

# Feuille de Travaux Dirigés n° 6

## Analyse factorielle des correspondances

### Exercice VI.1. Présidentielles

1. Récupérer les données dans R en exécutant les instructions suivantes. Penser à remplacer "C :\\..." par le répertoire dans lequel vous avez enregistré le fichier que vous souhaitez ouvrir.

```
> Chemin <- "C:\\..."
> presid <- read.csv(paste(Chemin, "Presidentielle.CSV",
+   sep = ""), row.names = 1)
```

2. Quelles sont les différentes variables reproduites dans le tableau au verso ? Quelle est leur nature ? Quelles sont les modalités sur lesquelles on va faire porter l'analyse factorielle des correspondances ? Obtenir les statistiques descriptives, en utilisant la fonction `summary`, et les représentations graphiques suivantes du jeu de données.

|   | Sarkozy          | Bayrou           | Royal            | Le.Pen          |
|---|------------------|------------------|------------------|-----------------|
| 1 | Min. : 56819     | Min. : 18979     | Min. : 33493     | Min. : 23432    |
| 2 | 1st Qu. : 289229 | 1st Qu. : 165316 | 1st Qu. : 217393 | 1st Qu. : 91391 |
| 3 | Median : 362391  | Median : 214259  | Median : 322212  | Median :143926  |
| 4 | Mean : 492330    | Mean : 293479    | Mean : 408817    | Mean :166272    |
| 5 | 3rd Qu. : 544817 | 3rd Qu. : 379599 | 3rd Qu. : 554790 | 3rd Qu. :182804 |
| 6 | Max. :1931429    | Max. :1143081    | Max. :1593033    | Max. :430553    |

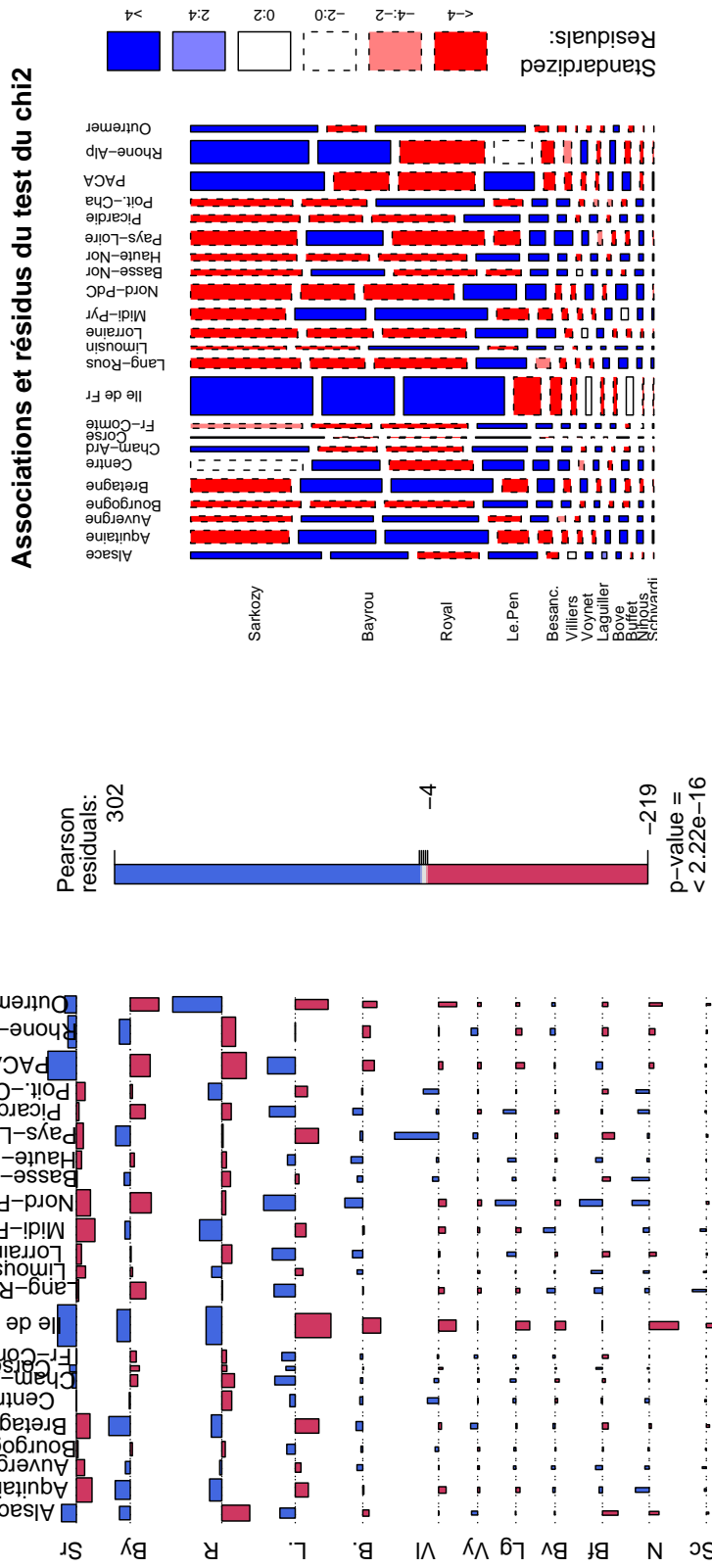
|   | Besanc.         | Villiers        | Voynet         | Laguiller      |
|---|-----------------|-----------------|----------------|----------------|
| 1 | Min. : 5941     | Min. : 1908     | Min. : 2119    | Min. : 1346    |
| 2 | 1st Qu. : 37473 | 1st Qu. : 21474 | 1st Qu. :13284 | 1st Qu. :13690 |
| 3 | Median : 58312  | Median : 28166  | Median :20382  | Median :18623  |
| 4 | Mean : 64976    | Mean : 35469    | Mean :24793    | Mean :21152    |
| 5 | 3rd Qu. : 83281 | 3rd Qu. : 40957 | 3rd Qu. :29837 | 3rd Qu. :25618 |
| 6 | Max. :181247    | Max. :107895    | Max. :89885    | Max. :52965    |

