

PROJET PLS

Modélisation des résultats présidentiels 2022

Enseignant :
Christian DERQUENNE

Auteurs
Anas HAOUD
Mamadou DIEDHIOU

Question 1 :

Explorer les données avec les statistiques descriptives (univariées et bivariées) et fournir un état des lieux de celles-ci (univariée, scatterplots, cartographie des corrélations), sur les nombres bruts et sur les proportions. Quelles différences de résultats remarquez-vous ?

Réponse :

➤ Statistiques descriptives

Avant de se lancer dans l'analyse des données, on a d'abord vérifié s'il y avait des valeurs manquantes dans notre base. Résultat : aucune variable n'est concernée par ce problème, ce qui est une bonne nouvelle. La majorité des variables sont numériques, sauf le **Numéro de département** et le **Nom de département**, qui sont des chaînes de caractères.

On a ensuite regardé les statistiques descriptives des variables du premier tour, et plusieurs éléments ressortent :

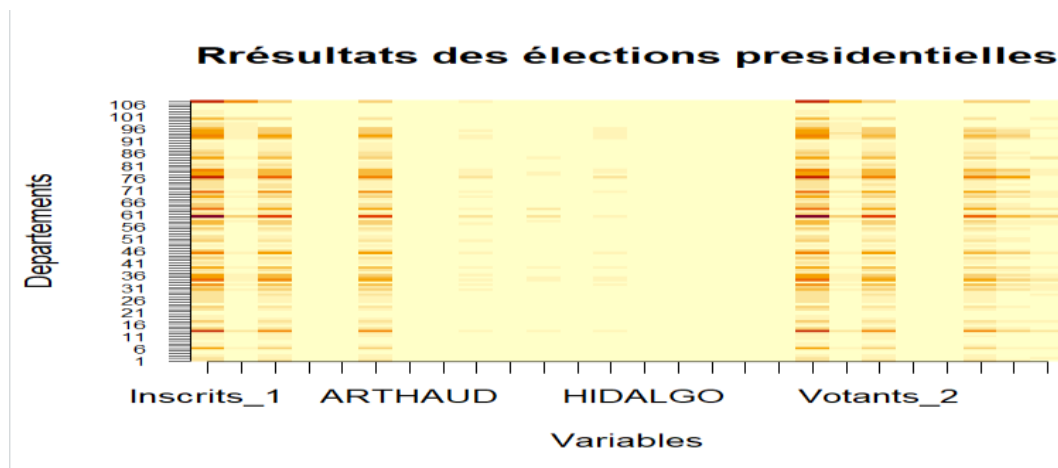
- Le nombre d'inscrits varie énormément d'un département à l'autre, allant de **5 045** à **1 813 906**, avec une moyenne de **455 588 inscrits**. Une telle différence montre qu'il y a une vraie inégalité dans la taille des électorats départementaux. Le **Nord** est le département avec le plus grand nombre d'inscrits, alors que **Saint-Pierre-et-Miquelon** est celui qui en compte le moins.
- Du côté des **abstentions**, la moyenne est de **119 852 personnes**, avec un minimum de **2 272** et un maximum de **931 455 abstentions**. Comme on pouvait s'y attendre, **les Français de l'étranger** affichent un taux d'abstention plus élevé, alors que **Saint-Pierre-et-Miquelon** enregistre peu d'abstentions.
- Concernant le **nombre de votants** (y compris bulletins blancs et nuls), il s'élève en moyenne à **335 736 personnes**, avec une variation de **2 773** à **1 299 873** votants selon les départements. Le **Nord** arrive encore une fois en tête, tandis que **Saint-Pierre-et-Miquelon** se retrouve en bas du classement.
- Pour les **bulletins blancs et nuls**, on remarque que les **votes blancs** sont plus fréquents, avec une moyenne de **5 080**, contre **2 310** pour les votes nuls.
- Enfin, en ce qui concerne les **votes exprimés** (c'est-à-dire les votes valides, après retrait des blancs et nuls), la moyenne est de **328 345 votes**, avec des extrêmes allant de **2 701** à **1 274 781** voix. Cela montre que les choix électoraux varient fortement selon les départements.

En regardant les résultats du scrutin, on remarque que le **Nord** est le département où il y a eu le plus de voix exprimées, tandis que **Saint-Pierre-et-Miquelon** en a eu le moins.

Au **premier tour**, Le Pen a obtenu le plus de voix dans le Nord, alors que **Arthaud** a enregistré le plus faible score.

Lors du **second tour**, la répartition reste similaire en termes d'inscrits et de votants, mais cette fois, **Macron** arrive en tête dans le Nord, contrairement au premier tour où c'était **Le Pen**.

➤ Cartographie avant standardisation



➤ Visualisation des données brutes

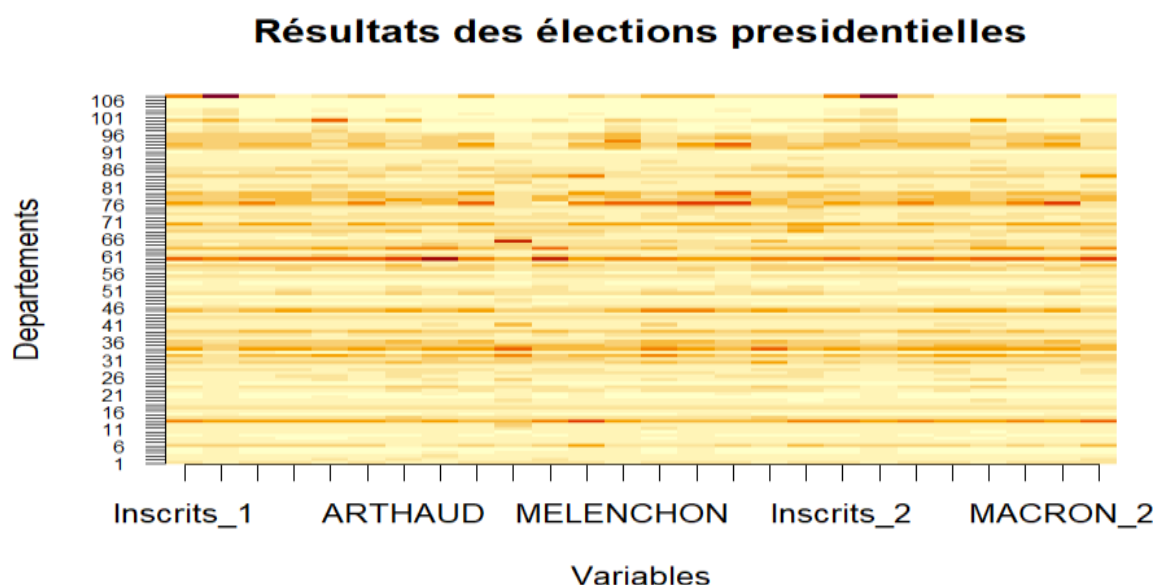
On a ensuite cartographié les données avant toute transformation pour voir leur répartition sur l'ensemble des départements. Ce qu'on remarque, c'est qu'il existe des valeurs atypiques qui se démarquent des tendances générales.

Les couleurs utilisées sur la carte nous permettent de mieux visualiser ces écarts :

- **Les tons clairs (jaune pâle)** correspondent aux valeurs faibles.
- **Les teintes foncées (marron foncé)** indiquent des valeurs plus élevées.
- **Les valeurs extrêmes**, qu'elles soient très hautes ou très basses, apparaissent en **très clair ou très foncé**, ce qui met en évidence des disparités significatives entre les départements.

On peut donc déjà identifier certains départements qui sortent du lot avant même d'avoir appliqué des transformations aux données.

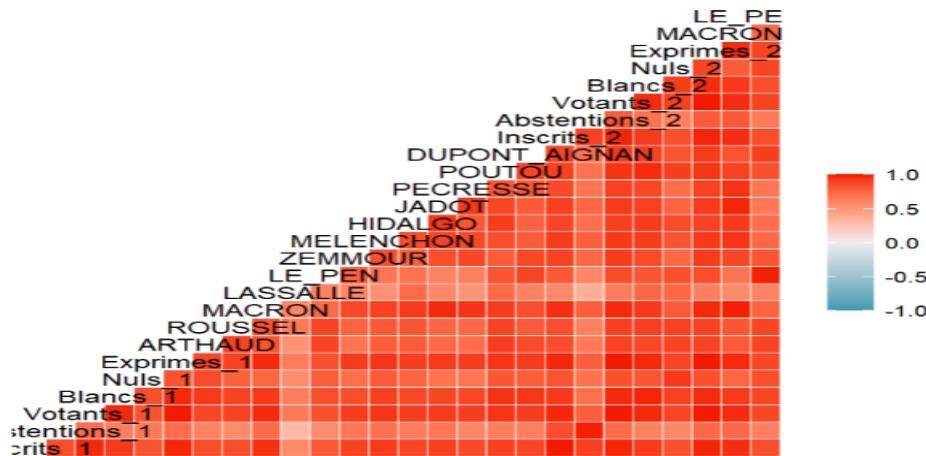
➤ Cartographie après standardisation



Cartographie des données standardisées

Après avoir appliqué la standardisation aux données, on observe des changements notables dans leur visualisation. Les valeurs atypiques, qui ressortaient nettement dans la base brute, sont désormais moins nombreuses. Cela permet d'obtenir une représentation plus homogène des variables et d'éviter que certaines valeurs extrêmes influencent trop les analyses.

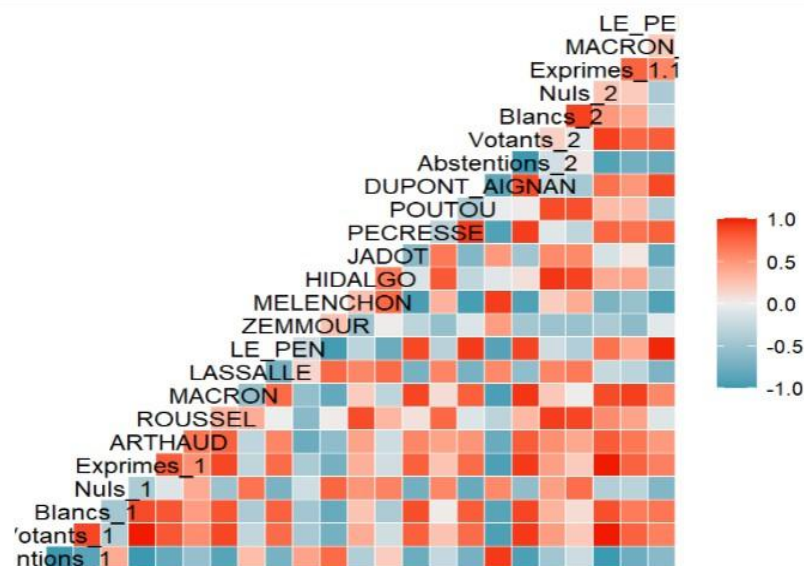
➤ Matrice de corrélation standardisée sans proportion



Matrice de corrélation des données standardisées sans pourcentage

Une fois les données standardisées, nous avons généré une matrice de corrélation afin d'analyser les relations entre les différentes variables. Celle-ci montre des **corrélations très fortes et positives** entre certaines variables. Cela signifie que plusieurs indicateurs évoluent de manière similaire, ce qui confirme la présence de **multicolinéarité** dans notre dataset.

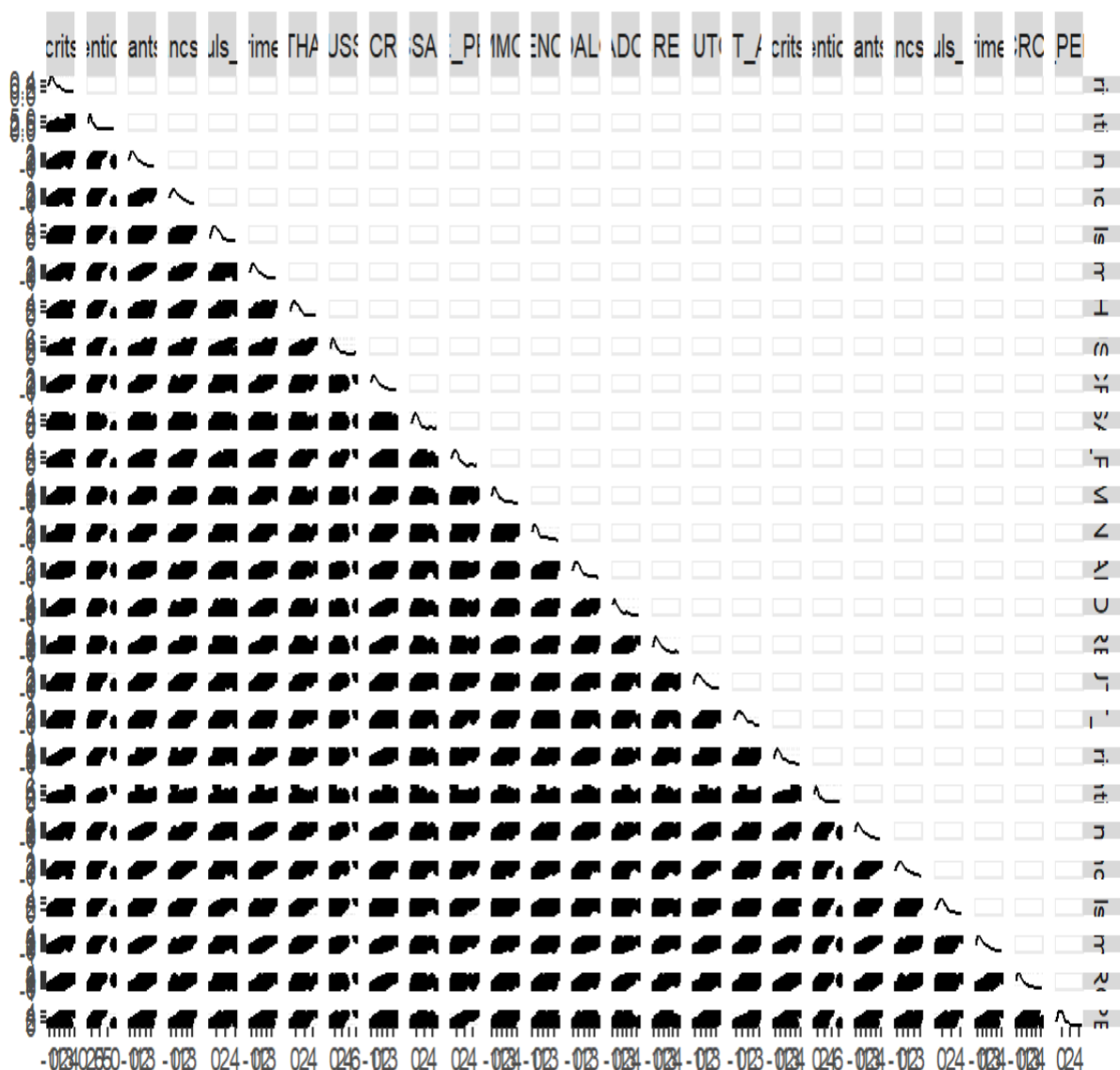
➤ Matrice de corrélation standardisée avec proportion



Matrice de corrélation des données standardisées en pourcentage

En recalculant les corrélations sur les **données exprimées en proportion**, on observe des différences notables. Contrairement à la matrice précédente, où la majorité des variables étaient fortement corrélées entre elles, ici, certaines relations sont beaucoup plus faibles, tandis que d'autres restent significatives. Cette transformation permet de mieux distinguer les tendances réelles en évitant que les effets de grandeurs absolues biaisent l'analyse des corrélations.

➤ Scatterplot



Question 2 :

Sur chacun des deux tours puis sur les deux tours simultanément, quels types d'analyses multidimensionnelles peut-on pratiquer ? Justifier vos choix et appliquer les méthodes retenues.

Réponse :

➤ Analyse en Composantes Principales (ACP)

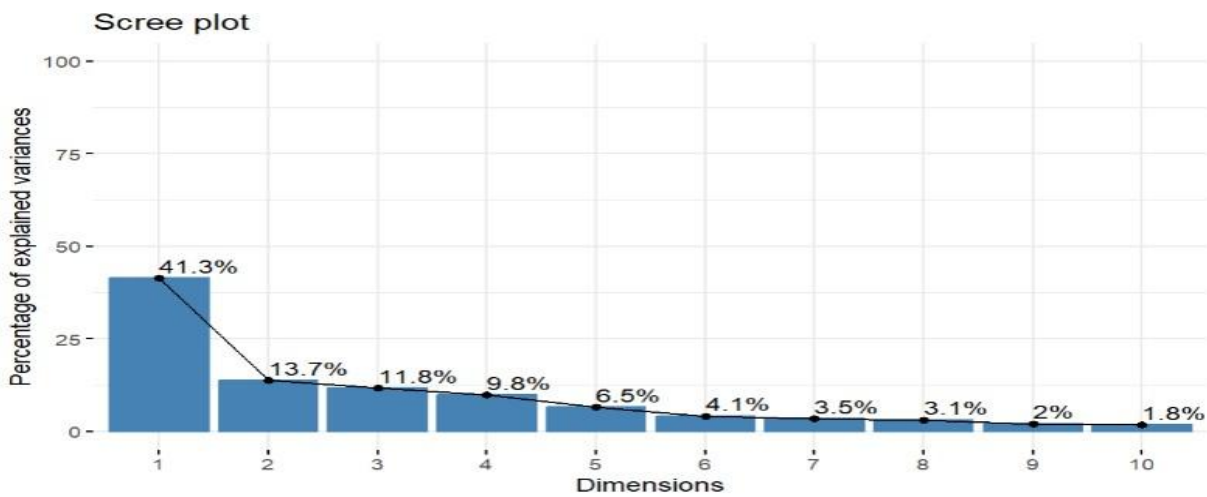
Pour analyser nos données, on a choisi d'appliquer une Analyse en Composantes Principales (ACP). Comme toutes les variables sont numériques, cette méthode est idéale pour réduire la dimensionnalité tout en conservant un maximum d'informations.

➤ Premier tour

D'après le scree plot, on observe que :

- La première composante principale (Dim1) explique 41,3 % de la variance totale.
- La deuxième composante principale (Dim2) en explique 13,7 %.

Avec un total de 17 variables, la méthode du coude nous indique que ces deux premières composantes résument une grande partie des informations contenues dans les données.

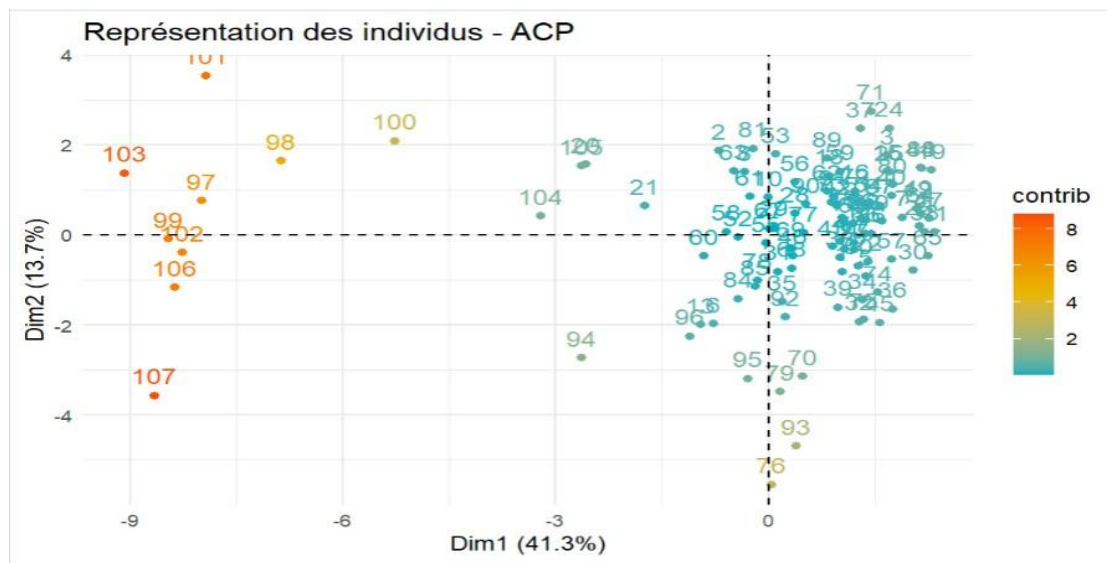


Pourcentage de variance expliquée sur les données du premier tour

➤ Interprétation des contributions :

Les cercles plus grands sur la carte des contributions montrent les variables qui influencent le plus chaque axe.

Plus on avance vers des dimensions supérieures, plus leur impact devient moins significatif.



➤ Analyse des départements :

Certains départements, situés près de l'origine (en bleu clair), ont une faible contribution à l'analyse.

D'autres, plus éloignés, jouent un rôle clé dans la variance, notamment les départements 102, 106, 103 et 99 pour Dim1 et 76, 93, 97 pour Dim2.

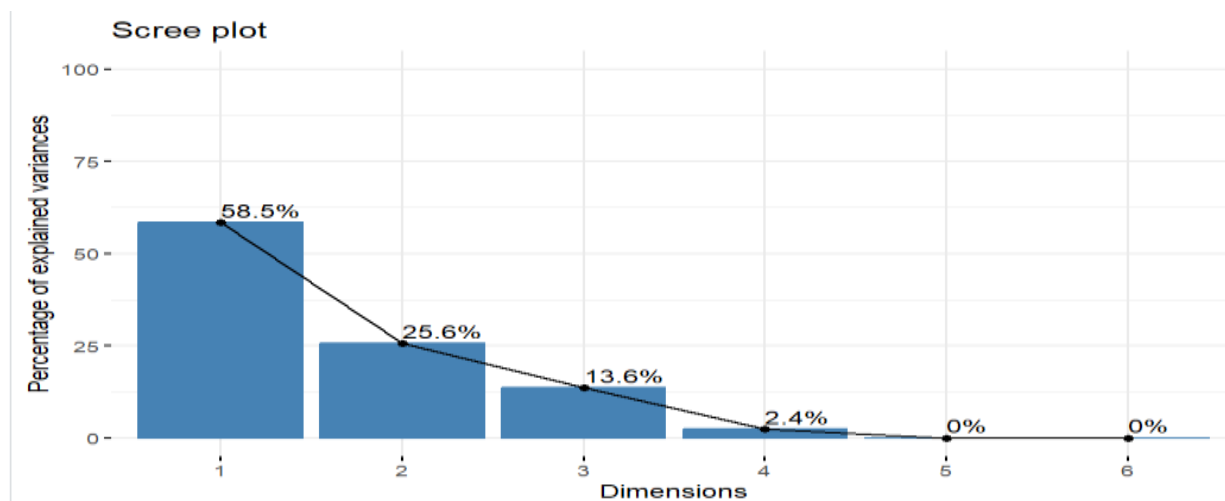
Les départements 107 et 101, situés loin des deux axes, ont une influence notable sur l'ensemble de la variance.

➤ Deuxième tour

En passant au second tour, les résultats évoluent :

La première composante principale explique 58,5 % de la variance totale, alors que la deuxième composante en explique 25,6 %.

Ces deux composantes résument donc une large majorité de l'information contenue dans les 7 variables du second tour.

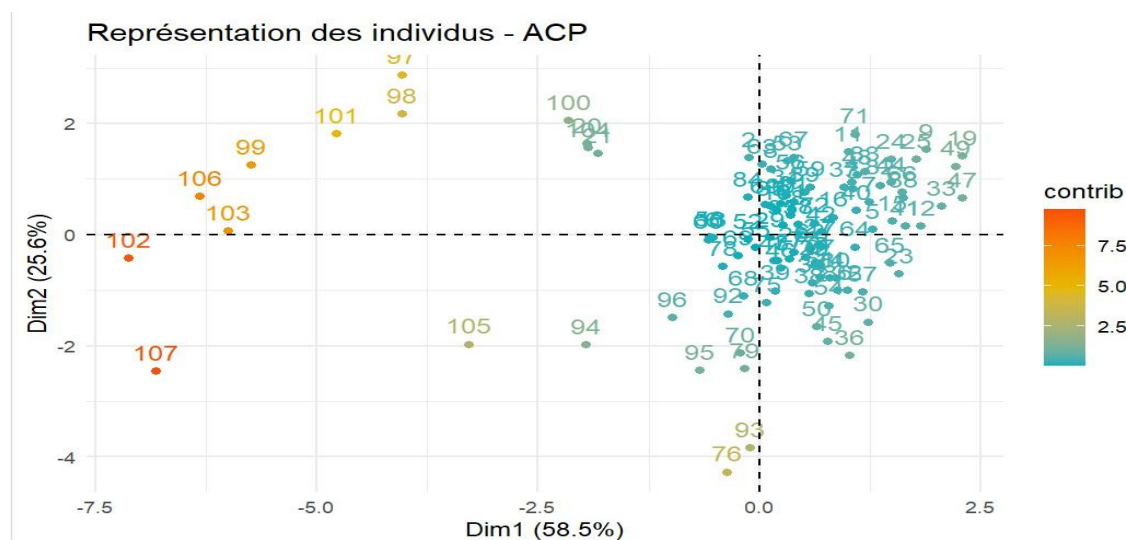
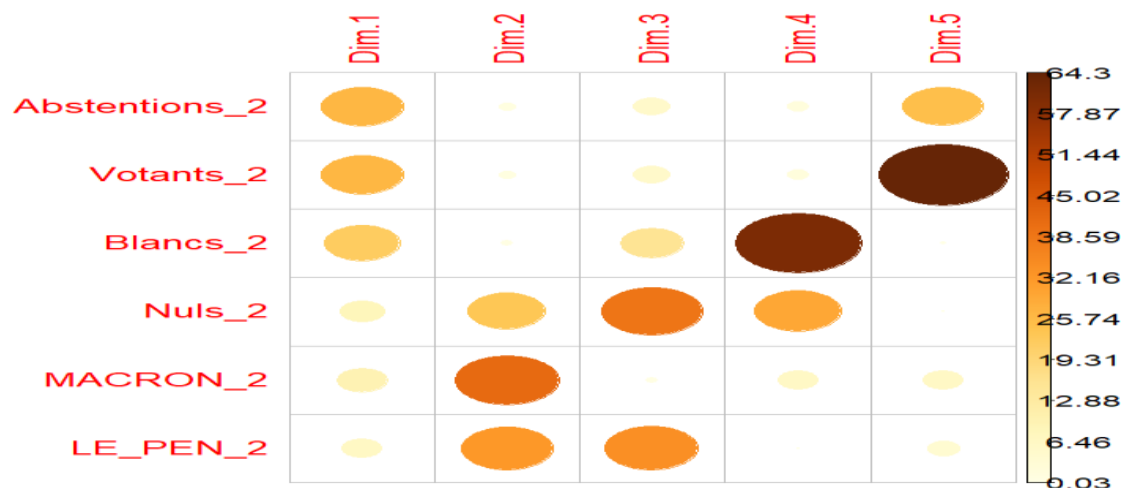


➤ Contribution des variables :

Comme au premier tour, la répartition des contributions est cohérente avec le scree plot.

Certains départements restent faiblement impliqués (ceux en bleu clair), alors que d'autres, comme 102, 106, 103 et 99, influencent fortement Dim1, et 76, 93 et 97 contribuent davantage à Dim2.

Encore une fois, 107, 101, 97 et 98 se démarquent avec des contributions élevées sur plusieurs axes.

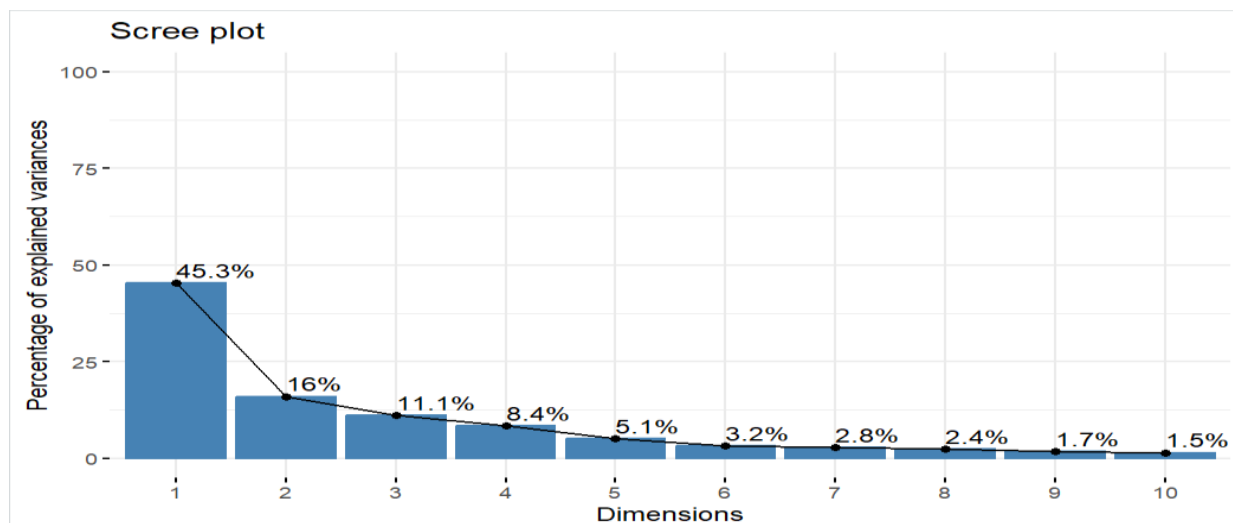


➤ Analyse combinée des deux tours

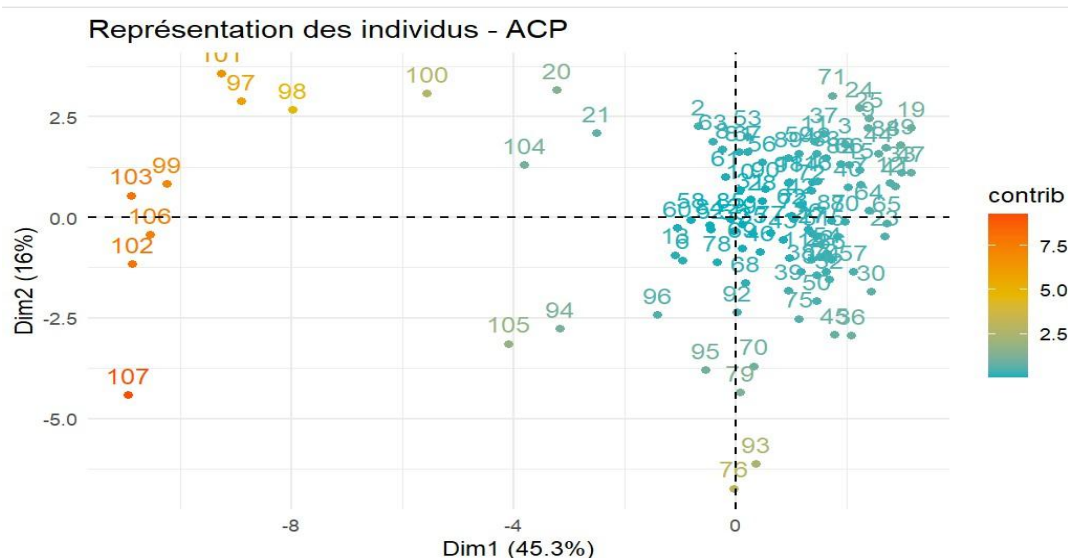
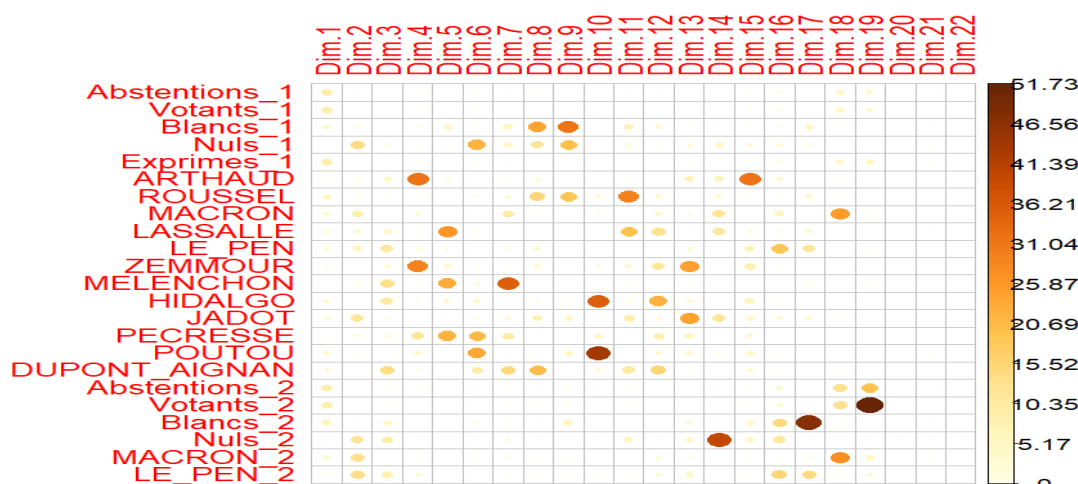
Quand on prend en compte les deux tours simultanément, le scree plot nous montre que :

La première composante principale explique 47,1 % de la variance.

La deuxième en explique 15,3 %.



Cela confirme que ces deux premières dimensions captent l'essentiel des tendances électorales des deux tours combinés.



➤ Observations finales :

Les départements situés près du centre de l'axe ont une contribution faible.

Ceux qui sont éloignés de Dim1 (102, 106, 103, 99) ou de Dim2 (107, 101, 97) jouent un rôle clé dans les variations observées.

Question 3

Correlation

```
##               cor_Abstentions_2 cor_Blancs_2 cor_MACRON_2 cor_LE_PEN_2
## Abstentions_1      0.82430342  -0.63881671  -0.51845232   0.24287345
## Blancs_1           -0.39378083   0.53352328   0.09938664  -0.01735018
## Nuls_1              0.01738509   0.20285398  -0.42293649   0.34118092
## ARTHAUD            -0.44058516   0.09268787  -0.08341726   0.29273363
## ROUSSEL            -0.19639415   0.63338212  -0.14868402   0.09002858
## MACRON             -0.49891716   0.01114929   0.88972810  -0.64985741
## LASSALLE           -0.16981695   0.68748587  -0.22501178   0.14606409
## LE_PEN             -0.16203150  -0.18818340  -0.81634553   0.97091923
## ZEMMOUR            0.35730747  -0.28495060  -0.32757199   0.22224083
## MELENCHON          0.19582248   0.20002183   0.51747339  -0.67777978
## HIDALGO            -0.55751403   0.78010002   0.30554686  -0.21041426
## JADOT              -0.22606624   0.09817222   0.83564709  -0.74852696
## PECRESSE           -0.16863333   0.17825304   0.37862875  -0.35048254
## POUTOU             -0.43245235   0.54679601   0.04426909   0.05272973
```

Ici nous faisons la corrélation entre les variables cibles et les variables explicatives afin de pouvoir faire des comparaisons avec les régressions.

ols

```
##
## Call:
## lm(formula = Abstentions_2 ~ ., data = dt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0032989 -0.0010399 -0.0000761  0.0009771  0.0044335
##
## Coefficients: (3 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.13782    0.16016  -0.861   0.3922
## Abstentions_1  1.17329    0.16128   7.275 2.50e-10 ***
## Votants_1      NA         NA      NA      NA
## Exprimes_1     1.01439    0.16659   6.089 4.15e-08 ***
## Blancs_1       1.47736    0.30485   4.846 6.39e-06 ***
## Nuls_1         NA         NA      NA      NA
## ARTHAUD       -0.70412    0.51998  -1.354   0.1797
## ROUSSEL       -0.21088    0.08894  -2.371   0.0202 *
## MACRON         0.11116    0.11585   0.959   0.3403
## LASSALLE       0.04645    0.10104   0.460   0.6470
## LE_PEN         0.24499    0.09849   2.487   0.0150 *
## ZEMMOUR        0.17588    0.10123   1.737   0.0863 .
## MELENCHON      0.09062    0.10221   0.887   0.3781
## HIDALGO       -0.16856    0.13089  -1.288   0.2017
## JADOT          0.14862    0.12726   1.168   0.2465
## PECRESSE       0.02014    0.10567   0.191   0.8493
## POUTOU        -0.11663    0.32008  -0.364   0.7166
## DUPONT_AIGNAN  NA         NA      NA      NA
## Blancs_2      -0.98963    0.08347 -11.855 < 2e-16 ***
## MACRON_2      -0.97890    0.03523 -27.789 < 2e-16 ***
## LE_PEN_2      -1.09727    0.04485 -24.464 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.001598 on 77 degrees of freedom
## Multiple R-squared:  0.9978, Adjusted R-squared:  0.9973
## F-statistic: 2058 on 17 and 77 DF, p-value: < 2.2e-16
```

En raison de la forte multi-colinéarité entre les variables, nous observons des valeurs manquantes dans la sortie de la régression, par suite de la non-singularité de la matrice des explicatives. Nous allons alors les exclure dans nos analyses afin de tirer des conclusions plus fiables.

##	OLS_Abstentions_2	OLS_Blancs_2	OLS_MACRON_2	OLS_LE_PEN_2
## intercept	0.7422341	0.23025682	-0.9179138	0.7336570
## Abstentions_1	0.1657971	-0.27527857	1.0516948	-0.6962571
## Blancs_1	0.8230436	1.34327457	-0.6825774	-0.9307184
## Nuls_1	0.6583059	-0.91572539	-0.1826083	-0.5356195
## ARTHAUD	-2.7463169	-0.72498612	4.9150763	-1.8698474
## ROUSSEL	-0.6088281	0.20659447	0.5928093	-0.3525164
## MACRON	-0.8485045	-0.26359293	2.1780843	-0.8307995
## LASSALLE	-0.6705506	-0.04838026	1.1899864	-0.3645434
## LE_PEN	-0.8969870	-0.20862207	0.7779997	0.5348316
## ZEMMOUR	-0.4676878	-0.19931274	1.0105803	-0.1352873
## MELENCHON	-0.3949841	-0.15520218	1.4511421	-0.7120722
## HIDALGO	-1.1492661	0.04590500	1.7696087	-0.7263454
## JADOT	-0.7401324	-0.15259403	1.9063646	-0.7531261
## PECRESSE	-0.7382574	-0.14815357	1.5308028	-0.5408793
## POUTOU	-1.6774136	1.64239013	-0.7107576	0.5752433

OLS_Abstentions_2

Intercept : 0.7422341 indique la valeur de base de la variable cible lorsque toutes les autres variables explicatives sont à zéro. Abstentions_1 : 0.1657971 suggère qu’une augmentation d’une unité dans “Abstentions_1” est associée à une augmentation de 0.1657971 unités dans la variable cible “OLS_Abstentions_2”, toutes autres choses égales. Les autres coefficients positifs (comme Blancs_1 et Nuls_1) indiquent une relation positive avec la variable cible. OLS_Blancs_2 intercepts : 0.23025682 signifie que la variable cible commence à cette valeur lorsque toutes les autres variables sont à zéro. Les coefficients négatifs (par exemple, Abstentions_1) suggèrent qu’une augmentation de ces variables est associée à une diminution de la variable cible “OLS_Blancs_2”. Les coefficients positifs (par exemple, Blancs_1) indiquent une relation positive avec la variable cible.

OLS_MACRON_2

intercept: -0.9179138 indique que la variable cible “OLS_MACRON_2” commence sous zéro si toutes les variables explicatives sont à zéro. MACRON : 2.1780843 indique une relation positive forte, suggérant que la variable “MACRON” a une influence positive et significative sur la variable cible. Des coefficients négatifs comme celui de LASSALLE indiquent une relation négative avec la variable cible.

OLS_LE_PEN_2

intercept: 0.7336570 sert de point de départ pour la prédiction de la variable cible lorsque toutes les variables explicatives sont à zéro. LE_PEN : 0.5348316 suggère une relation positive modérée entre la variable “LE_PEN” et la variable cible “OLS_LE_PEN_2”. Les coefficients négatifs (comme Abstentions_1) suggèrent une relation négative avec la variable cible. Pour une interprétation plus précise, il est essentiel de connaître le contexte des données, par exemple, si les variables sont des scores de votes ou des pourcentages, ou d’autres mesures de résultats électoraux. De plus, la signification statistique de ces coefficients (généralement évaluée par des tests t et des valeurs p) est cruciale pour comprendre si les relations observées sont significatives ou pourraient être dues au hasard.

##	OLS_Abstentions_2	OLS_Blancs_2	OLS_MACRON_2	OLS_LE_PEN_2
## intercept	0.07191291	0.0797137274	1.292328e-03	0.0027561082
## Abstentions_1	0.67950092	0.0337824873	2.062392e-04	0.0037674121

## Blancs_1	0.44356893	0.0001718582	3.478141e-01	0.1395528984
## Nuls_1	0.36567440	0.0001560037	7.099756e-01	0.2082528142
## ARTHAUD	0.24443589	0.3344355869	2.642951e-03	0.1750525248
## ROUSSEL	0.12110641	0.0992486554	2.679037e-02	0.1239424986
## MACRON	0.03843532	0.0435035002	8.429667e-12	0.0006873385
## LASSALLE	0.12898261	0.7293531965	1.282022e-04	0.1568273881
## LE_PEN	0.03807706	0.1278137594	8.340101e-03	0.0341602337
## ZEMMOUR	0.30658504	0.1727614113	1.499744e-03	0.6113621183
## MELENCHON	0.33769507	0.2379887138	1.271492e-06	0.0038156924
## HIDALGO	0.03176072	0.7848837415	3.632635e-06	0.0203245133
## JADOT	0.14737905	0.3468715864	3.271537e-07	0.0126337867
## PECRESSE	0.09022665	0.2835454489	1.180473e-06	0.0343024733
## POUTOU	0.23878597	0.0004728802	4.592115e-01	0.4874647411

Une p-value faible (typiquement ≤ 0.05) indique que vous pouvez rejeter l'hypothèse nulle, qui est l'hypothèse que le coefficient est égal à zéro (pas d'effet). Cela suggère que la variable associée a un effet significatif sur la variable cible. Une p-value élevée (typiquement > 0.05) indique que vous ne pouvez pas rejeter l'hypothèse nulle et que le coefficient n'est pas statistiquement différent de zéro à ce niveau de confiance. En se basant sur une valeur seuil de 0.05 pour la significativité, voici l'interprétation des p-values pour chaque modèle :

OLS_Abstentions_2

Abstentions_1, Blancs_1, Nuls_1, ARTHAUD, etc. ont des p-values faibles, suggérant que ces variables sont statistiquement significatives dans le modèle de prédiction des abstentions.

OLS_Blancs_2

LASSALLE a une p-value extrêmement faible, indiquant une forte significativité statistique dans le modèle de prédiction des votes blancs. MACRON et LE PEN ont également des p-values significatives dans ce modèle.

OLS_MACRON_2

MACRON a une p-value très faible, ce qui signifie que la variable est hautement significative pour prédire le résultat associé à Macron. Cela a du sens si cette variable représente, par exemple, le pourcentage de votes ou la popularité de Macron dans un modèle électoral. Les autres variables telles que ARTHAUD, ROUSSEL, etc. présentent également des p-values qui indiquent leur significativité statistique dans le modèle.

OLS_LE_PEN_2

ZEMMOUR a une p-value faible, indiquant que c'est une variable significative pour prédire le résultat associé à Le Pen. D'autres variables comme LE PEN et ARTHAUD ont également des p-values indiquant leur significativité dans le modèle. Il est important de noter que même si une variable est statistiquement significative, cela ne nous dit rien sur la taille de son effet sur la variable cible, qui est représentée par le coefficient. Par exemple, un coefficient pourrait être très faible mais significatif si l'échantillon est grand. De plus, la signification statistique ne signifie pas nécessairement qu'il y a une causalité ou une importance pratique. Il est également crucial de considérer la multi-colinéarité dans un modèle qui inclut de nombreuses variables, car cela peut affecter la fiabilité des coefficients de régression et des tests statistiques associés.

##	R_squared_Abstentions_2	R_squared_Blancs_2	R_squared_MACRON_2
## 1	0.9474804	0.9311175	0.993718
##			R_squared_LE_PEN_2
## 1	0.9947176		

Malgré toutes les variables non significatives, nous avons des R2 Très grand, ce qui est un grand signe de la multi-colinéarité qui ne peut pas être traité par la méthode des moindres carrés ordinaires

- PCR

Calcul des composantes principales :

Détection des variables les plus corrélées avec les composantes principales

##	Num CP	Cor max
## Abstentions_1	2	0.7001
## Blancs_1	2	-0.7236
## Nuls_1	3	0.6668
## ARTHAUD	4	-0.6787
## ROUSSEL	2	-0.5342
## MACRON	1	-0.6499
## LASSALLE	4	0.7612
## LE_PEN	1	0.9709
## ZEMMOUR	2	0.7011
## MELENCHON	1	-0.6778
## HIDALGO	0	0.0000
## JADOT	0	0.0000
## PECRESSE	0	0.0000
## POUTOU	0	0.0000

Ici nous retenons 4 composantes principales sur lesquelles nous allons construire notre modèle.

##	PCR1_Abstentions_2	PCR1_Blancs_2	PCR1_MACRON_2	PCR1_LE_PEN_2
## Abstentions_1	0.1137	-0.1720	-0.1534	0.1386
## Blancs_1	-0.1512	0.0866	0.0458	0.0090
## Nuls_1	-0.0328	0.1057	-0.1460	0.1364
## ARTHAUD	-0.1320	-0.0507	-0.0093	0.0928
## ROUSSEL	-0.0618	0.2180	-0.0661	0.0452
## MACRON	-0.1216	-0.0217	0.2431	-0.1774
## LASSALLE	-0.0139	0.2608	-0.0900	0.0311
## LE_PEN	0.0283	-0.0881	-0.2106	0.2212
## ZEMMOUR	0.1449	-0.0515	-0.0825	0.0175
## MELENCHON	0.0162	0.0810	0.1418	-0.1691
## HIDALGO	-0.0956	0.2720	0.0654	-0.0835
## JADOT	-0.0613	0.0140	0.2481	-0.2224
## PECRESSE	-0.0757	0.0220	0.1186	-0.0918
## POUTOU	-0.1283	0.1276	0.0136	0.0228

PCR1_MACRON_2

MACRON a un coefficient positif significatif (0.2431), ce qui suggère que la première composante principale capte des variations dans les données qui sont positivement associées à la variable MACRON. C'est-à-dire, plus la première composante principale augmente, plus la valeur prédite pour la variable dépendante associée à Macron augmente, indiquant une influence positive et potentiellement importante.

JADOT a également un coefficient positif (0.2481), indiquant une association positive avec la variable prédite MACRON_2. Cela pourrait signifier que les variables JADOT et MACRON se comportent de manière similaire ou sont corrélées dans le contexte de la première composante principale.

LE_PEN a un coefficient négatif significatif (-0.2106), ce qui indique que, toutes choses égales par ailleurs, une augmentation de la variable LE_PEN est associée à une diminution de la variable dépendante MACRON_2. Cela pourrait être interprété comme une sorte d'effet opposé entre les

variables MACRON et LE_PEN dans la première composante principale.

PCR1_Blancs_2

LASSALLE a un coefficient positif significatif (0.2608), indiquant que l'augmentation de la variable LASSALLE est associée à une augmentation de la variable dépendante Blancs_2 dans le contexte de la première composante principale, suggérant une relation positive.

HIDALGO a également un coefficient positif élevé (0.2720), suggérant une influence significative sur la variable Blancs_2. Cela pourrait indiquer que HIDALGO est positivement corrélée avec les votes blancs dans le cadre de la PCR.

Abstentions_1 a un coefficient négatif (-0.1720), ce qui indique que lorsque la première composante principale augmente, les votes blancs prédits diminuent, suggérant une association négative entre les abstentions précédentes et les votes blancs.

PCR1_Abstentions_2

Abstentions_1 a un coefficient positif de 0.1137, ce qui indique que, toutes choses égales par ailleurs, une augmentation de la variable Abstentions_1 est associée à une légère augmentation des abstentions prédites par la première composante principale. Blancs_1 a un coefficient négatif de -0.1512, indiquant une association négative avec les abstentions prédites. ZEMMOUR a un coefficient positif de 0.1449, indiquant une association positive avec les abstentions prédites.

PCR1_LE_PEN_2

LE_PEN a un coefficient positif de 0.2212, ce qui suggère une forte association positive avec la variable LE_PEN prédite. Cela pourrait signifier que la première composante principale capture des variations dans les données qui sont fortement liées à LE_PEN. MACRON a un coefficient négatif de -0.1774, indiquant que la variable MACRON est négativement associée à la prédiction de LE_PEN dans le contexte de la PCR.

PLS1

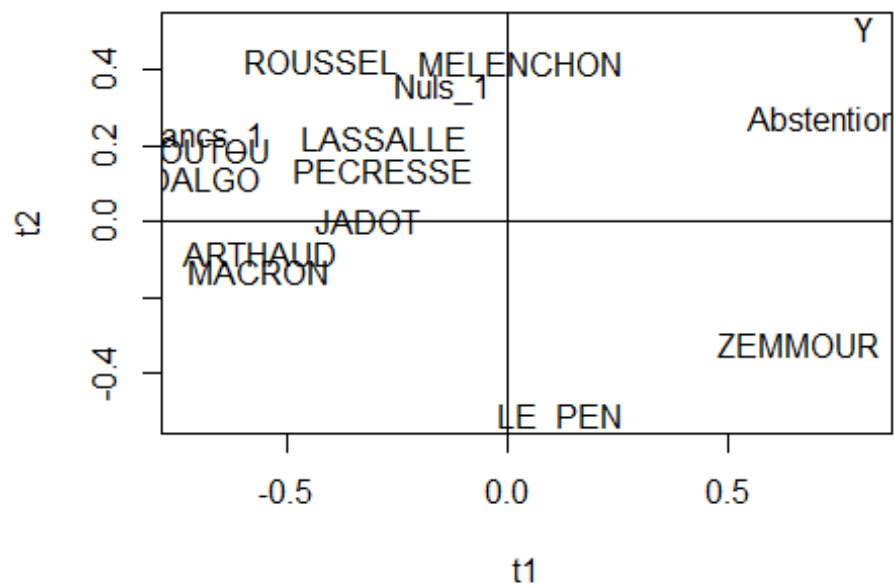
```
## [1] "Modèle pour la variable cible : PLS1_Abstentions_2"
##
## PLS Regression 1
## -----
## $x.scores      X-scores (T-components)
## $x.loads       X-loadings
## $y.scores      Y-scores (U-components)
## $y.loads       Y-loadings
## $cor.xyt       score correlations
## $raw.wgs       raw weights
## $mod.wgs       modified weights
## $std.coefs     standard coefficients
## $reg.coefs     regular coefficients
## $R2            R-squared
## $R2Xy          explained variance of X-y by T
## $y.pred        y-predicted
## $resid         residuals
## $T2            T2 hotelling
## $Q2            Q2 cross validation
## -----
##
```

```

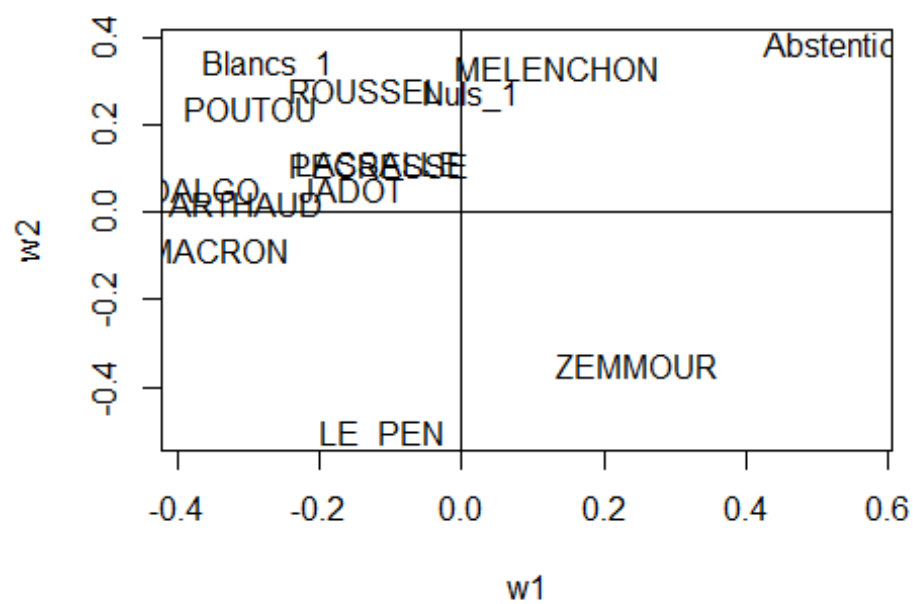
## [1] "R2 pour PLS1_Abstentions_2 : 0.657779175949834"
## [2] "R2 pour PLS1_Abstentions_2 : 0.257832806977265"
## [1] "R2Xy pour PLS1_Abstentions_2 : 0.601164327039413"
## [2] "R2Xy pour PLS1_Abstentions_2 : 0.488459607781274"
## [3] "R2Xy pour PLS1_Abstentions_2 : 0.0221076364965376"
## [4] "R2Xy pour PLS1_Abstentions_2 : 0.310713466939444"
## [5] "R2Xy pour PLS1_Abstentions_2 : 0.176299491017304"
## [6] "R2Xy pour PLS1_Abstentions_2 : 0.315694465844868"
## [7] "R2Xy pour PLS1_Abstentions_2 : 0.0789096999188867"
## [8] "R2Xy pour PLS1_Abstentions_2 : 0.014579939935907"
## [9] "R2Xy pour PLS1_Abstentions_2 : 0.437987798843564"
## [10] "R2Xy pour PLS1_Abstentions_2 : 0.000999382417900162"
## [11] "R2Xy pour PLS1_Abstentions_2 : 0.520588858784692"
## [12] "R2Xy pour PLS1_Abstentions_2 : 0.0985440454900058"
## [13] "R2Xy pour PLS1_Abstentions_2 : 0.0771620954971955"
## [14] "R2Xy pour PLS1_Abstentions_2 : 0.470728835952579"
## [15] "R2Xy pour PLS1_Abstentions_2 : 0.657779175949834"
## [16] "R2Xy pour PLS1_Abstentions_2 : 0.673433202882485"
## [17] "R2Xy pour PLS1_Abstentions_2 : 0.539710041022002"
## [18] "R2Xy pour PLS1_Abstentions_2 : 0.145162079806497"
## [19] "R2Xy pour PLS1_Abstentions_2 : 0.317083124251625"
## [20] "R2Xy pour PLS1_Abstentions_2 : 0.35693272086953"
## [21] "R2Xy pour PLS1_Abstentions_2 : 0.333032503465603"
## [22] "R2Xy pour PLS1_Abstentions_2 : 0.127884834833628"
## [23] "R2Xy pour PLS1_Abstentions_2 : 0.280224921938002"
## [24] "R2Xy pour PLS1_Abstentions_2 : 0.541885538902272"
## [25] "R2Xy pour PLS1_Abstentions_2 : 0.176278060764057"
## [26] "R2Xy pour PLS1_Abstentions_2 : 0.533556617181354"
## [27] "R2Xy pour PLS1_Abstentions_2 : 0.0985501790612112"
## [28] "R2Xy pour PLS1_Abstentions_2 : 0.0966174789944793"
## [29] "R2Xy pour PLS1_Abstentions_2 : 0.505908077995259"
## [30] "R2Xy pour PLS1_Abstentions_2 : 0.915611982927099"
## [1] "Coefficients de régression pour PLS1_Abstentions_2 : 0.226829944698952"
## [2] "Coefficients de régression pour PLS1_Abstentions_2 : 0.562985453359797"
## [3] "Coefficients de régression pour PLS1_Abstentions_2 : -0.00640402429005001"
"
## [4] "Coefficients de régression pour PLS1_Abstentions_2 : 2.72753265451469"
## [5] "Coefficients de régression pour PLS1_Abstentions_2 : -5.02505274325451"
## [6] "Coefficients de régression pour PLS1_Abstentions_2 : 0.360996543447149"
## [7] "Coefficients de régression pour PLS1_Abstentions_2 : -0.23787632113378"
## [8] "Coefficients de régression pour PLS1_Abstentions_2 : -0.0280672391526301"
## [9] "Coefficients de régression pour PLS1_Abstentions_2 : -0.240740429005555"
## [10] "Coefficients de régression pour PLS1_Abstentions_2 : -0.055316046378053"
## [11] "Coefficients de régression pour PLS1_Abstentions_2 : 0.223965616058837"
## [12] "Coefficients de régression pour PLS1_Abstentions_2 : -1.5156650669975"
## [13] "Coefficients de régression pour PLS1_Abstentions_2 : -0.239121078639606"
## [14] "Coefficients de régression pour PLS1_Abstentions_2 : -0.0635803460797285"
## [15] "Coefficients de régression pour PLS1_Abstentions_2 : -1.93068625956164"

