PROJECT PROBLEM STATEMENT

October 15, 2025

PREPARED FOR:
SEATTLE UNIVERSITY,
DATA 5100 25FQ
Foundations of Data Science

PREPARED BY:
Alyssa Zukas – alyzukas@seattu.edu
Ben Jessop – bjessop@seattleu.edu
Travis St Peter – tstpeter@seattleu.edu

This document outlines a project problem statement proposal that formulates the question to be answered, data sources utilized to research into the problem being researched, demonstrates our teams' ability to work with the data, outlines next steps, and calls out any risks to successful completion.

PROBLEM STATEMENT

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

Global carbon dioxide (CO_2) emissions have continued to rise despite record levels of investment in renewable energy technologies. According to the International Energy Agency (IEA, 2025), energy-related CO_2 emissions reached an all-time high in 2024, driven by growth in total energy demand. Previous research shows that renewable energy expansion can help reduce emissions intensity, but results vary by region and economic structure (Wang & Zhao, 2023; Ozturk & Acaravci, 2013). Other studies note that investment alone may not lead to immediate reductions, as infrastructure transitions and fossil fuel dependencies can delay impacts (Bölük & Mert, 2014).

However, there remains limited research examining the relationship between share of renewable energy investments and national levels of carbon emissions. This project aims to address that gap using cross-country data from 2000-2023 to explore the relationship between renewable energy shares and CO_2 outcomes. We will investigate whether countries with a higher share of renewable energy in their electricity mix emit less carbon overall or per capita. This question matters because decarbonizing national energy systems is central to limiting global warming. Renewable deployment is expanding rapidly, but we anticipate the extent to which it translates into measurable emission reductions will vary across regions. Quantifying this relationship provides evidence of the real-world effectiveness of renewable energy transitions.

DATA SOURCES

The data sources are not finalized, but we expect to draw on publicly available, international datasets including:

1. **Global Carbon Atlas**. (2025). Global Carbon Atlas Database: CO_2 Emissions by Country. https://globalcarbonatlas.org: Provides annual country-level carbon dioxide emissions (total, per capita, and per GDP) spanning several decades.
2. **International Energy Agency (IEA)**. (2025). Global Energy Review 2025: CO_2 Emissions. IEA. https://www.iea.org/reports/global-energy-review-2025: IEA's up-to-date global statistics on energy-related CO_2, trends, and sectoral breakdowns.
3. **World Bank**. (2025). World Development Indicators (WDI). The World Bank Group. https://databank.worldbank.org/source/world-development-indicators: Supplies complementary data on renewable electricity share (% of total electricity generation), GDP, and population.
4. **Energy Institute**. (2025). Statistical Review of World Energy 2025. Energy Institute. https://www.energyinst.org/statistical-review: Provided by the Energy Institute, this can offer us some insight into the demand for energy supply by fuel, region, etc.
5. **National Renewable Energy Laboratory (NREL)**. (2023). 100% Clean Electricity by 2035: Analysis and Modeling. U.S. Department of Energy. https://www.nrel.gov/analysis/100-percent-clean-electricity-by-2035-study.html: Provides scenario analysis for decarbonizing electricity supply.

We anticipate that these sources, in addition to others, will be sufficient for the analysis because they provide consistent data for many countries over several decades. Preprocessing will still be required—including merging datasets by year and country; handling missing values; and normalizing variable scales e.g., converting emissions to metric tons per capita. All the data sources we have identified so far are CSV files available via direct download.

ANALYTICAL APPROACH & TECHNIQUES

We will conduct a comparative observational study using quantitative, correlational analysis to examine the relationship between carbon emissions and the share of renewable energy in national electricity generation. Our analysis will use a country-year panel dataset, containing annual observations from multiple countries, to examine how renewable energy shares and emissions change over time.

Features will include renewable energy share, total CO_2 emissions, CO_2 per capita, and CO_2 per GDP, GDP per capita, population, and regional identifiers.

Initial analysis will involve descriptive statistics and visualization—scatter plots, time trends, and correlation matrices—to explore global patterns. For inference, we will likely use a panel regression model to estimate the association between renewable share and emissions, controlling GDP and time effects.

Further analyses may include clustering countries by energy mix and emissions trajectory, or changepoint analysis to identify when emissions began to decline following renewable adoption (since we anticipate there may be a time delay).

