CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Stratifying urban neighbourhoods by income: Comparison of a subjective participatory approach and an objective statistical analysis of deprivation indicators in Dar es Salaam, Tanzania
Authors	Msuya, Ibrahim;Boudou, Martin;Levira, Francis;Moshi, Irene;O'Dwyer, Jean;Killeen, Gerry
Publication date	2025-04-30
Original Citation	Msuya, I., Boudou, M., Levira, F., Moshi, I., O'Dwyer, J. and Killeen, G. (2025) 'Stratifying urban neighbourhoods by income: Comparison of a subjective participatory approach and an objective statistical analysis of deprivation indicators in Dar es Salaam, Tanzania', Environment and Planning B: Urban Analytics and City Science. https://doi.org/10.1177/23998083251334686
Type of publication	Article (peer-reviewed);journal-article
Link to publisher's version	10.1177/23998083251334686
Rights	© 2025, the Authors. Reprinted by permission of SAGE Publications. This is a post-peer-review, pre-copyedit version of the following article: The final authenticated version is available online at: https://doi.org/10.1177/23998083251334686
Download date	2026-11-06 03:24:55
Item downloaded from	https://hdl.handle.net/10468/17572



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Stratifying urban neighbourhoods by income: Comparison of a subjective participatory approach and an objective statistical analysis of deprivation indicators in Dar es Salaam, Tanzania

Key words*: Participatory Mapping, Neighbourhood Stratification, Income Inequality, Latent, Profile Analysis, Deprivation Indicators*

Abstract

Mapping out geographic heterogeneities across urban neighbourhoods can inform urban planning and intervention targeting. Despite their subjectivity, intuitive participatory stratification approaches (PSAs) are becoming increasingly popular because of their affordability and practicality. In contrast, more objective statistical approaches, like latent profile analysis (LPA), typically require rich survey data and advanced capacities that are often lacking in low and middle-income countries. This study therefore assessed these two distinct approaches, to compare their effectiveness, applicability and complementarity for identifying geoeconomic heterogeneities of urban neighbourhoods in a typical contemporary African city. This study assessed a PSA to stratify neighbourhoods across the Tanzania city of Dar es Salaam in terms of income, by comparing it with a complementary LPA of national census data to stratify them in terms of deprivation. A consultative community-based workshop was used for the PSA, while 15 selected deprivation indicators from the census data were used to profile them using LPA. While the PSA allocated neighbourhoods to five income strata, six clear deprivation strata could be distinguished by LPA. A strong positive correlation was observed between the stratum identified by the LPA and that obtained through the PSA ($\rho = 0.739$, $p < 0.0001$). Furthermore, paired comparison of the two sets of correlation coefficients between each deprivation indicator and the stratum assigned by each stratification approach revealed no difference ($V = 33$, p-value = 0.1354), confirming that the two approaches yielded very similar patterns of stratification. Also, the two approaches yielded broadly comparable cartographic pictures of the city, depicting similar spatial distribution of wealth and poverty. Overall, this evidence indicates that subjective community knowledge and lived experience may be invaluable for understanding built environment and for mapping out pockets of poverty and affluence at fine scales with minimum resources.

Introduction

More than half of the world's population now lives in towns and cities (United-Nations, 2016). While Africa and eastern and southern Asia have been urbanized most rapidly, the pace of urbanization in these regions is expected to slow over the coming decade (UN-Habitat, 2020). Meanwhile, limited government investment in urban planning has contributed to more extensive and overcrowded informal settlements, where approximately 47% of the world's 4.3 billion urban dwellers live (UN-Habitat, 2020). In Tanzania, for instance, it has been estimated that more than 75% of settlements have been informally developed (Magina et al., 2020). Indeed, Dar es Salaam is the largest city in Tanzania, where >75% of the population live in informal settlements with highly heterogeneous wealth distribution (Rasmussen, 2013).

Several historical studies on the early development of what is now the old city centre of Dar es Salaam have rationalized the distribution of settlement patterns in terms of colonial planning processes, which were heavily influenced by residential segregation policies. During the colonial era, the city was planned in a way that segregated people by race and income, with affluent neighbourhoods located near the city centre and along the Indian Ocean (Europeans and Indians) and the less affluent near and around the city centre (black Africans, referred to in that era as natives) (Brennan & Burton, 2007). The evident pattern in the city demonstrates that a significant portion of its development has unfolded through informal processes (Msuya et al., 2021). Informal settlements tend to be situated either within formal areas or near them, aligning with the contextual reality of the city's historical evolution. This historical development trajectory has contributed to an uneven distribution of neighbourhood attributes in terms of income and affluence, a disparity that persists today. However, despite the inclination to attribute these disparities solely to colonial influences, a more thorough investigation challenges this prevailing narrative. Contrary to the common assumption that the

current state of Dar es Salaam, beyond the old city, is a direct result of colonial planning (Brennan, 2007), in fact, it emerged from limited planning, especially during the city's rapid expansion from the 1970's onward (Brennan and Burton, 2007)

While both formal and informal settlements co-exist side-by-side in cities, they exhibit varying neighbourhood characteristics, constantly evolving at very fine scales. This results in a myriad of spatial, sociodemographic, economic, and cultural diversities coexisting within relatively small, but nevertheless, very heterogeneous geographic areas. This vibrant evolutionary process is particularly dynamic in unplanned, informal settlements, catalyzing different types of disparities, such as health inequality, social division, exclusion, segregation, which are typically assessed in terms of income inequality or socioeconomic deprivation, (Odoi et al., 2005). Despite the conceptual differences in how income and deprivation are respectively used to assess the two-sided coin of wealth and poverty (Supplementary Box S2), either may be readily applied to surveying the distribution of inequalities across urban areas.

For the first time within the context of internationally agreed development goals, the 2030 Agenda aims to reduce inequality based on income (UN, 2020b). While efforts are being made to bridge the gap between rich and poor (Cingano, 2014; Apel, 2020), there is still insufficient investment in methodologies for socioeconomic profiling of urban neighbourhoods (2020). In response to that methodological need, this study set out to evaluate a subjective participatory stratification approach (PSA) to classifying neighbourhoods across the Tanzanian city of Dar es Salaam in terms of income, by comparing it with an objective latent profile analysis (LPA) of deprivation based on national census data, see Supp Box 2, summary of the relevant evidence base and methodological considerations that informed the formulation of the study design.

Methods

Geographical context and study data

Dar es Salaam is the largest city in Tanzania, located on the Indian Ocean coast (Supp Fig 1). The city had a population of approximately 4.4 million in 2012 but currently has approximately 5.4 million inhabitants, distributed across an administrative area that has remained constant at 1,590 km^2 (URT and NBS, 2022). It is also the port in Tanzania, contributing substantially to the national economy and served as a government administrative centre before *de facto* relocation to the official capital, Dodoma, was practically implemented in 2016 (Todd et al., 2019).

The study reported herein used Tanzania's 2012 Population and Housing Census (PHC) data from the National Bureau of Statistics of Tanzania (NBS) to generate variables which were considered likely to represent deprivation in the study context (Supp Table 1). The data is presented via administrative units, starting from regions down to districts, wards (*kata*) and neighbourhoods (*mitaa*, literally translated as *streets* but actually means a larger, formalized administrative area with defined boundaries, rather than just a road and structures built along it). The data from PHC primarily consisted of socio-demographic information such as gender, age, disability, marital status, citizenship, place of residence, education status, migration patterns, economic activities, fertility rates, general mortality rates, and enrolment in social security programs. Other data include household with members who are living in diaspora, food and cash crops cultivation, and information about asset ownership and housing conditions. The current study used data aggregated at the neighbourhood level (Supp Fig 1C), the finest scale freely available to the public upon request for research purposes to derive deprivation

indicators based relevant variables in the PHC (Supp Table 1). The dataset included 452 neighbourhoods from three districts that comprised Dar es Salaam at the time (Supp Fig 1A)

Neighbourhoods in Dar es Salaam city are unevenly distributed, with Kinondoni District having the highest number of neighbourhoods, while Ilala has the lowest number of neighbourhoods (Supp Table 2). In terms of population, Kinondoni also has the highest share of residents living in this district, while Ilala is the least. However, from a historical perspective, Ilala stands as the most well-established district, comprising neighbourhoods with a rich legacy that dates back over a century. On the other hand, most parts of Temeke were relatively new and were experiencing rapid growth and development at the time.

Study design

A mixed method approach was adopted, based on considerations of the relevant methodological literature (Supplementary Box S2). Overall, this study sought to assess a subjective participatory stratification approach (PSA) to classifying neighbourhoods in terms of income level by comparing it with an objective statistical approach using latent profile analysis (LPA), to stratify neighbourhoods in terms of deprivation based on socioeconomic data from the 2012 PHC in the Tanzania city of Dar es Salaam. While the income which is referred by the PSA is inherently subjective and based on the lived experiences and knowledge of the city's residents, the LPA of deprivation strata in quantitative terms is a more objective socioeconomic metric derived from formal ways of living standards, conditions, and education at the neighbourhood level (Dirksen et al., 2022).

Participatory stratification approach (PSA)

The PSA involved active participation of community members and professionals in the stratification of Dar es Salaam neighbourhoods into different income strata. This one-day workshop, carried out on Friday, 22nd March 2019, from 09:00 AM to 17:00 PM at Ifakara Health Institute offices in Mikocheni, Dar es Salaam, involved 50 people, specifically five residents from each district recommended by the neighbourhood leaders, two town planners

from each district, three community development officers from each district (Supp Table 2), five practicing town planners from the private sector, two urban and regional planning lecturers from Ardhi University and 13 Ifakara Health Institute scientists working in urban projects in Dar es Salaam. The task started by collectively agreeing on the number of income strata to which the city's neighbourhoods should be assigned. Then, each of the 452 neighbourhoods were assigned to a respective stratum by each individual participant, following which a consensus was established through group discussion to collectively agree on the stratum it was finally assigned to (Supp Fig 4). Although the task fundamentally relied on the participants knowledge and lived experience of the city, some other complementary sources of information, such as satellite images, were also used to facilitate the process of informed debate and consensus building.

