

Analysis of the 3D shape of cities with respect to Urban Heat Island effect

Authors

Pradnya Raut ^{1*}

Cristian Román-Palacios ^{2*}

Affiliations

^{1*}MS Student at Information Science,

School of Information, University of Arizona, Tucson, Arizona

^{2*} Assistant Professor, School of Information, University of Arizona, Tucson, Arizona.

Abstract

The aim of this study was to explore how building shapes interact with urban policies and regulations. We plan to develop urban planning strategies aimed at mitigating Urban Heat Island effect as well as the impacts of climate change to urban areas. We explore the relationship between variables like urban morphology, building heights, vegetation, green distribution, shadow and the UHI effect/ climate change (H M Abdul Fattah, 2024, p2)

Research has shown that urban areas tend to have higher temperatures compared to the surrounding rural areas. This effect is known as the urban heat island (UHI) [1]. The temperature increase within urban areas is known to lead to various problems including human health issues, increase in energy consumption thereby increasing greenhouse gas emissions, as well as further contributions to climate change. To reduce these effects there is need to identify factors related to UHI and take appropriate actions to mitigate its effect in urban areas.

Work done by Mr. H M Abdul Fattah on this project [8] consisted of downloading city shape file, building height dataset, calculating city wise area vs building height distribution and classifying cities into one of the four categories. Dataset about the heights of buildings globally (GHS-BUILT-H - R2023A) was obtained from GHSL - Global Human Settlement Layer dataset [13]. Additionally, shapefiles, defining urban areas, were obtained from World Urban Areas, LandScan, 1:10 million (2012) [14]. Data of heights of buildings was filtered using shape files and analysed to find the distribution of building heights in the city. Based on the shapes of distribution (their skewness and dip statistic values), the shape of city was classified as unimodal right skew (Pyramid) shape, unimodal left skew (Inverse Pyramid) shape, or has no significant skew (Diamond) shape [15] [8]. In case of presence of multimodality, indicated by the dip statistic, shape was classified as "Hourglass." The "Pyramid" class signifies a concentration of shorter buildings, while "Inverse Pyramid" indicates a concentration of taller buildings. "Diamond" suggests a balanced distribution of building heights, while "Hourglass" implies varying patterns in building height distribution.

Work that needs to be done is getting UHI effects data for the year 2018 [9] and other variable data like green vegetation [10], solar radiation [11], surface water [12]. Since the latest dataset available for building height is of year 2018, the rest of the datasets will also be downloaded for the same year 2018 wherever possible. It was possible to download full global UHI data with resolution of 3000m/pixel. The resolution of building height dataset is around 30m/pixel whereas most of the datasets in googleearth engine are around 300m/pixel. Attempts can be made to improve resolution as close as 100 m or as close as possible. This downloading can be done either using Javascript scripts in google earth engine or using geemap api in Python

Introduction

Urban Heat Island (UHI) leads to increased vulnerability to human health issues like heat strokes, exhaustion, suicidal tendencies. It is also impacting air quality due to more amount of pollutants released in air and poor scattering of these pollutants. Water quality also gets impacted due to increase in water temperature affecting native aquatic life.[3]

Urban areas are densely populated with more people. Closely constructed building and skyscrapers mean a lot of waste energy is emitted and can not escape the area. Increase in temperature also causes increase in energy consumption thereby increasing greenhouse emissions for the city dwellers. [2]. According to David L. Chandler, Urban heat island effects also depend on a city's street and building layout. Some cities, such as New York and Chicago, are laid out on a precise grid, like the atoms in a crystal, while others such as Boston or London are arranged more chaotically, like the disordered atoms in a liquid or glass. The researchers found that the "crystalline" cities had a far greater buildup of heat compared to their surroundings than did the "glass-like" ones.[4] It is found that UHI is positively correlated with city area. Building materials which absorb and radiate heat back into the air gets trapped in the nearby vicinity in the area densely crowded with buildings instead of spreading out evenly[5]. Hence effect of heights of skyscrapers also needs to be studied apart from the area of city. This can help in planning for urban area expansions or in new urban area developments

According to Nyuk Hien Wong, Chun Liang Tan, Dionysia Denia Kolokotsa & Hideki Takebayashi [6], Green infrastructure acts to cool the urban environment through shade provision and evapotranspiration. Typically, greenery on the ground reduces peak surface temperature by 2–9 °C, while green roofs and green walls reduce surface temperature by ~17 °C, also providing added thermal insulation for the building envelope. However, the cooling potential varies markedly, depending on the scale of interest (city or building level), greenery extent (park shape and size), plant selection and plant placement . This can be a tool for mitigating UHI

Climate change is impacting cities and their residents in many ways, from poor air quality to flooding, biodiversity loss and extreme heat. Mackres et al. [7] with the help of a dashboard provides insight into connection between climate change and urban life which will be useful for city designing in a more sustainable and nature-positive ways to mitigate climate change

Objective – To determine the impact of building heights, city area expressed as city 3D shape on urban heat island effect observed in cities along with the impact of other variables like vegetation, water surfaces in the proximities of various cities spread across in the world

Scope There are various factors related to urbanization like building material building heights, industrial areas, city areas, NDVI, surface water which are responsible for urban heat island effect. The factors currently considered for this study are City shape which is based on building heights distribution, NDVI, surface water of around 6000 cities across the world

List of cities excluded in this study are ['Bathurst1', 'Bathurst2', 'DamanhÈr', 'Düsseldorf', 'JaĀn', 'Lüderitz', 'Münster', 'Cadiz1', 'Cadiz2', 'Newcastle', 'Ciudad Juárez', 'Osnabrück', 'PirailĀs', 'Saint GeorgeĀs', 'San Sebastián', 'St. JohnĀs', 'YpacaraiIta']

Limitations

Due to cloud cover many of the observations showed up as NA.

