Correlates and Policy Recommendations of DC's Urban Heat Island Effect

Evan Bowman

December 14, 2022

Introduction

A major discussion in today's politics are the consequences of climate change and shifting environmental trends if no meaningful attempt at mitigating atmospheric greenhouse gasses occurs. These consequences, on a meta level, range from higher mean temperatures, sea level rise, and changes in the ability to grow sustainable food. This poses large public health risks such as increases in heat related deaths and illness. This is cause for concern as about 25 percent of natural hazard mortality in the country is caused by heat exposure (Benz and Burney 2021). Furthermore, heat-related mortality in the USA causes roughly 1,500 more deaths annually than other severe weather events, and causes other health outcomes such as heat strokes, dehydration, loss of labor productivity, and decreased learning (Hsu et al. 2021). The potential increase in heat related health incidents requires a substantial need for policy to combat increases in mean temperatures.

Little discussion looks at the discrepancies these effects might have on urban populations. On the topic of increased global temperature averages, there is a need to look at the intersectionality of an already occurring phenomenon: urban heat island effects. This concept revolves around the key fact that urbanized city centers already create higher mean temperatures compared to their rural counterparts. Not only are these urban centers more likely to have higher mean temperatures than rural areas, there are clear sociodemographic discrepancies between populations within these cities. Few studies have been published analyzing these demographic discrepancies. However, in one report, research has found that roughly 77 percent of counties studied show that there are large differences in change of daytime temperature between census tracts with the highest and lowest Black population proportions. Furthermore, 81 percent of counties studied with lower difference in daytime temperature are majority non-Hispanic White. (Benz and Burney 2021). There is also empirical evidence that shows discrepancies in urban heat island effects targeting communities that are more impoverished. For example, a 2020 study found that there is a negative correlation of .88 between cities' urban heat island effects and income (Chakraborty et al. 2020). This is further cause for concern when looking at potential correlations between poverty and race in city census tracts.

These findings are consistent with research that shows majority minority populations in urban centers have less tree cover and larger proportions of impervious surfaces than majority non-Hispanic White populations. This shows a clear need for social policy that protects majority minority census tracts from being disproportionately affected by increased heat island effects. Research regarding the effects of urban heat island effects must be on a local level to adequately identify meaningful social demographic trends for areas within urban centers. These could be specific neighborhoods, districts, census tracts, etc. In the context of racial inequities in Washington DC, most research is done at the ward level. For this study, census tracts are chosen as sample observations to be able to identify which specific areas of DC's eight wards are most at risk of the urban heat island effect, and the demographic patterns that coincide with these risks.

Research Questions

What are reoccurring sociodemographic trends that are associated with higher heat sensitivity indexes? Who are the communities at a larger risk of heat-induced health events due to higher heat sensitivity index (HSI)? HSI is a metric that describes an individuals ability to adapt, cope, or recover from extreme heat and is explained by specific demographic and health variables of DC's census tracts. Identifying socioeconomic and health trends can assist policy makers in identifying the communities who would best benefit from policy implementation.

What are the environmental metrics that are correlated with the census tracts with the highest HEI? What environmental characteristics of certain census tracts create larger heat exposure indexes (HEI)? In the data, HEI is a metric that takes a census tract's mean ambient air temperature as the exposure variable (50 percent) and two physical variables (impervious surfaces: 25 percent, lack of tree cover: 25 percent) that describe a tract's heat retention. Understanding the environmental factors would help in identifying policy approaches that would mitigate the consequences of the urban heat island effect by assisting at risk census tracts with solutions to heat retention.

Do these findings highlight environmental justice issues that exist within Washington DC? Do the findings of this study concur with previous research regarding race inequities for heat exposure and sensitivity in urban centers? The ultimate goal in answering these two research questions is to find the intersectionality of environmental factors that exacerbate the urban heat island effect and the populations of urban centers most at risk.

