

TD7

Note : ce TD comprend une partie d'exercice sur table corrigée en séance.

La base de données est tirée du package `ade4` et présente les résultats d'une enquête menée auprès de 810 clients d'une banque. Le dictionnaire des données de la base peut être consulté [ici](#).

```
library(readr)
library(FactoMineR)
library(factorextra)
library(explor)
library(GDAtools)

setwd("/Users/mathieuferry/Documents/Enseignements/Année 2023-2024/L3 Sociologie Analyse d
enq<-read_csv("TD7.csv")
```

Pour réaliser l'ACM telle qu'indiquée sur l'énoncé, il faut sélectionner les variables présentes dans l'analyse.

```
enq1<-subset(enq,select=c(csp,age,sexe,
                        assurvi,cableue,eparliv,interdit,duree,credcon,soldevu,oppo,cred

str(enq1)
```

```
tibble [810 x 12] (S3: tbl_df/tbl/data.frame)
 $ csp      : chr [1:810] "ouvri" "cadsu" "cadsu" "inact" ...
 $ age      : chr [1:810] "ai75" "ai35" "ai75" "ai45" ...
 $ sexe     : chr [1:810] "hom" "hom" "hom" "fem" ...
 $ assurvi  : chr [1:810] "non" "non" "non" "non" ...
 $ cableue  : chr [1:810] "non" "non" "non" "non" ...
 $ eparliv  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ interdit: chr [1:810] "non" "non" "non" "oui" ...
 $ duree    : chr [1:810] "d48" "d24" "d48" "d24" ...
```

```
$ credcon : chr [1:810] "nul" "nul" "nul" "nul" ...
$ soldevu : chr [1:810] "n1" "n1" "p1" "p2" ...
$ oppo    : chr [1:810] "non" "non" "non" "non" ...
$ credhab : chr [1:810] "non" "non" "non" "non" ...
```

Il va falloir indiquer à la fonction MCA le numéro de colonne des variables supplémentaires. Ici, on a créé enq1 de telle sorte qu'elles soient dans les trois premières colonnes, donc on utilise l'argument quali.sup de la fonction MCA, où on indique le vecteur c(1,2,3).

```
res.acm1<-MCA(enq1,quali.sup=c(1,2,3),graph=F)
```

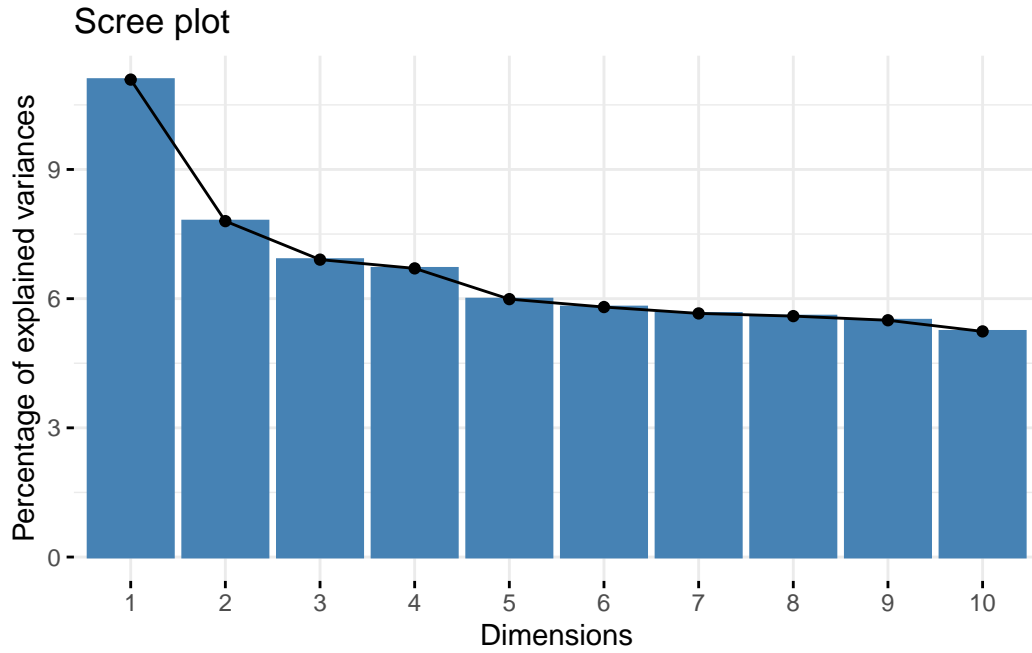
On peut vérifier nos résultats sur les valeurs propres que nous montrions tout à l'heure. Concernant le critère de Kaiser :

```
res.acm1$eig[,1]>mean(res.acm1$eig[,1])
```

```
dim 1  dim 2  dim 3  dim 4  dim 5  dim 6  dim 7  dim 8  dim 9  dim 10  dim 11
TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   FALSE  FALSE  FALSE
dim 12  dim 13  dim 14  dim 15  dim 16  dim 17  dim 18
FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
```

