



MAÎTRISE EN INFORMATIQUE

IFT870 : FORAGE DE DONNÉES

Analyse des correspondances des résultats du
premier tour des Élections Présidentielles françaises
de 2017 en fonction de la catégorie
socio-professionnelle des votants

Auteur :

Thibaud MERIEUX

Encadrant :

Bernard COLIN

1 Introduction

Dans les systèmes démocratiques, les élections sont l'un des moyens les plus utilisés par les citoyens pour participer au processus politique. Indépendamment des systèmes électoraux et des formes de gouvernement, elles permettent de sélectionner les décideurs publics et de les rendre responsables vis-à-vis des électeurs. Puisque le résultat des élections dépend du vote des citoyens éligibles qui vont effectivement aller voter, afin de bien comprendre la mise en place des politiques publiques, il est toujours intéressant de chercher des corrélations surprenantes entre le choix effectué dans l'isoloir et l'environnement des électeurs.

L'étude <https://stevedutch.net/research/elec2000/geoelec2000.htm> où des professeurs de l'université du Wisconsin (USA) ont trouvé un fort lien entre les résultats des Élections Présidentielles américaines et la richesse ou non des sols, motive en grande partie cette étude. On s'intéresse ici à la répartition des votes selon le type géologique des sols, répartis en cantons.

