VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY- UNIVERSITY OF SCIENCE
FACULTY OF INFORMATION TECHNOLOGY

March 24, 2025

Data Visualization

Report lab 02

Lecturers and Teaching assistants :

1. Mr. Bui Tien Len
2. Mr. Vo Nhat Tan

Students:

1. **Quach Tran Quan Vinh - 22127460**
2. **Nguyen Hoang Trung Kien - 22127478**

Contents

1 Project and team progress

1.1 Overall progress

#	**Tasks**	**Progress**
1	Data collection	100%
2	Data preprocessing	100%
3	Data visualization	100%

1.2 Team progress

#	**Student's ID**	**Name**	**Tasks**	**Progress**
1	22127460	Quach Tran Quan Vinh	Data collection	100%
2	22127478	Nguyen Hoang Trung Kien	Data preprocessing	100%
3	22127478	Nguyen Hoang Trung Kien	Data exploration	100%
4	22127478	Nguyen Hoang Trung Kien	Solve analysis objectives 1, 2 and 3	100%
5	22127460	Quach Tran Quan Vinh	Solve analysis objectives 4, 5 and 6	100%
6	22127460 and 22127478	Quach Tran Quan Vinh and Nguyen Hoang Trung Kien	Write report	100%

2 Implement steps

2.1 Importing libraries

Libraries to handle and visualize data.

- `matplotlib`, `seaborn`, `plotly` - used to visualize data.
- `pandas` - used to process data in table.
- `numpy` - interact with array more efficiently.
- `math` - get π constant to plot radar chart.
- `geopandas` - use world map to visualize data.

Libraries to handle missing values by algorithm.

- `enable_iterative_imputer` - to enable iterative imputer
- `IterativeImputer` - Iterative Imputer algorithm. `StandardScaler` - normalize data by standard scaling method.
- `LabelEncoder` - encode data by label encoder.
- `RandomForestRegressor` - Random Forest algorithm.

2.2 Data collection

Data is collected from World Development Indicators. We collect features based on environmental status around the world of all countries in 1990, 2000 and 2014 - 2023.

2.3 Data preprocessing

The raw data is in excel format, we read it into pandas. There are two final lines of the data has data information, so we will drop those two lines. We converted each Series Name value into each column. And the raw data now has 3192 rows and 34 columns.
Firstly, we check and saw that there are no duplicates so no need to handle this. Through the dataset, some missing values instead of being represented by NaN, they are represented by '..', so we converted them into NaN. Then, we dealt with missing values, here is the number of missing values of each columns:

Indicator	**Missing Values**
Country Name	0
Country Code	0
Year	0
Access to clean fuels and technologies for cooking (% of population)	822
Access to clean fuels and technologies for cooking, rural (% of rural population)	822
Access to clean fuels and technologies for cooking, urban (% of urban population)	822

Columns	Missing Values count
Adjusted net savings, including particulate emission damage (% of GNI)	1391
Adjusted savings: carbon dioxide damage (% of GNI)	815
Adjusted savings: consumption of fixed capital (% of GNI)	806
Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal)	1122
Annual freshwater withdrawals, domestic (% of total freshwater withdrawal)	1123
Annual freshwater withdrawals, industry (% of total freshwater withdrawal)	1139
Annual freshwater withdrawals, total (% of internal resources)	1140
Annual freshwater withdrawals, total (billion cubic meters)	1102
Carbon dioxide (CO2) emissions (total) excluding LULUCF (Mt CO2e)	180
Combustible renewables and waste (% of total energy)	2585
Forest area (% of land area)	367
Forest area (sq. km)	367
GDP (current US$)	138
Level of water stress: freshwater withdrawal as a proportion of available freshwater resources	1510
Methane (CH4) emissions (total) excluding LULUCF (Mt CO2e)	180
Nitrous oxide (N2O) emissions (total) excluding LULUCF (Mt CO2e)	180
People using at least basic drinking water services, rural (% of rural population)	1054
People using at least basic drinking water services, urban (% of urban population)	932
Renewable electricity output (% of total electricity output)	2128
Renewable energy consumption (% of total final energy consumption)	587
Renewable internal freshwater resources per capita (cubic meters)	914
Renewable internal freshwater resources, total (billion cubic meters)	926
Terrestrial and marine protected areas (% of total territorial area)	1425
Terrestrial protected areas (% of total land area)	1423
Total greenhouse gas emissions excluding LULUCF (Mt CO2e)	180
Total greenhouse gas emissions excluding LULUCF per capita (t CO2e/capita)	180
Urban land area (sq. km)	2403

Columns	Missing Values count
Water productivity, total (constant 2015 US$ GDP per cubic meter of total freshwater withdrawal)	1038

We can see that there are some columns which have high quantity of missing records. We dropped the columns that have over 50% missing values.
To handle missing values, we first identified which columns have below or above 10% missing values. If columns have low missing values, we just filled them by median. The other we filled them by Iterative Imputer with Random Forest Regressor estimator.
Reasons:

- Low missing columns' distribution will not be affected much when filling by median.
- Most of the columns have high missing values, and their distribution may vary.
- The columns are continuous variable so it is suitable to fill with Random Forest Regressor estimator.

We convert and make sure year is in int type first. Then we encoded Country Name and Country Code using LabelEncoder, converting category data into numerical ordinal data ([0, 1, 2...]). They are not ordinal data, but to enhance the process of filling missing values' runtime and space, we use it for simplicity. Using this instead of One hot Encoding makes the Iterative Imputer runs faster since the core of that method is to use loops through columns if it is just for filling missing values process. We calculated missing ratio of each columns and define whether they are high or low missing columns, with threshold is 10%, then filled low missing columns by median. And we also scaled columns using Standard Scaler to fill the other columns by Iterative Imputer.
We filled the remain missing columns by Iterative Imputer with Random Forest Regressor estimator.
How Iterative Imputer works: [1]

1. Initialization: The algorithm starts by initializing the missing values with a random or mean imputation.
2. Feature Selection: The algorithm selects a feature to impute, typically in a round-robin fashion.
3. Imputation: The selected feature is imputed using a regression model, which predicts the missing values based on the observed values of the other features.
4. Update: The imputed values are updated, and the process is repeated for the next feature.
5. Convergence: The algorithm continues until convergence, which is typically determined by a stopping criterion such as a maximum number of iterations or a tolerance threshold.

How Random Forest Regressor works:[2]
Random Forest Regression works by creating multiple of decision trees each trained on a random subset of the data. The process begins with Bootstrap sampling where random rows of data are selected with replacement to form different training datasets for each tree. After this we do feature sampling where only a random subset of features is used to build each tree ensuring diversity in the models.
After the trees are trained each tree make a prediction and the final prediction for regression tasks is the average of all the individual tree predictions.
We chose:
1. Iterative Imputer:

- `max_iter = 5`: Runs the imputation process for 5 iterations, balancing accuracy and speed.
- `random_state = 42`: Ensures consistent filling results.

2. Random Forest Regressor:

- `n_estimators = 30`: Uses 30 decision trees, balancing accuracy and speed.
- `n_job = -1`: Uses all available CPU cores for faster processing.

