COVID-19 Vaccination Visualization

Student Name: Chen SHEN ORCID: 0000-0002-2265-0253

Introduction

COVID-19 has been hitting the world with its high mortality risk and dramatic spreading speed since early 2020. According to the latest data from Wikipedia[1], a total of 154 million humans were infected and, 3 million were dead during this global pandemic.

My interest comes from the fact that the pandemic, generally speaking, has been well controlled in China, and the vaccination is now in the process.

However, how is vaccination going on around the world?
How many vaccines are popular these days?
Could I trace any vaccine-related information by investigating Twitter?

The structure of the story follows:

1) Comparison of death and vaccination density (at this moment)
   - ✧ Group by economic status or continents
2) Comparison of death and vaccination density (in time series)
   - ✧ Group by economic status
   - ✧ The country query is available
3) Supply and demand sides of the vaccines
   - ✧ Production: vaccines to countries, counted in million dose production
   - ✧ Use: vaccines to continents, counted in the number of the countries using
   - ✧ The country query is available
4) Twitter performance of Pfizer/BioNTech
   - ✧ The general trend in terms of positive and negative sentiment
   - ✧ Tweet query is available

Method

1. Data source

Death and vaccine data come from Our World in Data[2].
The detailed data of vaccines comes from airfinity[3].
The tweets data comes from Kaggle[4].

2. Methods

First, general approaches include cluster analysis, cross-sectional analysis, time trend analysis and sentiment analysis.

Second, statistical approaches include mean and linear interpolation. I found the problem of data missing is quite severe, and I used linear interpolation to replace the missed data under the assumption the data follows the linear form. However, some problems may still occur when making the country query.
Third, visualization approaches include maps, scatter plots, line charts, Sankey diagrams, etc.

3. Tools

Most data cleaning and processing is done by Python and Excel, and specifically, the VADER sentiment analysis[5] is used for the sentiment analysis, which is specifically attuned to sentiments expressed in social media.

Result

First, I started with the country-level analysis, comparing the latest death data and vaccination data, using the 'density number' instead of the 'total number'. I found generally:

- ✧ The discrepancy among the countries in terms of vaccination is large, for instance, Israel had vaccination 10 times higher than the world average rate, while some countries haven't started yet.
- ✧ Grouping countries by the status of the economy: advanced economies achieve a higher-level vaccination than the developing economies.
- ✧ Grouping countries by continents: European countries generally have high risk and high recovery speed, while African countries had especially little vaccination progress.

The extreme high vaccination happening in Israel makes it interesting to see the time trend data of Israel (but not only Israel).

Second, I switched the research focus from the cross-sectional comparison to the time trend development. I found:

- ✧ The best performance on vaccination achieved by Israel is not an immediate success, instead, it is a gradual process. In other words, Israel started early and continuously devoted to vaccination.
- ✧ The countries above the average-death level, mostly, faced a surge of increment in deaths during December 2020, which could be partially interpreted by the reunion in Christmas (pure guess and I skip this since death is not the focus of this project).

Third, comes the vaccine-level analysis. I found:

- ✧ From the supply side: China and US are the most productive countries, followed by Germany and India.
- ✧ From the demand side: different continents have different vaccines preferred (or the formal administration reasons). For example, excluding

the most popular ones (Pfizer/BioNTech and Oxford/AstraZeneca), European and North American countries use Moderna more, while Asian and African countries use Sinopharm more.

- ✧ From supply to the demand side: although Pfizer/BioNTech, Sinovac, Oxford/AstraZeneca and Moderna were high produced, Pfizer/BioNTech and Oxford/AstraZeneca were used most frequently, counted in several countries. (the difference caused by the different populations of countries covered by vaccines)

Fourth, to trace the text sentiment on Twitter (focused on Pfizer/BioNTech), I both aggregated the sentiment score and isolated the most influential tweet for each day. I found:

- ✧ Most positive sentiment tweets are about the first vaccination in the countries or the regular vaccination posts; negative sentiment tweets are highly likely to be something with the allergy or death caused by vaccination.
- ✧ There is a very strong negative sentiment wave happening on Jan 16th. On that day, 55 deaths, occurred among people receiving both the Moderna and Pfizer/BioNTech, in the US were identified.

Discussion

I perfectly confirm the research questions are well solved. The results of the project may give suggestions on how the countries take action when facing the outbreak next time. For example, start the vaccination early and devote continuous effort, take extreme care of the holidays where peoples take more chances to get together, watch out to the public attitudes on social media.

However, as I've mentioned at the beginning, the starting point of the project is the interest to the world, so this project may give more support on delivering 'what's happening right now' type a message to the ones interested.

