

CEE 492 Team Uhhh Project

This manuscript ([permalink](#)) was automatically generated from [uiceds/cee-492-term-project-fall-2022-team-uhhh@f884dec](#) on October 31, 2022.

Authors

- **Robert Joseph Wiggins**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [BobbyWiggles](#) ·  [bobbythewiggler](#)

Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign

- **YongjianMa**

 [0000-0001-8374-9401](#) ·  [YongjianMa](#)

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Muhammad**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [hanzla07](#)

Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign

Effects of Energy Consumption Industries on Carbon Emissions by Country

Written by: Yongjian Ma, Muhammad Hanzla Rafaqat, Robert Wiggins

Description:

The datasets of interest are housed within a collection of datasets titled "Compiled Datasets for CDP Analytics Competition." The dataset was published to Kaggle by user "seraphimstreets." and features data collected and cited from several reputable organizations, such as Organization for Economic Co-operation Development, The Institute for Health metrics and Evaluation, and The World Resource Institute.

The first dataset in question is titled "Percent of Energy Consumption by Country." The dataset is a CSV file containing 11 columns of data. The first three columns include the country and year the data in each row was obtained from, as well as the corresponding country codes. The next eight columns list the percentages of each industry's energy consumption within each country in a given year. The industries recorded include coal, gas, oil, hydroelectric, nuclear, solar, wind, and biomass.

The second dataset in question is titled "CO2 Emissions by Country." The dataset is also a CSV file containing 4 columns of data. The first three columns again include the country and year the data in each row was obtained from, as well as the corresponding country codes. The last column lists the CO2 emissions from each country in a given year.

Proposal:

With the increasing amount of CO2 emissions worldwide and the difficulty of establishing meaningful efforts to combat the increasing climate crisis, it is integral to pinpoint the major contributors to climate change and work to lower the CO2 emissions produced by these contributors as much as possible. With the dataset collection in question, there seems to be an abundance of data available to determine the major contributors to climate change, as well as the energy industries that contribute the least amount of CO2 emissions. Our group plans to use these datasets to analyze how each energy industry contributes to CO2 emissions within each country and determine which industries require the most changes and limitations to curb the increasing effects of climate change.

Furthermore, total annual CO2 emissions deviate between countries, as some countries have become more industrialized than others over the years. Therefore, our group is also interested in analyzing the contributions of each country's industries to climate change and determining the countries that require the most changes and limitations to carbon emissions as well, in order to curb the effects of climate change without directly targeting less industrialized countries who do not contribute nearly as much.

Source:

<https://www.kaggle.com/datasets/seraphimstreets/environmentequity-starterpack>

Description: The datasets of interest are housed within a collection of datasets titled "Compiled Datasets for CDP Analytics Competition." The dataset was published to Kaggle by user "seraphimstreets." and features data collected and cited from several reputable organizations, such as Organization for Economic Co-operation Development, The Institute for Health metrics and Evaluation, and The World Resource Institute.

The first dataset in question is titled "Percent of Energy Consumption by Country". The dataset is a CSV file containing 11 columns of data. The first three columns include the country and year the data in each row was obtained from, as well as the corresponding country codes. The next eight columns list the percentages of each industry's energy consumption within each country in a given year. The industries recorded include coal, gas, oil, hydroelectric, nuclear, solar, wind, and biomass.

The second dataset in question is titled "CO2 Emissions by Country." The dataset is also a CSV file containing 4 columns of data. The first three columns again include the country and year the data in each row was obtained from, as well as the corresponding country codes. The last column lists the CO2 emissions from each country in a given year.

Prior to directly analyzing any data, our group first decided to generate overarching questions about the nature of our data, the relationships we could identify between variables, and how these relationships reinforce our project goals. The following four questions were generated: □ Is there a strong correlation between energy consumption and CO2 emissions? If so, what energy industries are responsible for the most CO2 emissions? □ How have energy consumption and CO2 emissions changed since 1965? Have there any significant increases or decreases, or have these variables remained generally stagnant? □ How do changes in energy consumption and CO2 emissions over the years differ between countries? Have some countries made greater efforts to reduce their energy consumption and CO2 emissions? Have any countries increased these variables? While these questions do not necessarily account for all potential trends within our data, we believe they comprise the most important features of our data and cover our general project goals.

To have a better understanding of the dataset, I have to organize and clean datasets. Initially, the biggest problem of these datasets is the problem of the unit. Some sorts of energy consumption categories are listed by Terawatt-hour(TWh), while others are listed by Exajoule(EJ). So, clarifying and transferring the unit of energy is beneficial for data analysis and comparison.

Secondly, since this dataset is about energy consumption during 1965-2018, there are several data missed because some countries/regions didn't begin investigating the consumption of energy level at a very early age. These missing data will significantly disturb the calculation of the dataset and cause an error during code running. To solve this problem, I used "replace" function to change all missing data to 0.

Thirdly, these datasets contain several data representing a massive region like South Africa or Mid East. These data caused duplication problems when I tried to get the total energy consumption per year. To get a straightforward and convenient dataset, I only collected the data from a specific country or region like the United States or China. image.png !("C:(2) .png") Fig 1. Organize and Clean the datasets

Predictive Modeling: