

# STAT7008 Project Report

## Understanding COVID-19 Pandemic

Name: Lyu Baiyang

UID: 3035678222

### Introduction

Since December 2019, some hospitals in Wuhan City, Hubei Province have found multiple cases of unexplained pneumonia with a history of exposure to the seafood market in South China. It has now been confirmed as an acute respiratory infection caused by Corona Virus Disease (COVID-19) pandemic.

This disease is still spreading in some countries and has caused tens of thousands of deaths. This study aims to find the underlying pattern of the spread of COVID-19.

### Data Source and Data Processing

Here we get the data from WHO COVID-19 (<https://covid19.who.int/>) using Web scrapping method.

Input page:

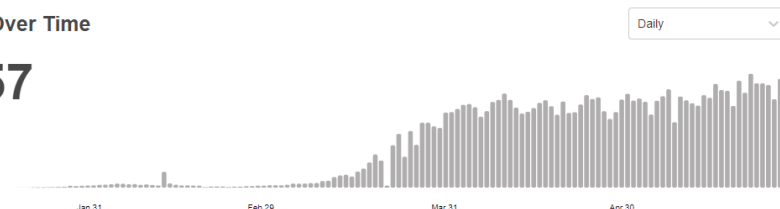
Globally, as of 9:37am CEST, 30 May 2020, there have been 5,796,257 confirmed cases of COVID-19, including 362,483 deaths, reported to WHO.

#### Confirmed Cases Over Time

5,796,257

confirmed cases

Source: World Health Organization

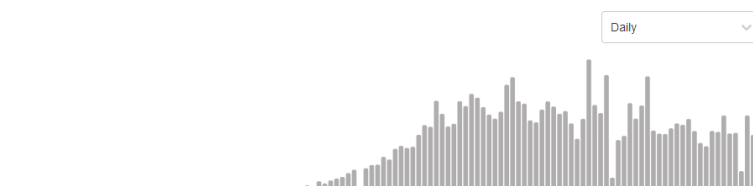


#### Deaths Over Time

362,483

deaths

Source: World Health Organization



After fetching the web page, we can find that each row consists of 7 variables: TimeStamp, Country, Region, Deaths, Cumulative Deaths, Confirmed Cases, Cumulative Confirmed Cases.

Here is an example:

```
[[1582502400000, 'AF', 'EMRO', 0, 0, 1, 1],
 [1582588800000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1582675200000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1582761600000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1582848000000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1582934400000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1583020800000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1583107200000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1583193600000, 'AF', 'EMRO', 0, 0, 0, 1],
 [1583280000000, 'AF', 'EMRO', 0, 0, 0, 1]]
```

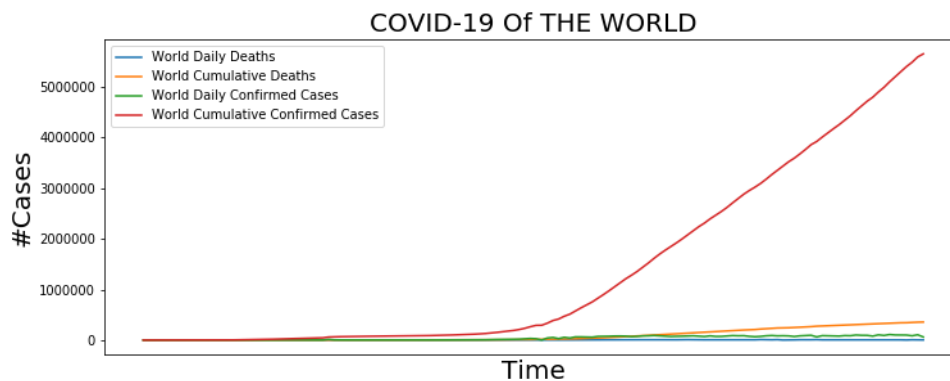
Then we need to translate the Country 2-character code into Country Name. So we get the iso-3166-1 code from web (<http://doc.chacuo.net/iso-3166-1>):

	code2	code3	Name
74	GL	GRL	Greenland
6	AM	ARM	Armenia

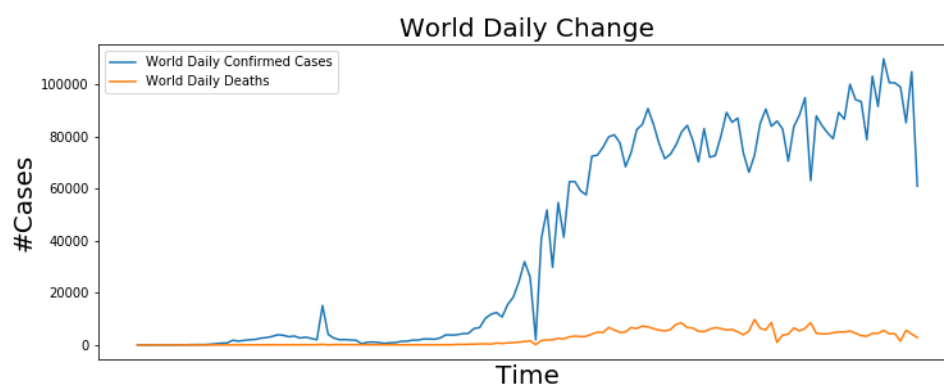
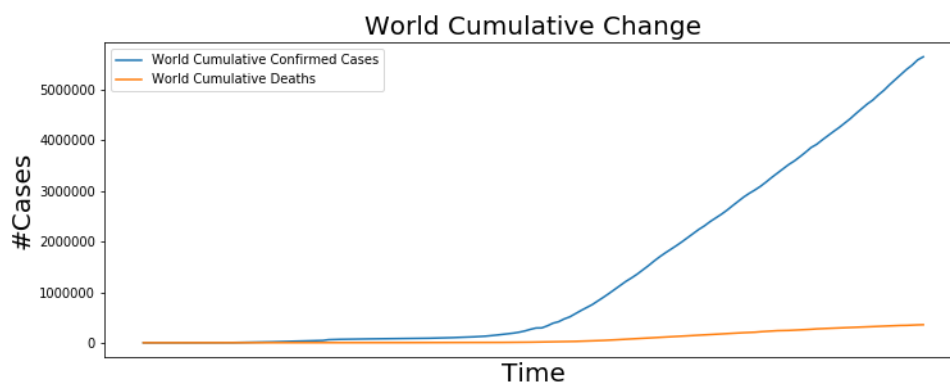
With a base date, we can also translate the timestamp to date (e.g. 2020-2.24). After processing, we get data like this:

	Country	Region	Date	Death	Cumulative Death	Confirmed	Cumulative Confirmed
0	Afghanistan	EMRO	2020-02-24	0	0	1	1
1	Afghanistan	EMRO	2020-02-25	0	0	0	1
2	Afghanistan	EMRO	2020-02-26	0	0	0	1
3	Afghanistan	EMRO	2020-02-27	0	0	0	1
4	Afghanistan	EMRO	2020-02-28	0	0	0	1
...	...	...	...	...	...	...	...
17844	Zimbabwe	AFRO	2020-05-25	0	4	0	56
17845	Zimbabwe	AFRO	2020-05-26	0	4	0	56
17846	Zimbabwe	AFRO	2020-05-27	0	4	0	56
17847	Zimbabwe	AFRO	2020-05-28	0	4	76	132
17848	Zimbabwe	AFRO	2020-05-29	0	4	0	132

## Analysis 1: COVID-19 in the whole world



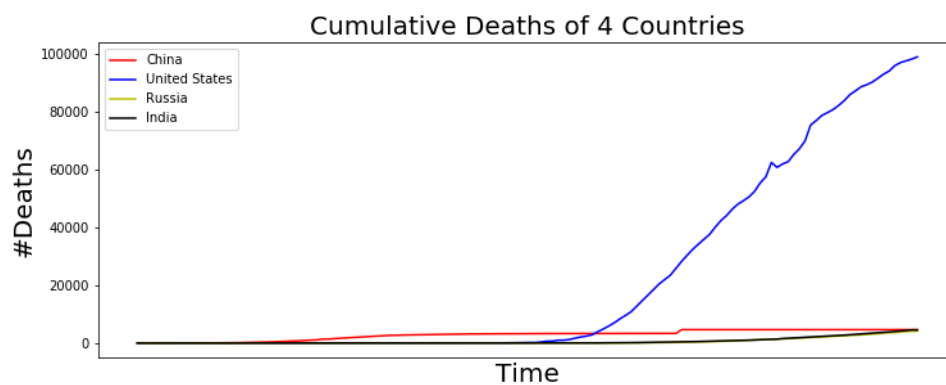
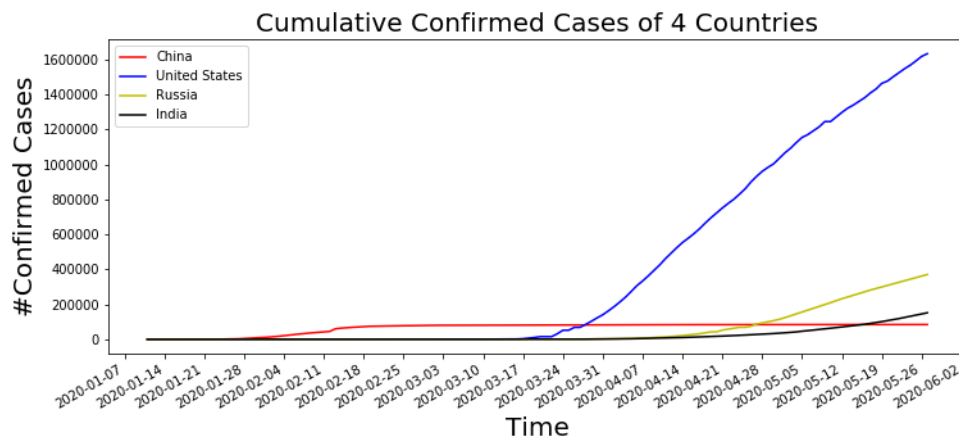
As the cumulative confirmed cases is much higher than other indicators, so we divide it into 2 graphs.



We can see that deaths and confirmed cases have similar trends. At first, it only spreads on a small scale. When the disease spreads to another country, the number of infections is increasing rapidly. There are nearly 5 million confirmed cases and 100 thousands deaths around the world until May 27<sup>th</sup>.

