Complementing Conventional with Innovative Data Sources

International Workshop on Sustainable Development Goals Indicators
26 to 29 June 2018

Arturo Martinez Jr.
Statistician
Asian Development Bank

Outline of Discussion

- Conventional Data and SDGs
- Complementing with Innovative Data
- Data for Development TA

SDG Monitoring

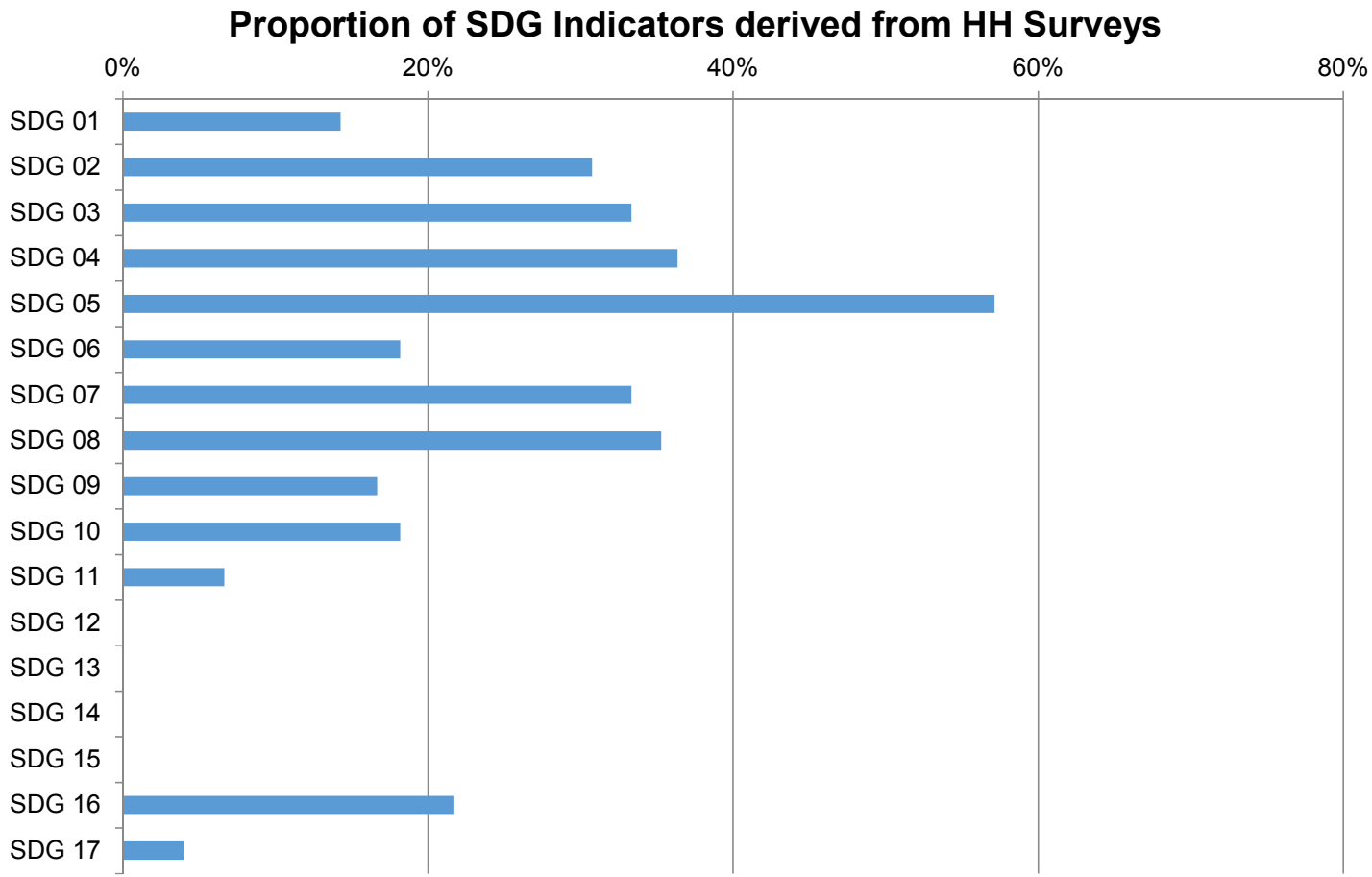
- Many countries in Asia and the Pacific have examined the link between their respective development targets and SDGs → strategic priorities
- Countries have also done data availability assessment → (i) immediately available, (ii) can be made available using conventional data sources, (iii) needs more effort

Conventional Data Sources for SDGs

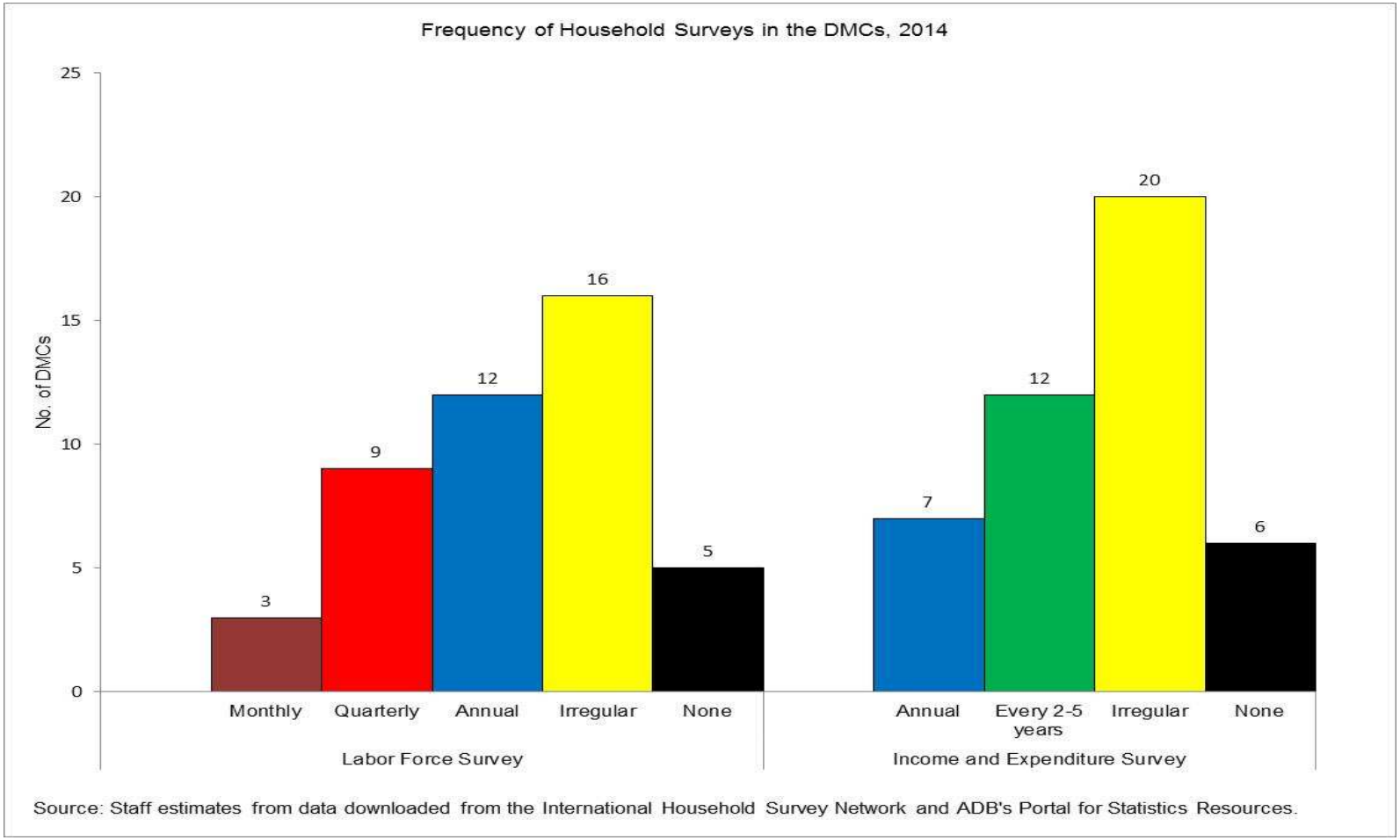
- Administrative Reporting Systems
  - Education
  - Health
  - Trade and Industry (e.g., business permits, letters of credits)
  - Civil Registration
- Censuses
  - Population and Housing
  - Agriculture
  - Enterprise
- Surveys
  - Households
  - Enterprises
  - Farm Holdings

