



PERCEPTION OF COVID-19 DISTANCING POLICIES

Team CuriousDuo

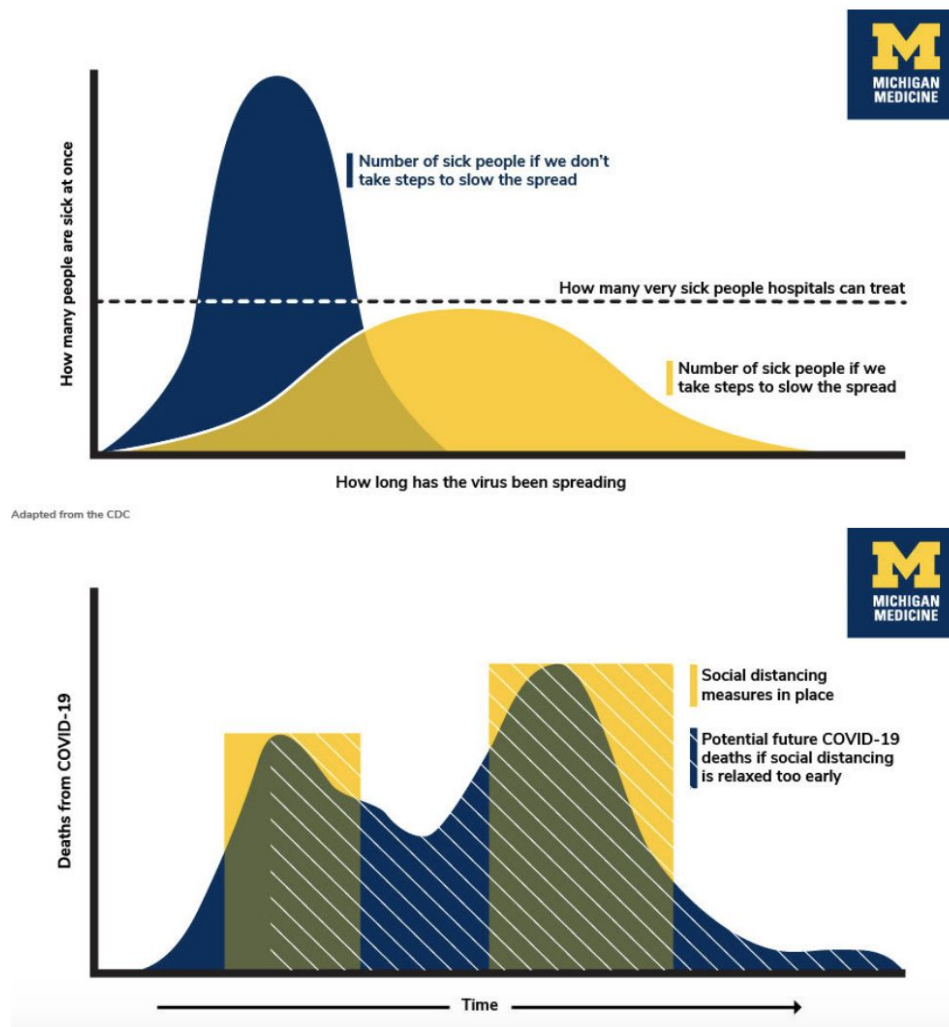
Vishwa Pardeshi, Varun Nair

Introduction

Coronavirus is a family of viruses that causes illness which can vary from common cold and cough which also leads to severe health conditions diarrhea, vomiting, chills, headache and much more.

SARS-CoV-2 (n-coronavirus) is the newest entry of the coronavirus family which was discovered in the year 2019. Moreover this virus was not identified in any humans before.

Due to the rate of spread, on March 11 2020, the coronavirus (COVID-19) outbreak was declared a global **pandemic** by **WHO**. As of now there are over 430k+ deaths and over 3M active cases.



The image above shows the recommended tactic in-order to control the cases.

As the world becomes more connected, opinions and information are easier to share and access thanks to several blogging & microblogging sites such as Twitter. Twitter is now being increasingly used to champion causes and provide support during crises which makes it a gold mine of first-hand opinions and sentiments of real human beings on current events. This can be used to understand how people perceive COVID-19 and the various policies in place to “flatten the curve”.

Naturally, twitter users discuss the latest development and react to local developments. However, this pandemic has generated a misinformation pandemic which has influenced the extent of seriousness towards COVID-19 and its policies. This was noticed predominantly in the United States of America, which now has the highest number of COVID-19 cases. While there were states which went into lockdown peacefully and willingly such as Washington, there were other states such as Florida where people vehemently protested against restrictive policies.

We believe that this sentiment will be captured in twitter data and hence can be utilized by the government at all levels to create effective strategies to inform the public better.

Objective

For the purpose of this hackathon, we will be focusing on two states- Washington and Florida for our analysis. The objective is to

- ❑ Identify & collect tweets from the states
- ❑ Identify the sentiment trends for the state-specific user and how this impacts the spread of COVID-19

Data

Data Collection & Preprocessing

The tweets collected by hydrating the tweet IDs which are available on this public [github repository](#) using Twarc & Twitters Developers Account. We collected approximately nearly a million tweets. However, there were a few inconsistencies such as

1. Location is not always recorded & not in a uniform format
2. Not all the tweets are in English language
3. Not all the tweets were from USA let alone Washington/Florida

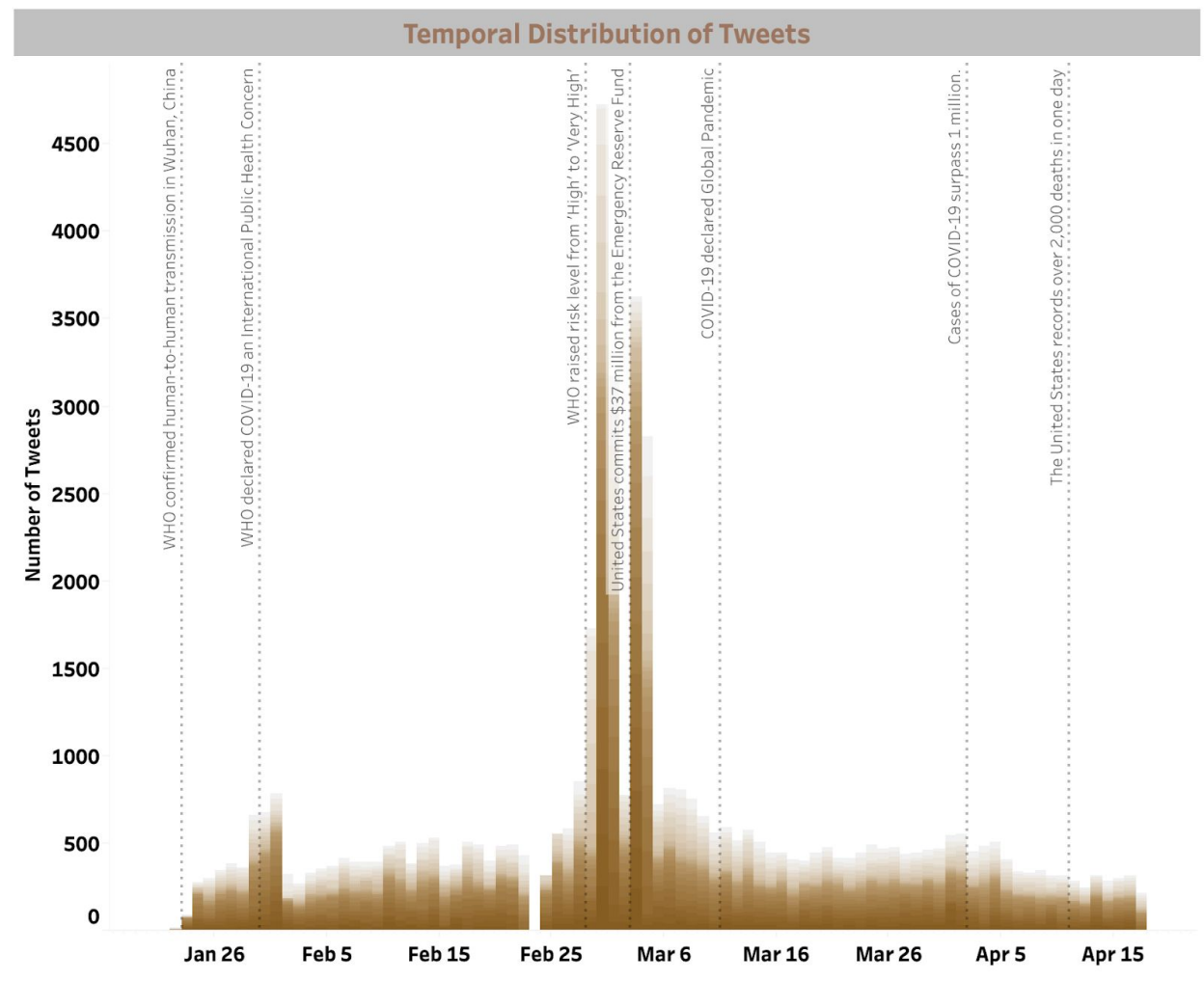
Apart from the tweets, another dataset was used inorder to find the correlation between the tweets and the targeted states which matches our solution of finding out as to how cases are rising in these two states particularly. The [kaggle dataset](#) used contained three different files that helped us discover more about the situation. It comprised datasets from counties, overall cases of each state combined and individual cases.

Exploratory Data Analysis

Tweets

Approximately 50,000 tweets were collected. Due to the limited time and restrictions of Twitter Developer Account we were able to get data from January 23, 2020 to April 17, 2020.

Temporal Distribution of Tweets



January 23, 2020 is chosen as the starting date as it is the date when WHO confirmed that human-to-human transmission has been recorded in Wuhan, China.

From the temporal distribution we notice that:

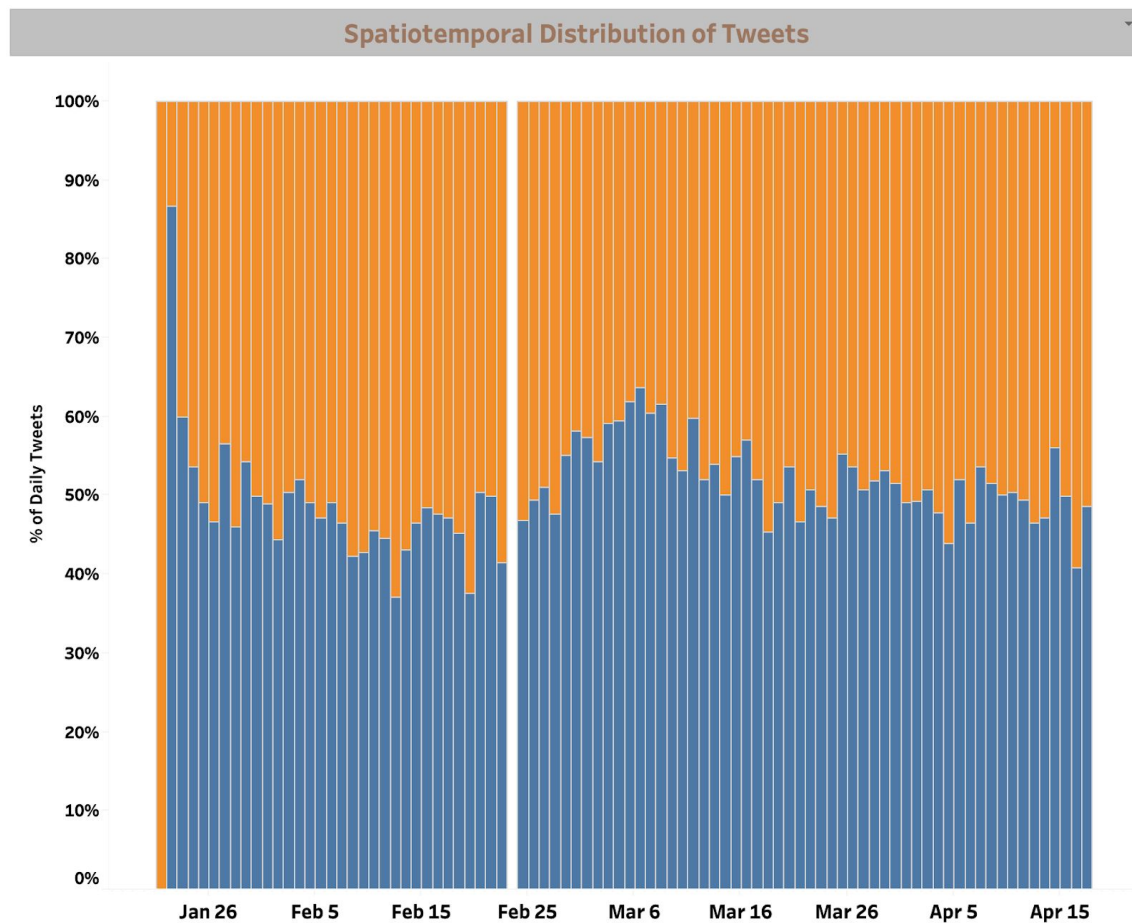
- Spike in the number of daily tweets usually correspond to major news being announced.
- Disproportionately high numbers of tweets are observed for two days -
 - Feb 28 - WHO raised risk of COVID-19 from 'High' to 'Very High'
 - March 3 - United States committed \$37 million to Emergency Reserve Fund

Note: As Florida is nicknamed the Sunshine State and Washington is infamous for being overcast, Florida is represented with Orange whereas Washington is represented with Blue.

States Coordinates

- Florida
- Washington

Spatiotemporal Distribution of Tweets



It can be noticed that there is an equal distribution of tweets between the two states - Washington and Florida.

Methodology

Our methodology is as follows:

1. Collect Tweets by hydrating
2. Filter tweets by location (reverse-geocoding used)
3. Exploratory Data Analysis
4. Sentiment Analysis

Twitter Sentiment Analysis

Sentiment Analysis is the interpretation and classification of emotions within text data using text analysis techniques. This allows us to understand public perception on specific issues and thus can be extended to understand the general public sentiment of Washington and Florida residents towards COVID and its policies.

For performing Sentiment Analysis, we used Valence Aware Dictionary and Sentiment Reasoner (VADER). Since the text we have is from Twitter, the texts can be classified as microtext. Microtext is text that is generally very short in length, semi-structured, and characterized by amorphous or informal grammar and language. Thus, it often contains emojis, capitalization which convey emotions and thus, should be considered while deriving the sentiment of a tweet.

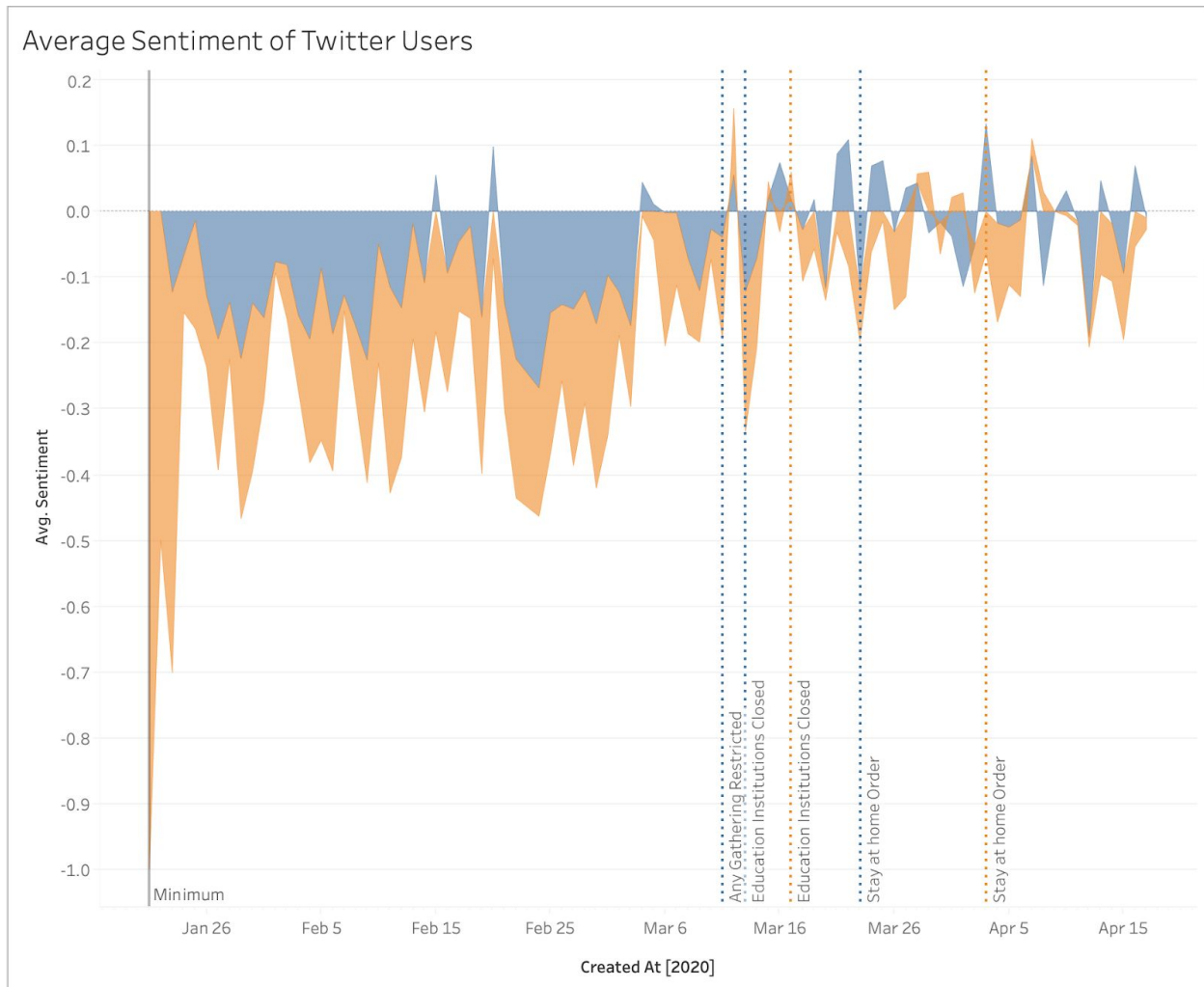
VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words), which are generally labeled according to their semantic orientation as either positive or negative. Vader uses a combination of capitalization, punctuation, word meaning, and emoji interpretation (which is used frequently in tweets) to give a percentage based score to each tweet according to the positive , negative, or neutral components contained within the tweet.

Insights and Results

Result I

It is important to note that we are working with the daily average sentiment for each state. To understand how the users responded to guidelines being imposed by the state government, here we have marked the important dates such as schools closing, gathering being restricted and stay at home order being imposed by both the states. While the sentiments were largely negative before the guidelines were imposed for either state, the trend changes once these were announced.

The average cases in Florida overtook Washington which initially was a hotspot. This can be attributed to public perception which varies greatly in the two states as is observed from the daily average sentiment trend.



For Washington Twitter Users:

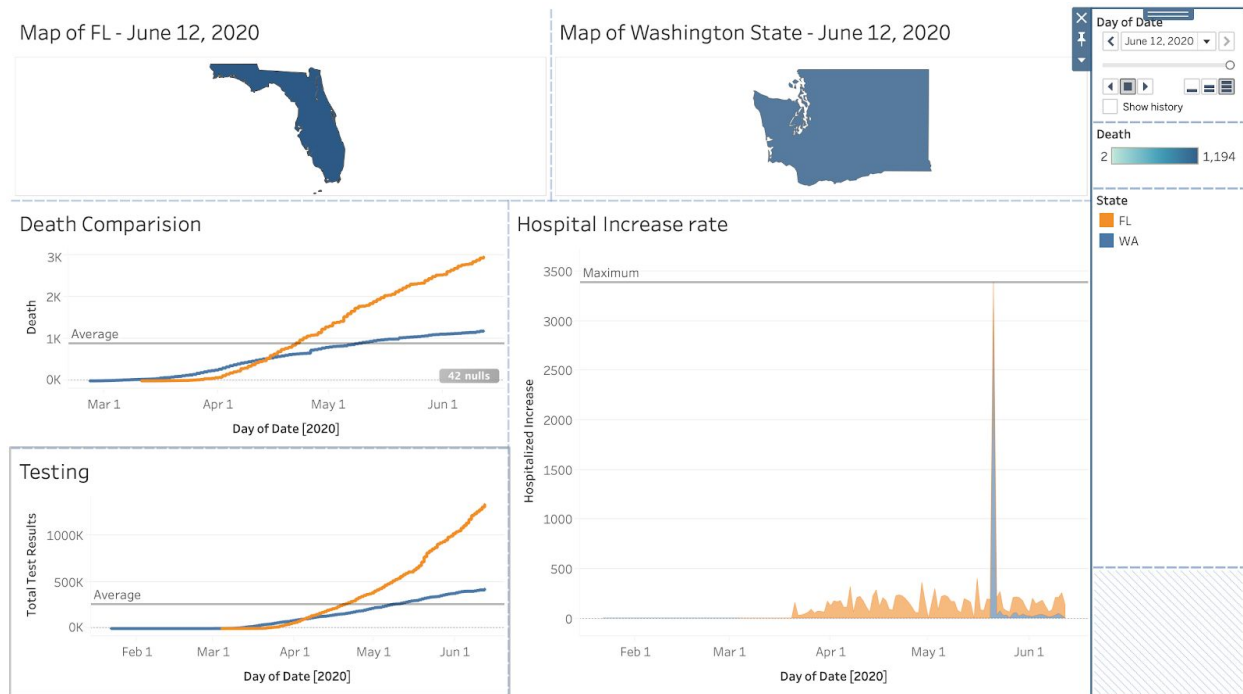
- The sentiment is largely negative before any local restrictions were announced
- The first blue line marks the advent of social distancing and is received with positive sentiments from Twitter Users.
- The second blue line marks the closure of educational institutions which elicited negative sentiment as there was uncertainty related to completion of the academic year and its impact on students residing in resident halls
- The third blue line marks the announcement of Stay-at-home order which is received positively

For Floridian Twitter Users:

- While Washington twitter users on average had more positive reactions as compared to Florida. This is made evident with the orange dips.
- Each of the announcements- highlighted by the orange vertical lines - is followed by an average negative sentiment.

Result II.

The below Dashboard shows the entire analysis of the two states - Florida and Washington



Based on the above Dashboard we have had the following observations based on -

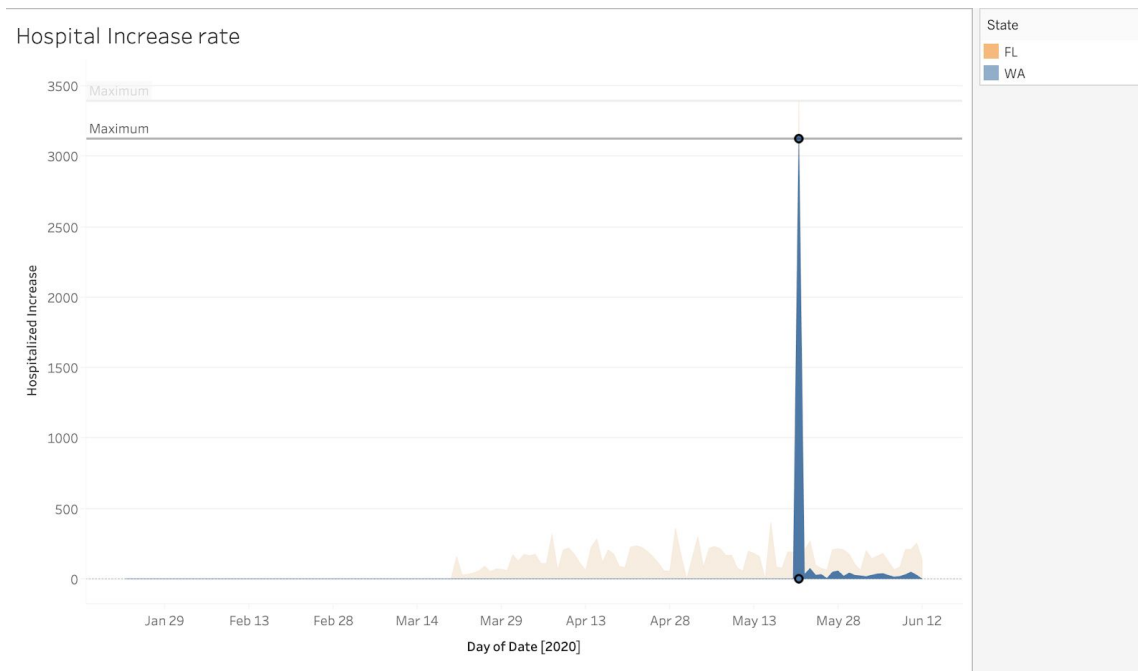
1. Hospital Admissions Rate
2. Death Comparison
3. Testing

Before going into the observations we have also noticed that although Washington started out having the cases, Florida overtook the Washington state rapidly post april where our tweet analysis suggests that most of the Floridians tweets the most during the month of april.

The case we are establishing is that the tweets have had the correlation between the cases being spiked up.

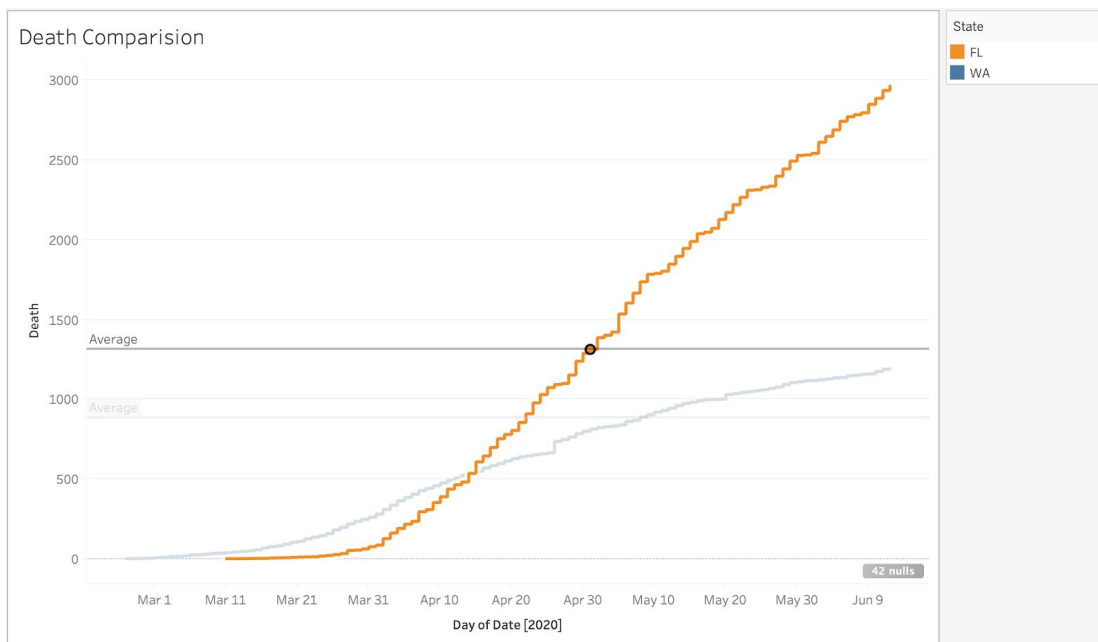
The following results support our analysis.

- 1) Hospital Admissions Rate



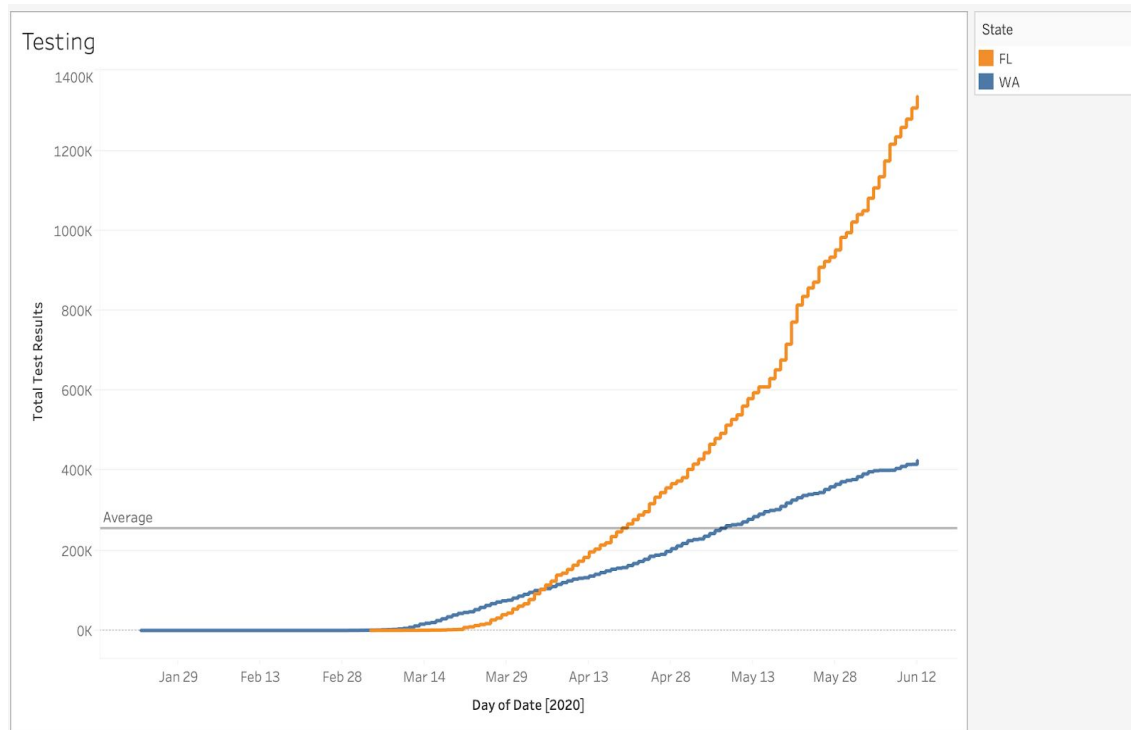
From the above visualization we can see that there was a spike for both the states regarding the hospital admits, especially during May mid week. Where Washington did fairly well compared to Florida.

2) Death Comparison



Similar to the Hospital admits, **right** when the month ended, we noticed how the death rates have risen in the state of Florida where the state of Washington started to flatten the curve.

3) Testing



Given the fact from our observations that both of the “Hospital admissions” and “Deaths” were peaked from the end of April, we have also seen the testing of the population being increased during the beginning of April.

The github link for our project -

<https://github.com/VarunNair4/COVID-19-Hackathon>

The tableau Dashboard could be accessed below -

<https://public.tableau.com/profile/varun.nair3055#!/vizhome/covid-19Flo-RidaandWashingtonanalysis/Dashboard1?publish=yes>

Conclusion

Based on the data and our results we can see there is a correlation between the tweets and the cases. Although it isn't safe to say that it's a strong correlation, we can still conclude that it accurately captures the sentiment which reflected the public perception towards COVID-19. This influences their attitude towards COVID-19 policies which involve following social distancing and staying at home which if not followed leads to greater spread of infection. Thus, this polar opposite perception of residents of the two states can be caught early to design effective, target informational campaigns to tackle both misinformation and COVID-19 pandemic.

