Quantifying the spatial aggregation bias of urban heat data

Esteban López Ochoa Ph.D., Kristen E. Brown Ph.D., Ryun Jung Lee Ph.D., Wei Zhai Ph.D.

Abstract

Every year, high temperatures send people to the hospital and morgue, and the combination of climate change and urbanization will increase extreme heat exposure. Cities are searching for ways to determine the most affected areas to begin addressing this pervasive issue. While we are living through the “big data” revolution, policy makers are still uncertain about what level of data is most useful. We evaluate the data loss from using data at different spatial resolutions to evaluate heat vulnerability, as both the definition of intra-urban heat and the resolution of the data affect the area identified and targeted for mitigation. Variance-based metrics provide many advantages, but when data is aggregated, these metrics are less able to represent the full range of urban heat. Using the case of Bexar County (home to San Antonio, TX), we find that increasing data aggregation increases both false positive and false negative identification of intra-urban heat islands, leading to unreliable results. Misclassification increases as aggregation increases, indicating that decisions should be made at the finest spatial resolution possible.

Keywords

intra-urban heat islands, San Antonio, spatial statistics, environmental management, ecological fallacy, heat island

1. Introduction

There is strong, scientific evidence for the adverse health impact of heat. Hospitalizations increase during extreme heat for cardiovascular disease, heat stroke, acute respiratory distress, hyperventilation/pulmonary stress, cognitive and organ dysfunction, dehydration, and can ultimately result in premature mortality (Anderson & Bell, 2009; Ebi et al., 2021; Li et al., 2015; Nakamura & Aruga, 2013). Epidemiological research that studies the heat-health link primarily relies on temperature measured at the city or county scale (Anderson & Bell, 2009; Li et al., 2015; Nakamura & Aruga, 2013; Scalley et al., 2015). However, municipalities often face the challenge of addressing public health concerns with limited resources. Some cities have already begun assessing how to allocate heat mitigation efforts to maximize the benefit. In Chelsea, MA (a suburb of Boston) temperatures experienced by residents were found to be several degrees higher than traditional weather station measurements, and differences in local scale vegetation had a significant impact on heat (Chang et al., 2021; Melaas et al., 2016; Milando et al., 2022). Meanwhile, in Phoenix, AZ (Gober et al., 2009, 2010; Guhathakurta & Gober, 2007) the tradeoff between the resource efficiency of compact urban form and urban heat islands (UHI) require planning that accounts for increasing heat. However, for many municipalities, the lack of access to spatially detailed data on heat exposure could result in interventions that are based on biased and incomplete information.

One obstacle with using coarse spatial resolution is loss of data. If data is collected at a low resolution, it may not be representative unless sampling procedures are specifically designed to

address this issue. If data is collected at a high resolution but later averaged, then sharp gradients may obscure important features. To demonstrate this effect, Figure 1 shows the land surface temperature (LST) at progressively coarser spatial resolutions. One row of pixels (30mx30m) collected by LandSat is plotted, and aggregations of the same data are plotted on top of that original data at progressively coarser levels of averaging. The black line aligns well with the traditional urban heat island, where the temperature within Bexar County (between B and C) is higher than that in the surrounding, more rural counties. It is also apparent that there is much more variation in temperature than just rural and urban regimes. Aside from this plot, only data from within Bexar County is considered in this paper. Once 100 pixels are averaged, much of the detail in the data has been lost, and several local maxima and minima are completely obscured.

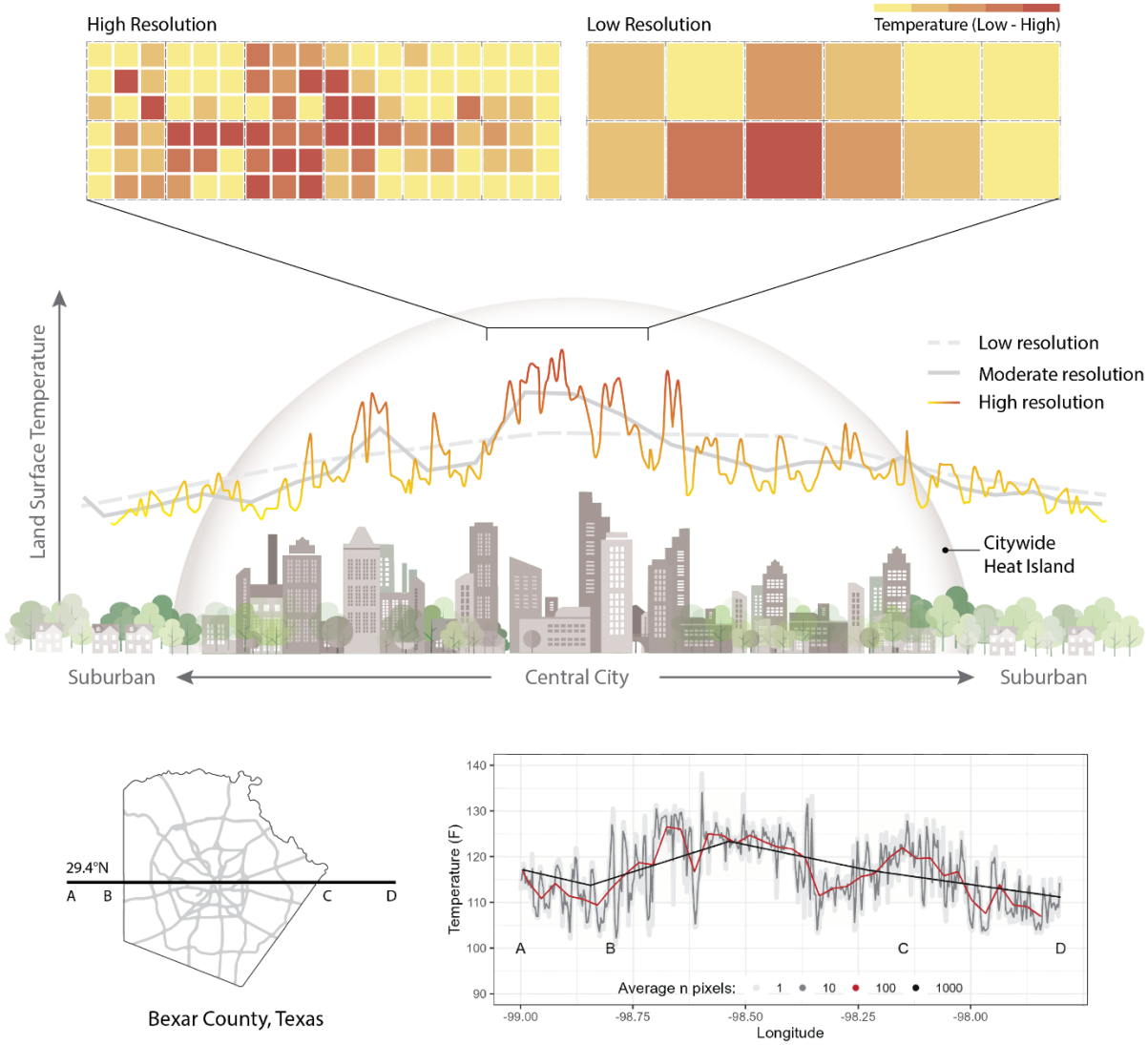

Figure 1. The effect of averaging on the accuracy of temperature. The top panels display the same representative areal LST at two resolutions. The middle panel demonstrates how discussion of IUHIs relate to traditional urban heat island definitions. The bottom

