 7, and 10 yield higher extremity level scale compared to Cluster 1, all other clusters are expected to have lower predicted scale than the baseline cluster. While Cluster 10 (Recommendation) has the largest increase (3.17), which can be understood that users in Cluster 10 are the most rhetorically extreme in expression, relative to Cluster 1, Cluster 9 (Polarization concern) decreases the extremity level the most compared to Cluster 1 (2.46). This result aligns with the finding from the equation of extremity level defined with PCA loadings. For instance, the variable "Distance to Cluster 10", which corresponds to responses centered on neutral content such as recommendations (e.g., "a friend recommended"), has a negative loading in the PCA. This means that users whose responses are farther from Cluster 10 tend to have lower values on PC1. Because Cluster 10 represents semantically neutral and emotionally minimal responses, greater distance from it suggests less neutrality and thus more likely to have rhetorical extremity. This interpretation is reinforced by the regression model: users in Cluster 9, which centered on concern about polarization, have a large negative coefficient, indicating they tend to have lower extremity scores (PC1). Together, these patterns support the conclusion that lower PC1 values correspond to more emotionally intense or rhetorically extreme responses, consistent with the conceptual design of the latent variable and revealing the direction of it.

5.3 Linear Regression Condition Check

Before interpreting the results of a linear regression model, it is essential to assess whether the fundamental assumptions of the method are satisfied. The following five standard conditions that underpin the validity of inference and interpretation in linear regression models are evaluated.

Linearity: The relationship between each explanatory variable and the outcome variable is assumed to be linear in form. This ensures that the model captures the correct functional form and does not systematically misestimate the effect of predictors.

Independence: The residuals (errors) should be independent of one another. Violations of this condition (e.g., due to repeated measures or grouping structures) can lead to underestimated standard errors and inflated Type I error rates.

Normality: The residuals are assumed to be normally distributed. This assumption primarily affects the validity of t-tests, F-tests, and confidence intervals, especially in smaller samples.

Equal Variance (Homoscedasticity): The variance of the response variable should be constant across all levels of the predictor variables. Heteroscedasticity can lead to inefficient estimates and unreliable hypothesis tests.

Multicolllinearity: The predictor variables should not exhibit strong linear relationships among themselves. High multicollinearity inflates the standard errors of the estimated coefficients, making it difficult to assess the individual effect of each predictor.

Checking these conditions is critical for two main reasons. If the model assumptions hold, many statistical tests such as t-tests, F-tests, and confidence intervals are trustworthy and therefore able to make a valid inference. Also, the estimated regression coefficients assume a linear and additive effect on the response variable, and these coefficients may not reflect true relationships when assumptions are violated (Kutner et al., 2005).

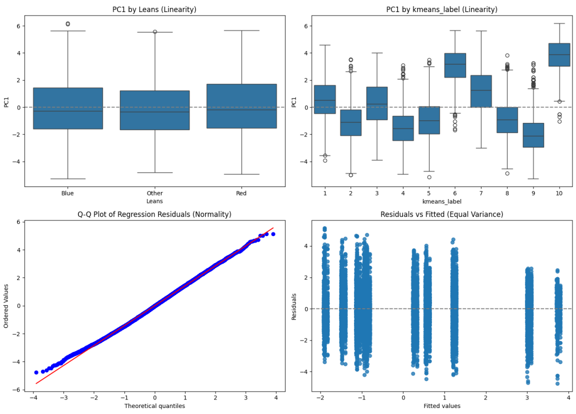


Figure 7: Regression Condition Visualization (Linearity (1st row) & Normality / Equal Variance (2nd row))

Visualizations and relevant measurements were utilized to assess each of the five core assumptions. For linearity, we examined the distribution of the response variable (PC1, the extremity level) across categorical predictors using boxplots. The boxplots by Leans were relatively similar, suggesting a weak or negligible linear association between political leaning and extremity level. In contrast, boxplots by K-Means cluster labels revealed more visibly distinct distributions, indicating a potentially stronger and more meaningful relationship between semantic cluster membership and the latent outcome. These results suggest that linearity is approximately satisfied, particularly for the cluster predictor.