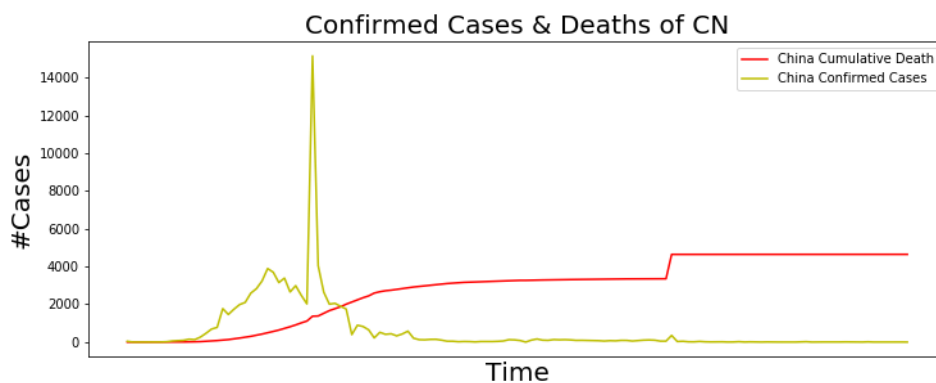
## Analysis 2: Comparison between countries

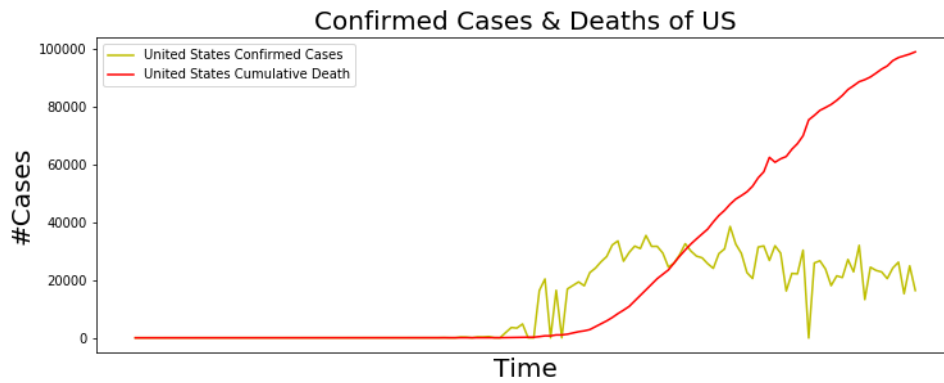
Then we compare the infections of the four major powers: China, US, Russia and India.



We can see that US has more confirmed cases and deaths than other three countries.

Specially, we compare China and US:





Here we can find two very different patterns. In China, there is an explosion of confirmed cases early, but deaths continued to increase day after day. But China took strong anti-epidemic measures to stop the infection promptly, so we can see the cumulative deaths didn't increase till now.

US has a total different pattern. It didn't have too many confirmed cases in January or February, but as the disease spread into US, we can find that the epidemic is not well controlled. The US still has many new confirmed cases everyday.

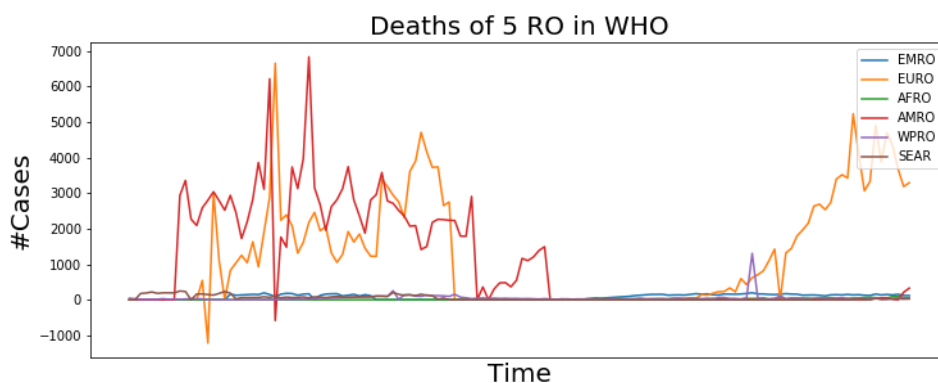
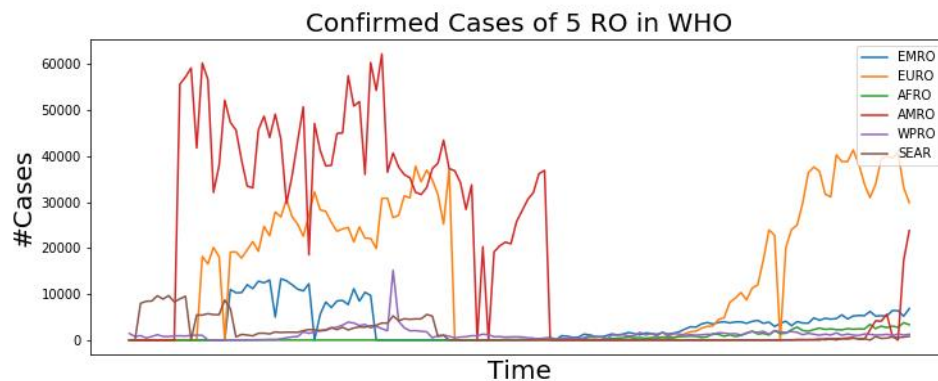
### Analysis 3: Comparison between 6 regions

There are 6 region organizations:

```
[ 'EMRO', 'EURO', 'AFRO', 'AMRO', 'WPRO', 'SEARO' ]
```

They are short for Eastern Mediterranean, Europe, Africa, America, West Pacific, Southeast Asia.

So we now compare these 6 regions and find the underlying pattern:



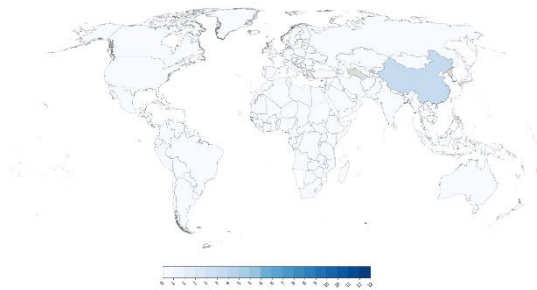
Here I need to explain that the negative cases are correction of misdiagnosis. We can find EURO and AMRO has serious situation in the first 2 months. Then the global seems to control the epidemic successfully. But the disease becomes widespread again in EURO in May.