Latent profile analysis (LPA)

Deprivation profiles were identified as latent (unobserved) variables, inferred from patterns in the observed socioeconomic indicators within the neighbourhood-level PHC data. These profiles were distinguished by analysing the relationships among relevant variables to uncover underlying groupings. The first step of the LPA was the identification of the most appropriate recorded variables. Based on the data and context, 15 variables in the PHC data that were intuitively expected to be closely associated with deprivation were identified. These included education indicators such as; *proportion completed any school level, proportion educated to university level, proportion educated to secondary level,* literacy and language proficiency; *proportion of literate, proportion of English literacy,* economic factors; *proportion of potentially economically active individuals, proportion enrolled in social security, proportion household with members in diaspora,* social and legal indicators*; proportion with registered birth certificate,* demographic factors; *average household size, proportion of female,*

proportion of single head households, and health and mortality statistics; *proportion with disability, proportion of death caused by diseases,* and *crude death rate.* Different indicators from the census data were used to derive these variables (Supp Table 1) for participating respondents or households, aggregated at neighbourhood level. To provide a standardised way to compare different quantities while allowing for more accurate and meaningful comparisons between variables in predicting outcomes or estimate probabilities, almost all variables were transformed into proportions, except for household size and crude death rate which were presented in their actual metrics units.

Descriptive statistics of the selected variables can be found in Figure S5 in the Supplementary Material, which shows prior expected general distribution of income and poverty across three study districts at the time, which are consistent with the actual existing situation in the city.

Depending on the nature of data, different six LPA models structures can be used which range from simple to complex, determined by how variance and covariance are treated and whether relationships between variables are considered, thus affecting model parsimony and complexity (Rosenberg et al., 2018). In model selection process, Pearson's rank correlation test was employed to quantify the level of correlation to inform optional specification of model assumption and maximize model performance. In addition, the correlations between variables can help with the interpretation of the resulting latent profiles, as such, if two variables exhibit high level of correlation it can suggest that these two variables may be driving the occurrence of profile and provide insights into the nature of the profile. Broadly speaking, a correlation coefficient $\rho \geq +/- 0.4$ and p value ≤ 0.05 was considered as an indication of strong correlation (Gogtay and Thatte, 2017). The Pearson's rank correlation analysis revealed that a significant correlation existed among ten out of the fifteen variables ($p \leq 0.05$), (Supp Fig 2 and 3). Hence, from the six tidyLPA models structures proposed by Rosenberg *et al.* (Rosenberg et al., 2019),

model structure six with varying variances and covariances was selected. It is the most complex model, so while it allows for the estimation of variances and covariances across profiles without constraints, potentially providing a deeper understanding of the relationships between the variables, it is also less parsimonious.

The LPA modelling was performed using R version 4.2.0 and RStudio version 2023.03.0+386 with the *tidyLPA* package. *tidyLPA* acts as an interface to the *mclust* package and provides the functionalities to estimate the commonly specified models used in the analysis, with different parameters depending on model complexity according to the structural forms developed by (Rosenberg et al., 2018). Seven criteria were used for best fit model selection (Supplementary table S3), specifically the Log-Likelihood (LogLik), Akaike Information Criterion (AIC), Difference in AIC values (ΔAIC), Consistent Akaike Information Criterion (CLC), Kullback-Leibler Information Criterion (KIC), Sample-Size Adjusted Bayesian Information Criterion (SABIC), and Entropy (Average Uncertainty or Randomness in Latent Class Assignments). Across the existing LPA literature, it has been widely reported that selection of the best fit model with the most informative number of distinct profiles remains a challenging task (Tein et al., 2013). Correspondingly, to validate the model optimization based on the fit statistics, an analytic hierarchy process (AHP), as proposed by (Akogul and Erisoglu, 2017) which helps in determining the correct number of profiles in a model based clustering was used.

Two approaches were used to interpret and assign meaning to each profile from the selected best fit model in relation to deprivation. Firstly, the distribution of the variables was plotted against every profile in a dot and violin plot. Secondly, using z-score or scaled mean of each variable on each profile. While median centre data around the second quartiles or 50% quantiles, z score means centre data around zero because the scaling involves standardizing each observed variable to have mean of zero and standard deviation of one ($\mu = 0$, $\sigma = 1$),

(DeVore, 2017). In the context of LPA, scaled means are often reported as part of the model output along with other fit statistics. Often, scaled means may have negative value due to scaling even though latent variable has only positive values, however, this does not necessarily indicate that the underlying construct is negative (Little et al., 2006). This helps in comparing the relative positions of the means across variables with different scales or variances, as well as for interpreting the magnitude and direction of the differences between the means of different profiles.

In LPA, after numbering profiles by ranking them according to their order of score, it is always a task to the investigators to interpret and assign their intuitive meaning (Ferguson et al., 2020). The ranked deprivation profiles were then each interpreted and assigned to named ranked deprivation stratum. The LPA deprivation strata were then compared to the results from the PSA using a non-parametric statistical correlation test (Spearman's rank correlation test), to assess the strength and direction of the relationship between the ranked strata from the two approaches. Further analyses were performed to cross validate the PSA income strata by plotting values using violin and dot plot for the same PHC variables broken down by PSA stratum. To test statistical significance difference between the two approaches, a Wilcoxon signed-rank test, was applied to compare the Spearman's correlation coefficients (ρ) obtained for each of these variables compared to either PSA or LPA.

Additionally, using the ranked results from the two approaches, neighbourhood income and deprivation strata from PSA and LPA, respectively, were mapped using ArcGIS Pro 3.1.2, to understand and compare stratification of the city neighbourhoods by the two approaches in terms of comparative spatial distribution. In spatial comparative analysis, ground truthing evidence of the existing conditions in the selected neighbourhoods, specifically, housing

density and quality of buildings, using in situ ground photographs and satellite images, were used to provide further descriptive insights into the meaning of the results.

Results

Latent profile identification

The best fit profile model identified by LPA of the census data was with 6-profiles: The goodness of fit consistently decreased as the number of profiles modelled increased, while entropy increased with a 98% classification quality reached for this final optimal 6-profile model that was the last one to converge (Supplementary Table S3). These model optimization results were confirmed using the AHP approach (Akogul and Erisoglu, 2017), which also indicated that the model with 6 profiles was optimal based on the fit statistics in Supplementary Table S3. The results, therefore, suggested that six clear deprivation strata could be identified based on census data, as compared to the five income strata obtained from the participatory stratification approach.

Interpreting the underlying meaning of identified latent profiles.

From the dot and violin plots in Fig 1, the value distributions of the most important individual indicators of deprivation show that ranking the numbered LPA profiles based on these same variables, yields an order of 2, 1, 3, 4, 5 and 6 for the profile numbers assigned by the LPA best fit model (Supp Table 3). Correspondingly, these were respectively assigned the following stratum names; *Highly Deprived, Deprived, Moderately Deprived, Moderately Affluent, Affluent,* and *Highly Affluent* (Table 1). To further examine and query these profile assignments based on variable value distributions, a calculated scaled mean for each variable was used (Supp Fig 6), consistent with other LPA studies that assigned interpreted profile meanings (Pastor et al., 2007; Yalçin et al., 2021; Voisin et al., 2018; Mathew and Doorenbos, 2022).

These scaled means for each variable within each stratum confirmed the same apparent profile order (Table 1) for most variables, as ranked from the most deprived to the most affluent (Supp Fig 6).

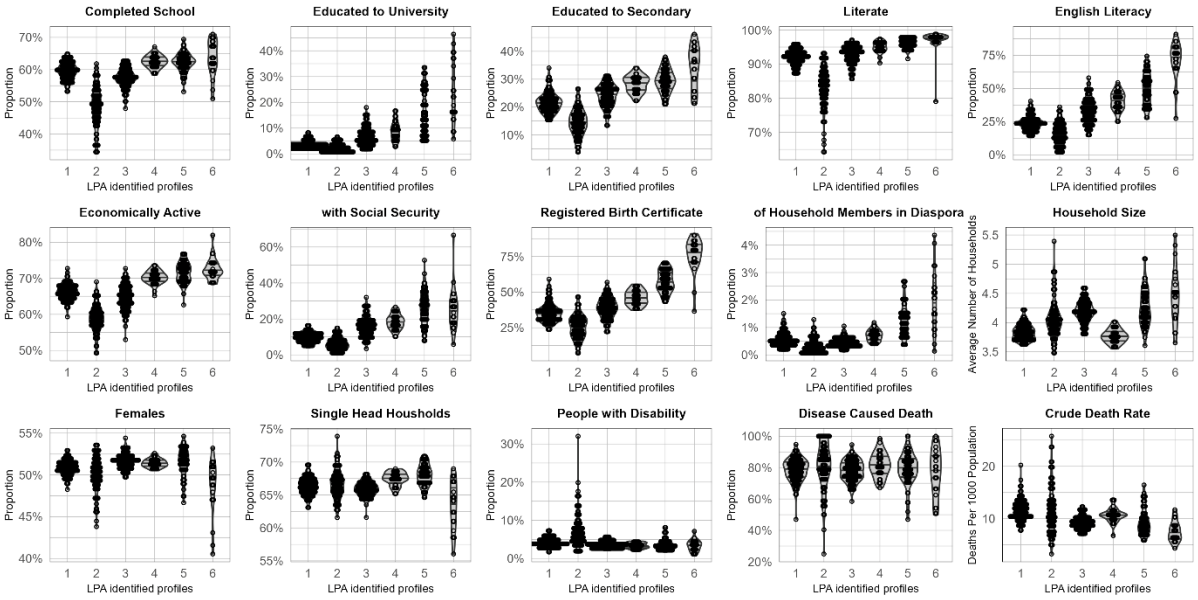


Fig 1: Violin and dot plot showing the distribution of values for 15 potential indicators of deprivation recorded by the 2012 national census, broken down across the six profiles distinguished by LPA analysis (Supp Table 3). Horizontal lines within the violin plot represent the first, second and third quartiles, or the 25%, 50% (median), 75% quantiles. The units for most of these variables are proportions of participating respondents or households, aggregated at neighbourhood level, the only two exceptions being household size and crude death rate.

Based on the interpretation of the LPA identified profiles (Supp Fig 6), six assigned deprivation strata could be resolved from the census data (Table 1). On the other hand, however, the results from participatory stratification approach revealed five income classes (Table 2) with *low income stratum* accounting for a remarkably low percentage of neighbourhoods while the *medium income stratum* had the highest.

