min_projet

EL_Hadrami

29/12/2020

```
library("FactoMineR")
library("factoextra")
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library("corrplot")
```

```
## corrplot 0.84 loaded
```

```
library("fpc")
```

```
etudiants <- read.csv("data/etudiants.csv",header = TRUE,sep = ";")
filiere <- etudiants[,1]
etudiants <- as.matrix(etudiants[2:13])
rownames(etudiants) <- filiere
etudiants.active <- as.data.frame(etudiants[,1:6])
head(etudiants.active,4)
```

```
##                                     Licence.F Licence.H Master.F Master.H
## Droit, sciences politiques              69373     37317    42371    21693
## Sciences economiques, gestion           38387     37157    29466    26929
## Administration economique et sociale    18574     12388     4183     2884
## Lettres, sciences du langage, arts      48691     17850    17672     5853
##                                     Doctorat.F Doctorat.H
## Droit, sciences politiques                4029       4342
## Sciences economiques, gestion             1983       2552
## Administration economique et sociale         0          0
## Lettres, sciences du langage, arts        4531       2401
```

```
summary(etudiants.active)
```

```
##    Licence.F       Licence.H        Master.F        Master.H
##  Min.   : 1779   Min.   :  726   Min.   : 1963   Min.   :  811
##  1st Qu.:19570   1st Qu.:15566   1st Qu.: 5910   1st Qu.: 3948
##  Median :31352   Median :19570   Median :15132   Median : 7155
##  Mean   :38901   Mean   :25490   Mean   :18238   Mean   :14341
##  3rd Qu.:59225   3rd Qu.:37277   3rd Qu.:26518   3rd Qu.:21382
##  Max.   :94346   Max.   :54861   Max.   :43016   Max.   :48293
##    Doctorat.F       Doctorat.H
##  Min.   :   0.0   Min.   :    0.0
##  1st Qu.: 600.8   1st Qu.:  472.8
##  Median :3006.0   Median : 2476.5
##  Mean   :3041.8   Mean   : 3424.0
```

```
##  3rd Qu.:4500.0   3rd Qu.: 5009.5
##  Max.   :7787.0   Max.   :11491.0
```

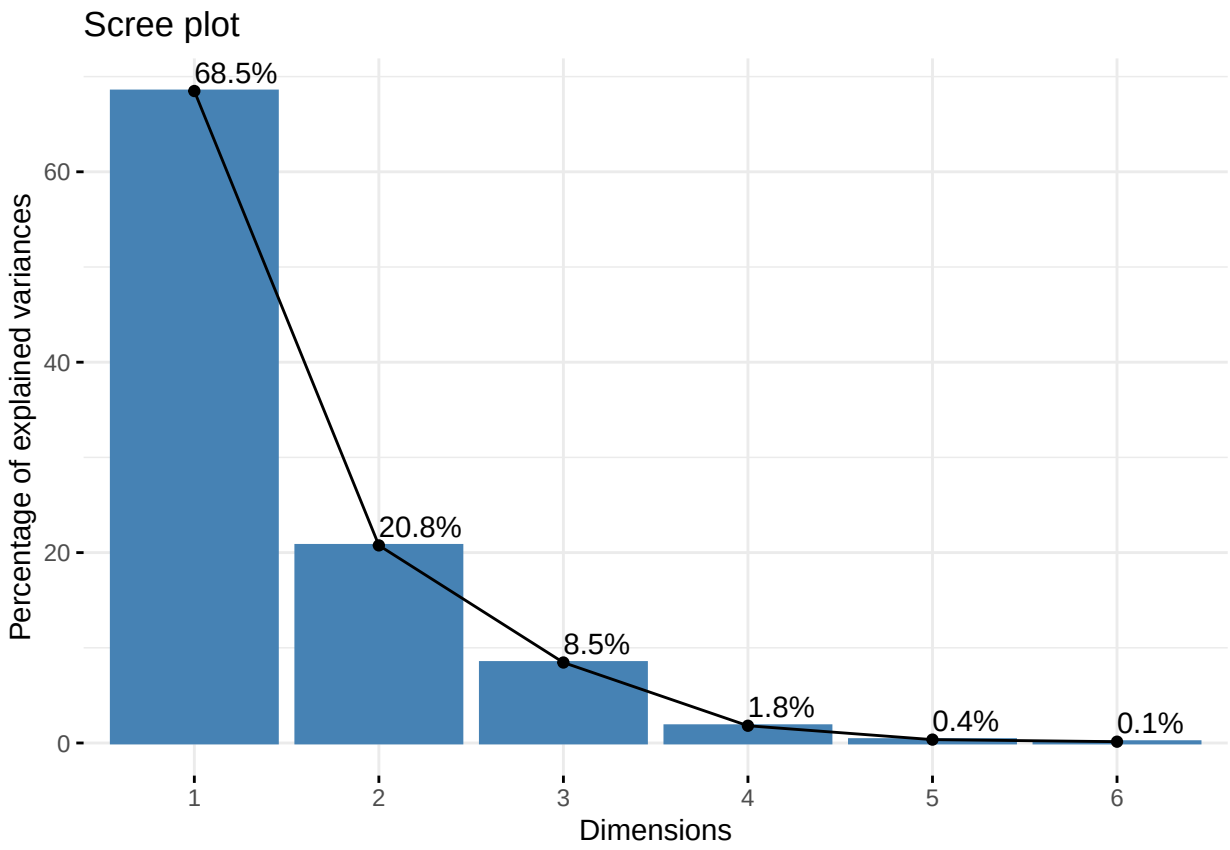
Realisation d'une ACP

```
#etudiants.active <- etudiants[,2:12]
res.acp <- PCA(etudiants.active,scale.unit = TRUE,graph = FALSE)
res.acp$eig
```

```
##         eigenvalue percentage of variance cumulative percentage of variance
## comp 1 4.109124790             68.4854132                           68.48541
## comp 2 1.245412990             20.7568832                           89.24230
## comp 3 0.507217178              8.4536196                           97.69592
## comp 4 0.108686680              1.8114447                           99.50736
## comp 5 0.021274439              0.3545740                           99.86193
## comp 6 0.008283924              0.1380654                          100.00000
```

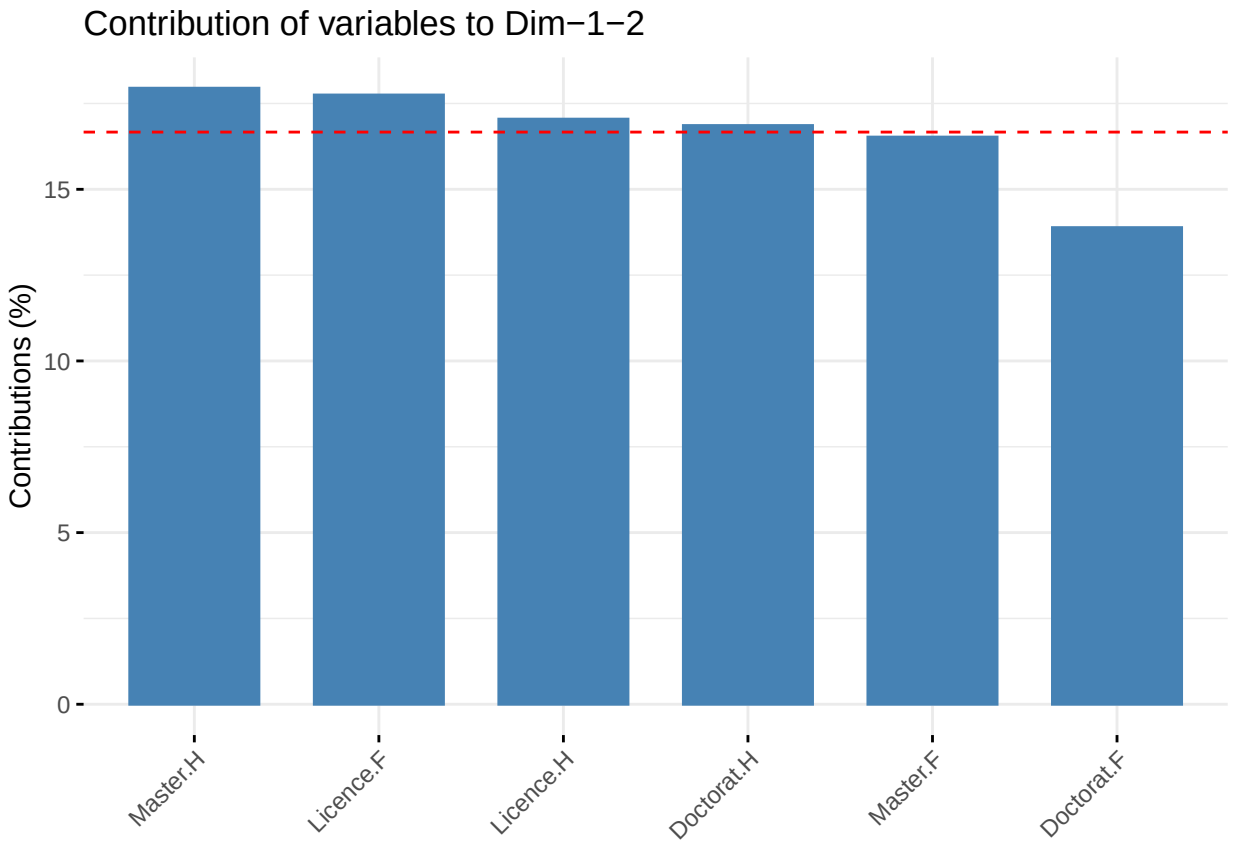
Graphe des valeurs propres

```
fviz_eig(res.acp, addlabels = TRUE)
```



