

Séminaire d'introduction à R

séance 5 - Analyse géométrique des données

Joël Gombin

CURAPP - UPJV

30 mars 2012

- 1 Introduction : l'analyse géométrique des données
- 2 L'analyse en composantes principales
- 3 L'analyse des correspondances multiples

Introduction : l'analyse géométrique des données

L'analyse géométrique des données (AGD) correspond à une famille de méthodes d'analyse des données, développées par des mathématiciens français (Jean-Paul Benzécri).

Comme beaucoup de méthodes statistiques, l'AGD a au départ partie liée avec la psychologie et la psychométrie, mais se développe rapidement en sociologie (chez Bourdieu notamment).

L'idée consiste à dégager, d'un tableau de données défini par un certain nombre de variables (colonnes) et d'individus (lignes), les structures sous-jacentes à ces données. Pour cela, on procède à partir de l'idée de distance entre deux points, dans un univers à n dimensions (n étant le nombre de variables considérées). Le statut des variables est symétrique : pas de notion de causalité.

On distingue deux grands types de méthodes, selon que les variables soient continues (analyse en composantes principales, ACP) ou discrètes (analyse des correspondances multiples, ACM).

Plan

- 1 Introduction : l'analyse géométrique des données
- 2 L'analyse en composantes principales
- 3 L'analyse des correspondances multiples

Commencer avec l'ACP

On utilisera, tout au long de cette séance, le package FactoMineR (<http://factominer.free.fr>).

A noter

Il existe une interface graphique pour utiliser FactoMineR. Il faut installer le package Rcmdr (voir procédure sur le site de FactoMineR).

```
library(FactoMineR)
load("/media/HDD/Dropbox/Thèse/séminaire R/séance 1/mini_picardie.Rdata")
data <- mini_picardie[, c(81, 85, seq(90, 112,
  2), 114, 115, 116, 118, 21)]
```

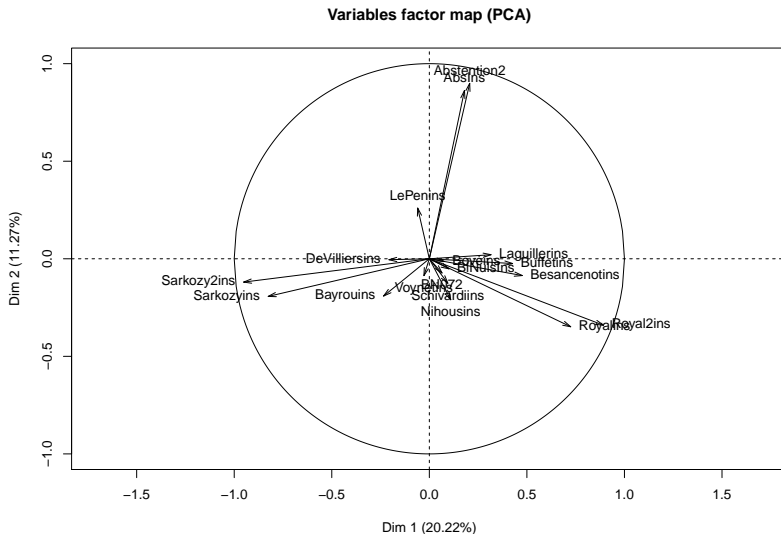
Mettre en place l'ACP

```
acp1 <- PCA(data, quali.sup = 19, graph = F, ind.sup = 1699)
```

On peut alors dessiner le cercle des corrélations :

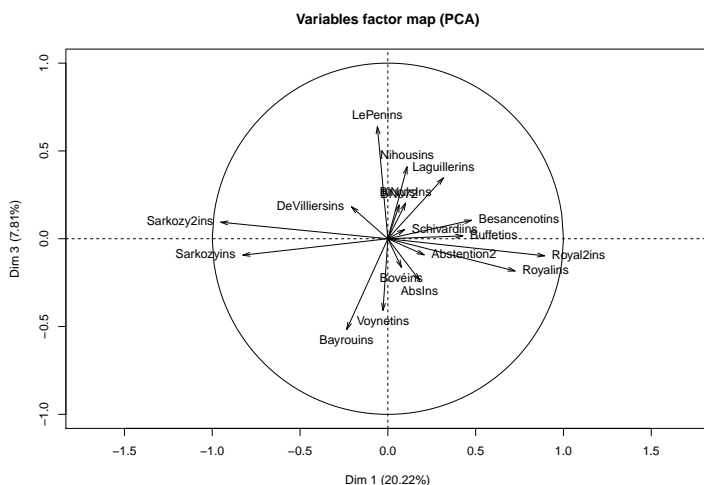
```
plot.PCA(acp1, choix = "var")
```

Mettre en place l'ACP



Mettre en place l'ACP

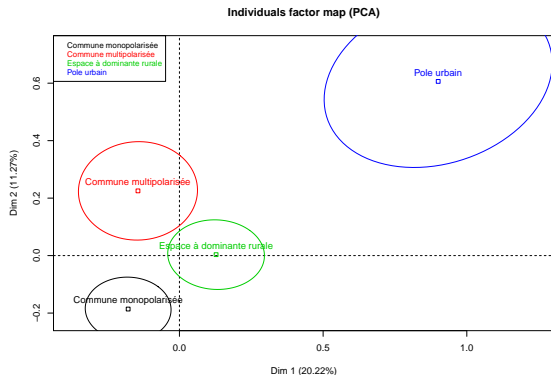
```
plot.PCA(acp1, choix = "var", axes = c(1, 3))
```



Mettre en place l'ACP

On peut aussi dessiner le graphe des individus :

```
concat <- cbind.data.frame(data[-1699, 19], acp1$ind$coord)
ellipse.coord <- coord.ellipse(concat, bary = T)
plot.PCA(acp1, choix = "ind", habillage = 19,
  invisible = c("ind", "ind.sup"), ellipse = ellipse.coord)
```



