Sommaire

Liste des Tableaux

Liste des Figures

1 Introduction

2 Présentation du contexte

2.1 Intérêt de l'étude

2.2 Cadre conceptuel de l'étude

2.3 Présentation des données

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

3 Méthodologie

3.1 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.

3.2 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.

3.2.1 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 :

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

3.2.2 Régression de Poisson

Utilisée pour modéliser des données de comptage ($y \sim Pois(\lambda)$), 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 $E[y] = \lambda$.

3.3 Modèles Linéaires Mixtes

Ces modèles ajoutent des termes d'effets aléatoires **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).

3.4 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).

3.5 Prédictions et Simulations

Les GLMM permettent deux types de prédictions :
- **Conditionnelles** : Basées sur les effets aléatoires spécifiques (**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 des résultats

4.1 Analyse descriptive

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

4.1.1 Analyse univariée

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1037    5993    9127   19130   17290  765833
```

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 éliminer cet effet en calculant le taux de consultations qui n'est autre que le nombre de consultations moyennes par personnes.

4.1.2 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 consulation et des variables telles que la population totale, la part des personnes agées (75 ans et plus), la part de quelques CSP (ouvriers et retraités).

4.1.2.1 Taux de consultation et population totale

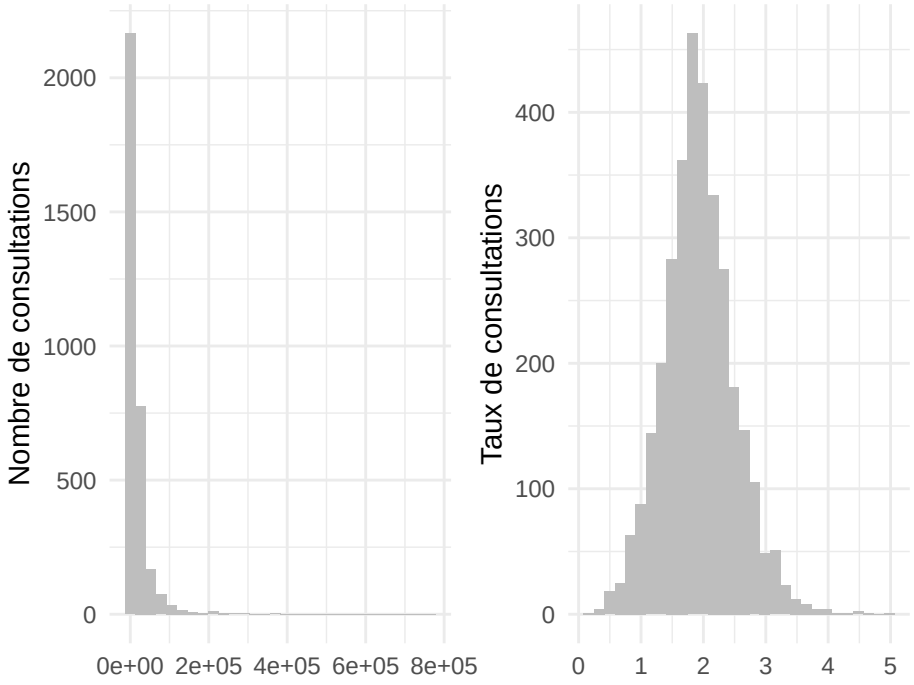

Figure 1: Répartition du nombre et du taux de consultations

```
## # A tibble: 3 x 2
##   taille_commune consultations_moyennes
##            <int>                  <dbl>
## 1              1                   1.38
## 2              2                   1.46
## 3              3                   1.53
```

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 consulations est élevé.

4.1.2.2 Taux de consultation et population âgée

```
## # A tibble: 2 x 2
##   grande_population_agee consultations_moyennes
##   <chr>                                  <dbl>
## 1 Non                                     1.50
## 2 Oui                                     1.41
```

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.

4.1.2.3 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.

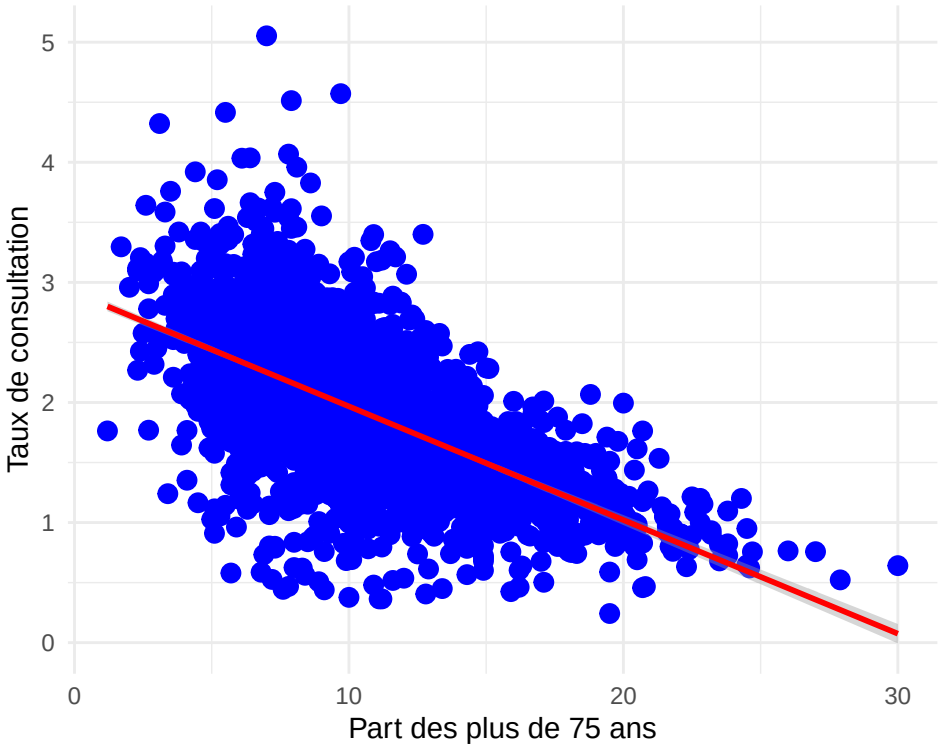

Figure 2: Relation entre taux de consultations et part des plus de 75 ans

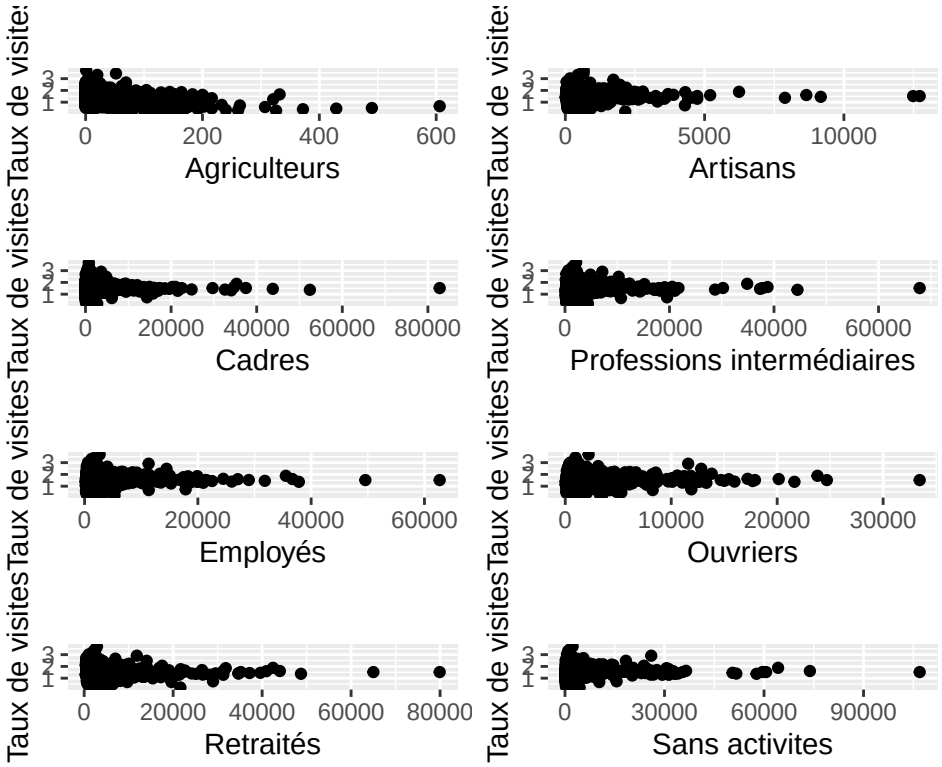

Figure 3: Relations entre le taux de consultatiosn et certaines le nombre de certaines catégories socioprofessionnelle

4.1.2.4 Analyse de corrélation Les résultats de la corrélation de Pearson sont consignées dans le tableau suivant :

Table 1: 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
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 agées 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.

4.1.3 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.

4.1.3.1 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.

4.1.3.2 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 constuit 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} $w_{ij} = 1$ 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.

4.1.3.3 Matrice de poids basée sur la distance de Haversine

4.1.3.3.1 Définition de 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.

4.1.3.4 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.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   312.8   515.7   531.6   688.7  7589.3
```

