ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Graph based nutrition guide

GROUP MEMBERS

AHMED SAADALLAH

FIRAS BEN OTHMAN

HANI BOUDABOUS

ASSISTANT

EFFROSYNI SIMOU

Introduction

- **Common food categorization:**
  - Dairy milk, butter, yogurt, cheese, cream and ice cream
  - Fruits
  - Grains beans and legumes
  - Vegetables
  - Sugary foods

But how could they be classified by their nutritional information?

Dataset

- A dataset of 2642 generic foods provided by the French government website.
- Contains 65 features including Macronutrients (Protein, Carbs, Fat) , Micro-nutrients (Vitamins and Minerals.) , Food name, Category and Category id.

	ORIGFDNM	ORIGGPFR	Protéines (g/100g)	Glucides (g/100g)	Lipides (g/100g)	Fibres alimentaires (g/100g)	Sucres (g/100g)	AG saturés (g/100g)	Cholestérol (mg/100g)
1030	Turbot sauvage, cru	Poissons et batraciens crus	16.6	0.0	2.43	0.0	0.0	0.720	48.0
1031	Cabillaud, cru	Poissons et batraciens crus	18.1	0.0	0.56	0.0	0.0	0.100	43.6
1032	Merlu, cru	Poissons et batraciens crus	17.4	0.0	1.00	0.0	0.0	0.200	35.0

Data Preprocessing

- **Clean dataset :**
  - Non-numerical data has values “-”, <(some small number) or “traces”.
  - From looking at the dataset, these values should mean inexistent nutrient or negligible amount.
    ➔ Set these values to 0
  - Some sub-categories have composed foods (e.g soup with vegetables and chicken or rice vegetables and some meat ready to be microwaved…).
    ➔ Delete These elements.

Data Preprocessing

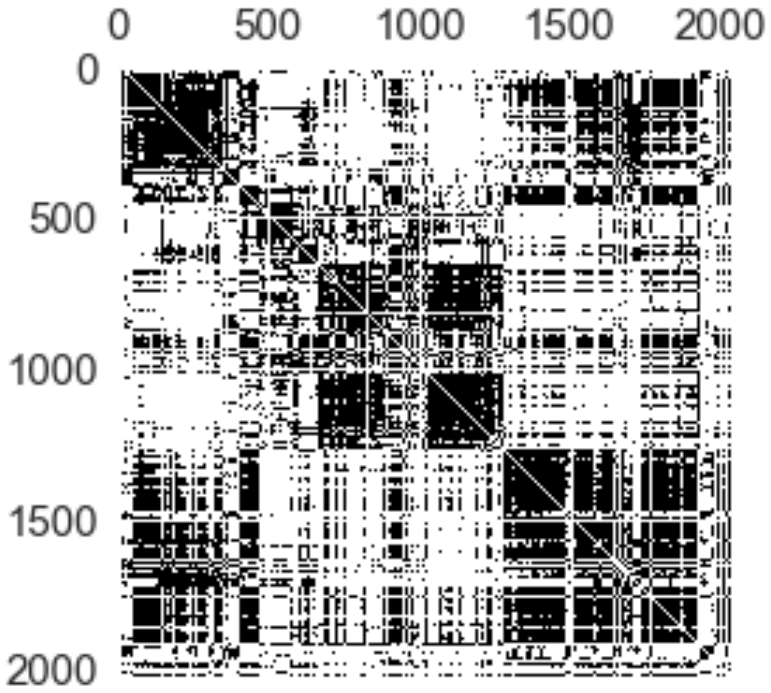
- **Edit categories**
  - The categories assigned in the dataset have some bad combinations of sub-categories.

    ➔Re-arrange sub-categories into new categories.

- **Standardize data**

Graph Construction

- **Graph:**
  - Nodes: Food items
  - Edges: similarity based on nutrients
- **Similarity measure:**
  - Cosine distance $s(u,v) = 1 - \frac{<u,v>}{\|u\|\|v\|}$
  - RBF kernel $k(d) = e^{-\frac{d^2}{2\sigma^2}}$



Number of clusters

- Eigengap to select number of clusters
- Choose k that maximizes Δ_k

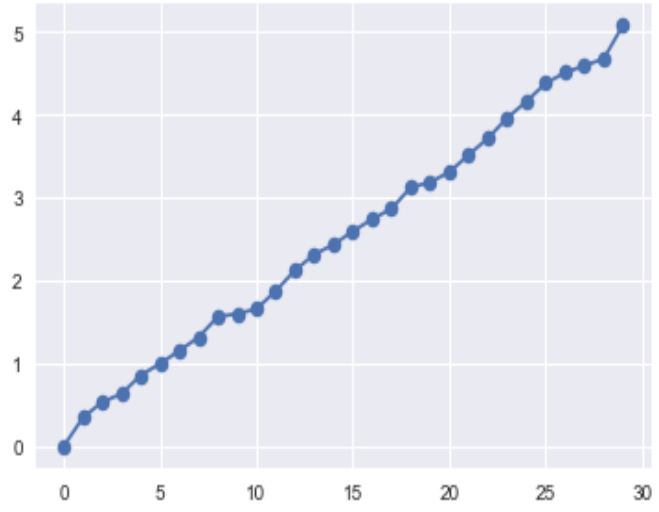
$$\Delta_k = \lambda_k - \lambda_{k-1}$$

- Heuristic and should be taken with a grain of salt
- Depends on the adjacency matrix

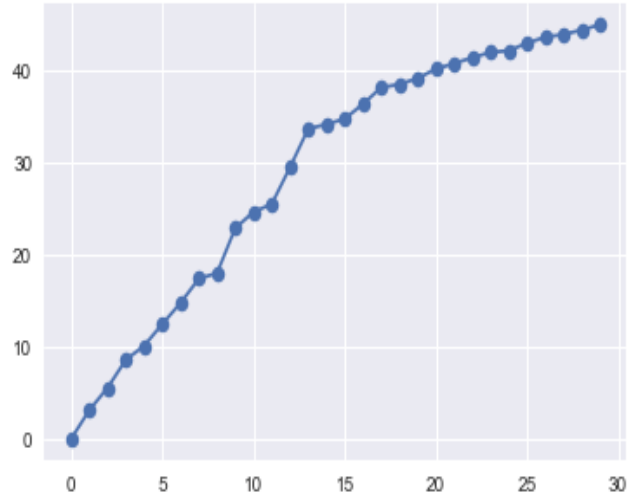
➔ choose number of neighbors that maximizes the eigengap

Selecting Neighbors

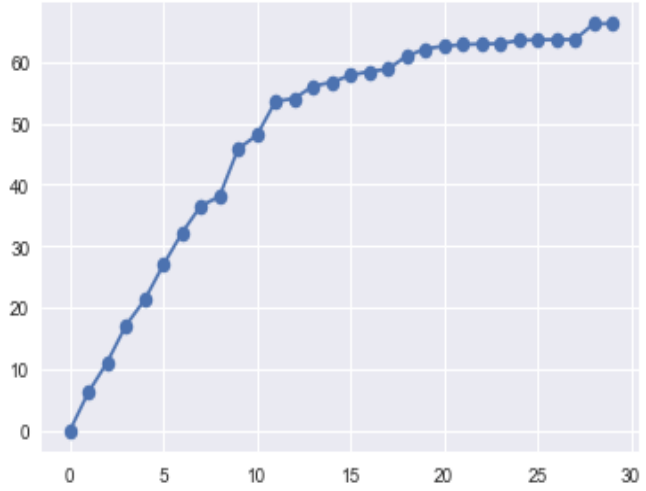
Number of neighbours : 20
jump : 0.404953889795, position : 29



