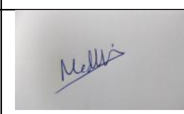
Assignment Coversheet – GROUP ASSIGNMENT

Please fill in your details below. Use one form for each group assignment.

Assignment Details:

Assignment Title	**OCMP5048 Report 2**				
Assignment number	**2**				
Unit of Study Tutor	**Maria Lim**				
Group or Tutorial ID	**Group 2**				
Due Date	**18 Nov 2023**	Submission Date	**18 Nov 2023**	Word Count	**3038**

Membership

Group Name/Number	Group 2				
Family Name	Given Name(s)	Student Number (SID)	Unikey	Contribution + percenage	Signature
Islam	Md Faiyam	490470604	fisl5259	30%	FIslam
Middleton	Daniel	530839956	dmid0500	30%	DM
Fernando	Medhavi	308172930	yfer0059	30%	
Tseng	Yun	500639001	tyun0514	10%	Christina Tseng

Declaration:

1. I understand that all forms of plagiarism and unauthorised collusion are regarded as academic dishonesty by the university, resulting in penalties including failure of the unit of study and possible disciplinary action.
2. I have completed the **Academic Honesty Education Module** on Canvas.
3. I understand that failure to comply with the Academic Dishonesty and Plagiarism in Coursework Policy can lead to the University commencing proceedings against me for potential student misconduct under Chapter 8 of the ***University of Sydney By-Law 1999*** (as amended).
4. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that it is not my own by acknowledging the source of that part or those parts of the work.
5. The assessment has not been submitted previously for assessment in this or any other unit, or another institution.
6. I acknowledge that the assessor of this assignment may, for the purpose of assessing this assignment may:
   a. Reproduce this assignment and provide a copy to another member of the school; and/or
   b. Use similarity detection software (which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking).

7. I have retained a duplicate copy of the assignment.	
Please type in your group number here to acknowledge this declaration:	**Group 2**

What can we do about GHG/CO2 emissions?

I. Introduction

In the ongoing discourse on environmental sustainability, understanding and mitigating Green House Gases (GHG)/ Carbon dioxide (CO2) emissions have become paramount. This report explores the various emission profiles across different countries, offering a visual analysis on sources which are driving environmental changes. Understanding trends and similarities of data obtained from these emission levels will assist nations in developing stronger implementations of key initiatives to alleviate the effects of climate change. Countries report their emissions by combining data on types of activity with the emissions [1]. Examining key variables such as GHG Energy, GHG Fugitive and GHG from fossil fuel combustion, we gain valuable insights into the specific contributors to atmospheric pollution.

II. Analysis of aggregate GHG/CO2 emissions

A. *Data Cleansing and Pre-processing*

The ‘GHGHighlights.xls’ dataset provided for this project, sourced from the Inernational Energy Agency (IEA), is a large and complex Excel Spreadsheet, which required pre-processing and cleansing to obtain a single Comma Separated Values (CSV) file with various qualitative and quantitative variables for analysis. The filtered dataset contains the Country, Year and a collection of GHG and CO2 socioeconomic measures between 1971-2022. The IEA dataset was used as input into a Python script, which incorportaed the melt function of the ‘pandas’ module to extract and transform the data from multiple worksheets and output to a single table in a CSV file (Figure 2).

