

1 **Evaluation of a participatory approach to stratifying**
2 **neighbourhoods across the Tanzanian city of Dar es**
3 **Salaam in terms of income by comparing with latent**
4 **profile analysis of deprivation based on national census**
5 **data.**

6 Ibrahim Msuya^{1,3}, Martin Boudou², Francis Levira¹, Irene Moshi¹, Jean O'Dwyer^{3,4}, Gerard
7 Killeen^{3,4}

8 ¹Department of Health Systems, Impact Evaluation and Policy, Ifakara Health Institute, Dar
9 es Salaam, Tanzania

10 ²Spatiotemporal Environmental Epidemiology Research (STEER) Group, Environmental
11 Sustainability and Health Institute, Technological University Dublin, Dublin, Ireland.

12 ³School of Biological, Earth and Environmental Sciences, University College Cork, Cork,
13 Ireland.

14 ⁴Environmental Research Institute, University College Cork, Cork, Ireland.

15 *Corresponding author

16 E-mail: imsuya@ihi.or.tz, imsuya@ucc.ie (IM)

17

18 **Abstract**

19 Studying geographic heterogeneities in the characteristics of city neighbourhoods, such as
20 population income and deprivation, has been a common practice in urban health studies for
21 tailored and targeted interventions, especially in highly developed and diversified cities.
22 While a range of different approaches has been applied to classify different parts of towns
23 and cities, participatory stratification approaches (PSAs) have become popular despite their
24 subjective basis because of their affordability, simplicity and practicality, all of which allow
25 them to be frequently updated. While more objective statistical approaches, such as latent
26 profile analysis (LPA), can also be used to stratify neighbourhoods using formal socio-
27 economic and demographic data, these rely on the availability of rich datasets and advanced
28 analytical capacities that are not always available in low and middle-income countries. This
29 study assessed a PSA to stratify neighbourhoods across the Tanzania city of Dar es Salaam in
30 terms of income, by comparing it with a complementary LPA using national census data from
31 2012 to stratify them in terms of deprivation. A consultative community-based workshop was
32 used for the PSA, while 15 selected deprivation indicators from the census data were used to
33 profile them using LPA. While the PSA allocated neighbourhoods to five income strata, six
34 clear deprivation strata could be distinguished by LPA. A strong positive correlation was
35 observed between the stratum identified by the LPA and that obtained through the PSA ($\rho =$
36 0.739 , $p < 0.0001$). Furthermore, paired comparison of the two sets of correlation coefficients
37 between each deprivation indicator and the stratum assigned by each stratification approach
38 revealed no difference ($V = 33$, $p\text{-value} = 0.1354$), confirming that the two approaches
39 yielded very similar patterns of stratification., Also, the two approaches yielded broadly
40 comparable cartographic pictures of the city, depicting similar spatial distribution of wealth
41 and poverty. Overall, this evidence indicates that subjective community knowledge and lived

- 42 experience may be invaluable for understanding the built environment and for mapping out
- 43 pockets of poverty and affluence at fine scales with minimum resources.

44 **Introduction**

45 More than half of the world's population now lives in towns and cities [1]. While Africa and
46 eastern and southern Asia have been urbanized most rapidly, the pace of urbanization in these
47 regions is expected to slow over the coming decade [2]. Meanwhile, limited government
48 investment in urban planning has contributed to more extensive and overcrowded informal
49 settlements, where approximately 47% of the world's 4.3 billion urban dwellers live [2]. In
50 Tanzania, for instance, it has been estimated that more than 75% of settlements have been
51 informally developed [3]. Indeed, Dar es Salaam is the largest city in Tanzania, where >75%
52 of the population live in informal settlements with highly heterogeneous wealth distribution
53 [4]. Several historical studies on the early development of what is now the old city centre of
54 Dar es Salaam have rationalized the distribution of settlement patterns in terms of colonial
55 planning processes, which were heavily influenced by residential segregation policies. During
56 the colonial era, the city was planned in a way that segregated people by race and income,
57 with affluent neighbourhoods located near the city centre and along the Indian Ocean
58 (Europeans and Indians) and the less affluent near and around the city centre (black Africans,
59 referred to in that era as natives) (Brennan & Burton, 2007). The evident pattern in the city
60 demonstrates that a significant portion of its development has unfolded through informal
61 processes [5]. Informal settlements tend to be situated either within formal areas or near
62 them, aligning with the contextual reality of the city's historical evolution. This historical
63 development trajectory has contributed to an uneven distribution of neighbourhood attributes
64 in terms of income and affluence, a disparity that persists today. However, despite the
65 inclination to attribute these disparities solely to colonial influences, a more thorough
66 investigation challenges this prevailing narrative. Contrary to the common assumption that
67 the current state of Dar es Salaam, beyond the old city, is a direct result of colonial planning

68 [6], in fact, it emerged from limited planning, especially during the city's rapid expansion
69 from the 1970's onward [7]

70 While both formal and informal settlements co-exist side-by-side in cities, they exhibit
71 varying neighbourhood characteristics. Urban neighbourhoods constantly evolve at very fine
72 scales, resulting in a myriad of spatial, sociodemographic, economic, and cultural diversities
73 coexisting within relatively small, but nevertheless, very heterogeneous geographic areas.
74 This vibrant evolutionary process is particularly dynamic in unplanned, informal settlements,
75 catalyzing different types of disparities, such as health inequality, social division, exclusion,
76 segregation, which are typically assessed in terms of income inequality or socioeconomic
77 deprivation, [8] . Despite the conceptual differences in how income and deprivation are used
78 to assess the two-sided coin of poverty and wealth (Box 1), either may be just as readily
79 applied to surveying the distribution of inequalities across urban areas.

Box 1. 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 [9].

Socioeconomic deprivation: On the other hand, *socioeconomic deprivation* refers to disadvantages that a person or group face in an inequitable manner [10]. 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 [11]. This can perpetuate a cycle of poverty, marginalization, and exclusion, leading to increased social inequality and decreased opportunities for upward mobility [12]. Semi-quantitative measures of socioeconomic deprivation often involves use multidimensional indices to account for these diverse factors affecting well-being [9].

80

81 For the first time within the context of internationally agreed development goals, the 2030
82 Agenda aims to reduce inequality based on income [13]. While efforts are being made to
83 bridge the gap between rich and poor [14, 15], there is still insufficient investment in
84 methodologies for socioeconomic profiling of urban neighbourhoods [16]. In response to that
85 methodological need, this study set out to evaluate a subjective participatory stratification
86 approach (PSA) to classifying neighbourhoods across the Tanzanian city of Dar es Salaam in
87 terms of income, by comparing it with an objective latent profile analysis (LPA) of
88 deprivation based on national census data (Box 2).

Box 2. 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 [17, 18]. 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 [19, 20], forcing people with comparable socio-economic characteristics to reside in similar spatial localities [21]. 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 [22-24]. 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 [25, 26].

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 [27]. 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 [28-31]. 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.

89

90 **Methods**

91 **Ethics statement**

92 This research was conducted as part of Centre for Sustainable, Healthy, and Learning Cities
93 and Neighbourhoods, supported by funding from UK Research and Innovation (UKRI)
94 through the Economic and Social Research Council (ESRC) under the UK Government's

95 Global Challenges Research Fund (GCRF) with project reference (ES/P011020/1). Ethical
96 clearance for the study, a component of the Sustainable Cities project at Ifakara Health
97 Institute (IHI), was granted by the Institutional Review Board of the Ifakara Health Institute
98 (IHI/IRB/NO: 26-2018) and the Medical Research Coordinating Committee at the National
99 Institute for Medical Research in Tanzania (NIMR/HQ/R.8a/Vol. IX/2954). Verbal consent,
100 in accordance with the approved protocols, was obtained from all human participants.

101 **Request for Access to Restricted Materials**

102 Certain materials, such as photographs featuring individuals or identifying body parts, have
103 been omitted from this publication (e.g., Figure 2, panels A, C, D, E, F) in accordance with
104 editorial guidelines. We recognise the significance of these materials to your understanding
105 of the study. Please contact the corresponding author for assistance if you require access to
106 these materials.

107 **Geographical context and study data**

108 Dar es Salaam is the largest city in Tanzania, located on the Indian Ocean coast (Fig 1). The
109 city had a population of approximately 4.4 million in 2012 but currently has approximately
110 5.4 million inhabitants, distributed across an administrative area that has remained constant at
111 1,590 km² [32]. It is also the port in Tanzania, contributing substantially to the national
112 economy and served as a government administrative centre before *de facto* relocation to the
113 official capital, Dodoma, was practically implemented in 2016 [33].

114

115

INSERT FIG 1A, 1B & 1C HERE.

116 **Fig 1: National and regional setting of Dar es Salaam city with district's boundaries in**
117 **2012 (A), wards boundaries (B), and neighbourhood boundaries, the finest**
118 **administrative scale which could be accessed and used in the study (C).**

119

120 The study reported herein used Tanzania's 2012 Population and Housing Census (PHC) data
121 from the National Bureau of Statistics of Tanzania (NBS) to generate variables which were
122 considered likely to represent deprivation in the study context (S1 Table). The data is
123 presented via administrative units, starting from regions down to districts, wards (*kata*) and
124 neighbourhoods (*mitaa*, literally translated as *streets* but actually means a larger, formalized
125 administrative area with defined boundaries, rather than just a road and structures built along
126 it). The data from PHC primarily consisted of socio-demographic information such as gender,
127 age, disability, marital status, citizenship, place of residence, education status, migration
128 patterns, economic activities, fertility rates, general and maternal mortality rates, and
129 enrolment in social security programs. Other data include household with members who are
130 living in diaspora, food and cash crops cultivation, and information about asset ownership
131 and housing conditions. The current study used data aggregated at the neighbourhood level
132 (Fig 1C), the finest scale freely available to the public upon request for research purposes to
133 derive deprivation indicators based relevant variables in the PHC (S1 Table). The dataset
134 included 452 neighbourhoods from three districts that comprised Dar es Salaam at the time
135 (Fig 1A)

136 Neighbourhoods in Dar es Salaam city are unevenly distributed, with Kinondoni District
137 having the highest number of neighbourhoods, while Ilala has the lowest number of
138 neighbourhoods (Table 1). In terms of population, Kinondoni also has the highest share of
139 residents living in this district, while Ilala is the least. However, from a historical perspective,
140 Ilala stands as the most well-established district, comprising neighbourhoods with a rich
141 legacy that dates back over a century. On the other hand, most parts of Temeke were
142 relatively new and were experiencing rapid growth and development at the time.

143 Table 1: 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		

144

145 Study design

146 Based on considerations of the relevant methodological literature (Box 2), the current study
147 sought to assess a subjective participatory stratification approach (PSA) to classifying
148 neighbourhoods in terms of income level by comparing it with an objective statistical
149 approach using latent profile analysis (LPA), to stratify neighbourhoods in terms of
150 deprivation based on socioeconomic data from the 2012 PHC in the Tanzania city of Dar es
151 Salaam. While the income which is referred by the PSA is inherently subjective and based on
152 the lived experiences and knowledge of the city's residents, the LPA of deprivation strata in
153 quantitative terms is a more objective socioeconomic metric derived from formal ways of
154 living standards, conditions, and education at the neighbourhood level [34].

155 **Participatory stratification approach (PSA)**

156 The PSA involved active participation of community members and professionals in the
157 stratification of Dar es Salaam neighbourhoods into different income strata. This one-day
158 workshop, carried out on Friday, 22nd March 2019, from 09:00 AM to 17:00 PM at Ifakara
159 Health Institute offices in Mikocheni, Dar es Salaam, involved 50 people, specifically five
160 residents from each district recommended by the neighbourhood leaders, two town planners
161 from each district, three community development officers from each district (Table 1), five
162 practicing town planners from the private sector, two urban and regional planning lecturers
163 from Ardhi University and 13 Ifakara Health Institute scientists working in urban projects in
164 Dar es Salaam. The task started by collectively agreeing on the number of income strata to
165 which the city's neighbourhoods should be assigned. Then, each of the 452 neighbourhoods
166 were assigned to a respective stratum by each individual participant, following which a
167 consensus was established through group discussion to collectively agree on the stratum it
168 was finally assigned to (Fig 2). Although the task fundamentally relied on the participants
169 knowledge and lived experience of the city, some other complementary sources of
170 information, such as satellite images, were also used to facilitate the process of informed
171 debate and consensus building.

172 **INSERT FIG 2 HERE**

173 **Fig 2: Images showing participants and activities conducted during the participatory**
174 **neighbourhood stratification workshop.** Panel A and B show individual participants
175 independently assigning neighbourhoods to one of five different income strata and sharing
176 their assignment using sticky pad notes. Panels C and D show how the assignments of
177 individual participants were collated and compared as basis for group discussion, by
178 assembling such sticky notes on a flip chart. Panels E and F illustrate how the investigators

179 then facilitated collective consensus building among the participants, based on group
180 discussions to compare and contrast their different neighbourhood assignments to specific
181 strata, and then reach final agreement for each neighbourhood.

182 **Latent profile analysis (LPA)**

183 Deprivation profiles were identified as unobserved variables (latent profiles), which were
184 distinguished based on the observed values for relevant socioeconomic variables in the
185 neighbourhood-level PHC data. The first step of the LPA was the identification of the most
186 appropriate recorded variables. Based on the data and context, 15 variables in the PHC data
187 that were intuitively to expected to be closely associated with deprivation were identified.
188 These included education indicators such as; *proportion completed any school level*,
189 *proportion educated to university level*, *proportion educated to secondary level*, literacy and
190 language proficiency; *proportion of literate*, *proportion of English literacy*, economic factors;
191 *proportion of potentially economically active individuals*, *proportion enrolled in social*
192 *security*, *proportion household with members in diaspora*, social and legal indicators;
193 *proportion with registered birth certificate*, demographic factors; *average household size*,
194 *proportion of female*, *proportion of single head households*, and health and mortality
195 statistics; *proportion with disability*, *proportion of death caused by diseases*, and *crude death*
196 *rate*. Different indicators from the census data were used to derive these variables (S1 Table)
197 for participating respondents or households, aggregated at neighbourhood level. To provide a
198 standardised way to compare different quantities while allowing for more accurate and
199 meaningful comparisons between variables in predicting outcomes or estimate probabilities,
200 almost all variables were transformed into proportions, except for household size and crude
201 death rate which were presented in their actual metrics units.

202 Descriptive statistics of the selected variables can be found in Fig 3, which shows prior
203 expected general distribution of income and poverty across three study districts at the time,
204 which are consistent with the actual existing situation in the city.

205 **INSERT FIG 3 HERE**

206 **Fig 3: Violin and dot plot showing the descriptive statistics of 15 potential indicators of**
207 **deprivation derived from the 2012 national census data, distributed across three study**
208 **districts at the time.** Horizontal lines within the violin plots represent the first, second and
209 third quartiles, or the 25%, 50% (median), and 75% quantiles. The units for most of these
210 variables are proportions of participating respondents or households, aggregated at
211 neighbourhood level, the only two exceptions being household size and crude death rate.

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

214 Depending on the nature of data, different six LPA models structures can be used which
215 range from simple to complex, determined by how variance and covariance are treated and
216 whether relationships between variables are considered, thus affecting model parsimony and
217 complexity [35]. In model selection process, Pearson's rank correlation test was employed to
218 quantify the level of correlation to inform optional specification of model assumption and
219 maximize model performance. In addition, the correlations between variables can help with
220 the interpretation of the resulting latent profiles, as such, if two variables exhibit high level of
221 correlation it can suggest that these two variables may be driving the occurrence of profile
222 and provide insights into the nature of the profile. Broadly speaking, a correlation coefficient
223 $\rho \geq \pm 0.4$ and p value ≤ 0.05 was considered as an indication of strong correlation [36]. The
224 Pearson's rank correlation analysis revealed that a significant correlation existed among ten
225 out of the fifteen variables ($p \leq 0.05$), (S2 Fig and S3 Fig). Hence, from the six tidyLPA
226 models structures proposed by Rosenberg *et al.* [29], model structure six with varying
227 variances and covariances was selected. It is the most complex model, so while it allows for
228 the estimation of variances and covariances across profiles without constraints, potentially
229 providing a deeper understanding of the relationships between the variables, it is also less
230 parsimonious.

231 The LPA modelling was performed using R version 4.2.0 and RStudio version
232 2023.03.0+386 with the *tidyLPA* package. *tidyLPA* acts as an interface to the *mclust* package
233 and provides the functionalities to estimate the commonly specified models used in the
234 analysis, with different parameters depending on model complexity according to the
235 structural forms developed by [35]. Seven criteria were used for best fit model selection:

236 Log-Likelihood (LogLik), Akaike Information Criterion (AIC), Difference in AIC values
237 (Δ AIC), Consistent Akaike Information Criterion (CLC), Kullback-Leibler Information
238 Criterion (KIC), Sample-Size Adjusted Bayesian Information Criterion (SABIC), and
239 Entropy (Average Uncertainty or Randomness in Latent Class Assignments). Across the
240 existing LPA literature, it has been widely reported that selection of the best fit model with
241 the most informative number of distinct profiles remains a challenging task [37].
242 Correspondingly, to validate the model optimization based on the fit statistics, an analytic
243 hierarchy process (AHP), as proposed by [38] which helps in determining the correct number
244 of profiles in a model based clustering was used.

245 Two approaches were used to interpret and assign meaning to each profile from the selected
246 best fit model in relation to deprivation. Firstly, the distribution of the variables was plotted
247 against every profile in a dot and violin plot. Secondly, using z-score or scaled mean of each
248 variable on each profile. While median centre data around the second quartiles or 50%
249 quartiles, z score means centre data around zero because the scaling involves standardizing
250 each observed variable to have mean of zero and standard deviation of one ($\mu = 0, \sigma = 1$),
251 [39]. In the context of LPA, scaled means are often reported as part of the model output along
252 with other fit statistics. Often, scaled means may have negative value due to scaling even
253 though latent variable has only positive values, however, this does not necessarily indicate
254 that the underlying construct is negative [40]. This helps in comparing the relative positions
255 of the means across variables with different scales or variances, as well as for interpreting the
256 magnitude and direction of the differences between the means of different profiles.

257 In LPA, after numbering profiles by ranking them according to their order of score, it is
258 always a task to the investigators to interpret and assign their intuitive meaning [41]. The
259 ranked deprivation profiles were then each interpreted and assigned to named ranked

260 deprivation stratum. The LPA deprivation strata were then compared to the results from the
261 PSA using a non-parametric statistical correlation test (Spearman's rank correlation test), to
262 assess the strength and direction of the relationship between the ranked strata from the two
263 approaches. Further analyses were performed to cross validate the PSA income strata by
264 plotting values using violin and dot plot for the same PHC variables used in broken down by
265 PSA stratum. To test statistical significance difference between the two approaches, a
266 Wilcoxon signed-rank test, was applied using to compare the correlation coefficient of the
267 Spearman's correlation (ρ) obtained by PSA and LPA.

268 Additionally, using the ranked results from the two approaches, neighbourhood income and
269 deprivation strata from PSA and LPA, respectively, were mapped using ArcGIS Pro 3.1.2, to
270 understand and compare stratification of the city neighbourhoods by the two approaches in
271 terms of comparative spatial distribution. In spatial comparative analysis, ground truthing
272 evidence of the existing conditions in the selected neighbourhoods, specifically, housing
273 density and quality of buildings, using in situ ground photographs and satellite images, were
274 used to provide further descriptive insights into the meaning of the results.

275 Results

276 Latent profile identification

277 The best fit profile model identified by LPA of the census data was with 6-profiles: The
278 goodness of fit consistently decreased as the number of profiles modelled increased, while
279 entropy increased with a 98% classification quality reached for this final optimal 6-profile
280 model that was the last one to converge (Table 2). These model optimization results were
281 confirmed using the AHP approach [38], which also indicated that the model with 6 profiles
282 was optimal based on the fit statistics in Table 2. The results, therefore, suggested that six
283 clear deprivation strata could be identified based on census data, as compared to the five
284 income strata obtained from the participatory stratification approach.

285 Table 2: Comparison of model fit statistics from different models with different numbers of
286 profiles using the tidyLPA complex model structure 6 with varying variance and covariance
287 [29], which were then evaluated using an Analytic Hierarchy Process (AHP) [38] in the
288 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

289

290 Interpreting the underlying meaning of identified latent profiles.

291 From the dot and violin plots in Fig , the value distributions of the most important individual
292 indicators of deprivation show that ranking the numbered LPA profiles based on these same
293 variables, yields an order of 2, 1, 3, 4, 5 and 6 for the profile numbers assigned by the LPA

294 best fit model (Table 2). Correspondingly, these were respectively assigned the following
295 stratum names; *Highly Deprived*, *Deprived*, *Moderately Deprived*, *Moderately Affluent*,
296 *Affluent*, and *Highly Affluent* (Table 3). To further examine and query these profile
297 assignments based on variable value distributions, a calculated scaled mean for each variable
298 was used (Fig 5), consistent with other LPA studies that assigned interpreted profile
299 meanings [28, 42-44]. These scaled means for each variable within each stratum confirmed
300 the same apparent profile order (Table 3) for most variables, as ranked from the most
301 deprived to the most affluent (Fig 5).

302

INSERT FIG 4 HERE

303 **Fig 4: Violin and dot plot showing the distribution of values for 15 potential indicators**
304 **of deprivation recorded by the 2012 national census, broken down across the six profiles**
305 **distinguished by LPA analysis (Table 2).** Horizontal lines within the violin plot represent
306 the first, second and third quartiles, or the 25%, 50% (median), 75% quantiles. The units for
307 most of these variables are proportions of participating respondents or households,
308 aggregated at neighbourhood level, the only two exceptions being household size and crude
309 death rate.

310

311

INSERT FIG 5 HERE

312 **Fig 5: Line and dot plot illustrating the distribution of calculated z-score means for each**
313 **contributing variable for each profile identified by latent profile analysis (LPA) and**
314 **subsequently attributed to a named stratum of deprivation (Table 3). Means were scaled**
315 **by standardizing observed variables (z-scores) within each latent profile and expressing them**
316 **in standard deviation units instead of the original metric. Scaling transforms the latent**
317 **variable to have mean of zero and standard deviation of one ($\mu = 0, \sigma = 1$). Negative means**
318 **do not imply negative values of the actual metric, but rather reflecting normalization around**
319 **zero by the scaling process.**

320 Based on the above interpretation of the LPA identified profiles, six assigned deprivation
 321 strata could be resolved from the census data (Table 3). On the other hand, however, the
 322 results from participatory stratification approach revealed five income classes (Table 4) with
 323 *low income stratum* accounting for a remarkably low percentage of neighbourhoods while the
 324 *medium income stratum* had the highest.

325 **Table 3: Latent profile analysis (LPA) assigned profiles numbers with investigator**
 326 **assigned ranked deprivation stratum names.** Based on interpretation of Fig and Fig 5,
 327 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

328

329 Table 4: Neighbourhood income classes obtained from participatory stratification approach
 330 (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

331

332 **Comparison of PSA with LPA of the census data**

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

336

INSERT FIG 6 HERE

337 **Fig 6: Jitter plot showing the relationship between the neighbourhood classification**
338 **obtained through the participatory stratification approach (PSA), and through**
339 **statistical latent profile analysis (LPA) stratification based on national census data.**

340 The same variables used to stratify the neighbourhoods by LPA yielded distribution of values
341 that were just as consistent with their income stratum as determined by the PSA (7) as they
342 were with their deprivation stratum as determined by the LPA (Fig and Fig 5).

343

INSERT FIG 7 HERE

344 **Fig 7: Violin and dot plot showing the distribution of values for 15 potential indicators**
345 **of deprivation recorded by the 2012 national census, broken down across the five**
346 **income strata obtained from the participatory stratification approach (PSA).** Horizontal
347 lines within the violin plot represent the first, second and third quartiles, or the 25%, 50%
348 (median), and 75% quantiles. The units for most of these variables are proportions of
349 participating respondents' households aggregated at neighbourhood level, the only two
350 exceptions being household size and crude death rate.

351 Further analysis of Spearman's rank correlation tests between each variable from the census
352 data and stratum rank as determined by each of the two different stratification approaches can
353 be seen in 8. A Wilcoxon signed-rank test to compare these two sets of correlation
354 coefficients in a paired manner revealed no difference between the two approaches ($V = 33$,
355 $p\text{-value} = 0.1354$), indicating that the two approaches yielded very similar patterns of
356 stratification.

357 **INSERT FIG 8 HERE**

358 **Fig 8: Dot plot showing the results of the Spearman's rank correlation test of the census**
359 **variables with each of the two stratification approaches, namely the participatory**
360 **stratification approach (PSA), and the latent profile analysis (LPA), with vertical axis**
361 **showing correlation coefficient (ρ) of each variable with the stratum ranking for each**
362 **stratification method.**

363

364 Overall, the two stratification approaches yielded relatively similar cartographic pictures of
365 wealth and poverty distribution across the city (Fig). High income, highly affluent
366 neighbourhoods are mainly located in or around the city centre and along the coast to the
367 north, while low income, deprived neighbourhoods are generally distributed along the
368 outskirts of the city. The distribution pattern of the medium income stratum, as determined by
369 PSA, corresponds well with the moderately deprived and moderately affluent strata identified
370 by LPA. Focusing on the city centre and areas immediately around it, the results reveal
371 patches of low income and deprived neighbourhoods located near high income and highly
372 affluent neighbourhoods, forming a pocket of poverty running from central Ilala into Temeke
373 District. However, in the western direction this pocket of deprived, medium income

374 neighbourhoods is enclosed by moderately deprived, mixed medium and high-income
375 neighbourhoods.

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

381 **INSERT FIG 9A & 9B HERE.**

382 **Fig 9: The spatial distribution of income and deprivation strata across Dar es Salaam**
383 **city neighbourhoods.** Panel A: Income strata obtained from the participatory stratification
384 approach (PSA) (Table 4). Panel B: deprivation strata generated from a statistical latent
385 profile analysis (LPA) (Table 3) of population census data.

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

393

INSERT FIG 10 HERE

394 **Fig 4: Satellite and building images from the Kigogo Kati neighbourhood showing fine-**
395 **scale variations in housing density and building quality within the neighbourhood.**

396 Kigogo Kati was classified as medium income by participatory stratification approach (PSA)
397 (Fig A), but as deprived by statistical latent profile analysis (LPA) of national census data
398 (Fig B). As mapped out in panel A, blocks of high-quality buildings (yellow bordered panels,
399 B, C, F, K, M, N and P), are contrasted with blocks with generally poor quality housing (blue
400 bordered panels D, E, G, H, I, J and L) in the same neighbourhood.

401

402 Additionally, while the example in Fig outlines the spatial distribution of different building
403 qualities within the same neighbourhood, similar variability is clearly noticeable within some
404 other neighbourhoods. For example, Fig illustrates part of the Msimbazi Bondeni
405 neighbourhood, where two housing blocks with obviously different building qualities are
406 found in the same neighbourhood, which has been classified as medium income by PSA but
407 deprived by LPA. Moreover, these two blocks with obviously different mean housing density
408 and quality lie at opposite sides of a steep topographic incline (Fig A and B), with the denser
409 informal settlements with generally smaller, poor quality housing being found in the flood-

410 prone valley to the north of it , while the carefully constructed apartment blocks and planned
411 housing developments (Fig D and E) were all found approximately 20 meters uphill to the
412 south. Indeed, as one walks southward along the transect depicted in Fig 7A and B, from the
413 valley up the hill onward along a relatively flat plateau, a clearly noticeable shift is observed
414 in building quality with elevation (Fig , panel D and E).

415

INSERT FIG 11 HERE

416 **Fig 5: Satellite and building images from the Msimbazi Bondeni neighbourhood,**
417 **showing fine scale variation in housing density and building quality within the**
418 **neighbourhood.** Msimbazi Bondeni was classified as medium income by participatory
419 stratification approach (PSA) (Fig A) but as deprived by statistical latent profile analysis
420 (LPA) of national census data (Fig B). As mapped out in Panel A, blocks of high-quality
421 buildings (yellow bordered panels, C, E, F and K) uphill and to the south of a steep
422 topographic gradient (Panel B) are contrasted with downhill blocks to the south where high-
423 density settlements have generally poor-quality housing (blue bordered panels D, G, H and I).
424 Panels A and B illustrate how building quality changes with elevation, with the vertical axis
425 in panel B projecting elevation above sea level, while the horizontal axis represents the
426 distance southward along the red vector in panel A.

427 Although there are some differences between the two approaches in the distribution patterns
428 of income and deprivation across neighbourhoods, with very few being classified as purely
429 low income by PSA (Table 4), the geographic distribution of poverty displayed by the two
430 approaches (Table 3 and Table 4) nevertheless follow a similar pattern (Fig).

431

432

433 Discussion

434 This study contributes to a growing area of research demonstrating similarities and
435 differences in the results obtained through quite different methodological approaches to
436 neighbourhood stratification. The use of LPA has provided valuable insights into the
437 classification of neighbourhoods, particularly in the assessment of deprivation profiles. The
438 results demonstrate LPA's ability to identify statistically distinct discrete neighbourhood
439 deprivation profiles using socioeconomic data, while highlighting its limitation in missing out
440 the complex realities existing at a finer spatial scales within neighbourhoods. This, however,
441 is just one facet of the broader challenge posed by profiling deprivation using aggregated data
442 at the level of the neighbourhoods in which people live, especially in diversified cities like
443 Dar es Salaam. Furthermore, through cross-validation of the two approaches, both
444 statistically and spatially, the results of the PSA yielded very similar patterns of heterogeneity
445 as the LPA, in terms of the distributions of variable values and neighbourhood classifications,
446 respectively. Furthermore, PSA proved capable of elucidating several hidden factors relying
447 on perceived and lived understanding of poverty, which were easily missed by LPA (**Fig 0**
448 **&Fig 1**). Akin to many participatory approaches [17, 20, 45-49] relying on the subjective
449 perspectives of contributing participants, the PSA developed in this study, proved to be cost-
450 effective and user-friendly. In contrast, while statistical analyses of formally collected data,
451 such as the LPA approach described herein, have the advantage of being objective, they are
452 also time-consuming, data-intensive and require advanced analytical skills.

453 Nevertheless, the subjectivity of participatory approaches has been confirmed by the
454 numbers of neighbourhoods attributed to each income stratum. While PSA classified only 4%
455 of the city's neighbourhoods as typically low-income (Table 4), LPA classified 22% as
456 highly deprived neighbourhoods (**Table 3**). On the other hand, PSA classified 44% of the

457 city's neighbourhoods as medium income, indicating a bias away from extreme classification
458 assignments. The surprisingly low frequency with which neighbourhoods were classified as
459 low income by PSA participants implies the subjectivity of the approach, as even some small
460 pockets of better housing in a neighbourhood known to be generally very poor often
461 prompted participants to classify it as medium income.

462 Furthermore, these PSA results seem incompatible with those of previous studies indicating
463 widespread that large African cities accommodate large deprived populations[50, 51]. .

464 Additionally, the observation that PSA classified some of the neighbourhoods as medium
465 income while LPA classified them as deprived, requires consideration of the fundamentally
466 different strengths and weakness of these two complementary approaches. A possible
467 explanation for this apparent bias of PSA towards moderate stratum assignment may well lie
468 in the diversity existing at fine scales within neighbourhoods. Such fine-scale heterogeneities
469 can easily be missed by the objective statistical approaches but captured, and perhaps
470 subjectively over-represented, by the qualitative understanding of the city obtained through
471 the lived experiences of the PSA participants, wherein even very small differences in housing
472 quality and infrastructures within neighbourhoods were considered in the stratification. For
473 example, the small blocks of high quality, low density housing that are highlighted in yellow
474 bordered panels in Figs 10 and 11 resulted in these neighbourhoods being classified as
475 medium income, despite being well known as poor neighbourhoods (Fig and Fig). Also, the
476 confirmatory stage of the consensus building process in the PSA workshop, resulted in
477 participants providing supplementary justification for their initial income stratum assignment,
478 taking into consideration additional factors that varied across coarser scales than income *per*
479 *se*, such as physical accessibility, closeness to the valley, road condition, roof age and
480 building density. These intuitive considerations, and ability of the PSA to accommodate

481 them in a flexible manner, differentiates the fine-scale lens of PSA from the coarser analysis
482 of the LPA, which unavoidably blurs such heterogeneities within neighbourhoods.

483 Such mixed contexts at fine scales might also help explain some of the more puzzling
484 outliers in Fig , because the nuanced, intuitive and holistic basis of the PSA could result in it
485 being subjectively influenced by the most obvious features and characteristics of a
486 neighbourhood in a disproportionate manner, whereas the LPA relies on an objective
487 representation of the area as a whole. Nevertheless, the two approaches yielded statistically
488 equivalent results, even though an intuitive conceptual rationale was used in selecting the
489 quantitative deprivation indicators used for LPA, while qualitative participants knowledge
490 were used in PSA. Hence, the two alternative approach can be applied to complement each
491 other and in different scenarios with different application requirements for stratifying
492 neighbourhoods [52].

493 This study study contributes to the growing area of research by demonstrating
494 methodological similarity and differences existing in neighbourhood stratification. It
495 complements the existing literature on economic inequality and urban segregation [33, 53-
496 59], which relies for the most part on statistical analysis of empirical data. Furthermore, it
497 suggests that this more subjective participatory approach, which exploits community
498 knowledge of the location, lived experiences and local expertise, could be used in the future
499 without any such data-intensive statistical approaches to validate it. Overall, this study
500 provides clear evidence that cost-effective PSA methodologies could have a variety of
501 practical applications, such as services planning and local tailoring of urban interventions.

502 While this research reassuringly presents neighbourhood-level income and deprivation
503 profiles, respectively assessed through a subjective PSA and an objective LPA of formal
504 socioeconomic data, some important study limitations are worth mentioning. For example,

505 the certainty of the findings may be somewhat limited by the fact that, at the time of this
506 study, there were no more recent census data available than those from the 2012 PHC, so the
507 results obtained with the two methods cannot be considered exactly contemporaneous.
508 Similarly, the retrospective *post hoc* examination of heterogeneities within neighbourhoods
509 (**Fig** and **Fig**) relies on images of contemporary Dar es Salaam, which has obviously
510 changed somewhat since the time that the PSA was conducted and the PHC data used for
511 LPA were collected.

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

519 Additionally, the study used area-based data aggregated at neighbourhood level, which was
520 the finest administrative scale made available to the investigators, so the stark heterogeneities
521 that clearly exist within individual neighbourhoods (**Fig 0** and **Fig 1**), are obviously blurred
522 around the neighbourhood average income by the LPA and cause underrepresentation of the
523 lowest income neighbourhoods by PSA (**Fig**). Correspondingly, further research is needed to
524 similarly compare subjective PSAs with objective statistical stratification based on formal
525 survey data at finer scales, which in the context of Dar es Salaam, for example, could relate to
526 ten cell unit housing clusters. Other avenues for future research can also include studies on
527 how the built environment influences income and deprivation, and how the physical
528 geography of the landscapes those human settlements are built upon affects patterns of
529 income, access to services, health and wellbeing among residents.

530 **Conclusion**

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

548 **Acknowledgements**

549 We gratefully acknowledge the support and contributions that made this research possible.
550 Special thanks to Prof. Ya Ping Wang and Keith Kintrea for their invaluable advice. We also
551 thank Dr. Victoria Mwakalinga for leading the study participants team and all participants
552 from DFID, DAWASA, Ardhi University, Humanitarian Open street Team, Ifakara Health
553 Institute, Ministry of Lands Housing and Human Settlements, Regional Administrative

554 Offices, TARURA, Twaweza and all five Dar es Salaam municipalities by then for their
555 participation in the neighbourhood stratification. This work was made possible by the
556 collective efforts of all those mentioned, and we appreciate their essential roles in the
557 successful completion of this study.

558

559

560 **References**

- 561 1. United-Nations. The new urban agenda. Draft outcome document of the United
562 Nations Conference on Housing and Sustainable Urban Development (Habitat III).
563 Conference on Housing and Sustainable Urban Development (Habitat III), Quito, 17–20
564 October 2016 Adoption of the final outcome of the Conference. 2016.
- 565 2. UN-Habitat UNHSP. World Cities Report 2020. 2020.
- 566 3. Magina FB, Kyessi A, Kombe WJ, editors. Benefits and Burdens Accrued in
567 Community-Led Informal Settlements Regularisation in Tanzania: The Case of Magengeni in Dar Es Salaam
568 City. Annual World Bank Conference; 2020.
- 569 4. Rasmussen MI, editor The power of Informal Settlements . The Case of Dar Es
570 Salaam , Tanzania2013.
- 571 5. Msuya I, Moshi I, Levira F. Land Pattern of Highly Urbanizing Cities: Change in
572 Built-up Area, Population Density and Spatial Development of the Sprawling Dar es Salaam
573 City. *Environment and Urbanization ASIA*. 2021;12(1_suppl):S165-S82.
- 574 6. Brennan JR, editor Between segregation and gentrification: Africans, Indians, and the
575 struggle for housing in Dar es Salaam, 1920-19502007.
- 576 7. Brennan JR, Burton A, editors. The emerging metropolis: a short history of Dar es
577 Salaam, circa 1862-20052007.
- 578 8. Odoi A, Wray R, Emo M, Birch S, Hutchison B, Eyles J, et al. Inequalities in
579 neighbourhood socioeconomic characteristics: potential evidence-base for neighbourhood
580 health planning. *International Journal of Health Geographics*. 2005;4:20 -
- 581 9. Charles V, Gherman T, Paliza JC. The Gini Index: A Modern Measure of Inequality.
582 In: Charles V, Emrouznejad A, editors. *Modern Indices for International Economic
583 Diplomacy*. Cham: Springer International Publishing; 2022. p. 55-84.
- 584 10. Trinidad SA, Brokamp C, Mor Huertas A, Beck AF, Riley C, Rasnik E, et al. Use Of
585 Area-Based Socioeconomic Deprivation Indices: A Scoping Review And Qualitative
586 Analysis. *Health affairs*. 2022;41 12:1804-11.
- 587 11. Drukker M, Os Jv. Mediators of neighbourhood socioeconomic deprivation and
588 quality of life. *Social Psychiatry and Psychiatric Epidemiology*. 2003;38:698-706.
- 589 12. Lamnissos D, Lambrianidou G, Middleton N. Small-area socioeconomic deprivation
590 indices in Cyprus: development and association with premature mortality. *BMC Public
591 Health*. 2019;19(1):627. doi: 10.1186/s12889-019-6973-0.
- 592 13. UN. The World Social Report. United Nations publication, 2020.
- 593 14. Cingano F. Trends in Income Inequality and its Impact on Economic Growth. 2014.
594 doi: [doi:https://doi.org/10.1787/5jxrjncwxv6j-en](https://doi.org/10.1787/5jxrjncwxv6j-en).
- 595 15. Apel H. Inequality in development: The 2030 Agenda, SDG 10 and the role of
596 redistribution. *Real-World Econ Rev*. 2020;92:228-37.
- 597 16. Nijman JE, Wei YD. Urban inequalities in the 21st century economy. *Applied
598 Geography (Sevenoaks, England)*. 2020;117:102188 -
- 599 17. Emmel ND, editor *Participatory Mapping: An innovative sociological method*2008.
- 600 18. Koti F. Confronting Sociospatial Exclusion on the Fringe of Africa’s Cities Using
601 Participatory GIS: Lessons from Athi River Town, Kenya. *Africa Today*. 2010;56:62 - 82.
- 602 19. Cornwall A, Jewkes R. What is participatory research? *Social science & medicine*.
603 1995;41(12):1667-76.
- 604

- 605 20. Israel BA, Eng E, Schulz AJ, Parker EA. Introduction to methods in community-
606 based participatory research for health. *Methods in community-based participatory research*
607 for health. 2005;3:26.
- 608 21. UN. The World Social Report. United Nations publication, 2020.
- 609 22. Diener E, Suh E. Measuring Quality of Life: Economic, Social, And Subjective
610 Indicators. *Social Indicators Research*. 1997;40(1):189-216. doi: 10.1023/A:1006859511756.
- 611 23. Wish NB. Are We Really Measuring the Quality of Life? Well-being Has Subjective
612 Dimensions, As Well As Objective Ones. *The American Journal of Economics and*
613 *Sociology*. 1986;45(1):93-9. doi: <https://doi.org/10.1111/j.1536-7150.1986.tb01906.x>.
- 614 24. Diener E, Suh E. Measuring quality of life: Economic, social, and subjective
615 indicators. *Social indicators research*. 1997;40:189-216.
- 616 25. Muñoz I, editor Piketty, Thomas. 2014. *Capital in the Twenty-First Century* The
617 Belknap Press of Harvard University Press, Cambridge 2014.
- 618 26. Jungman FM, editor *The Price of Inequality: How Today's Divided Society*
619 *Endangers Our Future*. Joseph E. Stiglitz 2013.
- 620 27. Oberski DL, van Kollenburg GH, Vermunt JK. A Monte Carlo evaluation of three
621 methods to detect local dependence in binary data latent class models. *Advances in Data*
622 *Analysis and Classification*. 2013;7:267-79.
- 623 28. Mathew A, Doorenbos AZ. Latent profile analysis – An emerging advanced statistical
624 approach to subgroup identification. *Indian Journal of Continuing Nursing Education*.
625 2022;23:127 - 33.
- 626 29. Rosenberg JM, Beymer PN, Anderson DJ, Van Lissa C, Schmidt JA. tidyLPA: An R
627 package to easily carry out latent profile analysis (LPA) using open-source or commercial
628 software. *Journal of Open Source Software*. 2019;3(30):978.
- 629 30. Ding C. Latent Profile Analysis. *The SAGE Encyclopedia of Research Design*. 2022.
- 630 31. Schmidt MN, Seddig D, Davidov E, Mørup M, Albers KJ, Bauer JM, et al., editors.
631 *Latent profile analysis of human values: What is the optimal number of clusters?* 2021.
- 632 32. URT, NBS. *The Initial Results of the Population and Housing Census 2022*. Dodoma:
633 Government Printers, 2022.
- 634 33. Todd G, Msuya I, Levira F, Moshi I. City Profile: Dar es Salaam, Tanzania.
635 *Environment Urbanization ASIA*. 2019;10(2):193-215.
- 636 34. Dirksen J, Pinilla-Roncancio M, Wehrmeister FC, Ferreira LZ, Vidaletti LP, Kirkby
637 K, et al. Exploring the potential for a new measure of socioeconomic deprivation status to
638 monitor health inequality. *International Journal for Equity in Health*. 2022;21(1):56. doi:
639 10.1186/s12939-022-01661-0.
- 640 35. Rosenberg J, Beymer P, Anderson D, Van Lissa CJ, Schmidt J. tidyLPA: An R
641 Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or
642 Commercial Software. *Journal of Open Source Software*. 2018;3(30):978. doi:
643 10.21105/joss.00978.
- 644 36. Gogtay NJ, Thatte UM. Principles of Correlation Analysis. *The Journal of the*
645 *Association of Physicians of India*. 2017;65 3:78-81.
- 646 37. Tein J-Y, Coxe S, Cham H. Statistical Power to Detect the Correct Number of Classes
647 in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*.
648 2013;20(4):640-57. doi: 10.1080/10705511.2013.824781.
- 649 38. Akogul S, Erisoglu M. An Approach for Determining the Number of Clusters in a
650 Model-Based Cluster Analysis. *Entropy*. 2017;19:452.
- 651 39. DeVore GR. Computing the Z Score and Centiles for Cross-sectional Analysis: A
652 Practical Approach. *Journal of Ultrasound in Medicine*. 2017;36(3):459-73. doi:
653 <https://doi.org/10.7863/ultra.16.03025>.

- 654 40. Little TD, Slegers DW, Card NA. A Non-arbitrary Method of Identifying and Scaling
655 Latent Variables in SEM and MACS Models. *Structural Equation Modeling: A*
656 *Multidisciplinary Journal*. 2006;13(1):59-72. doi: 10.1207/s15328007sem1301_3.
- 657 41. Ferguson SL, G. Moore EW, Hull DM. Finding latent groups in observed data: A
658 primer on latent profile analysis in Mplus for applied researchers. *International Journal of*
659 *Behavioral Development*. 2020;44:458 - 68.
- 660 42. Pastor DA, Barron KE, Miller BJ, Davis SL. A latent profile analysis of college
661 students' achievement goal orientation. *Contemporary Educational Psychology*.
662 2007;32(1):8-47. doi: <https://doi.org/10.1016/j.cedpsych.2006.10.003>.
- 663 43. Yalçın IK, Can N, Mançe Çalışır Ö, Yalcin S, Çolak B. Latent profile analysis of
664 COVID-19 fear, depression, anxiety, stress, mindfulness, and resilience. *Current Psychology*
665 (New Brunswick, Nj). 2021;41:459 - 69.
- 666 44. Voisin DR, Kim DH, Hong JS. A closer look at school bonding among African
667 American adolescents in low-income communities: A latent class analysis. *Journal of health*
668 *psychology*. 2018;23(11):1424-37.
- 669 45. Cook T. Community-based participatory research for health: advancing social and
670 health equity. *Educational Action Research*. 2018;27:144 - 5.
- 671 46. Lung-Amam WS, Dawkins C. The power of participatory story mapping: Advancing
672 equitable development in disadvantaged neighbourhoods. *Community Development Journal*.
673 2020;55(3):473-95.
- 674 47. Meenar MR. Using participatory and mixed-methods approaches in GIS to develop a
675 Place-Based Food Insecurity and Vulnerability Index. *Environment and Planning A*.
676 2017;49(5):1181-205.
- 677 48. Vaughn LM, Jacquez FM. Participatory Research Methods – Choice Points in the
678 Research Process. *Journal of Participatory Research Methods*. 2020.
- 679 49. Wallerstein N. Commentary on Community-Based Participatory Research and
680 Community Engaged Research in Health for *Journal of Participatory Research Methods*.
681 *Journal of Participatory Research Methods*. 2020.
- 682 50. Kuffer M. Digitalization and urban development in the Global South: Towards
683 reliable population data in deprived urban areas. *Österreichische Entwicklungspolitik*.
684 2020:73-82.
- 685 51. Thomson DR, Kuffer M, Boo G, Hati B, Grippa T, Elsey H, et al. Need for an
686 integrated deprived area “slum” mapping system (IDEAMAPS) in low-and middle-income
687 countries (LMICs). *Social Sciences*. 2020;9(5):80.
- 688 52. Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice.
689 *Journal of Black Psychology*. 2020;46(4):287-311. doi: 10.1177/0095798420930932.
- 690 53. Babere NJ, Chingwele I. Neighbourhood Unit Residential Segregation in the Global
691 South and Its Impact on Settlement Development: The Case of Dar es Salaam City.
- 692 54. Musterd S. Towards further understanding of urban segregation. *Handbook of urban*
693 *segregation*: Edward Elgar Publishing; 2020.
- 694 55. Smiley SL. Patterns of urban life and urban segregation in Dar es Salaam, Tanzania:
695 University of Kansas; 2007.
- 696 56. Smiley SL. Mental maps, segregation, and everyday life in Dar es Salaam, Tanzania.
697 *Journal of Cultural Geography*. 2013;30(2):215-44.
- 698 57. Vaughan L, Arbaci S. The challenges of understanding urban segregation. *Built*
699 *Environment*. 2011;37(2):128-38.
- 700 58. Villaça F. São Paulo: urban segregation and inequality. *Estudos avançados*.
701 2011;25:37-58.

702 59. Babere NJ, Chingwele I, editors. Neighbourhood Unit Residential Segregation in the
703 Global South and Its Impact on Settlement Development : The Case of Dar es Salaam
704 City2018.
705

706 **Supporting information**

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

709

710 **INSERT S1 Table HERE**

711

712 **INSERT S2 Fig HERE**

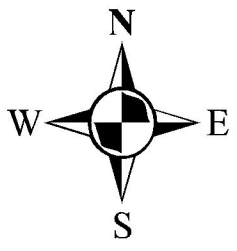
713 **S2 Fig: Pearson correlation plot with correlation coefficients and p-values, showing**
714 **numerical correlation coefficient labels only for pairs which are strongly correlated ($\rho \geq$**
715 **+/- 0.4) and statistically significant ($p \leq 0.05$) while insignificant variables are**
716 **unlabelled.**

717

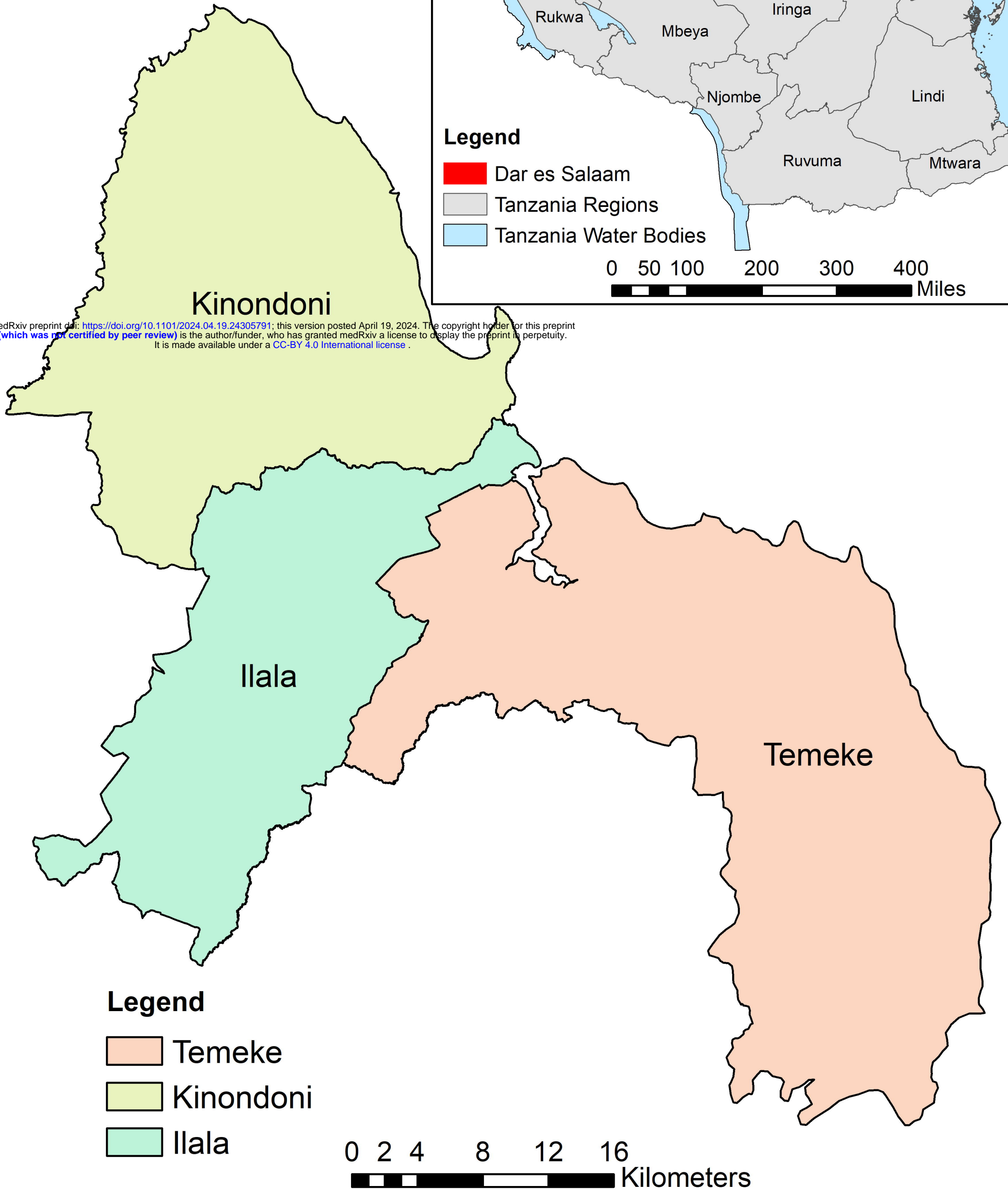
718 **INSERT S3 Fig HERE**

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

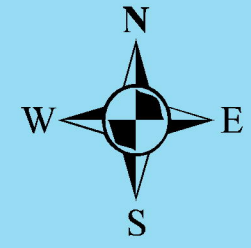
725



medRxiv preprint doi: <https://doi.org/10.1101/2024.04.19.24305791>; this version posted April 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



medRxiv preprint doi: <https://doi.org/10.1101/2024.04.19.24305791>; this version posted April 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](#).



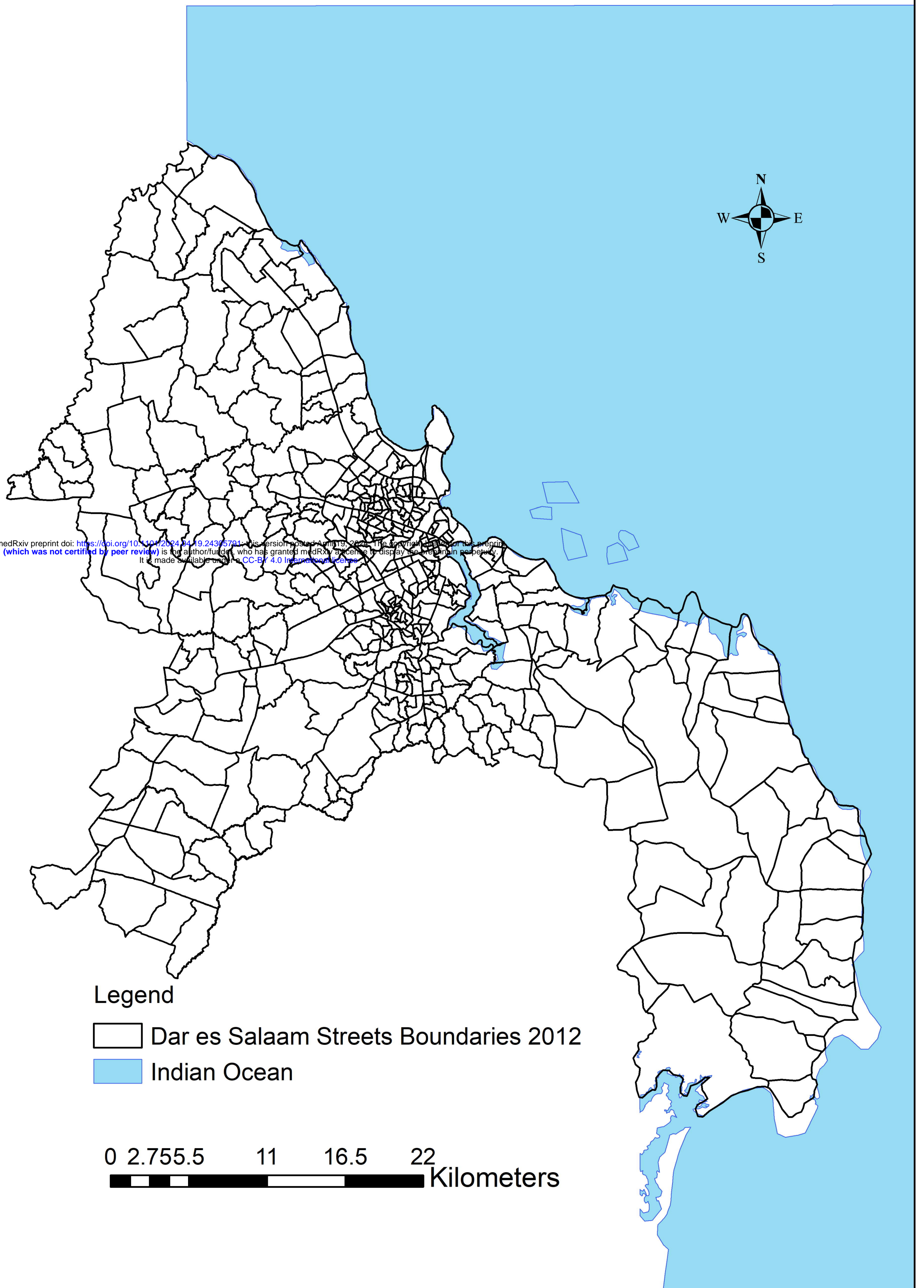
Legend

-  Dar es Salaam Ward Boundaries 2012
-  Indian Ocean

0 2.75 5.5 11 16.5 22 Kilometers



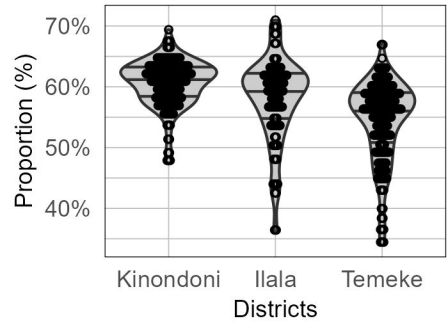
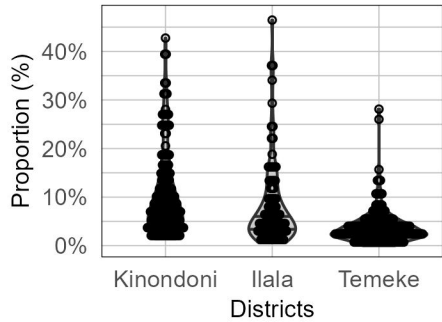
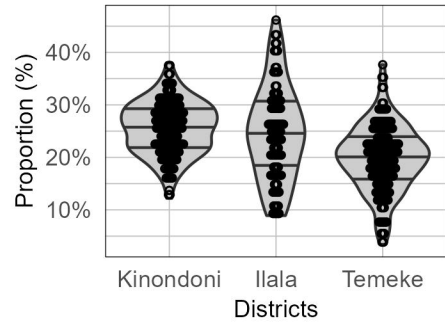
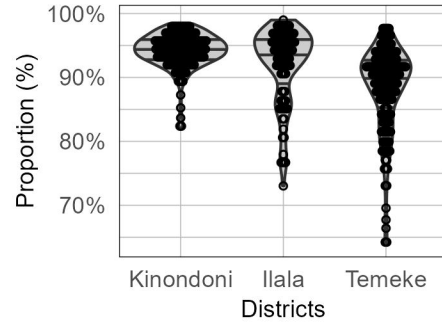
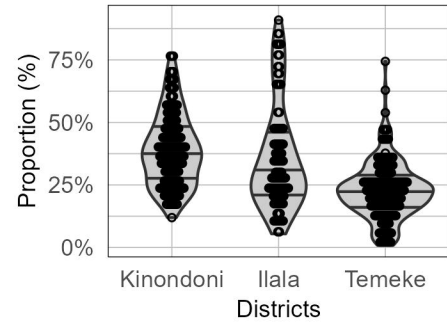
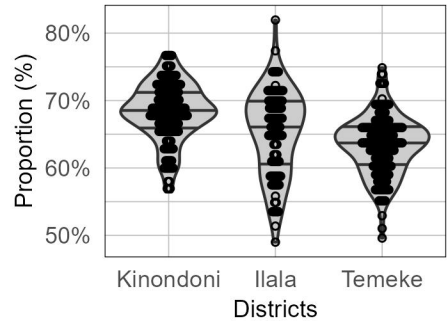
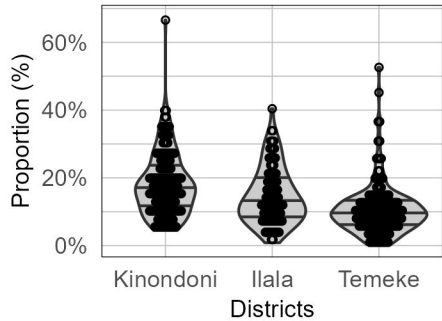
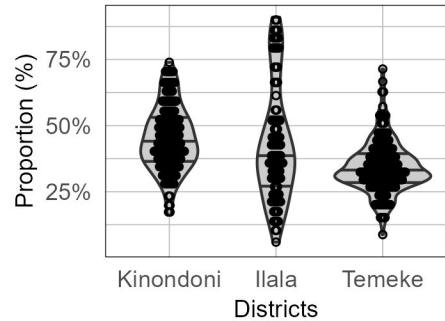
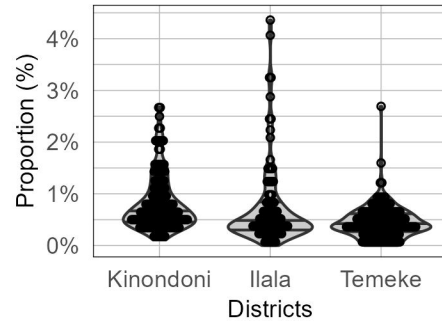
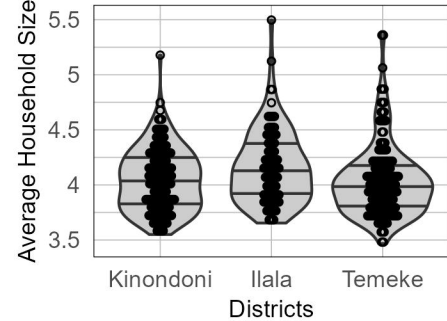
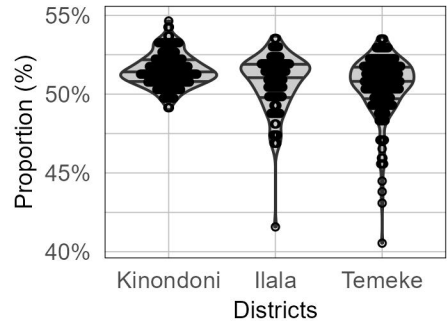
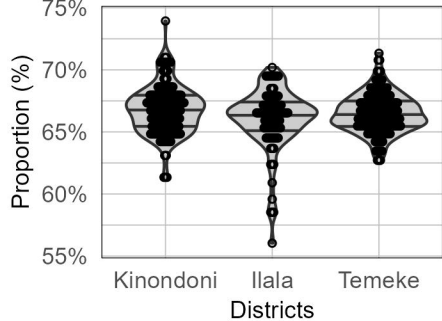
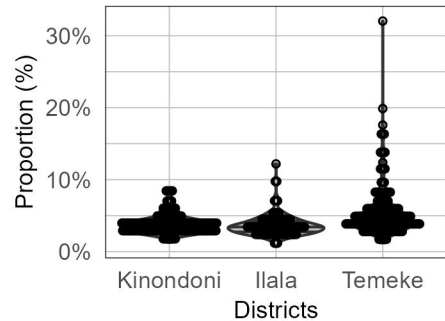
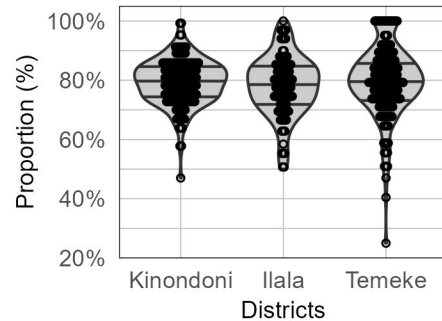
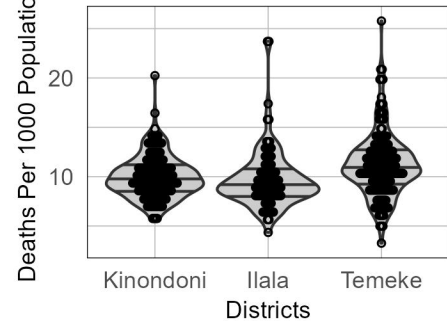
medRxiv preprint doi: <https://doi.org/10.1101/2020.04.19.24375791>; this version posted April 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

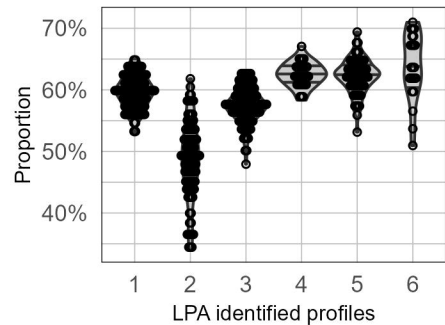
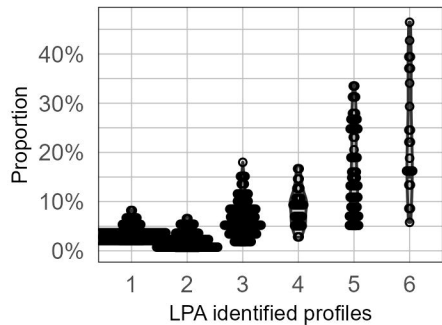
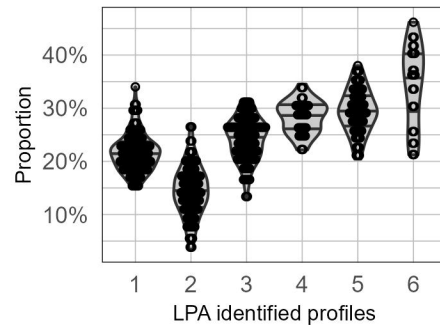
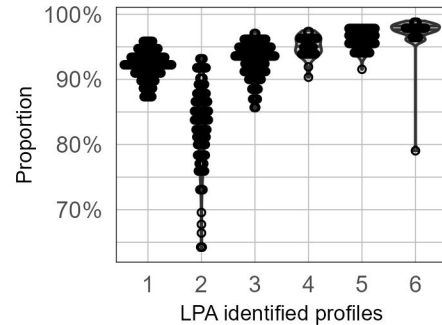
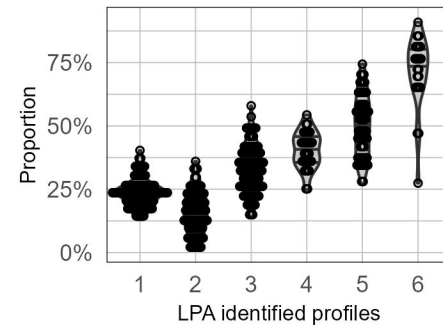
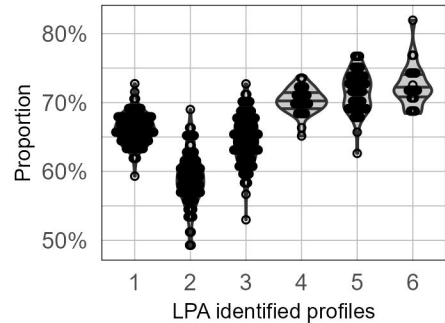
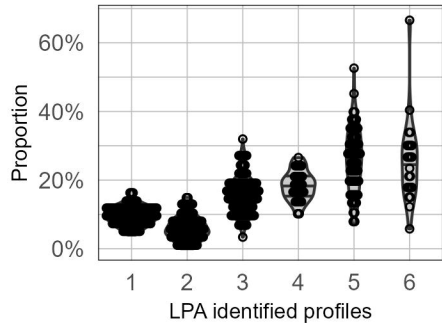
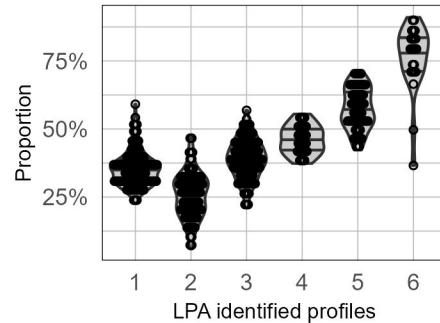
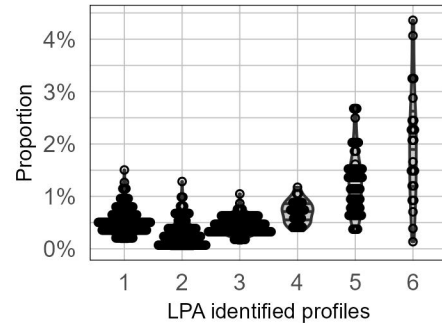
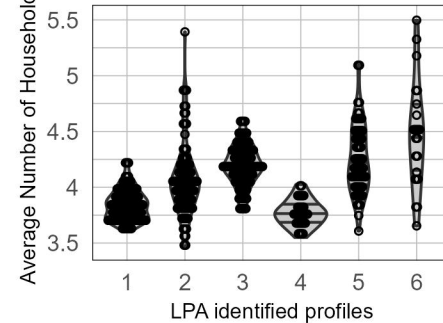
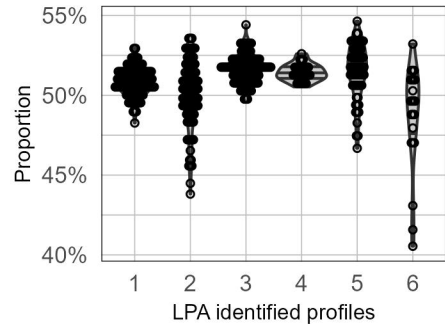
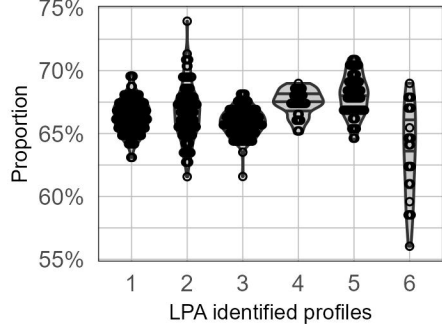
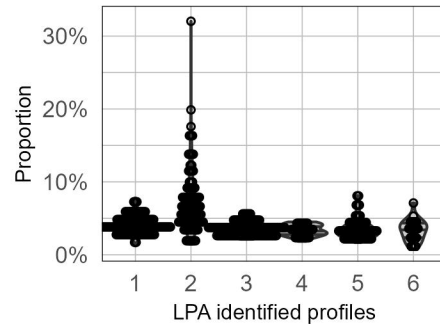
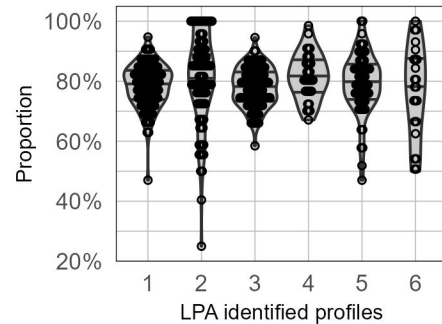
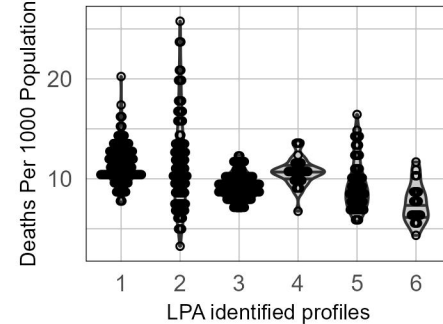


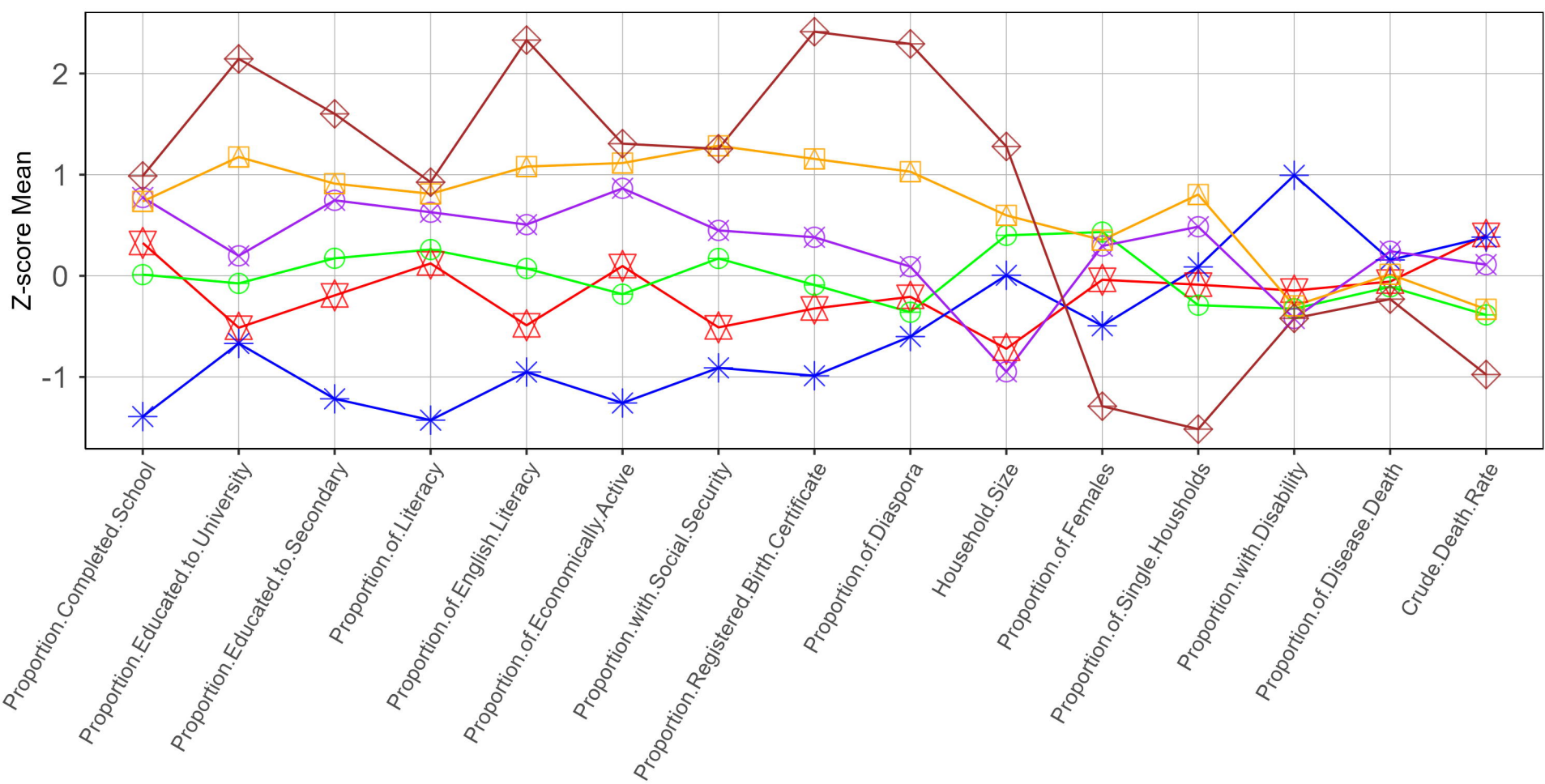
Legend

-  Dar es Salaam Streets Boundaries 2012
-  Indian Ocean

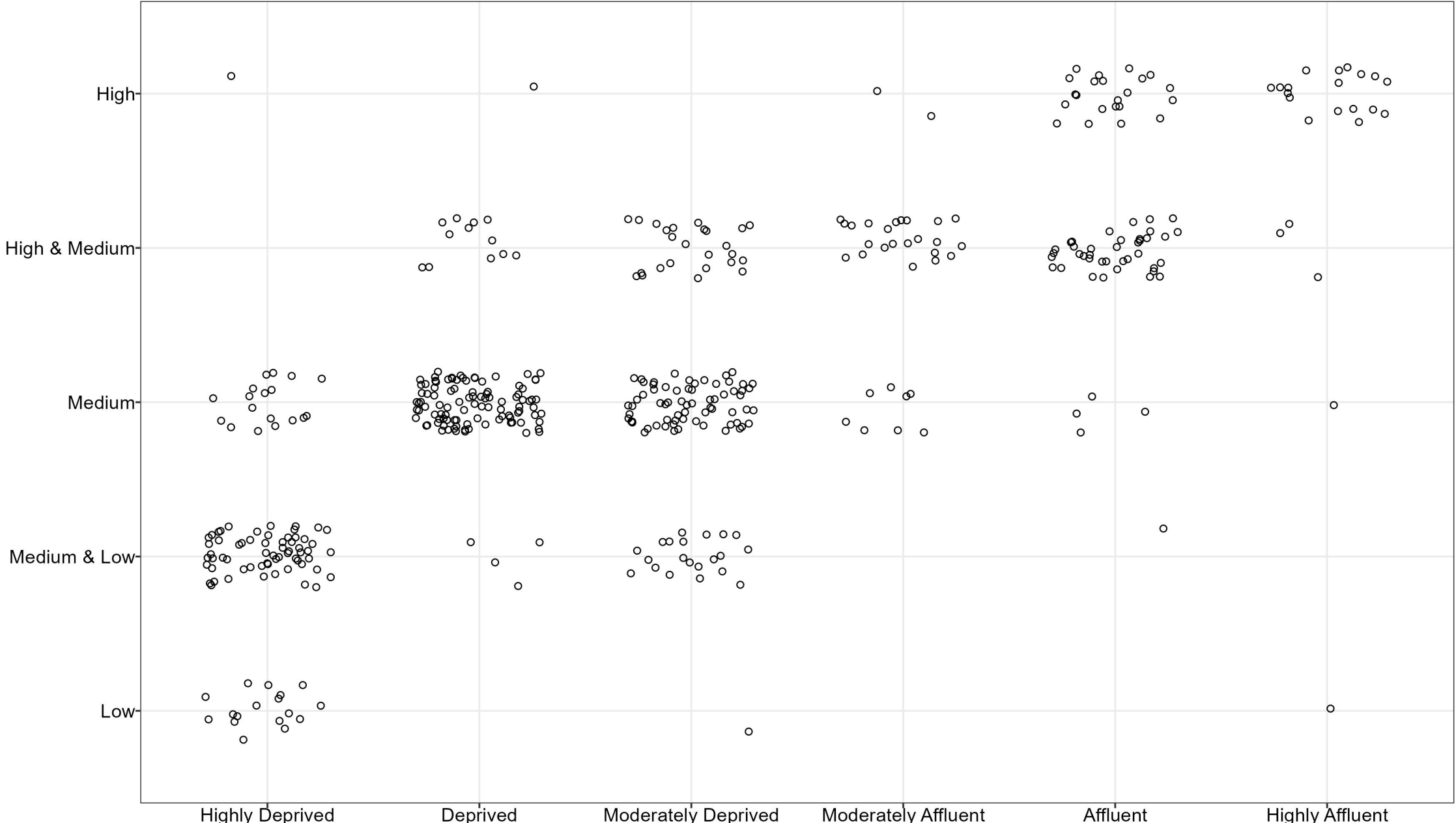
0 2.75 5.5 11 16.5 22 Kilometers

Proportion Completed School**Proportion Educated to University****Proportion Educated to Secondary****Proportion of Literacy****Proportion of English Literacy****Proportion of Economically Active****Proportion with Social Security****Proportion Registered Birth Certificate****Proportion of Diaspora****Household Size****Proportion of Females****Proportion of Single Households****Proportion with Disability****Proportion of Disease Death****Crude Death Rate**

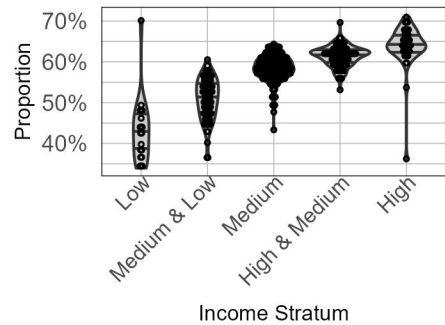
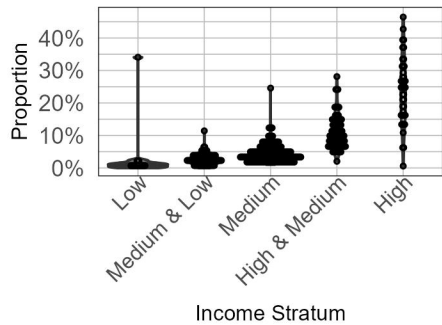
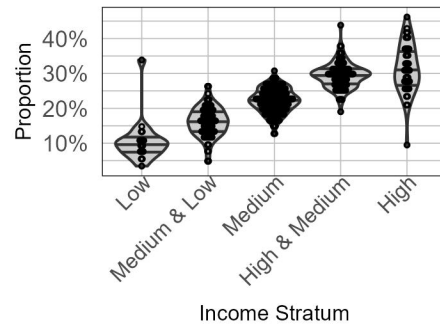
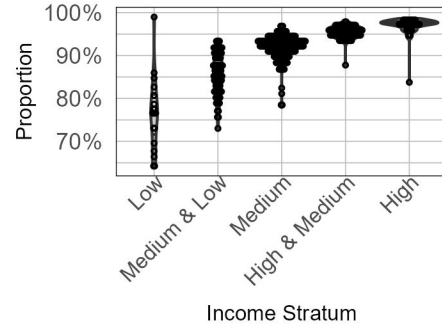
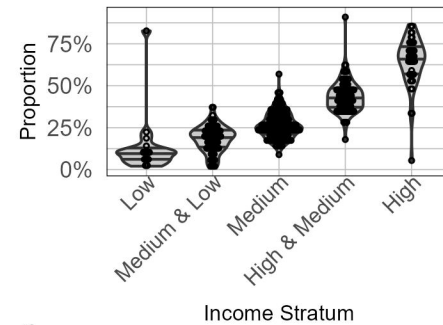
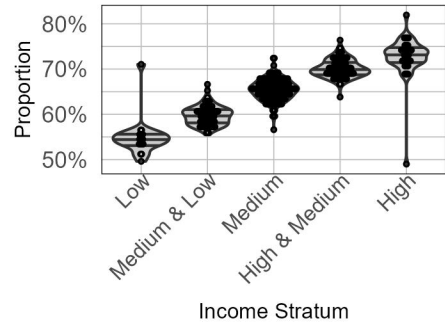
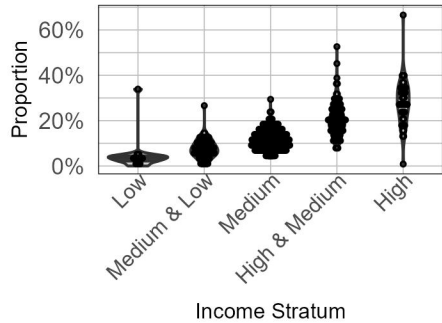
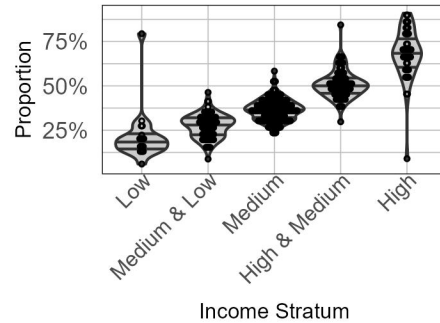
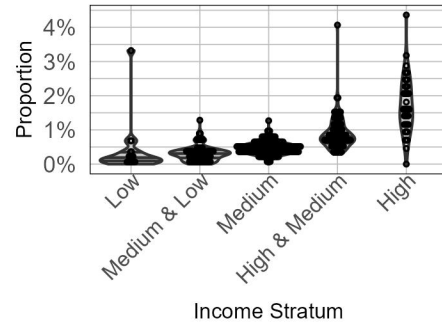
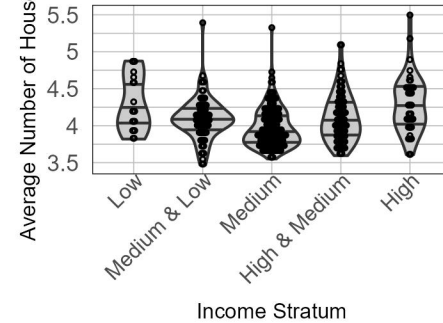
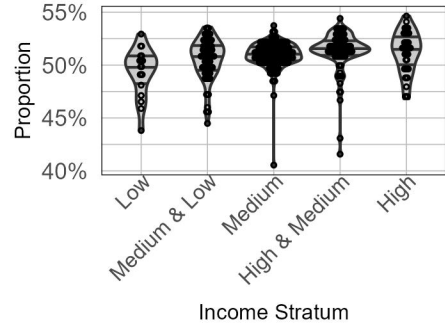
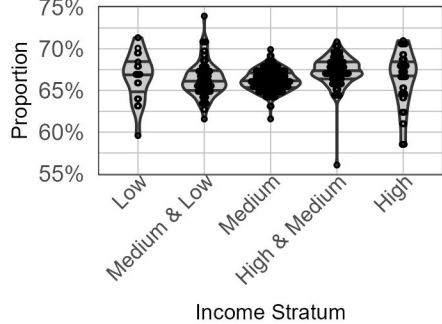
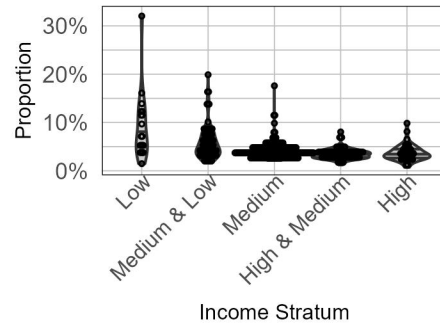
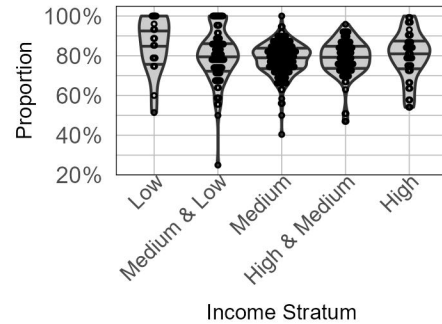
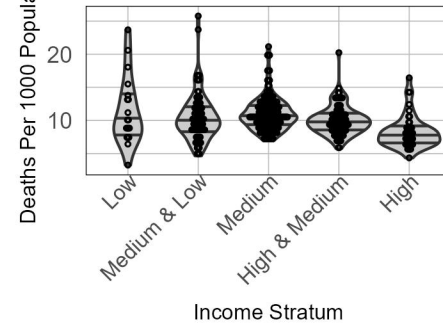
Completed School**Educated to University****Educated to Secondary****Literate****English Literacy****Economically Active****with Social Security****Registered Birth Certificate****of Household Members in Diaspora****Household Size****Females****Single Head Housholds****People with Disability****Disease Caused Death****Crude Death Rate**

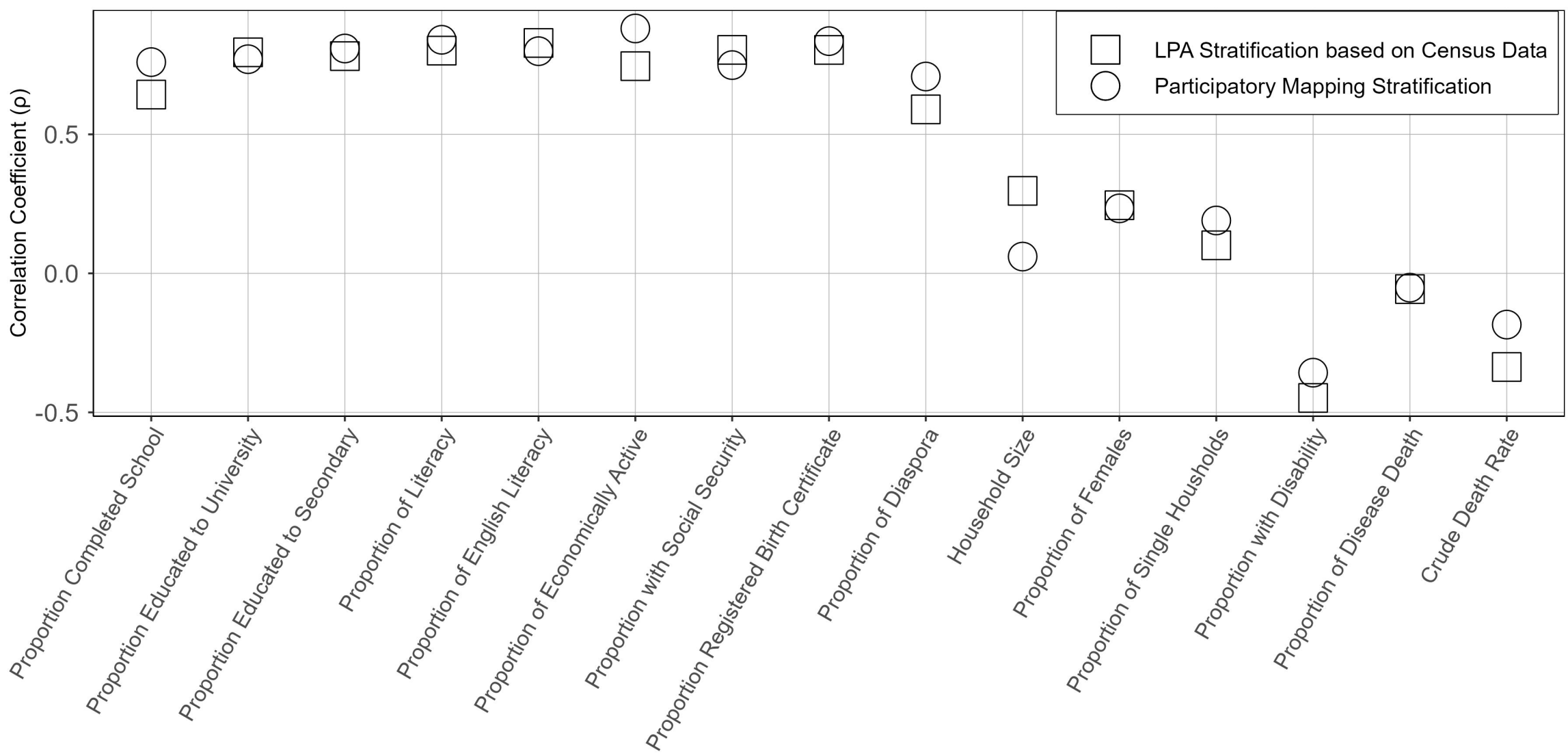


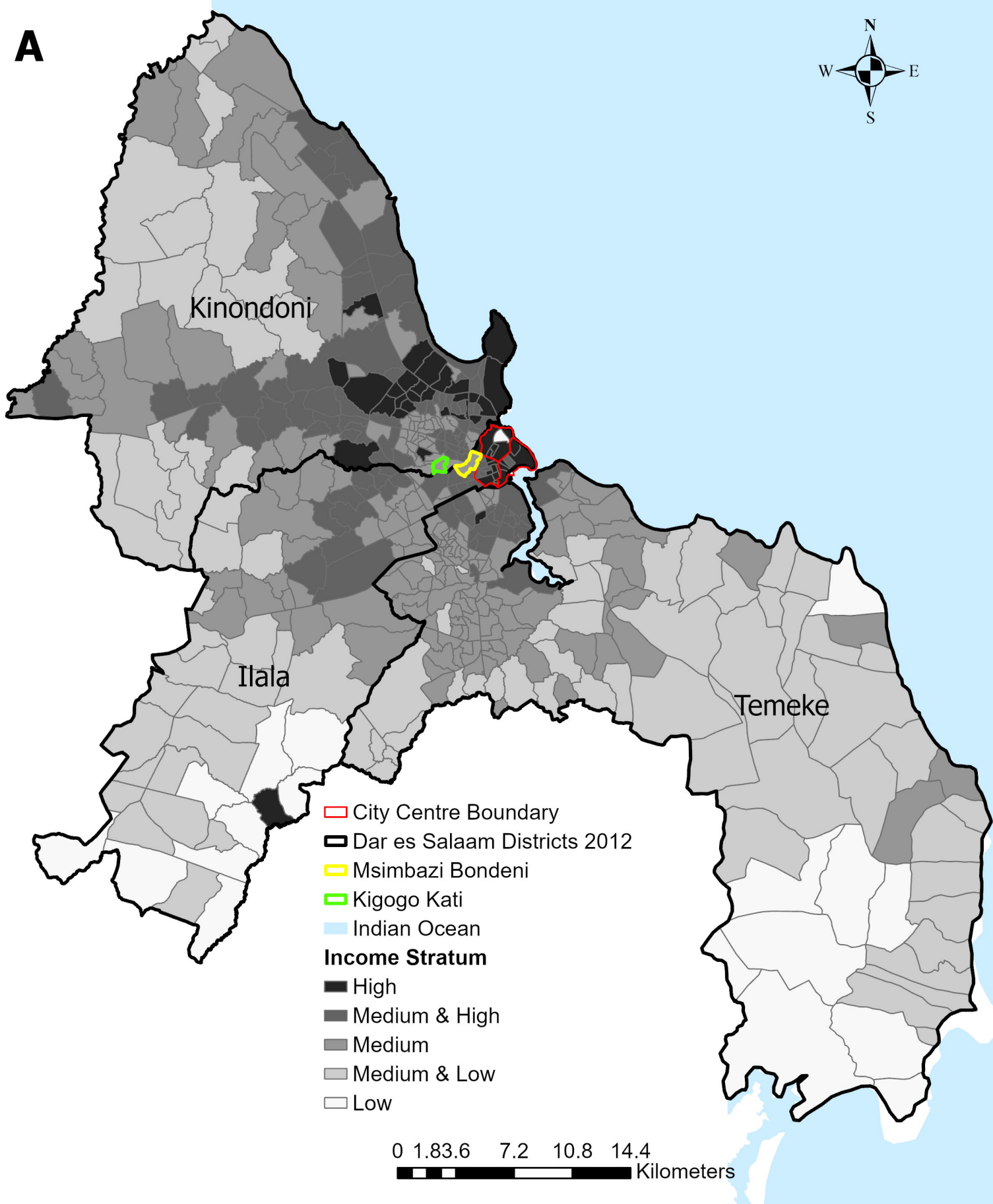
Income Stratum Identified by Participatory Stratification Approach



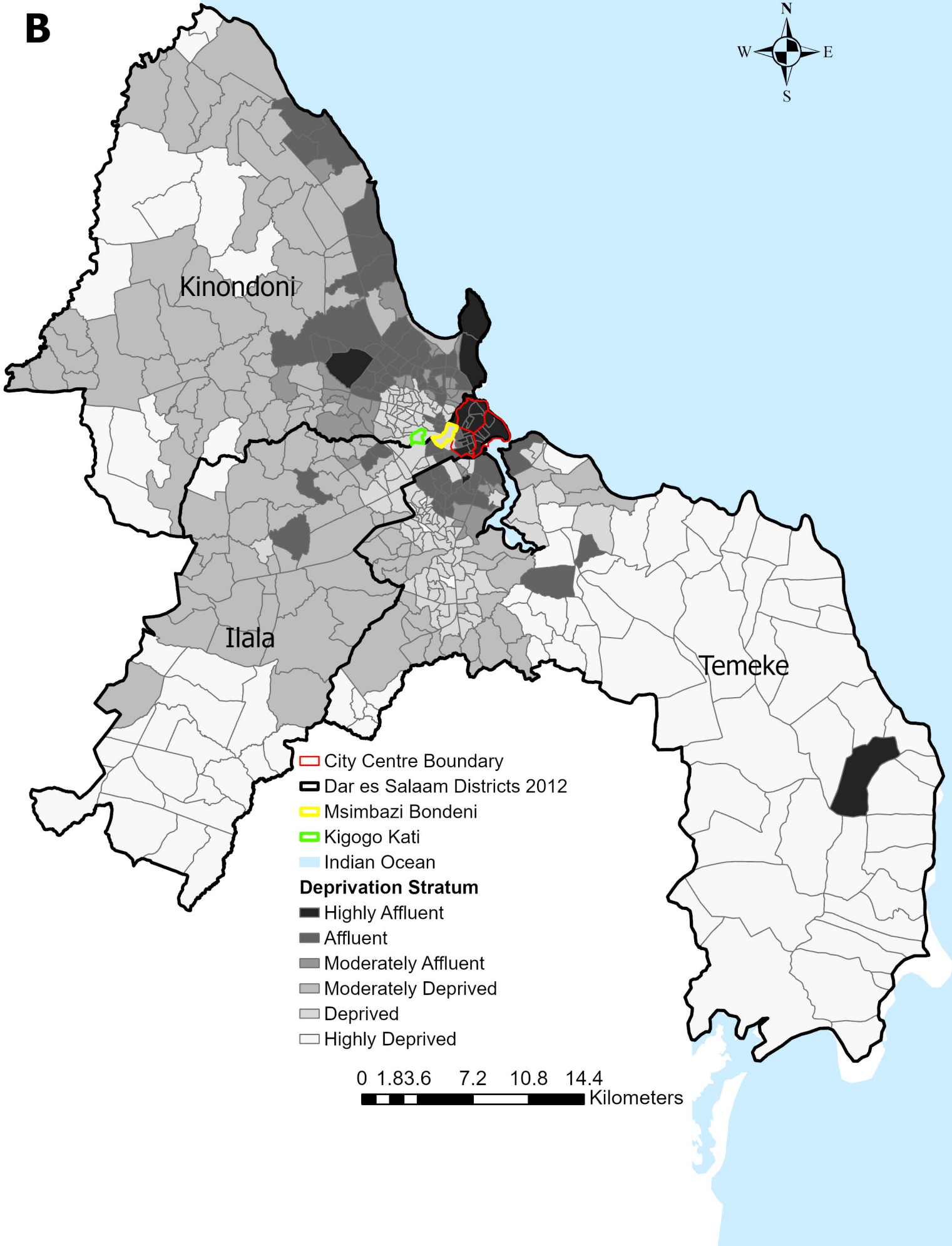
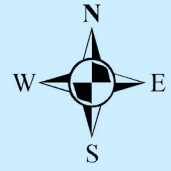
Deprivation Stratum Identified by Latent Profile Analysis of Census Data

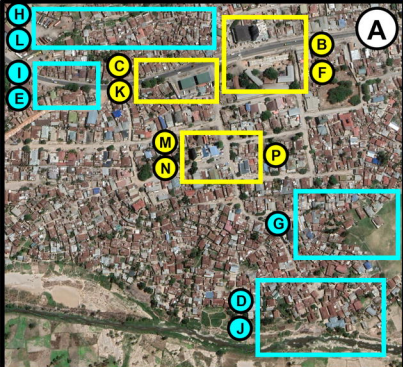
Completed School**Educated to University****Educated to Secondary****Literate****English Literacy****Economically Active****with Social Security****Registered Birth Certificate****of Household Members in Diaspora****Household Size****Females****Single Head Households****People with Disability****Disease Caused Death****Crude Death Rate**

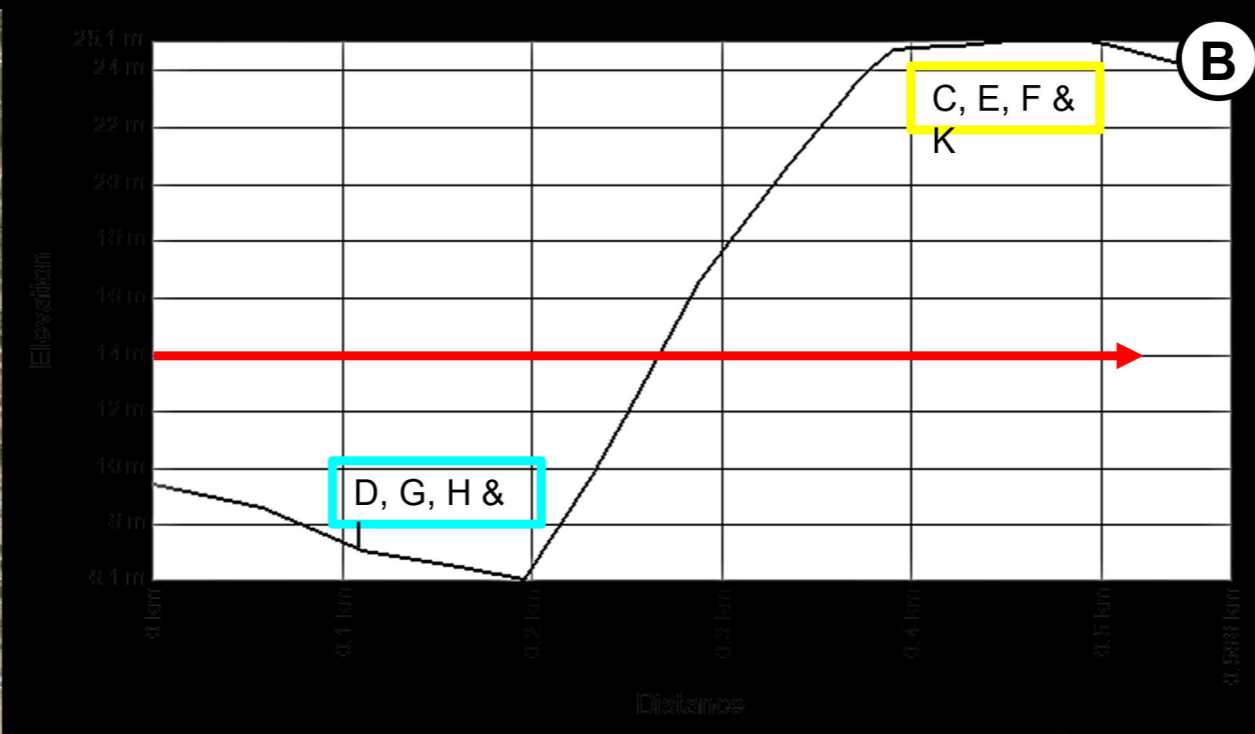
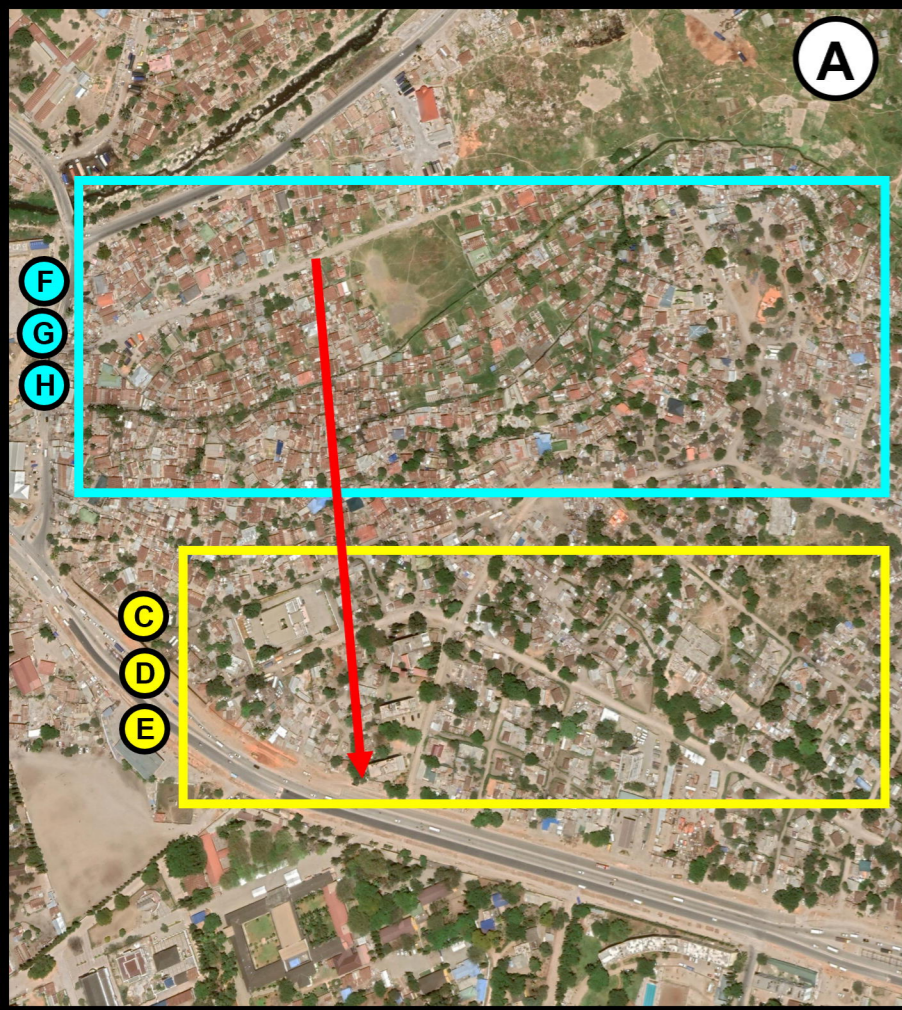




B







NEIGHBOURHOOD SELECTION

DAR ES SALAAM

TYPOLOGY / FORM / TYPE
URBAN SETTLE.

INCOME
Low / High / Medium
Mixed - High / Medium
Mixed - Low / Medium

LOCATION
CBD
POD CITIES
COMMERCIAL DISTRICT
SATELLITE TOWN

