

# COMM313 Final Project

## Computational Text Analysis of #MeToo Movement Rhetoric

Anna Callahan

10 May 2019

```
In [6]: %run 'COMM 313 Final Project (1).ipynb'
```

```
Extracting the-new-york-times-a-sexual-encounter_-4b36d050-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-harvey-weinstein-indicted-49e638c6-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-old-vic-theater-4a8ca4ea-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-comedy-clubs-are-49a55018-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-court-nominee_s-accuser-4b53217e-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-police-charity-invites-4b44e898-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-'whose-side-are-4b1b8b74-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-at-yale_-trying-4ab5a066-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-asia-argento-denies-49991352-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-when-the-news-49d6abfe-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-rose-mcgowan-and-495ba878-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-'it-can-happen-4ab75d7a-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-_whyididntreport_-survivors-of-496498a2-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-5-ways-harvey-4aa60fc0-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-bill-clinton-says-4ad7b034-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-trial-judge-rules-4a63112a-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-special-prosecutor-is-4b439ede-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-to-weaken-rape-4a6cbd4c-58f1-11e9-a271-0242ac130002.txt
```

```
Extracting the-new-york-times-on-politics-with-4a982c70-58f1-11e9-a271-0242ac130002.txt
```

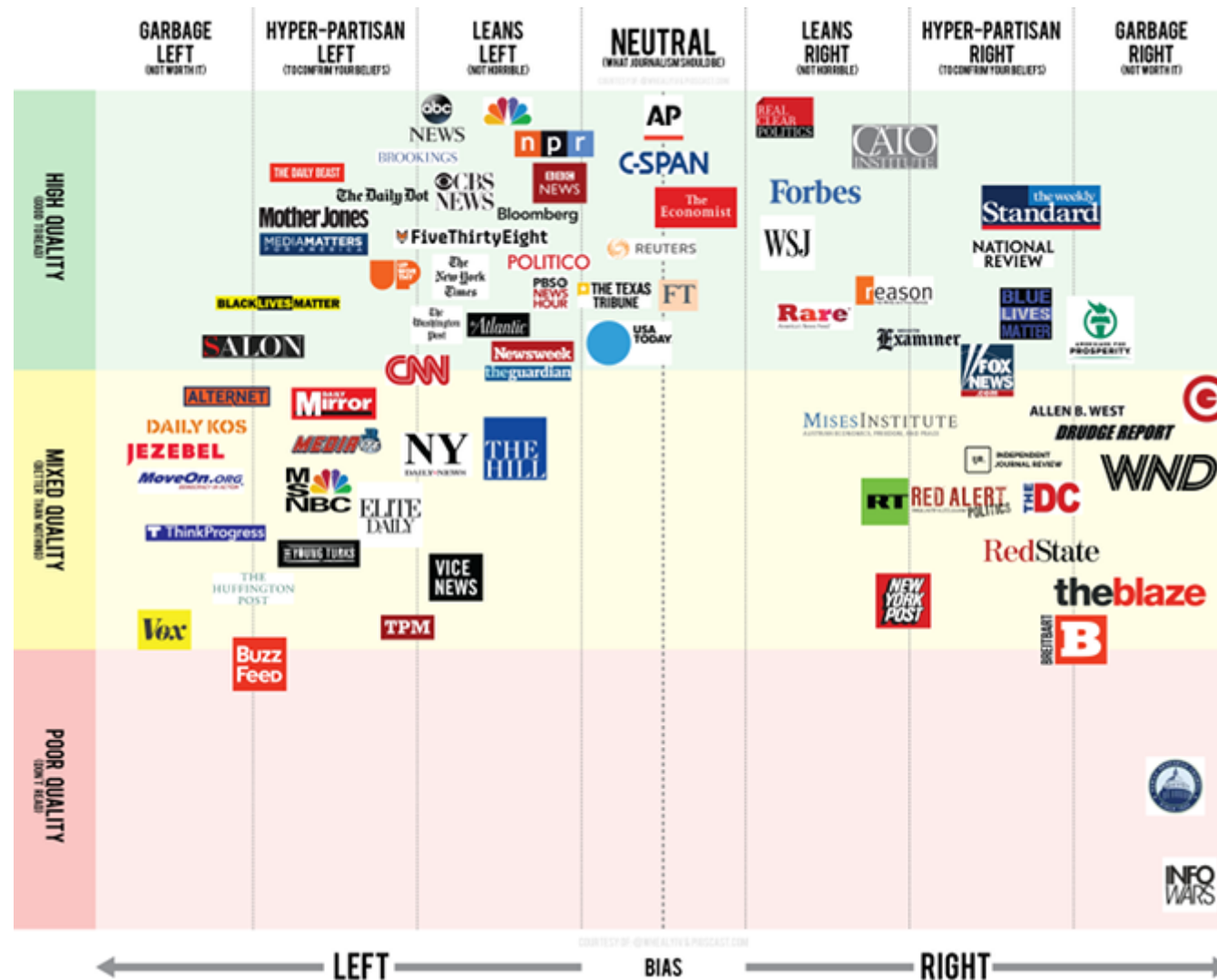


After a series of high-profile sexual assault allegations against now infamous Hollywood producer Harvey Weinstein flooded the news media in late 2017, the #MeToo movement began to spread virally.

The movement began as a hashtag on social media in an attempt to demonstrate the widespread prevalence of sexual assault and harassment, especially in the workplace. Tarana Burke, an American social activist and community organizer, began using the phrase 'Me Too' as early as 2006, and the phrase was later popularized by American actress Alyssa Milano, on Twitter in 2017.

Since the conception of the movement, hundreds of women (and men) have stepped into the spotlight to share their personal experiences of sexual harassment and assault with celebrities, partners, and coworkers alike. For several months, the mainstream news media was consumed with stories on these allegations. But the coverage of such a politicized modern social problem wasn't always objective.

#### For this project, I chose to delve deeply into the rhetorical choices made by journalists writing for media outlets positioned on opposite ends of the political spectrum: Fox News Network and The New York Times. The blog post that follows will explain how I accomplished this.

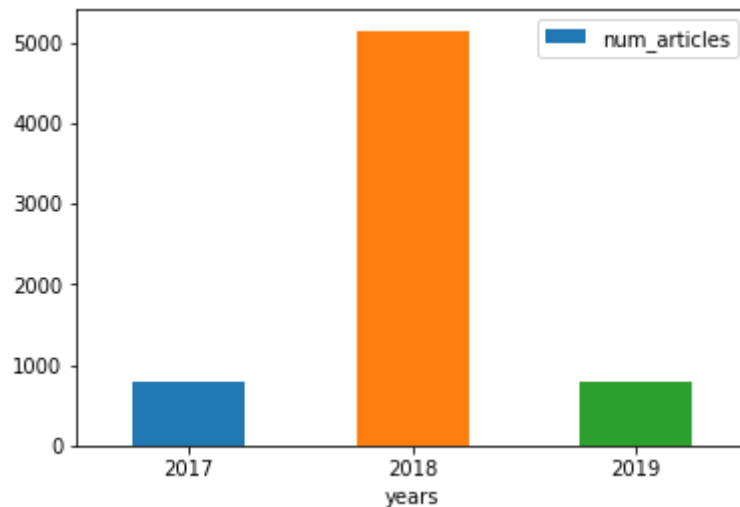


## The Data

I used the LexisNexis bulk download tool to accumulate a database of articles from Fox and The New York Times. The parameters of my search necessitates that articles contained the search term "#MeToo" as well as a number of characterizations of sexual assault (including victim, assault, accusation, etc.). The timeframe I set was January 2017 to present day, and the resulting corpus contained 6737 distinct documents.

Over the three years that the corpus spans, #MeToo was mentioned most in 2018. This makes sense because the movement didn't really take off until late in 2017 and it's only halfway through 2019. However, viewing the chart below it is clear that the number of articles mentioning #MeToo within the first five months of 2018 is much greater than in the first 5 months of 2019.

```
In [3]: import pandas as pd
import matplotlib
df = pd.DataFrame({'years':[2017, 2018, 2019], 'num_articles':[789, 5149, 799]})
ax = df.plot.bar(x='years', y='num_articles', rot=0)
```



## Collocation analysis

After tokenizing my data and adding key-value pairs for article texts and article tokens to each article in the corpus, the next step was embarking on collocation analysis.

Since my objective is comparing and contrasting linguistic choices of these two publications, I needed to divide my corpus into two subcorpora. Then I was able to investigate the collocates of words relevant to my research question for both publications. I started by comparing the lists of most commonly used tokens.

## THE NEW YORK TIMES

```
[('the', 272769), ('what', 32261),
 ('to', 168468), ('--', 32185),
 ('and', 148462), ('about', 31039),
 ('a', 120829), ('it's', 27098),
 ('of', 117484), ('all', 26744),
 ('that', 114284), ('think', 26559),
 ('you', 87602), ('know', 24525),
 ('is', 87498), ('at', 24229),
 ('i', 84050), ('as', 22778),
 ('in', 81602), ('going', 22621),
 ('this', 64871), ('do', 22492),
 ('it', 59602), ('president', 22090),
 ('for', 47577), ('there', 21772),
 ('on', 45774), ('people', 21218),
 ('we', 44576), ('if', 20142),
 ('they', 44176), ('now', 20009),
 ('he', 42977), ('who', 19841),
 ('have', 42870), ('has', 19470),
 ('are', 41249), ('from', 19302),
 ('was', 38485), ('just', 19064),
 ('but', 36539), ('right', 18287),
 ('not', 35975), ('she', 17793),
 ('with', 34147), ('that's', 17671),
 ('be', 33493), ('don't', 17133),
 ('so', 32532), ('his', 17064)]
```

## FOX NEWS NETWORK

```
[('the', 406017), ('not', 34984),
 ('to', 203843), ('an', 34354),
 ('of', 202026), ('had', 33678),
 ('a', 196463), ('has', 33336),
 ('and', 186798), ('but', 32916),
 ('in', 157514), ('from', 30731),
 ('that', 109491), ('be', 30326),
 ('for', 74473), ('women', 30157),
 ('on', 65885), ('ms', 29975),
 ('was', 63618), ('are', 29506),
 ('mr', 62736), ('about', 28662),
 ('is', 60955), ('they', 28512),
 ('said', 57835), ('sexual', 28336),
 ('he', 51288), ('this', 27619),
 ('with', 49989), ('new', 23423),
 ('she', 48765), ('or', 22855),
 ('as', 45926), ('their', 22253),
 ('it', 45279), ('been', 20705),
 ('her', 44576), ('one', 19791),
 ('by', 41527), ('were', 19716),
 ('who', 40619), ('when', 19113),
 ('his', 39409), ('you', 19032),
 ('at', 39200), ('more', 18295),
 ('have', 37107), ('-', 18183),
 ('i', 35177), ('what', 17600)]
```

Then I went on to compare the most common collocates of the words "weinstein" and "blasey" in both the NYT and FOX corpora. I found that

- NYT uses "Mr." in reference to HW more often than just "Harvey" or "Harvey Weinstein"

- "Mr" did not appear in the top ten collocates for "Weinstein"

In reference to "Blasey" (as in Christine Blasey Ford, the witness who alleged that she was sexually assaulted by then-prospective Supreme Court Justice Brett Kavanaugh)...

- NYT uses "Dr" most commonly to refer to CBF
- Fox uses "Christine" most commonly to refer to CBF

This pattern implies that there may be a distinction between the ways that Fox and The Times choose to refer women who come forward with allegations of sexual assault. The decision not to use a doctor's prefix indicates less of a sense of respect for that person. This is actually a broader societal trend, as [several studies have indicated \(https://www.liebertpub.com/doi/full/10.1089/jwh.2016.6044\)](https://www.liebertpub.com/doi/full/10.1089/jwh.2016.6044) that female doctors are referred to by their first names as opposed to their titles more often than male doctors.

I then conducted a series of KWIC analyses of the words "allegation", "accusation", "claim", and "against" with the goal of finding patterns as to how instances of sexual assault are described by these outlets.

#### THE NEW YORK TIMES

accusations	against	brett kavanaugh
	against	harvey weinstein
man who commits offenses	against	her
assault and inappropriate behavior	against	him
sex-crime claims	against	him
unanimous opinion	against	him judge
galvanizing opposition	against	him
harassment allegations	against	
accusations	against	judge
allegations	against	judge
accusation of sexual assault	against	
accusations	against	judge
allegations	against	justice
accusation	against	mr
accusation	against	mr

#### FOX NEWS NETWORK



accusation of sexual misconduct	against	judge
she came out	against	judge
accusations	against	kavanaugh
the new york times	against	kavanaugh
for them to vote	against	kavanaugh is a mistake
no corroboration no witnesses	against	poor brett kavanaugh
women that came out	against	weinstein
cautioning	against	a rush to judgment
sexual misconduct allegations	against	candidate
sexual misconduct allegations	against	
is now being weaponized	against	him in a way
were being thrown around	against	this esteemed judge
own ordinance by discriminating	against	chick-fil-a based on the
no corroboration no witnesses	against	poor brett kavanaugh these
the system is rigged	against	them listen this is

The vast majority of the ways that The Times and FOX characterized instances of sexual assault in articles that spoke of #MeToo and Harvey Weinstein were standard and uninteresting. But I did notice a few distinctions as well. While both the Times and Fox used phrases like "accusations against" and "allegations against," when referring to the cases against Weinstein or Kavanaugh, the Times never used the phrase "sexual misconduct" to describe the instances of assault that occurred. Fox, on the other hand, used that phrase frequently. Personally, I feel that using the phrase "sexual misconduct" to describe an instance of sexual assault or rape seems to me like a significant downplaying of the severity of the action that took place.

#### Typical Rhetoric Used

The New York Times	Fox News Network
accusations against	accusations against
allegations against	allegations of misconduct
accusation of sexual assault against	accusation of sexual misconduct against

From conducting KWIC analysis on the word "claim", it is interesting to note, that in the NYT sample, many of the times "claim" was used were situations in which the credibility of that statement was questionable or proven false. In comparison, the Fox "claim"s were not associated with particularly dubious statements, but rather more generally used.

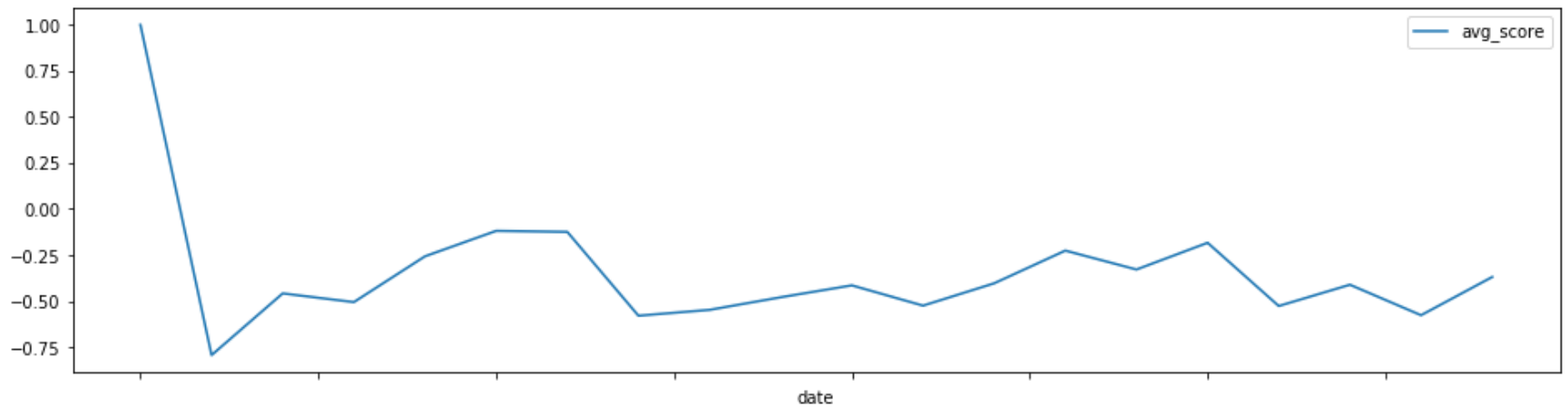
Since I am using a corpus of news articles which can sometimes be more noisy and less relevant to the research question at hand, I conducted a series of spot tests and concluded that because I used so many search parameters in addition to #MeToo during my LexisNexis search that the vast majority of the articles specifically pertain to my research question. I then conducted VADER analyses on text from both the NYT and Fox corpora.