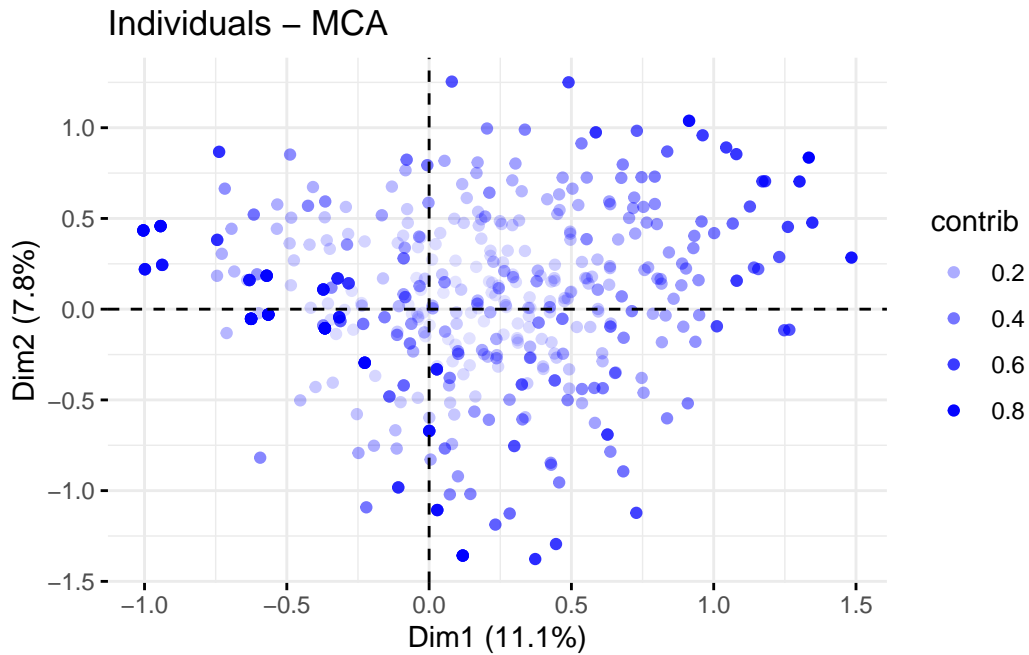
Concernant le critère du coude, on projette l'éboullis des valeurs propres.

```
fviz_eig(res.acm1)
```



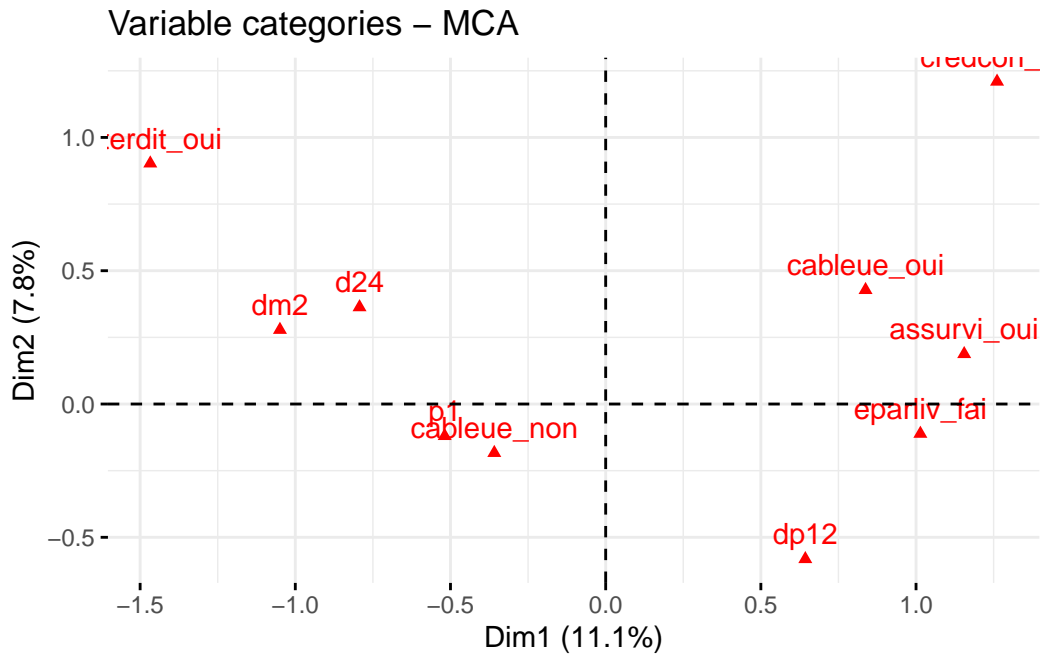
Examinons le nuage des individus sur le premier plan. Le nuage ne semble pas présenter de forme particulière sur le premier plan de l'ACM.

```
fviz_mca_ind(res.acm1,alpha.ind = "contrib",label="none")
```



