An analysis of the 3D shape of cities

Authors

H M Abdul Fattah,[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

We describe the three-dimensional shapes of cities using a dataset that summarizes information about the heights of buildings globally. To analyze the city shapes, we compute histograms of area versus building heights (hypsographic curves) for 5973 cities around the world. Overall, cities exhibited four distinct area-height patterns: unimodal right skew (linear/exponential decline in area with height; 56.27% of cities), bimodal distribution, irrespective of skew (a curve with peaks at low and high heights; 41.15%), unimodal with no skew (a normal curve with peak area at mid-height; 2.41%), and unimodal left skew (linear/exponential increase in area with height; 0.17%), termed as 'pyramid', 'hourglass', 'diamond', and 'inverse pyramid' shapes, respectively. To address concerns such as the impacts of urban heat islands and climate change, we anticipated greater variance in city shapes. However, this classification system may not be conducive to addressing these issues. We plan to use more sophisticated strategy considering various variables alongside city shapes to address the issues effectively.

Teaser

Most cities are either pyramid-shaped or hourglass-shaped, with few having a diamond shape, and the inverse pyramid shape is extremely rare.

Introduction

Understanding the three-dimensional shapes of cities, particularly the heights of their buildings, is crucial for effective urban planning. This information further allows us to address contemporary challenges associated with urbanization, including the association between the urban heat island (UHI) effects and contemporaneous climate change. Urban areas experience higher temperatures than surrounding rural areas is known as Urban Heat Island (UHI) effect [1]. Prior research have found that UHI effects depend on a city layout [2]. In fact, the way streets and buildings are arranged strongly influences how heat builds up. Similarly, building materials like concrete absorb and radiate heat, leading to dramatic temperature increases of up to 10 degrees Celsius in some cities. Elevated health problems and increased energy usage during hot weather, highlighting the importance of understanding and addressing the UHI effect. UHI is positively correlated with city size [3].

Climate change affects living conditions within cities. The effects of climate change are seen on air pollution, increased flood frequency, patterns of occurrence in animal and plant

species, and oftentimes, increased temperatures. In a recent article, Mackres et al. [4] presented a new tool to get information about how climate change is projected to affect particular cities. This tool helps cities see how climate change is connected to city life using different measurements of how sustainable things are likely to be. This helps decision-makers re-design and develop existing parts of their cities to provide more sustainable and equitable livelihood opportunities, and responsibly expand their cities in the most nature-sensitive ways possible.

Hughes et al. [5] studied how rising temperatures from climate change affect household energy use, with people relying on extra energy for heating and cooling. The Extra-Metabolic Scholander-Irving model (EMSI) describes this, but has only been studied in moderately warm cities, leaving out extremely hot temperatures. They demonstrate in Arizona, a hot semi-arid state, that household energy use increases in response to the hottest summer months, completing the EMSI model. Energy use is lowest in spring and fall, rising in colder temperatures in winter, with income disparities affecting cooling demand. Their findings provide insights into the challenges of sustaining thermoregulation amid global warming, studied in Yuma, Arizona, where even a small temperature rise has significant effects on cooling needs.

Elsen et al. [6] studied how species worldwide are responding to climate change by shifting their ranges in latitude and elevation. Montane species, in particular, face significant challenges due to their narrow elevational ranges and high rates of local endemism. Concerns arise over their fate under climate change, with expectations of further upslope shifts potentially leaving them with less habitable area. They analyzed the elevation availability of surface area for a global data set containing 182 of the world's mountain ranges. They show that, mountain range topographies exhibited four distinct area–elevation patterns: decreasing (decline in area with elevation), increasing (increase in area with elevation), a mid-elevation peak in area, and a mid-elevation trough in area. It suggests that many species may encounter increases in available area as they shift upslope. These findings challenge prevailing ecological assumptions. Further exploration of underlying mountain topography can inform strategies to mitigate the impacts of climate change on montane species and their habitats. For our work, we are taking inspiration from this work to get the classification categories for cities.

Our objective is to explore how building shapes interact with urban policies and regulations. In future, we plan to develop urban planning strategies aimed at mitigating climate change impacts. This study aims to contribute to this endeavor by exploring the relationship between urban morphology, building heights, and the UHI effect, with the goal of informing smarter and more sustainable urban development practices.

Materials and Methods

Dataset Description:

The project utilizes two datasets. One is a shape file outlining city boundaries, another one is building heights around the world. Below, we describe the sources and characteristics of each of these datasets.

