Sub-National Age and Sex Distributions: New data and methods for low-income countries

Mark R. Montgomery,* Emanuel Agu, Deborah Balk, Alessandra Carioli, Stefan Leyk, and Evgeny Noi

When a climate-related hazard strikes, its effects depend crucially on where, precisely, it strikes and who lives in the path of risk. Low- and middle-income countries often lack up-to-date information on their at-risk populations with enough spatial specificity to guide disaster planning and response—information that is routinely available in high-income countries. Poor countries seeking to respond effectively to the call of the Sustainable Development Goals for monitoring of diverse development targets and indicators at the sub-national level, also require similarly disaggregated sub-national time-series data. Any number of health risks—malaria among them—exhibit strong differentials by age, sex, and geographic space, so that risk assessment must be based on disaggregated demographic data. The waves of change in age structure producing what is known as the "demographic dividend" play out not only at the national level, but also in more localized labor markets. All these concerns underscore the need for sub-national socioeconomic data, ideally at fine spatial scales, especially in countries whose statistical systems have not in the past given high priority to sub-national data production.

To meet such challenges, physical and social scientists have joined forces over the past two decades with the common aim of producing internationally-comparable, high-resolution population grids (Tatem et al. 2013; Alegana et al. 2015; Leyk et al. 2019). For the first time, it has become possible to access credible estimates of total population and population density for finely-detailed sub-national geographic units. Because its estimates—derived from LANDSAT and SENTINEL satellite imagery—span the period from 1975 to the present, the most promising of these global datasets is the Global Human Settlement Layer (see European Commission 2024, for the 2023 GHSL revision).

Now that the first generation of population grids has arrived, we are in a position to further enrich their demographic content with microdata on a range of socioeconomic indicators. Using open-source individual-level data from population censuses and surveys, together with detailed sub-national tabulations from national statistical offices to fill gaps, we have compiled a very large collection of fine-resolution sub-national age and sex distributions for low- and middle-income countries, covering the period from the early 1980s to the present. This note presents the collection, with a focus on the multiple analytic challenges that such sub-national data bring to the fore. Once these census- and survey-based age and sex distributions have been vetted through comparisons with other sources, we will take steps to incorporate sub-national estimates of education and other core development indicators.

*Corresponding author: Economics Department, Stony Brook University, mark.montgomery@stonybrook.edu

1 The scope of our collection

Perhaps the closest analog to our collection is represented by the *Gridded Population of the World* CIESIN (2018), which in version 4.11 provides the most recent edition of a remarkable, large-scale, decades-long collaborative effort which links total population counts by sub-national administrative units (as reported by national statistical offices) to spatial boundary files ("shapefiles") for these sub-national units. Because some statistical offices do not consent to having their sub-national boundaries published, the GPW team converts the population data to raster form, gridding them in an output resolution of 30 arc-seconds (or about 1 km at the equator). In recent revisions, the GPW algorithm has also integrated GHSL rasters of built-up land, so that the administrative-unit population counts can be assigned to the spatial locations that are plausible sites of human habitation. The overall approach has a compelling logic, and is distinguished from other broadly similar efforts by its analytic transparency and commitment to data-sharing within the limits of its agreements with statistical offices. We take the GPW design as a model very much worthy of emulation.

Our collection enriches the GPW approach by incorporating *census and survey micro-data*, thereby enabling spatially-grounded demographic measures that can be described and analyzed in much greater depth than is possible with pre-tabulated aggregates of official data. We also maintain what amounts to a time-series of estimates by country, with data entered for each survey and census in our database, so as to provide a foundation for analysis of sub-national trends and lay the groundwork for similarly-detailed forecasts. A newly-released database of within-household (co-residential) relationships (Galeano et al. 2024) is similar in spirit to our own approach; it relies as we do on micro-data arranged according to the boundaries of sub-national administrative units. Age and sex distributions have not yet been included in the Galeano et al. compilation, but there are good prospects for data-sharing on that front. Likewise, in preparing for its upcoming new release, the GPW team is investing effort in adding official tabulations of age and sex, which will provide authoritative estimates against which ours can be assessed.

Our effort, and the comparable microdata work of Galeano et al. (2024), would simply not have been possible without the substantial resources invested in "spatializing" demographic data by the *Integrated Public-Use Microsamples–International* (IPUMS-I) program, by the pioneering *World Fertility Surveys* (WFS) program that went into the field in the early 1980s, and by the two major on-going demographic survey programs, the *Demographic and Health Surveys* (DHS) and the *Multiple Indicators Clusters Survey* (MICS), with the latter's spatial efforts being assisted by IPUMS-I over the past few years.

We now present a description (mainly via summaries in map form) of our collection, concentrating for the purposes of this short note on *low- and middle-income countries*, where the needs for sub-national demographic data are greatest. Our collection includes census microdata and national statistic office tabulations for high-income countries, but we do not comment further on such data in what follows.

1.1 GHSL raster data

Figure 1 indicates the potential of the GHSL raster data on built-up land area, with the graphs depicting the situation in Kenya in 1975 and 2020. As can be seen, in this country the north-eastern region bordering Somalia was only sparsely populated in 1975 and even by 2020 no dramatic change has taken place. Conventional summaries of demographic measures would assign one summary value to each administrative unit shown in this map, effectively hiding from view the relatively few locations within the unit where there is evidence of human settlement. Note that across Kenya, settled areas expanded greatly between the 1975 and 2020 end-points of satellite observation, although on closer inspection it can be seen that most of this areal expansion took place at very low levels of built-up. We will return to the Kenya case at the close of this note, to illustrate what can be gained by combining spatial with demographic information.

1.2 Census and national statistical office sources

One of the census-based sources we rely upon is similar to the materials used by the GPW program. The US Census Bureau's International Program supplies full age and sex data from national censuses in selected low- and middle-income countries, processing the data for level-1 and level-2 administrative units and in even more spatial detail when that is possible.[1] Figure 2 displays the availability of these tabulations by country. These are not microdata as such, but rather whole-census figures that can be used as a check on the microdata estimates from our other sources. The Census Bureau invests substantial effort in locating and editing sub-national spatial boundary files that fit the regional categories of each census. Our research team is supplementing these age–sex tabulations with similar data from national statistical offices, which have proven to be more than willing to equip the team with both demographic counts and tailored shapefiles.

For census microdata, we make use of the open-source, meticulously curated datasets provided by the IPUMS-I program, which takes simple random samples (sometimes weighted) of the full census records held in its vast collection. These are invaluable data, covering countries in Latin America and West Asia that are not otherwise adequately represented in our survey-based sources. Our research team currently works with some 258 census microdata samples from the range of low- and middle-income countries seen in Figure 3, and is integrating new releases from April 2024 with another tranche of data expected later this year. A number of countries are represented by two or more censuses, as the figure makes clear. The IPUMS-I program has invested heavily in linking the sub-national region variables in each census with a matching spatial boundaries file. In general this is an excruciatingly difficult, time-consuming task, because there exist few authoritative go-to sources for any but present-day regional boundaries.

1.3 Demographic surveys

World Fertility Surveys (WFS) The earliest of the large-scale demographic survey programs in low- and middle-income countries, the WFS provided results in such countries in the late 1970s and early 1980s. Our research team has spent considerable effort in locating spatial boundary files to match the survey-specific regional variables, and has had good if not universal success in linking such files to these early surveys. The limiting problem is that owing to their focus on reproductive-age women, a number of WFS surveys did not process data from their full household rosters, making it impossible to derive age and sex distributions. Even so, as Figure 4 indicates, enough surveys retained the household information to warrant inclusion in our collection, extending the time-series of estimates for several important but under-documented cases such as Tunisia, Syria, Sri Lanka, and South Korea. These very early WFS estimates help to anchor forecasts based on the full data-series (from other sources) for these countries.

Demographic and Health Surveys (DHS) The DHS program has released micro-data for no fewer than 359 surveys as of September 2024; our holdings are routinely updated as new surveys are made available. Figure 5 indicates the breadth and depth of the collection, with many countries having hosted quite a few DHS surveys over the course of the program.

Within-country boundaries come from several sources:

1. The default source is the *Spatial Data Repository* (SDR) at https://spatialdata.dhsprogram.com/home/, whose shapefiles correspond to the hv024 variable of a DHS household file or an equivalent. The units depicted in the shapefiles are generally first-level administrative units, but with substantial variation in the level and type of unit across surveys. Even across a single country's survey waves, sub-national boundaries tend to vary significantly in keeping with the country's revisions of its internal administrative structure, as we will shortly demonstrate.

[1] See https://www.census.gov/programs-surveys/international-programs.html

2. Most DHS surveys also provide an additional, more detailed geographic region variable, for which a shapefile has to be located and edited (if necessary) to correspond to the regional coding categories of this variable. In searching for appropriate spatial files, we have relied on several sources, including GADM (which unfortunately provides only the most recent boundaries), the IPUMS-I shapefiles (which are especially valuable in being keyed to the date of the census), and GAUL.[2]

In Figure 6 we take a closer look at the case of India, for which the *SDR* provides state boundaries for the country's first three surveys (note the emergence of new states by 2005) and district (admin-2) boundaries thereafter. Not many surveys offer spatial detail comparable to India's; some provide links from the survey records only to first-level administrative units, and especially in the early years of the DHS program, a number of surveys were limited to ecological regions with no particular administrative significance or to very broad aggregates of administrative units. With time, spatial coverage has tended to improve in specificity, but as the case of India suggests, there are numerous instances of back-sliding in which a spatially-detailed survey is followed by a survey having much less detail.

Uniquely among our spatial sources, for the vast majority of its surveys since the late 1990s, the DHS program has collected point coordinates approximating the locations of sampling clusters. With reference to Africa, Figure 7 documents which surveys are equipped with cluster coordinates. Although the coordinates are randomly displaced by a few kilometers to preserve respondent confidentiality, the displacement algorithm prevents the reported coordinate from being located in a different level-1 administrative unit (or the equivalent) than its actual location. For finer spatial divisions, however, a reported cluster does not necessarily lie in the unit in which the reported point is found. Our spatial methods limit the risks of such mis-classification, but the displacement errors cannot be entirely eliminated. Subject to this qualification, however, a sampling cluster point can be located in relation to geographies of considerably finer detail than provided by the *SDR* boundaries.

Multiple Indicator Cluster Surveys (MICS) The MICS survey program began in earnest around 1999, and is now comparable in scope to the DHS program. Figure 8 depicts the geographic sweep of the program, which is able to mount fieldwork in countries that have not been hospitable to the DHS. Unlike the DHS, links from MICS surveys to sub-national regions are almost always restricted to a single regional variable, and until recently no sub-national shapefiles were made available to accompany the MICS surveys. In a recent and welcome departure from previous practice, a collaboration has been forged with IPUMS-I that is now placing such boundary files into the public domain. Some surveys remain to be covered, but as with the DHS program, we have been able to locate acceptable shapefiles to pair up with MICS surveys in GADM or similar sources.

2 Research challenge I: Selective omission of age groups

Every trained demographer is aware of the potential for age mis-statement and age-heaping in demographic materials. Surprisingly, however, less attention has been given to age-specific omissions of broad age groups in survey and census data.[3] Relevant literature on the omission problem and the related issues of age-heaping includes Johnson et al. (2022) and Hauer, Baker, and Brown (2013), which ably summarize much earlier research. In preparing its estimates of national age and sex structures from series of population censuses, the UN Population Division has given close attention

[2]For GADM, see https://gadm.org/ and for GAUL, see ***. In checking the agreement of regional variables in an older survey with the boundaries in effect at the time of the survey, the www.statoids.com website has proven to be invaluable.

[3]Most attention has been devoted to age misreports on the part of interviewers to lighten their work-load by recording young women as a year or so too young, or older women just bit too old, to be administered the survey questions designed for reproductive-age women. Although we know of no similar studies for men, presumably the same interviewer motivations would apply to them.

to the possibility of omission of young children from the census household rosters. (Our research team is seeking advice from the UN on possible correction factors.) It seems likely that omission of children is also a problem in survey data, even when they are as well-implemented as the DHS and MICS surveys have been.

There is every reason to be concerned about potentially systematic biases in using unadjusted census and household survey rosters to estimate population age percentages, affecting not only children but also the elderly. To date we've seen almost no literature on old-age-selective omission in surveys of low-income countries: only Randall and Coast (2016) appear to have addressed the issue explicitly. In sub-Saharan Africa DHS surveys, Randall and Coast find evidence of greater omission percentages for elderly women than elderly men.

To be sure, for the elderly it can be exceedingly difficult to confirm that age-omission is indeed a problem. When the population percentage of the elderly is low, it remains possible that no household in a sampling cluster actually has an elderly member. Imagine that a household selected at random can have no more than 1 elderly member, with the probability of 1 such member being p and $1 - p$ being the probability of no elderly member. Further suppose that $n = 30$ households are surveyed independently to form each sampling cluster, as is the case with DHS sampling designs. Under these artificial assumptions, $(1 - p)^n$ is the probability that the cluster as a whole contains no elderly people. If $p = 0.05$ of households have 1 elderly person, the chance is about 0.21 that the cluster has no elderly residents, amounting to roughly one-fifth of all clusters in this illustrative example. Hence, it is possible that sampling variation coupled with "laws of small numbers" might explain the absence of the elderly. The same argument simply does not apply to children, however, since their percentage of the population is considerably larger in most low-income countries.

3 Research challenge II: Statistical "representativeness"?

An understandable concern in dividing survey samples into sub-national sub-samples is that the standard errors on the sub-national summary statistics may be too large to be tolerable. Unless a survey is designed to allow cross-unit comparisons at the sub-national level—at least for units more finely divided spatially than level-1 administrative areas, which often form the survey strata—contrasts among unit-specific means may not be based on enough observations to reach statistical significance. In other words, at the level of individual geographic units, it may be difficult to cleanly distinguish spatial signals from random noise.

Although a survey's technical documentation not uncommonly warns users that the survey is "not representative" below the level of its strata, the situation is not quite as simple as this phrase might suggest. The standard errors on a summary statistic for a small spatial unit depend crucially on the nature of the statistical model that is being brought to bear. In particular, if data are pooled across small units in the model, with covariates introduced to identify departures of a given unit from the mass of pooled data, this can have a major bearing on the standard errors. In effect, a pooled model "borrows strength" (as Bayesians say) from other units in the set being pooled, so as to better identify the distinctive levels of the unit of interest. In deciding how far to proceed with disaggregation, a balance needs to be struck between the assumptions used to justify data-pooling (an aggregating device) and the sample size underlying each unit's data.

A closely-related issue is how to extrapolate measures from surveyed spatial units to the nearby units that were not surveyed. In conventional treatments, this issue is effectively hidden from view: Sampling clusters in a few sections of a province, if they are sufficiently numerous, are taken to represent the province as a whole. But when the data are sub-divided spatially in finer units than is usually the case, the extrapolation issue comes out into the open, and needs to be explicitly addressed. Two recent technical reports from the DHS program review spatial methods for extrapolation (Burgert-Brucker, Dontamsetti, and Gething 2018; Janocha et al. 2021). We will implement and assess these methods in the final paper, applying them to a test set of countries—Kenya, Viet Nam, South Korea, Peru, and Spain—on which our research team has been focusing.

4 Research challenge III: Heterogeneity and change in spatial units

The variation in sub-national units over time seen in Figure 6 for India—a country-specific phenomenon evident throughout our collection—suggests that especially where near-term forecasts are needed, statistical methods for data integration will have to be developed. We do *not* advocate a "harmonization" approach whereby heterogenous units are somehow forced into a common spatial mold, whose boundaries would generally represent coarse "lowest-common-denominator" artificial sub-national areas. Such approaches discard meaningful variation and typically rest on hard-to-defend (or even unspoken) assumptions about the permissible limits on aggregation and disaggregation.

We prefer to leave the spatial units as they are, but where statistical summaries and forecasts are in order, to apply robust standard error methods in the spirit of Conley (1999) with panel-data extensions in Colella et al. (2019). The details will be presented in the paper and tested on a subset of countries (again, Kenya, Viet Nam, South Korea, Peru, and Spain) that present a variety of administrative-unit sizes and time-series.

5 An illustration using child-woman ratios

To preview how age-sex distributions can be refined spatially, we take the case of the child–woman ratio, which is generally expressed as the number of children (of both sexes) under 5 years of age, divided by the number of women in the reproductive age range from 15 to 49. Using Kenya as an example, we show in Figure 9 the sub-national distribution of the child–woman ratio from 1969 to 1999, basing the maps on IPUMS-I census microsamples summarized for level-1 administrative regions. The darker is the regional color in the map, the higher is its child–woman ratio, a measure that provides an indication of the level of regional fertility adjusted for child mortality. (For Kenya, national comparisons confirm that the child-woman ratio is highly correlated with total fertility rates—but TFRs as such require data on recent births, and therefore cannot be calculated directly for most censuses.) By the end of the period shown, the child-woman ratio had plummeted well below 1 surviving child per woman in Nairobi and the region just to its north, whereas in the north-east, one of the least-developed areas of the country (a site of persistent low-level civil conflict with ethnic Somalis through the 1970s and well into the 1980s), the child–woman ratio trended down but remained relatively high.

These conventional sub-national portrayals of demographic indicators do not capture the within-region variation apparent in Figure 10, which uses a GHSL raster to assign the level-1 child–woman ratio to within-unit built-up areas only. As can be seen, in the north-eastern region the vast majority of Kenya's terrain is uninhabited, with small concentrations of settlement near Wajir (a nexus of trade routes with Somalia); in the south, while there are outposts of population throughout, the heavier concentrations of settlement are situated on or near the coast, with Mombasa being especially prominent. Similar concentrations appear near Kisumu and Lake Victoria to the west. As the map shows, access to fine-resolution GHSL raster data (open-source, and globally available) uncovers striking within-region spatial features and highlights important detail that would otherwise be hidden from view.

Although the example we have just shown allocates level-1 demographic information to the raster cells within each such unit, Figure 11 demonstrates that meaningful demographic variation is available at finer spatial scales. In the final paper, we intend to experiment to see whether different types of spatial containers—the level-2 units are nested within the level-1 units—bring out new features of substantive demographic interest.

We think that the spatial assignment of demographic indicators to areas of built-up land (i.e., to raster cells having any detectable amount of built-up land cover) is quite reasonable as a first approximation—except for nomadic populations, the children and women of a child–woman ratio obviously must live in such areas. A more ambitious allocation model might further refine spatial placement by giving greater or lesser weight to within-region locations according to their level of

built-up (which ranges from low positive values to a theoretical maximum of 10,000 meters in a 100^2-meter raster cell). Such refinements are worth considering, if somewhat speculative given the thin current knowledge base linking demographic measures to measures of land cover. Of course, allocation models have their limits: There can be no more trustworthy variation in child–woman ratios than is assembled via on-the-ground surveys, censuses, and other demographic instruments.

6 Next steps

In producing the final paper for presentation at the meetings, we will provide analytic results based on the set of surveys and census micro-samples available as of January 2025. In the meantime, we will be consulting with colleagues at the United Nations Population Division, the Gridded Population of the World, and other major data-providers to understand how they perceive the groups of potential users of sub-national demographic data, and to receive critical feedback on our methods and research judgments.

References

[1] V. A. Alegana et al. "Fine resolution mapping of population age-structures for health and development applications". *Journal of The Royal Society Interface* 12.105 (2015), p. 20150073. DOI: 10.1098/rsif.2015.0073. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2015.0073. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2015.0073.

[2] Clara R. Burgert-Brucker, Trinadh Dontamsetti, and Peter W. Gething. "The DHS Program's Modeled Surfaces Spatial Datasets". *Studies in Family Planning* 49.1 (2018), pp. 87–92. DOI: https://doi.org/10.1111/sifp.12050. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12050. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/sifp.12050.

[3] CIESIN. *Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets.* Center for International Earth Science Information Network (CIESIN), Columbia University. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). 2018. DOI: 10.7927/H45Q4T5F.

[4] Fabrizio Colella et al. *Inference with arbitrary clustering*. IZA Discussion Paper Series No. 12584, IZA — Institute of Labor Economics, Bonn, Germany. 2019.

[5] T.G. Conley. "GMM estimation with cross sectional dependence". *Journal of Econometrics* 92 (1999), pp. 1–45.

[6] European Commission. *GHSL Data Package 2023*. JRC133256, Publications Office of the European Union, LuxembourgEuropean Commission, GHSL Data Package 2023, Publications Office of the European Union, Luxembourg, 2024, doi:10.2760/098587, JRC133256. Scientific contact: Thomas Kemper, European Commission - DG Joint Research Centre Space, Security and Migration Directorate, Disaster Risk Management Unit E.1. 2024. DOI: 10.2760/098587. URL: https://human-settlement.emergency.copernicus.eu/dataTools0verview.php.

[7] Juan Galeano et al. "CORESIDENCE: National and subnational data on household size and composition around the world, 1964-2021". *Scientific Data* 11.1 (2024), p. 145. ISSN: 2052-4463. DOI: 10.1038/s41597-024-02964-3. URL: https://doi.org/10.1038/s41597-024-02964-3.

[8] Matt Hauer, Jack Baker, and Warren Brown. "Indirect Estimates of Total Fertility Rate Using Child Woman/Ratio: A Comparison with the Bogue-Palmore Method". *PLOS ONE* 8.6 (June 2013), pp. 1–7. DOI: 10.1371/journal.pone.0067226. URL: https://doi.org/10.1371/journal.pone.0067226.

[9] Bradley Janocha et al. *Guidance and Recommendations for the Use of Indicator Estimates at Subnational Administrative Level 2*. DHS Spatial Analysis Reports 20, The DHS Program, ICF, Rockville, Maryland, USA. 2021.

[10] Peter Johnson et al. *Method protocol for the evaluation of census population data by age and sex*. United Nations Population DivisionUN DESA/POP/2022/TP/No.5. Dec. 2022.

[11] S. Leyk et al. "The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use". *Earth System Science Data* 11.3 (2019), pp. 1385–1409. DOI: 10.5194/essd-11-1385-2019. URL: https://essd.copernicus.org/articles/11/1385/2019/.

[12] Sara Randall and Ernestina Coast. "The quality of demographic data on older Africans". *Demographic Research* 34.5 (2016), pp. 143–174. DOI: 10.4054/DemRes.2016.34.5. eprint: https://www.demographic-research.org/volumes/vol34/5/34-5.pdf. URL: https://www.demographic-research.org/volumes/vol34/5/.

[13] Andrew J. Tatem et al. "Millennium development health metrics: Where do Africa's children and women of childbearing age live?" *Population Health Metrics* 11 (2013), pp. 1–11. URL: http://www.pophealthmetrics.com/content/11/1/11.

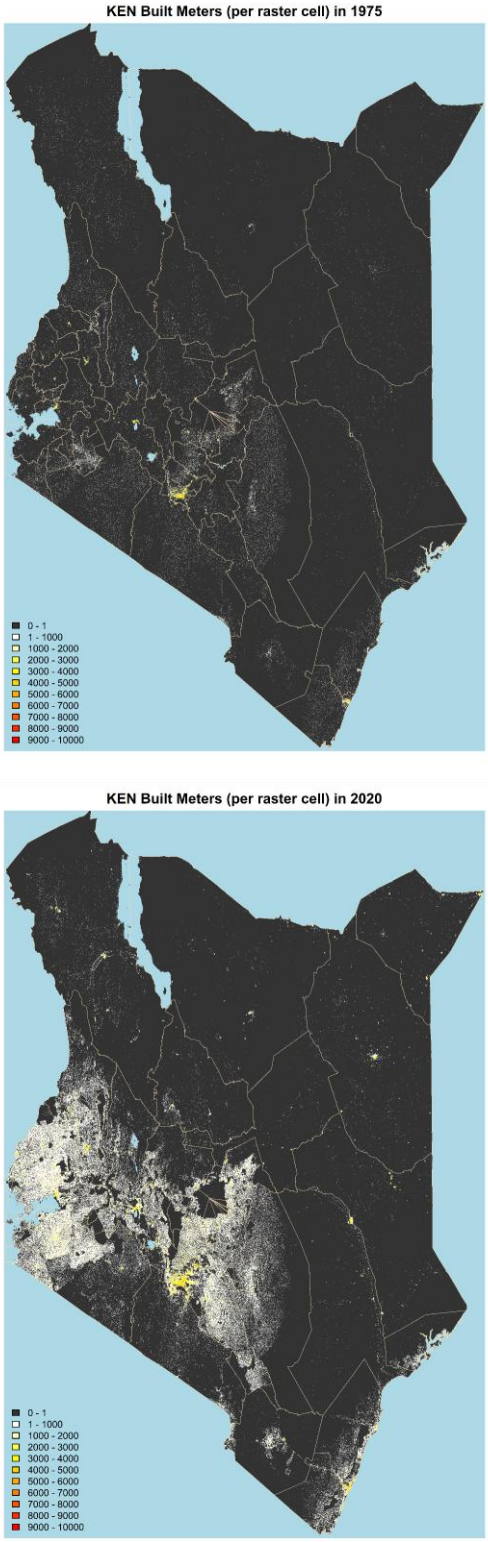

Figure 1: GHSL estimates of built-up land in Kenya per raster cell (each covering 100^2-meter units of habitable land), 1975–2020. Boundaries represent current level-1 administrative regions. Estimates of built-up for 1990, 2000, and 2015 are also available.

Figure 2: Countries with age and sex distributions for level-1 and level-2 sub-national administrative units. These are based on full-census tabulations by the U.S. Census Bureau, International Programs.

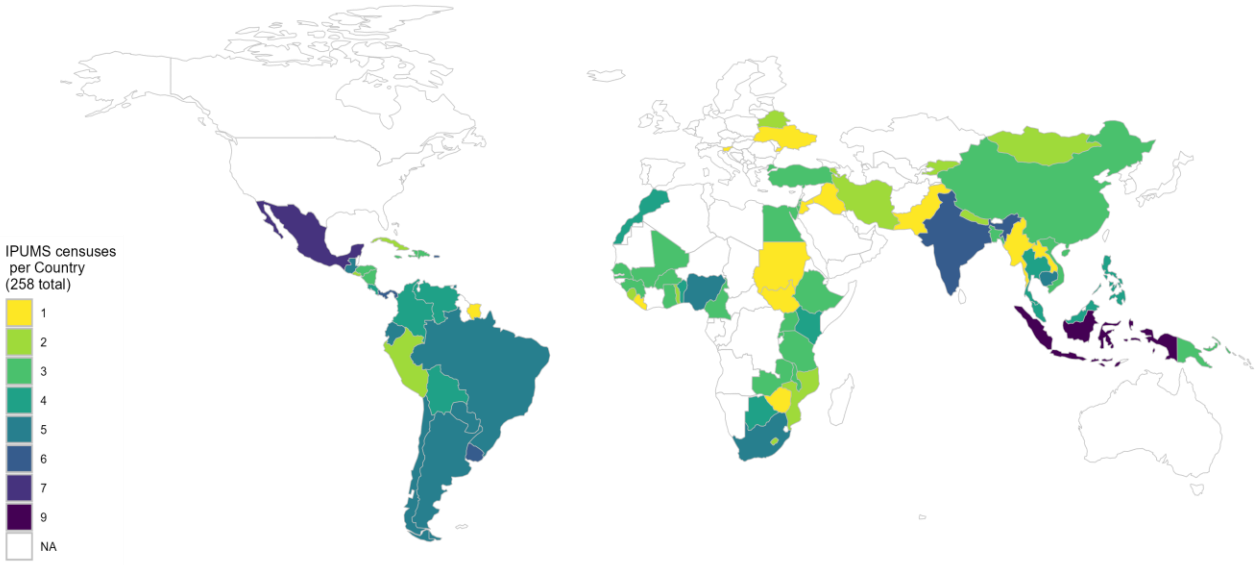

Figure 3: Countries with sub-national age and sex distributions from IPUMS-I, by the number of microdata samples per country. Each such sample is a random selection from full-census microdata records. The April 2024 data release is not yet fully incorporated in our holdings.

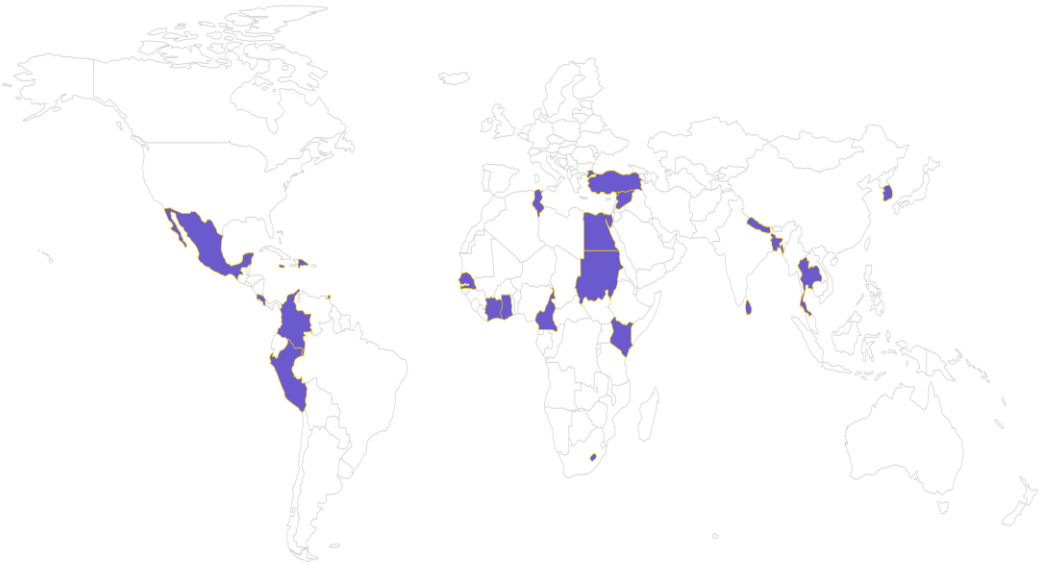

Figure 4: Countries with sub-national age and sex distributions from the WFS program, early 1980s. (Only 1 survey per country.)

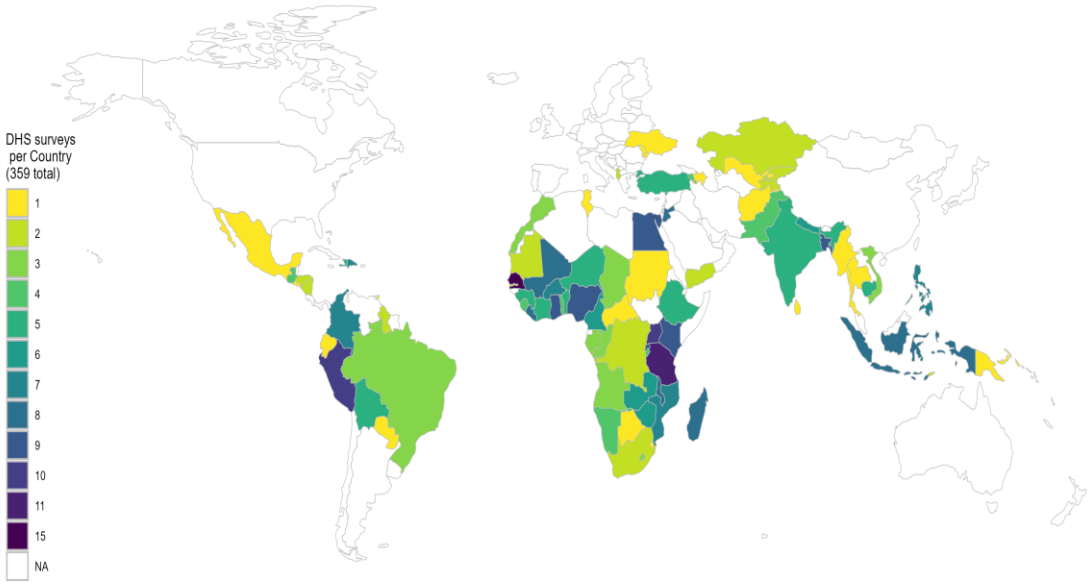

Figure 5: Countries with sub-national age and sex distributions from the DHS surveys, 1986–present, by number of surveys fielded per country

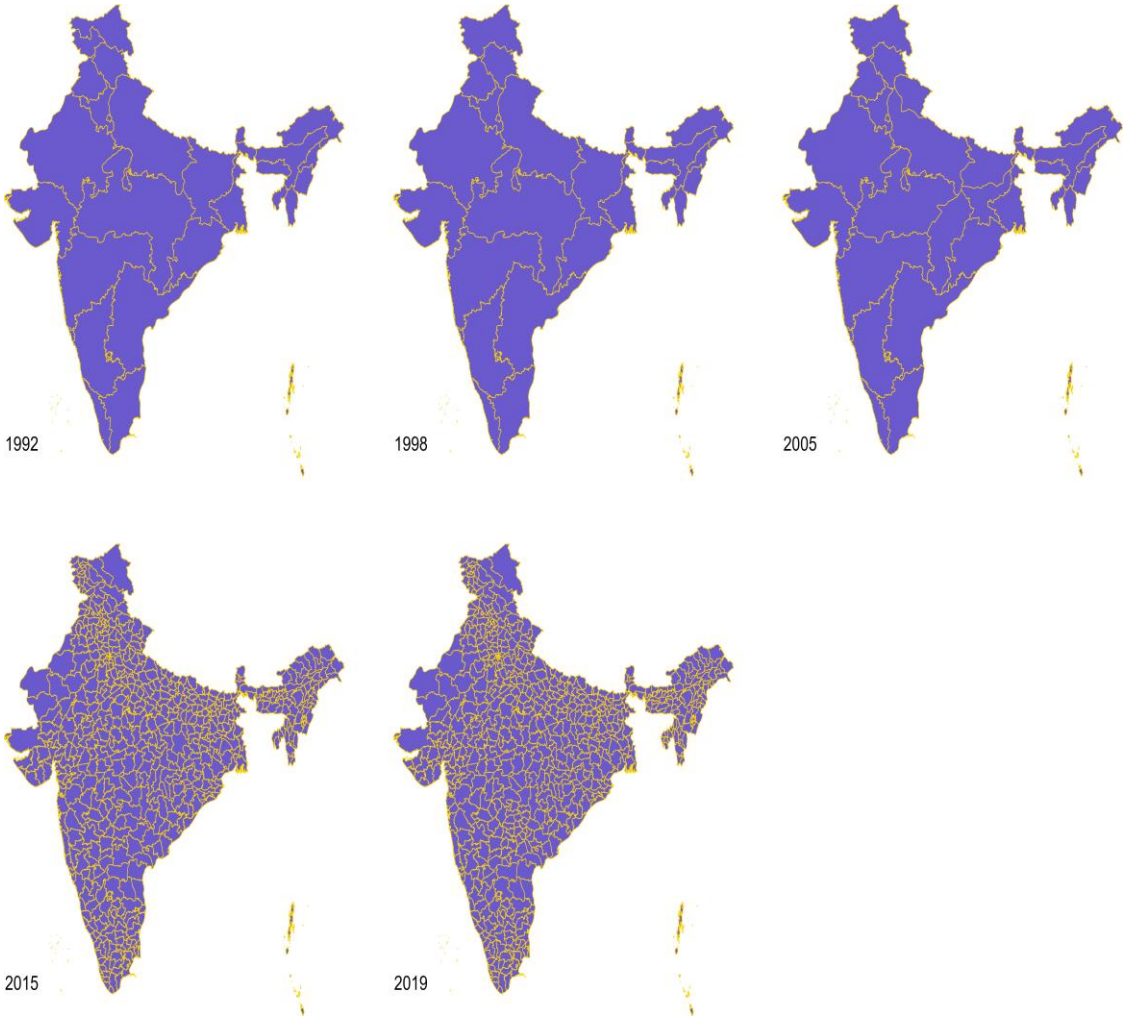

Figure 6: DHS surveys for India, displayed with sub-national boundaries from the *Spatial Data Repository*. District boundaries (i.e., second-level admin boundaries) can be obtained from other sources for the 1992 and 1998 surveys, but in the 2005 survey only the state of residence is identified in the survey, leaving no possibility for linkage to districts. New Indian states emerged in 2005 in conjunction with the 2001 census; additional new states and many new districts have come into being over the post-2011 census period. The 2015 and 2019 surveys were designed to be statistically representative at the district level.

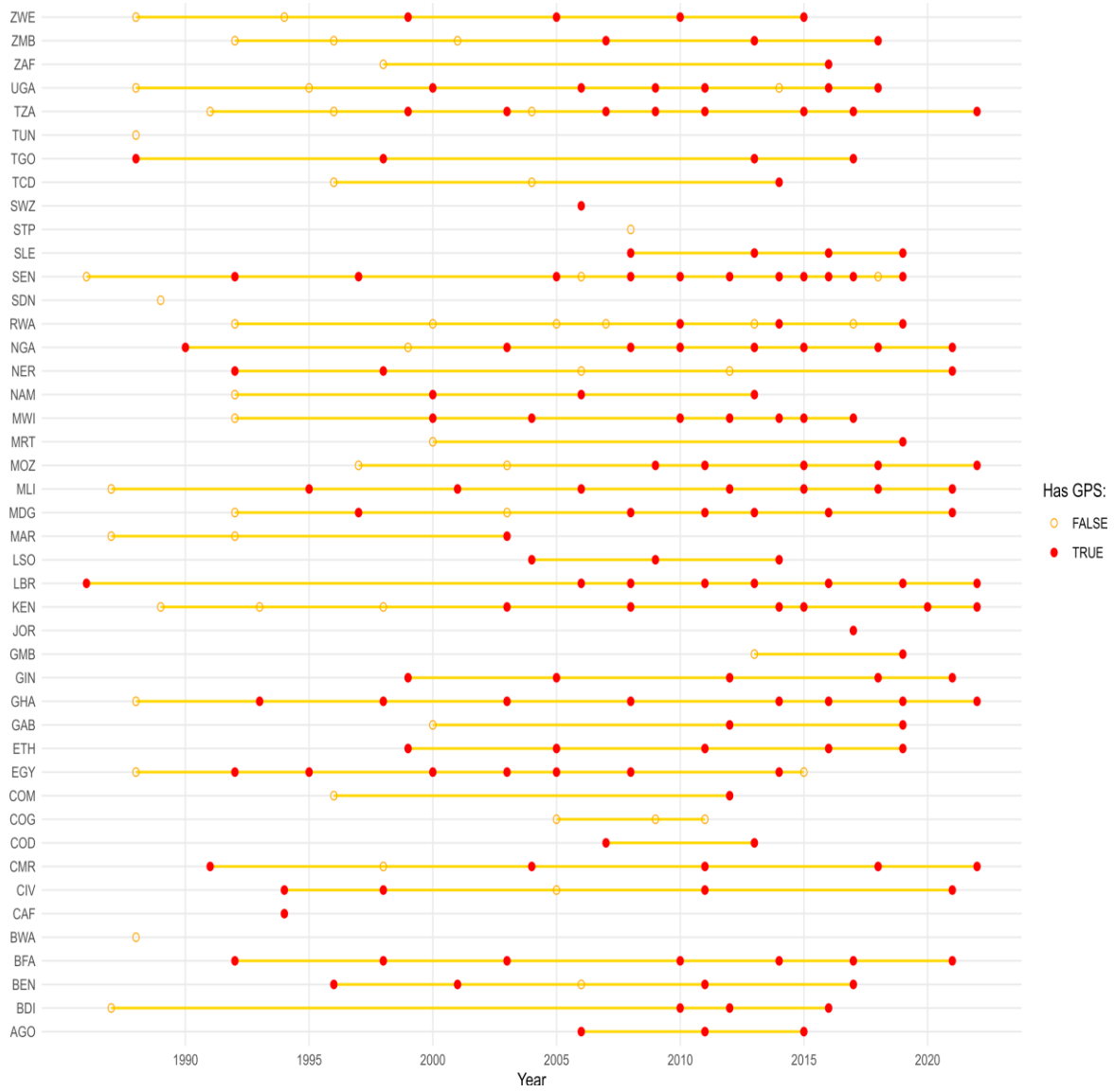

Figure 7: North African and sub-Saharan African DHS surveys with and without sampling cluster coordinates

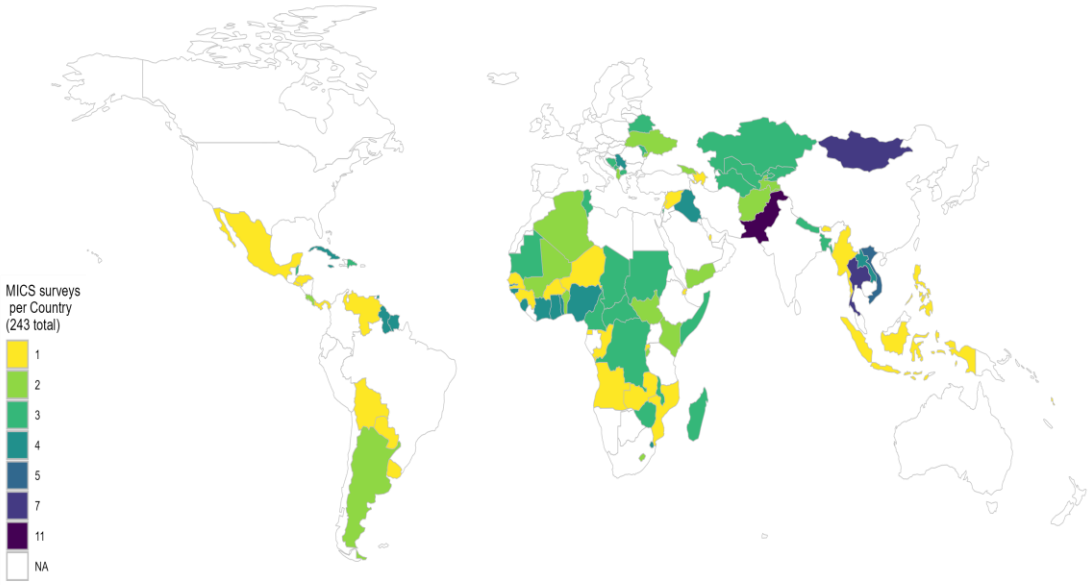

Figure 8: Countries with sub-national age and sex structure from the MICS program, 1999–present, by the number of surveys per country. The high count for Pakistan is misleading: In this country, the practice has been to spread MICS fieldwork across phases, each phase being devoted to one large province. The results are then released over a period of years on a province-by-province basis. For Pakistan, therefore, the count pertains to the number of provincial surveys that have been released.

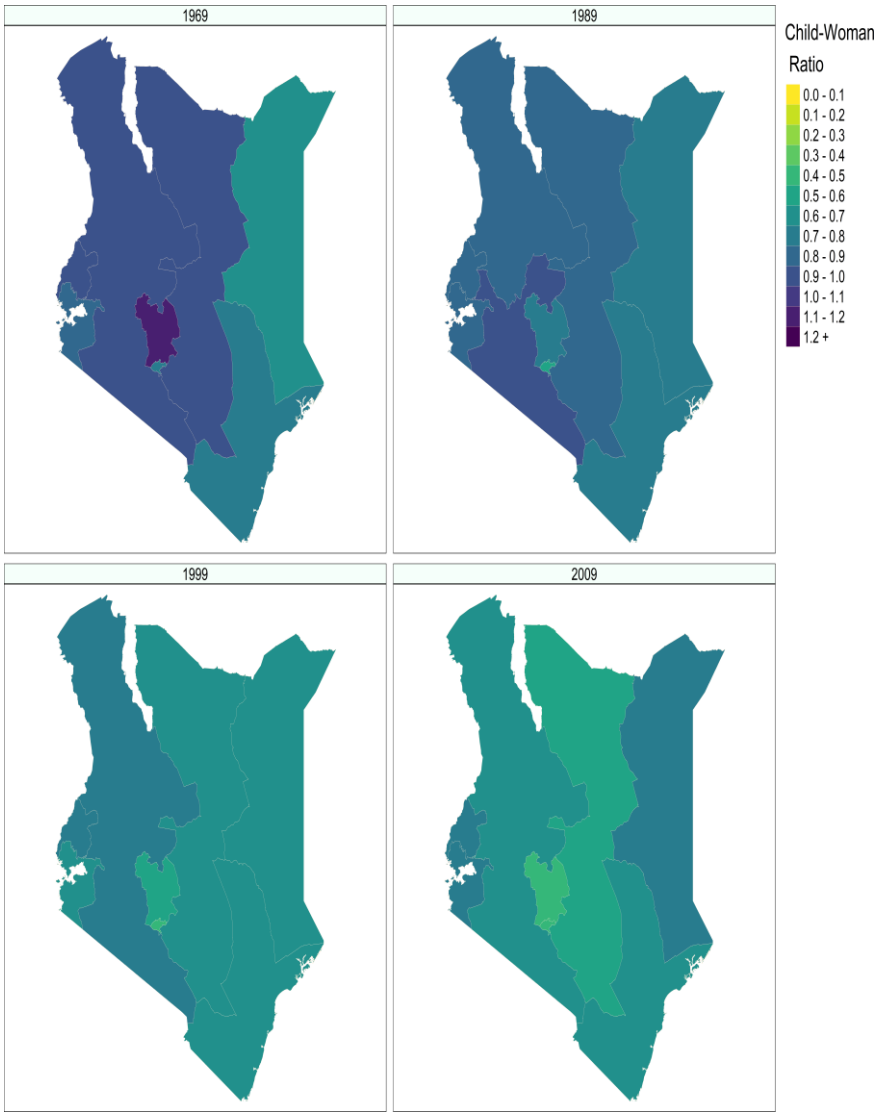

Figure 9: Child-woman ratios for Kenya from 1969 to 2009, by first-level administrative regions. The data shown are from IPUMS-I microsamples.

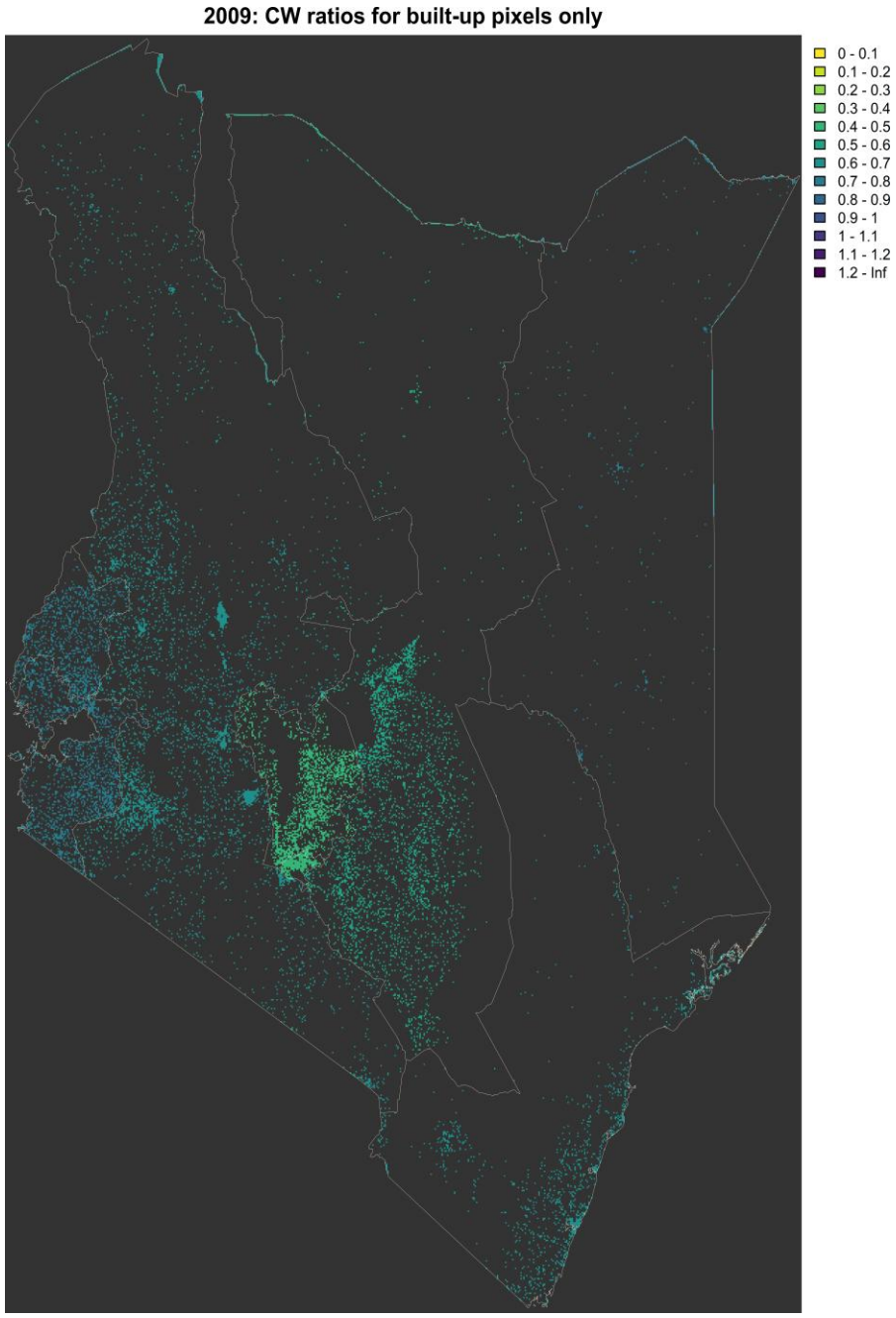

Figure 10: Child-woman ratios for Kenya in 2009, by region and raster cell within region. The ratios are first calculated for level-1 administrative regions (using IPUMS-I census microdata) and then, within each such administrative region, assigned spatially according to whether the raster cell (from GHSL) is built-up. In this approach, a demographic measure is allocated only to areas where there is evidence of human habitation.

Figure 11: Variation in child–woman ratios for Kenya *within* first-level administrative units, using IPUMS-I census microdata from 1969 to 2009.