

IRISX 2009–2019: Stable Census Tract Units for Longitudinal Spatial Analysis in France

Nadia Zargouni* Manon Garrouste[†] Miren Lafourcade[‡]

February 13, 2026

Abstract

This note documents the construction of the smallest stable geographical units in France for the period 2009–2019, called *IRISX*. *IRISX* are time-consistent spatial units derived from French census tracts (*IRIS*) whose boundaries may change over time due to population dynamics. The method relies on a graph-based approach that aggregates *IRIS* sharing the same code over the period, yielding spatial units that remain stable across years. The resulting data are designed for longitudinal empirical research, in particular urban economics applications requiring consistent neighborhood definitions.

1 Introduction and motivation

Census blocks are usually the smallest geographic units designed for the collection and diffusion of sub-municipal data. In France, this geographic unit was named *îlot* until 1999, then *IRIS* ("Aggregated Blocks for Statistical Information") from 1999 onward. They are defined by Insee (National Institute of Statistics and Economic Studies). In densely populated areas, *îlots* generally correspond to a city block, incorporating municipal or cantonal boundaries. In less dense areas (referred to as peripheral zones by Insee - "périphériques"), the equivalent is a district, typically bounded by roadways. Since 1999 *IRIS* units have been defined using a size homogeneity criterion, typically covering populations between 1 800 and 5 000 inhabitants.

For various reasons, the territorial division into *îlots* and later *IRIS* units has changed in size and coding between census waves. These changes can be attributed to factors such as intense urban development, population growth or decline, and municipal mergers or splits. To conduct longitudinal empirical work studying resident trajectories within neighborhoods, it is necessary to define a stable, ad hoc geographic framework over time, otherwise changing spatial definitions may confound true economic or demographic trends.

This project was developed for Garrouste and Lafourcade (2025) in the context of an empirical assessment of place-based policies secondary effect on population sorting across different

*Corresponding author. HEC Paris, ENSAE Paris. The author acknowledges the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-23-CE26-0001 (project URBOPP). Contact: nadia.zargouni@ensae.fr

[†]Université Paris-Saclay (RITM); E-mail: manon.garrouste@universite-paris-saclay.fr.

[‡]Université Paris-Saclay (RITM), Paris School of Economics and CEPR; E-mail: miren.lafourcade@universite-paris-saclay.fr

schooling systems. In order to build an accurate database intersecting census tracts with school district areas in a way that was robust to census tracts reorganization, I constructed spatial units that are consistent over time while remaining as disaggregated as possible. To that end, we introduce *IRISX*, a set of stable geographical units covering the period 2009–2019.

For a given period and reference year which will define the baseline geometry of IRIS, IRISX units correspond either to original IRIS from the reference year whose boundaries remained unchanged over a defined period, or to the smallest aggregated units that ensure temporal consistency when IRIS boundaries changed. In Garrouste and Lafourcade (2025) the period was 2009-2019 and the reference year was 2019. The full replication package reproduces the IRISX outputs used in the paper.

2 Methodology

2.1 Principle

The construction of IRISX follows a graph-based approach inspired by Behrens and Martin (2015). The key idea is to group together IRIS that have shared the same code at some point during the study period. Each IRIS code is treated as a node in a directed graph, and links are created whenever an IRIS in year t corresponds to an IRIS in year $t + 1$.

The resulting IRISX units can be interpreted as weakly connected components of this graph. Each IRISX is thus defined as a grouping of one or several IRIS from the reference year.

2.2 Contribution

Unlike Behrens and Martin (2015), our method does not rely on a pre-existing file documenting all boundary changes by date. Instead, we combine three sources. First, an annual IRIS correspondence table provided by Insee. Correspondence tables, albeit sometimes incomplete, have been made available by Insee starting in 1999. This table was created by replicating the work of Adélaïde et al. (2023), enriched by the latest Insee documentation on IRIS changes. Finally, we define IRISX boundaries using shapefiles of IRIS boundaries for the reference year, produced by IGN (National Institute of Geographic and Forest Information).

This information is sufficient to reconstruct the evolution of IRIS codes and to identify stable aggregations over time.

2.3 Limitations

One theoretical limitation of this approach is its sensitivity to IRIS code recycling, which occurs when a code previously assigned to an IRIS is reused for a different geographical area in later years. This issue is compounded by the lack of detailed documentation on the IRIS code assignment process. In practice, however, extensive testing on the correspondence tables shows that code recycling does not occur in the 2009-2019 data.

A more substantive caveat concerns the temporal depth of the method. Going too far back in time can lead to excessively large aggregations. This is particularly relevant before 1999, when census tracts (formerly *îlots*) were reorganized into IRIS, resulting in substantial boundary

changes. Aggregating units to ensure full temporal consistency in such contexts may yield spatial units approaching the municipality level, which is not well suited for urban economics research focused on neighborhood-level dynamics.

Our method deliberately does not exploit the spatial contiguity of IRIS units. This choice avoids introducing arbitrary spatial aggregation rules (for example, buffer distances or partial neighbor selection). While spatially informed aggregation could mitigate the “over-aggregation” issue, it would come at the cost of additional assumptions. We view this as a trade-off rather than a flaw of the approach.

3 Data and outputs

The algorithm produces two main outputs:

- `IRIS_historiques_IRISX.xlsx`, a historical table mapping each IRIS code to its corresponding IRISX and year of creation;
- `IRISX20092019.shp`, a shapefile aggregating 2019 IRIS boundaries by IRISX identifier.

The shapefiles are provided in EPSG:2154 and can be processed using standard Python geospatial libraries such as `geopandas`.

4 Conclusion

This work is intended for researchers analyzing panel data at the census tract level where tracts evolve over time. In the French context, for studies covering periods beyond the scope of ours, this method is fully tractable, provided that the appropriate correspondence tables and the IRIS boundaries shapefile for the reference year are supplied as input. In other contexts, the code can be adapted, though it fundamentally requires a correspondence table of census tracts and a shapefile of their boundaries. More generally, this algorithm is adaptable to contexts that involve stabilizing evolving nomenclatures over time.

To facilitate further academic research and data reuse, the complete codebase and associated inputs used in Garrouste and Lafourcade (2025) are made publicly accessible. Researchers utilizing these tools are invited to provide formal attribution as detailed in the repository’s citation metadata.

References

- Adélaïde, L., Stempfelet, M., Babut, C., Bulté, A., Ehrhart, B., and Jaccy, N. (2023). *HistorIRIS : table de passage des IRIS de 1999 à 2022*. Technical report, Commissariat Général au Développement durable (CGDD).
- Behrens, K. and Martin, J. (2015). *Concording large datasets over time: The c3 method*. Unpublished paper.

Garrouste, M. and Laffourcade, M. (2025). Place-based policies: A path to opportunity or a mark of stigma for targeted neighborhoods? In *Journal of the European Economic Association*. Conditionally accepted. Revised version of CEPR Discussion Paper No. 17750.