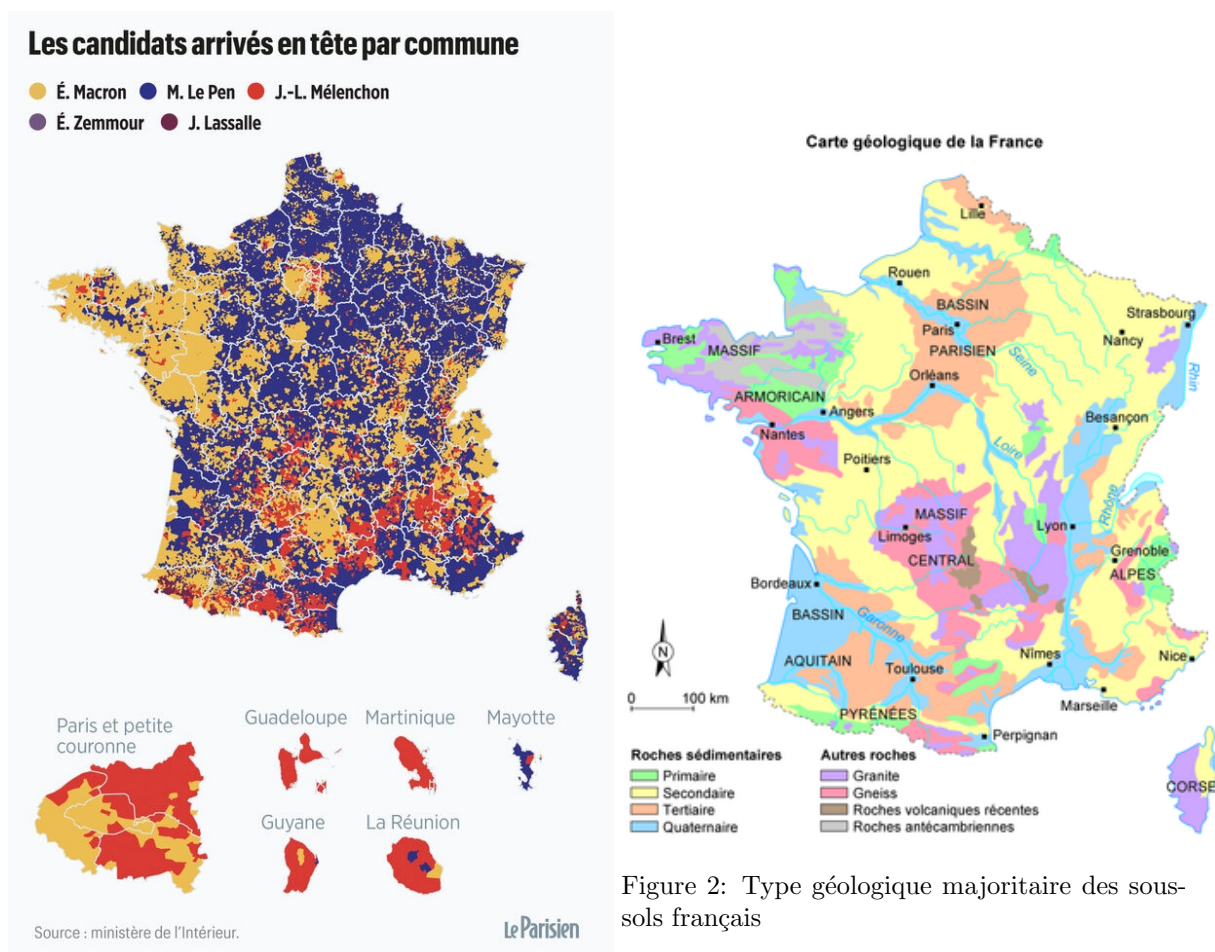


Figure 1: Résultats du 1er Tour des Élections Françaises 2022 par canton

On peut voir ici qu'il semble y avoir une certaine similarité entre les deux cartes, surtout visible en Bretagne et dans les Pays de la Loire

2 Position du problème

Il est plus précisément question de comprendre s'il existe une forte influence entre le sol d'une région et les votes dans cette dernière. Il existe naturellement un lien entre le sol et l'activité économique d'une région, nous cherchons maintenant à savoir, si au delà du cadre socio-professionnel d'une personne, on peut simplement résumer la carte politique française par la géologie.

3 Présentation des données

Il faut pour effectuer l'analyse créer le tableau de données : il est très difficile (voire impossible) de trouver de telles données déjà préparées. La solution est donc de trouver d'une part les résultats par circonscription des Élections, et ensuite les croiser avec le sol majoritaire de chaque circonscription.

Les données electorales sont trouvable sur le site officiel du Gouvernement Français data.gouv.fr

Les données géologiques sur fr.statista.com/statistiques/7039799067777/.

Pour croiser les données, on fait le choix d'associer à chaque circonscription un poids égal et non un poids dépendant de la population, car l'étude porte sur les sols et non la population, et d'en faire une moyenne.

NOTE : On remplace par la suite "Roches Volcaniques" par "R. Volc."

Voici le tableau final qui est utilisé seul dans la suite du document.

	Primaire	Secondaire	Tertiaire	Quaternaire	Granite	Gneiss	R. Volc
E. Macron	0,341	0,221	0,197	0,287	0,351	0,341	0,267
M. Le Pen	0,167	0,366	0,267	0,271	0,181	0,17	0,289
J-L. Mélenchon	0,276	0,198	0,181	0,221	0,231	0,316	0,15
E. Zemmour	0,064	0,095	0,192	0,07	0,079	0,05	0,11
V. Pécresse	0,041	0,025	0,063	0,044	0,05	0,009	0,047
Y. Jadot	0,036	0,023	0,03	0,031	0,038	0,029	0,032
J. Lassalle	0,03	0,016	0,018	0,023	0,021	0,021	0,02
R. Roussel	0,003	0,017	0,02	0,018	0,022	0,028	0,029
N. Dupont-Aignan	0,024	0,023	0,015	0,017	0,007	0,012	0,03
A. Hidalgo	0,009	0,007	0,005	0,009	0,008	0,013	0,01
P. Poutou	0,006	0,006	0,006	0,005	0,007	0,008	0,01
N. Arthaud	0,003	0,004	0,006	0,003	0,005	0,003	0,006
	100,0%	100,1%	100,0%	99,9%	100,0%	100,0%	100,0%

On note que les 100.1% et 99.1% sont explicables par les approximations des parts de vote à un chiffre significatif

4 Modèle

4.1 Choix du modèle

Puisque les données sont sous la forme d'entrées numériques d'un tableau à double-entrée, il paraît judicieux d'effectuer une analyse des correspondances du tableau des parts de vote.

4.2 Présentation du modèle

L'Analyse Factorielle est la pierre angulaire de l'analyse des correspondances. Les résultats d'une analyse des correspondances permettent d'explorer la structure expliquant le mieux possible les données catégorielles, ainsi que des métriques justifiant le crédit accordable aux analyses pouvant être effectuées.

Pour comprendre plus en profondeur cette méthode, le détail des calculs et les différentes applications possibles, on peut se référer à l'ouvrage de Greenacre (1984). Ces méthodes sont imaginées, en France, par Jean-Paul Benzécri au début des années 1960 et dans les années 1970 (on peut citer Benzécri(1973) ainsi que Lebart, Morineau, et Tabard(1977)).

4.3 Critique du modèle

Ici le travail est effectué avec des proportions de votes, c'est à dire que la somme de chaque colonne du tableau est égale à 100. Cette information peut sembler superflue mais elle est perdue dans l'analyse des correspondances. Entre un candidat à 50% et à 85%, il y a un écart bien plus important, presque adapté à une échelle logarithmique, que laisse penser la différence entre les nombres 0,85 et 0,50 qui est même plus petite qu'un facteur 2.

