Journal of Research High School
Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

JRHS Outstanding Research Paper Award

Analyzing U.S. State CO_2 Emission Variability (2010–2023): Regional and Population Differences Using Multivariate Modeling

Yerin Choi[1] *

[1]North Gwinnett High School, Suwanee, GA, USA
*Corresponding Author: irenechoi55@gmail.com

Advisor: Dr. Jangho Lee, jlee25@saintpeters.edu

Received March 27, 2026; Revised June 5, 2026; Accepted June 19, 2026

Abstract

This study examines the drivers of state-level CO_2 emissions across the United States from 2010 to 2023 using a multivariate and multi-scale analytical framework. Previous studies have often focused on national trends or individual emissions drivers, with less emphasis on regional and population-based differences across states. Addressing this gap is important for identifying geographic variability in emissions dynamics and in informing targeted policy approaches. Annual data on CO_2 emissions, conventional and renewable electricity generation, gross domestic product (GDP), population, temperature, and precipitation were collected for all U.S. states except Hawaii and Mississippi, due to insufficient data, from federal data sources. To capture geographic and demographic heterogeneity, states were further grouped by region and by population percentile. Temporal trend analysis, descriptive statistics, ANOVA, Pearson correlation, and multiple linear regression were conducted to evaluate patterns and relationships among variables. Results showed that national CO_2 emissions declined over the study period, similar to decreases in conventional electricity generation, while renewable generation and GDP increased. Pearson correlation analysis revealed that CO_2 emissions were most strongly and consistently associated with conventional energy generation across nationwide, regional, and population-based groupings, though the strength and direction of relationships involving renewable energy generation, GDP, and climate variables varied across subgroups. Multiple linear regression further showed that GDP was the strongest predictor in the nationwide model, while conventional energy generation remained the most consistent and statistically significant predictor across regional and population-based models. Climate variables showed less consistent effects. Overall, the study highlights the importance of multivariate, disaggregated analysis for understanding geographic and population-based heterogeneity in emissions dynamics.

Keywords: CO_2, Statistical modelling, Region, Population, Energy generation, GDP, Climate factors

1. Introduction

Greenhouse gases (GHGs) are atmospheric gases that absorb infrared radiation and affect the Earth's energy balance (Forster et al., 2007). Through this physical process, commonly referred to as the greenhouse effect, heat that would otherwise escape into space is retained within the climate system and thus contributing to warming at the Earth's surface. Main types of greenhouse gases include carbon dioxide (CO_2), methane (CH4) and nitrous oxide (N2O), and fluorinated gases (IPCC, 2021). Although it is an essential process in maintaining Earth's temperatures and supporting daily life, human activities have substantially increased atmospheric concentrations of greenhouse gases (NASA, 2024). These increases have led to global warming and climate change, manifested in rising global temperatures and widespread ecosystem disruption.

Among all greenhouse gases, CO_2 is the most significant contributor to anthropogenic climate change due to its high emission volume and long lifetime. Consistent with this role, atmospheric carbon dioxide concentrations continue

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

to rise at an accelerating rate. CO_2 emissions arise from a complex interaction of various sources. In the United States, national emissions trends are often discussed in aggregate, resulting in substantial variation across states due to differences in economic activity, population growth, energy systems, and climatic conditions (Auffhammer & Mansur, 2014). Thus, understanding the drivers of this variation is critical for designing effective and equitable climate and energy policy.

One strand of literature focused on renewable energy production as a mechanism for long-run emissions mitigation. Payne et al., (2024) examined whether per-capita renewable energy production across U.S. states converges over time in the absence of a unified national energy policy. Their study found limited convergence and evidence of persistent divergence among states. While this work did not directly model CO_2 emissions, it framed renewable energy expansion as a central long-run mechanism for reducing GHG emissions and challenged the assumption that states follow a common environmental trajectory (Payne et al., 2024).

At a broader scale, McCurdy & Rhodes, 2023 conducted a global scoping review of academic studies published between 2010 and 2019 to identify statistically significant drivers of GHG emissions. Across countries, economic activity (GDP) and energy consumption emerge as dominant predictors of emissions, though the authors also highlighted the growing importance of energy source substitution and technological innovation. (McCurdy & Rhodes, 2023).

Furthermore, empirical studies using multivariate frameworks further reinforced the relationship between economic activity, energy generation, and emissions. Gyamerah and Gil-Alana, 2023 analyzed the casual relationship between electricity consumption, economic growth, and CO_2 emissions in Western and Central Africa over a long historical period. Their findings indicated that rising electricity demand following economic growth contributes significantly to higher emissions, highlighting the policy challenge of balancing development goals with environmental sustainability. While their study focused on developing regions and long-run causality, it still reinforced the importance of modeling emissions as the outcome of multiple interactive forces rather than a single driver (Gyamerah & Gil-Alana, 2023).

The expansion of renewable energy is widely viewed as a key long-run pathway for emissions mitigation (Stern, 2011). Yet, uneven adoption across U.S. states suggests that structural and policy differences can lead states to transition to clean energy at different rates and in different ways (Payne et al., 2024). Such variation suggested that emissions dynamics are shaped by spatial heterogeneity and structural differences (Guo and Yu, 2024). This is particularly relevant in the United States, where states differ substantially in energy resource composition, economic structure, population scale, and climatic conditions. These differences underscore the importance of state-level analysis rather than reliance on national aggregates.

Despite the broad agreement that emissions are shaped by multiple variables including economic growth, energy consumption, and structural differences, gaps remain in the existing literature. Many U.S.-focused studies examine renewable energy production without incorporating broader structural or climate-related determinants (Wang et al., 2024). Moreover, climate variables such as temperature, despite their influence on electricity demand, are rarely considered into state-level emissions models (Auffhammer & Mansur, 2014). Additionally, only a few studies jointly examined economic, energy generation, and climate factors within a single empirical framework over the recent period (Liu et al., 2022).

This study seeks to address these gaps by analyzing CO_2 emissions across U.S. states from 2010 to 2023 using a multivariate framework. Specifically, it examines how economic activity (GDP), energy generation (renewable and conventional electricity), and climate conditions (temperature and precipitation) contribute to state-level emissions variability over time. The analysis additionally incorporates regional comparisons and population-based groupings. Regional analysis allows the study to capture climatic and energy-system differences across U.S. regions, where variations in temperature, resource availability, and electricity demand/generation may influence emission outcomes. Population-based comparisons further account for the differences in scale and intensity of human activity. States with larger populations tend to exhibit higher levels of economic output and energy demand, which may increase emission levels. By estimating the direction and magnitude of these relationships, this research complements prior convergence, behavioral, and international studies while providing a more integrated, quantitative assessment of recent U.S. state-level emissions dynamics.

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

2. Materials and Methods

2.1 Data Collection

To examine relationships between CO_2 emissions and energy generation, economic, demographic, and climatic variables, data were collected from several principal sources at the U.S. state level for the period 2010-2023. State-level CO_2 emissions data were taken from the U.S. Energy Information Administration, which provides annual estimates of CO_2 emissions estimates across U.S. states. Hawaii and Mississippi were excluded due to insufficient publicly available energy and emissions data across the study period. Specifically, the data were retrieved from the State Energy Data System by navigating to the data files module and selecting the dataset titled "Carbon dioxide (CO_2) emissions from energy consumption," that reports annual emissions by fuel type (EIA, 2025). Within the dataset, Total energy-related CO_2 emissions (TETCE) was used in order to include comprehensive emissions across all energy sources. The same source was also used to collect electricity generation data from Monthly Energy Review, Table 7.2 ("Total Electricity Net Generation"), September 2025 edition (EIA, 2025). Electricity generation was subsequently classified into conventional and renewable sources. Following fuel classifications by the U.S. Energy Information Administration, conventional generation includes coal, natural gas, and nuclear energy, while renewable energy includes geothermal, hydroelectric, solar, and wind sources. Furthermore, economic data, specifically gross domestic product (GDP), were obtained from the U.S. Bureau of Economic Analysis in the Regional Economic Accounts under Gross domestic product module (BEA, 2025). State-level population data were collected from the U.S. Census Bureau Population Estimates Program. Population estimates for 2010-2019 were taken from the 2019 vintage dataset, while estimates for 2020-2023 were taken from the 2023 vintage dataset. Lastly, climate data were obtained from the National Centers for Environmental Information Climate at a Glance database. Annual temperature and precipitation for each state were calculated using statewide time-series data with a 12-month time scale (NCEI, 2025).