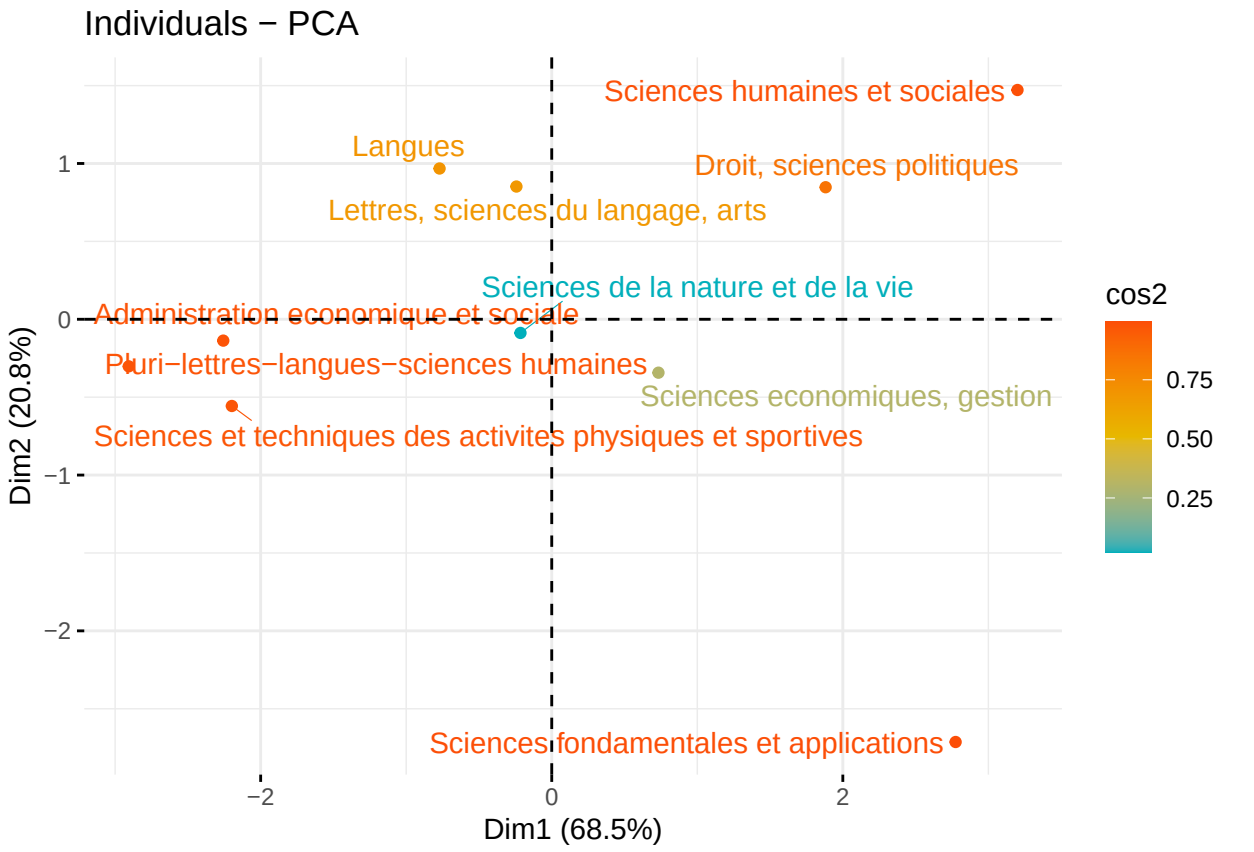
Contributions des variables

```
fviz_contrib(res.acp, choice = "var", axes = 1 :2)
```

