

# Trumps Tweets Concern Model

Guy Frankovits

19 6 2021

## Abstract:

Donald Trump's presence on social media attracted attention worldwide since he joined Twitter in 2009, with the handle @realDonaldTrump, having over 88.9 million followers by 2021. He frequently tweeted during the 2016 election campaign and as president, until his ban in the final days of his term. Over twelve years (from the creation of his account in May 2009 until his permanent ban in January 2021), Trump tweeted around 57,000 times, including more than 25,000 times during his presidency. A spokesman for Trump said that Trump's tweets were considered "official statements made by the President of the United States." In 2020, Twitter began hiding or adding fact-check labels to any of Trump's tweets that spread misinformation about the COVID-19 pandemic, or falsely suggested that postal voting or electoral fraud may compromise the presidential election.

## Research question:

In 2020, Twitter began hiding or adding fact-check labels to any of Trump's tweets that spread misinformation about the COVID-19 pandemic. Trump often posted controversial and false statements on Twitter. Trump's tweet undermined public health messaging and encouraged followers to disregard recommendations to prevent the spread of COVID-19. Scientific, medical, public health, and ethical experts, pandemic survivors, and the families those killed by COVID-19 expressed horror and dismay at Trump's attempt to downplay the COVID-19 pandemic in the United States, which at the time of Trump's tweet had killed at least 210,000 Americans.

Due to those facts, our research question is:

- Does Trump's tweets affect on the concern of republicans and democrats about COVID-19 pandemic?

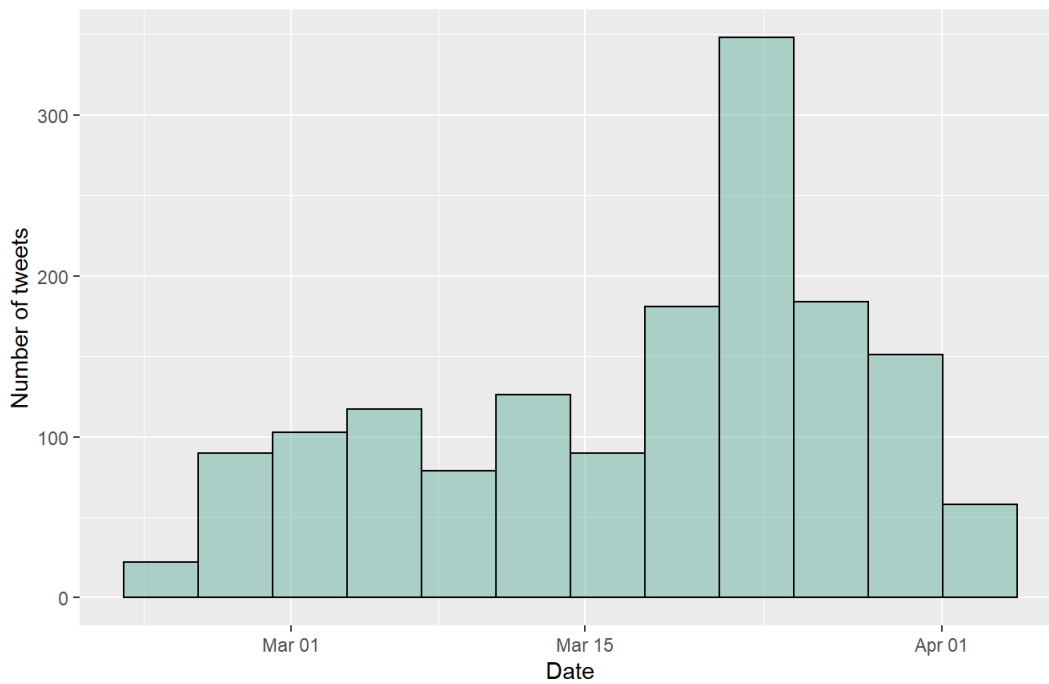
## Trump tweets data set:

Table 1: Descriptive trump tweets data set

| Variable  | Definition  | Example   |
|-----------|---|---|
| Date      | The dates in the data   | 2020-03-20 : 2020-04-04   |
| isRetweet | False means trump wrote the tweet, True means trump retweet someone else. | True or False   |
| retweets  | Number of retweets of the tweet   | Min = 30, Median = 10303, Max = 108116  |
| text      | The content of the tweet  | I was criticized by the Democrats when I closed the Country down to China many weeks... |

