Disaggregation of Statistics by Geography

Areal Interpolation as Paradigmatic Example

Eighth meeting of the Expert Group on the Integration of Statistical and Geospatial Information
together with the Seventh meeting of the Working Group on Geospatial Information of the IAEG-SDG

Nairobi, Kenya | Sept. 17th, 2024

BRYCE DAVENPORT (USA)

United States Census Bureau

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Why disaggregate?

"In the context of the ***Leave No One Behind*** principle of the 2030 Agenda, the need for disaggregated data is (i) to be able to identify vulnerable groups or populations that are most likely to be left behind, understanding the factors that keep them in or move them out of that position; and (ii) to report on the indicators for these groups to monitor their progress in achieving the development targets and goals."

- Practical Guidebook on Data Disaggregation for the Sustainable Development Goals, p.3

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

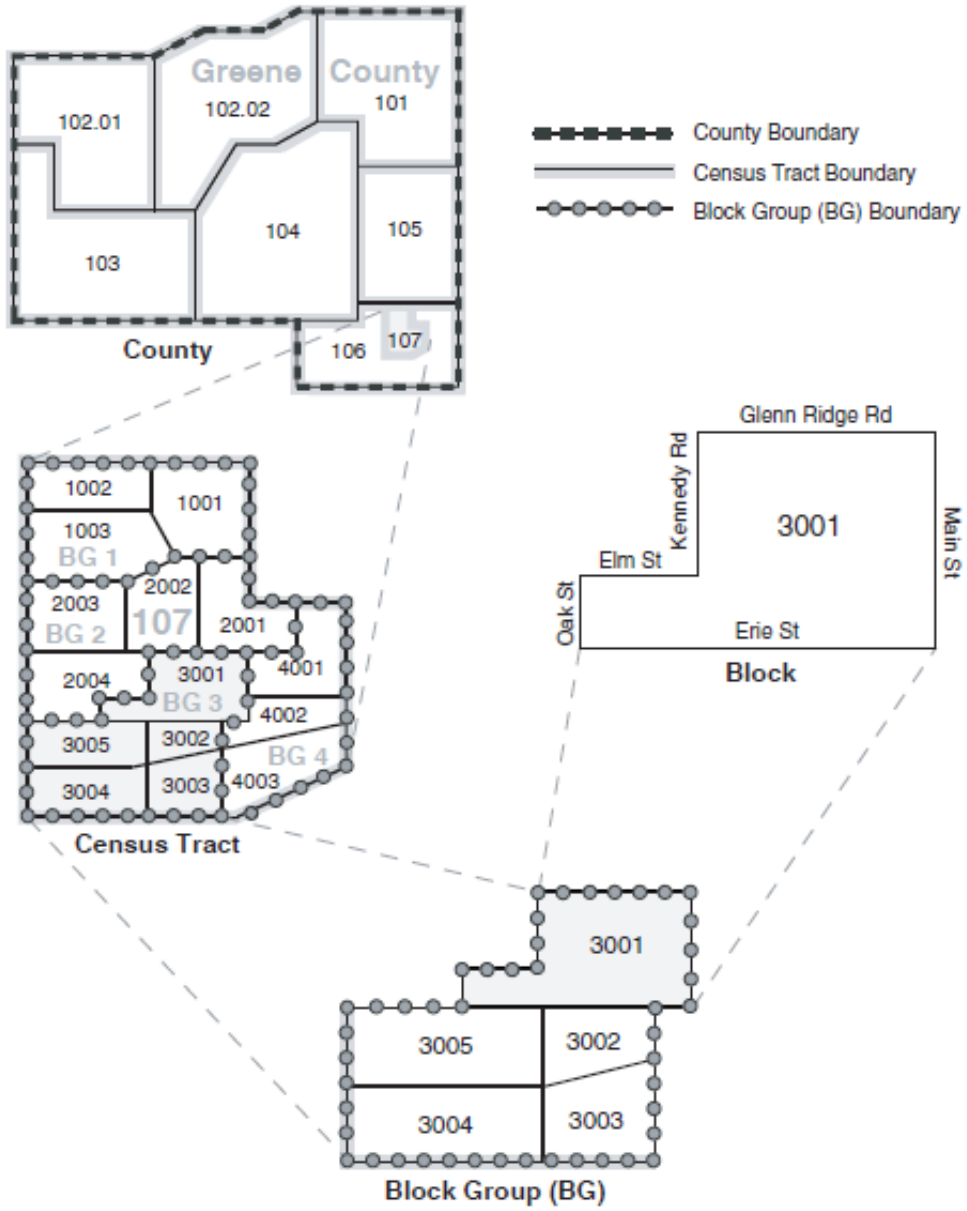
Why disaggregate?