Une visualtion de la densité de nos distance nous donne ceci.

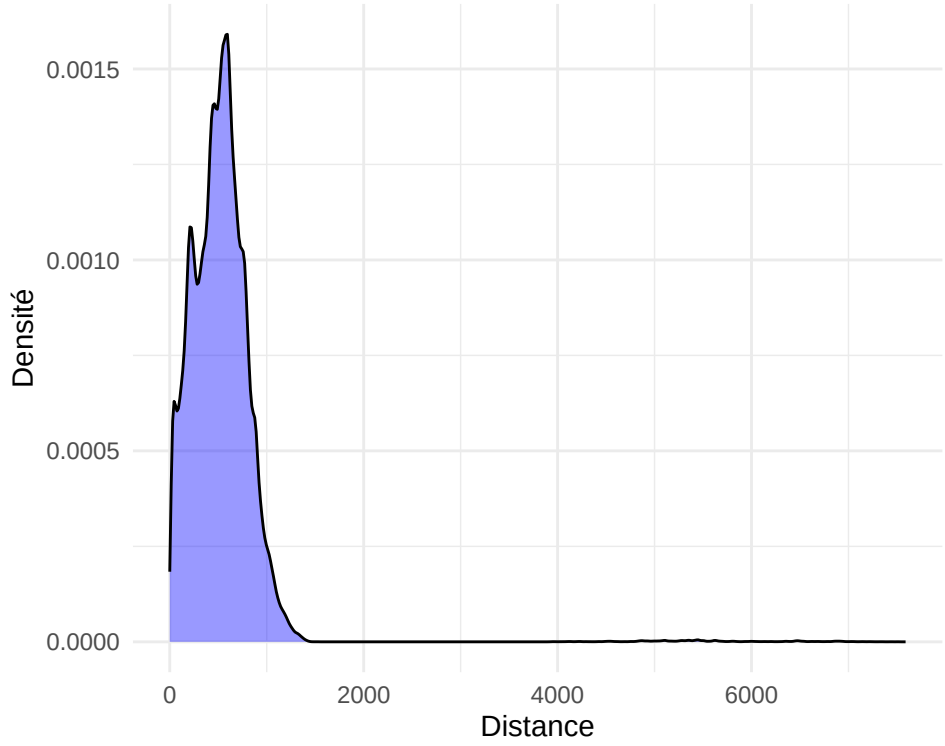

Figure 4: Densité des distances

4.1.3.5 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 2: 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 permi 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.