For independence of error, duplicate responses of "Why I joined" in the data were removed to reduce the risk of repeated measurements from the same individual. While the cleaned dataset likely satisfies the independence assumption across users, the responses may still exhibit temporal autocorrelation due to underlying political events occurring at specific points in time. For instance, the dataset includes "Member Start Date" variable, which could reflect waves of interest driven by national political moments. To investigate this, we inspected residuals over time and performed the Durbin–Watson test, a diagnostic tool for autocorrelation in residuals. The test yielded a value of approximately 1.95, suggesting weak or no evidence of autocorrelation, and thus supporting the assumption of independence (Durbin and Watson, 1950).

We examined the normality of residuals using a Q–Q plot. While the residual distribution is not perfectly aligned with the reference line, the majority of the points fall close to the diagonal, especially in the central portion of the distribution. This suggests that the residuals are approximately normally distributed, satisfying the assumption sufficiently for valid inference.

Fitted vs residual plot is the most widely used visualization to check equal variance. The residuals appeared to maintain a consistent spread across the range of fitted values, forming a roughly rectangular shape between -4 and 4. This pattern suggests that the variance of the errors is approximately constant. However, some vertical striping in the residual plot is present, likely due to the discrete nature of the cluster labels, which groups observations by cluster. While minor heteroscedasticity might be present, we could view the condition as no strong indication of systematic unequal variance.

Lastly, we computed the Variance Inflation Factor (VIF) for each predictor to check multicollinearity. A general rule of thumb is that VIF values above 5 (or 10) may indicate problematic multicollinearity (Montgomery et al., 2021). In this analysis, all VIF values were well below 2, indicating that the predictors are not excessively correlated with one another and that this assumption is well satisfied.

Overall, the regression model meets the key assumptions reasonably well, allowing for valid inference and interpretation of the extremity level as a function of political lean and semantic cluster.

5.4 Hypothesis Testing for Association

As described in the project's objective, identifying whether political polarization manifests in users' written responses is the central research goal. Specifically, we aim to determine whether political lean is associated with the extremity level, a latent variable constructed from linguistic features. To address this, we formally test two hypotheses.

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_1 : \text{At least one } \beta_i \neq 0 \text{ for } i \in \{1, 2\}$$

To test this hypothesis, we apply an F-test through Analysis of Variance (ANOVA), which evaluates whether the categorical predictor political lean has a significant effect on the response variable after accounting for all other covariates in the model (e.g., cluster labels). ANOVA is a standard technique for comparing group means in a linear regression framework and testing nested models (Kutner et al., 2005).

Variable	Sum Sq	df	F	p-value
C(Leans)	11.3801	2	2.8000	0.0608
C(kmeans_label)	34979.1237	9	1912.5487	<0.0001
Residual	29624.5571	14578	–	–

Table 3: ANOVA Table: Effects of Political Lean and Cluster on Extremity Level

Based on the summary table 3, the p-value is slightly greater than the significance threshold (0.0608), and thus fail to reject the null hypothesis. This indicates that there is no statistically significant association between political lean and extremity level when controlling for cluster labels. However, since the p-value is close to 0.05, suggesting weak evidence of a potential association that may merit further investigation with additional data or refined measures.

On the other hand, the second hypothesis touches slightly different direction.

$$H_0 : \beta_1 = \beta_2$$
$$H_1 : \beta_1 \neq \beta_2$$

To examine pairwise group differences, we conduct multiple comparisons using Tukey's Honest Significant Difference (HSD) procedure, which controls the family-wise error rate (FWER) when performing all pairwise tests (Montgomery et al., 2021). The comparison includes "Blue vs Red", "Blue vs Other", and "Red vs Other". The Tukey HSD test results are summarized below:

Group 1	Group 2	Mean Diff	p-adj	Lower	Upper	Reject
Blue	Other	-0.0745	0.35	-0.2007	0.0518	False
Blue	Red	0.1512	0.0072	0.0337	0.2687	True
Other	Red	0.2256	0.0025	0.0670	0.3843	True

Table 4: Tukey HSD Multiple Comparisons of Extremity Level by Political Lean

The mean differences in extremity level between Red and both Blue and Other groups are statistically significant, indicating that Red-leaning users tend to exhibit more rhetorically extreme responses. However, the difference between Blue and Other is not significant, which likely explains why the first hypothesis test did not yield statistical significance.