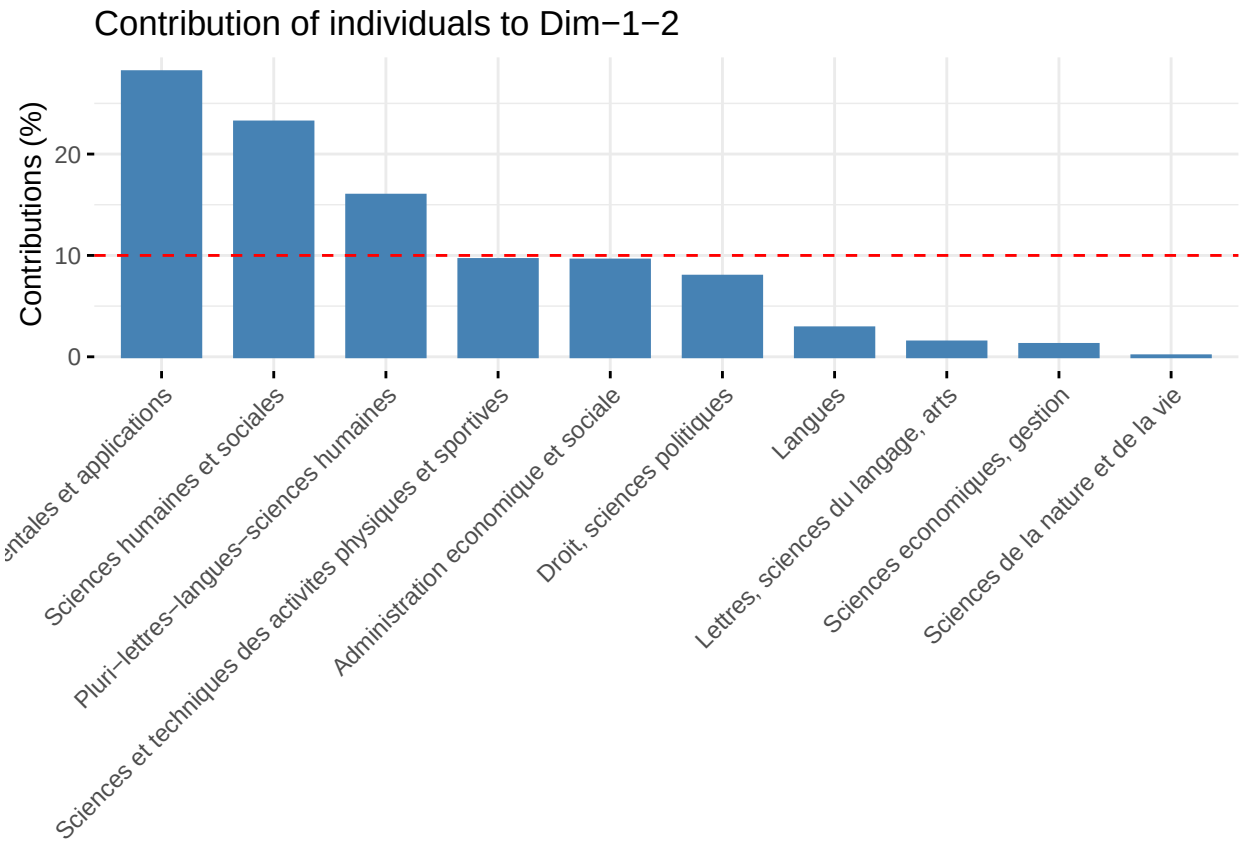


Graphiques des individus

```
ind <- get_pca_ind(res.acp)
fviz_pca_ind (res.acp, col.ind = "cos2",gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),repel = TRUE)
```

