Carbon Emission Analysis

DATA 5100, Final Project

By: Alyssa, Ben, Travis

Agenda

Introduction

- Domain Problem
- Libraries Used

Data Description

- Data Sources
- Collection + Preparation

Analysis

- Methodology
- Timeline of Analysis
- EDA Overview

Results

- Model Results
- Testing Results
- Expansions

Conclusion

- Key Takeaways

Introduction

- Domain Problem
- Libraries Used

What is Carbon Dioxide Emission?

https://www.myclimate.org/en/information/faq/faq-detail/what-is-co2-and-where-does-it-come-from/

Domain Problem

“Is higher share of renewable energy in total energy supply associated with lower per-capita CO_2 emissions?”

Python Libraries

Data Processing

Cleaning and Standardization

- **pandas** - ETL and Merging
- **numpy** - Numerical operations
- **pycountry** - ISO-3 codes

Visualization

Statistical plots and Styling

- **seaborn** - Heatmaps, Scatterplots
- **matplotlib** - Fine tuning, styling

Modeling

Regression and Features

- **Statsmodels** - Fixed Effects, Robust SE
- **Scikit-Learn** - Lagged features

Data

- Data Sources
- Data Preparation

Data Sources

- **International Energy Agency (IEA) - Global Energy Review**
  - Electricity generation by source
  - Annual country level data, 2010 - 2024
- **Global Carbon Atlas - Global Carbon Atlas Database**
  - CO2 emissions per capita
  - Cross country reporting
- **World Bank Group - World Development Indicators**
  - Gross Domestic Product (GDP)
  - Population
  - Country level indicators

Data Preparation

Standardization

- Cleaned individual data sets
- Applied ISO-3 Codes

Merge and Filter

- Kept only countries in all 3 sources.
- Removed regions / aggregates
- Merged on ISO-3 and Year

Engineering

- Log transformations
- Lagged features 1 and 2 years
- Final data set - 140 countries

Analysis

- Methodology
- Timeline of Analysis
- EDA Overview

Methodology

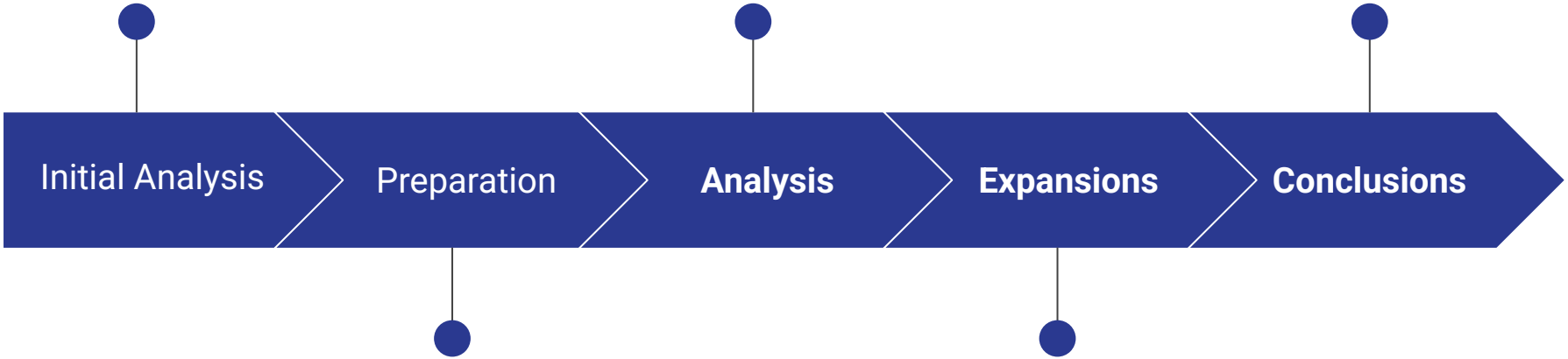
- **Study Design**
  - Comparative observational study
  - Quantitative, correlational analysis
  - Multi-country panel data set
- **Key Variables**
  - CO2 emissions per capita
  - Renewable energy share
  - Gross Domestic Product per Capita
  - Population
  - Country identifiers
    - ISO - 3 Codes
  - Year

Timeline of Analysis

Initial analysis of the data sources used.

Exploratory analysis of relationships between key variables + Testing accuracy.

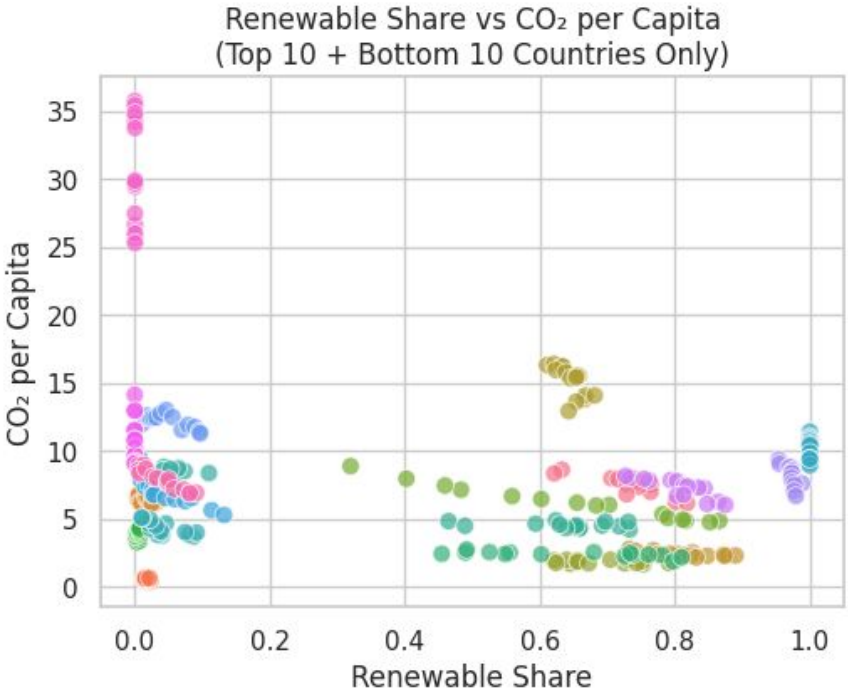
Key takeaways from our analysis

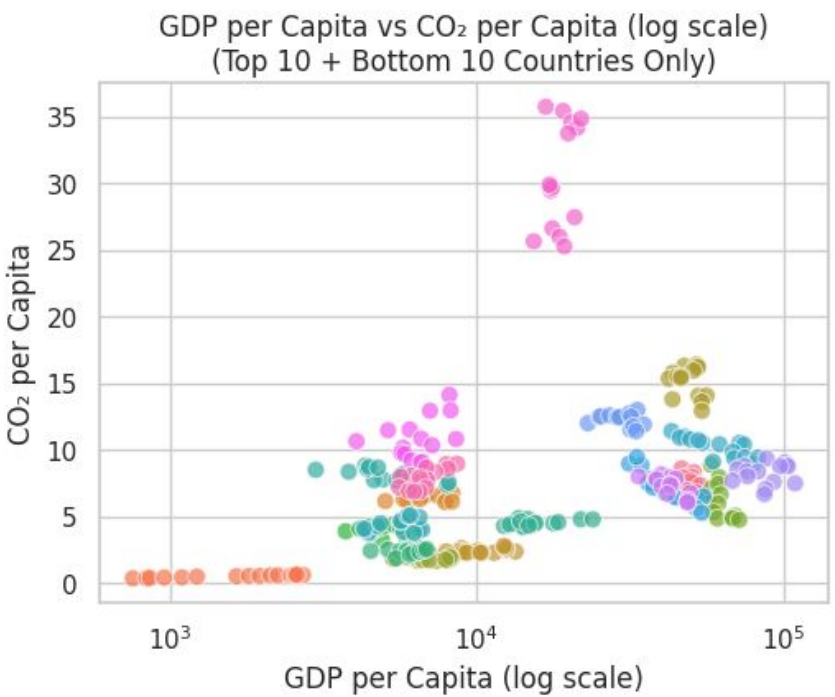


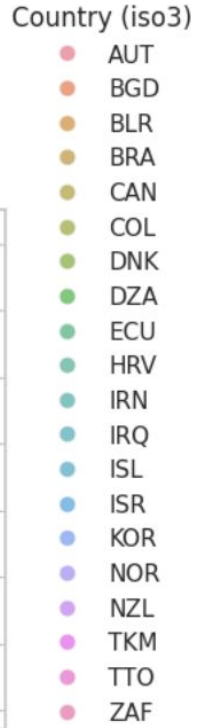
Data Collection & Data Preparation

Explore into the “why” and “how” to possibly explain/expand on observations.

Exploratory Data Analysis







Results

- EDA
- Fixed Effects Panel Regression
- Model Quality & Statistical Results
- Lag Analysis
- Robustness Checks
- Interpretation of Model Performance
- Expansion Analysis

Fixed Effects Panel Regression Results

Do Renewables Reduce CO_2 emissions?

- Country-year panel model (2010-2024)
- Controls for all time-invariant country factors + global year shocks
- Key predictors
  - Renewable energy share
  - GDP per capita
- **Main finding:** Higher renewable share → lower CO_2 emissions per capita
- Effects remain after controlling for economic conditions

Model Quality & Statistical Results

What's our evaluation of the model quality?

- High explanatory power ($R^2 \approx 0.985$ in baseline fixed effects model)
- Strong statistical significance for key variables
- Renewable share coefficient
  - Negative, statistically significant ($p < 0.001$)
  - Suggests measurable emissions reduction
- GDP per capita: positive and significant, consistent with higher energy demand

Lag Analysis

Do Renewables Reduce Emissions Immediately or Over Time?