Uses of Conventional Data Sources

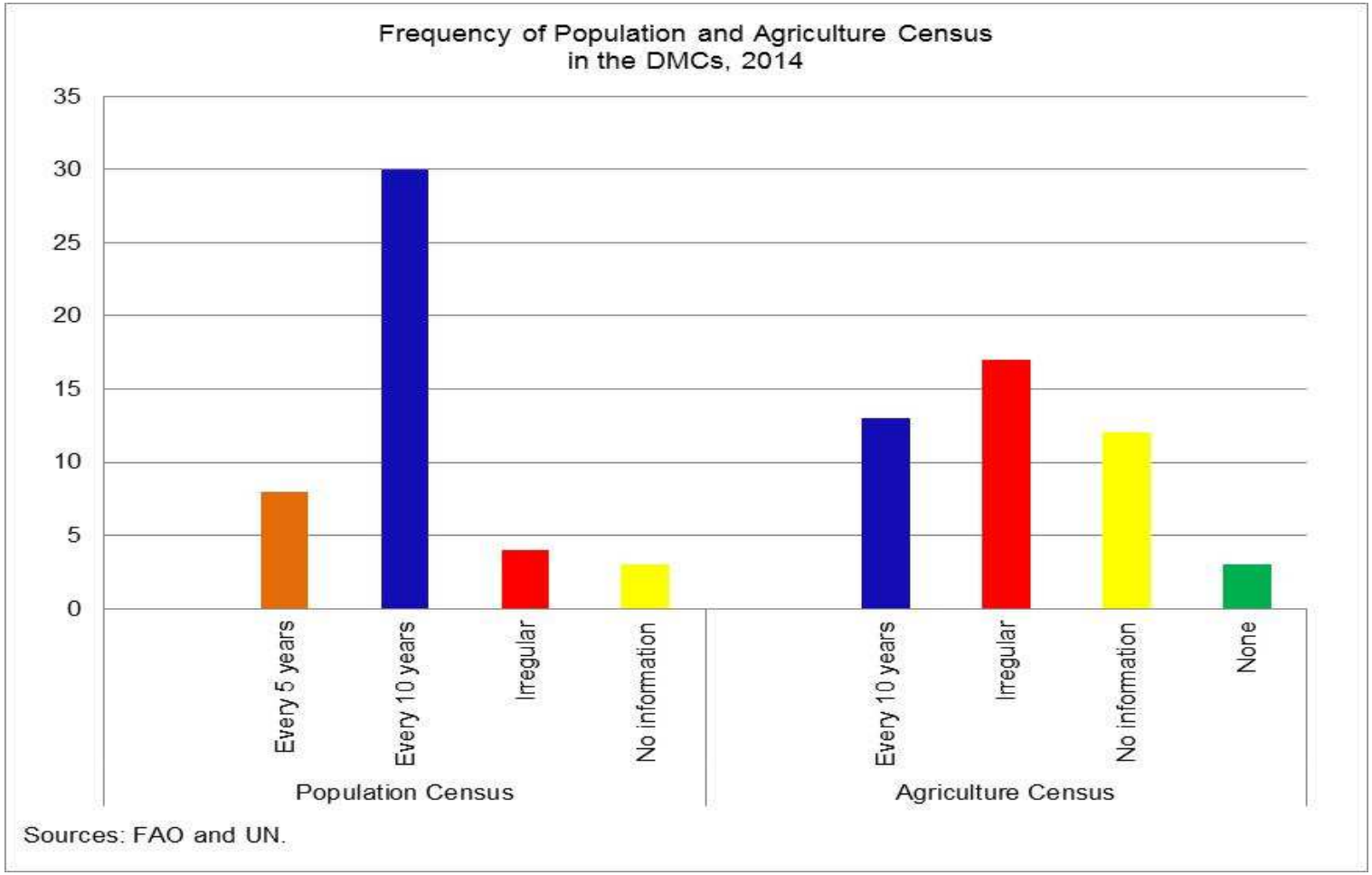
- About **20-50%** of SDG indicators can be derived from household surveys and censuses