Country	Year	GHG Energy	GHG Fugi	GHG FC	GHG FC - Coal	GHG FC - Oil	GHG FC - Gas	CO2 MARBUNK	CO2 AVBUNK	CO2-TES	CO2-GDP	CO2-GDP PPP	CO2-POP
Canada	1971	366.6118133	20994.32288	345.6174905	64.35598819	211.748862	68.3526568	3.099452497	1.265993236	57.486	0.719	0.702	15.492
Chile	1971	22.06574132	355.343975	21.71039734	5.167808865	14.78051651	1.324191205	0.606738605	0.438639753	57.724	0.466	0.278	2.161
Mexico	1971	118.9967321	22305.29285	96.69143921	5.206264152	70.06945173	19.58605364	0.258252769	1.403116037	52.102	0.335	0.176	1.754
United States	1971	4599.842353	260963.831	4338.878522	1112.829169	2024.817817	1197.765398	23.59087962	15.50762078	64.525	0.801	0.801	20.651
Australia	1971	152.1925554	6389.95525	145.8026002	75.65644928	65.2865156	4.005109765	5.152768116	1.59087494	66.342	0.451	0.499	10.863
Israel	1971	13.90319643	13.833375	13.88936306	0.022977732	13.83222933	0.032800045	0	1.810801528	57.262	0.283	0.285	4.517
Japan	1971	765.1442739	8359.744325	756.7845296	202.7778602	545.4191035	8.587565906	50.45768772	3.834008722	67.048	0.49	0.419	7.154
Korea	1971	66.52308347	11775.19714	54.74788633	23.77285235	30.97503398	0	1.542673945	0	74.507	0.751	0.569	1.61
New Zealand	1971	13.80086605	61.8577	13.73900835	4.034478086	9.458690582	0.245839685	1.053429755	0.645895128	47.232	0.226	0.232	4.714
Denmark	1971	55.96319869	8.884765	55.95431392	6.156243962	49.75611826	0	2.112617698	1.943215885	71.532	0.405	0.44	11.167
France	1971	433.3186661	1627.5393	431.6911268	142.7760271	268.2695294	17.75667797	12.89773437	4.617898976	63.769	0.442	0.396	8.074
Germany	1971	993.8918925	3181.157525	990.710735	565.5336944	385.9150593	38.53923864	13.13702067	7.646320011	76.585	0.676	0.583	12.485
Italy	1971	293.844143	1922.53775	291.9216053	32.98857589	234.8037602	24.11157756	23.10852055	3.504930044	65.581	0.328	0.269	5.352
Netherlands	1971	128.8385579	172.991224	128.6655667	15.42701401	65.83875586	47.39979678	28.62210313	2.026150312	59.942	0.444	0.399	9.678
Norway	1971	23.50249496	224.0175178	23.27847745	3.883623883	19.39485356	0	1.937973084	0.704060636	41.228	0.21	0.259	5.893
Poland	1971	293.139809	818.633655	292.3211754	258.7104788	21.85254658	10.31951854	1.649913924	0.525876995	79.735	2.298	1.076	8.761
Sweden	1971	83.12029424	14.625765	83.10566847	5.56744117	77.21846921	0	3.622786567	0.332045897	54.374	0.436	0.458	10.132
United Kingdom	1971	632.5675725	2456.448588	630.1111239	358.9141583	249.4743644	21.72260114	17.55185871	7.148212645	71.093	0.518	0.547	11.106

Figure 1 – Sample dataset with countries’ GHG/CO2 emissions from 1971-2022

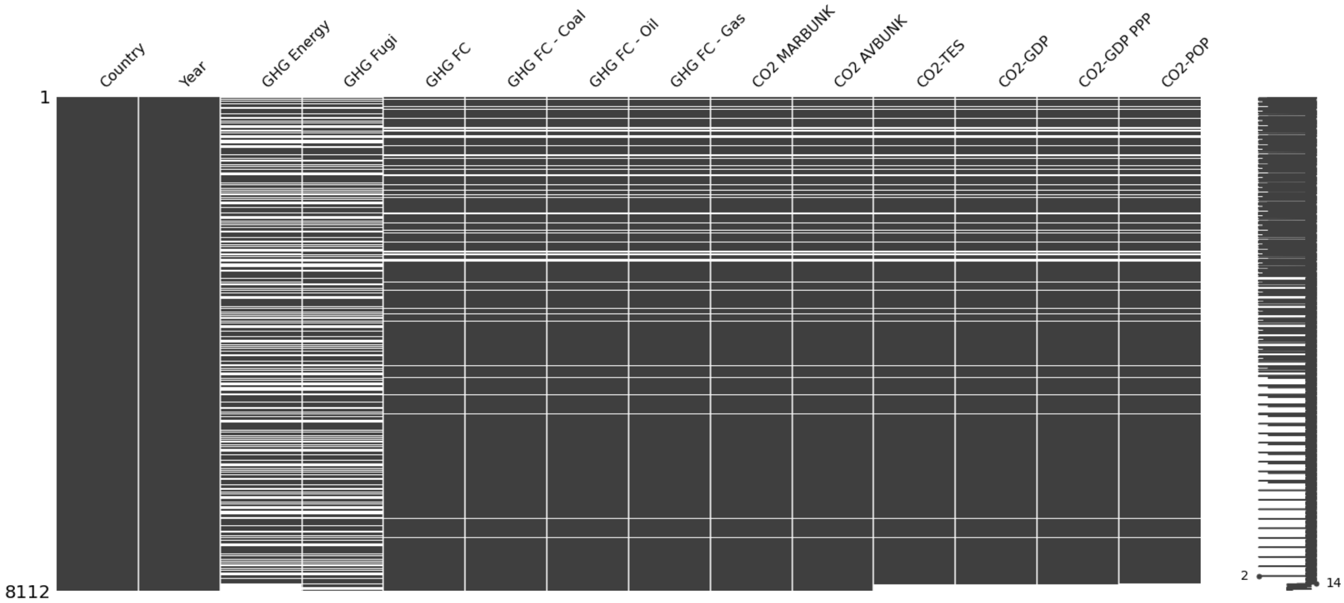


Figure 2 – A visualisation of the NA values present in the data frame

III. Training the data

Once the GHG/C02 emissions dataset was cleaned, it was required to reduce the multidimensional dataset to a two-dimensional space before visualisation. This was done by a Self-Organising Map (SOM), which is an unsupervised neural network that reduces the input dimensionality to represent its distribution as a map. Therefore, SOM forms a map where similar samples are clustered together [2]. The following attributes provided in the “Data” section of the report specifications for all countries were trained for each year after which they were animated to observe the trend in GHG/CO2 emissions from 1971 to 2022.