Shapefile data [7]: The dataset is a shapefile representing urban areas with dense human habitation worldwide. It emphasizes regional significance over population census for selection and includes population estimates derived from the Oak Ridge National Laboratory's LandScan population database for 90% of the depicted cities. These estimates

consider total metropolitan population rather than administrative boundaries. The dataset underwent a process of conversion from raster to vector, filtering out rural areas with fewer than 200 persons per square kilometer, and aggregating urban pixels into contiguous units. Thiessen polygons were created based on selected city points and used to intersect the contiguous city boundaries to produce bounded areas. Population totals were recalculated using aerial interpolation, assuming uniform population distribution within each pixel, at nested scales of 1:300 million, 1:110 million, 1:50 million, 1:20 million, and 1:10 million, with the data represented at a 1:10,000,000 scale. This dataset is part of the Natural Earth Collection and is the public domain.

GHS building height dataset [8]: Global Human Settlement Layer (GHSL) offers various datasets that have separate applications. Specifically, GHSL offers three spatial raster datasets GHS-BUILT-S, GHS-BUILT-H, and GHS-BUILT-V that differ in specific aspects of built-up areas they represent: GHS-BUILT-H: This dataset represents the spatial distribution of building heights per cell. GHS-BUILT-H specifically quantifies the vertical dimension of built-up areas, indicating the varying heights of structures.

GHS-BUILT-S: This dataset depicts the distribution of built-up surfaces, expressed in terms of the number of square meters. GHS-BUILT-S focuses on the horizontal extent or footprint of built-up areas, quantifying the surface area covered by structures.GHS-BUILT-V:

This dataset depicts the distribution of built-up volumes, expressed as the number of cubic meters. It reports on the built-up volume allocated to dominant non-residential (NRES) uses. GHS-BUILT-V focuses on the three-dimensional aspect of built-up areas, quantifying the volume of space occupied by structures.

We are using GHS-BUILT-H dataset in this project. Data is only available for the year 2018. GHS-BUILT-H offers two estimate models. They are (i) Average of the Gross Building Height (AGBH), (ii) Average of the Net Building Height (ANBH). The difference between Average Gross Building Height (AGBH) and Average Net Building Height (ANBH) lies in the considerations of certain architectural features and elements. While AGBH considers all architectural features contributing to the visual profile of a building, ANBH provides a more refined measurement by excluding certain non-habitable elements. For this project, we are using average gross building heights (AGBH) data.

Experimental Design:

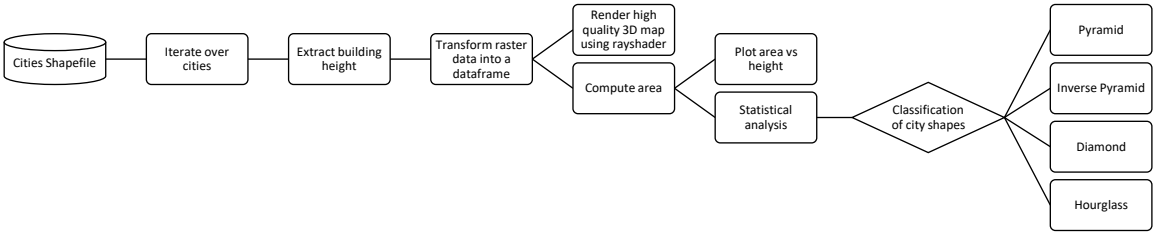

Fig. 1. Methodology of our work

We use a shapefile of urban cities to mark the boundaries of cities. These shapefiles have information on 6018 cities worldwide and their features. We utilize the 'sf' package [9] to extract this information from the shapefile. They formed the basis for our further analysis. Iterating over each city in the dataset, we extract the building height data using Terra package [10] and utilize it for visualization purposes. This involves saving a GeoTIFF file and generating a high-quality 3D map using rayshader package [11]. First, the raster data representing building heights is transformed into a dataframe. This conversion ensures that each cell's height value, along with its corresponding x and y coordinates, is organized into

a structured tabular format. We now render and save a 3D image for each city. This process is resource-intensive and time-consuming due to the need to render high-quality images for each city individually.

To visualize area vs height plot for cites, we estimated the area (measured in square meters) covered by each cell in the cropped raster. We computed the area covered by each cell in the raster using a built-in function called 'cellSize()' [12], resulting in a dataframe containing x and y coordinates along with the area information. Both dataframes are generated with the consideration of removing any cells with missing values (NA) to ensure data integrity.

