Emission Explorer

Alicia Dawn Taylor; Namra Ghafoor; Sharvari Sunil Mhatre; Shweta Saxena
Georgia Institute of Technology
{ataylor366; nghafoor6; smhatre34; ssaxena71}

Introduction

Our objective is to identify a visualization and data-driven approach that will facilitate strategies to achieve the United States Federal Government's goal of reducing Greenhouse Gases (GHGs) and achieving carbon net neutrality by 2050 (United States Department of State and United States Executive Office of the President, 2021). This federal goal is currently being implemented through laws like the Inflation Reduction Act, improved passenger vehicle standards, and Environmental Protection Agency regulations on gas and oil. Our focus will be on one of the major contributors to GHGs: energy generation. Our project will help stakeholders to identify additional potential for improvement in areas like transitioning to renewable energy sources and support collaborative learning by identifying success in similar areas and socio-environmental development by identifying demographic split and correlation.

Problem Definition

Identify the geographical areas with the least renewable energy and forecast future statistics using an analytical approach to create tools for visualization of the eGRID dataset to assist stakeholders with decision-making for continuous improvement.

Dataset Overview:

Every year, the Clean Air Market Division (CAMD), Office of Atmospheric Programs (OAP), and U.S. Environmental Protection Agency (EPA) release The Emissions & Generation Resource Integrated Database (eGRID). eGRID is a comprehensive data source on the environmental characteristics of all electric power generated in the United States. This dataset includes electricity generation by resources (fuel), emissions percentage for various GHGs, and classification into renewable and non-renewable resources. The latest edition of this dataset includes demographic data based on region from the census dataset. Currently, the dataset is available for the years 1996 to 2022. The published data is an Excel document with no interactive tool available and has the limitation of the most recent data being over a year old. As the eGrid dataset is not available for each year, additional data was pulled from the eia.gov API using Python code to unpack the dictionaries present in the API. The Excel functions Index and Match were then used to merge CO_2 emissions data from the API with population and GDP data.

Literature Survey

One benefit of providing visualizations and predictive analytics is that decision-makers can better understand the dynamics of energy demand and social implications, leading to more informed policy decisions (Walker, 2014). We will focus on best practices in User-Centered Design as explored by research from Academic Cartography (Tsou, 2013) and highlighting point features from research from the Cartographic Journal (Gedicke & Haunert, 2023).

Previous energy-related research around GHG emissions done in Ghana used Autoregressive Integrated Moving Average (ARMIA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) to forecast natural gas consumption (Yen Adams Sokama-Neuyam, 2023). Research from Iran used multiple linear regression (MLR) and multiple polynomial regression (MPR) analysis to predict CO_2emissions due to the unique energy-producing environment in Iran (Hosseini, Saifoddin, Shirmohammadi, & Aslani, 2019). The machine learning method long short-term memory (LSTM) was used alongside Season Autoregressive Moving Average with Exogenous Regressors (SARIMAX) and Holt-Winters and for CO_2 emissions prediction in India (Kumari & Singh, 2023). The Box Jenkins ARIMA model was used by researchers in Libya to predict the annual warming trend along the coast (El-Mallah & Elsharkawy, 2016). In a study of the Gulf countries various ARIMA models were used to estimate CO_2 emissions including ARIMA (1,1,1), ARIMA (1,1,2), and ARIMA (2,1,2) (Alam & Alarjani, 2021). Of the research, ARIMA was a very popular model as it is easy to interpret, and many variations can be used.

The demographic step of the project is inspired based on the article 'Demographic change and CO_2 emissions which explores the historical IPAT equation i.e. I=PxAxT where I stands for environmental impact, P stands for population, A stands for consumption and T stands for impact per consumption (Brian C O'Neill, 2012). A study from 2017 found that population growth has a much greater impact than economic growth on carbon emissions (Casey & Galor, 2017). In an article by Grossman, he reviews the shape of the relationship between income and environmental harm called the Environmental Kuznets curve (EKC) (Grossman, 2010). Other authors have proposed different shapes other than his U-shaped curve such as a linear relationship or an N-shaped relationship (Li & Lin, 2015), (Nassani, Aldakhil, Abro, & Zaman, 2017). These results will help stakeholders identify areas that need social and cultural awareness related to energy habits and environmental changes and show the location on the EKC curve.

