Does Economic Growth Increase CO_2 Emissions?

An Analysis of Global GDP and Environmental Impact

I. Introduction

Economic growth and environmental sustainability are often thought to be in conflict. As countries expand their economies, they tend to consume more energy and produce more goods, which usually leads to an increase in carbon dioxide (CO_2) emissions — one of the main drivers of climate change. For a long time, it was widely accepted that higher GDP almost always meant higher emissions. But that connection is starting to look more complicated.

In this project, I explore whether GDP is still a strong predictor of CO_2 emissions globally, using data from 1990 to 2023. I wanted to see not just if richer countries emit more, but also whether that pattern is changing — and if the world might be starting to decouple economic growth from environmental damage.

To do this, I went beyond a simple regression. I used log-log models to account for skewed variables, ran Monte Carlo simulations to test how emissions could respond to different GDP growth scenarios, and used bootstrapping to check the stability of my results. Based on feedback from my earlier work, I also added a fairness analysis — comparing countries not just by total emissions, but also by CO_2 per person and per dollar of GDP. This felt especially important given the way countries like India and China are often criticized for their total emissions, despite having much lower emissions per capita than developed nations.

This is a solo project, and my goal was to take a data-driven approach to a big global question: Is economic growth still tightly tied to CO_2 emissions, or are we finally seeing signs of decoupling?

II. Data and Methodology

For this analysis, I created a merged dataset of country-year observations spanning from 1990 to 2023. The data came from three public sources: GDP figures from the World Bank, CO_2 emissions data from Our World in Data, and population estimates from the same site. I cleaned and reshaped each dataset separately before merging them using country and year as keys.

The main outcome variable in my study is territorial CO_2 emissions, measured in million tonnes ($MtCO_2$). The primary predictor is GDP in current US dollars. To get a fuller picture and explore fairness and efficiency, I also calculated a few derived metrics:

- CO_2 emissions per capita
- GDP per capita

To prepare the data, I converted all three datasets to long format with columns for country, year, and value. I also standardized the country names (converted them to lowercase) to avoid mismatches during the merger. After combining the datasets, I removed rows with missing values for any of the key variables to ensure clean input for modeling.

In the early stages of exploration, I noticed that both GDP and CO_2 emissions were highly skewed — with a few very large economies dominating the totals. To address this, I applied a log transformation to both variables and used a log-log model to better capture their multiplicative relationship.

The main statistical method I used is linear regression. I started with both a standard linear model and a log-transformed version. Then, to go beyond a single model output, I ran a Monte Carlo simulation to estimate how CO_2 emissions might respond to different levels of GDP growth. I also used bootstrapping to understand the variability in the slope estimate of the model. Finally, I added fairness metrics like per capita and efficiency-based comparisons to get a more balanced view of which countries are contributing most to global emissions.

All data cleaning, transformation, and analysis were done using R. The sources for the datasets are listed below:

- GDP: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
- CO_2 Emissions: https://ourworldindata.org/co2-dataset-sources
- Population: https://ourworldindata.org/population-growth#explore-data-on-population-growth

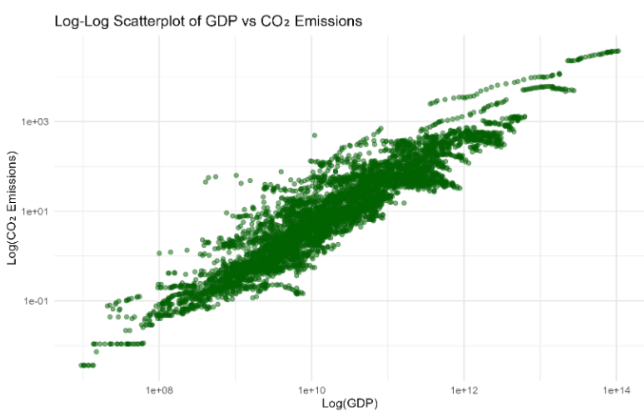
III. Modeling and Statistical Analysis

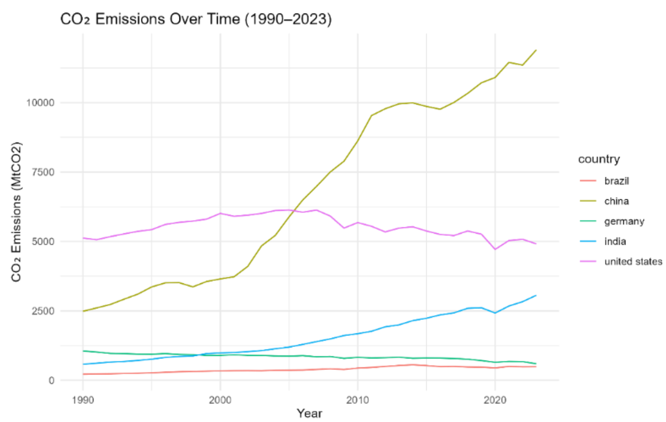
Exploratory Data Analysis

The dataset includes country-level observations for GDP, CO_2 emissions, and population spanning from 1990 to 2023. All three variables show substantial variation across countries and time. GDP values range from approximately $9.5 million to over $106 trillion, with a mean of $628 billion and a median of only $15.7 billion, indicating a strong right skew caused by a few extremely large economies. CO_2 emissions follow a similar pattern, ranging from less than 0.004 $MtCO_2$ to nearly 37,800 $MtCO_2$, with a median of 8.5 and a mean of 320 $MtCO_2$. Population values span from under 9,000 to over 8 billion people. These extreme differences in scale justify the later use of log transformations for modeling.

Visualizations provide a clearer understanding of trends across key countries. Line plots of GDP and CO_2 emissions over time for the United States, China, India, Brazil, and Germany show that while all have experienced steady economic growth, their emissions paths vary. For instance, China shows rapid increases in both GDP and CO_2, whereas the United States exhibits rising GDP alongside flat or slightly declining emissions since the late 2000s—potentially signaling the start of decoupling in advanced economies.

Scatterplots of GDP versus CO_2 emissions further illustrate the relationship. On the original scale, the plots are dominated by outliers, but log-log transformations reveal a much more linear and interpretable trend, supporting a multiplicative model. Colored scatterplots by year and faceted plots by decade suggest that the shape and slope of this relationship have shifted slightly over time, especially in the 2010s and 2020s. Population growth, analyzed through its own trend lines and scatterplots, reinforces the importance of adjusting for scale when comparing emissions across countries. These findings lay the groundwork for the fairness analysis explored later in the report.





Linear Model

The first step in modeling involved fitting a basic linear regression using GDP to predict CO_2 emissions. This model performed well, with an R^2 of 0.91, indicating that GDP alone accounts for over 90% of the variation in emissions. However, residuals from this model were highly skewed, and both GDP and CO_2 variables exhibited long right tails, making the use of a log transformation more appropriate.

A log-log model using $\log(CO_2) \sim \log(GDP)$ was then fitted to address this skewness. The transformed model showed a slightly lower R^2 of 0.87 but produced more balanced residuals and a more interpretable slope. The slope coefficient was approximately 0.96, suggesting that emissions increase nearly proportionally with GDP on a percentage basis. This transformation also helped stabilize variance across the range of fitted values.

