

Master 2 : ECAP

**Econométrie spatiale**

---

# **Taux d'inscription sur les listes électorales en 2022**

---

Auteurs :  
JAMIN Mathilde  
ONNO Lilou

Sous la direction de Madame TRAVERS

---

Février 2025

# Sommaire

<b>Introduction</b>	<b>3</b>
<b>I Partie économique</b>	<b>4</b>
<b>II Analyse exploratoire</b>	<b>8</b>
<b>III Modélisation par l'économétrie spatiale</b>	<b>20</b>
<b>IV Analyse cartographique complémentaire sous QGIS</b>	<b>37</b>
<b>Conclusion</b>	<b>39</b>

## **RÉSUMÉ**

Une analyse approfondie a été réalisée pour comprendre les facteurs influençant le taux d'inscription sur les listes électorales en France, en examinant des variables démographiques et socio-économiques au niveau départemental. Les principaux facteurs incluaient l'âge, le niveau d'éducation, la part de la population inactive, et active détenant au moins un bac+3, la part des résidences principales, la part des propriétaires et la part des ménages composés d'une personne seule. La cartographie et l'analyse exploratoire ont révélé des disparités géographiques, notamment des taux d'inscription plus élevés dans les départements plus âgés et incluant des personnes actives. Les relations spatiales ont été modélisées à l'aide de matrices de poids spatiaux, et l'analyse a révélé une autocorrélation spatiale positive significative. Les modèles d'économétrie spatiale ont montré que des variables telles que l'âge moyen et la structure familiale affectaient directement l'inscription, tandis que la densité de population et les services en dépendaient indirectement. Des disparités importantes dans les comportements d'inscription ont été observées, soulignant l'importance d'adapter les politiques locales. Des pistes d'amélioration, comme l'introduction de variables d'interaction ou des modèles géostatistiques, ont été suggérées pour approfondir l'analyse.

## **ABSTRACT**

An in-depth analysis was conducted to understand the factors influencing voter registration rates in France, focusing on demographic and socio-economic variables at the departmental level. Key factors included age, education level, homeownership rate, and family structure. Mapping and exploratory analysis revealed geographical disparities, with higher registration rates in older, more educated departments. Spatial relationships were modeled using spatial weight matrices, and the analysis found a significant positive spatial autocorrelation. Spatial econometrics models showed that variables such as average age and family structure had direct effects on registration, while population density and services were more influenced by indirect effects from neighboring departments. Significant disparities in registration behavior were observed, highlighting the need for localized policy adaptations. Suggestions for improvement, such as incorporating interaction variables or using geostatistical models, were proposed to deepen the analysis.

# Introduction

Les élections représentent un moment clé de la vie démocratique d'un pays, et la participation électorale constitue un indicateur fondamental de l'engagement civique des citoyens. En France, au 9 mars 2022, 48 803 175 électeurs étaient inscrits sur les listes électorales, selon l'INSEE.<sup>[1]</sup> Toutefois, ce chiffre global peut montrer des disparités importantes à l'échelle locale, avec des taux d'inscription variés selon les régions, les départements et même les communes. Ces différences de taux d'inscription peuvent être attribuées à une combinaison complexe de facteurs socio-économiques, démographiques et territoriaux.

L'objectif de cette étude est de comprendre ces écarts d'inscription en analysant les déterminants socio-économiques et géographiques du taux d'inscription électorale. Ainsi, l'étude s'appuie sur des variables telles que l'âge moyen de la population, la part des jeunes en formation et la structure du marché du travail, notamment la part de la population active détentrice d'un diplôme Bac+3 ou plus. Ces facteurs sont susceptibles d'influencer l'aptitude des citoyens à s'inscrire sur les listes électorales.

De plus, d'autres éléments structurels, liés aux conditions de logement et à la composition des ménages, jouent également un rôle crucial. Par exemple, la part des résidences principales, le taux de propriétaires et la proportion de ménages composés d'une personne seule sont des facteurs qui peuvent affecter la participation électorale en modifiant les dynamiques sociales et économiques à l'échelle locale. Les territoires où ces variables présentent des caractéristiques particulières peuvent ainsi connaître des comportements différents en matière d'inscription électorale.

Pour ce faire, nous débuterons notre analyse par une approche économique du sujet en expliquant le choix des variables sélectionnées. Par la suite, nous effectuerons une analyse exploratoire. Enfin, nous exposerons les différentes modélisations et procéderons à la sélection du modèle le plus performant.

# Chapitre I

## Partie économique

### I.1 Cadre économique

#### I.1.1 Taux d'inscrit sur les listes électorales

Le taux d'inscription sur les listes électorales représente le pourcentage de la population éligible inscrite sur les listes électorales dans une circonscription donnée. <sup>[2]</sup> Un taux élevé peut être associé à des populations stables et ancrées localement, souvent caractérisées par un âge moyen plus élevé, une part plus importante de personne propriétaire et résident dans leur logement, et ainsi une plus grande implication dans la vie civique.<sup>[3]</sup>

En revanche, un faible taux d'inscription peut être indicatif d'une population plus mobile, plus jeune, ou moins intégrée localement. Par exemple, dans les grandes villes ou les zones universitaires, les taux d'inscription peuvent être plus faibles en raison d'une population plus transitoire, <sup>[4]</sup> notamment composée d'étudiants et de jeunes actifs.<sup>[3]</sup>

L'analyse spatiale du taux d'inscription peut donc aider à identifier des régions où la population est plus susceptible de rester sur le long terme, ce qui pourrait avoir des répercussions immédiate sur les politiques publiques locales. De plus, en couplant ce taux avec des variables comme l'âge moyen ou la part des ménages composés d'une personne seule, nous pouvons obtenir une meilleure analyse de la population pour pouvoir mieux cibler par exemple des politiques publiques.

### **I.1.2 Age moyen**

L'âge moyen de la population est une mesure démographique importante qui reflète la structure d'une population en termes de répartition par âge. <sup>[5]</sup> En France, le phénomène de vieillissement de la population est bien documenté, <sup>[6]</sup> avec une augmentation de l'espérance de vie et une baisse de la natalité. Ce vieillissement entraîne une élévation progressive de l'âge moyen à l'échelle nationale et locale.

Une hypothèse intéressante à envisager est qu'un âge moyen plus élevé pourrait être lié à un taux d'inscription plus élevé sur les listes électorales. Les personnes âgées, souvent plus stables géographiquement et davantage impliquées dans la vie civique, ont tendance à s'inscrire et à participer plus activement aux élections, en raison de préoccupations spécifiques comme l'accès aux soins de santé ou aux logements adaptés. <sup>[7]</sup> En revanche, les jeunes adultes, notamment les étudiants, étant plus mobiles, pourraient avoir un taux d'inscription plus faible. Bien qu'il soit difficile d'établir un lien de cause à effet direct, il est plausible que le vieillissement de la population puisse conduire à une augmentation du taux d'inscription.

### **I.1.3 Part de la population inactive**

La part de la population inactive désigne les individus qui ne participent pas activement au marché du travail, tels que les élèves, les étudiants, les retraités, et les personnes au foyer. Elles n'ont donc ni un emploi et ne sont ni au chômage. <sup>[8]</sup> Cette catégorie représente souvent une proportion significative de la population dans les zones rurales ou dans des départements où l'emploi est moins accessible, ou des villes qui compte un grand nombres d'étudiants. L'inactivité peut aussi être le reflet de réalités socio-économiques ou démographiques spécifiques, comme un vieillissement de la population ou des choix de vie liés à l'éducation. <sup>[9]</sup>

Dans notre cas, cela représente seulement les élèves ainsi que les étudiants.

### **I.1.4 Part de la population active détenant au moins un Bac+3**

Le niveau d'éducation est un facteur déterminant dans l'engagement civique et politique. Les personnes ayant un niveau d'éducation élevé, comme un Bac+3 ou plus, sont généralement plus enclines à s'inscrire sur les listes électorales et à participer aux élections. Selon l'étude de Verba et al., un niveau d'éducation supérieur est associé à une plus grande probabilité de participation politique, notamment à travers l'inscription et le vote.<sup>[10]</sup>

En France, une analyse menée par Blais, a également montré que les individus possédant un diplôme universitaire sont plus susceptibles de s'inscrire sur les listes électorales, en raison de leur meilleur accès à l'information et de leur compréhension accrue des enjeux politiques.<sup>[11]</sup>

### **I.1.5 Part des résidences principales**

La proportion de résidences principales reflète la stabilité des habitants au sein d'un territoire. Une forte proportion de résidences principales est souvent associée à une plus grande implication civique, y compris l'inscription sur les listes électorales. Selon Highton, les individus ayant une plus grande stabilité résidentielle sont plus enclins à s'inscrire, car ils développent un sentiment d'ancrage dans leur communauté locale.<sup>[12]</sup>

En France, des études comme celles de Gougou et Sauger ont montré que les communes ayant une proportion élevée de résidences principales présentent des taux d'inscription électorale supérieurs à la moyenne nationale. Cela s'explique par une moindre mobilité des habitants, qui facilite les démarches administratives liées à l'inscription.<sup>[13]</sup>

### **I.1.6 Part des logements occupés par leur propriétaire**

La propriété est un indicateur clé de la stabilité socio-économique et résidentielle. Les propriétaires ont tendance à s'inscrire plus facilement sur les listes électorales que les locataires, notamment en raison de leur moindre mobilité. Verba et Nie ont démontré que les individus économiquement stables sont plus susceptibles de participer à la vie politique locale.<sup>[14]</sup>

En outre, Braconnier et Dormagenont ont constaté qu'en France, les quartiers à forte proportion de propriétaires affichent une meilleure participation électorale, y compris au niveau de l'inscription. Ces résultats soulignent l'importance de considérer cette variable dans l'analyse des déterminants territoriaux.<sup>[15]</sup>

### **I.1.7 Part des ménages composés d'une personne seule**

Les ménages composés d'une seule personne présentent des comportements politiques particuliers. Selon Campbell et al., les individus vivant seuls sont souvent moins engagés dans la vie communautaire, ce qui peut se traduire par une moindre inscription sur les listes électorales.<sup>[16]</sup>

Cependant, Putnam a montré que l'individualisation croissante de la société n'entraîne pas nécessairement un désengagement politique, car cela dépend aussi des contextes institutionnels et culturels. En France, des recherches récentes indiquent que les ménages composés d'une personne seule sont surreprésentés dans les zones urbaines, où les taux d'inscription électorale peuvent varier fortement.<sup>[17]</sup>

Cette revue de littérature met en évidence plusieurs facteurs économiques et démographiques ayant un impact sur le taux d'inscription sur les listes électorales. En particulier, des éléments tels que la stabilité résidentielle, le niveau d'éducation, la propriété du logement, l'âge moyen de la population et la composition des ménages jouent un rôle central dans l'engagement civique. Les individus ayant une plus grande stabilité géographique, une éducation supérieure, ou des liens forts avec leur communauté sont généralement plus enclins à s'inscrire et à participer aux élections et c'est ce que nous allons chercher à déterminer avec cette étude.



# Chapitre II

## Analyse exploratoire

Nous allons maintenant analyser nos données à travers des statistiques descriptives et des visualisations graphiques, telles que des histogrammes, des boxplots et des cartes. Il est important de noter que nos analyses sont réalisées à la fois sous RStudio et sous GeoDa.

### II.1 Statistiques descriptives

Table II.1 – Statistiques descriptives des variables

Variables	Minimum	1er quartile	Médiane	Moyenne	3ème quartile	Maximum
Taux d'inscrit	38.60	64.90	71.60	69.78	75.90	92.20
Age moyen	33.80	39.20	41.30	41.41	43.65	50.20
Part de la pop. inactive	4.4	6.1	7.1	7.71	8.6	18.9
Part de la pop. active avec un bac+3 et +	9.6	16.6	20.6	25.13	29.05	74.10
Part des résidences principales	33.7	80.75	86.70	83.60	90.75	95.40
Part des logements occupé par leur propriétaire	23.3	51.55	61.4	58.31	67.30	77.60
Part des ménages composés d'une personne seule	24.10	32	35.3	36.43	39.45	58.60

Les données portent sur 96 départements. Nous pouvons voir à travers ce tableau, que le taux d'inscription varie de 38.6% à 92.2%, avec une moyenne de 69.78% et une médiane de 71.6%. Ces résultats indiquent que la majorité des régions présentent un taux d'inscription relativement élevé, bien que certaines se situent nettement en dessous de la moyenne. Quant à l'âge moyen, celui-ci oscille entre 33.8 ans et 50.2 ans, avec une moyenne de 41.41 ans, ce qui montre une population relativement adulte et stable d'une zone à l'autre.

La part de la population inactive reste globalement faible, avec une moyenne de 7.71% et une médiane de 7.1%, bien que certaines zones enregistrent des taux d'inactivité très bas (4.4), d'autres atteignent des niveaux plus élevés (18.9). Cependant, la faible différence entre le 1er et le 3e quartile indique que la majorité des zones étudiées présentent des niveaux d'inactivité modérés.

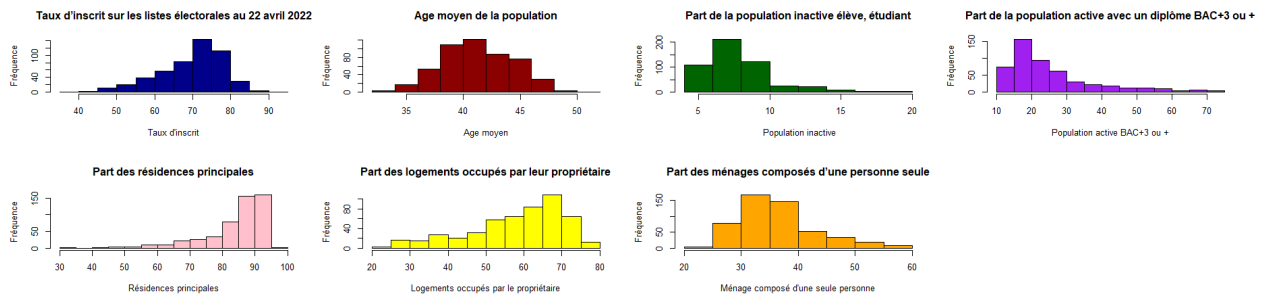
Concernant la part de la population active ayant un diplôme de niveau bac+3 et plus, les résultats montrent une grande hétérogénéité selon les zones : celle-ci varie de 9.6 à 74.1%, avec une moyenne de 25.13%. Cette disparité importante pourrait refléter des différences marquées en termes de niveau d'éducation ou d'accès à la formation supérieure selon les territoires.

En ce qui concerne les logements occupés par leurs propriétaires, la répartition est plus variée. La proportion de propriétaires oscille entre 23.3% et 77.6%, avec une moyenne de 58.31%. Ces résultats traduisent des situations contrastées, certaines zones comptant une majorité de locataires tandis que d'autres enregistrent une forte proportion de propriétaires. Cela peut se traduire aussi par le niveau de revenu qui peut différer d'une région à une autre.

Enfin, la part des ménages composés d'une personne seule varie de 24.1% à 58.6%, avec une moyenne de 36.43%. La faible dispersion observée indique que ce type de ménage est relativement courant et réparti de manière assez homogène dans l'ensemble des régions de France. Plusieurs facteurs peuvent expliquer ce phénomène, notamment l'évolution des modes de vie, le vieillissement démographique dans certaines régions, ainsi qu'une plus grande indépendance des individus, particulièrement dans les zones urbaines où l'accès au logement est plus facile.

## II.2 Distributions

Figure II.1 – Histogramme



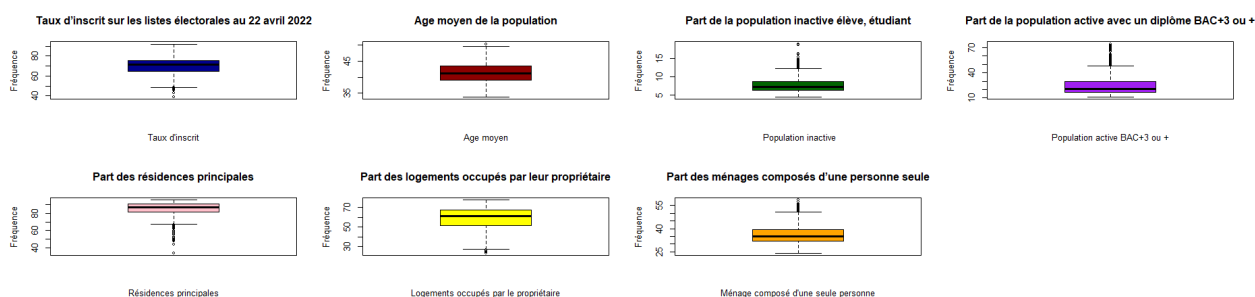
La figure II.1 illustre les histogrammes des différentes variables. Nous observons une concentration marquée du taux d'inscription entre 70 et 80%, ainsi qu'un âge moyen qui se situe autour de 40 ans. La part de la population inactive est faible, comme mentionné précédemment dans la section statistique descriptive, se situant entre 5 et 10%.

De plus, les distributions de la part des résidences principales et de la part des logements occupés par leurs propriétaires présentent des valeurs élevées, respectivement de 90% et de 70%. Enfin, la part des ménages composés d'une seule personne se situe en moyenne entre 30 et 40%.

## II.3 Outliers

Désormais, nous nous intéressons aux points atypiques que nous retrouvons ci-dessous.

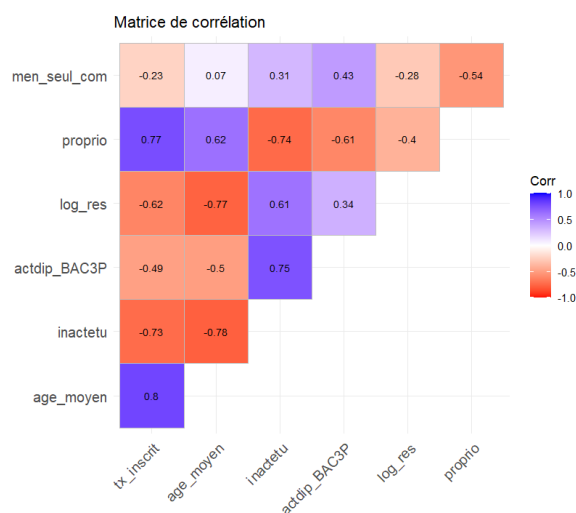
Figure II.2 – Boxplot



Toutes nos variables présentent des valeurs atypiques, seule la variable concernant l'âge moyen n'en compte qu'une seule. Nous avons décidé de conserver les départements présentant des points atypiques car leur exclusion serait incompatible avec une analyse spatiale. Ces valeurs, bien que potentiellement inhabituelles, peuvent refléter des caractéristiques spécifiques ou des contextes particuliers propres à ces départements. Les éliminer pourrait entraîner un biais dans notre analyse en écartant des informations potentiellement significatives.

## II.4 Corrélation

Figure II.3 – Matrice de corrélation



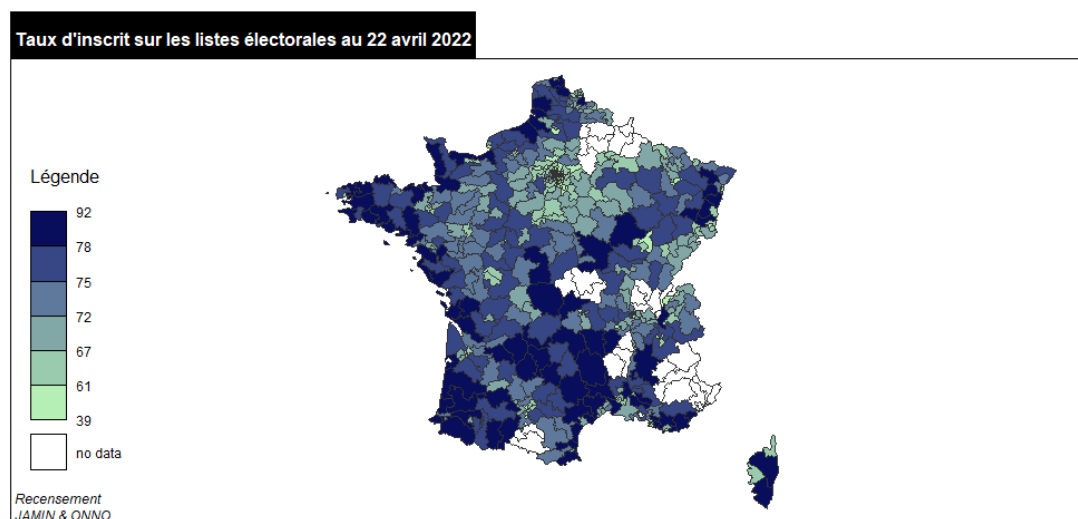
Plusieurs corrélations importantes ressortent sur cette matrice. Tout d'abord, nous observons deux corrélations positives avec la variable du taux d'inscrit sur les listes électorales. La première concerne la part des propriétaires (proprio), avec une corrélation de 0.77. Cela peut s'expliquer par le fait que les personnes propriétaires, généralement plus stables sur le plan résidentiel et financier, sont plus enclines à s'inscrire sur les listes électorales. Cette stabilité, en particulier géographique, favorise probablement un engagement civique plus fort. La seconde corrélation positive concerne l'âge moyen, avec un coefficient de 0.8. Comme évoqué dans les sections précédentes, il est plausible que des personnes plus âgées, souvent plus ancrées dans leur territoire et leurs habitudes, soient plus susceptibles de s'inscrire sur les listes électorales, ce qui explique cette association positive.

D'autre part, nous pouvons observer deux corrélations importantes et intéressantes avec la variable de la population inactive. La première, avec la part des propriétaires, présente un coefficient de -0.74. Cela suggère qu'une plus grande proportion de propriétaires et de résidents permanents est associée à une plus faible proportion de population inactive. En effet, les propriétaires ont souvent un statut socio-économique plus stable, ce qui les rend moins enclins à être inactifs. La seconde corrélation concerne la part de la population active détentrice d'un diplôme de niveau Bac+3 ou plus, avec une corrélation positive de 0.75. Cette relation peut indiquer que dans les départements où une plus grande proportion de la population est titulaire d'un diplôme supérieur, la population inactive est moins représentée.

## II.5 Représentation cartographique

### II.5.1 Taux d'inscription

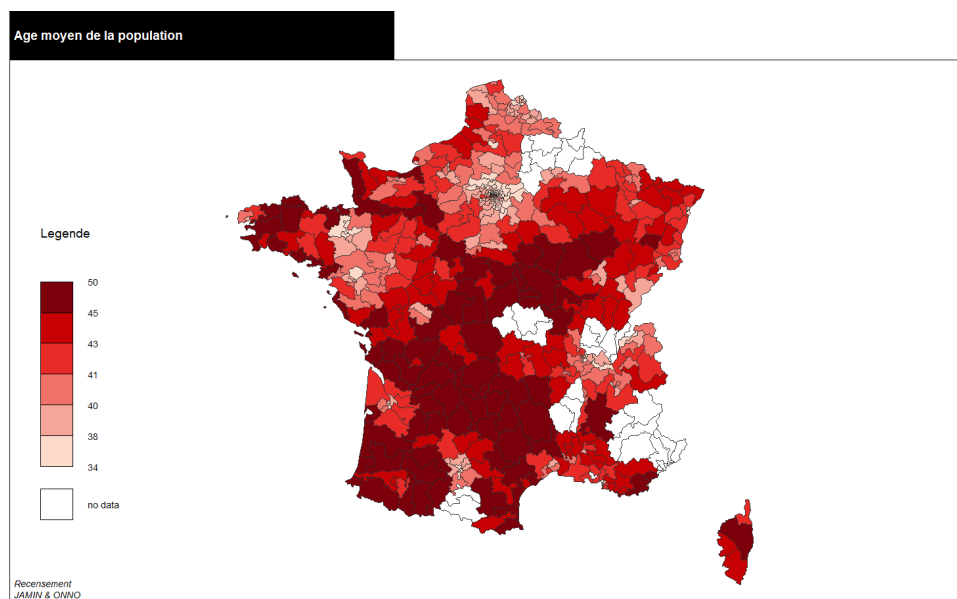
Figure II.4 – Cartographie du taux d'inscrit sur les listes électorales en 2022



Le taux d'inscription enregistré en 2022 sur les listes électorales, présente une concentration élevée principalement sur les côtés nord et ouest de la France, ainsi que les zones suivantes : Aquitaine, Limousin, l'ouest de l'Auvergne et le nord du Languedoc-Roussillon qui présentent généralement les taux d'inscription les plus élevés. Cette répartition pourrait refléter des différences socio-économiques ou démographiques, avec des populations potentiellement plus stables, plus âgées ou davantage impliquées dans les processus électoraux. Ces régions, souvent constituées de villes moins importantes, sont susceptibles d'accueillir davantage de familles. À l'inverse, Paris présente un taux d'inscription relativement faible, cela peut s'expliquer par le fait que la capitale est souvent considérée comme un centre d'attractivité pour les étudiants qui sont des groupes plus mobiles et moins enclins à s'inscrire sur les listes électorales. Cela pourrait également témoigner d'un phénomène de mobilité géographique, où la population parisienne, plus jeune et souvent plus transitoire, présente une moindre stabilité résidentielle, ce qui se traduit par un taux d'inscription plus bas.

## II.5.2 Age moyen

Figure II.5 – Cartographie de l'âge moyen de la population



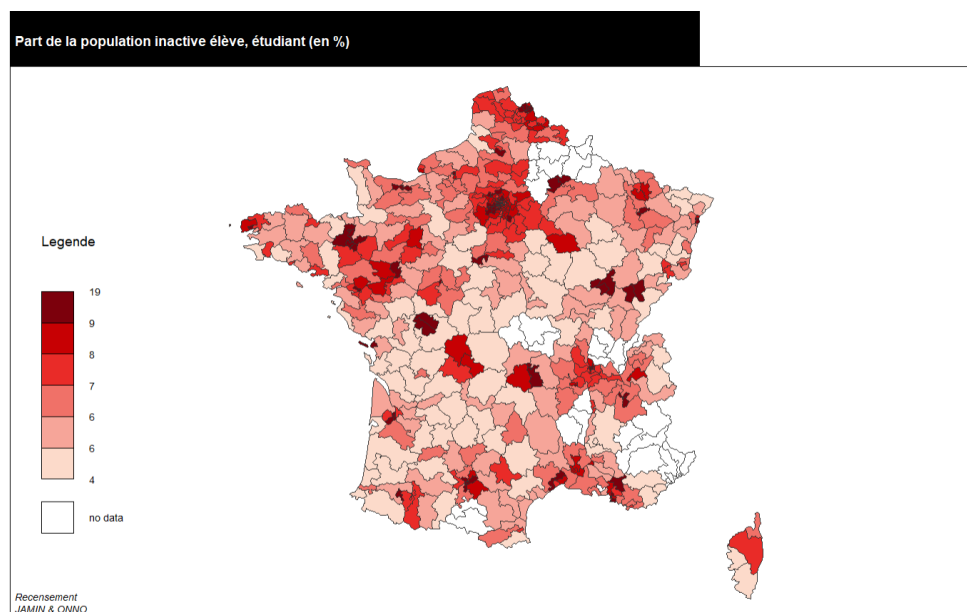
Nous observons un lien intéressant entre la cartographie du taux d'inscription et celle de l'âge moyen. Les zones d'Aquitaine, Limousin et le nord du Languedoc-Roussillon, qui présentent les taux d'inscription les plus élevés, coïncident également avec une moyenne d'âge plus élevée, avoisinant les 50 ans. Cette tendance pourrait s'expliquer par une population plus âgée et plus stable dans ces régions, ce qui favorise une participation électorale plus importante. En revanche, à Paris, la moyenne d'âge est relativement basse (environ 30 ans), ce qui pourrait être lié à une population plus jeune, moins stable géographiquement, et donc moins impliquée dans les démarches d'inscription électorale.

Dans le nord de la France, la population est généralement plus jeune, ce qui peut être attribué à la présence de nombreuses villes universitaires et de centres urbains dynamiques attirant une population étudiante ou en début de carrière. Les jeunes, souvent moins ancrés géographiquement, ont tendance à être moins nombreux à s'installer durablement dans ces régions, ce qui contribue à une moyenne d'âge plus basse.

En revanche, dans le sud de la France et en Corse, la population est plus âgée. Ces régions attirent souvent des individus en quête de qualité de vie, notamment des retraités ou des personnes cherchant à s'établir dans un environnement plus calme et ensoleillé. Ce phénomène entraîne une moyenne d'âge plus élevée, liée à une plus grande stabilité résidentielle et à une moindre mobilité géographique par rapport au nord du pays.

### II.5.3 Part de la population inactive

Figure II.6 – Part de la population inactive élève, étudiant



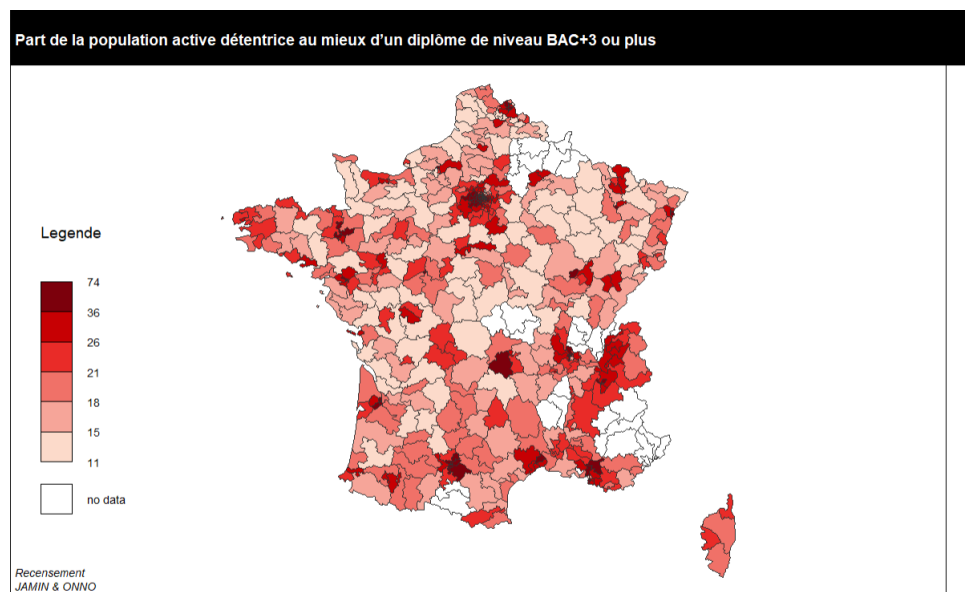
La cartographie de la variable population inactive met en évidence une concentration marquée de personnes inactives dans plusieurs zones spécifiques surtout proches des grandes villes, notamment à Paris, Nantes, ainsi qu'en Pas-de-Calais et dans le Nord. À Paris, cette forte proportion d'inactifs peut être attribuée à une combinaison de facteurs, tels que la présence de nombreuses populations jeunes, étudiantes, cela se confirme par ailleurs avec l'âge moyen qui se trouve être le plus faible dans cette région. Les régions dans le sud, notamment les zones touristiques, sont également caractérisées par une part significative de population inactive.

Lyon, étant une grande ville universitaire, présente également un taux d'inactivité notable, principalement dû à la forte proportion d'étudiants qui, bien que présents sur le marché du travail, sont souvent considérés comme inactifs en raison de leur statut académique. En effet, ces zones attirent de nombreux jeunes qui, pendant leur formation, sont moins actifs professionnellement. Ainsi, les villes universitaires et les grandes métropoles, avec une population étudiante importante, tendent à afficher des taux d'inactivité plus élevés.



## II.5.4 Part de la population active avec un bac+3 et +

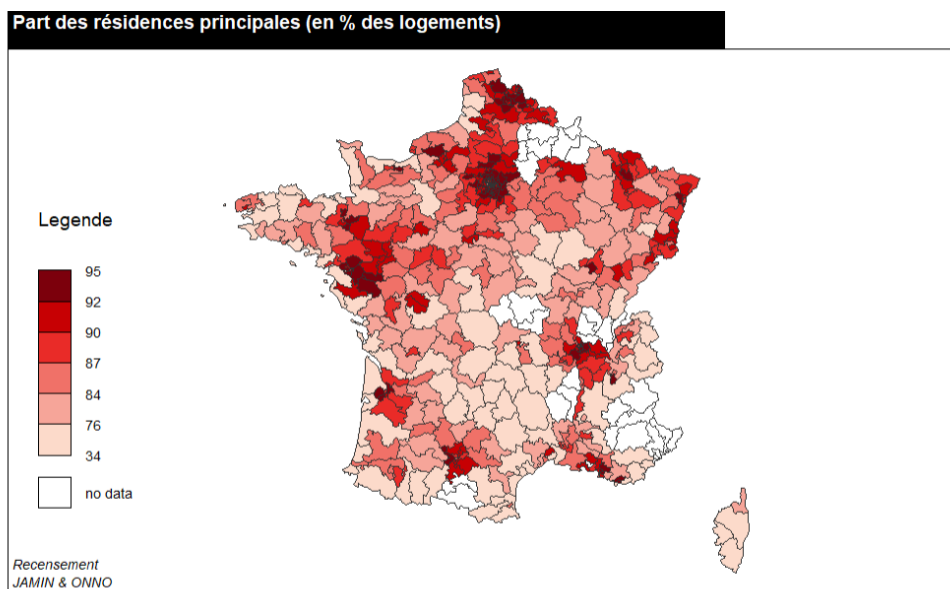
Figure II.7 – Part de la population active détentrice au mieux d'un diplôme de niveau BAC+3 et +



Concernant la variable part de la population active détentrice d'un Bac+3 ou plus, nous observons un taux particulièrement élevé à Paris, dans la région Rhône-Alpes, notamment en Isère et dans la Drôme, ainsi qu'en région Provence-Alpes-Côte d'Azur, autour de Marseille. Ces zones correspondent à des pôles économiques majeurs, où l'accès à des offres d'emploi qualifiées est plus important, notamment dans des secteurs comme les nouvelles technologies, les services, ou encore la recherche et l'innovation. Paris, en particulier, est un centre d'attraction pour les diplômés, en raison de la concentration d'entreprises et d'institutions qui offrent des opportunités adaptées aux qualifications supérieures. Cela n'est donc pas surprenant, ces zones ayant un environnement favorable à l'emploi pour les détenteurs de diplômes Bac+3 et plus.

