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

Jiawen Chen

2022-12-07

This is my PM566 Final Project website.

1.Backgroud story & Introduction

I was in the US at the beginning of 2020, and then back to China in March 2020 and quarentined. So I observed how did the pandemic spreaded so quickly and sharply increased around the world and specificly 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 seperatly to see the trend. The final step is to compare with those two countries in terms of confirmed and deaths number.

2.Methods:(include how and where the data were acquired, how you cleaned and wrangled the data, what tools you used for data exploration)

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.

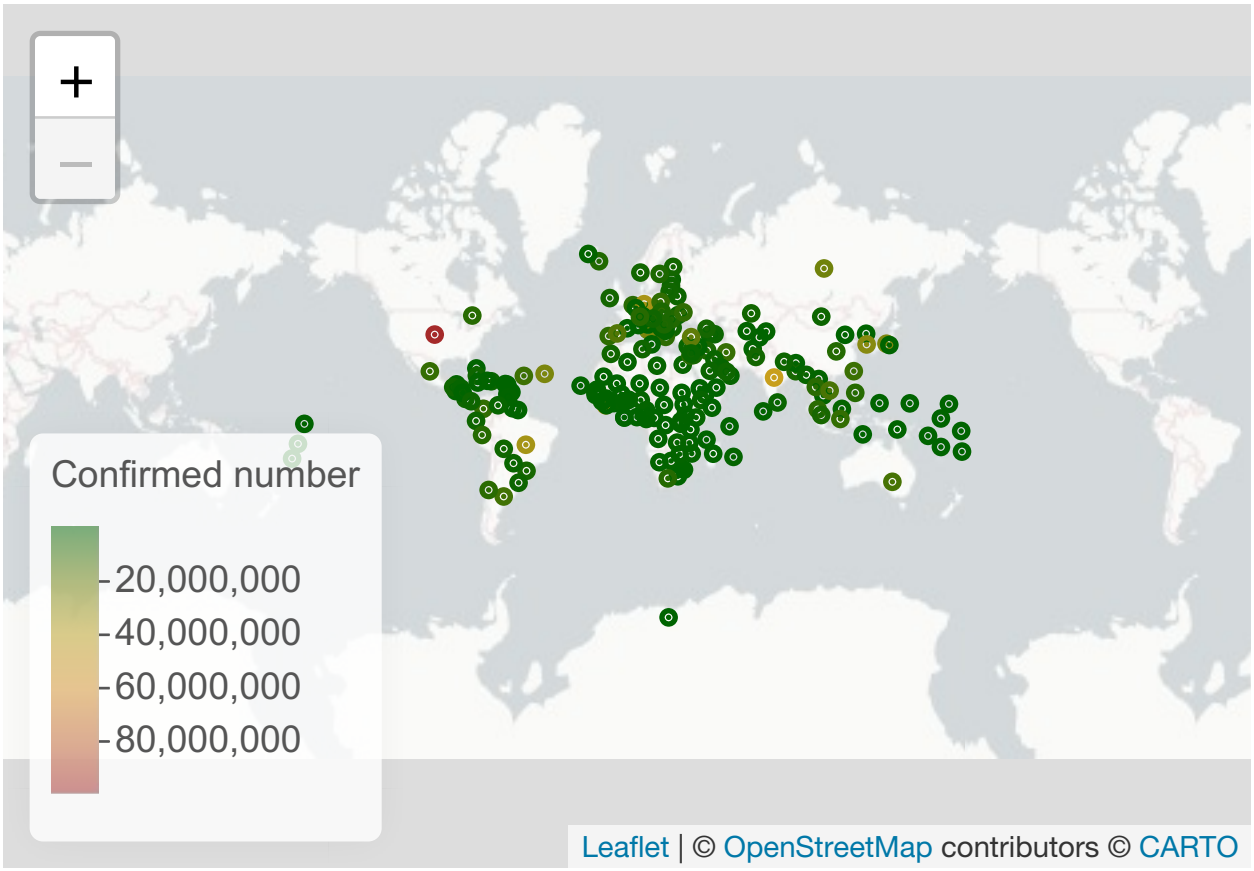
Preparation

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.