The resolution of UHI, NDVI, surface water was not as high as that of building data.

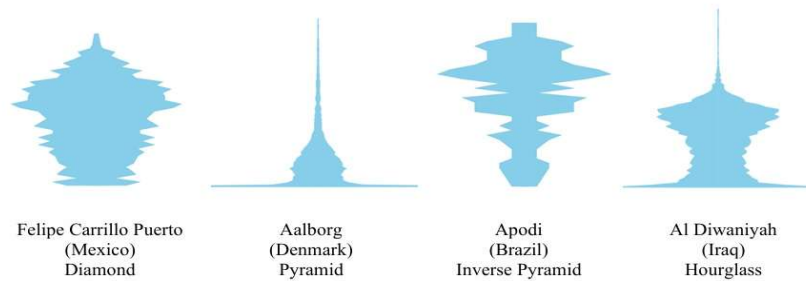
The latest data available for buildings was that of 2018.

Material and Method

Raw Dataset Description:

Building Height data – For determining city shape, dataset was downloaded from the website [13]. This data was available for year 2018 so all the rest of the variable data was downloaded for year 2018 whereas Shape file data was downloaded from World Urban Areas website [14]

Classification Results



Hypsographic curve for different cities from various classification level

Figure 1 City shapes - no significant skew (Diamond) shape, unimodal right skew (Pyramid) shape, unimodal left skew (Inverse Pyramid) shape, Varying pattern(Hourglass)

Urban Heat Island effect data - Urban Heat Island effect(UHI) dataset is available at the website [9] as geoTiff file. There are several different types of datasets of UHI available like summer, winter, annual, averaged for multiple year on the website. The dataset considered for this study is annual dataset for the year 2018. The resolution of file is 300m. This file has 2 bands 'Daytime' and 'Nighttime' for annual daytime average UHI value and annual nighttime average UHI value. Using this data, 3 mean values were derived per city viz. UHI daytime mean, UHI nighttime mean, UHI daynight time UHI mean which averages over 24-hour period of daytime and night time. The unit of measurement used for UHI is (Degree C) .

Out of 6000 records of cities, 3837 cities have non zero UHI value.

The highest value of both day and night UHI out of these cities was found to be 6.77057165329402 (° C) for city ' Pontianak' whereas lowest value for same was observed for city Aden with value -6.4956758212178 (° C)

The highest value of day UHI was found to be 5.83911699001886 (° C) for city ' Pontianak' whereas lowest value for same was observed for city Aden with value -6.77188387224751 (° C)

The highest value of night UHI was found to be 3.75197793743963 (° C) for city ' Tehran' whereas lowest value for same was observed for city Nagaoka with value -2.5011636231867 (° C)

Raster map of city Aalborg for City height distribution and corresponding Urban Heat Island effect, NDVI, surface water (SRC caption to be added)

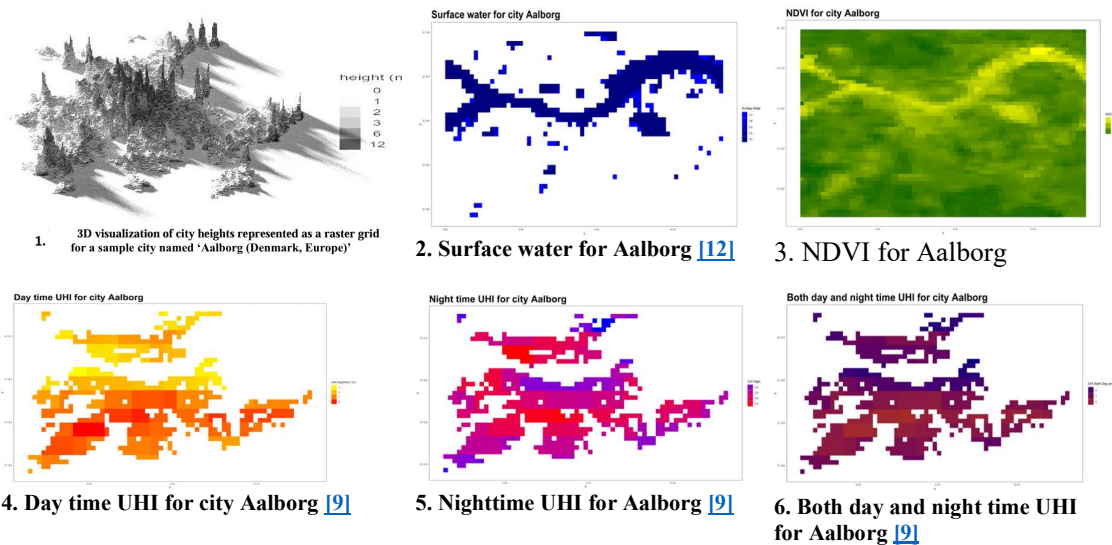


Figure 2 Raster maps of city of Aalborg displaying building heights(1), surface water(2), NDVI(3), Daytime UHI(4), Nighttime UHI(5), Both Day and Nighttime UHI

