COVID-19 Total Cases and Total Death Trends In a Few Countries

Chandra Adhikari

Abstract

Since its first appearance in a city of China in late 2019, the entire world is suffering from the COVID-19 outbreak. Using data available in "Coronavirus Source Data - Our World in Data," this study focuses on the total cases and the death cases in a few countries. A few python based packages such as Numpy, pandas, matplotlib, etc. are used for linear algebra, data processing, and plotting purposes. The total cases and the death cases due to COVID-19 follow a similar pattern. We use a Gaussian fit to observe number of cases per day in each of the countries of our selection. The Gaussian fit is not perfect, but it is a good approximation in a country where lockdown was effective.

Motivation

Almost all the people are affected by the COVID-19 directly or indirectly. The COVID-19 is hitting hard on human health as well as the socioeconomic aspect of most of the countries in the world. Thus studying the trends of the total COVID-19 cases and the death it caused is a hot topic. I have downloaded the data of June 28th, 2020 (the day I began to work on the project) to see the number of cases and the death cases. An insight into the trend of the COVID-19 instances can be useful to anyone.

Dataset

The data has been extracted from the following source:

- ❖ https://covid.ourworldindata.org/data/ecdc/total_cases.csv

- ❖ https://covid.ourworldindata.org/data/ecdc/total_deaths.csv

To reduce the notation clash, the downloaded csv files are renamed as 'total_deaths_June28.csv' and 'total_cases_June28.csv' (see the attached Jupiter notebook).

Data Preparation and Cleaning

❖ COVID-19 did not start at the same time in all the countries. Thus I replace NaN values to zero using fillna(0).

❖ There are more than 200 countries in the data, I have chosen only eight countries to study, namely, China, France, Germany, Italy, Spain, United States, Brazil, and United Kingdom.

❖ Number of cases and death cases are not in the same order in all countries. Thus tick marks got to be scaled in each plot based on the number.

❖ Some records appear to be outliers. But no data has been modified or dropped (because it is too early to say about these outliers). I have presented the data and findings as I get from the data.

Research Questions

In this work I aim to answer the following questions:

❖ Does the total cases of the infected person and the death cases due to COVID-19 follow the same trend?

❖ Whether the number of cases per day can be fitted using some countries, in which it appears to be more control, to predict the same in other countries.

Methods

❖ Data processing and Visualization has been carried out using a few python based packages such as Numpy, pandas, matplotlib.

❖ Number of cases per day has been fitted using the following (Gaussian) fit function:

$$f(x, \mu, \sigma, c) = \frac{c}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where μ, σ, c are mean, standard deviation and a constant.

❖ Why Gaussian Fit?

⟹ Beacsue the data of China, where the COVID-19 appeared first, agrees very well with the Gaussian fit.

Findings: Total cases and total deaths over a time

- ❖ Both the total cases and the total deaths follow the similar trend.

- ❖ There is a sharp jump in the total deaths in China in mid April, 2020. (Apparently, record(s) from a forgotten source was added.)

