

# Integrating Demographic and Spatial Data for the SDGs

## Analytic Use of Multiple Data Sources

Mark R. Montgomery

Stony Brook University and Population Council

October 5, 2015

## Outline

- 1 Overview
- 2 Mapping of demographic data: New resources
- 3 Extended example for Peru
- 4 Conclusions



## The “Inclusiveness” Theme in the SDGs

- To monitor whether development is taking place in an equitable and inclusive manner—perhaps the overriding theme of the SDGs—it will be essential to disaggregate national populations along the lines of socioeconomic subgroups and sub-national regions.
- **Sample surveys** provide vital information about socioeconomic groups, but (by themselves) can rarely support analysis and monitoring at the sub-national region level, especially below the first-order administrative level.
- Yet as governmental systems continue to **decentralize**, it becomes ever more important to provide trustworthy evidence for governments and civil society at **multiple sub-national levels**.
- The monitoring system for the SDGs must link **sample surveys with population censuses**, and link both to additional **spatially-referenced data**.

## Extreme-Event Threats Recognized Throughout the SDGs

- **Sudden-onset extreme weather events:** typhoons, heavy precipitation, coastal and interior flooding, landslides—these disasters believed to disproportionately harm **women and children** and there may be important differences in exposure, vulnerability, and resilience by **education** as well.
- **Gradual-onset conditions:** Less is known about **droughts** and in arid regions, **increasing water scarcity**. Implications for rural and urban dwellers (especially the poor) and for rural-urban migration.

## The Hyogo and Sendai Frameworks for Disaster Risk Reduction

The Hyogo Framework and its successor, the Sendai Framework for 2015–2030, aim to:

- Fully integrate disaster planning, early warning, preparedness, and vulnerability reduction into economic development strategies
- Ensure that multiple levels and units of government are well coordinated and linked with NGOs, community-based organizations, and civil society
- Repeatedly emphasizes **integration at the level of the community**, need to understand risk perceptions and constraints on responses at the level of families and neighborhoods
- But **what data** will support an **evidence-based** approach?

## What Role for Demographers?

### How Do We Link Our Socioeconomic Data to Hazard Exposure?

This is the major challenge facing demographers. We have much to contribute, but **we must organize our data** to join the adaptation and risk-reduction conversations.

**Population censuses** Can make a vital contribution—but only if they are **fully disaggregated**.

**The DHS, MICS, and other cross-sectional surveys** Can provide much more detail on risk exposure and behavior in sampled locations.

**Longitudinal surveys** If properly designed (or retrofitted) with spatial detail—can address **resilience** and identify short-term and longer-run consequences.

## Numerators and Denominators

$$\frac{\text{Population Harmed by Hazardous Event}}{\text{Population Exposed to Harm}}$$

We demographers do the denominators!

## Mappable Demographic Data

- 1 Sub-national census and survey data (IPUMS-I, DHS, MICS) provide information on population numbers and probe deeply into socioeconomic composition, perceptions, and behavior. The survey records include names of first-order administrative units (e.g., provinces)
- 2 DHS sampling cluster coordinates and other spatially referenced data: Open up valuable possibilities for data integration

## New, Complementary, Public-Domain Data Resources

- 1 Measures of built-up area (structures) from Martino Pesaresi's Global Human Settlements Layer (GHSL) rasters, from 1975 to the present using Landsat.
- 2 High resolution population density rasters, from Andy Tatem's *WorldPop* team, see <http://www.worldpop.org.uk/>
- 3 Measures of large-city population size and growth rates (UN Population Division) and increasing coverage of smaller cities and towns, via GRUMP and the upcoming GPW from CIESIN.
- 4 Economic activity and city spatial extents: Night-time lights from NOAA, both OLS and the newer VIIRS.

## Mappable Disasters Data

- 1 *EM-DAT*: The OFDA/CRED International Disaster Database.  
National-level disaster reports are available here:  
<http://www.emdat.be>. In recent years, sub-national units where disasters occurred are listed, but effects not disaggregated.
- 2 Much of the literature on disaster exposure and consequences has employed **whole countries as the units of analysis**. Poor, less poor, and higher-income countries; High, moderate, and low exposure to risks.
- 3 Helpful, but hardly enough.



