Executive Summary

Final Project
DS 710

Introduction

In the current political climate, there is an obvious divide between Liberals and Conservatives that can be seen in the comment section of many Twitter posts. With major news networks in the middle, I wanted to see if there's a difference in the proportion of positive comments a major Conservative news network, Fox News, receives compared to the proportion of positive comments a major Liberal news network, CNN, receives. I believe this analysis could be beneficial to the news networks themselves. If one is found as receiving less positive tweets than the other they can try and find out what may be causing the difference and find out ways they can gain more positive tweets.

My hypothesis is that there will be a difference in the proportion of positive tweets received from Fox News and positive tweets received from CNN.

Data Collection

The REST API method from the Tweepy python package was used to extract 10000 tweets that included mentions to @FoxNews and/or @CNN. I only wanted to view original tweets so retweets were filtered out. Also, I only wanted to look at tweets to the specified news organizations so I removed the tweets directly from their twitter accounts.

The information I found important to store was the tweet id string, to be used as an ID, and the text, the actual tweet comment. A data frame was then created to easily view, analyze, and add more necessary information. Using the text, each tweet was classified by the news organization it mentioned and that was added to a new column in the data frame. If a tweet included both news organizations or the tweet was cut off and the news organization it mentioned was unable to be viewed, it was marked as NaN.

Sentiment Analysis

To view which tweets were considered positive, a sentiment analysis was ran on each tweet. Before the sentiment analysis was ran, each tweet was cleaned, removing unnecessary information. After tweet cleaning, sentiment analysis was run to analyze the polarity of each tweet. The polarity of each tweet was added to a new column of the data frame.

Two-Sample Test for a Difference of Proportions / Conclusion

The information previously gathered was imported into R to be used in a Two-Sample Test for a Difference of Proportions. Shown in Figure 1, the proportion of Fox News tweets that were positive were 0.2352694 and the CNN tweets that were positive were 0.2844480. Using these proportions for the test and with a significance level of 0.05, at the p-value of 0.000005116, there is enough evidence to claim that there is a difference of proportions between Fox News tweets that are positive and CNN tweets that are positive.

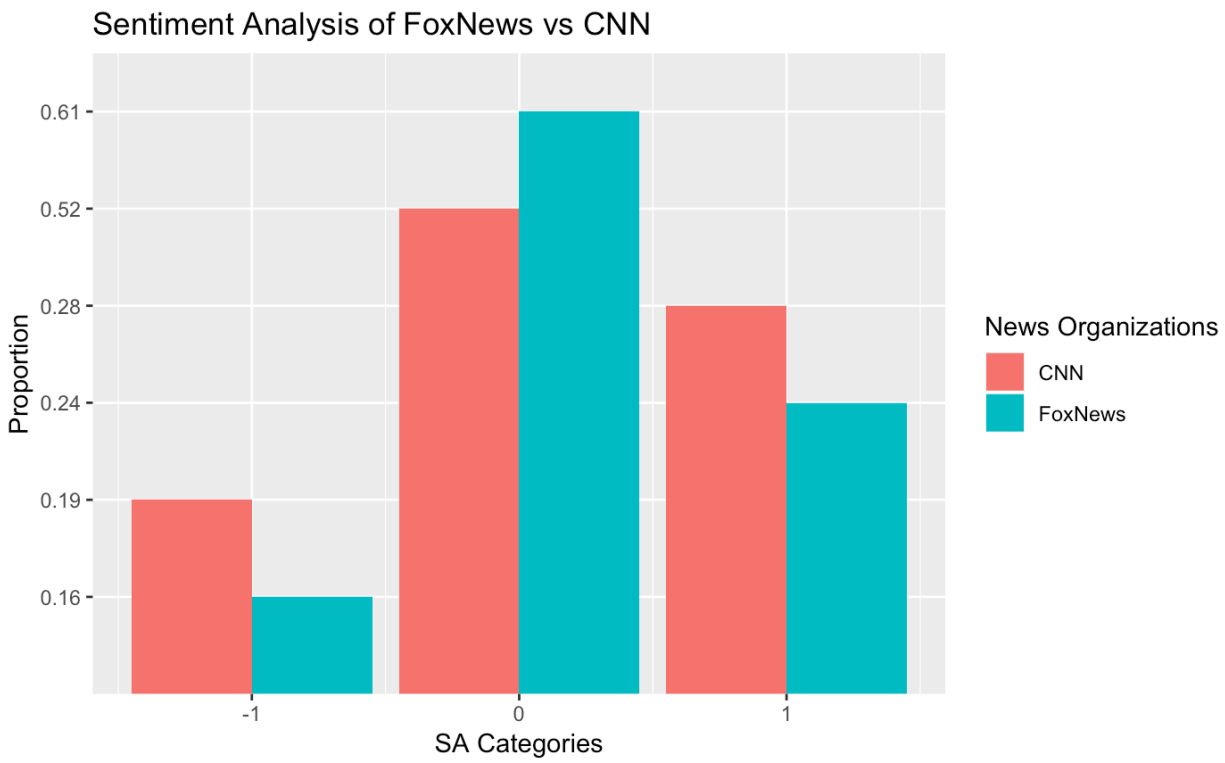

Figure 1