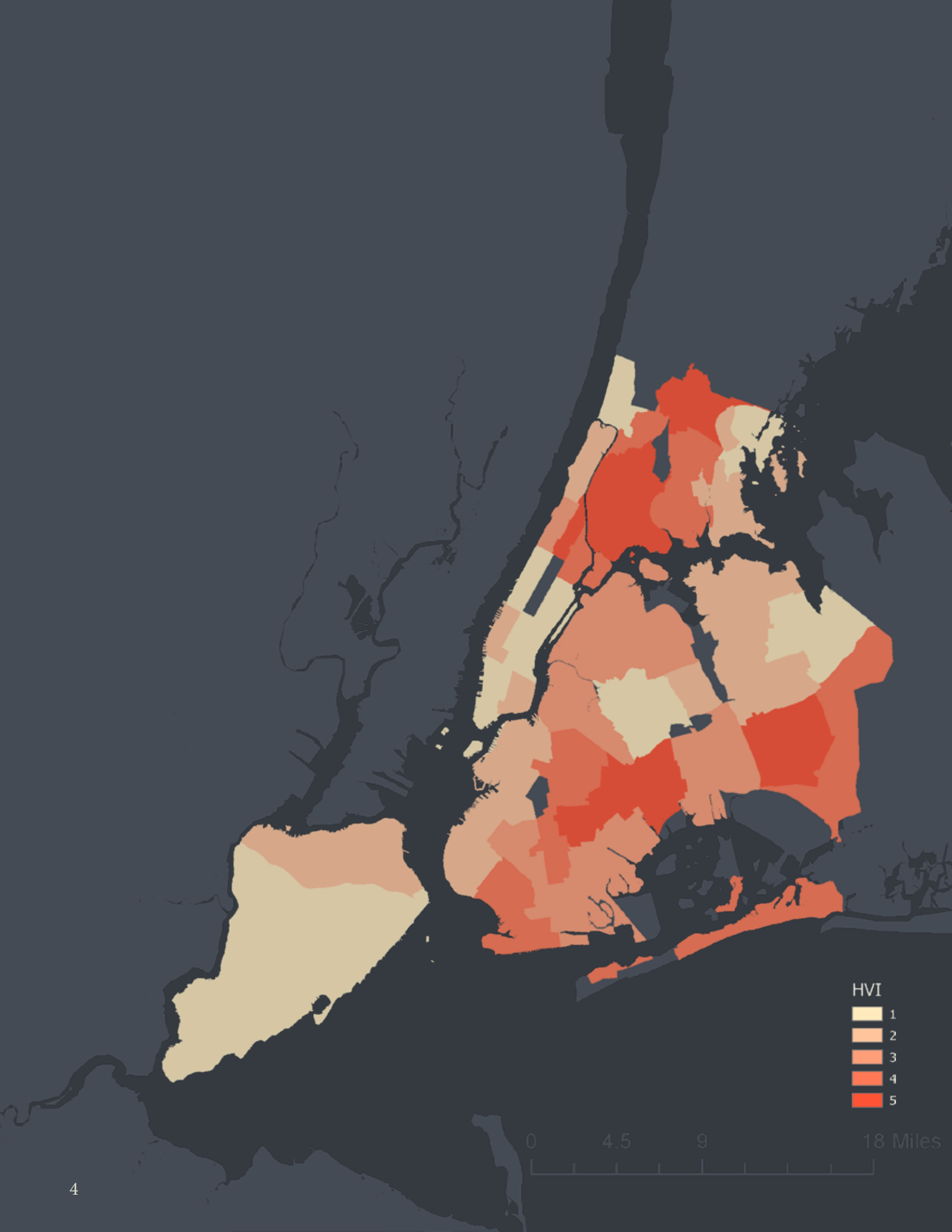
(Re)Detecting & (Re)Examining Heat Vulnerability Index

Mapping Vulnerability and Adaptive Capacity in New York City

Eric Xia, Hui Chen & Mina Wei
Advanced Spatial Analysis
Spring 24

Prof. Leah Meisterlin

Table of Content


HVI
1
2
3
4
5
0
4.5
9
18 Miles

INTRODUCTION

The assessment and mitigation of heat vulnerability in urban environments are critical considerations in contemporary geographic research, particularly in the context of climate change and urbanization. Drawing inspiration from McHarg's Ecological Method, which emphasizes the integration of natural and built environment factors in geographic analysis, this research endeavors to enhance our understanding of heat vulnerability by combining layered transparency mapping with advanced geospatial methodologies.

At the heart of this study lies the exploration of the Heat Vulnerability Index (HVI), While the HVI provided by the NYC Environmental & Health Data Portal offers valuable insights into neighborhood-level heat vulnerability, its current resolution at census tracts presents limitations for in-depth analysis. Moreover, the exclusion of certain demographic and socioeconomic variables underscores the need for refinement and augmentation of existing methodologies.

This research aims to address these limitations by employing a suite of geoprocessing techniques, including Principal Component Analysis (PCA), ISO Clustering, Anselin Local Moran's I, and Natural Break Reclassification. By leveraging multiple geographic information software and algorithms, we seek to elucidate the impact of different processing techniques on the spatial distribution and accuracy of the HVI. Additionally, we aim to investigate the influence of factors such as data resolution, interpolation algorithms, and modeling assumptions on the construction and interpretation of the HVI.

The Heat Vulnerability Index provided by NYC Environmenal & Health Data Portal provides an estimation of the heat vulnerability of neighborhoods in the city and charactertistcis that affect the vulnerability, including ***surface temperature****,* ***green space****,* ***access to home air conditioning****, and* ***non-white population****. Although the index provides useful infwormation for assuming heat vulnerability in the city, the data resolution is limited at census tracts, which are too coarse for an in-depth research. In addition, the current HVI does not factor into the impact of demographic composition, income, and disability in determining heat vulnerability, which leaves space for furthur modification and improvement with new methodologies.*

RESEARCH QUESTIONS

1. What is the impact of different geoprocessing techniques on the HVI?

2. How do different geoprocessing techniques infl uence the spatial distribution and accuracy of the HVI?

- To what extent do factors such as data resolution, interpolation algorithms, and modeling assumptions impact the construction and interpretation of the HVI?

- How can the choice of geoprocessing technique affect the identifi cation of high-risk areas and the prioritization of interventions to mitigate heat-related risks in urban environments?

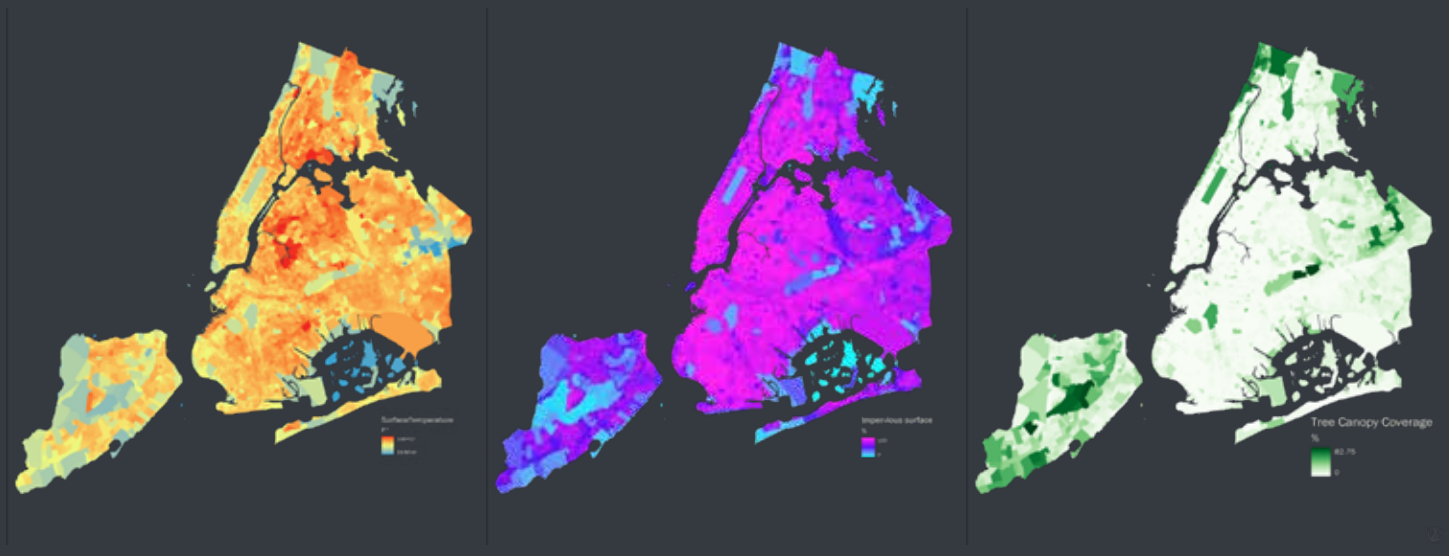
DATA CLEANING

To prepare essential data for indexing and weighting processes and future analysis, initial data cleaning is necessary for climate, built environment and demographic datasets. As climate and built environment data is by 15m x 15m resolution in raster, while demographic data is based on census block group unit in vectors, different data cleaning processes apply to the two data sources.

Cleaning and Preparing Raster Dataset

Using specifically tailored programs and codes, temporary surface temperature data is extracted from the surface temperature bundle in Landsat data provided by USGS. To estimate the highest potential vulnerability, mean surface temperature (in Fahrenheit) is calculated from July-to-August surface temperature data in 2021. As important components that can reflect factual/sensible heat in vulnerability assessment, the updated data in 2021 of impervious surface from NLCD Imperviousness Dataset and canopy coverage from NLCD Tree Cover Dataset are used in the analysis. These datasets are then clipped by the administrative boundary of New York City and water/land boundary, provided by the city planning department. To unify the scope of the weighting model, zonal statistics (using zonal average) are applied to turn continuous raster values into discrete values in the unit of census block groups in New York City.

After data cleaning, each census block groups are assigned data of 1) Surface temperature (in Fahrenheit); 2) Percentage of Impervious Surface (%); 3) Percentage of Tree Canopy Coverage (%).



