# The World COVID-19 Pandemic Report In 2020

Langyu Qie

November 5, 2020


## Abstract

Coronavirus disease 2019, i.e. COVID-19, is a contagious respiratory and vascular disease, caused by severe acute respiratory syndrome coronavirus 2(SARS-CoV-2). The first known human infections were in Wuhan, Hubei, China. The earliest date of onset of symptoms is reported to be December 1st ,2019. By December 2019, the spread of infection was almost driven by human-to-human transmission.

COVID-19 mainly spreads through the air when people are near each other long enough, primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or speaks. It can spread as early as two days before infected persons show symptoms (presymptomatic), and from asymptomatic (no symptoms) individuals. People remain infectious for up to ten days in moderate cases, and two weeks in severe cases.

We will look at the how the pandemic changes in 2020.


# Introduction

The dataset I use is "covid_19_data.csv". This data set contains 116,806 observations, showing the COVID-19 epidemic status of every country around world from January 22nd, 2020 to September 23rd, 2020. This data set counts confirmed cases, deaths, and recovered cases in every province(/state) in every country(/region) by day. In this case, we can use graphs to see how the epidemic in every country change over this period. We will also calculate the case fatality rate (CFR) of COVID-19 around world through time, which has been the statistic that was cared most to the public. It is defined as:

$$\frac{number\ of\ deaths\ from\ the\ disease}{number\ of\ diagnosed\ cases\ of\ the\ disease}$$

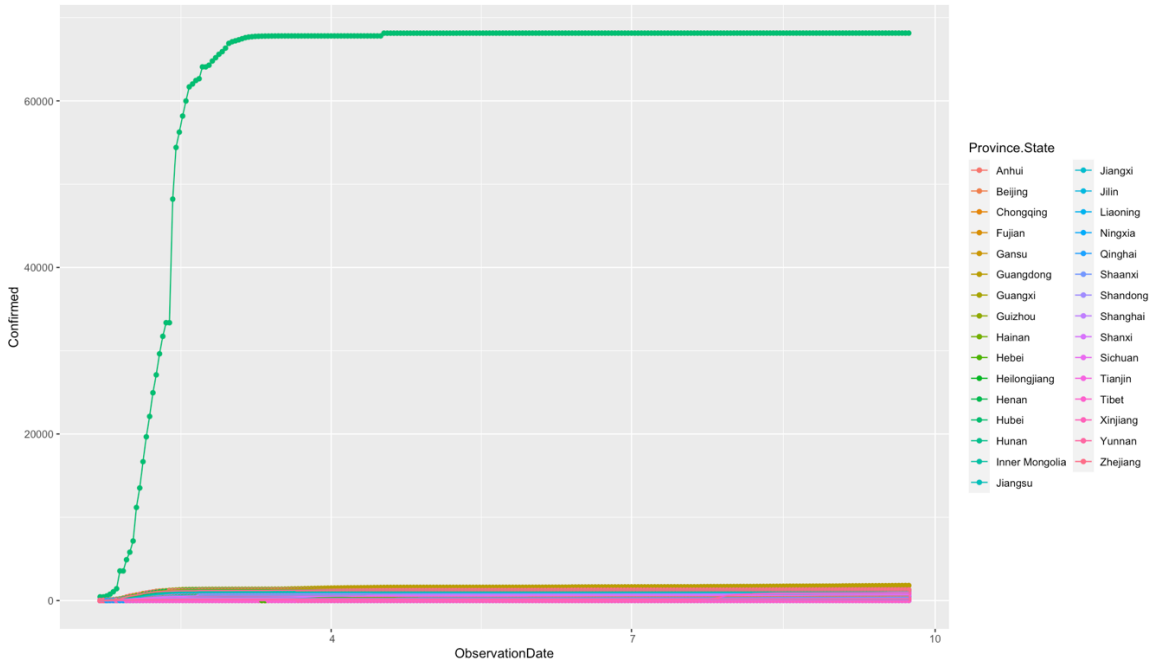
Origin:

https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset?select=covid_19_data.csv

# Data visualization

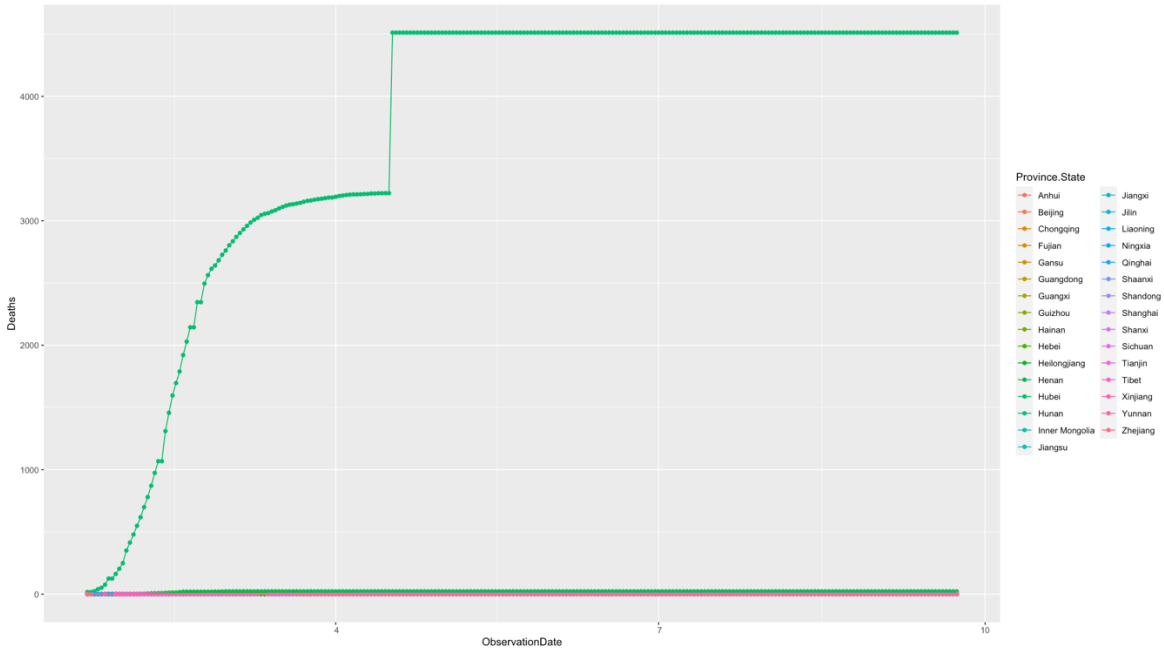
## China

As the first known human infection appeared in China, we first look at how the number of cases changes in China in 2020. Firstly, the confirmed cases in China:



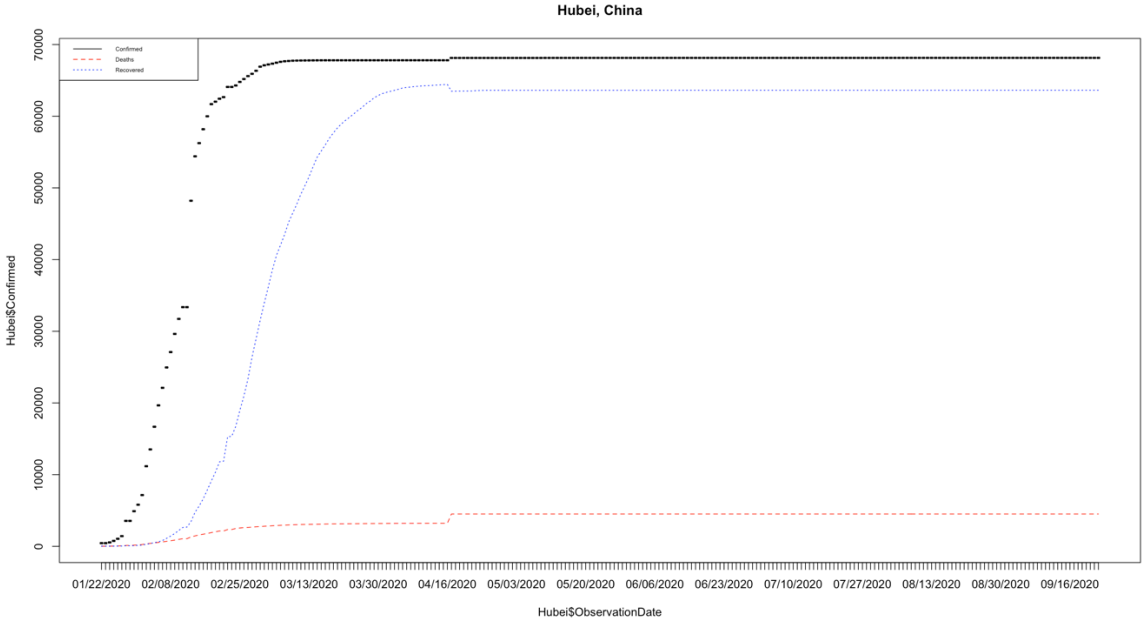

From the line graph we see that the line of Hubei province really stands out. The number has a sharp rise in February, and it kept increased to nearly 70,000. Which is said, almost 70,000 people are diagnosed positive for COVID-19 in Hubei in March. We also find that the number of confirmed cases in Hubei remained relatively stable from March to September, and it has been under 70,000 the whole time.

