

# Mini-Project: USDA Food Composition

CS-GY 6323

Caroline Barker, Omar Hammami, Andrew Liang

5/11/2020

# Overview

- Background information
- Goals
- Questions
  - SET 1 Analysis
  - SET 2 Analysis
  - Extended Analysis
- Summary
- Implications

# Background Information

- Food composition dataset provided by USDA FoodData Central
- Contains
  - 1836 food item entries
  - Food group labels
  - Macronutrient, vitamin, and mineral content
  - Other nutrition facts

GOAL:

Identify misleading classifications  
groups of food

# Questions

1. How do we group related foods together?
2. How do these groups compare to grouping by composition?
3. How does this apply to nutrient subsets?

# Question 1:

## How do we group related foods together?

(Set 1)

# Question 1: How do we group related foods together?

- Preprocessing
  - Selective Grouping
    - reduce assigned food group labels at our discretion
  - Normalize data
- Algorithmic Grouping
  - Categorize foods based on how closely they relate in terms of composition data
  - K-means clustering algorithm

# Question 1: How do we group related foods together?

Category Reduction: 24 Groups -> 19 Groups

## Meats

- ‘Poultry Products’
- ‘Beef Products’
- ‘Pork Products’
- ‘Lamb, Veal, and Game Products’

## Produce

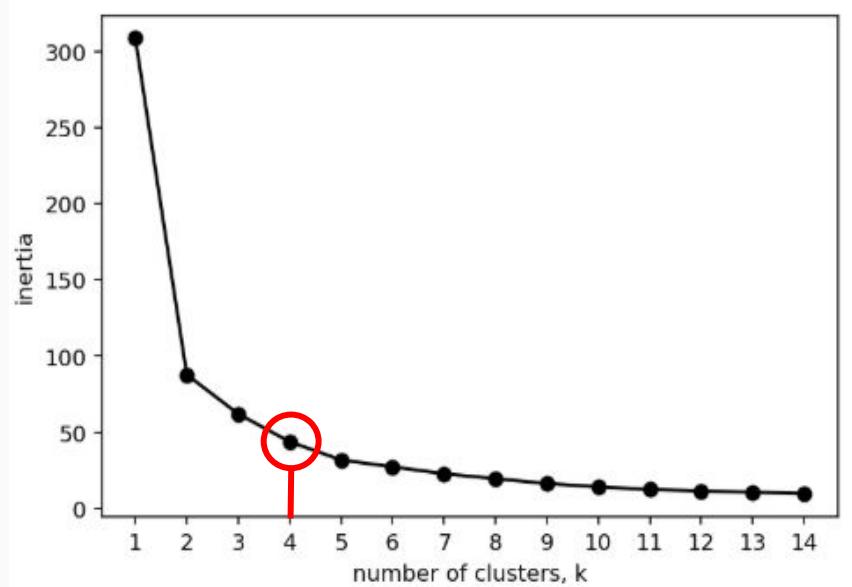
- ‘Fruits and Fruit Juices’
- ‘Vegetables and Vegetable Products’

## Meals

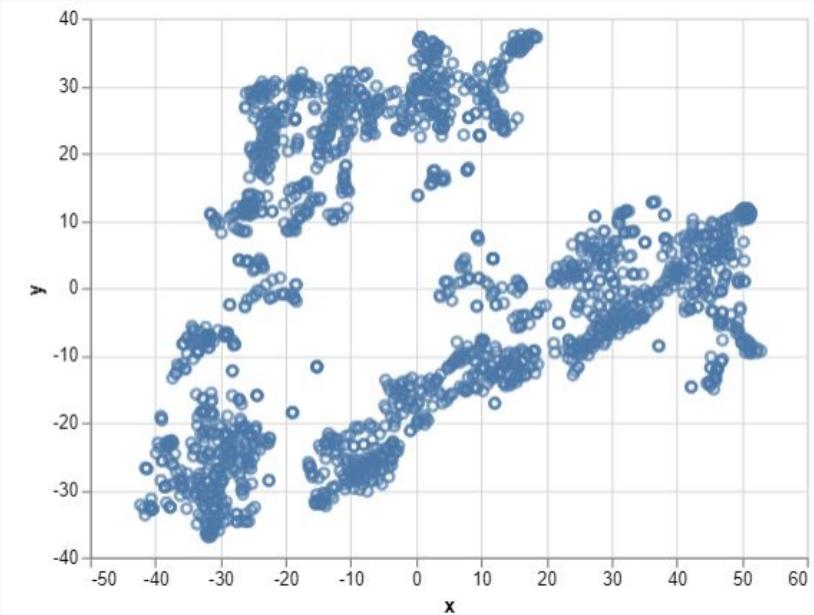
- ‘Meals, Entrees, and Side Dishes’
- ‘Restaurant Foods’

# Question 1: How do we group related foods together?

K-Means Clustering

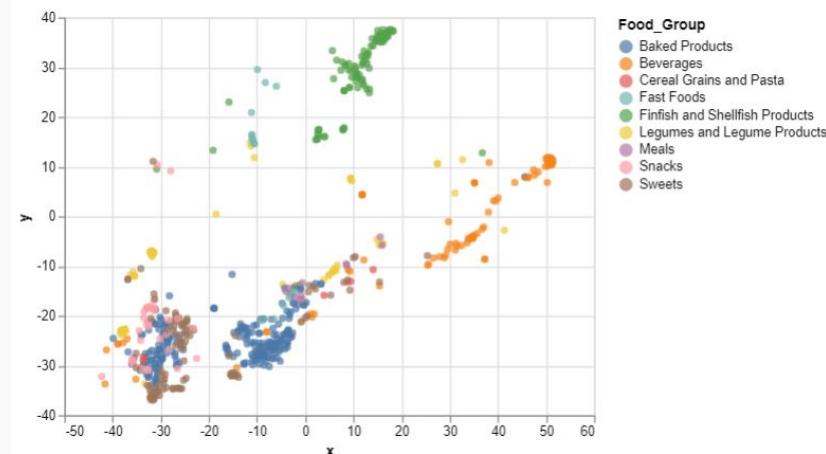
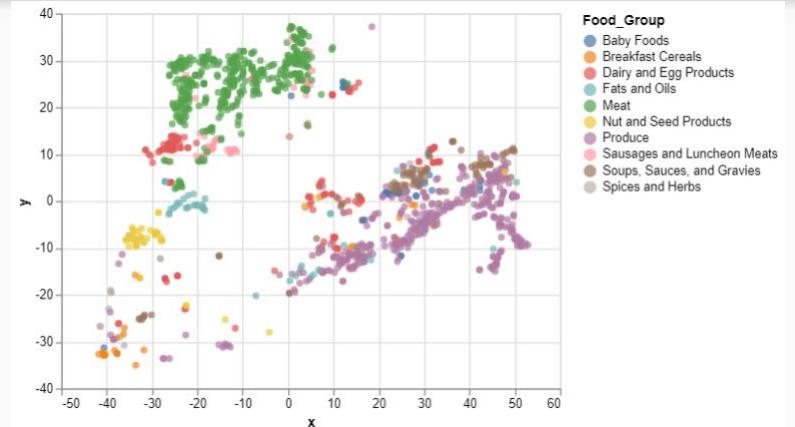
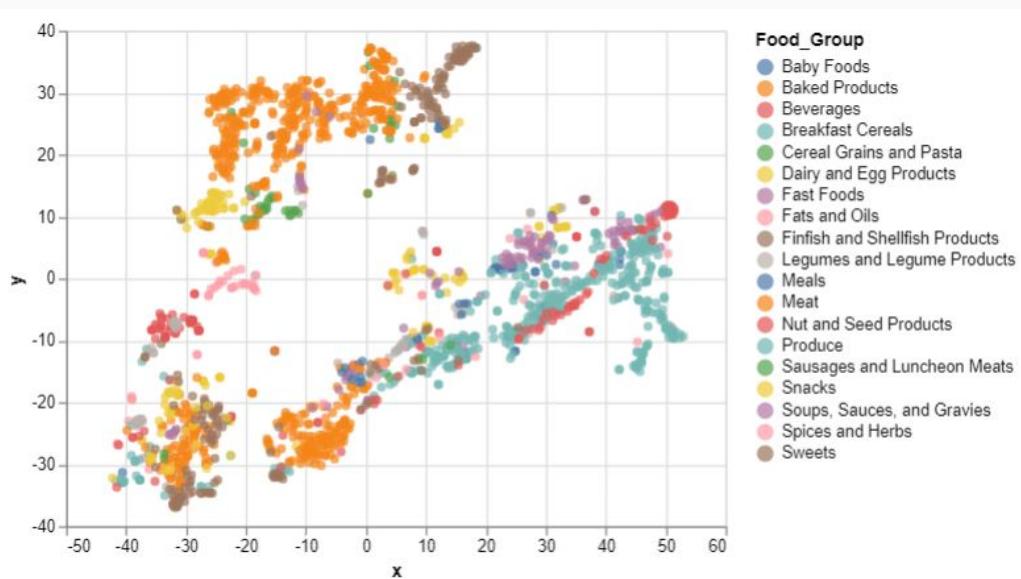