### II.5.5 Part des résidences principales

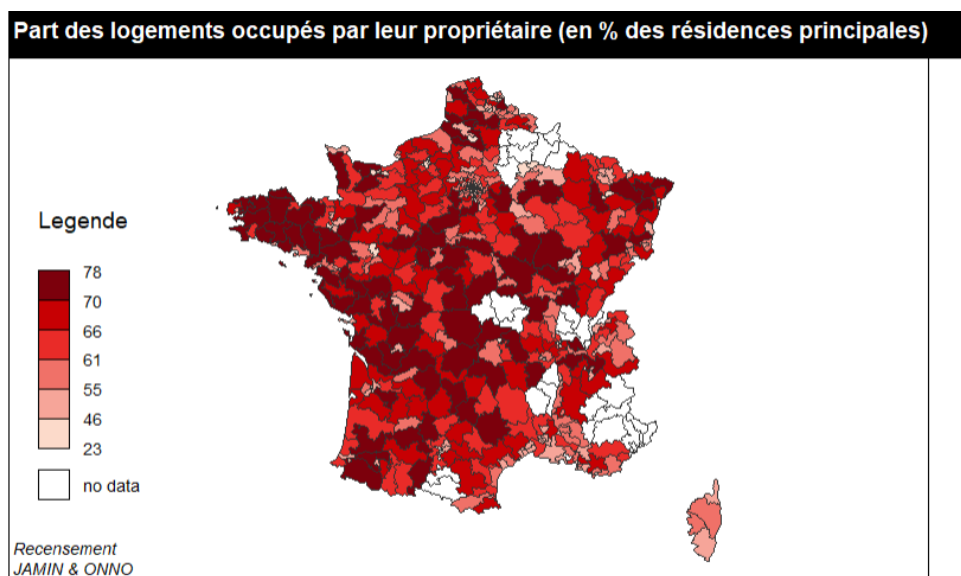
Figure II.8 – Part des résidences principales



En ce qui concerne la variable part des résidences principales, nous observons une forte concentration dans des zones comme Paris, le nord de la France (Somme, Nord-Pas-de-Calais), à Nantes en Pays de la Loire, à Lyon dans les Ardennes, et à Bordeaux en Nouvelle-Aquitaine. Cette tendance peut s'expliquer par un niveau élevé de stabilité géographique et socio-économique dans ces régions, où une plus grande proportion de la population vit dans des résidences principales, souvent en raison de la présence d'un tissu urbain développé et de fortes dynamiques résidentielles. En revanche, dans des régions comme l'Aquitaine, le Limousin et le nord du Languedoc-Roussillon ou encore la Bretagne, la part des résidences principales est relativement plus faible. Cela peut être dû à un plus grand nombre de résidences secondaires ou à une plus forte mobilité de la population dans ces zones, notamment liée à la nature de l'emploi ou aux pratiques démographiques. Il est intéressant de noter que ces régions, malgré un taux de résidences principales moins élevé, affichent un taux d'inscription plus important, ce qui pourrait refléter une population plus engagée politiquement et électoralement, en dépit d'une mobilité géographique plus marquée.

## II.5.6 Part des logements occupés par leur propriétaire

Figure II.9 – Part des logements occupés par leur propriétaire



Concernant la variable part des logements occupés par leur propriétaire, nous observons une concentration notable dans toute la France, avec une prédominance marquée dans certaines régions comme la Bretagne et le long de la côte ouest mais aussi dans le centre de la France. Dans ces zones, la part des logements occupés par leurs propriétaires est particulièrement élevée. La Bretagne, par exemple, est une région où l'on trouve de nombreuses petites communes, souvent rurales, et où la stabilité résidentielle est plus courante. Les propriétaires, dans ces régions, sont souvent des résidents permanents, ce qui peut être lié à des prix de l'immobilier plus accessibles, à un faible turnover locatif, et à une préférence marquée pour l'accession à la propriété.

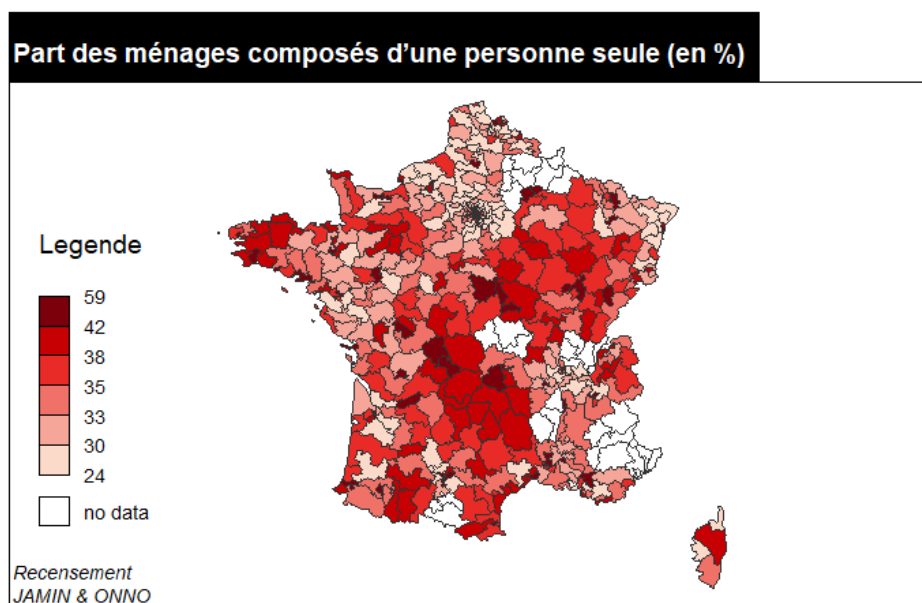
À l'inverse, dans les grandes agglomérations urbaines, comme Paris ou Lyon, la part des logements occupés par leurs propriétaires est généralement plus faible, en raison des prix élevés de l'immobilier, de la présence importante de logements locatifs, et de la mobilité accrue de la population. Toutefois, cette tendance à la forte proportion de propriétaires se retrouve globalement dans de nombreuses autres régions de France.

La part des logements occupés par leur propriétaire est plus faible en Normandie et dans le sud de la France, ce qui peut s'expliquer par la forte présence de résidences secondaires et de locations saisonnières. Ces régions, particulièrement attractives pour les vacanciers, voient de nombreuses maisons être louées pendant la saison estivale, réduisant ainsi la proportion

de résidences principales occupées par leurs propriétaires.

### II.5.7 Part des ménages composés d'une personne seule

Figure II.10 – Part des ménages composés d'une personne seule



Concernant la variable part des ménages composés d'une personne seule, nous observons une concentration plus marquée dans certaines zones spécifiques, notamment sur la diagonale du vide, qui traverse la France du sud-ouest au nord-est. Cette zone regroupe des régions où l'on retrouve une proportion significative de personnes vivant seules, souvent dans des départements moins urbanisés. Cette tendance peut s'expliquer par plusieurs facteurs démographiques et socio-économiques. D'une part, les grandes agglomérations urbaines, où les dynamiques familiales sont souvent plus diverses, peuvent présenter des modes de vie plus collectifs ou partagés, ce qui contraste avec certaines zones rurales ou périurbaines où les ménages sont souvent plus petits, notamment en raison de l'éloignement géographique et de la faible densité de population.

Cette répartition peut également être influencée par des facteurs tels que l'âge moyen plus élevé dans ces régions, où les individus sont plus susceptibles de vivre seuls après le départ des enfants ou la perte d'un conjoint. De plus, des conditions socio-économiques, comme l'accès plus limité à l'emploi ou aux infrastructures, peuvent contribuer à un mode de vie plus solitaire.

# Chapitre III

## Modélisation par l'économétrie spatiale

### III.1 Relations spatiales

#### III.1.1 Définition des matrice de poids

Les matrices de poids spatiaux sont des outils fondamentaux en économétrie spatiale, car elles permettent de modéliser les relations entre différentes unités géographiques ou spatiales (communes, régions, points géographiques...). Ces matrices définissent l'intensité des interactions spatiales entre les unités et sont utilisées pour capturer les dépendances spatiales dans les modèles économétriques.

Avant d'estimer les modèles spatiaux, il est essentiel d'établir les relations spatiales à travers la création d'une matrice de pondérations spatiales. Cette étape clé permet non seulement de définir la nature des liens entre les unités géographiques, mais aussi leur intensité respective. Nous allons présenter quatre type de matrice de poids.

#### **Matrice de poids de type tour**

La matrice de poids du tour repose sur le principe que deux unités sont connectées si elles partagent une frontière commune. Ce type de matrice est souvent utilisé lorsque les unités étudiées sont des zones géographiques bien définies, comme des communes ou des départements. Cette approche est nommée en référence à la pièce du jeu d'échecs, la "tour", qui se déplace horizontalement ou verticalement.

### **Matrice de poids de type fou**

La matrice de poids du fou établit une connexion entre deux unités lorsqu'elles se touchent uniquement par un sommet, c'est-à-dire un coin. Cette approche tire son nom du "fou", qui se déplace en diagonale aux échecs.

### **Matrice de poids de type reine**

La matrice de poids de la reine combine les caractéristiques des matrices du tour et du fou. Deux unités sont considérées comme connectées si elles partagent soit une frontière, soit un sommet. Cette méthode, inspirée des déplacements de la pièce "reine" aux échecs, est particulièrement adaptée lorsque toutes les formes possibles de relation géographique doivent être prises en compte.

### **Matrice de poids des plus proches voisins**

La matrice des plus proches voisins est utilisée lorsque la distance entre les unités est une variable clé. Dans ce cas, chaque unité est connectée à ses plus proches voisins en fonction d'une mesure de distance géographique, telle que la distance euclidienne. Ce type de matrice est souvent employé dans des contextes impliquant des données ponctuelles, comme des coordonnées GPS ou des localisations précises de villes.

### **Choix de la matrice de poids**

Le choix de la matrice de poids dépend de la structure des données et des objectifs de l'analyse. Par exemple, les matrices basées sur la contiguïté (tour, fou, reine) sont adaptées aux données zonales, tandis que les matrices des plus proches voisins conviennent mieux aux données ponctuelles.

Dans le cadre de notre analyse spatiale, nous avons donc choisi de poursuivre avec les matrices de poids de type reine et des plus proches voisins en raison de leur pertinence pour capturer efficacement les interactions spatiales dans nos données. La matrice de poids de la reine, qui prend en compte à la fois les frontières communes et les sommets partagés entre les unités géographiques, est particulièrement adaptée à notre échelle d'analyse. D'autre part, la matrice des plus proches voisins nous permet d'intégrer la notion de distance géographique directe, ce qui est essentiel lorsque les relations spatiales ne se limitent pas à la simple relation mais dépendent également de la proximité physique entre les unités. Ensemble, ces matrices offrent une représentation robuste et complète des dépendances spatiales.

