RISE RESEARCH INNOVATION SCHOLARSHIP ENTREPRENEURSHIP

Mentor Name: John Rachlin
Subject/Category: News Media
Class: DS3500
Group #: 53

Northeastern University

Same Story, Different Angles: Investigating News Media Bias

Catrina Herman, Mika McLean, Caroline Monaco

Background, Motivation, and Goals

- News articles on the same topic can be presented very differently depending on the source.
- Our goal was to explore and visualize how **sentiment**, **language**, and **tone** vary across major U.S. media outlets reporting on the same event.
- We aim to highlight media bias and framing by comparing how often certain words appear and how sentiment shifts sentence-by-sentence in each article.

Process and Methods

- Gathered 11 news articles from outlets like NPR, CNN, FOX, BBC, USA Today, etc., all reporting on **Cory Booker's 2025 Filibuster**
- Text files were processed using a custom Python class (TextAnalysis) that cleaned text, removed stopwords, and calculated metrics like: **Sentence- and word-level sentiment**, **word frequency** & **sentence structure stats**

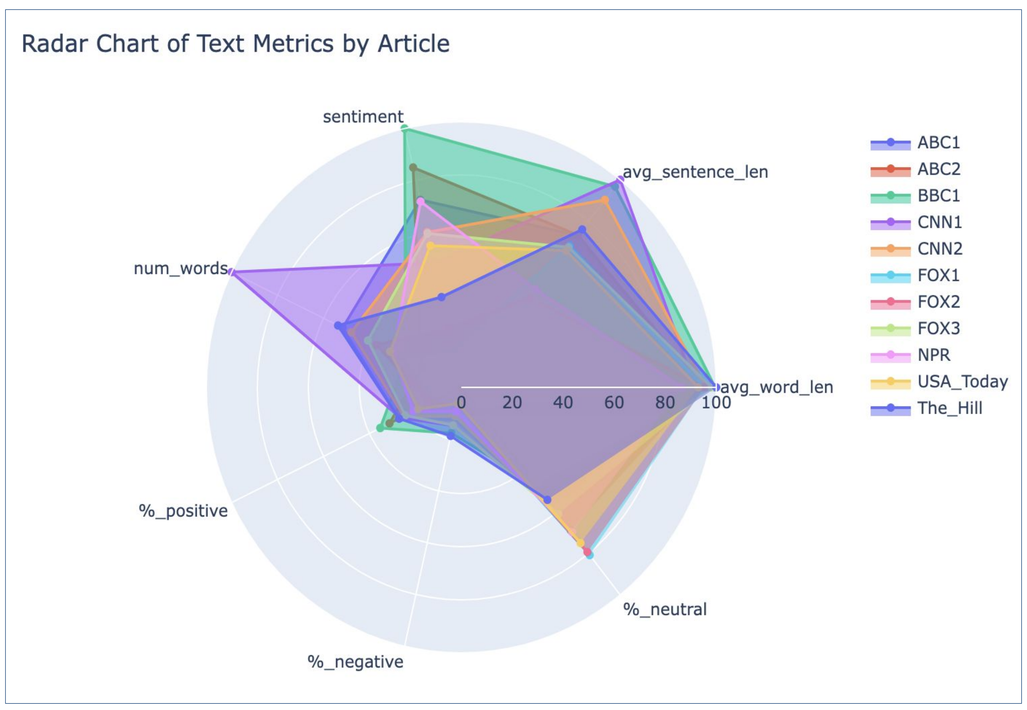

Figure 1: Text Metrics by Article
Each news article is represented by a different color on the radar chart. The metrics are normalized among all the articles, indicating that points along the border are the "highest" observed media sources for that metrix.