Heat Sensitivity Index and Socioeconomic Patterns

Visualizing Patterns

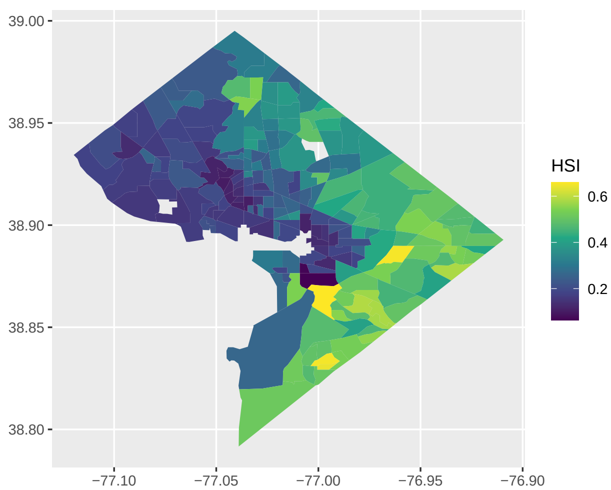


Fig. 1

As seen from the Figure 1, the census tracts with the highest HSI levels fall within areas along the Anacostia River, also known as Wards 7 and 8. These census tracts are a vast contrast when compared to census tracts with low HSI levels found in the Northwest portion of the city. Areas along the Anacostia River are typically known for being predominately majority minority as well as lower income compared to census tracts in the Northwest. This further contributes to the hypothesis of racial inequity in DC's Urban Heat Island Effect as the intersectionality of poverty and race exacerbates individuals ability to cope with extreme heat (Fig. 2). Analysis conducted in this section will look to define what specific patterns exist within tracts with high HSI levels.

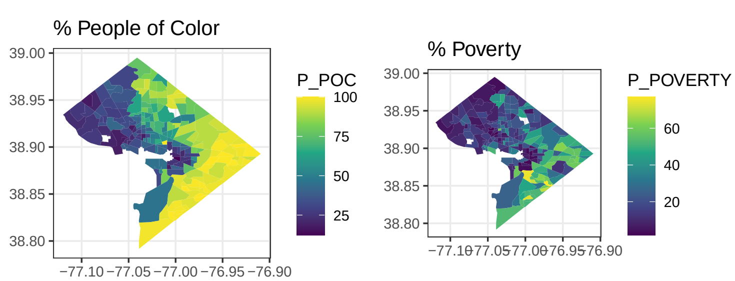


Fig. 2

Other interesting patterns found within the data are health trends and specific discrepancies between predominately non-Hispanic White census tracts and predominately Black census tracts. For all four health variables found within the data set, asthma, obesity, coronary heart disease (CHD), and disability, majority minority census tracts exhibit higher prevalence than predominately White tracts (Fig. 3). This highlights the increased risk of health related events that are exacerbated by a higher sensitivity to heat and the inability to cope or adapt to it.

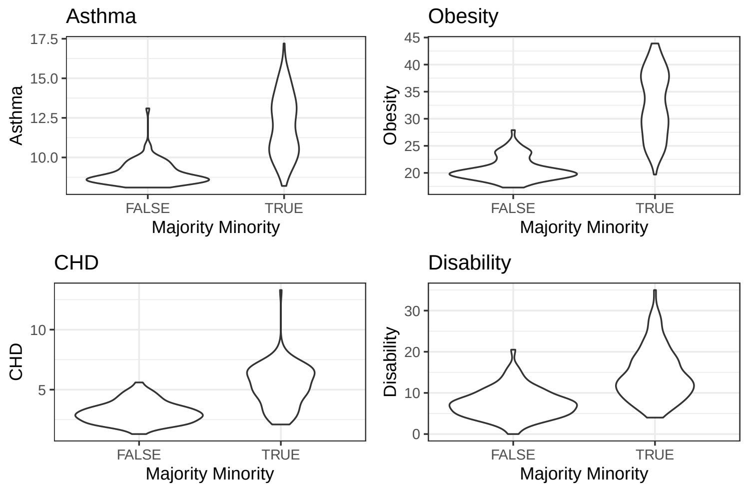


Fig. 3

With preliminary patterns relating to a census tract's Heat Sensitivity Index established, this paper will shift to creating a preliminary model that best exemplifies which variables best predict HSI for any given census tract.