On regarde la projection des modalités actives qui ont une contribution supérieure à la contribution moyenne sur l'axe 1. On a un axe structuré par une opposition entre instabilité et solvabilité bancaire.

```
contrib1<-rownames(res.acm1$var$contrib)[res.acm1$var$contrib[,1]>100/nrow(res.acm1$var$co
fviz_mca_var(res.acm1,axes=c(1,2),
select.var=list(name = contrib1))
```

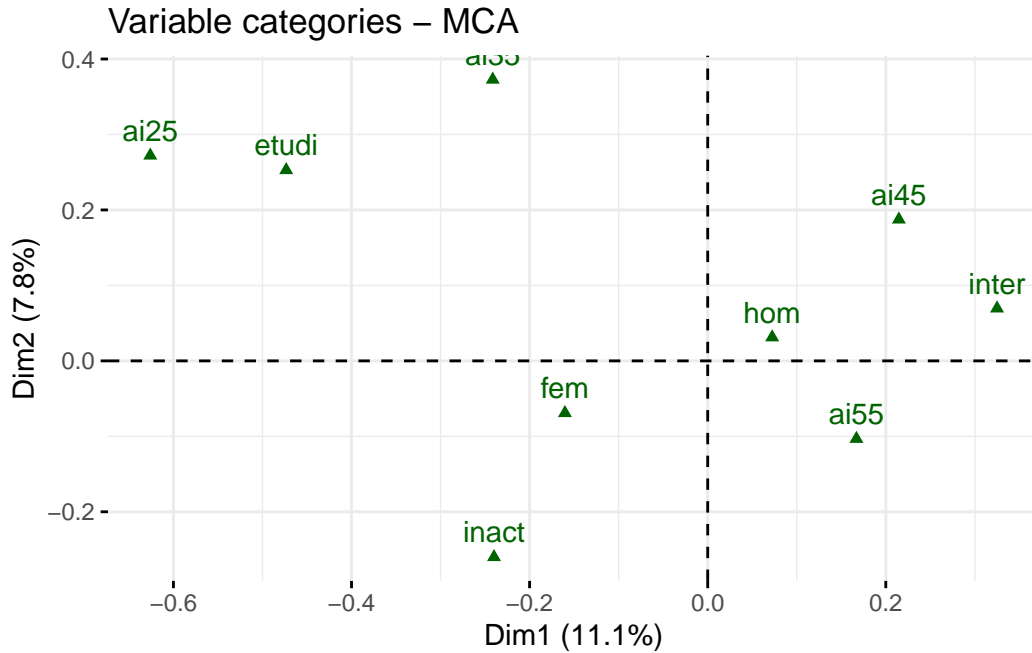


Les modalités supplémentaires significativement associées à l'axe peuvent être projetées de la manière suivante :

```

sigsup1<-rownames(res.acm1$quali.sup$v.test)[abs(res.acm1$quali.sup$v.test[,1])>1.96]
fviz_mca_var(res.acm1,axes=c(1,2),
