

# Rapport

*Chater\_Croce\_El Ouedghiri\_Pery*

*28 mars 2018*

## Projet Alimentation

Dans le contexte du projet statistiques de notre deuxième année de formation GIS, nous avons eu à analyser un jeu de données sur l'alimentation. Le but de ce projet était de classer divers produits alimentaires en fonction de leur teneur en nutriments essentiels : Energie, Eau, Vitamine A, Vitamine B1, Vitamine B2, Vitamine C etc. Avec l'aide de plusieurs analyses statistiques, nous allons ensuite définir une règle de décision pour aider les consommateurs à choisir un aliment en fonction de leurs préférences.

### Chargement des packages

```
## Loading required package: lattice
## Loading required package: colorspace
## Loading required package: grid
## Loading required package: data.table
## VIM is ready to use.
## Since version 4.0.0 the GUI is in its own package VIMGUI.
##
## Please use the package to use the new (and old) GUI.
## Suggestions and bug-reports can be submitted at: https://github.com/alexkova/VIM/issues
##
## Attaching package: 'VIM'
## The following object is masked from 'package:datasets':
##
## sleep
##
## Attaching package: 'ggplot2'
## The following objects are masked from 'package:psych':
##
## %+%, alpha
## Loading required package: survival
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
## The following object is masked from 'package:psych':
##
## describe
```

```
## The following objects are masked from 'package:base':
##
##      format.pval, round.POSIXt, trunc.POSIXt, units
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

## Chargement des données

```
rm(list=ls())
d <- read.table("/home/gis3/acroce/Linux/GIS 4/Modélisation/Projet/foodMissing.csv", sep="\t", dec=",", header=TRUE)
d = d[,-1]
#Definition du nom des lignes
row.names(d)=d[,1]
d = d[,-1]
head(d)
attach(d)
```

## Analyse préalable

Après avoir visualiser le jeu de données fourni, nous pouvons très rapidement remarquer que certaines valeurs vont nous poser problème pour effectuer une règle de décision efficace. Par exemple, les valeurs très différentes de Vitamines A faussent l'analyse car elles créent une classe constituée uniquement des individus ayant de très hautes valeurs pour cette variable. Nous avons donc décidé de la coder en tant que variable qualitative en trois niveaux : Null [0;1], Medium ]1;150] et High ]150;15000]. De cette manière, nous regroupons ces individus avec les autres individus présentant une forte teneur en vitamines A. Ce découpage est arbitraire mais basé sur ce que nous avons pu observer du comportement de la teneur en Vitamines A des individus.

```
d$Vitamin.A.levels. <- cut(d$Vitamin.A..mg., breaks=c(0,1,150,15000), include.lowest = TRUE)
levels(d$Vitamin.A.levels.) <- c("null","medium","high")
d <- d[,-5]
summary(d$Vitamin.A.levels.)
```

```
##      null medium   high
##      319      97     53
```

## Imputation des valeurs manquantes

Nous avons observé plusieurs valeurs aberrantes dans le jeu de données : la quantité de protéines dans les cornflakes qui est supérieure au poids total de l'échantillon et la teneur de vitamines B2 dans le whiting steamed. Nous allons supprimer ces valeurs et utiliser une technique d'imputation pour les remplacer. Nous aurions pu uniquement les supprimer, mais dans un soucis de qualité d'analyse, l'imputation est la meilleure solution.

## Tableau des données manquantes

```
md.pattern(d)
```

	Energy..Kcal.	Fat..g.	Water..g.	Vitamin.B1..mg.	Vitamin.C
## 467	1	1	1	1	1
## 1	1	1	1	1	1
## 1	1	1	1	1	1

```
##           0           0           0           0           0
##   Vitamin.A.levels. Protein..g. Vitamin.B2..mg.
## 467           1           1           1 0
##   1           1           0           1 1
##   1           1           1           0 1
##           0           1           1 2
```

Il y a 2 individus présentant des données manquantes. 1 individu avec 1 valeur manquante pour la variable Vitamin.B2..mg. Et pour finir 1 individu avec 1 valeur manquante pour la variable Protein..g..

## Modèle d'imputation

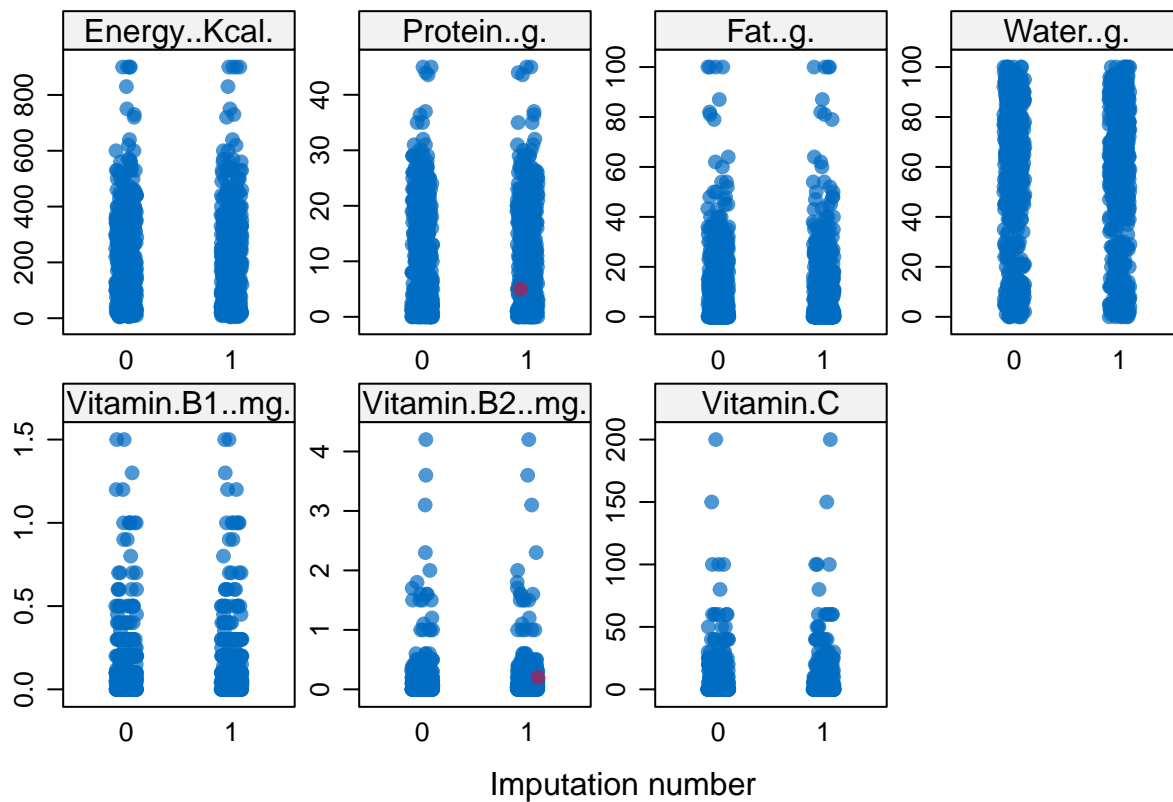
```
dm=mice(d,m=1,seed=10,print=FALSE,maxit=50)
print(dm$imp)
```

```
## $Energy..Kcal.
## NULL
##
## $Protein..g.
##           1
## Cornflakes 5
##
## $Fat..g.
## NULL
##
## $Water..g.
## NULL
##
## $Vitamin.B1..mg.
## NULL
##
## $Vitamin.B2..mg.
##           1
## Whiting, steamed 0.2
##
## $Vitamin.C
## NULL
##
## $Vitamin.A.levels.
## NULL
```

Le modèle d'imputation affiche les valeurs imputées pour les variables présentant des valeurs manquantes. La teneur en protéines des cornflakes est donc de 5g/100g. La teneur en vitamine B2 des steamed whiting est de 0.2mg/100g.

### Affichage du modèle d'imputation

```
stripplot(dm,pch=20,cex=1.2)
```



Sur le modèle d'imputation on peut voir les valeurs imputées par rapport aux autres valeurs.

### Remplissage des valeurs manquantes avec les valeurs imputées

```
d1 = complete(dm,1)
```

Ici, nous ajoutons les valeurs imputées à notre jeu de données.

## Analyse

Notre jeu de données contient donc des variables quantitatives et une variable qualitative. Par conséquent, nous ne pouvons pas uniquement nous baser sur une ACP normée ou sur une ACM. Nous avons besoin d'une nouvelle méthode d'analyse permettant d'avoir ces deux types de variables. Nous allons donc réaliser une FAMD (mélange d'une ACP et d'une ACM) afin de visualiser les composantes principales avant de réaliser la Classification Ascendante Hiérarchique correspondante. Nous utiliserons les packages FactoMineR et factoextra.

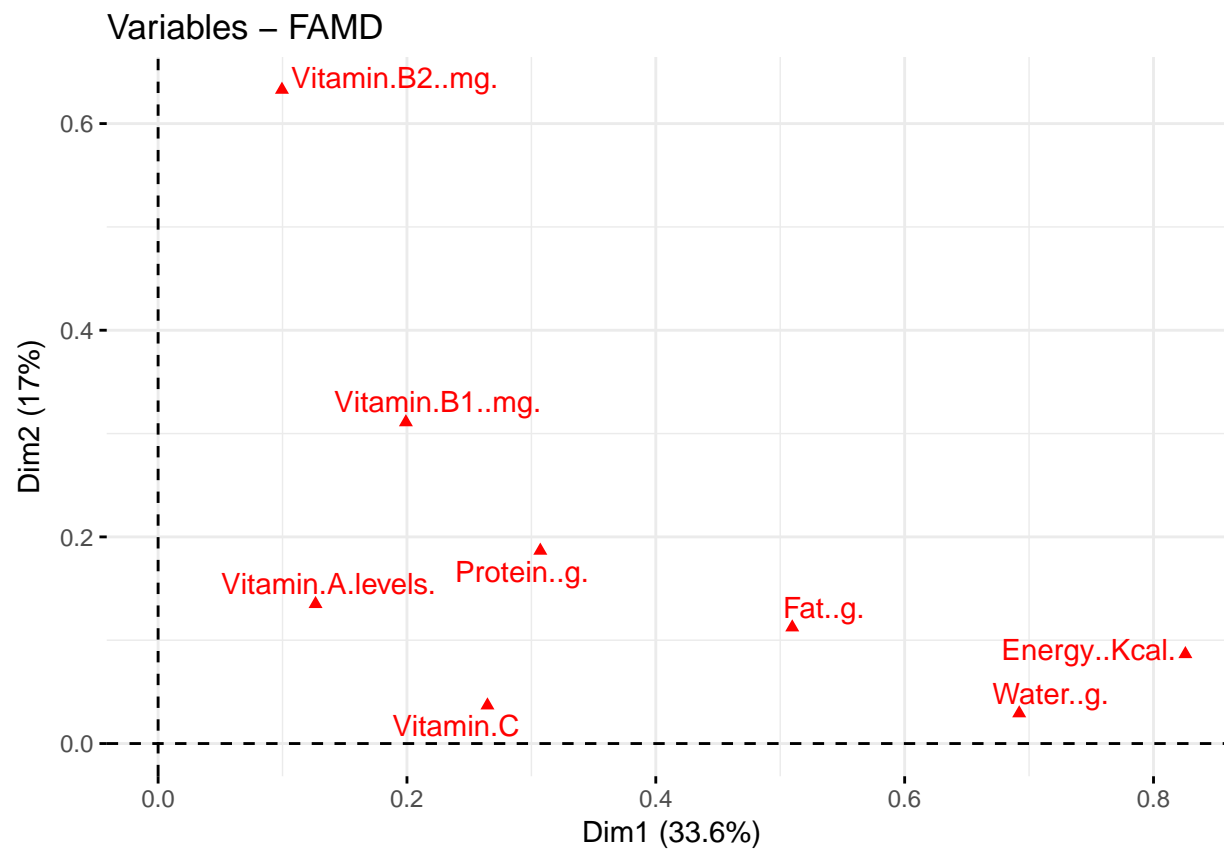
### Statistiques descriptives

```
summary(d1)
```

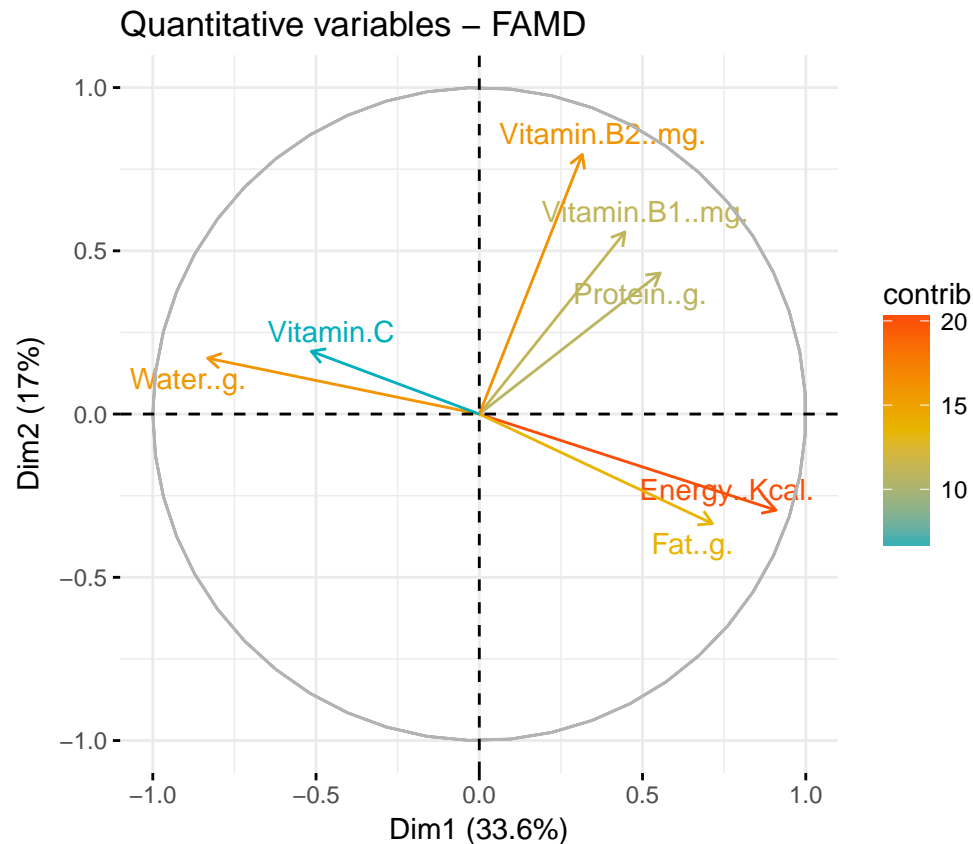
```
## Energy..Kcal.      Protein..g.      Fat..g.      Water..g.
## Min.      : 5.0    Min.      : 0.00    Min.      : 0.00    Min.      : 0.00
## 1st Qu.: 65.0    1st Qu.: 1.50    1st Qu.: 0.00    1st Qu.: 30.00
## Median :200.0    Median : 7.00    Median : 4.00    Median : 63.00
## Mean   :210.3    Mean   :10.27    Mean   : 11.03    Mean   : 55.58
## 3rd Qu.:320.0    3rd Qu.:18.00    3rd Qu.: 16.00    3rd Qu.: 80.00
## Max.    :900.0    Max.    :45.00    Max.    :100.00    Max.    :100.00
## Vitamin.B1..mg.    Vitamin.B2..mg.      Vitamin.C      Vitamin.A.levels.
## Min.      :0.0000    Min.      :0.0000    Min.      : 0.000    null      :319
## 1st Qu.:0.0000    1st Qu.:0.0300    1st Qu.: 0.000    medium: 97
## Median :0.0600    Median :0.1000    Median : 0.000    high   : 53
## Mean   :0.1295    Mean   :0.1924    Mean   : 6.387
## 3rd Qu.:0.1000    3rd Qu.:0.2000    3rd Qu.: 4.000
## Max.    :1.5000    Max.    :4.2000    Max.    :200.000
```

## FAMD

*#Corrélation entre les variables et contribution aux axes principaux*  
`fviz_famd_var` (`d1.famd`, `repel = TRUE`)



```
# Affichage des variables quantitatives sur le cercle de corrélation
fviz_famd_var(d1.famd, "quanti.var", col.var = "contrib",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              repel = TRUE)
```



Les variables quantitatives semblent être regroupées en 3 groupes : groupe 1 : Energy et Fat corrélées très positivement avec l'axe 1 et légèrement négativement avec l'axe 2. Groupe 2 : Vitamine B1, Vitamine B2 et Proteine corrélées positivement avec les axes 1 et 2. Groupe 3 : Water et Vitamin C corrélées négativement avec l'axe 1 mais légèrement positivement avec l'axe 2. La coloration de chaque variable indique son niveau de contribution.

### Valeurs Propres

##	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	3.0244147	33.604608	33.60461
## comp 2	1.5307188	17.007987	50.61259
## comp 3	1.2610717	14.011908	64.62450
## comp 4	0.9884953	10.983281	75.60778
## comp 5	0.7036339	7.818155	83.42594

On retient 4 Composantes principales afin d'expliquer 75% de la variation. En faisant ce choix, nous nous assurons que notre analyse permettra d'obtenir une règle de décision viable.

### Informations sur les variables quantitatives

#### Coordonnées

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Energy..Kcal.	0.9086374	-0.2942342	0.23745563	0.10071033	0.04925132
## Protein..g.	0.5542328	0.4320512	-0.32254588	-0.06404989	-0.23560830
## Fat..g.	0.7139052	-0.3353273	0.41479284	-0.01259931	-0.07809096
## Water..g.	-0.8319065	0.1711374	-0.04845151	-0.21251694	-0.18127902
## Vitamin.B1..mg.	0.4462523	0.5576532	-0.21165016	0.41131193	0.31572321
## Vitamin.B2..mg.	0.3154845	0.7954823	0.14097110	-0.05365842	-0.15322102
## Vitamin.C	-0.5144809	0.1921544	0.45079152	0.22497690	0.51259857

Axe 1 : on voit que les variables Energy, Fat et Protein sont très positivement corrélées avec l'axe tandis que les variables Water et Vitamin C sont très négativement corrélées.

Axe 2 : on voit que les variables Vitamin B1 et Vitamin B2 sont positivement corrélées avec l'axe tandis que les variables Energy et Fat sont un petit peu négativement corrélées. Cela contraste avec les variables de l'axe 1.

Axe 3 : on voit que la variable Vitamin C et Fat sont faiblement positivement corrélées tandis que Protein et Vitamin B1 sont un petit peu négativement corrélée.

Axe 4 : on voit que la variable Vitamin C est légèrement corrélée négativement avec l'axe. Les autres variables sont très peu corrélées avec l'axe.

### Qualité

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Energy..Kcal.	0.82562189	0.08657376	0.056385178	0.0101425702	0.002425692
## Protein..g.	0.30717399	0.18666825	0.104035846	0.0041023887	0.055511270
## Fat..g.	0.50966061	0.11244442	0.172053103	0.0001587426	0.006098198
## Water..g.	0.69206846	0.02928799	0.002347549	0.0451634497	0.032862082
## Vitamin.B1..mg.	0.19914109	0.31097708	0.044795790	0.1691775030	0.099681143
## Vitamin.B2..mg.	0.09953049	0.63279201	0.019872852	0.0028792264	0.023476681
## Vitamin.C	0.26469059	0.03692331	0.203212996	0.0506146077	0.262757297

Axe 1 : Les variables Energy et Water sont de très bonne qualité pour cet axe. Les variables Fat et Protein sont de qualité convenable.

Axe 2 : La variable Vitamin B2 est de très bonne qualité pour cet axe. La variable Vitamin B1 est de qualité convenable.

Axe 3 : La variable Vitamin C est de qualité moyenne pour cet axe.

Axe 4 : La variable Vitamin B1 est de qualité moyenne pour cet axe.

### Contribution

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## Energy..Kcal.	27.298568	5.655758	4.4712111	1.02606153	0.3447378
## Protein..g.	10.156477	12.194810	8.2497964	0.41501347	7.8892257
## Fat..g.	16.851545	7.345857	13.6434038	0.01605902	0.8666719
## Water..g.	22.882724	1.913349	0.1861551	4.56890884	4.6703377
## Vitamin.B1..mg.	6.584451	20.315755	3.5522001	17.11464898	14.1666194
## Vitamin.B2..mg.	3.290901	41.339532	1.5758701	0.29127365	3.3364907
## Vitamin.C	8.751796	2.412155	16.1143096	5.12036902	37.3428966

Axe 1 : L'axe 1 est principalement défini par les variables Energy, Water, Fat et Water.

Axe 2 : L'axe 2 est principalement défini par les variables Vitamin B2, Vitamin B1 et Protein.

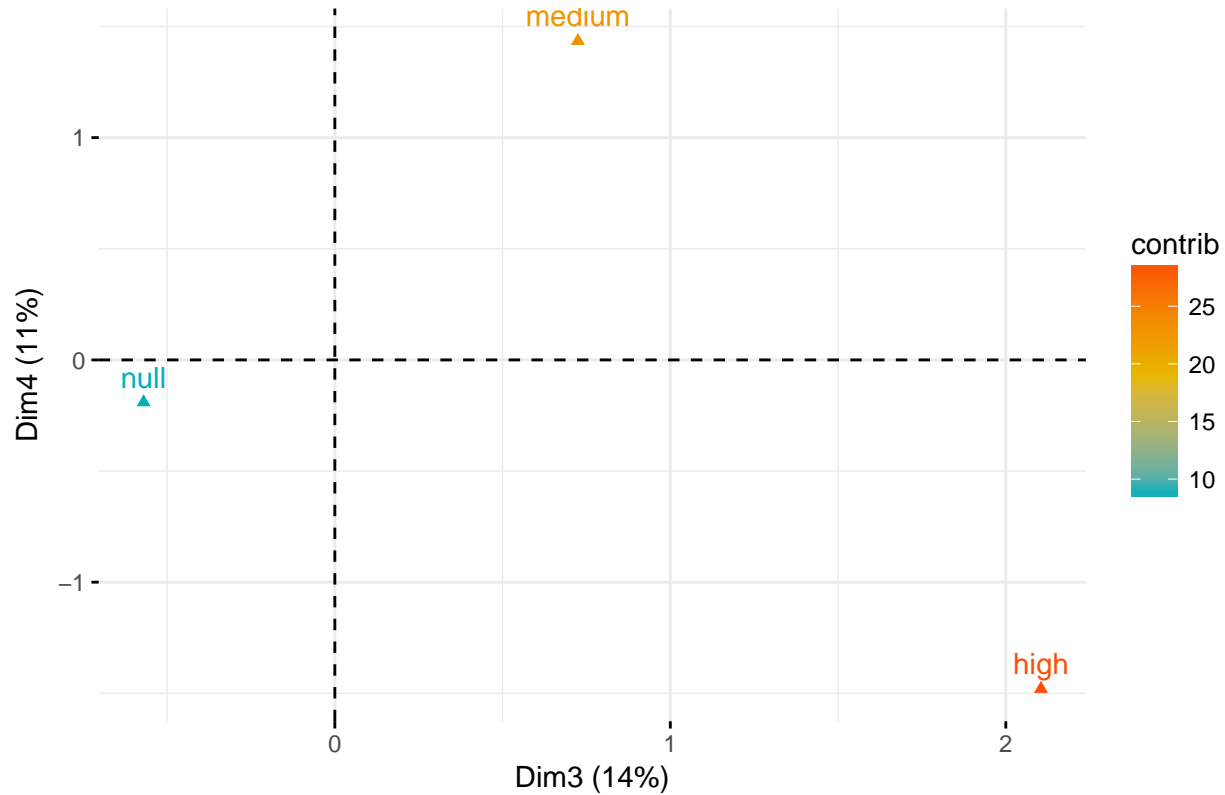
Axe 3 : L'axe 3 est principalement défini par les variables Vitamin C et Fat.



Axe 4 : L'axe 4 est principalement défini par la variable Vitamin B1.

### Informations sur la variable qualitative (Vitamin A)

#### Qualitative variable categories – FAMD



```
## $coord
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## null    0.2874620 -0.21470161 -0.5700313 -0.1905477  0.22670771
## medium -1.2059166  0.02567938  0.7245773  1.4351649 -0.76269032
## high    0.4768592  1.24526254  2.1048301 -1.4797411  0.03134343
##
## $cos2
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## null    0.14670176 0.0818361388 0.5768621 0.06445876 0.0912445819
## medium  0.31387536 0.0001423285 0.1133163 0.44455581 0.1255507065
## high    0.02551858 0.1740195613 0.4971762 0.24572406 0.0001102477
##
## $contrib
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## null    0.6144633 1.338127809 13.897471  2.527412  7.06084767
## medium  3.2881433 0.005820725  6.827927 43.596664 24.29974902
## high    0.2809316 7.478836127 31.481657 25.323590  0.02242346
##
## $v.test
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## null    5.214737 -5.4746988 -16.014061 -6.046291  8.5263966
## medium -7.660091  0.2292842  7.127746 15.946002 -10.0441263
## high    2.117306  7.7719036 14.473093 -11.492450  0.2885276
```

La variable quantitative Vitamine A est représentée sur le plan des axes (3,4) car c'est sur ces axes qu'elle est le mieux représentée.

### **Coordonnées**

Axe 1 : Une teneur moyenne en Vitamine A est corrélée négativement avec l'axe 1. Une teneur nulle ou haute en Vitamine A est légèrement positivement corrélée avec l'axe 1.

Axe 2 : Une teneur haute en Vitamine A est positivement corrélée avec l'axe 2. Une teneur nulle en Vitamine A est légèrement négativement corrélée avec l'axe 1.

Axe 3 : Une teneur haute en Vitamine A est très positivement corrélée avec l'axe 3. Une teneur moyenne a une légère corrélation positive et une teneur nulle a une légère corrélation négative.

Axe 4 : Une forte teneur en Vitamine A est négativement corrélée avec l'axe 4. Une teneur moyenne est positivement corrélée.

### **Qualité**

Axe 1 : Une teneur moyenne en Vitamine A est un indicateur de moyenne qualité pour l'axe 1.

Axe 2 : La Vitamine A n'est globalement pas un indicateur de bonne qualité pour l'axe 2.

Axe 3 : Une teneur nulle ou élevée en Vitamine A est un indicateur de bonne qualité pour l'axe 3.

Axe 4 : Une teneur moyenne en Vitamine A est un indicateur de bonne qualité pour l'axe 4. Une teneur forte en Vitamine A est un indicateur de moyenne qualité.

### **Contribution**

Axe 1 : La teneur en Vitamine A ne contribue que très peu à l'axe 1.

Axe 2 : La teneur en Vitamine A ne contribue que très peu à l'axe 2.

Axe 3 : La teneur en Vitamine A contribue à environ 50% de l'axe 3. Plus particulièrement une forte teneur et une teneur nulle.

Axe 4 : La teneur en Vitamine A contribue à environ 70% de l'axe 4, plus particulièrement une moyenne et une forte teneur.

### **Information sur les individus**

#### **Coordonnées**

Voici les individus qui ont les coordonnées les plus élevées et les plus faibles sur chaque axe.

Axe 1 : + Oil vegetable, Dripping, Fat Cooking, Lard.

- Blackcurrants, raw green peppers, Brussels sprouts raw .

Axe 2 : + Foie frit, foie mijoté, foie cru.

- Dripping, Fat Cooking, Lard, Oil vegetable

Axe 3 : + Margarine, beurre, persil, cassis.

- Soya flour low fat, Bemax, Bacon gammon rashers, Ready Brek.

Axe 4 : + Blackcurrants, Peas split dried raw ready brek, Roe cod - Pumpkin raw, Carrots boiled, Carrots raw, Endive.

## **Qualité**

Voici les individus les mieux représentés sur chaque axe.

Axe 1 : + Fried streaky bacon, roasted and salted peanuts, peanut butter, Salami

Axe 2 : + fried kidney, stewed kidney, raw heart

Axe 3 : + frozen fish fingers, steamed Plaice, raw Plaice, raw Salmon, cream cheese

Axe 4 : + raw chickpeas, rip plantain, raw dried peas

## **Contribution**

Voici les individus qui contribuent le plus à chaque axe.

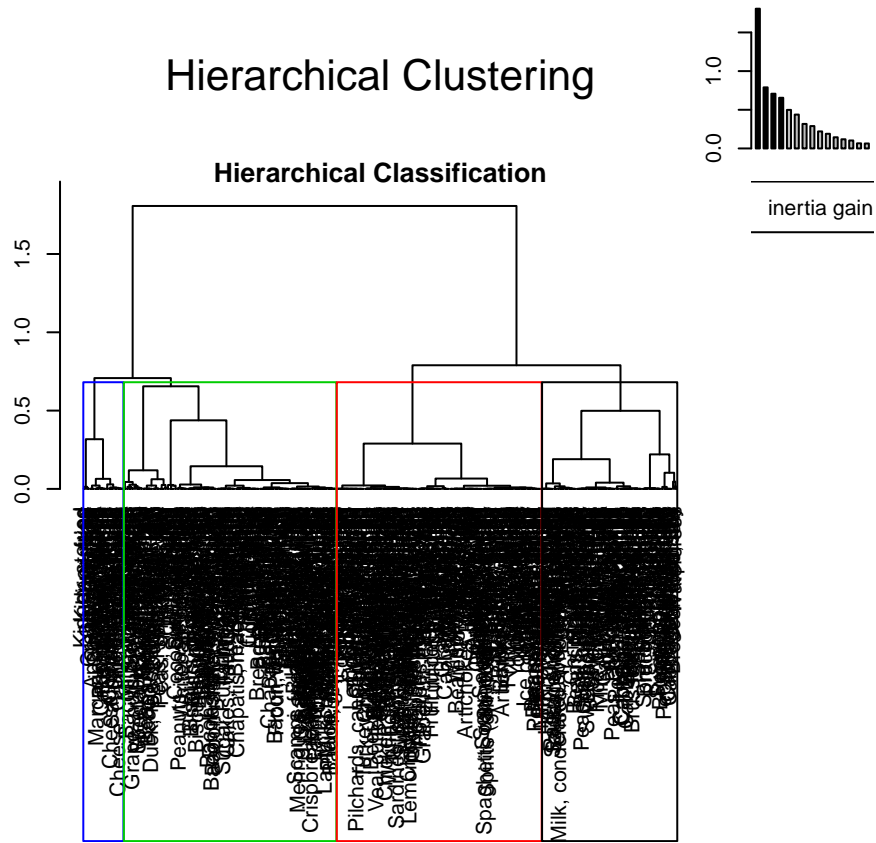
Axe 1 : + blackcurrants, vegetable oil, Lard, cooking fat

Axe 2 : + fried liver, stewed liver, raw liver, fried kidney

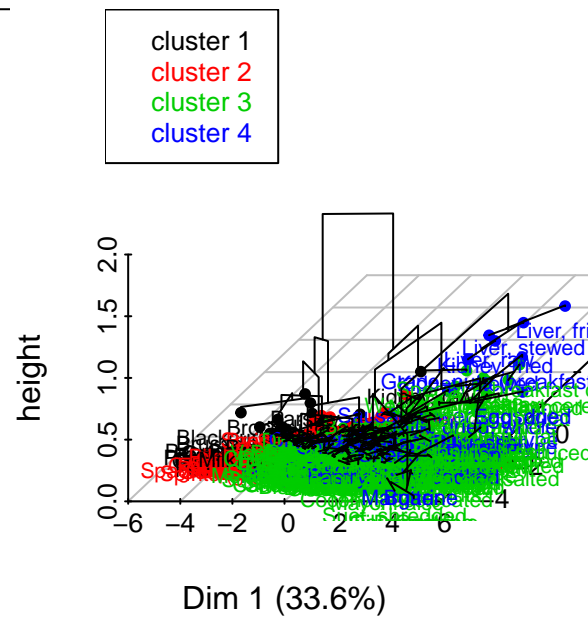
Axe 3 : + blackcurrants, parsley, butter, margarine

Axe 4 : + blackcurrants, cod roe, ready brek, dried raw peas

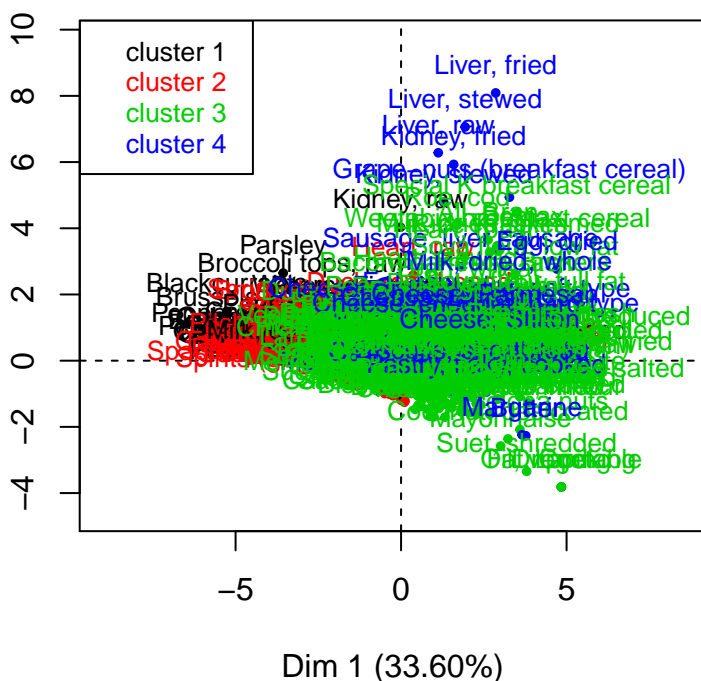
```
d1.hcpc = HCPC(d1.famd, nb.clust = 4)
```



## Hierarchical clustering on the



## Factor map



Une fois qu'on a réussi à modéliser nos données de la façon souhaitée, nous avons procédé à une classification hiérarchique afin de pouvoir répartir nos variables en différentes classes qui permettront de faciliter la tâche à l'utilisateur pour décider de l'alimentation qu'il veut. Cette classification nous permet donc de répartir les types d'aliments sous 4 classes chacune possédant ses propres caractéristiques selon les nutriments contenus dans ces aliments.

## Description par les Composantes Principales

```
d1.hcpc$desc.axes
```

```
## $quanti.var
##           Eta2           P-value
## Dim.1 0.6671507 1.179048e-110
## Dim.4 0.6318475 1.732451e-100
## Dim.3 0.4829319 3.014055e-66
## Dim.5 0.1416710 2.466256e-15
## Dim.2 0.1096613 1.084792e-11
##
## $quanti
## $quanti$`1`
##           v.test Mean in category Overall mean sd in category Overall sd
## Dim.4 12.992340      1.2511915 3.468263e-16      0.5362813 0.9942310
## Dim.3  7.333228      0.7976534 -2.095280e-16      0.8303806 1.1229745
## Dim.5 -7.097239     -0.5766496 -1.932756e-16      1.3484189 0.8388289
## Dim.1 -10.299218     -1.7348983 -8.403607e-18      1.1741778 1.7390844
##
##           p.value
## Dim.4 1.352288e-38
## Dim.3 2.246737e-13
## Dim.5 1.272737e-12
## Dim.1 7.103677e-25
```

```
##
## $quanti$`2`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.5    3.086033          0.1482601 -1.932756e-16      0.4625413  0.8388289
## Dim.3   -8.508572         -0.5472389 -2.095280e-16      0.7464218  1.1229745
## Dim.1   -9.280030         -0.9243159 -8.403607e-18      0.8522659  1.7390844
## Dim.4  -10.504611         -0.5981604  3.468263e-16      0.4032971  0.9942310
##          p.value
## Dim.5  2.028460e-03
## Dim.3  1.760881e-17
## Dim.1  1.694309e-20
## Dim.4  8.226109e-26
##
## $quanti$`3`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.1  14.602055          1.6159332 -8.403607e-18      1.0338973  1.7390844
## Dim.4   5.125066          0.3242467  3.468263e-16      0.8156193  0.9942310
## Dim.5   4.232452          0.2259198 -1.932756e-16      0.6714940  0.8388289
## Dim.2  -3.279403         -0.2581851 -9.818757e-17      1.3908898  1.2372222
## Dim.3  -3.775473         -0.2697926 -2.095280e-16      0.8853409  1.1229745
##          p.value
## Dim.1  2.725059e-48
## Dim.4  2.974329e-07
## Dim.5  2.311569e-05
## Dim.2  1.040271e-03
## Dim.3  1.597043e-04
##
## $quanti$`4`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.3  11.808838          2.1585643 -2.095280e-16      0.6713465  1.1229745
## Dim.2   6.733531          1.3560587 -9.818757e-17      2.5199505  1.2372222
## Dim.1   6.071328          1.7186691 -8.403607e-18      1.1293364  1.7390844
## Dim.5  -2.899950         -0.3959604 -1.932756e-16      0.5961725  0.8388289
## Dim.4  -8.954734         -1.4491986  3.468263e-16      0.4720182  0.9942310
##          p.value
## Dim.3  3.513820e-32
## Dim.2  1.655944e-11
## Dim.1  1.268568e-09
## Dim.5  3.732223e-03
## Dim.4  3.405553e-19
##
##
## attr("class")
## [1] "catdes" "list "
```

```
dim1<- d1.hcpc$desc.axes$quanti$`1`[,1]
dim1
```

```
##          Dim.4          Dim.3          Dim.5          Dim.1
##  12.992340    7.333228   -7.097239  -10.299218
```

```
dim2<- d1.hcpc$desc.axes$quanti$`2`[,1]
dim2
```

```
##          Dim.5          Dim.3          Dim.1          Dim.4
##    3.086033   -8.508572   -9.280030  -10.504611
```

```
dim3<- d1.hcpc$desc.axes$quanti$`3`[,1]
dim3
```

```
##      Dim.1      Dim.4      Dim.5      Dim.2      Dim.3
## 14.602055  5.125066  4.232452 -3.279403 -3.775473
```

```
dim4<- d1.hcpc$desc.axes$quanti$`4`[,1]
dim4
```

```
##      Dim.3      Dim.2      Dim.1      Dim.5      Dim.4
## 11.808838  6.733531  6.071328 -2.899950 -8.954734
```

Nous avons effectué cette partie afin de déterminer la contributions des différentes classes sur les axes qui définissent les composantes principales. On rappellera que dans une première partie, nous avons retenu les 4 premiers axes en composantes principales. Nous remarquons ainsi que les aliments de la classe 1 possèdent des coordonnées élevées sur le 3ème et 4ème axe, ceux de la classe 2 ont des coordonnées faibles sur les axes 1,3 et 4, ceux de la classe 3 contribuent bien au 1er et 4ème axe et ceux de la classe 4 contribuent plus aux 1ère,2ème et 3ème axe.

### Description par les variables

```
par(mfrow=c(1,1))
#Test d'indépendance des variables
d1.hcpc$desc.var$test.chi2
```

```
##                                p.value df
## Vitamin.A.levels. 2.257594e-141  6
## On rejette l'hypothèse d'indépendance h0.
```

```
d1.hcpc$desc.var$category
```

```
## $`1`
##              Cla/Mod  Mod/Cla  Global      p.value
## Vitamin.A.levels.=medium 85.5670103 95.402299 20.68230 9.422718e-72
## Vitamin.A.levels.=high   5.6603774  3.448276 11.30064 5.740291e-03
## Vitamin.A.levels.=null   0.3134796  1.149425 68.01706 2.546883e-51
##              v.test
## Vitamin.A.levels.=medium 17.912488
## Vitamin.A.levels.=high   -2.762259
## Vitamin.A.levels.=null   -15.070110
##
## $`2`
##              Cla/Mod  Mod/Cla  Global      p.value
## Vitamin.A.levels.=null  53.29154 91.89189 68.01706 3.058203e-21
## Vitamin.A.levels.=medium 0.00000 0.00000 20.68230 2.572267e-25
##              v.test
## Vitamin.A.levels.=null   9.460723
## Vitamin.A.levels.=medium -10.396493
##
## $`3`
##              Cla/Mod  Mod/Cla  Global      p.value
## Vitamin.A.levels.=null  46.39498 91.358025 68.01706 6.283463e-17
## Vitamin.A.levels.=medium 14.43299  8.641975 20.68230 1.001516e-06
## Vitamin.A.levels.=high   0.00000 0.000000 11.30064 3.173784e-11
##              v.test
```

```
## Vitamin.A.levels.=null      8.359781
## Vitamin.A.levels.=medium -4.891340
## Vitamin.A.levels.=high    -6.638274
##
## $`4`
##               Cla/Mod Mod/Cla   Global      p.value    v.test
## Vitamin.A.levels.=high  66.03774      100 11.30064 7.916517e-40 13.207741
## Vitamin.A.levels.=medium 0.00000        0 20.68230 2.117325e-04 -3.704591
## Vitamin.A.levels.=null   0.00000        0 68.01706 2.315819e-19 -8.997184
```

```
d1.hcpc$desc.var$quanti
```

```
## $`1`
##               v.test Mean in category Overall mean sd in category
## Vitamin.C      8.714603      21.41379310      6.3869936      34.20679227
## Water..g.      6.699828      74.97931034      55.5754797      19.10360824
## Vitamin.B1..mg. -2.766045      0.07137931      0.1294883      0.09227859
## Fat..g.        -4.423752      3.94482759      11.0347548      5.85794987
## Protein..g.    -5.614325      4.93563218      10.2720682      6.16290305
## Energy..Kcal.  -7.074630      93.75862069     210.3432836     92.75464676
##               Overall sd      p.value
## Vitamin.C      17.8020393 2.917793e-18
## Water..g.      29.9002402 2.086647e-11
## Vitamin.B1..mg. 0.2168876 5.674077e-03
## Fat..g.        16.5463389 9.700154e-06
## Protein..g.     9.8130634 1.973306e-08
## Energy..Kcal.  170.1329953 1.498478e-12
##
## $`2`
##               v.test Mean in category Overall mean sd in category
## Water..g.      10.501096      73.55837838      55.5754797      18.98312330
## Protein..g.    -4.341764      7.83189189      10.2720682      8.96672583
## Vitamin.B2..mg. -4.583594      0.08664865      0.1924307      0.12092263
## Vitamin.B1..mg. -5.589698      0.06005405      0.1294883      0.07657886
## Fat..g.        -8.805362      2.69027027      11.0347548      4.38212739
## Energy..Kcal.  -11.172187     101.48108108     210.3432836     73.05489176
##               Overall sd      p.value
## Water..g.      29.9002402 8.538273e-26
## Protein..g.     9.8130634 1.413432e-05
## Vitamin.B2..mg. 0.4029540 4.570508e-06
## Vitamin.B1..mg. 0.2168876 2.274645e-08
## Fat..g.        16.5463389 1.304308e-18
## Energy..Kcal.  170.1329953 5.579017e-29
##
## $`3`
##               v.test Mean in category Overall mean sd in category
## Energy..Kcal.   14.450720      366.7901235     210.3432836     136.4583069
## Fat..g.         9.406452      20.9388889      11.0347548      20.5289174
## Vitamin.B1..mg.  7.770228      0.2367284      0.1294883      0.3079559
## Protein..g.      7.491358      14.9500000      10.2720682      9.5302394
## Vitamin.C       -5.488317      0.1697531      6.3869936      1.2947706
## Water..g.      -14.562473      27.8679012      55.5754797      21.9639313
##               Overall sd      p.value
## Energy..Kcal.   170.1329953 2.480781e-47
## Fat..g.         16.5463389 5.131670e-21
```



```
## Vitamin.B1..mg.    0.2168876 7.834527e-15
## Protein..g.       9.8130634 6.816485e-14
## Vitamin.C         17.8020393 4.057822e-08
## Water..g.         29.9002402 4.866478e-48
##
## $`4`
##               v.test Mean in category Overall mean sd in category
## Vitamin.B2..mg.  8.740296      0.7657143    0.1924307    1.055785
## Fat..g.          5.899052     26.9228571    11.0347548    18.214485
## Energy..Kcal.    5.094543     351.4285714   210.3432836   149.878863
## Protein..g.      2.823969     14.7828571    10.2720682    11.255589
## Water..g.       -3.089849     40.5371429    55.5754797    22.591226
##               Overall sd      p.value
## Vitamin.B2..mg.  0.402954 2.325009e-18
## Fat..g.          16.546339 3.655968e-09
## Energy..Kcal.    170.132995 3.495828e-07
## Protein..g.      9.813063 4.743292e-03
## Water..g.        29.900240 2.002585e-03
```

Nous commençons par émettre une hypothèse  $H_0$  selon laquelle les variables (nutriments) sont indépendantes. On effectue ensuite un test de  $\chi^2$  afin de déterminer la validité de l'hypothèse. La valeur obtenue par ce test est inférieure à 0.05 ce qui nous pousse à rejeter l'hypothèse  $H_0$ . On peut donc dire que les variables sont corrélées entre elles.

Nous faisons ensuite une description des groupes par les variables, on arrive donc à définir les variables caractérisant le plus chaque groupe :

Classe 1 : Il s'agit ici de Vitamine A en teneur moyenne, Vitamine C et de l'eau

Classe 2 : Pour ce groupe il s'agit de la Vitamine A en teneur faible ainsi que l'eau

Classe 3 : Ici plus de variables interviennent, on a donc la Vitamine A en teneur moyenne et faible avec une plus grande tendance vers la teneur faible, avec l'énergie, la graisse, la Vitamine B1 et les protéines.

Classe 4 : Il s'agit ici de la Vitamine A en grande teneur, la Vitamine B2, la graisse, l'énergie et les protéines

On peut donc remarquer que les groupes 1,2 et les groupes 3,4 se rapprochent un peu avec quelques différences nous permettant ainsi de dissocier les différents aliments selon les différents groupes. Ainsi nous pouvons dire que la classe 1 va contenir les aliments à base des nutriments la caractérisant le plus et ainsi de suite pour les 3 autres classes.

## Description des classes

### #Variables quantitatives

```
x1<- d1.hcpc$desc.var$quanti$`1`
x2<- d1.hcpc$desc.var$quanti$`2`
x3<- d1.hcpc$desc.var$quanti$`3`
x4<- d1.hcpc$desc.var$quanti$`4`
```

### #Variables qualitatives

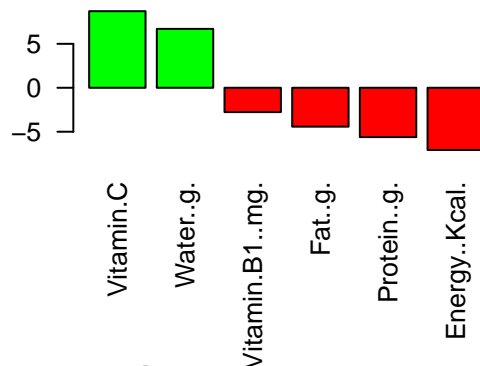
```
y1<- d1.hcpc$desc.var$category$`1`
y2<- d1.hcpc$desc.var$category$`2`
y3<- d1.hcpc$desc.var$category$`3`
y4<- d1.hcpc$desc.var$category$`4`
```

### #Affichage des histogrammes

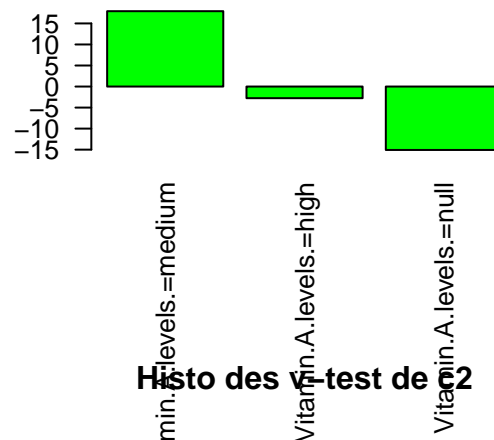
```
par(mfrow=c(2,2))
```

```
barplot(x1[,1], col=ifelse(x1>0,'green','red'), names.arg=row.names(x1[,0]), las=2, main='Histo des v-t  
barplot(y1[,5], col=ifelse(y1>0,'green','red'), names.arg=row.names(y1[,0]), las=2, main='Histo des v-t  
barplot(x2[,1], col=ifelse(x2>0,'green','red'), names.arg=row.names(x2[,0]), las=2, main='Histo des v-t  
barplot(y2[,5], col=ifelse(y2>0,'green','red'), names.arg=row.names(y2[,0]), las=2, main='Histo des v-t
```

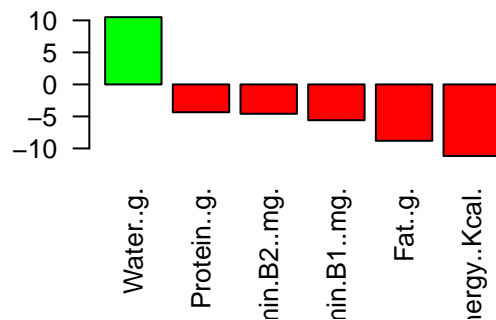
Histo des v-test de c1



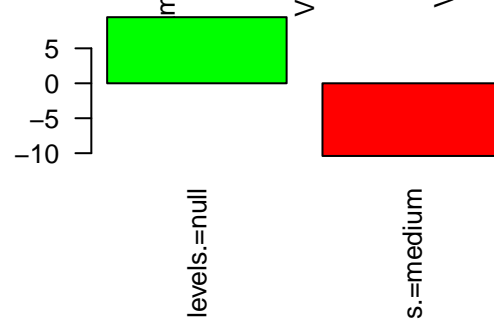
Histo des v-test de c1



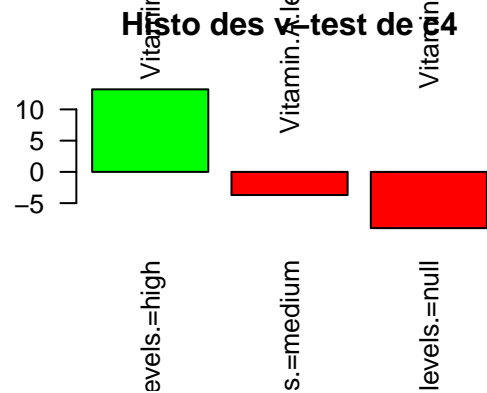
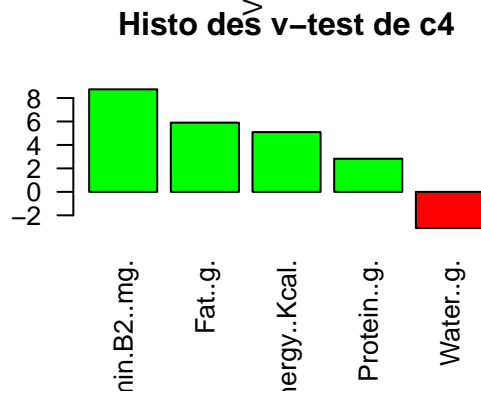
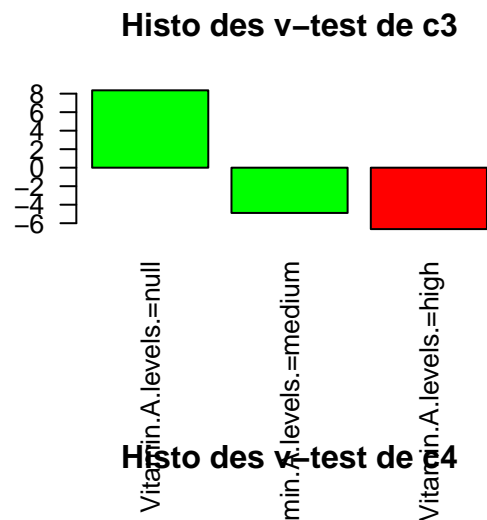
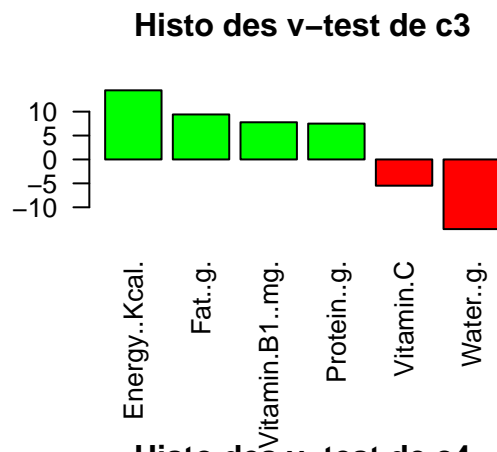
Histo des v-test de c2



Histo des v-test de c2



```
barplot(x3[,1], col=ifelse(x3>0,'green','red'), names.arg=row.names(x3[,0]), las=2, main='Histo des v-t  
barplot(y3[,5], col=ifelse(y3>0,'green','red'), names.arg=row.names(y3[,0]), las=2, main='Histo des v-t  
barplot(x4[,1], col=ifelse(x4>0,'green','red'), names.arg=row.names(x4[,0]), las=2, main='Histo des v-t  
barplot(y4[,5], col=ifelse(y4>0,'green','red'), names.arg=row.names(y4[,0]), las=2, main='Histo des v-t
```



Nous avons essayé dans cette partie d'expliciter les analyses effectuées dans la partie précédente. En les illustrant sous forme de graphiques, nous pouvons voir que l'histogramme des valeurs test de chaque classe nous confirme la répartition que nous avons effectué.

### Description par les Individus

*#Description par les Individus*

d1.hcpc\$desc.ind

```
## $para
## Cluster: 1
##   Beans, broad, boiled      Tomatoes, fried      Asparagus, boiled
##           0.5182234           0.5480245           0.5693954
##   Blackberries Sweetcorn, kernels only
##           0.6380259           0.6901649
## -----
## Cluster: 2
##   Yoghurt, flavoured  Beans, mung, cooked  Beans, baked, canned
##           0.3578854           0.3721395           0.4231180
##   Lentils, boiled  Beans, butter, boiled
##           0.4403788           0.4539361
## -----
## Cluster: 3
##   Bacon, collar joint  Bacon, collar joint, boiled
##           0.4190213           0.4740462
##   Cocoa powder          Pork pie
##           0.4819278           0.6494201
## Bacon rashers, streaky, raw
```

```

##                                0.6609291
## -----
## Cluster: 4
##                Egg yolk                Cheese, Danish blue
##                0.3833039                0.6376534
## Cheese, Camembert, soft type                Cheese, Edam
##                0.7394580                0.8103877
##                Haggis, boiled
##                0.8134505
##
## $dist
## Cluster: 1
##                Blackcurrants                Parsley                Broccoli tops, raw
##                9.796636                8.002808                5.862208
## Peppers, green, raw Brussels sprouts, raw
##                5.423656                5.366618
## -----
## Cluster: 2
## Broccoli tops, boiled Spring greens, boiled                Mangoes, fresh
##                3.603772                3.542600                3.462879
##                Melons                Spinach, boiled
##                3.440300                3.388431
## -----
## Cluster: 3
##                Bemax                Ready Brek                Dripping Fat, Cooking                Lard
##                6.999234                6.456603                6.364901                6.364901                6.364901
## -----
## Cluster: 4
## Liver, fried Liver, stewed                Liver, raw                Kidney, fried Kidney, stewed
##                9.329717                8.288821                7.480712                6.581292                5.804053

```

Cette partie nous montre quels sont les aliments les plus représentatifs ou les moins représentatifs de chaque classe.

On trouve ainsi pour la classe 1 que les aliments pour lesquels s'applique le mieux notre hiérarchisation sont : Beans broad or boiled, Tomatoes fried, Asparagus boiled, Blackberries Sweetcorn et kernels only. Ceux qui sont le plus loin du barycentre de la classe sont : Blackcurrants, Parsley, Broccoli tops raw et Peppers green raw.

Les aliments : Yoghurt flavoured, Beans mung cooked, Beans baked canned, Lentils boiled sont les plus représentatifs de la classe 2 tandis que Broccoli tops boiled, Spring greens boiled, Mangoes fresh sont les moins significatifs.

Pour la classe 3, Bacon boiled, Cocoa powder, Pork pie et Bacon rashers streaky sont les plus proches du barycentre tandis que les plus éloignés sont Bemax, Ready Brek et Dripping.

Les aliments Cheese Camembert, Cheese Edam et Haggis boiled sont ceux pour lesquels s'applique le mieux cette classification tandis que Liver fried, Liver stewed, Liver raw sont les plus éloignés du barycentre.

## Règle de décision

Une personne voulant donc des aliments contenant plus d'eau et de Vitamine C, avec une Vitamine A en teneur moyenne, une valeur nutritive très faible en Energie et en protéines, une valeur nutritive relativement faible en graisse se servira donc des aliments appartenants à la classe 1. Une personne qui aimerait s'alimenter à base d'eau et de Vitamines A uniquement devra s'intéresser à la classe 2. On notera que les aliments présents dans cette classe ont une teneur très faible en Energie, graisse et Vitamine B1. La classe 3 intéressera les personnes qui veulent s'alimenter de produits contenant de la Vitamine A en teneur moyenne, de l'Energie, de la graisse, de la Vitamine B1 et des protéines (à l'opposé complètement de la classe 1 ) et donc avec des

valeurs nutritives faibles en ce qui concerne l'eau et la Vitamine C. la classe 4 sera pour ceux qui veulent des aliments à base de Vitamine B2, graisse, Energie et Protéines et qui ne sont pas intéressés par la présence de l'eau dans ce qu'ils mangent.