After all the preprocessing steps above, when looking through the data, we saw that Year column is in float type, we converted it into int type.
Finally we stored the preprocessed data into csv.

2.4 Data exploration

The data now has 3192 rows and 31 columns.
Columns and their meanings:

#	**Column Name**	**Description**
1	Country Name	Country's name
2	Country Code	Country's Code
3	Year	Year
4	Access to clean fuels and technologies for cooking (% of population)	Access to clean fuels and technologies for cooking is the proportion of total population primarily using clean cooking fuels and technologies for cooking. Under WHO guidelines, kerosene is excluded from clean cooking fuels
5	Access to clean fuels and technologies for cooking, rural (% of rural population)	Access to clean fuels and technologies for cooking, rural is the proportion of rural population primarily using clean cooking fuels and technologies for cooking. Under WHO guidelines, kerosene is excluded from clean cooking fuels
6	Access to clean fuels and technologies for cooking, urban (% of urban population)	Access to clean fuels and technologies for cooking, urban is the proportion of urban population primarily using clean cooking fuels and technologies for cooking. Under WHO guidelines, kerosene is excluded from clean cooking fuels
7	Adjusted net savings, including particulate emission damage (% of GNI)	Adjusted net savings are equal to net national savings plus education expenditure and minus energy depletion, mineral depletion, net forest depletion, and carbon dioxide and particulate emissions damage
8	Adjusted savings: carbon dioxide damage (% of GNI)	Cost of damage due to carbon dioxide emissions from fossil fuel use and the manufacture of cement, estimated to be US$40 per ton of CO2 (the unit damage in 2017 US dollars for CO2 emitted in 2020) times the number of tons of CO2 emitted

#	Column Name	Description
9	Adjusted savings: consumption of fixed capital (% of GNI)	Consumption of fixed capital represents the replacement value of capital used up in the process of production
10	Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for agriculture are total withdrawals for irrigation and livestock production. Data are for the most recent year available for 1987-2002
11	Annual freshwater withdrawals, domestic (% of total freshwater withdrawal)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for domestic uses include drinking water, municipal use or supply, and use for public services, commercial establishments, and homes. Data are for the most recent year available for 1987-2002
12	Annual freshwater withdrawals, industry (% of total freshwater withdrawal)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for industry are total withdrawals for direct industrial use (including withdrawals for cooling thermoelectric plants). Data are for the most recent year available for 1987-2002

#	Column Name	Description
13	Annual freshwater withdrawals, total (% of internal resources)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for agriculture and industry are total withdrawals for irrigation and livestock production and for direct industrial use (including withdrawals for cooling thermoelectric plants). Withdrawals for domestic uses include drinking water, municipal use or supply, and use for public services, commercial establishments, and homes. Data are for the most recent year available for 1987-2002
14	Annual freshwater withdrawals, total (billion cubic meters)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for agriculture and industry are total withdrawals for irrigation and livestock production and for direct industrial use (including withdrawals for cooling thermoelectric plants). Withdrawals for domestic uses include drinking water, municipal use or supply, and use for public services, commercial establishments, and homes. Data are for the most recent year available for 1987-2002
15	Carbon dioxide (CO2) emissions (total) excluding LULUCF (Mt CO2e)	A measure of annual emissions of carbon dioxide (CO2), one of the six Kyoto greenhouse gases (GHG), from the agriculture, energy, waste, and industrial sectors, excluding LULUCF. The measure is standardized to carbon dioxide equivalent values using the Global Warming Potential (GWP) factors of IPCC's 5th Assessment Report (AR5)
16	Forest area (% of land area)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens

#	Column Name	Description
17	Forest area (sq. km)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens
18	GDP (current US$)	GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used
19	Level of water stress: freshwater withdrawal as a proportion of available freshwater resources	The level of water stress: freshwater withdrawal as a proportion of available freshwater resources is the ratio between total freshwater withdrawn by all major sectors and total renewable freshwater resources, after taking into account environmental water requirements. Main sectors, as defined by ISIC standards, include agriculture; forestry and fishing; manufacturing; electricity industry; and services. This indicator is also known as water withdrawal intensity
20	Methane (CH4) emissions (total) excluding LULUCF (Mt CO2e)	A measure of annual emissions of methane (CH4), one of the six Kyoto greenhouse gases (GHG), from the agriculture, energy, waste, and industrial sectors, excluding LULUCF. The measure is standardized to carbon dioxide equivalent values using the Global Warming Potential (GWP) factors of IPCC's 5th Assessment Report (AR5)
21	Nitrous oxide (N2O) emissions (total) excluding LULUCF (Mt CO2e)	A measure of annual emissions of nitrous oxide (N2O), one of the six Kyoto greenhouse gases (GHG), from the agriculture, energy, waste, and industrial sectors, excluding LULUCF. The measure is standardized to carbon dioxide equivalent values using the Global Warming Potential (GWP) factors of IPCC's 5th Assessment Report (AR5)

#	Column Name	Description
22	People using at least basic drinking water services, rural (% of rural population)	The percentage of people using at least basic water services. This indicator encompasses both people using basic water services as well as those using safely managed water services. Basic drinking water services is defined as drinking water from an improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes or tubewells, protected dug wells, protected springs, and packaged or delivered water
23	People using at least basic drinking water services, urban (% of urban population)	The percentage of people using at least basic water services. This indicator encompasses both people using basic water services as well as those using safely managed water services. Basic drinking water services is defined as drinking water from an improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes or tubewells, protected dug wells, protected springs, and packaged or delivered water
24	Renewable energy consumption (% of total final energy consumption)	Renewable energy consumption is the share of renewables energy in total final energy consumption
25	Renewable internal freshwater resources per capita (cubic meters)	Renewable internal freshwater resources flows refer to internal renewable resources (internal river flows and groundwater from rainfall) in the country. Renewable internal freshwater resources per capita are calculated using the World Bank's population estimates
26	Renewable internal freshwater resources, total (billion cubic meters)	Renewable internal freshwater resources flows refer to internal renewable resources (internal river flows and groundwater from rainfall) in the country
27	Terrestrial and marine protected areas (% of total territorial area)	Terrestrial protected areas are totally or partially protected areas of at least 1,000 hectares that are designated by national authorities as scientific reserves with limited public access, national parks, natural monuments, nature reserves or wildlife sanctuaries, protected landscapes, and areas managed mainly for sustainable use. Marine protected areas are areas of intertidal or subtidal terrain—and overlying water and associated flora and fauna and historical and cultural features—that have been reserved by law or other effective means to protect part or all of the enclosed environment. Sites protected under local or provincial law are excluded