             select.var=list(name = sigsup1))

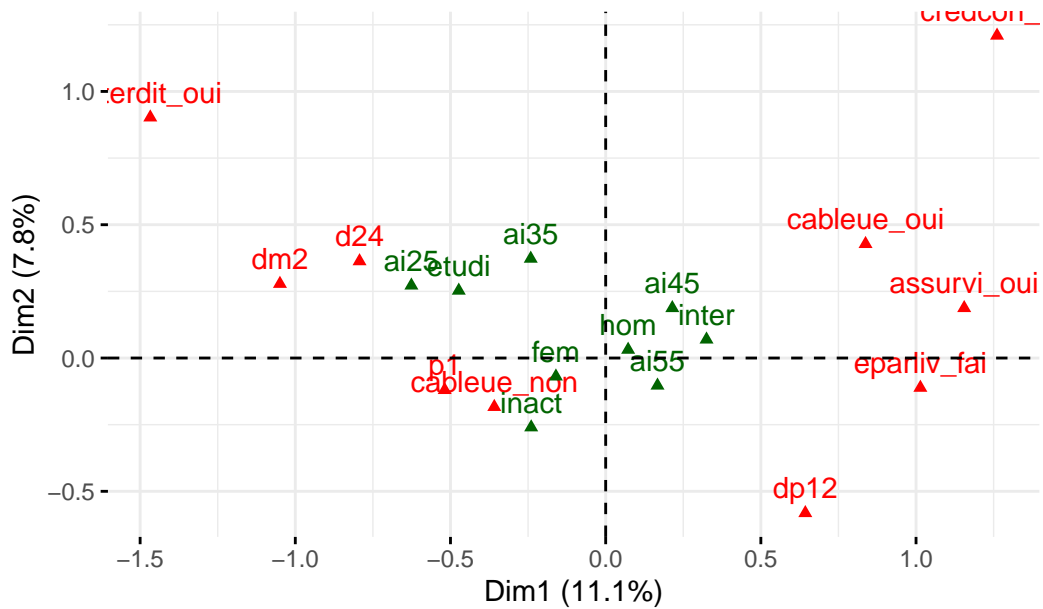
```



Il peut être judicieux de représenter les variables actives et supplémentaires sur le même plan. A noter que les variables actives sont beaucoup plus dispersées que les variables supplémentaires, ainsi on n’explique pas l’entièreté de la variabilité des pratiques bancaires avec les variables sociodémographiques prises en compte ici.

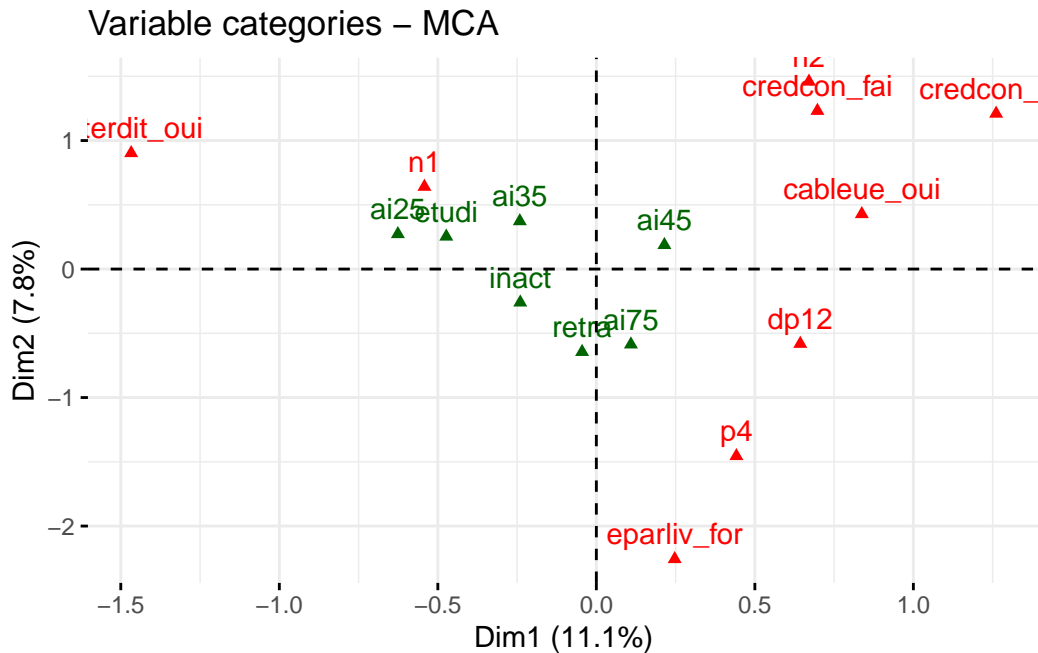
```
fviz_mca_var(res.acm1, axes=c(1,2),
             select.var=list(name = c(sigsup1, contrib1)))
```

Variable categories – MCA



Faisons la même chose pour l'axe 2.

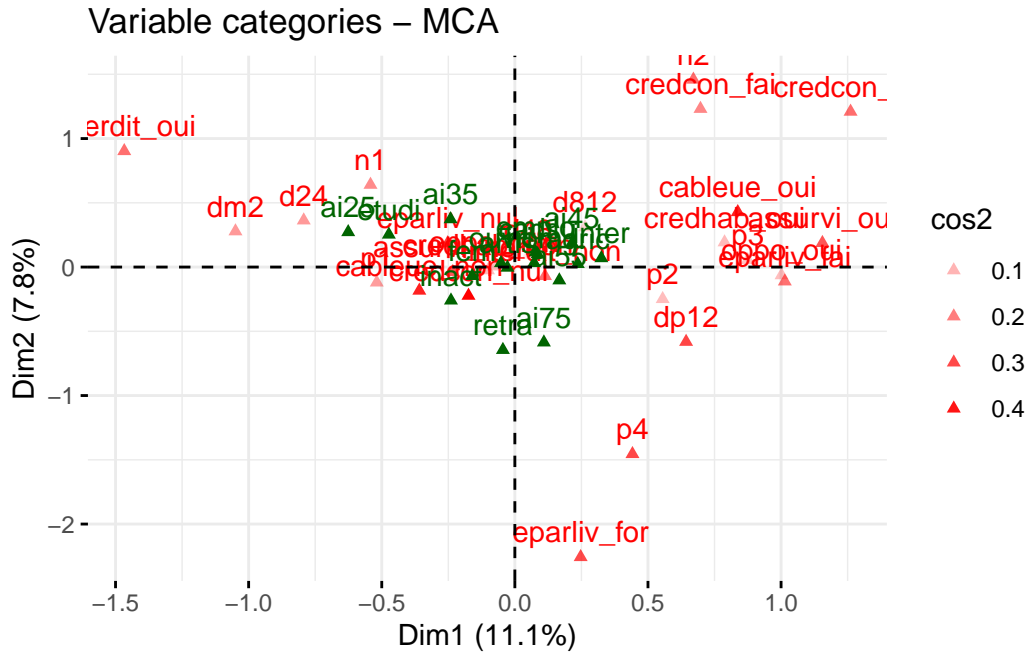
```
contrib2<-rownames(res.acm1$var$contrib)[res.acm1$var$contrib[,2]>100/nrow(res.acm1$var$contrib)]
sigsup2<-rownames(res.acm1$quali.sup$v.test)[abs(res.acm1$quali.sup$v.test[,2])>1.96]
fviz_mca_var(res.acm1,axes=c(1,2),
             select.var=list(name = c(sigsup2,contrib2)))
```



Cet axe oppose des *emprunteurs* (en haut) qui ne sont pas solvables et sont aussi plutôt interdits bancaires, à des *épargnants* avec beaucoup de réserve financière (et une grande ancienneté dans la banque) en bas. L'axe est surtout structuré par l'âge : on emprunte en début de cycle de vie, tandis qu'on est plus épargnants en fin de cycle de vie.

On projette finalement l'ensemble du nuage des variables, en distinguant l'intensité des modalités suivant la qualité de représentation des modalités.

```
fviz_mca_var(res.acm1, axes=c(1,2), alpha.var = "cos2" )
```

Examinons le taux modifié des premiers axes de l'ACM qui permet de mieux voir l'importance relative du premier axe. Il faut utiliser la fonction `modif.rate` du package `GDAtools`. On voit que pour le premier axe il correspond à 78% de l'information, contre 13% pour le second axe. Le premier axe structure donc très fortement les données.

```
mrate<-modif.rate(res.acm1)
mrate$modif
```

	mrate	cum.mrate
dim 1	78.376274789	78.37627
dim 2	12.906256376	91.28253
dim 3	4.672854554	95.95539
dim 4	3.373297051	99.32868
dim 5	0.481427051	99.81011
dim 6	0.159532959	99.96964
dim 7	0.026445496	99.99609
dim 8	0.003911724	100.00000

Appuyons-nous maintenant sur toutes les variables du fichier `enq` pour réaliser une seconde ACM (21 variables !).

str(enq)

```
spc_tbl_ [810 x 21] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ csp      : chr [1:810] "ouvri" "cadsu" "cadsu" "inact" ...
 $ duree    : chr [1:810] "d48" "d24" "d48" "d24" ...
 $ oppo     : chr [1:810] "non" "non" "non" "non" ...
 $ age      : chr [1:810] "ai75" "ai35" "ai75" "ai45" ...
 $ sexe     : chr [1:810] "hom" "hom" "hom" "fem" ...
 $ interdit: chr [1:810] "non" "non" "non" "oui" ...
 $ cableue  : chr [1:810] "non" "non" "non" "non" ...
 $ assurvi  : chr [1:810] "non" "non" "non" "non" ...
 $ soldevu  : chr [1:810] "n1" "n1" "p1" "p2" ...
 $ eparlog  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ eparliv  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ credhab  : chr [1:810] "non" "non" "non" "non" ...
 $ credcon  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ versesp  : chr [1:810] "non" "non" "non" "oui" ...
 $ retresp  : chr [1:810] "fai" "fai" "fai" "fai" ...
 $ remiche  : chr [1:810] "nul" "nul" "nul" "fai" ...
 $ preltre  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ prelfm  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ viredeb  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ virecre  : chr [1:810] "nul" "nul" "nul" "nul" ...
 $ porttit  : chr [1:810] "nul" "nul" "nul" "moy" ...
- attr(*, "spec")=
 .. cols(
 ..   csp = col_character(),
 ..   duree = col_character(),
 ..   oppo = col_character(),
 ..   age = col_character(),
 ..   sexe = col_character(),
 ..   interdit = col_character(),
 ..   cableue = col_character(),
 ..   assurvi = col_character(),
 ..   soldevu = col_character(),
 ..   eparlog = col_character(),
 ..   eparliv = col_character(),
 ..   credhab = col_character(),
 ..   credcon = col_character(),
 ..   versesp = col_character(),
 ..   retresp = col_character(),
 ..   remiche = col_character(),
```

```

.. preltre = col_character(),
.. prelfm = col_character(),
.. viredeb = col_character(),
.. virecre = col_character(),
.. porttit = col_character()
.. )
- attr(*, "problems")=<externalptr>

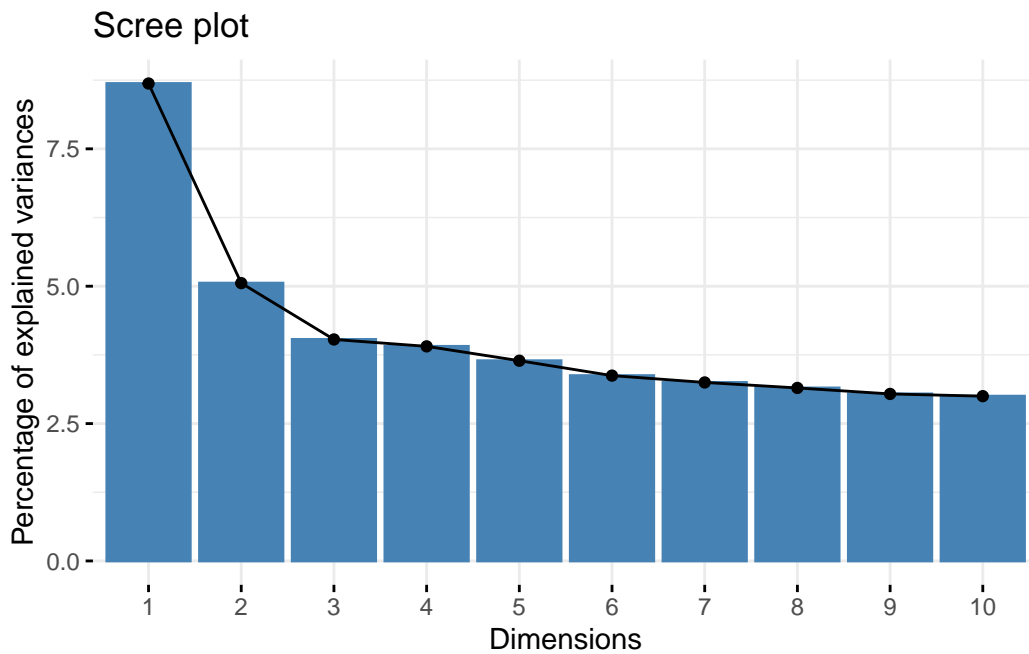
```

Cette fois, les variables de position sociale que nous souhaitons positionner en supplémentaire sont en position 1, 4, 5 du fichier enq.

```
res.acm2<-MCA(enq,quali.sup=c(1,4,5),graph=F)
```

Regardons de nouveau combien d'axes analyser. D'après le critère du coude, on a plus nettement deux axes à garder ici.

```
fviz_eig(res.acm2)
```



```
res.acm2$eig[,1]>mean(res.acm2$eig[,1])
```

```
dim 1 dim 2 dim 3 dim 4 dim 5 dim 6 dim 7 dim 8 dim 9 dim 10 dim 11
```

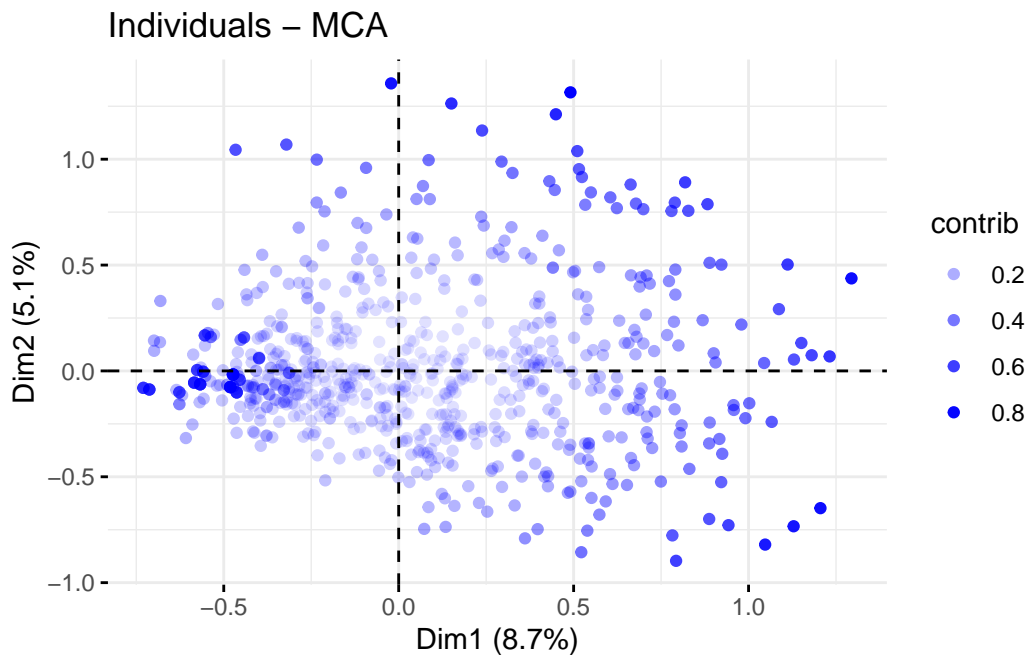
```

TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
dim 12 dim 13 dim 14 dim 15 dim 16 dim 17 dim 18 dim 19 dim 20 dim 21 dim 22
TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE
dim 23 dim 24 dim 25 dim 26 dim 27 dim 28 dim 29 dim 30 dim 31 dim 32 dim 33
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
dim 34 dim 35 dim 36 dim 37 dim 38 dim 39
FALSE FALSE FALSE FALSE FALSE FALSE

```

Le nuage des individus ne semble pas non plus présenter de structure particulière. Si on pinaillait, on irait regarder la contribution des individus à l'ACM car on a l'impression que certains d'entre eux contribuent bien davantage.

```
fviz_mca_ind(res.acm2,alpha.ind = "contrib",label="none")
```



