UNITED NATIONS

ECLAC

# Poverty mapping and disaggregated estimates using Small Area Estimation in ECLAC

## Statistics Division

## 2022

**Rolando Ocampo Gutiérrez**
**Director - Statistics Division - ECLAC**
rolando.ocampo@un.org

# Disaggregation and the measurement of the Sustainable Development Goals


SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS

# Some targets related to SDG 1 (No Poverty)

- **1.1** By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day.

- **1.4** By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance

# Fundamental principle of data disaggregation

*Sustainable Development Goal indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics*

***Global indicator framework for the Sustainable Development Goals A/RES/71/313***

# Household surveys limitations and the use of auxiliary information

# ¿What is it all about?

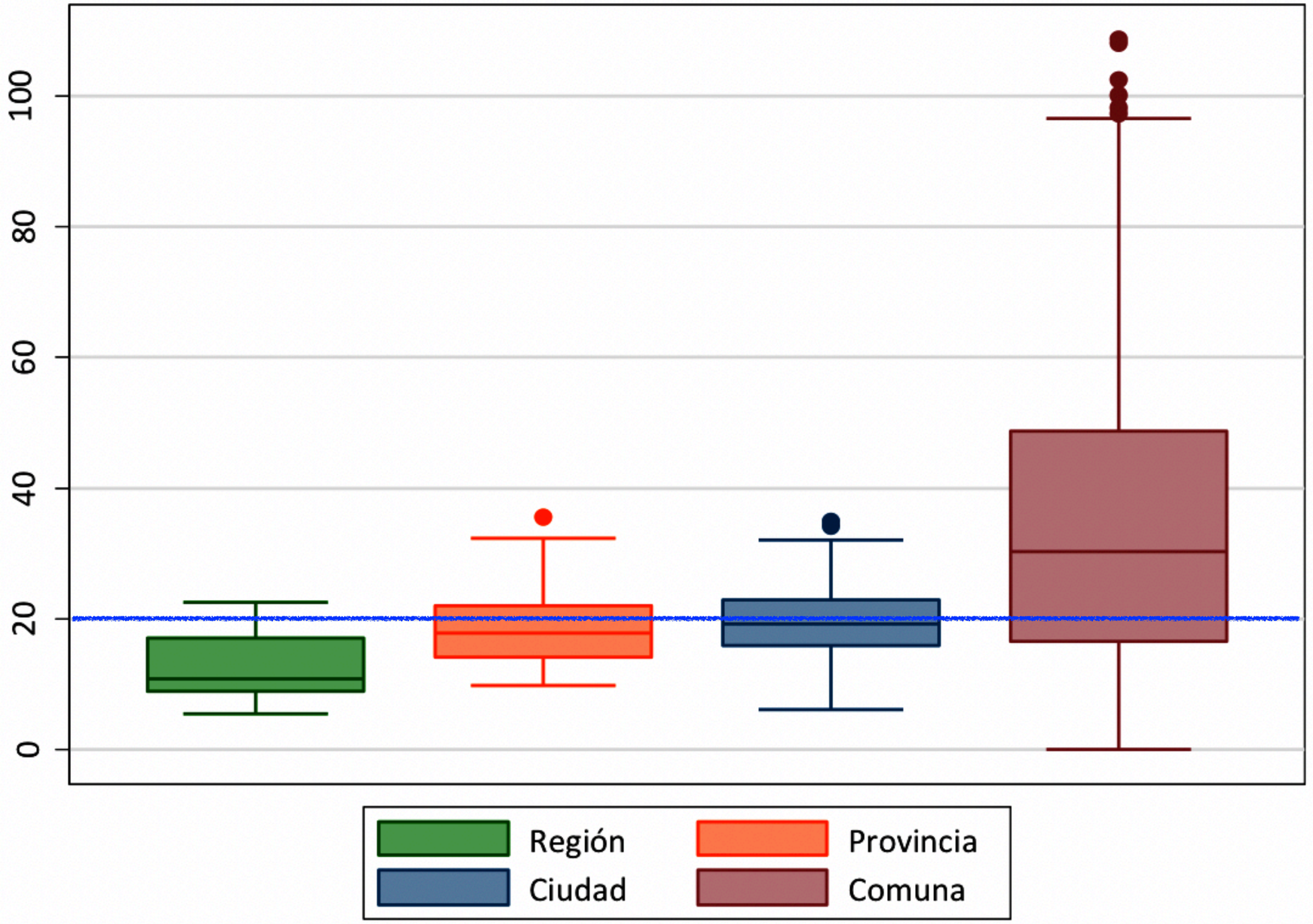
Surveys that depend on large sample size and a proper sampling strategy (sampling design and estimator) also rely on a robust inferential system that provides precise and exact estimation in planned domains.

When the sample size of the survey is not enough to support the statistical inference required for some subgroups of interest, it is necessary to resort to external auxiliary information (censuses, administrative records, satellite images) so that together (surveys and external data) a precise and exact inferential system can be built.

# ¿What is a small area?

An area (or domain) is small if the sample size is insufficient to support the direct inference process (based on the representativity principle of survey sampling design) with adequate precision.

The term **small** does not refer to the subgroup's absolute size; i.e.,  states or provinces can be considered small areas if the sample size is insufficient.




# Coefficient of variation at different disaggregation levels

Source: NSI - Chile

# Disaggregation levels

The parameters of interest may be required in geographic breakdowns (which can be observed on a map) or in crossings of sociodemographic and income subgroups.

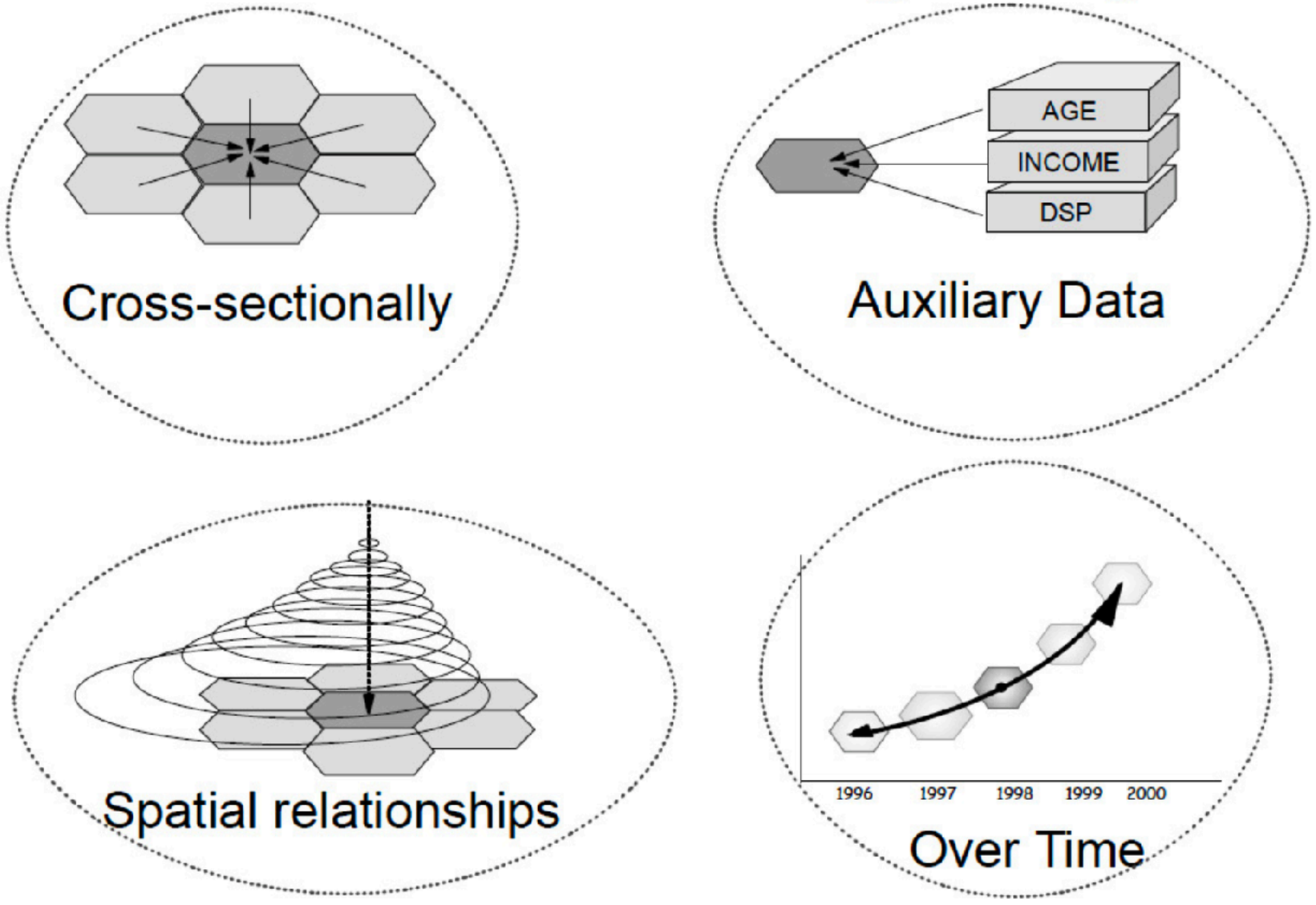
- Geographic: regions, states, municipalities, and scholar districts.
- Specific subgroups: age × sex × ethnicity × immigration status.

In general, if the subgroups are not part of the survey design domains, their sample size is not planned and will be random, which increases the estimate's uncertainty.

# Optimal solution

When the sample size does not allow obtaining reliable direct estimates for some domains of interest, the following options can be addressed:

1. Increase the sample size: this option raises costs, and it is unfeasible.
2. Use statistical methodologies that involve external auxiliary information to obtain reliable estimates (not direct) in the subgroups of interest while keeping the survey sample size.




# *Borrowing strength*: different ways to include auxiliary information

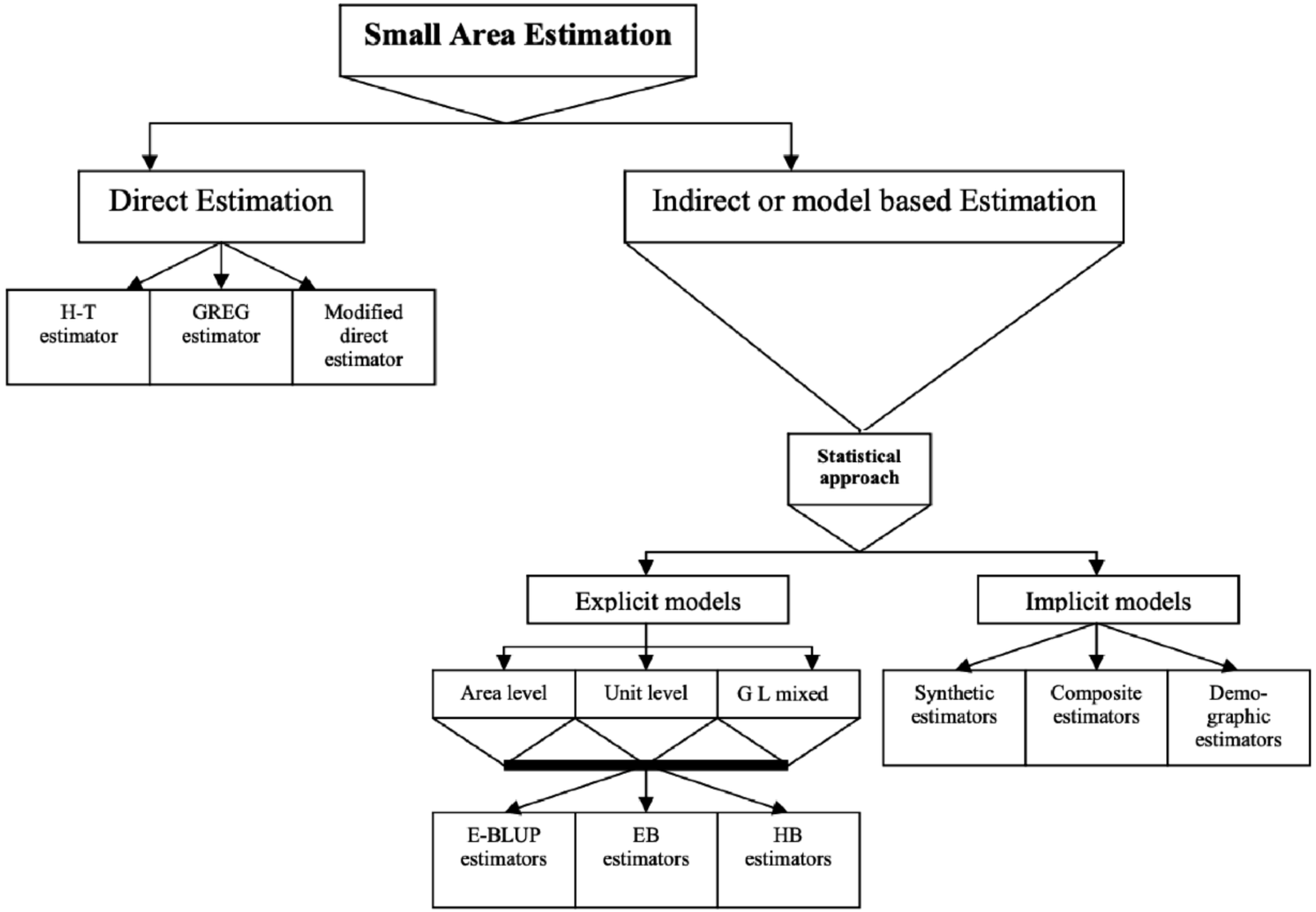
Source: Methodology of Modern Business Statistics (2014)

# SAE methodologies

SAE estimators are divided into three main types:

- Indirect synthetic estimators
- Estimators based on area models
- Estimators based on unit models

The choice of the method that should be used in the estimation of the domains of interest is made depending on the level at which the auxiliary information is found (at the domain or aggregation level - at the household or person level)




# Different possibilities when fitting SAE models

Source: adaptation from Rahman (2008)

# Area-level model (Fay-Herriot)

Fay & Herriot (1979) analyzed per capita income for small areas with less than 1000 inhabitants.

Link model:

$$\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + u_d \qquad u_d \sim_{ind} N(0,\, \sigma_u^2)$$

Sampling model:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d = \mathbf{x}_d'\boldsymbol{\beta} + u_d + e_d \qquad e_d \sim_{ind} N(0,\, \psi_d)$$

# Unit-level model (*Empirical Best/Bayes Predictor*)

Molina & Rao (2010) estimated poverty rates and gaps at the crossroads of gender and province in Spain using the following model.

Link model:

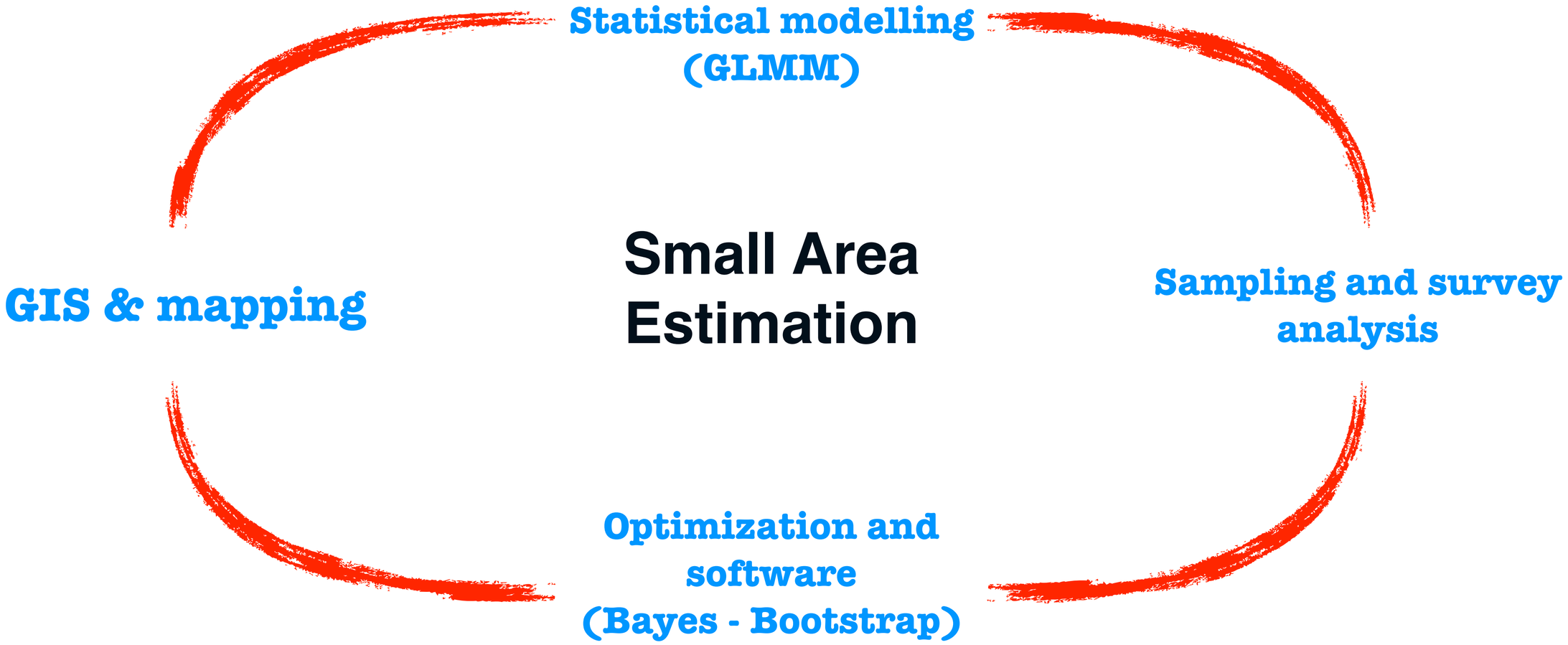
$$\tilde{\delta}_d^B(\boldsymbol{\theta}) = E_{\mathbf{y}_{dr}}[\delta_d(\mathbf{y}_d) \,|\, \mathbf{y}_{ds}; \boldsymbol{\theta}]$$

$$\mathbf{y}_{dr} \,|\, \mathbf{y}_{ds} \sim_{ind} N(\boldsymbol{\mu}_{dr|s}, \mathbf{V}_{dr|s})$$

Conditional estimation:

$$\boldsymbol{\mu}_{dr|s} = \mathbf{X}_{dr}\boldsymbol{\beta} + \gamma_d(\bar{y}_{da} - \bar{\mathbf{x}}_{da}^T\boldsymbol{\beta})\mathbf{1}_{N_d - n_d}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1-\gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}_{N_d-n_d}^T + \sigma_e^2 \text{diag}_{i \in r_d}(k_{di}^2)$$




**Synthesis of the processes involved in the production of data with SAE**

Source: adaptation from Kolenikov (2014)

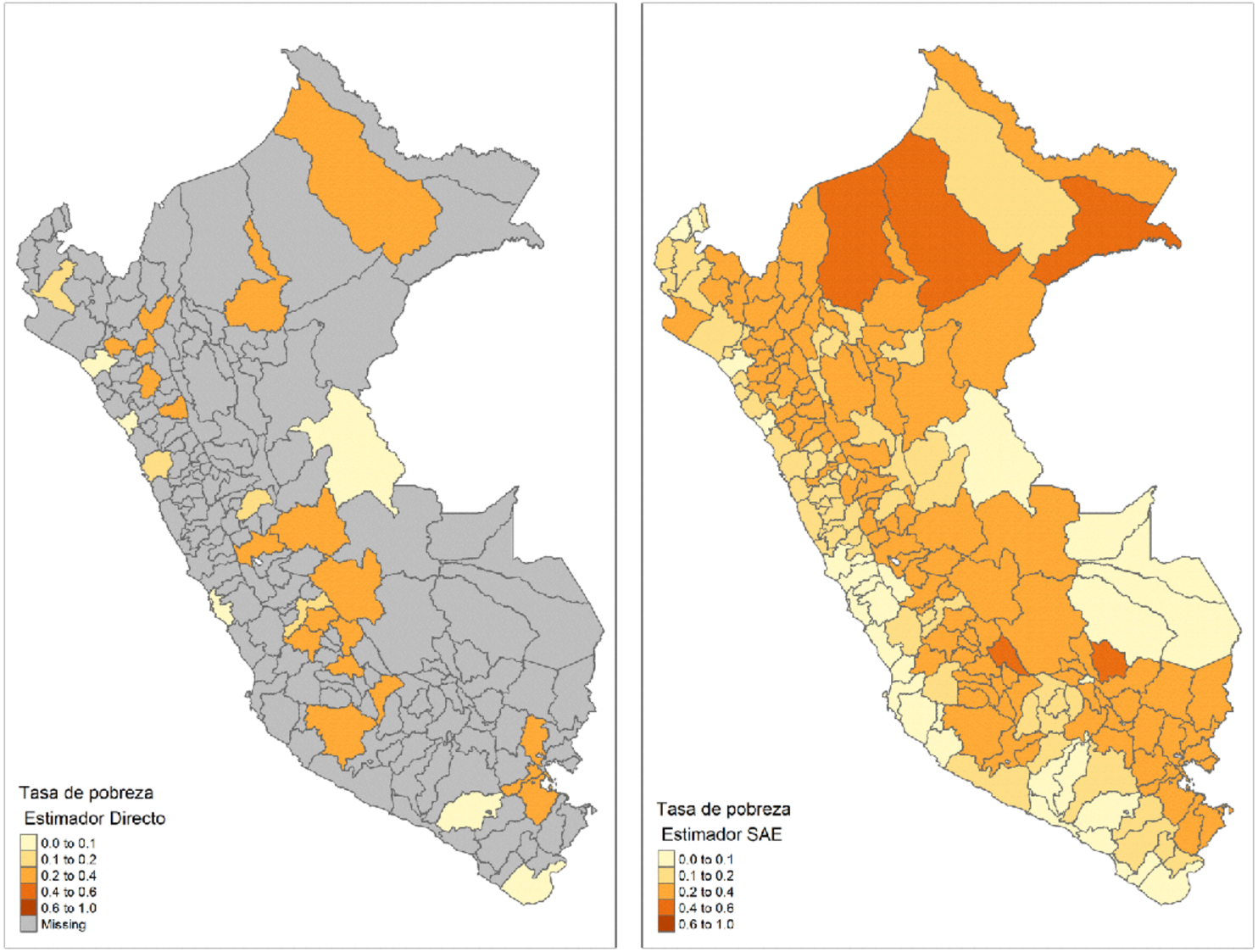
# ECLAC’s poverty maps in Latin-America

# Poverty mapping

As the 2020 Round of Population & Housing Censuses (2015-2024) is being performed, census data-bases are being released. ECLAC uses that data to create and update poverty maps in conjunction with the regular household surveys of the countries in the region.

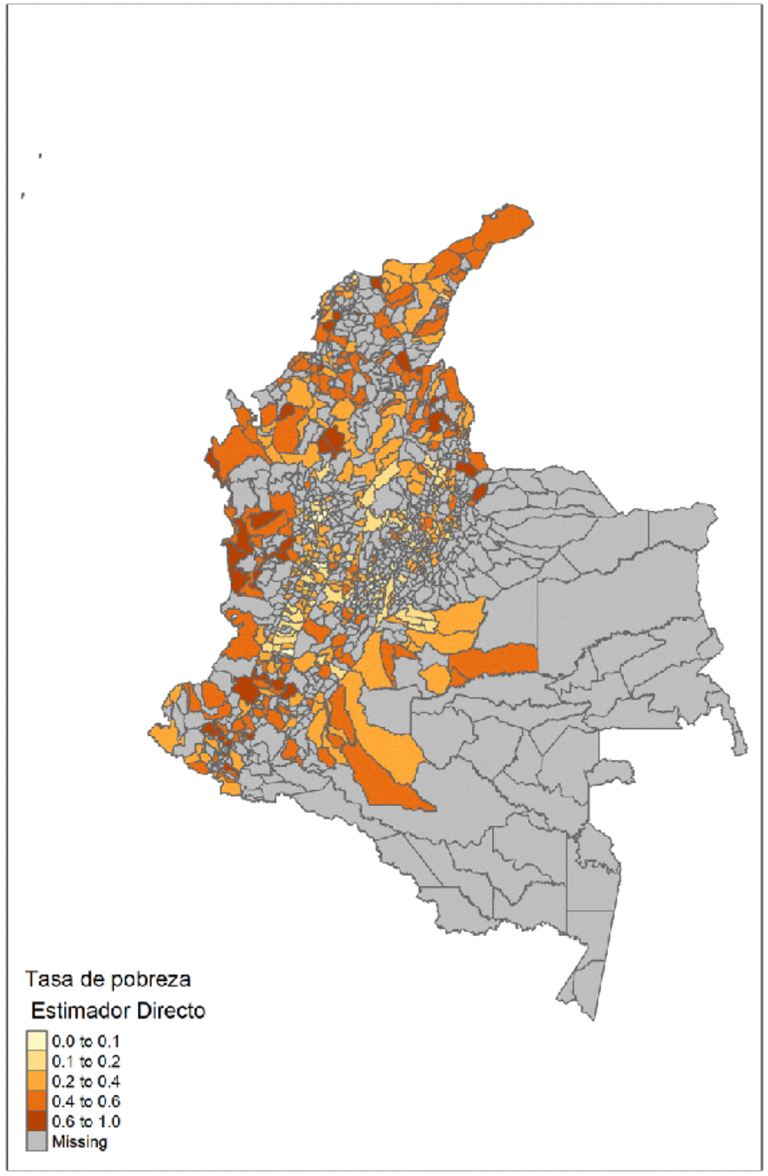
At this moment, we have achieved to create poverty maps for all of the countries that conducted a census:

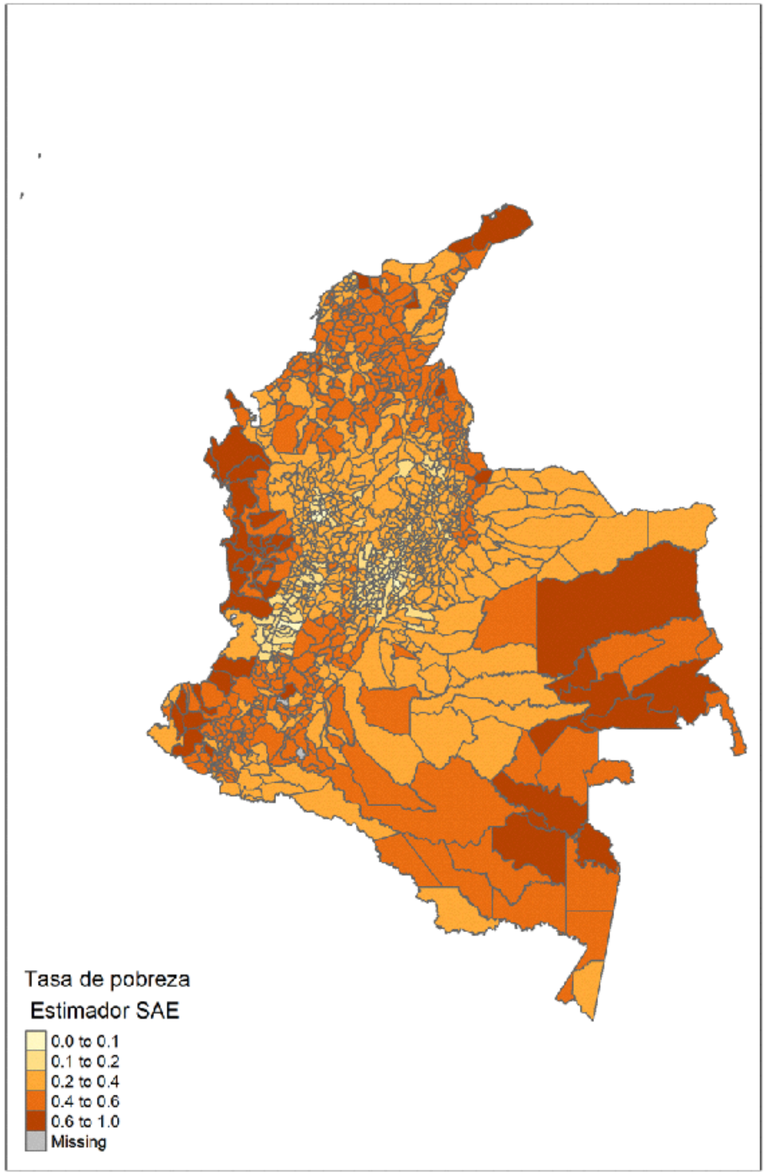
- Colombia
- Perú
- Chile
- Guatemala




# Poverty map in Perú

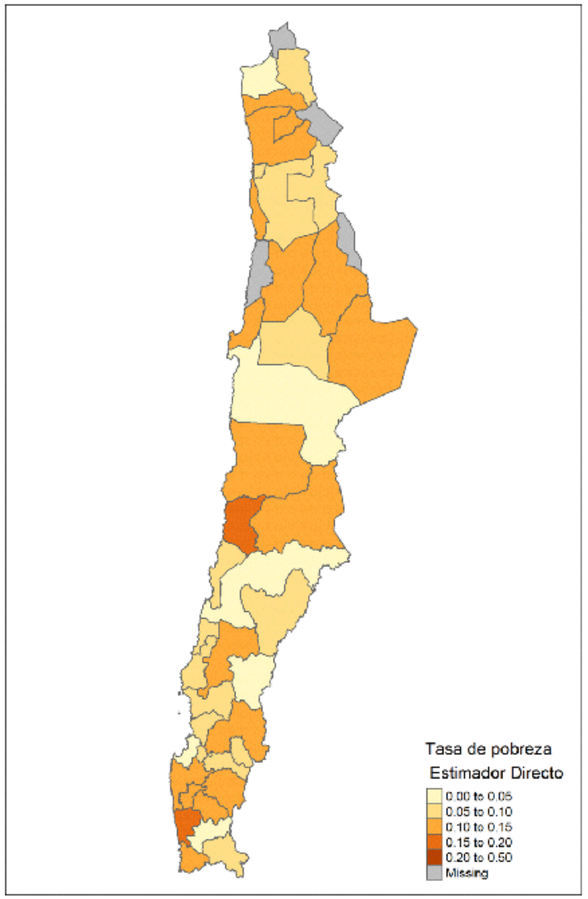
**Source: ECLAC**

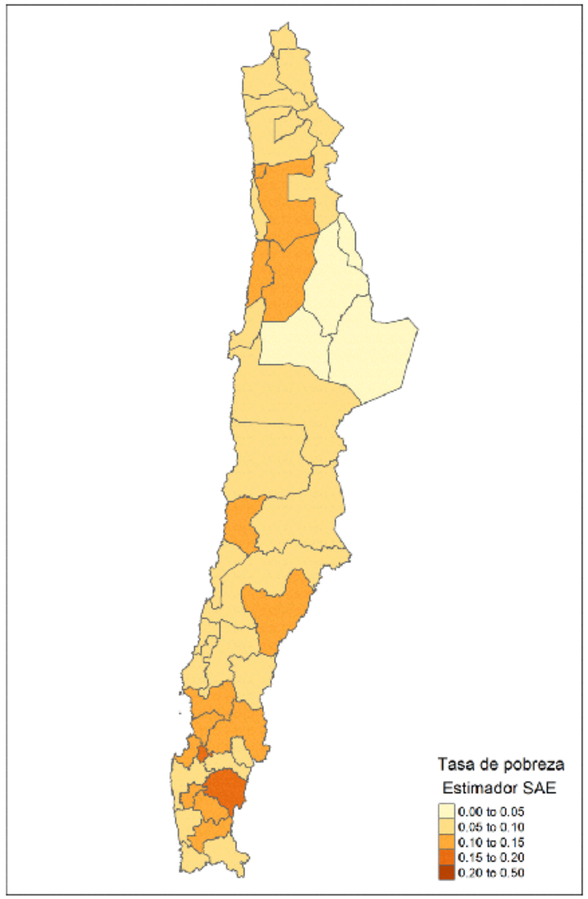


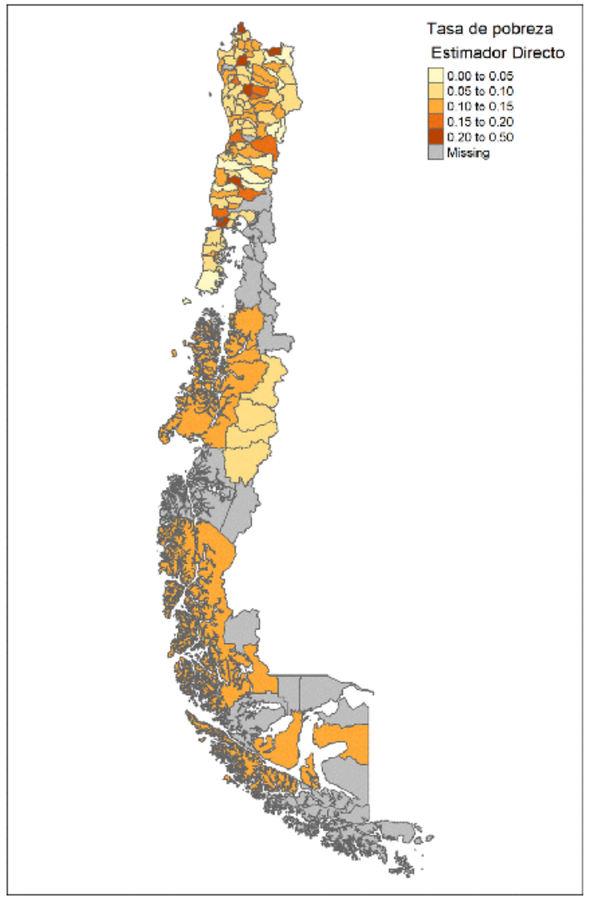



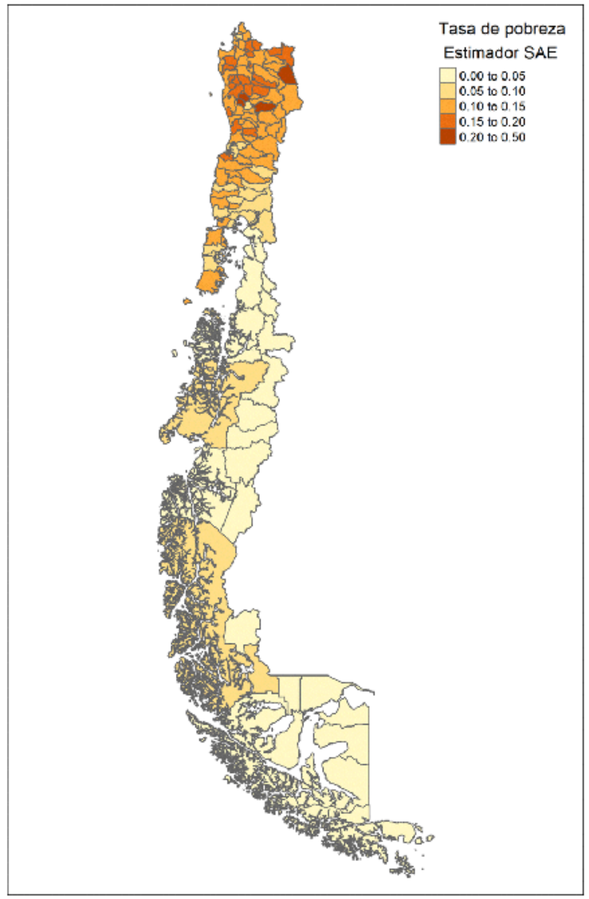

# Poverty map in Colombia

**Source: ECLAC**




# Poverty map in Chile

**Source: ECLAC**

# Capacity building and technical assistance

# SAE applied to other social indicators
# Technical assistance

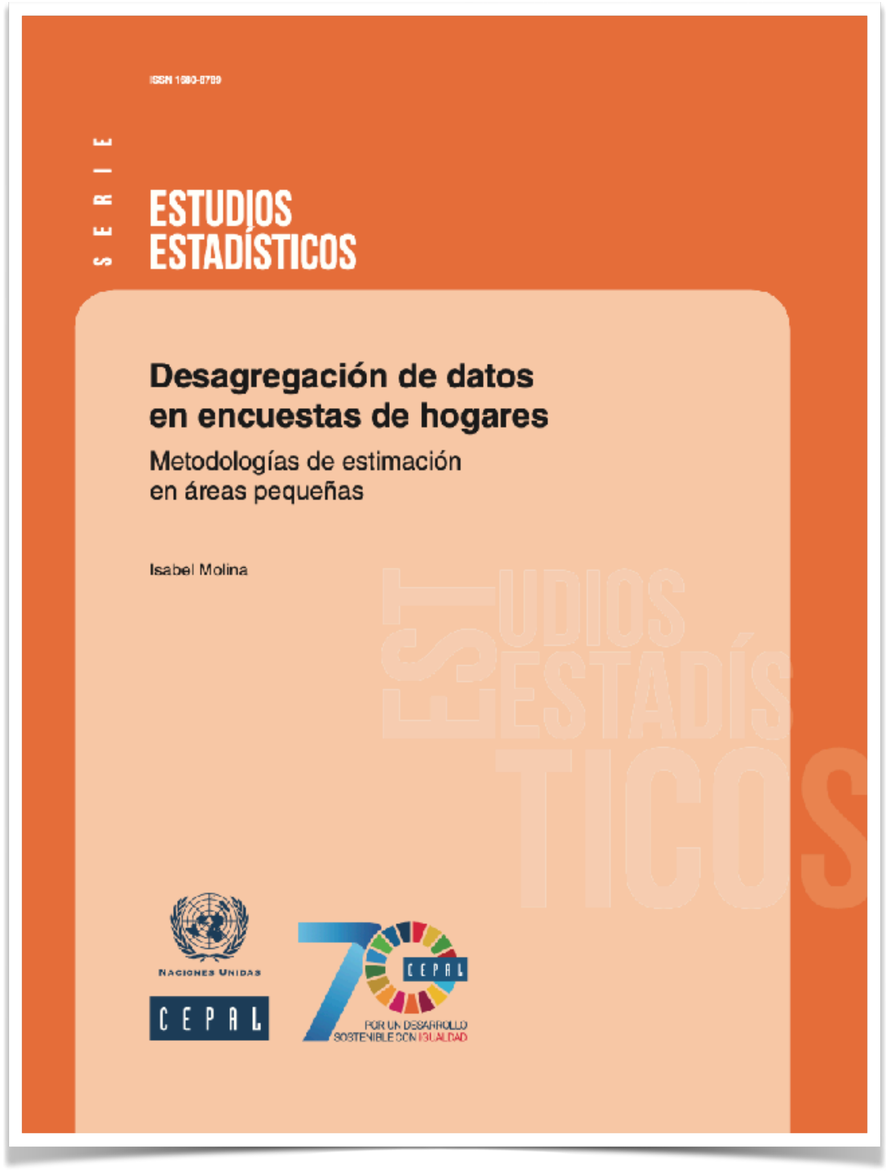
1. Ecuador: malnutrition rates  at the municipality level (SDG 2).
   - FH and BHF models.
2. Chile: poverty mapping at the municipality level (SDG 1).
   - FH model.
3. Chile: victimization rates at the municipality level (SDG 5 and SDG 16).
   - FH model.
4. Perú: family planning indicators at the municipality level (SDG 3).
   - GLMM and plug-in estimation.
5. Colombia: access to justice indicators at the municipality level (SDG 16).
   - FH model.

# Training

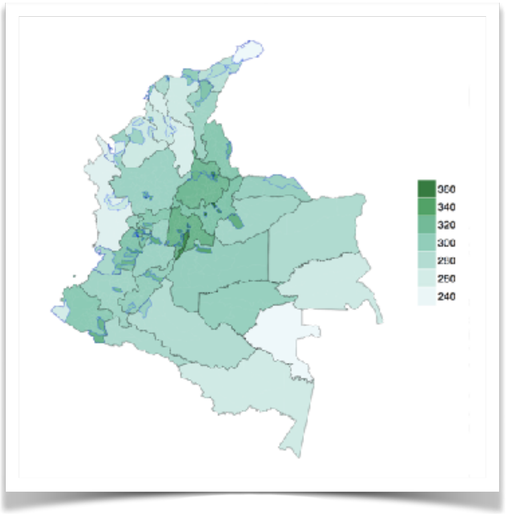
ECLAC's Statistics Division has led face-to-face training in the following countries:

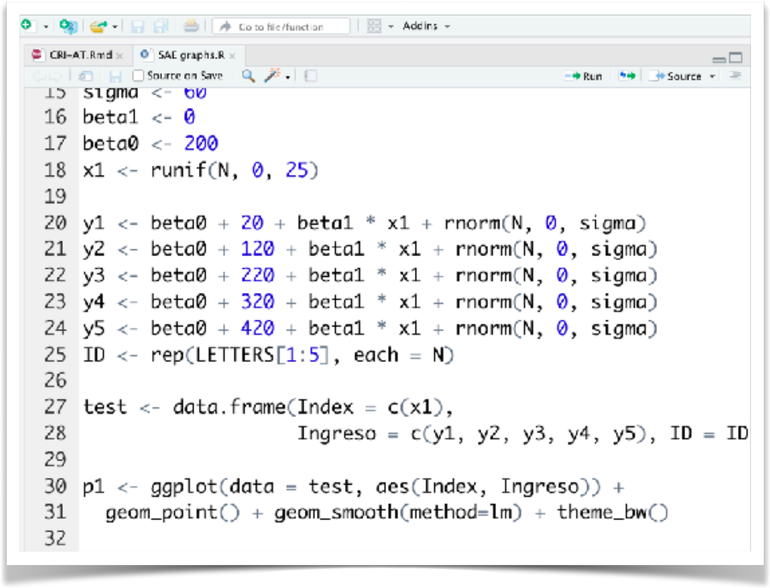
- Colombia
- Uruguay
- El Salvador

In cooperation with UNPFA, the International Course on Data Disaggregation using R will be held, as of the first quarter of 2021, through the ECLAC Moodle platform, hoping to reach all of the countries in the region.

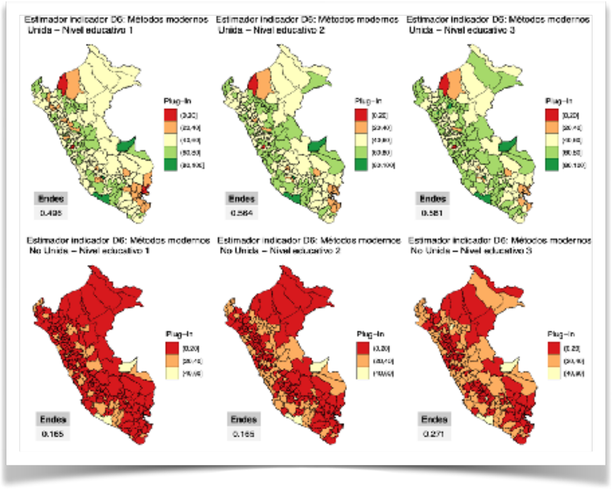






```
15 sigma <- 60
16 beta1 <- 0
17 beta0 <- 200
18 x1 <- runif(N, 0, 25)
19
20 y1 <- beta0 + 20 + beta1 * x1 + rnorm(N, 0, sigma)
21 y2 <- beta0 + 120 + beta1 * x1 + rnorm(N, 0, sigma)
22 y3 <- beta0 + 220 + beta1 * x1 + rnorm(N, 0, sigma)
23 y4 <- beta0 + 320 + beta1 * x1 + rnorm(N, 0, sigma)
24 y5 <- beta0 + 420 + beta1 * x1 + rnorm(N, 0, sigma)
25 ID <- rep(LETTERS[1:5], each = N)
26
27 test <- data.frame(Index = c(x1),
28                     Ingreso = c(y1, y2, y3, y4, y5), ID = ID
29
30 p1 <- ggplot(data = test, aes(Index, Ingreso)) +
31   geom_point() + geom_smooth(method=lm) + theme_bw()
32
```



# Development of a SAE estimation system in ALC

**Source: ECLAC**

# Thank you!