

Sommaire

1	Le carroyage du recensement de la population : contexte.....	2
2	Production des données au carreau.....	2
2.1	<i>Champ.....</i>	2
2.2	<i>Les variables disponibles.....</i>	2
2.3	<i>Les carreaux.....</i>	2
2.4	<i>Méthode de carroyage du recensement de la population.....</i>	3
2.4.1	Connaître les coordonnées géographiques de chaque adresse.....	3
2.4.2	Estimation des résultats du recensement au carreau.....	4
3	Le traitement de la confidentialité des données au carreau.....	5
4	Mise à disposition des données : mode opératoire et précaution d'utilisation.....	5
4.1	<i>Mode opératoire.....</i>	5
4.2	<i>Précautions d'utilisation.....</i>	5

L'Insee produit des données carroyées de population basée sur la source du recensement de la population depuis le RP 2021. Pour ce premier millésime du carroyage, quelques variables de population sont disponibles (voir partie 2.2.) sur des carreaux de 1 km de côté, sur le champ de la France métropolitaine.

1 Le carroyage du recensement de la population : contexte

Carroyer le recensement de la population consiste à découper la France en carreaux pour y diffuser les résultats statistiques à un niveau faiblement agrégé. Les carreaux permettent de s'affranchir des limites administratives habituelles (Iris, communes, départements, régions) et peuvent être assemblés pour construire ou approcher une zone d'intérêt.

Les premiers résultats du recensement au carreau ont été produits dans le cadre de la réponse aux règlements¹ encadrant le recensement européen 2021 (Census 2021). À l'échelle nationale, l'Insee s'inscrit également dans cette optique de mise à disposition de données carroyées, afin de répondre à une demande croissante des utilisateurs d'accéder à des données finement localisées.

2 Production des données au carreau

2.1 Champ

Les données au carreau sont disponibles sur le champ de la France métropolitaine. La production de ces données aux carreaux n'est actuellement pas possible dans les départements et régions d'Outre-Mer (DROM). La méthode employée nécessite de pouvoir localiser tous les logements d'habitation sur l'intégralité du territoire de chaque commune de plus de 10 000 habitants, ce qui n'est pas encore le cas dans les DROM.

La population considérée ici est la population municipale, à l'exception des personnes résidant habituellement dans des habitations mobiles et des personnes sans abri puisqu'elles n'ont pas de localisation fixe. Sont donc comptabilisées les personnes ayant leur résidence habituelle sur le territoire de la commune dans un logement ordinaire ou une communauté.

2.2 Les variables disponibles

Le carroyage du recensement de la population concerne les variables suivantes :

- population totale
- population selon le sexe (femme / homme)
- population selon l'âge (3 classes d'âge : moins de 15 ans, 15-64 ans, 65 ans ou plus)
- population selon le lieu de naissance (3 groupes : France, pays de l'Union européenne autre que la France, pays hors de l'Union européenne)
- population selon le lieu de résidence habituelle un an auparavant (3 groupes : même résidence, déplacement à l'intérieur de la France, déplacement depuis l'extérieur de la France)

Ces variables de population sont les mêmes que celles disponibles pour le recensement européen.

2.3 Les carreaux

Le recensement de la population 2021 est diffusé sur des carreaux de 1km de côté. Les identifiants des carreaux sont établis à partir des normes fixées par la [directive Inspire](#). Il se décompose de la manière suivante :

« CRS » pour « coordinate reference system »

+code_crs (code projection EPSG)

+« RES » pour « résolution » +« 1000m »

+ « N » pour Nord + coordonnée_y_coin_inférieur_gauche

+ « E » pour Est + coordonnée_x_coin_inférieur_gauche.

Par exemple, le carreau identifié CRS3035RES1000mNyyyyyyExxxxxx est associé à un carreau de

¹ La Commission européenne a adopté un règlement d'exécution relatif à l'établissement d'une action statistique directe temporaire pour la diffusion de thèmes sélectionnés du recensement de la population et des logements de 2021 géocodés selon une grille de 1 km² ([règlement n°2018/1799](#) du 21 novembre 2018).

1000 mètres de côté dont le coin inférieur gauche a pour latitude yyyyyyy et pour longitude xxxxxxx dans le système de projection LAEA (EPSG 3035).

La projection utilisée est la projection Lambert 93 (EPSG 2154). Toutefois, la grille de carreaux a été produite à partir des données projetées en LAEA (EPSG 3035) qui est la projection utilisée au niveau européen. Les contours des carreaux ainsi obtenus ont ensuite été reprojetés en Lambert 93. L'identifiant Inspire du carreau décrit les coordonnées du coin en bas à gauche du carreau selon la projection LAEA.

2.4 Méthode de carroyage du recensement de la population

2.4.1 Connaître les coordonnées géographiques de chaque adresse

Le carroyage du recensement de la population nécessite de connaître les coordonnées géographiques de l'ensemble des logements.

Dans les **communes de 10 000 habitants ou plus**, il existe un répertoire de bâtiments d'habitation géolocalisé² : l'Insee et les communes mettent à jour en continu un répertoire qui liste tous les bâtiments d'habitation de la commune et contient les coordonnées géographiques de chacun de ces bâtiments.

Dans les **communes de moins de 10 000 habitants**, ce type de répertoire n'existe pas. Afin de renseigner les coordonnées géographiques des logements, les données du recensement sont rapprochées des fichiers fiscaux (du cadastre notamment). Deux méthodes sont mises en œuvre :

- la première s'appuie sur des éléments d'adressage (numéro, voie, complément d'adresse, lieu-dit, commune). Lorsque ceux-ci sont assez précis, on peut associer à chaque bâtiment enquêté dans le recensement des coordonnées géographiques issues des fichiers fiscaux. Mais lorsque les éléments d'adressage ne pas assez précis, par exemple s'ils se limitent au nom d'un lieu-dit, il n'est pas possible d'obtenir les coordonnées précises d'un bâtiment.
- une deuxième méthode est alors mise en œuvre. Celle-ci repose sur les caractéristiques des individus résidant dans chaque bâtiment. À partir des informations sur les personnes (date de naissance, sexe, commune de naissance), il est possible de rapprocher un bâtiment enquêté par le recensement à un bâtiment présent dans les fichiers fiscaux dès lors que les deux bâtiments partagent un certain nombre d'individus avec des caractéristiques très proches (même sexe, date de naissance et lieu de naissance, au sein d'une même commune de résidence)³.

Ces deux méthodes sont combinées pour garder les coordonnées géographiques les plus fiables pour chaque logement. Pour les logements pour lesquels aucune des deux méthodes ne fournit des coordonnées géographiques de qualité, les coordonnées sont interpolées de manière linéaire entre deux adresses encadrantes géolocalisées⁴.

Il existe par ailleurs un répertoire des **communautés** (pour l'ensemble des communes), constitué par l'Insee et mis à jour tous les ans à partir de sources administratives, renseignant les coordonnées géographiques de chaque communauté⁵.

2.4.2 Estimation des résultats du recensement au carreau

Le carroyage nécessite également de disposer de données suffisamment fines pour obtenir des résultats au carreau de bonne qualité.

2 Répertoire des immeubles localisés (RIL)

3 À présent, le rapprochement des données du recensement et des fichiers fiscaux géolocalisés se fait à partir du code statistique non significatif (CSNS). Voir : Bénichou Y., Espinasse L., Gilles S., « [Le code statistique non significatif \(CSNS\) : un service pour faciliter les appariements](#) », Courrier des statistiques N°9-2023.

4 Pour en savoir plus : Gallic G., Pages J., « [La géolocalisation du recensement de la population dans les communes métropolitaines de moins de 10 000 habitants](#) », Insee, article pour les journées de méthodologie statistique, mars 2022.

5 Une communauté est un ensemble de locaux d'habitation relevant d'une même autorité gestionnaire et dont les habitants partagent à titre habituel un mode de vie commun (par exemple, les établissements hébergeant des élèves ou des étudiants, les maisons de retraite, les établissements pénitentiaires, etc.).

Le calcul de la population d'une année donnée s'appuie sur cinq années d'enquête du recensement (cycle de cinq ans) et prend comme référence l'année médiane du cycle : ainsi, la population de l'année N-2 est calculée à partir des résultats des collectes réalisées en N-4, N-3, N-2, N-1 et N⁶.

Dans les **communes de moins de 10 000 habitants**, le recensement de la population vivant en logement ordinaire est exhaustif au cours d'un cycle de cinq ans : la production de données au carreau consiste donc simplement à sommer les données de population de chaque adresse du carreau.

Dans les **communautés** et les **résidences touristiques** (quelle que soit la taille de la commune), la population est également recensée exhaustivement au cours d'un cycle de cinq ans : là aussi, la production de données au carreau consiste à agréger les données de chaque communauté du carreau.

Dans les **communes de 10 000 habitants et plus**, en revanche, le recensement de la population vivant en logement ordinaire n'est pas exhaustif. Il s'agit d'un recensement par sondage : tous les ans, ces communes sont enquêtées à hauteur de 8 % de leurs logements, de telle sorte que 40 % de leur population est recensée au cours d'un cycle de cinq ans.

Ce recensement par sondage permet d'obtenir des estimations de population dont la qualité est contrôlée à l'échelle des communes et à la maille infra-communale des Iris⁷. Chaque Iris représentant en moyenne 2000 personnes, le taux de sondage permet d'avoir suffisamment d'observations dans chacun d'entre eux⁸. En revanche, cela ne garantit pas une taille d'échantillon suffisante dans chaque carreau. Contrairement aux Iris, la superficie des carreaux ne s'adapte pas à la densité de l'habitat. Dans certains carreaux, il peut donc y avoir très peu d'adresses recensées, voire aucune, remettant dès lors en cause la robustesse de la méthode d'estimation usuelle du recensement dans les carreaux de ces communes.

Une méthode d'estimation spécifique⁹ a donc été mise en œuvre afin d'obtenir une estimation de la population fiable à l'échelle des carreaux de 1 km de côté. Elle met à contribution de manière plus complète la connaissance de la localisation et du nombre de logements de chacun des bâtiments du Ril. La méthode classique d'estimation de la population affecte un poids d'estimation à chaque bâtiment enquêté en ne tenant compte que du zonage par Iris, qui est l'échelle d'estimation infra-communale actuellement privilégiée.

La méthode d'imputation consiste à répartir ce poids, ainsi que les caractéristiques des bâtiments enquêtés, aux bâtiments non recensés : chaque bâtiment enquêté conserve ses caractéristiques (et un poids de 1) et distribue le reste de son poids à un ou plusieurs bâtiments non enquêtés, ainsi qu'une fraction de ses caractéristiques, à hauteur du poids distribué. Les bâtiments recensés donnent leur poids et caractéristiques à des bâtiments qui leur ressemblent, en termes de nombre moyen de personnes par logement et de structure par sexe et âge (information auxiliaire exhaustive issue des données fiscales). Au final, chaque bâtiment, recensé ou non, est doté de son vrai nombre de logements et de sa vraie localisation (connus grâce au RIL) et dispose de caractéristiques venant du recensement (les siennes ou celles de bâtiments enquêtés donneurs).

Cette méthode garantit de retrouver une cohérence d'ensemble avec les chiffres diffusés par ailleurs à la maille des communes et des Iris.

3 Le traitement de la confidentialité des données au carreau

La diffusion d'informations sur des carreaux peut poser des problèmes de confidentialité à différents égards :

- lorsque les carreaux contiennent trop peu de ménages (risque de ré-identification dans les carreaux ayant une faible population),
- ou lorsque la grande majorité des individus d'un carreau partagent les mêmes caractéristiques sensibles (risque de divulgation d'attributs sensibles),

6 Pour en savoir plus : « **La détermination de la population légale des communes** ».

7 L'Iris est une partie de commune correspondant, le plus souvent, à un quartier d'environ 2 000 habitants.

8 Le tirage des adresses est équilibré sur le nombre de logements au sein de chaque Iris.

9 Pour en savoir plus : Chevalier M., Pilorge C., Gallic G., Guymarc G., Guillo C., « **Méthodes de carroyage du recensement de la population dans les communes de 10 000 habitants et plus** », Insee, article pour les journées de méthodologie statistique, mars 2022

- ou encore lorsque le croisement d'informations à deux mailles géographiques différentes peut conduire à retrouver une information confidentielle sur des intersections (risque de divulgation par différenciation géographique).

Ces trois risques de rupture du secret sont pris en compte et traités afin de ne pas diffuser des données confidentielle à l'échelle du carreau. Ainsi, les carreaux définis comme étant à protéger sont de trois types :

- les carreaux comptant moins de 11 ménages¹⁰ ;
- les carreaux (de 11 ménages ou plus) dans lesquels au moins 80 % de la population partage une même caractéristique sensible ;
- les carreaux pour lesquels un risque de différenciation géographique (croisement d'un carreau et d'un Iris, ou d'un carreau et d'une commune) est identifié¹¹.

Parmi les variables du recensement de la population diffusées au carreau, les informations sensibles sont le fait qu'un individu soit né à l'étranger ou alors qu'il résidait à l'étranger une année auparavant.

La méthode mise en œuvre pour traiter la confidentialité consiste en premier lieu à repérer les carreaux à risque pour l'un des trois risques évoqués ci-dessus. Une seconde étape consiste alors à regrouper ces carreaux entre eux ou avec des carreaux supplémentaires non considérés comme étant à risque, dans un voisinage défini par des grilles emboîtées de différentes résolutions (de 64 km de côté, à 1 km de côté)¹². Les groupes de carreaux ainsi constitués comptent toujours plus de 11 ménages.

Au sein de chaque groupe, les sous-populations sont ensuite ventilées au prorata de la population des carreaux (la population totale du carreau n'étant pas confidentielle). Ainsi, pour les carreaux traités pour confidentialité, la population totale du carreau correspond à la « vraie » population du carreau. En revanche, les autres variables (nombre de femmes, nombre d'hommes, nombre de personnes nées à l'étranger, etc.) sont perturbées. Les variables les plus structurelles, de sexe et d'âge, sont néanmoins peu affectées. Par ailleurs les perturbations ainsi introduites pour garantir la confidentialité préservent les structures spatiales ainsi que la reconstruction d'estimations plus agrégées comme l'échelle départementale.

4 Mise à disposition des données : mode opératoire et précaution d'utilisation

4.1 Mode opératoire

Les fichiers sont disponibles en plusieurs formats : shapefile (shp) et geopackage (gpkg) pour les systèmes d'information géographique et en format csv pour les traitements de bases de données.

4.2 Précautions d'utilisation

Plus de 50 % des carreaux font l'objet de perturbations dans le cadre du traitement de la confidentialité. Par ailleurs, le recensement reposant sur un échantillon dans les grandes communes, les populations calculées par carreaux reposent sur des estimations¹³, ce qui implique une certaine vigilance, en particulier lors d'analyses sur des zones peu denses. La moitié des carreaux ont moins de 9 ménages et moins de 21 individus.

Les données d'un carreau ne doivent pas être interprétées de manière isolée. En zone urbaine, du fait

10 Cette règle de gestion de la confidentialité, consistant à ne pas diffuser l'information brute dans les carreaux comptant moins de 11 ménages, est la même que celle appliquée pour les données fiscales.

11 Voir Costemalle V., « [Detecting geographical differencing problems in the context of spatial data dissemination](#) », *SJI*, vol. 35, n° 4, p. 559-568, déc. 2019

12 Il s'agit d'une méthode de type *quadtree* (subdivision successive d'un arbre hiérarchique). Cette méthode des grilles superposées est déjà utilisée à l'Insee pour gérer la confidentialité des données carroyées de Filosofi (cf. [Branchu et al. \(2018\), « Données carroyées et confidentialité », Insee, article pour les journées de méthodologie statistique de 2018, mars 2018](#)). La méthode est appliquée à partir du **package gridy** du logiciel R.

13 Voir la fiche sur « [La précision des résultats du recensement](#) ».

des fortes densités, on peut considérer que les données sont fiables pour un territoire comptant un nombre relativement réduit de carreaux. En zone rurale, en revanche, il est recommandé de ne pas travailler à un niveau en deçà de la taille d'un canton et de lisser les données.

Il ne faut en aucun cas comparer la population des carreaux issue du recensement à celle issue des données fiscales Filosofi, produites par l'Insee. Ces deux effectifs seront nécessairement différents du fait de la différence de concepts dans ces deux sources.