```

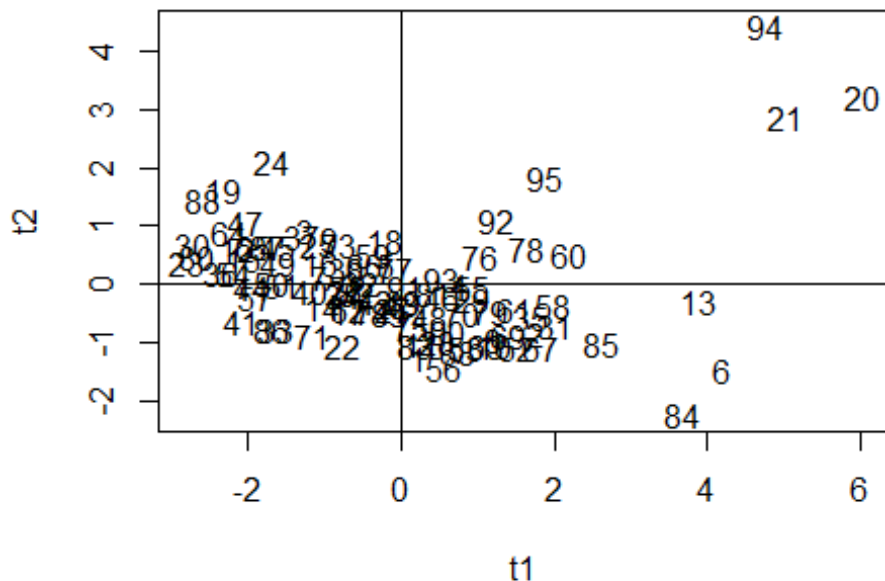
Premier plan des variables pour PLS1_Abstentions



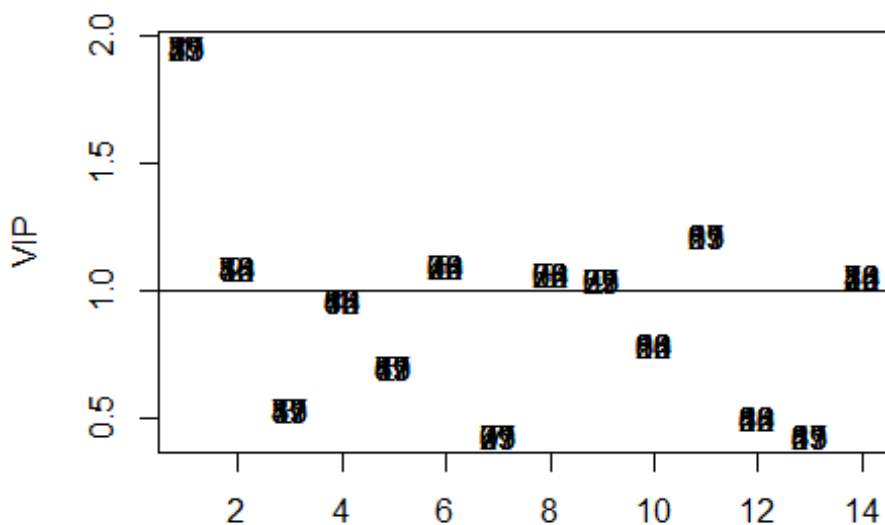
.S1_Abstentions_2 Premier plan des poids variables



PLS1_Abstentions_2 Premier plan des individus (t1)



Représentation des VIP



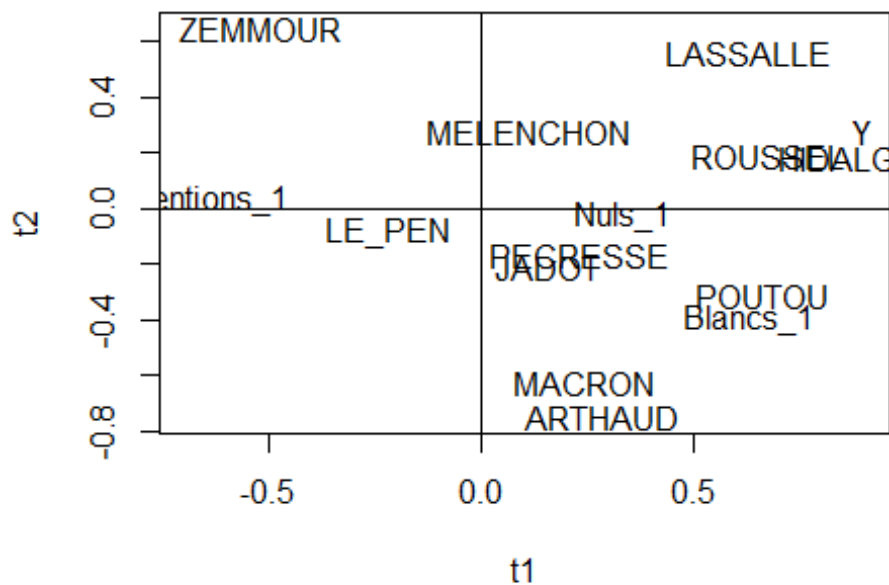
```
## [1] "Modèle pour la variable cible : PLS1_Blancs_2"
##
## PLS Regression 1
## -----
## $x.scores      X-scores (T-components)
## $x.loads       X-loadings
## $y.scores      Y-scores (U-components)
## $y.loads       Y-loadings
## $cor.xyt       score correlations
## $raw.wgs       raw weights
## $mod.wgs       modified weights
## $std.coefs     standard coefficients
## $reg.coefs     regular coefficients
```

```

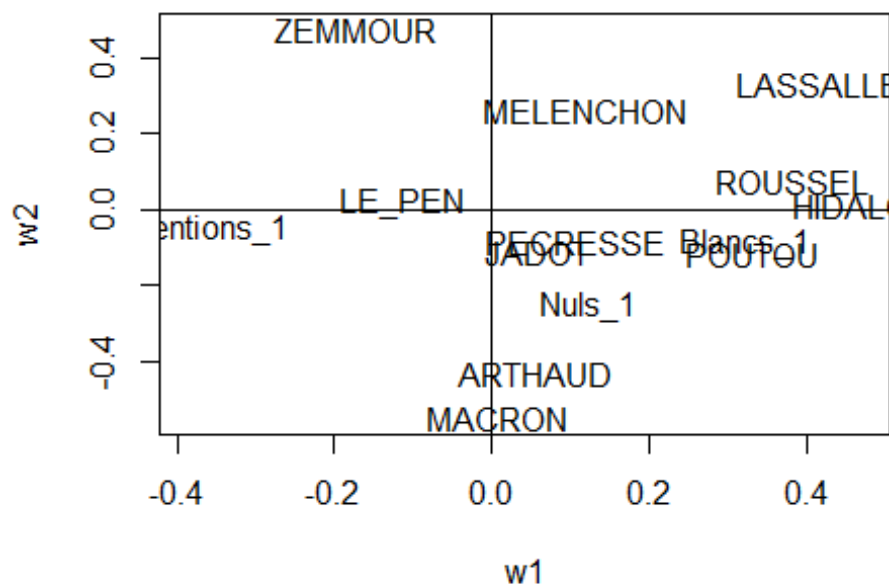
## $R2          R-squared
## $R2Xy        explained variance of X-y by T
## $y.pred      y-predicted
## $resid       residuals
## $T2          T2 hotelling
## $Q2          Q2 cross validation
## -----
##
## [1] "R2 pour PLS1_Blancs_2 : 0.804007730251244"
## [2] "R2 pour PLS1_Blancs_2 : 0.0776605018806844"
## [1] "R2Xy pour PLS1_Blancs_2 : 0.478743395439975"
## [2] "R2Xy pour PLS1_Blancs_2 : 0.397095253539272"
## [3] "R2Xy pour PLS1_Blancs_2 : 0.11020453579402"
## [4] "R2Xy pour PLS1_Blancs_2 : 0.0807600251986839"
## [5] "R2Xy pour PLS1_Blancs_2 : 0.454701584930757"
## [6] "R2Xy pour PLS1_Blancs_2 : 0.0591174881480393"
## [7] "R2Xy pour PLS1_Blancs_2 : 0.394517280570816"
## [8] "R2Xy pour PLS1_Blancs_2 : 0.0477132370909713"
## [9] "R2Xy pour PLS1_Blancs_2 : 0.267155413289418"
## [10] "R2Xy pour PLS1_Blancs_2 : 0.0127045809474446"
## [11] "R2Xy pour PLS1_Blancs_2 : 0.747777518422281"
## [12] "R2Xy pour PLS1_Blancs_2 : 0.0241283260090013"
## [13] "R2Xy pour PLS1_Blancs_2 : 0.0542474714241546"
## [14] "R2Xy pour PLS1_Blancs_2 : 0.433162351277518"
## [15] "R2Xy pour PLS1_Blancs_2 : 0.804007730251244"
## [16] "R2Xy pour PLS1_Blancs_2 : 0.479698679599191"
## [17] "R2Xy pour PLS1_Blancs_2 : 0.552188928018593"
## [18] "R2Xy pour PLS1_Blancs_2 : 0.110883686290429"
## [19] "R2Xy pour PLS1_Blancs_2 : 0.642905806672078"
## [20] "R2Xy pour PLS1_Blancs_2 : 0.490459321989694"
## [21] "R2Xy pour PLS1_Blancs_2 : 0.449319272372071"
## [22] "R2Xy pour PLS1_Blancs_2 : 0.704508023117677"
## [23] "R2Xy pour PLS1_Blancs_2 : 0.053991917624758"
## [24] "R2Xy pour PLS1_Blancs_2 : 0.682566368376936"
## [25] "R2Xy pour PLS1_Blancs_2 : 0.088480369612219"
## [26] "R2Xy pour PLS1_Blancs_2 : 0.780526335618474"
## [27] "R2Xy pour PLS1_Blancs_2 : 0.0681925311672267"
## [28] "R2Xy pour PLS1_Blancs_2 : 0.0807366142648384"
## [29] "R2Xy pour PLS1_Blancs_2 : 0.528470542162715"
## [30] "R2Xy pour PLS1_Blancs_2 : 0.881668232131929"
## [1] "Coefficients de régression pour PLS1_Blancs_2 : 0.0322065534137466"
## [2] "Coefficients de régression pour PLS1_Blancs_2 : -0.0619303835705117"
## [3] "Coefficients de régression pour PLS1_Blancs_2 : 0.824754868828741"
## [4] "Coefficients de régression pour PLS1_Blancs_2 : 0.073288849542195"
## [5] "Coefficients de régression pour PLS1_Blancs_2 : -0.457554256041744"
## [6] "Coefficients de régression pour PLS1_Blancs_2 : 0.391471075437176"
## [7] "Coefficients de régression pour PLS1_Blancs_2 : -0.0280711870908833"
## [8] "Coefficients de régression pour PLS1_Blancs_2 : 0.137773946714669"
## [9] "Coefficients de régression pour PLS1_Blancs_2 : -0.0116745690754272"
## [10] "Coefficients de régression pour PLS1_Blancs_2 : 0.00381580040424613"
## [11] "Coefficients de régression pour PLS1_Blancs_2 : 0.0294112579037742"
## [12] "Coefficients de régression pour PLS1_Blancs_2 : 0.502828927892268"
## [13] "Coefficients de régression pour PLS1_Blancs_2 : 0.00886824023495297"
## [14] "Coefficients de régression pour PLS1_Blancs_2 : 0.0419351431355149"
## [15] "Coefficients de régression pour PLS1_Blancs_2 : 1.26219947024454"

```

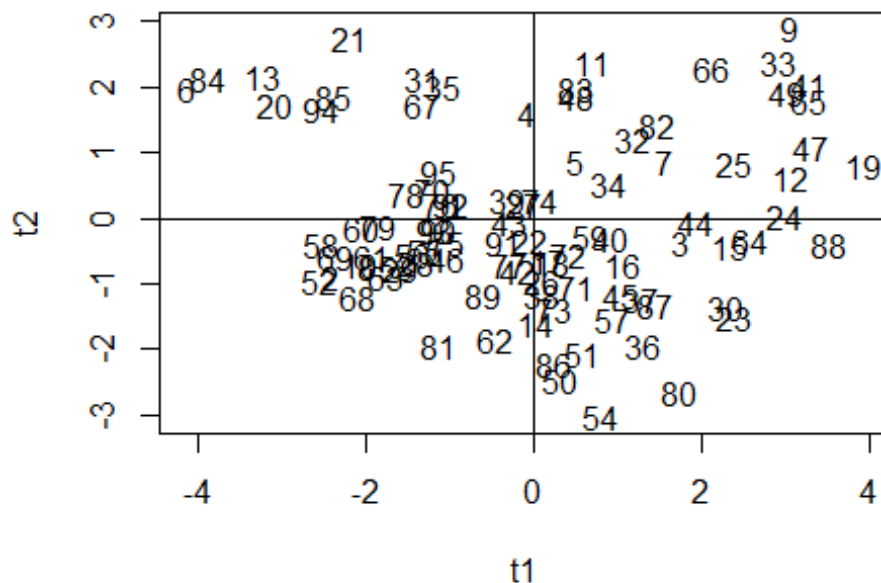
Premier plan des variables pour PLS1_Blancs_2



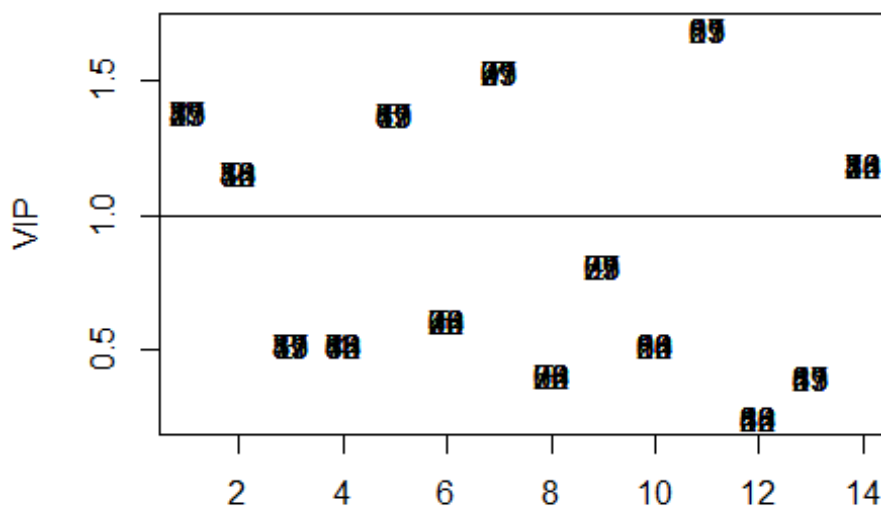
PLS1_Blancs_2 Premier plan des poids variables (t'



PLS1_Blancs_2 Premier plan des individus (t1,t2)