Phase 1: Visualization and Interactive Tool

For the visualization portion of our project, we have selected Tableau due to its modular design and user-friendly interface. Leveraging Tableau allowed us to seamlessly integrate various maps simplifying the visualization process. Our interactive dashboard, shown in Figure 1, titled 'Emissions Explorer: US Regional Energy and Environmental Analysis Dashboard', offers insights into key environmental metrics across US regions. With four dynamic maps, users can track Total CO_2 Emissions, Electricity Generation by Renewable and Non-Renewable sources, CO_2 Emissions by Sector, and Population trends over the years. Users can easily filter data by Year, Energy Type, and Sectors for deeper analysis.

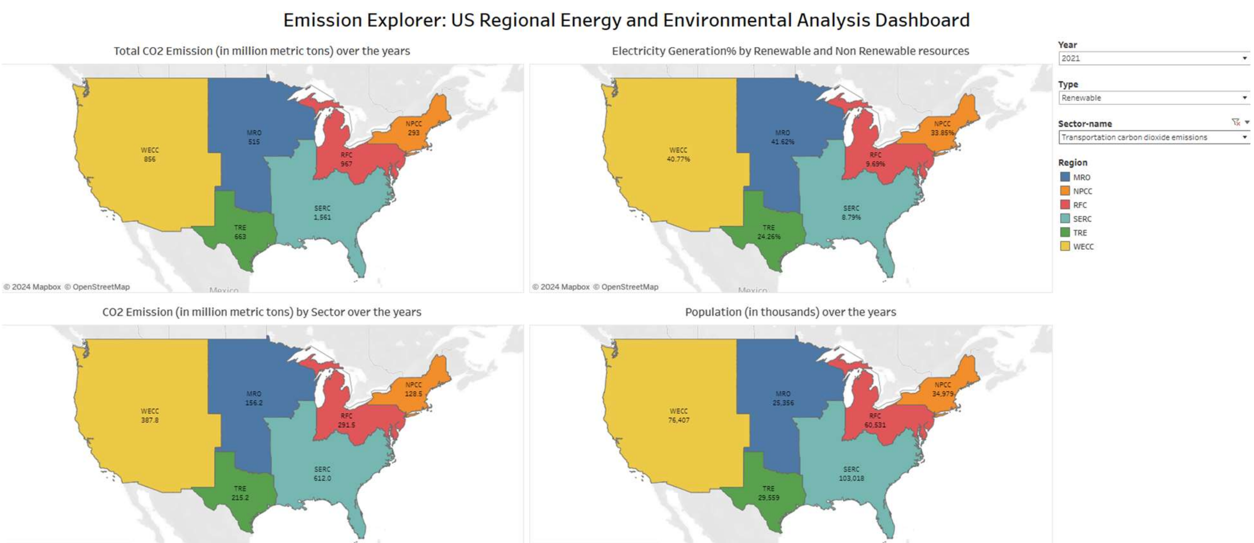

Figure 1: Tableau Dashboard showing Total CO_2 Emissions, Electricity Generation Percentage by Renewable and Non-Renewable Resources, CO_2 Emissions by Sector and Population

Historical trends of energy generation by type vary significantly among the six regions. While some regions like WECC prioritize renewable sources, others like RFC rely more heavily on non-renewable sources. Overall, there's a trend toward increasing renewable energy generation, but the pace and extent differ across regions due to factors like resource availability and policy frameworks. By using the filters in the dashboard, the user can see historical trends in CO_2 emissions in the United States vary by region and sector due to factors such as changes in energy consumption patterns, economic activity, energy efficiency measures, and shifts in energy sources. Generally, there's a trend of decreasing CO_2 emissions per unit of electricity generated, driven by increased energy efficiency, a shift to cleaner energy sources, and emissions reduction policies. Our innovation in our Tableau maps involves special data visualization by plotting data points on maps using longitude and latitude coordinates, interactive filters that allow the user to select specific data subsets based on year, region, and sector identifiers.

Phase 2: Modeling Regions with Autoregressive Integrated Moving Average

For the second phase, we experimented with ARIMA and exponential smoothening models to forecast future carbon emissions. As an innovation to our model, we trained data for each region independently as trends, seasonality, governance policy, and environmental conditions are different than the United States as a whole. After cleaning the data to remove missing values, excluding total fuel as each region does not use each fuel, and filtering out summary rows, data from 1996 to 2012 was classified as training data, and data after 2012 was used as test data.

For models we plotted the Auto Correlation Function (ACF) and Partial Correlation Function (PACF), to identify if the data is stationary and decide on the stationarity of the dataset. Based on the result, we concluded that data for all regions is stationary and decided Integrated factor (d) for the first two manually selected ARIMA models as 0. Therefore, for all regions, we trained two manually selected models, ARIMA (1,0,0) also known as the auto-regressor model, and ARIMA (0,0,2) also known as the moving average model over 2 periods. The third ARIMA model we selected automatically using the pmdarima package in Python. We also trained data using three Exponential Smoothening models by varying smoothening levels as none, 0.2, and 0.5.

Region	Selected Model
Southeast Reliability Corporation (SERC)	Exponential smoothing with smoothing level of 0.2
Midwest Reliability Organization (MRO)	Exponential smoothing with smoothing level none
Northeast Power Coordinating Council (NPCC)	ARIMA (1,0,1)
Reliability First Corporation (RFC)	ARIMA (0,0,2)
Western Electricity Coordinating Council (WECC)	Exponential smoothing with a smoothing level of 0.5
Texas Regional Entity (TRE)	ARIMA (0,0,2)

Table 1: Models Selected for Each Region

We trained 6 different models for each regional data. For the selection of the best model for each region, we used an evaluation matrix including mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and R-squared score. For the selection of the model, we looked for lower values of MSE, RMSE, and MAPE and higher values of the r-squared score as these metrics show the accuracy, precision, and predictive power of the models. Based on the result we selected a model for each region as seen above in Table 1.

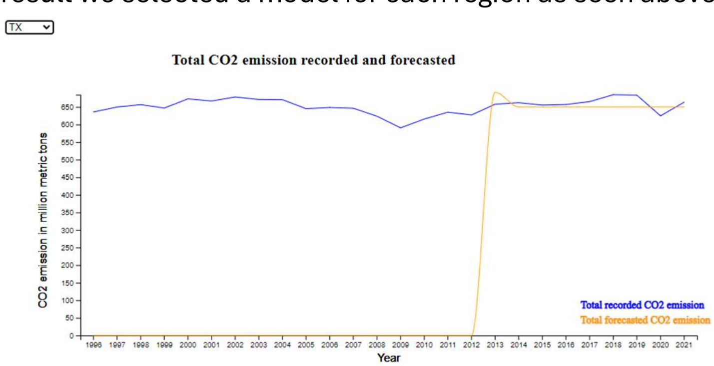

Figure 2: Image of the D3 Output showing forecasted and recorded values of CO_2 emissions for the selected model of the Texas Energy Region

To visualize the predictive power of these selected models we are using D3 to show the results of the model. An example of the D3 is shown in Figure 3. The user will be able to select a region of study from the top left dropdown menu and a corresponding graph of recorded values vs forecasted values for CO_2 emissions will be displayed.

Phase 3: Demographic in Economic Environmental Impact

For the final phase, we constructed an Economic Kuznets Curve (EKC) for each region to compare the shapes between regions. Traditionally an Economic Kuznets curve has environmental degradation as the dependent variable and a combination of Global Domestic Product (GDP) and population as independent variables, however, since the United States has a fully industrialized economy, as our innovation to this method we opted to not use GDP and are just using population as the independent variable and are using CO_2 for the dependent variable.