Table 1: Latent profile analysis (LPA) assigned profiles numbers with investigator assigned ranked deprivation stratum names. Based on interpretation of Fig 1 and Supp Fig 6, with the number and proportion of neighbourhoods attributed to each stratum.

Profile Number	Assigned Profile Name	No. of Neighbourhoods (n)	%
2	Highly Deprived Neighbourhoods	98	22
1	Deprived Neighbourhoods	117	26
3	Moderately Deprived Neighbourhoods	115	25
4	Moderately Affluent Neighbourhoods	33	7
5	Affluent Neighbourhoods	66	15
6	Highly Affluent Neighbourhoods	23	5

Table 2: Neighbourhood income classes obtained from participatory stratification approach (PSA), with the number and proportion of neighbourhoods attributed to each stratum.

Class Number	Assigned Classes Name	No. of Neighbourhoods (n)	%
5	Low Income	19	4
4	Mixed Low and Medium Income	88	19
3	medium income	199	44
2	Mixed Medium and High	102	23
1	High Income	44	10

Comparison of PSA with LPA of the census data

The results of the spearman rank correlation test shows that there is a strong positive correlation ($\rho = 0.739$, $S = 4015865$, $p < 0.001$) between the stratum identified by the LPA of census data and the results of the PSA (Supp Fig 7).

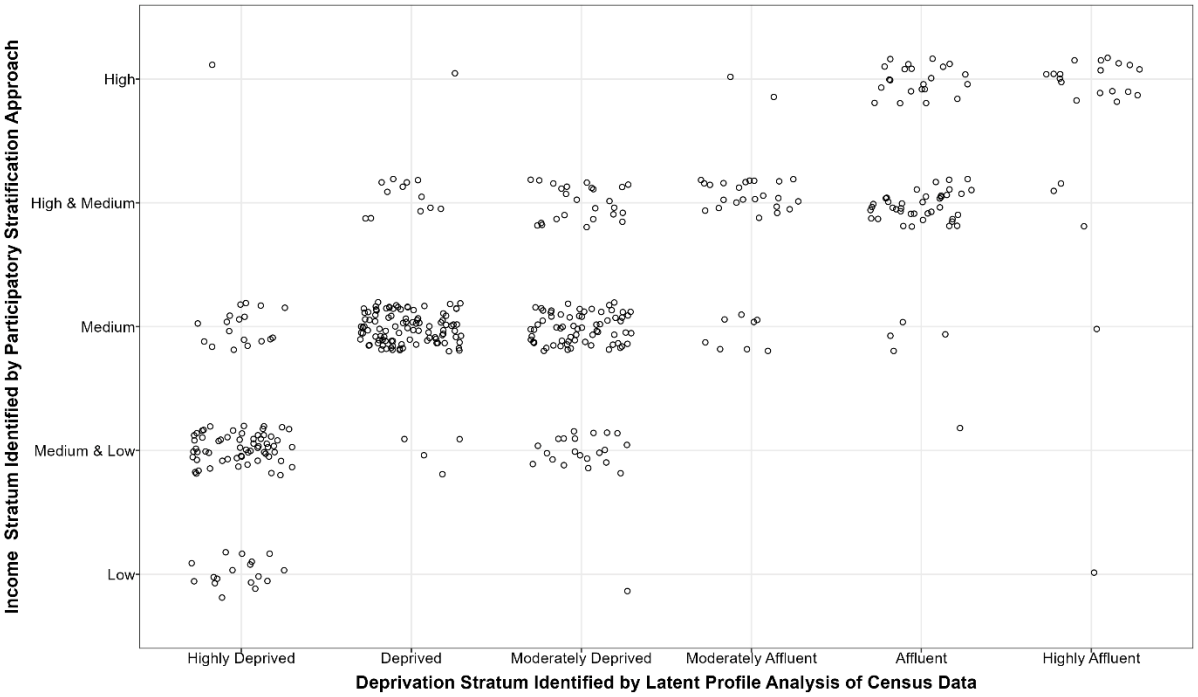


Fig 2: Jitter plot showing the relationship between the neighbourhood classification obtained through the participatory stratification approach (PSA), and through statistical latent profile analysis (LPA) stratification based on national census data.

The same variables used to stratify the neighbourhoods by LPA yielded distribution of values that were just as consistent with their income stratum as determined by the PSA (Supp Fig 8) as they were with their deprivation stratum as determined by the LPA (Fig 1 and Supp Fig 8).

Further analysis of Spearman's rank correlation tests between each variable from the census data and stratum rank as determined by each of the two different stratification approaches can be seen in Supp Figure S7 in the Supplementary Material. A Wilcoxon signed-rank test to compare these two sets of correlation coefficients in a paired manner revealed no difference between the two approaches ($V = 33$, p-value = 0.1354), indicating that the two approaches yielded very similar patterns of stratification.

Overall, the two stratification approaches yielded relatively similar cartographic pictures of wealth and poverty distribution across the city (Fig 3). High income, highly affluent

neighbourhoods are mainly located in or around the city centre and along the coast to the north, while low income, deprived neighbourhoods are generally distributed along the outskirts of the city. The distribution pattern of the medium income stratum, as determined by PSA, corresponds well with the moderately deprived and moderately affluent strata identified by LPA. Focusing on the city centre and areas immediately around it, the results reveal patches of low income and deprived neighbourhoods located near high income and highly affluent neighbourhoods, forming a pocket of poverty running from central Ilala into Temeke District. However, in the western direction this pocket of deprived, medium income neighbourhoods is enclosed by moderately deprived, mixed medium and high-income neighbourhoods.

The overall distribution, on the other hand, shows that most of the neighbourhoods classified by PSA as medium income in Fig 3A, such as Msimbazi Bondeni and Kigogo Kati, have been classified as deprived neighbourhoods by LPA in Fig 3B. Many of these neighbourhoods are found in the same pocket of deprived neighbourhoods described above that cuts across all three districts but mostly sits in northern Temeke.

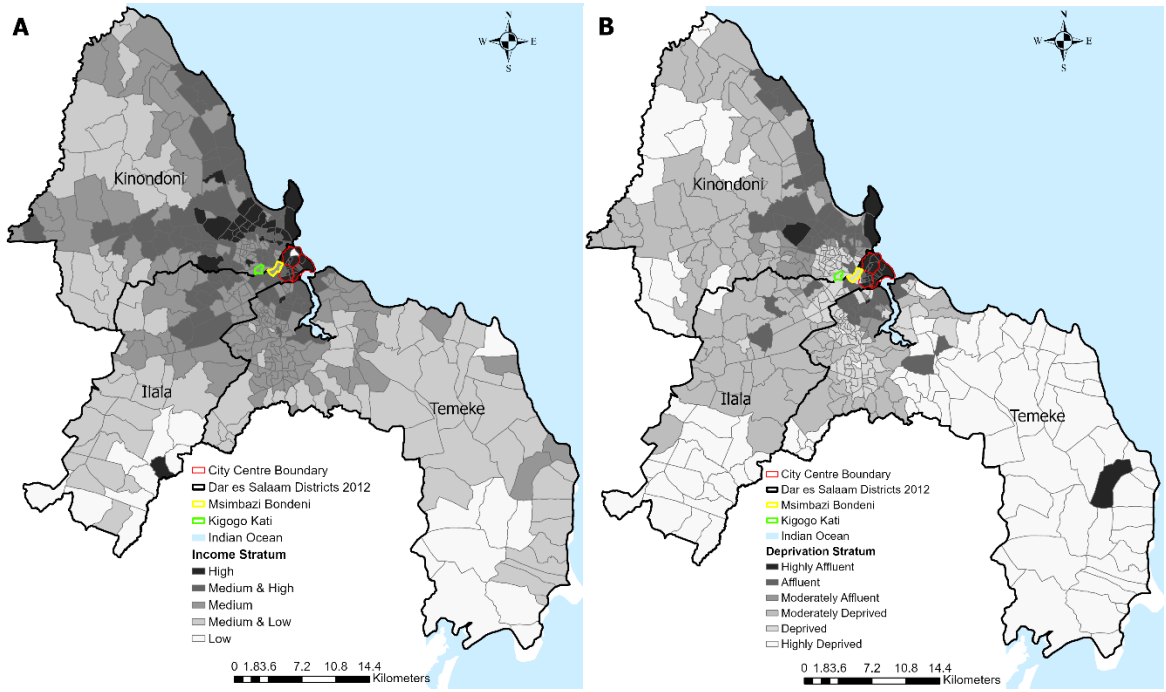


Fig 3: The spatial distribution of income and deprivation strata across Dar es Salaam city neighbourhoods. Panel A: Income strata obtained from the participatory stratification approach (PSA) (Table 2). Panel B: deprivation strata generated from a statistical latent profile analysis (LPA) (Table 1) of population census data.

Looking at the heterogeneous settlement patterns of these neighbourhoods sheds light on how they were classified differently by the two different methods (Supp Fig 9 and Supp Fig 10). For example, the varying density and quality of the buildings in Kigogo Kati illustrated in Supp Fig 9 shows that, while participants in the PSA stratification process considered the mixed housing qualities and density that are clearly visible in panel A to be a reasonable basis for classifying it as medium income, LPA averaged out the diversity existing at fine scale within the neighbourhood and consequently classified it simply as being deprived.

Additionally, while the example in Supp Fig 5 outlines the spatial distribution of different building qualities within the same neighbourhood, similar variability is clearly noticeable within some other neighbourhoods. For example, Supp Fig 10 illustrates part of the Msimbazi Bondeni neighbourhood, where two housing blocks with obviously different building qualities are found in the same neighbourhood, which has been classified as medium income by PSA but deprived by LPA. Moreover, these two blocks with obviously different mean housing density and quality lie at opposite sides of a steep topographic incline (Supp Fig 10 A and B), with the denser informal settlements with generally smaller, poor quality housing being found in the flood-prone valley to the north of it , while the carefully constructed apartment blocks and planned housing developments (Supp Fig 10 D and E) were all found approximately 20 meters uphill to the south. Indeed, as one walks southward along the transect depicted in Supp Fig 10 A and B, from the valley up the hill onward along a relatively flat plateau, a clearly noticeable shift is observed in building quality with elevation (Supp Fig 10, panel D and E).