- Production of small area estimates
- Demand for finer geographic detail of published statistics
- Data production for various subpopulations.
- Current methods do not provide large enough samples to produce reliable direct estimates for small areas such as counties or even most states.
- The use of valid statistical models can provide small area estimates with greater precision of statistical models;

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Case study:
US Census Blocks

- Small area geographies parameterized by population proxy, not area.
- Ideally formed as organic units, bounded by streets, rivers, or other visible features.
- Hierarchically integrated with all other US Census geographies – Census blocks are building blocks.
- Statistical records with location attributes geocoded to a common geography – what's left?



UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

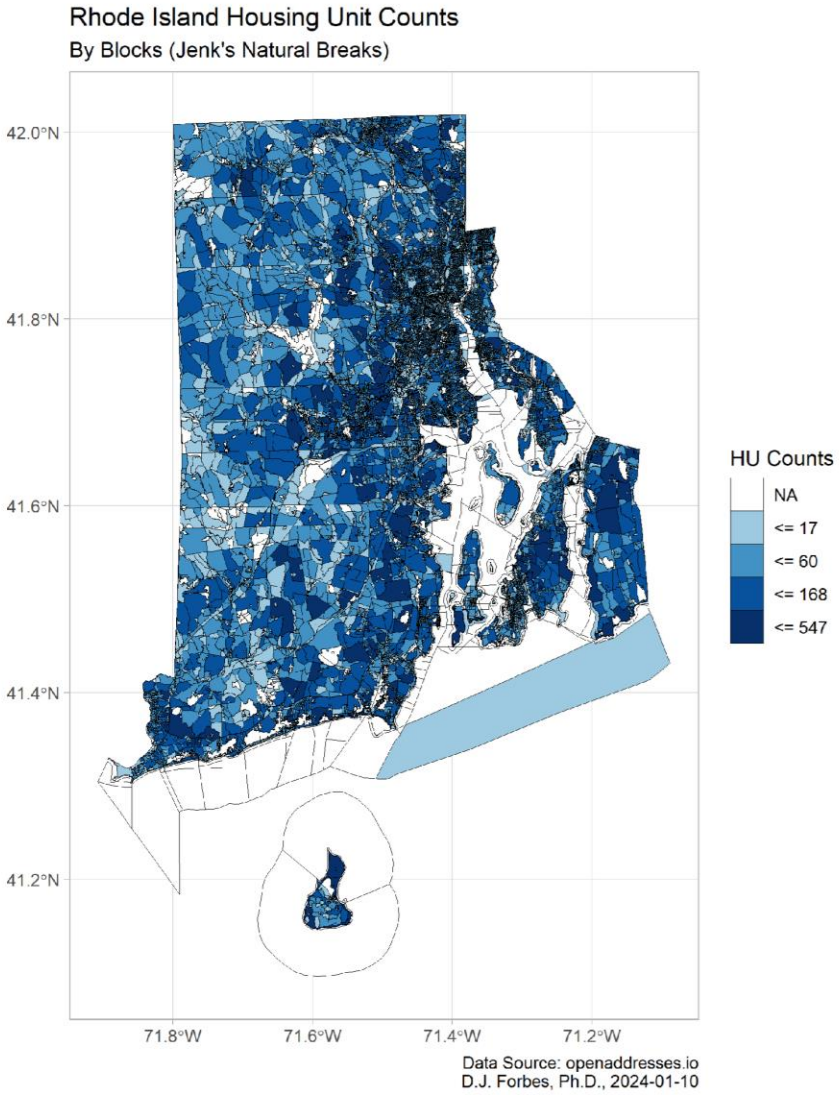
The Modifiable Areal Unit Problem (MAUP)

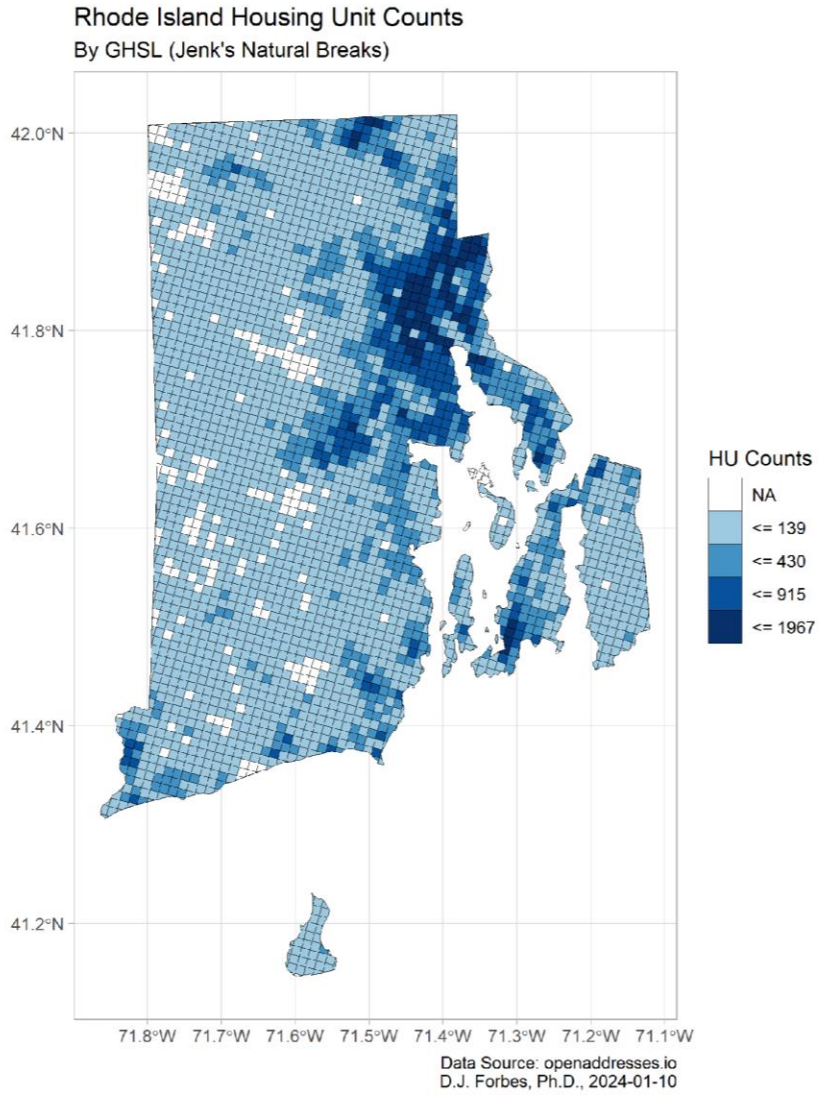
The results of spatial analysis can change when different aggregation methods are used:

- Differences in scale – parsing an area into a larger or smaller number of units can obscure attributes under study. Smaller units can offer more precision at the risk of amplifying locally salient variables outside of their importance to the larger population, but also risk disclosure issues.
- Differences in zoning – data collected in one geography, i.e. administrative units, may not map neatly into political boundaries. Sampling units that are fitted to population metrics and have variable area can also disguise density measures
- Variables of interest in fields that are consumers of statistical frequently have spatial patterning that is independent of the sampling units.

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Census Blocks - MAUP

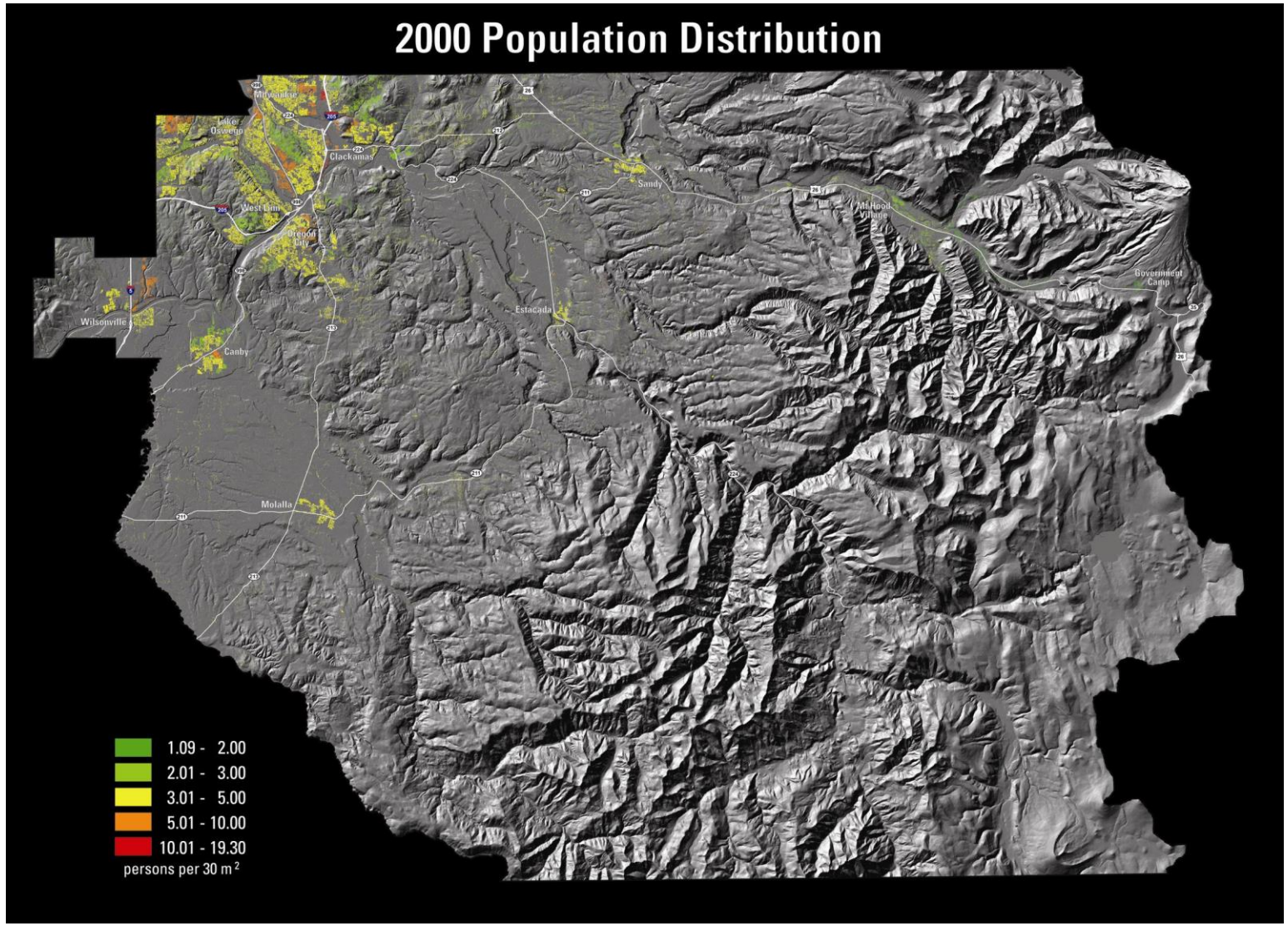




UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Geographic Disaggregation

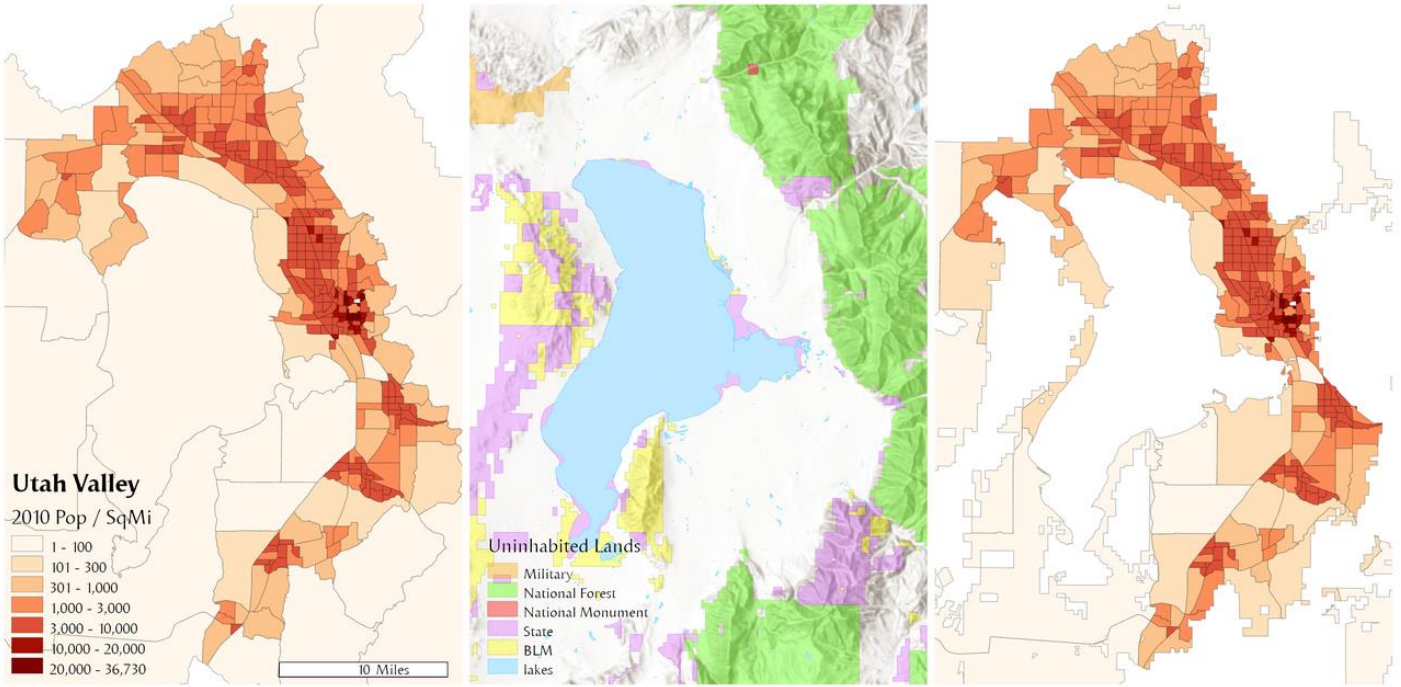
- Emergency planning and disaster relief agencies, among many others, often use remotely sensed earth observation data as foundational datasets – high resolution regular grids.
- Census blocks are variable in size and distribution – how do we bridge from our administrative geography to useful population density?



