



# COVID-19 Tweet Topic Modeling

By Sullivan Myer and Brayton Hall

# Data Collection and Purpose

## Twitter

- Used the python 'twint' library to **scrape almost 100,000 tweets related to 'coronavirus'**
- We then filtered the tweets, using Vader sentiment analyzer, to find **sentimentally non-neutral tweets**

## Topic Modelling

With UNSUPERVISED LEARNING:

Find the most intelligibly-distinct Twitter communities based on natural language processing, e.g. tweets related to: **'Trump', 'Health', 'Economy', 'Employment', 'Hope'**

## Sentiment Analysis

After finding distinct topic communities on Twitter, we intended to find the distribution of sentiments across those communities

Trump-Tweets:

POSITIVE?

Health-Tweets:  
NEGATIVE?

Economy-Tweets:

**CRISIS?**

# Cleaning and Pre-processing

## Stop Words

**Remove the least informative words from the set, such as:**

- a, the, and, but
- health, virus, test
- coronavirus, corona, covid-19, covid19, covid2019

## Regex

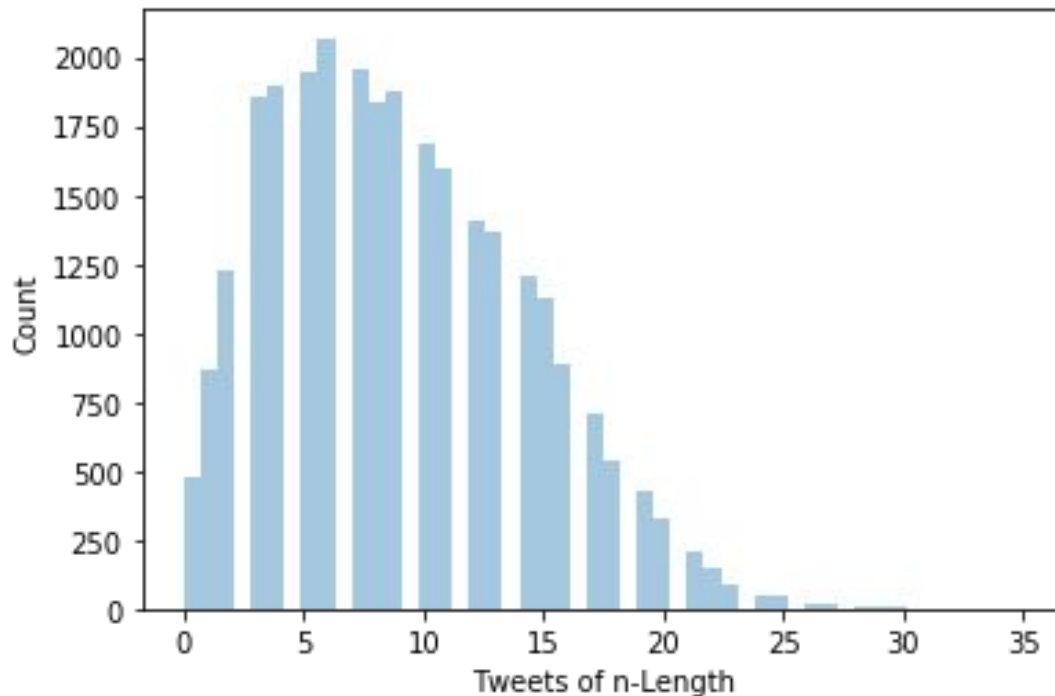
- Remove new line characters
- Remove Emails
- Remove quotes

## Tokenize

**Finally, convert the entire set of tweets into a BAG OF WORDS using a pipeline, resulting in the 'corpus' to use in modeling**

# Tweet Distribution

- Mean length of Tweets is 9 interesting words (after stop words were removed)
- Total of 27,890 Tweets after filtering for sentiment ( $x < -.40$  or  $x > .40$ ) and stop words
- Right-skewed, so most tweets fall between 4 to 12 interesting words



# The LDA Model

(Latent Dirichlet Allocation)

LDA is a 'generative probabilistic model' which is similar to K-means clustering, but allows for *multiple memberships*, or 'fuzzy' memberships.

