Final Project

Joseph Lobowicz

So What?

Political news outlets don't just report facts, they frame them. This project explores how word choice, sentiment, and topic emphasis vary across left-leaning, neutral, and right-leaning news outlets. We ask:

What linguistic and thematic differences emerge among outlets with different political alignments?

Using a 30% sample from a 17,362-article corpus labeled left, neutral, or right, we apply a sequence statistical analyses to reveal how each camp talks about politics and, more importantly, what that tells us about media framing.

Exploratory Data Analysis

From our sampled proportion of 5206 articles, 45% are neutral, 23% are left biased, and 32% are right biased.

```
# A tibble: 3 x 2
  label    count
  <fct>    <int>
1 neutral  2340
2 left     1198
3 right    1668
```

We can generate word clouds for the left, right, and neutral, respectively, to reveal the dominant terms.

After looking at the word clouds for each label, it is clear they have several common words. Let's compare the relative frequencies of these common words that all stances share, and compute their relative frequencies on the plot of Figure 4. Those with a significant differences are labeled.
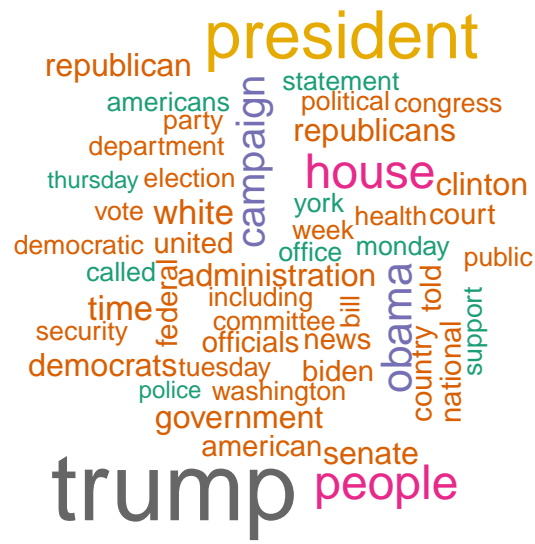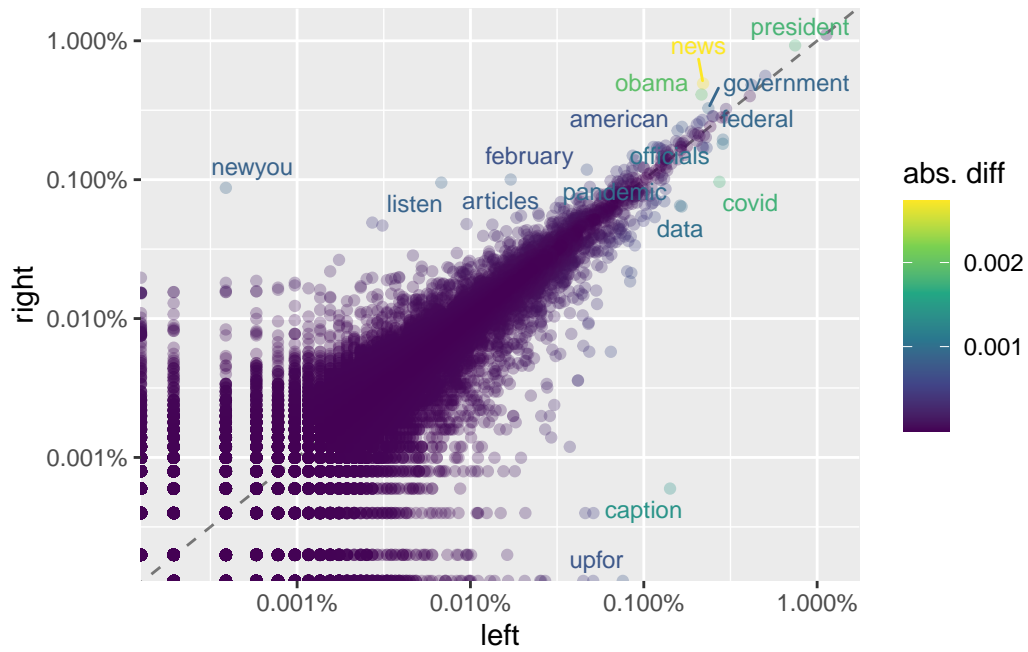
Figure 1