For NDVI (Normalized Difference Vegetation Index) and surface water data was downloaded from google earth engine.

NDVI data-

NDVI files are available on the website[10] as 16 day files . So 24 files -two per month were averaged for year 2018 and then downloaded from google earth engine. The download is in the form of 21 .tif tiles. There is only one band 'NDVI' in the file. Resolution is 30 m

Out of 6000 cities, 5731 records have non zero NDVI mean values

The highest value of NDVI was found to be 8383.38 for city Ikela whereas lowest non zero value for same was observed for city Mejillones with value 136.73

Surface water data –

Surface water dataset was found at the google earth engine website [12] and downloaded as geoTiff file for the year 2018. The file is available as annual file containing only one band 'waterClass'. Resolution is 30 m originally but data was downloaded at 300m resolution to enable download without errors . Since the file contains a larger geometry, the file was downloaded as 6 tiles each of with latitudes ranging from -85 degrees to +85 degrees and longitudes of size 60 degrees each with resolution of 300m

For Surface water, values were averaged pixelwise for the city area. 5487 cities out of 6000 have non zero mean value. Surface water measures the seasonality of water with value

- 1 indicating no water,
- 2 indicating seasonal water and
- 3 indicating water permanent water throughout the year

The highest value of surface water of 3 was observed for various cities like Brochet, Port Hope Simpson, Jaque etc indicating permanent water whereas lowest value of 1 for same was

observed for cities like Schefferville, Apodi, Lethem etc indicating 'no water'. Value of 2 indicates seasonal water.

Data processing Method- For UHI data processing, using `Terra::crop()` in R program, the .geoTiff data file is opened ,cropped and saved as individual city raster .tif files in rasters folder. Also the raster is converted to data frame and used to calculate mean values per city viz. daytime, nighttime , both day and night time UHI mean values . The program was run in batches of 600 cities at a time and all the mean values are stored in Citywise Means[0-10].csv files These are then later merged with NDVI, water surface, geographical coordinates data files in a single ML_input_data_coordinates.csv file using a python program

For NDVI and surface water data was processed similarly for creating rasters. Using `Terra::crop()`, the .geoTiff files are opened ,cropped and saved as individual raster .tif files in rasters folder. Also, the raster is converted to data frame and used to calculate mean values per city in R program. Since the data for both NDVI and surface water is spread in multiple .tif files, for finding the tile/tiles containing city data, CSV files containing list of cities and their corresponding tile is accessed. For NDVI data, the csv file is 'City_NDVI_files.csv' where as for surface water data it is 'City_surface.csv' (These CSVs were created using python program by comparing shape file with individual tiles in case of NDVI data or the the coordinate locations in case of surface water data)

All NDVI and surface water mean values are stored in Citywise Means NDVI[0-10].csv , Citywise Means NDVI Surface water [0-10].csv file respectively, and then later merged with UHI data files in ML_input_data_coordinates.csv file, ML_input_data.csv using python program.

Geographical coordinate information was extracted from shape file containing city data and this information was merged in ML_input_data_coordinates.csv, ML_input_data.csv file using python program

Statistical Data Analysis

Distribution of city shapes across world

World map is plotted using `ggplot`, `geom_sf` (sf package) and distribution of around 6000 cities under study and their shapes is examined.

Majority of cities are either pyramid (identified by yellow triangle) or hour glass (identified by red hourglass). Location of cities under study is represented as blue dots and are spread all over the world.

Diamond shaped cities identified by pink diamond shapes

Cities of Inverted pyramid shape are found more closer to equator region

City Shapes

Based on building height distribution

shape_color [Hourglass] [Pyramid] [Inverse Pyramid] [Diamond] [Unknown]