SOLUTIONS TECHNOLOGY

We will use Python as the primary language, specifically: pandas and NumPy for data wrangling and merging; matplotlib and seaborn for visualization; stats models and scikit-learn for regression and exploratory modeling; and Jupyter Notebook for reproducible analysis and documentation.

We do not expect the final dataset to exceed 10,000 rows (200 countries x 40 years would result in 8,000 rows). Hence, the dataset will easily be manageable on our personal computers.

CHALLENGES

One of the main challenges will be interpreting the relationship between renewable energy use and carbon emissions appropriately. While we expect to identify statistical associations, determining whether these reflect causal results will be more complex, as emissions depend on many interrelated factors—such as economic growth, energy demand, and national policy interventions. To address this, we will include relevant control variables and interpret our findings cautiously, focusing on observed relationships rather than definite cause-and-effect conclusions.

A secondary challenge will involve handling missing or inconsistent data, particularly for developing countries. This will require thoughtful preprocessing, including decisions about whether and how to impute missing values. These choices could influence the robustness of our conclusions, so we will conduct sensitivity checks to ensure the results are reliable.

Finally, interpreting results responsibly will be essential. Even statistically significant findings may be shaped by factors outside our dataset, and visualizations or results can be easily misinterpreted if not contextualized. Careful framing and transparent discussion of assumptions will therefore be an essential part of the project.

RELEVANT CITATIONS

When studying renewable energy and carbon emissions, researchers frequently mention total energy supply (TES) or total primary energy supply (TPES) because it provides context for how big a country's renewable energy share actually is.

Sources that may be used for reference during our project are as follows:

1. **Intergovernmental Panel on Climate Change (IPCC)**. (2011). Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change (SRREN). Cambridge University Press. https://www.ipcc.ch/report/renewable-energy-sources-and-climate-change-mitigation/This is something we can investigate before and after we analyze the data to guide us through foundational discussion of renewables, mitigation potential, and challenges.
2. **Wang, Q., & Zhao, M**. (2023). Can renewable energy investment reduce carbon dioxide emissions? Evidence from scale and structure. Energy Economics, 112, 106215. https://www.sciencedirect.com/science/article/abs/pii/S0140988322003334A theoretical paper that addresses one of our core questions by linking the renewable energy investment to CO_2 emissions while decomposing effects into scale, technique, and structure.
3. **Liu, Y., Zhang, H., & Li, X**. (2023). The effect of clean energy investment on CO_2 emissions. Energy Economics, 123, 107048. https://www.sciencedirect.com/science/article/pii/S014098832300498XA recent study published by Science Diet, diving into quantifying the elasticity between clean energy investment and emissions.
4. **Chen, Z., Li, F., & Wu, S**. (2022). The effect of renewable energy on carbon emissions through globalization. Frontiers in Environmental Science, 10, 960795. https://www.frontiersin.org/articles/10.3389/fenvs.2022.960795/fullPublished by PMC Central, this report assesses direct and indirect (mediated) effects of renewable energy on emissions
5. **Bölük, G., & Mert, M**. (2014). Fossil and renewable energy consumption, economic growth, and carbon emissions: An empirical analysis. Energy Policy, 74, 471–479. https://ideas.repec.org/a/eee/energy/v74y2014icp439-446.htmlProvides us with panel data for the period of 1990-2008.

6. **Ozturk, I., & Acaravci, A**. (2013). The long-run and causal analysis of energy, growth, openness, and financial development on carbon emissions in Turkey. Energy Economics, 36, 262–267. https://econpapers.repec.org/RePEc:eee:eneeco:v:36:y:2013:i:c:p:262-267

Note: Not all citations will be referenced in our final report, and there may be additional sources identified throughout the process.

GROUP DYNAMICS

- Communication: We are communicating through Microsoft Teams, chat link: https://teams.microsoft.com/v2/
- Sharing data and code: SharePoint Website (file housing, links to GitHub and Jupyter Notebook, etc.): https://redhawks.sharepoint.com/sites/data5100groupproject?spStartSource=spappbar
- Scheduled syncs: Ad hoc, due to complicated schedules.
- Next steps: finalize data sources (Travis); merge datasets (Alyssa); handle missing data (Ben).