# 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
- location
- date
- total_cases
- new_cases
- new_cases_smoothed
- total_deaths
- new_deaths
- new_deaths_smoothed
- total_cases_per_million
- new_cases_per_million
- new_cases_smoothed_per_million
- total_deaths_per_million
- new_deaths_per_million
- new_deaths_smoothed_per_million
- reproduction_rate
- icu_patients
- icu_patients_per_million
- hosp_patients
- hosp_patients_per_million
- total_tests
- new_tests
- total_tests_per_thousand
- new_tests_per_thousand
- new_tests_smoothed
- new_tests_smoothed_per_thousand
- tests_per_case
- positive_rate
- tests_units

- stringency_index
- population
- population_density
- median_age
- aged_65_older
- aged_70_older
- gdp_per_capita
- extreme_poverty
- cardiovasc_death_rate
- diabetes_prevalence
- female_smokers
- male_smokers
- handwashing_facilities
- hospital_beds_per_thousand
- life_expectancy