Surface Temperature *Impervious Surface* *Tree Canopy Coverage*

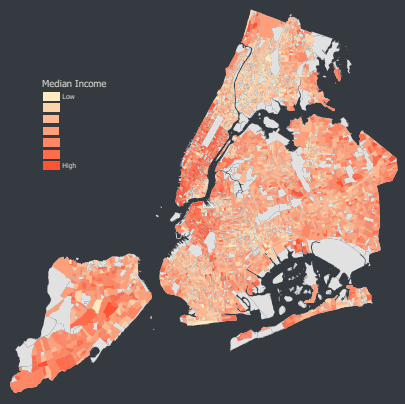
Cleaning and Preparing Vector Dataset

Demographic and socio-economic data used in the research are provided by two sources: The ACS Community Survey Data (most recent version used) provides information related to age and sex, racial components, income information, and disability conditions (with limited statistical information); The Decennial Census Data (the 2020 report used) provides basic population data and documents change in census block groups over the decades. The datasets are processed with unified target identification codes for table joins. Similar to the raster processing operations, data are clipped by administrative boundaries and water/land lines. In addition, census block groups with zero population, where no vulnerability signals can be reflected, are deleted from the final datasets. The final vector data table, in unit of census block groups, contains 1) Percentage of infant/child population under 5 (%); 2) Percentage of elderly population above 65 (%); 3) Average median household income (in $, inflation factored); 4) Percentage of non-white population (%); 5) Percentage of population below poverty line (%); 6) Percentage of disability among population of workforce age (20-64, in %).

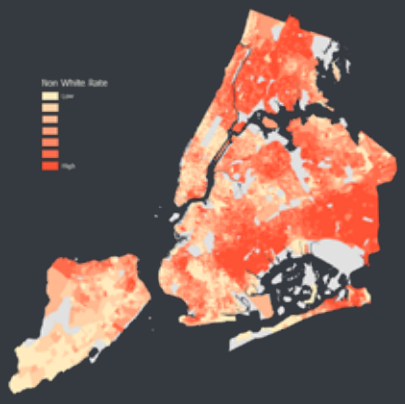
In addition to the vector data outcomes for analysis that follows, the NYC Heat Hazard Map provided by the city's Environment Office is re-mapped with the same metrics as other maps in the research, serving as a visual reference for difference observation. The social vulnerability map provided by USCDC (in the New York area), and obesity rate data provided by NYCCHS are also incorporated as backup references during the following stages.

Disability Rate

Elderly Population



Median Income



Non-White Rate

Poverty Rate

Youth Population

METHODOLOGY

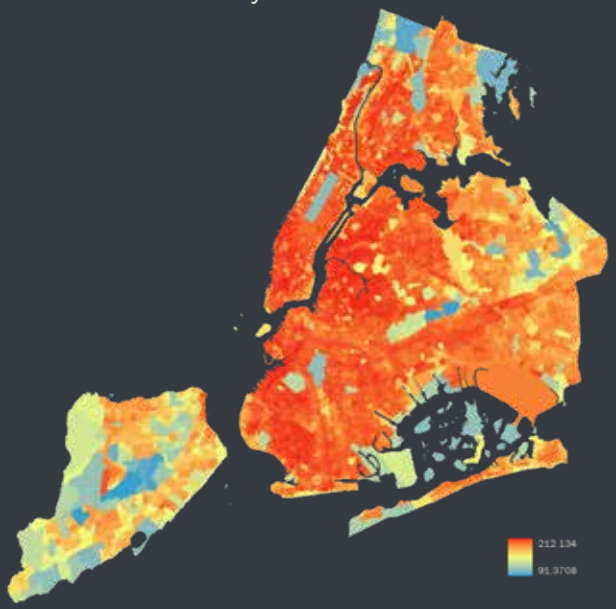
Raster Method 1: Principle Component Analysis (PCA)

Principal Component Analysis (PCA) is a powerful technique used in geospatial analysis to show the underlying patterns and structures within multiband raster datasets. It is commonly employed to reduce the dimensionality of such datasets while preserving the essential information contained within them. PCA operates as an unsupervised method, making it particularly useful for exploring complex relationships within data without the need for labeled samples.

PCA works by transforming the original raster bands into a new set of variables called principal components. These components are linear combinations of the original bands and are ordered such that the first component captures the maximum variance present in the data, with subsequent components capturing progressively less variance. This allows for the identification of the most influential factors driving variability within the dataset. The number of principal components chosen determines the dimensionality of the output multiband raster, with each component representing a unique aspect of the data's variability. By examining the percentage of variance explained by each component, analysts can gain insights into the correlation of different bands in contributing to overall dataset variability. This information is valuable for understanding the key drivers behind geospatial phenomena and can guide subsequent analyses or decision-making processes.

Raster Method 2: ISO Clustering

ISO Clustering is an unsupervised geospatial machine-learning technique that can be applied to determine the characteristics of canopy cover, surface type, and temperature. Similar to PCA, the ISO Clustering method predicts the maximum likelihood of classification types with reference to data performance. Instead of examining the predominant impact factors in PCA, ISO Clustering method creates multiple scenarios with orders where data performs alike. The clusteirng scenarios, which reflect comprehensive heat impact, are reclassified into 14 scores , where the higher value reflects scenarios with higher heat vulnerability.



PCA Results

ISO Clustering Results

Vector Method 1: Clustering Analysis through Anselin Local Moran's I

Anselin Local Moran's I is a local spatial autocorrelation statistic based on Moran's I statistics, which is expected to give an indication of the extent of significant spatial clustering of similar values, and acts as a proportional indicator of global spatial association (NKU, 1999). Each of the six demographic data is used as the point value for Local Moran's I analysis. The geoprocessing is set with default value of maximum study distance and the number of bands. Results from Local Moran's I will recognize five types of clustering based on P values and Z scores (and their relationship), which are high-high clusters, high-low outliers, low-high outliers, low-low clusters, and points with no significant relationship (ESRI, 2024).

Compared to observing absolute values in forming HVI, Anselin Local Moran's I analysis provides researchers with opportunities to detect clusters of risky factor concentrations that better signal real-world vulnerability. Within the research context, high-high (HH) clusters reflect the most heat-vulnerable areas that deserve extra attention, high-low (HL) outliers reflect areas with significant heat vulnerability due to certain risky demographic factors (low-high (LH) outliers reflect the opposite situation), while low-low (LL) clusters reflect regions that are relatively safe from extreme heat. Focusing on extreme value and vulnerable region clusters, HH block groups are assigned a score of 7 in the model, HL with 5, LH with 3, and LL (the safest scenario) assigned 1. Blocks with no significant clusters, with fitting P values, are regarded as communities with medium vulnerability conditions and assigned a score of 3.

Vector Method 2: Natural Break Reclassification

Natural break is a set of algorithms developed by ESRI that can document and detect large differences in data values automatically and can be used to reclassify data into desired ranks and classes. Compared to the equal interval or quantile classification method, natural break classification displays the trend and distribution of HVI by levels more obviously, making it easier for researchers to classify heat vulnerability (de Smith et al. 2021). The research set 7 classes using the natural break method and assigns scores from 1 to 7 to the reclassified demographic data, where 1 reflects the lowest rank (showing low risk) and 7 reflects the highest rank (showing highest heat vulnerability).

Weighted HVI Score Modelling

Based upon previous research works that manually assign weights to different HVI impact factors, the research recognizes the prevalent influence of poverty and disability, which are two dominant factors in determining individual vulnerability (Johnson et al. 2012). Considering that the existing data on disability rates is inconclusive in which child and elderly disability rates are not included, the model assigns the heaviest weight (base score x3) on the poverty rate and heavier weight on disability rate (base score x2) in the final calculation stage. Average income and non-white composition are unweighted (x1), while the influence of child and elderly population percentages are combined (x0.5 each). The total possible score for demographic data in the model is thus 56.

The impact of natural climate and built environment factors have a more profound impact in determining heat vulnerability, as shown in studies conducted by Chen et al. (2022) in Taipei and Yoo et al. (2023) in Seoul. Following previous simplified weighted models, the research assigns a 2x importance for any climate-related and/or built-environment factors, hereby the processed raster data, to the final model. The total possible score for natural data in the model is 112, which is eight times of the base reclassification score after combining three different impact sources.

Combining these two sectors, the final model reflects an adjusted Heat Vulnerability Index with a total possible score of 168. While detailed influential factors vary, a higher score here shows higher vulnerability in the face of heat activites.

COMPARISON OF HVI CONSTRUCTION METHODS

To quantifiably compare the combination outcome of two vector methods – Anselin Local Moran's I and Natural Break, and two raster methods – ISO clustering and Principal Component Analysis, the research uses a 2x2 matrix that include all potential combination outcomes (one raster method plus one vector method). The matrix makes it easier to compare the data performance with the variation of each individual method and can demonstrate clearly howA difference in vector or raster methodologies lead to different data behaviors. Potential outcomes in the matrix include:

1. HVI index using Anselin Local Moran's I to evaluate scoio-economic factors, and PCA for natural and built-environment data.
2. HVI index using Anselin Local Moran's I to evaluate scoio-economic factors, and ISO Clustering for natural and built-environment data.
3. HVI index using Natural Break Classification to evaluate socio-economic factors, and PCA for natural and built-environment data.
4. HVI index using Natural Break Classification to evaluate socio-economic factors, and ISO Clustering or natural and built-environment data.

For all the HVI outcomes calculated, a higher value, which is reflected by deeper reddish color on the map, shows that the specific area is more vulnerable to heat activities and heat extremes, including but are not limited to heat waves, continuous days of high temperature, and non-seasonal heat abnormalities.

However, each of these methods applied in the research has slightly different targets of detection: While the natural break classification look for the distribution and ranking of absolute values (in single census block groups), the Anselin Local Moran's I detects clustering that reflects regions of concentrated heat vulnerability; While the PCA analysis figures out the single most predominant built-environment factors that construct the heat index, ISO clustering finds the pattern of data distribution with relevance to all of the natural built-environment factors. Therefore, different outcome of HVI patterns symbolize a difference in central focus in constructing the index (out of the selection of variables) and will be analyzed both individually and collaboratively.

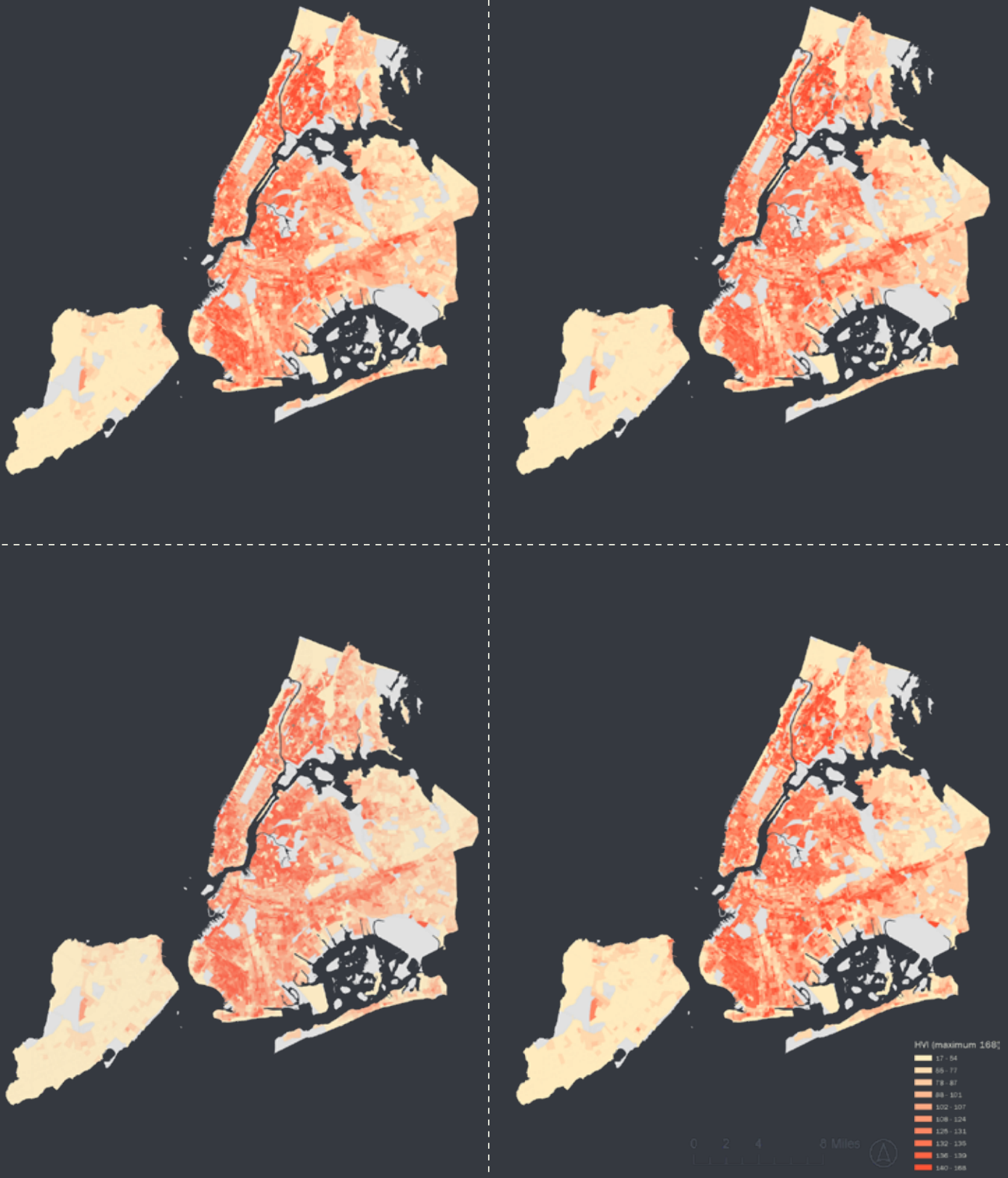
Visualizing all HVI outcomes through different combination of spatial analysis methods, the research will specially focus on the central Bronx area, where all historical heat vulnerability indices have shown high heat vulnerability. Zooming into the area of high risk, it is easier and more useful to compare the functionality of different methods as well as their outcomes and the social impacts.

RASTER METHOD 1: PRINCIPAL COMPONENT ANALYSIS

RASTER METHOD 2: ISO CLUSTERING

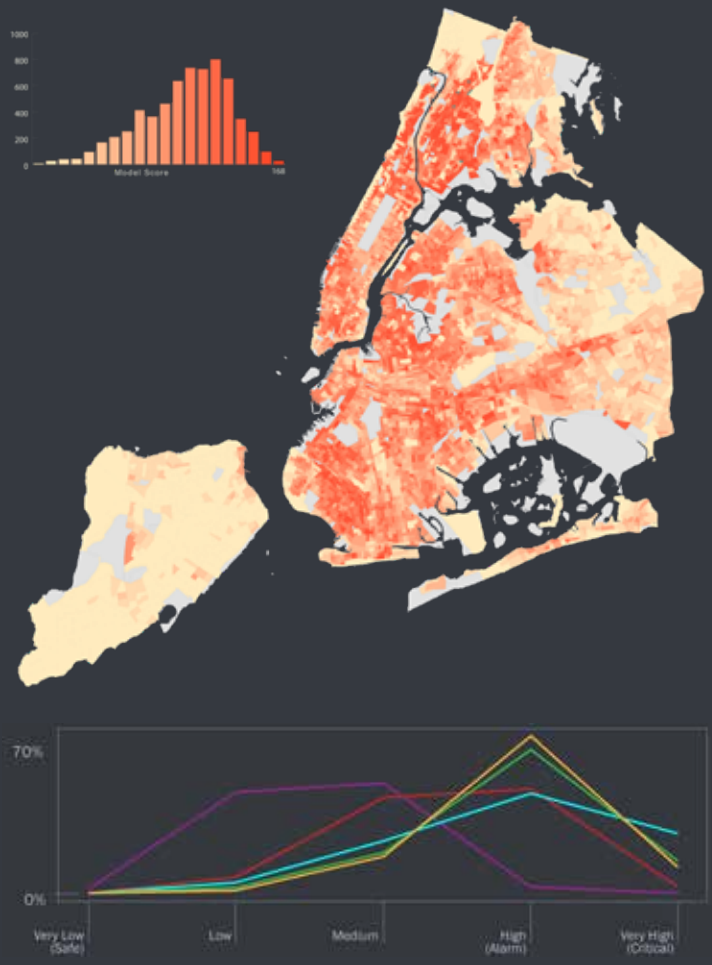
VECTOR METHOD 1: MORAN'S I

VECTOR METHOD 2: NATURAL BREAK CLASSIFICATION

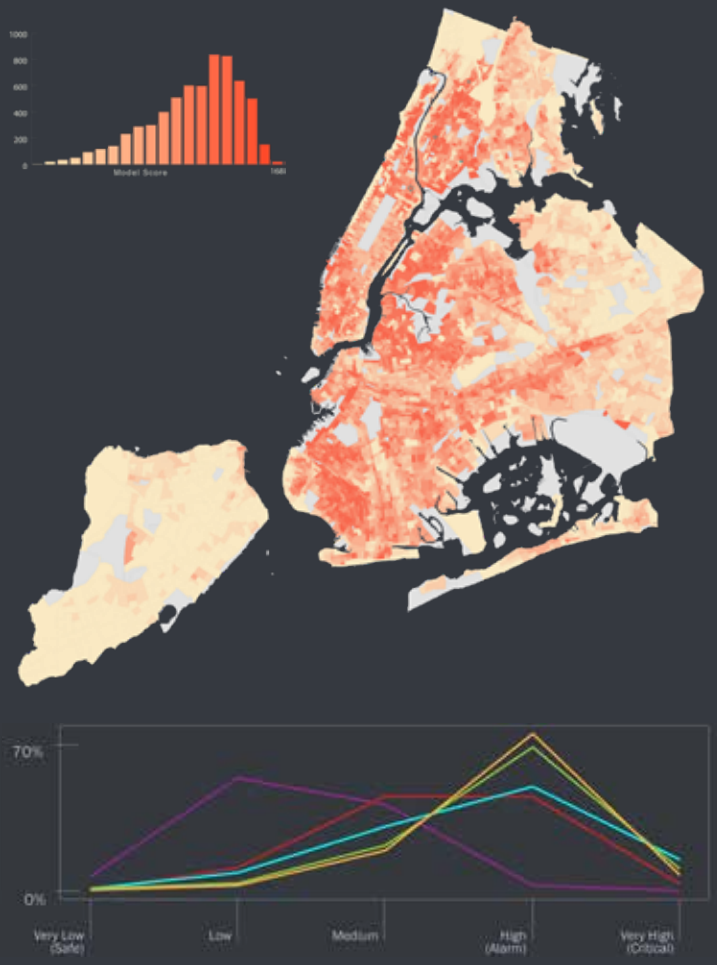


ALL MAPS

PCA - MORAN'S I



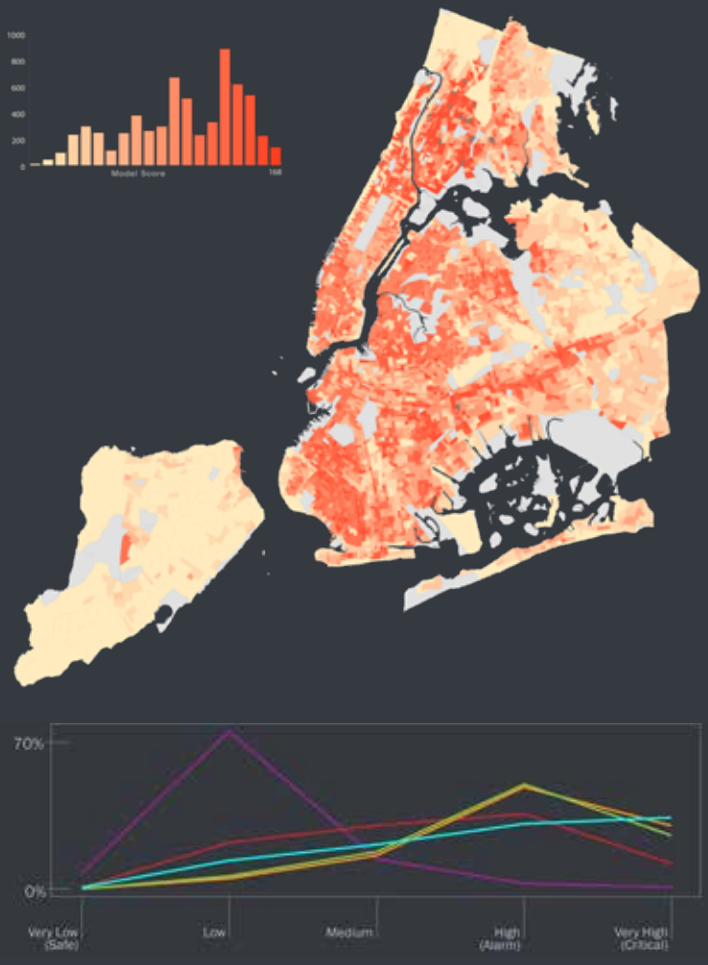
PCA - NATURAL BREAK



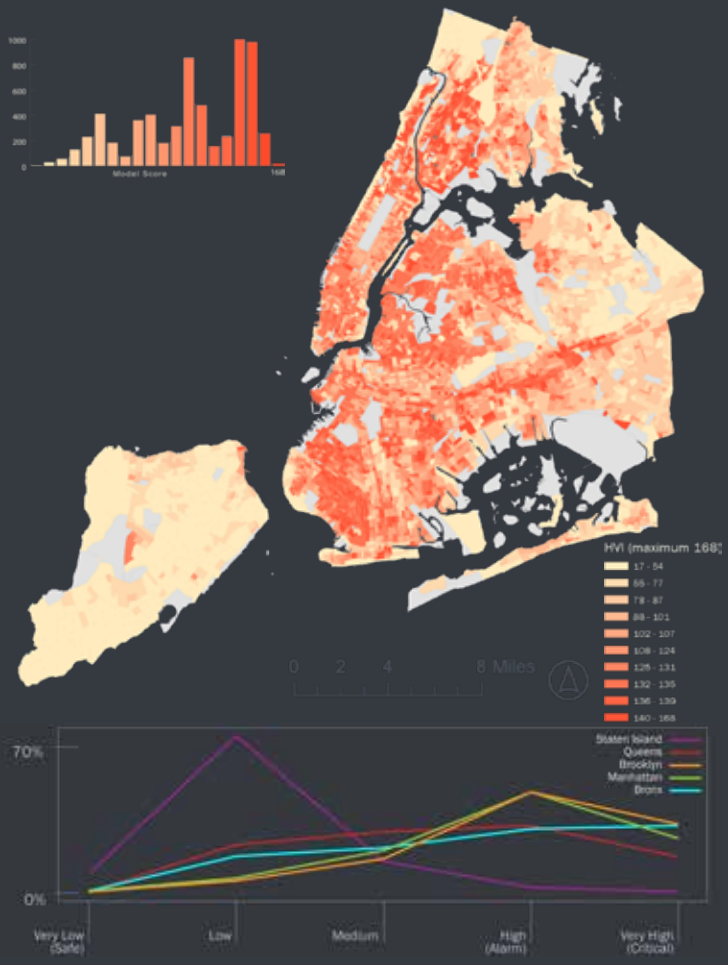
The mapping of HVI across the New York City provides a clear visualization of heat vulnerability by individual census block groups and by regions in five boroughs. All four methods witness high HVI values in Central and Southern Bronx, Lower-West Manhattan, Southern Brooklyn, and Central Queens, indicating that communities in these regions are more vulnerable to heat-related risks and extreme events. Among these areas, Central and Southern Bronx has the highest overall heat vulnerability, reflecting the urgency of heat mitigation efforts to be made in the region. By comparison, all four models have shown that most census block groups in Staten Island have a lower HVI, along with area in Eastern Queens and Northwestern Bronx, indicating that these neighborhoods are more resilient to heat-related risks and extreme events.

A high HVI shown on the map may be caused by different combination of heat-risky factors. For example, selected Lower-West Manhattan neighborhoods witness a high heat vulnerability mostly due the dominant presence of impervious surface, very low tree canopy coverage, and (causally) a high surface temperature. By comparison, other regions with a high HVI see a larger

ISO - MORAN'S I



ISO - NATURAL BREAK



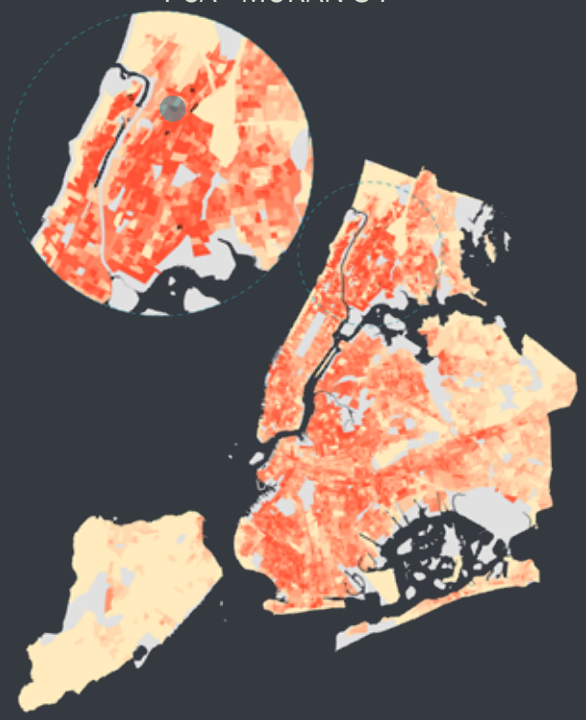
contribution of socio-demographic factors, which are largely contributed by combinations of potentially high percentage of the elderly and children population, low median household income, high poverty rate, high disability rate (across working age population), and high non-white population percentage.

Despite similarities in overall trends, each model reflects a subtle difference in top-ranked census block groups, as well as difference in data distribution. Indices built upon PCA model shows relatively normally distributed data, while the PCA-NaturalBreak combination shows a more right-tilting distribution pattern with more census block groups being categorized into the lower value sphere. On the other hand, indices built upon ISO model shows a more separated data distribution, while the ISO-NaturalBreak combination reflects the most separated and isolated pattern.

ALL MAPS - BRONX

A more detailed focus on Central and Southern Bronx reflects the minor HVI difference by different modelling methods under a larger scale. As shown in the enlarged maps, the four indexing methods indicate different census block groups with highest HVI index scores, due to the variance in classification methodology and detection targets. The detailed difference and selected characteristics of top-scorers are visualized below.

PCA - MORAN'S I



PCA - NATURAL BREAK

Demographic Info
(Highest HVI CBGs)

160/168

Block Group 3; Census Tract 401

Total population: 507

Children (<5): 16.8%; Elderly (>65): 47.9%

Non-white: 66.9%; Disability:14.5%

Median Income: $28.638; Poverty:20.7%

151/168

Block Group 1; Census Tract 197

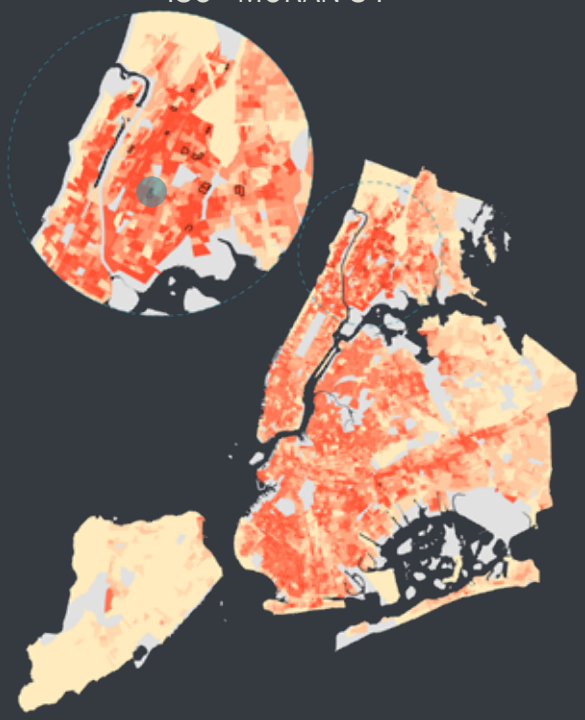
Total population: 2149

Children (<5): 11.2%; Elderly (>65): 9.7%

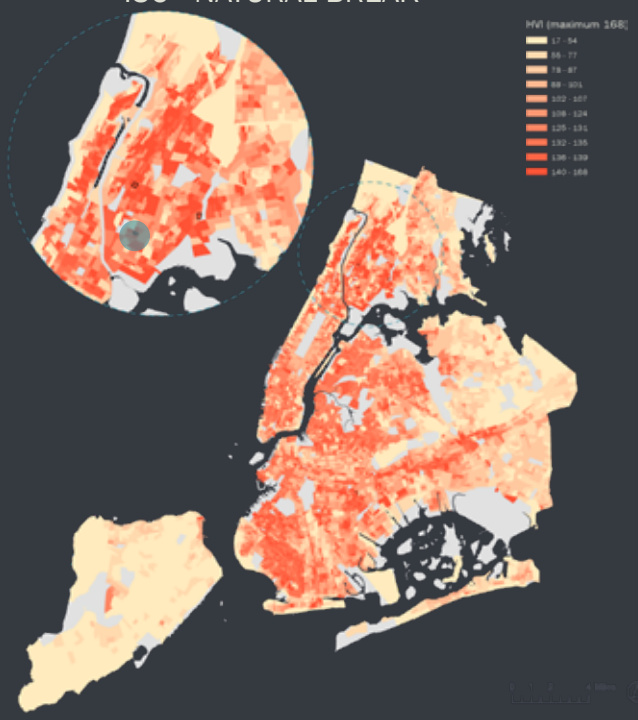
Non-white: 97.3%; Disability:27.6%

Median Income: $11,787; Poverty:75.4%

ISO - MORAN'S I



ISO - NATURAL BREAK



160/168

Block Group 3; Census Tract 177.01

Total population: 2409

Children (<5):11.2%; Elderly (>65): 18.0%

Non-white: 91.3%; Disability: 22.0%

Median Income: $32,935; Poverty: 45.6%

155/168

Block Group 5; Census Tract 67

Total population: 331

Children (<5): 0%; Elderly (>65): 69.8%

Non-white: 83.4%; Disability: 62.3%

Median Income: $10,440; Poverty: 92.7%

ANALYSIS & COMPARISON

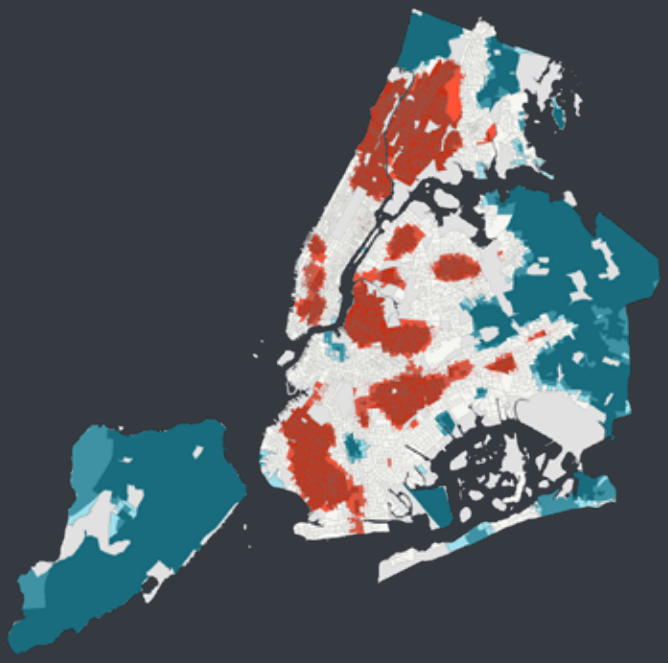
In addition to visualizing and comparing the distribution of HVI in the New York City, quantitative analysis techniques are required to further detect and categorize the similarity and difference in data performance across different HVI value. After multiple attempts and discussions, the research chooses to use similarity value (contrast), Gi hot-spot detection and comparison, as well as kernel density functions to compare the performance and effectiveness of these HVI construction methods more precisely. The comparison sets in this chapter (A total of 5) will follow the visualization matrix, which include:

1. With both HVI using PCA to analyze natural and built-environment components, comparing between Asselin Local Moran's I versus Natural Break Reclassification in shaping HVI values.
2. With both HVI using ISO clustering to analyze natural and built-environment components, comparing between Asselin Local Moran's I versus Natural Break Reclassification in shaping HVI values.
3. With both HVI using Anselin Local Moran's I to analyze socio-economic components, comparing between PCA and ISO Clustering in shaping HVI values.
4. With both HVI using Natural Break Classification to analyze socio-economic components, comparing between PCA and ISO Clustering in shaping HVI values.

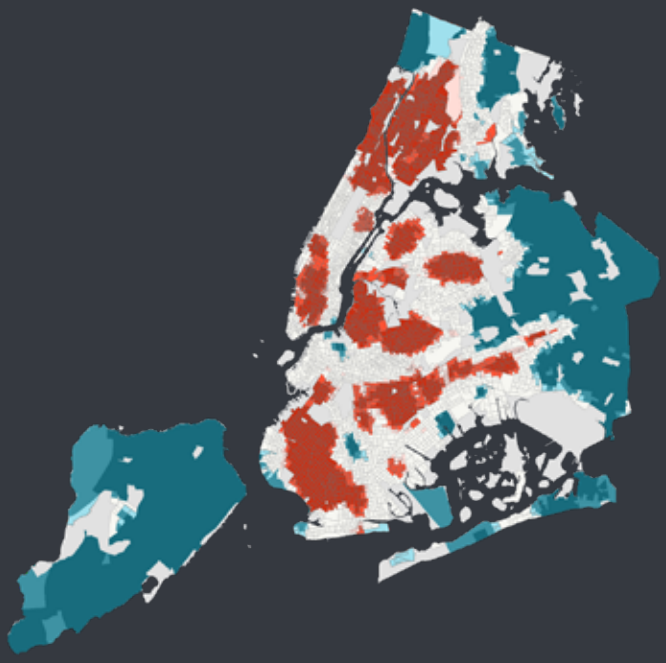
The five models used for forming the comparison matrix can be divided into three categories, which are 1) Similarity comparison, in which the similarity between HVI values (per census block groups is calculated; 2) Hotspot comparison, in which hotspots for HVI index value (by census block groups) are detected and compared; 3) Kernel density comparison, in which the hotspots for HVI index value (unbounded) are created and compared. Each way of evaluating data characterless yield different outcomes and results.

This chapter will end with a comprehensive cross-comparison across 4 indexes using ROC curves and preset thresholds to reflect on the overall performance difference across all the dataset, and lead to the start of the conclusion and policy suggestions chapter.

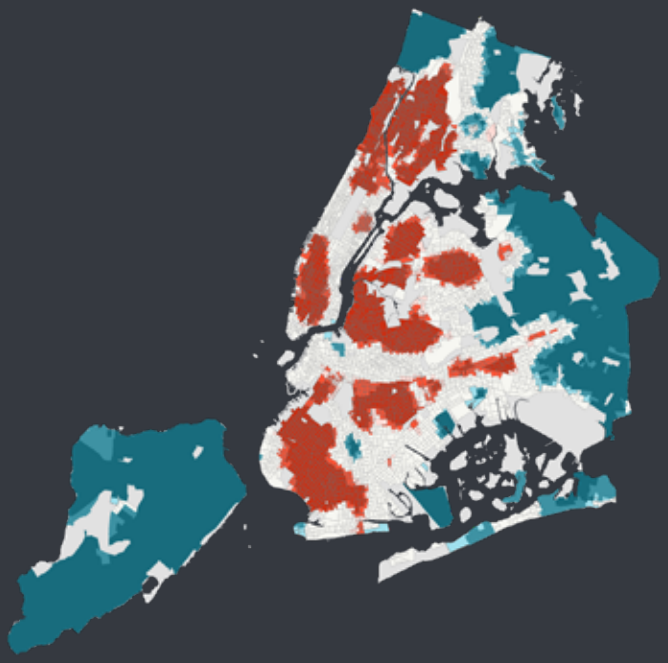
Hotspot of Concentration



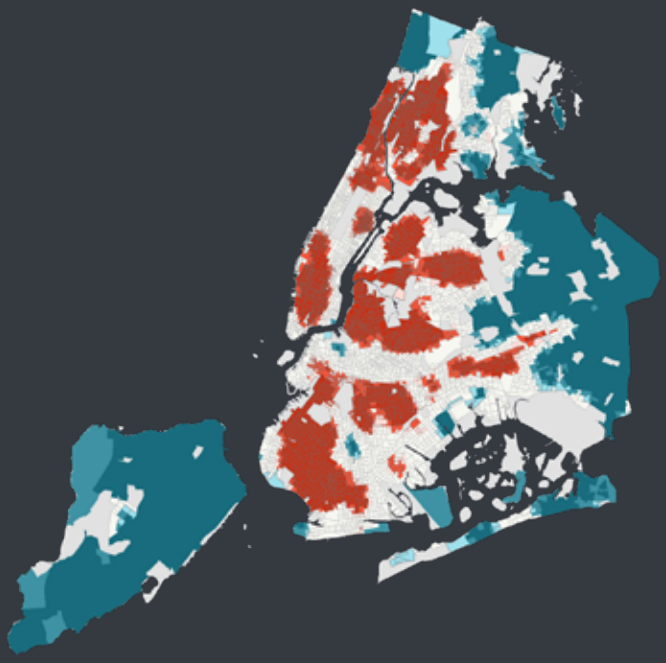
PCA - MORAN'S I



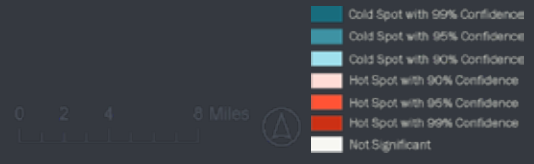
ISO - MORAN'S I



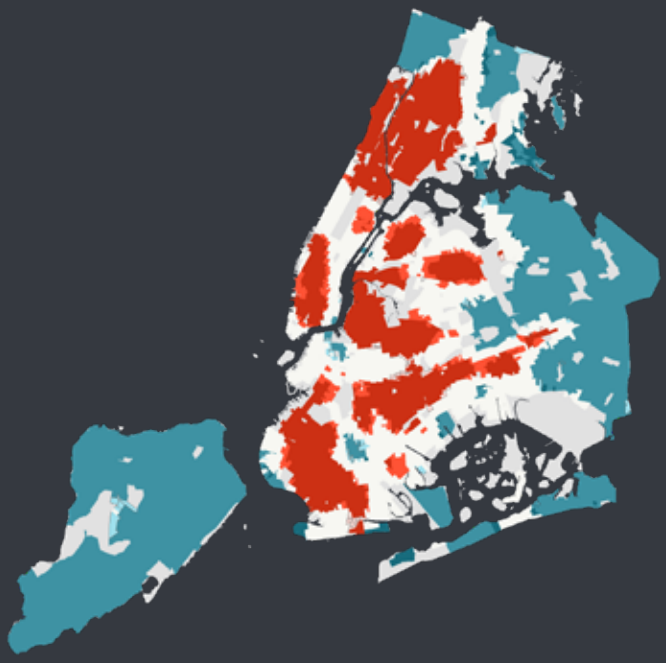
PCA - NATURAL BREAK



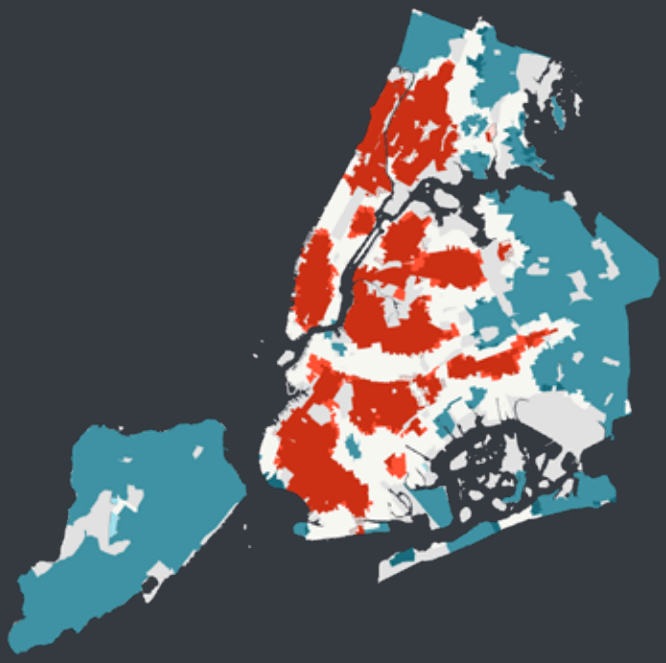
ISO - NATURAL BREAK



MORAN: PCA vs ISO



NB: PCA vs ISO



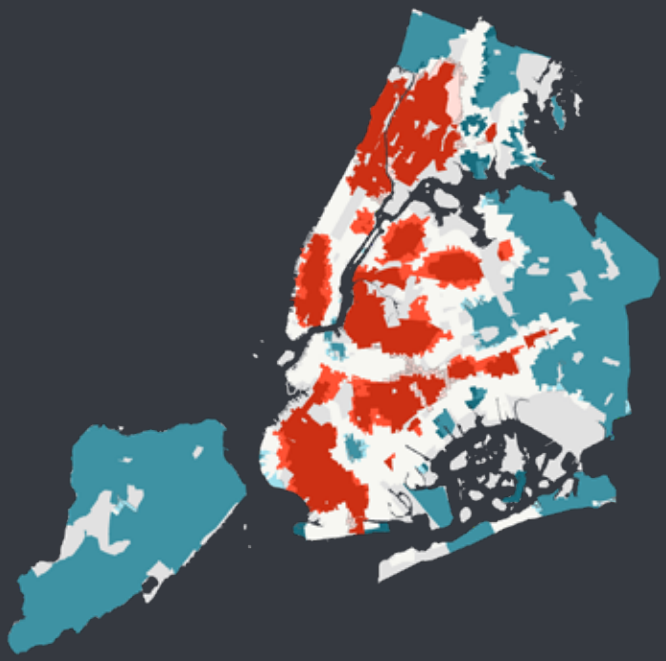
Hotspot Change Value

- Cold to Not Significant
- Cold to Cold
- Not Significant to Cold
- Hot to Cold
- Hot to Not Significant
- Cold to Hot
- Not Significant to Hot
- Hot to Hot
- Not Significant

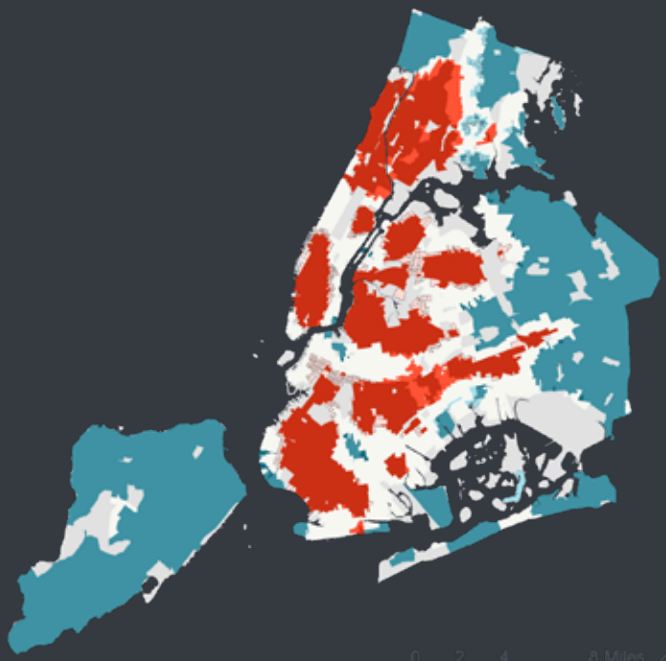
The Hot Spot Analysis tool calculates the Getis-Ord Gi* statistic for each individual score (by census block groups) in the model. It works by looking at each feature within the context of neighboring features. A census block group is considered spatially significant with a high value that is surrounded by other groups with high values as well. The FDR applied adjusted statistical significance to account for multiple testing and spatial dependency

By comparing two hot spot outcomes, the hot spot analysis comparison tool examines the similarity and association between the hot spot result layers with different significant level categories. Clearly knowing that models do not need to associate with each other, the research focuses on observing the potentially existing similarities. Using default weight, the research processed a visualized comparison between four sets of models, documenting

PCA: MORAN vs NB



ISO: MORAN vs NB

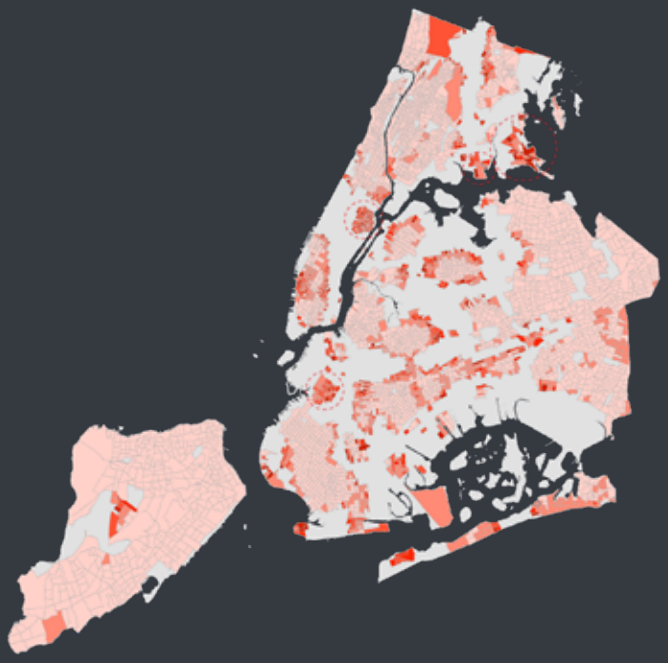


any changes from hot spots to cold spots, or vice versa.

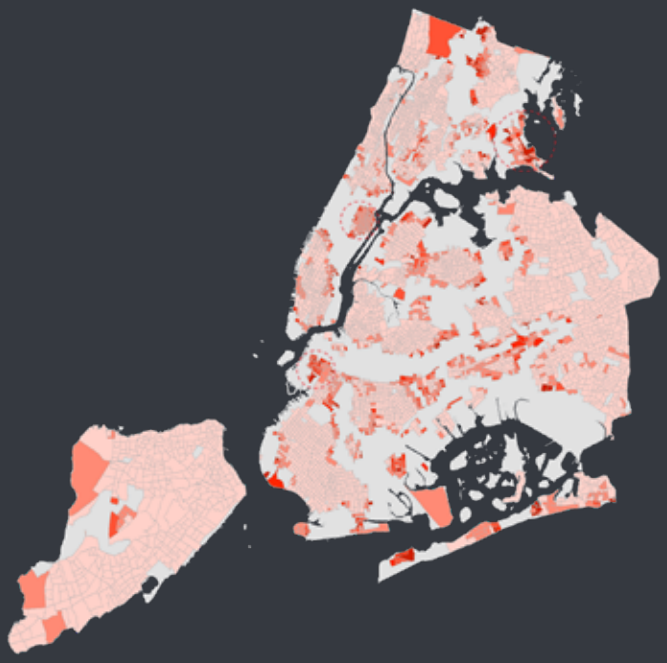
The direct visualization of HVI hotspots by Gi method reflects clearly the clustering of cold spots (HVI is relatively high) and hot spots (where heat vulnerability is relatively low). All of the four outcomes indicate relatively heat-resistant areas across most parts of the Staten Island, Eastern Queens (near the city boundary), and Western/ Northeastern Bronx. The area of Central and Southern Bronx, Southern Brooklyn, Eastern Brooklyn, Northern Manhattan (Harlem and Above), Lower-West, and Southern/Central Queens see a higher clustering of heat vulnerability. While the four outcomes largely resemble each other, minor differences, specifically across hot spot regions, are obvious: The outcome using natural break classification and ISO clustering indicates that the neighborhood between Long Island City and Central Queens has high vulnerability, while other models do not; Compared to the other three methods, Moran's I and PCA Anaysis outcome detects the lowest numbers of hot spot concentration. Most of these differences are caused by the detection focus upon clustering vs. absolute value, while some occur due to classification boundaries that are not continuous.

Comparing between the hot spots and cold spots using different methods, it can be found that the results are mostly kept the same, with some variances that occur mostly along the edge of hot spots, which is caused by the same two reasons above (as data classification boundary was strict, while clustering analysis often led to a wider higher value concentration). The difference between Anselin Local Moran's I and Natural Breaks are more significant compared to cross-raster methodological differences.

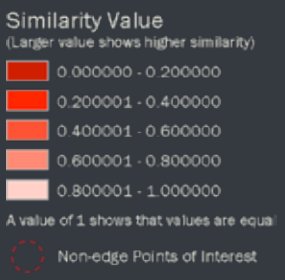
MORAN: PCA vs ISO



NB: PCA vs ISO



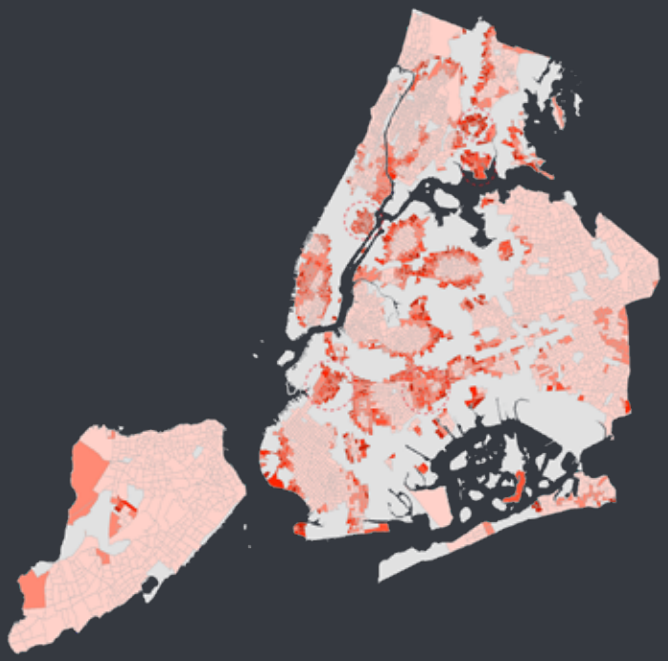
Similarity Value



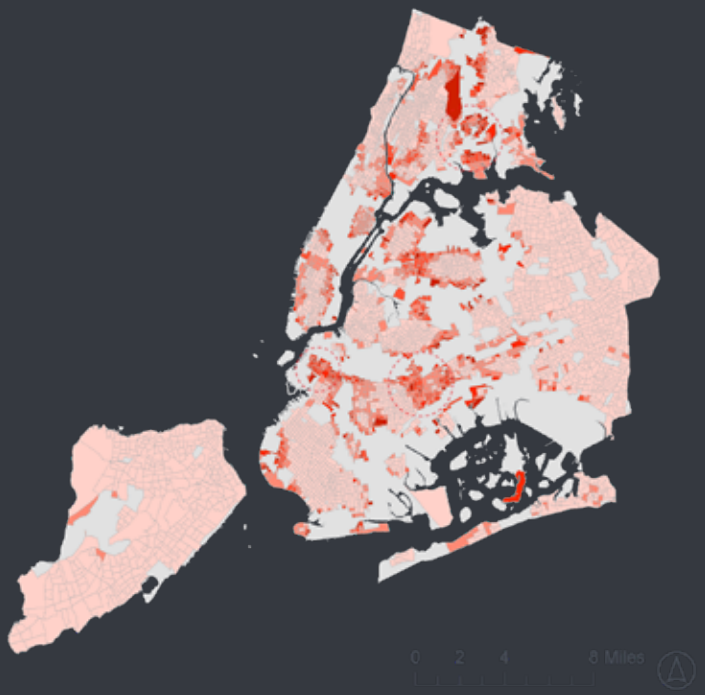
Similarity value comparison is another important factor in determining how two index models perform alike. Comparing the difference of four target sets, a score of 0 to 1 is given to all census block groups: If many corresponding features in both results have the same significance level, the value will be close to 1. If many corresponding features do not have matching significance levels, the value will be close to 0 (https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/hot-spot-comparison.htm). As it would be impossible to reach complete data match, a similarity value over 0.8 is considered partial match in data, and is defined as ‘similar’ in the research.

Following observtaions during the hotspot comparison stage, a large portion of census block groups that witness obvious difference in similarities are along the edges of hotspot/coldspot regions due to the methodological

PCA: MORAN vs NB



ISO: MORAN vs NB



concentration on clustering versus single values. The edge effect is strongest in the scenario where Anselin Moran's I is compared with Natural Breaks within the same PCA environment, showing that PCA (as a continuous rather than discrete option) outcome does not have a mitigation effect in constraining the large marginal difference between Moran's I versus Natural Breaks.

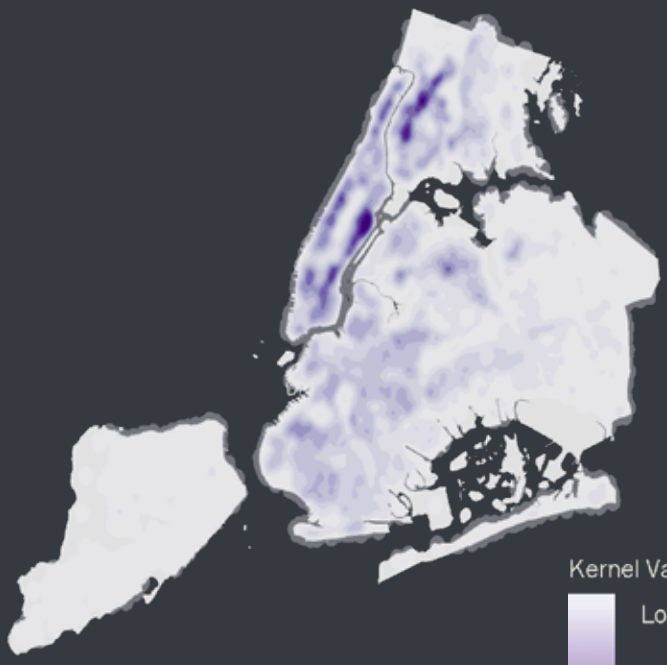
The comparison of HVI similarities also reflects clustering of census block groups with high value difference that are not along the edges of hotspot/coldspot groups (designated as 'non-edge points of interest' in the research). Despite a difference in hotspot/coldspot areas, four comparison maps reflect a similar range of non-edge points of interest, including eastern Bronx, coastal eastern Bronx, upper-east Manhattan, southern Downtown Brooklyn, and Eastern Brooklyn. This is potentially due to that census block groups within these places see more than one values (of socio-economic and built environment data) that are near the margin of different data classes, leading to a large difference in their final score outcomes (as they fall into different score categories with different model combinations). The reason can also explain why hotspot/coldspot regions near these highly contrasting places are different.

Kernel Density

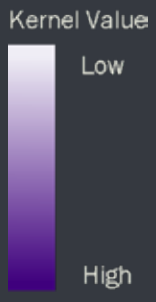
PCA - MORAN'S I

ISO - MORAN'S I

PCA - NATURAL BREAK

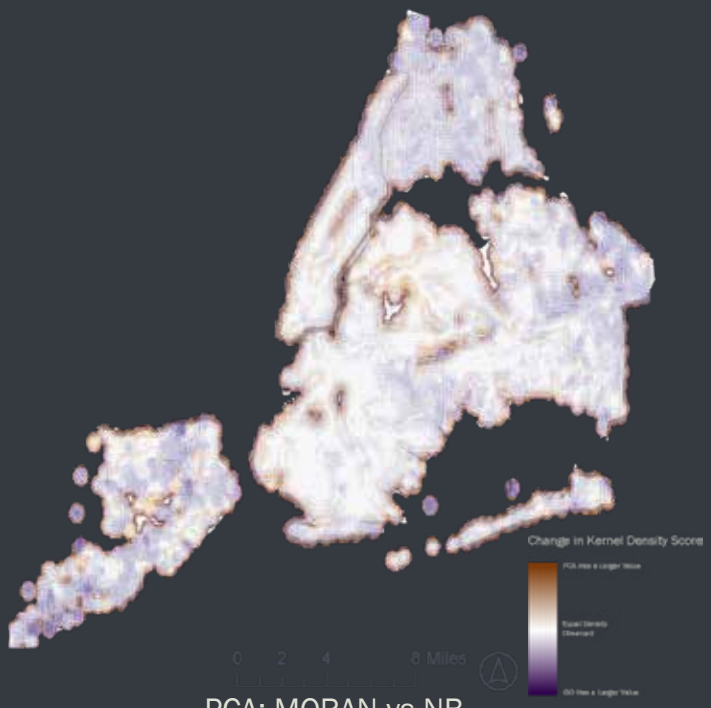


ISO - NATURAL BREAK



Density of Difference

MORAN: PCA vs ISO



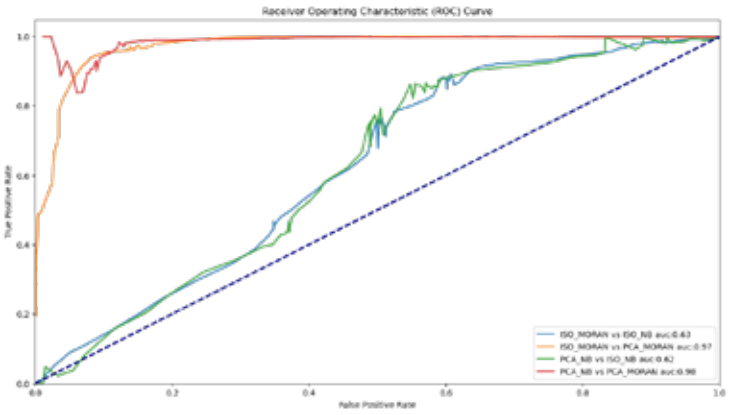
PCA: MORAN vs NB

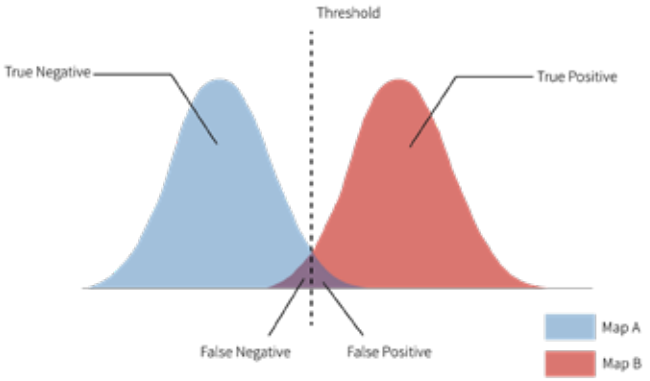
The Kernel Density tool is applied to calculate the density of heat vulnerability index in the city using values of individual census tract groups. The research applies default density, search radius and weight value during the automated calculation process. Natural environment differences, such as rivers and oceans in the case of New York City, act as barriers in the calculation process.

