

Sommaire

1	Introduction	4
2	Présentation du contexte	4
2.1	Intérêt de l'étude	4
2.2	Cadre conceptuel de l'étude	4
2.3	Revue de littérature	5
3	Méthodologie	6
3.1	Présentation des données	6
3.2	Motivation	6
3.3	Modèles Linéaires Généralisés	6
3.3.1	Régression Logistique	6
3.3.2	Régression de Poisson	7
3.4	Modèles Linéaires Mixtes	7
3.5	Modèles Linéaires Généralisés à Effets Mixtes (GLMM)	7
3.6	Prédictions et Simulations	8
4	Analyse descriptive	8
4.1	Analyse univariée	8
4.2	Analyse bivariée	12
4.2.1	Taux de consultation et population totale	12
4.2.2	Taux de consultation et population âgée	12
4.2.3	Taux de consultation et CSP	12
4.2.4	Analyse de corrélation	12
4.3	Autocorrélation	13
4.3.1	Définition de l'indice de Moran	13
4.3.2	Formule de l'indice de Moran	13
4.3.3	Matrice de poids basée sur la distance de Haversine	13
4.3.4	Formule de la distance de Haversine	16
4.3.5	Construction de la matrice de poids	16
ANNEXES		17

Liste des Tableaux

1	Taux de consultations selon la taille de la commune	12
2	Taux de consultations selon la population âgée	12
3	Corrélations de Pearson entre le taux de consultation et les autres variables	12
4	Résultats du test de Moran	17

Liste des Figures

1	Pyramide des âges	9
2	Taux de Natalité et Taux de Mortalité	10
3	Répartition du nombre et du taux de consultations	11
4	Relation entre taux de consultations et part des plus de 75 ans	14
5	Corrélations entre le nombre de visite et quelques variables	15
6	Densité des distances	16
7	Carte du nombre de consultations par commune	18
8	Carte du taux de consultations par commune	18
9	Carte du taux de consultations par commune pour les plus de 19 ans	19

1 Introduction

La répartition géographique des besoins en soins de santé est un enjeu majeur pour les politiques publiques, notamment en ce qui concerne l'accès aux services de médecine de ville. Les inégalités territoriales dans l'offre et la demande de soins peuvent entraîner des disparités significatives en matière de santé, affectant particulièrement les populations vivant dans des zones sous-dotées en professionnels de santé. Comprendre ces dynamiques spatiales et socio-démographiques est essentiel pour identifier les zones prioritaires et orienter les décisions en matière d'allocation des ressources.

Dans ce contexte, ce travail propose une modélisation du nombre de consultations en médecine de ville à l'échelle communale, en tenant compte des caractéristiques démographiques, socio-économiques et spatiales des communes. L'objectif est double : d'une part, analyser les facteurs influençant la demande de soins, et d'autre part, identifier les zones susceptibles de dépasser un seuil critique de "désert médical". Pour ce faire, nous nous appuyons sur une base de données riche et variée, comprenant des informations issues du Système National des Données de Santé (SNDS) pour la période 2018-2022, ainsi que des indicateurs socio-démographiques et géographiques.

Notre approche méthodologique repose sur une combinaison de techniques statistiques et spatiales. Nous commencerons par une analyse descriptive et cartographique des données pour visualiser les tendances et les disparités territoriales. Ensuite, nous utiliserons des modèles de régression de Poisson pour modéliser le nombre de consultations annuelles, en tenant compte des effets fixes (caractéristiques des communes) et des effets aléatoires (variations spatiales). Enfin, une régression logistique binaire sera employée pour évaluer la probabilité qu'une commune dépasse un seuil prédéfini de "désert médical".

Ce travail s'inscrit dans une perspective à la fois académique et opérationnelle. Sur le plan académique, il contribue à l'étude des inégalités territoriales en santé en proposant une méthodologie robuste pour l'analyse spatiale des données de soins. Sur le plan opérationnel, il fournit des outils pour identifier les zones prioritaires et soutenir la prise de décision en matière de politiques de santé publique.

2 Présentation du contexte

2.1 Intérêt de l'étude

2.2 Cadre conceptuel de l'étude

Dans cette partie, nous allons définir certaines notions clés qui apparaissent dans notre étude entre autres, le nombre de visites, le nombre de visites espérés ainsi que le taux de visite.

1. Nombre de visites espérés

Le nombre de visite espérés, terme qui apparaîtra dans notre modélisation est le nombre de visite qu'il y aurait eu dans chaque commune si le taux de visite était le même dans ces dernières. En d'autres termes si r est le taux moyen de visite alors, le nombre de visite escompté noté μ est calculé par :

$$\mu_i = r * P_i$$

où P_i est le nombre d'habitants dans la commune i

2. Taux de visite

Le taux de visite n'est rien d'autre que le nombre moyen de visite dans chaque commune. IL est calculé en divisant le nombre de visite par la population de la commune en question. En d'autres termes, il s'agit du nombre de visites que chaque habitant de la commune a effectué en moyenne.

$$\tau_i = \frac{n_i}{P_i}$$

où τ_i et n_i sont respectivement le taux et le nombre de visite de la commune i .

2.3 Revue de littérature

La modélisation des visites dans les hôpitaux est cruciale pour comprendre les dynamiques de la santé publique et pour optimiser la gestion des ressources médicales. Dans ce contexte, les modèles de régression de type Poisson généraux, en particulier les modèles mixtes généralisés (GLMM), sont souvent utilisés pour gérer les données de comptage, en tenant compte des spécificités démographiques et géographiques des agglomérations, ici définies par les communes.

Les recherches antérieures, comme celles de Mohebbi et al. (2011), montrent l'importance d'intégrer des effets spatiaux dans les modèles de comptage, surtout lorsqu'on évalue des phénomènes tels que le cancer œsophagien (EC) dans des provinces spécifiques. Leur étude souligne comment la non-prise en compte de l'autocorrélation spatiale peut conduire à une estimation biaisée des effets des variables socio-économiques sur l'incidence des maladies. Ce phénomène pourrait également s'appliquer à la modélisation des visites à l'hôpital, où certains facteurs jouent un rôle significatif.

Selon l'article, trois structures d'autocorrélation peuvent être appliquées dans le cadre de la régression Poisson lorsqu'on traite des données de comptage. Ces modèles comprennent :

1. **Modèle Poisson avec effets aléatoires non spatiaux** : Bien qu'utiles, ces modèles négligent l'autocorrélation spatiale, ce qui peut conduire à une sous-estimation des erreurs standard.
2. **Modèle avec effets aléatoires spatiaux basés sur la distance** : Ce modèle prend en compte l'influence des agglomérations voisines, mais peut ne pas capturer efficacement l'hétérogénéité locale.
3. **Modèle avec effets aléatoires de type voisinage** : Utilisant la structure conditionnelle autorégressive (CAR), ce modèle est le plus adapté pour les données spatiales en intégrant les interactions entre les communes. Pour notre étude, le modèle avec effets aléatoires de type voisinage est recommandé pour la modélisation du nombre de visites à l'hôpital en France, car il permet de mieux prendre en compte l'effet de proximité entre communes.

