Rhea Mane

12/2/2022

Research Outline

How Covid Death Statistics become Politicized on Social Media

Background

The COVID-19 pandemic gripped the world in an unprecedented wave. People were thrown into a lockdown almost overnight without enough time to comprehend what exactly was going on in the world. As healthcare providers rushed to keep up with the inflow of patients and the people affected, the number of Covid deaths significantly rose. According to the CDC guidelines, a death which had some role of covid-19 as its cause was to be considered a Covid 19 fatality. With a rise in the number of deaths and a variety of opinions about the do's and don'ts for the pandemic from local authorities, the number of deaths soon became a debate about politics. According to NPR, "People living in counties that went 60% or higher for Trump in November 2020 had 2.73 times the death rates of those that went for Biden. Counties with an even higher share of the vote for Trump saw higher COVID-19 mortality rates."(Brumfiel, Wood) Politicizing deaths, and the death count led to opinions like the death count being exaggerated, and misinformation being spread on social media platforms like Twitter, which became forums for people to come and discuss these issues. The public felt like they were kept in the dark as a lot of the statistics were misrepresented or had errors and were straight up false. A general feeling of being cheated and lied to from the governing bodies rose in the public and hence they started blaming the lawmakers for the statistical errors. All conversations about the COVID death numbers became infected with political outbursts. My aim in this project is to try to understand how public opinions about the covid death count become politicized by using social media like Twitter.

Hypothesis and Research Question

I hypothesize that `:

1.1- How the public reacts to the death count due to Covid-19 on Twitter and whether they believe that the numbers are misrepresented or not ?

1.2- How most tweets about the covid death counts become politicized and people deem the numbers to be fake because of certain political affiliations ?

Methodology and Data

For this project, I mainly had to conduct a lot of manual data collection, and web scraping. I did initially try and use the Twitter API and the python library Tweepy to look up tweets, but because of limited access, I could not find tweets older than a week which were not the most relevant to my research question as I wanted to focus on discussions when death rates due to covid were very high, around the peaks of the waves.

With the API, I used the 'search-recent-tweets' function with the following parameters:

```
client.search_recent_tweets("(death #Covid19) lang:en -is:retweet",
max_results=100)
```

I kept my query as broad as possible for the maximum amount of tweets possible, but would end up with a very sparse dataset with an average of 2-5 tweets.

I then moved onto manual data collection. I used the advanced search feature on the Twitter website which gave me access to older tweets.

I divided my research timeline into 3 waves:

      July 7,2020- August 13, 2020 (38 days) (First wave)
      December 22nd, 2020 - February 13, 2021 (54 days) (Delta Variant)
      December 20,2021 - February 11,2022 (54 days) (O-micron)

Using these timeframes, I divulged into the tweets manually with other filters which included the following queries:

Covid deaths -UK -Sweden -Canada min_faves:20 lang:en until:2020-08-11 since:2020-07-07
Covid death -UK -England -Victoria -Africa -Sweden -Indonesia -India min_replies:10 min_faves:50 lang:en until:2020-08-13 since:2020-07-07

I tried to constrain myself to a specific timeline, (here) July 7th,2020- August 13th, 2020, and tried to focus on tweets about the US by eliminating mentions of other countries. I set a minimum of 20- 50 likes on the search filter so that the tweets I was going over had a basic level of engagement, and because I was looking into conversations focussed about the topic, I also set up a minimum of 10 replies for the tweets so that I could evaluate what the public was talking about.

While manually searching for relevant tweets which would contribute to my dataset, here are a few general keywords I looked for. #COVIDdeaths, #CovidIsNotOver, #CovidRelatedDeaths, #ExcessCovidDeaths, #COVID19, #deaths, #misinformation, #fake news

For tweets with some political opinion I looked for these keywords in addition to the ones mentioned above: Doug Ducey, Republican, republican politicians, governor, health officials, White house, Biden Administration, CDC, Democrats, government, Leftists, Election

For tweets without a politicized angle, which were only talking about the statistics, here are some keywords which helped me identify significant pieces of information: False numbers, fake numbers, statistics, death rate, hiding numbers

[Here](#) is the link to my final dataset.

Results :

Average % of political replies in a thread:
1st Timeline: 19.2%
2nd Timeline: 10.33%
3rd Timeline: 12.21%

Discussion :

- I expected the percentages to steadily decline through the timelines. Which means that the public became less political as the pandemic moved through its variants and as people got used to it.
- That is not what I saw, I saw an increase in the number of politicized tweets for the 3rd time after an expected decline in the second wave.

Some possible facts to consider while explaining the results:
- On January 9th, 2021 (midway in the second timeline), Donald Trump's Twitter account gets suspended, hence, a lot of the conversations which were initiated by him did disappear.
- The second timeline is also when the US government changed power.
- The third timeline covers a time when Joe Biden has been the POTUS for a year, which might have led to some inflammatory conversations.

Conclusion:

This project is still underway and this is what I plan to achieve going forward:

- I think a big accomplishment would be using the API and generalizing my research over a larger and more robust dataset.
- I am interested to see if the results will substantially change, and will the trend I saw in the politicization of tweets change.
- I would also like to include more parameters in my data collection and make it more streamlined and consistent.

Works Cited:

Brumfield Geoff, Wood Daniel 'Pro-Trump counties now have far higher COVID death rates. Misinformation is to blame', *NPR,* December 5, 2021. https://www.npr.org/sections/health-shots/2021/12/05/1059828993/data-vaccine-misinformation-trump-counties-covid-death-rate

Ding Jason, Kananack Claude, "The Effects of Blame and Politicization on Responses to Pandemics", *Journal of Student Research,* 2021. https://www.jsr.org/hs/index.php/path/article/view/1320

Gisondi Michael, "A Deadly Infodemic: Social Media and the Power of COVID-19 Misinformation", *Journal of Medical Internet Research,* https://www.jmir.org/2022/2/e35552/authors

Bagherpour Amir, Nouri Ali, "COVID Misinformation Is Killing People:This "infodemic" has to stop", *Scientific American.* October 11, 2020, https://www.scientificamerican.com/article/covid-misinformation-is-killing-people1/

Mitropoulos Arielle, "For red and blue America, a glaring divide in COVID-19 death rates persists 2 years later" *ABC News,* March 28, 2022, https://abcnews.go.com/Health/red-blue-america-glaring-divide-covid-19-death/story?id=83649085