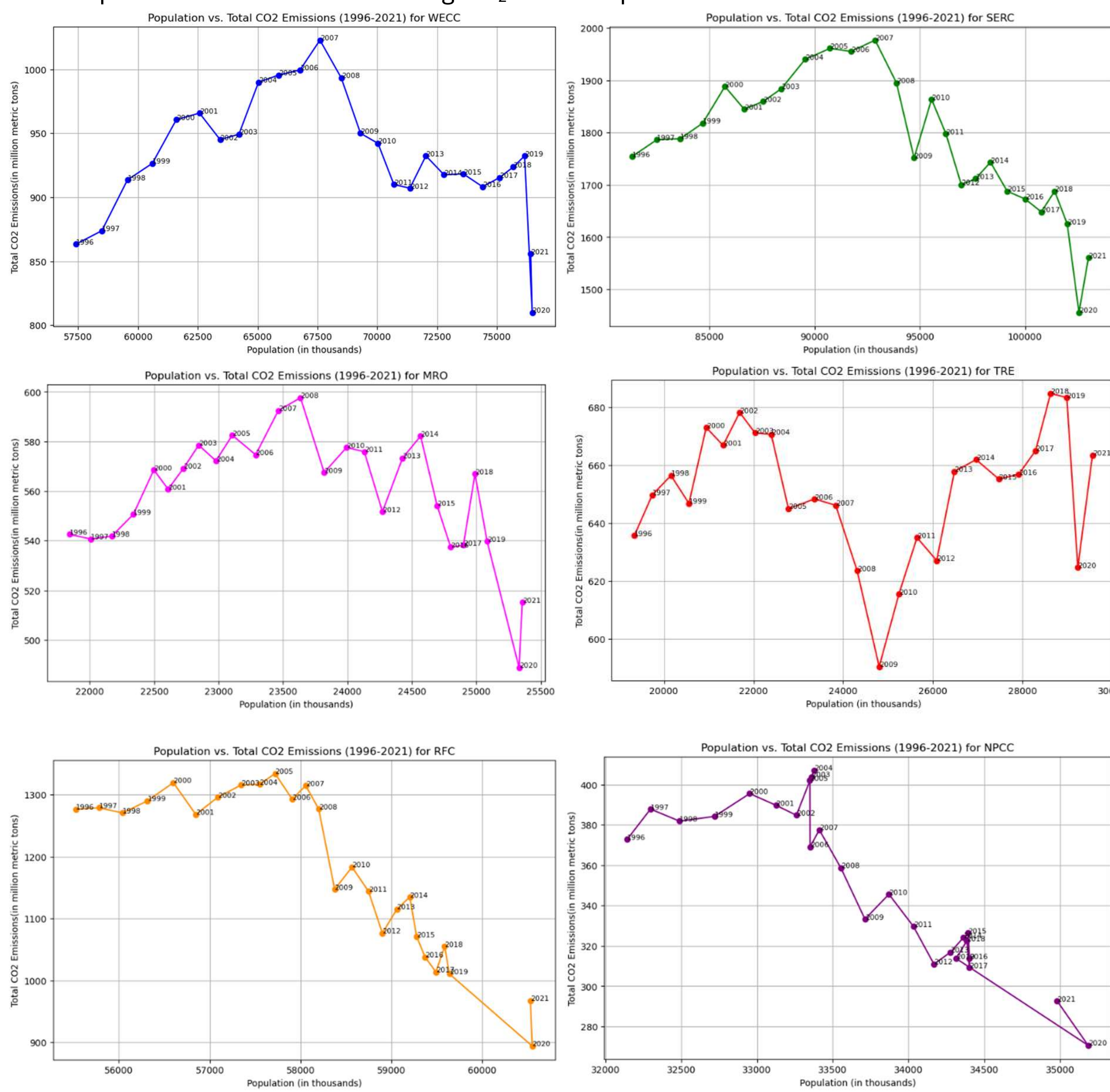

Figure 3: EKC for each of the six regions WECC shown in blue, SERC shown in green, MRO shown in pink, TRE shown in red, RFC shown in gold, NPCC shown in purple

The graphs above show the EKC curves for the WECC, MRO, and SECC regions in blue, green, and pink showing the typical inverted "U" shape for the graph. These three regions peaked in CO_2 emissions around the 2007-2008 financial crisis and have since continued to decrease in emissions while the population has continued to increase. These three regions have favorable weather conditions for wind and solar and have implemented government programs that provide incentives for switching to renewable energy.