Figure 2

Figure 3

Figure 4

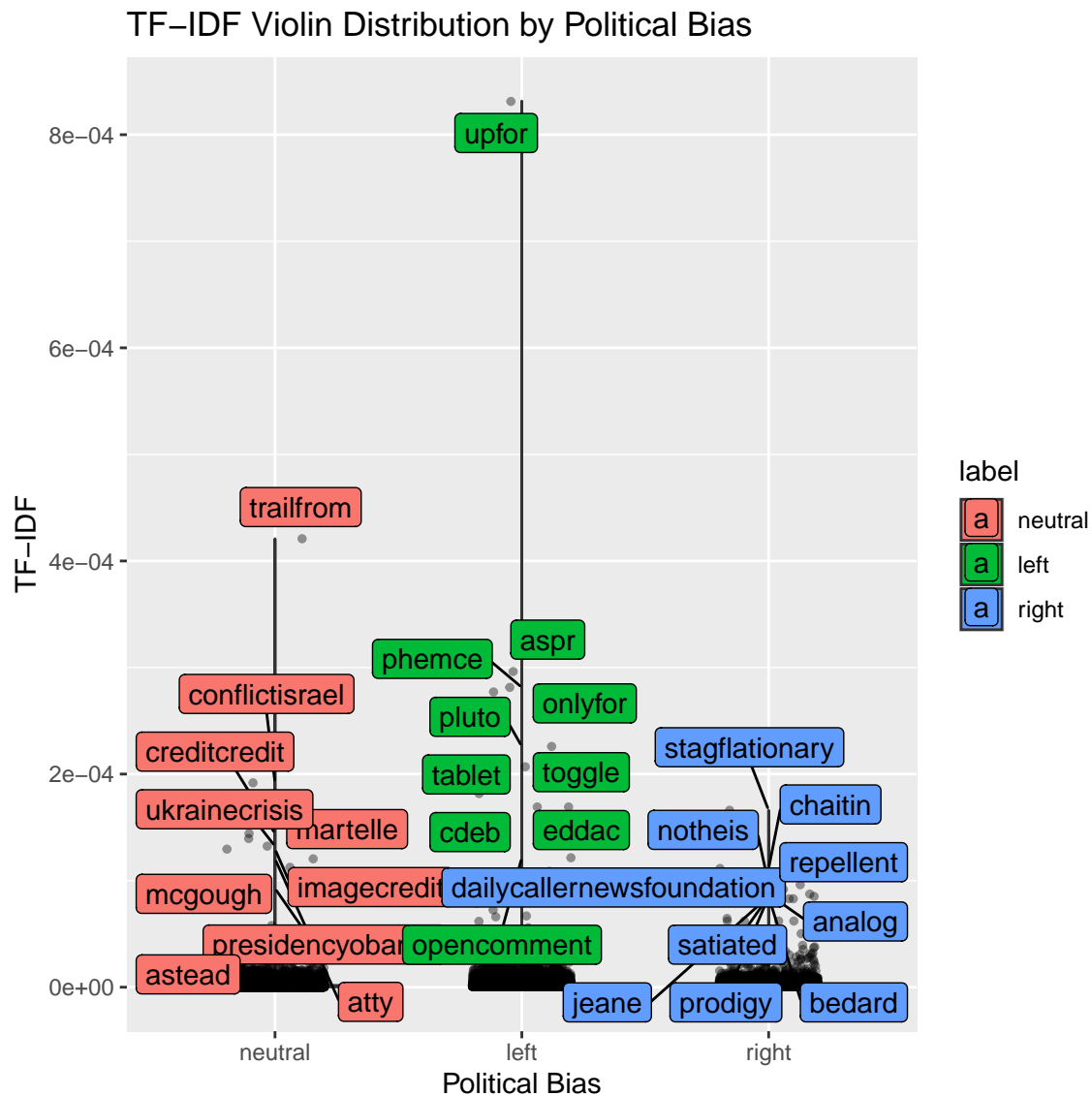Using tf-idf, we can identify the top ten words uniquely important to each label, and show the distribution below.

Figure 5

Further, we plot the top 15% of words with the greatest absolute difference.

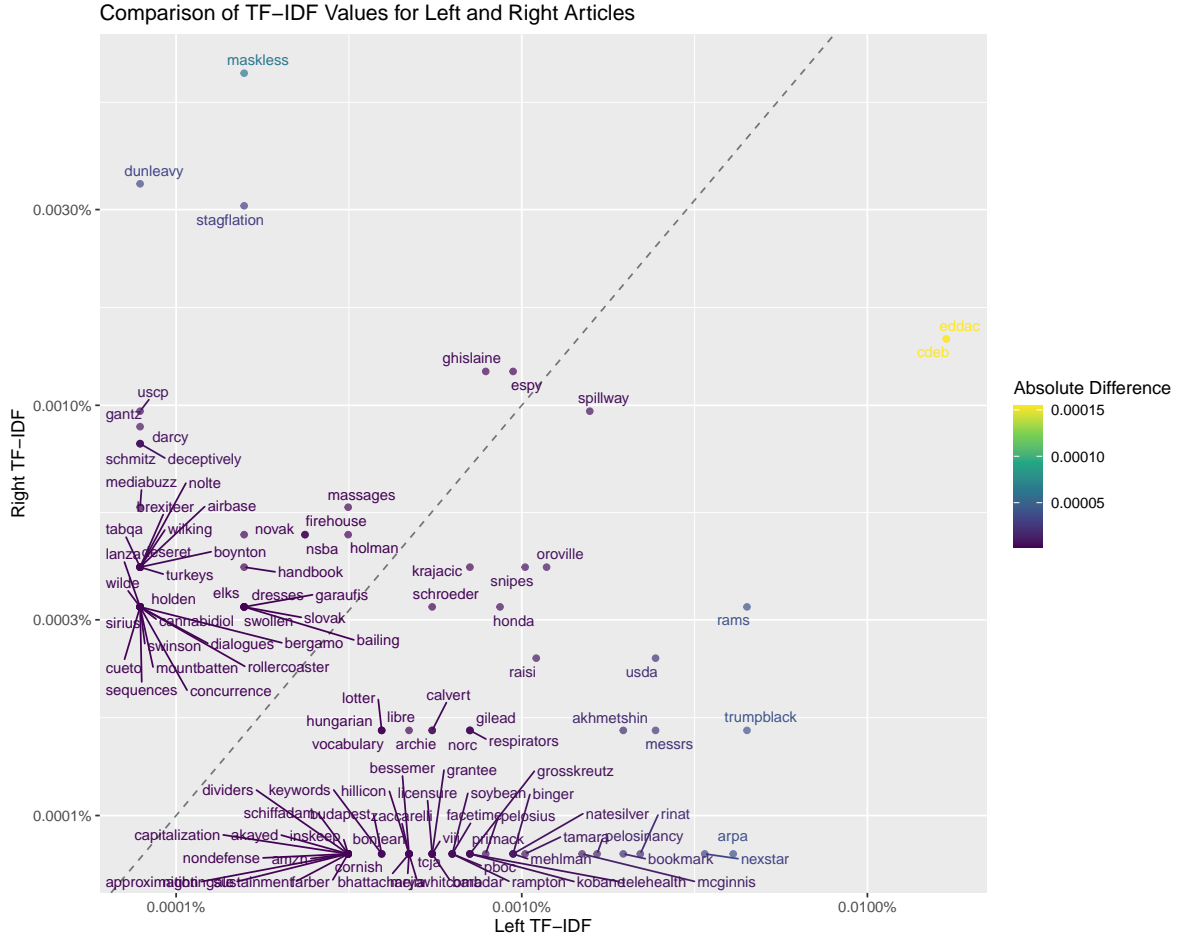Comparison of TF–IDF Values for Left and Right Articles

Figure 6

Sentiment Around Key Figures

News outlets choose which voices to elevate and how to color them. How does the tone around key political figures differ by media bias? The two main figures for each political party from my time has arguably been Trump and Obama.

For Figure 7, we isolated every sentence across our sampled corpus that mentions the entity "obama," then tokenized those sentences and joined them to the Bing sentiment lexicon. Then, by converting "positive"/"negative" labels into +1/–1 scores, we computed the average sentiment of all "obama" contexts within left-, neutral-, and right-leaning articles.

We mirrored our *Obama* workflow for the entity *Trump* and computed the average sentiment score and count of mentions for each political bias category.
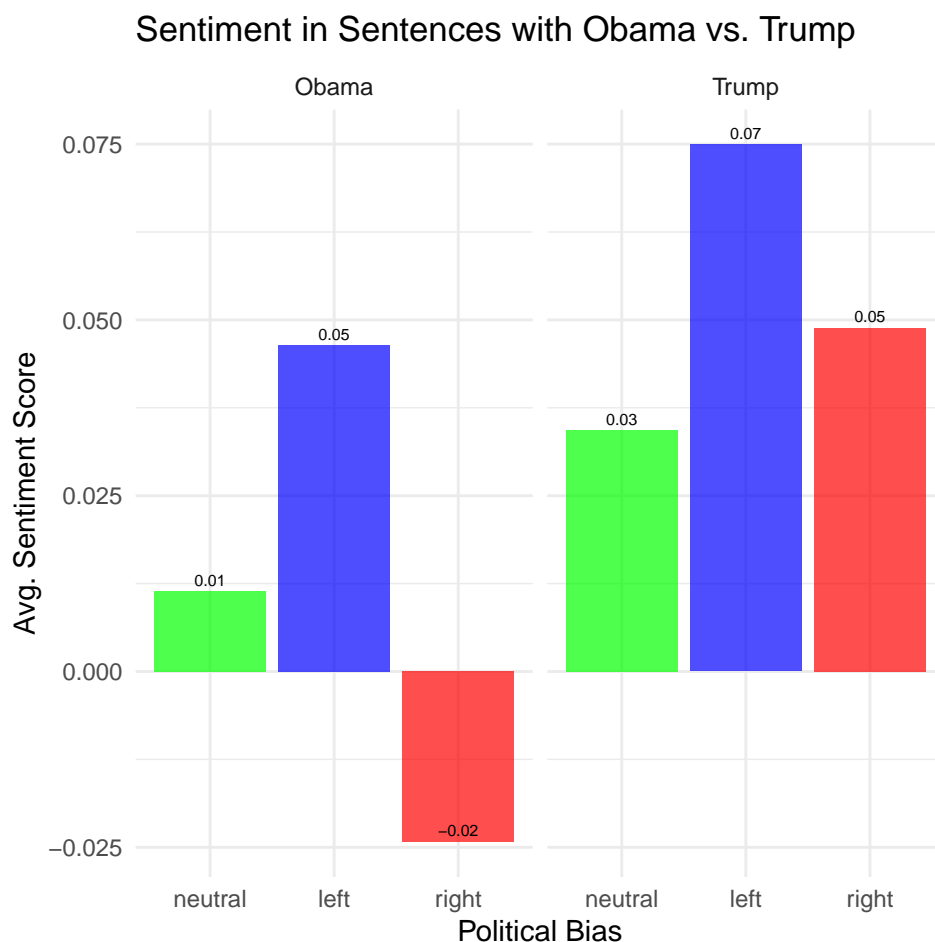
Figure 7

- Neutral articles mention "Obama" 104,330 times, with an average sentiment score of +0.0403.
- Left-leaning articles mention "Obama" 18,794 times, with an average sentiment score of +0.0299.
- Right-leaning articles mention "Obama" 22,532 times, with an average sentiment score of +0.0094.

In the case of *Obama*, neutral-labeled articles mention him with a modestly positive average sentiment score. Left-leaning outlets reference him far less frequently (18,794 mentions) and at a slightly lower positive sentiment (0.0299), while right-leaning outlets mention him 22,532 times and with a very small positive mean sentiment (0.0094), essentially a neutral tone. Thus, although all three camps speak of Obama in broadly positive terms, neutral outlets give him the greatest prominence and a tone that is slightly more favorable than partisan sources. The left's smaller volume and moderate positivity may reflect a focus on specific policies or

achievements, whereas the right's relatively lower sentiment may suggest a more restrained framing.

- Neutral articles mention "Trump" 123,710 times, with an average sentiment score of +0.0612.
- Left-leaning articles mention "Trump" 45,566 times, with an average sentiment score of +0.0849.
- Right-leaning articles mention "Trump" 40,668 times, with an average sentiment score of +0.0291.

By contrast, the overall conversation around "Trump" displays both higher mention counts and more pronounced tonal differences. Neutral outlets again lead in sheer volume, mentioning Trump with an average sentiment of 0.0612, higher than their sentiment toward Obama, but overall pose both entities with positive references. Left-leaning outlets mention Trump substantially more than right-leaning ones, and one might expect left-leaning outlets to cast Trump in a negative light, yet they do so with a highly positive tone. They outlets mention him with the strongest positive average sentiment (0.0849), suggesting that when they do cover Trump, they often discuss events that register positively under a simple lexicon count. Right-leaning outlets reference Trump with a mean sentiment that is more muted than neutral coverage, but overall still positive.

Framing Through Co-occurrence: Which Words Surround Our Key Figures?

We have shown how outlets talk about "Obama" and "Trump"; Figure 8 address the natural next question, "With which words do they talk about them?"