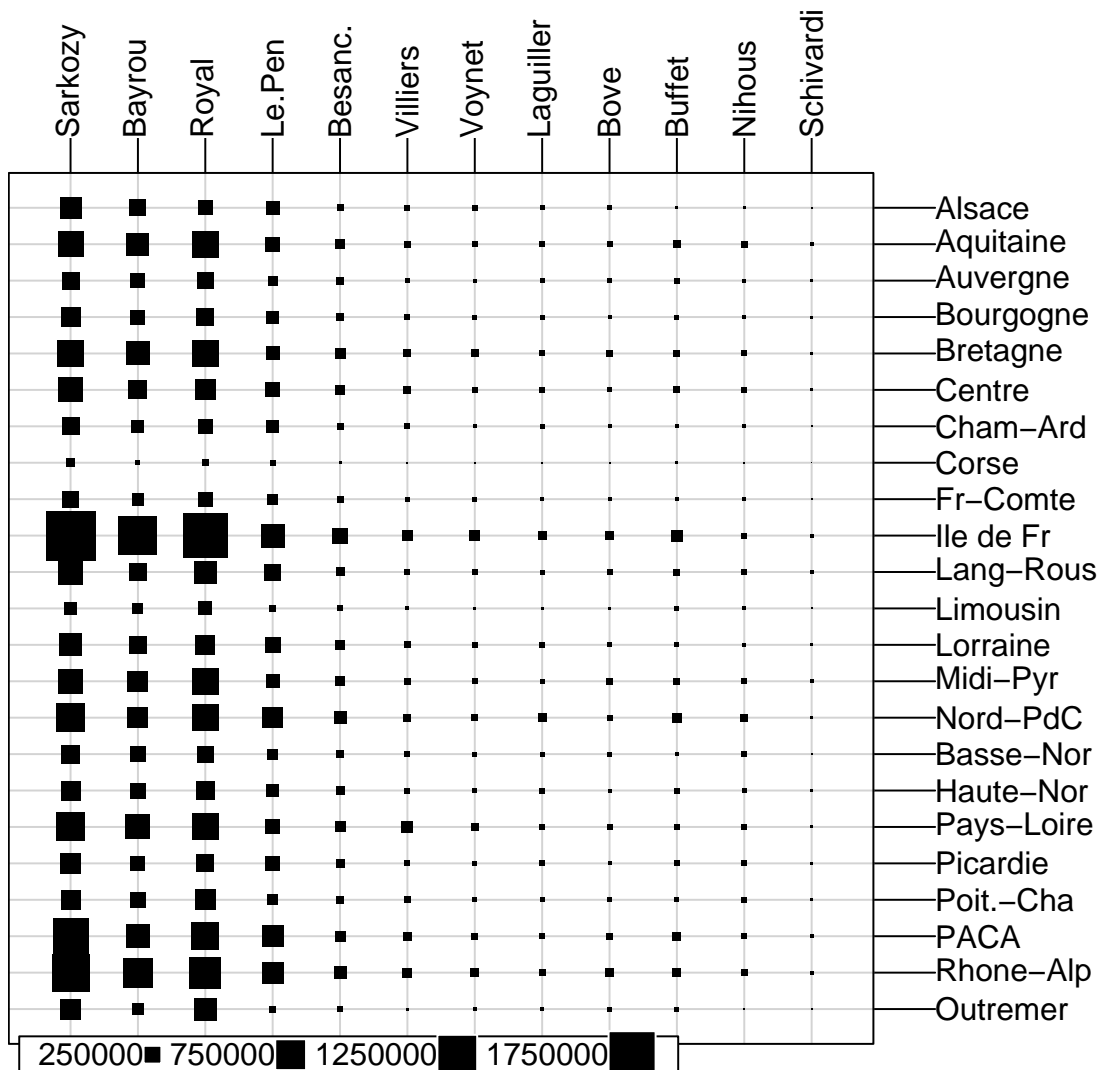
```
> library(graphics)
> install.packages("vcd")
> library(vcd)
> presidtable <- as.table(as.matrix(presid[, -c(13, 14)]))
> assoc(presidtable, shade = T, las = 2, rot_labels = 90,
+   labeling_args = list(abbreviate = c(A = TRUE), varnames = F,
+   main = "Associations et résidus du test du chi2"))
```

|   | Bove           | Buffet          | Nihous         | Schivardi      |
|---|----------------|-----------------|----------------|----------------|
| 1 | Min. : 1659    | Min. : 5163     | Min. : 2260    | Min. : 450     |
| 2 | 1st Qu. :12430 | 1st Qu. : 14878 | 1st Qu. : 9697 | 1st Qu. : 2806 |
| 3 | Median :14904  | Median : 22334  | Median :20439  | Median : 4207  |
| 4 | Mean :20832    | Mean : 30673    | Mean :18265    | Mean : 5361    |
| 5 | 3rd Qu. :27677 | 3rd Qu. : 36333 | 3rd Qu. :23426 | 3rd Qu. : 7026 |
| 6 | Max. :57453    | Max. :110967    | Max. :43595    | Max. :12386    |

|   | Sarkozy_2        | Royal_2          |
|---|------------------|------------------|
| 1 | Min. : 93771     | Min. : 62209     |
| 2 | 1st Qu. : 470508 | 1st Qu. : 408900 |
| 3 | Median : 641324  | Median : 583406  |
| 4 | Mean : 817371    | Mean : 723216    |
| 5 | 3rd Qu. : 915508 | 3rd Qu. : 987746 |
| 6 | Max. :2885762    | Max. :2612107    |

