

Wilford Woodruff Papers

Nicole Alfaro, Stefanie Alfaro, Erik Sanders, and Aubrey
Shuga

Spell Checker Algorithm

Results:

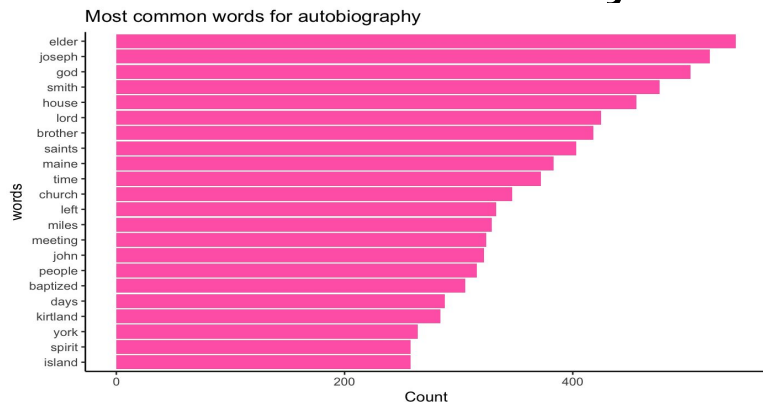
The spell checking algorithm corrects approximately 45% of misspelled words.

Things to Consider:

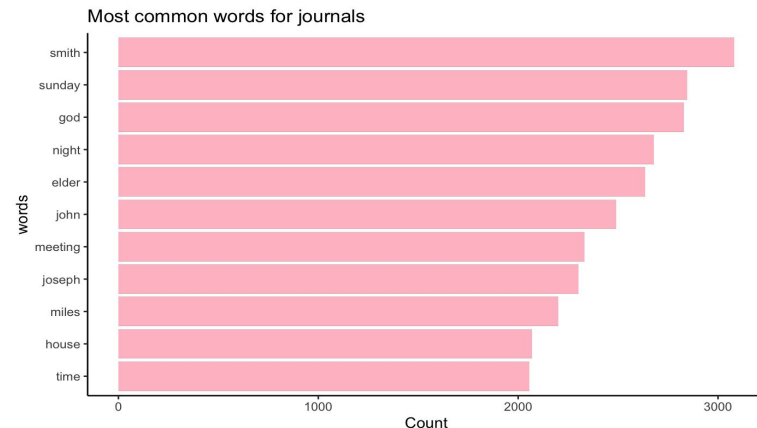
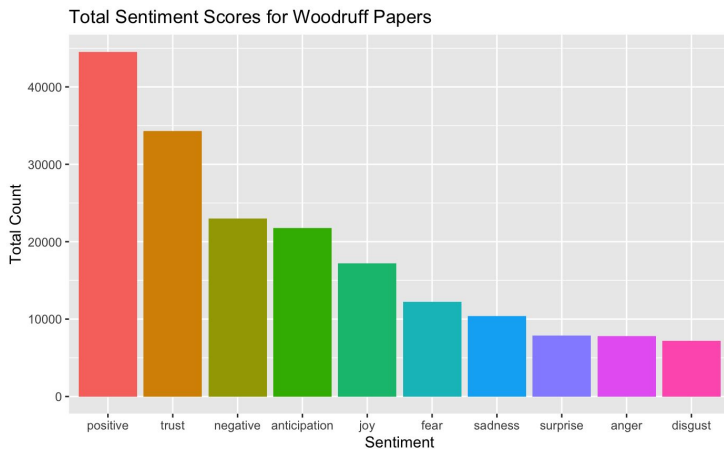
Some words do not need to be corrected (names, places, etc)

It doesn't account for conjoined words or split up words.

Initial Sentiment Analysis Exploration



We started our analysis by finding the most frequent words for some of the document types. We also plotted the top overall sentiments for the all the data



negative



negative



negative



We created word clouds for different document types since we thought the level of intimacy would differ. There is a difference in the smaller words, but bigger words like sick, death and faith are consistent throughout document type.

Log Likelihood for Collocation Analysis

We decided to do collocation analysis for some of the top words in our word cloud and words that were important to the church during that time.

Dunning's Log Likelihood compares two hypotheses about “word one” and “word two.”

Hypothesis 1 : $P(w_2|w_1) = p = P(w_2| \text{not } w_1)$

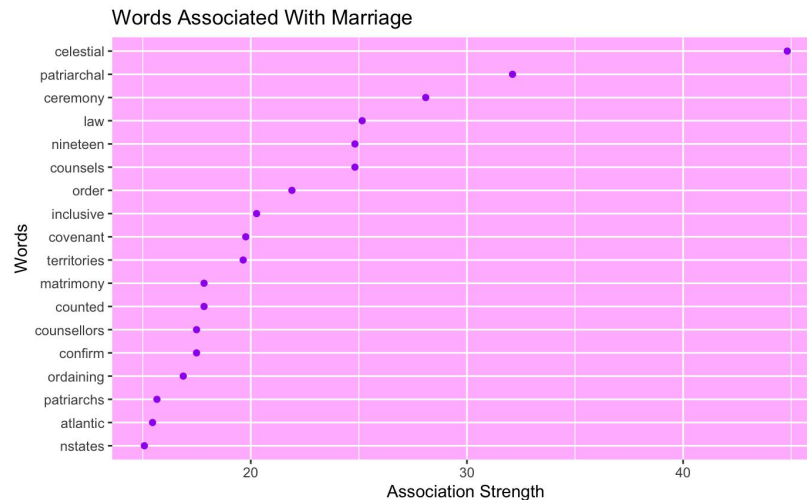
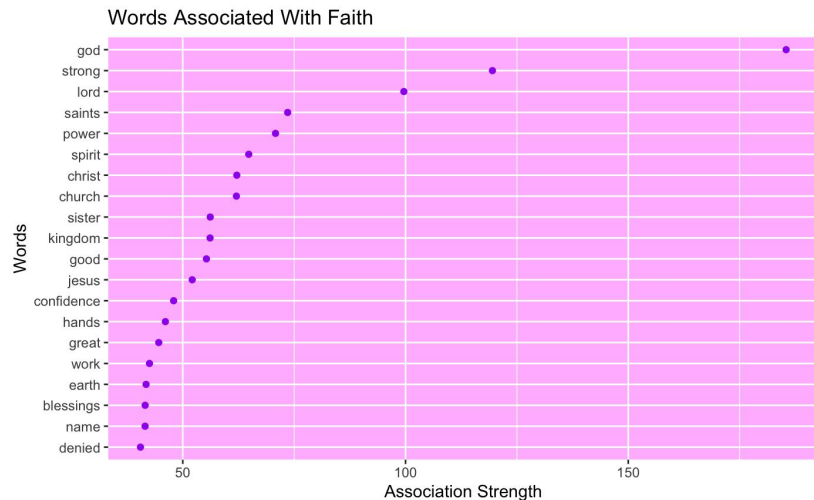
Hypothesis 2 : $P(w_2|w_1) = p_1 \neq P(w_2|\text{not } w_1)$

Formula: $\log\lambda = L(H_1) / L(H_2)$

The higher the value, the higher the likelihood that word one and word two are collocations.

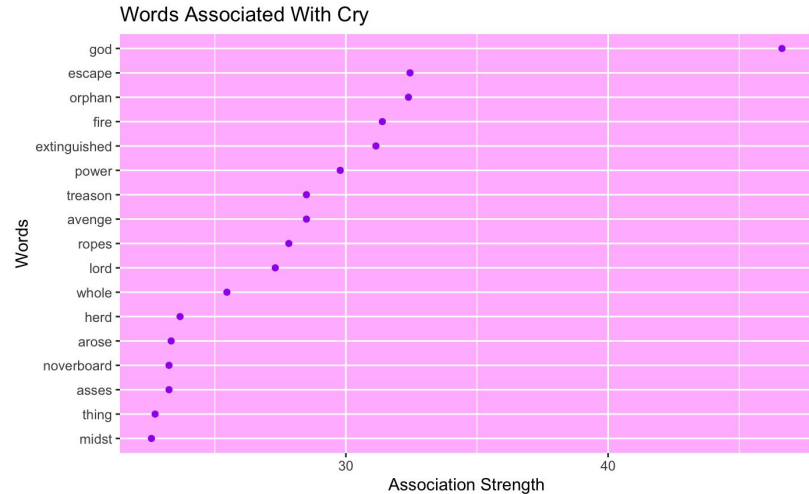
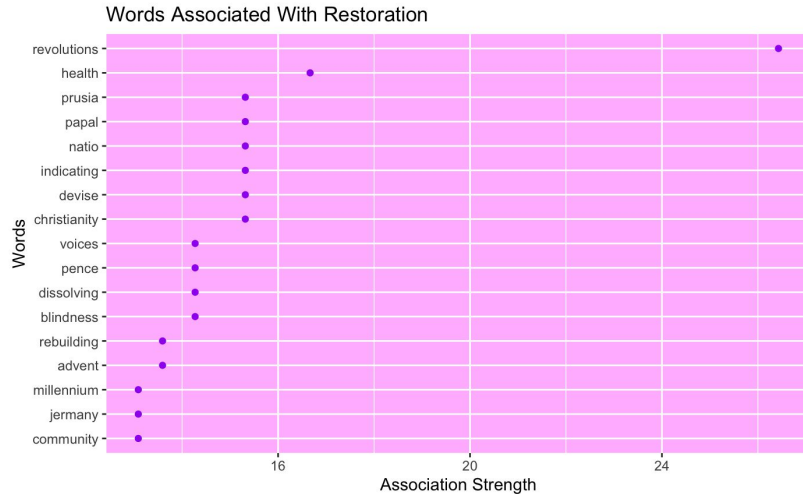
Due to time restrictions, we were unable to look at statistical significance using Log Likelihood. We would have taken the Log Likelihood value and used the chi squared table to find significance.

