



HAL
open science

Developing a contextual database using census and other administrative data. The exemple of the “Trajectoires et Origines 2” database

Giulia Ferrari

► To cite this version:

Giulia Ferrari. Developing a contextual database using census and other administrative data. The exemple of the “Trajectoires et Origines 2” database. PAA 2023 Annual Meeting, Apr 2023, New Orleans, Louisiana, United States. 2023. hal-04331761

HAL Id: hal-04331761

<https://hal.science/hal-04331761>

Submitted on 8 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DEVELOPING A CONTEXTUAL DATABASE USING CENSUS AND OTHER ADMINISTRATIVE DATA

The "Trajectoires et Origines 2" Contextual Database



Giulia Ferrari 

Institut National d'Études Démographiques (INED)
giulia.ferrari@ined.fr








BACKGROUND

- Demographic behavior is influenced by the surrounding context
- Information about the context in which people live is increasingly important in understanding demographic phenomena (Caporali et al., 2016)
- National -sometimes regional- macro indicators are commonly used in demography
 - Eg: Fertility: Campisi et al. 2020, Nisén et al. 2020
- Studies integrating a finer contextual perspective are rarer
 - Eg: Alcohol-related mortality and suicide: Winkleby et al. 2006, Martikainen et al. 2004, Agerbo et al. 2006. Migration: Beauchemin & Bouquier 2005. Fertility: Campisi et al. 2022
- Neighborhood and municipality level indicators are difficult to obtain and to access
- There is a need for contextual data encompassing different geographic levels
 - Family, health and education policies are often decided at the regional level
 - Policies regarding childcare facilities are often left to municipalities

OBJECTIVES

- A contextual database for the "Trajectories and Origines 2" (TeO2) survey, INED & INSEE (French Statistical Office) 2019-2020
 - Build a set of indicators characterizing the place of residence of respondents **at the time of the survey**
- Develop a method to build a **database of geocoded contextual indicators** derived from the French population census and other French administrative sources

DATA SOURCES

- 2018 French Population Census** 
 - Access to census micro data is needed to compute contextual indicators
 - In France, access to individual-level census data is provided through a secure data hub after an individual application has been approved
 - Permanent Base of Facilities** 
 - Covers 188 different types of services and facilities at the district level
 - Indicator of "Localized Potential Accessibility to Medical Facilities" 
 - Measure of spatial adequacy between supply and demand for primary health care at the municipal level
 - French Deprivation Index** 
 - Indicator of material and social disadvantage
 - Income, living standards and poverty indicators** 
 - Population density**
 - Housing supply/demand imbalance**
- French administrative sources
Freely accessible & already computed

METHOD

- Census-based indices are percentages and rates obtained by aggregating individual-level information over the selected geographical levels
- The same computation method is applied for each indicator
 - Example: **% of immigrants** in a given district
 - Sum of the weighted number of immigrants residing in a given district /
 - Weighted number of people residing in that district

$$i = 1, \dots, N \quad (\text{population of the district})$$

$$k = 1, \dots, J \quad (\text{districts})$$

$$iw : \text{survey weight}$$

$$immi = \begin{cases} 1 & \text{if immigrant} \\ 0 & \text{otherwise} \end{cases}$$

POP: Population size of the district

$$pop_k = \sum_{i=1}^{n_k} I_i$$

POP_w: Weighted population size of the district

$$pop_w_k = \sum_{i=1}^{n_k} iw_i$$

PCT_IMMI: % of immigrants

$$pct_immi_k = \frac{\sum_{i=1}^{n_k} immi_i * iw_i}{\sum_{i=1}^{n_k} iw_i} \times 100$$

Aggregated index

$$pct_immi_L = \frac{\sum_{k=1}^{n_L} pct_immi_k * pop_w_k}{\sum_{k=1}^{n_L} pop_w_k}$$

DATABASE CONTENT

359x6 indicators at the district, municipal and broader geographical levels

Their choice relies primarily on the survey's main topic (i.e. migration) and on the availability of administrative data at a subnational level

Census-based measures (189 indicators)

- Socioeconomic structure
 - Educational attainment
 - Activity rate
 - Employment rate
 - Unemployment rate
 - Status of employed persons
- Age structure
- Structure by nationality
- Housing
 - Type of dwelling
 - Social housing
 - Homeownership
 - Density of the dwelling
 - Number of persons in the dwelling
 - Household type

Equipment and services (148 indicators)

- Personal services
- Businesses
- Educational facilities
- Health and social services
- Transport & travel
- Sports, leisure and culture
- Tourism

Localized Potential Accessibility to Medical Facilities (11 indicators)

- Spatial adequacy between supply and demand for primary health care at the municipal level

French Deprivation Index

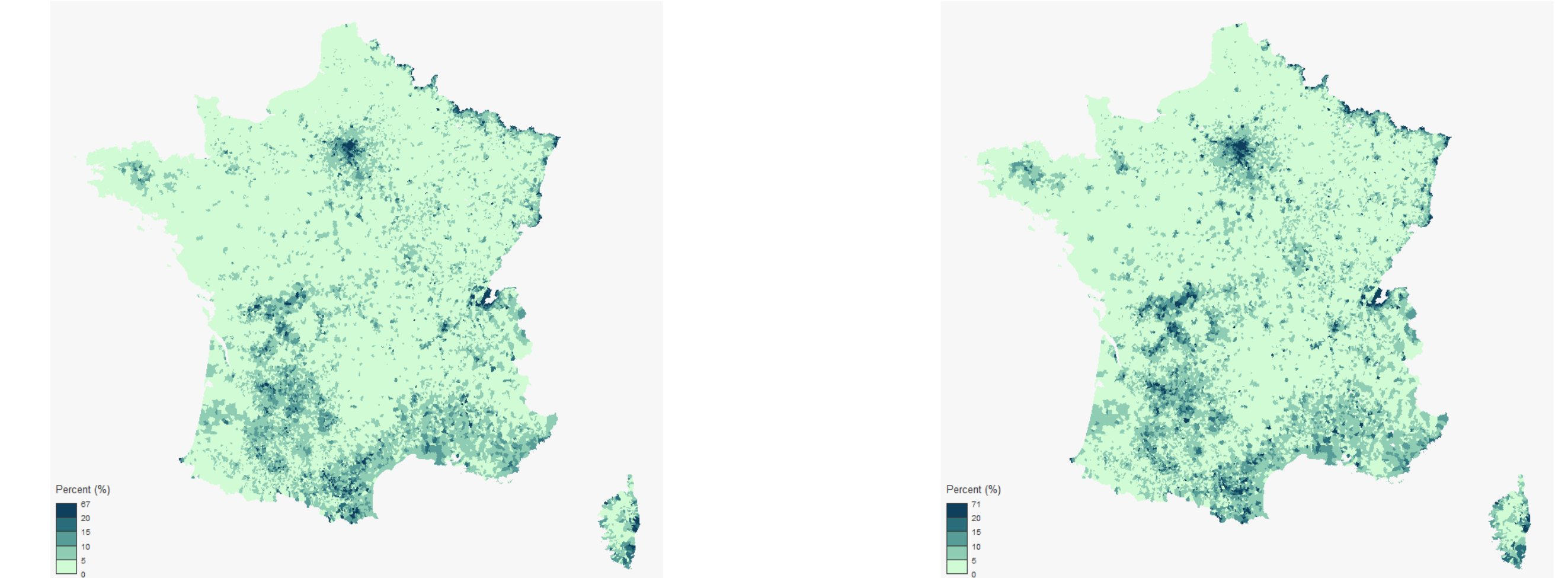
- Indicator of disadvantage considered as a combination of material and social disadvantages at the municipal level

CDB LINKAGE TO SURVEY DATA

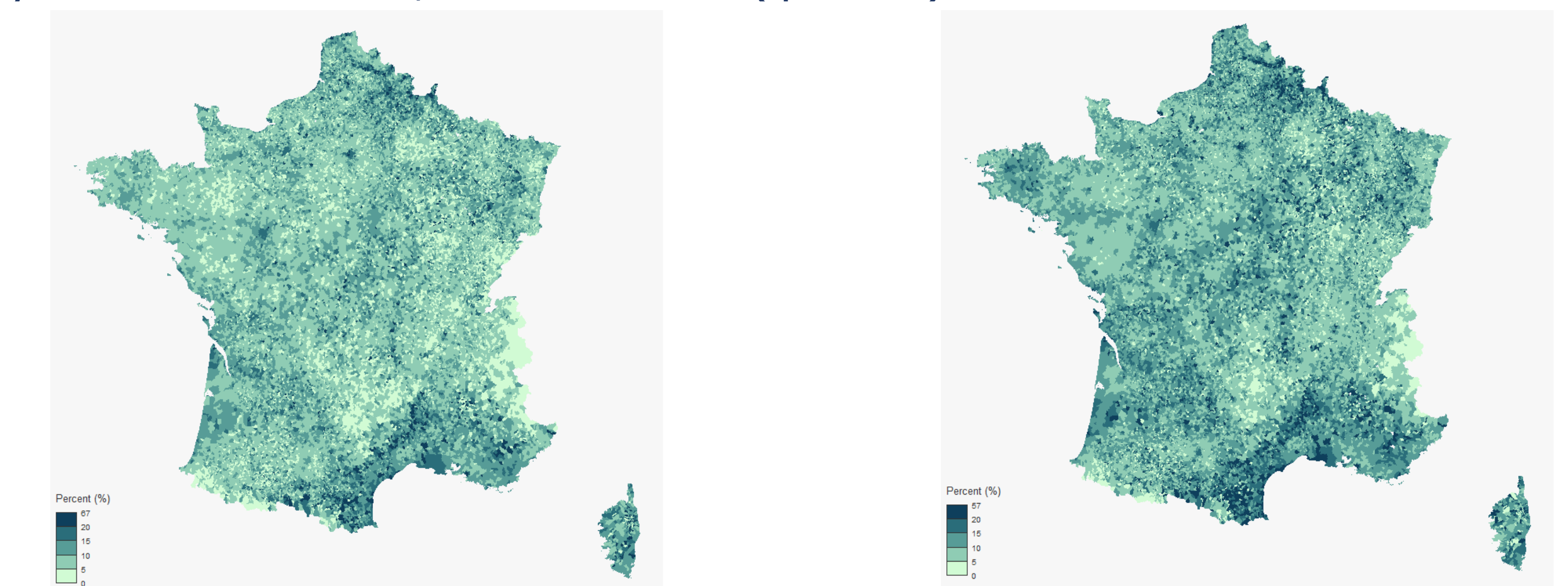
- The contextual database contains indicators for the entire French territory
- It has been **linked to survey information**, by means of the geographical code of respondent's place of residence
- The linked database will soon become available to scientists willing to contextualize individual-level analyses

VISUALIZATION

Immigrant population in France, 2008 and 2018 (quintiles)



Unemployment rate in France, 2008 and 2018 (quintiles)



FURTHER DEVELOPMENTS & APPLICATIONS

VISUALIZATION

- We are developing a Shiny App for **dynamic data visualization** (with Arno Muller, INED)

DATA AVAILABILITY

- Make it available for **record linkage to other geocoded surveys**

RETROSPECTIVE & PROSPECTIVE CDB

- The current version of the CDB describes the French context in 2018
 - It characterizes the survey respondents' current place of residence
- "Which context is it that matter? At which point at individual life stage?" (Goldscheider, 2000)
 - Rosa Weber (INED) is currently **expanding the temporal coverage of the CBD**, using this method and harmonized French population census microdata **since 1968**
 - Computation of the same indicators across a longer time span
 - Linked TeO2 survey respondents' place of residence at age 15, first independent, and previous residence
- We aim at **updating the CDB in the future**, at each yearly census release

A BROADER RANGE OF POSSIBLE APPLICATIONS OF THIS METHOD

- Computation of **new indicators**: extension to other domains, for fertility and mortality research
- Potential **replication** of the database in **other countries**