

NLP PROJECT

CORONAVIRUS TWEETS: APRIL 2020

Sandra Paredes

INTRODUCTION

- ▶ **Motivation:** Twitter was at the center of information sharing and conversations early on in the COVID pandemic, including misinformation. [1, 2, 3]
- ▶ **Research Question:** What were Americans tweeting about coronavirus and COVID-19 in April 2020?
- ▶ **Impact Hypothesis:** Inform CDC's communication strategy for future pandemics.

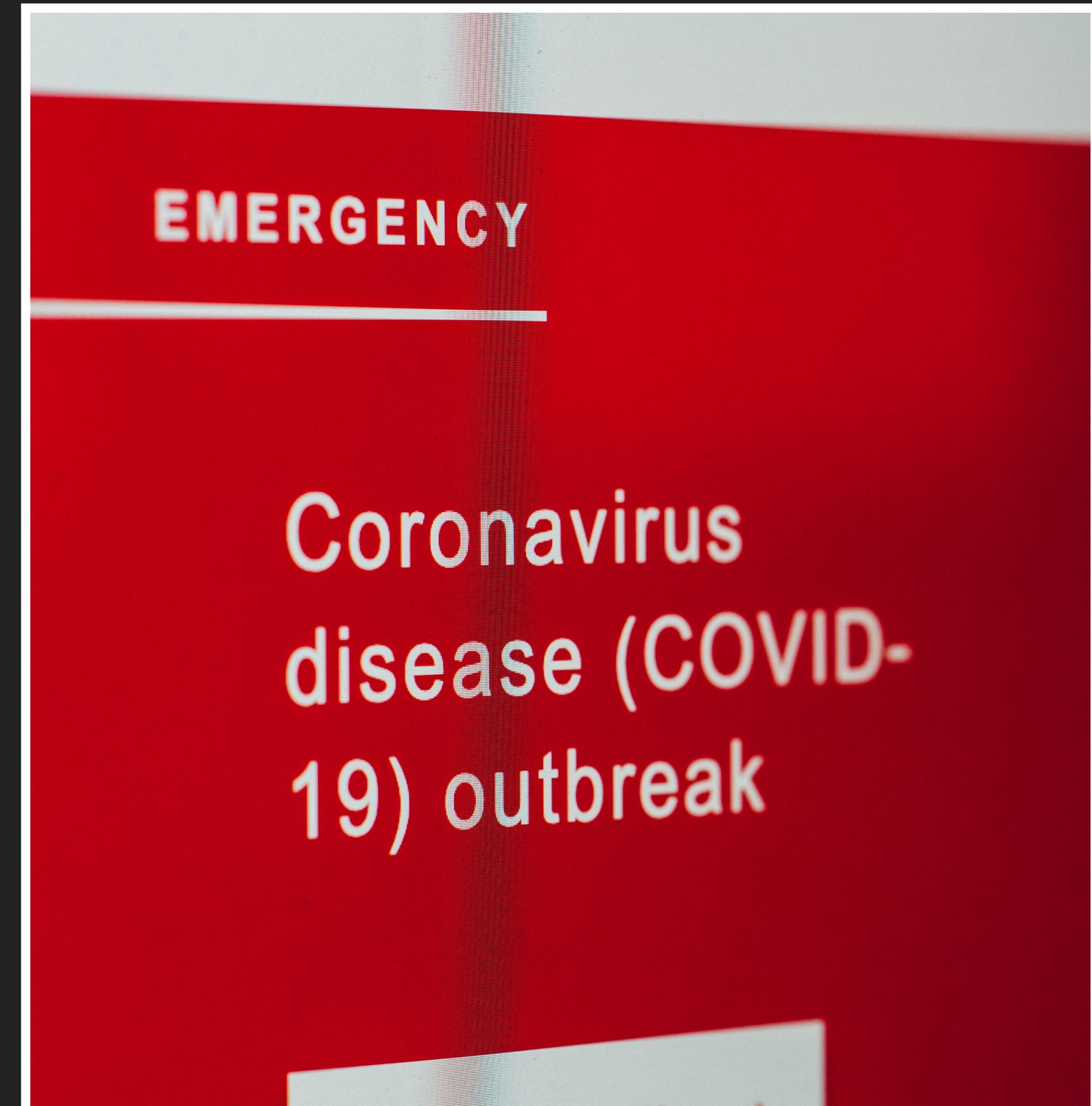


Photo by Markus Spiske on Unsplash

TWITTER DATA

METHODOLOGY

- ▶ **Dataset**
 - ▶ Kaggle:
 - ▶ **Coronavirus Tweets** [2, 3]
 - ▶ April 2020
 - ▶ **Tweets = 138,796**
 - ▶ Twitter handles = 54,279
 - ▶ Preprocessing:
 - ▶ Remove punctuation
 - ▶ Stopwords
 - ▶

The image displays four separate Twitter posts arranged in a grid-like fashion:

- Post 1:** A tweet from **uche blackstock, md** (@uche_blackstock) dated 8:00 PM · Apr 5, 2020 from Brooklyn, NY. It discusses the high death rate in New York City during the pandemic. The post includes a photo of a person walking on a city street at night.
- Post 2:** A tweet from **Redding Rancheria** (@RRancheria) dated 1:27 AM · Apr 16, 2020 from Anderson, CA. It mentions the **#LightItBlue** campaign and the **#COVID19** outbreak, showing a building illuminated with blue lights.
- Post 3:** A tweet from **Annie M. Dance** (@AnnieMDance) dated 7:28 PM · Apr 23, 2020 from Super Stop & Shop. It asks about face mask ordinances in Danbury and shows a sign on a store door stating "By Order Of The City Of Danbury All Customers Must Wear A Face Mask Or Face Covering To Enter This Building. Thank You".

