

Mapping population in Niger using voters registration lists

Grégoire Lurton

December 13, 2016

Knowing where people live is essential to the design and the evaluation of policies or interventions targetting populations. Population maps is thus an important tool for policy makers, or for actors of social policy or relief actions. Meanwhile, in data poor settings, the quest for precise and exact mapping of population has long proved an illusory goal, for reasons that touch both on the paucity of data, and on technical mapping issues.

This project is exploring an innovative approach to mapping of populations in resource-limited settings. Using voters registration lists, a data source seldom used in demography, we will endeavour to offer a map of population in Niger that will be easily callable and usable by different actors intervening in Niger. To do this, we will challenge current practice in population mapping, and will aim at producing a point map, more adapted to daily usage than raster surfaces.

In the rest of this documents, we will first take a quick historical detour to better understand how our work is justified by an under-investment in population data sources in former French colonies, and how it is also situated in a long standing tradition of population mapping methods. We will then present the data sources we are using, and will discuss the methods use. Finally, we will describe the outputs we hope to produce by the end of the quarter.

1 Mapping populations in sub-Saharan : a historical challenge

1.1 Data issues

The first population map can be traced to Crome's 1785 *Groessen-Karte von Europa* [?], in the time at which Foucault dates the "discovery of population"[?], thus hinting to how the concept of a governed population is consubstantial to the localization of this population. Meanwhile, the production of these maps necessitates the availability and the processing of amounts of data that grow exponentially with the desired level of details. In countries that were formerly colonized by the French

powers, the investment necessary to produce this geolocalized data on population has seldom been done. In the first movements of colonization, maps were mainly produced by the military, and in 1902, Paul Pelet's very first *Atlas des colonies française*[?] did not include a lot on information outside of topographic data[?]. Additionally, there was a radical choice made to use spellings for places in colonies that were adapted to metropolitan French rather than to local languages[?] The second period of colonisation was more interested in economic and commercial exploitation, and thus, the 1922's *Exposition Coloniale* showed some progress in the mapping of geologic and natural resources, but still no population mapping was produced[?]. In 1929, a first map of the population density was produced for Sénégal, using indirect methods to estimate the size of communities, using the amount collected for personal taxes by local authorities[?]. This undervaluation of human data was prolonged in the structures that were left after the decolonizations[?], and reinforced by structural underinvestment in statistical systems in sub-saharan Africa[?]. As a result, the main source of data for population size estimation in most sub-Saharan countries are census, that are performed at best every ten years. In countries with dynamic demographics, this means that most population estimations used are either outdated, or heavily modelled for projection purposes. There is a blatant need to find new data sources to estimate the size of local communities, that are updated more frequently, and are valid at local level.

1.2 Methodological questions

Mapping populations is a tricky task. In contrast to geological data or natural features, population is non continuous in space, and is changing very quickly. In 1935, Fawcett discerned three facts one may want to describe in a population map[?] :

1. The actual number of the people within given areas
2. The density of the population
3. The grouping, or arrangement, of the population.

Each of these facts require a different type of mapping. The number of people in an area can be communicated through discretization of space, and association of a number of inhabitants in each subzone. Density mapping necessitates a more continuous approach to mapping, in terms of scale and of color scale. Arrangement of population can be shown through the plotting of points representing settlements, and with size representing the number of inhabitants. Each of these require different amounts and nature of primary data, and different computational approaches.

The latest advances in population mapping are geared towards the production of density surfaces, presenting a continuous description of where populations live on a territory. This new approach is made possible by the availability of large aggregate datasets for land use, and other usable covariates and the computational ability to interpolate these different data sources for population distribution[?].

Meanwhile, this approach presents two main problems that make its result of little use in country like Niger. First, the output format if these maps, a raster of population distribution, is of little use at a local level, where actors typically use places name and not GPS coordinates. Second, in countries with little urbanization, and poor population data, these rasters end up displaying an overlay of covariate layers more than they present a credible distribution of populations.

Moreover, if this top-down approach is useful to provide macro-level perspectives on population distribution, it is not useful for local level use. Local actors typically think in terms of places names, more than in terms of GPS coordinates. Additionally, attractivity of a urban center for rural population is hard to model through macro level covariates, as it often depends on such factors as tradition, habits or administrative border drawing, as is evident in Figure 1.

In order to offer more useful maps for local actors, we will use an approach supported by very minimal modelling of primary population data distribution, and geared towards the anchoring of population in callable localisation names.

2 Project concept

2.1 Data sources

Voters list as a demographic Datasource A data source that is, to our knowledge, seldom used to inform population mapping for public health purposes, is voters registration lists. There is meanwhile a case to be made for the use of voters' registration data to estimate size of populations. By definition, voters' registration should aim at being as complete as possible a register of adults in the nation. Moreover, in most democracies, some form of national elections are held at least every five years, leading to an update at least partial of voters' registrations. In sub-Saharan Africa, between the years 2015 and 2016, 27 countries were supposed to hold national elections, leading to a theoretical registration of more than half of the adult population of the continent. Finally, for transparency and accountability reasons, electors registries are usually supposed to be easily accessible.

Due to the sensitive and political use of these data, the quality of voters registries are often described as not being trustworthy. On other hand, for the same sensitivity reasons, voters registries are receiving a very high level of scrutiny from many different actors, and are audited sometimes multiple times before validation. This level of scrutiny before validation is much higher than the attention given to a lot of studies or other often used data sources.

The Niger 2016 elections voters registry In Niger, presidential and parliamentary elections were held in February 2016. Voters lists were updated during the second half of the year 2015, under the supervision and control of a mission of the Office International de la Francophonie (OIF). The operations for registration of voters were conducted during the third quarter of 2015. A first version of the voters list was

Table 1: Comparison of different data sources

Data Source	Population Data	GPS quality	Completeness
Voters List	Individual level. Only adults. Age.	None	Exhaustive
RENALOC	Everybody. Locality Level. Gender	Bad	Quasi Exhaustive
RENACOM	Everybody. Locality Level. Gender	Good	Quasi Exhaustive
OSM	None	Perfect	Sparse

published on December 21, 2015, tallying 7,569,172 voters, out of 8,569,309 that were expected based on the 2012 census.

Final lists were validated in early January 2016 after being corrected for some incoherencies noted by the supervisory body. A final report on these lists was published in may 2016. The Comission Electorale Nationale Independante (CENI) later made these lists fully available on its website, from which we extracted, anonymized and formatted the lists.

RENALOC and RENACOM The *Répertoire National des Localités* (RENALOC) is a geolocalized repertory of all localities in Niger. The 2012 version was downloaded as a pdf file from the *Institut National de la Statistique* (INS) website. The tables were extracted in bulk from this file using the Tabula Package, and then processed in Python to recompose the geographic structure of the document. The final data consists in 34507 localities, for which the INS provides the number of inhabitants, by gender, as well as the number of households, and the number of agricultural households. For most of the localities, a GPS coordinate is recorded, as well as the type of locality (neighborhood, village, camp, water well, hamlet).

The 2006 version, named RENACOM, was retrieved in tabular format directly from the INS website.

OpenStreetMap Additional geolocalization method will be extracted from OpenStreetMap (OSM), using the python api for OSM.

2.2 Methods

2.2.1 Localities mapping

In order to map localities in Niger, we can rely on three data sources : the RENALOC, the RENACOM and OpenStreetMap. These three data sources contain similar data but not identical. As noted in Table 1, we see there are different