- Tested 1-year and 2-year lags of renewable share
- **Current-year coefficient:** large, negative significant
- **1-year lag:** smaller but still negative and significant
- **2-year lag:** negative but not statistically significant
- Interpretation: strongest impact occurs in year of adoption and the following year

Robustness Checks

Do the Results Hold Under Different Assumptions?

- Removed outlier countries (very high/low emissions or renewable share)
- Re-estimated the lag model
- **Effects remain negative and stable**
  - Signs of coefficients unchanged
  - Magnitudes highly similar
  - Main conclusions remain intact
- Confirms findings are not driven by a few extreme cases
- Strengthens confidence in overall conclusions

Interpretation of Model Performance

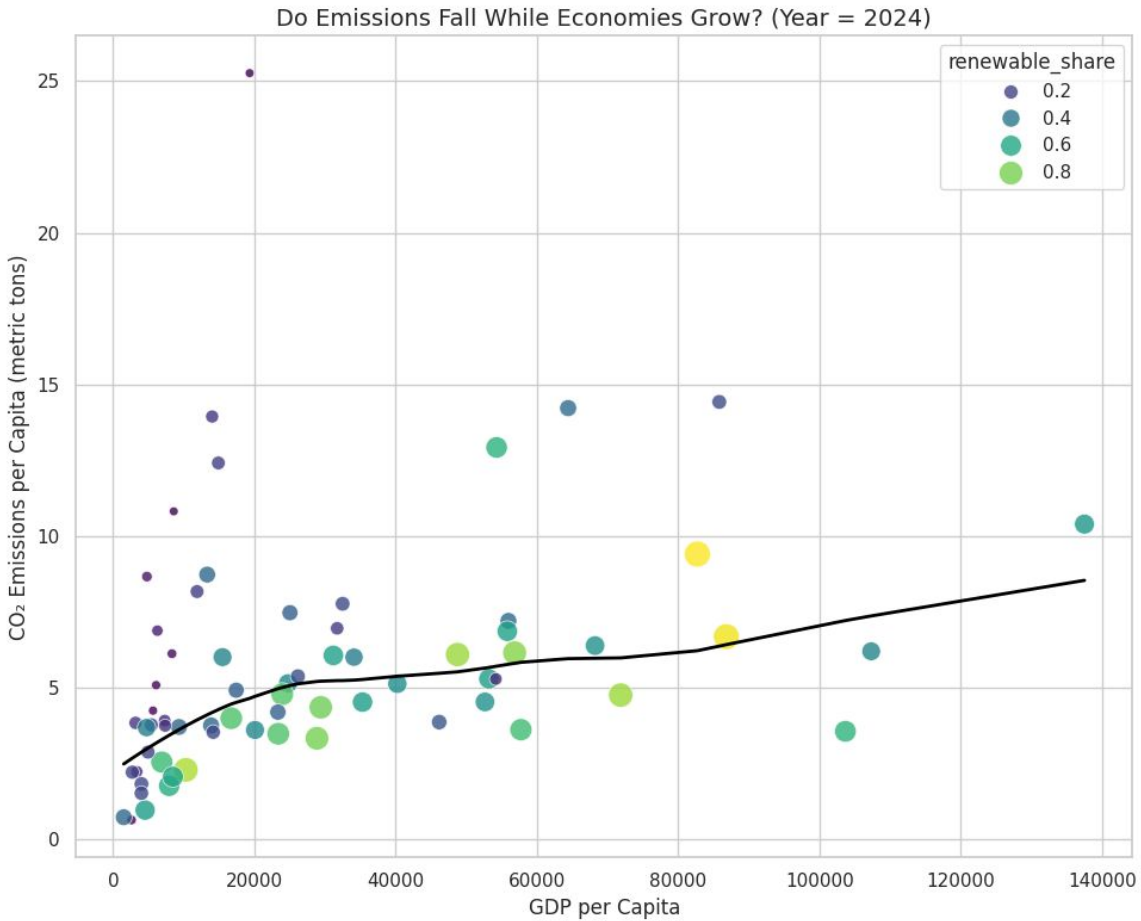
How Well Do the Models Answer the Research Question?

- Clear and consistent evidence across all models
- Renewable share increase → lower CO_2 per capita
- Effect appears to be both immediate and with a short-term lag
- Supported by statistical significance and robustness
- Models effectively answer the central research question

Expansions

Is it possible for CO_2 emissions to fall even when economies grow?

- Higher renewable share → generally lower CO_2 per capita
- Emissions can flatten or decline at mid/high GDP levels
- Economic growth **does not automatically mean higher emissions**
- Clean energy makes “decoupling” growth from emissions possible



Conclusion

- Key Takeaways

Key Takeaways

- Higher renewable shares → lower CO_2 emissions per capita
- Emissions can decline **even with economic growth**
- **Lag analysis** shows reductions persist over future years
- **“Decoupling”** the relationship: Renewable energy breaks the historic GDP–emissions link
- Clean energy expansion supports sustainable development