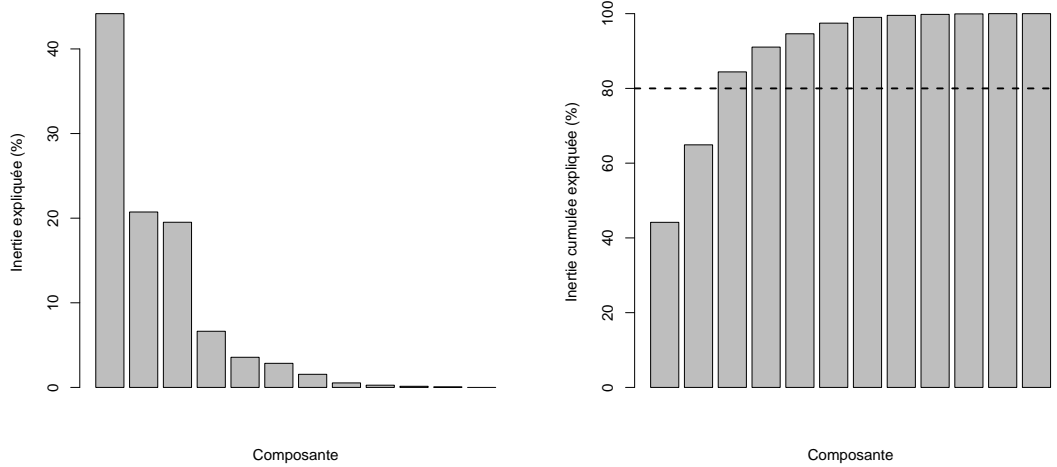
```
> mosaicplot(presidtable, type = "pearson", shade = T,
+   las = 2, main = "Associations et résidus du test du chi2")
> install.packages("ade4")
> library(ade4)
> table.value(presidtable, grid = T, col.labels = names(presid))
```





3. Faire l'analyse factorielle des correspondances du tableau de données en traitant les résultats du second tour, colonnes Sarkozy\_2 et Royal\_2, comme des colonnes supplémentaires, puis construire les diagrammes des valeurs propres suivants. Par combien d'axes l'information est-elle résumée de manière satisfaisante ?

```
> library(FactoMineR)
> res.ca <- CA(presid, col.sup = c(13, 14), graph = FALSE)
> barplot(res.ca$eig$per, ylab = "Inertie expliquée (%)",
+         xlab = "Composante")
> barplot(res.ca$eig$cum, ylab = "Inertie cumulée expliquée (%)",
+         xlab = "Composante")
> abline(h = 80, lty = 2, lwd = 2)
```



```
> res.ca$call$marge.row
```

| res.ca.call.marge.row |         |
|-----------------------|---------|
| Alsace                | 0.02751 |
| Aquitaine             | 0.05360 |
| Auvergne              | 0.02333 |
| Bourgogne             | 0.02677 |
| Bretagne              | 0.05507 |
| Centre                | 0.04073 |
| Cham-Ard              | 0.02070 |
| Corse                 | 0.00422 |
| Fr-Comte              | 0.01886 |
| Ile de Fr             | 0.15695 |
| Lang-Rous             | 0.04237 |
| Limousin              | 0.01284 |
| Lorraine              | 0.03747 |
| Midi-Pyr              | 0.04787 |
| Nord-PdC              | 0.06293 |
| Basse-Nor             | 0.02502 |
| Haute-Nor             | 0.02899 |
| Pays-Loire            | 0.05936 |
| Picardie              | 0.03006 |
| Poit.-Cha             | 0.02964 |
| PACA                  | 0.07500 |
| Rhone-Alp             | 0.09433 |
| Outremer              | 0.02637 |

```
> res.ca$call$marge.col
```