Regression Definition

Preliminary Regression Identification

To best understand to what magnitude factors affect HSI levels, preliminary regression models will be created to identify variables of statistical significance. To achieve the best possible model, best subsets regression will be conducted and the best model will be taken following the analysis of specific model fitting metrics.

All possible models were fitted using K-fold cross validation on the full data. The number of folds were set to ten and fitting was iterated across all potential models. Mean cross-validation errors across all fold were then obtained to identify the model with the lowest RMSE.

Table 1: RMSE Error of Best Subsets

Variables	CV Error
1	0.0538886
2	0.0492850
3	0.0420452
4	0.0249455
5	0.0209644
6	0.0202020
7	0.0133773
8	0.0071626
9	0.0003290
10	0.0003236
11	0.0003244

Variables	CV Error
12	0.0003263
13	0.0003279

The model including ten variables was identified as the lowest mean RMSE across the ten folds. This model was then fitted with the whole data for coefficient identification (Table 2). This model, while not generative, can be useful in identifying the most important predictors of HSI for any given census tract. The inclusion of both P_POC and P_POV further contribute to the patterns found in initial data exploration. The inclusion of health variables makes intuitive sense as these illnesses would mitigate an individual's ability to adapt or deal with extreme heat. The statistical significance of P_TREECOVER is interesting as it is the only environmental metric that is included in the regression model. However, the idea of tree cover acting as a solution to overheating makes this seem plausible.

Table 2: Coefficients of Identified Best Subsets Model

Predictor	Coefficient
(Intercept)	-0.1944454
P_POC	0.0012666
P_CHILD	0.0056873
P_ELDERLY	0.0028569
P_POVERTY	0.0011316
P_DISABILITY	0.0031852
P_LIMENG	0.0031199
ASTHMA	0.0122109
CHD	0.0090995
OBESITY	0.0040330
P_TREECOVER	0.0003942

Shrinkage Regression

While best subsets regression identified the best model as including 10 variables, there appears to be covariation that may affect the magnitude of the coefficients (Appendix Tab. 6). As a result, a LASSO regression model will be fitted and compared to the regression model identified by the best subsets method. LASSO takes into account the covariation of the predictors, and adjusts the coefficients accordingly. The LASSO model was tuned using K-fold Cross Validation. Summary statistics of the LASSO models are given below.

Table 3: CV LASSO Results

	Lambda	Index	MSE	SE	Variables
min	0.0003943	64	6.119e-05	1.448e-05	9
1se	0.0015918	49	7.348e-05	1.738e-05	8

There are two possible lambda values that the model could use to tune: the minimum or one standard error. Lambda min tends to over fit the data as exemplified by an extra predictor variable when compared to that of the 1se method. The 1se method gives the most regularized model while ensuring that the CV error is within one standard error of the minimum. As a result, there are differences between the two models besides the exclusion of P_TREECOVER in the one standard error lambda value. The difference between the magnitude of shrinkage of variables given certain lambda values is interesting and highlights the robustness of LASSO regression given different lambda values that differ by a very small factor. (Appendix Fig.3).

Further analysis of the different lambda values can occur in the future to further explore the robustness of LASSO.

Table 4: Coefficients of LASSO Models

	Min	1 se
(Intercept)	-0.1874	-0.1753
P_POC	0.0016	0.0016
P_CHILD	0.0060	0.0057
P_ELDERLY	0.0023	0.0021
P_POVERTY	0.0010	0.0010
P_DISABILITY	0.0037	0.0036
P_LIMENG	0.0030	0.0027
ASTHMA	0.0186	0.0177
CHD	0.0129	0.0135
P_TREECOVER	-0.0018	0.0000

Also, by obtaining the MSE of the two lambda values we can compute the RMSE and compare to that of the best subsets method. The comparison of the RMSE and MSE of the three models highlights that the best subsets model has the best predictive power. However, it appears that all three have relatively good predicting power. This model, while not generative, offers the framework of identifying significant predictors of HSI levels for census tracts.

