Coursework

Applied Statistics & Data Visualisation

MSc. Data Science

Exploring the Relationship Between Greenhouse Gas Emissions and Economic Indicators in Europe and Asia

Yash Panchal

@00664404

16th December 2022

TABLE OF CONTENT

Interactive Dashboard Design

Introduction

Greenhouse gases, such as carbon dioxide and methane, are major contributors to global warming and climate change. In recent years, there has been a growing need for data visualization and interactive dashboard design to help policymakers and the public understand the trends and patterns in greenhouse gas emissions. This report presents an analysis of greenhouse gas emissions in Germany, the United Kingdom, Ireland, Norway, the Netherlands, France, Bangladesh, China, India, South Korea, the Philippines, and Thailand from 1997 to 2018.

The data for this analysis were obtained from the World Bank's World Development Indicators database. The data include annual greenhouse gas emissions in metric tons of CO2 equivalent for each country. The emissions data were visualized using line charts and interactive dashboards, which allow users to explore the data and compare emissions trends across countries.

The results of the analysis show that greenhouse gas emissions have increased over time in most of the countries studied. Germany, and the United Kingdom have seen the largest decrease in emissions. Among the developing countries studied, China and India have had the largest increases in emissions, while Bangladesh and the Philippines have had the smallest increases.

One of the key findings of this analysis is that greenhouse gas emissions are highly correlated with economic growth. As countries experience economic growth, their greenhouse gas emissions tend to increase. This suggests that efforts to reduce emissions must be accompanied by policies that promote sustainable economic growth.

In conclusion, data visualization and interactive dashboard design are powerful tools for understanding and communicating trends in greenhouse gas emissions. The analysis presented in this report provides valuable insights into the trends and patterns in emissions for Germany, the United Kingdom, Ireland, Norway, the Netherlands, France, Bangladesh, China, India, South Korea, the Philippines, and Thailand from 1997 to 2018.

Background Research

The current state of the art in the design of interactive dashboards has seen significant advancements in recent years, with a focus on user-centered design and the integration of various technologies. This report aims to explore these developments and provide an overview of current perspectives on the composition, layout, and design principles of single-screen dashboards.

One key aspect of modern dashboard design is the incorporation of user-centered design principles. This involves understanding the needs and goals of the user and designing the dashboard to meet those needs most effectively. This can be achieved through a variety of methods, such as user research, usability testing, and user feedback. By considering the user's perspective, designers can create dashboards that are intuitive, easy to use and provide the necessary information clearly and concisely.

Another important aspect of dashboard design is the layout and composition of the information on the screen. A well-designed dashboard should provide a clear visual hierarchy, with the most important information presented prominently and other information organized in a logical and easy-to-follow manner. This can be achieved with visual elements such as colours, fonts, and imagery to draw the user's attention to important information and guide them through the dashboard.

In addition to user-centered design and effective layout, several design principles are commonly used in the development of interactive dashboards. One of these principles is the use of clear and concise labelling and data visualizations. This helps to ensure that the information on the dashboard is easy to understand and interpret, allowing users to quickly make decisions and act based on the data. Other principles include the use of minimalism, which helps to prevent information overload, and the use of animation and interactivity to engage the user and provide a more dynamic and engaging experience.

Several current works provide insight into the design and development of interactive dashboards. One such work is "Designing Dashboards: A Guide for Business and Technology" by Stephen Few (2012), which provides a comprehensive overview of the principles and best practices of dashboard design. Another valuable resource is "Information Dashboard Design: The Effective Visual Communication of Data" by Stephen Few (2006), which offers practical advice on creating effective and engaging dashboards.

In conclusion, the current state of the art in the design of interactive dashboards involves a focus on user-centered design and the integration of various technologies. Effective design involves considering the needs and goals of the user, creating a clear visual hierarchy, and utilizing design principles such as clear labelling and data visualizations. By following these principles, designers can create dashboards that are intuitive, easy to use and provide the necessary information clearly and concisely.

Exploration of data

Data visualisation is a powerful tool for data exploration, as it can help quickly and easily identify patterns, trends, and relationships in the data. By creating visualizations of the data, a better understanding can be gained of its characteristics and uncover insights that may not be immediately apparent from looking at the raw data.

There are many different types of visualizations that can be used for data exploration, depending on the type and characteristics of the data. Some common types of visualizations include histograms, scatter plots, line graphs, bar charts, and pie charts.

To create a visualization, a suitable type of chart or graph needs to be chosen, and a software tool or library to generate the visualization used. Power BI has been chosen for Data visualisation and implementation of an Interactive Dashboard for this assignment.

This chance has been used for exploring the relationship between the indicators, intended to be used in this assignment. First, the relationship between Greenhouse Gas Emissions & GDP will be explored.

A scatter plot is a useful tool for understanding the relationship between two variables. By plotting the values of the two variables on a graph, a scatter plot can help identify patterns and trends in the data. If the variables are strongly correlated, the points on the scatter plot will tend to fall along a line or curve. This can help identify a linear or nonlinear relationship between the variables. Scatter plots can also be used to identify outlier points, which can be useful for identifying potential errors in the data. Overall, scatter plots are a valuable tool for understanding the relationship between variables and can aid in the analysis of data.

For this analysis, the development status of the countries in the dataset needs to be determined. The development status of a country is typically determined by a variety of factors, including economic indicators such as gross domestic product (GDP), unemployment rates, and inflation; social indicators such as literacy rates, life expectancy, and access to education and healthcare; and environmental indicators such as access to clean water and air quality. In addition, a country's level of political stability, rule of law, and level of corruption may also be considered when determining its development status. The United Nations Development Programme (UNDP) publishes annual reports that rank countries based on their development status, using a composite index known as the Human Development Index (HDI). The HDI combines a variety of factors to provide a comprehensive measure of a country's development status. But the development status of a country can also be crudely determined by their GDP per capita. There is no set level that defines whether an economy is considered developed or developing. Some economists believe that a per capita GDP of $12,000 to $15,000 is sufficient for a country to be considered developed, while others believe that a per capita GDP above $25,000 or $30,000 is necessary for a country to be considered developed. The specific criteria for determining a country's development status can vary among economists (Majaski C, 2022). As such $25,000 has been considered as the GDP per Capita amount for this assignment above which a country is considered developed, above which the country is considered developed. And a new conditional column is made in the Power BI table which puts labels on the country based on the GDP per capita for the year.

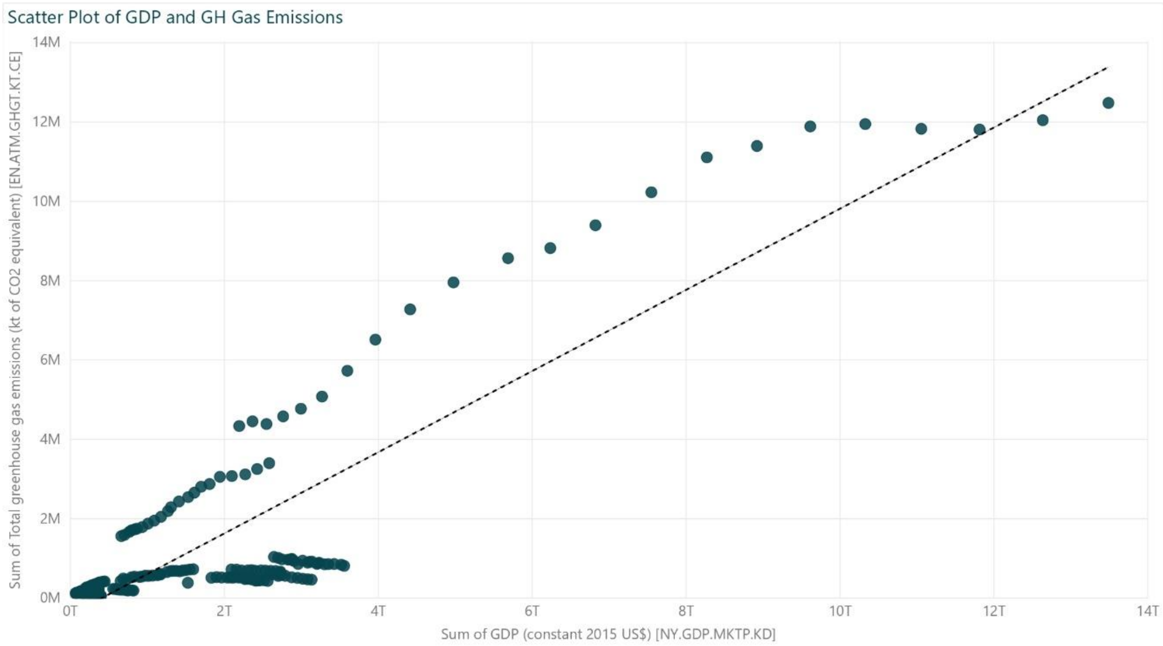

Figure 1

Overall, the scatter plot (Figure 1) shows a positive correlation between GDP and Greenhouse Emissions, but I suspect this relation is greatly influenced by India and China. Hence, I further plotted separate scatter plots for developed and developing countries to deeper analyse the relation.

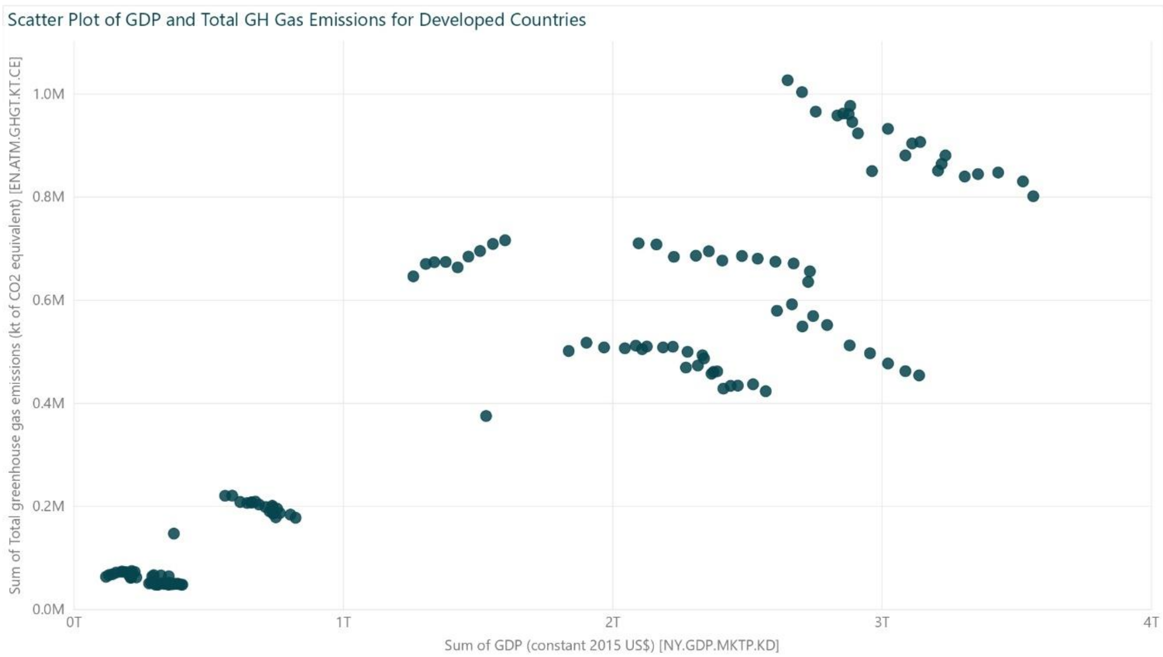

Figure 2

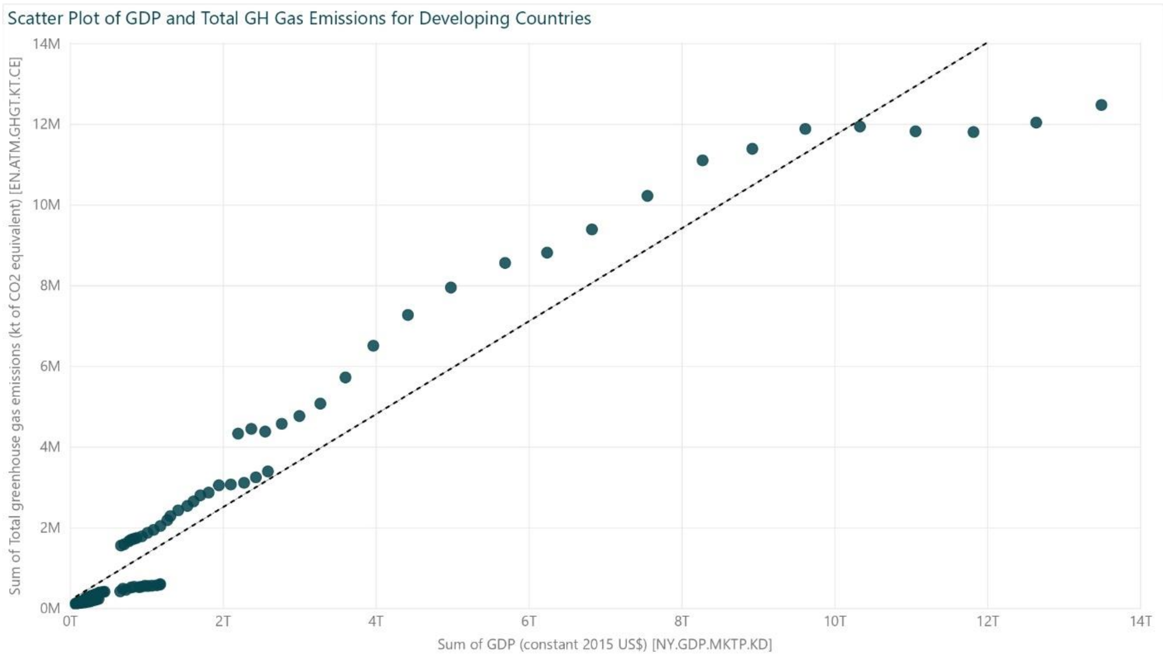

Figure 3

The plot for developing countries (Figure 3) shows a clear positive correlation between the indicators. But the plot for the developed countries in figure 4 makes the picture a lot muddier, as such I dived deeper and plotted the graphs for all the developed countries in the dataset individually.

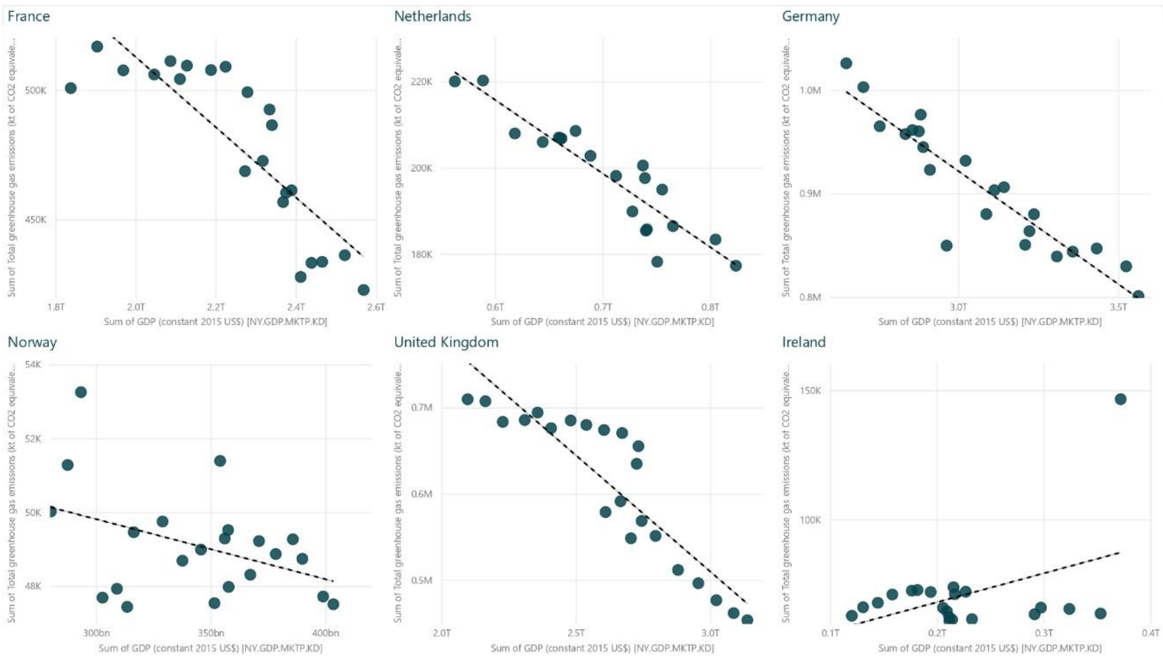

Figure 4

And these plots give us a much clearer picture that all the developed countries have a negative correlation for these indicators. This tells us that the correlation between GDP and Greenhouse Gas Emissions for a country is greatly dependent on the development status of the country in question.

Next, the relationship between Greenhouse Gas Emissions and Population has been explored.

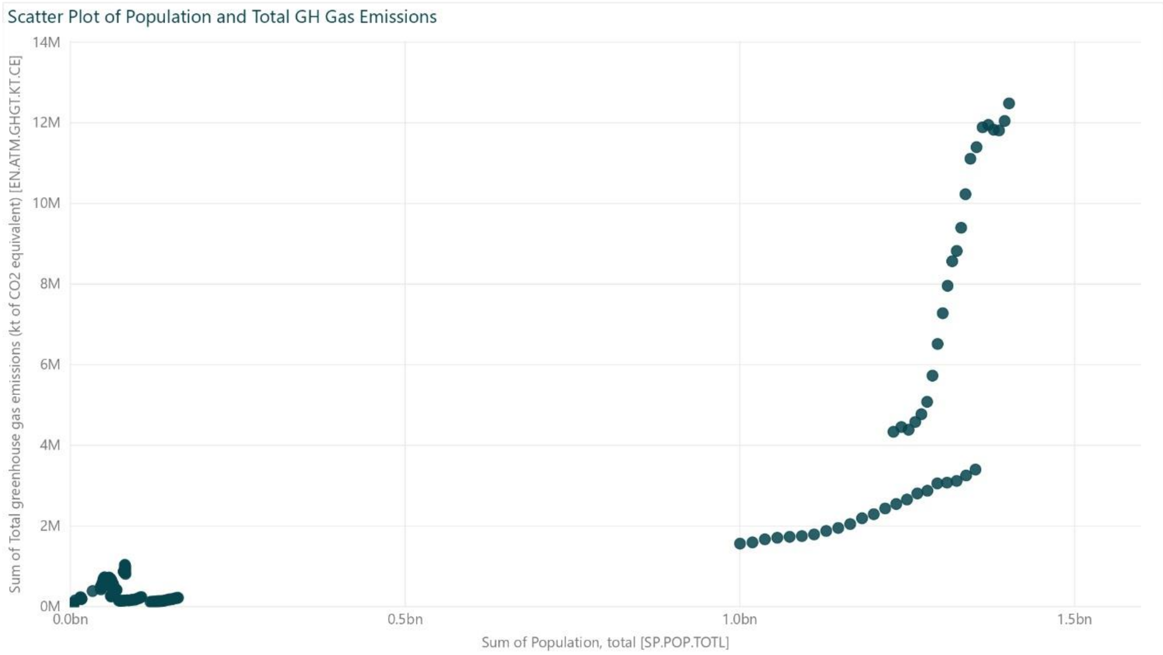

Figure 5

This plot in (Figure 5) shows a positive correlation between Population and Greenhouse Gas Emissions, but it also shows 3 separate clusters. These clusters are identified as China, India, and the rest of the countries in the last cluster. I further investigated the same plots for developed and developing countries but without India and China.

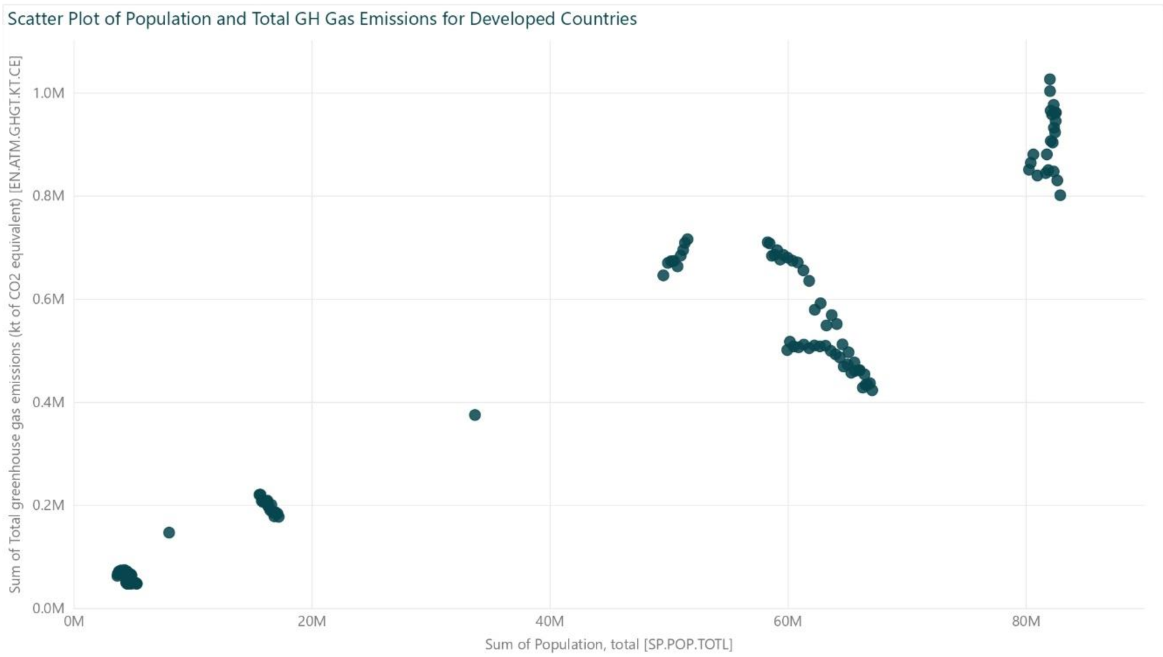

Figure 6

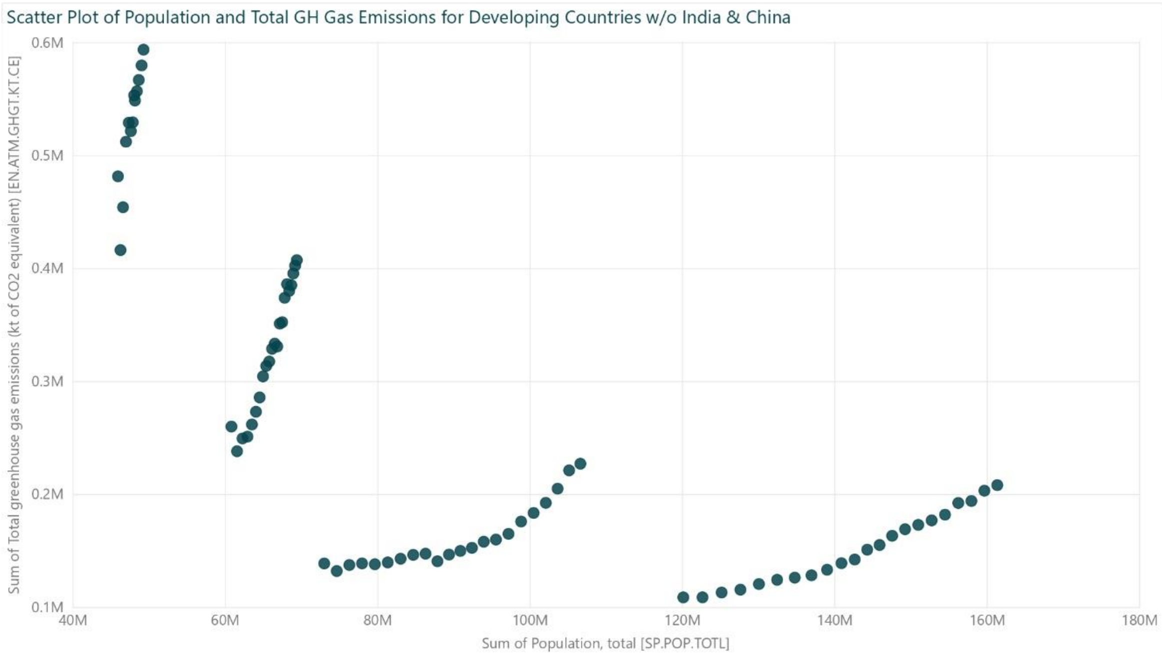

Figure 7

The plot for both developed and developing countries in (Figure 6 & 7) again shows clear clusters for different countries, individually all the developed countries show a negative correlation, whereas the developing countries show a positive correlation. The plots for India and China show a positive correlation as well.

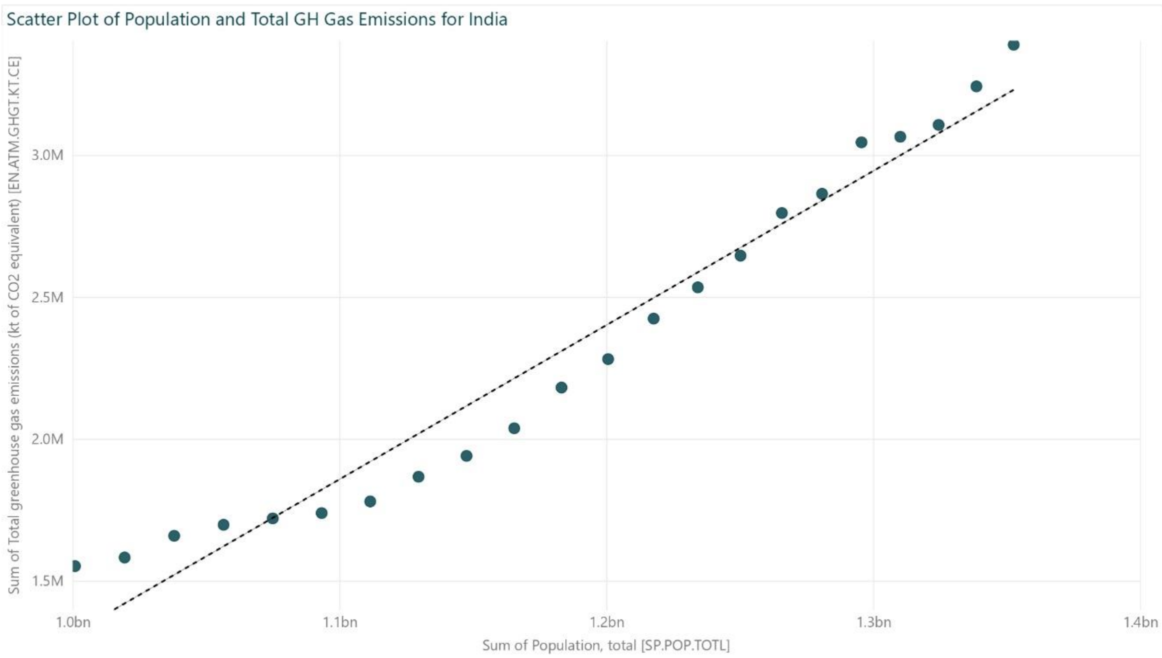

Figure 8

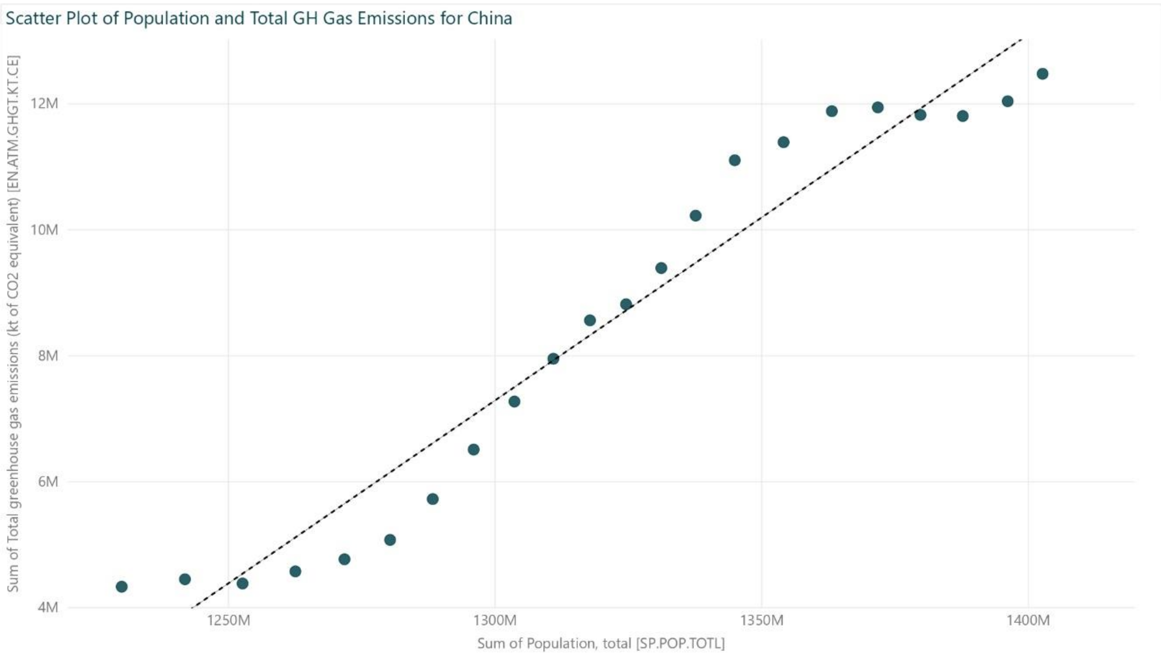

Figure 9

Again, just as with GDP and Greenhouse Gas Emissions, Populations and Greenhouse Gas Emissions show a clear relationship between these indicators, the nature of this relationship is highly dependent on the development status of the country in question.

Investigation of Data Workflows & Proposal for Design of Dashboard

As stated above the dataset contains data on various economic indicators for Germany, the United Kingdom, Ireland, Norway, the Netherlands, France, Bangladesh, China, India, South Korea, the Philippines, and Thailand from 1997 to 2018. These indicators include CO2 emissions, GDP, GNI, imports of goods and services, nitrous oxide emissions, total greenhouse gas emissions, and population. The data is taken from the World Bank's World Development Indicators database.

