HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 3 (2024)

Examining the Relationship Between Environmental Hazards and Socioeconomic Factors in the United States Using Machine Learning Methods

William Zhou

Hunter College High School, USA

ABSTRACT

This study investigates the intricate relationship between environmental factors and socioeconomic indicators across all U.S. states, utilizing data from the EPA EJScreen. With environmental challenges like climate change and air pollution continuing to persist, it is essential to understand not only their overall impact but also who is most affected. Disadvantaged groups, such as low-income individuals and people of color, often bear the brunt of these issues, experiencing heightened exposure to pollutants and other hazards. Through comprehensive analysis using advanced machine learning methods such as correlation analysis, Pearson correlation coefficient, linear regression, XGBoost, and neural networks, this research identifies statistically significant correlations between environmental and socioeconomic factors in certain states, potentially linked to historical patterns of systemic discrimination in policymaking. Conversely, in other states, this relationship appears less pronounced. Further investigation is warranted to elucidate the underlying factors contributing to these variations, including geographical features and demographic characteristics. While providing valuable insights, this study acknowledges limitations such as the reliance on state-level data and the use of correlation analysis, which does not establish causation. Moving forward, future research should explore these relationships at finer geographical scales and employ methodologies beyond correlation analysis to better understand causal relationships. By addressing these limitations and continuing to explore the complexities of environmental justice, policymakers can inform more targeted and effective interventions to address environmental disparities and advance equity for all communities.

Introduction

As environmental problems like climate change and air pollution continue to persist, it is vital to not only look at the overall impact of environmental issues but also see who is impacted the most. Many disadvantaged groups are disproportionately affected by environmental problems. For example, low-income people of color are more heavily exposed to air pollution [Jbaily, 2022]. This project seeks to show a statistically significant relationship between environmental factors (including exposure to particulate matter 2.5, ozone, diesel particulate matter, air toxics cancer risk, air toxics respiratory hazard index, toxic releases to air, traffic proximity, lead paint, superfund proximity, RMP facility proximity, hazardous waste proximity, underground storage tanks, and wastewater discharge) and socioeconomic factors (including the percentage of low-income people, unemployment rate, percentage of people of color, percentage of households with limited English Speaking, percentage of people who have less than a high school education, percentage of people under age 5, and percentage of people over age 64).

According to the United States Environmental Protection Agency (EPA), environmental justice is "the just treatment and meaningful involvement of all people, regardless of income, race, color, national origin, Tribal affiliation, or disability, in agency decision-making and other Federal activities that affect human health and the environment." Enhancing environmental justice within solutions to environmental crises is paramount as it guarantees equitable protection for all individuals, regardless of background or socioeconomic status, from the detrimental impacts of environmental hazards. As early as 1991, grassroots movements deeply troubled by environmental crises had been advocating for justice to be an integral component of solutions. A notable instance is the fervent call for environmental justice voiced during the 1991 People of Color Environmental Leadership Summit (First National People of Color Environmental Leadership Summit, 1991). More recently, environmental and sustainability scientists have established environmental justice as a foundational principle of current and future scientific and policy solutions (Rockström, 2023).

Researchers have begun to use machine learning both to understand and explain trends in environmental injustice, as well as to inform policy-making. Ho et al., for example, used machine learning to analyze data from six metropolitan U.S. counties to measure how socioeconomic factors including race, renter-occupied housing, and elder population shape disparities in environmental hazard levels like urban heat, floods, and air pollution. They found that these socioeconomic factors vitally shape exposure to environmental hazards and their machine-learning data can be used to inform data-driven and analytics-informed urban development strategies and solutions (Ho et. al, 2023). Clark et al. also proposed using natural language processing (NLP) on data sets of previous legislation, accompanying texts like debate transcripts and news articles to provide context, and parliamentary procedures like election timeframes and parliamentary composition to predict voting outcomes for climate legislation decisions (Clark et al., 2023). This project utilizes data from EPA EJscreen, which is a mapping and screening tool developed by the EPA that has collected data on various environmental and demographic socioeconomic indicators throughout the United States. This data has already been used in some studies to investigate environmental exposures of various demographics. Kumar et al. focused on the data from Ohio's metropolitan areas and found that some of the above-stated environmental risks are significantly more prevalent in areas with higher proportions of low-income and minority populations than in areas with lower proportions of these groups (Kumar et. al, 2022).

