202620 - DATA-101-21894

Introduction to Project 1

R. Miranda
10/9/2025

Dataset and question

Instructions: Provide the following:
1- Name of the dataset
2- Source of dataset.
3- Your research question.
4- The variables (name of columns)you will use to answer your questions
Write your answers on a word document and attach it here.

1. Name of the dataset: **Energy**, part of *Our World in Data*: ourworldindata.org/energy
The download is stored at https://github.com/owid/energy-data as a .csv file:
https://nyc3.digitaloceanspaces.com/owid-public/data/energy/owid-energy-data.csv

2. Sources:
The authors and curators are Hannah Ritchie, Pablo Rosado, and Max Roser
(https://ourworldindata.org/team/hannah-ritchie, https://ourworldindata.org/team/pablo-rosado, and https://ourworldindata.org/team/max-roser)

Their motivation, as stated in Our World in Data, is: "The world lacks a safe, low-carbon, and cheap large-scale energy infrastructure. Until we scale up such an energy infrastructure, the world will continue to face two energy problems: hundreds of millions of people lack access to sufficient energy, and the dominance of fossil fuels in our energy system drives climate change and other health impacts such as air pollution."

The data sources they used were:
(a) Statistical review of world energy (Energy Institute, EI) found at
https://www.energyinst.org/statistical-review,

(b) More details on metrics of energy consumption per capita (a key variable in the dataset) are given here:
https://ourworldindata.org/explorers/energy?overlay=sources&Total+or+Breakdown=Total&Energy+or+Electricity=Primary+energy&Metric=Per+capita+consumption&country=USA~GBR~CHN~OWID_WRL~IND~BRA~ZAF

(c) The variable definitions are at https://github.com/owid/energy-data/blob/master/owid-energy-codebook.csv

It consists of 130 variables over the time period: 1900-2024, and the main variables include:

1) countries or regions of the world,
2) population,
3) GDP,
4) total energy production,
5) total energy consumption,
6) type of energy produced or consumed,
7) carbon content of each type of energy.

They also include general classes of energy, such as
8) fossil,
9) biorenewable,
10) renewable (other than biorenewable),

as well as specific types of energy within those classes, such as
11) oil,
12) gas,
13) coal,
14) nuclear,
15) wind,
16) electricity,
17) solar (portion converted to electricity),
18) fraction of other primary sources converted to electricity (such as oil, gas, nuclear).

The table does not explicitly mention geothermal or marine sources, but I assume they are part of renewable (other than biorenewable) energy.

3. Research questions:

The main questions are based on the energy and technological revolutions that happened over the period covered by this dataset, 1900-2024. In general, we understand that the increasing availability of energy made it possible for the world to increase its overall standard of living over the last 124 years. Such standard of living is linked to GDP, one of the key variables reported in the dataset. I note that "standard of living" is better measured by energy consumed than by energy produced, as some countries are high energy producers yet export much of it.

The leading hypothesis is that cumulative GDP and cumulative energy produced and energy consumed follow the same trend; a strict statistical correlation analysis will not be carried out in this first project; but I expect that a graphical depiction will clearly show the trends.

A second important hypothesis is that different regions of the world developed their energy infrastructure at different rates and possibly with different consequences on their GDP. Thus the next part of this study will examine the historical energy evolution (production capacity) and consumption rate (related to standard of living) by world region and by groups of countries (to be defined). I expect to reveal how growth in energy production and in energy consumption are not parallel, but rather, follow each other, with an induction period in the growth of consumption. I also expect a more complex reality, where growth of production and consumption in certain countries follow different rates. I intend to explore in general the historical connotations for selected countries, such as wars or crises in certain years that caused perturbations to the energy production, energy consumption, and GDP curves.

I do not intend to explore in this limited study the interesting evolution from fossil to renewable energies, because such must have been studied to death by the curators of this dataset, since such was their main objective.

Rather, I'll focus on the general trends in the relations between total energy (production and consumption), GDP (economic power of nations), and standard of living (related to energy consumption).

4. Variables to be used:

Based on the rationale provided, I will aggregate total energy, which includes fossil energy of various types and renewable energy of various types as mentioned in part 2(c) above. I will use the countries column and create regions (such as continents or subcontinents) and groups of countries based on the World Bank's classification of countries based on degree of development. I will create a modified GDP (per capita) for those regions and groups of countries. Thus I will explore the hypotheses and unknowns described in part 3.