**Stat 5870 Project Proposal**

**Sentiment Analysis of Joe Biden and Donald Trump’s Tweets**

Elizabeth Chambers

**Datasets**

This project will utilize two datasets which each consist of tweets from Donald Trump (@realdonaldtrump) and Joe Biden (@JoeBiden). Each dataset contains variables such as the tweet “id”, “content”, number of “retweets” and number of “favorites”.

**Applications**

* Sentiment analysis: to classify each user’s tweets for opinion polarity (positive, negative, neutral).
  + In addition, the total count of sentiments for each user’s tweets will be plotted and analyzed to determine whether or not one user is more “positive” or “negative”.
* Word frequency: to analyze which words and expressions each user implements most frequently.
* Collocation: to identify words that commonly co-occur in each user’s tweets.
* Visualization: to better understand and represent trends in each user’s tweets.

**Methods and Tools**

In this project, I will utilize the aforementioned applications in order to compare and contrast the sentiment, sophistication and frequency of Joe Biden and Donald Trump’s tweets from 2009 to 2020. Sentiment analysis, word frequency and collocation will be conducted by use of the *TextBlob* package. Each tweet will be converted into a *TextBlob* object and the polarity of each tweet will be determined through the *sentiment()* function. Word frequency and collocation will be conducted through the *word\_counts()* and *ngrams()* functions, respsectively. The *matplotlib* and *seaborn* packages will be implemented in order to visualize, as well as compare and contrast the counts of words, word-pairs, and “positive”/”negative”/”neutral” tweets.

**Citations**

Reese, A. (2020, June). Trump Tweets. https://www.kaggle.com/austinreese/trump-tweets.

Vopani. (2020, October 31). Joe Biden Tweets (2007-2010). https://www.kaggle.com/rohanrao/joe-biden-tweets.