

Data Visualization Analysis of COVID-19 Epidemic Situation

Yang Tao

Nanchang Institute of Science and Technology
Nanchang, China
E-mail: taoyangxp@163.com

Abstract—2019 Novel Coronavirus (COVID-19) has brought immeasurable losses and huge impact to the world. For human health, many Centres for Disease Control(CDC) in various countries around the world are actively collecting data and doing a good job in virus prevention and control. The real-time release of the epidemic situation, with analysis and prediction, is a very effective method to combat the epidemic. By studying the situation of epidemic data, based on Jupyter Notebook, this paper gives the visual analysis process of COVID-19 epidemic data, and carries out specific analysis and implementation. And then it estimates the coronavirus converges roughly using sigmoid fitting. Although the sigmoid fitting tend to underestimate the curve, its actual value tend to be more than sigmoid curve estimation. The proposed data visualization analysis method could effectively display the status of the COVID-19 epidemic situation, hoping to help control and reduce the impact of the COVID-19 epidemic.

Keywords—COVID-19; Data Visualization; Situation Dashboard; Jupyter Notebook; Epidemic Situation

I. INTRODUCTION

2019 Novel Coronavirus (2019-nCoV) is a virus (more specifically, a coronavirus) identified as the cause of an outbreak of respiratory illness. A growing number of patients reportedly have indicated person-to-person spread is occurring. At this time, it's unclear how easily or sustainably this virus is spreading between people.

Therefore, it is very important to visually analyse the COVID-19 Epidemic Situation, which helps to control the impact of the COVID-19 epidemic and reduce losses.

II. LOAD DATA

The dataset has daily level information on the number of affected cases, deaths and recovery from 2019 novel coronavirus. This is a time series data and so the number of cases on any given day is the cumulative number. The data is available from 22 Jan, 2020. We can download latest data from Johns Hopkins University github repository: <https://github.com/CSSEGISandData/COVID-19>[1]. We can also grab data from various Centres for Disease Control [2-6].

The data folder contains the previously posted dashboard case reports from Jan 21 to Feb 14, 2020 for the coronavirus COVID-19 (formerly known as 2019-nCoV). We will refer to the data provided in the new folder, entitled "csse_covid_19_data folder". Moving forward they will be updating daily case reports into this new folder. Additionally, the previously uploaded data from Jan 21-Feb 14, 2020 is also included in the new folder, and it has been cleaned and re-formatted to address inconsistencies in the time zone and update frequency that resulted during the transition from our manual updates to automated updates (which took place on Feb 1, 2020). The new folder now includes one case report per day, from the same time of day. This will be the standard moving forward (as of Feb 14, 2020). That is the data we will load for visualization analysis.

Main file in this dataset is covid_19_data.csv and the detailed descriptions are below.

- ✧ Sno - Serial number
- ✧ ObservationDate - Date of the observation in MM/DD/YYYY. We will convert ObservationDate and Last Update to datetime since they are currently taken as object.

- ✧ Province/State - Province or state of the observation (Could be empty when missing)
- ✧ Country/Region - Country of observation
- ✧ Last Update - Time in UTC at which the row is updated for the given province or country. (Not standardised and so please clean before using it)
- ✧ Confirmed - Cumulative number of confirmed cases till that date
- ✧ Deaths - Cumulative number of deaths till that date
- ✧ Recovered - Cumulative number of recovered cases till that date

III. VISUALIZATION ANALYSIS

For the purpose of data visualization, we mainly use the Python-based tools of Jupyter Notebook[7] and plotly[8]. The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more. The plotly visualization is heavy used in this kernel so that we can interactively see the figure, map etc. As a side effect, it might take a little bit more time to initialize the Python environment and to load the kernel. Then grab data from the Internet and load the data.

A. Worldwide Trend

When we see the confirmed cases in worldwide, it just look like exponential growth curve. The number is increasing very rapidly especially recently. As a further matter, daily new confirmed cases started not increasing from April 4. After that, flat trend continues so far, as shown in Figure 1.