Although there are some differences between the two approaches in the distribution patterns of income and deprivation across neighbourhoods, with very few being classified as purely low income by PSA (Table 2), the geographic distribution of poverty displayed by the two approaches (Table 1 and Table 2) nevertheless follow a similar pattern (Fig 3).

Discussion

This study contributes to a growing area of research demonstrating similarities and differences in the results obtained through quite different methodological approaches to neighbourhood stratification. The use of LPA has provided valuable insights into the classification of neighbourhoods, particularly in the assessment of deprivation profiles. The results demonstrate LPA's ability to identify statistically distinct discrete neighbourhood deprivation profiles using socioeconomic data, while highlighting its limitation in missing out the complex realities existing at a finer spatial scale within neighbourhoods. This, however, is just one facet of the broader challenge posed by profiling deprivation using aggregated data at the level of the neighbourhoods in which people live, especially in diversified cities like Dar es Salaam. Furthermore, through cross-validation of the two approaches, both statistically and spatially, the results of the PSA yielded very similar patterns of heterogeneity as the LPA, in terms of the distributions of variable values and neighbourhood classifications, respectively. Furthermore, PSA proved capable of elucidating several otherwise hidden factors relying on perceived and lived understanding of poverty, which were missed by LPA for understandable reasons (Supp Fig 9 and Supp Fig 10). Akin to many participatory approaches (Cook, 2018; Emmel, 2008; Israel et al., 2005; Lung-Amam and Dawkins, 2020; Meenar, 2017; Vaughn and Jacquez, 2020; Wallerstein, 2020) relying on the subjective perspectives of contributing participants, the PSA developed in this study, proved to be cost-effective and user-friendly. In contrast, while statistical analyses of formally collected data, such as the LPA approach

described herein, have the advantage of being objective, they are also time-consuming, data-intensive and require advanced analytical skills.

Nevertheless, the subjectivity of participatory approaches has been confirmed by the numbers of neighbourhoods attributed to each income stratum. While PSA classified only 4% of the city's neighbourhoods as typically low-income (Table 2), LPA classified 22% as highly deprived neighbourhoods (Table 1). On the other hand, PSA classified 44% of the city's neighbourhoods as medium income, indicating a bias away from extreme classification assignments. The surprisingly low frequency with which neighbourhoods were classified as low income by PSA participants implies the subjectivity of the approach, as even some small pockets of better housing in a neighbourhood known to be generally very poor often prompted participants to classify it as medium income.

Furthermore, these PSA results seem incompatible with those of previous studies indicating that large African cities scattered all across the continent accommodate large deprived populations (Kuffer, 2020; Thomson et al., 2020). Additionally, the observation that PSA classified some of the neighbourhoods as medium income while LPA classified them as deprived, requires consideration of the fundamentally different strengths and weakness of these two complementary approaches. A possible explanation for this apparent bias of PSA towards moderate stratum assignment may well lie in the diversity existing at fine scales within neighbourhoods. Such fine-scale heterogeneities can easily be missed by the objective statistical approaches but captured, and perhaps subjectively over-represented, by the qualitative understanding of the city obtained through the lived experiences of the PSA participants, wherein even very small differences in housing quality and infrastructures within neighbourhoods were considered in the stratification. For example, the small blocks of high-quality, low density housing that are highlighted in yellow bordered panels in Supp Fig 5 and 11 resulted in these neighbourhoods being classified as medium income, despite being well

known as poor neighbourhoods (Supp Fig 9 and Supp Fig 10). Also, the confirmatory stage of the consensus building process in the PSA workshop, resulted in participants providing supplementary justification for their initial income stratum assignment, taking into consideration additional factors that varied across coarser scales than income *per se*, such as physical accessibility, closeness to the valley, road condition, roof age and building density. These intuitive considerations, and ability of the PSA to accommodate them in a flexible manner, differentiates the fine-scale lens of PSA from the coarser analysis of the LPA, which unavoidably blurs such heterogeneities within neighbourhoods.

Such mixed contexts at fine scales might also help explain some of the more puzzling outliers in Fig 2, because the nuanced, intuitive and holistic basis of the PSA could result in it being subjectively influenced by the most obvious features and characteristics of a neighbourhood in a disproportionate manner, whereas the LPA relies on an objective representation of the area as a whole. Nevertheless, the two approaches yielded nearly statistically equivalent results, even though extensively surveyed semi-quantitative indicators of deprivation were used for the objective LPA, whereas intuitive qualitative knowledge of much smaller numbers of participants were solicited for the more subjective PSA. This study therefore provides an illustrative example of how these two alternative approaches can be applied to complement and cross-validate each other and in various scenarios with different application requirements for stratification (Weller et al., 2020).

This study study contributes to the growing area of research by demonstrating methodological similarity and differences existing in neighbourhood stratification. It complements the existing literature on economic inequality and urban segregation (Babere and Chingwele; Musterd, 2020; Smiley, 2007; Smiley, 2013; Todd et al., 2019; Vaughan and Arbaci, 2011; Villaça, 2011; Babere and Chingwele, 2018), which relies for the most part on statistical analysis of empirical data. Furthermore, it suggests that this more subjective participatory approach, which

exploits community knowledge of the location, lived experiences and local expertise, could be used in the future without any such data-intensive statistical approaches to validate it. Overall, this study provides clear evidence that cost-effective PSA methodologies could have a variety of practical applications, such as services planning and local tailoring of urban interventions.

While this research reassuringly presents neighbourhood-level income and deprivation profiles, respectively assessed through a subjective PSA and an objective LPA of formal socioeconomic data, some important study limitations are worth mentioning. For example, the certainty of the findings may be somewhat limited by the fact that, at the time of this study, there were no more recent census data available than those from the 2012 PHC, so the results obtained with the two methods cannot be considered exactly contemporaneous. Similarly, the retrospective *post hoc* examination of heterogeneities within neighbourhoods (Supp Fig 9 and Supp Fig 10) relies on images of contemporary Dar es Salaam, which has obviously changed somewhat since the time that the PSA was conducted and the PHC data used for LPA were collected.

Moreover, the absence of data relating to variables like employment, household incomes, neighbourhood incomes, property ownership, and living conditions (including housing, water, and electricity) restricted our ability to establish a conclusive relationship between income and living standards on one hand and deprivation on the other. These factors are traditionally utilized as indicators for assessing deprivation across various geographical and poverty dimensions. Consequently, future studies should explore alternative poverty indicators and incorporate updated data when accessible.

Additionally, the study used area-based data aggregated at neighbourhood level, which was the finest administrative scale made available to the investigators, so the stark heterogeneities that clearly exist within individual neighbourhoods (Supp Fig 9 and Supp Fig 10), are obviously

blurred around the neighbourhood average income by the LPA and cause underrepresentation of the lowest income neighbourhoods by PSA (Fig 2). Correspondingly, further research is needed to similarly compare subjective PSAs with objective statistical stratification based on formal survey data at finer scales, which in the context of Dar es Salaam, for example, could relate to ten cell unit housing clusters. Other avenues for future research can also include studies on how the built environment influences income and deprivation, and how the physical geography of the landscapes those human settlements are built upon affects patterns of income, access to services, health and wellbeing among residents.

Conclusion

The purpose of this study was to evaluate a PSA to stratifying neighbourhoods across the Tanzanian city of Dar es Salaam in terms of income, by comparing with LPA of deprivation based on national census data. The findings revealed the potential of both LPA and PSA methods and set a precedent for their further application as complementary approaches to neighbourhood classification. With reference to PSA, this study highlights ample room for advancing our understanding of the complex relationships between poverty, location, and socio-spatial dynamics in urban settings. To achieve a more comprehensive perspective, we recommend moving beyond reliance on relatively coarse geographic units for aggregating large-scale datasets. Instead, we propose adopting a finer-grained approach that can capture previously overlooked social, economic, physical, and environmental factors at the very fine scales they can vary across. Overall, such participatory methodologies may be valuable, not only for policy makers, but also for urban planning practitioners and researchers who seek a deeper understanding of the diverse dynamics occurring within towns and cities across various scales. These findings confirm that subjective PSA methodologies offer more affordable and practical alternatives to conventional objective statistical analyses of formal survey data, with

446 which to obtain detailed geographic information to inform planning and intervention targeting
447 in towns and cities all across low and middle-income countries.

References

Akogul S and Erisoglu M (2017) An Approach for Determining the Number of Clusters in a Model-Based Cluster Analysis. *Entropy* 19: 452.

Apel H (2020) Inequality in development: The 2030 Agenda, SDG 10 and the role of redistribution. *Real-World Econ. Rev* 92: 228-237.

Babere NJ and Chingwele I Neighbourhood Unit Residential Segregation in the Global South and Its Impact on Settlement Development: The Case of Dar es Salaam City.

Babere NJ and Chingwele I (2018) Neighbourhood Unit Residential Segregation in the Global South and Its Impact on Settlement Development : The Case of Dar es Salaam City.

Brennan JR (2007) Between segregation and gentrification: Africans, Indians, and the struggle for housing in Dar es Salaam, 1920-1950.

Brennan JR and Burton A (2007) The emerging metropolis: a short history of Dar es Salaam, circa 1862-2005.

Charles V, Gherman T and Paliza JC (2022) The Gini Index: A Modern Measure of Inequality. In: Charles V and Emrouznejad A (eds) *Modern Indices for International Economic Diplomacy*. Cham: Springer International Publishing, pp.55-84.

Cingano F (2014) Trends in Income Inequality and its Impact on Economic Growth.

Cook T (2018) Community-based participatory research for health: advancing social and health equity. *Educational Action Research* 27: 144 - 145.

Cornwall A and Jewkes R (1995) What is participatory research? *Social science & medicine* 41(12): 1667-1676.

DeVore GR (2017) Computing the Z Score and Centiles for Cross-sectional Analysis: A Practical Approach. *Journal of Ultrasound in Medicine* 36(3): 459-473.

Diener E and Suh E (1997a) Measuring quality of life: Economic, social, and subjective indicators. *Social indicators research* 40: 189-216.