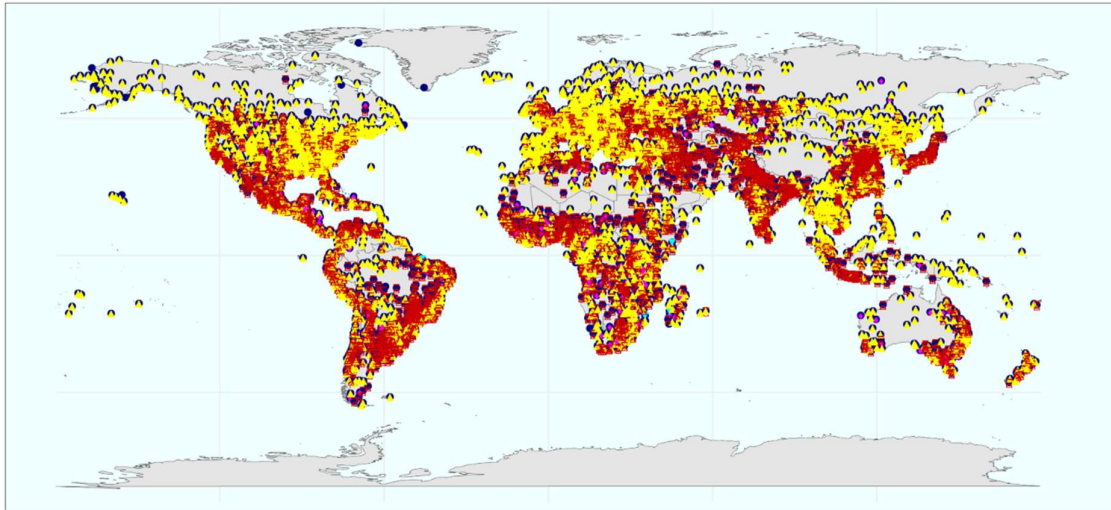


Figure 3 City shape distribution on world map

In 4 categories of city shapes, value comparison from box plot

Boxplots are plotted using ggplot and facetwrap with ML_input_data_coordinates.csv to examine value distribution of various variables in an R program

The below figure shows distribution of various variable values viz. UHI, NDVI, Surface water across 3 different city shape categories (City shape categories 'NA' and 'Unknown' were dropped and 'Inverse pyramid' cities did not had NA UHI values so those records were deleted)

UHI values are highest in hourglass city shapes followed by Pyramid shapes. Diamond city shapes have the lowest UHI values but the outliers in pyramid shapes are lowest for daytime and both day and night whereas hourglass has lowest outlier values for nighttime. Drop in nighttime UHI values is significant in Pyramid shape cities.

NDVI values are highest in pyramid shapes whereas Diamond shape has lowest median NDVI value among 3 shapes

Surface water median values are highest in pyramid shapes whereas Diamond shape has lowest median surface water value among 3 shapes. Plots below in figure 4,5,6 show various city mean values and log of mean values distributed across 3 shapes viz. Diamond, hourglass and pyramid.