Table 5: Prediction Metrics of LASSO Models

	Best	Lambda Min	Lambda 1se
RMSE	0.0003236	0.0078224	0.0085720
MSE	0.0000001	0.0000612	0.0000735

Environmental Factors of Heat Exposure Indexes

It is important to understand not only socioeconomic factors that contribute to a census tract's HSI statistic, but also the environmental factors that contribute to their Heat Exposure Index. However, unlike HSI, HEI is defined for each census tract as 50 percent mean air temperature, 25 percent lack of tree canopy, and 25 percent impervious surface. As a result, analysis will be centered around patterns of the three variables as opposed to any model building to predict HEI values.

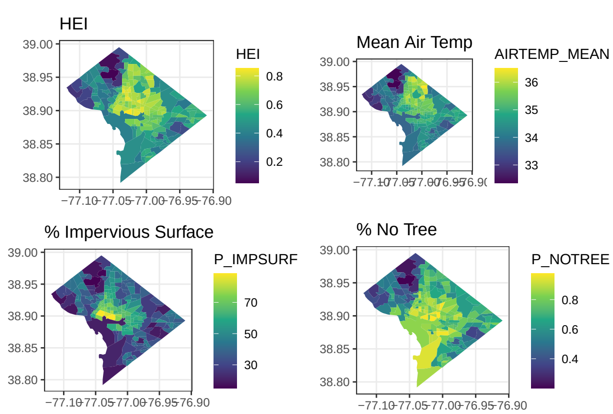


Fig. 4

As shown by Fig. 4, there are fewer patterns in discrepancies between environmental factors and census tracts compared to that of the HSI data. Predominately Black census tracts are not disproportionately affected by any of the environmental variables (Fig. 5). These maps highlight areas such as downtown being at the highest risk of heat exposure. This makes intuitive sense as this is the city center of DC which includes high amounts of impervious surface and lacks cooling characteristics such as large parks or trees. When compared to areas of interest in the HSI section, census tracts that were poorer and majority minority are relatively similar to more White Census tracts. However, given the urban nature of the city, proportions of census tracts that are impervious surfaces and the lack of tree cover is a cause for concern given the socioeconomic patterns discussed previously. This creates the need for both environmental and social policy to adequately mitigate the risk of heat induced health incidents.

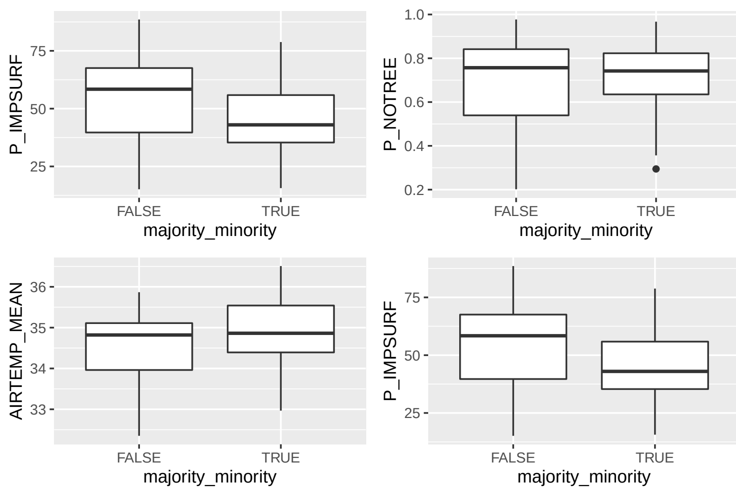


Fig. 5

Potential Policy Recommendations

Environmental Policy Recomendations

Policy recommendations revolving around mitigating mean air temperature of census tracts is centered around both impervious surfaces and tree canopy cover. Both of these factors are main contributors of heat retention due to the inability to deflect solar rays. Advancing infrastructure improvements such as cool roofs or cool pavements is key in reducing mean air temperature. Infrastructure projects such as these would need to prioritize census tracts that are majority minority or poor. These areas are at most need of infrastructure improvements already and would benefit far more than other tracts due to their increased sensitivity to heat. Utilizing the predictive model could also help identify what census tracts should be prioritized in policy implementation.

Social Policy Recomendations

Policy recommendations relating to socioeconomic factors must center around enabling individuals in census tracts with high HSI metrics to be able to adapt and cope to instances of extreme heat. One clear avenue is to strengthen DC's Heat Emergency Plan (HEP) which opens resources when temperature exceed 95 degrees Fahrenheit. These resources are deemed cooling centers and range from pools, libraries, shelters, etc. located across the city (Appendix Fig. 2). However, the city must be able to give clear and concise information regarding where these centers are located and the hours they are open. Furthermore, the city must be able to offer more cooling centers that are accessible during non-business hours. As of right now, the city only advertises four options that are open 24/7 in cases of extreme heat. While temperatures do decrease during the night hours, individuals with low heat tolerance are still at adverse risk of heat induced illness Opening more accessible cooling centers in census tracts with high HSI will drastically reduce the possibility of heat induced health incidents.

Conclusions and Future Steps

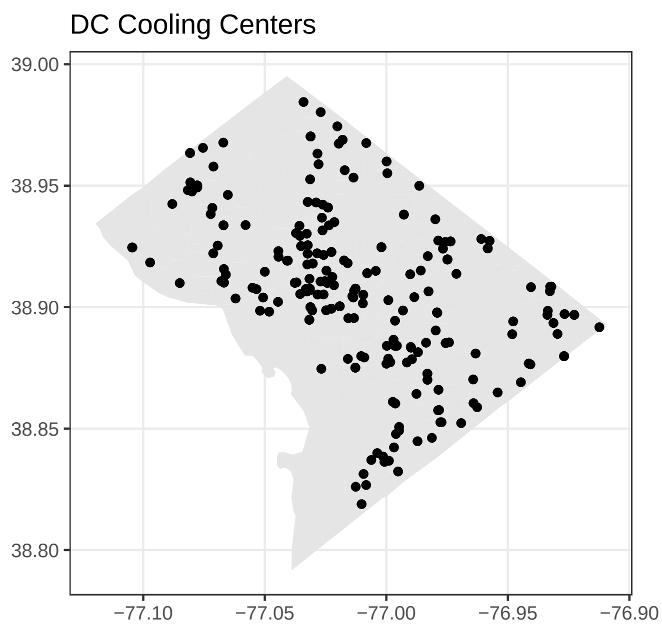
The scope of the findings of this study lends itself to the first and third research questions posed at the beginning of this report. Best subsets and LASSO modeling using K-fold Cross Validation identified specific variables that are statistically significant in predicting a census tracts HSI level. From this modeling, and exploratory data analysis, it was found that specific census tracts with higher proportions of socioeconomic groups such as people of color and poverty were more likely to have harder times adapting and coping with extreme heat. Not only were race and poverty deemed to be statistically significant, but also health metrics such as asthma, obesity, and CHD. Through correlation tables and mapping visuals, it was also identified that there are clear patterns between some of these predictors such as census tracts with a higher proportion of minority populations are also more likely to have higher proportions of poverty. These findings highlight clear demographic discrepancies that allow us to view DC's Urban Heat Island Effect as an environmental justice issue.

The study failed to adequately answer the second question of the report due the inability to create a generative model regarding the causes of heat exposure. This was a difficult task due definition of HEI, which is simply 50 percent mean air temperature, 25 percent percent-impervious surfaces, and 25 percent percent-no tree cover. This would require any meaningful model aimed at identifying the causes of heat to utilize mean air temperature as the response. In the case of this study, that would only leave percent tree cover, percent no tree cover, and percent impervious surfaces as potential predictor variables. The small quantity of variables and the inverse relationship between percent tree cover and percent no tree cover drastically reduced the possibility of any meaningful model being created. Future analysis for this research question needs to be centered around finding statistically significant environmental variables that explain the causes of extreme heat. Further data is needed to offer any sound conclusions to this question.

