Yoseph Habtu

DATA 205 (CRN 34669)

Prof. Perine

12 May 2025

Rethinking Bioenergy – Final Report

**Introduction & Project Overview**

Bioenergy is a renewable energy source derived from biomass, or organic material (e.g., plants, wood, food waste). Though widely used, questions remain about its true sustainability and economic competitiveness. Fossil fuels, including coal, petroleum, and natural gas, still dominate the U.S. energy sector but pose well-known environmental and cost-related challenges.

This project aims to evaluate whether bioenergy is a viable long-term substitute for fossil fuels by comparing the two groups across four sustainability metrics:

* Carbon Intensity = Emissions per unit of energy generated
* Energy Conversion Efficiency = Energy produced relative to fuel consumed
* Infrastructure Impact = Construction cost per generation capacity
* Economic Performance = Fuel cost/price trends and volatility over time

**Data Overview**

* EPA Greenhouse Gas Emissions by Fuel Source (2010–2023): Facility-level greenhouse gas emissions in the U.S. categorized by fuel source and unit type; includes location and year data
* EIA Electricity Generation by Energy Source (2014–2023): Annual totals of net electricity generation for various energy sources (renewable and nonrenewable) across all sectors in the U.S. grouped by facility type (e.g., utility-scale, small-scale)
* EIA Energy Consumption by Source (1949–2023): Historical data on primary energy consumption (measured in quadrillion Btu) in the U.S. categorized by energy source, including fossil fuels (coal, natural gas, petroleum) and renewables (biomass, among others)
* EIA Construction Cost Data for Electric Generators by Major Energy Source (2013–2022): Annual construction costs for electric generators in the U.S, categorized by energy source; also contains generator capacity data and installation totals EIA Densified Biomass Fuel Report (2016–2024): Prices and volumes of densified biomass fuel and feedstock
* EIA Fossil Fuel Costs (2013–2023): Annual average cost of coal, gas, and petroleum fuels

**Research Goal & Questions**

The goal of this project is to assess whether bioenergy can serve as a viable alternative to fossil fuels for electricity generation in the U.S. To do this, the project compares the performance of bioenergy and fossil fuels across various environmental and economic metrics.

* Environmental Analysis
  + How does carbon intensity (emissions per unit of energy) compare between biomass and fossil fuels over time?
  + How efficiently does each fuel type convert feedstock into usable electricity?
* Economic Analysis
  + How does the relationship between construction cost and generation capacity differ between biomass and fossil fuel plants?
  + What are the trends in fuel prices/costs over time, and how stable are they?

**Tools & Methods**

* Tools: Google Colab, Excel, GitHub
* Programming language: Python
  + Libraries: Pandas, NumPy, Matplotlib, Seaborn, SciPy, Statsmodels
* Statistical Methods:
  + Carbon Intensity:
    - Comparative analysis via Student’s t-test and Cohen’s d (for effect size), after conducting standard assumption checks (Shapiro-Wilk test for normality, Levene’s test for homogeneity of variances)
    - Trend analysis: Used regression modeling for both fuel groups (linear for fossil fuels, quadratic for biomass) only to fit trend lines on a scatter plot of carbon intensity values over time, with no intent to forecast future values; evaluated model performance metrics (R², MAE, RMSE)
  + Energy Efficiency
    - T-test and Cohen’s d to compare mean efficiency values between the two fuel groups of interest; conducted prior assumption testing (Shapiro-Wilk, Levene’s)
  + Infrastructure Impact
    - Correlation analysis (groupwise scatter plots of construction cost per generation capacity vs. total capacity; complementary correlation matrices for each fuel group’s cost-capacity relationship)
  + Fuel Cost Performance (Trends + Stability)
    - Visual trend analysis (dual-axis groupwise line plot of annual fuel prices/costs for biomass and fossil fuels, including annotations for year-over-year percent changes)
    - Mann-Whitney U tests on year-over-year (YoY) relative and absolute price changes

**Data Cleaning & Preprocessing**

* Merged multi-year tables and reshaped datasets for consistency (e.g., wide to long format)
* Handled missing values using median imputation or row removal based on context
* Created group-level categories (e.g., All Biomass, All Fossil Fuels) for aggregation
* Categorized ambiguous or overlapping fuel types using EPA/EIA documentation and domain research​
* Harmonized units, column names, and energy source names across datasets for integration and comparability
* Constructed summary tables that aggregated data by fuel type (biomass vs. fossil fuels) and year for each area of analysis
* Derived normalized variables (e.g., carbon intensity, energy efficiency) from the summary data