Boxplot of UHI, NDVI, Surface water mean values

Based on city shapes

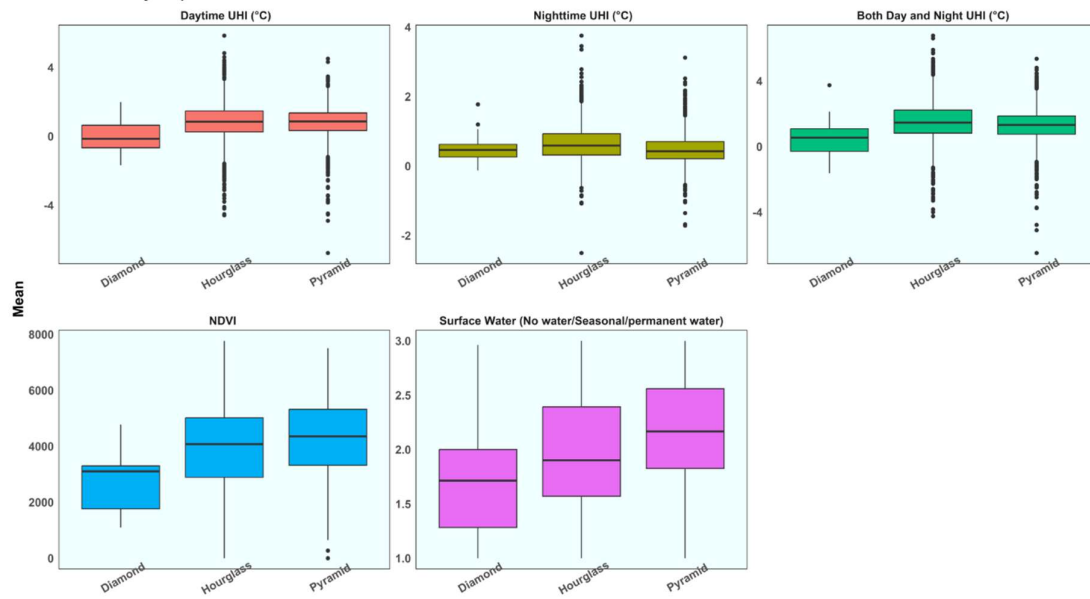


Figure 4 Boxplots UHI, NDVI, surface water across city shapes

Log transformed values

Boxplot of UHI, NDVI, Surface water mean log values

Based on city shapes

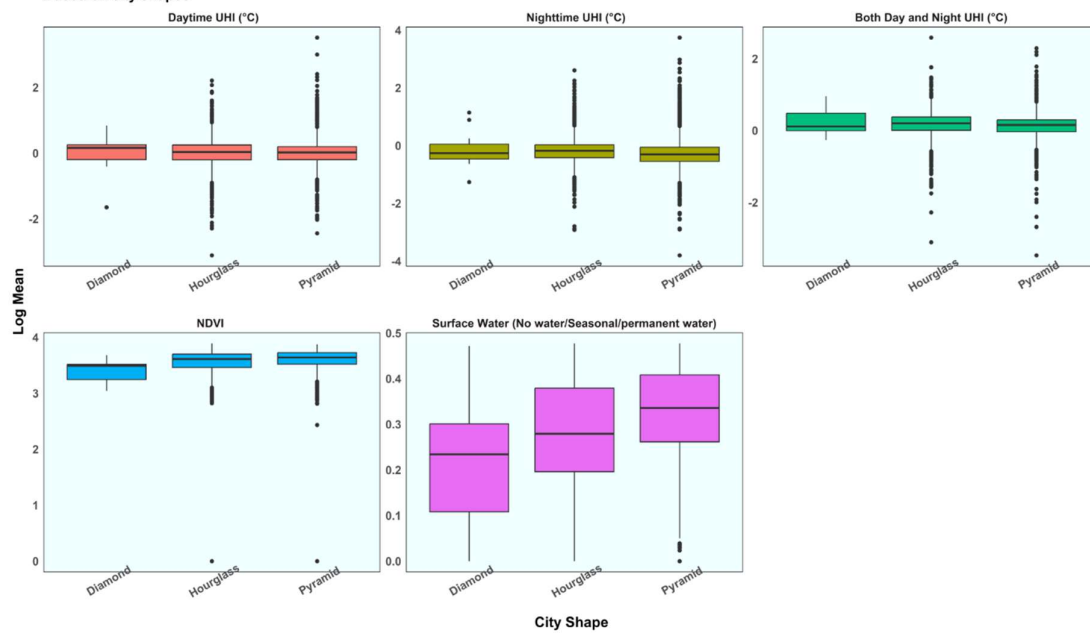


Figure 5 Boxplot log values of UHI, NDVI, surface water across city shapes

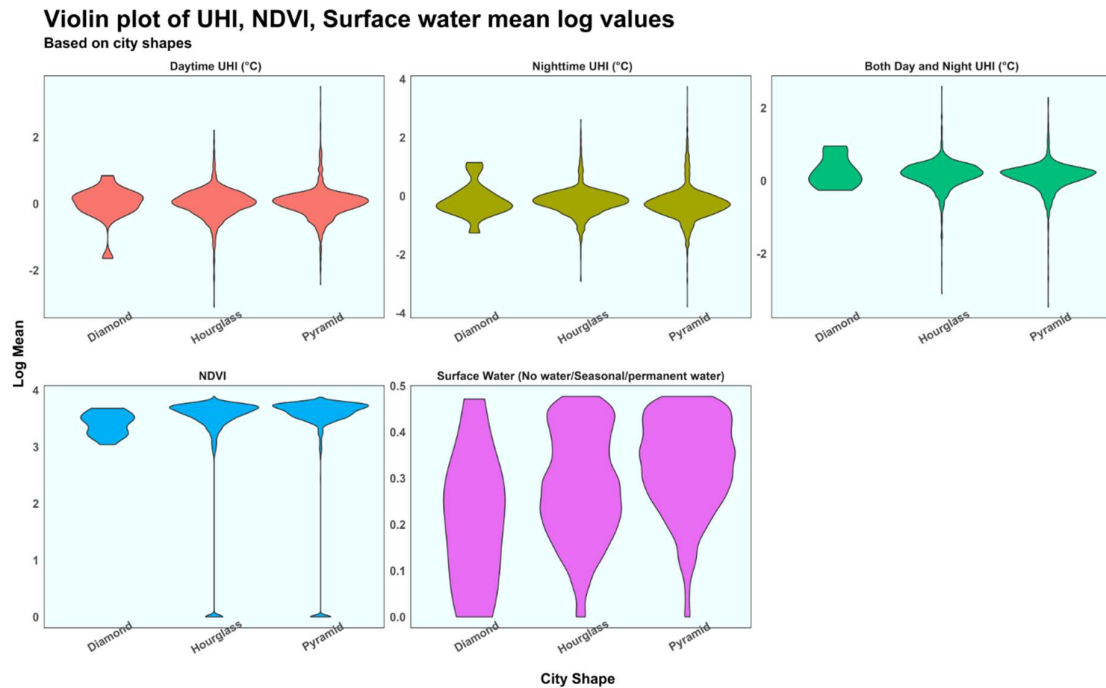


Figure 6 Violin plots for Log transformed values of UHI, NDVI, surface water across city shapes