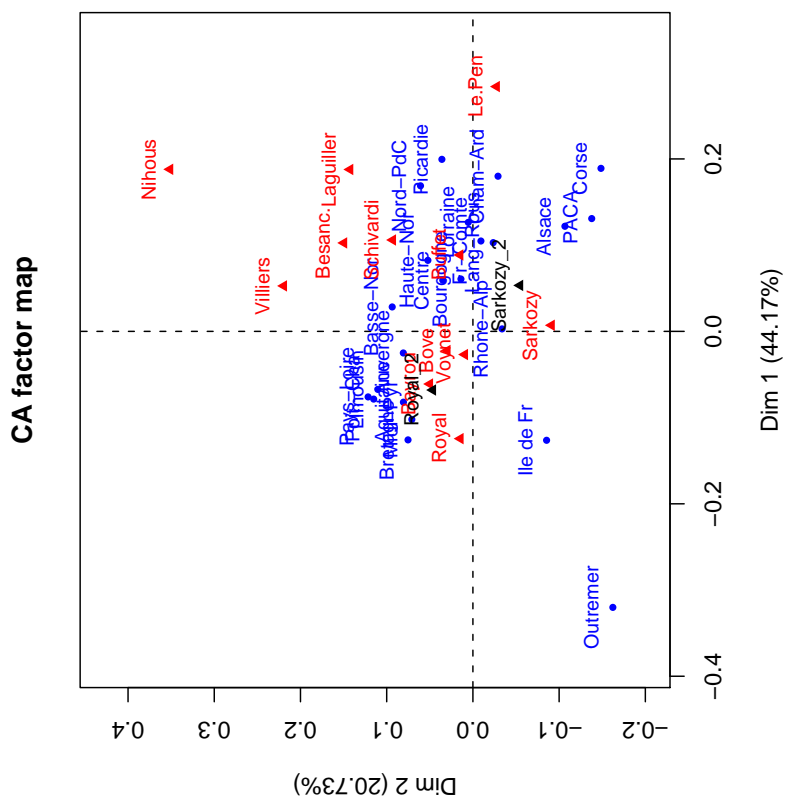
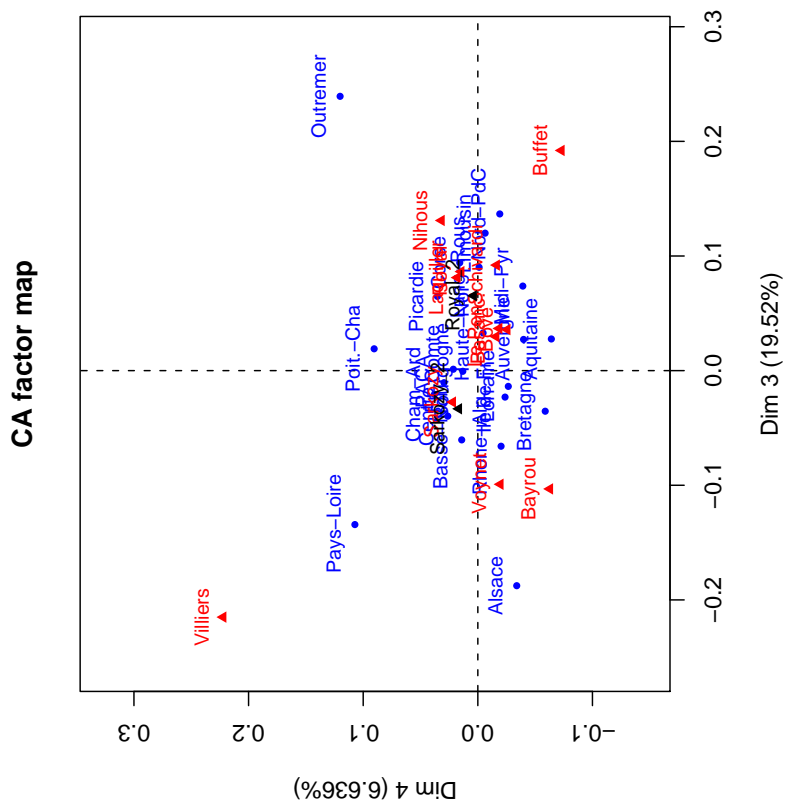
3. Représenter le premier et le second plan factoriel.

---

|           | res.ca.call.marge.col |
|-----------|-----------------------|
| Sarkozy   | 0.31113               |
| Bayrou    | 0.18546               |
| Royal     | 0.25835               |
| Le.Pen    | 0.10507               |
| Besanc.   | 0.04106               |
| Villiers  | 0.02241               |
| Voynet    | 0.01567               |
| Laguiller | 0.01337               |
| Bove      | 0.01316               |
| Buffet    | 0.01938               |
| Nihous    | 0.01154               |
| Schivardi | 0.00339               |

---

```
> plot(res.ca, col.col.sup = "black", cex = 0.8)
> plot(res.ca, col.col.sup = "black", axes = c(3, 4),
+      cex = 0.8)
```



Commenter la qualité de la représentation obtenue sur les quatre premiers axes factoriels en analysant le contenu des tableaux suivants :

- `res.ca$col$coord`, `res.ca$col$cos2`, `res.ca$col$contrib`.
- `res.ca$row$coord`, `res.ca$row$cos2`, `res.ca$row$contrib`.