Number of neighbours : 80
jump : 4.87420505834, position : 9



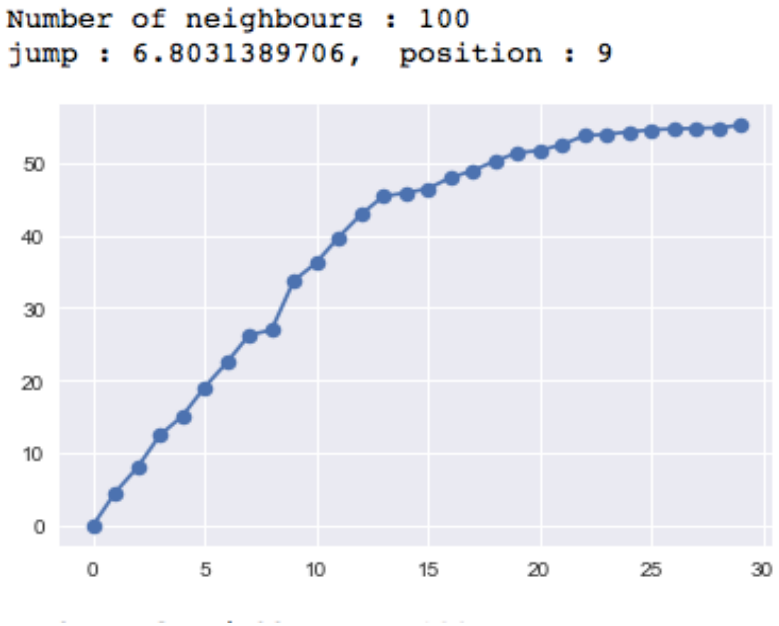
Number of neighbours : 120
jump : 7.75731491777, position : 9

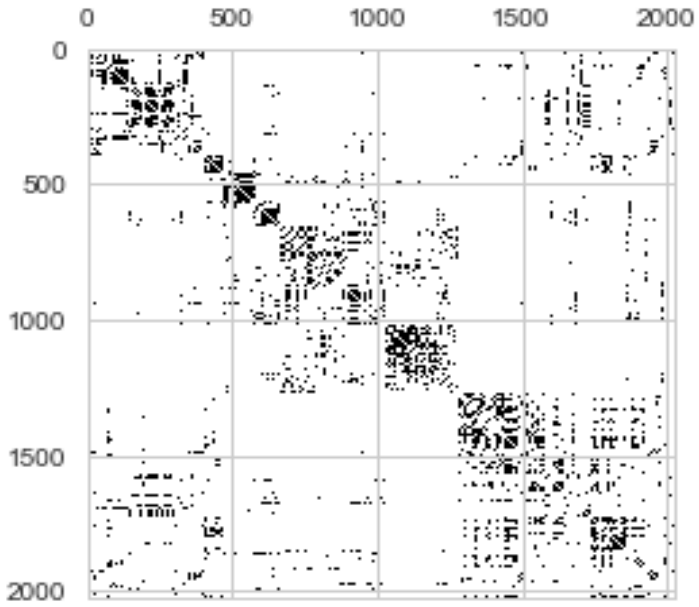


Perform a grid-search on the number of neighbors to find best jump of eigenvalue at the best position.

Spectral Clustering

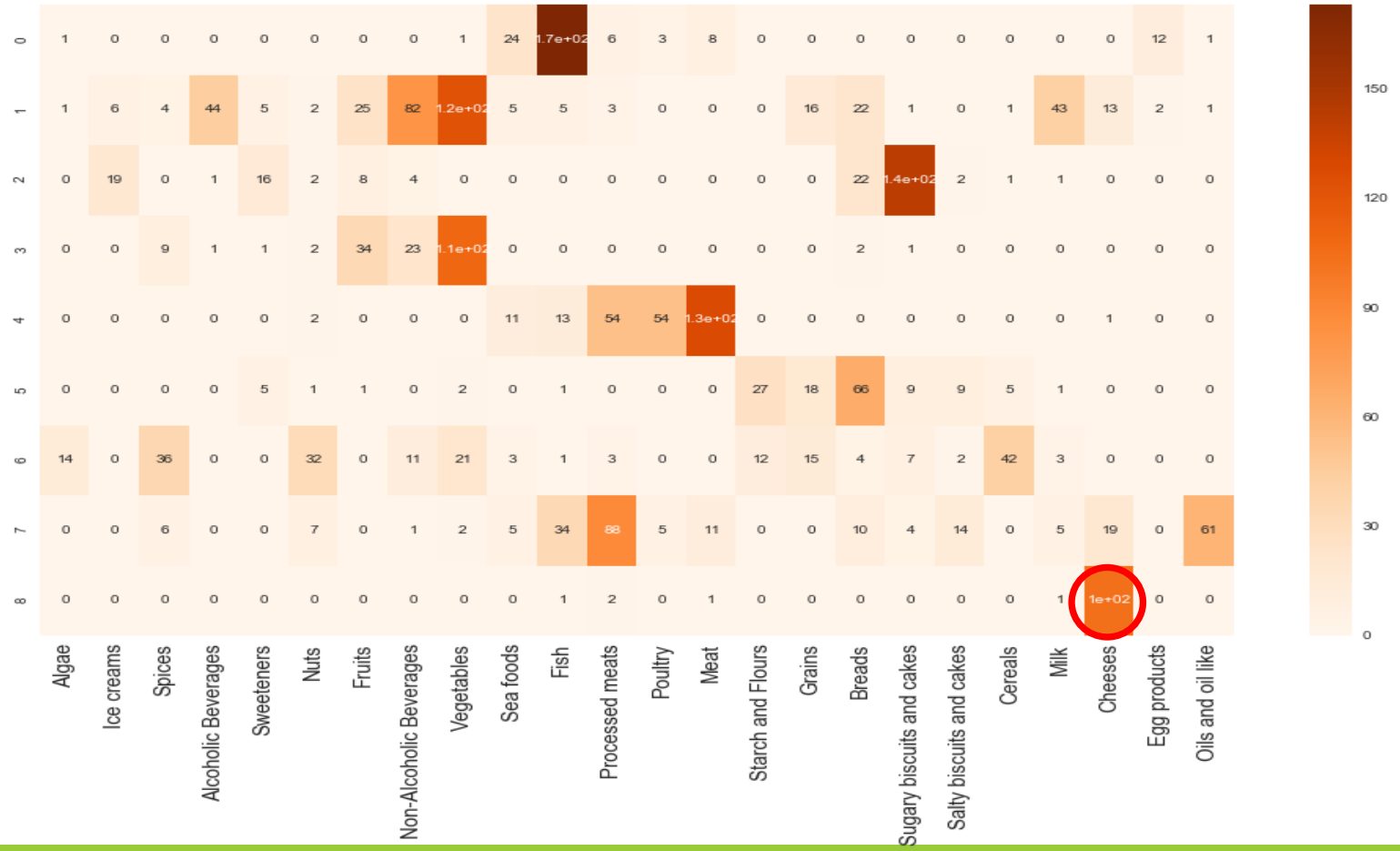
Based on the grid-search and graphs, we estimated that selecting 100 most similar neighbors per is the best choice.