We developed a custom function to link each city with its continent because the dataset do not have continent details. This addresses the fact that there are multiple cities with different names located on different continents. This function figures out which continent each city belongs to. It looks at the center of each city and compares it with the continents to find the closest one. If a city's shape intersects with a continent, it's labeled with that continent. If not, it gets the continent with the closest distance. It provides geographic context to the cities, enabling further exploration and understanding of the dataset in terms of continental distributions. We utilize Natural Earth's 1:50m public domain data [13] to extract the boundaries of continents.

Statistical analysis:

We examined particular properties that could potentially contribute to identifying the categories for classifying city shapes. We calculate key summary statistics like average building height, skewness [19], dip statistic [20], and modality for each city. Cities with higher average building heights may indicate denser urban development. A positive skewness suggests that the distribution is biased towards taller buildings, while a negative skewness indicates a bias towards shorter buildings. This can provide insights into the overall trend of building height distribution within a city. The dip statistic measures the presence of multi-modality in the distribution of building heights. A high dip statistic suggests the presence of multiple peaks or modes in the distribution, indicating areas with different types of buildings or urban development patterns within the city. Modality refers to the number of peaks or modes in the distribution of building heights. By analyzing these summary statistics, we gain insights into the variation, trends, and patterns of building heights within cities.

Building height distributions within cities are categorized into different classes based on their skewness and dip statistic values. Skewness, which measures the asymmetry of the distribution, is utilized to determine whether the distribution is unimodal right skew (Pyramid), unimodal left skew (Inverse Pyramid), or has no significant skew (Diamond) [6]. Additionally, the presence of multimodality, indicated by the dip statistic, results in classification 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. Through this classification, the analysis gains insights into the diverse patterns of building height distributions within urban areas.

Results

For each city, we generated a high-quality 3D map using rayshader.

Fig. 2. 3D visualization of city heights represented as a raster grid for a sample city named 'Aalborg (Denmark, Europe)'

Figure 2 displays a 3D visualization of city heights for a sample city named 'Aalborg', where each pixel in the grid corresponds to a location in the city, with the color (shade of gray) indicating the height of the building at that location. Lighter colors indicating higher elevations and darker colors indicating lower elevations. The plot includes shadows to add depth and realism to the 3D visualization, enhancing the perception of elevation differences. The light source is positioned at an angle of 315 degrees, illuminating the terrain from the top left corner of the plot. The plot is rotated at a specific angle (theta = 0), providing a particular view of the city terrain. We can tune these parameters to have different angles or shadows. The figure is rendered using rayshader's plot_gg function, which combines the capabilities of ggplot2 for data visualization with rayshader for 3D rendering.

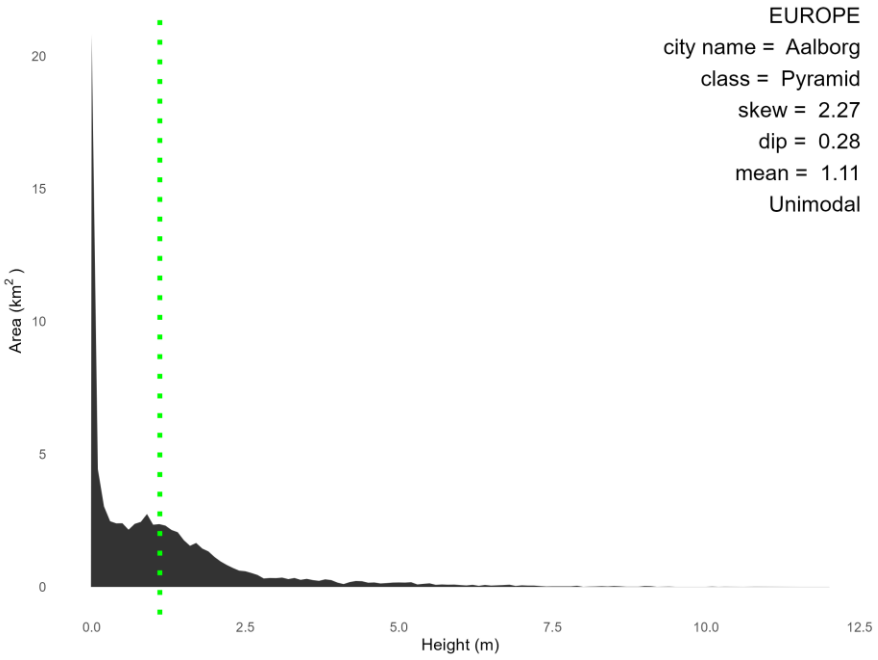

Fig. 3. Distribution of building heights and corresponding area coverage for a sample city named Aalborg, Denmark (Europe)

Figure 3 shows a filled area plot representing the relationship between height (on the x-axis) and the corresponding area covered (on the y-axis) for the buildings within the city boundaries. Additionally, annotations provide information about the city, such as its name, continent, class (e.g., Pyramid, Inverse Pyramid, Diamond, and Hourglass), skewness, dip statistic, and modality. The green dotted vertical line indicates the mean height of the buildings in the city. The height values in the dataframe are rounded to two decimal places. This ensures consistency and precision in the representation of height measurements. The area values, originally in (m^2), are converted to (km^2) by dividing each value by 1,000,000. This conversion simplifies the interpretation of the area data, making it more intuitive and easier to compare with other geographical measurements. The area values in the dataframe are aggregated based on unique rounded height values. This aggregation sums up the areas covered by buildings at each distinct height level. The plot is saved as an image file (".png") with a filename based on the city name.

Classification Results

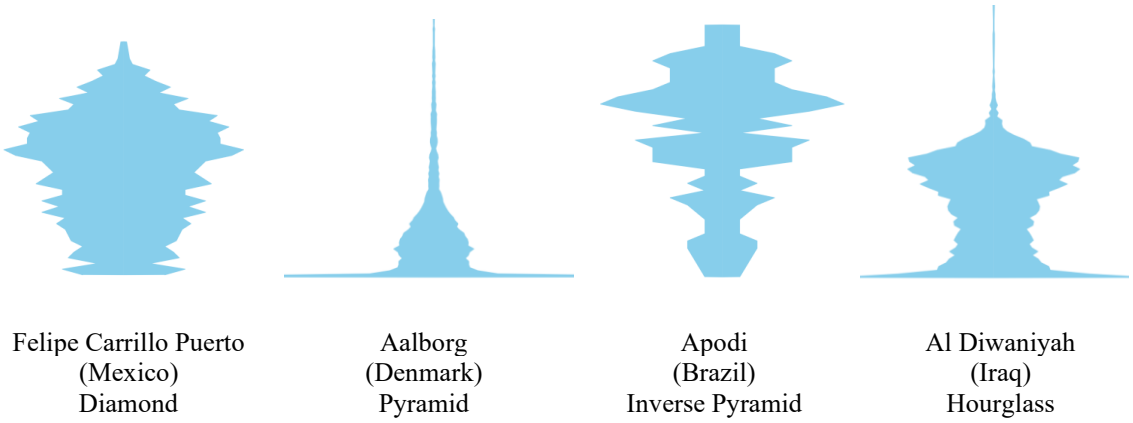

Fig. 4. Hypsographic curve for different cities from various classification level

We computed histograms of area versus height, also known as hypsographic curvse for each city (Fig. 4). We then classified shapes into four categories, representing the full diversity of hypsographic patterns observed, by analyzing the skewness and modality of the height–area profiles for each city.

Table 1. Classification report for all cities

Classification	No. of cities	Percentage
Pyramid	3363	56.27 %
Hourglass	2456	41.15 %
Diamond	144	2.41 %
Inverse Pyramid	10	0.17 %

Categories were determined as unimodal right skew (that is, linear/exponential decline in area with height; 56.27% of ranges), bimodal distribution, irrespective of skew (that is, a curve with peaks at low and high heights; 41.15%), unimodal with no skew (that is, a normal curve with peak area at mid-elevation; 2.41%), and unimodal left skew (that is,

linear/exponential increase in area with height; 0.17%; See Table 1). In terms of topography, we describe the four hypsographic patterns as ‘pyramid’, ‘hourglass’, ‘diamond’ and ‘inverse pyramid’ city shapes, respectively. Cities are most commonly categorized as having either a pyramid-shaped or hourglass-shaped structure, with few exhibiting a diamond shape, and the inverse pyramid shape being exceptionally rare.

Discussion

In this project, we look at how cities are shaped around the world. We used data about building heights are in cities and maps that show city borders. We found that cities have four different shapes based on their statistical measures. The majority of cities exhibit either a pyramid shape or an hourglass shape. Few cities have a diamond shape. The inverse pyramid shape is very rare, with only 10 cities exhibiting this form. Out of 6018 cities, raster file for 45 are broken or corrupted. The errors occurred due to several technical constraints, including limitations on disk space and file access permissions, which must be resolved before incorporating them into the study. In this analysis, we excluded those 45 cities. The existing classification system lacks the complexity needed to tackle challenges such as urban heat islands and climate change due to the limited variation in city shapes observed. It suggests the necessity of a more sophisticated strategy considering various variables alongside city shapes to address these issues effectively.