4. La fonction `dimdesc` permet d'obtenir une description automatique des axes de l'AFC. Commenter ses résultats lorsqu'elle est appliquée à `res.ca`.



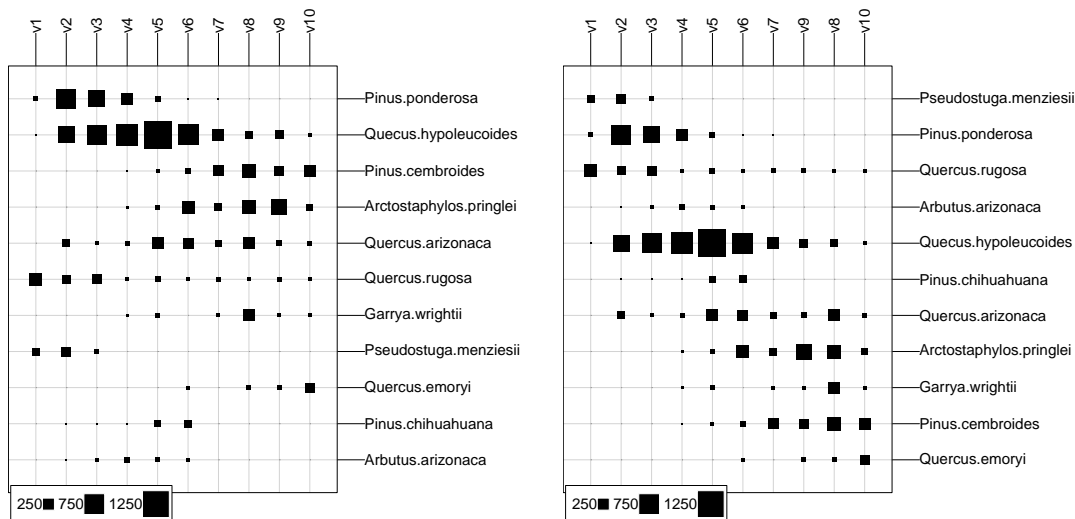
|            | Sarkozy | Bayrou  | Royal   | Le Pen | Besanc. | Villiers | Voynet | Laguiller | Bove  | Buffet | Nilhous | Schivardi | Sarkozy_2 | Royal_2 |
|------------|---------|---------|---------|--------|---------|----------|--------|-----------|-------|--------|---------|-----------|-----------|---------|
| Alsace     | 362391  | 214259  | 171282  | 135730 | 33310   | 22492    | 20382  | 13821     | 13758 | 6100   | 5142    | 2522      | 641324    | 337781  |
| Aquitaine  | 532127  | 417546  | 557300  | 168664 | 78230   | 34028    | 28285  | 22046     | 27941 | 41791  | 35300   | 7572      | 909798    | 971807  |
| Auvergne   | 238152  | 169395  | 225477  | 78704  | 41522   | 18730    | 12090  | 12936     | 13532 | 21920  | 12474   | 4207      | 405352    | 415650  |
| Bourgogne  | 297544  | 175213  | 241094  | 119041 | 42246   | 24971    | 13690  | 14440     | 12296 | 18154  | 12079   | 3608      | 515827    | 434478  |
| Bretagne   | 557507  | 451988  | 564100  | 143926 | 94205   | 41212    | 39026  | 25662     | 28484 | 31860  | 21207   | 5169      | 921218    | 1023094 |
| Centre     | 460425  | 278175  | 345352  | 168912 | 65347   | 45720    | 22655  | 22279     | 17395 | 30003  | 20567   | 5696      | 795959    | 650777  |
| Cham-Ard   | 246680  | 122642  | 160280  | 114527 | 33424   | 20455    | 10727  | 13560     | 7414  | 12465  | 9016    | 2270      | 429708    | 304449  |
| Corse      | 56819   | 18979   | 33493   | 23432  | 5941    | 1908     | 2119   | 1346      | 1659  | 5163   | 2260    | 450       | 93771     | 62209   |
| Fr-Comte   | 212358  | 114148  | 165243  | 94212  | 30672   | 16361    | 12879  | 10880     | 10365 | 9204   | 7814    | 2446      | 371156    | 297492  |
| Ile de Fr  | 1931429 | 1143081 | 1593033 | 430553 | 181247  | 89498    | 89885  | 52965     | 57453 | 110967 | 19890   | 12386     | 2885762   | 2612107 |
