

## Projet Analyse de données

ANDONISSAMY Rémi  
MAMIHARILALA Judicha  
FAZUL Loutfi

### I.Introduction

Notre base de données porte sur les élections présidentielles de 2022 en France métropolitaine. Cette base nous transmet différentes informations sur ces élections, pour chaque ville de chaque département, telles que : le nombre d'inscrits par ville, le nombre de votants ou encore le nombre de votes relatif à chacun des candidats du premier et du deuxième tour.

Notre problématique est la suivante : Quelles sont les tendances politiques en fonction des zones géographiques en France métropolitaine ?

Pour répondre à notre problématique nous allons considérer seulement les variables relatives aux pourcentages de voix, et mettre les pourcentage de voix du second tour en variables supplémentaires. De plus, en raison du nombre de villes qui est conséquent, nous allons plutôt considérer la moyenne des votes par département (en se basant sur les votes de leurs villes), plutôt que de considérer les votes de toutes les villes.

### II. Statistique descriptive/bivariée

#### Statistiques descriptives univariées

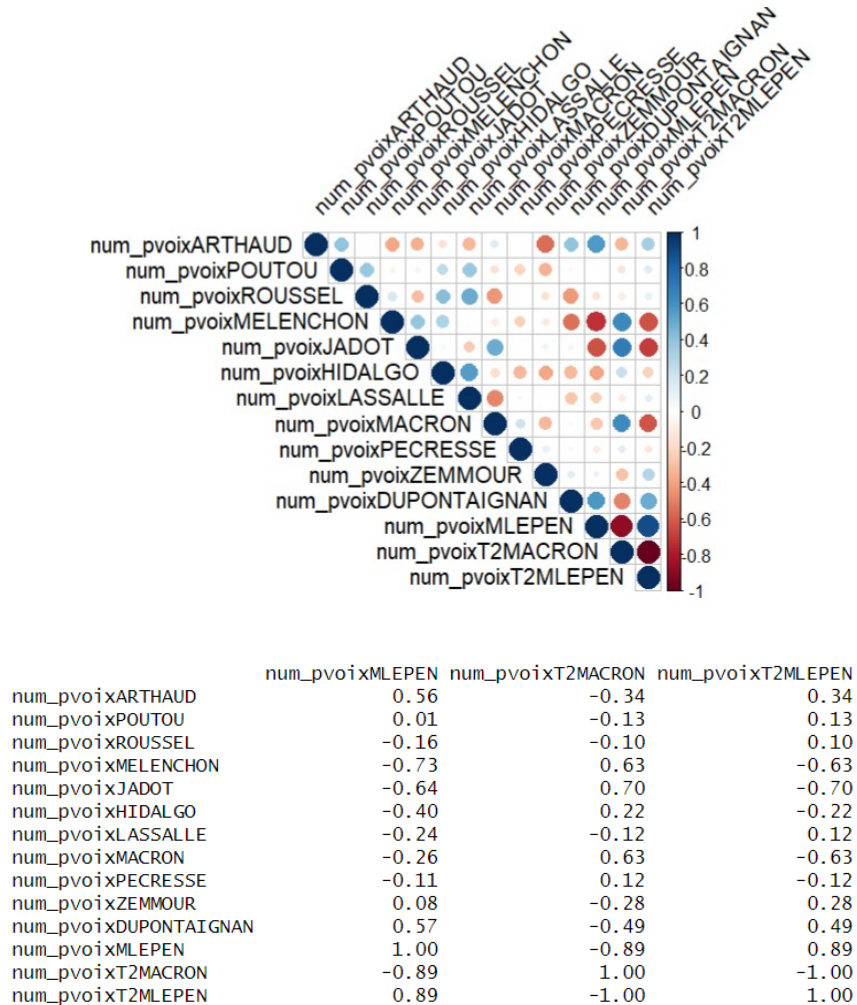
```
> #Moyenne des pourcentage de voix pour Emmanuel Macron et Marine Le Pen au second tour.  
> mean(sum_dep$num_pvoixT2MACRON, na.rm = TRUE)  
[1] 0.5170778  
> mean(sum_dep$num_pvoixT2MLEPEN, na.rm = TRUE)  
[1] 0.4829222  
> #Médianes  
> median(sum_dep$num_pvoixT2MACRON, na.rm = TRUE)  
[1] 0.5058715  
> median(sum_dep$num_pvoixT2MLEPEN, na.rm = TRUE)  
[1] 0.4941285
```

Lors du second tour qui oppose Emmanuel Macron à Marine Le Pen, environ 52% des votes étaient en faveur d'Emmanuel Macron, contre environ 48% pour Marine Le Pen.

