

Correction du devoir 5 – script R

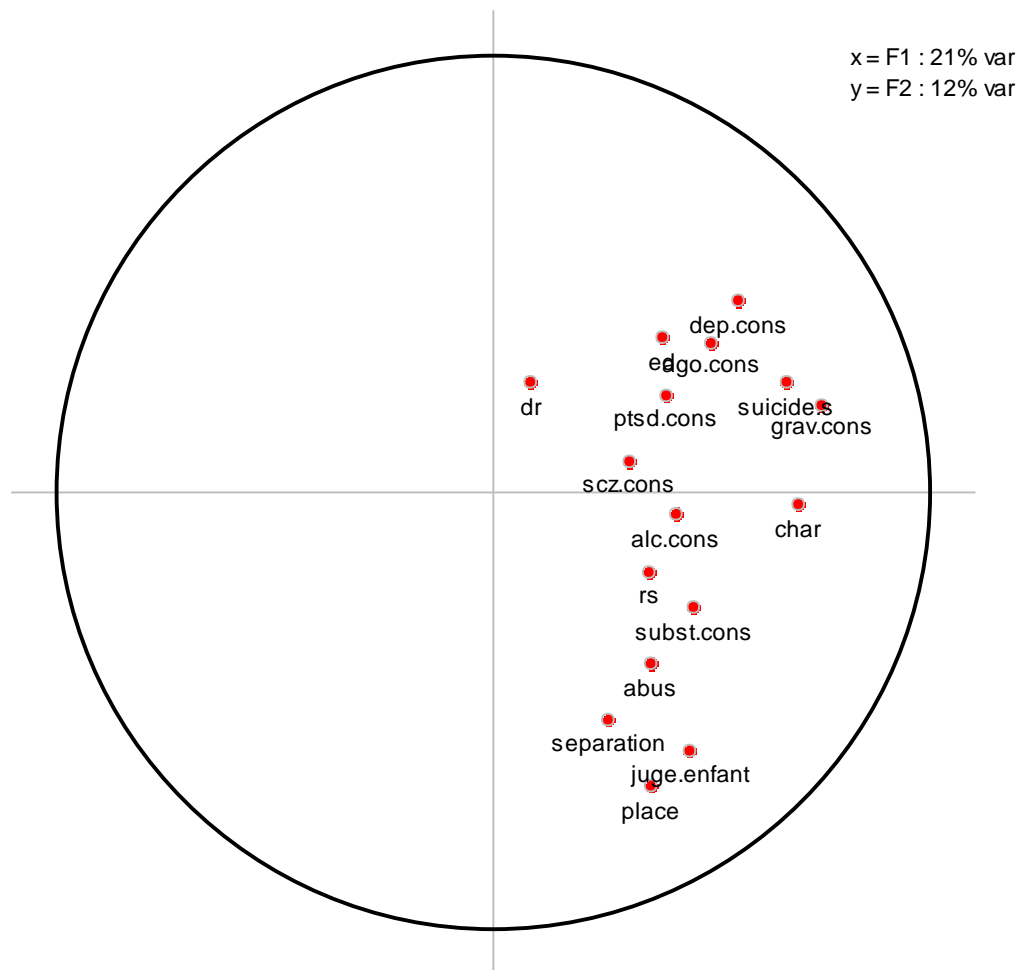
Représentez graphiquement, de façon exploratoire, les relations entre les variables quantitatives et binaires suivantes du fichier smp2 (étude Santé Mentale en Prison)

“separation” (séparation du milieu parental),
“juge.enfant” (contact avec un juge pour enfant),
“place” (placement dans l’enfance),
“abus” (antécédents d’abus dans l’enfance),
“grav.cons” (score de gravité global),
“dep.cons” (dépression),
“ago.cons” (agoraphobie),
“ptsd.cons” (syndrome de stress post traumatique”,
“alc.cons” (alcoolisme),
“subst.cons” (abus/dépendance aux substances),
“scz.cons” (schizophrénie),
“char” (score de trouble de la personnalité),
“rs” (recherche de sensation),
“ed” (éviterement du danger”,
“dr” (dépendance à la récompense),
“suicide.s” (score de risque suicidaire)