5.5 Characteristics of "Typical" Users Representing Each Lean

As an additional analysis, we examined linguistic characteristics of "typical" users from each political lean by selecting five responses per group whose extremity level (PC1) values were closest to that group's mean. We then computed summary statistics of several key features — including sentiment score, absolutist/modal language proportion, sentence length, type-token ratio, and the combined complexity metric.

Feature	Blue	Other	Red
Sentiment Score	0.271 (0.529)	0.264 (0.529)	0.119 (0.575)
Absolutist/Modal Prop.	0.0091 (0.0203)	0.0000 (0.0000)	0.0043 (0.0095)
Sentence Length	18.0 (19.43)	16.6 (8.14)	35.2 (16.68)
Type-Token Ratio	0.917 (0.125)	0.929 (0.081)	0.806 (0.061)
Complexity	−0.102 (0.158)	−0.080 (0.319)	−0.157 (0.645)

Table 5: Linguistic Characteristics of Typical Users by Political Lean (Standard Deviation in parentheses)

Several notable patterns can be detected from the table 5 and appendix A12, A13, and A14. First, Red-leaning users tend to use much longer sentences, averaging nearly double the length of Blue and Other groups. This is accompanied by a lower type-token ratio, indicating relatively less lexical diversity, and the lowest sentiment scores, suggesting their tone tends to be more emotionally neutral or negative. They also show greater variance in complexity and usage of absolutist/modal terms.

In contrast, Blue and Other users are more similar to each other, with nearly identical sentiment scores and type-token ratios, and lower sentence lengths and complexity than Red users. Notably, Other-leaning users exhibited no absolutist/modal language usage in the selected responses, though this may be due to the small sample size.

These findings offer a linguistic snapshot of how different political leans may express themselves in their reasons for joining the organization — a potentially valuable layer in understanding political discourse beyond just topical content.

6 Discussion

6.1 Conclusion

This project set out to investigate whether political polarization could be detected through the open-ended "Why I joined" responses submitted by members of a nonprofit organization. Building on previous work, we incorporated contextual embeddings using a pre-trained Sentence-BERT model to revisit clustering analysis and introduced a novel latent construct—extremity level—to quantify rhetorical intensity in user responses.

In the first analysis, we applied K-Means clustering to the contextual embeddings and identified ten thematic clusters. The results showed interpretable groupings (e.g., concern about polarization, workshop participation, organizational recommendation), and dimensionality reduction using UMAP supported these structural patterns. Compared to clustering on TF-IDF features from the prior project, the contextual embeddings yielded a similar but slightly clearer segmentation of users' motivations.

In the second part, we defined a latent variable extremity level by combining linguistic features such as sentiment, use of absolutist/modal language, complexity, and semantic distance from cluster centroids. Applying Principal Component Analysis (PCA), we extracted the first principal component as a proxy for rhetorical extremity. We then fit a linear regression model to test whether this extremity level was associated with political lean, while controlling for semantic clusters.

While the overall regression model was statistically significant, the omnibus ANOVA F-test for political lean was marginally non-significant ($p = 0.06$), suggesting weak evidence for a direct association. However, pairwise comparisons via Tukey's HSD revealed that Red-leaning users expressed significantly higher extremity than both Blue and Other users. These findings suggest that while the broad relationship between lean and extremity may be nuanced, certain political identities are linked to more rhetorically intense language in voluntary political engagement.

6.2 Limitation

While the findings from this project offer valuable insights into political expression and polarization through language, several limitations should be acknowledged.

First, the dataset is skewed in terms of political lean, with a dominant number of users identifying as Blue. This imbalance may reduce the generalizability of results and diminish the statistical power when comparing groups, particularly when estimating differences involving Red or Other-leaning users.

Second, although contextual embeddings from general-purpose sentence transformers were used effectively, we were unable to locate a pre-trained language model specifically fine-tuned on political discourse for sentence similarity tasks. Such a model could potentially yield more accurate clustering and feature extraction in politically nuanced language and improve downstream inference.

Lastly, while the latent variable "Extremity level" was carefully constructed using multiple linguistic indicators, it remains a proxy measure. Its relationship to political polarization is indirect, and it is possible that extremity in rhetoric does not fully capture deeper ideological or affective polarization. Future work might refine this construct or evaluate it against behavioral or attitudinal ground truths to assess its validity.

6.3 Future work

Several promising directions remain for extending this analysis.

First, incorporating the "Profession" feature could provide valuable insight into how occupational identity correlates with political extremity. However, this was not pursued due to challenges in preprocessing; many responses were written as free-form descriptions rather than consistent labels (e.g., "retired teacher," "work in public health," etc.). Future work could explore methods such as occupation classification models or rule-based entity extraction to normalize and categorize these responses. Given that prior research has established links between profession and political ideology, this feature could significantly enrich the analysis.

Plus, the inclusion of geographic information, such as users' ZIP codes or states, would offer another dimension to investigate political polarization. Regional differences are well-documented in political behavior and could allow for spatial analysis or mapping of linguistic extremity. Unfortunately, such data were not available in the current dataset, but if obtained, it could help uncover how geography intersects with rhetorical expression and political alignment.

Acknowledgement

I would like to thank Professor Jonathan Dunn for the lectures, insightful guidance, constructive feedback, and encouragement for the semester-long project. His expertise in corpus linguistics and commitment have been invaluable in shaping the direction and depth of this work.

References

Mohammed Al-Mosaiwi and Tom Johnstone. 2018. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 6(4):529–542.

Christopher A. Bail, Lisa P. Argyle, and Taylor W. Brown. 2021. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 118(50).

Pew Research Center. 2014. Political polarization in the american public.

Dorottya Demszky, Nikhil Garg, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. In *Proceedings of NAACL*, pages 2970–3005.

David Dorn, Gordon Hanson, and David Autor. 2020. When work disappears: Manufacturing decline and the falling marriage market value of young men. *National Bureau of Economic Research*.

James Durbin and Geoffrey S Watson. 1950. Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409–428.

Gallup. 2021. Presidential job approval ratings from ike to biden.

Jesse Graham, Jonathan Haidt, and Brian A Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5):1029–1046.

C.J. Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 216–225.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2018. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1113–1122.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer.

Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What does bert learn about the structure of language? In *Proceedings of ACL*, pages 3651–3657.

Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li. 2005. *Applied Linear Statistical Models*, 5th edition. McGraw-Hill/Irwin.

Douglas C Montgomery, Elizabeth A Peck, and G. Geoffrey Vining. 2021. *Introduction to Linear Regression Analysis*, 6th edition. Wiley.

James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Trends in Cognitive Sciences*, 7(12):547–551.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3982–3992.

American National Election Studies. 2020. Anes time series studies.

Karolina Sylwester and Matthew Purver. 2015. Twitter language use reflects psychological differences between democrats and republicans. *PLOS ONE*, 10(9):e0137422.

Wenhui Wang, Furu Wei, Li Dong, Hang Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:5776–5788.

A Appendix

Model	Transformer	Input Parameter
all-MiniLM-L6-v2	Sentence (BERT)	'max_seq_length'=256 'do_lower_case'=False 'word_embedding_dimension'=384 'pooling_mode_cls_token'=False 'pooling_mode_mean_tokens'=True 'pooling_mode_max_tokens'=False 'pooling_mode_mean_sqrt_len_tokens'=False 'pooling_mode_weightedmean_tokens'=False 'pooling_mode_lasttoken'=False 'include_prompt'=True

Table A1: Input parameters of language model for contextual embeddings generation

Cluster 1 – Representative Responses
• I am motivated to step into the work to bridge the divides separating human beings from one another to bring about greater peace and community cohesion. • Desire to bridge divides and create better understanding and engagement • Your attempt to bring sides together to build understanding and cooperation • Desire to build understanding and be a peacemaker, to bring people together • My life's work has centered on bringing people together and building trusted relationships to create change in our world.

Table A2: Representative responses closest to Cluster 1 centroid

Cluster 2 – Representative Responses
• I want to learn to facilitate civilized, productive conversations between people with opposing points of view. • Wanting to learn how to have civil conversations with others who have different views • Really believe in civil discourse and want to be part of the solution; want to have ways to hear others perspectives in a meaningful way • Want to be part of conversations that bring people with different political views together and keep them connected in some way in a time of extreme polarization. I believe my work as a therapist has given me skills to play a helpful role in these types of • I have become very interested in improving communication between people who disagree.