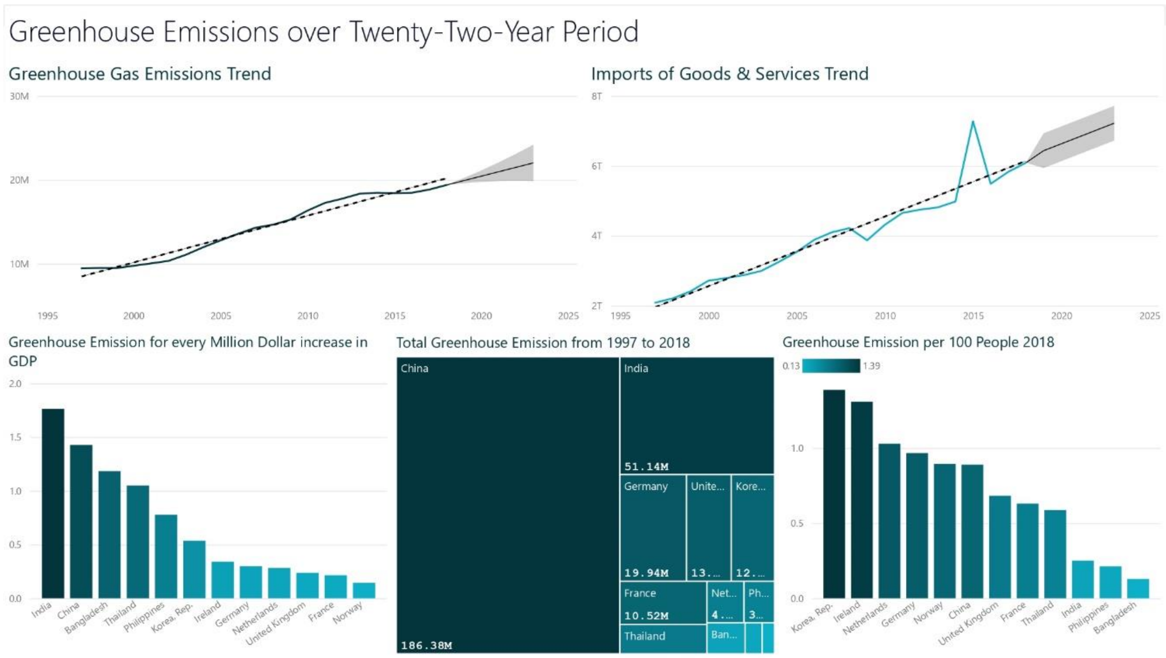

Figure 10

First, a treemap of greenhouse gas emissions for the countries was created in Power BI. A treemap of greenhouse gas emissions for the countries listed above shows the relative sizes of each country's emissions over the time in question. China is the largest country on the treemap, with 59% of total emissions, followed by India with 16% of emissions. Germany, the United Kingdom, and Korea are all represented by similar-sized rectangles, each with around 4% of emissions. France, Thailand, the Netherlands, the Philippines, and Bangladesh are all represented by smaller rectangles, each with around 3% of emissions or less. Ireland and Norway are the smallest countries on the treemap, with close 0% of total emissions.

However, a treemap of greenhouse gas emissions of countries does not give the full picture for a few reasons. First, a treemap only shows the relative sizes of the different countries' emissions but does not provide any information about the absolute levels of emissions. This means that it may be difficult to compare the emissions of different countries and understand their impact on the environment. Finally, a treemap does not consider a country's population or the size of its economy, which can also be important factors in understanding its overall emissions as we found earlier that these factors have a definite relation. But it is still important to know the relative size of a country’s emissions.

Next, a line graph of total greenhouse gas emissions for all selected countries over the specified period showing the trend in emissions over time was built. The x-axis of the graph represents the years from 1997 to 2018, and the y-axis represents the total emissions in millions of metric tons of carbon dioxide equivalent. The line on the graph starts at the point representing the emissions in 1997, which were 9434920.01 million metric tons of carbon dioxide equivalent, and then rises over the next few years before reaching a peak in 2010 when emissions were 16330629.72 million metric tons of carbon dioxide equivalent. After 2010, the line on the graph begins to decline, reaching a low point in 2016, when emissions were 18455160.18 million metric tons of carbon dioxide equivalent. From 2016 onwards, the line on the graph rises again, reaching 19383179.77 million metric tons of carbon dioxide equivalent in 2018. This line graph shows the overall trend in emissions over the period and highlights any changes or fluctuations in emissions.

Next, another line graph of imports of goods and services for all selected countries over the specified period showing the trend in imports over time was built. The x-axis of the graph represents the years from 1997 to 2018, and the y-axis represents the imports in millions of US dollars. The line on the graph starts at the point representing the imports in 1997, which were 2.08441 trillion US dollars, and then rises over the next few years before reaching a peak in 2015 when imports were 7.27447 trillion US dollars. After 2015, the line on the graph begins to decline, reaching a low point in 2016, when imports were 5.48685 trillion US dollars. From 2016 onwards, the line on the graph rises again, reaching 6.0904 trillion US dollars in 2018. This line graph shows the overall trend in imports over the period and highlights any changes or fluctuations in imports.

Next, some measures (calculated columns), we created. The first one is the factor of greenhouse gas (GH) emissions to gross domestic product (GDP) it is a measure of the relationship between a country's GH emissions and its GDP. It shows the increase in greenhouse gas emissions for every million-dollar increase in GDP. It is calculated by dividing the total GH emissions of a country by its GDP and multiplying the result by 1000000. This calculation can provide insight into the environmental impact of a country's economic activities and can be used to compare the GH emissions and GDP of different countries. A lower score on this measure indicates better performance by the country.

Next, the measure, greenhouse gas (GH) emissions per 100 people, was created, it is a measure of the average GH emissions per 100 people in a country. It is calculated by dividing the total GH emissions of a country by its population and multiplying the result by 100. This calculation can provide insight into the environmental impact of a country's population and can be used to compare the GH emissions and population of different countries. A lower score on this measure indicates better performance.

Next, the graphs for the calculated fields, were plotted, a column chart was chosen for both these measures. Column charts were selected to visualise these measures as they are especially useful for showing comparisons between different factors, in our case countries, and they are easily understood by most people.

The plot shows the average factor of greenhouse gas (GH) emissions to gross domestic product (GDP) for a variety of countries. The plot shows that, on average, Bangladesh, China, India, Korea, and the Philippines have higher GH emissions per unit of GDP compared to other countries such as France, Germany, Ireland, the Netherlands, and the United Kingdom. However, the table only provides a snapshot of each country's GH emissions and GDP, i.e., an average over 22 years, and does not account for changes over time or other factors that may affect a country's GH emissions.

The plot shows the average greenhouse gas (GH) emissions per 100 people for the countries. The plot shows that, on average, Bangladesh, India, and the Philippines have lower GH emissions per person compared to other countries such as China, France, Germany, Ireland, the Netherlands, and the United Kingdom. However, the table only provides a snapshot of each country's GH emissions and population, i.e., an average over 22 years, and does not account for changes over time or other factors that may affect a country's GH emissions.

Once all the plots are made, the dashboard can be interacted with, by selecting one of the countries, from either the treemap or from one of the column charts, to get more insight into greenhouse gas emission and import of goods and services trends for the country.

When selecting one of the various countries from the treemap or one of the column charts, it is observed that China & India and other developing countries have a substantial increase in greenhouse gas emissions throughout the period, whereas the developed countries have in some cases massively reduced their greenhouse gas emissions or have managed to plateau it. At the same time, it is observed that there has been a huge increase in the import of goods and services for all countries other than China, given that China does not publish these figures we can comment on this.

While looking through the plots for different countries, it is further observed that the European countries have done a really respectable job at reducing their greenhouse gas emissions. The European Union has made noteworthy progress in reducing greenhouse gas emissions in recent years. According to the European Environment Agency, the EU has reduced its emissions by 23% between 1990 and 2017 and is on track to meet its target of reducing emissions by 40% by 2030 (European Environment Agency, 2018).

One of the key policies driving this progress is the EU Emissions Trading System (ETS), which sets a cap on the amount of carbon dioxide that can be emitted by industries in the EU. This cap is gradually reduced over time, creating a financial incentive for industries to invest in cleaner technologies and reduce their emissions (European Commission, 2018). Another important policy is the Renewable Energy Directive, which requires EU countries to derive a certain percentage of their energy from renewable sources. This has led to significant investments in wind and solar power and has helped to reduce the EU's reliance on fossil fuels (European Commission, 2018). In addition to these policies, many European countries have implemented their initiatives to reduce emissions. For example, Germany has invested heavily in renewable energy and has introduced measures to promote energy efficiency (Federal Ministry for Economic Affairs and Energy, 2018). France has also introduced a carbon tax, which has encouraged industries to reduce their emissions to avoid the tax (Ministère de la Transition Écologique et Solidaire, 2018). Overall, the efforts of the EU and its member countries have been successful in reducing greenhouse gas emissions. However, there is still more work to be done to meet the EU's ambitious climate targets and avoid the worst impacts of climate change.