**Basic Descriptive Statistics**

* KDE plots were plots were created to visualize and compare distributions of values for each raw/basic metric (e.g., emissions, consumption, electricity generation) between the two fuel groups of interest; excluded fuel costs/prices since biomass and fossil fuels use different units

**Initial Insights:**

* Fossil fuels dominate in scale but show high variability across emissions, consumption, and generation, potentially due to their diverse sources and long-term shifts.
* Biomass is consistent but small-scale, with tight clustering across all metrics and minimal fluctuation.
* Construction cost per MW is higher for biomass, suggesting higher investment needs despite lower output.
* Overall, these plots reveal greater stability in biomass operations vs. fossil fuels' more volatile and transitional nature.
* Summary tables of annual values for core sustainability metrics (e.g., carbon intensity, energy efficiency), one table for each fuel group; including only the years 2016, 2019, and 2021, due to limited time range overlap between the datasets for each metric (these 3 years were the best available option since they are mostly evenly spread out)

**Initial Insights:**

* Biomass has consistently higher carbon intensity than fossil fuels in all selected years (roughly 5-7 times higher)
* Fossil Fuels convert fuel to electricity much more efficiently, producing 2.7-3x more output per unit of energy
* Biomass electric generators cost significantly more to build per MW of capacity (~ 2.5-3x higher construction costs)
* Biomass fuel prices rose steadily from 2016 to 2021, while fossil fuel costs fluctuated and spiked in 2021
* These descriptive statistics show a clear performance gap favoring fossil fuels; however, deeper evaluation is needed to understand the full context and magnitude of these differences.

**Final Data Products**

Results, visualizations, and conclusions for each research area:

* Carbon Intensity
  + Exploratory visualizations: 1) Grouped bar chart of GHG emissions by fuel type in 2013 vs. 2023; 2) Line plot of raw total emissions over time (2014-2023) for biomass vs. fossil fuels
* Comparative analysis:
  + T-test results: t = 54.6337, p = 0.0000 (rounded to 4 decimal places)
    - Extremely significant result 🡪 Strong evidence of a difference in carbon intensity between biomass and fossil fuel sources.
  + Cohen’s d: 24.4329
    - Huge effect size 🡪 The difference between the two groups is not only statistically significant, but practically very large as well.
* Trend analysis:
  + Visual groupwise scatter plot of carbon intensity over time for biomass and fossil fuels, with fitted trend lines created from regression models
  + Table of performance metrics for each fuel group’s model

**Interpretation and Implications of Results:**

* Fossil fuels show a clear and steady decline across the decade, especially from 2014 to 2020, with a small uptick in 2021 before continuing downward. This trend aligns with a shift toward cleaner-burning fuels and is well-modeled linearly (R² ≈ 0.93).
* Biomass, in contrast, shows a more complex pattern. There’s a gradual decline in the earlier years, but it flattens out and slightly increases from around 2018 onward. The curved trajectory is better captured by a quadratic fit (R² ≈ 0.81).
* These trends contrast those shown in the earlier total emissions plot: although biomass has higher carbon intensity, it emits far less in total, likely due to overall energy (i.e., electricity) production.
* Energy Efficiency
* Exploratory visualizations: 1) Line plot of percent change in electricity generation since 2014 for biomass vs. fossil fuels; 2) Line plot of long-term (1949-2023) energy consumption by source (coal, natural gas, petroleum, biomass)
* Comparative analysis
  + Grouped bar chart of energy efficiency values in the 2020s, biomass vs. fossil fuels
  + T-test results: t = 36.5894, p = 0.0000 (rounded to 4 decimal places); Cohen’s d: 25.876

**Interpretation and Implications of Results:**

* Across the last four years, fossil fuels have shown much higher energy conversion efficiency than biomass — with a consistent gap of about 2x to 3x.
* Biomass efficiency has steadily declined from 2020 to 2023, while fossil fuel efficiency has remained relatively stable.
* This persistent and widening gap is reflected in the statistical analysis:
* Student’s t-test confirms the difference in group means is statistically significant (t = 36.59, p < 0.001).
* Cohen’s d = 25.87 indicates an extremely large effect size, supporting a strong performance divide.
* When paired with earlier findings (e.g., stagnant generation growth), the low output-to-input ratio suggests biomass currently lags behind fossil fuels in terms of energy return, which is a key metric of long-term sustainability.
* Infrastructure Impact
* Exploratory visualization: boxplot of generator construction costs per generation capacity ($/MW) by energy source (including both renewable and nonrenewable)
* Correlation analysis:
  + Biomass cost-capacity correlation: 0.33
  + Fossil fuels cost-capacity correlation: - 0.31