Dimension Reducing: T-SNE



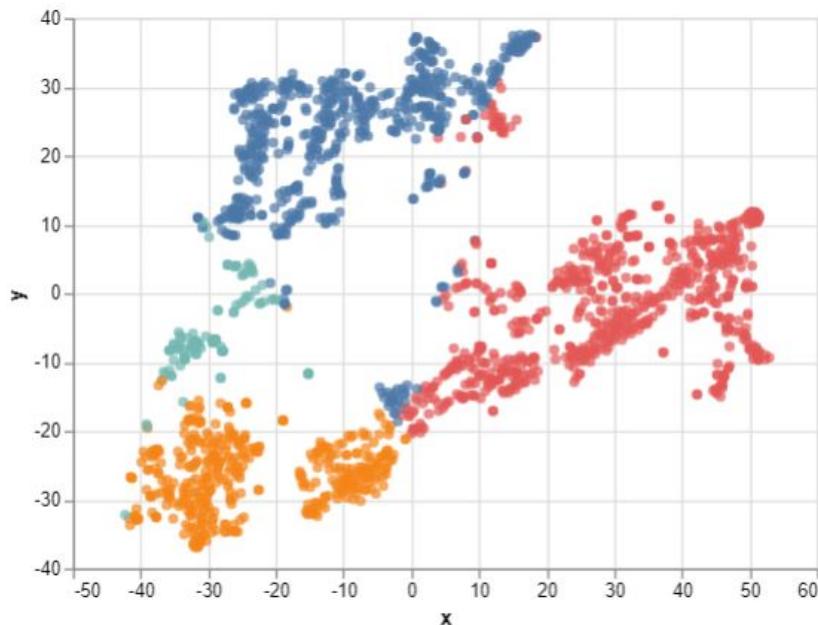
# Question 1: How do we group related foods together?

Colored by Food Group (Split because of Color limitations)