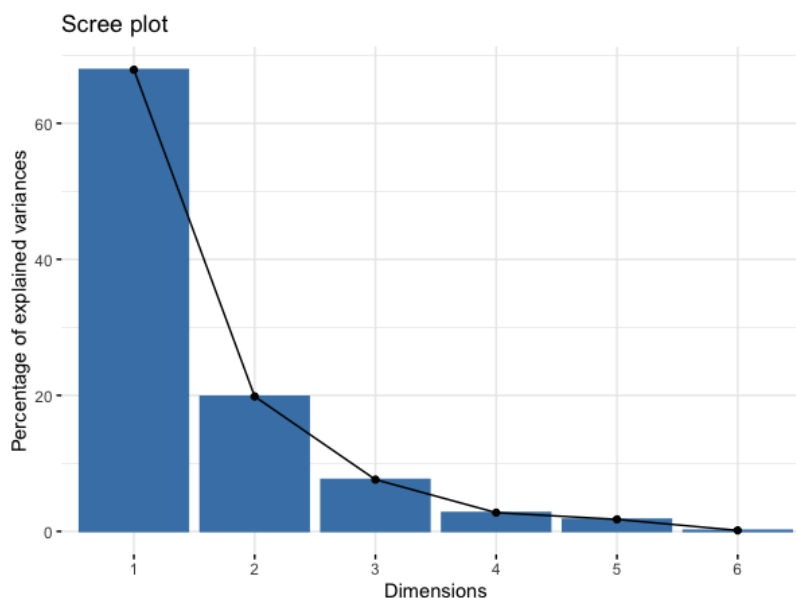
5 Analyse des résultats

5.1 Traitement informatique

Toutes le support informatique est offert par le **language R** épaulé des librairies **FactoMineR**, **factoextra**, **ggplots**.

5.2 Analyse des données et interprétation des résultats

On commence par effectuer l'analyse en composantes principales du jeu de données pour mieux comprendre s'il existe une forme de structure pouvant justifier l'analyse en correspondances sur le premier plan principal. L'histogramme suivant montre la variance expliquée par ordre de valeur propre. Le premier couple propre explique 87,8%, ce qui est très convaincant pour la suite des analyses.



Variance expliquée par ordre de valeur propre

Les deux histogrammes suivants montrent l'importance dans le calcul du premier plan principal. On peut remarquer que les candidats qui semblent être le mieux expliqués par la géologie présente chez leur électeurat. Peut-être pourra-t-on lire que les sols riches en charbon du Nord, fief du Rassemblement

National, ou que les sols riches de l'ancien Océan qui couvrait la Provence il y a des centaines de millions d'années ont conditionné la vie des habitants, justifiant les chiffres du RN dans la région PACA. On remarque aussi que les sols les plus déterminants sont le tertiaire ainsi que le granite, ce qui est surprenant car il est assez peu présent en France. Le prochain graphique montre la qualité de