The next three EKCs are for the TRE, RFC and NPCC regions shown in red, gold, and purple. The TRE region is somewhat of an outlier since it is only one state (Texas). Texas continues to have a high dependence on non-renewable resources and the variants in the graph appear to reflect a decrease in CO_2-causing activities rather than a switch to renewable resources as it lines up with the 2008 crisis and the COVID-19 pandemic in 2020. For the RFC and NPCC regions in the northeast section of the United States, there slower increase in CO_2 emissions at the beginning of the graph and both had a steep decline after 2008. While some of this is due to an increase in wind and solar energy like the previous three regions, the dependence on non-renewal energy for RFC also suggests that some of this is a decrease in CO_2-producing activities.

To further see how the region's energy is related we ran a correlation analysis. Pearson's Correlation summarizes the strength of the linear relationship between data. This coefficient returns a value between -1 and 1, where the value of -1 is a negative correlation, 1 is a positive correlation and 0 represents no correlation (Brownlee, 2023). In Table 2 below, the correlation from 1996-2021 shows the SERC, NPCC, and RFC with a strong negative correlation which means in general that as Population growth increases, CO_2 emissions have decreased over the years. This can be justified as these regions have shown growth in renewable energy production year over year. The correlation from 2008-2021 becomes slightly more negative for these regions. In the full correlation, MRO and WECC have a slight negative correlation which indicates efforts to reduce CO_2 emissions, but at a slower rate per capita, but there is a much stronger correlation for the 2008-20201 correlation. The TRE region shows approximately no correlation, which is also indicated by the EKC.

Region	Correlation 1996-2021	Correlation 2008-2021
Southeast Reliability Corporation (SERC)	-0.6301	-0.8804
Midwest Reliability Organization (MRO)	-0.2946	-0.8117
Northeast Power Coordinating Council (NPCC)	-0.8612	-0.9271
Reliability First Corporation (RFC)	-0.8562	-0.9284
Western Electricity Coordinating Council (WECC)	-0.2282	-0.6884
Texas Regional Entity (TRE)	0.0282	0.7115

Table 2: Regional Correlation for 1996-2021 and Regional Correlation for 2008-2021

Conclusions and Discussion

In a study from 2018, analyzing individual state contributions to carbon emissions, the impact of half of the United States cooperating with this study's plan would result in a 68% decrease in carbon reduction (Galan-Martin, et al., 2018). The visualization, forecasting, and demographic tools from this project, could raise this number even higher. By employing advanced analytics and visualization techniques, we aim to provide stakeholders with a deeper understanding of regional energy dynamics. This will help identify areas for improvement and facilitate evidence-based decision-making to achieve carbon neutrality goals.

All team members have contributed a similar amount of effort.

References

Alam, T., & Alarjani, A. (2021). A Comparative Study of CO2 Emission Forecasting in the Gulf Countries Using Autoregressive Integrated Moving Average, Artificial Neural Network, and Holt-Winters Exponential Smoothing Models. *Advances in Meteorology*.

Brian C O'Neill, B. L. (2012). Demographic change and carbon dioxide emissions. *The Lancet volume 380, issue 9837*, P157-164.

Casey, G. P., & Galor, O. (2017). Is faster economic growth compatible with reductions in carbon emissions? The role of diminished population growth. *Environmental Research Leters*.

CO2 emissions of all world countries. (2022). Retrieved from EDGAR - Emissions Database for Global Atmospheric Research: https://edgar.jrc.ec.europa.eu/report_2023

Daim, T. U., Amer, M., & Brenden, R. (2012). Technology Roadmapping for wind energy: case of the Pacific Northwest. *Journal of Cleaner Producction*, 27-37.

