

Comparison of Covid-19 confirmed and death cases between the US and China

Jiawen Chen

07 December 2022

1. Background story & Introduction

I was in the US at the beginning of 2020, and then back to China in March 2020 and quarantined. So I observed how did the pandemic spread so quickly and sharply increased around the world and specifically in the US. On the other side, after the few first months in 2020, the pandemic was considered as well controlled. Therefore, in this project, firstly I will briefly check through the overall COVID-19 situation in the world wide. Then I analyzed China and the US separately to see the trend. The final step is to compare with those two countries in terms of confirmed and deaths number.

2. Methods

About the data

I downloaded the datasets from “<https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases?>” in CSV format and updated daily and then chose the global confirmed and deaths throughout the whole pandemic (1/22/20-recent two days). The datas are updated everyday by John Hopkins center.

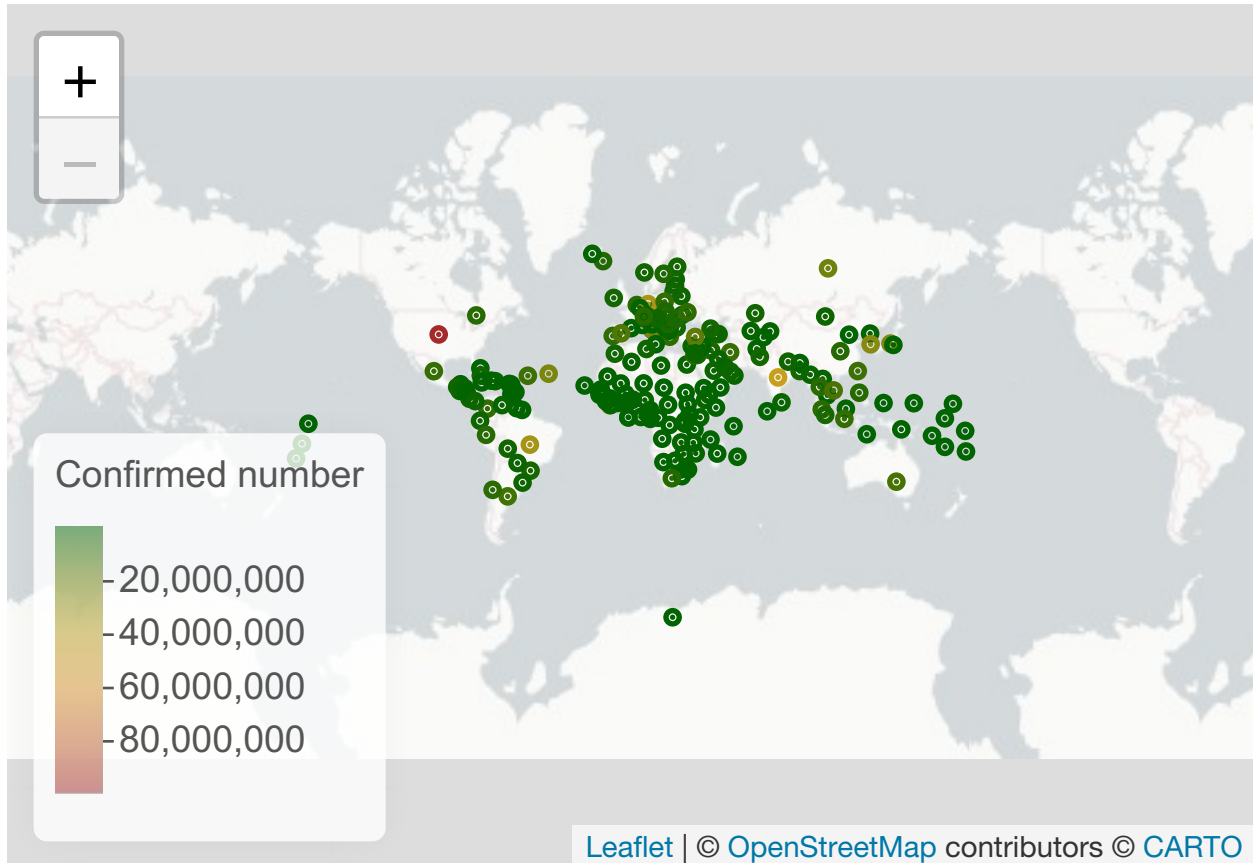
Preparation

However, the format of two datasets are about the accumulated-confirmed and death cases of different countries. So I did some filter out steps of countries/regions, cleaned and normalized that data, for example tidying dates and consolidating them into normalized time series. We have variables called “province” “Country/Region”, “dates” and “Lat” and “Long” representing different regions. Notice that each day’s count are in separate columns. For this analysis, it would be nicer to have a column for date and a column for count instead. Also I select and transformed these two countries from the total death dataset into data frames. They have variables called “province” “Country/Region”, “Lat”, “Long” and different dates after transforming.

3. Results

Figure 1. The heat map is the relationship between confirmed number and every Country/Region that the dataset has. The confirmed cases spread in the world wide. The red color means higher confirmed number. The green color means smaller confirmed number.

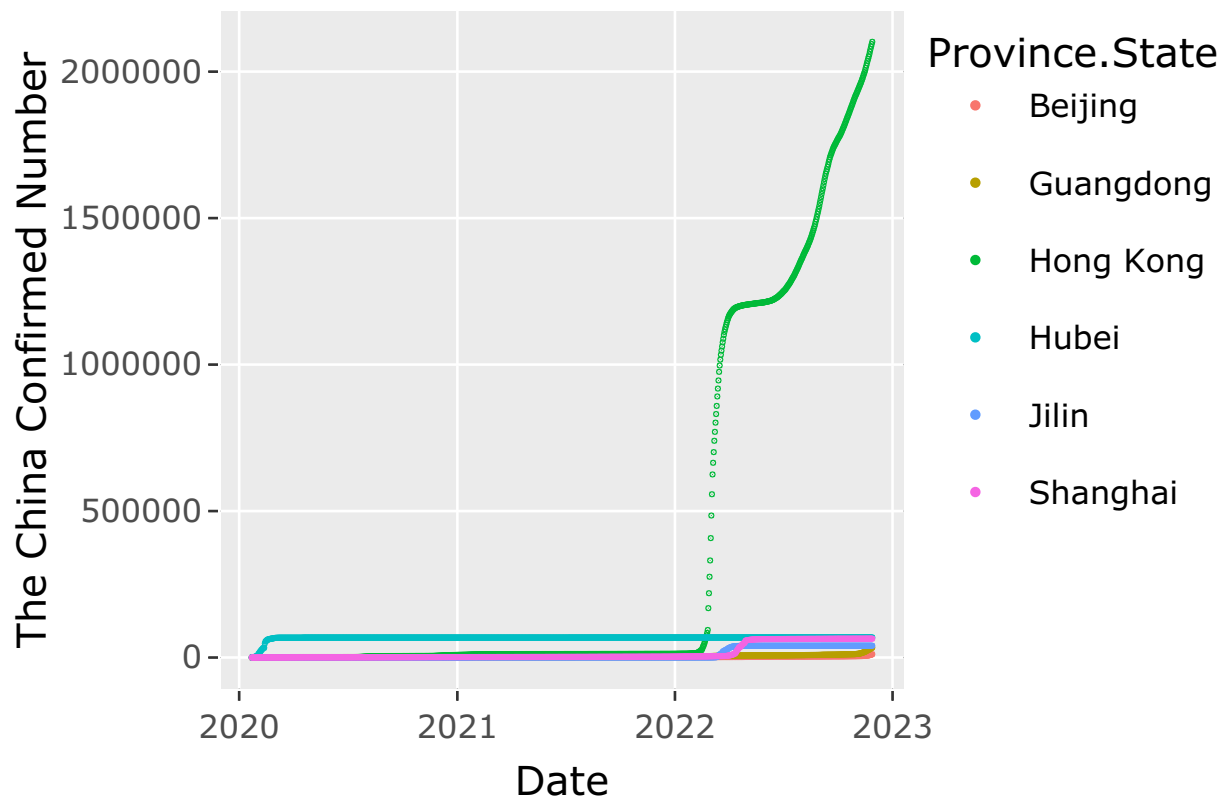
##	Province.State	Country.Region	Lat	Long	sumconfirmed
## 1	<NA>	Afghanistan	33.93911	67.70995	206414
## 2	<NA>	Albania	41.15330	20.16830	333472
## 3	<NA>	Algeria	28.03390	1.65960	271128
## 4	<NA>	Andorra	42.50630	1.52180	47446
## 5	<NA>	Angola	-11.20270	17.87390	104808
## 6	<NA>	Antarctica	-71.94990	23.34700	11

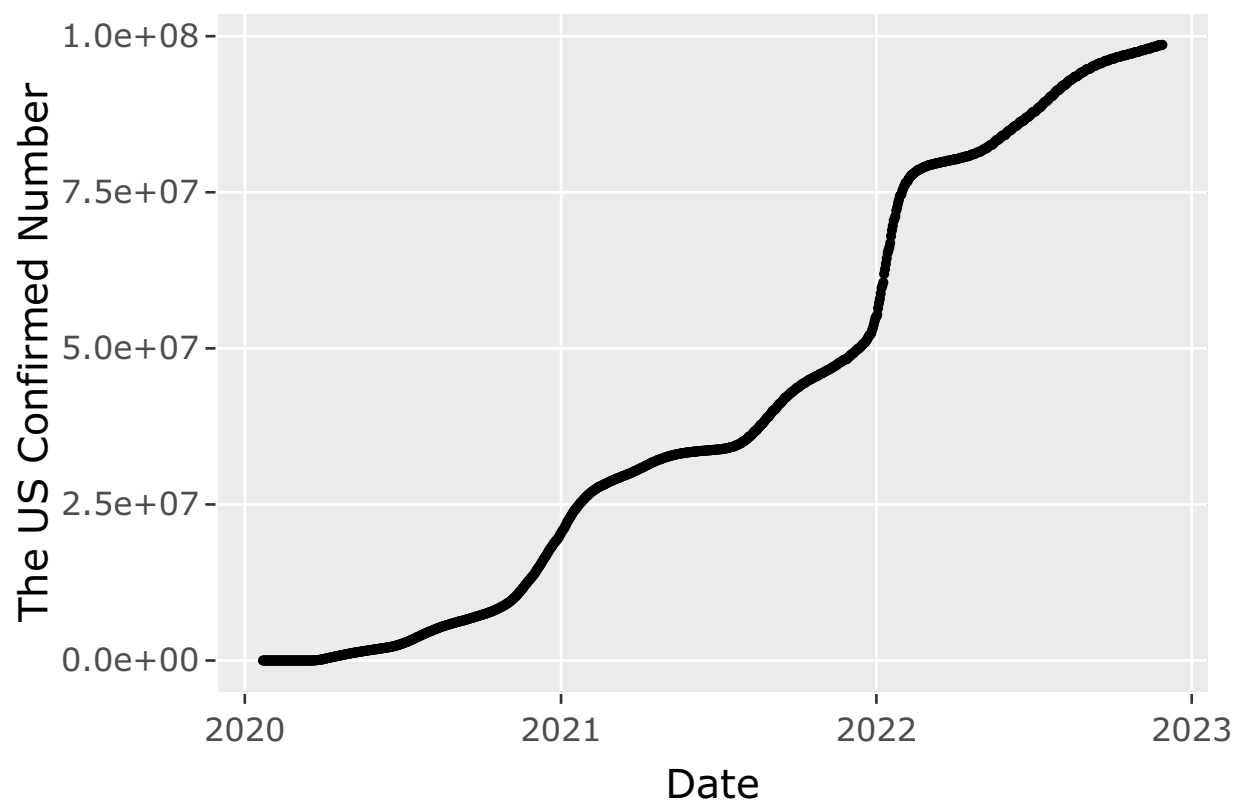


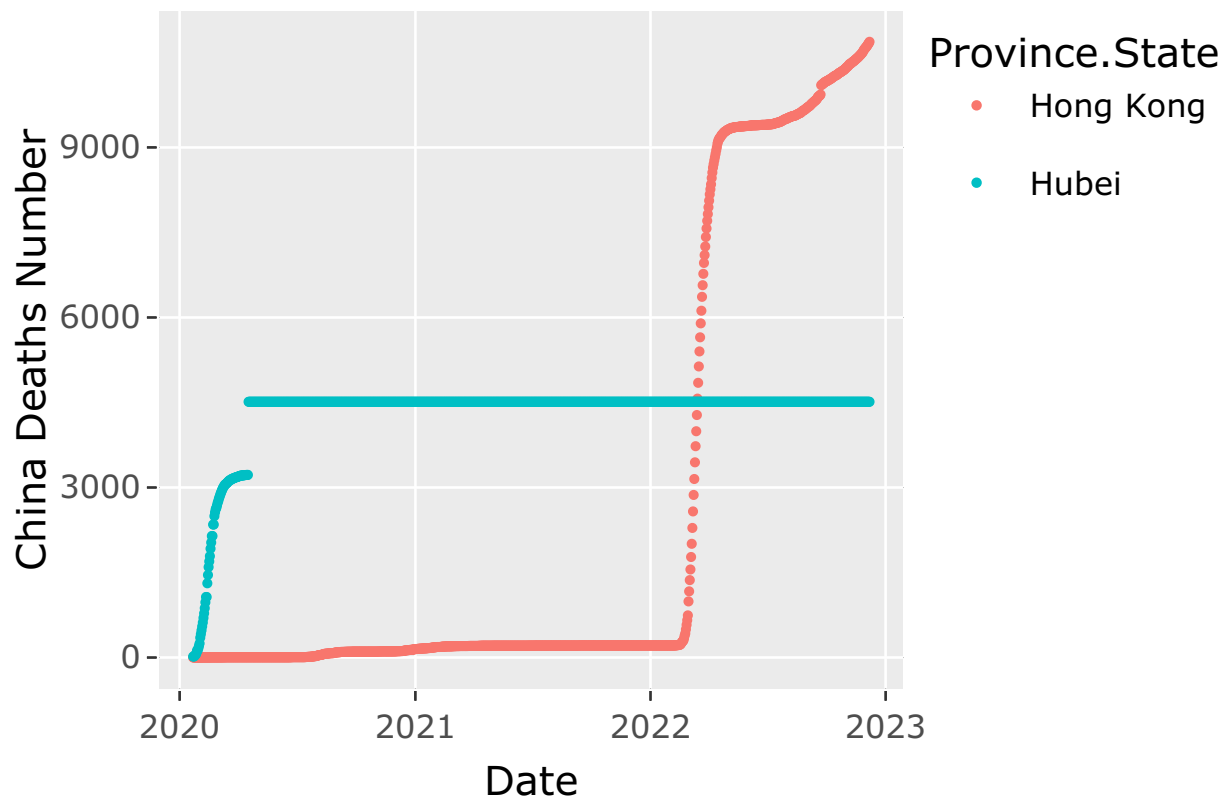
Here it shows how the confirmed cases spread in the world wide. From the map, the legend shows the range of confirmed number. We can see that there America has the red dot, then India, Brazil and Germany