Diener E and Suh E (1997b) MEASURING QUALITY OF LIFE: ECONOMIC, SOCIAL, AND SUBJECTIVE INDICATORS. *Social Indicators Research* 40(1): 189-216.

Ding C (2022) Latent Profile Analysis. *The SAGE Encyclopedia of Research Design*.

Dirksen J, Pinilla-Roncancio M, Wehrmeister FC, et al. (2022) Exploring the potential for a new measure of socioeconomic deprivation status to monitor health inequality. *International Journal for Equity in Health* 21(1): 56.

Drukker M and Os Jv (2003) Mediators of neighbourhood socioeconomic deprivation and quality of life. *Social Psychiatry and Psychiatric Epidemiology* 38: 698-706.

Emmel ND (2008) Participatory Mapping: An innovative sociological method.

Ferguson SL, G. Moore EW and Hull DM (2020) Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development* 44: 458 - 468.

Gogtay NJ and Thatte UM (2017) Principles of Correlation Analysis. *The Journal of the Association of Physicians of India* 65 3: 78-81.

Israel BA, Eng E, Schulz AJ, et al. (2005) Introduction to methods in community-based participatory research for health. *Methods in community-based participatory research for health* 3: 26.

Jungman FM (2013) The Price of Inequality: How Today's Divided Society Endangers Our Future. Joseph E. Stiglitz.

Koti F (2010) Confronting Sociospatial Exclusion on the Fringe of Africa's Cities Using Participatory GIS: Lessons from Athi River Town, Kenya. *Africa Today* 56: 62 - 82.

494 Kuffer M (2020) Digitalization and urban development in the Global South: Towards reliable
495 population data in deprived urban areas. *Österreichische Entwicklungspolitik*: 73-82.
496 Lamnisos D, Lambrianidou G and Middleton N (2019) Small-area socioeconomic deprivation
497 indices in Cyprus: development and association with premature mortality. *BMC Public*
498 *Health* 19(1): 627.
499 Little TD, Slegers DW and Card NA (2006) A Non-arbitrary Method of Identifying and Scaling
500 Latent Variables in SEM and MACS Models. *Structural Equation Modeling: A*
501 *Multidisciplinary Journal* 13(1): 59-72.
502 Lung-Amam WS and Dawkins C (2020) The power of participatory story mapping: Advancing
503 equitable development in disadvantaged neighbourhoods. *Community Development*
504 *Journal* 55(3): 473-495.
505 Magina FB, Kyessi A and Kombe WJ (2020) Benefits and Burdens Accrued in Community-
506 Led Informal
507 Settlements Regularisation in Tanzania: The Case of Magengeni in Dar Es Salaam
508 City. *Annual World Bank Conference.*
509 Mathew A and Doorenbos AZ (2022) Latent profile analysis – An emerging advanced
510 statistical approach to subgroup identification. *Indian Journal of Continuing Nursing*
511 *Education* 23: 127 - 133.
512 Meenar MR (2017) Using participatory and mixed-methods approaches in GIS to develop a
513 Place-Based Food Insecurity and Vulnerability Index. *Environment and Planning A*
514 49(5): 1181-1205.
515 Msuya I, Moshi I and Levira F (2021) Land Pattern of Highly Urbanizing Cities: Change in
516 Built-up Area, Population Density and Spatial Development of the Sprawling Dar es
517 Salaam City. *Environment and Urbanization ASIA* 12(1_suppl): S165-S182.
518 Musterd S (2020) Towards further understanding of urban segregation. *Handbook of urban*
519 *segregation*. Edward Elgar Publishing.
520 Muñoz I (2014) Piketty, Thomas. 2014. Capital in the Twenty-First Century The Belknap Press
521 of Harvard University Press, Cambridge.
522 Nijman JE and Wei YD (2020) Urban inequalities in the 21st century economy. *Applied*
523 *Geography (Sevenoaks, England)* 117: 102188 - 102188.
524 Oberski DL, van Kollenburg GH and Vermunt JK (2013) A Monte Carlo evaluation of three
525 methods to detect local dependence in binary data latent class models. *Advances in Data*
526 *Analysis and Classification* 7: 267-279.
527 Odoi A, Wray R, Emo M, et al. (2005) Inequalities in neighbourhood socioeconomic
528 characteristics: potential evidence-base for neighbourhood health planning.
529 *International Journal of Health Geographics* 4: 20 - 20.
530 Pastor DA, Barron KE, Miller BJ, et al. (2007) A latent profile analysis of college students'
531 achievement goal orientation. *Contemporary Educational Psychology* 32(1): 8-47.
532 Rasmussen MI (2013) The power of Informal Settlements . The Case of Dar Es Salaam ,
533 Tanzania.
534 Rosenberg J, Beymer P, Anderson D, et al. (2018) tidyLPA: An R Package to Easily Carry Out
535 Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. *Journal*
536 *of Open Source Software* 3(30): 978.
537 Rosenberg JM, Beymer PN, Anderson DJ, et al. (2019) tidyLPA: An R package to easily carry
538 out latent profile analysis (LPA) using open-source or commercial software. *Journal of*
539 *Open Source Software* 3(30): 978.
540 Schmidt MN, Seddig D, Davidov E, et al. (2021) Latent profile analysis of human values: What
541 is the optimal number of clusters?

Smiley SL (2007) *Patterns of urban life and urban segregation in Dar es Salaam, Tanzania.* University of Kansas.
Smiley SL (2013) Mental maps, segregation, and everyday life in Dar es Salaam, Tanzania. *Journal of Cultural Geography* 30(2): 215-244.
Tein J-Y, Coxe S and Cham H (2013) Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal* 20(4): 640-657.
Thomson DR, Kuffer M, Boo G, et al. (2020) Need for an integrated deprived area "slum" mapping system (IDEAMAPS) in low-and middle-income countries (LMICs). *Social Sciences* 9(5): 80.
Todd G, Msuya I, Levira F, et al. (2019) City Profile: Dar es Salaam, Tanzania. *Environment Urbanization ASIA* 10(2): 193-215.
Trinidad SA, Brokamp C, Mor Huertas A, et al. (2022) Use Of Area-Based Socioeconomic Deprivation Indices: A Scoping Review And Qualitative Analysis. *Health affairs* 41 12: 1804-1811.
UN (2020a) The World Social Report. Reportno. Report Number|, Date. Place Published|: Institution|.
UN (2020b) The World Social Report. Reportno. Report Number|, Date. Place Published|: Institution|.
UN-Habitat UNHSP (2020) World Cities Report 2020.
United-Nations (2016) The new urban agenda. Draft outcome document of the United Nations Conference on Housing and Sustainable Urban Development (Habitat III). *Conference on Housing and Sustainable Urban Development (Habitat III), Quito, 17–20 October 2016. Adoption of the final outcome of the Conference*.
URT and NBS (2022) The Initial Results of the Population and Housing Census 2022. Reportno. Report Number|, Date. Place Published|: Institution|.
Vaughan L and Arbaci S (2011) The challenges of understanding urban segregation. *Built Environment* 37(2): 128-138.
Vaughn LM and Jacquez FM (2020) Participatory Research Methods – Choice Points in the Research Process. *Journal of Participatory Research Methods*.
Villaça F (2011) São Paulo: urban segregation and inequality. *Estudos avançados* 25: 37-58.
Voisin DR, Kim DH and Hong JS (2018) A closer look at school bonding among African American adolescents in low-income communities: A latent class analysis. *Journal of health psychology* 23(11): 1424-1437.
Wallerstein N (2020) Commentary on Community-Based Participatory Research and Community Engaged Research in Health for Journal of Participatory Research Methods. *Journal of Participatory Research Methods*.
Weller BE, Bowen NK and Faubert SJ (2020) Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology* 46(4): 287-311.
Wish NB (1986) Are We Really Measuring the Quality of Life?Well-being Has Subjective Dimensions, As Well As Objective Ones. *The American Journal of Economics and Sociology* 45(1): 93-99.
Yalçin IK, Can N, Mançe Çalışır Ö, et al. (2021) Latent profile analysis of COVID-19 fear, depression, anxiety, stress, mindfulness, and resilience. *Current Psychology (New Brunswick, N.j.)* 41: 459 - 469.

588

Supplementary materials

Table S1: Equation used to derive census data and transform selected variables into proportion for each sample neighbourhood.

Category	Variable	Equation
Education	Proportion completed any school level	$= \left(\frac{CompleSch}{(CompleSch + AttendingSch + PartlyAttended + NeverAttended)}\right)$
	Proportion educated to university level	$= \left(\frac{UniversityOther}{(NurseryEdu + Primary\{1,8\} + TrainingAfterPri + PreFormOne + Secondary\{1,6\} + TrainingAfterSec + UniversityOther)}\right)$
	Proportion educated to secondary level	$= \left(\frac{Secondary\{1,6\}}{(NurseryEdu + Primary\{1,8\} + TrainingAfterPri + PreFormOne + Secondary\{1,6\} + TrainingAfterSec + UniversityOther)}\right)$
Literacy and language proficiency	Proportion of literacy	$= \left(\frac{(Total\ Population - Illiterate)}{Total\ Population\ total}\right)$
	Proportion of English literacy	$= \left(\frac{(English + SwahiliEnglish)}{(Swahili + English + SwahiliEnglish + Othelanguage)}\right)$
Economic factors	Proportion of economically active individuals	$= \left(\frac{\text{Nage}\{15,64\}}{Nage\{0,80+\}}\right)$
	Proportion with social security	$= \left(\frac{socialsec}{(socialsec + nosocialsec)}\right)$
Social and legal indicators	Proportion registered birth certificate	$= \left(\frac{BirthCertificate}{(BirthCertificate + BirthNotification + noCertNot + bircernotknow)}\right)$
Economic factors	Proportion household members in diaspora	$= \left(\frac{Diaspora}{Total\ Population}\right)$
Demographic factors	Average household size	$= \left(\frac{Total\ Population}{Total\ Household\ Heads}\right)$
	Proportion of female in a neighbourhood	$= \left(\frac{Females}{Total\ Population}\right)$
	Proportion of single head households	$= \left(\frac{(NeverMarried + Divorced + Separated\& + Widowed)}{(NeverMarried + Married + LivingTogether\& + Divorced + Separated + Widowed)}\right)$
Health and mortality statistics	Proportion with disability	$= \left(\frac{(Albino + Seeing + Hearing + Walking + Remembering + SelfCare + otherDisability)}{Total\ Population}\right)$
	Proportion of death caused by diseases	$= \left(\frac{Total\ Disease\ Death}{Total\ Death}\right)$
	Crude death rate	$= \left(1000 * \left(\frac{Total\ Death}{Total\ Population}\right)\right)$