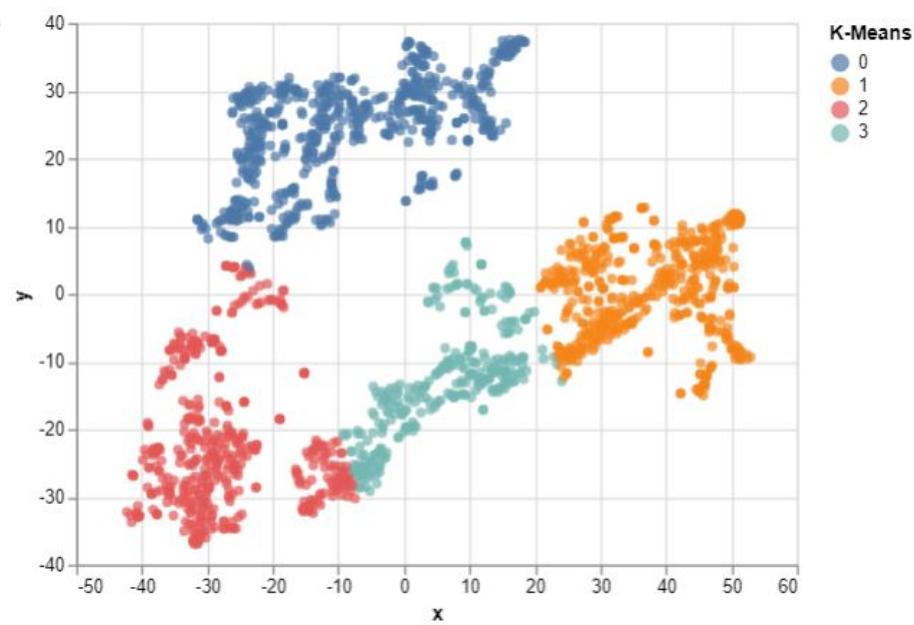


# Question 1: How do we group related foods together?

Colored with High Dimensionality  
K-Means Clustering

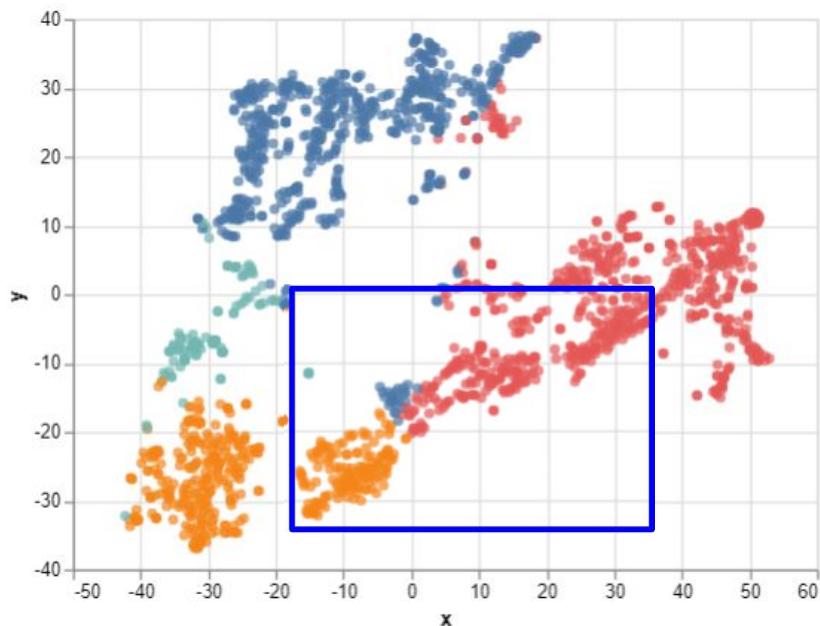


Colored with Low Dimensionality  
K-Means Clustering

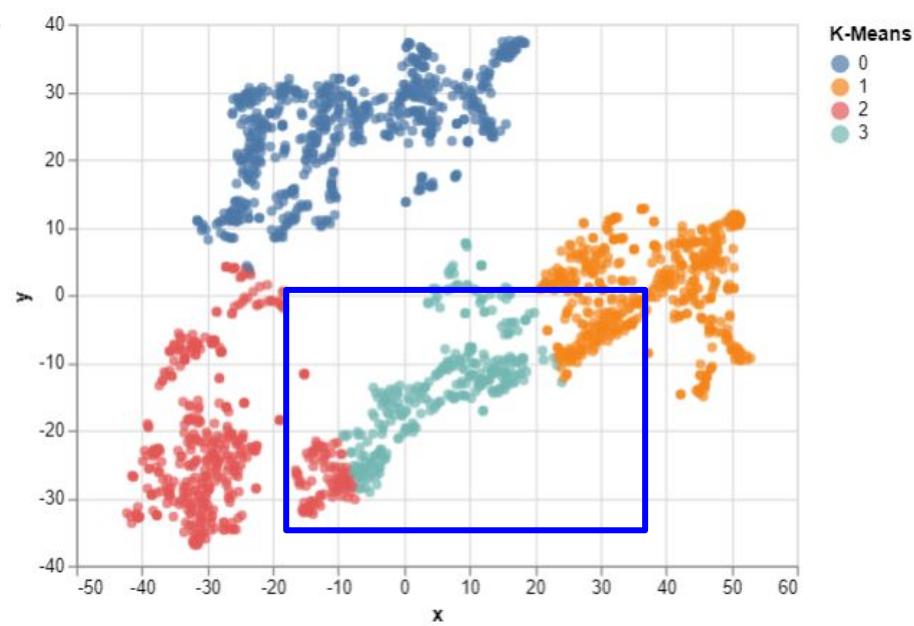