- ❖ Deaths and infected cases due to the COVID-19 are apparently under control in China, it is towards the saturation in France and United Kingdom, however there is no sign of saturation in Brazil.
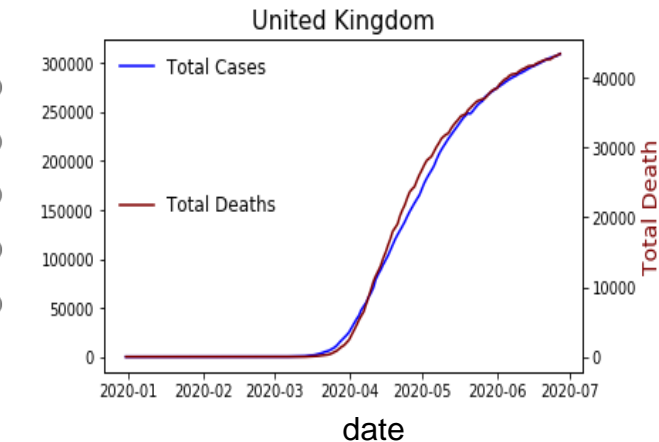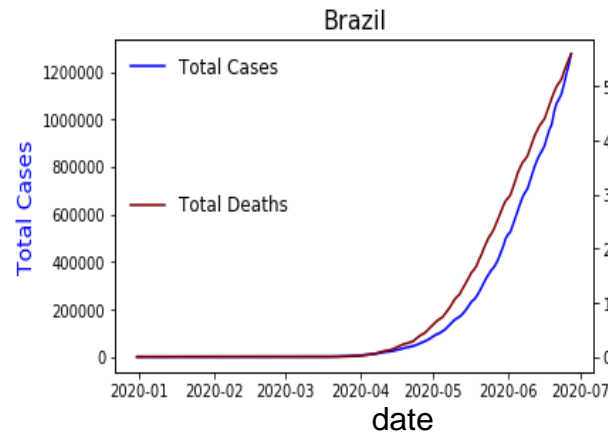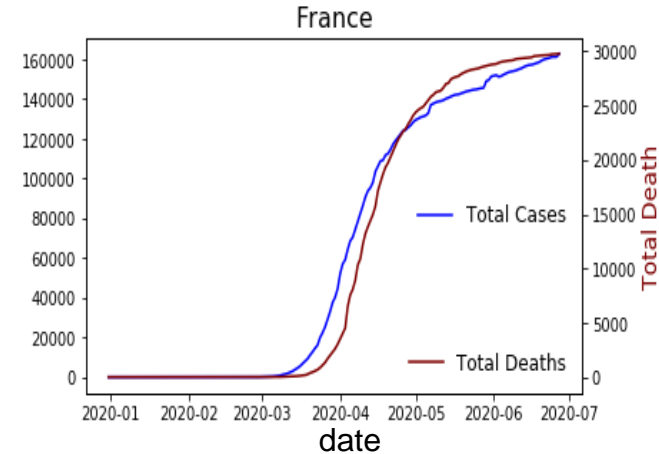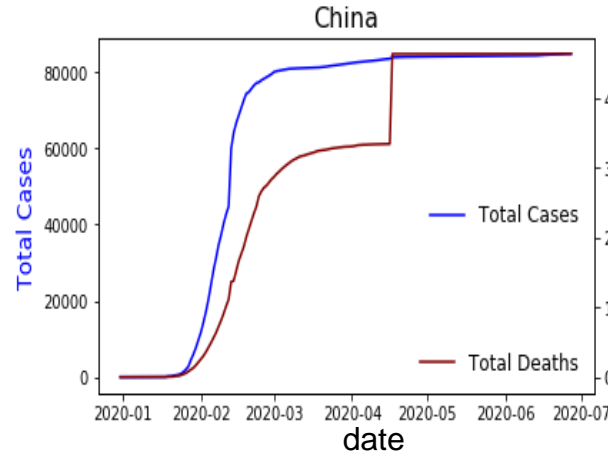
Findings: Total cases and total deaths over a time

❖ Both the total cases and the total deaths follow the similar trend as in the previous slide.

❖ Deaths and infected cases due to the COVID-19 are apparently toward the saturation in Germany, Italy, and Spain, however, it is too early to say about it in the case of the United States.