So we can come to a conclusion, countries in EURO ignore the disease prevention which caused lots of deaths.

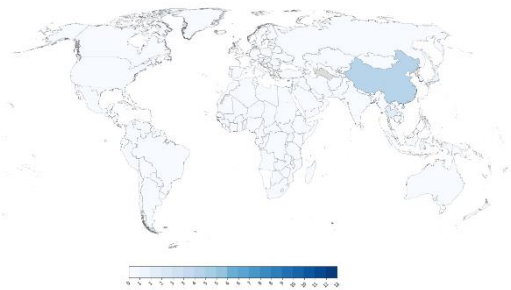
## Analysis 4: Choropleth Map from January to May

Here I make choropleth maps per week from January 12<sup>th</sup> to May 27<sup>th</sup>.  
Confirmed Cases:

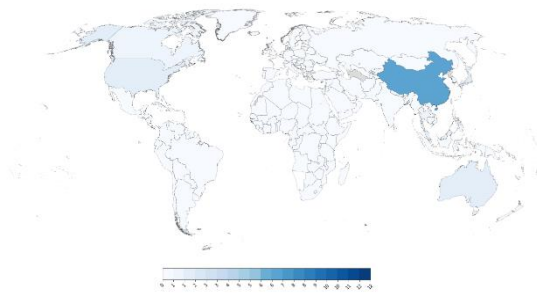
Confirmed on 2020-01-12



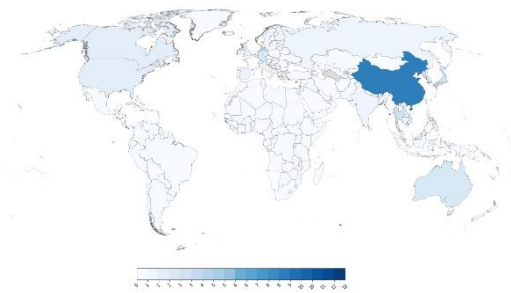
Confirmed on 2020-01-19



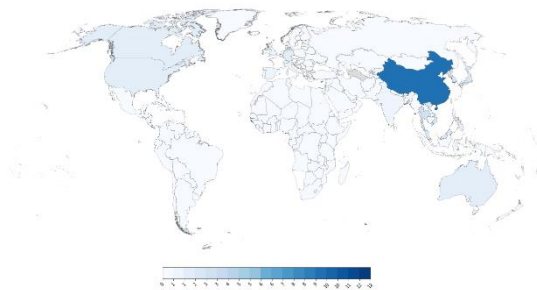
Confirmed on 2020-01-26



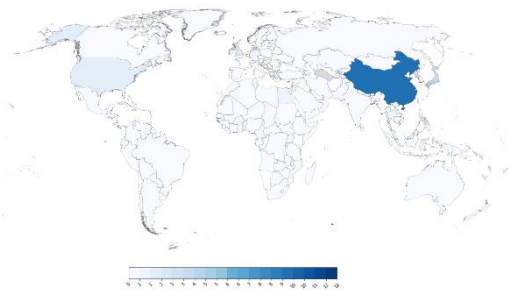
Confirmed on 2020-02-02