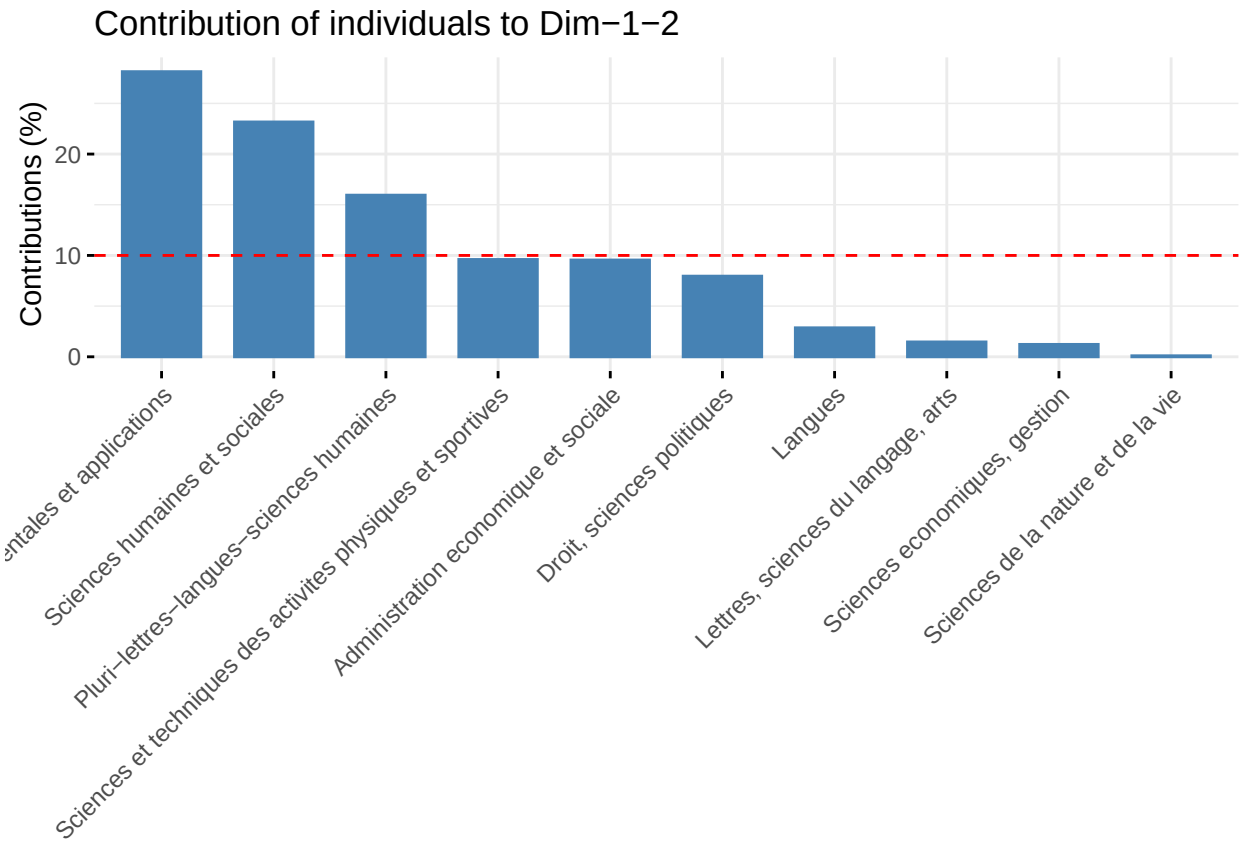



```
fviz_contrib(res.acp, choice = "ind", axes = 1 :2)
```