**Interpretation and Implications of Results:**

* Fossil fuels show a weak negative correlation (r = –0.31), suggesting that larger plants may achieve slightly lower per-MW costs — a possible, though modest, indication of economies of scale.
* Biomass shows a weak positive correlation (r = 0.33), implying that cost per MW may increase with capacity, possibly due to logistical, technological, or site-specific inefficiencies in scaling biomass projects.
* However, both correlation values are very weak by conventional standards (|r| < 0.4) and neither is statistically significant.
* Given these limitations, the observed relationships are suggestive at best — not strong enough to claim that one group has a “better” or more efficient cost-capacity dynamic.
* Still, the directional differences hint at structural challenges in scaling biomass infrastructure relative to fossil fuels.
* Fuel Cost Performance (Trends + Stability)
* Exploratory visualizations: 1) Groupwise histogram (with KDE plot) of densified biomass fuel prices by market type (domestic vs. export); 2) grouped bar chart of annual (2013-2023) average fossil fuel costs by source
* Trend analysis:
  + Dual-axis groupwise line plot of annual fuel prices/costs for biomass and fossil fuels

**Summary & Interpretation of Results:**

* Biomass prices followed a steady upward trend, with moderate year-over-year increases nearly every year. This consistency suggests relative stability in bioenergy markets over time.
* In contrast, fossil fuel costs showed a much more volatile trajectory, with a dramatic spike during the COVID-19 period (2021–2022), followed by a steep decline in 2023.
* The divergence in behavior is especially clear in 2021, when fossil fuel costs surged by over 70%, while biomass prices rose less than 12%.
* These patterns indicate that fossil fuel markets are more vulnerable to short-term shocks and global disruptions, whereas biomass has shown more predictable price dynamics.
* Although biomass prices are not immune to increases, their smoother trajectory could offer greater price stability, an important consideration for long-term energy planning and investment.
* Comparative analysis of year-over-year fuel cost percent-changes:
  + Two-sided Mann-Whtiney U test # 1 results (using **raw/relative** percent-change values): test statistic = 27.0, p-value = 0.8048
  + Two-sided Mann-Whtiney U test # 1 results (using **absolute** percent-change values): test statistic = 8.0, p-value = 0.03788

**Interpretation and Implications of Results:**

* When comparing the raw percent-change values, there was no statistically significant difference between biomass and fossil fuels (p = 0.80). This suggests that the direction and magnitude of yearly changes varied similarly across groups when considering increases and decreases together.
* However, when looking at the absolute percent-change values — which capture volatility regardless of direction — the difference was statistically significant (p = 0.038).
* These findings support the earlier visual observation that fossil fuel costs exhibited greater year-to-year volatility than biomass fuel prices over the 2017-2023 period.

**Project Conclusions:**

* + Environmental Tradeoff: Biomass did not outperform fossil fuels in carbon intensity — a key emissions metric — and currently emits more CO₂ per unit of electricity generated. This challenges the assumption that bioenergy is inherently cleaner, despite its renewable classification.
  + Efficiency Barrier: Biomass also lags behind fossil fuels in energy conversion efficiency. The sizable performance gap presents a significant obstacle to large-scale substitution unless technical advancements in biomass systems are achieved.
  + Infrastructure Parity: No strong cost-capacity relationship was found for either group, suggesting that biomass projects are not inherently more cost-efficient or scalable than fossil fuel plants — a reminder that infrastructure decisions are highly context-specific.
  + Economic Stability Advantage: Where biomass does show clear strength is in fuel cost stability. Unlike fossil fuels, which experienced extreme price volatility in recent years, biomass prices followed a steady, predictable trajectory — a critical asset for long-term energy planning.
  + Overall Viability Judgment: Bioenergy cannot fully replace fossil fuels in its current form due to clear disadvantages in carbon emissions and conversion efficiency. However, its long-term viability remains plausible — particularly as a stable, renewable complement in a diversified energy mix. If technical improvements can close the performance gap, biomass could play a meaningful role in the U.S. transition away from fossil fuels.

**References & Acknowledgements**

* Data Sources:
  + U.S. Energy Information Administration (EIA)
    - Construction Costs, Electricity Generation, Biomass Prices, Fossil Fuel Costs, Primary Energy Consumption
  + U.S. Environmental Protection Agency (EPA)
    - Greenhouse Gas Reporting Program (GHGRP)
* Acknowledgements:
  + Data Science faculty at Montgomery College for their support and guidance throughout my learning journey:
    - Prof. Perine
    - Prof. Mohamed
    - Prof. Saidi
    - Prof. Evans