592 Table S2: Population and neighbourhoods across districts in Dar es Salaam

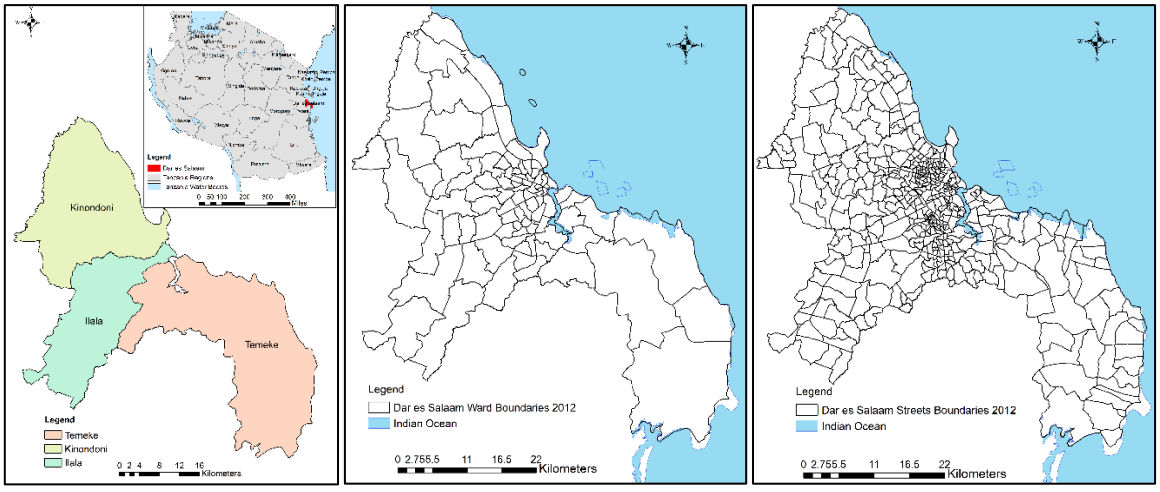
District	No. Neighbourhoods	% Neighbourhoods	Population	% Population	Area (sq. km)	Population Density (people/area sq. km)
Kinondoni	171	37.83	1,775,049	40.67	512.71	3462.07
Ilala	101	22.35	1,220,611	27.97	389.63	3132.72
Temeke	180	39.82	1,368,881	31.36	728.65	1878.64
Total	**452**	**100.00**	**4,364,541**	**100.00**		

593

594 Table S3: Comparison of model fit statistics from different models with different numbers of
595 profiles using the tidyLPA complex model structure 6 with varying variance and covariance
596 (Rosenberg et al., 2019), which were then evaluated using an Analytic Hierarchy Process
597 (AHP) (Akogul and Erisoglu, 2017) in the selection of the best fit model.

No. Profile in a Model	LogLik	AIC	ΔAIC	CLC	KIC	SABIC	Entropy
1	-6128	12527	N/A	12259	12665	12654	1
2	-4738	10017	2510	9477	10291	10272	0.96
3	-4486	9786	231	8974	10196	10169	0.95
4	-3945	8977	809	7893	9523	9487	0.97
5	-3677	8712	265	7356	9394	9350	0.98
6	**-3435**	**8500**	**212**	**6872**	**9318**	**9267**	**0.98**

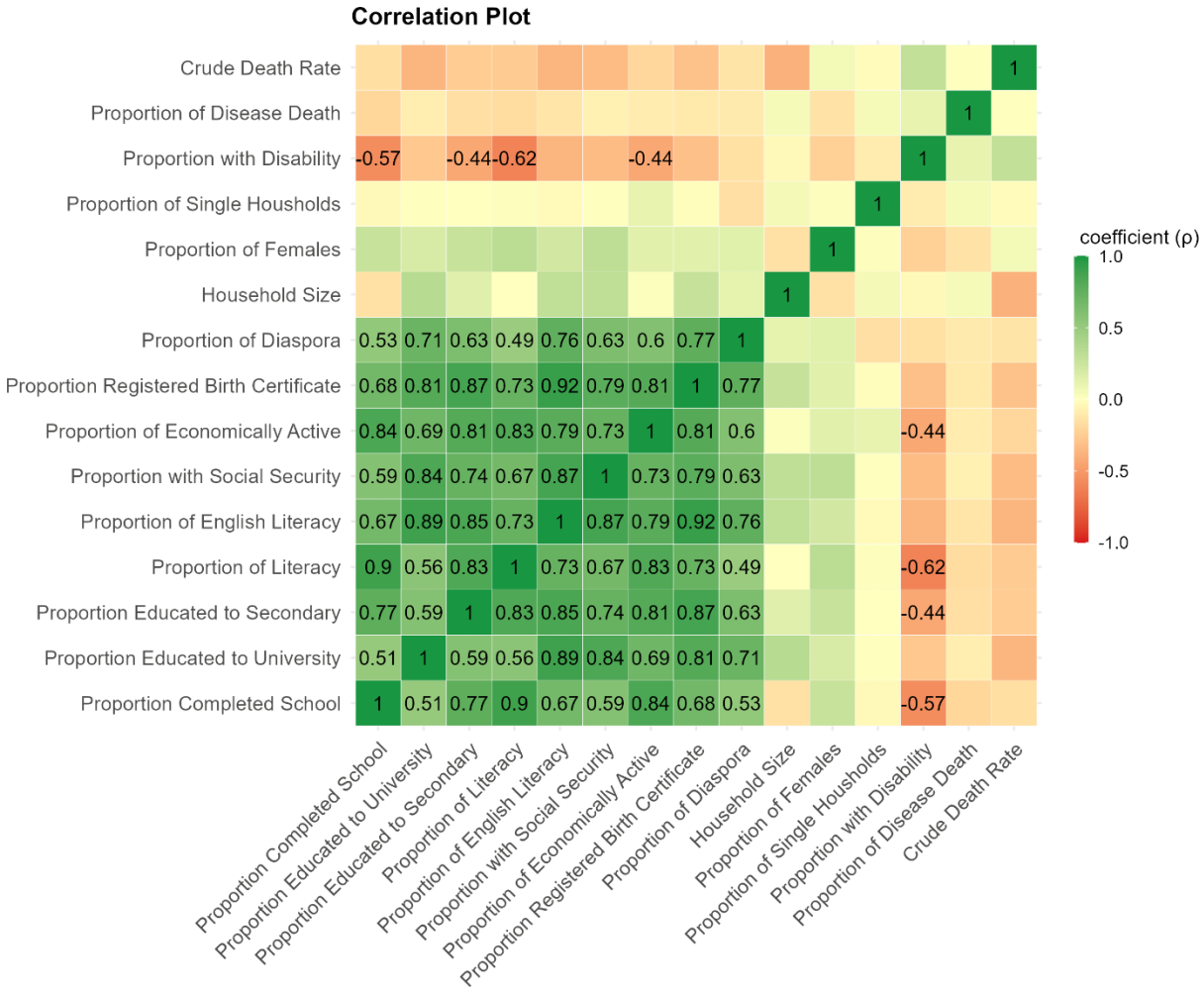
598



599

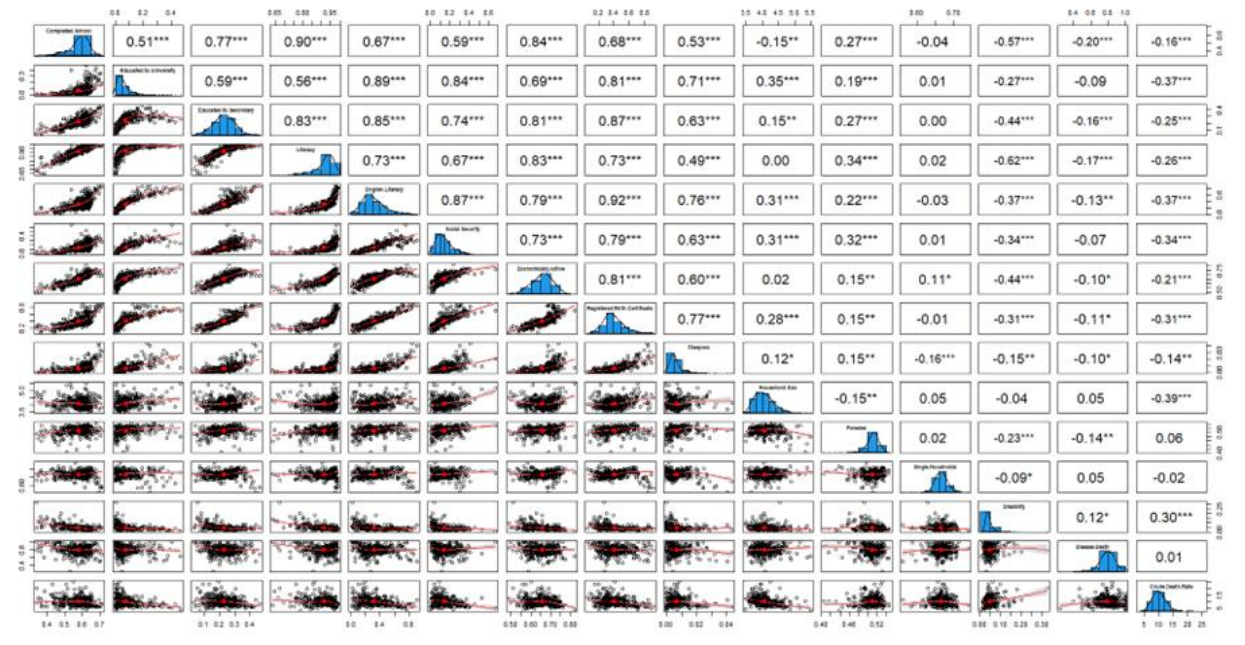
600 Figure S1: National and regional setting of Dar es Salaam city with district's boundaries in
601 2012 (A), wards boundaries (B), and neighbourhood boundaries, the finest administrative scale
602 which could be accessed and used in the study (C).

603



604

605 Figure S2: Pearson correlation plot with correlation coefficients and p-values, showing
606 numerical correlation coefficient labels only for pairs which are strongly correlated ($\rho \geq +/-$
607 0.4) and statistically significant ($p \leq 0.05$) while insignificant variables are unlabelled.



608

609 Figure S3: Pairwise scatter plots of all the variables as well as histograms, locally smoothed
610 regressions, and Pairwise Pearson's correlation test matrix with correlation coefficients (ρ) and
611 p values (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). The x axis in each scatter plot represents the
612 column variable, the y axis the row variable. The units for most of these variables are
613 proportions of participating respondents or households, aggregated at neighbourhood level, the
614 only two exceptions being household size and crude death rate.



Figure S4: Images showing participants and activities conducted during the participatory neighbourhood stratification workshop. Panel A and B show individual participants independently assigning neighbourhoods to one of five different income strata and sharing their assignation using sticky pad notes. Panels C and D show how the assignations of individual participants were collated and compared as basis for group discussion, by assembling such sticky notes on a flip chart. Panels E and F illustrate how the investigators then facilitated collective consensus building among the participants, based on group discussions to compare and contrast their different neighbourhood assignments to specific strata, and then reach final agreement for each neighbourhood.

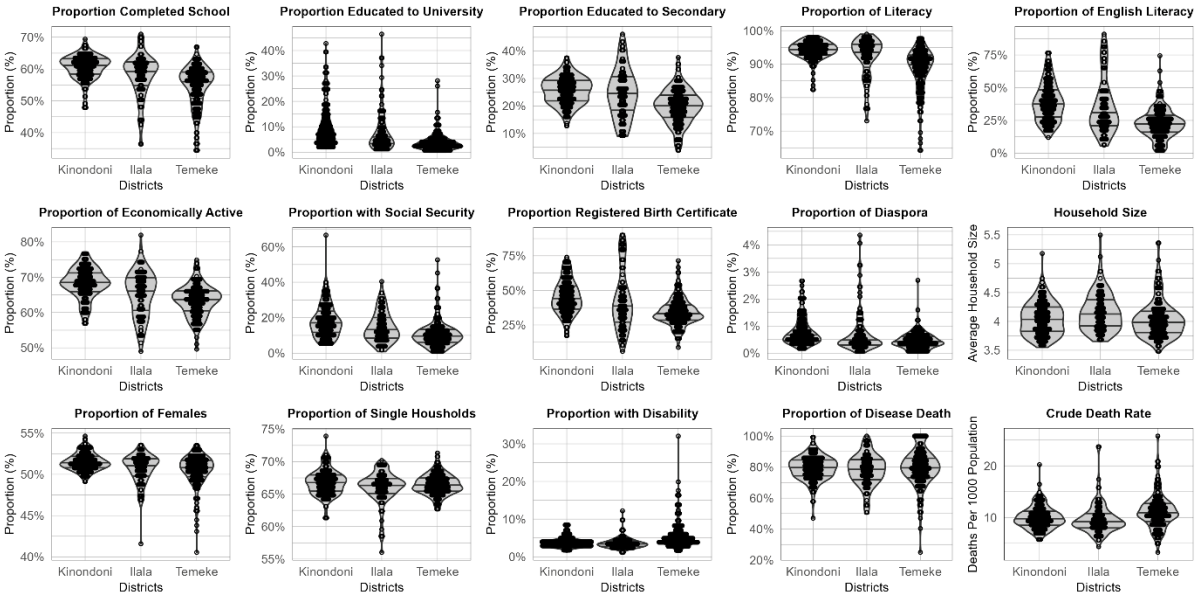


Figure S5: Violin and dot plot showing the descriptive statistics of 15 potential indicators of deprivation derived from the 2012 national census data, distributed across three study districts at the time. Horizontal lines within the violin plots represent the first, second and third quartiles, or the 25%, 50% (median), and 75% quantiles. The units for most of these variables are proportions of participating respondents or households, aggregated at neighbourhood level, the only two exceptions being household size and crude death rate.

Using the selected 15 variables from the neighbourhoods, LPA was used to profile neighbourhoods into statistically distinct strata.

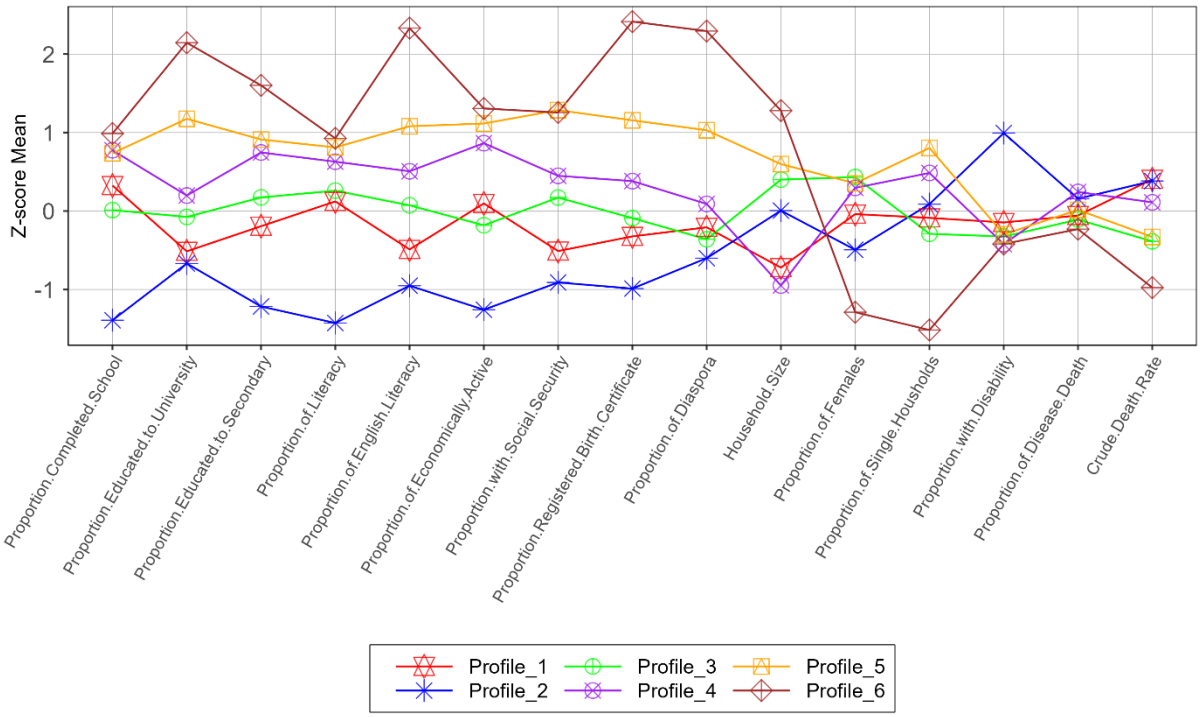
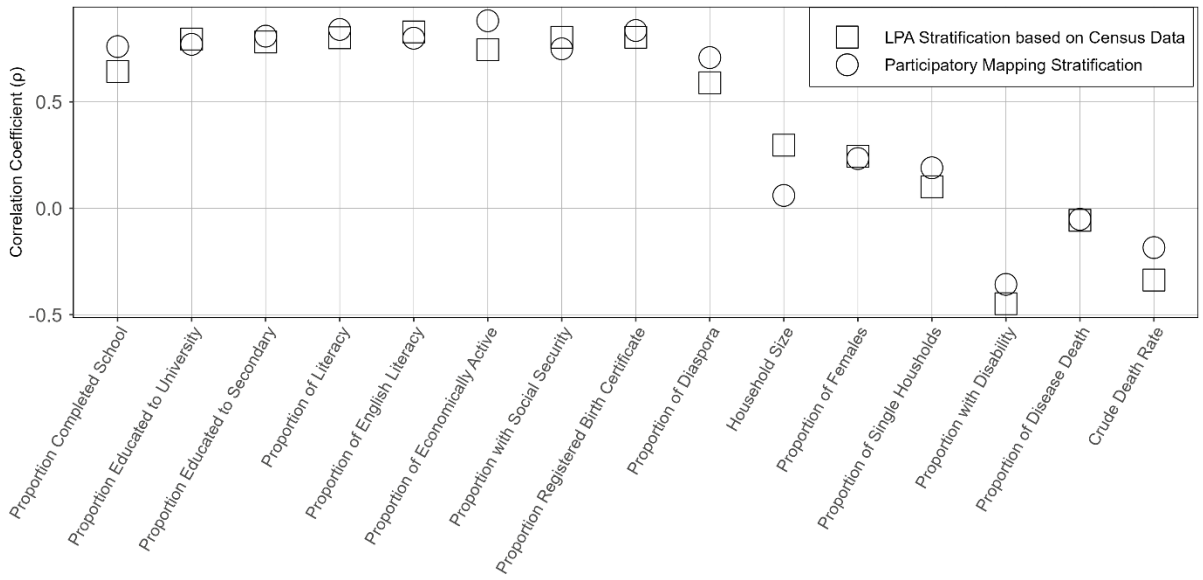


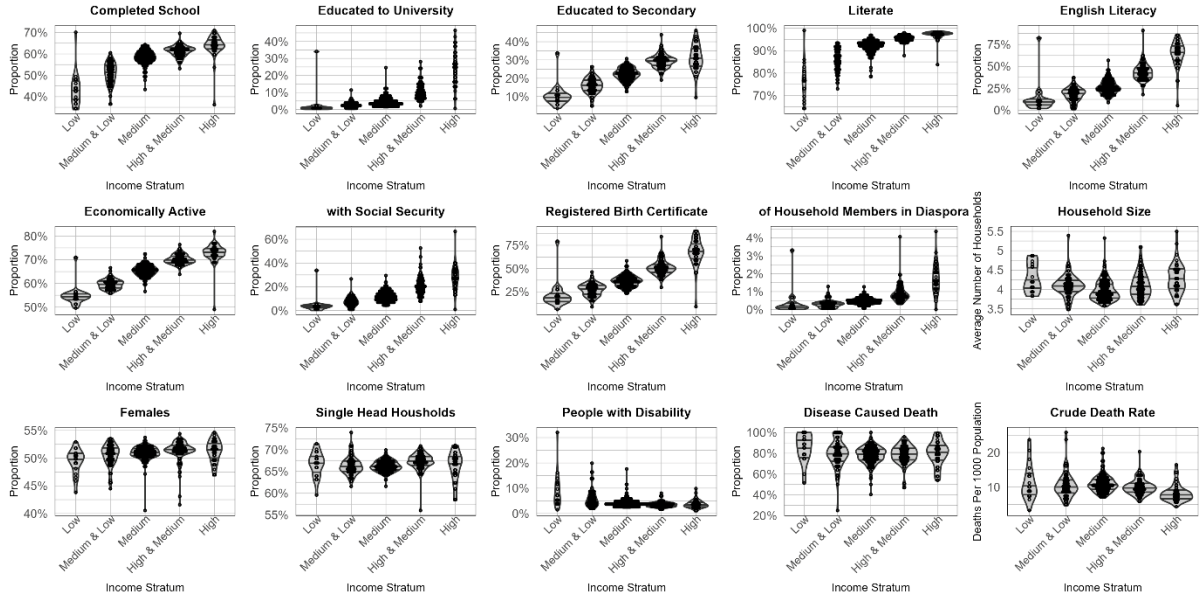
Figure S6: Line and dot plot illustrating the distribution of calculated z-score means for each contributing variable for each profile identified by latent profile analysis (LPA) and subsequently attributed to a named stratum of deprivation (Table 1). Means were scaled by standardizing observed variables (z-scores) within each latent profile and expressing them in standard deviation units instead of the original metric. Scaling transforms the latent variable to have mean of zero and standard deviation of one ($\mu = 0$, $\sigma = 1$). Negative means do not imply negative values of the actual metric, but rather reflecting normalization around zero by the scaling process.



645

646 **Figure S7: Dot plot showing the results of the Spearman's rank correlation test of the**
647 **census variables with each of the two stratification approaches, namely the participatory**
648 **stratification approach (PSA), and the latent profile analysis (LPA), with vertical axis**
649 **showing correlation coefficient (ρ) of each variable with the stratum ranking for each**
650 **stratification method.**

651



652

Figure S8: Violin and dot plot showing the distribution of values for 15 potential indicators of deprivation recorded by the 2012 national census, broken down across the five income strata obtained from the participatory stratification approach (PSA). Horizontal lines within the violin plot represent the first, second and third quartiles, or the 25%, 50% (median), and 75% quantiles. The units for most of these variables are proportions of participating respondents' households aggregated at neighbourhood level, the only two exceptions being household size and crude death rate.



Figure S9: Satellite and building images from the Kigogo Kati neighbourhood showing fine-scale variations in housing density and building quality within the neighbourhood. Kigogo Kati was classified as medium income by participatory stratification approach (PSA) (Fig 3A), but as deprived by statistical latent profile analysis (LPA) of national census data (Fig 3B). As mapped out in panel A, blocks of high-quality buildings (yellow bordered panels, B, C, F, K, M, N and P), are contrasted with blocks with generally poor-quality housing (blue bordered panels D, E, G, H, I, J and L) in the same neighbourhood.

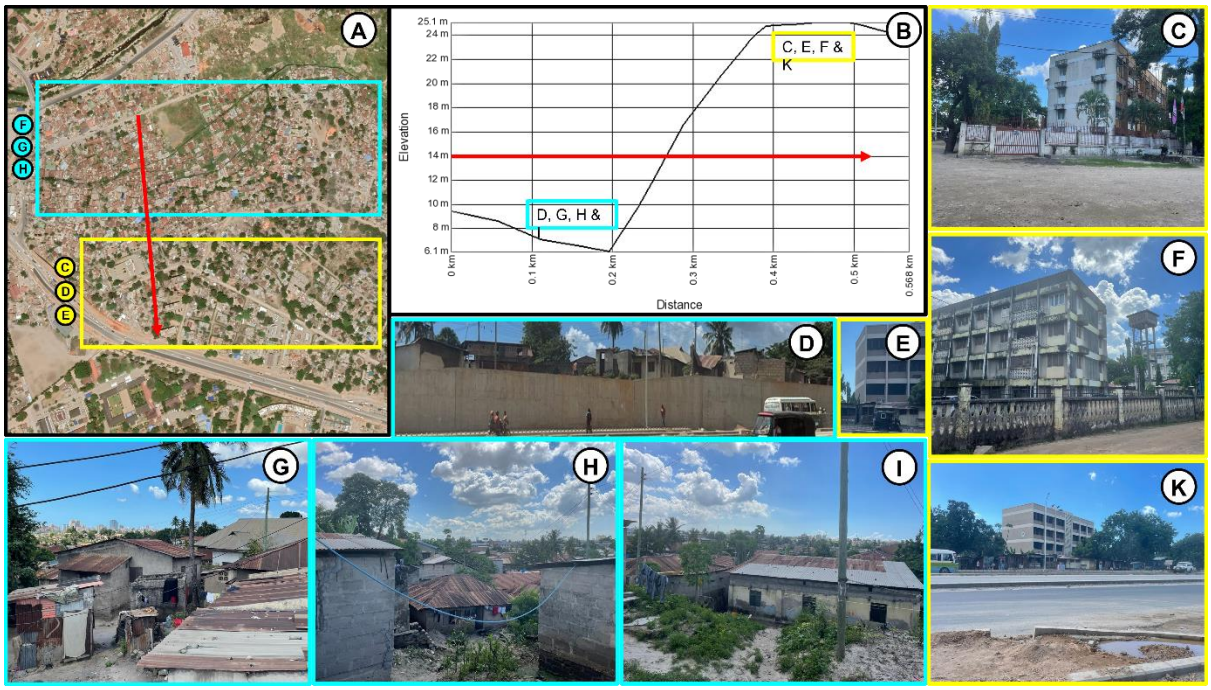


Figure S10: Satellite and building images from the Msimbazi Bondeni neighbourhood, showing fine scale variation in housing density and building quality within the neighbourhood. Msimbazi Bondeni was classified as medium income by participatory stratification approach (PSA) (Fig 3A) but as deprived by statistical latent profile analysis (LPA) of national census data (Fig 3B). As mapped out in Panel A, blocks of high-quality buildings (yellow bordered panels, C, E, F and K) uphill and to the south of a steep topographic gradient (Panel B) are contrasted with downhill blocks to the south where high-density settlements have generally poor-quality housing (blue bordered panels D, G, H and I). Panels A and B illustrate how building quality changes with elevation, with the vertical axis in panel B projecting elevation above sea level, while the horizontal axis represents the distance southward along the red vector in panel A.

Box S1. Summary of the conceptual and methodological distinctions between *income inequality* and *socioeconomic deprivation*.

Income inequality: The concept of *income inequality* specifically targets the unequal allocation of income among individuals or households within a given population. Surveyed indices of income inequality express the disparity between the earnings of the wealthiest and the poorest members of society in semi-quantitative terms, often employing statistical summary measures, such as the Gini coefficient or income quintiles (Charles et al., 2022).

Socioeconomic deprivation: On the other hand, *socioeconomic deprivation* refers to disadvantages that a person or group face in an inequitable manner (Trinidad et al., 2022). Such disadvantages can be linked to geographic factors, such as living in a certain neighbourhood or community, or other factors like housing and built environment quality. These disadvantages can take various forms, such as limited access to job opportunities, education, healthcare, and other basic needs. Essentially, it describes the relative absence of power that certain individuals or groups have over their economic and social circumstances (Drukker and Os, 2003). This can perpetuate a cycle of poverty, marginalization, and exclusion, leading to increased social inequality and decreased opportunities for upward mobility (Lamnisos et al., 2019). Semi-quantitative measures of socioeconomic deprivation often involves use multidimensional indices to account for these diverse factors affecting well-being (Charles et al., 2022).

Box S2. Summary of the relevant evidence base and methodological considerations that informed the formulation of the study design.

Participatory Stratification Approaches (PSA): Participatory stratification approaches involve active, collaborative engagement and involvement of participants who are residents of the study area in question (Emmel, 2008; Koti, 2010). These participants may sometimes be influential or authoritative stakeholders, in the research investigation or decision-making process. PSA can be used to understand socioeconomic, geographical inequalities and the hierarchical arrangement of social groups within a given society or system, and is especially useful where socio-economic classes are distributed very unevenly among people within a population (Cornwall and Jewkes, 1995; Israel et al., 2005), forcing people with comparable socio-economic characteristics to reside in similar spatial localities (UN, 2020a). However, it is essential to recognise the subjective nature of PSA, which can be influenced by relative income assessments within the social circles or communities of individual participants, drawing from lived experiences and personal knowledge (Diener and Suh, 1997b; Wish, 1986; Diener and Suh, 1997a). Nonetheless, such subjective impressions of individual members within a community can yield qualitative insights that complement objective quantitative measurements, thus enriching an understanding of the broader economic landscape within cities (Muñoz, 2014; Jungman, 2013).

Latent Profile Analysis (LPA): A more objective way to understand such socioeconomic strata within in urban spatial and administrative subunits, is through statistical analysis of more structured empirical data from surveys of the major drivers and consequences of such heterogeneities. Specifically, LPA is a statistical method used to identify unobserved subgroups, or latent class or profiles within a sample of observed entities, based on similar patterns observed variables (Oberski et al., 2013). LPA has been sometimes referred to as an

advanced model-based clustering approach, due to its ability to use different fit statistics to select the best profile, class, or cluster. Over the last decade, LPA has become more widely applied as an alternative to previously existing methods for identifying clusters (eg. K-means clustering, Hierarchical clustering, Fuzzy C-means clustering, and Partitioning Around Medoids), due its flexibility in model criterion selection, such as variance and covariance estimation, and the objectivity of its best fit model selection procedures (Mathew and Doorenbos, 2022; Rosenberg et al., 2019; Ding, 2022; Schmidt et al., 2021). LPA of detailed data from the 2012 national Population and Housing Census (PHC) for Tanzania was therefore adopted as the gold standard objective statistical methodology, with which the PSA was compared and contrasted as a means to assess the advantages and disadvantages of the latter relative to the former.

684