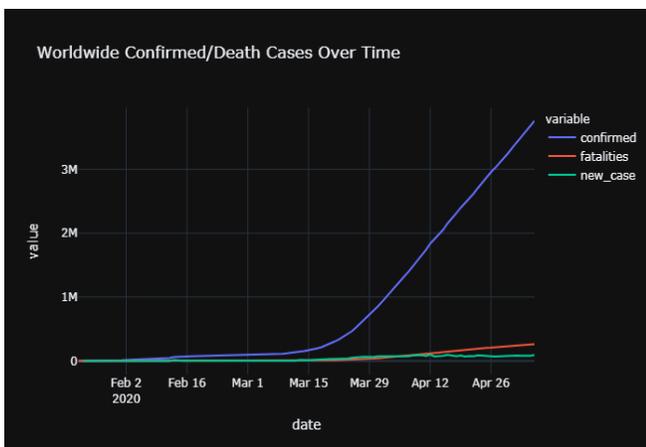


Figure 1. Worldwide Confirmed/Death Cases Over Time

Moreover, when we check the growth in log-scale below figure, we can see that the speed of confirmed cases growth rate slightly increases when compared with the beginning of March and end of March. In spite of the Lockdown policy in Europe or US, the number is still increasing rapidly, as shown in Figure 2.



Figure 2. Worldwide Confirmed/Death Cases Over Time (Log scale)

It looks like fatalities curve is just shifted the confirmed curve to below in log-scale, which means mortality rate is almost constant. We see that mortality rate is kept almost 3%, however it is slightly increasing gradually to go over 7% at the end of April. Europe & US has more seriously infected by Coronavirus recently, and mortality rate is high in these regions, as shown in Figure 3. It might be because when too many people get coronavirus, the country cannot provide enough medical treatment.



Figure 3. Worldwide Mortality Rate Over Time

B. Country-wise Growth

There are 187 countries in the dataset. How's the distribution of number of confirmed cases by country? It is difficult to see all countries so let's check top countries as shown in Figure 4.

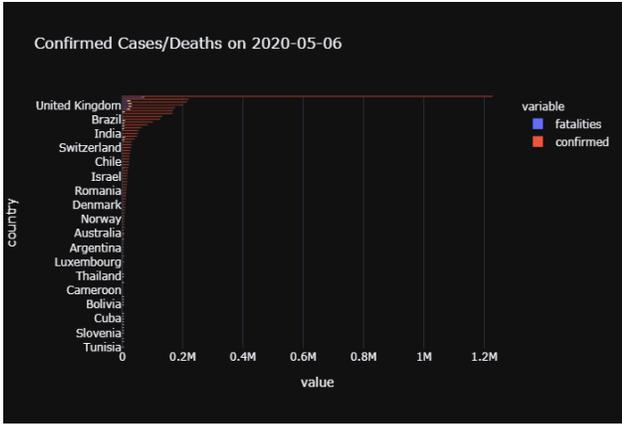


Figure 4. Confirmed Cases/Deaths on 2020-05-06

Now US, Italy and Spain has more confirmed cases than China, and we can see many Europe countries in the top. Korea also appears in relatively top despite of its population, this is because Korea executes inspection check aggressively.

Let's check these major country's growth by date.

As we can see, Coronavirus hit China at first but its trend is slowing down in March which is good news. Bad news is 2nd wave comes to Europe (Italy, Spain, Germany, France, UK) at March. But more sadly 3rd wave now comes to US, whose growth rate is much faster than China, or even Europe. Its main spread starts from middle of March and its speed is faster than Italy. Now US seems to be in the most serious situation in terms of both total number and spread speed. Now let's see the confirmed cases for the top 30 countries, as shown in Figure 5.