#	Column Name	Description
28	Total greenhouse gas emissions excluding LULUCF (Mt CO2e)	A measure of annual emissions of the six greenhouse gases (GHG) covered by the Kyoto Protocol (carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulphurhexafluoride (SF6)) from the energy, industry, waste, and agriculture sectors, standardized to carbon dioxide equivalent values. This measure excludes GHG fluxes caused by Land Use Change Land Use and Forestry (LULUCF), as these fluxes have larger uncertainties. The measure is standardized to carbon dioxide equivalent values using the Global Warming Potential (GWP) factors of IPCC's 5th Assessment Report (AR5)
29	Total greenhouse gas emissions excluding LULUCF per capita (t CO2e/capita)	Total annual emissions of the six greenhouse gases (GHG) covered by the Kyoto Protocol (carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulphurhexafluoride (SF6)) from the energy, industry, waste, and agriculture sectors, standardized to carbon dioxide equivalent values divided by the economy's population. This measure excludes GHG fluxes caused by Land Use Change Land Use and Forestry (LULUCF), as these fluxes have larger uncertainties
30	Water productivity, total (constant 2015 US$ GDP per cubic meter of total freshwater withdrawal)	Water productivity is calculated as GDP in constant prices divided by annual total water withdrawal

2.5 Data visualization and analysis

2.5.1 Visualization objective 1:

Question: How gases emission (CO2, N2O and CH4) affects the greenhouse effect?
Benefits:

- Helps climate researchers understand the causes of climate change, which is crucial for developing effective mitigation strategies.
- Supports industries and businesses in understanding their impact and adopting cleaner technologies.

Columns used in data:

- 'Year'
- 'Carbon dioxide (CO2) emissions (total) excluding LULUCF (Mt CO2e)'
- Methane (CH4) emissions (total) excluding LULUCF (Mt CO2e)

- Nitrous oxide (N2O) emissions (total) excluding LULUCF (Mt CO2e)‘
- Total greenhouse gas emissions excluding LULUCF (Mt CO2e)‘

Visualization used: Area chart.
Reason: To show the trend over time, highlight contribution of different gases and helps understand the total emissions while breaking it down by gas type.
Analyzing and Visualizing steps:
First, we extracted gases emission and total greenhouse gas emissions column. We filtered the data with Country Name is world to analyze greenhouse effect of the world. There are still other gas types, so we calculated by substracting Total Emissions with sum of the emissions of the gas types that we currently had.
We plotted the quantity of Total emissions and emissions of different types of gases throughout the year using area chart.

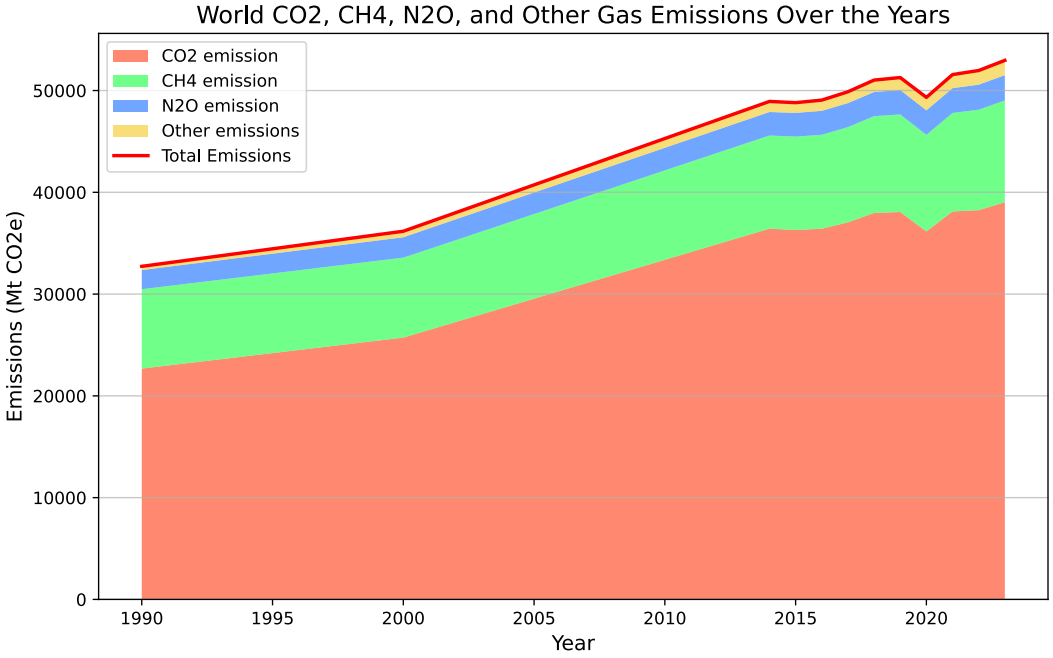

Figure 1: World CO2, CH4, N2O, and Other Gas Emissions Over the Years

From the area chart, we can clearly see that Total Emissions and the gas types rise from 1990 to 2023. CO2 emissions takes the highest proportion throughout the years. All the emissions seem to have the same trend.
Additionally, the world faces the presence of the other emissions types.
Conclusion:
The continuous increase in CO2 levels suggests that human activities, such as fossil fuel combustion, industrial expansion, and deforestation, remain major contributors to climate change.
Furthermore, the presence of “‘Other Emissions“‘ highlights additional greenhouse gases that also contribute to climate change. While they make up a smallest portion of total emissions, their long-term effects should not be overlooked.

2.5.2 Visualization objective 2:

Question: Which regions or continents contribute the most to global GHG emissions in 2023?
Benefits:

- Assists international organizations in prioritizing emission reduction policies and global climate agreements.
- Helps researchers analyze trends in emissions and develop regional strategies for sustainable development.

Columns used in data:

- Total greenhouse gas emissions excluding LULUCF (Mt CO2e)
- Country Name
- Country Code
- GDP (current US$)
- Renewable energy consumption (% of total final energy consumption)

Visualizations used: World heatmap and line graph
Reasons:

- For world heatmap: There are a lot of countries, we cannot just visualize them on normal chart since it is hard to make further analysis on a plot that has many different elements. To identify the high emissions area and visualize global emissions distribution.
- For line graph: To show the trend of GDP and Renewable Energy Consumption over time.

Analyzing and Visualizing steps:
We want to visualize this on a world map, so our strategy is to load a shapefile containing world country boundaries into a GeoDataFrame using the geopandas (gpd), and rename the column in it to match with our original data.
Link of the shapefile data: **[3]** We downloaded the data in Admin 0 - Countries section.
We then filtered the data in 2023 and plotted the Global greenhouse gas emissions in 2023 using world heatmap.

