Derrick Yong, Alonso Hernandez, Florian Henrichs, Gianna Shen

Professor Tekgurler

Digital Humanities 101

01 July 2025

Proposed Amendments to the US Constitution from 1787 to 2014

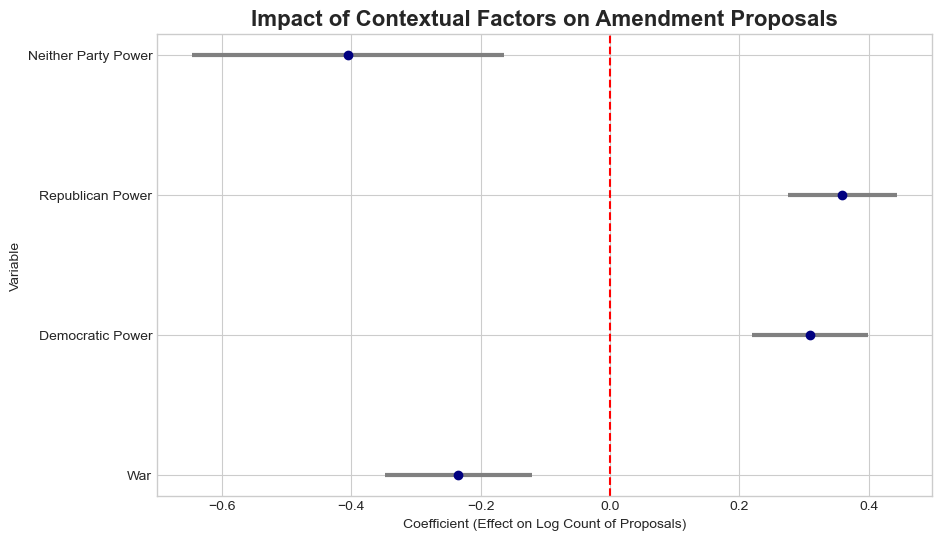
**Research Questions (Derrick):** How have the dominant themes of proposed amendments to the The Constitution changed over time from 1787 to 2014? What do proposed amendments reveal about the political and social evolution in the United States? Which ideas were dominant or suppressed during specific historical periods based on the context of proposed amendments?

**Data (Derrick):** We used the [dataset](https://www.archives.gov/open/amending-america/us-nara-amending-america-dataset-raw-2016-02-25.csv) *Amending America: Proposed Amendments to the United States Constitution, 1787 to 2014* sourced from the [National Archives](https://www.archives.gov/open/dataset-amendments). It contains more than 11,000 proposed amendments to the constitution from 1787 to 2014. For each amendment, the data includes sub sections of: source of information, title or description, congressional details, sponsor information, and committee of referral. It is a CSV (comma-separated values) file. In addition, a data dictionary is included to explain each subsection in the data.

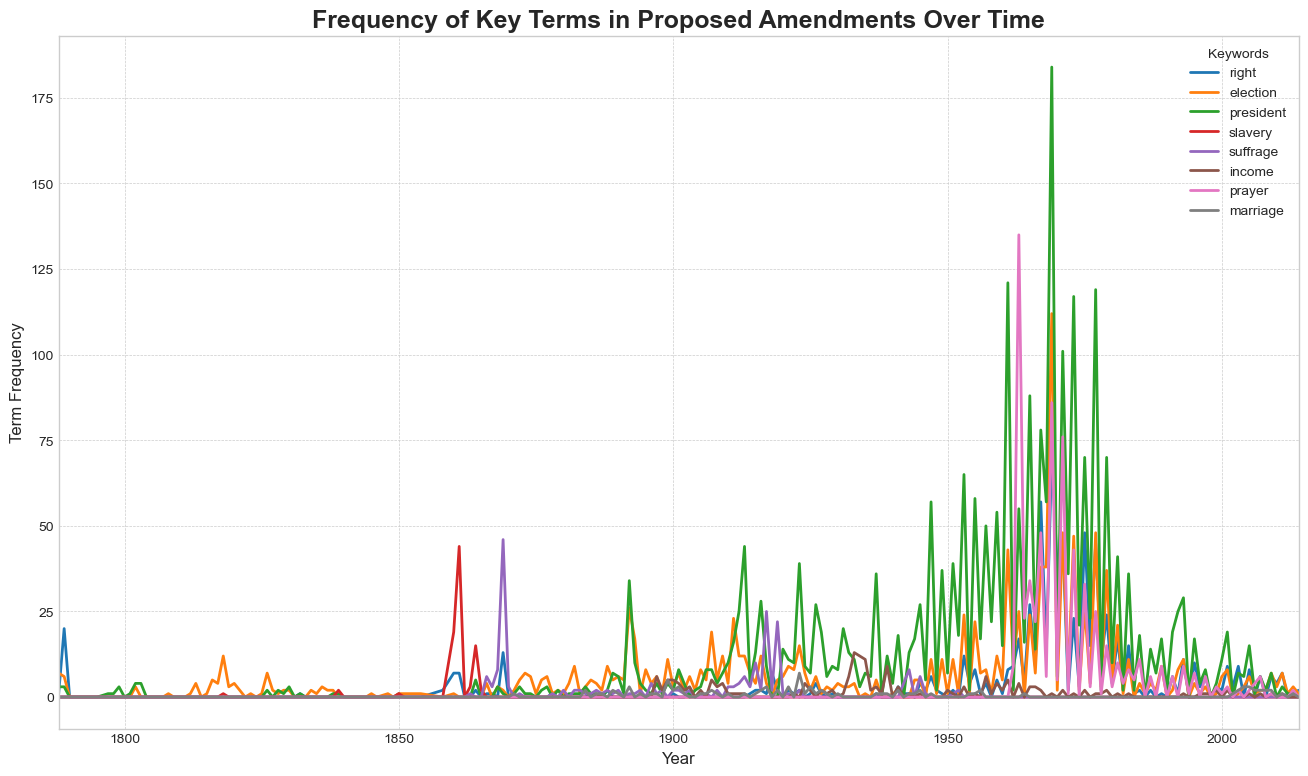
**Methods and Approaches (Alonso):** For this project, we did a focus on using Natural Language Processing and Text Analysis. Once we loaded the dataset, we first preprocessed and cleaned it up by removing unnecessary data and also made a new dataset that removed the duplicates within the original. This was done by tokenizing the description of each proposed amendment and only retrieving the latest unique entry from those tokens, removing all others before that were duplicates. This managed to reduce the dataset from 11,797 to 4,627 entries. These tokens were manipulated further. They were stripped of stop words so common English words that did not tell much of what the amendment was about were removed. Then, they were lemmatized to help with grouping of words. Using these new tokens, the frequency of each word was identified within all the proposed amendments which was made into a TF-IDF. This was used to identify words that were unique and important to specific amendments. In a different analysis, the dataset was grouped by decade. Another dataset was merged with the original here to show whether the United States was at war and which political party held power at a specific decade. When determining which political party was in power during each decade, we identified the political affiliation of the U.S. president at the time. If a decade included presidents from more than one party, we assigned the party of the president who served the majority of that decade. Using this information, a linear regression model was first constructed, followed by a Poisson Regression after not achieving satisfactory results. We chose Poisson Regression as the second attempt after realizing our proposal counts may not be normally distributed, and a Poisson distribution is suitable because our outcome (proposal counts) is going to be non-negative integers.

**General Reasoning or Discussion (Forian/Derrick):** Since the goal of our project was to uncover political and social trends in the United States at the time of a proposed amendment, we used Natural Language Processing (NLP) and Text Analysis with various statistical such as the Poisson Regression modeltec to yield data that would be quantitatively larger and more relevant. By working with a smaller and more relevant dataset, this allowed us to gain insight into how amendment proposals increase when the nation is dominated by a single political party during a presidential administration and decrease when the nation is at war due to focuses on more pressing issues like national defense.

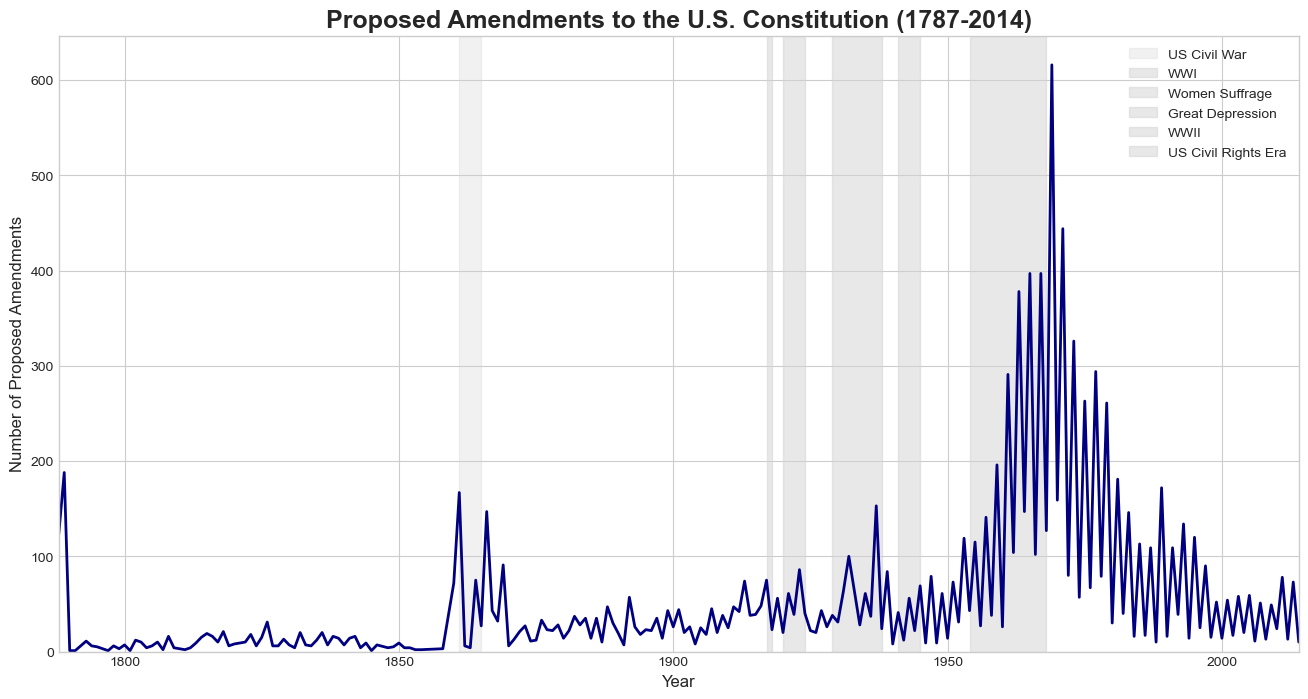
**Results (Gianna):** In the last method mentioned above, where we try to predict the number of amendment proposals with factors like political party in power and the country’s war status, we removed the option of both democratic and republican party in power to avoid multicollinearity in our models. After running our linear regression, we achieved a low R-squared value of 0.098, meaning our linear regression model only explains about 9.8% of the variance in our number of proposals. This suggests weak explanatory power in the variables we used, political power and war status. To confirm the linear regression model isn’t suitable for our prediction, we also looked into the RMSE, and found it to be 131.42, compared to our average proposal, 197.08. This means that on average, the model’s predictions are off by about 131 proposals per decade. This relative error of 0.667 suggests that our model’s average error is about two-thirds the size of the actual average value, which is pretty high. This confirms the unreliability of our linear regression approach. In our second attempt, we ran a Poisson Regression and found much better results. For all coefficients, we obtained very low p-values, meaning they’re statistically significant at any standard threshold level (0.05, 0.01, etc). This means that our interpretations are reliable when making predictions on amendment proposal counts based on the political party in power and war status of the country at the time. Given the Poisson Regression summary, we found that proposal count is 36% higher in democratic eras compared to when both parties are in power, 33% lower when neither parties are in power, 43% higher in Republican eras compared to when both parties in power, and that during war periods, proposal counts are 21% lower on average. From that, we concluded two trends: When a single party dominates the presidency, amendment proposal activity increases substantially; During war periods, legislative focus shifts to national defense, stability, and urgent issues, and not so much long-term structural change. To show these trends, a coefficient plot was made:



Top 8 Most Frequent Words Over Time Visualization:



# of Proposed Amendments Per Year Visualization W/ Key Events Highlighted:



Histogram of the Frequency of Proposed Amendments:

