ANALYZING POTENTIAL INCREASED EXPOSURE OF DEMOGRAPHIC AND SOCIOECONOMIC POPULATIONS TO THE URBAN HEAT ISLAND IN MILWAUKEE, WISCONSIN

Nathan Walker
Geography 401: Capstone Seminar
May 20, 2021

Rising global climate temperatures capture the attention of scientists, politicians, and the general public, threatening the survival of the polar ice caps and biome diversity. While fears for these changing temperatures remain valid, other implications in urban areas can cause disastrous consequences, particularly for those who may not have the resources to stay cool. Traditionally, built urban environments demonstrate higher temperatures than surrounding rural areas, due to the higher structural density and lack of green vegetation to absorb heat energy (Mitchell and Chakraborty 2021). These areas of warmer temperatures centered around built urban environments are commonly referred to as urban heat islands. While urban heat islands may not initially seem significant when addressing public health policy, abnormal heat has potential to result in human death, alongside decreased learning, labor productivity, and dehydration. There are even more heat related deaths in the US than other weather causes (Hsu et. al 2021). With a climbing percentage of the US population living in urban areas, it becomes important to understand and recognize the potential impacts (Mitchell and Chakraborty 2021).

Examining urban heat islands from a social justice perspective reveals additional complications to the issue. Rarely are large cities uniform in their spatial population demographics, due to a host of historical and socioeconomic factors. Several case studies have found that marginalized groups in US urban spaces experience higher heat intensity, due to the built infrastructure and environmental conditions of the neighborhood (Benz and Burney 2021). This research will seek to apply the same techniques and methods as previously performed studies, but apply them to Milwaukee, Wisconsin. Milwaukee has been identified as one of the most racially segregated cities in the US by numerous research articles (Athey et al 2021). It stands to reason that it is likely the community will follow the trend established by the literature as a result. Using a regression model and bivariate correlation tests, this study will use Census and LANDSAT data products for determining if race or other socioeconomic factors have significant impact on individual exposure to significantly higher temperatures caused by urban heat islands.

Methods:

Milwaukee County was chosen as the extent of this analysis, primarily chosen due to previous research on the prevalence of segregation in the city. Defining the study area as the county saved valuable research time, simplifying the data preparation process. This extent could also potentially be used to offer policy results at the county level if significant results are found. Figure 1 displays the research extent in a greater spatial context.

