

# P3 : Préparez des données pour un organisme de santé publique

---

Xiaofan LEI

# Ordre du jour

## Présentation du projet

### Données

- Présentation du jeu de données
- Vérification des données
- Nettoyage des données

### Le nutri-score

- Comment est-il calculé?
- Attribution des lettres
- Que prend en compte le calcul?
- Le nutri-score dans le monde

### Repères PNNS

- Bien choisir grâce aux groupes PNNS
- Où trouver les vitamines dans l'alimentation?
- Où trouver les minéraux dans l'alimentation?

### Fruits et légumes

- Est-ce toutes les formes de fruits et de légumes se valent-elles ?
- Apport en nutriments

## Emballage des produits alimentaires

## Exploration libre



# Présentation du projet



## Contexte



Ce projet a été constitué pour répondre à un appel à projets lancé par l'Agence "Santé publique France"

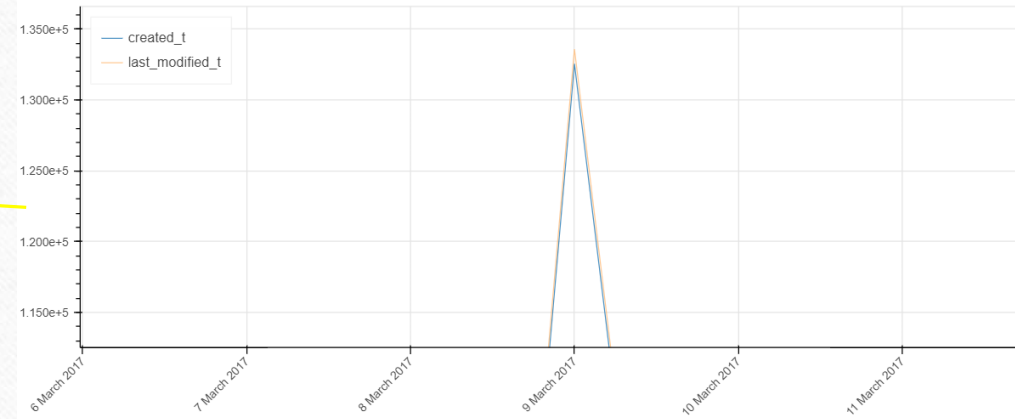
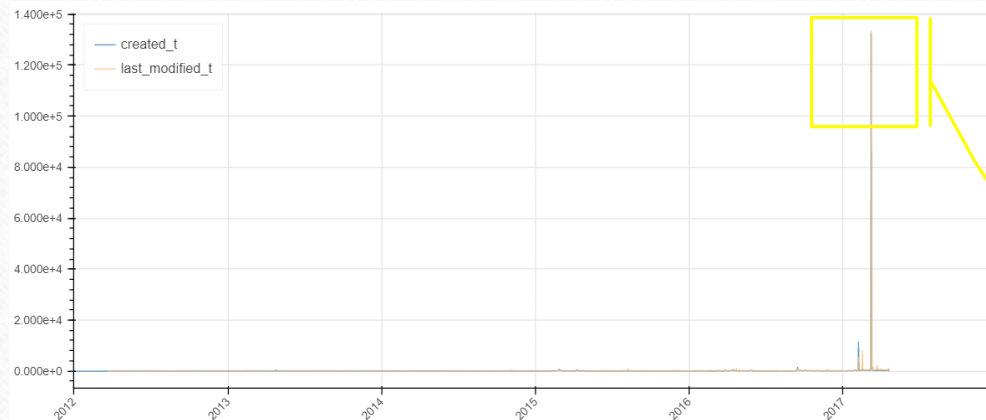
## Objectif



Ce projet a pour objectif de rendre les données de santé exploitables par les agents de L'agence "Santé publique France" afin que les choix favorables à la santé soient fondés sur des données scientifiques.

Une première analyse est réalisée en basant sur les informations nutritionnelles des produits alimentaires.

# Présentation du jeu de données



- Ce prototype est mise en regard avec les données d'Open Food , dont les produits alimentaires ont été majoritairement injectés et mises à jour en début 2017.

# Vérification des données

	no_nutriments	additives_n	ingredients_from_palm_oil_n	ingredients_from_palm_oil	ingredients_that_may_be_from_palm_oil_n	ingredients_that_may_be_from_palm_oil	nutrition_grade_uk	butyric-acid_100g	caproic-acid_100g	caprylic-acid_100g	capric-acid_100g	lauric-acid_100g	trans-fat_100g	cholesterol_100g	carbohydrates_100g	sugars_100g
count	0.0	248939.0	248939.0	0.0	248939.0	0.0	0.0	0.0	0.0	1.0	2.0	4.0	143298.0	144090.0	243588.0	244971.0
mean	-	1.936024	0.019659	-	0.055246	-	-	-	-	7.4	6.04	36.136182	0.073476	0.020071	32.073981	16.003484
std	-	2.502019	0.140524	-	0.269207	-	-	-	-	-	0.226274	24.101433	1.540223	0.358062	29.731719	22.327284
min	-	0.0	0.0	-	0.0	-	-	-	-	7.4	5.88	0.04473	-3.57	0.0	0.0	-17.86
25%	-	0.0	0.0	-	0.0	-	-	-	-	7.4	5.96	34.661183	0.0	0.0	6.0	1.3
50%	-	1.0	0.0	-	0.0	-	-	-	-	7.4	6.04	47.6	0.0	0.0	20.6	5.71
75%	-	3.0	0.0	-	0.0	-	-	-	-	7.4	6.12	49.075	0.0	0.02	58.33	24.0
max	-	31.0	2.0	-	6.0	-	-	-	-	7.4	6.2	49.3	369.0	95.238	2916.67	3520.0
mode	-	0.0	0.0	-	0.0	-	-	-	-	7.4	5.88	0.04473	0.0	0.0	0.0	0.0
nb na	320772	71833	71833	320772	71833	320772	320772	320772	320772	320771	320770	320768	177474	176682	77184	75801
% na	100.00%	22.39%	22.39%	100.00%	22.39%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	55.33%	55.08%	24.06%	23.63%

- Comprend 162 colonnes et 320772 lignes, , dont certaines colonnes sont entièrement vides ou très peu rempli (>99,5% na)
- Des valeurs aberrantes se sont glissées dans les informations nutritionnelles



# Nettoyage des données

Suppression

Suppression de 69 colonnes très peu remplies (avec >99,5% de na)

Traitement des doublons

Suppression de 22 lignes en doublon en basant sur le « code », et les lignes avec le plus de valeurs remplies sont conservées.

Traitement des valeurs aberrantes

Ce traitement ne concerne que les informations nutritionnelles : à l'exception de energy\_100g (qui se mesure par kj ou cal), les valeurs négatives ou supérieures à 100 sont effacées.

Affichage

Unités de mesure: litres>l, Kilogramme>kg, etc

Description :

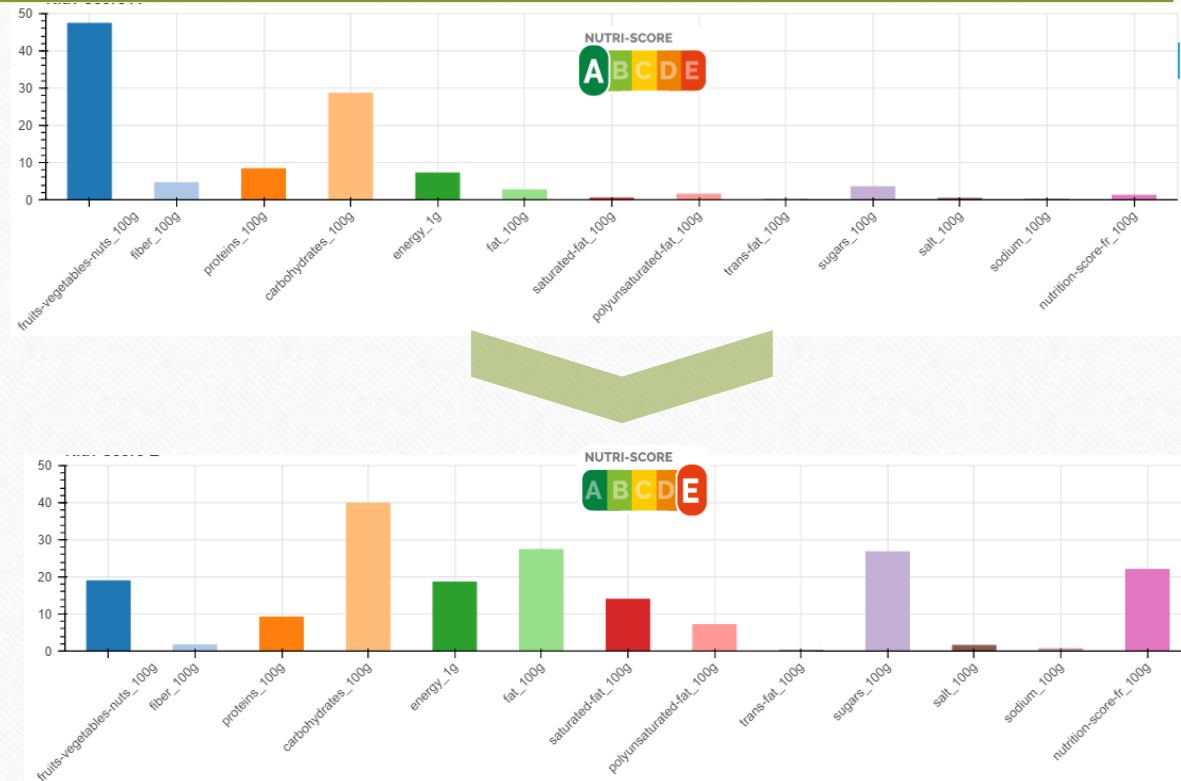
- Première lettre en majuscule
- Espace entre les mots à la place de '-'
- Traduction pour certains mots

Imputation des valeurs aux cellules vides

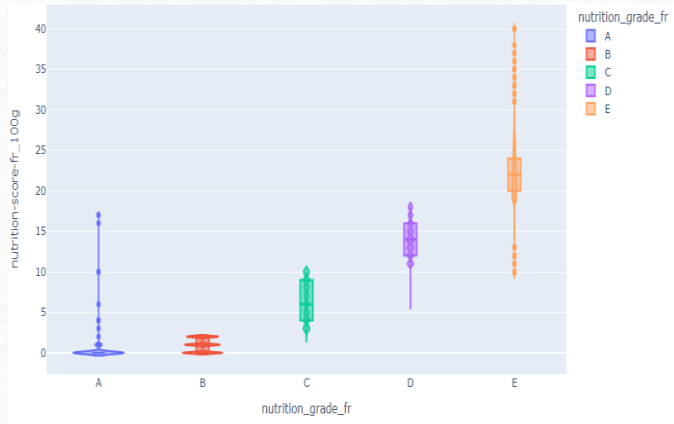
Cette opération sera réalisée au fur et à mesure en fonction des analyses

# Le nutri-score : Comment est-il calculé?

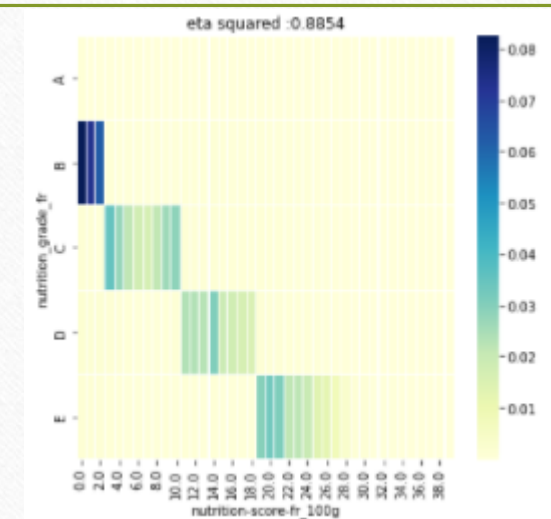
- Le **score** d'un produit résulte de l'attribution de points positifs (teneur en fruits et légumes, fibres) et de points négatifs (valeur énergétique du produit, teneur en sucre, sel et acides gras saturés).



# Le nutri-score : Attribution des lettres



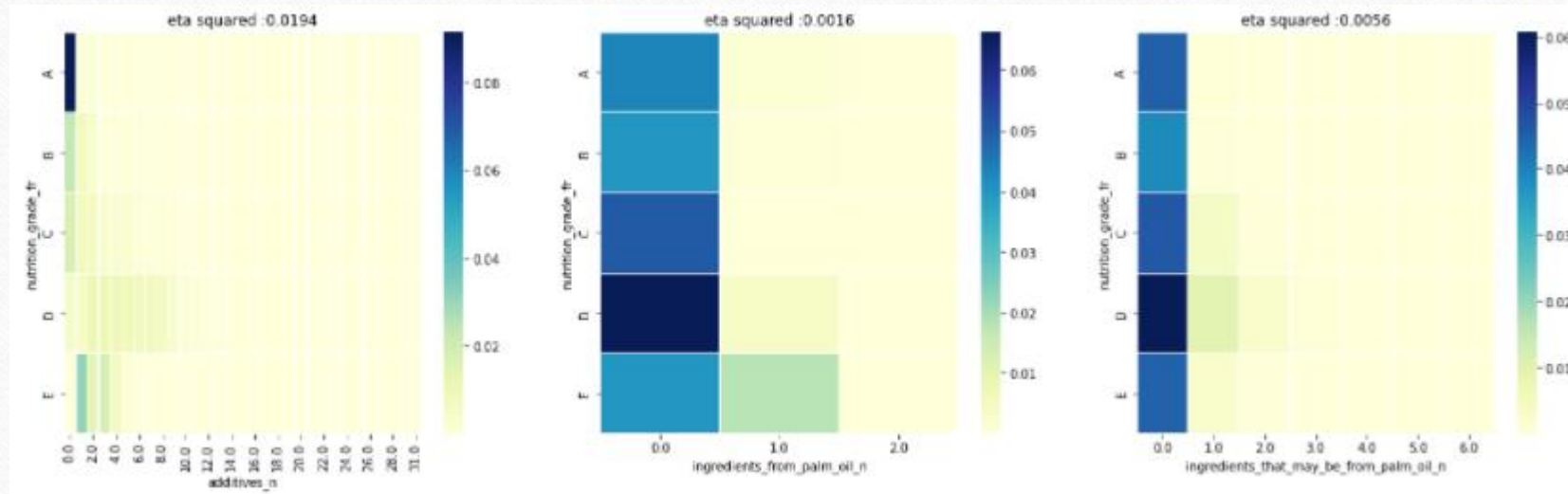
- Le résultat du score permet d'attribuer une lettre, du plus favorable (A) au moins favorable (E).
- Le résultat de l'éta carré (0,88) ainsi que la carte de chaleur nous confirment cette forte corrélation.



- Néanmoins, tous les produits, quel que soit leur score, ont leur place dans une alimentation équilibrée. Un produit avec un score élevé (D, E) invitera à une consommation plus occasionnelle ou en quantité raisonnable, alors qu'un produit avec un score bas (A, B) pourra être consommé plus fréquemment.

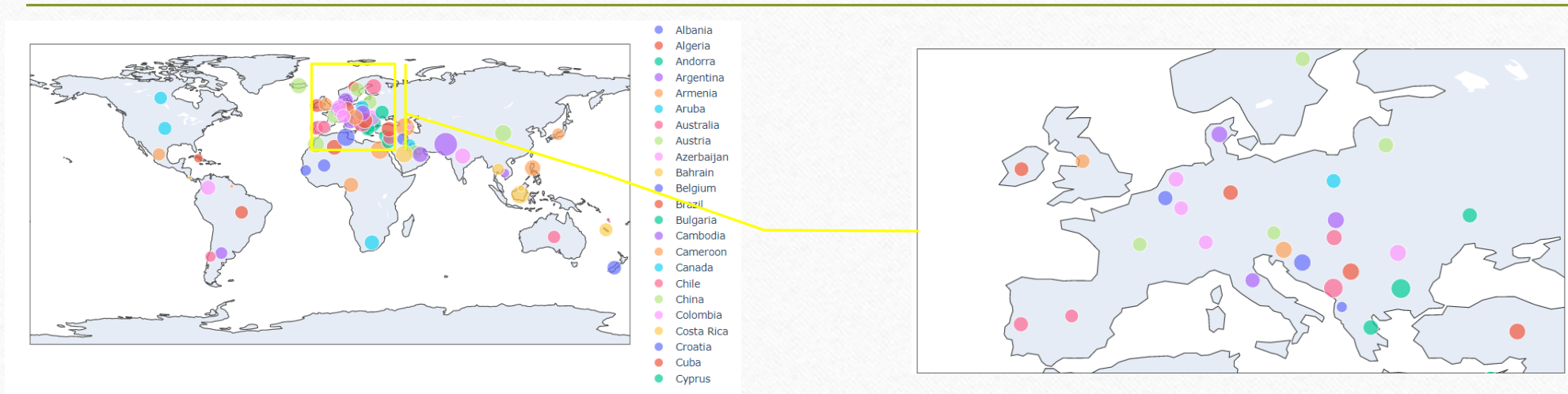


# Le nutri-score : Que prend en compte le calcul?



- Seuls les aspects nutritionnels sont pris en compte dans calcul du Nutri-Score. Il n'inclut pas d'autres dimensions comme les additifs ou les ingrédients potentiellement nocifs.

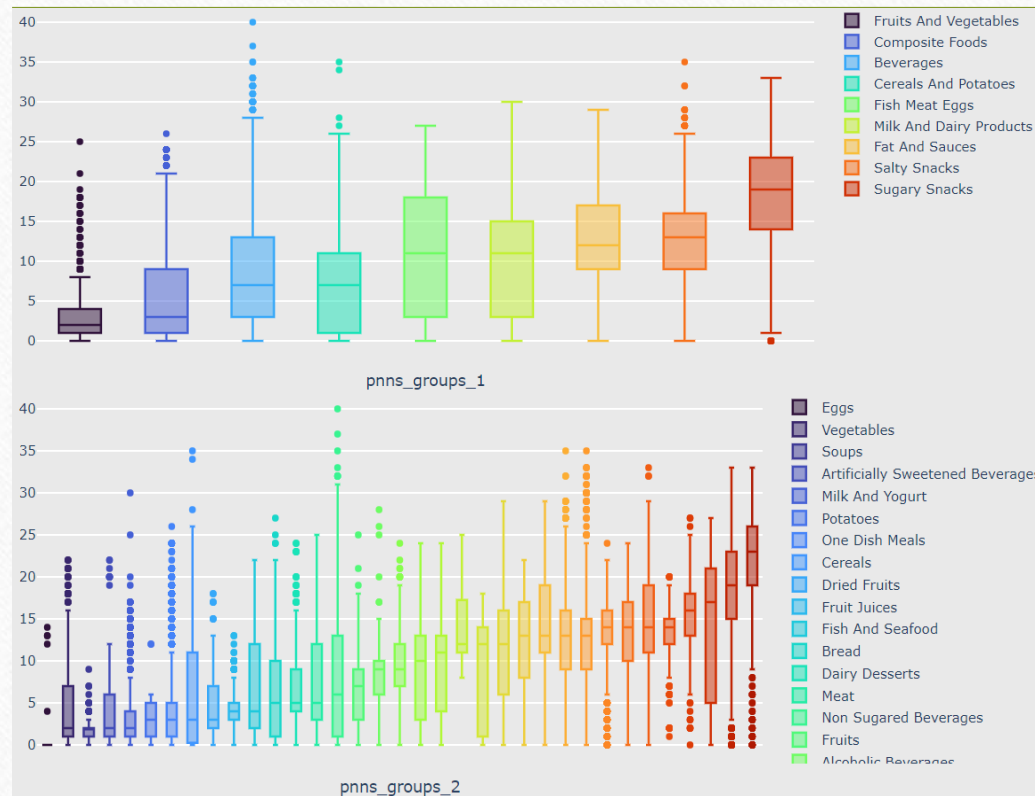
# Le nutri-score dans le monde



- Le nutri-score a déjà gagné du terrain en Europe, mais a encore du chemin à faire pour s'internationaliser.
- Selon les scores, les produits vendus dans les pays de l'Europe de l'ouest sont légèrement plus sains que l'Europe de l'Est.

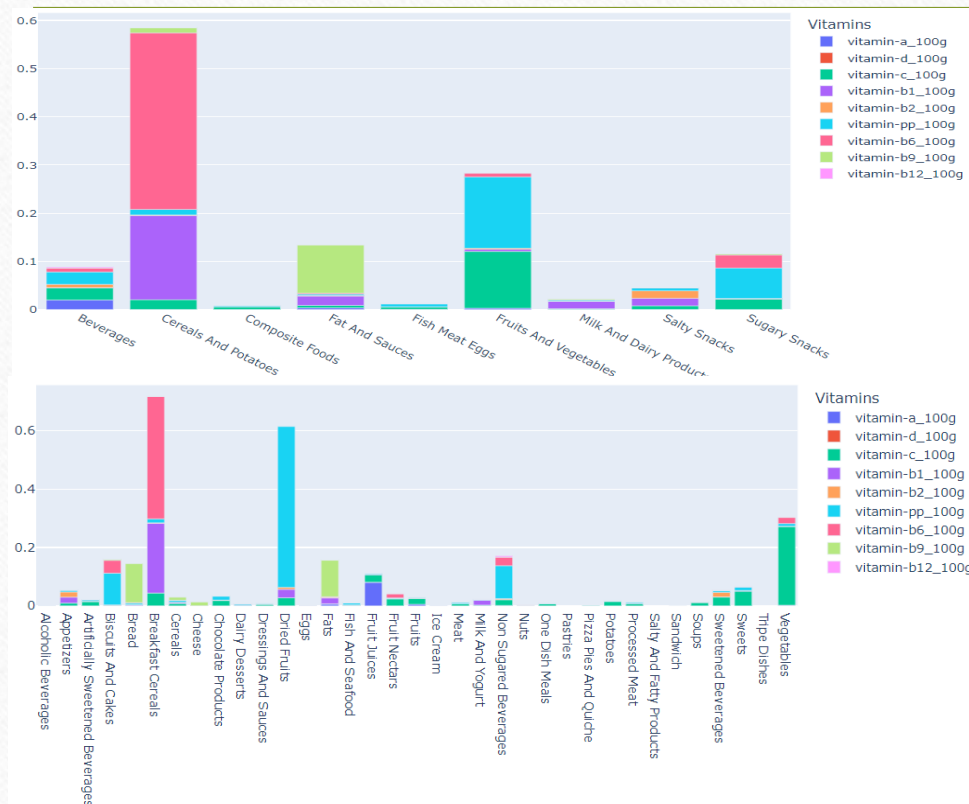


# PNNS : Bien choisir grâce aux groupes PNNS



- En France, les recommandations sur l'alimentation et l'activité physique sont définies et diffusées par les pouvoirs publics depuis la mise en œuvre du Programme national nutrition-santé (PNNS) en 2001.
- Les fruits, légumes et produits céréaliers sont à privilégier, compte tenu de leurs intérêts nutritionnels
- Les snacks, que ce soient sucrés ou salés, doivent être limités le plus possible, mais il n'est pas nécessaire de faire l'impasse sur ces produits

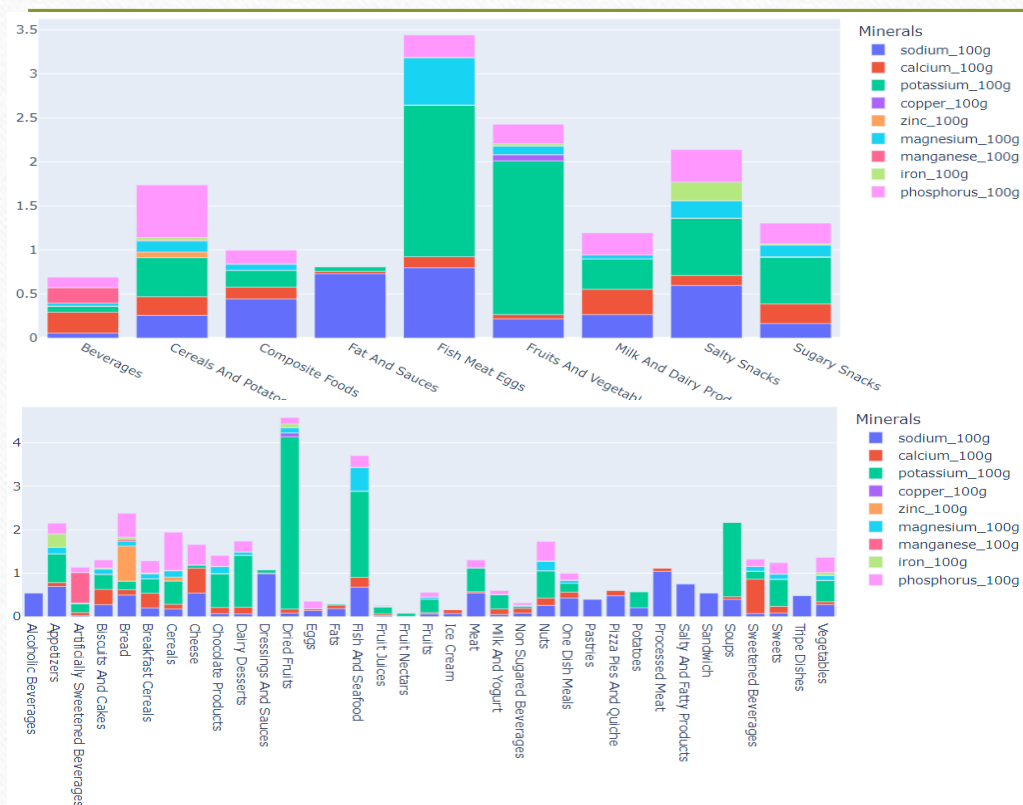
# PNNS : Où trouver les vitamines dans l'alimentation?



- Il est possible de compter sur l'alimentation pour compléter les besoins en vitamine x :
  - Les céréales, fruits et légumes sont riches en vitamines.
  - Les légumes, riches en vitamine C, qui favorise la lutte contre la fatigue, et participe au bon fonctionnement des défenses immunitaires.
  - les **fruits secs** regorgent de vitamine pp. Aussi appelé vitamine B3, elle contribue au métabolisme énergétique normal. Très peu stockée dans l'organisme ce qui implique sa consommation quotidienne.
  - L'enrichissement des **céréales** du petit déjeuner porte sur vitamine B6 , qui nous donne de l'énergie pour la journée.

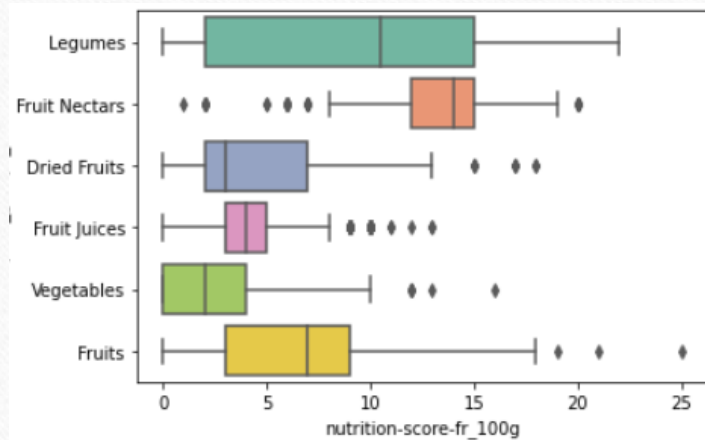


# PNNS : Où trouver les minéraux dans l'alimentation?



- Le **potassium** est nécessaire au fonctionnement normal des cellules, des nerfs et des muscles, les meilleures sources alimentaires en potassium sont les fruits, légumes, les fruits de mer, le poisson et les œufs.
- Le **sodium**, élément chimique que l'on retrouve dans la composition du sel, se retrouve en grande quantité dans de nombreux aliments, notamment des aliments transformés ou préparations industrielles.
- Le **phosphore** est une composante importante des os et des dents. Il se trouve dans la plupart des aliments riches en protéines tel que les produits laitiers, les fromages, la viande, le poisson, les œufs, les fruits de mer.

# Fruits et légumes : Est-ce toutes les formes de fruits et de légumes se valent-elles ? (1/2)



- Etape 1 : Déterminer si les différences entre les moyennes des groupes sont statistiquement significatives

	sum_sq	df	F	PR(>F)
fruitsVeggies	25732.659788	5.0	466.415909	0.0
Residual	31613.016980	2865.0	NaN	NaN

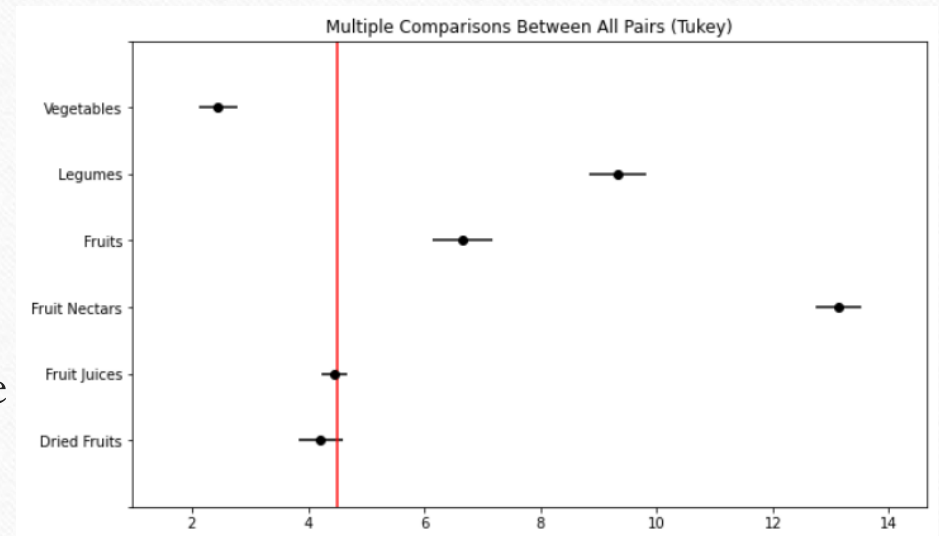
- Dans ces résultats de l'analyse de la variance (ANOVA), l'hypothèse nulle stipule que les valeurs de nutri-score moyen des différentes formes de fruits et légumes sont égales. La valeur de p (0,0) étant définitivement inférieure au seuil de signification de 0,05, nous pouvons rejeter l'hypothèse nulle et en conclure que les fruits et les légumes ont des moyennes différentes en Nutri-score.



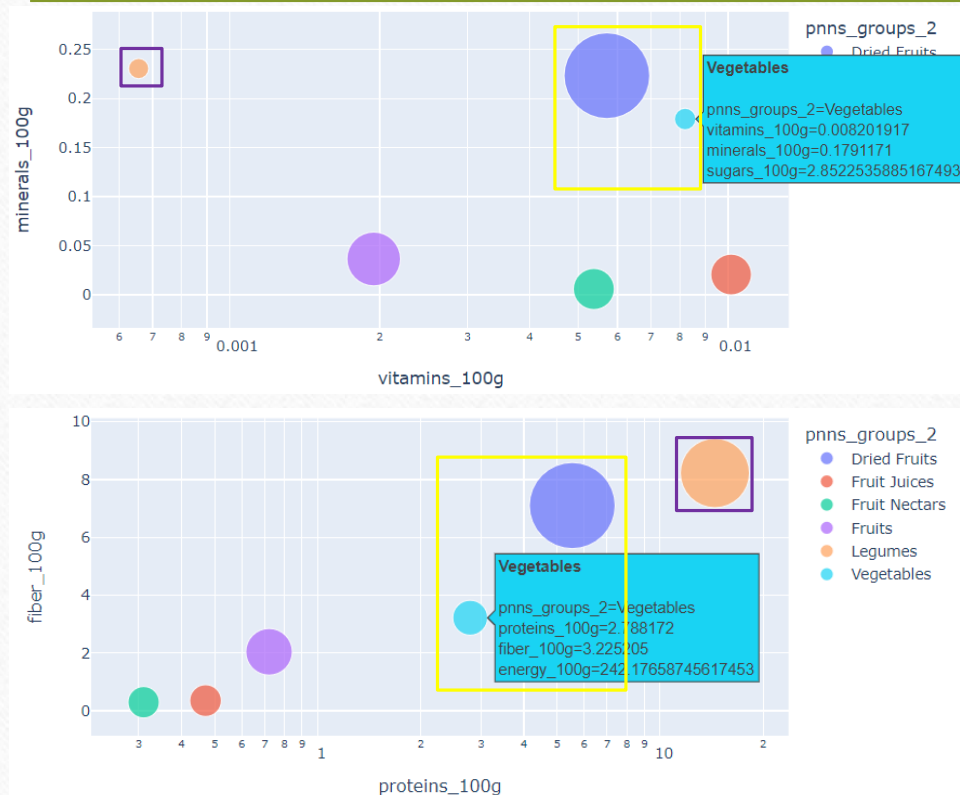
# Fruits et légumes : Est-ce toutes les formes de fruits et de légumes se valent-elles ? (2/2)

- Etape 2 : Comparer les moyennes de groupes par paire
  - Les intervalles de confiance pour les paires de moyennes des légumes, du jus de fruit et des fruits secs comprennent tous la valeur 4,5 , ce qui indique que les différences ne sont pas statistiquement significatives.
  - Le restant des paires indiquent que la différence est statistiquement significative.

Informations de groupement avec la méthode de Tukey et un niveau de confiance de 95 %:



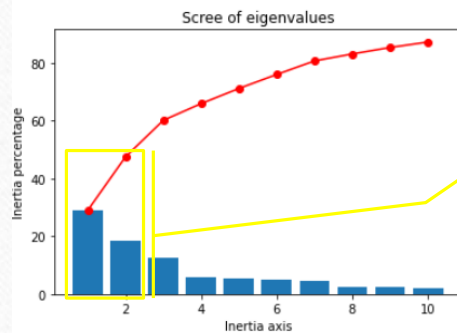
# Fruits et légumes : apport en nutriments (1 / 3)



- Les fruits secs et les légumes renferment l'essentiel des nutriments, en particulier les vitamines et les minéraux.
  - Les légumes contiennent très peu de sucre et d'énergie, à privilégier.
  - Le teneur en sucre et en énergie est très élevée dans les fruits secs, à consommer modérément.
- La principale caractéristique nutritionnelle des légumineuses ou légumes secs ( haricots, pois chiches et lentilles ) est leur **richesse** en **glucides**

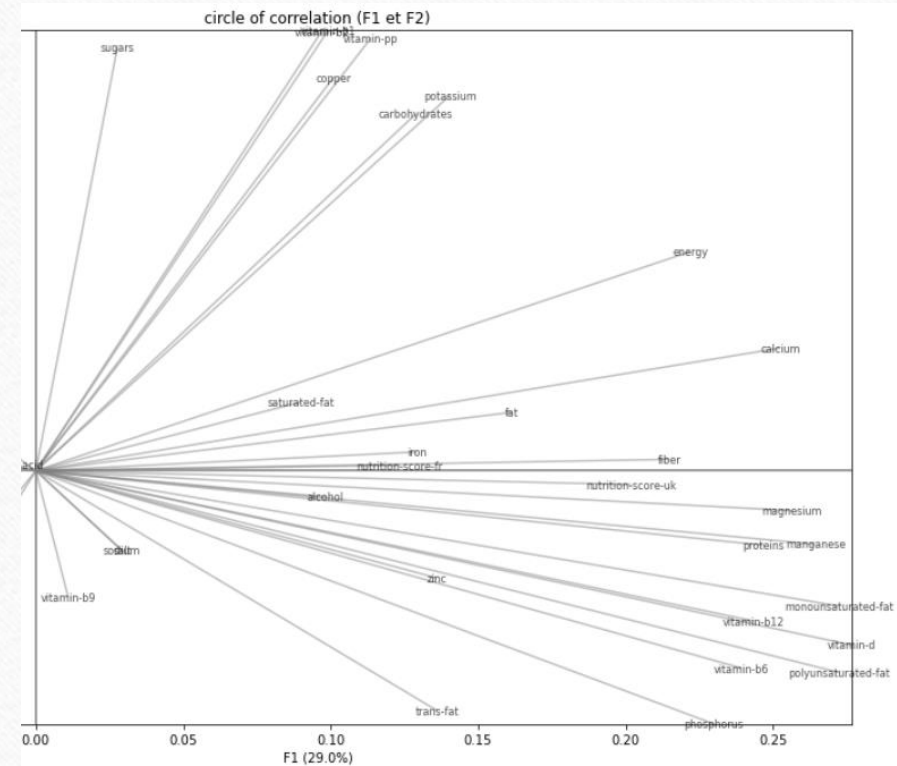


# Fruits et légumes: apport en nutriments (2/3)

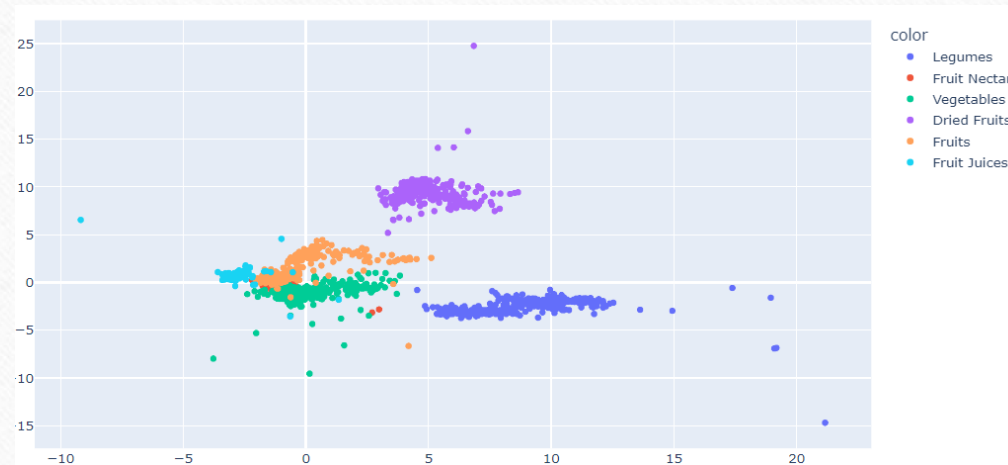


- Le « coude » de l'éboulis des valeurs propres :  $K=3$
- Les 2 premiers axes factoriels sont retenus, expliquant aux alentours de 50% de l'inertie totale.

- Les minéraux, les vitamines B, D, les fibres, et les protéines sont fortement corrélées entre eux, «synthétisables» par F1.
- Et la variable la plus corrélée à F2 est le sucre.



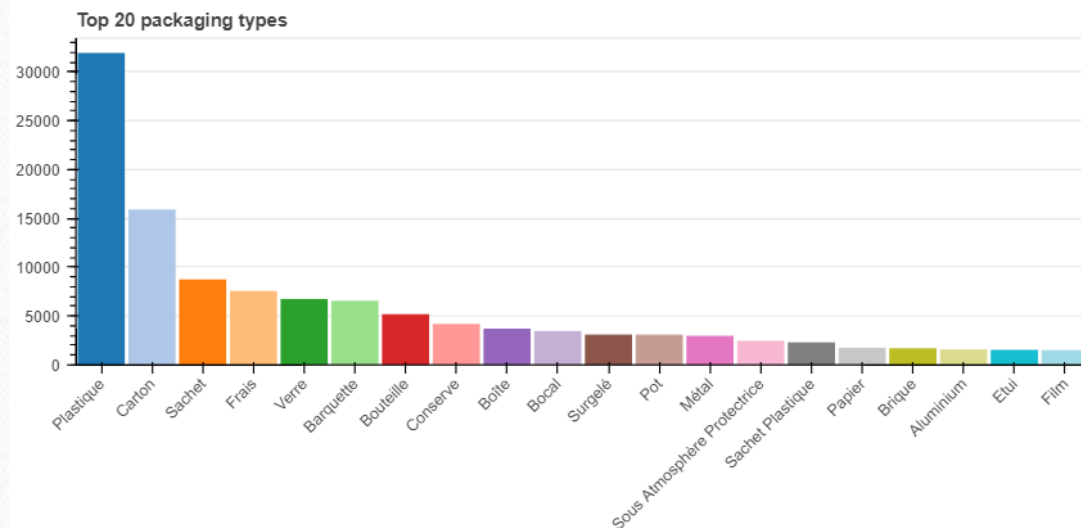
# Fruits et légumes: apport en nutriments (3/3)



- Nous pouvons constater que les fruits secs, plus étalés sur l'axe F2, sont plus riches en sucre. Alors que les légumineuses renferment plus de minéraux.



# Emballage des produits alimentaires



- Ce statistique nous relève le succès commercial des plastiques dans l’emballage des produits alimentaires, du film aux applications rigides comme les barquettes ou encore les bouteilles.
- Heureusement, des solutions alternatives comme le carton, qui n’altèrent pas le goût ou la qualité des aliments, commencent à remplacer le plastique de nos emballages.

# Exploration libre

countries

☐ countries ☐ countries\_fr ☐ countries\_tags

nutrition indicators: ☒ without nutrition-100g details  
☐ with nutrition-100g details

duplicates action: ☐ no clean up action  
☒ drop duplicates by product\_name, brands, quantity

[show dataset](#)

option1 without nutrition-100g details

option2 drop duplicates by product\_name

product ---all products---

	product_name	brands	quantity	nutrition_grade_fr
0	Farine de blé noir	Ferme ty R'nao	1kg	0
1	Banana Chips Sweetened (Whole)	NaN	nan	D
2	Peanuts	Torn & Glasser	nan	B
3	Organic Salted Nut Mix	Grizzlies	nan	D
4	Organic Polenta	Bob's Red Mill	nan	0
...	...	...	...	...
320767	Tomato & ricotta	Panzani	1	0
320768	Mint Melange Tea A Blend Of Peppermint, Lemon ...	Trader Joe's	nan	B
320769	乐吧泡菜味薯片	乐吧	50g	0
320770	Tomates aux Vermicelles	Knorr	67g	0
320771	Sugar Free Drink Mix, Peach Tea	Market Pantry	nan	0

287863 rows x 4 columns

32887 duplicate products have been removed from the dataset.

path\file.csv: C:\Users\Fan\Desktop\subset.csv

[save dataset to csv](#)

- La dernière section est dédiée aux agents plus aguerris pour des fins d'exploration libre de données.
- En plus de la visualisation directe en ligne par produit, le jeu de données est téléchargeable en local.