

Projet final (UAOM05)  
Licence Science des données (LG04201A)

Adam Benali (N.élève : 100227598)

## Contents

<b>1</b>	<b>Introduction.</b>	<b>4</b>
<b>2</b>	<b>La base de données.</b>	<b>6</b>
2.1	Importation de la base de données : AGRIBALYSE® - Synthèse. . . . .	6
2.2	Les unités statistiques. . . . .	6
2.3	Les variables. . . . .	6
2.4	Préparation du jeu de données. . . . .	6
<b>3</b>	<b>Analyse univariée.</b>	<b>7</b>
3.1	Variables quantitatives. . . . .	7
3.2	Variables qualitatives. . . . .	11
3.3	Conclusion de l'analyse univariée. . . . .	15
<b>4</b>	<b>Analyse bivariée.</b>	<b>16</b>
4.1	Croisement des variables quantitatives. . . . .	16
4.2	Croisement des variables qualitatives. . . . .	21
4.3	Croisement des variables quantitatives et qualitatives. . . . .	23
4.4	Conclusion de l'analyse bivariée. . . . .	25
<b>5</b>	<b>Analyse en composantes principales.</b>	<b>26</b>
5.1	Choix des variables. . . . .	26
5.2	Valeurs propres et choix du nombre d'axes. . . . .	26
5.3	Analyse des individus. . . . .	27
5.4	Analyse des variables. . . . .	27
5.5	Interprétation des axes. . . . .	28
<b>6</b>	<b>Classification.</b>	<b>30</b>
6.1	Pré-traitement. . . . .	30
6.2	Méthode Mixte. . . . .	30
6.3	Analyse. . . . .	31

<b>7 Conclusion.</b>	<b>33</b>
<b>8 Appendice : préparation des données.</b>	<b>34</b>
8.1 Unités statistiques. . . . .	34
8.2 Variables. . . . .	34
<b>9 Sources.</b>	<b>36</b>
9.1 Sitographie. . . . .	36
<b>10 Annexes.</b>	<b>37</b>
10.1 Annexe 1 : variables qualitatives. . . . .	37
10.2 Annexe 2 : variables quantitatives. . . . .	38
10.3 Annexe 3 : statistiques descriptives. . . . .	41
10.4 Annexe 4 : box-plots. . . . .	43
10.5 Annexe 5 : histogramme des variables de pollution. . . . .	47
10.6 Annexe 6 : histogramme et droite de Henry des transformations logarithmiques. . . . .	51
10.7 Annexe 7 : histogramme et droite de Henry des transformations de Box-Cox. . . . .	53
10.8 Annexe 8 : tests de Shapiro-Wilks. . . . .	58
10.9 Annexe 9 : nombre de modalités. . . . .	60
10.10 Annexe 10 : tableaux des fréquences. . . . .	61
10.11 Annexe 11 : pollution des modalités rares. . . . .	63
10.12 Annexe 12 : diagramme en barre des variables qualitatives. . . . .	66
10.13 Annexe 13 : nuages de points des corrélations. . . . .	67
10.14 Annexe 14 : coefficients de corrélation de Spearman. . . . .	69
10.15 Annexe 15 : tableau des variables quantitatives des produits 52 et 656 . . . . .	70
10.16 Annexe 16 : image des coefficients de pondération pour le “Score EF”. . . . .	72
10.17 Annexe 17 : image des coefficient de Normalisation “Score EF”. . . . .	73
10.18 Annexe 18 : tableau des coefficients. . . . .	74
10.19 Annexe 19 : table des variables quantitatives des produits 51 et 52. . . . .	75
10.20 Annexe 20 : tableau de contingence. . . . .	77
10.21 Annexe 21 : tableaux des valeurs théoriques pour le test du khi-deux. . . . .	78
10.22 Annexe 22 : test du khi-deux. . . . .	79
10.23 Annexe 23 : tableaux des valeurs théoriques du test du khi-deux. . . . .	81
10.24 Annexe 24 : test du khi-deux avec “simulate p-value”. . . . .	83
10.25 Annexe 25 : matrice des corrélations entre variables qualitatives et quantitatives. . . . .	84
10.26 Annexe 26 : valeurs propres de l’ACP. . . . .	86
10.27 Annexe 27 : contribution et $\cos^2$ . . . . .	88
10.28 Annexe 28 : coordonnées axe 1. . . . .	89

10.29	Annexe 29 : coordonnées axe 2. . . . .	90
10.30	Annexe 30 : graphe des individus selon le groupe d'aliments. . . . .	91
10.31	Annexe 31 : $\cos^2$ et corrélation. . . . .	92
10.32	Annexe 32 : classe 1. . . . .	93
10.33	Annexe 33 : classe 2. . . . .	94
10.34	Annexe 34 : classe 3. . . . .	95

# 1 Introduction.

L'impact du secteur agricole est catastrophique sur l'environnement selon Arnaud Gauffier, responsable de l'agriculture chez WWF France, « *À elle seule, l'agriculture occupe environ un tiers de la surface terrestre totale, est la cause de 80 % de la déforestation mondiale et pèse pour près de 70 % de la consommation d'eau* » et « *30 % de l'empreinte carbone en France est liée à l'alimentation.* » (Garric, 2016). L'élevage est responsable, par exemple, selon le modèle de calcul, de 14.5 % des émissions de gaz à effet de serre. La surpêche, quant à elle, diminue, entre autres, l'abondance des populations de poissons, réduisant la diversité des espèces : « *D'après une enquête menée par le Fonds Mondial pour la nature, ainsi que WWF, 50% des espèces d'animaux marins auraient disparu en 40 ans à cause de l'activité de l'homme* » (Marine Stewardship Council (MSC), s.d.).

L'Agence de l'Environnement et de la Maîtrise de l'Énergie<sup>1</sup> est chargée de la transition écologique de la France vers la neutralité carbone à horizon 2050. L'ADEME a, entre autre, un rôle de recherche et de diffusion de l'information dans différents domaines liés à l'environnement. Depuis plusieurs décennies, maintenant, la sensibilisation de la population aux impacts environnementaux des systèmes de productions est un levier de la transition écologique. Dans un rapport de la Commission européenne sur le comportement des consommateurs européens, 67 % des consommateurs ont déclaré qu'ils achetaient des produits plus respectueux de l'environnement même s'il coûte plus cher (Commission européenne). C'est pourquoi, l'ADEME met à disposition du public ses données.

Nous nous sommes procurés la base de données « AGRIBALYSE® - Synthèse » (ADEME, 2022) concernant l'impact environnemental des produits alimentaires vendus en France. L'AGRIBALYSE référence l'impact environnemental d'un produit fondé sur la méthode de l'Analyse du Cycle de la Vie <sup>2</sup>.

L'objectif de cette base de données s'inscrit dans la recherche et dans la diffusion de l'information. En donnant une note de l'impact environnemental d'un produit, L'ADEME espère toucher la sensibilité du consommateur sur l'impact environnemental d'un produit alimentaire, à l'instar du nutri-score.

Un produit n'est pas l'estimation de l'impact d'une marque spécifique, mais l'estimation de l'impact d'un produit moyen. L'exemple donné par AGRIBALYSE pour illustrer cela est le suivant : « l'impact d'une pizza Margherita « standard », constituée de tomates « standards » conventionnelles, de gruyère et de jambon standards « conventionnels », issus des systèmes de production majoritaires aujourd'hui, ainsi que de l'emballage majoritaire observé pour ce type de produit (c'est-à-dire du carton).» (Agribalyse, 2022). Le tableau de résultat synthétise par un produit moyen la pollution de produits alimentaires.

Nous allons essayer de décrire au mieux la base de données en s'organisant autour de la problématique suivante, quelles informations intéressantes peut-on extraire de cette base de données ? Pour cela, nous nous poserons plusieurs questions sur ce jeu de données : quelle est la tendance des variables ? La dispersion ? Comment nos unités statistiques se distribuent-elles sur nos variables ? Nos variables, suivent-elles une loi statistique ? Existe-t-il des liens entre nos variables ? Comment le score de pollution est-il construit pour résumer l'information ? Quelle est la nature des liaisons entre nos variables ? Peut-on trouver des produits qui se ressemblent ? Et sur quels critères peut-on dire qu'ils se ressemblent ? Peut-on distinguer des produits selon le type de pollution qu'ils émettent ? Peut-on faire des groupes de produits selon les types de pollution qu'ils émettent ?

Après avoir présenté et mis en forme notre jeu de données, nous décrirons, dans un premier temps, les variables selon les indicateurs statistiques de tendances et de dispersion. Nous regarderons les fréquences des variables qualitatives et la distribution des variables quantitatives. En somme, nous ferons une analyse univariée de notre jeu de données.

Dans un second temps, nous chercherons des liens entre variables quantitatives et entre variables qualitatives, mais aussi entre variables quantitatives et qualitatives, nous ferons une analyse bivariée. De plus, nous chercherons à savoir comment est construit la variable "Score EF" par rapport aux autres variables.

<sup>1</sup> ADEME : établissement public créé en 1991, affilié au ministère de la Transition écologique et solidaire.

<sup>2</sup> L'ACV essaye de prendre en compte, dans la quantification, des impacts d'un produit sur l'environnement : la production agricole, la transformation, le transport et la logistique, l'emballage, la préparation, et le recyclage du produit. Ce qu'on appelle du champ à l'assiette. (Agribalyse, 2022)

Dans un troisième temps, nous essayerons de visualiser au mieux nos données en effectuant une analyse en composante principale. Nous chercherons à trouver une classification de nos unités statistiques.

## 2 La base de données.

### 2.1 Importation de la base de données : AGRIBALYSE® - Synthèse.

Le 3 novembre 2022, après le début du projet, le fichier AGRIBALYSE a été mis à jour. Dans cette nouvelle version, certains produits ont des meilleures notes DQR<sup>3</sup>, et deux variables ont été ajoutées. Nous utilisons cette dernière pour étudier les produits alimentaires.

```
df <- read.csv(file = "Data/Agribalyse_Synthese.csv",
               header = TRUE,
               sep = ",",
               dec = ".",
               fill = TRUE,
               encoding="UTF-8")
```

### 2.2 Les unités statistiques.

Notre jeu de données contient 2517 unités statistiques. Les unités statistiques sont des produits alimentaires consommés en France et des produits agricoles produits en France.

### 2.3 Les variables.

La base de données présente 11 variables qualitatives (**Annexe 1**) et 18 variables quantitatives (**Annexe 2**). Les variables qualitatives portent sur les références et les caractéristiques des produits. On retrouve parmi les caractéristiques : le groupe alimentaire, le sous-groupe alimentaire, l'emballage, le type de livraison, la préparation, la saisonnalité du produit. Les variables qualitatives portent quant à elles sur la pollution d'une part, un score calculé à partir des autres variables de pollution ("Score EF"), d'autre part, et enfin, une note de qualité sur les données (DQR).

### 2.4 Préparation du jeu de données.

Après la préparation de nos variables (**Appendice**) la base de données se compose de 2743 unités statistiques et de 25 variables.

---

<sup>3</sup>Data quality ratio, note donnée entre 0 et 5, plus la note est proche de 0 plus la qualité des informations sur le produit est bonne.

### 3 Analyse univariée.

Notre base de données présente 18 variables quantitatives et 7 variables qualitatives. Séparons notre analyse en deux temps.

#### 3.1 Variables quantitatives.

Commençons par regarder les statistiques des variables de pollution et de la variable “Score EF” (**Annexe 3**).

Le résumé des variables quantitatives est présenté selon, le minimum, le premier quartile, la médiane, la moyenne, le troisième quartile, le maximum, l’écart-type.

Nous pouvons tirer de ce tableau plusieurs informations intéressantes.

En particulier sur l’interprétation de nos résultats :

- En moyenne, un produit alimentaire de notre base de données rejette 5.97 kg CO<sub>2</sub> eq. de la production à la consommation. Et 75 % des produits rejettent moins de 6.165 CO<sub>2</sub> eq. Le maximum est de 52.236 CO<sub>2</sub> eq.
- 75 % des produits ont un score EF inférieur à 0.874 mPt/kg par produit. Le maximum étant de 5.974 mPt/kg par produit.

Mais aussi sur la distribution de nos variables :

- Les variables, “Appauv. ozone”, “Particules”, “Eff. non cancer.” et “Eff. cancer.” ont des valeurs très petites. Nous constatons des échelles différentes selon les mesures, ce qui indique que nous devons sans doute normaliser par la suite nos unités statistiques.
- Pour aucune des variables la moyenne et la médiane ne correspondent pas.
- Nous avons, pour certaines variables, des maximums éloignés de notre troisième quartile. Par exemple, la variable “Rayo. ionisants”, à un troisième quartile à 1.332 et un maximum à 22,807.
- Les variables “Uti. sol” et “Epuis. eau” ont des valeurs négatives.

Certaines questions se posent quant à la distribution de nos variables :

- Notre base de données, présente-t-elle des données aberrantes ?

Les maximums de certaines variables semblent trop éloignés du troisième quartile.

- Nos données, suivent-elles une loi de probabilité ?

Les moyennes et les médianes ne correspondent pas.

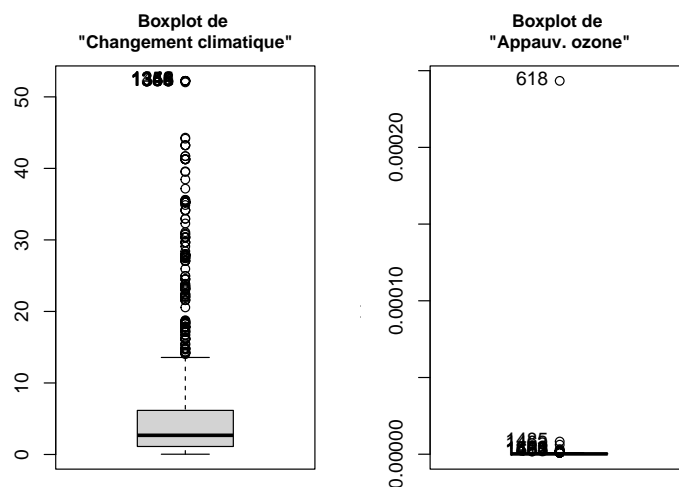
- Y a-t-il une minorité de produits avec des variables très élevées ?

Si on regarde le score EF, 75 % des produits ont une note inférieure à 0.874 mPt/kg de produit.

Regarder la distribution de nos variables nous permettra de répondre visuellement dans un premier temps.

### 3.1.1 Dispersion des variables.

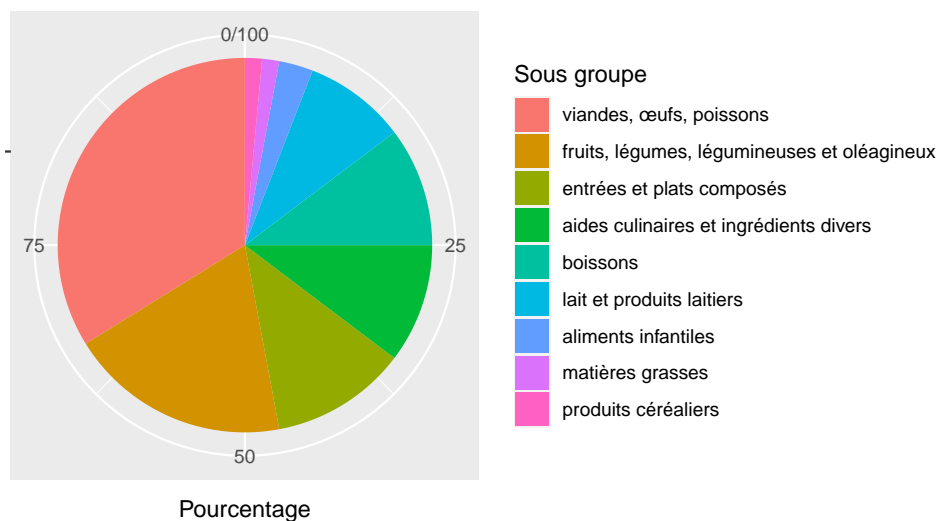
Pour confirmer ou infirmer nos à priori, nous regardons les box plots de nos variables de pollution : on regardera en particulier les variables “Changement climatique” et “Appauv. ozone”, les diagrammes en boîte des autres variables sont disponibles en **annexe 4**.



Comme attendues, nos variables présentent des produits sortant des box plots. De plus, nous pouvons constater que certaines unités statistiques ont des valeurs si grandes qu’elles “écrasent” nos boîtes, cela est particulièrement marquant pour la variable “Appauv. ozone”. Parmi les 166 unités statistiques sortant des boxplots nous retrouvons souvent les mêmes. En enlevant les doublons, on trouve 68 produits sortant des diagrammes en boîtes.

Regardons comment ces données se répartissent les données en-dehors des diagrammes en boîte selon le groupe d’aliments.

Digramme circulaire des produits sortant des boxplots selon le groupe d'aliments





Plus du quart des produits viennent du groupe “viandes, œufs, poissons”. La question de savoir, “quel groupe d’aliments pollue le plus ?”, reviendra plus tard dans notre analyse, mais nous avons une petite idée du groupe avec les pires scores EF.

Nous pouvons déjà commencer à apporter des arguments supplémentaires aux questions posées précédemment :

- Il y a effectivement des unités statistiques sortant des diagrammes en boîte, mais peut-on parler de données aberrantes ?

Une donnée aberrante est effectivement une donnée éloignée de la distribution “normale” de nos variables. Mais ici, nous avons supprimé les données de mauvaises qualités (**Appendice**). Les produits sortants des boxplots, sont en réalité des produits polluants énormément. Il y a effectivement des données polluant beaucoup en comparaison des autres produits.

- Nos données suivent-elles une loi normale ?

Les moyennes et les médianes ne correspondent pas. Nous pouvons aussi, nous appuyer sur les diagrammes en boîte d’une variable, pour supposer la normalité d’une variable. Si la barre de la médiane se trouve au centre de notre box-plot, alors on peut à priori continuer notre analyse pour savoir s’il s’agit bien d’une loi normale. Ici les boxplots semblent clairement s’éloigner d’une loi normale. Pour nous en convaincre regardons les histogrammes des variables de pollution uniquement.

### 3.1.2 Distribution des variables quantitatives.

**3.1.2.1 Histogramme des variables quantitatives.** Dans l’annexe 5, visuellement, il semble difficile de croire que nos variables suivent une distribution normale, surtout en regardant la ligne décrivant la densité. Les variables sont plus susceptibles de suivre une distribution log-normale. (Une variable aléatoire  $X$  à valeurs dans  $]0, +\infty[$  suit une loi log-normale de paramètres  $N(m, \sigma^2)$ , si  $Y = \ln(X)$  suit la loi  $N(m, \sigma^2)$ .)

Nous allons donc effectuer une transformation pour essayer de revenir à une loi normale. Pour cela, nous allons passer par une transformation de Box-Cox, afin de nous assurer qu’une transformation log-normale est adapté.

**3.1.2.2 Transformation de Box-Cox.** Nous devons trouver pour chaque variable la transformation qui semble la plus pertinente et analyser leur variabilité :

La transformation de Box-Cox donne la fonction suivante :

$$f(x, \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \log(x) & (\lambda = 0) \end{cases} \quad \forall x \in \mathbb{R}_+, \quad \forall \lambda \in \mathbb{R}$$

En fonction de la valeur de lambda, la transformation de nos variables change, ce qui donne les transformations suivantes :

$\lambda$	Transformation :
-2	$\frac{1}{x^2}$
-1	$\frac{1}{x}$
-0.5	$\frac{1}{\sqrt{x}}$
0	$\log(x)$
0.5	$\sqrt{x}$
1	$x$

$\lambda$	Transformation :
2	$x^2$

Si l'estimation de  $\lambda$  est proche d'une des transformations suggérées, il est conseillé de prendre l'une des transformations ci-dessus, sinon il faut prendre la transformation passant par  $\lambda$ .

Nous allons automatiser nos transformations :

- Création d'une fonction "trans\_bc", qui prend une variable en entrée et renvoie en fonction de  $\lambda$  la variable transformée selon le tableau et les explications ci-dessus.
- Création de la matrice "bc\_res" avec en colonne :
  - Le lambda.
  - le lambda arrondi.
  - La transformation effectuée.
- Dans une boucle, nous estimons le  $\lambda$  d'une variable avec la fonction "powerTransform". Ensuite, la fonction effectue la transformation à partir du lambda arrondie. Enfin, on stocke dans la variable "trans\_quanti" le résultat de la transformation.

Table 2: Transformation de Box-Cox

	lambda	lambda arrondi	Transformation
Changement climatique	-0.0545	-0.0545	Box-cox
Appauv. ozone	-0.1367	-0.1367	Box-cox
Rayo. ionisants	0.0133	0	log(x)
Forma. photo. ozone	-0.0426	-0.0426	Box-cox
Particules	-0.0248	0	log(x)
Eff. non cancer.	0.0546	0.0546	Box-cox
Eff. cancer.	0.0999	0.0999	Box-cox
Acid. terre./eaux d.	0.0219	0	log(x)
Eutro. ed	0.0552	0.0552	Box-cox
Eutro. marine	0.1147	0.1147	Box-cox
Eutro. terre	0.0551	0.0551	Box-cox
Écotoxi. aqua ed	0.1868	0.1868	Box-cox
Uti. sol	NA	NA	Valeurs négatives
Épuis. eau	NA	NA	Valeurs négatives
Épuis. éner.	-0.0065	0	log(x)
Épuis. miné.	-0.1088	-0.1088	Box-cox

Deux variables ont des valeurs négatives, quatre variables ont une transformation logarithmique, et le reste des variables ont une transformation de Box-Cox.

**3.1.2.2.1 Distribution log-normale.** Après la transformation logarithmique des variables "Rayo. ionisants", "Particules", "Acid. terre./eaux d.", "Épuis. éner.", regardons leurs histogrammes et leurs droites de Henry (**Annexe 6**).

Les histogrammes ne ressemblent malheureusement pas à une distribution d’une loi normale après la transformation. Cependant la droite de Henry semble bien alignée, sauf aux extrémités. Il semble aussi que certaines variables aient des mesures identiques sur beaucoup de produits. Vérifions avec un test de Shapiro-Wilks si ces variables suivent une loi normale.

---

**Test de Shapiro-Wilks sur la variable transformée “Rayo. ionisants”.**

---

Shapiro–Wilk normality test

data: Rayo. ionisants W = 0.98248, p-value = 9.088e-14

---

**Test de Shapiro-Wilks sur la variable transformée “Particules” :**

---

Shapiro–Wilk normality test

data: Particules W = 0.99134, p-value = 1.215e-08

---

**Test de Shapiro-Wilks sur la variable transformée “Acid. terre./eaux d.” :**

---

Shapiro–Wilk normality test

data: Acid. terre./eaux d. W = 0.99164, p-value = 2.003e-08

---

**Test de Shapiro-Wilks sur la variable transformée “Épuis. éner.” :**

---

Shapiro–Wilk normality test

data: Épuis. éner. W = 0.99146, p-value = 1.494e-08

---

Après transformation, nous pouvons conclure que les variables ne suivent pas de loi normale, cependant, la transformation nous permet de constater le caractère multi-modal de nos variables.

**3.1.2.2.2 Box-cox :** Comme pour les variables précédentes, regardons les variables ayant subi une transformation de Box-Cox (**Annexe 7**) et vérifions la normalité de nos variables (**Annexe 8**).

Après transformation, aucune variable ne suit une loi normale et comme précédemment, nous constatons le caractère multimodal de nos variables.

L’explication la plus probable est l’absence d’une variable expliquant la multi-modalité pour nos variables.

## 3.2 Variables qualitatives.

En fonction du nombre de modalités, la représentation graphique ne va pas être la même. Par exemple, la variable “Sous groupe aliment” contient 55 modalités (**Annexe 9**), on ne représentera pas cette variable de la même façon que la variable “Préparation” qui en a 7.

### 3.2.1 Tableau des fréquences.

Le tableau des fréquences (**Annexe 10**) nous permet de dire que certaines variables présentent peu de produits dans certaines modalités. Les variables “Saisonnalité” et “Transport par avion” contiennent chacune une modalité avec plus de 99 % des produits. Nous ne visualiserons donc pas graphiquement les résultats. Nous pourrions regarder graphiquement, via un diagramme en barre, la répartition des produits selon les variables “Groupe aliment”, “Livraison” et “Préparation”. Enfin, nous regarderons avec un diagramme circulaire les modalités les plus fréquentes des deux dernières variables.

### 3.2.2 Modalités des variables “Saisonnalité” et “Transport par avion”.

Nous avons noté, précédemment, que ces deux variables ont des modalités avec peu de produits. La variable “Saisonnalité” a 2 produits “de saison” et 2 autres produits “hors saison”, le reste des aliments se trouvent dans la catégorie “mix de consommation FR”. Ensuite, dans la variable “Transport par avion” nous avons 2 autres produits importés “par avion”.

Quels produits prennent ces modalités ?

Table 3: Produits “de saison” et “hors saison” (0 : hors saison ; 1 : de saison ; 2 : mix de consommation FR)

	Saisonnalité
Fraise de saison, crue	1
Fraise hors saison, crue	0
Tomate de saison, crue	1
Tomate hors saison, crue	0

Table 4: Produits “transportés par avion” code avion (1 : par avion)

	Transport par avion
Haricot vert importé par avion, cru	1
Mangue importée par avion, pulpe, crue	1

De ce tableau certaines questions se posent naturellement :

- Les fraises et tomates hors saison, polluent-elles plus que les fraises et tomates de saison ?

Les fraises et les tomates hors saison ont des paramètres de pollution plus élevés que les mêmes produits de saison (**Annexe 11**).

- Les produits transportés par avion polluent-ils plus que les autres produits ?

Pour cela, on regarde le score EF, il est censé résumer nos variables de pollutions.

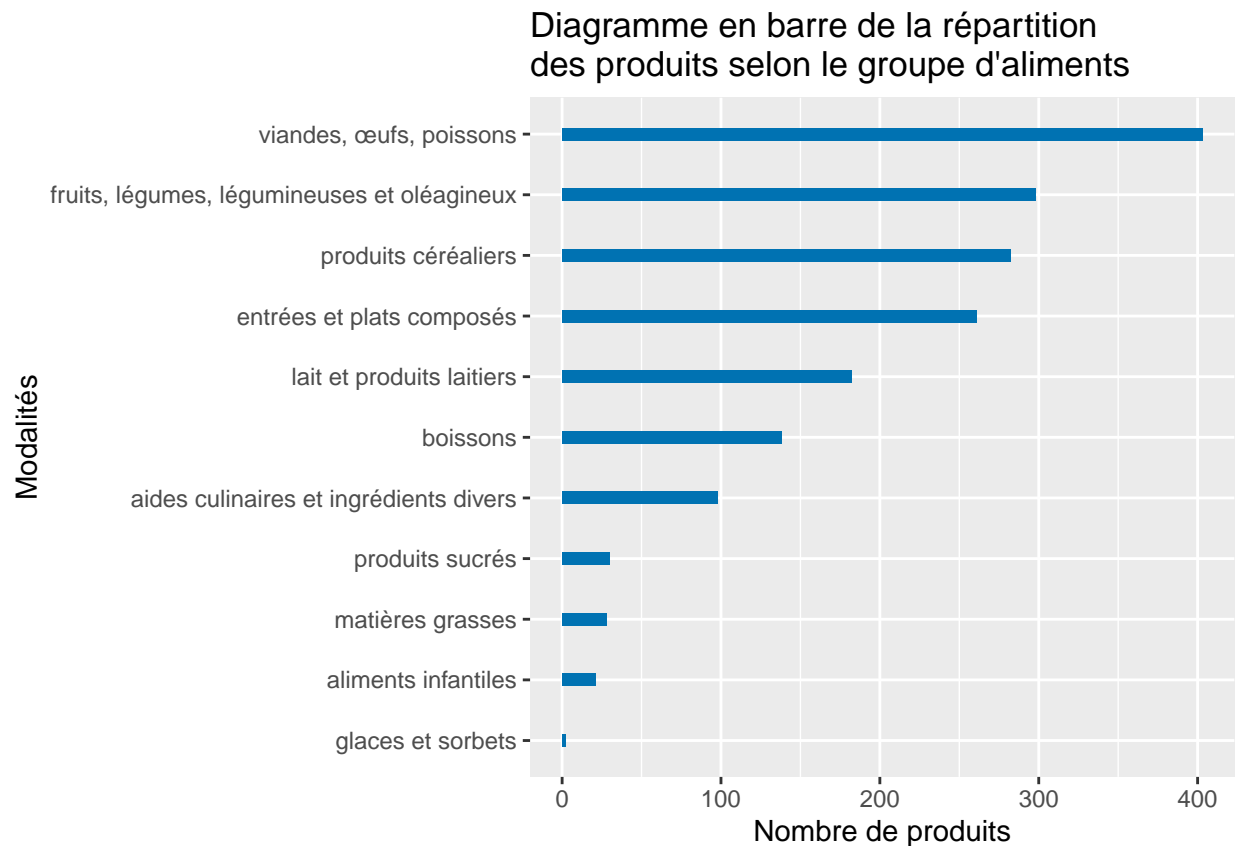
Table 5: Tableau des modalités rares

	Score EF
Haricot vert importé par avion, cru	0.5500296
Mangue importée par avion, pulpe, crue	0.9117978

La mangue, importée par avion, pollue globalement plus que la moyenne des produits (0.689 mPt/kg de produit), et les haricots globalement moins. Cependant, on ne peut pas en conclure que les produits importés par avion polluent plus ou moins que les autres aliments à partir de ces informations.

### 3.2.3 Diagrammes en barre.

Visualisons les diagrammes en barre des variables “Groupe aliment”, “Livraison” et “Préparation” :



Les modalités les plus représentées, pour la variable “Groupe aliment” sont les produits “viande, œufs, poissons”, les produits “fruits, légumes, légumineuses et oléagineux” et les “produits céréaliers”.

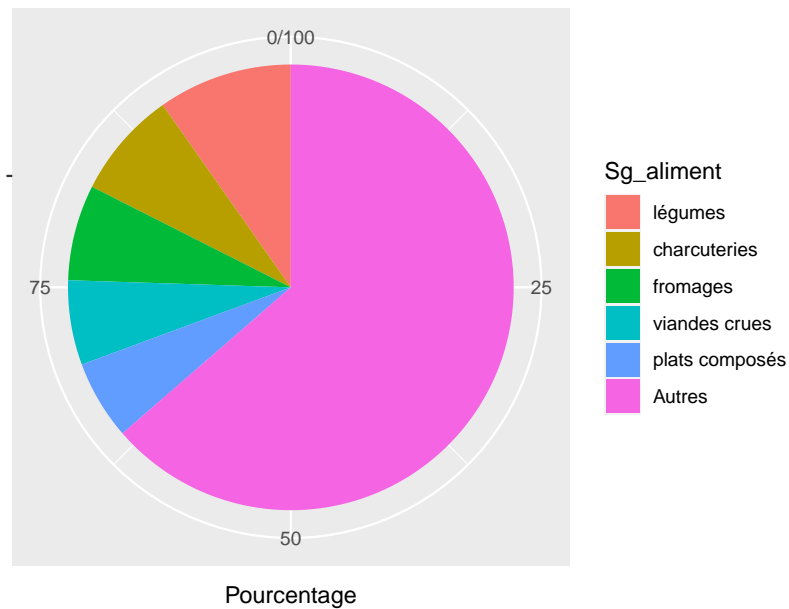
En annexe 12, pour les variables “Livraison” et “Préparation”, presque 50% des aliments sont livrés glacé et 66 % des aliments ne nécessitent pas de préparation.

Des questions se posent déjà sur la pollution de certains groupes d'aliments : existe-t-il des groupes d'aliments qui polluent plus que d'autres ? Ensuite, existe-t-il des relations entre les modalités de certaines variables ? (Par exemple, la livraison et le score EF.)

### 3.2.4 Diagramme circulaire.

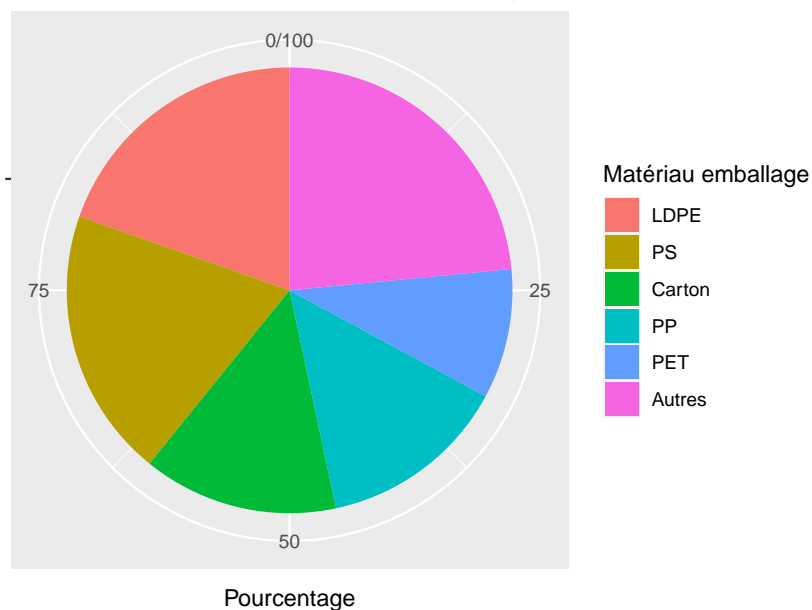
Nous allons essayer de représenter dans des diagrammes circulaires la part de chaque modalité des variables “Sous groupe aliment” et “Matériau emballage”.

Digramme circulaire de la répartition des produits selon le sous groupe d'aliments



Nous avons donc agrégé dans le facteur “Autres” les modalités des 5 plus grandes catégories. Nous avons ensuite, affichées le diagramme circulaire. Les 3 plus grands sous-groupes : “légumes”, “charcuteries” et “fromages” se partagent presque 25 % des produits.

Digramme circulaire de la répartition des produits selon le matériau d'emballage



Nous avons appliqué la même méthode que pour la variable “Sous groupe aliment”. Les 3 premiers matériaux

d'emballage utilisés sont les LDPE<sup>4</sup>, les PS<sup>5</sup> et les PP<sup>6</sup>. A eux trois, ils représentent plus de 50 % des matériaux d'emballage.

### 3.3 Conclusion de l'analyse univariée.

Plusieurs informations intéressantes ressortent de l'analyse univariée. Nous en apprenons plus sur la distribution des variables. En effet, nous savons maintenant qu'une minorité de produits de variables polluent énormément. Il semble que parmi les produits polluants énormément, beaucoup appartiennent au groupe "viande, œufs, poissons". Ensuite, nous n'avons pas réussi à ramener nos variables à une distribution connue.

Sur les variables qualitatives "Saisonnalité" et "Transport par avion", nous avons très peu de produits pour certaines modalités. On a donc pas assez de données pour savoir si, en fonction de la saisonnalité ou du type de transports, les produits sont plus ou moins polluants. Par la suite, nous avons vu quelles étaient les modalités les plus fréquentes, pour chaque variable.

On se rend compte assez rapidement des limites de la simple analyse univariée. Beaucoup de questions, que nous pouvons nous poser, tendent vers le croisement des variables.

---

<sup>4</sup>Polyéthylène basse densité (PEBD ou LDPE) est malléable et flexible. Sa grande pureté permet l'emploi de ce matériau plastique dans de nombreuses applications industrielles. Il est notamment très utilisé pour la fabrication des sacs plastique bretelles que l'on connaît aussi sous le nom de sac de caisse. (Mon-Emballage.com, s.d.)

<sup>5</sup>Le polystyrène est un thermoplastique. Le polystyrène de base, appelé PS cristal, est une matière dure et cassante, pouvant être transparente ou colorée. Il peut servir à fabriquer des produits rigides ou en mousse. Le plastique PS est très utilisé pour la fabrication des barquettes plastique sans couvercle, comme les barquettes de frites. (Mon-Emballage.com, s.d.)

<sup>6</sup>Le polypropylène est une variété de plastique très répandue dans l'agroalimentaire. C'est un matériau alliant rigidité et souplesse lors des chocs. Il reprendra naturellement sa forme après un choc de moyenne amplitude. Le plastique PP est très utilisé pour les barquettes plastique micro-ondables et les barquettes plastique scellables. Il est également résistant pour passer au congélateur. (Mon-Emballage.com, s.d.)

## 4 Analyse bivariable.

Nous l'avons vu au cours de la première partie de cette analyse, nous nous sommes posés des questions qui s'écartaient de la simple analyse univariée. Nous allons essayer, dans cette seconde partie, de mieux comprendre, les liens entre nos variables. Dans un premier temps, nous allons commencer comme notre première partie, avec les variables quantitatives.

### 4.1 Croisement des variables quantitatives.

#### 4.1.1 Nuages de points.

Nous commençons par une analyse des nuages de points pour visualiser le croisement entre nos variables quantitatives. (**Annexe 13**)

Nous trouvons plusieurs types de nuages de points pour chaque variable :

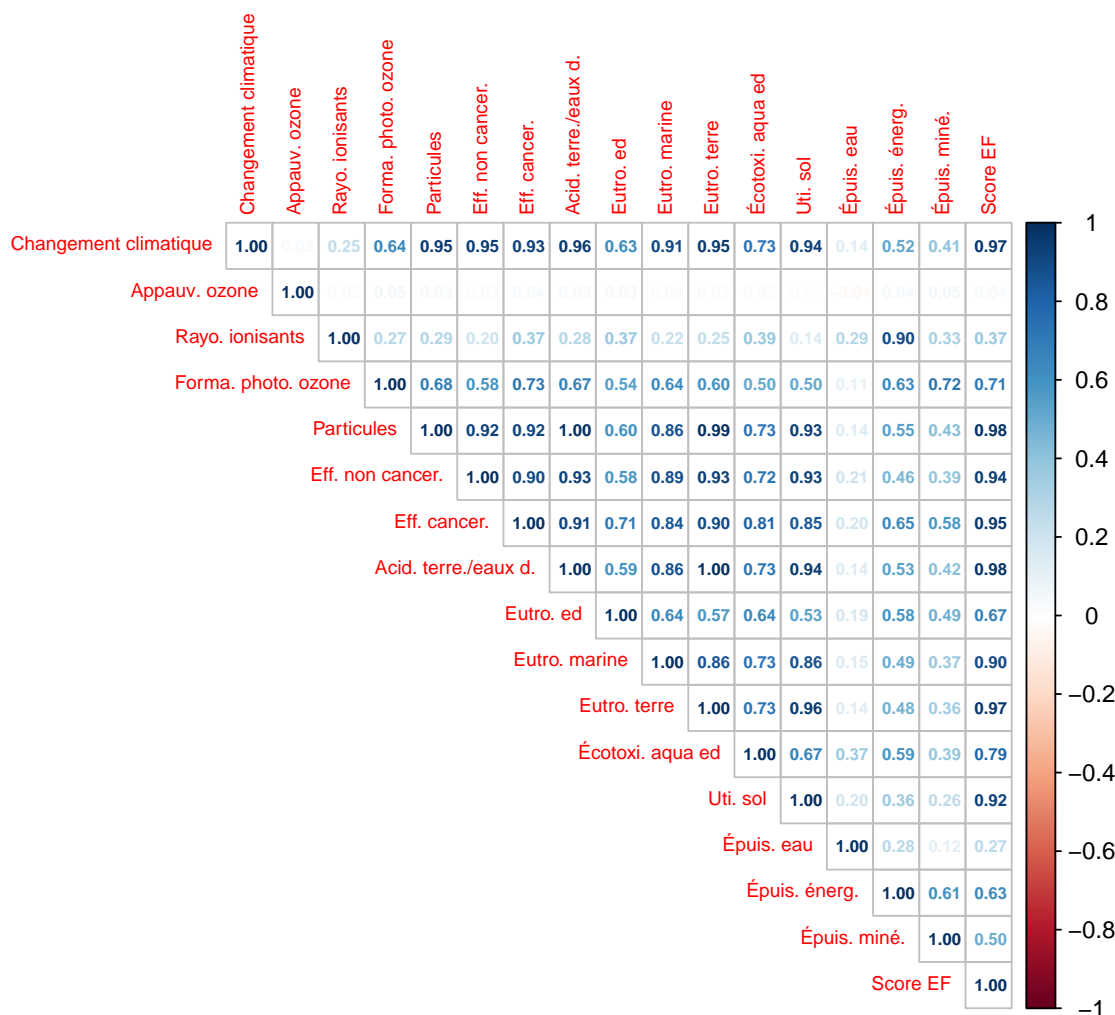
- Le lien linéaire : le croisement des variables, donne une ligne, signe que les variables ont un lien linéaire (positif quand c'est le cas). Le croisement des variables, "Particules" et "Eutro. terre.", est un exemple.
- Les nuages présentant plusieurs liens linéaires différents. Par exemple, sur le croisement des variables "Acid. terre./eaux d." et "Eutro. marine" nous pouvons voir plusieurs liens linéaires.
- Les nuages présentant peu ou pas de lien linéaire. Les exemples les plus frappants sont les variables "Appauv. ozone", qui donne un résultat constant sauf pour une valeur aberrante.

Pour confirmer ce que l'on constate visuellement, et pour avoir un ordre de grandeur de l'intensité de la liaison entre nos variables, nous regardons les coefficients de corrélation. Dans un premier temps les coefficients de corrélation de Pearson.



#### 4.1.2 Matrice des corrélations.

**Table des corrélations (Pearson)**

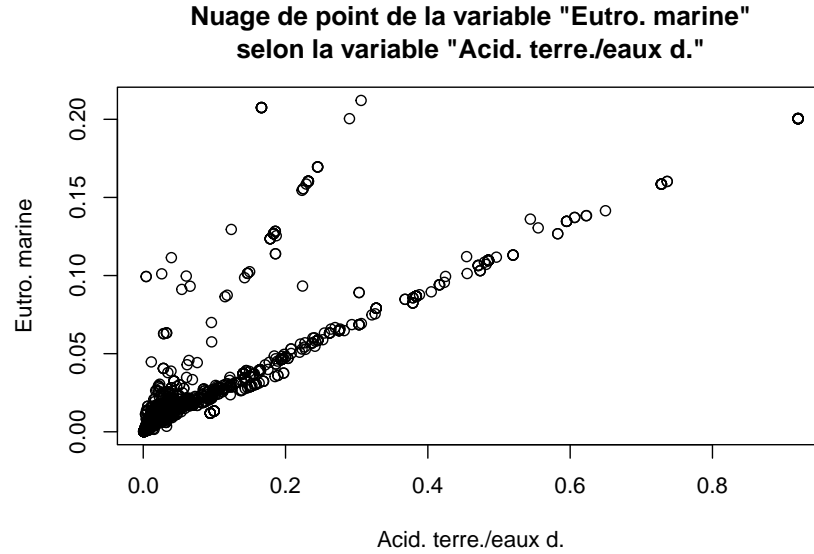


L'analyse visuelle, nous a donné les tendances entre nos variables. Les variables "Epuis eau" et "Appauv ozone" ont des coefficients proches de 0 avec les autres variables. Donc il n'y a pas de liens avec les autres variables. Nous pouvons par contre rajouter la variable "Epuis. eau", qui a des coefficients faibles, et la variable "Rayo. ionisants", qui a uniquement une corrélation forte avec la variable "Epuis. éner."

Nous retrouvons également, à l'inverse, des corrélations fortes entre certaines variables. Par exemple, "Particules" et "Acid. terre./eaux d." ont un coefficient égal à 1. Ou encore comme nous l'avons constaté visuellement : un coefficient de corrélation de 0.99 entre les variables "Particules" et "Eutro. terre."

Visuellement, nous ne constatons pas d'autres liaisons que des liens linéaires entre les variables. Mais par prudence, nous pouvons regarder les coefficients de Spearman (**Annexe 14**).

Il y a avec la méthode de Spearman, une augmentation de certains coefficients qu'il faut prendre en compte. En effet, l'augmentation, ou la baisse des coefficients, indiquent un lien monotone entre nos variables. Mais n'infirmons pas notre analyse visuelle. Effectivement, lorsque l'on regarde de plus près certains nuages de points, ceux-ci semblent avoir des liens linéaires, ce qui influe sur le coefficient de corrélation selon Pearson ou Spearman. Par exemple, les variables "Acid. terre./eaux d." et "Eutro. marine":



Le croisement semble donner deux liens linéaires différents, et ce n'est pas la première fois que nous le constatons. Nous avons vu dans notre analyse univariée que nos variables suivent des lois multimodales. Et ici, nous constatons que les unités statistiques, ont deux liens linéaires différents.

Une variable que nous n'avons que très peu analysé depuis le début de notre analyse, le score EF, est intéressant à regarder dans cette partie de notre analyse. En effet, ce score agrège les 16 variables de pollutions essayons de décomposer cette variable.

#### 4.1.3 Variable "Score EF".

En premier lieu, nous savons que la variable "Score EF" est la somme de chacune des variables multipliée par deux facteurs, le premier, le facteur de normalisation, le deuxième, le facteur de poids.

La formule ainsi donnée sur Eco-balyse :

$$Score\ EF = \sum_{i=1}^{16} \frac{Impact_i}{Normalisation_i} \times Pondration_i$$

Avec  $Impact_i$  la variable de pollution  $i$  du produit,  $Normalisation_i$  étant le coefficient de normalisation de la variable, et  $Pondration_i$  le facteur de pondération (de poids) de la variable.

Appliquer une régression linéaire multiple est possible ici, on pourra retrouver les coefficients pour trouver le score EF. Cela permettra en plus de juger de la qualité de la base de données en retrouvant les résultats de la base de données.

**4.1.3.1 Régression linéaire multiple.** En effectuant la régression linéaire multiple, nous trouvons les résultats suivants :

- $R^2$  : 1
- $R^2$  ajusté : 1
- Erreur résiduelle standard : 9.812457e-05 avec 1726 degrés de liberté

- Résidus

Table 6: Résidus du modèle

Statistic	Min	Pctl(25)	Median	Pctl(75)	Max
rlm1.residuals	−0.00003	−0.00000	−0.00000	−0.00000	0.004

- Coefficients :

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.0000	0.0000	1.15	0.2515
‘Changement climatique’	0.0279	0.0000	19204.46	0.0000
‘Appauv. ozone’	1176.1541	0.4043	2909.03	0.0000
‘Rayo. ionisants’	0.0119	0.0000	1421.52	0.0000
‘Forma. photo. ozone’	1.1775	0.0005	2522.11	0.0000
Particules	150556.0211	82.4261	1826.56	0.0000
‘Eff. non cancer.’	142943.9343	56.2072	2543.16	0.0000
‘Eff. cancer.’	1233368.2200	2284.1066	539.98	0.0000
‘Acid. terre./eaux d.’	1.1156	0.0008	1475.44	0.0000
‘Eutro. ed’	17.4245	0.0038	4528.42	0.0000
‘Eutro. marine’	1.5143	0.0002	7919.78	0.0000
‘Eutro. terre’	0.2099	0.0001	1657.80	0.0000
‘Écotoxi. aqua ed’	0.0003	0.0000	3762.50	0.0000
‘Uti. sol’	0.0001	0.0000	6120.95	0.0000
‘Épuis. eau’	0.0074	0.0000	25750.31	0.0000
‘Épuis. éner.’	0.0013	0.0000	3609.73	0.0000
‘Épuis. miné.’	1186.1647	0.1027	11551.09	0.0000

Table 7: Tableau des coefficients du modèle de régression

Pour l’interprétation de nos coefficients et la p-value on fixe un risque  $\alpha$  à 5%.

Alors effectivement, nous avons un  $R^2$  de 1, c’est-à-dire, que le coefficient de corrélation entre les variables explicatives et la variable expliquée est très fort, la qualité d’ajustement est très bonne. De plus, on ne tient pas compte de l’“Intercept”, la p-value est supérieure à 5%, le coefficient n’est pas significatif. Il n’est pas significativement différent de 0. Les autres coefficients sont significatifs au seuil  $\alpha$ .

Mais nous pouvons constater un problème : les résidus sont différents de zéro. En effet, nous savons que le résultat est un calcul sans coefficient aléatoire, l’erreur devrait donc être de 0.

**4.1.3.2 Résidus.** Pour essayer d’identifier d’où vient l’erreur, regardons tout d’abord les produits avec la plus grosse erreur.

Il semble que le produit “Sauce au poivre, condimentaire, froide, préemballée” est le produit avec la plus grande erreur (0.004049003). Le second produit est “Sauce au poivre, condimentaire, froide, préemballée” avec 0.0004467406 d’erreur.

Plusieurs pistes de recherche, pour savoir d’où provient l’erreur, sont envisageables :

- les produits ont des variables de pollution avec des valeurs extrêmes ?

Les produits n’ont pas de valeurs extrêmes.

- Les produits ont des notes “DQR” très proches de 3 ?

Les produits n’ont pas des notes “DQR” proches de 0, mais elles sont acceptables. (**Annexe 15**)

Ces pistes ne donnant rien, nous avons mis entre parenthèses l’analyse bivariable pour comprendre d’où l’erreur peut venir. Pour cela, nous sommes allés chercher les vrais coefficients sur Internet.

**4.1.3.3 Calcul “Score EF” :** Après de nombreuses recherches, nous trouvons les coefficients de normalisation et de pondération. (**Annexe 16** et **annexe 17**). (Notons que la variable avec le plus grand poids dans le “Score EF” est “Changement climatique”).

Pour essayer de comprendre d’où peut venir l’erreur, nous effectuons le calcul selon ce que nous avons trouvé. Nous avons donc la formule suivante :

$$Score\ EF = \sum_{i=1}^{16} \frac{Impact_i}{\frac{Normalisation_i}{1000}} \times Pondration_i$$


La division par 1000 de notre coefficient de normalisation vient de l’échelle de mesure de notre score EF, la normalisation est exprimée en Pt, or nous avons le score EF exprimé en mPt dans notre base de données.

Si l’on regarde les coefficients de la formule et les coefficients de la régression linéaire, on trouve presque la même chose (**Annexe 18**). Cependant, cela ne résout pas la problématique de nos coefficients. Si l’on regarde le produit avec la plus grande erreur, on trouve

dans la base de données un “Score EF” de : 0.1974744 mPt/kg de produit, dans la régression un “Score EF” de 0.1934254 mPt/kg de produit et avec le calcul “à la main” 0.1934045.

On trouve des résultats proches pour la régression linéaire et le calcul “à la main”. Même si l’erreur entre notre variable “Score EF” et notre régression est minime, elle rend notre modèle faux. Le modèle, qui devait nous renseigner sur la qualité de notre jeu de données, et en particulier sur la qualité, nous aura permis de nous mettre sur la piste d’une erreur plus globale sur notre jeu de données.

**4.1.3.4 Le problème du “Score EF”.** Pour comprendre le problème de notre variable “Score EF”, il nous a fallu regarder la version Excel de notre jeu de données. Et nous avons premièrement constaté la chose suivante : nous avons exactement le même score EF pour les deux produits alors que les indicateurs de pollutions sont différents (**Annexe 19**). De plus, lorsque l’on regarde la base de données au format Excel, on remarque que le “Code AGB” a été modifié et que le nom de la variable “LCI Name” est le même.



**Synthèse**

						Score unique EF 3.1
Code AGB	Code CIQUAL	Groupe d'aliment	Sous-groupe d'aliment	Nom du Produit en Français	LCI Name	mPt/kg de produit
11182	11182	aides culinaire	sauces	Sauce au poivre, chaude, préemballé	Sauce, pepper, prepacked	1,97E-01
11212	11212	aides culinaire	sauces	Sauce au poivre, condimentaire, froid	Sauce, pepper, prepacked	1,97E-01

Figure 1: Score EF des produits “Sauce au poivre”

Du point de vue de la formule donnée, on peut en conclure qu’il y a une erreur dans les “Score EF” pour ce produit et on peut étendre cette remarque à d’autres produits, car ce produit n’est pas le seul à avoir

été modifié (**ici** le fichier Excel). Nous n'avons pas d'informations allant dans le sens d'une rectification à posteriori des scores EF, ce qui remet en doute la qualité de la variable "Score EF".

Cependant, cela ne devrait pas impacter le reste de notre analyse, mais cette erreur est à prendre en compte.

## 4.2 Croisement des variables qualitatives.

### 4.2.1 Tableau de contingence "Préparation"/"Livraison".

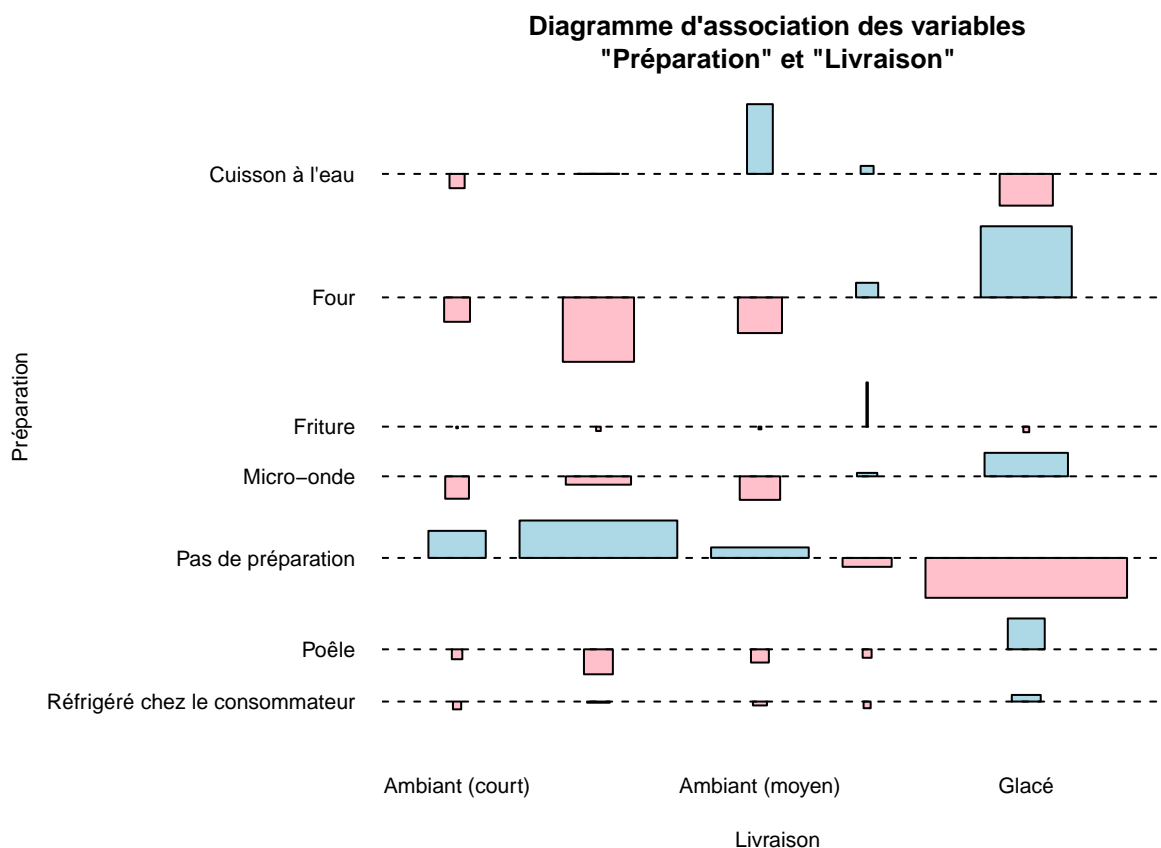
La question est de savoir s'il y a indépendance entre ces deux variables qualitatives.

Le tableau de contingence est disponible en **annexe 20**. Nous pouvons aussi regarder le tableau de contingence des fréquences, selon les marges du tableau. Ici, nous pouvons le faire selon les deux marges, car les deux marges sont libres.

Tous les résultats du tableau de contingence sont interprétables, un des plus intéressants étant : si la livraison est de type "Ambiant (court)" alors le produit ne nécessite pas de préparation. Nous pouvons faire ce constat sur d'autres croisements de modalités du tableau. Mais cela ne nous renseigne pas davantage sur le lien possible entre deux variables.

### 4.2.2 Représentation graphique de l'indépendance.

Nous pouvons utiliser le diagramme d'association pour représenter l'indépendance. Le diagramme d'association représente pour chaque couple de modalités l'écart entre la valeur théorique et la valeur observée. Ici, chaque rectangle représente cet écart. La couleur dépend du signe de l'écart.



Pour ces deux variables, nous observons quelques rectangles avec un air proche de 0. Mais, à priori, nous ne pouvons pas dire que les deux variables sont indépendantes. La représentation graphique aide à l'interprétation, mais ne permet pas de conclure non plus. Pour trancher, il faut faire un test d'indépendance du khi deux.

### 4.2.3 Test d'indépendance du khi-deux.

Nous souhaitons effectuer un test d'indépendance du khi-deux, on formule les hypothèses :

$$\begin{cases} H_0 : \text{La livraison est indépendante du groupe d'aliments} \\ H_1 : \text{La livraison dépend du groupe d'aliments} \end{cases}$$

Nous fixons un risque  $\alpha$  à 5%.

---

```
## Warning in chisq.test(tbl): 'L'approximation du Chi-2 est peut-être
      incorrecte
```

Pearson's Chi-squared test

data: tbl X-squared = 438.6, df = 24, p-value < 2.2e-16

---

Nous constatons un message d'erreur quand nous voulons connaître le résultat du test du khi-deux. Il peut y avoir plusieurs raisons, généralement, les valeurs théoriques sont inférieures à 5, c'est ce que l'on remarque ici (**Annexe 21**).

Le premier réflexe est d'utiliser le test exact de Fisher. Cependant, nous ne pouvons pas l'effectuer, car le tableau est trop grand. Nous utilisons le paramètre "simulate.p.value" pour effectuer notre test.

En quoi consiste le "simulate.p.value" ? Ce paramètre permet d'évaluer la p-value avec la méthode MCMC (Markov chain Monte Carlo) pour supprimer la contrainte des effectifs théoriques. Cette méthode simule la loi de distribution de la statistique de test et détermine sa probabilité d'exister. L'idée est de comparer la statistique de test sur nos observations à celle obtenue à partir des simulations tirées aléatoirement avec remise. (lemakistatheux, s.d.)

---

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: tbl X-squared = 438.6, df = NA, p-value = 0.0004998

---

Résultat du test : La p-value est la probabilité d'obtenir ce que l'on observe, ou pire, sous l'hypothèse d'indépendance. La probabilité d'obtenir ce  $X^2$ , ou pire, est de 0.0004998 sous l'hypothèse d'indépendance, autrement dit, il est peu probable d'avoir ce résultat, nous rejetons  $H_0$  au seuil  $\alpha$  de 5 %.

### 4.2.4 Les autres variables.

Après avoir effectué le test pour le couple de variable "Préparation" et "Livraison", nous pouvons l'effectuer pour les autres variables. On exclut les variables "Saisonnalité" et "Transport par avion", car pour ces deux variables, une modalité pèse pour 99 % des produits. Ensuite, nous enlevons la variable "Sous groupe aliment", elle est redondante avec le groupe d'aliments. De plus, elle comporte 55 variables ce qui est très

grand, ça nous forcera à construire des tableaux de contingence très grands, et nous aurons encore des valeurs théoriques inférieures à 5.

Nous fixons les mêmes hypothèses que précédemment, et le même risque  $\alpha$  (**Annexe 22**). Comme précédemment, nous retrouvons des valeurs théoriques inférieures à 5 (**Annexe 23**). En utilisant, la même méthode, nous pouvons dire que toutes les variables qualitatives testées ont un lien entre elles (**Annexe 24**).

### 4.3 Croisement des variables quantitatives et qualitatives.

Dans cette dernière partie de l’analyse bivariable, nous allons croiser les variables qualitatives et quantitatives. Avant de parler de lien entre les variables, certaines questions plus simples peuvent se poser.

#### 4.3.1 Pollution des produits par groupe d’aliments :

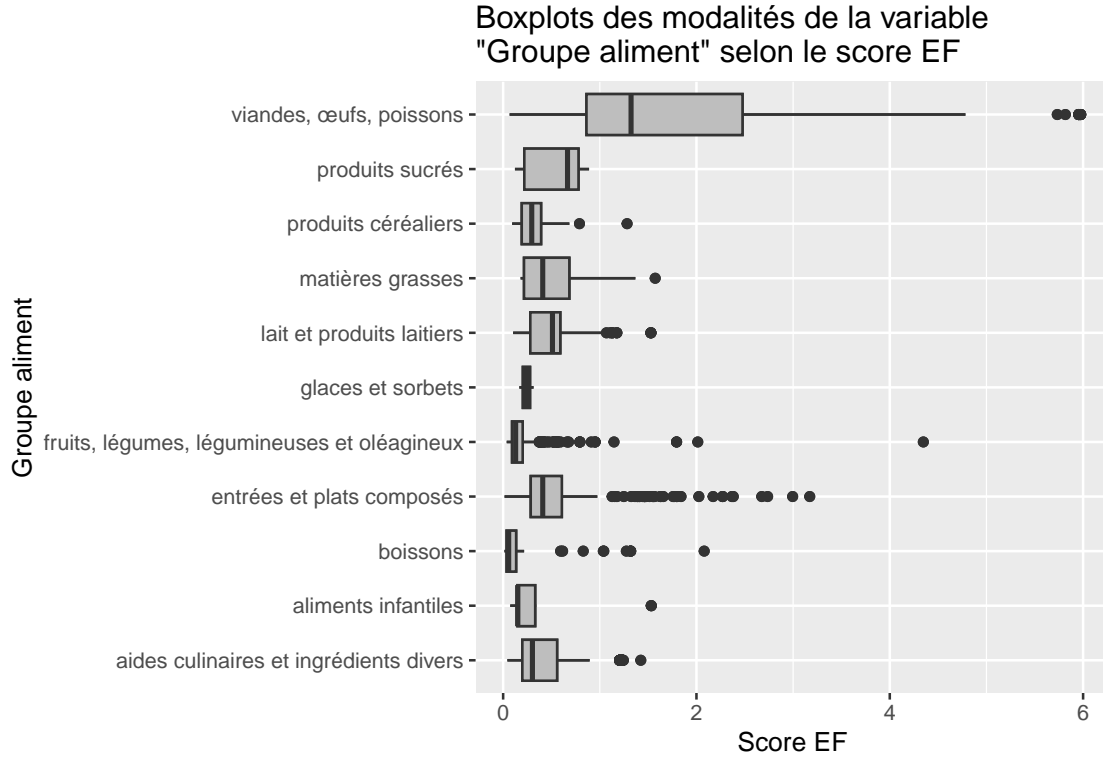
Quel groupe d’aliments pollue le plus en moyenne ? Le “Score EF” est intéressant à regarder dans cette situation, il permet de comparer les groupes d’aliments entre eux.

Table 8: Moyenne des scores EF par groupe d’aliments

Groupe aliment	Score_EF
aides culinaires et ingrédients divers	0.4714951
aliments infantiles	0.3185550
boissons	0.1550608
entrées et plats composés	0.5842308
fruits, légumes, légumineuses et oléagineux	0.2156464
glaces et sorbets	0.2409859
lait et produits laitiers	0.5207939
matières grasses	0.5205077
produits céréaliers	0.3016930
produits sucrés	0.5324882
viandes, œufs, poissons	1.7859768

Le groupe “viande, œufs, poisson” est le groupe qui pollue le plus avec un score EF en moyenne de 1.78 mPt/kg de produit et le groupe boisson est le groupe polluant le moins en moyenne.

Il peut être intéressant de regarder les diagrammes en boîte par groupe d’aliments. Cela permettra de mieux voir la dispersion de nos variables.



Il semble que globalement, le score EF soit plus grand quand le produit appartient à la modalité “viandes, œufs, poisson”. Il est possible qu’il existe une corrélation entre la variable “Groupe aliment” et “Score EF”.

#### 4.3.2 Corrélation entre les variables “Groupe aliment” et “Score EF”.

Le rapport de corrélation  $\eta^2$  calcule la corrélation entre une variable qualitative et une variable quantitative. C’est-à-dire, nous allons comparer la variation totale de la variable quantitative, à la variation de cette variable pour chaque modalité.

Calculons  $\eta^2$  pour la variable “Groupe aliment” et “Score EF” avec la formule :

$$\eta^2 = \frac{\sum_{k=1}^p n_k (\bar{x}_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Où  $x_i$  la valeur du produit  $i$ ,  $p$  le nombre de modalités,  $\bar{x}_k$  la moyenne du groupe  $k$ ,  $n_k$  le nombre d’individus appartenant au groupe  $k$  et  $\bar{x}$  la moyenne globale.

Si  $\eta^2 = 0$  alors il n’y a pas de lien entre les variables. Si  $\eta^2 = 1$  alors il y a un lien fort entre les variables.

La corrélation est de 0.4263954.

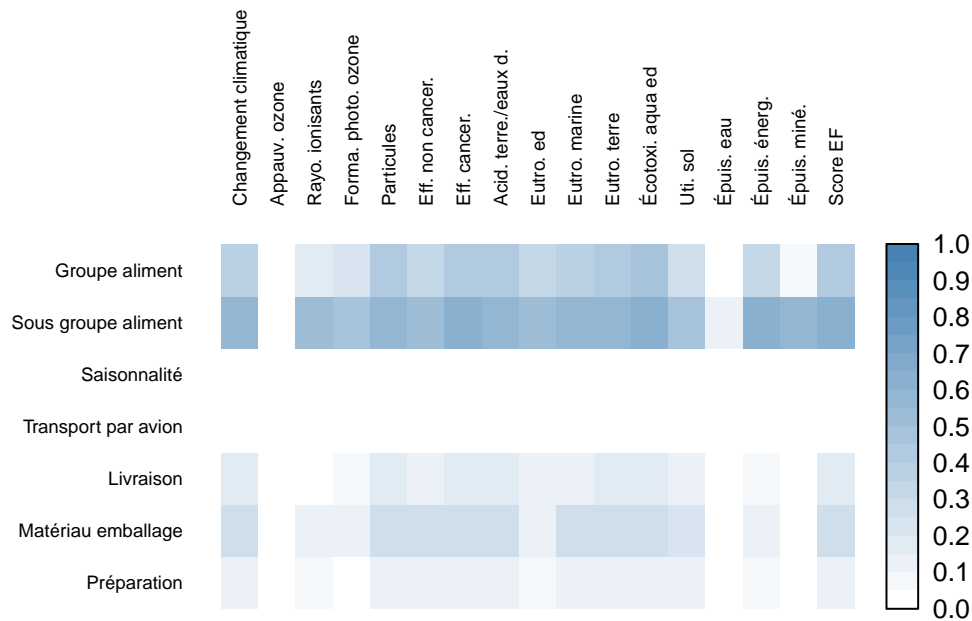
On ne peut pas dire qu’il existe une corrélation forte entre les deux variables, mais elle n’est pas inexistante.

#### 4.3.3 Matrice des corrélations.

Calculons les coefficients de corrélation entre les variables qualitatives et quantitatives. ( **Annexe 25** )



## Tableau des corrélations entre les variables quantitatives et qualitatives



benal/2023-05-08

Il est normal, que les variables qualitatives “Saisonnalité” et “Transport par avion”, aient peu de corrélation avec les variables de pollutions, elles ont toutes deux une modalité cumulant 99.9 % des produits. Les variables “Livraison”, “Matériau emballage” et “Préparation” ont des corrélations assez faibles avec les variables de pollutions. Les variables “Appauv. ozone” et “Epuis. éner.” ne sont corrélées avec aucune variable qualitative. Les variables “Écotoxi. aqua ed” et “Sous groupe aliment” ont le taux de corrélation le plus élevé :  $\eta^2 = 0.63$ .

Globalement, les corrélations semblent faibles, sauf pour la variable “Sous groupe aliment” et certaines variables de pollutions.

### 4.4 Conclusion de l’analyse bivariée.

Nous avons mis en avant dans la première partie de l’analyse bivariée les relations des variables quantitatives entre elles. Certaines variables sont fortement liées, d’autres ont plusieurs liens linéaires et certaines variables ne sont liées à aucune autre variable.

En particulier, nous avons voulu souligner le lien entre la variable “Score EF” et les autres variables. Nous savions qu’une formule était utilisée pour le calcul du “Score EF” et nous voulions la retrouver. Cependant, nous sommes tombés sur une erreur sur les “Score EF”. Il semble que les scores EF soient légèrement différents que ceux donnés.

Ensuite, nous avons essayé de mettre en évidence les liens entre les variables qualitatives. Nous avons pu constater que toutes les variables qualitatives sont liées. Par exemple, le type de transport dépend du groupe d’aliments.

Enfin, nous avons croisé les variables qualitatives et quantitatives. Nous avons essayé de savoir s’il existait un lien entre les variables qualitatives et quantitatives.

## 5 Analyse en composantes principales.

L'objectif de l'ACP (Analyse en Composantes Principales) est de décrire nos données par des graphes et de les synthétiser. Pour cela, nous allons étudier les individus et les variables de notre jeu de données.

### 5.1 Choix des variables.

La variable "Score EF", nous l'avons vu, résume nos variables selon un score, elle est redondante avec nos autres variables. Les 16 autres variables de pollutions sont nécessaires pour notre ACP. Pour les variables supplémentaires, nous choisissons de garder les variables : "Groupe aliment" et "Sous groupe aliment". Il peut être intéressant de regarder à la fin de notre analyse le graphique des individus selon des variables qualitatives.

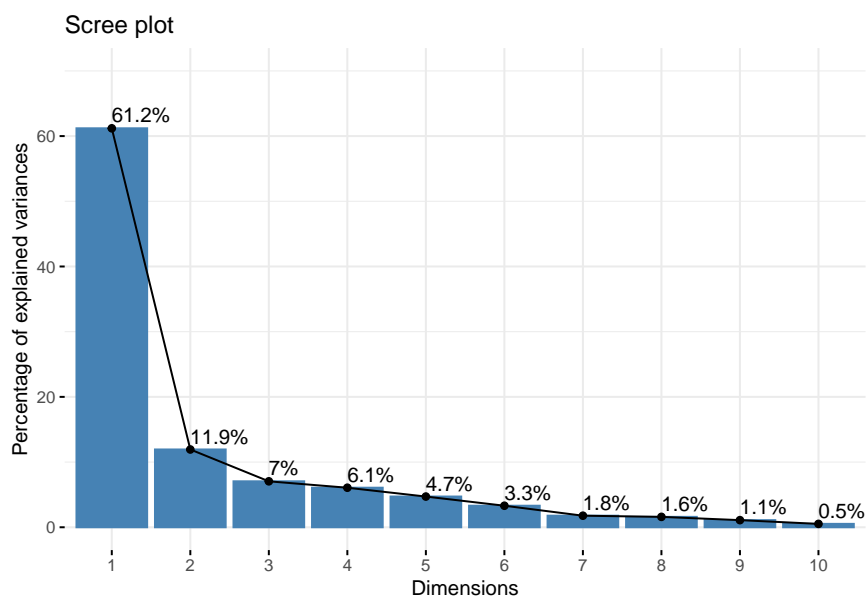
### 5.2 Valeurs propres et choix du nombre d'axes.

Le package FactoMineR est utilisé pour effectuer l'ACP, on ajoute les variables qualitatives et le paramètre "scale.unit" pour normaliser nos données, les données sont à des échelles différentes. (**Annexe 26**)

La première composante porte énormément d'informations. Comparons les différentes techniques pour décider du choix d'axes, mais à priori une analyse sur 2 dimensions serait meilleure :

- Le critère de Kaiser suggère d'interpréter les dimensions avec une inertie supérieure à 1, en regardant les valeurs propres. Ici, nous avons les 3 premières composantes qui sont supérieures à 1.
- La règle du coude, quant à elle, suggère de trouver une cassure dans l'éboulement des valeurs propres. Ici, nous avons une cassure entre la 2ème et la 3ème composante. Cette méthode nous suggère de prendre les 2 premières composantes.

Nos deux règles se contredisent, l'une suggère de prendre les 3 premières composantes et l'autre les 2 premières composantes. Pour trancher regardons le pourcentage cumulé de nos composantes.

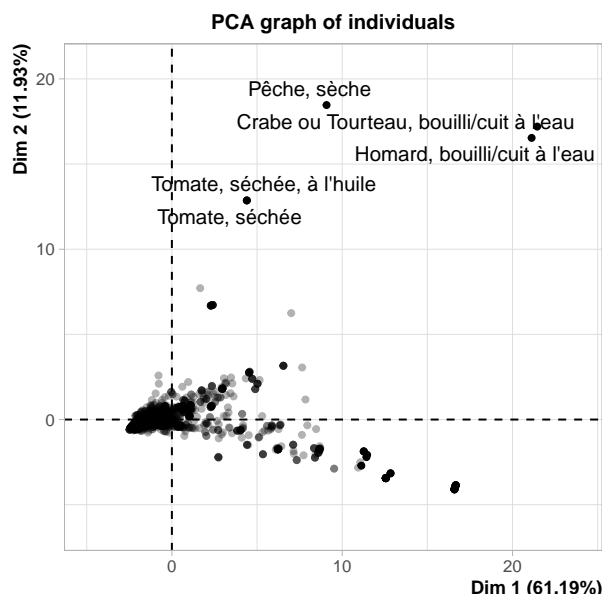


Avec 2 axes, nous retenons 73 % de l'information, ce qui est élevé en comparaison du nombre de variables, c'est résumer 73 % du jeu de données en dimension 16. Limitons-nous à 2 axes, à la fois parce que nous résumons beaucoup d'informations et pour ne pas complexifier l'interprétation de nos axes.

### 5.3 Analyse des individus.

En croisant les informations de contribution et de qualité de projection (**Annexe 27**) il semble qu'un individu contribue énormément à la formation de l'axe 2, mais à une qualité de représentation assez faible sur les dimensions 1 et 2. La "Pêche, sèche" à la plus forte contribution et a pourtant un indice de non-déformation de 0.07 sur l'axe 1 et de 0.31 sur l'axe 2. La qualité de représentation de cet individu n'est pas très grande en comparaison des autres observations. Nous conservons ce produit, car sa contribution n'est pas conséquente pour notre analyse. Il faudra faire preuve de prudence s'il l'on souhaite analyser cet individu en particulier sur nos deux axes.

La projection de nos observations sur nos 2 premiers axes donne quelques unités statistiques excentrées par rapport au reste des produits.



Il est intéressant de regarder quels individus s'opposent sur chaque axe pour développer de premières hypothèses quant à l'interprétation de nos dimensions. (**Annexe 28** et **annexe 29**)

Sur l'axe 1, les premiers produits sont des crustacés et des produits d'élevage d'animaux, de la viande d'agneau (côtelette, cote filet, gigot, selle, ...) sous plusieurs formes (grillé, braisé, rôti, ...). Donc des produits qui se ressemblent énormément au niveau de leur composition. À l'opposé, nous avons des produits buvables, comme du thé, de la soupe, et de l'eau.

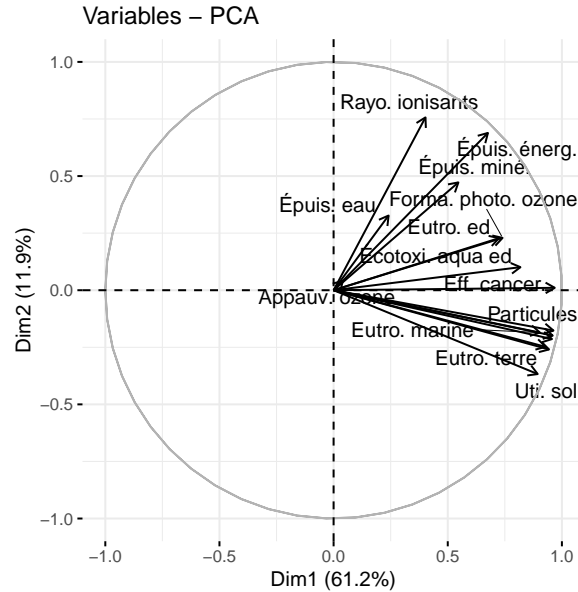
Sur l'axe 2, les produits avec les plus grandes coordonnées ont la particularité d'être séchés ou déshydratés, à l'exception de "Crabe ou Tourteau, bouilli/cuit à l'eau" et "Homard, bouilli/cuit à l'eau". À l'opposé de cet axe, nous trouvons beaucoup de produit vus sur l'axe 1, des produits composés d'agneau et du chevreau.

On retrouve en coordonnées négatives sur l'axe 2, les produits avec des coordonnées positives sur l'axe 1.

Enfin, si l'on regarde le graphe des individus selon les groupes d'aliments, il semble effectivement que l'on retrouve principalement des produits issus du groupe "viandes, œufs, poisson" sur l'extrémité droite de l'axe 1 et quelques produits de la catégories "fruits, légumes légumineuses et oléagineux" sur l'axe 2. (**Annexe 30**)

### 5.4 Analyse des variables.

Pour l'interprétation des axes selon les variables, on projette les variables dans le cercle des corrélations.



Sur le graphique, nous pouvons voir que : les flèches de certaines variables s'éloignent de l'extrémité du cercle des corrélations pour ces deux axes. Quand une variable est bien projetée, pour les axes considérés, la flèche est proche du cercle et quand la flèche est proche du centre du cercle, la variable est mal projetée.

La variable "Appauv. ozone" est très mal représentée dans ce plan. On retrouve, d'autres variables un peu mieux représentées, mais toujours distantes du cercle, comme les variables "Épuis. eau", "Forma. photo. ozone" et "Eutro. ed". On peut corroborer ça avec des chiffres. L'indice de non-déformation de la variable "Appauv. ozone" est de 0.003 sur le plan, autrement dit extrêmement faible.

L'autre information intéressante à regarder est la corrélation de chaque variable avec les dimensions du plan. Sur le graphique, plus une flèche est proche d'un axe, plus elle est corrélée avec cet axe. La variable la plus corrélée avec l'axe 1 est "Eff. cancer" (si l'on regarde les chiffres, le coefficient de corrélation est de 0.96). Mais on retrouve, pour l'axe 1 les variables de pollutions ayant un impact direct, le climat ("Changement climatique"), la biodiversité ("Acid. terre./eaux d.", "Particules", "Eutro. ed", "Eutro. marine", "Eutro. terre", "Écotoxi. aqua ed", "Uti. sol") et la santé environnementale ("Forma. photo. ozone", "Eff. non cancer", "Eff. cancer.>").

Les variables les plus liées à l'axe 2 sont les variables "Rayo. ionisants" et "Épuis. énerg.". On constate qu'elles aussi sont proches du cercle des corrélations. La variable "Rayo. ionisants" forme presque un angle droit avec les variables corrélées à l'axe 1.

(En **annexe 31** on retrouve les informations de qualité et de corrélation des variables sur ce plan)

## 5.5 Interprétation des axes.

### 5.5.1 Axe 1.

Notre ACP, oppose, sur l'axe 1, d'une part les produits affectant principalement le climat, la biodiversité et la santé environnementale, et d'autre part les produits n'ayant pas d'impacts sur ces trois catégories de pollutions. Les produits les plus à droite augmentent ce type de pollution. Nous pouvons remarquer, aussi, l'opposition qu'il semble exister entre les produits issus de l'élevage animalier et l'agriculture. Pour être plus précis, il semble y avoir une opposition entre la production de viande, issue de l'élevage et les légumes et les boissons.

### 5.5.2 Axe 2.

Nous trouvons sur l'axe 2 des produits liés aux “Rayo. ionisants” et à “Épuis. énerg.”, ces deux variables mesurent en particulier l'utilisation des énergies électriques produites par l'uranium et de l'utilisation d'énergie non-renouvelable. Il semble que certaines catégories de produits soient liées à l'axe 2. La deuxième dimension semble opposer des produits marins (“algues” et “mollusques et crustacés cuits”) à des produits provenant de l'élevage animalier.

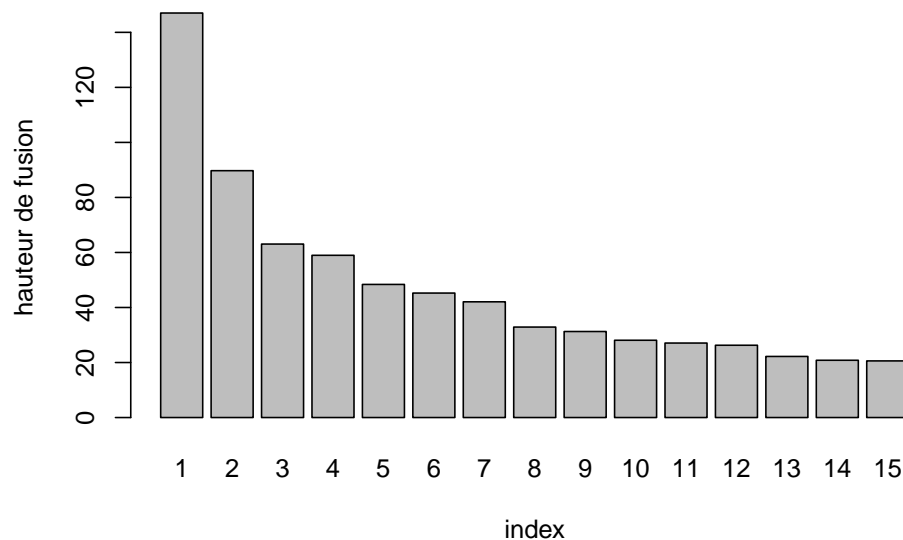
## 6 Classification.

### 6.1 Pré-traitement.

Regardons de nouveau les composantes de l'ACP, mais cette fois, il faut essayer d'aller chercher un maximum d'informations. Pour cela, regardons le tableau des valeurs propres et le diagramme des valeurs propres pour déterminer le nombre de composantes. Il faut, au minimum 5 composantes, pour atteindre 90 % de l'information. Sur le diagramme des valeurs propres, nous constatons un coude entre la composante 6 et 7. Finalement, nous retenons 6 composantes pour la suite de notre classification. (**Annexe 26** : Valeurs propres ACP.)

### 6.2 Méthode Mixte.

Nous souhaitons donc utiliser une méthode pour classer nos unités statistiques dans des groupes. Pour cela, nous allons effectuer dans un premier temps, une CAH (classification ascendante hiérarchique) sur les partitions obtenues et enfin une consolidation des classes obtenue par k-means.



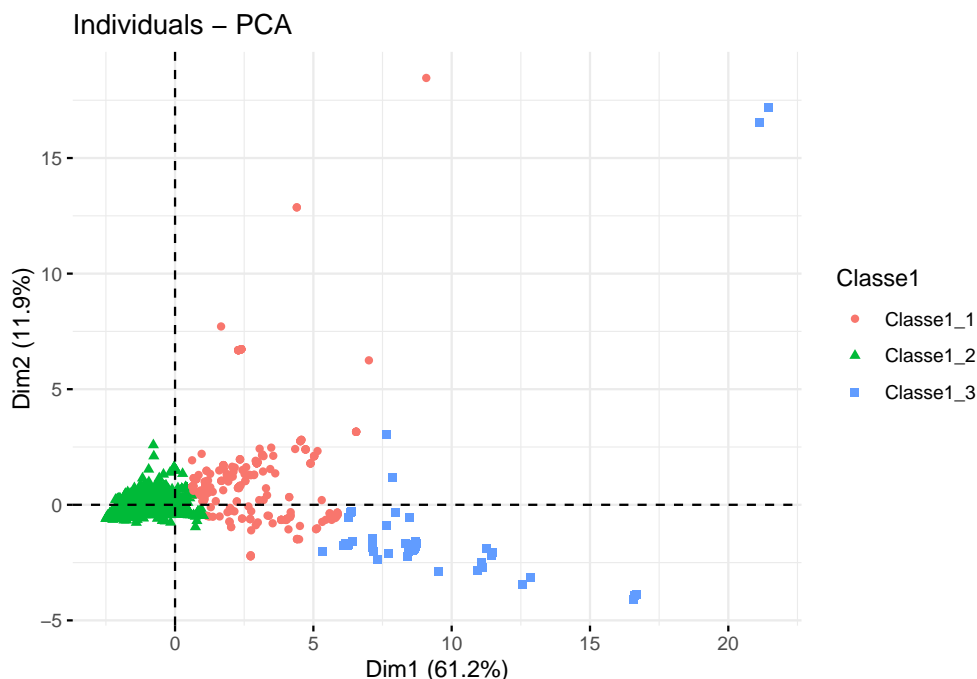
On constate un coude lorsque l'on passe de 2 à 3 classes. Le gain d'inertie intra-classe est élevé pour passer d'une partition de 2 à 3 classes. Ce graphique suggère de conserver 3 classes, mais il peut être intéressant, en fonction de notre analyse, de rajouter des classes par la suite.

Table 9: Répartition des produits par classe

	n	%	val%
1	290	16.6	16.6
2	1344	77.1	77.1
3	109	6.3	6.3

### 6.3 Analyse.

Pour l'analyse de nos classes, nous utiliserons en plus de la fonction "catdes" la projection de nos individus sur les axes de notre ACP selon la classe.



Nous retrouvons uniquement 6 classes, les 2 autres classes sont composées uniquement d'individus supplémentaires".

Nous retrouvons le message indiquant que l'approximation du khi-deux est peut-être incorrecte, nous n'évaluerons donc pas nos classes selon les catégories.

Nous nous concentrons donc sur les variables quantitatives pour décrire nos classes.

#### 6.3.1 Classe 1 : 290 produits.

Si l'on regarde le graphique, la classe 1 semble réunir les produits impactant les 3 catégories de pollution précédemment et les produits utilisant beaucoup de ressources énergétiques. Si l'on regarde cette classe par rapport à l'ensemble de nos produits (**Annexe 32**), notre groupe semble concentrer en particulier les produits avec des valeurs fortes sur les variables "Épuis.énerg.", "Rayo.ionisants" et "Écotoxi. aqua. ed".

#### 6.3.2 Classe 2 : 1344 produits.

La classe 2 réunit la plus grande partie des produits, cette classe regroupe les produits polluants le moins sur toutes les variables. Si l'on regarde les informations de notre groupe en comparaison de l'ensemble du jeu de données (**Annexe 33**), la moyenne de cette classe est inférieure à l'ensemble du jeu de données sur toutes les variables de notre jeu de données.

#### 6.3.3 Classe 3 : 109 produits.

Il semble que la dernière classe réunit les produits avec une forte pollution impactant la santé environnementale : la biodiversité et le climat. Les informations sur cette classe (**Annexe 34**), montre que ce groupe est

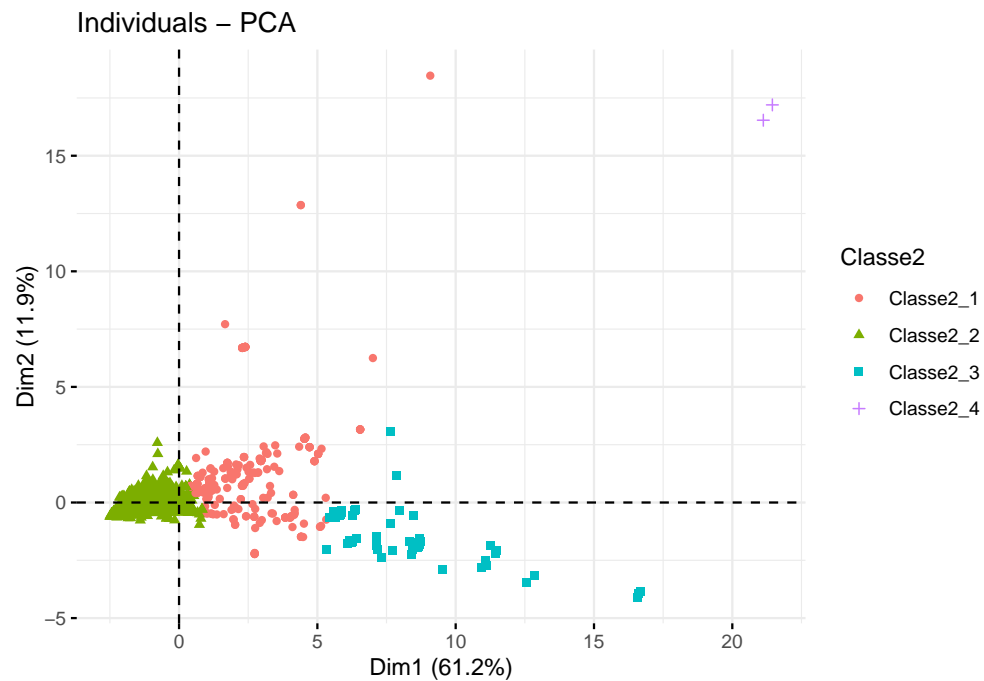
très lié aux variables “Eff. cancer”, “Eutro..terre”, “Changement climatique”, ... mais ne consomment que très peu d’énergie électrique. Cependant, deux produits semblent éloignés de ce groupe : les deux produits en haut à droite de notre graphique. Il faut prendre en compte ces deux produits.

Si l’on reproduit notre méthode, mais cette fois-ci avec 4 classes :

Table 10: Répartition des produits sur les 4 classes

	n	%	val%
1	299	17.2	17.2
2	1323	75.9	75.9
3	119	6.8	6.8
4	2	0.1	0.1

On obtient une 4ème classe avec 2 produits et la classe 1 à augmenter tandis que la classe 2 a diminué.



Notre analyse ne change pas pour les 3 premières classes, la 4ème classe réunit 2 valeurs extrêmes sur la dimension 1 et 2. Les deux produits polluent fortement et consomment beaucoup d’énergies.



## 7 Conclusion.

Nous avons essayé en explorant la base de données de tirer des informations intéressantes de notre base de données :

- Les variables de pollutions de notre jeu de données ne suivent pas de lois statistiques particulières.
- Certaines variables qualitatives ont des modalités concentrant tous les produits.
- Le croisement de certaines variables qualitatives donne plusieurs liens linéaires.
- Grâce à une régression linéaire multiple, nous avons trouvé des erreurs dans la variable “Score EF”. Il semble que le score EF n’ait pas été recalculé après les avoir synthétisés.
- Nous avons réussi à montrer que toutes les variables qualitatives sont en lien les unes avec les autres.
- Le groupe de produits provenant de l’élevage est le groupe qui pollue le plus selon le score EF.
- La variable “Sous groupe aliment” est la variable la plus corrélée aux variables de pollutions.
- En effectuant une ACP, nous avons essayé de résumer les variables de pollutions sur 2 axes :
  - Le premier axe correspond à la pollution des produits sur 3 catégories de pollution, le climat, la biodiversité et la santé environnementale.
  - Le second axe correspond à la consommation énergétique d’un produit.
- Enfin, avec la classification, nous avons mis en évidence 3 groupes de produits selon les variables de pollutions.

Cette analyse descriptive de notre jeu de données donne une idée générale de comment se répartissent nos données selon les variables de pollution. Cependant, il faut rappeler que ce jeu de données se base sur de nombreuses hypothèses et approximations. Nous nous sommes concentrés sur la synthèse d’un jeu de données plus vaste qui détaille les étapes du cycle de la vie. La base de données donne un ordre d’idée sur les catégories de produits polluant le plus. De plus, elle donne un ordre d’idée sur les types de pollutions associées aux produits.

Bien que cette méthode s’appuie sur des hypothèses et approximations, elle permettra sans doute bientôt de nous aider dans nos achats. De plus, la base de données AGRIBALYSE a été construite selon la base de données CIQUAL. Cette base de données donne les valeurs nutritionnelles d’un produit. La santé environnementale est indissociable de la santé physique. Mais est-il possible de mettre en relation les enjeux environnementaux et nutritionnels pour toucher la sensibilité des consommateurs ?

## 8 Appendice : préparation des données.

Nous préparons nos données pour l'analyse. À la fin du fichier des lignes ont été ajouté (par erreur ?) avec uniquement des virgules. Nous les avons supprimées manuellement pour plus de simplicité. Nous préparons les données statistiques en deux temps. D'abord les unités statistiques, ensuite les variables.

### 8.1 Unités statistiques.

Notre jeu de données a une quantité conséquente d'unité statistique, nous souhaitons filtrer les unités statistiques pour plusieurs raisons :

- Renommons les lignes de notre jeu de données, et supprimons la colonne correspondante :

---

```
rownames(df) <- df$Nom.du.Produit.en.Français
df <- df[, -5]
```

---

- La Commission européenne recommande d'utiliser avec prudence les produits avec une note "DQR" supérieure à 3. Par précautions, nous les enlevons, la qualité des mesures étant moins bonne :

---

```
df <- df[df$DQR < 3,]
```

---

- Supprimons les doublons.

---

```
df <- df[!duplicated(df),]
```

---

Après avoir exécuté nos filtres, la base de données ne contient plus que 28 variables et 1766 lignes.

### 8.2 Variables.

Nous souhaitons utiliser la base de données, mais nous constatons plusieurs problèmes :

- « Code CIQUAL » et « Code AGB » : ces codes sont uniques, mais nous avons déjà le numéro de ligne. Les codes sont associés à la base de données CIQUAL, mais nous ne l'utiliserons pas.
- « LCI Name », le nom du produit en anglais, fait doublon avec le « Nom du Produit en français ».
- Les colonnes ne sont pas dans un ordre qui nous convient.
- Les variables "Sous groupe aliment", "Saisonnalité", "Transport par avion", "Livraison", "Matériau emballage" et "Préparation" sont de type "chr" quand elles devraient être de type "factor"

Nous faisons donc un certain nombre de modifications pour simplifier notre dataframe :

---

```
# Nous supprimons les colonnes inutiles
df <- df[, -c(1,2,5)]

# On trie les colonnes
df <- df[, c(1:7, 10:25, 9, 8)]

# On modifie le nom des colonnes.
colnames(df) <- c("Groupe_aliment",
                  "Sous_groupe_aliment",
                  "Saisonnalité",
                  "Transport_par_avion",
                  "Livraison",
                  "Matériau_emballage",
                  "Préparation",
                  "Changement_climatique",
                  "Appauv._ozone",
                  "Rayo._ionisants",
                  "Forma._photo._ozone",
                  "Particules",
                  "Eff._non_cancer.",
                  "Eff._cancer.",
                  "Acid._terre./eaux_d.",
                  "Eutro._ed",
                  "Eutro._marine",
                  "Eutro._terre",
                  "Écotoxi._aqua_ed",
                  "Uti._sol",
                  "Épuis._eau",
                  "Épuis._énerg.",
                  "Épuis._miné.",
                  "Score_EF",
                  "DQR")

# On remplace en supprimant les espaces
df$Préparation[df$Préparation == "Pas_de_préparation_"] <- "Pas_de_préparation"

df$Préparation[df$Préparation == "Four_"] <- "Four"
df$Préparation[df$Préparation == "Micro-onde_"] <- "Micro-onde"

# On modifie le type des variables en factor
df$'Groupe aliment' <- as.factor(df$'Groupe aliment')
df$'Sous groupe aliment' <- as.factor(df$'Sous groupe aliment')
df$Saisonnalité <- as.factor(df$Saisonnalité)
df$'Transport par avion' <- as.factor(df$'Transport par avion')
df$Livraison <- as.factor(df$Livraison)
df$'Matériau emballage' <- as.factor(df$'Matériau emballage')
df$Préparation <- as.factor(df$Préparation)
```

---

## 9 Sources.

### 9.1 Sitographie.

ADEME. (2022, Novembre 3). *Portail open data de l'ADEME*. Récupéré sur AGRIBALYSE® - Synthèse: <https://data.ademe.fr/datasets/agribalyse-31-synthese>

Agribalyse. (2022, Novembre). *Méthode de l'Analyse du Cycle de Vie*. Récupéré sur Documentation Agribalyse: <https://doc.agribalyse.fr/documentation/methodologie-acv>

Commission européenne. (s.d.). *Données clés sur les consommateurs en 2020*. Récupéré sur [https://ec.europa.eu/info/sites/default/files/120321\\_key\\_consumer\\_data\\_factsheet\\_fr.pdf](https://ec.europa.eu/info/sites/default/files/120321_key_consumer_data_factsheet_fr.pdf)

Ecobalyse. (s.d.). *Indicateurs d'impacts ACV*. Récupéré sur Ecobalyse: <https://fabrique-numerique.gitbook.io/ecobalyse/alimentaire/impacts-consideres>

European Commission. (2022, Novembre). *Developer Environmental Footprint (EF)*. Récupéré sur European Platform on LCA | EPLCA: <https://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml>

Garric, A. (2016, 10 27). La moitié des vertébrés a disparu en quarante ans. *Le Monde*. Récupéré sur [https://www.lemonde.fr/biodiversite/article/2016/10/27/la-moitie-des-vertebres-a-disparu-en-quarante-ans\\_5020936\\_1652692.html](https://www.lemonde.fr/biodiversite/article/2016/10/27/la-moitie-des-vertebres-a-disparu-en-quarante-ans_5020936_1652692.html)

lemakistatheux. (s.d.). *Le test du Chi2 de Pearson*. Récupéré sur <https://lemakistatheux.wordpress.com/>: <https://lemakistatheux.wordpress.com/category/tests-statistique-indices-de-liaison-et-coefficients-de-correlation/le-test-du-chi2-de-pearson/>

Marine Stewardship Council (MSC). (s.d.). *Quelles sont les conséquences de la surpêche ?*. Récupéré sur [www.msc.org](https://www.msc.org/fr/informations-sur-le-msc/consequences-de-la-surpeche): <https://www.msc.org/fr/informations-sur-le-msc/consequences-de-la-surpeche>

Mon-Emballage.com. (s.d.). *Guide des emballages alimentaires plastique*. Récupéré sur Mon-Emballage.com: <https://www.mon-emballage.com/content/131-guide-des-emballages-alimentaires-plastique>

Pr Lobry, J. R. (s.d.). *Normalisation d'une variable quantitative : la transformation de Box et Cox*. Récupéré sur [pbil.univ-lyon1.fr](https://pbil.univ-lyon1.fr/R/pdf/tdr29.pdf): <https://pbil.univ-lyon1.fr/R/pdf/tdr29.pdf>

## 10 Annexes.

### 10.1 Annexe 1 : variables qualitatives.

Nom variable	Type de variable	Description
Code AGB	Qualitative	Code du produit
Code CIQUAL	Qualitative	Autre code du produit
Groupe d'aliment	Qualitative	L'AGRIBALYSE place un produit dans une catégorie alimentaire. Par exemple : fruits, légumineuses et oléagineux.
Sous-groupe d'aliment	Qualitative	Il s'agit d'une description plus détaillée du produit
Nom du Produit en Français	Qualitative	Nom du produit en Français
LCI Name	Qualitative	Nom du produit en Anglais
Saisonnalité	Qualitative	Pour les fruits et légumes par exemple, la saisonnalité indique si le produit est « de saison », « hors saison ». S'il est composé de plusieurs ingrédients ou si on ne peut pas trouver de saisonnalité pour le produit, le produit est classé comme « mix de consommation FR »
Transport par avion	Qualitative	Si un produit est transporté en avion, il est noté 1 sinon il est noté 0.
Livraison	Qualitative	Type de livraison effectué pour un produit, si le produit est transporté dans une température ambiante et la distance que le produit parcourt (long, moyen, court) si le produit est « congelé » ou « glacé ».
Matériau d'emballage	Qualitative	On retrouve le type d'emballage du produit.
Préparation	Qualitative	Indique le type de préparation du plat.

## 10.2 Annexe 2 : variables quantitatives.

Nom variable	Type de variable	Description
DQR - Note de qualité de la donnée	Quantitative continue	Data Quality Ratio (le rapport de qualité des données) : note de qualité allant de 1, très bon, à 5 très mauvaise. Le DQR est associé à chaque produit agricole et alimentaire pour lequel Agribalyse fournit des inventaires de cycle de vie et des indicateurs d'impacts. La Commission européenne recommande de la prudence dans l'utilisation des données avec des DQR supérieurs à 3. Dans la base de données AGRIBALYSE, 67 % des données ont un DQR jugé bon ou très bon (1 à 3). La note DQR est la moyenne de 4 variables, la TeR : Représentativité technique, la GeR : Représentativité géographique, la TiR : Représentativité temporelle, la P : Incertitude des paramètres. Ces données ne sont pas disponibles dans notre base de données.
Score unique EF	Quantitative continue	Score indiquant l'impact du produit sur l'environnement, il est exprimé en mpt/kg c'est-à-dire en milli-point par kilo. Plus ce score est élevé plus l'impact environnemental est fort. 1 point correspond à l'impact environnemental moyen d'un citoyen européen sur 1 an.
Changement climatique	Quantitative continue	Correspond à la modification du climat, affectant l'écosystème global. Exprimé en kg CO <sub>2</sub> eq. (Conversion des quantités des divers gaz émis en la quantité équivalente de dioxyde de carbone ayant le même potentiel de réchauffement planétaire.)
Appauvrissement de la couche d'ozone	Quantitative continue	La couche d'ozone est située en haute altitude dans l'atmosphère, elle protège des rayons ultra-violet solaires. Son appauvrissement augmente l'exposition de l'ensemble des êtres vivants à ces radiations négatives (cancérogènes en particulier). Exprimé en kg CFC-11 eq.
Rayonnements ionisants	Quantitative continue	Correspond aux effets de la radioactivité. Cet impact correspond aux déchets radioactifs résultants de la production de l'électricité nucléaire. Exprimé en kBq(def) U-235(def) eq. (Kbq U-235 : kilobecquerel d'Uranium 235)
Formation photochimique d'ozone	Quantitative continue	Correspond à une dégradation de la qualité de l'air, principalement via la formation de brouillard de basse altitude nommé "smog". Il a des conséquences néfastes sur la santé. Exprimé en kg NMVOC (def) eq (NMVOC : Non-methane volatile organic compounds ou composés organiques volatils non-méthaniques)

Nom variable	Type de variable	Description
Particules	Quantitative continue	Les particules fines pénètrent dans les organismes, notamment via les poumons. Elles ont un effet sur la santé humaine. Exprimé en disease inc.
Effets toxicologiques sur la santé humaine : substances non-cancérogènes	Quantitative continue	Impact sur la santé lié à l'exposition de la population générale aux contaminants chimiques via les pollutions du milieu (air, eau, sol). Ces contaminants émis dans l'environnement sont notamment des pesticides, des métaux lourds ou encore des polluants industriels. L'exposition par ingestion directe d'un aliment contenant des résidus de pesticides n'est pour l'instant pas intégrée dans les calculs d'Analyse du Cycle de Vie. (Exprimé en CTUh)
Effets toxicologiques sur la santé humaine : substances cancérogènes	Quantitative continue	La définition est la même que " Effets toxicologiques sur la santé humaine : substances non-cancérogènes "
Acidification	Quantitative continue	Résulte d'émissions chimiques dans l'atmosphère qui se redéposent dans les écosystèmes. Cette problématique est connue en particulier via le phénomène des pluies acides. Exprimé en mol H <sup>+</sup> eq.
Eutrophisation terrestre	Quantitative continue	Comme dans l'eau, l'eutrophisation terrestre correspond à un enrichissement excessif du milieu, en azote en particulier, conduisant à un déséquilibre et un appauvrissement de l'écosystème. Ceci concerne principalement les sols agricoles. Exprimé en mol N eq.
Eutrophisation eaux douces	Quantitative continue	Correspond à un enrichissement excessif des milieux naturels en nutriments, ce qui conduit à une prolifération et à une asphyxie (zone morte). C'est ce phénomène qui est à l'origine des algues vertes. On peut le retrouver en rivière et dans les lacs également. Exprimé en P eq.
Eutrophisation marine	Quantitative continue	Correspond à un enrichissement excessif des milieux naturels en nutriments, ce qui conduit à une prolifération et à une asphyxie (zone morte). C'est ce phénomène qui est à l'origine des algues vertes. Exprimé en N eq. (N : Azote de nitrate)
Utilisation du sol	Quantitative continue	Les terres sont une ressource finie, qui se partage entre milieux "naturels" (forêts), productifs (agricultures) et urbains. L'usage des terres et les habitats déterminent dans une large mesure la biodiversité. Cette catégorie reflète donc l'impact d'une activité sur la dégradation des terres, en référence à « l'état naturel ». Exprimé en point.

Nom variable	Type de variable	Description
Écotoxicité pour écosystèmes aquatiques d'eau douce	Quantitative continue	Ecotoxicité d'eau douce, Toxicité humaine cancérigène et non cancérigène. Indicateurs de toxicité via la contamination de l'environnement. Ces indicateurs sont encore peu robustes actuellement. Exprimé en CTUe (CTUe : unité de toxicité des écosystèmes)
Épuisement des ressources en eau	Quantitative continue	Correspond à la consommation d'eau et son épuisement dans certaines régions. Cette catégorie tient compte de la rareté (cela a plus d'impact de consommer un litre d'eau au Maroc qu'en Bretagne). Exprimé en m3 world eq.
Épuisement des ressources énergétiques	Quantitative continue	Correspond à l'épuisement des ressources énergétiques non-renouvelables : charbon, gaz, pétrole, uranium, etc. Exprimé en MJ/kg de produit
Épuisement des ressources minéraux	Quantitative continue	Correspond à l'épuisement des ressources minérales non-renouvelables : cuivre, potasse, terres rares, sable, etc. Exprimé en Sb eq/kg de produit.



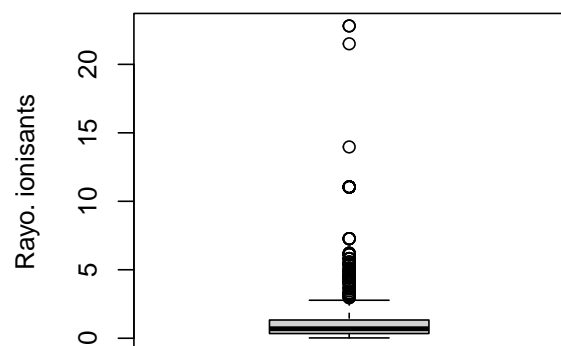
### 10.3 Annexe 3 : statistiques descriptives.

Table 13: Statistiques descriptives

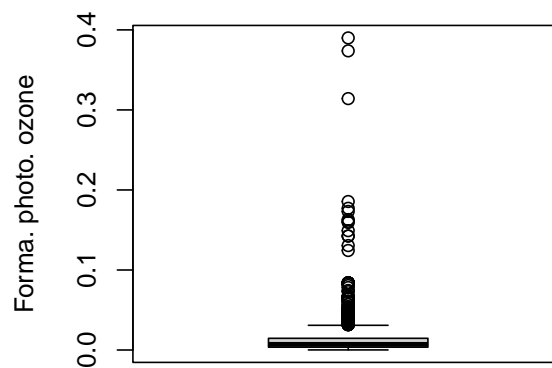
Statistic	N	Min	Pctl(25)	Median	Mean	Pctl(75)	Max	St. Dev.
Changement climatique	1,743	0.039	1.120	2.680	5.972	6.165	52.236	8.765
Appauv. ozone	1,743	0.000	0.00000	0.00000	0.00000	0.00000	0.0002	0.00001
Rayo. ionisants	1,743	0.023	0.343	0.695	1.158	1.332	22.807	1.720
Forma. photo. ozone	1,743	0.0001	0.004	0.007	0.014	0.015	0.390	0.024
Particules	1,743	0.000	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000
Eff. non cancer.	1,743	0.000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Eff. cancer.	1,743	0.000	0.000	0.000	0.000	0.000	0.00000	0.000
Acid. terre./eaux d.	1,743	0.0005	0.010	0.032	0.082	0.094	0.920	0.137
Eutro. ed	1,743	0.00001	0.0002	0.0005	0.001	0.001	0.024	0.001
Eutro. marine	1,743	0.0002	0.005	0.013	0.025	0.026	0.212	0.037
Eutro. terre	1,743	0.001	0.035	0.127	0.344	0.354	4.069	0.598
Écotoxi. aqua ed	1,743	0.678	14.338	37.188	57.139	78.819	484.430	59.774
Uti. sol	1,743	-40.044	55.523	154.995	401.020	337.945	5,579.824	761.196
Épuis. eau	1,743	-23.128	1.256	2.389	4.726	5.450	331.230	11.201
Épuis. énerg.	1,743	1.710	20.944	31.003	48.697	60.184	612.111	53.206
Épuis. miné.	1,743	0.00000	0.00001	0.00001	0.00002	0.00002	0.001	0.00004
Score EF	1,743	0.013	0.163	0.364	0.701	0.780	5.974	0.936

## 10.4 Annexe 4 : box-plots.

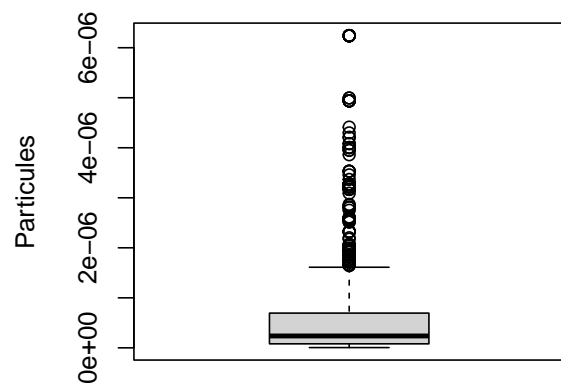
**Boxplot de Rayo. ionisants**



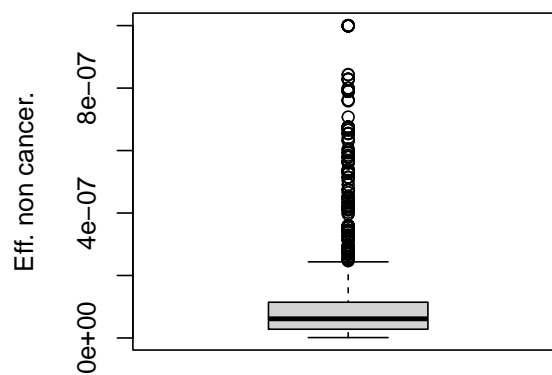
**Boxplot de Forma. photo. ozone**



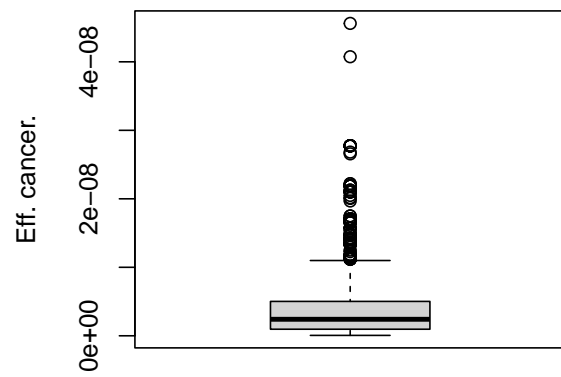
**Boxplot de Particules**



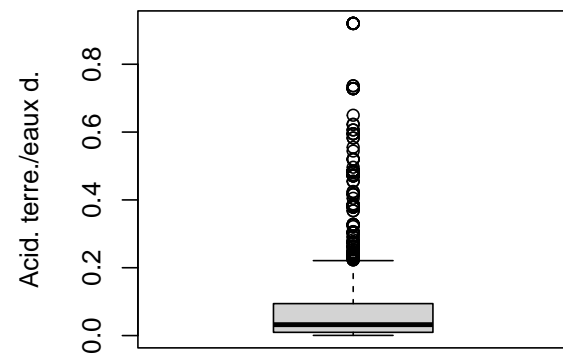
**Boxplot de Eff. non cancer.**



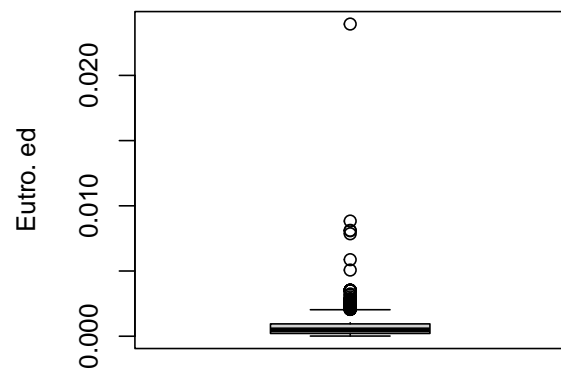
**Boxplot de Eff. cancer.**



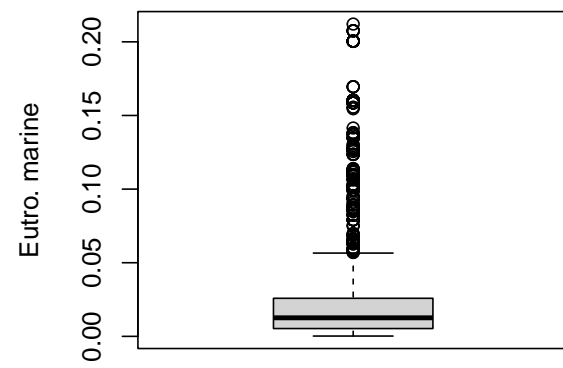
**Boxplot de Acid. terre./eaux d.**



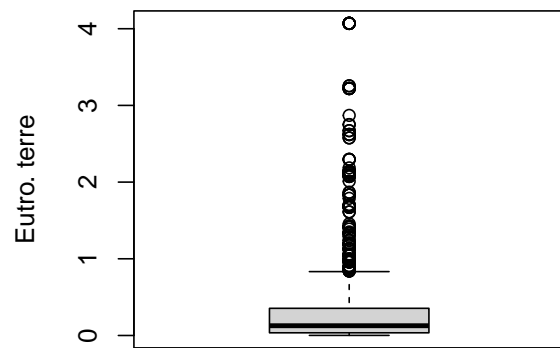
**Boxplot de Eutro. ed**



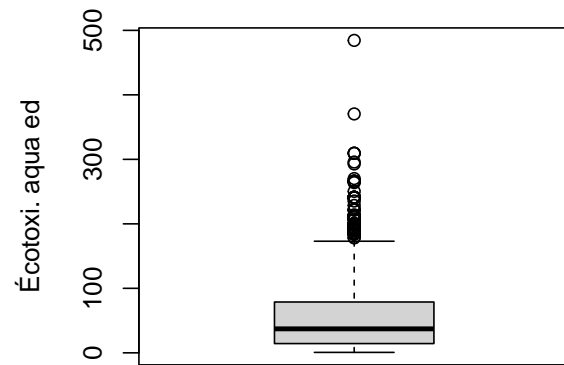
**Boxplot de Eutro. marine**



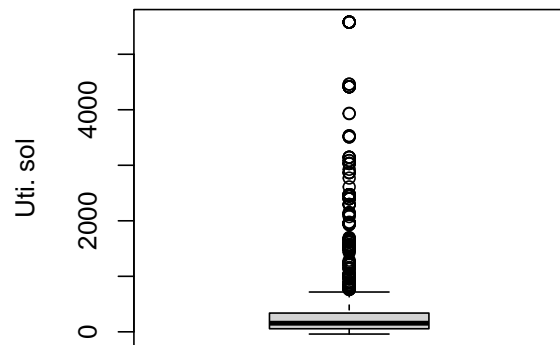
**Boxplot de Eutro. terre**



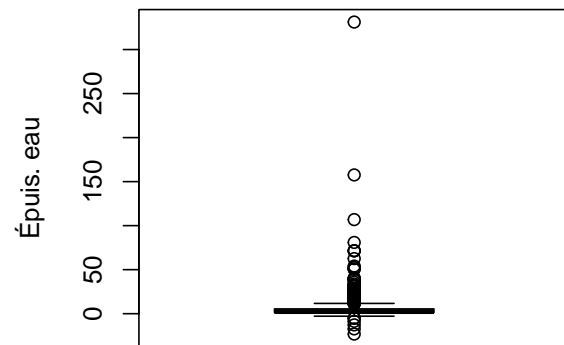
**Boxplot de Écotoxi. aqua ed**



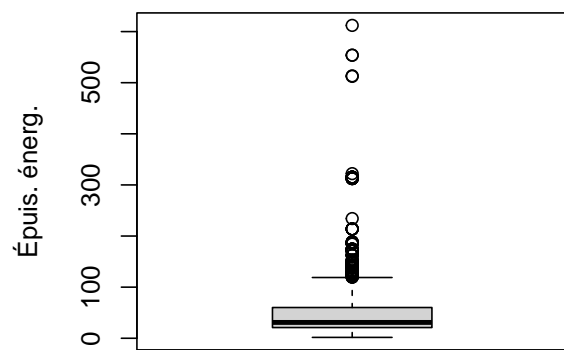
**Boxplot de Uti. sol**



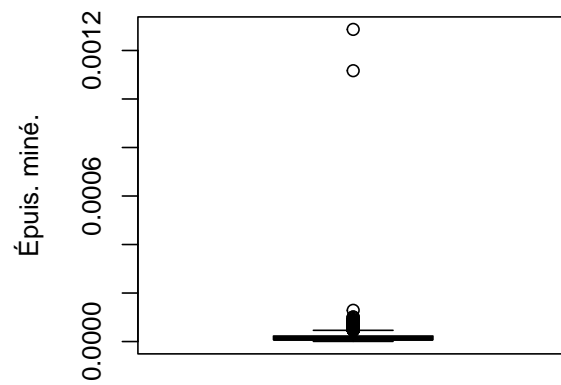
**Boxplot de Épuis. eau**



**Boxplot de Épuis. éner.**

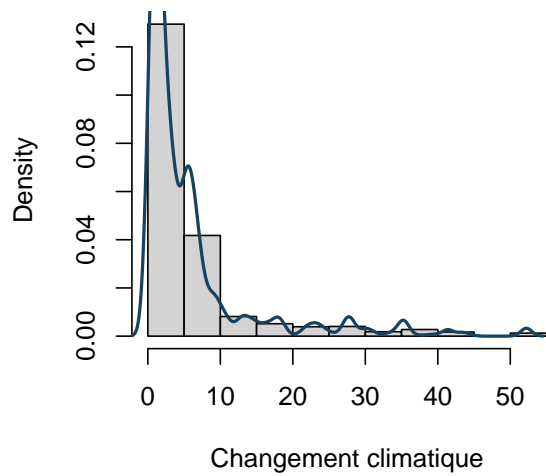


**Boxplot de Épuis. miné.**

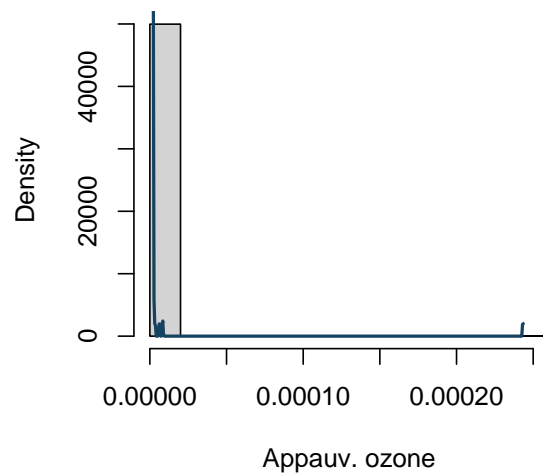


## 10.5 Annexe 5 : histogramme des variables de pollution.

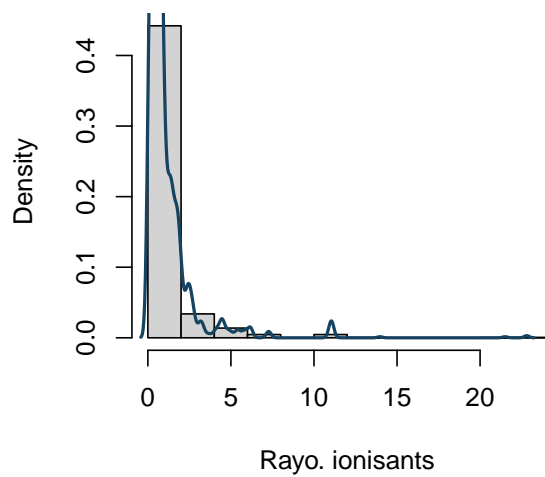
**Histogramme de Changement climatique**



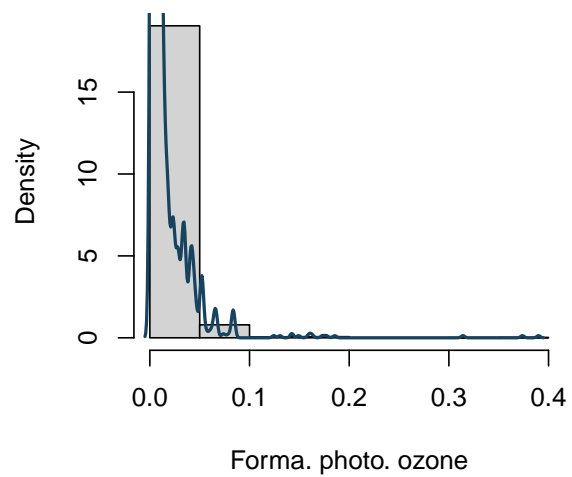
**Histogramme de Appauv. ozone**



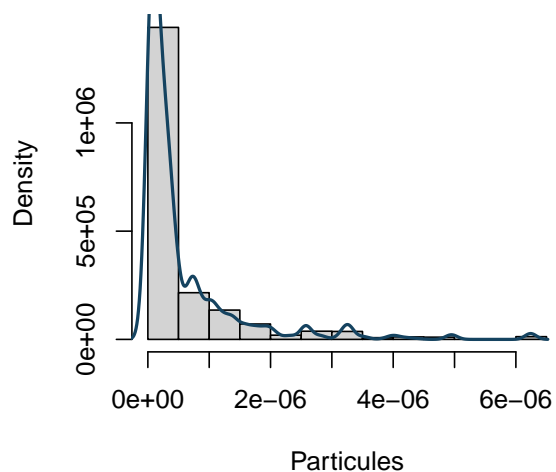
**Histogramme de Rayo. ionisants**



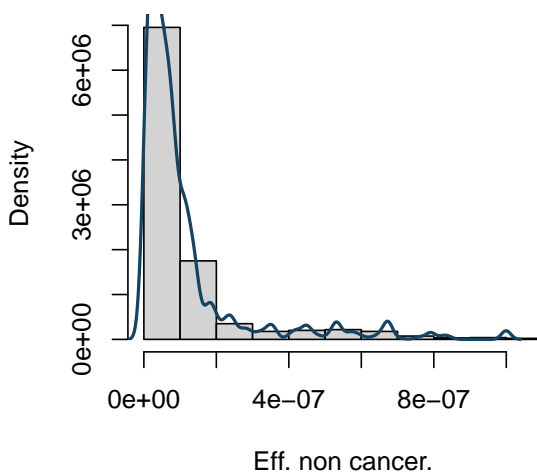
**Histogramme de Forma. photo. ozone**



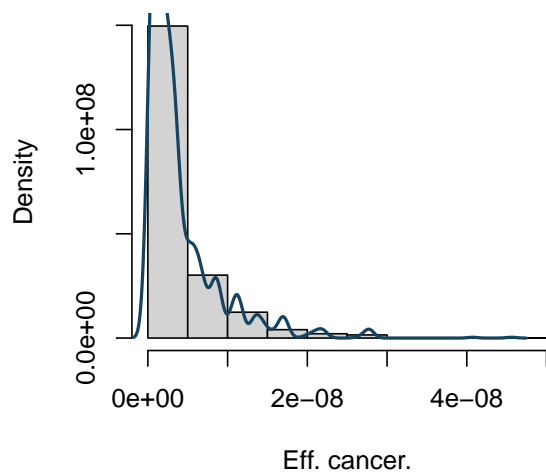
**Histogramme de Particules**



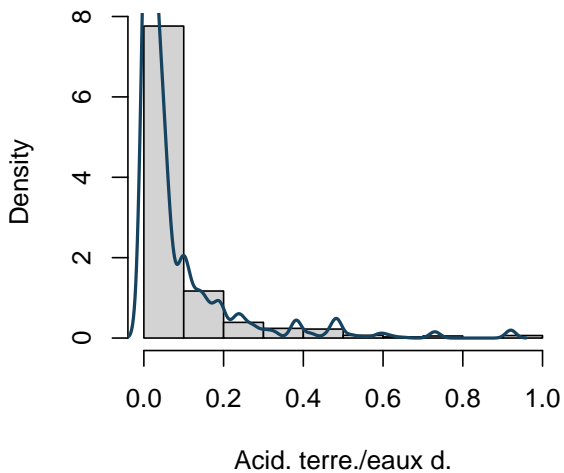
**Histogramme de Eff. non cancer.**



**Histogramme de Eff. cancer.**

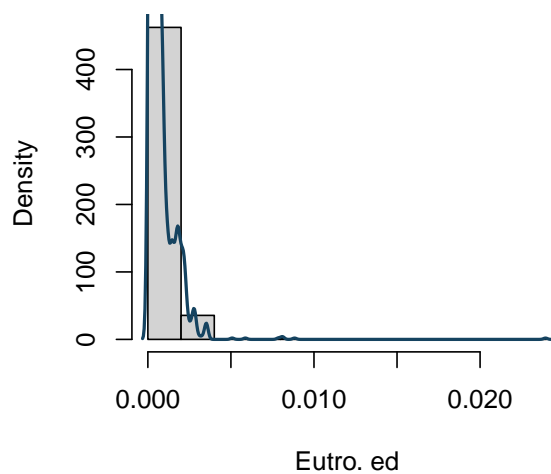


**Histogramme de Acid. terre./eaux d.**

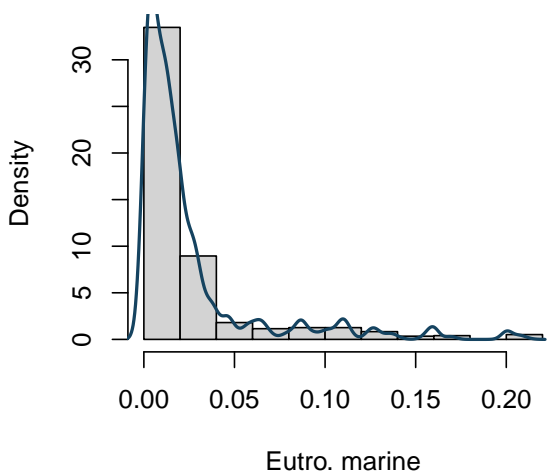




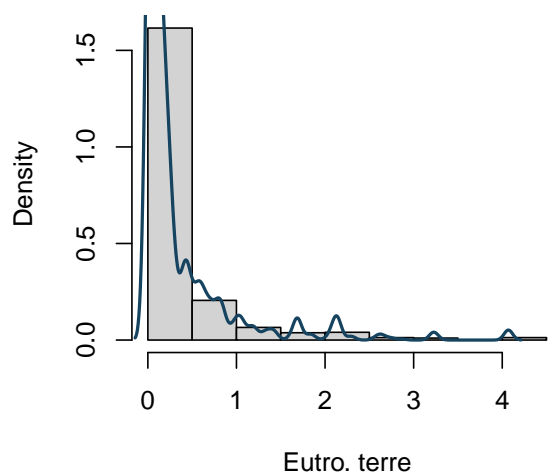
**Histogramme de Eutro. ed**



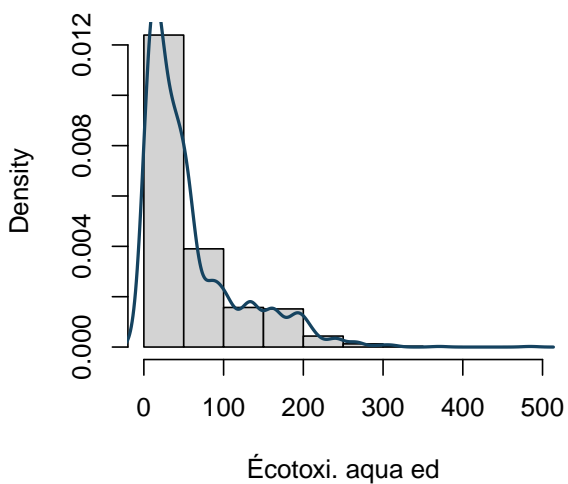
**Histogramme de Eutro. marine**



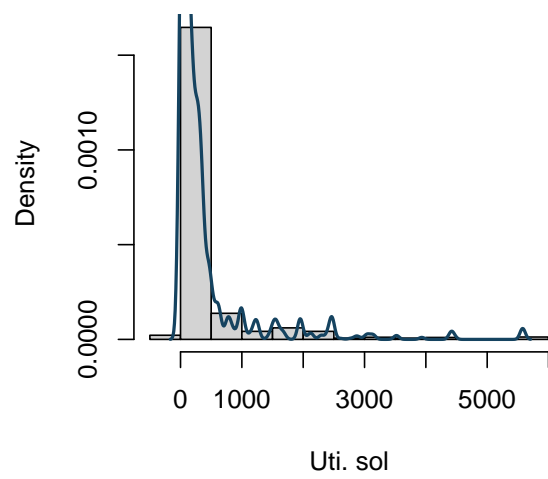
**Histogramme de Eutro. terre**



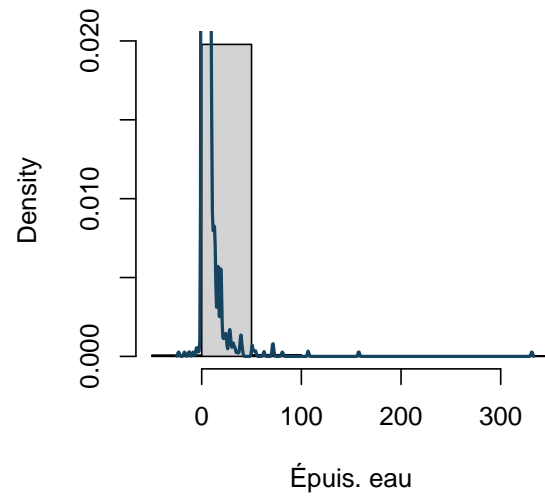
**Histogramme de Écotoxi. aqua ed**



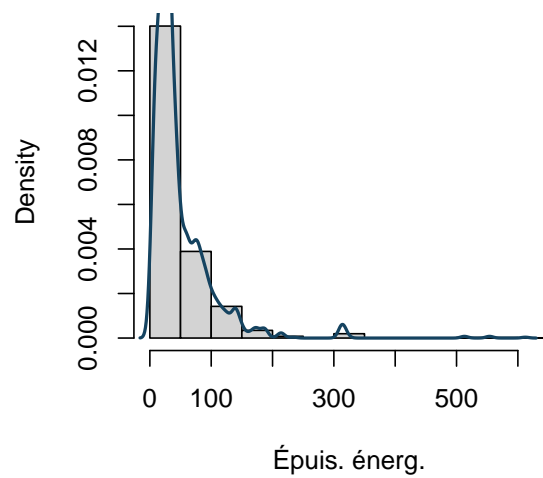
**Histogramme de Uti. sol**



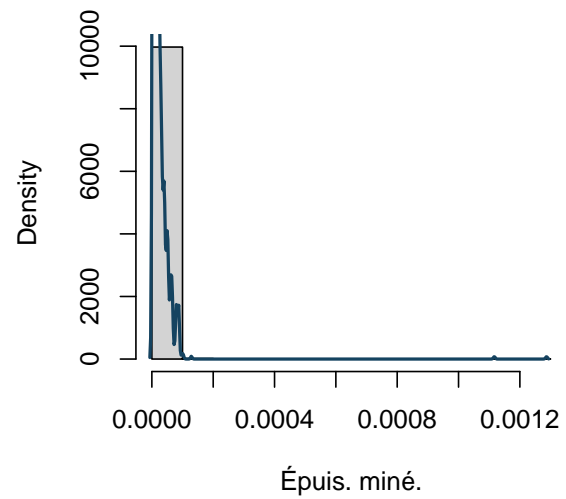
**Histogramme de Épuis. eau**



**Histogramme de Épuis. éner.**



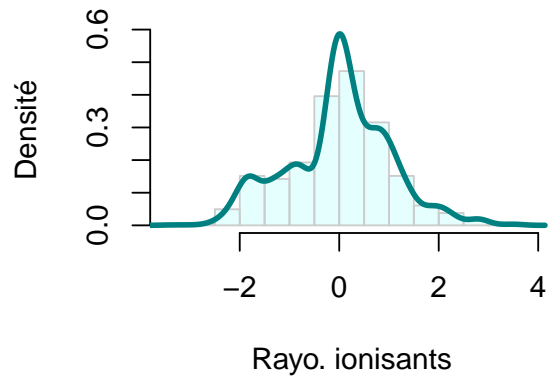
**Histogramme de Épuis. miné.**



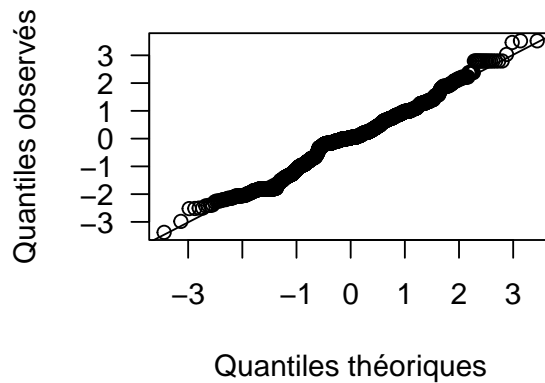
## 10.6 Annexe 6 : histogramme et droite de Henry des transformations logarithmiques.

Fonction myplot reprise et modifié du cours du *Pr Lobry Jean R.* (Ajouter lien)

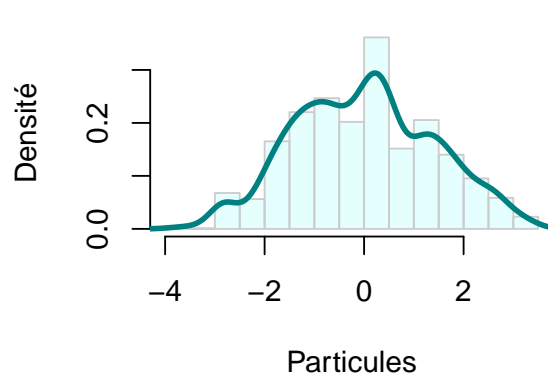
**Histogramme de " Rayo. ionisants ".  
Transformation logarithmique.**



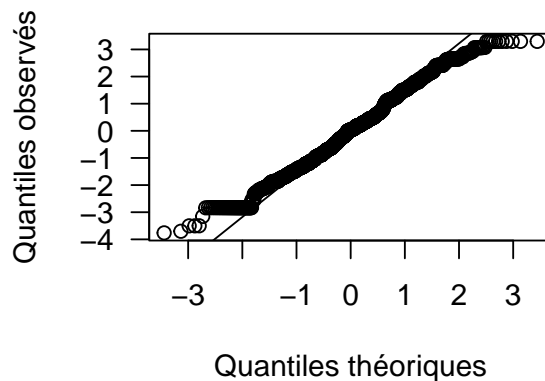
**Droite de Henry**



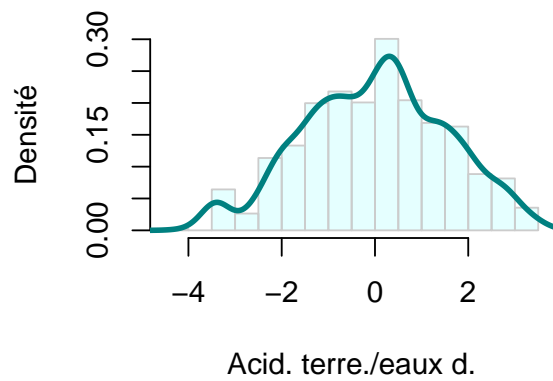
**Histogramme de " Particules ".  
Transformation logarithmique.**



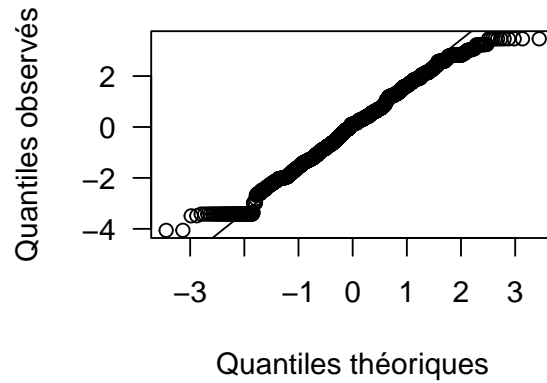
**Droite de Henry**



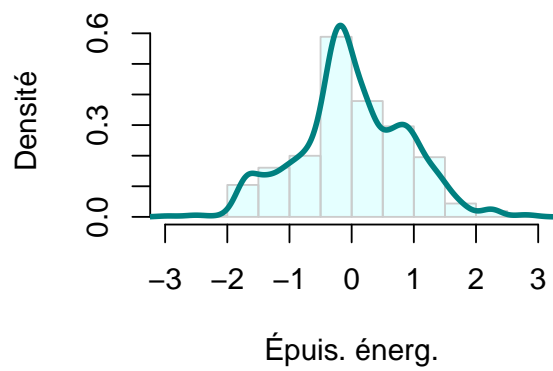
Histogramme de " Acid. terre./eaux d. ".  
Transformation logarithmique.



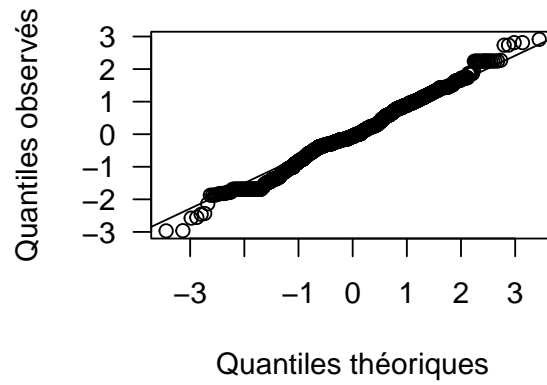
Droite de Henry



Histogramme de " Épuis. éner. ".  
Transformation logarithmique.

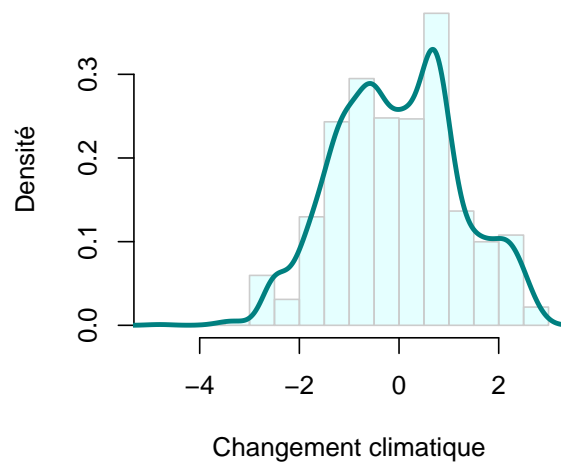


Droite de Henry

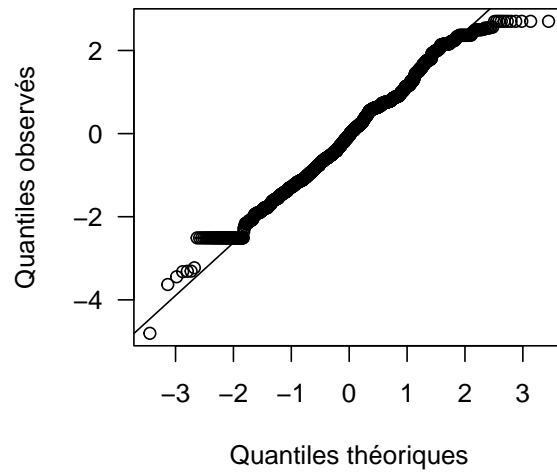


## 10.7 Annexe 7 : histogramme et droite de Henry des transformations de Box-Cox.

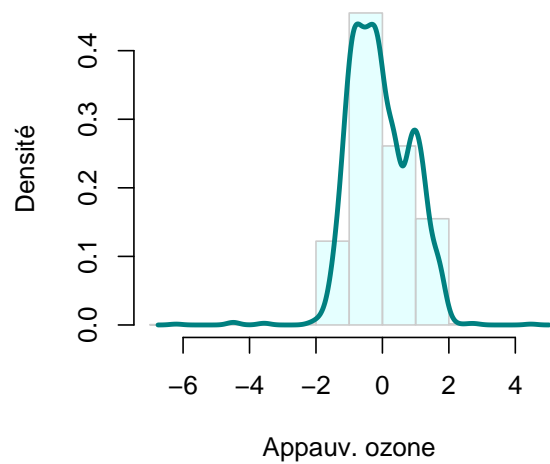
Histogramme de " Changement climatique ".  
Transformation de Box-Cox.



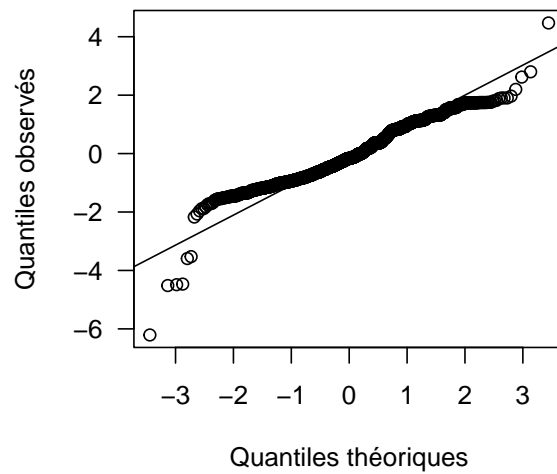
Droite de Henry



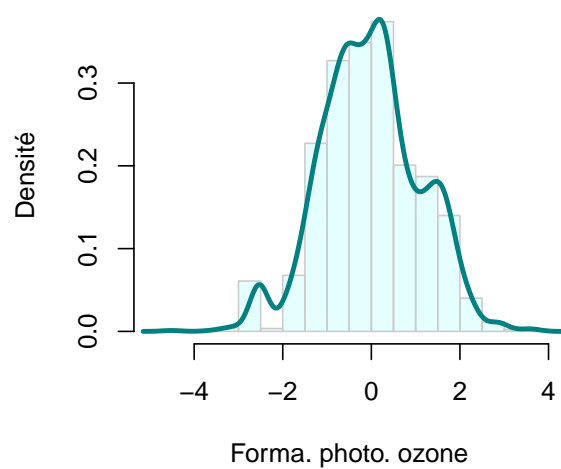
Histogramme de " Appauv. ozone ".  
Transformation de Box-Cox.



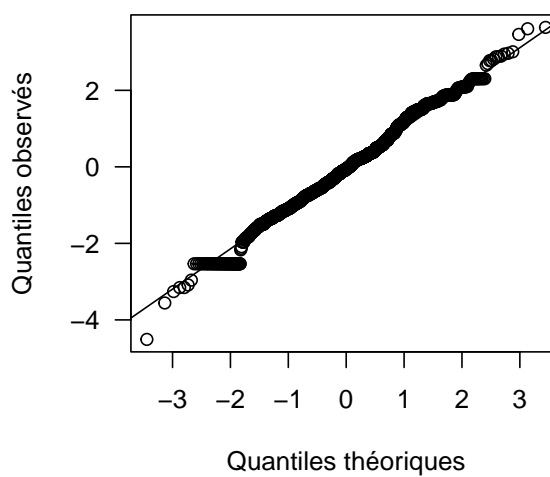
Droite de Henry



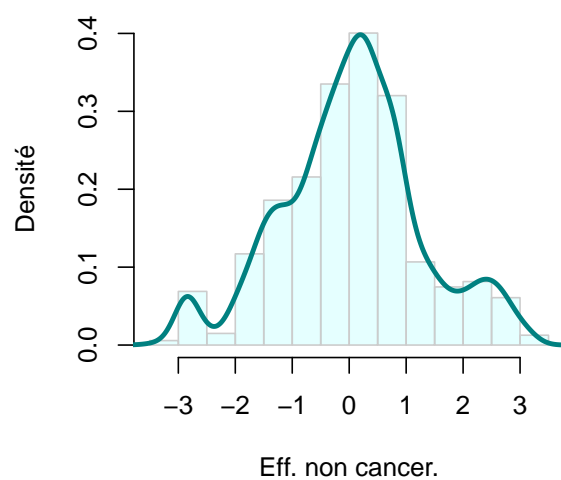
Histogramme de " Forma. photo. ozone ".  
Transformation de Box-Cox.



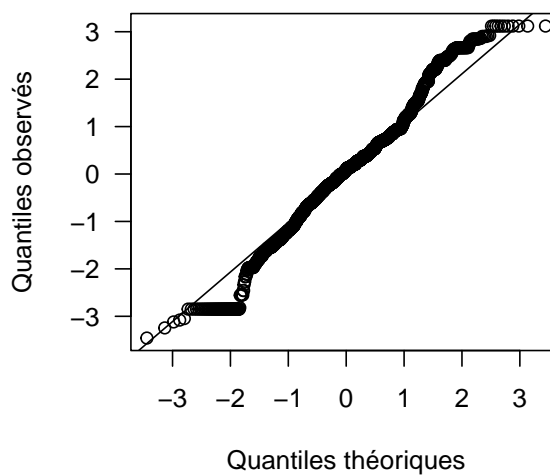
Droite de Henry



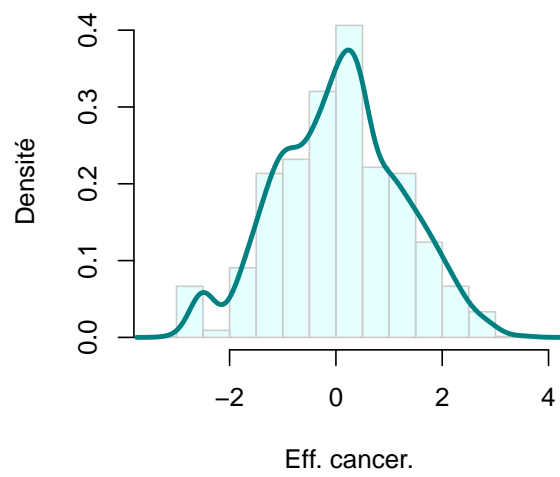
Histogramme de " Eff. non cancer. ".  
Transformation de Box-Cox.



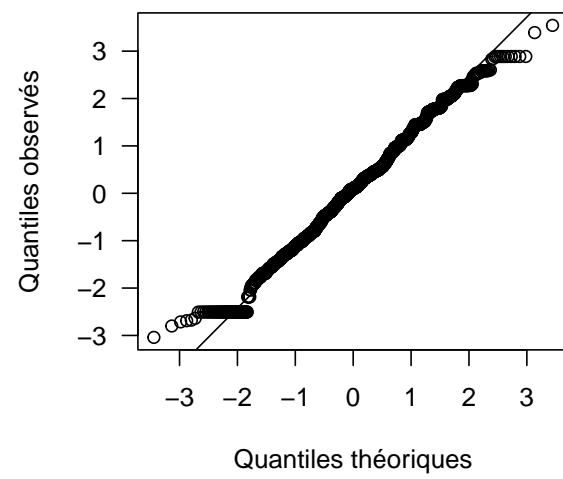
Droite de Henry



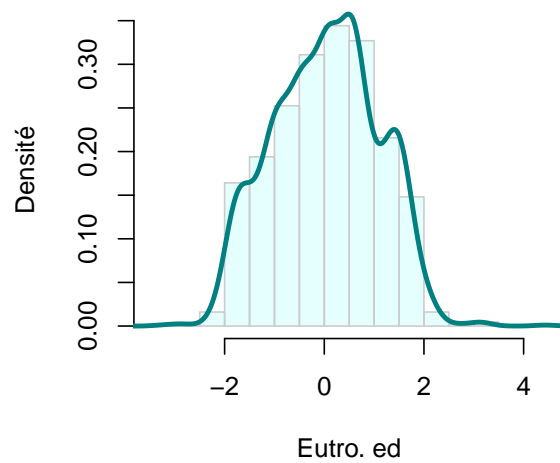
Histogramme de " Eff. cancer. ".  
Transformation de Box-Cox.



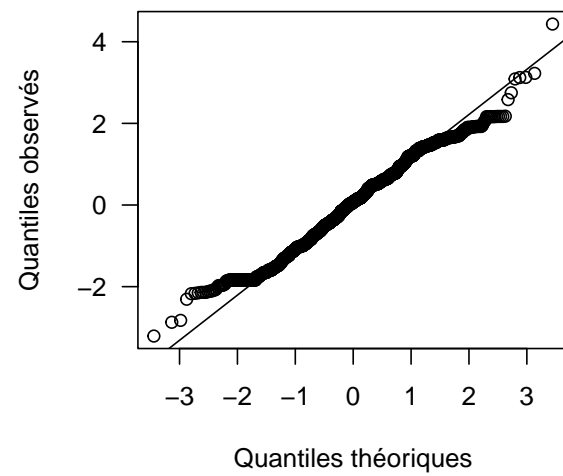
Droite de Henry

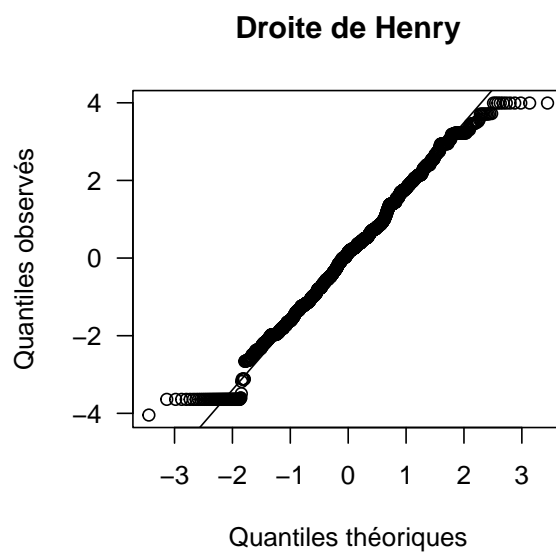
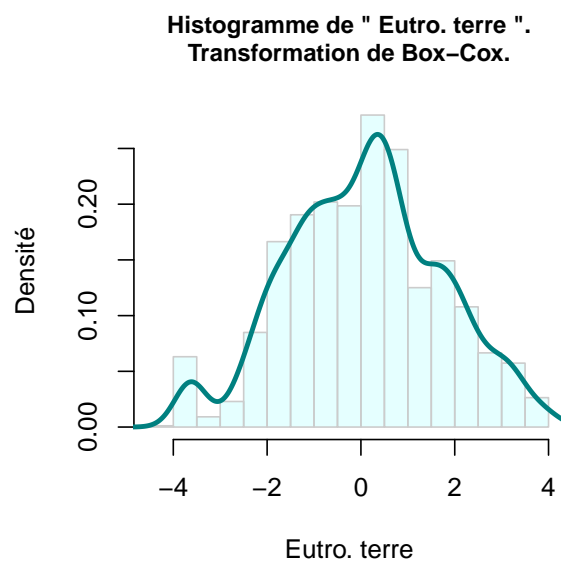
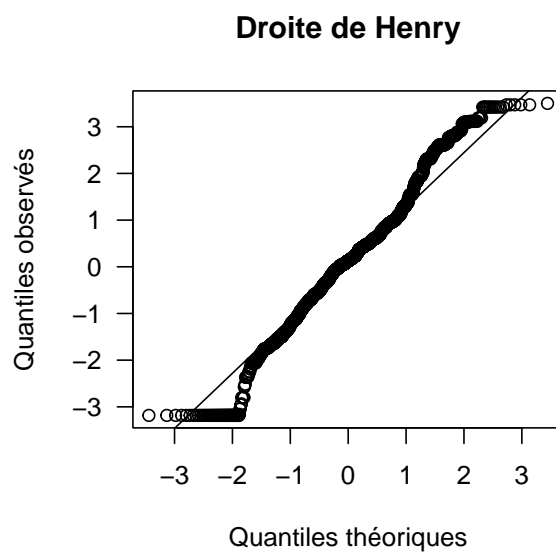
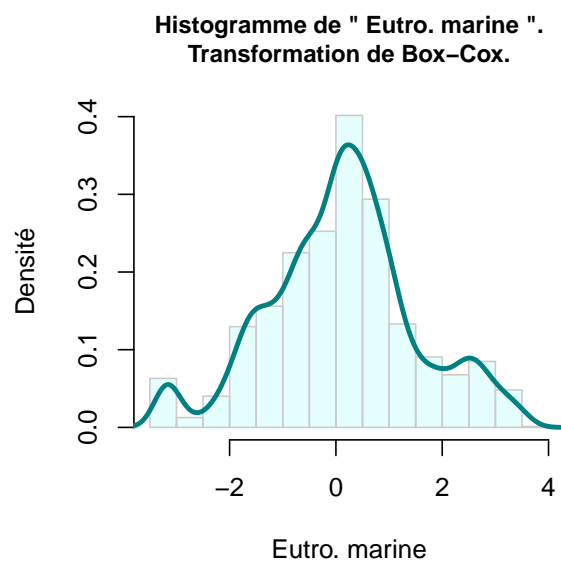


Histogramme de " Eutro. ed ".  
Transformation de Box-Cox.



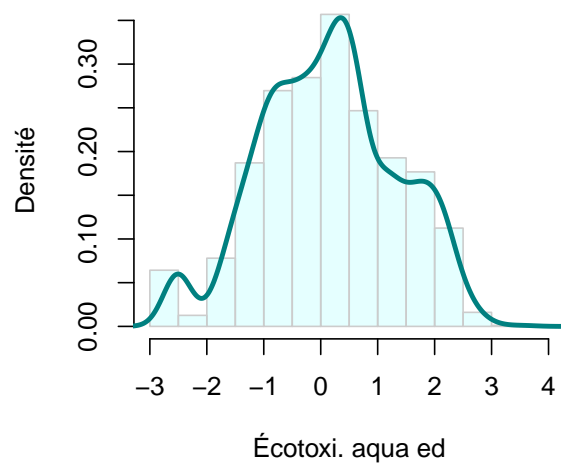
Droite de Henry



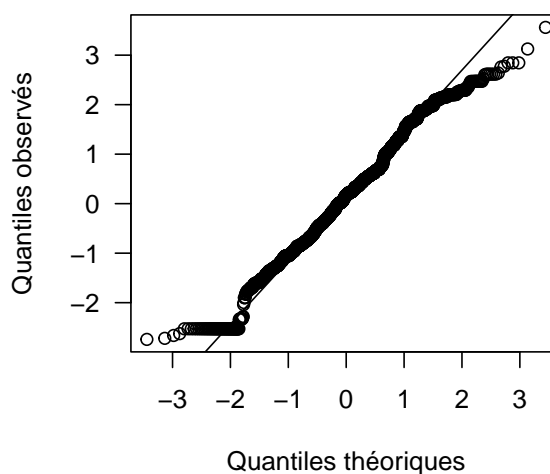




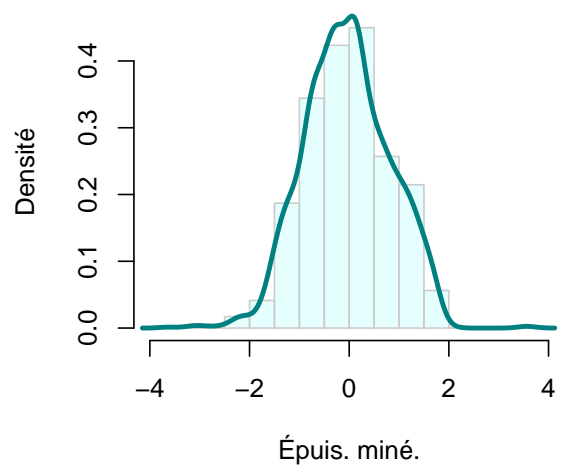
Histogramme de " Écotoxi. aqua ed ".  
Transformation de Box-Cox.



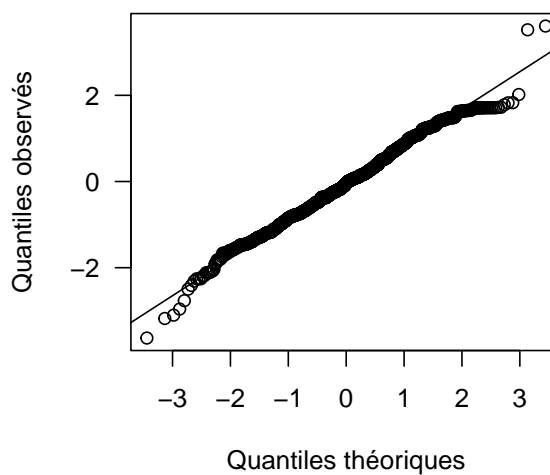
Droite de Henry



Histogramme de " Épuis. miné. ".  
Transformation de Box-Cox.



Droite de Henry



## 10.8 Annexe 8 : tests de Shapiro-Wilks.

### Test sur la variable transformée “Changement climatique” :

---

Shapiro–Wilk normality test

data: Changement climatique W = 0.99109, p-value = 7.958e-09

---

### Test sur la variable transformée “Appauv. ozone” :

---

Shapiro–Wilk normality test

data: Appauv. ozone W = 0.96169, p-value < 2.2e-16

---

### Test sur la variable transformée “Forma. photo. ozone” :

---

Shapiro–Wilk normality test

data: Forma. photo. ozone W = 0.99364, p-value = 8.118e-07

---

### Test sur la variable transformée “Eff. non cancer.” :

---

Shapiro–Wilk normality test

data: Eff. non cancer. W = 0.98689, p-value = 1.711e-11

---

### Test sur la variable transformée “Eff. cancer.” :

---

Shapiro–Wilk normality test

data: Eff. cancer. W = 0.99455, p-value = 5.24e-06

---

### Test sur la variable transformée “Eutro. ed” :

---

Shapiro–Wilk normality test

data: Eutro. ed W = 0.9881, p-value = 8.755e-11

---

### Test sur la variable transformée “Eutro. marine” :

---

Shapiro–Wilk normality test

data: Eutro. marine  $W = 0.98603$ ,  $p\text{-value} = 5.644\text{e-}12$

---

**Test sur la variable transformée “Eutro. terre” :**

---

Shapiro–Wilk normality test

data: Eutro. terre  $W = 0.99236$ ,  $p\text{-value} = 7.116\text{e-}08$

---

**Test sur la variable transformée “Écotoxi. aqua ed” :**

---

Shapiro–Wilk normality test

data: Écotoxi. aqua ed  $W = 0.98909$ ,  $p\text{-value} = 3.578\text{e-}10$

---

**Test sur la variable transformée “Épuis. miné.” :**

---

Shapiro–Wilk normality test

data: Épuis. miné.  $W = 0.99326$ ,  $p\text{-value} = 3.811\text{e-}07$

---

## 10.9 Annexe 9 : nombre de modalités.

```
## 'data.frame':    1743 obs. of  7 variables:
## $ Groupe aliment      : Factor w/ 11 levels "aides culinaires et
    ingrédients divers",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Sous groupe aliment: Factor w/ 55 levels "aides culinaires",...: 2 2 2 27
    27 47 2 48 2 27 ...
## $ Saisonnalité        : Factor w/ 3 levels "0","1","2": 3 3 3 3 3 3 3 3 3 3
    ...
## $ Transport par avion: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ Livraison           : Factor w/ 5 levels "Ambiant (court)",...: 2 2 2 2 2
    5 2 2 2 2 ...
## $ Matériau emballage  : Factor w/ 17 levels "Acier","Aluminium",...: 8 8 8 3
    3 16 8 16 8 3 ...
## $ Préparation        : Factor w/ 7 levels "Cuisson à l'eau",...: 5 5 5 5 5
    5 5 5 5 5 ...
```

## 10.10 Annexe 10 : tableaux des fréquences.

Le nombre de modalités étant trop grand pour la variables "Sous groupe aliment" on ne peut pas l'éditer dans un tableau.

Table 14: Tableau de fréquence de "Groupe aliment"

	n	%	val%
aides culinaires et ingrédients divers	98	5.6	5.6
aliments infantiles	21	1.2	1.2
boissons	138	7.9	7.9
entrées et plats composés	261	15.0	15.0
fruits, légumes, légumineuses et oléagineux	298	17.1	17.1
glaces et sorbets	2	0.1	0.1
lait et produits laitiers	182	10.4	10.4
matières grasses	28	1.6	1.6
produits céréaliers	282	16.2	16.2
produits sucrés	30	1.7	1.7
viandes, œufs, poissons	403	23.1	23.1

Table 15: Tableau des fréquences de "Saisonnalité"

	n	%	val%
0	2	0.1	0.1
1	2	0.1	0.1
2	1739	99.8	99.8

Table 16: Tableau des fréquences de "Transport par avion"

	n	%	val%
0	1741	99.9	99.9
1	2	0.1	0.1

Table 17: Tableau des fréquences de "Livraison"

	n	%	val%
Ambiant (court)	71	4.1	4.1
Ambiant (long)	537	30.8	30.8
Ambiant (moyen)	206	11.8	11.8
Congelé	52	3.0	3.0
Glacé	877	50.3	50.3

Table 18: Tableau des fréquences de "Matériau emballage"

	n	%	val%
Acier	66	3.8	3.8
Aluminium	9	0.5	0.5
Carton	247	14.2	14.2
Carton et Aluminium	2	0.1	0.1
Carton et LDPE	1	0.1	0.1
HDPE	2	0.1	0.1
HPDE	18	1.0	1.0
LDPE	342	19.6	19.6
Papier	32	1.8	1.8
Pas d'emballage	139	8.0	8.0
PET	164	9.4	9.4
PP	240	13.8	13.8
PP/PE	27	1.5	1.5
PPE/PE	1	0.1	0.1
PS	341	19.6	19.6
PVC	65	3.7	3.7
Verre	47	2.7	2.7

Table 19: Tableau des fréquences de "Préparation"

	n	%	val%
Cuisson à l'eau	81	4.6	4.6
Four	237	13.6	13.6
Friture	1	0.1	0.1
Micro-onde	199	11.4	11.4
Pas de préparation	1162	66.7	66.7
Poêle	39	2.2	2.2
Réfrigéré chez le consommateur	24	1.4	1.4

## 10.11 Annexe 11 : pollution des modalités rares.

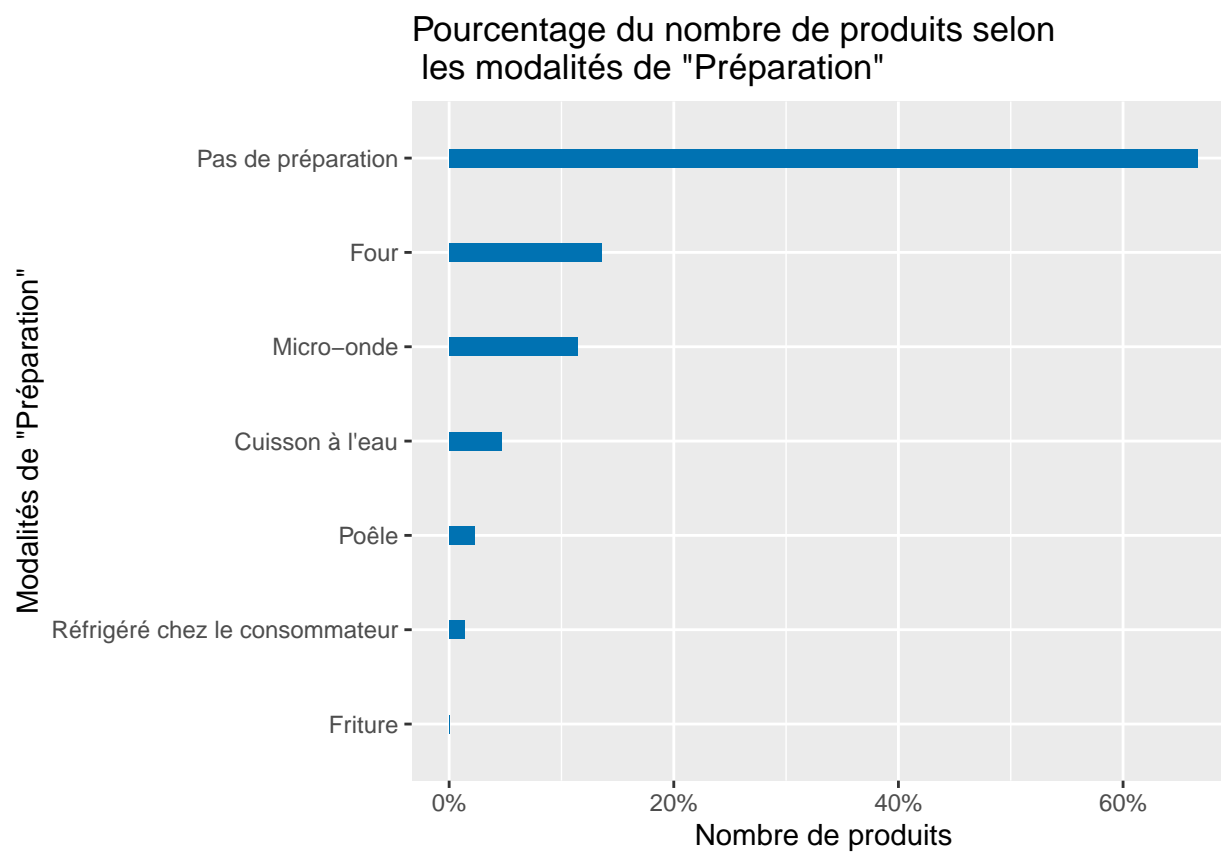
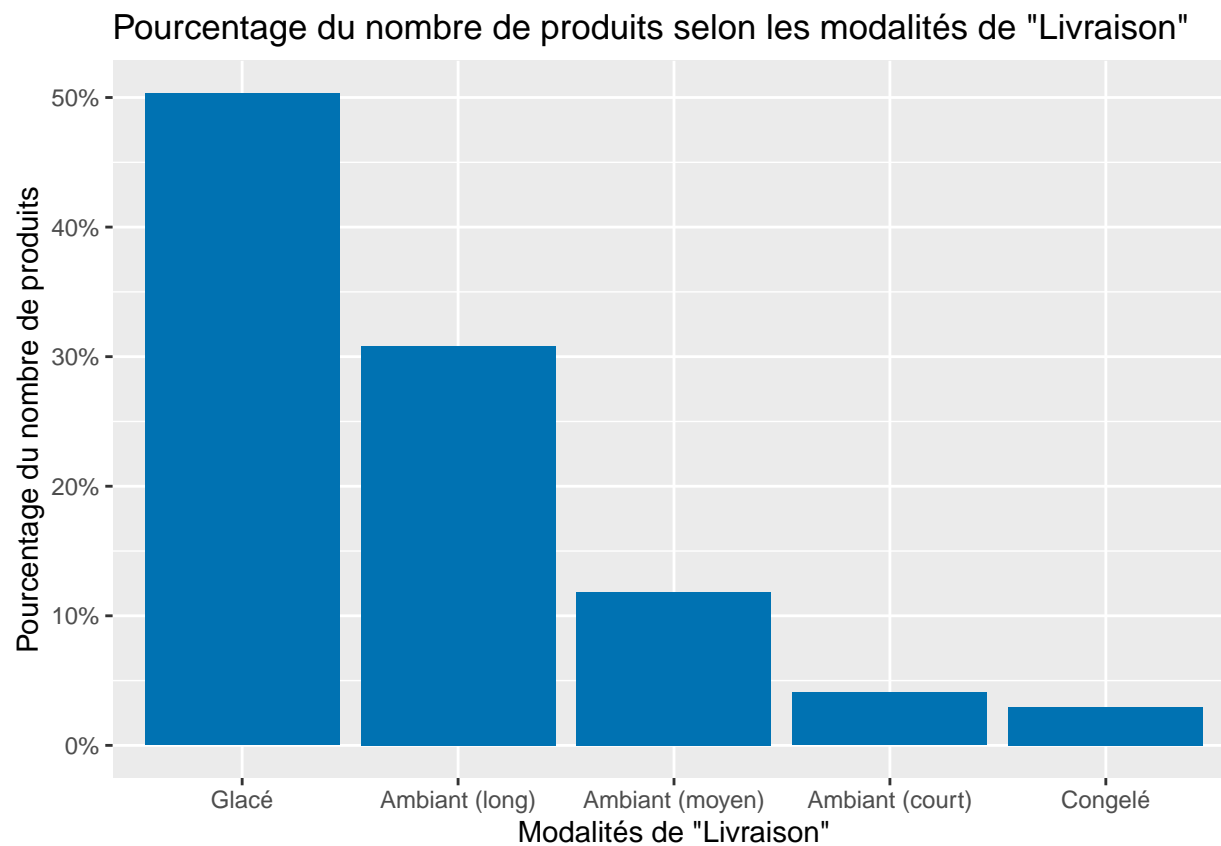
Table 20: Pollution des modalités rares

	Changement climatique	Appauv. ozone	Rayo. ionisants	Forma. photo. ozone	Particules	Eff. non cancer.	Eff. cancer.	Acid. terre./eaux d.	Eutro. ed	Eutro. marine	Eutro. terre	Écotoxi. aqua ed	Uti. sol	Épuis. eau	Épuis. énerg.	Épuis. miné.	Score EF
Fraise hors saison, crue	0.5503879	1e-07	0.1107712	0.0030853	1e-07	0	0	0.0076556	0.0001030	0.0052574	0.0308022	11.676213	114.054430	6.887865	8.764672	3.50e-06	0.1401284
Fraise de saison, crue	0.4785599	1e-07	0.1038577	0.0022991	1e-07	0	0	0.0071888	0.0000987	0.0050011	0.0290570	11.327946	62.517971	6.885207	7.631611	3.20e-06	0.1287045
Tomate hors saison, crue	1.9609115	3e-07	0.1797814	0.0044287	1e-07	0	0	0.0081612	0.0004115	0.0028475	0.0238541	18.298315	6.720441	2.613154	28.951932	1.07e-05	0.1887158
Tomate de saison, crue	0.5828724	1e-07	0.1077282	0.0016345	0e+00	0	0	0.0028253	0.0000720	0.0014082	0.0100447	3.633278	-2.266646	1.893178	8.724414	4.30e-06	0.0675736





## 10.12 Annexe 12 : diagramme en barre des variables qualitatives.



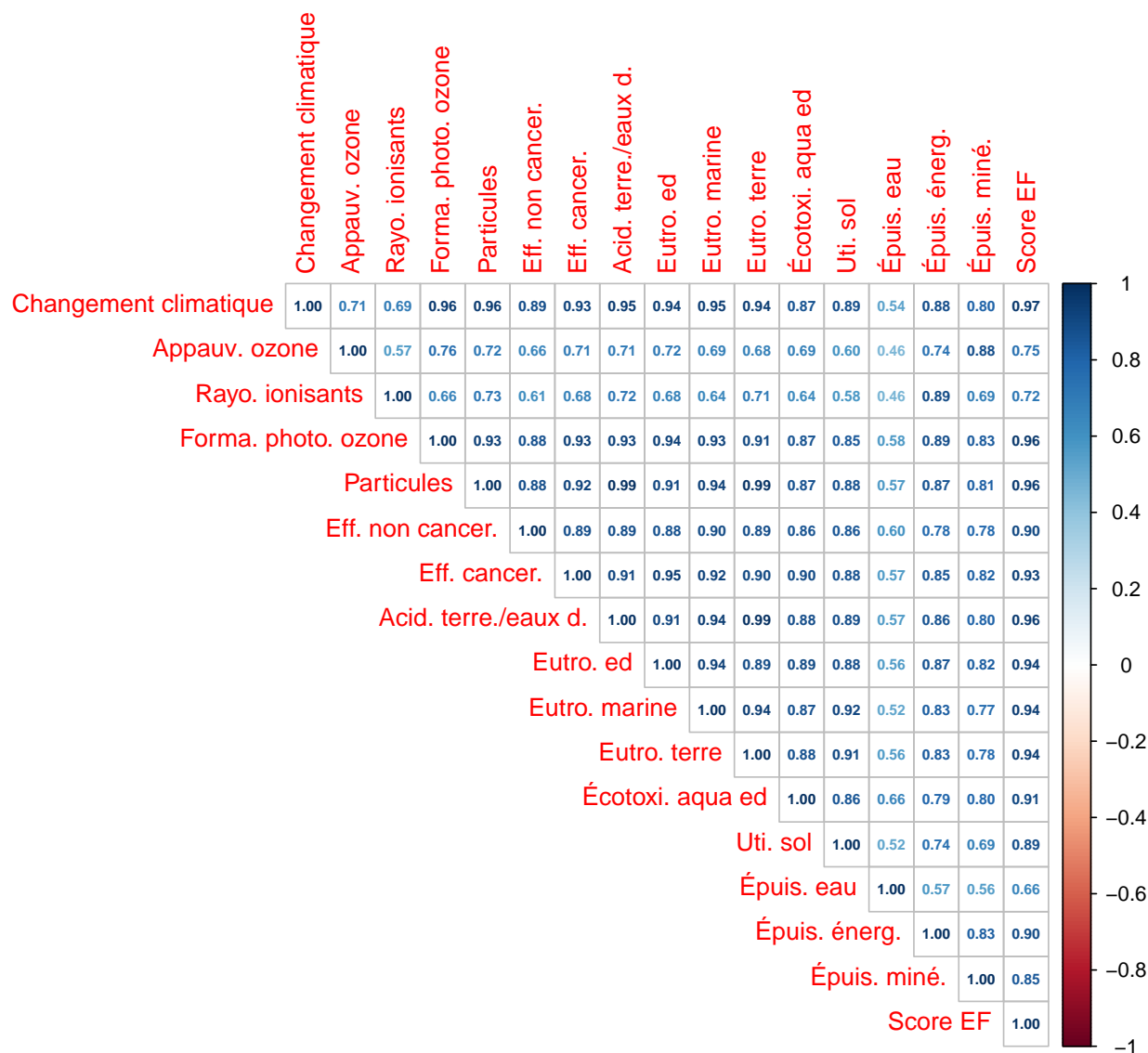
### 10.13 Annexe 13 : nuages de points des corrélations.



Figure 2: Nuages de points entre les variables

## 10.14 Annexe 14 : coefficients de corrélation de Spearman.

**Tableau des corrélations de Spearman**



## 10.15 Annexe 15 : tableau des variables quantitatives des produits 52 et 656

Table 21: Table quantitative des produits 52 et 656

	Changement climatique	Appauv. ozone	Rayo. ionisants	Forma. photo. ozone	Particules	Eff. non cancer.	Eff. cancer.	Acid. terre./eaux d.	Eutro. ed	Eutro. marine	Eutro. terre	Écotoxi. aqua ed	Util. sol	Épuis. eau	Épuis. énerg.	Épuis. miné.	Score EF	DQR
Sauce au poivre, condimentaire, froide, préemballée	1.666308	2e-07	0.8496285	0.0049490	1e-07	0	0	0.0134942	0.0002428	0.0046722	0.0526539	16.764912	63.08010	0.5682808	28.25513	1.43e-05	0.1974744	2.14
Haricot vert, surgelé, cru	0.704000	1e-07	1.0269408	0.0020435	1e-07	0	0	0.0077222	0.0001720	0.0032351	0.0298410	7.694813	86.97883	6.7631934	27.69177	7.80e-06	0.1750000	2.59

10.16 Annexe 16 : image des coefficients de pondération pour le “Score EF”.

**The recommended weighting set, robustness factors and final weighting factors for all midpoint impact categories<sup>1</sup>**

	Aggregated weighting set	Robustness factors	Intermediate Coefficients	Final weighting factors (incl. robustness)
	(A)	(B)	C=A*B	C scaled to 100
Climate change	12.90	0.87	11.18	<b>21.06</b>
Ozone depletion	5.58	0.60	3.35	<b>6.31</b>
Human toxicity, cancer effects	6.80	0.17	1.13	<b>2.13</b>
Human toxicity, non-cancer effects	5.88	0.17	0.98	<b>1.84</b>
Particulate matter	5.49	0.87	4.76	<b>8.96</b>
Ionizing radiation, human health	5.70	0.47	2.66	<b>5.01</b>
Photochemical ozone formation, human health	4.76	0.53	2.54	<b>4.78</b>
Acidification	4.94	0.67	3.29	<b>6.20</b>
Eutrophication, terrestrial	2.95	0.67	1.97	<b>3.71</b>
Eutrophication, freshwater	3.19	0.47	1.49	<b>2.80</b>
Eutrophication, marine	2.94	0.53	1.57	<b>2.96</b>
Ecotoxicity freshwater	6.12	0.17	1.02	<b>1.92</b>
Land use	9.04	0.47	4.22	<b>7.94</b>
Water use	9.69	0.47	4.52	<b>8.51</b>
Resource use, minerals and metals	6.68	0.60	4.01	<b>7.55</b>
Resource use, fossils	7.37	0.60	4.42	<b>8.32</b>

Figure 3: Tableau des pondérations (Agribalyse, 2022)



## 10.17 Annexe 17 : image des coefficient de Normalisation “Score EF”.

Impact category	NFs in EF3.0	NFs in 3.1 interim for AGB
Climate change [kg CO <sub>2</sub> eq./person]	8.10E+03	7.55E+03
Ozone depletion [kg CFC-11 eq./person]	5.36E-02	5.36E-02
Human toxicity, cancer [CTUh/person]	1.69E-05	1.73E-05
Human toxicity, non-cancer [CTUh /person]	2.30E-04	1.29E-04
Particulate matter [disease incidences/person]	5.95E-04	5.95E-04
Ionising radiation, human health [kBq U235 eq./person]	4.22E+03	4.22E+03
Photochemical ozone formation - human health [kg NMVOC eq./person]	4.06E+01	4.06E+01
Acidification [mol H <sup>+</sup> eq./person]	5.56E+01	5.56E+01
Eutrophication, terrestrial [mol N eq./person]	1.77E+02	1.77E+02
Eutrophication, freshwater [kg P eq./person]	1.61E+00	1.61E+00
Eutrophication, marine [kg N eq./person]	1.95E+01	1.95E+01
Ecotoxicity, freshwater [CTUe/person]	4.27E+04	5.67E+04
Land use [pt/person]	8.19E+05	8.19E+05
Water use [m <sup>3</sup> water eq. of deprived water/person]	1.15E+04	1.15E+04
Resource use, minerals and metals [kg Sb eq./person]	6.36E-02	6.36E-02
Resource use, fossils [MJ/person]	6.50E+04	6.50E+04

Figure 4: Tableau des coefficients de normalisation pour la version 3.0 et 3.1 (European Commission, 2022)

## 10.18 Annexe 18 : tableau des coefficients.

Table 22: Table des coefficients "à la main" et de régression

	coeffs	coeff_rlm
'Changement climatique'	2.788270e-02	2.789440e-02
'Appauv. ozone'	1.176186e+03	1.176154e+03
'Rayo. ionisants'	1.187160e-02	1.187800e-02
'Forma. photo. ozone'	1.177299e+00	1.177507e+00
Particules	1.504904e+05	1.505560e+05
'Eff. non cancer.'	1.429285e+05	1.429439e+05
'Eff. cancer.'	1.234575e+06	1.233368e+06
'Acid. terre./eaux d.'	1.115719e+00	1.115620e+00
'Eutro. ed'	1.742537e+01	1.742448e+01
'Eutro. marine'	1.514440e+00	1.514274e+00
'Eutro. terre'	2.098951e-01	2.099463e-01
'Écotoxi. aqua ed'	3.385000e-04	3.385000e-04
'Uti. sol'	9.690000e-05	9.690000e-05
'Épuis. eau'	7.420200e-03	7.419700e-03
'Épuis. énerg.'	1.279900e-03	1.279500e-03
'Épuis. miné.'	1.186356e+03	1.186165e+03

**10.19 Annexe 19 : table des variables quantitatives des produits 51 et 52.**

	Changement climatique	Appauv. ozone	Rayo. ionisants	Forma. photo. ozone	Particules	Eff. non cancer.	Eff. cancer.	Acid. terre./eaux d.	Eutro. ed	Eutro. marine	Eutro. terre	Écotoxi. aqua ed	Uti. sol	Épuis. eau	Épuis. éner.	Épuis. miné.	Score EF
Sauce au poivre, chaude, préemballée	1.677353	2e-07	0.9194510	0.0049808	1e-07	0	0	0.0135611	0.0002466	0.0046869	0.0527683	16.80822	63.15416	0.5733867	29.78333	1.47e-05	0.1974744
Sauce au poivre, condimentaire, froide, préemballée	1.666308	2e-07	0.8496285	0.0049490	1e-07	0	0	0.0134942	0.0002428	0.0046722	0.0526539	16.76491	63.08010	0.5682808	28.25513	1.43e-05	0.1974744

## 10.20 Annexe 20 : tableau de contingence.

Table 23: Tableau de contingence entre les variables "Préparation" et "Livraison".

	Cuisson à l'eau	Four	Friture	Micro-onde	Pas de préparation	Poêle	Réfrigéré chez le consommateur
Ambiant (court)	0	0	0	0	71	0	0
Ambiant (long)	25	3	0	53	448	1	7
Ambiant (moyen)	37	4	0	9	153	1	2
Congelé	4	12	1	7	28	0	0
Glacé	15	218	0	130	462	37	15

## 10.21 Annexe 21 : tableaux des valeurs théoriques pour le test du khi-deux.

```
## Warning in chisq.test(tbl): 'L'approximation du Chi-2 est peut-être  
incorrecte
```

Table 24: Table théorique du khi-deux entre les variables "Préparation" et "Livraison".

	Cuisson à l'eau	Four	Friture	Micro-onde	Pas de préparation	Poêle	Réfrigéré chez le consommateur
Ambiant (court)	3.30	9.65	0.04	8.11	47.33	1.59	0.98
Ambiant (long)	24.96	73.02	0.31	61.31	358.00	12.02	7.39
Ambiant (moyen)	9.57	28.01	0.12	23.52	137.33	4.61	2.84
Congelé	2.42	7.07	0.03	5.94	34.67	1.16	0.72
Glacé	40.76	119.25	0.50	100.13	584.67	19.62	12.08

## 10.22 Annexe 22 : test du khi-deux.

**Test d'indépendance du khi-deux entre les variables "Groupe aliment" et "Livraison":**

---

```
## Warning in chisq.test(table('Groupe aliment', Livraison)): 'L'approximation  
du  
## Chi-2 est peut-être incorrecte  
  
Pearson's Chi-squared test  
  
data: table(Groupe aliment, Livraison) X-squared = 2399.5, df = 40, p-value < 2.2e-16
```

---

**Test d'indépendance du khi-deux entre les variables "Groupe aliment" et "Matériau emballage":**

---

```
## Warning in chisq.test(table('Groupe aliment', 'Matériau emballage')):  
## 'L'approximation du Chi-2 est peut-être incorrecte  
  
Pearson's Chi-squared test  
  
data: table(Groupe aliment, Matériau emballage) X-squared = 4134.7, df = 160, p-value < 2.2e-16
```

---

**Test d'indépendance du khi-deux entre les variables "Groupe aliment" et "Préparation":**

---

```
## Warning in chisq.test(table('Groupe aliment', Préparation)):  
'L'approximation du  
## Chi-2 est peut-être incorrecte  
  
Pearson's Chi-squared test  
  
data: table(Groupe aliment, Préparation) X-squared = 1197.4, df = 60, p-value < 2.2e-16
```

---

**Test d'indépendance du khi-deux entre les variables "Livraison" et "Matériau emballage":**

---

```
## Warning in chisq.test(table(Livraison, 'Matériau emballage')):  
'L'approximation  
## du Chi-2 est peut-être incorrecte  
  
Pearson's Chi-squared test  
  
data: table(Livraison, Matériau emballage) X-squared = 1859.1, df = 64, p-value < 2.2e-16
```

---

**Test d'indépendance du khi-deux entre les variables "Matériau emballage" et "Préparation":**

---

```
## Warning in chisq.test(table('Matériau emballage', Préparation)):  
  'Lapproximation  
## du Chi-2 est peut-être incorrecte  
  
Pearson's Chi-squared test  
  
data: table(Matériau emballage, Préparation) X-squared = 1460.9, df = 96, p-value < 2.2e-16
```

---



## 10.23 Annexe 23 : tableaux des valeurs théoriques du test du khi-deux.

```
## Warning in chisq.test(table(quali$ 'Groupe aliment ', quali$Livraison)) :
## 'L'approximation du Chi-2 est peut-être incorrecte
```

Table 25: Table théorique du khi-deux entre les variables "Groupe aliment" et "Livraison".

	Ambiant (court)	Ambiant (long)	Ambiant (moyen)	Congelé	Glacé
aides culinaires et ingrédients divers	3.99	30.19	11.58	2.92	49.31
aliments infantiles	0.86	6.47	2.48	0.63	10.57
boissons	5.62	42.52	16.31	4.12	69.44
entrées et plats composés	10.63	80.41	30.85	7.79	131.32
fruits, légumes, légumineuses et oléagineux	12.14	91.81	35.22	8.89	149.94
glaces et sorbets	0.08	0.62	0.24	0.06	1.01
lait et produits laitiers	7.41	56.07	21.51	5.43	91.57
matières grasses	1.14	8.63	3.31	0.84	14.09
produits céréaliers	11.49	86.88	33.33	8.41	141.89
produits sucrés	1.22	9.24	3.55	0.90	15.09
viandes, œufs, poissons	16.42	124.16	47.63	12.02	202.77

```
## Warning in chisq.test(table(quali$ 'Groupe aliment ', quali$ 'Matériau
## emballage ')) : 'L'approximation du Chi-2 est peut-être incorrecte
```

Table 26: Table théorique du khi-deux entre les variables "Groupe aliment" et "Matériau emballage".

	Acier	Aluminium	Carton	Carton et Aluminium	Carton et LDPE	HDPE	HDPE	LDPE	Papier	Pas d'emballage	PET	PP	PP/PE	PPE/PE	PS	PVC	Verre
aides culinaires et ingrédients divers	3.71	0.51	13.89	0.11	0.06	0.11	1.01	19.23	1.80	7.82	9.22	13.49	1.52	0.06	19.17	3.65	2.64
aliments infantiles	0.80	0.11	2.98	0.02	0.01	0.02	0.22	4.12	0.39	1.67	1.98	2.89	0.33	0.01	4.11	0.78	0.57
boissons	5.23	0.71	19.56	0.16	0.08	0.16	1.43	27.08	2.53	11.01	12.98	19.00	2.14	0.08	27.00	5.15	3.72
entrées et plats composés	9.88	1.35	36.99	0.30	0.15	0.30	2.70	51.21	4.79	20.81	24.56	35.94	4.04	0.15	51.06	9.73	7.04
fruits, légumes, légumineuses et oléagineux	11.28	1.54	42.23	0.34	0.17	0.34	3.08	58.47	5.47	23.76	28.04	41.03	4.62	0.17	58.30	11.11	8.04
glaces et sorbets	0.08	0.01	0.28	0.00	0.00	0.00	0.02	0.39	0.04	0.16	0.19	0.28	0.03	0.00	0.39	0.07	0.05
lait et produits laitiers	6.89	0.94	25.79	0.21	0.10	0.21	1.88	35.71	3.34	14.51	17.12	25.06	2.82	0.10	35.61	6.79	4.91
matières grasses	1.06	0.14	3.97	0.03	0.02	0.03	0.29	5.49	0.51	2.23	2.63	3.86	0.43	0.02	5.48	1.04	0.76
produits céréaliers	10.68	1.46	39.96	0.32	0.16	0.32	2.91	55.33	5.18	22.49	26.53	38.83	4.37	0.16	55.17	10.52	7.60
produits sucrés	1.14	0.15	4.25	0.03	0.02	0.03	0.31	5.89	0.55	2.39	2.82	4.13	0.46	0.02	5.87	1.12	0.81
viandes, œufs, poissons	15.26	2.08	57.11	0.46	0.23	0.46	4.16	79.07	7.40	32.14	37.92	55.49	6.24	0.23	78.84	15.03	10.87

```
## Warning in chisq.test(table(quali$ 'Groupe aliment ', quali$Préparation)) :
## 'L'approximation du Chi-2 est peut-être incorrecte
```

Table 27: Table théorique du khi-deux entre les variables "Groupe aliment" et "Préparation".

	Cuisson à l'eau	Four	Friture	Micro-onde	Pas de préparation	Poêle	Réfrigéré chez le consommateur
aides culinaires et ingrédients divers	4.55	13.33	0.06	11.19	65.33	2.19	1.35
aliments infantiles	0.98	2.86	0.01	2.40	14.00	0.47	0.29
boissons	6.41	18.76	0.08	15.76	92.00	3.09	1.90
entrées et plats composés	12.13	35.49	0.15	29.80	174.00	5.84	3.59
fruits, légumes, légumineuses et oléagineux	13.85	40.52	0.17	34.02	198.67	6.67	4.10
glaces et sorbets	0.09	0.27	0.00	0.23	1.33	0.04	0.03
lait et produits laitiers	8.46	24.75	0.10	20.78	121.33	4.07	2.51
matières grasses	1.30	3.81	0.02	3.20	18.67	0.63	0.39
produits céréaliers	13.10	38.34	0.16	32.20	188.00	6.31	3.88
produits sucrés	1.39	4.08	0.02	3.43	20.00	0.67	0.41
viandes, œufs, poissons	18.73	54.80	0.23	46.01	268.67	9.02	5.55

```
## Warning in chisq.test(table(quali$Livraison , quali$ 'Matériau emballage ')) :
## 'L'approximation du Chi-2 est peut-être incorrecte
```

Table 28: Table théorique du khi-deux entre les variables "Livraison" et "Matériau emballage".

	Acier	Aluminium	Carton	Carton et Aluminium	Carton et LDPE	HDPE	HPDE	LDPE	Papier	Pas d'emballage	PET	PP	PP/PE	PPE/PE	PS	PVC	Verre
Ambiant (court)	2.69	0.37	10.06	0.08	0.04	0.08	0.73	13.93	1.30	5.66	6.68	9.78	1.10	0.04	13.89	2.65	1.91
Ambiant (long)	20.33	2.77	76.10	0.62	0.31	0.62	5.55	105.37	9.86	42.82	50.53	73.94	8.32	0.31	105.06	20.03	14.48
Ambiant (moyen)	7.80	1.06	29.19	0.24	0.12	0.24	2.13	40.42	3.78	16.43	19.38	28.36	3.19	0.12	40.30	7.68	5.55
Congelé	1.97	0.27	7.37	0.06	0.03	0.06	0.54	10.20	0.95	4.15	4.89	7.16	0.81	0.03	10.17	1.94	1.40
Glacé	33.21	4.53	124.28	1.01	0.50	1.01	9.06	172.08	16.10	69.94	82.52	120.76	13.59	0.50	171.58	32.71	23.65

```
## Warning in chisq.test(table(quali$ 'Matériau emballage', quali$Préparation))
:
## 'L'approximation du Chi-2 est peut-être incorrecte
```

Table 29: Table théorique du khi-deux entre les variables "Matériau emballage" et "Préparation".

	Cuisson à l'eau	Four	Friture	Micro-onde	Pas de préparation	Poêle	Réfrigéré chez le consommateur
Acier	3.07	8.97	0.04	7.54	44.00	1.48	0.91
Aluminium	0.42	1.22	0.01	1.03	6.00	0.20	0.12
Carton	11.48	33.59	0.14	28.20	164.67	5.53	3.40
Carton et Aluminium	0.09	0.27	0.00	0.23	1.33	0.04	0.03
Carton et LDPE	0.05	0.14	0.00	0.11	0.67	0.02	0.01
HDPE	0.09	0.27	0.00	0.23	1.33	0.04	0.03
HPDE	0.84	2.45	0.01	2.06	12.00	0.40	0.25
LDPE	15.89	46.50	0.20	39.05	228.00	7.65	4.71
Papier	1.49	4.35	0.02	3.65	21.33	0.72	0.44
Pas d'emballage	6.46	18.90	0.08	15.87	92.67	3.11	1.91
PET	7.62	22.30	0.09	18.72	109.33	3.67	2.26
PP	11.15	32.63	0.14	27.40	160.00	5.37	3.30
PP/PE	1.25	3.67	0.02	3.08	18.00	0.60	0.37
PPE/PE	0.05	0.14	0.00	0.11	0.67	0.02	0.01
PS	15.85	46.37	0.20	38.93	227.33	7.63	4.70
PVC	3.02	8.84	0.04	7.42	43.33	1.45	0.90
Verre	2.18	6.39	0.03	5.37	31.33	1.05	0.65

## 10.24 Annexe 24 : test du khi-deux avec “simulate p-value”.

**Test d’indépendance du khi-deux avec “simulate.p.value” entre les variables “Groupe aliment” et “Livraison”:**

---

Pearson’s Chi-squared test with simulated p-value (based on 2000 replicates)

data: table(Groupe aliment, Livraison) X-squared = 2399.5, df = NA, p-value = 0.0004998

---

**Test d’indépendance du khi-deux avec “simulate.p.value” entre les variables “Groupe aliment” et “Matériau emballage”:**

---

Pearson’s Chi-squared test with simulated p-value (based on 2000 replicates)

data: table(Groupe aliment, Matériau emballage) X-squared = 4134.7, df = NA, p-value = 0.0004998

---

**Test d’indépendance du khi-deux avec “simulate.p.value” entre les variables “Groupe aliment” et “Préparation”:**

---

Pearson’s Chi-squared test with simulated p-value (based on 2000 replicates)

data: table(Groupe aliment, Préparation) X-squared = 1197.4, df = NA, p-value = 0.0004998

---

**Test d’indépendance du khi-deux avec “simulate.p.value” entre les variables “Livraison” et “Matériau emballage”:**

---

Pearson’s Chi-squared test with simulated p-value (based on 2000 replicates)

data: table(Livraison, Matériau emballage) X-squared = 1859.1, df = NA, p-value = 0.0004998

---

**Test d’indépendance du khi-deux avec “simulate.p.value” entre les variables “Matériau emballage” et “Préparation”:**

---

Pearson’s Chi-squared test with simulated p-value (based on 2000 replicates)

data: table(Matériau emballage, Préparation) X-squared = 1460.9, df = NA, p-value = 0.002499

---

**10.25 Annexe 25 : matrice des corrélations entre variables qualitatives et quantitatives.**

Table 30: Matrice des corrélation entre variables qualitatives et quantitatives

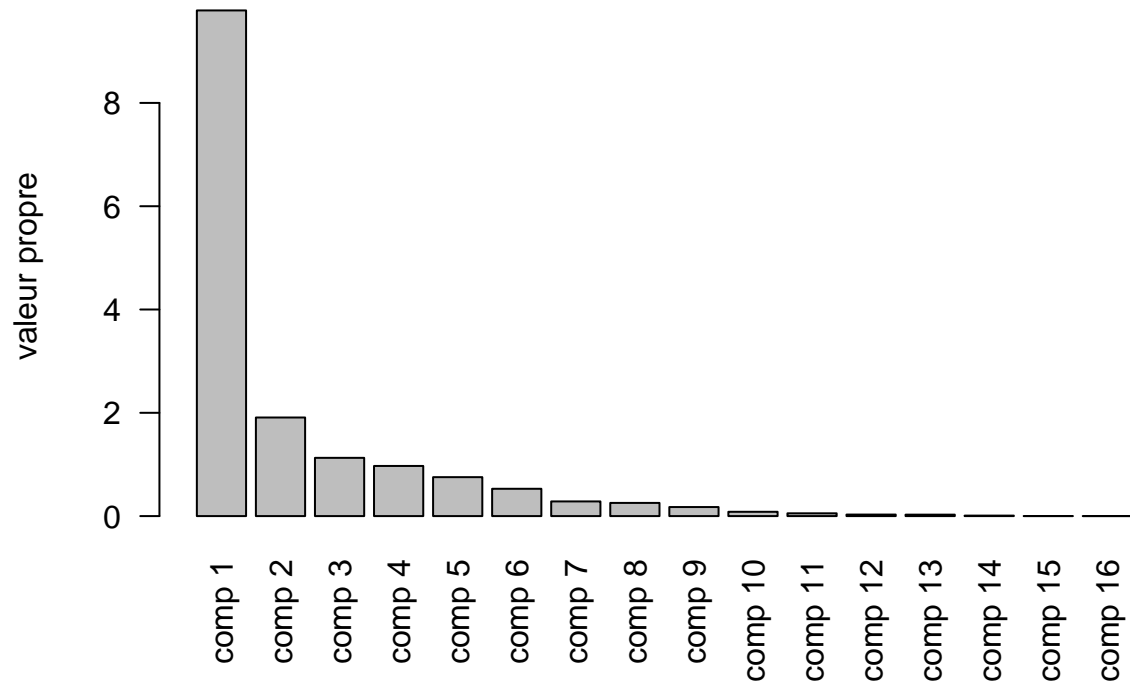
	Changement climatique	Appauv. ozone	Rayo. ionisants	Forma. photo. ozone	Particules	Eff. non cancer.	Eff. cancer.	Acid. terre./eaux d.	Eutro. ed	Eutro. marine	Eutro. terre	Écotoxi. aqua ed	Uti. sol	Épuis. eau	Épuis. énerg.	Épuis. miné.	Score EF
Groupe aliment	0.39	0.00	0.18	0.20	0.42	0.33	0.44	0.41	0.30	0.37	0.41	0.49	0.29	0.03	0.31	0.10	0.43
Sous groupe aliment	0.59	0.01	0.51	0.48	0.58	0.55	0.61	0.57	0.52	0.58	0.57	0.63	0.49	0.13	0.61	0.55	0.60
Saisonnalité	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transport par avion	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Livraison	0.16	0.00	0.04	0.07	0.18	0.15	0.15	0.18	0.11	0.14	0.18	0.17	0.12	0.02	0.09	0.04	0.17
Matériau emballage	0.28	0.00	0.11	0.10	0.26	0.26	0.26	0.26	0.15	0.26	0.27	0.28	0.24	0.02	0.13	0.03	0.27
Préparation	0.14	0.00	0.06	0.05	0.14	0.14	0.13	0.14	0.07	0.13	0.14	0.13	0.13	0.00	0.08	0.03	0.14

## 10.26 Annexe 26 : valeurs propres de l'ACP.

Table 31: Table des valeurs propre de l'ACP

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	9.7908142	61.1925885	61.19259
comp 2	1.9090088	11.9313047	73.12389
comp 3	1.1278248	7.0489048	80.17280
comp 4	0.9705453	6.0659080	86.23871
comp 5	0.7539891	4.7124318	90.95114
comp 6	0.5288511	3.3053195	94.25646
comp 7	0.2840833	1.7755209	96.03198
comp 8	0.2551264	1.5945399	97.62652
comp 9	0.1751131	1.0944567	98.72097
comp 10	0.0833177	0.5207355	99.24171
comp 11	0.0536248	0.3351551	99.57687
comp 12	0.0293608	0.1835052	99.76037
comp 13	0.0280415	0.1752594	99.93563
comp 14	0.0091152	0.0569699	99.99260
comp 15	0.0008313	0.0051953	99.99780
comp 16	0.0003528	0.0022048	100.00000

**Diagramme en barre des valeurs propres  
porté par chaque axe**



## 10.27 Annexe 27 : contribution et $\cos^2$ .

Table 32: Tableau des  $\cos^2$  et des contributions selon les contributions maximales des produits sur l'axe 1

	cos2.Dim.1	cos2.Dim.2	contrib.Dim.1	contrib.Dim.2
Crabe ou Tourteau, bouilli/cuit à l'eau	0.34	0.22	2.70	8.89
Homard, bouilli/cuit à l'eau	0.38	0.23	2.61	8.22
Agneau, collier, braisé ou bouilli	0.93	0.05	1.63	0.45
Agneau, gigot, braisé	0.93	0.05	1.63	0.45
Agneau, gigot, rôti/cuit au four	0.93	0.05	1.63	0.45
Agneau, selle, partie maigre, rôtie/cuite au four	0.93	0.05	1.63	0.45
Brochette d'agneau	0.93	0.05	1.62	0.46
Chevreau, cuit	0.93	0.06	1.61	0.50
Agneau, côte filet, grillée/poêlée	0.93	0.06	1.61	0.50
Agneau, côte première, grillée/poêlée	0.93	0.06	1.61	0.50
Agneau, gigot, grillé/poêlé	0.93	0.06	1.61	0.50
Agneau, selle, partie maigre, grillée/poêlée	0.93	0.06	1.61	0.50
Agneau, côtelette, grillée	0.93	0.06	1.61	0.50
Agneau, épaule, maigre, rôtie/cuite au four	0.93	0.06	0.97	0.30
Agneau, épaule, rôtie/cuite au four	0.93	0.06	0.97	0.30
Agneau, collier, cru	0.92	0.07	0.92	0.36
Agneau, côte filet, crue	0.92	0.07	0.92	0.36
Agneau, côte première, crue	0.92	0.07	0.92	0.36
Agneau, gigot, cru	0.92	0.07	0.92	0.36
Agneau, selle, crue	0.92	0.07	0.92	0.36

Table 33: Tableau des  $\cos^2$  et des contributions selon les contributions maximales des produits sur l'axe 2

	cos2.Dim.1	cos2.Dim.2	contrib.Dim.1	contrib.Dim.2
Pêche, sèche	0.08	0.31	0.48	10.25
Crabe ou Tourteau, bouilli/cuit à l'eau	0.34	0.22	2.70	8.89
Homard, bouilli/cuit à l'eau	0.38	0.23	2.61	8.22
Tomate, séchée	0.08	0.64	0.11	4.97
Tomate, séchée, à l'huile	0.08	0.64	0.11	4.97
Oignon, séché	0.03	0.68	0.02	1.79
Ao-nori (Enteromorpha sp.), séchée ou déshydratée	0.09	0.71	0.03	1.36
Kombu ou kombu japonais (Laminaria japonica), séchée ou déshydratée	0.09	0.71	0.03	1.36
Kombu royal (Saccharina latissima), séchée ou déshydratée	0.09	0.71	0.03	1.36
Nori (Porphyra sp.), séchée ou déshydratée	0.09	0.71	0.03	1.36
Spiruline (Spirulina sp.), séchée ou déshydratée	0.09	0.71	0.03	1.36
Wakamé (Undaria pinnatifida), séchée ou déshydratée	0.09	0.71	0.03	1.36
Gracilaire ou ogonori (Gracilaria verrucosa), séchée ou déshydratée	0.09	0.71	0.03	1.36
Agar (algue), cru	0.09	0.71	0.03	1.36
Dulse (Palmaria palmata), séchée ou déshydratée	0.08	0.71	0.03	1.34
Fucus vésiculeux (Fucus serratus ou Fucus vesiculosus), séché ou déshydraté	0.08	0.71	0.03	1.34
Kombu breton (Laminaria digitata), séchée ou déshydratée	0.08	0.71	0.03	1.34
Laitue de mer (Ulva sp.), séchée ou déshydratée	0.08	0.71	0.03	1.34
Lichen de mer ou piocha ou goémon rouge (Chondrus crispus), séché ou déshydraté	0.08	0.71	0.03	1.34
Wakamé atlantique (Alaria esculenta), séchée ou déshydratée	0.08	0.71	0.03	1.34



## 10.28 Annexe 28 : coordonnées axe 1.

Table 34: Tableau des coordonnées les plus fortes sur l'axe 1

	Dim.1	Dim.2
Crabe ou Tourteau, bouilli/cuit à l'eau	21.45	17.20
Homard, bouilli/cuit à l'eau	21.12	16.54
Agneau, collier, braisé ou bouilli	16.67	-3.86
Agneau, gigot, braisé	16.67	-3.86
Agneau, gigot, rôti/cuit au four	16.67	-3.86
Agneau, selle, partie maigre, rôtie/cuite au four	16.67	-3.86
Brochette d'agneau	16.62	-3.92
Chevreau, cuit	16.59	-4.09
Agneau, côte filet, grillée/poêlée	16.59	-4.09
Agneau, côte première, grillée/poêlée	16.59	-4.09
Agneau, gigot, grillé/poêlé	16.59	-4.09
Agneau, selle, partie maigre, grillée/poêlée	16.59	-4.09
Agneau, côtelette, grillée	16.58	-4.10
Agneau, épaule, maigre, rôtie/cuite au four	12.84	-3.16
Agneau, épaule, rôtie/cuite au four	12.84	-3.16
Agneau, collier, cru	12.56	-3.44
Agneau, côte filet, crue	12.56	-3.44
Agneau, côte première, crue	12.56	-3.44
Agneau, gigot, cru	12.56	-3.44
Agneau, selle, crue	12.56	-3.44

Table 35: Tableau des coordonnées les moins fortes sur l'axe 1

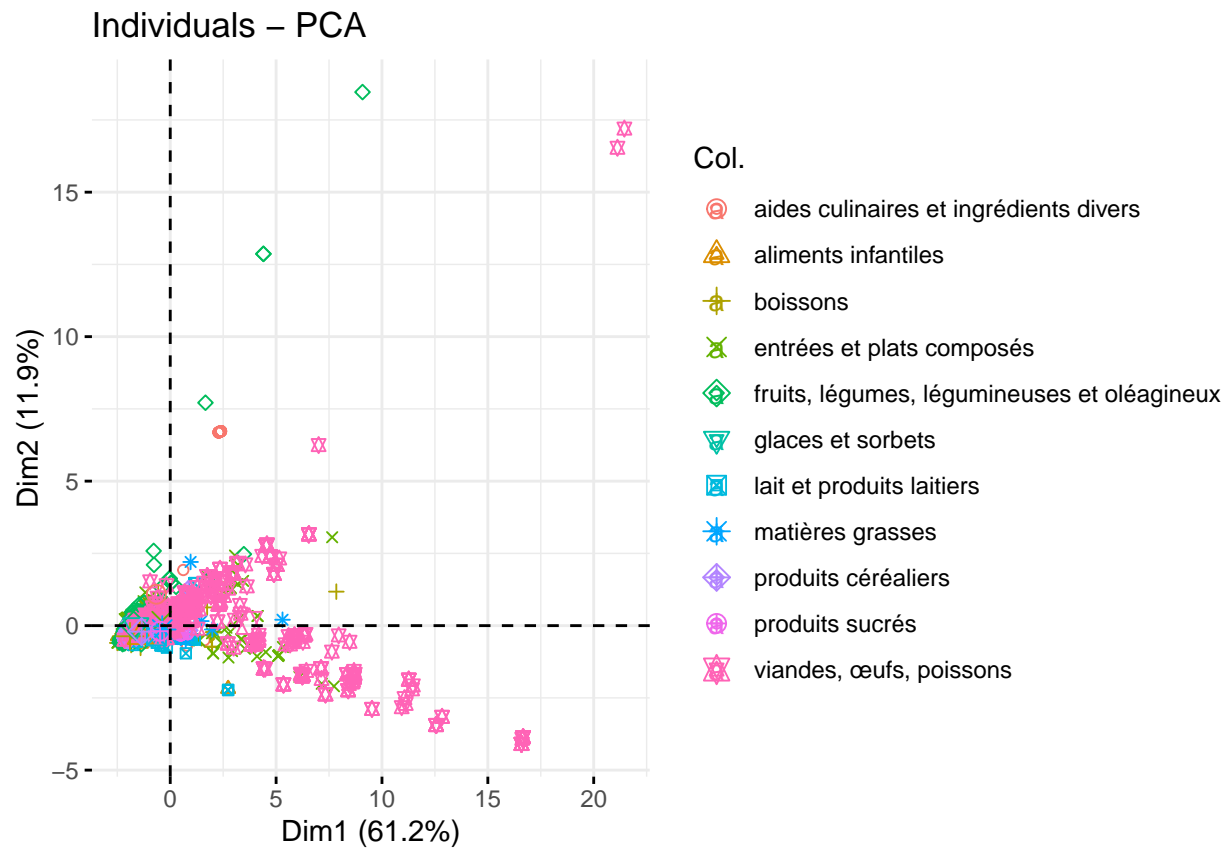
	Dim.1	Dim.2
Thé infusé, non sucré	-2.50	-0.60
Bouillon de volaille	-2.48	-0.59
Bouillon de boeuf, déshydraté reconstitué	-2.48	-0.59
Bouillon de légumes, déshydraté reconstitué	-2.48	-0.58
Boisson préparée à partir de sirop à diluer type menthe, fraise, etc, sucré, dilué dans l'eau	-2.48	-0.60
Soupe à la volaille et aux légumes, déshydratée reconstituée	-2.47	-0.58
Chicorée, poudre soluble	-2.46	-0.50
Eau de source Cristaline, embouteillée, non gazeuse	-2.42	-0.48
Eau de source Ogeu, embouteillée, faiblement minéralisée (Ogeu, 64)	-2.42	-0.48
Eau embouteillée de source	-2.42	-0.48
Eau minérale Abatilles, embouteillée, non gazeuse, faiblement minéralisée (Arcachon, 33)	-2.42	-0.48
Eau minérale Aix-les-Bains, embouteillée, non gazeuse, faiblement minéralisée (Aix-les-Bains, 73)	-2.42	-0.48
Eau minérale Amanda, embouteillée, non gazeuse, fortement minéralisée (St-Amand, 59)	-2.42	-0.48
Eau minérale Appollinaris, embouteillée, non gazeuse, fortement minéralisée (Allemagne)	-2.42	-0.48
Eau minérale Avra, embouteillée, non gazeuse, faiblement minéralisée (Grèce)	-2.42	-0.48
Eau minérale Beckerich, embouteillée, non gazeuse, faiblement minéralisée (Luxembourg)	-2.42	-0.48
Eau minérale Biovine, embouteillée, non gazeuse, faiblement minéralisée (Dax, 40)	-2.42	-0.48
Eau minérale Carola, embouteillée, gazeuse ou non gazeuse, moyennement minéralisée (Ribeauville, 68)	-2.42	-0.48
Eau minérale Celtic, embouteillée, gazeuse ou non gazeuse, très faiblement minéralisée (Niederbronn, 67)	-2.42	-0.48
Eau minérale Chambon, embouteillée, non gazeuse, faiblement minéralisée (Chambon, 45)	-2.42	-0.48

## 10.29 Annexe 29 : coordonnées axe 2.

Table 36: Tableau des coordonnées les plus fortes sur l'axe 2

	Dim.1	Dim.2
Pêche, sèche	9.08	18.46
Crabe ou Tourteau, bouilli/cuit à l'eau	21.45	17.20
Homard, bouilli/cuit à l'eau	21.12	16.54
Tomate, séchée	4.40	12.87
Tomate, séchée, à l'huile	4.40	12.87
Oignon, séché	1.66	7.71
Ao-nori ( <i>Enteromorpha</i> sp.), séchée ou déshydratée	2.39	6.73
Kombu ou kombu japonais ( <i>Laminaria japonica</i> ), séchée ou déshydratée	2.39	6.73
Kombu royal ( <i>Saccharina latissima</i> ), séchée ou déshydratée	2.39	6.73
Nori ( <i>Porphyra</i> sp.), séchée ou déshydratée	2.39	6.73
Spiruline ( <i>Spirulina</i> sp.), séchée ou déshydratée	2.39	6.73
Wakamé ( <i>Undaria pinnatifida</i> ), séchée ou déshydratée	2.39	6.73
Gracilaire ou ogonori ( <i>Gracilaria verrucosa</i> ), séchée ou déshydratée	2.39	6.73
Agar (algue), cru	2.39	6.73
Dulse ( <i>Palmaria palmata</i> ), séchée ou déshydratée	2.29	6.69
Fucus vésiculeux ( <i>Fucus serratus</i> ou <i>Fucus vesiculosus</i> ), séché ou déshydraté	2.29	6.69
Kombu breton ( <i>Laminaria digitata</i> ), séchée ou déshydratée	2.29	6.69
Laitue de mer ( <i>Ulva</i> sp.), séchée ou déshydratée	2.29	6.69
Lichen de mer ou pioca ou goémon rouge ( <i>Chondrus crispus</i> ), séché ou déshydraté	2.29	6.69
Wakamé atlantique ( <i>Alaria esculenta</i> ), séchée ou déshydratée	2.29	6.69

### 10.30 Annexe 30 : graphe des individus selon le groupe d'aliments.



### 10.31 Annexe 31 : $\cos^2$ et corrélation.

Table 37: Tableau des  $\cos^2$  des variables avec les 2 premiers axes

	Dim.1	Dim.2
Changement climatique	0.915	0.046
Appauv. ozone	0.002	0.002
Rayo. ionisants	0.162	0.573
Forma. photo. ozone	0.548	0.053
Particules	0.928	0.031
Eff. non cancer.	0.862	0.062
Eff. cancer.	0.939	0.000
Acid. terre./eaux d.	0.921	0.040
Eutro. ed	0.524	0.052
Eutro. marine	0.817	0.034
Eutro. terre	0.890	0.067
Écotoxi. aqua ed	0.670	0.010
Uti. sol	0.799	0.135
Épuis. eau	0.058	0.107
Épuis. éner.	0.457	0.475
Épuis. miné.	0.299	0.223

Table 38: Tableau des corrélations des variables avec les 2 premiers axes

	Dim.1	Dim.2
Changement climatique	0.957	-0.214
Appauv. ozone	0.041	0.041
Rayo. ionisants	0.402	0.757
Forma. photo. ozone	0.740	0.229
Particules	0.963	-0.177
Eff. non cancer.	0.929	-0.248
Eff. cancer.	0.969	0.011
Acid. terre./eaux d.	0.960	-0.199
Eutro. ed	0.724	0.228
Eutro. marine	0.904	-0.185
Eutro. terre	0.944	-0.259
Écotoxi. aqua ed	0.818	0.101
Uti. sol	0.894	-0.367
Épuis. eau	0.241	0.327
Épuis. éner.	0.676	0.689
Épuis. miné.	0.547	0.472

## 10.32 Annexe 32 : classe 1.

Table 39: Description de la classe 1 selon les variables quantitatives

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Épuis..énerg.	23.884229	116.8293724	48.6965308	77.5003309	53.1906803	0
Rayo..ionisants	21.662384	3.1554007	1.1579875	3.3319523	1.7193011	0
Écotoxi..aqua.ed	21.631679	126.4642998	57.1391238	58.4827893	59.7573069	0
Eutro..ed	18.713374	0.0017251	0.0007547	0.0015398	0.0009670	0
Eutro..marine	15.688059	0.0562782	0.0254397	0.0407647	0.0366534	0
Eff..cancer.	15.187533	0.0000000	0.0000000	0.0000000	0.0000000	0
Forma..photo..ozone	13.363033	0.0310454	0.0141326	0.0270247	0.0235994	0
Particules	11.720990	0.0000012	0.0000006	0.0000005	0.0000009	0
Changement.climatique	11.457880	11.3565150	5.9719589	5.6104285	8.7626583	0
Acid..terre./eaux.d.	11.139875	0.1634316	0.0818147	0.0706089	0.1366123	0
Eutro..terre	10.099577	0.6674268	0.3437980	0.3315070	0.5974951	0
Eff..non.cancer.	8.619744	0.0000002	0.0000001	0.0000001	0.0000002	0
Épuis..miné.	8.154911	0.0000395	0.0000203	0.0000193	0.0000440	0
Épuis..eau	6.785093	8.8013239	4.7264481	22.6267336	11.1982155	0
Uti..sol	6.655023	672.6213733	401.0203114	485.5333483	760.9780783	0

### 10.33 Annexe 33 : classe 2.

Table 40: Description de la classe 2 selon les variables quantitatives

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Épuis..eau	-8.294622	3.5138743	4.7264481	6.6993174	11.1982155	0
Épuis..miné.	-15.690787	0.0000112	0.0000203	0.0000072	0.0000440	0
Rayo..ionisants	-20.886073	0.6892049	1.1579875	0.5473902	1.7193011	0
Forma..photo..ozone	-24.397959	0.0066160	0.0141326	0.0065265	0.0235994	0
Uti..sol	-25.603028	146.6735502	401.0203114	144.5444702	760.9780783	0
Eutro..ed	-26.721414	0.0004174	0.0007547	0.0003138	0.0009670	0
Eff..non.cancer.	-27.752280	0.0000001	0.0000001	0.0000000	0.0000002	0
Épuis..énerg.	-28.419162	28.9627839	48.6965308	17.2990308	53.1906803	0
Eutro..terre	-28.876239	0.1185620	0.3437980	0.1291745	0.5974951	0
Acid..terre./eaux.d.	-29.741058	0.0287741	0.0818147	0.0299905	0.1366123	0
Changement.climatique	-30.056553	2.5337062	5.9719589	2.2050766	8.7626583	0
Particules	-30.074365	0.0000002	0.0000006	0.0000002	0.0000009	0
Eutro..marine	-30.412247	0.0108876	0.0254397	0.0091939	0.0366534	0
Écotoxi..aqua.ed	-31.151042	32.8380013	57.1391238	31.3900260	59.7573069	0
Eff..cancer.	-31.426105	0.0000000	0.0000000	0.0000000	0.0000000	0

### 10.34 Annexe 34 : classe 3.

Table 41: Description de la classe 3 selon les variables quantitatives

	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Eff..non.cancer.	34.897182	0.0000007	0.0000001	0.0000002	0.0000002	0.0000000
Eutro..terre	34.571291	2.2599883	0.3437980	0.7966376	0.5974951	0.0000000
Changement.climatique	34.530115	34.0406783	5.9719589	8.5666444	8.7626583	0.0000000
Acid..terre./eaux.d.	34.471804	0.5186751	0.0818147	0.1755629	0.1366123	0.0000000
Uti..sol	34.189815	2814.5775505	401.0203114	1247.4266548	760.9780783	0.0000000
Particules	34.156325	0.0000035	0.0000006	0.0000012	0.0000009	0.0000000
Eff..cancer.	31.169862	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Eutro..marine	28.640753	0.1228237	0.0254397	0.0373474	0.0366534	0.0000000
Forma..photo..ozone	21.781015	0.0618160	0.0141326	0.0520203	0.0235994	0.0000000
Écotoxi..aqua.ed	20.780653	172.3355238	57.1391238	50.1470106	59.7573069	0.0000000
Eutro..ed	17.583139	0.0023319	0.0007547	0.0010163	0.0009670	0.0000000
Épuis..miné.	14.683184	0.0000803	0.0000203	0.0001543	0.0000440	0.0000000
Épuis..énerg.	12.575614	110.7481986	48.6965308	67.5978036	53.1906803	0.0000000
Épuis..eau	3.956394	8.8364031	4.7264481	3.2462934	11.1982155	0.0000761
Rayo..ionisants	2.921763	1.6239872	1.1579875	0.7039404	1.7193011	0.0034806