# Question 1: How do we group related foods together?

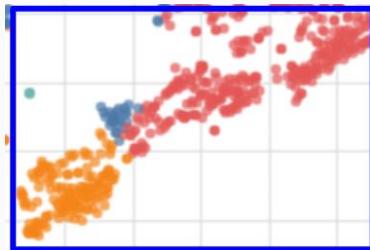
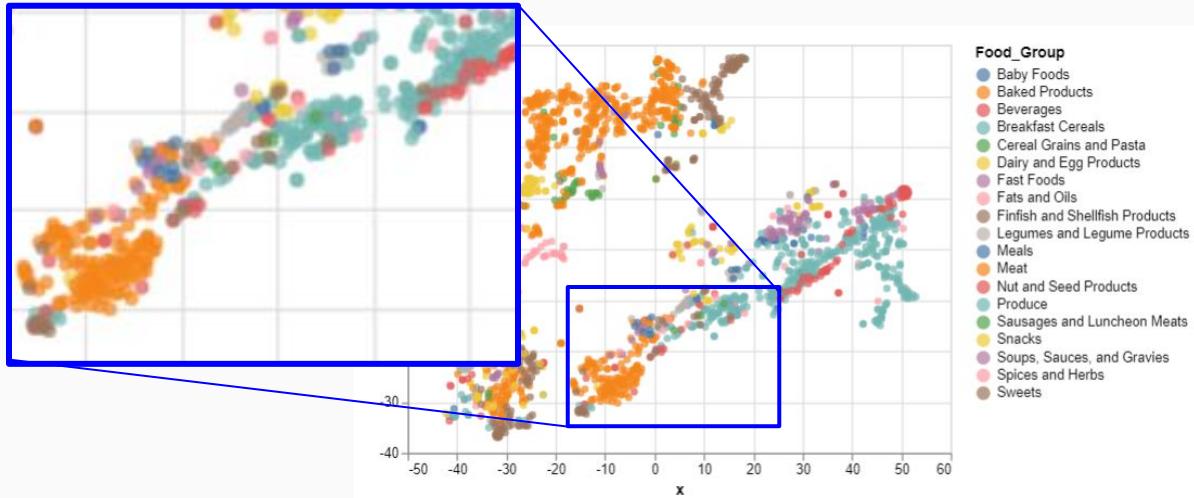
Colored with High Dimensionality  
K-Means Clustering



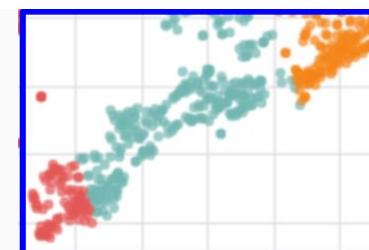
Colored with Low Dimensionality  
K-Means Clustering