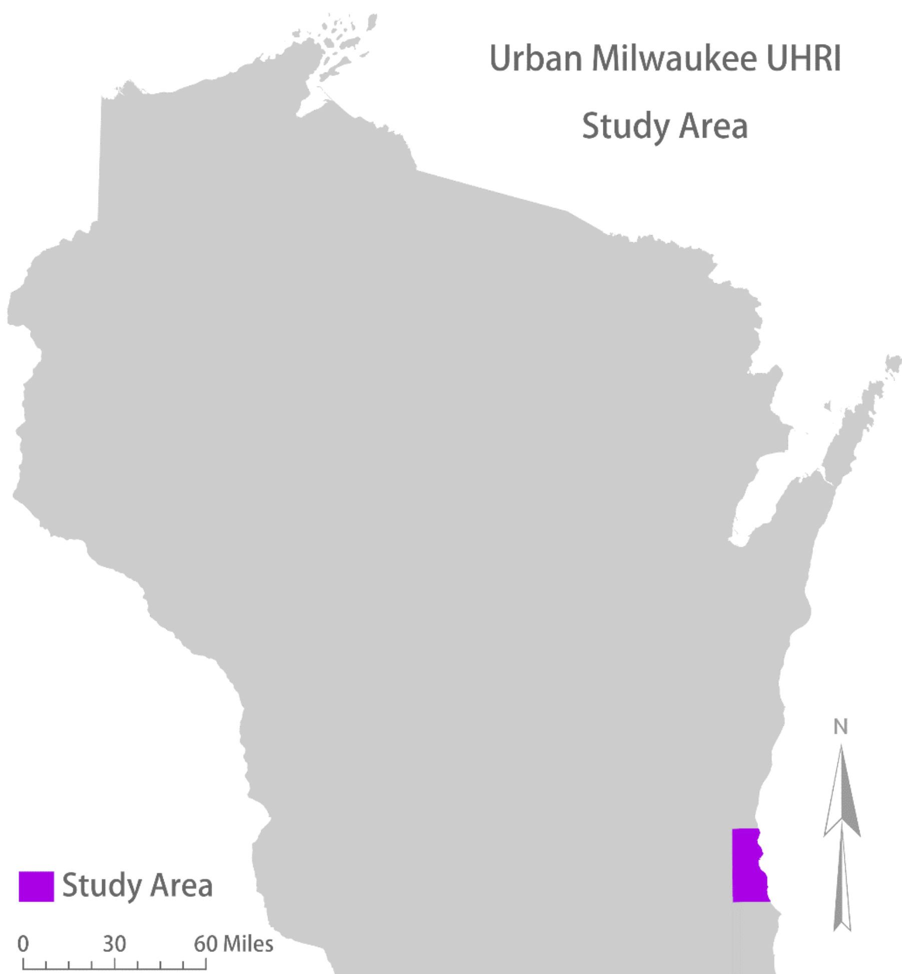

Figure 1: A map displaying the study area in the greater overall context of Wisconsin.

In order to accurately assess the potential for differences in experiences, this study will perform a modified version of the methodology created by Mitchell and Chakraborty (2015). To conduct the analysis, US Census data was primarily obtained from the NHGIS database, while remotely sensed products were pulled from USGS and NASA. Table 1 describes these datasets in greater detail.

Data	Source	Date	Resolution/Level
LANDSAT 8	USGS	August 2020	30 Meter
MODIS Land Surface Temperature	NASA	August 2020	1 Kilometer
Race Populations	NHGIS – Decennial Census	2020	Census Tract
Hispanic Populations	NHGIS - Decennial Census	2020	Census Tract
Age	NHGIS – Decennial Census	2020	Census Tract
Median Income	NHGIS – ACS 5-Year-Estimates	2019	Census Tract
Education Attained	NHGIS – ACS 5-Year-Estimates	2019	Census Tract
Census Tract Shapefile	NHGIS	2020	--

Table 1: A descriptive table of datasets used to perform the analysis.

With these datasets collected, the next step will be to derive the dependent variable, an indicator of the estimated risk of potential exposure to excessive urban heat island temperatures. To accomplish this, this study will calculate an Urban Heat Risk Index (UHRI). UHRI is based on the land surface temperature, while also accounting for the environmental factors that contribute and mitigate urban heat islands. This is shown in the following equation:

$$UHRI = (LST + NDBI) - NDVI$$

LST refers to the land surface temperature of the location and is the most basic initial estimator when testing for the prevalence of an urban heat island. However, due to changes in climate and seasonal temperatures, land surface temperature can vary greatly. To stabilize this and create a more constant index, the additional factors of NDBI and NDVI are accounted for. NDBI, or Normalized Difference Built-up Index, approximates the amount of infrastructure present in a given location. This includes concrete or other factors that may contribute to an urban heat island. The counterpart, NDVI, or Normalized Difference Vegetation Index, is an estimate for the amount of vegetation present. This includes trees, grasses, or other plants that would absorb energy and create a cooler environment. For this study, MODIS data provides the land surface temperature, while LANDSAT 8 data from a clear day that minimizes cloud cover will be used to derive the NDVI and NDBI for Milwaukee County. The results before calculation along with the equation for UHRI are shown in Figure 2.