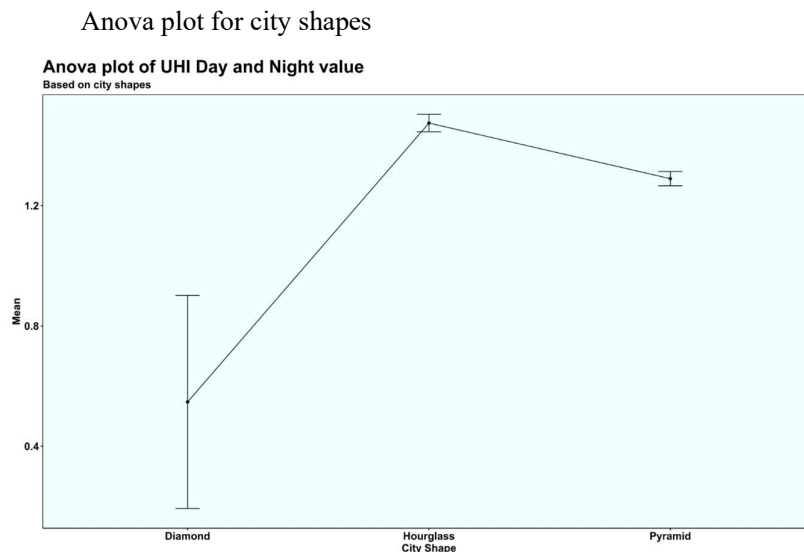


Figure 7 Anova plot of City shape and UHI both day and night mean values

Anova plot of UHI both day and night mean values against City shape provides an insight into distribution of UHI day and night mean values among three major city shapes. UHI values are spread across different ranges suggesting correlation among city shapes and UHI values

city shapes are converted to numerical values and pairplot of all variables using `ggpairs()` is plotted below which suggest correlation among variables like UHI-NDVI

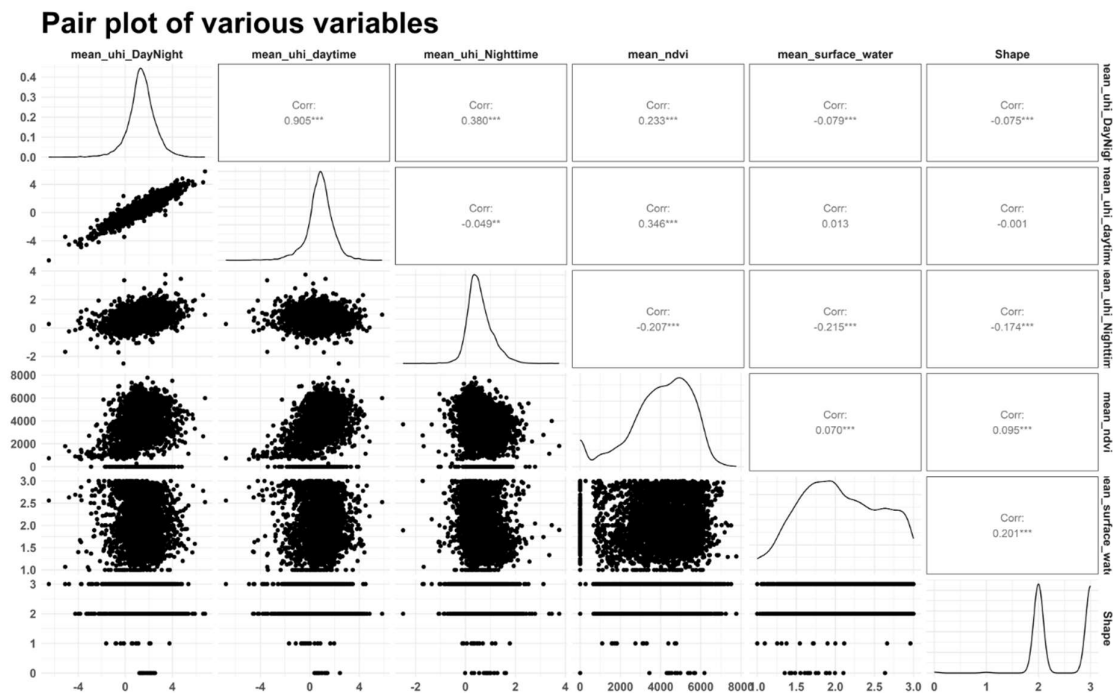


Figure 3 Correlogram displaying relationship among variables

Programs used and corresponding outputs generated by them -

There is a python program written using ee, geemap API to download surface water data from Google earth engine as 6 tiles.

NDVI and surface data is in the form of multiple tiles. So, a python program is written to determine the list of cities included in each tile. And in R program only the tile containing city is opened at a time in order to reduce memory usage.

Then R program reads all the data files and crops the data based on the shape file, calculates means of cities and merges NDVI, UHI, surface means into a single csv

Python program was written to merge City shape csv created previously with this csv containing various means Also city coordinates information was added to this csv. A single file containing all means along with coordinates is created

R program was used for analysis using Boxplots, Anova and Kruskal-Wallis test Also logistic regression ML model was used in a python program to explore relationship between UHI and city shape, NDVI, surface water

ML model and correlation discussion

Correlations suggested in anova plot become more clear with one way Anova test shown below in Fig.9. P-value obtained from Anova test by comparing mean UHI day and night mean value with city shape is lesser than 0.05. This indicates UHI day and night mean value is correlated to city shape.

