

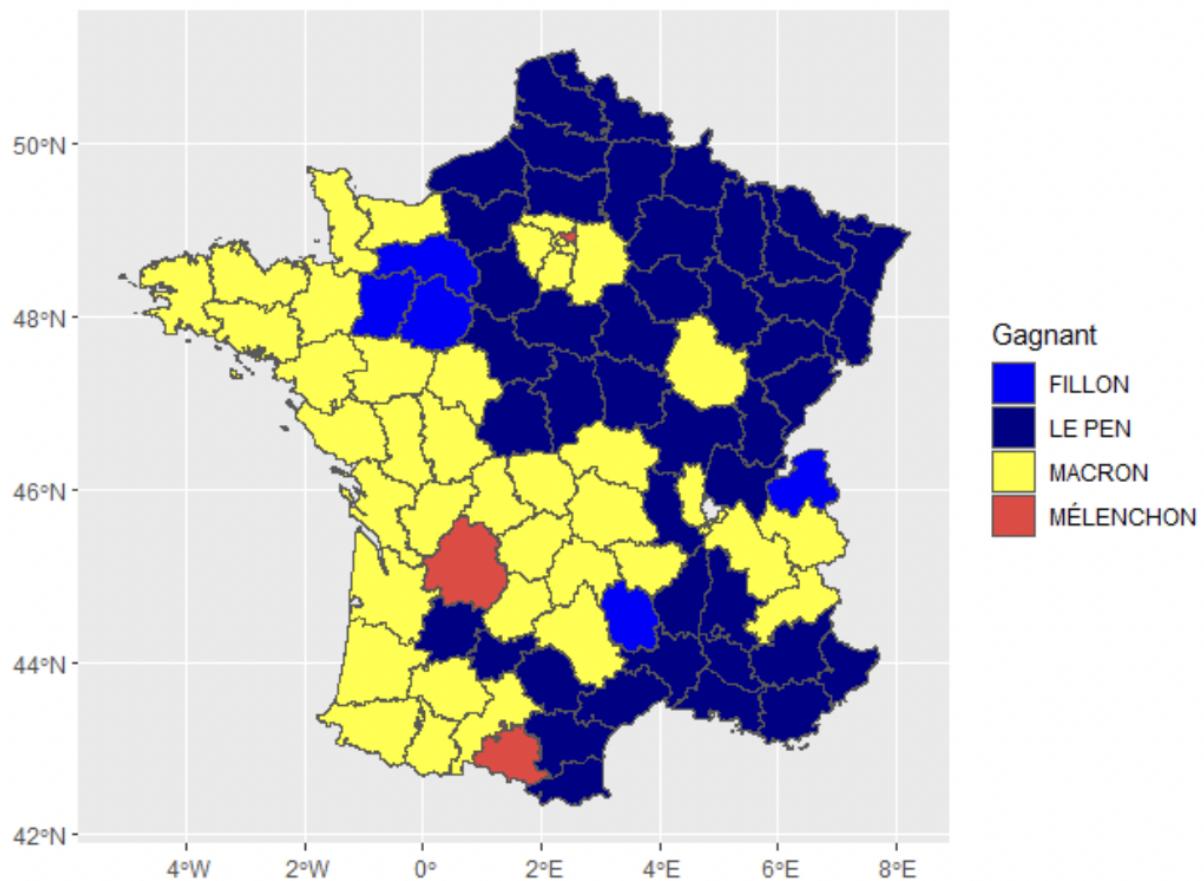
# PROJET GÉOSTATISTIQUE

## RAPPORT DE PROJET

---

### Comprendre les élections présidentielles sur le territoire Français

---



*Auteurs :* Thomas BERNARD, Paul GUARDIOLA, Noé LANGLAIS  
ING3

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Contexte et Méthodologie</b>	<b>3</b>
2.1	État de l'art . . . . .	3
2.2	Données et outils utilisés . . . . .	3
<b>3</b>	<b>Analyse spatiale</b>	<b>4</b>
3.1	Carte des votes majoritaires . . . . .	4
3.2	Calcul du point moyen pondéré . . . . .	4
3.3	Recomposition de l'Assemblée avec la proportionnelle en fonction de leur position géographique . . . . .	6
3.4	La gauche à gauche, la droite à droite . . . . .	7
<b>4</b>	<b>Explication des résultats</b>	<b>11</b>
4.1	La quantification des déplacements . . . . .	11
4.2	Explications des déplacements . . . . .	12
<b>5</b>	<b>Conclusion</b>	<b>14</b>

# Introduction

En France, nous observons une répartition inégale des votes selon le territoire qui peut être expliquée par plusieurs facteurs socio-économiques tels que le niveau d'éducation, le revenu moyen, la densité de population ou encore l'âge moyen des habitants.

L'analyse statistique permet de comprendre les dynamiques de vote à travers les années. Grâce aux données historiques, on peut identifier les tendances sur une période donnée. On peut y voir là un moyen de mettre en évidence l'évolution des partis sur le territoire Français ainsi que les potentiels liens entre les votes et les localités. L'analyse statistique des données électorales nous permettra de suivre l'évolution des préférences politiques à travers le territoire français.

