November 16

Alex: Built a Naive Bayes model to characterize sentiment of tweet language (compared with an elastic net, random forest and gradient boosted classifier, naive bayes performed the best) using the labeled Sentiment140 dataset from Stanford. Applied to our data to construct a sentiment score for indexing outrage.
Clarifying the need to use ideology bins: As an exploratory portion of our analysis we will also be looking into the type of engagement associated with different tweets considering their outrage level. If we look at the engagement rates by the retweeters ideology bin (i.e. if a strong liberal is retweeting a strong conservative post, it may be coming from a place of mockery) we might be able to understand if there's some sort of level where outrage goes over a 'reasonable' threshold and users then retweet to mock to delegitimize the original post.

Rob: After a suggestion to consider tweet2vec, a character-level word embedding model, to handle our single character and stop word anomalies in our original analyses I looked more into word embeddings again for their effectiveness in our modeling. I came to a similar conclusion from my initial research that word embeddings are extremely useful for neural network modeling and performing cosine similarities. However, for topic modeling the current preferred models for understandability are the LSI, and even moreso LDA modeling techniques. But once the topic modeling and extraction is complete it seems like it could be very useful to create a word embedding space and find words that are similar to our topic words in an attempt to further enhance our topic selection. In addition I will consider bigrams for modeling in our space to test any gains in topic modeling in place of entity recognition or other more advanced/unreliable methods.