# BC2406 Analytics I: Visual & Predictive Techniques

**Group Report**

**SEMINAR 02, TEAM 3**


1. **Au Yew Rong Roydon (U2021424J)**
2. **Benjamin Chung Zhi Yong (U2021629B)**
3. **Lau Shing Hung (U2021155C)**
4. **Yeo Haw Lin (U2020768H)**


**SUBMITTED ON ________________**

# Effects of Environmentally Friendly Policies on Maintaining a Healthy GDP Growth Rate

Machine Learning: Proof of Concept

# Table of Contents

# 1: Executive Summary

**Studies on Environmental Factors:**
This report studies environmental factors, primarily renewable energy, and their relationship with two variables – GDP per capita, and GDP per capita growth, over a period of 28 years (1991 to 2018). Due to the presence of many strong outliers for GDP per capita growth, the relationship was studied by converting it into a categorical variable (Acceptable growth).

**Linear Regression and Continuous CART Model (GDP per capita):**
The study was conducted using Linear regression and continuous CART models. We tested different models and explored how the results varied across different datasets (Splitting by economy type). Upon many comparisons, we have found that continuous CART was the better model of the two for all datasets (Lower RMSE). Using CART, the results are shown below:
Renewable Energy Consumption and Population Growth are found to be the two most significant variables for predicting GDP per capita for all datasets. Moreover, environmental variables generally have high importance with CO2 emissions being top 4. It was also observed that CO2 emissions had a negative weight for developed countries
Upon further studying the variable importance of Renewable electricity output and Renewable energy consumption over three time periods (1991-1999, 2000-2008, 2009-2018) we observed:
**Overall dataset**: Renewable electricity output's importance increased over the years.
**Developed Countries dataset**: Renewable energy consumption's importance increased
**Developing Countries dataset**: Both variables' importance remained constant.
Hence, Renewable energy consumption is a strong predictor for the GDP per capita of a country, and its importance in the developed countries model will likely increase over time.

**Logistic Regression, Categorical CART Model (Acceptable growth):**
The study was conducted using Logistic regression and categorical CART models. Just like above, we compared the models and found that CART has a better prediction accuracy for all datasets except for developing countries. Using CART and logistic models, the results are:
**CART:**
In general, environmental variables are strong predictors of Acceptable growth for all datasets with CO2 emissions being a relatively strong predictor.
**Overall dataset**: Population total and nitrous oxide emissions are the two most significant variables. Moreover, both renewable energy variables are good predictors.
**Developed Countries dataset:** Population total and CO2 emissions are the two most significant variables.
Just like above, we studied the trend of the two renewable variables over three time periods:
**Overall dataset**: Renewable electricity output's variable importance increases over time.
**Developed Countries dataset**: Renewable energy consumption's importance increased over time.
**Logistic Regression (Developing countries dataset):**
CO2 emissions and greenhouse gases are good predictors of Acceptable growth. Other environmental variables are also close to statistical significance. Moreover, Renewable energy consumption's variable importance and strength as a predictor increases over time.
Hence, from the two models, we can conclude that both renewable energy variables are strong predictors for Acceptable growth for the overall dataset model. Moreover, renewable energy consumption's strength as a predictor increases over time.

# 2: Introduction

## 2.1: Why environmental factors?

### 2.1.1: Rising importance of climate change:

We believe that climate change is a significant issue in the 21st century that leaders need to pay close attention to[1]. Many countries are growing at the expense of the environment which results in many extreme weather conditions as well as alarming issues. These issues not only affect the country itself, but also on a regional or even global scale, such as an increase in global temperatures and rising sea levels.

To preserve our future, countries should aim for green and sustainable growth as it affects not just GDP growth in the economic sense, but also has positive effects on the overall political and social well-being of a country.

### 2.1.2: Lack of environment considerations/components in the EIU report:

The main EIU report focuses on economic growth indicators based on the implicit assumption that the global environment and networks for business will remain status quo for the foreseeable future. However, we believe that climate change and global warming are likely to have an impact on these figures (e.g., GDP). In the long-term, focusing on sustainable growth creates a conducive environment for business and economic development, which aligns with the goals of the EIU.

## 2.2: Justification

In 2020, the World Economic Forum (WEF) published its 15th Global Risks Report[2]. It was the first time that environmental factors were brought up as the highest risk to the world's economic well-being. Some of these factors include the failure of climate change mitigation and the loss of biodiversity. Hence, we have decided to study the relationship between environmental factors and economic indices, namely GDP per capita, and its growth.

## 2.3: Literature Review & Hypothesis Development

We have found many studies detailing the ways in which economic factors affect the environment. For example, there has been the study of the Environmental Kuznets Curve (EKC), which is a hypothetical relationship between per capita income and certain environmental indicators[3,4,5,6] Due to the lack of studies on the reverse relationship, our team has decided to explore whether environmental factors affect GDP per capita and its growth.

# 3: Data and Approach

## 3.1: Intended Approach

We intend to use machine learning models (linear regression, logistic regression, and CART) to explore the relationship between environmental factors, primarily renewable energy and a country's GDP per capita and its growth. (Refer to **Appendix 3.1** for flowchart)

There is a limited extent as to how accurate statistical models can predict economic indices accurately due to the large number of variables and data[7]. Moreover, they are mostly used to study univariate relationships. By using our machine learning models, it could aid EIU in studying multivariate relationships between economic indices and environmental variables. Moreover, it would increase the accuracy of prediction for GDP per capita and its growth since more variables and data can be used.

Our prediction models would also help countries better visualise the impacts of their actions and make it easier for them to track how renewable energy could potentially improve their GDP and growth rate. Through this report, government officials and key stakeholders will be able to identify the major areas in which they will need to improve on (e.g., prioritise increasing of renewable energy as a source of energy or decreasing of $CO_2$ emissions).

## 3.2: Data used

Data was sourced from the World Bank Data Bank.

### 3.2.1: Removal of Certain Countries and 1970-1990, 2019-2020 data:

Through our initial data observations, it was seen that certain countries had almost no data for many of our independent variables. A large majority of these countries are small island nations. Hence, due to the lack of data, we removed countries such as American Samoa, Aruba etc.

It was also observed that many nations, especially smaller nations, did not have any data for the periods above. Hence, it was decided to remove these data altogether as there would have been insufficient data during the years above to do any meaningful analysis.

### 3.2.2: Addition of Columns for further analysis:

We added two columns economy stage and regions to widen our analysis of how certain variables vary across regions and different stages of economies. Countries were split into 3 different sections, developed countries, developing countries and economies in transition. The economy stage column provides us with a fair comparison between countries who are in the same stage since some variables such as GDP per capita could vary between countries at different stages.

The split by regions would provide us with a better perspective of how different regions fare in terms of GDP per cap and renewable energy usage. It allows us to view from a bigger perspective as compared to individual countries to implement environmentally friendly measures on a larger scale.

### 3.2.3: Data used in the report:

We have decided to use **GDP per Capita** and **GDP Growth per Capita** instead of **GDP** and **GDP Growth** of the entire country as **per capita** provides us with a fairer comparison between countries of different sizes. Moreover, the skewness of **GDP** and **GDP Growth** is also significantly higher than **per capita.** (**Table 1**)

| GDP (current US$) | GDP growth (annual %) | GDP per capita (current US$) | GDP per capita growth (annual %) |
|---|---|---|---|
| 9.83212 | 3.927242 | 3.495805 | 3.818661 |

***Table 1 (Section 3.2.3)***

We decided to remove CO2 emissions (metrics ton per capita) since it is very similar to CO2 emissions (kt) which is not scaled by per capita. Most of our other independent variables are also not scaled by per capita, hence we decided to keep CO2 emission (kt). (Refer to **Appendix 12** for more details on the original variables and **Appendix 3.2** for current variables.

# 4: Data Exploration

Before we delve into data cleaning, we conducted data exploration. Below are some noteworthy graphs and datasets that help provide better data visualisation. Exploratory analysis helps determine how we should approach data cleaning and helps identify important relationships.

## 4.1: Initial Exploratory Analysis

### 4.1.1: CO2 Emissions (kt) vs GDP per capita (current US$):

Many developing countries, especially those in Asia, have a high amount of CO2 emissions per capita, while having a lower GDP per capita. (**Appendix 4.1A**)

When split into different economy types, we observed: (**Appendix 4.1A**)
Developed Countries: The gradient of increase in GDP per capita decreases as CO2 emissions increases. This implies that the impact of CO2 emissions on the increase in GDP per capita decreases as CO2 emissions increases.

Developing Countries: An increase in CO2 emissions increases GDP per capita. Ignoring outliers, the gradient remains roughly the same.

Economies in Transition: Due to the small data size, a trend cannot really be seen since GDP per capita shows a drastic drop, followed by a sharp increase as CO2 emissions increases.

### 4.1.2: Renewable Energy Consumption vs GDP per capita:

Developing countries generally have a low GDP per capita, while having a spread of Renewable energy consumption from 0 to 100%. Developed economies tend to have a smaller spread of Renewable energy consumption. (**Appendix 4.1B**)

## 4.2: Distribution of Dependent Variables

### 4.2.1: Distribution of GDP per capita (current US$):

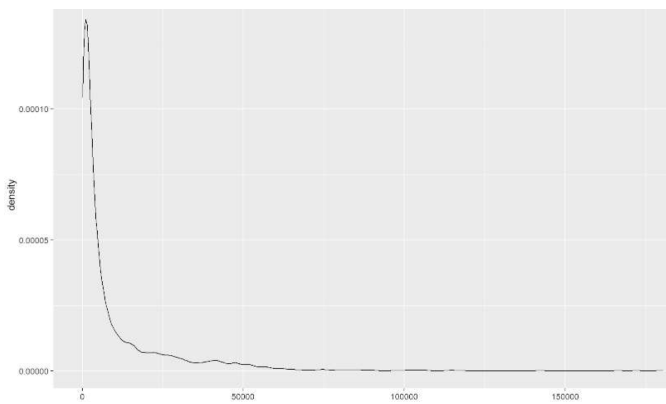


*Figure 1: Distribution Plot (Section 4.2.1)*

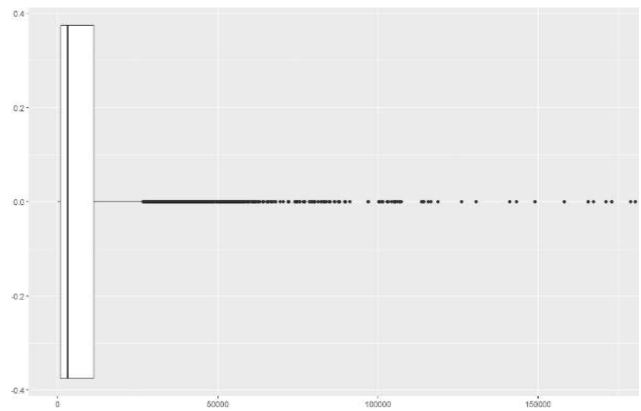


*Figure 2: Boxplot (Section 4.2.1)*

The density plot of the GDP per capita (current US$) shows that the data is extremely right skewed. Upon inspection, it is found that its mean value is 10,251.7, with median being 3,149.7. The minimum value is 22.8, with maximum value being 180,366.7. Hence, we have decided to log the values reduce the impact of the outliers.

### 4.2.2: Distribution of GDP per capita growth (annual %):

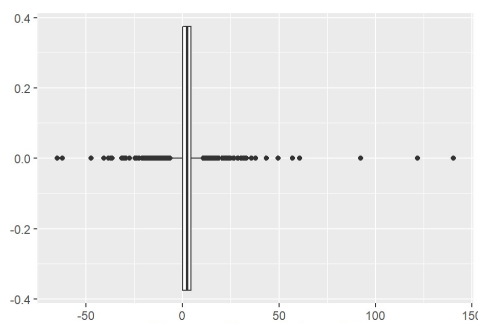



*Figure 1 (Section 4.2.2)*

The boxplot shows that GDP per capita growth is skewed just slightly to the right, but has a very wide range of values, from a minimum of -65% to a maximum of 140%.

## 4.3: Overall Summary of Data

We further analysed the summary statistic of variables namely:

| Variable | Minimum Value | First Quartile | Median | Mean | Third Quartile | Maximum Value |
|---|---|---|---|---|---|---|

Refer to Appendix **4.3** for full table

# 5: Data Cleaning

## 5.1: Independent Variable Data Cleaning

### 5.1.1: Removal of Columns

Based on summary statistics, we were able to identify independent variables that contains many missing values for a significant number of countries and proceeded to remove them.
These variables include: (1) Agricultural methane emissions (% of total), (2) Energy related methane emissions (% of total), (3) Energy use (kg of oil equivalent per capita) and (4) Fossil fuel energy consumption (% of total).

### 5.1.2: Imputation of NA Values

Next, we began the imputation of missing values for some independent variables. Some of these values do not exist due to reasons including, but not limited to, the Fall of the Soviet Union, and internal conflicts in certain countries.

Most data from 2016 to 2018 were missing from (1) Renewable electricity output and (2) Renewable energy consumption. Since the missing data was on the tail end of the years, the latest values (from 2015) were used to replace them, as an average reading will not be representative of the overall trend of data.

The imputation is done with respect to **each country** instead of the **entire column**. Hence when we are imputing with average, it would be the averages **with respect to each country** instead of averages of the **entire column**.

## 5.2: Dependent Variable Data Cleaning

### 5.2.1: Removal of Rows with NA values

We have decided to remove the rows which contain NA values in our dependent variables of primarily GDP per capita and GDP growth per capita. This is because imputation of values for dependent variables could introduce biases into our models. Moreover, most of the NA values are random from 1991 onwards.

### 5.2.2: Checking of inconsistencies

We made a check with regards to some obvious data inconsistencies such as negative GDP per capita which would not make sense. No inconsistencies were found as the values looked reasonable.

### 5.2.4: Analysing our outliers

With regards to GDP per capita and GDP Growth per capita, the graphs are very right skewed. However, these outliers are important to analysis, and we can't remove them. The reason for the high skewness is due to the wide range of countries with different GDP per capita and GDP Growth per capita. There are also many other external factors, for example conflict and natural disasters, which may hinder GDP Growth. Hence when running our models, we have logged the variables to reduce the impact of outliers on the model.

# 6: Correlation Analysis

## 6.1: Studying Correlations with Subsets of Data

### 6.1.1: Correlation between variables: (Full dataset)

When comparing between different variables, there are some notable findings from the correlation data. (For full table on pairwise correlations, refer to **Appendix 6.1A)**

In general, methane emissions are the only variable with a strong linear correlation to greenhouse gas emission. Moreover, greenhouse gases have a negative correlation with renewable energy consumption and renewable electricity output implying that an increase in renewable energy and electricity would decrease greenhouse gas emissions. (**Table 1**)

| Correlation between greenhouse gases against other variables | | | | | |
|---|---|---|---|---|---|
| **Variable** | **CO2 emission (kt)** | **Methane emissions** | **Nitrous oxide emissions** | **Renewable electricity output** | **Renewable energy consumption** |
| **Total greenhouse gas emissions** | 0.149 | 0.685 | 0.248 | -0.295 | -0.492 |

***Table 1 (Section 6.1.1)***

**GDP per capita** has a negative correlation with both renewable energy consumption and renewable electricity output. Hence, we have decided to conduct further analysis over time and economy to view possible changes in correlation. Greenhouse gas emissions have a stronger correlation to GDP per capita than CO2 emissions. (**Table 2)**

On the contrary, **GDP per capita growth** has a positive correlation with both renewable energy consumption and electricity output. Moreover, CO2 emissions has a positive correlation with GDP per capita growth while greenhouse gas emissions have negative correlation. (**Table 2**)

This implies that GDP per capita and GDP per capita growth are affected very differently by different variables.

| Correlation between GDP and Growth against other variables | | | | |
|---|---|---|---|---|
| **Variable** | **CO2 emissions** | **Total greenhouse gas emissions** | **Renewable electricity output** | **Renewable energy consumption** |
| **GDP per capita (current US$)** | 0.125 | 0.511 | -0.083 | -0.333 |
| **GDP per capita growth (annual %)** | 0.045 | -0.038 | 0.015 | 0.002 |

***Table 2 (Section 6.1.1)***

### 6.1.2: Correlation over time:

When comparing between 3 datasets from different time periods (**1991-1999**, **2000-2008**, **2009-2018)**, there are some notable findings from the correlation data.

Although renewable electricity output and renewable energy consumption were negatively correlated to GDP per capita in the previous analysis, they are becoming less negative overtime. This implies that they could eventually be positively correlated to GDP per capita. (**Table 1**)

Despite greenhouse gas emissions being highly correlated to GDP per capita, it is becoming less positively correlated overtime which shows a decrease in positive relationship.

| GDP per capita (current US$) against other variable (Over 3 periods) | | | |
|---|---|---|---|
| **GDP per capita (current US$) vs** | **1991-1999** | **2000-2008** | **2009-2018** |
| CO2 emissions (metric tons per capita) | 0.208 | 0.136 | 0.089 |
| Total greenhouse gas emissions (kt of CO2 equivalent) | 0.180 | 0.116 | 0.078 |
| Renewable electricity output (% of total electricity output) | -0.152 | -0.099 | -0.045 |
| Renewable energy consumption (% of total final energy) | -0.389 | -0.340 | -0.322 |

***Table 1 (Section 6.1.2)***

The same trend can be seen for renewable electricity output and GDP per capita growth as it becomes positively correlated overtime. Despite the increase in negative correlation for renewable energy consumption in 2000-2008, there is an overall positive correlation trend as it becomes positively correlated in 2009-2018. This implies that on a whole, the positive relationships of both renewable variables and both dependent variables grow stronger over time.

| GDP per capita growth (annual %) against other variable (Over 3 periods) | | | |
|---|---|---|---|
| **GDP per capita growth (Annual %) vs** | **1991-1999** | **2000-2008** | **2009-2018** |
| CO2 emissions (metric tons per capita) | 0.046 | 0.059 | 0.070 |
| Total greenhouse gas emissions (kt of CO2 equivalent) | 0.014 | 0.067 | 0.073 |
| Renewable electricity output | -0.035 | 0.015 | 0.093 |
| Renewable energy consumption | -0.007 | -0.092 | 0.103 |

***Table 2 (Section 6.1.2)***

## 6.1.3: Comparing Economy Types:

When comparing between the three different economy types - **Developing, Economies In Transition**, and **Developed**, there are some notable findings from the correlation data. As a country becomes increasingly developed, the positive linear relationship between its emissions and GDP per capita increases, but so does the relationship between its renewable energy indicators and GDP per capita. This implies that, the relationships of both renewable variables and both dependent variables become more positively correlated to GDP per capita as a country becomes more developed. (**Table 1**)

On the contrary, the positive linear relationship between the GDP per capita growth and greenhouse gas emissions decreases as a country becomes more developed. Moreover, the positive linear relationship between the GDP per capita growth and renewable energy consumption increases as a country becomes more developed. This implies that there could be a stronger positive relationship between renewable energy consumption and GDP per capita growth as a country becomes more developed. (**Table 2**)

| GDP per capita (current US$) against other variable (Over economy types) | | | |
|---|---|---|---|
| **GDP per capita (current US$) vs** | **Developing** | **In Transition** | **Developed** |
| CO2 emissions | 0.027 | 0.033 | 0.123 |
| Total greenhouse gas emissions | 0.018 | -0.007 | 0.119 |
| Renewable electricity output | -0.262 | -0.032 | 0.238 |
| Renewable energy consumption | -0.469 | -0.165 | 0.176 |

***Table 1 (Section 6.1.3)***

| GDP per capita growth (annual %) against other variable (Over economy types) | | | |
|---|---|---|---|
| **GDP per capita growth (current US$) vs** | **Developing** | **In Transition** | **Developed** |
| $CO_2$ emissions | 0.080 | 0.416 | -0.060 |
| Total greenhouse gas emissions | 0.082 | 0.041 | -0.061 |
| Renewable electricity output | 0.011 | 0.036 | 0.019 |
| Renewable energy consumption | -0.009 | 0.038 | 0.062 |

***Table 2 (Section 6.1.3)***

# 7: Common Preparation For Models

## 7.1: Preparing variables for models

We have decided to **drop** the **year column** and **country column** as they would not be used as predictors in our models. We further one hot encoded **categorical columns** of **economy type** and **region**. One hot encoding was used instead of label encoding as our categorical variables had no ranking. We have also logged GDP per Capita due to its skewness.

The next step was to perform **train-test split** for **linear regression** and **logistic regression model** with a ratio of **70:30** (**Train: Test**).

## 7.2: Final Variables to be used:

Independent Variables:

| | | | | | |
|---|---|---|---|---|---|
| Agricultural land (% of land area) | Agricultural land (sq km) | Agricultural methane emissions | $CO_2$ emissions (kt) | Methane emissions | Nitrous oxide emissions |
| Population growth (annual %) | Population, total | Renewable electricity output | Renewable energy consumption | Total greenhouse gas emissions | Hot encoded Economies And Regions |

Dependent Variables:

| | | | |
|---|---|---|---|
| GDP per capita (current US$) | GDP per capita growth (annual %) | Log GDP per capita | Healthy Growth |

## 7.3: Analysing using 3 datasets:

For all models, we will be analysing 3 datasets namely, the original dataset, developing countries dataset and developed countries dataset.