Représentation des VIP



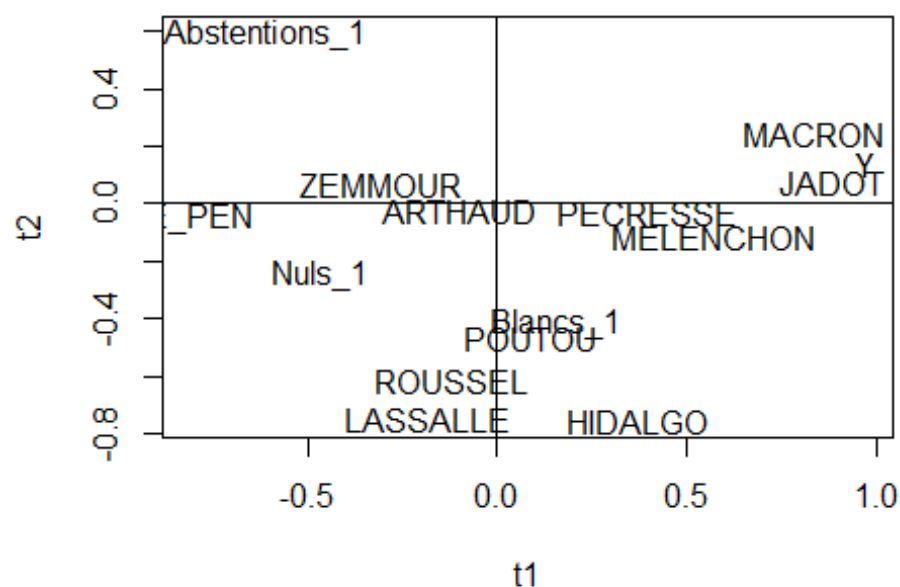
```
## [1] "Modèle pour la variable cible : PLS1_MACRON_2"
##
## PLS Regression 1
## -----
## $x.scores      X-scores (T-components)
## $x.loads       X-loadings
## $y.scores      Y-scores (U-components)
## $y.loads       Y-loadings
## $cor.xyt       score correlations
## $raw.wgs       raw weights
## $mod.wgs       modified weights
## $std.coefs     standard coefficients
## $reg.coefs     regular coefficients
```

```

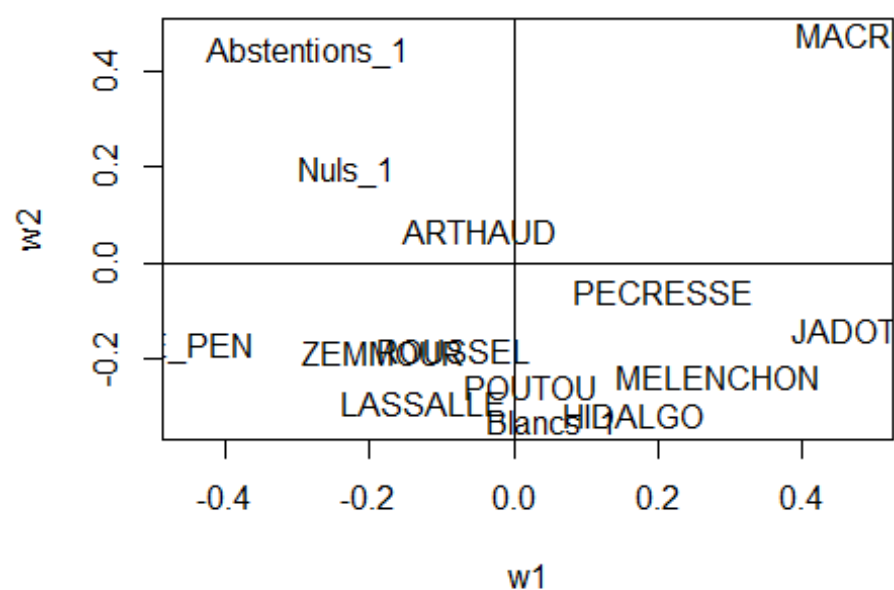
## $R2          R-squared
## $R2Xy        explained variance of X-y by T
## $y.pred      y-predicted
## $resid       residuals
## $T2          T2 hotelling
## $Q2          Q2 cross validation
## -----
##
## [1] "R2 pour PLS1_MACRON_2 : 0.945388088538547"
## [2] "R2 pour PLS1_MACRON_2 : 0.0204629155822294"
## [1] "R2Xy pour PLS1_MACRON_2 : 0.370660303283525"
## [2] "R2Xy pour PLS1_MACRON_2 : 0.0254618507093045"
## [3] "R2Xy pour PLS1_MACRON_2 : 0.21913673993683"
## [4] "R2Xy pour PLS1_MACRON_2 : 0.00953703411060266"
## [5] "R2Xy pour PLS1_MACRON_2 : 0.0149254492928661"
## [6] "R2Xy pour PLS1_MACRON_2 : 0.694325246630185"
## [7] "R2Xy pour PLS1_MACRON_2 : 0.0329477159941121"
## [8] "R2Xy pour PLS1_MACRON_2 : 0.6559651551341"
## [9] "R2Xy pour PLS1_MACRON_2 : 0.0931258450577784"
## [10] "R2Xy pour PLS1_MACRON_2 : 0.32806450319579"
## [11] "R2Xy pour PLS1_MACRON_2 : 0.135926237980112"
## [12] "R2Xy pour PLS1_MACRON_2 : 0.780387057828807"
## [13] "R2Xy pour PLS1_MACRON_2 : 0.159373016510677"
## [14] "R2Xy pour PLS1_MACRON_2 : 0.00805468010875261"
## [15] "R2Xy pour PLS1_MACRON_2 : 0.945388088538547"
## [16] "R2Xy pour PLS1_MACRON_2 : 0.724774227331919"
## [17] "R2Xy pour PLS1_MACRON_2 : 0.195033424795224"
## [18] "R2Xy pour PLS1_MACRON_2 : 0.279034050060234"
## [19] "R2Xy pour PLS1_MACRON_2 : 0.0100850257415129"
## [20] "R2Xy pour PLS1_MACRON_2 : 0.394261761921422"
## [21] "R2Xy pour PLS1_MACRON_2 : 0.755751426073673"
## [22] "R2Xy pour PLS1_MACRON_2 : 0.589510286309278"
## [23] "R2Xy pour PLS1_MACRON_2 : 0.658331659005826"
## [24] "R2Xy pour PLS1_MACRON_2 : 0.0980121999818996"
## [25] "R2Xy pour PLS1_MACRON_2 : 0.340747798547229"
## [26] "R2Xy pour PLS1_MACRON_2 : 0.707751223026308"
## [27] "R2Xy pour PLS1_MACRON_2 : 0.786779558207479"
## [28] "R2Xy pour PLS1_MACRON_2 : 0.160034085573514"
## [29] "R2Xy pour PLS1_MACRON_2 : 0.223195766470964"
## [30] "R2Xy pour PLS1_MACRON_2 : 0.965851004120776"
## [1] "Coefficients de régression pour PLS1_MACRON_2 : 0.354535998962201"
## [2] "Coefficients de régression pour PLS1_MACRON_2 : -0.195376642025657"
## [3] "Coefficients de régression pour PLS1_MACRON_2 : -0.44319121093791"
## [4] "Coefficients de régression pour PLS1_MACRON_2 : -3.71930928026798"
## [5] "Coefficients de régression pour PLS1_MACRON_2 : -0.870239310914362"
## [6] "Coefficients de régression pour PLS1_MACRON_2 : -0.813678880390936"
## [7] "Coefficients de régression pour PLS1_MACRON_2 : 0.572517253419286"
## [8] "Coefficients de régression pour PLS1_MACRON_2 : -0.337313820621266"
## [9] "Coefficients de régression pour PLS1_MACRON_2 : -0.375076951742402"
## [10] "Coefficients de régression pour PLS1_MACRON_2 : -0.588635349521949"
## [11] "Coefficients de régression pour PLS1_MACRON_2 : 0.212213635742888"
## [12] "Coefficients de régression pour PLS1_MACRON_2 : 0.697177003001722"
## [13] "Coefficients de régression pour PLS1_MACRON_2 : 1.51966744192024"
## [14] "Coefficients de régression pour PLS1_MACRON_2 : 0.744168488892938"
## [15] "Coefficients de régression pour PLS1_MACRON_2 : -1.07655662245469"

```

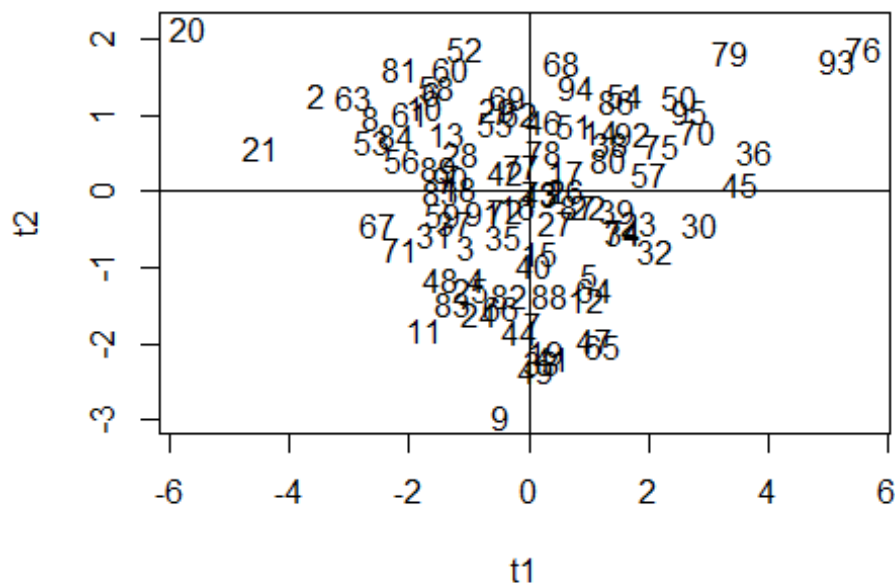
Premier plan des variables pour PLS1_MACRON_



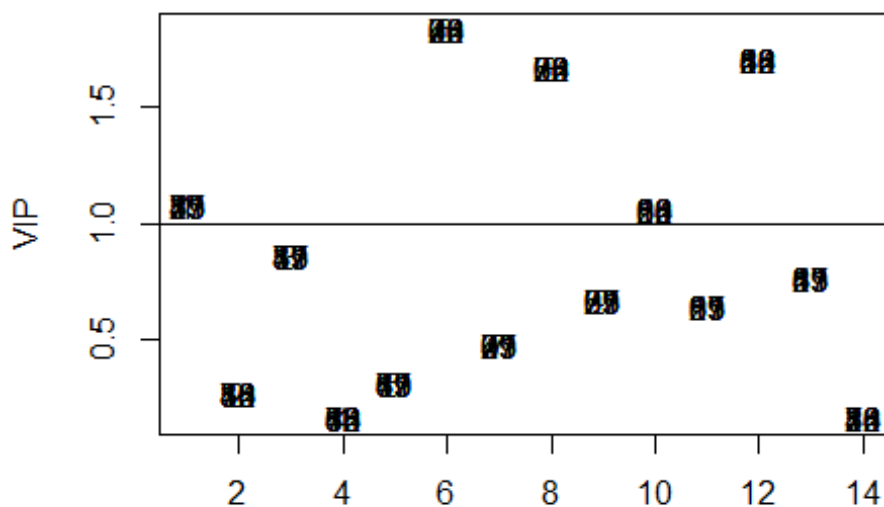
PLS1_MACRON_2 Premier plan des poids variables (



PLS1_MACRON_2 Premier plan des individus (t1,t



Représentation des VIP



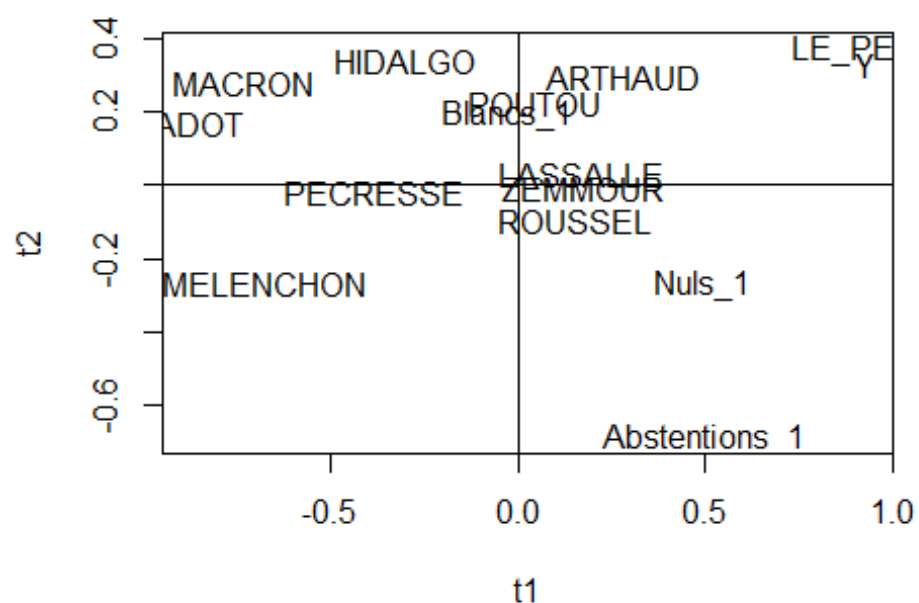
```
## [1] "Modèle pour la variable cible : PLS1_LE_PEN_2"
##
## PLS Regression 1
## -----
## $x.scores      X-scores (T-components)
## $x.loads       X-loadings
## $y.scores      Y-scores (U-components)
## $y.loads       Y-loadings
## $cor.xyt       score correlations
## $raw.wgs       raw weights
## $mod.wgs       modified weights
## $std.coefs     standard coefficients
## $reg.coefs     regular coefficients
```

```

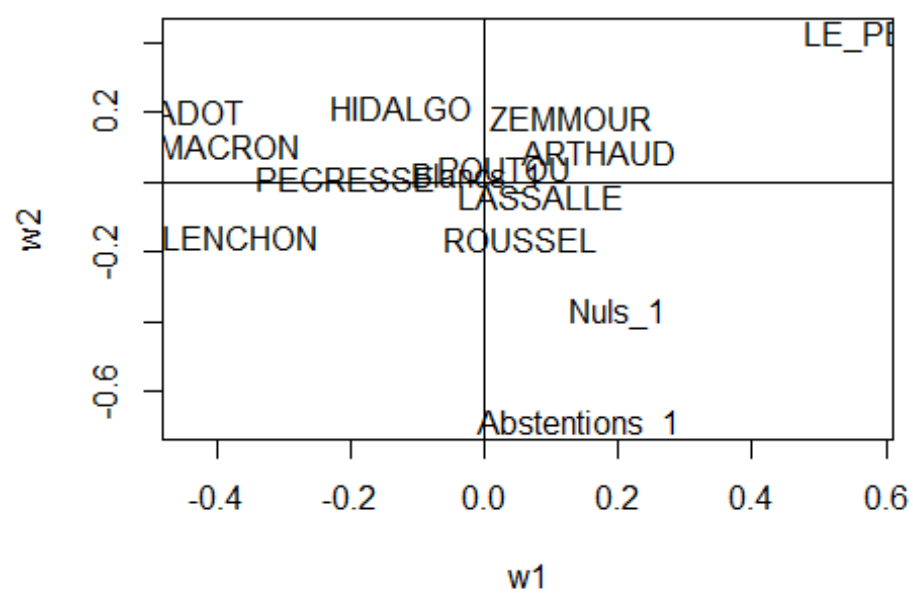
## $R2          R-squared
## $R2Xy        explained variance of X-y by T
## $y.pred      y-predicted
## $resid       residuals
## $T2          T2 hotelling
## $Q2          Q2 cross validation
## -----
##
## [1] "R2 pour PLS1_LE_PEN_2 : 0.864434012537074"
## [2] "R2 pour PLS1_LE_PEN_2 : 0.109258421120394"
## [1] "R2Xy pour PLS1_LE_PEN_2 : 0.248775741562892"
## [2] "R2Xy pour PLS1_LE_PEN_2 : 0.00055058808274033"
## [3] "R2Xy pour PLS1_LE_PEN_2 : 0.245734187458101"
## [4] "R2Xy pour PLS1_LE_PEN_2 : 0.0808767717809401"
## [5] "R2Xy pour PLS1_LE_PEN_2 : 0.0233482414925477"
## [6] "R2Xy pour PLS1_LE_PEN_2 : 0.539244714793148"
## [7] "R2Xy pour PLS1_LE_PEN_2 : 0.0288019987596844"
## [8] "R2Xy pour PLS1_LE_PEN_2 : 0.808571357110297"
## [9] "R2Xy pour PLS1_LE_PEN_2 : 0.0302609788138103"
## [10] "R2Xy pour PLS1_LE_PEN_2 : 0.455654157747297"
## [11] "R2Xy pour PLS1_LE_PEN_2 : 0.0906741403565075"
## [12] "R2Xy pour PLS1_LE_PEN_2 : 0.767260478453401"
## [13] "R2Xy pour PLS1_LE_PEN_2 : 0.144718298695067"
## [14] "R2Xy pour PLS1_LE_PEN_2 : 0.00181612892520232"
## [15] "R2Xy pour PLS1_LE_PEN_2 : 0.864434012537074"
## [16] "R2Xy pour PLS1_LE_PEN_2 : 0.71872519714617"
## [17] "R2Xy pour PLS1_LE_PEN_2 : 0.039150571493942"
## [18] "R2Xy pour PLS1_LE_PEN_2 : 0.318039803140575"
## [19] "R2Xy pour PLS1_LE_PEN_2 : 0.16837159123733"
## [20] "R2Xy pour PLS1_LE_PEN_2 : 0.032544234123042"
## [21] "R2Xy pour PLS1_LE_PEN_2 : 0.616800582514775"
## [22] "R2Xy pour PLS1_LE_PEN_2 : 0.0298019372230071"
## [23] "R2Xy pour PLS1_LE_PEN_2 : 0.945970277302505"
## [24] "R2Xy pour PLS1_LE_PEN_2 : 0.0302858961262975"
## [25] "R2Xy pour PLS1_LE_PEN_2 : 0.526825842756753"
## [26] "R2Xy pour PLS1_LE_PEN_2 : 0.206556272546423"
## [27] "R2Xy pour PLS1_LE_PEN_2 : 0.795411109459567"
## [28] "R2Xy pour PLS1_LE_PEN_2 : 0.144989495060562"
## [29] "R2Xy pour PLS1_LE_PEN_2 : 0.0526914737090563"
## [30] "R2Xy pour PLS1_LE_PEN_2 : 0.973692433657468"
## [1] "Coefficients de régression pour PLS1_LE_PEN_2 : 0.374708442017606"
## [2] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.293992391997922"
## [3] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.0361929219014491"
## [4] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.490169590094681"
## [5] "Coefficients de régression pour PLS1_LE_PEN_2 : 6.64273293281368"
## [6] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.300872135314359"
## [7] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.310201037672299"
## [8] "Coefficients de régression pour PLS1_LE_PEN_2 : 0.11440853408594"
## [9] "Coefficients de régression pour PLS1_LE_PEN_2 : 0.641305468432474"
## [10] "Coefficients de régression pour PLS1_LE_PEN_2 : 0.654655030332088"
## [11] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.460837126170777"
## [12] "Coefficients de régression pour PLS1_LE_PEN_2 : 0.0524430795660882"
## [13] "Coefficients de régression pour PLS1_LE_PEN_2 : -1.14008726758303"
## [14] "Coefficients de régression pour PLS1_LE_PEN_2 : -0.773692697717755"
## [15] "Coefficients de régression pour PLS1_LE_PEN_2 : 1.73139570869594"

```

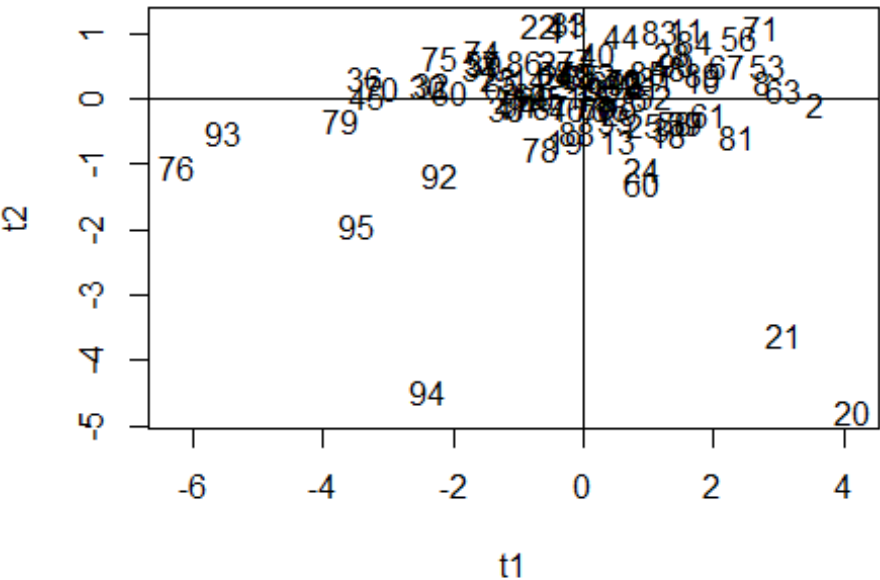

Premier plan des variables pour PLS1_LE_PEN_1



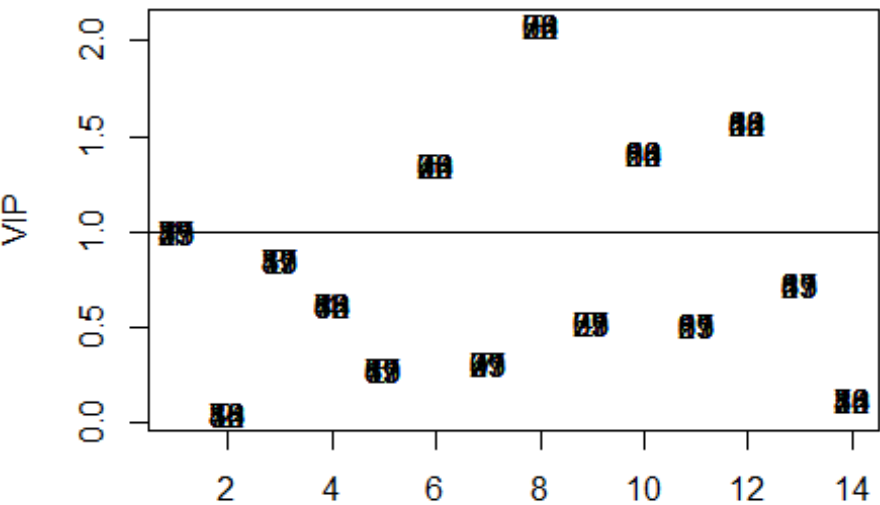
PLS1_LE_PEN_2 Premier plan des poids variables (t



PLS1_LE_PEN_2 Premier plan des individus (t1,t2)



Représentation des VIP



##	PLS1_Abstentions_2	PLS1_Blancs_2	PLS1_MACRON_2	PLS1_LE_PEN_2
## Abstentions_1	0.5490459691	-0.216987542	-0.09745560	-0.155838127
## Blancs_1	-0.0003362206	0.155566369	-0.01190104	-0.001032812
## Nuls_1	0.1431821743	0.013822175	-0.09986284	-0.013985933
## ARTHAUD	-0.1794347545	-0.058698640	-0.01589378	0.128925501
## ROUSSEL	0.0566964249	0.220887839	-0.06536244	-0.025683884
## MACRON	-0.2584911138	-0.109590947	0.31820397	-0.183216147
## LASSALLE	-0.0164834436	0.290693183	-0.10132229	0.036520168
## LE_PEN	-0.3257398370	-0.056752134	-0.25957618	0.471642445
## ZEMMOUR	-0.0217367723	0.005387026	-0.11830766	0.139824259
## MELENCHON	0.2520992653	0.118938738	0.12217606	-0.281944602
## HIDALGO	-0.2141646810	0.255260938	0.05038605	0.004027716
## JADOT	-0.0698179245	0.009302636	0.22694481	-0.180931116
## PECRESSE	-0.0173015999	0.040997843	0.10357555	-0.114434762
## POUTOU	-0.0653079939	0.153391794	-0.01862577	0.031832963

Selon le modèle pour Abstentions_2 Les départements atypique sont principalement les 94,21,20,13,85,6,84, de plus ce sont les variables Abstentions1 et ZEMOUR ZERBO et LE_PEN qui contribue le plus à la première composante principale.