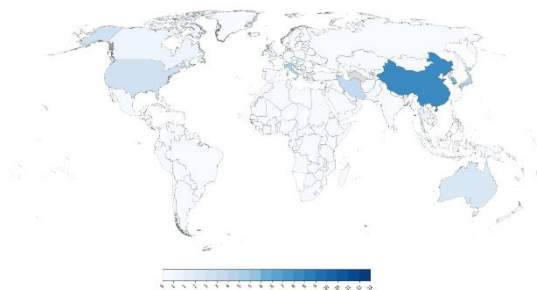
Confirmed on 2020-02-09



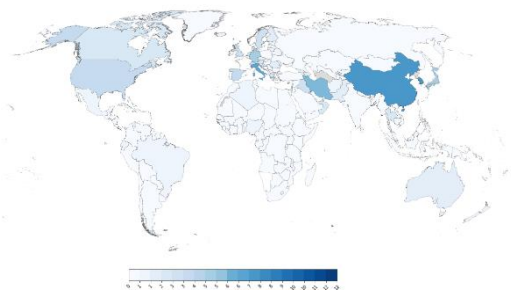
Confirmed on 2020-02-16



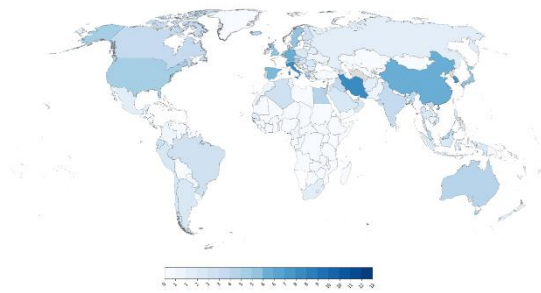
Confirmed on 2020-02-23



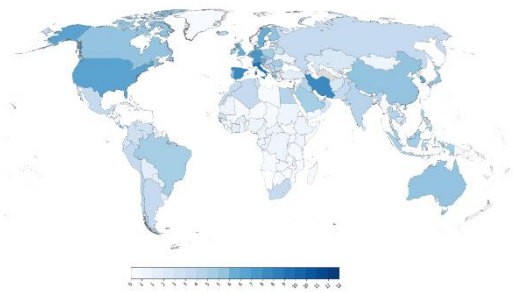
Confirmed on 2020-03-01



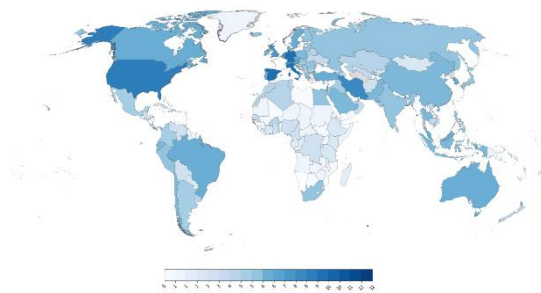
Confirmed on 2020-03-08



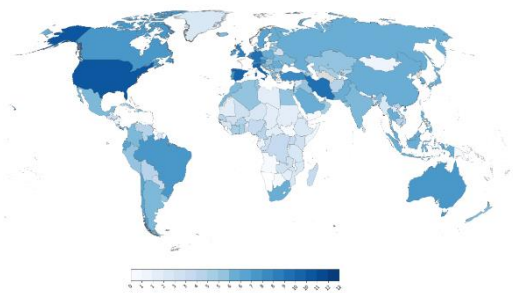
Confirmed on 2020-03-15



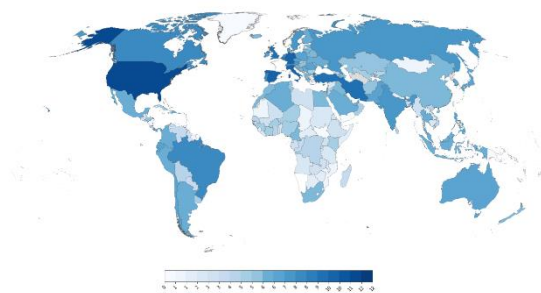
Confirmed on 2020-03-22



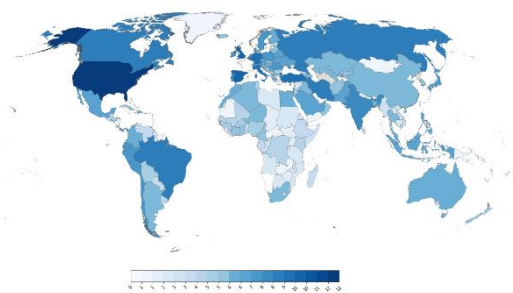
Confirmed on 2020-03-29



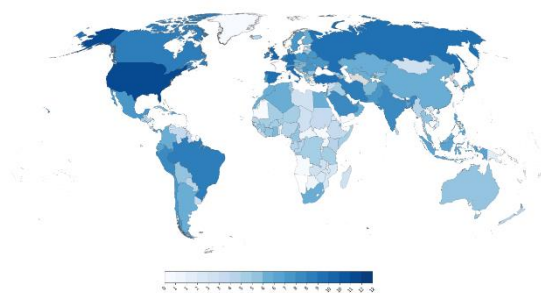
Confirmed on 2020-04-05



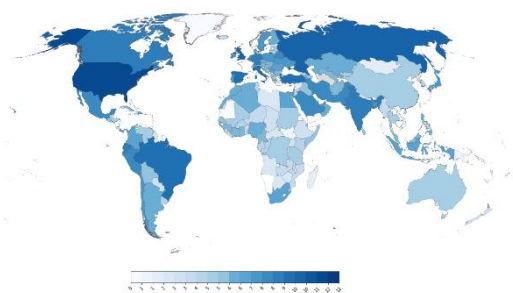
Confirmed on 2020-04-12



Confirmed on 2020-04-19

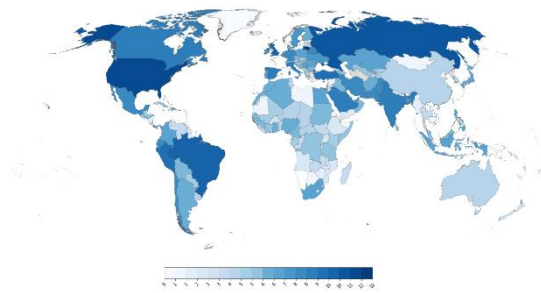


Confirmed on 2020-04-26

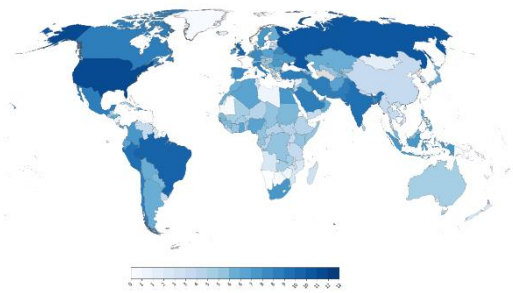




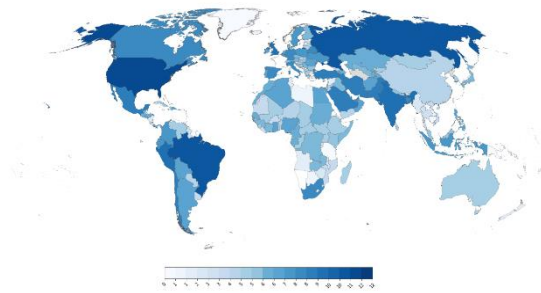
Confirmed on 2020-05-03



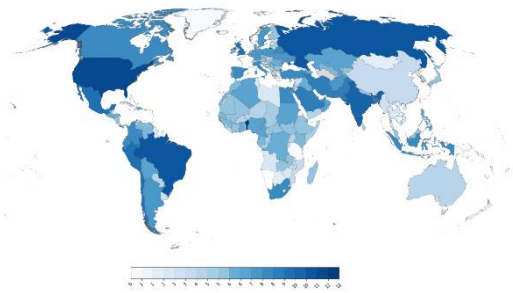
Confirmed on 2020-05-10



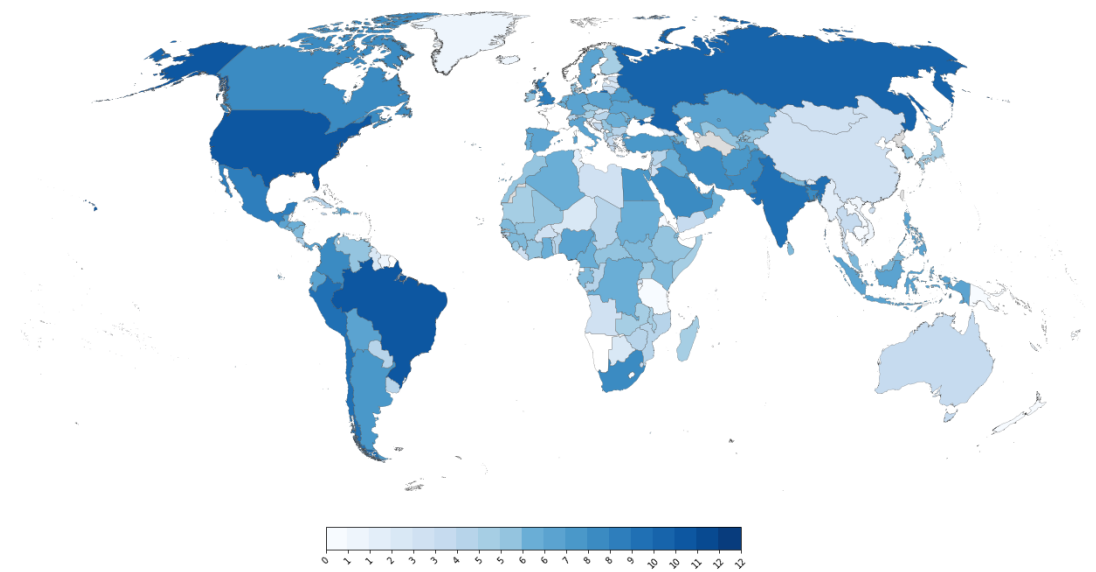
Confirmed on 2020-05-17



Confirmed on 2020-05-24

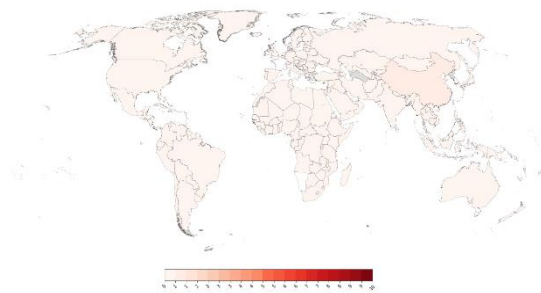


Confirmed on 2020-05-31

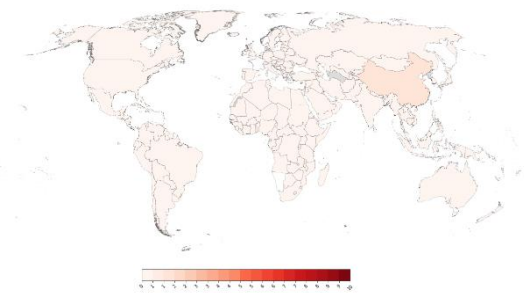


Death Cases:

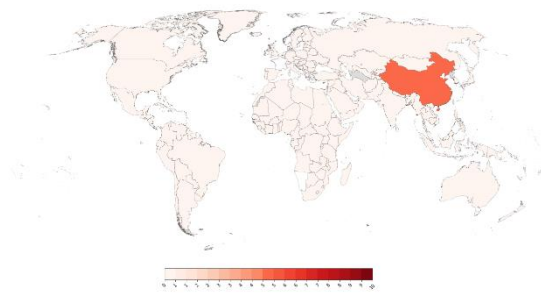
Deaths on 2020-01-12



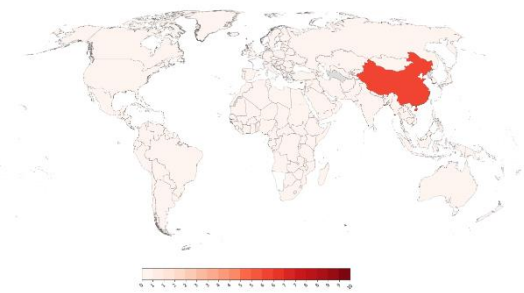
Deaths on 2020-01-19



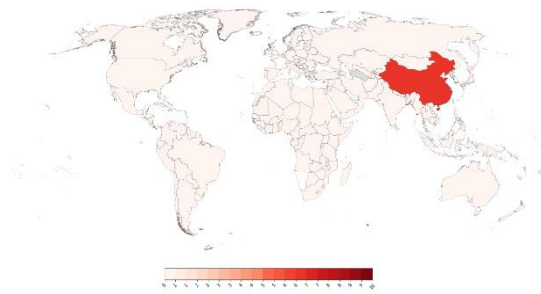
Deaths on 2020-02-02



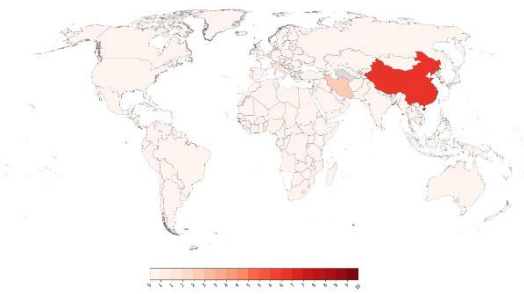
Deaths on 2020-02-09



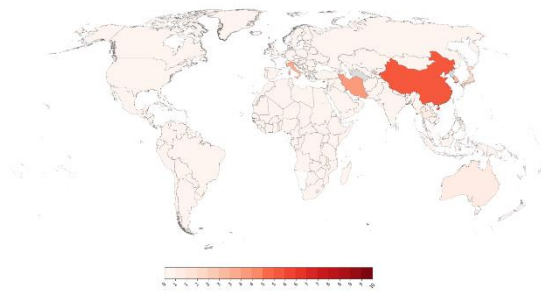
Deaths on 2020-02-16



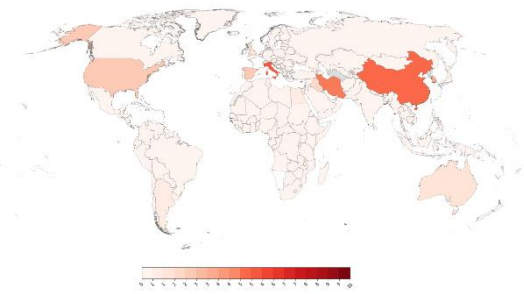
Deaths on 2020-02-23



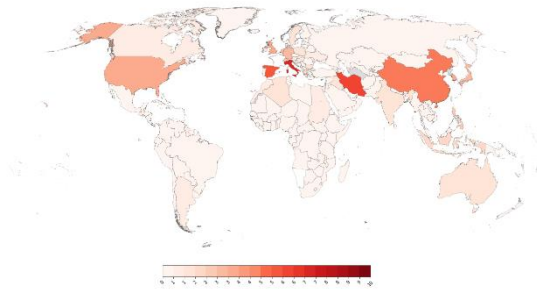
Deaths on 2020-03-01



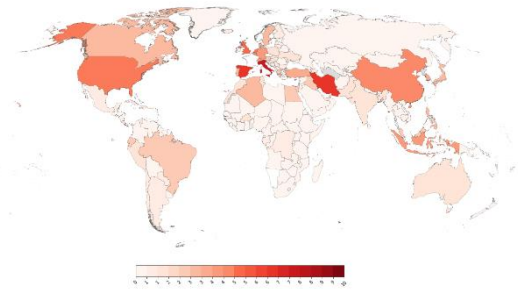
Deaths on 2020-03-08



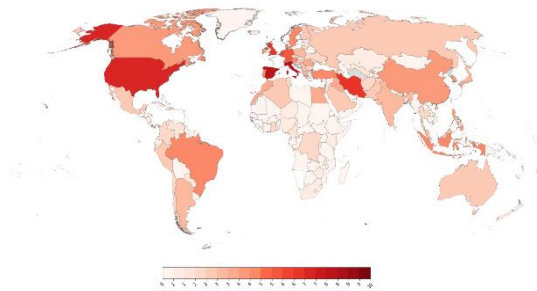
Deaths on 2020-03-15



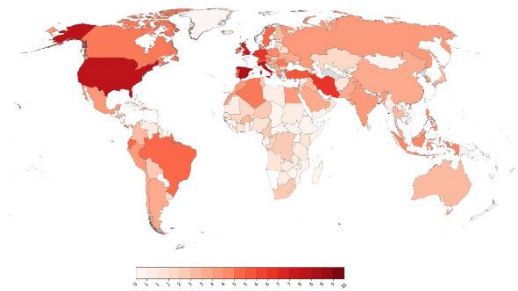
Deaths on 2020-03-22



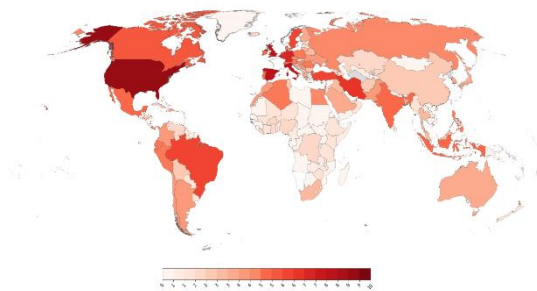
Deaths on 2020-03-29



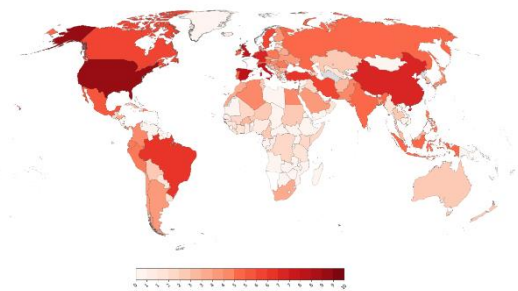
Deaths on 2020-04-05



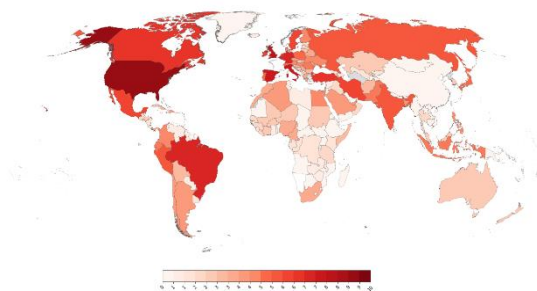
Deaths on 2020-04-12



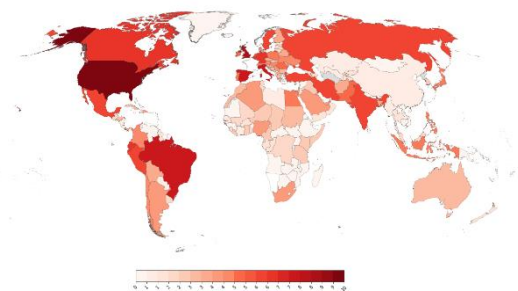
Deaths on 2020-04-19



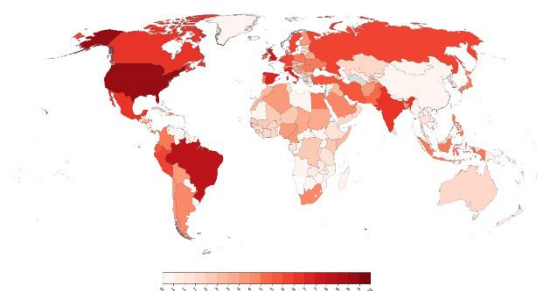
Deaths on 2020-04-26



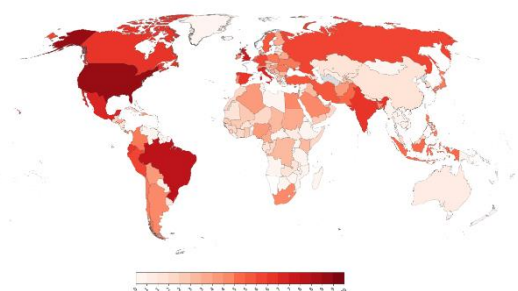
Deaths on 2020-05-03