Dans ce projet de géostatistiques nous tenterons de comprendre comment la composante géographique intervient dans les élections présidentielles. Pour cela nous étudierons dans un premier temps le déplacement des votes sur le territoire au cours de ces dernières années. Nous verrons aussi le positionnement géographique de l'assemblée national et des bords politiques. Enfin, nous verrons comment expliquer cette répartition géographique.

# Contexte et Méthodologie

## 2.1 État de l'art

Les élections présidentielles en France font l'objet de nombreuses analyses statistiques visant à comprendre les dynamiques complexes qui influencent le vote à l'échelle nationale et locale. Ces analyses intègrent une variété de données, telles que les résultats précédents, les caractéristiques démographiques, les tendances socio-économiques, et les évolutions politiques, pour éclairer les comportements électoraux. Elles sont cruciales pour comprendre les nuances politiques au niveau local et national, ainsi que pour anticiper les tendances électorales à venir. On retrouve principalement des cartographies de résultat dans les différents média mais aussi certaines analyses spatiales.

Les données de votes sont géoréférencées à l'échelle des bureaux de vote ce qui permet une analyse fine sur le territoire. Dans les articles on retrouve généralement des cartes de résultat. On peut par exemple représenter graphiquement les tendances électORAUX à travers des cartes choroplèthes, montrant des variations territoriales des votes. On peut retrouver ce type de carte sur le site de l'Observatoire des votes [2] par exemple. On y retrouve notamment la répartition des votes sur le territoire Français par candidat, mais aussi les écarts entre candidats. Les taux de variation des votes sont également disponibles. La représentation des votes électORAUX sur un fond de cartes est ce qu'il y de plus parlant pour le public mais cela ne permet pas toujours d'en déduire des résultats.

On retrouve aussi l'utilisation de méthodes comme la spatialisation des données pour étudier les corrélations spatiales des votes, mettant en évidence des clusters, des similarités ou des différences entre régions.

Généralement les analyses statistiques présentent la répartition du vote sur le territoire. L'étude des géostatistiques des votes en France est un domaine en évolution constante, profitant des avancées technologiques et des approches analytiques plus sophistiquées pour mieux comprendre les dynamiques électORAUX à l'échelle géographique.

## 2.2 Données et outils utilisés

Les données utilisées pour ce projet proviennent toutes du site [data.gouv.fr](http://data.gouv.fr). On y retrouve des données publiques fournies par le Ministère de l'Intérieur et des Outre-Mer. Ces données représentent les résultats électORAUX au niveau des bureaux de vote. On y retrouve le nombre de votes pour chaque candidat par bureau de vote. Les années sélectionnées pour l'analyse statistique s'étendent de 2007 à 2022 pour garder une certaine invariabilité dans les candidatures. Ainsi nous pourrons suivre l'évolution des votes pour un candidat ou un parti. Aucun nettoyage de la donnée n'a été nécessaire pour effectuer les analyses statistiques étant donné la fiabilité et l'exhaustivité de la source.

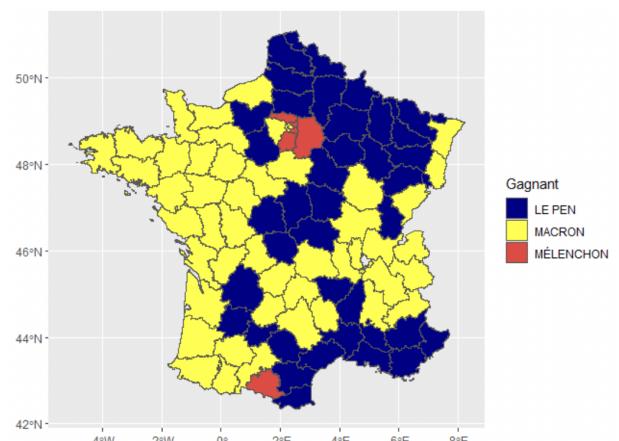
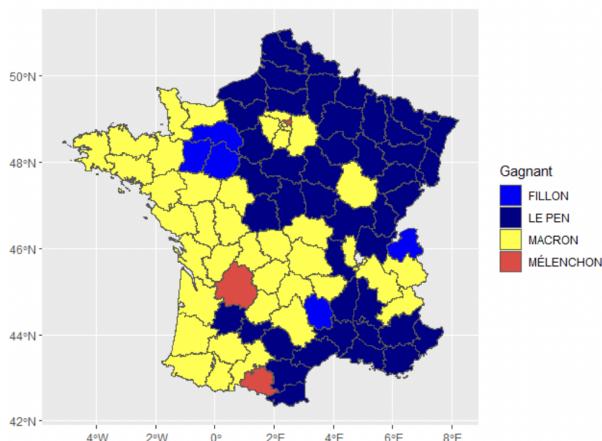
Pour ce qui est de la partie script nous utilisons le language de programmation R où nous faisons appel aux bibliothèques suivantes...

# Analyse spatiale

Dans cette partie nous verrons s'il existe un lien entre la géographie et les votes. Nous observerons notamment la répartition des candidats majoritaires dans les différents départements puis nous chercherons à savoir ce que dit cette répartition sur la propagation des idées politiques.

## 3.1 Carte des votes majoritaires

Pour avoir une première idée de la répartition des votes sur le territoire nous avons observé la donnée récupérée. Pour cela nous avons produit des cartes de vote majoritaire par département pour les 4 dernières élections. Ces cartes sont similaires à celles que l'on peut retrouver dans les médias lors des présidentielles. Grâce à ces cartes nous allons pouvoir faire nos premières suppositions pour la suite.



Ces cartes (figure 3.1 et figure 3.2) nous montrent qu'il existe bien un lien entre la géographie et le vote. En 2017, l'ensemble du Nord-Est de la France et la côte méditerranéenne ont voté à Droite alors que l'Ouest ont voté à Gauche. Nous retrouvons la même dynamique à quelques départements près en 2022. Nous pouvons alors nous demander comment le vote varie-t-il géographiquement entre les années et s'il existe un véritable lien entre la géographie et le vote.

## 3.2 Calcul du point moyen pondéré

Nous cherchons ici à associer géographiquement le vote à un seul point. En effet, nous avons pu observer plus haut que certains candidats avaient tendance à avoir plus de vote dans certaines zones géographiques que dans d'autre. Le calcul d'un point moyen pondéré par candidat permettrait de voir où se situe géographiquement son vote.

Le point moyen pondéré est le barycentre d'un nuage de points, pondéré par un poids. Dans notre cas d'étude, le nuage de point est les centroïdes des départements. Le poids est le nombre de votes des candidats dans le département. En effet, le point moyen pondéré va permettre de visualiser où se positionne le candidat à l'échelle de la France. Dans notre étude nous ne prendrons que les résultats de la France métropolitaine pour obtenir des points en France. L'objectif de cette étude est de calculer le point sur plusieurs années pour voir l'évolution des idées sur le territoire Français. Une évolution de ces points en dira long sur la capacité d'un candidat à faire part de ses idées politiques. Nous nous attendons notamment à voir un regroupement de ces points en fonction du bord politique des candidats.

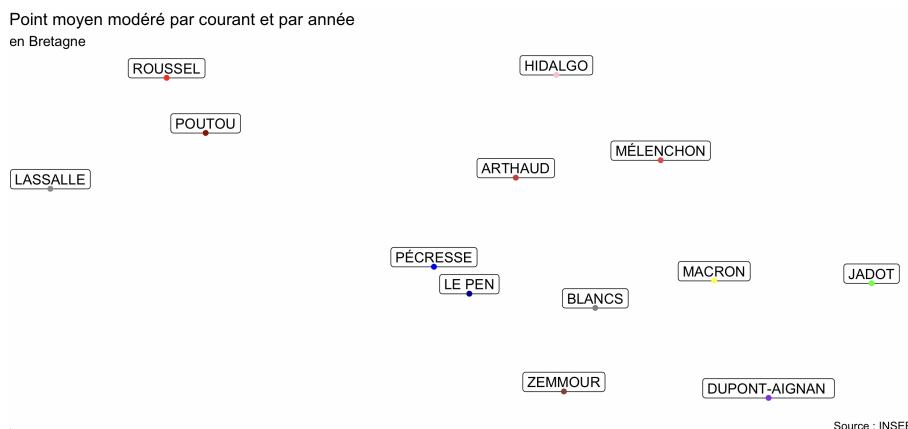


FIGURE 3.3 – Représentation des points moyens modérés par candidat en Bretagne en 2022

La figure 3.3 présente ces points pour l'année 2022 à l'échelle de la Bretagne. On y observe effectivement un rassemblement des candidats en fonction de leur bord politique.

En calculant, ce point sur plusieurs élections nous observons le déplacement des idées politiques des candidats.

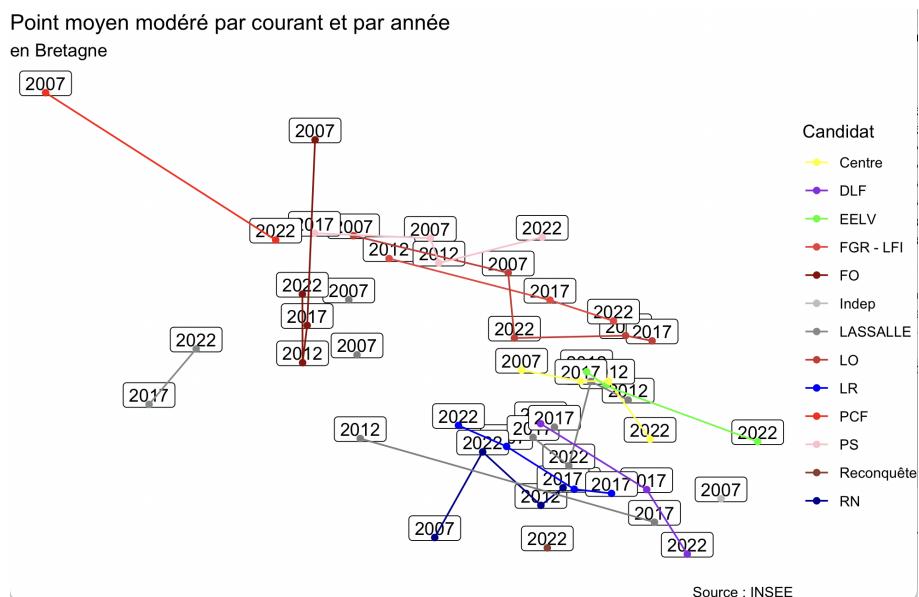


FIGURE 3.4 – Représentation des points moyens modérés par candidat en Bretagne en 2007, 2012, 2017 et 2022

Sur la carte ci-dessus nous avons dilaté le résultat des points moyens pondérés pour donner une carte plus parlante. En effet à l'échelle de la France, les points sont très concentré au centre du territoire et ne sont étalé qu'à l'échelle d'un département.

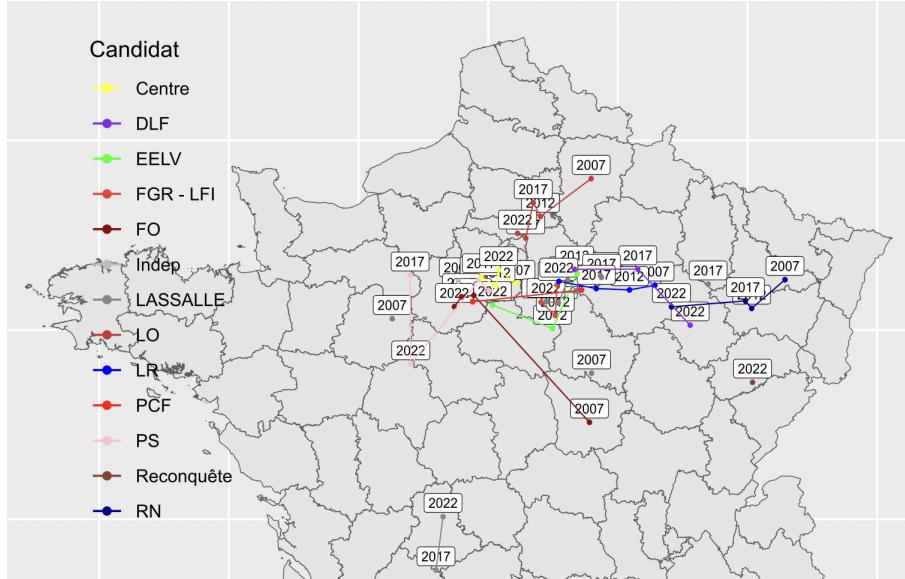


FIGURE 3.5 – Représentation des points moyens modérés par candidat en France en 2007, 2012, 2017 et 2022

Nous observons plusieurs tendances qui confirme la publication de La Revue Française de Science Politique [3]. Les petits partis sont porté par leur bastion d'origine. Le candidat Jean Lassalle est l'exemple le plus parlant dans ce cas. En effet ses points sont très tiré vers les pyrénées. Toutefois, il y a une propagation des idées de ces petits candidats en France visible par des points qui se rapprochent du centre du territoire en 2022.

Cette représentation dit aussi l'importance de l'implantation des idées de Gauche dans l'Ouest. Ce qui confirme l'article du Figaro concerant l'évolution de la gauche sur le territoire Français [5].

Enfin, le rassemblement national se propage de plus en plus sur le territoire avec un déplacement de son point moyen pondéré de l'Est vers le centre de la France. En effet, comme il est dit dans l'article du Monde [1] les électeurs de LR se sont reporté sur le RN. On retrouve principalement cette tendance avec les élections de 2022 où les électeurs de François Fillon se sont tourné vers Marine Le Pen.

### 3.3 Recomposition de l'Assemblée avec la proportionnelle en fonction de leur position géographique

La répartition du vote sur la carte de France montre une bonne répartition du vote Droite-Gauche. Notamment, Zemmour semble proche de Le Pen, Le Pen semble proche de Dupont-Aignan, Dupont-Aignant semble proche de Pécresse, etc... Cela semble indiquer qu'il est possible de recomposer la position idéologique des candidats par leurs positions géographique de leurs points moyens pondérés. Pour essayer de visualiser ce résultat, nous allons recomposer une Assemblée Nationale de deux façon. Premièrement, en partant du centre, en prenant ceux qui sont le plus proche d'Emmanuel Macron géographiquement et en les positionnant ainsi dans l'Assemblée. Deuxiemement, en partant d'un extrême et en listant les candidats du plus proche au plus loin de lui géographiquement.

Nous pouvons contruire les deux assemblées ci-dessus (figure 3.7). La première est très similaire à l'Assemblée Nationale actuelle. Nous retrouvons exactement les positions idéologique du centre à l'extrême droite avec Pécresse, Zemmour, Dupont-Aignan et Le Pen. A gauche, nous avons du centre à l'extreme Jadot, Roussel, Poutou, Mélenchon, Hidalgo, Arthaud et Lassale. Nous avons

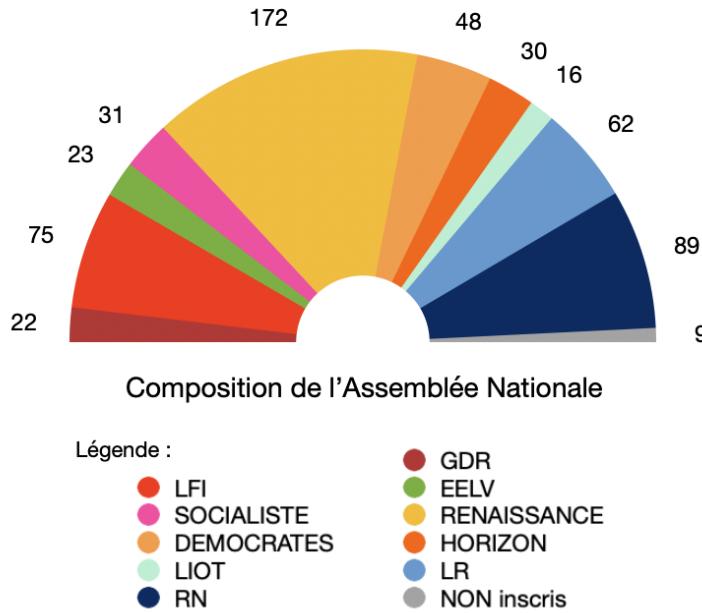


FIGURE 3.6 – Composition de l'Assemblée Nationale actuelle

Lassale et Arthaud qui sont très décentré en raison de leur faible score et donc de leur forte dépendance géographique à leur bastion d'origine. Ensuite la seule faute et le PS qui se retrouve à l'extrême gauche. Cette erreur peut s'expliquer par son faible score en 2022 et donc aussi sa forte dépendance géographique à l'Ouest. D'un point de vue général, l'emplacement géographique droite gauche en prenant le centre comme point de départ décrit assez bien l'emplacement idéologique des candidats. Nous pouvons noter qu'il y a une proximité géographique entre le PCF, les partis trotskistes et le centre d'un point de vue géographique comme nous pouvons le voir dans d'autres représentations des clivages Droite-Gauche comme avec le politoscope. [4]

La deuxième Assemblée Nationale montre la proximité en partant de Marine Le Pen. Nous remarquons qu'il est beaucoup plus dur d'interpréter ce résultat. Nous pouvons remarquer que l'ordre des partis est respecté sauf LFI qui se retrouve proche du RN. De même, cette représentation permet de mettre en lumière le vote blanc au centre et son importance.

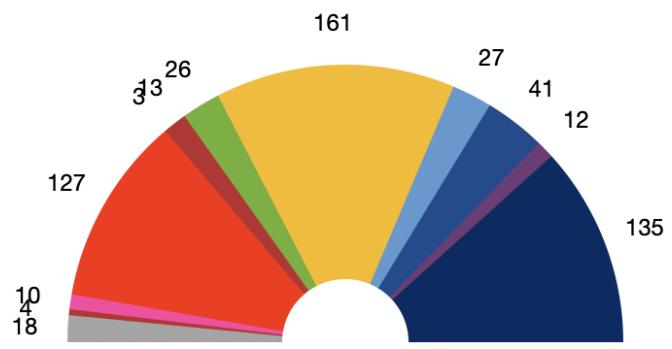
### 3.4 La gauche à gauche, la droite à droite

En conclusion, la composition de l'Assemblée Nationale par la proportionnelle en fonction de leur position géographique en partant du centre montre qu'il y a une bonne répartition géographique du vote de la Droite et de la Gauche.

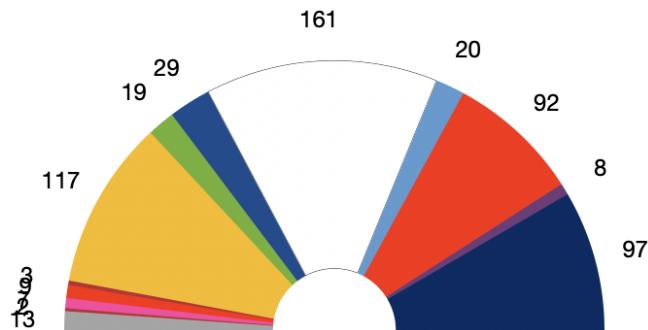
La figure 3.8 est la variation des points moyens pondérés mais colorisé en fonction du positionnement idéologique des candidats. Nous remarquons qu'il y a en Bretagne une véritable partition entre le droite et la gauche séparé par le centre. Nous obtenons le même résultat au niveau national.

Nous allons essayer de répartir la gauche et la droite grâce à un algorithme de classification. Le nombre de données ne nous permet pas d'envisager un algorithme de Machine Learning mais nous pouvons utiliser un algorithme de k-means. Nous allons juste lui demander de séparer le jeu en deux jeux de données en fonction des coordonnées des points moyens.

Nous obtenons le résultat ci-dessus (figure 3.10). Nous remarquons que ce premier résultat est satisfaisant, la majorité de la droite est bien attrapé par le cluster droit et le centre-gauche par



Composition de l'Assemblée Nationale avec la proportionnelle au présidentielle positionné en fonction de leur position géographique par rapport à Renaissance



Composition de l'Assemblée Nationale avec la proportionnelle au présidentielle positionné en fonction de leur position géographique par rapport au Rassemblement National avec le vote blanc

Légende :

- POUTOU
- ROUSSEL
- MACRON
- ZEMMOUR
- PECRESSE
- DUPONTAINAN
- LASSALLE
- HIDLAGO
- ARTHAUD
- JADOT
- BLANC
- MELONCHON
- LE PEN

FIGURE 3.7 – Composition de l'Assemblée Nationale avec la proportionnelle en fonction de leur position géographique

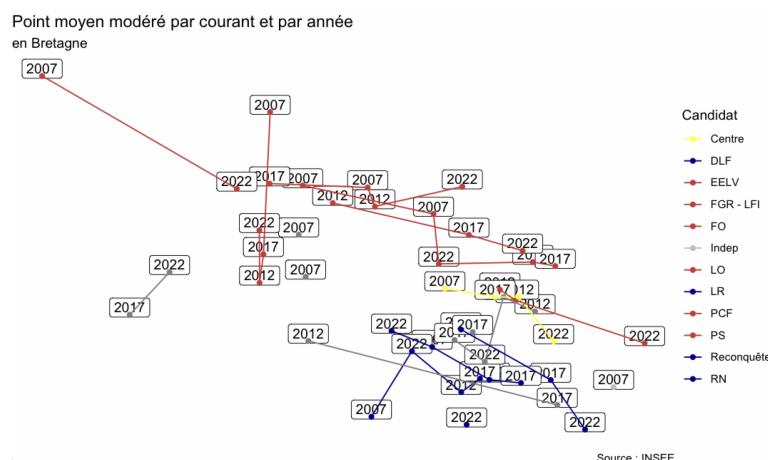


FIGURE 3.8 – Composition de l'Assemblée Nationale actuelle

le cluster de gauche. Nous avons fait le calcul du tableau suivant grâce à plusieurs test de 2-means car le résultat n'est pas le même en fonction des positions des points de départs des centroïdes des

Point moyen modéré par courant et par année  
en France

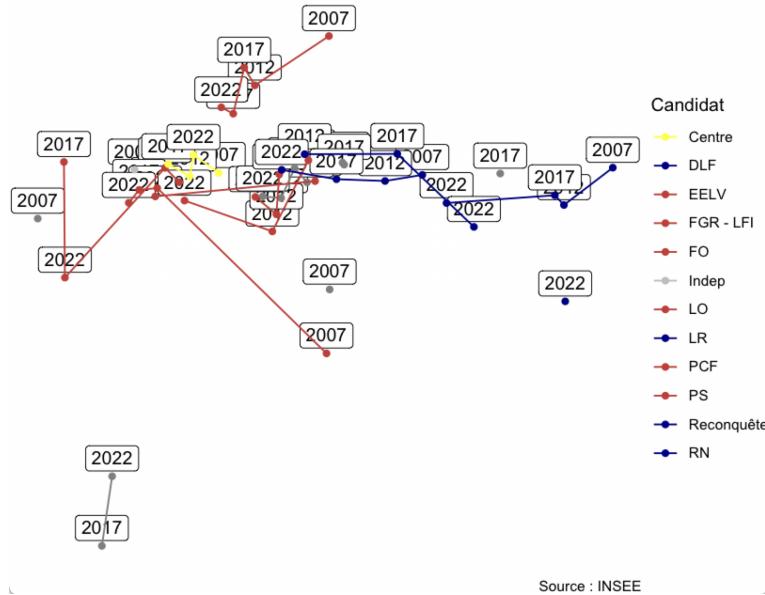


FIGURE 3.9 – Composition de l’Assemblée Nationale actuelle

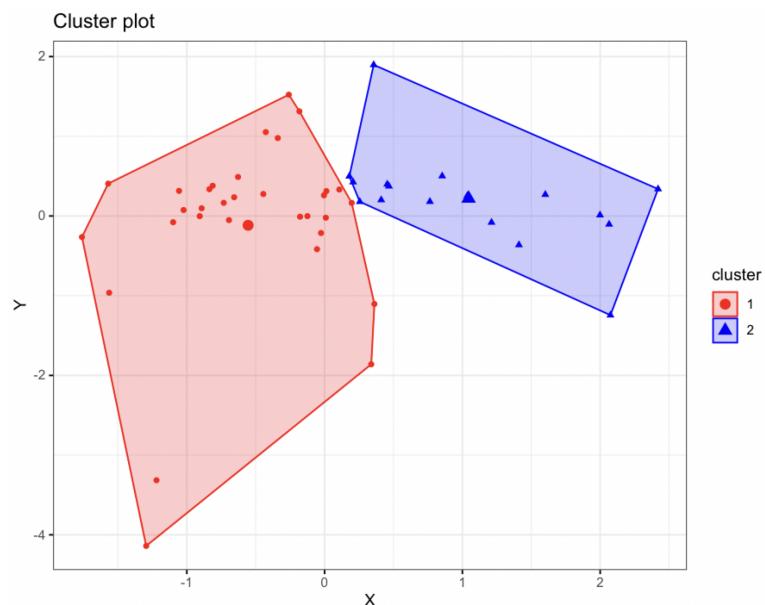


FIGURE 3.10 – Clusters ressortant après l’exécution d’un 2-means sur les points modérés

clusters.