We have variables called "province" ""Country/Region", "Lat", "Long" and different dates

3. Here it shows how the confirmed cases spread in the world wide

```
##   Province.State Country.Region       Lat     Long sumconfirmed
## 1           <NA>    Afghanistan  33.93911 67.70995       206331
## 2           <NA>        Albania  41.15330 20.16830       333455
## 3           <NA>        Algeria  28.03390  1.65960       271122
## 4           <NA>        Andorra  42.50630  1.52180        47219
## 5           <NA>         Angola -11.20270 17.87390       104750
## 6           <NA>     Antarctica -71.94990 23.34700           11
```



From the map, the legend shows the range of confirmed number. The red color means higher 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.

4. Table to check total confirmed cases for each country

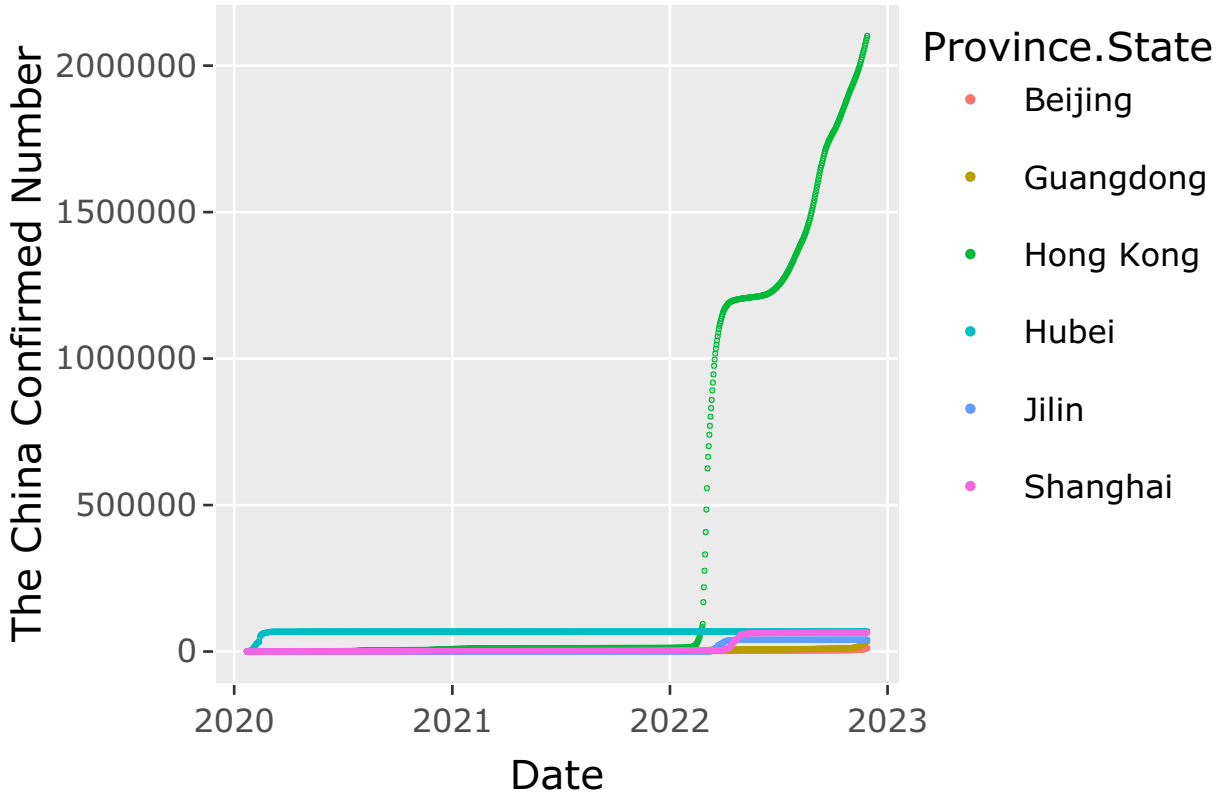
Province.State	Country.Region	Lat	Long	sumconfirmed

