* total number of COVID-19 cases reported in each country,
* infection and death rates,
* socio-economic data for each country describing its overall population,
* population density,
* climate,
* GDP,
* GDP per capita, and wealth level.
* In addition, we aligned COVID-19 rates with the data on the governmental containment measures.
* Finally, we used the mobility dataset that for each country reports the percent change in visits to public places:
* retail and recreation places,
* groceries and pharmacies,
* parks,
* transit stations,
* workplaces, and
* residential areas

The datasets have the following properties:

1. COVID-19 infection and death rates are reported in absolute numbers per country per day.

2. Countries may have different methodologies of reporting COVID-19 infected cases (this depends on how ubiquitous the testing system is).

3. Countries may also have different methodologies of reporting COVID-19 death cases (e.g., in Germany every patient that died while diagnosed with COVID-19 is counted as dead from COVID-19 vs. in Russia patients that died from comorbidities while also having COVID-19 are not part of the official statistics).

4. It is beyond the scope of this project to normalise COVID-19-related data with respect to the issues from (2) and (3). We rely on the official numbers provided by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

5. Containment measures are also reported per country per day. The data collected by the Oxford COVID-19 Government Response Tracker (OxCGRT) classifies all possible measures into 19 types, of which 13 types are of particular interest for us, as they deal with containment, closure policies, and the healthcare system, e.g., schools closure, contact tracing, or ban of international travels. Here is the full list of containment measures used in this report together with their codes that are used both in the original dataset:

C: containment and closure policies

1. C1: school closing
2. C2: workplace closing
3. C3: public events cancellation
4. C4: restrictions on gatherings
5. C5: public transport closure
6. C6: stay-at-home requirements 4
7. C7: restrictions on internal movement
8. C8: international travel controls

H: health system policies

1. H1: public information campaigns
2. H2: testing policy
3. H3: contact tracing
4. H4: emergency investment in healthcare
5. H5: investment in vaccines

6. For each containment measure the data is reported by country by case, both as a binary flag (a given measure was in place at a particular date) and a grade between 0 and 2 or 3 (0 meaning the given measure is not present and 3 being the strictest implementation of the measure).

7. We do not consider economic measures that helped sustain the economies; this is outside the scope of our project. We focus on containment measures only.

8. The percentage of missing values is very low, and we treat missing values as 0.

9. The mobility data is presented in the form of a time series between January and October 2020. 10. Demographic data is also reported per country. The population data is reported per country per year, up until 2019.

11. Income is reported using one categorical value per country: each country is classified either as High income, Upper middle income, Lower middle income, or Low income. Additionally, for each country we used the 2019 figures of GDP and GDP per capita, as reported by the World Bank.

12. Finally, we factored in the climate of each country, as some scientists claim that the virus is more likely to spread in colder temperatures. We used historic climate data from the Climatic Research Unit that reports the average yearly temperature per country as a single value.