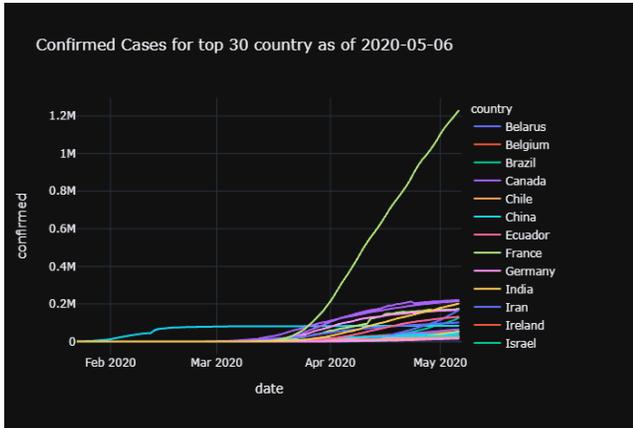


Figure 5. Confirmed Cases for top 30 country as of 2020-05-06

In terms of number of fatalities, Europe & US are serious situation now, as shown in Figure 6. Many countries have more fatalities than China now, including US, Italy, Spain, France, UK, Iran Belgium,

Germany, Brazil, Netherlands. US's spread speed is the fastest, US's fatality cases become top1 on Apr 10th.

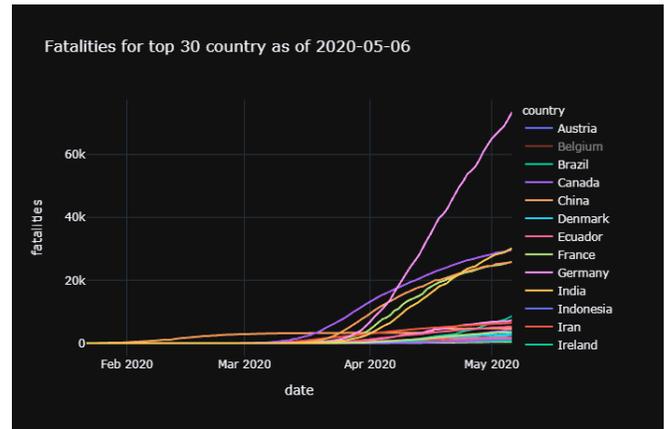


Figure 6. Fatalities for top 30 country as of 2020-05-06

Now let's see mortality rate by country, as shown in Figure 7.

Italy is the most serious situation, whose mortality rate is over 10% as of 2020/3/28. We can also find mortality rate countries, as shown in Figure 7. Iran/Iraq from Middle East, Philippines & Indonesia from tropical areas. Spain, Netherlands, France, and UK form Europe etc. It shows this coronavirus is really worldwide pandemic.

The countries whose mortality rate is low are shown in Figure 8.

By investigating the difference between above & below countries, we might be able to figure out what is the cause which leads death.

Be careful that there may be a case that these country's mortality rate is low due to these country does not report/measure fatality cases properly.

