

Mini Projet ACP

SEFFANE Asmaa

2023-03-04

Dans ce mini projet, je vais analyser un aspect du jeu de données “protéines”.

L’axe choisit pour réaliser une ACP sera défini tout au long du projet.

```
library(reader)

## Loading required package: NCmisc

##
## Attaching package: 'reader'

## The following objects are masked from 'package:NCmisc':
##
##   cat.path, get.ext, rmv.ext

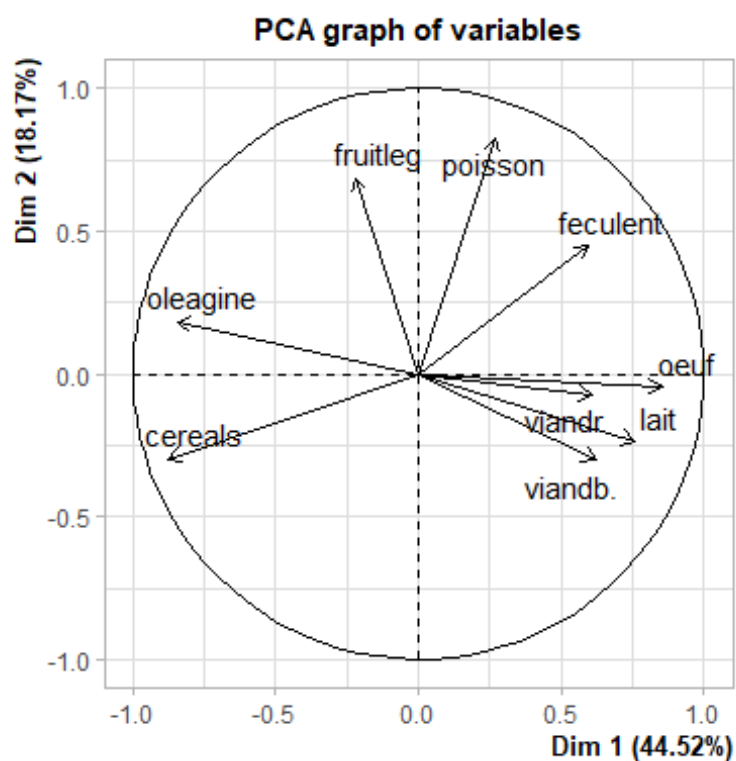
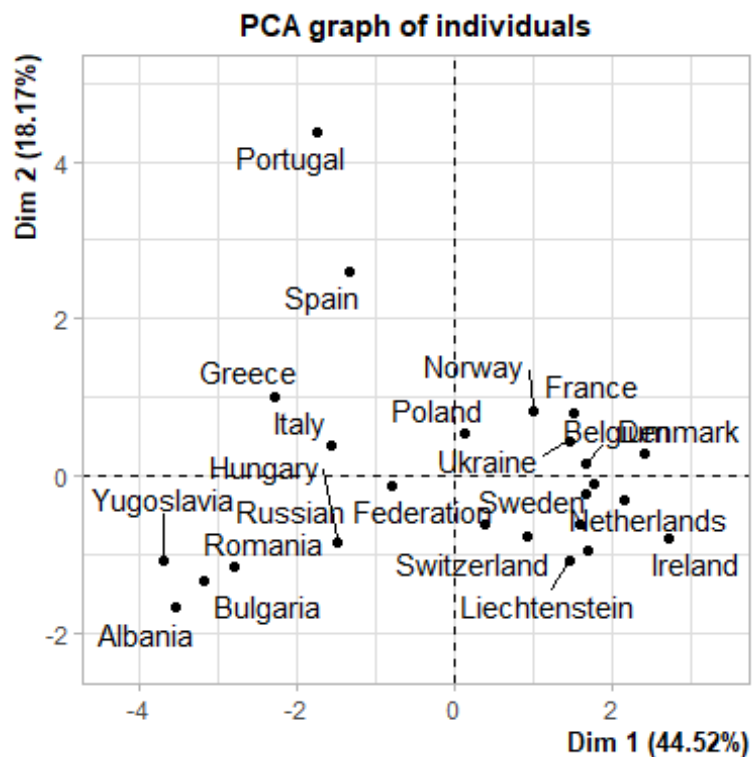
proteine=read.table(file="proteines.dat", dec=",", col.names=c("country",
"viandr","viandb ", "oeuf", "lait", "poisson", "cereals", "feculent",
"oleagine", "fruitleg"))

x = proteine[,-1]
rownames(x) = proteine$country

n=nrow(x)
nvar=ncol(x)

library(FactoMineR)
res.pca <- PCA(x, ncp = 9)

## Warning: ggrepel: 4 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
res.pca$eig
```