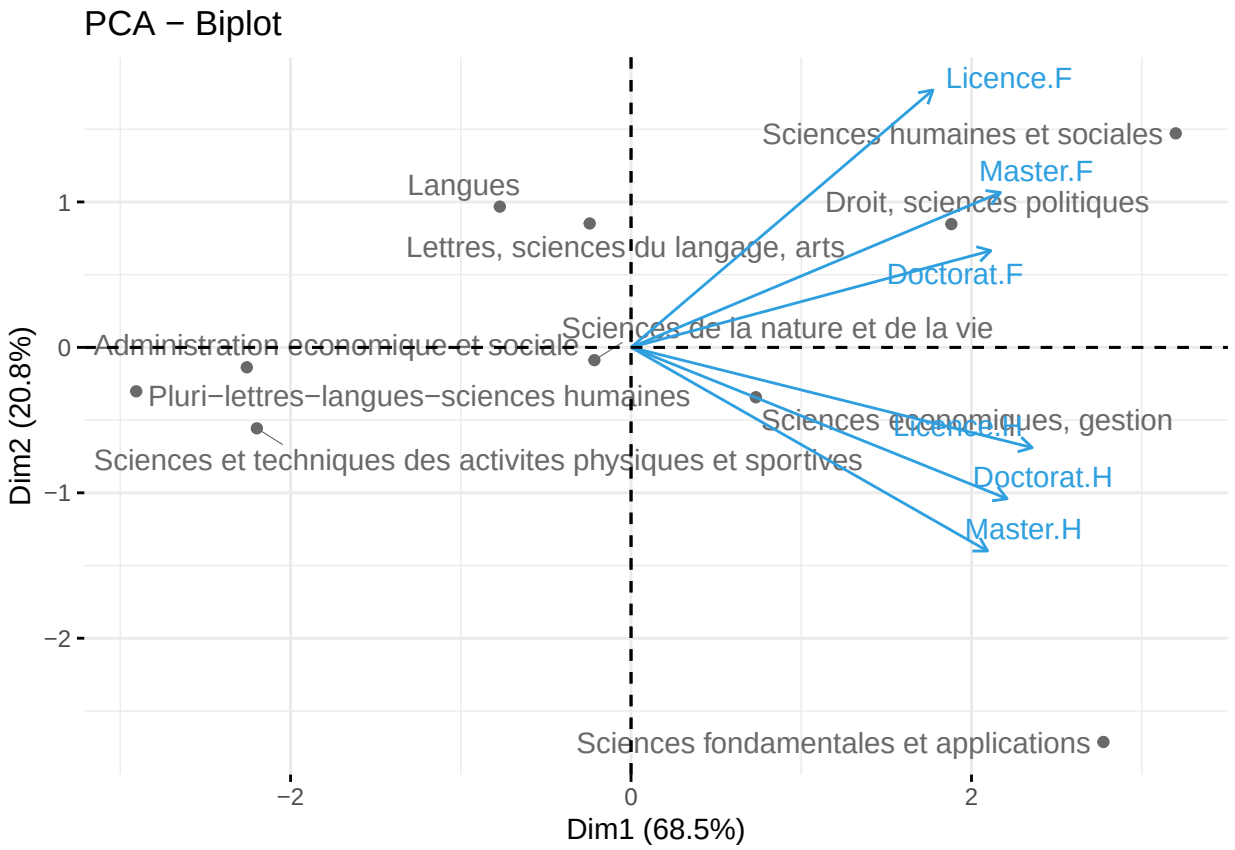
Graphiques des variables

```
fviz_contrib(res.acp, choice = "ind", axes = 1 :2)
```