To further refine the model, log(Population) was added as a second predictor in a multiple log-log regression. This extended model slightly increased the R^2 to 0.88 and reduced the residual standard error. Both predictors were statistically significant. The coefficient for GDP dropped to 0.84, while the population coefficient was around 0.16, implying that emissions grow with both economic and demographic scale, though not at a one-to-one rate.

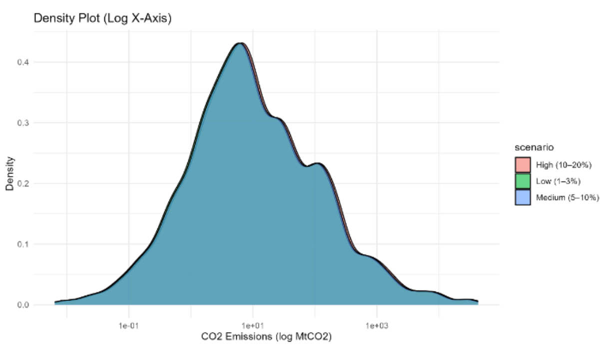
An ANOVA test comparing the single-predictor and two-predictor log-log models showed a statistically significant improvement ($F = 409$, $p < 2.2e-16$) when population was included. As a result, the multiple regression model was used as the foundation for all simulations and resampling procedures in the later parts of the project.

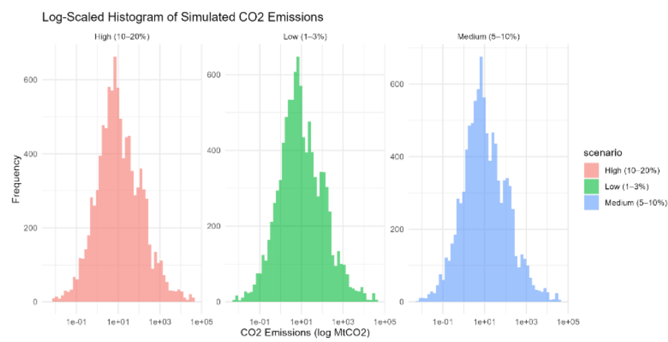
Monte-Carlo Simulation

To explore the uncertainty in predicted CO_2 emissions, a Monte Carlo simulation was run using the log-log regression model of CO_2 emissions on GDP. Three economic growth scenarios were defined: low (1–3%), medium (5–10%), and high (10–20%). For each scenario, 1,000 GDP values were randomly sampled from the dataset. Each sampled value was then increased by a random percentage within the specified range, and CO_2 emissions were predicted using the log-log model. The predictions were then exponentiated to return them to the original scale.

The resulting distributions were heavily right-skewed across all three scenarios. Most predicted emissions fell below 300 $MtCO_2$, but some values, particularly in the high-growth scenario, exceeded 10,000 $MtCO_2$. Summary statistics showed that while the mean increased steadily with the growth rate—from around 287 $MtCO_2$ in the low-growth case to about 322 $MtCO_2$ in the high-growth case—the median remained much lower, between 8 and 9 $MtCO_2$. This large gap between mean and median highlights the influence of outliers and supports the use of log transformation to manage skewed data.

These simulation results offer a probabilistic view of how CO_2 emissions might evolve under different economic growth conditions. The outcomes also reinforce the idea that economic growth and emissions do not scale linearly in practice, and that large economies play a disproportionately large role in shaping global emission totals.

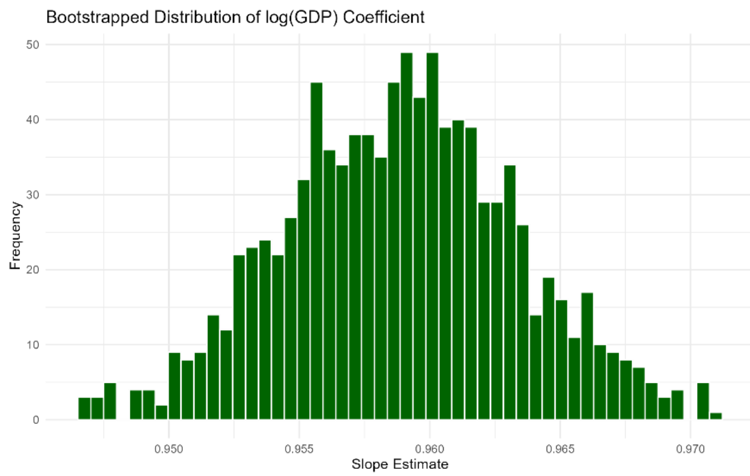




Bootstrapping

To test the stability of the log-log regression model ($\log(CO_2) \sim \log(GDP)$), a bootstrapping procedure was performed using 1,000 resampled datasets. This allowed for the estimation of confidence intervals for the model's parameters. The 95% confidence interval for the intercept ranged from approximately -20.68 to -20.26, while the slope ranged from 0.950 to 0.968. The average slope across all bootstrapped models was about 0.959, showing that the relationship between GDP and CO_2 emissions remains consistently strong across repeated samples. These narrow intervals suggest that the model is robust and not highly sensitive to fluctuations in the data.

The bootstrapped models were also used to predict CO_2 emissions at three representative GDP levels: the 25th percentile (low GDP), 50th percentile (median GDP), and 90th percentile (high GDP). The results showed average predicted emissions of about 1.83 $MtCO_2$, 7.75 $MtCO_2$, and 204.7 $MtCO_2$, respectively. Even though these GDP levels aren't drastically different in relative terms, the predicted emissions differ by more than two orders of magnitude. This reflects the exponential structure of the model and highlights how much more heavily emissions are concentrated in high-GDP countries. These results also tie into broader fairness concerns, especially when considering how responsibility for emissions is distributed in global climate policy discussions.

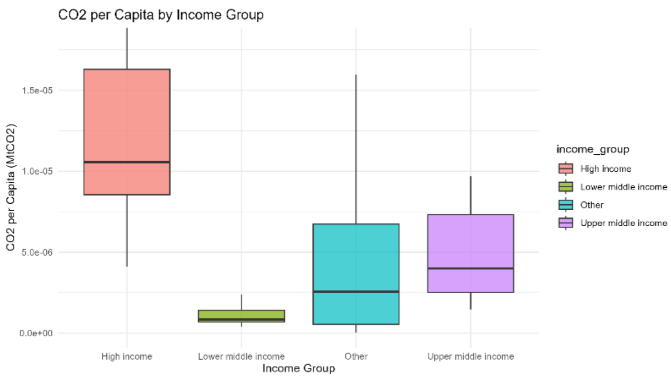


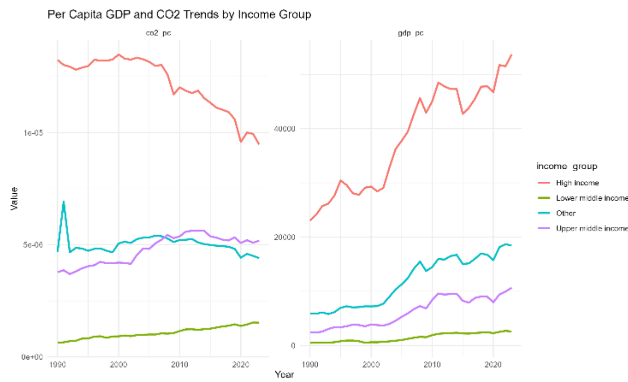
Fairness Analysis

To explore fairness in emissions and income across countries, I calculated both GDP per capita and CO_2 emissions per capita for every country-year observation. Countries were then grouped into four broad income categories—High income, Upper middle income, Lower middle income, and Other—based on well-known examples. This grouping was used to highlight differences in economic output and environmental impact across development levels.