My research delves into the data within EPA EJScreen to explore the interplay between socioeconomic indicators and environmental factors across all U.S. states. Employing advanced statistical methodologies such as correlation analysis, Pearson correlation coefficient, linear regression, XGBoost, and neural networks, I aim to uncover patterns that elucidate the relationship between socioeconomic status and the risk of exposure to environmental hazards. The overarching objective of this study is to identify regions where socioeconomic disparities contribute significantly to heightened susceptibility to environmental risks. By pinpointing states where environmental threats, spanning from climate change impacts to the prevalence of substandard housing, disproportionately affect low-income communities and people of color, our research endeavors to furnish policymakers with empirically driven insights while providing some historical context. By equipping decision-makers with data-driven evidence, we aspire to guide targeted interventions in states most urgently requiring attention amid escalating environmental challenges.

Materials and Methods

Data

The dataset utilized in this study was sourced from the EPA EJScreen, which comprehensively covers environmental and socioeconomic data across all U.S. states and territories, encompassing Washington D.C. and Puerto Rico. It adopts the US Census block groups as the fundamental geographic unit for analysis. Socioeconomic data was extracted from the Census Bureau's ACS 2017-2021 5-year Summary, while environmental data was aggregated from diverse sources, as illustrated in Figure 2. The dataset comprises 7 socioeconomic indicators and 13 environmental indicators, as delineated in Figure 1, which provides an overview of the socioeconomic metrics recorded within the EPA EJScreen.

Name of indicator	Definition
People of Color (P_PEOPCOLORPCT)	The percentage of people who are not non-Hispanic white-alone individuals.
Low Income (P_LOWINCPCT)	The percentage of the population whose income is less than or equal to twice the poverty level.
Unemployment Rate (P_UNEMPPCT)	The percentage of people who did not have a job, were looking for a job, and were available for work.
Limited English-speaking Households (P_LINGISOPCT)	The percentage of households where all individuals aged 14 and older are unable to speak English proficiently.
Less than High School Education (P_LESSHSPCT)	The percentage of people who are aged 25 or over and do not have a high school diploma.
Under Age 5 (P_UNDER5PCT)	The percentage of people under age 5.
Over Age 64 (P_OVER64PCT)	The percentage of people over age 64.

Figure 1. Table of socioeconomic indicators (U.S. Environmental Protection Agency, 2023).

The environmental indicators encompassed in this dataset serve to gauge residents' exposure to a range of pollutants. These metrics offer direct assessments of recognized hazardous pollutants present within the environment of a census block, alongside indirect measurements that account for exposure amplifiers like proximity to motor traffic.

Name of indicator	Source	Definition
PM 2.5 (P_PM25)	2019 data from EPA's Office of Air Quality Planning and Standards (OAQPS)	The annual average concentration of inhalable particles 2.5 micrometers or smaller in air.
Ozone (P_OZONE)	2019 data from OAQPS	The annual mean of the 10 highest MDA8 O_3 concentrations.

