

COVID-19 Data Set

Source: The source of this data is from Kaggle on a new coronavirus designated 2019-nCoV which was first identified in Wuhan, the capital of China's Hubei province.

Collection: The data was compiled and scraped from the <https://www.worldometers.info/coronavirus/> website so should be fairly dependable.

Contents: There are 16 columns and 209 rows in original data set

Limitations & Ethics: Different countries may have varying reporting practices, and there could be delays or discrepancies in the data reported by different sources. Some regions may have more robust reporting systems than others. Also COVID-19 situation is dynamic, and the status of the pandemic can change rapidly. Therefore, the data available on Worldometer may not always reflect the most current situation.

Relevance: I believe this data set meets the necessary requirements for this project as it is open source, includes a geospatial component, and meets the size and variable requirements.

Data Profile

Variable	Time Variant/Invariant	Structured/ Unstructured	Quantitative/ Qualitative	Nominal/Ordinal /Discrete/Continuous
Country/Region	Invariant	Structured	Qualitative	Nominal
Continent	Invariant	Structured	Qualitative	Nominal
Population	Variant	Structured	Quantitative	Continuous
Total Cases	Variant	Structured	Quantitative	Continuous
Total Deaths	Variant	Structured	Quantitative	Continuous
Total Recovered	Variant	Structured	Quantitative	Continuous
Active Cases	Variant	Structured	Quantitative	Continuous
Serious,Critical	Variant	Structured	Quantitative	Continuous
Total Tests	Variant	Structured	Quantitative	Continuous
WHO region	Invariant	Structured	Qualitative	Nominal

Questions to explore

- Number of total covid cases, deaths, and recovery?
- What continent were most impacted by COVID-19?
- Within continent, what country were most impacted by COVID-19?
- Does population size affect the percentage of cases and deaths?
- Which country had the fastest recovery rate?