➔ Perform Spectral Clustering **with 100 neighbors** and **9 clusters**

Occurrence of categories in clusters



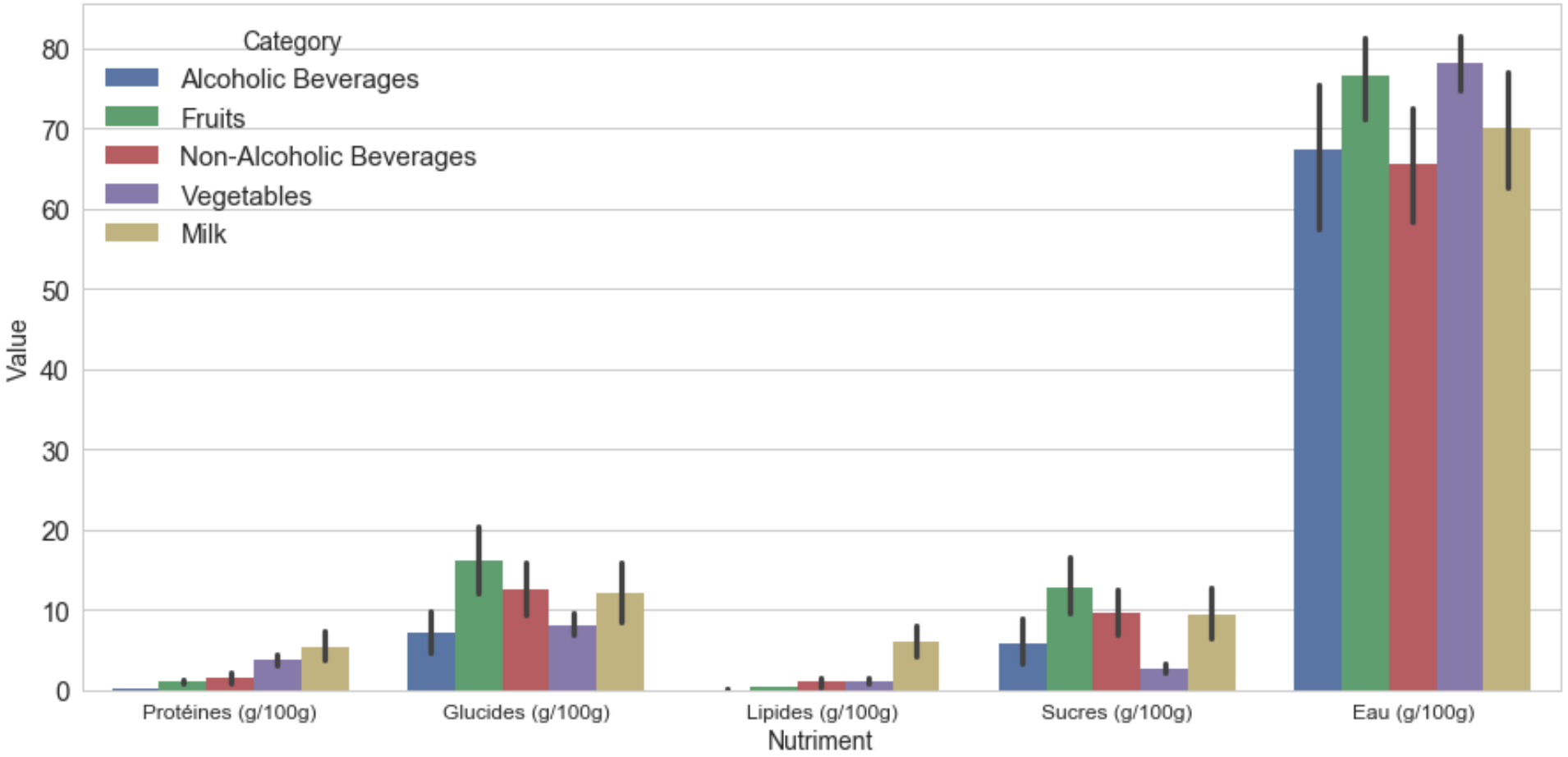
Occurrence of categories in clusters