Now we look at the number of deaths in China during this period:




It's clear that number of deaths in provinces other than Hubei are trivial. The number of deaths in Hubei has been increasing since January, and it has gone beyond 3,000 in March. There is a jump of the number in April, and it went to approximately 4,500 from less than 3,300 all of a sudden. It stays at approximately this number from then till September.

From the two plots above we found that there have been no major outbreaks in other provinces of China. So, we want to look at cases only in Hubei:

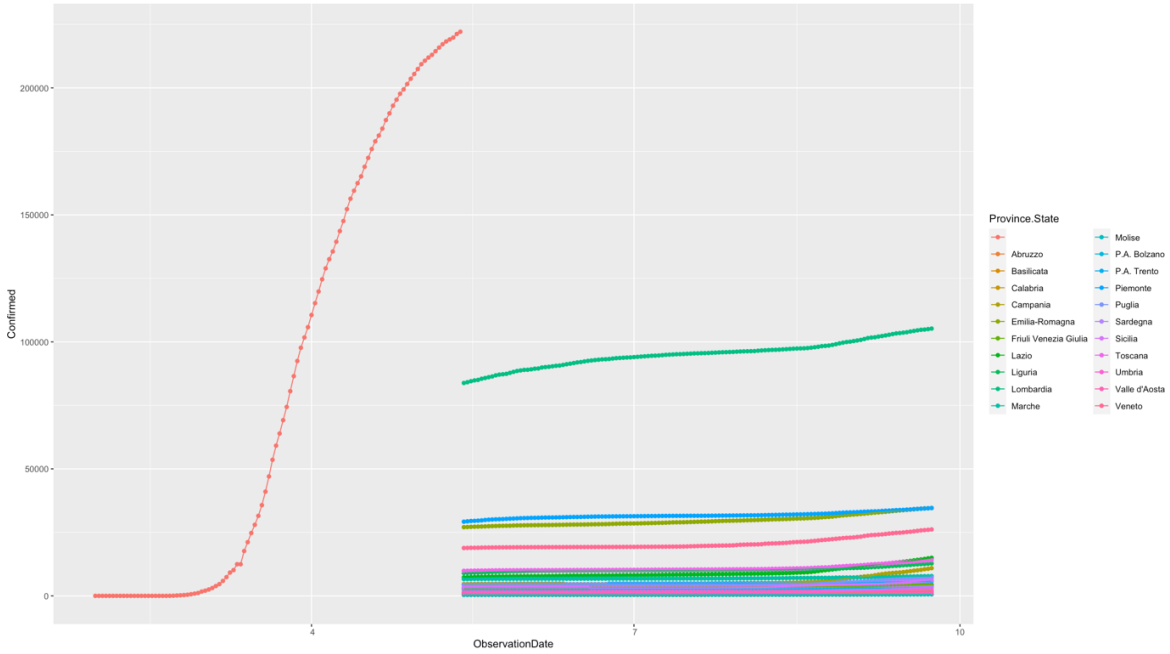



Evidently, from the graph above we see that Hubei has done a great job in treating patients diagnosed positive for COVID-19, most of the patients have recovered in September. Since

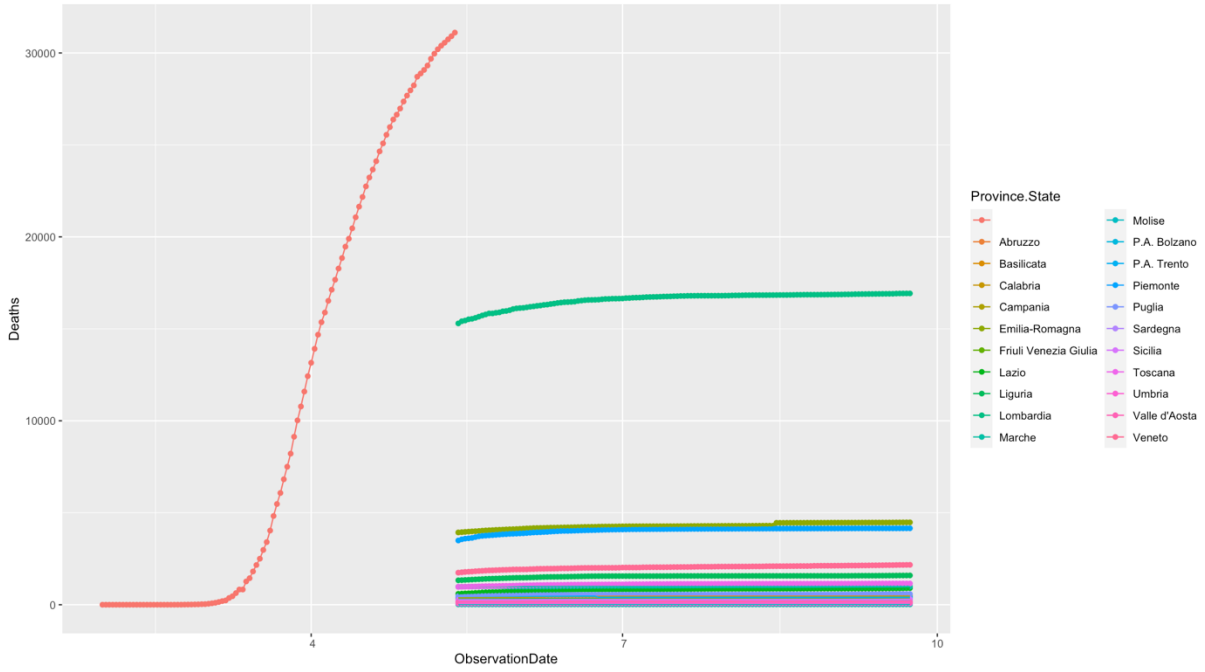
the confirmed cases aren't increasing, we can boldly conclude that the COVID-19 spread has been under control in Hubei.

## Italy

Since COVID-19 started to spread man-to-man, Italy has been the worst-hit country in Europe. Therefore, we now look at confirmed cases in each province across Italy during this period.

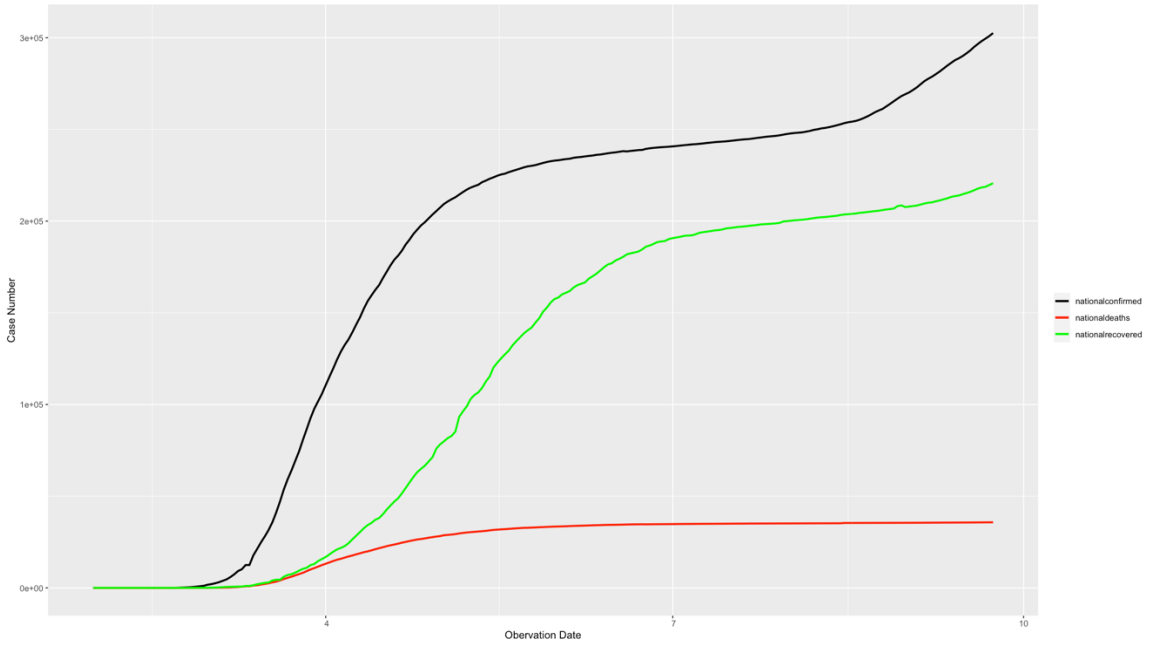



From the plot we can see that province data are lost before May, however, the number of confirmed cases kept climbing since February. It has increased to around 225,000 in May. After the confirmed cases are reported with province data, we see that among all the provinces in Italy, Lombardia has the highest number of confirmed cases, around 100,000. Other provinces all have confirmed cases under 50,000. Now we look at deaths across Italy:




Deaths in Italy started from several days after confirmed cases appear in Italy. With the number of confirmed cases going up from February to May, deaths also became more and more. Number of deaths has reached 30,000 in May. Among all the provinces, Lombardia has the highest number of deaths, which is consistent with it having the largest number of confirmed cases.

Now we look at the number of all confirmed cases, deaths and recovered cases across Italy over these 8 months:

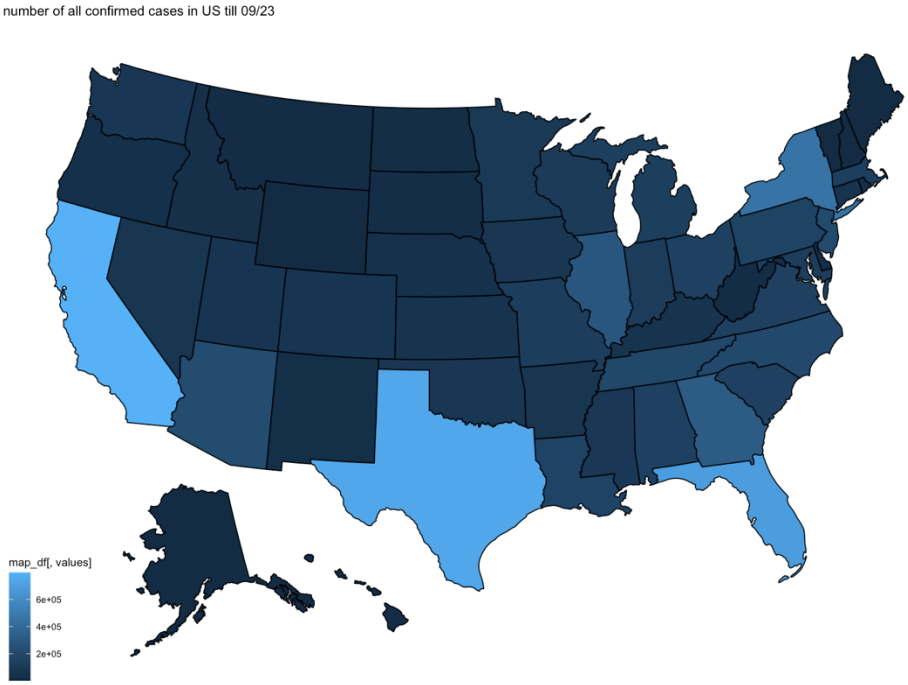



Now we look at time series case count in Italy. From the plot we see that, although recovered cases kept going up, number of confirmed cases has also been increasing. Then we look at the growing rates of number of confirmed cases and recovered cases, we see that confirmed cases has been growing faster than recovered cases in most of the time. They both became slower

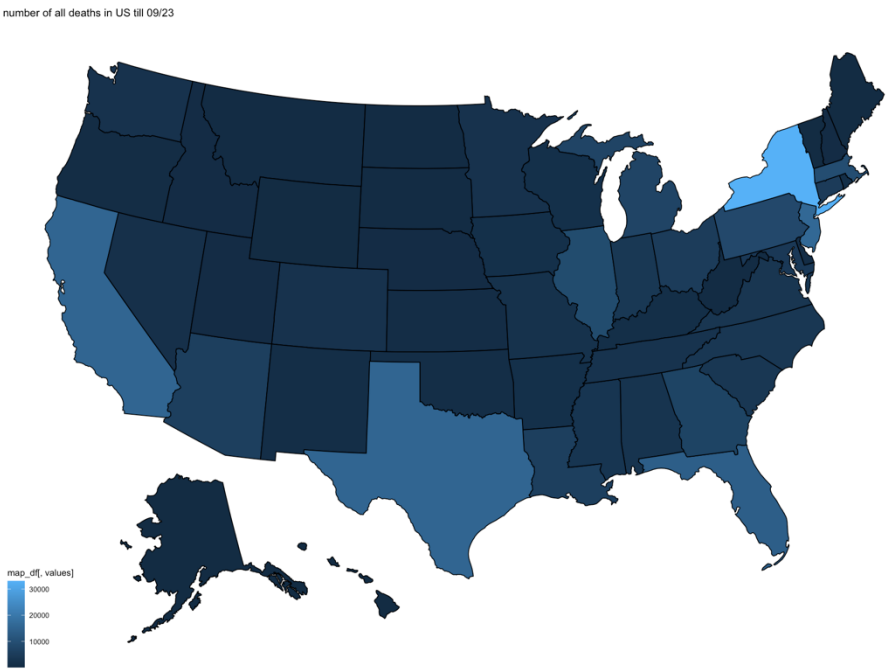
between June and July for a short time. The number of confirmed cases is still growing on September 23rd, and the rate of growing is even faster than it was in August. The rate of growing in confirmed cases is faster than the rate of growing in recovered cases. Hence, we can conclude that number of patients still suffering from COVID-19 disease (diagnosed but not cured) will still be growing after September in Italy.