Questions

1) Faites une analyse en composantes principales. Commentez le résultat.

```
smp <- read.csv2("D:/MOOC/Data/smp2.csv")  
  
noms <-  
c("separation", "juge.enfant", "place", "abus", "grav.cons",  
  "dep.cons", "ago.cons", "ptsd.cons", "alc.cons", "subst.cons",  
  "scz.cons", "char", "rs", "ed", "dr", "suicide.s")  
  
library(psy)  
  
mdspca(smp[,noms])
```

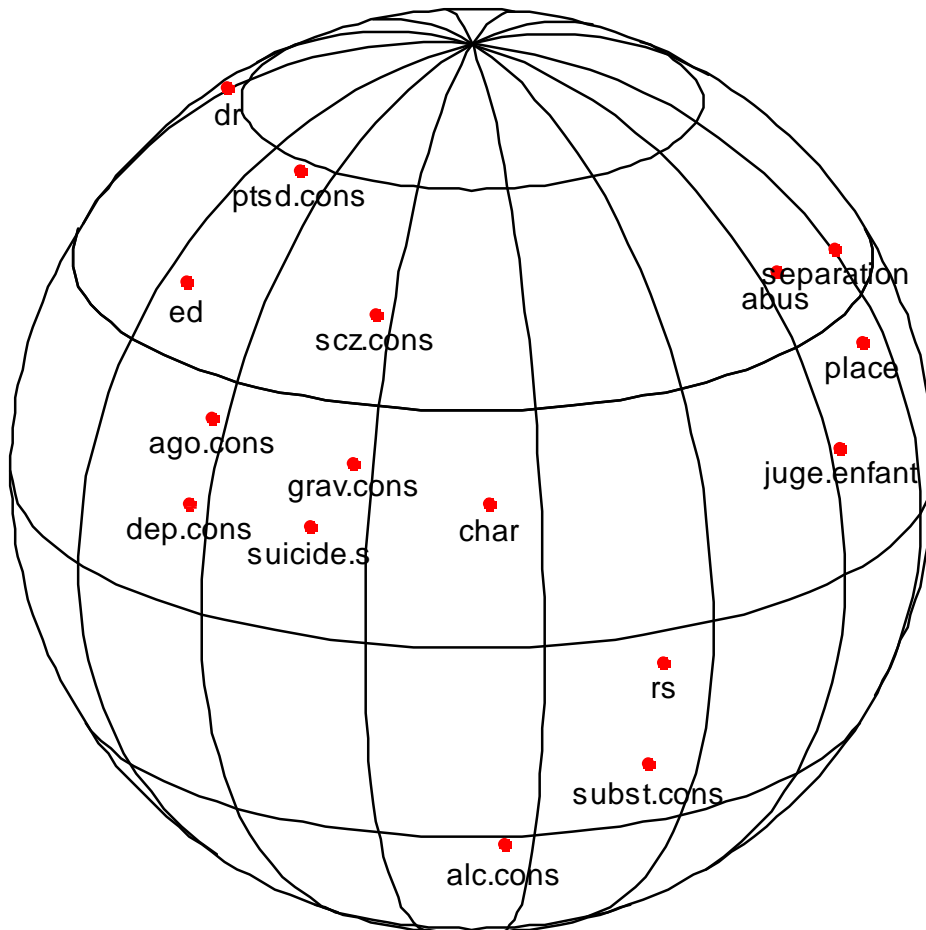


Interprétation :

- Globalement la plus grande prudence s'impose dans cette interprétation, les points variables étant trop proche du centre du cercle. C'est en particulier vrai « dr » et scz.cons ».
- On constate cependant dans le cadran inférieur droit 4 variables en liens avec une enfance que l'on pourrait qualifier de « difficile » (« abus », « separation », « juge.enfant », « place »)
- Dans le cadran supérieur droit, il y a un groupe de variables en lien avec la pathologie anxio/dépressive, le risque suicidaire et la gravité globale du détenu (ainsi que l'évitement du danger « ed » qui est également souvent trouvé élevé chez des sujets anxio/dépressifs)
- Entre ces deux groupes nous trouvons des variables en lien avec l'abus de substance et d'alcool, associées à la recherche de sensation « rs » (ce qui est classique également).

