Technical brief 1: Energy

Data Sources: Both datasets come from Our World in Data. Dataset 1 was downloaded from the OWID Github page, and dataset 2 from OWID energy access.

- Dataset 1: Energy data. The dataset is fairly big with 128 variables. A description for each can be found in the owid-energy-codebook.csv. Some data was recorded as early as 1900. Most recent recordings are from 2021.
- Dataset 2: Energy access data. This dataset contains information about the number of people with and without electricity access between 1990 and 2019 across different countries/ continents/ income brackets/ worldwide.
- Extra materials: owid-energy-codebook.csv describes the variables in dataset 1, and ISO_codes.csv (obtained via statisticstimes.com) provides an overview of ISO codes region and continent information.

You have been hired by a philanthropist who wants to solve the energy crisis by donating funds to energy sources.

Your report should contain statistics and visualisations that you interpret to provide the following key insights:

1. Averaging across time, which is the energy source (e.g. fossil fuels, nuclear power, and renewables) that generates the highest amount of electricity worldwide. How much electricity has been generated from each of these sources?
2. How has primary energy consumption (in terawatt-hours) changed worldwide across time? What are the key trends?
3. Focussing on human energy consumption from specific sources (i.e. coal, oil, gas, nuclear, hydropower, solar, wind, and biofuel). How much energy has been consumed per continent for each year between 1970 and 2020? Comment on the trends you observe.
4. What is the percentage of people with/ without electricity access for different income categories? Comment on the trends for each year between 1990 and 2019, and focus on income categories labelled High income, Upper middle income, Lower middle income, and Low income.
5. One original insight of your own, based on your exploration of the data.

Based upon all the insights above, how would you advise the philanthropist to donate money if they want to contribute solving the energy crisis and why?

Notes.

1. Before you start writing any code, go through the code book and identify which variables you will need to work with to answer the questions. This will take longer than you think but will save you huge amounts of time in the long-run.
2. Definitions and calculations:
    a. Fossil fuel: coal, goas, oil
    b. Renewables: hydropower, solar, wind, biofuel
    c. Low-carbon: Renewables + nuclear
3. Primary energy is calculated based on the 'substitution method' which takes account of the inefficiencies in fossil fuel production by converting non-fossil energy into the energy inputs required if they had the same conversion losses as fossil fuels. For more information on different calculation methods, see OWID article by Hannah Ritchie.

4. The geographical information contained within both datasets is quite complex – take the time to review the data to ensure you know what is in each variable:
    a. Both datasets have a code column (dataset 1: “iso_code”, dataset 2: “Code”) containing ISO country codes. Individual countries have an ISO code, the higher-level groupings of regions, unions, and income groups or discontinued countries (e.g. East Germany, Yugoslavia) may not have any data in this column or have a prefix of “OWID_” added. This makes the data difficult to work with, however, this is a realistic example of messy data, and as such, we have not altered the ISO Code columns.
    b. “Entity” (dataset 2) contains geographical information. Importantly, there are several different units represented in this variable: individual countries, World Bank Income Regions, geographical regions, and political unions.
    c. The World Bank provides a list of the current Country and Lending Group income classifications should you want more information.
5. In reality, more information would be needed to provide sound advice to inform the decision of where to donate. For the purposes of this assignment and to ensure a level playing field across all students and backgrounds, your decision should be based on the data alone.