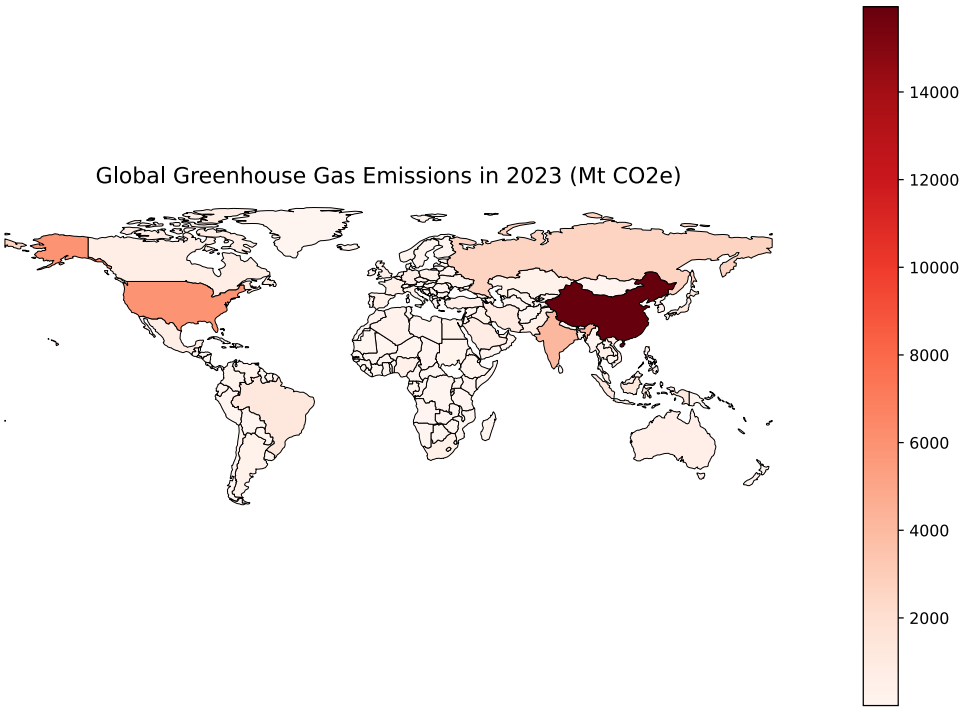

Figure 2: Global Greenhouse Gas Emissions in 2023 (Mt CO2e)

From the world heatmap, we can clearly observe that China, United States and India have the highest emissions. This is consistent with historical data and global reports, which indicate that these countries are the largest contributors.
But China seems to have highest greenhouse gas emissions. So we make further analysis by diving into China to find out the key factors leading to this.
Thinking about it, we all know that China is a country with high-level economy, with fast GDP growth. So we will analyze if this is true so that can understand how it contributes to its emissions. We plotted line graph to show the GDP growth of China.

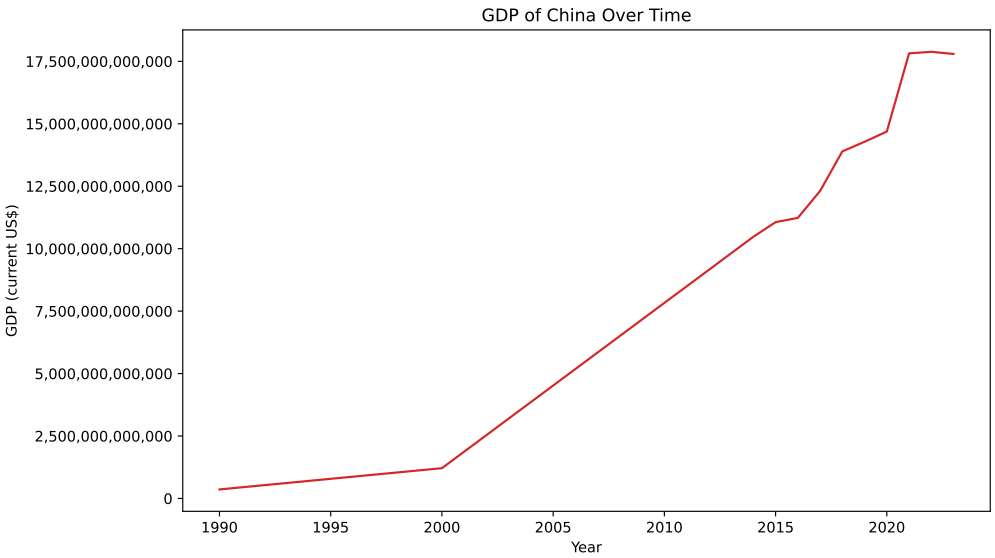

Figure 3: GDP of China Over Time

So it is the fact that China has fast GDP growth. It mostly rises significantly over time. Due to its growing economy, it may have high emissions by the presence of industrialization, urbanization, and increased energy consumption.
Another factor should be taken into consideration is China's renewable energy consumption. If it has highest emissions, maybe the policies of protecting environment and preserving energy is lack of effectiveness.
We plotted line graph to show the Renewable Energy Consumption of China.

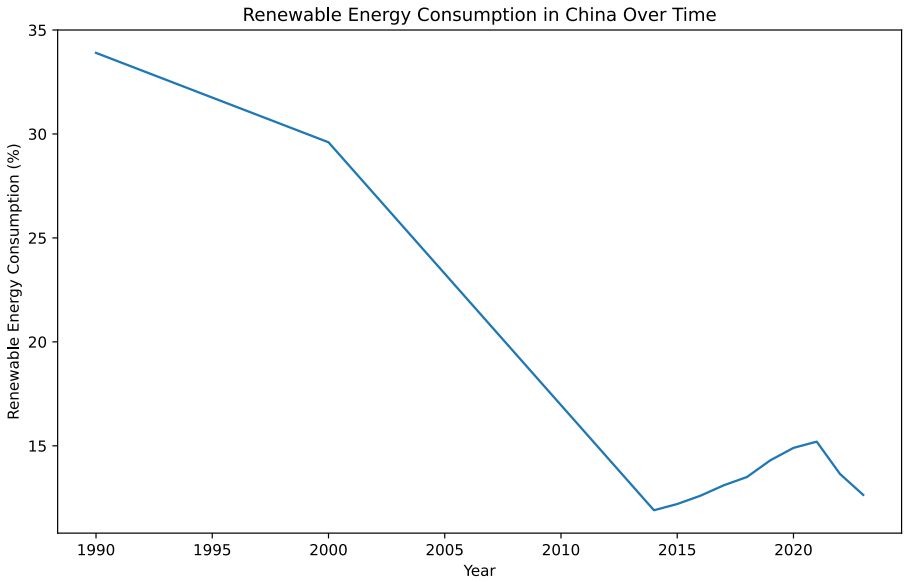

Figure 4: Renewable Energy Consumption of China Over Time

So it is true that Renewable energy consumption of China reduces over time. This decline may be contributed to China's dependence on fossil fuels, particularly coal, which remains a primary energy source for its industrial and economic growth.

Conclusion:
The high-ranked economy countries (especially China) are likely to contribute high greenhouse gas emissions. As their economies grow, they use more energy for industry, transportation, and urban development, leading to more pollution.
China's rapid GDP growth has increased its energy demand, relying heavily on coal and fossil fuels. Although the country invests in renewable energy, it may not be enough to keep up with overall energy consumption.
To reduce emissions, high-emission countries need to focus on clean energy, better efficiency, and stronger environmental policies for a more sustainable future.

2.5.3 Visualization objective 3:

Question: What is the status of deforestation over years around the world?
Benefits:

- Supports conservation efforts by identifying areas at high risk of forest loss.
- Helps governments and NGOs measure the impact of deforestation on climate change, biodiversity, and local communities.

Columns used in data:

- Forest area (% of land area)
- Forest area (sq. km)
- Year
- Country Name
- Country Code

Visualizations used: World heatmap, line graph and horizontal bar chart
Reasons:

- For line graph: To show the Forest area over time.
- For world heatmap: World heatmap helps to highlight the percentage that forest area cover all over the world, also show where deforestation occurs the most.
- For horizontal bar chart: To make comparison clearer between 10 countries, effectively rank 10 countries.

Analyzing and Visualizing steps: We extracted the world data and plotted forest area over time of the world by line graph.

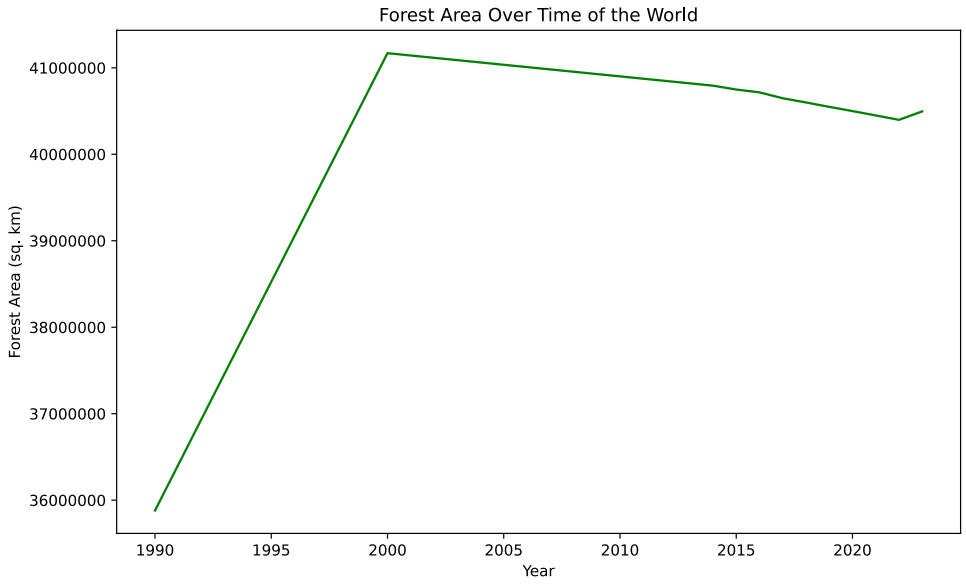

Figure 5: Forest Area Over Time of the World

As can be seen from the line graph, Forest Area rises from 1990 to 2000, then experiences a decline until after 2020. The decline in Forest Area indicates that deforestation occurs in that period.
The peak of the line graph is at 2000, and the nearest year of the data is 2023, so we compare the Forest Area (% of Land Area) of the world in these 2 years.
We compared Global Forest Area in 2000 with 2023 by using world heat map to visualize.

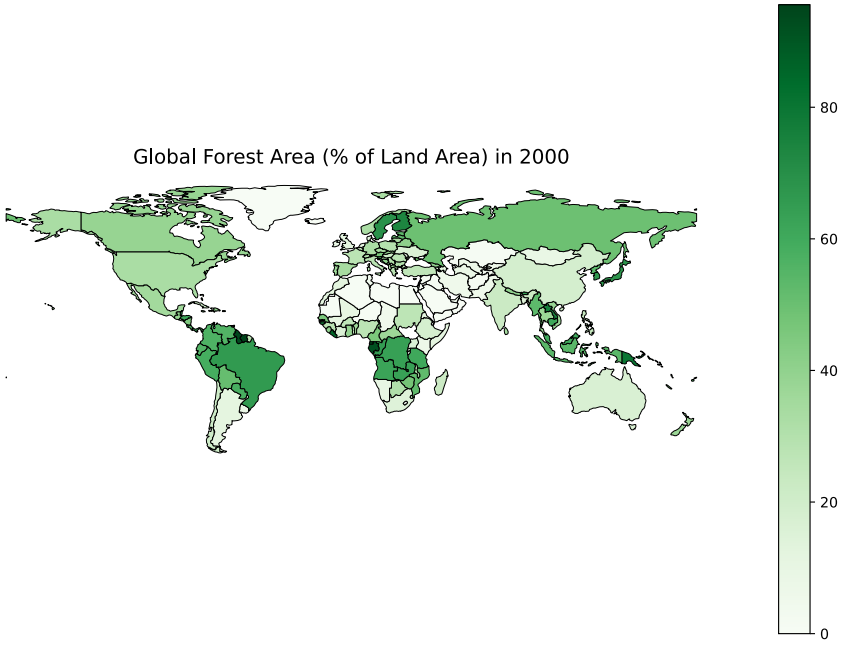

Figure 6: Global Forest Area (% of Land Area) in 2000

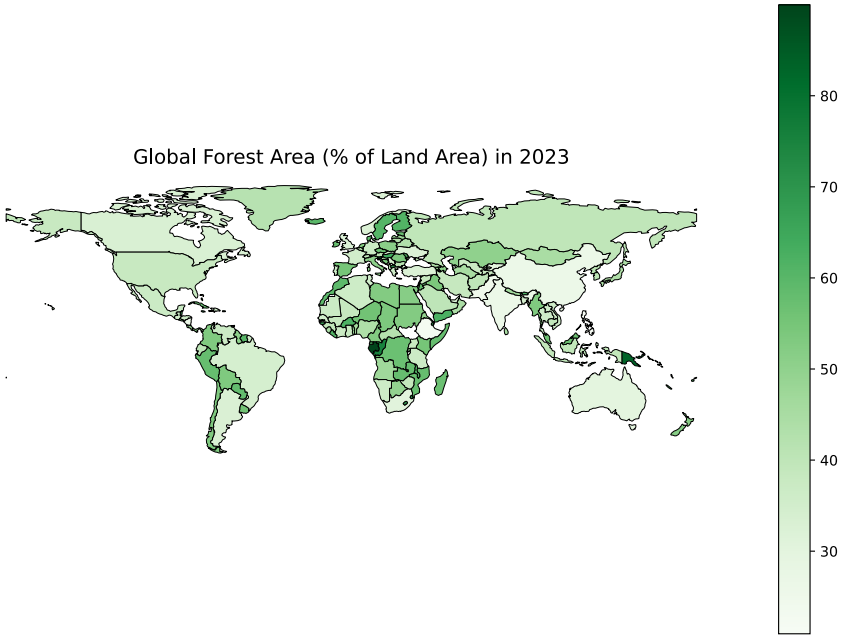

Figure 7: Global Forest Area (% of Land Area) in 2023

It is clear that world has experienced a huge reduction of forest area for the past 23 years. Deforestation mostly occurs in the large countries. This can be due to their population growth, industrialization and the need of the land space and forest based material.
However, some specific regions or countries still have risen area of forest, especially those in North Africa. This growth can be due to reforestation efforts, afforestation projects, or improved conservation policies.
To make further exploration, we identified and analyzed the top 10 countries that have highest forest loss rate over years. This can help to indicate if there are some specific regions that has high proportion of deforestation.
We extracted necessary columns, calculated the rate of deforestation change over years. Then, we calculated each country's mean of that rate and identify the top 10 countries by average deforestation rate.

Forest Loss Rate (%)

$$\text{Forest Loss Rate}(\%) = \left(\frac{\text{Forest Area}_{\text{previous year}} - \text{Forest Area}_{\text{current year}}}{\text{Forest Area}_{\text{previous year}}} \right) \times 100 \tag{1}$$

Average Deforestation Rate (%)

$$\text{Avg Deforestation Rate}(\%) = \frac{\sum \text{Forest Loss Rate (\%)}}{\text{Number of years}} \tag{2}$$

We plotted the top 10 countries' Forest Loss Rate by horizontal bar charts.

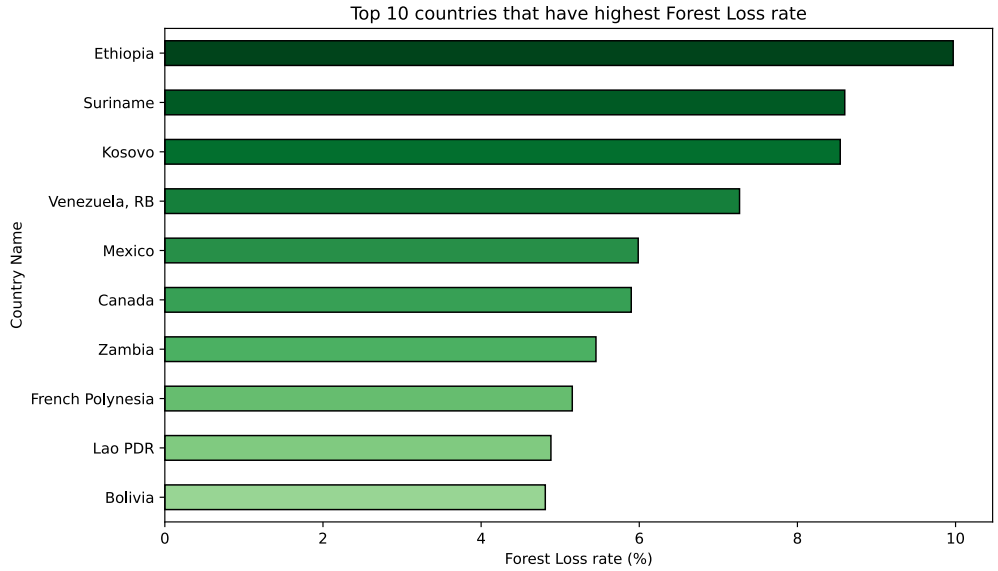

Figure 8: Top 10 countries that have highest Forest Loss rate

From the chart we can see that Ethiopia has the highest deforestation rate, while Bolivia is the lowest.
When tracking down these countries, we can see that deforestation occurs all over the world. This shows that deforestation is a global issue, affecting multiple continents, especially South America, Africa, and Asia.

Conclusion:

Deforestation has been a major global issue over the past 23 years, with a significant decline in forest

area since 2000.
While large countries face deforestation due to population growth and industrialization, some regions, like North Africa, have seen forest recovery through conservation efforts. Ethiopia has the highest deforestation rate, while Bolivia has the lowest. This highlights the widespread impact of deforestation, particularly in South America, Africa, and Asia.
Without immediate action, continued deforestation could lead to severe environmental consequences, including biodiversity loss, climate change, and disrupted ecosystems.

2.5.4 Visualization objective 4:

Question: How are environmental protection policies implemented in Southeast Asian countries, and what impact do they have on the greenhouse effect in these nations?
Benefits:

- Allows governments to learn from successful environmental policies and implement similar measures.
- Helps investors and businesses identify regions with strong sustainability practices for eco-friendly investments.

Columns used in data:

- Access to clean fuels and technologies for cooking (% of population)
- Renewable energy consumption (% of total final energy consumption)
- Total greenhouse gas emissions excluding LULUCF (Mt CO2e)
- Year
- Country Name

Visualizations used: Twin plot
Reasons: A twin plot is a visualization technique that combines two different charts within a single frame, allowing for the comparison of two related datasets with distinct scales. We use stacked bar chart and line chart

- For stacked bar chart: There are a lot of countries, we cannot just visualize them on normal chart since it is hard to make further analysis on a plot that has many different elements. It effectively show two different characteristics within the same figure, making it easier to compare them.
- For line chart: Additionally, combining a line chart with a bar chart helps visualize the relationship between features more clearly.

Analyzing and Visualizing steps:
Firstly, we define the Southeast Asian countries. After that we group the data by Country Name with the average value of *Access to Clean Fuels and Technologies* and *Renewable energy consumption.* Then we plot the data to the twin plot.

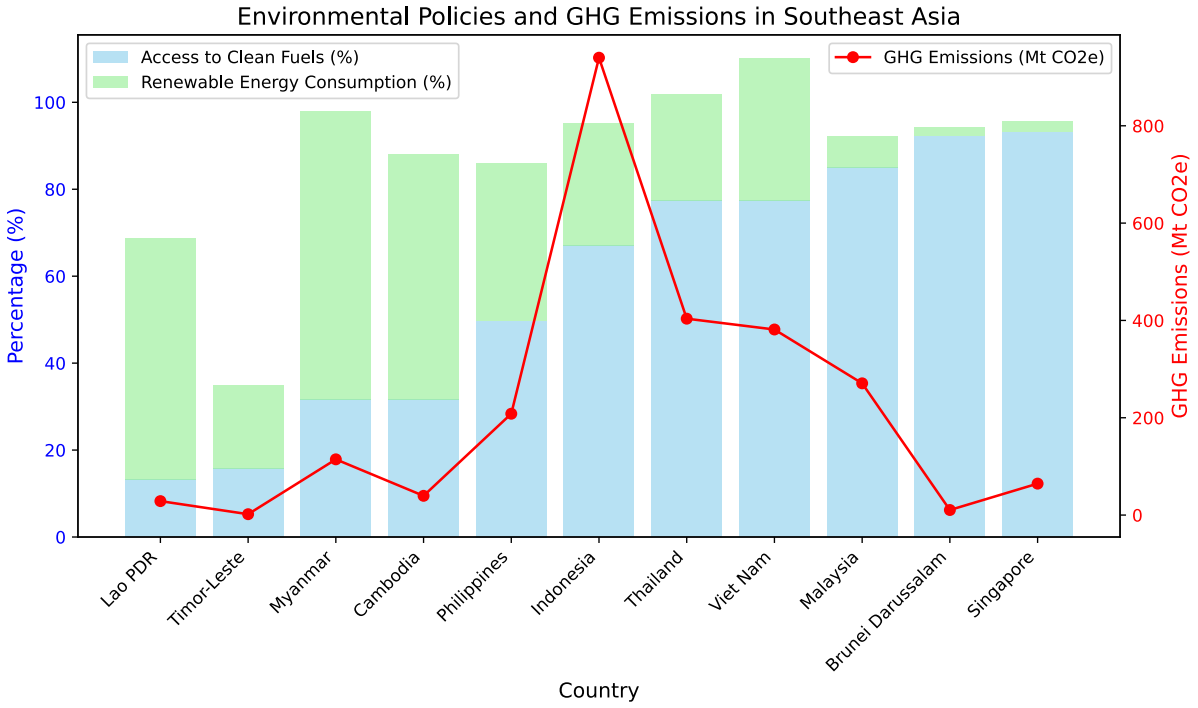

Figure 9: Environmental Policies and GHG Emissions in Southeast Asia

This twin plot compares environmental policies (Access to Clean Fuels, Renewable Energy Consumption) and GHG Emissions in Southeast Asia (2021-2023).

Malaysia, Brunei Darussalam, and Singapore have high access to clean fuels (nearly 100%), while the Philippines, Cambodia, and Lao PDR lead in renewable energy consumption (50-70%).

Indonesia has the highest emissions (800 Mt CO2e), whereas Brunei has the lowest.

Conclusion:

As we can see, 'Access to Clean Fuels and Technologies for Cooking' and 'Renewable Energy Consumption' do not directly determine 'greenhouse gas emissions'. For instance, countries like Singapore and Brunei, which have the highest access to clean fuels and technology, exhibit low greenhouse gas emissions. In contrast, Indonesia, despite using a significant amount of clean fuels and technology, still has emissions exceeding 800 Mt CO2e. This suggests that greenhouse gas emissions are influenced by multiple factors beyond just clean energy access and using renewable resources.

2.5.5 Visualization objective 5:

Question: Based on greenhouse gas emissions in Southeast Asia from the previous section, how are these countries impacted by emission-related damage?

Benefits:

- Helps economists assess the financial burden of climate change on different economies.
- Encourages businesses and policymakers to invest in green technologies to minimize long-term costs.

Columns used in data:

- Adjusted savings: carbon dioxide damage (
- Year

- Country Name

Visualizations used: World heatmap and bar chart
Reasons:

- For world heatmap: : World heatmap helps to highlight the average amount of carbon dioxide damage all over the world during the surveyed period in Southeast Asia.
- For bar chart: Bar chart is effective to compare average carbon dioxide damage between regions.

Analyzing and Visualizing steps:
We have been analyzed the amount of greenhouse gas emission in the previous section. Now in this section, let's see how Southeast Asian countries incur the cost due to emission-related damage.

First, let's have a quick look on carbon dioxide damage in this area over years. We will use interactive world map to plot the data.

Figure 10: CO2 Damage (% of GNI) in Southeast Asia Over the Years

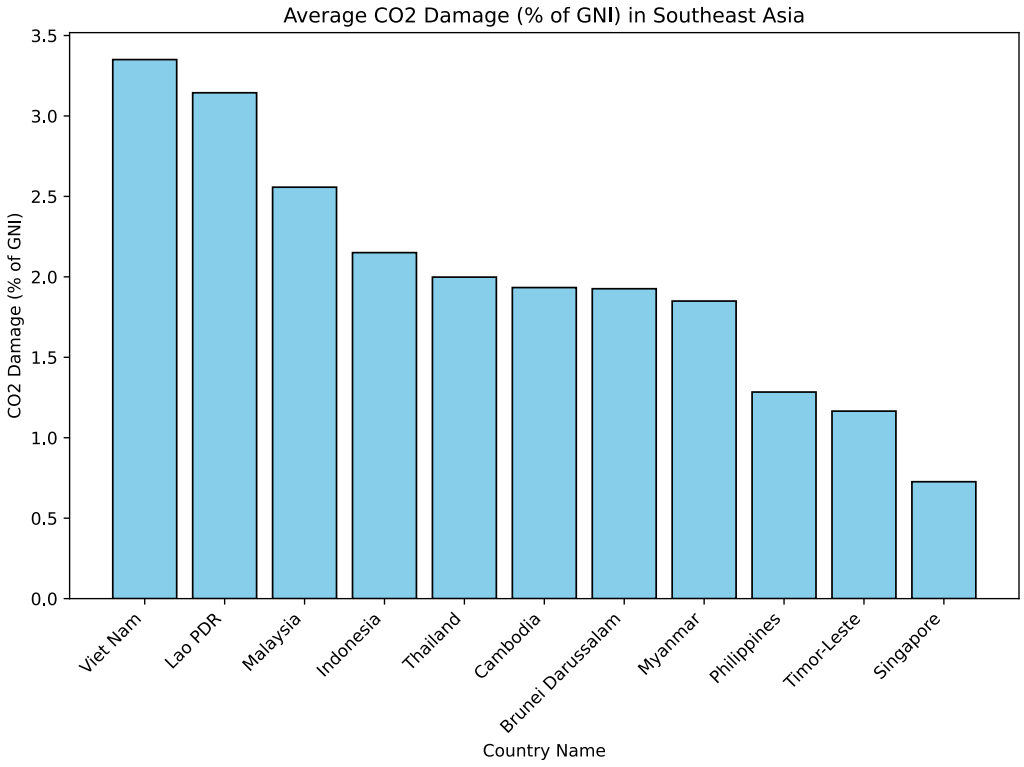

Figure 11: Average CO2 Damage (% of GNI) in Southeast Asia

Conclusion:
Countries with high GHG emissions tend to suffer greater CO damage. Indonesia, with the highest emissions (around 900 Mt COe), experiences CO damage of 2.5% GNI, while Vietnam (400 Mt COe) faces the highest damage at 3.5% GNI due to its lower economic output.

In contrast, Singapore, with the lowest emissions (50 Mt COe), has CO damage of only 0.5% GNI, thanks to effective environmental management. Lao PDR (100 Mt COe) exhibits high CO damage

(3.2% GNI) despite relatively lower emissions, highlighting the disproportionate economic impact on smaller economies. CO damage in Southeast Asia reflects the interplay between GHG emissions and economic structure, with highly industrialized nations like Indonesia and Vietnam bearing greater losses.

2.5.6 Visualization objective 6:

Question: What is the status of water supply among income groups (Freshwater used on daily basis, level of water stress)?
Benefits:

- Helps governments and water management authorities develop better water distribution and conservation plans.
- Supports humanitarian efforts to ensure access to clean water in water-stressed regions.
- Analyze the amount of freshwater is used on daily basis.

Columns used in data:

- Level of water stress: freshwater withdrawal as a proportion of available freshwater resources
- People using at least basic drinking water services, rural (% of rural population)
- People using at least basic drinking water services, urban (% of urban population)
- Water productivity, total (constant 2015 US$ GDP per cubic meter of total freshwater withdrawal)
- Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal)
- Annual freshwater withdrawals, industry (% of total freshwater withdrawal)
- Annual freshwater withdrawals, domestic (% of total freshwater withdrawal)

Visualizations used: Radar chart and pie charts
Reasons:

- For radar chart: : Radar chart is the best option for displaying multiple characteristics and compare features among groups.
- For pie charts: It is easy to visualize the differences in the proportions of freshwater withdrawals among income groups.