2) Comparez avec les résultats obtenus à l'aide d'une représentation sphérique.

```
sphpca (smp[,noms])
```

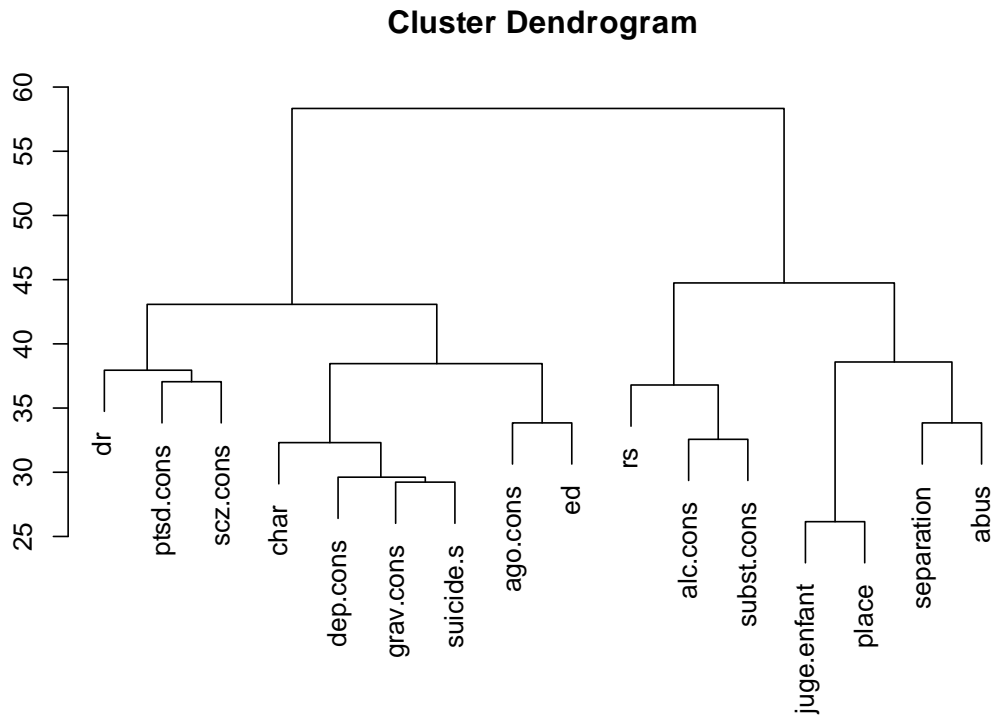


Interprétation

- Elle est d'autant plus délicate que nous ne pouvons pas savoir ici la qualité de représentation de chacun des points variables
- On retrouve cependant, en plus clair, la même segmentation en trois groupes que dans l'ACP (enfance difficile / abus d'alcool et de substance / pathologie anxio-dépressive)

3) Comparez avec les résultats obtenus à l'aide d'une classification hiérarchique.

```
cha <- hclust(dist(t(scale(smp[,noms]))),method="ward")  
plot(cha, xlab="", ylab="")
```



Interprétation

- Ici la question de la qualité de représentation de chaque point variable ne se pose pas, mais nous perdons en facilité de lecture (la représentation est moins synthétique). Par ailleurs, des variables éloignées peuvent être corrélées négativement ou non corrélées, sans qu'on puisse le différencier.
- Il y a en première lecture deux groupes de variables : pathologie anxio-dépressive d'un côté, enfance difficile et abus de substance et d'alcool de l'autre.
- Ce dernier groupe se segmente à son tour dans un second temps (enfance difficile d'un côté, abus de substance et d'alcool de l'autre). Nous retrouvons alors la segmentation de l'ACP et de la représentation sphérique.
- On notera par ailleurs dans le cluster « anxio-dépressif » un sous cluster associé à la dépression au suicide et à la gravité globale, et un autre sous cluster associé à l'agoraphobie et à l'évitement du danger. Cette segmentation plus fine était juste ébauchée sur la représentation sphérique.