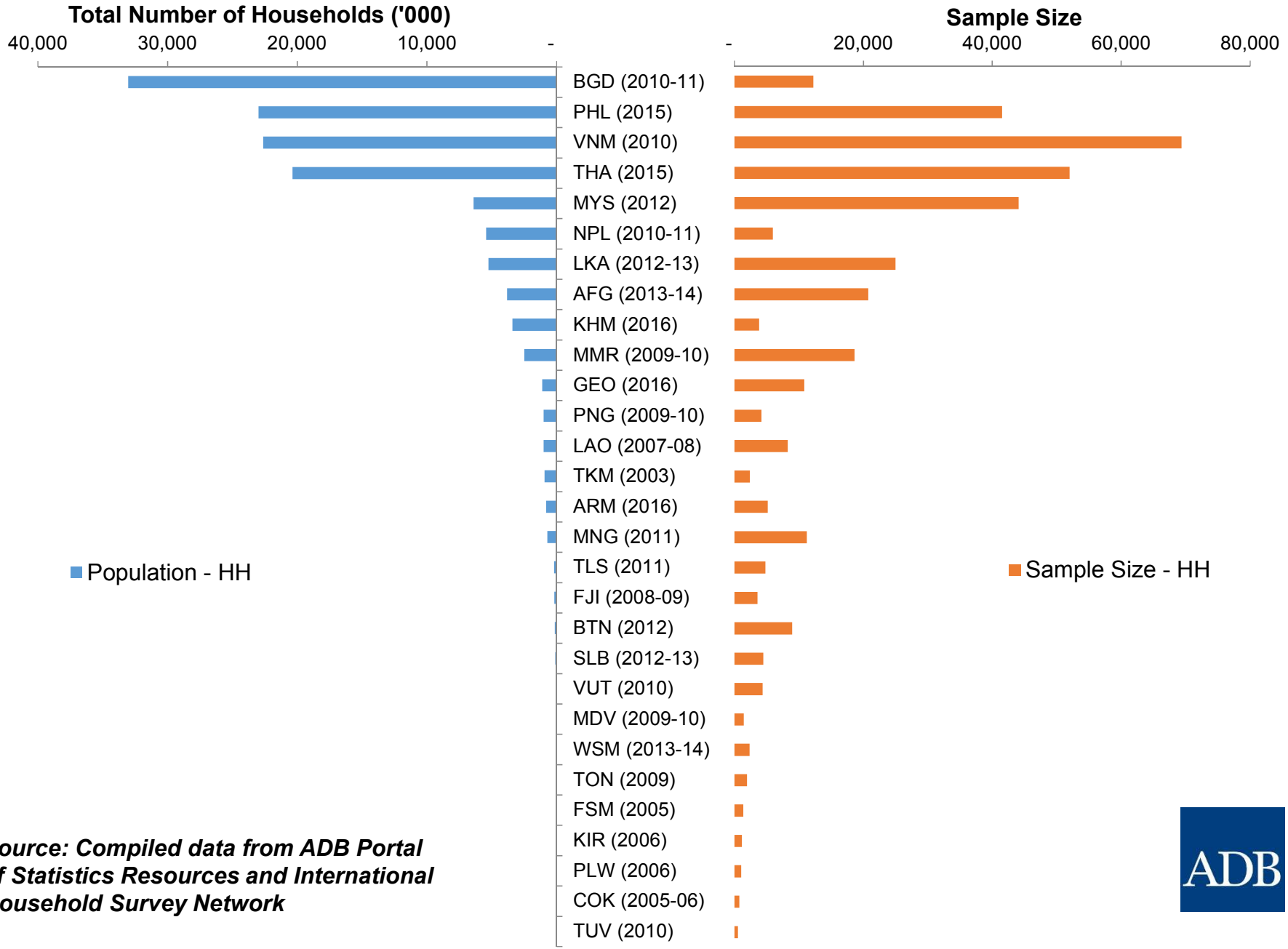
Limitations of Conventional Data Sources



Limitations of Conventional Data Sources

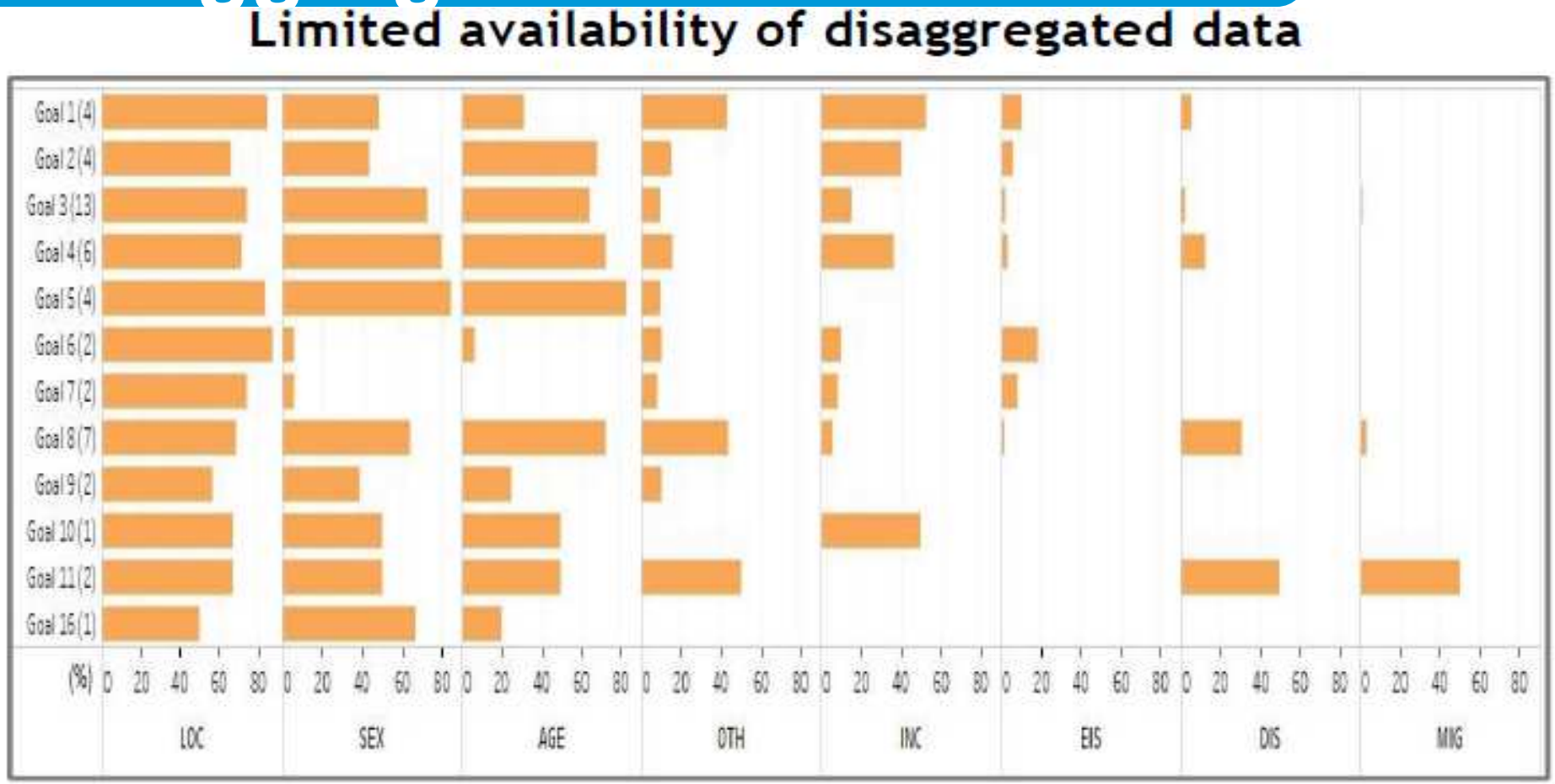


Sample size of latest household surveys vs. total number of households of select ADB DMCs



