

Désagrégation spatiale des données de population à l'échelle du bâti



Xavier CLEDAT
Lou HENNENFENT

Tom MONNET
Thomas PELTIER

Orlane FELLER
Pierre-Marie LE GRAND

Julia MORTER
Pablo RIALGIL

Sommaire :

Remerciements	2
Introduction.....	3
I - Cadrage de notre projet.....	4
a. Notre mission	4
b. Nos commanditaires.....	4
c. Les acteurs.....	4
d. Les livrables	4
e. Le Périmètre du Projet.....	5
II- L'étude des données	6
a. Étude bibliographique.....	6
b. Étude des données des bâtiments.....	10
c. Étude des données des populations et des logements	18
d. Exploration des jeux de données inutilisées dans la méthodologie.....	19
III- Ventilation Rennes Métropole.....	29
a. Méthodes tests	29
b. Méthode finale.....	30
c. Perspectives.....	37
IV- Ventilation reproductible (Saint-Malo)	38
a. Données.....	38
b. Méthode du filtrage des bâtiments et du calcul des informations nécessaires à la ventilation	38
c. Méthode de ventilation classique.....	39
d. Méthode de ventilation d'estimation de la population.....	40
e. Méthode de ventilation avancée	43
f. Perspectives.....	46
Bilan global de l'atelier	47
Bibliographie :.....	48
Annexe :.....	49

Remerciements

Avant de démarrer, nous tenons à remercier l'agence de l'Audiar et notamment nos deux commanditaires Johan Poquet et Emmanuel Bouriau qui par leurs bienveillances, leurs conseils et leurs critiques nous ont guidé dans l'exploration des méthodologies. Nous tenons également à remercier le master SIGAT et notamment Boris Mericksay, François Leprince et Erwan Quesseveur pour leur accompagnement durant cet atelier. Sans vous, tout ce travail n'aurait clairement pas été le même.

Introduction

En 2017, date du dernier recensement de la population, la métropole Rennaise comptait 447 429 habitants, répartis dans les 43 communes de l'agglomération. La question que l'on se pose ici est l'échelle à laquelle nous disposons de cette information de la population.

De nos jours, cette donnée est accessible partout à des échelles variées comme le département, la métropole, la commune ou bien à des échelles infracommunales comme les quartiers ou les IRIS. Or, nous ne disposons pas d'informations en libre-accès de populations à des échelles de granularités plus fines, comme par exemple les îlots ou bien encore les bâtiments. Cela permet de garder une donnée confidentielle et anonyme.

Néanmoins, cette information à l'échelle du bâtiment pourrait être déterminante dans les processus d'aménagement du territoire et cela permettrait à des structures, qu'elles soient publiques ou privées, d'affiner leur compréhension du territoire dans le but de l'aménager.

C'est pour cela que l'agence d'urbanisme Rennaise, l'AUDIAR, nous a demandé de travailler sur des méthodologies permettant de connaître l'information de la population à l'échelle des bâtiments.

I - Cadrage de notre projet

a. Notre mission

C'est dans ce contexte de recherche d'information à l'échelle du bâtiment que ce projet est né. Ce dernier est porté par un groupe de 8 étudiants en 1ère année de Master SIGAT de l'Université de Rennes 2. Nous avons pour but d'établir plusieurs méthodologies permettant de définir la composition socio-démographique des ménages à l'échelle du bâtiment. En d'autres termes, notre objectif principal était de déterminer le nombre d'habitants à la maille du bâtiment.

b. Nos commanditaires

L'agence d'urbanisme AUDIAR est le commanditaire du projet. C'est une association créée en 1972 par l'État et la métropole de Rennes afin de remplir des missions de services publics sur le territoire de la métropole. L'AUDIAR intervient à l'échelle de l'aire urbaine de Rennes afin de favoriser la cohérence territoriale et de faciliter la coordination des politiques publiques. Ce projet va permettre de faciliter leurs recherches et de les aider dans leurs prochains travaux.

c. Les acteurs

AUDIAR

Emmanuel BOURIAU
Responsable du pôle Environnement
- Foncier - Transition énergétique

Johan POQUET
Energie - Environnement

UNIVERSITÉ Rennes 2

Xavier CLEDAT
Master 1 SIGAT

Orlane FELLER
Master 1 SIGAT

Lou HENNENFENT
Master 1 SIGAT

Pierre-Marie LE GRAND
Master 1 SIGAT

Tom MONNET
Master 1 SIGAT

Julia MORTIER
Master 1 SIGAT

Thomas PELTIER
Master 1 SIGAT

Pablo RIAL GIL
Master 1 SIGAT

d. Les livrables

Nous nous sommes engagés à fournir aux commanditaires du projet à minima deux méthodologies de ventilation du nombre d'habitants par bâtiment :

- Une méthodologie à partir de la base AUDIAR sur le territoire de Rennes Métropole (non reproductible). Cette méthode sera fournie avec une chaîne de traitement et la cartographie du résultat.
- Une méthodologie reproductible à partir de la BD TOPO v3 IGN disponible sur l'ensemble du territoire français. Elle sera testée en partie sur le territoire de Saint-

Malo Agglomération. Une chaîne de traitement automatisée sera fournie ainsi que la cartographie du résultat.

Ces deux méthodologies seront documentées et justifiées avec les diverses publications étudiées lors de ce projet. Le tout sera complété par un dossier sous format PDF.

e. Le Périmètre du Projet

Ce projet porte sur les territoires de Rennes Métropole et de Saint-Malo Agglomération. Les données dont nous disposons ne sont pas les mêmes sur ces deux étendues. Les processus de ventilations seront donc différents sur les deux EPCI.

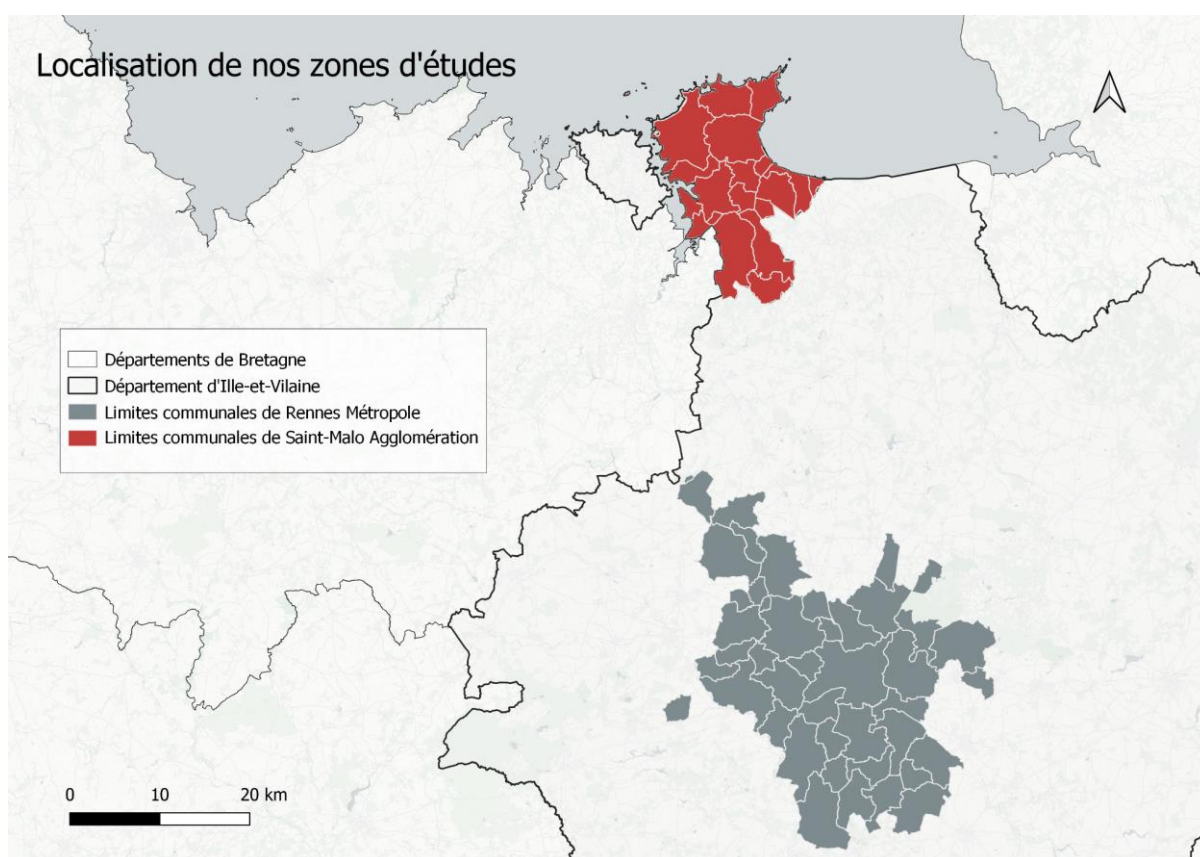


Figure 1 : Localisation de nos zones d'études

Nous vous avons présenté l'étendue de nos missions ainsi que les différents acteurs. Désormais nous allons entrer dans le vif du sujet en vous présentant les données et la bibliographie que nous avons mobilisée dans le cadre de ce projet de ventilation de la population.

II- L'étude des données

Pour commencer sur de bonnes bases, la première étape consistait à étudier les différentes sources de données disponibles pour mener à bien le projet.

a. Étude bibliographique

1. Development of Micropopulation Census through Disaggregation of National Population Census¹

Cette méthode consiste, à partir du nombre de ménages, de population et des bâtiments à ventiler la population.

La première étape consiste à extraire les bâtiments qui contiennent des ménages, en fonction des attributs du bâti (son usage, sa surface, son nombre d'étages, ses pièces etc). Puis, il s'agit d'estimer la surface plancher des bâtiments avec la formule suivante :

$$E_n = \frac{F_n S_n}{Somme_n}$$

Avec, *Somme*, le nombre de pièces dans le bâtiment, *F*, le nombre d'étages et *S*, la surface du bâtiment. Pour les bâtiments collectifs, contenant des espaces communs, il faut multiplier cette surface par un facteur, ce qui permet d'éliminer ces espaces de la surface réelle d'habitation. L'allocation des informations sur les ménages (type de famille et taille du ménage) dans les bâtiments se fait grâce à un tableau de liaison entre la surface du logement et le type de famille qui a le plus de probabilité d'y vivre.

Cette méthodologie, bien que pertinente, requiert une échelle de recensement plus fine que celle de l'IRIS, des informations détaillées des bâtiments qui ne sont pas disponibles dans les bases de données ouvertes françaises ainsi que des tableaux de liaisons n'existant pas en tant que tel en France.

2. Densibati 2009 et 2012²

Avant 2000, l'INSEE recensait les populations par îlots de bâtiments. Depuis, elle le fait selon un découpage par IRIS. Il y a donc une nécessité à trouver une nouvelle méthodologie pour une ventilation des habitants aux bâtiments. L'Institut d'Urbanisme et d'Aménagement (IAU) d'Île-de-France a donc mis au point une nouvelle méthodologie avec des données multiples.

Données en entrées :

- Découpages îlots-1999 et IRIS-2009
- La couche bâtie de la BD TOPO : Fournit les emprises au sol des bâtiments et leur hauteur.

¹ https://cupum2013.geo.uu.nl/download/usb/contents/pdf/shortpapers/110_Akiyama.pdf

² https://geo.valdemarne.fr/portail/system/files_force/note_methodologique_densimos2009_et_densibati2009_2.pdf?download=1

- La couche MOS : Elle ne contient que des bâtiments habités, donc on peut éliminer les bâtiments de la BD TOPO qui sont en trop. Elle fournit également le type d'occupation du sol.

Traitements préliminaires :

- Recalage de la couche IRIS 2009 (sur le réseau routier et les limites communales)
- Enrichissement de la BD TOPO par le MOS en y ajoutant les codes d'occupations du sol
- Estimation du nombre d'étages à 4 m (on le fait pour tous les bâtis puis on gardera uniquement les bâtis habités).
 - Hauteur / 4 = nombre d'étages
- Estimation de la surface plancher de chaque bâti
 - Emprise au sol * nombre d'étages

Méthodologie :

- Identification du nombre de "Maison de ville" ou "Maison individuelle" par polygone :
 - Sélection des bâtis avec codes MOS relatifs à l'habitat individuel et à l'habitat continu bas, à l'habitat autre et aux bâtiments de moins de trois étages.
 - On détermine ensuite le nombre de logements par polygone en fonction de sa localisation géographique afin de tenir compte des disparités réelles de surface des maisons selon les territoires.
- Identification du nombre "d'habitats collectifs" par polygone :
 - Même méthode que précédemment
- On a ainsi le nombre de logements par IRIS, îlots ou communes. On le compare au nombre de logements recensés. Si on a des blancs, soit on essaye de faire concorder avec d'autres postes MOS, soit on ajoute 0 (et on écartera ces bâtiments)
- Ensuite, chaque logement se voit affecter un nombre de personnes (0 possible, car certains logements ne sont pas occupés). On tient compte de :
 - la surface du logement
 - la taille moyenne des ménages estimés à l'îlot / IRIS / commune par types de bâtis
 - population totale des ménages par IRIS ou commune
- La population a été affectée en priorité aux bâtiments de type habitat lorsque cela était possible. Lorsque les densités étaient trop fortes, le reste de population a été affecté sur les postes de type : équipement, administration, santé, école.

Bilan :

Cette méthode est relativement pertinente et présente une multitude d'avantages. C'est notamment à l'aide de la couche d'occupation du sol (MOS) qui se veut être très précise, cette dernière indique directement si le logement est collectif ou individuel. Or, ce MOS est disponible seulement à l'échelle de l'Île-de-France.

3. CERTU - Traitements géomatiques par carreaux pour l'observation des territoires³

Cette méthode a un objectif différent du nôtre, puisqu'il s'agit ici de ventiler la population dans un maillage plus fin et régulier que l'IRIS. Pour autant il présente certaines méthodes géostatistiques de ventilation intéressantes :

- par clipping zonal ou superposition de couche. Cette méthode ne fonctionne que si les mailles de destination sont plus fines que la surface du polygone initial. Cette méthode sous-entend que les données sont réparties de manière homogène, elle ne se suffit donc pas à elle seule et nécessite la juxtaposition d'autres traitements. Mais combinée elle a l'avantage de proposer une ventilation géométrique et proportionnelle, impliquant donc un faible taux d'erreur.
- par noyau / méthode de Kernel. Dans cette méthode, une zone à forte densité de population verra sa densité renforcée par ce calcul, et une zone isolée à forte densité ne sera maintenue que si cette densité est beaucoup plus importante que celle des zones voisines. Il existe diverses fonctions de noyau, avec chacune leurs avantages et leurs inconvénients (distribution uniforme, fonction quadratique, fonction triangulaire, fonction exponentielle, fonction normale, etc.) Au final, cette méthode a tendance à surestimer la population dans des zones de grande surface composées de secteurs industriels, portuaires ou avec un fort relief et peu de bâtiments habités.
- par Krigeage. Cette méthode est un modèle statistique qui comprend de l'auto-corrélation spatiale, c'est-à-dire qu'on considère que les objets rapprochés spatialement tendent à posséder des caractéristiques similaires. Cette méthode permet de transférer à une maille des informations en provenance de mailles voisines par calcul d'une moyenne pondérée. Ici, la méthode a tendance à sous-estimer les IRIS du centre-ville et surestimer les IRIS de périphérie, avec une surface plus grande.
- par interpolation inversement proportionnelle à la distance. Ici la valeur d'une maille est lissée en calculant la moyenne pondérée des mailles voisines. Cette méthode n'est que peu satisfaisante dans les cas où les données en entrée sont clairsemées ou inégales, le lissage est alors trop important et la répartition de la population en est trop uniforme.
- par Spline. Ici, la méthode consiste à interpoler une surface à partir de points, en utilisant une technique de courbure minimum. En réalité, cette méthode est adaptée aux surfaces sur lesquelles les variations sont peu importantes, et en définitive les IRIS du centre-ville sont sous-estimés et ceux de périphéries surestimés.

Finalement, ces méthodes, excepté celle par clipping zonale, apportent trop d'imprécisions et ne seront pas retenues. Le principe d'auto-corrélation spatiale pourrait néanmoins être retenu.

³ <https://hal-lara.archives-ouvertes.fr/hal-02156445/>

4. “Méthode Japonaise”⁴

- Bâtiments
 - Hauteur
 - Nombre d'étages
 - Surface Plancher
- Population
- **Ventilation** → $\text{Pop IRIS} * (\text{Surface Plancher du Bâtiment} / \text{Surface Plancher IRIS})$

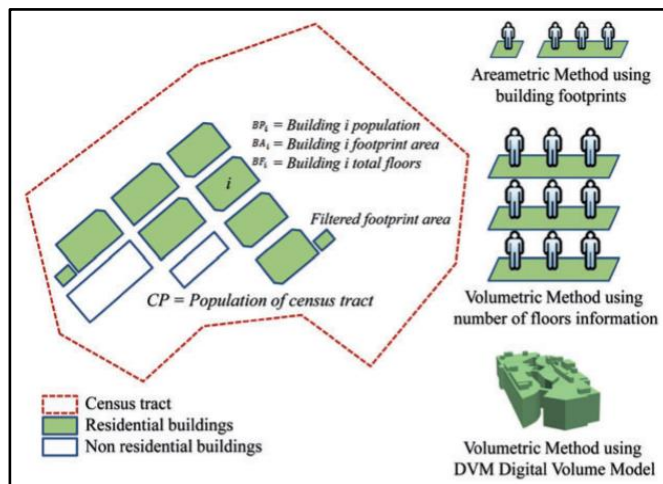


Figure 2 : Graphical illustration of equations

Dans cette étude, les auteurs proposent deux méthodes différentes, une méthode areamétrique et une méthode volumétrique. Dans la méthode volumétrique qui est le plus complexe et le plus fiable, nous trouvons ce que nous avons précédemment défini comme “ventilation classique”, dans laquelle nous utilisons la surface plancher et les étages. Et on retrouve une nouvelle forme, très similaire à cette dernière, qui prend en compte le volume du bâtiment.