2.2 Data Processing and Group Classification

First, the data were organized by geographic region and population size to examine regional- and population-based differences in the influence of variables to CO_2 emissions. As shown in Supplementary Table S1, states were grouped into four regions based on the U.S. Census regional classifications, with the Census Northeast region referred to as "East" for analytical clarity. The four regions used were West, East, Midwest, and South (U.S. Census Bureau, 2021).

Moreover, to account for differences in population scale across states, population was further used to categorize states into four population percentile groups. Table S2 portrays that the 25th percentile group corresponds to states with average populations less than or equal to 1,886,126, the 50th percentile group includes states with populations greater than 1,886,126 and less than or equal to 4,766,790, the 75th percentile group includes states with populations greater than 4,766,790 and less than or equal to 7,598,968 and the 100th percentile group includes states with populations exceeding 7,598,968.

2.3 Temporal and Statistical Analyses

To evaluate temporal trends, time-dependent changes were examined using state-averaged values, regional and population-based averaged values. For statistical analyses, R software was utilized, a widely adopted software environment for statistical computing and data analysis. (Ihaka & Gentlemen, 1996). Descriptive statistics, including mean, standard deviation, minimum, and maximum values, were calculated for all states, each regional and population group to compare differences among these categories. One-way analysis of variance (ANOVA) was conducted separately for regional groups and population-based groups to evaluate statistically significant differences.

Pearson correlation analysis was conducted to assess the strength and direction of linear associations between pairs of variables of interest. The degree of association was quantified using Pearson's correlation coefficient (1), which ranges from -1, indicating perfect negative linear relationship, to +1, indicating perfect positive linear

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

relationship. Statistical significance was evaluated by testing the null hypothesis that the population correlation coefficient is equal to zero (Bewick et al., 2003; Hazra & Gogtay, 2016).

To evaluate the independent effects of multiple explanatory variables, including conventional energy generation, renewable energy generation, GDP, population, average temperature, and precipitation, on CO_2 emissions, multiple linear regression (MLR) analysis was conducted. . Multiple Linear Regression estimates regression coefficients by minimizing the sum of squared residuals, enabling the independent effect of each explanatory variable to be evaluated while holding all others constant (Pal & Bharati, 2019). Model assumptions, such as linearity, independence, homoscedasticity, normality of residuals, and multicollinearity among predictors, were assessed using standard regression diagnostics. Statistical significance for correlation coefficients and regression estimates was evaluated at significance level of 0.05. The general form of the multiple linear regression model used in this study can be expressed as:

$$CO_2=\beta_0 + \beta_1 CE + \beta_2 RE + \beta_3 GDP + \beta_4 TEMP + \beta_5 PREC \tag{1}$$

Where CO_2 represents the CO_2 emissions per capita, CE represents conventional energy generation per capita, RE represents renewable energy generation per capita, GDP represents gross domestic product per capita, TEMP represents average temperature, and PREC represents precipitation.

3. Results

3.1 Temporal Trends in CO_2 Emissions and Related Variables (2010–2023)

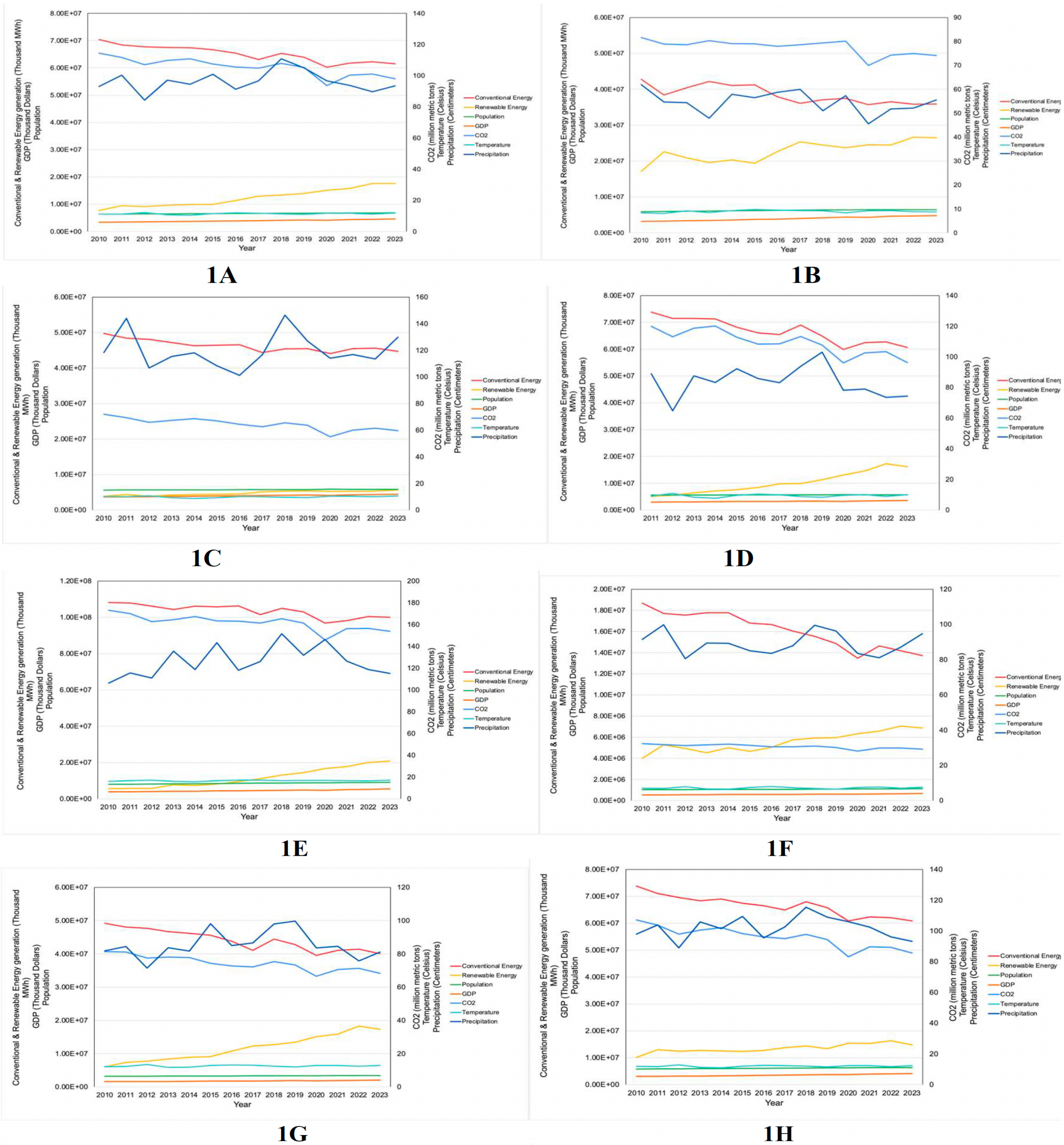
At the national level, time-series analysis revealed long-term trends in CO_2 emissions, energy generation, and economic activity from 2010 to 2023 (Figure 1A). National CO_2 emissions declined over the study period. Conventional energy generation also decreased over time by approximately 12.72%. In contrast, renewable energy generation increased, rising by approximately 128%. Gross domestic product (GDP) also increased with an overall growth of approximately 35.18%. Population levels and climatic variables, including temperature and precipitation, showed limited variation over time.