Source: Compiled data from ADB Portal of Statistics Resources and International Household Survey Network

Availability of Disaggregated Data



- *LOC: Location/ spatial disaggregation*
- *SEX: Sex/gender*
- *AGE:Age*
- *INC: Income Quintiles/ deciles*
- *DIS: Disability*
- *EIS: Ethnicity/indigenous status*
- *MIG: Migratory status*
- *OTH: e.g. education, occupation, religion etc.*

Source: Serrao (2017) – Presentation on SDG Data Compilation

Situation in Asia and the Pacific

- Major surveys and censuses in some countries conducted only if donor funds are available in many countries
  - donor dependence 70-80% budget in some countries
- Poor coverage and quality of administrative reporting systems
  - both economic and social increasing the dependence on surveys
- For disaggregated data, surveys alone may not sufficient
  - administrative data such as from civil registration and administrative registries need strengthening for long term sustainability

ADB

Small Area Estimation

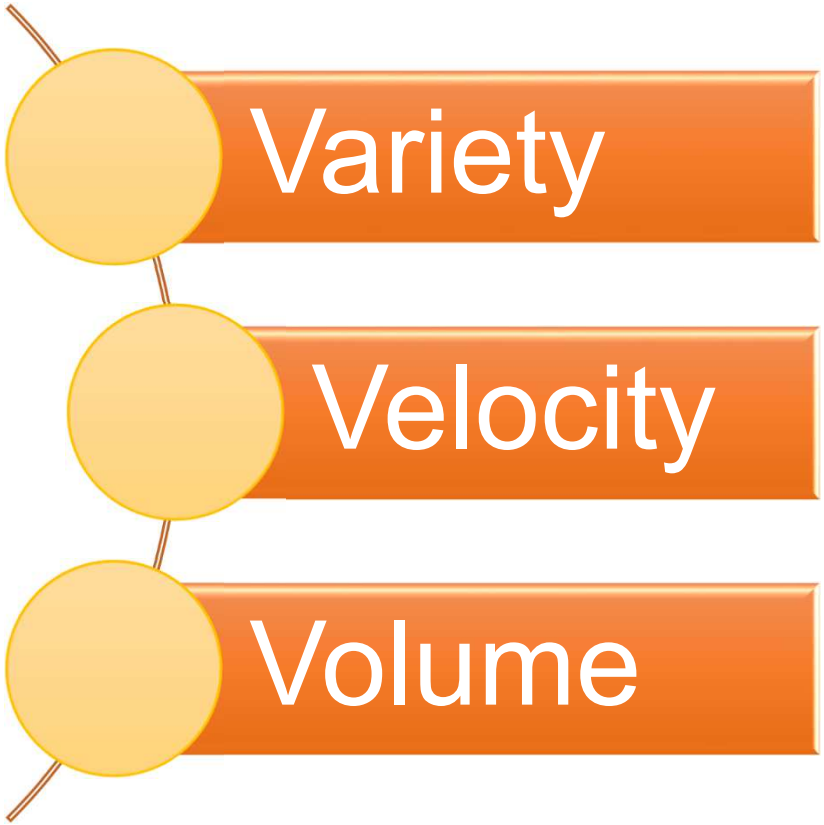
- Let's focus on indicators that are compiled through surveys… but surveys are unable to provide reliable estimates at very granular level!
- A small area is small geographical area or a population segment for which reliable statistics of interest cannot be produced due to survey's limitations
- Small area estimation techniques borrow strength from other 'auxiliary information'

Limitations of SAE methods

- Good auxiliary data usually required
- Gap between census and survey years can increase model error
  - Only time invariant variables can be added in the model if the gap between census and survey periods is too wide

ADB

Big data can help address the limitations of SAE



Big data can be much faster in providing granular auxiliary data than conventional data sources.

How can big data be useful for data disaggregation?

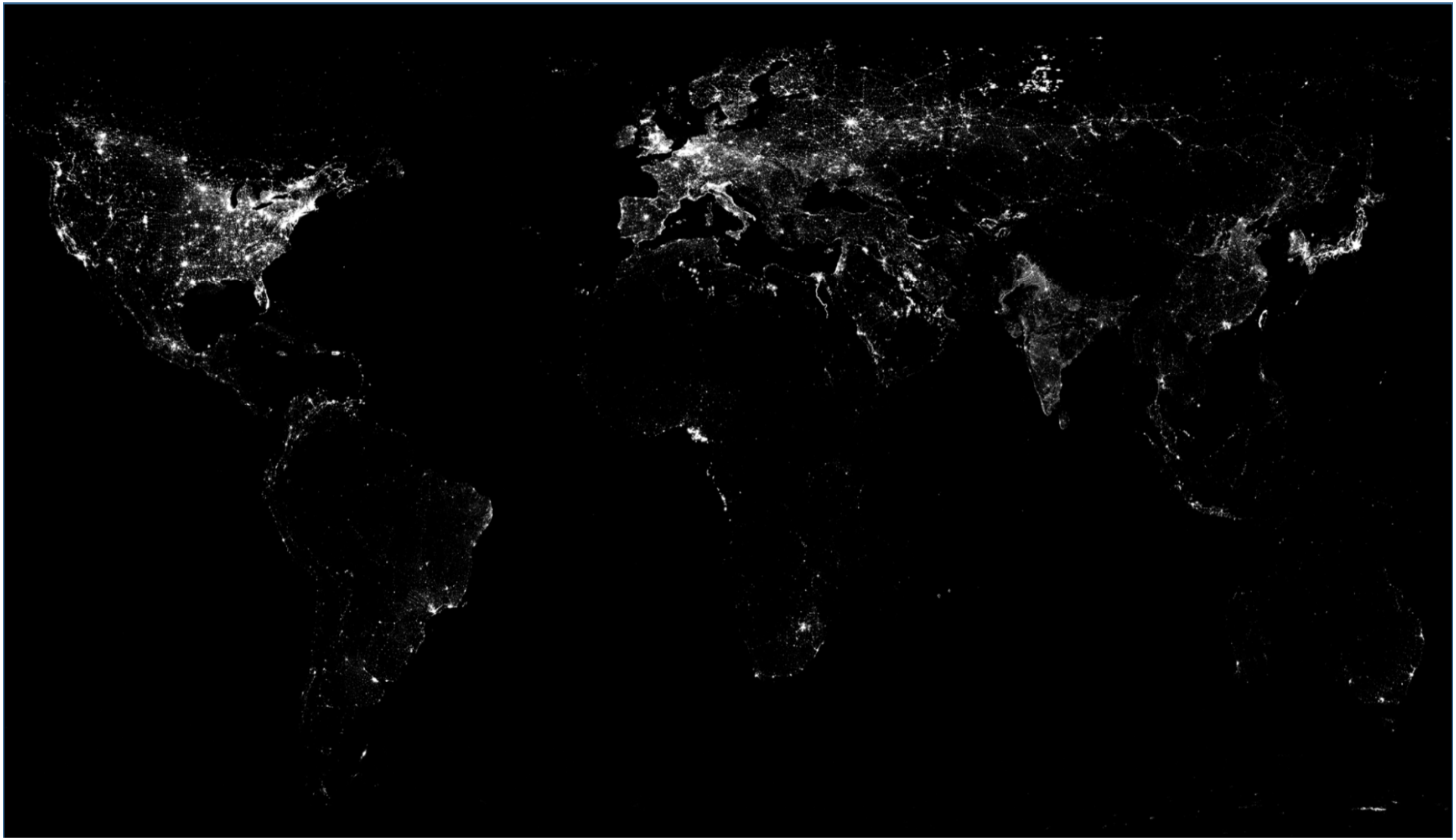
- To create proxy indicators
- To generate new covariates for small area models
- To validate small area estimates