have yellow dots. However, one drawback of this dataset is that some country, for example China's data is updated by Province, but the data from the US as a country total.

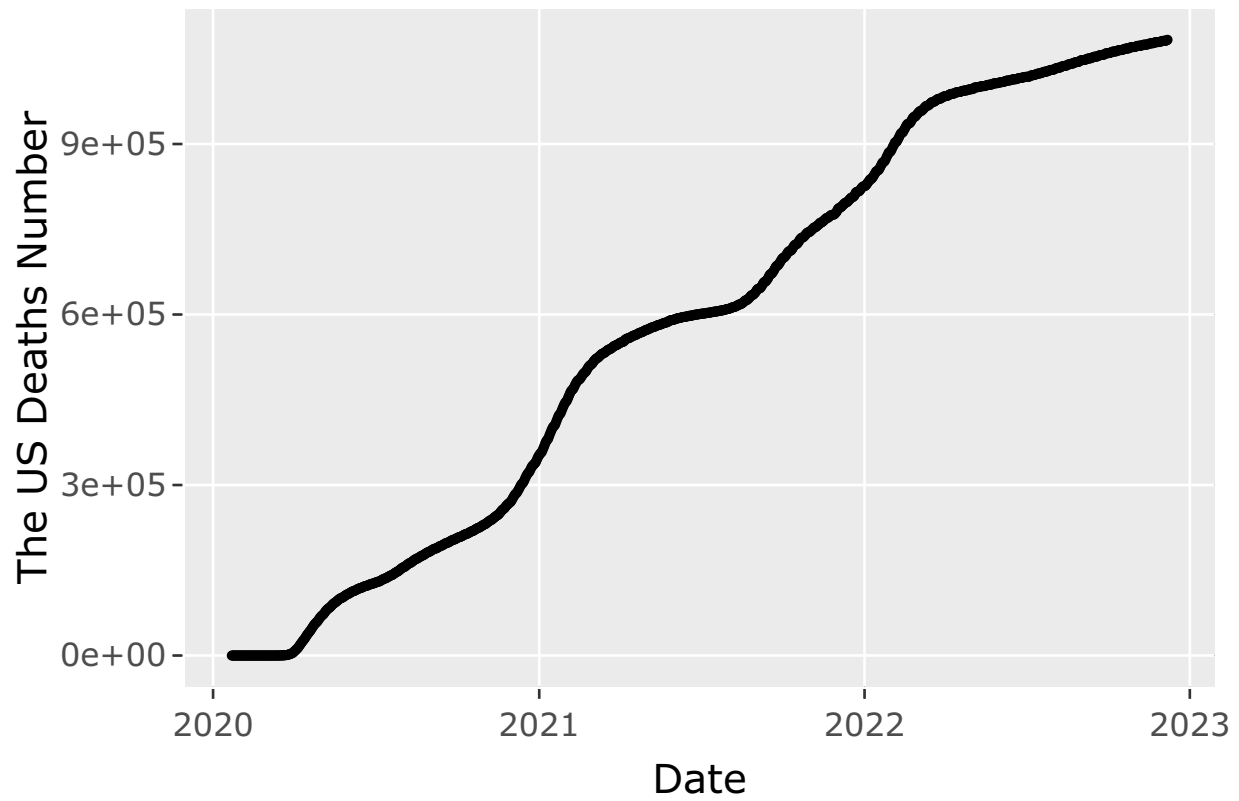
Figure 2. Graphs of each province in china and US confirmed and deaths cases.