So a tweet involving 'Trump' and 'Economy' will, based on **other words** in the tweet, have a maximum probability corresponding to one of them

---

# LDA Results

Health ->

U.K. ->

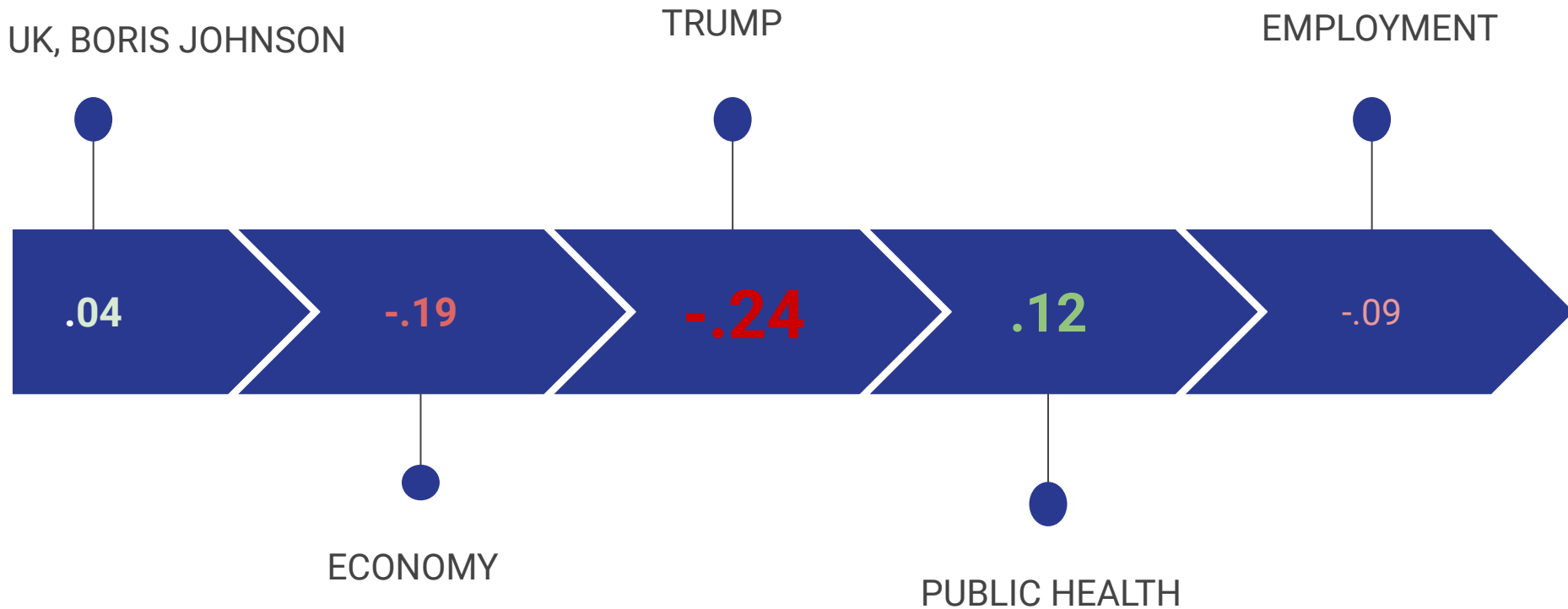
Economy ->

Trump ->

Employment ->

```
1 lda_model.show_topics()  
  
[(0,  
  '0.022*home" + 0.016*support" + 0.014*event" + 0.014*be" + 0.014*give" + 0.014*hospital" + 0.014*business" + 0.014*health" + 0.013*safe" + 0.011*get'),  
 (1,  
  '0.057*virus" + 0.021*sign" + 0.019*boris_johnson_demand" + 0.014*fear" + 0.013*article" + 0.012*come" + 0.012*thank" + 0.011*make" + 0.011*relief" + 0.010*week'),  
 (2,  
  '0.028*outbreak" + 0.025*say" + 0.018*fight" + 0.017*petition" + 0.016*death" + 0.014*s" + 0.014*way" + 0.013*economy" + 0.012*live" + 0.011*really'),  
 (3,  
  '0.029*crisis" + 0.026*trump" + 0.017*case" + 0.014*think" + 0.013*die" + 0.013*watch" + 0.013*update" + 0.011*video" + 0.010*life" + 0.009*use'),  
 (4,  
  '0.030*need" + 0.023*tackle" + 0.017*lockdown" + 0.013*job" + 0.012*bad" + 0.011*free" + 0.011*share" + 0.010*number" + 0.010*money" + 0.010*call')]
```