	Centre - Gauche	Droite
Centre - Gauche	27	3
Droite	2	13
	Recall	Precision
Centre - Gauche	0.90	0.93
Droite	0.86	0.65
Average	0.88	0.79
F score	0.83	

Le F score est correct, nous remarquons que seul la précision de la droite est inférieur à 0.85. Nous pouvons expliquer ce résultat par le fait que seul les points situés à la frontière du cluster

sont des erreurs et également les points extrêmes comme le point du parti communiste qui est celui le plus au Nord.

# Explication des résultats

Dans cette partie, nous analysons des schémas de déplacement des votes à travers les élections, décelant des nuances et des tendances singulières chez les différents partis politiques. De plus, nous explorons la corrélation entre les choix électoraux et des indicateurs socio-économiques clés, tels que le revenu fiscal par département.

## 4.1 La quantification des déplacements

Après analyse, nous remarquons que le vote ne se déplace pas de la même façon pour l'ensemble des candidats. Le Parti Socialiste a un déplacement beaucoup moins rectiligne que l'ensemble des partis politiques de droite. Une façon d'analyser ces déplacements est de voir la distance parcouru entre les années ou la régularité de ces déplacements au niveau de sa norme. Pour cela nous allons calculer un indice avec les normes de vecteurs et l'écart-type :

Soit  $Vect$  une liste de vecteur conduisant d'un point à un autre pour plusieurs années et pour un même candidat, de longueur  $n$ .  $Vect_{tot}$  est le vecteur allant de la première année à la dernière année.

$$Indi_{norme} = \frac{||\vec{Vect\_tot}||}{\sum_i^n ||\vec{Vect}_i||}$$

Nous avons un indice de norme important lorsque les points moyens pondérés d'un candidat peuvent se mettre sur une même droite. L'indice vaut 1 lorsque c'est le cas et 0 lorsque ce n'est pas le cas. Nous interprétons cet indice que lorsque l'indice vaut 1 les idées s'imprègnent de façon continue. Lorsque l'indice est proche de zéro nous avons un candidat qui à un électoral qui varie change fortement entre les années. Rajouter l'écart-type permet de voir si les changement de vote se font de façon régulier ou non.

Nous pouvons remarquer que ces indices doivent être mis en perspective avec le score aux élections. Le RN et les LR ont tous les deux un fort indice de norme et un faible écart-type. C'est à dire que les deux parties varient de façon continue dans l'espace et de façon régulière. Sauf que le RN c'est pour une progression et une propagation de leurs idées alors que les LR c'est pour une recession et une retraction de leurs idées. Nous pouvons même avancer qu'il s'agit d'un lien de cause à effet d'un vote LR qui s'est déplacé vers le RN. [1]. Nous avons LFI qui a un indice de norme très faible donc un électoral très varibale avec une propagation des idées très discontinue. Le parti avec le plus grand écart-type est FO, c'est à dire que le parti a connu des changements d'électoral très important ces 20 dernières années.