To see not just how outlets talk about "Obama" and "Trump," but what language they use around those names, we extracted 10-word context windows for each entity, removed stop words, and tallied the most frequent co-occurring terms separately for left, neutral, and right outlets. This analysis extends our exploration of framing by revealing the narrative vocabulary each bias camp deploys whenever these figures appear.
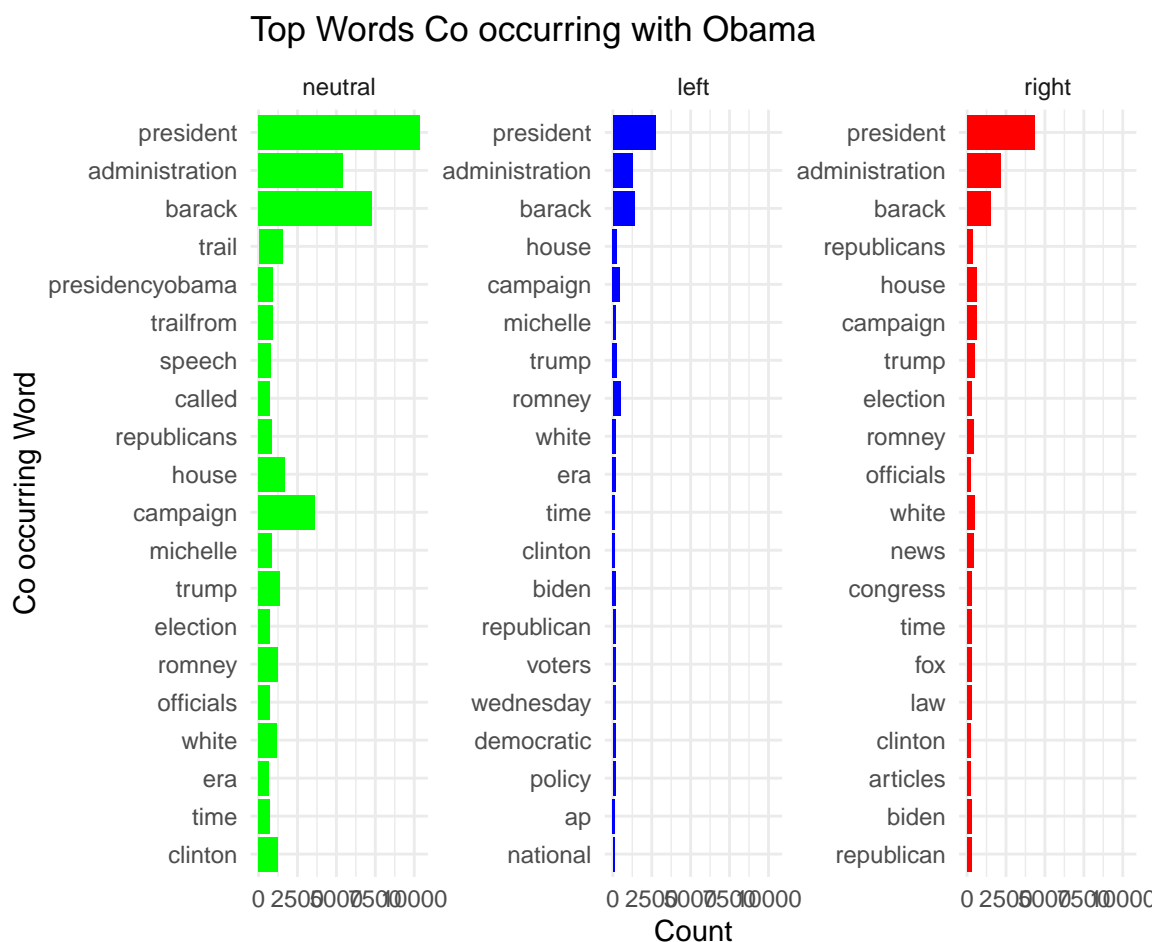
Figure 8

For Obama, left-leaning sources likewise emphasize his official role ("president", "administration,"), but then shift more quickly to policy and values terms such as "policy" and "health". Right-leaning coverage, while also foregrounding title and biography ("president," "barack,"), brings in outlet-specific and partisan context labels like "michelle".
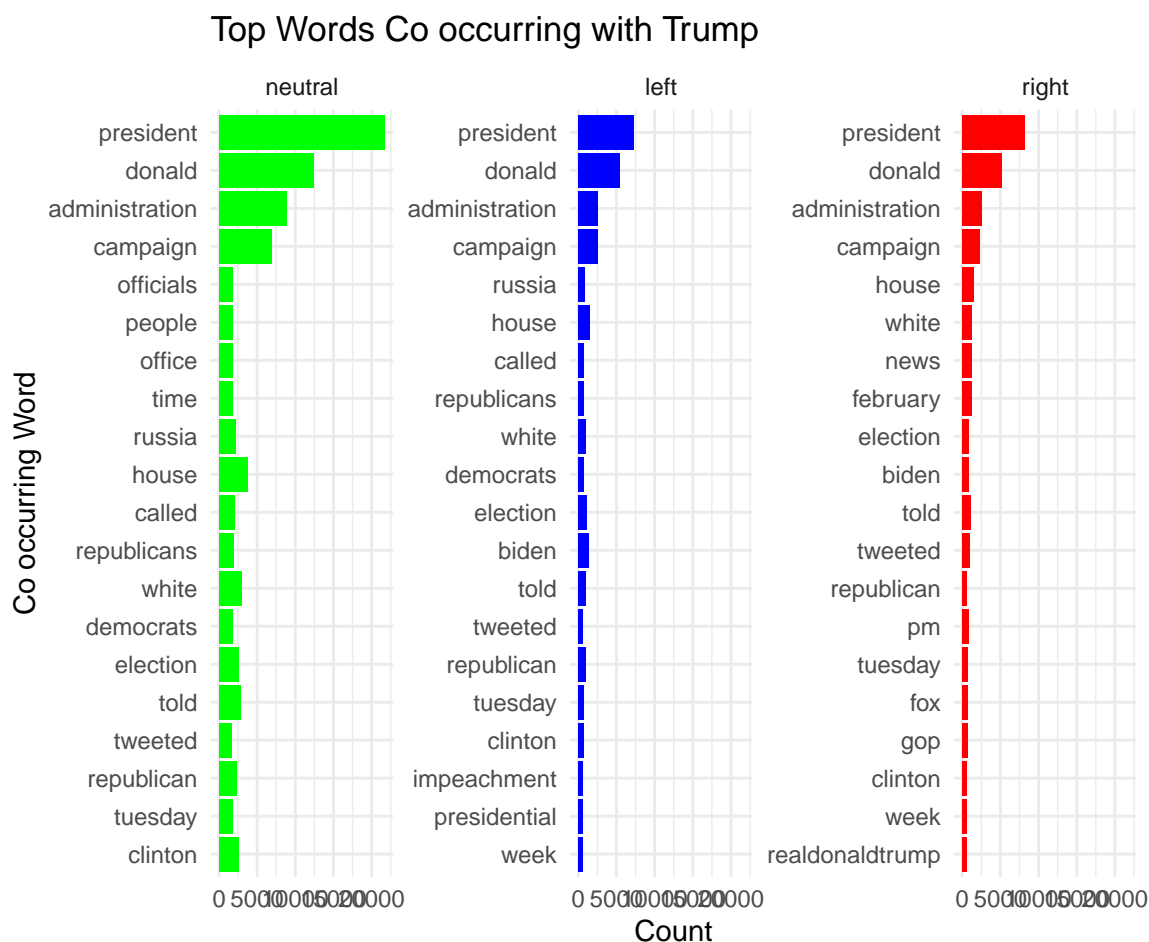
Figure 9

