Census Thematic Analysis: Identifying and selecting topics

Willis Odek, PhD, Population and Development Advisor, UNFPA ASRO Arab States Regional Workshop on Census Thematic Analysis for Enhanced Data Utilization, Amman, Jordan, 19-21 May 2025

Outline

- When and how to identify thematic topics
- The role of user consultations in topic identification
- Factors to consider in the selection of topics
- Core census topics
- Common topics covered in thematic analyses
- Other less common thematic analysis topics
 - Geospatial analysis
 - Small area estimation
 - Degree of Urbanization



When to identify census thematic topics

- Choosing the most useful topics for census thematic analysis is crucial for ensuring that users' needs are met.
- Potential topics for thematic analysis should be established early in the census planning and closely tied to the questionnaire design phase.
- The census questionnaire(s) provides the overall framework for the contents of the census and thus generates the 'raw materials' that will be used for conducting thematic analysis.
- There is <u>no limit</u> to the number of topics for thematic analysis. UNFPA's assessment across NSOs globally shows that the topics range from a few to more than 15.

How to identify thematic topics

- User consultations is critical to ensure thematic analysis addresses their specific needs and concerns, and thus enhances utilization of the results
- User consultations happen in two phases:
 - Before census enumeration to ensure appropriate variables are collected
 - After census enumeration to ensure the right analysis and packaging of information for users
- Some of the users can also be engaged in the thematic analysis by leveraging their research/analytical capacity or for quality assurance.
- Multidisciplinary task teams, including line ministries and agencies, can be established to identify potential thematic analysis topics and establish guidelines for their development before census data collection



Questions to guide user consultations for selecting thematic analysis topics

Gaps	Data sources	Data requirements	Relevance
 What are the evidence gaps to be addressed by the thematic analysis? What are the priority information needs for users? In what formats do they need this information? 	• What other data sources exist that can contribute to indicators of interest and how can these data sources be integrated with census data for comprehensive analysis on any given topic?	• What are the legal and statutory requirements for collecting the data required for thematic analysis?	 Are the topics being identified of relevance to a wider audience or intended purely for statistical purposes? Collecting and analyzing data for sectarian or political purposes must be avoided.

Factors to consider in the selection of thematic topics

JNFP



JNFPA

Factors to consider in the selection of thematic topics

- National priority: The needs of the broad range of data users in the country at both the national and local area level needs to be taken into account.
 - Response to overarching development frameworks, national priorities, regional agenda e.g. Sustainable Development Goals (SDGs); ICPD Program of Action, etc.
- International comparability: Achievement of the maximum degree of international comparability, both within regions and on a global basis is critical for both census main and thematic analysis
- **Suitability:** Consideration is needed for the sensitivity of the topics and the willingness and ability of the public to give accurate information on them
- **Resources:** Resources available for conducting the census should guide the selection of topics; an efficient collection of accurate data for a manageable number of topics, followed by prompt tabulation and publication should be emphasized
- Alternative sources: Topics for which no alternative sources exist should be given higher priority

Population census topics: UN Principles and Recommendations

- UN Principles and Recommendations identifies 31 core topics on population:
 - 26 are direct questions asked in census questionnaire
 - 4 topics derived from the core topics
- There are 17 additional topics that can be included depending on user needs

Core topics

Topics

- Place of usual residence
- Place where present at time of census
- Place of birth
- Duration of residence
- Place of previous residence
- Place of residence at a specified date in the past
- Country of birth
- Country of citizenship

Topics

- Year or period of arrival
- Relationship to the reference person of household
- Household and family composition
- Sex
- Age
- Marital status
- Disability status
- Children ever born alive
- Children living

Topics

- Date of birth of last child born alive
- Household deaths in the past 12 months
- Literacy
- School attendance
- Educational attainment
- Labor force status
- Status in employment
- Occupation
- Industry
- Participation in ownuse production of goods

Common topics covered for thematic analysis

Demographic trends

- Fertility (often combined with nuptiality)
- Internal migration/ urbanization
- Mortality (sometimes includes the estimation of maternal mortality as a separate theme)
- Population projections

Demographic groups

- Older persons/aging
- Children/adolescents
- Youth
- Disability status

Socio-economic themes

- Gender
- Housing
- Education
- Employment/ economic conditions

Other less common thematic analysis topics

Multidimensional poverty

Water and sanitation

Geospatial analysis (e.g. access to services)

Population vulnerability to climate change

Population health status

Financial inclusion

Refugees and displaced persons

Ethnicity, language skills, nationality, religion

Small area estimation

Degree of urbanization (DEGURBA)

Geospatial analysis

- The integration of geospatial data during the 2020 census round has significantly expanded the possibilities for advanced analysis and informed decision-making.
 - accessibility to infrastructure
 - strategic location of new infrastructure
 - mapping populations exposed to natural disasters and climate change by integrating census data with geospatial and environmental datasets
 - Service area mapping e.g. geographic service coverage within radius of key facilities such as hospitals or schools
- In-depth analysis of inequality by highlighting disparities in infrastructure access across regions, between rural and urban areas, and among specific demographic groups.

Small area estimation (SAE)

- Small area estimation refers to a variety of statistical techniques for producing granular estimates for subpopulations and smaller geographic entities.
- Representative household sample surveys are usually designed to provide national estimates but do not have large enough samples to produce reliable direct estimates for small areas such as districts.
- Statistical models can provide small area estimates within limits of error by combining census and survey data.
- Small area estimation can also be used to generate population estimates for areas not covered by a census.



Small Area Estimation (SAE)

Survey Data - can be used to compute indicators for the monitoring of different development agenda

BUT

 Surveys are not representative at the lowest administratives units

Census Data - representative at the lowest administrative levels BUT

 Census does not include all data needed to compute indicators of interest

	CENSUS	SURVEYS
LOWEST ADMIN	YES	NO
MOST INDICATORS	NO	YES

We need indicators at the lowest administrative level



General Principles of SAE

- Low resolution indicator data e.g. from representative household surveys
- High-resolution predictor data e.g. from census, satellite imagery, social network data
- Prediction model







Mean household income correlates with road access



Multiple predictors can be brought into a model

SAE for improving census data

- Population models to complete the census quality and coverage.
- Resulting estimates to assess and potentially correct the possible coverage errors identified.
- Estimates offer robust results at national level with potential to produce estimates at subnational level if the data input available allow for it (quality, quantity and coverage)



JNFPA

How SAE works for population estimation

Goal: estimate population in areas where census information is incomplete.

Model establishes connections between observed population data obtained from the census or other surveys with additional datasets that cover the entire study area, referred to as covariates.



JNFPA

(SPM implementation - Dominican Republic -Household counts at municipality level from previous household listing exercise)

How SAE works for population estimation

In multilinear models, the aim is to identify a set of covariates that assist in predicting the population density.



JNFPA

Population = Light Intensity + Travel time to closest school + Error

How SAE works for population estimation

From Census dataset we need to determine which part of the dataset **inform/train** the model and where to apply the predictions.



Vacant

 \checkmark

V

- Incompleted Population <> 0
- Not Covered / Refused



UNFPA



Regression Modeling



Data input and covariates

Human Settlements

This type of spatial covariates help the model predict where and to which degree population is distributed at subnational level.



UNFPA

SAE process

Data collection / compilation

- Population data, number of households
- Census data from previous collections
- Geospatial data (e.g. night lights, roads, relief, etc)
- Administrative Records (employment, school enrolment, etc)
- Settlements, household locations.

Workflow

- Data processing
- Covariates database
 processing and
 consolidation
- Model Definition
- Model parameters estimation

JNFPA

- Model validation
- Results and dissemination

Key considerations

• **Crucial role of statistical models:** Provide population estimates by sex and age, including uncertainty levels.

UNFPA

- Not a substitute: Cannot replace censuses or demographic reconciliations but serve as valuable additional inputs.
- **Improving quality:** Enhance demographic reconciliation and population projections.
- **Filling data gaps:** Aid in understanding population dynamics, especially where demographic reconciliation is lacking.
- Interpreting results: Data are estimates with intervals and confidence levels; should be documented and presented accordingly.









WorldFop

Statistics Division

Application of the DEGURBA Approach to Generate Comparable Urban Data in Select Countries – Overview of UNFPA Led Implementation

.

ۯۯ፨ۯ♀ۯ身ۯۯ؞؞ۯ፼ۯшۯۯᆇۯ₡ۯۯѷ҂҉ҫѽѽѡѽҫҌ҉҄҄Ҁ҄҉Ҫ҄ѽѷѽ҂҄҉ҫ҄Ҫ҄Ҍ҉Ҁ҄ѽѽ҂҄ӷҨѼш҄Ҫ҄҂҉Ҁ҄Ҁ҄ѽѽ҂҄ҫҨ҄Ҭ҄Ҁ҄ѽҧѽѽҧҀѽ҂҉ҫ҄҂҄ҀҨ҄Ҁ



The Degree of Urbanization (DEGURBA)

- A global, people-based definition of Urban Degree: cities, towns and rural areas
- Developed by six international organization (European Commission, OECD and World Bank. Later, FAO, ILO and UN-Habitat)
- Endorsed by the UN Statistical Commission
- The global definition complements, but does not replace the national ones
- To improve global statistical comparability





Advantages of DEGURBA (1)

- Starts from a population grid to reduce the bias generated by the different shapes and sizes of spatial units
- 2. Measures population clusters directly, instead by proxy through building clusters
- 3. Uses the same population size and density thresholds across the globe



Advantages of DEGURBA (

- 4. Captures the urban-rural continuum in harmonised manner
- 5. Defines areas to monitor access to services, not areas defined by access to services
- 6. Proposes a cost-effective approach



A city Lagos, Nigeria

A town Chaguarama s, Venezuela





A village Ayeyarwady region Myanmar

Population by Degree of Urbanization v.s. Nationally Defined Urban Area



Example of Country definition v.s. Degurba definition National Definition of New Global Standardized Urban Degree Definition, Urban/Rural, Lilongwe, Malawi Lilongwe, Malawi New definition Government New definition definition DEGURBA L1 DEGURBA L1 Rural Rural rural MALAW Urban Towns and semiurban dense areas EA boundary EA boundary Cities EA boundary V.S (Source: Census2017) (Source:EC-JRC) *based on global census data *based on alobal census data *based on global census data

The Degree of Urbanisation

- <u>Manual</u> published in 2021, available in English, French and Spanish
- Free online tools and data and training
- UNFPA is planning a regional workshop for countries that can apply DEGURBA to recent census data.

Applying the Degree of Urbanisation

A METHODOLOGICAL MANUAL TO DEFINE CITIES, TOWNS AND RURAL AREAS FOR INTERNATIONAL COMPARISONS

2021 edition