L'application des principes issus de l'article de Mohebbi et al. peut enrichir notre étude sur la fréquentation des hôpitaux en France. En intégrant des effets aléatoires spatiaux adaptés à la structure des données, nous pouvons réaliser des estimations plus précises et utiles pour la planification hospitalière.

En ce qui concerne les facteurs explicatifs du nombre de visite dans les hôpitaux, plusieurs études se sont penchées sur ce sujet. De nombreuses études ont montré que le nombre de consultations médicales est influencé par divers facteurs sociodémographiques, allant des caractéristiques individuelles aux contextes socio-économiques et territoriaux.

1. Influence de l'âge et du sexe

L'âge constitue un déterminant majeur du recours aux soins. Les personnes âgées, en particulier celles de 65 à 79 ans, consultent plus fréquemment en raison de la prévalence accrue de maladies chroniques et du suivi médical nécessaire à leur prise en charge [1](#). En revanche, la population jeune et en bonne santé présente une utilisation plus sporadique des services médicaux. Le sexe est également un facteur différenciant important. De manière générale, les femmes consultent plus fréquemment que les hommes. Cette différence est attribuée aux besoins spécifiques en santé reproductive, mais aussi à une plus grande propension à rechercher des soins préventifs [2](#). En revanche, les hommes, notamment dans les catégories socio-professionnelles les plus actives, ont tendance à sous-utiliser les services de soins, ce qui peut entraîner des diagnostics plus tardifs et des complications médicales accrues.

2. Impact du statut socio-économique et du niveau d'éducation

Le revenu et la précarité économique influencent considérablement l'accès aux soins. Les individus à revenu élevé bénéficient généralement d'un meilleur accès aux consultations médicales, notamment grâce à une plus grande couverture sociale et des assurances complémentaires leur permettant de réduire les coûts associés aux soins [4](#). À l'inverse, les personnes en situation de précarité rencontrent des obstacles financiers, administratifs et culturels qui limitent leur recours aux soins, malgré des besoins souvent accrus en raison de conditions de vie plus précaires. Le niveau d'éducation joue un rôle clé dans la fréquentation des services de santé. Une instruction plus élevée est associée à une meilleure connaissance des risques sanitaires et à une adoption plus

proactive des comportements de prévention, ce qui entraîne un recours plus fréquent aux soins médicaux 4. En revanche, un faible niveau d'éducation est souvent corrélé à un moindre suivi médical et à une utilisation plus tardive des services de soins, notamment en cas de complications.

3. Influence du contexte familial et de l'environnement social

L'état matrimonial influence également la fréquence des consultations médicales. Les personnes mariées ou vivant en couple consultent davantage, bénéficiant du soutien d'un conjoint qui peut inciter à prendre soin de sa santé et à consulter régulièrement un médecin 1. Par ailleurs, l'accès à un médecin traitant ou de famille constitue un déterminant important. Les patients disposant d'un suivi médical régulier sont plus enclins à effectuer des consultations préventives et à être orientés rapidement vers des spécialistes si nécessaire 1.

4. Perception de l'état de santé et accès géographique aux soins

La perception de la santé est un facteur déterminant du recours aux soins. Les individus qui considèrent leur état de santé comme excellent ou très bon consultent rarement, tandis que ceux qui ont une perception négative de leur état de santé ont tendance à multiplier les visites médicales 1. Enfin, les inégalités spatiales dans l'accès aux soins modulent également la fréquence des consultations. En milieu urbain, la densité médicale plus élevée facilite l'accès aux services de soins, tandis qu'en zones rurales ou médicalement sous-dotées, les délais d'attente et les distances à parcourir constituent des freins majeurs 3.

3 Méthodologie

3.1 Présentation des données

Les données que nous avons utilisées nous proviennent de ...

3.2 Motivation

Les modèles linéaires généralisés à effets mixtes (GLMM) combinent :

- Les caractéristiques des modèles linéaires généralisés (GLM) pour modéliser des variables non-normalement distribuées.
- Les propriétés des modèles à effets mixtes pour gérer des données groupées ou hiérarchiques.

3.3 Modèles Linéaires Généralisés

Un GLM relie le prédicteur linéaire η à la moyenne μ de la réponse à travers une fonction de lien g :

$$g(\mu) = \eta = \beta_0 + \sum_{i=1}^m \beta_i x_i$$

Les distributions possibles incluent :

- **Normale** : Régression linéaire classique, avec lien identité.
- **Binomiale** : Régression logistique pour données binaires, avec lien logit.
- **Poisson** : Régression de Poisson pour données de comptage, avec lien logarithmique.

3.3.1 Régression Logistique

Modélise une réponse binaire ($y \sim B(n, p)$), où p est la probabilité de succès :

$$P(y|n, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

La probabilité p est reliée au prédicteur par la fonction logistique :

$$p = \frac{1}{1 + e^{-\eta}} \quad \text{où} \quad \eta = \beta_0 + \sum_{i=1}^m \beta_i x_i.$$

Le log-vraisemblance est exprimé comme :

$$\ell(\beta) = \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log (1 - p_i)]$$

3.3.2 Régression de Poisson

Utilisée pour modéliser des données de comptage ($y \sim \text{Pois}(\lambda)$), où λ est la moyenne et la variance :

$$P(y|\lambda) = \frac{\lambda^y}{y!} e^{-\lambda}$$

Le lien logarithmique assure $\lambda > 0$:

$$\log \lambda = \beta_0 + \sum_{i=1}^m \beta_i x_i$$

L'espérance est $E[y] = \lambda$.

3.4 Modèles Linéaires Mixtes

Ces modèles ajoutent des termes d'effets aléatoires \mathbf{Zu} au prédicteur linéaire :

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \varepsilon,$$

avec :

- $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, les effets aléatoires.
- $\varepsilon \sim N(\mathbf{0}, \mathbf{R})$, les résidus.

La matrice de covariance totale est :

$$\text{Var}(\mathbf{y}) = \mathbf{ZGZ}^T + \mathbf{R}.$$

Les paramètres sont estimés par maximum de vraisemblance (ML) ou par vraisemblance restreinte (REML).

3.5 Modèles Linéaires Généralisés à Effets Mixtes (GLMM)

Un GLMM étend les GLM en intégrant des effets aléatoires :

$$g(\mu) = \mathbf{X}\beta + \mathbf{Zu},$$

où :

- $g(\cdot)$ est la fonction de lien.
- $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ est le vecteur d'effets aléatoires.