From Figure 9, in the case of "Trump," left-leaning sources also highlight his title ("president," "administration,") but quickly pivot to legal and political scrutiny with words like "impeachment", "cohen" (who was Trump's lawyer), and "russia", reflecting coverage of investigatory narratives. Right-leaning outlets, while overwhelmingly focusing on the presidency itself, favor press-release terms like "news", "tweeted" and network identifiers such as "fox", underscoring a framing rooted in media statements and official communications.

Each outlet selects surrounding words that reinforce particular storylines: left outlets foreground policy and legal scrutiny; and right outlets emphasize media platforms and direct communications. In this way, co-occurrence analysis demonstrates that word choice around key figures could itself be a form of editorial framing.

Topic Modeling by Bias

To move from individual words and sentiments to the broad narrative frames that guide political coverage, we fit a single LDA model (k = 10) on our entire sampled corpus. By examining the top ten terms for each topic, we see the themes that underlie every article, regardless of political alignment.
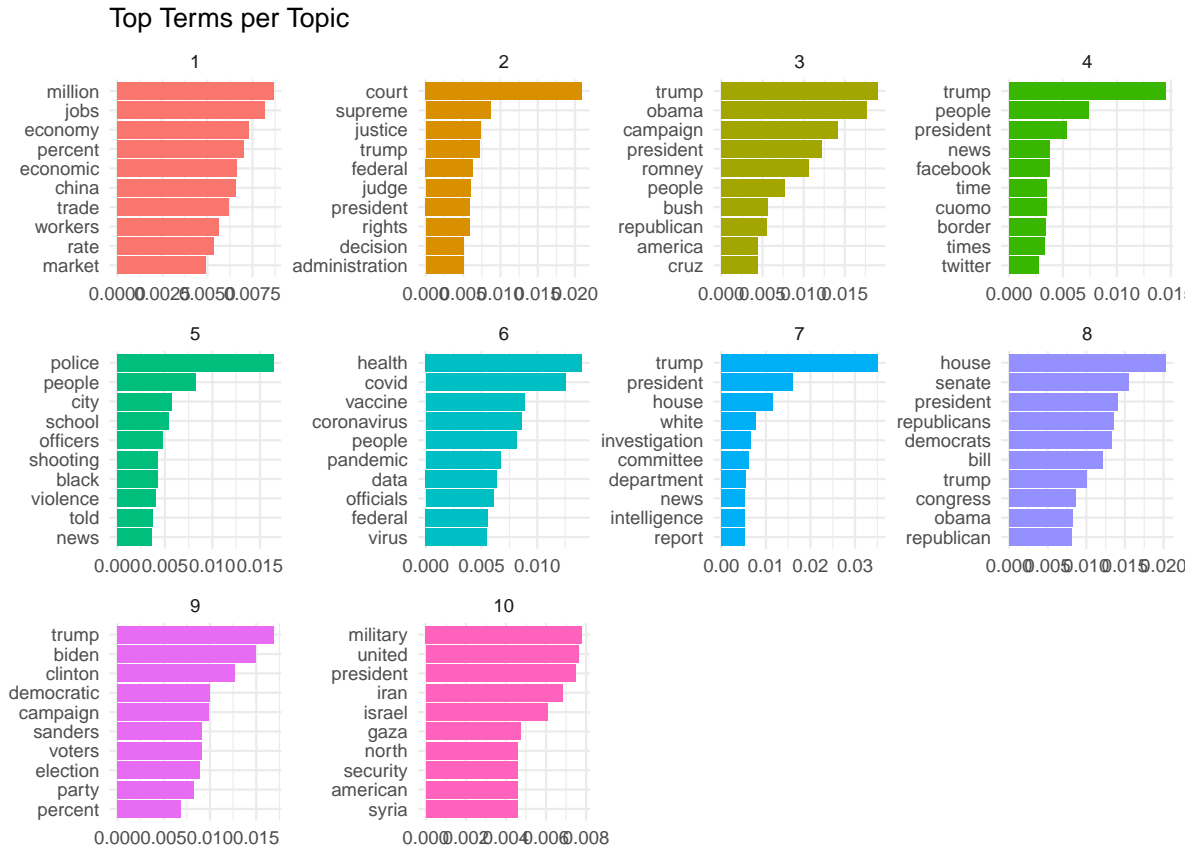