IV. Visually Grouping via similar countries

To see clustering within the dataset, a Self-Organizing map (SOM) was employed to reduce the dimensionality to 2 variables. The resulting graph shows clusters when all variables were included, but not applicable, that is, null data (NA) have been removed. The way our data was structured means that any year that had an NA will be removed for an entire country, which is a slight drawback. The mean result for each year was calculated for each country and plotted, which demonstrates the clustering over the period. Here we can clearly see the clusters of countries which give a great overview of the period. Additionally, an interactive visualization has been created which demonstrates the clusters found within each year if the user wants to take a more granular approach.

A. Axis arrangement and Visual variables

The x and y axis have been used by the two variables that are the reduced dimensions. This allows the viewer to clearly depict clustering. A visual variable of colour has been added to plot 1 to give the viewer some extra context of the total energy from GHG. We decided to leave the size of the points constant so it is easy to see where they are and which countries are close together.Text has been added to the first plot of the country name, to allow the viewer to instantly see which countries are grouped together. I have omitted the text from the second plot as the viewer is encouraged to hover over the cluster they are interested in to get the list of countries that are in it.The size of the point in the second plot is used to show how many countries are found in that 'cluster' and the interaction assists the analytical process as it enables the viewer to find out who are in the clusters and also how they change over time.

B. Analysis of visualisation 1

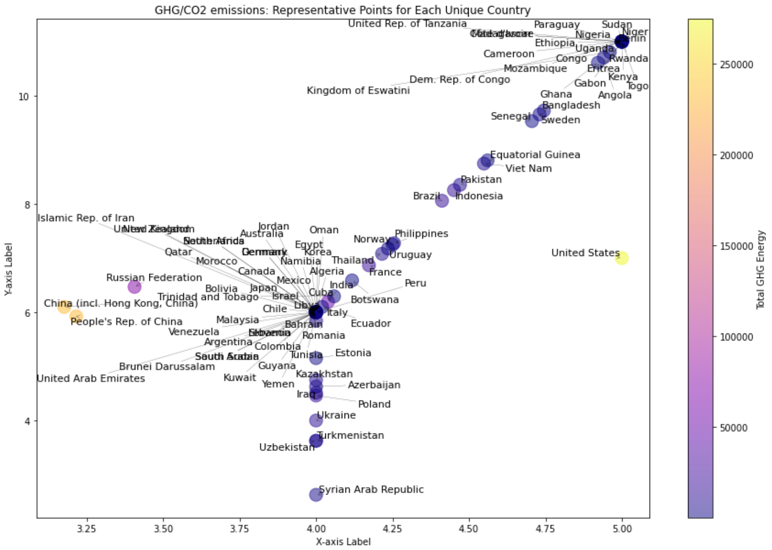


Figure 3 – Overall representative points for each country in respect to the 12 attributes

The above plot is the overall representation for each country. It is immediately noticed that there are some clusters in the middle and top right of the plot, as well as a small number of countries that don't fit exactly into the clusters and are in their own spots. The colouring of the point allows the viewers to also ascertain the scale of each countries greenhouse gas emissions and can play an important role in understanding the graph. This is an important part as the dimensions have been reduced from 12 to 2, the x and y axis can be hard to tell what exactly they mean in terms of absolute emissions.

C. Analysis of visualisation 2

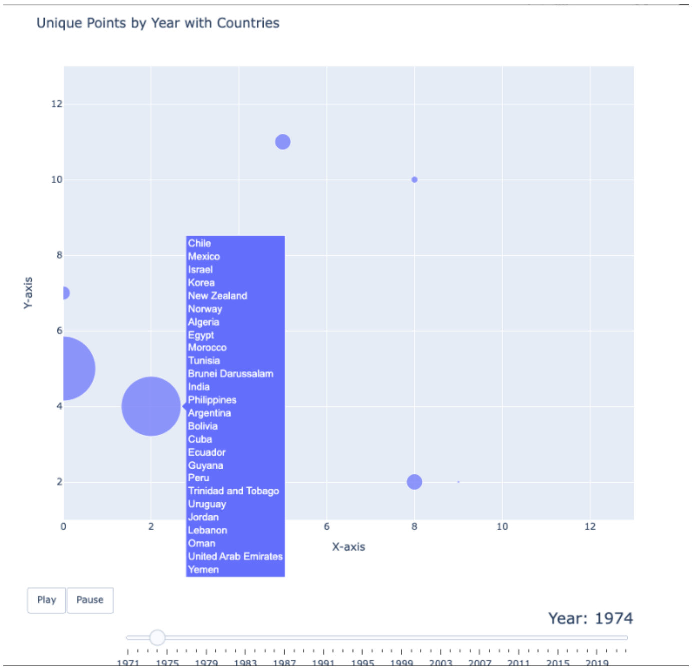


Figure 4 – Interactive plot to see the clustering within each year

This interactive visualization allows the viewer to dig deeper into the distribution of each country and see where each country is placed for each year. I have left the colouring off as it will distract the viewer from the main purpose behind the visualisation of spotting groups of similar countries with respect to the twelve attributes. In conjunction with the first plot, viewers can see where a country is in a certain year and then see groups of countries where they 'aimed' to be.

