

Pascal AUREGAN

INE 0G5DRJ1IK04

Pascal.auregan@gmail.com

**Projet STA101 - Analyse des données : méthodes descriptives**

**Analyse descriptive des principaux nutriments contenus dans 214 aliments.**

Table des matières

[Introduction 3](#_Toc443353989)

[Présentation des Données 4](#_Toc443353990)

[Obtention et nettoyage des données 4](#_Toc443353991)

[Présentation des variables 4](#_Toc443353992)

[Présentation des individus 7](#_Toc443353993)

[Analyse descriptive 8](#_Toc443353994)

[Statistiques de base 8](#_Toc443353995)

[ACP 9](#_Toc443353996)

[Classification 17](#_Toc443353997)

[Caractérisations des classes 18](#_Toc443353998)

[Conclusion et perspectives 22](#_Toc443353999)

[Annexes 23](#_Toc443354000)

[A1. Matrice des corrélations de Pearson des 21 premières variables 23](#_Toc443354001)

[A2. Description des méthodes utilisées dans SPAD 24](#_Toc443354002)

[A3. Description des scripts R 25](#_Toc443354003)

[A4. Coordonnées, contributions et cosinus carrés des individus 25](#_Toc443354004)

[A5. Coordonnées des individus actifs contribuant fortement à l’axe 4 25](#_Toc443354005)

[Bibliographie 26](#_Toc443354006)

# Introduction

Dans un contexte sociétal où une partie de la population veut connaitre avec de plus en plus de détail ce dont elle se nourrit, il nous apparaissait intéressant d’étudier les nutriments d’une grande partie des aliments que nous utilisons pour préparer nos repas.

Dans le cadre de ce projet, nous allons analyser, décrire, une population de 214 aliments à travers les 32 principaux nutriments qui les composent. Les aliments choisis pour cette étude sont des aliments bruts. C'est-à-dire qu’ils n’ont fait l’objet d’aucune transformation autre que leur préparation pour le commerce.

Nous décrirons en un premier temps, à travers leur présentation, de quelle manière ces données sur ces aliments ont été choisis et obtenus. Ce sera aussi l’occasion de décrire les aliments et les nutriments choisis pour cette étude. On peut ici informer le lecteur que l’ensemble de l’étude a été menée grâce au logiciel SPAD lorsque cela a été possible. Le lecteur trouvera par ailleurs en annexe une A2. Description des méthodes utilisées dans SPAD dans le cadre de ce projet ainsi qu’une A3. Description des scripts R ayant servi à cette analyse de données.

En un second temps, nous décrirons les données grâce tout d’abord aux statistiques de base puis à une analyse en composantes principales. L’analyse de la classification nous permettra enfin de tenter à la question ayant suscité pour nous l’intérêt de cette étude : est-ce que la classification naturelle des aliments basée sur la classification phylogénétique peut se juxtaposer à une classification basée sur leur composition nutritionnelle ? [[1]](#footnote-1)

# Présentation des Données

### Obtention et nettoyage des données

Lorsque, dans google, vous tapez « nutrition apricot », une fonctionnalité de google propose la décomposition de l’aliment en quelques nutriments.

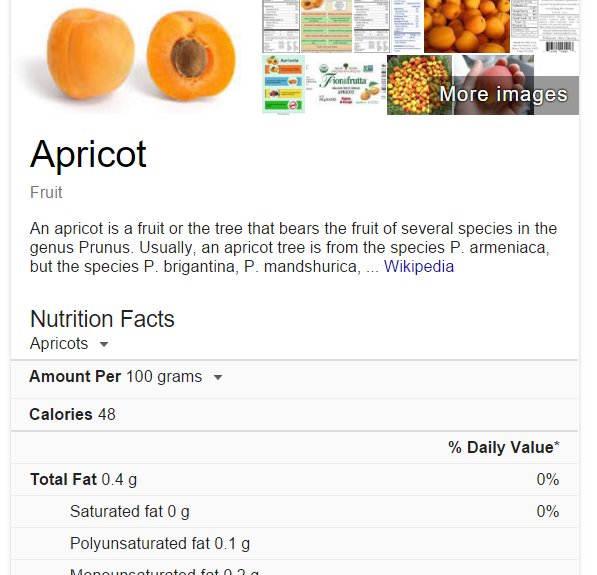


Figure 1 : Affichage de Google pour « nutrition apricot »

Cette fonctionnalité utilise la base de données de l’USDA : United States Department of Agriculture. L’ensemble (version SR28) de la base de données est disponible sous format texte sur le site <http://www.ars.usda.gov/Services/docs.htm?docid=25700>. Cette base de données est normalisée, chaque table étant représentée par une fichier texte au format non standard [1]. Une première étape a donc été transformer ces fichiers au format CSV que peut gérer SPAD. Un simple remplacement des « ^ » et « ~ » par « ; » et « ″ »   a été faite. Lorsqu’une valeur manquante est rencontrée, l’individu est supprimé. Cette action a été fait en dehors de SPAD. La licence empêchant de travailler sur plus de 50 000 cellules. Une autre action a été de ne garder que les aliments bruts. Les catégories des plats transformés ont donc été enlevé en dehors de SPAD. Le jeu de données est en anglais. Pour garder l’homogénéité, il ne sera traduit en français que au cas par cas au besoin et à chaque fois qu’un terme sera cité dans ce document.

### Présentation des variables

En prévision de l’ACP à effectuer sur ces données, il a semblé opportun de ne pas intégrer à l’ACP l’eau (Water\_.g) et les graisses Saturées (FA\_Sat\_.g). L’eau contenue dans les aliments n’étant pas spécialement intéressante d’un point de vue nutritionnel. Les graisses saturées, même si indispensables à un équilibre nutritionnel, ne sont pas recherchées car source de « mauvais » cholesterol. Ces deux variables seront mises en variable illustratives. Les variables représentent la composition en un nutriment donné pour 100 grammes d’un individu donné.

###### VARIABLES CONTINUES ACTIVES

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 41 VARIABLES |  |  |  |  |  |  |  |
| -------------------------------------------------------------------------------------------------------------- | | | | | | |  |
| Nom | Type | Unité | Effectif | Moyenne | Ecart-type (N-1) | min | max |
| 2 . Energ\_Kcal | CONTINUE | en Kilo calorie | 214 | 120.3738318 | 127.7903964 | 11 | 718 |
| 3 . Protein\_.g. | CONTINUE | en grammes | 214 | 9.244579439 | 9.224298009 | 0.26 | 37.81 |
| 4 . Lipid\_Tot\_.g. | CONTINUE | en grammes | 214 | 4.92088785 | 11.41797712 | 0.03 | 81.11 |
| 5 . Ash\_.g. | CONTINUE | en grammes | 214 | 1.155233645 | 0.962505383 | 0.12 | 7.2 |
| 6 . Carbohydrt\_.g. | CONTINUE | en grammes | 214 | 10.54551402 | 16.84399249 | 0 | 79.95 |
| 7 . Fiber\_TD\_.g. | CONTINUE | en grammes | 214 | 2.228504673 | 3.571428168 | 0 | 25 |
| 8 . Sugar\_Tot\_.g. | CONTINUE | en grammes | 214 | 2.864205607 | 3.639905596 | 0 | 16.26 |
| 9 . Calcium\_.mg. | CONTINUE | en milligrammes | 214 | 47.79439252 | 83.73305878 | 1 | 674 |
| 10 . Iron\_.mg. | CONTINUE | en milligrammes | 214 | 1.324859813 | 1.949881356 | 0 | 15.7 |
| 11 . Magnesium\_.mg. | CONTINUE | en milligrammes | 214 | 33.80841121 | 48.95217293 | 1 | 429 |
| 12 . Phosphorus\_.mg. | CONTINUE | en milligrammes | 214 | 130.8878505 | 126.6303502 | 5 | 734 |
| 13 . Potassium\_.mg. | CONTINUE | en milligrammes | 214 | 317.4065421 | 288.900308 | 24 | 2515 |
| 14 . Sodium\_.mg. | CONTINUE | en milligrammes | 214 | 78.09813084 | 165.2613372 | 0 | 1146 |
| 15 . Zinc\_.mg. | CONTINUE | en milligrammes | 214 | 1.256962617 | 2.926077447 | 0.02 | 39.3 |
| 16 . Copper\_mg. | CONTINUE | en milligrammes | 214 | 0.189219626 | 0.370385275 | 0 | 2.92 |
| 17 . Manganese\_.mg. | CONTINUE | en milligrammes | 214 | 0.413060748 | 1.573043129 | 0 | 21.306 |
| 18 . Selenium\_.µg. | CONTINUE | en microgrammes | 214 | 11.48224299 | 15.90570402 | 0 | 90.6 |
| 19 . Vit\_C\_.mg. | CONTINUE | en milligrammes | 214 | 14.75654206 | 27.40622816 | 0 | 242.5 |
| 20 . Thiamin\_.mg. | CONTINUE | en milligrammes | 214 | 0.134060748 | 0.203429565 | 0.005 | 1.17 |
| 21 . Riboflavin\_.mg. | CONTINUE | en milligrammes | 214 | 0.14871028 | 0.194134966 | 0.01 | 1.778 |
| 22 . Niacin\_.mg. | CONTINUE | en milligrammes | 214 | 2.217 | 2.803925004 | 0.022 | 18.475 |
| 23 . Panto\_Acid\_mg. | CONTINUE | en milligrammes | 214 | 0.579275701 | 0.628389142 | 0.046 | 6.233 |
| 24 . Vit\_B6\_.mg. | CONTINUE | en milligrammes | 214 | 0.209799065 | 0.196243601 | 0.002 | 1.235 |
| 25 . Folate\_Tot\_.µg. | CONTINUE | en microgrammes | 214 | 52.03271028 | 108.8430573 | 0 | 633 |
| 27 . Food\_Folate\_.µg. | CONTINUE | en microgrammes | 214 | 50.99065421 | 108.1871093 | 0 | 633 |
| 28 . Folate\_DFE\_.µg. | CONTINUE | en microgrammes | 214 | 52.76168224 | 110.5562302 | 0 | 633 |
| 29 . Choline\_Tot\_..mg. | CONTINUE | en milligrammes | 214 | 40.20093458 | 46.3731578 | 0 | 293.8 |
| 30 . Vit\_B12\_.µg. | CONTINUE | en microgrammes | 214 | 1.108785047 | 2.684953977 | 0 | 20 |
| 31 . Vit\_A\_IU | CONTINUE | en unité internationale | 214 | 937.2943925 | 2885.535015 | 0 | 27521 |
| 32 . Retinol\_.µg. | CONTINUE | en microgrammes | 214 | 44.37383178 | 249.5233211 | 0 | 3290 |
| 33 . Alpha\_Carot\_.µg. | CONTINUE | en microgrammes | 214 | 45.1635514 | 365.0323625 | 0 | 4016 |
| 34 . Beta\_Carot\_.µg. | CONTINUE | en microgrammes | 214 | 442.7336449 | 1596.235433 | 0 | 16194 |
| 35 . Beta\_Crypt\_.µg. | CONTINUE | en microgrammes | 214 | 16.76635514 | 67.02386646 | 0 | 589 |
| 36 . Lycopene\_.µg. | CONTINUE | en microgrammes | 214 | 48.64485981 | 387.6612856 | 0 | 4532 |
| 37 . Lut.Zea\_..µg. | CONTINUE | en microgrammes | 214 | 420.4392523 | 1708.455304 | 0 | 13610 |
| 38 . Vit\_E\_.mg. | CONTINUE | en milligrammes | 214 | 0.674626168 | 1.056051909 | 0 | 8.93 |
| 39 . Vit\_D\_IU | CONTINUE | en unité internationale | 214 | 42.12616822 | 150.46291 | 0 | 1123 |
| 40 . Vit\_K\_.µg. | CONTINUE | en milligrammes | 214 | 26.95560748 | 97.8869437 | 0 | 830 |
| 42 . FA\_Mono\_.g. | CONTINUE | en grammes | 214 | 1.74078972 | 4.14235601 | 0 | 28.041 |
| 43 . FA\_Poly\_.g. | CONTINUE | en grammes | 214 | 0.77017757 | 1.940959667 | 0.003 | 15.558 |
| 44 . Cholestrl\_.mg. | CONTINUE | en milligrammes | 214 | 41.34579439 | 110.9740149 | 0 | 884 |

La variable Vit\_A\_IU était doublée de sa valeur en RAE(Retinol Activity Equivalent).

* 1 IU retinol = 0.3 mcg RAE [2]

Seule la valeur en unité internationale a été gardée.

La variable Vit\_D\_IU était doublée de sa valeur en microgrammes.

* 1000 IU = 25µg [3]

Seule la valeur en unité internationale a été gardée.

###### VARIABLES CONTINUES ILLUSTRATIVES

2 VARIABLES

----------------------------------------------------------------------------------------------------------------------------------

1 . Water\_.g. ( CONTINUE ) en grammes

41 . FA\_Sat\_.g. ( CONTINUE ) en grammes

Ces deux variables illustratives seront utilisées aussi pour créer deux variables qualitatives issues d’un découpage en classe. Le découpage en classe a été choisi pour qu’aucune classe n’ait d’effectif nul. La variable eau (Water\_g.) a été découpée en 6 classes et la variable graisses saturées (Fa\_Sat\_g.) en 5 classes.

La variable EdGrp\_Desc est la variable représentant les groupes de la classification naturelle des aliments.

###### VARIABLES NOMINALES ILLUSTRATIVES

3 VARIABLES 21 MODALITES ASSOCIEES

----------------------------------------------------------------------------------------------------------------------------------

45 . FdGrp\_Desc ( 10 MODALITES )

46 . Rec\_Water\_.g. ( 6 MODALITES )

47 . Rec\_FA\_Sat\_.g. ( 5 MODALITES )

EdGrp\_desc

|  |  |
| --- | --- |
| 1. | Beef Products 2. | Cereal Grains and Pasta 3. | Dairy and Egg Products 4. | Finfish and Shellfish Pr 5. | Fruits and Fruit Juices 6. | Lamb, Veal, and Game Pro 7. | Legumes and Legume Produ 8. | Pork Products 9. | Poultry Products 10. | Vegetables and Vegetable | C:\workspace\Nutrition100Food\04_rapport\graphiques\GalerieGraphiqueXYAvantSuppressionIndividus\Secteurs - Histogrammes.png |

Ce jeu de données fait la différence entre les légumineuses et les autres aliments végétaux. Les aliments végétaux (Vegetables), les poissons et coquillages (Finfish and shellfish), fruits et jus de fruits (Fruits and Fruit juices). Ces trois modalités constituent plus de 50% des données.

Water\_g.

|  |  |
| --- | --- |
| 1. < 17.700 2. [17.700;33.700[ 3. [33.700;49.700[ 4. [49.700;65.700[ 5. [65.700;81.700[ 6. >= 81.700 |  |

La plupart des données (170/206) ont une concentration en eau >=65.700g

Fa\_SAT\_g.

|  |  |
| --- | --- |
| 1. < 5.944 2. [5.944;11.883[ 3. [11.883;17.821[ 4. [17.821;23.760[ 5. >= 23.760 |  |

La majorité des aliments de ce jeu de données contiennent moins de 5.944 grammes de graisses saturées.

La variable cholestérol a été gardée comportant le mauvais cholestérol comme le bon.

### Présentation des individus

Le but de cette analyse est d’étudier les aliments principaux qui constituent notre alimentation. Une suppression empirique des catégories Beverages(boissons alcoolisées), Fats and Oils, Nut and seed Products, Sprices and Herbs, Sweets. Ces catégories ne sont pas consommées seules et ne présente donc pas d’intérêt pour cette étude.

Une première analyse grâce à une galerie de nuages de points montre des aliments présentant une concentration exceptionnelle en certains nutriments.

* Farine de soja (SOY FLOUR, Full-FAT,RAW) -> Protéines (Protein), Magnésium, Potassium
* Grains de soja (SOYBEANS,MATURE SEEDS,RAW) -> Protéines (Protein), Magnésium, Potassium
* Huitres (OYSTER,EASTERN) -> Zinc
* Pois chiches (CHICKPEAS) -> Manganèse
* Piment (PEPPERS,HOT CHILI) -> Vitamin C
* Foie de poulet (Chicken,LIVER) -> Acide Pantothenic (Panto Acide) vitamine B5 [4]
* Riz blanc long (RICE,WHITE,LONG-GRAIN) -> Acide Folique (FOLIC ACID) vitamine M / vitamine B9 [5]
* Potiron (PUMPKIN) -> Alpha carotene
* Carotte (CARROT,RAW) -> Alpha carotene
* Feuille de vigne (GRAPE,LEAVES,RAX) -> Beta carotene
* Pastèque (WATERMELON,RAW) -> Lycopène
* Tomate (TOMATOES) -> Lycopène
* Papaye (PAPAYAS) -> Lycopène
* Raison rouge (GRAPEFRUITS,PINK&RED) -> Lycopène
* Les œufs (EGGS) -> Cholesterol

# Analyse descriptive

### Statistiques de base

Les coefficients de corrélation R seront significativement différents de 0 si :

au seuil de

= 0.13

N = 214 >>30

A titre d’exemple en Annexe : A1. Matrice des corrélations de Pearson des 21 premières variables

Aussi bien la matrice des corrélations de Pearson ou de Spearman nous montrent des coefficients de corrélations plutôt significatif.

Nous appellerons par la suite :

* Rp : coefficient de Pearson,
* Rs : coefficient de Spearman,
* Rk : coefficient de Kendall.

La matrice des coefficients de Pearson nous montre que la variable Alpha carotène n’est significativement corrélée linéairement qu’à la vitamine A et au beta carotène

* Alpha carotène ; Vitamine A : Rp = 0.471, Rs =0.47706872, Rk = 0.37844203,
* Alpha carotène ; Beta carotene : Rp = 0.403, Rs = 0.5633994, Rk = 0.47725697.

Ce qui est attendu : les principales provitamines A sont l’alpha carotène et le beta carotène [2].

Le lycopène est quant à lui corrélé linéairement au béta carotène : Rp = 0.229, Rs = 0.21644976, Rk = 0.18583. Ceci s’explique car le lycopène est un important intermédiaire dans la biosynthèse du beta carotene [6].

En fichier annexe, vous pouvez trouver les différentes matrices des coefficients de corrélation (Pearson, Spearman, Kendall) et deux matrices supplémentaires permettant de :

* Déterminer quelles pourraient être les corrélations de Pearson significatives dues sans doute à des points atypiques,
* Déterminer si une relation autre que linéaire existe entre deux variables.

Fichiers obtenus à partir du script Nutrition100Food\02\_dataCleaning\create\_correlation\_matrix.R :

* correlations\_pearson.xlsx
* correlations\_kendall.xlsx
* correlations\_spearman.xlsx
* correlations\_due\_to\_atypic\_points.xlsx
* correlations\_non\_linear.xlsx

Ces fichiers sont utilisés uniquement à titres informatifs et une différence limite arbitraire de 0.40 a été choisie pour faciliter la lecture des Rs, Rk >> à Rp ou Rp >> Rs, Rk dans les tableaux de 41\*41 variables.

Ces tableaux ont donc été utiles pour déterminer la liste des individus contenant des doses exceptionnelles dans certains nutriments décrite dans la section «  Présentation des individus »

### ACP

Pour synthétiser les 41 variables continues actives, il a été choisi d’effectuer une analyse en composantes principales afin de faciliter la description des données.

L’ACP a été faite sur données normées : les variables sont d’unité différentes et présentent des ordres de grandeur différents. Le poids de tous les individus est uniforme.

Voici donc ce que donne l’ACP effectuée avec SPAD.

##### VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 41.0000

SOMME DES VALEURS PROPRES .... 41.0000

HISTOGRAMME DES 41 PREMIERES VALEURS PROPRES

+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+

| NUMERO | VALEUR | POURCENTAGE | POURCENTAGE | |

| | PROPRE | | CUMULE | |

+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+

| 1 | 10.2500 | 25.00 | 25.00 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 2 | 5.5743 | 13.60 | 38.60 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 3 | 3.2444 | 7.91 | 46.51 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 4 | 2.8446 | 6.94 | 53.45 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 5 | 2.0388 | 4.97 | 58.42 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 6 | 1.9284 | 4.70 | 63.12 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 7 | 1.4635 | 3.57 | 66.69 | \*\*\*\*\*\*\*\*\*\*\*\* |

| 8 | 1.3830 | 3.37 | 70.07 | \*\*\*\*\*\*\*\*\*\*\* |

| 9 | 1.2132 | 2.96 | 73.03 | \*\*\*\*\*\*\*\*\*\* |

| 10 | 1.0560 | 2.58 | 75.60 | \*\*\*\*\*\*\*\*\* |

| 11 | 1.0143 | 2.47 | 78.07 | \*\*\*\*\*\*\*\* |

| 12 | 0.9607 | 2.34 | 80.42 | \*\*\*\*\*\*\*\* |

| 13 | 0.9049 | 2.21 | 82.62 | \*\*\*\*\*\*\*\* |

| 14 | 0.7122 | 1.74 | 84.36 | \*\*\*\*\*\* |

| 15 | 0.7045 | 1.72 | 86.08 | \*\*\*\*\*\* |

| 16 | 0.6580 | 1.60 | 87.69 | \*\*\*\*\*\* |

| 17 | 0.6338 | 1.55 | 89.23 | \*\*\*\*\* |

| 18 | 0.5673 | 1.38 | 90.61 | \*\*\*\*\* |

| 19 | 0.5248 | 1.28 | 91.89 | \*\*\*\*\* |

| 20 | 0.4503 | 1.10 | 92.99 | \*\*\*\* |

| 21 | 0.4019 | 0.98 | 93.97 | \*\*\*\* |

| 22 | 0.3355 | 0.82 | 94.79 | \*\*\* |

| 23 | 0.2934 | 0.72 | 95.51 | \*\*\* |

| 24 | 0.2768 | 0.68 | 96.18 | \*\*\* |

| 25 | 0.2443 | 0.60 | 96.78 | \*\* |

| 26 | 0.2125 | 0.52 | 97.30 | \*\* |

| 27 | 0.1918 | 0.47 | 97.76 | \*\* |

| 28 | 0.1664 | 0.41 | 98.17 | \*\* |

| 29 | 0.1415 | 0.35 | 98.52 | \*\* |

| 30 | 0.1237 | 0.30 | 98.82 | \* |

| 31 | 0.1053 | 0.26 | 99.07 | \* |

| 32 | 0.0970 | 0.24 | 99.31 | \* |

| 33 | 0.0804 | 0.20 | 99.51 | \* |

| 34 | 0.0670 | 0.16 | 99.67 | \* |

| 35 | 0.0527 | 0.13 | 99.80 | \* |

| 36 | 0.0405 | 0.10 | 99.90 | \* |

| 37 | 0.0228 | 0.06 | 99.95 | \* |

| 38 | 0.0179 | 0.04 | 100.00 | \* |

| 39 | 0.0014 | 0.00 | 100.00 | \* |

| 40 | 0.0000 | 0.00 | 100.00 | \* |

| 41 | 0.0000 | 0.00 | 100.00 | \* |

+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+

Il y a 11 axes représentant plus d’une variable (valeur propre > 1) et représentant 78% de l’information. Néanmoins, on note une cassure après le premier axe, le deuxième et le quatrième. La question se pose de faire l’analyse avec 4 axes 53% de l’information est peu.

Voyons ce que donne la recherche des valeurs propres en supprimant les individus exceptionnels décrit en « Présentation des individus »

HISTOGRAMME DES 41 PREMIERES VALEURS PROPRES

+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+

| NUMERO | VALEUR | POURCENTAGE | POURCENTAGE | |

| | PROPRE | | CUMULE | |

+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+

| 1 | 10.0763 | 24.58 | 24.58 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 2 | 6.9934 | 17.06 | 41.63 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 3 | 3.8203 | 9.32 | 50.95 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 4 | 3.0824 | 7.52 | 58.47 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 5 | 2.1867 | 5.33 | 63.80 | \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* |

| 6 | 1.4858 | 3.62 | 67.43 | \*\*\*\*\*\*\*\*\*\*\*\* |

| 7 | 1.3638 | 3.33 | 70.75 | \*\*\*\*\*\*\*\*\*\*\* |

| 8 | 1.1269 | 2.75 | 73.50 | \*\*\*\*\*\*\*\*\* |

| 9 | 1.1120 | 2.71 | 76.21 | \*\*\*\*\*\*\*\*\* |

| 10 | 1.0136 | 2.47 | 78.69 | \*\*\*\*\*\*\*\*\* |

| 11 | 0.9864 | 2.41 | 81.09 | \*\*\*\*\*\*\*\* |

| 12 | 0.8339 | 2.03 | 83.13 | \*\*\*\*\*\*\* |

| 13 | 0.7320 | 1.79 | 84.91 | \*\*\*\*\*\* |

| 14 | 0.6577 | 1.60 | 86.52 | \*\*\*\*\*\* |

| 15 | 0.6158 | 1.50 | 88.02 | \*\*\*\*\* |

| 16 | 0.5682 | 1.39 | 89.40 | \*\*\*\*\* |

| 17 | 0.5591 | 1.36 | 90.77 | \*\*\*\*\* |

| 18 | 0.4583 | 1.12 | 91.88 | \*\*\*\* |

| 19 | 0.4311 | 1.05 | 92.94 | \*\*\*\* |

| 20 | 0.3912 | 0.95 | 93.89 | \*\*\*\* |

| 21 | 0.3388 | 0.83 | 94.72 | \*\*\* |

| 22 | 0.2834 | 0.69 | 95.41 | \*\*\* |

| 23 | 0.2773 | 0.68 | 96.08 | \*\*\* |

| 24 | 0.2505 | 0.61 | 96.70 | \*\* |

| 25 | 0.2051 | 0.50 | 97.20 | \*\* |

| 26 | 0.1871 | 0.46 | 97.65 | \*\* |

| 27 | 0.1842 | 0.45 | 98.10 | \*\* |

| 28 | 0.1350 | 0.33 | 98.43 | \*\* |

| 29 | 0.1227 | 0.30 | 98.73 | \* |

| 30 | 0.1122 | 0.27 | 99.00 | \* |

| 31 | 0.0997 | 0.24 | 99.25 | \* |

| 32 | 0.0935 | 0.23 | 99.47 | \* |

| 33 | 0.0677 | 0.17 | 99.64 | \* |

| 34 | 0.0543 | 0.13 | 99.77 | \* |

| 35 | 0.0406 | 0.10 | 99.87 | \* |

| 36 | 0.0369 | 0.09 | 99.96 | \* |

| 37 | 0.0148 | 0.04 | 100.00 | \* |

| 38 | 0.0012 | 0.00 | 100.00 | \* |

| 39 | 0.0000 | 0.00 | 100.00 | \* |

| 40 | 0.0000 | 0.00 | 100.00 | \* |

| 41 | 0.0000 | 0.00 | 100.00 | \* |

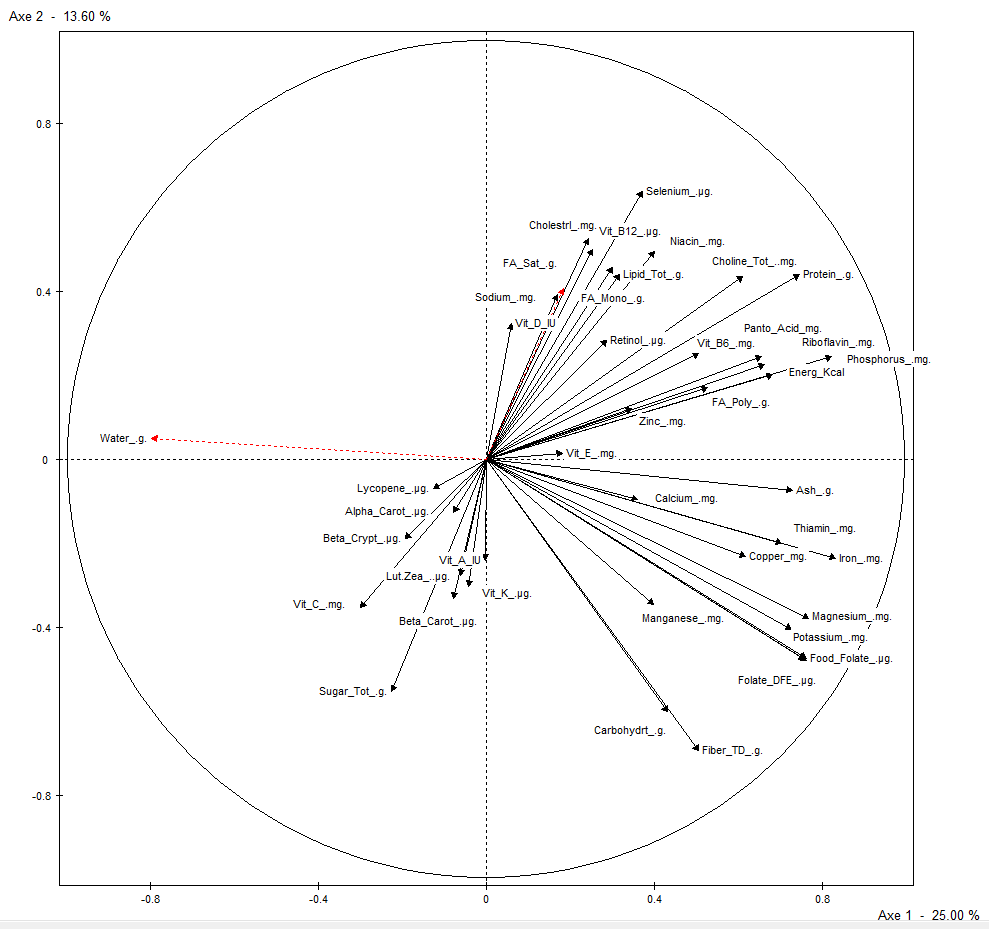
+--------+------------+-------------+-------------+----------------------------------------------------------------------------------+

Le comportement est semblable et il serait alors possible de faire l’analyse avec 5 axes et 63% de l’information.

Il ne nous semble pas opportun de supprimer ces individus. Ce sont des individus tout à fait habituels d’un point de vue alimentation et nous souhaitons qu’ils participent à l’ACP. Ceci signifie accepter de faire l’analyse avec 10% de l’information en moins alors qu’elle est assez faible (53%/63%). De plus, une fois ces individus retirés, d’autres apparaissent comme exceptionnels.

L’interprétation et l’analyse seront donc par la suite sur 4 axes sur la totalité des individus.

Axe 1 et Axe2



L’axe 1 est l’axe de taille. Les individus contribuant le plus à cet axe sont d’après leur distance à l’origine/contribution/cos² : foie de poulet (chicken liver) ; haricots noirs(BEANS,BLACK); haricot rouge de kidney(BEANS,KIDNEY); haricots blancs(BEANS,WHITE); fèves (BROADBEANS); pois chiches(CHICKPEAS); haricot long(COWPEAS); lentilles (LENTILS,RAW); haricot mungo(MUNG BNS); cacahuètes(PEANUTS); grains de soja(SOYBEANS); farine de soja(SOY FLOUR); orge(OAT BRAN); (cf. A4. Coordonnées, contributions et cosinus carrés des ).

Résumons que le foie de poulet d’une part et tout ce qui est légumineuses d’autre part contribuent le plus fortement à l’axe1 et sont les plus riches en nutriments autre que les nutriments liés à la vitamine A (caroténoïdes : β carotene, lutéine-zéaxanthine, β-Cryptoxanthin), Vitamine C, Vitamine K, Sucre. Le lycopene et l’α-carotene ne sont pas bien représentés sur ce plan 1-2.

Rappelons que les valeurs des variables sont les masses de quantités de nutriments pour 100grammes d’un aliment donné. Il n’est donc pas étonnant de voir l’eau en négatif sur cet axe de taille. En effet, sur 100 grammes d’un aliment donné, plus la quantité en eau est grande, moins il y a de nutriment. Ceci explique aussi les individus contribuant le plus à l’axe 1. Il s’agit d’aliment compacts (foie/grains). Les végétaux contribuant le plus à l’axe 1 sont les légumineux (végétaux à grain) faisant partie de la famille des Fabacées.

L’axe 2 sépare quant à lui tous les individus riches en nutriments propres aux aliments d’origine animales (cholestérol, énergie, protéines, graisses, sélénium, niacine (vitamine B3)) des autres nutriments : fibres, carbonhydrates. La superposition avec la variable qualitative illustrative « groupe d’aliments(Fd\_Grp\_desc) » dans les graphiques Figure 2 et Figure 3 le confirme. A noter l’individu cacahuètes (peanuts) qui se différencie de cette analyse. C’est sa teneur en niacine, graisses mono insaturées (FA\_Mono), lipides et protéines qui explique cette position sur ce plan 1-2. La niacine (vitamine B3) est principalement retrouvée dans les volailles. On peut observer sur le cercle des corrélations une corrélation inverse entre le faisceau sélénium et le sucre et vitamine C. Ceci est cohérent avec l’opposition sur le plan 1-2 viandes/légumes et viandes/fruits visible aussi sur le graphique des individus actifs. Les fruits (riches en vitamine C et sucres) vont se retrouver en bas à gauche du graphique des individus sur le plan 1-2. Les légumes un peu plus proches de l’origine sur l’axe 1.

Le foie de poulet (en haut à droite) est remarquable sur ce plan 1-2 de par sa richesse en nutriment (fer, selenium, choline, cuivre) très bien représentés sur l’axe 1 et 2.

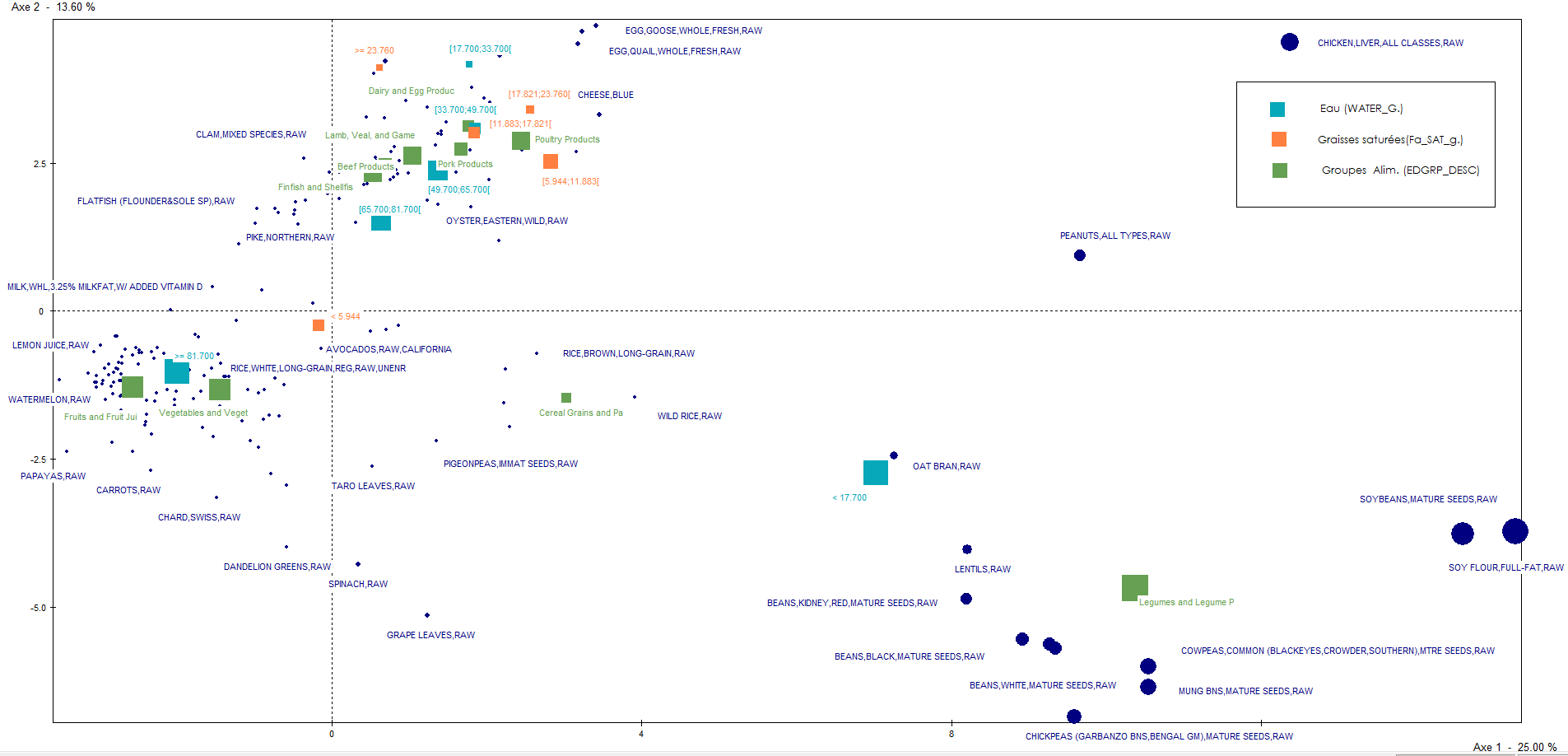


Figure 2 Visualisation des individus sur le plan 1-2 : Tailles de points des individus proportionnelle à leur contribution

Les individus légumineuses(Legumes, fabacées) précédemment remarqués contribuent fortement. L’ACP sans ces individus révèlent d’autres individus appartenant aussi au groupe des légumineuses. Ce groupe d’individus semble être le plus riche en nutriment. On note là aussi une juxtaposition de l’opposition des produits riches et pauvres en haut sur l’axe 1. On constate aussi la présence des aliments avec la teneur en graisses saturées la plus forte en positif sur l’axe 2 et assez éloignée de l’origine.

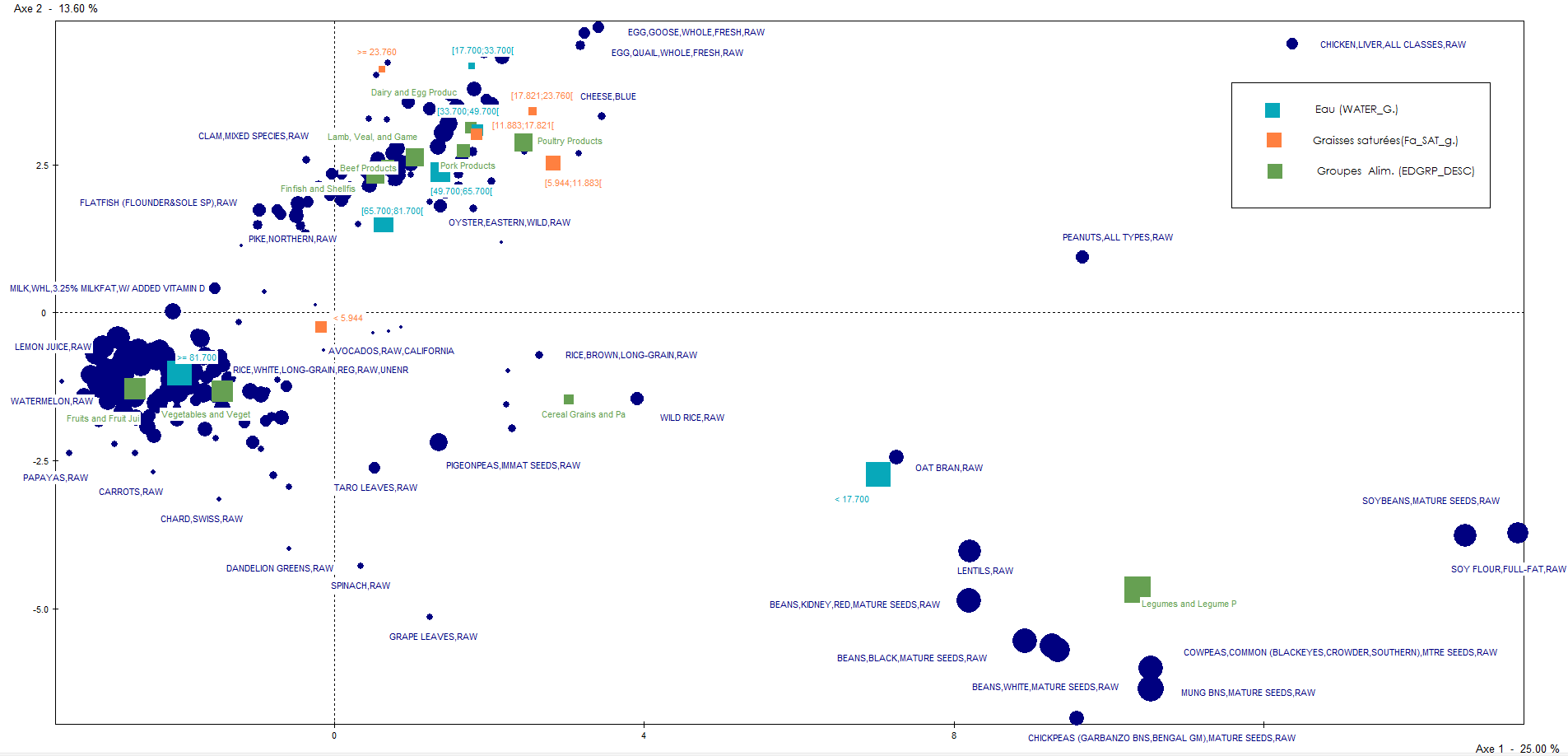
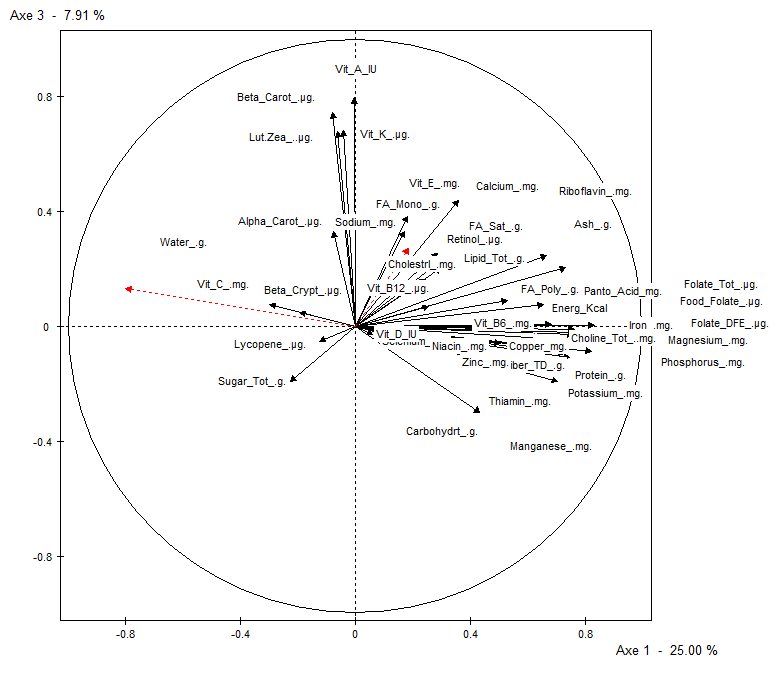


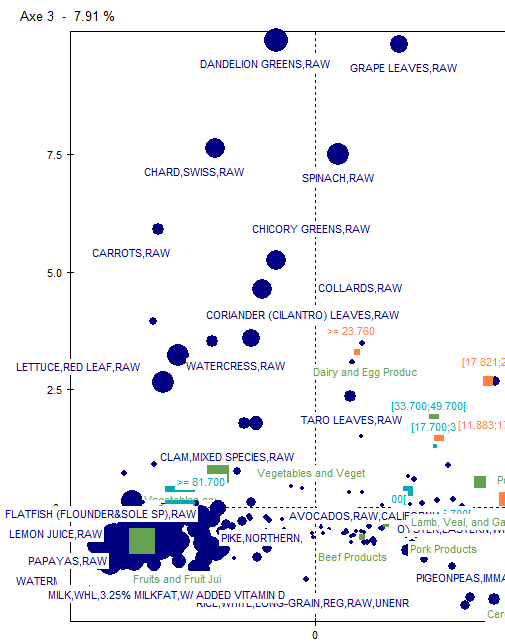
Figure 3 Visualisation des individus sur le plan 1-2 : Tailles des points représentant les individus proportionnelles à leur cos²

Sur l’axe 2, les aliments d’origine animale sont en haut du graphique alors qu’à l’opposé on retrouve nos légumineuses et aussi les autres aliments végétaux qui sont aussi bien représentés d’après leur distance à l’origine et leur cos².

Axe 3

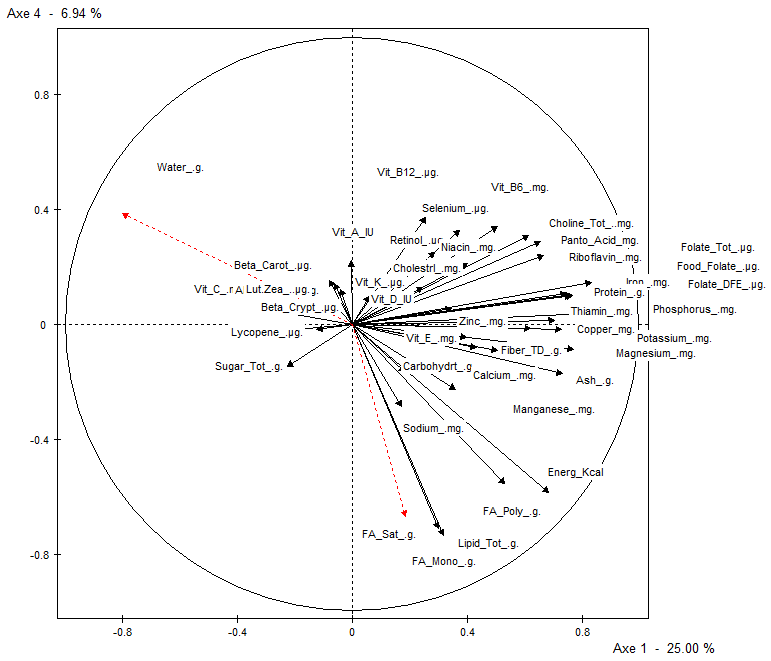


L’axe 3 représente très bien les variables Vitamine A, β-carotene, lutéine-zéaxanthine (Lut-Zea), vitamine K. Ce sont des nutriments que l’on retrouve ensemble dans les aliments de type feuille. Le graphique des individus actif nous montre une contribution.



On peut citer la feuille de pissenlit (Dandelion greens), les feuilles de vigne (Grape leaves), la blette (swiss chard), les épinards (spinash), chicorée (chicory greens), chou vert (collards), coriandre feuille (coriander leaves), cresson (watercress), laitue (lettuce) mais aussi carottes(carrots).

Axe4



Sur l’axe 4 se distingue un faisceau constitué des graisses mono insaturées, lipides, et un autre de graisses poly insaturées et énergie calorifique dans la partie négative de l’axe. La variable illustrative vient se juxtaposer sur ce faisceau. Cette axe va déterminer surtout les aliments qui sont surtout gras. La partie positive de l’axe se distingue surtout par la vitamine B12, B6, sélénium, niacyne (Vitamine B3), choline, panto-acide (vitamine B5). La partie supérieure de l’axe 4 concerne donc principalement les vitamines B.

Grâce au graphique des individus actifs et au tableau en annexe « A5. Coordonnées des individus actifs contribuant fortement à l’axe 4 », on peut identifier les aliments beurre salé et non salé (BUTTER W/ SALT et BUTTER WITH SALT), graisse de porc, peau de poulet, viande de canard et les fromages de vache. C’est principalement des aliments gras sans beaucoup d’autre nutriments. On note aussi la présence des cacahuètes (PEANUTS) bien représentées sur l’axe 4 (cf. annexe A5) dans la partie négative c'est-à-dire riche en gras. Leur bonne représentation sur l’axe 1 les différencie des aliments uniquement gras par une plus grande présence d’autres nutriments (Protéines, Magnésium, phosphore) comme le montre l’extrait Excel ci-dessous mais aussi en nyacine (12.066 mg)





Le foie de poulet est encore très bien représenté sur l’axe 4 avec 17% de contribution et un cos² de 0.207.

Sur 4 axes étudiés, le foie de poulet a une contribution et un cos² assez importants sur 3 axes (1, 2, 4). Il aurait peut-être été intéressant de le mettre en tant qu’individu illustratif. Sur les autres axes, le foie de poulet s’illustre sur les axes 5 et 10 grâce notamment à sa forte teneur en rétinol. Il a été décidé de le garder actif dans cette analyse.

### Classification

La classification ascendante hiérarchique a été faite sur les 11 axes créés à partir d’une valeur propre > 1 pour ne pas perdre trop d’information.

SUR LES 11 PREMIERS AXES FACTORIELS

DESCRIPTION DES 50 NOEUDS D'INDICES LES PLUS ELEVES

NUM. AINE BENJ EFF. POIDS INDICE HISTOGRAMME DES INDICES DE NIVEAU

378 283 9 4 4.00 0.03384 \*

379 361 342 14 14.00 0.03387 \*

380 370 358 7 7.00 0.03848 \*

381 191 236 3 3.00 0.03909 \*

382 374 301 47 47.00 0.03979 \*

383 348 346 21 21.00 0.04080 \*

384 379 316 17 17.00 0.04194 \*

385 355 193 5 5.00 0.04355 \*

386 350 353 5 5.00 0.04360 \*

387 371 375 34 34.00 0.04578 \*

388 380 354 9 9.00 0.04843 \*

389 376 132 3 3.00 0.04906 \*

390 383 100 22 22.00 0.04966 \*

391 363 152 3 3.00 0.05432 \*

392 334 67 3 3.00 0.05965 \*

393 357 5 5 5.00 0.06187 \*

394 366 331 9 9.00 0.06507 \*

395 368 367 7 7.00 0.06681 \*

396 385 210 6 6.00 0.07588 \*

397 362 117 3 3.00 0.08923 \*

398 384 372 23 23.00 0.08954 \*

399 393 129 6 6.00 0.09396 \*

400 387 220 36 36.00 0.10156 \*\*

401 394 66 10 10.00 0.10656 \*\*

402 388 326 11 11.00 0.10689 \*\*

403 395 171 8 8.00 0.10918 \*\*

404 396 381 9 9.00 0.11640 \*\*

405 392 266 5 5.00 0.13433 \*\*

406 390 340 27 27.00 0.15677 \*\*

407 397 138 4 4.00 0.18623 \*\*

408 398 373 28 28.00 0.18915 \*\*

409 403 391 11 11.00 0.21367 \*\*\*

410 382 400 83 83.00 0.21906 \*\*\*

411 401 389 13 13.00 0.25092 \*\*\*

412 351 195 3 3.00 0.40458 \*\*\*\*\*

413 406 386 32 32.00 0.42211 \*\*\*\*\*

414 409 408 39 39.00 0.46953 \*\*\*\*\*

415 377 402 14 14.00 0.60459 \*\*\*\*\*\*\*

416 414 413 71 71.00 0.62550 \*\*\*\*\*\*\*

417 410 411 96 96.00 0.79550 \*\*\*\*\*\*\*\*\*

418 404 412 12 12.00 0.86634 \*\*\*\*\*\*\*\*\*

419 416 181 72 72.00 0.93849 \*\*\*\*\*\*\*\*\*\*

420 415 407 18 18.00 1.13499 \*\*\*\*\*\*\*\*\*\*\*\*

421 419 378 76 76.00 1.16102 \*\*\*\*\*\*\*\*\*\*\*\*

422 399 405 11 11.00 1.17293 \*\*\*\*\*\*\*\*\*\*\*\*

423 418 14 13 13.00 1.94359 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

424 421 422 87 87.00 2.12389 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

425 417 420 114 114.00 2.15759 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

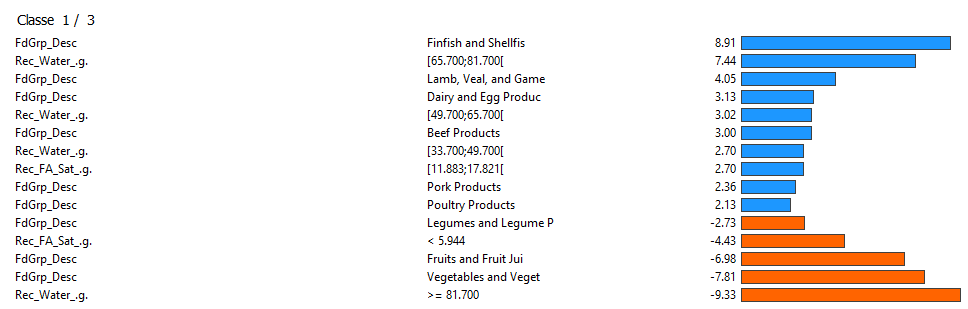
426 425 424 201 201.00 5.22097 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

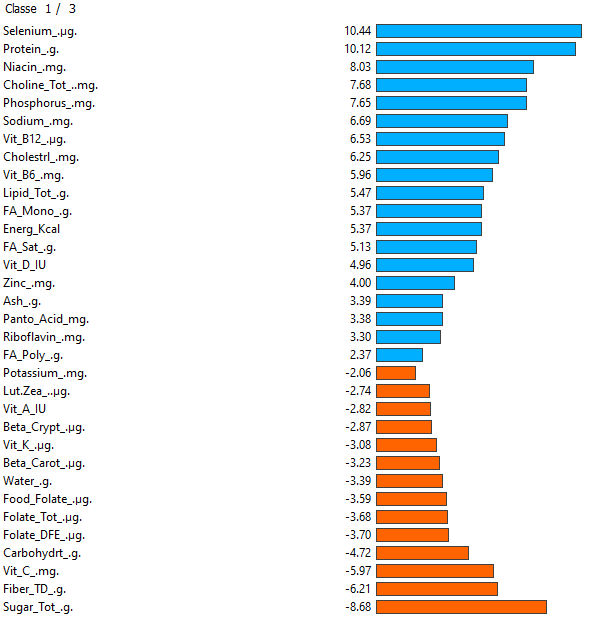
427 423 426 214 214.00 7.89576 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Le nombre de classe sera de 3 ou 6 classes. L’inertie inter-classe est de 13.17 dans le cas de 3 classes et de 19.65 dans le cas de 6 classes. Le partitionnement en 6 classes isole un aliment. C’est un aliment particulier riche en nutriments : le foie de poulet. Nous choisissons le partitionnement en 3 classes. Les classes restent nettes sur tous les axes.

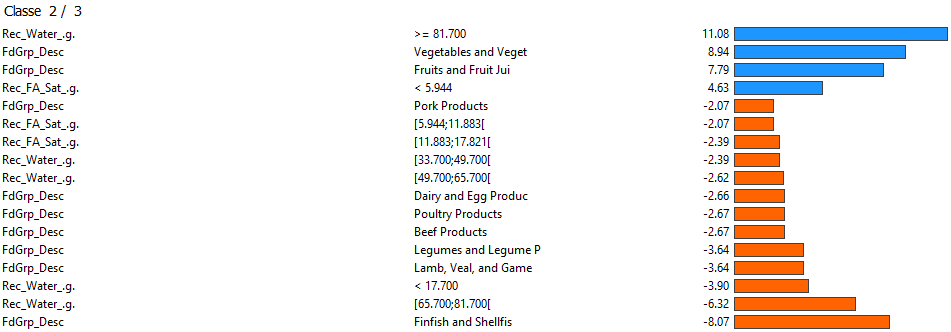
Le dendrogramme n’apparait pas dans ce rapport du fait du grand nombre d’individu et n’apporte ici pas d’élément visuel particulier.

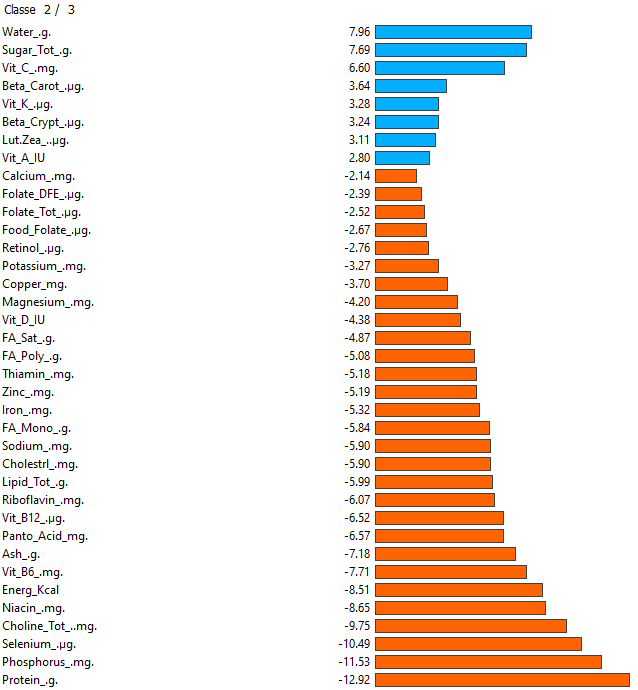
### Caractérisations des classes



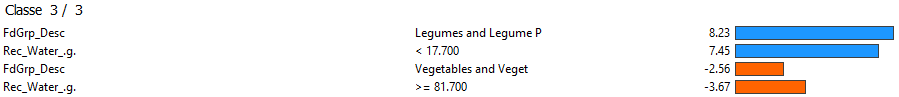


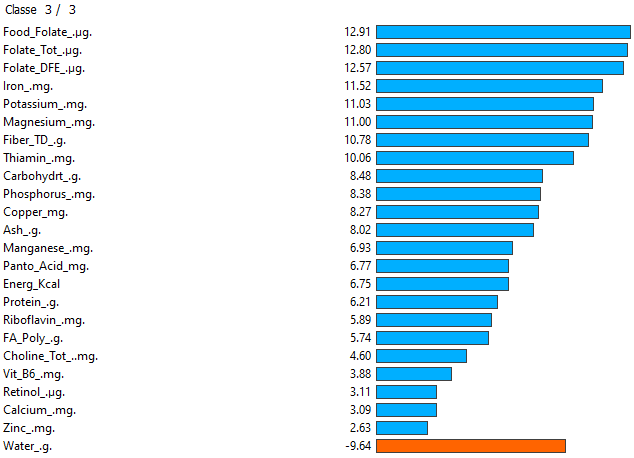
La classe 1 est caractérisée par tous les aliments d’origine animale (finfish, lamb, veal, dairy and egg, beef, pork, poutlry) riches en sélénium, protéine, vitamines B (niacin, choline, vitamin b12, vitamine b6, panto acide), graisses, phosphore, sodium.



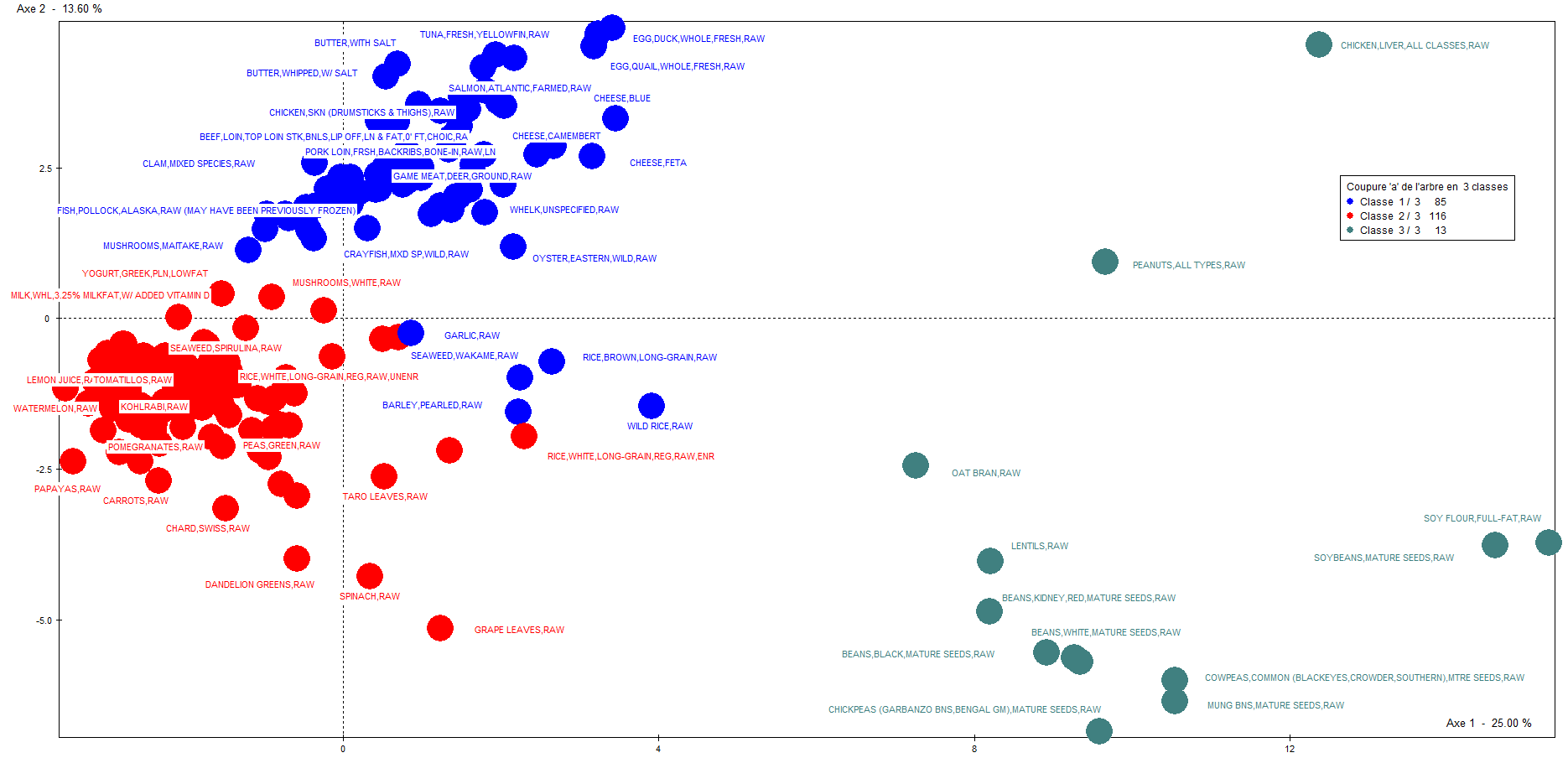


La classe 2 est caractérisée par tous les aliments qui ont beaucoup d’eau, qui appartiennent aux légumes végétaux ou aux fruits. Ce sont des aliments peu riches en graisses saturées. Ils peuvent être par contre riches en vitamine C, sucres, β-carotene, vitamine K, β-cryptoxanthin, vitamine K, lutéine zéaxanthine, vitamine A. Ce sont les aliments fruits et légumes non légumineux (qui n’appartiennent pas à la famille des graines de fabacées).





La classe 3 est caractérisées par les légumineuses, aliments qui n’ont pas beaucoup d’eau et riches en nutriments (folates, fer, potassium, magnesium, fibres, thiamine, carbohydrates, phosphore, cuivre, cendres, manganese, acide pantothénique, energétiques, protéines, riboflavine, graisses poly insaturées, choline, vitamine B6, rétinol, calcium, zinc). Le foie de poulet a aussi une certaine importance dans la caractérisation de cette classe.



Les classes sont assez nettes et ceci sur les 4 axes interprétés (ici 1-2). Ceci dit c’est surtout l’axe1 qui sépare les classes. Les aliments végétaux se trouvant aux frontières entre la classe 1 et la classe 2 sont des végétaux riches en protéines, phosphore, vitamine B6, cendres ou sodium qui sont des nutriments plutôt trouvés dans les aliments d’origine animale. Le cas des champignons est aussi intéressant pour l’anecdote : il y a de champignons dans la classe 1 comme dans la classe 2. Dans la classification phylogénétique, les champignons sont ne font ni partie du règne des végétaux (plantae) ni de celui des animaux (métazoaires). C’est un règne à part (fungi).

# Conclusion et perspectives

L’analyse de données a permis de classifier les aliments en trois grandes classes : aliments faibles en nutriments par rapport à leur masse de 100g, aliments avec une teneur en nutriment moyenne, et les aliments avec une proportion élevée en nutriments. Ces classes se juxtaposent assez facilement avec la classification phylogénétique usuelle mais grossière : végétaux (fruits et légumes), viandes, graines de légumineuses.

Le choix de garder un individu tel que le foie de poulet semble discutable car cet aliment est remarquable de par sa richesse en nutriments.

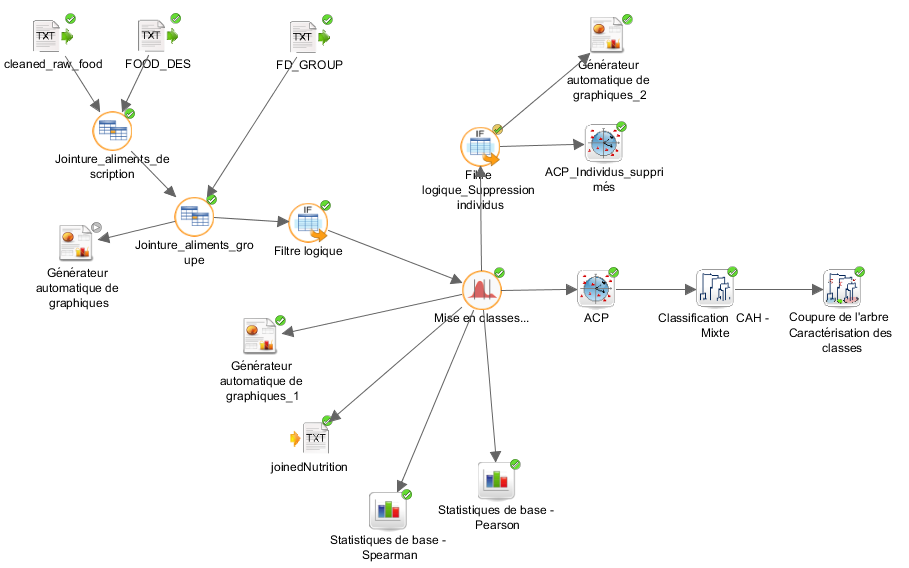
De manière générale, il semble que le sujet sous-jacent aux données analysées mérite des connaissances dans le domaine afin d’aider à l’interprétation.

# Annexes

### A1. Matrice des corrélations de Pearson des 21 premières variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables** | **Energ\_Kcal** | **Protein\_.g.** | **Lipid\_Tot\_.g.** | **Ash\_.g.** | **Carbohydrt\_.g.** | **Fiber\_TD\_.g.** | **Sugar\_Tot\_.g.** | **Calcium\_.mg.** | **Iron\_.mg.** | **Magnesium\_.mg.** | **Phosphorus\_.mg.** | **Potassium\_.mg.** | **Sodium\_.mg.** | **Zinc\_.mg.** | **Copper\_mg.** | **Manganese\_.mg.** | **Selenium\_.µg.** | **Vit\_C\_.mg.** | **Thiamin\_.mg.** | **Riboflavin\_.mg.** | **Niacin\_.mg.** |
| **Energ\_Kcal** | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Protein\_.g.** | 0.535 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Lipid\_Tot\_.g.** | 0.806 | 0.271 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Ash\_.g.** | 0.491 | 0.526 | 0.261 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Carbohydrt\_.g.** | 0.386 | 0.026 | -0.143 | 0.276 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Fiber\_TD\_.g.** | 0.283 | 0.067 | -0.100 | 0.381 | 0.745 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Sugar\_Tot\_.g.** | -0.191 | -0.445 | -0.222 | -0.207 | 0.263 | 0.266 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Calcium\_.mg.** | 0.277 | 0.222 | 0.218 | 0.666 | 0.087 | 0.191 | -0.059 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Iron\_.mg.** | 0.427 | 0.499 | 0.074 | 0.561 | 0.455 | 0.545 | -0.072 | 0.255 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |
| **Magnesium\_.mg.** | 0.436 | 0.454 | 0.067 | 0.642 | 0.553 | 0.665 | 0.024 | 0.307 | 0.708 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |
| **Phosphorus\_.mg.** | 0.560 | 0.859 | 0.233 | 0.660 | 0.264 | 0.280 | -0.356 | 0.370 | 0.593 | 0.631 | 1.000 |  |  |  |  |  |  |  |  |  |  |
| **Potassium\_.mg.** | 0.372 | 0.446 | 0.023 | 0.566 | 0.465 | 0.643 | 0.022 | 0.235 | 0.695 | 0.834 | 0.522 | 1.000 |  |  |  |  |  |  |  |  |  |
| **Sodium\_.mg.** | 0.308 | 0.240 | 0.419 | 0.600 | -0.209 | -0.239 | -0.281 | 0.552 | -0.039 | -0.040 | 0.286 | -0.170 | 1.000 |  |  |  |  |  |  |  |  |
| **Zinc\_.mg.** | 0.209 | 0.313 | 0.098 | 0.184 | 0.070 | 0.056 | -0.173 | 0.112 | 0.349 | 0.183 | 0.289 | 0.150 | 0.070 | 1.000 |  |  |  |  |  |  |  |
| **Copper\_mg.** | 0.299 | 0.337 | 0.048 | 0.428 | 0.340 | 0.412 | 0.015 | 0.198 | 0.616 | 0.668 | 0.393 | 0.617 | -0.039 | 0.606 | 1.000 |  |  |  |  |  |  |
| **Manganese\_.mg.** | 0.248 | 0.144 | 0.001 | 0.287 | 0.445 | 0.423 | 0.147 | 0.108 | 0.325 | 0.346 | 0.244 | 0.280 | -0.051 | 0.082 | 0.246 | 1.000 |  |  |  |  |  |
| **Selenium\_.µg.** | 0.205 | 0.668 | 0.103 | 0.203 | -0.170 | -0.234 | -0.485 | -0.014 | 0.204 | 0.085 | 0.559 | 0.023 | 0.237 | 0.184 | 0.127 | -0.029 | 1.000 |  |  |  |  |
| **Vit\_C\_.mg.** | -0.304 | -0.406 | -0.206 | -0.214 | -0.020 | 0.036 | 0.266 | -0.063 | -0.126 | -0.119 | -0.360 | -0.066 | -0.199 | -0.170 | -0.112 | -0.042 | -0.337 | 1.000 |  |  |  |
| **Thiamin\_.mg.** | 0.447 | 0.477 | 0.096 | 0.421 | 0.500 | 0.508 | -0.071 | 0.104 | 0.622 | 0.605 | 0.596 | 0.587 | -0.116 | 0.166 | 0.378 | 0.330 | 0.182 | -0.139 | 1.000 |  |  |
| **Riboflavin\_.mg.** | 0.319 | 0.504 | 0.222 | 0.456 | -0.010 | 0.067 | -0.193 | 0.362 | 0.539 | 0.379 | 0.527 | 0.347 | 0.228 | 0.207 | 0.366 | 0.098 | 0.307 | -0.165 | 0.349 | 1.000 |  |
| **Niacin\_.mg.** | 0.312 | 0.653 | 0.187 | 0.106 | -0.084 | -0.116 | -0.387 | -0.158 | 0.157 | 0.151 | 0.463 | 0.136 | -0.044 | 0.199 | 0.076 | -0.014 | 0.538 | -0.279 | 0.267 | 0.342 | 1.000 |
| **Panto\_Acid\_mg.** | 0.339 | 0.477 | 0.164 | 0.341 | 0.134 | 0.114 | -0.260 | 0.116 | 0.474 | 0.274 | 0.507 | 0.242 | 0.145 | 0.153 | 0.250 | 0.173 | 0.370 | -0.220 | 0.404 | 0.696 | 0.398 |
| **Vit\_B6\_.mg.** | 0.272 | 0.609 | 0.041 | 0.262 | 0.101 | 0.065 | -0.307 | 0.053 | 0.326 | 0.238 | 0.495 | 0.324 | -0.024 | 0.169 | 0.130 | 0.157 | 0.449 | -0.102 | 0.387 | 0.422 | 0.668 |
| **Folate\_Tot\_.µg.** | 0.377 | 0.309 | 0.005 | 0.551 | 0.569 | 0.679 | 0.098 | 0.245 | 0.756 | 0.637 | 0.416 | 0.673 | -0.054 | 0.135 | 0.481 | 0.463 | -0.056 | -0.039 | 0.580 | 0.422 | 0.044 |
| **Folic\_Acid\_.µg.** | 0.131 | -0.016 | -0.026 | -0.037 | 0.283 | -0.018 | -0.052 | -0.016 | 0.105 | -0.012 | -0.009 | -0.048 | -0.030 | -0.004 | 0.006 | 0.029 | 0.016 | -0.037 | 0.149 | -0.035 | 0.048 |
| **Food\_Folate\_.µg.** | 0.361 | 0.313 | 0.009 | 0.559 | 0.532 | 0.686 | 0.106 | 0.249 | 0.746 | 0.643 | 0.419 | 0.684 | -0.050 | 0.137 | 0.483 | 0.462 | -0.059 | -0.034 | 0.562 | 0.430 | 0.038 |
| **Folate\_DFE\_.µg.** | 0.384 | 0.303 | 0.003 | 0.539 | 0.587 | 0.667 | 0.092 | 0.240 | 0.754 | 0.626 | 0.408 | 0.658 | -0.056 | 0.133 | 0.474 | 0.459 | -0.054 | -0.042 | 0.585 | 0.412 | 0.048 |
| **Choline\_Tot\_..mg.** | 0.316 | 0.654 | 0.170 | 0.278 | -0.054 | -0.003 | -0.349 | 0.016 | 0.469 | 0.279 | 0.582 | 0.294 | 0.103 | 0.249 | 0.278 | 0.101 | 0.567 | -0.287 | 0.307 | 0.585 | 0.386 |
| **Vit\_B12\_.µg.** | 0.065 | 0.373 | 0.069 | 0.093 | -0.231 | -0.259 | -0.312 | -0.023 | 0.262 | -0.037 | 0.278 | -0.064 | 0.213 | 0.258 | 0.118 | -0.035 | 0.551 | -0.186 | -0.022 | 0.330 | 0.293 |
| **Vit\_A\_IU** | -0.101 | -0.164 | -0.030 | 0.038 | -0.045 | 0.092 | 0.010 | 0.237 | 0.073 | 0.023 | -0.129 | -0.006 | -0.009 | -0.072 | -0.010 | 0.048 | -0.124 | 0.082 | -0.071 | 0.171 | -0.068 |
| **Retinol\_.µg.** | 0.171 | 0.115 | 0.229 | 0.071 | -0.102 | -0.111 | -0.130 | 0.045 | 0.217 | -0.050 | 0.142 | -0.071 | 0.170 | 0.027 | 0.009 | -0.033 | 0.211 | -0.053 | 0.031 | 0.555 | 0.182 |
| **Alpha\_Carot\_.µg.** | -0.079 | -0.105 | -0.049 | -0.027 | -0.013 | 0.006 | 0.036 | 0.012 | -0.032 | -0.036 | -0.086 | 0.001 | -0.036 | -0.040 | -0.025 | -0.007 | -0.087 | -0.011 | -0.044 | -0.025 | -0.059 |
| **Beta\_Carot\_.µg.** | -0.151 | -0.198 | -0.096 | 0.026 | -0.016 | 0.134 | 0.041 | 0.243 | 0.017 | 0.046 | -0.171 | 0.017 | -0.056 | -0.081 | -0.009 | 0.064 | -0.187 | 0.099 | -0.080 | 0.017 | -0.121 |
| **Beta\_Crypt\_.µg.** | -0.139 | -0.216 | -0.089 | -0.150 | -0.006 | -0.013 | 0.278 | -0.050 | -0.096 | -0.088 | -0.203 | -0.080 | -0.100 | -0.091 | -0.083 | -0.047 | -0.167 | 0.329 | -0.093 | -0.098 | -0.142 |
| **Lycopene\_.µg.** | -0.088 | -0.116 | -0.052 | -0.105 | -0.022 | -0.044 | 0.097 | -0.054 | -0.069 | -0.056 | -0.114 | -0.069 | -0.057 | -0.049 | -0.048 | -0.029 | -0.087 | 0.033 | -0.062 | -0.077 | -0.084 |
| **Lut.Zea\_..µg.** | -0.160 | -0.174 | -0.089 | 0.053 | -0.061 | 0.041 | -0.085 | 0.164 | 0.044 | 0.052 | -0.141 | 0.049 | -0.019 | -0.069 | -0.024 | 0.010 | -0.153 | 0.158 | -0.045 | 0.011 | -0.131 |
| **Vit\_E\_.mg.** | 0.155 | 0.066 | 0.218 | 0.147 | -0.048 | 0.055 | -0.043 | 0.060 | 0.106 | 0.176 | 0.101 | 0.123 | 0.045 | 0.000 | 0.165 | 0.067 | 0.050 | 0.035 | 0.075 | 0.043 | 0.103 |
| **Vit\_D\_IU** | 0.038 | 0.220 | 0.057 | 0.012 | -0.160 | -0.148 | -0.196 | -0.078 | -0.083 | -0.028 | 0.216 | -0.016 | -0.013 | -0.033 | -0.083 | -0.065 | 0.217 | -0.145 | 0.031 | 0.016 | 0.289 |
| **Vit\_K\_.µg.** | -0.157 | -0.168 | -0.086 | 0.086 | -0.059 | 0.066 | -0.096 | 0.189 | 0.068 | 0.088 | -0.139 | 0.086 | -0.016 | -0.066 | 0.006 | 0.022 | -0.172 | 0.166 | -0.037 | 0.020 | -0.137 |
| **FA\_Mono\_.g.** | 0.759 | 0.278 | 0.952 | 0.181 | -0.162 | -0.100 | -0.239 | 0.128 | 0.062 | 0.041 | 0.214 | 0.011 | 0.291 | 0.100 | 0.028 | -0.016 | 0.116 | -0.208 | 0.104 | 0.177 | 0.253 |
| **FA\_Poly\_.g.** | 0.652 | 0.379 | 0.674 | 0.285 | 0.033 | 0.138 | -0.070 | 0.130 | 0.357 | 0.466 | 0.397 | 0.394 | 0.027 | 0.118 | 0.389 | 0.176 | 0.115 | -0.162 | 0.333 | 0.319 | 0.278 |
| **Cholestrl\_.mg.** | 0.221 | 0.265 | 0.298 | 0.064 | -0.216 | -0.234 | -0.262 | 0.038 | 0.160 | -0.099 | 0.234 | -0.116 | 0.229 | 0.082 | -0.011 | -0.083 | 0.408 | -0.186 | -0.002 | 0.414 | 0.092 |

### A2. Description des méthodes utilisées dans SPAD



* Jointure\_aliments\_description : sert à faire la jointure entre les fichiers de données sur les nutriments (cleaned\_raw\_food) et FOOD\_DES
* Jointure\_aliments\_groupe : sert à faire la jointure entre le groupe auquel appartient l’individu et l’individu contenu dans le tableau issu de Jointure\_aliments\_description
* Filtre logique est le filtre qui sert à supprimer les groupes d’aliments qui ne seront pas utilisés pour cette analyse
* ACP : Analyse en composante principale retenue pour cette analyse
* Filtre logique\_suppression individus : supprime les individus exceptionnels décrits en « Présentation des individus »
* ACP\_Individus\_supprimés : ACP effectuée avec ces individus supprimés. Meilleure qualité mais non préférée pour cette analyse. Laissée au lecteur pour information et reproductibilité.

Les autres éléments sont supposés familiers du lecteur.

### A3. Description des scripts R

Fichiers situés dans Nutrition100Food\02\_dataCleaning

* dataCleaning.Rproj : fichier projet R s’ouvrant avec RStudio pour le confort du lecteur
* clean.R : supprime 5 variables non utiles à cette analyse (GmWt\_1, GmWt\_Desc1, GmWt\_2, GmWt\_Desc2, Refuse\_Pct) qui décrivent la manière dont ont été faites les mesures ; supprime les individus ayant des valeurs manquantes ; supprime les variables non utilisées dans le fichier FOOD\_DES (pour l’analyse, seul l’identifiant du groupe d’aliment était important) ; écrit les fichiers résultants
* create\_correlation\_matrix.R : détermine les matrices de corrélation de Pearson, Kendall, Spearman, une matrice des corrélations pouvant être dues à des points atypiques, une matrice des corrélations pouvant être des relations non linéaires.

### A4. Coordonnées, contributions et cosinus carrés des individus

AXES 1 A 5

+---------------------------------------+-------------------------------+--------------------------+--------------------------+

| INDIVIDUS | COORDONNEES | CONTRIBUTIONS | COSINUS CARRES |

|---------------------------------------+-------------------------------+--------------------------+--------------------------|

| IDENTIFICATEUR P.REL DISTO | 1 2 3 4 5 | 1 2 3 4 5 | 1 2 3 4 5 |

+---------------------------------------+-------------------------------+--------------------------+--------------------------+

| CHICKEN,LIVER,ALL CLASSE 0.47 506.28 | 12.37 4.54 4.45 10.23 12.69 | 7.0 1.7 2.8 17.2 36.9 | 0.30 0.04 0.04 0.21 0.32 |

| BEANS,BLACK,MATURE SEEDS 0.47 126.30 | 8.92 -5.54 -2.08 -0.17 -0.57 | 3.6 2.6 0.6 0.0 0.1 | 0.63 0.24 0.03 0.00 0.00 |

| BEANS,KIDNEY,RED,MATURE 0.47 103.57 | 8.19 -4.86 -2.01 0.09 -0.57 | 3.1 2.0 0.6 0.0 0.1 | 0.65 0.23 0.04 0.00 0.00 |

| BEANS,WHITE,MATURE SEEDS 0.47 142.53 | 9.34 -5.69 -1.52 -0.24 -1.10 | 4.0 2.7 0.3 0.0 0.3 | 0.61 0.23 0.02 0.00 0.01 |

| BROADBEANS (FAVA BEANS), 0.47 138.71 | 9.27 -5.63 -2.21 0.16 -0.22 | 3.9 2.7 0.7 0.0 0.0 | 0.62 0.23 0.04 0.00 0.00 |

| CHICKPEAS (GARBANZO BNS, 0.47 293.19 | 9.59 -6.85 -2.28 -0.54 2.49 | 4.2 3.9 0.7 0.0 1.4 | 0.31 0.16 0.02 0.00 0.02 |

| COWPEAS,COMMON (BLACKEYE 0.47 169.04 | 10.54 -6.00 -2.10 0.79 1.17 | 5.1 3.0 0.6 0.1 0.3 | 0.66 0.21 0.03 0.00 0.01 |

| LENTILS,RAW 0.47 109.63 | 8.21 -4.02 -2.00 1.12 1.24 | 3.1 1.4 0.6 0.2 0.4 | 0.61 0.15 0.04 0.01 0.01 |

| MUNG BNS,MATURE SEEDS,RA 0.47 171.50 | 10.55 -6.34 -1.95 0.70 1.25 | 5.1 3.4 0.5 0.1 0.4 | 0.65 0.23 0.02 0.00 0.01 |

| PEANUTS,ALL TYPES,RAW 0.47 234.20 | 9.66 0.93 1.51 -7.65 -0.41 | 4.3 0.1 0.3 9.6 0.0 | 0.40 0.00 0.01 0.25 0.00 |

| SOYBEANS,MATURE SEEDS,RA 0.47 280.58 | 14.60 -3.76 -0.28 -2.44 -1.45 | 9.7 1.2 0.0 1.0 0.5 | 0.76 0.05 0.00 0.02 0.01 |

| SOY FLOUR,FULL-FAT,RAW 0.47 340.17 | 15.28 -3.71 -0.08 -2.39 -1.95 | 10.6 1.2 0.0 0.9 0.9 | 0.69 0.04 0.00 0.02 0.01 |

| OAT BRAN,RAW 0.47 122.13 | 7.26 -2.44 -2.08 -0.90 -2.60 | 2.4 0.5 0.6 0.1 1.5 | 0.43 0.05 0.04 0.01 0.06 |

### A5. Coordonnées des individus actifs contribuant fortement à l’axe 4



# Bibliographie

|  |  |
| --- | --- |
| [1] | United States Department of Agriculture, "National Nutrient Database for Standard Reference," *http://www.ars.usda.gov/Services/docs.htm?docid=8964.* |
| [2] | National Institudes of Health, "Vitamin A," *https://ods.od.nih.gov/factsheets/VitaminA-HealthProfessional/.* |
| [3] | National Institudes of Health, "Vitamin D," *https://ods.od.nih.gov/factsheets/VitaminD-HealthProfessional/.* |
| [4] | National Institutes of Health, "Vitamin B5," *https://www.nlm.nih.gov/medlineplus/druginfo/natural/853.html.* |
| [5] | Wikipédia, Communauté, "Acide Folique," *https://en.wikipedia.org/wiki/Folic\_acid.* |
| [6] | Wikipedia, Communauté, "Lycopene," *https://en.wikipedia.org/wiki/Lycopene.* |
| [7] | National Institutes of health, "DIETARY SUPPLEMENT FACT SHEETS," *https://ods.od.nih.gov/factsheets/list-all/.* |
| [8] | Wikipédia, Communauté, "Acides gras," *https://fr.wikipedia.org/wiki/Acide\_gras.* |
| [9] | Brigitte ESCOFIER & Jérôme PAGES, Analyses factorielles simples et multiples, DUNOD, 2002. |
| [10] | P.-L. GONZALEZ, "STA101 - Analyse des données : méthodes descriptives," *http://maths.cnam.fr/spip.php?article50,* 2015-2016. |
| [11] | François HUSSON, Sébastien LE, Jérôme PAGES, Analyse de données avec R, Presses Universitaires de Rennes, 2015. |

1. Si cette étude n’est sans doute pas nouvelle, elle aura tout de même pour intérêt d’intéresser un informaticien de la finance un peu curieux. [↑](#footnote-ref-1)