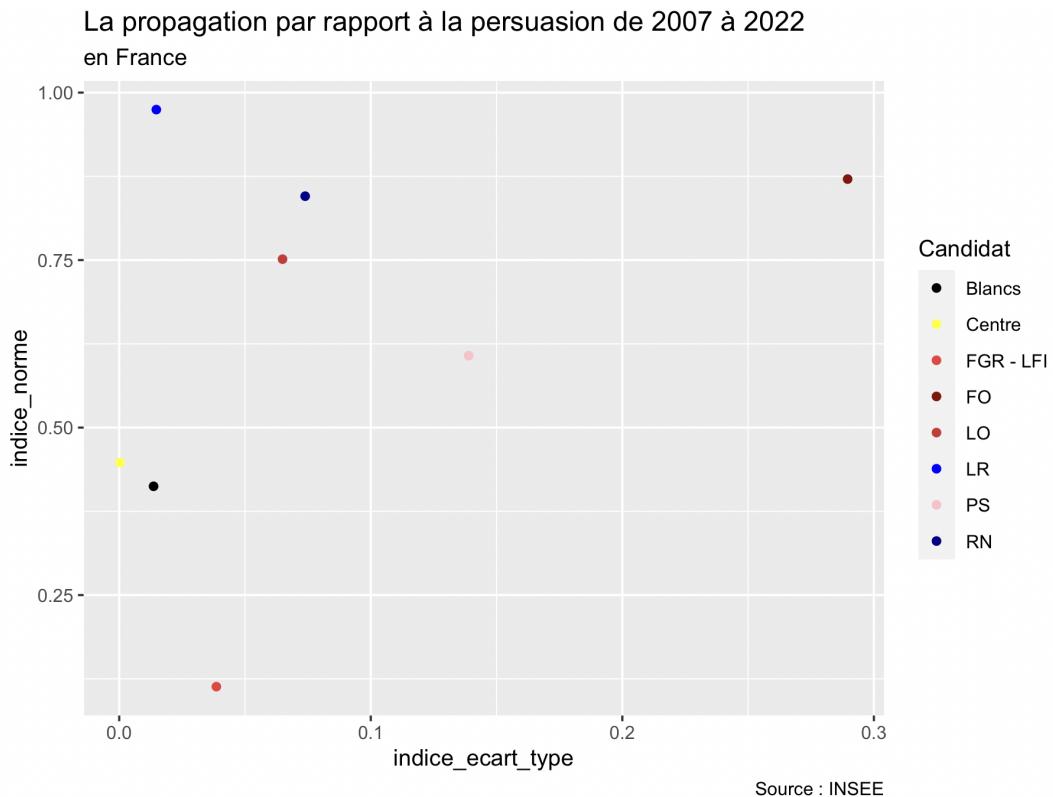


FIGURE 4.1 – indice de norme en fonction de l'écart-type au niveau national

## 4.2 Explications des déplacements

On a pu observer au cours de ce travail que le vote des français semblait très lié à leur géolocalisation. Nous avons donc cherché à déterminer quels étaient les causes de cette liaison car il semble incongru que la seule différence de territoire justifie cela. Nous nous sommes donc penchés sur les données socio-économiques que l'on a pu récupérer sur data.gouv avec notamment celles sur les revenus fiscaux des français par département. Pour étudier le lien possible avec nos données de vote nous avons décidé d'étudier la corrélation pour chaque candidat entre son nombre de vote et le revenu fiscal moyen par département. Cela nous a alors donné les trois tableaux suivants :

Grâce à ces indices de corrélations, nous pouvons observer que, d'une élection à l'autre, les partis gardent plus ou moins la même corrélation ce qui semble indiquer que les gens ont plus ten-

Nom	2012	2017
Arthaud	0.22	0.26
Le Pen	0.29	0.24
Poutou	0.39	0.37
Mélenchon	0.50	0.46
Holl / Hamon	0.49	0.63
Dupont-Aignan	0.49	0.55
Cheminade	0.55	0.59
Bayrou / Macron	0.66	0.70
Sarkozy / Fillon	0.67	0.74

FIGURE 4.2 – Correlation entre le revenu fiscal par foyer fiscal par département et les votes par candidats en 2017 et 2012

dance à voter pour un parti ou des idées que pour une personne. En outre on observe également la forte corrélation entre le revenu et le vote de droite "historique" (UMP et LR) ainsi que pour le MO-DEM et EnMarche ce qui semble indiquer que ces partis vont chercher leur électorat surtout chez les personnes aisées. A l'inverse on remarque que d'autres partis semblent prendre des voix peu importe les revenus des électeurs comme Lutte Ouvrière, Chasse et Pêche ou encore le Rassemblement National.

Il semble ainsi réducteur de déterminer un vote par rapport au revenu du votant car cela ne s'applique qu'à certains partis. Il serait donc intéressant d'associer à cela d'autres données socio-économiques comme le chômage, le coût de la vie ou bien encore le caractère rural ou citadin.

# Conclusion

Ce projet a mis en évidence l'interaction entre géographie, idéologie politique et comportements électoraux en France. Premièrement, les cartographies des points moyens pondérés par candidat ont révélé une distribution territoriale significative des votes, offrant une représentation visuelle des tendances politiques. Cette visualisation a permis d'observer des regroupements de candidats en fonction de leurs affiliations politiques, mettant en évidence des liens géographiques entre les positions idéologiques.

Deuxièmement, la recomposition hypothétique de l'Assemblée Nationale en se basant sur les positions géographiques des candidats a démontré la corrélation entre la position géographique et l'orientation politique des candidats. Cette recomposition, à partir du centre ou en partant d'un extrême, a révélé des schémas intéressants dans la répartition des forces politiques, reflétant parfois la position droite-gauche actuelle notamment lorsque nous repartissons la droite et la gauche en partant du centre.

L'analyse des déplacements du vote au fil du temps a mis en évidence des différences significatives dans les schémas de déplacement des partis politiques. Notamment, le RN a montré une propagation régulière de leurs idées dans les terres, tandis que d'autres partis ont connu des déplacements plus importants, comme le PS soulignant ainsi des changements brutaux dans le choix des électeurs de gauche.

Enfin, la corrélation entre le revenu fiscal par foyer et les votes pour certains partis a été observée, indiquant une tendance de certains partis à attirer davantage un électoralat spécifique en fonction du niveau de revenu. Ceci est l'une des raisons qui peut expliquer pourquoi il existe une dynamique géographiques du vote.

Les résultats soulignent la nécessité d'une approche multidimensionnelle pour comprendre pleinement les comportements électoraux, invitant à explorer davantage les multiples facteurs socio-économiques qui façonnent ces dynamiques.

# Bibliographie

- [1] Pierre Breteau et Luc Martinon. En cartes : visualisez la recomposition politique entre les élections présidentielles de 2017 et 2022. *Le Monde*, 2022.
- [2] Site Internet [En ligne]. L'observatoire des votes, 2023.
- [3] Céline Colange Michel Bussi, Jérôme Fourquet. Analyse et compréhension du vote lors des élections présidentielles de 2012. *Revue française de science politique*, 2012.
- [4] David Chavalarias Noé Gaumont, Maziyar Panahi. Reconstruction of the socio-semantic dynamics of political activist twitter networks—method and application to the 2017 french presidential election. 2018.
- [5] Albert Zennou. En 31 ans, la gauche est allée vers l'ouest. *Le Figaro*, 2012.