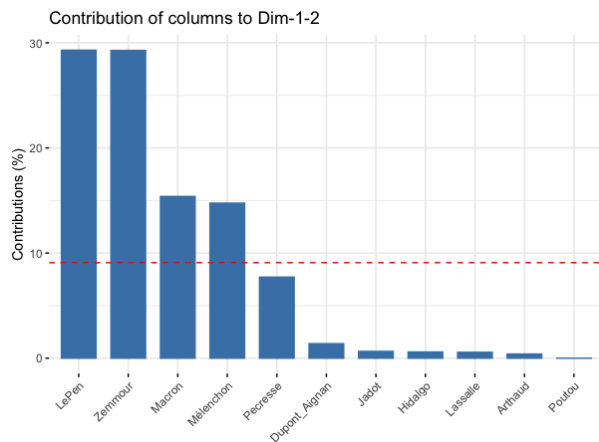


Figure 3: Influence des candidats dans le calcul du premier plan principal

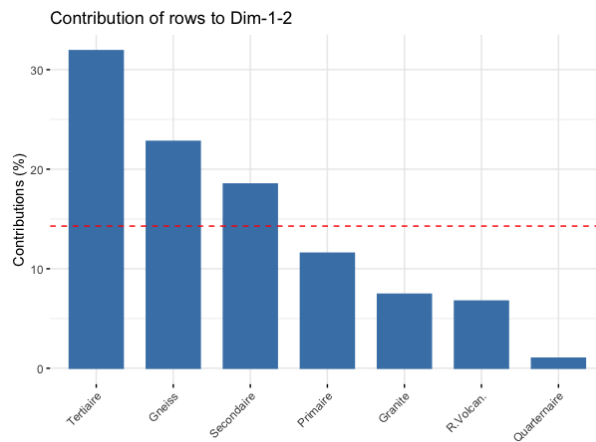
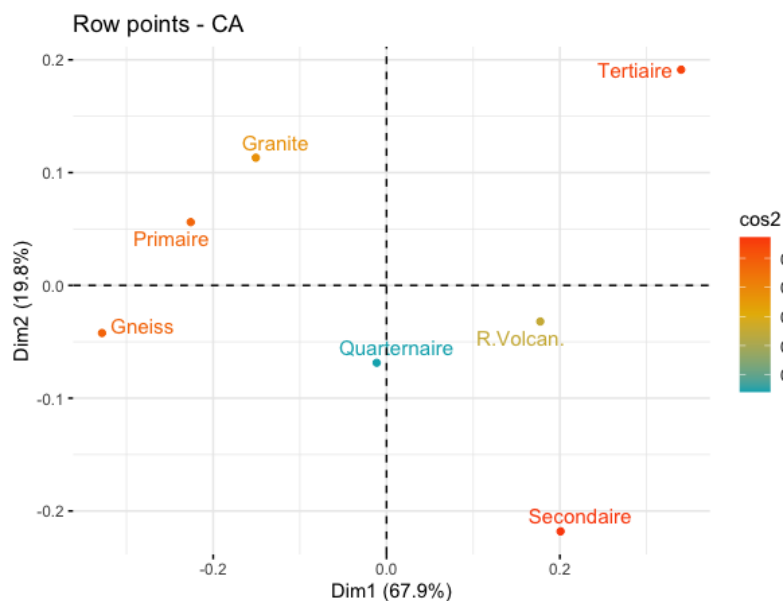


Figure 4: Influence de la géologie dans le calcul du premier plan principal

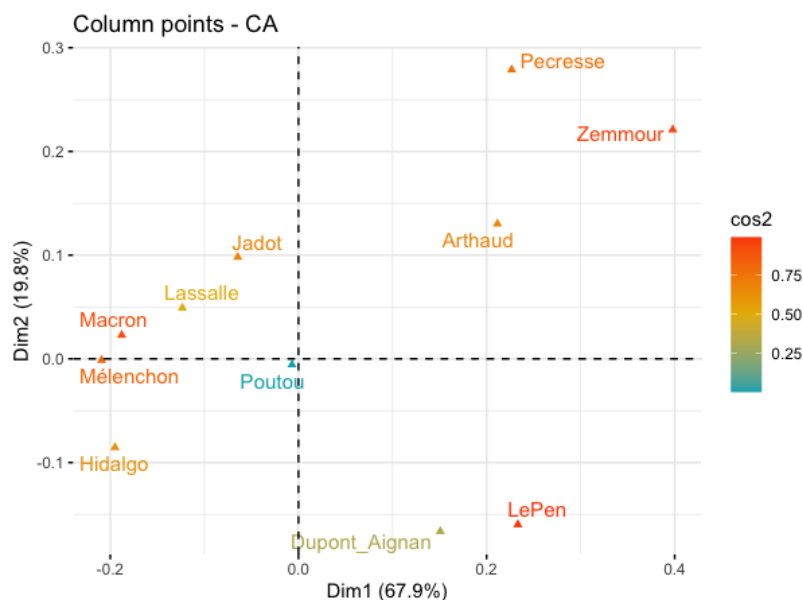
représentation des différents types de sol. Couplé à la figure 3, ce graphique montre bien que les sols tertiaires sont très bien représentés, suivis par les sols secondaires, le Gneiss (lui aussi très peu présent en France), et les sols primaires. Il est très difficile de dégager deux interprétations géologiques des axes principaux sans un expert.



Analyse des correspondances

Le coeur de l'étude porte sur l'explicabilité des votes. Le graphique ci-dessous permet de comprendre quels sont les candidats les plus sujet au facteur géologique. Puisque les villes sont des berceaux de

population et que l'électorat principal d'E. Macron ainsi que de J-L. Mélenchon y est niché, seront-ils impactés par la géologie ? On peut supposer que les villes ne sont pas implantées dans des cantons dont la géologie Préhistorique a une grande influence : ce serait négliger l'apport de la région alentour dans le développement de ces dernières. Ici on remarque la très juste représentation des candidats plébiscités par les campagnes : M. Le Pen et É. Zemmour. On peut lire que E. Macron et J-L. Mélenchon sont bien représentés, mais ce serait oublier que ce sont aussi des candidats favoris.