Diesel PM (P_DSLPM)	2019 data from OAQPS	The estimated concentration of diesel particulate matter.
Air Toxics Cancer Risk (P_CANCER)	2019 data from OAQPS	The estimated lifetime inhalation cancer risk from analyzed carcinogens in ambient outdoor air over a 70-year lifetime.
Air Toxics Respiratory HI (P_RESP)	2019 data from OAQPS	The Respiratory Hazard Index from analyzed toxic pollutants in ambient outdoor air.
Toxic Releases to Air (P_RSEI_AIR)	EPA's Office of Pollution Prevention and Toxics (OPPT) from 2021 Risk-Screening Environmental Indicators (RSEI) modeled results	The relative potential human health impacts of chemicals released into the air by facilities listed on the list of toxic chemicals established under section 313 of the Emergency Planning and Community Right-to-Know Act (EPCRA).
Traffic Proximity and Volume (P_PTRAF)	2020 Highway Performance Monitoring System (HPMS)	The annual average daily traffic (AADT) count divided by distance in meters from the Census block centroid within 500 meters.
Lead Paint (P_LDPNT)	Census Bureau's ACS 2017-2021 5-year Summary	The percentage of occupied housing units built before 1960.
Superfund Proximity (P_PNPL)	Superfund Enterprise Management System (SEMS) database, November 3, 2022	The total count of sites proposed and listed on the National Priorities List in each block group within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group.
RMP Facility Proximity (P_PRMP)	EPA's Facility Registry Service (FRS) by selecting active facilities included in the RMP National Program System on October 31, 2022	The total count of active RMP facilities in each block group within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group.
Hazardous Waste Proximity (P_PTSDF)	Treatment, Storage, and Disposal Facilities (TSDFs) from RCRAInfor and Large Quantity Generators (LQGs) from the 2021 Biennial Reports (BR) on February 9, 2023	Hazardous waste proximity, the total count of hazardous waste facilities in each block group within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group.
Underground Storage Tanks (P_UST)	EPA's Office of Underground Storage Tanks on February 2, 2023	Underground storage tanks (UST) and leaking UST (LUST), the relative risk of being affected by a LUST for a block group, derived by the weighted sum of active LUSTs and the

		sum of active and temporarily out-of-service USTs within a certain distance from a block group.
Wastewater Discharge (P_PWDIS)	EPA's OPPT on November 23, 2022 from 2020 RSEI modeled results	Wastewater discharge, the block group's relative risk of exposure to pollutants in downstream water bodies, derived using toxicity-weighted concentrations in stream reach segments within 500 meters of a block centroid, divided by distance in meters, presented as the population-weighted average of blocks in each block group.

Figure 2. Table of environmental indicators (U.S. Environmental Protection Agency, 2023).

Some of the environmental indicator data were aggregated at the census tract level and then extrapolated for each block group within that tract. Additionally, there were instances of missing values, which I addressed in my analysis by excluding any Census block containing missing data. Furthermore, EJScreen provided a percentile representation of each indicator, ranging from 0 to 100%, where the 50th percentile denoted the median value for that indicator. I opted to utilize the percentile version of each indicator during data analysis.

Machine Learning Methods

In my experiments, I employed various analytical techniques to scrutinize the dataset. Specifically, I utilized the Pearson correlation coefficient, linear regression, extreme gradient boosting (XGBoost), and neural networks to analyze the data and extract meaningful insights.

Linear Regression

Linear regression makes predictions by creating a line that minimizes a loss function for the data. This loss function that I used was the root mean squared error (RMSE), which is the square root of the mean of the squares of the differences between the predicted values and the actual values. I used the Pearson correlation coefficient to measure the linear correlation between one environmental indicator and one socioeconomic indicator.

Decision Trees

Decision trees are a nonparametric supervised method, capable of handling both classification and regression tasks. The model is structured in a hierarchical tree, featuring a root node, branches, and leaf nodes. The leaf nodes of a decision tree enumerate all possible outcomes from the model. In regression tasks, the leaf nodes provide scores. Utilizing XGBoost, I employ the paradigm of gradient boosting, as defined by Friedman, to optimize the search within the function space of trees (Chen and Guestrin, 2016; Friedman, 2001).

Neural Networks

Neural networks are composed of layers of interconnected computation nodes, which can be used to predict a wide range of outcomes or classifications by discerning intricate patterns within input data (LeCun, Bengio and Hinton, 2015). Their strength lies in effectively handling large datasets and capturing nonlinear relationships, making them invaluable tools in predictive modeling. I used a multilayer perceptron (MLP), which is a fully connected multi-layer neural network. In this project, I used MLPs to predict environmental indicators based on socioeconomic indicators.

Experimental Setup

I first separated and organized the EJscreen data by the different states and territories. Next, I paired each socioeconomic indicator with every environmental indicator and assessed each pair's correlation within each state using the Pearson correlation coefficient. This enabled me to identify the pair of one socioeconomic factor and one environmental factor with the greatest correlation within each state or territory. Subsequently, I sorted the states based on the value of that correlation. I used the machine learning methods, which were linear regression, XGboost, and neural networks, on the five states with the highest correlation (Rhode Island, Illinois, Indiana, Wisconsin, New York) and five states with the lowest correlation (South Dakota, Montana, New Mexico, Puerto Rico, and Wyoming) to predict one environmental factor(Diesel particulate matter) using all of the socioeconomic indicators.

Results

This study uncovered varying degrees of linear correlation between environmental and socioeconomic factors across different states. Within the comprehensive list of states (excluding Guam, Northern Mariana Islands, American Samoa, and U.S. Virgin Islands due to incomplete data), correlation values ranged from 0.202 to 0.693. Figure 3 displays the ten states exhibiting the lowest correlation between these factors, whereas Figure 4 illustrates the ten states with the highest correlation.

Name	Factors	Correlation
South Dakota	[P_UST, P_LOWINCPCT]	0.202078
Montana	[P_UST, P_LOWINCPCT]	0.278391
New Mexico	[P_LDPNT, P_LOWINCPCT]	0.282355
Puerto Rico	[P_LDPNT, P_LOWINCPCT]	0.284978
Wyoming	[P_DSLPM, P_PEOPCOLORPCT]	0.292056
Hawaii	[P_UST, P_LINGISOPCT]	0.296153
North Dakota	[P_LDPNT, P_OVER64PCT]	0.314342
Maine	[P_LDPNT, P_LOWINCPCT]	0.328154
Idaho	[P_LDPNT, P_LOWINCPCT]	0.330361
South Carolina	[P_LDPNT, P_LOWINCPCT]	0.336016

Figure 3. Ten states with the lowest correlation between environmental and socioeconomic factors

Name	Factors	Correlation
Rhode Island	[P_DSLPM, P_PEOPCOLORPCT]	0.693379
Illinois	[P_DSLPM, P_PEOPCOLORPCT]	0.657355
Indiana	[P_DSLPM, P_PEOPCOLORPCT]	0.627712
Wisconsin	[P_DSLPM, P_PEOPCOLORPCT]	0.620256
New York	[P_DSLPM, P_PEOPCOLORPCT]	0.607241
Connecticut	[P_PTSDF, P_PEOPCOLORPCT]	0.605291
Kentucky	[P_PTSDF, P_PEOPCOLORPCT]	0.587363
Missouri	[P_DSLPM, P_PEOPCOLORPCT]	0.576977
Ohio	[P_DSLPM, P_PEOPCOLORPCT]	0.573925
Pennsylvania	[P_DSLPM, P_PEOPCOLORPCT]	0.572905

Figure 4. Ten states with the highest correlation between environmental and socioeconomic factors.

Upon examining the pairs of socioeconomic factors exhibiting the strongest correlation within each state, I observed a pattern where certain indicators recurred more frequently than others. To visually represent this trend, histograms were employed to depict the frequency of occurrence for each indicator.