Occurrence of categories in clusters

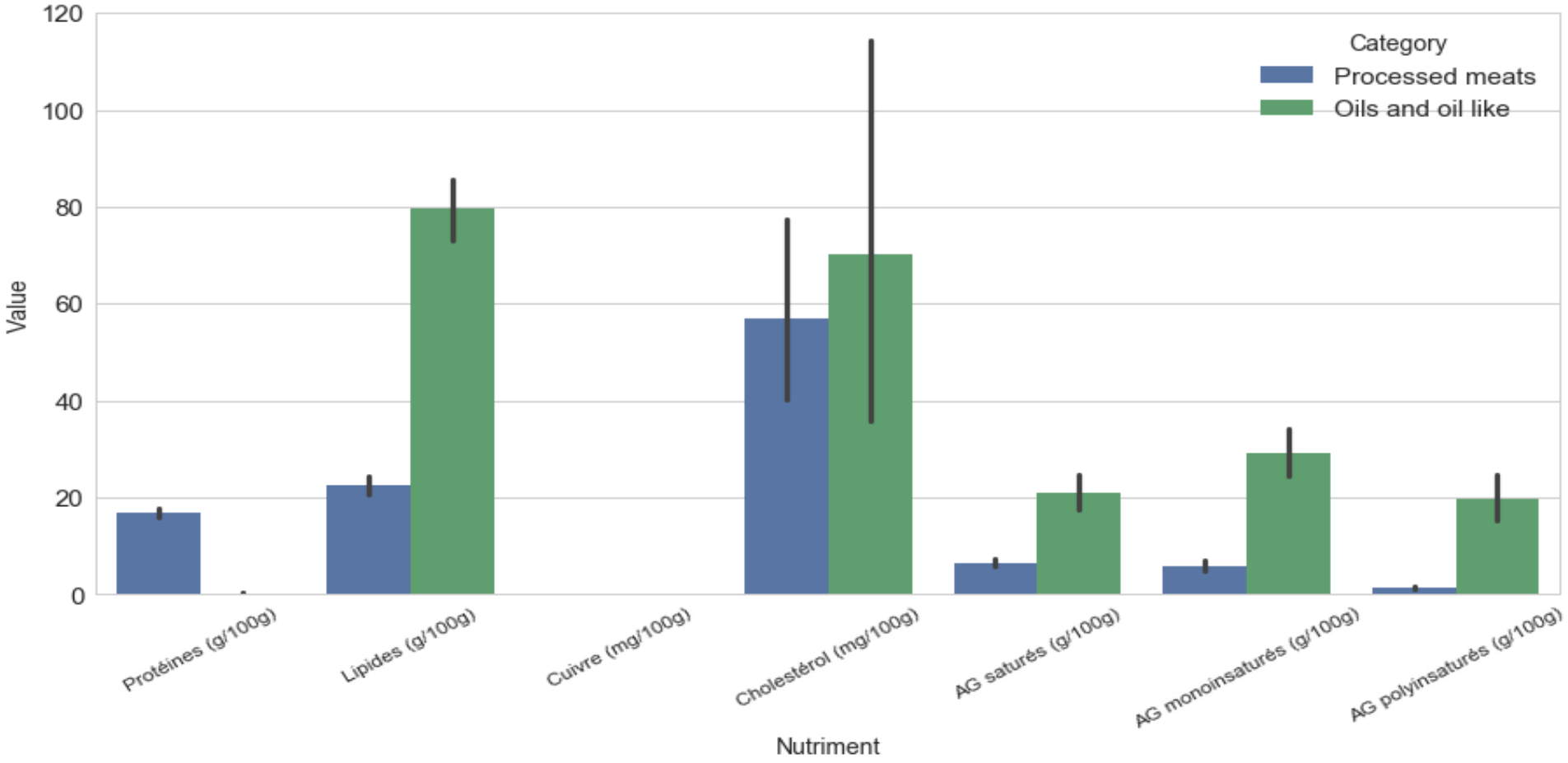
Occurrence of categories in clusters

Occurrence of categories in clusters

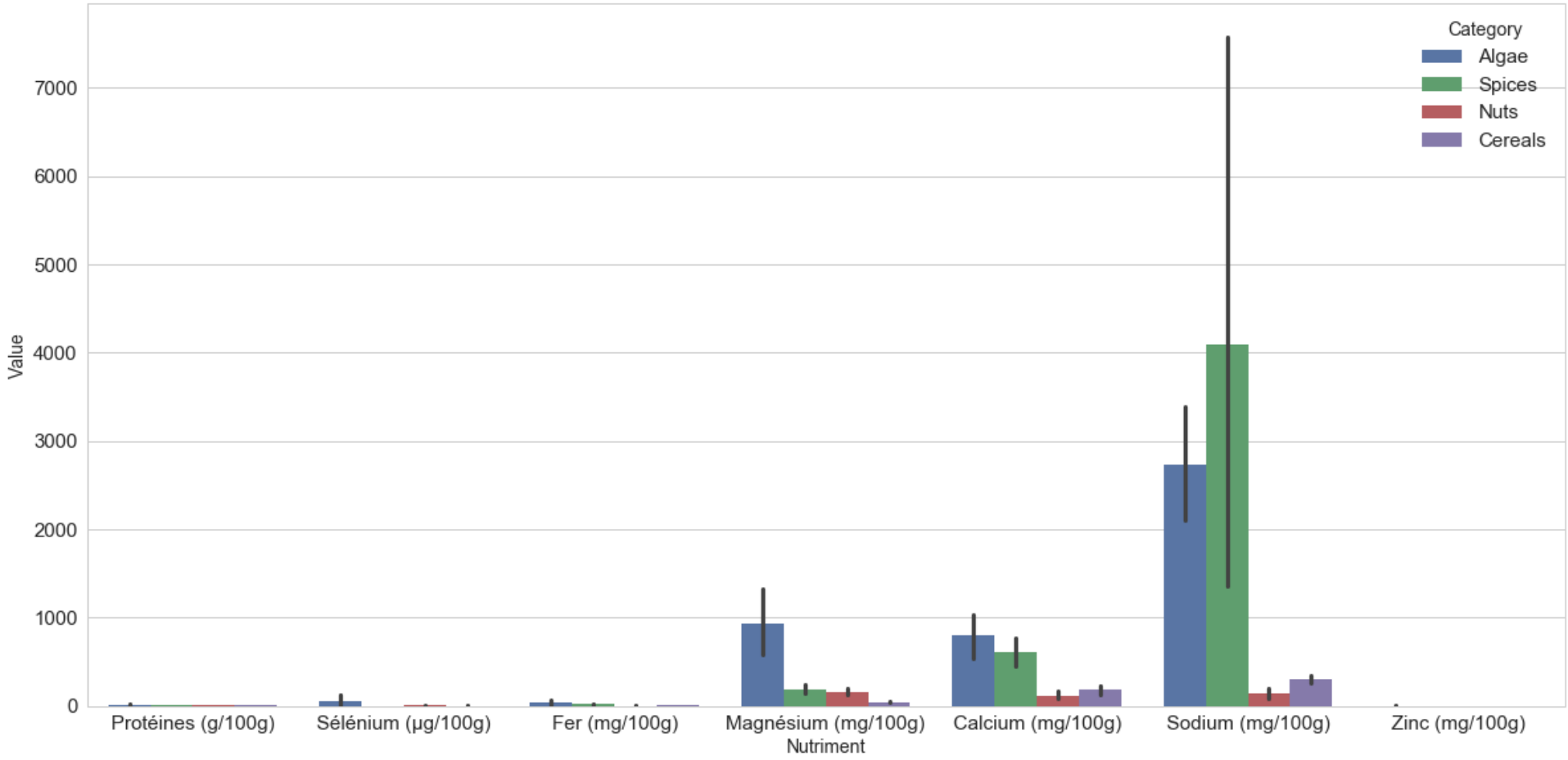
Analyzing clusters :