panel plots actual data of a single row of pixels. The location of the row is overlaid on a map of Bexar County and the major roads within. The letters A-D on the map and data correspond to the edges of the data included and the county boundary. The temperature of the row of pixels is plotted at the pixel level (30m), for the average of 10 pixels (300m), the average of 100 pixels (3km), and the average at 1,000 pixels (30km). By the 3km resolution there are already several miscategorized areas and the minimum and maximum temperatures are obscured.

Bexar County, located in south-central Texas, was chosen as a case study for this analysis as it is home to San Antonio, one of the largest and fastest-growing cities in the United States, and regularly experiences extreme temperatures. San Antonio is characterized by a humid subtropical climate, transitioning to a semi-arid climate towards the city's western regions. The city's geographical location makes it susceptible to these rapid weather changes, necessitating a well-prepared approach to weather-related challenges and infrastructure resilience. Also, San Antonio is identified as one of the ten worst cities in the United States in terms of the heat island effect, exposing more than 67% of its residents to a temperature rise of at least 8°F due to the urban heat island effect (Climate Central, 2023). The two million residents in San Antonio typically experience multiple days over 100°F (38°C) each year. Bexar County has a varied landscape that shifts from semi-arid vegetation on an elevated terrain to a more humid and densely vegetated prairie and grassland at lower elevation. (Climate Central, 2023)

Increased spatial resolution allows for areas of highest need to be more precisely identified for targeted prioritization of heat management and mitigation interventions. However, there are obstacles to increasing resolution. Historically, the major obstacle to fine resolution analyses has been data availability. However, as the age of ‘big data’ and increased computing capability now allow access to much more detailed information, it is imperative to reassess the decision-making paradigm. While data is now broadly available for LST across the globe thanks to satellite missions, there are still limitations.

First, reconciling LST data with other important variables (e.g., socioeconomic and built environment) may force researchers to aggregate LST data to coarser spatial resolutions. For example, researchers (e.g., (Eisenman et al., 2016) have used the widely known social vulnerability index (SVI) created by the CDC (Centers for Disease Control and Prevention et al., 2022) for identifying heat areas for targeting mitigations. However, the SVI is only available at the census tract level, forcing researchers to aggregate LST data, without having an assessment of the bias incurred when doing so. Second, although computing power could be available for most researchers and policymakers, not all of them have the data manipulation expertise to use remote sensing data and embark in data fusion projects to reconcile the data at a fine spatial resolution. For example, spatial interpolation (Schroeder & Van Riper, 2013) and locally weighted regression techniques (Brunsdon et al., 1996; Cleveland & Devlin, 1988) can be used to disaggregate tract-level median income estimates from the American Community Survey (ACS) to block groups or even blocks. In sum, increasing spatial resolution remains an elusive task that is often discarded because the size of the aggregation (or disaggregation) bias is not well known.

Although the issues with using aggregated data and the bias that it generates have been widely shown in the fields of spatial and geographical analysis (see Holt et al. (1996) for an initial

review), there are still a number of recent articles analyzing urban heat phenomena that continue using aggregated LST measures to inform both the academic and policy communities (Babak Jalalzadeh et al., 2021; Huang et al., 2011; Jagarnath et al., 2020; Nayak et al., 2018; Reid et al., 2009). Disregarding the aggregation bias when using LST averages is dangerous for three main reasons. First, using aggregated data misleads data users to wrongly derive conclusions about the individual-level observations composing those aggregations, a problem known as the ecological fallacy (Piantadosi et al., 1988). Second, results and interpretations based on aggregated data are sensitive to the aggregation boundaries, a problem known as the Modifiable Area Unit Problem (MAUP) (Fotheringham & Wong, 1991). Finally, deriving indices from aggregated data disregards the spatial nature of the underlying distribution resulting in misleading results. For example, as Niu et al. (2021) points out, the well-known Heat Vulnerability Index (HVI) is calculated in most cases using the Principal Component Analysis (PCA) method on aggregated data. However, Cartone & Postiglione (2021) have recently shown how PCA is vulnerable spatial dependence issues on the underlying distribution and hence needs to be corrected by adding spatial weights to the method. This paper provides an assessment of the aggregation bias on LST measurements. Several aggregation levels are considered within multiple definitions of intra-urban heat islands (IUHI), which are areas of increased temperatures within an urban heat island (Hoffman et al., 2020).

2. Methods

2.1 Land Surface Temperature (LST)

This study used land surface temperature data from the Landsat project. Specifically, we used the Landsat Collection 2 analysis ready data (ARD) tiles freely available from the United States Geological Survey (USGS), which are available at 30m x 30m resolution (Dwyer et al., 2018; USGS, 2023) from https://earthexplorer.usgs.gov/. LST is often referred to as the “skin” temperature and is different from the air temperature (measured near the surface) that is routinely measured at meteorological stations. LST is generally correlated to air temperature, but typically higher. The Landsat satellite images for the United States are analysis-ready with minimal measurement errors, especially for cloud-free days (Cook et al., 2014; Malakar et al., 2018). Thus, this study selected satellite images on clear days as cloud cover could cause errors in LST readings (Cook et al., 2014). We selected days in which the cloud cover over Bexar County (not the entire tile captured) was less than 0.2%. After dismissing those satellite images obscured by clouds, the available data points during the summer of 2022 were: May 20, May 28, June 29, and July 7. The satellite collects the information along consistent routes at 12:03PM on the respective dates.

2.2 Categorization

There is disagreement in the literature about how to define IUHIs (Buyantuyev & Wu, 2009; Gu & You, 2022; Martin et al., 2014; Stewart, 2011; Wong et al., 2016). The literature has traditionally used non-inferential metrics (i.e. metrics that do not rely in analytical or permutation-based approaches to derive the full distribution of the statistic for deriving statistical inference under a null hypothesis) to define heat islands by calculating central tendency statistics and defining a cut-off threshold. In order to provide an assessment of these metrics in terms of their

sensitivity and ability to capture the full distribution of LST values across different aggregation levels, we used three non-inferential metrics to define areas that can be categorized as IUHIs within Bexar County. The first metric used a uniform threshold and defined all areas with an LST above 120˚F as IUHI. This method is seemingly straightforward, but ignores documented acclimatization effects in which the same temperature induces a different health response in different locations and different times of year (Scalley et al., 2015). In general, the use of an unchanging threshold measurement is problematic due to acclimatization; but for mathematical comparison, it can be informative.

We also calculated IUHI based on the county level distribution of temperatures at the time of measurement. As such, we used the Urban Thermal Field Variance Index (UTFVI) to evaluate definitions of IUHIs beyond direct assessment of the LST. This is particularly useful for evaluating across multiple dates as LST cannot usefully be combined, but UTFVI can. UTFVI is a simple raster-based metric that calculates the percentage of thermal variance of a particular raster cell (*i*) with respect to the mean (average temperature) such as:

$$UTFVI_i = \frac{T_i - \bar{T}}{\bar{T}} \quad (1)$$

where $\bar{T}$ refers to the mean temperature across all the data under consideration. In other words, the UTFVI indicates how much higher the temperature of an individual pixel is relative to the mean temperature of a given area. This means that the threshold temperature is different not only for each day, but for each resolution. Identification of IUHI in the literature that use UTFVI include all values where UTFVI is greater than 0.02 (Alfraihat et al., 2016; Guha, 2017; Sharma et al., 2021).

A final metric to define hot spots, introduces z-scores as a modification of the UTFVI based on standard deviations.

$$Z_i = \frac{T_i - \bar{T}}{\sigma(T)} \quad (2)$$

where σ refers to the standard deviation. This provides a statistical perspective of thermal variation that can be related to the notion of confidence intervals. As such, IUHIs can here be defined as being 1 standard deviation over the mean ($Z_i > 1$), indicating that these areas are of interest with 68% confidence.

2.3 Aggregation

To calculate the aggregated temperatures, we used ExactExtractR (Baston, 2023), the shapefiles associated with the 2020 census boundary definitions (tracts, block groups, and blocks), and the 2020 zip code tabulation areas (US Census Bureau, n.d.; Walker & Rudis, 2023). This ensures a consistent and reproducible aggregation. For this work, we used the mean to represent the underlying data, although alternative functions are available, most notably median. These aggregation methods were used across definitions for all calculations not performed at the pixel level. The LST was aggregated to each level and then the z-scores

and UTFVI were recalculated for each resolution, so that the mean and standard deviation vary for each resolution (see Figure 3).

The selected polygons also represent common aggregations used to merge multiple data sets as urban data and the notion of neighborhoods is often referenced to census tracts or zip code spaces. Census block groups are created as aggregations of discrete census blocks, and census tracts consist of discrete combinations of block groups. This will be particularly useful in assessing the variation in resolution as the only difference will be the level of aggregation. Although zip code boundaries do not necessarily align to the census boundaries, and are developed for reasons other than data collection, the zip code tabulation areas provide an example of a larger area unit commonly used by researchers. Using the zip code tabulation area allows for comparison between blocks and zip codes, but the zip codes cannot be aligned to block groups or tracts. The relative shape and size of the different aggregations are shown for the study area in Figure 2. The scale bar enables comparison to the 30m pixel width.

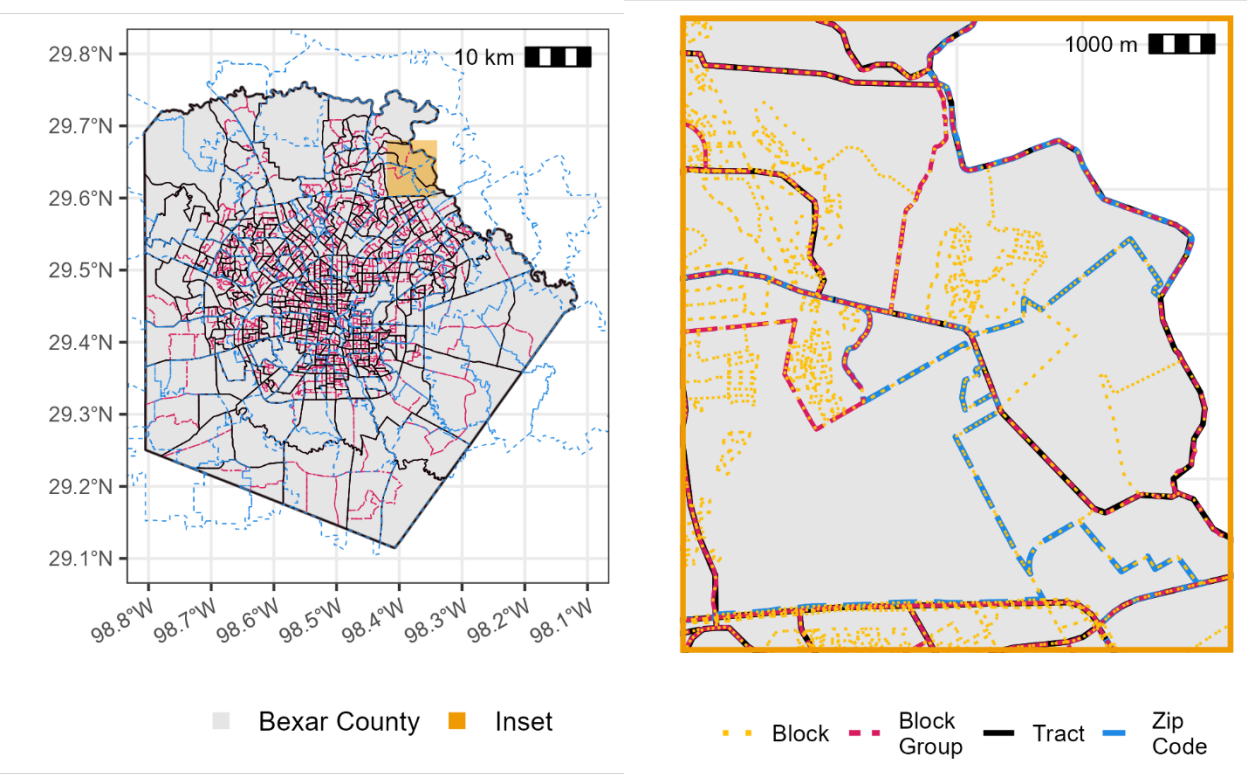

Figure 2 Lines demonstrate the relative size and orientation of the different spatial arrangements considered for the study area. The block boundaries are only shown in the inset map as they represent such a small area.

2.4 Statistical Methods

To aid in the comparison across definitions and resolutions, we utilize a statistical test designed to make inference, the Moran's I. This index can be calculated in its global version to test if there are global patterns of spatial autocorrelation, but in its local version is categorized as a Local Indicator of Spatial Association (LISA) (Anselin, 1995) as it measures the level of

association between an observation's value (e.g., temperature in census block i) and the variable values of i's geographical neighbors (e.g., temperature in census blocks surrounding i). As argued in the seminal article of Anselin, 1995, LISA indices have strong advantages over other indices because they can a) provide an assessment of significant local clustering around specific areas or observations (same as the Gi and Gi* statistics) and b) the identification of areas with specific spatial processes (non-stationarity, outliers, or spatial regimes). Additionally, the LISA version of the Moran's I allows to derive permutation-based inference hence relaxing the normality assumptions needed to derive analytical p-values in other spatial association indices. As such, besides measuring spatial association, the Moran's I can be used to statistically test for the presence of spatial clusters. In its local version, it can be defined as:

$$I_i = z_i \sum w_{ij} z_j \quad (3)$$

where z is a random variable (temperature in this case) measured in standard deviations from the mean or z-scores as defined above, and w_{ij} are spatial weights defined using contiguity-based queen criteria. The main advantage of the Moran's I over other non-inferential statistics (e.g., UTFVI), is that it accounts for not only the location of a specific measurement and its location, but also the variable values (e.g. temperature) of its neighbors (defined by the contiguity matrix). As such, the local Moran's I can provide four clusters: *High-High*, high values surrounded by high value neighbors; *Low-Low*, low values surrounded by low value neighbors; *High-Low*, high values surrounded by low value neighbors; and *Low-high*, low values surrounded by high value neighbors.

This allows for stronger, statistically supported claims to the definition of an IUHI if it belongs to the High-High cluster. This calculation was done at the census tract and census block levels to determine the level of misclassification that occurs when averaging over larger areas. The comparison is performed between blocks and tracts as those are the smallest and largest areas that are fully coincident (i.e. the tract boundaries fully contain a discrete number of blocks).

3. Results and Discussion

The distributions of LST at the varying spatial resolutions are presented in Figure 3. It is apparent that at progressively coarser resolution, the detail is compressed such that the minimum and maximum temperatures are closer to the median. This confirms known artefact introduction through data aggregation. However, in the use case of allocating resources based on relative temperatures, more than the single highest point is of interest.

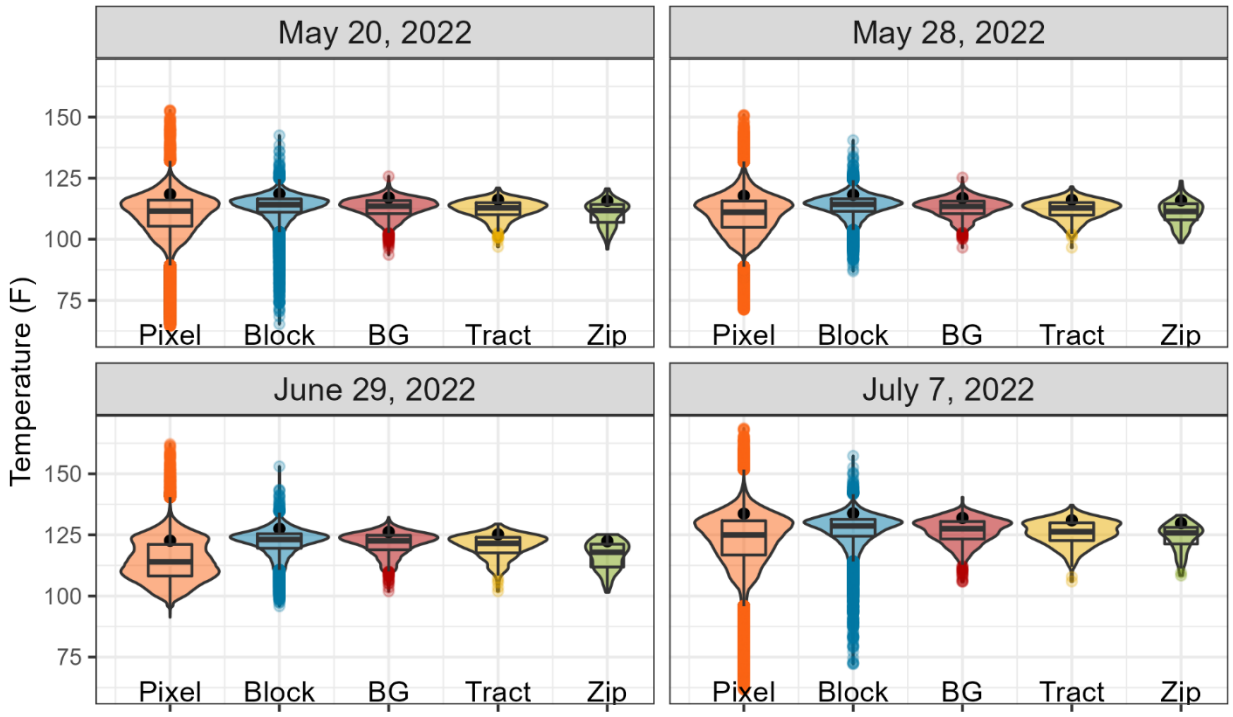

Figure 3: Distribution of LST for four days after spatial aggregation. The box plots show the interquartile range and a black dot on each violin represents the mean temperature plus one standard deviation. The resolution decreases as you move to the right in the figure, with "Pixel" indicating the full raster data for Bexar county, "Block" indicating the Census blocks within the county, "BG" representing the Census Block Groups, "Tract" representing the Census tracts, and "Zip" representing the zip codes in the county.

The total area of Bexar County defined as IUHI using each of the categorization methods described above are presented in Table 1. These calculations represent the complete (pixel-level) data set, and the spatial resolution is examined further in Figure 4. Differences between the methods show up even at this fine resolution. The effect of seasonality on a constant threshold is immediately obvious, as over half the county has a temperature above 120°F by July. We also observe seasonality for UTFVI > 0.02. This means that the rate of temperature increase is not constant across all areas. Considering only temperatures more than 1 standard deviation above the mean temperature takes the distribution itself into account more so than other definitions. For instance, we can see that the bimodal distribution observed on June 29 leads to a larger IUHI as the cluster of hot areas around 125°F in Figure 3 fall outside the normal distribution. This statistical definition is very beneficial in terms of defining the hottest areas in an intuitive way but will also be most affected by a decrease in spatial resolution. From Figure 3, it is apparent that this detail in the distribution is removed for all aggregations.

Table 1: Area (km^2) determined to be IUHI using different definitions at the full raster resolution.

Definition	*May 20*	*May 28*	*June 29*	*July 07*
LST > 120°F	278	266	926	2150
UTFVI >0.02	1451	1227	792	721
Z > 1	446	485	638	435

Figure 4 presents the variation in these IUHI calculations across spatial resolutions. Each definition responds to the loss of data differently. The issues with a constant temperature threshold (dashed line) persist as the data is aggregated to coarser resolution. When we aggregate up to the zip code level, there are far fewer values. Therefore, the relative metrics (Z and UTFVI), may not capture any area. This aggregation level may be useful when evaluating a much larger total area. For instance, a state level evaluation would still have many values at this aggregation level, so this is not an inherent issue with the aggregation itself or of relative metrics but rather an indication that the total area under consideration is an important factor in determining the parameters.

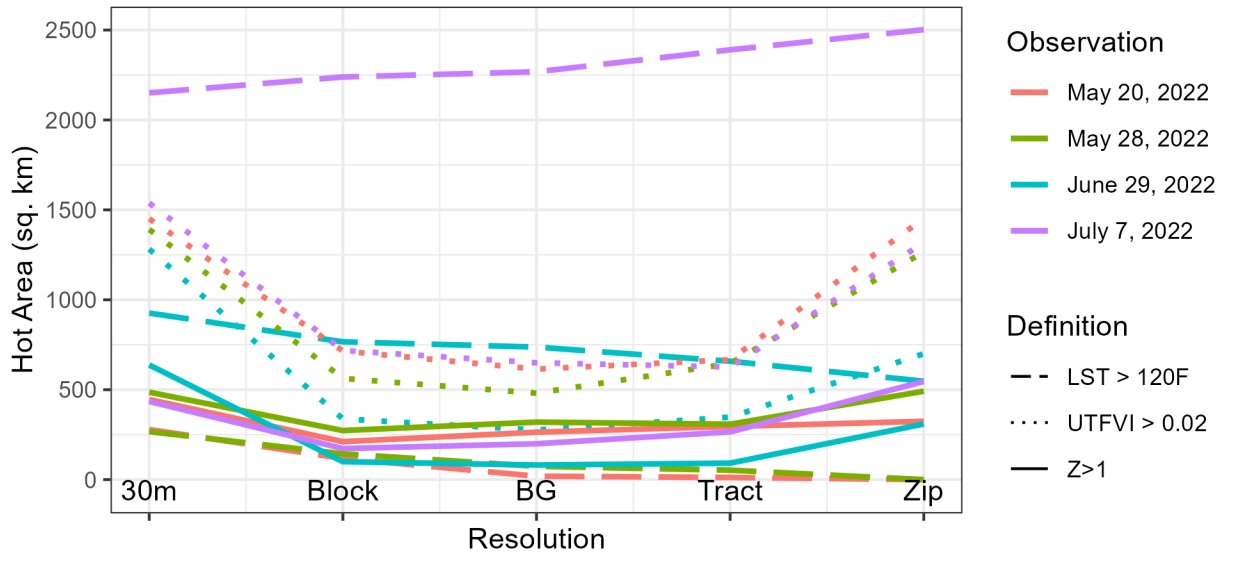

Figure 4: The area identified as IUHI according to three different definitions. The dashed lines represent IUHI identified as those with a temperature greater than 120°F. The solid lines represent the hot areas defined as Z>1. The dotted lines represent hot areas defined with a UTFVI > 0.02. The total county area is 3253 km^2.

Although understanding the size of IUHI is informative, it is also necessary to understand the accuracy of the identification. This section uses the Moran's I index as a way to assess the level of bias incurred when aggregating data to less detailed spatial resolution[1]. These comparisons are made only between the census blocks, block groups, and tracts to ensure that only the spatial resolution is changing. Block groups are made up of a combination of discrete blocks, and tracts are composed of a discrete combination of block groups. First, we compare the

[1] We also conducted the global version of this test yielding the expected results of revealing the presence of global spatial association, which was a supportive indication of specific local clustering identified with the local version of the Moran's I.

smallest and largest of these representations of the county. Figure 5 shows the different area sizes (in km^2) being classified across the different types of significant spatial clusters (inner axis) between two different spatial aggregation levels (outer axis). Consistently identified areas would appear diagonally from top left to bottom right. Of particular importance in determining bias is the area that is identified as High-High (IUHI) at the block level (outer y-axis), that is not significant at the tract level (outer x-axis), or vice versa. As Bexar County moves into warmer months, a larger area classified as High-High at the block level is no longer significant at the tract level. Understanding the variation in temperature may be even more important during the warmest months as the hottest areas are even more hazardous to health at this time.

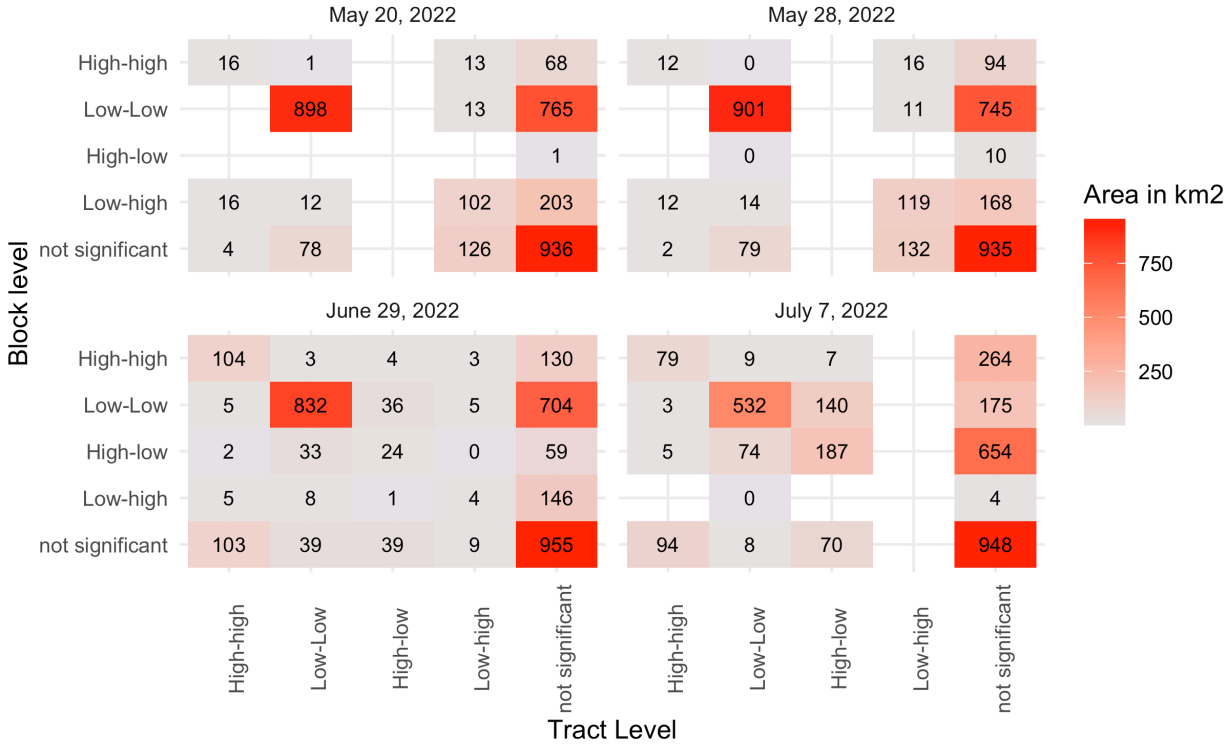

Figure 5. Misclassification area after aggregating from Census Blocks to Tracts. Blocks that are consistently identified appear along the diagonal (top left to bottom right). Areas that lie off this diagonal indicate bias due to aggregation.

This bias can be summarized by counting the number of census blocks that change significance status when aggregating from a finer to coarser level of spatial aggregation. For example, out of the 13,483 block-level IUHI observations, only 46.7% of blocks are correctly classified in the high-high heat cluster (indicating IUHI) at the Tract level. In other words, spatial aggregation creates a 53.3% error as LST values are averaged across census tracts. Additional insight into the intermediate aggregation is also provided in Table 2, which demonstrates that the data loss increases at each step. Note that the group to tract column is aggregating from moderate to high aggregation, while the other columns are considering low aggregation (block) as the base. There is less data loss here as there are at most 9 block groups in a tract compared to tens to hundreds of blocks within the other aggregations. Aggregation introduces both false positive and false negative indicators at nearly equal rates, meaning that there is a true loss of

information. This makes sense based on the distributions in Figure 3, where it is clear that increasing aggregation leads to a loss of both the high and low temperatures. The false positive (type 1) errors indicate that the calculation at the larger spatial resolution would categorize areas as significant that were not. The false negative (type 2) errors indicate that the calculation at the larger spatial resolution would have reported areas as insignificant that were significant at the finer resolution. In addition, we report the IUHI loss as the portion of blocks (or groups) that should have been classified as high-high clusters but were not. An example of this is shown in Figure 6.

Table 2: Total misclassification due to aggregation.

	Block to Block Groups	Groups to Tracts	Block to Tracts	Block to Zip
Total Type 1 Error (False Positive)	34%	25%	43%	44%
Total Type 2 Error (False Negative)	32%	22%	40%	44%
IUHI Loss	51%	21%	53%	54%

4. Conclusions

Irrespective of resolution, certain IUHI definitions demonstrate superior informativeness and consistency. Notably, metrics of variation, such as z-scores and the Urban Thermal Field Variance Index (UTFVI), offer enhanced capabilities for evaluating across different dates, allowing for comparisons of departure from the mean. This stands in contrast to absolute temperature values, which lack this comparative dimension. However, it is crucial to acknowledge that when the resolution becomes too coarse relative to the total area under consideration, insufficient data may impede meaningful comparisons. Given that the misclassification does not have predictable bias in a certain direction, there is no opportunity to improve the performance of identification with coarse data based on simple assumptions. As expected, misclassification increases with aggregation, so each increase in resolution will improve the results. The introduction of both false positives and negatives in IUHI identification, coupled with the observed increase in misclassification during spatial aggregation, underscores the importance of leveraging the finest spatial resolution feasible for decision-making.

As urban heat mitigation gains prominence in city planning and research communities, the limitations inherent in the data must be transparently communicated alongside its applications. Decision-makers relying on data-driven approaches to mitigate IUHI should be aware of the nuanced challenges introduced by varying spatial resolutions. For example, at the neighborhood level, existing studies (Müller et al., 2014; Santamouris et al., 2018; Taleghani et al., 2019; Wang & Li, 2016) have examined heat island mitigation strategies, such as surface material and tree canopy (e.g. (Coccolo et al., 2018; Lee & Mayer, 2018; Zhao et al., 2018), but none of them have answered if the scale of study area matters for the effectiveness of analyzed interventions.

Therefore, we expect further investigation of these issues. Additional analyses should investigate whether the temperature data itself represent socioeconomic variables well, as this would potentially reduce the need to aggregate the finer resolution data. Other investigations that could enable better characterization with existing data include examination of urban morphology such as building density and vegetative cover. Additionally, alternative methods of aggregation, such as weighted averaging could be considered. Although existing analyses show that LST is correlated with air temperatures, further investigation on the representativeness of LST to felt temperature would be useful for improving human health. Additionally, epidemiological studies should also be improved by taking the variation in temperature into account when associating health impacts to exposure. In addition to the errors associated with data aggregation, using zip codes for analysis introduces additional errors, as described by Grubesic (2008). Zip codes also do not align with county boundaries and other policy-level jurisdictions, meaning that implementing changes in response to research findings will be more difficult.

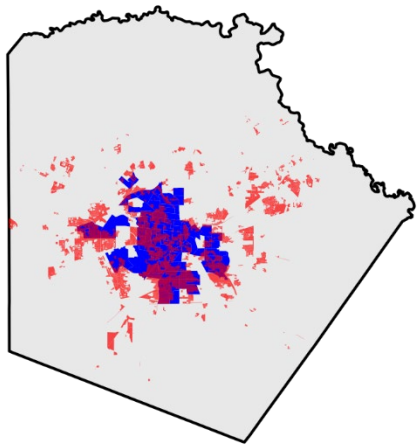

Figure 6. The blue areas are identified as high-high clusters at the tract level on June 29, while the red areas are identified as high-high clusters at the block level. This visually represents the values presented in Figure 5 and Table 2.

Future studies should be conducted to explore the relationship between urban morphology (e.g., building density, vegetative cover) and heat mitigation under different UHI scales. We could also evaluate the effectiveness of various heat mitigation strategies (e.g., increased green spaces, reflective materials, urban layout changes) at different spatial scales, which could help determine the scale at which certain interventions are most effective and how they can be optimized for different urban settings. Also, we should investigate methods to increase the spatial resolution of data collection to minimize aggregation bias. This could involve the use of more sophisticated remote sensing technologies or the integration of multiple data sources to create a higher resolution dataset. This also includes studying the impact of data aggregation on detecting trends, seasonality, and extreme events in urban temperature datasets. Last but not least, researchers need to further investigate methods to integrate LST data with socioeconomic and built environment variables without significant data aggregation bias. This could include

developing new spatial interpolation techniques or using advanced statistical models to better understand the interplay between urban heat, social vulnerability, and urban planning.

Acknowledgments:

The researchers are thankful to the City of San Antonio and its Sustainability Office for the financial support provided at the later stages of this study.

5. References

Alfraihat, R., Mulugeta, G., & Gala, T. S. (2016). Ecological Evaluation of Urban Heat Island in Chicago City, USA. *Journal of Atmospheric Pollution*.

Anderson, B. G., & Bell, M. L. (2009). Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Epidemiology (Cambridge, Mass.)*, *20*(2), 205–213. https://doi.org/10.1097/EDE.0b013e318190ee08

Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Babak Jalalzadeh, F., Rezaul, M., Michael, J. H., Clinton, M. R., Azar, A., Martha, S., Sharon, M., Rachel, L., & Jesse Eugene, B. (2021). Mapping Heat Vulnerability Index Based on Different Urbanization Levels in Nebraska, USA. *Earth and Space Science Open Archive ESSOAr*. https://doi.org/10.1002/essoar.10507450.1

Baston, D. (2023). *exactextractr: Fast Extraction from Raster Datasets using Polygons* [R]. https://github.com/isciences/exactextractr

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, *28*(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Buyantuyev, A., & Wu, J. (2009). Urban heat islands and landscape heterogeneity: Linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology*, *25*(1), 17–33. https://doi.org/10.1007/s10980-009-9402-4

Cartone, A., & Postiglione, P. (2021). Principal component analysis for geographical data: The role of spatial effects in the definition of composite indicators. *Spatial Economic Analysis*, *16*(2), 126–147. https://doi.org/10.1080/17421772.2020.1775876

Centers for Disease Control and Prevention, Geospatial Research, Analysis, and Services Program, & Agency for Toxic Substances and Disease Registry. (2022, December 22). *CDC/ATSDR Social Vulnerability Index 2020 Database US*. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

Chang, Y., Xiao, J., Li, X., Frolking, S., Zhou, D., Schneider, A., Weng, Q., Yu, P., Wang, X., Li, X., Liu, S., & Wu, Y. (2021). Exploring diurnal cycles of surface urban heat island intensity in Boston with land surface temperature data derived from GOES-R geostationary satellites. *Science of The Total Environment*, *763*, 144224. https://doi.org/10.1016/j.scitotenv.2020.144224

Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, *83*(403), 596–610. https://doi.org/10.1080/01621459.1988.10478639

Climate Central. (2023, July 26). Urban Heat Hot Spots. *Climate Matters*. https://www.climatecentral.org/climate-matters/urban-heat-islands-2023

Coccolo, S., Pearlmutter, D., Kaempf, J., & Scartezzini, J.-L. (2018). Thermal Comfort Maps to estimate the impact of urban greening on the outdoor human comfort. *Urban Forestry & Urban Greening*, *35*, 91–105. https://doi.org/10.1016/j.ufug.2018.08.007

Cook, M., Schott, J. R., Mandel, J., & Raqueno, N. (2014). Development of an Operational Calibration Methodology for the Landsat Thermal Data Archive and Initial Testing of the Atmospheric Compensation Component of a Land Surface Temperature (LST) Product from the Archive. *Remote Sensing*, *6*(11), Article 11. https://doi.org/10.3390/rs61111244

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sensing*, *10*(9), Article 9. https://doi.org/10.3390/rs10091363

Ebi, K. L., Capon, A., Berry, P., Broderick, C., de Dear, R., Havenith, G., Honda, Y., Kovats, R. S., Ma, W., Malik, A., Morris, N. B., Nybo, L., Seneviratne, S. I., Vanos, J., & Jay, O. (2021). Hot weather and heat extremes: Health risks. *Lancet (London, England)*, *398*(10301), 698–708. https://doi.org/10.1016/S0140-6736(21)01208-3

Eisenman, D. P., Wilhalme, H., Tseng, C.-H., Chester, M., English, P., Pincetl, S., Fraser, A., Vangala, S., & Dhaliwal, S. K. (2016). Heat Death Associations with the built environment, social vulnerability and their interactions with rising temperature. *Health & Place*, *41*, 89–99. https://doi.org/10.1016/j.healthplace.2016.08.007

Fotheringham, A. S., & Wong, D. W. S. (1991). The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A: Economy and Space*, *23*(7), 1025–1044. https://doi.org/10.1068/a231025

Gober, P., Brazel, A., Quay, R., Myint, S., Grossman-Clarke, S., Miller, A., & Rossi, S. (2009). Using Watered Landscapes to Manipulate Urban Heat Island Effects: How Much Water Will It Take to Cool Phoenix? *Journal of the American Planning Association*, *76*(1), 109–121. https://doi.org/10.1080/01944360903433113

Gober, P., Kirkwood, C. W., Balling, R. C., Ellis, A. W., & Deitrick, S. (2010). Water Planning Under Climatic Uncertainty in Phoenix: Why We Need a New Paradigm. *Annals of the Association of American Geographers*, *100*(2), 356–372. https://doi.org/10.1080/00045601003595420

Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, *42*(2), 129–149. https://doi.org/10.1016/j.seps.2006.09.001

Gu, Y., & You, X. (2022). A spatial quantile regression model for driving mechanism of urban heat island by considering the spatial dependence and heterogeneity: An example of Beijing, China. *Sustainable Cities and Society*, *79*, 103692. https://doi.org/10.1016/j.scs.2022.103692

Guha, S. (2017). Dynamic analysis and ecological evaluation of urban heat islands in Raipur city, India. *Journal of Applied Remote Sensing*, *11*(03), 1. https://doi.org/10.1117/1.JRS.11.036020

Guhathakurta, S., & Gober, P. (2007). The Impact of the Phoenix Urban Heat Island on Residential Water Use. *Journal of the American Planning Association*, *73*(3), 317–329. https://doi.org/10.1080/01944360708977980

Hoffman, J. S., Shandas, V., & Pendleton, N. (2020). The Effects of Historical Housing Policies on Resident Exposure to Intra-Urban Heat: A Study of 108 US Urban Areas. *Climate*, *8*(1), Article 1. https://doi.org/10.3390/cli8010012

Holt, D., Steel, D. G., Tranmer, M., & Wrigley, N. (1996). Aggregation and Ecological Effects in Geographically Based Data. *Geographical Analysis*, *28*(3), 244–261. https://doi.org/10.1111/j.1538-4632.1996.tb00933.x

Huang, G., Zhou, W., & Cadenasso, M. L. (2011). Is everyone hot in the city? Spatial pattern of land surface temperatures, land cover and neighborhood socioeconomic characteristics in Baltimore, MD. *Journal of Environmental Management*, *92*(7), 1753–1759. https://doi.org/10.1016/j.jenvman.2011.02.006

Jagarnath, M., Thambiran, T., & Gebreslasie, M. (2020). Heat stress risk and vulnerability under climate change in Durban metropolitan, South Africa—Identifying urban planning priorities for adaptation. *Climatic Change*, *163*(2), 807–829. https://doi.org/10.1007/s10584-020-02908-x

Lee, H., & Mayer, H. (2018). Maximum extent of human heat stress reduction on building areas due to urban greening. *Urban Forestry & Urban Greening*, *32*, 154–167. https://doi.org/10.1016/j.ufug.2018.04.010

Li, M., Gu, S., Bi, P., Yang, J., & Liu, Q. (2015). Heat Waves and Morbidity: Current Knowledge and Further Direction-A Comprehensive Literature Review. *International Journal of Environmental Research and Public Health*, *12*(5), Article 5. https://doi.org/10.3390/ijerph120505256

Malakar, N. K., Hulley, G. C., Hook, S. J., Laraby, K., Cook, M., & Schott, J. R. (2018). An Operational Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(10), 5717–5735. https://doi.org/10.1109/TGRS.2018.2824828

Martin, P., Baudouin, Y., & Gachon, P. (2014). An alternative method to characterize the surface urban heat island. *International Journal of Biometeorology*, *59*(7), 849–861. https://doi.org/10.1007/s00484-014-0902-9

Melaas, E. K., Wang, J. A., Miller, D. L., & Friedl, M. A. (2016). Interactions between urban vegetation and surface urban heat islands: A case study in the Boston metropolitan region. *Environmental Research Letters*, *11*(5), 054020. https://doi.org/10.1088/1748-9326/11/5/054020

Milando, C. W., Black-Ingersoll, F., Heidari, L., López-Hernández, I., de Lange, J., Negassa, A., McIntyre, A. M., Martinez, M. P. B., Bongiovanni, R., Levy, J. I., Kinney, P. L., Scammell, M. K., & Fabian, M. P. (2022). Mixed methods assessment of personal heat exposure, sleep, physical activity, and heat adaptation strategies among urban residents in the Boston area, MA. *BMC Public Health*, *22*(1), 2314. https://doi.org/10.1186/s12889-022-14692-7

Müller, N., Kuttler, W., & Barlag, A.-B. (2014). Counteracting urban climate change: Adaptation measures and their effect on thermal comfort. *Theoretical and Applied Climatology*, *115*(1), 243–257. https://doi.org/10.1007/s00704-013-0890-4

Nakamura, S., & Aruga, T. (2013). *Epidemiology of Heat Illness*. *56*(3).

Nayak, S. G., Shrestha, S., Kinney, P. L., Ross, Z., Sheridan, S. C., Pantea, C. I., Hsu, W. H., Muscatiello, N., & Hwang, S. A. (2018). Development of a heat vulnerability index for New York State. *Public Health (London)*, *161*, 127–137. https://doi.org/10.1016/j.puhe.2017.09.006

Piantadosi, S., Byar, D. P., & Green, S. B. (1988). THE ECOLOGICAL FALLACY. *American Journal of Epidemiology*, *127*(5).

Reid, C. E., O'Neill, M. S., Gronlund, C. J., Brines, S. J., Brown, D. G., Diez-Roux, A. V., & Schwartz, J. (2009). Mapping Community Determinants of Heat Vulnerability. *Environmental Health Perspectives*, *117*(11), 1730–1736. https://doi.org/10.1289/ehp.0900683

Santamouris, M., Haddad, S., Saliari, M., Vasilakopoulou, K., Synnefa, A., Paolini, R., Ulpiani, G., Garshasbi, S., & Fiorito, F. (2018). On the energy impact of urban heat island in Sydney: Climate and energy potential of mitigation technologies. *Energy and Buildings*, *166*, 154–164. https://doi.org/10.1016/j.enbuild.2018.02.007

Scalley, B. D., Spicer, T., Jian, L., Xiao, J., Nairn, J., Robertson, A., & Weeramanthri, T. (2015). Responding to heatwave intensity: Excess Heat Factor is a superior predictor of health service utilisation and a trigger for heatwave plans. *Australian and New Zealand Journal of Public Health*, *39*(6), 582–587. https://doi.org/10.1111/1753-6405.12421

Schroeder, J. P., & Van Riper, D. C. (2013). Because Muncie's Densities Are Not Manhattan's: Using Geographical Weighting in the Expectation–Maximization Algorithm for Areal Interpolation. *Geographical Analysis*, *45*(3), 216–237. https://doi.org/10.1111/gean.12014

Sharma, R., Pradhan, L., Kumari, M., & Bhattacharya, P. (2021). Assessing urban heat islands and thermal comfort in Noida City using geospatial technology. *Urban Climate*, *35*, 100751. https://doi.org/10.1016/j.uclim.2020.100751

Stewart, I. D. (2011). A systematic review and scientific critique of methodology in modern urban heat island literature. *International Journal of Climatology*, *31*(2), 200–217. https://doi.org/10.1002/joc.2141

Taleghani, M., Crank, P. J., Mohegh, A., Sailor, D. J., & Ban-Weiss, G. A. (2019). The impact of heat mitigation strategies on the energy balance of a neighborhood in Los Angeles. *Solar Energy*, *177*, 604–611. https://doi.org/10.1016/j.solener.2018.11.041

US Census Bureau, U. C. (n.d.). *TIGER/Line Shapefiles*. Census.Gov. Retrieved September 25, 2023, from https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

USGS. (2023). *Landsat Collection 2 U.S. Analysis Ready Data* [dataset]. Landsat Collection 2 U.S. Landsat Analysis Ready Data (ARD) Level-2. https://doi.org/10.5066/P960F8OC

Walker, K., & Rudis, B. (2023). *tigris: Load Census TIGER/Line Shapefiles* [Computer software]. https://cran.r-project.org/package=tigris

Wang, X., & Li, Y. (2016). Predicting urban heat island circulation using CFD. *Building and Environment*, *99*, 82–97. https://doi.org/10.1016/j.buildenv.2016.01.020

Wong, M. S., Peng, F., Zou, B., Shi, W. Z., & Wilson, G. J. (2016). Spatially Analyzing the Inequity of the Hong Kong Urban Heat Island by Socio-Demographic Characteristics. *International Journal of Environmental Research and Public Health*, *13*(3), 317. https://doi.org/10.3390/ijerph13030317

Zhao, Q., Sailor, D., & Wentz, E. (2018). Impact of tree locations and arrangements on outdoor microclimates and human thermal comfort in an urban residential environment. *Urban Forestry and Urban Greening*, *32*, 81–91. https://doi.org/10.1016/j.ufug.2018.03.022