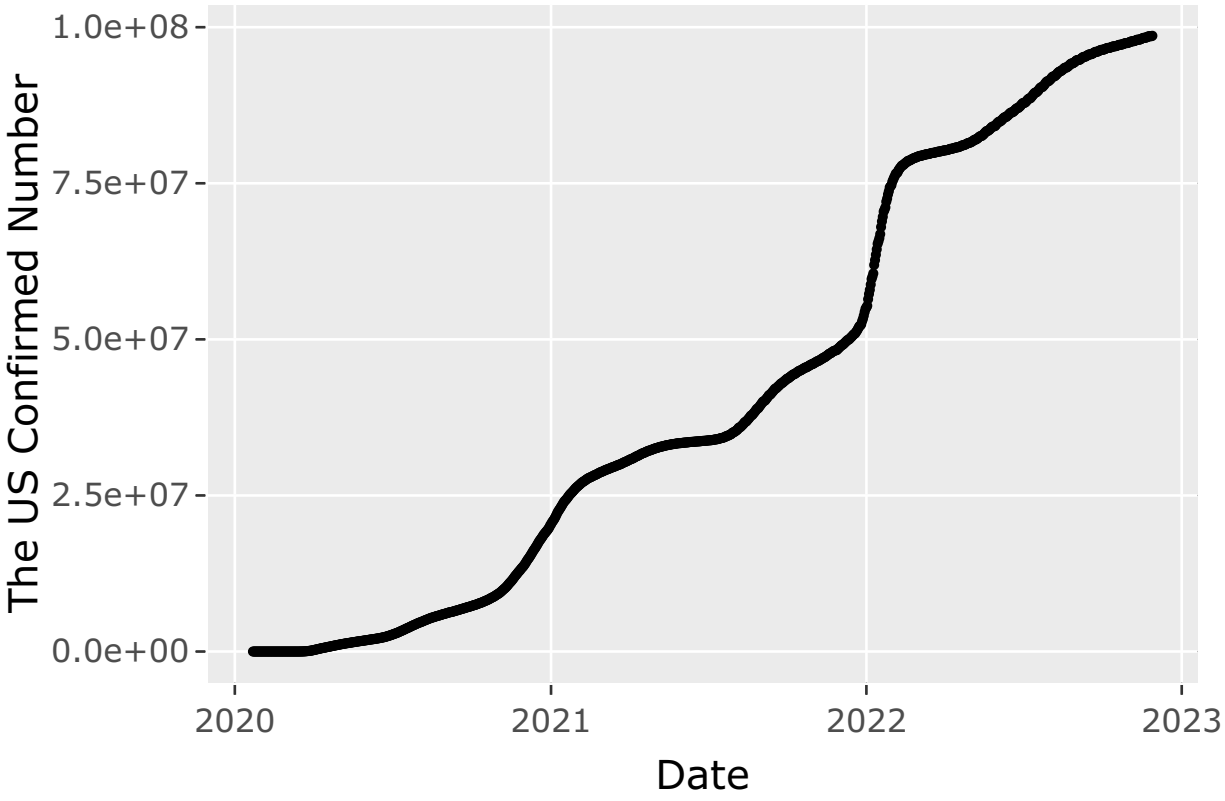
We can also check and sort which are the top countries have the greatest or smallest cases number. Corresponding with the previous heat map, the top five countries have the greatest cases number are US, India and France and Germany

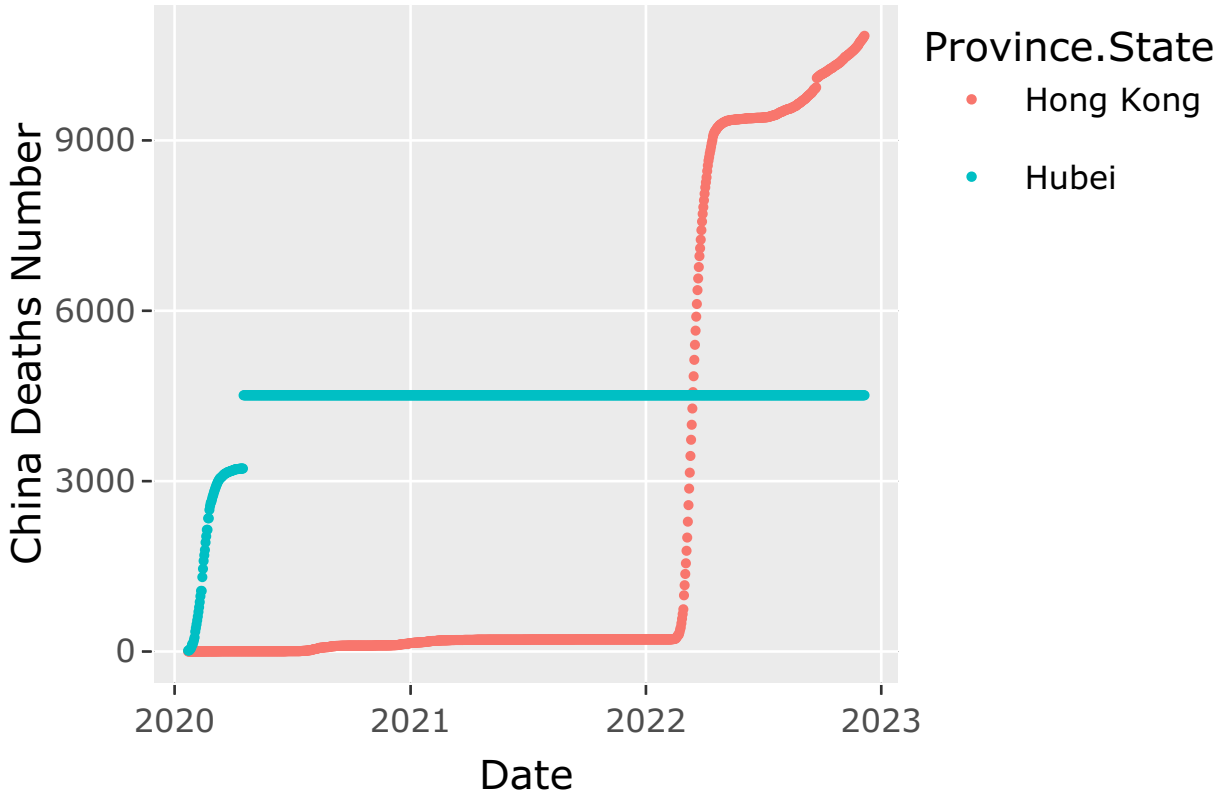
5. Transform the total death dataset, confirmed and deaths data of these two countries into dataframes

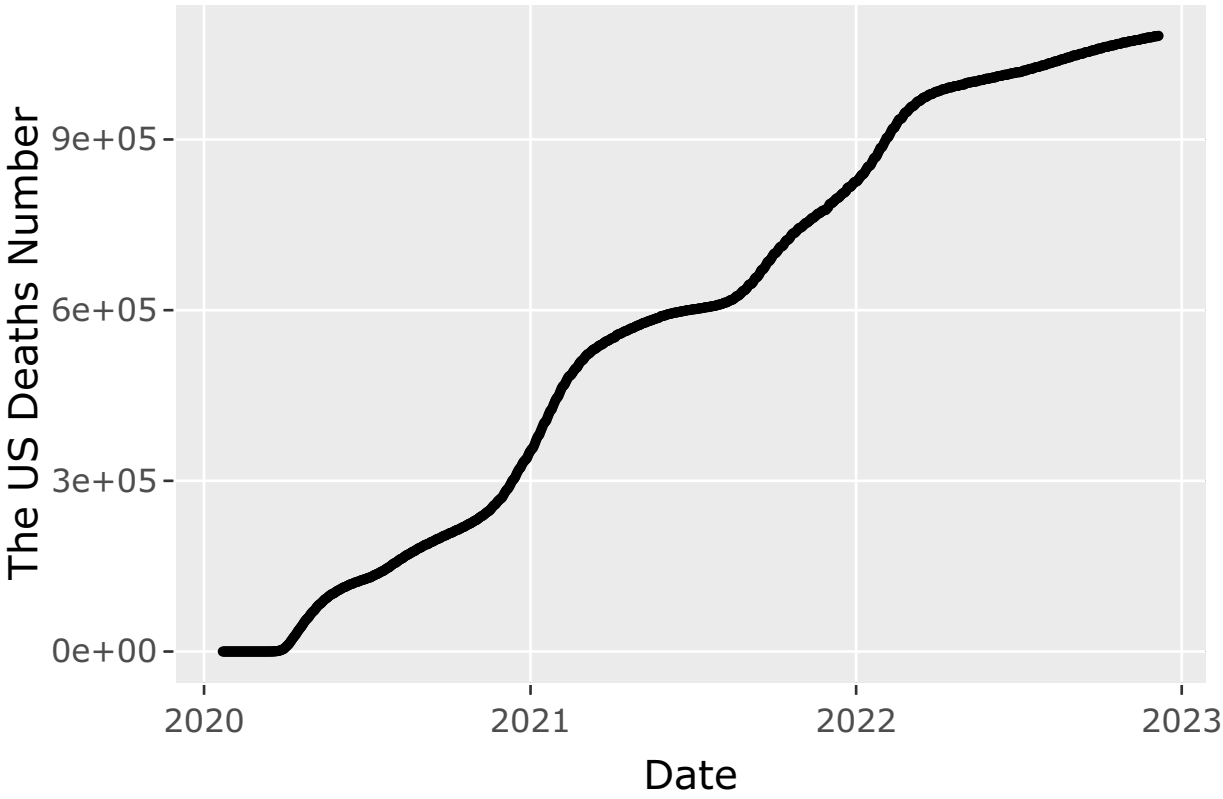
6.graphs of each province in china and US confirmed and deaths cases.