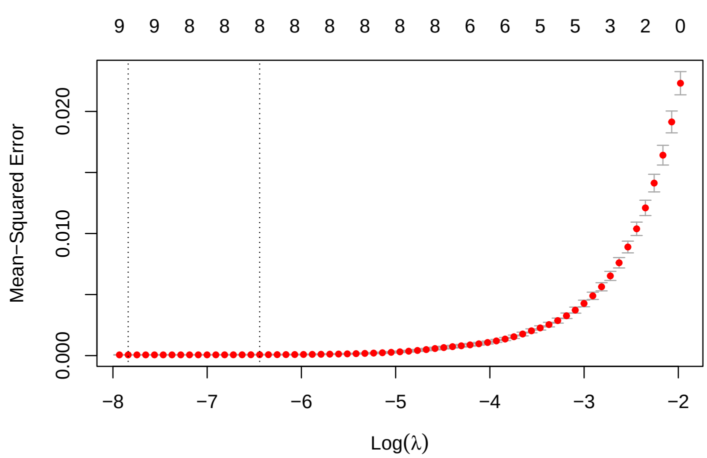
Appendix

Table 6: Correlation table of HSI variables

	P_POC	P_CHI	P_ELD	P_POV	P_DIS	P_LIM	ASTHMA	CHD	OBESE	HSI	P_TREE	P_NOTREE	P_IMPSURF	AIRTEMP
P_POC	1.00	0.35	-0.03	0.76	0.68	0.17	0.86	0.75	0.93	0.93	-0.01	0.01	-0.36	0.11
P_CHI	0.35	1.00	-0.06	0.32	0.12	-0.02	0.38	0.29	0.40	0.47	0.12	-0.12	-0.30	0.00
P_ELD	-0.03	-0.06	1.00	-0.13	0.24	-0.04	-0.11	0.40	-0.05	0.15	0.39	-0.39	-0.39	-0.39
P_POV	0.76	0.32	-0.13	1.00	0.63	0.10	0.89	0.61	0.82	0.84	-0.09	0.09	-0.26	-0.01
P_DIS	0.68	0.12	0.24	0.63	1.00	0.00	0.65	0.65	0.71	0.77	-0.02	0.02	-0.32	-0.06
P_LIM	0.17	-0.02	-0.04	0.10	0.00	1.00	-0.02	0.01	0.03	0.15	-0.01	0.01	0.09	0.15
ASTHMA	0.86	0.38	-0.11	0.89	0.65	-0.02	1.00	0.69	0.93	0.90	-0.02	0.02	-0.38	-0.02
CHD	0.75	0.29	0.40	0.61	0.65	0.01	0.69	1.00	0.78	0.84	0.25	-0.25	-0.56	-0.18
OBESE	0.93	0.40	-0.05	0.82	0.71	0.03	0.93	0.78	1.00	0.95	-0.02	0.02	-0.39	0.03
HSI	0.93	0.47	0.15	0.84	0.77	0.15	0.90	0.84	0.95	1.00	0.07	-0.07	-0.46	-0.04
P_TREE	-0.01	0.12	0.39	-0.09	-0.02	-0.01	-0.02	0.25	-0.02	0.07	1.00	-1.00	-0.67	-0.71
P_NOTREE	0.01	-0.12	-0.39	0.09	0.02	0.01	0.02	-0.25	0.02	-0.07	-1.00	1.00	0.67	0.71
P_IMPSURF	-0.36	-0.30	-0.39	-0.26	-0.32	0.09	-0.38	-0.56	-0.39	-0.46	-0.67	0.67	1.00	0.64
AIRTEMP	0.11	0.00	-0.39	-0.01	-0.06	0.15	-0.02	-0.18	0.03	-0.04	-0.71	0.71	0.64	1.00



Appendix Fig. 2



Appendix Fig. 3

References

Benz, Susanne Amelie, and Jennifer Anne Burney. 2021. "Widespread Race and Class Disparities in Surface Urban Heat Extremes Across the United States." *Earth's Future* 9 (7): e2021EF002016.

Chakraborty, T, A Hsu, D Manya, and G Sheriff. 2020. "A Spatially Explicit Surface Urban Heat Island Database for the United States: Characterization, Uncertainties, and Possible Applications." *ISPRS Journal of Photogrammetry and Remote Sensing* 168: 74–88.

Hsu, Angel, Glenn Sheriff, Tirthankar Chakraborty, and Diego Manya. 2021. "Disproportionate Exposure to Urban Heat Island Intensity Across Major US Cities." *Nature Communications* 12 (1): 1–11.