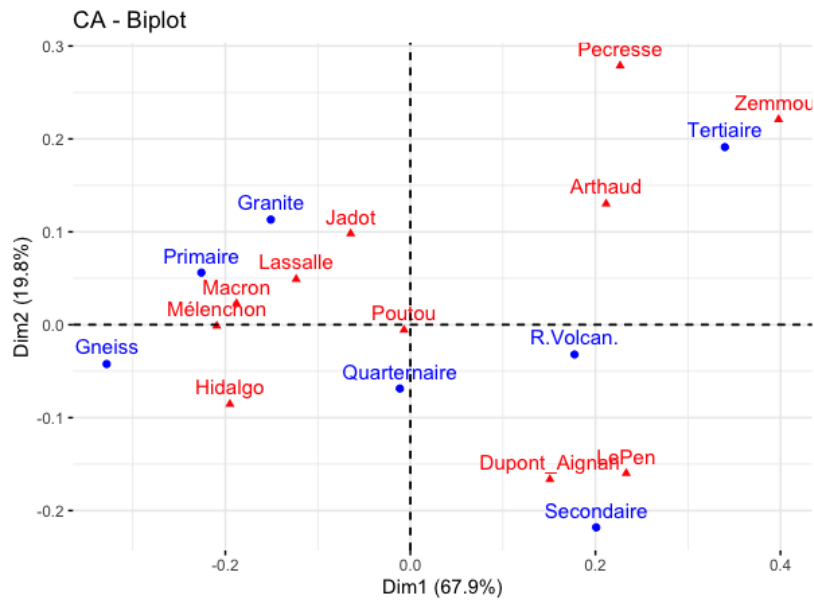


Analyse des correspondances

Le dernier graphique permet enfin de mettre tous les anciens calculs en relief : nous pouvons comprendre les pôles d'attractivité géologique de l'électorat de chaque candidat.

On peut déjà lire que les électors d'E. Macron et J-L. Mélenchon sont très proches : ces deux candidats sont majoritaires en Vendée par exemple, où les sols sont majoritairement primaires. Ils sont épaulés par A. Hidalgo. On peut essayer de comprendre un certain bloc de Gauche sur la gauche du graphique.

Aussi, M. Le Pen et N. Dupont-Aignan, plutôt proches sur l'échiquier politique, se retrouvent voisins ici aussi, portés par les sols du Secondaire. E.Zemmour est aussi à droite du graphique, ce qui peut laisser supposer un axe horizontal, qui rappelons le explique près de 68% des données, serait un axe Gauche-Droite édulcoré ? Bien que cette notion ne soit plus d'actualité (E. Macron et J-L. Mélenchon ne sont pas du tout du même bord politique, mais peut-être touchent la même population), on la retrouve en partie dans cette étude. Par ailleurs, on ne sait pas expliquer l'électorat de P. Poutou. Cependant, E.Zemmour est très proche des sols de l'ère tertiaire mais de façon surprenante, V. Pécresse, dont l'électorat est à l'opposé de celui d'É. Zemmour, se retrouve proche de ce dernier. On découvre alors le pot-au-roses : comment justifier la proximité entre E. Zemmour et V. Pécresse ? Les conclusions de cette étude sont à prendre avec de grandes précautions.



Analyse des correspondances

6 Conclusion

Il est difficile de tirer une conclusion : d'un côté, certains résultats semblent probants (grande explication des données dans le premier plan principal), mais surtout il semblerait que la répartition des votes selon la géologie des sols ne soit pas si facilement explicable. Quelques informations sont bonnes à prendre : proximité des électors de N. Dupont-Aignan et M. Le Pen avec les sols secondaires, par exemple. Mais l'analyse poussée de cette étude requiert des compétences de géologie bien plus poussées, il faudrait aussi croiser ces données avec les données économiques associées à chaque région. On pourrait ainsi étudier et élaborer en profondeur un modèle bien plus complexe qui dépasse le cadre de cours.

7 Bibliographie

<https://stevedutch.net/research/elec2000/geolelec2000.htm>
data.gouv.fr
fr.statista.com/statistiques/7039799067777/.
 Greenacre (1984)
 Benzécri(1973)
 Lebart, Morineau, et Tabard (1977)