The results showed clear disparities. High-income countries had a much higher average GDP per capita (around $12,344) and emitted significantly more CO_2 per person (about 5.13e-6 $MtCO_2$). Median values told a similar story, confirming that the distribution is skewed toward wealthier nations. Boxplots of CO_2 per capita by income group made this even more visible, with high-income countries consistently clustering at the upper end of the emissions scale.

Trends over time added more context. While high-income countries have maintained their economic lead, their per capita emissions have slightly declined since the early 2000s. In contrast, both upper- and lower-middle-income countries have seen steady increases in GDP per capita and emissions. This suggests that while developing economies are making progress, they're doing so with a growing environmental footprint. These findings point to the challenges of designing climate policies that are fair—not just in terms of total output, but also in how emissions are distributed relative to income and population.





IV. Conclusion and Discussion

This project set out to examine the relationship between economic growth and CO_2 emissions using a global dataset covering 1990 to 2023. The results showed that GDP remains a strong predictor of emissions, with both standard and log-log regression models explaining a large share of the variance. Resampling methods like bootstrapping and Monte Carlo simulation helped confirm the stability of that relationship under different scenarios and assumptions.

At the same time, there were signs that this link may be starting to weaken—at least in some high-income countries. In places like the United States and parts of Europe, emissions have begun to flatten or even decline despite continued economic growth. This suggests that decoupling may be underway, possibly driven by cleaner energy, stricter climate policies, or structural changes in the economy. But this isn't happening everywhere. In many lower- and middle-income countries, emissions are still rising right alongside GDP. So, while decoupling is real in some cases, it hasn't reached a global scale.

The fairness analysis added another layer to the story. High-income countries still emit far more CO_2 per person than lower-income ones, even though the latter are often under more pressure to reduce emissions. Countries like India and Indonesia contribute relatively little on a per capita basis but are still expected to meet the same global targets. This raises important questions about how responsibility is measured—and who carries the weight of climate action.

In the end, this project showed that while economic growth still drives emissions globally, the relationship is starting to shift in some regions. Whether decoupling can become a global trend will depend on international collaboration, policy, and support for sustainable growth in developing economies. Modeling, simulation, and fairness-based comparisons all played a role in helping me understand that dynamic more clearly—and they're tools I'll continue to use in future work.

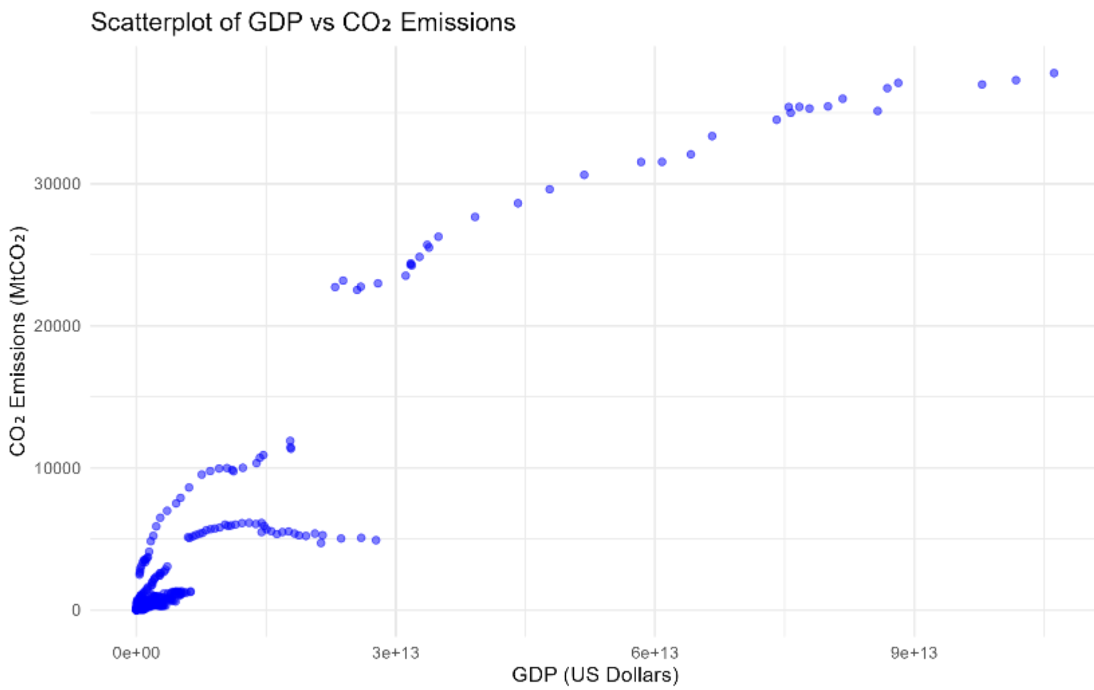
V. Appendix

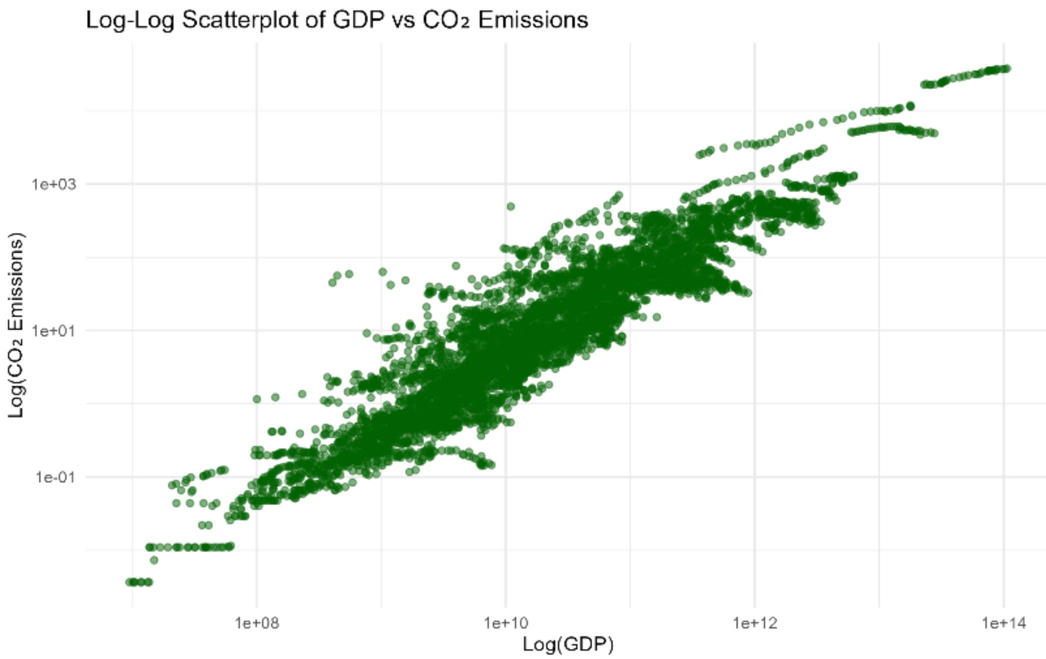
Exploratory Data Analysis

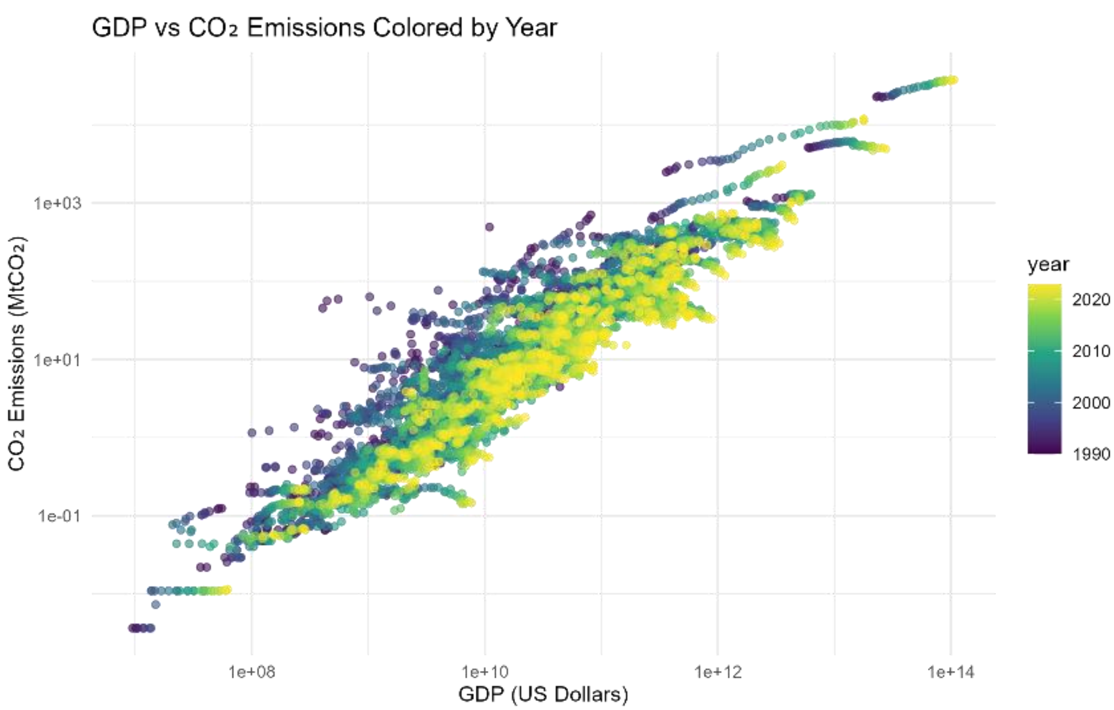
Summary Statistics

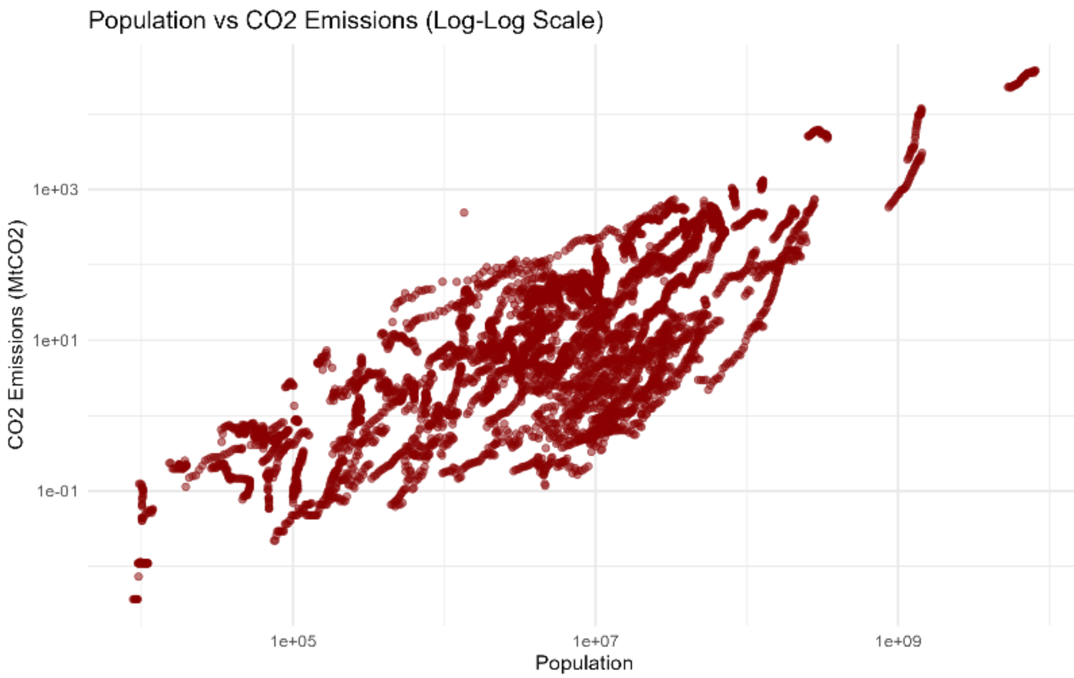
Variable	Min	Max	Mean	Median
GDP	9542900.901	1.06172E+14	6.2792E+11	15695598704
CO2 Emissions	0.003664	37791.57	319.881966	8.5151635
Population	8778	8091734853	73208456.87	6796328.5

Data Analysis Visualizations

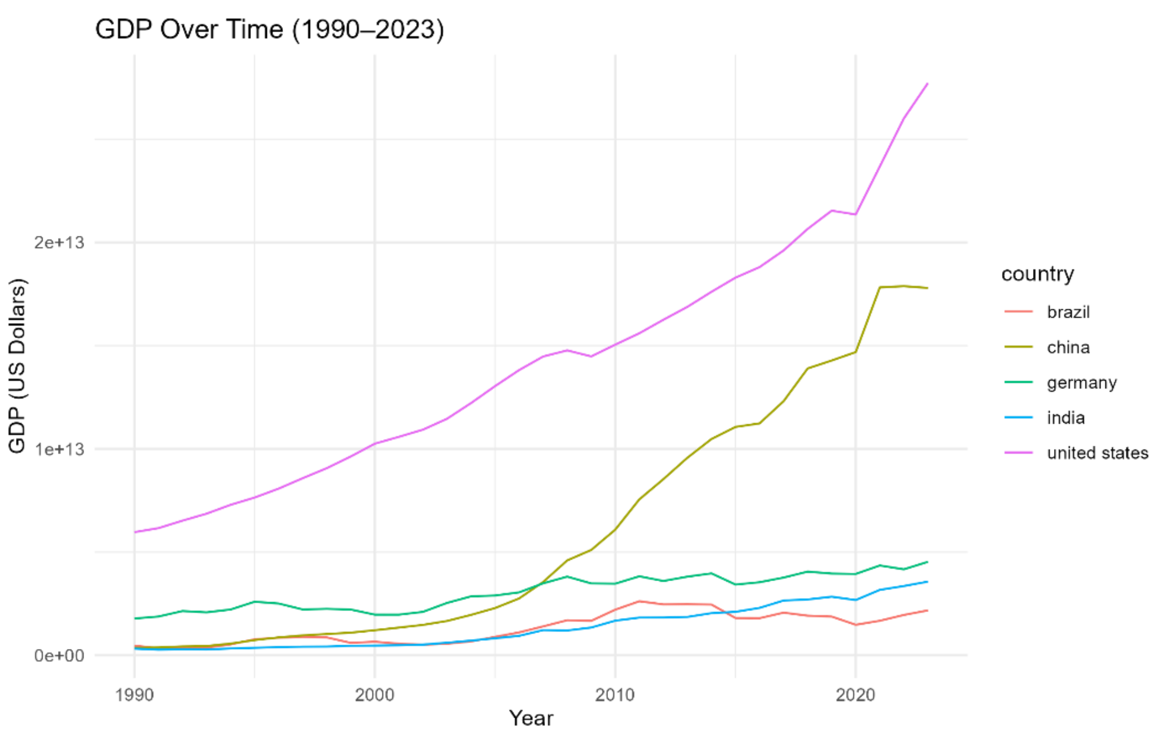



Log-Log Scatterplot of GDP vs CO_2 Emissions
Log(CO_2 Emissions)
1e+03
1e+01
1e-01
1e+08
1e+10
1e+12
1e+14
Log(GDP)

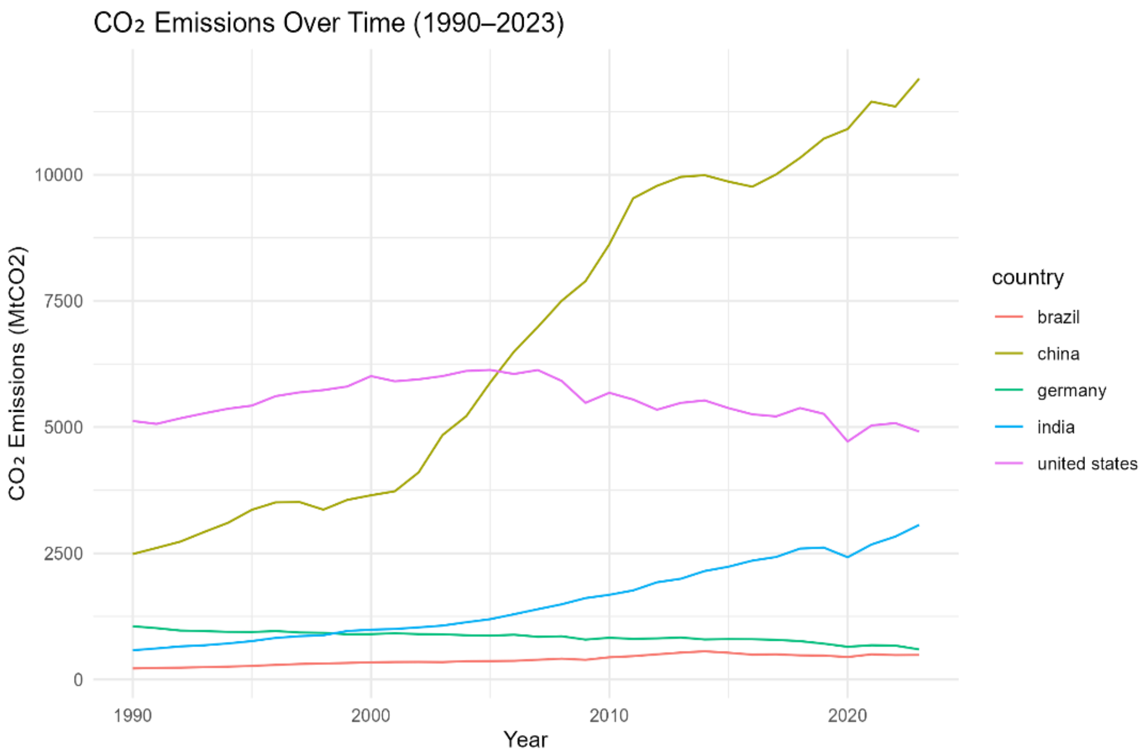

GDP vs CO_2 Emissions Colored by Year
CO_2 Emissions ($MtCO_2$)
1e+03
1e+01
1e-01
1e+08
1e+10
1e+12
1e+14
GDP (US Dollars)
year
2020
2010
2000
1990



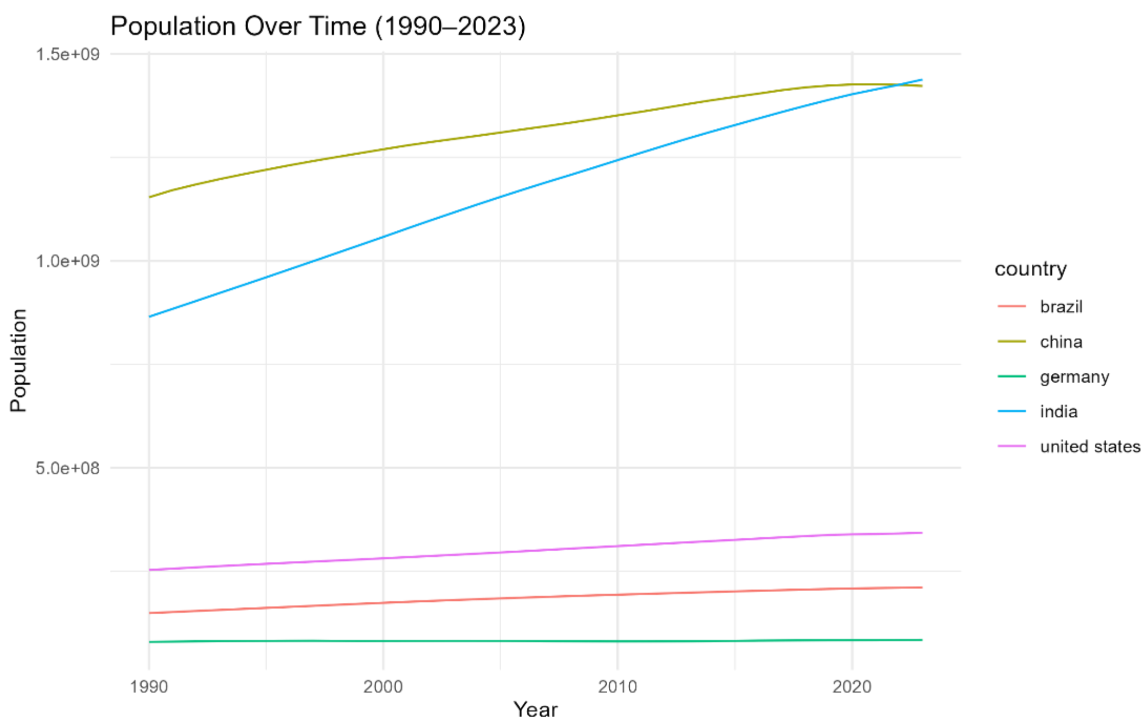
Population vs CO2 Emissions (Log-Log Scale)
CO2 Emissions (MtCO2)
1e+03
1e+01
1e-01
1e+05
1e+07
1e+09
Population