As I mentioned above, the dataset only has the whole US data, so I just present the US as a country total. China has 50+ provinces in total and thus can be hard to represent them all, and I choose the top 6 provinces that have the largest number of confirmed cases. For the death cases, I only choose the provinces have more than 1000 death cases. The total confirmed cases in the US is around 100M and deaths cases is around 1M. Both confirmed and deaths trends of the US are continuous increasing, near to a certain slope.



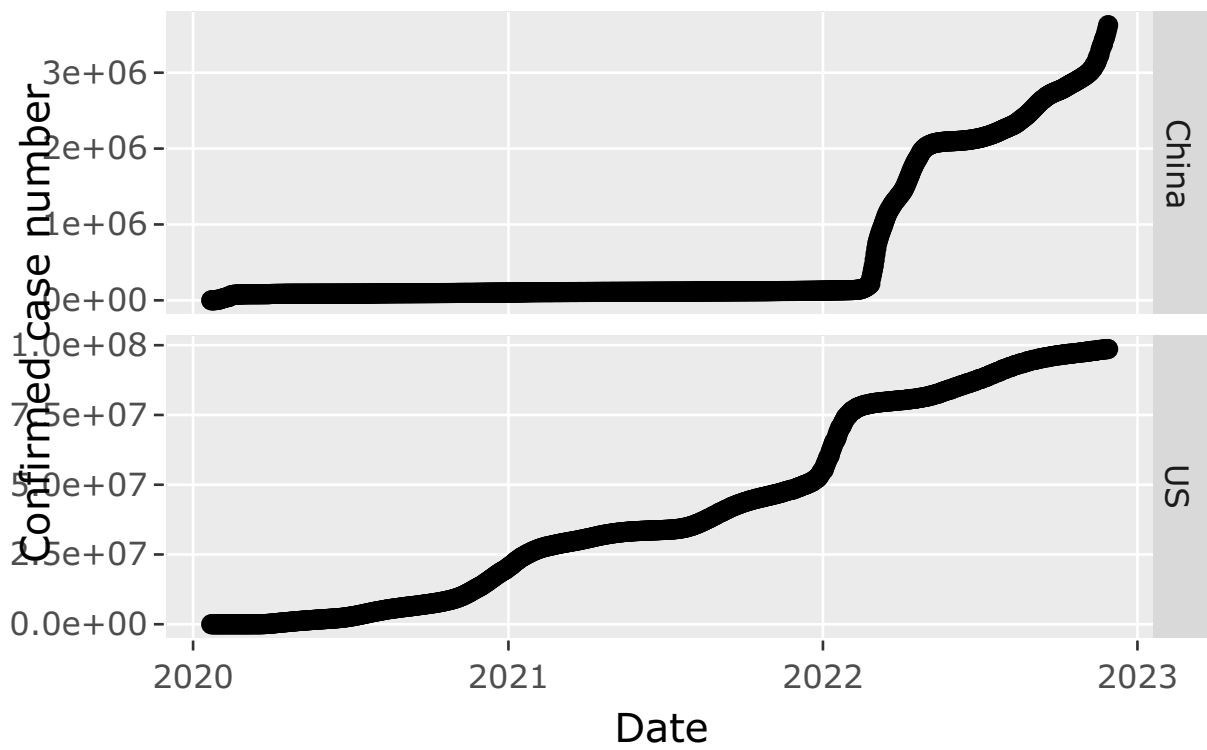


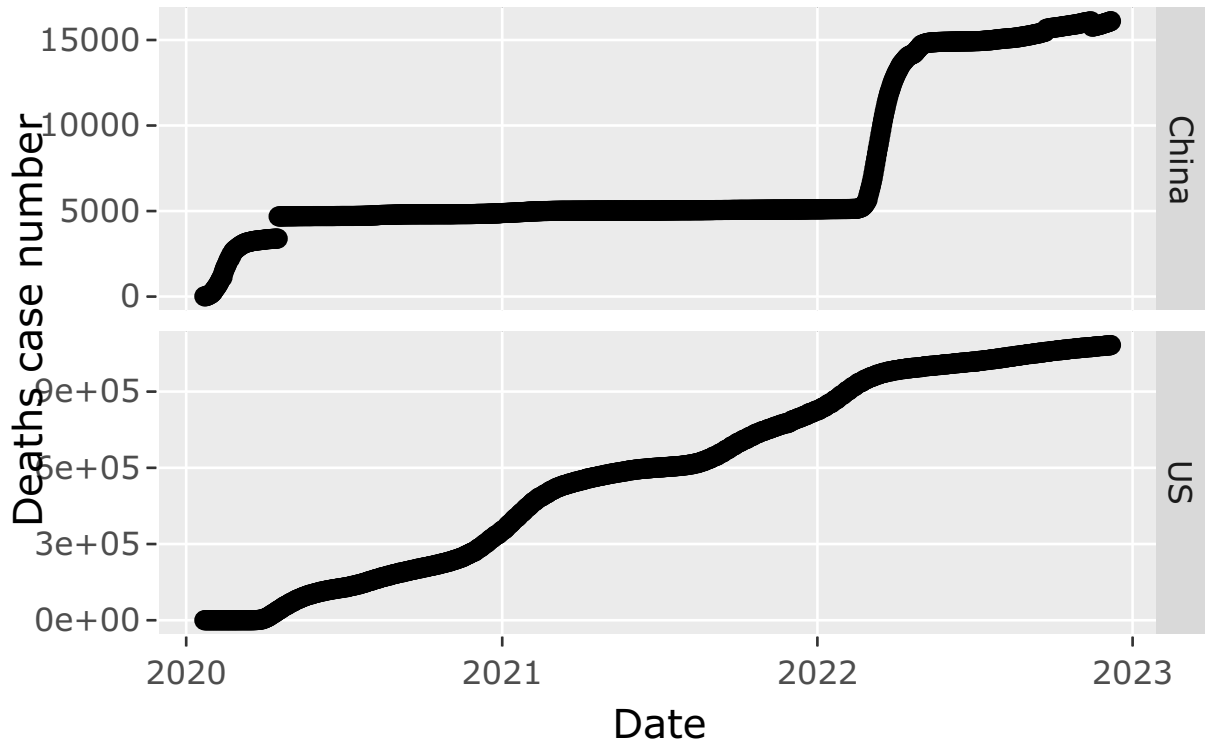




The X-axis is the date(year) and Y-axis is Confirmed/Deaths number. The top 6 provinces/region that have the largest number confirmed case are Hongkong, Hubei, Shanghai, Jilin, Guangdong and Beijing. Hongkong is the city with the largest COVID-19 confirmed case of 200 million according to the graphs, and it is the only region with more than 5000 death cases. The second one is Hubei, with around 68000 confirmed case number, 4500 deaths case number.

Figure 3. Two contries in comparison. The top graph is the comparison of confirmed cases number and the under one is the deaths cases number. The black line shows how does the number increase by date.





From these two graphs, we can conclude from 2020 to the beginning of 2022, China had a slowly increasing confirmed case number while the US rapidly increased. However, after 2022, this number of China sharply increased while the US has a relatively mild increasing trend.

4. Conclusion and Summary

We can see China started confirming cases earlier than the US but the confirmed and deaths cases increased slightly until the beginning of 2022. However, the overall confirmed and deaths cases in the US are significantly higher than China (100M VS 1.81M, 1M VS 5,235). Part of the reason is probably because of the strict COVID-19 policies. In terms of HongKong, it executed different policies compared with the mainland of China. Nowadays COVID-19 is evolving to be easier to be spreaded, this may result in the recent increasing trend in China.

5. Source

- <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>
- <https://news.google.com/covid19/map?hl=en-US&mid=%2Fm%2F0fbp0&gl=US&ceid=US%3Aen>