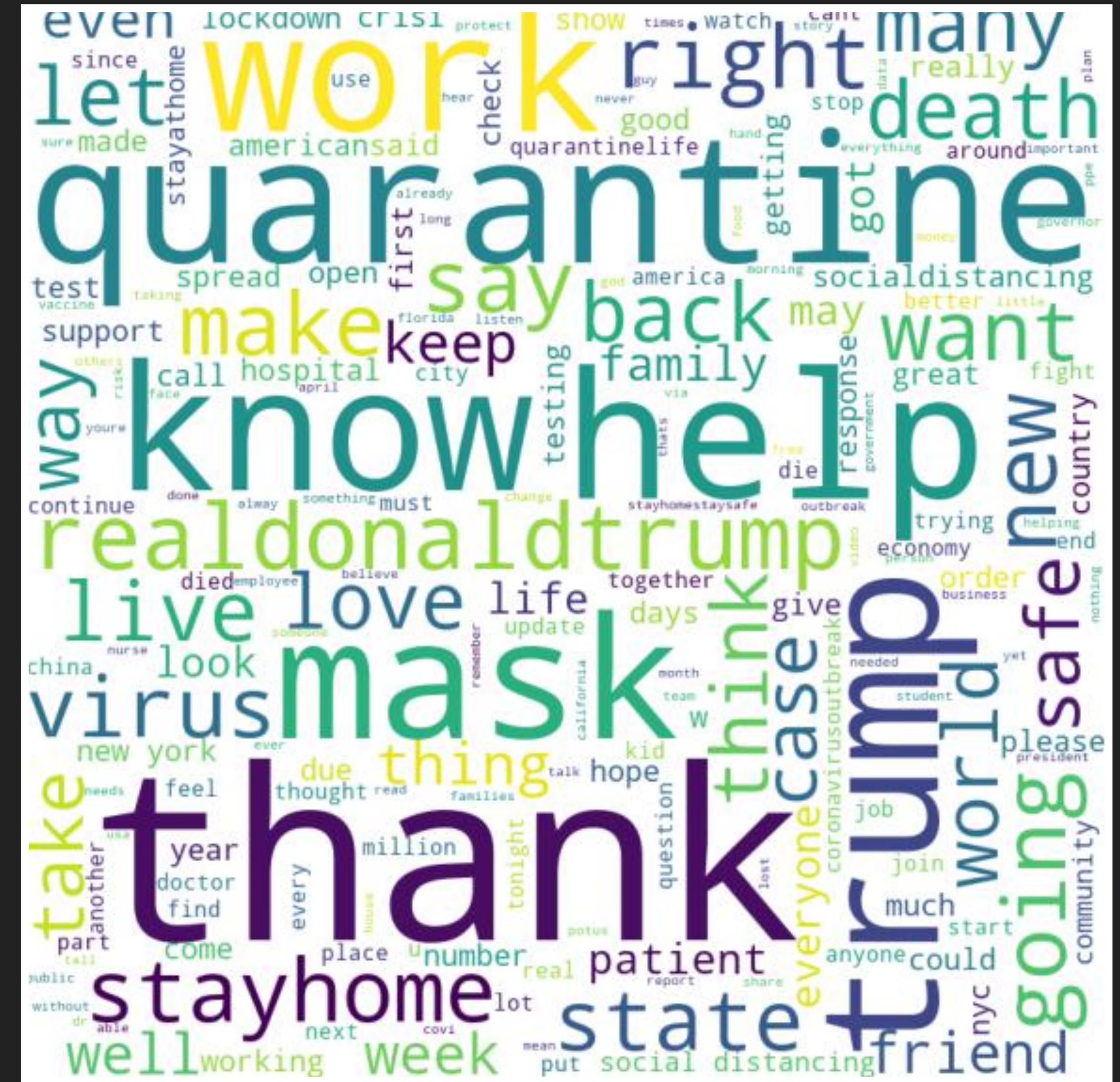
TWITTER DATA

RESULTS

Topics

- 1. Trump and grab bag COVID terms
 - ▶ trump, realdonaldtrump, help, thank, know, going, work, please, virus, health
 - 2. Stay home
 - ▶ quarantine, stayhome, quarantinelife, socialdistancing, stayathome, staysafe, lockdown, california, love, stayhomestaysafe
 - ▶ New York City crisis
 - ▶ new, york, cases, deaths, city, state, nyc, county, positive, april

Figure 1. Word Cloud



RESULTS

Sentiment

- ▶ NaiveBayes to predict sentiment and compared to Vader Sentiment compound score mapped to label
- ▶ Sentiment prediction mean accuracy score:
 - ▶ Gaussian = 0.598
 - ▶ Multinomial = 0.7345

Figure 2. Vader Sentiment

		text	compound_score	sentiment
0		sundayfunday coronavirus style vino cheers	0.4767	positive
1	this pandemic has confirmed my worst fears most people dont know how to make entertaining videos sinceivebeenquarantined americasunniestvideos wrestlemania tonyaharding		-0.6124	negative
2		is this true \n \n ecuadorenemergencia coronaviruspandemic	0.4215	positive
3	many us thought it was wuhan province but i uld never be us then it was italy but i uld never be us now it is here one newyorker died every minutes from over this weekend absolutely devastating \n\n		-0.9230	negative
4		ah coronavirus humor	0.2732	positive
5	miami and south florida in general are also staying home \n\nnorth florida thinks their immune to oipfvqymvc		0.2960	positive
6	how can president trump be flip about a question abou ntinuity of power and contracting the coronavirus when i am around him i dont breathe and we wonder why other americans question and defy the quarantine coronavirustaskforce		0.0000	positive