USGS Western Geographic Science Center,
Population Density over DEM Hillshade for Clackamas County, OR

Dasymetric Modeling

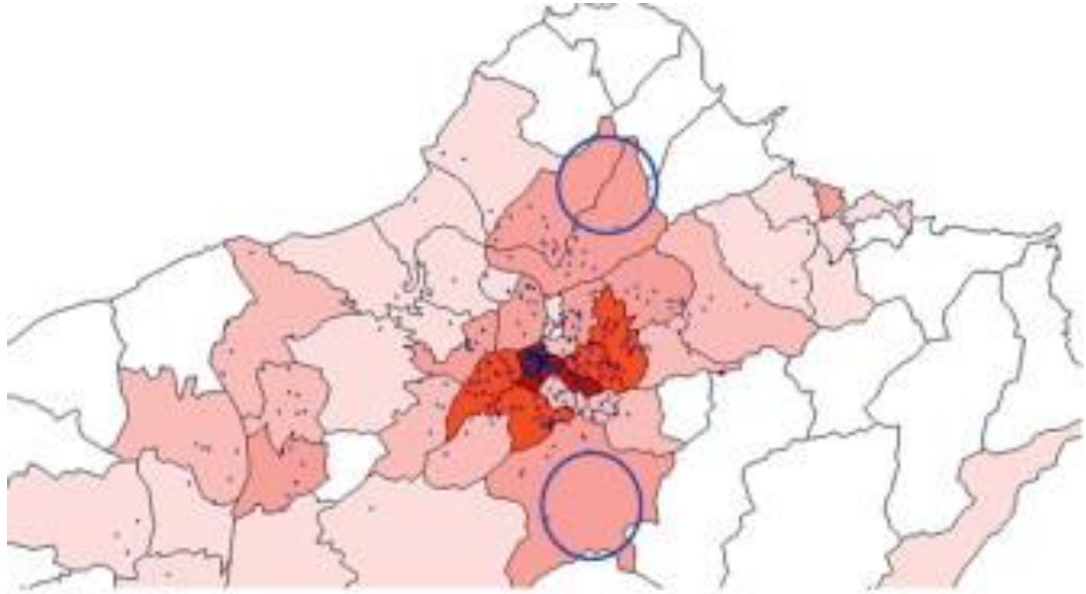
- Areal interpolation technique for mapping population density against ancillary datasets (frequently land cover).
- At simplest – removing areas known to be uninhabited from calculations.
- More sophisticated models distribute population attributes to grid cells using weighted values.



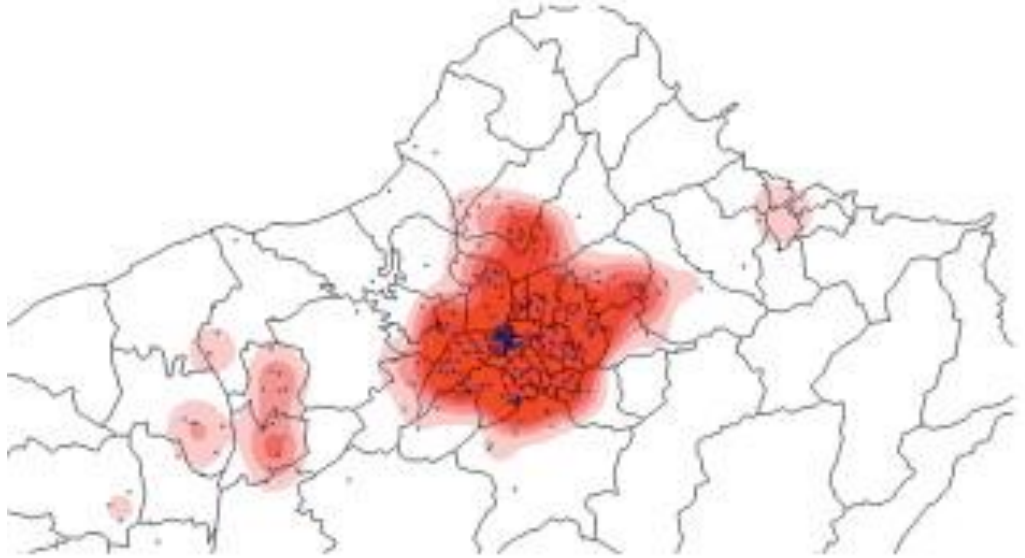
Wikimedia Commons:
Utah Valley dasymetric map - Dasymetric map - Wikipedia

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Decoupling observations from geographic units (without releasing protected data)



