Analyzing Partisan Framing of Immigration in U.S. Presidential Speeches Using Structural Topic Modeling and Word Embeddings

**Analyzing Partisan Framing of Immigration in U.S. Presidential Speeches**

**Overview**

This project analyzes how U.S. presidents from different political parties (Democratic and Republican) have framed immigration in their speeches from 1789 to 2023. The project uses Natural Language Processing (NLP) techniques, including Structural Topic Modeling (STM) and word embeddings, to explore the differences in how each party discusses immigration.

**Methods**

1. **Data Collection and Preprocessing**
   * **Data Source**: The dataset contains presidential speeches from various U.S. presidents. The speeches were tokenized using the quanteda package, which breaks the text into words (tokens), while removing punctuation, symbols, numbers, and separators.

toks <- tokens(pres\_df\_stm$clean\_text, remove\_punct = TRUE, remove\_symbols = TRUE, remove\_numbers = TRUE, remove\_separators = TRUE)

* + **Stopword Removal**: Common English stopwords (e.g., "the", "and") were removed to focus on the most meaningful words. Additionally, words with fewer than three characters were excluded.

toks\_nostop <- tokens\_select(toks, pattern = stopwords("en"), selection = "remove", min\_nchar = 3)

* + **Bigrams and Frequency Count**: For deeper insight, bigrams (pairs of adjacent words) were analyzed to understand commonly co-occurring word pairs in the speeches. The analysis highlights the most frequent bigrams used by each president.

pres\_df %>%

unnest\_tokens(bigram, transcript, token = "ngrams", n = 2) %>%

filter(!word1 %in% stop\_words$word, !word2 %in% stop\_words$word) %>%

count(president, bigram, sort = TRUE)

1. **Structural Topic Modeling (STM)** STM was used to explore the topics in presidential speeches, particularly how the topics evolved over time. The STM model helps identify topics that commonly appear in speeches and tracks the prevalence of these topics over different decades.
   * **Model Building**: After cleaning and preparing the data, STM was applied to the subset of speeches by recent presidents (Biden, Trump, Obama, etc.). This approach allows for temporal and topic-based analysis of the speech content.

pres\_stm <- stm(docs, vocab, K = 0, prevalence = ~ s(date\_clean), data = meta, init.type = "Spectral")

* + **Topic Effects Over Time**: The estimateEffect function was used to estimate the effect of time (date) on the prevalence of different topics.

pres\_stm\_effects <- estimateEffect(1:73 ~ s(date\_clean), stmobj = pres\_stm, metadata = meta, uncertainty = "None")

1. **Word Embeddings and Context Analysis** To focus specifically on immigration, word embeddings were used to explore the contextual differences in how the terms "immigration", "immigrants", and "immigrate" are discussed by different political parties. This approach allows for a deeper understanding of the ideological differences between the Democratic and Republican parties.
   * **Contextual Tokens**: A window of 8 words was used around immigration-related terms to capture the context in which these terms appear.

immig\_toks <- tokens\_context(x = toks, pattern = c("immigration", "immigrants", "immigrant", "immigrate"), window = 8L)

immig\_dfm <- dfm(immig\_toks)

* + **Cosine Similarity**: Cosine similarity was calculated between the Democratic and Republican parties’ use of immigration-related words, allowing for a comparison of the semantic differences between the two parties.

immig\_cos\_sims <- cos\_sim(immig\_party, pre\_trained = cr\_glove\_subset, features = immig\_party@features, as\_list = FALSE)

**Key Findings**

1. **Partisan Framing of Immigration**:
   * **Republican Party**: The analysis revealed that Republican presidents are more likely to use words like "border", "enforcement", "homeland", and "patrol", framing immigration in terms of **security** and **law enforcement**.
   * **Democratic Party**: Democratic presidents, on the other hand, frequently use words such as "trillion", "spending", and "taxes", suggesting an emphasis on **economic** discussions related to immigration.
2. **Cosine Similarity Analysis**:
   * The cosine similarity between the word embeddings for each party demonstrated a significant difference in the way the two parties frame immigration. Republican presidents' discussions are more focused on **border control**, while Democratic presidents link immigration to **economic policy** and **public spending**.
3. **Visualizing Partisan Differences**: The visualization of cosine similarities revealed clear ideological divides, with certain words strongly associated with either party. The Republican Party, for instance, emphasizes immigration enforcement, while the Democratic Party discusses immigration in the context of taxes and the economy.

**Conclusion**

This project highlights the stark differences in how Democratic and Republican presidents have framed the issue of immigration in their speeches. By using STM and word embeddings, this analysis provides insight into the evolving nature of immigration rhetoric and how partisan ideologies shape the discourse. This divergence in framing is consistent with broader ideological differences, with Republicans viewing immigration primarily as a security issue and Democrats framing it in terms of economic impact and reform.