Source: Marchetti et al. (2015) – Small Area Model-Based Estimators Using Big Data Sources

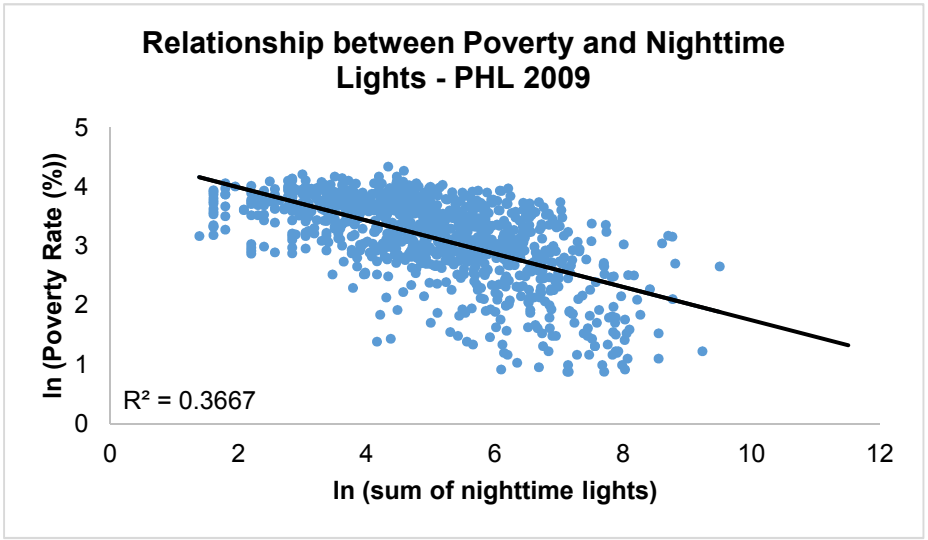
Create Proxy Indicators

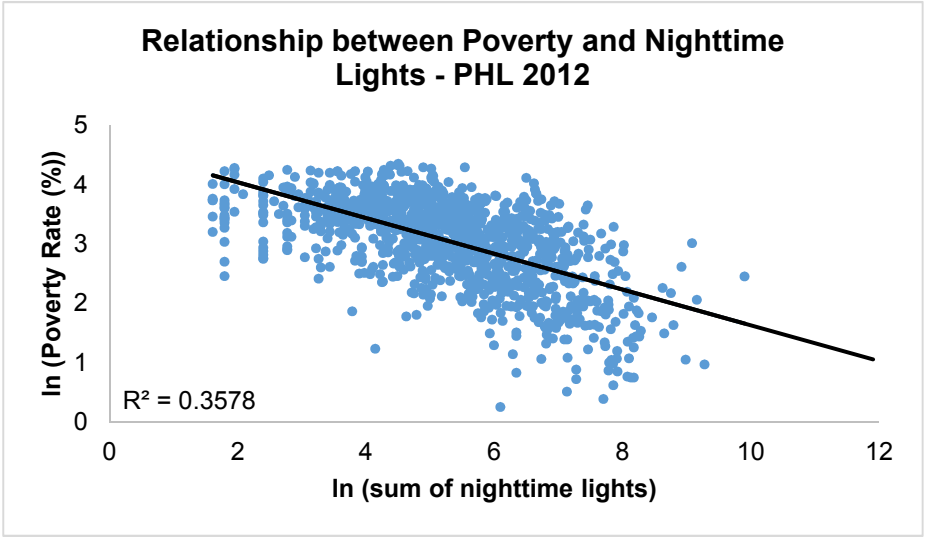
Video source: https://www.youtube.com/watch?v=rXFVejLDGAA

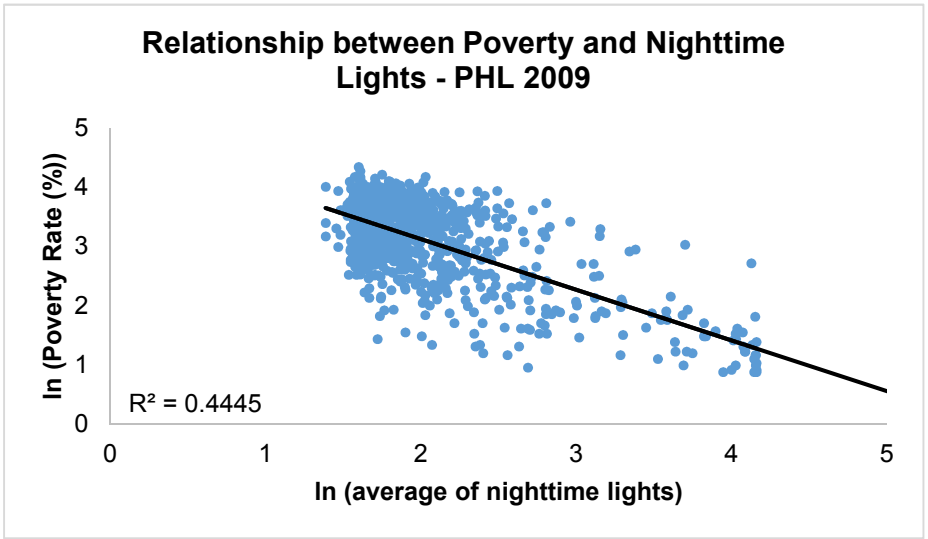
Create Proxy Indicators

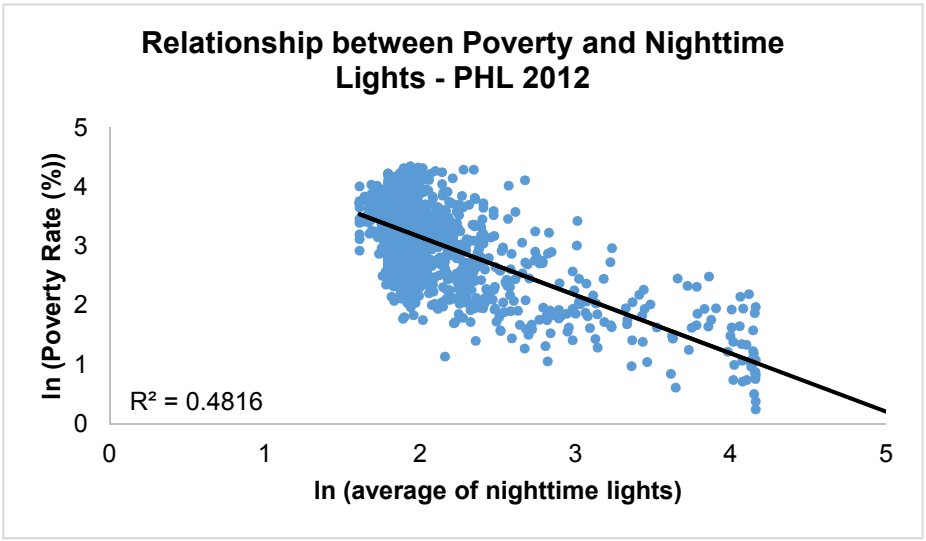


Create Proxy Indicators

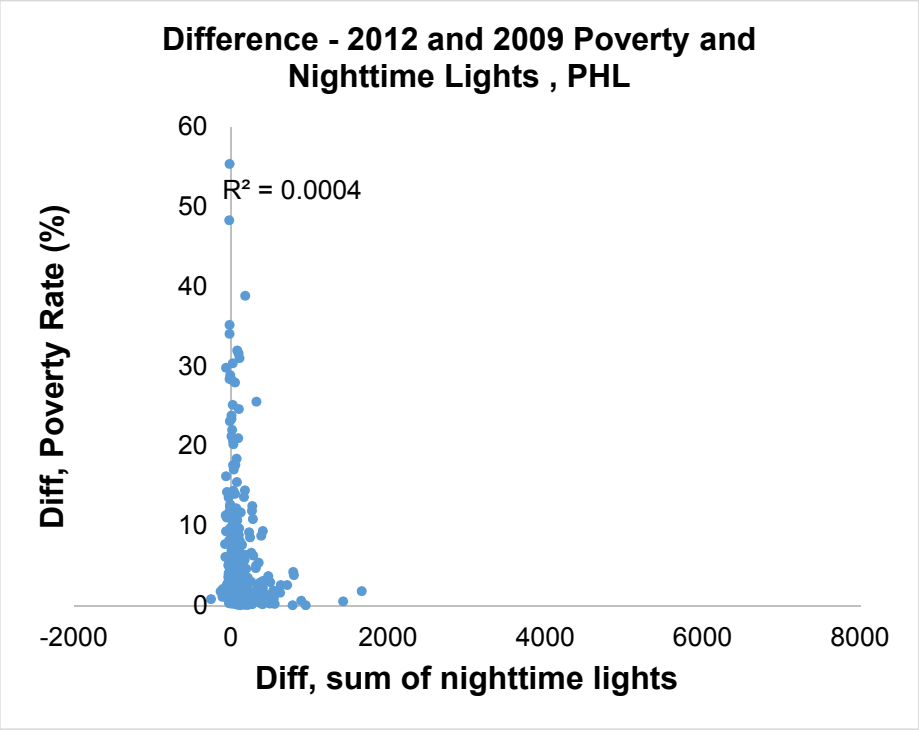


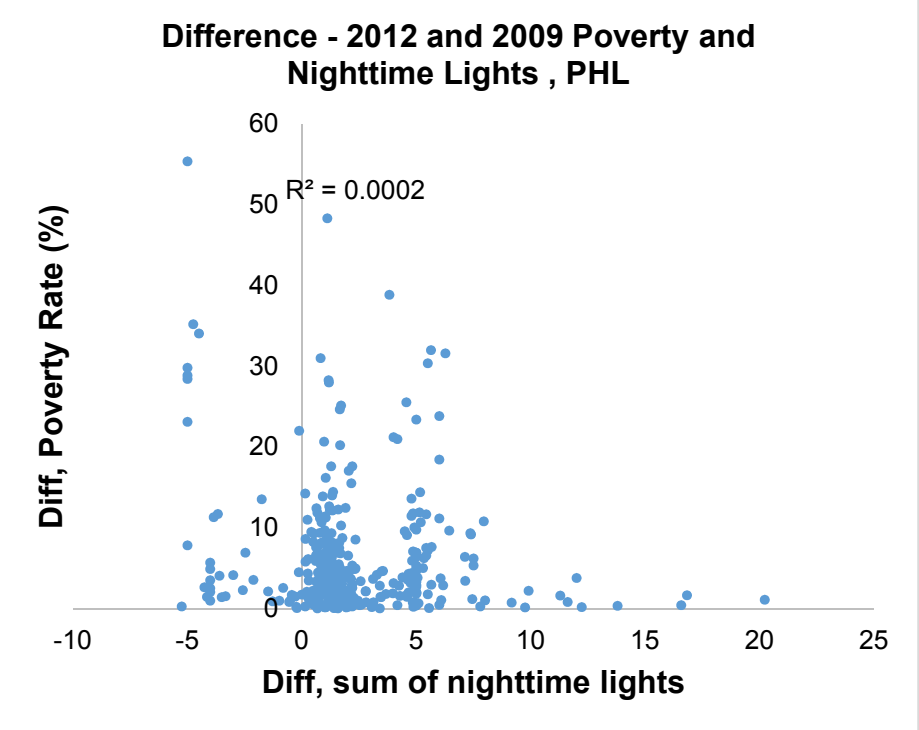






Create Proxy Indicators

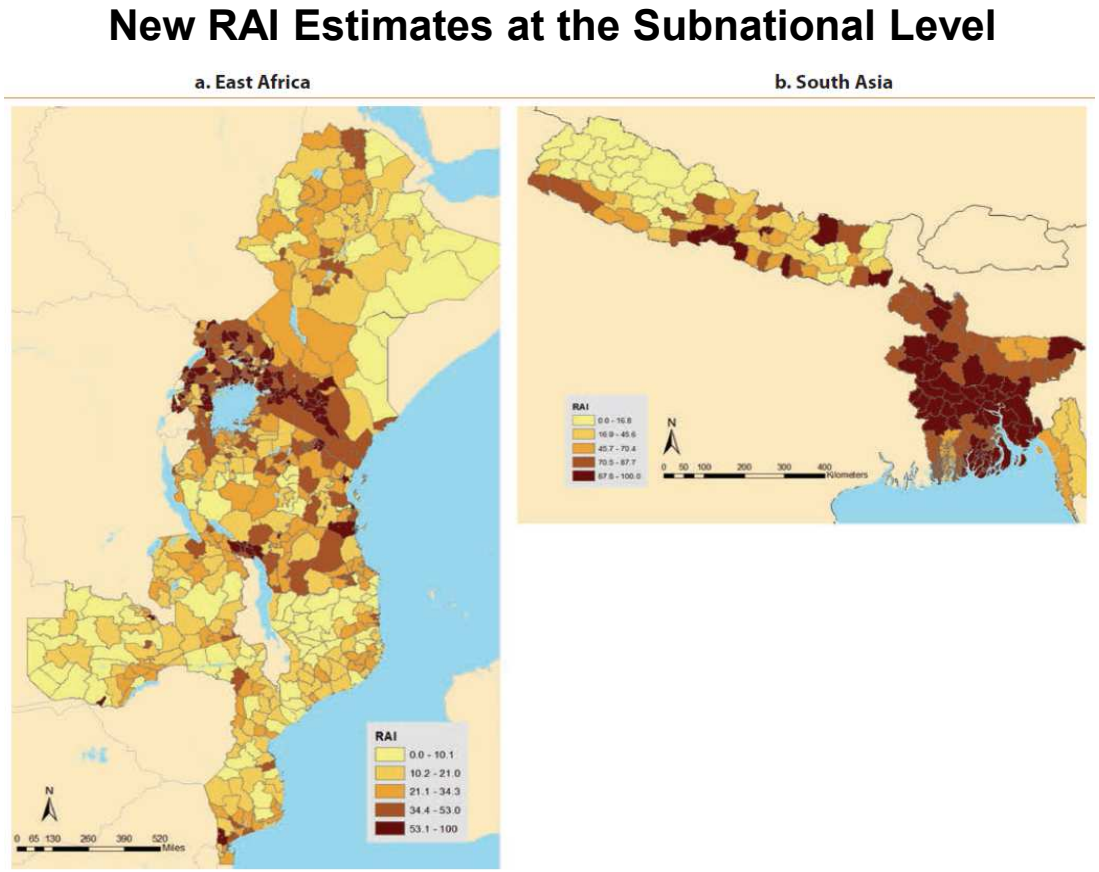




Generate new covariates

SDG 9.1.1. Proportion of the rural population who live within 2 km of an all-season road

- Indicator can be generated using satellite images
- Method to calculate rural access index (RAI) requires the following data:
  - Population distribution (e.g. Landscan and WorldPop gridded population maps)
  - Road network (e.g. OpenStreetMap)
  - Road condition (e.g. satellite image, user reports from apps)



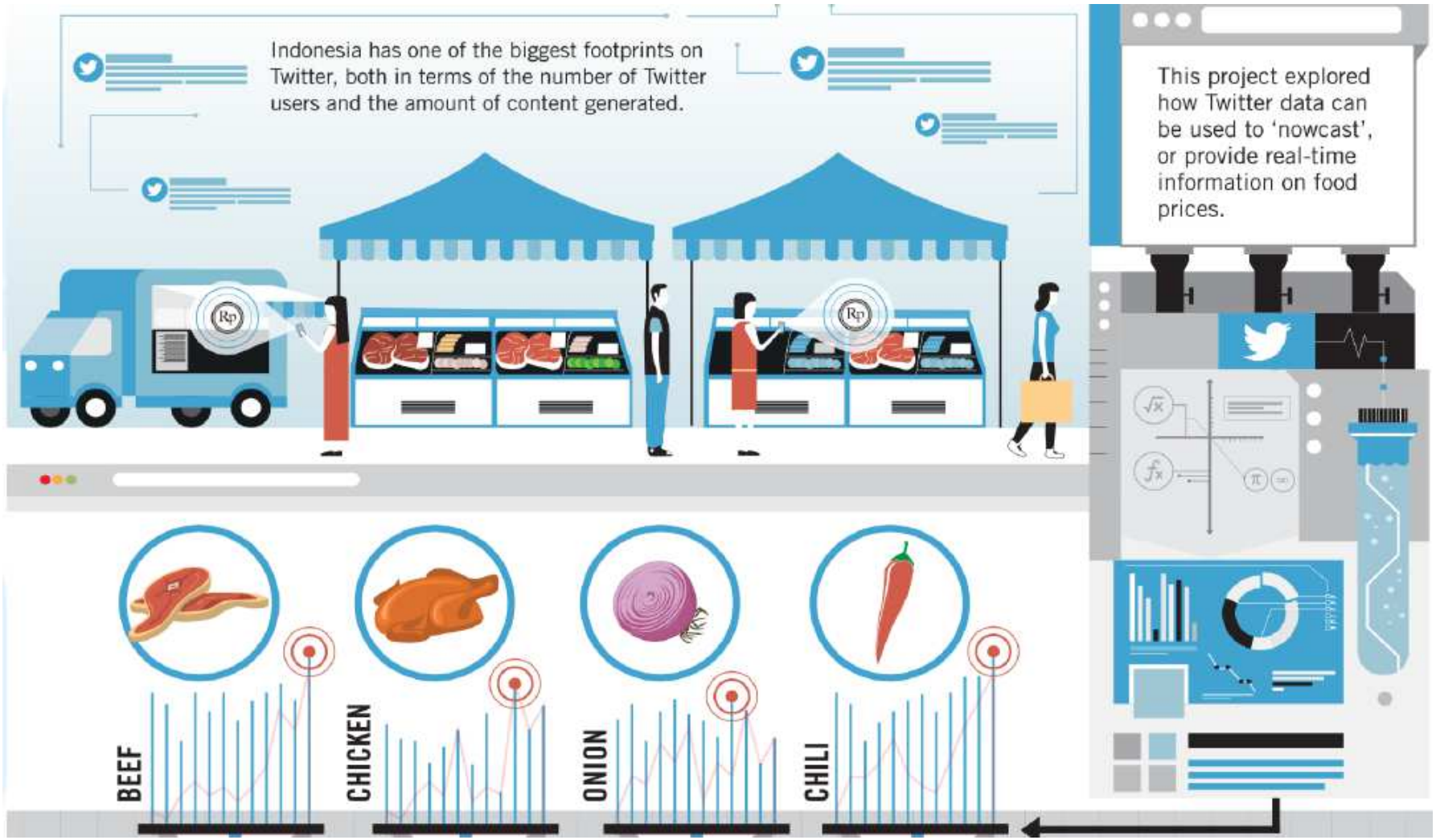
Validate Small Area Estimates

- Socioeconomic indicators derived from big data can be compared to similar measures obtained from survey data
  - E.g. poverty estimates based on satellite image vs small area estimates based on household expenditure surveys

Source: Marchetti et al. (2015) – Small Area Model-Based Estimators Using Big Data Sources

ADB

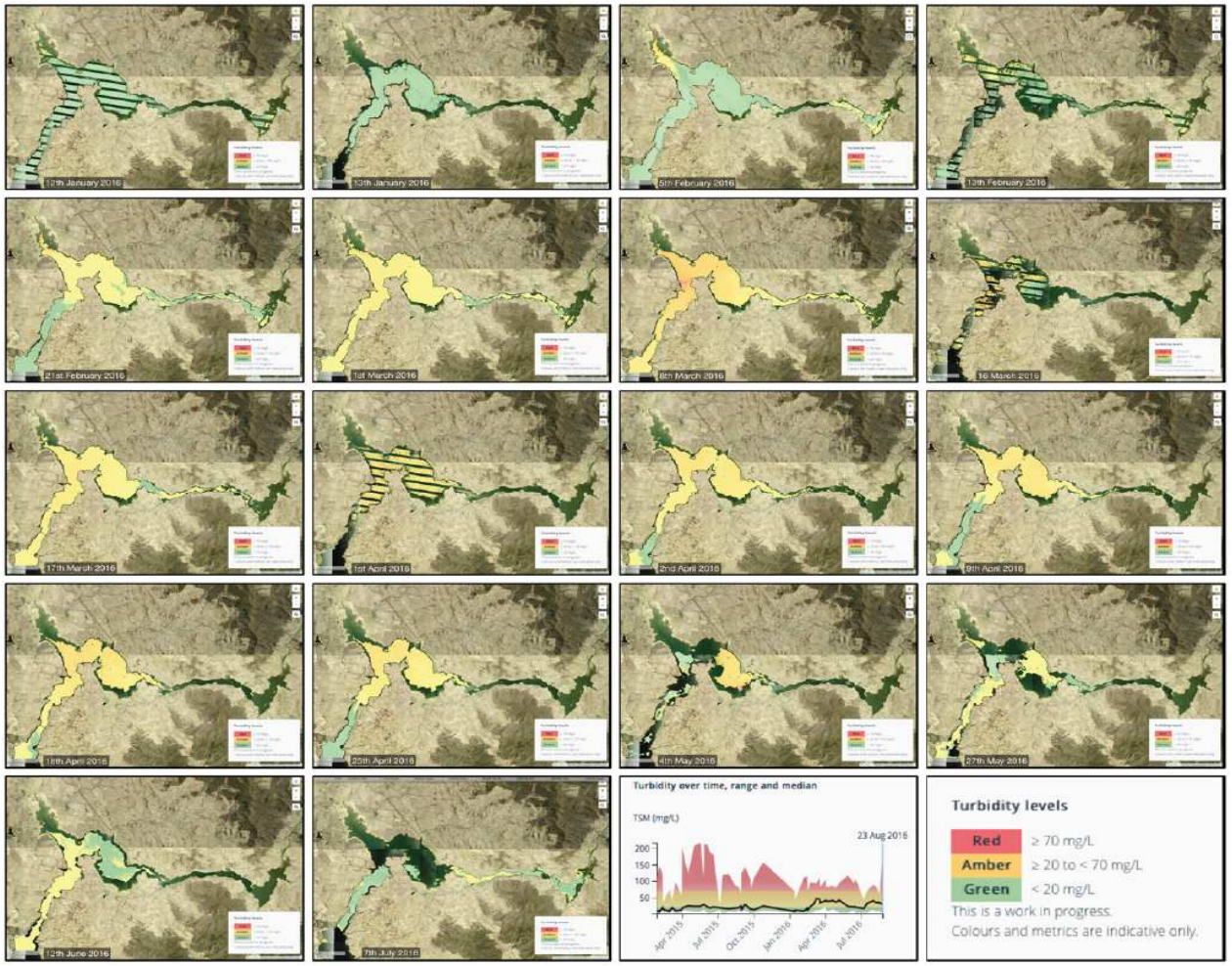
Other uses of Big Data: Now-casting Food Prices in Indonesia Using Social Media Signals



Source: UN Global Pulse

Other Uses of Big Data: Algal Bloom Early Warning Alert System

6 CLEAN WATER AND SANITATION



Source: Group on Earth Observations, 2017

ADB

Key Considerations

- Some types of big data may not be representative of the whole population of interest (self-selection bias)
- Other types of big data are held by private sector
- Needs a different technological infrastructure

Data for Development Technical Assistance

Aims to build the capacity of DMCs in compiling disaggregated data for select indicators of the SDGs using combination of traditional and innovative forms of data in accordance with the SDGs' "leave no one behind" principle's granular data requirements.

ADB is collaborating with UNESCAP, PARIS21 and other development partners

Data for Development Technical Assistance

Country-Specific Case Studies on Data Disaggregation and Big Data Analytics

- Issue: What is the benefit of complementing conventional with innovative data sources?

Data for Development Technical Assistance

- Technical Manual on Disaggregation of Official Statistics and SDGs
- Strategically-designed training workshops targeted to NSO staff
- Online Course Modules on SAE and Big Data Analytics

Thank you very much!

email: amartinezjr@adb.org

ADB's Statistics Capacity Building Efforts and Some Lessons

First statistics capacity building project in 1970s (for Singapore on national accounts)

Approximately 100 technical assistance projects on various topics since then

- Statistics management and strengthening of national statistical systems
- Development of statistics master plan
- Strengthening of selected areas in statistics (national accounts, financial statistics, social statistics, etc.)
- Improving data collection strategies (household surveys, administrative reporting system, dissemination practices)

Established partnerships with other development agencies in the region.

ADB

ADB's Statistics Capacity Building Efforts and Some Lessons

International Comparison Programme for Asia and the Pacific

Updating and Constructing the Supply and Use Tables for Selected Developing Member Economies

Statistical Business Registers (SBR) for Improved Information on Small, Medium-Sized, and Large Enterprises

Evidence and Data for Gender Equality (EDGE)

Innovative data collection methods for agricultural and rural statistics

Implementing Information and Communication Technology Tools to Improve Data Collection and Management of National Surveys in Support of the Sustainable Development Goals

ADB