Urban Poverty Mapping: Data and Methods

Alex de Sherbinin
CIESIN, The Earth Institute at Columbia University
Population-Environment Research Network (PERN)
adesherbinin@ciesin.columbia.edu

PRIPODE Workshop on Urban Population, Development, and Environment Dynamics
11-13 June 2007, Nairobi, Kenya

Population-Environment Research Network

CIESIN Columbia University

The Earth Institute AT COLUMBIA UNIVERSITY

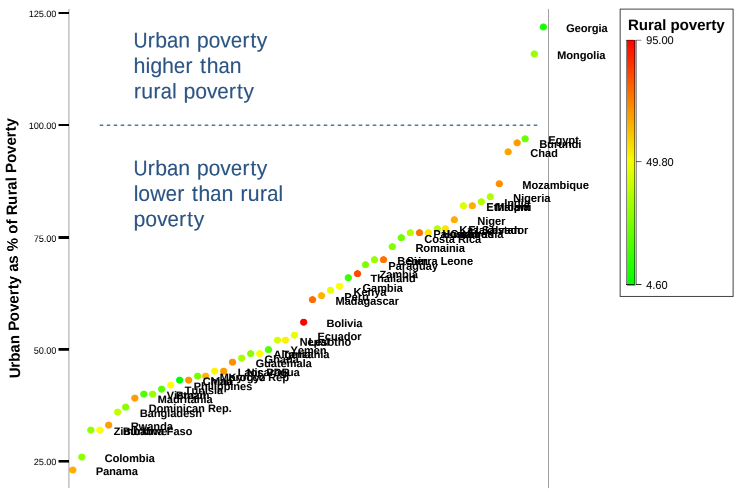
Presentation Outline

- Measurement of poverty and well-being
- Advantages of high spatial resolution poverty data
- Sources of urban poverty data
- Statistical issues
- Research examples

Measurement of Poverty and Well-Being

Current State of Global Poverty?

- Close to 1 billion people (1/5 world's population) lives on less than $1 day (estimate is 985.5 million as of 2004)
  - Good News: Some parts of the world are on track (East Asia) or have met (China) MDG 1 target for halving percentage of population living on less than $1 day.
  - Bad news: Others are at severely risk of failing (Sub-Saharan Africa).

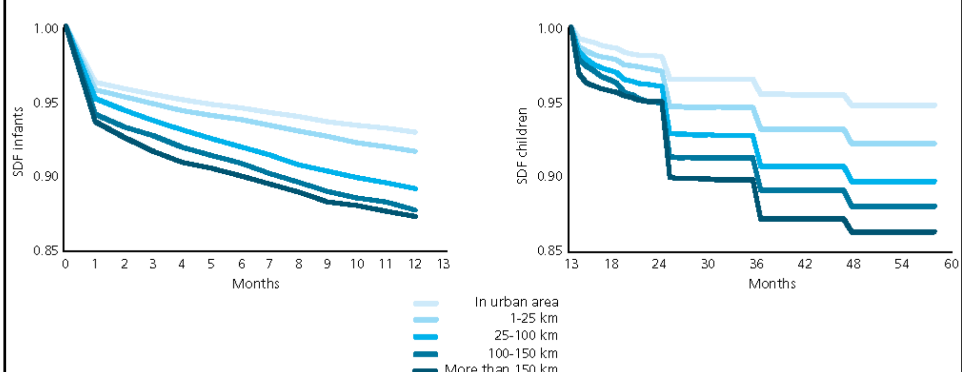


There is great variation in urban-rural poverty differentials globally, but, at a national scale, urban dwellers are generally better off.

Still, there are spatial pockets of poverty *within* urban areas

Analysis of infant and child mortality

For both infant and children, the chances of survival decrease monotonically the future one resides from a city (of 50,000 persons or more), in a 10-country West Africa study



Steps to Measure Poverty

- First define an indicator of welfare
  - Most often measured according to income or consumption levels (expenditures on goods and services).
  - Data Typically Collected using Household Surveys
- Establish a minimum acceptable standard of welfare to separate the poor and the non-poor (often known as a poverty line)
- Poverty lines vary in time and place, and each country uses lines which are appropriate to its level of development, societal norms and values.

How do you determine the poverty line?

- Absolute Poverty Line: Determine amount of income or consumption expenditures necessary to meet minimum standards of living (food requirements, plus basic non-food expenditures such as housing, transportation and clothing)
- Relative: Use average standards of income or consumption as a yardstick with which to compare welfare of others

Other Measures of Well Being

- Household income/consumption expenditures
  - may not necessarily reflect attainment of minimum standards of well being
  - Food expenditures are an input, but the actual desired outcome is adequate nutrition. Better to measure nutrition directly
- Non-monetary indicators of well being
  - Malnutrition
  - Unsatisfied Basic Needs
  - Infant Morality Rates

Malnutrition

- Requires establishing a baseline measurements related to body size and composition (weight for age, height for age, or weight for height) against which to judge whether individual is malnourished.
  - A child is considered malnourished if any of these indices falls below two standard deviations of the median value of international reference population. Severe malnutrition is when the indexes fall below three standard deviations.

Unsatisfied Basic Needs Indicators

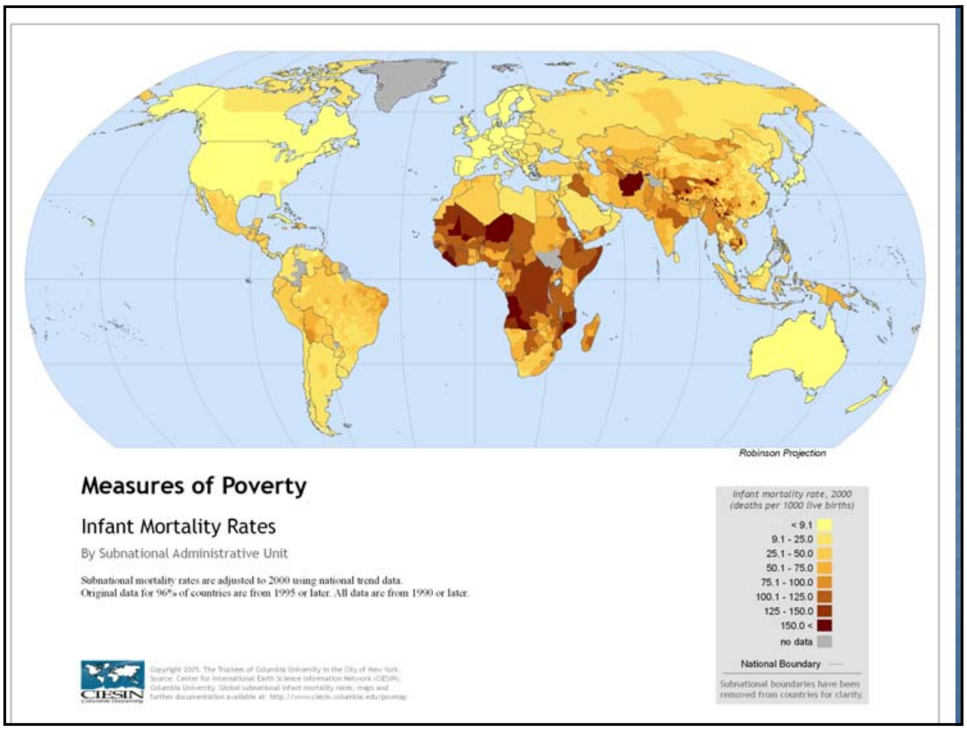
- Household possessions of goods or services that are associated with well-being are used to measure poverty
- The basic needs indicators are constructed by combining census level household measures such as access to adequate housing conditions, water, electricity, sanitation, and education, into a composite indicator of wellbeing for small administrative areas
- Typically UBN indicators are reported for administrative units as the proportion of households that have one, two, three or four basic needs unmet

Pros and Cons of UBN

- Pros
  - The indicators can be calculated from typical census data
  - Yields geographically disaggregated data for areas as small as the census agency is willing to report
  - Indicators can be compared across countries
- Cons
  - Users of UBN are limited by the time between census dates, usually 10 years
  - UBN measures are inputs to poverty, not outcomes

Infant Mortality Rate: number of infant deaths per one thousand live births

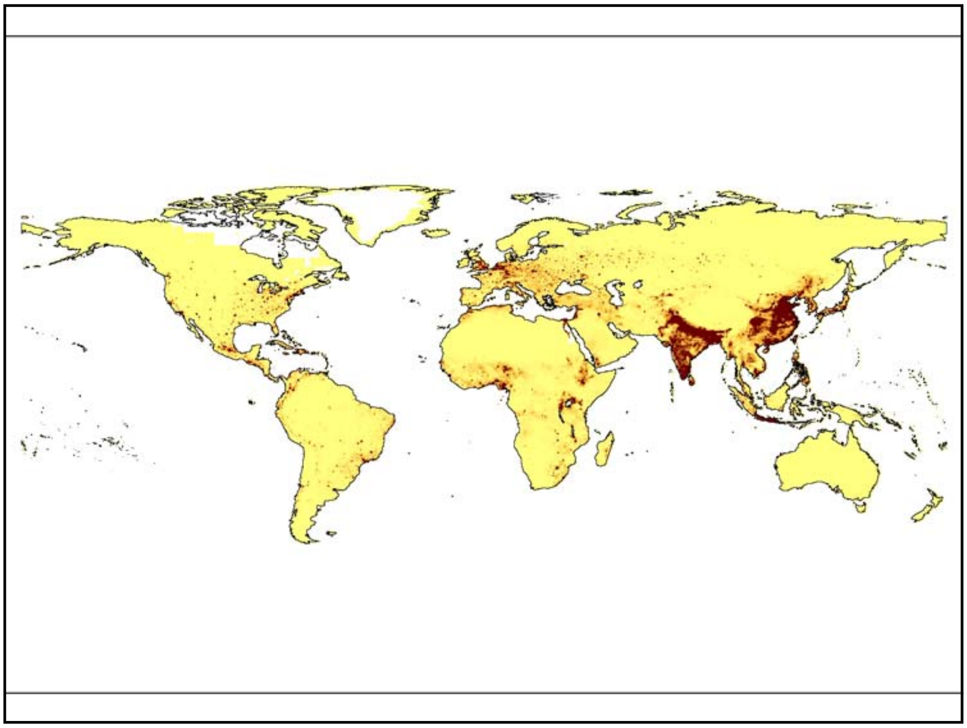
- Pros
  - Data availability: National infant mortality rates are reported for nearly every country in the world.
  - Definitions and measurements associated with infant mortality (i.e., deaths and live births) are well standardized across countries
  - Established Methods to methods to adjust national rates to account for reporting and definitional differences across countries
- Cons
  - The indicators typically calculated in developing countries from survey data (not vital registration systems)
    - Surveys estimating infant deaths require large samples, because such incidences are uncommon and representative households cannot ordinarily be identified for sampling
    - Frequency of surveys: Every 2-3 years
  - Measures only one dimension of poverty
  - Since it is a rate, it is not straight forward to develop a head-count

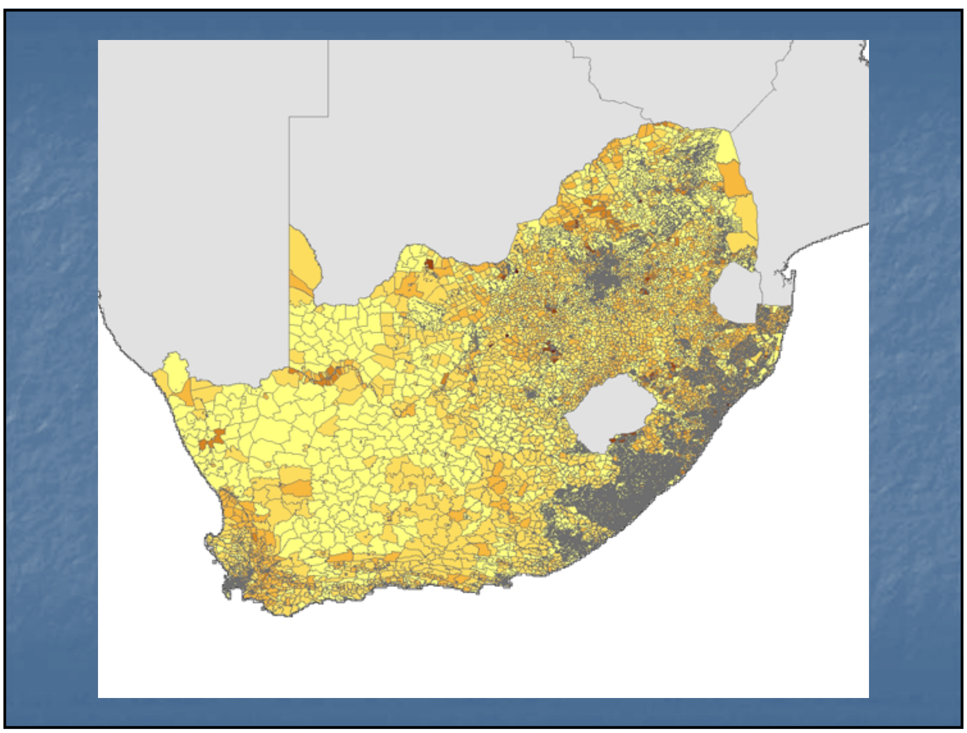


The Advantages of Higher Resolution Socioeconomic Data

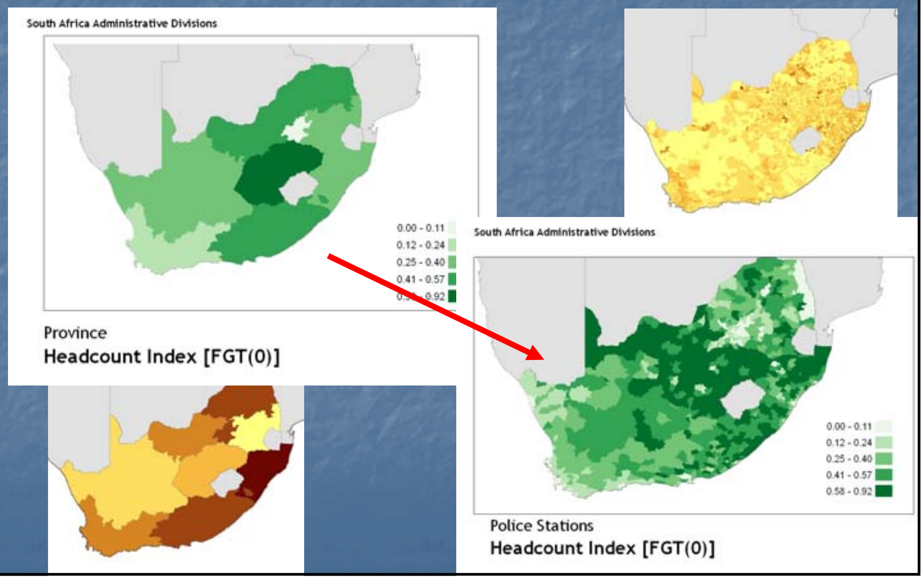
Why are high spatial resolution socioeconomic data desirable?

- More efficient policies, strategies, and resource allocation
  - Disaggregated data can help to identify highest priority areas
  - If high priority units are clustered, the strategy will be different than if they are spread out
- Improved data for research purposes (understanding the correlates of poverty)
  - Can utilize high resolution poverty data in combination with high resolution biophysical, infrastructure (roads, health/education facilities), or other data

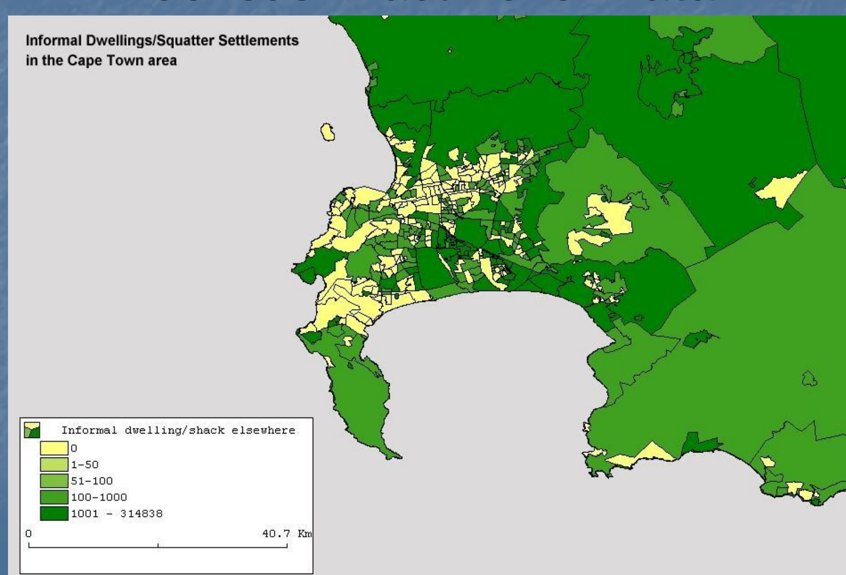


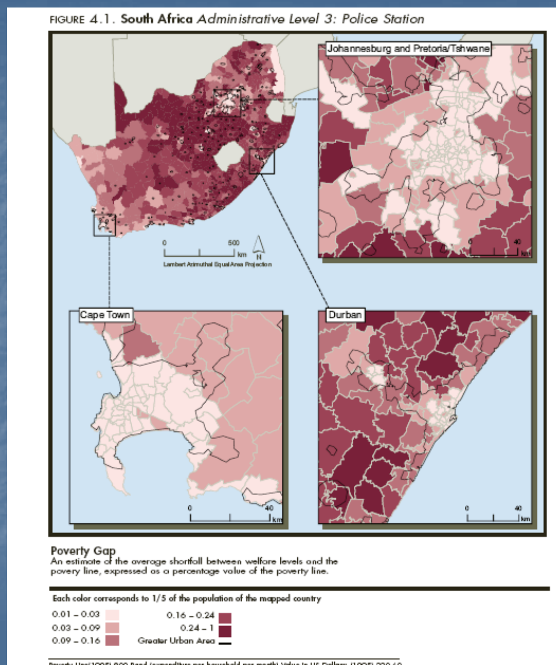


Small Area Estimates of Poverty

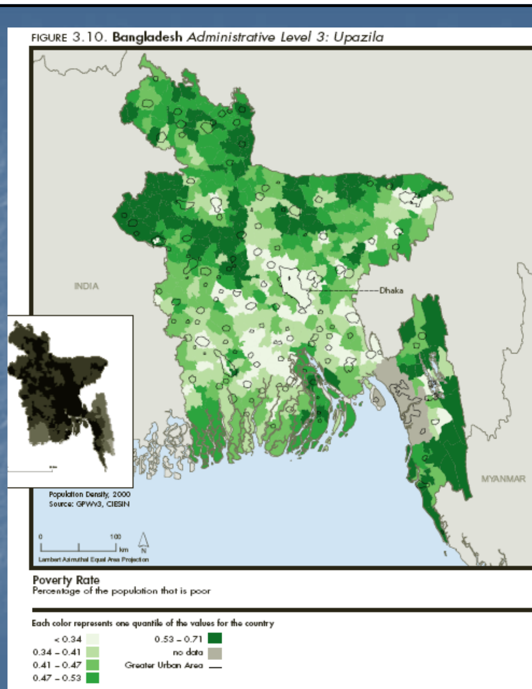


Census Tract Level Data

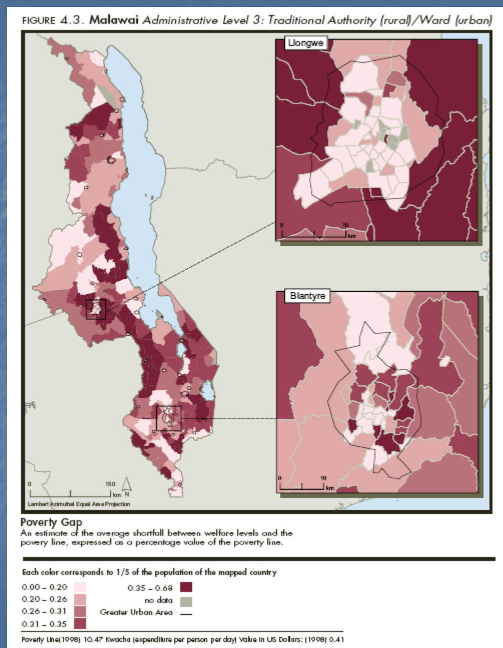




The urban-rural poverty differential that applies to a country as whole is not necessarily reflected uniformly across all urban-rural gradients within the country



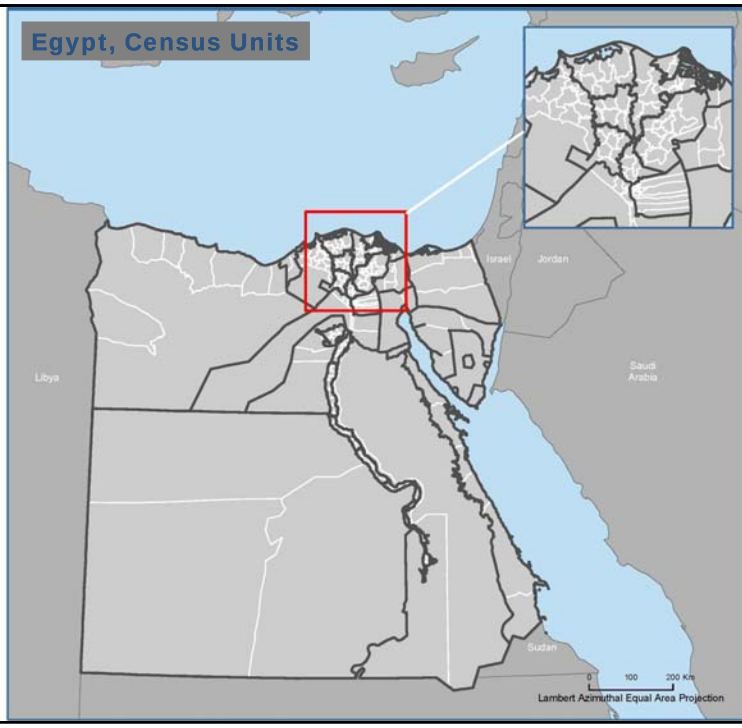
In Bangladesh, the pattern of poverty rates is primarily shaped by proximity to the capital Dhaka. In this map we can see that in general, poverty rates rise as one moves increasingly far from Dhaka. We can also see how the coastal remote areas are less disadvantaged than the inland remote areas. For the country as whole, the urban areas tend to be less poor than their rural counterparts.



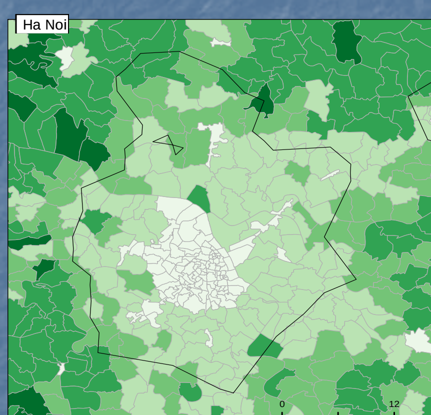
This map compares poverty gaps in the capital, Llongwe, and Blantyre. Both cities are comparable in population (approximately a half million). Llongwe has far less poverty within its limits, but is surrounded by regions of very high poverty. Blantyre, by contrast, has very high poverty within its limits, but is surrounded by regions of only moderate poverty.

Comparisons between Regions

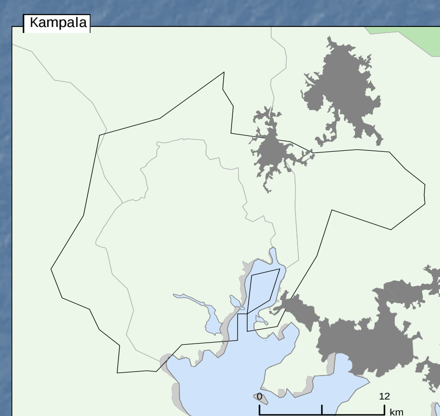
- Geographic resolution of data may be different within or between countries
  - Census units are generally delineated based on population size – units are smaller, more detailed, in urban areas vs. rural areas
  - Detail of survey / census data differs between countries
- Poverty lines will differ depending on urban vs. rural residence



Non-Uniformity Size of Units, Across Countries

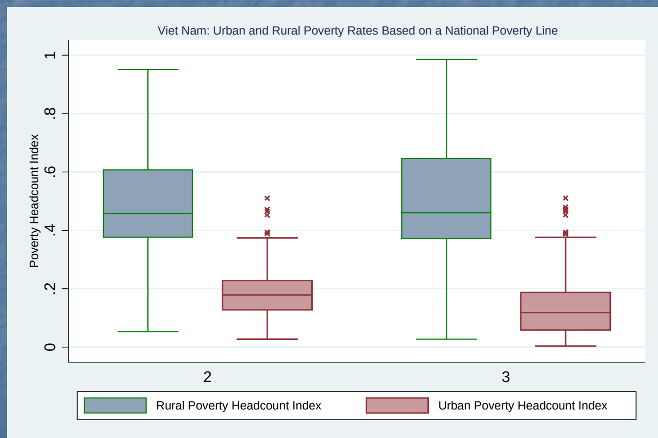


Intra-urban definition in data

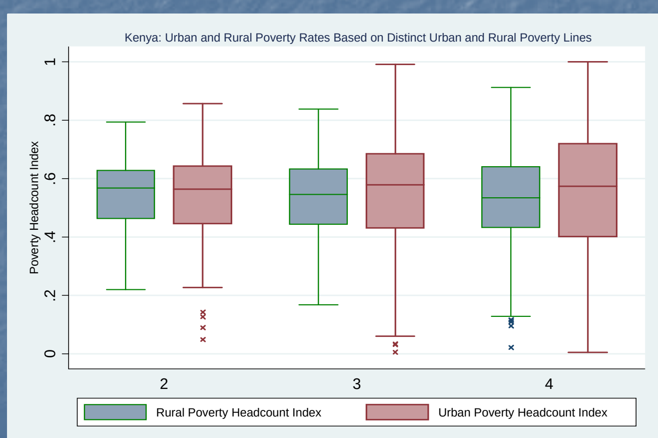


No intra urban information available

One national poverty line with information on urban and rural from survey respondents



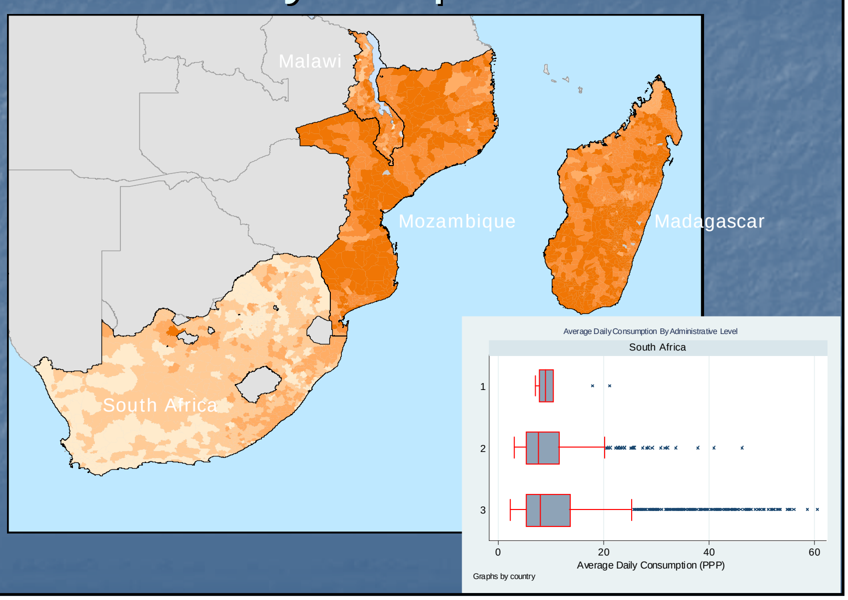
Separate urban and rural poverty lines



Comparisons between Regions

- Comparing measures of poverty between countries
  - What can be compared
    - Consumption: *if* defined the same way and measured in PPP
    - Inequality: relative measure
  - What is more difficult to compare
    - Poverty rates: different poverty lines, different inputs into defining the poverty line.

Country Comparison



Urban Poverty Data Sources

Other sources of information on urban populations

- Censuses
- Household surveys (DHS, MICS, LSMS)
- Blended model results linking surveys, censuses (small area estimation techniques)

Census vs. Survey

Census

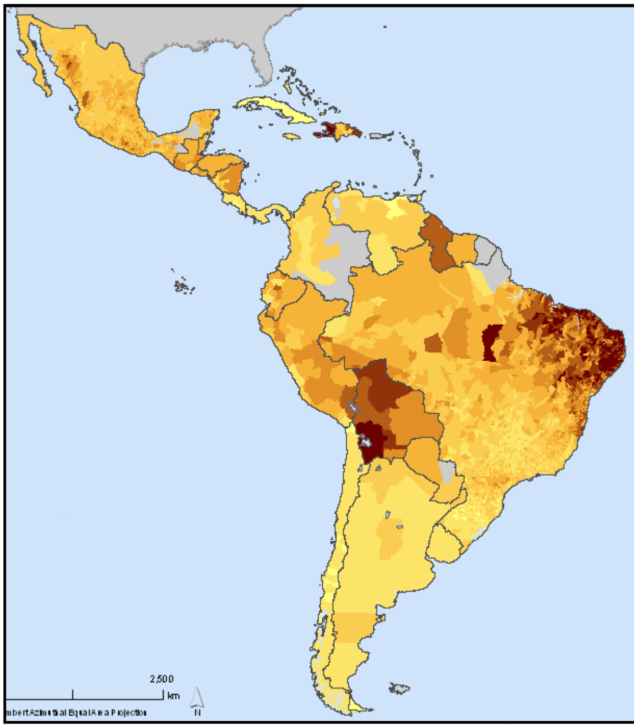
- PROS
  - Results can be representative for small geographic areas
- CONS
  - Costly to implement
  - Decennial
  - Limited number of questions

Survey

- PROS
  - Relatively lower costs
  - Annual / every few years
  - Can cover larger number or more focused set of issues
- CONS
  - Representative only for larger geographic areas

Census Data - Varies widely in quality and detail

- Geographic detail
  - US includes over 1 million geographic units; avg size of unit is less than 1km^2
  - South Africa includes over 83,000 geographic units; avg size of unit is less than 1 km^2
  - Chad includes 14 geographic units; avg size of unit is 298 km^2
  - Indonesia, close to 70,000 geographic units; avg size of unit is about 5 km^2
  - Saudi Arabia includes 13 geographic units; avg size of unit is 386 km^2



- Some countries—e.g., Brazil, Mexico, China—have data at moderate-high resolution IMR and other data (e.g., e_0) based on censuses
  - These countries tend not to have current DHS/MICS surveys.
- New methods should be flexible enough to think about data poor and data-wealthier regions.

Survey Data – Strengths, Weaknesses

- Flexibility – can focus the survey questions specific a particular research interest – HIV/AIDS, beliefs etc.
- Generalizable – with good survey design, can get data from sample and apply information to the larger population.
- Can be implemented more often and cheaper than full census.

- Dependent on response rate
  - may have inherent bias due to self selection
- Comparability – different sets of questions may or may not be comparable between study areas.
  - Translation to different languages may lead to different meanings

Demographic and Health Surveys

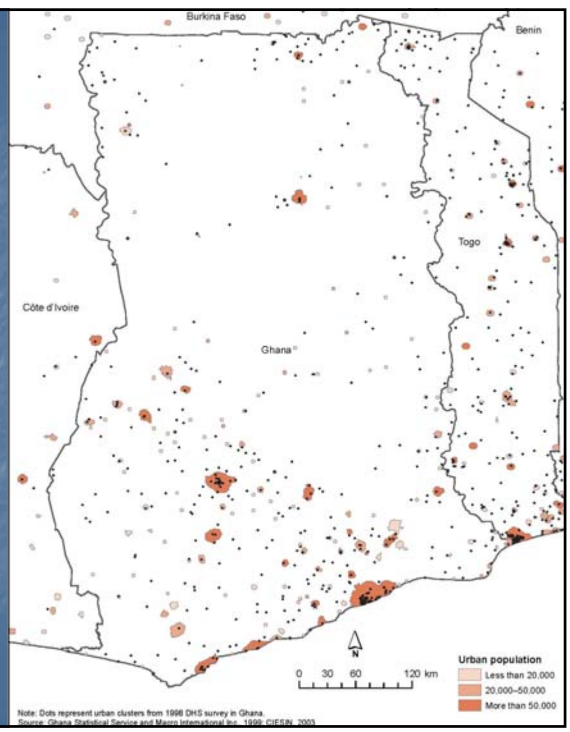
Number of city-regions surveyed	DHS Countries (n=54, since1990)	MICS Countries (n=22, All Africa MICS 2)
0	32	13
1	20	8
2	1 Zimbabwe	1 The Gambia
3	0	
4	1 Egypt	

These DHS surveys have city modules with higher sampling.

Good boundary files are not necessarily available for all urban areas

DHS clusters + GRUMP

- Colored footprints are GRUMP urban extents
- Black dots are DHS clusters
  - Clusters average 30 households
  - Potential for grouping into urban areas, by size of area, has some promise



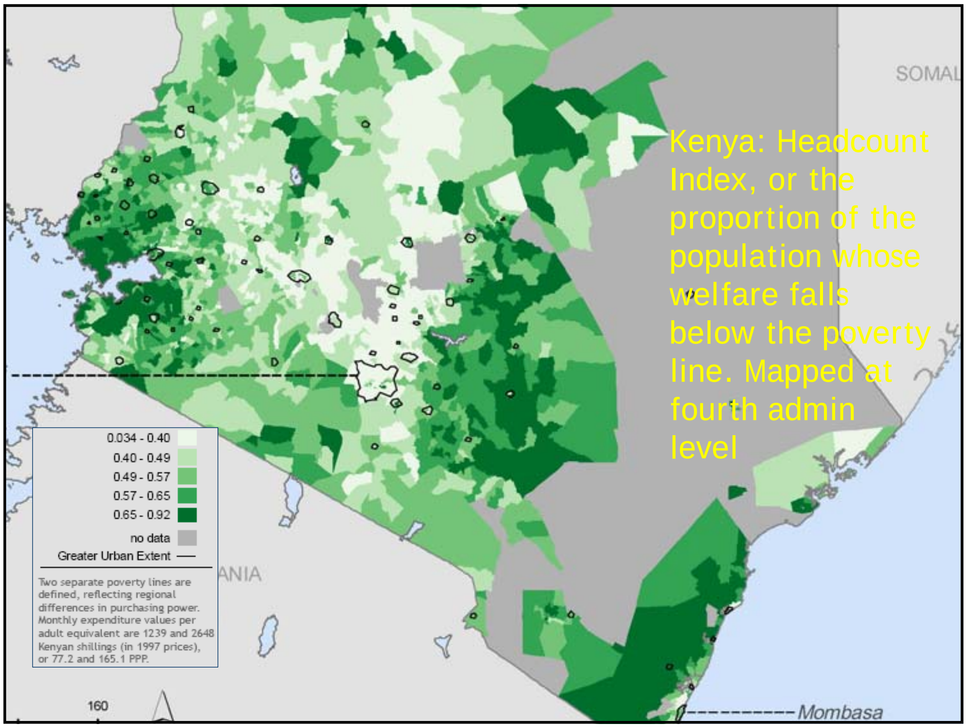
SAE Methodology (1)

- The methodology of small area estimation involves imputing into population census—which does not have consumption data—a measure of per capita consumption from household survey data
- Relies on statistical relationships between common variables in census and survey data

- Detailed description in Elbers et al. (2003), *Econometrica.*

SAE Methodology (2)

- Three Stages:
  - *zero stage*: establish comparability of data sources; identify common variables; understand sampling strategy.
  - *First stage*: estimate model of consumption in the household survey based on common variables.
  - *second stage*: take parameter estimates to census, predict consumption. Use this to estimate aggregate poverty and inequality indicators for small areas.



Statistical Issues

Spatial Statistics

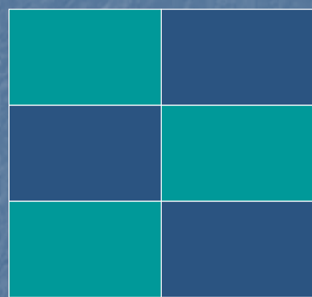
- Spatial characteristics of the dataset are important to take into account when building a model
- Standard OLS regression assumptions include:
  - Normality
  - Independence
  - Homoskedasticity
- Independence among observations is generally violated when building statistical models for spatial data sets

Spatial Autocorrelation (SA)

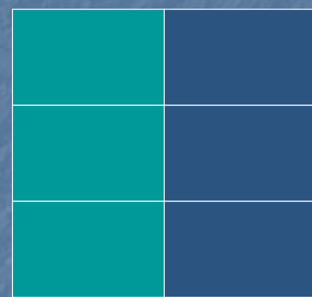
- The extent to which an occurrence of an event constrains or makes more likely an event in a a neighboring unit
- Like serial autocorrelation (in time series data), the events are not independent, and thus violates Gauss-Markov assumptions*
- Estimated coefficients are biased and inconsistent
- Residuals/Standard Errors are artificially deflated leading to type I errors (inproper rejection of null hypothesis)

* According to Lembo (undated): "If the observations… are spatially clustered in some way, the estimates obtained from the correlation coefficient or OLS estimator will be biased and overly precise. They are biased because the areas with higher concentration of events will have a greater impact on the model estimate and they will overestimate precision because, since events tend to be concentrated, there is actually a fewer number of independent observations than are being assumed."

Example of test for spatial randomness



Negative autocorrelation

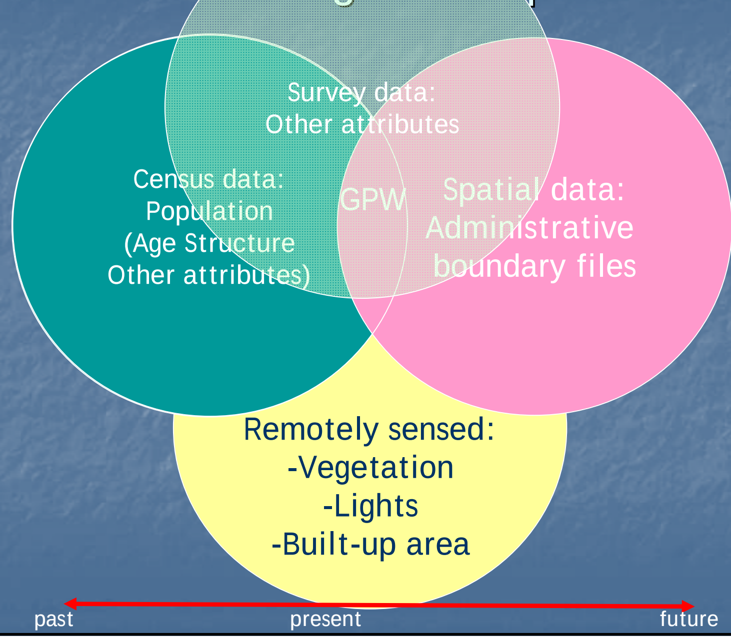


Positive autocorrelation

Moran's *I* is a weighted correlation coefficient used to detect departures from spatial randomness. Looks for spatial autocorrelation

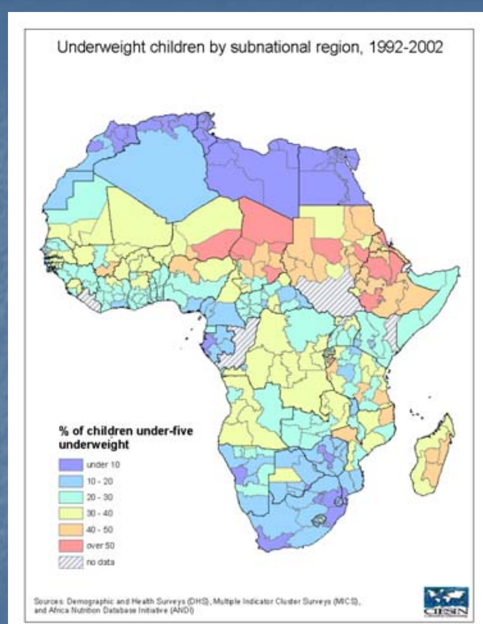
Research Examples

Data integration: inputs



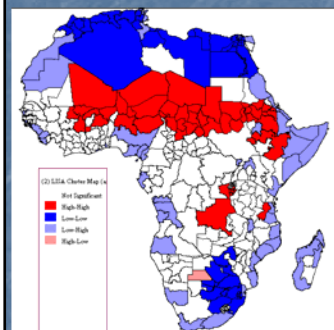
Example of Data Integration Using Spatial Error Model Using Malnutrition Data

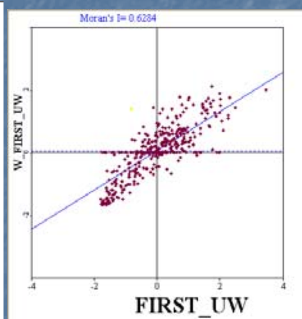
- Obtained percent of children underweight* from DHS and MICS surveys
- Match survey data to boundary data
- 377 sub-national units (SNUs)

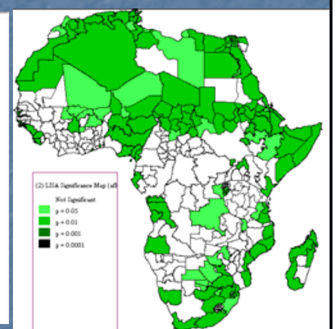


* Children are defined as underweight if their weight-for-age z-scores are below minus two standard deviations (-2 SD) from the median of the NCHS/CDC/WHO International Reference Population.

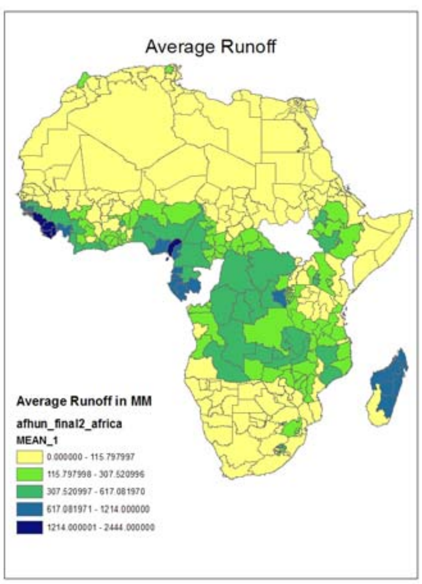
GeoDa for Assessing Spatial Autocorrelcation

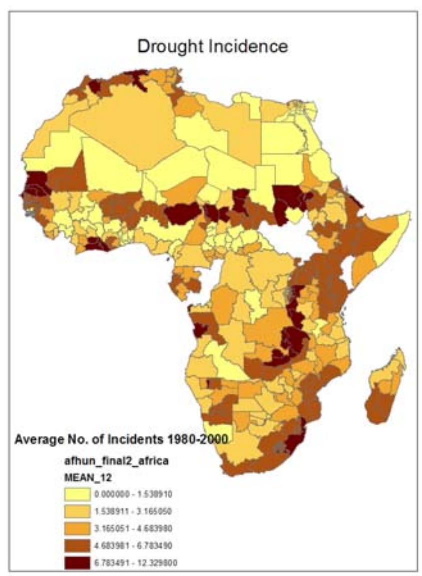


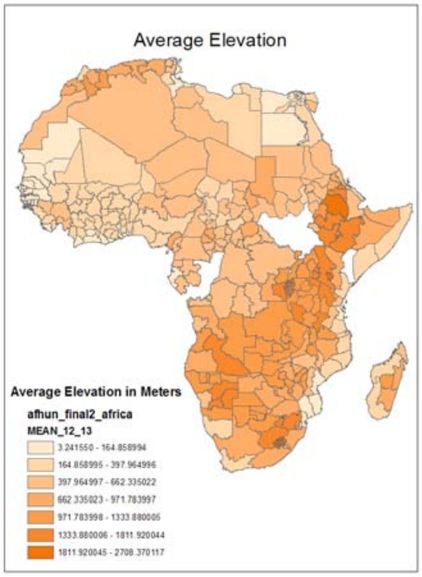


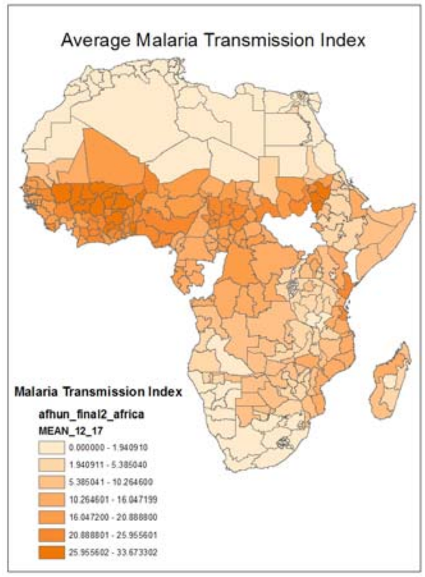


Moran's I is similar to correlation coefficient, varying between –1.0 and + 1.0. When autocorrelation is high, the coefficient is high. A positive I value indicates positive autocorrelation.









Spatial Error Model Results

Dependent Variable: % of Children Underweight	**Unstandardized Betas**
Constant	22.132 ***
GDP per capita	-0.002 ***
Log of Average Runoff	0.348
Log of Average Elevation	1.05 *
Log of Average Malaria Transmission	0.246
Average No. of Drought Incidents (1980-2000)	0.684 ***
Proportion of SNU <2km from road	-13.436 ***
North Africa Dummy	-4.807 *
Ethiopia Dummy	10.943 **
High Agricultural Constraints Dummy	3.22 **
Lambda (autoregressive error term)	1.005 ***

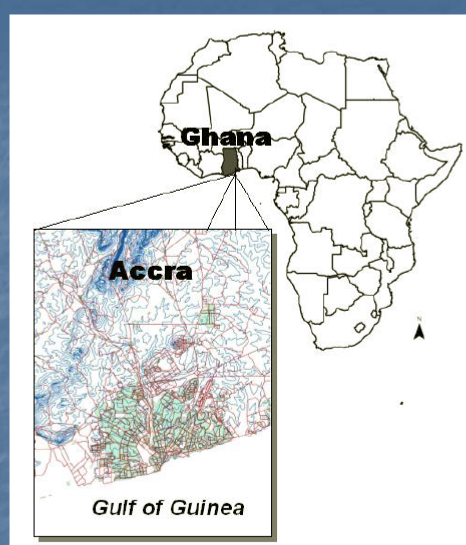
* *p* <.05, ** *p* <.01, *** *p* <.001
Pseudo R^2 = .74
N = 374

Correlates of Hunger: Conclusions

- **What does all this mean?**
  - **Higher elevation areas tend to have higher levels of child malnutrition (even when controlling for the "Ethiopia effect"). This may reflect greater isolation, or constrained agricultural systems due to high slopes**
  - **Overall water availability is less important that the perturbations to agricultural systems from frequent drought (deviations from the mean)**
  - **High road density means greater access to markets, but may also be a proxy for wealth and accessibility to health and other services**
  - **SNUs that face the highest climate, soil and slope constraints to agriculture have significantly higher child malnutrition**
- **Limitations: scale dependence, coarse spatial resolution, error in the measures, lack of other household variables as controls**

Poverty, Health & Land Cover in Accra, Ghana

- Study used:
  - Census data at tract level
  - Women's health survey data
  - Remotely sensed data



Source: Weeks *et al.* 2005. "Intra-urban differentials in poverty and health in Accra, Ghana." Paper presented at the IUSSP International Population Conference, July 2005.

Poverty & Health: Low Correlation

Table 4. Regression of Poverty on Health Indicators

Variable	Standardized Beta Coefficient	t-score
Poverty Score	.361	-2.42

Dependent variable is self-reported health score

R = .361

Adjusted R2 = .108

No outliers; little evidence of heteroscedasticity; Moran's I for residuals = .33; Z(I) = 1.69 at 1500m

Poverty & Vegetation Cover: Higher Correlation

Table 5. Vegetation as a Predictor of Poverty Levels by Locality

Variable	Standardized Beta Coefficient	t-score
Pct Vegetation	-.793	-7.92

Dependent variable is poverty score for locality

Adjusted R2 = .62

No outliers; evidence of heteroscedasticity;
Moran's I for residuals=.43, Z(I)=2.26 at 1500m