Figure III.1 – Matrice de poids Reine

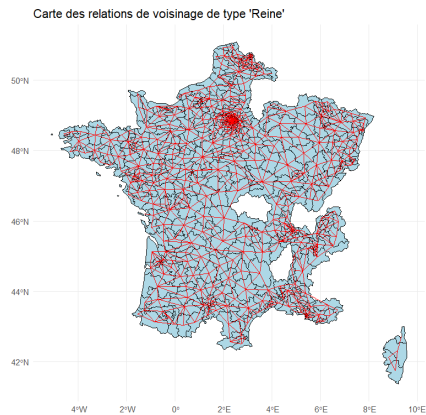
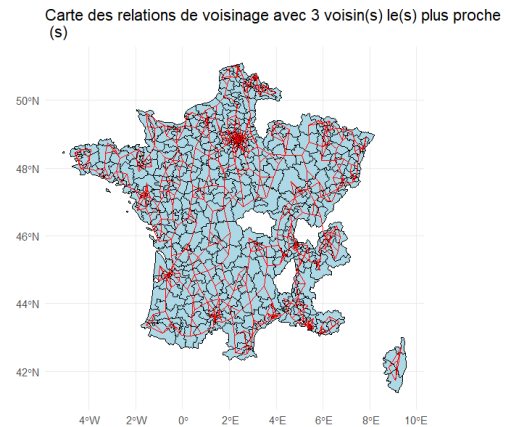


Figure III.2 – Matrice de poids PPV3



La carte concernant la matrice de type Reine montre un réseau de voisinage dense où chaque unité géographique est connectée à un grand nombre d'autres unités situées à proximité. Cette densité reflète une forte interaction spatiale, particulièrement visible dans les zones urbaines ou densément peuplées comme l'Île-de-France, où les connexions sont nombreuses et entrelacées. Ce type de réseau met en évidence les interdépendances régionales à grande échelle, mais peut aussi masquer les spécificités locales en raison du grand nombre de relations.

La seconde carte, basée sur la matrice de poids des Plus Proches Voisins, présente un réseau moins dense, où chaque unité est limitée à des connexions avec ses trois voisins les plus proches. Cette structure met davantage l'accent sur les interactions locales, avec des liens plus espacés dans les zones rurales ou faiblement peuplées. Elle permet de distinguer les dynamiques locales en réduisant la complexité du réseau, tout en montrant une moindre interconnexion globale par rapport à la première carte.

### III.1.2 Mesure de l'auto-corrélation

#### Auto-corrélation spatiale globale

Nous venons de définir les relations spatiales entre les unités géographiques grâce aux matrices de poids. A présent, nous allons analyser les liens entre ces zones en fonction du taux d'inscrit à l'aide de l'autocorrélation spatiale.

Le test de Moran I permet de mesurer si des zones proches géographiquement partagent des valeurs similaires ou, au contraire, dissimilaires. Une valeur positive indique un regroupement de valeurs proches tandis qu'une valeur négative révèle des différences marquées entre voisins. Enfin, une valeur proche de zéro traduit une absence de lien spatial. Ce test aide à identifier des schémas spatiaux significatifs et à comprendre si la répartition de la variable étudiée est influencée par la géographie.

Nous testons cette autocorrélation spatiale en comparant l'hypothèse nulle d'absence d'autocorrélation à l'hypothèse alternative indiquant la présence d'un lien spatial significatif.

Table III.1 – Test de Moran

Type de poids	Statistiques de Moran	p.value
Reine	0.5828747401	< 2.2e-16
Plus proches voisins	0.555964045	< 2.2e-16

L'étude de l'autocorrélation spatiale du taux d'inscription, réalisée à l'aide du test de Moran I, a permis de mettre en évidence une structure spatiale significative. Les analyses effectuées avec les deux matrices de poids spatiaux, Reine et Plus Proches Voisins, montrent des coefficients de Moran I respectivement égaux à 0,5829 et 0,5559 accompagnés de p-valeurs inférieures à 2,2e-16. Ces résultats confirment une autocorrélation spatiale positive, indiquant que les zones géographiquement proches présentent des comportements similaires en termes de taux d'inscription. La matrice Reine apparaît comme plus pertinente, avec une statistique de Moran légèrement plus élevée que celle observée avec les Plus Proches Voisins. Ces observations mettent donc en lumière l'influence des relations spatiales sur la répartition des taux d'inscription.

Nous allons maintenant visualiser nos données à l'aide du diagramme de Moran afin d'analyser la corrélation spatiale des données, c'est-à-dire pour vérifier si les valeurs d'une variable dans une zone sont liées à celles de ses zones voisines. Ce diagramme permet de voir si les zones proches ont des valeurs similaires et donc une corrélation positive, opposées



(corrélation négative) ou indépendantes. Selon la disposition des points correspondant aux départements, nous pouvons identifier des zones où les valeurs sont regroupées ou dispersées.

Figure III.3 – Diagramme de Moran Reine

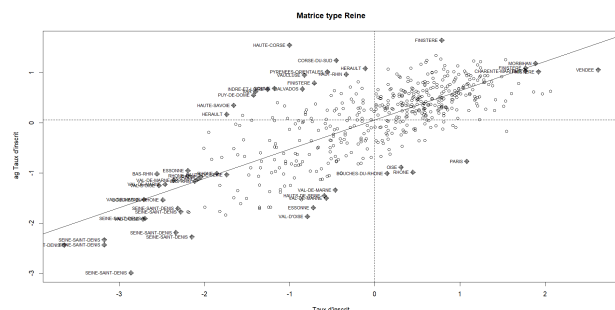
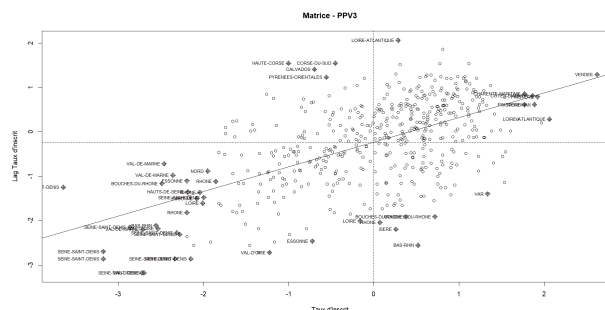


Figure III.4 – Diagramme de Moran PPV



Les deux graphiques illustrent une relation positive entre le taux d'inscription et son lag spatial, avec des tendances similaires mais des niveaux de dispersion différents. Le premier graphique montre une légère plus grande variabilité autour de la droite de régression, mettant en évidence des écarts marqués entre certaines régions et leurs lags spatiaux. Des régions comme la Seine-Saint-Denis se distinguent par des valeurs très faibles, tandis que d'autres, comme la Vendée ou encore Paris, affichent des écarts nettement positifs. Cette dispersion reflète une hétérogénéité notable entre les dynamiques locales.

En revanche, le second graphique présente une distribution plus concentrée des points autour de la tendance générale, suggérant une relation plus uniforme entre les variables. Si les régions atypiques comme la Seine-Saint-Denis et la Vendée restent visibles, leur comportement semble plus modéré et mieux intégré dans la tendance globale. Cela suggère une perception plus homogène des interactions spatiales, où les variations locales sont atténuées par une structure plus équilibrée.

### Auto-corrélation spatiale locale

À présent, nous allons visualiser et analyser l'autocorrélation spatiale locale à l'aide des indicateurs LISA (Local Indicators of Spatial Association). Ces derniers permettent d'identifier les zones géographiques présentant des structures spatiales similaires ou différentes, en mettant en évidence les clusters de valeurs élevées ou faibles. Cela nous aide à mieux comprendre les relations spatiales locales entre les observations dans un ensemble de données géographiques. Pour réaliser ces cartes, nous avons utilisé GeoDa (cf Annexe 2).

Figure III.5 – Représentation des LISA de type Reine

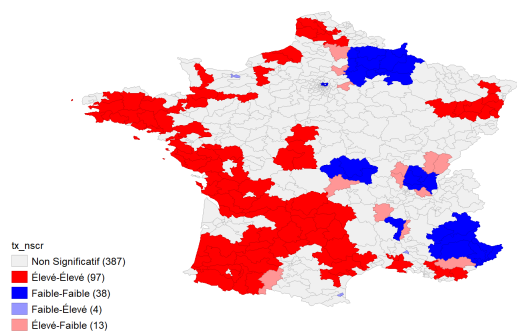
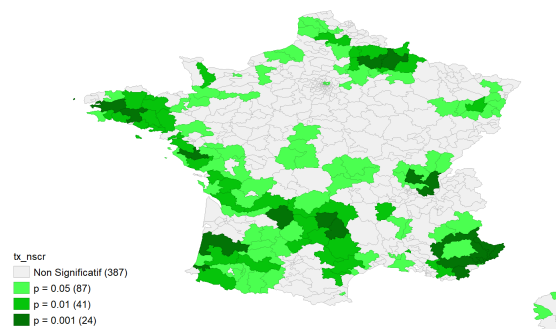


Figure III.6 – Significativités



La carte ci-dessus permet d'examiner les valeurs du taux d'inscription sur les listes électorales dans une zone donnée ainsi que celles de ses voisins, en identifiant des clusters spatiaux significatifs grâce à l'analyse LISA. Nous remarquons notamment un groupe distinct, intitulé "Élevé-Élevé", situé principalement dans l'ouest de la France. Cela indique que le taux d'inscription est particulièrement élevé dans ces régions, et que ces départements sont entourés de zones présentant des caractéristiques similaires. Ces clusters pourraient être associés à des facteurs tels qu'une population relativement stable et propriétaire, favorisant une meilleure participation électorale.

À l'inverse, les zones identifiées comme "Faible-Faible", principalement localisées dans certaines parties du centre et du sud-est, mettent en évidence des territoires où les taux d'inscription sont faibles et entourés par des départements aux mêmes caractéristiques. Ces résultats pourraient refléter des contextes socio-économiques moins favorables, une population plus jeune et mobile, ou encore un désintérêt pour les élections dans ces régions.

Enfin, les zones de décalage spatial, identifiées comme "Élevé-Faible" ou "Faible-Élevé", soulignent des disparités locales. Par exemple, un département à fort taux d'inscription entouré de zones à taux faible pourrait refléter des spécificités administratives, culturelles, ou économiques nécessitant une analyse approfondie. Il convient également de noter que certaines zones ne montrent pas de clusters significatifs, ce qui peut indiquer une absence de relation spatiale marquée dans ces territoires.

Cette analyse est complétée par une seconde carte qui illustre la significativité statistique des clusters identifiés. Les zones marquées en vert foncé ( $p = 0.001$ ) sont celles où les clusters sont les plus robustes et statistiquement fiables. Cela renforce la pertinence des résultats obtenus pour les régions "Élevé-Élevé" de l'ouest et les régions "Faible-Faible" identifiées.

Figure III.7 – Représentation des LISA de type PPV3

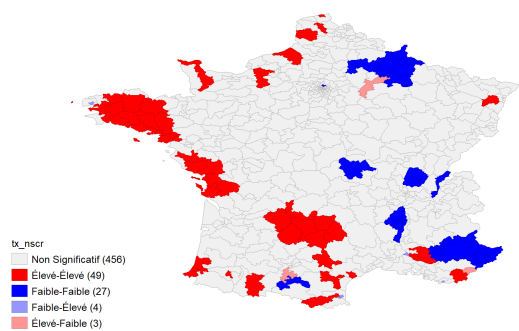
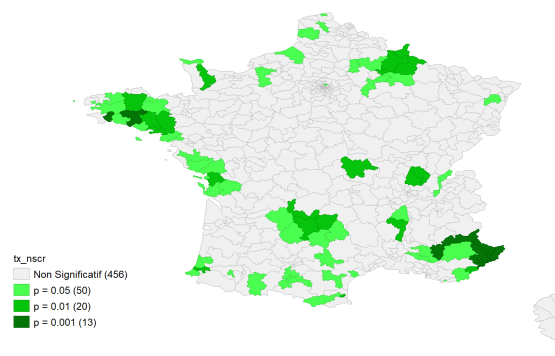


Figure III.8 – Significativités



La carte des LISA ci dessus représente une matrice de poids des Plus Proches Voisins. Les clusters "Élevé-Élevé" se concentrent principalement dans l'ouest et le sud-ouest de la France, reflétant des zones où l'inscription sur les listes électorales est forte, entourées de territoires similaires.