- Volume des bâtiments
- Population
- **Ventilation** → $\text{Pop IRIS} * (\text{Volume du Bâtiment} / \text{Volume totale des bâtiments IRIS})$

5. “Méthode du Doubs”⁵

Dans le département français de Doubs ils ont fait un travail de ventilation de la population, qu'ils ont appelé “estimation statistique de la population par bâtiments ou groupe de bâtiment”, et cela introduit des critères différents par rapport aux méthodes vues précédemment. Ils utilisent des données de population de l'INSEE et données de la Direction Générale des Finances Publiques (DGFIP).

- Parcelle Cadastreale
 - Surface habitable
 - Nombre de logements
 - Nombre de pièces
- Bâtiments
- Population

⁴ <https://www.cdema.org/virtuallibrary/images/A%20GIS%20Approach%20to%20Estimation%20of%20Building.pdf>

⁵ <https://bit.ly/2QYvz1T>

- Ventilation → Population de l'IRIS distribuée dans chaque bâtiment en pondérant à poids égal la surface, le nombre de logements et le nombre de pièces.

Compte tenu du caractère confidentiel des données utilisées dans cette méthode, nous n'avons pas accès à la formule exacte qui a été utilisée ni à une partie des données. Cependant, cet exemple nous a inspiré pour ajouter et rechercher de nouvelles données dans les données ouvertes.

Pour faire un bilan sur les différentes bibliographies que nous avons pu explorer, il en ressort que pour ventiler ou estimer la population, on utilise les caractéristiques physiques du bâtiment et notamment la surface plancher (ou parfois le volume). Nous allons voir par la suite si nous avons besoin de nous affranchir de cette méthode de ventilation ou bien en utiliser une autre. Tout dépendra des jeux de données mis à notre disposition (open-data ou par le biais de l'Audiar)

b. Étude des données des bâtiments

La seconde étape de notre projet a été de réaliser un diagnostic des différentes données que nous possédions. L'objectif de cette étape a été de faire un tri dans les données utilisées afin d'avoir une base de données prête pour la réalisation des méthodes de ventilation.

1. Les données qui sont à notre disposition

- **Base Bâtiment Audiar** : Cette base bâtiments nous a été transmise par l'Audiar qui l'a elle-même créée. C'est une base interne à leur structure qui date de juin 2020 (elle n'est donc pas disponible en accès-libre). Elle possède de nombreuses informations plus ou moins intéressantes dont nous avons fait l'étude au préalable. Les données de cette base sont à la fois des fichiers fonciers dit "majic", produites par le Cerema, des fichiers de la BD TOPO v3, produites par l'IGN et également d'autres sources de données variées. Cette base de données sera celle que nous allons utiliser pour la réalisation de la ventilation pour Rennes métropole.

BD TOPO v3, couche "BÂTIMENT" : La BD TOPO est constituée d'un ensemble de thématiques diverses renseignées à l'échelle nationale. Nous avons utilisé celle du bâti. Elle nous a été utile pour le territoire de Saint-Malo puisque la base Audiar est renseignée uniquement sur le territoire de Rennes métropole. C'est une donnée qui est produite par l'IGN et la version que nous avons utilisé est la V3 de BDTopo qui date de janvier 2021.

- **Données Arcep** : Nous avons également fait l'étude de la base de données ARCEP (Autorité de régulation des communications électroniques et des Postes). C'est une base créée à l'origine pour le déploiement de la fibre. Contrairement aux autres, c'est une base de point et non de polygone qui date du quatrième trimestre 2020.
- **Données carroyées de 200 m et 1 km** : ces données proviennent de l'INSEE. Elle se différencie des autres puisque le territoire est découpé en carreaux. Chacun de ces

carreaux diffuse une information qui est finement agrégée selon la taille du carreau. Dans chacun d'entre eux sont détaillées des informations statistiques que nous avons analysées.

2. Nos secteurs test

Pour l'étude des données des bâtiments, nous avons choisi de tester les différentes sources sur trois secteurs différents : le quartier centre de Rennes composé de 10 iris, une commune avec trois iris de grandes zones commerciales et une commune avec un iris unique qui est plus rurale.

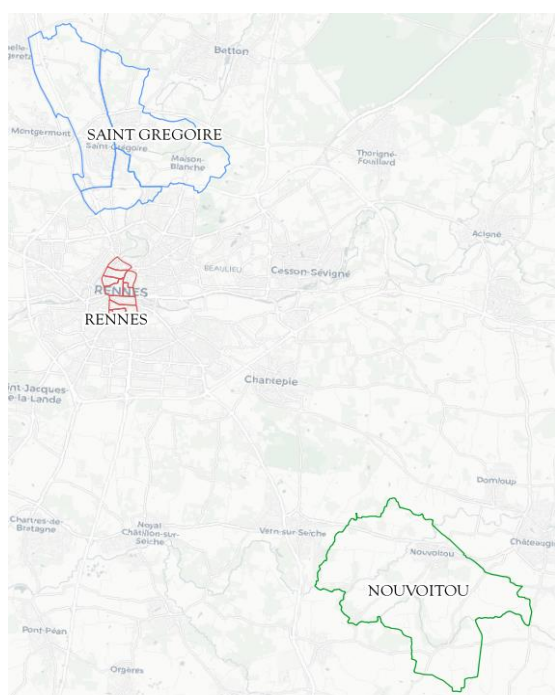


Figure 3 : Localisation de nos secteurs

Saint-Grégoire

3 Iris

Ville moyenne périurbaine avec zones commerciales

Rennes quartier Centre

10 Iris

Centre-ville

Nouvoitou

1 Iris

Ville rurale

3. Étude des géométries

Dans l'exploration des données bâtiments à disposition (BD TOPO et Base Audiar), nous avons fait un diagnostic de ces deux bases de données. En ce qui concerne la géométrie, on a remarqué que les bâtiments n'étaient pas modélisés de la même façon. En effet, pour la BD TOPO, on peut dire qu'un même bâtiment peut être fractionné en plusieurs portions de tailles variées. Les géométries de la BDTopo V3 sont multiples pour un même bâtiment. Or, il est impossible de regrouper ces portions de bâtiments ensemble, car elles ne possèdent pas d'ID communs. On peut le voir sur la figure en dessous, la géométrie BD TOPO est fractionnée (bâtiment en marron clair). Pour la base bâtiment Audiar, les géométries sont plus simples puisqu'un bâtiment correspond à une entité (contour rouge sur la figure ci-dessous).



Figure 4 : Différence de géométrie entre la base Audiar et la BD TOPO V3

Cette différence s'aperçoit lorsque l'on compare le nombre de géométries sur une même emprise. C'est le cas de la commune de St-Grégoire qui compte 3258 entités dans la BD TOPO et 2 865 dans la base Audiar.

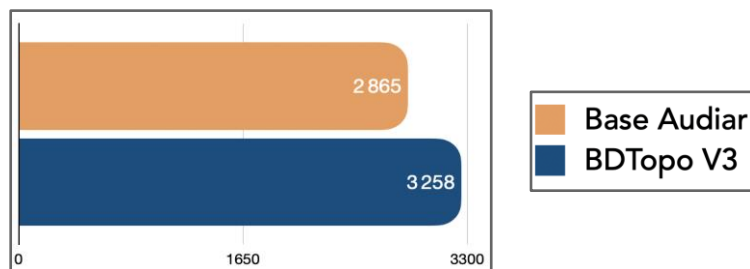


Figure 5 : Nombre de géométries sur nos 2 bases bâtiments sur la commune de St-Grégoire

Les géométries des deux couches sont assez similaires, cependant afin de faciliter les traitements lors de la ventilation, nous avons favorisé la base audiar pour le territoire de Rennes métropole. Pour notre second territoire qui est celui de Saint-Malo, nous n'avons pas eu le choix que de prendre les géométries de la BDTopo bâti puisque c'est la seule donnée bâtie que nous possédons.

4. Étude des surfaces des bâtiments

Pour la surface des bâtiments de la BD TOPO, nous avons étudié la surface minimale d'un logement individuel. En définissant ce seuil, cela permet de supprimer les petites entités et ainsi de réduire le risque d'avoir les entités (bâtiments) sans logements. Les surfaces de moins de 30 m² ont été supprimé de la BDTopo. Pour la base de l'Audiar, nous avons fait le choix de garder toutes les surfaces, car la modélisation était fiable.

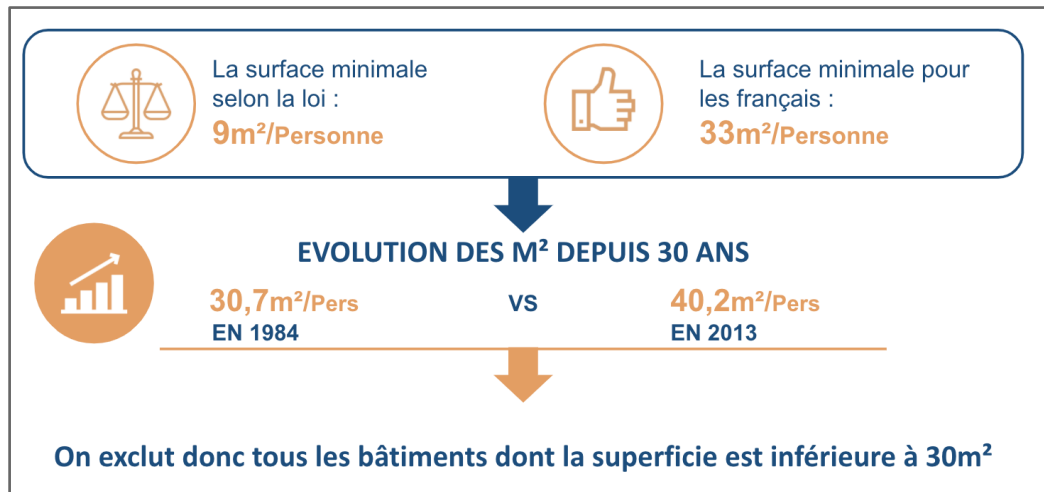


Figure 6 : choix de la surface minimum par bâtiment

5. Usage des bâtiments

Nous nous sommes ensuite concentrés sur les différents usages affectés aux bâtiments pour la base de données de la BD TOPO ainsi que la base Audiar.

BD TOPO v3

La BD TOPO v3 renseigne l'usage des bâtiments sur trois champs :

- Champ "Nature" qui n'est pas exhaustif à 100% donc inutile
- Champ "Usage 1" exhaustives à 100% sur les 3 zones
- Champ "Usage 2" exhaustives à 50% sur les 3 zones
 - Intérêt d'utiliser "Usage 1" en complément de "Usage 2" quand c'est possible

Usage 1

Usage 2

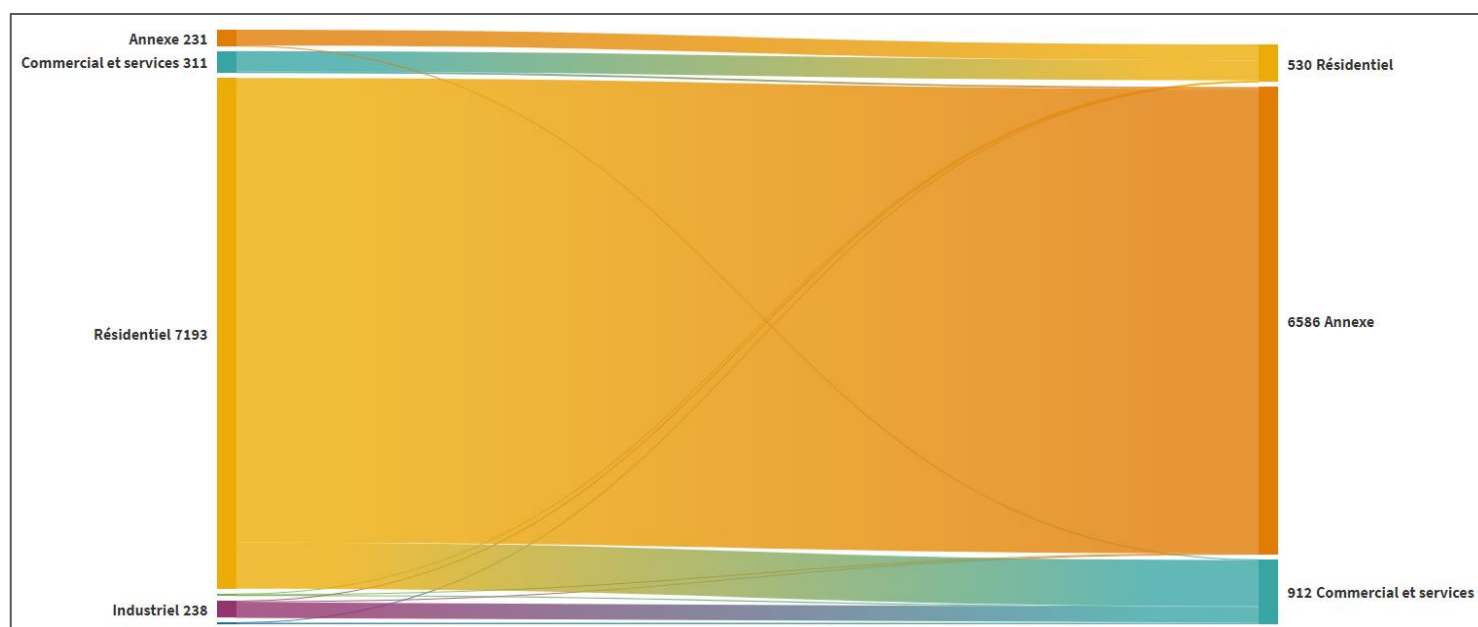


Figure 7 : Diagramme de la complémentarité en "Usage 1" et "Usage 2"

Ce graphique montre la complémentarité qu'il peut y avoir entre ces 2 champs de la BD TOPO. On peut donc utiliser les deux usages en complément afin d'affiner les méthodologies de ventilations prochaines. En effet, on pourrait supprimer les bâtiments qui ne correspondent pas à des bâtiments qui accueillent de la population. On peut le voir sur les tableaux suivants. Ces derniers montrent les modalités des usages 1 et 2 et le nombre sur le territoire de Saint-Malo Agglomération.

Avec complémentarité

Usages 1	Usage 2	Nb modalités	Utilité
Sportif	Résidentiel	1	
Agricole	Résidentiel	14	
Industriel	Commercial et services	219	
Industriel	Annexe	12	
Industriel	Résidentiel	7	
Résidentiel	Commercial et services	649	
Commercial et services	Résidentiel	285	
Agricole	Commercial et services	16	
Annexe	Résidentiel	223	
Résidentiel	Annexe	6544	
Commercial et services	Annexe	26	
Annexe	Commercial et services	8	
Sportif	Commercial et services	20	
Agricole	Annexe	4	

Figure 9 : Liste des modalités avec complémentarité

Sans complémentarité

Usage 1	Usage 2	Nb modalités	Utilité
Industriel	/	306	
Agricole	/	494	
Religieux	/	73	
Annexe	/	5684	
Commercial et services	/	1679	
Indifférencié	/	13731	
Résidentiel	/	24053	
Sportif	/	31	

Figure 8 : Liste des modalités sans complémentarité

Base AUDIAR

En ce qui concerne la couche de bâtiments de l'Audiar, nous avons six champs qui renseignent sur l'usage des bâtiments. Nous avons étudié les usages propres à cette base de données (qui ne proviennent pas de l'IGN) :

- Modèle
- usage_bati 1
- usage_bati 2

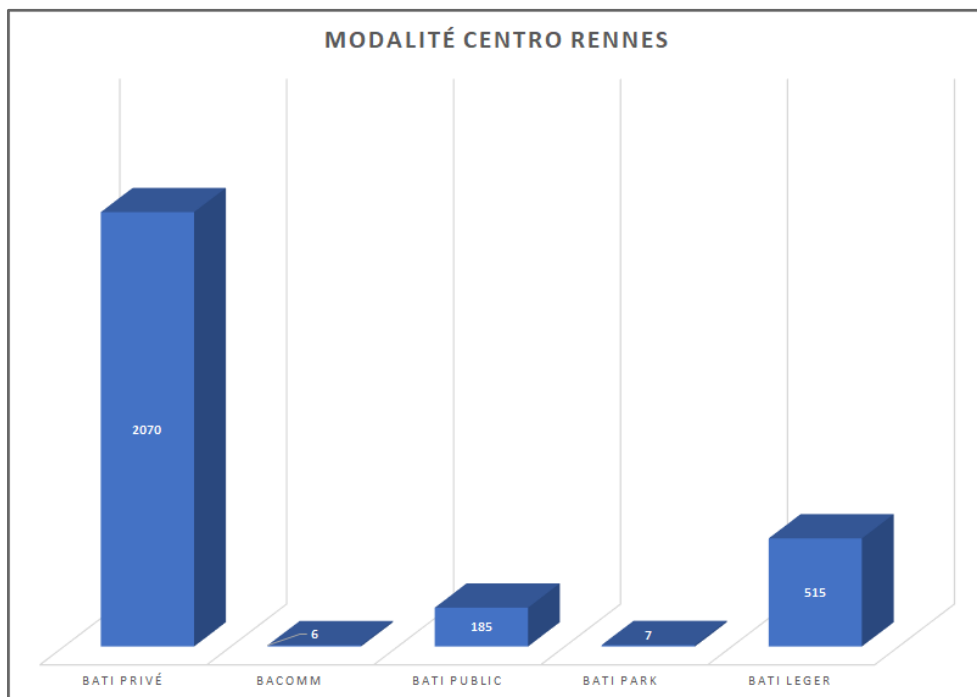


Figure 10 : Modalités du champ "Modèle" de la base Audiar dans le quartier Rennes Centre

Ce graphique représente les différentes modalités de la colonne "**modèle**", que nous avons pour le centre de la ville de Rennes, nous décidons de ne pas l'utiliser, car elle n'est pas assez précise (même si elle est exhaustive à 100%).

Ce graphique représente les différentes modalités de la colonne **"Usage Bâti"**, que nous avons pour le centre de la ville de Rennes (exhaustivité de 100%).

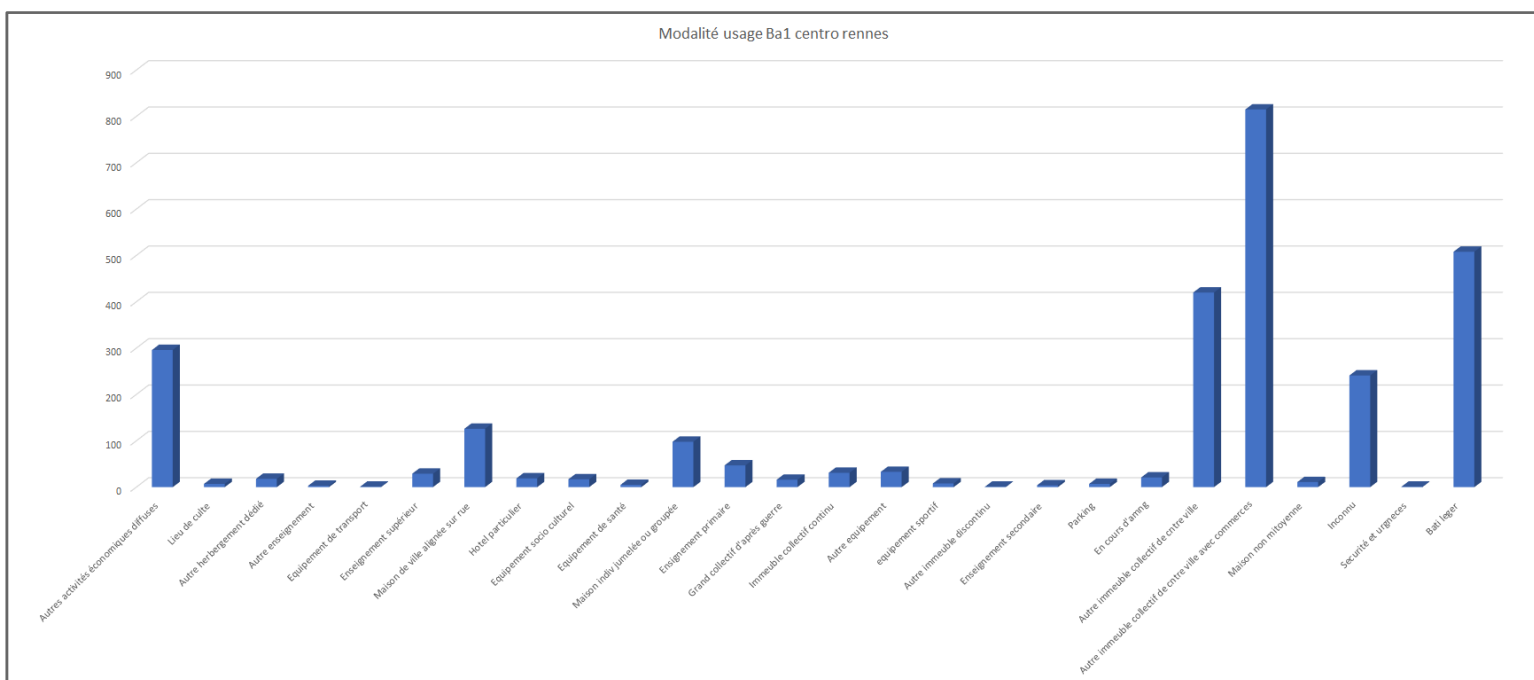


Figure 11 : Modalités du champ "Usage_bati 1" de la base Audiar dans le quartier Rennes Centre

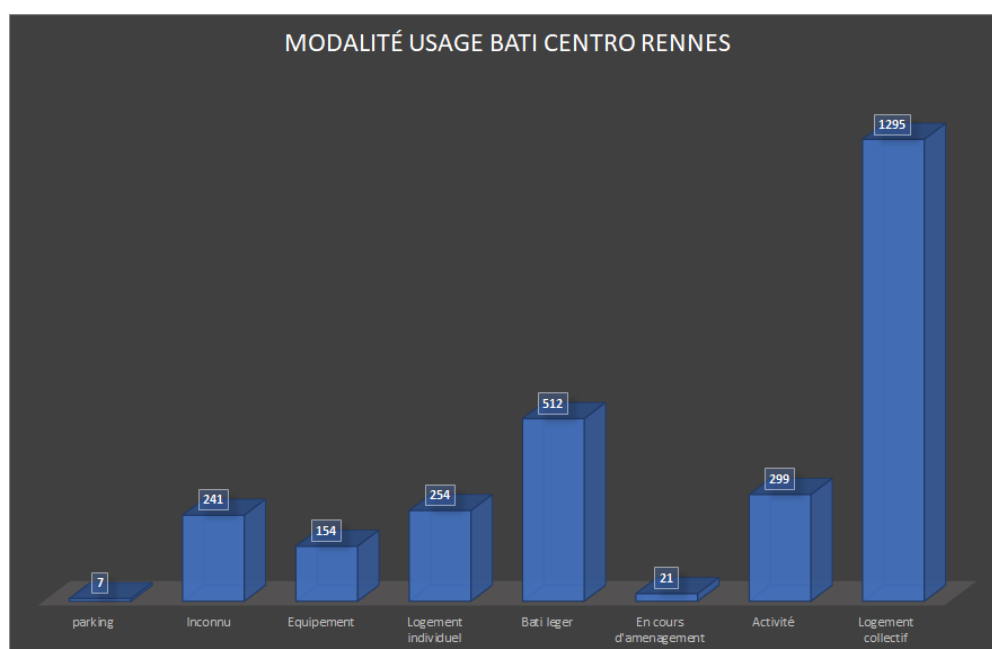


Figure 12 : Modalité du champ "usage_bati 2" de la base Audiar dans le quartier Rennes Centre

Ce graphique représente les différentes modalités de la colonne “**Usage Bati1**”, que nous avons pour le centre de la ville de Rennes. Les modalités sont trop nombreuses, mais cette colonne nous permet de connaître beaucoup plus d’indications sur les bâtiments. De plus, nous avons également une exhaustivité de 100%.

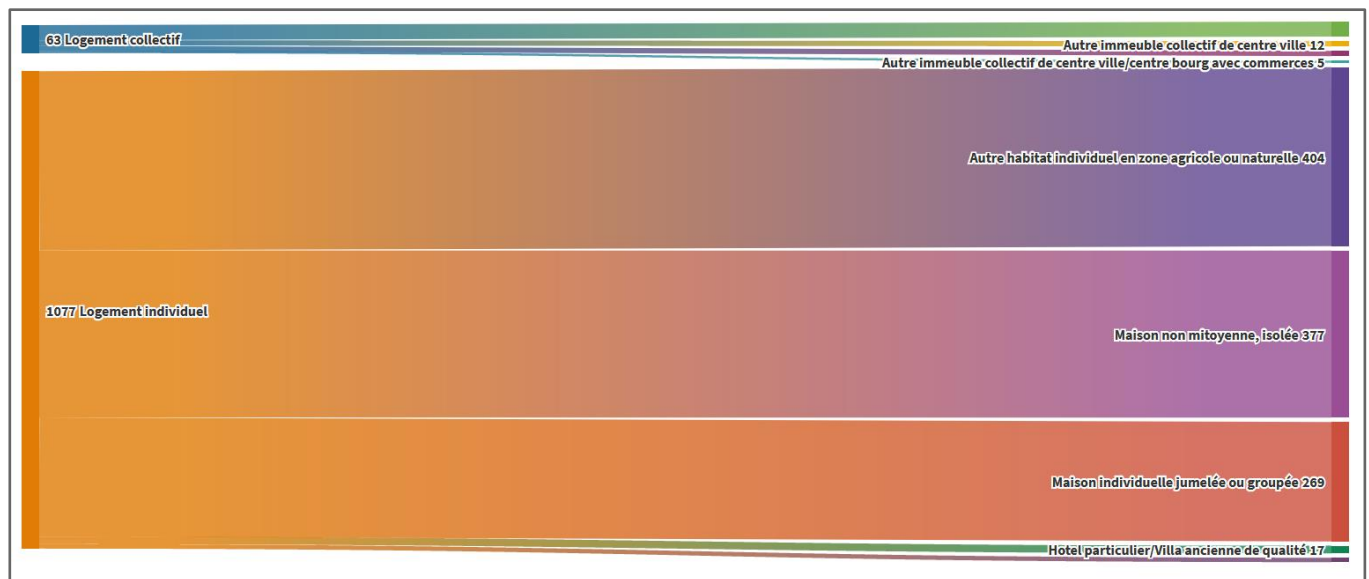


Figure 13 : Diagramme de complémentarité entre les champs "Usage_bati 1" et "Usage_bati 2" dans la base bâtiments de l'Audiar pour la commune de Nouvoitou

En conclusion de cette base de données bâtiment de l'Audiar, on peut dire que les usages Bâti 1 et 2 sont cohérents et ne se contredisent pas l'un l'autre. En outre, on s'est rendu compte que le champ "Usage_bati 1" était largement suffisant pour nous renseigner sur le type de logement auquel on faisait face (logement collectif / individuel).

c. Étude des données des populations et des logements

1. Base infracommunale 2017 du recensement de la population

La base infracommunale 2017 du recensement de la population, est, comme son nom l'indique, issue du recensement de la population française effectué par l'INSEE tous les 5 ans de manière exhaustive pour les communes de moins de 10 000 habitants et par échantillonnage de 40% de la population pour les villes de plus de 10 000 habitants. Cette base contient ainsi de multiples informations relatives à la population et aux logements : sexe, âge, activité, professions exercées, caractéristiques des ménages ou encore la taille et le type de logement.

Dans le cadre de notre projet de ventilation de la population, nous avons utilisé cette base de données pour identifier les personnes provenant des communautés. Cela englobe les résidents d'EHPAD et maison de retraite, les foyers de jeunes travailleurs ou encore les prisons. Nous expliquerons dans la partie suivante comment ce fichier nous a permis d'identifier ces personnes.

2. Fichier de détail logement

Le fichier de détail logement est un fichier assez méconnu de l'INSEE, dérivé du fichier Logement classique. Son nom complet est "Logements ordinaires en 2017 Recensement de la population - Fichiers détail". Ici, chaque entité du fichier fait référence à un bâtiment résidentiel de la zone d'étude. Cependant, les données ne sont pas spatialisées afin de garantir l'anonymisation. Dans ce fichier, on retrouve des informations comme le type de bâtiment (individuel / collectif), si c'est du parc privé ou social, mais aussi le nombre de personnes/ménages y résidant ou encore l'IRIS dans lequel se trouve le bâtiment.

Grâce à des tableaux croisés dynamiques de comptage, avec comme variable de regroupement l'IRIS, nous avons pu réaliser une synthèse des logements et des habitants à l'IRIS. Ainsi, ces traitements nous ont permis d'obtenir les informations suivantes :

- Le nombre de ménages et d'individus habitant des bâtiments résidentiels
- Le nombre de logements collectifs et individuels
- Le nombre de logements du parc privé et du parc social

Ce fichier détail retravaillé nous a ensuite servi de socle pour ventiler notre population dans les logements, mais nous a également permis, avec la base infracommunale du recensement de la population, d'identifier les personnes des communautés. En effet, en faisant (Population municipale issue de la base infracommunale - population des logements issue du fichier détail) on obtient ces personnes des communautés.

d. Exploration des jeux de données inutilisées dans la méthodologie

1. Les données ARCEP

Nous avons ensuite essayé d'utiliser les données ARCEP. Cette base de données distingue les bâtis de type "pavillon" et "immeuble" ce qui nous intéressait tout particulièrement. Pour utiliser l'information des points ARCEP nous avons besoin de les joindre à notre couche bâtie. Le problème avec cette jointure est que les points ARCEP peuvent se retrouver en dehors des bâtis ou alors nous pouvons nous retrouver avec plusieurs points par bâti. Nous avons donc tenté de trouver une méthode pour affecter nos points à nos polygones. Nous avons en entrée 147305 entités de bâtis et 154686 entités de points sur notre emprise test de Rennes Métropole.

Si nous effectuons une simple sélection par localisation nous voyons que plus d'un tiers des bâtiments ne contiennent pas de points.

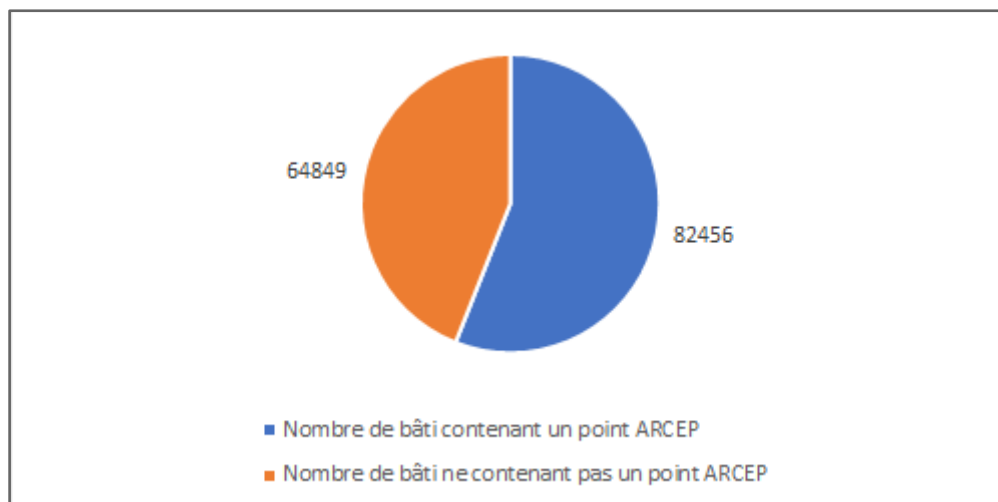


Figure 14 : Nombre de bâtiments contenant un point ARCEP : Sélection par localisation (contient)

Nous avons donc utilisé avec nos bâtis ne contenant pas de point ARCEP à l'étape précédente l'outil "Joindre les attributs par le plus proche" avec un nombre maximum de voisins les plus proches à 1 et une distance maximale de 10 mètres. Nous voyons qu'encore plus d'un quart des entités polygonales ne se voyait pas attribuer de point.

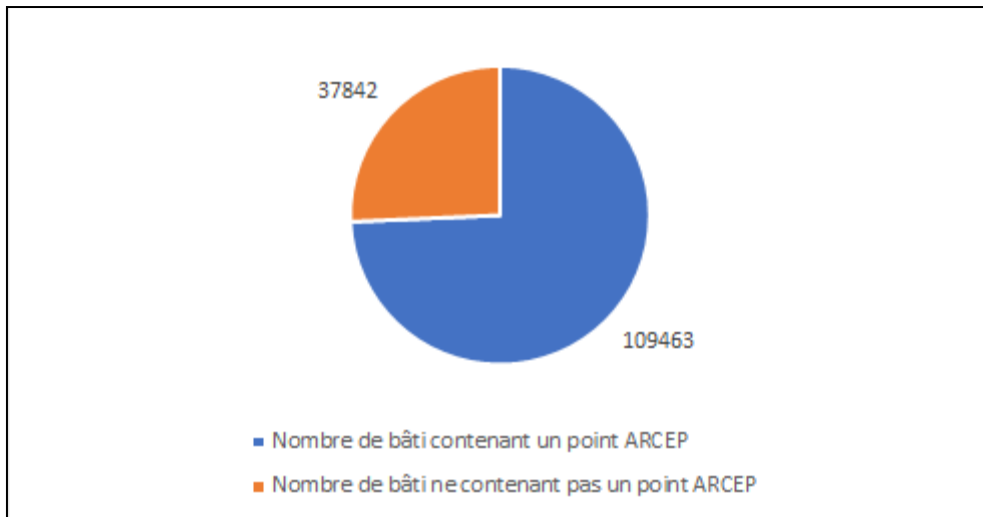


Figure 15 : Nombre de bâtiments contenant un point ARCEP : Jointure attributaire par le plus proche (1 voisin /10m)

Nous avons donc avec nos bâtis ne contenant pas de point ARCEP à l'étape précédente une nouvelle fois augmenté le nombre maximum de voisins les plus proches à 10 avec une distance maximale de 30 mètres.

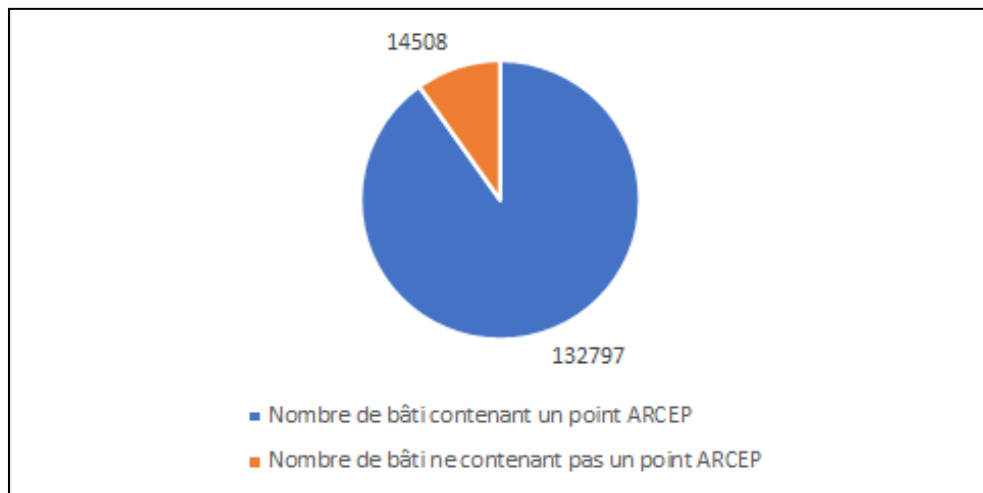


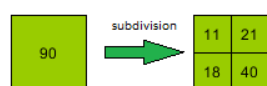
Figure 16 : Nombre de bâtiments contenant un point ARCEP : Jointure attributaire par le plus proche (10 voisins / 30m)

2. Les données carroyées

Le découpage territorial des données carroyées de l'INSEE quadrille le territoire français par des mailles dites "naturelles", de 200 m de côté ou de 1 km de côté. Cette base comprend 18 variables⁶ sur la structure par âge des individus, les caractéristiques des ménages (locataire/propriétaire, etc.) et de leurs logements et les revenus au 31 décembre 2010. Ces données sont récoltées par la Direction Générale des Finances Publiques dans le cadre du dispositif sur les revenus localisés sociaux et fiscaux (FiLoSoFi), ce qui permet, en définitive, une observation des caractéristiques socio-démographiques et des caractéristiques des logements à un niveau territorial infra-communal.

Carroyage naturel :

La grille de niveau naturel correspond à un partitionnement du territoire en carreaux de différentes tailles (200 m, 1 km, 2 km, 4 km, 8 km, 16 km ou 32 km) permettant de diffuser toutes les informations sans imputation des données, tout en respectant le secret fiscal. Le niveau naturel forme un pavage complet du territoire habité, c'est-à-dire que chaque ménage appartient à un et un seul carreau du niveau naturel. Cette grille a été construite de telle sorte qu'aucun de ces carreaux ne comporte moins de 11 ménages fiscaux .



Concernant la méthode de production du carroyage naturel, il part d'une grille régulière formée de carreaux de 32 km de côté. il est ensuite subdivisé :

- si pour chacun des quatre carreaux suivants, le nombre de ménages est toujours supérieur ou égal à 11, c'est validé
- dans le cas contraire (entre 1 et 10 ménages, 0 ménages ne comptant pas), il n'est pas subdivisé.

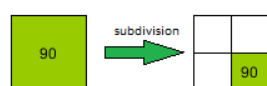


Figure 17 : Explication des carreaux

Les données du niveau naturel sont celles issues de l'agrégation des carreaux de 200m. Or les valeurs des variables ont été arrondies à une décimale pour le niveau de 200m (dans le cas des carreaux imputés). Cela a pour conséquence que pour certains carreaux du niveau naturel, les valeurs diffusées diffèrent légèrement des valeurs qu'on aurait obtenues directement ou sans arrondie.

Finalement, de par l'irrégularité des mailles et les éventuelles différences de valeurs liées aux arrondies, ce carroyage ne fut pas retenu pour nos tests préliminaires.

⁶ Cf annexe 1 : Tableau récapitulatif des champs des carreaux 200 m et 1 km de l'INSEE

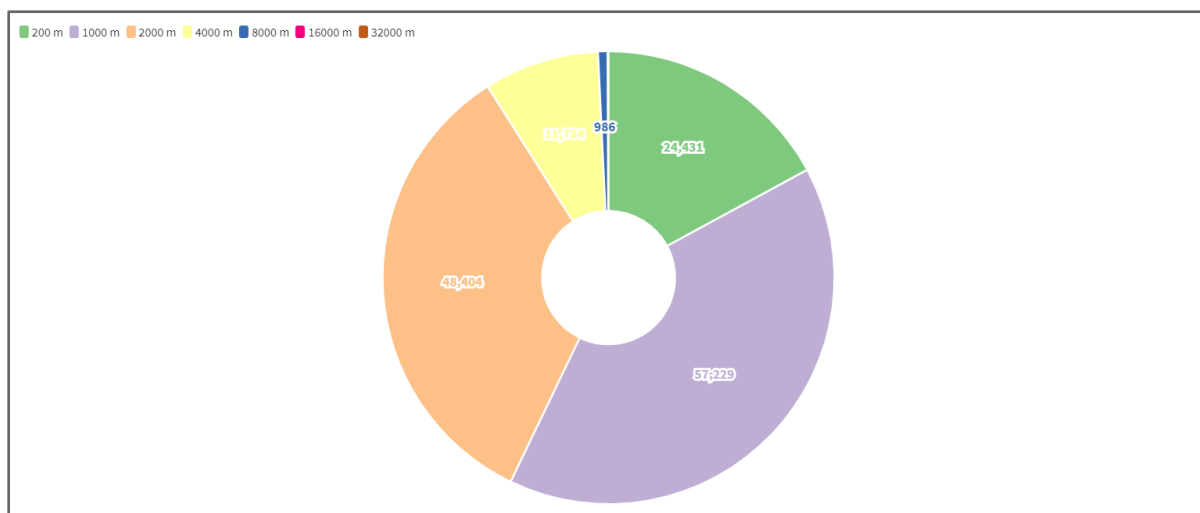


Figure 18 : Taille des carreaux de France métropolitaine en mètre

Carroyage 1 km :

Concernant le niveau à 1 km, c'est les valeurs du carroyage de 200 m qui sont utilisées et agrégées selon la variable `Id_carr1km`, en faisant pour chaque variable de diffusion la somme des valeurs présentes au niveau 200 m (que ces valeurs soient issues d'une imputation ou non). Finalement, 64% des carreaux de 1 km font l'objet d'une imputation. Ces derniers représentent 9% de la population. Le carroyage de 1 km sera utilisé pour les tests préliminaires.

Carroyage 200 m :

La plupart des carreaux (80%) contiennent moins de 11 ménages fiscaux. Pour ces carreaux, il n'est pas possible de diffuser directement l'information des ménages qui y résident. Pour pouvoir le faire, il faut regrouper ces carreaux entre eux jusqu'à obtenir des groupes de carreaux rassemblant au total au moins 11 ménages fiscaux. La variable `Groupe` indique ainsi à quel regroupement appartient chaque carreau de 200 m. Chacun de ces groupes rassemble au moins 11 ménages fiscaux.

Les carreaux de 11 ménages ou plus sont généralement tout seuls dans leur groupe mais il arrive, pour certains d'entre eux, qu'ils soient rassemblés dans des groupes avec d'autres carreaux de peu d'effectifs : cela est dû au fait qu'on prend en compte la proximité des carreaux pour réaliser les groupes. Les dits groupes ne sont pas formés directement à partir des carreaux de 200 m. Dans la pratique, il faut partir du niveau naturel et former des groupes à chaque niveau, en allant vers des niveaux de plus en plus fins.

Si le niveau naturel est arrivé au niveau d'un carreau de 200 m, alors ce carreau contient bien 11 ménages ou plus et un groupe est créé seulement pour lui (aucun autre carreau n'appartiendra à son groupe). Au contraire, si le niveau naturel s'est arrêté à un niveau supérieur à 200 m, alors le carreau en question est divisible en plusieurs sous-carreaux. Si au moins un de ces sous-carreaux contient moins de 11 ménages par construction. Il faut rassembler tous les sous-carreaux contenant moins de 11 ménages dans un groupe. Si cet ensemble de carreaux rassemble 11 ménages ou plus, alors il est possible de s'arrêter à cette étape. En revanche, si cet ensemble contient moins de 11 ménages, il faut ajouter un des sous-carreaux restant, contenant au moins 11 ménages, de façon que le groupe ainsi formé

contienne au moins 11 ménages. Il est choisi, dans ce cas, parmi les sous carreaux contenant au moins 11 ménages, celui qui compte le moins de ménages.

La progression est itérative, en passant des niveaux les plus agrégés au niveau le plus fin (carreaux de 200 m). Un groupe peut lui-même être subdivisé en plusieurs sous-groupes lorsqu'on passe à un niveau plus fin.

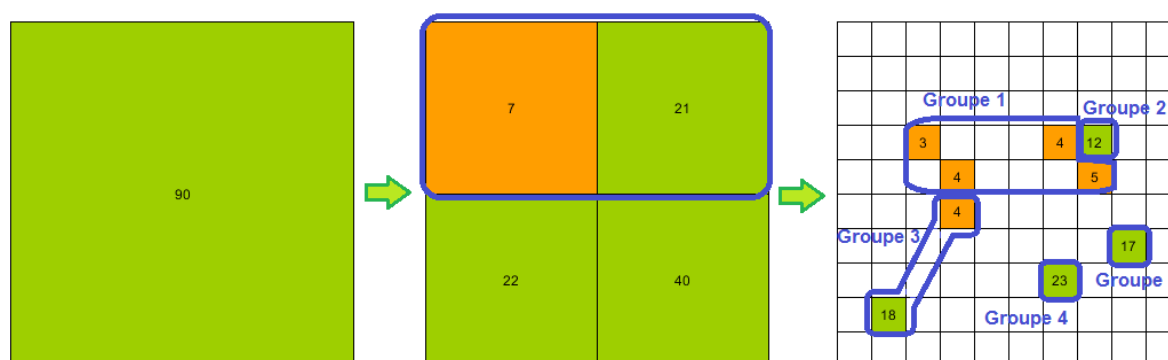


Figure 19 : Désagrégation des carreaux

Exemple : le carreau de 2 km comprend 90 ménages. Lors de la subdivision en 4 carreaux de 1 km, on voit qu'un carreau est sous le seuil de 11 ménages. Il va alors être regroupé provisoirement avec le carreau le moins peuplé des 3 carreaux restants. À l'étape suivante, l'ensemble des sous-carreaux de 200 m du carreau de 7 restera dans le même groupe, mais ils seront regroupés avec seulement 2 autres carreaux (de 4 et 5 ménages) du carreau de 1 km précédent. Le sous-carreau de 12 ménages de ce carreau de 1 km restera indépendant.

Si le carreau qu'on cherche à subdiviser est déjà affecté dans un groupe, alors tous les sous carreaux contenant moins de 11 ménages de ce carreau seront affectés au même groupe que celui du carreau du départ. C'est pour cette raison que certains groupes peuvent contenir plus de 11 ménages.

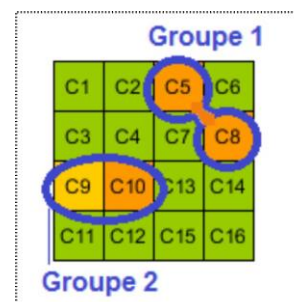
Concernant l'imputation des données, une fois les groupes déterminés, il est imputé à chaque carreau, et pour toutes les variables de diffusion, la valeur moyenne du groupe pondérée par le nombre d'individus. Considérons ainsi une variable Y de diffusion, autre que le nombre de personnes. Connaissant la valeur de cette variable Y sur chaque groupe et connaissant le nombre de personnes par carreau, on peut imputer la valeur de Y sur les carreaux en répartissant la valeur de Y du groupe proportionnellement au nombre de personnes de chaque carreau. (Le nombre de personnes par carreau est la seule variable diffusée sans restriction, quel que soit leur nombre de ménages).

Par exemple, pour le groupe 1, si :

- le carreau C5 comporte 7 ménages, 17 personnes et $Y = Y_a$ sur ce carreau ;
- le carreau C8 comporte 4 ménages, 8 personnes et $Y = Y_b$ sur ce carreau ;

On estime que la variable Y vaut :

- $(Y_a + Y_b) * (17 / (17 + 8))$ pour C5 ;
- $(Y_a + Y_b) * (8 / (17 + 8))$ pour C8.



Ainsi, la somme des estimations est bien égale à $Y_a + Y_b = Y_g$ la valeur de Y pour le groupe. Cette valeur Y_g n'est pas confidentielle car elle porte sur au moins 11 ménages fiscaux. Finalement, de par sa granularité et les méthodes statistiques employées pour la diffusion de l'information, le carroyage de 200 m sera utilisé pour les tests préliminaires.

Bien que les données commencent à être obsolètes (2010), nous avons essayé de les exploiter afin de profiter de leur granularité fine. Trois variables nous intéressaient principalement dans cette base de données: le nombre d'individus par carreau, le nombre de ménages par carreaux et le nombre de logements par carreau.

Pour vérifier si ces données étaient encore utilisables, nous les avons comparés aux données disponibles à l'IRIS de trois bases de l'INSEE différentes, considérés dans ces travaux comme notre référence :

- Population 2017.
- Logement 2017.
- Couples Familles Ménages 2017.

Méthode n°1 :

La première méthode de comparaison se base sur l'agrégation des carreaux à l'échelle de l'IRIS pour effectuer la comparaison. L'utilisation des centroïdes de ces carreaux induit néanmoins des imprécisions. Les écarts ont été transposés en pourcentage, entre les valeurs à l'IRIS (des bases population, logement et couples familles ménages) et les valeurs des carreaux agrégés à l'IRIS.

Pour les variables de population et des ménages, des écarts trop importants ont été mesurés entre les carreaux et les Iris. En effet, les écarts mesurés allaient au-delà de 20% (en valeur absolue) pour certaines communes. Voici nos résultats pour les communes pilotes:

POPULATION				
NOM	IRIS	CARROYAGE	IRPO-CAPO	IRPO-CAPO%
Nouvoitou	3099	3143	-44	-1
Saint-Grégoire	9638	11941	-2303	-24
Centre	20411	18849	1562	8

Figure 20 : Population présente dans les données INSEE et dans les carroyages par IRIS

MENAGES				
NOM	IRIS	CARROYAGE	IRME-CAME	IRME-CAME%
Nouvoitou	1223	1176	47	4
Saint-Grégoire	4133	4934	-801	-19
Centre	13154	10826	2328	18

Figure 21 : Nombre de ménages présents dans les données INSEE et dans le carroyage par IRIS

Pour ce qui est des logements, les écarts étaient un peu plus faibles, autour de 10% en valeur médiane, ce qui constituait notre seuil minimal. Cependant, après analyse cartographique des résultats, nous avons remarqués que les écarts importants étaient concentrés sur les Iris de Rennes et aux alentours, comme nous pouvons le voir sur la carte ci-dessous, qui nous présente les écarts mesurés entre les valeurs à l'iris et les valeurs des carreaux :

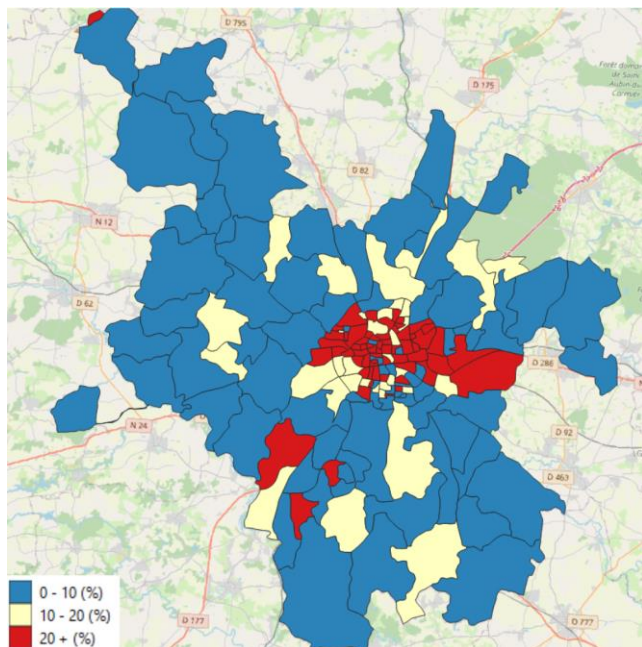


Figure 22 : Différence entre le nombre de logements des données carroyées et des données INSEE

Méthode n°2 :

La seconde méthode de comparaison se veut un peu plus avancée. Dans la mesure du possible, les tests ont été effectués sur de grandes échelles territoriales, afin d'anticiper au mieux les éventuels cas particuliers. Ici deux options ont été envisagées en fonction de la taille du carroyage et de l'échelle de la couche INSEE de comparaison :

- dans une comparaison entre le carroyage 200 m et les informations de population, de logement et de ménages à l'échelle des communes, la construction des données carroyées nous permet d'agréger les informations selon le champ Depcom. Dans les faits, chaque carreau est associé à une commune de référence dans laquelle la majorité des ménages du carreau réside.
- dans une comparaison entre le carroyage 200 m et les informations de l'INSEE à l'échelle de l'IRIS, ou dans une comparaison entre le carroyage 1 km et les informations de l'INSEE à l'échelle de l'IRIS ou de la commune, il est nécessaire d'introduire de la rigueur statistique. Pour cela, on passe par une couche transitoire, la couche des bâtiments. En attribuant une fraction égale des informations qu'on cherche

à comparer dans chacun des bâtiments du carreau, il est ensuite possible d'agréger ces informations à l'échelle souhaitée, commune ou IRIS.

Concernant les logements :

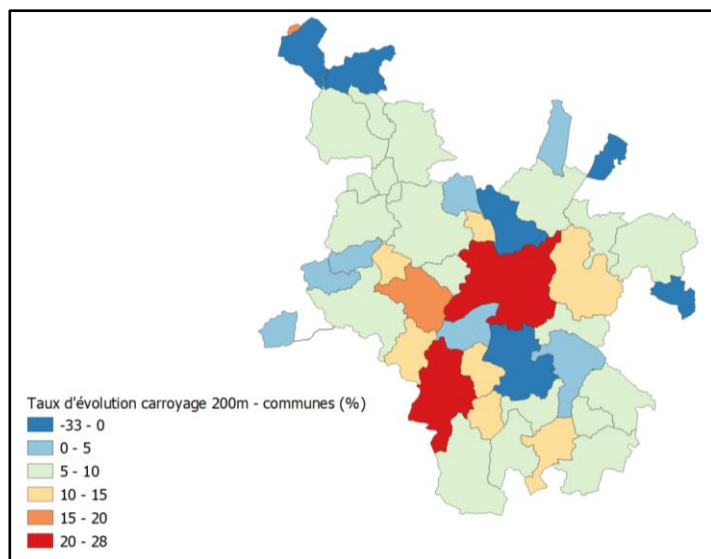


Figure 23 : Différence entre le nombre de logements des données carroyées et des données INSEE pour les communes

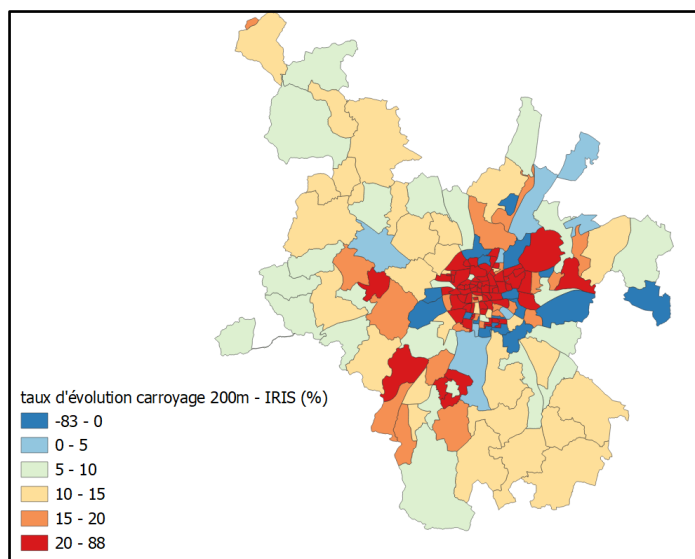


Figure 24 : Différence entre le nombre de logements des données carroyées et des données INSEE pour les IRIS

Les deux cartes ci-dessus permettent de faire ressortir la différence de logements qu'il y a entre le carroyage 200 m et la base de données logement de l'INSEE. Les pourcentages sont le résultat du taux d'évolution suivant : $((\text{données INSEE} - \text{données carroyées}) / \text{données de référence}) * 100$. Que ce soit pour les iris ou bien les communes les forts écarts sont principalement concentrés à Rennes. Ces écarts sur cet espace peuvent être expliqués notamment par la forte population étudiante y résidant. Cependant, l'IRIS comparé aux communes présente des écarts beaucoup plus importants. Cela nous amène alors comme pour les ménages à nous questionner sur la possibilité d'exploiter ces données sur des espaces urbains. L'exploitation de ces données serait plus adaptée sur un milieu rural.

Concernant la population :

Différence de population entre le recensement de la population de l'INSEE et le Carroyage 200 m :

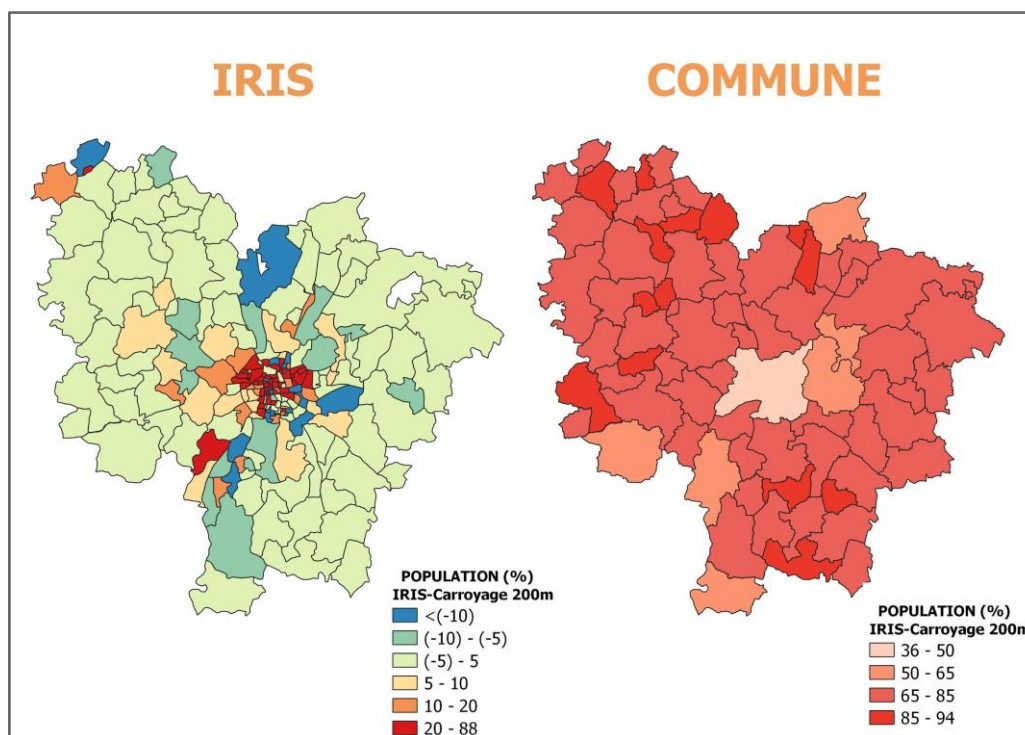


Figure 25 : Différence entre le nombre de logements des données carroyées et des données INSEE pour les communes et IRIS

On observe dans les tons vert et orange clair des différences entre $\pm 10\%$ entre le recensement de la population de l'INSEE et les données démographiques des fichiers fonciers "MAJIC", agrégées dans le carroyage de 200 m. Sur la carte de gauche, les données du recensement de la population de l'INSEE sont agrégées au niveau des IRIS et à droite au niveau des Communes. Dans le reste des tons on voit des différences avec la population de référence (recensement de la population de l'INSEE de 2017) supérieures en valeur absolue à 10%, atteignant dans certains cas plus de 80% d'écart.

Quand on compare le carroyage de 200 m avec les données agrégées au niveau de la commune, on constate que ces différences sont fortement accentuées et généralisées. Concernant les logements pour le carroyage 1 km et les informations INSEE à l'échelle des communes : Comme nous l'avons vu dans la comparaison avec le carroyage de 200 m, plus la surface d'agrégation des données démographiques de l'Insee est grande, plus les différences sont importantes par rapport au carroyage. Ces différences sont encore plus accentuées lorsque nous travaillons avec le carroyage de 1 km. Pour représenter ces différences, nous avons regroupé notre surface d'étude en une seule surface et l'avons comparée aux données carroyées. Dans la carte suivante, nous pouvons voir un résumé général, sur une plus grande surface, des différences entre l'objet de données démographiques de notre étude.

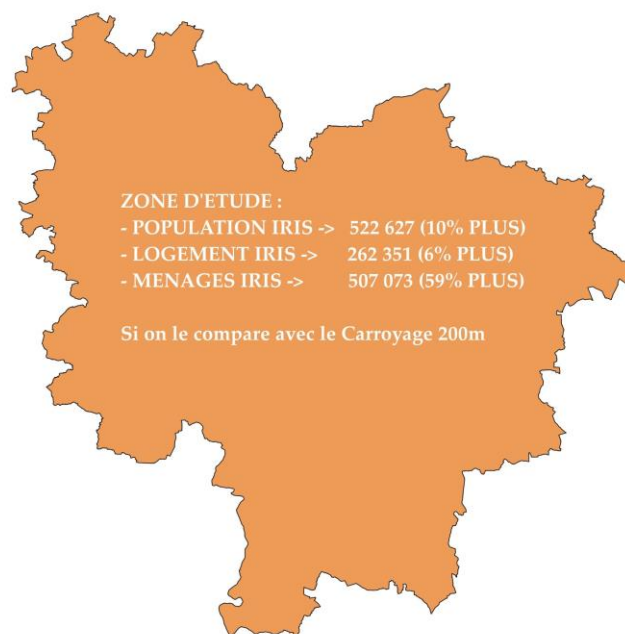


Figure 26 : Résumé de nos résultats

Il en ressort, en définitive, que le carroyage de 1 km induit trop d'écarts et d'imprécisions lorsque comparé avec nos bases de données de référence. Le carroyage de 200 m, de par sa granularité, aurait pu être central pour nos méthodes de ventilation, mais les écarts constatés, en particulier dans les zones urbaines telles que Rennes ou Saint-Malo nous forcent à devoir mettre de côté les données carroyées. Pour autant, il semblerait pertinent d'effectuer des tests plus approfondis dans des prochains travaux sur cette thématique de valorisation du bâti, en déployant une meilleure répartition des informations des carreaux dans les bâtiments, en ne prenant en compte que les bâtiments résidentiels et en effectuant une proto-ventilation, par exemple.

Dans cette seconde partie, nous avons partagé l'ensemble des données et méthodes que nous avons explorées, analysées et comparées. Certaines ont été mises de côté, on pense notamment aux données carroyées qui commencent à dater. D'autres, en revanche, se sont révélées à haut potentiel comme le fichier détail. Cette phase exploratoire désormais terminée, nous allons passer au cœur de notre sujet: la ventilation de la population à Rennes Métropole.

III- Ventilation Rennes Métropole

a. Méthodes tests

Nous avons effectué cette méthode de ventilation afin de comparer celle-ci avec la méthode de ventilation avancée que nous aborderons plus tard dans le rapport.

1. Données

Dans cette méthode de ventilation que l'on peut surnommer "classique". Nous avons pris en compte plusieurs jeux de données différents.

- La donnée des bâtiments que l'agence d'urbanisme Audiar nous renseigne contient plein d'informations relatives aux bâtiments.
- Le recensement intracommunal qui vient de l'INSEE
- Les contours des IRIS qui proviennent de l'IGN. Ce sont en effet les découpages administratifs que nous utiliserons.

2. Ventilation "classique"

Lors de cette ventilation, nous avons dans un premier temps calculé la surface par bâtiments et nous l'avons ajouté dans une nouvelle colonne de la table attributaire. Puis nous avons calculé la hauteur du bâtiment à l'aide de la colonne "h_faitage" dans la base bâti Audiar. Pour calculer le nombre d'étages, nous avons choisi de diviser par 3 la hauteur calculée de nos bâtiments. Nous avons déjà une colonne nombre d'étages dans la base Audiar, cependant la hauteur calculée était plus exhaustive et reflétée plus la réalité. Une fois la surface et le nombre d'étages des bâtiments calculés, nous pouvons alors déterminer à l'aide d'une formule, la surface plancher :

Avec cette surface de planchers que nous avons pour chaque bâtiment, cela nous servira à ventiler la population dans chacun des bâtiments plus tard. Ensuite, nous avons extrait les centroïdes des bâtiments de la base de données bâti de l'Audiar. Nous avons réalisé ce géotraitement afin de pouvoir ensuite faire une jointure avec la couche IRIS de l'IGN pour ajouter la population de chaque IRIS dans chaque bâtiment. Nous avons eu besoin de faire des centroïdes de bâtiments afin qu'il n'y ait pas de superposition entre les contours des IRIS et les bâtiments. Pour répartir cette information dans les bâtiments, nous avons fait une jointure avec le code IRIS qui était présent dans la table attributaire des bâtiments. Enfin, la dernière partie de notre ventilation classique est évidemment la formule de ventilation que l'on peut appliquer à l'aide de la calculatrice de champs dans notre base bâtiments de l'Audiar. Le rendu final de cette ventilation sera alors le nombre de personnes que l'on pourra avoir à peu près par bâtiment mais nous verrons que dans les prochaines méthodes, les résultats seront plus précis.

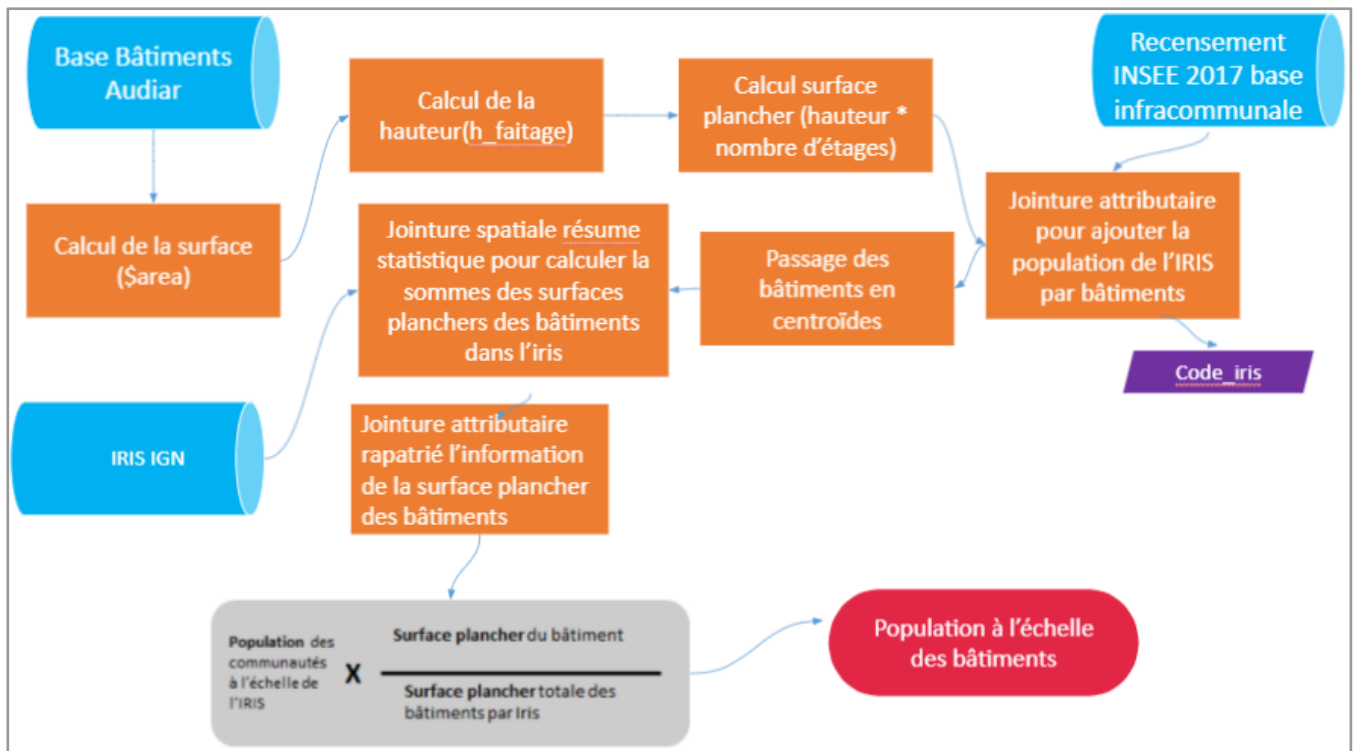


Figure 27 : Chaîne de traitement de la méthode "classique"

b. Méthode finale

1. Données

C'est cette méthode de ventilation sur laquelle nous nous sommes penchés afin d'avoir l'information de la population la plus précise possible. Pour cela, nous avons utilisé plusieurs jeux de données différents :

- La donnée des bâtiments interne à l'agence Audiar qui nous donne des informations relatives à ces mêmes bâtiments.
- Le contour IRIS qui provient de l'IGN. C'est le découpage administratif que nous allons utiliser.
- Les fichiers détails logement de l'INSEE du recensement 2017 qui nous donne l'information de la population à l'échelle du logement.

Ces jeux de données ci-dessus, nous ne les avons pas utilisés à l'état brut, on les a retravaillés pour établir notre méthode de ventilation.

2. Préparation des données

Donnée population (sous Excel)

Comme dit précédemment, ce jeu de donnée de l'INSEE nous donne le nombre d'habitants à l'échelle du logement. De plus, il nous donne des informations supplémentaires comme l'IRIS du logement, le type de logement (appartement/maison, parc privé/social). Ce que nous avons fait ensuite, c'est une agrégation statistique de cette donnée à l'échelle de

l'IRIS grâce notamment au champ "Code_IRIS". Cela nous donne l'information de la population à l'échelle des IRIS.

Ensuite, cette information de la population, il est possible de la reclassifier en 4 classes bien distinctes :

- Les populations vivantes en appartements privés
- Les populations vivantes en appartements sociaux (HLM)
- Les populations vivantes en maisons privées
- Les populations vivantes en maisons sociales (HLM)

Après cette étape, on s'est rendu compte que la somme des populations de ce jeu de donnée sur la métropole de Rennes était de 413 040 habitants. Or, on sait que selon les chiffres du recensement (base infracommunale) 2017 de l'INSEE, le nombre d'habitants à Rennes était de 447 429 en 2017, soit une différence de 34 389. Cette différence s'explique par le fait que les populations présentes dans les fichiers de détails logements sont les populations qui ne sont pas recensées dans les logements listés ci-dessus (appartements/maisons). Typiquement, ces populations peuvent être caractérisées en quatre catégories :

- Les étudiants.
- Les résidents des maisons de retraites et EHPAD.
- Les personnes résidentes en établissements pénitenciers.
- Les personnes résidentes en foyer du jeune travailleur (FJT)

Ces populations sont donc aux nombres de 34 389 sur le territoire de Rennes Métropole (soit la différence juste au-dessus). Pour simplifier, nous avons décidé d'appeler ces populations les "communautés". Ce sont donc des habitants supplémentaires à répartir dans nos bâtiments rennais. Finalement, nous avons réalisé une chaîne de traitement pour résumer cette phase de préparation de la donnée de population.

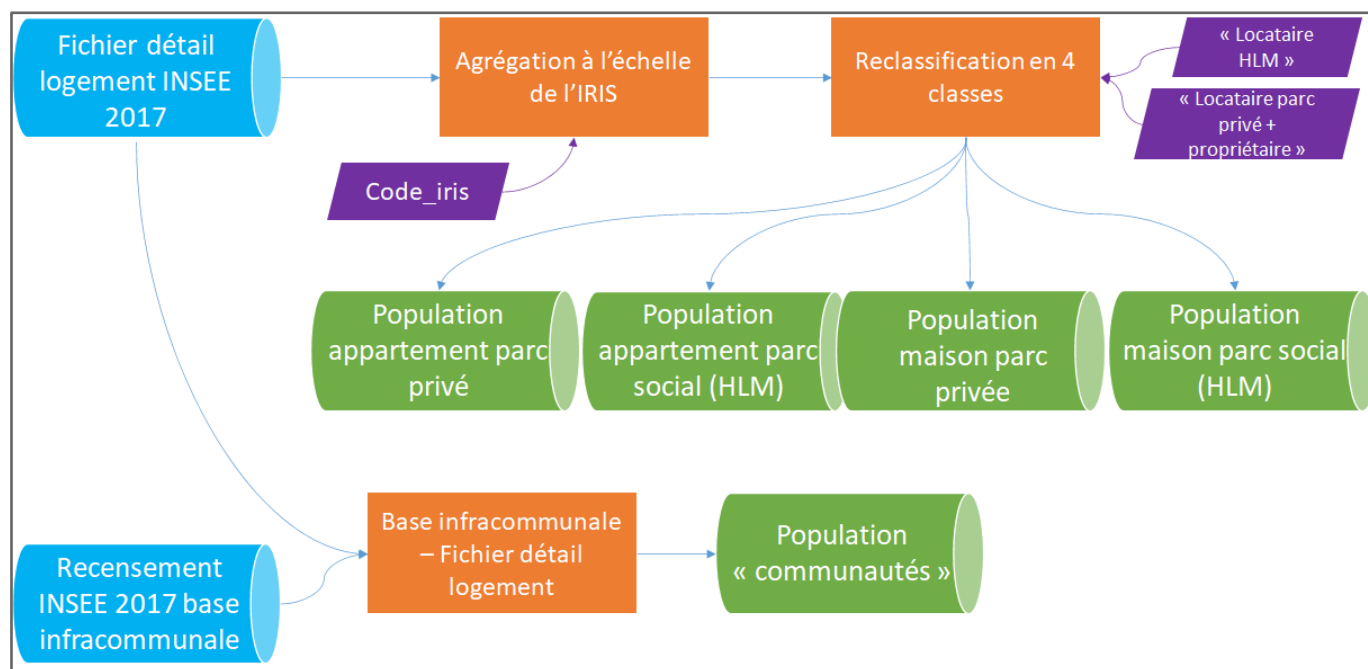


Figure 28 : Chaîne de traitement de la préparation des données sous Excel

Donnée bâtiments (sous Qgis)

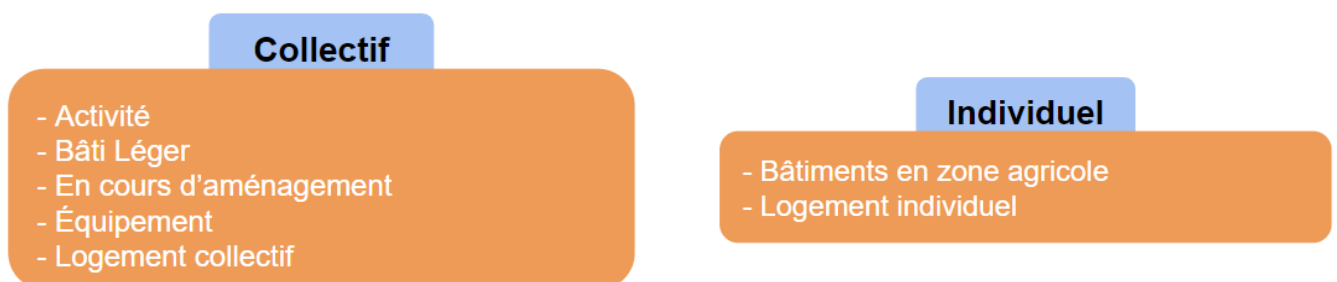
Nous avons également retravaillé cette base bâtiment de l'Audiar en utilisant les informations diverses qui sont à dispositions dans cette base de données. Nous avons ainsi mobilisé le champ "NB_logement" qui nous renseigne sur le nombre de logements au sein du bâtiment. Nous avons donc décidé de retirer tous les bâtiments qui ne comptaient aucun logements en leur sein.

Ensuite, nous devons reclassifier ces bâtiments en 5 classes :

- Les appartements privés
- Les appartements sociaux (HLM)
- Les maisons privées
- Les maisons sociales (HLM)
- Les bâtiments "communautaires"

En ce qui concerne les bâtiments "communautaires", nous avons pu reclassifier ces bâtiments en faisant directement des sélections sur la couche de bâtiments de l'audiar avec l'aide d'autres sources de données comme OSM, google maps ou la Base Permanente des Équipements. En outre, nous n'avons pas pu automatiser cette sélection des bâtiments communautaires avec la BPE car la géolocalisation des entités dans la BPE était imprécises. Nous avons opté ainsi pour une méthode de sélection "à la volée" mais qui ce veut être tout de même fiable.

Puis, pour les autres types de bâtiments, nous avons utilisé directement les informations de la base bâtiments de l'Audiar pour reclasser ces derniers. C'est le cas de la colonne "Usage_bati" qui nous donne une information sur l'usage du bâtiment que nous avons reclassifier pour savoir si c'est un logement collectif (appartement) ou un logement individuel (maison). Les modalités de "Usage_bati" ont donc été reclassifier de la sorte :



Pour les bâtiments individuels, nous avons vu que les bâtiments agricoles étaient des bâtiments présents sur les exploitations agricoles et pouvaient parfois accueillir de la population, il nous a donc semblé logique de placer cela en tant que logement individuel. Pour les logements collectifs, les modalités que l'on a reclassifiées correspondent à des bâtiments situés dans la ville de Rennes. On a vu que c'était très souvent des bâtiments de types appartements.

Ensuite, il a fallu encore une fois reclassifier en fonction du type de parc (privé / social). Pour cela, nous avons utilisé le champ "ps_PRG_LIB" qui nous renseigne du nom du logement lorsque celui-ci est du parc social. Nous pouvons ainsi différencier les logements privés ("ps_PRG_LIB" non rempli) des logements HLM / social ("ps_PRG_LIB" rempli). Nous avons

également réalisé une chaîne de traitement pour résumer cette phase de préparation de la donnée des bâtiments.

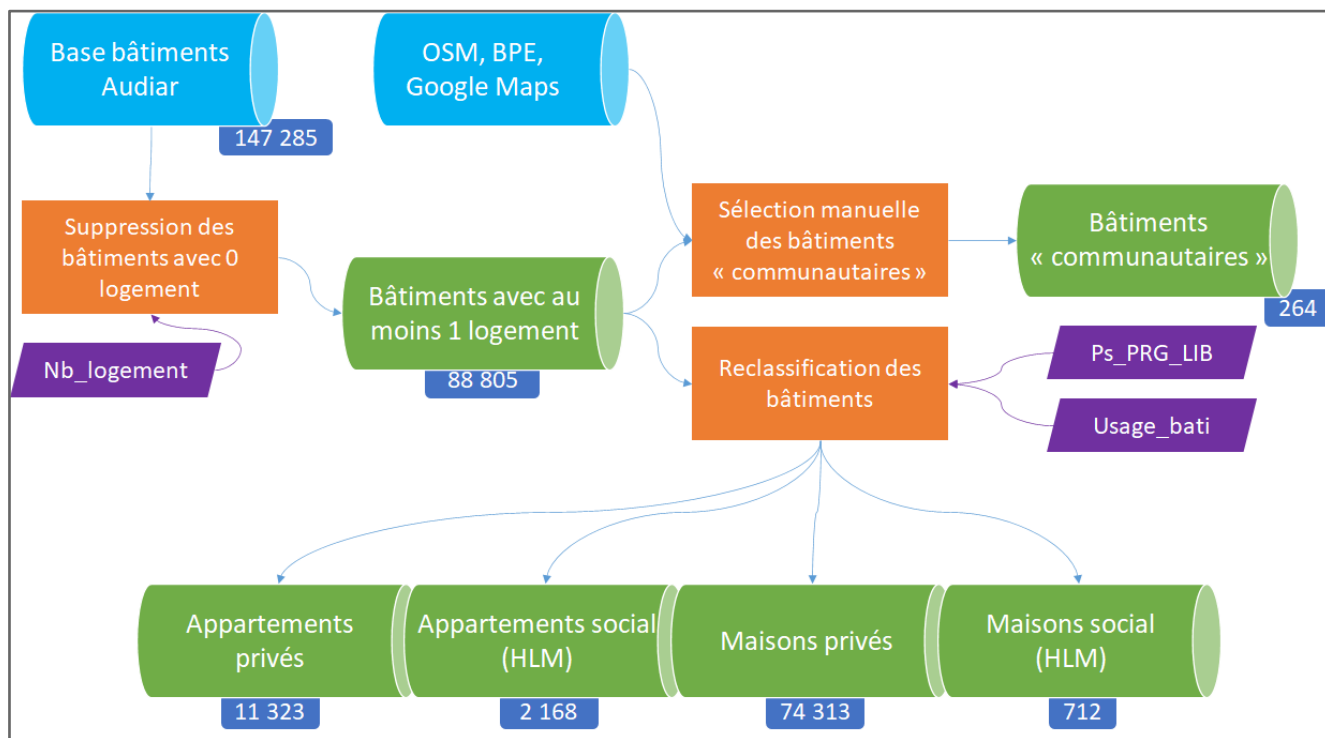


Figure 29 : Chaîne de traitement pour la préparation de la donnée des bâtiments sous Qgis

3. Ventilation avancée (sous Qgis)

Nous avons donc notre jeu de donnée de population réagréé et celui de bâtiment reclassifier. Nous allons pouvoir répartir la population dans ces mêmes bâtiments. Nous allons faire ça en 2 étapes en commençant par les bâtiments des logements (appartements / maisons).

Étape 1 :

Pour commencer, on utilisera la colonne “NB_logement” pour répartir notre population dans ces bâtiments. En effet, cette fois-ci et contrairement à notre méthode test, nous avons utilisé une information qui renseigne directement sur la présence de population dans un bâtiment pour répartir notre population. Cette méthode est beaucoup plus logique à mobiliser au lieu d'utiliser les caractéristiques physiques du bâti (surface plancher).

Dans un premier temps, nous allons prendre l'exemple des appartements privés dans l'optique de la ventilation. Cette façon de faire est la même que pour les appartements HLM, maisons publiques / privés. Nous prenons juste cet exemple dans un souci de clarté et de compréhension. Donc, pour les appartements privés, nous allons calculer la somme des logements de ces appartements pour chaque IRIS de Rennes Métropole (jointure par localisation avec résumé statistique). Il faut au préalable passer nos bâtiments en centroïdes afin que la jointure puisse se faire correctement. Ensuite, cette information du nombre de logements totale des appartements privés, nous l'avons rapatrié dans nos logements avec

une jointure attributaire en utilisant le code IRIS comme élément de jointure. Puis, nous faisons de même avec l'information de la population des "Fichier détail logement de l'INSEE", que nous allons rapatrier dans nos bâtiments avec une jointure attributaire avec le code IRIS. Enfin, nous pouvons déterminer la population contenue dans nos bâtiments en utilisant la calculatrice de champs de Qgis et en utilisant cette formule :

$$\text{Population des appartements privés de l'IRIS} \times \frac{\text{Nombre de logements du bâtiment}}{\text{Nombre de logements total des appartements privés de l'IRIS}}$$

En quelque sorte, nous avons multiplié la population totale des appartements privés de l'IRIS par le ratio de ce que représente le bâtiment (en termes de logement) sur tous les autres appartements privés de l'IRIS. Cette manière de faire, nous l'avons fait pour les appartements HLM, les maisons privées et les maisons HLM. Cela nous donne l'information de la population dans ces bâtiments, maintenant il nous reste les bâtiments communautaires.

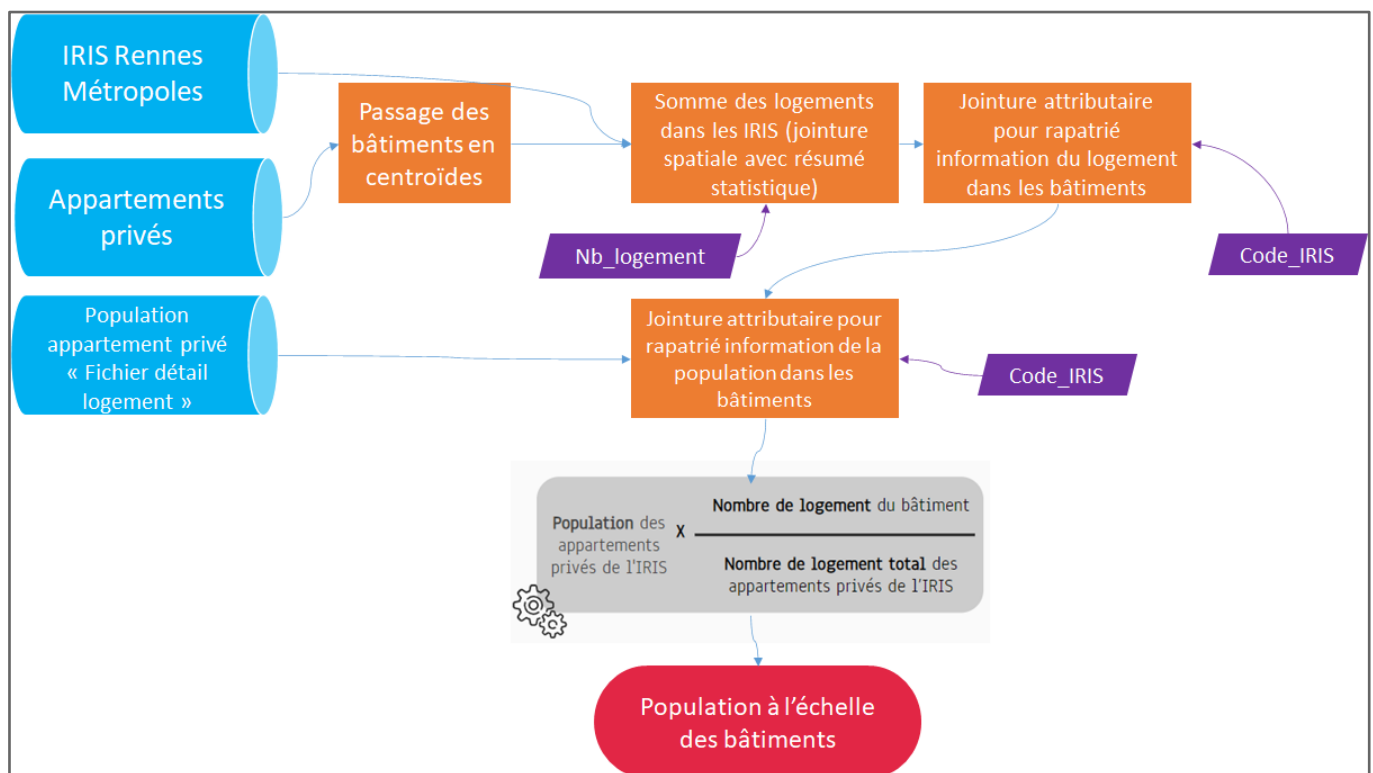


Figure 30 : Chaîne de traitement de la méthode de ventilation des logements privés pour Rennes Métropole

Étape 2

Pour cela, on va utiliser cette fois-ci la surface plancher des bâtiments pour répartir la population. En effet, nous n'avons pas pu utiliser le champ "Nb_logement" car on s'est rendu compte que pour les bâtiments "communautaires" exclusivement, cette information n'était pas fiable. Par exemple, on pouvait voir des résidences étudiantes avec seulement 3 logements, alors que pour certaines d'entre elles, il y avait plus de 100 logements. Nous avons donc été contraint de choisir une caractéristique physique du bâti, car nous n'avons pas d'autres choix pour ventiler la population.

Ainsi, nous avons dans un premier temps calculé la surface plancher de, tous ces bâtiments en utilisant la surface du bâtiment que nous avons recalculé afin d'être sur à 100 % (il pouvait y avoir un léger décalage entre ce qui était indiqué et les valeurs de Qgis avec le \$area). De plus, nous avons déterminé le nombre d'étages en utilisant la hauteur du bâtiment contenue dans le champ "h_faitage", et nous l'avons divisé par trois (taille moyenne d'un étage en France) pour le nombre de niveaux du bâti. Encore une fois, il y avait un champ "Nb_niveau" déjà renseigné dans la base bâtiment, mais on a remarqué quelques incohérences (certains bâtiments avait moins d'étage qu'en réalité que sur la base et vice-versa) et nous avons préféré recalculé ça. Enfin, la surface plancher est obtenue en multipliant la surface par le nombre d'étages.

Ensuite, nous allons déterminer la population "communautaire" de Rennes Métropole. Pour cela, il faut faire la différence entre la population de la base infracommunale à l'échelle de Rennes Métropole et celle du Fichier détail logement à la métropole également. Ce calcul nous donne la population "communautaire" pour les IRIS de Rennes Métropole. Cette information de la population "communautaire", nous l'avons rapatrié dans nos logements avec une jointure attributaire avec le code IRIS en élément de jointure. Puis, nous allons calculer la somme des surfaces planchers de tous les bâtiments "communautaires" de la métropole avec une jointure par localisation avec résumé statistique. Encore une fois, il faut rapatrier cette information dans les bâtiments. Enfin, nous avons toutes les informations pour calculer la population dans ces bâtiments en utilisant la calculatrice de champ de Qgis :

$$\text{---} \times \frac{\text{Surface plancher du bâtiment}}{\text{Surface plancher totale de tous les bâtiments "communutaires" de RM}}$$

Là encore, nous avons utilisé un ratio (surface plancher / surface plancher totale) pour ventiler notre population. Enfin, nous n'avons plus qu'à fusionner tous les bâtiments ensembles pour avoir notre couche de bâtiments avec l'information de population à l'échelle de Rennes Métropole.

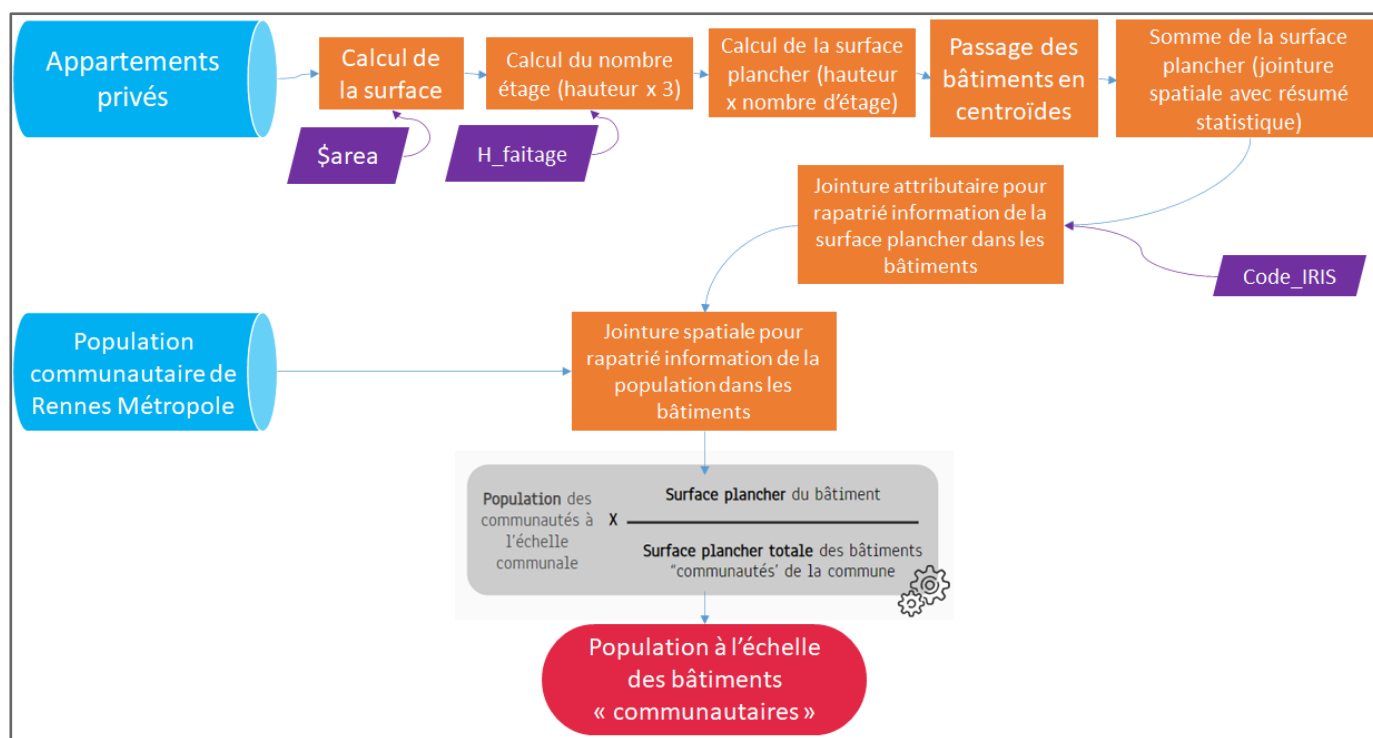


Figure 31 : Chaîne de traitement de de la méthode de ventilation des logements "communautaires" pour Rennes Métropole

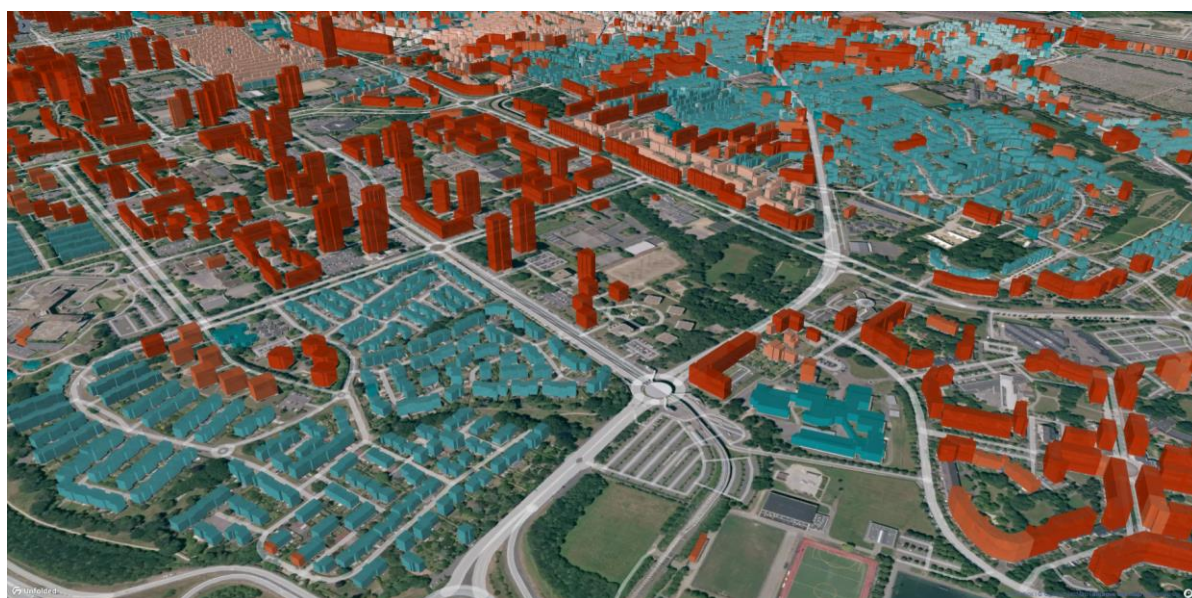


Figure 32 : Représentation de l'information de la population à l'échelle des bâtiments du quartier de la Poterie

Nous avons fait également quelques rendus cartographiques pour rendre compte de notre travail. Ici, plus les bâtiments du quartier de la poterie sont peuplés, plus ils tendent vers le rouge. On peut nettement voir la différence entre maison et immeuble.

c. Perspectives

Les deux méthodes proposées font office de méthodes perfectibles qui posent les bases pour des travaux futurs. L'ensemble des données explorées nous a permis de constituer un socle de connaissances pour des problématiques de ventilation des populations. La première méthode, plus simple et plus rapide, permet de ventiler la population avec peu de données en entrée, mais ne tient pas compte des caractéristiques des bâtiments. La seconde est celle qui nous paraît la plus proche de la réalité, mais nécessite beaucoup de données et de connaissances techniques.

Concernant les outils utilisés, nous avons pu voir toute l'importance des SIG pour mettre en place et tester nos méthodes de ventilation, cependant, la taille des fichiers de données aura exposé les limites de l'outil. Afin de gérer cette quantité importante d'information, des bases de données géographiques (PostGre avec l'extension PostGIS par exemple) que l'on connecte ensuite à QGIS, pourraient être utilisées pour fluidifier le travail. Cela permettrait de gagner un temps énorme en traitement, mais aussi d'améliorer la stabilité globale de notre méthode.

Cette ventilation ouvre des portes à de multiples usages parmi lesquels on retrouve l'étude de la consommation énergétique, la gestion des risques pour les cabinets d'assurance, le géomarketing ou encore des études en aménagement. À l'avenir, on pourrait espérer que l'Insee établisse un recensement à l'échelle du bâtiment pour les logements collectifs, bien que cela semble peu probable au regard des restrictions de l'anonymisation.

Nous en avons fini avec les deux méthodes de ventilation des populations pour Rennes Métropole. Deux ventilations ont été proposées dont une qui prend en compte les particularités du bâtiment. Désormais, nous allons nous attarder sur la méthode reproductible de ventilation pour Saint-Malo Agglomération.

IV- Ventilation reproductible (Saint-Malo)

a. Données

Voici la liste des données que nous avons utilisées pour les 3 méthodologies de ventilation de population que nous avons établies :

- Une emprise des IRIS de la zone d'étude : on peut la trouver sur le site de l'IGN
- Base Bâtiment BD TOPO V3 millésime janvier 2021 : c'est une base bâtiment en accès-libre depuis janvier 2021. Elle est mise à disposition par l'IGN.
- Recensement INSEE 2017 base infracommunale : c'est une base de données qui renseigne sur la population (base identique à celle utilisée sur Rennes Métropole)
- Fichier détail logement 2017 INSEE : c'est également une base de données de population qui renseigne sur la population à l'échelle du logement (c'est aussi le même que pour Rennes Métropole)
- Corine Land Cover 2018 : c'est une couche d'occupation du sol en accès-libre et mise à disposition par l'Agence Européenne de l'Environnement. Cette couche est disponible à l'échelle de l'Europe et est exhaustive à 100%.
- Adresse premium: base de données qui permet de facilement faire des liens entre des adresses et des informations provenant d'autres BD comme la BD topo V3.

b. Méthode du filtrage des bâtiments et du calcul des informations nécessaires à la ventilation

Trois couches ont été mobilisées pour effectuer la première ventilation de Saint-Malo. Nous allons nous attarder ici à vous présenter les traitements qui ont été réalisés pour arriver à notre couche finale de répartition des populations sur Saint-Malo.

Tout d'abord, il a fallu mettre au propre la couche bâtiment issue de la BD topo. Pour cela, on supprime en premier lieu les bâtiments qui ne font pas partie de notre zone d'étude. Ensuite, il a fallu supprimer les entités qui n'étaient pas de l'ordre du résidentiel (renseigné par la variable usage), les bâtiments légers ainsi que les bâtiments qui ne sont pas en service. À noter que l'information sur l'usage des bâtiments a été croisée avec les informations du Corine Land Cover (code 111 et 112 = résidentiel) quand le premier était "indifférencié".

Une fois ce premier filtrage réalisé, des variables de mesure ont été créées. La surface au sol a été calculée. La hauteur, quand celle-ci était nulle, a été déduite, soit par le nombre d'étages, soit par les altitudes minimales et maximales au sol et au toit des bâtiments. Si aucune de ces deux solutions n'étaient possibles, le bâtiment a été supprimé. Le nombre d'étages a ensuite été calculé à partir de ce champ hauteur (hauteur divisée par 3). Nous avons ensuite éliminé ceux dont la hauteur n'est pas renseignée ainsi que ceux dont la surface au sol est inférieure à 30 m², ce qui correspond à un seuil minimal d'habitation individuelle qui nous a permis d'éliminer les portions de bâtiments. Sur le territoire d'étude, il y a également

Enfin, la surface plancher totale du bâtiment a été calculée en multipliant la surface au sol au nombre d'étages. Également, d'autres informations attributaires ont été ajoutées aux bâtiments depuis les fichiers à l'Iris (jointure attributaire via le code Iris).



Cette méthode est la même que la ventilation classique effectuée sur Rennes Métropole à l'exception de la base de données des bâtiments qui est cette fois-ci la BD TOPO v3 et non la base bâtiment de l'Audiar. On peut donc ainsi se référer à ce qui a été dit auparavant.

d. Méthode de ventilation d'estimation de la population

Dans cette méthode, les données de départ sont le nombre de logements par IRIS, que nous avons ventilé au sein de chaque bâtiment en fonction de sa surface plancher. Une chaîne de traitement a été faite en annexe⁷. Ensuite, nous avons calculé différents indicateurs démographiques en utilisant les données disponibles dans la base de données de l'INSEE :

- Logements en 2017 (IRIS)
- Population en 2017 (Recensement)

En utilisant ces statistiques démographiques et les caractéristiques du logement, nous ajustons notre résultat final.

1. Synthèse des traitements de la méthode :

- Créer la Colonne Population Potentiel (PPOT)
- Type de Bâtiment (Individuel-Collectif)
- Facteur Type de Logement (FTL)
- Facteur Résidences Secondaires et Logements Vacants (FRSLV)
- Facteur Classe d'Âge Population (FCAP)
- Facteur Indice Urbanisation IRIS(FIU)
- FACTEUR DE PONDERATION FINAL (FPF)
- VENTILATION

- Facteur Population Potentiel (PPOT)

Facteur de pondération basé sur la capacité de logement moyenne, il est obtenu en ventilant les logements IRIS en fonction de la surface plancher du bâtiment et en multipliant par le nombre moyen de personnes vivant dans un logement (2,2 en moyenne nationale).

- $2,2 * (\text{Logements IRIS} * (\text{Surface Plancher Bâtiment} / \text{Surface Plancher IRIS}))$

- Type de Bâtiment (Individuel-Collectif)

Identifier le type de bâtiment, Individuel Maison. Si le nombre de logements et le nombre d'étages est inférieur à 3, nous identifions ce bâtiment comme une maison. Tout le reste est collectif. Après avoir essayé différentes variantes, cette fonction est celle qui se rapproche le plus du nombre réel de maisons figurant dans l'INSEE.

- Maisons INSEE : 31 464 Estimation : 31 432.00
- Différence en valeur absolue (surestimation + sous-estimation) : 1581.70 (5,03%)

- Facteur Type de Logement (FTL)

Facteur de pondération tenant compte du type de bâtiment. Compte tenu des statistiques nationales de l'INSEE à ce sujet, elles ont été ajustées comme suit.

- Maison : 1 Collectif : 1,4 (Plus de population par surface dans le collectif)

⁷ Cf annexe 2 : Chaîne de traitement méthode "estimation de la population"

- **Résidences Secondaires et Logements Vacants (FRSLV)**

Facteur de pondération tenant compte du pourcentage de résidences secondaires et de logements vacants par IRIS.

- $1 - ("résidences\ secondaires" + "logements\ vacants") / "logement\ iris"$

Il s'obtient en comparant ces hébergements avec le total de l'IRIS.

- **Facteur Classe d'Âge Population (FCAP)**

Facteur de pondération tenant compte du nombre de personnes de chaque classe d'âge par IRIS. La logique derrière est que plus la personne est âgée, plus elle a un logement étendu. Selon l'INSEE : " les ménages dont la personne de référence a 65 ans ou plus disposent en moyenne d'environ 60 m² par personne, alors que ceux de moins de 30 ans disposent d'un peu plus de 30 m² par personne."

La différence de surface par personne selon ce critère est remarquable, nous avons donc décidé de l'inclure dans notre méthode. Il est calculé en attribuant un poids différent pour chaque classe d'âge, on obtient une moyenne pondérée qui permet d'ajuster la population à ventiler en fonction de l'âge moyen de l'IRIS :

- $1 - ((POP_30-44 * 0.1 + POP_45-59 * 0.2 + (POP_60-74 + POP_75+) * 0.4) / (POP_30-44 + POP_45-59 + POP_60-74 + POP_75+))$

- **Facteur Indice Urbanisation IRIS(FIU)**

Facteur de pondération tenant compte de la surface plancher de tous les bâtiments par IRIS.

- $1 + (0.5 * (1 - ("SURF_IRIS" - "SURF_PLAN_sum") / "SURF_IRIS"))$

Selon l'INSEE : "la surface par personne des logements augmente à mesure que la taille de l'agglomération diminue. Dans l'unité urbaine de Paris, la surface moyenne par personne est de 31 m², elle est de 47 m² dans les communes rurales."

Pour mesurer cette urbanisation, nous utilisons la surface plancher comme indicateur, en le comparant à la surface totale de l'IRIS.

- **FACTEUR DE PONDÉRATION FINAL (FPF)**

Nous multiplions tous les facteurs de pondération obtenus jusqu'à présent entre eux :

- $FPF = PPOT * FTL * FRSLV * FCAP * FIU$

- **VENTILATION**

Nous ajoutons d'abord la population totale de tous les IRIS et le FPF de tous les bâtiments de notre zone d'étude. Ensuite, nous faisons la ventilation selon la formule :

- $Population\ Total\ IRIS * (FPF / FPF\ Total)$

2. Avantages et inconvénients :

Cette méthode nous permet d'obtenir un facteur de pondération théorique en réduisant la pollution et les biais inhérents aux modèles de ventilation utilisés par l'INSEE lors de la conduite de ses études et du regroupement des données.

La donnée principale de base utilisée est le nombre de logements, qui est une donnée plus précise, facile à calculer et qui varie moins dans le temps.

Cette méthode est divisée en plusieurs parties, dont chacune peut être améliorée individuellement, exclue si elle n'est pas jugée pratique, et permet d'inclure facilement de nouveaux coefficients qui pourraient grandement améliorer la méthode.

Comme inconvénient pour mettre en évidence la valeur de départ, en l'occurrence 2,2 personnes par logement (moyenne nationale). Ce chiffre varie selon les régions, le modèle devrait pouvoir s'adapter en fonction du code IRIS, ou ajuster automatiquement sa valeur en fonction des données IRIS.

e. Méthode de ventilation avancée

Cette dernière méthode de ventilation reproductible proposée a été développée sous RStudio. Le langage de programmation est donc le langage R, destiné aux statistiques et aux sciences des données. Ce changement d'environnement de travail a pour principale motivation l'exploitation de fonctionnalités avancées, indisponible dans l'environnement de travail de QGIS et de son modèleur graphique.

De plus, la base infracommunale 2017 du recensement de la population de l'INSEE sera mise de côté, pour des questions de finesse de l'information et sera remplacée par le fichier de détail logement. Également, une nouvelle base de données sera exploitée : la base Adresse Premium de l'IGN.

Aparté sur la base Adresse Premium de l'IGN : Il s'agit d'un produit complémentaire aux bases de données de l'IGN et il est composé de différents thèmes :

- **A_ADR_PARC** : qui comprend les objets géométriques permettant de faire le lien entre les adresses de la BD TOPO et les parcelles du produit Parcellaire Express (PCI).
- **B_ADR_BATI** : qui comprend les objets géométriques permettant de faire le lien entre les adresses de la BD TOPO et les bâtiments de la BD TOPO.
- **C_BATI_PARC** : qui comprend les objets géométriques permettant de faire le lien entre les bâtiments de la BD TOPO et les parcelles du produit Parcellaire Express (PCI).
- **D_ADR_IRIS** : qui comprend des tables attributaires sans géométrie permettant de faire le lien entre les identifiants des adresses de la BD TOPO et les identifiants des IRIS de l'INSEE.
- **E_ADR_HEXA** : qui comprend des tables attributaires sans géométrie permettant de faire le lien entre les identifiants des adresses de la BD TOPO et les codes Hexavia et Hexaclé de la Poste.

Ici, on ne retiendra donc que le thème **B_ADR_BATI**, afin d'enrichir notre base bâtiment du nombre d'adresses reliées au bâtiment, afin de mieux estimer le nombre potentiel de logements par la suite. La modélisation du **B_ADR_BATI** est en deux dimensions linéaire, en réalité sa modélisation n'est que peu importante puisque les seuls champs que nous retiendrons pour cette méthode sont le champ ID_BAT, permettant d'effectuer le lien avec la BD TOPO et le champ NB_ADR renseignant sur le nombre d'adresses reliées au bâtiment. L'étude de cette base de données n'ira pas plus loin dans ce rapport, pour autant il semblerait pertinent d'analyser plus en profondeur les atouts et les inconvénients de ces thèmes.

