

The background of the slide is a collage of mathematical and scientific formulas. Visible formulas include  $(x+15)(x-6)=0$ ,  $\log_a(m) = \log_a m + \log_a n$ ,  $y = qx$ ,  $0 < q < 1$ ,  $\frac{\Delta v}{\Delta t} = \frac{v-u}{t}$ ,  $\frac{\Delta v}{\Delta t} = \frac{v^2 - u^2}{25}$ ,  $\frac{H-C=C-H}{H-C=C-H}$ ,  $CH_4 + xO_2 \rightarrow CO_2 + 2H_2O$ ,  $xH_2SO_4 \rightarrow xH_2O + xSO_2$ ,  $3Cu + 8HNO_3 \rightarrow 3Cu(NO_3)_2 + 2NO + 4H_2O$ ,  $2Cr(OH)_3 + 2OH^- + 3H_2O_2 \rightarrow 2CrO_4^{2-} + 8H_2O$ ,  $\{y \in \mathbb{R}^+ \times \mathbb{R} \mid x = a^y; a > 0, a \neq 1\}$ ,  $n = 360^\circ / \theta - 1$ ,  $5 + \frac{1}{3} = \frac{(15 \times 4) + 10}{15}$ ,  $f = \frac{R}{2}$ ,  $U = \frac{vF}{v-f}$ ,  $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$ ,  $v = \frac{uf}{u-f}$ ,  $g = \frac{6m_e}{R_B^2}$ ,  $O: 3y = 16 + y + \frac{x}{2}$ ,  $MO_4S_7 + 8NO_3^- \rightarrow MO_4 + 7$ ,  $N^{2+} + Cl_2 + H_3O^+ \rightarrow 2Cl^- + NO_3^- + CO_2 +$ ,  $+ C_2O_4^{2-} + H_2O$ ,  $\frac{2l}{n}$ ,  $v = \sqrt{\frac{T}{\mu}}$ ,  $^1_1H + ^1_0n \rightarrow ^2_1H + 2.23$ ,  $mev$ ,  $\vec{v}_1$ ,  $n_1$ ,  $\vec{v}_2$ .

# Projet

# Préparez des données pour un

# organisme de santé publique

Par Antony Coco

nino.one@hotmail.fr

# Challenge de ce projet ?

## contexte:

- Un appel à projets pour rendre les données de santé plus accessibles.
- Fournir un prototype simplifié de l'exploration sous forme de page web. L'utilisation de fonctionnalités de graphiques interactifs pour rendre le prototype aussi réaliste que possible.



# Le nutri-score c'est quoi ?



C'est le calcul d'un score (pour 100 grammes de produit) qui est évalué en fonction de la teneur :

- en nutriments et aliments à favoriser (fibres, protéines, fruits et légumes...) ,
- et en nutriments à limiter (graisses, acides gras saturés, sucres, sel, calories...).



# La pyramide alimentaire

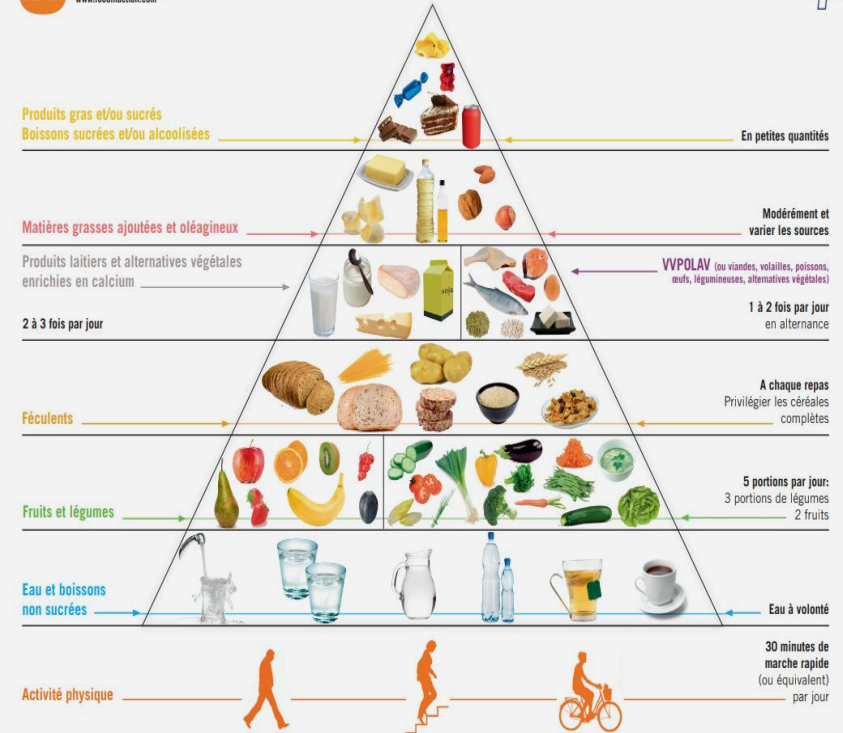
cette pyramide définit  
les besoins alimentaire  
de l'être humain. Mais  
est-ce suffisant ?



Copyright Food In Action 2011  
www.foodinaction.com

## La pyramide alimentaire

Avec la collaboration de l'Institut Paul Lambin



Éditeur responsable: Dr. Guyonneau - Institut Paul Lambin



# Création d'un référentiel nutri- santé



C'est quoi ?

- sur la base d'étude faite comme par "Que Choisir" sur l'évaluation risque des additifs
- sur le classement Nova



Application d'un coefficient sur le Nutri-score favorisant l'utilisation d'additifs non-controversés

# Statistiques descriptives et représentation graphique

## Présentation du jeu de données:

- . 56 variables qualitatives
- . 106 variables quantitatives

### Pour les variables quantitatives:

- Les variables protéines et protéines brutes,
- Les glucides (11), sucre, amidon, polyols et alcool,
- Les lipides (38) dont les acides gras saturés (au nombre de 17)
- Les minéraux (8)
- Les oligoéléments (6)
- L'eau
- Les vitamines (14)

### catégories des individus

Fruits and vegetables  
Sugary snacks  
Cereals and potatoes  
Composite foods  
Fish Meat Eggs  
Beverages  
Fat and sauces  
Milk and dairy products  
Salty snacks



# ALIMENTAIRE

01

## Attribution des points à différents facteurs nutritionnels



### Nutriments à limiter

Points N	Score N	Exemple (g)	Score N	Exemple (g)	Score N	Exemple (g)
0	≤ 335	≤ 4,5	≤ 0	≤ 0	≤ 1	≤ 10
1	> 335	> 4,5	≤ 30	≤ 1,5	> 1	> 10
2	> 970	> 9	≤ 60	≤ 3	> 2	> 20
3	> 1000	> 11,5	≤ 90	≤ 4,5	> 3	> 30
4	> 1340	> 16	≤ 120	≤ 6	> 4	> 40
5	> 1875	> 22,5	≤ 150	≤ 7,5	> 5	> 50
6	> 2010	> 27	≤ 180	≤ 9	> 6	> 60
7	> 2140	> 31	≤ 210	≤ 10,5	> 7	> 70
8	> 2580	> 36	≤ 240	≤ 12	> 8	> 80
9	> 3015	> 40	≤ 270	≤ 13,5	> 9	> 90
10	> 3350	> 45	≤ 300	≤ 15	> 10	> 100
Score N	0 à 10	0 à 10	0 à 10	0 à 10	0 à 10	0 à 10

### Nutriments, aliments à encourager

Points P	Score P	Exemple (g)	Score P	Exemple (g)	Score P	Exemple (g)
0	≤ 40	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
1	> 40	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
2	> 60	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
3	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
4	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
5	> 80	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
6	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
7	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
8	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
9	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
10	-	≤ 0,7	≤ 1,6	≤ 0,7	≤ 1,6	≤ 1,6
Score P	0 à 5	0 à 5	0 à 5	0 à 5	0 à 5	0 à 5

02

## Choix de la méthode de calcul du score final

Points Négatifs N ≥ 11

Points Négatifs N < 11 ou pour les fromages

Score des fruits et légumes = 5

Score des fruits et légumes < 5

Points N - Points P

Points N - (points fibres + points fruits et légumes)

Points N - Points P

03

## Attribution d'une couleur et d'une lettre

Quels produits concernés ?

Score final variant de -15 (qualité nutritionnelle élevée) à 40 (faible qualité nutritionnelle)

Aliments solides	Boissons	Logo
Min à -1	Eaux toujours en A	A
0 à 2	Min à 1	B
3 à 10	2 à 5	C
11 à 18	6 à 9	D
19 à max	10 à max	E

- Tous les aliments transformés, excepté les herbes aromatiques, thés, cafés, levures...
- Toutes les boissons, excepté les boissons alcoolisées
- Excepté les produits dont la face la plus grande a une surface inférieure à 25 cm<sup>2</sup>

## Hypothèse:

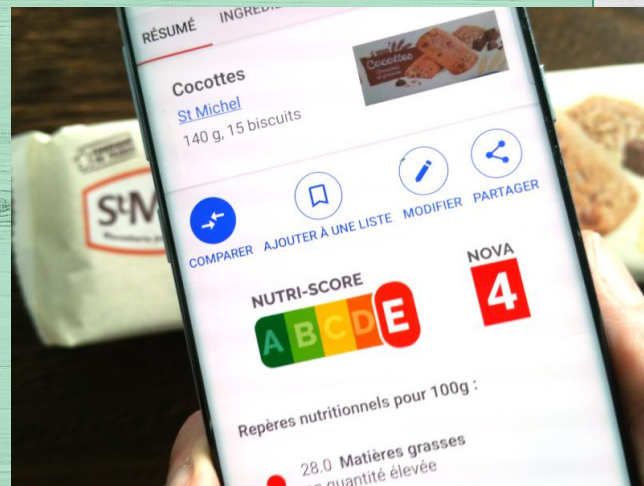
Quel est l'influence du groupe d'aliment sur le nutri-score ?

### Facteurs négatifs :

- les kiloJoules
- les sucres
- graisse saturé
- le sodium

### Facteurs positifs :

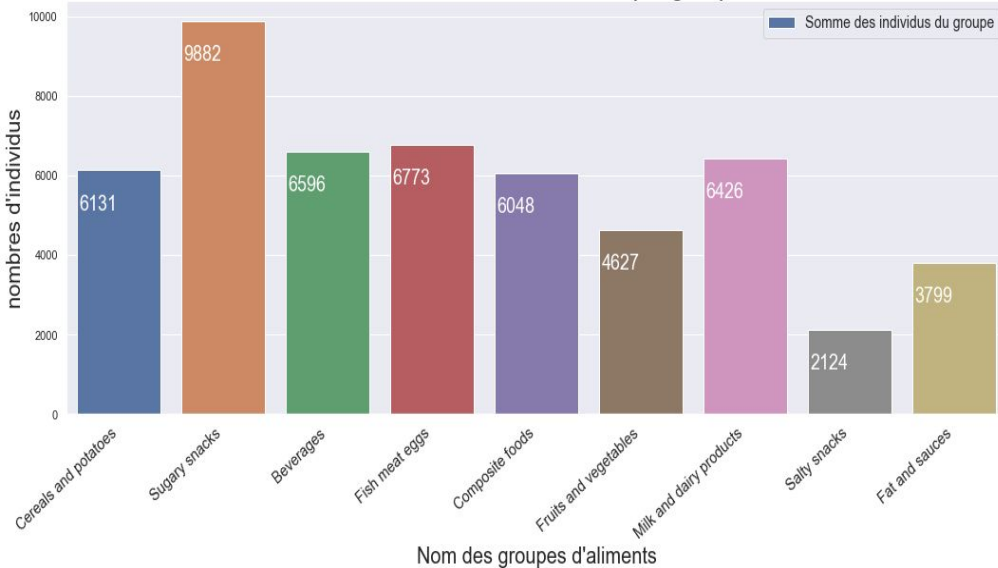
- taux de fruits & légumes
- masse de fibre
- masse de protéines



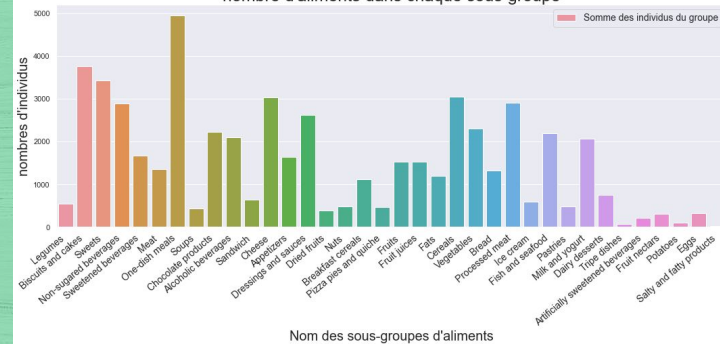
# Analyse des variables qualitatives

## Diagrammes en Bâtons du nombre d'aliments

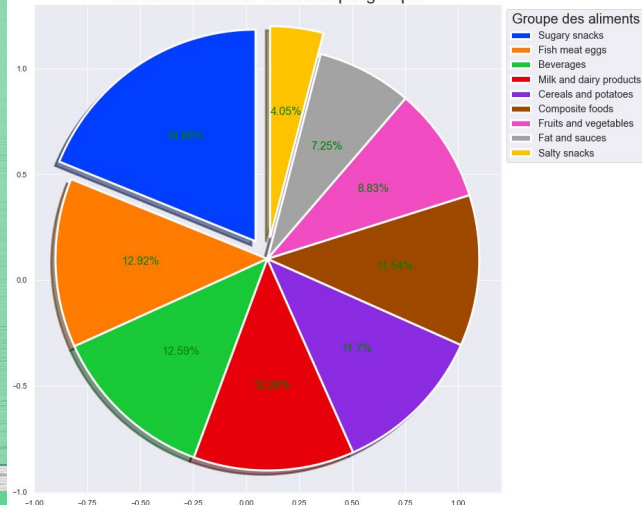
Histogramme correspondant au nombre d'aliments dans chaque groupe



Histogramme correspondant au nombre d'aliments dans chaque sous-groupe



Camembert représentatif du nombre d'aliments dans chaque groupe

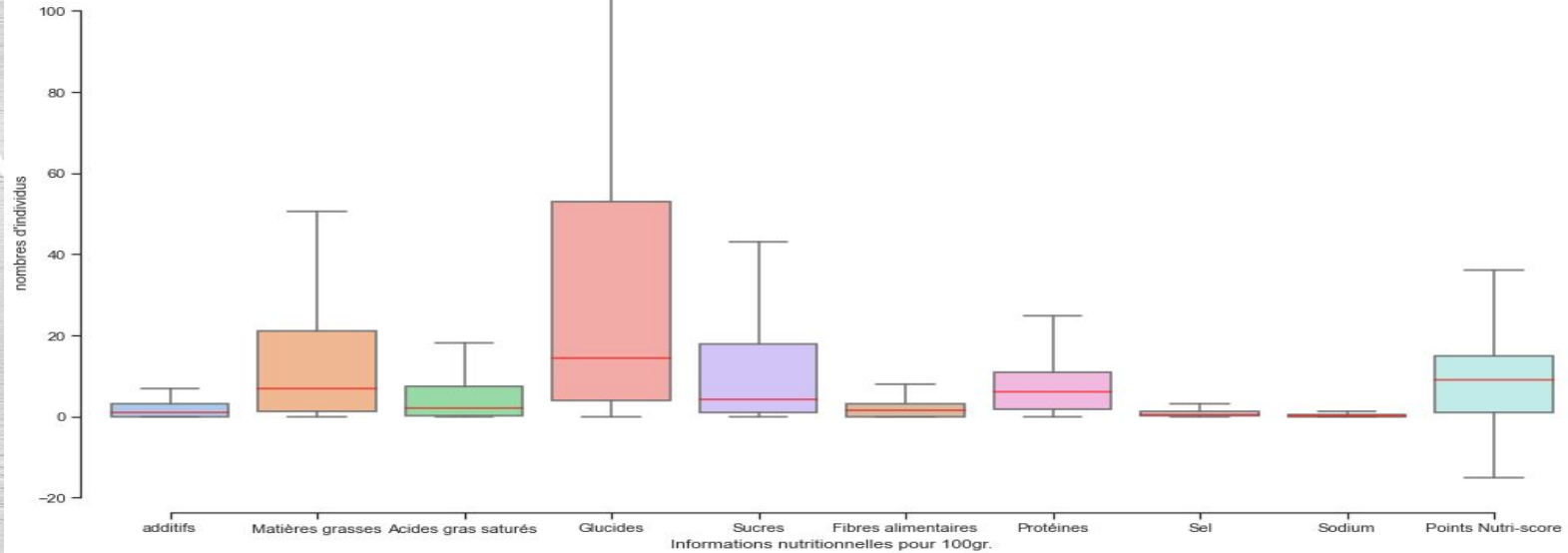




# Analyse et représentation des variables quantitatives



Représentation Uni-varié des quantitatives avant imputation par boxplot

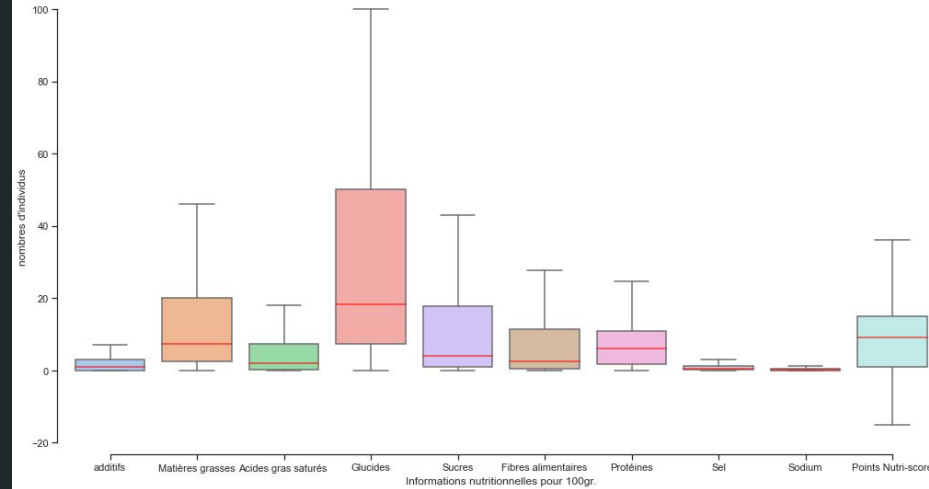


# Analyse et représentation des variables quantitatives après traitement

scikit  
learn

Imputation des valeurs manquantes avec des variantes de Iterative Imputer

Représentation Uni-variée des quantitatives imputation itérative par boxplot



lors de cette exercice, j'ai pu confronter l'algo Scikit avec des fonctions que j'ai créé dont une imputation par la médiane ou moyenne, et une autre fonction qui impute par l'utilisation de la droite de régression linéaire



# La valeur Énergétique des aliments

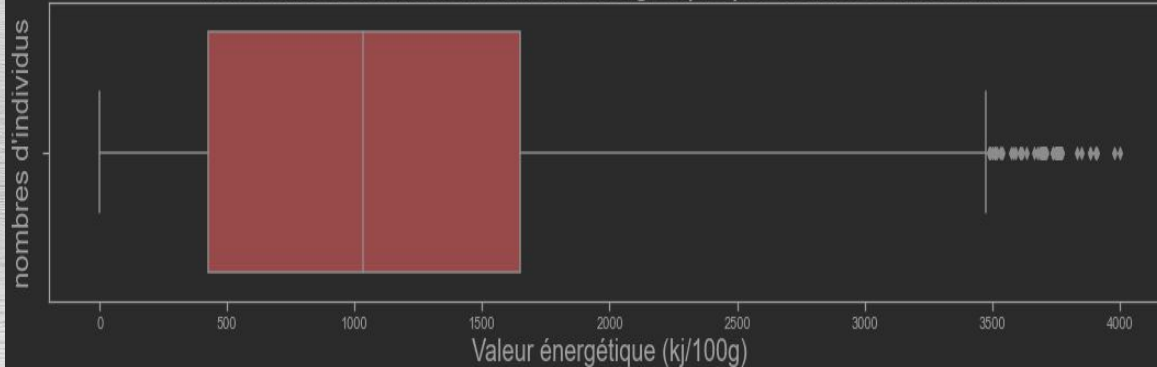
Règlement UE N 1169/2011 (kcal/100g)

D'après les informations fournies par l'Anses sur la table, pour l'ensemble des aliments, la valeur énergétique a été calculée en utilisant les coefficients suivants :

- pour les lipides : 37 kJ/g (9 kcal/g) ;
- pour l'alcool (éthanol) : 29 kJ/g (7 kcal/g) ;
- pour les protéines : 17 kJ/g (4 kcal/g) ;
- pour les glucides (à l'exception des polyols) : 17 kJ/g (4 kcal/g) ;
- pour les acides organiques : 13 kJ/g (3 kcal/g) ;
- pour les polyols : 10 kJ/g (2,4 kcal/g) ;
- pour les fibres alimentaires : 8 kJ/g (2 kcal/g).

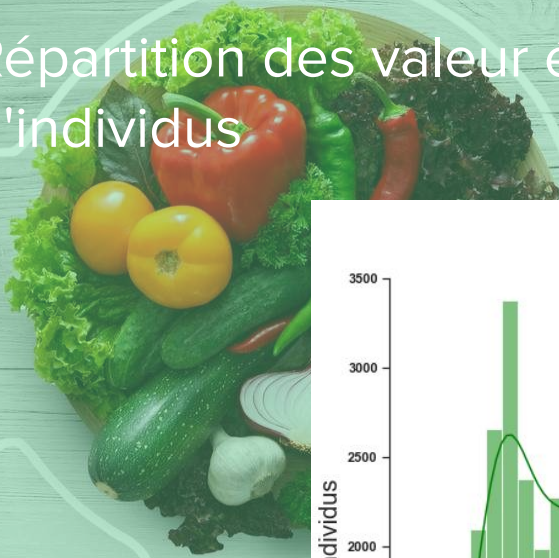
Calcul des principales grandeurs statistiques

boite à moustache de la valeur énergétique par nombres d'individus

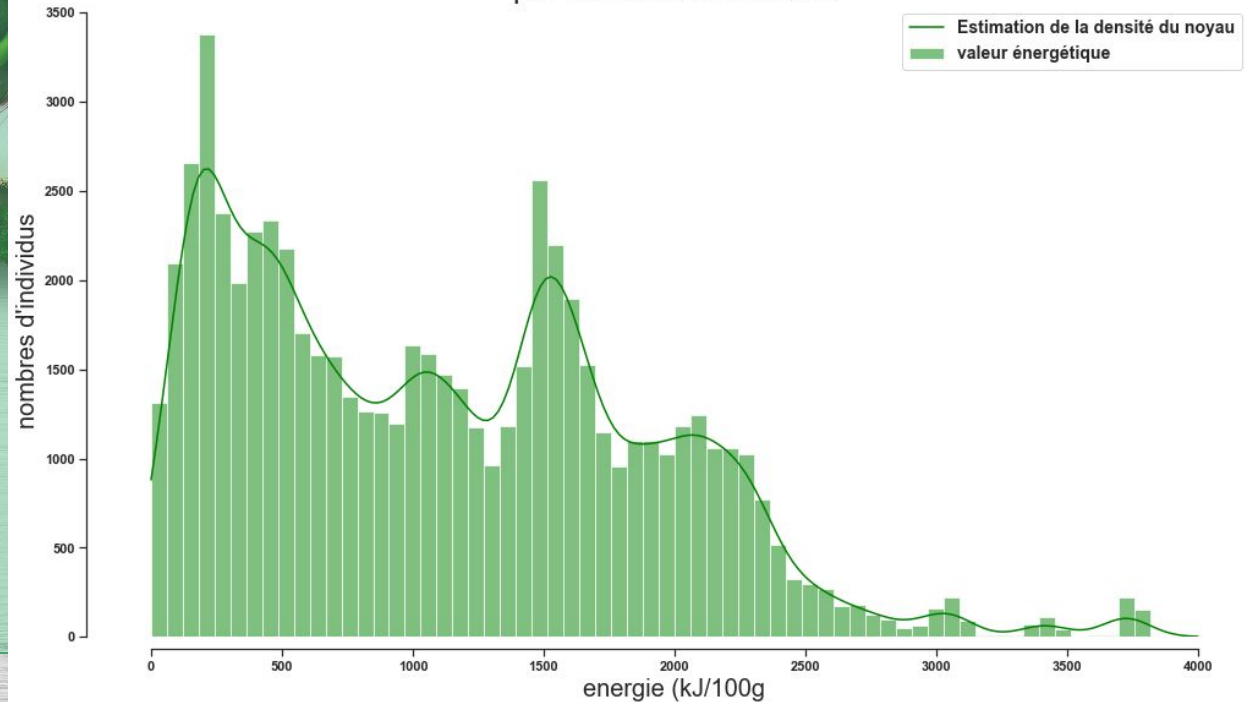


count	64533.000000
mean	1111.933423
std	774.159143
min	0.000000
25%	427.000000
50%	1033.000000
75%	1648.000000
max	4000.000000

# Répartition des valeur énergétique par nombres d'individus

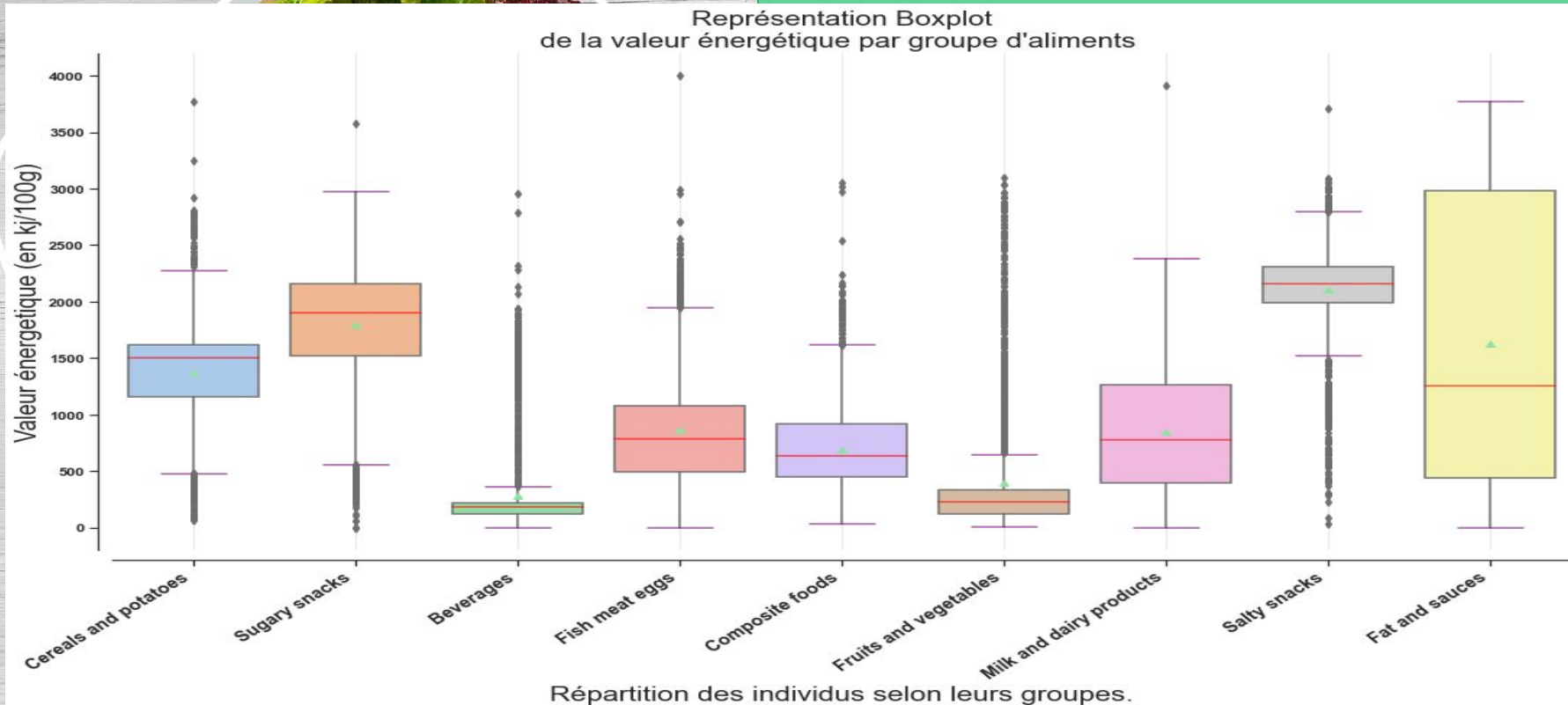


Repartition des valeur énergétique  
par nombres d'individus

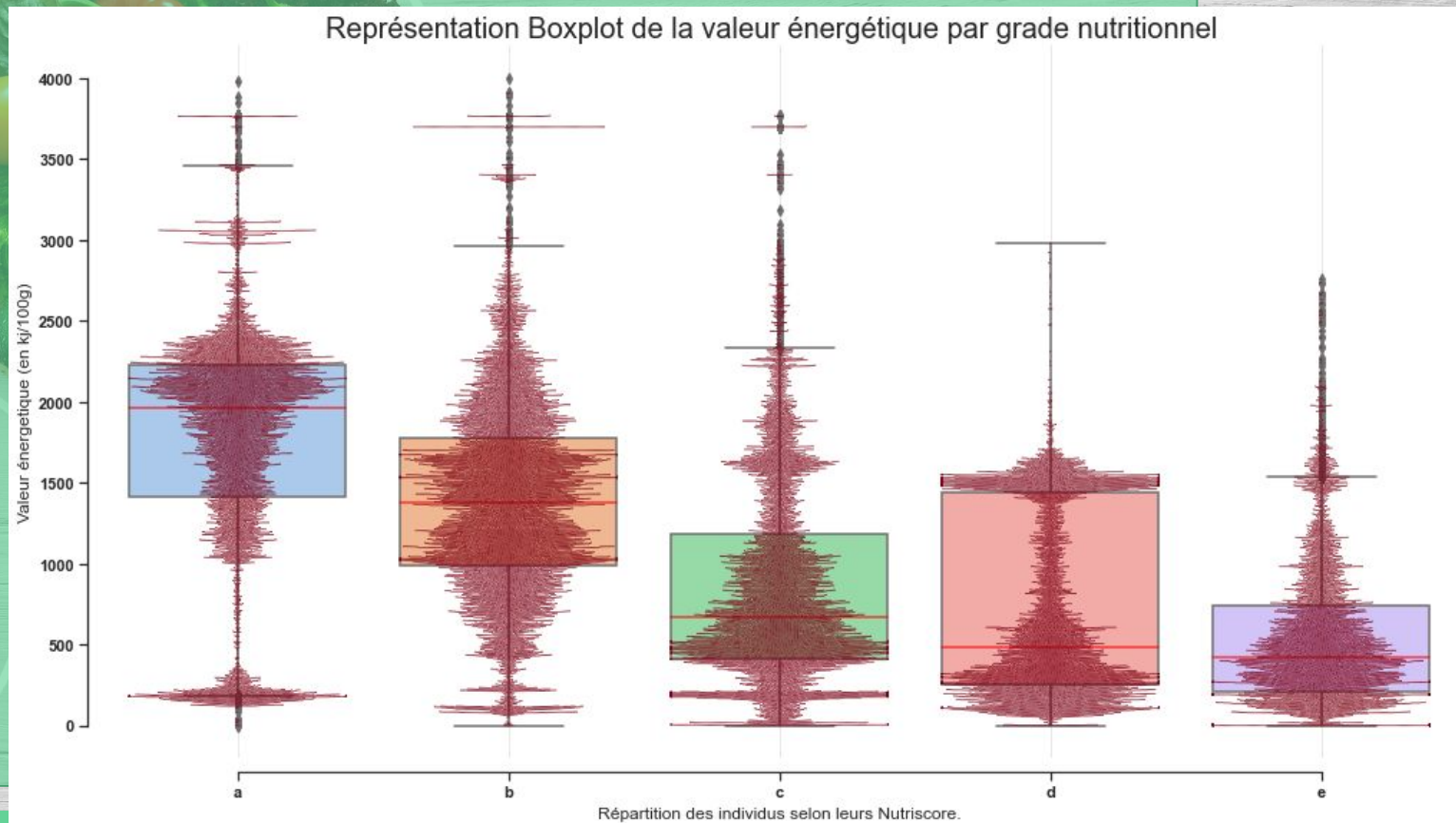




# Représentation de la valeur énergétique par groupe d'aliment



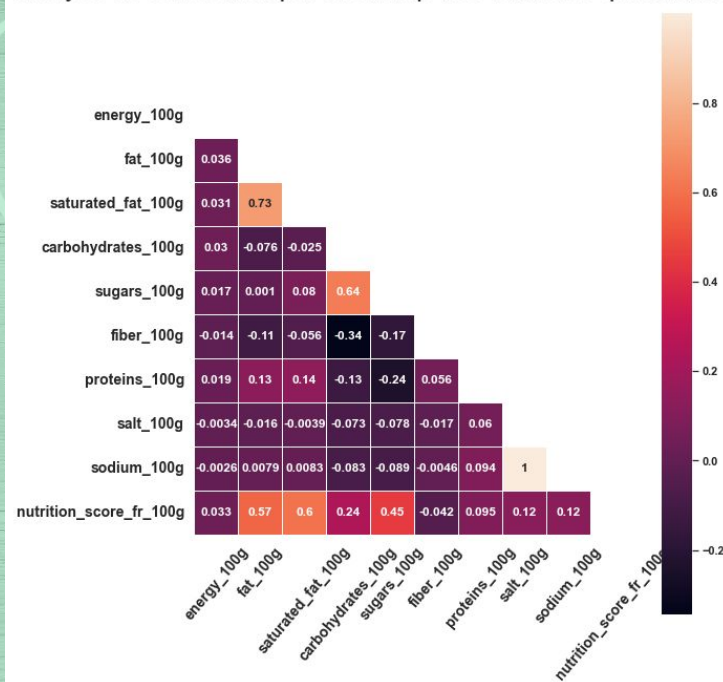
# Représentation de la valeur énergétique par leur nutri-score





# Etude de la corrélation entre les variables nutritionnelles

Analyse de corrélation par Heatmap des variables quantitatives

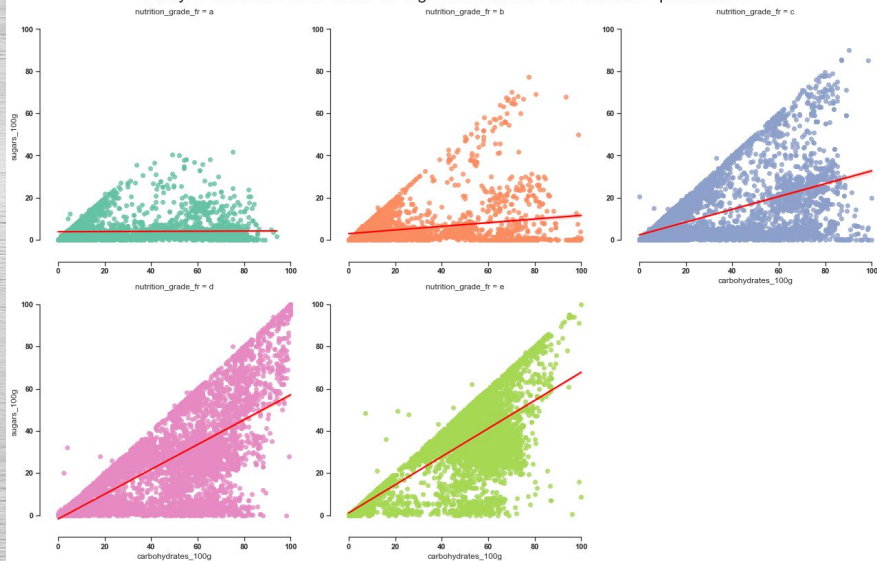


corrélation :

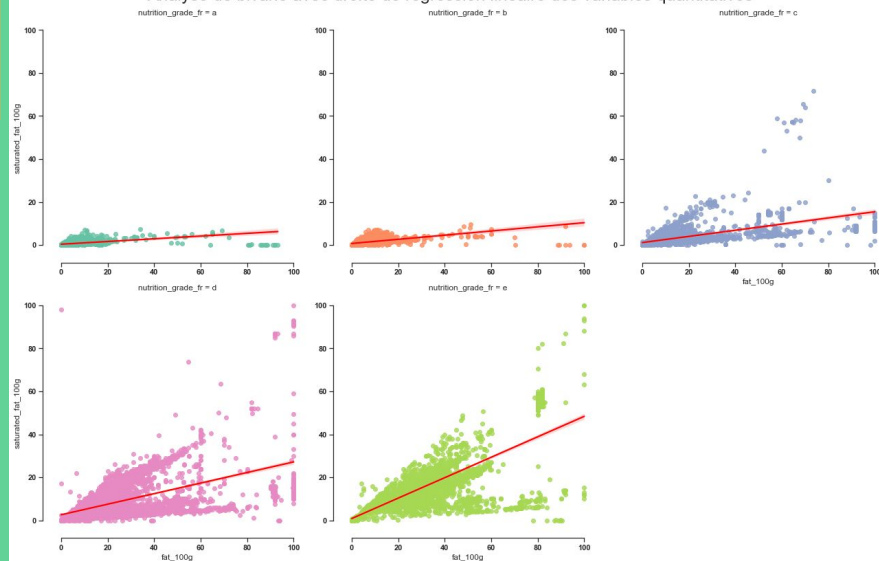
- entre graisse et graisse saturée
- entre glucose et sucre
- forte entre sel et sodium

# Régression linéaire

Analyse de bivarié avec droite de régression linéaire des variables quantitatives



Analyse de bivarié avec droite de régression linéaire des variables quantitatives

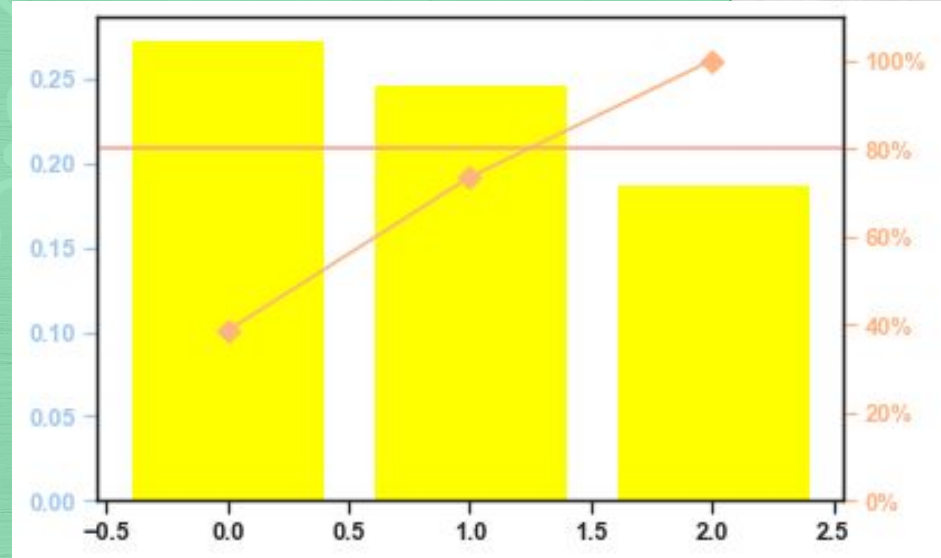
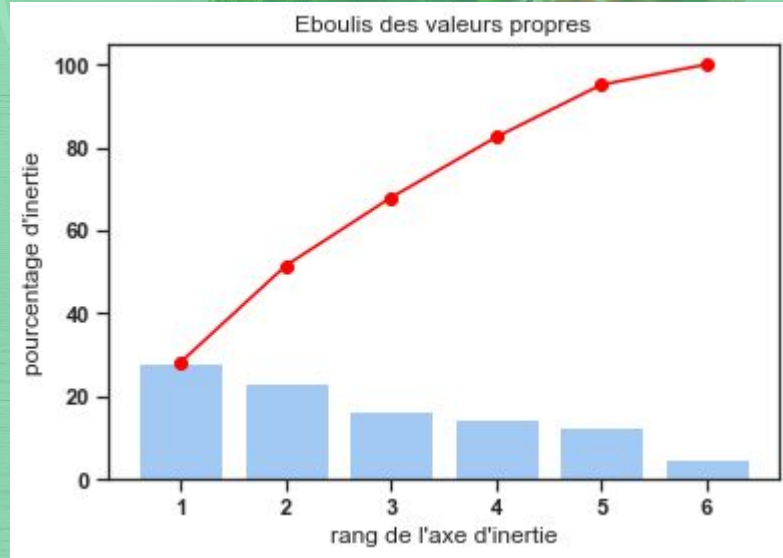


- faible pente pour les nutri-score A
- à contrario elle augmente selon le déclassement
- la répartition des individus n'est pas homogène dans les différent groupe



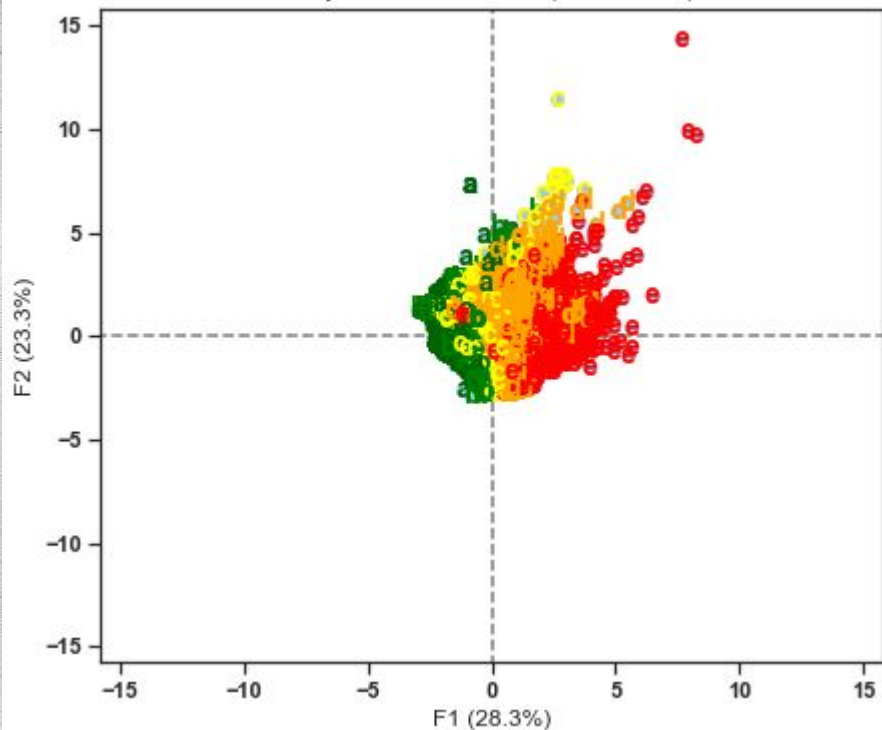
# Analyse en composante principale

- Eboulis des valeurs propres

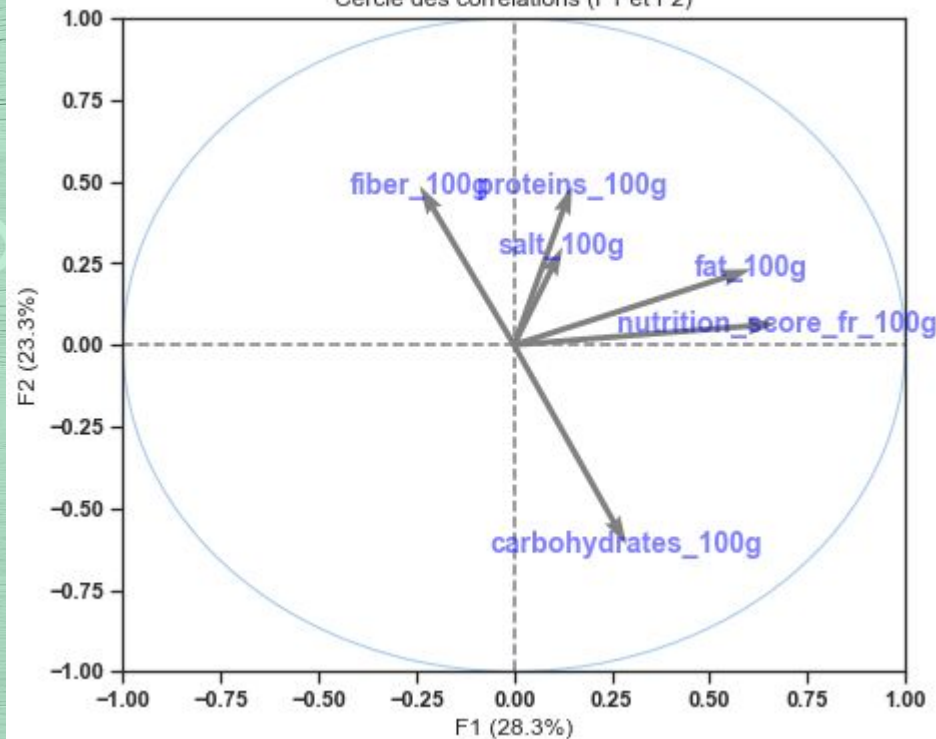


## Projection des individus par groupe de nutri-score et son cercle de corrélation

Projection des individus (sur F1 et F2)

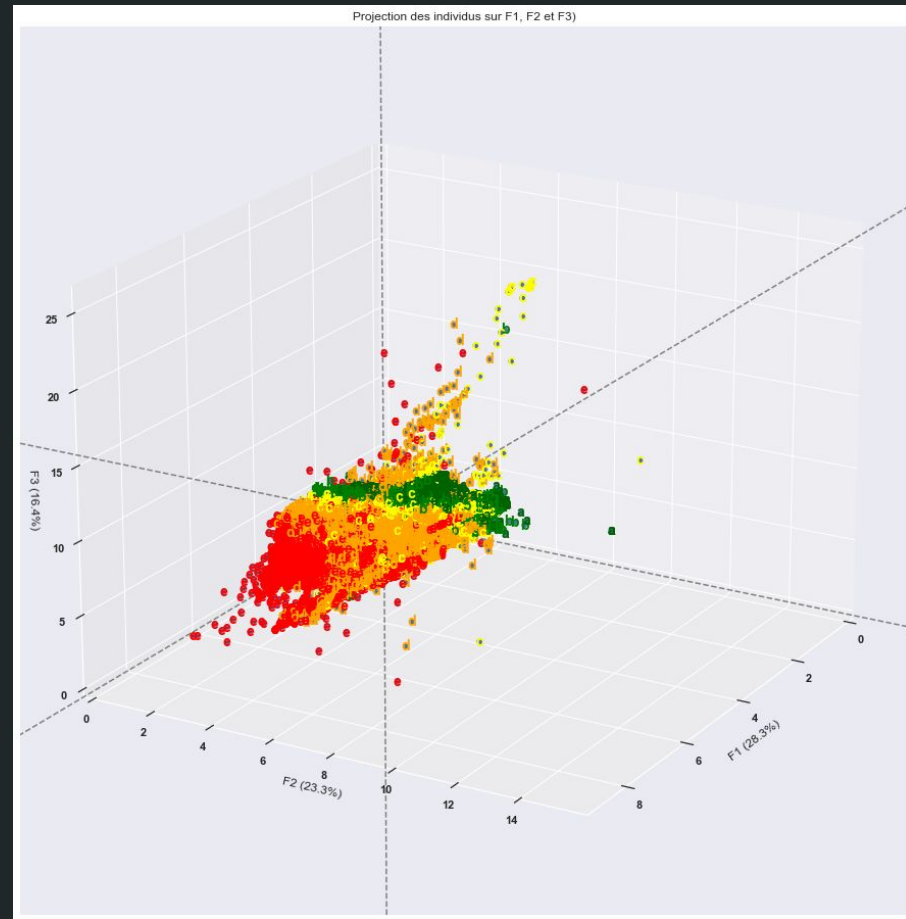


Cercle des corrélations (F1 et F2)

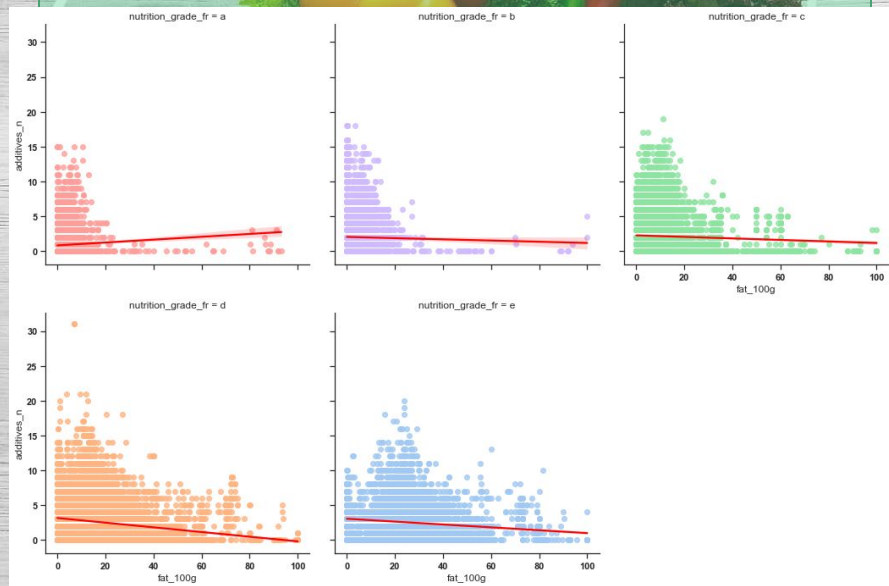




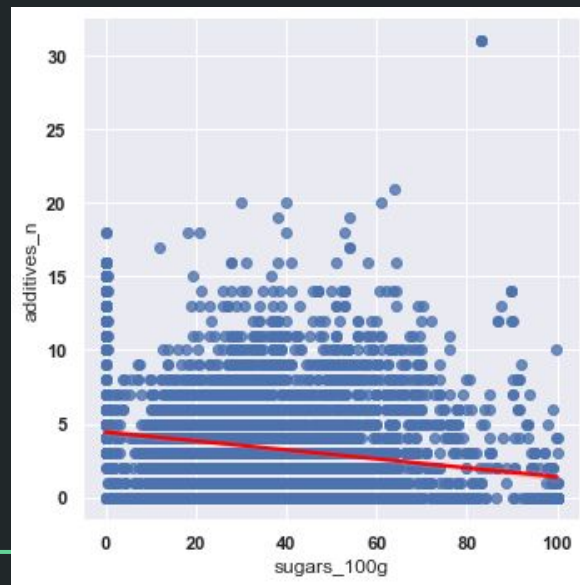
# Représentation 3D de la projection des individus



# analyse du taux de gras selon les additifs.

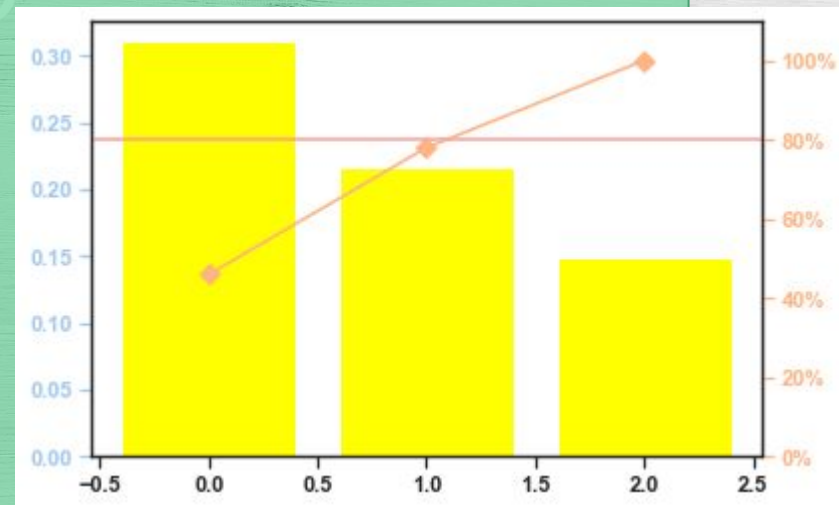
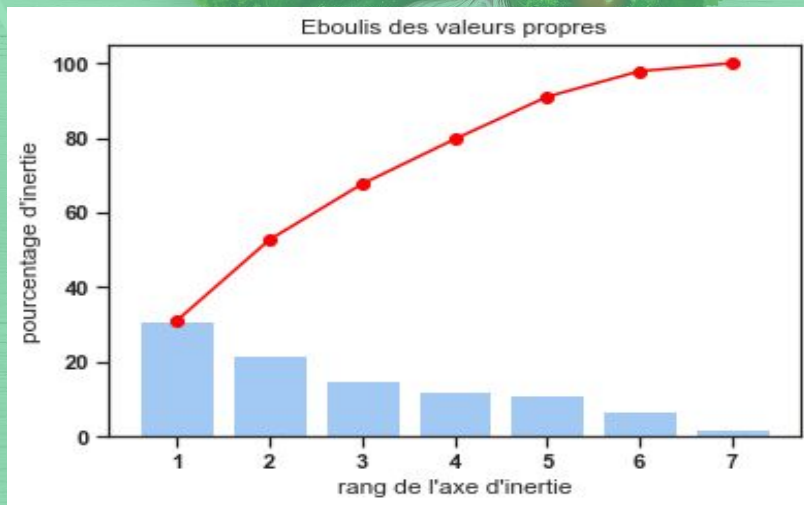


analyse du taux de sucre  
selon les additifs dans les  
snacks sucrées





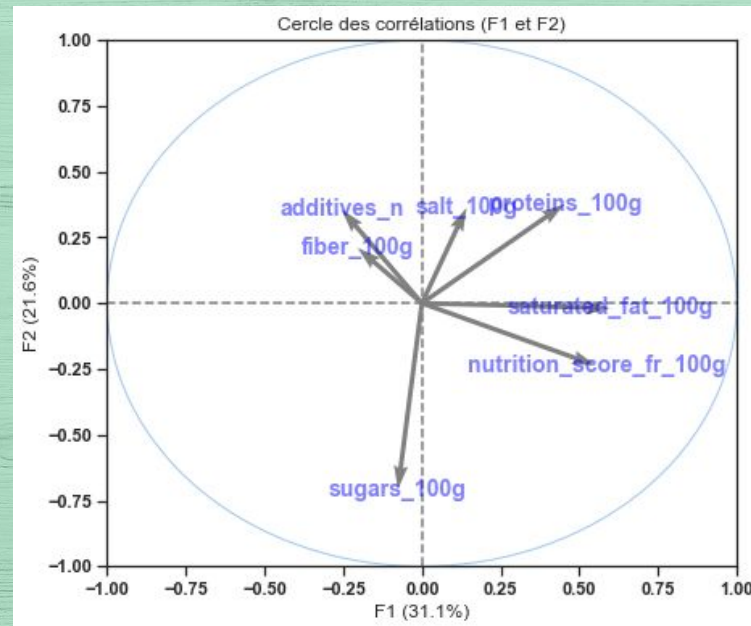
# Eboulis des valeurs propres snack sucrées



# Cercle de corrélation

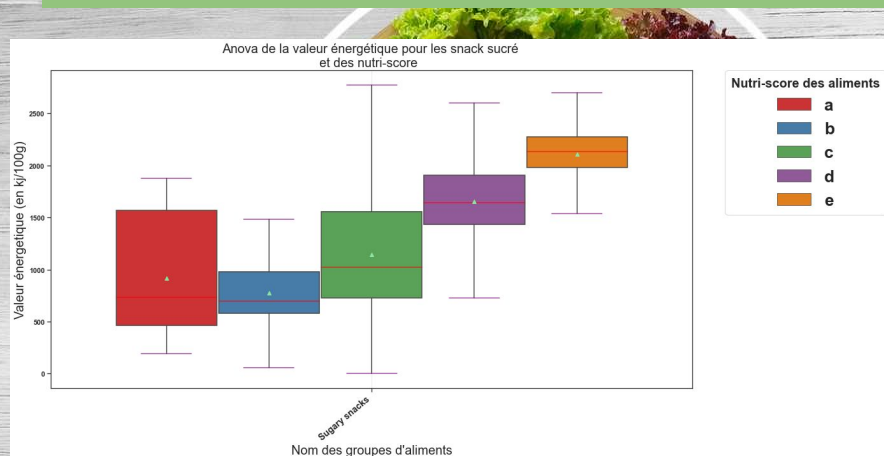
- anti-corrélation entre les additifs et le gras saturé sur F1
- anti-corrélation entre les additifs et le sucre sur F2

**Mais le plus important, c'est la confirmation de l'hypothèse de départ que les additifs sont la pour réduire le nutri-score.**





# Analyse de la variance des snacks sucrés



## OLS model

	sum_sq	df	F	PR(>F)
C(nutrition_grade_fr)	1.021807e+09	4.0	2314.246039	0.0
Residual	8.562351e+08	7757.0	NaN	NaN

	df	sum_sq	mean_sq	F	PR(>F)
C(nutrition_grade_fr)	4.0	1.021807e+09	2.554517e+08	2314.246039	0.0
Residual	7757.0	8.562351e+08	1.103822e+05	NaN	NaN

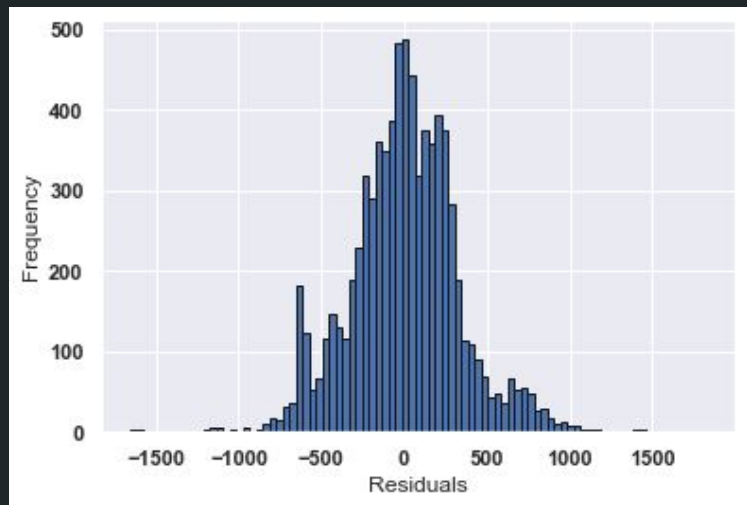
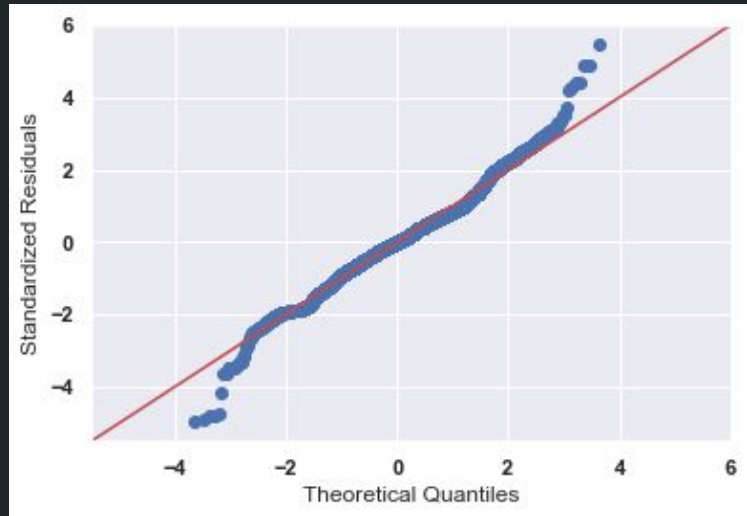
## HSD de Tukey

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	d	e	453.238467	431.035544	475.441390	78.766040	0.0010
1	d	c	507.074514	468.501729	545.647300	50.723965	0.0010
2	d	b	878.538009	819.589069	937.486948	57.505180	0.0010
3	d	a	738.008235	608.734400	867.282069	22.027894	0.0010
4	e	c	960.312981	922.558005	998.067958	98.143378	0.0010
5	e	b	1331.776476	1273.359389	1390.193562	87.965800	0.0010
6	e	a	1191.246701	1062.214524	1320.278878	35.622639	0.0010
7	c	b	371.463494	305.074897	437.852092	21.589623	0.0010
8	c	a	230.933720	98.102276	363.765165	6.708244	0.0010
9	b	a	140.529774	0.417996	280.641553	3.870046	0.0489

# Tester les hypothèses de l' ANOVA

Test de Levene

	Parameter	Value
0	Test statistics (W)	227.4985
1	Degrees of freedom (Df)	4.0000
2	p value	0.0000



## ...Fin de Ma présentation



**Merci pour votre attention**

- Avez-vous des question ?  
suivant présentation de graphique avec  
Voila