Directorate-Generate for Climate Action. (2023, October 24). *European Comission*. Retrieved from State of the Energy Union 2023: Further action needed to accelerate climate action: https://climate.ec.europa.eu/news-your-voice/news/climate-action-progress-report-2023-2023-10-24_en

El-Mallah, E., & Elsharkawy, S. (2016, February 24). Time-Series Modeling and Short Term Prediction of Annual Temperature Trend on Coast Libya Using the Box-Jenkins ARIMA Model. *Advances in Research*, pp. 1-11.

Galan-Martin, A., Pozo, C., Azapagic, A., Grossmann, I., Mac Dowell, N., & Guillen-Gosalbez, G. (2018). Time for global action: an optimised cooperative approach towards effective climate change mitigation. *Royal Society of Chemistry*, 572-581.

Gedicke, S., & Haunert, J.-H. (2023, August 15). An Empirical Study on Interfaces for Presenting Large Sets of Point Features in Mobile Maps. *The Cartographic Journal*, pp. 25-42.

Gillingham, K., Boyd, P., & Gentry, B. (2020, April 188). *Yale Experts Explain Carbon Neutrality*. Retrieved from Yale Sustainability: https://sustainability.yale.edu/explainers/yale-experts-explain-carbon-neutrality

Goldstein, B., Gounaridis, D., & Newell, J. P. (2020, August 11). *The carbon footprint of household energy use in the United States*. Retrieved from PubMed Central: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7431053/

Grossman, G. M. (2010, August 04). Pollution and growth: what do we know? *The Economics of Sustainable Development*, pp. 19-46.

Hosseini, S. M., Saifoddin, A., Shirmohammadi, R., & Aslani, A. (2019). Forecasting of CO2 emissions in Iran based on time series and regression analysis. *Energy Reports*, 619-631.

Jacobson, M. Z., Delucchi, M. A., Bazouin, G., Bauer, Z. A., Heavey, C. C., Fisher, E., . . . Yeskoo, T. W. (2015). 100% clean and renewable wind, water, and sunlight (WWS) all-sector energy roadmaps for the 50 United States. *Energy & Environmental Science*.

Knox, J. H. (2018, June 6). The Paris Agreement As a Human Rights Treaty. *Human Rights and 21st Century Challenges: Poverty, Conflict, and the Environment*, p. 26.

Kumari, S., & Singh, S. K. (2023). Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research*, 116601-116616.

Li, K., & Lin, B. (2015, August 25). Impacts of urbanization and industrialization on energy. *Renewable and Sustainable Energy Reviews*, pp. 1107-1122.

Li, Y. (2020, February 20). Forecasting Chinese carbon emissions based on a novel time series prediction method. *Energy Science & Engineering*, pp. 2274-2285.

Nassani, A. A., Aldakhil, A. M., Abro, M. M., & Zaman, K. (2017, April 5). Environmental Kuznets curve among BRICS countries: Spot lightening. *Journal of Cleaner Production*, pp. 474-487.

Scheffers, B. R., De Meester, L., Bridge, T. C., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. T., . . . Watson, J. E. (2016, November 11). The broad footprint of climate change from genes to biomes to people. *Science*.

Sofia O'Connor, J. M. (2019). Corporate Renewable Energy Goals: What Does " 100% Renewable" Really Mean? *Law Journal Library*, 10648- 10669.

Tsou, M.-H. (2013, March 14). Revisiting Web Cartography in the United States: the Rise of User-Centered Design. *Cartography and Geographic Information Science*, pp. 250-257.

United States Department of State and United States Executive Office of the President. (2021). *The Long-Term Strategy of the United States: Pathways.* Washington DC.

Walker, G. (2014, March). The dynamics of energy demand: Change, rhythm and synchronicity. *Energy Research & Social Science*, pp. 49-55.

Yen Adams Sokama-Neuyam, S. M. (2023, December 07). Predictive modeling of energy-related greenhouse gas emissions in Ghana towards a net-zero future. *Greenhouse Gases: Science and Technology (Volume 14, Issue 1)*, pp. p. 42-61.