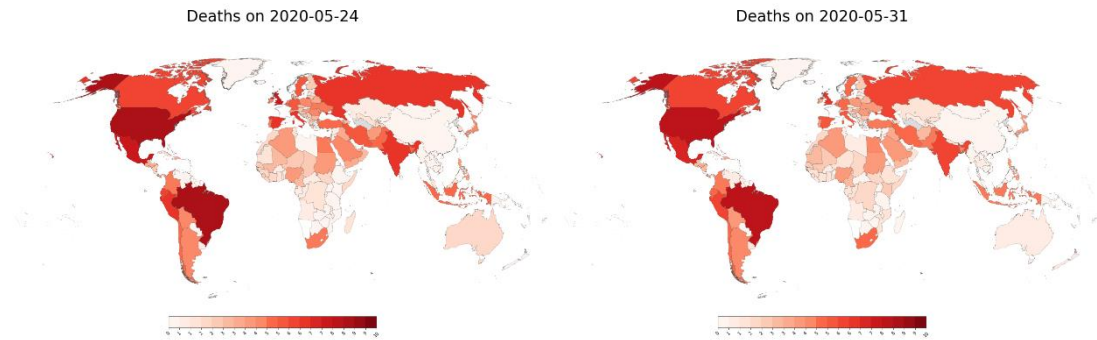


Deaths on 2020-05-10



Deaths on 2020-05-17





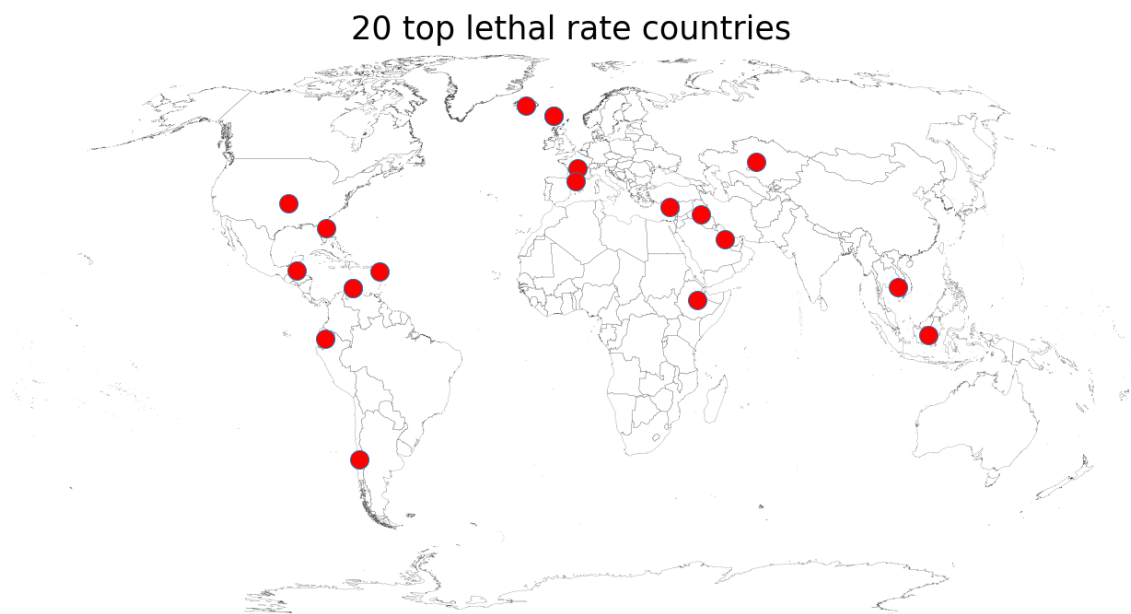
From the maps above, we can find that China has won a lot of time for the world. Other countries could have take measures to actively prevent epidemic. However, it was their failure to neglect this disease that caused tens of thousands death.

## Analysis 5: Top 20 lethal rate countries

We then find the top lethal rate countries, the top 5 countries are: If we

```
['Kazakhstan', 'Cambodia', 'Aruba', 'Cyprus', 'Iraq']
```

If we plot them in a world map:



## Conclusion

This study aims to do the visualization of epidemic data around the world. We can find that COVID-19 outbreaked in China in January. However, other countries, especially European countries, neglected this disease which caused tens of thousands death around the world. They have a wrong understanding of the virus.

## Appendix

Key code to do the visualization is submitted along with this report as .ipynb file.