But, at the same time looking at the trend for the imports of goods and services could point to a case of carbon leakage, which would mean the reduction in greenhouse gas emissions observed are not entirely real. Carbon leakage refers to the practice of developed countries offloading their greenhouse gas emissions to developing countries with weaker environmental regulations. This can occur when industries in developed countries relocate their polluting operations to developing countries to avoid stricter regulations and reduce their emissions.

According to a study by the European Commission, carbon leakage is a significant problem in the European Union. The study found that, without adequate policies in place to prevent carbon leakage, the EU's efforts to reduce emissions could be undermined, and global emissions could increase (European Commission, 2013).

One of the key policies implemented by the EU to prevent carbon leakage is the EU Emissions Trading System (ETS), which sets a cap on the amount of carbon dioxide that can be emitted by industries in the EU. This cap is gradually reduced over time, creating a financial incentive for industries to reduce their emissions and invest in cleaner technologies.

However, the effectiveness of the ETS in preventing carbon leakage has been criticized, and some have argued that more robust policies are needed to effectively address this problem (Stavins, 2013).

Overall, carbon leakage is a significant challenge for the European Union and its efforts to reduce greenhouse gas emissions. Further efforts are needed to address this problem and ensure that the EU's climate targets are not undermined by the offshoring of polluting industries.

Discussion

The dashboard provided before outlines a series of steps for analysing a dataset containing economic indicators for several countries from 1997 to 2018. The report proposes building a treemap of greenhouse gas emissions for the countries in the dataset and creating line graphs of total greenhouse gas emissions and imports of goods and services over time. It also suggests calculating measures such as the factor of greenhouse gas emissions to gross domestic product and the factor of greenhouse gas emissions to population.

One potential weakness of the proposed approach is that it relies heavily on visualizations, such as the treemap and line graphs, to convey information about the data. While these visualizations can be useful for identifying trends and patterns in the data, they do not provide a comprehensive understanding of the underlying economic indicators and their relationship to each other. For example, the treemap only shows the relative sizes of each country's emissions but does not provide any information about the absolute levels of emissions or how they compare to other countries. The line graphs also only show the overall trend in emissions and imports over time, but do not provide detailed information about how these indicators are changing on a year-to-year basis.

Furthermore, the dashboard does not mention any potential limitations of the dataset or the potential biases that may be present in the data. Economic indicators can be complex and difficult to measure, and it is important to consider potential sources of error or bias when analysing this type of data. For example, the data may only include certain types of economic indicators or may be based on estimates rather than actual measurements. These limitations could affect the accuracy and reliability of the analysis.

Overall, the dashboard provides a basic plan for analysing the dataset, but it would benefit from a more comprehensive approach that considers the broader context and limitations of the data. A more thorough analysis would include additional information about the economic indicators and how they are related. It would also consider potential sources of bias or error in the data and consider how these factors could affect the reliability of the findings.

Conclusion

In conclusion, the analysis presented in this report provides valuable insights into the trends and patterns in greenhouse gas emissions for Germany, the United Kingdom, Ireland, Norway, the Netherlands, France, Bangladesh, China, India, South Korea, the Philippines, and Thailand from 1997 to 2018. The results show that emissions have increased over time in most of the countries studied, with the largest decrease in emissions seen in Germany, the United Kingdom, and Ireland. Among the developing countries studied, China and India have had the largest increases in emissions, while Bangladesh and the Philippines have had the smallest increases. One of the key findings of the analysis is the strong correlation between greenhouse gas emissions and economic growth. As countries experience economic growth, their emissions tend to increase. This suggests that efforts to reduce emissions must be accompanied by policies that promote sustainable economic growth. Data visualization and interactive dashboard design are powerful tools for understanding and communicating trends in greenhouse gas emissions. The use of user-centred design principles, effective layout, and design principles such as clear labelling and data visualizations can help to create dashboards that are intuitive, easy to use, and provide valuable insights into the data. By utilizing these tools and techniques, policymakers and the public can better understand the trends and patterns in emissions and take action to address the challenges of climate change.

Statistical Analysis

Introduction

Greenhouse gases, such as carbon dioxide and methane, are major contributors to climate change. The amount of greenhouse gas emissions produced by a country is related to its level of economic development and population size. This assignment contains analysis of the greenhouse gas emissions of twelve countries, 6 European and 6 Asian, and exploring the relationship between their emissions and gross domestic product (GDP) and population size.

To begin with, we will collect data on the greenhouse gas emissions, GDP, and population size of the 12 countries in question. We will then use statistical techniques to analyse this data and identify any trends or patterns that exist. For example, we may find that there is a positive correlation between a country's GDP and its greenhouse gas emissions, meaning that as a country's economy grows, so too do its emissions.

Once we have identified any trends or patterns in the data, we will then use this information to explore the relationship between greenhouse gas emissions and GDP and population size. We may find, for example, that countries with larger populations tend to have higher levels of emissions, while those with smaller populations have lower emissions.

Overall, this assignment will provide us with a deeper understanding of the relationship between greenhouse gas emissions and economic and population factors. By better understanding the factors that contribute to emissions, we can work towards developing strategies for reducing them and mitigating the impact of climate change.

Background Research

Statistical analysis is an essential tool for understanding and interpreting data in many fields, including the study of greenhouse gas emissions. This assignment will explore the use of several statistical methods for analysing the relationship between greenhouse gas emissions, gross domestic product (GDP), and population in six European and six Asian countries.

First, the theoretical background of some of the key statistical methods that will be used in this analysis will be discussed. This will include a discussion of the mean, median, and standard deviation, as well as more advanced techniques such as skewness, kurtosis, and correlation analysis. Regression analysis, including both simple and multiple regression, and time series analysis and forecasting using the Holt-Winters model will be discussed. Additionally, the Mann-Kendall hypothesis testing and the Willcox test, which are commonly used in the analysis of time series data will be discussed.

Next, the data will be presented, and the research design used in this assignment will be described. This will include a description of the six European and six Asian countries that were selected for the analysis, as well as the specific data sources and time periods used for the analysis.

Once the data and research design have been presented, the assignment will proceed to the statistical analysis itself. This will involve calculating summary statistics for the greenhouse gas emissions, GDP, and population data for each of the twelve countries. These summary statistics will include the mean, median, and standard deviation, as well as measures of skewness and kurtosis.

Next, a correlation analysis to examine the relationship between greenhouse gas emissions, GDP, and population will be conducted. This will involve calculating the Pearson correlation coefficient for each pair of variables, as well as conducting a hypothesis test to determine whether the correlation is statistically significant.

After the correlation analysis, the assignment will move on to regression analysis. This will involve fitting a simple linear regression model to the data to examine the relationship between greenhouse gas emissions and GDP or population. We will also conduct a multiple regression analysis to examine the relationship between greenhouse gas emissions, GDP, and population simultaneously.

Finally, time series analysis and forecasting using the Holt-Winters model will be conducted. This will involve fitting the model to the data and using it to generate forecasts of greenhouse gas emissions for the six European and six Asian countries. We will also use the Mann-Kendall hypothesis testing and the Willcox test to assess the statistical significance of any trends or shifts in the data.

In conclusion, this assignment has presented an overview of the use of several statistical analysis methods for examining the relationship between greenhouse gas emissions, GDP, and population in six European and six Asian countries. The results of the analysis suggest that there are strong correlations between these variables, and that regression and time series analysis can be useful tools for understanding and forecasting greenhouse gas emissions in these countries. Further research is needed to confirm and expand upon these findings.

Exploration of Dataset

The data used in this study was obtained from the World Bank's World Development Indicators Data Bank (World Bank, 2022). It contains information on total greenhouse gas emissions in CO2 equivalent kilo-tonnes, gross domestic product adjusted for inflation in 2015 USD, and population for a period of 22 years from 1997 to 2018 for 12 countries. These countries include six Asian nations (Bangladesh, China, India, the Republic of Korea, the Philippines, and Thailand) and six European countries (France, Germany, Ireland, Netherlands, Norway, and the United Kingdom). The selection of countries provides a balanced representation of both regions and levels of development.