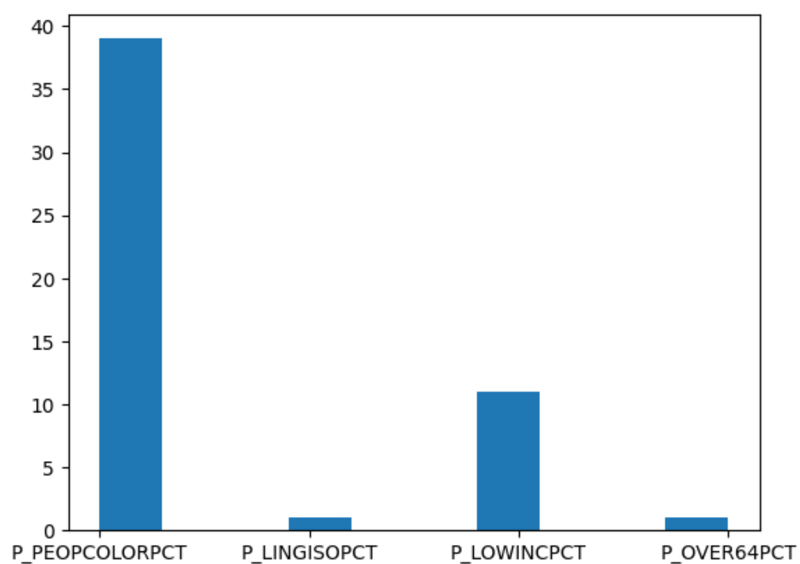

Figure 5. This histogram illustrates the frequency of socioeconomic factors within the list of attributes showing the highest correlations between one socioeconomic factor and one environmental factor in a specific state or territory.

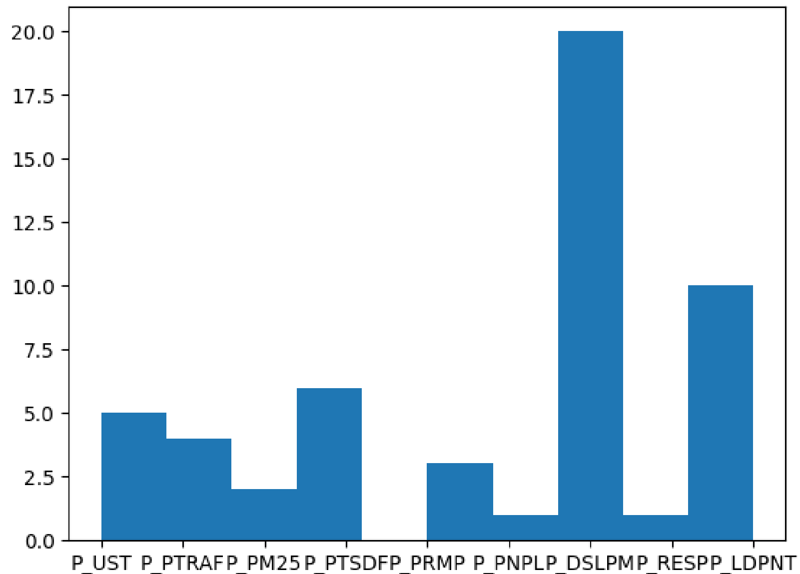

Figure 6. This histogram illustrates the frequency of environmental factors within the list of attributes showing the highest correlations between one socioeconomic factor and one environmental factor in a specific state or territory

Following the application of machine learning methods to analyze the data from the five states exhibiting the highest correlation and the five states with the lowest correlation, I computed the R-squared scores to evaluate the predictive performance of each method. Interestingly, I observed that the R-squared values tended to be lower for states characterized by lower correlations, contrasting with states exhibiting higher correlation levels.

Name	r2 Score
Rhode Island	0.426450
Illinois	0.558832
Indiana	0.417888
Wisconsin	0.425281
New York	0.487609

Figure 7. The r-squared scores of the predictions made by linear regression on the five states with the highest correlation.

Name	r2 Score
Wyoming	0.119569
Puerto Rico	0.151060
New Mexico	0.043657
Montana	0.126536
South Dakota	0.086798

Figure 8. The r-squared scores of the predictions made by linear regression on the five states with the lowest correlation.

Name	r2 Score
Rhode Island	0.307164
Illinois	0.561343
Indiana	0.347086
Wisconsin	0.397582
New York	0.581807

Figure 9. The r-squared scores of the predictions made by XGboost on the five states with the highest correlation.

Name	r2 Score
Wyoming	-0.249078
Puerto Rico	-0.031374
New Mexico	-0.031273
Montana	-0.061037
South Dakota	-0.123455

Figure 10. The r-squared scores of the predictions made by XGboost on the five states with the lowest correlation.

Name	r2 Score
Rhode Island	0.370665
Illinois	0.600019
Indiana	0.413659
Wisconsin	0.455097
New York	0.602328

Figure 11. The r-squared scores of the predictions made by a neural network on the five states with the highest correlation.

Name	r2 Score
Wyoming	0.026698
Puerto Rico	0.102071
New Mexico	-0.016749
Montana	0.045022
South Dakota	-0.099217

Figure 12. The r-squared scores of the predictions made by a neural network on the five states with the lowest correlation.

Discussion

People of Color vs. Diesel Particulate Matter

In my findings regarding the pair of environmental and socioeconomic indicators with the highest correlation value within each state, I discovered that "People of Color" and "Diesel Particulate Matter" emerged as a remarkably frequent pairing, occurring twenty times across the dataset. Tigue explains that many highways have been placed through Black and Brown neighborhoods, increasing traffic like freight trucks and school buses which burn large amounts of diesel fuel (Tigue, 2021). During the 1950s and 1960s, after the enactment of the Federal Highway Act of 1956, local politicians and business leaders capitalized on federal funding, which covered a substantial 90% of highway costs, to carry out what were termed "urban renewal" projects. In these initiatives, numerous black neighborhoods were systematically demolished, and historically Black communities were often cleaved from white neighborhoods by the construction of highways. Urban historian Raymond Mohl has highlighted instances such as Interstate 95's route through Overtown, a Black neighborhood in Miami, and Interstate 40's division of a Black community in north Nashville, as well as instances in New Orleans and Kansas City where freeways were redirected from white neighborhoods through predominantly Black areas. Consequently, a significant portion of Black residents in urban areas found themselves residing near highways, where large amounts of diesel particulate matter are dispersed (Sullivan, 2021).

Research by Park et al. underscores significant racial disparities in exposure to air pollution, stemming from systemic discrimination in housing, public transportation, and environmental policies. This disproportionately impacts Black residents in inner cities compared to predominantly white neighborhoods in the suburbs. For instance, in Atlanta, Georgia, 23.7% of Black residents reside within 200 meters of a major road between 12-3 pm, while only 17.1% of white residents do. During the hours of 9 pm to 3 am, 12.2% of Black residents are situated near major roads, in contrast to just 6.4% of white residents (Park and Kwan, 2020).

The prevalence of "People of Color" and "Diesel Particulate Matter" as a recurrent pairing underscores the systemic inequities ingrained within urban development, particularly evident in the historical placement of highways through Black and Brown neighborhoods. This perpetuates environmental injustice, amplifying exposure to harmful pollutants such as diesel particulate matter, disproportionately affecting communities of color, and highlighting the urgent need for equitable policy interventions.

People of Color vs. Particulate Matter 2.5 (Nevada and Arizona)

Another pair of indicators that appeared were People of Color and PM 2.5, which was the pair with the highest correlation in Nevada and Arizona. Rowland-Shea et al. find that in the United States, communities of color are three times more likely than white communities to live in nature-deprived areas (Rowland-Shea et. al, 2020). Recent journalism has suggested that Black, brown, and low-income communities have significantly fewer trees than wealthier, whiter communities. In Southern Nevada, for example, some suggest that 560,000 trees would need to be planted to reach “tree equity,” the number of trees that all residents can benefit from carbon reductions. According to a Nevada newspaper, trees provide shade to just 1% of the Sunrise Manor neighborhood in North Las Vegas. This area, characterized by a poverty rate of 71% and a population composed of 82% people of color, experiences limited tree coverage (Solis, 2021). Nowak et al. find that trees and forests, especially in urban areas, cause many positive health impacts, such as reducing mortality because trees intercept particulate matter and absorb pollutants. The lack of green spaces in low-income neighborhoods and those predominantly occupied by people of color plays a significant role in the heightened levels of particulate matter 2.5 (PM2.5) observed in these communities. Greenery, known for its ability to absorb pollution and particulate matter, is often insufficient or lacking altogether in areas predominantly occupied by people of color (Nowak et. al, 2014). Hence, the higher concentrations of PM2.5 in these neighborhoods can be attributed to the scarcity of green spaces, exacerbating environmental health disparities.

Low Income vs. Lead Paint

Lead paint is widely recognized as a hazardous substance, yet it continues to be used in the construction and renovation of many homes. Fortunately, increasing awareness and evolving legislation at the state level are driving the adoption of safer alternatives, such as titanium dioxide, which poses substantially fewer health risks.

Although the federal government banned the consumer use of lead-based paint in 1978, its legacy persists in countless homes, often buried beneath layers of newer coatings. The real danger arises when lead-based paint deteriorates, manifesting in peeling, chipping, chalking, cracking, or becoming damaged or damp, releasing toxic particles into the environment (U.S. Environmental Protection Agency, 2024). While this issue may not directly stem from broader climate change phenomena, the persistent presence of lead paint in residences inhabited by people of color exacerbates their vulnerability to health risks linked to environmental hazards.

Statistics reveal the extent of the issue: approximately 24% of homes constructed between 1960 and 1977, 69% of those built between 1940 and 1959, and a staggering 87% of homes constructed before 1940 contain lead-based paint. As a result of historic redlining and discriminatory lending practices, approximately 60 percent of Black Americans live in places that were redlined before 1968 and have subsequently declined in value. A study spanning from 1995 to 2013 revealed a consistent trend: Black children between the ages of one and five consistently exhibited higher blood lead levels in comparison to their white peers (Sampson and Winter, 2016). Another study found that as recently as 2006, 28 percent of Black households were exposed to housing-related lead risks, in contrast to 20 percent of white families (U.S. Department of Housing and Urban Development, 2011). Addressing this widespread problem is crucial for safeguarding public health and ensuring the safety of residential environments.

Conclusions

In this study, notable states such as Rhode Island, Illinois, Wisconsin, and New York exhibited statistically significant correlations between environmental factors and socioeconomic indicators. This phenomenon may be attributed to historical patterns of systemic segregation, classism, and racism in policymaking. These practices have disproportionately placed low-income individuals and people of color in neighborhoods with heightened exposure to environmental hazards, including lead paint, particulate matter, and diesel emissions. Furthermore, in states such as South Dakota, Montana, Puerto Rico, New Mexico, and Wyoming, the relationship between environmental factors and socioeconomic indicators appeared to be less pronounced.

Further research is warranted to elucidate the factors contributing to the observed variations among states in terms of the relationships between environmental and socioeconomic indicators. Potential avenues for exploration include geographical features such as proximity to large bodies of water, as well as demographic factors such as whether a state's population is predominantly rural or urban. The considerable diversity in correlation and R-squared values across states suggests that certain state-specific characteristics may influence the degree of association between environmental and socioeconomic factors, warranting further analysis.

This project has several limitations worth noting. Firstly, focusing solely on state-level data may overlook nuanced relationships present at finer geographical scales, such as cities or counties, as well as distinctions between urban and rural areas. Secondly, the reliance on correlation analysis for much of the investigation poses a limitation, as correlation does not inherently imply causation and therefore cannot fully elucidate cause-and-effect relationships.

In conclusion, as environmental challenges like climate change and air pollution persist, it becomes imperative to not only assess their overall impact but also to understand who bears the brunt of these issues. Disadvantaged groups, such as low-income individuals and people of color, are disproportionately affected by environmental problems, experiencing heightened exposure to pollutants and other hazards. This project aimed to explore the relationship between environmental factors and socioeconomic indicators across all U.S. states, utilizing data from the EPA EJScreen.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Chen, T., & Guestrin, C. (2016, August). XGBoost: A Scalable Tree Boosting System. In KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. ACM. https://doi.org/10.1145/2939672.2939785

Clark, J., Wan, M., & Santos-Rodríguez, R. (2023). Understanding Climate Legislation Decisions with Machine Learning. In Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023. https://research-information.bris.ac.uk/ws/portalfiles/portal/384223295/CCAI_NeurIPS_23_Climate_Legislation.pdf.

Environmental Justice Network. (1991, October). Principles of Environmental Justice. Retrieved February 25, 2024, from https://www.ejnet.org/ej/principles.html

First National People of Color Environmental Leadership Summit (1991, October). Principles of Environmental Justice. Retrieved February 25, 2024, from https://www.ejnet.org/ej/principles.html

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189–1232.

Ho, Y.-H., Liu, Z., Lee, C.-C., & Mostafavi, A. (2023). ML4EJ: Decoding the Role of Urban Features in Shaping Environmental Injustice Using Interpretable Machine Learning. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4598975.

Jbaily, A., Zhou, X., Liu, J., Lee, T. H., Kamareddine, L., Verguet, S., & Dominici, F. (2022). Air pollution exposure disparities across US population and income groups. Nature, 601(7892), 228–233. https://doi.org/10.1038/s41586-021-04190-y

Kumar, A., Kuruppuarachchi, L. N., & Madiraju, S. V. H. (2022). An Application of EJSCREEN for the Examination of Environmental Justice in Metropolitan Areas of Ohio, USA. In G. Venkatesan, S. L. Prabu, & M. Rengasamy (Eds.), Sustainability Studies: Environmental and Energy Management (pp.112–128. BENTHAM SCIENCE PUBLISHERS. https://doi.org/10.2174/9789815039924122010008

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Nowak, D. J., Hirabayashi, S., Bodine, A., & Greenfield, E. (2014). Tree and forest effects on air quality and human health in the United States. Environmental pollution (Barking, Essex: 1987), 193, 119–129. https://doi.org/10.1016/j.envpol.2014.05.028

Park, Y. M., & Kwan, M. P. (2020). Understanding Racial Disparities in Exposure to Traffic-Related Air Pollution: Considering the Spatiotemporal Dynamics of Population Distribution. International journal of environmental research and public health, 17(3), 908. https://doi.org/10.3390/ijerph17030908

Rockström, J., Gupta, J., Qin, D., Lade, S. J., Abrams, J. F., Andersen, L. S., Armstrong McKay, D. I., Bai, X., Bala, G., Bunn, S. E., Ciobanu, D., DeClerck, F., Ebi, K., Gifford, L., Gordon, C., Hasan, S., Kanie, N., Lenton, T. M., Loriani, S., Liverman, D. M., … Zhang, X. (2023). Safe and just Earth system boundaries. Nature, 619(7968), 102–111. https://doi.org/10.1038/s41586-023-06083-8

Rowland-Shea, J., Doshi, S., Edberg, S., & Robert, F. (2020, July 21). The Nature Gap. Center for American Progress. https://www.americanprogress.org/article/the-nature-gap/

Sampson, R. J., & Winter, A. S. (2016). The Racial Ecology of Lead Poisoning: Toxic Inequality in Chicago Neighborhoods, 1995-2013. Du Bois Review 13(2), 261–283. https://doi.org/10.1017/S1742058X16000151

Solis, J. (2021, July 27). Southern Nevada needs half a million trees to reach 'tree equity.' Nevada Current. https://nevadacurrent.com/2021/07/27/southern-nevada-needs-half-a-million-trees-to-reach-tree-equity/

Sullivan, A. (2021, May 25). U.S. freeways flattened Black neighborhoods nationwide. Reuters. https://www.reuters.com/world/us/us-freeways-flattened-black-neighborhoods-nationwide-2021-05-25/

Tigue, K. (2021, October 7). Diesel Emissions in Major US Cities Disproportionately Harm Communities of Color, New Studies Confirm. InsideClimate News. https://insideclimatenews.org/news/27102021/diesel-pollution-environmental-justice/

U.S. Department of Housing and Urban Development. (2011, April). American Healthy Homes Survey Lead and Arsenic Findings. https://www.hud.gov/sites/documents/AHHS_REPORT.PDF

U.S. Environmental Protection Agency (EPA). (2024). Protect Your Family From Sources of Lead. Washington, D.C. https://www.epa.gov/lead/protect-your-family-sources-lead.