| Lang-Rous  | 470017  | 234739  | 395509  | 214468 | 62597   | 28166    | 20787  | 17175     | 27412 | 38590  | 21356   | 11436     | 809868    | 687158  |
| Limousin   | 123870  | 82445   | 142237  | 38525  | 24040   | 10789    | 6144   | 6566      | 6629  | 15695  | 8029    | 2284      | 207618    | 249470  |
| Lorraine   | 403919  | 250195  | 315596  | 196696 | 70940   | 29183    | 21562  | 25574     | 16094 | 19229  | 10378   | 4211      | 715247    | 593561  |
| Midi-Pyr   | 458093  | 341651  | 542038  | 154777 | 69177   | 30850    | 25267  | 18623     | 33687 | 34076  | 25280   | 8778      | 787351    | 898194  |
| Nord-PdC   | 639390  | 340679  | 573071  | 335855 | 127881  | 40702    | 31388  | 52695     | 24591 | 74027  | 43595   | 6348      | 1131904   | 1117952 |
| Basse-Nor  | 280914  | 184256  | 209308  | 88569  | 43997   | 25722    | 14389  | 14652     | 13180 | 11211  | 21449   | 3090      | 476255    | 402149  |
| Haute-Nor  | 309924  | 184615  | 257664  | 126795 | 58312   | 25692    | 15707  | 20048     | 12563 | 26376  | 13717   | 3604      | 543404    | 492154  |
| Pays-Loire | 636934  | 457560  | 552280  | 158844 | 93685   | 107895   | 38952  | 28481     | 25811 | 26737  | 26674   | 6481      | 1075685   | 1003684 |
| Picardie   | 331053  | 161236  | 251862  | 168699 | 57769   | 26731    | 14219  | 23982     | 11619 | 22334  | 20439   | 4189      | 582715    | 487508  |
| Poit.-Cha  | 304493  | 194126  | 322212  | 88138  | 45638   | 38735    | 16333  | 13949     | 13934 | 15901  | 21571   | 3882      | 516282    | 542882  |
| PACA       | 1010234 | 419161  | 579036  | 377831 | 88331   | 54851    | 38339  | 26467     | 36963 | 61968  | 26538   | 9935      | 1662131   | 1025774 |
| Rhone-Alp  | 1121615 | 689984  | 807220  | 360646 | 123776  | 75959    | 63032  | 38114     | 51441 | 57654  | 32624   | 10969     | 1856430   | 1440224 |
| Outremer   | 337711  | 103933  | 398110  | 36714  | 22159   | 5139     | 12383  | 10234     | 14904 | 14062  | 2698    | 1772      | 464760    | 583406  |