❖ Apparently, the total deaths got decreased for a month in Spain. This is because of a confusion whether to add a record from a source in the national record or not.
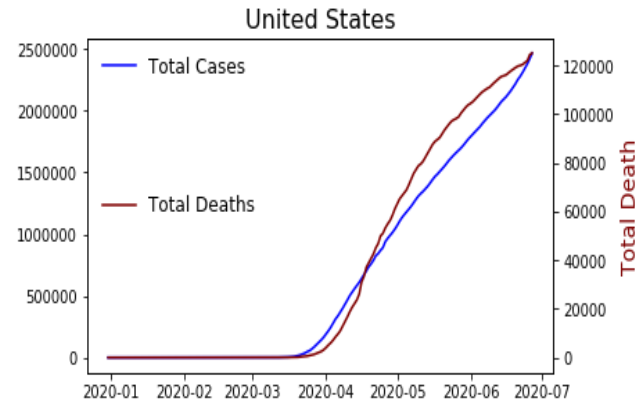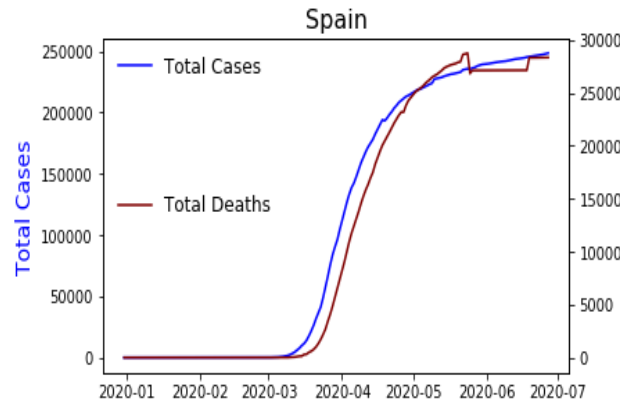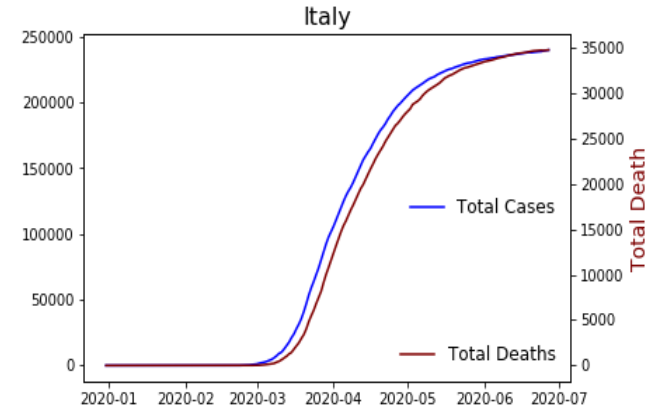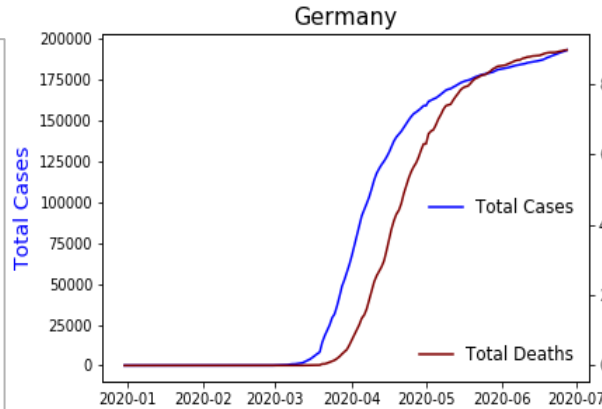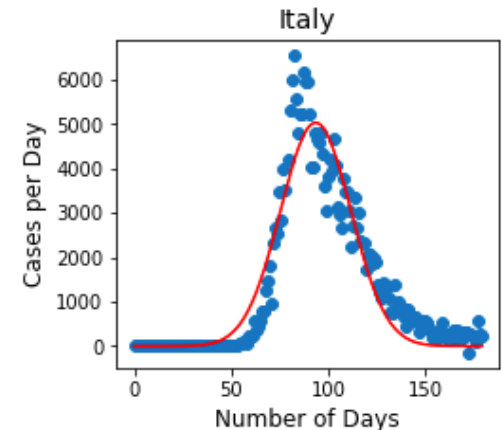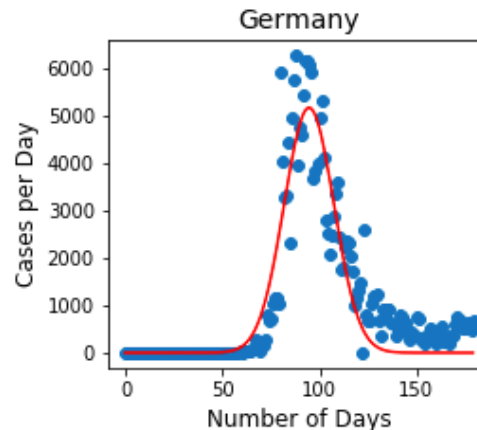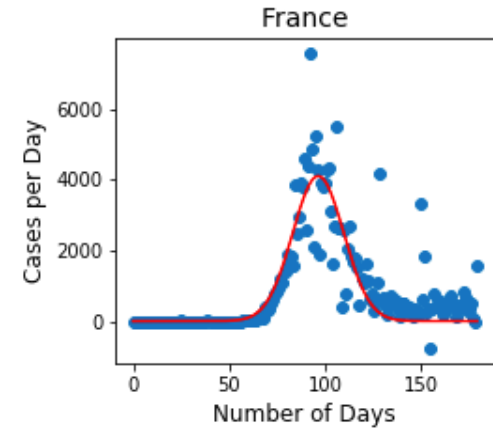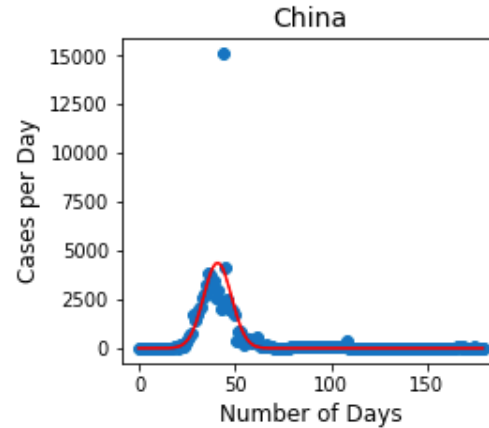
Findings: Number of Cases vs. Number of Days