Les clusters "Faible-Faible" apparaissent plutôt dans le centre et l'est, suggérant des contextes locaux moins favorables à l'inscription. Les anomalies spatiales "Élevé-Faible" et "Faible-Élevé", bien que rares, soulignent des disparités locales spécifiques.

Enfin, de nombreuses zones sans clusters significatifs indiquent que les variations d'inscription dans ces territoires sont davantage influencées par des facteurs internes que par leurs relations spatiales immédiates.

En somme, les analyses LISA avec les matrices de poids Reine et PPV montrent des résultats similaires, avec des clusters "Élevé-Élevé" à l'ouest et "Faible-Faible" au centre et sud-est de la France. Cependant, l'approche des Plus Proches Voisins offre une analyse plus précise des zones locales, mettant en évidence des décalages comme "Élevé-Faible" et "Faible-Élevé". La matrice Reine donne une vue plus globale et homogène, tandis que celle des Plus Proches Voisins capte mieux les variations précises et les spécificités locales.

## III.2 Modèles économétriques spatiaux

Après avoir défini et visualisé les relations spatiales, nous allons développer des modèles économétriques. Dans un premier temps, nous en préciserons les spécifications, puis procéderons à leur estimation. Enfin, nous sélectionnerons le modèle le plus performant et l'analyserons.

### III.2.1 Définitions

#### Modèle SAR (Spatial Autoregressive Model)

Le modèle SAR introduit une dépendance spatiale dans la variable dépendante. La valeur de la variable à expliquer dans une région dépend directement des valeurs observées dans les régions voisines. Autrement dit, la valeur de  $y$  dans une unité spatiale peut avoir un impact sur la valeur de  $y$  dans une unité spatiale voisine. Ce modèle est particulièrement adapté pour analyser des phénomènes de contagion ou d'interaction directe entre régions, où les comportements locaux sont influencés par ceux des zones voisines.

L'équation s'écrit comme suit :

$$y = \rho Wy + X\beta + \varepsilon$$

où  $y$  est la variable à expliquer,  $Wy$  représente le lag spatial de  $y$ ,  $\rho$  est le coefficient de dépendance spatiale,  $X\beta$  sont les effets des variables explicatives, et  $\varepsilon$  est le terme d'erreur.

#### Modèle SEM (Spatial Error Model)

Le modèle SEM suppose que la dépendance spatiale réside dans le terme d'erreur, c'est-à-dire que les résidus  $\varepsilon$  dans une unité spatiale peuvent avoir un impact sur les résidus dans une unité spatiale voisine. Cela signifie que des facteurs non observés, affectant la variable dépendante, sont spatialement corrélés. Ce modèle est utile lorsque des variables omises ou des effets non mesurés suivent une structure géographique, influençant indirectement les résultats.

Ce modèle s'exprime à travers l'équation suivante :

$$y = X\beta + \varepsilon$$

$$\varepsilon = \lambda W\varepsilon + u$$

où  $\lambda$  est le coefficient de dépendance spatiale dans l'erreur,  $W\varepsilon$  représente le lag spatial de l'erreur, et  $u$  est une erreur non corrélée.

## Modèle SDM (Spatial Durbin Model)

Le modèle SDM combine les deux approches précédentes. Il inclut à la fois la dépendance spatiale dans la variable dépendante et celle des variables explicatives. Cela permet d'analyser simultanément les interactions entre régions et l'impact des caractéristiques des zones voisines.

L'équation s'écrit comme suit :

$$y = \rho W y + X\beta + WX\theta + \varepsilon$$

où  $WX\theta$  représente l'effet spatial des variables explicatives des régions voisines, et les autres termes ont la même signification que dans le modèle SAR.

## Modèle SLX (Spatial Lag of X)

Le modèle SLX introduit une dépendance spatiale dans les variables explicatives. Cela signifie que les variables explicatives dans une unité spatiale peuvent être influencées par celles des unités voisines. Autrement dit, les caractéristiques des régions voisines peuvent affecter la variable dépendante de la région observée. Ce modèle est utile lorsqu'on suppose que les facteurs spatiaux influencent directement les variables explicatives, et que cela se répercute sur la variable à expliquer.

Nous pouvons représenter ce modèle avec l'équation suivante :

$$y = X\beta + \rho WX + \varepsilon$$

où  $y$  est la variable à expliquer,  $X$  représente les variables explicatives de l'unité spatiale,  $WX$  est le lag spatial des variables explicatives des régions voisines,  $\rho$  est le coefficient de dépendance spatiale des variables explicatives, et  $\varepsilon$  est le terme d'erreur.

Ces quatre modèles permettent de capturer différentes formes de dépendances spatiales et sont adaptés à des contextes variés selon la nature des données et les dynamiques géographiques étudiées.

### **III.2.2 Choix du modèle**

Afin de déterminer le modèle le plus adapté à notre analyse, nous adopterons une approche systématique et rigoureuse. En premier lieu, nous utiliserons la méthode ascendante proposée par Le Gallo en 2002 pour faire le choix entre les modèles SAR et SEM. Cette approche nous permettra de comparer ces deux modèles en fonction de la structure spatiale présente dans nos données. Plus précisément, elle nous aidera à évaluer la pertinence de chaque modèle en tenant compte des spécificités de la dépendance spatiale, qu'elle réside dans la variable dépendante (SAR) ou dans le terme d'erreur (SEM).

Ensuite, dans le cas où l'analyse nécessiterait une comparaison plus approfondie, nous appliquerons l'approche mixte développée par Elhorst en 2010. Cette méthode permet de sélectionner le modèle optimal parmi les quatre modèles spatiaux présentés ci dessus, en tenant compte de différents critères d'évaluation. Elle offre une flexibilité supplémentaire pour adapter le choix du modèle en fonction des caractéristiques uniques de chaque matrice de pondération spatiale utilisée dans notre étude.

En combinant ces deux approches, nous serons en mesure de choisir le modèle le plus approprié, en garantissant une évaluation complète et rigoureuse. Cette démarche nous permettra d'assurer la robustesse et la fiabilité des résultats, en prenant en considération toutes les spécificités de la structure spatiale de nos données.

### III.2.3 Application

Pour commencer l'approche ascendante, la première étape consiste à estimer un modèle par Moindres Carrés Ordinaires (MCO). Cette analyse initiale est cruciale, car elle nous permettra de réaliser le test de Moran qui est indispensable pour détecter la présence éventuelle d'autocorrélation spatiale dans nos données.

Table III.2 – Test de Moran sur les résidus du modèle du MCO

Type de poids	Observed Moran	p.value
Reine	0.5225964355	< 2.2e-16
Plus proches voisins	0.542686554	< 2.2e-16

Les résultats des tests de Moran appliqués aux résidus de notre modèle de régression montrent la présence d'une autocorrélation spatiale significative, ce qui suggère que les erreurs du modèle ne sont pas indépendantes d'une unité spatiale à l'autre. En effet, dans les deux tests réalisés, la statistique de Moran I observée est positive et bien au-dessus de zéro (0.5226 pour Reine et 0.5427 pour Plus Proches Voisins), ce qui indique une autocorrélation spatiale positive. Cela signifie que les résidus dans les unités spatiales voisines ont tendance à être similaires. De plus, les p-values associées aux tests sont extrêmement faibles, ce qui permet de rejeter l'hypothèse nulle d'absence d'autocorrélation spatiale et confirme l'existence d'une dépendance spatiale significative dans les résidus. Ces résultats suggèrent que le modèle de régression utilisé n'explique pas entièrement la structure spatiale des données, et qu'il existe des effets spatiaux non pris en compte par les variables explicatives.

Ainsi, il est essentiel d'explorer les modèles d'autocorrélation spatiale afin de mieux appréhender les relations entre les variables et de détecter les éventuels effets spatiaux qui échappent au modèle actuel avec les MCO. Étant donné l'autocorrélation spatiale significative observée dans les résidus, il serait judicieux d'envisager l'utilisation d'un modèle spatial, comme les modèles SAR ou SEM, pour prendre en compte cette dépendance spatiale.

## Modèles SAR VS SEM

Table III.3 – Résultats des tests de diagnostic de multiplicateur de Lagrange

Test	p-value : Reine	p-value : Voisin
RSerr	< 2.2e-16	< 2.2e-16
RSlag	< 2.2e-16	2.567e-11
adjRSerr	< 2.2e-16	3.431e-14
adjRSlag	0.08491	0.8823

Les résultats des tests de multiplicateur de Lagrange révèlent des niveaux significatifs d'autocorrélation spatiale et d'erreurs de spécification spatiale dans les deux modèles analysés, utilisant à la fois les poids de Reine et de voisinage. Les p-values des tests RSerr et RSlag sont extrêmement faibles, ce qui indique une autocorrélation spatiale significative dans les résidus et dans la variable dépendante. Cependant, les p-values plus faibles observées avec les poids de Reine suggèrent une autocorrélation spatiale plus marquée, soulignant que cette matrice de poids capture mieux la dépendance spatiale dans les données.

De plus, bien que les tests ajustés comme adjRSlag indiquent une absence de dépendance spatiale significative dans la variable dépendante après ajustement, les erreurs du modèle sont toujours susceptibles d'être autocorrélées spatialement, ce qui confirme l'existence d'effets spatiaux non captés par les variables explicatives. En raison de cette autocorrélation significative et des erreurs de spécification spatiale, l'adoption du modèle SEM apparaît comme la solution la plus appropriée. L'utilisation des poids de Reine, avec des p-values plus faibles, renforce cette recommandation, car elle permet de mieux intégrer la dépendance spatiale et d'obtenir des estimations plus précises. Ce modèle spatial offrira un meilleur ajustement aux données, en traitant explicitement les effets spatiaux, ce qui améliorera la robustesse des résultats.



## Modèles SEM VS SDM

Nous allons à présent comparer les modèles SEM et SDM afin de déterminer lequel est le plus approprié à nos données.

Table III.4 – Test de Moran sur les résidus du modèle du MCO

Modèle	AIC
SEM	2540.8
SDM	2524.3

L'AIC qui est un indicateur essentiel pour évaluer l'ajustement relatif des modèles va nous permettre de déterminer notre meilleur modèle. Dans cette analyse, le modèle SEM présente un AIC de 2540.8, tandis que le modèle SDM affiche un AIC plus bas, à 2524.3. Un AIC plus faible suggère, en règle générale, un meilleur ajustement aux données ce qui indique que le modèle SDM parvient à capturer de manière plus précise les relations spatiales entre les variables. Ainsi, le modèle SDM semble offrir une performance supérieure en termes d'ajustement global.

Table III.5 – Test de Rapport de Vraisemblance entre les Modèles SDM et SEM

Mesure	Valeur
Log-vraisemblance du SDM	-1247.153
Log-vraisemblance du SEM	-1261.401
p-value	7.576e-05

Le test de rapport de vraisemblance entre le modèle SDM et le modèle SEM compare la qualité d'ajustement des deux modèles en évaluant leurs log-vraisemblances respectives. Le test une p-value de 7.576e-05, qui est significativement inférieure à 0.05. Cela signifie qu'il y a une différence statistiquement significative entre les deux modèles, et que le modèle SDM offre un meilleur ajustement aux données que le modèle SEM.