Column	Details
Year	1997 - 2018
Country	Bangladesh, China, France, Germany, India, Ireland, Republic of Korea, Netherlands, Norway, Philippines, Thailand & United Kingdom
GDP (constant 2015 US$)	Gross Domestic Product of the corresponding Country and Year adjusted for inflation to 2015 USD for better comparability
Total greenhouse gas emissions (kt of CO2 equivalent)	Total Greenhouse Gas Emissions for the corresponding Country and Year represented in CO2 equivalent kilo-tonne unit
Population	The population of the corresponding Country and Year

The Greenhouse Gas Emissions, Population and GDP of China and India show up as outliers in the data. And as they are an important aspect of the datasets, they cannot be removed. One approach for dealing with this type of dataset would be to use robust statistical methods that are less sensitive to outliers, such as the median instead of the mean. Another approach is to transform the data using a technique such as a log transformation to reduce the impact of the outliers. Additionally, it can be helpful to include a description of the outliers in the analysis and to consider the potential reasons for their existence. In some cases, it may also be appropriate to create separate models or analyses for the data with and without the outliers. This assignment uses separate models for different countries to deal with the outliers.

Analysis

The script submitted alongside the report performs a statistical analysis of a dataset containing information on various indicators for different countries and years. The first step in the analysis is to read in the data from an Excel file and check the structure of the data. We then convert the "Year" and "Country" columns to factors and rename some of the columns for clarity.

After checking for missing values, we calculate the correlations among the columns using the cor function. The resulting correlations are then plotted using the corrplot function. This helps us to identify which variables are related to each other and the strength of those relationships.

	GDP	GHGE	Population
GDP	1	0.876665	0.57082
GHGE	0.876665	1	0.81385
Population	0.57082	0.81385	1

The table presents the correlation between three variables: GDP, GHGE, and population. The values in the table indicate a strong positive correlation between GDP and GHGE, as well as between GHGE and population. This suggests that as one variable increases, the other tends to increase as well. The data in the table can be used to further study the relationship between these variables and their potential impact on each other.

Next, a new data frame is created with the mean and median values for each column. This provides an overview of the central tendency of the data and helps to identify any potential outliers or skewed distributions.

This table presents information about the mean and median values of three indicators: GDP, GHGE, and population. The mean and median values are provided for each indicator. For example, the mean GDP value is 1587896310186, and the median GDP value is 753179320810. This information can be used to understand the average and typical values of these indicators and to compare them with each other.

Two functions are then defined to calculate the skewness and kurtosis of a vector, which can provide further insight into the shape of the data. The dplyr package is then used to group the data by country and calculate a variety of summary statistics for each group. This helps to understand the variation in the data across different countries.

This table provides a wealth of information about the GDP, GHGE, and population of several countries. The table includes statistics such as the mean, median, standard deviation, variance, skewness, and kurtosis for each variable for each country. This information can be used to compare the economic and environmental conditions of different countries and to understand how these variables vary within each country. For example, the table shows that India has a higher mean GHGE value than Germany, but also a higher mean population value. This information can be used to study the relationship between economic growth, environmental impact, and population size.

After exploring the overall dataset, we create scatter plots to visualize the relationships between different variables. We create a scatter plot to visualize the relationship between greenhouse gas emissions (GHGE) and year, and another scatter plot to visualize the relationship between GHGE and gross domestic product (GDP). These plots help us to identify trends or patterns in the data.

```
Call:
lm(formula = df$GHGE ~ df$Population)

Coefficients:
  (Intercept)  df$Population
  57119.50964        0.00442


Call:
lm(formula = df$GHGE ~ df$Population)

Residuals:
     Min       1Q   Median       3Q      Max
-3190963  -278243     7980   257652  6213416

Coefficients:
                   Estimate    Std. Error t value            Pr(>|t|)
(Intercept)    57119.509637 100917.039860   0.566               0.572
df$Population      0.004420      0.000195  22.671 <0.0000000000000002
***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1423000 on 262 degrees of freedom
Multiple R-squared:  0.6624, Adjusted R-squared:  0.6611
F-statistic:   514 on 1 and 262 DF,  p-value: < 0.00000000000000022
```

Then the linear regression models are fitted to investigate the relationships between GHGE and GDP, as well as GHGE and population. Linear regression models are useful for identifying the extent to which one variable can be predicted from another. We then print the fitted models and their summaries, which provide information on the strength of the relationship and the statistical significance of the model. We also add the fitted regression lines to the scatter plots to help visualize the relationship between the variables.

Then the multiple linear regression model is fitted to investigate the relationship between GHGE and GDP and population and create diagnostic plots for the model using the avPlots function.

The assignment then filters the data to include only observations for four specific countries - India, China, Germany, and the United Kingdom - and fit multiple linear regression models for each of these countries. The fitted models and their summaries are printed, and diagnostic plots are created for each model.

This output from the linear regression analysis examines the relationship between GHGE and GDP. The model estimates the coefficients for an equation of the form GHGE = intercept + GDP * coefficient. The output shows that the estimated coefficient for GDP is 0.000001023, with a standard error of 0.00000003468. This suggests that there is a positive relationship between GHGE and GDP, such that as GDP increases, GHGE tends to increase as well. The output also shows that the model has a high degree of fit, with an R-squared value of 0.7685 and an adjusted R-squared value of 0.7677. This indicates that the model explains a significant portion of the variation in GHGE.

The regression analysis of GHGE on population found a significant positive relationship between the two variables, with a p-value less than 0.0000000000000002. The model explained 66% of the variance in GHGE and had a low residual standard error. The coefficient for the population was 0.004420, indicating a small but statistically significant increase in GHGE with each unit increase in population.

The regression analysis of GHGE on population and GDP found significant positive relationships between all three variables, with p-values less than 0.0000000000000002. The model explained 91% of the variance in GHGE and had a low residual standard error. The coefficient for the population was 0.0025251133, indicating a small but statistically significant increase in GHGE with each unit increase in population. The coefficient for GDP was 0.00000071327, indicating a very small but statistically significant increase in GHGE with each unit increase in GDP.

Multiple regression analyses were performed to investigate the relationship between GHGE, population, and GDP for four countries: India, China, Germany, and the UK. In India and the UK, significant positive relationships were found for both population and GDP, with p-values less than 0.05. In China, a significant positive relationship was found for population, but not for GDP. In Germany, no significant relationships were found for either variable. The models for India and the UK had high R-squared values and low residual standard errors, indicating that they explained a large proportion of the variance in GHGE. The coefficients for population and GDP varied among the countries but generally indicated small to moderate effects on GHGE. Overall, these results suggest that both population and GDP may play a role in determining GHGE, but the exact relationship may differ among countries.

Then the focus is shifted to time series data, with Holt-Winters model fit to the greenhouse gas emissions data for the overall dataset and for each of the individual countries. The fitted values from these models are plotted.

Finally, the analysis includes two statistical tests - a Mann-Kendall trend test and a Wilcoxon rank-sum test - to evaluate the relationship between greenhouse gas emissions and other variables.

Two vectors are created containing the names of countries in Europe and Asia and subset the data frame to only include data for these countries.

Next, a Mann-Kendall trend test is performed on the GHGE data for Europe and Asia using the Mann-Kendall function. This test is used to determine whether there is a significant trend in the data over time. We also perform the trend test on individual countries to examine whether there are significant trends in the GHGE data for those countries.

```
# Mann-Kendall Test Result for Asia on Greenhouse Gas Emissions
tau = 0.991, 2-sided pvalue =< 0.000000000000000222
# Mann-Kendall Test Result for Europe on Greenhouse Gas Emissions
tau = -0.931, 2-sided pvalue =0.000000001596
# Mann-Kendall Test Result for India on Greenhouse Gas Emissions
tau = 1, 2-sided pvalue =< 0.000000000000000222
# Mann-Kendall Test Result for China on Greenhouse Gas Emissions
tau = 0.948, 2-sided pvalue =< 0.000000000000000222
# Mann-Kendall Test Result for Germany on Greenhouse Gas Emissions
tau = -0.853, 2-sided pvalue =0.000000032613
# Mann-Kendall Test Result for United Kingdom on Greenhouse Gas
Emissions
tau = -0.922, 2-sided pvalue =0.0000000022595
```

The trend in greenhouse gas emissions (GHGE) was evaluated using the Mann-Kendall test for several regions: Asia, Europe, India, China, Germany, and the UK. The results showed a negative trend in GHGE for Europe, Germany, and the UK, with p-values less than 0.05. This suggests that GHGE has decreased over time in these regions. In contrast, the test results showed a positive trend in GHGE for India and China, with p-values less than 0.000000000000000222. This indicates that GHGE has increased over time in these regions. For Asia, the test showed the same trend in GHGE, with a p-value smaller than 0.05. This suggests that the relationship between GHGE and time is clear in this region.

These findings have important implications for climate change and environmental policies. The negative trend in GHGE for Europe, Germany, and the UK suggests that efforts to reduce GHGE in these regions have been successful. This may be due to a variety of factors, such as increased use of renewable energy sources and improved energy efficiency. In India and China, where GHGE has increased, further efforts may be needed to reduce GHGE and mitigate the impacts of climate change.

Additionally, the variation in the trend among different regions highlights the need for region-specific policies and strategies to address GHGE. For example, while the negative trend in Europe, Germany, and the UK may be due to a combination of factors, such as renewable energy and energy efficiency policies, the positive trend in India and China may be influenced by different factors, such as economic growth and industrialization. Therefore, it is important to consider the unique characteristics and circumstances of each region to develop effective policies to reduce GHGE.

Overall, the results of the Mann-Kendall test indicate that the trend in GHGE varies among different regions and that successful strategies to reduce GHGE may differ among these regions. Further research is needed to identify the factors contributing to the trend in GHGE, and to develop targeted policies to reduce GHGE and mitigate the impacts of climate change.

Then a new column created called "DevStatus" in the data frame, which indicates whether a country is considered "developed" or "developing" based on its GDP per capita. Then perform a Wilcoxon rank-sum test on the GHGE data, using the "DevStatus" column as the grouping variable. This test is used to determine whether there is a significant difference in the GHGE data between developed and developing countries.

Finally, another new column is created called "Continent" in the data frame, indicating whether a country is in Asia or Europe. We then perform another Wilcoxon rank-sum test on the GHGE data, using the "Continent" column as the grouping variable. This test is used to determine whether there is a significant difference in the GHGE data between countries in Asia and Europe.

```
	Wilcoxon rank sum test with continuity correction

data:  df$GHGE by df$DevStatus
W = 6466, p-value = 0.0003667
alternative hypothesis: true location shift is not equal to 0


	Wilcoxon rank sum test with continuity correction

data:  df$GHGE by df$Continent
W = 11414, p-value = 0.0000133
alternative hypothesis: true location shift is not equal to 0
```

The Wilcoxon rank sum test was used to evaluate the relationship between GHGE and development status and continent. The test results showed a significant difference in GHGE between developed and developing countries, with a p-value less than 0.05. This suggests that GHGE is higher in developed countries compared to developing countries. The test also showed a significant difference in GHGE among different continents, with a p-value less than 0.05. This indicates that GHGE varies among different continents.

These findings have important implications for climate change and environmental policies. The higher GHGE in developed countries suggests that efforts to reduce GHGE may need to focus on these countries. Additionally, the variation in GHGE among different continents indicates the need for region-specific policies to address GHGE. Further research is needed to identify the factors contributing to the differences in GHGE, and to develop targeted policies to reduce GHGE and mitigate the impacts of climate change.

Discussion

This assignment used a variety of statistical techniques to analyse the data and draw conclusions about the relationship between greenhouse gas emissions and various factors such as GDP and population.

One of the first statistical techniques used in the assignment is the calculation of column-wise and country-wise mean and median. This allows us to summarize the data and get a sense of the typical values for each variable. The calculation of the standard deviation, variance, skew, and kurtosis also provides useful information about the distribution of the data and helps us understand how the data is distributed around the mean.

Scatter plots were also used to visualize the relationship between the different variables. Scatter plots are a useful tool for exploring the relationship between two variables and can help us identify any potential trends or patterns in the data. This assignment used scatter plots to explore the relationship between greenhouse gas emissions and GDP & population.

Simple and multiple regression analyses were also performed to understand the relationship between greenhouse gas emissions and other variables. Regression analysis is a statistical technique that allows us to estimate the strength and direction of the relationship between two or more variables. This assignment used simple regression to analyse the relationship between greenhouse gas emissions and GDP & population, and then use multiple regression to include additional variables.

In addition to regression analysis, time series analysis was also performed on the data. Time series analysis is a statistical technique that allows us to analyse data over time and identify any trends or patterns that may be present. This assignment used time series analysis to explore the relationship between greenhouse gas emissions and time.

Finally, the Mann-Kendall Test and the Wilcox Test are performed on the greenhouse gas emissions data. These statistical tests are used to evaluate the significance of the relationship between two variables and can help us determine whether any observed trends in the data are statistically significant.

Overall, the assignment uses a variety of statistical techniques to analyse the relationship between greenhouse gas emissions and other factors such as GDP, population, and development status. I have used mean and median calculations, scatter plots, regression analysis, time series analysis, and statistical tests to explore the data and draw conclusions about the relationship between these variables. These statistical methods are widely used in many fields and can provide valuable insights into complex data sets.

Conclusions

In conclusion, this study used a variety of statistical techniques to analyze the relationship between greenhouse gas emissions (GHGE) and other factors such as GDP, population, and development status. The authors found that GHGE is positively correlated with GDP and population, indicating that as these variables increase, GHGE tends to increase as well. Multiple regression analysis showed that both GDP and population have a statistically significant effect on GHGE in some countries, but the exact relationship may vary among countries. Time series analysis revealed a general trend of increasing GHGE over time, which was confirmed by statistical tests. These findings highlight the importance of considering economic and population factors in efforts to reduce GHGE and address climate change. Further research is needed to better understand the specific mechanisms and factors that influence GHGE and to develop effective strategies for mitigating its effects.

References

- World Bank. (2022). World development indicators. Washington, DC: World Bank. Retrieved from https://datacatalog.worldbank.org/dataset/world-development-indicators
- Few, S. (2012). Designing dashboards: A guide for business and technology. O'Reilly Media, Inc.
- Few, S. (2006). Information dashboard design: The effective visual communication of data. O'Reilly Media, Inc.
- Majaski, C. (2022, October 2). What is a developed economy? definition, how it works, HDI index. Investopedia. Retrieved December 13, 2022, from https://www.investopedia.com/terms/d/developed-economy.asp
- European Commission. (2018). EU Emissions Trading System (ETS). Retrieved from https://ec.europa.eu/clima/policies/ets_en
- European Commission. (2018). Renewable Energy Directive. Retrieved from https://ec.europa.eu/energy/en/topics/renewable-energy
- European Environment Agency. (2018). GHG emissions in the EU. Retrieved from https://www.eea.europa.eu/themes/climate/greenhouse-gas-emissions/ghg-emissions-in-the-eu
- Federal Ministry for Economic Affairs and Energy. (2018). Energy transition in Germany. Retrieved from https://www.bmwi.de/EN/Topics/energy/energy-transition.html
- Ministère de la Transition Écologique et Solidaire. (2018). Carbon tax in France. Retrieved from https://www.ecologique-solidaire.gouv.fr/en/carbon-tax-france
- European Commission. (2013). Assessment of the risk of carbon leakage. Retrieved from https://ec.europa.eu/clima/policies/ets/reform/docs/com_2013_448_en.pdf
- Stavins, R. (2013). Can the European Union's emissions trading system prevent carbon leakage? Environmental and Resource Economics, 56(2), 267-286. Retrieved from https://doi.org/10.1007/s10640-012-9612-4