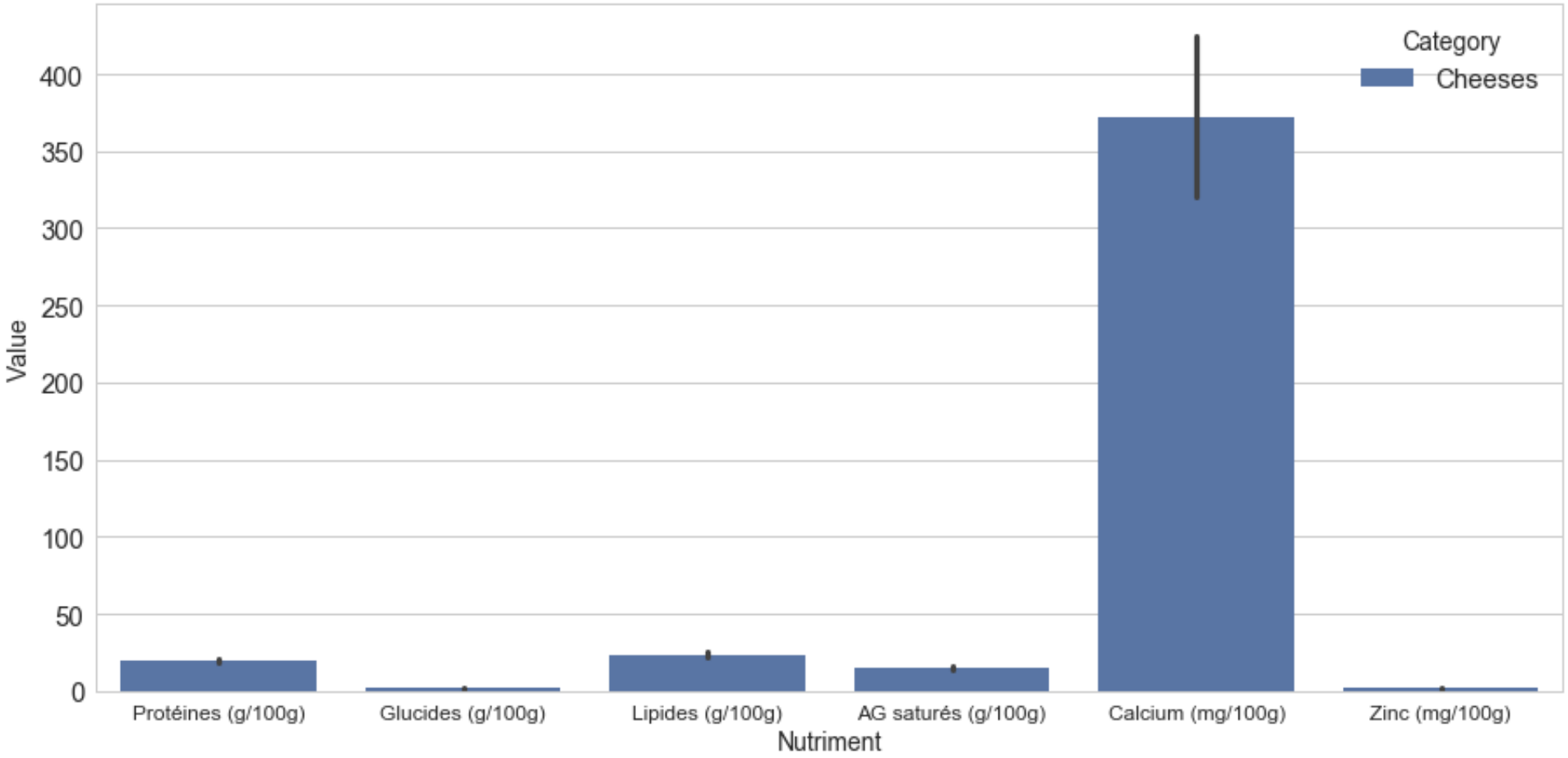
Analyzing clusters :



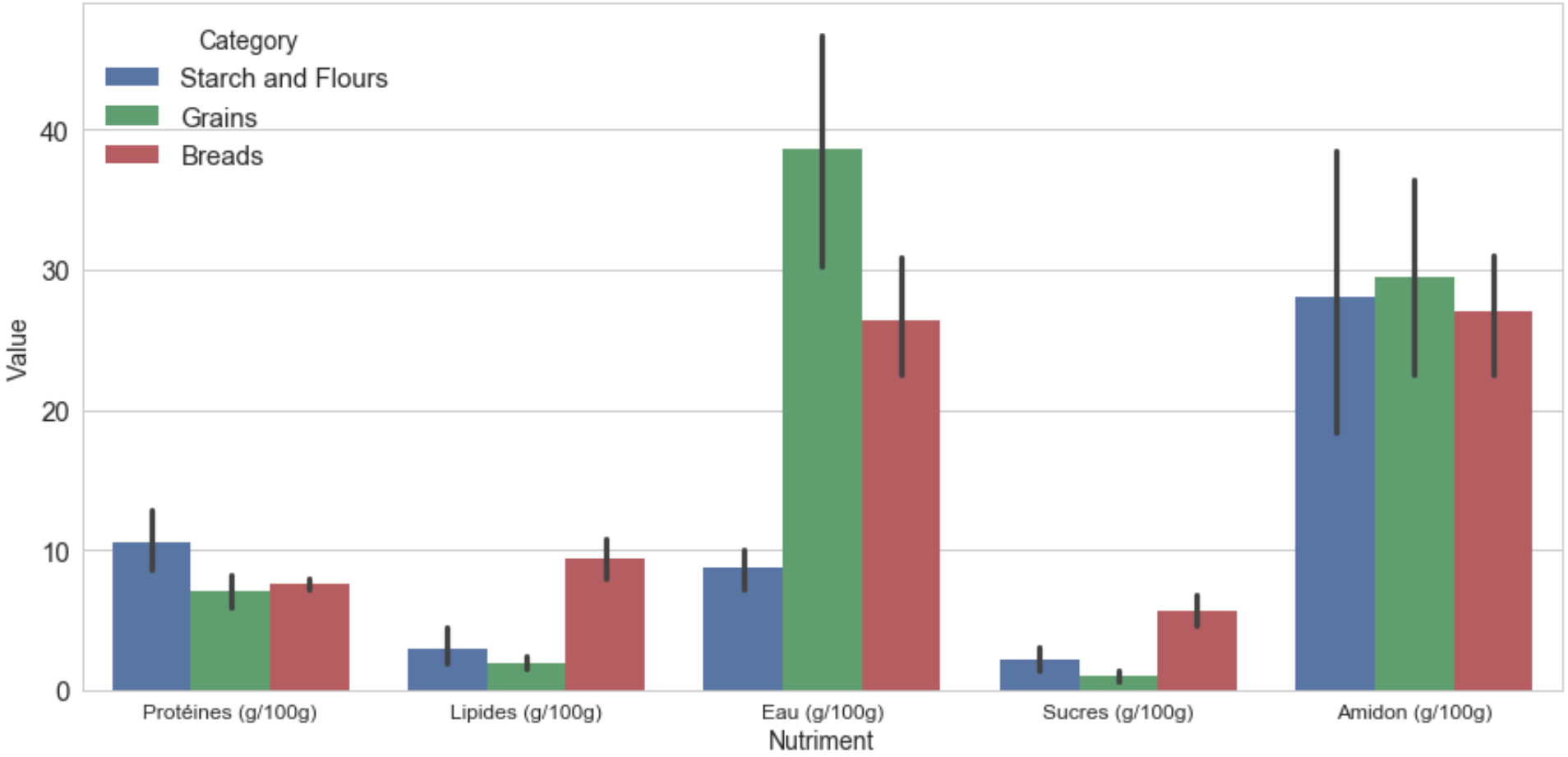
Analyzing clusters :