Despite previous studies, there's a need for updated understanding and enhanced mechanisms for UHI effects, considering the diverse characteristics of cities in terms of size, climate, and economic development [21]. Hu et al. [22] explored the influence of density and morphology on the Urban Heat Island (UHI) intensity in urban areas. By simulating urban climates of various cities under identical weather conditions, they proposed a model combining logarithmic city area and logarithmic gross building volume to capture the effects of building density. The authors found that UHI intensity is not only related to city size but also to density and the amplifying effect of urban sites on each other. We plan to combine our finding about morphological shape of cities with existing models to create more robust mechanisms to address UHI effect.

This research [23] employs boosted regression trees to assess how 2D and 3D urban indicators influence Urban Heat Island (UHI) dynamics across seasons in Beijing's Olympic Area. Key factors like NDBI, NDVI, and building height exhibit complex, nonlinear relationships, with implications for urban planning and environmental management, particularly regarding the strategic integration of green infrastructure such as evergreen trees. In our study, we are storing the calculated city properties in a csv file. We plan to apply some machine learning techniques to find some meaningful patterns that could provide valuable insights.

Acknowledgments

Author contributions:

Conceptualization: CRP, HMAF
Methodology: HMAF, CRP
Code Writing: HMAF
Visualization: HMAF
Supervision: CRP
Writing—original draft: HMAF, CRP
Writing—review & editing: CRP, HMAF

Data and materials availability: All data are available in the main text or the supplementary materials.

References

[1] Urban heat island, Wikipedia, Available at https://en.wikipedia.org/wiki/Urban_heat_island

[2] 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

[3] 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

[4] 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.

[5] Hughes, H. B., Breshears, D. D., Cook, K. J., Keith, L., & Burger, J. R. (2023). Household energy use response to extreme heat with a biophysical model of temperature regulation: An Arizona case study. PLOS Climate, 2(4), e0000110.

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

[7] Kelso, N.V. and Patterson, T. (2012). World Urban Areas, LandScan, 1:10 million (2012). Made with Natural Earth, online at http://www.naturalearthdata.com

[8] Pesaresi, Martino; Politis, Panagiotis (2023): GHS-BUILT-H R2023A - GHS building height, derived from AW3D30, SRTM30, and Sentinel2 composite (2018). European Commission, Joint Research Centre (JRC) [Dataset] doi: 10.2905/85005901-3A49-48DD-9D19-6261354F56FE

[9] sf: Simple Features for R, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/sf/index.html

[10] terra: Spatial Data Analysis, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/terra/index.html

[11] rayshader: Create Maps and Visualize Data in 2D and 3D, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/rayshader/index.html

[12] cellSize: Area covered by each raster cell, The Comprehensive R Archive Network, URL: https://search.r-project.org/CRAN/refmans/terra/html/cellSize.html

[13] Physical Labels, Natural Earth, Avaialable at: https://www.naturalearthdata.com/downloads/50m-physical-vectors/50m-physical-labels/

[14] tidyverse: Easily Install and Load the Tidyverse, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/tidyverse/index.html

[15] dplyr: Introduction to dplyr, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html

[16] raster: Geographic Data Analysis and Modeling, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/raster/index.html

[17] giscoR: Download Map Data from GISCO API - Eurostat, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/giscoR/index.html

[18] gridExtra: Miscellaneous Functions for "Grid" Graphics, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/gridExtra/index.html

[19] e1071: Misc Functions of the Department of Statistics, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/e1071/index.html

[20] diptest: Hartigan's Dip Test Statistic for Unimodality, The Comprehensive R Archive Network, URL: https://cran.r-project.org/web/packages/diptest/index.html

[21] City module - City cluster and urban heat islands (Europe), URL: https://www.pik-potsdam.de/cigrasp-2/city-module/heat-island-cluster/index.html

[22] Li, Y., Schubert, S., Kropp, J. P., & Rybski, D. (2020). On the influence of density and morphology on the Urban Heat Island intensity. Nature communications, 11(1), 2647.

[23] Hu, Y., Dai, Z., & Guldmann, J. M. (2020). Modeling the impact of 2D/3D urban indicators on the urban heat island over different seasons: A boosted regression tree approach. Journal of environmental management, 266, 110424.

Supplementary Materials

GitHub: https://github.com/datadiversitylab/city-3D-shapes-hmfattah