Visualisation and Analysis for GHG Emission Data Individual Report for Task 4

COMP5048/4448 Assignment 2

line 1: Rayne Zhu
line 2: *530551405*

line 4: CC-01
line 5: Vadi Sai Sakthivel

I. Processing the Data

A. *Group the data in one file*

The raw datasets contain 12 files. Each file describes one specific emission category or one socioeconomic indicator from the year 1971 to 2022 for all countries. In the newly generated dataset, the emission categories and socioeconomic indicators can be seen as 12 attributes. The values of year can be seen as one attribute named "Year". According to this design, the country names are located in the first column. The following columns are the values of the year and 12 attributes, corresponding to the country. The cleaned dataset will be referred to as the cleaned dataset in the following analysis.

B. *Understanding the attributes*

Understanding accurately the 12 attributes in the cleaned dataset is crucial for defining the "highest emissions" and "lowest emissions" for later analysis. The Database Documentation of Greenhouse Gas Emissions, 2023 edition, issued by the International Energy Agency(IEA), elaborates the definitions of names used in the dataset.

It is suggested to give all attributes abbreviated for clear reading and interpretation during the investigation.

It is noticed that the 2023 edition replaced the IEA CO2 emissions from the fuel combustion database with expanded content as of 2021. The old and new definitions are both indicated in the 2023 edition documentation, helping to understand the exact meaning of each attribute. Under the new regulation rules, CO2 fuel combustion is renamed as GHG from fuel combustion by checking the constituents in the emission equation from IEA documentation. They have the same constituents. Thus, for attributes starting with "CO2 emissions", refer to the "CO2 emissions from fuel combustion", which is equal to "GHG from fuel combustion". However, the unit of CO2 emissions from fuel combustion is different from GHG from fuel combustion. In conclusion, the CO2 emissions from fuel combustion and GHG from fuel combustion refer to the same item but in different units.

The GHG emission from Energy(total), the GHG emissions from fuel combustion(total), the GHG emissions from fuel combustion(oil), the GHG emissions from fuel combustion(gas), the GHG emissions from fuel combustion(coal), and the CO2 emissions from international (aviation bunkers), CO2 emissions from international (marine bunkers) have the same units, as "million tonnes of CO2 eq. The GHG fugitive emissions(total) has unit of "kilo tonnes of CO2 eq). They are excluded from the green gas emissions calculations for national use. The value of these two is decided by the travelling's departure and destination located in a country.

The IEA 2023 database document on page 59 and page 92 explains the unit conversions. The data below attribute GHG fugitive emissions(total) should be divided by 1000 for consistency. CO2 emissions/TES(CO2TES), CO2 emissions/GDP using exchange rates(CO2GDP), CO2 emissions/GDP using purchasing power parities(CO2PPP), CO2 emissions/population, are considered as socioeconomic indicators(CO2PPP). CO2GDP using exchange rate and CO2PPP have the same units as kgCO2/ US dollars(2015 prices). CO2GDP is computed using the total CO_2 emissions from fuel combustion and GDP is calculated using nominal exchange rates, The value measures the amount of CO_2 emissions produced per unit of GDP. CO_2PPP provides a more comparable measure across countries by accounting for differences when comparing countries in cost of living and inflation rates. TES is the total energy supply, representing the total amount of energy available in a country or region, which consists of energy production, energy imports, and exports, except international marine bunkers and international aviation bunkers, and including or removing stock changes in the country. CO_2TES measures the carbon intensity of the energy supply, less CO_2 is being emitted per unit of energy supplied and the energy supply is becoming less carbon-intensive, which is beneficial for reducing overall greenhouse gas emissions. CO_2POP measures the amount of CO_2 emissions produced per person.

Table I. Names and units of attributes.

Name	Unit	Abbreviated name
GHG emissions from Energy (total)	million tonnes of CO2 eq	GHGEnergy
GHG fugitive emissions (total)	million tonnes of CO2 eq (by unit conversion)	GHGFugi
GHG emissions from fuel combustion(total)	million tonnes of CO2 eq	GHGFC
GHG emissions from fuel combustion(gas)	million tonnes of CO2 eq	GHGFCGas
GHG emissions from fuel combustion(oil)	million tonnes of CO2 eq	GHGFCOil
GHG emissions from fuel combustion(coal)	million tonnes of CO2 eq	GHGFCCoal
CO2 emissions from international(Marie bunkers)	million tonnes of CO2 eq	CO2MARBU
CO2 emissions from international(Aviation bunkers)	million tonnes of CO2 eq	CO2AVBUNK
CO2 emissions/TES	tonnes CO2/terajoule	CO2TES
CO2 emissions/GDP using exchange rates	kgCO2/ US dollars (2015 prices)	CO2GDP
CO2 emissions/GDP using purchasing power parities	kgCO2/ US dollars (2015 prices)	CO2PPP
CO2 emissions/population	tonnes CO2/capita	CO2PPP

The socioeconomic indicators' units are not converted due to their specific meaning for interpreting the social and economic impacts of the emissions. It will be considered when during the investigation but does not need to be converted to the same unit as the value of the emission volume as the two cannot be compared.

C. *Investigation of the relationships between attributes*

The GHG emissions from Energy(Total) are the sum of the GHG fugitive emissions and the GHG emissions from fuel combustion(total).

$$\begin{aligned}&\text{GHG emissions from Energy(Total)}\\ =\;&\text{GHG fugitive emissions} + \text{GHG emissions from fuel combustion(total)}\end{aligned}$$

The GHG emissions from fuel combustion(total) is approximately sum of the the GHG emissions from fuel combustion(oil), the GHG emissions from fuel combustion(gas), and the GHG emissions from fuel combustion(coal). Oil, gas, and coal are considered as main productions of energy resources. There are other tiny parts calculated for GHG emissions from fuel combustion according to the formula provided by IEA documentation[2].

GHG emissions from fuel combustion(total)

=fuel combustion(oil)

+GHG emissions from fuel combustion(gas)

+GHG emissions from fuel combustion(coal)

II. PINPOINT THE TOP FIVE COUNTRIES THAT HAVE THE LARGEST EMISSIONS

A. *Technique 1: finding the top **eight** countries*

The cleaned dataset has 205 countries and 38 regional aggregates in the country column. Get all the data when the year equals 2021. The implementation of codes sorting the countries' values from the largest to the smallest of each attribute. We decided to find the top eight countries rather than five. If we define country number as five, there will be more than five countries represented in the results. It will be hard to recognize the countries that generally always performed high values of each attribute. When we use eight countries, even if some countries are not in rank top five, at least they will be at the position of five to ten, which still indicates they have high value for the attribute.

B. *Technique 2: divide 12 attributes into three groups*

The attributes indicate the green gas emission in different aspects. The 12 attributes will be divided into “emission volume attributes”, “International fuel combustion” and “socioeconomic indicator”. The most considerable group for decision-making is the “emission volume attributes” includes GHGEnergy, GHGFugi, GHGFC, GHGFCGas, GHGFCCoal, and GHGFCOil as group 1. The “International fuel combustion” includes CO2MARBUNK and CO2AVBUNK as group 2. The “socioeconomic indicator” includes CO2TES, CO2GDP, CO2PPP, and CO2POP as group 3. The group 1 is the most considerable because it directly represent the question for the goal of looking for the country having the highest “emission volume in the year 2021.

The result of ranking the countries’ frequency

With the “Top Eight countries” technique, we can count the countries' frequency according to the code output. The five most frequent countries will be the final solution of the “ top five countries that have the largest emissions.” The group 1 includes China (6 times), the United States (6 times), Russian(6 times), Japan (5 times), India (5 times), Iran (4 times). The group 2 includes:: China (2 times), the United States (2 times), United Arab Emirates (2 times). There is no count for a 1-time country because there are only two attributes in this group. The count for countries is rather 1 or 2. Counting as 1 for a country does not mean very much. The group 3 includes: North Korea(3 times), Mongolia (3 times). These are indicators.

The group 1 is the most considerable because it directly represent the question for the goal of looking for the country having the highest “emission volume in the year 2021. In group 2, the country with high values of CO2MARBUNK and CO2AVBUNK mostly indicates the country has ports business and services and has efficient import and export business. China and the U.S. are on the list again. For group 3, the assessment is according to the unit emission. Unit emissions are not representative of a country's overall emissions, regardless of the values of these attributes.

Consequently, the top five countries are China, the United States, Russia, Japan, and India.

GRAPH.I. The codes and corresponding partial output as an example.

```
import pandas as pd

df = pd.read_csv("processed_df_clean.csv")

df_2021 = df.loc[df['Year'] == 2021].copy()
df_2021['df_GHGEnergy'] = pd.to_numeric(df_2021['df_GHGEnergy'], errors='coerce')

not_attr = ['Unnamed: 0', 'Country', 'continents', 'oced_country', 'Year']
attr = [col for col in df_2021.columns if col not in not_attr]

for column in attr:
    df_2021[column] = pd.to_numeric(df_2021[column], errors='coerce')
df_2021.head()

top_8_values = {}
for column in attr:
    top_8_values[column] = df_2021.nlargest(8, column)

for column, top_values in top_8_values.items():
    print(f"Top 8 values for {column}:")
    for index, row in top_values.iterrows():
        print(f"{row['Country']} : {row[column]}")
    print()
```

```
Top 8 values for df_GHGEnergy:
China : 11314.48859
United States : 5018.45847
India : 2427.365033
Russian Federation : 2227.925656
Japan : 1012.031969
Iran : 814.3316204
Indonesia : 678.5563836
Germany : 638.3590046

Top 8 values for df_GHGFCCoal:
China : 8426.499593
India : 1636.882149
United States : 999.8745071
Russian Federation : 436.7223224
Japan : 404.9242464
South Africa : 326.4016087
Indonesia : 287.5638848
South Korea : 262.115743
```

Fig. 2. The example for explaining how to count: The partial output can have a counting result: China(2 times), The United States(2 times), India(2 times), Russian Federation: (2 times), Japan(2 times).

III. SELECT THE UNITED STATES TO FIND ITS MINIMUM EMISSION YEAR

A. *Create visulization*

Use tableau to create visualization and out several graphs in a dashboard. The interaction of each graph is navigated by the year on the right-up corner. The live interaction is in the group video and uploaded tableau file. When modifying the year, every graph will have reflections. It will be helpful for identifying the dot location of the year in the lines.

GRAPH II. The dashboard and general visualization

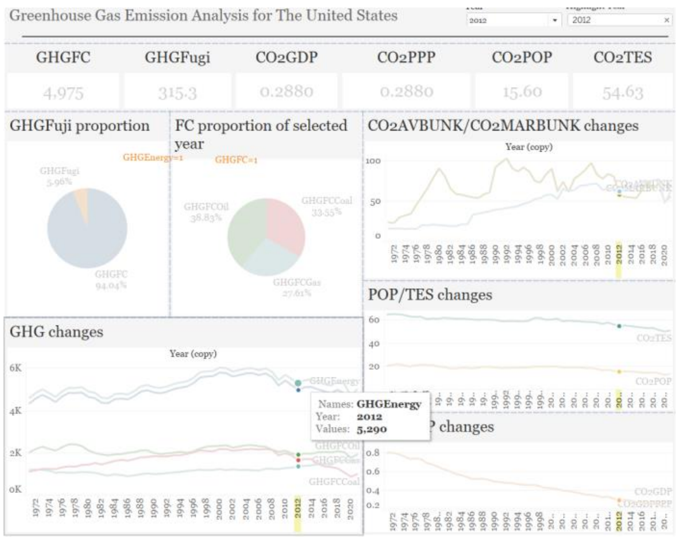

Fig. 3. The live interaction is in the group video and uploaded tableau file. Each graph and its attributes can be selected.