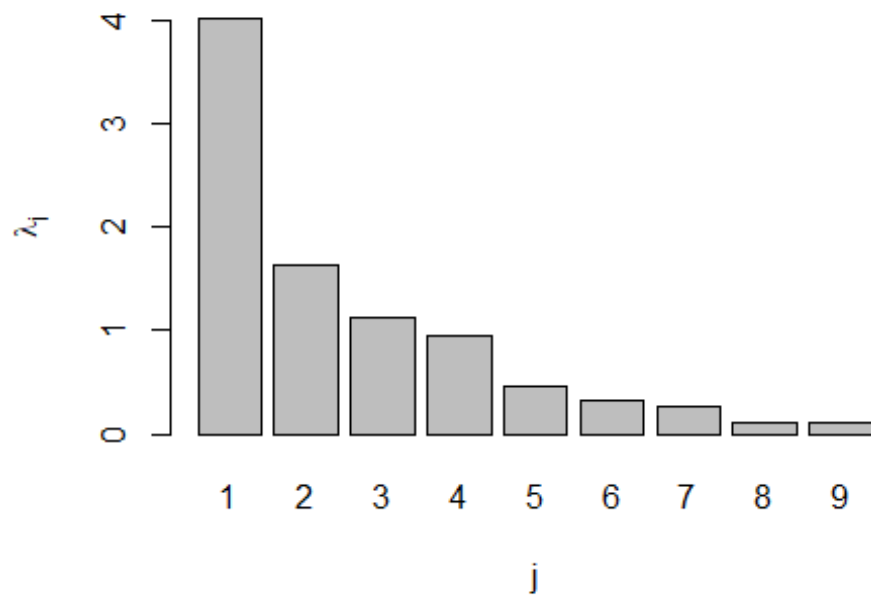
```
##          eigenvalue percentage of variance cumulative percentage of variance
## comp 1   4.0064376          44.515973          44.515973
```

## comp 2	1.6349994	18.166661	62.68263
## comp 3	1.1279195	12.532439	75.21507
## comp 4	0.9546640	10.607377	85.82245
## comp 5	0.4638384	5.153760	90.97621
## comp 6	0.3251310	3.612566	94.58878
## comp 7	0.2716063	3.017848	97.60662
## comp 8	0.1162919	1.292132	98.89876
## comp 9	0.0991119	1.101243	100.00000

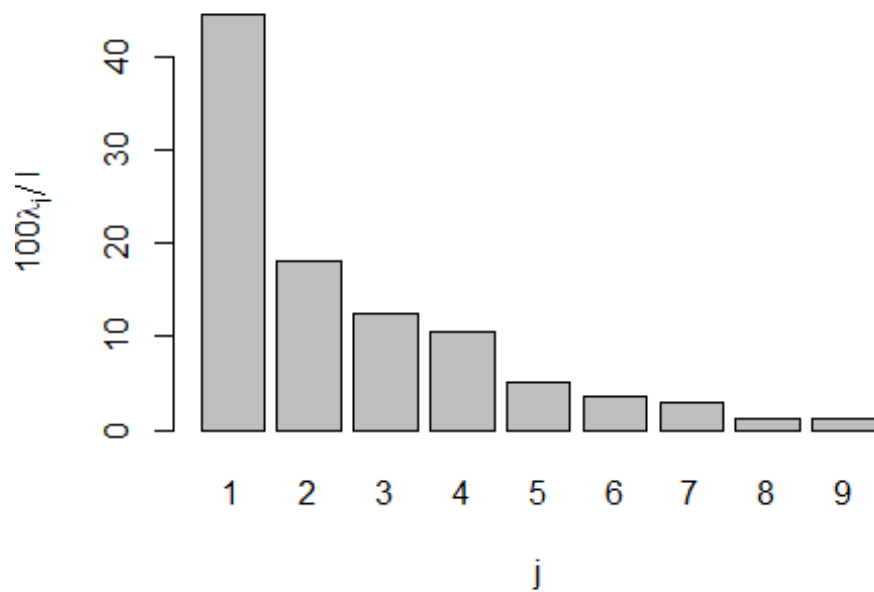
Nous avons neuf valeurs propres, on remarque qu'on pourra prendre les quatre premières valeurs propres pour atteindre les 85%, ou cinq pour atteindre les 90%.

Regardant maintenant le tracé des valeurs propres et leur pourcentage de variance.

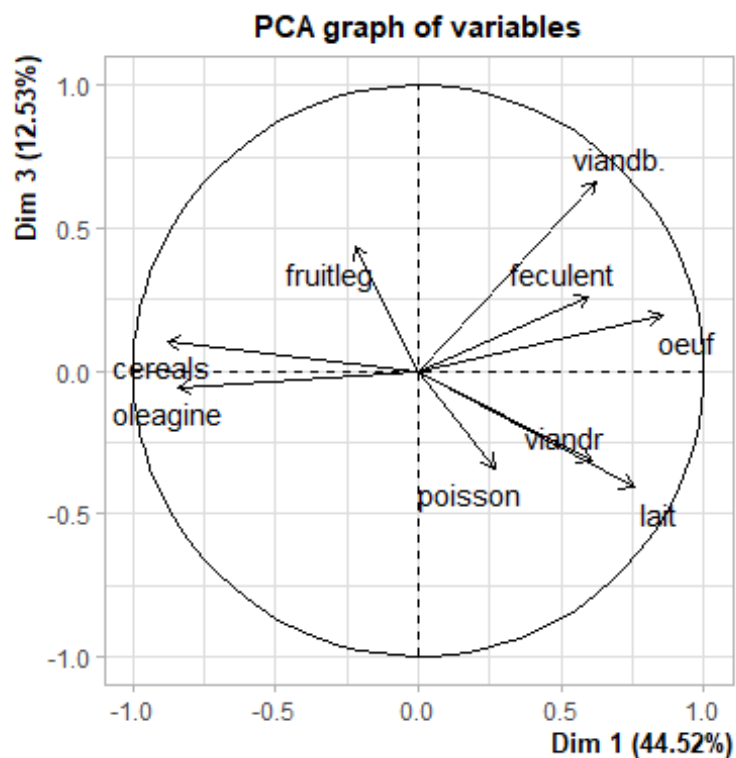
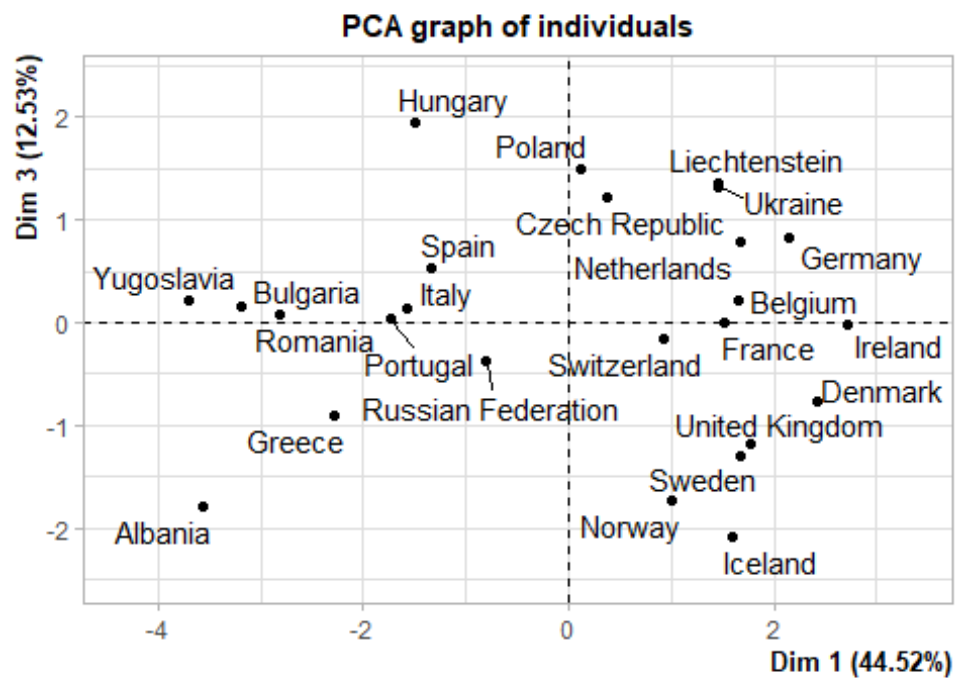
Tracé des valeurs propres



Part de variance expliquée par chaque axe



Ayant déjà vu auparavant la projection sur le plan des deux axes 1 et 2, Je regarde directement la projection sur les axes 1 et 3 avec un pourcentage de variance accumulé de 57%.



Analyse: Je remarque que par rapport à l'axe 3, il y a une consommation de la protéine végétal (à l'exception de la viande blanche) plus que la consommation de la protéine animale.

La distance entre les variables et l'origine mesure la qualité de représentation des variables. Les variables qui sont loin de l'origine sont bien représentées par l'ACP, donc on a les viandes blanches, céréales, lait et oleagine qui sont bien présentés par l'ACP.

Je remarque aussi qu'il y a une grande consommation de viande blanche et œuf pour Germany et Neitherland et Liechtenstein, et nos données confirme ceci.

Iceland, Norway et Sweden ont une consommation très importante du lait, de la viande rouge et du poisson. Les données confirment bien cela.

Ensuite :

```
dimdesc(res.pca)

## $Dim.1
##
## Link between the variable and the continuous variables (R-square)
##
=====
====
##          correlation      p.value
## oeuf          0.8540434 5.631033e-08
## lait           0.7560622 1.234504e-05
## viandb.        0.6216121 9.101673e-04
## viandr         0.6057056 1.333313e-03
## feculent       0.5949735 1.706062e-03
## oleagine      -0.8413450 1.376380e-07
## cereals       -0.8761911 9.505468e-09
##
## $Dim.2
##
## Link between the variable and the continuous variables (R-square)
##
=====
====
##          correlation      p.value
## poisson        0.8270696 3.441990e-07
## fruitleg       0.6856106 1.552354e-04
## feculent       0.4511479 2.360026e-02
##
## $Dim.3
##
## Link between the variable and the continuous variables (R-square)
##
=====
====
##          correlation      p.value
## viandb.        0.6626010 0.000307451
## fruitleg       0.4328391 0.030682052
## lait          -0.4095822 0.042026638
```

On regarde maintenant l'axe 3 : les protéines qui ont un effet important sont viande blanche, fruit légume et lait.

La viande blanche et fruit légume sont les plus consommés sur cet axe car corrélation positive, le lait est quant à lui moins consommé sur ce même axe avec une corrélation négative.

Même si l'axe 3 ne représente que 12,53% d'information, mais il nous fournit des indications intéressantes sur la consommation des protéines. Ces éléments sont complémentaires aux déductions faites depuis les axes 1 et 2.