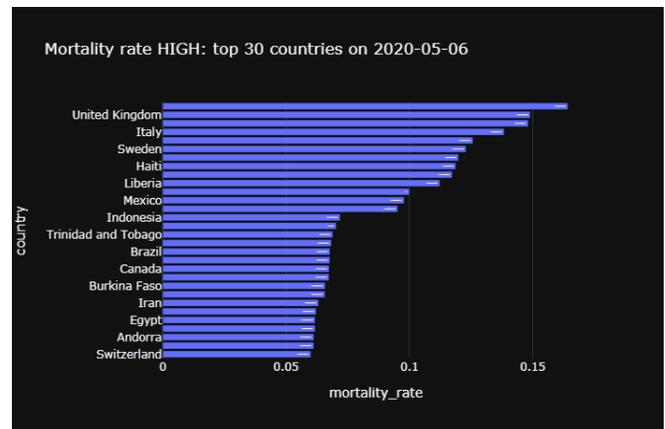


Figure 7. Mortality rate HIGH: top 30 countries on 2020-05-06

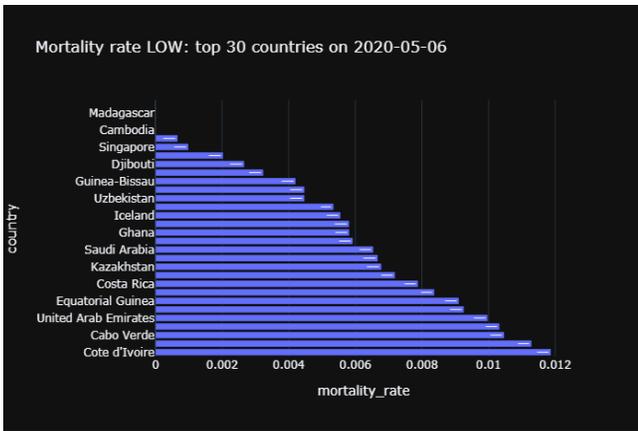


Figure 8. Mortality rate HIGH: top 30 countries on 2020-05-06

Let's see number of confirmed cases on map. Again we can see Europe, US, Middle East (Turkey, Iran) and Asia (China, Korea) are red, as shown in Figure 9.



Figure 9. Countries with Confirmed Cases on 2020-05-06

The number of fatalities on map is shown as Figure 10 and the mortality rate map is shown as Figure 11.

When we see mortality rate on map, we see Europe (especially Italy) is high. Also we notice Middle East (Iran, Iraq) is high. When we see tropical area, I wonder why Philippines and Indonesia are high while other countries (Malaysia, Thai, Vietnam, as well as Australia) are low. For Asian region, Korea's mortality rate is lower than China or Japan, I guess this is due to the fact that number of inspection is quite many in Korea[9-10].

From the mortality rate map, it seems that mortality rate is especially high in Europe region, compared to US or Asia.

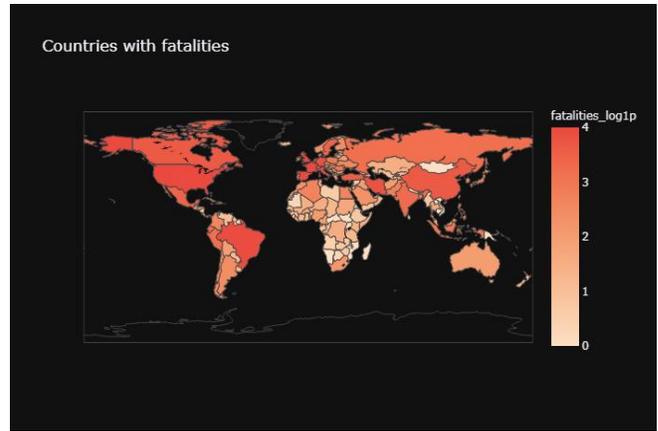


Figure 10. Countries with fatalities on 2020-05-06

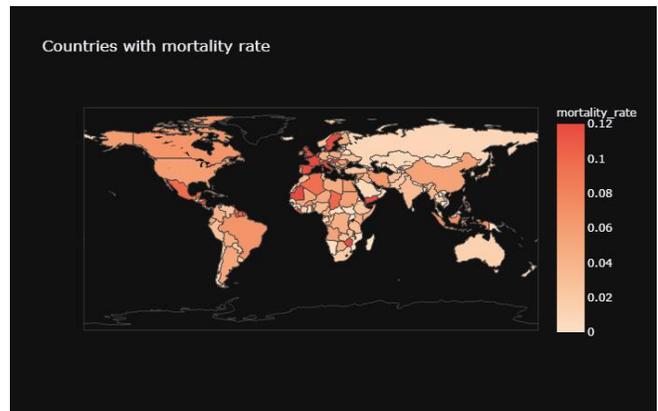


Figure 11. Countries with mortality rate on 2020-05-06

Why mortality rate is different among country? What kind of hint is hidden in this map? Especially mortality rate is high in Europe and US, is there some reasons? There is one interesting hypothesis that BCG vaccination[11].

C. Daily NEW Confirmed Cases Trend

Let's see the DAILY new cases trend as shown in Figure12.