❖ The figure shows how many days each of these eight countries took to reach a particular number of cases.

❖ Dec 31, 2020 is the first day of data recording.

❖ In China, it took just three days for total cases to reach 100, while for Brazil, it took 71 days. (please see the notebook for the other detail. I have written only the extrema.)

❖ The number of cases reached to 50,000 in 33 days in China while it took 94 in Brazil.

❖ The number of cases raised sharply in Spain, Italy, and Brazil than other five countries namely, China, France, Germany, United States and United Kingdom.
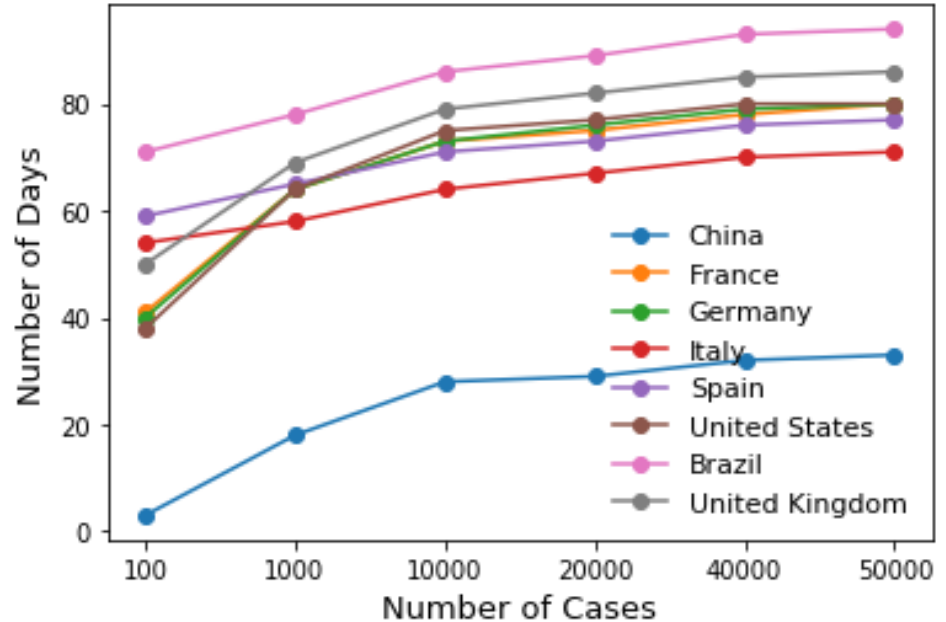
Findings: Cases per Day vs. Number of Days

- ❖ Big dots are the actual values, and the red curve is the fitted one.

- ❖ The number of cases per day versus the number of days agrees very well with the Gaussian fit for China's data, where the COVID-19 case first appeared. There is an outlier, which arose because a record of a different source for a period of time has been added to national count on that day.

- ❖ Although the fits are not perfect, the Gaussian fits agree pretty well for the data of Germany and Italy as well. The fit does not look bad in the case of France as well.

Findings

- ❖ The Gaussian fit agrees pretty well for the data of Spain and United Kingdom.

- ❖ It is too early to say about the fitting in the case of Brazil as the number of cases has not been reach in a peak.