La médiane du pourcentage de votes pour Emmanuel Macron au second tour est d'environ 51 %. Cela signifie qu'il y a 50 % des départements qui ont un pourcentage de votes pour Macron inférieur à 51 %, et 50 % ont un pourcentage de votes supérieur à 51 %.

La médiane du pourcentage de votes pour Marine Le Pen au second tour est d'environ 49 %. C'est-à-dire qu'il y a 50 % des départements qui ont un pourcentage de vote pour Marine Le Pen inférieur à 49 %, et 50 % des départements qui ont un pourcentage de votes supérieur à 49 %.

### Matrice de corrélation



La matrice de corrélation permet de voir les corrélations entre les variables, c'est-à-dire les pourcentages de voix par candidat. Par exemple le pourcentage de voix pour Jean-Luc Mélenchon et le pourcentage de voix pour Marine Le Pen sont fortement corrélés négativement (-0.73). On peut donc se dire qu'un département qui a tendance à principalement voter pour Marine Le Pen, n'aura pas tendance à également voter pour Jean Luc Mélenchon, ce qui peut montrer une forte opposition entre ces deux candidats d'un point de vue politique.

Un autre exemple serait la corrélation entre le pourcentage de voix au premier tour pour Yannick Jadot, et le pourcentage de voix pour Emmanuel Macron au deuxième tour, il s'agit d'une corrélation positive plutôt élevée (0.70). Cela peut vouloir dire qu'un département qui a principalement voté pour Yannick Jadot au premier tour, a eu tendance à plutôt voter pour Emmanuel Macron au second tour.

### III. Analyse géométrique des données

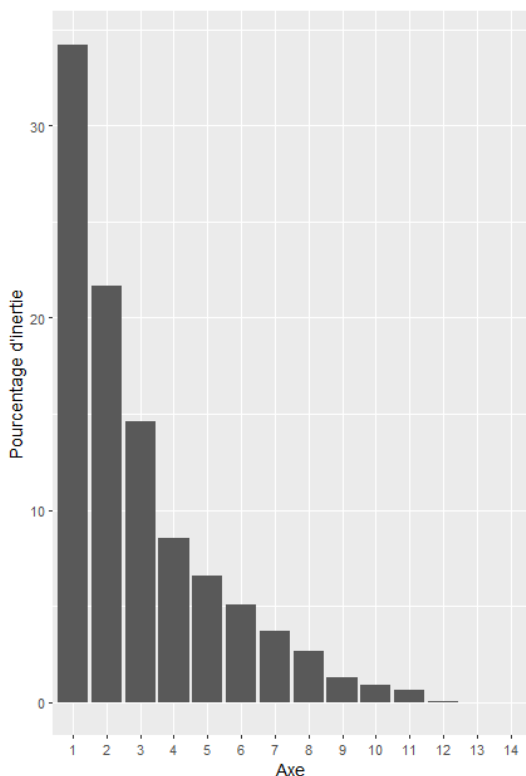
#### Quel type d'analyse factorielle ? :

On utilise ici une Analyse en Composante Principale (ACP), car nous travaillons sur un tableau contenant un grand nombre de données quantitatives telles que les pourcentages de voix relatifs aux candidats politiques pour toutes les villes.

Nous voulons observer les tendances politiques des différentes zones géographiques de la France, en nous basant sur ces données quantitatives. L'ACP est particulièrement adaptée pour ce que l'on cherche à faire.

#### Quels axes retenir ?

Histogramme des valeurs propres



Sur cet histogramme des valeurs propres nous pouvons constater que la pente du graphique chute à partir du quatrième axe. Si l'on se base sur le critère du coude, il est pertinent de retenir les trois premiers axes.

Il est également possible d'utiliser le critère de Kaiser pour savoir quels axes retenir. D'après ce critère, il faut garder les axes qui ont une valeur propre supérieure à 1.

On peut constater que l'axe 1 (v.p = 4.79), l'axe 2 (v.p = 3.03), l'axe 3 (v.p = 2.04) et l'axe 4 (v.p = 1.19), sont les axes avec une valeur propre supérieure à 1.

Le critère de Kaiser étant plus précis, nous pouvons donc retenir les quatre premiers axes.