Figure 10

Based on the highest-probability words from Figure 10, it would not be far-fetched to label the ten topics as follows:

- Topic 1: Republican Border News
- Topic 2: War & Goepolitical Conflicts
- Topic 3: Security & Military Operations
- Topic 4: Legal & Investigations
- Topic 5: Policing & Civil Unrest

- Topic 6: Elections & Party Politics
- Topic 7: Health & Pandemic
- Topic 8: Global Trade
- Topic 9: Covid Economy
- Topic 10: Congress & Legislation

These themes show the frames through which political outlets present news: from foreign affairs (Topics 2 & 8) and pandemic coverage (Topic 7), to domestic governance (Topic 10) and culture-war flashpoints (Topic 5). By defining a shared topic space, we can then compare how much each outlet emphasizes each narrative. This could help address our core question of how theme varies across political alignments.

When we compare the average topic proportions by label, distinct patterns emerge because each side selects different narrative frames to emphasize.
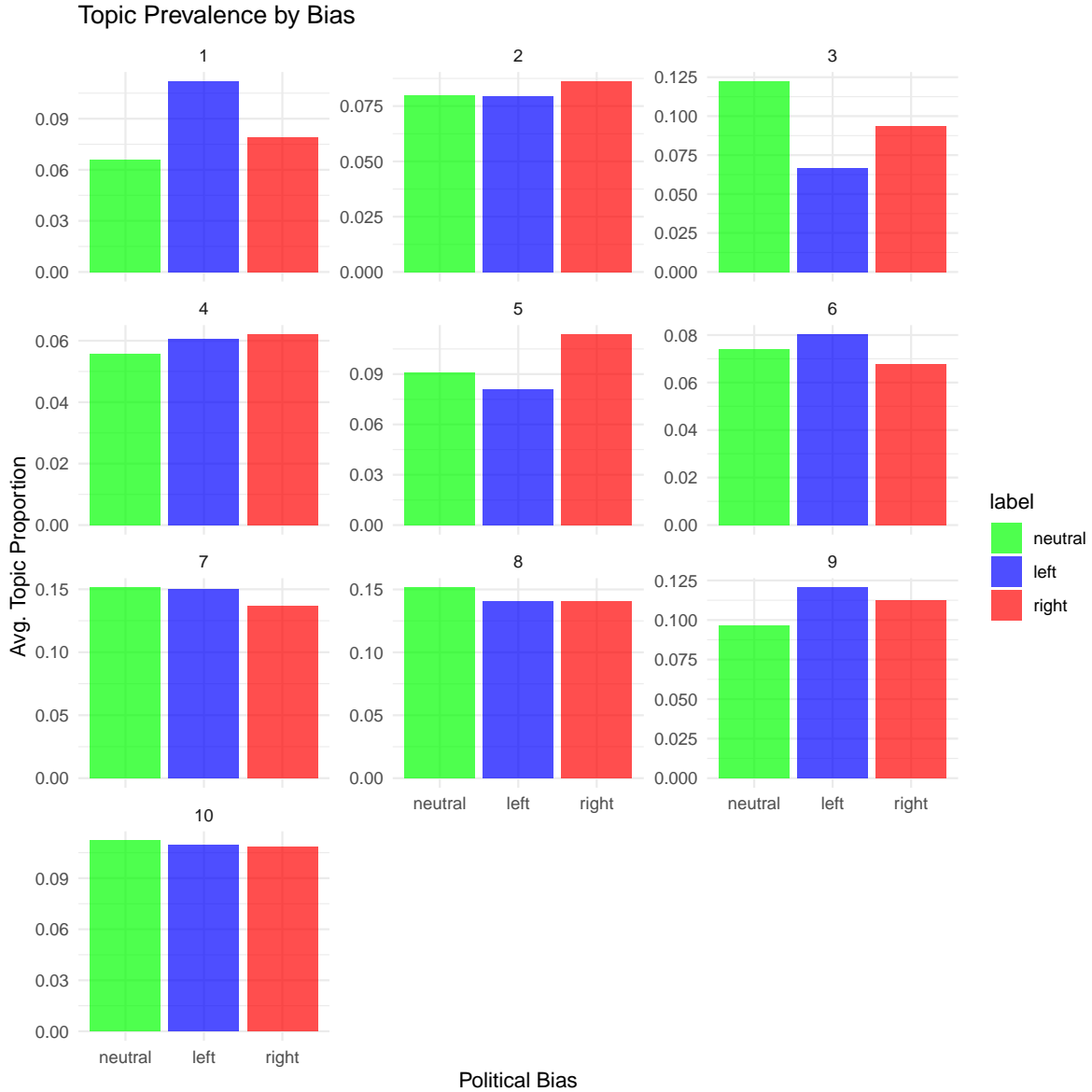
Figure 11