GDP Over Time (1990–2023)
GDP (US Dollars)
2e+13
1e+13
0e+00
1990
2000
2010
2020
Year
country
brazil
china
germany
india
united states



CO₂ Emissions Over Time (1990–2023)
CO₂ Emissions (MtCO2)
10000
7500
5000
2500
0
1990
2000
2010
2020
Year
country
brazil
china
germany
india
united states



Population Over Time (1990–2023)
Population
1.5e+09
1.0e+09
5.0e+08
1990
2000
2010
2020
Year
country
brazil
china
germany
india
united states

Linear Model

Simple Linear Model Coefficients

term	estimate	std.error	statistic	p.value
(Intercept)	23.92158167	9.475002453	2.524704536	0.01160556
gdp	4.71E-10	1.92E-12	245.8796852	0

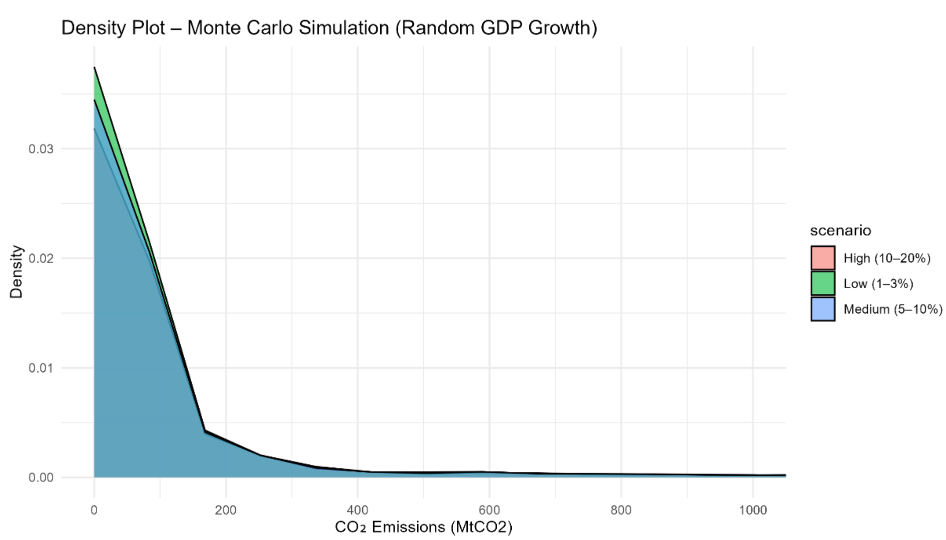
Log Model Coefficients

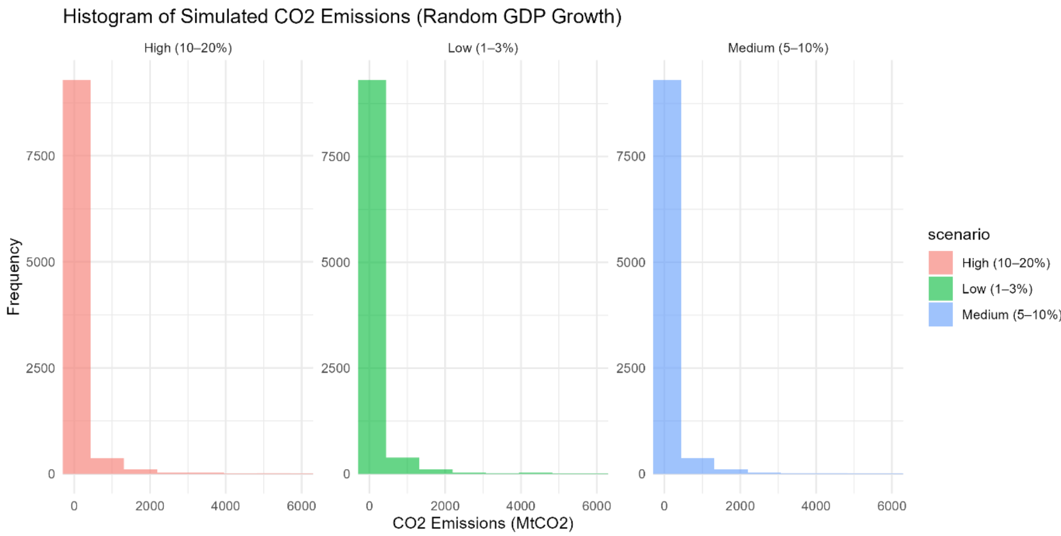
term	estimate	std.error	statistic	p.value
(Intercept)	-20.4658878	0.113586417	-180.1790071	0
log_gdp	0.958950114	0.004771359	200.9804858	0