## Exploring the data:

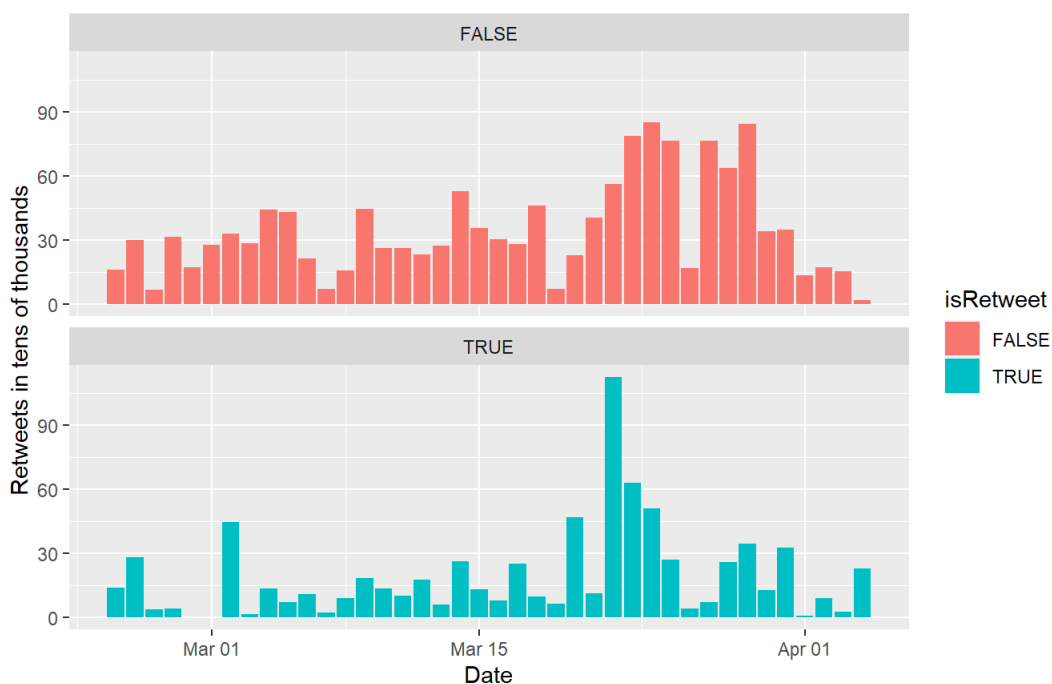
Amount of Trump's tweets  
between 2020-03-20 - 2020-04-04



Explain-

We can see that the peak of the amount of tweets is in the end of March, when the virus start to spread faster and killed people.

Amounts of Trump's tweets by isRetweet feature  
between 2020-03-20 to 2020-04-04



Explain-

We can see a sharp increase in the number of times Trump retweeted in late March, when the virus start to spread faster and killed people.

## Data analysis:

In order to work with accurate information, we decided to find Corona-related tweets. In order to decide which tweet is related to Corona and which is not, we decided to look for the most common words, and from them to choose words related to Covid19. The words we searched for were one-gram, two-gram and three-gram words. We did this twice - after the first time we saved the tweets in which we found

all these words, and performed a repeat search on them - meaning we searched again for words only this time on tweets that are certainly related to Corona.

The explanatory variables:

1 - Log on the number of times has been done retweet per day - Log(retweets)

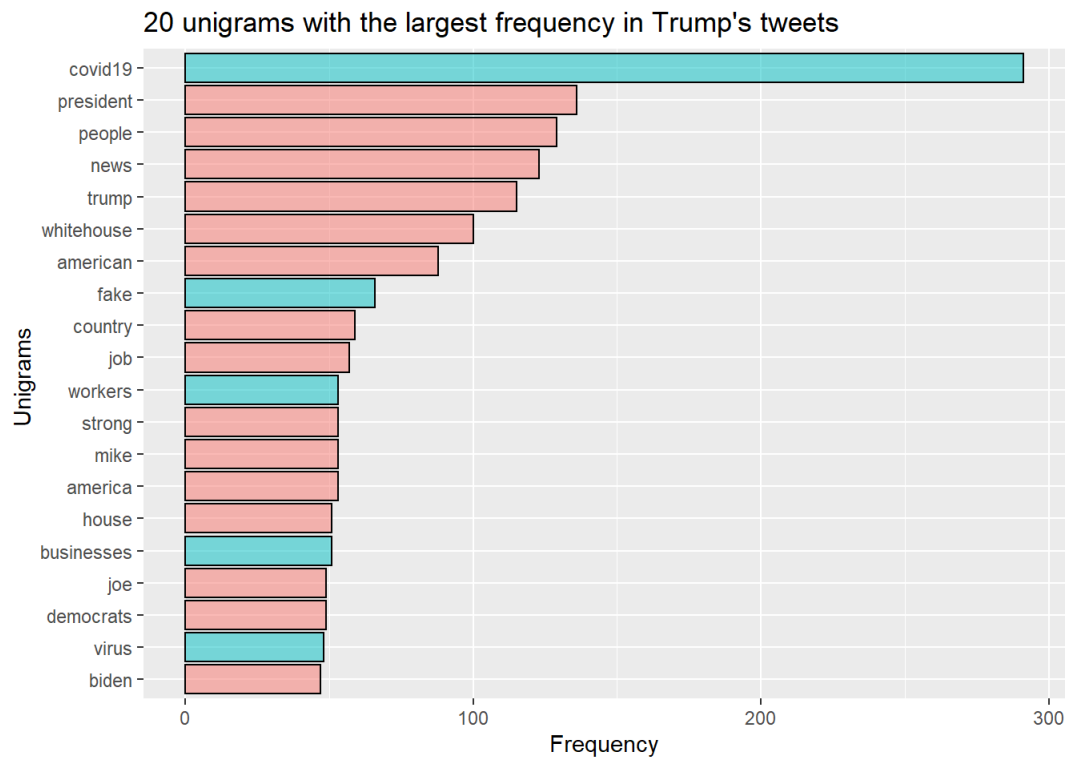
2 - Number of tweets per day

Linear regression model results for each column in the concern data set without filter tweets

| Politic_view | pearson   |
|--------------|-----------|
| Demokrats    | 0.7951256 |
| Republicans  | 0.8406380 |
| Differance   | 0.8565861 |

