**Final Project**

**Data 698**

**Candidate/Publication Sentiment Analysis**

**Question: Does media sentiment differ by news outlet at either the candidate or publication level?**

Null Hypothesis: media is not bias for candidates. Sentiment is normally distributed for each candidate for each news outlet.

Hypothesis: Media is bias, and sentiment differs per candidate per news outlet.

**Data Sources:**

1. **Newspaper Articles (100s of publications):**
   1. Select up to (7) outlets. The assortment should be a mixed of traditionally left-leaning, right-leaning and neutral syndicates
      1. Access: public API or HTML web scraping
   2. Select up to (5) candidates. The selected candidates should range from progressive (i.e. Warren/Sanders) to more moderate ( Biden/O’Rouke)
2. **Twitter**: corroborating evidence to prove hypothesis
   1. Are the candidates and affiliates re-tweeting or citing news syndicates?
   2. What is the sentiment of the tweets for the news sentiment for a candidate?
      1. Word maps [frequency] and/or network analysis
   3. What is the intensity of candidate coverage and associated responses?
3. **SEC Data**: find individuals and/or parent companies of the selected news outlets
   1. Is that individual backing a candidate?
      1. If so, could that skew the sentiment of the publication?
   2. Can the intensity of the sentiment and the donation amount be proportional?

Helpful Research:

* Canadian election paper <http://rpubs.com/justin_herman_42/523150>
  + Reach out to the people who wrote the medium(?)

Testing

* Chi-Squared Test

Toolkit

* Python (primary programming language) NLP/News API/Visualization
* SQL environment (SQLite?); Google Cloud if space is an issue

Next Steps

* Pick candidates and news outlets to research
* Pull the news API and Twitter data for the selected subset