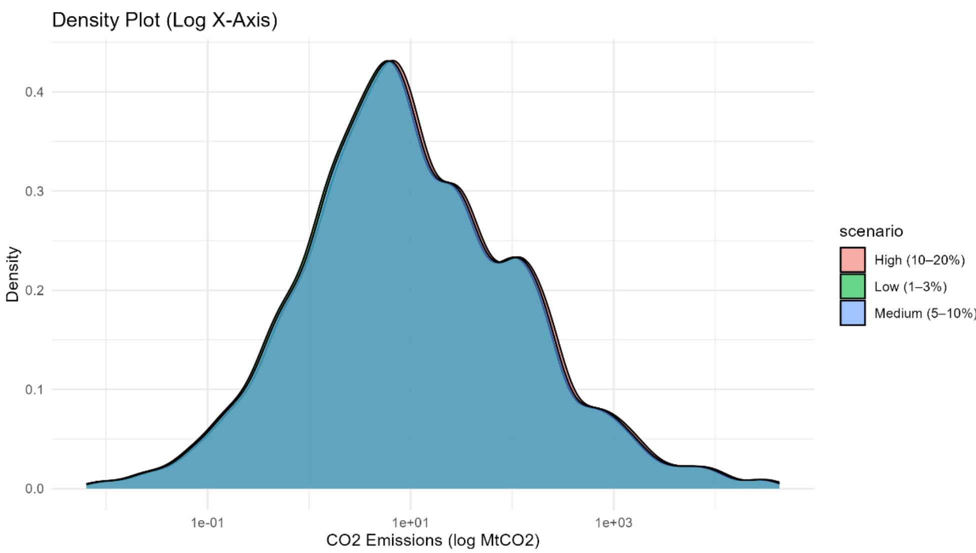
Multi Log Model Coefficients

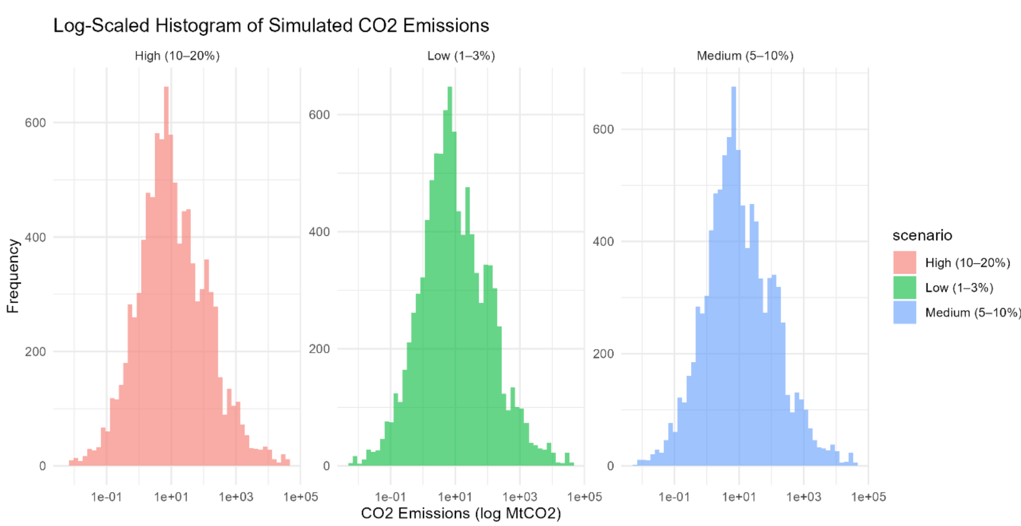
term	estimate	std.error	statistic	p.value
(Intercept)	-20.23942925	0.110420541	-183.2940598	0
log_gdp	0.843494334	0.007340459	114.9103043	0
log_population	0.162946811	0.008056866	20.22458881	5.41E-88

Monte-Carlo Simulation




Histogram of Simulated CO2 Emissions (Random GDP Growth)
High (10–20%)
Low (1–3%)
Medium (5–10%)
Frequency
CO2 Emissions (MtCO2)
scenario
High (10–20%)
Low (1–3%)
Medium (5–10%)

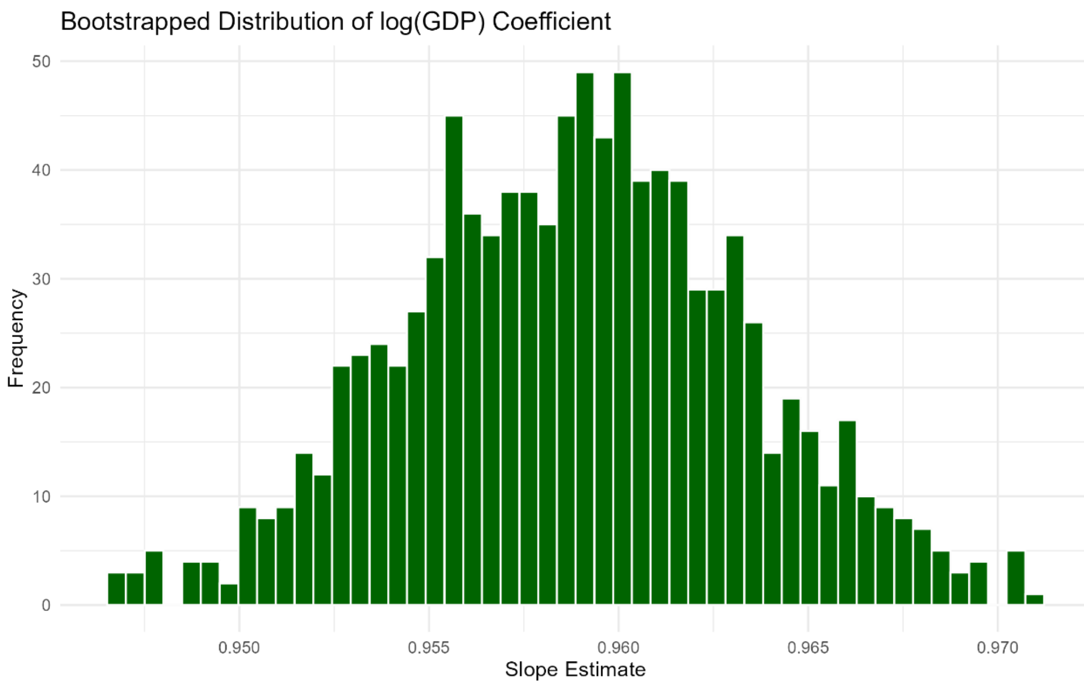

Density Plot (Log X-Axis)
Density
CO2 Emissions (log MtCO2)
scenario
High (10–20%)
Low (1–3%)
Medium (5–10%)


Log-Scaled Histogram of Simulated CO2 Emissions
High (10–20%)
Low (1–3%)
Medium (5–10%)
Frequency
CO2 Emissions (log MtCO2)
scenario
High (10–20%)
Low (1–3%)
Medium (5–10%)

scenario	mean	sd	min	max	p10	median	p90
High (10-20%)	321.960426	2177.8909	0.00701847	42986.758	0.5560062	8.897927	253.47381
Low (1-3%)	286.824523	1936.861	0.00641446	37838.783	0.4942609	7.927744	226.65235
Medium (5-10%)	301.538811	2034.5854	0.00665008	39185.452	0.5187923	8.34163	238.21362