Pour le modèle Blancs2 on constate une répartition équitable des différents départements, et ce sont les variables ZEMOUR et Abstention1 qui pour le modèle Macron2.

##	OLS_Abstentions_2	PCR1_Abstentions_2	PLS1_Abstentions_2
## Abstentions_1	0.1657971	0.1137	0.5490459691
## Blancs_1	0.8230436	-0.1512	-0.0003362206
## Nuls_1	0.6583059	-0.0328	0.1431821743
## ARTHAUD	-2.7463169	-0.1320	-0.1794347545
## ROUSSEL	-0.6088281	-0.0618	0.0566964249
## MACRON	-0.8485045	-0.1216	-0.2584911138
## LASSALLE	-0.6705506	-0.0139	-0.0164834436
## LE_PEN	-0.8969870	0.0283	-0.3257398370
## ZEMMOUR	-0.4676878	0.1449	-0.0217367723
## MELENCHON	-0.3949841	0.0162	0.2520992653
## HIDALGO	-1.1492661	-0.0956	-0.2141646810
## JADOT	-0.7401324	-0.0613	-0.0698179245
## PECRESSE	-0.7382574	-0.0757	-0.0173015999
## POUTOU	-1.6774136	-0.1283	-0.0653079939
##	cor_Abstentions_2		
## Abstentions_1	0.82430342		
## Blancs_1	-0.39378083		
## Nuls_1	0.01738509		
## ARTHAUD	-0.44058516		
## ROUSSEL	-0.19639415		
## MACRON	-0.49891716		
## LASSALLE	-0.16981695		
## LE_PEN	-0.16203150		
## ZEMMOUR	0.35730747		
## MELENCHON	0.19582248		
## HIDALGO	-0.55751403		
## JADOT	-0.22606624		
## PECRESSE	-0.16863333		
## POUTOU	-0.43245235		

Abstentions_1: Dans OLS et PCR, le coefficient est positif, ce qui suggère une relation positive avec Abstentions_2. Dans le modèle PLS, le coefficient est également positif et relativement plus grand, indiquant une forte association positive. Cela est en accord avec la corrélation positive élevée, ce qui signifie que les augmentations dans Abstentions_1 sont associées à des augmentations dans Abstentions_2.

Blancs_1: Pour OLS, le coefficient est positif, alors qu'il est négatif dans PCR et très proche de zéro dans PLS. La corrélation est négative, ce qui suggère que les augmentations dans Blancs_1 sont généralement associées à des diminutions dans Abstentions_2. Le modèle PCR est en accord avec cette corrélation, contrairement à OLS.

Nuls_1: Le coefficient est positif dans OLS, positif mais plus faible dans PLS, et négatif dans PCR. La corrélation étant positive mais faible, le modèle PLS semble être plus en accord avec cette corrélation.

ARTHAUD: Tous les modèles montrent un coefficient négatif, en accord avec la corrélation négative, indiquant que des augmentations dans ARTHAUD sont associées à des diminutions dans Abstentions_2.

ROUSSEL: De même que pour ARTHAUD, tous les modèles indiquent une relation négative, ce qui est cohérent avec la corrélation négative.

MACRON: Les coefficients dans tous les modèles sont négatifs et concordent avec la corrélation négative. Cela suggère que des augmentations dans MACRON sont associées à des diminutions dans Abstentions_2.

LASSALLE: Les coefficients sont négatifs dans OLS et PLS, et très proche de zéro dans PCR, ce qui suggère une relation négative ou non significative avec Abstentions_2. Cela est partiellement en accord avec la corrélation négative.

LE_PEN: Les coefficients dans OLS et PLS sont négatifs, alors qu'il est positif dans PCR. La corrélation négative suggère que le modèle PCR n'est pas en accord avec cette tendance.

ZEMMOUR: Le coefficient dans OLS est négatif, dans PCR positif, et très proche de zéro dans PLS. La corrélation est positive, ce qui indique que seul le modèle PCR est en accord avec cette tendance.

MELENCHON: Les coefficients dans OLS sont négatifs, positifs dans PLS et PCR. La corrélation positive suggère que les modèles PLS et PCR sont en accord avec la tendance observée.

En résumé, il semble que le modèle PLS soit généralement plus en phase avec les tendances de corrélation, à l'exception de quelques cas comme Blancs_1. Le modèle PCR est parfois en accord avec les corrélations (ZEMMOUR, MELENCHON), tandis que le modèle OLS présente des incohérences par rapport aux corrélations données pour certaines variables (Blancs_1, LE_PEN). Cela peut indiquer que la méthode de PLS, qui cherche à maximiser la covariance entre les variables explicatives et la variable cible, capture mieux certaines des relations sous-jacentes.

##	OLS_Blancs_2	PCR1_Blancs_2	PLS1_Blancs_2	cor_Blancs_2
## Abstentions_1	-0.27527857	-0.1720	-0.216987542	-0.63881671
## Blancs_1	1.34327457	0.0866	0.155566369	0.53352328
## Nuls_1	-0.91572539	0.1057	0.013822175	0.20285398
## ARTHAUD	-0.72498612	-0.0507	-0.058698640	0.09268787
## ROUSSEL	0.20659447	0.2180	0.220887839	0.63338212
## MACRON	-0.26359293	-0.0217	-0.109590947	0.01114929
## LASSALLE	-0.04838026	0.2608	0.290693183	0.68748587
## LE_PEN	-0.20862207	-0.0881	-0.056752134	-0.18818340
## ZEMMOUR	-0.19931274	-0.0515	0.005387026	-0.28495060
## MELENCHON	-0.15520218	0.0810	0.118938738	0.20002183
## HIDALGO	0.04590500	0.2720	0.255260938	0.78010002
## JADOT	-0.15259403	0.0140	0.009302636	0.09817222
## PECRESSE	-0.14815357	0.0220	0.040997843	0.17825304
## POUTOU	1.64239013	0.1276	0.153391794	0.54679601

Abstentions_1: Tous les modèles présentent des coefficients négatifs, ce qui est cohérent avec la corrélation négative élevée. Cela suggère que des taux d'abstention plus élevés sont associés à moins de votes blancs.

Blancs_1: OLS affiche un coefficient positif important, PLS un coefficient positif mais bien plus petit, et PCR un coefficient légèrement positif, ce qui est en accord avec la corrélation positive modérée. Cela indique que plus il y a de votes blancs au premier tour, plus il est probable qu'il y en ait au second.

Nuls_1: OLS montre un coefficient négatif important, PLS un coefficient positif très faible, et PCR un coefficient positif, cohérent avec la corrélation positive faible. Cela suggère une tendance légère vers plus de votes nuls associés à plus de votes blancs.

ARTHAUD: Tous les modèles ont des coefficients négatifs, mais avec une corrélation positive faible. Le modèle PCR semble mieux correspondre à la corrélation, bien que légère.

ROUSSEL: Les coefficients dans OLS et PCR sont positifs, et celui de PLS est encore plus positif, en accord avec la forte corrélation positive. Cela indique que l'augmentation de cette variable est associée à une augmentation des votes blancs.

MACRON: Tous les modèles montrent des coefficients négatifs, mais la corrélation est presque nulle, suggérant une faible relation directe entre MACRON et Blancs_2.

LASSALLE: Les coefficients de PLS et PCR sont positifs, et celui de PCR est particulièrement élevé, ce qui correspond à la corrélation positive forte. Cela indique que LASSALLE est positivement associé à Blancs_2.

LE_PEN: Les modèles OLS et PCR ont des coefficients négatifs, et la corrélation est également négative, ce qui suggère une association négative entre LE_PEN et Blancs_2.

ZEMMOUR: OLS et PCR montrent des coefficients négatifs, tandis que PLS montre un coefficient légèrement positif, mais presque nul. La corrélation négative modérée indique une tendance opposée entre ZEMMOUR et Blancs_2.

MELENCHON: OLS montre un coefficient négatif, tandis que PLS et PCR montrent des coefficients positifs, ce qui correspond mieux à la corrélation positive faible.

Le modèle PLS semble, en général, mieux correspondre aux tendances des corrélations, sauf pour quelques cas tels que ARTHAUD. Le modèle PCR correspond bien aux corrélations pour des variables comme ROUSSEL et LASSALLE, tandis que les coefficients du modèle OLS ne correspondent pas toujours à la direction des corrélations observées, ce qui peut indiquer que le modèle OLS ne capte pas

aussi bien les nuances de la relation entre les variables explicatives et la variable Blancs_2.

##	OLS_MACRON_2	PCR1_MACRON_2	PLS1_MACRON_2	cor_MACRON_2
## Abstentions_1	1.0516948	-0.1534	-0.09745560	-0.51845232
## Blancs_1	-0.6825774	0.0458	-0.01190104	0.09938664
## Nuls_1	-0.1826083	-0.1460	-0.09986284	-0.42293649
## ARTHAUD	4.9150763	-0.0093	-0.01589378	-0.08341726
## ROUSSEL	0.5928093	-0.0661	-0.06536244	-0.14868402
## MACRON	2.1780843	0.2431	0.31820397	0.88972810
## LASSALLE	1.1899864	-0.0900	-0.10132229	-0.22501178
## LE_PEN	0.7779997	-0.2106	-0.25957618	-0.81634553
## ZEMMOUR	1.0105803	-0.0825	-0.11830766	-0.32757199
## MELENCHON	1.4511421	0.1418	0.12217606	0.51747339
## HIDALGO	1.7696087	0.0654	0.05038605	0.30554686
## JADOT	1.9063646	0.2481	0.22694481	0.83564709
## PECRESSE	1.5308028	0.1186	0.10357555	0.37862875
## POUTOU	-0.7107576	0.0136	-0.01862577	0.04426909

Abstentions_1: OLS montre un coefficient positif important, ce qui est en désaccord avec la corrélation négative élevée. PLS et PCR montrent tous deux des coefficients négatifs, ce qui est en accord avec la corrélation négative.

Blancs_1: OLS montre un coefficient négatif, PLS montre un coefficient très proche de zéro et PCR un coefficient légèrement positif. La corrélation positive faible est plus en phase avec le modèle PCR.

Nuls_1: Tous les modèles montrent des coefficients négatifs, ce qui est cohérent avec la corrélation négative modérée.

ARTHAUD: OLS présente un coefficient très positif, tandis que PLS et PCR présentent des coefficients négatifs mais très proches de zéro. Cela contraste avec la corrélation négative faible, suggérant un modèle OLS en désaccord avec la corrélation.

ROUSSEL: OLS montre un coefficient positif, tandis que PLS et PCR montrent des coefficients négatifs, ce qui est en accord avec la corrélation négative faible.

MACRON: Tous les modèles montrent des coefficients positifs, avec OLS montrant le plus grand. C'est en accord avec la corrélation positive très élevée, indiquant que lorsque la variable MACRON augmente, la variable cible MACRON_2 tend également à augmenter, ce qui suggère une forte association positive.

LASSALLE: OLS montre un coefficient positif, tandis que PLS et PCR montrent des coefficients négatifs. La corrélation négative modérée suggère que les modèles PLS et PCR sont plus en ligne avec la tendance observée.

LE_PEN: Tous les modèles présentent des coefficients positifs, à l'exception de PCR qui montre un coefficient négatif significatif. Cependant, la corrélation est fortement négative, ce qui signifie que seule la représentation PCR correspond à la tendance de corrélation, indiquant que les valeurs élevées pour LE_PEN sont associées à des valeurs plus faibles pour MACRON_2.

ZEMMOUR: OLS montre un coefficient positif, tandis que PLS et PCR montrent des coefficients négatifs. La corrélation négative indique que les modèles PLS et PCR correspondent mieux à la corrélation.

MELENCHON: Tous les modèles montrent des coefficients positifs, ce qui est en accord avec la corrélation positive modérée, indiquant que les valeurs élevées pour MELENCHON sont associées à des valeurs élevées pour MACRON_2.

En général, les coefficients de régression dans les modèles PLS et PCR semblent souvent mieux refléter les corrélations sous-jacentes entre les variables explicatives et la variable cible MACRON_2 par rapport au modèle OLS. Le modèle PLS, en particulier, semble harmoniser avec le sens des corrélations (positive ou négative) pour la majorité des variables, bien que l'amplitude des coefficients dans le modèle PLS ne soit pas toujours aussi marquée que dans le modèle OLS.

##	OLS_LE_PEN_2	PCR1_LE_PEN_2	PLS1_LE_PEN_2	cor_LE_PEN_2
## Abstentions_1	-0.6962571	0.1386	-0.155838127	0.24287345
## Blancs_1	-0.9307184	0.0090	-0.001032812	-0.01735018
## Nuls_1	-0.5356195	0.1364	-0.013985933	0.34118092
## ARTHAUD	-1.8698474	0.0928	0.128925501	0.29273363
## ROUSSEL	-0.3525164	0.0452	-0.025683884	0.09002858
## MACRON	-0.8307995	-0.1774	-0.183216147	-0.64985741
## LASSALLE	-0.3645434	0.0311	0.036520168	0.14606409
## LE_PEN	0.5348316	0.2212	0.471642445	0.97091923
## ZEMMOUR	-0.1352873	0.0175	0.139824259	0.22224083
## MELENCHON	-0.7120722	-0.1691	-0.281944602	-0.67777978
## HIDALGO	-0.7263454	-0.0835	0.004027716	-0.21041426
## JADOT	-0.7531261	-0.2224	-0.180931116	-0.74852696
## PECRESSE	-0.5408793	-0.0918	-0.114434762	-0.35048254
## POUTOU	0.5752433	0.0228	0.031832963	0.05272973

Abstentions_1: OLS et PLS présentent des coefficients négatifs, alors que PCR présente un coefficient positif. La corrélation positive indique que PCR est en accord avec la direction de la relation observée, suggérant que des taux d'abstention plus élevés sont associés à une augmentation pour LE_PEN_2.

Blancs_1: Les coefficients de tous les modèles sont négatifs ou proches de zéro, ce qui est généralement en accord avec la corrélation négative faible.

Nuls_1: OLS présente un coefficient négatif, PCR un coefficient positif, et PLS un coefficient négatif. La corrélation positive suggère que PCR correspond mieux à la relation observée.

ARTHAUD: Les coefficients de OLS sont négatifs, alors que PCR et PLS sont positifs. La corrélation positive indique que les modèles PCR et PLS sont en phase avec la direction de la corrélation, suggérant une relation positive entre ARTHAUD et LE_PEN_2.

ROUSSEL: OLS présente un coefficient négatif, PCR un coefficient positif, et PLS un coefficient négatif. La corrélation positive faible suggère une relation légèrement positive, que le modèle PCR capture mieux.

MACRON: Tous les modèles présentent des coefficients négatifs, ce qui est cohérent avec la corrélation négative forte. Cela indique que des valeurs élevées pour MACRON sont associées à des valeurs plus faibles pour LE_PEN_2.

LASSALLE: Les coefficients de OLS et PLS sont négatifs et PCR positif. La corrélation positive indique que le modèle PCR est en accord avec la relation observée.

LE_PEN: Tous les modèles montrent des coefficients positifs, ce qui est en parfait accord avec la corrélation très forte positive. Cela indique que des valeurs élevées pour LE_PEN sont fortement associées à des valeurs élevées pour LE_PEN_2, ce qui était attendu.

ZEMMOUR: OLS présente un coefficient négatif, tandis que PCR et PLS sont positifs. La corrélation positive indique que les modèles PCR et PLS reflètent mieux la relation positive observée.

MELENCHON: Les coefficients de OLS et PLS sont négatifs, et PCR est également négatif. La corrélation négative forte suggère que ces modèles capturent correctement la direction de la relation, indiquant que des valeurs élevées pour MELENCHON sont associées à des valeurs plus faibles pour LE_PEN_2.

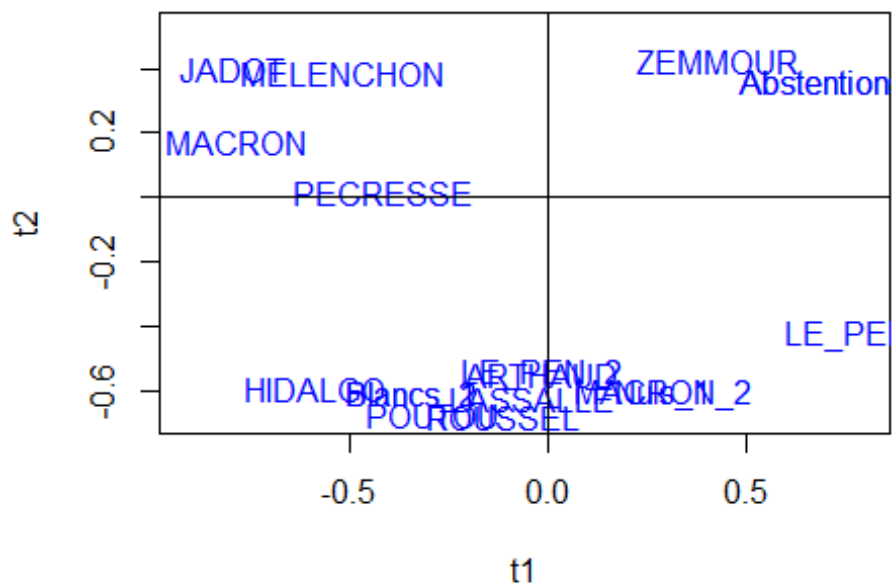
En somme, pour la variable LE_PEN_2, les modèles PCR et PLS semblent souvent en accord avec les signes des corrélations fournies, suggérant qu'ils peuvent mieux capturer la relation entre les variables explicatives et la variable cible dans cet ensemble de données. Le modèle OLS, quant à lui, présente des incohérences dans la direction de certaines relations, par exemple pour Abstentions_1 et ZEMMOUR.