B. *Finding the year of minimum GHG emission*

1) *Technique: divide attributes into several groups.*

To find the the year of minimum GHG emission, divide attributes into several groups and find their minimum values and corresponding years. The decision-making strategy to divide attributes to several groups based on the equations and relationships. In each group, there will be one or several fitted points for the minimum value and its year. After analyzing all

12 attributes, divided into several groups, the most frequent year represented for the year of minimum emissions will be chosen as the year we are looking for.

2) *Group of "Emission volumes"*

a) GHGEnergy, GHGFC and GHGFugi

According to the equations, the GHGEnergy, GHGFC and GHGFugi are analyzed together because the sum of GHGFugi and GHGFC is GHGEnergy. The proportion of GHGFuji out of GHGEnergy in 1983 was 4.87%. [1] From both graphs, the percentage of GHGFuji in GHGEnergy is small, so the trend and value changes of GHGFuji could not significantly affect the total GHG emissions. The range of change for GHGFuji is much smaller than the change for GHGFC. This is the reason why the GHGFC and GHGEnergy look like having almost the same trends. The minimum of GHGFC is equal to GHGEnergy. The right graph of the single line for GHGFuji is in **1983.** The minimum points directly seen from the visualization are the year 1971, and the year 1983. From 2007 to today, the general trend of GHGEnergy has decreased reaching the lowest emission in **2020**.

GRAPHS.III THE VISUALIZATION OF GHGFuji, GHGFC and GHGEnergy.

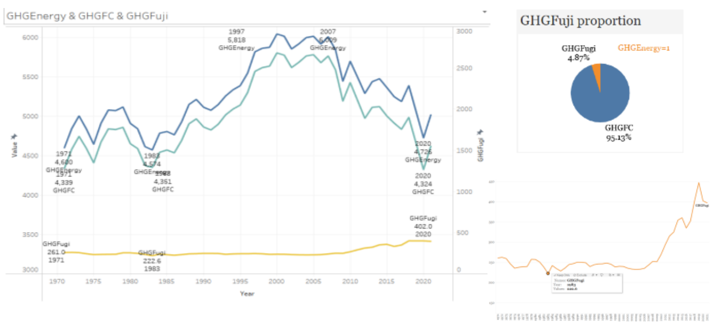

Fig. 4 The several years and the values of minimum emissions of attributes are labelled, the pie chart in 1983.

b) GHGFC, GHGFCOil, GHGFCGas, GHGFCCoal.

Although the sum of GHG emissions made from oil, gas and coal is GHGFC, the trends of each partial and their minimum values should be investigated. The graph reveals an upward trend in GHGFCCoal from 1971 to around the 2000s, then a long-term decline from the early 2000s until 2020. The year **2020** has the lowest GHGFCCoal emissions. The minimum point of GHGFCOil is in **2020**. The GHGFCGas first decreased until **1986** to reach the minimum then increased continually to 2020. The right pie charts show the proportion of each production out of GHGFC, compared between 2020 and 1986.

GRAPHS.III THE VISUALIZATION OF GHGFCOil, GHGFCGas and GHGFCCoal.

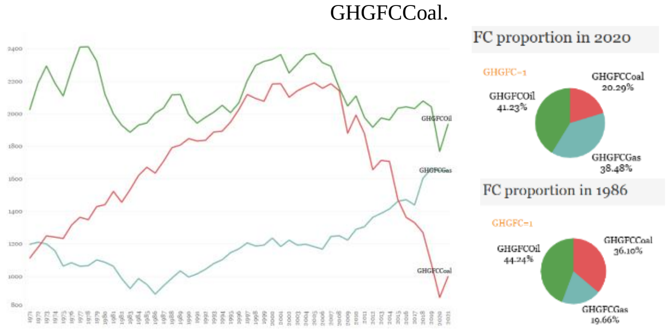

Fig.4-3. The years and the values of minimum emissions of attributes are labelled in line charts and pie charts.

c) CO2MARBUNK and CO2AVBUNK

CO2MARBUNK and CO2AVBUNK are excluded from national use in a country for domestic transportation and represent emissions caused by international navigation departure or arrival in the United States. These attributes affect the country's international transportation status in the world in terms of tourism. Besides, it is affected by the import and export of the United States.

GRAPHS.IV THE VISUALIZATION OF CO2MARBUNK AND CO2AVBUNK

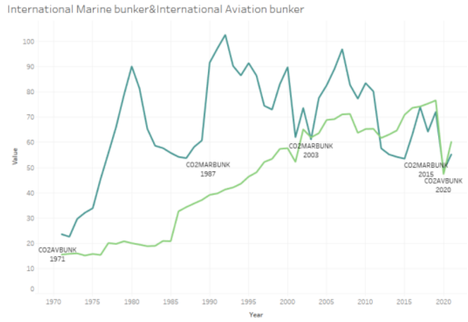

Fig.5. The several years and the values of minimum emissions of CO2MARBUNK and CO2AVBUNK.

The emissions in the year **1971** for both CO2AVBUNK and CO2MARBUNK are the minimum. The two attributes use million tonnes of CO2eq, having values below 100. It is non-essential compared to 4500 to 6000 million tonnes of CO2eq for the overall emission volume of the United States.

d) The *GHGFCGas and GHGFuji*

The volume of GHGFugi almost remained but fluctuated from 1971 to 2009 around 240 to 250 million tonnes of CO2 eq, but increased rapidly from 2009, and reached a peak in 2019 at 447.4 million tonnes. The GHGFuji and GHGFCGas are the two attributes increase in the past 20 years. They have very similar trends.

GRAPHS.V THE VISUALIZATION GHGFCGas and GHGFuji

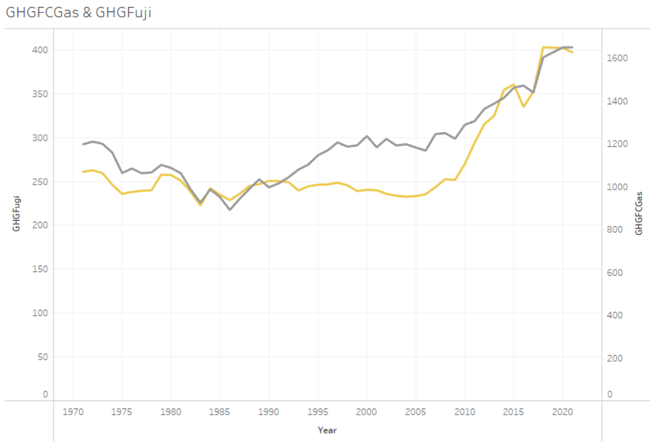

Fig.5. The trends of GHGFCGas and GHGFuji are very similar.

The reason is that the GHGFuji are usually produced by the use of Gas. The gas is treated as a cleaner energy resource than Coal and Oil. The decrease in greenhouse gas emissions I make up by the increase in production from using of gas.

3) *Group of "Socioeconomic indicators*

The description of the definition of CO2TES is a crucial metric for evaluating the carbon efficiency of the total energy supply. The CO2GDP and CO2PPP have the same values because we selected the United States as the target. The ratio for calculating the CO2PPP uses the United States as a standard for other countries. CO2TES is the description of carbon intensity. TES stands for Total Energy Supply, which states the energy resources for the country's use. The value of CO2TES demonstrates the relationship between energy supply and the emission of the resource produced. CO_2POP measures the amount of CO_2 emissions produced per person. The CO2TES and CO2POP could reflect the Energy resource efficiency.