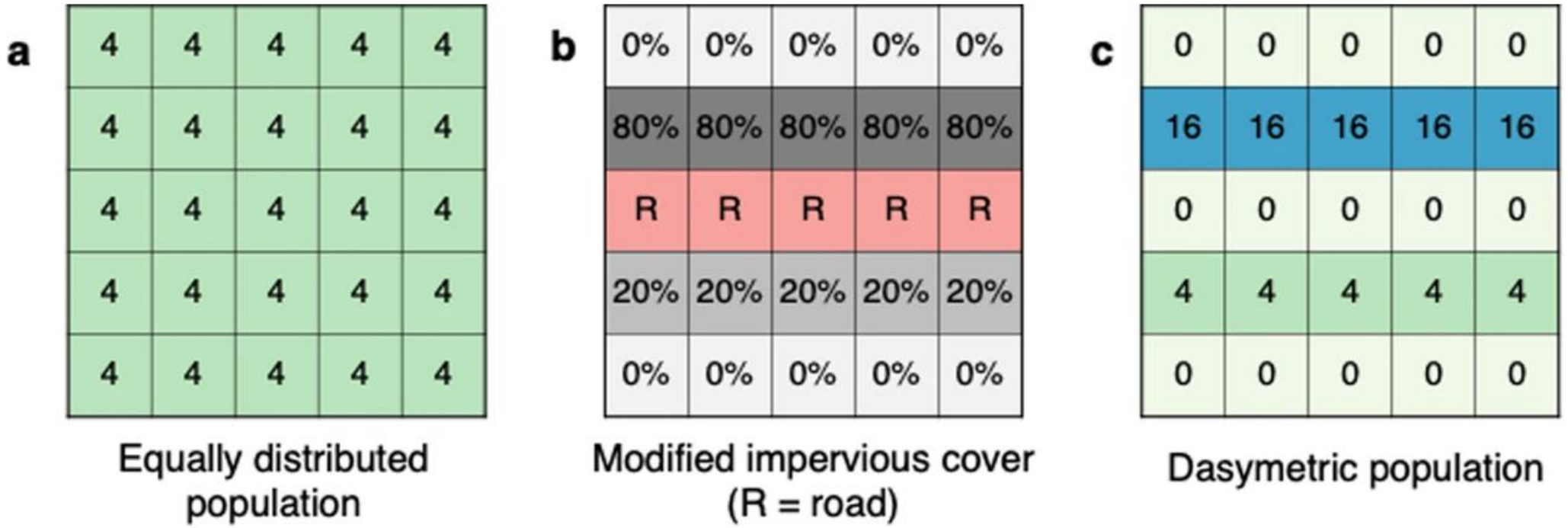
(a) Choropleth map of SARS cases in Taiwan (2003)



(b) Density map of SARS cases in Taiwan (2003)

Fig 6. from M.D. Su, M.-C. Lin, T.H. Wen. Spatial Mapping and Environmental Risk Identification. In Encyclopedia of Environmental Health, Editor(s): J.O. Nriagu. Elsevier, 2011, Pages 228-235

Case Study: Land cover as density classifier



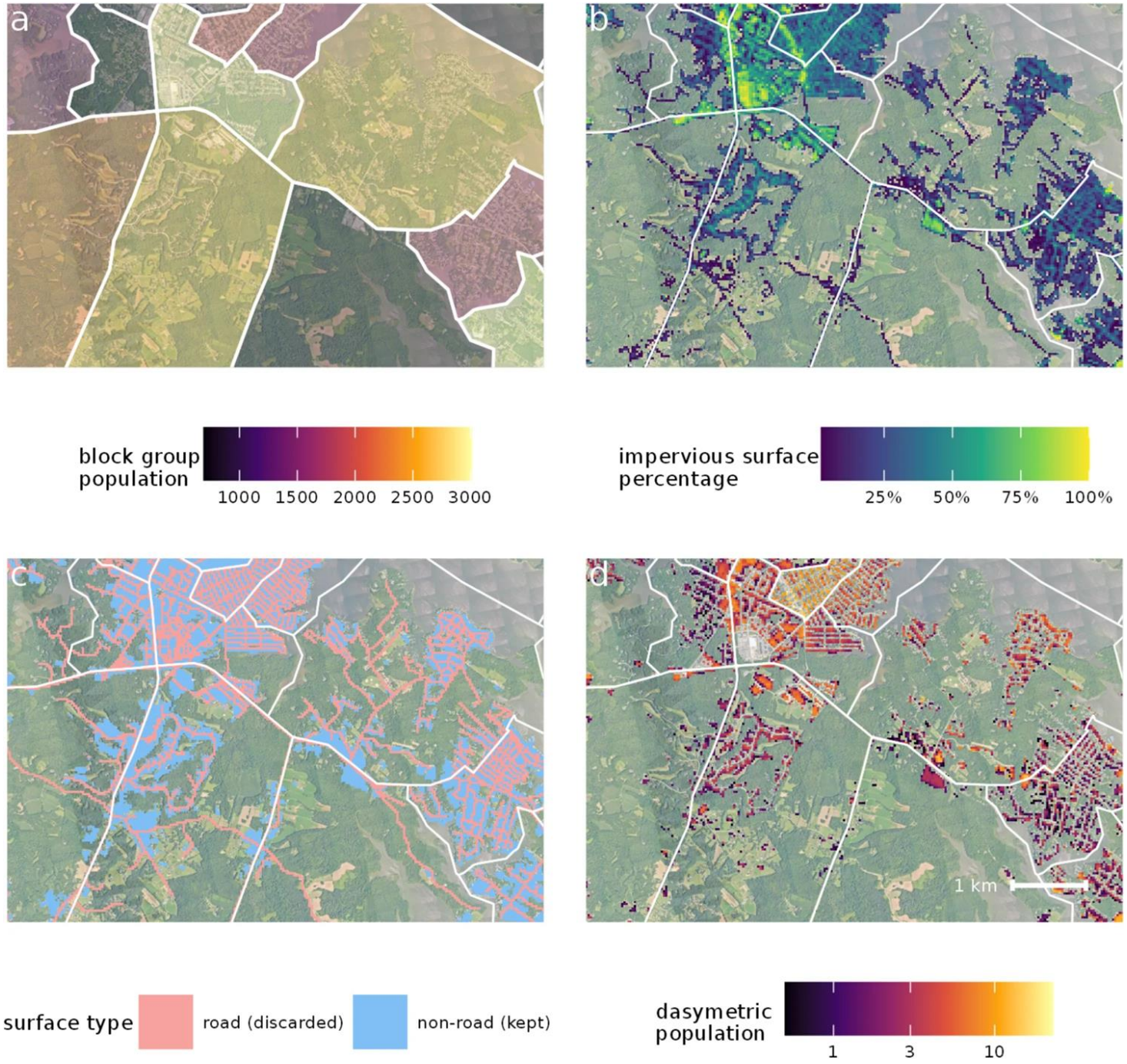
Dasymetric population mapping based on US census data and 30-m gridded estimates of impervious surface. Rachel H. Swanwick, Quentin D. Read, Steven M. Guinn, Matthew A. Williamson, Kelly L. Hondula & Andrew J. Elmor. Scientific Data volume 9, Article number: 523 (2022). Fig. 2.