The analytical results from kernel density of the HVI scores show similarities among all four combinations, supporting the previous claim that overall HVI performance in the research is close and accurate. The kernel density process further demonstrates that heat-vulnerable regions are located in central Bronx, upper Manhattan, lower-East Manhattan, and part of Central Queens, echoing with visual outcomes based upon census block groups. A smaller difference in Kernel density, compared to vector similarity evaluation and hotspot change detection, reflects that HVI indices proposed in the research share an overall similar trend but differs in extreme values.

Due to limited data range and duplicated data performance, the report only shows two kernel density comparisons showing difference between PCA and ISO Clustering (Within Moran's I) and between Moran's I and Natural Breaks (Within PCA). The result shows that while general trend in kernel density is similar, the detailed value still remains different. While comparing absolute value is not advisable and useful in kernel raster analysis, the two results show that there is a larger difference between PCA and ISO method (in determining raster) compared to vector method differences.

Cross-Comparison Analysis ROC Curve





To provide a more comprehensive analysis that can cross-compare four outcomes, ROC (Receiver Operating Characteristic Curve) technique is applied to compare the value of HVI by each census block groups. In this research, a desirable ROC curve is tilted towards the top-left corner in the coordinate, showing the capability of high true positive rate with low false positive incidences, which indicates that the data for comparison is highly similar.

The Area under the ROC Curve (AUC), which is an indicator showing data similarity, is 0.63 for the comparison between Natural Break vs. Moran's I under the same ISO Clustering method, and 0.62 for the comparison between PCA vs. ISO under the same Natural Break Reclassification method, which is slightly higher than random data grouping (around 0.5 in the curve), showing that the data similarity of HVI is relatively low for these two sets of data.

By contrast, a high similarity can be observed between the PCA-Moran combination method and the ISO-Moran method (AUC 0.97) as well as the PCA-NB method (AUC 0.98), showing a relatively higher data combability for the PCA-Moran model. While the ROC curve can provide a clearer comparison between datasets, in this research it does not indicate the effectiveness or accuracy of HVI evaluation model.

LIMITATIONS

Although the research calculates and visualizes multiple sets of Heat Vulnerability Index in the New York City with relatively high precision and accuracy, several limitations exist in the research method and outcome, which require further studies to be conducted for a more comprehensive analysis. Limitations in the study include:

1. The natural/built environment and demographic data to be included in the model weighting process, along with the weighting distribution, are selected and chosen based upon established models and previous research, which may be biased and incomprehensive, and may yield HVI scores that do not fully reflect heat risk and vulnerability.
2. Neither the PCA nor the ISO Clustering method used for model construction has gone through autocorrelation analysis and covariate elimination process (which is usually conducted with machine learning techniques) due to time and scale restrictions, which may add inaccuracy in the unsupervised categorization process, negatively impacting the final score outcome.
3. The HVI index, regardless of methods used, is only a rough estimate of the heat vulnerability for an average demographic profile in certain region. It does not represent how vulnerable individuals (with different characteristics) are in face of extreme heat activities. Therefore, policies and interventions shall not be made solely based upon the HVI index.
4. Only one weighting structure is applied (for each related risk factors) during the indexing stage due to time and scale limitations. The limited weighting diagram may increase the bias in the model with fewer options for viewer/ policymakers to choose from in the final stage. Future research should include multiple sets of weighting diagrams or correlation scores to construct a more comprehensive index.

CONCLUSION

Extreme heat events and extended heat waves in the urban area has become a major climate challenge in the New York City, which is exacerbated by climate change, global temperature rise, and urban island effect. The research constructs a model to reflect community-level heat vulnerability in the city. To compare between different spatial analytical methods, the research applies Principle Component Analysis (PCA), ISO Clustering, Natural Break Reclassification, and Anselin Local Moran's I and provides four sets of HVI outcomes with different combinations. All four models yield similar outcomes showing that communities in Central Bronx, Lower-West Manhattan, Southern Brooklyn and Central Queens are highly vulnerable to heat extremes but are different in detailed scores and clustering performance. Difference across the models is caused by different ways of categorization, data boundary differences, and the difference in clustering-detection methods versus absolute-value-based methods. While each model has its focus and cannot be compared through its effectiveness or accuracy, the HVI index constructed upon PCA and Local Moran's I show a relatively normal data distribution, well-allocated data values, and the capability to track regional clustering. Therefore, the PCA-Moran model is recommended for most analytical scenarios, as results of clustering can reflect real-world situation more accurately. By contrast, the ISO-Moran model can better reflect differences in areas with lower heat vulnerability, which proves useful for more detailed analysis in city boroughs such as Queens or Staten Island.

The research is significant in mitigating heat impacts in the city as it provides evaluation and estimation of HVIs with a greater level of precision (in census block groups) compared to the original official map used by the city (in census blocks). In addition, the models point out multiple communities with very high vulnerability with precision, which enables specific intervention and emergency responses during sudden heat events. A comparison between different spatial analytical methodologies also shows a variety of data performances and model focuses, providing insights for future research in the area. More importantly, the methods of calculating HVI index and evaluating heat vulnerability can be applied globally and be easily adopted in other cities with similar heat-related issues.

It is also recommended that users of the HVI model should compared among all potential method combinations and their outcomes between making observations and decisions. Considering conclusions and insights above, it is suggested that policymakers and urban planners in the NYC should prioritize their resources to mitigate heat in Central Bronx, Lower-West Manhattan, Southern Brooklyn, and Central Queens, while the coverage of heat-mitigation strategies should be enlarged (if funding permits) to the scales which all HVI index models agree on to eliminate potential biases and errors. A proper emergency response management system, along with intervention strategies such as provision of air-conditioning, issue of heat warnings, construction of heat shelter (or use of existing public spaces), and facilitation of mutual-help in the community, can better help the city mitigate heat impacts and reduce heat vulnerability in the future.

REFERENCE

Chen, T.-L., Lin, H., & Chiu, Y.-H. (2022). Heat vulnerability and extreme heat risk at the metropolitan scale: A case study of Taipei metropolitan area, Taiwan. Urban Climate, 41, 101054. https://doi.org/10.1016/j.uclim.2021.101054

Coffel, E. (2018). *Extreme heat and its impacts in a changing climate*. https://doi.org/10.7916/D88358JX

de Smith, Goodchild, Longley (2021). *Univariate classification schemes* in Geospatial Analysis—A Comprehensive Guide, 6th edition;

Eugenio Pappalardo, S., Zanetti, C., & Todeschi, V. (2023). Mapping urban heat islands and heat-related risk during heat waves from a climate justice perspective: A case study in the municipality of Padua (Italy) for inclusive adaptation policies. *Landscape and Urban Planning*, 238, 104831. https://doi.org/10.1016/j.landurbplan.2023.104831

Guardaro, M., Hondula, D. M., Ortiz, J., & Redman, C. L. (2022). Adaptive capacity to extreme urban heat: The dynamics of differing narratives. *Climate Risk Management*, 35, 100415. https://doi.org/10.1016/j.crm.2022.100415

Johnson, D. P., Stanforth, A., Lulla, V., & Luber, G. (2012). Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Applied Geography*, 35(1), 23–31. https://doi.org/10.1016/j.apgeog.2012.04.006

Kim, Y., Li, D., Xu, Y., Zhang, Y., Li, X., Muhlenforth, L., Xue, S., & Brown, R. (2023). Heat vulnerability and street-level outdoor thermal comfort in the city of Houston: Application of google street view image derived SVFs. *Urban Climate*, 51, 101617. https://doi.org/10.1016/j.uclim.2023.101617

Kuras, E. R., Richardson, M. B., Calkins, M. M., Ebi, K. L., Hess, J. J., Kintziger, K. W., Jagger, M. A., Middel, A., Scott, A. A., Spector, J. T., Uejio, C. K., Vanos, J. K., Zaitchik, B. F., Gohlke, J. M., & Hondula, D. M. (2017). Opportunities and Challenges for Personal Heat Exposure Research. *Environmental Health Perspectives*, 125(8), 085001. https://doi.org/10.1289/EHP556

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*, 161, 127–137. https://doi.org/10.1016/j.puhe.2017.09.006

North Kentucky University NKU (1999). “Local Moran’s I”. https://www.nku.edu/~longa/geomed/ppa/doc/LocalI/LocalI.htm

Ortiz, L., Mustafa, A., Cantis, P. H., & McPhearson, T. (2022). Overlapping heat and COVID-19 risk in New York City. *Urban Climate*, 41, 101081. https://doi.org/10.1016/j.uclim.2021.101081

Wolf, T., & McGregor, G. (2013). The development of a heat wave vulnerability index for London, United Kingdom. *Weather and Climate Extremes*, 1, 59–68. https://doi.org/10.1016/j.wace.2013.07.004

Yoo, C., Im, J., Weng, Q., Cho, D., Kang, E., & Shin, Y. (2023). Diurnal urban heat risk assessment using extreme air temperatures and real-time population data in Seoul. *iScience*, 26(11), 108123. https://doi.org/10.1016/j.isci.2023.108123

Acknowledgement

We would like to express our sincere gratitude to **Professor Leah Meisterlin** for her invaluable guidance and support throughout this project. Her expertise and insights greatly contributed to its success.

We also extend our heartfelt thanks to **Ruju Joshi**, our teaching assistant, for their assistance and encouragement at every stage of the project.

We would like to express our sincere gratitude to the following individuals for their invaluable contributions to this project:

Eric Xia & Hui Chen:
INTRODUCTION
DATA CLEANING
METHODOLOGY
ANALYSIS & COMPARISON
LIMITATIONS & CONCLUSION
MAP DESIGN & LAYOUT

Mina Wei:
Cross-Comparison Analysis ROC Curve