In summary, groupings both whole period and within the same year can be found clearly with the above visualisations. The interactive visualization allows the viewer to navigate between years of interest to dig deeper and the countries that are grouped together are easy to see.

V. VISUALLY GROUPING VIA SIMILAR TRENDS

A. Axis arrangement and Visual variables

The purpose of this plot is to explore, interactively, the aggregate values of all the socioeconomic measures of GHG/CO2 emissions in the y-axis amongst a subset of the countries across years in the x-axis. It is noteworthy that a separate feature 'GHG Total' was constructed to represent the aggregate sum of all the socioeconomic measures. Figure 1 demonstrates the trend of these socioeconomic measures, warranting a line chart to be selected as the framework. The 'Year' variable is ordinal, and the aggregate values are ratio data. To differentiate between different levels of emissions across countries, each country is represented with a different colour.

B. Analysis of visualisation

Evidence from various studies have indicated that energy accounts for more than three-quarters of total greenhouse gas emissions globally [3]. An effective analysis and exploration of regular and interactive visualisations will be imperative in determining trend and groups of similar countries in their emissions of GHG/CO2. Worldwide GHG emissions' data have 2 core purposes – first to identify the main cause of climate change and second, to monitor progress and achieve targets [4].

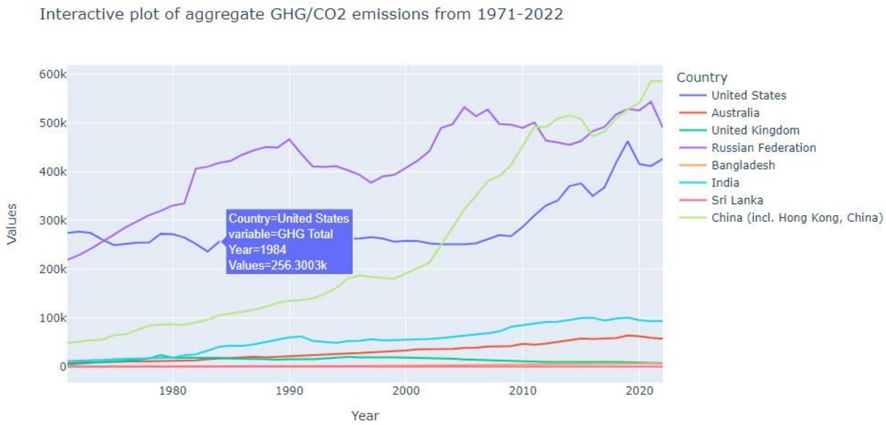


Figure 5 – Aggregate measures of GHG/CO2 emissions from 1971-2022

The above line chart showcases that China has the largest considerable GHG/CO2 emissions with an exponential growth between 2000 to 2022. Although this representation does not provide clear evidence of similar countries in their performance of GHG/CO2 emissions, this plot provides an initial framework of our approach in trend analysis using visual analytical techniques. From figure 1, we can theoretically group China, Russian Federation and United States based on their similar performance, however it's difficult to group these countries based on trend, as the historical data suggests a different pattern with China having far less emissions than Russia and United States. To view the physical emission data, we can hover above each trend line and view their corresponding GHG total value, as shown in figure 1 of the United States aggregate emission.

C. Trend analysis of clusters from 1971 – 2022

From the output of the trained SOM data, a few clusters of countries were identified, and their trends were observed over time. Figure 5 illustrates two groups of countries (labelled Cluster A and B) from the SOM training data.

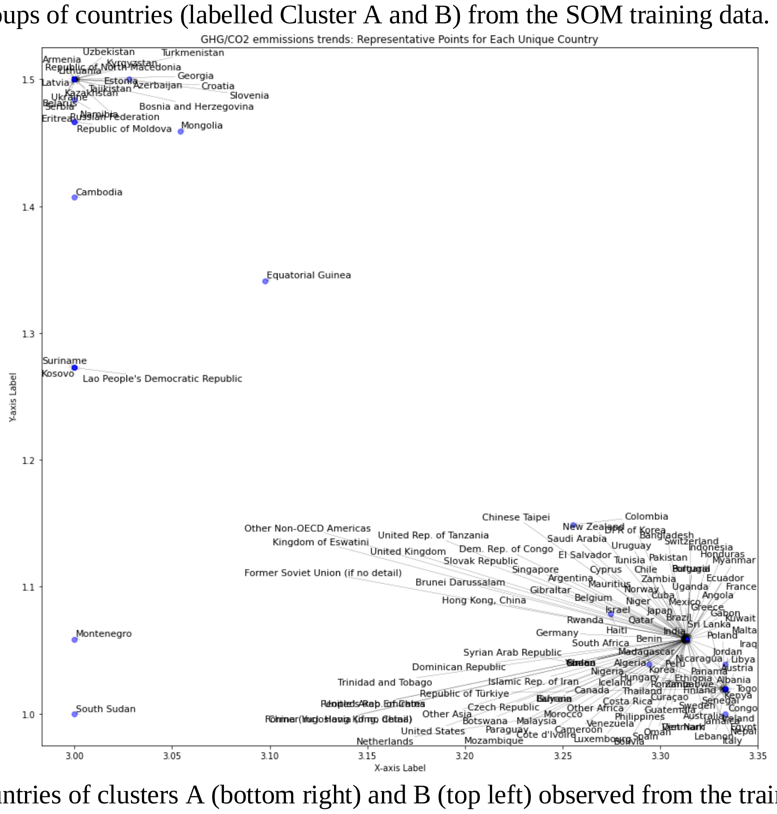


Figure 6 – Countries of clusters A (bottom right) and B (top left) observed from the trained SOM data.

The socioeconomic measures were plotted against time for clusters A and B in Tableau as illustrated in figures 6 and 7 respectively. Almost all countries of these clusters demonstrated a similar trend from 1971 to 2022, signifying the importance and high accuracy of the SOM's training methodology for this dataset.

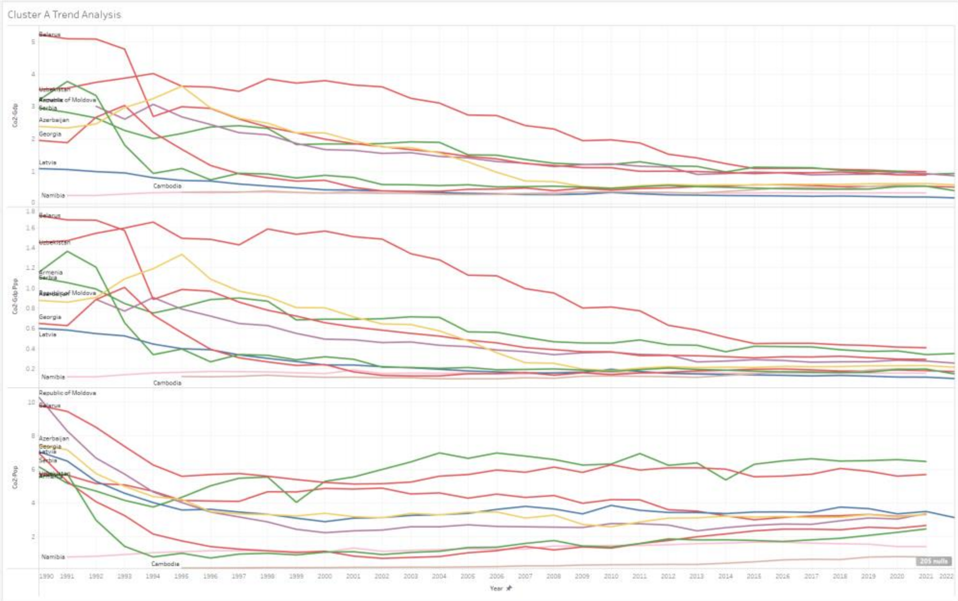


Figure 7 – GHG/CO2 emission trends of countries in Cluster A

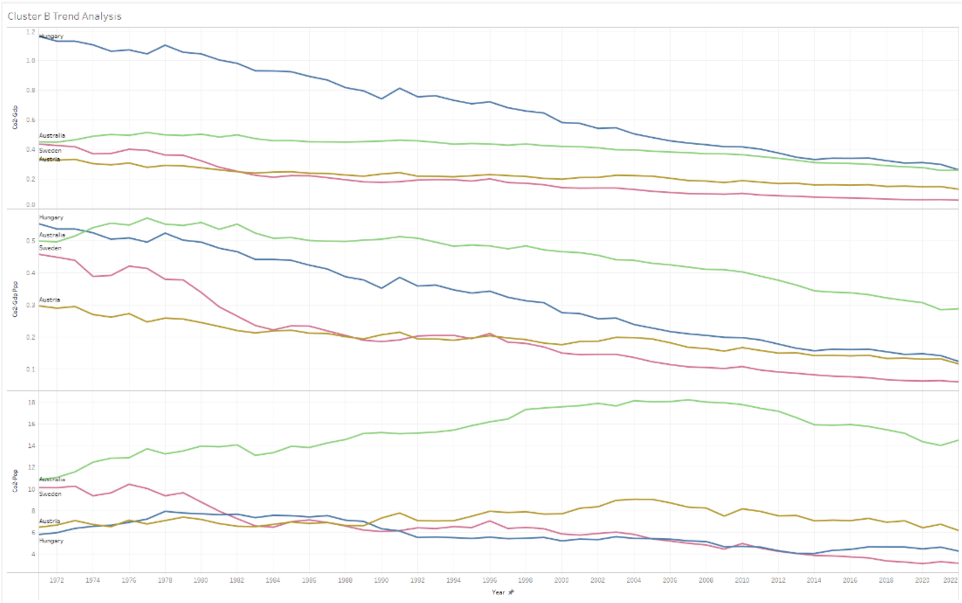


Figure 8 – GHG/CO2 emission trends of countries in Cluster B

VI. Improving Socioeconomic Measures in a Country's Profile

A. Data Cleansing and Pre-processing

In this section, we explore a country's profile from the year which yielded the minimum GHG/CO2 emission and the year with the maximum GHG/CO2 emission. Firstly, the dataset as shown in figure 2 was further enhanced by filtering our chosen country as Australia as well as a variable 'GHG Total' explained in section III. Using Python code, we can easily determine the highest and lowest emission for Australia, where 1971 yields the lowest total GHG emission and 2019 the highest. In the below visualisations, these two years will be visually highlighted for an appropriate and easier analysis in planning gradual improvement on various measures.

B. Axis arrangement and Visual variables

The below 4 plots visualises the GHG and CO2 emissions profile for Australia, however we have grouped these socioeconomic measures into 3 categories – GHG emissions from fuel combustion, international emission and CO2 emissions. Each plot has been produced to provide a 'plan' to gradually improve these features in the scope of improving Australia's environment and tackle the challenges of climate change. The x-axis exemplifies the Year in which GHG/CO2 emissions were measured, which is classified as ordinal data. A stacked bar plot is effective for displaying the total quantity of GHG emissions based on the fuel combustion which is a contribution to each subgroup to the total GHG. Using distinct colours of the stacked bar plot, we can differentiate the difference between each emission source. A blue colour scheme was utilised, where the largest contributor to GHG emissions is much darker compared to emission sources which are contributing much less to these emissions overall. Furthermore, for the scope of our analysis, the graph in figure 8 contains a red mark in the bottom of each bar for 1971 and 2019, showcasing the lowest and highest GHG total emissions for Australia, respectively.

C. Analysis of visualisation - GHG emissions from Fuel Combustion (Coal, Oil, Gas)

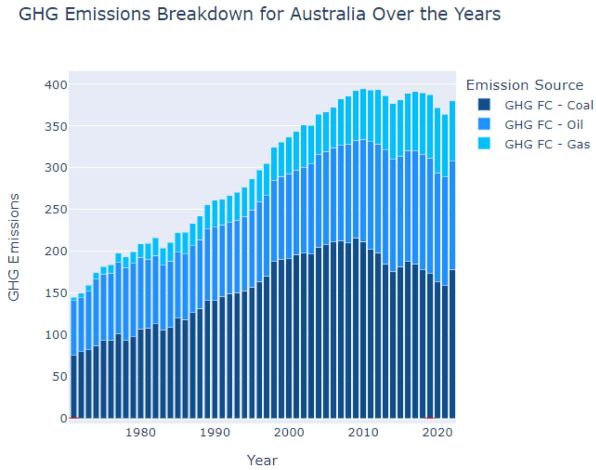


Figure 9 – GHG Emissions profile from fuel combustions for Australia

The above plot encompasses a stacked bar chart of total GHG emissions originating from the fuel combustion emission source. Upon inspection, coal contributes the most in terms of emissions. The year 1971 recorded the lowest coal combustion emissions, with the total GHG increasing over time till 2009 which reached its peak for coal fuel combustion emissions. Since then, the trend has become slightly negative. For oil fuel combustion emissions, we observe the second highest contributor compared to coal and gas. In terms of trend, from 1971 was again the lowest recorded measure with 65.28 and the highest was in 2013, comprising of 136.06 of oil combustion emission. Finally, gas contributed to the least GHG emissions, but it's noteworthy that the increase in gas emissions is far more drastic over time compared to the other sources. For example, in 1971, only 4.00 amount of gas was recorded and in 2019 it was recorded to be 75.45. This dramatic increase highlights the significance of industries and the Australian government to be vigilant about the consumption and use of gas within society to mitigate these emissions and plan for a more sustainable change.