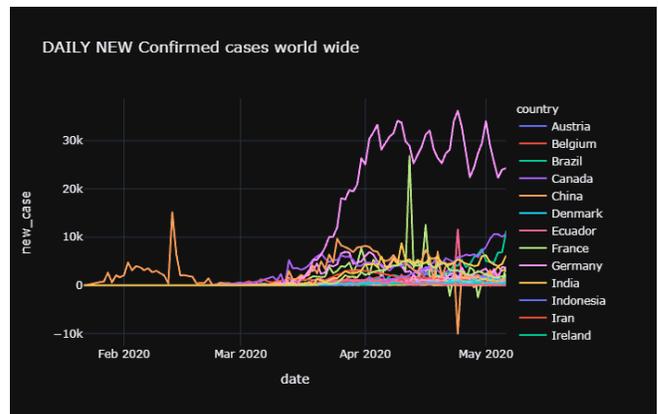


Figure 12. DAILY NEW Confirmed cases worldwide

We find from the figure 12:

- ✧ China has finished its peak at Feb 14, new confirmed cases are suppressed now.
- ✧ Europe&US spread starts on mid of March, after China slows down.
- ✧ As effect of lock down policy in Europe (Italy, Spain, Germany, France) now comes on the figure, the number of new cases are not so increasing rapidly at the end of March.
- ✧ Current US new confirmed cases are the worst speed, recording worst speed at more than 30k people/day at peak. Daily new confirmed cases start to decrease from April 4 or April 10.
- ✧ After that we can see a weekly trend that the confirmed cases becomes small on Monday. I think this is because people don't (or cannot) get medical care on Sunday so its reporting number is low on Sunday or Monday.

D. Zoom up to US

As we can see, the spread is fastest in US now, at the end of March. Let's see in detail what is going on in US. When we see inside of the US, we can see only New York, and its neighbour New Jersey dominates its spread and are in serious situation. The number of New York confirmed cases is over 50k, while other states are less than about 5k confirmed cases, as shown in Figure 13.

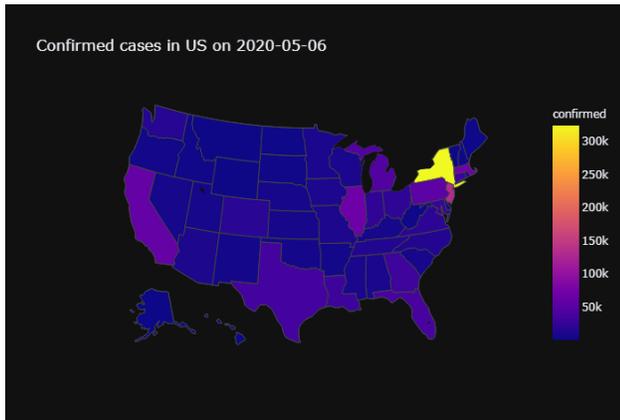


Figure 13. Confirmed cases in US on 2020-05-06

Mortality rate in New York seems not high, around 2% for now, as shown in Figure 14.

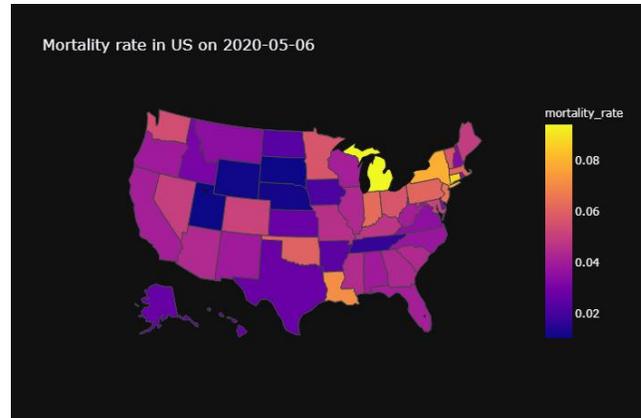


Figure 14. Mortality rate in US on 2020-05-06

All state is US got affected from middle of March, and now growing exponentially. In New York, less than 1k people are confirmed on March 16, but more than 50k people are confirmed on March 30. 50 times explosion in 2 weeks! The confirmed cases by state in US is show in Figure 15.

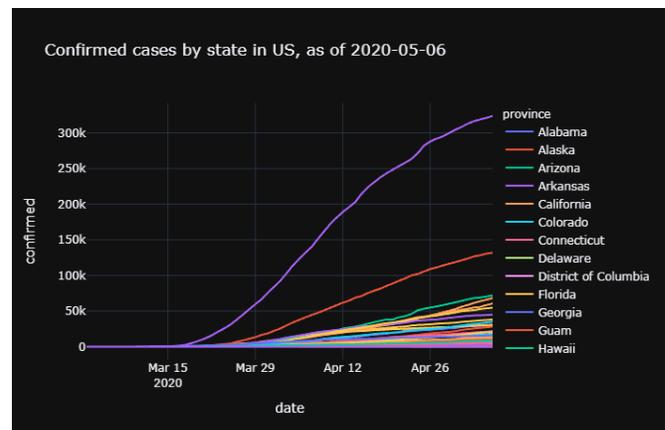


Figure 15. Confirmed cases by state in US, as of 2020-05-06

E. Zoom up to Europe

When we look into the Europe, its Northern & Eastern areas are relatively better situation compared to Eastern & Southern areas. The map of European Countries with Confirmed Cases is shown as Figure 16 and Figure 17.



Figure 16. European Countries with Confirmed Cases, as of 2020-05-06

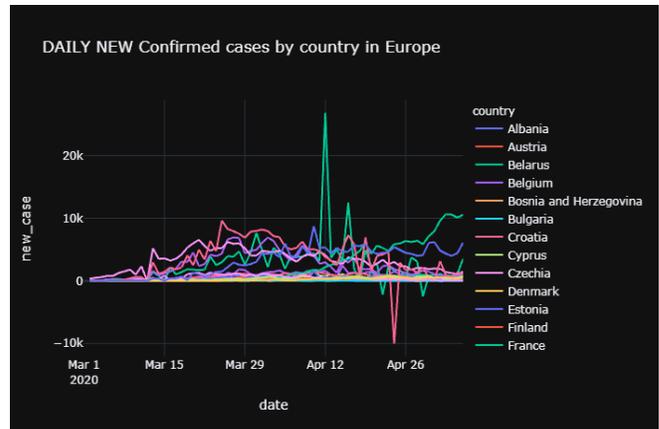


Figure 18. DAILY NEW Confirmed cases by country in Europe

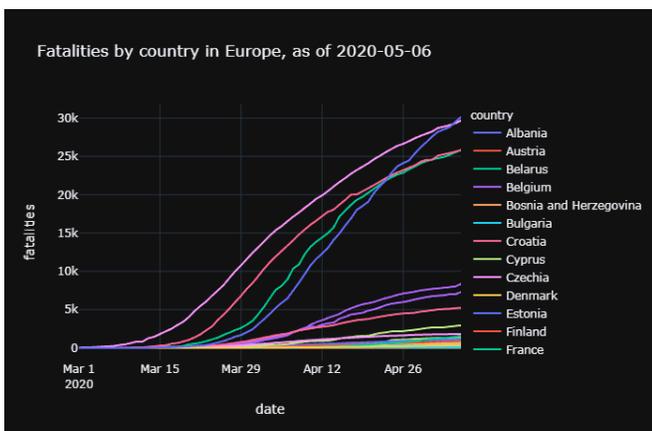


Figure 17. Confirmed cases by country in Europe, as of 2020-05-06

Especially Italy, Spain, German, France, UK are in more serious situation. Number of confirmed cases rapidly increasing in Russia now (as of May 1), Russia is now potentially very dangerous situation.

When we check daily new cases in Europe(as shown in Figure 18), we notice:

- ✧ UK and Russia daily growth are more than Italy now. These countries are potentially more dangerous now.
- ✧ Italy new cases are not increasing since March 21, maybe due to lock-down policy is started working.