Les quatre premiers axes résument environ 79 % de l'inertie totale. (34.22 + 21.66 + 14.63 + 8.51)

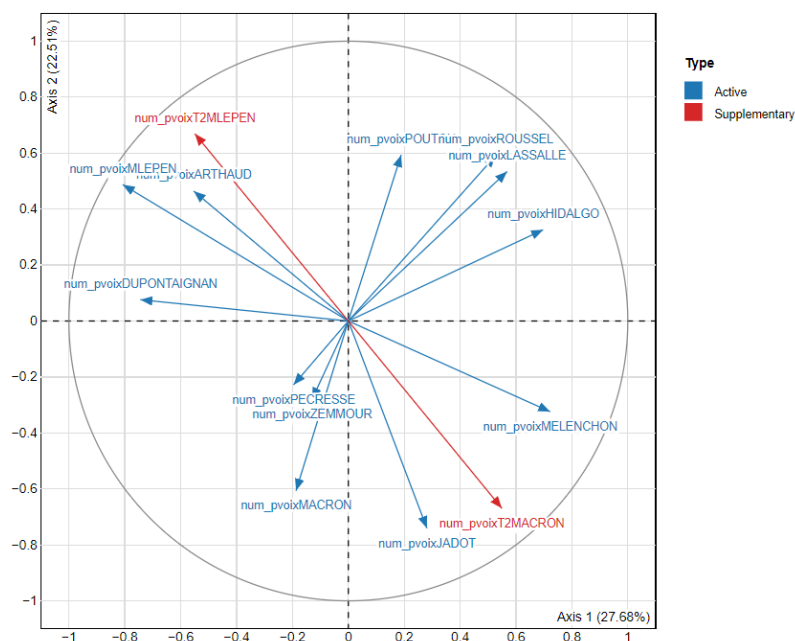
|         | eigenvalue   | percentage of variance | cumulative percentage of variance |
|---------|--------------|------------------------|-----------------------------------|
| comp 1  | 4.790523e+00 | 3.421802e+01           | 34.21802                          |
| comp 2  | 3.032973e+00 | 2.166409e+01           | 55.88211                          |
| comp 3  | 2.048275e+00 | 1.463054e+01           | 70.51265                          |
| comp 4  | 1.191996e+00 | 8.514254e+00           | 79.02691                          |
| comp 5  | 9.171584e-01 | 6.551132e+00           | 85.57804                          |
| comp 6  | 7.158468e-01 | 5.113191e+00           | 90.69123                          |
| comp 7  | 5.238615e-01 | 3.741868e+00           | 94.43310                          |
| comp 8  | 3.766466e-01 | 2.690333e+00           | 97.12343                          |
| comp 9  | 1.841241e-01 | 1.315172e+00           | 98.43860                          |
| comp 10 | 1.223559e-01 | 8.739704e-01           | 99.31257                          |
| comp 11 | 8.892487e-02 | 6.351776e-01           | 99.94775                          |
| comp 12 | 7.315071e-03 | 5.225051e-02           | 100.00000                         |
| comp 13 | 1.242163e-16 | 8.872595e-16           | 100.00000                         |
| comp 14 | 9.080111e-17 | 6.485794e-16           | 100.00000                         |

```
> sum(ACP$eig[,1])
```

```
[1] 14
```

```
> explor(ACP)
```

## Interprétation des résultats



**L'axe 1** résume 27.68 % de l'inertie totale. Cet axe semble opposer les candidats de gauche (à droite du cercle) aux candidats de droite (à gauche du cercle).

**L'axe 2** quant à lui, résume 22.51 % de l'inertie totale. Cet axe semble opposer, en haut, les candidats proches des préoccupations des classes populaires et des travailleurs, et, en bas, les candidats moins proches de ce type de préoccupations.

**Pour l'axe 1**, on trouve majoritairement des départements situés au sud (Hérault, Aveyron, Gironde, Pyrénées-Atlantique...). On peut donc dire que, en France métropolitaine, la plupart des départements du sud votent pour des partis politiques de gauche.

Les départements à gauche de l'axe sont plutôt situés au centre et au nord de la France (Yvelines, Eure-et-Loire, Seine-et-Marne, Aube...).

La plupart des départements du nord de la France votent pour des partis politiques de droite.

**Pour l'axe 2**, les départements du bas sont plutôt situés au nord de la France (Yvelines, Seine-Saint-Denis, Val-de-Marne...).

A l'opposé de cela nous avons en haut, des départements qui sont plutôt situés au centre et au nord de la France (Haute-Vienne, Aube, Indre-et-Loire...). On n'observe donc pas une très grande différence en fonction de la géographie pour l'axe 2, contrairement à l'axe 1.

## IV. Conclusion

Pour conclure, nos analyses, notamment via l'ACP, ont mis en lumière des tendances politiques globales en France lors du premier tour des élections présidentielles de 2022. La France semble divisée en deux grandes zones :

- Au nord, une nette tendance en faveur des candidats de droite ou d'extrême droite.
- Au sud, une préférence plus marquée pour les candidats de gauche.

Cette polarisation nord-sud pourrait refléter des disparités socio-économiques et culturelles nécessitant d'autres études.