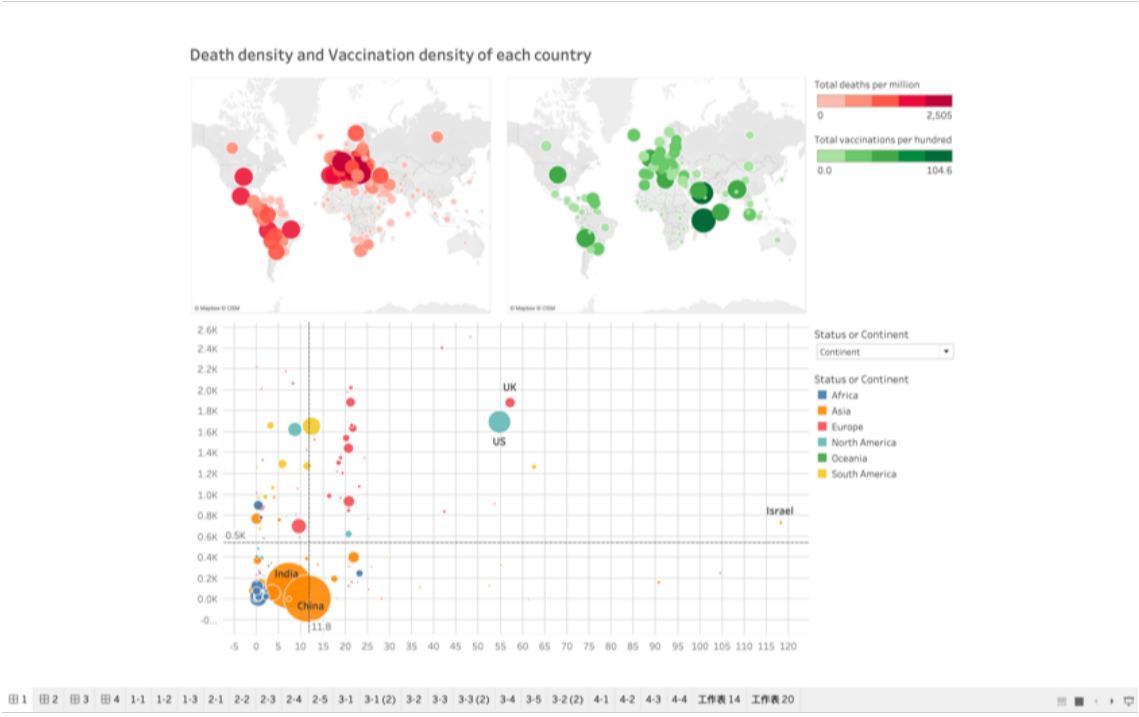
The limitation comes from the data per se. The data of COVID-19 is just started to be collected and ideally, should cover all the countries. However, first, not all the countries have the records in terms of the basic death and vaccination data; second, not all the data is well recorded for each unit of time, i.e., data missing problem is severe; third, detailed information has either not been collected systematically or been restricted from the ordinary researchers.

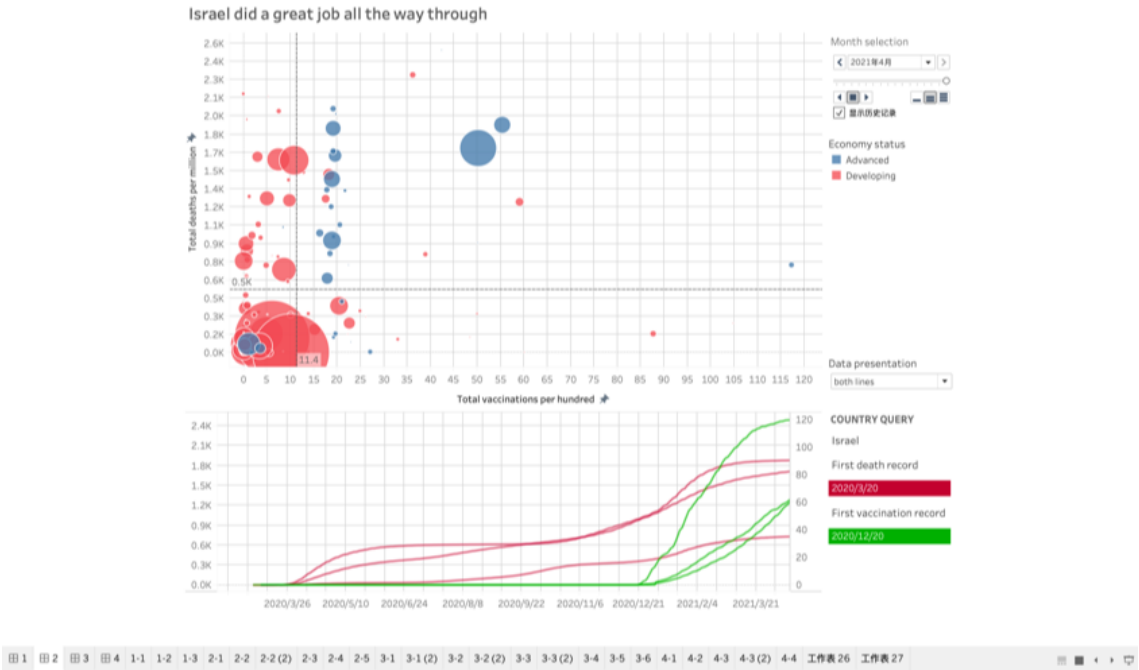
It would be better, if further analysis could be made on the supply and demand sides, focusing on the transportation of vaccines among the countries. Besides, building connections between the public attitudes and the number of daily vaccinations within each vaccine category would be much more helpful for the government interest.

References:

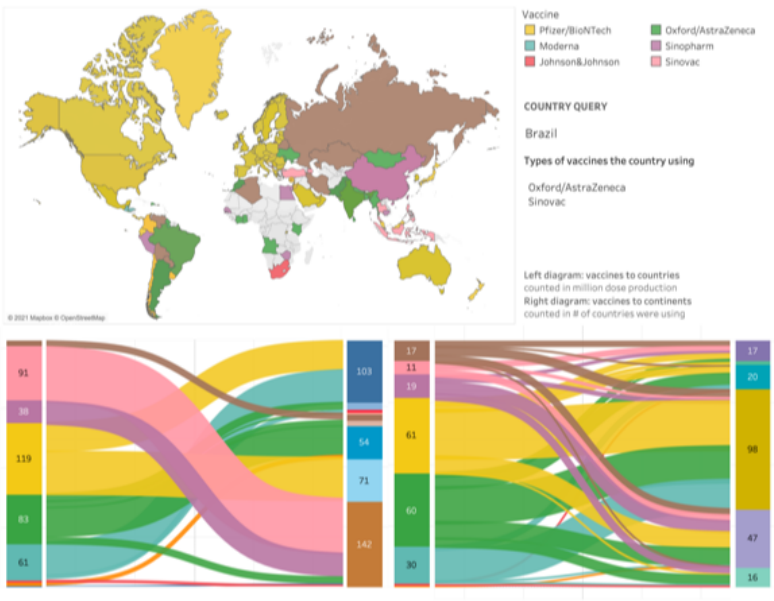
1. Wikipedia. (2021). Template: COVID-19 pandemic data. Retrieved April 29, 2021, from: https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data
2. Max, R., Hannah, R., Esteban, O.O. and Joe, H. (2020). Coronavirus Pandemic (COVID-19). Retrieved April 29, 2021, from: https://ourworldindata.org/coronavirus
3. Airfinity. COVID-19 Vaccine Production. Retrieved April 29, 2021, from: https://www.ifpma.org/wpcontent/uploads/2021/03/Airfinity_global_summit_master_final.pdf
4. Kaggle. Pfizer Vaccine Tweets. Retrieved April 29, 2021, from: https://www.kaggle.com/gpreda/pfizer-vaccine-tweets
5. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Figures:


Death density and Vaccination density of each country
Total deaths per million
0
2,505
Total vaccinations per hundred
0.0
104.6
Status or Continent
Continent
Status or Continent
Africa
Asia
Europe
North America
Oceania
South America
UK
US
Israel
India
China
11.8
0.5K


Israel did a great job all the way through
Total deaths per million
Total vaccinations per hundred
Month selection
2021年4月
显示历史记录
Economy status
Advanced
Developing
Data presentation
both lines
COUNTRY QUERY
Israel
First death record
2020/3/20
First vaccination record
2020/12/20
11.4
0.5K

Pfizer/BioNTech and Oxford/AstraZeneca are the most popular vaccines



1 2 3 4 1-1 1-2 1-3 2-1 2-2 2-2 (2) 2-3 2-4 2-5 3-1 3-1 (2) 3-2 3-2 (2) 3-3 3-3 (2) 3-4 3-5 3-6 4-1 4-2 4-3 4-3 (2) 4-4 工作表 26 工作表 27

Story of Pfizer/BioNTech with the investigation on tweets

TWEET QUERY

User Name	User Location	User Followers
The Epoch Times	New York, USA	403105
Date	**Favorites**	**Retweets**
2021/1/16	654	678

URL https://t.co/bxezEVXQX8

Content

55 people in the US have died after receiving a #Vaccine for #COVID19, according to reports submitted to a federal... https://t.co/bxezEVXQX8

1 2 3 4 1-1 1-2 1-3 2-1 2-2 2-2 (2) 2-3 2-4 2-5 3-1 3-1 (2) 3-2 3-2 (2) 3-3 3-3 (2) 3-4 3-5 3-6 4-1 4-2 4-3 4-4 4-5 4-6 工作表 26 工作表 27