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Population Density as a Geospatial Dataset

- Disaggregation techniques scale in utility with sophistication of geospatial infrastructure - additional ancillary layers, geocoding down to the level of point coordinates, et c.
- With sufficient disclosure avoidance methods, these techniques allow aggregating up from raw point data for user-defined geographies.
- Highlights spatial relationships that may previously been obscured.

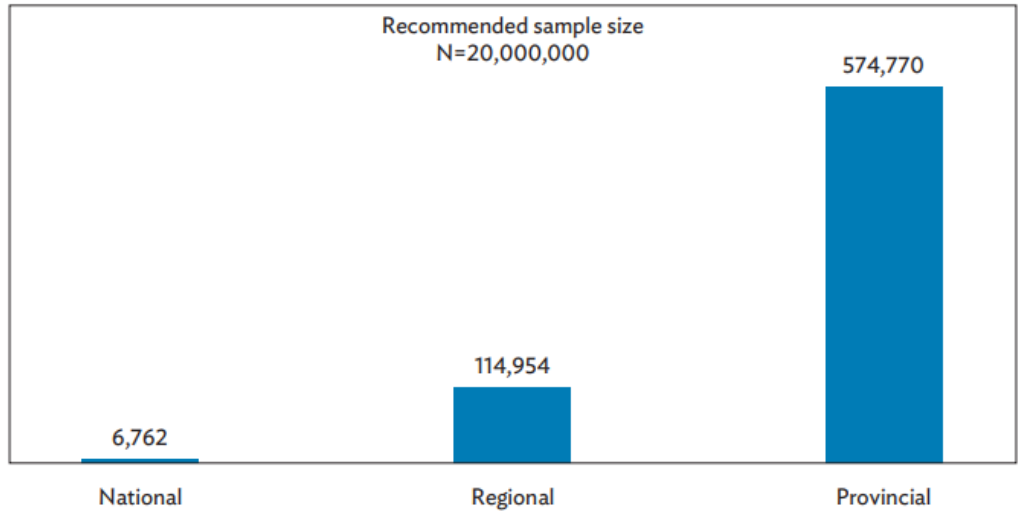
Rachel H. Swanwick, Quentin D. Read, Steven M. Guinn, Matthew A. Williamson, Kelly L. Hondula & Andrew J. Elmor. 2022 *Dasymetric population mapping based on US census data and 30-m gridded estimates of impervious surface.* Scientific Data volume 9, Article number: 523 (2022). Fig 1.



Representative Samples

- When modeling demographic attributes, smaller scale geographies may not have these characteristics available in published materials – cost, disclosure avoidance, and survey instrument design may all be factors.
- Ancillary datasets can add constraints to spatial reallocation of survey data to smaller scales, estimating sampling weights that better approximate population values.
- Techniques such as Penalized Maximum Entropy and Expectation Maximization can quantify the uncertainty of allocated estimates in comparison to the original dataset (which may contain its own uncertainty).

Figure 3.1: Recommended Sample Size for Different Levels of Geographic Disaggregation



Source: Adapted from Asian Development Bank. 2020. *Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistics Offices*. Manila. Available from https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf.

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION

Validation of Estimates

- Internal validation withholds some input data from the model for later comparison – danger of not being independent test, but for many applications internal validation is the only available
- External validation compares outputs to another survey similar scope and design – here, estimates from the Bangladesh Demographic and Health Survey, designed to be representative of the country's 7 primary administrative divisions, have been reallocated to map to the 64 second-order districts. The results of the Penalized Maximum Entropy Dasymetric Model agree favorably with the results of the 2011 National Census.

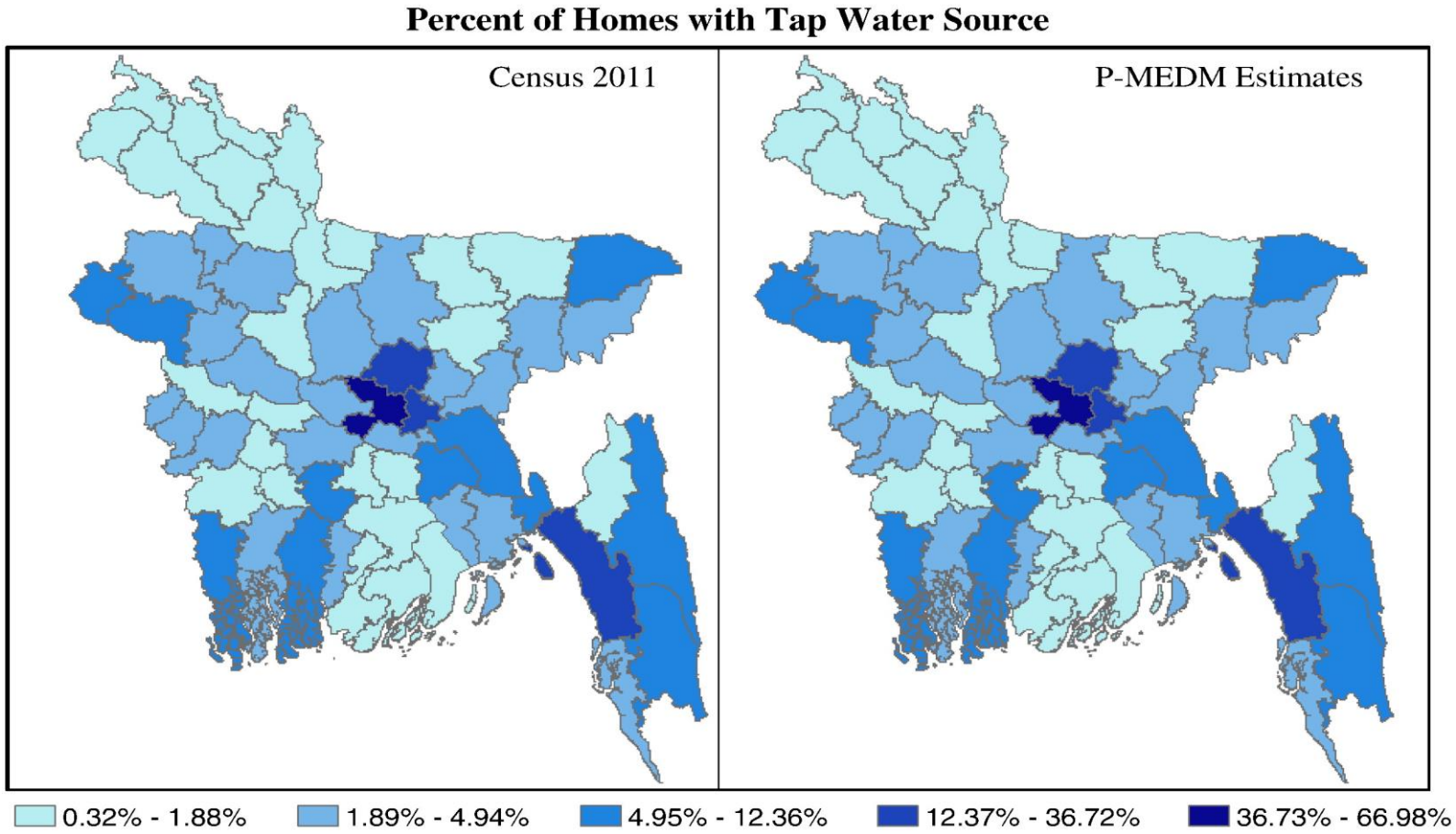


Fig. 5 from Amy N. Rose, Nicholas N. Nagle. Validation of spatiodemographic estimates produced through data fusion of small area census records and household microdata. Computers, Environment and Urban Systems, Volume 63,2017, Pages 38-49, https://www.sciencedirect.com/science/article/pii/S0198971516301338

Motivations

Areal interpolation methods of disaggregation

- Directly address the needs of SDG localization by providing not only the required granularity but by the inclusion of ancillary datasets that complement Official Statistics and give more accurate spatial patterning.
- Accelerate the production of geospatially integrated statistics, even in circumstances where the foundational geospatial infrastructure is not yet fully realized.
- Cultivates work pipelines that necessitate close collaboration between NSOs and NGIAs.
- The geospatial community has a sense of “ownership” over this domain – we have the expertise to promote this within statistics circles as a productive way of approaching Small Area Estimates.
- Powerful visualization techniques provide a throughline to policy and decisionmakers.

UN·EG-ISGI
UNITED NATIONS • EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION