Does Economic Growth Increase CO_2 Emissions?

An Analysis of Global GDP and Environmental Impact

Background

This project looks at the relationship between economic growth and CO_2 emissions. As countries get richer, their industries and energy use often increase, which can lead to more emissions. I want to check if there is a clear pattern between GDP and CO_2 emissions from 1990 to 2023. This period includes rapid global development, especially in countries like China and India, as well as major events like economic crises and international climate agreements.

Other studies have shown that GDP and CO_2 emissions are usually positively related, but recently some countries have started to grow without increasing emissions as much. This is known as "decoupling," where economic activity grows while emissions remain stable or decline. Using real-world data, I will explore if this trend holds globally or if economic growth still causes more CO_2 emissions overall.

Data Sources and Cleaning

GDP Data: World Bank - World Development Indicators.
CO_2 Emissions Data: Our World in Data - CO_2 and Greenhouse Gas Emissions.

Both datasets contain yearly country-level data. GDP is in current US dollars; CO_2 is in $MtCO_2$

Steps taken**:**

- Removed extra rows and cleaned column names
- Reshaped into long format (Country, Year, Value)
- Dropped countries with too much missing data (e.g., British Virgin Islands, South Sudan)
- Merged on Country and Year
- Removed any remaining rows with missing values

Final dataset: 1990–2023, no missing values.

Exploratory Data Analysis

Summary Statistics

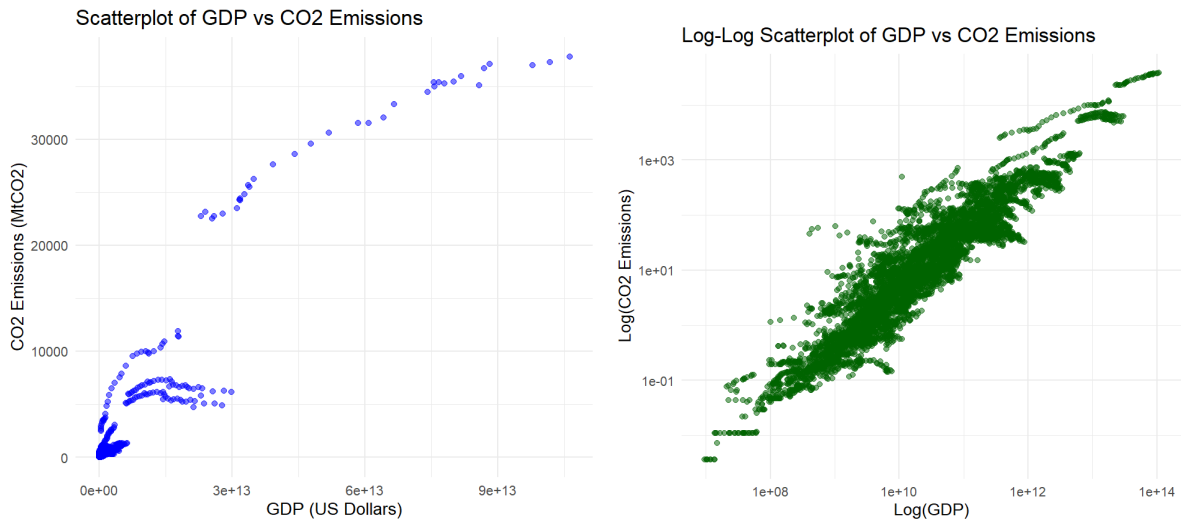
GDP:

- Min: $9.54 million | Max: $106 trillion
- Mean: $712 billion | Median: $15.98 billion
- Highly skewed — large economies raise the average

CO_2 Emissions:

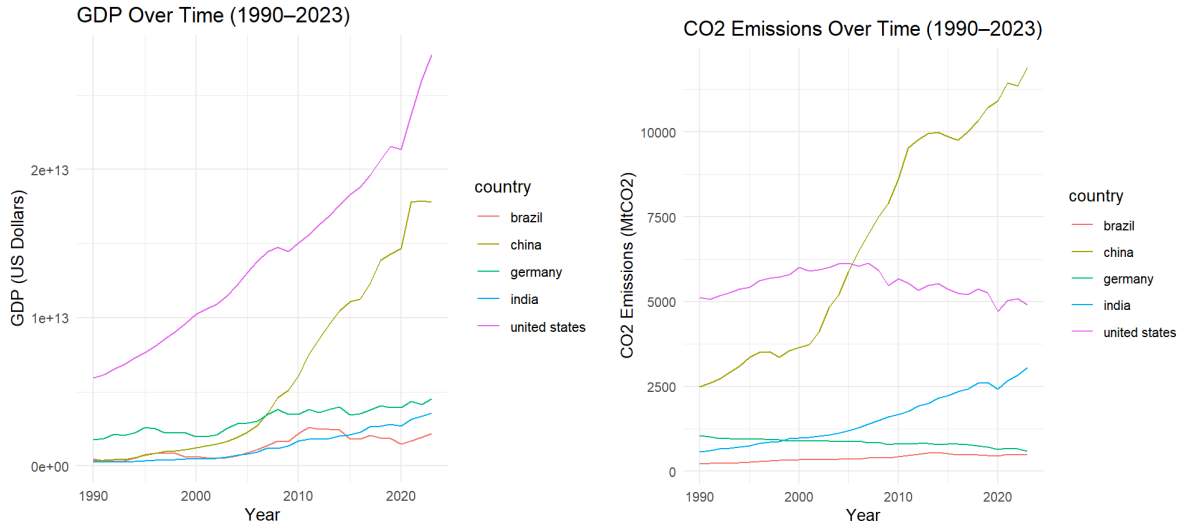
- Min: 0.00366 $MtCO_2$ | Max: 37,792 $MtCO_2$
- Mean: 356 $MtCO_2$ | Median: 8.69 $MtCO_2$
- Also highly skewed — dominated by China, USA, etc

Scatterplots of GDP vs CO_2 Emissions (Raw and Log-Log Scales)



The first scatterplot shows a positive relationship between GDP and CO_2 emissions, with high-GDP countries producing more emissions. However, the spread is wide and many lower-GDP countries are clustered near the origin, making it hard to interpret. The second plot uses a log-log scale, which reveals a much clearer linear trend. This transformation reduces skewness and highlights that the GDP–CO_2 relationship may be exponential. The log-log version is more suitable for regression and supports using a log-transformed model in later analysis.

GDP and CO_2 Emissions Over Time (1990–2023)



These two line plots show how GDP and CO_2 emissions have changed over time for five major economies. China's GDP and emissions both rose rapidly after 2000, while India also shows steady growth in both. The United States remained the largest economy throughout the period but saw its emissions slightly decline after the late 2000s. Germany and Brazil show more stable GDP growth, with relatively flat or declining emissions. Together, the trends suggest that while economic growth is generally linked with rising emissions, some countries may be starting to decouple GDP from CO_2 output.

Model: Simple Linear Regression

I fit a linear regression model using GDP as the predictor and CO_2 emissions as the response variable to assess the strength and direction of their relationship. The model produced a strong positive relationship, with an R-squared value of 0.905. This means that approximately 90.5% of the variation in CO_2 emissions can be explained by differences in GDP across countries and years, which is a very high level of explanatory power. The coefficient for GDP was both positive and highly statistically significant ($p < 2.2e\text{-}16$), confirming that, in general, countries with higher economic output tend to emit more CO_2.

Despite the strong fit, the residuals from the model ranged widely, and the residual standard error was relatively high (757.2). This suggests that while GDP is a strong predictor, the linear model may not fully capture the complexity or curvature in the relationship—especially due to the large scale differences between smaller and larger economies. These findings, along with the earlier scatterplots showing strong skewness, support the use of a log-log transformation in the next phase of modeling. A log-log regression is likely to provide a better fit and offer more stable variance across different levels of GDP.

Future Plans

In the next stage of this project, I plan to go beyond visualizations and simple modeling by applying techniques directly related to statistical computing, as emphasized in this course.

- **Log-Log Regression Model**:
  Based on the log-log scatterplot and the skewness of the data, I will fit a log-transformed model using $\log(CO_2) \sim \log(GDP)$ and compare its performance to the simple linear model. This should improve model fit and reduce heteroscedasticity.

- **Monte Carlo Simulation**:
  I will use Monte Carlo methods to simulate future GDP increases (e.g., +5% or +10%) and compute predicted CO_2 emissions using the regression model. This allows me to quantify the potential environmental impact of economic growth using simulation.

- **Bootstrapping**:
  If time permits, I may implement bootstrapping to estimate the sampling distribution of the regression slope. This will help quantify uncertainty in the GDP–CO_2 relationship without relying strictly on parametric assumptions.

GitHub Repository

The full project code, cleaned datasets, R scripts, plots, and report drafts are available at:

https://github.com/sushanthvk02/gdp-co2-emissions

Questions

One question I have for the final report is:

> If I organize all R scripts, plots, and datasets into separate folders in my GitHub repository (for example, /scripts/, /plots/, /data/), would I still need to attach supplementary materials like R code, figures, and tables inside the final report document itself, or is linking to the GitHub repository enough?