2019/2020

Faculté des sciences et techniques MARRAKECH

Master SDAD

Enseignant: Pr. A. Ouaarab

[ATELIER 1 : ACP EN R

Réalisé par:

ELMRHARI Yassine

INTRODUCTION

L'objectif de cet atelier est d'étudier des exemples d'application de l'analyse en composantes principales avec le logiciel R ainsi pratiquer cette étude sur un exercice d'application.

EXEMPLE D'APPRENTISSAGE

Pour débuter la phase d'apprentissage on commence par charger les bibliothèques nécessaires :

```
#Installing environnement packages
install.packages("FactoMineR")
library(FactoMineR)
install.packages("ggplot2")
library(ggplot2)
install.packages("factoextra")
library(factoextra)
```

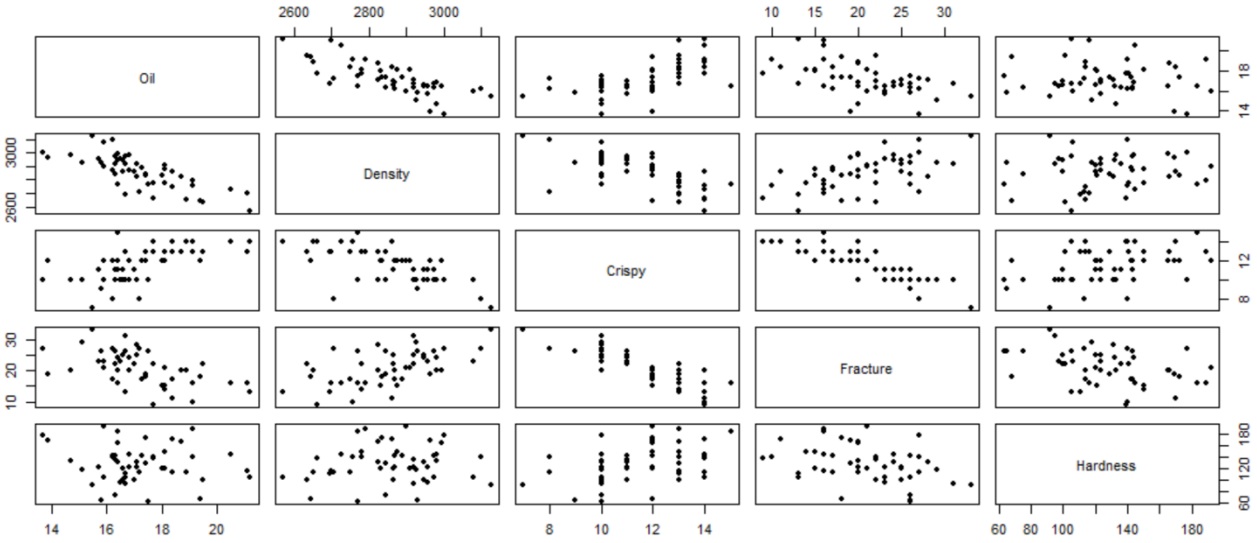
- **FactoMineR :** Bibliothèque de la fonction ACP.
- **ggplot2 :** Bibliothèque des affichages graphiques nécessaires.
- **factoextra :** Bibliothèque de la visualisation des nuages de points ACP.

Ensuite on va charger les données à travailler avec :

```
#Loading data
food <- read.csv("C:/Users/PC Gamer/Desktop/food.csv")
#Cheking data
head(food)
```

On affichage le nuage des points par couples de variables :

```
#Variables display
plot(food[,-1], pch=16, cex=0.8)
```



Application de la fonction ACP sur les données :

```
#Appliying ACP on data
food.pca<-PCA(food[,-1], scale.unit=TRUE, graph=F)
#Displaying ACP summary
summary(food.pca)
#Other way of displaying ACP informations
str(food.pca)
```

```
Call:
PCA(X = food[, -1], scale.unit = TRUE, graph = F)

Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
Variance               3.031   1.296   0.310   0.242   0.121
% of var.             60.624  25.914   6.201   4.838   2.422
Cumulative % of var.  60.624  86.538  92.739  97.578 100.000

Individuals (the 10 first)
                Dist    Dim.1     ctr    cos2    Dim.2     ctr    cos2    Dim.3     ctr
1           |  1.659 |  -1.397   1.288   0.710 |  -0.626   0.604   0.142 |  -0.407   1.066
2           |  3.096 |   2.823   5.258   0.831 |   0.357   0.197   0.013 |  -1.087   7.621
3           |  1.109 |   0.240   0.038   0.047 |   0.870   1.169   0.616 |  -0.618   2.462
4           |  2.403 |  -1.959   2.532   0.664 |  -1.153   2.053   0.230 |   0.442   1.262
5           |  1.567 |  -1.281   1.082   0.668 |   0.670   0.693   0.183 |   0.442   1.258
6           |  2.676 |   2.018   2.686   0.568 |   1.312   2.655   0.240 |   1.113   7.987
7           |  1.664 |   1.503   1.491   0.816 |  -0.639   0.630   0.147 |  -0.276   0.491
8           |  2.582 |  -0.837   0.463   0.105 |  -2.407   8.941   0.869 |  -0.252   0.410
9           |  1.344 |  -1.197   0.946   0.794 |   0.354   0.193   0.069 |  -0.430   1.193
10          |  1.096 |  -1.004   0.664   0.838 |   0.354   0.194   0.105 |   0.104   0.070
              cos2
1           0.060 |
2           0.123 |
3           0.310 |
4           0.034 |
5           0.079 |
6           0.173 |
7           0.027 |
8           0.010 |
9           0.102 |
10          0.009 |

Variables
             Dim.1     ctr    cos2    Dim.2     ctr    cos2    Dim.3     ctr    cos2
Oil       |  0.797  20.934   0.635 |  -0.422  13.722   0.178 |   0.367  43.432   0.135 |
Density   | -0.834  22.920   0.695 |   0.406  12.727   0.165 |   0.009   0.026   0.000 |
Crispy    |  0.927  28.344   0.859 |   0.225   3.907   0.051 |  -0.100   3.200   0.010 |
Fracture  | -0.878  25.450   0.771 |  -0.252   4.895   0.063 |   0.302  29.407   0.091 |
Hardness  |  0.267   2.353   0.071 |   0.916  64.749   0.839 |   0.272  23.935   0.074 |
```

```
List of 5
 $ eig : num [1:5, 1:3] 3.031 1.296 0.31 0.242 0.121 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:5] "comp 1" "comp 2" "comp 3" "comp 4" ...
  .. ..$ : chr [1:3] "eigenvalue" "percentage of variance" "cumulative percentage of var
iance"
 $ var :List of 4
  ..$ coord  : num [1:5, 1:5] 0.797 -0.834 0.927 -0.878 0.267 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:5] "Oil" "Density" "Crispy" "Fracture" ...
  .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
  ..$ cor    : num [1:5, 1:5] 0.797 -0.834 0.927 -0.878 0.267 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:5] "Oil" "Density" "Crispy" "Fracture" ...
  .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
  ..$ cos2   : num [1:5, 1:5] 0.6345 0.6947 0.8592 0.7714 0.0713 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:5] "Oil" "Density" "Crispy" "Fracture" ...
  .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
  ..$ contrib: num [1:5, 1:5] 20.93 22.92 28.34 25.45 2.35 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:5] "Oil" "Density" "Crispy" "Fracture" ...
  .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
 $ ind :List of 4
```

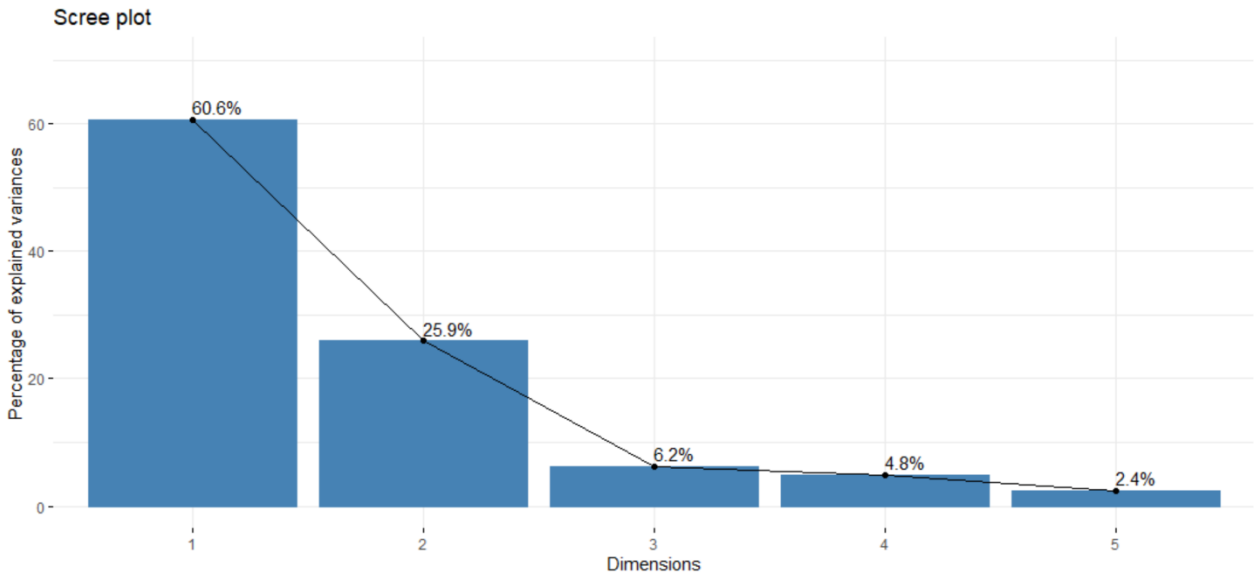
Génération des valeurs propres/variances :

```r
#Proper values calculation
eig.val<-get_eigenvalue(food.pca)
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.0312132	60.624263	60.62426
Dim.2	1.2957058	25.914115	86.53838
Dim.3	0.3100493	6.200987	92.73937
Dim.4	0.2419201	4.838402	97.57777
Dim.5	0.1211116	2.422233	100.00000

Affichage du graphe des pourcentages des variances :

```r
#Displaying the percentage of explained variances
fviz_eig(food.pca, addlabels = TRUE, ylim = c(0, 70))
```



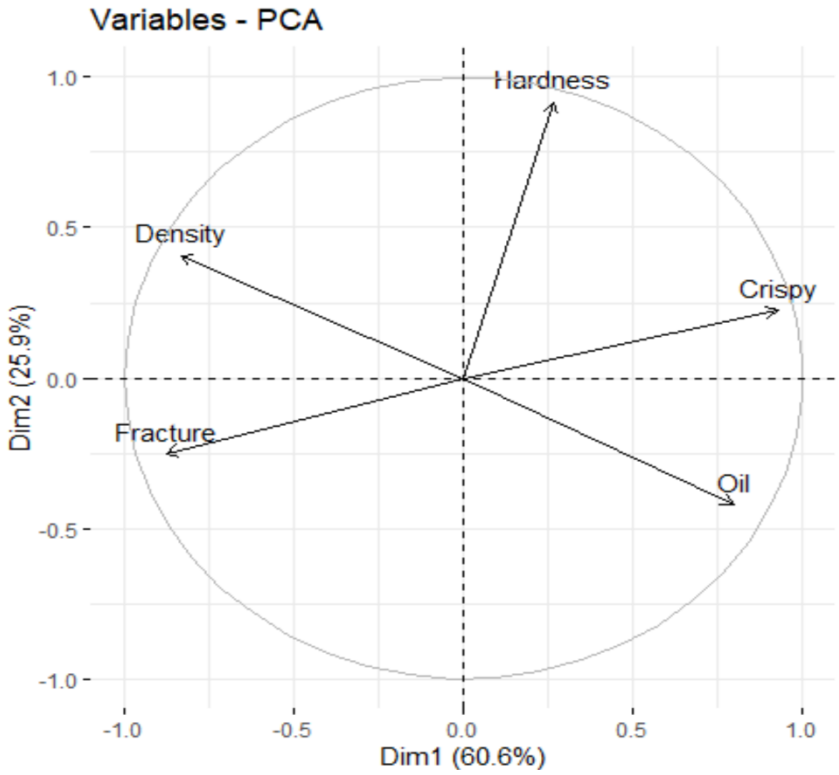
Génération des résultats pour les variables de sortie ACP :

```
#Active variables results
food.var<-get_pca_var(food.pca)
```

```
Principal Component Analysis Results for variables
 ===================================================
  Name       Description
1 "$coord"   "Coordinates for the variables"
2 "$cor"     "Correlations between variables and dimensions"
3 "$cos2"    "Cos2 for the variables"
4 "$contrib" "contributions of the variables"
```

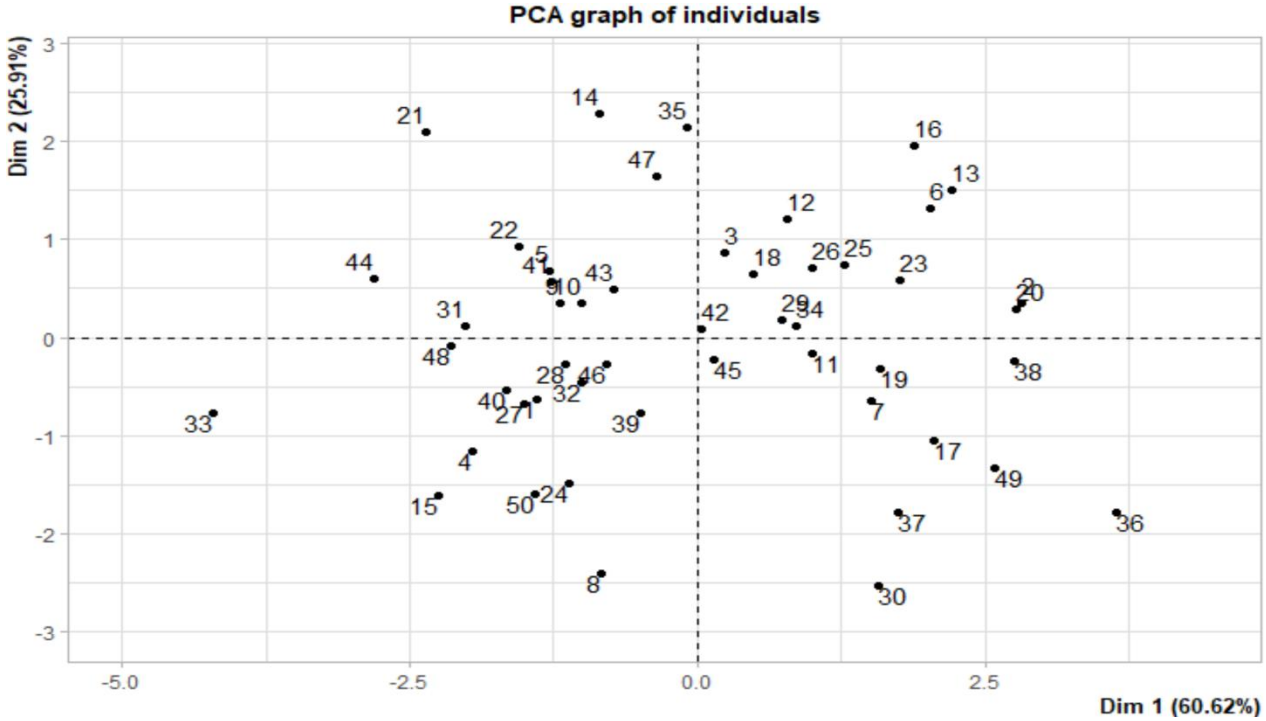
Affichage du graphe de la corrélation des variables :

```
#Displaying variables correlation
fviz_pca_var(food.pca, col.var = "black")
```



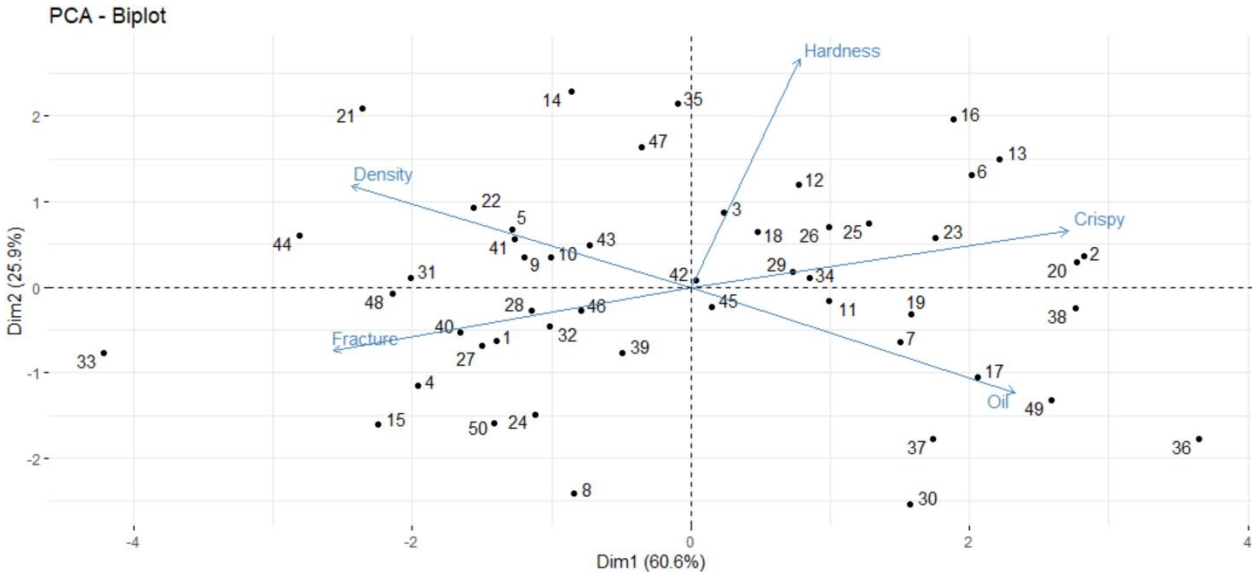
Affichage du graphe de la corrélation des individus :

```
#Displaying individuals
plot(food.pca)
```



Affichage du graphe de la corrélation des individus et des variables en même temps :

```
#Displaying individuals and variables
fviz_pca_biplot(food.pca, repel=TRUE)
```



EXERCICE D'APPLICATION

On commence par charger les 30 premières lignes des données de « quakes » :

```
#Cheking Data
quakes<-head(quakes,30)
```

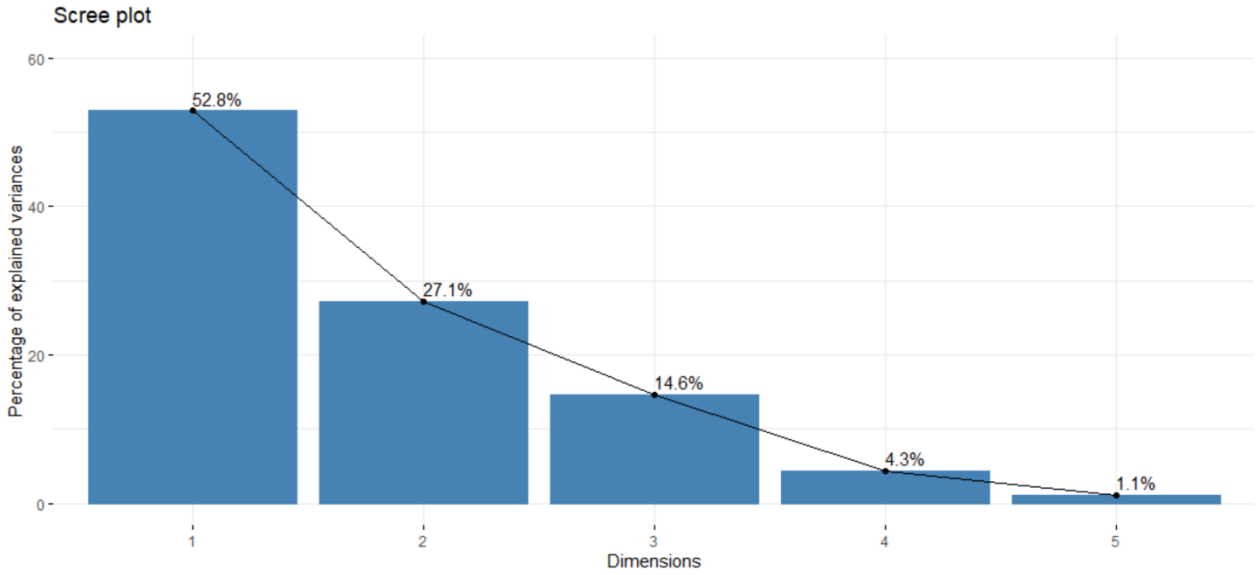
Application de la fonction ACP sur les données :

```
#Appliying ACP on data
quakes.pca=PCA(quakes, scale.unit=TRUE, graph=F)
```

```
Call:
PCA(X = quakes[1:30, ], scale.unit = TRUE, graph = F)


Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
Variance               2.642   1.357   0.730   0.215   0.055
% of var.             52.850  27.145  14.606   4.294   1.105
Cumulative % of var.  52.850  79.994  94.600  98.895 100.000

Individuals (the 10 first)
              Dist    Dim.1    ctr   cos2    Dim.2    ctr   cos2
1         |  1.163 | -0.078  0.008  0.005 |  0.597  0.874  0.263 |
2         |  1.592 | -1.442  2.622  0.820 | -0.236  0.136  0.022 |
3         |  2.896 |  1.186  1.776  0.168 |  2.320 13.224  0.642 |
4         |  1.628 | -1.320  2.196  0.657 | -0.605  0.900  0.138 |
5         |  1.907 | -1.800  4.085  0.891 | -0.386  0.366  0.041 |
6         |  1.924 | -0.986  1.227  0.263 | -0.123  0.037  0.004 |
7         |  3.240 |  2.590  8.459  0.639 | -1.819  8.123  0.315 |
8         |  2.230 | -0.739  0.688  0.110 |  1.331  4.353  0.357 |
9         |  2.226 | -0.020  0.000  0.000 |  1.898  8.851  0.728 |
10        |  1.373 | -0.921  1.070  0.450 | -0.712  1.244  0.269 |

Variables
            Dim.1    ctr   cos2    Dim.2    ctr   cos2
lat       |  0.375  5.318  0.141 | -0.851 53.400  0.725 |
long      | -0.746 21.077  0.557 |  0.560 23.103  0.314 |
depth     | -0.699 18.478  0.488 | -0.094  0.653  0.009 |
mag       |  0.870 28.630  0.757 |  0.419 12.956  0.176 |
stations  |  0.837 26.497  0.700 |  0.366  9.888  0.134 |
```

Affichage du graphe des pourcentages des variances :

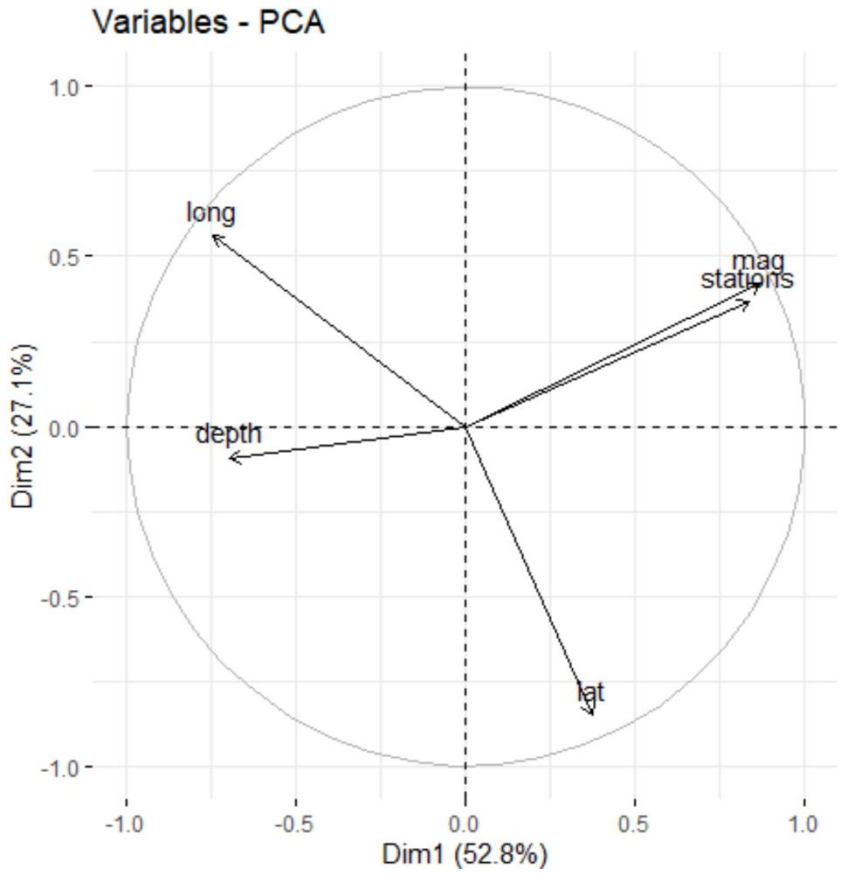
```
#Displaying the percentage of explained variances
fviz_eig(quakes.pca, addlabels = TRUE, ylim = c(0, 60))
```



Continuer l'analyse avec deux composantes principales est suffisant car on a la plupart d'inertie totale sur le premier et le deuxième axe.

Affichage du graphe de la corrélation des variables :

```
#Displaying variables correlation
fviz_pca_var(quakes.pca, col.var = "black")
```

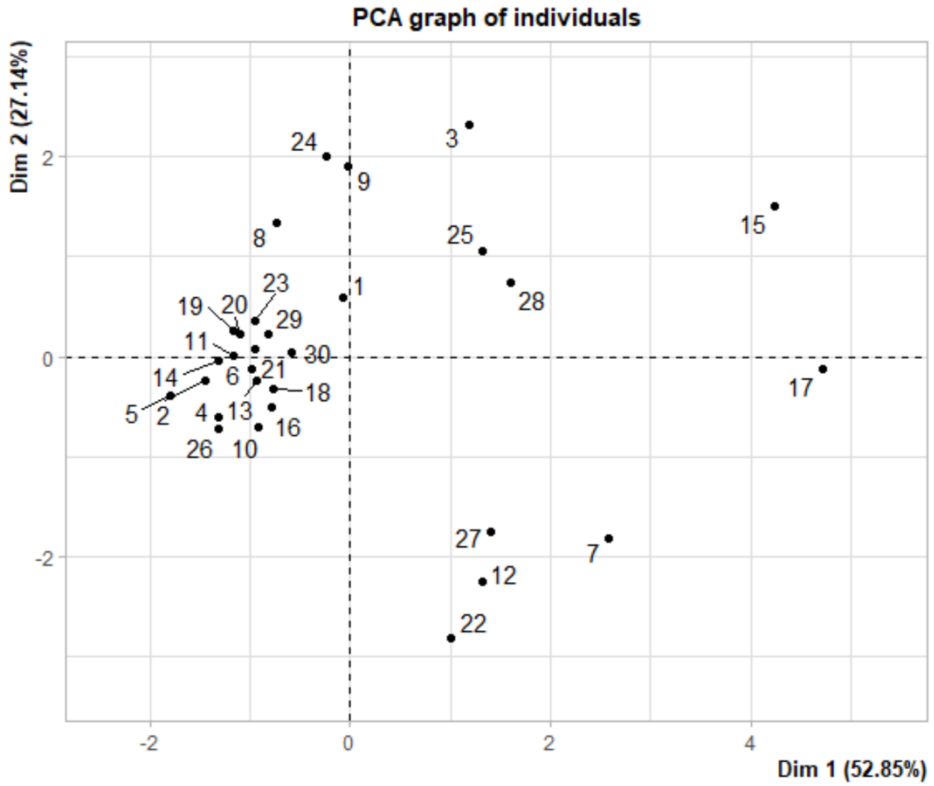



```
> quakes.var$contrib
              Dim.1      Dim.2     Dim.3     Dim.4      Dim.5
lat        5.318355 53.4001866  9.717830 29.0296906  2.5339379
long      21.077181 23.1032349  2.439826 51.5265728  1.8531859
depth     18.477540  0.6530227 63.971971 16.5221774  0.3752884
mag       28.629742 12.9559030  5.062486  0.7282807 52.6235875
stations  26.497182  9.8876528 18.807887  2.1932785 42.6140003
```

Les variables importantes pour la première composante sont : Long, Mag, Stations (x > 20) et pour la deuxième composante : Lat, Long (x > 20)

Affichage du graphe de la corrélation des individus :

```
#Displaying individuals correlation
plot(quakes.pca)
```



Affichage du graphe de la corrélation des individus et des variables :

```
#Displaying individuals and variables
fviz_pca_biplot(quakes.pca, repel=TRUE)
```