Author Contributions
Catrina - TextAnalysis class and generic parser for text pre-processing, radar chart, constructing poster
Caroline - Created bar charts comparing sentiment percentages and a scatter plot of positive vs. negative percentages, constructed poster
Mika - Built sankey diagram, created sentence sentiment subplots, constructed poster

Visualizations and Findings

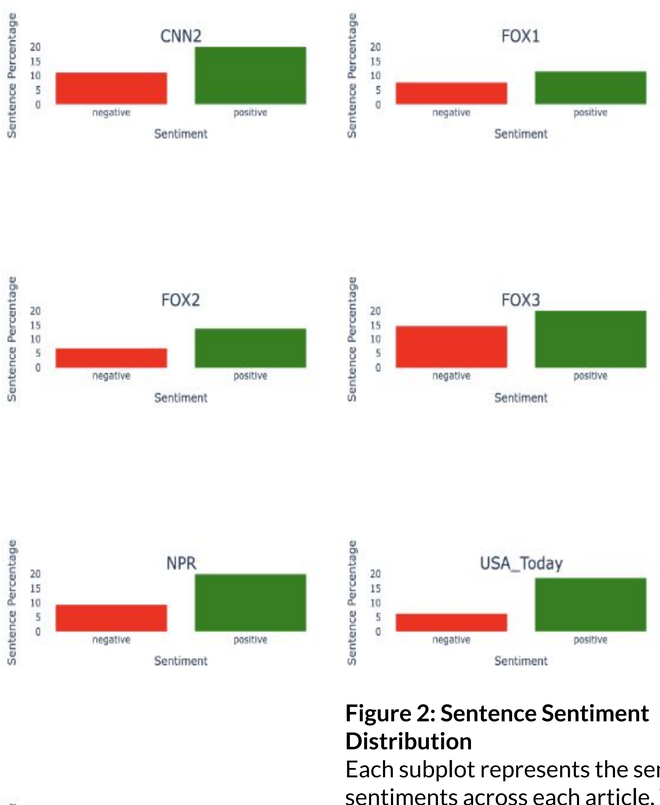

Figure 2: Sentence Sentiment Distribution
Each subplot represents the sentence sentiments across each article. The "red" bar indicates the % of sentences with negative polarity while the "green" bar represents the % of sentences with positive polarity.

Analysis

- **CNN1** stands out with a very high num_words count, suggesting a much longer article that may include more elaboration than others
- On average, **ABC** had the most positive broadcasting on this event
- Conversely, **Fox** & **The Hill** had the most negative broadcasting on average
- The key words used, as reflected in the Sankey, were ***Booker***, ***Senate***, and ***Speech***, which shows that the news media largely stayed on topic, focusing on the factual and procedural aspects of the event rather than using dramatic language.
- Overall, the percentage of positive sentences in each article exceeded the negatives, indicating that none of the articles conveyed had an overall "negative" sentiment.

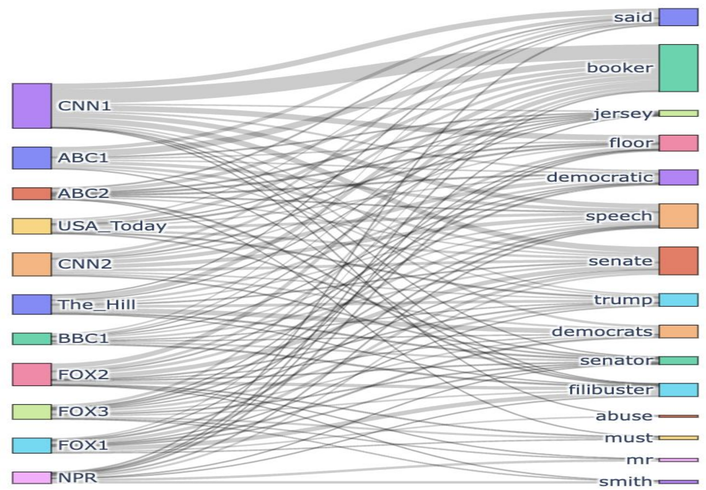

Figure 3: Sankey Text-to-Word Diagram
Each article maps to its 3 most common words with a "thickness" of each word count. The words on the right are the 15 most common words among the 11 articles, indicating lots of overlap.

Conclusion and Next Steps

- Overall, we identified many similarities in sentiment and text structure across articles from different sources.
- We initially presumed that we would see bias in different news media sources, but from our findings we can note that news media today tends to be more subjective than objective.
- In the future, it could be interesting to take out the "neutral" words for an event to further analyze the existence of opinions in the sources.