7	this is our most desperate hour help us obi wan kenobi you re our only hope starwars weneedaleader		0.5058	positive
8	face your fears \n business love coronavirus nyc ny la california zoom happy prayer		0.7269	positive
9	first defense nasal screen coming up on abcsharktank during the outbreak and all of a sudden the guy doesnt seem crazy \n\nsharks redeemed themselves w mega offers \n\npos rona airing mcuban thesharkdaymond kevinolearytv is gonna make a killing replaces mask need		-0.1441	negative
10		total tested day ny state \n	0.0000	positive
11	finally turned out dining room dog crate from hell room into an office took coronavirus and quarantinelife to do it but im glad i did i need a new desk but it it will do the job acceptancehq		0.2960	positive
12	the many takes on restaurant takeout coronavirus has curbed our orders but our enthusiasm qaozqvchuz		0.5927	positive
13	fox news along with cnn and msnbc parents companies have been sued over their coverage of coronavirus by an ethics organization cddrtqipxk		0.0000	positive
14	is zoonotic bronxzoo confirms numerous cases of coronavirus in tigers and lions i thought life couldnt get any more bizarre after tigerking and then this natgeo piece \n		-0.3804	negative

CONCLUSIONS

Insights

- ▶ CDC was not among the top 10 terms in the 3 topics.
- ▶ Tweets from CDC do not appear in the corpus, nor any mention of the agency
- ▶ However, the president at the time was in the top 2 terms in the first topic.

Recommendations

- ▶ Join Twitter conversation early.
- ▶ Designate hashtag to uniformly use across U.S. health agencies (i.e., FDA, NIH, CMS).



Photo by Sandie Clarke on Unsplash

FUTURE WORK

- ▶ Analyze tweets that mention @CDCgov and their sentiment.
- ▶ Cluster analysis to visualize topics.
- ▶ Time series analysis of topics to see how the conversation on Twitter changed during April 2020.
- ▶ Expand corpus through December of 2020 to include mask mandates, vaccines, and variants.



Photo by Edwin Hooper on Unsplash

APPENDIX

- ▶ Summary, data, and slides are available at github.com/slp22/nlp-project



Photo by Sarah Kilian on Unsplash

APPENDIX: SOURCES

1. Twitter study tracks early days of COVID-19 pandemic in U.S.: <https://www.binghamton.edu/news/story/2415/twitter-study-tracks-early-days-of-covid-19-pandemic-in-u.s>
2. The vital role of Twitter in responding to Covid: <https://www.brookings.edu/blog/techtank/2022/01/04/the-vital-role-of-twitter-in-responding-to-covid/>
3. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7170811>
4. Kaggle: <https://www.kaggle.com/datasets/smld80/coronavirus-covid19-tweets-early-april>
5. Kaggle: <https://www.kaggle.com/datasets/smld80/coronavirus-covid19-tweets-late-april>
6. Name: link
7. Name: link
8. Name: link

APPENDIX: CDC COVID-19 TIMELINE

April 3, 2020

At a White House press briefing, CDC announces new mask wearing guidelines and recommends that all people wear a mask when outside of the home.

April 6, 2020

Hundreds of doctors and civil rights groups urge CDC and U.S. government to release race and ethnicity data on COVID-19 cases in order to ensure the proper response in black communities.

April 8, 2020

U.S. Health and Human Services announces first contract for ventilator production goes to General Motors under the Defense Production Act.

cdc.gov

March 31, 2020

At a White House Press Briefing, Dr. Anthony Fauci and Dr. Deborah Brix announce that 100,000 to 240,000 deaths in the U.S. are expected even if social distancing and public health measures are perfectly enacted.

April 3, 2020

CDC launches COVIDView, a weekly report that summarizes and interprets key indicators from a number of existing surveillance systems

April 7, 2020

A report from the Chicago Tribune notes that 68% of the COVID-19 related deaths in Chicago occur among the city's African American community, illuminating racial disparities of the pandemic in the U.S.

April 13, 2020

Most U.S. states report widespread cases of COVID-19.

April 26, 2020

Clinicians in the United Kingdom note increased reports of previously healthy children presenting with severe inflammatory syndrome with Kawasaki disease-like features. The cases occurred in children testing positive for current or recent infection with SARS-CoV-2. This condition would later be known as Multisystem inflammatory syndrome in children (MIS-C), an inflammatory condition that affects children with COVID-19.

April 30, 2020

President Trump launches Operation Warp Speed, an initiative to produce a vaccine for the coronavirus as quick as possible with CDC as an integral member.

April 10, 2020

The U.S. surpasses Italy as the global leader for reported deaths due to COVID-19 (23,036 deaths)

April 13, 2020

At a White House press briefing, President Trump announces that the U.S. will cease funding to the World Health Organization, shaking the public health community during the pandemic.

April 24, 2020

Georgia, Alaska, and Oklahoma begin to partially reopen their states despite concerns from health experts saying it was too early to reopen.