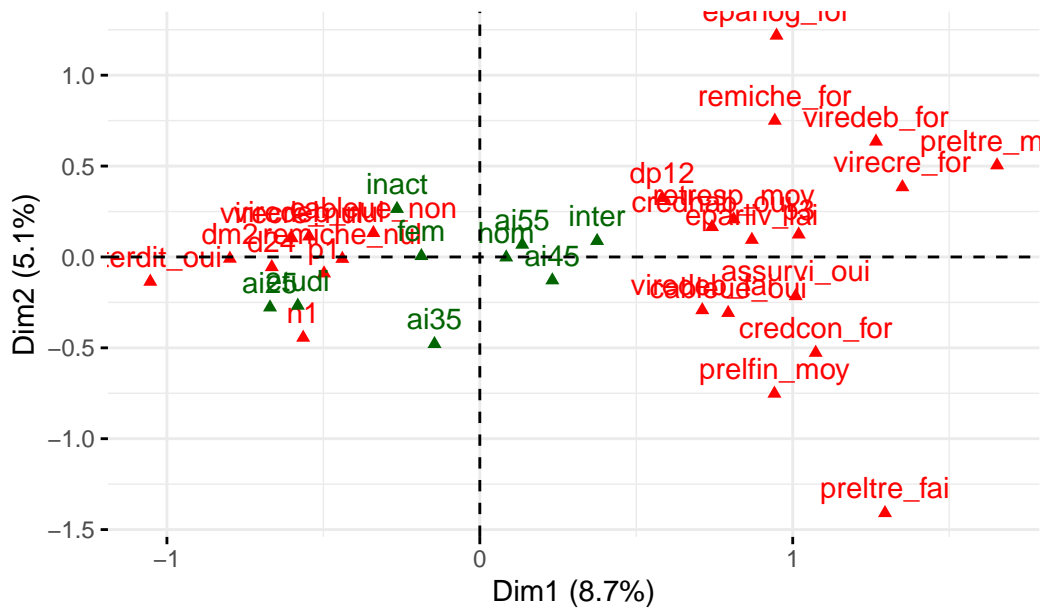
Le premier axe, interprétation à discuter :

```

contrib1<-rownames(res.acm2$var$contrib)[res.acm2$var$contrib[,1]>100/nrow(res.acm2$var$co
sigsup1<-rownames(res.acm2$quali.sup$v.test)[abs(res.acm2$quali.sup$v.test[,1])>1.96]
fviz_mca_var(res.acm2,axes=c(1,2),
             select.var=list(name = c(sigsup1,contrib1)))

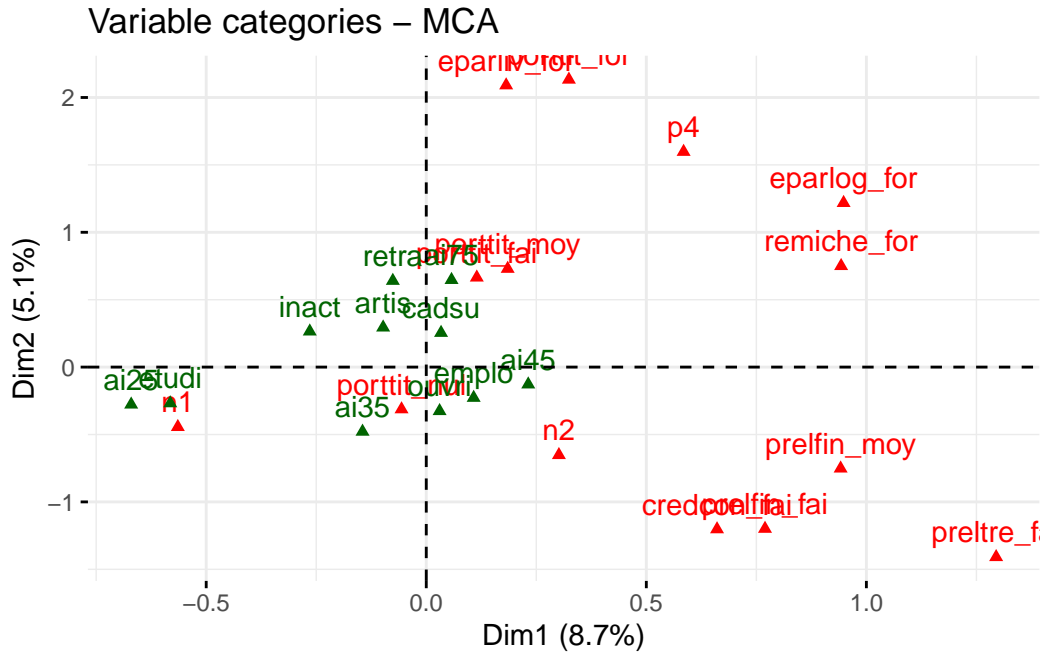
```

Variable categories – MCA



Le second axe, interprétation à discuter :

```
contrib2<-rownames(res.acm2$var$contrib)[res.acm2$var$contrib[,2]>100/nrow(res.acm2$var$co
sigsup2<-rownames(res.acm2$quali.sup$v.test)[abs(res.acm2$quali.sup$v.test[,2])>1.96]
fviz_mca_var(res.acm2,axes=c(1,2),
             select.var=list(name = c(sigsup2,contrib2)))
```



L'ensemble des modalités sur le premier plan factoriel, en distinguant l'opacité suivant la qualité de représentation (cos2) :

```
fviz_mca_var(res.acm2, axes=c(1,2), alpha="cos2")
```

Variable categories – MCA

