Exploratory analysis on the contribution of mobile data to the collection phase of the upcoming Population and Housing census

Bogotá D.C, Colombia

November 9, 2017

MINTIC

DANE INFORMACIÓN ESTRATÉGICA

TODOS POR UN NUEVO PAÍS
PAZ EQUIDAD EDUCACIÓN

Topics

1. Background

2. Proposed indicators

3. Lessons learned and conclusions

MINTIC

- The access to data maintained by private entities has been the main challenge due to security, confidenciality and privacy concerns.

- Under the **"Measuring the information society" project,** led by the International Telecommunication Union (ITU) and with the participation of the Colombian Ministry of Information and Communication Technologies (Mintic) and CLARO (Mobile Phone Operator) there is the oportunity to have access to mobile data.

- DANE is exploring the potencial contribution of the mobile data source for the collection phase of the upcoming National Population and Housing Census.

The National Population and Housing Census is the main statistical operation in a country, it provides fundamental and basic socio-demographic information.

FASES CNPV
1:1.800.000

eCenso

Fase	# Municipalities
1	100

Field operation collection strategy

Phase	Municipalities	% Municipalities	Population 2017	% Population
2	480	42,78%	30.357.320	61,60%
3	551	49,11%	16.958.728	34,40%
4	91	8,11%	1.975.561	4,00%
Total	1.122	100%	49.291.609	100%

Dwellings   **Households**   **Persons**

Input: Summary table generated by CLARO: Total events and duration in intervals of 1 hour by BTS (Antenna) for 3 months.
Secure transmission of data through an SFTP Service between DANE and CLARO.

Events:

connections between the device and the BTS

We are working in three indicators considering mobile phone activity and traditional complementary sources to provide additional information for the collection of the Census

I01: Classification of municipalities based on the behavior of their daily activity

I02: DANE – Classification of residential and non-residential areas

I03: DANE – Calculation of activity densities disaggregated by

To provide information on:

- Cities with similar behavior in terms of daily activity
- Identification of residencial areas of interes for the population census
- Hours of the day where is more activity within residential areas as proxy for population density

I01: Classification of municipalities based on the behavior of their daily activity

Objective: Identify clusters of municipalities that share similar characteristics based on their daily activity and complementary socioeconomic and land use variables.

Methodology

Smoothed activity curves - Monday
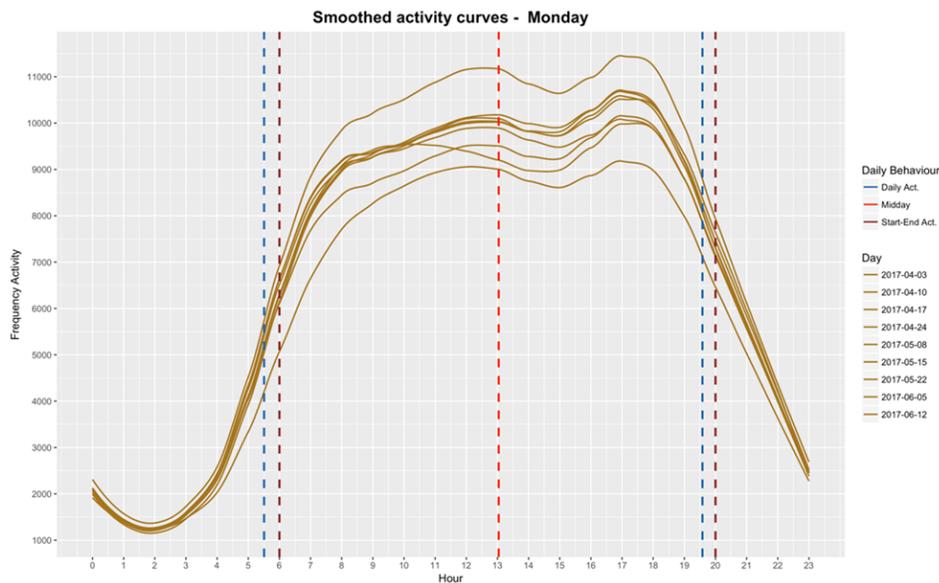
Outcome: Maps

Complemented by:

Socio-economic and Administrative Records

Land areas are identified both by statistics pertaining to their nighttime and daytime activity, as well as their daily activity duration.
Fuzzy clustering methods must be applied over the statistics

I02: DANE – Classification of residential and non-residential areas

Objective: To contribute to the achievement of coverage and quality with respect to the Population and Housing Census from the classification of residential and non-residential areas.

Methodology

Outcome: Maps

$$X_n = Y_n ++ Z_n$$

V_j
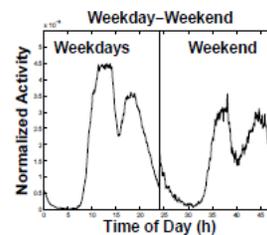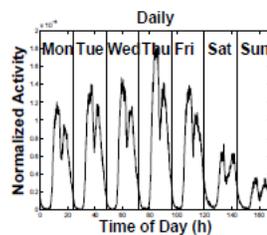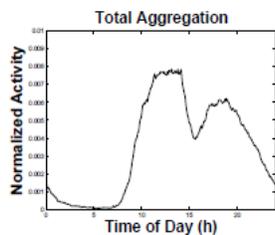
Complemented by:

Socio-economic Administrative Records and cadastral information
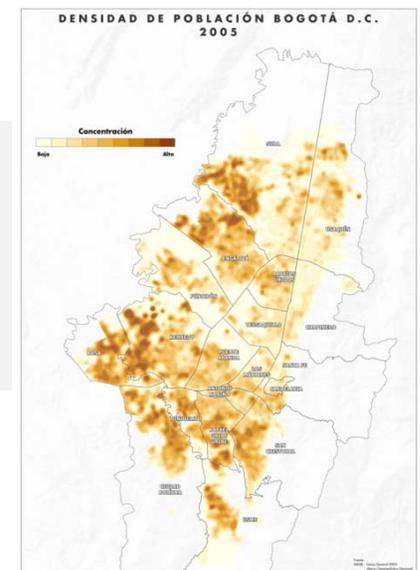
For each city or cluster of regions fuzzy clustering is applied on the entirety of BTS that have a land coverage which can be associated with a land use area. In accordance with the behavior of the weekday and weekend curve.

Objective: To provide complementary information, for the data collection of the population and housing census from the calculation of activity densities disaggregated by areas

Methodology

Outcome: daily profiles
Density indicator
Disaggregation by hour

Each V_j

and its coverage area (Voronoi) is associated with a land use area

A population's activity density indicator is obtained according to land use. Square meters of land might be used.

- Robust Land Use Characterization of Urban Landscapes using Cell Phone Data. Soto, Frías-Martínez **[ ]**

- Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. M. Tiru, Tartu University and Ghent university et al. **[ ]**

- Time patterns, geospatial clustering and mobility statistics based on mobile phone network data. Statistics Netherlands **[ ]**

Conclusions and lessons learned

1. 80% of the **project time** was spent in the agreements, 20% of the time was left for the thematic experts to work with the data.

2. It is possible to **obtain valuable indicators** and insights from aggregated mobile data.

3. We worked closely with the Mobile Phone Operator in order to ensure the quality of the data and the **consistency of the results**.

4. It is important to **complement mobile data** with traditional sources and administrative records.

@DANEColombia @DANE_Colombia /DANEColombia /DANEColombia DANEColombia