One Way Anova test

```
res.aov <- aov(Mean ~ Shape, data = msel)
# Summary of the analysis
summary(res.aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Shape          2      42  21.157    15.82 1.44e-07 ***
## Residuals    3745    5008    1.337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig 9 One way Anova test between city shape and UHI both day and night mean values

Kruskal Wallis test- Kruskal Wallis test further provides comparison among the city shapes as shown below. The pairwise comparison shows that, only Hourglass and Pyramid are significantly different ($p < 0.05$).

```
kruskal.test(Mean ~ Shape, data = msel)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Mean by Shape
## Kruskal-Wallis chi-squared = 35.428, df = 2, p-value = 2.027e-08
```

```
pairwise.wilcox.test(msel$Mean, msel$Shape,
  p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  msel$Mean and msel$Shape
##
##              Diamond Hourglass
## Hourglass 0.0047 -
## Pyramid 0.0060 5e-07
##
## P value adjustment method: BH
```

Fig 10. Kruskal Wallis test and wilcox test showing comparison among city shapes

Machine learning models-

Logistic regression model is fitted on variables Mean day and night UHI against city shape, surface water, NDVI, mean height, dip statistic, skewness in a python program . Another similar model is fitted between city shape and Mean day and night UHI. On fitting logistic regression machine learning model, the coefficient summary of two models is listed below. p-value of NDVI is 0.005 where as p-value of mean height is 0.03 which is lesser than 0.05. First table column indicates a

correlation between mean NDVI, mean height and UHI where as in second column , p-value of ‘both day and night UHI mean’ is lesser than 0.05 indicating correlation among two.

Target variable - mean_uhi_DayNight
Other variables - mean_ndvi + Class +mean_surface_water
+Mean_Height+skewness+dip statistic
(After deleting all NA values)
Logit Regression Results

Dep. Variable: mean_uhi_DayNight No. Observations: 3748

Model: Logit Df Residuals: 3741

Method: MLE Df Model: 6

Date: Tue, 03 Dec 2024 Pseudo R-squ.: -1.009

Time: 03:10:53 Log-Likelihood: -2456.9

converged: True LL-Null: -1222.9

Covariance Type: nonrobust LLR p-value: 1.000

coef std err z P>|z| [0.025 0.975]

Intercept -0.0566 0.200 -0.282 0.778 -0.449 0.336

mean_ndvi 0.5181 0.183 2.833 0.005 0.160 0.877

Class -0.0265 0.408 -0.065 0.948 -0.826 0.773

mean_surface_water -0.1737 0.138 -1.258 0.208 -0.444 0.097

Mean_Height 0.8087 0.378 2.138 0.033 0.067 1.550

Skewness 1.1729 0.844 1.390 0.165 -0.481 2.827

Dip_Statistic -0.0315 0.121 -0.261 0.794 -0.268 0.205

Target variable - Class
Other variables - mean_uhi_DayNight
Optimization terminated successfully.
Current function value: 0.501699
Iterations 5
Logit Regression Results

Dep. Variable: Class No. Observations: 3819

Model: Logit Df Residuals: 3817

Method: MLE Df Model: 1

Date: Mon, 09 Dec 2024 Pseudo R-squ.: -0.3264

Time: 16:19:55 Log-Likelihood: -1916.0

converged: True LL-Null: -1444.5

Covariance Type: nonrobust LLR p-value: 1.000

coef std err z P>|z| [0.025 0.975]

Intercept -1.6500 0.256 -6.450 0.000 -2.151 -1.149

mean_uhi_DayNight 0.9911 0.424 2.335 0.020 0.159 1.823

Correlation matrix plotted using corrgram() can display correlation between variables.

Darker colors indicate a stronger correlation whereas lighter colors indicate a weaker correlations, Figure 12 indicates a weaker correlation between City shape and both day and night mean UHI value

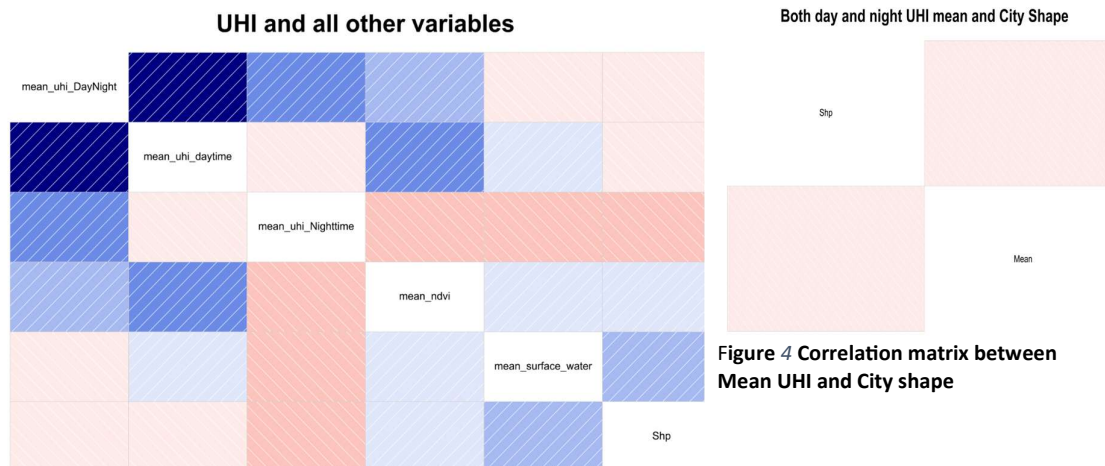


Figure 11 Correlation matrix using corrgram

Inference

Results of Anova test, Kruskal Wallis test, logistic regression model summary suggest that city shapes are weakly correlated to UHI day and night mean values

Acknowledgement

I would like to thank Mr. Ian Estacio for helping in obtaining NDVI raw data. The high resolution NDVI mean data calculated taking care of cloud cover greatly enhanced the quality of this report.