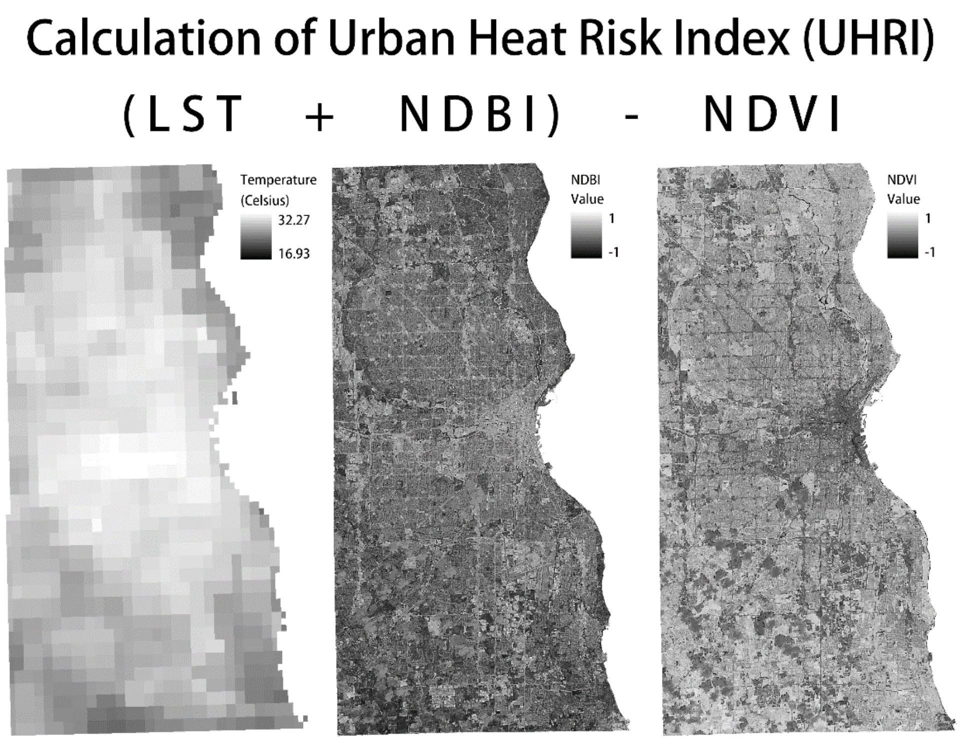

Figure 2: Maps displaying the contributing factors to the calculation of the UHRI.

The raster datasets shown in Figure 2 can then be converted to vector points, each attributed with the value of the respective pixel. These values are then aggregated to the Census tracts, with the Census tract being assigned the mean value of all the points contained within. Since the MODIS dataset was at such a low resolution, it needs to be resampled at a higher resolution to ensure that every Census tract contains at least one point with the land surface temperature. Once the Census tracts are attributed with their respective remotely-sensed products, the values will then be converted into Z-scores, to ensure that no one variable has a greater influence on the overall UHRI. These Z-scores will then be subject to the UHRI calculation. Figures 3 and 4 display the distribution of the index and a map of the resulting UHRI values respectively.

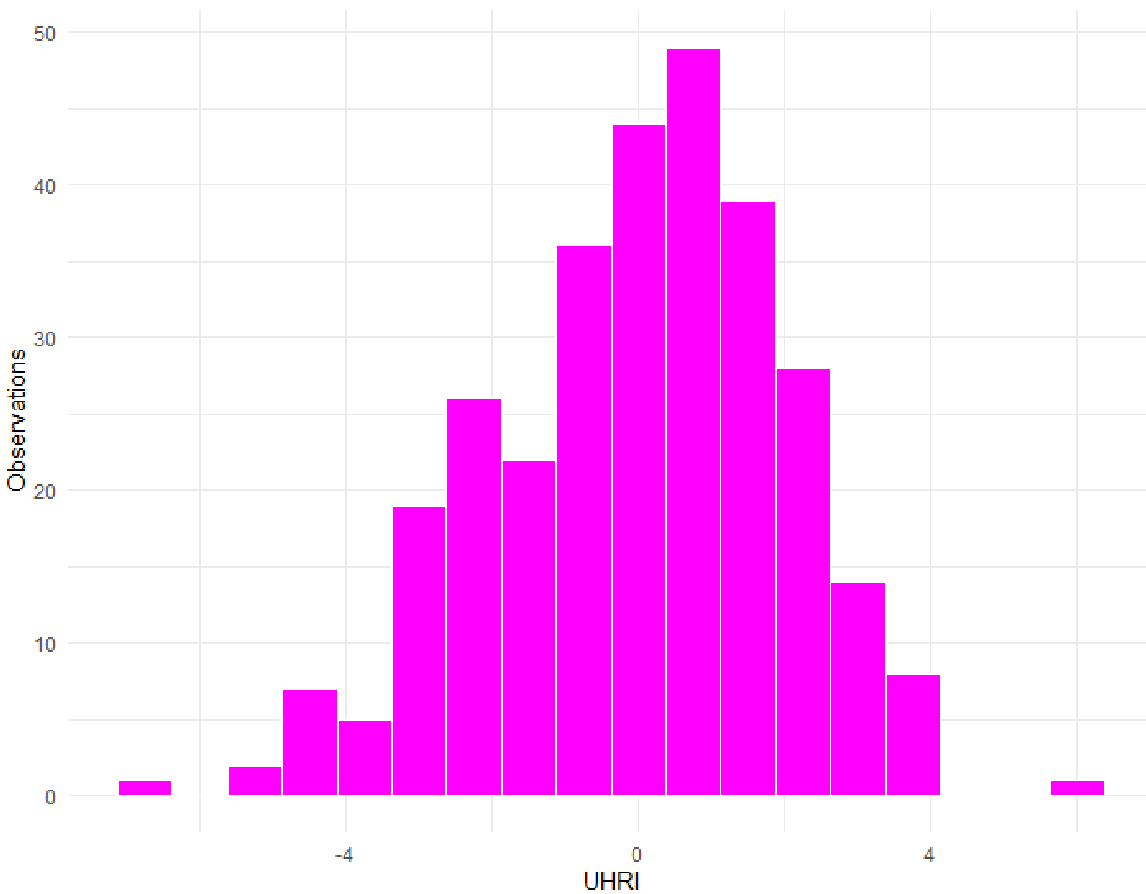

Figure 3: Histogram displaying the distribution of the Census tract UHRI scores.

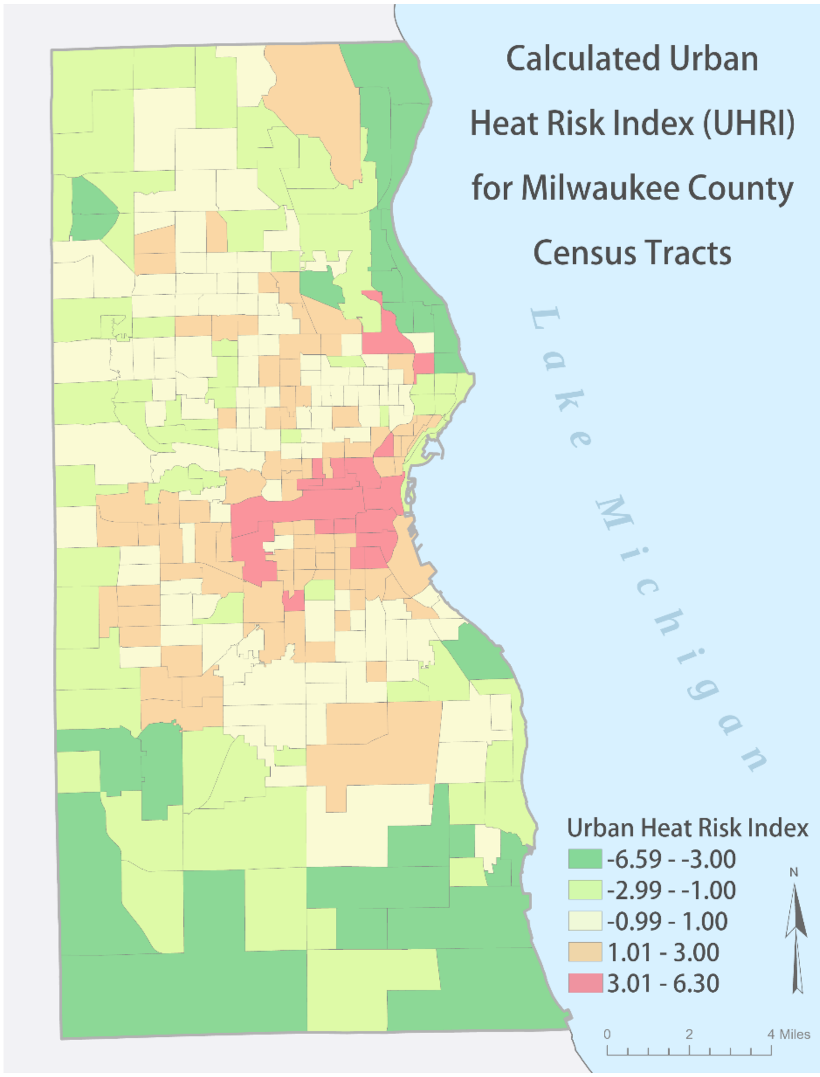

Figure 4: Map of the calculated UHRI for each Census tracts.

With the UHRI serving as the dependent variable, the regression model will then be created. First, the model will incorporate racial and ethnic factors. These will be added by calculating the percentage of the population that is Black, Asian, or Hispanic. Literature suggests that these groups may be experiencing significantly higher temperatures, and by including them as explanatory variables, this occurrence can be modeled and tested for. Another factor of interest is age. Another demographic population of interest are age extremities. For this study, that will be defined as percentage of the Census trac population that is children under the age of 5 and adults over the age of 65. For economic factors, two dependent variables will be included, median household income, and the percentage of the population that has attained at least a high school diploma or equivalent. Finally, population density will be included as an additional variable. Using these factors, an ordinary least squares (OLS) regression model will be created, and used to test for multicollinearity, heteroskedasticity, the overall goodness-of-fit of the regression and spatial autocorrelation.

Table 2 displays the complete list of dependent variables and their matching statistics.

Variables	Coefficient	Std. Error	t-Value	Probability	VIF
Intercept	1.297	0.892	1.455	0.14682	--
Median Income	-3.103e-05	5.486e-06	-5.656	3.79e-08	2.2379
% Asian	8.709e-03	2.081e-02	0.418	0.67598	1.1875
% Black	-1.737e-03	5.107e-03	-0.340	0.73409	3.4356
% Hispanic	2.525e-02	7.638e-03	3.306	0.00107	2.7569
% Under 5	-9.031e-02	3.649e-02	-2.475	0.01391	1.3957
% Over 65	-8.058e-02	1.968e-02	-4.095	5.52e-05	1.8019
% High School Graduate	1.982e-02	1.013e-02	1.957	0.05137	2.5931
Population Density	5.777e-05	2.245e-05	2.574	0.01058	1.5563

Table 2: Output results of the OLS regression analysis. Significant results are shown in blue.

With an R-squared of .4038, the model only explains on average 40% of the variance in the UHRI variable. This indicates a poor to moderate model. The variance inflation factors for the OLS regression indicates that there is only low to moderate multicollinearity occurring, with the likely culprit being the percent Black variable associated with the economic factors. The model exhibits significant evidence of multicollinearity, and as a result any interpretation of the model should incorporate a degree of caution. When a global Moran’s I test is performed on the residuals, there is significant evidence of spatial autocorrelation. As a result, the model will have to be appropriately modified.

To account for the evidence of spatial autocorrelation, a Spatially Lagged X (SLX) regression model will be performed. Using a weight matrix, based on queen's contiguity, the variables of the surrounding tracts will be considered as additional variables. In interpreting the model, the coefficients are representative of the total impacts of the both the independent variables of the tracts and the surrounding variables. This method is successful in reducing the spatial autocorrelation of the residuals, and slightly increasing the adjusted R-squared value to .4524.

In addition to the SLX model, another, simpler tool will be used to test if certain populations are at increased risk for urban heat islands. Bivariate correlation tests do not indicate causation but will illuminate if Census tract demographic values exhibit any correlation with the UHRI. These can also be used to evaluate the results of the model, and if a variable should be removed, as well as the potential causes of multicollinearity.

Results:

The UHRI appears to be an accurate indicator of the potential for increased temperatures due to urban heat islands. The heavily built-up downtown area, likely to be the warmest, was attributed with the highest risk index. The tracts near the lake also have lower indexes, accurate as to what would be expected. Based on these observations, it is likely safe to assume that the UHRI calculated in this study is a safe proxy for risk to urban heat island exposure.

With this assumption established, the results of the correlation tests will first be presented. Table 3 displays the results of the bivariate correlation tests.

Variables	Correlation (Pearson's r)	Correlation (Pearson's r p-value)
Median Income	-0.48	< 2.2e-16
% Asian	-0.01	0.6715
% Black	0.09	0.0546
% Hispanic	0.39	5.074e-12
% Under 5	0.01	0.8718
% Over 65	-0.44	5.09e-15
% High School Graduate	-.35	4.744e-09
Population Density	0.44	2.2e-16

Table 3: A table of the bivariate correlation test results. Significant findings are shown in blue.

Several variables are not initially significant when looking solely at the correlation results. Percent Asian, percent Black, and percent under 5 all have low correlation, and insignificant p-values at the 95% confidence level. This seems to suggest that these variables are not correlated with increased exposure to urban heat islands. The percent black variable is especially unexpected, considering this finding is not consistent with the literature on the subject.

Median income is moderately inversely correlated with the risk index. This indicates that as the median income of a tract decreases, the expected risk of exposure to urban heat tends to increase. That is, there is significant evidence to suggest that lower income populations are associated with higher potential risk of urban heat in Milwaukee.

The percentage of the population that is Hispanic is positively correlated with the risk index. Census tracts with higher populations of people identifying as Hispanic also tend to exhibit higher risk indexes. The correlation tests provide significant evidence that Hispanics tend to experience higher risk to the Milwaukee urban heat island.

While Hispanics and lower income people seem to be at increased risk of exposure, the correlation tests provide evidence that elderly populations over 65 that are inherently at increased risk are not in locations with higher UHRI values, at least for Milwaukee County. This is consistent with the results derived by Mitchell and Chakraborty in New York City, Los Angeles, and but contradicts with other literature findings (2015).

Next, the percentage of the population of at least high school graduates is negatively correlated with the urban heat risk index. As this percentage of the population decreases, the expected risk to the urban heat island increases. In Milwaukee, the correlation tests provide evidence that those with more education are at decreased risk to urban heat islands, and those with less are at increased risk.

Finally, the population density of the Census tracts is positively associated with the calculated urban heat risk index. This is a logical finding, since as the number of people per square mile increases, so does the infrastructure necessary to support those people. This infrastructure is directly accounted for in the UHRI calculation through the NDBI and is likely responsible for the resulting correlation.

The correlation tests reveal a basic understanding of the datasets; however, they do not provide a holistic understanding of the independent variables alongside each other, nor do they account for previously addressed issue of spatial autocorrelation. The SLX regression provides another tool for evaluating the relationships found. The results of the SLX model are shown in Table 4.

The SLX model results in similar findings to the correlation tests, with a few exceptions. Percent Asian and percent Black are still not significant, and inconsistent with previous

Variables	Coefficient	St. Errors	Z-score	Probability
Intercept	2.452	1.446	1.696	0.091005
Median Income	-4.646473e-05	8.199095e-06	-5.6670558	1.4527e-08
% Asian	3.854639e-02	3.216961e-02	1.1982238	0.2308299
% Black	-7.323023e-04	8.248422e-03	0.0887809	0.9292560
% Hispanic	4.114248e-02	1.261096e-02	3.2624387	0.0011046
% Under 5	-2.333174e-01	8.351817e-02	-2.7936123	0.0052123
% Over 65	-1.492804e-01	3.815931e-02	-3.9120321	9.1523e-05
% High School Graduate	4.356954e-02	1.685506e-02	2.5849530	0.0097392
Population Density	9.925718e-06	3.737438e-05	0.2655754	0.7905662

Table 4: A table of the SLX model outputs. Significant findings are shown in blue.

literature. Given the segregation present in Milwaukee, this is especially unanticipated for the percent Black variable. Median income, percent over 65, and the percentage that have graduated high school are still significant and follow the same trends as the correlation tests.

In the model, percent under 5 is calculated to be significant and to have an impact in predicting the UHRI for the Census tract. However, the correlation tests indicated next to no relationship between the two variables. This seems to indicate that the variable is falsely being attributed by the model as having a significant impact and should likely be removed.

The population density is also not significant in the SLX model. This can also be easily explained since the SLX model is accounting for spatial autocorrelation. Population density tends to be clustered in its distribution and is a prime example of the occurrence of spatial autocorrelation. When this aspect is accounted for, population density no longer has a direct correlation with the calculated urban heat risk index.

Conclusions:

Statistically, combined the two tests provided evidence of relationships with the urban heat risk index for four of the eight variables tested. Median income was found to be negatively correlated higher risk values, providing evidence that lower income areas are more likely to be at risk to urban heat island exposure. Significant evidence was found that Census tracts with higher Hispanic populations also have a higher risk of exposure. Similarly, populations with lower percentage of people that have at least attained a high school education are at increased risk, according to the results of the tests. These three socio-economic factors all indicate examples of potential injustices present in the form of greater urban heat island temperatures and are consistent with previous literature on the topic.

One group was found to actually be at a decreased risk to urban heat islands in Milwaukee, populations over 65. As the older populations increased, the urban heat risk index decreased. Previous studies have found inconsistent results regarding the environmental impact of increased risk of elderly populations, as discussed previously.

These findings are not without limitations and cautions. One of the major challenges presented was the heavy cloud cover for the available datasets in desired time period. The LANDSAT 8 data used mitigated cloud cover; however, it was still present to a limited degree and may have affected the derivation of the NDVI and NDBI in some areas. This also forced the study extent to be limited to Milwaukee County, since other areas would have been even more inaccurate. This may have affected the study results, by using a study area that does not allow for the full variability of some of the variables. For example, the study area may not have been able to capture the full degree to which segregation occurs over the Greater Milwaukee area. This study extent does not capture the suburbs, like other studies may have accomplished. With a larger, more expansive dataset, it is possible the relationship between populations and the risk index can be better understood and possibly result in findings more consistent with the literature.

Another potential limitation with the remotely sensed products is the MODIS data used for the land surface temperature. The data set is at a low spatial resolution, with each pixel based one-kilometer intervals. This low a resolution could result in overgeneralization when working with political areas as small as Census tracts. It is possible to calculate the land surface temperature using LANDSAT 8 data, however due to time constraints this was not an option.

The SLX model used also exhibited evidence of needing reform. While none of the calculated variance inflation factors were higher than 8, the standard benchmark of multicollinearity indicating they should be disregarded, the percent black variable still had a moderately high value of 3.4. While not directly problematic, this is still higher than what would be preferred for attempting to draw conclusions. According to the Breusch-Pagan test performed on the OLS model, the significant heteroskedasticity present also should be addressed. However, after reviewing the literature, there appears to be no current method to create an SLX model that also accounts for heteroskedasticity in R. As such, this study is forced to accept the limitation and proceed with caution. In future work, it may be worthwhile to develop a model that can account for these issues.

While some transformations were tested on the variables, primarily logarithmic versions, none increased the goodness-of-fit estimates for the model. It is possible still that other transformations not tested could potentially increase the accuracy of the model and improve the calculated R-squared. These other options remain to be explored in future studies.

Finally, the model likely should also be improved through the modification of chosen explanatory variables. For example, the correlation tests indicated little to no correlation between young populations and the urban heat risk index. As such, it should be removed from the model in future tests. Due to time constraints, only the variables discussed in the study were used. However, previous studies have integrated many more potential factors involved, such as Gini coefficients, household ownership, and non-English speaking populations (Mitchell and Chakraborty 2015). Any of these variables could also be potentially identified in Milwaukee, to attempt to improve the model and create a better understanding of the environmental impacts on these groups.

Assuming none of these limitations invalidate the results, this study provided at least three examples of potential injustices present as a result of urban heat islands: disproportionate exposure to low-income, uneducated, and Hispanic populations. These all suggest that a "climate gap" is present, where populations with fewer economic resources do not have the necessary resources to combat the higher temperatures found in urban centers (Shonkoff et al. 2011). As a result, new policies are needed to address this disparity present the urban space of Milwaukee. To accomplish this, two primary approaches can be used.

The first strategy is to mitigate the effects of the Milwaukee urban heat island directly. This can be done through policies that overall improve the energy efficiency of the cities, and by optimizing the urban landscape through the use of reflective materials in construction, green spaces, and roof vegetation (Yang et. al 2016). These policies all adjust the previously identified built-up and vegetation indices that contribute to the creation of the urban heat islands. To implement these adjustments, previous studies have suggested greater education for public workers and policy makers (Malley et. al 2014). With the knowledge that these environmental factors exist and how they can be mitigated, the impacts of the urban heat island can be reduced to decrease any potential disparities.

The other approach is more of an indirect solution. By addressing the distributions of socioeconomic disparities in Milwaukee, this can try and ensure that no groups are disproportionately exposed to the effects. This would be considered a less ideal approach, since the urban heat island still exists, it just affects more people with adequate resources to control for the impacts. This approach is also rather slow in span and will take time to change the history of redlining and segregation discussed previously. To accomplish this will require increased coordination between public health professionals and urban planners (Corburn 2005). Public policies and zoning laws can also be created, modified, or remove to potentially create a less segregated Milwaukee.

A true solution to the urban heat island will likely require a strategy that occurs at multiple levels of scale and government, whether the strategy attempts to combat national contribution to greenhouse gas emissions, or the placement of a new road in a local Milwaukee neighborhood. The findings of this study highlight just one of the potential impacts of both segregated spaces in US urban environments and climate change overall. More studies and analysis are needed to identify and address the coming issues as both wealth disparities and temperatures continue to increase, in order to avoid or mitigate potentially disastrous economic and social justice outcomes.

Works Cited

Benz, S. A., and J. A. Burney. 2021. Widespread race and class disparities in surface urban heat extremes across the United States. *Earth's Future* 9 (7).

Bivand, Roger S. and Wong, David W. S. (2018) Comparing implementations of global and local indicators of spatial association TEST, 27(3), 716-748. URL https://doi.org/10.1007/s11749-018-0599-x

Bivand, R., G. Millo, and G. Piras. 2021. A Review of Software for Spatial Econometrics in R. Mathematics 9 (11):1276. https://doi.org/10.3390/math9111276.

Corburn, J. 2005. Urban Planning and Health Disparities: Implications for research and Practice. Planning Practice and Research 20 (2):111-126

Hsu, A., G. Sheriff, T. Chakraborty, and D. Manya. 2021. Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications* 12 (1).

Imdad, M. U. & Aslam, M. (2018). mctest: Multicollinearity Diagnostic Measures. URL https://CRAN.R-project.org/package=mctest, R package version 1.3.1

Manson, S., J. Shroeder, D. Van Riper., T. Kugler, and S. Ruggles. 2021. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS.

Mitchell, B. C., and J. Chakraborty. 2015. Landscapes of thermal inequity: Disproportionate exposure to urban heat in the three largest US cities. Environmental Research Letters 10 (11):115005.

Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Shonkoff, S. B., R. Morello-Frosch, M. Pastor, and J. Sadd. 2011. The Climate Gap: Environmental Health and equity implications of climate change and mitigation policies in California—a review of the literature. *Climatic Change* 109 (S1):485–503.

Wickham, H., Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Yang, L., F. Qian, D.-X. Song, and K.-J. Zheng. 2016. Research on urban heat-island effect. *Procedia Engineering* 169:11–18.

O'Malley, C., P. A. E. Piroozfarb, E. R. P. Farr, and J. Gates. 2014. An investigation into minimizing Urban Heat Island (UHI) effects: A UK perspective. Energy Procedia 62:72-80