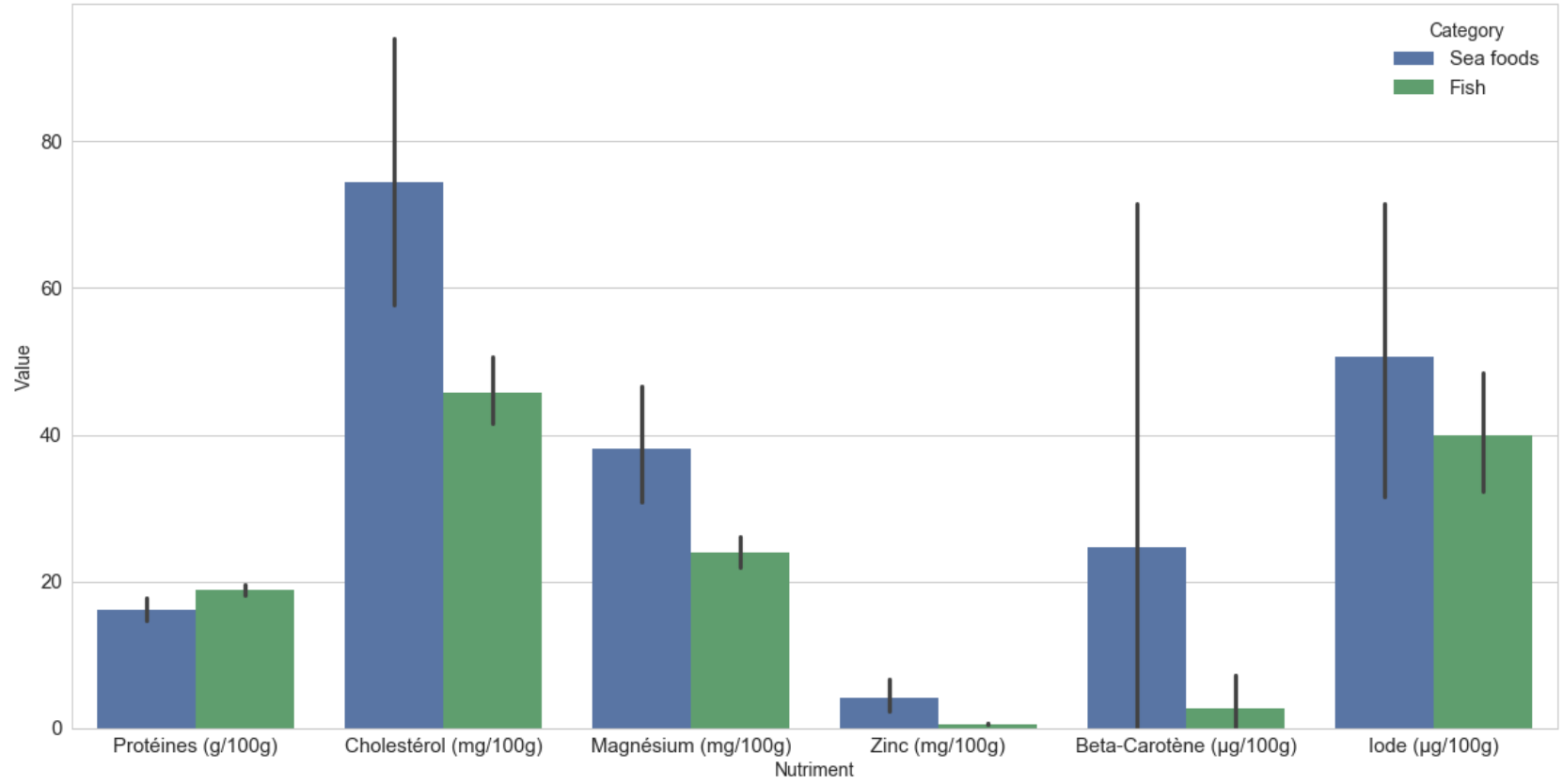
Analyzing clusters :



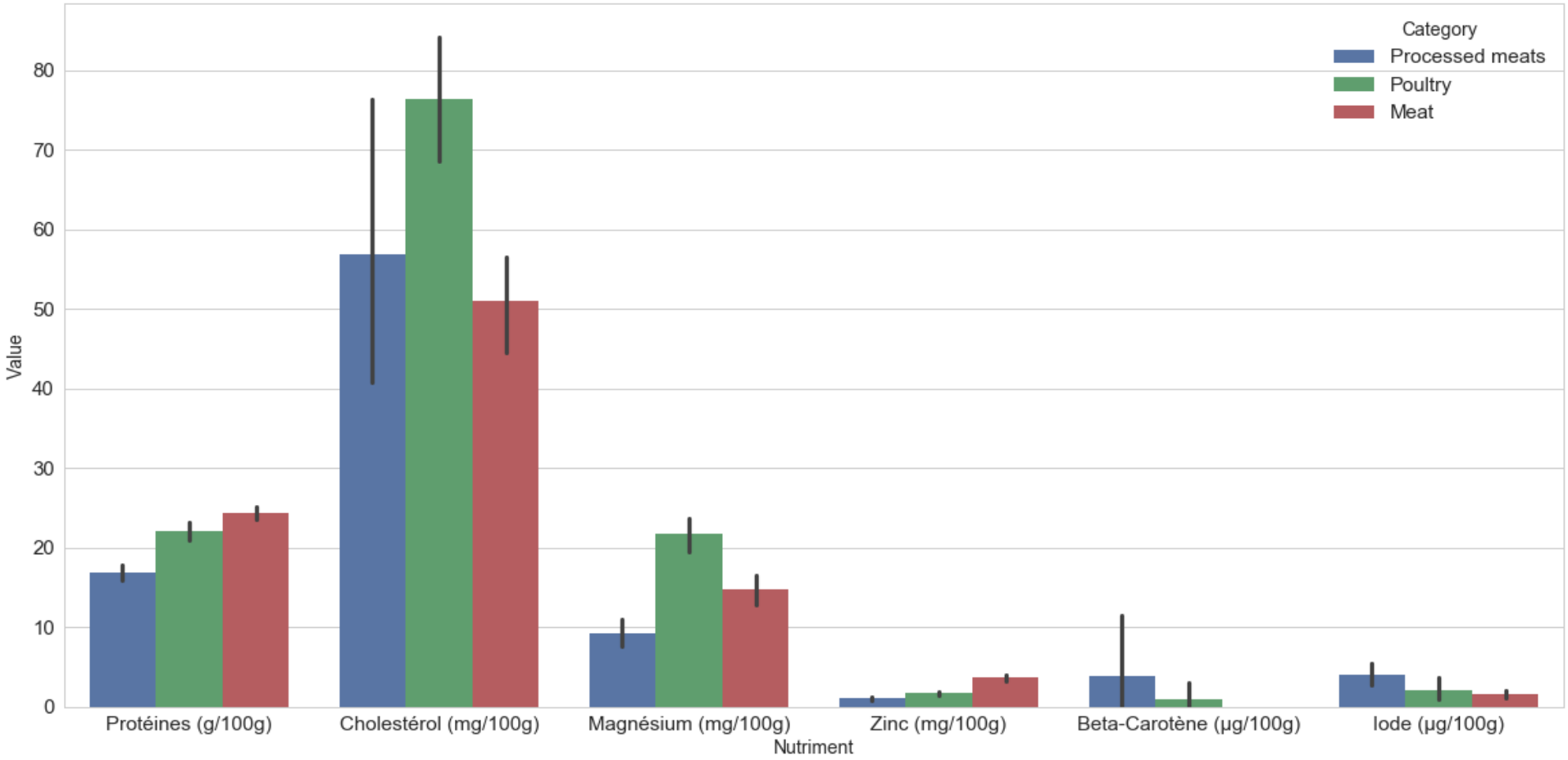
Analyzing clusters :



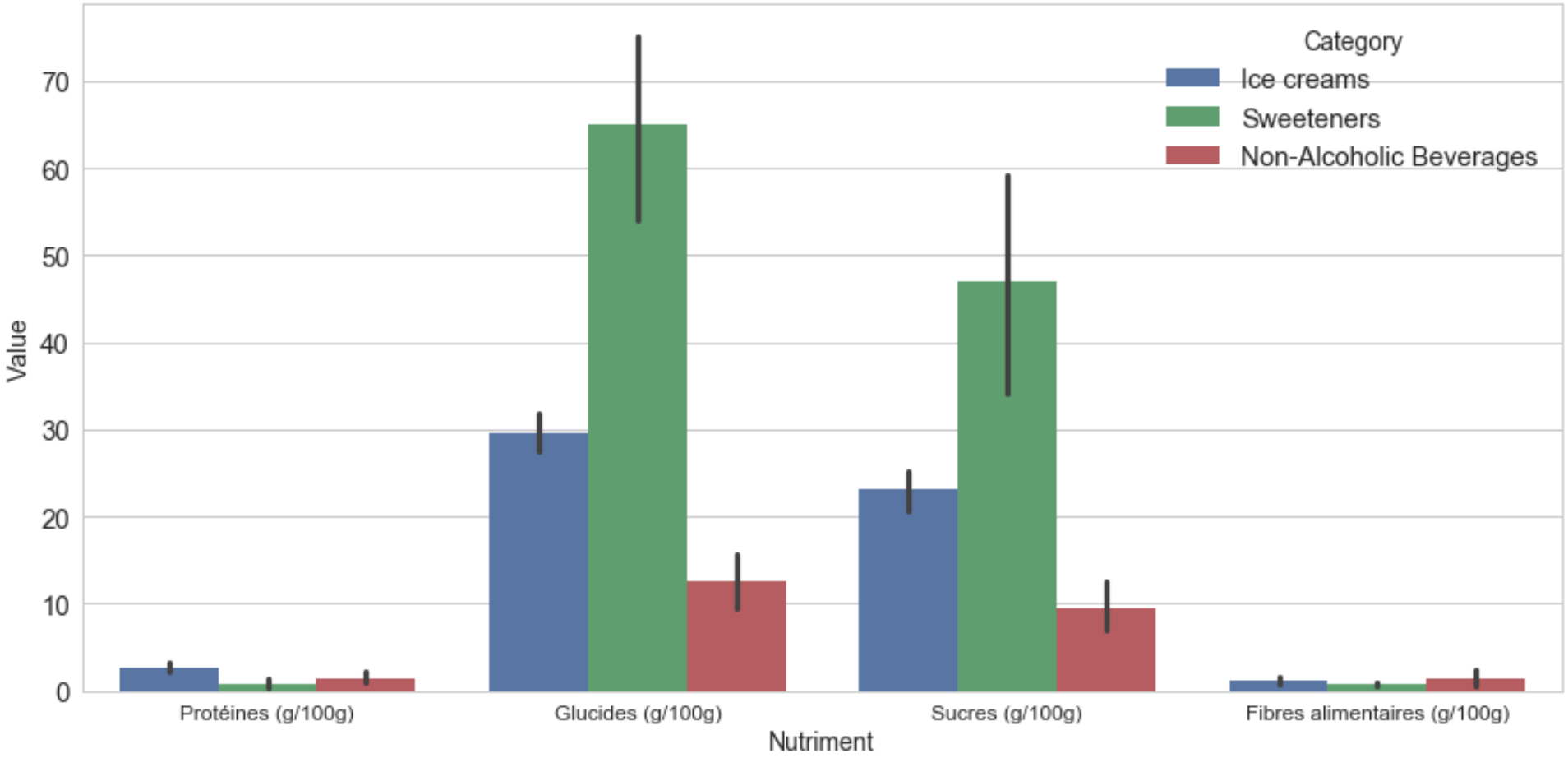
Analyzing clusters :



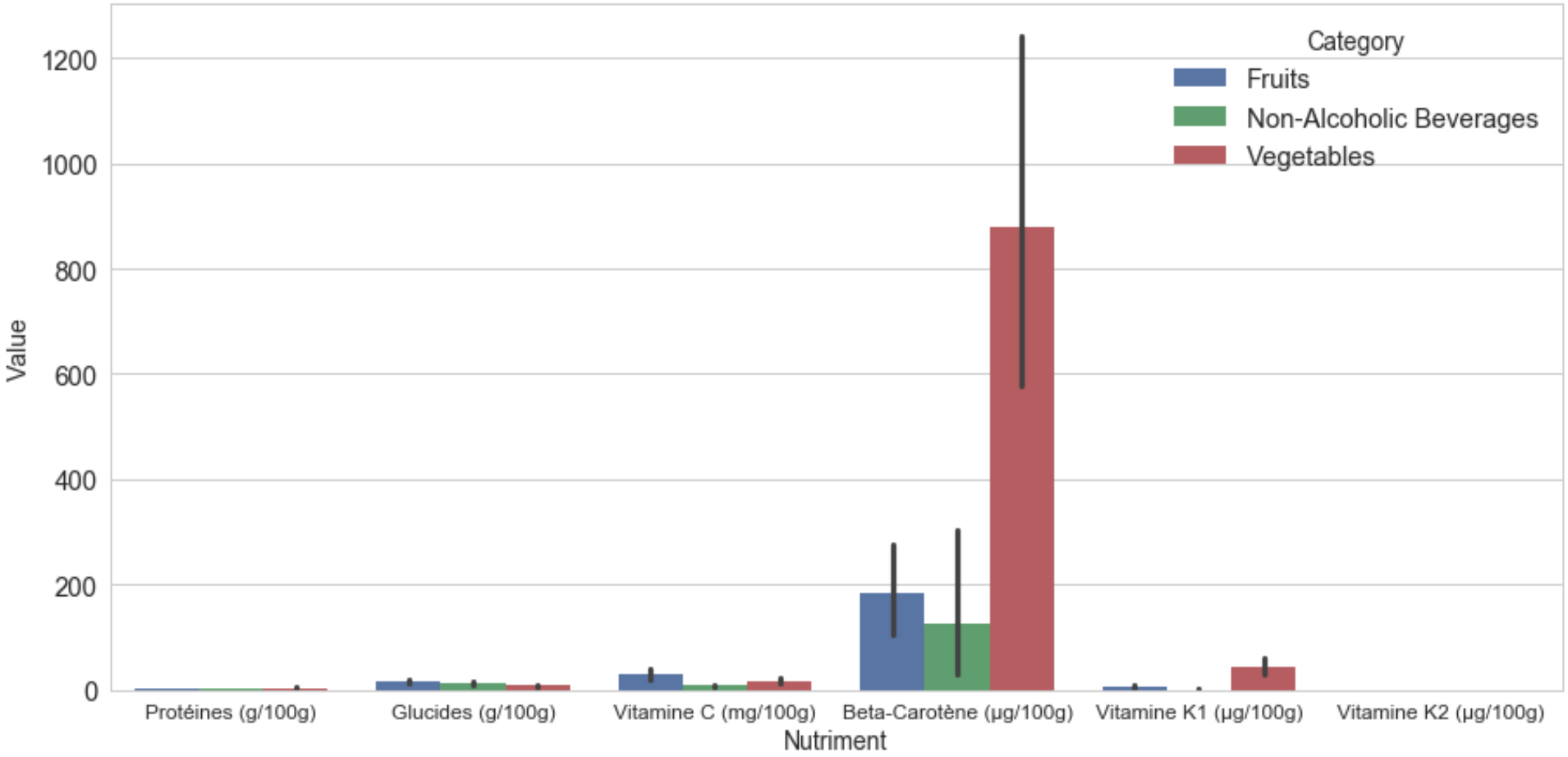
Analyzing clusters :



