

Final project dataset report

Danyu Zhang, Limingrui Wan and Daniel Alonso

November 16th, 2020

The COVID-19 dataset

Coronavirus disease 2019 (COVID-19) is a contagious respiratory and vascular disease caused by severe acute respiratory syndrome coronavirus 2. 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. There are currently no proven vaccines or specific treatments for COVID-19, though several are in development.

Variables

The dataset contains the following variables:

- continent: Continent of the geographical location
- location: Geographical location, country
- total_cases: Total confirmed cases of COVID-19 at the location
- new_cases: New confirmed cases of COVID-19 at the location
- new_cases_smoothed: New confirmed cases of COVID-19 (7-day smoothed), the average of new cases during 7 days
- total_deaths: Total deaths attributed to COVID-19 of the location
- new_deaths: New deaths attributed to COVID-19 of the region
- new_deaths_smoothed: New deaths attributed to COVID-19 (7-day smoothed), the average of new deaths during 7 days
- total_cases_per_million: Total confirmed cases of COVID-19 per 1,000,000 people of the location
- new_cases_per_million: New confirmed cases of COVID-19 per 1,000,000 people of the location
- new_cases_smoothed_per_million: New confirmed cases of COVID-19 (7-day smoothed) per 1,000,000 people, the average of new cases per million during 7 days
- total_deaths_per_million: Total deaths attributed to COVID-19 per 1,000,000 people of the location
- new_deaths_per_million: New deaths attributed to COVID-19 per 1,000,000 people
- stringency_index: Government Response Stringency Index: composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response)
- population: Population of the location in 2020
- population_density: Number of people divided by land area, measured in square kilometers
- median_age: Median age of the population of the location, UN projection for 2020
- aged_65_older: Share of the population that is 65 years and older in 2015 of the location
- aged_70_older: Share of the population that is 70 years and older in 2015 of the location
- gdp_per_capita: Gross domestic product at purchasing power parity of the location (constant 2011 international dollars), most recent year available

- extreme_poverty: Share of the population living in extreme poverty of the location
- cardiovasc_death_rate: Death rate from cardiovascular disease in 2017 of each location (annual number of deaths per 100,000 people)
- diabetes_prevalence: Diabetes prevalence of each location (% of population aged 20 to 79) in 2017
- hospital_beds_per_thousand: Hospital beds per 1,000 people of the location, most recent year available since 2010
- life_expectancy: Life expectancy at birth in 2019 of each location
- human_development_index: Summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living