

COVID-19 Visualization with Python

Lisa Saunders - DAFT REM MAR21
April 29, 2021

1 The Data

This explanatory data analysis is based on an open access data set by Our World in Data¹. Our World in Data aggregates relevant variables related to COVID-19 from different, credible sources. It is updated daily, measured on a country level and provides variables on confirmed COVID-19 cases as well as socioeconomic and health related variables, such as GDP per capita and life expectancy.

2 Exploring the Data

2.1 Overview

It may be interesting to start with a trend analysis to showcase the development of COVID-19 cases since the start of the pandemic. Figure 1 shows the trend of total confirmed COVID-19 cases per million from January 2020 until April 2021. The figure shows the contrasting evolution between continents with North America, South America and Europe having the overall highest occurrences of COVID-19 and Asia, Africa and Oceania showing a flatter growth.

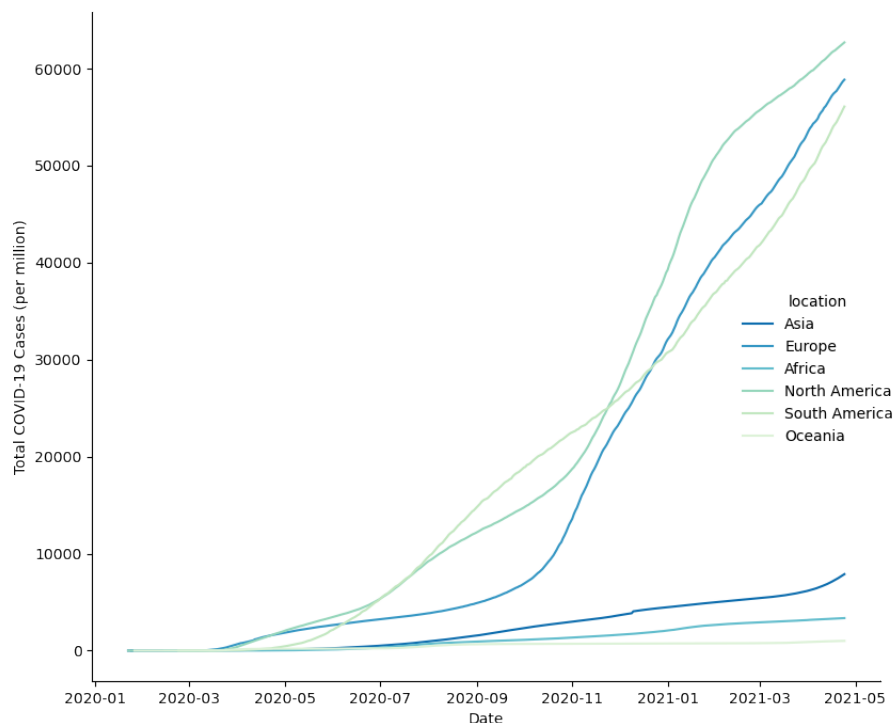


Figure 1: Total COVID-19 cases per Million

¹<https://github.com/owid/covid-19-data/tree/master/public/data>

2.2 Monthly Evolution of COVID-19

The monthly data show an overall growth since the first confirmed COVID-19 case in January 2020. There has been a strong increase in cases with a slight dip in December 2020 and a larger dip in February 2021, likely due to the strong lockdown measures around the world over the winter period. The graph also shows the strong year over year difference between January and April of 2020 and 2021.

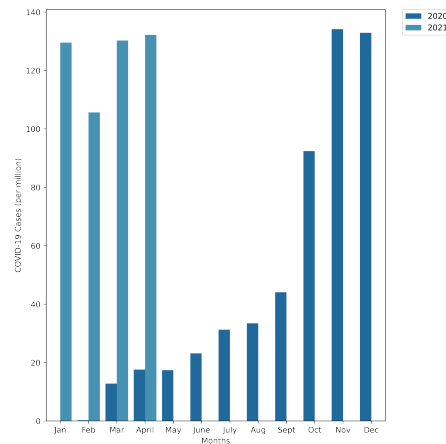


Figure 2: Total Monthly confirmed COVID-19 cases by Month

2.3 Most affected Countries

Taking a closer look at the current affect of COVID-19 on countries worldwide shows, that the highest COVID-19 death rates occurred in Europe, as well as Latin America and Central America. Although it should be noted that it is likely that confirmed cases are not reported correctly for some developing countries such as India, which is currently experiencing a crisis and unprecedented death rate due to COVID-19.