D. Analysis of visualisation – CO2 emissions from international (MARBUNK and AVBUNK)

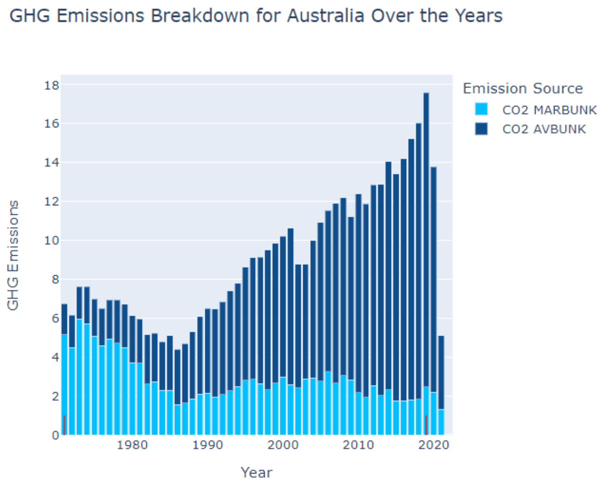
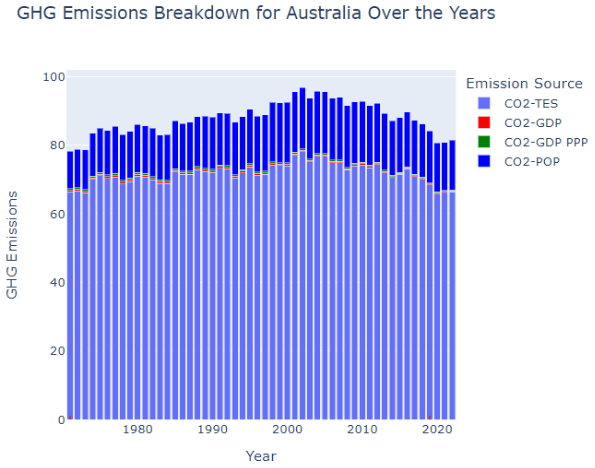


Figure 10 – International CO2 emissions profile for Australia

The GHG emissions breakdown for the international CO2 sources are showcased in figure 10 for Australia. The red vertical line in the 1971 and 2019 bars are to indicate the lowest and highest recorded emissions as per our objective. The emission source which contributes the most, is coloured in a darker shade of blue to indicate this visually to the viewer. For CO2 AVBUNK, a portmanteau for Aviation bunkers, the highest recorded emission is 15.11 in 2019 and the lowest is 1971 with 1.59. In terms of trend, CO2 pollution from 1971 till 1986 remains stable until there is a considerable increase till 2019. In addition, between 1971 and 2019, the largest recorded emission for CO2 MARBUNK, a portmanteau for CO2 Marine bunkers, is in 1973 and the lowest is in 1986. An overall downward trend is observed however a small increase is measured in the early 2000s till the bars level off to around 2 in terms of GHG emissions. For climate change activists and workers in the aviation industry, to gradually decrease these two groups of measures, they need to concentrate their efforts in decreasing CO2 emissions from Aviation bunkers and continue to maintain their levels in Marine bunkers.