From Figure 11, we can draw out some interesting relations:

- Left-leaning outlets had the greatest proportion of articles on *Republican Border News* (Topic 1) and *Health & Pandemic* (Topic 7).
- Right-leaning outlets had the greatest proprtion of articles on *Legal & Investigations* (Topic 4) and *Covid Economy* (Topic 9). They also wrote more on *Security & Military Operations* than left outlets.

- Neutral outlets led in the other topic categories.
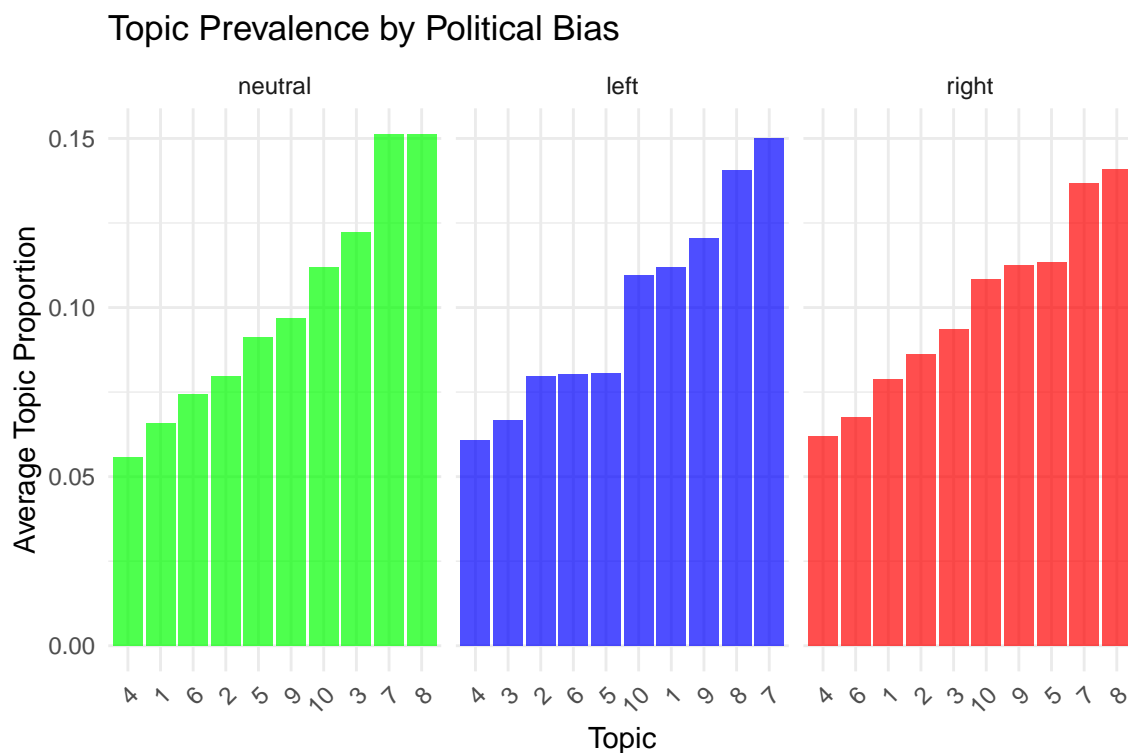
Topic Prevalence by Political Bias

Figure 12

From Figure 12, left outlets focus more on highlighting *Republican Border News* (Topic 1), right outlets on *Congress & Legislation* (Topic 10), and neutrals on *Elections & Party Politics* (Topic 6). Taken together, these topic differences underscore our core question: political news outlets don't just report facts; they frame them by choosing which narratives to foreground.