## US

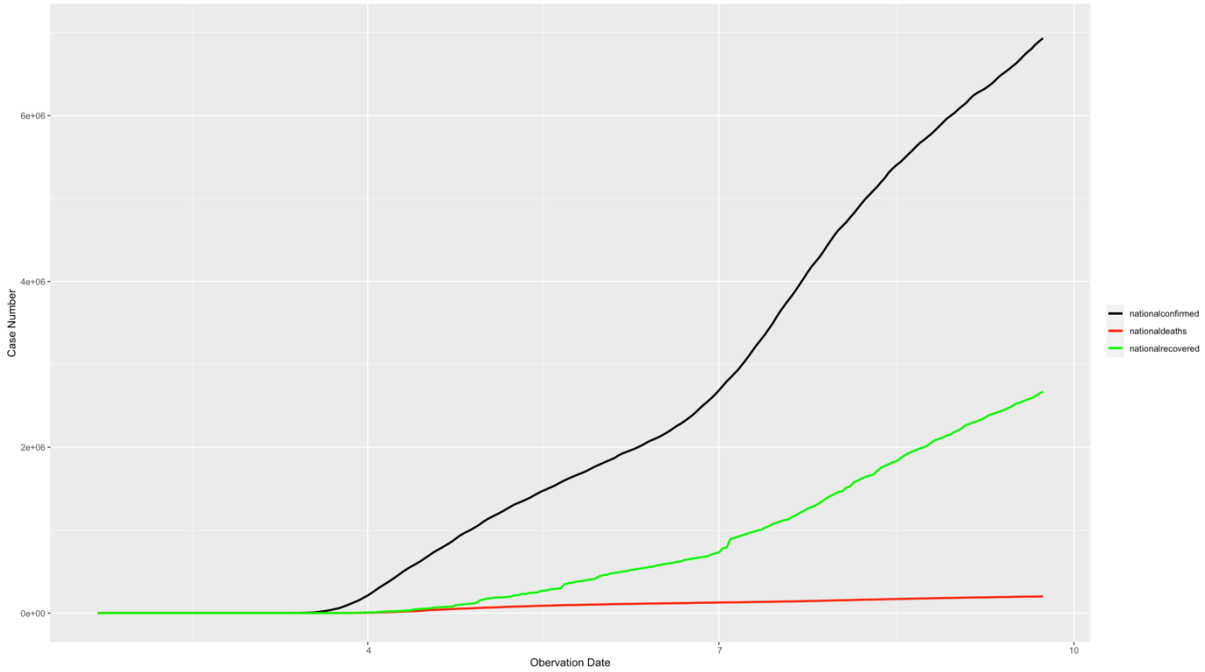
First, we take a look at the number of confirmed cases across US at the end of these 8 months in 2020.




We can see from the map that California took the worst hit (796,436 confirmed). Taxes, Florida and New York also have confirmed cases more than 400,000 (742,913 in Texas, 690,499 in Florida, 451,892 in New York). Then we look at number of deaths across the country in 8 months in 2020:




Although California and Taxes have more confirmed cases than New York does, they have less deaths than New York does. New York has more than 30,000 deaths over 8 months (33,090 to be exact). California, Texas, Florida have deaths over 10,000 (15,291 in California, 15,372 in Texas, 13,618 in Florida). Now we look at how the number of all confirmed cases, deaths and recovered cases across US had been changing over 8 months:

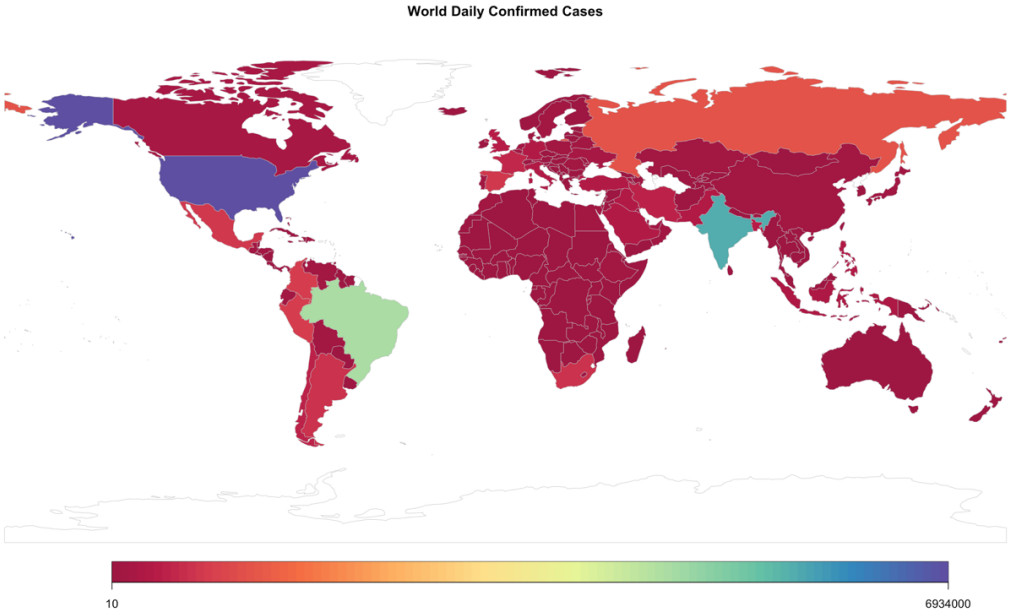



It's clear that the number of confirmed cases has been increasing from May to September, however, the increasing rate didn't change too much. The rate appears to go a little slower in June, but it went right back up in July. For the number of recovered， when we compare it to the number of confirmed cases, it doesn't seem optimistic. When it came to September, the number of recovered cases across US has not reach half of the number of confirmed cases

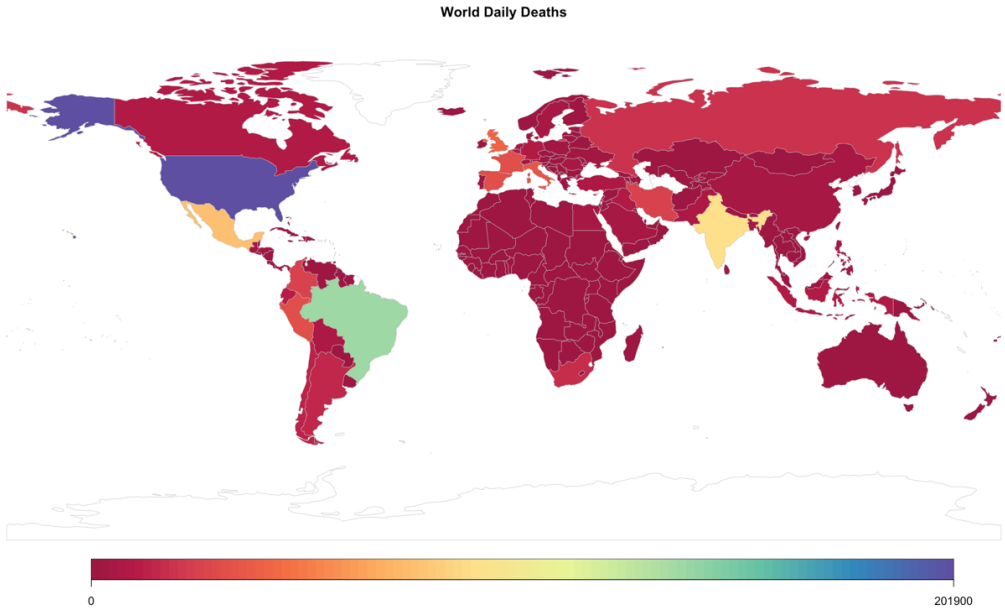
(6,933,548 confirmed, 2,670,256 recovered). Up till September, 201,884 people are reported dead from COVID-19.

## Cases around the world in 2020

The following map shows the number of confirmed cases around the world from January 22nd to September 23rd. In this map we can see the contrast of number of confirmed cases in countries all over the world:




We can see that US has the most confirmed cases (6,933,548 cases) among all the countries in the world. India has the second most confirmed cases (5,646,010 cases) in the world and Brazil has the third most confirmed cases in the world (4,591,364 cases). Then we look at number of deaths around world:

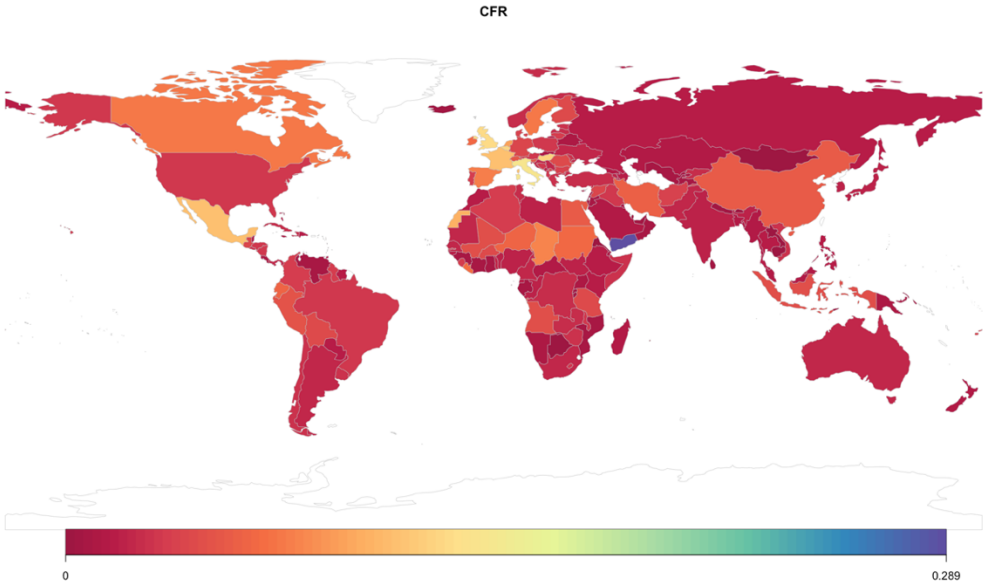



From the map we can see that US has the most deaths around the world. Brazil comes the second (138,105 deaths) and India comes the third (90,020 deaths). Mexico also has a considerably number of deaths (74,949 deaths).

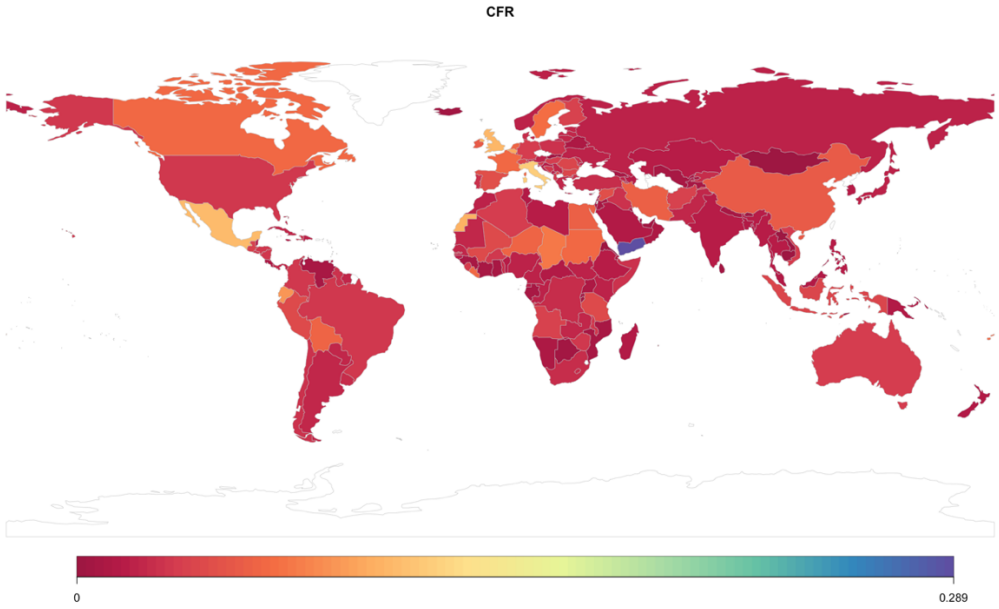
## CFR

From the definition of CFR, we know that CFR is not a constant. It changes with time and locations. However, we are not going to show how CFR changes by day due to the errors in measurements and low quantities in sample in the first half of year. Instead, we are going to choose two days and calculate the CFR in each country in each day. With acknowledgement of these severe limitations, we continue to use `deaths/confirmed` as a very rough proxy of CFR.

First, we choose the date: August 23rd. We calculate CFR on this day in each country and present them in the following map:




Then we choose date: September 23rd, and calculate CFR in each country on this day, present in map:




From comparison of two maps above, we found that there are no prominent distinctions between two maps. This suggests that CFR in each country has become stable in the last two months. Also, From the color palette we see that only a few countries, for example, Yemen, have CFR that's conspicuously higher than other countries in the world. Of course, CFR is also affected by factors like age, diabetes, obesity and so on. Differences in proportion of these factors in different areas also lead to regional disparity.

Although there are countries having CFR extremely low, we can't lower our guard for the risk of dying from COVID-19. In most countries, 2%~3% confirmed cases of COVID-19 are dead.

# Conclusion

COVID-19 is a deadly virus that has already taken so much human lives all over the world. After it's first discovered in China, China has done a great job in containing the virus. However, since confirmed cases started to appear in other countries, especially US, people tend to give less concern about the virus. People are not strict with the stay-home order and still not wearing face mask in public places, which aggravates the transmission of the virus from human to human. Now US has become the country that suffers from the most human deaths due to COVID-19 around the world. This also led to a great hit on economics and job offer shortage in US.