As I mentioned above, we only have 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, so 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 China Confirmed Number
2000000
1500000
1000000
500000
0
2020
2021
2022
2023
Date
Province.State
Beijing
Guangdong
Hong Kong
Hubei
Jilin
Shanghai

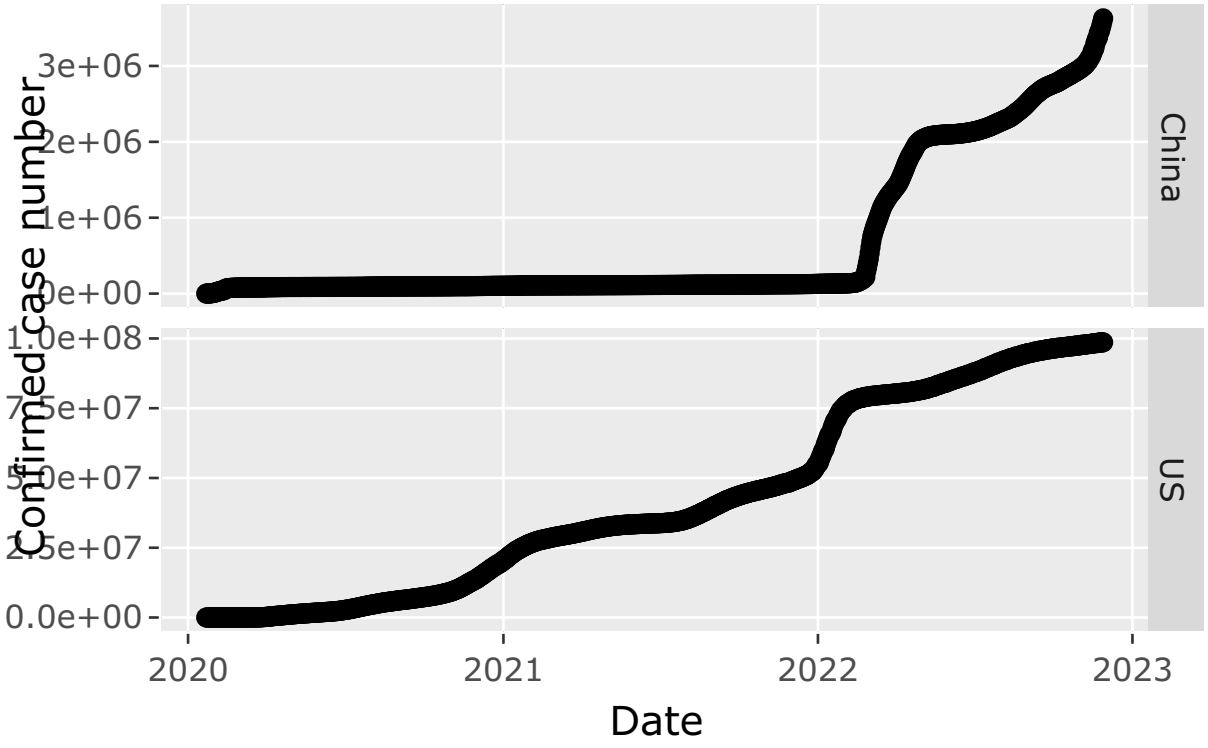

1.0e+08
7.5e+07
5.0e+07
2.5e+07
0.0e+00
The US Confirmed Number
2020
2021
2022
2023
Date

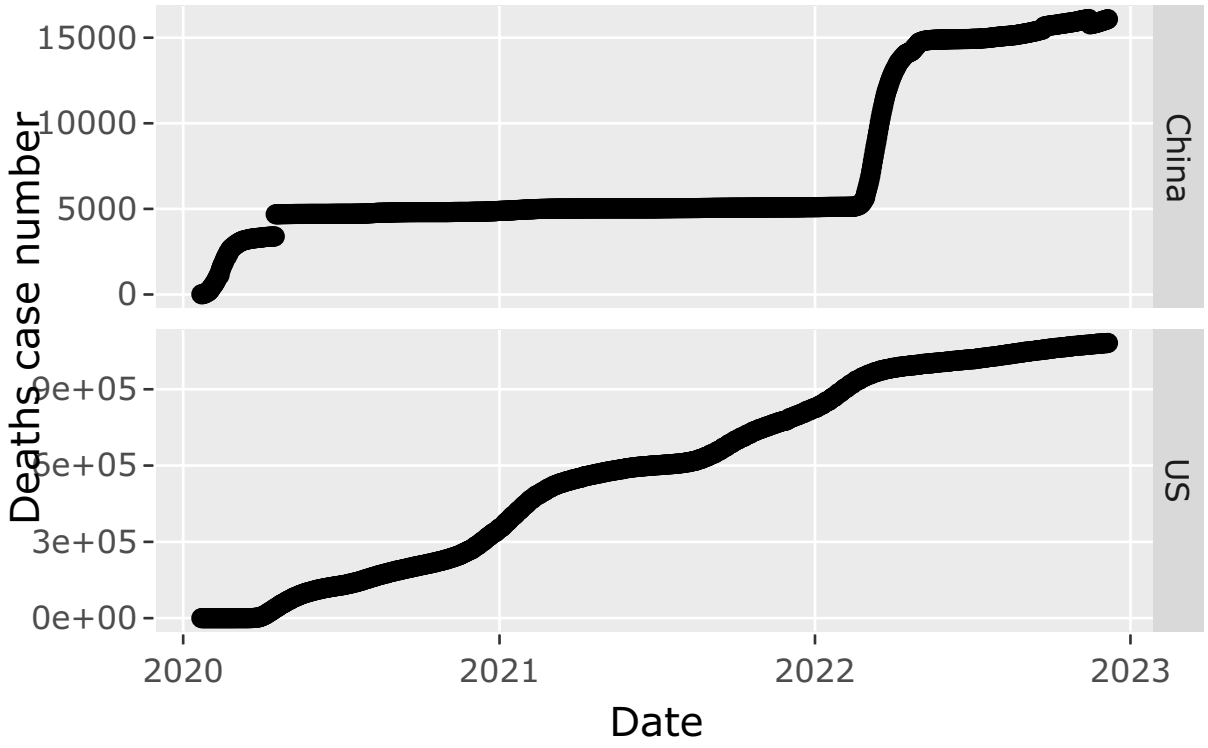

China Deaths Number
9000
6000
3000
0
2020
2021
2022
2023
Date
Province.State
Hong Kong
Hubei



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 most COVID-19 confirmed cases according to the graphs, and it is the only region with more than 5000 death cases.

7. let's make those two contries in comparision





From these two pairs of comparision, 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.

Conclusion and Summary

We can see China started confirming cases ealier 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 (Until 11/29/2022, deaths cases is 1.09M, from Wikipedia) are significantly higher than China (until 11/29/2022 deaths cases is 5233, from Wikipedia). 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