

Spatial Analysis of Urban Heat Island in Twin Cities & Phoenix

December 20, 2023

Abstract

The urban heat island effect (UHI) causes serious harm to human health and beyond and is unevenly distributed across American cities. Specifically, demographic trends point to disproportionate burden among certain groups, such as low-income individuals, renters, and people of color. This report seeks to provide a rigorous spatial analysis of these trends in Minneapolis-St. Paul and Phoenix and to compare model similarity and performance between these two cities. Our analysis confirms that race, income, and house value are significantly connected to UHI distribution, but individual spatial models can help to pinpoint unique distribution patterns by city that can then be used to inform policy implementation.

1 Introduction

The Urban Heat Island effect (UHI) is a phenomenon that occurs when urban areas have higher temperatures compared to the surrounding rural landscapes. This thermal contrast results from a combination of human activities, urban infrastructure, and a lack of greenery in the form of trees, parks, and other vegetation (1). Specifically, the heavy presence of pavement, concrete, and buildings re-emit the sun's heat at rates greater than the surrounding areas. These impervious surfaces absorb more heat and allow less ground storage of rainfall, creating an “island” of high temperatures. This process can be seen below.

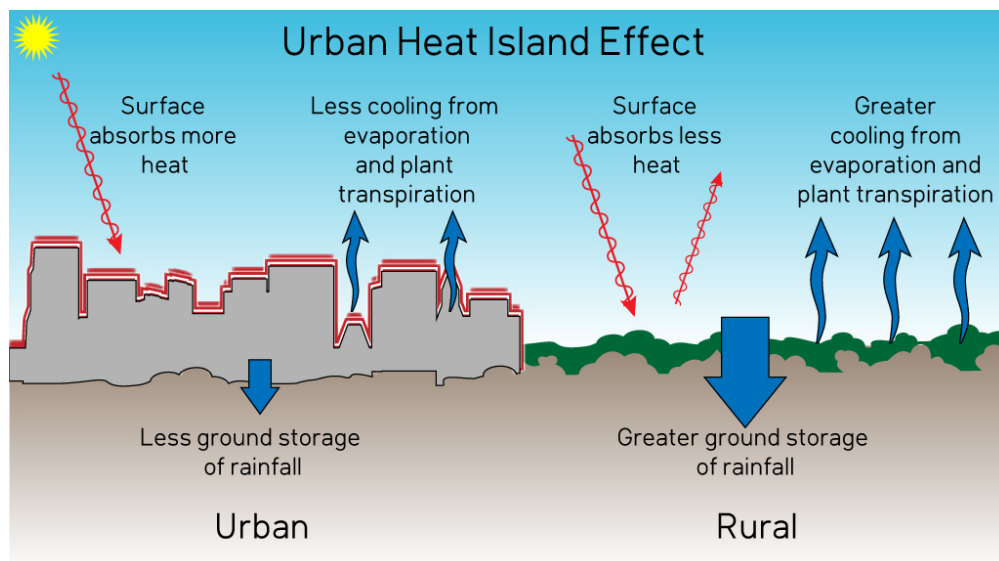


Figure 1: Diagram modeling the process underlying the urban heat island. From DENIAL101x, 2.4.2.1: Heat in the City, with Kevin Cowtan.

In the United States overall, daytime temperatures in urban areas are about 1–7°F higher than those in outlying areas and nighttime temperatures are about 2–5°F higher. In addition to diurnal variations, UHI varies seasonally, reaching its zenith in summertime during the day due to the direct overhead proximity of the sun. This has several extremely harmful effects, including increased incidences of heat-related illnesses (2), longer and more severe heat waves, higher energy costs, harmful effects on aquatic ecosystems (3), and unpleasantly hot living situations for much of the summer (4). Furthermore, the spatial distribution of UHI both between and within cities is uneven. In the United States, cities in the Southwest tend to exhibit the highest effects of UHI, while those in the Midwest tend to exhibit the lowest (5). Additionally, UHI distribution within cities, specifically when analyzed through the lens of natural vegetation and greenspace, covaries with a number of factors, such as age, socioeconomic status, race, ethnicity, and income. UHI also exhibits significantly higher values in formerly redlined neighborhoods—areas which were designated by the Home Owners’ Loan Corporation (HOLC) as “risky” investments largely on the basis of their racial and ethnic composition between 1933–1954 and face ongoing disinvestment (6). Thus, studying the spatial distribution of UHI within American cities is crucial in understanding if inequities persist and informing future policy decisions surrounding the mitigation of their harmful effects.

1.1 Research Questions

To shed light on these issues, we seek to identify which demographic factors are related to the spatial distribution of UHI within American cities and to better understand how these factors vary city by city. Spatial analysis can provide insights into the multi-faceted ways in which social and economic demographics are connected to ongoing environmental harms, such as UHI. Thus, we engage in a two-part analysis: first, we create individual spatial models of UHI distribution for two cities from different regions of the United States, then compare model similarity and performance across selected cities. Developing these models and carrying out comparative analysis is crucial for identifying vulnerable communities that may be disproportionately affected by higher temperatures, which can then allow for targeted policy interventions to mitigate potential health risks associated with extreme heat exposure. Addressing these disparities not only fosters more equitable living conditions but also promotes overall community well-being, resilience, and the creation of healthier urban environments for all residents.

1.2 Data Sources

We incorporate data related to UHI from United States Surface Urban Heat Island database (7). This dataset includes land surface temperature and the calculated UHI both seasonally and diurnally across major American urban areas. This dataset empirically calculates UHI, taking into account measures of land surface temperature gathered from satellite imaging, surface reflectance, elevation and land cover data, and tree canopy. The value of these factors combined is then subtracted from the calculated UHI of surrounding rural areas. Thus, a negative UHI indicates urban cooling, meaning that the urban area exhibits a low heat index compared to the rural average. Conversely, positive values indicate the presence of UHI.

We also utilized the 2019 American Community Survey (ACS) for demographic and socioeconomic measurement, started by the U.S. Census Bureau in 2005. This survey gathers information on social, economic, housing, and demographic facets by surveying roughly 3 million households annually using detailed questionnaires. Through its comprehensive data, the ACS offers extensive insights into the U.S. population, facilitating analyses and estimations regarding housing market trends and demographic patterns (8).

1.3 City Selection

Due to the disparity that has been identified between cities in Midwest and the Southwest, we choose to analyze one city from each region (5). Our objective is to compare model similarity and performance across selected cities and to see if we can identify a difference in important spatial-demographic factors that underlie UHI between the two cities. Due to data availability and personal interest, we choose Minneapolis-St. Paul and Phoenix as the two cities of study.

The above figure shows the spatial layout of the average UHI during summer days in Minneapolis-St. Paul and Phoenix. In the Twin Cities, high levels of UHI can be seen in the downtown areas of both Minneapolis and St. Paul, while outlying areas tend to exhibit lower effects. In Phoenix, much of downtown exhibits high UHI although there are some tracts in higher-income parts of the area that experience significantly lower effects. It is important to note that, although it is seemingly counterintuitive, the range of UHI in Minneapolis-St. Paul tends toward much more positive values than Phoenix. This is due to the difference in vegetation levels of the surrounding rural areas in

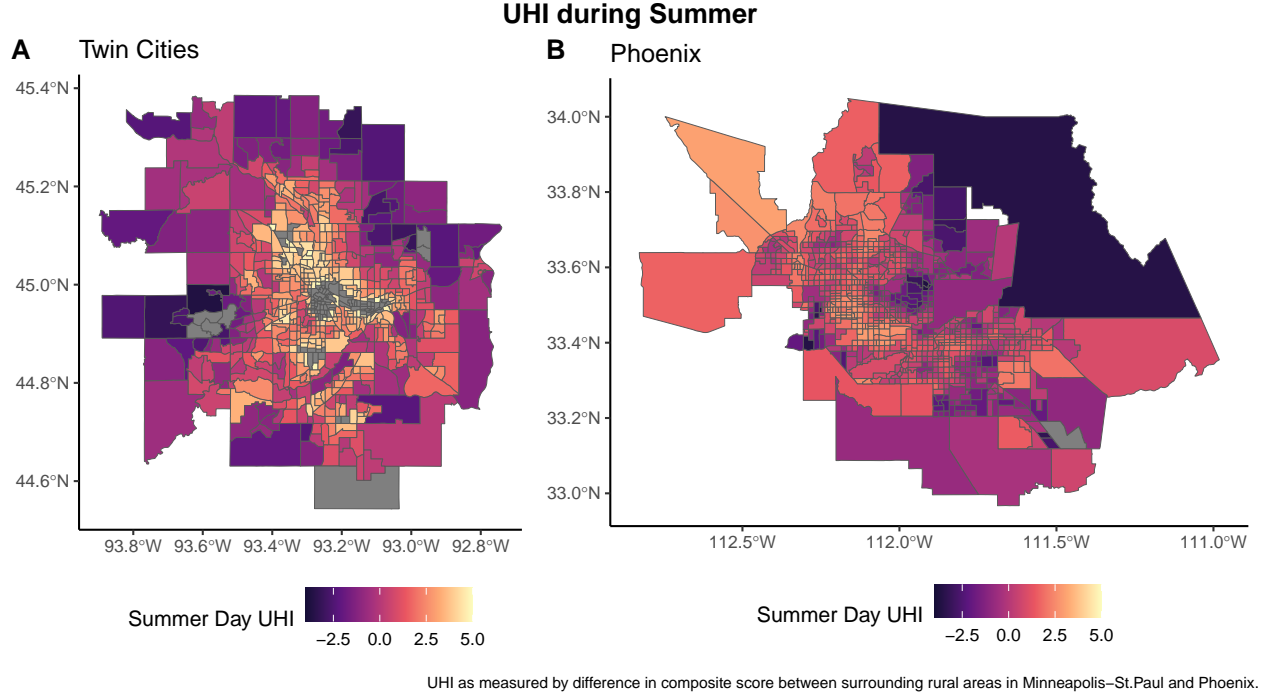


Figure 2: Urban heat island measured during summer 2019 in Minneapolis–St.Paul and Phoenix

the Twin Cities compared to Phoenix—the former is largely composed of coniferous and deciduous forest, the latter is primarily desert, and thus has a much higher baseline UHI.

2 Methods

We begin by extracting relevant census data from the 2019 ACS; guided by the literature, we consider age, birthplace, race, ethnicity, income, home value, and predominant industry by census tract in the model building process. For each of the cities studied, we use a random forest model to aid in the variable selection process (9), with summer day UHI the outcome variable. Additionally, we create maps and other visualizations to understand the spatial relationship of each predictor and to ensure that we avoid repeated information in the model. For example, the proportion of White residents and proportion of Black residents are highly correlated, as they both provide information about the racial makeup of the census tracts. Similarly, variables like home value and income are highly correlated; confirming potential instances of multicollinearity using maps is crucial in ensuring that there is no repeated information in the model.

2.1 Random Forest

In the Twin Cities, the random forest shows that proportion of Black residents, proportion of White residents, and proportion of residents born in-state of census tracts are the most highly predictive variables; in Phoenix, it shows that proportion of Black residents, proportion of White residents, and average income of census tracts are the most highly predictive. This confirms the findings of previous literature: race is a particularly strong predictor of UHI, particularly in predominantly

Black neighborhoods, which tend to be positively associated with UHI. Additionally, we find that measures of the wealth of a census tract are important predictors in UHI. In Minneapolis-St. Paul, income, average home value, proportion of residents owning their own home, and proportion of residents who work in manufacturing and industrial sectors were all moderately strong predictors of UHI. In Phoenix, many of the predictors held similar predictive power including average home value, proportion of residents born in-state, age, and proportion of residents owning their own home. These factors were taken into account in the model creation, balancing predictive power and the need to eliminate overlapping information between predictors.

2.2 Mean Model

Following the determination of the most relevant variables in each city, we fit OLS models using combinations of these predictors and analyze goodness of fit, individual predictor significance, and overall sensibility. Although the spatial data is highly correlated, OLS still provides us with an initial understanding of coefficient value for each variable, thanks to the unbiasedness of these estimates. After building the mean model, we use Global Moran's I to test if the residuals are independent or not to determine if an additional spatial model is necessary.

2.3 Neighborhood Structure

In the context of this study, we specifically choose four types of neighborhood structures, including K Nearest Neighbors (KNN), and Distance Nearest Neighbors, to capture the potential spatial correlation between census tracts.

The Rook and Queen structures define neighborhood relationships differently in spatial analysis. The Rook method defines neighbors as polygons that share at least one edge, while the Queen method takes a more inclusive approach, considering any touching points between irregular polygons (10). The Queen method offers a broader view of spatial relationships across various shapes and sizes, making it simpler to implement compared to more complex methods like K Nearest Neighbors (KNN) or Distance Nearest Neighbors. However, this inclusivity can also be a drawback, as it might include distant polygons as neighbors without considering their actual distance or characteristics.

In contrast, the Distance Nearest Neighbors approach sets specific minimum and maximum centroid distances to identify neighborhoods. This method can provide a more refined delineation of neighborhoods, especially in areas with distinct boundary features like highways or bodies of water. To illustrate the difference, we applied a maximum centroid distance of 3 km. The Queen structure displays a more uniform distribution across neighborhoods, while the distance-based structure highlights neighboring correlations mainly in the central area, leaving the surrounding regions more isolated.

