ShengYun Peng JunYan Mao GuanChen Meng ZeFang Liu YuXuan Wang HuiLi Huang

COVID-19 GLOBAL VISUALIZATION

INTRODUCTION AND MOTIVATION

Twitter, as a leading social platform, generates over 800 million Tweets in a single day. Tweets on COVID-19 are too overwhelming for users to extract useful information. Also, current COVID-related websites mainly provide 2-D COVID-19 visualizations. Main **contributions** and **novelties**:

- Demonstrate how COVID-19 spread across the world
- Manifest the correlation between trend in tweets and COVID-19 daily cases
- Present a word cloud of most frequently occurring tokenized words

By performing the above visualizations, the project goal wishes to:

- Evaluate the relationship between the COVID-19 cases and number of tweets in each country
- Provide data visualization to assist users make decisions regarding pandemic
- Give users an overview of the most popular topics about COVID-19 on Twitter

DATASETS

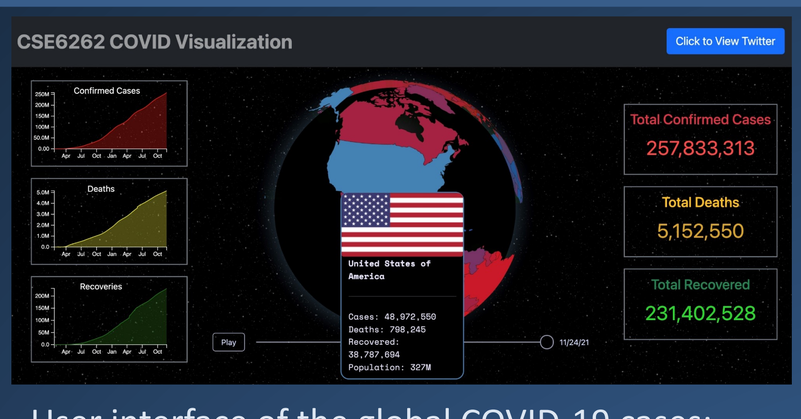
- Dataset-1: COVID-19 daily cases

- Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which has COVID-19 cases by day, month and year

- Dataset-2: COVID-19 Twitter chatter dataset

- Panacea Lab at GSU
- Consists of COVID-19-related tweets acquired from the Twitter Stream
- Provides top 1000 frequent terms, bigrams, and trigrams. Data used is March 1 to April 15, 2021 (to capture the second spike of COVID-19 cases in the U.S.)

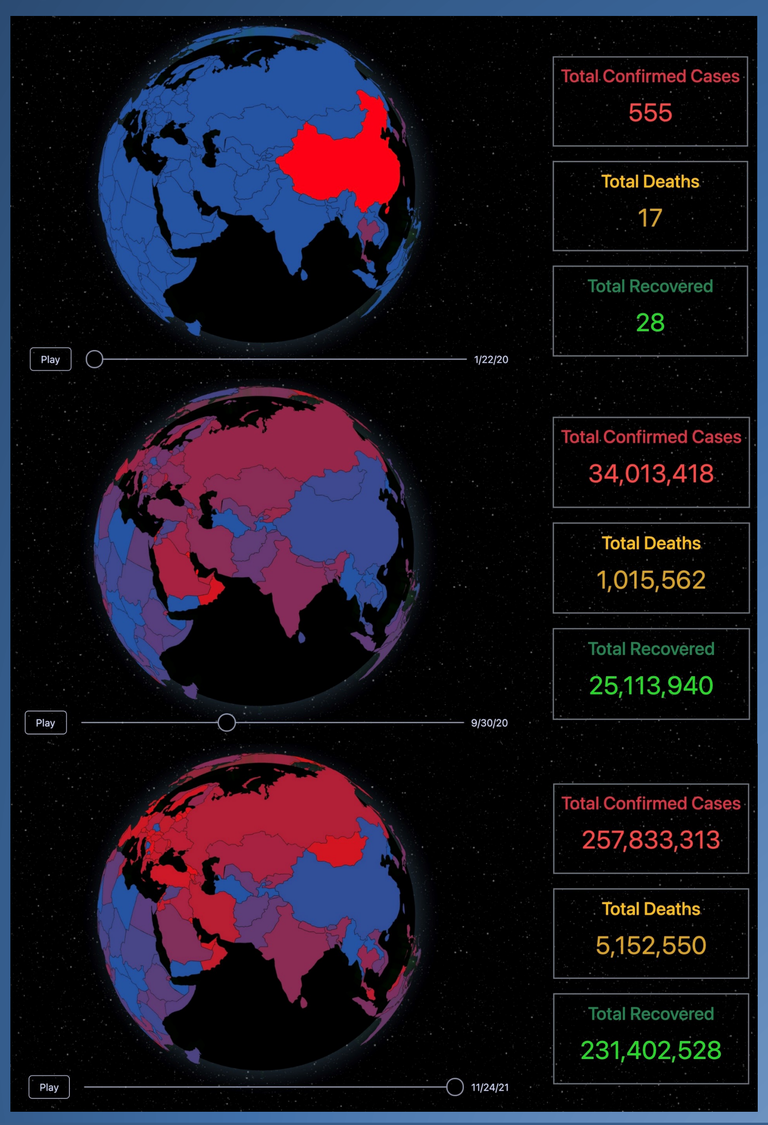
GLOBAL COVID-19 VISUALIZATION



User interface of the global COVID-19 cases:

1) line charts
2) total number
3) country-wise cases of the selected date
4) the navigation bar

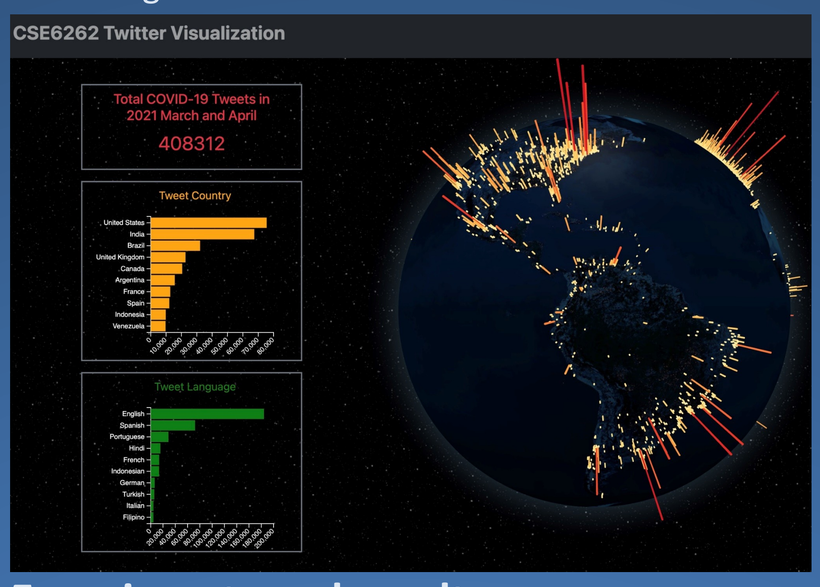
Dragging along the slide bar displays daily COVID-19 case data, and each country switches to corresponding color scheme. Meanwhile, the total cases are dynamically displayed on the right.



COVID-19 TWEET VISUALIZATION

Approaches:

- Select 6000 tweets with geographic coordinates from 408312 COVID-19 related tweets in March and April 2021
- Use world globe for visualizing tweet locations effectively
- Display higher and redder bars for more tweets in one position
- Visualize twitter data by geographic coordinates, which innovatively present tweet numbers among countries



Experiments and results:

- Evaluate legibility of the geographic data, density of coordinates, and distinctiveness of colors
- Display two bar charts for the most common tweet countries and languages
- World globe with proper longitude and latitude data giving the clearest result

CONCLUSION

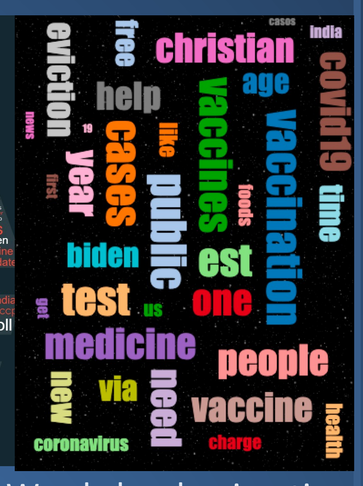
- The 3-D global COVID-19 visualizer provides accurate and necessary information. Interactive functions display the evolution of the pandemic.
- The number of COVID-19 tweets from one location has a positive relation with its number of COVID-19 cases visually.
- Word cloud reflects the most concerned areas, severity, and people's coping method towards COVID-19.

COVID-19 TWITTER WORD CLOUDS

Method:

- Words are selected from frequently discussed topics on Twitter
- Word sizes are positively correlated to the number of counts
- Non-English words are filtered out by the langdect library
- Words with repetitive meanings are manually filtered out. E.g., "covid19", "covid-19", and "coronavirus19"
- Dates and numbers are filtered out. E.g., "million", "000", "19"

COVID-19 second wave



Word cloud animation

Experiments and findings:

- Generate word cloud of the COVID-19 second wave.
- Generate animated word cloud to show the most popular COVID-19 topic in each month from Jan-Apr 2021.
- Prove the change of Twitter topics is related to COVID-19 situation.
- Reflect the correlation between the epidemic and Twitter data
- Find people's changing ways of coping COVID-19 pandemic, from "hard immunity" to "vaccination", from "social distance" to "boost shot"
- Find the meaning of locations appeared in the word cloud, usually the most infected area by COVID-19