



PROJET 3

Préparez des données pour un organisme
de santé publique



PLAN

- Problématique
- Solution
- Résultat
- Dashboard (Prototype)

Problématique



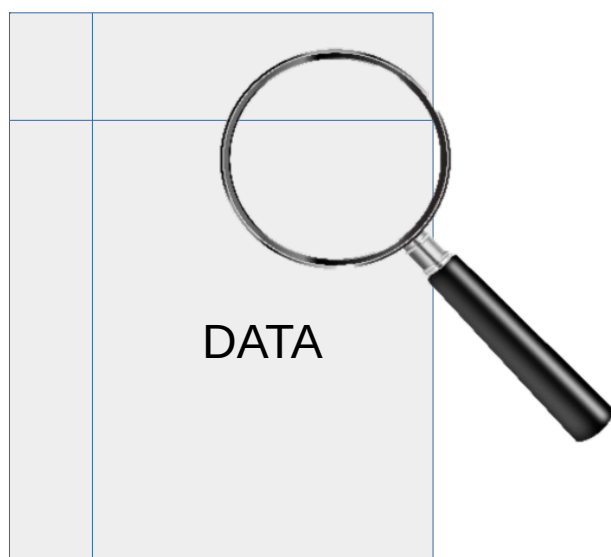
- Appel à projets



Problématique



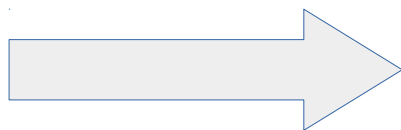
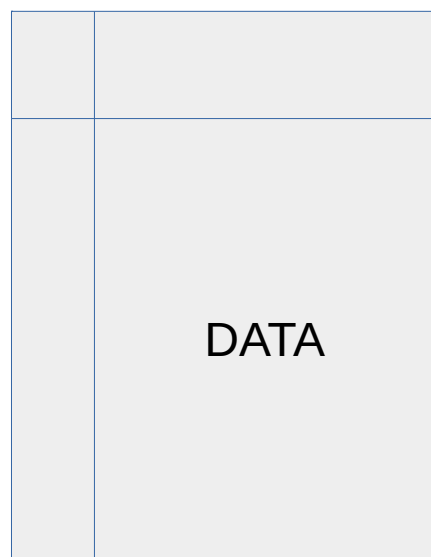
- I) explorer + visualiser



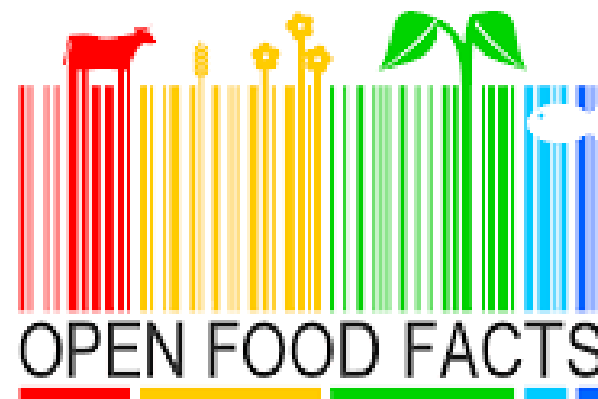
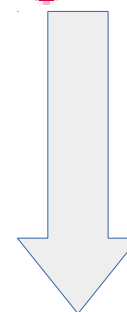
Problématique



- II) Dashboard

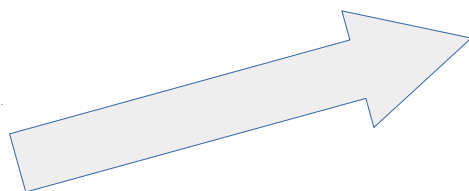
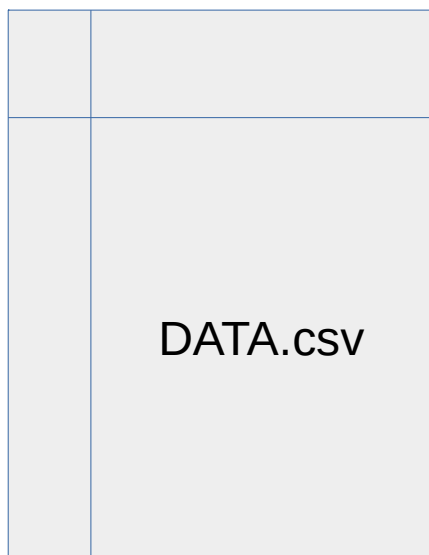


Problématique

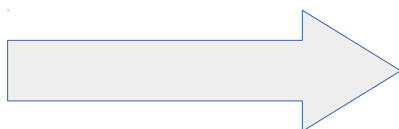


DATA.csv

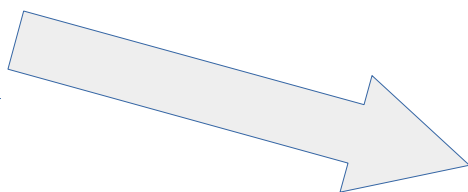
Problématique



Les informations générales
(nom, date de modification)

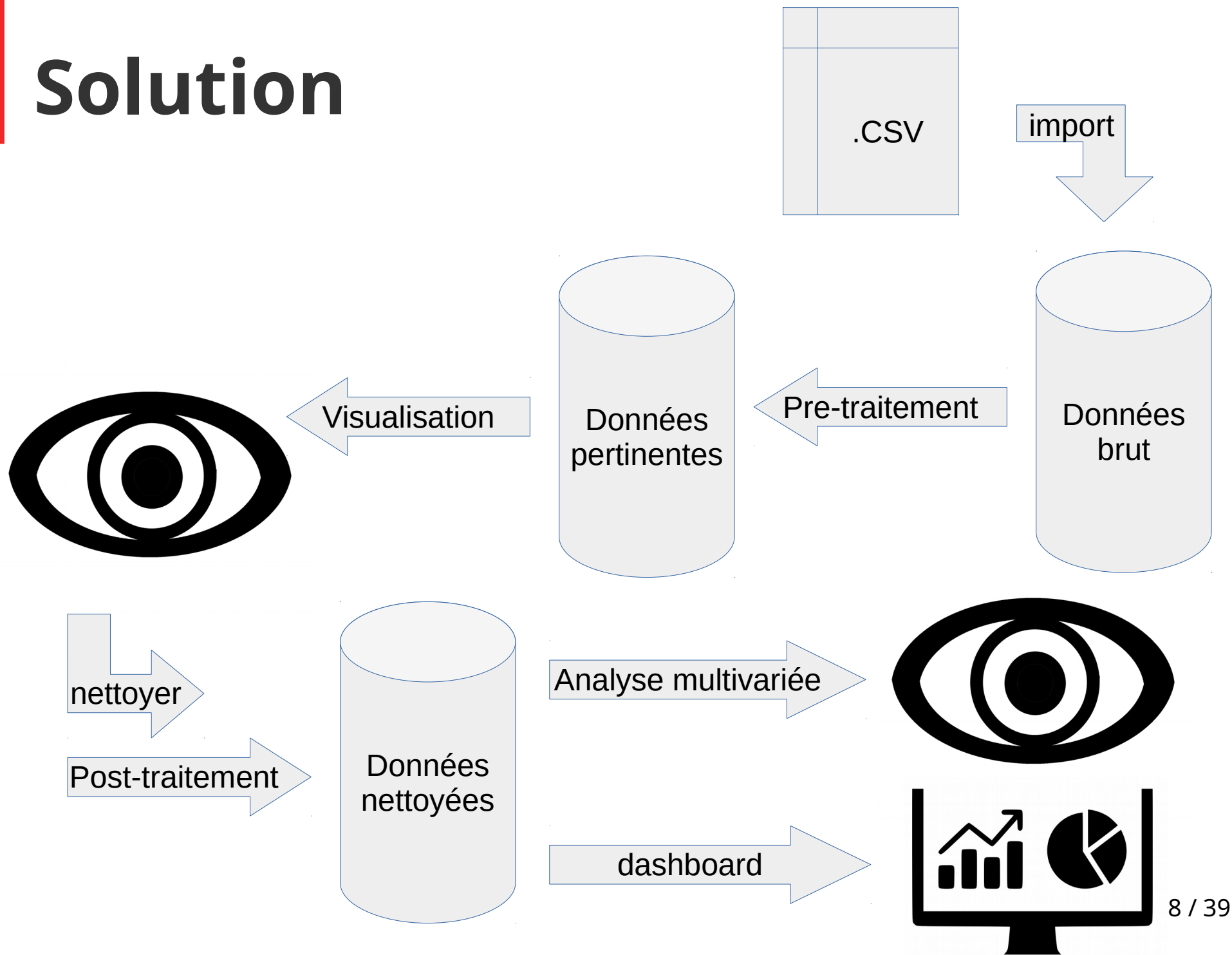


Un ensemble de tags
(catégorie du produit, localisation, origine)



Les quantités de nutriments
(10g sucre/100g)

Solution



Solution



- virtualenv



voilà

matplotlib

seaborn



- missingno
- ipywidgets



Résultat

- Hypothèses
- Pre-traitement
- I Présentation des données
- II Nettoyage des données
- III Analyse statistique multivariée
- Conclusion

Résultat / Hypothèses



- Le nutri-score de nos aliments est-il fiable ?
- Qu'en est-il en cas d'absence de ce logo ?

Résultat / Hypothèses

NUTRI-SCORE



Fibres
Protéines
Fruits et légumes

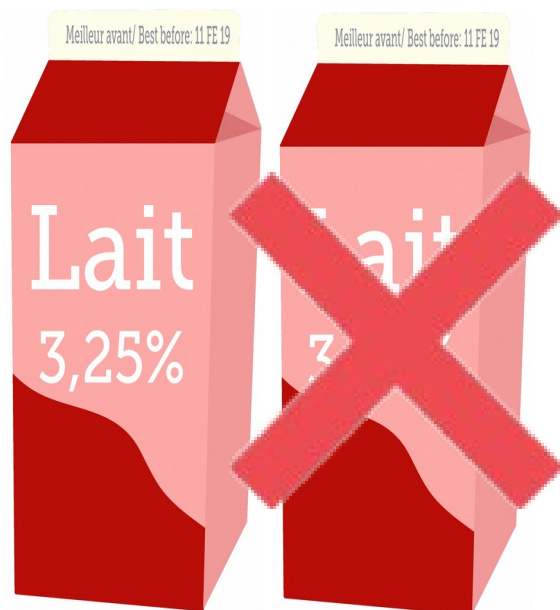


Énergie
Acides gras saturés
Sucres
Sel

Résultat / Pre-traitement



XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
xxxx	NaN	NaN	NaN	xxxx	NaN



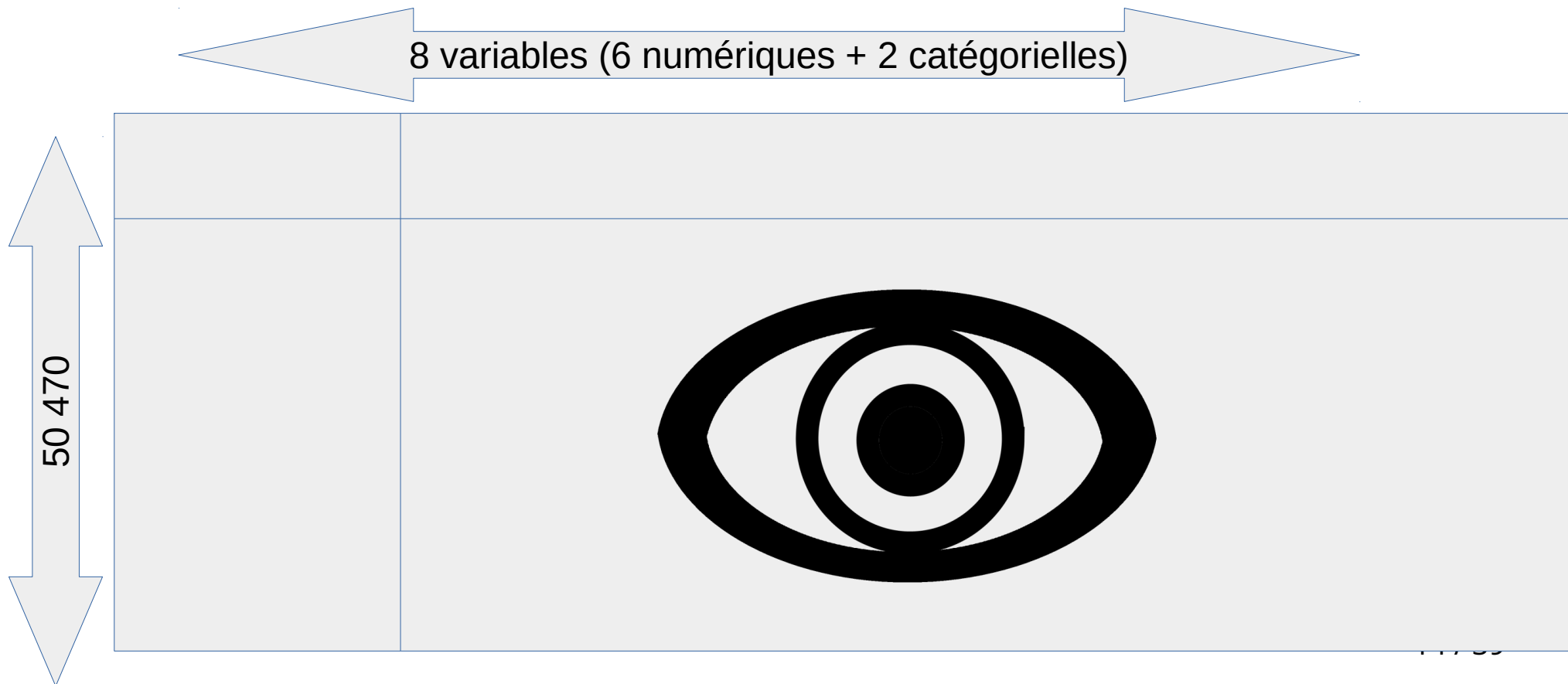
Fibres
Protéines
Fruits et légumes



Énergie
Acides gras saturés
Sucres
Sel

Résultat / I Présentation ...

- I Présentation des données



Résultat / I Présentation ...

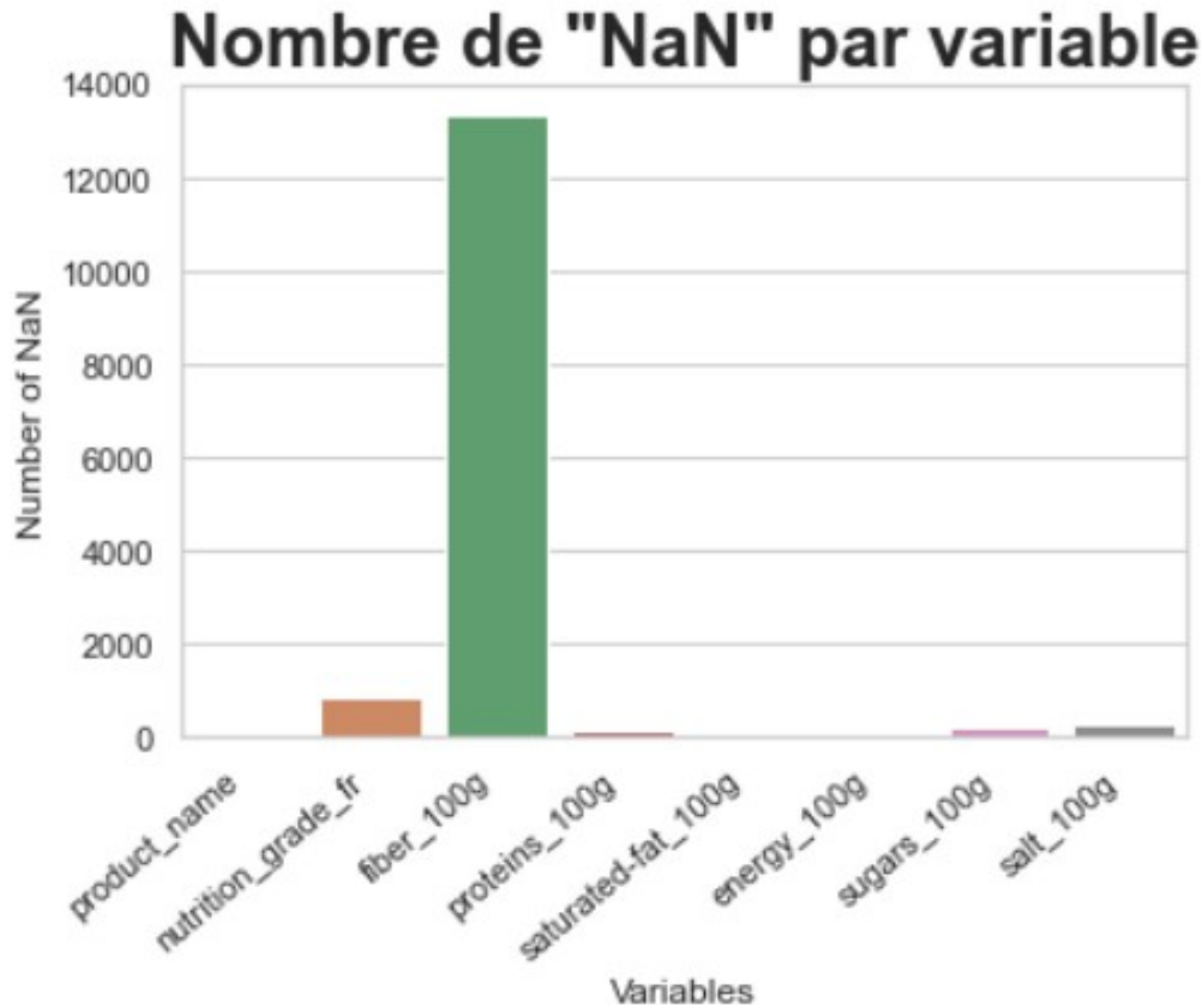
- I Présentation des données

8 variables (6 numériques + 2 catégorielles)

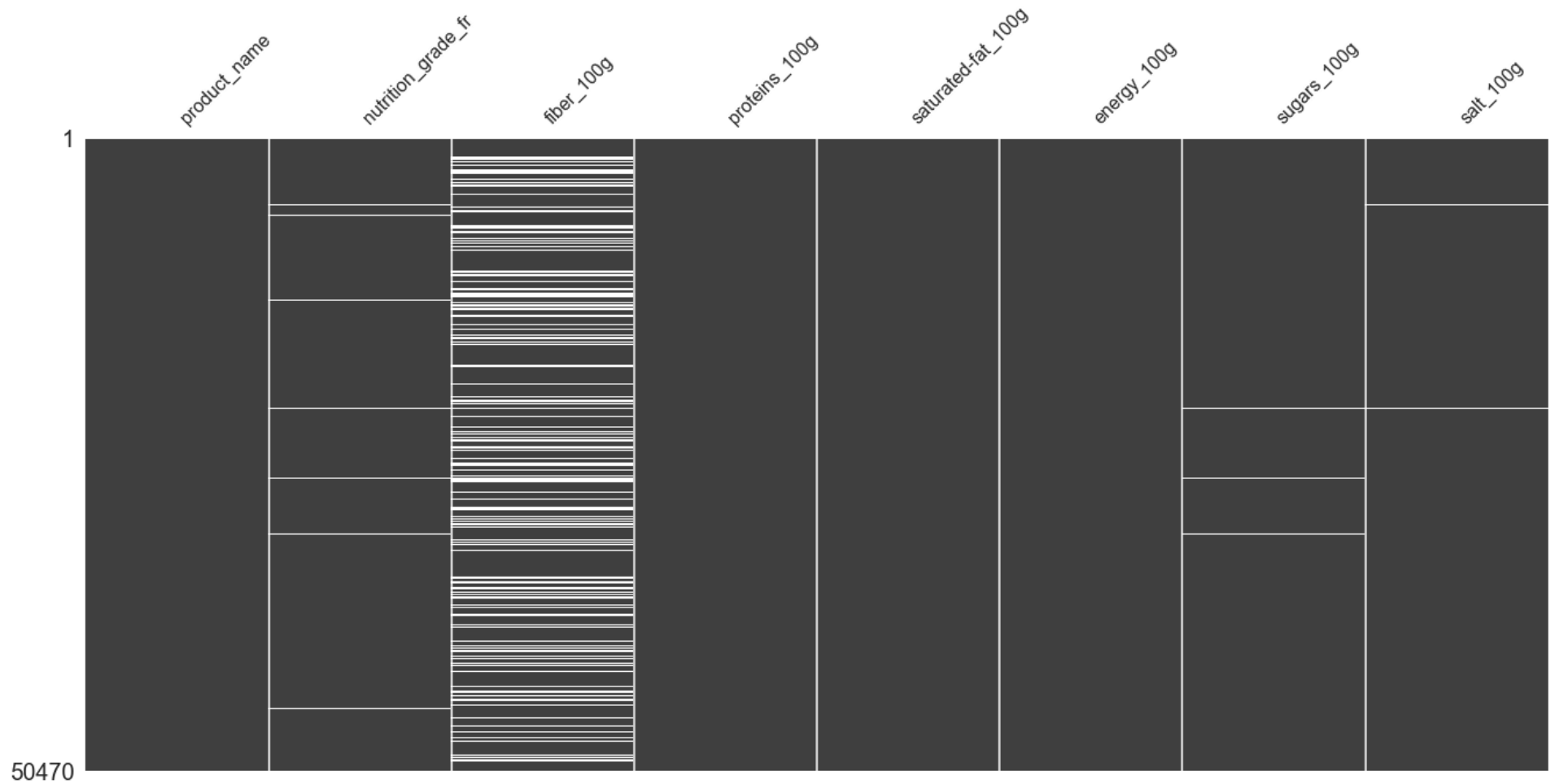
50 470

	product_name	nutrition_grade_fr	fiber_100g	proteins_100g	saturated-fat_100g	energy_100g	sugars_100g	salt_100g
106	Lion Peanut x2	e	2.5	2.5	12.5	1883.0	57.5	0.09652
190	Mini Confettis	d	0.9	0.6	0.8	1753.0	87.7	0.01000
191	Praliné Amande Et Noisette	d	3.9	9.5	2.9	2406.0	50.3	0.00300
226	Pepsi, Nouveau goût !	e	0.0	0.0	0.0	177.0	10.4	0.02540
234	Tarte Poireaux Et Lardons	d	1.4	7.5	11.0	1079.0	1.0	0.80000
...
320734	Nature cheddar and red onion crisps	c	5.6	7.8	3.4	2155.0	1.8	1.35000
320740	Verrine Cheesecake Myrtille	d	0.0	8.7	12.0	1084.0	10.5	0.29000
320741	Fiche Brevet	b	10.0	10.0	1.0	4.0	1.0	10.00000
320751	Tartines craquantes bio au sarrasin	a	5.9	13.0	0.6	1643.0	2.6	0.68000
320763	Thé vert Earl grey	c	0.2	0.5	0.2	21.0	0.5	0.02540

Résultat / I Présentation ...



Résultat / I Présentation ...



Résultat / I Présentation ...

- Variables numériques

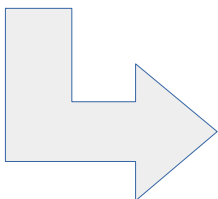
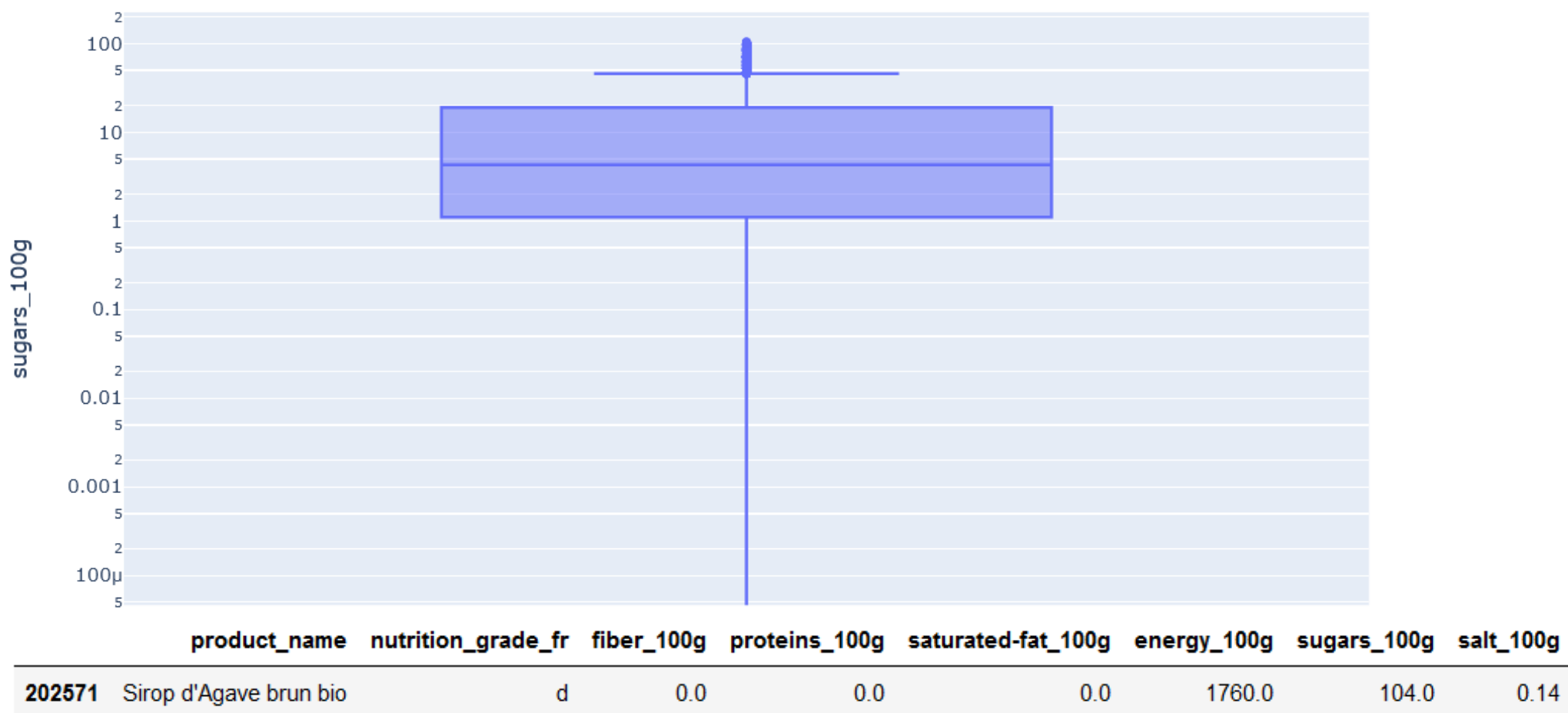
	variable	moyenne	ecart-type	max	quartile 1	quartile 2	quartile 3	exemple
0	fiber_100g	2.557448	4.649453	178.0	0.00000	1.30	3.20	2.50000
1	proteins_100g	7.774573	7.819603	100.0	2.10000	6.00	10.80	2.50000
2	saturated-fat_100g	5.461012	8.449843	210.0	0.35000	2.00	7.50	12.50000
3	energy_100g	1129.189446	1125.585183	182764.0	445.00000	1041.00	1665.00	1883.00000
4	sugars_100g	13.573569	18.821918	104.0	1.10000	4.30	19.00	57.50000
5	salt_100g	1.139029	4.148035	211.0	0.08875	0.55	1.22	0.09652

Résultat / I Présentation ...

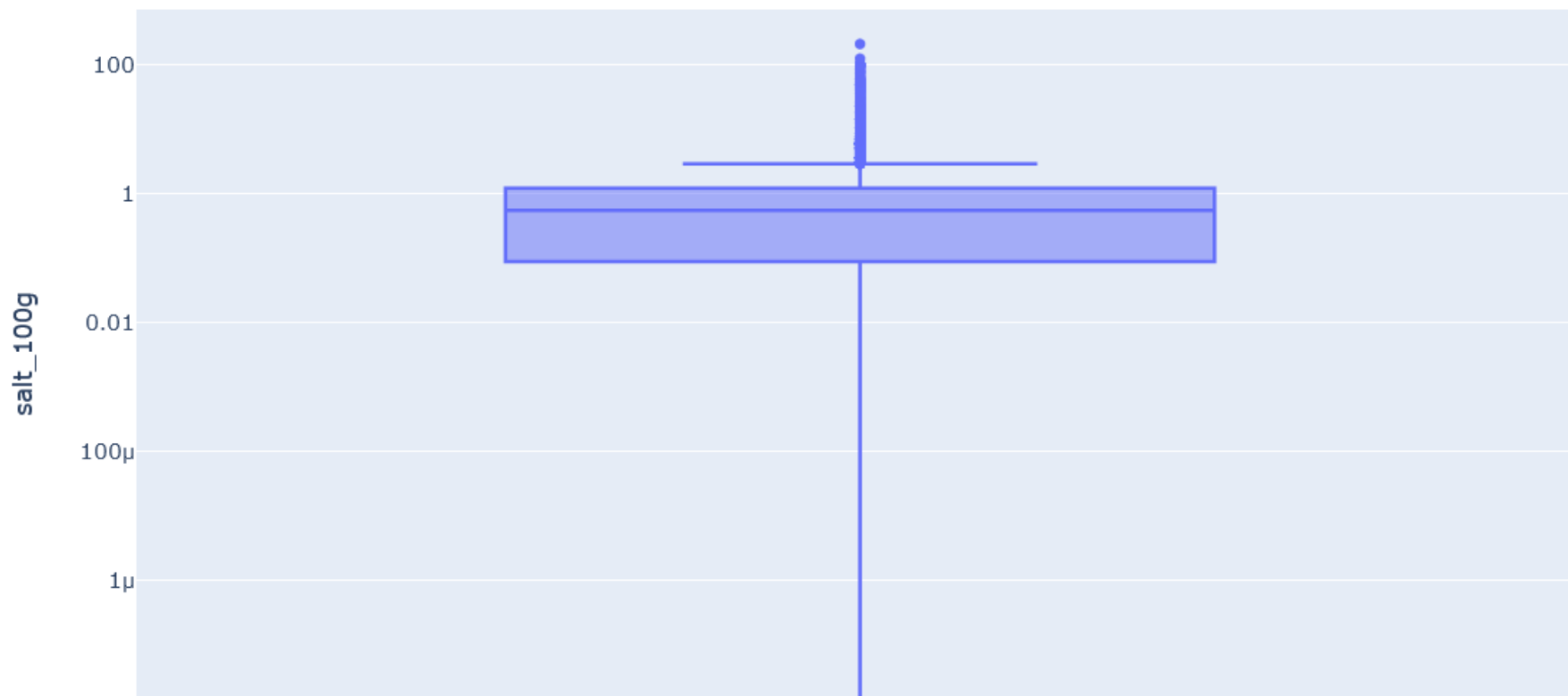
- Variables numériques

	variable	moyenne	ecart-type	max	quartile 1	quartile 2	quartile 3	exemple
0	fiber_100g	2.557448	4.649453	178.0	0.00000	1.30	3.20	2.50000
1	proteins_100g	7.774573	7.819603	100.0	2.10000	6.00	10.80	2.50000
2	saturated-fat_100g	5.461012	8.449843	210.0	0.35000	2.00	7.50	12.50000
3	energy_100g	1129.189446	1125.585183	182764.0	445.00000	1041.00	1665.00	1883.00000
4	sugars_100g	13.573569	18.821918	104.0	1.10000	4.30	19.00	57.50000
5	salt_100g	1.139029	4.148035	211.0	0.08875	0.55	1.22	0.09652

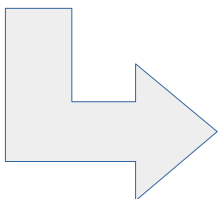
Résultat / I Présentation ...



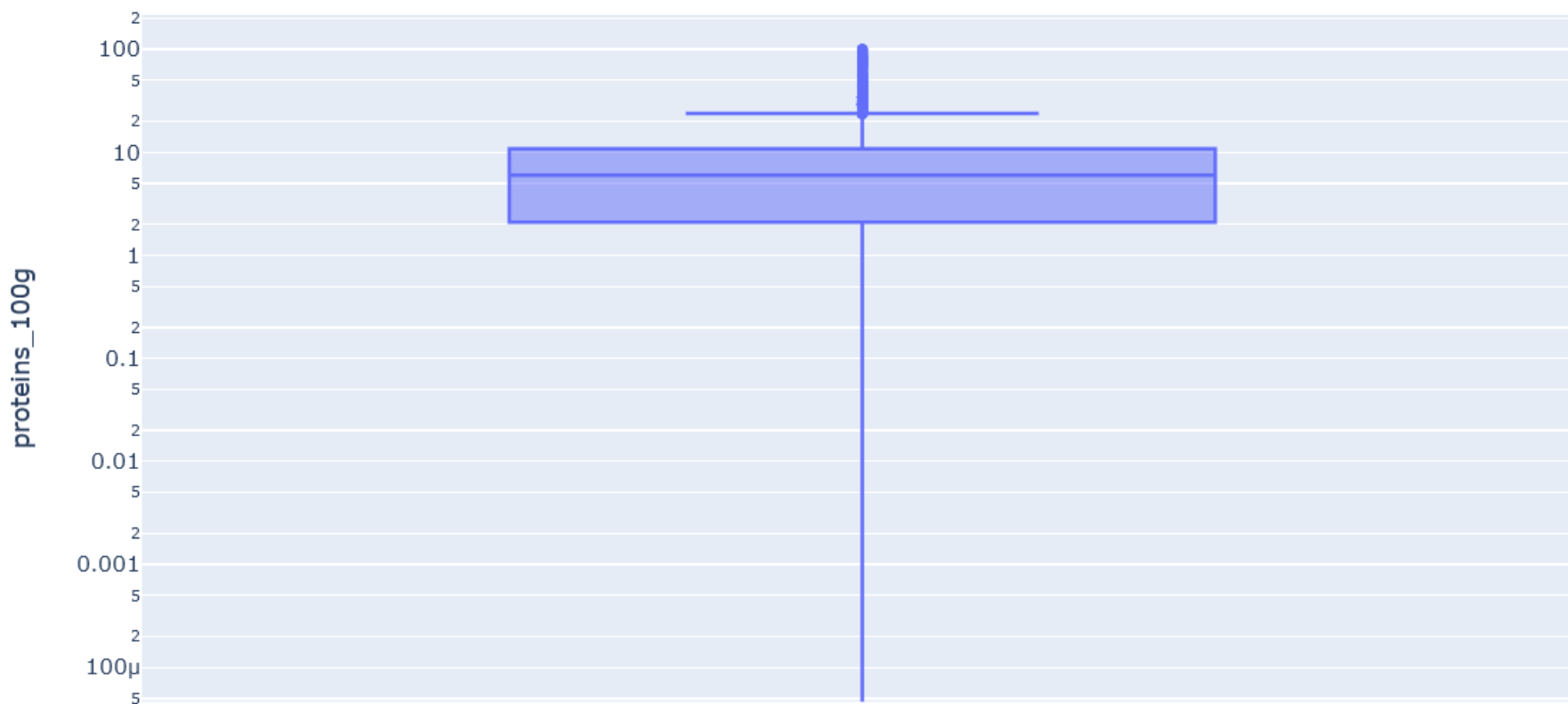
Résultat / I Présentation ...



	product_name	nutrition_grade_fr	fiber_100g	proteins_100g	saturated-fat_100g	energy_100g	sugars_100g	salt_100g
232239	Keeny Bio	e	0.741	6.3	3.33	1720.0	26.7	211.0



Résultat / I Présentation ...



	product_name	nutrition_grade_fr	fiber_100g	proteins_100g	saturated-fat_100g	energy_100g	sugars_100g	salt_100g
127525	Savarez - Jeu De Cordes	a	100.0	100.0	0.1	4.0	0.1	0.0
170118	Harry Potter & The Deathly Hallows Radcliffe /...	a	100.0	100.0	0.0	4.0	0.0	0.0
273543	Le Hobbit : La Bataille Des Cinq Armées - Versi...	a	100.0	100.0	0.0	4.0	0.0	0.0
285429	Marshall - Major Noir	c	100.0	100.0	5.0	21.0	5.0	50.0

?

Résultat / I Présentation ...



Protéines ?



Résultat / I Présentation ...

- Les erreurs faibles

Nutrient levels for 100 g ⓘ

- 1.2 g **Fat** in low quantity
- 0.9 g **Saturated fat** in low quantity
- 2 g **Sugars** in low quantity
- 1.4 g **Salt** in moderate quantity

Valeurs nutritionnelles		
Pour 1 portion de 50 g :		Pour 100 g :
énergie :	84 kJ / 20 kcal	énergie : 163 kJ / 39 kcal
matières grasses	0,5 g	matières grasses 0,9 g
dont acides gras saturés	0,1 g	dont acides gras saturés 0,1 g
glucides	3,4 g	glucides 6,8 g
dont sucres	1,3 g	dont sucres 2,5 g
fibres alimentaires	0,5 g	fibres alimentaires 0,5 g
protéines	0,3 g	protéines 0,6 g
sel	0,43 g	sel 0,84 g

Ce produit contient 4 portions de 50 g

Résultat / I Présentation ...

- Les erreurs fortes

Nutrient levels for 100 g 

- 54 g **Fat** in high quantity
- 15 g **Saturated fat** in high quantity
- 0 g **Sugars** in low quantity
- 160 g **Salt** in high quantity

Nutrition facts	
Nutrition facts	As sold for 100 g / 100 ml
Energy (kJ)	?
Energy (kcal)	10 kcal
Energy	42 kJ (10 kcal)
Fat	1 g
- Saturated fat	0 g
Carbohydrates	1 g
- Sugars	0.9 g
Proteins	100 g
Salt	1 g
Sodium	0.4 g



Valores Nutricionales Declaração Nutricional	Por 100g
Valor energético / Energia	1998 kJ 486 kcal
Grasa, de las cuales / Lipidos, dos quais	54 g
Saturadas / Saturados	15 g
Mono-insaturadas / Monoinsaturados	27 g
Poli-insaturadas / Polinsaturados	12 g
Hidratos de carbono /	0 g
De los cuales azúcares / Dos quais açúcares	0 g
Proteínas	0 g
Sal	0,16 g
Vitamina E	12mg (100% IR -)
Omega 3 - ácido alfa-linolénico	3,6 g

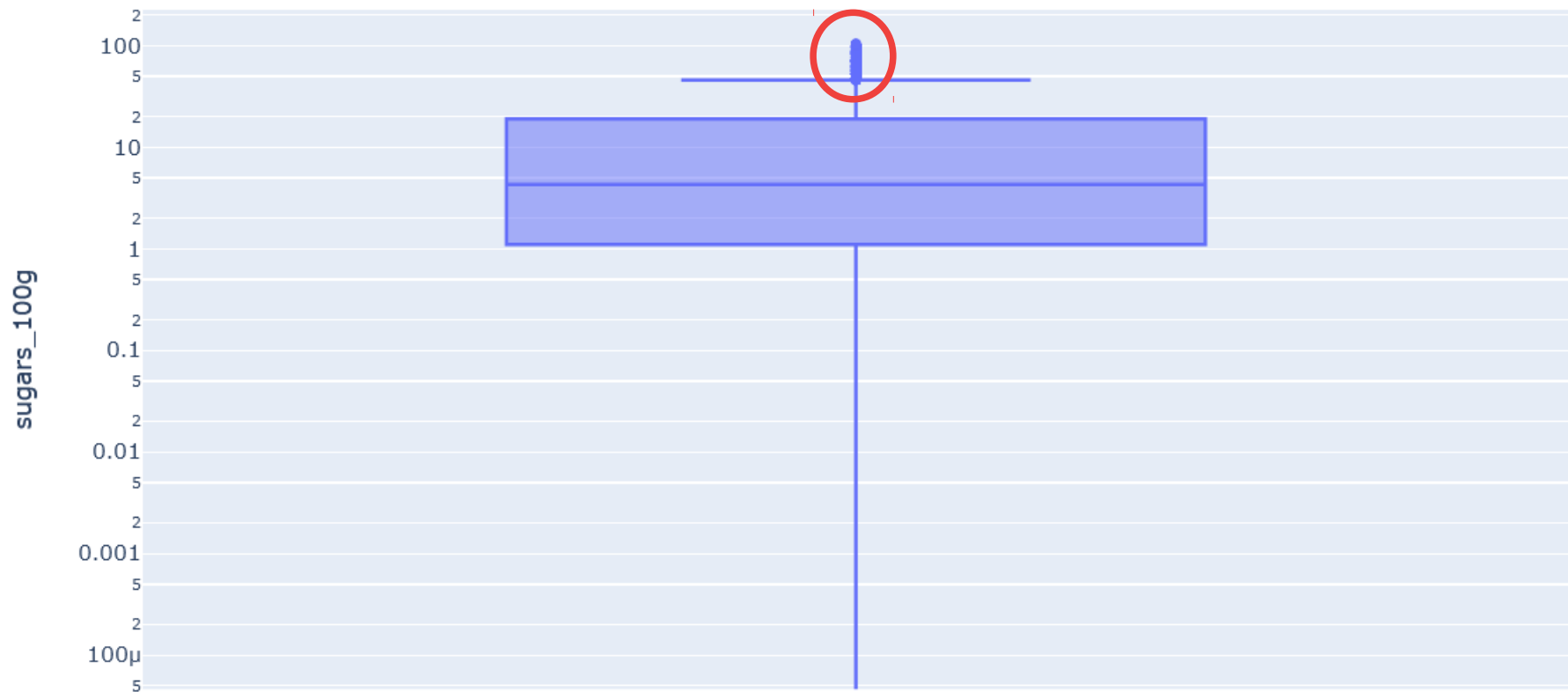
Résultat / II Nettoyage ...

- II Nettoyage des données

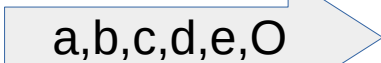


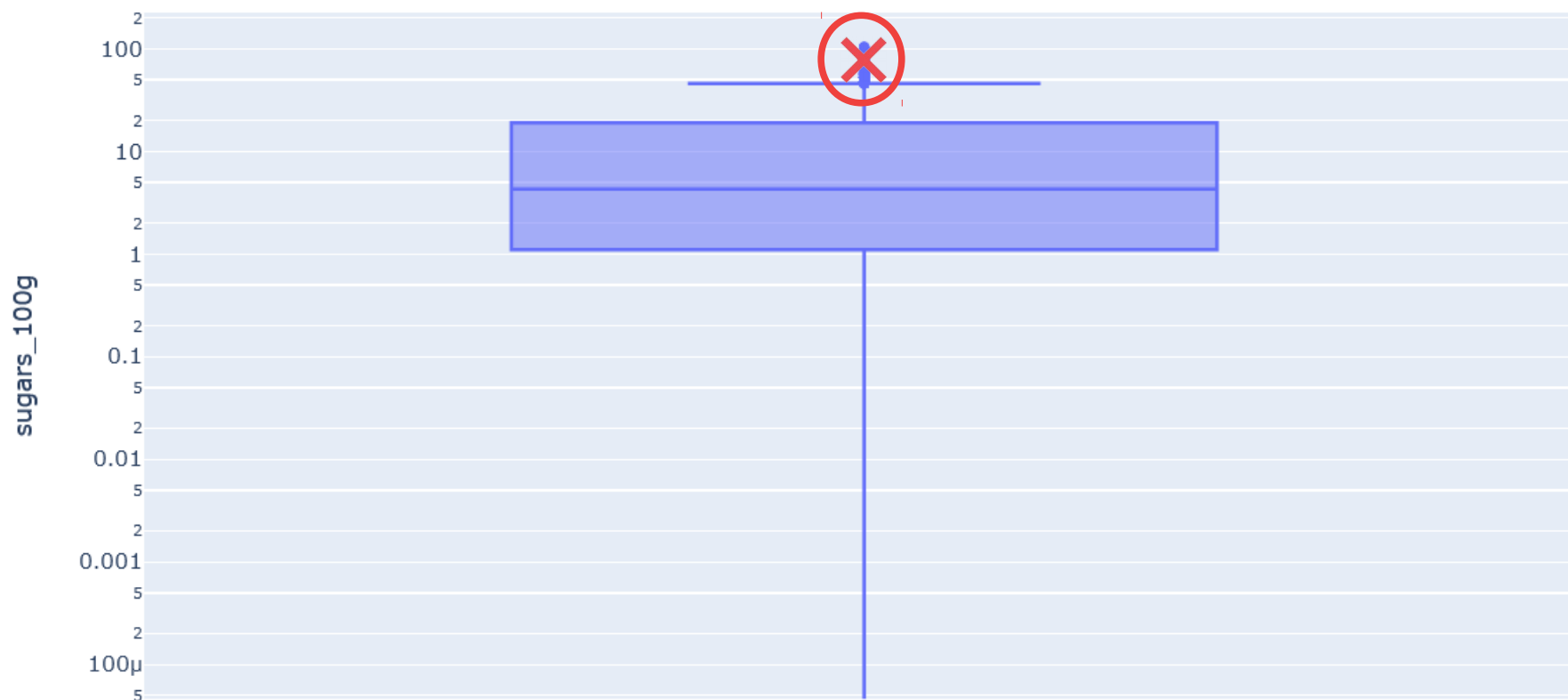
Résultat / II Nettoyage ...

- Valeurs aberrantes → écart interquartile



Résultat / II Nettoyage ...

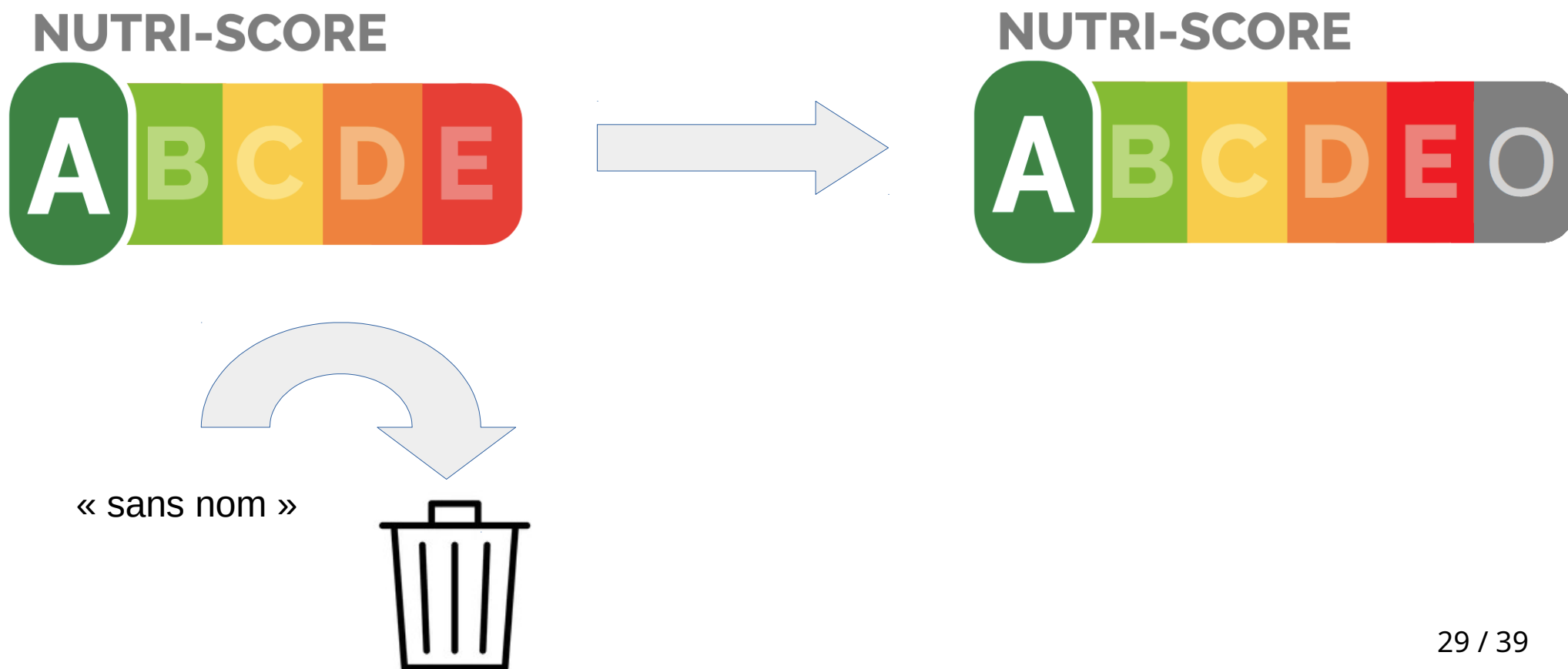
- Valeurs aberrantes  écart interquartile



- 50 000  - 22 % 39 000 individus

Résultat / II Nettoyage ...

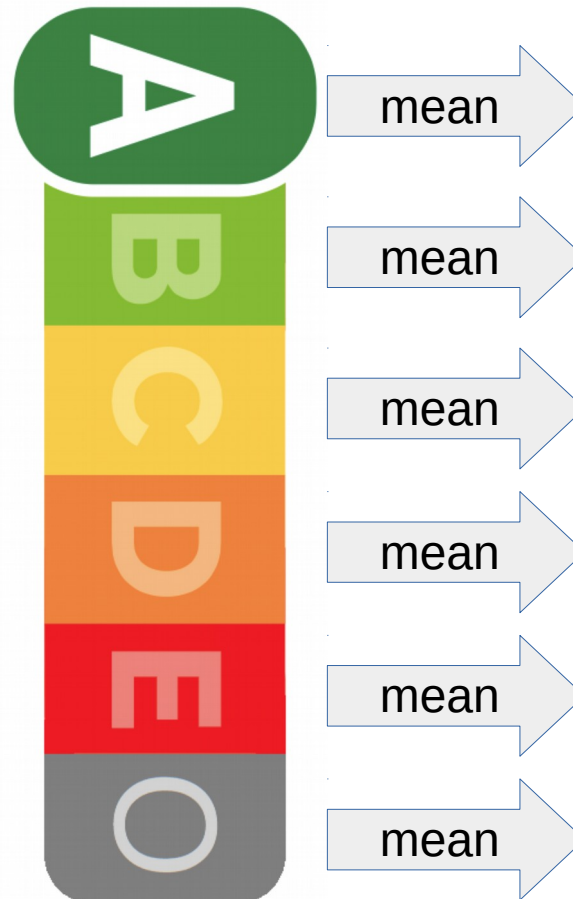
- Imputer : Valeurs manquantes (catégorielles)



Résultat / II Nettoyage ...

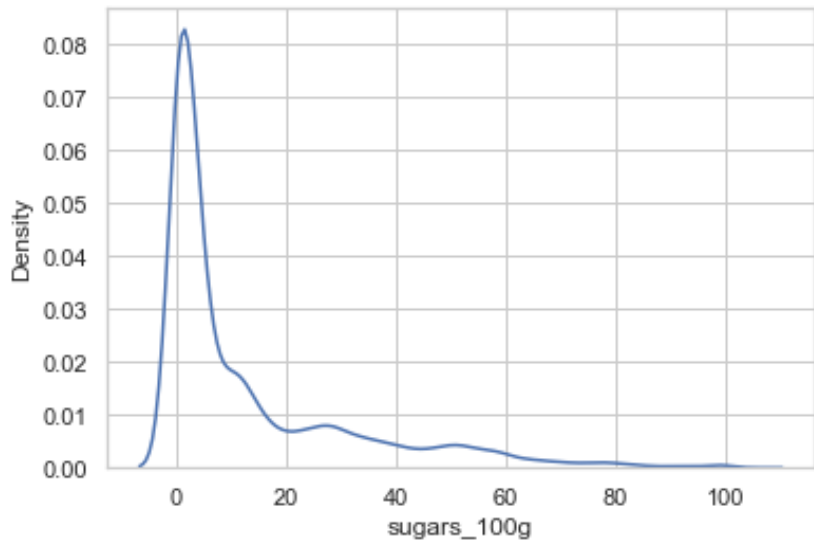
- Imputer : Valeurs manquantes (numériques)

NaN

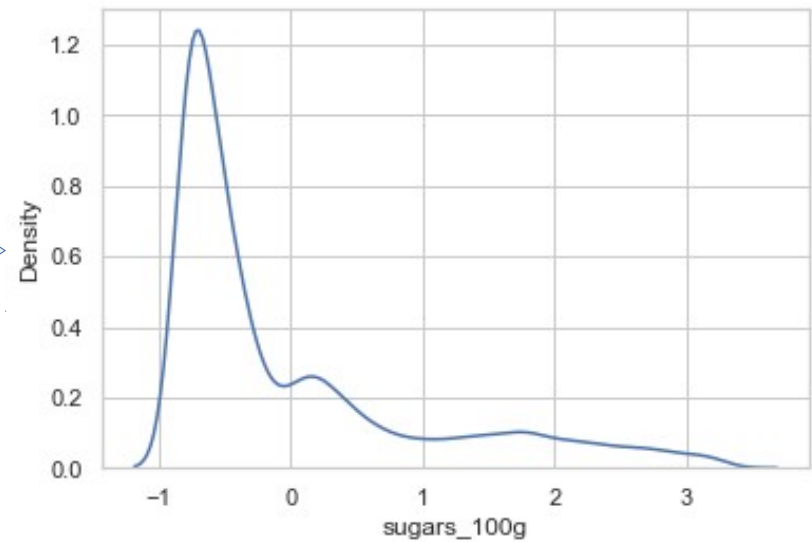


Résultat / II Nettoyage ...

- Post-traitement

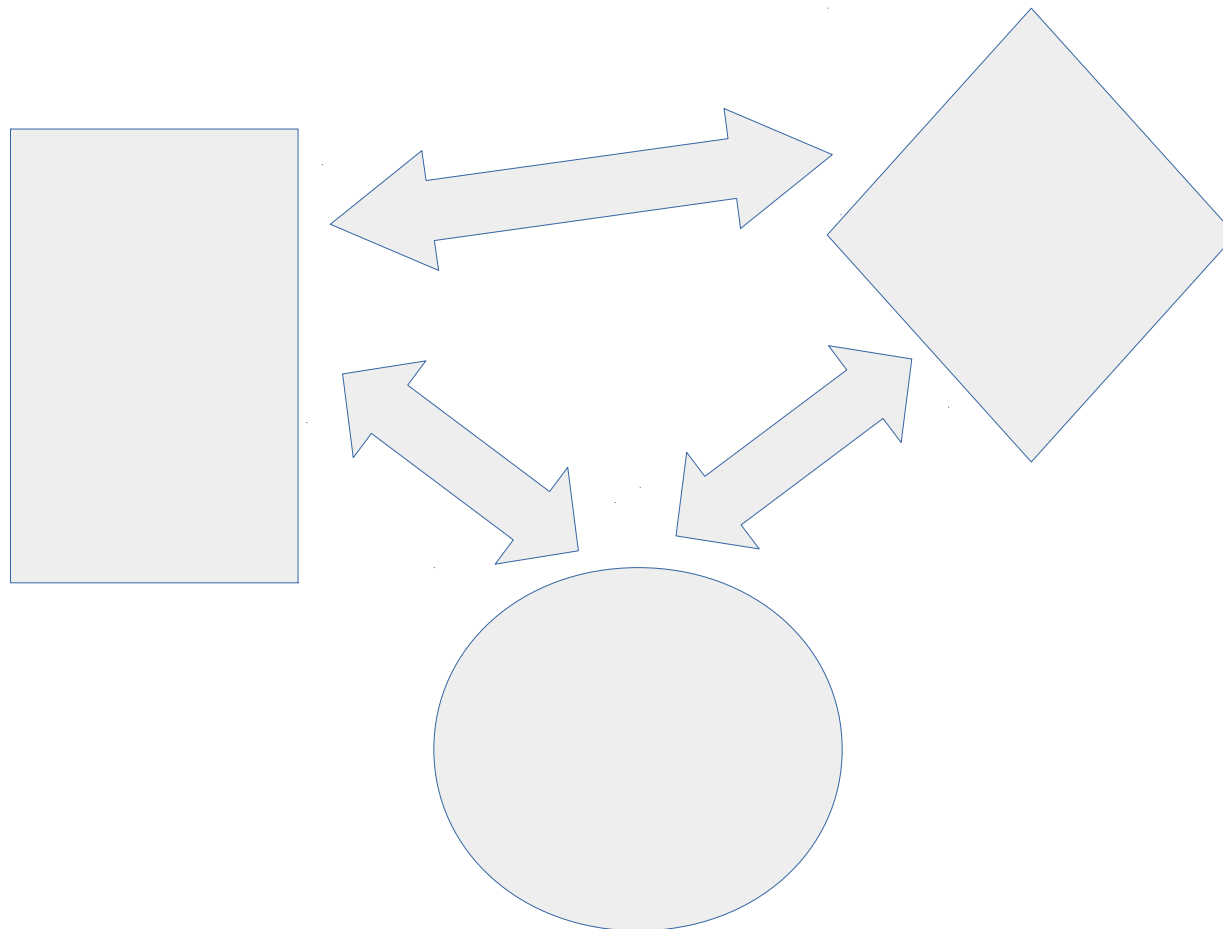


Centrer réduire



Résultat / III Analyse ...

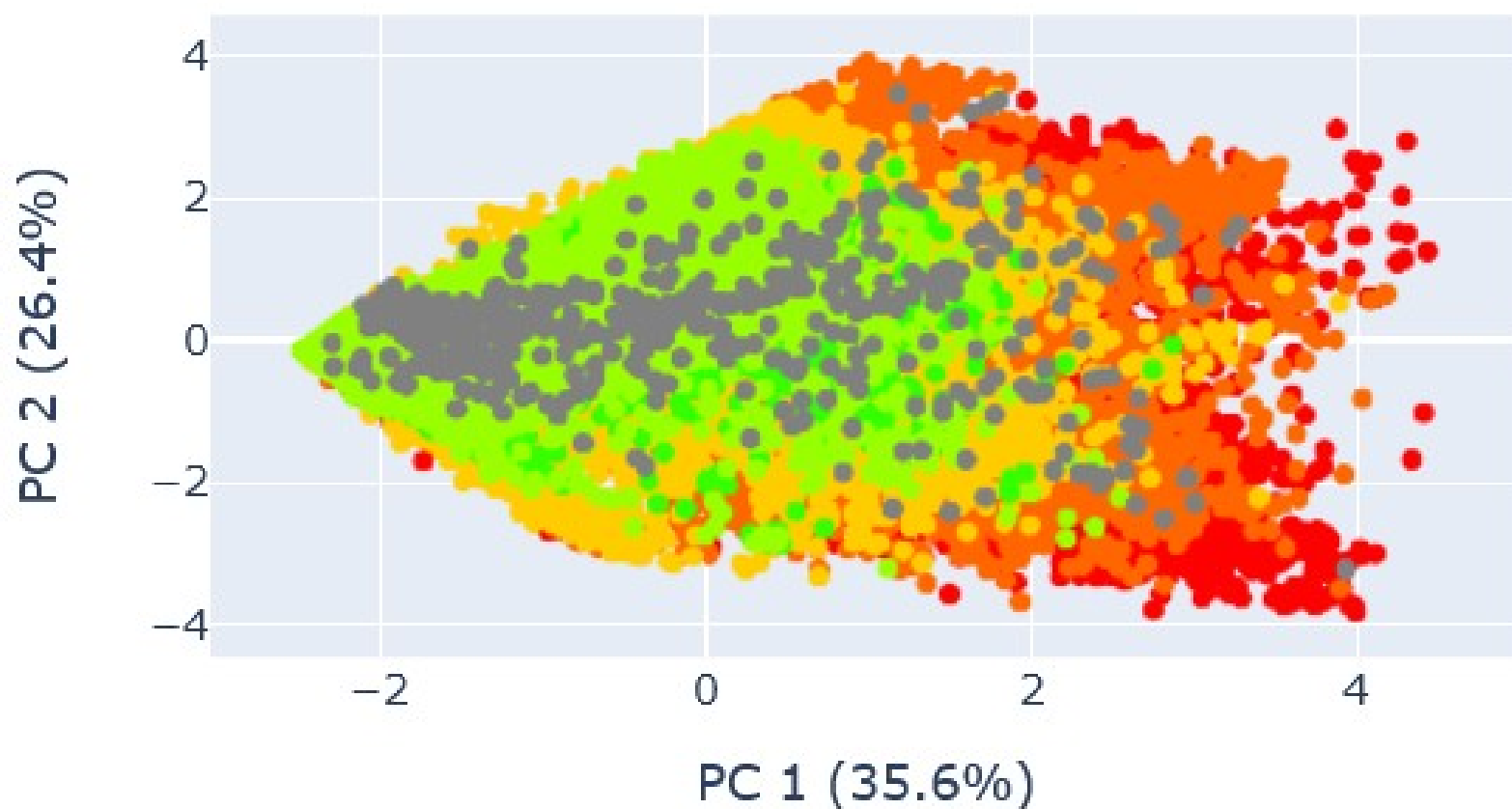
- III Analyse statistique multivariée



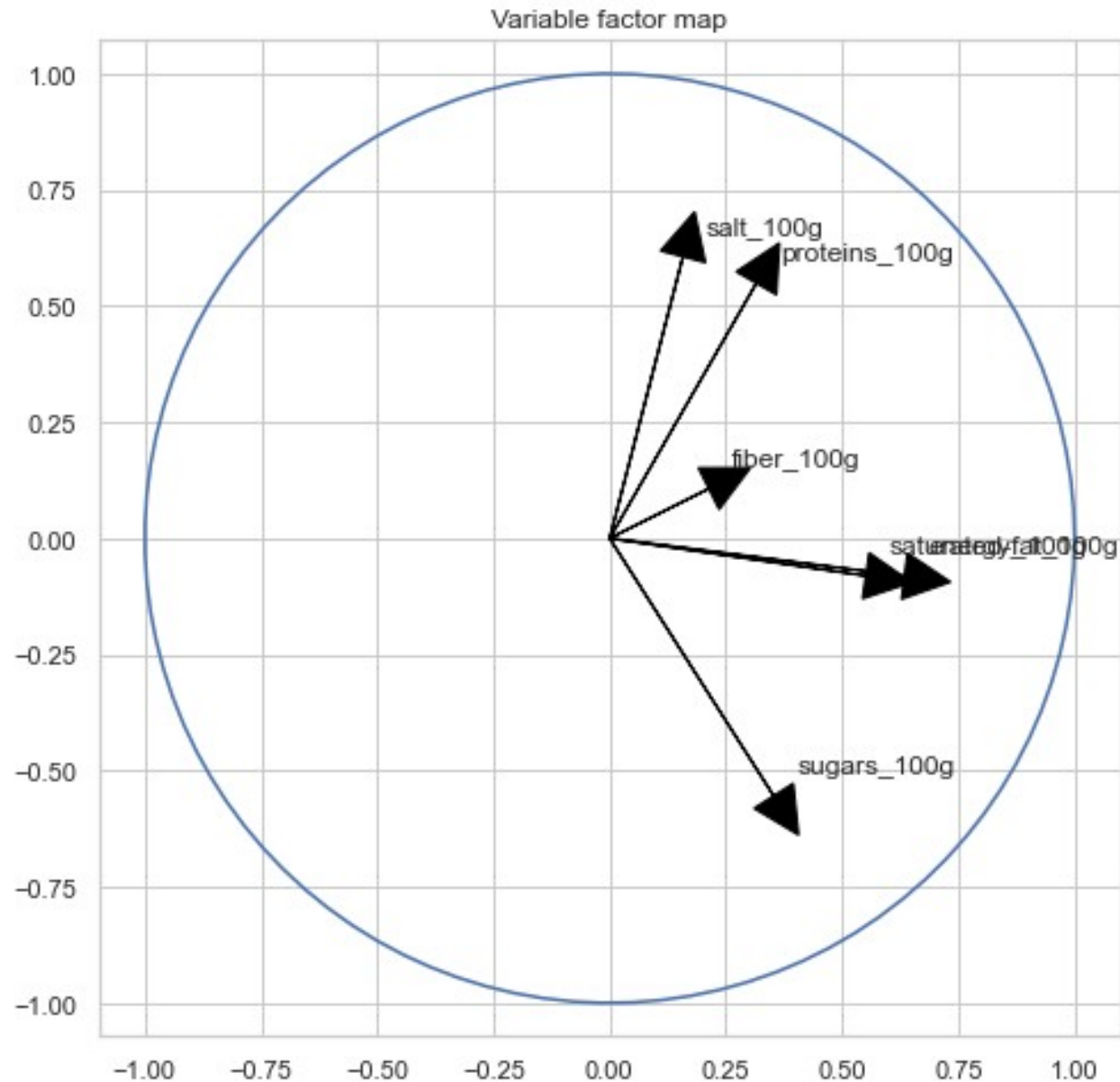
Résultat / III Analyse ...

- ACP

NUTRI-SCORE

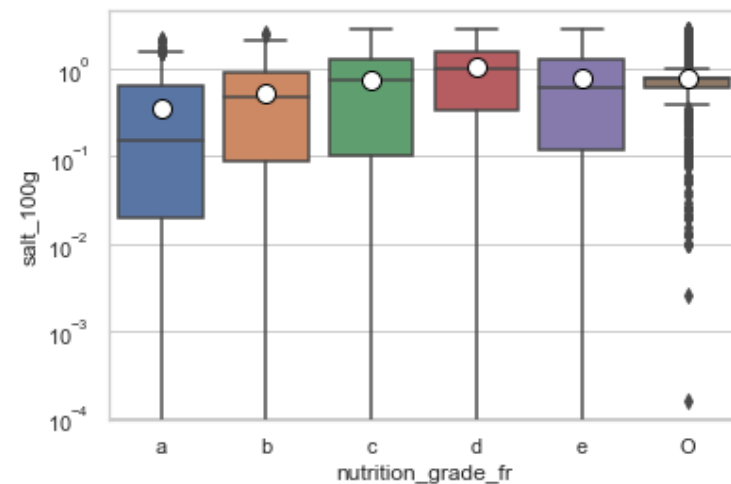
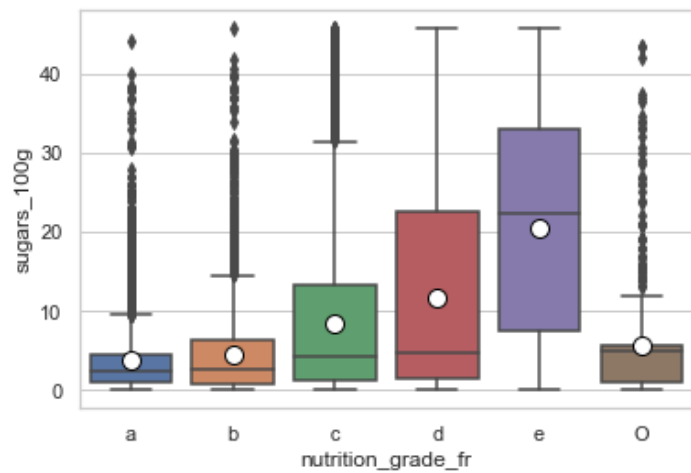
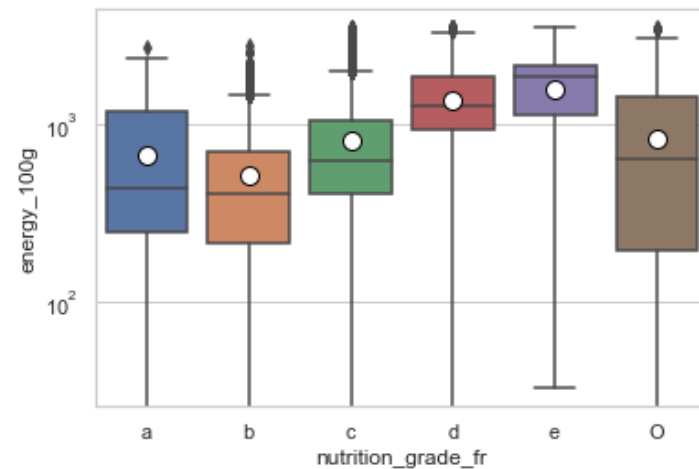
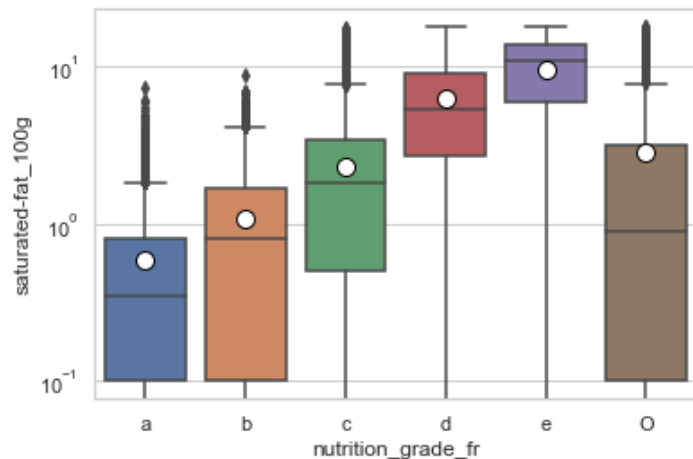


Résultat / III Analyse ...

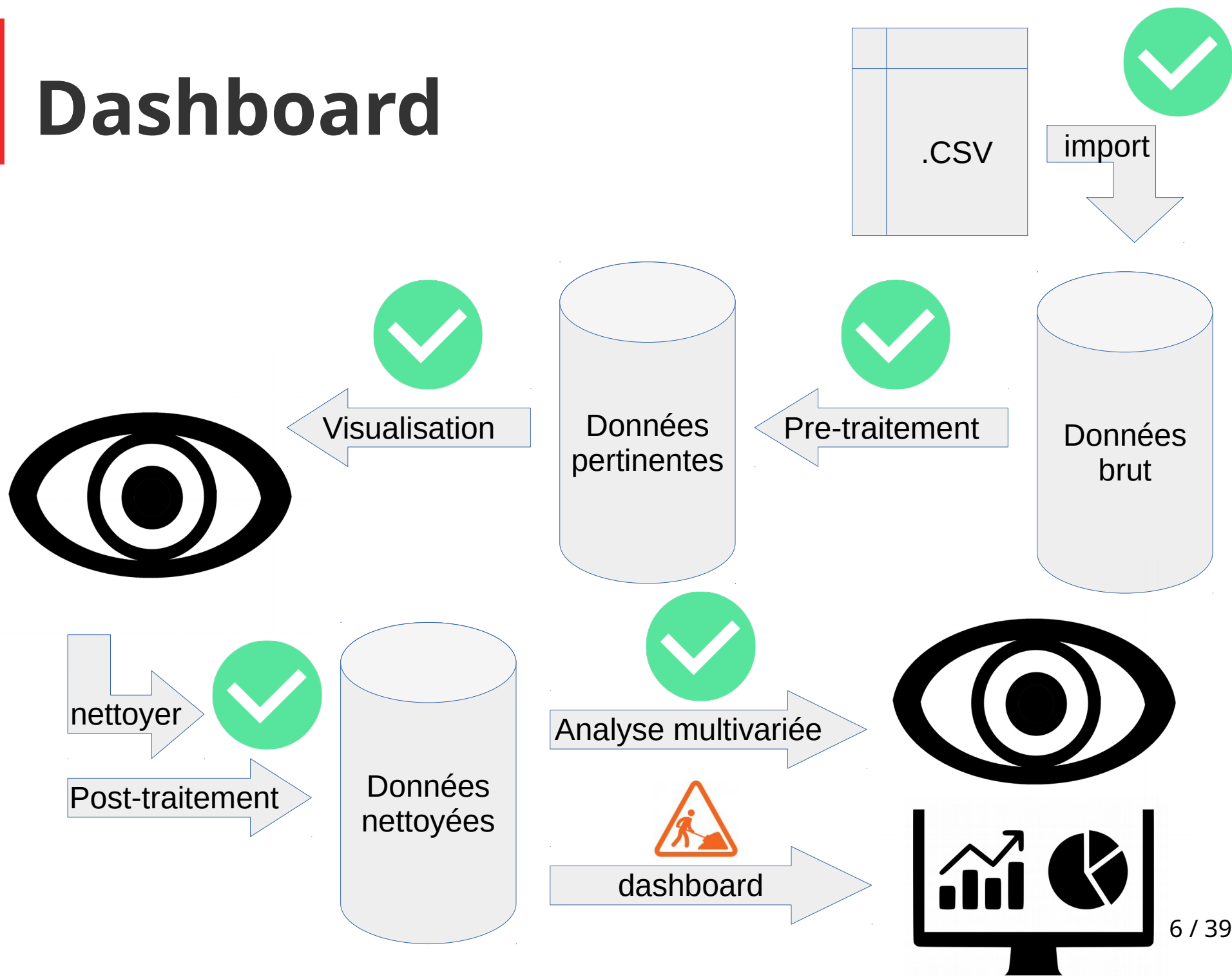


Résultat / III Analyse ...

- ANOVA



Dashboard






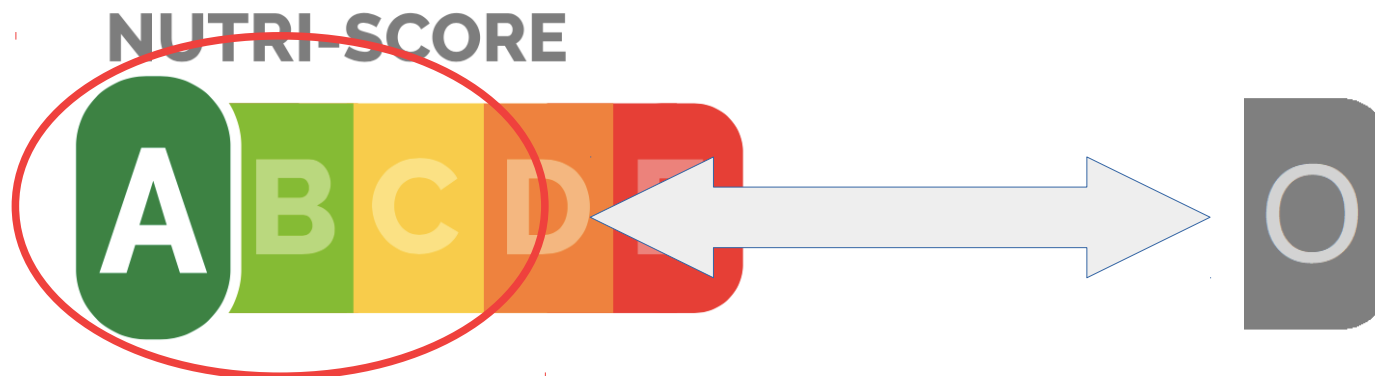
Dashboard

voilà

Conclusion



- Le nutri-score de nos aliments est-il fiable ?
Différences notables (ACP, ANOVA)
Variances  (ANOVA)
- Qu'en est-il en cas d'absence de ce logo ?
(ACP, ANOVA, dashboard)



Perspective