Note:  
By 7 day moving average

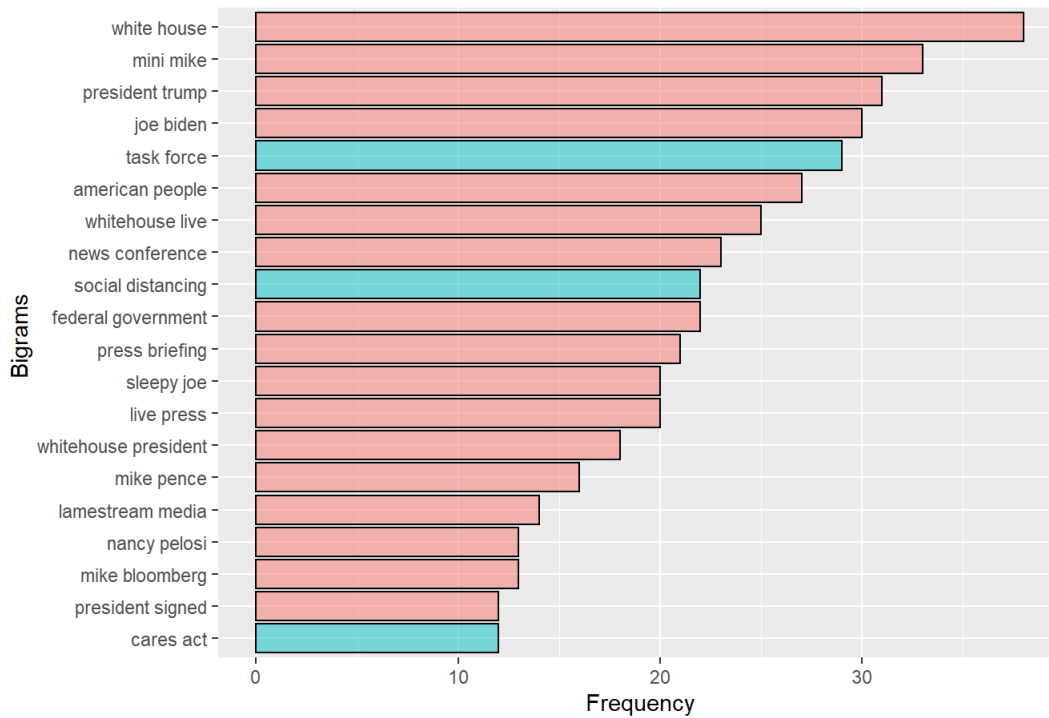
# Step 1 - find words related to Covid19 and save the tweets contains those words



Explain-

The blue words represent unigrams words we decided related that to covid19.

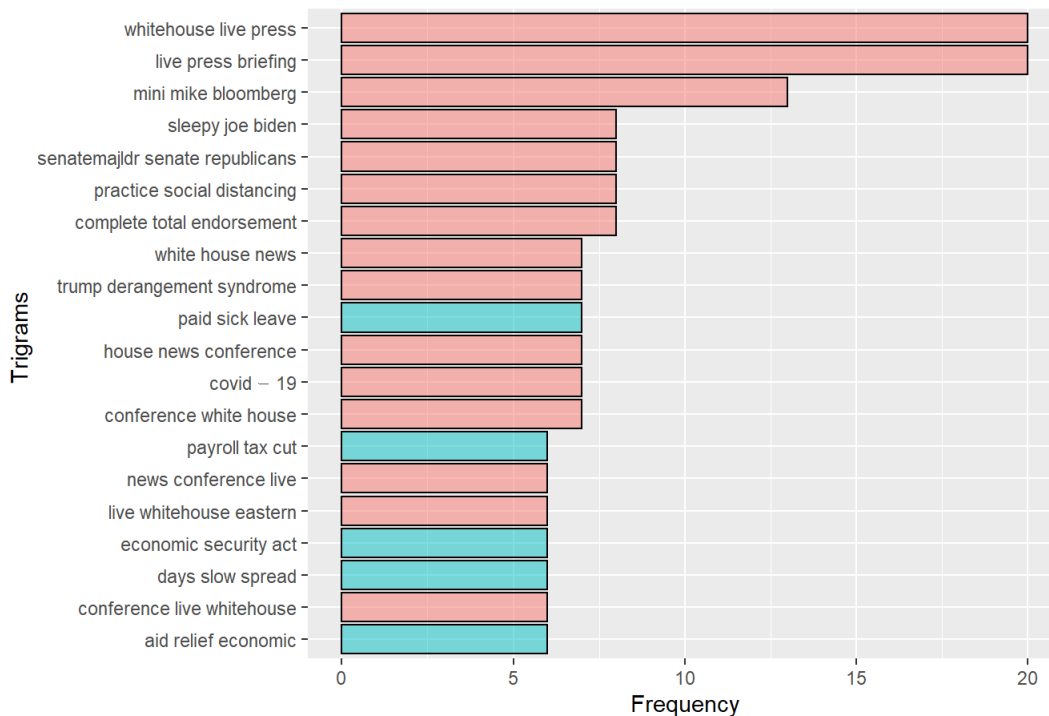
20 Bigrams with the largest frequency in Trump's tweets