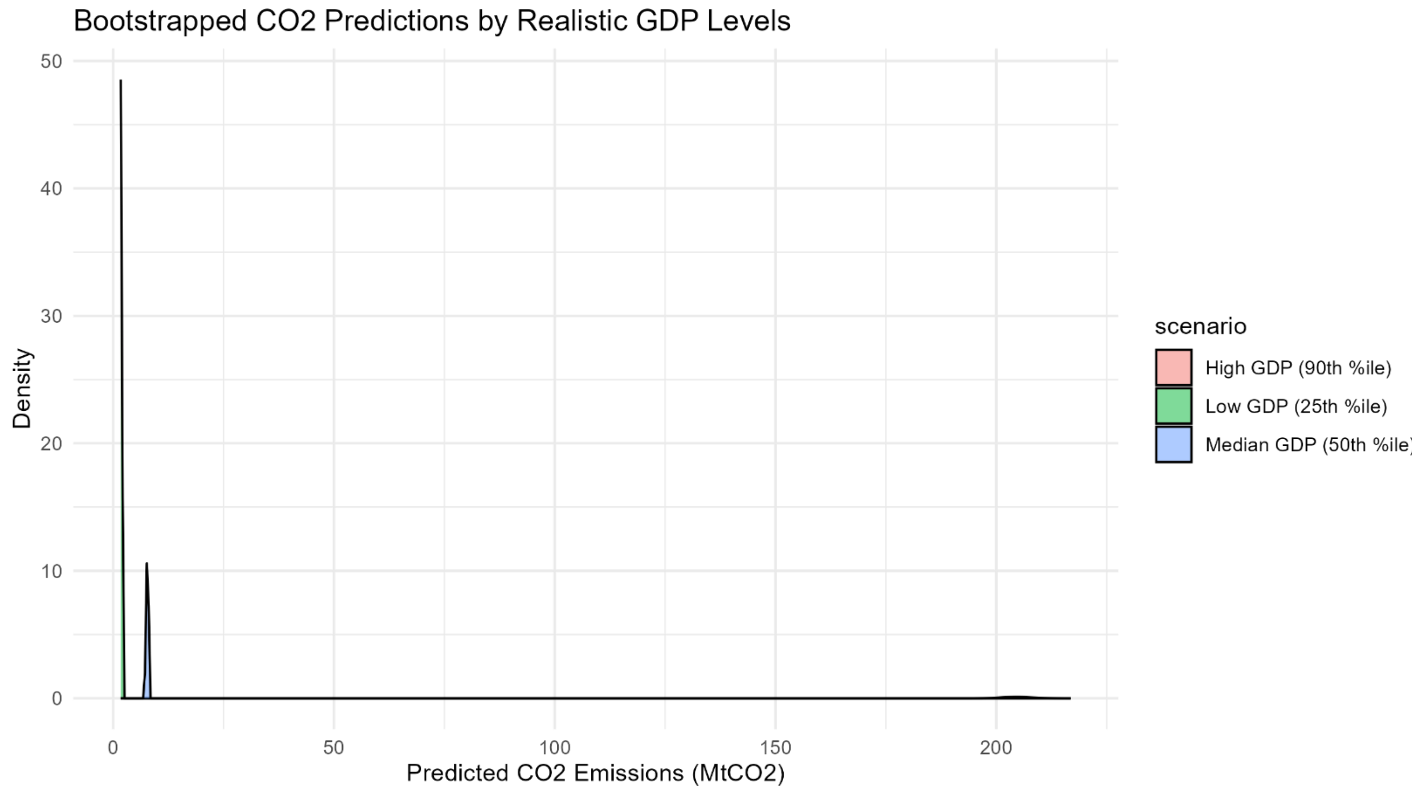
Bootstrapping

Bootstrapping Regression Coefficients



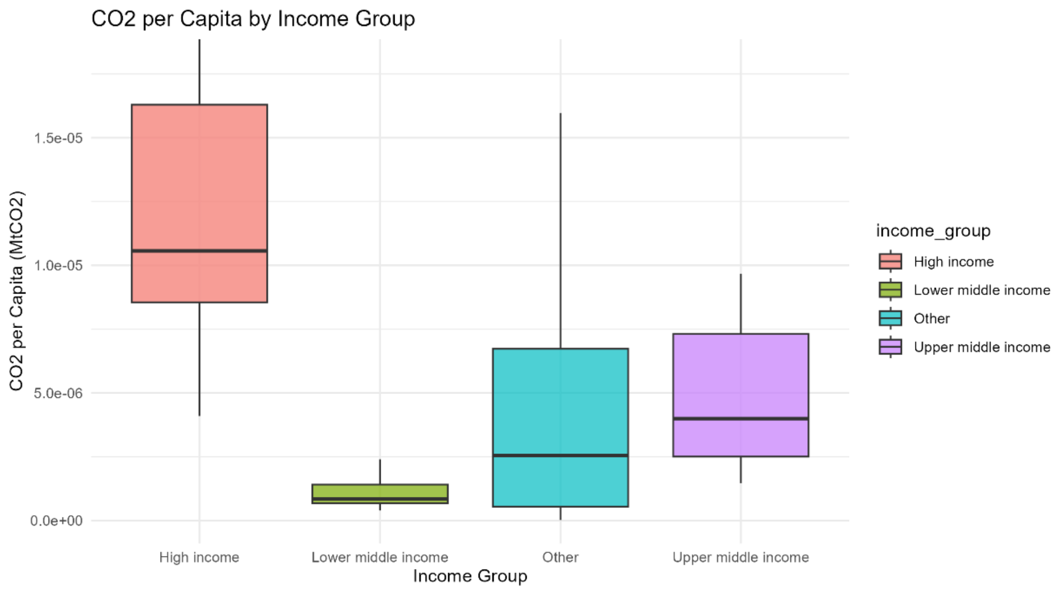
intercept_mean	intercept_lower	intercept_upper	slope_mean	slope_lower	slope_upper
-20.46263768	-20.68251437	-20.25735192	0.958833936	0.950131537	0.967729183

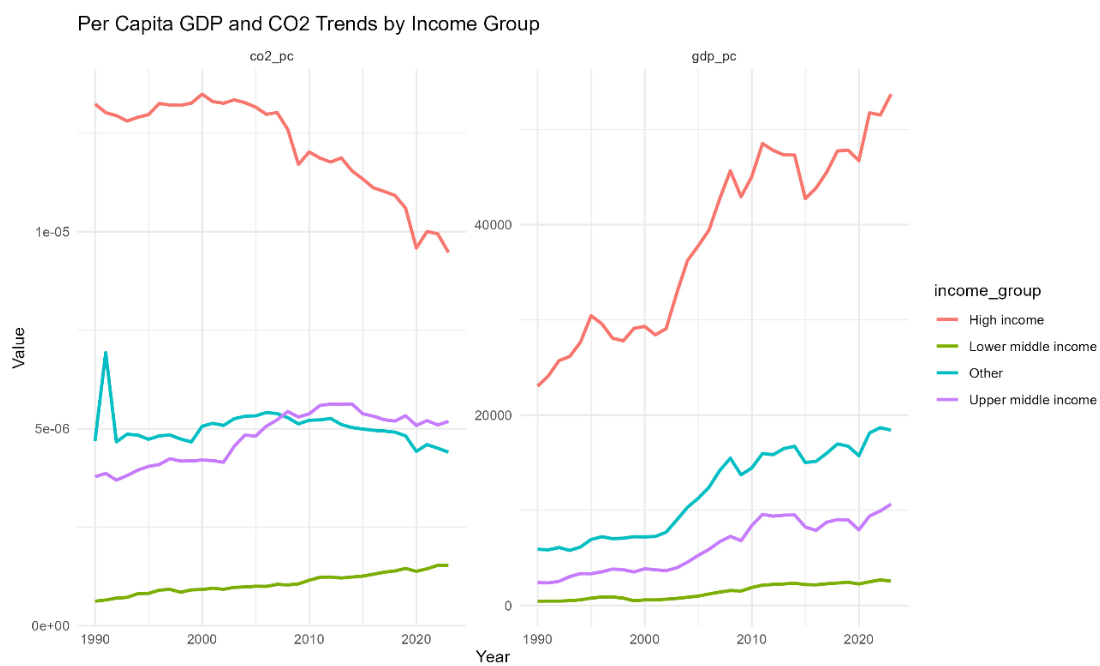
Bootstrapped Prediction Distributions



scenario	mean	sd	min	max	p10	median	p90
High GDP (90th %ile)	204.74134	3.6228632	192.75015	216.86906	200.1718	204.6371	209.4968009
Low GDP (25th %ile)	1.82540357	0.0266918	1.7446755	1.9071644	1.7895153	1.825985	1.859086628
Median GDP (50th %ile)	7.74957246	0.0919579	7.4629085	8.0850069	7.6314527	7.749077	7.86910739

Fairness Analysis





Fairness Summary

gdp_pc_mean	gdp_pc_median	co2_pc_mean	co2_pc_median
12343.738	3950.88309	5.13E-06	2.68E-06

Authors & Repository Information

This project was completed by Viswa Sushanth Karuturi as part of the requirements for DATA 375: Statistical Computing under Professor Niu.

All data, analysis, and code used in this project are available in our GitHub repository: https://github.com/sushanthvk02/ gdp-co2-emissions

The repository is organized as follows:

- data: Raw input datasets, cleaned datasets, and the final merged dataset.
- docs: Project related reports
- plots: All plots and diagnostic graphics generated from our models and computing
- scripts: All R scripts used for data cleaning, visualization, modelling, computing, etc.
- tables: CSVs of model outputs, coefficient tables, test metrics, and more.

This structure ensures that the entire analysis is fully transparent and reproducible.