Table A3: Representative responses closest to Cluster 2 centroid

Cluster 3 – Representative Responses
• I have appreciated being involved with Braver Angels for several years. • I have been a Braver Angels member in the past. I have been active in my community promoting and participating in Braver Angels initiatives. • I have been trying to reach out across our political divide for years but was not aware of Braver Angels until a couple of years ago. • I have been trying on my own to promote respectful political discourse, so when I read about Braver Angels in the NYT today, I decided to join. • Conversations with a member of Braver Angels

Table A4: Representative responses closest to Cluster 3 centroid

Cluster 4 – Representative Responses
• I want to help reduce the division in our country, restore civil discourse and discussion of the issues, and get back to the foundation of what makes us Americans and one nation.
• Wanting to be part of the solution to bring people together to heal the increasing divide and polarization among us in this country and in my community.
• I want to strengthen our democracy and contribute to healing the toxic divides in our nation.
• Wanting to help bridge the divide and save our democracy
• Interested in helping to unite America, and heal the political divide.

Table A5: Representative responses closest to Cluster 4 centroid

Cluster 5 – Representative Responses
• The growing partisan divide and animosity within families & communities; plus my own need to listen & speak with those of differing views.
• I am sickened by the political polarization in our country, media & community dialog and want to learn to have and foster dialog that is respectful of differences and able to reach compromise for the benefit of all.
• I am disturbed by the growing lack of civil discourse in the US today, and the inability and in some cases unwillingness of people to really listen and try to understand others' points of view. I would like to cross the partisan divide and reach greater
• One of your communication tools was recently shared with me. I have struggled to talk calmly to friends and family about political issues. I want to be part of the solution to bridge the divide yet I am finding it challenging.
• I, too, am troubled by growing political polarization. I am a trained facilitator and think my skills might be helpful. I also lead a group in Liberty Twp, OH, called the Community Forum. I moderate weekly sessions on the tough issues.

Table A6: Representative responses closest to Cluster 5 centroid

Cluster 6 – Representative Responses
• I respect the work for do and what prompted joining today was hearing one it is founders (Bill Doherty, I think) interviewed by Dan Harris for the Ten Percent Happier podcast.
• Presentation by Bill Doherty
• Presentation by Bill Doherty
• Heard an interview on NPR
• Agree with the mission and heard an interview on NPR with John Wood

Table A7: Representative responses closest to Cluster 6 centroid

Cluster 7 – Representative Responses
• Did a workshop and have always been interested in the organization.
• The division in our nation. I have attended a couple of workshops in my community. Not sure how I want to be involved.
• I attended a workshop recently and was interested in membership after that. I am a dialogue facilitator and am interested in contributing in some way to depolarizing our political and civic culture
• Heard about the organization on a podcast I listen to, a friend invited me to attend a debate and a workshop. Both were positive experiences.
• I love this organization and tout its benefits to many I encounter. I believe I joined a few years ago/ want to continue to support this essential effort!

Table A8: Representative responses closest to Cluster 7 centroid

Cluster 8 – Representative Responses
• Concerned about the divide in our country.
• I am profoundly concerned about the political division in this country.
• Concern over the increasing deep divide our nation suffers from
• The current political division in our country
• Concern about the political divide in our country

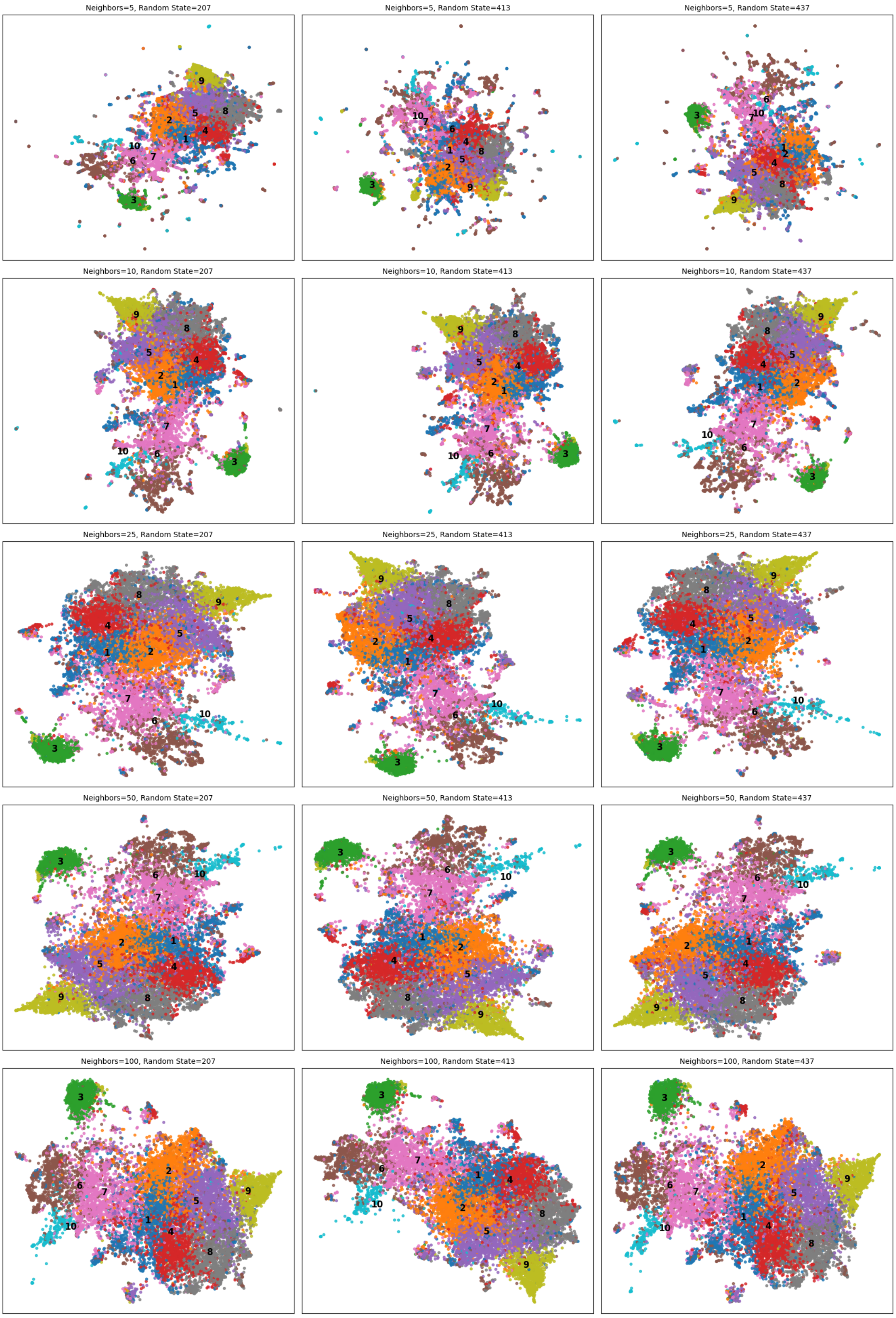
Table A9: Representative responses closest to Cluster 8 centroid

Cluster 9 – Representative Responses
• Our political climate of polarization
• I am passionate about mitigating polarization in America
• I am concerned about the prevalence of polarization in America and think my perspective could be valuable in bridging the divide.
• I see our polarization as the biggest problem facing this country and want to improve my ability to work against it.
• Deep concern over the increasing polarization in our country and world.

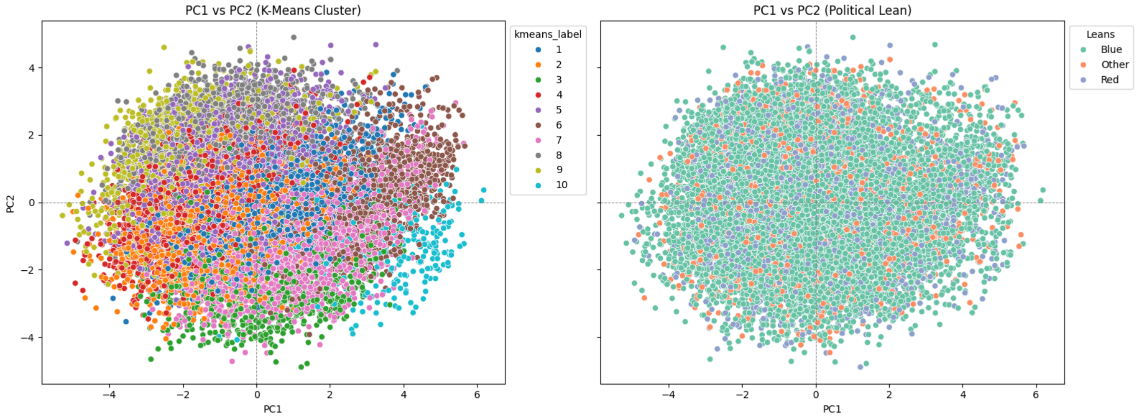
Table A10: Representative responses closest to Cluster 9 centroid

Cluster 10 – Representative Responses
• A friend recommended
• A friend recommended.
• A friend’s recommendation
• A friend’s recommendation
• A friend’s recommendation

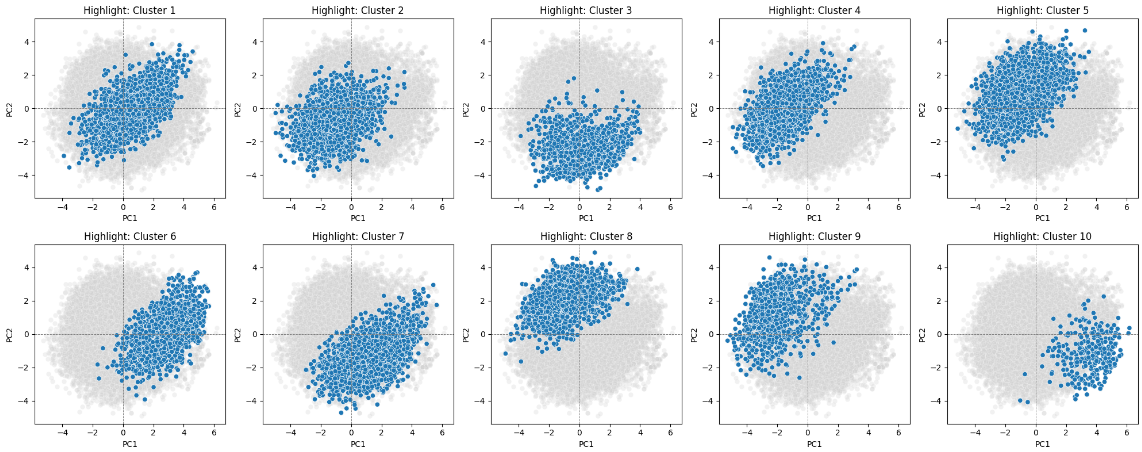
Table A11: Representative responses closest to Cluster 10 centroid



Appendix A1: UMAP Visualizations (Contextual Embeddings)



Appendix A2: Visualizations of PC1 vs PC2 by Cluster Label and Lean



Appendix A3: Visualizations of PC1 vs PC2 Highlight by Cluster Label

Representative Responses of Blue-Leaning Users (5 Closest to Average Extremity Level)
• We are the United States, free to disagree, but not to bully. We need to hear each other's points of view respectfully.
• Support the mission of finding ground and unity.
• I believe in the power of nonviolent communication. I am learning about it now. Then I was invited to check out Braver Angels by a member of an interspiritual group I am a member of. I looked carefully at the website and watched the video. I was sold. I t
• Politics!
• my fear about the possible demise of our democracy

Table A12: Typical responses closest to Blue average extremity level

Representative Responses of Other-Leaning Users (5 Closest to Average Extremity Level)
• I am a civically engaged singer/songwriter actively writing and releasing music about community life and engagement. I look forward to being part of the Braver Angles community.
• Trepidation watching the fear & divisiveness industry's momentum.
• To support a group that works on bringing people together.
• Absolute concern over the state of division and the vehemence of that division expressed everywhere in the country at this time.
• My organization participated in a debate on sex work and sex traffic. I Have followed Braver Angels since.

Table A13: Typical responses closest to Other average extremity level

Representative Responses of Red-Leaning Users (5 Closest to Average Extremity Level)
• the importance of bringing the country back to the center and away from the extremes
• Just the tenets of cooling the temperature and getting the legislative branch back to legislating instead of tv stars.
• I am tired of watching friends unfriend each other on social media because of different political views. I am tired of being afraid of losing my job, my clients, or my friends because I am not free to express my point of view without retaliation. The Medi
• I am passionate about bridge building and getting to know and love people who think/believe differently than I do. I want to learn more skills/tools for helping myself and others be better at this. I am also working to develop a TV series that will help mo
•Was a member of BA after the Riots and pandemic and our Nation needs help now more than ever. Iv'e been involved in many orgs before but the partisan divide is not helping humans and the lack of goodwill is on life support. Reached out about the Ambassad

Table A14: Typical responses closest to Red average extremity level