Across Regions, figures 1B-1E present temporal trends in CO_2 emissions and related variables. In the West, CO_2 emissions declined slightly over time by 9.20%, while conventional energy generation decreased more noticeably by 16.07%. Both renewable energy generation and GDP increased steadily throughout the period.

Notably, CO_2 emissions, conventional energy generation, GDP, and precipitation exhibited similar trend patterns around 2020, with a sudden drop in values. In the East, CO_2 emissions and conventional energy generation both decreased over time. Renewable energy generation and GDP increased, while population remained largely unchanged. In the Midwest, CO_2 emissions and conventional energy generation both showed declining trends, with a decrease around 2020. Renewable energy generation increased overall, reaching a peak around 2022 before exhibiting a slight decline. GDP showed an upward trend with a temporary decrease around 2020. In the South, CO_2 emissions and conventional energy generation followed similar declining trends, again with a drop around 2020. Population increased over time and GDP showed an overall upward pattern.

Figures 1F-1I display time-series trends in CO_2 emissions and related variables for each population group. In the 25th percentile group, CO_2 emissions and conventional energy generation both decreased over time, including a marked decrease in 2020 followed by a rebound in 2021. Renewable energy generation increased over time, reaching peak levels around 2022. GDP increased with a slight decline around 2020, while population showed an upward trend. In the 50th percentile group, CO_2 emissions and conventional energy generation also declined overall, with minor increases around 2018 followed by further decreases. Renewable energy generation increased by approximately 189%, also peaking around 2022. GDP showed a slight upward trend, and population remained largely stable. In the 75th percentile group, CO_2 emissions and conventional energy generation both declined over time, by 20.09% and 17.50%, respectively. In the 100th percentile group, CO_2 emissions and conventional energy generation again displayed similar overall patterns, though conventional generation included larger year-to-year variability. Renewable energy generation increased markedly, with 2015 marking the beginning of a continuous growth.

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]



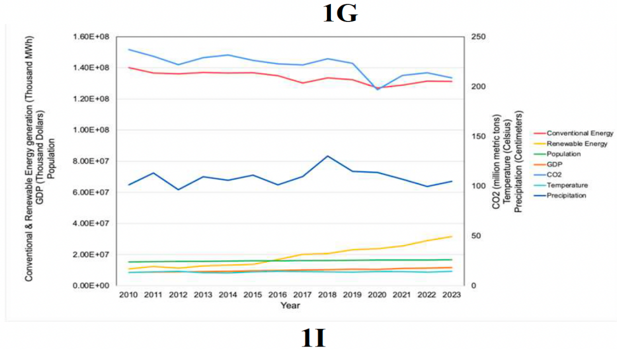


Figure 1. Time-Series Trends Across Nationwide, Regional, and Population Percentile Groups (2010–2023). Energy generation, GDP, and population are shown on the left y-axis, while CO_2 emissions, temperature, and precipitation are shown on the right y-axis. Panels represent (A) Nationwide, (B) West, (C) East, (D) Midwest, (E) South, (F) 25th Percentile, (G) 50th percentile, (H) 75th percentile, and (I) 100th percentile.

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

3.2 Descriptive Statistics of Model Variables

Descriptive statistics in Table 1 reveal regional differences in emissions, energy generation, demographics, and climate. The West (mean = 2.60 × 10-5) and Midwest (2.46 × 10-5) exhibit the highest mean CO_2 emissions, whereas the East has the lowest mean emissions (1.10 × 10-5). Conventional energy generation is highest in the South (mean = 16.08), exceeding the national mean (12.63), followed by the Midwest (13.30), the West (13.18). In contrast, the East exhibits lower conventional energy generation (7.22). Renewable energy generation is highest in the West (mean = 5.61), compared with the national mean (3.06), and lowest in the East (1.00). The mean population size is the largest in the South (8.61 × 106) and smallest in the Midwest (5.67 × 106). The East records the highest mean of GDP per capita (6.49 × 10-2), while the South exhibits the lowest (4.99 × 10-2). Climatic variables differed across regions. The South is the warmest region (mean temperature = 16.94), exceeding the national mean (11.42), whereas the West (8.99) and Midwest (9.43) remain below the national mean. Precipitation is lowest in the West (54.68), substantially below the national mean of 96.24, and highest in the South (126.08), followed by the East (119.77). Population-based descriptive statistics show differences in emissions, energy generation, and climate. The highest mean observed in the 25th percentile group (mean = 3.28 × 10-5) and the lowest in the 100th percentile group (1.41 × 10-5). Additionally, the highest value in the 25th percentile group (mean = 16.45) and the lowest in the 100th percentile group (9.42). Renewable energy generation was the highest among smaller population states (mean = 5.77 in the 25th percentile group) and lowest among the largest states (mean = 0.87 in the 100th percentile group). Population variability is highest in the 100th percentile group (standard deviation = 8.92 × 106), whereas GDP variability was limited across groups.

Table 1. Regional and Population-based Descriptive Statistics for CO_2 Emissions, Energy Generation, Economic Activity, and Climate Variables (2010–2023). CO_2 Emissions, Conventional Energy Generation, Renewable Energy Generation, and GDP are reported on a per capita basis. Note: *** = $p < 0.001$; Stdev: Standard Deviation

		All States	West	East	Midwest	South	25th Percentile	50th Percentile	75th Percentile	100th Percentile
CO_2 (Per Capita), Million Metric Tons ***	Mean	2.12×10^{-5}	2.60×10^{-5}	1.10×10^{-5}	2.46×10^{-5}	2.23×10^{-5}	3.28×10^{-5}	2.20×10^{-5}	1.59×10^{-5}	1.41×10^{-5}
	Stdev	1.78×10^{-5}	2.70×10^{-5}	2.74×10^{-6}	1.60×10^{-5}	1.15×10^{-5}	3.05×10^{-5}	8.63×10^{-6}	5.60×10^{-6}	4.34×10^{-6}
	Min	7.10×10^{-6}	7.58×10^{-6}	7.10×10^{-6}	1.31×10^{-5}	9.59×10^{-6}	8.34×10^{-6}	8.67×10^{-6}	7.45×10^{-6}	7.10×10^{-6}
	Max	1.19×10^{-4}	1.19×10^{-4}	2.01×10^{-5}	8.03×10^{-5}	5.43×10^{-5}	1.19×10^{-4}	4.55×10^{-5}	3.36×10^{-5}	2.48×10^{-5}
Conventional Energy Generation, Thousand MWh (Per Capita)***	Mean	12.63	13.18	7.22	13.30	16.08	16.45	13.29	11.36	9.42
	Stdev	11.29	17.59	4.30	7.89	8.75	19.99	4.53	6.62	4.64
	Min	1.95×10^{-3}	7.54×10^{-1}	1.95×10^{-3}	2.66	1.84×10^{-1}	1.95×10^{-3}	3.06	2.03	1.84×10^{-1}
	Max	80.70	80.70	17.80	42.20	42.30	80.70	21.40	30.00	17.80
Renewable Energy Generation, Thousand MWh (Per Capita)***	Mean	3.06	5.61	1.00	4.17	1.43	5.77	3.55	2.06	0.87
	Stdev	4.08	4.45	1.20	5.29	1.71	5.57	3.57	2.99	0.86
	Min	2.84×10^{-3}	4.71×10^{-1}	2.84×10^{-3}	3.94×10^{-2}	1.37×10^{-2}	2.84×10^{-3}	7.31×10^{-2}	1.55×10^{-1}	5.89×10^{-3}
	Max	23.20	18.40	4.80	23.20	9.80	23.20	14.70	14.40	4.88
Population***	Mean	6.63×10^{6}	6.24×10^{6}	5.75×10^{6}	5.67×10^{6}	8.61×10^{6}	1.07×10^{6}	3.31×10^{6}	6.11×10^{6}	1.60×10^{7}
	Stdev	7.26×10^{6}	1.01×10^{7}	5.81×10^{6}	3.90×10^{6}	7.32×10^{6}	3.98×10^{5}	8.23×10^{5}	7.87×10^{5}	8.92×10^{6}
	Min	564,487	564,487	623,657	674,715	1,770,071	564,487	1,829,542	4,635,649	8,023,699
	Max	39,512,220	39,512,223	20,104,710	12,895,129	30,503,301	1,964,726	4,678,135	7,812,880	39,512,223
GDP (Per Capita), Thousand Dollars***	Mean	5.74×10^{-2}	5.75×10^{-2}	6.49×10^{-2}	5.86×10^{-2}	4.99×10^{-2}	5.78×10^{-2}	5.43×10^{-2}	5.74×10^{-2}	6.02×10^{-2}
	Stdev	1.09×10^{-2}	1.13×10^{-2}	1.16×10^{-2}	7.40×10^{-3}	7.00×10^{-3}	1.22×10^{-2}	9.33×10^{-3}	1.10×10^{-2}	9.96×10^{-3}
	Min	3.87×10^{-2}	3.89×10^{-2}	4.44×10^{-2}	4.47×10^{-2}	3.87×10^{-2}	3.87×10^{-2}	3.87×10^{-3}	4.07×10^{-2}	4.45×10^{-2}
	Max	9.16×10^{-2}	8.59×10^{-2}	9.16×10^{-2}	8.16×10^{-2}	7.01×10^{-2}	8.16×10^{-2}	7.83×10^{-2}	8.78×10^{-2}	9.16×10^{-2}
Temperature, Celsius***	Mean	11.42	8.99	9.95	9.43	16.94	7.19	12.61	12.04	13.83
	Stdev	4.68	4.68	2.64	2.86	2.82	3.87	3.33	4.14	4.36
	Min	-4.28	-4.28	4.83	3.33	10.44	-4.28	7.5	3.67	5.5
	Max	23.00	17.22	14.72	14.78	23	14.72	20.89	18.56	23
Precipitation, Centimeters**	Mean	37.89	21.53	47.15	33.03	49.64	89.24	85.37	102.00	108.34
	Stdev	15.12	11.06	6.56	9.86	11.24	35.82	44.33	38.26	29.41
	Min	5.86	5.86	36.73	12.92	14.7	27.84	14.88	16.66	20.14
	Max	68.70	52.87	64.76	55.95	68.67	165.35	170.97	174.42	173.61

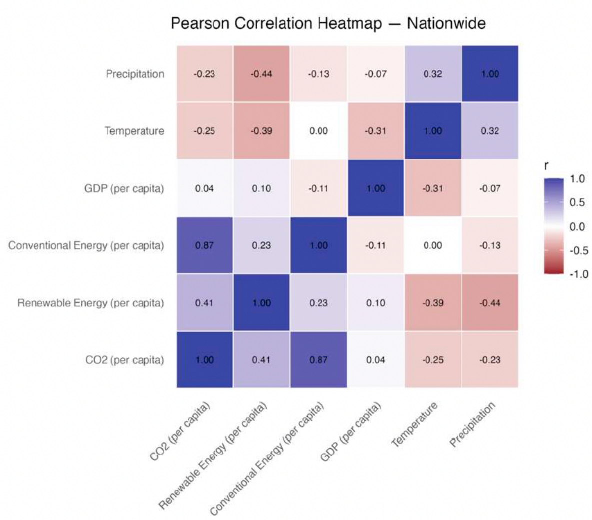
3.3 Pearson Correlations of CO_2 Emissions and Related Variables

The Pearson correlation results show the direction and relative magnitude of relationships between variables.

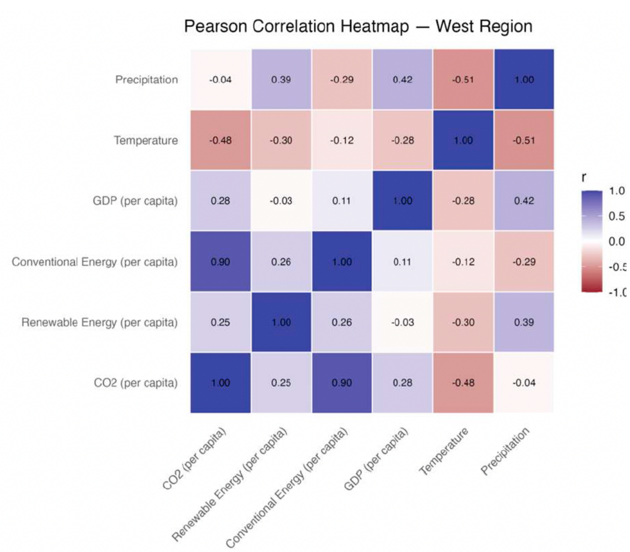
Results presented in Figures 2A-2I show that CO_2 and conventional energy generation exhibit a very strong positive correlation nationwide ($r = 0.87$). These two variables represent the strongest relationship across all groupings. CO_2 and renewable energy generation show a moderate positive correlation nationwide ($r = 0.41$). CO_2 and GDP show almost no correlation nationwide, despite many subgroups showing strong positive or negative relationships. CO_2 and precipitation show a weak negative correlation nationwide, but larger negative correlations are observed in several regions and percentiles. Climate variables, on the other hand, show weak to moderate correlations nationwide.

Regional and population-based differences were also observed. Regionally, the Midwest shows the strongest CO_2 and conventional energy generation correlation ($r = 0.94$). This value is higher than the nationwide value (0.87), and all other regions. The West also shows an extremely strong CO_2 and conventional energy generation correlation ($r = 0.90$). The West shows a stronger negative correlation ($r = -0.48$), between CO_2 and temperature compared to other regions, however. The East has the weakest CO_2 and conventional energy generation correlation among the four regions, and is the only region where renewable energy generation is strongly negatively correlated with temperature ($r = -0.79$). The South shows strong positive correlation between CO_2 and conventional energy generation as well, but very weak correlations involving renewable energy generation, unlike other regions. Renewable energy generation relationships are strongest in the Midwest, and weakest in the South.

Differences across population percentiles were also observed. All population groups show a strong positive correlation between CO_2 and conventional energy generation, with the r values ranging from 0.67 to 0.91. The 25th percentile group shows the strongest CO_2 to conventional energy generation correlation ($r = 0.91$). The group also shows the strongest renewable and precipitation negative correlation. The 50th percentile group is the only group that shows a strong positive correlation between CO_2 and temperature, unlike other population groups ($r = 0.66$). The 75th percentile group showed strong negative correlations between GDP and both CO_2 ($r = -0.59$) and conventional energy generation ($r = -0.83$). The 100th percentile group demonstrated stronger positive correlations between renewable energy generation and both population and GDP than all other groups. Beyond relationships involving CO_2, several strong correlations were observed between other variable pairs. In the East, renewable energy generation and temperature show a strong negative correlation ($r = -0.79$), which is the strongest climate-energy correlation among the regions. In the 25th percentile population group, renewable energy generation and precipitation demonstrate a strong negative correlation as well ($r = -0.84$). GDP and conventional energy generation show a very strong negative correlation in the 75th percentile group ($r = -0.83$). However, none of these relationships reach the $|r| \geq 0.90$ threshold (Dohoo et al., 2009). Accordingly, while several strong linear associations exist, they do not meet the level at which multicollinearity would be considered a major concern in this analysis.

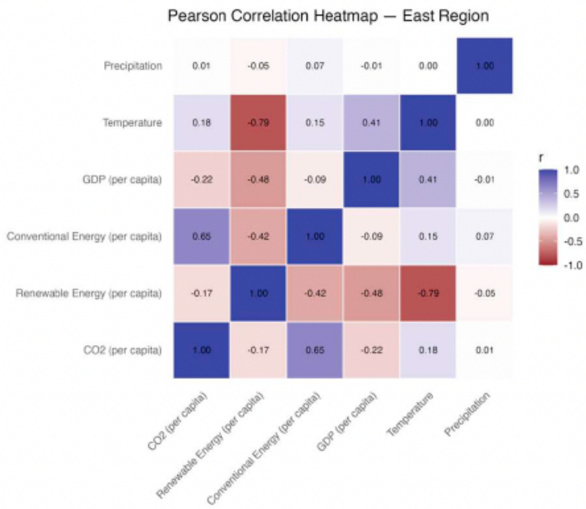


2A

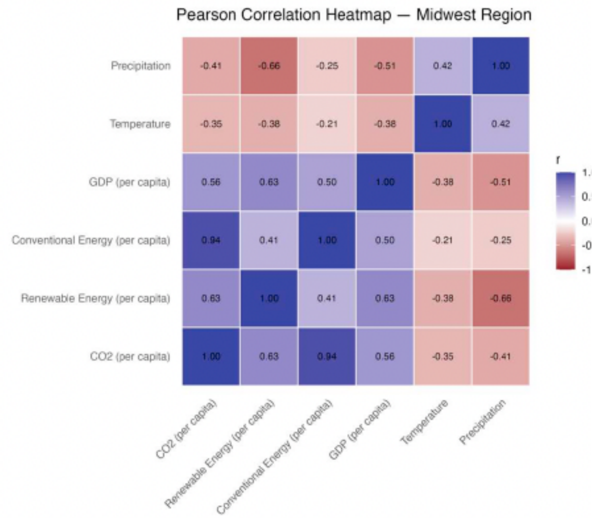


2B

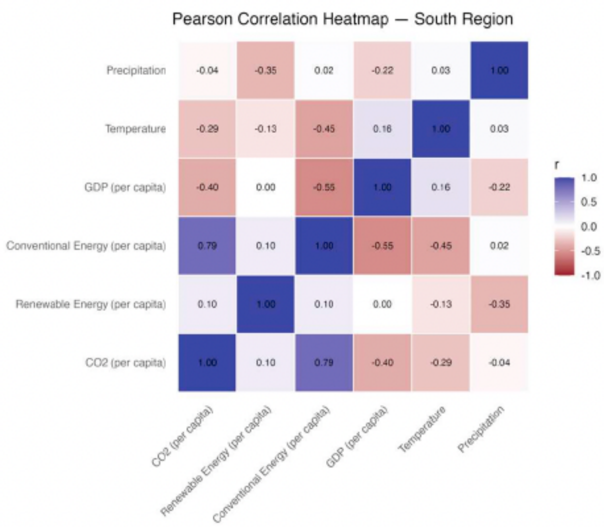
Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]



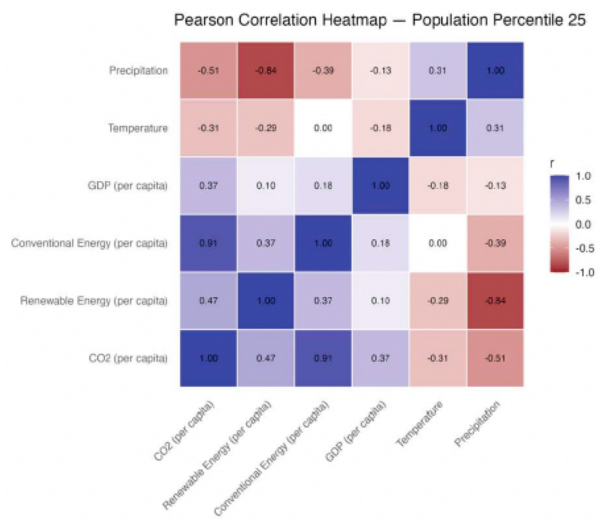
2C



2D

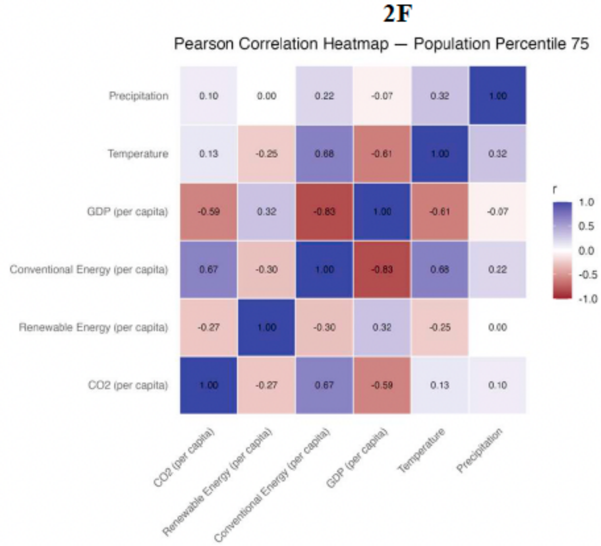


2E



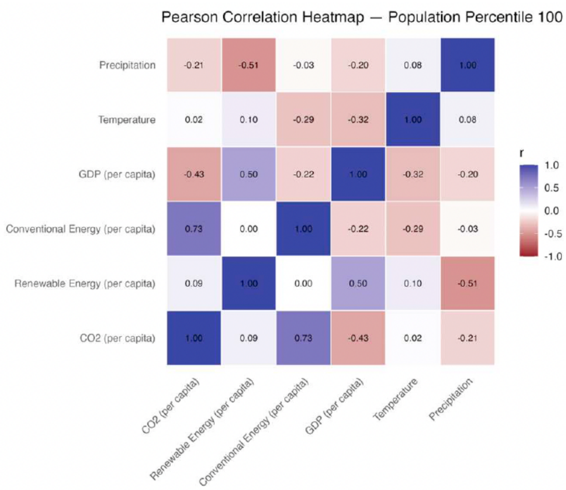
2F

2G



2H

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]



2I

Figure 2. Pearson Correlation Matrices Across Nationwide, Regional, and Population Percentile Groups (2010–2023). Panels represent (A) Nationwide, (B) West, (C) East, (D) Midwest, (E) South, (F) 25th Percentile, (G) 50th percentile, (H) 75th percentile, and (I) 100th percentile.

3.4 Multivariate Regression Analysis of CO_2 Emission Drivers

As shown in Table 2 (also shown in supplementary Table S3), GDP had the largest coefficient among predictors in the nationwide model ($\beta = 1.13 \times 10^{-4}$). Conventional energy generation was also a significant positive predictor (β = 1.33 × 10-6). Renewable energy generation was positive, although its effect was approximately half the magnitude of conventional energy generation. Temperature was a strong negative predictor (β = -3.65 × 10-7). Precipitation, however, was not significant (β = -8.24 × 10-9, $p > 0.05$). In the West, the estimated coefficient for conventional energy generation (1.39 × 10-6) was similar in magnitude to the nationwide effect. The Midwest had the largest conventional energy generation coefficient among the regions, while in the South, conventional energy generation was the only strong and consistent predictor. Notably, renewable energy generation in the West became a significant negative contributor, in the opposite direction from nationwide. Temperature had the largest negative coefficient among regions (β = -1.14 × 10-6). Precipitation was a significant positive contributor, unlike at the national level. In the East, renewable energy generation had a stronger positive effect than conventional energy generation, while GDP was no longer a significant predictor. Temperature also shifted direction, emerging as a significant positive contributor, opposite to the nationwide and Western trends. In the Midwest, renewable energy generation was strongly positive and significant. GDP and temperature remained a significant negative contributor, while precipitation had no meaningful effect. In contrast, the South displayed the least diversified set of predictors, as all renewable energy generation, GDP, temperature, and precipitation were not significant. Across population percentiles, conventional energy generation's magnitude declined as population increased, from the least populated states (1.30 × 10-6) to the most populated (6.30 × 10-7). GDP was a major positive contributor in the 25th percentile group, while temperature and precipitation were both strong negative predictors. At the 50th percentile group, conventional energy generation reached its largest effect (β = 1.87 × 10-6). Renewable energy generation was also a positive and statistically significant predictor (β = 7.81 × 10-7). In the 75th percentile group, renewable energy generation and GDP both were significant negative contributors. Renewable energy generation had the largest positive coefficient in the 100th percentile group. GDP was strongly negative, and precipitation was significant and negative, unlike nationwide. Conventional energy generation remained statistically significant across all models, with p-values consistently below 0.001, including the nationwide ($p < 0.001$) and Midwest ($p < 0.001$) samples. While renewable energy generation and GDP were significant predictors in most contexts, in the West ($p < 0.001$) and Midwest ($p < 0.001$), neither of them reached significance in the South ($p > 0.05$). Temperature effects were highly significant in most regions ($p < 0.01$) but they failed to reach significance in the South and the 100th percentile group ($p > 0.05$). Precipitation was the least consistent factor, appearing as a significant predictor only in specific cases, such as the West ($p < 0.05$) and the 100th percentile group ($p < 0.001$). The magnitude of t-values varies substantially by predictor and group. While most predictors have very large absolute t-values exceeding 2, conventional energy generation exhibits especially large t-values (Magnus, 2019). Both nationwide (52.0) and Midwest (53.8) have high t-values. However, renewable energy generation remains statistically significant in some regions, including the East (t = 4.34, $p < 0.001$) and the Midwest (t = 15.35, $p < 0.001$) but not in others including the South (t = 2.52 × 10-1, $p > 0.05$) and the 25th percentile group (t = 5.14 × 10-1, $p > 0.05$). Climate variables also display mixed patterns, as temperature yields statistically significant effects in several models while remaining insignificant in others.

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

Table 2. Multiple Linear Regression Coefficients Across Nationwide, Regional, and Population Percentile Groups for CO_2 Emissions, Energy Generation, Economic Activity, and Climate Variables (2010-2023).

	Nationwide	West	East	Midwest	South	25th Percentile	50th Percentile	75th Percentile	100th Percentile
Intercept	1.57×10^{-5}	5.70×10^{-5}	-6.45×10^{-6}	2.35×10^{-5}	-9.07×10^{-6}	4.47×10^{-5}	-2.46×10^{-5}	3.99×10^{-5}	1.70×10^{-5}
Conventional Energy Generation, Thousand MWh (Per capita)	1.33×10^{-6}	1.39×10^{-6}	4.93×10^{-7}	1.71×10^{-6}	1.12×10^{-6}	1.30×10^{-6}	1.87×10^{-6}	7.50×10^{-7}	6.31×10^{-7}
Renewable Energy Generation, Thousand MWh (Per capita)	5.74×10^{-7}	-7.02×10^{-7}	1.16×10^{-6}	9.44×10^{-7}	8.41×10^{-8}	8.31×10^{-8}	7.81×10^{-7}	-1.78×10^{-7}	8.06×10^{-7}
GDP, Thousand Dollars (Per capita)	1.13×10^{-4}	1.04×10^{-4}	-3.03×10^{-5}	-1.95×10^{-4}	7.32×10^{-5}	3.95×10^{-4}	9.34×10^{-5}	-1.16×10^{-4}	-1.59×10^{-4}
Temperature, Celsius	-3.65×10^{-7}	-1.14×10^{-6}	3.01×10^{-7}	-2.89×10^{-7}	1.94×10^{-7}	-1.15×10^{-6}	2.83×10^{-7}	-5.05×10^{-7}	5.48×10^{-8}
Precipitation, Centimeters	-8.24×10^{-9}	1.61×10^{-7}	-6.30×10^{-9}	1.10×10^{-11}	-5.07×10^{-8}	-1.37×10^{-7}	-4.57×10^{-8}	4.06×10^{-8}	-7.35×10^{-8}

4. Discussions

4.1 Structural Changes in Electricity Generation and CO_2 Emissions Trends

The overall observed temporal trends in CO_2 emissions reflect broader changes in energy production over the study period. The decline in CO_2 emissions occurred alongside decreases in conventional energy generation and increases in renewable generation. At the same time, GDP increased substantially over the study period. The simultaneous increase in GDP and decrease in CO_2 emissions suggests a partial decoupling of economic growth from emissions. These findings suggest that economic expansion can occur alongside reductions in carbon intensity. This divergence indicates a structural transition toward lower-carbon electricity energy sources. The U.S. power sector provides supporting evidence for this interpretation, as available data indicate that CO_2 emissions have fallen as the generation mix has moved away from coal toward natural gas and renewables, with renewables contributing to emissions reductions as their share rises (EIA, 2021). These findings are consistent with prior work emphasizing renewable energy expansion as one of the key long-run mechanisms driving emissions mitigation. For example, the findings of Payne et al. (2024) identify renewable expansion as the primary long-run mechanism through which states differ in their decarbonization pathways, a pattern consistent with the emissions trend observed in this study (Payne et al., 2024). By explicitly estimating emissions alongside changes in both conventional and renewable energy generation, this study directly links these structural shifts in energy production to observed emissions declines. The pronounced declines around 2020 across CO_2 emissions, conventional energy generation, and GDP suggest that short-term economic disruptions can temporarily accelerate emissions reductions. However, the subsequent rebound indicates that long-term emissions trends appear to be more strongly associated with structural changes in energy systems rather than by temporary economic shocks. Additionally, a multivariate framework is used, accounting for economic, demographic, and climatic factors. This allows emissions dynamics to be analyzed as the result of multiple interacting forces, rather than a single structural mechanism. These findings align with broader literature showing that economic and energy activities are major determinants of greenhouse gas emissions (McCurdy & Rhodes, 2023). The U.S. emissions decline documented reflects both changes in economic activity and a structural shift in the electricity generation mix. This reinforces the importance of integrated, multivariate frameworks for understanding long-run emissions dynamics.

4.2 Regional and Demographic Heterogeneity in CO_2 Emissions

Regional and population-based differences highlight the importance of spatial and demographic context in shaping emission dynamics across the United States. Across all regions, declines in CO_2 emissions closely mirrored declines in conventional energy generation and increases in renewable generation. However, the magnitude and pace of these changes varied regionally. The West exhibited the highest average renewable energy generation, whereas the East exhibited the lowest. This highlights substantial variation in regional energy generation profiles and may contribute to explain regional differences in emissions patterns. These regional differences are consistent with

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

variations in resource availability, policy environments, and historical energy dependence, as energy transitions often produce uneven outcomes across regions depending on existing structures and policy designs (Carley & Konisky, 2020). Population size further shaped emissions levels and variability. Interestingly, states in the highest population percentile exhibited the lowest mean per-capita CO_2 emissions, while states in the lowest population percentile exhibited the highest. This pattern suggests that population size alone does not determine emissions intensity and that other factors such as energy efficiency, urban infrastructure, and energy composition also influence how population translates into emissions outcomes. Nevertheless, population remains an important factor in emissions dynamics, as previous research has shown that population growth and economic activity can increase overall emissions when fossil fuels constitute a substantial portion of the energy mix (Dong et al., 2018). This observation suggests that renewable energy generation expansion can partially offset the emissions effects typically associated with economic and population growth (Jeon, 2022). Climate variables such as temperature and precipitation remained relatively stable across population groups. The stronger variation observed in energy-related variables across groups suggest that differences in energy structure may play a larger role in shaping emissions patterns than climatic conditions alone. Long-run regional as well as demographic emissions differences were driven more by structural energy and economic factors than by climatic variability.

4.3 Correlation Patterns of Emissions Drivers Across Regions and Population Groups

The Pearson correlation results show that CO_2 emissions are shaped not only by national-level trends but also by distinct regional patterns of energy production. The consistently strong positive correlation between conventional energy generation and CO_2 emissions across nearly all models indicates the dominant role of fossil-based electricity generation in determining emissions variability. The changing role of renewable energy across regions and population groups, shown by changes in the directions of correlations, shows that renewables do not always replace fossil fuels and may instead grow alongside rising energy demand. In several cases, renewable generation increased alongside overall electricity demand rather than directly replacing fossil fuel generation.

Differences in GDP, population, average temperature, and precipitation suggest that they mainly affect emissions through their influence on energy demand and energy consumption, rather than directly causing emissions. These results demonstrate that national-level analysis may obscure important regional variation in emissions drivers. The near-zero nationwide relationship between GDP and CO_2 emissions, despite stronger relationships within several subgroups, highlights how national-level analyses can mask important regional and demographic differences. Differences in Pearson correlation reflect underlying heterogeneity in how energy, economic, and demographic factors interact. U.S. state-level research shows that the relationships among GDP, renewable energy generation, and CO_2 emissions vary substantially across states due to differences in energy mix, industrial structure, and policy context (Gurney et al., 2019). Similarly, state-level emissions are shaped not only by internal economic conditions but also by geographic and structural characteristics, which may help explain why the strength and direction of correlations differ across regions, such as the stronger CO_2–conventional energy generation relationship in the Midwest compared with the East (Burnett et al., 2013).

Studies of carbon emission drivers also indicate population scale increases heterogeneity in emissions outcomes. Densely populated areas show different emission responses than less dense ones, aligning with the observation of different r values across population percentiles (Xu et al., 2025). Urban scaling research further highlights that the relationship between population size, density, and carbon emissions varies across regions and is shaped by urban form including transportation systems and economic structure. These differences help explain why population-based subgroups exhibit distinct correlation patterns (Fragkias et al., 2013). Finally, broader analyses of carbon emission heterogeneity underscore how demographic factors such as population growth and urbanization interact with economic conditions to influence emissions in different regions. The interaction between factors reinforces the importance of considering multiple scales when interpreting statistical relationships (O'Neill et al., 2010).

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

4.4 Multivariate Determinants of CO_2 Emissions

The multiple linear regression results help explain why Pearson correlation results differ across nationwide, regional, and population scales. GDP emerges as the strongest predictor in the nationwide model, highlighting the influence of economic scale on aggregate emissions. However, conventional energy generation remains the most consistent and dominant driver of CO_2 emissions across regional and population models. This is consistent with research showing that conventional energy intensity of electricity systems strongly determines state-level emission outcomes (Davis & Socolow, 2014). In contrast, the effects of renewable energy generation and GDP per capita vary by region and population group, reflecting differences in energy mix and industrial composition across states, in line with research showing that the link between economic activity and emissions varies across space and depends on local structural conditions (Guo & Yu, 2024). Population scale and urbanization also shape emissions responses. Larger, denser areas tend to exhibit different relationships between demographic change, economic activity, and emissions than smaller, less urbanized areas. This aligns with findings that urban form, density, and infrastructure systems affect how demographic and economic change translate into CO_2 emissions (Zhang et al., 2021).

Conventional energy generation is the most consistent driver of emissions across regional and population contexts because it directly reflects dependence on fossil-based energy systems. Notably, conventional energy generation remained statistically significant across all nationwide, regional, and population-based models, whereas the effects of renewable energy generation, GDP, and climate variables varied considerably across contexts. By contrast, renewable energy generation and GDP behave differently due to their interactions with region-specific energy systems and development patterns. These differences help explain why some MLR coefficients diverge from Pearson correlation results, as multiple regression captures independent effects while Pearson reflects associations among correlated variables (Waldorp & Marsman, 2021).

Overall, the results demonstrate that both regional energy structure and population scale shape how economic and climatic factors translate into emissions outcomes. These findings suggest that emissions dynamics should be interpreted across multiple analytical levels rather than relying solely on a single national pattern.

4.5 Limitations

This study has potential limitations. The analysis relied on state-level aggregated data, which may mask important within-state variation in emissions sources, urbanization, activity, and energy consumption patterns. Additionally, the study focused primarily on electricity generation and does not directly account for other major contributors to CO2 emissions, such as transportation and agriculture. Although the multiple linear regression results identified statistical relationships among variables, the results do not directly establish causation. Differences in data availability also led to the exclusion of Hawaii and Mississippi, which may slightly limit the comprehensiveness of nationwide comparisons. Finally, policy variables, technology innovation, and regulatory differences were not directly incorporated into the models despite their potential influence on emissions dynamics.

5. Conclusion

This study provides a multi-scale examination of U.S. CO_2 emissions from 2010 to 2023 utilizing temporal trends, descriptive statistics, Pearson correlation, and multiple linear regression. The analysis evaluates how energy structure, economic activity, demographic scale, and climate factors shape emissions dynamics across nationwide, regional, and population-based contexts. The multiple linear regression results reveal that GDP emerges as the strongest predictor in the nationwide model, underscoring the importance of economic scale in explaining aggregate emissions patterns. The findings indicate that declines in CO_2 emissions over the study period closely paralleled conventional energy generation reductions, highlighting fossil fuel displacement within the electricity sector as the dominant structural mechanism driving emissions reductions. While renewable energy generation expanded substantially, its relationship with emissions varied across regions and population groups, reflecting differences in infrastructure, resource availability, and energy demand. Economic and demographic growth did not produce uniform emissions responses across contexts. This suggests that structural differences in energy systems mediate how GDP and population changes translate into emissions outcomes. Substantial heterogeneity was observed across contexts, demonstrating that national averages often mask meaningful variation in energy dependence and economic structure. Conventional energy generation remained the most consistent predictor across models, whereas GDP dominated the nationwide regression. This indicates that economic scale explains national emissions patterns while energy system structure shapes variation at more localized levels. Differences between Pearson correlation results and multiple linear regression results further underscore the importance of modeling independent effects. Overall, the results reinforce that long-term decarbonization largely reflects structural transitions within energy systems. Mitigating emissions effectively therefore depends less on limiting economic or demographic growth and more on accelerating fossil fuel displacement within regional electricity systems. Understanding emissions behavior therefore requires multivariate approaches that account for geographic and demographic heterogeneity as well as the different roles of economic scale and energy structure in shaping emissions patterns.

Acknowledgement

The author would like to thank Dr. Jangho Lee for his guidance and feedback throughout the research process.

References

Auffhammer, M., & Mansur, E. T. (2014). Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Economics*, 46, 522–530. https://doi.org/10.1016/j.eneco.2014.04.017

U.S. Bureau of Economic Analysis. (n.d.). Regional data download. https://apps.bea.gov/regional/downloadzip.htm

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical care (London, England)*, 7(6), 451–459. https://doi.org/10.1186/cc2401

Bureau, U. C. (2021, December 16). Geographic Terms and Definitions. Census.gov. https://www.census.gov/programs-surveys/popest/about/glossary/geo-terms.html

Burnett, J. C., Bergstrom, J. C., & Dorfman, J. H. (2013). A spatial panel data approach to estimating U.S. state-level energy emissions. *Energy Economics*, 40, 396–404. https://doi.org/10.1016/j.eneco.2013.07.02

Carley, S., Konisky, D.M. The justice and equity implications of the clean energy transition. *Nat Energy* 5, 569–577 (2020). https://doi.org/10.1038/s41560-020-0641-6

U.S. Census Bureau. (2011). nst-est2019-alldata.csv [Data set]. https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/national/totals/nst-est2019-alldata.csv

U.S. Census Bureau. (2020). NST-EST2023-ALLDATA.csv [Data set]. https://www2.census.gov/programs-surveys/popest/datasets/2020-2023/state/totals/NST-EST2023-ALLDATA.csv

Davis, S. J., & Socolow, R. H. (2014). Commitment accounting of CO 2 emissions. *Environmental Research Letters*, 9(8), 084018. https://doi.org/10.1088/1748-9326/9/8/084018

Dohoo, I. R., Martin, W., & Stryhn, H. (2014). Model-building strategies. In Veterinary epidemiologic research (2nd ed.). VER Inc.

Dong, K., et al, (2018). CO2 emissions, economic and population growth, and renewable energy: Empirical evidence across regions. *Energy Economics*, 75, 180–192. https://doi.org/10.1016/j.eneco.2018.08.017

EIA. (2025). U.S. Energy Information Administration (EIA) - Total Energy Monthly Data. Eia.gov. https://www.eia.gov/totalenergy/data/monthly/

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

Forster, P., et al. (2007). Changes in atmospheric constituents and in radiative forcing. In Climate change 2007: The physical science basis (Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change). Intergovernmental Panel on Climate Change https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter2-1.pdf

Fragkias, M., et al., (2013). Does Size Matter? Scaling of CO2 Emissions and U.S. Urban Areas." *PLoS ONE*, vol. 8, no. 6, e64727, https://doi.org/10.1371/journal.pone.0064727

Guo, C., Yu, J. (2024) Determinants and their spatial heterogeneity of carbon emissions in resource-based cities, China. *Scientific Reports* 14, 5894, https://doi.org/10.1038/s41598-024-56434-2

Gurney, K., et al. (2019). Comparison of Global Downscaled Versus Bottom-Up Fossil Fuel CO 2 Emissions at the Urban Scale in Four U.S. Urban Areas. *Journal of Geophysical Research: Atmospheres*, 124(5), 2823–2840. https://doi.org/10.1029/2018jd028859

Gyamerah, S. A., & Gil-Alana, L. A. (2023). A multivariate causality analysis of CO2 emission, electricity consumption, and economic growth: Evidence from Western and Central Africa. *Heliyon*, 9(1), e12858. https://doi.org/10.1016/j.heliyon.2023.e12858

Hazra, A., & Gogtay, N. (2016). Biostatistics Series Module 6: Correlation and Linear Regression. *Indian journal of dermatology*, 61(6), 593–601. https://doi.org/10.4103/0019-5154.193662

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314. https://doi.org/10.2307/1390807

Jeon, H. (2022). CO2 emissions, renewable energy and economic growth in the US. *The Electricity Journal*, 107170. https://doi.org/10.1016/j.tej.2022.107170

Liu, W., et al. (2022). Age Structure and Carbon Emission with Climate-Extended STIRPAT Model-A Cross-Country Analysis. *Frontiers in Environmental Science*, vol. 9, 25, https://doi.org/10.3389/fenvs.2021.719168

Magnus, J. R. (2019). On Using the t-Ratio as a Diagnostic. Econometrics, 7(2), 24. https://doi.org/10.3390/econometrics7020024

McCurdy, J. D., & Rhodes, E. (2023). What Drives Greenhouse Gas Emissions? An International Scoping Review of Academic Studies in 2010–2019. Climate Resilience and Sustainability, vol. 2, no. 3, https://doi.org/10.1002/cli2.52

McGrath, G. (2021, June 9). Electric power sector CO_2 emissions drop as generation mix shifts from coal to natural gas - Today in Energy - U.S. Energy Information Administration (EIA). Www.eia.gov. https://www.eia.gov/todayinenergy/detail.php?id=48296

NASA. (2024, October 23). The Causes of Climate Change (K. Velev, Ed.). NASA. https://science.nasa.gov/climate-change/causes/

NCEI.Monitoring.Info@noaa.gov. (2026). Climate at a Glance | National Centers for Environmental Information (NCEI). Noaa.gov. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series/45/tavg/12/12/2010-2023

Pal, M., Bharati, P. (2019). Introduction to Correlation and Linear Regression Analysis. In: Applications of Regression Techniques. Springer, Singapore. https://doi.org/10.1007/978-981-13-9314-3_1

Payne, J. E., et al. (2024) Renewable Energy Production across U.S. States: Convergence or Divergence? *Energy Economics*, vol. 140, 108015, https://doi.org/10.1016/j.eneco.2024.108015.

Vol. 2026 (8) 660 – 674
ISSN 2688-3651 [online]

Stern, D.I. (2011), The role of energy in economic growth. *Annals of the New York Academy of Sciences*, 1219: 26-51. https://doi.org/10.1111/j.1749-6632.2010.05921.x

Waldorp, L., & Marsman, M. (2022). Relations between Networks, Regression, Partial Correlation, and the Latent Variable Model. *Multivariate behavioral research*, 57(6), 994–1006. https://doi.org/10.1080/00273171.2021.1938959

Wang, C., et al. (2024). Navigating Sustainability in the US: A Comprehensive Analysis of Green Energy, Eco-Innovation, and Economic Policy Uncertainty on Sectoral CO2 Emissions. *Energy Reports*, vol. 11, 5286–5299, https://doi.org/10.1016/j.egyr.2024.05.014.

Xu, Y., et al. (2025). Spatial Heterogeneity of Carbon Emissions and Low-Carbon Optimization for Small Towns Based on the Geographically Weighted Regression Model. *Ecosystem Health and Sustainability*, vol. 11, https://doi.org/10.34133/ehs.0424.

Zhang, S., et al. (2021). Gauging the Impacts of Urbanization on CO2 Emissions from the Construction Industry: Evidence from China. *Journal of Environmental Management*, vol. 288, 112440, https://doi.org/10.1016/j.jenvman.2021.112440.

Supplementary information

https://docs.google.com/document/d/1aKrpbF0oPp0k51Qeegx_aQlwxrNKmjZe/edit?usp=sharing&ouid=107269886536271979834&rtpof=true&sd=true

- Table S1. Regional Classification of U.S. States and Number of Observations (n). Regions are defined according to the study's classification framework. Hawaii and Mississippi were excluded due to insufficient data availability.

- Table S2. Population-based Classification of U.S. States and Number of Observations (n). States were categorized into population percentile groups (25th, 50th, 75th, and 100th) based on their population levels within the dataset. Hawaii and Mississippi were excluded due to insufficient data availability.

- Table S3. Multiple Linear Regression Results Across Nationwide, Regional, and Population Percentile Groups (2010–2023). The table presents coefficients, standard errors (SE), t-statistics, p-values, and 95% confidence intervals for predictors of CO_2 emissions per capita, including energy generation, GDP, temperature, and precipitation. Variables are measured on a per capita basis, and statistical significance is indicated by p-values. Panels represent (A) Nationwide, (B) West, (D) East, (D) Midwest, (E) South, (F) 25th Percentile, (G) 50th percentile, (H) 75th percentile, and (I) 100th percentile.

- Multiple Linear Regression Equations