F. Zoom up to Asia

In Asia, China & Iran have many confirmed cases, followed by South Korea & Turkey. Asian Countries with Confirmed Cases is as shown in Figure 19.

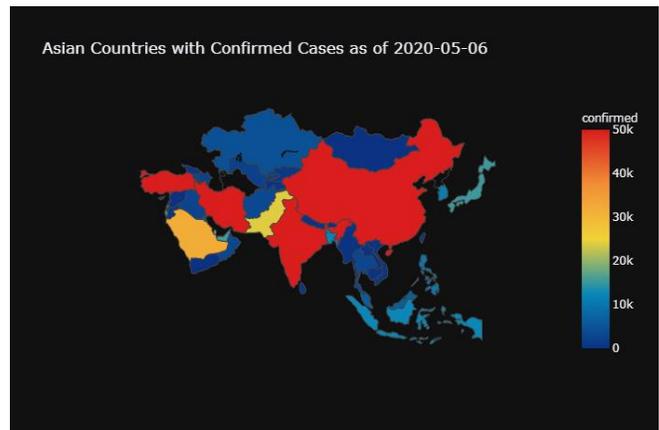


Figure 19. Confirmed cases by country in Asia, as of 2020-05-06

The coronavirus hit Asia in early phase, how is the situation now?

China & Korea is already in decreasing phase. Unlike China or Korea, daily new confirmed cases were kept increasing on March or April, especially in Iran or Japan. But the number is started to decrease now on these country as well, as shown in Figure 20.

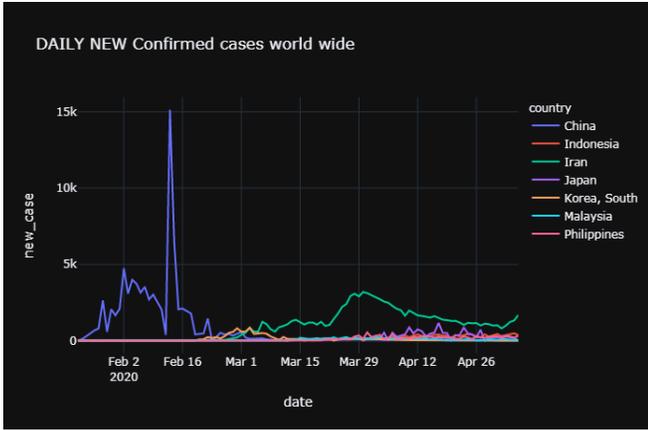


Figure 20. DAILY NEW Confirmed Cases in Asia, as of 2020-05-06

IV. ESTIMATION

Of course everyone is wondering when the coronavirus converges. Let's estimate it roughly using sigmoid fitting.

I referenced two kernels[12-13] for original ideas. The fitting result is shown in Figure 21.

Sigmoid fitting with all latest data

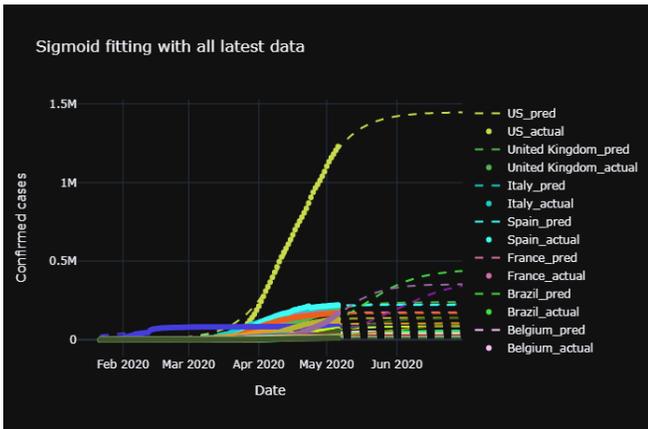


Figure 21. Sigmoid fitting with all latest data

If believe above curve, the number of confirmed cases is slowing down now and it will be converging around the beginning of May in most of the country. It might take until beginning on June in US.

Let's try validation by excluding last 7 days data, as shown in Figure 22.

