

Étude de la démographie française du XIXe siècle à partir de données collaboratives de généalogie

A. Charpentier^a and E. Gallic^b

^aFaculté de Sciences Économiques de l'Université de Rennes 1 & CREM UMR CNRS 6211

7 Place Hoche, CS 86514

35065 Rennes Cedex.

`arthur.charpentier@gmail.com`

^bInstitut Louis Bachelier & CREM UMR CNRS 6211

7 Place Hoche, CS 86514

35065 Rennes Cedex.

`ewen.gallic@gmail.com`

Mots clefs : Généalogie, Données collaboratives, Longévité, Migration.

Les travaux initiés par Henry (1956) au milieu du XXe siècle fondent le cadre conceptuel de la démographie historique et sont suivis par de nombreuses études menées dans les années 1960 visant à caractériser la population française, à l'aide de l'outil statistique. La démographie historique, à la croisée de plusieurs disciplines (démographie, économie, biologie, sociologie) se penche sur :

- la reconstitution de la population (Blayo, 1975) ;
- la structure des familles (Matthijs & Moreels, 2010) ;
- la migration (Bourdieu, Postel-Vinay, Rosental, & Suwa-Eisenmann, 2004) ;
- les transferts héréditaires (Gavrilov & Gavrilova, 2001).

Les données sont au cœur de la démographie historique. Traditionnellement, les recherches sont menées sur les retranscriptions de registres (paroissiaux, d'Etat-civil, militaires). Les données de généalogie sont également exploitées dans la littérature. Leur intérêt s'est accru avec l'ère du développement des techniques informatiques : la révolution du numérique permet d'accéder dans des temps courts, pour des coûts moindres et des efforts réduits à des quantités conséquentes d'informations relatives à la généalogie. Les sites proposant aux amateurs de réaliser leur arbre généalogique fleurissent et les chercheurs ont commencé à s'emparer de ces données collaboratives (Kaplanis et al., 2018).

Problématique

Nous souhaitons explorer les données collaboratives de généalogie d'un site populaire en Europe pour savoir si le travail de collecte et de saisie effectué par les généalogistes amateurs peut potentiellement être réutilisé en démographie historique pour compléter la connaissance du passé de nos ancêtres.

Données

Ce travail s'appuie sur des données de généalogie issues du site web Geneanet. Les utilisateurs du site créent leur arbre généalogique et renseignent pour leurs ancêtres : les naissances, mariages et décès. À chaque événement est associé un lieu et une date. Pas moins de 700 millions d'enregistrements sont présents dans notre échantillon. Ces enregistrements nécessitent un important travail de nettoyage et de couplage. En effet, comme chaque utilisateur du site construit son propre arbre, il faut apparier les arbres et supprimer les doublons, tout en prenant compte des erreurs concernant à la fois les orthographes erronées des noms et

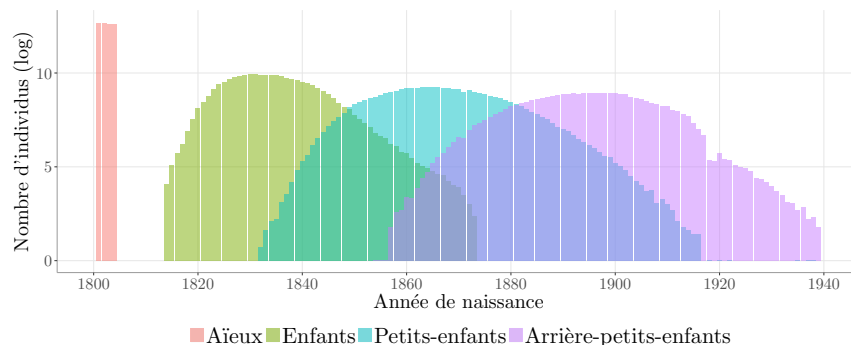


Figure 1: FIGURE 1. Distribution des années de naissance de l'échantillon par génération.

prénoms, ainsi que les dates. Toutes ces tâches sont réalisées à l'aide du logiciel R. Une fois la base nettoyée, nous effectuons un focus sur les individus nés entre 1800 et 1804 en France et sur leurs descendants sur trois générations, ce qui représente 2 457 450 individus. La distribution par génération des années de naissance de ces individus est reportée sur la Figure 1.

Résultats principaux

Nous montrons qu'il est possible de retrouver, bien qu'il subsiste parfois quelques biais, certains résultats de la littérature. Nous proposons d'explorer les caractéristiques temporelles contenues dans les arbres généalogiques pour étudier la longévité, puis de se saisir des caractéristiques spatiales pour analyser les phénomènes migratoires internes en France.

Un document de travail est disponible sur HAL ((Charpentier & Gallic, 2018)), une annexe méthodologique est proposée sur GitHub (https://3wen.github.io/genealogie_fr/).

Références

- Blayo, Y. (1975). La mortalité en France de 1740 a 1829. *Population*, 30, 123. <https://doi.org/10.2307/1530647>
- Bourdieu, J., Postel-Vinay, G., Rosental, P.-A., & Suwa-Eisenmann, A. (2004). La dispersion spatiale des familles: un problème de taille. Les solidarités familiales de 1800 à 1940. *Recherches et prévisions*, 77(1), 63-72.
- Charpentier, A., & Gallic, E. (2018, mars). Étude de la démographie française du XIX^e siècle à partir de données collaboratives de généalogie. Repéré à <https://hal.archives-ouvertes.fr/hal-01724269>
- Gavrilov, L. A., & Gavrilova, N. S. (2001). Etude biodémographique des déterminants familiaux de la longévité humaine. *Population*, 56(1/2), 225.
- Henry, L. (1956). Anciennes familles genevoises. Etude démographique: XVI^{ème} - XX^{ème} siècle. *Population*, 11(2), 334.
- Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., ... Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*. <https://doi.org/10.1126/science.aam9309>
- Matthijs, K., & Moreels, S. (2010). The Antwerp COR*-database: A unique Flemish source for historical-demographic research. *The History of the Family*, 15(1), 109-115.