


Data Science in Journalism: News Bias Project

Talia Berler

Plan

- Compare the bias in headlines of two news sources
 - Sources should be similar in style, on different sides of political spectrum
 - Take tweets containing a keyword (ie. Bernie, Trump) using twitter scraper
 - Definitive time period
 - Analyze frequency of appearance in tweets
 - Analyze emotional context using vader
 - Compare the frequency and emotional context around that subject for both subjects
- 

Goals and Relevance to RE

- Application to Journalism course
- “Everything is political”
- Puts evidence behind the knowledge that there is bias in the news
- Can be applied to any current event
- Helps anyone learn to be more meticulous about the way they digest the news



Dataset of all tweets

	index	date	tweet	retweet	user
0	783	2020-03-04 08:00:06	Once known for spoiling on grocery shelves, ca...	False	Wall Street Journal
1	782	2020-03-04 08:04:59	U.S. stock futures rise sharply as investors d...	False	Wall Street Journal
2	781	2020-03-04 08:15:05	Travelers returning from coronavirus hot spots...	False	Wall Street Journal
3	780	2020-03-04 08:30:11	For now, most people's health insurance will l...	False	Wall Street Journal
4	779	2020-03-04 08:45:08	Dozens of lawsuits accuse some of the country'...	False	Wall Street Journal
5	778	2020-03-04 09:00:00	Coronavirus fears have driven U.S. Treasury yi...	False	Wall Street Journal
6	777	2020-03-04 09:15:01	In South Korea, a rise in coronavirus cases un...	False	Wall Street Journal
7	776	2020-03-04 09:30:11	Thousands of refugees from Middle Eastern conf...	False	Wall Street Journal
8	775	2020-03-04 09:45:06	Protesters in Hong Kong have launched bomb att...	False	Wall Street Journal
9	774	2020-03-04 10:00:10	Lawmakers and federal officials, alarmed by th...	False	Wall Street Journal

Dataset of tweets with keyword

```
Bernie = df[df['tweet'].str.contains('Bernie')]  
Bernie.head()
```

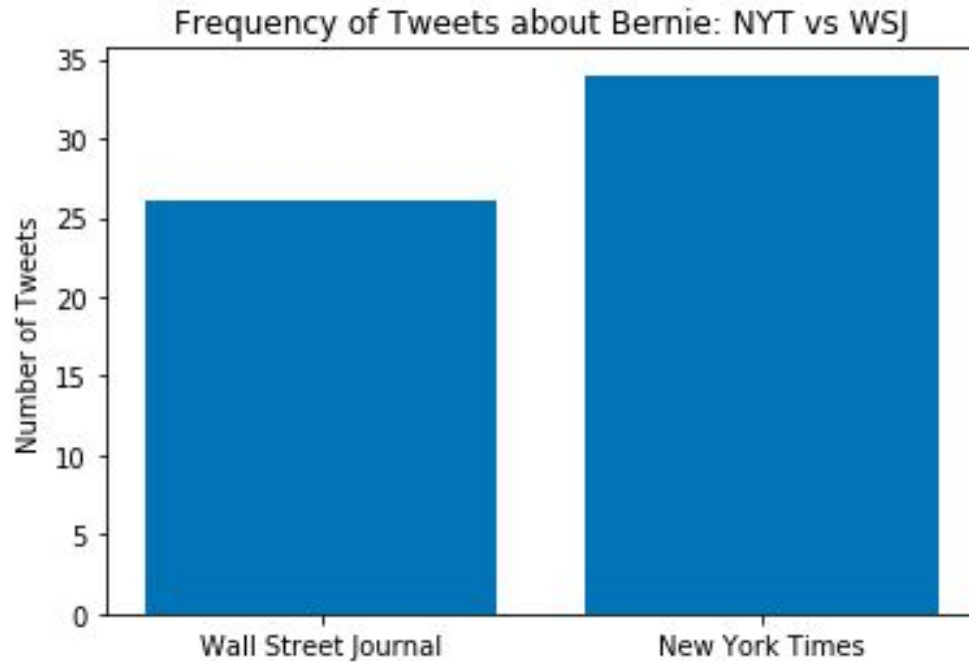
	index	date	tweet	retweet	user
10	791	2020-02-29 14:45:06	Some Southern Democrats fear that Bernie Sande...	False	Wall Street Journal
61	740	2020-03-01 08:00:09	Bernie Sanders uses the term proudly. Opponent...	False	Wall Street Journal
80	721	2020-03-01 14:00:09	Bernie Sanders lost badly in South Carolina, r...	False	Wall Street Journal
86	715	2020-03-01 16:08:05	If Bernie Sanders won the presidential electio...	False	Wall Street Journal
94	707	2020-03-01 19:30:07	Some on Wall Street don't consider Bernie Sand...	False	Wall Street Journal