Collocation Analysis



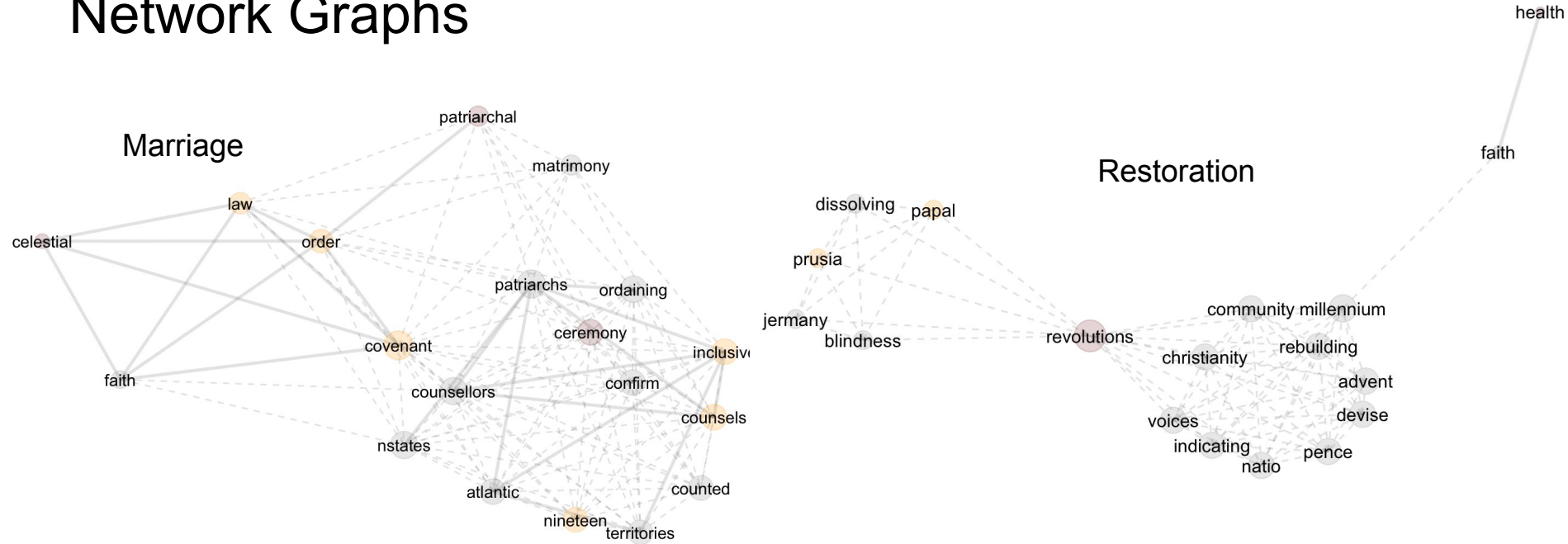
For words associated with “faith,” “god” and “strong” had the strongest association. “Celestial” and “patriarchal” had the highest association strength to “marriage.”

Collocation Analysis



The words with the strongest association to “Restoration” were “revolutions,” “health,” and “purisa”(historical region of Germany). As for “cry,” it was interesting to see that the words with the strongest association were “God,” “escape,” and “orphan.”

Network Graphs



These network graphs shows how related words are or how unrelated they are. The light red nodes have the words that are highly associated with the term while the yellow nodes have the words that are mildly associated with the terms. Finally, the grey nodes had the least association strength. Here, you can see how they cluster into groups. For example, celestial is related to “law” and “faith” but not directly related to “territories.”