Les paramètres sont estimés via des méthodes comme :

- Approximations Laplaciennes.
- Quadrature gaussienne adaptative.
- Méthodes MCMC (chaînes de Markov Monte Carlo).

3.6 Prédictions et Simulations

Les GLMM permettent deux types de prédictions :

- **Conditionnelles** : Basées sur les effets aléatoires spécifiques (\mathbf{u}).
- **Marginales** : En intégrant sur les effets aléatoires.

Les simulations utilisent des approches paramétriques pour évaluer la variabilité et tester les hypothèses. Une approche courante est le bootstrap paramétrique :

1. Générer des données simulées basées sur les paramètres estimés.
2. Réajuster le modèle pour chaque jeu de données simulé.
3. Analyser la distribution des estimations obtenues.

4 Analyse descriptive

1. Description de la population d'étude

Notre population d'étude est une population assez homogène en matière d'âge. Cependant plus on dépasse les 75 ans et moins on rencontre de personnes. D'autres part notre population est fortement masculine avec une forte proportion des hommes quelle que soit la tranche d'âge à l'exception des tranches du troisième âge.

Dans cette partie, nous allons réaliser quelques statistiques descriptives sur nos données.

4.1 Analyse univariée

1. Taux et Nombre de visites

L'analyse des statistiques descriptives sur le nombre de consultations annuelles de médecin généraliste entre 2018 et 2022 révèle une distribution fortement asymétrique à droite, avec une grande dispersion des données. La moyenne de 19130 consultations, nettement supérieure à la médiane de 9127, indique la présence de valeurs extrêmes tirant la distribution vers le haut. Cette asymétrie est confirmée par l'écart considérable entre le minimum de 1037 et le maximum de 765833 consultations par an.

La moitié des médecins généralistes effectuent entre 5993 et 17290 consultations annuellement, ce qui suggère une variabilité importante dans la charge de travail. La médiane de 9127 consultations par an, équivalant à environ 25 consultations par jour ouvrable, semble plus représentative de l'activité typique d'un médecin généraliste que la moyenne influencée par les valeurs extrêmes. Ces statistiques mettent en lumière la diversité des pratiques et des charges de travail parmi les médecins généralistes, avec potentiellement quelques cas atypiques présentant un volume de consultations exceptionnellement élevé.

Le nombre de visites pouvant potentiellement être influencé par la taille de la commune et donc par sa population, nous avons éliminé cet effet en calculant le taux de consultations qui n'est autre que le nombre de consultations moyennes par personnes.

2. Taux de mortalité et de Natalité

Dans les communes étudiées, le taux de natalité et de mortalité sont un peu élevés avec la plupart des taux variant entre 5 et 15 pour 1000 en ce qui concerne la natalité et 0 et 20 pour 1000 pour la mortalité. On remarque une corrélation négative entre ces deux taux. Néanmoins cette corrélation n'a à priori aucun sens. Par ailleurs, l'observation des distributions permet de constater que la natalité est de façon générale élevée par rapport à la mortalité dans les communes étudiées.

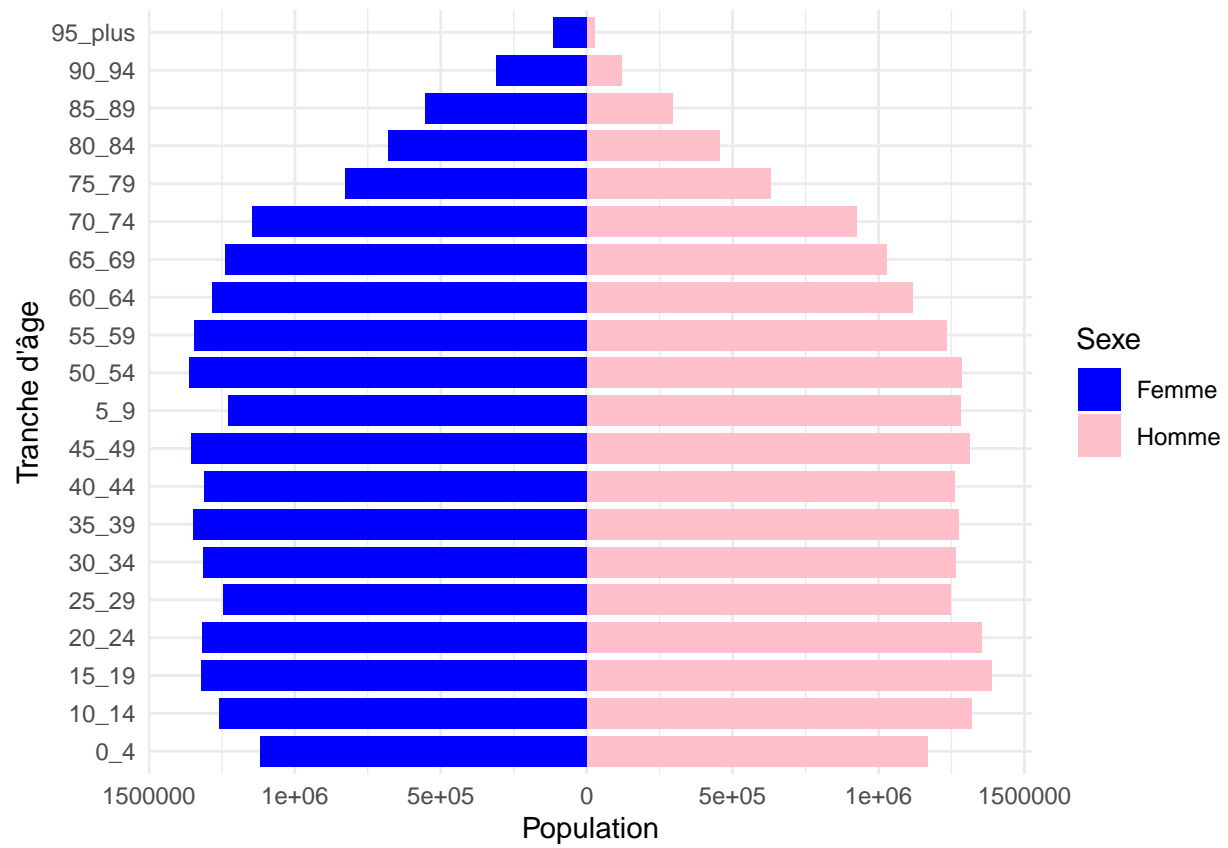


Figure 1: Pyramide des âges

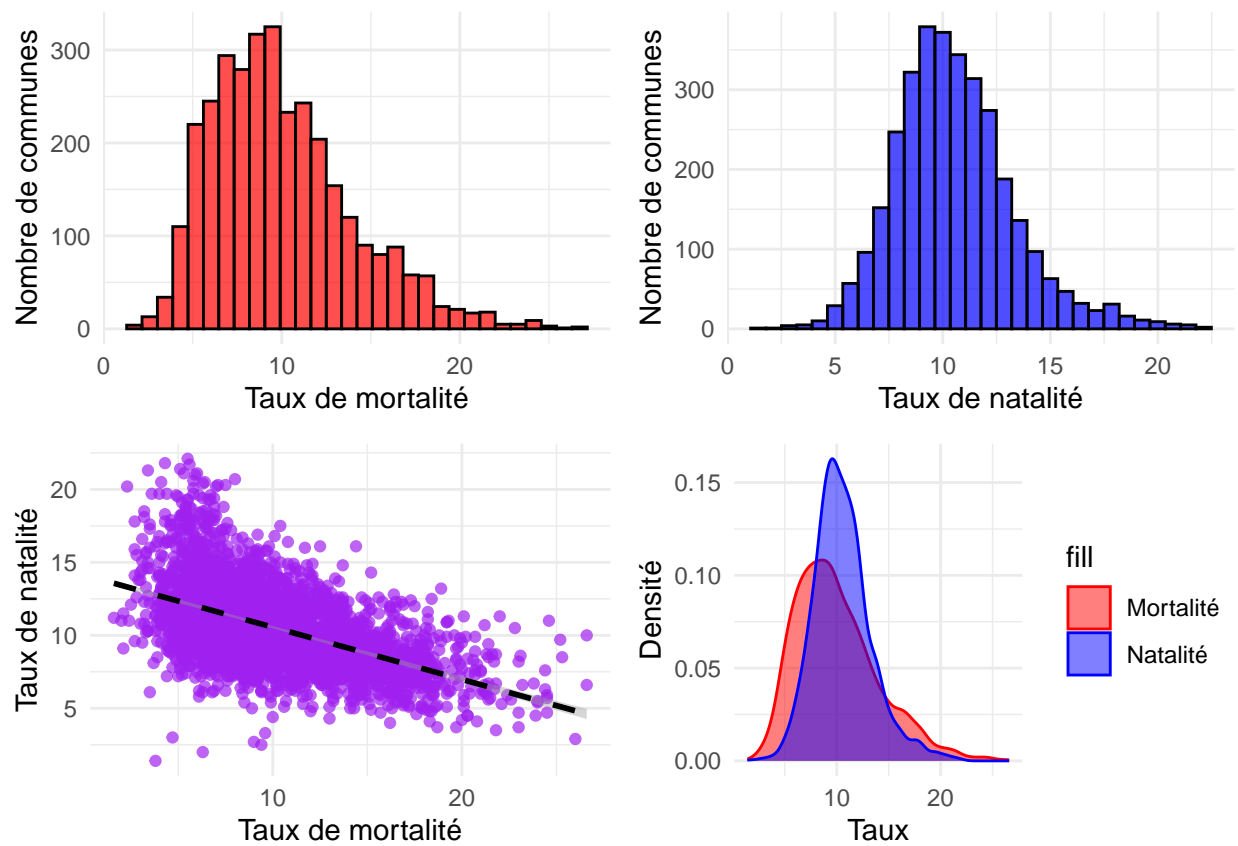


Figure 2: Taux de Natalité et Taux de Mortalité

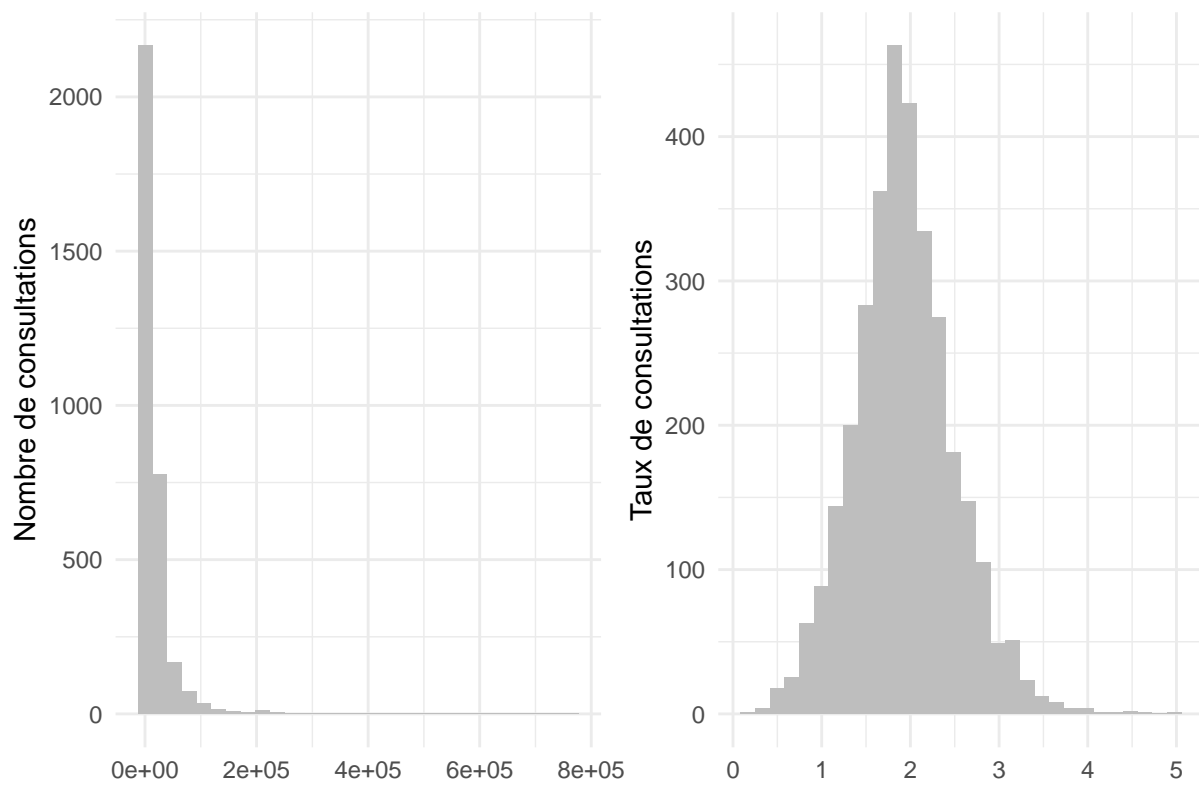


Figure 3: Répartition du nombre et du taux de consultations

4.2 Analyse bivariée

Nous allons ici, voir s'il y a un lien à priori entre le taux de consultation et certaines de nos variables explicatives. Ainsi, nous avons d'abord réalisé une analyse descriptive bivariée puis nous avons calculé la corrélation de Pearson pour évaluer le lien linéaire entre le taux de consultation et des variables telles que la population totale, la part des personnes âgées (75 ans et plus), la part de quelques CSP (ouvriers et retraités).

4.2.1 Taux de consultation et population totale

Table 1: Taux de consultations selon la taille de la commune

taille_commune	Taux de consultations
Grande (> 8974)	1.526810
Moyenne (4849 - 8974)	1.456356
Petite (<= 4848)	1.383861

En divisant les communes en trois groupes égaux (ou presque égaux) en fonction de la population totale, il ressort qu'en moyenne, plus la taille de la commune est importante plus le taux de consultations est élevé.

4.2.2 Taux de consultation et population âgée

Table 2: Taux de consultations selon la population âgée

population_agee_importante	consultations_moyennes
Non (<= 670)	1.501111
Oui (> 670)	1.410213

Les communes avec une population âgée importante (communes dont la population âgée de 75 ans ou plus est supérieure à la médiane) ont en moyenne un taux de consultations plus faible.

4.2.3 Taux de consultation et CSP

Aucune catégorie ne semble montrer une relation linéaire évidente avec le taux de visite. Par ailleurs, pour toutes les catégories socio-professionnelles, la majorité des communes se situent dans une plage de proportions faibles, ce qui limite la variabilité observable dans les relations. Une analyse statistique supplémentaire, comme le calcul de corrélations, serait nécessaire pour confirmer ou infirmer les relations observées visuellement.

4.2.4 Analyse de corrélation

Les résultats de la corrélation de Pearson sont consignés dans le tableau suivant :

Table 3: Corrélations de Pearson entre le taux de consultation et les autres variables

	Variable	Correlation	P_value	Significatif
cor	population_municipale_2021_x	0.0765022	0.0000118	Oui
cor1	part_des_pers_agees_de_75_ans_ou_2021	-0.6258560	0.0000000	Oui

	Variable	Correlation	P_value	Significatif
cor2	population_de_15_ans_ou_selon_la_csp_2021_retraites	-0.0285517	0.1024362	Non
cor3	population_de_15_ans_ou_selon_la_csp_2021_ouvriers	0.1077559	0.0000000	Oui

Les résultats nous montrent que le taux de consultation est positivement corrélé à la population ainsi qu'à celle de plus de 15 ans. Cependant la corrélation est faible. Par ailleurs, la corrélation est négative avec la part des personnes âgées de plus de 75 ans. Cela dit, plus la part des plus de 75 ans augmente moins est le taux de consultations dans une commune. Cela peut vouloir dire que les personnes de plus de 75 ans sont ceux qui ne se consultent pas assez.

4.3 Autocorrélation

L'autocorrélation spatiale est une mesure essentielle pour analyser la dépendance entre des observations géographiques. Dans notre étude nos données sont des données portant sur des communes. Ainsi il peut exister une dépendance entre nos taux de consultations du fait de la proximité des communes ou de l'appartenance à un même département ou région. Ainsi nous allons mesurer cette dépendance en évaluant l'autocorrélation spatiale. Dans ce contexte, **l'indice de Moran** est largement utilisé pour quantifier cette dépendance en fournissant une mesure globale de l'autocorrélation spatiale.

4.3.1 Définition de l'indice de Moran

L'indice de Moran (I) évalue la similitude des valeurs d'une variable entre différentes entités géographiques (par exemple, des communes) en fonction de leur proximité spatiale. Il se base sur la matrice de poids spatiale (W), qui définit les relations entre ces entités.

4.3.2 Formule de l'indice de Moran

La formule mathématique de l'indice de Moran est la suivante :

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Où :

- n : Nombre total d'entités spatiales (Ici, le nombre de communes).
- x_i, x_j : Valeurs observées de la variable pour les entités i et j (Ici le taux de consultations)
- \bar{x} : Moyenne de la variable x .
- w_{ij} : Poids spatial définissant la relation entre i et j .

La matrice de W peut être construite sur la base du voisinage entre les deux communes ou soit de la distance entre les deux communes. Dans le premier cas alors $w_{ij} = 1$ si i et j sont voisins et $w_{ij} = 0$ sinon. Dans le second cas $w_{ij} = d_{ij}$. Nous allons dans notre cas utiliser une matrice de poids basée sur la distance, notamment celle d'Haversine.

4.3.3 Matrice de poids basée sur la distance de Haversine

La distance de Haversine est une mesure de la distance entre deux points sur une sphère, basée sur leurs coordonnées géographiques (*latitude et longitude*). Elle est particulièrement utile pour les données géographiques projetées sur une surface sphérique, comme la Terre.

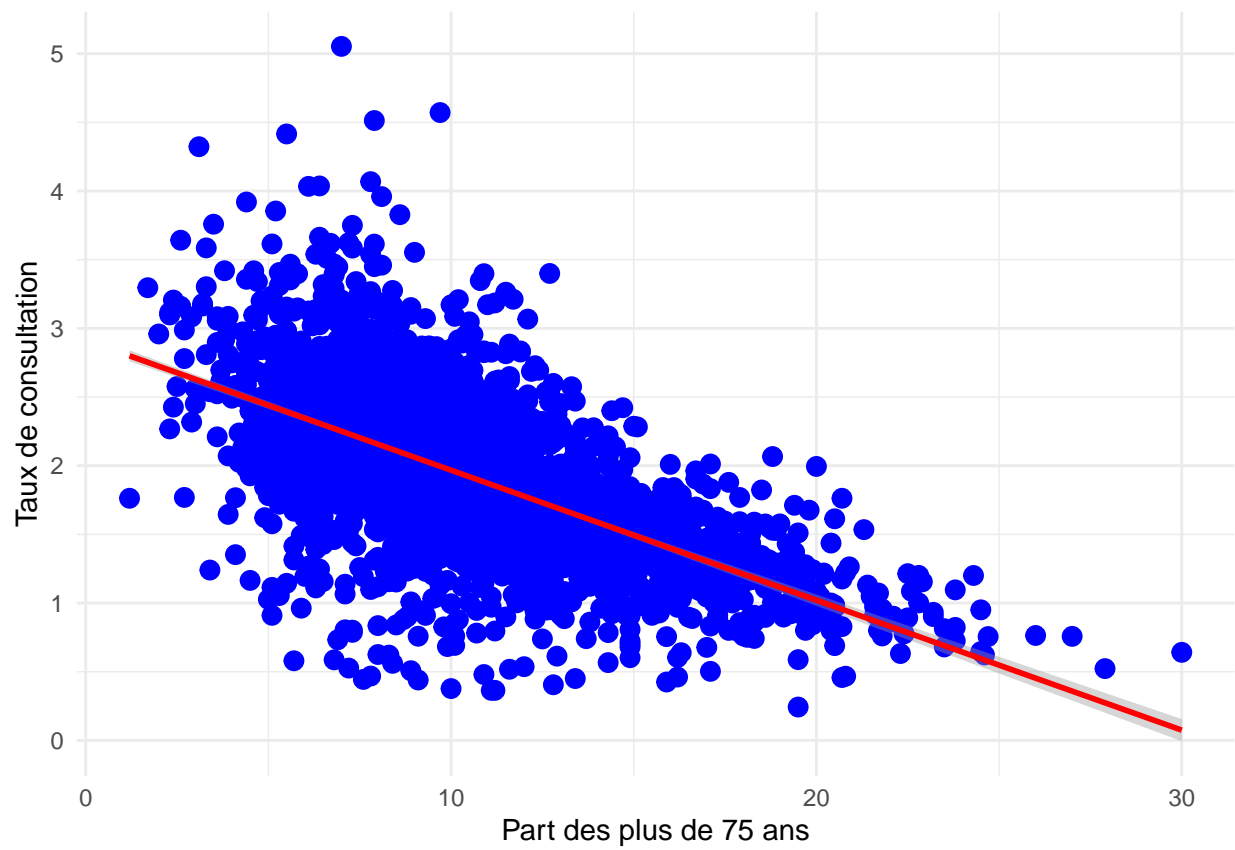


Figure 4: Relation entre taux de consultations et part des plus de 75 ans

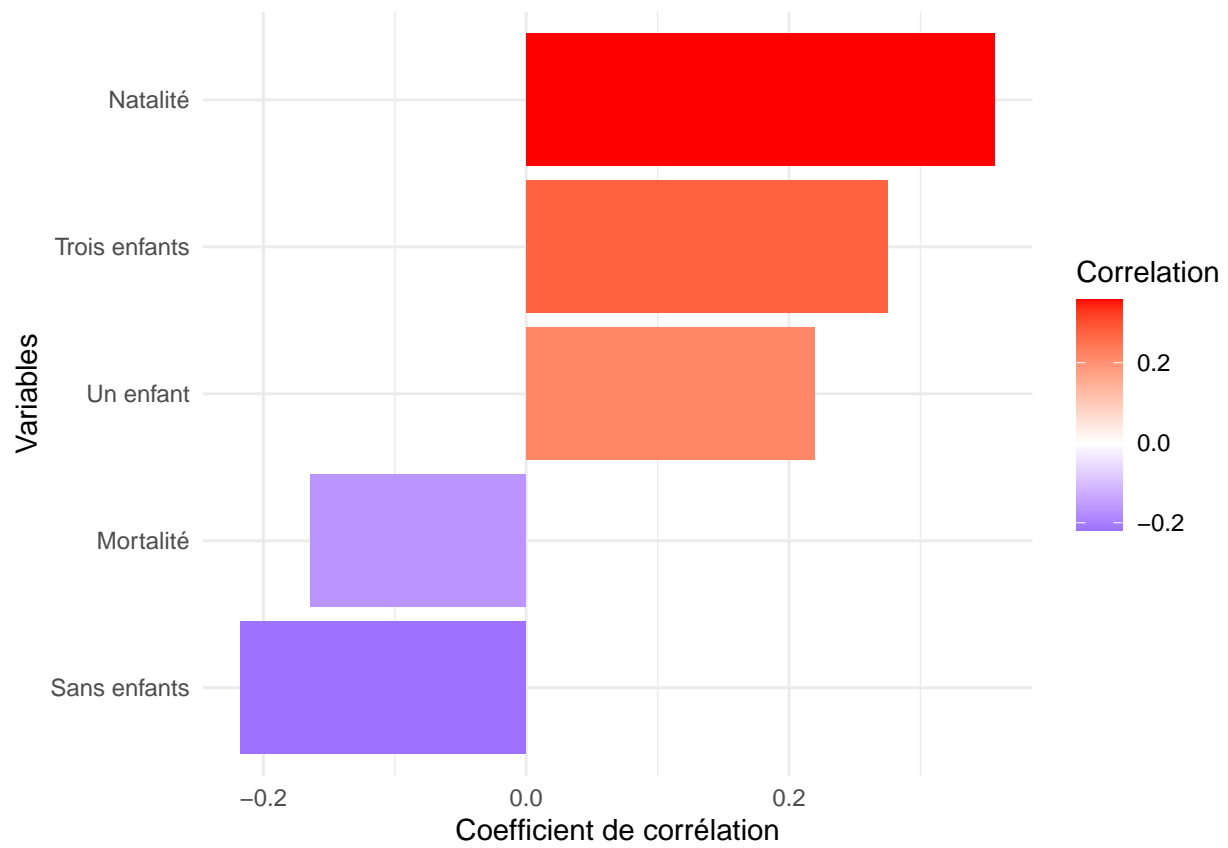


Figure 5: Corrélations entre le nombre de visite et quelques variables

4.3.4 Formule de la distance de Haversine

Si l'on considère deux points (i) et (j), la distance (d_{ij}) entre ces deux points sur la surface d'une sphère de rayon (r) est donnée par :

$$d_{ij} = 2r \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_j - \phi_i}{2} \right) + \cos(\phi_i) \cos(\phi_j) \sin^2 \left(\frac{\lambda_j - \lambda_i}{2} \right)} \right)$$

Où :

- r : Rayon de la Terre (environ 6371 km).
- ϕ_i, ϕ_j : Latitudes des points i et j (en radians).
- λ_i, λ_j : Longitudes des points i et j (en radians). Après calcul nous avons ces statistiques sur nos distances.

Une visualisation de la densité de nos distance nous donne ceci, indiquant une forte asymétrie à gauche de la distribution. En d'autres termes, les communes étudiées sont assez rapprochées les unes des autres pour la plupart.

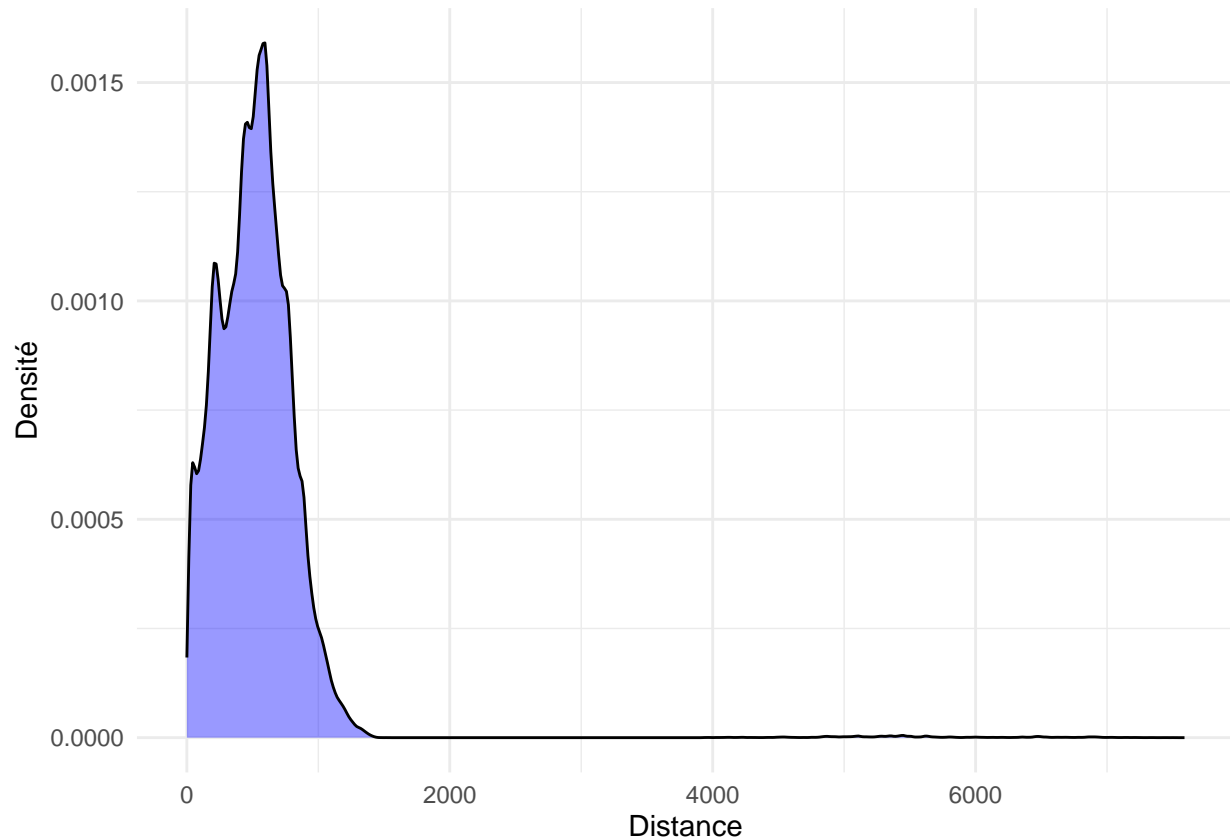


Figure 6: Densité des distances

4.3.5 Construction de la matrice de poids

Pour construire la matrice de poids, nous avons alors suivi ces étapes.

1. Calculer les distances de Haversine entre chaque paire d'entités.
2. Définir un seuil de distance maximale (d_{max}) :
 - Si $d_{ij} < d_{max}$, $w_{ij} = \frac{1}{d_{ij}}$.
 - Sinon, $w_{ij} = 0$.
3. Normaliser les poids pour que chaque ligne de la matrice ait une somme égale à 1 :

$$w_{ij}^{norm} = \frac{w_{ij}}{\sum_j w_{ij}}.$$

Table 4: Résultats du test de Moran

	x
Moran I statistic	0.1275832
Expectation	-0.0003056
Variance	0.0000029

Ainsi dans notre étude, nous avons trouvé un indice de Moran égale à 0.1275832. Le test nous a permis d'obtenir une p-value de 0. Ce qui permet de conclure qu'il y a effectivement une autocorrélation positive et significative entre les communes selon leur taux de consultations.

Références

- Statistique Canada. Fréquence des consultations médicales et facteurs sociodémographiques. 2022.
- Office fédéral de la santé publique. Santé des femmes et accès aux soins en suisse. 2024.
- Irdes. Inégalités spatiales d'accessibilité aux soins médicaux. 2020.
- BVS Santé. Facteurs influençant l'accès aux soins médicaux en milieu défavorisé. 2023.

ANNEXES

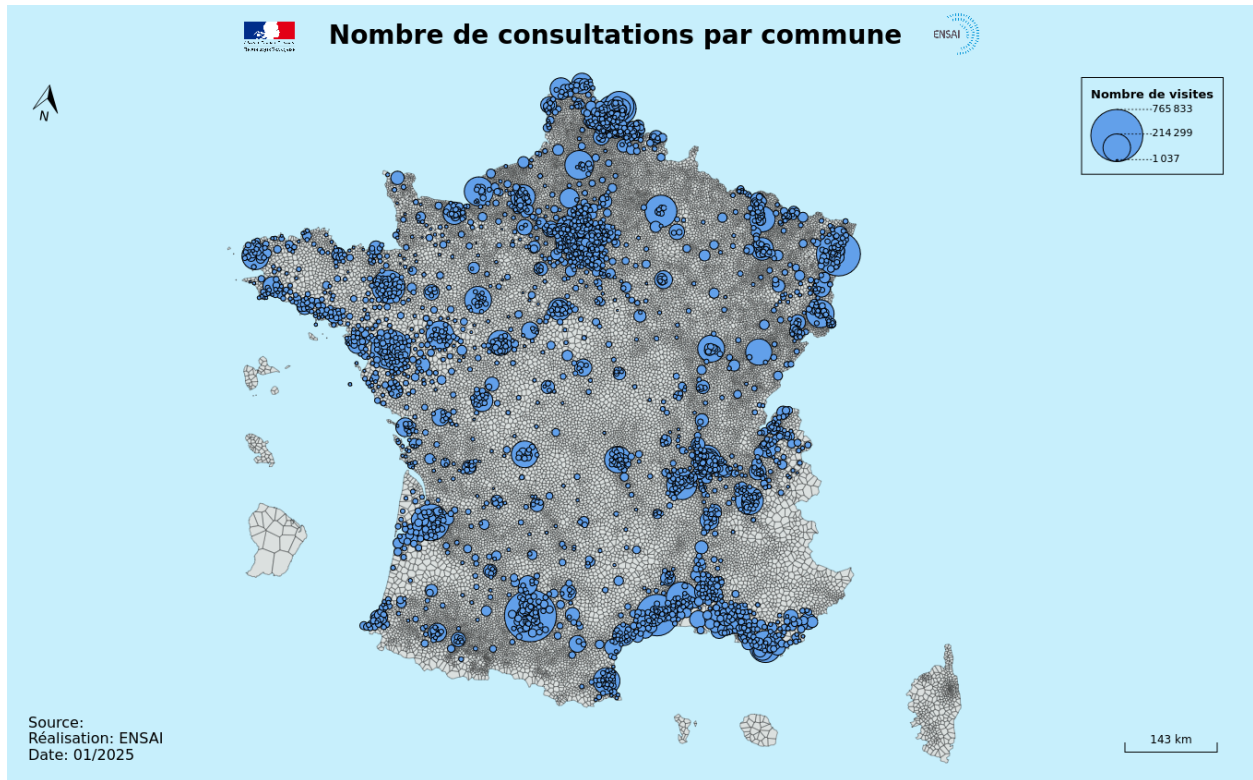


Figure 7: Carte du nombre de consultations par commune

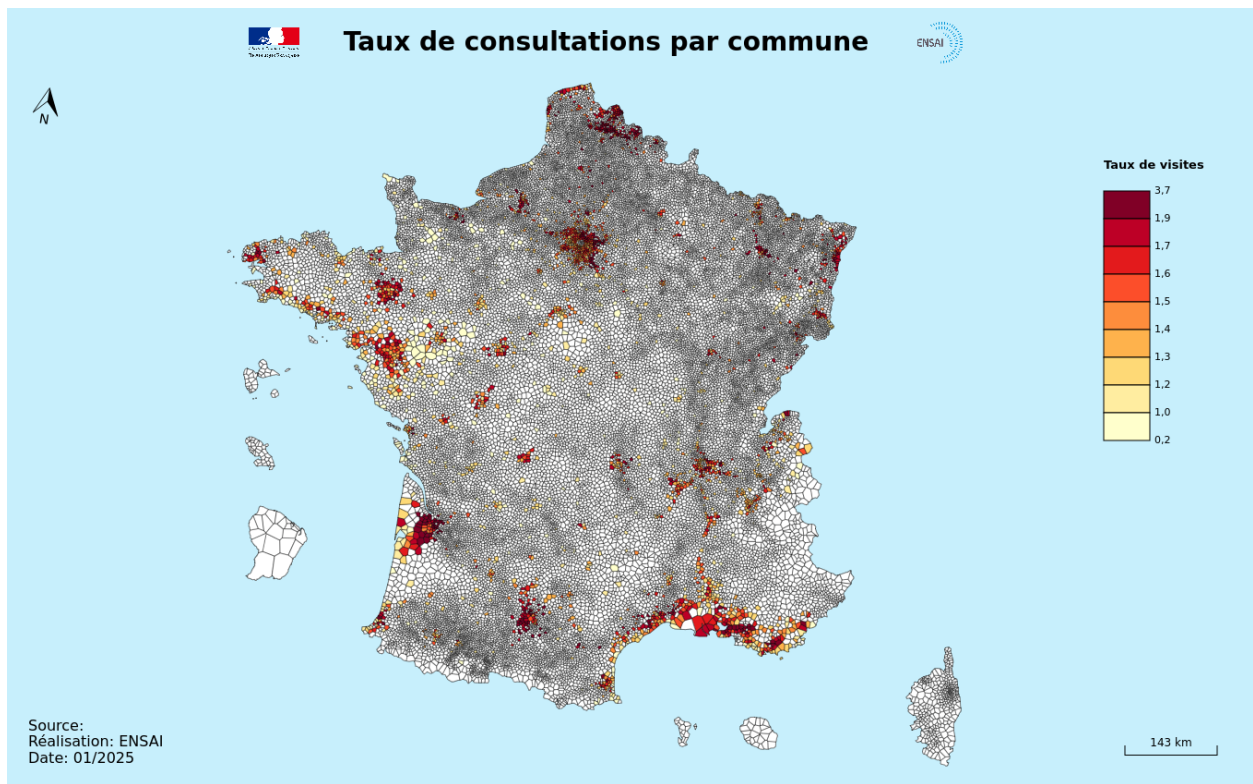


Figure 8: Carte du taux de consultations par commune

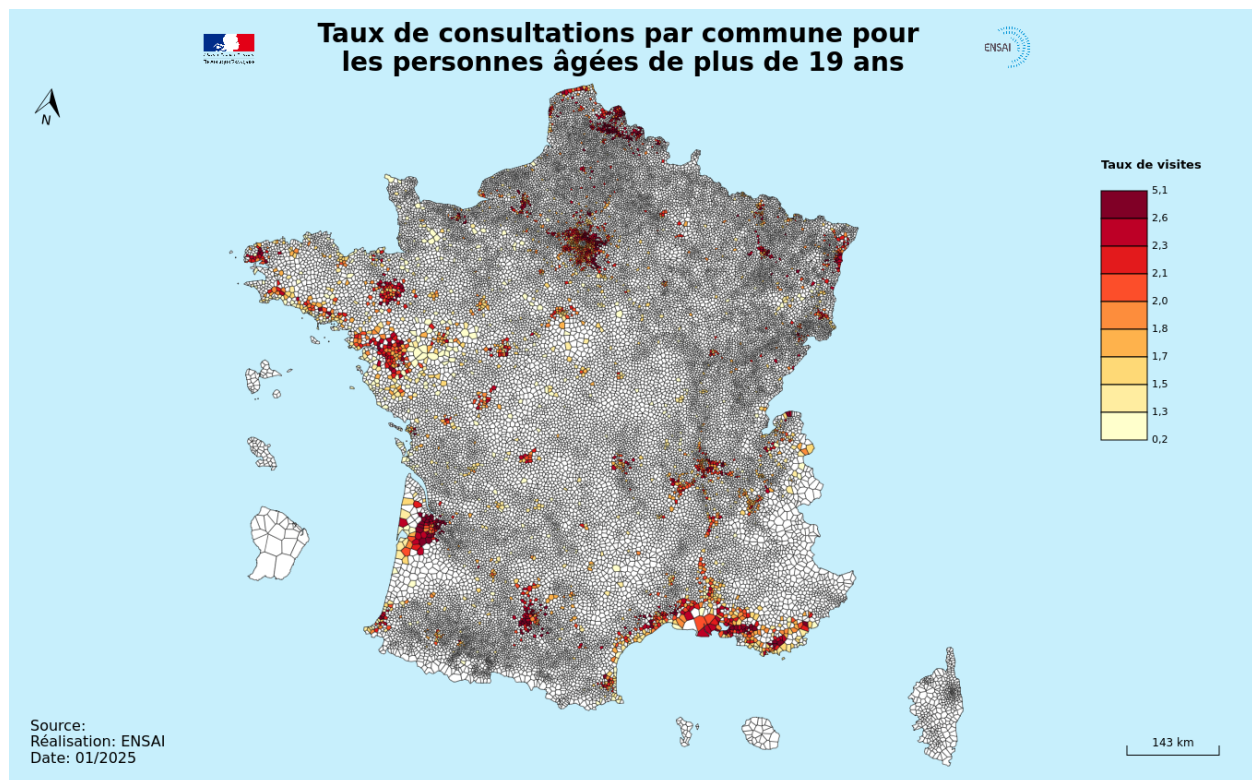


Figure 9: Carte du taux de consultations par commune pour les plus de 19 ans