### VADER Polarity Scores

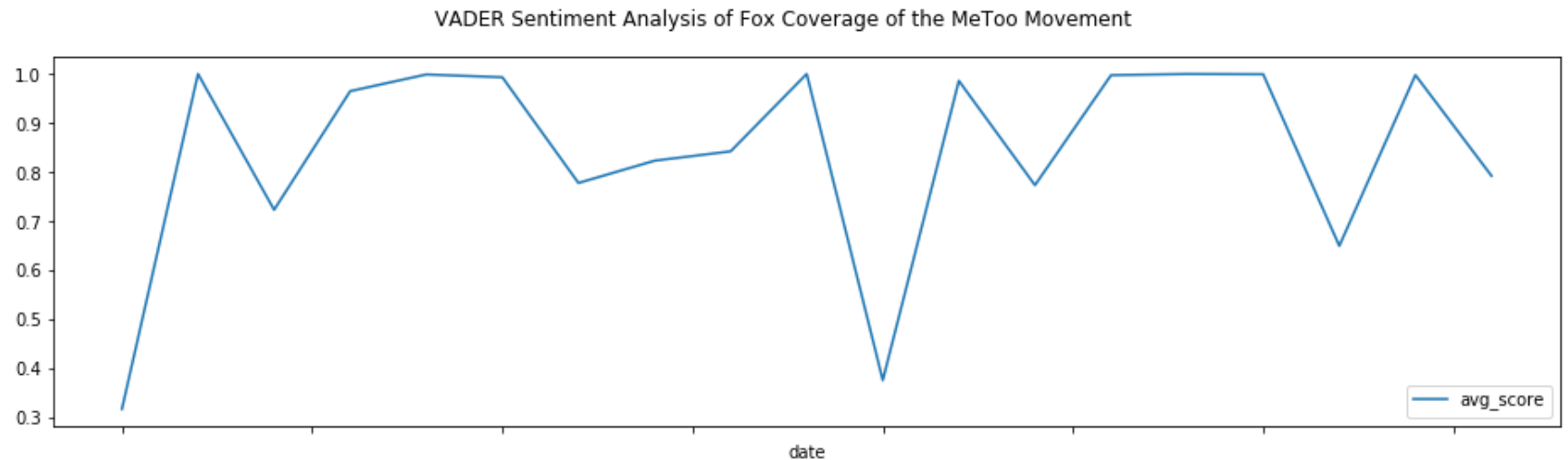
**The New York Times**    **Fox News Network**

-0.3883453621981673    0.8390683060109336

VADER Sentiment Analysis of NYT Coverage of the MeToo Movement







From these charts, we can see an interesting trend in the compound polarity scores of the texts about the #MeToo movement respectively. All of these patterns alone are inconclusive, but together they begin to paint a picture representing the general feeling towards the #MeToo movement held by journalists and executives at these two top publications.

The use of the word "claim" in addition to "allegation" and "accusation" implies a dubiousness about the truthfulness of the story. Referring to a female doctor by her first name as opposed to her official title connotes a lack of respect if intentional and gender-based discrimination if subconscious. Lastly, a positive sentiment score when describing allegations of sexual assault may be indicative of indifference or dubiousness in place of using sympathetic language that emphasizes the gravity of the alleged crimes.

These were all trends noted in the Fox News Network subcorpus, supporting my theory that identity politics surrounding the #MeToo movement influence the way that conservative media outlets describe sexual assault.