# 8: Linear Regression Models

## 8.1: Using Linear Regression to predict GDP per capita

We began by building our first Linear Regression model that predicts the value of GDP per Capita of countries. To ascertain the reason for high RMSE, we analysed the model diagnostic plots (**Appendix 8.1A: Before log**). Through the plots, we found out that the QQ plot was unfavourable since it showed that GDP per capita was right skewed which led to our decision of using the log GDP per capita instead.
We then proceeded to try out different methods to further improve our RMSE and Adjusted R-Squared to obtain the best model. The results are shown below. **(Table 1)** We decided to proceed with the Log Model **(Best Model)** as it has the lowest RMSE and highest adjusted r-squared. (**Appendix 8.1A: After log**)

By examining the p-values in the best model, we found that most of the emission variables were statistically significant in predicting **log GDP per capita**. Besides greenhouse gases, most emission-related variables have a negative weight which suggests that a **decrease in** them **increases GDP per capita** (**such as CO2 emissions**). (**Appendix 8.1B**).
Both renewable electricity output and renewable energy consumption are statistically significant. The weight of renewable electricity is positive; hence we can conclude that when **renewable electricity output increases, GDP per capita increases**. (**Table 1**)
On the other hand, the weight of renewable energy consumption was negative which suggests the opposite. As a result, we decided to conduct further analysis over economy types and different time periods to look for possible differences in significance and weights.

| **Linear Regression Model (Predicting GDP per Capita)** | | | | | |
|---|---|---|---|---|---|
| | **Trainset RMSE** | **Testset RMSE** | **Trainset Adjusted R-Squared** | **Renewable Electricity Output (p-value)** | **Renewable Energy Consumption (p-value)** |
| GDP Model | 11,280 | 14,564 | 0.479 | - | - |
| Log GDP Model **(Best Model)** | 0.921 | 0.938 | 0.658 | $8.89*10^{-14}$ (+ve) | $2*10^{-16}$ (-ve) |
| Removed Independent Variables with p-value > 0.05 (**Based on best model**) | 0.923 | 0.939 | 0.657 | - | - |
| Removed Independent Variables with VIF (>10) (**Based on best model**) | 1.127 | 1.150 | 0.489 | - | - |

***Table 1 (Section 8.1)***

Due to the popular adoption of renewable energy occurring later during our data timeframe, the impact of renewable energy on the overall dataset was likely to be imbalanced. Hence, we began our analysis of p-value and weights over three split time periods.
Since CART has a higher accuracy, we did not further evaluate the results, but they can be found in (**Appendix 8.1C**). The same analysis over time periods were also done in CART.

To conclude, both renewable electricity output and renewable energy consumption are significant predictors of GDP per capita for all countries. However, due to the negative weight for renewable energy consumption, more analysis must be done over the years to check for possible trends of it becoming less negative. Most emission variables are also statistically significant and a decrease in them helps increase GDP per capita (e.g., CO2 emissions)

## 8.2: Splitting by Developed vs Developing

For further analysis, we decided to split the original dataset into 2 sections for Developing and Developed Countries. This would help us examine if the significance of variables differs for different types of economy.
We first began by analysing the data over 28 years with the split of Developed vs Developing Countries. Countries that have economies in transition are placed together with developing countries due to the small data size. We repeated the same steps above to obtain the best model for developed and developing countries dataset (**Appendix 8.2A, 8.2B**).

Using our best model, we analysed the p-value and weights for developed and developing models. (**Appendix 8.2C**) For both datasets, CO2 emissions is a strong predictor and has a negative weight. With regards to developed countries, not many environmental variables were statistically significant. (**Appendix 8.2C**) However, renewable energy consumption is very close to statistical significance. Since both the weights of renewable variables are also positive, an increase in them would **increase GDP per capita.** (**Table 1**)
With regards to developing countries, many environmental variables are statistically significant. (**Appendix 8.2C**) Moreover, both renewable energy variables are statistically significant. (**Appendix 8.2C**) Since the weight of renewable electricity output is positive, an **increase** in the output **would increase GDP per capita**. (**Table 1**)

| **Linear Regression Model (Developed vs Developing)** | | | | | |
|---|---|---|---|---|---|
| | **Trainset RMSE** | **Testset RMSE** | **Trainset Adjusted R-Squared** | **Renewable Electricity Output (p-value)** | **Renewable Energy Consumption (p-value)** |
| **Developed** | 0.722 | 0.630 | 0.569 | 0.8693 (+ve) | 0.0976 (+ve) |
| **Developing** | 0.866 | 0.864 | 0.582 | $8.31*10^{-13}$ (+ve) | $2*10^{-16}$ (-ve) |

***Table 1 (Section 8.2)***

To further view if there could be a possible increase in variable importance of renewable energy consumption such that it would be statistically significant in the future, we proceeded to analyse how the variable importance changes over 3 periods of time (1990-1999, 2000-2008, 2009-2018). Since CART is a better model **(lower RMSE**), we did not further evaluate the results but included them in (**Appendix 8.2D**).

To conclude, an increase in renewable electricity output and renewable energy consumption helps increase GDP per capita for developed countries. Renewable energy consumption is a fairly strong predictor of GDP per capita with a p-value that is very close to 0.05. With regards to developing countries, renewable electricity output is a strong predictor for GDP per capita (p-value < 0.05) and an increase helps increase GDP per capita.

## 8.3: Using Linear Regression to predict GDP Growth per Capita

We built another Linear Regression model that predicts the value of GDP Growth per Capita of countries. However, the model had extremely high RMSE and extremely low Adjusted R-Squared. We proceeded to attempt to improve the model by translating the GDP Growth per Capita followed by logging it. Below are the results: (**Table 1**)

| **Linear Regression Model (GDP Growth Per Capita)** | | | |
|---|---|---|---|
| | **Trainset RMSE** | **Testset RMSE** | **Trainset Adjusted R-Squared** |
| **Original Model (Before logging)** | 5.780 | 6.368 | 0.019 |
| **Translate and log Model (Translate and log GDP per capita)** | 0.048 | 0.049 | 0.008 |

***Table 1 (Section 8.3)***

The RMSE improved tremendously, however Adjusted R-Squared decreased. We further analysed the model diagnostic plots (**Appendix 8.3**) and found out that there were many outliers in the dataset. The residual vs leverage graph had many influential outliers even after the translation and log of the dependent variable which would have minimised their effect. Since these outliers are important and shouldn't be removed, we have decided to use logistic regression to predict if a country has acceptable growth using the data from GDP Growth Per Capita in place of the linear regression model.

# 9: Continuous CART

## 9.1: Using Continuous CART to predict GDP per capita

Since linear regression is fitting a line to the data points while CART has binary splits, the weights of the variables in linear regression tell us an overall picture of how the independent variable affects the dependent variable. On the other hand, CART is useful in showing how independent variables affects dependent variables at the respective splits. Hence, the weights of the variables in linear regression are used to tell us about the overall direction between the dependent and independent variables. Just as before, we begin by finding the best model for CART. Since CART can handle missing values with surrogates, we decided to run CART using both the original data and cleaned data to observe possible differences. We then chose the best possible model with the lowest RMSE and proceeded to remove variables that are less significant (variables with significance of less than 1% significance). However, it provided a higher RMSE compared to the previous two models. (**Table 1**) (Full table in **Appendix 9.1A**) Using our best model, we observed that the two most significant variables were Renewable energy consumption (at 26%), and Population growth (at 10%) when analysed across all countries (Full table in **Appendix 9.1B**). This implies that Renewable energy consumption is a strong predictor for GDP per capita. Renewable electricity output was the sixth highest in terms of significance which implies that it is still relatively important. Most environmental variables were also strong predictors. (**Appendix 9.1B**)

A decrease in renewable energy consumption (**Appendix 9.1C**) leads to an increase in GDP per capita, so we decided to split the dataset into developed and developing economies to check if the type of economy would affect this trend.

| | Original Data | Cleaned Data | After Removal of Variables |
|---|---|---|---|
| **RMSE** | 0.8521 | 0.8171 | 0.8807 |

***Table 1 (Section 9.1)***

## 9.2: Splitting by Developed vs Developing

As above, we began by finding the best model for the two different economy datasets. The results show that the Cleaned Data model is better for both datasets. (**Table 1**) For Developed economies, the top 2 factors are (1) Population growth and (2) Renewable energy consumption. On the contrary, the top 2 factors for Developing economies are (1) Renewable energy consumption and (2) Population growth. This implies that population growth affects the GDP of developed countries more than developing countries. The GDP per capita of developing countries, on the other hand, might be more affected by the amount of Renewable energy they consume. This could be due to many new countries formed being labelled as developing countries whereas the source of growth for developed countries have been largely due to fossil fuels etc for long periods of time.
Despite the negative weight of Renewable energy consumption with GDP per capita, CART shows that for Developed economies, an increase in Renewable energy consumption leads to an increase GDP per capita. (**Appendix 9.2A**)

However, for developing economies, a decrease in Renewable energy consumption generally leads to an increase in GDP per capita, while an increase in $CO_2$ emissions leads to an increase in GDP per capita. (**Appendix 9.2B**) This could be due to the capabilities of a country.

| | **Original Data** | **Cleaned Data** |
|---|---|---|
| **Developed RMSE** | 0.5519 | 0.5236 |
| **Developing RMSE** | 0.8933 | 0.7242 |

***Table 1 (Section 9.2)***

## 9.3: Time Series Trends (For both renewable energy variables)

By comparing Linear Regression and CART, it is found that CART is a better model for predicting GDP per capita. (Compared using respective best models) (**Table 1)**

| **Best Linear Trainset RMSE** | **Best Linear Testset RMSE** | **Best CART RMSE** |
|---|---|---|
| 0.9210 | 0.9380 | 0.8171 |

***Table 1 (Section 9.3)***

Using CART, we proceeded to split the datasets into 3 time periods to identify possible trends. Primarily, we focused on how the variable importance of Renewable energy consumption and Renewable electricity output variables changes over time. (**Table 1**)

We observed that the variable importance of renewable electricity output increases over the years for the overall dataset. For developed countries, the Renewable energy consumption importance increases over the years. (Overall increasing trend since the year of 2009-2018 is the highest). Lastly, the importance of both variables remains constant for developing countries. (**Table 2**)

To conclude renewable energy consumption is an important predictor for GDP per capita across all countries. When broken down into economy types, renewable energy consumption is a stronger predictor for GDP per capita for developing countries than developed countries. However, the renewable energy consumption seems to have an increasing variable importance over the years for developed countries. Moreover, the variable importance of renewable electricity output also had an overall increase in the all-countries dataset which could mean that it could be a stronger predictor in the future.

| **Variable Importance** | **All years** | **1991-1999** | **2000-2008** | **2009-2018** |
|---|---|---|---|---|
| Overall Data - Renewable energy consumption | 26% | 29% | 28% | 27% |
| Overall Data - Renewable electricity output | 6% | 8% | 7% | 9% |
| Developed - Renewable energy consumption | 10% | 0% | 0% | 8% |
| Developed - Renewable electricity output | 7% | 4% | 0% | 0% |
| Developing - Renewable energy consumption | 27% | 29% | 26% | 29% |
| Developing - Renewable electricity output | 9% | 10% | 9% | 10% |

***Table 2 (Section 9.3)***

# 10: Logistic Regression Models

## 10.1: Variables for Logistic Regression

Instead of using linear regression to predict GDP per capita growth, we have decided to change it into a categorical variable. We do so by adding an extra variable (Acceptable growth). To generate the new variable, we would need to decide the percentage of growth that would imply healthy growth. However, the percentage depends on the country's economy type - developing or developed. A developing country should have a GDP per capita growth rate of at least 7% to be considered having healthy growth[8]. However, as 7% growth is difficult for a country to achieve, we will set a target of 4.7% for developing countries to hit[9]. A developed country should have a GDP per capita growth rate of between 2% to 3% to have a healthy growth[10]. With regards to economies in transition, we consider them to fall under the developing country category, hence the percentage used would be the same percentage as developing countries.

## 10.2: Using Logistic Regression to predict Acceptable Growth

To arrive at the best prediction model for Acceptable Growth, we decided to use a sampling of the original data to ensure a balanced number of Yes and No responses for Acceptable Growth. However, we proceeded to use the model with no rebalancing as the accuracy was severely affected by it and the prediction of "Yes" and "No" had equal importance. There was no removal of insignificant variables as it resulted in a drop in accuracy. (**Table 1 and 2**)

| **Logistic Regression Model (At threshold of 0.5)** | **Trainset Accuracy** | **TPR** | **TNR** | **FPR** | **FNR** |
|---|---|---|---|---|---|
| Original Model | **0.7360** | **0.285** | **0.921** | **0.079** | **0.715** |
| Re-balancing of number of Nos and Yes to have equal proportion in **Acceptable Growth** | 0.4610 | 0.285 | 0.921 | 0.079 | 0.715 |
| Removing of **Insignificant Variables** | 0.7260 | 0.125 | 0.970 | 0.030 | 0.875 |

***Table 1 (Section 10.2)***

| **Logistic Regression Model (At threshold of 0.5)** | **Testset Accuracy** | **TPR** | **TNR** | **FPR** | **FNR** |
|---|---|---|---|---|---|
| Original Model (**Best Model**) | **0.732** | **0.270** | **0.921** | **0.079** | **0.730** |
| Re-balancing of number of Nos and Yes to have equal proportion in **Acceptable Growth** | 0.6070 | 0.278 | 0.927 | 0.073 | 0.722 |
| Removing of **Insignificant Variables** for **scaled model** | 0.7230 | 0.122 | 0.968 | 0.032 | 0.878 |

***Table 2 (Section 10.2)***

Despite the low TPR, since there was no difference in value of predicting "Yes" or "No" we were evaluating the models based on overall accuracy instead. Using our best model, we further tried out different threshold values to obtain the highest possible accuracy. Upon doing so, we found out that 0.5 is still the best threshold for predictions. (**Appendix 10.2A**).

With our best model, we looked at the odds ratio and p-value of the independent variables. (**Appendix 10.2B**). Most environmental variables are statistically significant and strong predictors for the model. (**Appendix 10.2B**) The most statistically significant variable would be population growth followed by agricultural land. However, renewable energy consumption and **renewable electricity output** are not far behind. Renewable energy consumption is also statistically significant while renewable electricity output is very close. (**Table 3**)

| **Overall Dataset (No splits)** | | | | |
|---|---|---|---|---|
| **Logistic Regression Model (At threshold of 0.5)** | **Variable Importance of Renewable Electricity Output (p-value, odds ratio)** | **Variable Importance of Renewable Energy Consumption (p-value, odds ratio)** | **Trainset Accuracy** | **Testset Accuracy** |
| **Overall (All years)** | p-value: 0.0649 | p-value: 0.0410 | 0.7360 | 0.7320 |

***Table 3 (Section 10.2)***

Due to the p-value of renewable electricity output being really close to 0.05 (statistical significance), we decided to analyse how its significance changes over time. Due to CART model having a higher accuracy, we did not evaluate the changes over time in this section, but the results can be found in (**Appendix 10.2C**).

To conclude, renewable energy consumption is a good predictor for acceptable growth when analysed across all countries. Renewable electricity output is very close to being statistically significant however as of now no trend can be predicted.

## 10.3: Splitting by Developed vs Developing

In the same vein as linear regression, we tested our prediction model further on the sub-tables created, categorizing Developed and Developing variables with the three different time periods. Just like above, we began by finding the best models for each respective dataset (**Appendix 10.3A and 10.3B**). In the process of doing so, we found out that the best model for the developed countries dataset would be the model where all insignificant variables were removed and the only two significant variables were **population growth and population total.** With regards to the developing countries dataset, the best model was the model without removing insignificant variables. To compare the renewable variables, the weaker model was chosen for developed countries dataset.

For both developed and developing countries, renewable electricity output and renewable energy consumption was found to be statistically insignificant for the logistic regression model. For developed countries, there were no environmental variables that were statistically significant. Whereas for developing countries, some environmental variables such as CO2 emissions and total greenhouse gas emissions are statistically significant. The other environmental variables are very close to statistical significance. This implies that environmental variables are stronger predictors for developing countries than developed countries. (**Appendix 10.3C**) This includes renewable variables which has a higher statistical significance for developing countries than developed countries. (**Table 1**) Despite the high p-value, just as before, we decided to further analyse if the significance changes over time. Since CART had a higher accuracy for **the developed countries dataset**, we did not evaluate the changes over time in this section, but the results can be found in (**Appendix 10.3D**).

With regards to developing countries, the statistical significance of renewable energy consumption seems to have an increasing trend over the years with 2009-2018 being the best where it attained statistical significance (p-value < 0.05). (**Table** 2) This implies that despite the renewable energy consumption being statistically insignificant using the entire dataset, the significance would increase overtime. The insignificance could also be due to the importance of renewable energy being recognised fairly late in the years. Unfortunately, no trend could be seen for renewable electricity output and its statistical significance may not improve.

| Developed vs Developing Countries (Data over 28 years) | | | | |
|---|---|---|---|---|
| **Logistic Regression Model (At threshold of 0.5)** | **Renewable Electricity Output (p-value)** | **Renewable Energy Consumption (p-value)** | **Trainset Accuracy** | **Testset Accuracy** |
| **Developed** | p-value: 0.809 | p-value: 0.799 | 0.6340 | 0.6140 |
| **Developing** | p-value: 0.6235 | p-value: 0.1479 | 0.7780 | 0.7850 |

***Table 1 (Section 10.3)***

| Developing Countries | | | | |
|---|---|---|---|---|
| **Logistic Regression Model (At threshold of 0.5)** | **Renewable Electricity Output (p-value)** | **Renewable Energy Consumption (p-value)** | **Trainset Accuracy** | **Testset Accuracy** |
| **1991 - 1999** | p-value: 0.7095 | p-value: 0.5672 | 0.7710 | 0.7680 |
| **2000 - 2008** | p-value: 0.01708 | p-value: 0.35004 | 0.7240 | 0.7020 |
| **2009 - 2018** | p-value: 0.5328 | p-value: $1.08 * 10^{-5}$ | 0.8320 | 0.8440 |

***Table 2 (Section 10.3)***

To conclude, the strength of renewable energy consumption as a predictor of acceptable growth increases over time for developing countries and in the most recent period of 2009-2018, it became a strong predictor of acceptable growth.

# 11: Categorical CART

## 11.1: Process for CART

### 11.1.1: Process for Continuous CART

We will use the 10-fold Cross Validation (CV), which does random train-test splits in order to find the best model. CART compared a continuous log GDP per capita variable against our independent variables.

1. Run CART with the original set of data, which includes the NA values. As CART provides surrogates for missing values, it will return a model, with a certain Mean Square Error (MSE).
2. Run CART with the cleaned set of data, which have no NA values. Take the MSE.
3. By looking at the importance of the independent variables from the above models, we will drop the insignificant variables, before running the cleaned set of data again.
4. Run the best of the above models for the Developed and Developing countries sets.

### 11.1.2: Process for Categorical CART

For CART, similar to logistic regression, we have decided to compare the independent variables with whether a country is experiencing a healthy growth rate. A developing country should have a GDP per capita growth rate of at least 4.7% to be considered having healthy growth. While a developed country should have a GDP per capita growth rate of between 2% to 3% to have a healthy growth. As a trivial number, only 35 countries have an Economy type of Economies in transition, we will streamline them and Developing economies into the same category, with economies in transition falling under the developing country category. Categorical CART compares a categorical "Acceptable Growth" variable against our independent variables. Below are a few steps we took:

1. We implemented a train-test split approach on our dataset at a ratio of (70: 30).
2. Run CART with the train set data to obtain a tree.
3. Used the tree model to run the test set data to obtain the predicted result of each row (whether it is acceptable or not acceptable growth)
4. Obtain the confusion matrix of the test set data, which will show the true negative and true positive rates, as well as the overall accuracy of the model.

## 11.2: CART for Categorical Dependent Variable

Unlike logistic regression, CART uses categorical splits and hence we can tell how the dependent variable is affected by the independent variable at certain splits.