- ❖ The plot for the United States shows that the United States has been hit already by the second wave. Thus, probably we got to go the a multiple peaks fitting function rather than a single peak Gaussian.
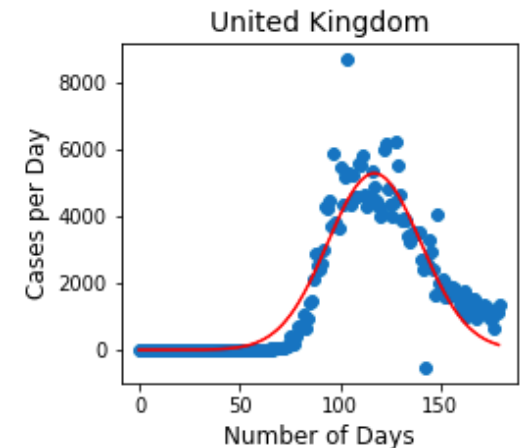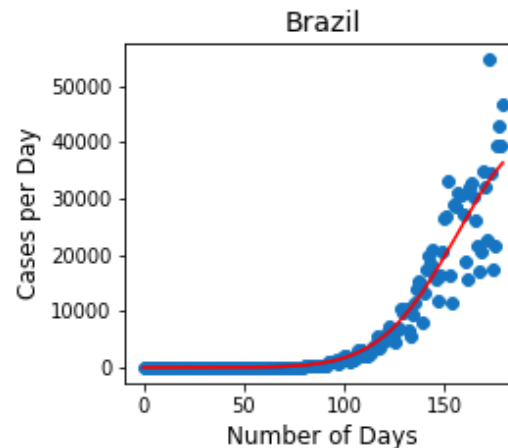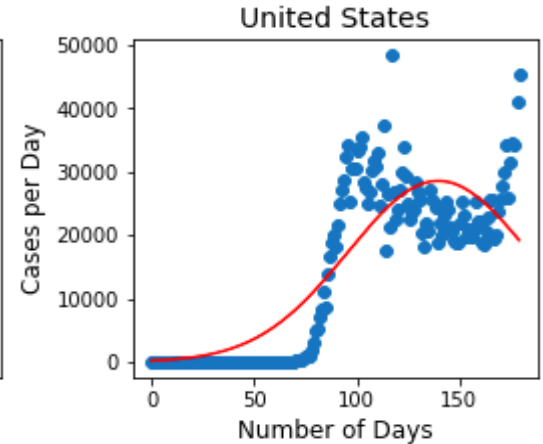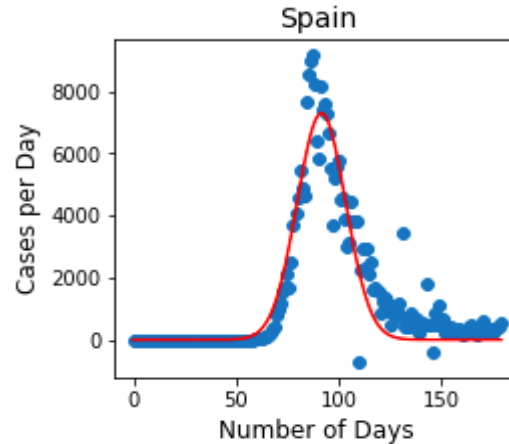
Limitations

❖ This study is essential, and this kind of study on COVID-19 may enlighten our mind for a definitive solution in the future. Still, the study suffers from a limitation that no clear conclusion can be drawn for a moment.

❖ A socioeconomic factor and a political belief may also lead to an outbreak. Thus, the findings based on a country may not work to the other country.

❖ The number of cases saturation also depends on the effectiveness of the lockdown in a country, which is mathematically abstract.

❖ The different viruses spread differently; thus, the approach we have followed here for COVID-19 may not apply to the other outbreak(s).

Conclusions

❖ The total cases of the infected person and the death cases due to COVID-19 follow a similar trend in all the eight countries.

❖ The number of cases per day can be fitted well for China as a function of the number of days. Taking China's data and Gaussian fit as a reference, the same analysis has been done for the other seven countries. The fitting works pretty well (although it is not perfect) for France, Germany, Italy, Spain, and the United Kingdom. Brazil is not in saturation yet, and the United States has already been hit by the second wave of COVID-19.

Acknowledgements

- I worked with the dataset myself downloading it from "Our World in Data" (https://ourworldindata.org/coronavirus-source-data).

- I acknowledge Hannah Ritchie of "Our World in Data", who has collected the data of COVID-19 and stated that anyone can download and work with the data.

- Some of my arguments, such as the cause of outlier in the data of China, a death count drop in Spain are based on the news I read, and I heard. There are many news sources I follow. I acknowledge them collectively.

Thank you for reading my work!