Analyzing and Visualizing steps:
First, let's analyze the overall water status:

How do different income groups differ in terms of water stress levels, the proportion of water withdrawals for domestic use, access to safe water in rural and urban areas, and water productivity?

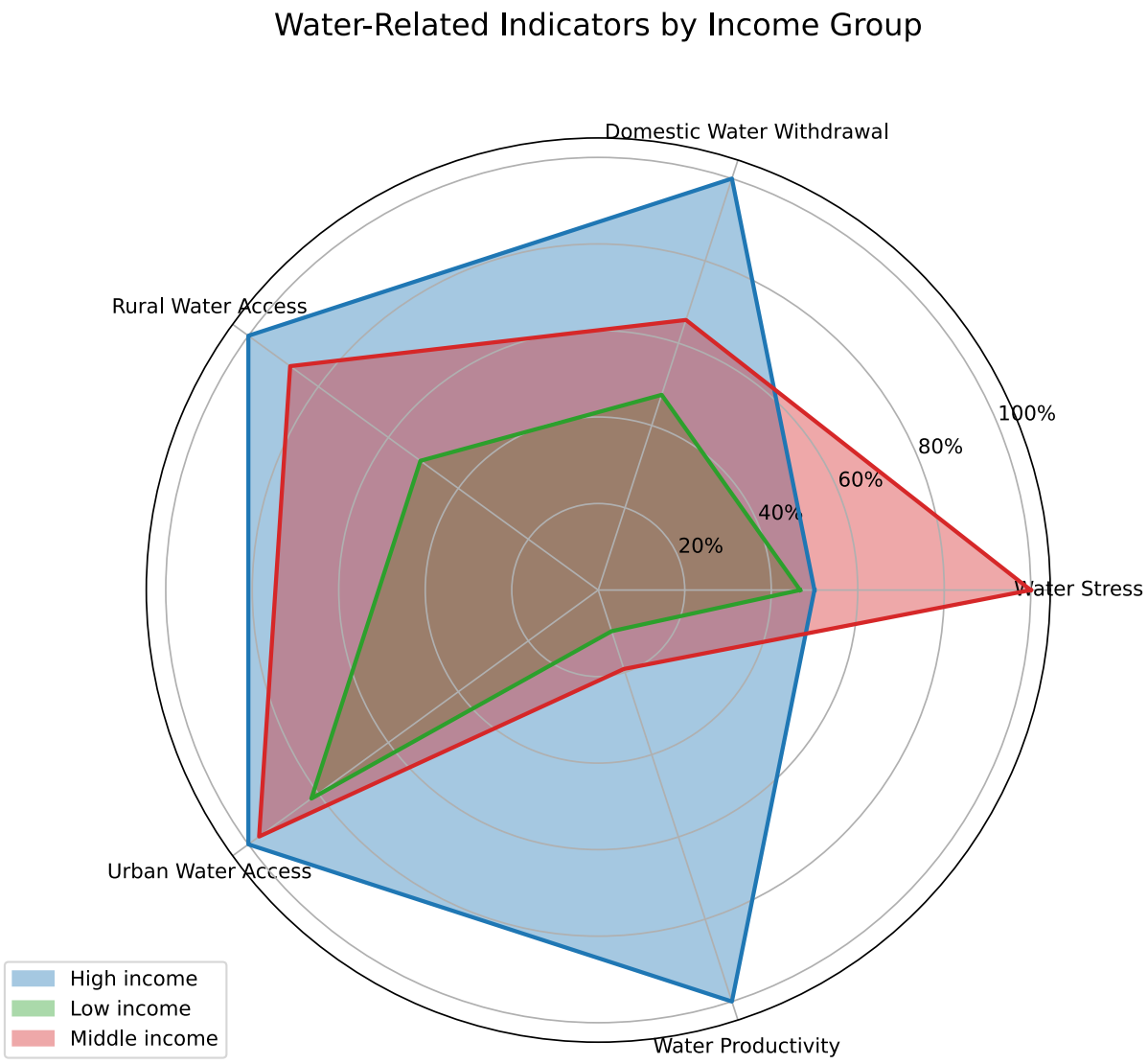

Figure 12: Water-Related Indicators by Income Group

The Radar Chart clearly illustrates the stark differences in water resource management among three income groups.
High-income countries excel with the highest Water Productivity (100%) and nearly perfect access to clean water, thanks to advanced technology and well-developed infrastructure.

In contrast, middle-income countries face the highest Water Stress (100%) but have low Water Productivity (20%), highlighting the need for more efficient water use.

Meanwhile, low-income countries lag behind in all indicators, especially Water Productivity (10%) and access to clean water (only 50% in rural areas), primarily due to limited resources and inadequate infrastructure.

Next, let's see **What was the allocation rate of freshwater for agriculture, domestic use, and industry for each income group in the most recent year?**

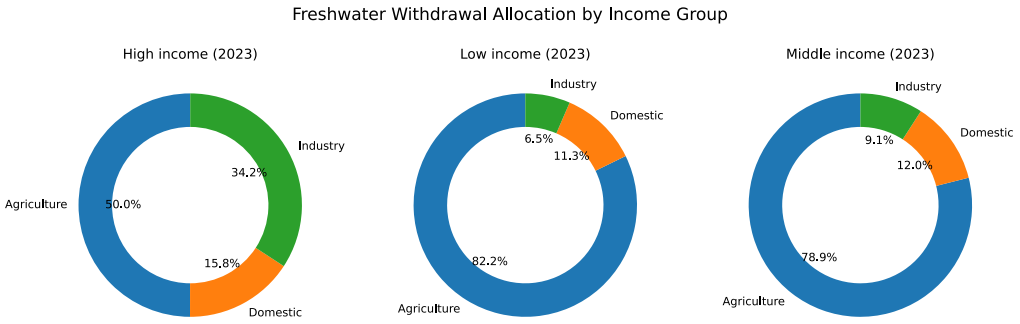

Figure 13: Freshwater Withdrawal Allocation by Income Group

The following donut charts infer that agriculture is the sector with the highest freshwater withdraw among 3 groups of income.

Specifically, *low income* group allocate the highest amount of freshwater consumption for Agriculture since the low-income countries tend to depend on agriculture. The ratio of freshwater for domestic use higher than that of industry.

Middle-income group takes after the trend of low income, but with the smaller proportion of freshwater allocation and higher rate in industry.

High-income group are likely to invest more freshwater for industry than domestic use than the others.

Conclusion:
The water situation highlights clear inequality: *High-income* countries manage water efficiently with a balanced allocation, *middle-income* countries face significant resource pressure, while *low-income* countries struggle with access to clean water and efficient water use, requiring strong investment in infrastructure and technology.

References

[1] https://www.geeksforgeeks.org/handling-missing-data-with-iterativeimputer-in-scikit-learn/, Accessed Date: March 21st

[2] https://www.geeksforgeeks.org/random-forest-regression-in-python/ , Accessed Date: March 21st

[3] https://www.naturalearthdata.com/downloads/110m-cultural-vectors/ , Accessed Date: March 19th