Question 4

PLS2

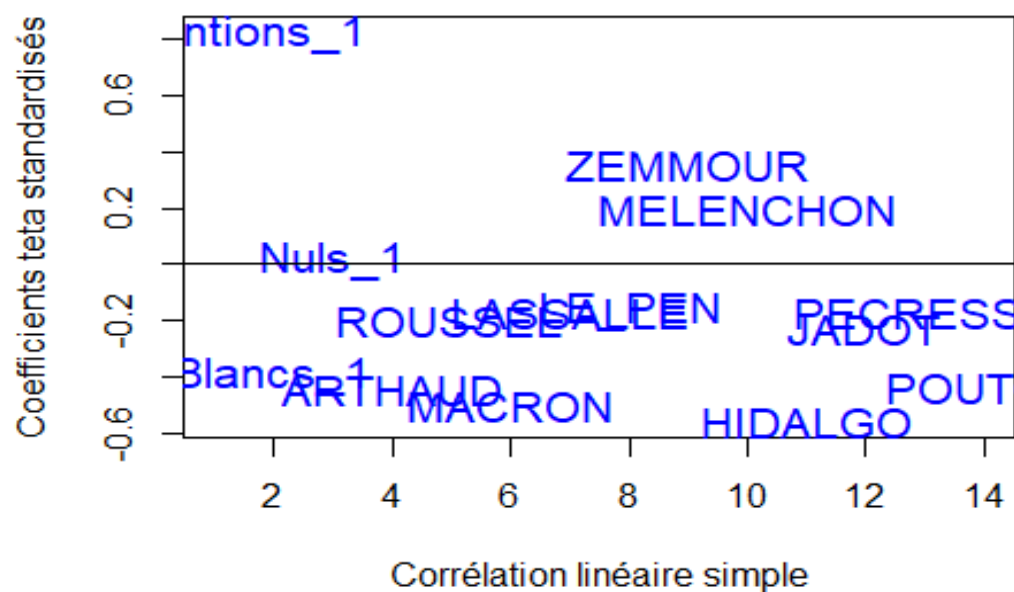
```
##
## PLS Regression 2
## -----
## $x.scores      X-scores (T-components)
## $x.loads       X-loadings
## $y.scores      Y-scores (U-components)
## $y.loads       Y-loadings
## $cor.xt        X,T correlations
## $cor.yt        Y,T correlations
## $cor.xu        X,U correlations
## $cor.yu        Y,U correlations
## $cor.tu        T,U correlations
## $raw.wgs       raw weights
## $mod.wgs       modified weights
## $std.coefs     standard coefficients
## $reg.coefs     regular coefficients
## $y.pred        Y-predicted
## $resid         residuals
## $expvar        explained variance
## $VIP           variable importance for projection
## $Q2            Q2 index
## $Q2cum         cumulated Q2
## -----
```


Premier plan des variables (t1,t2)

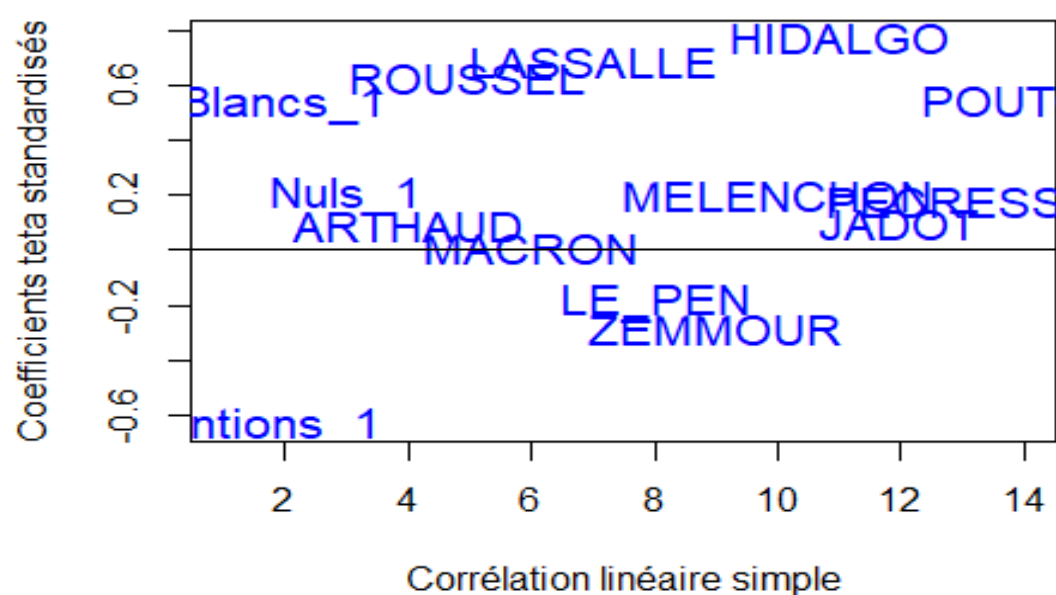


##	Abstentions_2	Blancs_2	MACRON_2	LE_PEN_2
## Abstentions_1	0.5768	-0.2380	-0.0915	-0.1668
## Blancs_1	-0.0192	0.1203	0.0337	-0.0583
## Nuls_1	0.1684	0.0411	-0.1040	-0.0004
## ARTHAUD	-0.1559	-0.0523	-0.0075	0.1025
## ROUSSEL	0.0455	0.2145	-0.0547	-0.0255
## MACRON	-0.2002	-0.0688	0.2698	-0.1494
## LASSALLE	-0.0400	0.2772	-0.0987	0.0532
## LE_PEN	-0.3409	-0.0392	-0.2609	0.4656
## ZEMMOUR	-0.0519	-0.0028	-0.1230	0.1602
## MELENCHON	0.2236	0.0876	0.1512	-0.2989
## HIDALGO	-0.2339	0.2701	0.0534	0.0059
## JADOT	-0.0824	0.0134	0.2399	-0.2024
## PECRESSE	-0.0160	0.0412	0.1118	-0.1165
## POUTOU	-0.0769	0.1360	0.0106	-0.0064

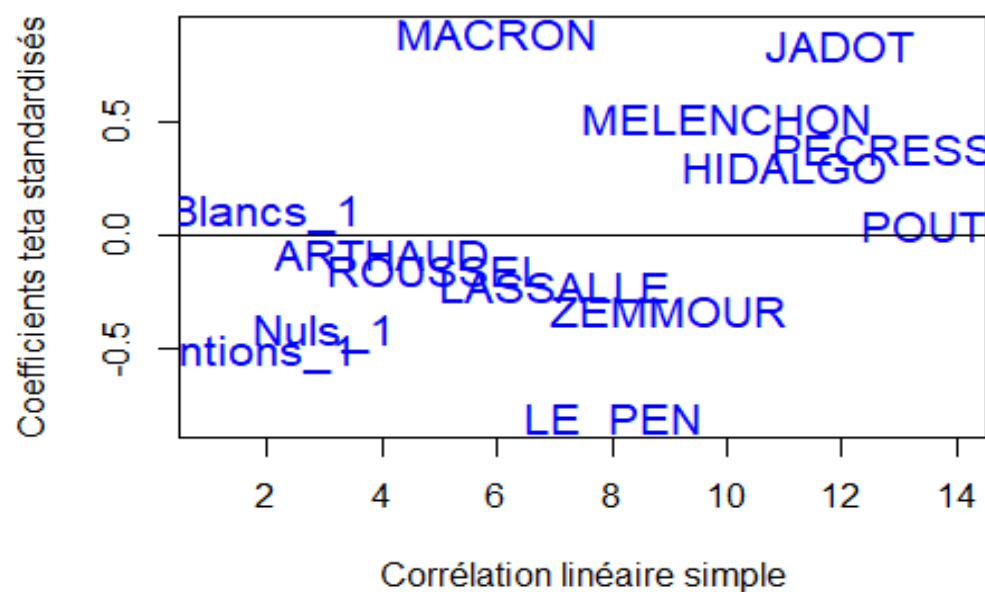
Abstentions_2



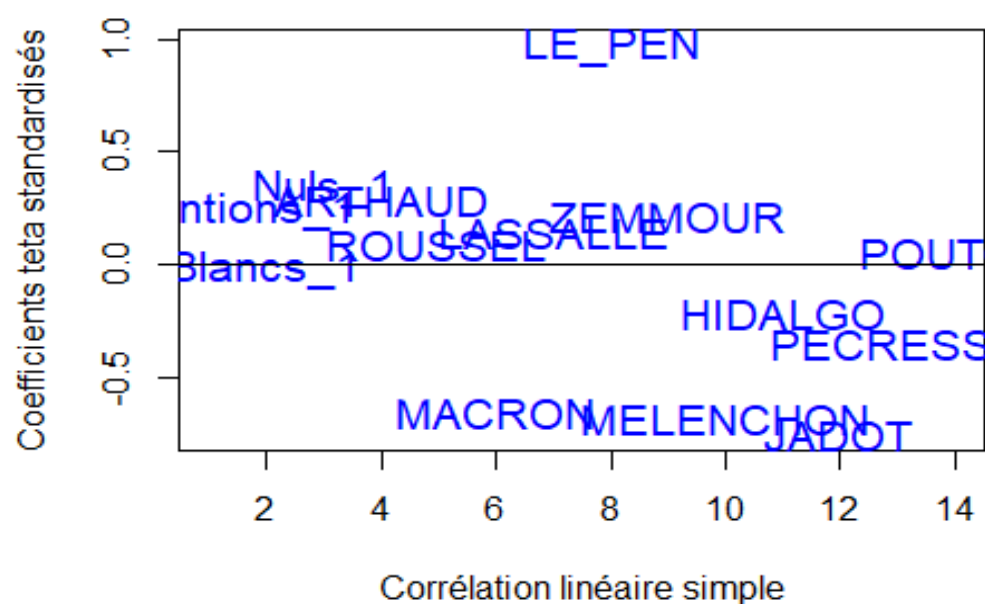
Blancs_2

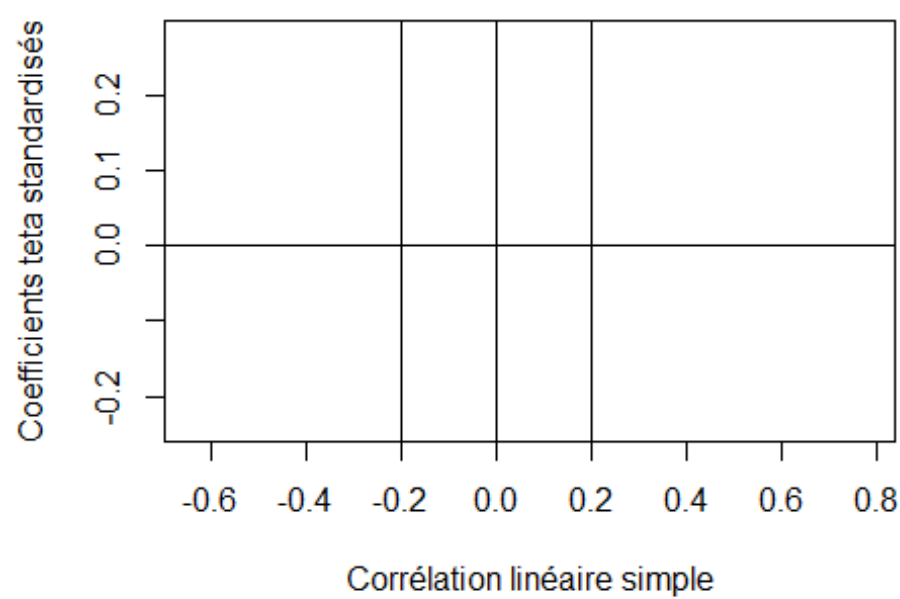
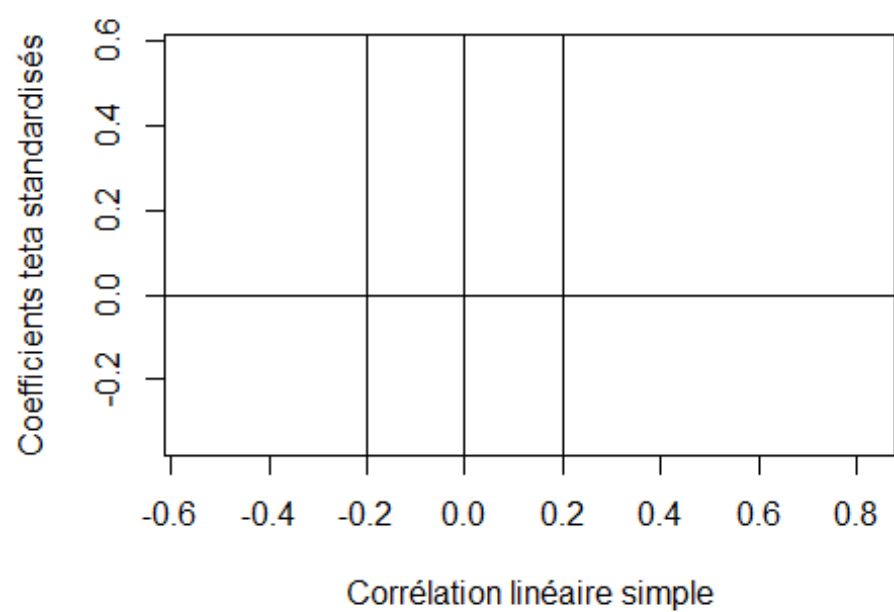


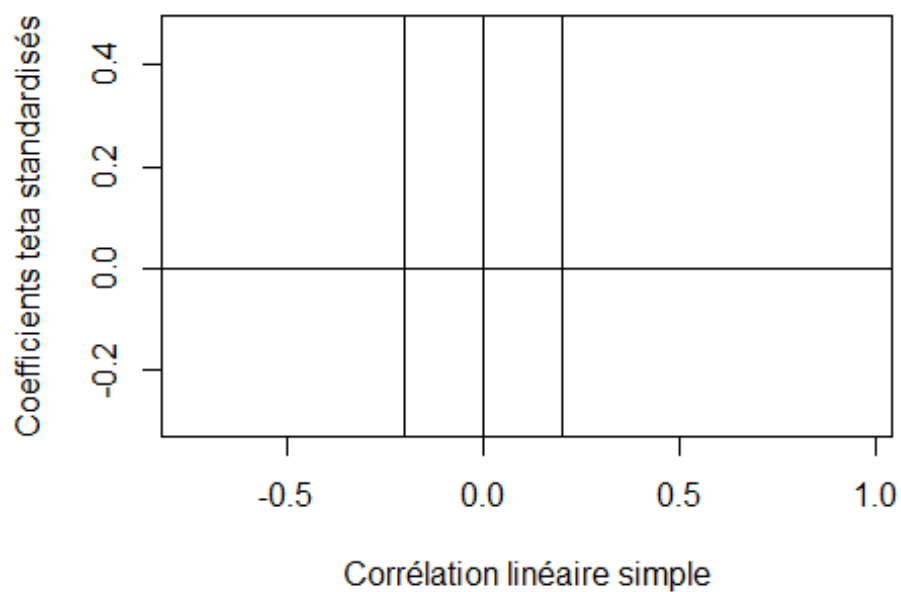
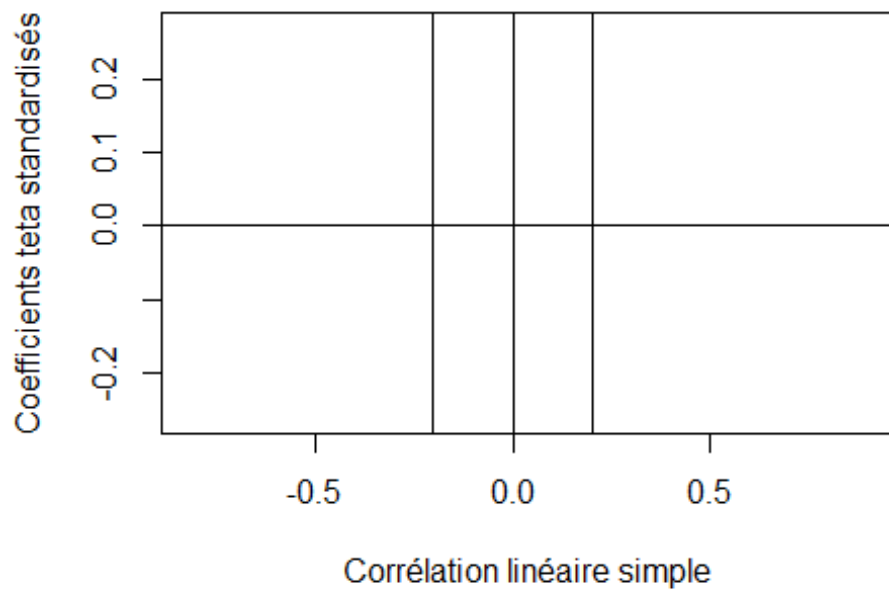
MACRON_2