5 Discussion

6 Conclusion

7 Références bibliographiques

8 Annexes

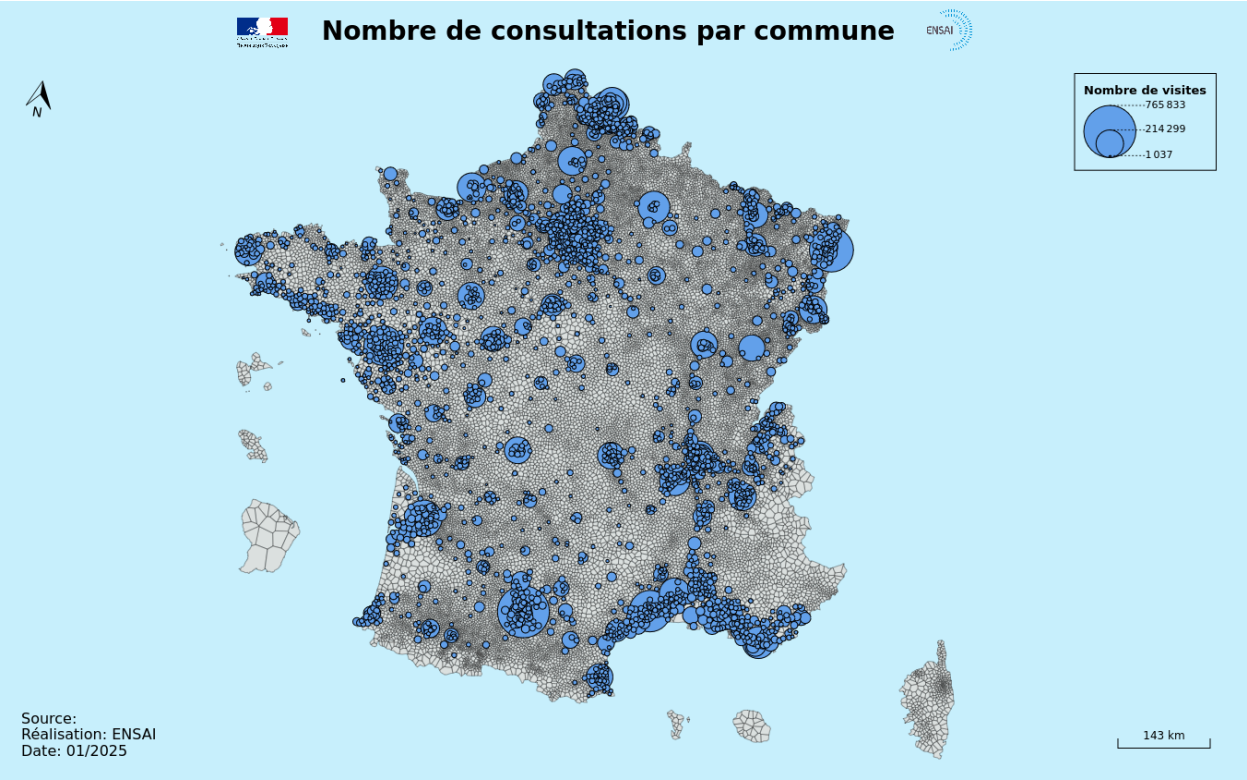

Figure 5: Carte du nombre de consultations par commune

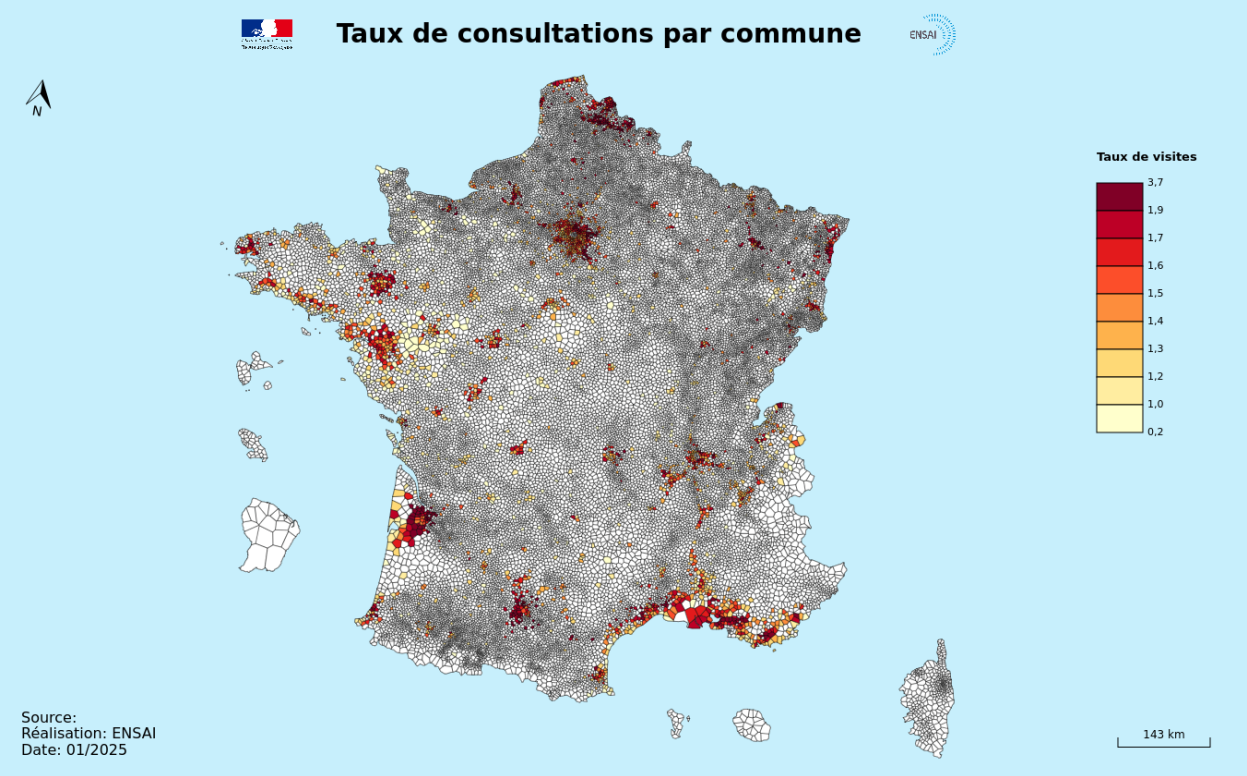