.....

## Exercice VI.2. Effet Guttman

L'effet Guttman traduit une association entre les deux variables qualitatives dont nous étudions la table de contingence. En écologie, il est alors d'usage de parler d'ordination indirecte. Le jeu de données utilisé ci-dessous est un ensemble de Densités à l'hectare de 11 espèces d'arbres par classe de valeurs de l'humidité du sol (moyenne sur plusieurs stations par classe).

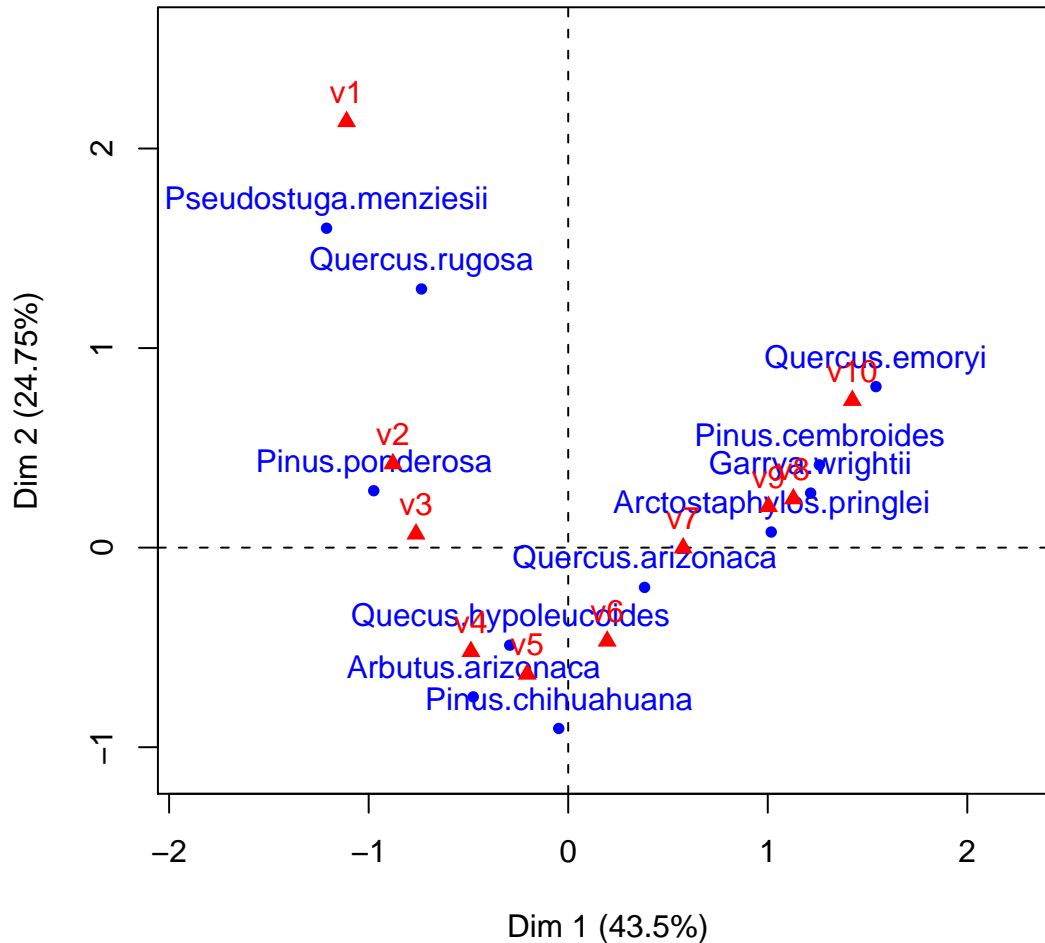
```
> library(FactoMineR)
> data(santacatalina, package = "ade4")
> sant = santacatalina
> coasanta = CA(sant, ncp = 1, graph = F)
> table.value(sant, csize = 1.5)
> table.value(sant[order(coasanta$row$coord), order(coasanta$col$coord)],
+           csize = 1.5)
```



À gauche, les données brutes et à droite, le tableau traité : les sites et les espèces sont ordonnés en fonction des sites, les colonnes, qui sont les classes du gradient.

```
> santa.coa <- CA(santacatalina, ncp = 4, graph = F)
> plot(santa.coa)
```

## CA factor map



Le package `ade4` propose des représentations complémentaires. Voir l'article de J. Oksanen<sup>1</sup> pour plus de détails.

```
> library(ade4)
> santa.coa <- dudi.coa(santacatalina, scan = F)
> scatter(santa.coa, met = 2)
> scatter(santa.coa, met = 3)
```

1. Exécuter les lignes de commande ci-dessus.
2. Traiter l'exemple qui vous a été proposé en cours.

<sup>1</sup>J. Oksanen. Problems of joint display of species and site scores in correspondence analysis. *Vegetatio*, 72 : 51–57, 1987.