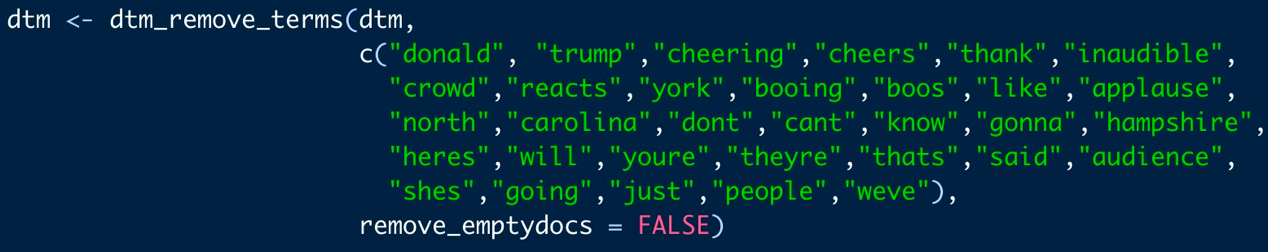
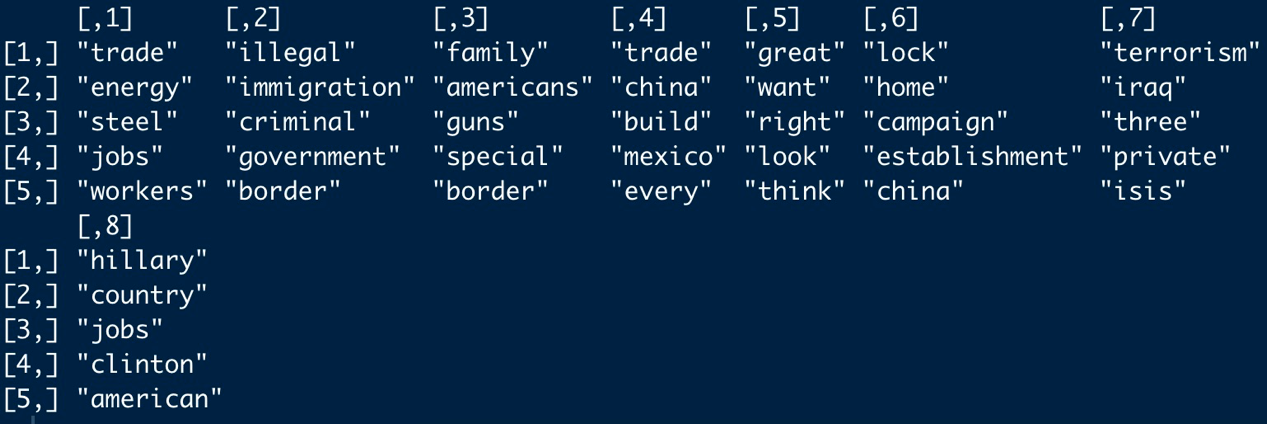
**Group assignment 2 – group 8**

**Cleaning and word limits**

For cleaning, we removed words such as state names and: These words repeat quite a lot but do not have actual meanings. We manually removed the words york, carolina, north and hampshire because states names that were bigrams were removed.

For word limits, we removed the words that were longer than 20 characters and shorter than 4 characters. If we had kept the shorter words in the analysis, then there would have been more “fill” words.

**Identify a topic that should relate to speech location (e.g. “wall” should relate to southern boarder), What do you expect?**



We narrowed it down to 8 topics because when we had more, a few topics were quite similar. The topic we select is topic 8, and the most frequent words are Hillary, country, jobs, Clinton, and America. We believe this topic is related to trump speaking badly about Hillary and how he can do better than her and we expect him to mention her more in the states that are more likely to vote for her.

**If Trump adjusts his speeches to the location, where does he talk more/less about the topic you identify**

He talks less about Hillary in the middle part of America. This might be because those states are more conservative, and he does not need to talk badly about Hillary to get more votes for himself. We also believe that the states that voted for him (Figure 2), but where he talked more about Hillary (Figure 1) may be states where he was not sure whether he would win the election there or not. These states are Mississippi, North Carolina, Florida, and Ohio. Trump adjusts his speeches according to people’s preference.