## The DESINVENTAR Program

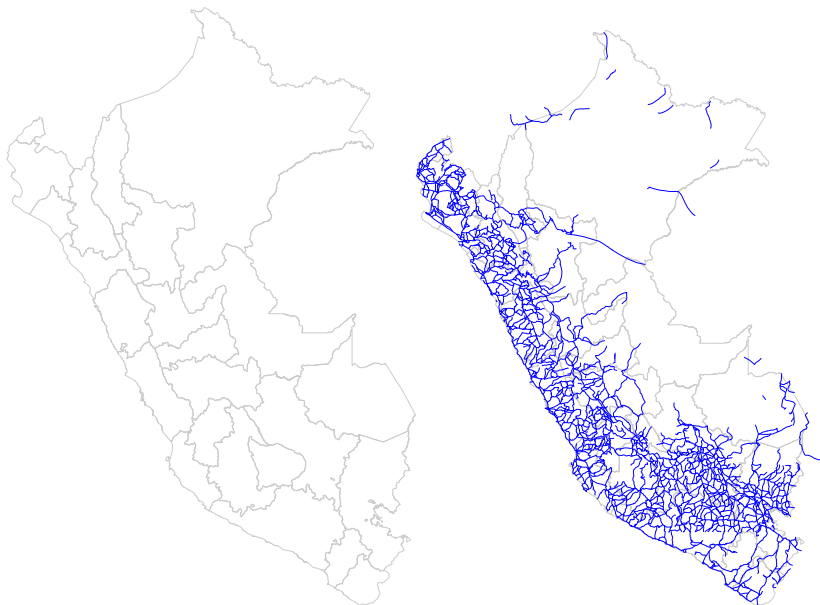
Since the late 1980s in Latin America and a smattering of other countries in Africa and Asia, a coherent system of **comprehensive, small-scale disaster reporting** has been in place, with protocols, software, and public-domain data: <http://www.desinventar.org>.

Curiously under-exploited by researchers to date, apart from rather thin summaries in Global Assessment Reports on Disaster Risk Reduction (2013).

## Example: Peru 2009 DHS

- DHS surveys generally record first-level administrative region (sometimes second-level) and are designed to be “statistically representative” at that level.
- But such large geographic units say little about (for example) household access to services via roads (which are also important for humanitarian relief logistics).
- And first-order units are simply too large for meaningful extreme-event risk profiles

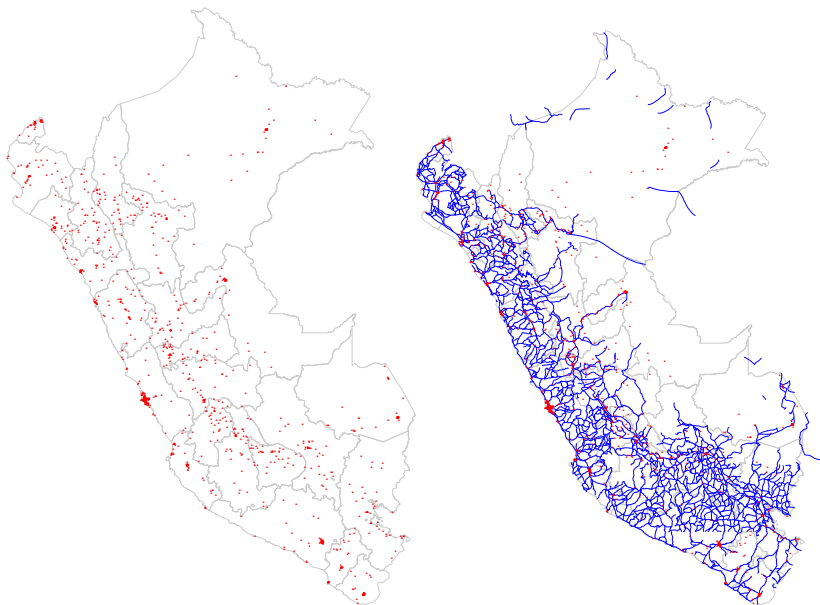
## Peru DHS and Global Roads Data (SEDAC)



## Most of the Recent DHS Collect Cluster Coordinates

- The DHS program now discourages questions about finer levels of geographic location in the questionnaire
- These questions would be redundant if all surveys collected cluster coordinates
- Not all surveys do (yet) and some countries do not permit coordinates to be distributed
- The DHS program randomly displaces coordinates (2 km. in urban, 5 km. in rural) to protect confidentiality. The true coordinates are never released.

## Cluster Coordinates Add Substantial Value

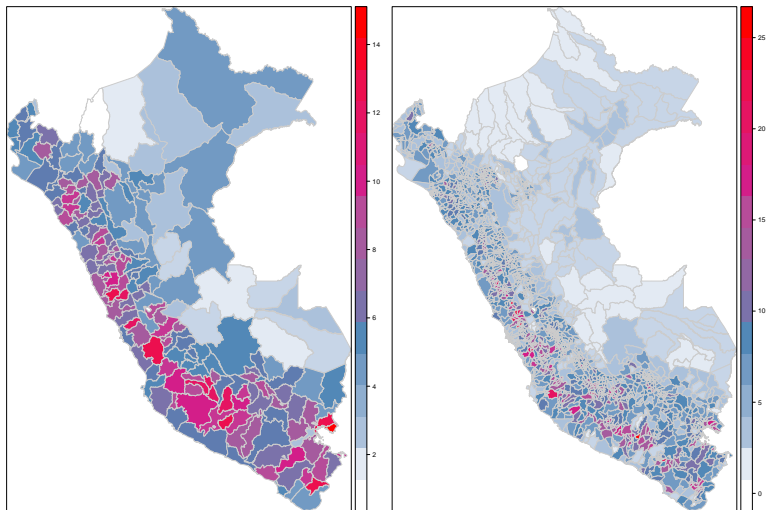


## Coordinates Enable Links to Finer Administrative Levels



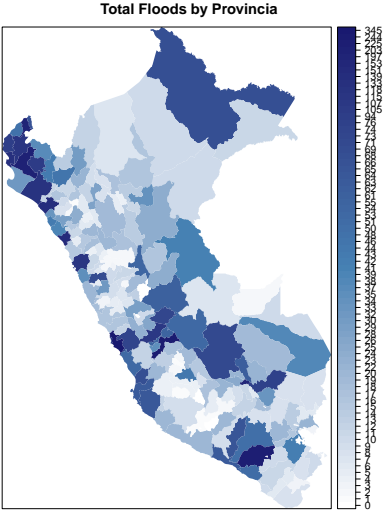
## These Links Allow Integration of Fine-Scale Census Data

Percent of Provincia and Distrito Population Aged 65+, 2007 Census Microdata



# Census Data Indicate Population Exposure to Extreme-Event Risks

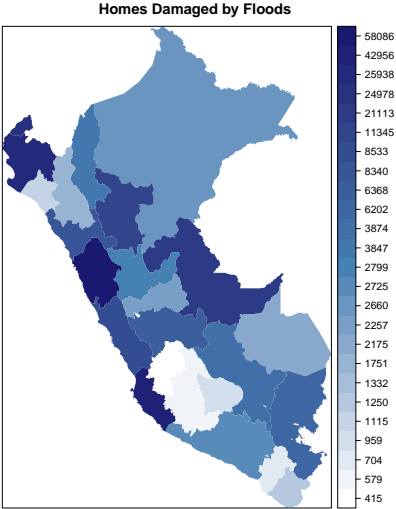
Flooding in Peru, from DESINVENTAR





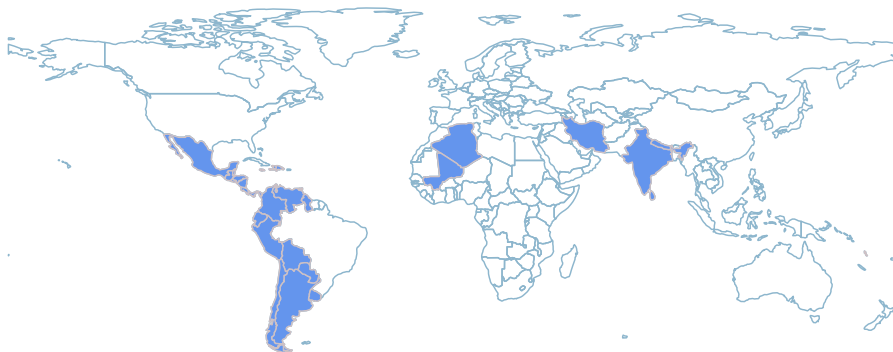
# Allowing Loss Data on Homes and Lives to be Interpreted

Homes Damaged by Floods in Peru, from DESINVENTAR



## The Reach of the DESINVENTAR Program

Other countries (not shown) have tested parts of the protocol: see GAR (2013)



## The DESINVENTAR Approach

- Event occurs (e.g., storm)
  - No human harm  $\Rightarrow$  No record of event (**Design bias**)
  - Human harm occurs:
    - But not reported in the media  $\Rightarrow$  No record of harm (**Reporting biases**)
    - Reported in the media  $\Rightarrow$  Disaster record(s) created via a well-conceived process of:

Verification by two or more coders

Spatial pinpointing with a record created for each spatial (administrative) unit affected

**Coding at most disaggregate level possible** given the verified data

Notes on street addresses often included for urban events

Multiple types of harm detailed for each unit

## Adapt DESINVENTAR for SDG Monitoring?

- Its **finer spatial scale is essential**. DESINVENTAR methods and protocols warrant close attention.
- But the DESINVENTAR reporting system is media-based. What is the reach of the media?
- What events and harms are deemed “newsworthy”? Droughts vs. floods comparison.
- Does not address the fundamental problem affecting disaster data:  
**Who is harmed?** Sex, age, education, poverty, . . . .

## What More Should Demographers Be Doing?

Encourage collection of coordinates in all data-gathering Even the DHS program falls short of gathering coordinates in all its surveys.

Allow use of true (non-displaced) cluster coordinates, in an enclave Not all research questions require precise location data—but a number of extreme-event analyses do! Low-elevation coastal zone, for example. And in urban areas, 2 kilometers of displacement prevents any meaningful linkage to neighborhood and within-city census and services data. That is not acceptable in a majority-urban world. Use of enclaves is standard survey practice where confidential data are concerned.

Establish enclaves for analysis of census micro-data IPUMS-I shows how much can be done with census micro-data. As Tom LeGrand mentioned, for some issues and disadvantaged population groups, there is no substitute for the full individual records, analyzed in a secure setting.