Le script R fourni avec ce rapport propose le filtrage des bâtiments ainsi que l'agrégation du fichier de détail logement. Nous ne rentrons pas ici dans les détails de ces traitements puisqu'ils sont similaires à ce qui a été expliqué ci-devant dans les sections Fichier de détail logement et Méthode du filtrage des bâtis. Nous nous concentrerons donc sur la méthodologie de ventilation⁸.

Sur la couche de bâtiments filtrée, on va d'abord importer le code IRIS depuis la couche "Contours... IRIS" de l'IGN. Pour ce faire, on transforme l'ensemble de nos polygones

⁸ Annexe 3 : Chaîne de traitement méthode de ventilation "avancée"

de bâtiments en simples points, pour éviter les éventuels problèmes d'appartenance à plusieurs IRIS d'un bâtiment et on joint par localisation (soit dans quel IRIS est contenu le point) ce code IRIS. Ce code IRIS sert de filtre pour le territoire d'étude, un bâtiment sans code IRIS sera, de fait, supprimé.

La base de données Adresse Premium est simplifiée. On ne retient que les deux champs, présentés ci-devant, qui nous intéressent et on simplifie en ne retenant que les associations uniques identifiant bâtiment - nombre d'adresses. On joint ensuite, par valeur de champ (soit l'identifiant du bâtiment) cette information sur le nombre d'adresses. Dans les cas où le bâtiment ne reçoit aucune adresse de la base de données, une seule et unique adresse lui est attribuée d'office. On en profite également pour calculer la somme du nombre d'adresses par code IRIS.

On joint finalement les informations pertinentes (le nombre de maisons en résidence principale et secondaire, la population y résidant, le nombre d'appartements en résidence principale et secondaire et la population y résidant) du fichier de détail logement dans les bâtiments par valeur de champ (soit le code IRIS).

Une fois toutes ces informations contenues dans la base bâtiments, on effectue un ajustement dans certains IRIS, selon la condition suivante : la somme du nombre d'appartements (en résidence principale et secondaire) divisée par 2 (d'après la définition d'un bâtiment collectif, qui de fait, contient au moins deux logements) est inférieure à la somme du nombre d'adresses à laquelle on soustrait la somme du nombre de maisons (en résidence principale et secondaire). Dans ces cas précis, on ne peut pas coïncider avec les chiffres du fichier de détail logement et à la définition d'un bâtiment collectif. Alors on procède comme suit au sein de chaque IRIS concerné : une table temporaire est créée, contenant l'ensemble des bâtiments de l'IRIS et tous ces bâtiments sont triés par surface plancher croissante. On va ensuite parcourir un à un nos bâtiments et vérifier notre test. Tant que le test reste vrai, on va supprimer le bâtiment en question et mettre à jour le nombre d'adresses de l'IRIS. Une fois l'ajustement fait, on va fusionner cette table temporaire avec les bâtiments qui ne nécessitent pas d'ajustement et répéter ces opérations jusqu'à avoir traité tous les bâtiments des IRIS concernés. Puis, on distingue nos bâtiments individuels de nos bâtiments collectifs. Pour ce faire, on va, par code IRIS, trier de nouveau nos bâtiments par surface plancher croissante. Ici, on va attribuer, en commençant par la surface plancher la plus faible, le type individuel. Une variable de surface plancher pour le type individuel est créée, elle se verra attribuer sa valeur initiale auxquelles on ajoute la surface plancher du bâtiment en question. Le nombre de logements attribués vaut le nombre d'adresses reliées au bâtiment. On va répéter ces opérations jusqu'à avoir placé tous les logements individuels, selon les chiffres du fichier de détail logement. Le reste de nos bâtiments sont, de facto, des bâtiments collectifs. Finalement, on va ventiler :

- la population résidant en maison, en multipliant cette valeur de population par un rapport entre la surface plancher du bâtiment à ventiler sur la somme des surfaces plancher des bâtiments individuels ;
- les logements collectifs, en multipliant le nombre de logements collectifs par un rapport entre la surface plancher du bâtiment à ventiler sur la somme des surfaces plancher des bâtiments collectifs ;

- la population résidant en appartement, en multipliant cette valeur de population par un rapport entre la surface plancher du bâtiment à ventiler sur la somme des surfaces plancher des bâtiments collectifs.

La ventilation étant terminée, l'export de ces informations se fait sous fichier .csv.

Il en ressort de cette méthode plusieurs avantages. L'environnement RStudio apporte bien plus de fonctionnalités, de reproductibilité et d'automatisation que l'environnement QGIS. Il est bien plus simple de modifier le script ou de le réécrire dans un langage différent (Python par exemple). Il a également l'avantage d'être plus stable, dans des phases de test finales, il a été possible d'effectuer une ventilation à l'échelle de l'Ille-et-Vilaine (soit environ un million de bâtiments) sans soucis techniques. Enfin, la formule de calcul induit nécessairement une équivalence entre l'agrégat des valeurs ventilées et les valeurs initiales du fichier de détail logement.

Pour autant, cette méthode a l'inconvénient de nécessiter des compétences techniques de programmation plus poussées. De plus, les traitements d'ajustement et de ventilation sont longs de par le parcours de table ligne par ligne indispensable. Enfin, cette méthode implique de traiter des cas particuliers. La question des logements en bâtiments collectifs a pu être traitée, en supprimant toutefois des bâtiments considérés comme résidentiels. Mais il reste encore des cas où le nombre d'adresses dans l'IRIS est égale aux nombres de logements individuels. Dans ces cas précis, il nous est impossible de ventiler la population résidant dans logements collectifs et aucune solution efficace n'a, à ce jour, été trouvée.



Figure 34 : Représentation de l'information de la population à l'échelle des bâtiments du centre de Saint-Malo et de Bain-de Bretagne

f. Perspectives

Au travers de la ventilation sur Saint-Malo Agglomération, nous avons pu voir les différentes possibilités de ventilation de la population en utilisant à 100% les données en open data. Basé sur un tandem de données Insee-IGN principalement, les méthodes proposées ont chacune leurs particularités. La première, clé en main, rapide et facile à prendre en main. La seconde qui permet d'ajuster les facteurs de pondération en fonction des territoires. La troisième, enfin, qui présente une méthode novatrice, mais avec des complexités techniques de par la méthode et l'environnement de travail.

Nous avons pu avoir ainsi les limites inhérentes à l'outil QGIS, c'est-à-dire la gestion de données massives. Pour y palier, nous nous sommes tournés vers un logiciel libre (R) et qui permet la gestion plus rapide et stable de ces volumes de données, mais aussi une meilleure reproductibilité.

Ces méthodes de ventilation ouvrent, elles aussi, la porte à de nombreux usages dans l'énergie, le géomarketing, l'aménagement du territoire, bien plus d'ailleurs que les méthodes présentées pour Rennes Métropole, car celles-ci sont composées à 100% d'éléments provenant du libre.

Nous en avons fini avec la présentation des différentes méthodes de ventilation et plus globalement de l'atelier Audiar. Nous vous avons présenté au travers ce rapport plusieurs méthodes ainsi que les données que nous avons mobilisées. Nous allons désormais faire un bilan du travail effectué en collaboration avec l'Audiar.

Bilan global de l'atelier

En essayant de prendre du recul sur ce qui a été fourni en marge de cet atelier, on peut dire que le travail effectué a été très enrichissant pour nous étudiants. Nous avons exploré une multitude de jeux de données, provenant de sources variées. Tout d'abord les données issues de l'Insee, qui nous ont permis de nous familiariser avec les mesures statistiques issues de la population et que l'on a étudiées à différentes échelles (IRIS, communes, données, carroyage). Il y a également les données du bâtiment, qui nous ont forgé notre esprit critique du fait des problèmes géométriques et de données attributaires parfois incohérentes. Il y a également les données issues du CLC, d'OSM, ... qui ont renforcé notre culture de la donnée spatialisée.

Désormais, nous avons une vision plus globale et un esprit plus aguerri sur toutes les données disponibles librement en France. De plus, nous avons gagné en maturité et en expérience quant à la conduite de projet en géomatique. C'était pour la plupart d'entre nous la première fois que l'on s'exerçait à une telle tâche. Certes, nous avons eu quelques difficultés au démarrage de cet atelier, notamment en termes de coordination, de prises de décisions et d'expertises. Mais nous avons su nous mobiliser afin de fournir un travail qui répondait aux attentes du commanditaire.

Nous tenons encore une fois à remercier l'agence de l'Audiar et notamment nos 2 commanditaires Johan Poquet et Emmanuel Bouriau, qui, par leurs bienveillances, leurs conseils et leurs critiques, nous ont guidés dans l'exploration des méthodologies. Nous tenons également à remercier le Master SIGAT et notamment Boris Mericksay, François Leprince et Erwan Quesseveur pour leur accompagnement durant cet atelier.

Merci à tous !

Bibliographie :

- Régis Dugué, Michel Hénin. Note méthodologique sur Densimos 2009 et Densibati 2009. IAU île-de-France. Janvier 2014. https://geo.valdemarne.fr/portail/system/files_force/note_methodologique_densimos2009_et_densibati2009_2.pdf?download=1
- François-Xavier Robin, Vincent Demeules. Méthodes d'estimations de population : comparaisons et seuils de validité. [Rapport de recherche] Centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques (CERTU). 2005, 88 p., photos, illustrations en couleurs, figures, graphiques, tableaux. <https://hal-lara.archives-ouvertes.fr/hal-02156445>
- Lwin, K. and Y. Murayama. "A GIS Approach to Estimation of Building Population for Micro-spatial Analysis." Trans. GIS 13 (2009): 401-414. <https://www.cdema.org/virtuallibrary/images/A%20GIS%20Approach%20to%20Estimation%20of%20Building.pdf>
- Estimation de la population dans le Doubs en 2014. <https://bit.ly/2QYvz1T>
- Akiyama Y, Takada T, Shibasaki R (2013) Development of micropopulation census through disaggregation of national population census. In: CUPUM2013 conference papers, #110. https://cupum2013.geo.uu.nl/download/usb/contents/pdf/shortpapers/110_Akiyama.pdf

Annexe :

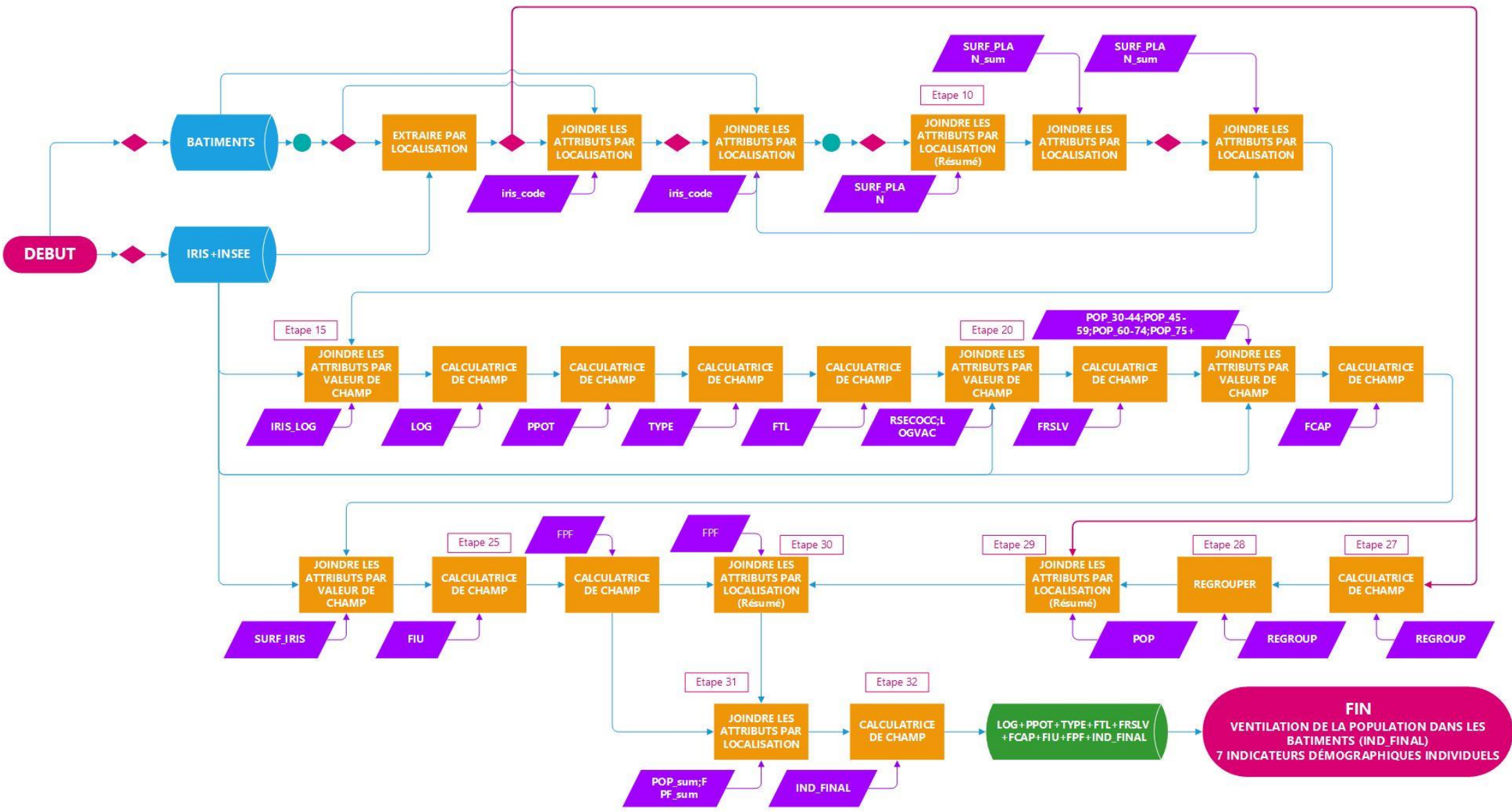
Annexe 1 : Tableau récapitulatif des champs des carreaux 200 m et 1 km de l'INSEE

Nom du champ	Description
fid	ID du carreau
Ind	Nombre d'individus
Men	Nombre de ménages
Men_pauv	Nombre de ménages pauvres
Men_1ind	Nombre de ménages d'un seul individu
Men_5ind	Nombre de ménages de 5 individus ou plus
Men_prop	Nombre de ménages propriétaires
Men_fmp	Nombre de ménages monoparentaux
Ind_snv	Somme des niveaux de vie winsorisés des individus
Men_surf	Somme de la surface des logements* du carreau
Men_coll	Nombre de ménages en logement collectif
Men_mais	Nombre de ménages en maison
Log_av45	Nombre de logements* construits avant 1945
Log_45_70	Nombre de logements* construits entre 1945 et 1969
Log_70_90	Nombre de logements* construits entre 1970 et 1989
Log_ap90	Nombre de logements* construits depuis 1990
Log_inc	Nombre de logements* dont la date de construction est inconnue
Log_soc	Nombre de logements* sociaux
Ind_0_3	Nombre d'individus de 0 à 3 ans
Ind_4_5	Nombre d'individus de 4 à 5 ans
Ind_6_10	Nombre d'individus de 6 à 10 ans
Ind_11_17	Nombre d'individus de 11 à 17 ans
Ind_18_24	Nombre d'individus de 18 à 24 ans
Ind_25_39	Nombre d'individus de 25 à 39 ans
Ind_40_54	Nombre d'individus de 40 à 54 ans
Ind_55_64	Nombre d'individus de 55 à 64 ans
Ind_65_79	Nombre d'individus de 65 à 79 ans
Ind_80p	Nombre d'individus de 80 ans et plus
Ind_inc	Nombre d'individus dont l'âge est inconnu
Variables complémentaires de la grille de niveau naturel	
Id_carr_n	Identifiant basé sur la norme Inspire
l_pauv	Nombre de carreaux de 200m compris dans le carreau qui ont été traités pour respecter la confidentialité sur le nombre de ménages pauvres
t_maille	Taille en mètres du côté du carreau
Variables complémentaires de la grille de 1 km	
Id_carr_1km	Identifiant basé sur la norme Inspire

I_est_1km	Vaut 2 si le carreau donne les vraies valeurs mais le nombre de ménages affiché apparaît comme inférieur à 11 pour des questions d'arrondis. Vaut 1 si le carreau est imputé par une valeur approchée. Vaut 0 dans les autres cas.
I_pauv	Nombre de carreaux de 200m compris dans le carreau de 1km qui ont été traités pour respecter la confidentialité sur le nombre de ménages pauvres.
Variables complémentaires de la grille de 200 m	
IdINSPIRE	Identifiant basé sur la norme Inspire
Id_carr1km	Identifiant du carreau de 1km auquel appartient le carreau de 200 m
Id_carr_n	Identifiant du carreau de niveau naturel auquel appartient le carreau de 200 m
Id_car2012	L'identifiant Inspire du carreau de 200m présent dans la base de données carroyées à 200 m issu de la source RFL2010 (il diffère en nombre de caractères).
I_est_cr	Vaut 1 si le carreau est imputé par une valeur approchée, 0 sinon
I_est_1km	Vaut 1 si le carreau de 1km auquel est rattaché le carreau de 200m est imputé par une valeur approchée, 0 sinon
Groupe	Numéro du groupe auquel appartient le carreau
Depcom	Code commune, selon le code officiel géographique 2019, auquel sont rattachés la majorité des ménages du carreau.
I_pauv	Vaut 1 si le carreau a été traité pour respecter la confidentialité sur le nombre de ménages pauvre (valeur ramenée à 80% du nombre de ménages)
	Champs retenus
	Champs non retenus
	Champs incertains

Annexe 2 : Chaîne de traitement méthode “estimation de la population” - méthode reproductible

DIAGRAMME DE TRAITEMENT DE LA METHODE DE POPULATION POTENTIAL (2EME METHODE REPRODUCTIBLE)



Annexe 3 : Chaîne de traitement méthode de ventilation “avancée” - méthode reproductible