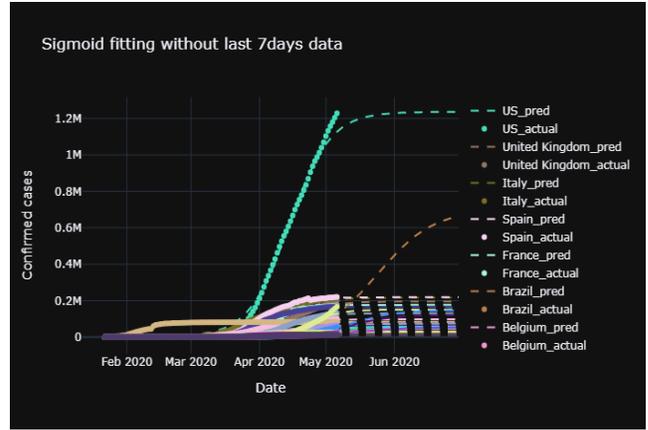


Figure 22. Sigmoid fitting without last 7days data

Now noticed that sigmoid fitting tend to underestimate the curve, and its actual value tend to be more than sigmoid curve estimation.

Therefore, need to be careful to see sigmoid curve fitting data; actual situation is likely to be worse than the previous figure trained with all data.

V. CONCLUSION

Based on data available on May 6, the paper showed the visualization of the COVID-19 Epidemic Situation, including the worldwide trend, country-wide growth, and so on. Then it estimated when the coronavirus converges roughly using sigmoid fitting. The model's estimates and predictions closely match reported confirmed cases. Therefore the proposed data visualization analysis method could effectively display the status of the COVID-19 epidemic situation, hoping to help control and reduce the impact of the COVID-19 epidemic.

The next steps include applying the method to global COVID-19 death data into small regions, as provinces. The method of visualization analysis could also be used to evaluate population mortality and the spread of other diseases.

ACKNOWLEDGMENTS

The authors wish to thank Corochann, who is an Engineer at Preferred Networks in Tokyo. This work was supported in part by the Ph.D. Research Initiation Fund of Nanchang Institute of Science and Technology with the Project (No. NGRCZX-18-01) and supported by Robot and Intelligent System Research Centre of Artificial Intelligence College of Nanchang Institute of Science and Technology (No.NGYZY-20-005).

REFERENCES

- [1] <https://www.kaggle.com/benhamner/covid-19-forecasting-challenges-week-2-data-prep>.
- [2] China CDC (CCDC): <http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm>
- [3] Taiwan CDC: <https://sites.google.com/cdc.gov.tw/2019ncov/taiwan?authuser=0>
- [4] US CDC: <https://www.cdc.gov/coronavirus/2019-ncov/index.html>
- [5] Government of Canada: <https://www.canada.ca/en/public-health/services/diseases/coronavirus.html>
- [6] European Centre for Disease Prevention and Control (ECDC): <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>
- [7] Jupyter notebook - Project Jupyter | Home. <https://jupyter.org/>
- [8] Plotly: The front-end for ML and data science models. <https://plotly.com/>
- [9] South Korea launches 'drive-thru' coronavirus testing facilities as demand soars. <https://www.japantimes.co.jp/news/2020/03/01/asia-pacific/science-health-asia-pacific/south-korea-drive-thru-coronavirus/#.XoAmw4j7RPY>
- [10] Coronavirus: Why Japan tested so few people. <https://asia.nikkei.com/Spotlight/Coronavirus/Coronavirus-Why-Japan-tested-so-few-people>
- [11] If I were North American/West European/Australian, I will take BCG vaccination now against the novel coronavirus pandemic. <https://www.jsatonotes.com/2020/03/if-i-were-north-american-european-austral.html>
- [12] Sigmoid per country. <https://www.kaggle.com/group16/sigmoid-per-country-no-leakage>
- [13] COVID-19 growth rates per country. <https://www.kaggle.com/mikestubna/covid-19-growth-rates-per-country>.