2.4 SAR vs CAR

To account for spatial autocorrelation, we can use either Simultaneous Autoregressive (SAR) and Conditional Autoregressive (CAR) models. Both models are used following a Global Moran's I value indicating that the residuals of the OLS model are spatially correlated (11). SAR models account for this autocorrelation by incorporating the weighted average of neighboring observations for each spatial unit. Thus, SAR would effectively model the UHI as a function of the UHI of

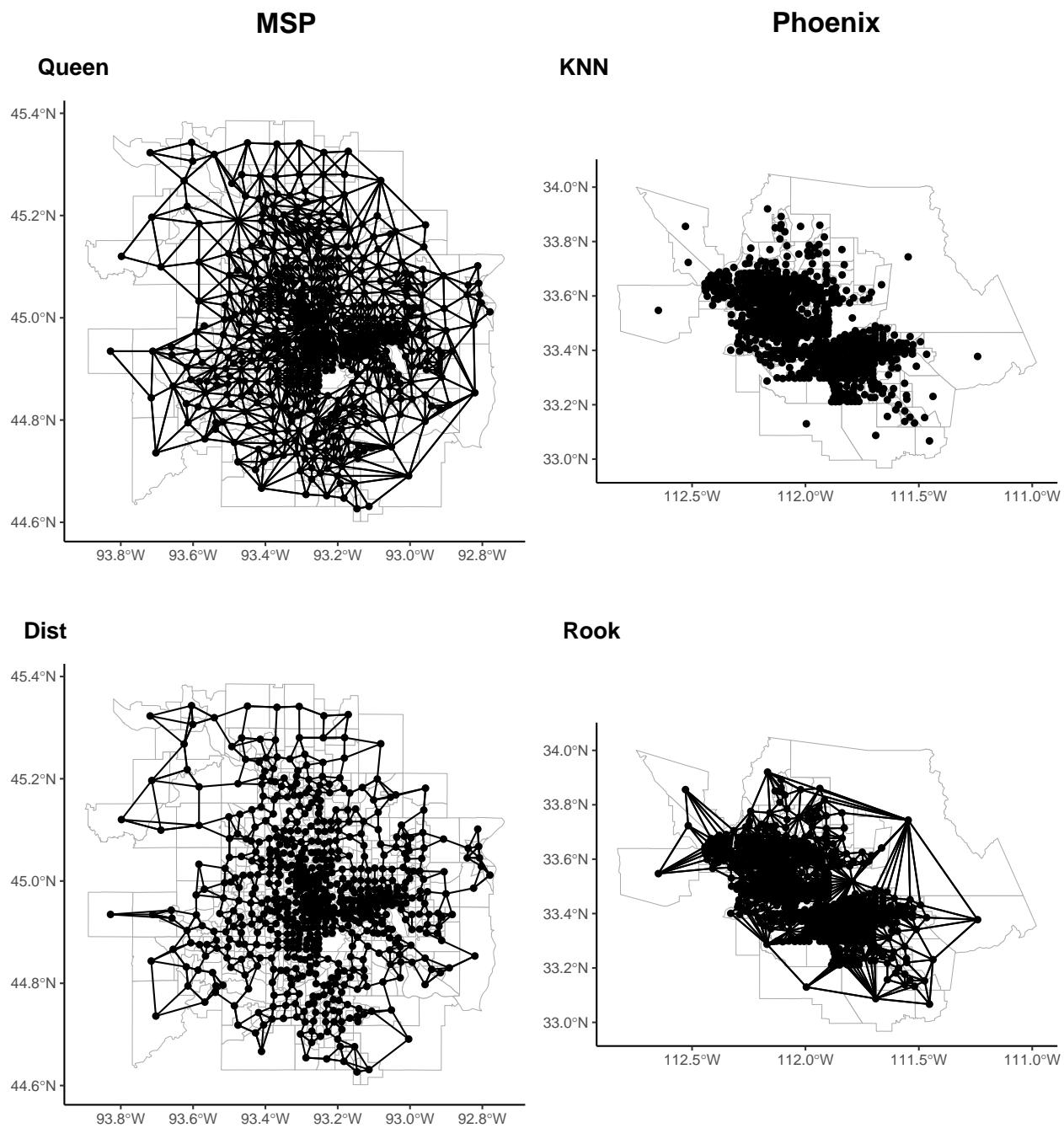


Figure 3: Four types of neighborhood structures used to access the spatial correlation in Minneapolis-St. Paul and Phoenix

neighboring census tracts (12). CAR, on the other hand, models autocorrelation in a more local setting, taking into account the average value of predictors in neighboring units. A CAR model in this context would effectively model the UHI as a function of the average of relevant predictors, such as income, of neighboring tracts.