Graphiques des individus et des variables

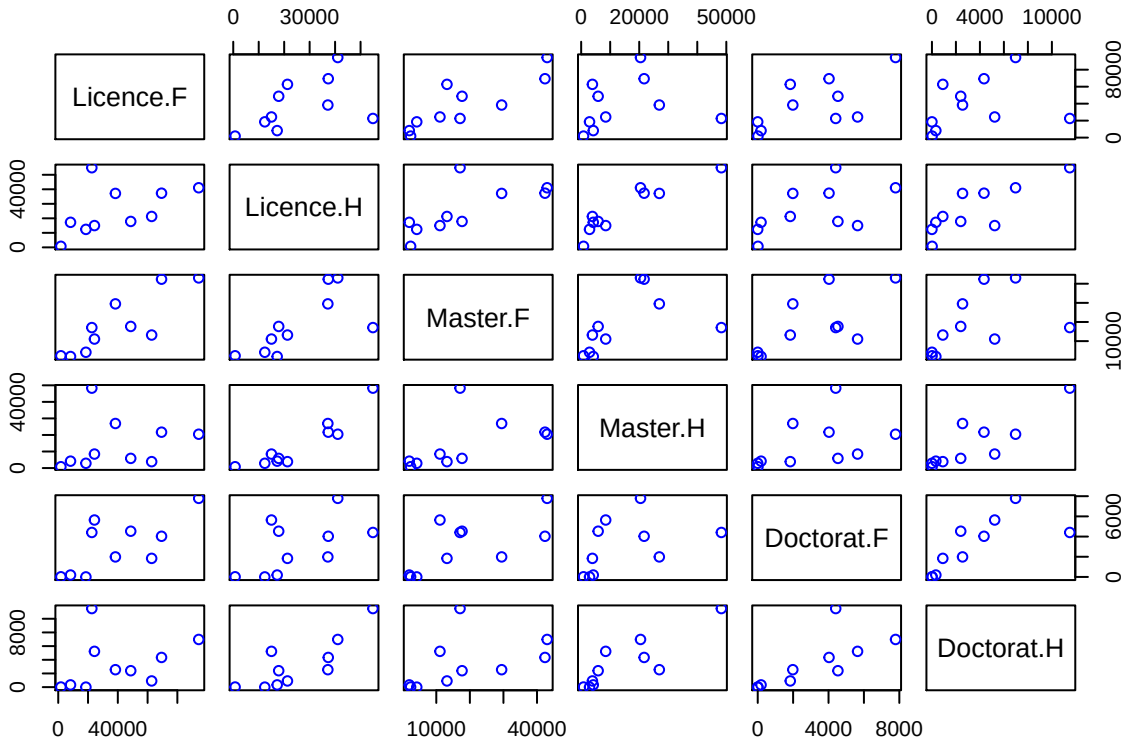
```
fviz_pca_biplot(res.acp,
repel = TRUE,col.var = "#2E9FDF", # Couleur des variables
col.ind = "#696969")
```



Classification hiérarchique ascendante des données

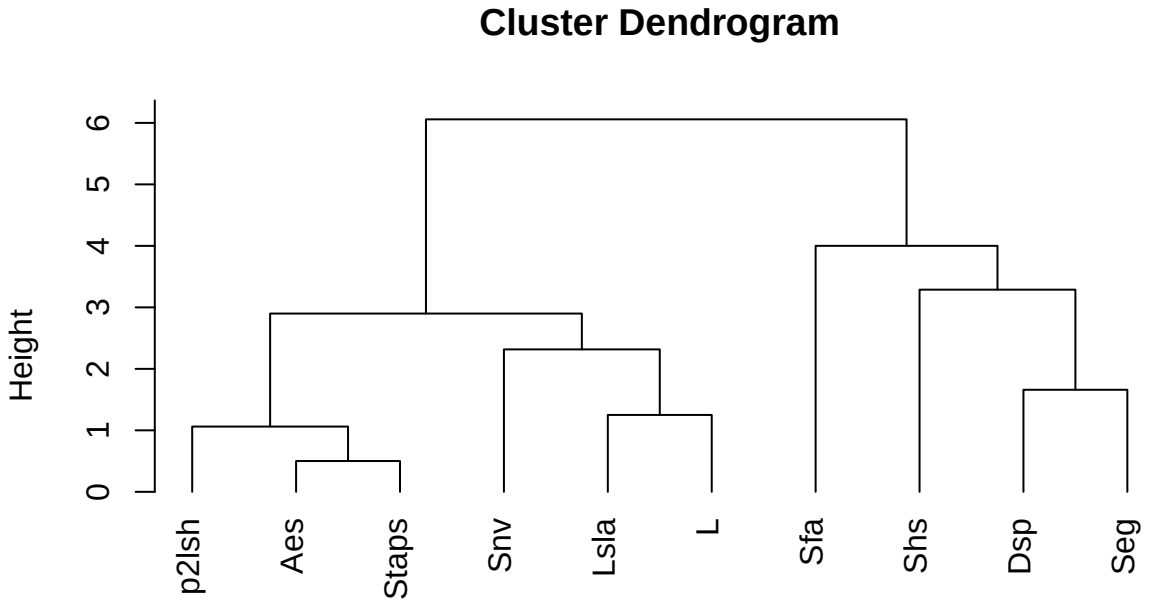
- Realisation d'un plot pour tous les individus

```
plot(etudiants.active,col = "blue")
```