Explain-

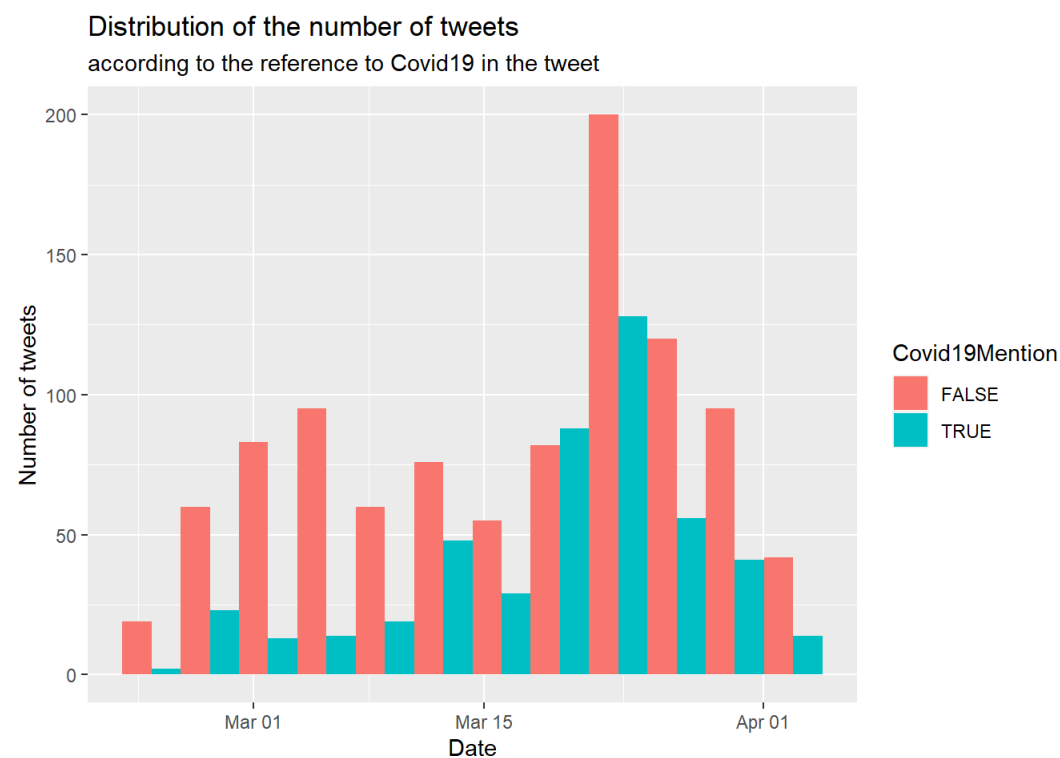
The blue words represent bigrams words we decided related that to covid19.

20 trigrams with the largest frequency in Trump's tweets



Explain-

The blue words represent trigrams words we decided related that to covid19.



Explain-

A different trend can be seen in the number of tweets associated with covid19 and those unrelated after the first filter.

Linear regression model results after the first filter of tweets

| Politic_view | pearson   |
|--------------|-----------|
| Demokrats    | 0.8684099 |
| Republicans  | 0.8809145 |
| Differance   | 0.8604277 |

