

# Analyzing Twitter Conversations About Vaccines

By Michaela Brydon  
Kenyon College

## Introduction

As of today, there are two COVID-19 vaccines being distributed to healthcare workers across the country. Some are thrilled at this news, hopeful that vaccines will put an end to the quarantines and social distancing. Others are more skeptical, worried at the prospect of mandatory injections. According to the experts, for the US to reach herd immunity, 60-70% of the population would need to be immune to the disease. As such, if not enough people are on board, then vulnerable populations will still be at risk. In light of this dilemma, I decided to focus my final project on vaccination. Specifically, I collected recent vaccine-related tweets and explored them using natural language processing techniques.

## Preparing the Data

1. First, I used the Python package Twint to collect tweets from four separate queries, each using the terms "Vaccine", "CoronavirusVaccine", "CovidVaccine", and "Covid19Vaccine", respectively. I started out with around 10,000 tweets.
2. After combining the raw collected data into one file, I performed data cleaning. This included dropping empty columns, filtering out rows with NULL or non-English tweets, and removing duplicate entries.
3. Next I focused on preparing the tweets for further analysis. I started by tidying up information within the tweets, swapping emojis for their text representations with the emot package, replacing URLs with the string "URL" and mentions with "MENTION" using re, and formalizing contractions.
4. With the help of the nltk package, I then tokenized the cleaned tweets, transforming them into lists of terms, removed stop words from this list, and lemmatized the resulting terms.
5. Once the pre-processing was complete, there were 8799 tweets remaining for analysis.

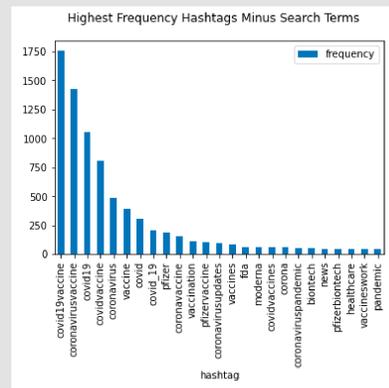
## Acknowledgements

I would like to Professor Chun and Professor Elkins for their support with this project. Thanks for a great class!

## Frequency Analysis

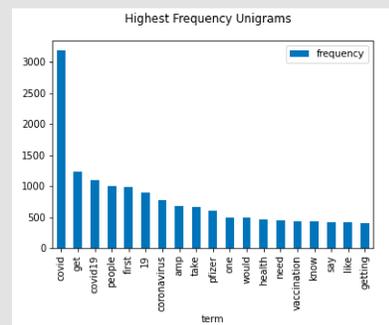
Once the pre-processing was done, I performed frequency analysis, identifying top hashtags, mentions, and terms.

Unsurprisingly, the top hashtags are either related to coronavirus, vaccines, or vaccine-producing companies like Pfizer. Though, it is notable that these are all very neutral. The only polarized term to make the top twenty was "VaccinesWork", and even it is only used 43 times.



The top mentioned users include politicians, news outlets, health organizations, and pharma companies. As none of these are surprising, I will not show this figure for the sake of space.

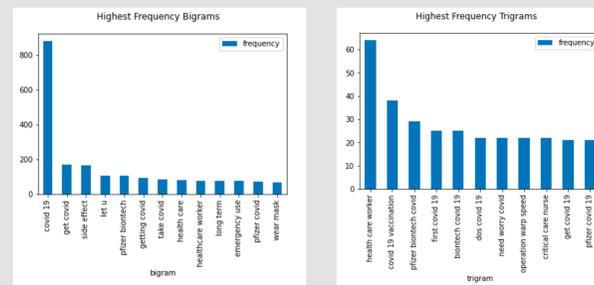
Next I considered the most common unigrams, bigrams and trigrams in sample of tweets, disregarding removing URLs, mentions, and any of the four search terms I used to find the tweets. The following bar chart displays the most used unigrams in tweets.



Most of them are coronavirus-related nouns, though there are some verbs, such as 'get' and 'need'. While the first does not indicate one's vaccine preference, I suspect that latter is more likely to pop up in pro vaccine tweets, as people encourage the collective to get vaccinated.

## Top Uni-, Bi-, and Trigrams

Next we will take a look at notable high-frequency bigrams and trigrams. While "covid 19" is by far the most commonly used bigram, we also notice phrases such as "side effect" and "long term", which may come from people concerned about the potential risks associated with the new vaccines.

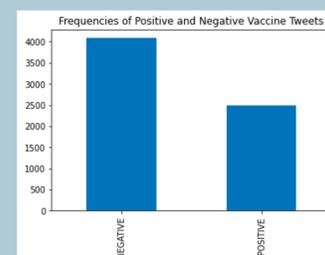


Given that "Covid 19" was the most common bigram by a long shot, it is no surprise that the phrase shows up in most of the top trigrams as well. With phrases like "health care worker" and "critical care nurse" showing up near the top, it is clear that people are talking about medical personal in the context of vaccines. With that being said, further investigation is needed to determine what these conversations look like.

## Sentiment Analysis

I also used Flair to perform sentiment analysis on the tweets, labeling each positive or negative.

In my dataset, tweets about vaccines were usually negative. Specifically, 62% of the tweets surveyed were labelled negative, while 38% were labelled as positive.



## Conclusion

First, it is no surprise that coronavirus is flooding the twitter conversation about vaccines. With all of the news coming out on the pandemic and the new vaccines hitting the shelves, there are plenty of tweets sharing facts and information which are drowning out public debate. Based on my analysis and according to the Flair model used, tweets about vaccines tend to be negative. However, this does not say much about how the public feels about vaccines, only how they feel when they discuss vaccines. The frequency analysis on hashtags and terms revealed that neutral words and phrases can be seen frequently in vaccine tweets. With that being said, this work was purely exploratory and meant to inspire further investigation.

## Future Work

Given more time, there is plenty I would have liked to include in this project. I would have liked to produce word clouds instead of the histograms presented. Also, I would have presented a network of mentions, determining which users were the most connected.

I also would have like to do more with the sentiment analysis. My original plan was to use sentiment scores as a proxy for a tweet's feelings towards vaccines, but this proved too unreasonable. In the future, I could train a model to be able to categorize tweets by their sentiment towards vaccines. I also considered using Entity based Sentiment Analysis, but I struggled to find code which would accomplish what I hoped to achieve.

## References

- Code modified from:  
<https://github.com/programminghumanity/iph200fall2020>
- <https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f>
- <https://towardsdatascience.com/a-complete-exploratory-data-analysis-and-visualization-for-text-data-29fb1b96fb6a>
- <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html>
- <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/herd-immunity-and-coronavirus/art-20486808>
- <https://www.muhealth.org/our-stories/covid-19-vaccine-key-reaching-herd-immunity>
- <https://www.raps.org/news-and-articles/news-articles/2020/3/covid-19-vaccine-tracker>