Utilisation de la fonction hclust pour la classification

```
filiere <- c("Dsp","Seg","Aes","Lsla","L","Shs","p2lsh","Sfa","Snv","Staps")
rownames(etudiants.active) <- filiere
etudiants.cr <- scale(etudiants.active,center=T,scale=T)
d.etudiants <- dist(etudiants.cr)
tree <- hclust(d.etudiants)
plot(tree,hang = -1)
```

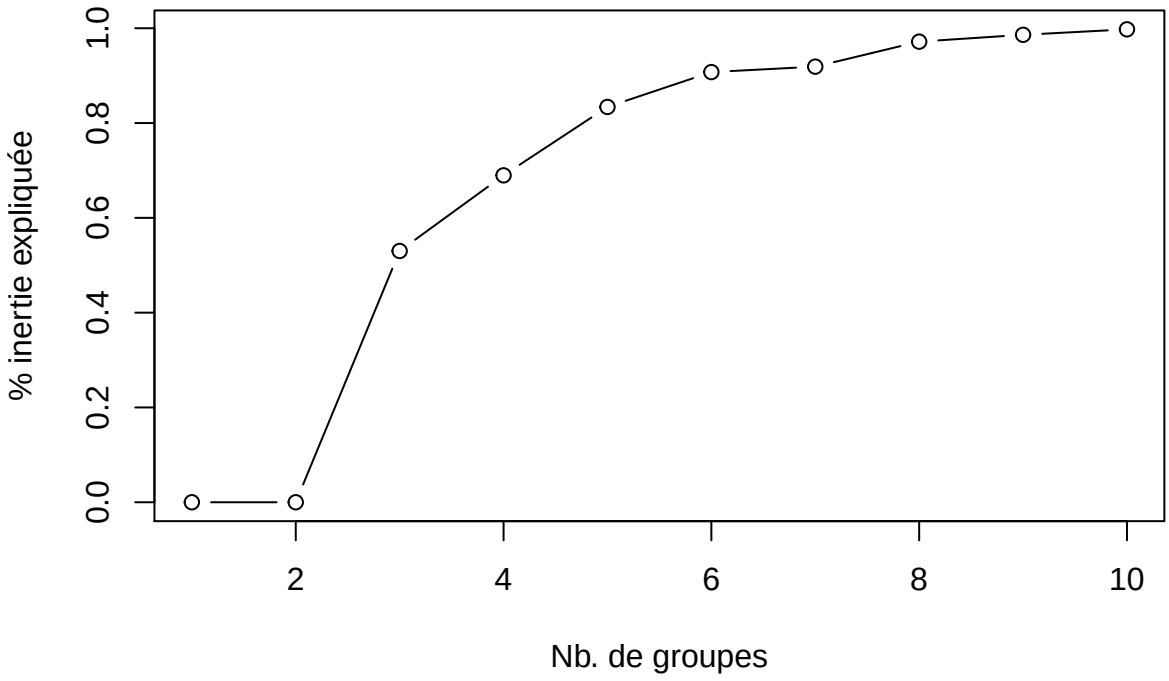



```
print(sort(cutree(tree,k=4)))
```

```
##   Dsp   Seg   Aes  Lsla     L p2lsh   Snv Staps   Shs   Sfa
##     1     1     2     2     2     2     2     2     3     4
```

Methode de K-means

```
groupes.kmeans <- kmeans(etudiants.cr,centers=4,nstart=5)
inertie <- rep(0,times=10)
for (k in 2:10){
  group <- kmeans(etudiants.cr,centers = k - 1,nstart=5)
  inertie[k] <- group$betweenss/group$totss
}
plot(1:10,inertie,type="b",xlab="Nb. de groupes",ylab="% inertie expliquée")
```


% inertie expliquée
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
Nb. de groupes