Urban Heat Island Disparities

Introduction

Urban areas experience higher temperatures than their rural counterparts, which is referred to as the urban heat island (UHI) effect (cite). Urbanization alters land cover, replacing natural landscapes with impervious surfaces (cite). For example, parking lots and roadways constructed with asphalt and concrete may replace trees and other vegetation. However, impervious surfaces have low land surface albedo, so they absorb more energy from solar radiation and re-emit it as heat (cite). In contrast, natural surfaces, such as vegetation and water bodies, reflect solar radiation, shade man made surfaces, and release moisture, which can cool the local microclimate. Moreover, building infrastructure, vehicles, and local industry generate heat and greenhouse gas emissions, contributing to the UHI effect (cite). As a result, the UHI effect can be more pronounced in more densely populated areas (cite). Concerningly, the UHI effect is expected to exacerbate as urban areas continue to expand in area and population and modify land cover (cite), impacting at least 83 percent of the US population (cite). In general, increasing heat stress as a result of climate change leaves urban residents vulnerable to a myriad of hazardous health outcomes, including heat stroke, hospitalization, and mortality. In fact, heat mortality is the leading cause of weather-related death, and it is an increasing concern in U.S. cities due to climate change (cite).

Emerging research demonstrates that across the United States marginalized groups overwhelmingly face higher burdens of heat stress, exacerbated by the built environment, with respect to the distributions of the urban heat island effect and intra-urban heat. One study of UHIs in 175 contiguous U.S. cities found that in 169 cities, or 96.5% of the studied cities, people of color on average reside in census tracts with a greater surface urban heat island intensity (SUHII) (cite). In addition, the study found that on average those below the federal poverty line live in higher places with greater SUHII intensity compared to those above twice the poverty line (cite). Furthermore, researchers have sought to examine the extent to which the built environment plays a role in SUHII intensity disparities. A study, representing 300 million people living in U.S. urban areas, found that low-income and less educated tracts face higher surface heat than their more affluent and educated counterparts; in addition, Black, Hispanic, and Asian communities experience higher SUHI intensity than non-Hispanic white communities, even after controlling for income (cite). These widespread differences occurred due to greater built-up intensity, less vegetation, and to a smaller degree population density (cite). Moreover, a large-scale study on intra-urban heat in 108 historically redlined cities found that in 94 percent of cities, redlined areas possessed higher land surface temperatures than their non-redlined counterparts, coinciding with lower percentages of tree canopy in redlined area (cite). Nationally, redlined areas experienced 2.6 degrees Celsius higher land surface temperatures on average (cite). Further, a study, accounting for 167 million people in urban areas, found that urban temperature disparities co-occur with tree cover differences (cite). That is, low income census blocks in urban areas have fewer trees and warmer microclimates (cite). The average low-income block contained 15.2 percent fewer trees while being 1.5^{0}C warmer compared to blocks with higher income (cite). These widespread disparities are of particular concern since the burden of adverse heat-related public health outcomes disproportionately impacts marginalized groups (cite, cite, cite, cite, cite, cite).

Goal

You are a quantitative analyst at the White House Council on Environmental Quality. Your team is currently working on a project about urban heat island disparities nationally. Your supervisor has asked you to study the environmental factors that contribute to SUHI disparities in marginalized communities. You suggest a machine learning approach. Machine learning enables us to develop a non-linear, non-parametric, and interpretable model, which helps us to compare several continuous variables at once. Given the continuous and non-linear nature of environmental and SUHII data, this approach comes in handy for your analysis.

You gather data from a previous study (cite) and the Climate and Economic Justice and Screening Tool to examine the extent to which environmental factors exacerbate SUHII disparities. The environmental variables include green space (i.e., Normalized Difference Vegetation Index), built-up extent (i.e., Normalized Difference Built-Up Index), and impervious surface (i.e., percent of a tract covered by impervious surface.). The demographic variables include race (Black, Hispanic/Latino, and White) and socioeconomic status (below poverty).

After gathering the data, it is your job to perform data exploration and model development. Your data exploration includes descriptive statistics, correlation analysis, and probability density estimation. Your model development focuses on developing a tree-based model and gaining insights through SHAP values and partial dependence plots. Lastly, you will summarize your findings for your team.

Data Exploration

Spearman Correlation Analysis

How does race and poverty correlate with environmental factors and SUHII? Which variables are most correlated with SUHII? How does this compare to the research discussed in the introduction? Describe any other interesting correlations.

Probability Density Estimation
What do the distributions reveal about race, socioeconomic status, environmental factors, and SUHII? How does this compare to the research discussed in the introduction?

Data Dissection
Divide the data using the low income indicator, race bins (e.g., 0-20%, 20-40%,...,80-100%), and environmental bins. Find the average SUHII for each bin. Is there a monotonic pattern? Count how much of the data falls into each bin. Feel free to use kernel density estimates as a tool to examine the different groups as well. Communicate your results.

Regional Examination
Compute the county and state average SUHII; county and state weighted average SUHII by race; and county and state weighted average SUHI by race and low income indicator. Which regions have the highest average SUHII? Which regions have the highest disparities?

Freestyle
Formulate a question that can be answered through data exploration. Perform 1-2 extra data exploration techniques to answer your question. Discuss your findings.

Model Development

Model Performance
RMSE is the standard deviation of the residuals. Based on the RMSE, how did the model perform? Based on the R^2, how well did the model fit the data? How does model performance impact your communication of the model to your team? Create a scatter plot of the residuals by race and poverty status. How did model performance vary across groups? Discuss your findings.

Predictive Factors of SUHII
Using the SHAP values, discuss the most predictive factors of SUHII. Why do you think these are the most predictive factors? How does this compare to the previous research discussed in the introduction?

Multi-Dimensional Analysis
Using SHAP scatter plots, discuss how racial composition of a tract and environmental factors impact SUHII (i.e., use environmental factors as a heat map by race). Do the same for poverty status and environmental factors. Explain the disparities that these plots illuminate. How does this compare to the previous research discussed in the introduction?

Summary

Summarize the findings for your team. Based on your analysis, how do environmental disparities by race and socioeconomic status influence SUHII?