# Sentiment Scores for Topics





# UK and Boris Johnson

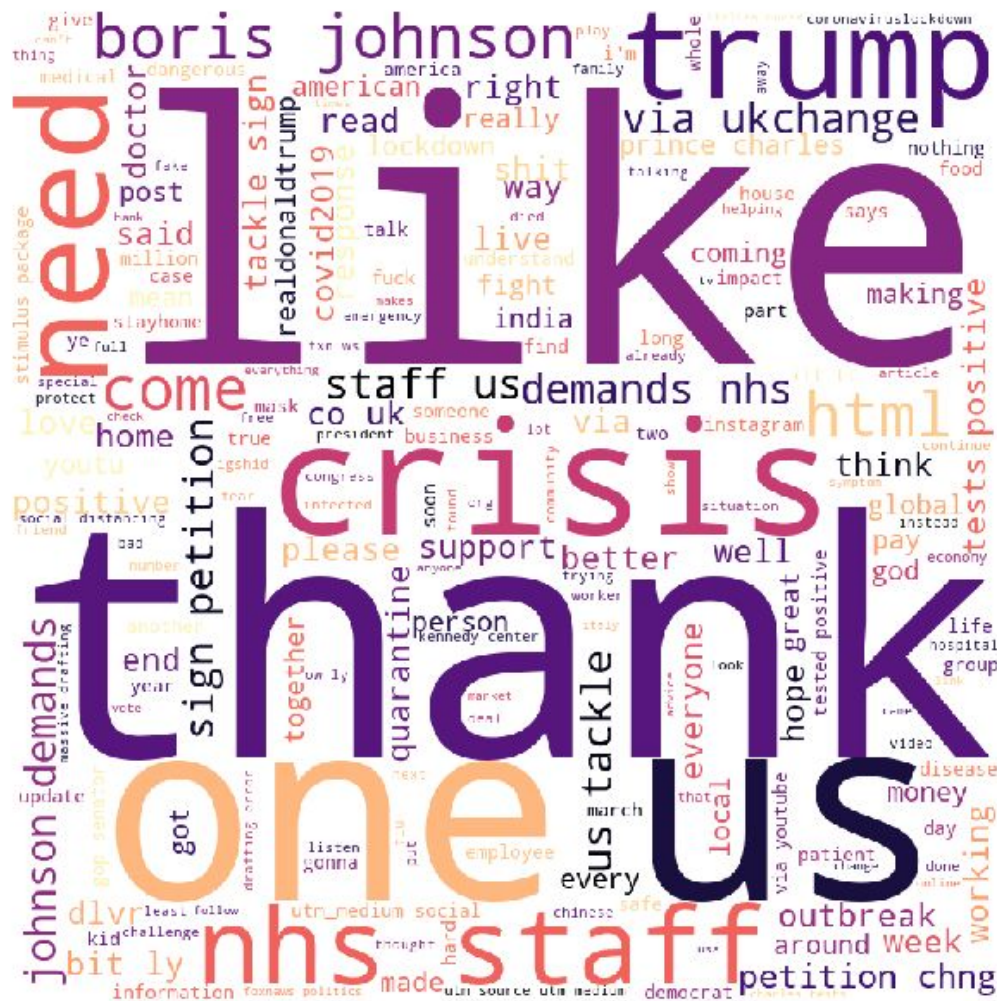
## Most Negative

Coronavirus: 560,000 offer to volunteer in the UK's health service" #Coronavirus

[https://twitter.com/ievents/1219057585707315201](https://twitter.com/ievents/1219057585707315201?lang=en) @TheKatieForbes How Are You Helping Planet Earth Besides Posting Insensitive Offensive Shit About This #Heartless

## Most Positive

The only absolute positive about coronavirus is the impact it has had on the launch of 'just Harry and Meghan' 😄😄😄😄😄😄😄 I know she will be cursing this global pandemic for no other reason that to its detriment of HER launch!!!!!! 😄😄😄😄😄😄😄 BRILLIANT!!!!