LE_PEN_2







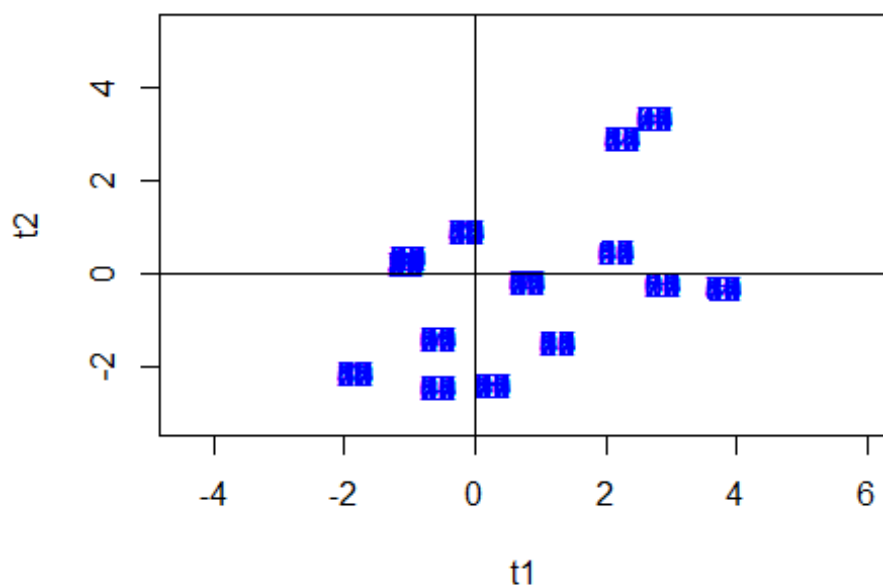
```
##
## PLS Regression 2
## -----
## $x.scores      X-scores (T-components)
## $x.loads       X-loadings
## $y.scores      Y-scores (U-components)
## $y.loads       Y-loadings
## $cor.xt        X,T correlations
## $cor.yt        Y,T correlations
## $cor.xu        X,U correlations
## $cor.yu        Y,U correlations
## $cor.tu        T,U correlations
## $raw.wgs       raw weights
```

```

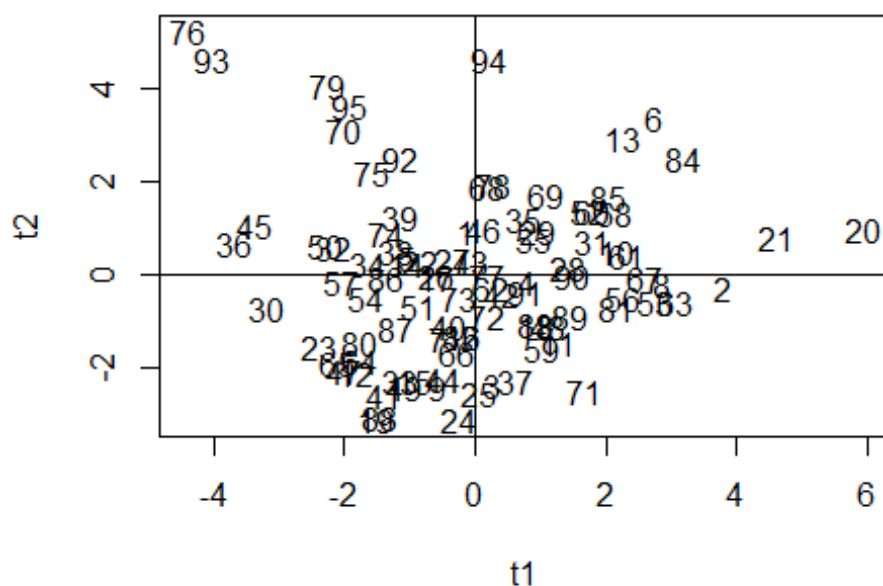
## $mod.wgs      modified weights
## $std.coefs     standard coefficients
## $reg.coefs     regular coefficients
## $y.pred       Y-predicted
## $resid        residuals
## $expvar       explained variance
## $VIP          variable importance for projection
## $Q2           Q2 index
## $Q2cum        cummulated Q2
## -----

```

Premier plan des variables (t1,t2)

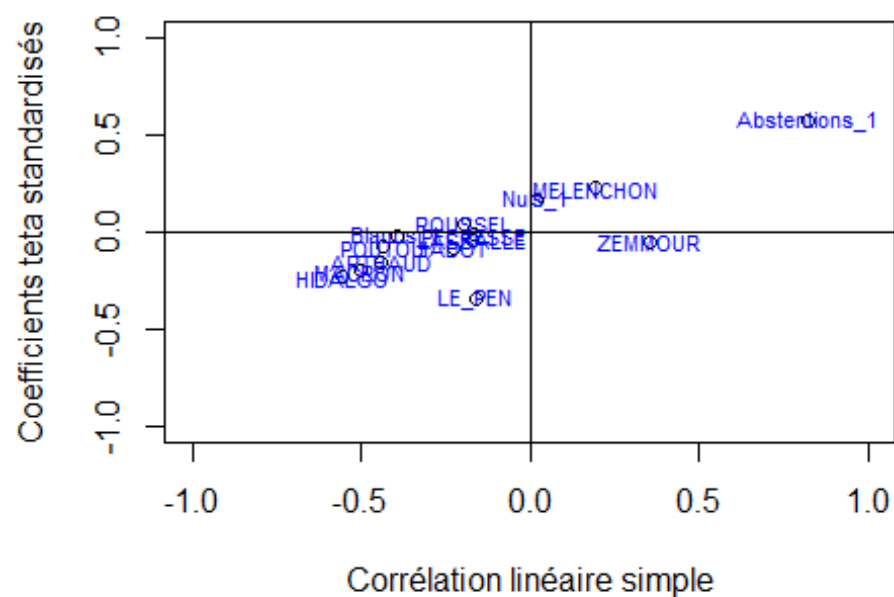


Premier plan des individus (t1,t2)

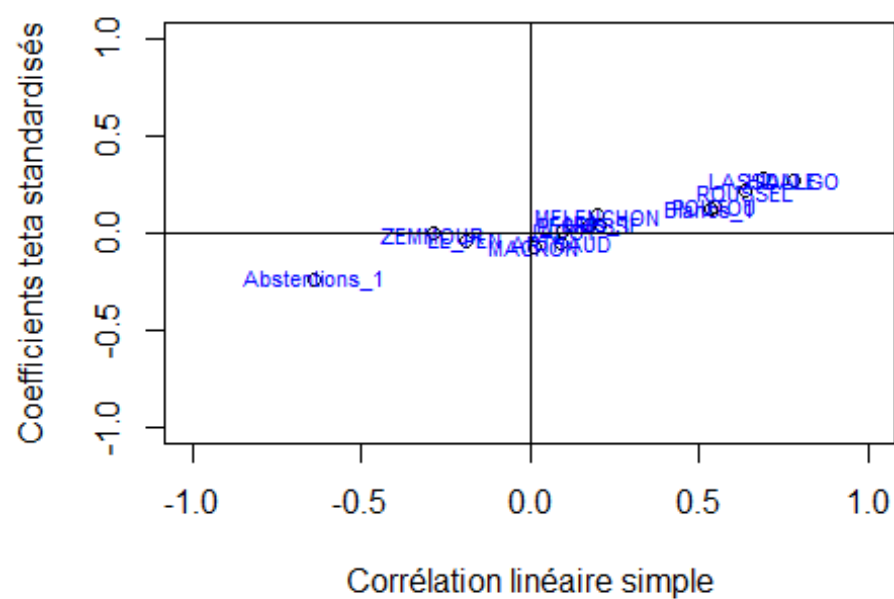


Nous détectons alors que les départements 2, 76, 93, 20 et 21 comme atypiques vis-à-vis du graphique des individus.

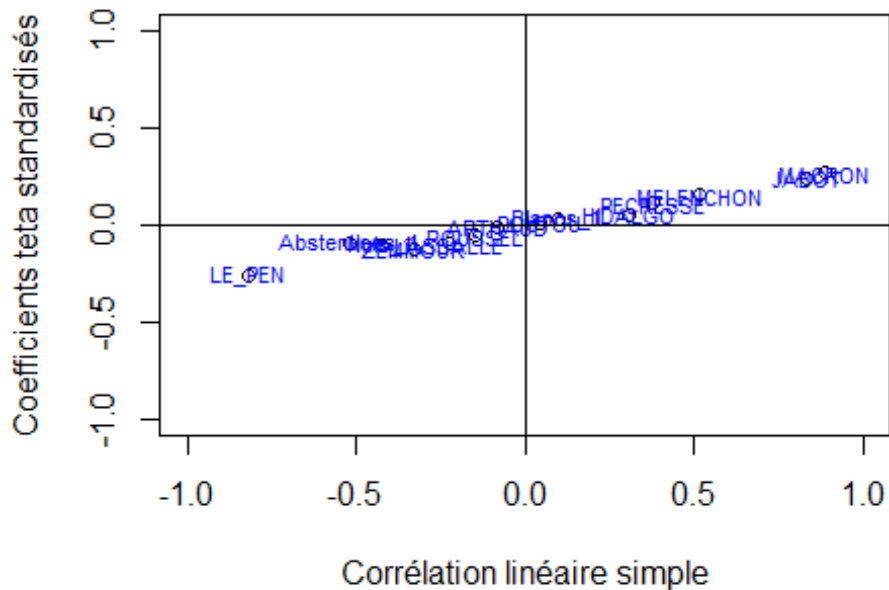
Abstentions_2



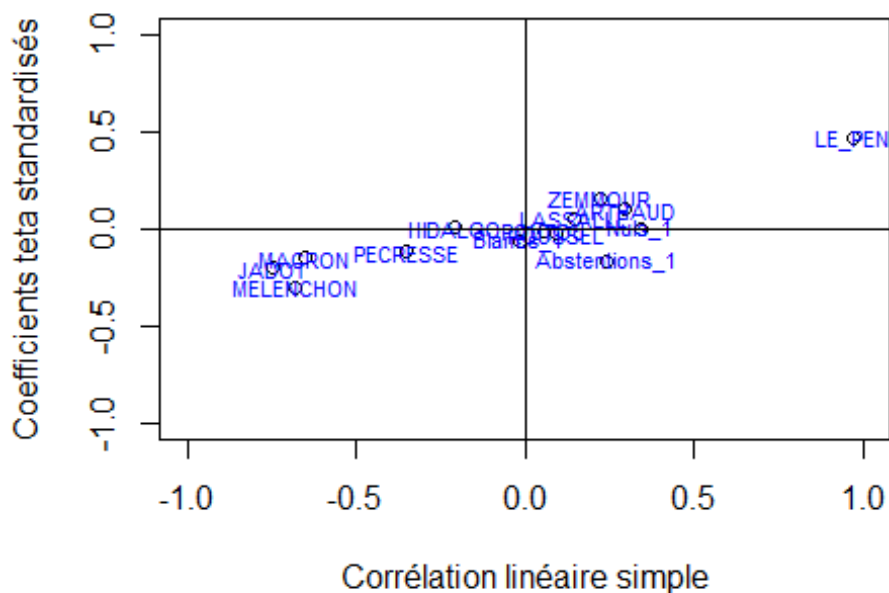
Blancs_2



MACRON_2



LE_PEN_2



Pour les différents graphiques, on constate que le modèle PCR2 est plus concluant car non seulement il respecte l'évolution des signes entre les corrélations et le coefficient mais aussi ils sont des valeurs similaires.

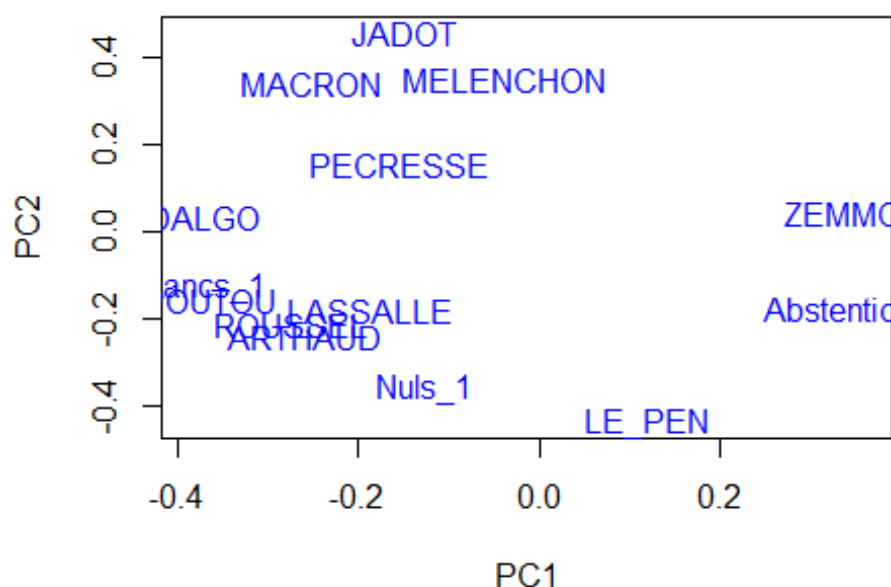
PCR2

```
##
## Call:
## lm(formula = Y[, dep_var] ~ X_pca)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.045655 -0.013176 -0.003671  0.009024  0.096280
##
## Coefficients:
```



```
##          Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2537707  0.0023517 107.911 < 2e-16 ***
## X_pcaPC1     0.0106577  0.0012175   8.754 1.12e-13 ***
## X_pcaPC2    -0.0002087  0.0012688  -0.164  0.8697
## X_pcaPC3     0.0031044  0.0016466   1.885  0.0626 .
## X_pcaPC4     0.0008690  0.0020811   0.418  0.6772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02292 on 90 degrees of freedom
## Multiple R-squared:  0.4718, Adjusted R-squared:  0.4483
## F-statistic: 20.1 on 4 and 90 DF, p-value: 7.462e-12
```

Premier plan des variables (PC1,PC2)



```
##          Abstentions_2      Blancs_2      MACRON_2      LE_PEN_2
## (Intercept)  0.2537707031  0.0504729101  0.376450550  0.3002024707
## X_pcaPC1     0.0106577395 -0.0033275587 -0.013453942  0.0080506037
## X_pcaPC2    -0.0002087053 -0.0003713372  0.027680289 -0.0255308493
## X_pcaPC3     0.0031044379  0.0029957438 -0.005851207 -0.0024715992
## X_pcaPC4     0.0008690354  0.0002982993 -0.001081450  0.0001690841
```

Intercept : La valeur de l'intercept est de 0.2537707, avec une p-value inférieure à 2e-16, ce qui indique que l'intercept est statistiquement significatif. Cela signifie que quand toutes les composantes principales (PC) sont à zéro, la valeur attendue pour la variable dépendante est de 0.2537707.

Première composante principale (X_pcaPC1) : Elle a un coefficient de 0.0106577 et est statistiquement significative ($p < 0.001$). Cela indique que pour chaque augmentation d'une unité de PC1, on s'attend à ce que la variable dépendante augmente de 0.0106577 unités, en tenant compte de l'effet des autres variables.

Deuxième composante principale (X_pcaPC2) : Avec un coefficient proche de zéro (-0.0002087) et une p-value de 0.8697, cette composante n'est pas statistiquement significative. Elle n'a donc pas d'impact notable sur la variable dépendante dans ce modèle.

Troisième composante principale (X_pcaPC3) : Le coefficient est de 0.0031044, avec une p-value juste au-dessus du seuil de signification (0.0626), ce qui suggère une possible mais pas forte association avec

la variable dépendante.

Quatrième composante principale (X_pcaPC4) : Elle a également un coefficient proche de zéro (0.0008690) avec une p-value élevée (0.6772), indiquant qu'elle n'est pas statistiquement significative pour la prédiction de la variable dépendante.

Les R-squared et Adjusted R-squared valeurs de 0.4718 et 0.4483 respectivement indiquent que le modèle explique environ 47.18% de la variance de la variable dépendante, et cette proportion reste proche une fois ajustée pour le nombre de prédicteurs dans le modèle. Le R-squared ajusté est une mesure plus précise de la qualité d'ajustement, car il prend en compte le nombre de prédicteurs et la taille de l'échantillon.

Le Residual standard error de 0.02292 montre la taille typique des erreurs (écarts entre les valeurs prédites et observées).

Le F-statistic est utilisé pour tester l'hypothèse que tous les coefficients de régression sont égaux à zéro. La p-value associée à la statistique F (7.462e-12) indique que nous pouvons rejeter cette hypothèse au niveau de confiance donné, ce qui signifie que le modèle dans son ensemble est significatif.

Les coefficients listés à la fin sont ceux pour différentes variables dépendantes ('Abstentions_2', 'Blancs_2', 'MACRON_2', 'LE_PEN_2') obtenus en utilisant les composantes principales comme prédicteurs dans des modèles séparés pour chaque variable dépendante. Ces coefficients s'appliquent à chaque modèle spécifique pour la variable dépendante respective et reflètent l'influence de chaque composante principale sur ces variables.

PC1 : Il capture la plus grande variance parmi les variables. On voit que LE_PEN et ZEMMOUR sont positivement associés avec PC1, suggérant que ces variables ont une forte variance en commun qui est capturée par PC1. Inversement, des variables comme Nuls_1, LASSALLE, et ARTHAUD sont négativement associées avec PC1.

PC2 : Il capture la deuxième plus grande variance après PC1. JADOT, MELENCHON, et MACRON sont positivement associés avec PC2, indiquant que ces variables contribuent significativement à la variance capturée par PC2. Au contraire, Blancs_1, POUTOU, et d'autres sont en bas, indiquant une association négative avec PC2.

Proximité entre les variables : Les variables qui sont proches les unes des autres sont probablement corrélées dans l'ensemble de données original. Par exemple, MELENCHON et MACRON sont proches, suggérant une corrélation positive entre ces deux variables.

Opposition entre les variables : Les variables qui sont sur des côtés opposés du biplot (par rapport à l'origine) sont probablement négativement corrélées. Par exemple, LE_PEN et Nuls_1 semblent être négativement corrélés puisqu'ils sont opposés par rapport à l'origine sur l'axe PC1.

Relation avec les axes : Les variables qui sont alignées le long d'un axe sont fortement influencées par cet axe. LE_PEN est aligné avec PC1, ce qui signifie que les variations de LE_PEN sont une source majeure de la variance capturée par PC1.

Importance relative des axes : PC1 semble avoir une plus grande dispersion des variables que PC2, ce qui suggère que PC1 explique une plus grande part de la variance totale dans les données.

Conclusion

La régression PLS (Partial Least Squares), y compris ses variantes PLS1 et PLS2, offre une solution robuste au problème de multi-colinéarité souvent présent dans les ensembles de données avec un grand nombre de prédicteurs corrélés. Elle se distingue par sa capacité à fournir des résultats interprétables qui permettent de comprendre les relations entre les variables explicatives et les réponses. Grâce à l'algorithme NIPALS, la PLS peut gérer efficacement les données manquantes et maintenir la structure de liaison entre les variables, ce qui est particulièrement avantageux lorsqu'il y a des dépendances entre les variables à expliquer.

En contraste, la régression sur composantes principales (PCR), bien que traitant également la multi-colinéarité par réduction de dimensionnalité, sélectionne des composantes basées sur la variance au sein des prédicteurs plutôt que sur la covariance avec les variables de réponse. Alors que la PLS ajuste les composantes avec un œil sur la variable de réponse dès le début, la PCR pourrait manquer des informations pertinentes pour la prédiction si la variance des prédicteurs ne reflète pas les relations avec la variable de réponse.

La PLS2 est particulièrement efficace pour modéliser des réponses multivariées, contrairement à la PCR2 qui pourrait ne pas capturer complètement la corrélation entre les variables réponses si elles sont interdépendantes. En effet, la modélisation séparée de variables réponses sans tenir compte de leur structure de liaison peut entraîner une perte d'information.

Néanmoins, il est essentiel de reconnaître que les coefficients obtenus par PCR1, PCR2, PLS1 et PLS2 ne sont pas de variance minimale et sont influencés par le processus de réduction dimensionnelle, ce qui introduit un biais. Par conséquent la loi sous l'hypothèse nulle : $\beta = 0$, par exemple, à l'aide de la statistique de Student n'est plus utilisable. Seuls sont valides les tests réalisés sur les composantes principales et PLS.