*Note:*

By 7 day moving average

**Step 2 - From covid19 found tweets look for new common words that are also related to covid19**

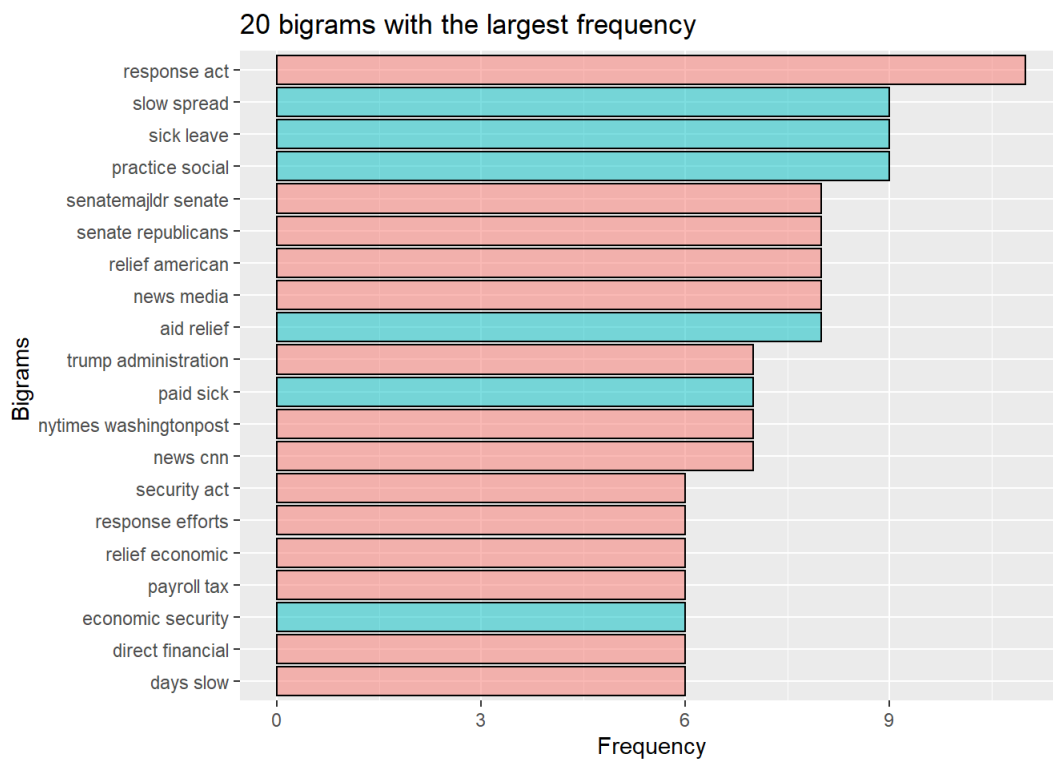
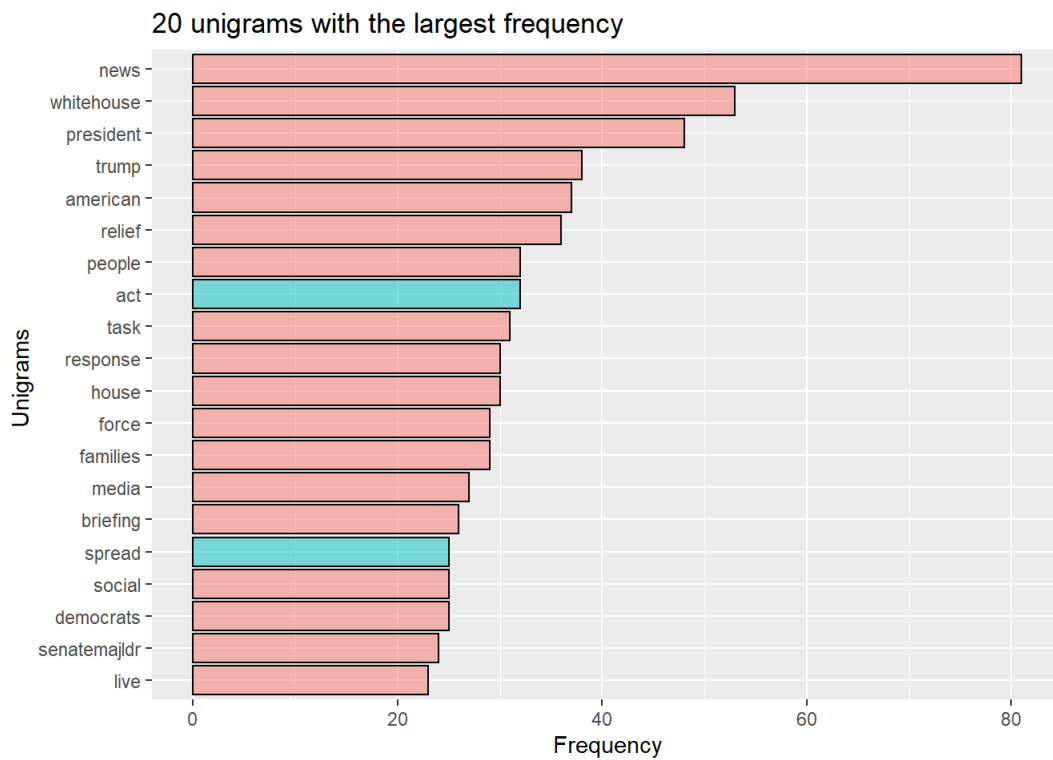


Table 2: All words that has relation to COVID-19 in Trump's tweets.

**words**

act

spread

social distancing

task force

cares act

covid19

## words

fake

workers

businesses

virus

social distancing

task force

cares act

days slow spread

aid relief economic

paid sick leave

economic security act

relief economic security

payroll tax cut

bill

cdcgov

lamestream media

healthcare

h1n1

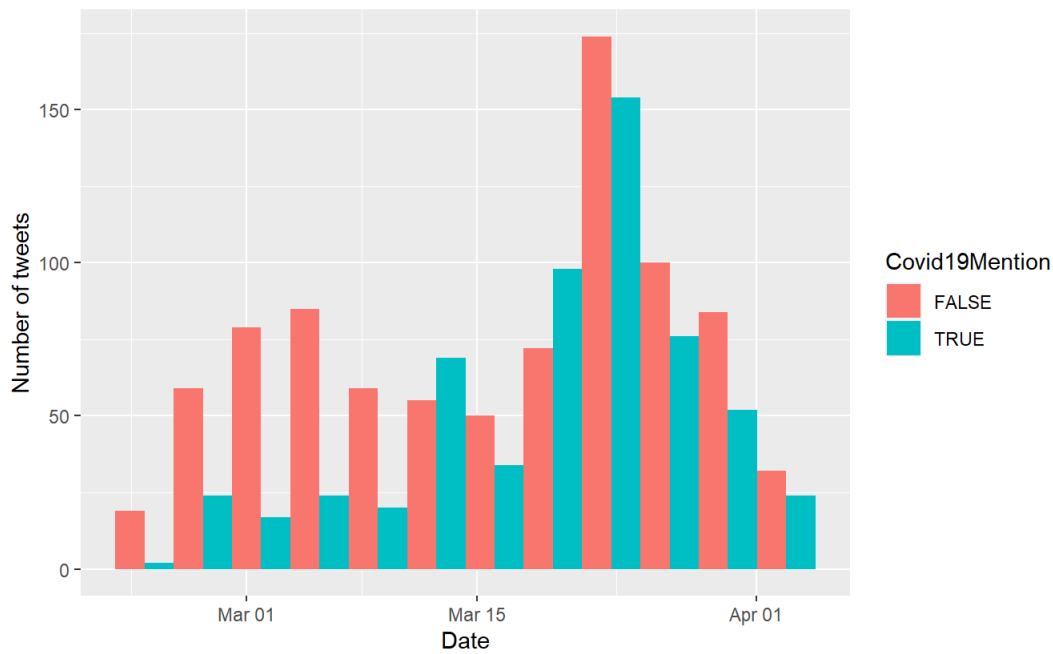
china

epidemic

emergency

free testing

Distribution of the number of tweets  
according to the reference to Covid19 in the tweet



We can see the increase of number of tweets the relation to COVID-19

Explain-

We can see the increase in the number of tweets related to Covid19 relative to the first step.

Linear regression model results after the second filter of tweets

| Politic_view | pearson   |
|--------------|-----------|
| Demokrats    | 0.9105394 |
| Republicans  | 0.9215314 |
| Difference   | 0.8960374 |

*Note:*

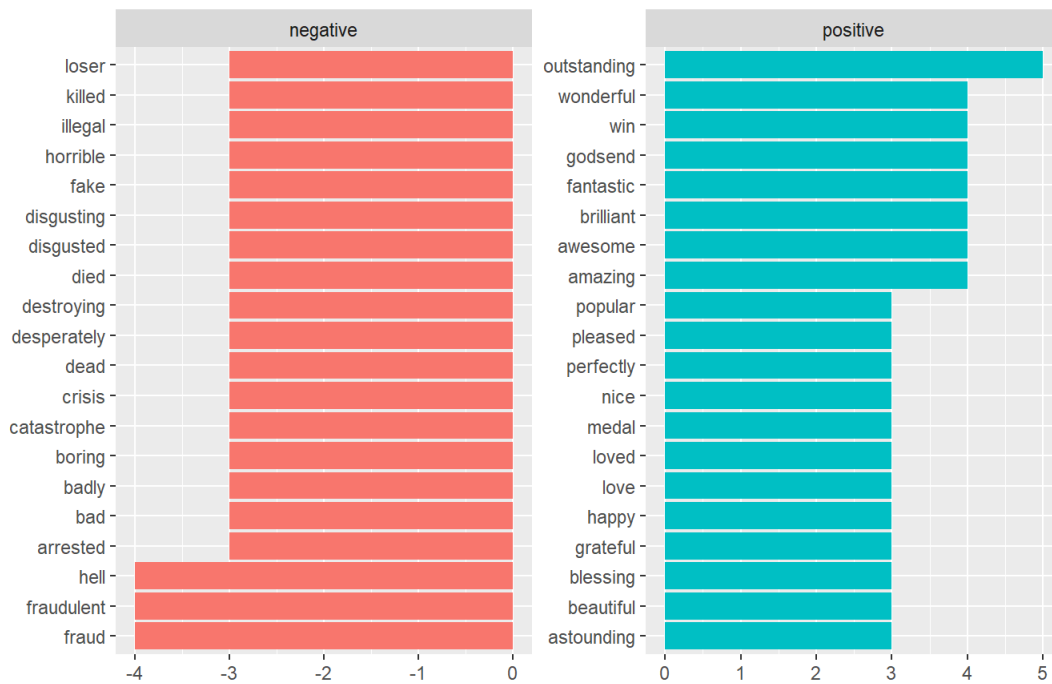
By 7 day moving average

## Step 3 - Sentiment analysis:



## Sentiment of words in Trump's filter tweets

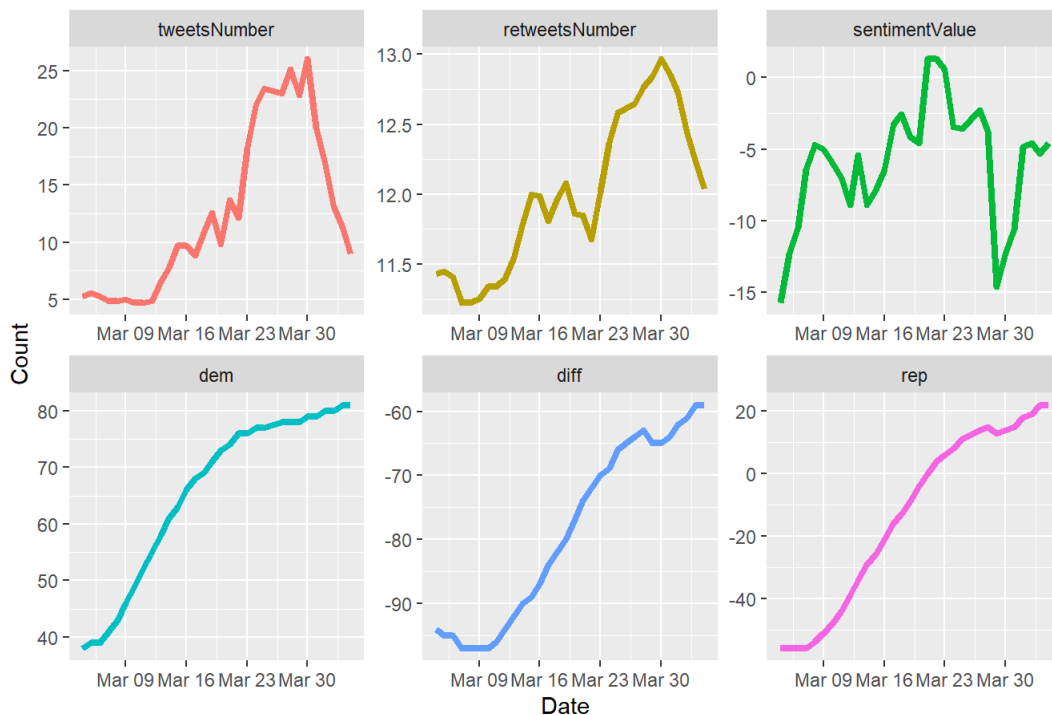
Bing lexicon



Explain-

Each word has a value between -5 and 5 in Trump's tweets, so we will sum up for each tweet his word values, and the same schema action we will perform for all the tweets that day, and we will add this as another feature. \*\*NOTE - We will change word value(\*-1) that have been linked as words with the opposite meaning as: positive(negative in our case), healthy(negative meaning)\*\*

## Compare the trends of concern and the 3 features



Explain-

It can be seen that there is a relationship in the behavior of the three features and each of the variables of concern.

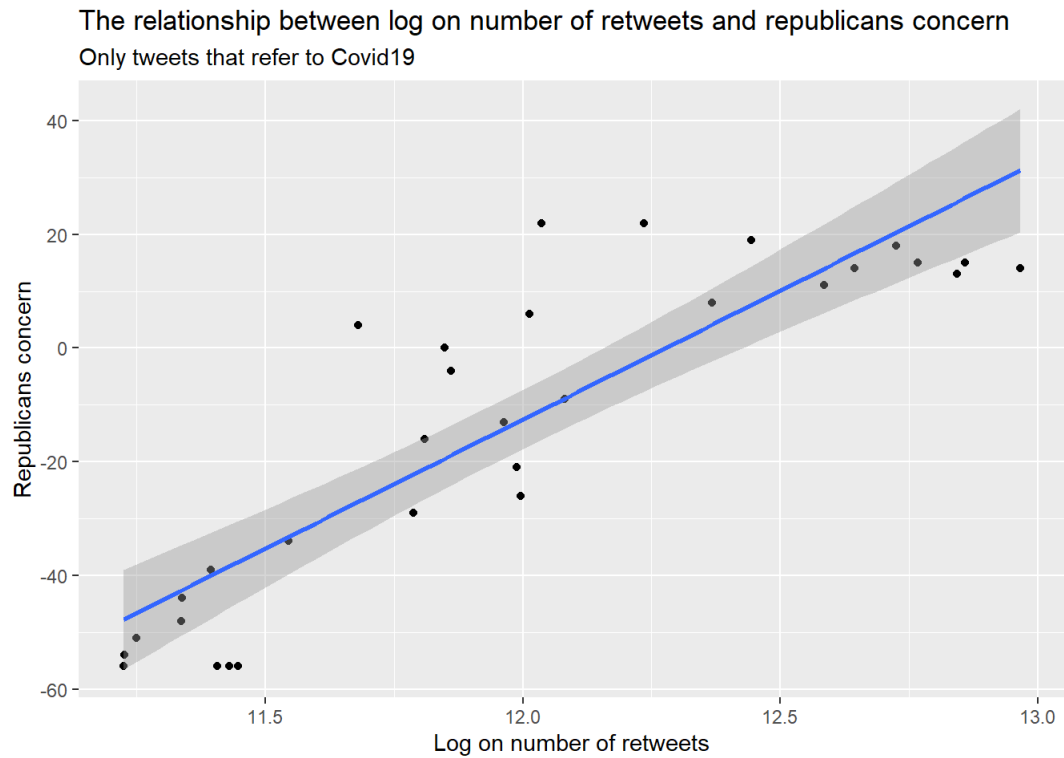
## Findings

Table 3 shows that the tweets features are highly correlated with the concern of the democrats and the republicans. The linear regression model is estimated with moving average of 7 days of features like number of Donald Trump's tweets that related to the virus, the log number of the number of time people retweets Trump tweets, and the by the positive/negative score of Trump tweets according to afin method of score.

Table 3: Linear regression model test results for each column in the concern data set

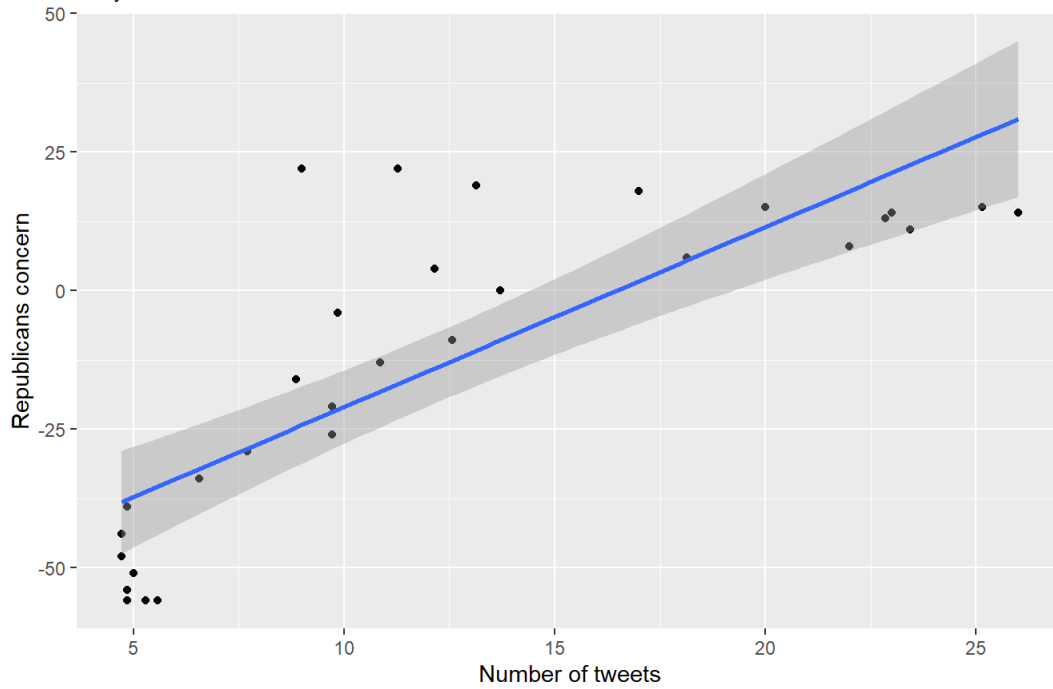
| Politic_view | correlationCoefficint | rmse     | std_err   |
|--------------|-----------------------|----------|-----------|
| Demokrats    | 0.9454334             | 5.564980 | 0.9682779 |
| Republicans  | 0.9463457             | 9.947749 | 2.3438785 |
| Differance   | 0.9260291             | 6.121527 | 1.2696246 |

Note:  
By 5-cross valudation



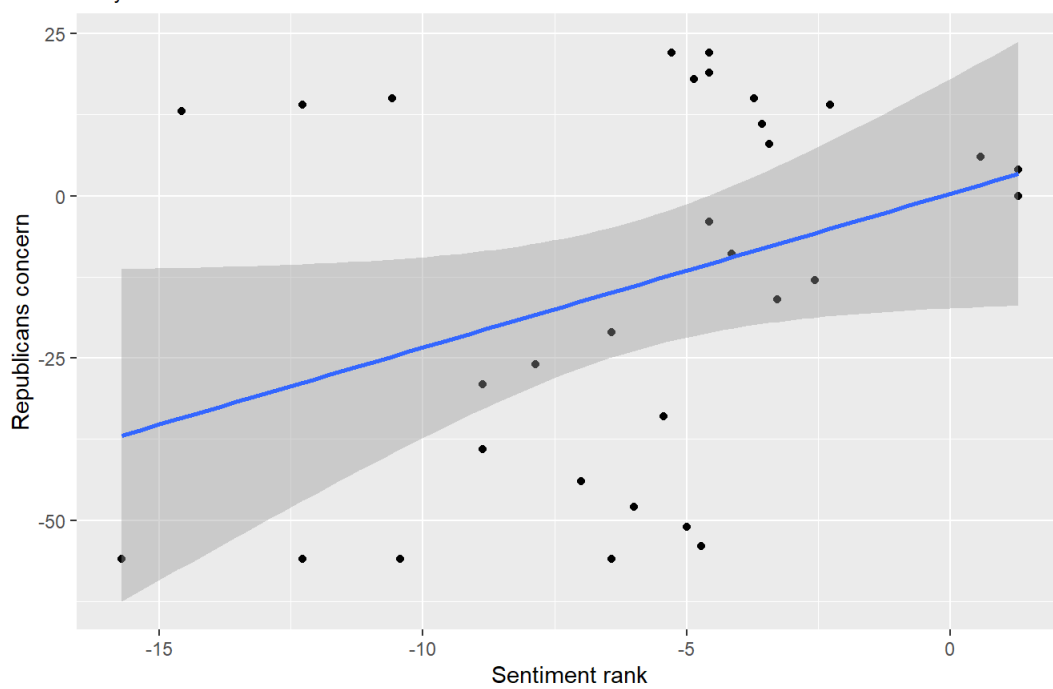
### The relationship between number of tweets and republicans concern

Only tweets that refer to Covid19



### The relationship between sentiment rank and republicans concern

Only from tweets that refer to Covid19



Result1 - Results of estimating regression model. Dependent variable is republican concern rate.

| Variable             | Coefficient Estimate | p-value |
|----------------------|----------------------|---------|
| Intercept            | -715.621190          | < 0.05  |
| Tweets Number        | -1.348580            | 0.052   |
| Log(Retweets number) | 61.332835            | < 0.05  |
| Sentiment value      | 2.673695             | < 0.05  |

Result2 - Results of estimating regression model. Dependent variable is demokrats concern rate.

| Variable             | Coefficient Estimate | p-value |
|----------------------|----------------------|---------|
| Intercept            | -312.9671596         | < 0.05  |
| Tweets Number        | -0.8857699           | < 0.05  |
| Log(Retweets number) | 33.3363153           | < 0.05  |
| Sentiment value      | 1.6620400            | < 0.05  |

Result3 - Results of estimating regression model. Dependent variable is difference between political view concern.

| Variable             | Coefficient Estimate | p-value |
|----------------------|----------------------|---------|
| Intercept            | 402.6540300          | < 0.05  |
| Tweets Number        | 0.4628106            | 0.266   |
| Log(Retweets number) | -27.9965193          | < 0.05  |
| Sentiment value      | -1.0116550           | < 0.05  |

An explanation of the three models-

It can be seen in the three models - when each model is represented by a different response variable from the concern table, that the variable of log on the number of retweets on the same day has the strongest effect on the model. It can also be seen that the p-value of each of the variables is almost always less than 0.05 meaning that there is a relationship between them and the effect on the response variable.