Les log-vraisemblances pour les deux modèles sont également fournies : pour le modèle SDM, la log-vraisemblance est -1247.153, tandis que pour le modèle SEM, elle est -1261.401. Par conséquent, le modèle SDM semble avoir une log-vraisemblance plus favorable et s'avère être un meilleur choix par rapport au modèle SEM selon ce test de rapport de vraisemblance.

Ainsi, nous allons retenir le modèle SDM pour la suite de l'analyse.

### III.2.4 Interprétation

Nous allons maintenant analyser les différents facteurs qui influent sur le taux d'inscription aux listes électorales en France.

Table III.6 – Résumé du modèle SDM

Variables	Estimation	p-value
(Intercept)	-11.16	0.0565
Age moyen	1.17	<2.2e-16
Pop inactive	-0.23	0.0725
Pop active bac+3	0.25	<2.2e-16
Residence principale	-0.19	<2.2e-16
Propriétaire	0.24	7.934e-13
Menage seul	-0.28	5.933e-06
lag age moyen	-0.69	7.118e-05
lag pop inactive	0.28	0.1846
lag pop active bac+3	-0.19	2.297e-07
lag residence principale	0.21	3.538e-10
lag propriétaire	-0.09	0.0649
lag menage seul	0.24	0.0039
Rho	0.7111	<2.22e-16
Asymptotic standard error	0.035261	<2.22e-16
Wald statistic	406.69	<2.22e-16
Log likelihood	-1247.153	
ML residual variance ( sigma squared)	7.1511	
Number of observations	505	
Number of parameters estimated	15	
AIC	2524.3	

Les résultats du modèle SDM révèlent plusieurs relations significatives entre les variables explicatives et le taux d'inscription, mettant en lumière les influences clés de certains facteurs sur cette variable.

Tout d'abord, l'âge moyen se distingue comme une variable fortement significative et positive. Cette relation suggère qu'une augmentation de l'âge moyen dans un département est associée à une hausse du taux d'inscription. Une explication possible pourrait être que dans les zones où la population est plus âgée, les comportements d'inscription sur les listes électorales sont plus marqués. Les jeunes adultes ou les populations plus âgées peuvent en effet être plus enclins à s'inscrire en raison de facteurs socio-économiques ou d'un plus grand engagement civique.

La proportion de la population active détentrice d'un diplôme de niveau Bac+3 ou plus montre également une relation positive et significative avec le taux d'inscription. Cela indique qu'un département avec un pourcentage plus élevé de personnes ayant un niveau Bac+3 ou plus tend à avoir un taux d'inscription plus élevé. Cette tendance pourrait être expliquée par le fait que les individus plus instruits sont davantage enclins à s'inscrire sur les registres électoraux, probablement en raison d'une meilleure compréhension de l'importance de l'inscription et d'un engagement civique plus fort.

En outre, la proportion de propriétaires dans un département présente un effet positif et significatif sur le taux d'inscription. Cela suggère que les régions avec une plus grande proportion de propriétaires voient des taux d'inscription plus élevés. Cette relation peut être liée à une stabilité socio-économique accrue dans ces zones, où les propriétaires sont plus investis et plus susceptibles de participer aux démarches administratives liées aux élections.

En revanche, la proportion de ménages composés d'une seule personne montre un coefficient négatif, ce qui signifie que les départements ayant un plus grand nombre de ménages solitaires tendent à avoir des taux d'inscription plus faibles. Cela pourrait suggérer que les individus vivant seuls sont moins enclins à s'inscrire sur les listes électorales, peut-être en raison de facteurs sociaux, comme un manque d'incitation à participer à la vie publique.

Les effets retardés des variables offrent également des informations intéressantes. En particulier, le lag de l'âge moyen est significatif, suggérant que l'âge moyen des régions voisines a un effet négatif sur le taux d'inscription dans une zone donnée. Cela indique que l'augmentation de l'âge moyen dans les départements voisins semble exercer une influence réduite sur l'inscription dans le département étudié.

De même, le lag de la population active détentrice au mieux d'un diplôme de niveau Bac+3 ou plus et le lag de la part des résidences principales s'avèrent significatifs, ce qui suggère une influence notable de ces variables sur le taux d'inscription dans une zone donnée.

Plus précisément, le lag de la population active qualifiée met en évidence une dynamique intéressante : la présence de travailleurs hautement diplômés dans les régions voisines semble jouer un rôle indirect sur les comportements ou les choix d'inscription, probablement en influençant les dynamiques économiques ou culturelles. Cela pourrait refléter un effet d'attractivité ou, au contraire, de concurrence entre territoires. Quant au lag de la part des résidences principales, sa significativité indique que la composition résidentielle des départements environnants agit comme un facteur structurant. Une proportion élevée de résidences principales dans les zones voisines peut être interprétée comme un indicateur de stabilité dé-

mographique ou de dynamisme local, qui peut influencer les décisions d'inscription dans le département étudié.

Un autre point notable est le coefficient Rho de 0.7111, qui indique une forte dépendance spatiale entre les observations. La p-value associée à Rho, qui est extrêmement faible, confirme que cette dépendance spatiale est statistiquement significative. Cela justifie l'utilisation d'un modèle spatial comme le SDM, où les effets des régions voisines sont pris en compte pour une meilleure précision des résultats.

Enfin, la statistique Wald de 406.69, accompagnée d'une p-value très faible, montre que l'ensemble du modèle est significatif. Cela signifie que les variables explicatives, y compris les effets spatiaux, exercent un impact significatif sur les taux d'inscription.

En conclusion, les résultats de ce modèle confirment les observations de la littérature, validant les relations identifiées. Ils mettent en évidence l'influence majeure des variables socio-économiques, telles que l'âge moyen, le niveau d'éducation, le statut de propriétaire et la structure familiale, sur le taux d'inscription sur les listes électorales. De plus, l'analyse confirme que la dépendance spatiale joue un rôle crucial dans les comportements d'inscription, les caractéristiques des départements voisins influençant fortement les taux d'inscription dans un département donné.

Table III.7 – Effets des variables sur le résultat

<b>Variables</b>	<b>Effet direct</b>	<b>Effet indirect</b>	<b>Effet total</b>
Age moyen	0.4561	0.3883	0.8444
Pop inactive	-0.0895	0.9007	0.8111
Pop active bac+3	0.0994	0.8959	0.9953
Residence principale	-0.0727	0.9077	0.8349
Propriétaire	0.0945	0.8983	0.9928
Menage seul	-0.1104	0.89	0.7796
lag age moyen	-0.2708	0.7334	0.462
lag pop inactive	0.1095	0.8905	1
lag pop active bac+3	-0.0727	0.9077	0.8350
lag residence principale	0.0815	0.9042	0.9857
lag propriétaire	-0.0335	0.9183	0.8848
lag menage seul	0.0947	0.8982	0.9929

Les résultats du modèle SDM révèlent trois types d'effets pour chaque variable : directs, indirects et totaux.

L'effet direct représente l'impact immédiat d'une variable sur le taux d'inscription sur les listes électorales dans un département. Par exemple, l'âge moyen a un effet direct positif de 0.45, ce qui signifie qu'une augmentation de l'âge moyen dans un département entraîne une hausse du taux d'inscription. De même, le statut de propriétaire a un effet direct de 0.0945, indiquant qu'une plus grande proportion de propriétaires dans un département est associée à un taux d'inscription plus élevé.

L'effet indirect capture l'impact des variables des départements voisins sur l'inscription. Par exemple, l'effet indirect de la population inactive est très élevé, à 0.9007, ce qui suggère que les comportements d'inscription dans les départements voisins influencent fortement les comportements locaux. L'effet indirect du statut de propriétaire est aussi important, à 0.8983, ce qui montre que la proportion de propriétaires dans les régions voisines a également une grande influence sur le taux d'inscription.

Enfin, l'effet total combine l'effet direct et indirect, donnant une vue d'ensemble de l'impact global de chaque variable. Par exemple, l'âge moyen a un effet total de 0.8444, soulignant l'influence combinée de la variable dans le département et ses voisins. De même, l'effet total du statut de propriétaire est 0.9928, montrant une relation très forte avec le taux d'inscription sur les listes électorales, en tenant compte à la fois des effets directs et des interactions spatiales.

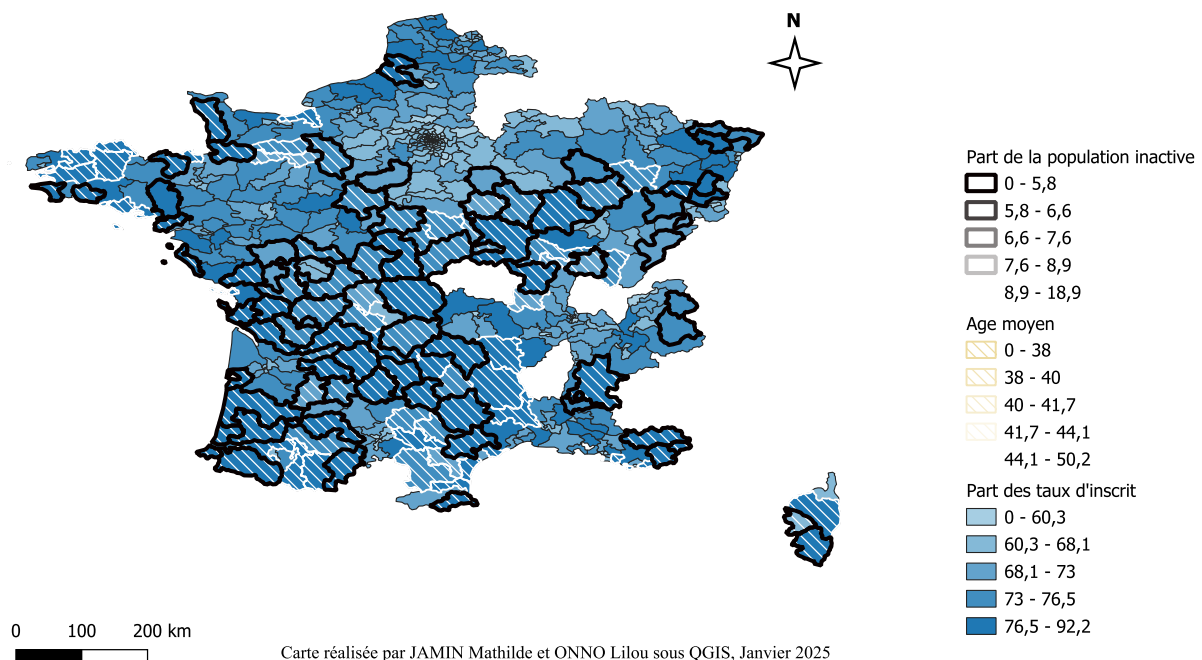
Ainsi, les effets directs, indirects et totaux montrent l'importance des facteurs socio-économiques locaux et des interactions entre départements dans l'explication du taux d'inscription sur les listes électorales en 2022.

## Chapitre IV

# Analyse cartographique complémentaire sous QGIS

Figure IV.1 – Cartographie

Une répartition inégale des départements dans le taux d'inscription sur les listes électorales en 2022



Nous avons choisi de représenter le taux d'inscription sur les listes électorales en 2022 en fonction de la part de la population inactive et de l'âge moyen. Sur notre carte, nous avons décidé de ne montrer que les individus les plus âgés et les plus actifs afin d'analyser leur comportement vis-à-vis de ces inscriptions. Nous avons fait ce choix en tenant compte des résultats obtenus dans nos modèles d'économétrie spatiale. En effet, ces deux variables étaient celles ayant le plus d'impact sur le taux d'inscription sur les listes électorales en 2022, comme en témoignent leur p-value et leurs coefficients associés.