GRAPHS.V THE VISUALIZATION OF SOCIOECONOMIC INDICATORS

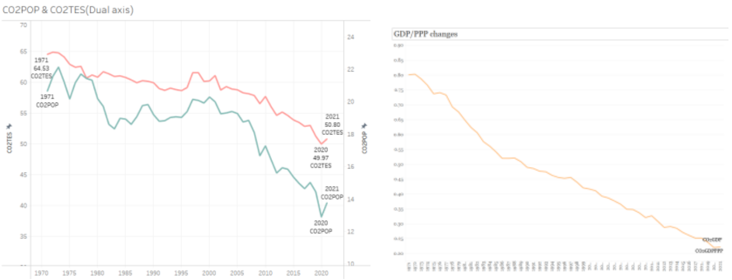

Fig.4-5. The trends of several socioeconomic indicators.

The trends of both indicators generally decrease from 1971 with a flat increase from around 1995 to 2000, then continue decreasing until 2020, reaching the lowest points. The graph went all the way down from 1971 to **2020**.

Based on the previous analysis, the most frequent and representative year is 2020. We will choose 2020 as the year with the lowest U.S. emissions

IV. DEPLOY A PLAN TO REVERSE FROM YEAR 2021 TO 2020

A. *Looking for the reason cause the minimum emissions in the year 2020 in the United States.*

1) Technique 1: recall back personal experience

What happened in 2020? Obliviously it was the COVID-19 pandemic. The year 2020 is only one year before 2021. The change in greenhouse gas emissions is extremely huge. The emission of 2021 might not be the highest in history but the year 2020 is considered as the lowest emission during the past 50 years. From personal experience, it is not possible to directly get the reason for "why the greenhouse gas emissions diminish. However, human life experiences change a lot during the pandemic. This could lead to some ideas of how people reduce greenhouse gas emissions, such as reducing unnecessary transportation, continuing work from home, etc.

2) Technique 2: The period of another peak, rather than the year 2021, to 2020 can be considered.

It noticed that the year 2021 is not the highest emissions year. The experience from previous strategies for reducing emissions can be continued to promote. The emissions from coal and oil fluctuated decrease over the past 20 years. The technology to process the energy resource should be innovated and continued to be used, for example, the Carbon Capture and Storage (CCS) [3]. Furthermore, the policy regulated in the past 50 years should be continued and supervised by other countries from all over the world. Besides, the socioeconomic indicators lead to the consideration of global energy resource pricing. As a result, the strategies reversed from the year 2021 to the year 2020 can be considered the strategies processed from the 2000s to 2020. Details strategies are listed in the group project.

V. EVALUATION

The analysis for task 4 is complicated and heavy, so it is necessary to have a questionnaire and interview to improve the analysis. In the process, 5 classmates and 5 students from other courses are invited to participate in the evaluation process. The invitation to classmates who take the COMP5048 could have expert views and knowledge, providing professional evaluations of the analysis. The other 5 students from other courses who do not have professional visualization skills could efficiently reflect whether the information is clear, the design is pleasing, and the interaction is easy to understand. The questionnaires can be designed combination of close-end and open-end formats. The levels start from 0 to 10 indicating the weak to perfect for the analysis. The average mark from students having professional backgrounds is higher than students without data visualization knowledge. It might reflect the complexity of the analysis. However, they both believe the labels showing on graphs and interactions are pleasing to view.

According to the evaluation results, the classmates of COMP5048 point out that the scale of the x-axis of each graph is not perfect for identifying trends. There are several ways to solve the problem: changing the x-axis range, using the dual-axis method, and separating the attributes to another graph. The changes have been made and the improved graphs are shown currently in both individual and group reports. The students without professional knowledge give positive feedback however are sometimes confused to the multiple attributes' meaning. In the future study, adding the data dictionary could effectively improve the understanding of the audience.

REFERENCES

[1] IEA."Global Energy Review 2021." IEA.https://www.iea.org/reports/global-energy-review-2021 (accessed 2021).

[2] IEA."Global Energy Review 2021." IEA.https://www.iea.org/reports/global-energy-review-2021 (accessed 2021).

[3] Holappa, L. A General Vision for Reduction of Energy Consumption and CO2 Emissions from the Steel Industry. Metals 2020, 10,1117. https://doi.org/10.3390/met10091117 (accessed 2023)