Vader analysis output

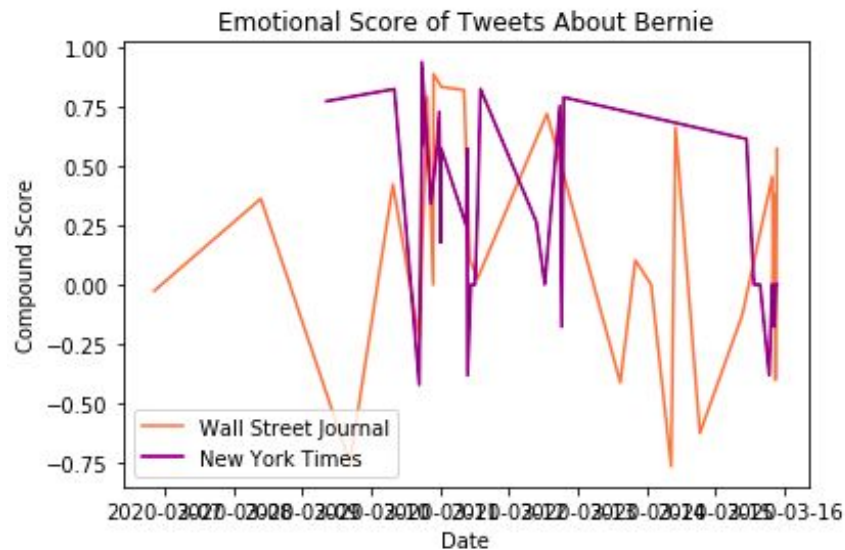
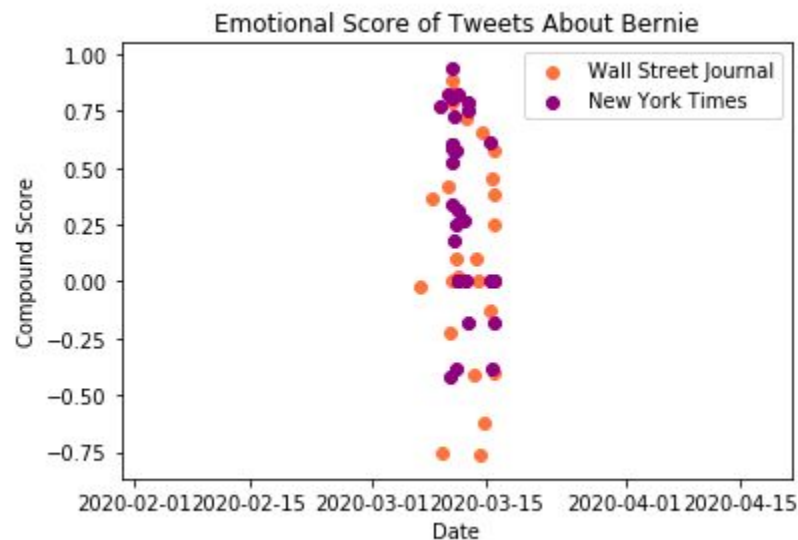
- Returns a dictionary
 - Neg = proportion negative emotion
 - Neu = proportion neutral emotion
 - Pos = proportion positive emotion
 - Compound is the emotion of the text sample overall
- What the numbers mean:
 - positive sentiment: compound score ≥ 0.05
 - neutral sentiment: (compound score > -0.05) and (compound score < 0.05)
 - negative sentiment: compound score ≤ -0.05

```
{ 'neg': 0.105, 'neu': 0.757, 'pos': 0.138, 'compound': -0.0 }  
{ 'neg': 0.0, 'neu': 0.882, 'pos': 0.118, 'compound': 0.5574 }  
{ 'neg': 0.162, 'neu': 0.838, 'pos': 0.0, 'compound': -0.6597 }
```

Frequency of Tweets



Emotional Scores



Moving Forward

- Find a way to pull tweets based on date posted, not page of twitter
- Explore other news sources (discuss with Margini)
- Choose a specific (controversial) topic to focus on
 - Election, trump, climate change
- Organize data and graphs into a more understandable format