### 11.2.1: Comparing Between Different Models

Just like above we tested out different datasets and models to obtain the best model. We found that the model works best with the cleaned dataset and has the highest accuracy after removal of insignificant variables **(Table 1).** Using the best model, we found out that Population, total (10%), and Nitrous oxide emissions (thousand metric tons of CO2 equivalent) (10%) are the 2 most significant variables **(Table 1).** Moreover, most of the environmental variables have high variable importance and are good predictors of Acceptable growth for the overall dataset. With regards to the 2 renewable energy variables, although they may not be in the top 10 variables, they were not far behind with a variable importance of **5%**. (**Appendix 11.2A**) This implies that overall, they are still important as predictors since the most significant variable only has an importance of **11%**. As it is unlikely for us to observe all the splits of the renewable energy variables and CO2 emission variable throughout the entire fully grown tree, we only observed the ones present in the optimal tree. We observed that when Renewable energy consumption is more or equal than 1, the split is directed to having an acceptable growth, suggesting that an increase in renewable energy consumption increases the likelihood of having an acceptable growth. (**Appendix 11.2B**)

| | **Overall Accuracy** | **Most Significant Variable (%)** | **Second Most Significant Variable (%)** |
|---|---|---|---|
| **Original Data** | 0.7517 | Agricultural methane emissions (thousand metric tons of CO2 equivalent) (10%) | Population, total (10%) |
| **Cleaned Data** | 0.7520 | Agricultural methane emissions (thousand metric tons of CO2 equivalent) (10%) | Population, total (10%) |

| | | | |
|---|---|---|---|
| **After Removal of Insignificant Variables (Using cleaned data)** | 0.7570 | Population, total (11%) | Nitrous oxide emissions (thousand metric tons of CO2 equivalent) (10%) |

***Table 1 (Section 11.2.1)***

## 11.2.2: Comparing Between Economy Types

Just as before we found the best model with respect to developing countries dataset and developed countries dataset. We found out that the best model with the highest overall accuracy for Developed economies is the original dataset. **(Table 1)** While for Developing economies, the dataset after removal of insignificant variables has the highest overall accuracy. **(Table 2)**

Using the respective best models, we were able to observe that most environmental variables had high variable importance and are good predictors of Acceptable growth for both developed and developing countries datasets. With regards to developed countries, the 2 most significant variables for developed countries dataset are (1) Population, total, (2) CO2 emissions. Unfortunately, the 2 renewable energy variables have low variable importance. As a result, we conducted further analysis by splitting the dataset into 3 different time periods as per before. (**Table 1**)

With regards to developing countries dataset, the 2 most significant variables are (1) Nitrous oxide emissions, (2) Agricultural land. We also observed that renewable energy consumption has a good variable importance of 8% which is not too far from the most significant variable. (**Table 2**)

| | **Overall Accuracy** | **Most Significant Variable (%)** | **Second Most Significant Variable (%)** |
|---|---|---|---|
| **Original Data-Developed** | 0.7079 | Population, total (16%) | CO2 emissions (kt) (15%) |
| **Cleaned Data-Developed** | 0.6934 | Population, total (14%) | CO2 emissions (kt) (11%) |
| **After Removal of Insignificant Variables (Based on original data)** | 0.6934 | Population, total (17%) | CO2 emissions (kt) (13%) |

***Table 1 (Section 11.2.2)***

| | **Overall Accuracy** | **Most Significant Variable (%)** | **Second Most Significant Variable (%)** |
|---|---|---|---|
| **Original Data-Developing** | 0.7774 | Nitrous oxide emissions (thousand metric tons of CO2 equivalent) (13%) | Population, total (13%) |
| **Cleaned Data-Developing** | 0.7774 | CO2 emissions (kt) (14%) | Total greenhouse gas emissions (kt of CO2 equivalent) (13%) |
| **After Removal of Insignificant Variables (Based on clean data)** | 0.7785 | Nitrous oxide emissions (thousand metric tons of CO2 equivalent) (15%) | Agricultural land (% of land area) (13%) |

***Table 2 (Section 11.2.2)***

### 11.2.3: Time Series Trends (For both renewable energy variables)

We began by running our best model on the overall dataset that was split into 3 years. Upon doing so, we observed that the variable importance of renewable energy consumption remained roughly constant while the variable importance of renewable energy output increased tremendously reaching 8% in 2009-2018. This implies that the importance of renewable electricity output as a predictor for Acceptable growth could likely increase over time. (**Table 1**)

| Variable Importance | Over all years | 1991-1999 | 2000-2008 | 2009-2018 |
|---|---|---|---|---|
| **Overall Data - Renewable energy consumption** | 67.29 (5%) | 6.43 (2%) | 11.58 (4%) | 0.87 (<1%) |
| **Overall Data - Renewable electricity output** | 68.40 (5%) | 2.79 (<1%) | 7.94 (3%) | 12.37 (8%) |

***Table 1 (Section 11.2.3)***

Just as above, we split our developed and developing countries dataset over 3 time periods. With regards to developed countries, we observed that the variable importance of renewable energy consumption increases over time. Despite the low variable importance of renewable energy consumption, the increasing trend shows that its strength as a predictor could increase over time. (**Table 2**)

With regards to developing economies, logistic regression has a higher accuracy and hence we did not evaluate the results in this section. However, they are included in (**Appendix 11.2C**) Results are drawn from logistic regression instead.

| Variable Importance | Over all years | 1991-1999 | 2000-2008 | 2009-2018 |
|---|---|---|---|---|
| **Developed - Renewable energy consumption** | 6.31 (2%) | 1.54 (<1%) | 2.25 (<1%) | 6.42 (4%) |
| **Developed - Renewable electricity output** | 5.16 (2%) | 3.34 (<1%) | 1.37 (<1%) | 0 (0%) |

***Table 2 (Section 11.2.3)***

# 12: Evaluation of Models

## 12.1: Continuous CART vs Linear Regression

The continuous CART model has a lower RMSE compared to the linear regression model for all 3 datasets. This could be due to CART using a 10-fold cross validation, which splits the data into 10 sets. It trains the model using 9 of the sets and tests it on the remaining one. This is done 10 times to find a summary of the evaluation scores. Moreover, it could also be due to the data being non-linear. Hence, it could be difficult for linear regression to fit a line well onto the dataset. (**Table 1**)

Evaluation of Continuous CART vs Linear Regression for overall dataset:

| Best Models | Linear Regression Train Set RMSE | Linear Regression Test Set RMSE | CART RMSE |
|---|---|---|---|
| **Overall Data** | 0.826 | 0.840 | 0.817 |
| **Developing** | 0.794 | 0.799 | 0.724 |
| **Developed** | 0.659 | 0.592 | 0.524 |

***Table 1 (Section 12.1)***

## 12.2: Categorical CART vs Logistic Regression

Categorical CART has a higher overall accuracy compared to Logistic Regression for both the overall dataset and Developed economies dataset. **(Table 1 and 2)** However, for the dataset of Developing economies, Logistic regression model has a higher overall accuracy. **(Table 3)** Therefore, we plan to draw conclusions on Renewable electricity output and Renewable energy consumption for the overall dataset and Developed economy dataset from categorical CART, while findings for Developing economies dataset will be drawn from the Logistic regression model as it has higher overall accuracy.

Evaluation of Categorical CART vs Logistic Regression for overall dataset:

| Overall Data | Overall Accuracy |
|---|---|
| **Best Model for Categorical Cart** | 0.7570 |
| **Best Model for Logistic Regression** | 0.7320 |

***Table 1 (Section 12.2)***

Evaluation of Categorical CART vs Logistic Regression for Developed Economy dataset:

| Economy Type | Overall Accuracy |
|---|---|
| **Categorical CART for Developed Economies** | 0.7079 |
| **Logistic Regression for Developed Economies** | 0.6140 |

***Table 2 (Section 12.3)***

Evaluation of Categorical CART vs Logistic Regression for Developing Economy dataset:

| Economy Type | Overall Accuracy |
|---|---|
| **Categorical CART for Developing Economies** | 0.7785 |
| **Logistic Regression for Developing Economies** | 0.7850 |

***Table 3 (Section 12.3)***

# 13: Possible Improvements and Pilot Studies

## 13.1: Data Sources

The data used in this report was obtained from World Bank, which has many missing data. One possible improvement is to source for better data and combine them together with the data from World Bank. With more data, the accuracy of the models would improve especially in the training process.

## 13.2: Outlier Analysis

During data exploration, we found many significant outliers in some of our variables, especially the dependent variables. For example, for GDP per capita growth, Libya was a source of two outliers, with its growth rate at -62% in 2011 and +122% in 2012. This was due to the First Libyan Civil War, which hindered its economy. These outliers skew the data, which could result in a worst accuracy for models. However, they are still important and should not be removed. One improvement we could make would be to split the data by identifying the outliers. Another model would then be trained using the outliers. The two different models would then have better accuracy and serve different purposes.

## 13.3: Different Regions and per capita variables

Different regions might have different trends and different relationships between economic indices and environmental variables. Hence it might be useful for EIU to conduct pilot studies for different region types using similar machine learning techniques.

Since our current analysis has a mix between per capita variables (e.g., GDP per capita) and variables of absolute values (all our independent variables), the accuracies of the models and trends observed could change by changing all variables to the same scale of per capita.

# 14: Components for EIU

From our models, we observed that environmental variables are strong predictors for both GDP per capita and GDP per capita growth (Acceptable growth). We propose to add two components into the report.

Based on our best models, it can be observed that **CO2 emissions** have been consistently strong as a predictor across most models (**Appendix 9.1B, 10.3C). Renewable energy consumption** has also been a strong predictor and its strength generally grows over time for most models. Hence, we believe that it would be useful to add these two components to the EIU report.

# 15: Conclusion

With the increasing severity of climate change on the economy and the world, everyone has a part to play to help reduce global warming and environmental degradation. We believe that the addition of these two components would greatly benefit the environment as countries would be more aware of their actions since these components would impact their GDP and economic growth. With that being said, a country must be developed to a certain extent before renewable energy could become its main source of energy[11].

# References


1: *Climate change widespread, rapid, and intensifying*. IPCC. (n.d.). Retrieved November 6, 2021. From https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/.

2: *The Global Risks Report 2020*. World Economic Forum. (n.d.). Retrieved November 6, 2021. From http://reports.weforum.org/global-risks-report-2020/

3: Grossman, G., Krueger, A., 1991. *Environmental Impacts of a North American Free Trade Agreement*. NBER, Working Paper 3914.

4: Shafik, N., Bandyopadhyay, S., 1992. *Economic Growth and Environmental Quality: Time Series and Cross-country Evidence Policy Research Working Paper*. World Development Report, WPS 904.

5: Panayotou, T., 1993. *Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development*. Working Paper, Technology and Environment Programme. International Labour Office, Geneva.

6: Selden, T.M., Song, D., 1994. *Environmental quality and development: is there a Kuznets curve for air pollutions emissions? Journal of Environmental Economics and Management 27 (2)*, 147

7: Menon, S., Ranjan, R. (n.d.). *The evolution of forecasting techniques: Traditional versus machine learning methods*. Retrieved November 6, 2021. From https://www.genpact.com/insight/technical-paper/the-evolution-of-forecasting-techniques-traditional-versus-machine-learning-methods

8: Meurs, M., Seidelmann, L. & Koutsoumpa, M. *How healthy is a 'healthy economy'? Incompatibility between current pathways towards SDG3 and SDG8*. Global Health 15, 83 (2019). From https://doi.org/10.1186/s12992-019-0532-4

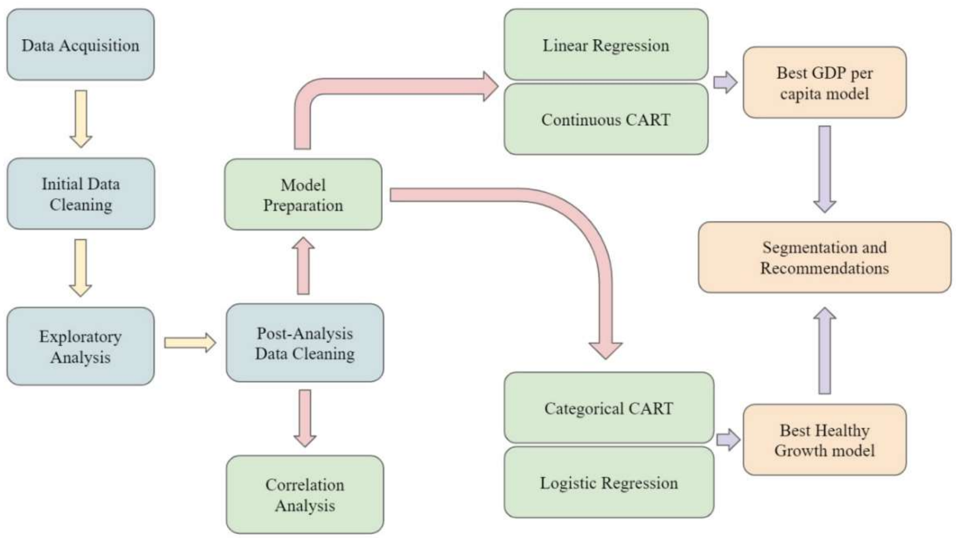
9: *UNCTAD Handbook of Statistics 2019 - Economic trends*. UNCTAD. (n.d.). Retrieved November 6, 2021. From https://unctad.org/system/files/official-document/tdstat44_FS07_en.pdf

10: O'Neill, A. (May 25, 2021). *Singapore: Growth rate of the real gross domestic product (GDP) from 2016 to 2026**. Retrieved November 6, 2021. From https://www.statista.com/statistics/378630/gross-domestic-product-gdp-growth-rate-in-singapore/

11: Silverstein, K. (March 16, 2021). *How Rich Countries Must Help Developing Economies Afford The Clean Energy Transition*. Retrieved November 6, 2021. From https://www.forbes.com/sites/kensilverstein/2021/03/16/can-emerging-countries-afford-to-make-the-clean-energy-transition/?sh=50b1f33d383a

# Appendix (Numbers below correspond to the section numbers in main report)

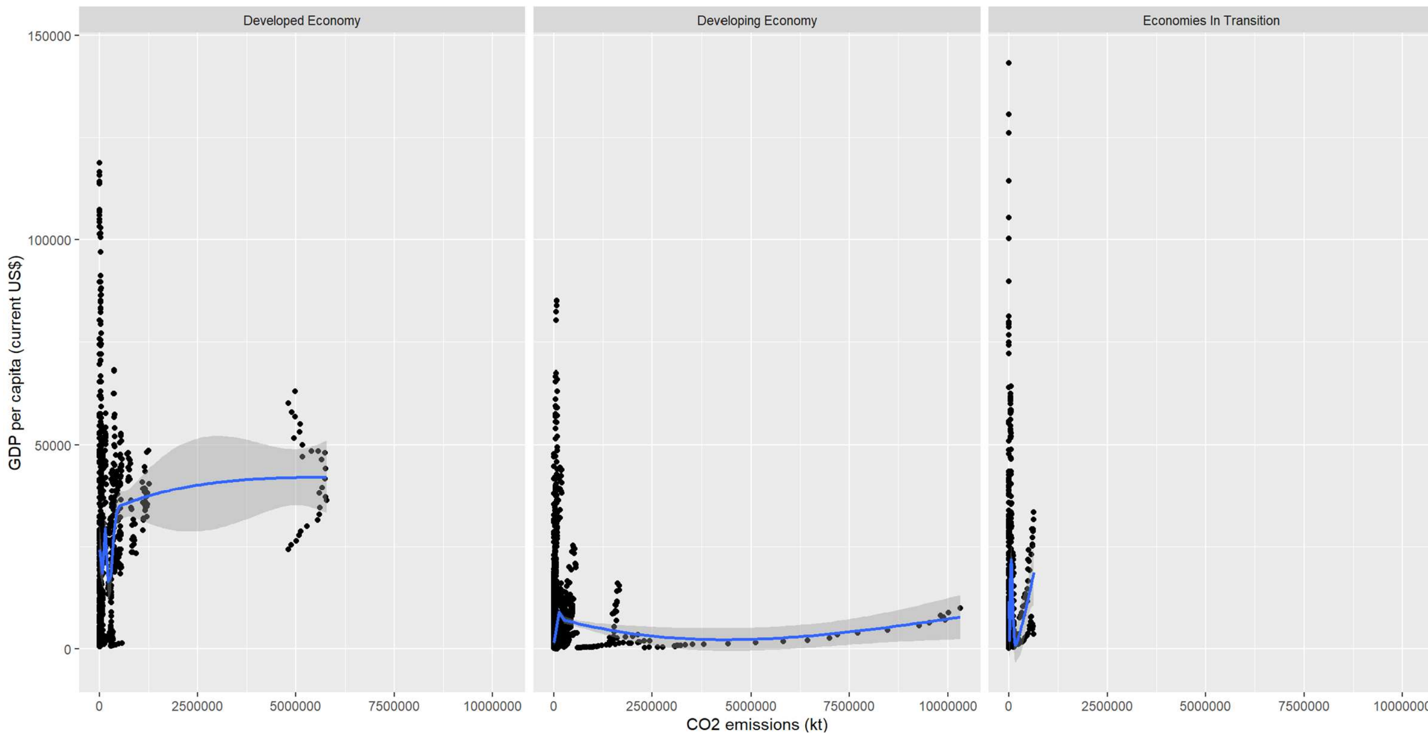
## 3.1: Flow Chart Diagram of Intended Approach

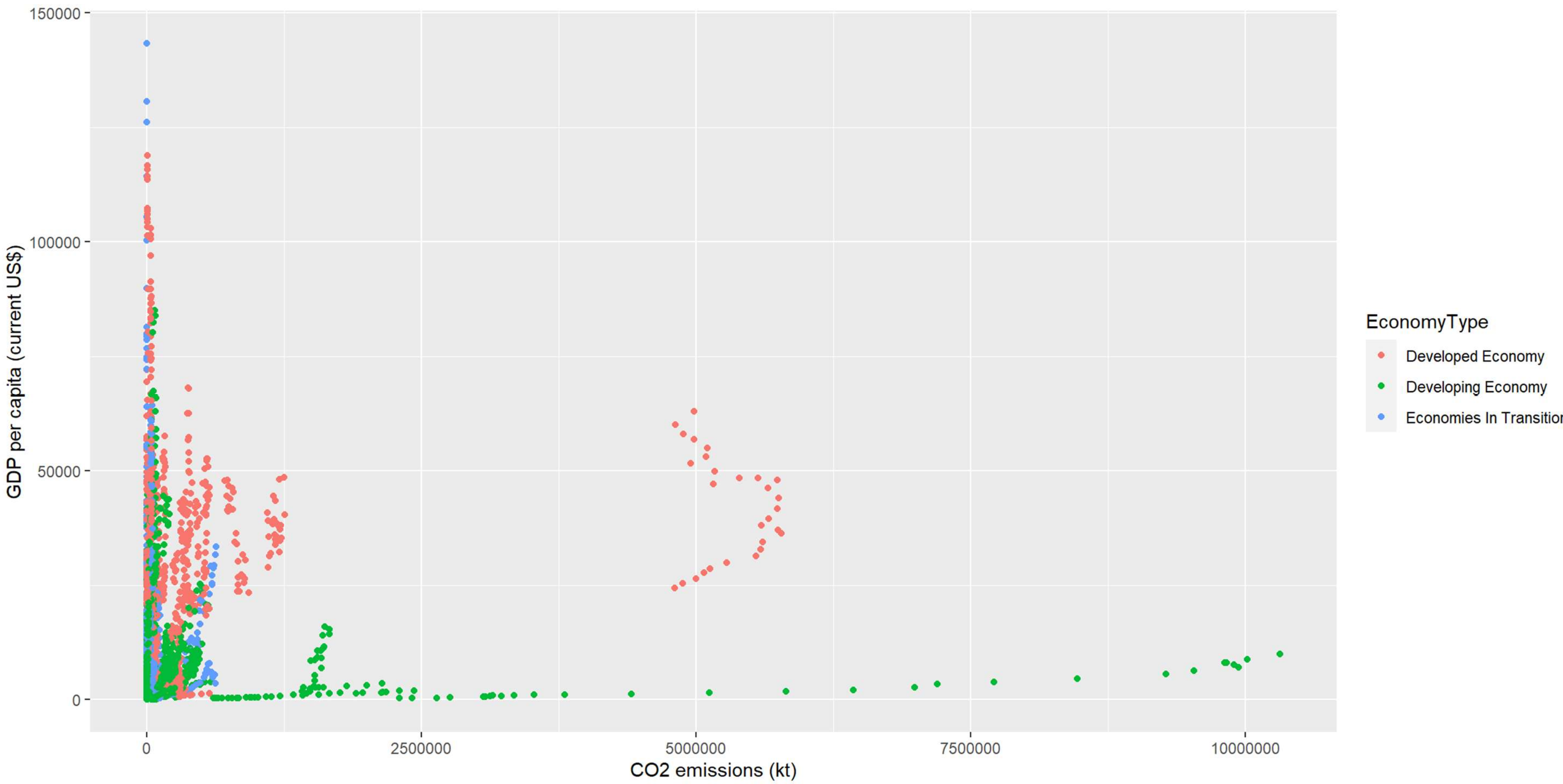


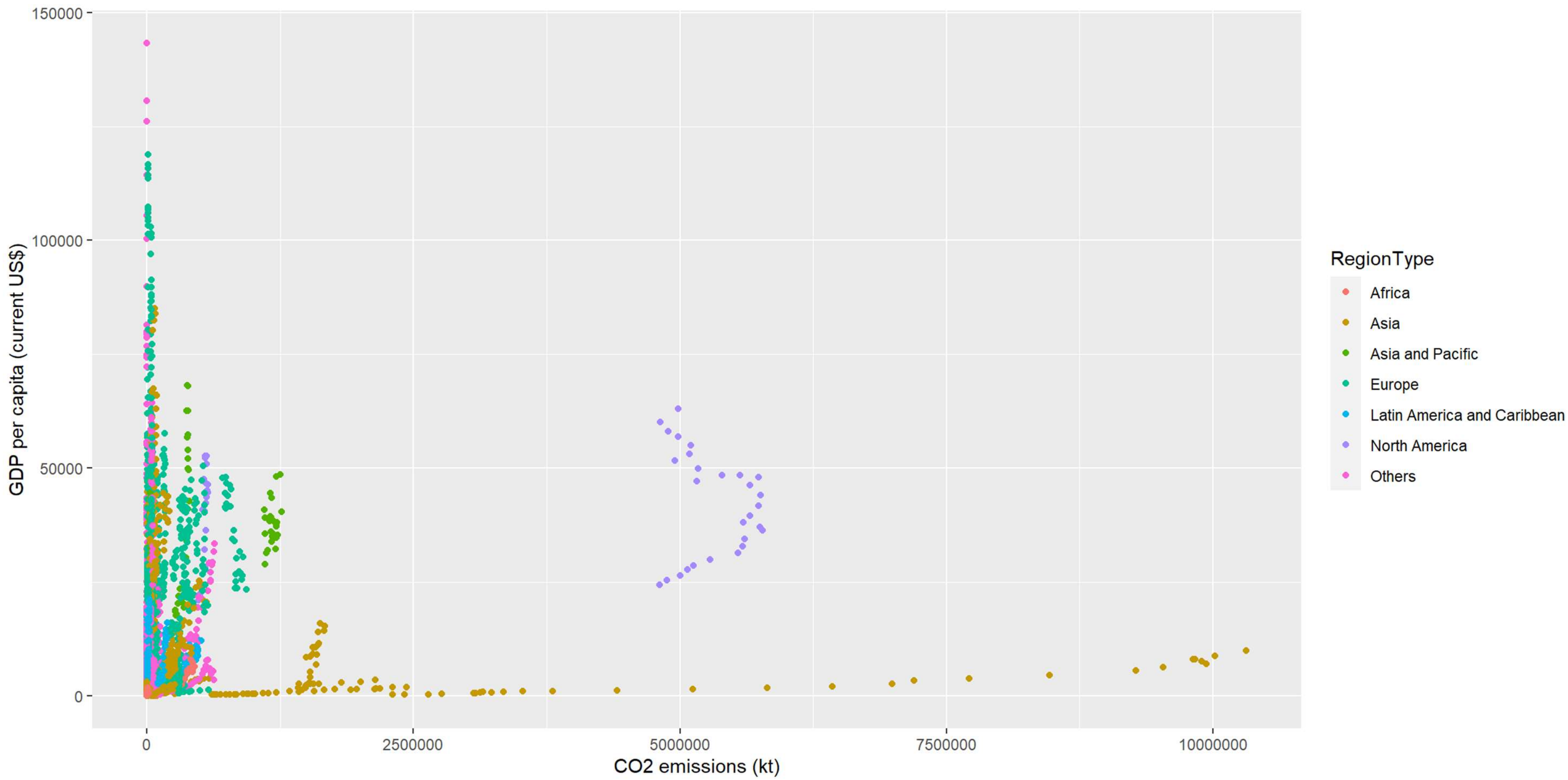
## 3.2: Variables before Data Exploration and Data Cleaning

| INDEPENDENT VARIABLES | INDEPENDENT VARIABLES | DEPENDENT VARIABLES |
|---|---|---|
| • Agricultural land (% of land area)<br>• Agricultural land (sq. km)<br>• Agricultural methane emissions (% of total)<br>• Agricultural methane emissions (thousand metric tons of $CO_2$ equivalent)<br>• $CO_2$ emissions (metric tons per capita)<br>• Energy related methane emissions (% of total)<br>• Energy use (kg of oil equivalent per capita)<br>• Fossil fuel energy consumption (% of total)<br>• Methane emissions (kt of $CO_2$ equivalent) | • Nitrous oxide emissions (thousand metric tons of $CO_2$ equivalent)<br>• Population growth (annual %)<br>• Population, total<br>• Renewable electricity output (% of total electricity output)<br>• Renewable energy consumption (% of total final energy consumption)<br>• Total greenhouse gas emissions (kt of $CO_2$ equivalent)<br>• EconomyType (from UN)<br>• RegionType (from UN) | • GDP per capita (current US$)<br>• GDP per capita growth (annual %) |

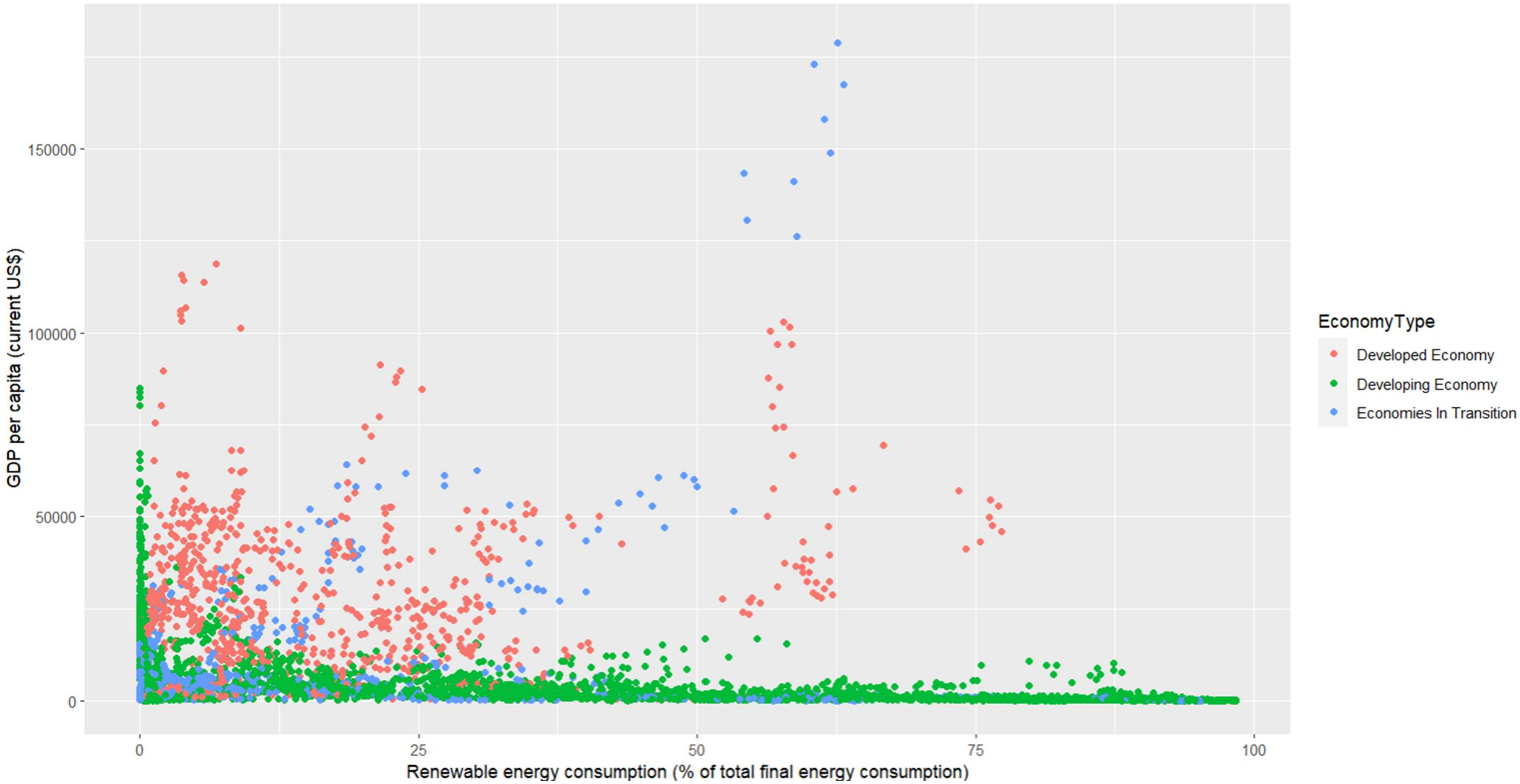
## 4.1A: CO2 Emissions (kt) vs GDP per capita (current US$)




150000
100000
50000
0
GDP per capita (current US$)
0
2500000
5000000
7500000
10000000
CO2 emissions (kt)
EconomyType
Developed Economy
Developing Economy
Economies In Transition


150000
100000
50000
0
GDP per capita (current US$)
0
2500000
5000000
7500000
10000000
CO2 emissions (kt)
RegionType
Africa
Asia
Asia and Pacific
Europe
Latin America and Caribbean
North America
Others

## 4.1B: Renewable Energy Consumption vs GDP per capita (current US$)



## 4.3: Summary Statistics of variables

| Variable | Minimum Value | First Quartile | Median | Mean | Third Quartile | Maximum Value | Number of NA |
|---|---|---|---|---|---|---|---|
| **Agricultural land (% of land area)** | 0.4487 | 21.1776 | 40.0238 | 39.1379 | 55.9789 | 85.4874 | 69 |
| **Agricultural land (sq. km)** | 7 | 7717 | 39490 | 252995 | 214865 | 5290386 | 69 |
| **Agricultural methane emissions (% of total)** | 0 | 23.38 | 50.00 | 46.95 | 70.02 | 100.00 | 1883 |
| **Agricultural methane emissions (thousand metric tons of CO2 equivalent)** | 0 | 557.5 | 3640.0 | 17328.1 | 13955.0 | 498490.0 | 0 |
| **CO2 emissions (kt)** | 0 | 1490 | 9305 | 141587 | 58872 | 10313460 | 0 |
| **Energy related methane emissions (% of total)** | 0 | 5.728 | 14.159 | 22.865 | 30.573 | 99.337 | 1883 |
| **Energy use (kg of oil equivalent per capita)** | 9.548 | 563.641 | 1318.765 | 2374.157 | 3144.547 | 22120.430 | 1878 |
| **Fossil fuel energy consumption (% of total)** | 0 | 43.69 | 74.74 | 65.12 | 89.81 | 100.00 | 1956 |
| **GDP per capita (current US$)** | 22.8 | 929.6 | 3149.7 | 10251.7 | 11243.2 | 180366.7 | 182 |
| **GDP per capita growth (annual %)** | -64.9924 | 0.0697 | 2.1953 | 2.0972 | 4.4102 | 140.3670 | 251 |
| **Methane emissions (kt of CO2 equivalent)** | 0 | 2230 | 8455 | 38752 | 27730 | 1242150 | 0 |
| **Nitrous oxide emissions (thousand metric tons of CO2 equivalent)** | 0 | 587.5 | 3460.0 | 13949.4 | 10400.0 | 546990.0 | 20 |
| **Population growth (annual %)** | -6.7661 | 0.5086 | 1.3859 | 1.4948 | 2.4370 | 17.5122 | 12 |
| **Population, total** | 9014 | 1935010 | 7278490 | 34258249 | 22515917 | 1392730000 | 10 |
| **Renewable electricity output (% of total electricity output)** | 0 | 0.6913 | 17.0453 | 32.0559 | 59.5894 | 100.0000 | 561 |
| **Renewable energy consumption (% of total final energy consumption)** | 0 | 5.398 | 24.184 | 33.769 | 59.295 | 98.343 | 617 |
| **Total greenhouse gas emissions (kt of CO2 equivalent)** | 10 | 7530 | 28460 | 197498 | 92168 | 12355240 | 0 |

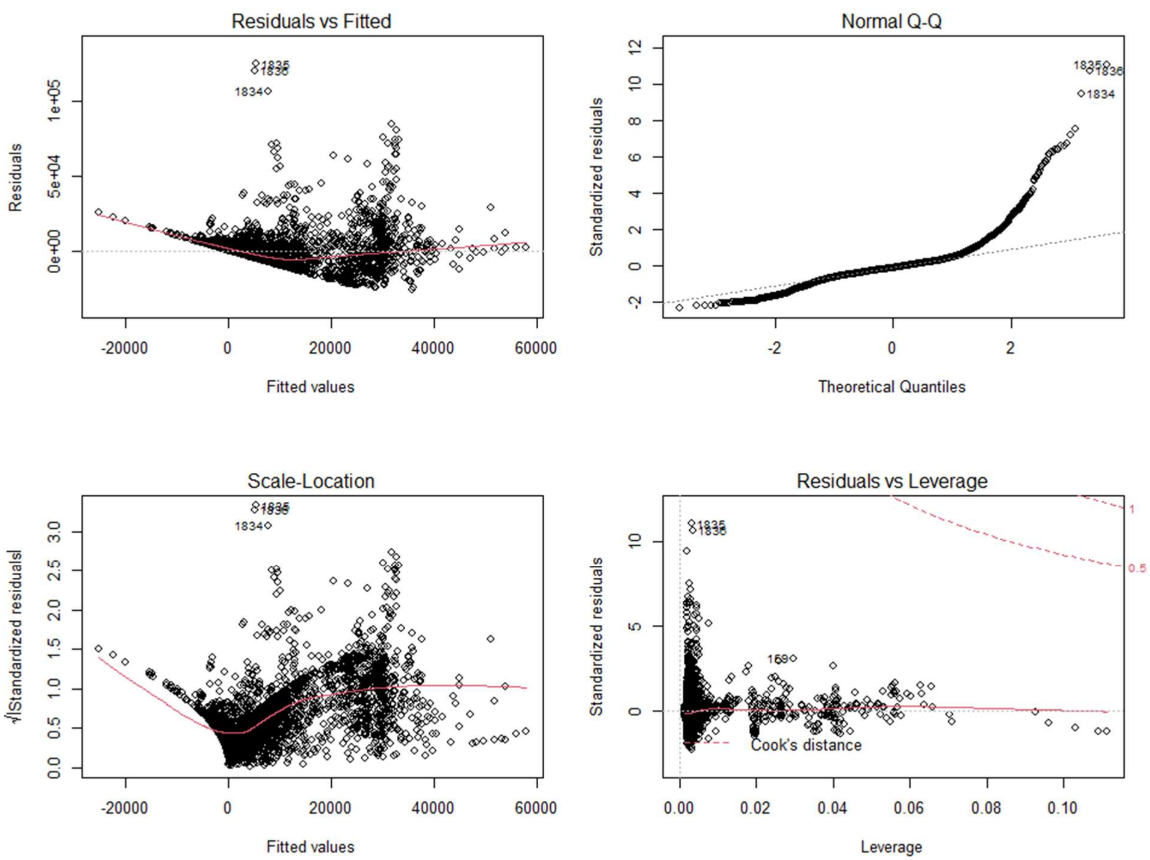
## 6.1A: Pairwise Correlation Table of Variables (Full dataset)

| | AGRICULTURAL LAND (% OF LAND AREA) | AGRICULTURAL LAND (SQ. KM) | AGRICULTURAL METHANE EMISSIONS | CO2 EMISSIONS (KT) | GDP PER CAPITA (CURRENT US$) | GDP PER CAPITA GROWTH (ANNUAL %) | METHANE EMISSIONS | NITROUS OXIDE EMISSIONS | POP GROWTH (ANNUAL %) | POP, TOTAL | RENEWABLE ELECTRICITY OUTPUT (% OF TOTAL OUTPUT) | RENEWABLE ENERGY CONSUMPTION (% OF TOTAL ENERGY) | TOTAL GREENHOUSE GAS EMISSIONS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AGRICULTURAL LAND (% OF LAND AREA)** | 1 | | | | | | | | | | | | |
| **AGRICULTURAL LAND (SQ. KM)** | 0.19 | 1 | | | | | | | | | | | |
| **AGRICULTURAL METHANE EMISSIONS (THOUSAND METRIC TONS CO2 EQUIV)** | 0.11 | 0.72 | 1 | | | | | | | | | | |
| **CO2 EMISSIONS (KT)** | 0.06 | 0.73 | 0.61 | 1 | | | | | | | | | |
| **GDP PER CAPITA (CURRENT US$)** | -0.14 | 0.04 | -0.01 | 0.12 | 1 | | | | | | | | |
| **GDP PER CAPITA GROWTH (ANNUAL %)** | 0 | 0.03 | 0.06 | 0.05 | -0.05 | 1 | | | | | | | |
| **METHANE EMISSIONS (KT OF CO2 EQUIV)** | 0.06 | 0.8 | 0.83 | 0.83 | 0.02 | 0.05 | 1 | | | | | | |
| **NITROUS OXIDE EMISSIONS (THOUSAND METRIC TONS CO2 EQUIV)** | 0.09 | 0.82 | 0.86 | 0.88 | 0.04 | 0.06 | 0.9 | 1 | | | | | |

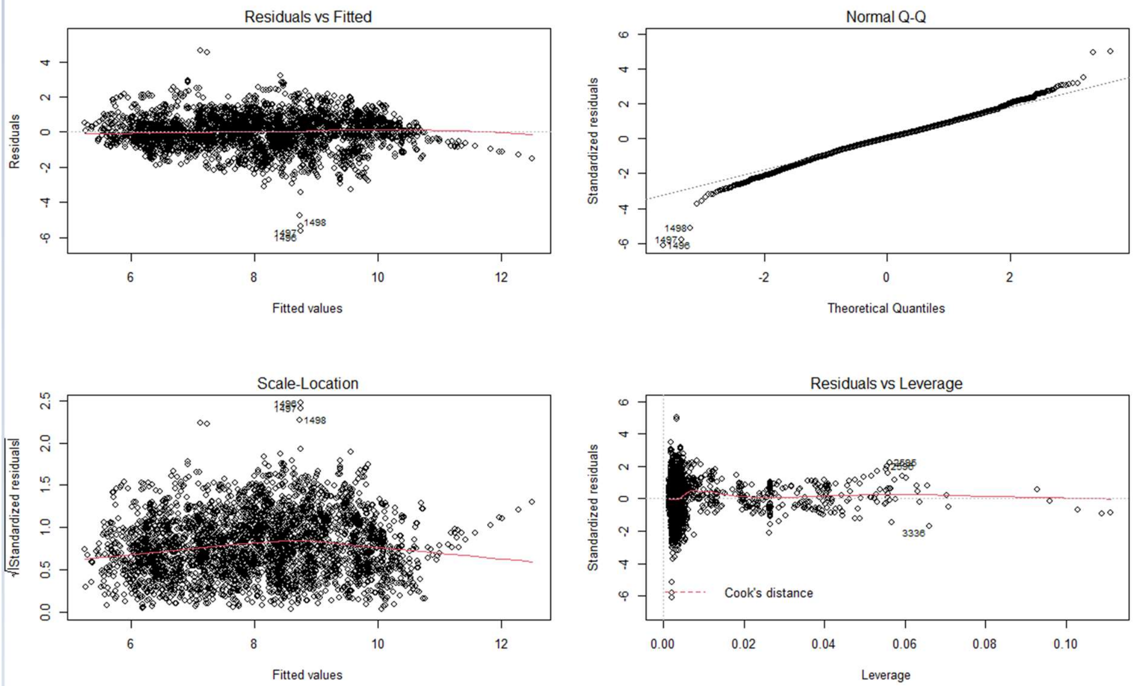
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **POPULATION GROWTH (ANNUAL %)** | -0.15 | -0.03 | -0.04 | -0.1 | -0.07 | -0.1 | -0.08 | -0.06 | 1 | | | | |
| **POPULATION, TOTAL** | 0.11 | 0.64 | 0.89 | 0.71 | -0.04 | 0.08 | 0.82 | 0.88 | -0.04 | 1 | | | |
| **RENEWABLE ELECTRICITY OUTPUT (% OF TOTAL ELECTRICITY OUTPUT)** | -0.02 | -0.06 | 0.01 | -0.1 | -0.08 | 0.01 | -0.06 | -0.02 | 0.04 | -0.05 | 1 | | |
| **RENEWABLE ENERGY CONSUMPTION (% OF TOTAL ENERGY CONSUMED)** | 0.02 | -0.09 | 0.01 | -0.14 | -0.33 | 0.001 | -0.08 | -0.03 | 0.3 | -0.01 | 0.58 | 1 | |
| **TOTAL GREENHOUSE GAS EMISSIONS (KT OF CO2 EQUIV)** | 0.06 | 0.76 | 0.67 | 0.99 | 0.11 | 0.05 | 0.88 | 0.91 | -0.1 | 0.75 | -0.09 | -0.13 | 1 |

## 8.1A: Model Diagnostics Plots - GDP growth per capita

Before log:



After log:


Residuals vs Fitted
Residuals
Fitted values
1498
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
3336

# 8.1B: P-values and Weights of Log Model (Linear Regression) (Best Model)

```
Coefficients:
                                                                            Estimate Std. Error
(Intercept)                                                                9.023e+00  5.320e-02
`Agricultural land (% of land area)`                                      -1.441e-02  8.088e-04
`Agricultural land (sq. km)`                                               1.042e-07  5.165e-08
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`  1.845e-06  1.044e-06
`CO2 emissions (kt)`                                                      -4.059e-05  4.093e-06
`Methane emissions (kt of CO2 equivalent)`                                -4.053e-05  3.867e-06
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`        -4.082e-05  4.017e-06
`Population growth (annual %)`                                             9.298e-02  1.271e-02
`Population, total`                                                        8.841e-11  3.729e-10
`Renewable electricity output (% of total electricity output)`             4.426e-03  5.912e-04
`Renewable energy consumption (% of total final energy consumption)`      -3.051e-02  8.049e-04
`Total greenhouse gas emissions (kt of CO2 equivalent)`                    3.978e-05  3.959e-06
`Developed Economy`                                                        1.469e+00  2.016e-01
`Developing Economy`                                                      -3.250e-01  2.527e-01
Africa                                                                     2.448e-01  2.544e-01
Asia                                                                       1.473e-01  2.525e-01
`Asia and Pacific`                                                         3.328e-01  2.182e-01
Europe                                                                     2.781e-01  2.002e-01
`Latin America and Caribbean`                                              6.673e-01  2.498e-01
                                                                          t value Pr(>|t|)
(Intercept)                                                               169.599  < 2e-16 ***
`Agricultural land (% of land area)`                                      -17.812  < 2e-16 ***
`Agricultural land (sq. km)`                                                2.017   0.0438 *
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`   1.766   0.0775 .
`CO2 emissions (kt)`                                                       -9.915  < 2e-16 ***
`Methane emissions (kt of CO2 equivalent)`                                -10.482  < 2e-16 ***
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`        -10.162  < 2e-16 ***
`Population growth (annual %)`                                              7.313 3.22e-13 ***
`Population, total`                                                         0.237   0.8126
`Renewable electricity output (% of total electricity output)`              7.487 8.89e-14 ***
`Renewable energy consumption (% of total final energy consumption)`      -37.907  < 2e-16 ***
`Total greenhouse gas emissions (kt of CO2 equivalent)`                    10.047  < 2e-16 ***
`Developed Economy`                                                         7.289 3.85e-13 ***
`Developing Economy`                                                       -1.286   0.1984
Africa                                                                      0.962   0.3359
Asia                                                                        0.583   0.5598
`Asia and Pacific`                                                          1.525   0.1273
Europe                                                                      1.389   0.1648
`Latin America and Caribbean`                                               2.671   0.0076 **
---
```

## 8.1C: P-values and Weights of Log Model (Best Model) (Over 3 periods)

| Linear Regression Model (Analysis over the years) (Overall) | | | | | |
|---|---|---|---|---|---|
| | Trainset RMSE | Testset RMSE | Trainset Adjusted R Squared | Renewable Electricity Output (p-value) | Renewable Energy Consumption (p-value) |
| 1991-1999 | 0.918 | 0.879 | 0.648 | 0.0178 (+ve) | $2*10^{-16}$ (-ve) |
| 2000-2008 | 0.866 | 0.879 | 0.706 | 0.00836 (+ve) | $2*10^{-16}$ (-ve) |
| 2009-2018 | 0.753 | 0.772 | 0.718 | $3.53*10^{-5}$ (+ve) | $2*10^{-16}$ (-ve) |

Trend observable for Renewable electricity output, increasing significance over time

Year 1991-1999:

```
Coefficients:
                                                                          Estimate Std. Error
(Intercept)                                                              8.629e+00  1.031e-01
`Agricultural land (% of land area)`                                    -1.501e-02  1.426e-03
`Agricultural land (sq. km)`                                            -1.012e-09  9.225e-08
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`  2.337e-06  2.183e-06
`CO2 emissions (kt)`                                                    -3.422e-05  1.286e-05
`Methane emissions (kt of CO2 equivalent)`                              -3.613e-05  1.240e-05
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`      -2.646e-05  1.205e-05
`Population growth (annual %)`                                           7.582e-02  2.610e-02
`Population, total`                                                     -9.785e-10  8.642e-10
`Renewable electricity output (% of total electricity output)`           2.574e-03  1.084e-03
`Renewable energy consumption (% of total final energy consumption)`    -2.731e-02  1.539e-03
`Total greenhouse gas emissions (kt of CO2 equivalent)`                  3.371e-05  1.250e-05
`Developed Economy`                                                      9.007e-01  5.980e-01
`Developing Economy`                                                    -1.236e+00  6.540e-01
Africa                                                                   1.143e+00  6.525e-01
Asia                                                                     9.702e-01  6.538e-01
`Asia and Pacific`                                                       9.269e-01  5.843e-01
Europe                                                                   6.757e-01  5.885e-01
`Latin America and Caribbean`                                            1.634e+00  6.469e-01
                                                                        t value Pr(>|t|)
(Intercept)                                                              83.662  < 2e-16 ***
`Agricultural land (% of land area)`                                    -10.521  < 2e-16 ***
`Agricultural land (sq. km)`                                             -0.011  0.99125
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`   1.071  0.28459
`CO2 emissions (kt)`                                                     -2.661  0.00791 **
`Methane emissions (kt of CO2 equivalent)`                               -2.914  0.00364 **
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`       -2.197  0.02825 *
`Population growth (annual %)`                                            2.905  0.00375 **
`Population, total`                                                      -1.132  0.25778
`Renewable electricity output (% of total electricity output)`            2.374  0.01780 *
`Renewable energy consumption (% of total final energy consumption)`    -17.747  < 2e-16 ***
`Total greenhouse gas emissions (kt of CO2 equivalent)`                   2.697  0.00712 **
`Developed Economy`                                                       1.506  0.13230
`Developing Economy`                                                     -1.890  0.05906 .
Africa                                                                    1.752  0.08015 .
Asia                                                                      1.484  0.13816
`Asia and Pacific`                                                        1.586  0.11294
Europe                                                                    1.148  0.25117
`Latin America and Caribbean`                                             2.526  0.01168 *
```

Year 2000 – 2008:

```
Coefficients:
                                                                         Estimate Std. Error
(Intercept)                                                             8.873e+00  8.546e-02
`Agricultural land (% of land area)`                                   -1.334e-02  1.325e-03
`Agricultural land (sq. km)`                                            2.198e-08  8.914e-08
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` 4.646e-07  1.719e-06
`CO2 emissions (kt)`                                                   -7.424e-05  1.317e-05
`Methane emissions (kt of CO2 equivalent)`                             -7.264e-05  1.259e-05
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`     -7.208e-05  1.270e-05
`Population growth (annual %)`                                          1.488e-01  1.817e-02
`Population, total`                                                     6.815e-10  6.632e-10
`Renewable electricity output (% of total electricity output)`          2.642e-03  1.000e-03
`Renewable energy consumption (% of total final energy consumption)`   -3.270e-02  1.325e-03
`Total greenhouse gas emissions (kt of CO2 equivalent)`                 7.264e-05  1.282e-05
`Developed Economy`                                                     1.460e+00  3.687e-01
`Developing Economy`                                                   -3.768e-01  4.729e-01
Africa                                                                  4.011e-01  4.760e-01
Asia                                                                    1.004e-01  4.739e-01
`Asia and Pacific`                                                      6.592e-01  3.972e-01
Europe                                                                  4.186e-01  3.649e-01
`Latin America and Caribbean`                                           8.153e-01  4.687e-01
                                                                       t value Pr(>|t|)
(Intercept)                                                            103.828  < 2e-16 ***
`Agricultural land (% of land area)`                                   -10.069  < 2e-16 ***
`Agricultural land (sq. km)`                                             0.247  0.80529
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` 0.270  0.78700
`CO2 emissions (kt)`                                                    -5.637 2.18e-08 ***
`Methane emissions (kt of CO2 equivalent)`                              -5.769 1.03e-08 ***
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`      -5.674 1.77e-08 ***
`Population growth (annual %)`                                           8.191 6.91e-16 ***
`Population, total`                                                      1.028  0.30439
`Renewable electricity output (% of total electricity output)`           2.642  0.00836 **
`Renewable energy consumption (% of total final energy consumption)`   -24.681  < 2e-16 ***
`Total greenhouse gas emissions (kt of CO2 equivalent)`                  5.666 1.85e-08 ***
`Developed Economy`                                                      3.960 7.97e-05 ***
`Developing Economy`                                                    -0.797  0.42573
Africa                                                                   0.843  0.39957
Asia                                                                     0.212  0.83226
`Asia and Pacific`                                                       1.659  0.09730 .
Europe                                                                   1.147  0.25155
`Latin America and Caribbean`                                            1.739  0.08222 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Year 2009 – 2018

```
Coefficients:
                                                                          Estimate Std. Error
(Intercept)                                                              9.419e+00  7.099e-02
`Agricultural land (% of land area)`                                    -1.405e-02  1.135e-03
`Agricultural land (sq. km)`                                             2.015e-07  7.858e-08
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` 3.250e-06  1.761e-06
`CO2 emissions (kt)`                                                    -2.489e-05  4.653e-06
`Methane emissions (kt of CO2 equivalent)`                              -2.523e-05  4.345e-06
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`      -3.041e-05  4.735e-06
`Population growth (annual %)`                                           3.535e-02  1.941e-02
`Population, total`                                                     -1.617e-10  5.594e-10
`Renewable electricity output (% of total electricity output)`           3.500e-03  8.430e-04
`Renewable energy consumption (% of total final energy consumption)`    -2.571e-02  1.148e-03
`Total greenhouse gas emissions (kt of CO2 equivalent)`                  2.455e-05  4.483e-06
`Developed Economy`                                                      1.234e+00  2.355e-01
`Developing Economy`                                                    -3.321e-01  3.245e-01
Africa                                                                   1.112e-01  3.284e-01
Asia                                                                     2.674e-01  3.233e-01
`Asia and Pacific`                                                       4.912e-01  2.743e-01
Europe                                                                   5.291e-01  2.344e-01
`Latin America and Caribbean`                                            6.970e-01  3.216e-01
                                                                        t value Pr(>|t|)
(Intercept)                                                             132.671  < 2e-16 ***
`Agricultural land (% of land area)`                                    -12.381  < 2e-16 ***
`Agricultural land (sq. km)`                                              2.564   0.0105 *
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`  1.846   0.0652 .
`CO2 emissions (kt)`                                                     -5.348 1.05e-07 ***
`Methane emissions (kt of CO2 equivalent)`                               -5.806 8.10e-09 ***
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`       -6.421 1.91e-10 ***
`Population growth (annual %)`                                            1.821   0.0688 .
`Population, total`                                                      -0.289   0.7725
`Renewable electricity output (% of total electricity output)`            4.151 3.53e-05 ***
`Renewable energy consumption (% of total final energy consumption)`    -22.387  < 2e-16 ***
`Total greenhouse gas emissions (kt of CO2 equivalent)`                   5.475 5.27e-08 ***
`Developed Economy`                                                       5.238 1.90e-07 ***
`Developing Economy`                                                     -1.023   0.3064
Africa                                                                    0.339   0.7350
Asia                                                                      0.827   0.4084
`Asia and Pacific`                                                        1.791   0.0736 .
Europe                                                                    2.257   0.0242 *
`Latin America and Caribbean`                                             2.168   0.0304 *
---
```

## 8.2A: Obtaining best model for Developed countries dataset

| Linear Regression Model (**Developed**) (Predicting GDP per Capita) | | | | | |
|---|---|---|---|---|---|
| | Trainset RMSE | Testset RMSE | Trainset Adjusted R-Squared | Renewable Electricity Output (p-value) | Renewable Energy Consumption (p-value) |
| Log Model **(Best Model)** | 0.722 | 0.630 | 0.569 | 0.8693 (+ve) | 0.0976 (+ve) |
| Removed Independent Variables with p-value > 0.05 (**Based on best model**) | 0.743 | 0.662 | 0.544 | - | - |
| Removed Independent Variables with high VIF (>10) (**Based on best model**) | 0.814 | 0.718 | 0.455 | - | - |

## 8.2B: Obtaining best model for Developing countries dataset

| Linear Regression Model (**Developing**) (Predicting GDP per Capita) | | | | | |
|---|---|---|---|---|---|
| | Trainset RMSE | Testset RMSE | Trainset Adjusted R-Squared | Renewable Electricity Output (p-value) | Renewable Energy Consumption (p-value) |
| Log Model **(Best Model)** | 0.866 | 0.864 | 0.582 | $8.31*10^{-13}$ (+ve) | $2*10^{-16}$ (-ve) |
| Removed Independent Variables with p-value > 0.05 (**Based on best model**) | 0.883 | 0.867 | 0.568 | - | - |

| | | | | | |
|---|---|---|---|---|---|
| Removed Independent Variables with high VIF (>10) (**Based on best model**) | 0.891 | 0.883 | 0.560 | - | - |

## 8.2C: P-values and Weights of Log Models (Best Model) (Developed vs Developing)(Section 8.2)

Developed:

```
Coefficients:
                                                                       Estimate Std. Error
(Intercept)                                                           9.517e+00  1.275e-01
`Agricultural land (% of land area)`                                 -3.331e-03  2.147e-03
`Agricultural land (sq. km)`                                         -2.132e-07  9.297e-08
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`  1.125e-05  6.950e-06
`CO2 emissions (kt)`                                                 -1.205e-05  5.599e-06
`Methane emissions (kt of CO2 equivalent)`                           -3.047e-05  5.271e-06
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`    1.367e-05  8.722e-06
`Population growth (annual %)`                                        7.675e-01  3.722e-02
`Population, total`                                                  -2.700e-09  2.223e-09
`Renewable electricity output (% of total electricity output)`        3.731e-04  2.267e-03
`Renewable energy consumption (% of total final energy consumption)`  7.412e-03  4.468e-03
`Total greenhouse gas emissions (kt of CO2 equivalent)`               1.268e-05  5.439e-06
                                                                     t value Pr(>|t|)
(Intercept)                                                           74.621  < 2e-16 ***
`Agricultural land (% of land area)`                                  -1.551   0.1213
`Agricultural land (sq. km)`                                          -2.293   0.0222 *
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`   1.619   0.1058
`CO2 emissions (kt)`                                                  -2.152   0.0317 *
`Methane emissions (kt of CO2 equivalent)`                            -5.781 1.14e-08 ***
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`     1.567   0.1176
`Population growth (annual %)`                                        20.619  < 2e-16 ***
`Population, total`                                                   -1.215   0.2249
`Renewable electricity output (% of total electricity output)`         0.165   0.8693
`Renewable energy consumption (% of total final energy consumption)`   1.659   0.0976 .
`Total greenhouse gas emissions (kt of CO2 equivalent)`                2.331   0.0201 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7271 on 664 degrees of freedom
Multiple R-squared:  0.5759,	Adjusted R-squared:  0.5689
F-statistic: 81.97 on 11 and 664 DF,  p-value: < 2.2e-16
```

Developing:

```
Coefficients:
                                                                              Estimate Std. Error
(Intercept)                                                                  9.022e+00  5.614e-02
`Agricultural land (% of land area)`                                        -9.101e-03  9.686e-04
`Agricultural land (sq. km)`                                                -7.768e-08  6.603e-08
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`  2.800e-06  1.147e-06
`CO2 emissions (kt)`                                                        -3.594e-05  5.498e-06
`Methane emissions (kt of CO2 equivalent)`                                  -3.532e-05  5.206e-06
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`          -3.109e-05  5.261e-06
`Population growth (annual %)`                                               3.435e-02  1.419e-02
`Population, total`                                                         -1.832e-09  4.317e-10
`Renewable electricity output (% of total electricity output)`               4.576e-03  6.357e-04
`Renewable energy consumption (% of total final energy consumption)`        -3.304e-02  7.509e-04
`Total greenhouse gas emissions (kt of CO2 equivalent)`                      3.521e-05  5.319e-06
                                                                            t value Pr(>|t|)
(Intercept)                                                                 160.698  < 2e-16 ***
`Agricultural land (% of land area)`                                         -9.396  < 2e-16 ***
`Agricultural land (sq. km)`                                                 -1.176   0.2396
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)`    2.441   0.0147 *
`CO2 emissions (kt)`                                                         -6.536 7.81e-11 ***
`Methane emissions (kt of CO2 equivalent)`                                   -6.785 1.49e-11 ***
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`           -5.909 3.97e-09 ***
`Population growth (annual %)`                                                2.421   0.0156 *
`Population, total`                                                          -4.243 2.30e-05 ***
`Renewable electricity output (% of total electricity output)`                7.199 8.31e-13 ***
`Renewable energy consumption (% of total final energy consumption)`        -44.000  < 2e-16 ***
`Total greenhouse gas emissions (kt of CO2 equivalent)`                       6.619 4.53e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8688 on 2200 degrees of freedom
Multiple R-squared:  0.5845,    Adjusted R-squared:  0.5824
F-statistic: 281.3 on 11 and 2200 DF,  p-value: < 2.2e-16
```

## 8.2D: Developed vs Developing (Analysis over time periods)

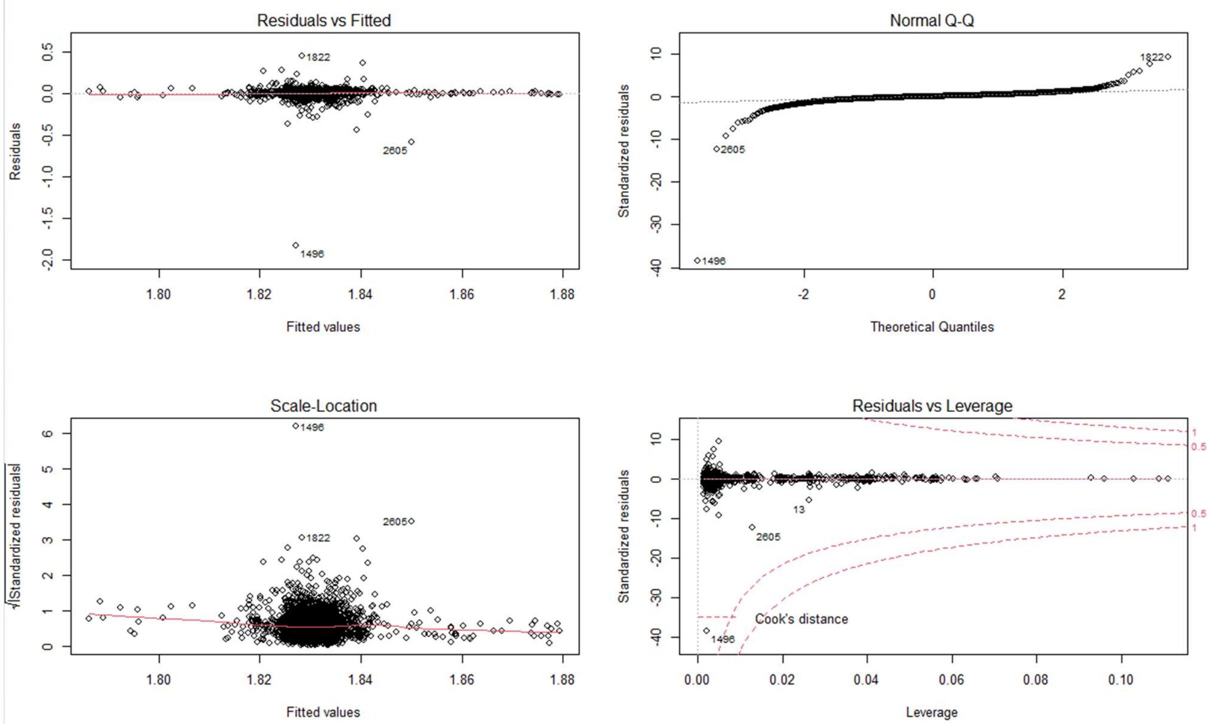
Developed:

| Linear Regression Model (Analysis over the years) (Developed Countries) | | | | | |
|---|---|---|---|---|---|
| | Trainset RMSE | Testset RMSE | Trainset Adjusted R-Squared | Renewable Electricity Output (p-value) | Renewable Energy Consumption (p-value) |
| 1991-1999 | 0.705 | 0.846 | 0.611 | 0.0582 (+ve) | 0.3866 (-ve) |
| 2000-2008 | 0.617 | 0.695 | 0.618 | 0.000749 (+ve) | 0.004929 (-ve) |
| 2009-2018 | 0.403 | 0.436 | 0.734 | 0.01387 (-ve) | $7.65*10^{-6}$ (+ve) |

Developing:

| Linear Regression Model (Analysis over the years) (Developing Countries) | | | | | |
|---|---|---|---|---|---|
| | Trainset RMSE | Testset RMSE | Trainset Adjusted R-Squared | Renewable Electricity Output (p-value) | Renewable Energy Consumption (p-value) |
| 1991-1999 | 0.892 | 0.898 | 0.531 | 0.0582 (+ve) | $2*10^{-16}$ (-ve) |
| 2000-2008 | 0.885 | 0.814 | 0.598 | 0.000749 (+ve) | $2*10^{-16}$ (-ve) |
| 2009-2018 | 0.755 | 0.728 | 0.619 | 0.007252 (+ve) | $2*10^{-16}$ (-ve) |

## 8.3: Linear Regression Model (GDP per capita growth) (Model diagnostic plots)



## 9.1A: Continuous CART Model Comparisons

| | Min. CP | Root Node Error | MSE | RMSE | Most Significant Variable Before Pruning (%) | Second Most Significant Variable Before Pruning (%) | Most Significant Variable After Pruning (%) | Second Most Significant Variable After Pruning (%) |
|---|---|---|---|---|---|---|---|---|
| **Original Data (Unscaled)** | 0.0019071 | 2.53 | 0.72611 (0.287*2.53) | 0.85212 | Renewable energy consumption (% of total final energy consumption) (25%) | Population growth (annual %) (11%) | Renewable energy consumption (% of total final energy consumption) (25%) | Population growth (annual %) (11%) |
| **Original Data - Developed** | 0.024606 | 1.12 | 0.30464 (1.12*0.272) | 0.55194 | Population growth (annual %) (54%) | Agricultural land (% of land area) (8%) | Population growth (annual %) (63%) | Agricultural land (% of land area) (8%) |
| **Original Data - Developing** | 0.016942 | 1.99 | 0.79800 (1.99*0.401) | 0.8933 | Renewable energy consumption (% of total final energy consumption) (31%) | Population growth (annual %) (12%) | Renewable energy consumption (% of total final energy consumption) (32%) | Africa (12%) |
| **Cleaned Data** | 0.0023952 | 2.51 | 0.66766 (0.266*2.51) | 0.8171 | Renewable energy consumption (% of total final energy consumption) (26%) | Population growth (annual %) (10%) | Renewable energy consumption (% of total final energy consumption) (26%) | Population growth (annual %) (10%) |

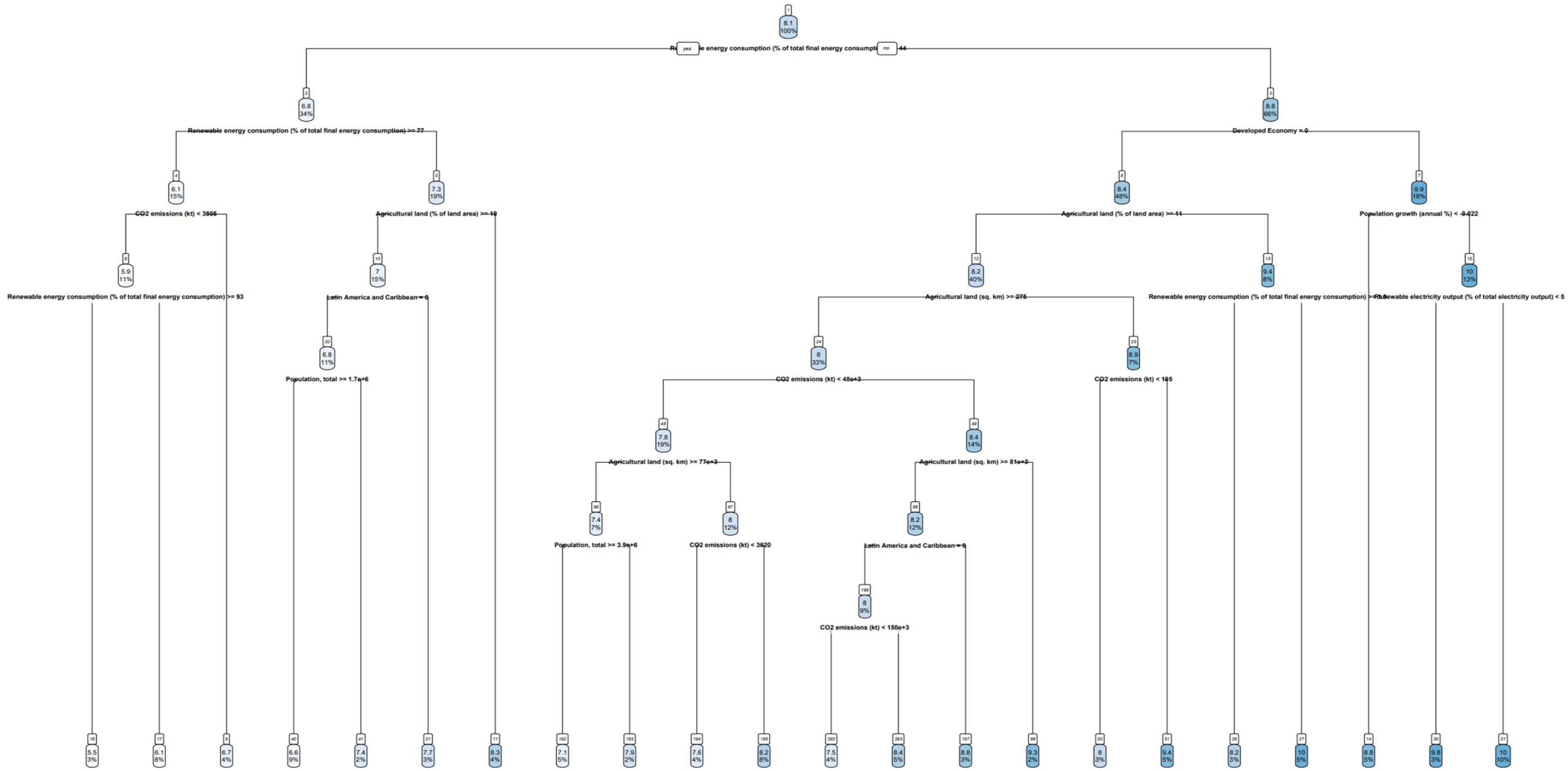
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **After Removal of Variables** | 0.0031058 | 2.51 | 0.77559 (2.51*0.309) | 0.88068 | Renewable energy consumption (% of total final energy consumption) (29%) | CO2 emissions (kt) (14%) | Renewable energy consumption (% of total final energy consumption) (30%) | CO2 emissions (kt) (14%) |
| **Cleaned Data - Developed** | 0.0067565 | 1.11 | 0.27417 (1.11*0.247) | 0.52361 | Population growth (annual %) (32%) | Renewable energy consumption (% of total final energy consumption) (10%) | Population growth (annual %) (33%) | Renewable energy consumption (% of total final energy consumption) (10%) |
| **Cleaned Data - Developing** | 0.0034704 | 1.98 | 0.52452 (1.88*0.279) | 0.72424 | Renewable energy consumption (% of total final energy consumption) (26%) | Population growth (annual %) (10%) | Renewable energy consumption (% of total final energy consumption) (27%) | Population growth (annual %) (10%) |

## 9.1B: Top 10 variable importance for Categorical CART

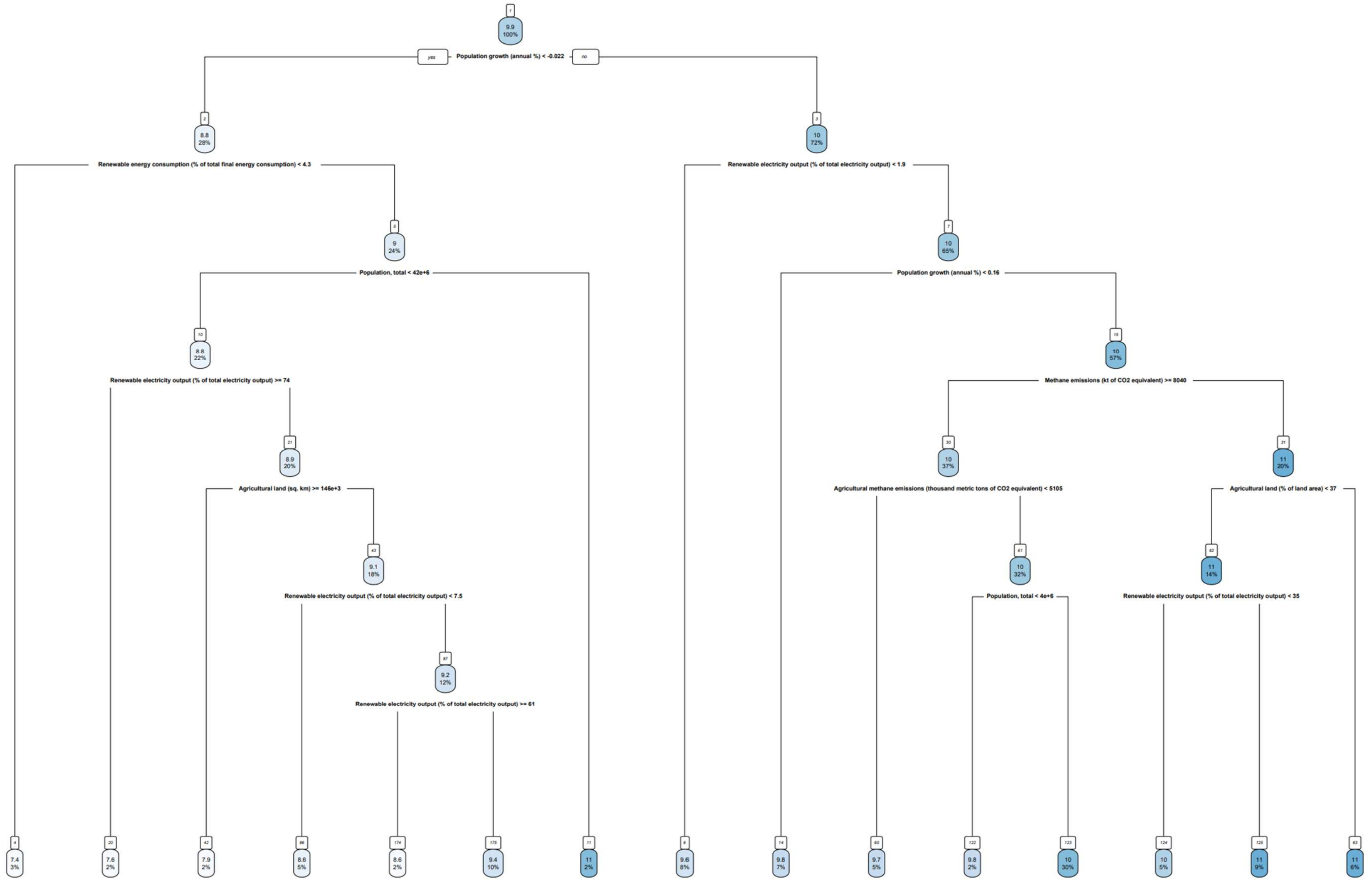
| Variable Importance | Overall Data | Developed | Developing |
|---|---|---|---|
| 1 | Renewable energy consumption (**26%**) | Population growth (**33%**) | Renewable energy consumption (27%) |
| 2 | Population growth (**10%**) | Renewable energy consumption (**10%**) | Population growth (**10%**) |
| 3 | Africa (**9%**) | CO2 emissions (kt) (**8%**) | Africa (**10%**) |
| 4 | CO2 emissions (kt) (**9%**) | Methane emissions (8%) | CO2 emissions (kt) (**9%**) |
| 5 | Developed Economy (**6%**) | Renewable electricity output (7%) | Renewable electricity output (9%) |
| 6 | Renewable electricity output (6%) | Agricultural land (sq. km) (**7%**) | Nitrous oxide emissions (7%) |
| 7 | Europe (**5%**) | Agricultural land (% of land area) (**6%**) | Total greenhouse gas emissions (6%) |
| 8 | Total greenhouse gas emissions (5%) | Population, total (**6%**) | Agricultural land (sq. km) (**5%**) |
| 9 | Nitrous oxide emissions (4%) | Agricultural methane emissions (6%) | Methane emissions (5%) |
| 10 | Agricultural land (% of land area) (**4%**) | Nitrous oxide emissions (5%) | Population, total (**4%**) |

Highlighted in Green are Environmental Factors

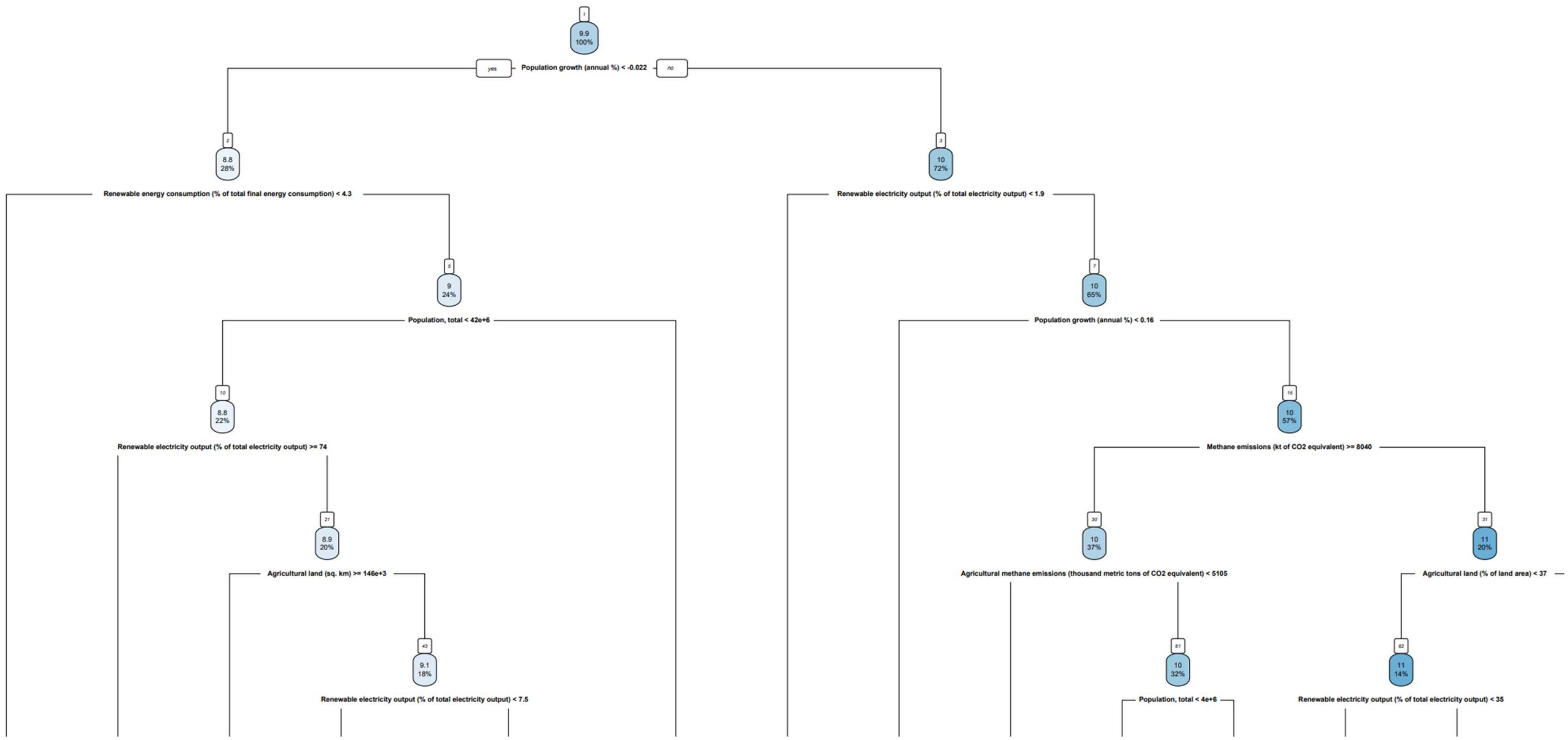
## 9.1C: Overall (Cleaned) Pruned CART Tree



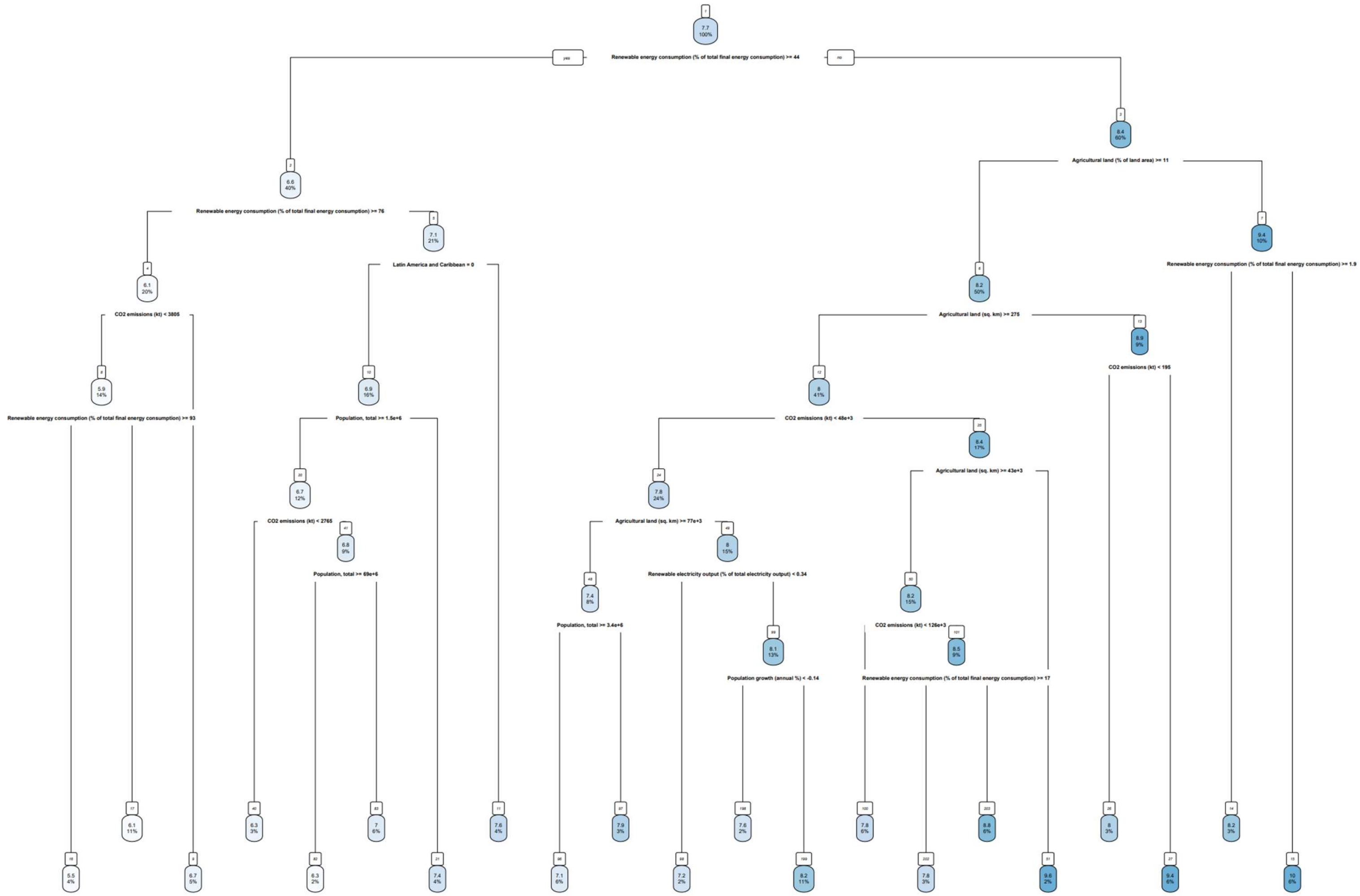
# 9.2A(1): Developed (Cleaned) Pruned CART Tree



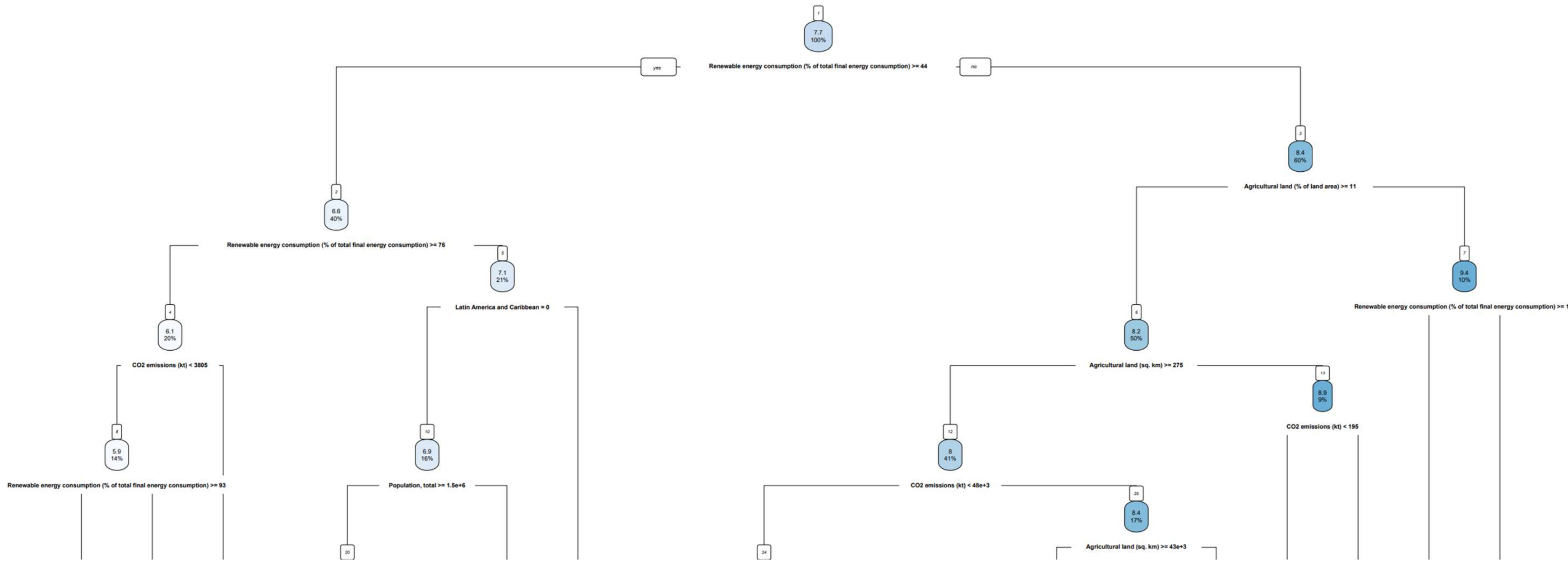
# 9.2A(2): Developed (Cleaned) Pruned CART Tree (Zoomed in)



# 9.2B (1): Developing (Cleaned) Pruned CART Tree



<u>9.2B (2): Developing (Cleaned) Pruned CART Tree (Zoomed in)</u>



## 10.2A: Logistic Regression Model (Thresholds)

| | Threshold | Trainset Accuracy | TPR | TNR | FPR | FNR |
|---|---|---|---|---|---|---|
| Original Model (**Best Model**) | 0.6 | 0.733 | 0.106 | 0.988 | 0.012 | 0.894 |
| | **0.5** | **0.736** | **0.285** | **0.921** | **0.079** | **0.715** |
| | 0.4 | 0.722 | 0.423 | 0.844 | 0.156 | 0.577 |
| | Threshold | Testset Accuracy | TPR | TNR | FPR | FNR |
| Original Model (**Best Model**) | 0.6 | 0.719 | 0.088 | 0.977 | 0.023 | 0.912 |
| | **0.5** | **0.732** | **0.270** | **0.921** | **0.079** | **0.730** |
| | 0.4 | 0.705 | 0.376 | 0.839 | 0.161 | 0.624 |

## 10.2B: Logistic Regression Model (p-value and odds ratio)

```
Coefficients:
                                                                          Estimate Std. Error
(Intercept)                                                             -8.967e-01  1.358e-01
`Agricultural land (% of land area)`                                    -1.417e-03  2.056e-03
`Agricultural land (sq. km)`                                             3.165e-07  1.287e-07
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` -3.757e-06  3.129e-06
`CO2 emissions (kt)`                                                     3.243e-05  1.113e-05
`Methane emissions (kt of CO2 equivalent)`                               3.073e-05  1.055e-05
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`       3.193e-05  1.159e-05
`Population growth (annual %)`                                          -2.973e-01  4.073e-02
`Population, total`                                                      2.800e-09  1.119e-09
`Renewable electricity output (% of total electricity output)`           2.952e-03  1.599e-03
`Renewable energy consumption (% of total final energy consumption)`     4.533e-03  2.218e-03
`Total greenhouse gas emissions (kt of CO2 equivalent)`                 -3.172e-05  1.079e-05
`Developed Economy`                                                      2.409e-01  5.005e-01
`Developing Economy`                                                     5.051e-02  5.980e-01
Africa                                                                  -5.089e-01  6.072e-01
Asia                                                                     3.487e-01  5.982e-01
`Asia and Pacific`                                                      -5.841e-02  5.312e-01
Europe                                                                   7.145e-01  4.937e-01
`Latin America and Caribbean`                                           -5.274e-01  5.931e-01
                                                                        z value Pr(>|z|)
(Intercept)                                                              -6.606 3.96e-11 ***
`Agricultural land (% of land area)`                                     -0.689  0.49070
`Agricultural land (sq. km)`                                              2.459  0.01395 *
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` -1.201  0.22987
`CO2 emissions (kt)`                                                      2.915  0.00356 **
`Methane emissions (kt of CO2 equivalent)`                                2.912  0.00359 **
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`        2.754  0.00589 **
`Population growth (annual %)`                                           -7.299 2.89e-13 ***
`Population, total`                                                       2.502  0.01233 *
`Renewable electricity output (% of total electricity output)`            1.846  0.06493 .
`Renewable energy consumption (% of total final energy consumption)`      2.044  0.04098 *
`Total greenhouse gas emissions (kt of CO2 equivalent)`                  -2.939  0.00330 **
`Developed Economy`                                                       0.481  0.63032
`Developing Economy`                                                      0.084  0.93268
Africa                                                                   -0.838  0.40195
Asia                                                                      0.583  0.55992
`Asia and Pacific`                                                       -0.110  0.91245
Europe                                                                    1.447  0.14783
`Latin America and Caribbean`                                            -0.889  0.37388
```

## 10.2C: Logistic Regression Model (Variable importance overtime)

| Overall Data (Split by years) | | | | |
|---|---|---|---|---|
| Logistic Regression Model (At threshold of **0.5**) | Variable Importance of Renewable Electricity Output (p-value, odds ratio) | Variable Importance of Renewable Energy Consumption (p-value, odds ratio) | Trainset Accuracy | Testset Accuracy |
| 1991 - 1999 | p-value: 0.0712<br>odds ratio: 0.9949 | p-value: 0.6537<br>odds ratio: 1.00177 | 0.744 | 0.730 |
| 2000 - 2008 | p-value: 0.0393<br>odds ratio: 1.0079 | p-value: 0.3878<br>odds ratio: 0.9969 | 0.699 | 0.716 |
| 2009 - 2018 | p-value: 0.7949<br>odds ratio: 1.00079 | p-value: 0.000754<br>odds ratio: 1.0137 | 0.797 | 0.787 |

## 10.3A: Logistic Regression Model (Comparing of accuracy for developed countries dataset)

| Developed Countries | | | | | |
|---|---|---|---|---|---|
| Logistic Regression Model (At threshold of **0.5**) | **Trainset Accuracy** | TPR | TNR | FPR | FNR |
| Original Model (**Best Model**) | 0.634 | 0.679 | 0.586 | 0.414 | 0.321 |
| Removing of **Less Significant Variables (Based on best Model)** | **0.725** | **0.084** | **0.986** | **0.014** | **0.915** |

| Logistic Regression Model Developed Countries (At threshold of **0.5**) | **Testset Accuracy** | TPR | TNR | FPR | FNR |
|---|---|---|---|---|---|
| Original Model (**Best Model**) | 0.614 | 0.696 | 0.528 | 0.472 | 0.304 |
| Removing of **Less Significant Variables (Based on best Model)** | **0.716** | **0.071** | **0.979** | **0.021** | **0.929** |

## 10.3B: Logistic Regression Model (Comparing of accuracy for developing countries dataset)

| Developing Countries | | | | | |
|---|---|---|---|---|---|
| Logistic Regression Model (At threshold of **0.5**) | **Trainset Accuracy** | TPR | TNR | FPR | FNR |
| Original Model (**Best Model**) | **0.778** | **0.057** | **0.994** | **0.006** | **0.943** |
| Re-balancing of number of Nos and Yes to have equal proportion in Acceptable Growth (**Based on Best Model**) | 0.633 | 0.477 | 0.785 | 0.215 | 0.523 |
| Removing of **Less Significant Variables (p-value > 0.05)** (**Based on best model**) | 0.775 | 0.059 | 0.989 | 0.011 | 0.941 |

| Logistic Regression Model Developing Countries (At threshold of **0.5**) | **Testset Accuracy** | TPR | TNR | FPR | FNR |
|---|---|---|---|---|---|
| Original Model (**Best Model**) | **0.785** | **0.083** | **0.995** | **0.005** | **0.917** |
| Re-balancing of number of Nos and Yes to have equal proportion in Acceptable Growth (**Based on best model**) | 0.651 | 0.530 | 0.768 | 0.232 | 0.470 |
| Removing of **Less Significant Variables (p-value > 0.05)** (**Based on best model**) | 0.783 | 0.069 | 0.996 | 0.004 | 0.931 |

## 10.3C: Logistic Regression Model (Developed vs developing) (p-value and odds ratio)

**Developed:**

```
Coefficients:
                                                                          Estimate Std. Error
(Intercept)                                                              6.703e-01  3.764e-01
`Agricultural land (% of land area)`                                     4.364e-03  6.366e-03
`Agricultural land (sq. km)`                                             1.769e-07  2.824e-07
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` -3.359e-05  2.069e-05
`CO2 emissions (kt)`                                                    -2.668e-05  1.825e-05
`Methane emissions (kt of CO2 equivalent)`                              -2.669e-05  1.716e-05
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`      -1.082e-06  2.629e-05
`Population growth (annual %)`                                          -5.930e-01  1.212e-01
`Population, total`                                                     -3.374e-08  7.178e-09
`Renewable electricity output (% of total electricity output)`          -1.602e-03  6.626e-03
`Renewable energy consumption (% of total final energy consumption)`    -3.316e-03  1.305e-02
`Total greenhouse gas emissions (kt of CO2 equivalent)`                  2.747e-05  1.773e-05
                                                                        z value Pr(>|z|)
(Intercept)                                                               1.781    0.075 .
`Agricultural land (% of land area)`                                      0.686    0.493
`Agricultural land (sq. km)`                                              0.627    0.531
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` -1.624    0.104
`CO2 emissions (kt)`                                                     -1.462    0.144
`Methane emissions (kt of CO2 equivalent)`                               -1.556    0.120
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`       -0.041    0.967
`Population growth (annual %)`                                           -4.895 9.85e-07 ***
`Population, total`                                                      -4.701 2.59e-06 ***
`Renewable electricity output (% of total electricity output)`           -0.242    0.809
`Renewable energy consumption (% of total final energy consumption)`     -0.254    0.799
`Total greenhouse gas emissions (kt of CO2 equivalent)`                   1.550    0.121
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 938.19  on 676  degrees of freedom
Residual deviance: 859.00  on 665  degrees of freedom
AIC: 883

Number of Fisher Scoring iterations: 4
```

**Developing:**

```
Coefficients:
                                                                          Estimate Std. Error
(Intercept)                                                             -5.471e-01  1.548e-01
`Agricultural land (% of land area)`                                    -5.834e-03  2.788e-03
`Agricultural land (sq. km)`                                             8.238e-08  1.932e-07
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` -2.265e-07  3.730e-06
`CO2 emissions (kt)`                                                     4.165e-05  2.115e-05
`Methane emissions (kt of CO2 equivalent)`                               3.905e-05  2.005e-05
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`       4.016e-05  2.066e-05
`Population growth (annual %)`                                          -3.355e-01  4.944e-02
`Population, total`                                                      1.437e-09  1.370e-09
`Renewable electricity output (% of total electricity output)`          -8.852e-04  1.803e-03
`Renewable energy consumption (% of total final energy consumption)`     3.191e-03  2.206e-03
`Total greenhouse gas emissions (kt of CO2 equivalent)`                 -4.035e-05  2.048e-05
                                                                        z value Pr(>|z|)
(Intercept)                                                              -3.533  0.00041 ***
`Agricultural land (% of land area)`                                     -2.092  0.03640 *
`Agricultural land (sq. km)`                                              0.426  0.66981
`Agricultural methane emissions (thousand metric tons of CO2 equivalent)` -0.061  0.95158
`CO2 emissions (kt)`                                                      1.969  0.04892 *
`Methane emissions (kt of CO2 equivalent)`                                1.948  0.05144 .
`Nitrous oxide emissions (thousand metric tons of CO2 equivalent)`        1.944  0.05195 .
`Population growth (annual %)`                                           -6.785 1.16e-11 ***
`Population, total`                                                       1.049  0.29434
`Renewable electricity output (% of total electricity output)`           -0.491  0.62354
`Renewable energy consumption (% of total final energy consumption)`      1.447  0.14789
`Total greenhouse gas emissions (kt of CO2 equivalent)`                  -1.970  0.04884 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2389.3  on 2212  degrees of freedom
Residual deviance: 2264.1  on 2201  degrees of freedom
AIC: 2288.1

Number of Fisher Scoring iterations: 5
```

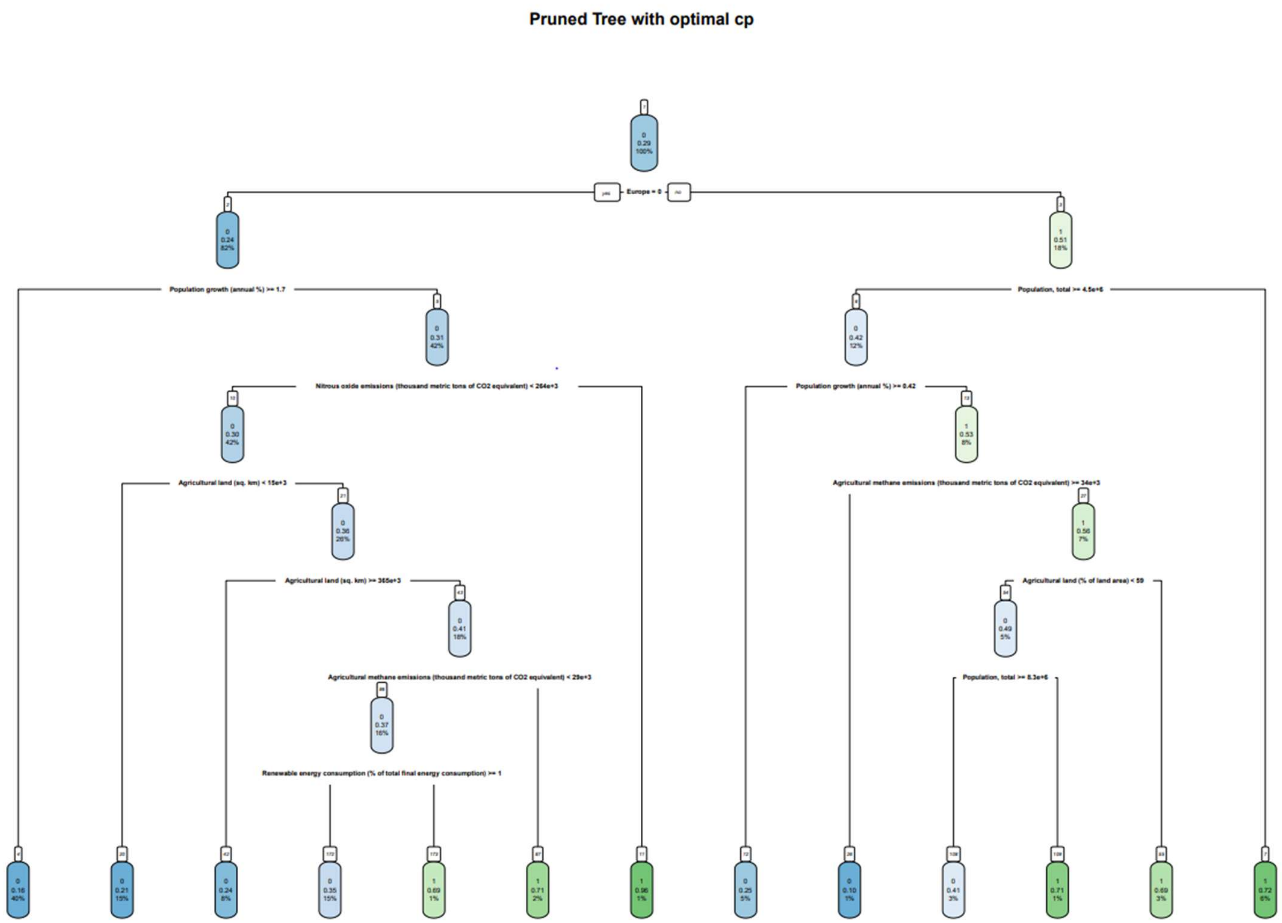
## 10.3D: Logistic Regression Model (Developed countries split over years)

| Developed Countries | | | | |
|---|---|---|---|---|
| Logistic Regression Model (At threshold of **0.5**) | Variable Importance of Renewable Electricity Output (p-value, odds ratio) | Variable Importance of Renewable Energy Consumption (p-value) | Trainset Accuracy | Testset Accuracy |
| 1991 - 1999 | p-value: 0.6089<br>odds ratio: 0.9924 | p-value: 0.3697<br>odds ratio: 1.0288 | 0.809 | 0.647 |
| 2000 - 2008 | p-value: 0.9848<br>odds ratio: 0.9997 | p-value: 0.6909<br>odds ratio: 0.9879 | 0.678 | 0.711 |
| 2009 - 2018 | p-value: 0.5097<br>odds ratio: 1.008 | p-value: 0.272<br>odds ratio:0.973 | 0.726 | 0.722 |

## 11.2A: Variable Importance of Renewable Energy Variables for Categorical CART

| | Renewable Energy Consumption | Renewable Electricity Output |
|---|---|---|
| Variable Importance | 67.29 (5%) | 68.40 (5%) |

<u>11.2B: Optimal Tree for Categorical CART</u>



## 11.2C: Variable Importance of Renewable Energy Variables for Categorical CART

| | | | | |
|---|---|---|---|---|
| Developing - Renewable energy consumption (% of total final energy consumption) Variable Importance | 34.66 (8%) | 0 (0%) | 23.17 (12%) | 5.47 (2%) |
| Developing - Renewable electricity output (% of total electricity output) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |

## 12: Summary Details of Variables (Section 4.4)

All the data are sourced from The World Bank Data Bank, and are World Development Indicators.

| Name | Long definition | Source | Statistical concept and methodology | Development relevance | Limitations and exceptions |
|---|---|---|---|---|---|
| Agricultural land (% of land area) | Agricultural land refers to the share of land area that is arable, under permanent crops, and under permanent pastures. Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary | Food and Agriculture Organization, electronic files and web site. | Agriculture is still a major sector in many economies, and agricultural activities provide developing countries with food and revenue. But agricultural activities also can degrade natural resources. Poor farming practices can cause soil erosion and loss of soil fertility. Efforts to increase productivity by using chemical fertilizers, pesticides, and intensive irrigation have environmental costs and health impacts. Excessive use of chemical fertilizers can alter the chemistry of soil. Pesticide poisoning is common in developing countries. And salinization of irrigated land diminishes soil fertility. Thus, | Agricultural land covers more than one-third of the world's land area, with arable land representing less than one-third of agricultural land (about 10 percent of the world's land area). Agricultural land constitutes only a part of any country's total area, which can include areas not suitable for agriculture, such as forests, mountains, and inland water bodies. In many industrialized countries, agricultural land is subject to zoning regulations. In the context of zoning, agricultural land (or more properly agriculturally zoned land) refers to plots that may be used for agricultural | The data are collected by the Food and Agriculture Organization of the United Nations (FAO) from official national sources through annual questionnaires and are supplemented with information from official secondary data sources. The secondary sources cover official country data from websites of national ministries, national publications and related country data reported by various international organizations.. The FAO tries to impose standard definitions and reporting |

| | | | | | |
|---|---|---|---|---|---|
| | meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded. Land under permanent crops is land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber. This category includes land under flowering shrubs, fruit trees, nut trees, and vines, but excludes land under trees grown for wood or timber. Permanent pasture is land used for five or more years for | | inappropriate use of inputs for agricultural production has far-reaching effects. Agricultural land is also sometimes classified as irrigated and non-irrigated land. In arid and semi-arid countries agriculture is often confined to irrigated land, with very little farming possible in non-irrigated areas. Land abandoned as a result of shifting cultivation is excluded from Arable land. Data on agricultural land are valuable for conducting studies on a various perspectives concerning agricultural production, food security and for deriving cropping intensity among others uses. Agricultural land indicator, along with land-use indicators, can also elucidate the environmental sustainability of countries' agricultural practices. Total land area does not include inland water bodies such as major rivers and lakes. Variations from year to year may be due to updated or revised data rather than to change in area. | activities, regardless of the physical type or quality of land. FAO's agricultural land data contains a wide range of information on variables that are significant for: understanding the structure of a country's agricultural sector; making economic plans and policies for food security; deriving environmental indicators, including those related to investment in agriculture and data on gross crop area and net crop area which are useful for policy formulation and monitoring. There is no single correct mix of inputs to the agricultural land, as it is dependent on local climate, land quality, and economic development; appropriate levels and application rates vary by country and over time and depend on the type of crops, the climate and soils, and the production process used. | methods, but complete consistency across countries and over time is not possible. Thus, data on agricultural land in different climates may not be comparable. For example, permanent pastures are quite different in nature and intensity in African countries and dry Middle Eastern countries. Data on agricultural employment, in particular, should be used with caution. In many countries much agricultural employment is informal and unrecorded, including substantial work performed by women and children. To address some of these concerns, this indicator is heavily footnoted in the database in sources, definition, and coverage. |

| | | | | | |
|---|---|---|---|---|---|
| | forage, including natural and cultivated crops. | | | | |
| Agricultural land (sq. km) | Agricultural land refers to the share of land area that is arable, under permanent crops, and under permanent pastures. Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded. Land under permanent crops is land cultivated with crops that occupy | Food and Agriculture Organization, electronic files and web site. | Agricultural land constitutes only a part of any country's total area, which can include areas not suitable for agriculture, such as forests, mountains, and inland water bodies. Three components of the agricultural land are a) arable land - land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow, b) permanent pasture - land used for five or more years for forage, including natural and cultivated crops, and c) and under permanent crops - land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber; land under flowering shrubs, fruit trees, nut trees, and vines is included, but land under trees grown for wood or timber is not. Agricultural land is also sometimes classified as irrigated and non-irrigated land. In arid and semi-arid countries agriculture is often confined to irrigated land, with very little farming possible in non-irrigated areas. Land abandoned as a result of shifting cultivation is excluded from arable land. Data on agricultural land are valuable for conducting studies on a | Agricultural land covers more than one-third of the world's land area. In many industrialized countries, agricultural land is subject to zoning regulations. In the context of zoning, agricultural land (or more properly agriculturally zoned land) refers to plots that may be used for agricultural activities, regardless of the physical type or quality of land. FAO's agricultural land data contains a wide range of information on variables that are significant for understanding the structure of a country's agricultural sector; making economic plans and policies for food security; and deriving environmental indicators, including those related to investment in agriculture and data on gross crop area and net crop area which are useful for policy formulation and monitoring. Agriculture is still a major sector in many economies, and agricultural activities provide developing countries with food and revenue. But agricultural activities also can degrade natural resources. Poor farming practices can cause soil erosion and loss of soil fertility. Efforts to increase productivity by using chemical fertilizers, pesticides, and intensive irrigation have | The data are collected by the Food and Agriculture Organization of the United Nations (FAO) through annual questionnaires. The FAO tries to impose standard definitions and reporting methods, but complete consistency across countries and over time is not possible. Thus, data on agricultural land in different climates may not be comparable. For example, permanent pastures are quite different in nature and intensity in African countries and dry Middle Eastern countries. Data on agricultural employment, in particular, should be used with caution. In many countries much agricultural employment is informal and unrecorded, including substantial work performed by women and children. To address some of these concerns, this indicator is heavily footnoted in the database in sources, definition, and coverage. The secondary sources cover official country data from websites of national ministries, national publications and related |

| | | | | | |
|---|---|---|---|---|---|
| | the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber. This category includes land under flowering shrubs, fruit trees, nut trees, and vines, but excludes land under trees grown for wood or timber.<br>Permanent pasture is land used for five or more years for forage, including natural and cultivated crops. | | various perspectives concerning agricultural production, food security and for deriving cropping intensity among others uses. Agricultural land indicator, along with land-use indicators, can also elucidate the environmental sustainability of countries' agricultural practices. | environmental costs and health impacts. Excessive use of chemical fertilizers can alter the chemistry of soil. Pesticide poisoning is common in developing countries. And salinization of irrigated land diminishes soil fertility. Thus, inappropriate use of inputs for agricultural production has far-reaching effects. There is no single correct mix of inputs to the agricultural land, as it is dependent on local climate, land quality, and economic development; appropriate levels and application rates vary by country and over time and depend on the type of crops, the climate and soils, and the production process used. | country data reported by various international organizations. |
| Agricultural methane emissions (% of total) | Agricultural methane emissions are emissions from animals, animal waste, rice production, agricultural waste burning (nonenergy, on- | World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment | IPCC category 4 = Agriculture. Methane emissions result largely from agricultural activities, industrial production landfills and wastewater treatment, and other sources such as tropical forest and other vegetation fires. The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective contributions of different gases to be compared. A kilogram of methane is 21 times as | The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average | National reporting to the United Nations Framework Convention on Climate Change that follows the Intergovernmental Panel on Climate Change guidelines is based on national emission inventories and covers all sources of anthropogenic carbon dioxide emissions as well as carbon sinks (such as forests). To estimate emissions, the countries that are Parties to the Climate Change |

| | | | | | |
|---|---|---|---|---|---|
| | site), and savanna burning. | Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/. | effective at trapping heat in the earth's atmosphere as a kilogram of carbon dioxide within 100 years. | emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), Sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working | Convention (UNFCCC) use complex, state-of-the-art methodologies recommended by the Intergovernmental Panel on Climate Change (IPCC). |

| | | | | | |
|---|---|---|---|---|---|
| | | | | towards curbing CO2 emissions globally. | |
| Agricultural methane emissions (thousand metric tons of CO2 equivalent) | Agricultural methane emissions are emissions from animals, animal waste, rice production, agricultural waste burning (nonenergy, on-site), and savanna burning. | Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: [https://www.climatewatchdata.org/ghg-emissions](https://www.climatewatchdata.org/ghg-emissions). | IPCC category 4 = Agriculture. Expressed in CO2 equivalent using the GWP100 metric of the Second Assessment Report of IPCC and include CH4 (GWP100=21). Methane emissions result largely from agricultural activities, industrial production landfills and wastewater treatment, and other sources such as tropical forest and other vegetation fires. The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective contributions of different gases to be compared. A kilogram of methane is 21 times as effective at trapping heat in the earth's atmosphere as a kilogram of carbon dioxide within 100 years. The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective contributions of different gases to be compared. | The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and | National reporting to the United Nations Framework Convention on Climate Change that follows the Intergovernmental Panel on Climate Change guidelines is based on national emission inventories and covers all sources of anthropogenic carbon dioxide emissions as well as carbon sinks (such as forests). To estimate emissions, the countries that are Parties to the Climate Change Convention (UNFCCC) use complex, state-of-the-art methodologies recommended by the Intergovernmental Panel on Climate Change (IPCC). |

| | | | | | |
|---|---|---|---|---|---|
| | | | | climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), Sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working towards curbing CO2 emissions globally. | |
| CO2 emissions (kt) | Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring. | Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate | Carbon dioxide emissions, largely by-products of energy production and use, account for the largest share of greenhouse gases, which are associated with global warming. Anthropogenic carbon dioxide emissions result primarily from fossil fuel combustion and cement manufacturing. In combustion different fossil fuels release different amounts of carbon dioxide for the same level of energy use: oil releases about 50 percent more carbon dioxide than natural gas, and coal releases about twice as much. Cement manufacturing releases about half a metric ton of carbon dioxide for each metric ton of cement produced. Data for carbon dioxide emissions include gases from | Carbon dioxide (CO2) is naturally occurring gas fixed by photosynthesis into organic matter. A byproduct of fossil fuel combustion and biomass burning, it is also emitted from land use changes and other industrial processes. It is the principal anthropogenic greenhouse gas that affects the Earth's radiative balance. It is the reference gas against which other greenhouse gases are measured, thus having a Global Warming Potential of 1. Burning of carbon-based fuels since the industrial revolution has rapidly increased concentrations of atmospheric carbon dioxide, increasing the rate of global warming and causing anthropogenic | The U.S. Department of Energy's Carbon Dioxide Information Analysis Center (CDIAC) calculates annual anthropogenic emissions from data on fossil fuel consumption (from the United Nations Statistics Division's World Energy Data Set) and world cement manufacturing (from the U.S. Department of Interior's Geological Survey, USGS 2011). Although estimates of global carbon dioxide emissions are probably accurate within 10 percent (as calculated from global average fuel chemistry and use), country estimates may have larger error |

| | | | | | |
|---|---|---|---|---|---|
| | | Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. | the burning of fossil fuels and cement manufacture, but excludes emissions from land use such as deforestation. The unit of measurement is kt (kiloton). Carbon dioxide emissions are often calculated and reported as elemental carbon. The were converted to actual carbon dioxide mass by multiplying them by 3.667 (the ratio of the mass of carbon to that of carbon dioxide). | climate change. It is also a major source of ocean acidification since it dissolves in water to form carbonic acid. The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and | bounds. Trends estimated from a consistent time series tend to be more accurate than individual values. Each year the CDIAC recalculates the entire time series since 1949, incorporating recent findings and corrections. Estimates exclude fuels supplied to ships and aircraft in international transport because of the difficulty of apportioning the fuels among benefiting countries. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working towards curbing CO2 emissions globally. | |
| CO2 emissions (metric tons per capita) | Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring. | Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate | Carbon dioxide emissions, largely by-products of energy production and use, account for the largest share of greenhouse gases, which are associated with global warming. Anthropogenic carbon dioxide emissions result primarily from fossil fuel combustion and cement manufacturing. In combustion different fossil fuels release different amounts of carbon dioxide for the same level of energy use: oil releases about 50 percent more carbon dioxide than natural gas, and coal releases about twice as much. Cement manufacturing releases about half a metric ton of carbon dioxide for each metric ton of cement produced. Data for carbon dioxide emissions include gases from | Carbon dioxide (CO2) is naturally occurring gas fixed by photosynthesis into organic matter. A byproduct of fossil fuel combustion and biomass burning, it is also emitted from land use changes and other industrial processes. It is the principal anthropogenic greenhouse gas that affects the Earth's radiative balance. It is the reference gas against which other greenhouse gases are measured, thus having a Global Warming Potential of 1. Burning of carbon-based fuels since the industrial revolution has rapidly increased concentrations of atmospheric carbon dioxide, increasing the rate of global warming and causing anthropogenic | The U.S. Department of Energy's Carbon Dioxide Information Analysis Center (CDIAC) calculates annual anthropogenic emissions from data on fossil fuel consumption (from the United Nations Statistics Division's World Energy Data Set) and world cement manufacturing (from the U.S. Department of Interior's Geological Survey, USGS 2011). Although estimates of global carbon dioxide emissions are probably accurate within 10 percent (as calculated from global average fuel chemistry and use), country estimates may have larger error |

| | | | | | |
|---|---|---|---|---|---|
| | | Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. See SP.POP.TOTL for the denominator's source. | the burning of fossil fuels and cement manufacture, but excludes emissions from land use such as deforestation. | climate change. It is also a major source of ocean acidification since it dissolves in water to form carbonic acid. The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. Emission intensity is the average emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and | bounds. Trends estimated from a consistent time series tend to be more accurate than individual values. Each year the CDIAC recalculates the entire time series since 1949, incorporating recent findings and corrections. Estimates exclude fuels supplied to ships and aircraft in international transport because of the difficulty of apportioning the fuels among benefiting countries. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working towards curbing CO2 emissions globally. | |
| Energy related methane emissions (% of total) | Methane emissions from energy processes are emissions from the production, handling, transmission, and combustion of fossil fuels and biofuels. | World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research | IPCC category 1 = Energy. Methane emissions result largely from agricultural activities, industrial production landfills and wastewater treatment, and other sources such as tropical forest and other vegetation fires. The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective contributions of different gases to be compared. A kilogram of methane is 21 times as effective at trapping heat in the earth's atmosphere as a kilogram of carbon dioxide within 100 years. The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective | The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - | National reporting to the United Nations Framework Convention on Climate Change that follows the Intergovernmental Panel on Climate Change guidelines is based on national emission inventories and covers all sources of anthropogenic carbon dioxide emissions as well as carbon sinks (such as forests). To estimate emissions, the countries that are Parties to the Climate Change Convention (UNFCCC) use complex, state-of-the-art methodologies recommended by the Intergovernmental Panel on Climate Change (IPCC). |

| | | | | | |
|---|---|---|---|---|---|
| | | (EDGAR): http://edgar.jrc.ec.europa.eu/. | contributions of different gases to be compared. | emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), Sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working towards curbing CO2 emissions globally. | |
| Energy use (kg of oil | Energy use refers to use of primary energy before | IEA Statistics © OECD/IEA 2014 | Total energy use refers to the use of primary energy before transformation to other end-use fuels (such as electricity | In developing economies growth in energy use is closely related to growth in the modern sectors - industry, | The IEA makes these estimates in consultation with national statistical offices, oil companies, |

| | | | | | |
|---|---|---|---|---|---|
| equivalent per capita) | transformation to other end-use fuels, which is equal to indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport. | (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/ | and refined petroleum products). It includes energy from combustible renewables and waste - solid biomass and animal products, gas and liquid from biomass, and industrial and municipal waste. Biomass is any plant matter used directly as fuel or converted into fuel, heat, or electricity. World Bank population estimates are used to calculate per capita data. Energy data are compiled by the International Energy Agency (IEA). IEA data for economies that are not members of the Organisation for Economic Co-operation and Development (OECD) are based on national energy data adjusted to conform to annual questionnaires completed by OECD member governments. Data for combustible renewables and waste are often based on small surveys or other incomplete information and thus give only a broad impression of developments and are not strictly comparable across countries. The IEA reports include country notes that explain some of these differences. All forms of energy - primary energy and primary electricity - are converted into oil equivalents. A notional thermal efficiency of 33 percent is assumed for converting nuclear electricity into oil equivalents and 100 percent efficiency for converting hydroelectric power. | motorized transport, and urban areas - but energy use also reflects climatic, geographic, and economic factors (such as the relative price of energy). Energy use has been growing rapidly in low- and middle-income economies, but high-income economies still use almost five times as much energy on a per capita basis. Governments in many countries are increasingly aware of the urgent need to make better use of the world's energy resources. Improved energy efficiency is often the most economic and readily available means of improving energy security and reducing greenhouse gas emissions. | electric utilities, and national energy experts. The IEA occasionally revises its time series to reflect political changes, and energy statistics undergo continual changes in coverage or methodology as more detailed energy accounts become available. Breaks in series are therefore unavoidable. |

| | | | | | |
|---|---|---|---|---|---|
| Fossil fuel energy consumption (% of total) | Fossil fuel comprises coal, oil, petroleum, and natural gas products. | IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/ | Energy data are compiled by the International Energy Agency (IEA). IEA data for economies that are not members of the Organisation for Economic Co-operation and Development (OECD) are based on national energy data adjusted to conform to annual questionnaires completed by OECD member governments. Data for combustible renewables and waste are often based on small surveys or other incomplete information and thus give only a broad impression of developments and are not strictly comparable across countries. The IEA reports include country notes that explain some of these differences. All forms of energy - primary energy and primary electricity - are converted into oil equivalents. A notional thermal efficiency of 33 percent is assumed for converting nuclear electricity into oil equivalents and 100 percent efficiency for converting hydroelectric power. | Fossil fuels are non-renewable resources because they take millions of years to form, and reserves are being depleted much faster than new ones are being made. In developing economies growth in energy use is closely related to growth in the modern sectors - industry, motorized transport, and urban areas - but energy use also reflects climatic, geographic, and economic factors (such as the relative price of energy). Energy use has been growing rapidly in low- and middle-income economies, but high-income economies still use almost five times as much energy on a per capita basis. Total energy use refers to the use of primary energy before transformation to other end-use fuels (such as electricity and refined petroleum products). It includes energy from combustible renewables and waste - solid biomass and animal products, gas and liquid from biomass, and industrial and municipal waste. Biomass is any plant matter used directly as fuel or converted into fuel, heat, or electricity. | The IEA makes these estimates in consultation with national statistical offices, oil companies, electric utilities, and national energy experts. The IEA occasionally revises its time series to reflect political changes, and energy statistics undergo continual changes in coverage or methodology as more detailed energy accounts become available. Breaks in series are therefore unavoidable. |
| GDP per capita (current US$) | GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added | World Bank national accounts data, and OECD National | - | - | - |

| | | | | | |
|---|---|---|---|---|---|
| | 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. | Accounts data files. | | | |
| GDP per capita growth (annual %) | Annual percentage growth rate of GDP per capita based on constant local currency. Aggregates are based on constant 2010 U.S. dollars. GDP per capita is gross domestic product divided by midyear population. GDP | World Bank national accounts data, and OECD National Accounts data files. | - | - | - |

| | | | | | |
|---|---|---|---|---|---|
| | 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. | | | | |
| Methane emissions (kt of CO2 equivalent) | Methane emissions are those stemming from human activities such as agriculture and from industrial methane production. | Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, | Methane emissions are those stemming from human activities such as agriculture and from industrial methane production. Expressed in CO2 equivalent using the GWP100 metric of the Second Assessment Report of IPCC and include CH4 (GWP100=21). The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective contributions of different gases to be compared. A kilogram of methane is 21 times as effective at | The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average emission rate of a given pollutant from | National reporting to the United Nations Framework Convention on Climate Change that follows the Intergovernmental Panel on Climate Change guidelines is based on national emission inventories and covers all sources of anthropogenic carbon dioxide emissions as well as carbon sinks (such as forests). To estimate emissions, the countries that are Parties to the Climate Change Convention (UNFCCC) use |

| | | | | | |
|---|---|---|---|---|---|
| | | Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. | trapping heat in the earth's atmosphere as a kilogram of carbon dioxide within 100 years. | a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), Sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working | complex, state-of-the-art methodologies recommended by the Intergovernmental Panel on Climate Change (IPCC). |

| | | | | | |
|---|---|---|---|---|---|
| | | | | towards curbing CO2 emissions globally. | |
| Nitrous oxide emissions (thousand metric tons of CO2 equivalent) | Nitrous oxide emissions are emissions from agricultural biomass burning, industrial activities, and livestock management. | Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. | Nitrous oxide emissions are mainly from fossil fuel combustion, fertilizers, rainforest fires, and animal waste. Nitrous oxide is a powerful greenhouse gas, with an estimated atmospheric lifetime of 114 years, compared with 12 years for methane. The per kilogram global warming potential of nitrous oxide is nearly 310 times that of carbon dioxide within 100 years. The emissions are usually expressed in carbon dioxide equivalents using the global warming potential, which allows the effective contributions of different gases to be compared. | The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the largest share of the greenhouse gases contributing to global warming and | National reporting to the United Nations Framework Convention on Climate Change that follows the Intergovernmental Panel on Climate Change guidelines is based on national emission inventories and covers all sources of anthropogenic carbon dioxide emissions as well as carbon sinks (such as forests). To estimate emissions, the countries that are Parties to the Climate Change Convention (UNFCCC) use complex, state-of-the-art methodologies recommended by the Intergovernmental Panel on Climate Change (IPCC). |

| | | | | | |
|---|---|---|---|---|---|
| | | | | climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), Sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working towards curbing CO2 emissions globally. | |
| Population growth (annual %) | Annual population growth rate for year t is the exponential rate of growth of midyear population from year t-1 to t, expressed as a percentage . Population is based on the de facto definition of population, which counts all residents regardless of | Derived from total population. Population source: (1) United Nations Population Division. World Population Prospects: 2019 Revision, (2) Census reports and other statistical publications from national statistical | Total population growth rates are calculated on the assumption that rate of growth is constant between two points in time. The growth rate is computed using the exponential growth formula: r = ln(pn/p0)/n, where r is the exponential rate of growth, ln() is the natural logarithm, pn is the end period population, p0 is the beginning period population, and n is the number of years in between. Note that this is not the geometric growth rate used to compute compound growth over discrete periods. For information on total population from which the growth rates are calculated, see total population (SP.POP.TOTL). | - | - |

| | | | | | |
|---|---|---|---|---|---|
| | legal status or citizenship. | offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Reprot (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme. | | | |
| Population, total | Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates. | (1) United Nations Population Division. World Population Prospects: 2019 Revision. (2) Census reports and other statistical publications from national statistical | Population estimates are usually based on national population censuses. Estimates for the years before and after the census are interpolations or extrapolations based on demographic models. Errors and undercounting occur even in high-income countries. In developing countries errors may be substantial because of limits in the transport, communications, and other resources required to conduct and analyze a full census. The quality and reliability of official demographic data are also affected by public trust in the | Increases in human population, whether as a result of immigration or more births than deaths, can impact natural resources and social infrastructure. This can place pressure on a country's sustainability. A significant growth in population will negatively impact the availability of land for agricultural production, and will aggravate demand for food, energy, water, social services, and infrastructure. On the other hand, decreasing population size - a result of fewer births than deaths, and people | Current population estimates for developing countries that lack (i) reliable recent census data, and (ii) pre- and post-census estimates for countries with census data, are provided by the United Nations Population Division and other agencies. The cohort component method - a standard method for estimating and projecting population - requires fertility, mortality, and net migration data, often collected from sample surveys, |

| | | | | | |
|---|---|---|---|---|---|
| | | offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Reprot (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme. | government, government commitment to full and accurate enumeration, confidentiality and protection against misuse of census data, and census agencies' independence from political influence. Moreover, comparability of population indicators is limited by differences in the concepts, definitions, collection procedures, and estimation methods used by national statistical agencies and other organizations that collect the data. The currentness of a census and the availability of complementary data from surveys or registration systems are objective ways to judge demographic data quality. Some European countries' registration systems offer complete information on population in the absence of a census. The United Nations Statistics Division monitors the completeness of vital registration systems. Some developing countries have made progress over the last 60 years, but others still have deficiencies in civil registration systems. International migration is the only other factor besides birth and death rates that directly determines a country's population growth. Estimating migration is difficult. At any time many people are located outside their home country as tourists, workers, or refugees or for other reasons. Standards for the duration and purpose of international moves that qualify as migration vary, and estimates require information on | moving out of a country - can impact a government's commitment to maintain services and infrastructure. | which can be small or limited in coverage. Population estimates are from demographic modeling and so are susceptible to biases and errors from shortcomings in both the model and the data. In the UN estimates the five-year age group is the cohort unit and five-year period data are used; therefore interpolations to obtain annual data or single age structure may not reflect actual events or age composition. Because future trends cannot be known with certainty, population projections have a wide range of uncertainty. |

| | | | | | |
|---|---|---|---|---|---|
| | | | flows into and out of countries that is difficult to collect. Population projections, starting from a base year are projected forward using assumptions of mortality, fertility, and migration by age and sex through 2050, based on the UN Population Division's World Population Prospects database medium variant. | | |
| Renewable electricity output (% of total electricity output) | Renewable electricity is the share of electricity generated by renewable power plants in total electricity generated by all types of plants. | IEA Statistics © OECD/IEA 2018 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/ | - | - | - |
| Renewable energy consumption (% of total final energy consumption ) | Renewable energy consumption is the share of renewables energy in total final energy consumption. | World Bank, Sustainable Energy for All (SE4ALL) database from the SE4ALL Global Tracking Framework led jointly by the World Bank, International Energy Agency, and the Energy | - | - | - |

| | | | | | |
|---|---|---|---|---|---|
| | | Sector Management Assistance Program. | | | |
| Total greenhouse gas emissions (kt of CO2 equivalent) | Total greenhouse gas emissions in kt of CO2 equivalent are composed of CO2 totals excluding short-cycle biomass burning (such as agricultural waste burning and savanna burning) but including other biomass burning (such as forest fires, post-burn decay, peat fires and decay of drained peatlands), all anthropogenic CH4 sources, N2O sources and F-gases (HFCs, PFCs and SF6). | Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: [https://www.climatewatchdata.org/ghg-emissions](https://www.climatewatchdata.org/ghg-emissions). | The GHG totals are expressed in CO2 equivalent using the GWP100 metric of the Second Assessment Report of IPCC and include CO2 (GWP100=1), CH4 (GWP100=21), N2O (GWP100=310) and F-gases (c-C4F8 GWP=8700, C2F6 GWP=9200, C3F8 GWP=7000, C4F10 GWP=7000, C5F12 GWP=7500, C6F14 GWP=7400, C7F16 GWP=7820, CF4 GWP=6500, HFC-125 GWP=2800, HFC-134a GWP=1300, HFC-143a GWP=3800, HFC-152a GWP=140, HFC-227ea GWP=2900, HFC-23 GWP=11700, HFC-236fa GWP=6300, HFC-245fa GWP=858, HFC-32 GWP=650, HFC-365mfc GWP=804, HFC-43-10-mee GWP=1300, SF6 GWP=23900). | The addition of man-made greenhouse gases to the Atmosphere disturbs the earth's radiative balance. This is leading to an increase in the earth's surface temperature and to related effects on climate, sea level rise and world agriculture. Emissions of CO2 are from burning oil, coal and gas for energy use, burning wood and waste materials, and from industrial processes such as cement production. Emission intensity is the average emission rate of a given pollutant from a given source relative to the intensity of a specific activity. Emission intensities are also used to compare the environmental impact of different fuels or activities. The related terms - emission factor and carbon intensity - are often used interchangeably. The carbon dioxide emissions of a country are only an indicator of one greenhouse gas. For a more complete idea of how a country influences climate change, gases such as methane and nitrous oxide should be taken into account. This is particularly important in agricultural economies. The environmental effects of carbon dioxide are of significant interest. Carbon dioxide (CO2) makes up the | National reporting to the United Nations Framework Convention on Climate Change that follows the Intergovernmental Panel on Climate Change guidelines is based on national emission inventories and covers all sources of anthropogenic carbon dioxide emissions as well as carbon sinks (such as forests). To estimate emissions, the countries that are Parties to the Climate Change Convention (UNFCCC) use complex, state-of-the-art methodologies recommended by the Intergovernmental Panel on Climate Change (IPCC). |

| | | | | | |
|---|---|---|---|---|---|
| | | | | largest share of the greenhouse gases contributing to global warming and climate change. Converting all other greenhouse gases (methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), Sulphur hexafluoride (SF6)) to carbon dioxide (or CO2) equivalents makes it possible to compare them and to determine their individual and total contributions to global warming. The Kyoto Protocol, an environmental agreement adopted in 1997 by many of the parties to the United Nations Framework Convention on Climate Change (UNFCCC), is working towards curbing CO2 emissions globally. | |