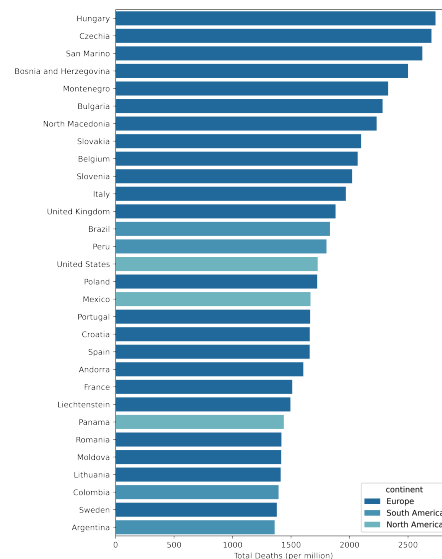


Figure 3: 30 Most affected Countries

3 Relationships

More severe COVID-19 illness in humans is associated with certain health conditions, such as respiratory diseases. The data by Our World in Data offers various COVID-19 related variables to measure the relationships between health and socioeconomic variables. Figure 4 shows the relationship between the variable *total cases per million* and the following variables: *GDP per capita*, *median age*, *aged 65 older*, *aged 70 older*, *extreme poverty*, *cardiovascular death rate*, *diabetes prevalence*, *female smokers*, *male smokers*, *hospital beds per thousand*, *life expectancy*, *human development index*. After plotting the data, we can see a clear relationship between *total cases per million* and *GDP per capita*, *median age*, *aged 65 older*, *aged 70 older*, *female smokers*, *hospital beds per thousand*, *life expectancy*, *human development index*. There is a slight relationship between *total cases per million* and *cardiovascular death rate*, *diabetes prevalence* and *male smokers* and no visible, linear relationship between *total cases per million* and *extreme poverty*.

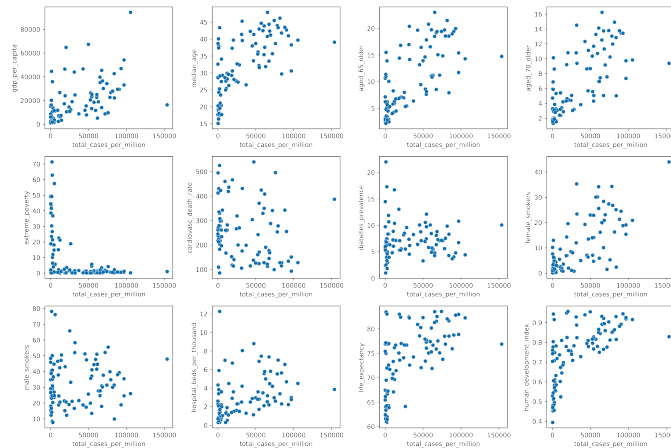


Figure 4: Relationships between Confirmed Total COVID-19 Cases and COVID-19 related Variables

Figure 5 confirms the positive relationships between all the variables in the data set. Total deaths per million and medium age have a correlation of 0.66. The correlation increases slightly for aged 65 older and aged 75 older. The correlation of total cases per million and female smokers is the largest for this variable with a correlation of 0.71. Total deaths per million is highly correlated with aged 65 older (correlation: 0.71), aged 70 older (correlation: 0.72) and female smokers (correlation: 0.73).

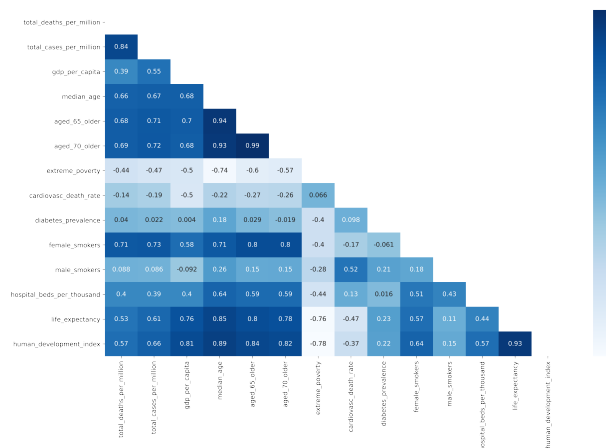


Figure 5: Correlation Matrix of COVID-19 related Variables