D. Analysis of visualisation – CO2 emissions breakdown for Australia



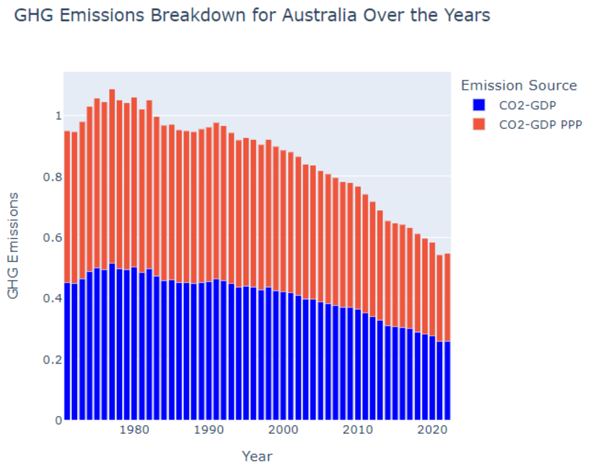


Figure 11 – CO2 emissions breakdown for Australia per source

Figure 12 – CO2-GDP/PPP emissions breakdown for Australia

To begin our analysis of CO2 measures, we scrutinise the trends from figure 11, where CO2-TES and CO2-POP dominates the chart, making it difficult to understand the patterns to analyse for CO2-GDP and CO2-GDP PPP. To cater our analysis for these two variables, we have constructed another stacked bar plot in figure 12. The highest CO2-POP emission is in 2002 with 17.90 and the lowest is in 1971 with 10.86 value of emission. When observing the trends for CO2-TES, the highest recorded emission is 78.01 value for 2002, whereas the lowest is in 1971 with 66.34. Although there are no noteworthy increases or decreases between 1971 to 2019, the highest and lowest pertaining to each measure being the same is an interesting observation which will be useful for the Australian government in understanding and potentially clustering the variables associated with the highest recorded mission being in 2002. Grouping CO2-TES and CO2-POP will be useful in planning to gradually decrease these two emissions for a lower trajectory in the future for Australia. In figure 12, we clearly observe a negative relationship between the aggregate GHG emissions from CO2-GDP and CO2-GDP PPP, which is different to any visualisation produced in previous sections. A lower CO2-GDP ratio generally considered more favourable for Australia, as it suggests that the economy is producing less carbon emissions per unit of economic activity. As for CO2-GDP PPP, this representation is useful when comparing the environmental efficiency of different countries, as it accounts for the economic disparities and cost of living variations that may exist between them. Figure 11 thus provides a more nuanced and internationally comparable perspective which seems less prevalent to ‘plan’ to gradually decrease these metrics, Australia would require to maintain this pattern over-time.

VII. Final visualisastion to improve each socioeconomic measure for Asutralia

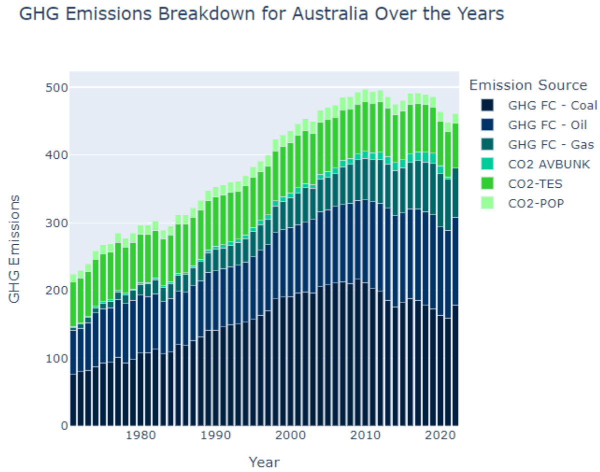


Figure 13 – Socioeconomic measures to improve for Australia

Thorough analysis from previous sections has yielded this stack bar plot which will be useful for climate change activists, workers in the industry who are producing these high emissions as well as the Australian government to initiate actions for greater sustainability. The choice of these emission sources as shown in figure 13 was decided based on the increase in trend of these measures over the years. The interactivity of this plot will provide information for the viewer and stakeholder to understand the variability in each emission source and thus construct a plan to reduce these gradually.

VIII. Evaluation

A. Evaluation Method

The assessment method is “Analytic Examination”. We refer to the “Benefits of Visualization” in the course and made sure all points are preserved in our visualization. Also, we use the “Gestalt principle” to explain how we use it to make our visualization better.

B. Evaluation of each Visualization

1) *Evaluation of Question1 Visualization:* We used SOM to reduce dimensionality and generate the plot. It helps to offload work to our perception system and makes us can find similar coutry groups easily and quickly. We also used Gestalt Laws to review our visualization results. It can be seen that although the points in the figure have the same color, we can still easily identify the same group because similar entities are closely connected. In order to make the plot more readable, we apply the principle of closure to have all countries in the same group in one dot. Also, this is an interaction chart that if you hover over each point, it tells you about the countries within the cluster and how big it is, and we also use animation to show the changes over time.
2) *Evaluation of Question2 Visualization*: In this part, we need to observe trends in similar countries. Therefore, we use the X-axis to represent each year, the Y-axis to represent the aggregate values, and observe the trend over the years with line chart. We use different colors to represent different countries. With the lines, we can observe the trends over the years in each country and this is an interaction chart for us to get detailed information of each country in each year easily.
3) *Evaluation of Question3 Visualization*: To compare the measures changes over time, we used stacked bar chart. We use X-asis as years and Y-axis to represent the value of GHG emission. In addition, we use darkness of color to represent different indicators, largest contributor to GHG emissions is much darker than others. To indicate the lowest and highest emission years, we mark them with red lines.

TABLE 1. Result of Evaluation and Opportunities for Improvement

References

[1] How do countries measure greenhouse gas emissions?. UN environment programme. (2022, September 13). https://www.unep.org/news-and-stories/story/how-do-countries-measure-greenhouse-gas-emissions

[2] Sacco, G. Motta, L.-l. You, N. Bertolazzo, F. Carini, T.-y. Ma, Smart cities, urban sensing, and big data: mining geo-location in social networks. Big Data and Smart Service Systems (2017,) Chapter 5, Pages 59-84

[3] Greenhouse Gas Emissions from Energy Data Explorer. iea. (2023, August 2). https://www.iea.org/data-and-statistics/data-tools/greenhouse-gas-emissions-from-energy-data-explorer

[4] Busch, T., Cho, C.H., Hoepner, A.G.F. et al. Corporate Greenhouse Gas Emissions’ Data and the Urgent Need for a Science-Led Just Transition: Introduction to a Thematic Symposium. J Bus Ethics 182, 897–901 (2023). https://doi.org/10.1007/s10551-022-05288-7