Avoir des résumés numériques

```
dimdesc(acpl, axes = 1)

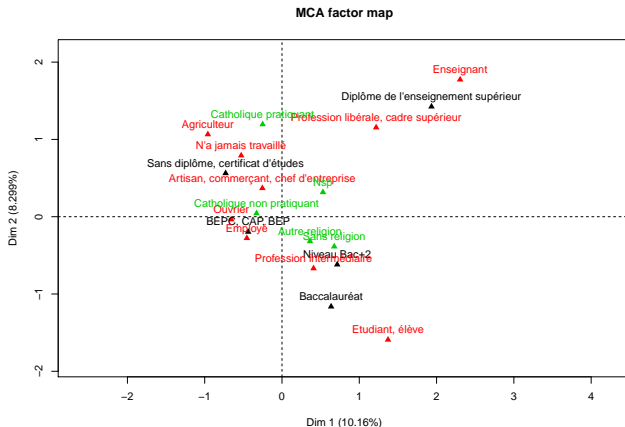
## $Dim.1
## $Dim.1$quanti
##          correlation      p.value
## Royal2ins      0.89271  0.000e+00
## Royalins       0.72464  0.000e+00
## Besancenotins  0.47652  4.282e-130
## Buffetins      0.42699  4.278e-102
## Laguillierins  0.31645  2.092e-54
## Abstention2    0.20695  1.456e-23
## AbsIns         0.18009  3.879e-18
## Nihousins      0.10966  1.445e-07
## BlNulsIns      0.10118  1.231e-06
## Schivardiins   0.09495  5.348e-06
## Bovéins        0.07758  2.031e-04
## BN072          0.06512  1.825e-03
## LePenins       -0.05963  4.319e-03
## DeVilliersins  -0.20723  1.272e-23
## Bayrouins      -0.23464  5.319e-30
## Sarkozyins     -0.82593  0.000e+00
## Sarkozy2ins    -0.95192  0.000e+00
##
## $Dim.1$quali
##          R2      p.value
## type_urbain 0.01631 3.471e-08
##
## $Dim.1$category
##          Estimate      p.value
## Pole urbain      0.7241 3.493e-09
## Commune multipolarisée -0.3206 2.319e-04
## Commune monopolarisée -0.3547 1.134e-06
##
##
```

Plan

- 1 Introduction : l'analyse géométrique des données
- 2 L'analyse en composantes principales
- 3 L'analyse des correspondances multiples**

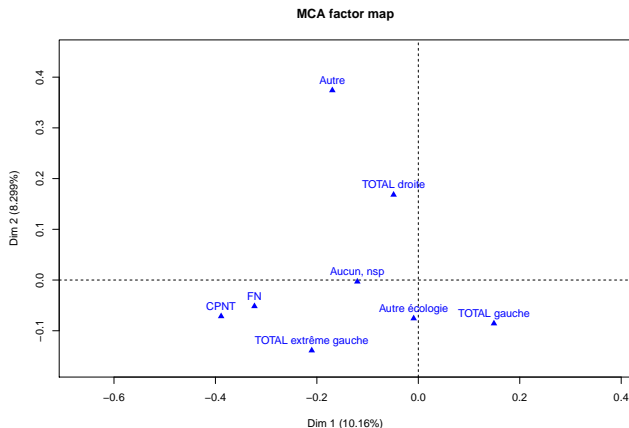
Initier l'ACM

```
load("/media/HDD/Dropbox/Thèse/séminaire R/séance 4/BPFv3.Rdata")
data2 <- BPFv3[, c("RCRS2", "RCRS7", "RRS15",
  "RC40")]
acm1 <- MCA(data2, quali.sup = 4, graph = F, ind.sup = 1699)
plot(acm1, choix = "ind", invisible = c("ind",
  "ind.sup", "quali.sup"), habillage = "quali")
```



Avec la variable illustrative

```
plot(acm1, choix = "ind", invisible = c("ind",  
    "ind.sup", "var"), habillage = "quali")
```



Description des dimensions

```
dimdesc(acml, axes = 1)

## $'Dim 1'
## $'Dim 1'$quali
##           R2      p.value
## RCRS2 0.72137 0.000e+00
## RCRS7 0.70019 0.000e+00
## RRS15 0.20337 9.559e-254
## RC40  0.02491 4.561e-25
##
## $'Dim 1'$category
##                                     Estimate      p.value
## Enseignant                        1.4957 0.000e+00
## Diplôme de l'enseignement supérieur 1.1117 0.000e+00
## Etudiant, élève                   0.8092 1.050e-295
## Profession libérale, cadre supérieur 0.6960 1.802e-264
## Niveau Bac+2                      0.2142 6.199e-45
## Baccalauréat                      0.1560 8.628e-36
## Profession intermédiaire           0.1002 9.396e-13
## Sans religion                      0.3521 4.347e-10
## TOTAL gauche                      0.2128 1.823e-05
## FN                                -0.1347 3.373e-02
## CPNT                               -0.1830 5.747e-03
## Catholique pratiquant              -0.3297 3.757e-08
## Catholique non pratiquant          -0.3880 3.039e-12
## Artisan, commerçant, chef d'entreprise -0.3873 3.217e-66
## N'a jamais travaillé               -0.5889 6.431e-202
## Agriculteur                       -0.9065 4.222e-231
## Ouvrier                           -0.6835 0.000e+00
## Employé                           -0.5350 0.000e+00
## BEPC, CAP, BEP                    -0.6335 0.000e+00
## Sans diplôme, certificat d'études -0.8485 0.000e+00
##
##
```

Et maintenant... Exerçons-nous !