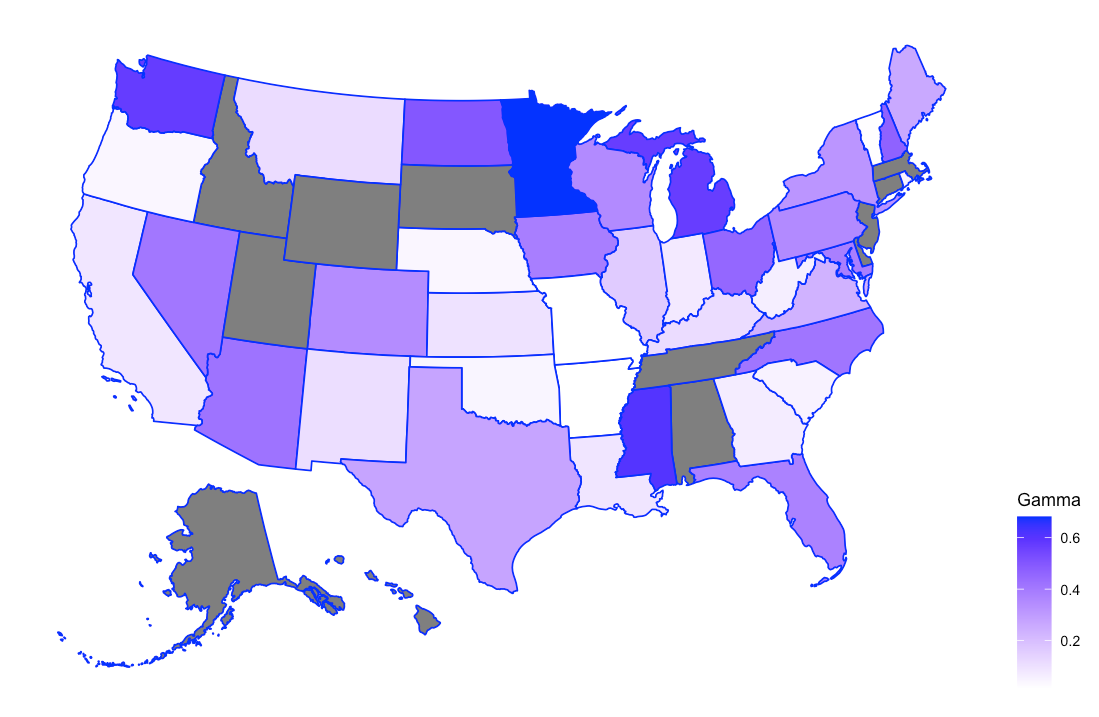


Figure : How much Trump talked about Hillary Clinton per state

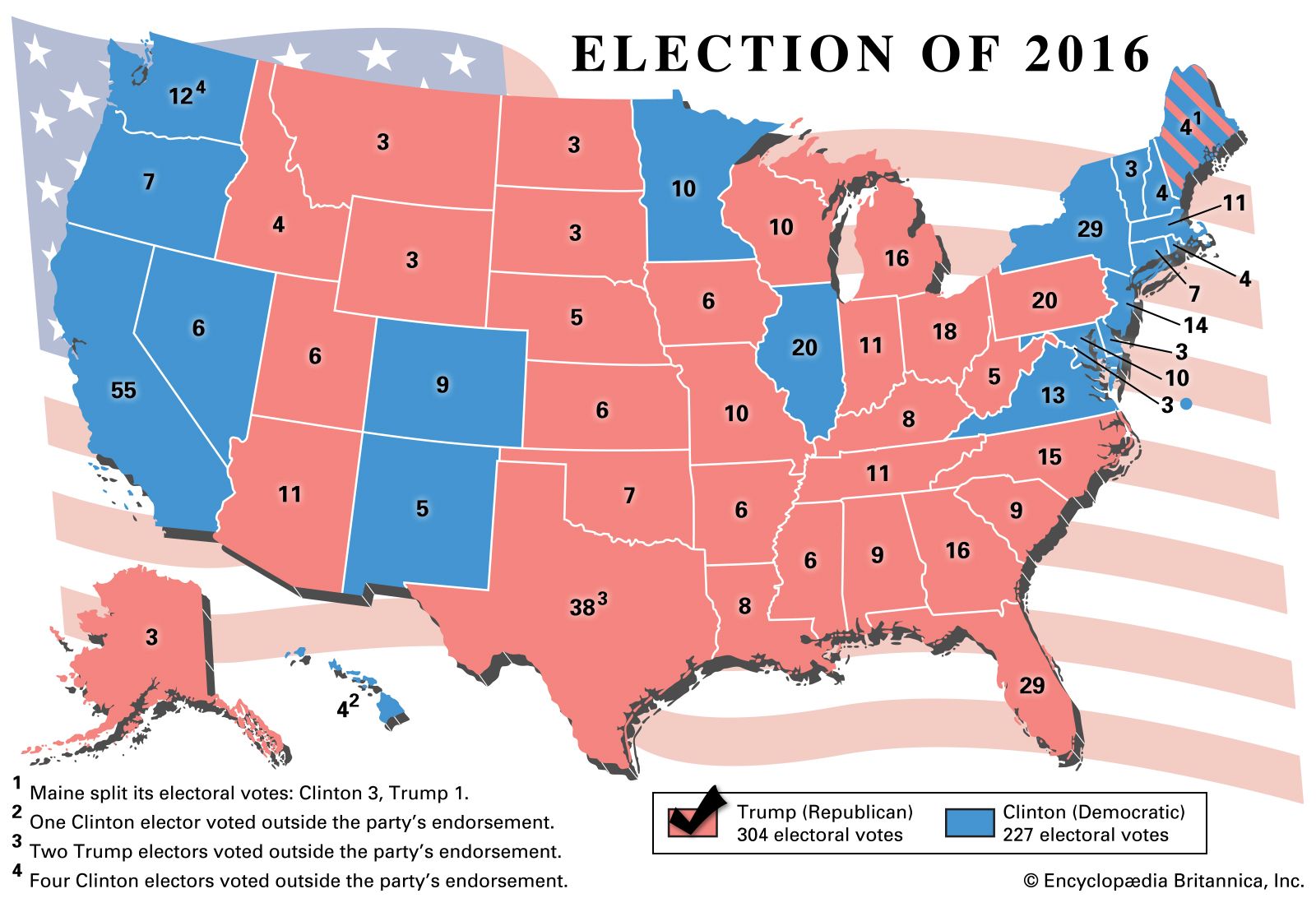
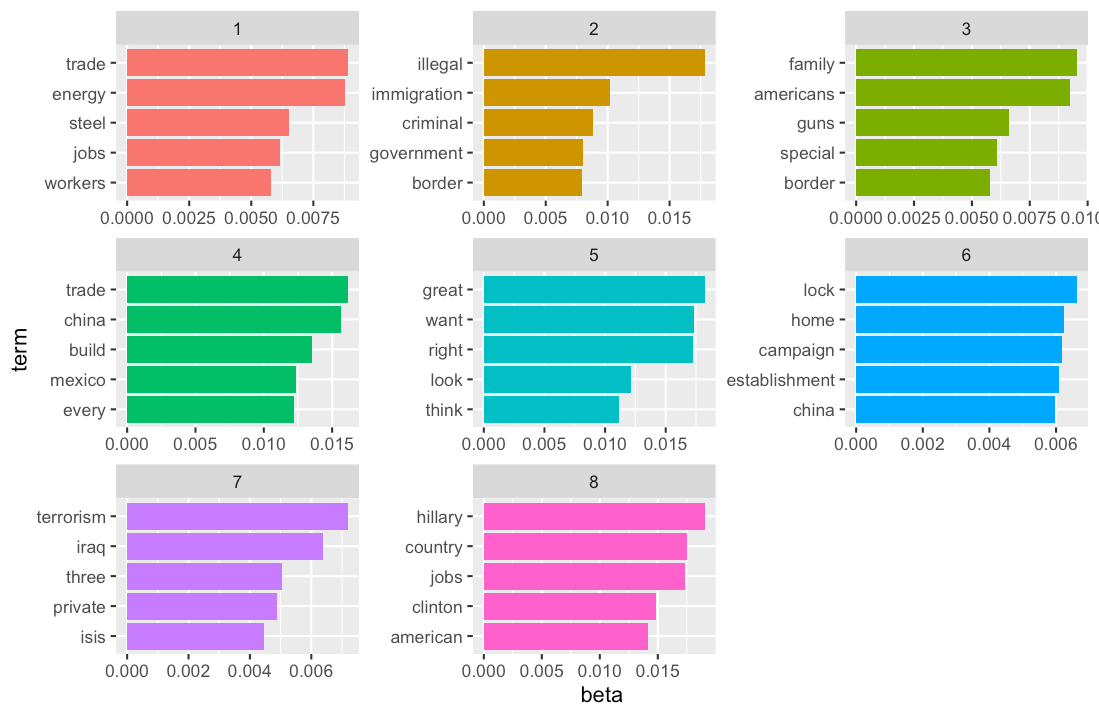


Figure : Source: Beckwith (2021).

**What do you think is the right number of topics?**

We used 8 topics. We think the right number of topics should match the actual number of topics that Trump could mention in his speeches for his presidential campaign. If too many topics are selected, then the topics will be very sparse and not clear. However, if too few topics are chosen, then some essential topics are not included.



**Does the result look better if you estimate the model on speeches or if you split each individual speech in paragraphs/parts?**

If we split into paragraphs, then each paragraph will be very clear and related to one topic, as opposed to a speech which can include multiple topics. Additionally, we would not need to remove as many words, because we could simply remove the introduction and conclusion paragraphs. We believe that the results may look better because of this.

**References**

Beckwith, D. C. (2021, October 28). United States Presidential Election of 2016. Encyclopedia Britannica. https://www.britannica.com/topic/United-States-presidential-election-of-2016