Les résultats issus de la cartographie sont donc cohérents avec ceux de nos modèles ainsi qu'avec la littérature économique. En effet, les individus les plus susceptibles de s'inscrire sur les listes électorales en 2022 sont donc les personnes plus âgées et les plus actives, selon nos données. Ces groupes se concentrent principalement dans le Sud-Ouest de la France notamment dans la région Nouvelle-Aquitaine.

# Conclusion

Pour conclure, nous avons réalisé une analyse approfondie des facteurs influençant le taux d'inscription sur les listes électorales en France en 2022, en prenant en compte diverses variables démographiques et socio-économiques au niveau départemental. Les variables clés comprenaient l'âge moyen, le niveau d'éducation, la proportion de propriétaires, et la structure familiale. Nous avons d'abord cartographié le taux d'inscription pour visualiser sa distribution géographique et avons mené une analyse exploratoire pour identifier d'éventuelles hétérogénéités ou autocorrélations spatiales. Cette étape a révélé des disparités notables, avec des taux d'inscription plus élevés dans certains départements caractérisés par une population plus âgée et plus active.

Les relations spatiales entre départements ont été modélisées en utilisant différentes matrices de poids spatiaux. Les tests d'autocorrélation spatiale, notamment le test de Moran, ont révélé une autocorrélation positive significative, suggérant que les modèles de régression classique n'étaient pas adaptés sans ajustement pour l'interdépendance spatiale. Nous avons ensuite employé des modèles d'économétrie spatiale, en privilégiant le modèle SDM, qui a montré que les variables telles que l'âge moyen et la structure familiale avaient des effets directs sur l'inscription électorale, tandis que la densité de population était influencé par des effets indirects des départements voisins.

Nos résultats ont mis en lumière des disparités significatives dans les comportements d'inscription, où les départements avec une plus grande proportion de personnes âgées et une densité de population plus élevée se sont révélés avoir des taux d'inscription plus faibles. Ces résultats ont des implications importantes pour les politiques d'incitation à l'inscription, suggérant qu'une attention particulière doit être portée à la diversité des structures socio-économiques locales et aux influences spatiales.

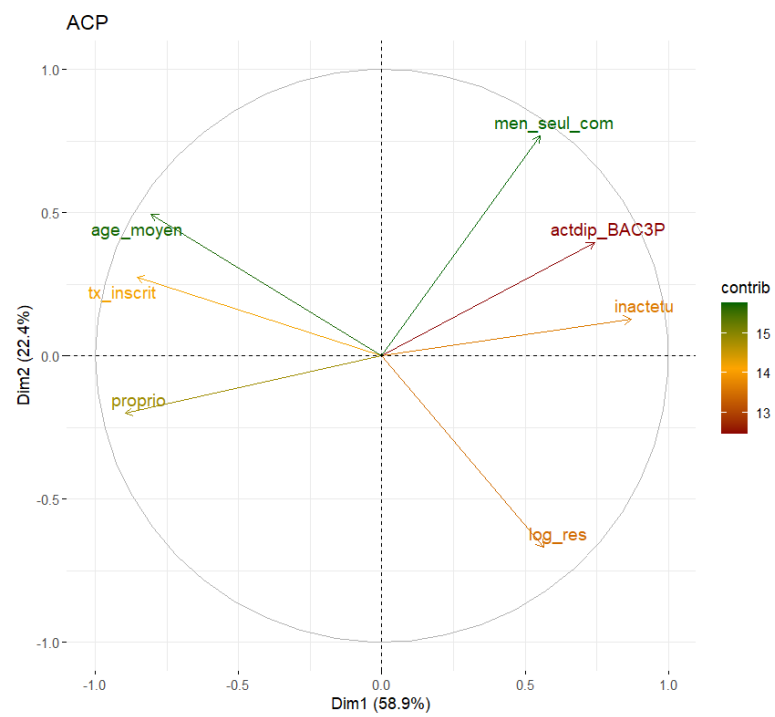
Pour finir, quelques limites ont été identifiées dans notre approche. Par exemple, certaines variables, telles que la structure familiale, pourraient ne pas saisir complètement les



nuances des comportements d'inscription en fonction des spécificités régionales. Pour affiner ces résultats, une stratégie possible serait d'intégrer des variables d'interaction ou de recourir à des modèles de régression géostatistique qui ajustent les coefficients en fonction des caractéristiques régionales. Ces approches permettraient d'approfondir la compréhension des dynamiques d'inscription et de mieux cerner les facteurs qui influencent le comportement électoral à l'échelle locale.

# Annexes

Annexe 1 : ACP



## Annexe 2 : Manipulation sous GeoDa

Pour commencer, nous avons importé notre base de données au format .shp, contenant les différentes variables ainsi que tous les fichiers associés à la géospatialisation.

### **Matrice de poids**

Nous avons accédé au menu "Outils", puis sélectionné "Gestionnaire de Pondérations" et cliqué sur "Créer". Lors de cette étape, nous avons choisi ID comme variable identifiant. Pour construire la matrice, nous avons opté pour la méthode de contiguïté Queen (Matrice de Contiguïté), que nous avons sauvegardée. Une fois enregistrée, nous avons visualisé le graphe de connectivité.

De manière similaire, nous avons créé une matrice de poids basée sur les plus proches voisins. Pour cela, nous avons sélectionné "Poids de distance", puis dans la sous-section "Méthode", choisi K-Plus Proches Voisins avec un paramètre fixé à 3 voisins

### **Local Indicators of Spatial Association**

Pour l'analyse LISA (Local Indicators of Spatial Association), nous avons sélectionné le menu "Espace", puis cliqué sur "I de Moran Local Univarié". Nous avons ensuite choisi notre variable d'intérêt, à savoir le taux d'inscription. À cette étape, nous avons sélectionné nos deux matrices de poids, d'abord celle construite selon la méthode de la Reine, puis celle des Plus Proches Voisins avec 3.

Enfin, nous avons activé les trois options disponibles afin d'afficher : l'indice de Moran, la carte des clusters LISA et la carte indiquant leur signification statistique.

# Bibliographie

- [1] INSEE, "Inscriptions et radiations depuis l'élection présidentielle de 2022," Mai 2024.
- [2] INSEE, "48,7 millions d'électeurs inscrits pour l'élection présidentielle 2022," 2022.
- [3] INSEE, "Les jeunes votent moins que leurs aînés," 2022.
- [4] C. FRANCE, "La mobilité étudiante dans le monde," 2023.
- [5] INSEE, "Âge moyen et âge médian de la population," 2024.
- [6] V.Territoires, "Vieillesse de la population et territoires," 2023.
- [7] Drees, "Les personnes âgées et leurs ressources," 2022.
- [8] INSEE, "Inactifs," 2023.
- [9] INSEE, "Tableaux de l'économie française," 2019.
- [10] S.Verba, "Participation in america : Political democracy and social equality," 1978.
- [11] A. Blais, "The role of education in voter turnout," 2004.
- [12] B.Highton, "The contextual causes of issue and party voting in american presidential elections," Decembre 2010.
- [13] F.Gougou and N.Sauger, "The 2017 french election study (fes 2017) : a post-electoral cross-sectional survey," 2017.
- [14] S.Verba and N.Nie, "Participation in america, political democracy and social equality," 1987.
- [15] C.Braconnier and J-Y.Dormagen, "La démocratie de l'abstention," 2007.
- [16] A. Campbell and all, "The american voter," 2008.
- [17] J-L.Thiébaud, "Les travaux de robert d. putnam sur la confiance, le capital social, l'engagement civique et la politique comparée," 2003.

# Liste des tableaux

II.1	Statistiques descriptives des variables . . . . .	8
III.1	Test de Moran . . . . .	23
III.2	Test de Moran sur les résidus du modèle du MCO . . . . .	30
III.3	Résultats des tests de diagnostic de multiplicateur de Lagrange . . . . .	31
III.4	Test de Moran sur les résidus du modèle du MCO . . . . .	32
III.5	Test de Rapport de Vraisemblance entre les Modèles SDM et SEM . . . . .	32
III.6	Résumé du modèle SDM . . . . .	33
III.7	Effets des variables sur le résultat . . . . .	35

# Table des figures

II.1	Histogramme . . . . .	10
II.2	Boxplot . . . . .	11
II.3	Matrice de corrélation . . . . .	12
II.4	Cartographie du taux d'inscrit sur les listes électorales en 2022 . . . . .	13
II.5	Cartographie de l'âge moyen de la population . . . . .	14
II.6	Part de la population inactive élève, étudiant . . . . .	15
II.7	Part de la population active détentrice au mieux d'un diplôme de niveau BAC+3 et + . . . . .	16
II.8	Part des résidences principales . . . . .	17
II.9	Part des logements occupés par leur propriétaire . . . . .	18
II.10	Part des ménages composés d'une personne seule . . . . .	19
III.1	Matrice de poids Reine . . . . .	22
III.2	Matrice de poids PPV3 . . . . .	22
III.3	Diagramme de Moran Reine . . . . .	24
III.4	Diagramme de Moran PPV . . . . .	24
III.5	Représentation des LISA de type Reine . . . . .	25
III.6	Significativités . . . . .	25
III.7	Représentation des LISA de type PPV3 . . . . .	26
III.8	Significativités . . . . .	26
IV.1	Cartographie . . . . .	37
IV.2	Annexe 1 : ACP . . . . .	41

# Table des matières

<b>Introduction</b>	<b>3</b>
<b>I Partie économique</b>	<b>4</b>
I.1 Cadre économique . . . . .	4
I.1.1 Taux d'inscrit sur les listes électorales . . . . .	4
I.1.2 Age moyen . . . . .	5
I.1.3 Part de la population inactive . . . . .	5
I.1.4 Part de la population active détenant au moins un Bac+3 . . . . .	6
I.1.5 Part des résidences principales . . . . .	6
I.1.6 Part des logements occupés par leur propriétaire . . . . .	6
I.1.7 Part des ménages composés d'une personne seule . . . . .	7
<b>II Analyse exploratoire</b>	<b>8</b>
II.1 Statistiques descriptives . . . . .	8
II.2 Distributions . . . . .	10
II.3 Outliers . . . . .	11
II.4 Corrélation . . . . .	12
II.5 Représentation cartographique . . . . .	13
II.5.1 Taux d'inscription . . . . .	13
II.5.2 Age moyen . . . . .	14
II.5.3 Part de la population inactive . . . . .	15
II.5.4 Part de la population active avec un bac+3 et + . . . . .	16
II.5.5 Part des résidences principales . . . . .	17
II.5.6 Part des logements occupés par leur propriétaire . . . . .	18
II.5.7 Part des ménages composés d'une personne seule . . . . .	19
<b>III Modélisation par l'économétrie spatiale</b>	<b>20</b>
III.1 Relations spatiales . . . . .	20
III.1.1 Définition des matrice de poids . . . . .	20

III.1.2	Mesure de l'auto-corrélation . . . . .	23
III.2	Modèles économétriques spatiaux . . . . .	26
III.2.1	Définitions . . . . .	27
III.2.2	Choix du modèle . . . . .	29
III.2.3	Application . . . . .	30
III.2.4	Interprétation . . . . .	33
<b>IV</b>	<b>Analyse cartographique complémentaire sous QGIS</b>	<b>37</b>
	<b>Conclusion</b>	<b>39</b>