3 Modeling Results

With a thorough understanding of SAR and CAR models, we implement both types of models on all four neighborhood structures and choose the final model based on the lowest BIC index, serving as an estimate of a model’s predictive power (13).

Starting with the final spatial model for the Twin Cities, we conclude that the SAR model using the Queen neighborhood structure performs the best, reflected by the smallest BIC index among all. Analyzing the model’s coefficients reveals key insights into the UHI trends within the Twin Cities.

In regions with a White resident percentage below 75%, a positive correlation is observed, resulting in a 0.196 increase in the UHI index, indicating a connection between residents of color and a higher UHI index, while holding other variables constant. Additionally, homes valued below 250k exhibit a positive correlation with a higher UHI index, with a coefficient of 0.088, compared to those valued between 250k and 500k. Conversely, houses with values exceeding 500k demonstrate a negative association with a higher UHI index, marked by a coefficient of -0.377. Moreover, the percentage of owner-occupied houses significantly contributes to the UHI index, displaying a negative association with a coefficient of -0.472, incorporating partial information about the building structure in the area. Lastly, a slight negative association is observed between the age of residents and a higher UHI index.

In the context of the Twin Cities, the lambda value of 0.95856, accompanied by a standard error of 0.012, signifies a strong spatial correlation in the model. Employing the Moran’s I test on the remaining residuals yields a p-value of 0.002735 (below the 0.05 threshold), indicating some lingering minor residual correlation not captured by our model. However, the Moran I statistic of 0.0635, proximate to 0, suggests relative spatial randomness in the residuals. The residual map for the Twin Cities illustrates that rural areas tend to exhibit lower UHI indices, while downtown areas register higher UHI indices.

Table 1: SAR Model Result for MSP using Queen Neighborhood

	Coefficient	SE	P-value
Intercept	3.417	0.782	0.000
% Race White (Below 0.75)	0.196	0.093	0.035
House Value (Below 250k)	0.088	0.083	0.288
House Value (Above 500k)	-0.377	0.173	0.030
% Owner Occupied	-0.472	0.226	0.037
Age	-0.018	0.007	0.013

In the context of Phoenix, our analysis, based on the Bayesian Information Criterion (BIC) index, indicates that the SAR model with a Rook neighborhood structure outperforms other models.

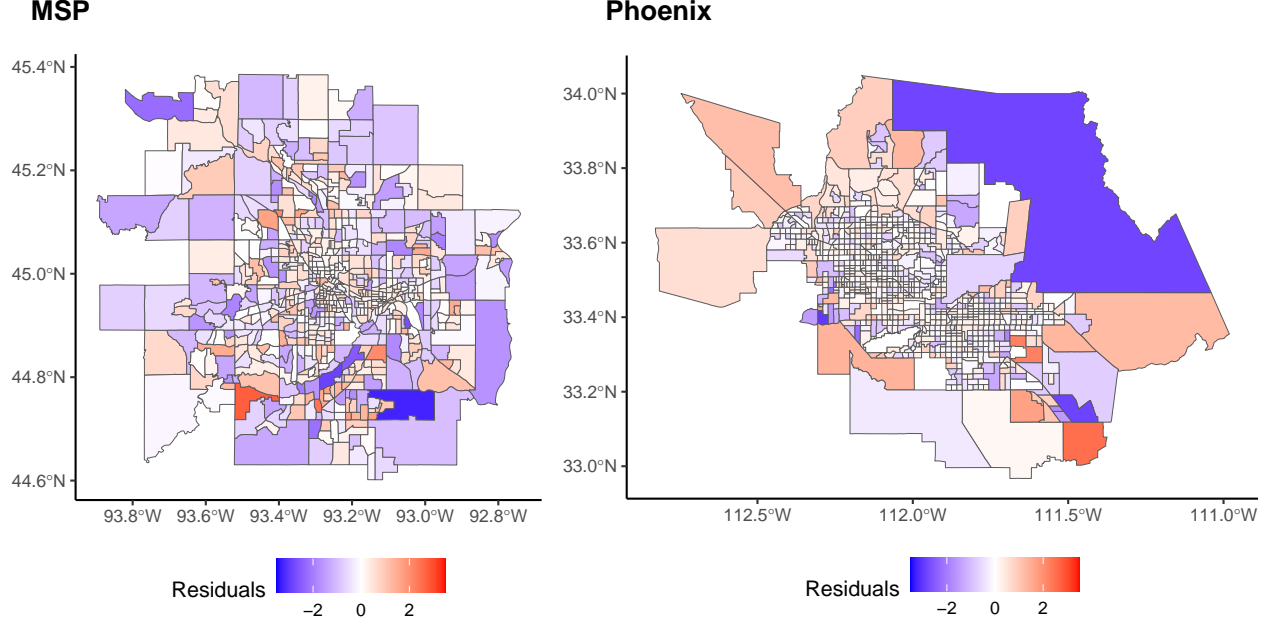


Figure 4: Residuals after modeling

Notably, the statistically significant variables differ from those in the Twin Cities, contributing to a more succinct model for Phoenix.

Specifically, in comparison to houses valued below 250k, houses in the 250k to 500k range exhibit a negative impact on a higher Urban Heat Island (UHI) index, with a coefficient of -0.178. This trend intensifies further for houses valued above 500k, marked by a coefficient of -0.568. Additionally, the average age of residents in Phoenix appears to result in a slight decrease in the UHI index.

For Phoenix, the lambda value of 0.93174, with a standard error of 0.0128, implies a strong spatial correlation within the model. The p-value of 0.1821 (greater than 0.05) for the Moran's I test on residuals indicates that our model effectively captures the spatial correlation of the UHI index in Phoenix. The residual maps for Phoenix also suggest a higher UHI index in downtown areas compared to surrounding rural areas, echoing the conclusions drawn for the Twin Cities.

Table 2: SAR Model Result for Phoenix using Rook Neighborhood

	Coefficient	SE	P-value
Intercept	0.794	0.272	0.003
House Value (Below 250k)	0.178	0.053	0.001
House Value (Above 500k)	-0.390	0.096	0.000
Age	-0.007	0.002	0.002

4 Discussion

The model for Minneapolis-St. Paul indicates that there is an association between UHI and the racial composition, home value, percentage of owner-occupied homes, and age of a given census

tract. Specifically, less White, non-owner occupied, lower home value, and younger census tracts are predicted to have higher UHI. SAR modelling using distance-based neighborhood structure still leaves some spatial autocorrelation unaccounted for, indicating that potential improvement could be made to describe the UHI distribution of Minneapolis-St. Paul using demographic factors. This remaining spatial autocorrelation is a moderate limitation of the model for Minneapolis-St. Paul; although coefficient estimates are unbiased and can still be used to inform targeted interventions, the exact extent of the overall spatial relationship as it pertains to the predictors remains unclear.

In Phoenix, several mean models performed very similarly, leading to the choice of a relatively simple linear model. This model indicates that census tracts with a higher value of homes and that are older on average tend to be predicted to have a lower UHI. SAR modelling using Rook-based neighborhood structures leads to independent residuals, indicating that there is no remaining spatial autocorrelation and that the model more accurately describes UHI distribution in Phoenix. However, although this is the best performing model in terms of limiting BIC and eliminating spatial autocorrelation, its practical applicability remains somewhat dubious—although it is important to know that UHI is modulated by age and house value of a census tract in Phoenix, it may be difficult to pinpoint exact strategies for UHI mitigation in the identified neighborhoods.

Random forest models in both cities show that the racial composition of census tracts, particularly the proportion of Black residents, and income are crucial factors underlying UHI distribution. Our model for Minneapolis-St. Paul directly accounts for this difference, whereas the model for Phoenix does not explicitly include this factor. Furthermore, the model for Minneapolis-St. Paul is more complex than the model for Phoenix, which only includes two significant predictors. This suggests that the UHI distribution in Minneapolis-St. Paul is more complicated than that of Phoenix, but this could in part be due to the differing neighborhood structures used between the two models.

Despite the finding of the random forest that the proportion of Black residents and income highly predictive of UHI across both cities, they were not found to be significant in the model, and led to higher BIC. This likely means that, although they are significant predictors of UHI, they may overlap with other predictors that are included in the model, such as home value in Phoenix. This finding could also highlight the extent to which the legacy of racial segregation continues to be visible in American cities, as higher BIC values for these models could be indicative of a fundamentally inequitable layout of cities that stems from discriminatory practices such as redlining. Further studies should analyze spatial autocorrelation models as it pertains to the racial makeup of census tracts and determine if this finding holds true across other cities.

Our analysis provides insight into where future policy interventions could be directed. Foremost, UHI mapping pinpoints areas in which the effect is the highest and suggests the need for more investment into urban greenspace and sustainable and eco-friendly buildings. Specifically for the Twin Cities, we believe that more efforts should be made to improve UHI in underserved and predominantly Black communities, especially in the central urban areas. For Phoenix, the significant contribution of house value to the model reflected the unbalanced distribution of income & race, leading to dramatic differences in UHI compared to rural counterparts. Thus, policy interventions should focus more on socioeconomically disadvantaged communities with significantly lower average annual income.

Appendix: Model Input Visualizations

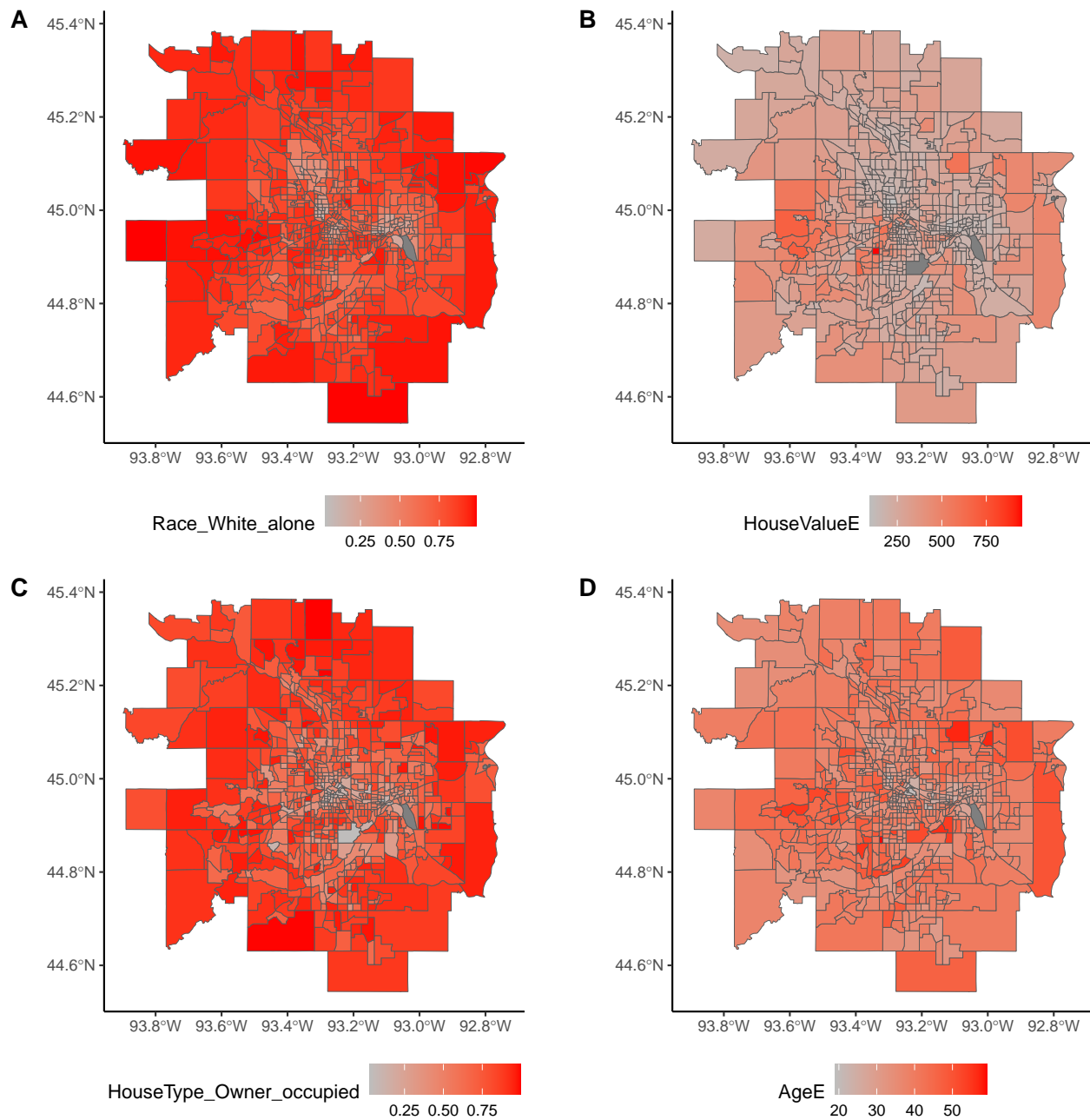


Figure 5: Model Inputs for Minneapolis-St.Paul

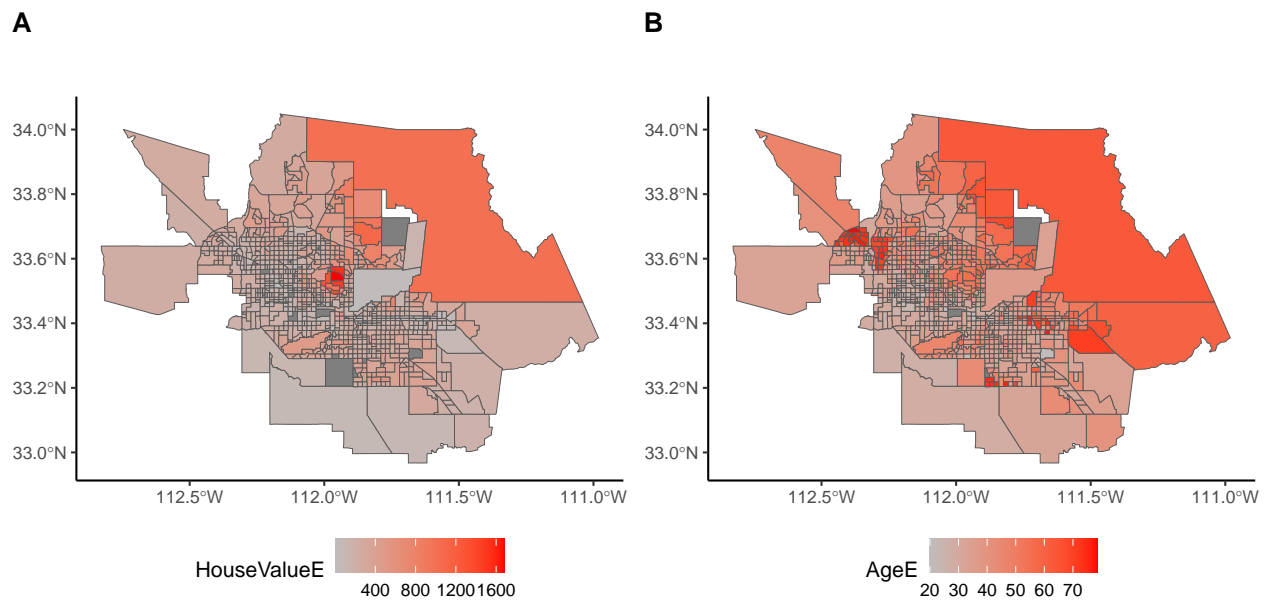


Figure 6: Model Inputs for Phoenix

References

1. Wai CY, Muttill N, Tariq MA, Paresi P, Nnachi RC, Ng AW. Investigating the relationship between human activity and the urban heat island effect in melbourne and four other international cities impacted by COVID-19. *Sustainability*. 2021;14(1):378.
2. Kovats RS, Hajat S. Heat stress and public health: A critical review. *Annual Review of Public Health* [Internet]. 2008;29(1):41–55. Available from: <https://doi.org/10.1146/annurev.publhealth.29.020907.090843>
3. Broadbent AM, Krayenhoff ES, Georgescu M. The motley drivers of heat and cold exposure in 21st century US cities. *Proceedings of the National Academy of Sciences*. 2020;117(35):21108–17.
4. Fuladlu K, Riza M, Ilkan M. THE EFFECT OF RAPID URBANIZATION ON THE PHYSICAL MODIFICATION OF URBAN AREA. 2018 May;
5. Hoffman JS, Shandas V, Pendleton N. The effects of historical housing policies on resident exposure to intra-urban heat: A study of 108 US urban areas. *Climate*. 2020;8(1):12.
6. Casey J, James P, Cushing L, Jesdale B, Morello-Frosch R. Race, ethnicity, income concentration and 10-year change in urban greenness in the united states. *International Journal of Environmental Research and Public Health*. 2017;14(12):1546.
7. Chakraborty T, Hsu A, Manya D, Sheriff G. A spatially explicit satellite-derived surface urban heat island database for urbanized areas in the united states: Characterization, uncertainties, and possible applications. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020;168:74–88.
8. U.S. Census Bureau. QuickFacts - ramsey county, minnesota [Internet]. U.S. Census Bureau; [cited 2023 Nov 20]. Available from: <https://www.census.gov/quickfacts/fact/table/ramseycountyminnesota/PST045222>
9. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognition Letters*. 2010;31(14):2225–36.
10. Earnest A, Morgan G, Mengersen K, et al. Evaluating the effect of neighbourhood weight matrices on smoothing properties of conditional autoregressive (CAR) models. *International Journal of Health Geographics*. 2007;6:54.
11. Li H, Calder CA, Cressie N. Beyond moran’s i: Testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis*. 2007;39(4):357–75.
12. Wall MM. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*. 2004;121(2):311–24.
13. Bollen KA, Harden JJ, Ray S, Zavisca J. BIC and alternative bayesian information criteria in the selection of structural equation models. *Struct Equ Modeling*. 2014;21(1):1–19.