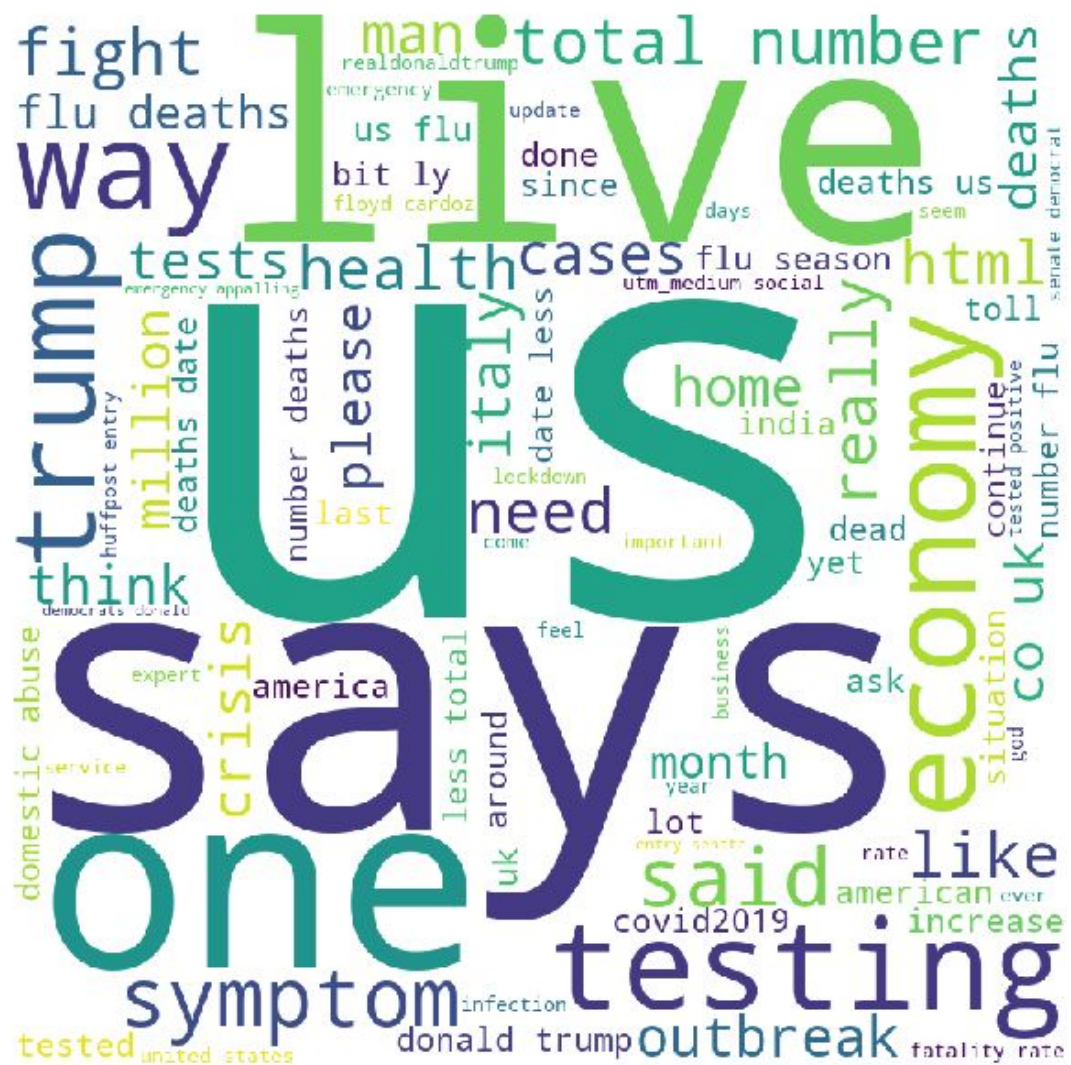
# Economy

## Most Negative

We are at war! And sadly, against our own Democrat 'lawmakers' who literally weaponized the #coronavirus to push their agenda and to destroy President @realDonaldTrump's hard work that saved the economy. What they are doing is disgusting! This November they will  
LOSE

## Most Positive

Lets all take a moment to commend the marvelous job glorious leader Kim Jong Un has done in containing the coronavirus in the Democratic People's Republic of Korea. 0 cases and 0 deaths to date, an inspiration for every nation who have been unable to contain this.





# Employment

## Most Negative

Not many people are talking about it but it needs to be said.

Unemployment links to more suicide deaths. Social isolation only worsens it. The containment strategy for the coronavirus will sadly lead to many tragedies. I am urging you to check on your loved ones. IT IS VITAL rn.

## Most Positive

Hey good people, I will probably not be able to get a job (thanks to coronavirus 🙄🙄) so if you want to help a broke college student not starve to death and pay some of this off, then please send something or RT. Thanks 💜



# Public Health

## Most Negative

In coronavirus pandemic, Trump allies say they're ready to die Did these guys miss their appointments last week? They are talking about geriatric genocide.How evil can they get?These guys are manics, fools, and evil!!!

## Most Positive

haha quarantine so funny haha  
remember to wash ur hand haha  
coronavirus haha funny haha its  
funny guys hahahahhha





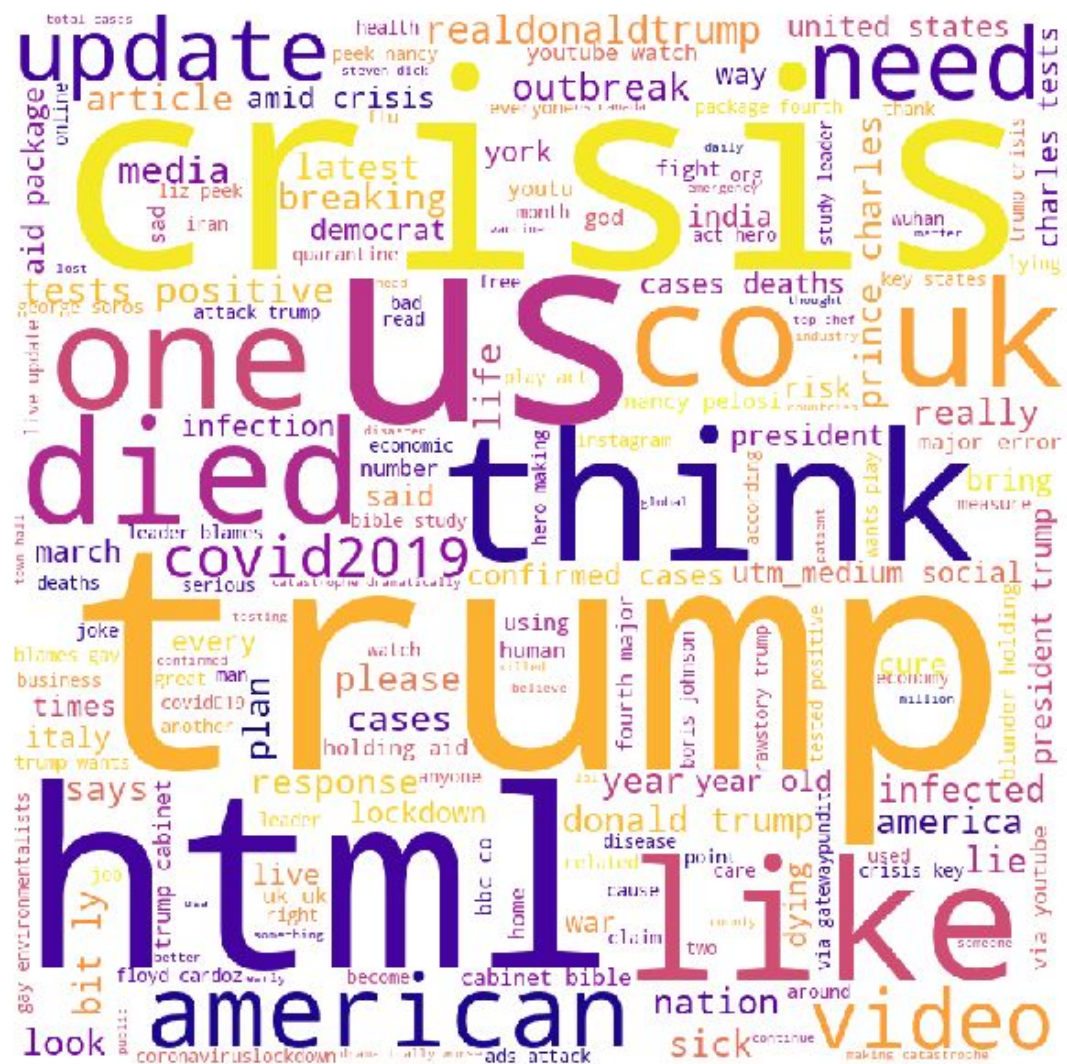
# Trump

## Most Negative

Ignoring the Coronavirus Trumps followers are like Hitler and his generals they followed a MAD MAN and his idiot leadership into a devastating war that left millions dead the only difference between Hitler and Trump was that Hitler was an educated fool and Trump is just a fool

## Most Positive

'Rob Reiner loves the spotlight. He is irrelevant....no one cares! Trump is our great President who is handling everything just fine! We the people love the job he is doing for all American Citizens. He will win the war with the Coronavirus! We will re-elect Trump 2020!



## TODAY:

# March 27th, all of Twitter around 10am, after Boris Johnson Tests 'Positive' for Covid-19



# Topic Modeling Conclusion

- Useful for unsupervised clustering in NLP and social media analysis
- Further analysis could use Word2Vec to find deeper patterns in Tweets based on word associations and sentence completions with LSTM, for example
- Incorporating this data into a time series would be useful to track sentiment *over time* in respect to particular Twitter communities