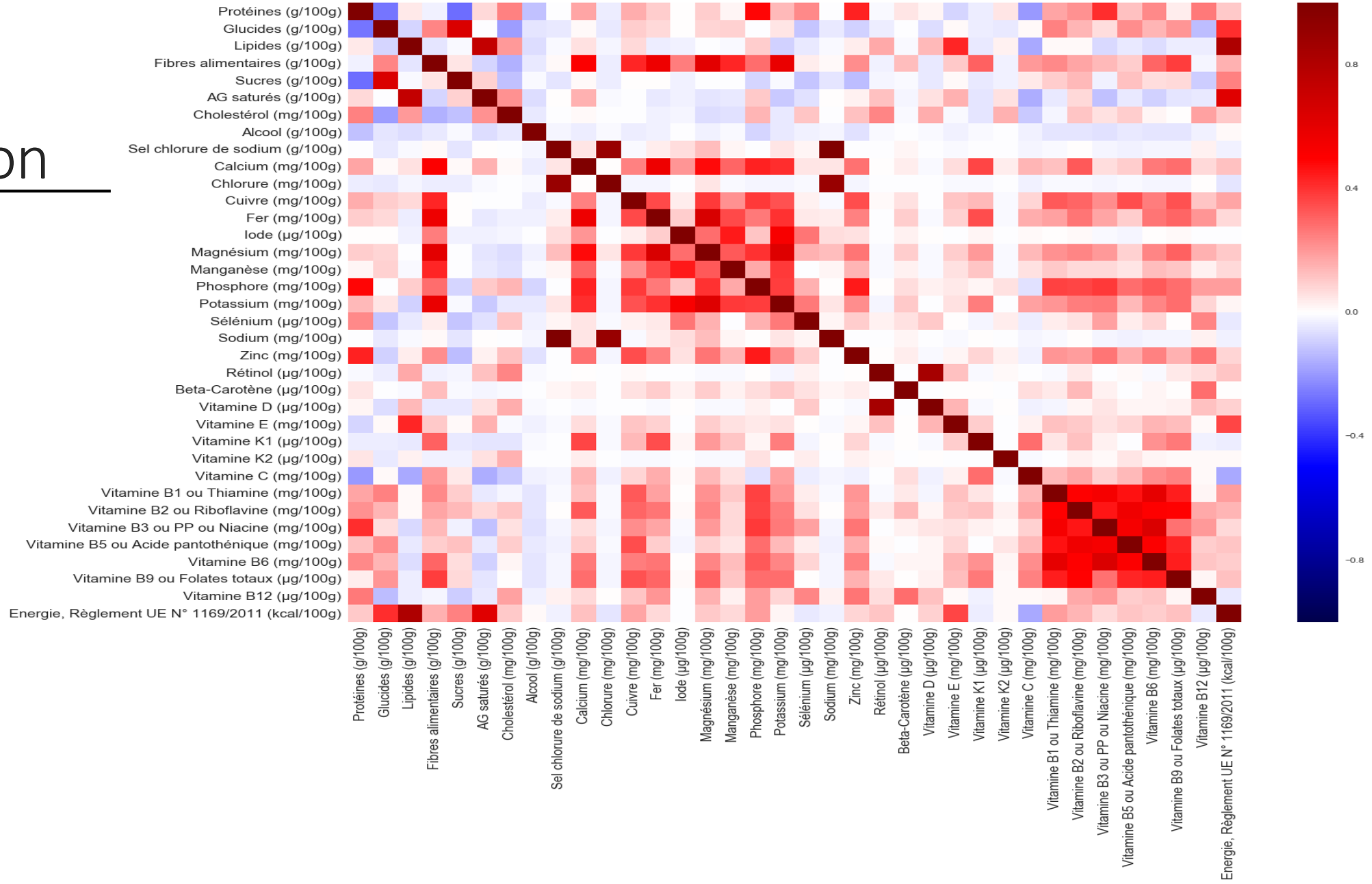
Analyzing clusters :



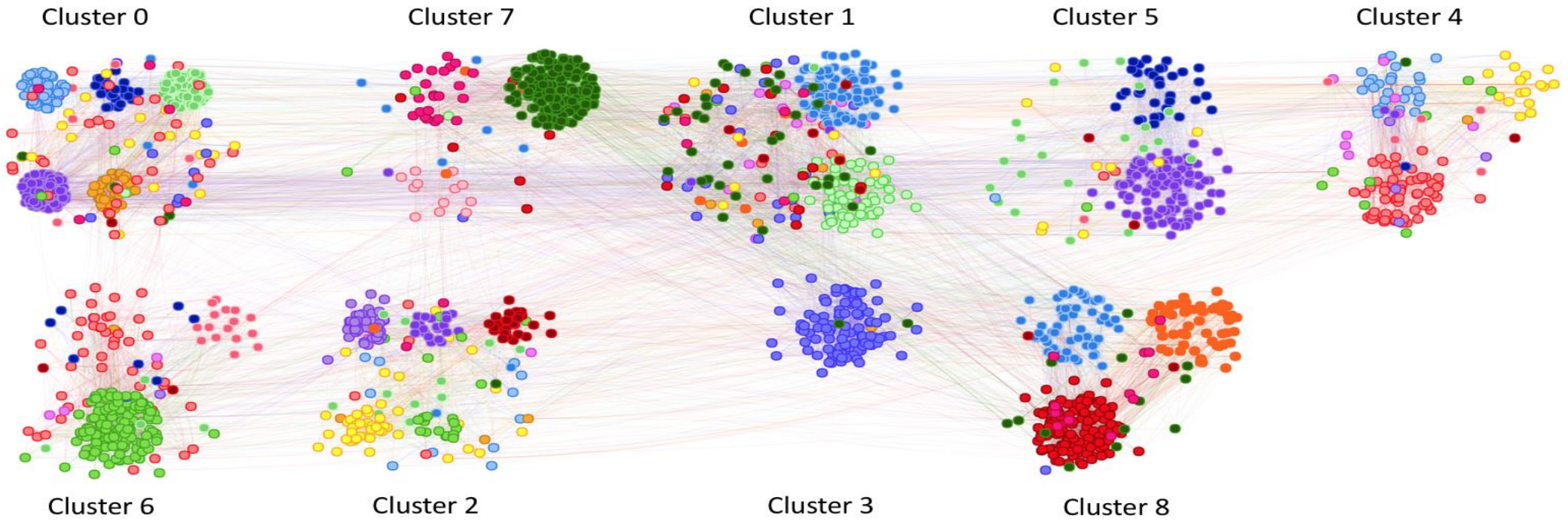
Analyzing clusters :



Features correlation



Visualizing clusters



Live Demo :

https://www.boudabou.github.io/ntds_project

Thank you for your attention !