Figure 6: Carte du taux de consultations par commune

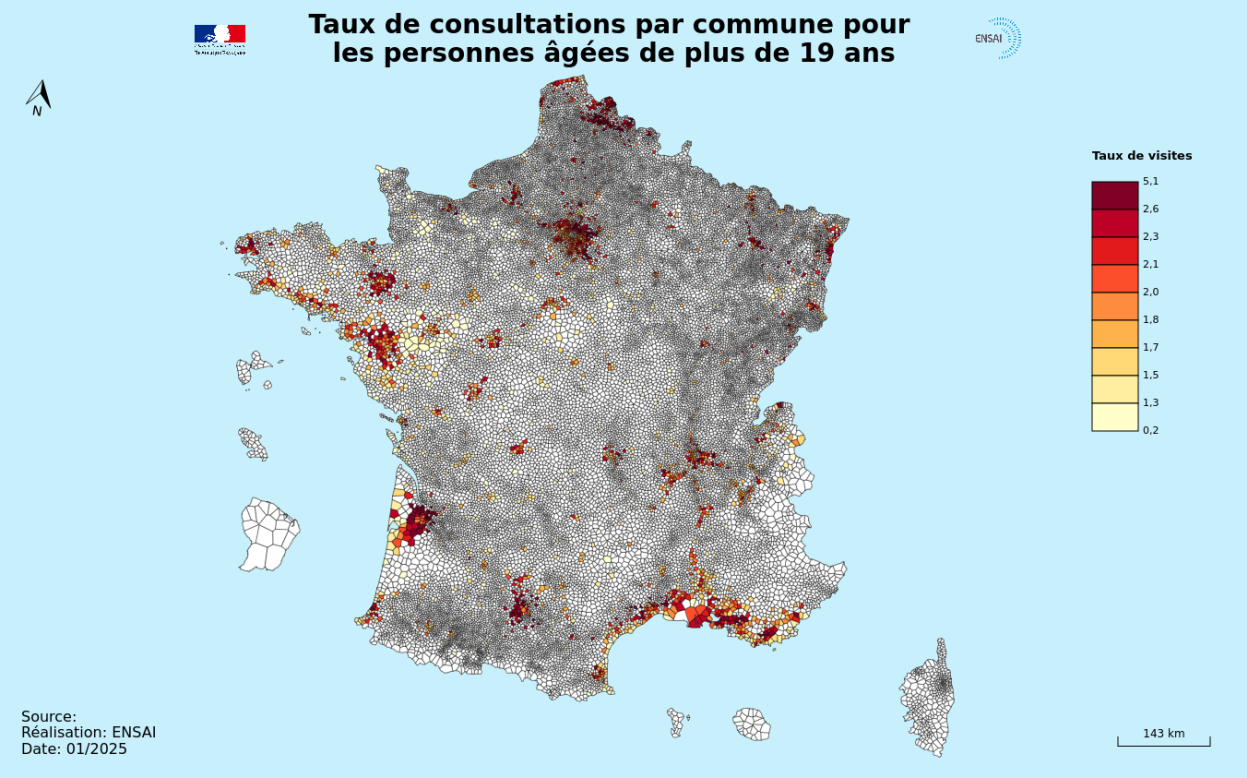

Figure 7: Carte du taux de consultations par commune pour les plus de 19 ans