I would like to thank Dr Greg Chism and Dr Cristian Román-Palacios for their guidance and support throughout the project

Author contributions:

Conceptualization: CRP, PR

Methodology: PRR, CRP

Code Writing: UHI, NDVI, surface water- PRR, City Heights data-HMAF

Visualization: UHI, NDVI, surface water- PRR, City Heights data-HMAF

Supervision: CRP

Writing—original draft: PRR, CRP

Writing—review & editing: CRP, PRR

References

- [1] Urban heat island, Wikipedia, Available at https://en.wikipedia.org/wiki/Urban_heat_island
- [2] Nidhi Singh, Saumya Singh, R.K. Mall. Urban ecology and human health: implications of urban heat island, air pollution and climate change nexus. Available at <https://www.sciencedirect.com/science/article/abs/pii/B9780128207307000173#:~:text=The%20UHI%20effect%20has%20caused%20an%20increase%20in,in%20nearby%20areas%20thus%20developing%20new%20ecological%20implications.>
- [3] Urban Heat Island (nationalgeographic.org) Available at <https://education.nationalgeographic.org/resource/urban-heat-island/>
- [4] David L. Chandler, Urban heat island effects depend on a city's layout, Available at <https://news.mit.edu/2018/urban-heat-island-effects-depend-city-layout-0222>
- [5] City module - City cluster and urban heat islands (Europe), Available at <https://www.pik-potsdam.de/cigrasp-2/city-module/heat-island-cluster/index.html>
- [6] Nyuk Hien Wong, Chun Liang Tan, Dionysia Denia Kolokotsa & Hideki Takebayashi Available at <https://www.nature.com/articles/s43017-020-00129-5>

[7] Mackres, E., Pool, J. R., Shabou, S., Wong, T., Anderson, J., Gillespie, C., & Alexander, S. New Data Dashboard Helps Cities Build Urban Resilience in a Changing Climate. Available at <https://www.wri.org/update/new-data-dashboard-shows-climate-change-risks-in-cities>

[8] H. M. Fattah Capstone Report, Available at https://github.com/datadiversitylab/city-3D-shapes/tree/main/docs/Capstone_report_for_H_M_Abdul_Fattah.pdf

[9] Urban Heat Island dataset, Various UHI datasets available at <https://developers.google.com/earth-engine/datasets/tags/uhi> or direct dataset download - <https://sedac.ciesin.columbia.edu/data/set/sdei-yceo-sfc-uhi-v4/data-download>

Citation- Center for International Earth Science Information Network - CIESIN - Columbia University. 2018. *Gridded Population of the World, Version 4.11 (GPWv4): Population Count*, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4JW8BX5>. Accessed 10/15/2024

[10] NOAA CDR AVHRR NDVI: Normalized Difference Vegetation Index, Version 5 dataset, available at MODIS Combined 16-Day NDVI available at

https://developers.google.com/earth-engine/datasets/catalog/MODIS_MCD43A4_006_NDVI

[11] MCD18A1.061 Surface Radiation Daily/3-Hour dataset, Available at

https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MCD18A1

[12] JRC Global Surface Water Mapping Layers, v1.4 dataset. Available at

https://developers.google.com/earth-engine/datasets/catalog/JRC_GSW1_4_YearlyHistory

Citation- 'Source: EC JRC/Google'

[13] GHSL - Global Human Settlement Layer. Available at

<https://human-settlement.emergency.copernicus.eu/datasets.php>

(GHS-BUILT-H - R2023A dataset used for building height calculation)

[14] Kelso, N.V. and Patterson, T. (2012). World Urban Areas, LandScan, 1:10 million (2012). Available at <https://geo.nyu.edu/catalog/stanford-yk247bg4748>

[15] Elsen, PR., & Tingley, MW. (2015). Global mountain topography and the fate of montane 296 species under climate change. *Nature Climate Change*, 5(8), 772-776 Available at

<https://www.nature.com/articles/nclimate2656>

[16] One-Way ANOVA Test in R - Easy Guides - Wiki - STHDA Available at

<https://www.sthda.com/english/wiki/one-way-anova-test-in-r>

[17] Kruskal-Wallis Test in R - Easy Guides - Wiki - STHDA Available at

<https://www.sthda.com/english/wiki/kruskal-wallis-test-in-r>

[18] Correlation matrix with ggally – the R Graph Gallery__ Available at

<https://r-graph-gallery.com/199-correlation-matrix-with-ggally.html>