Group 4: Lucas Mueller, Aaron Janaszak, Christian Buonfiglio, Addison Gangwish

Underweight adults

From the WHO, this dataset gives the percentage of individuals in a country who are underweight, where underweight is described as having a BMI (body mass index) of less than 18. We will use this dataset to predict the mortality rate among adults. We specifically focus on adults since they had the mortality rate data. We have data from 1975-2019, however for ML models we cut it to 2000-2016 like other data.

https://ghoapi.azureedge.net/api/NCD_BMI_18A

Access to safe drinking water

From the WHO, this dataset gives the percentage of individuals in a country with access to safe drinking water. It has data for each country for each year from 2000-2016. We plan to predict adult mortality rate from this dataset.

Like underweight adults, this dataset had several null values for countries where data could not be taken (namely Sudan and South Sudan). These nulls were interpolated with the means from their regions (in the case of Sudan, East Africa).

https://ghoapi.azureedge.net/api/WSH_WATER_BASIC

Adult mortality rate

This dataset shows the adult mortality rate from 2000-2016 for each country. This mortality rate is defined as the probability of dying between the ages of 15 to 60. That is, the number of people who have turned 15 who will not reach their 60^{th} birthday. We used this as a metric of a country's health.

https://ghoapi.azureedge.net/api/WHOSIS_000004

Malaria Incidence Rate

This dataset shows the number of people per 1,000 that are infected with malaria in a given country in a given year, with a date range of 2000-2016. This helps predict mortality rate. Many countries have no data for malaria, as they have cured it and case rates are too low to be detected, and we set these to 0.

Tuberculosis Incidence Rate

This dataset shows the number of people per 100,000 that are infected with tuberculosis in a given country in a given year. We have a date range of 2000-2016. We used this as another variable to help predict mortality rate.

Regions

We are using the following site to group different countries. This is used for interpolation and also so we can more easily perform analyses that go beyond national borders.

https://statisticstimes.com/geography/countries-by-continents.php

Census

We are using the yearly population by state. We can compare states' population to that of different nations. For instance, we saw the underweight adults and population of Iraq are similar to that of California in 1975, but the mortality rate and water availability of the former were much worse than the latter.

https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-total.html