# Question 1: How do we group related foods together?

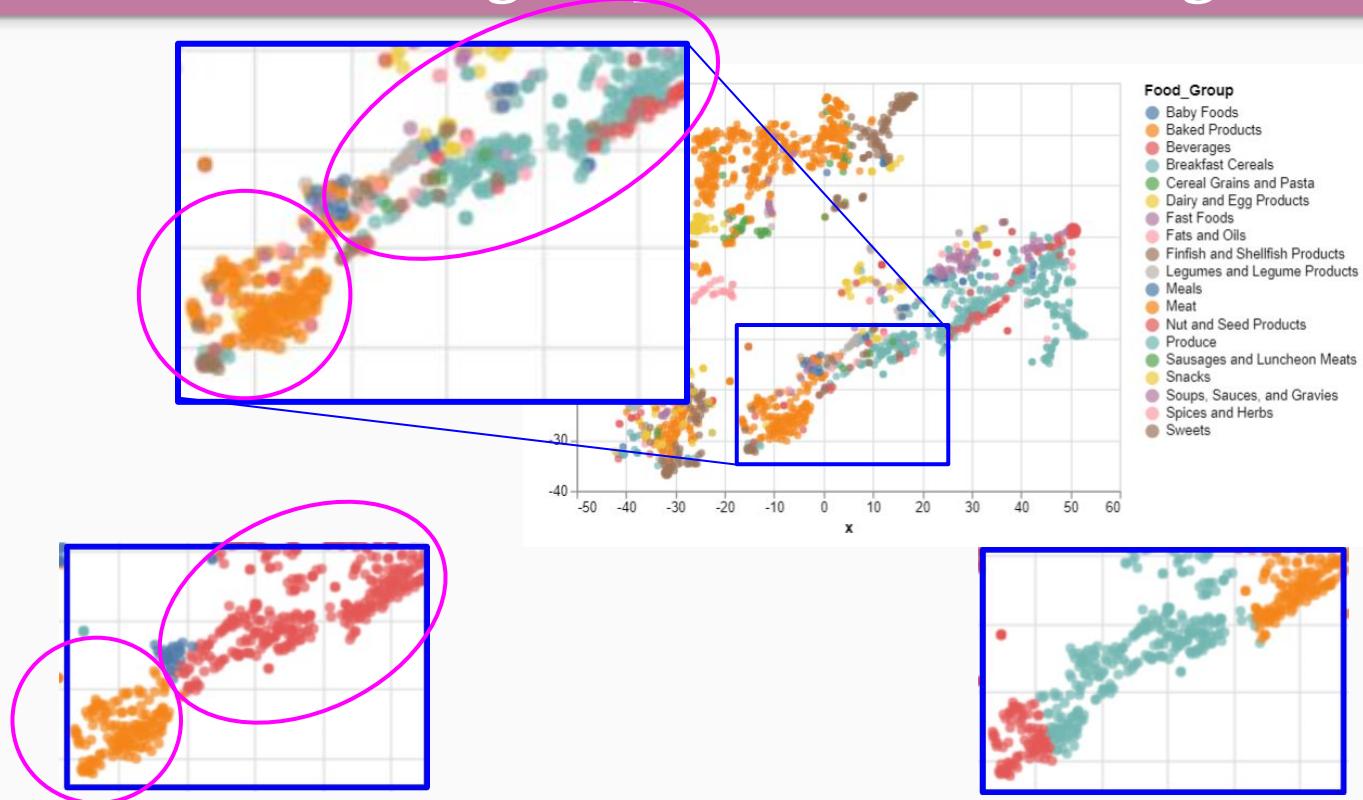


Colored with High Dimensionality  
K-Means Clustering



Colored with Low Dimensionality  
K-Means Clustering

# Question 1: How do we group related foods together?



Colored with High Dimensionality  
K-Means Clustering

Colored with Low Dimensionality  
K-Means Clustering

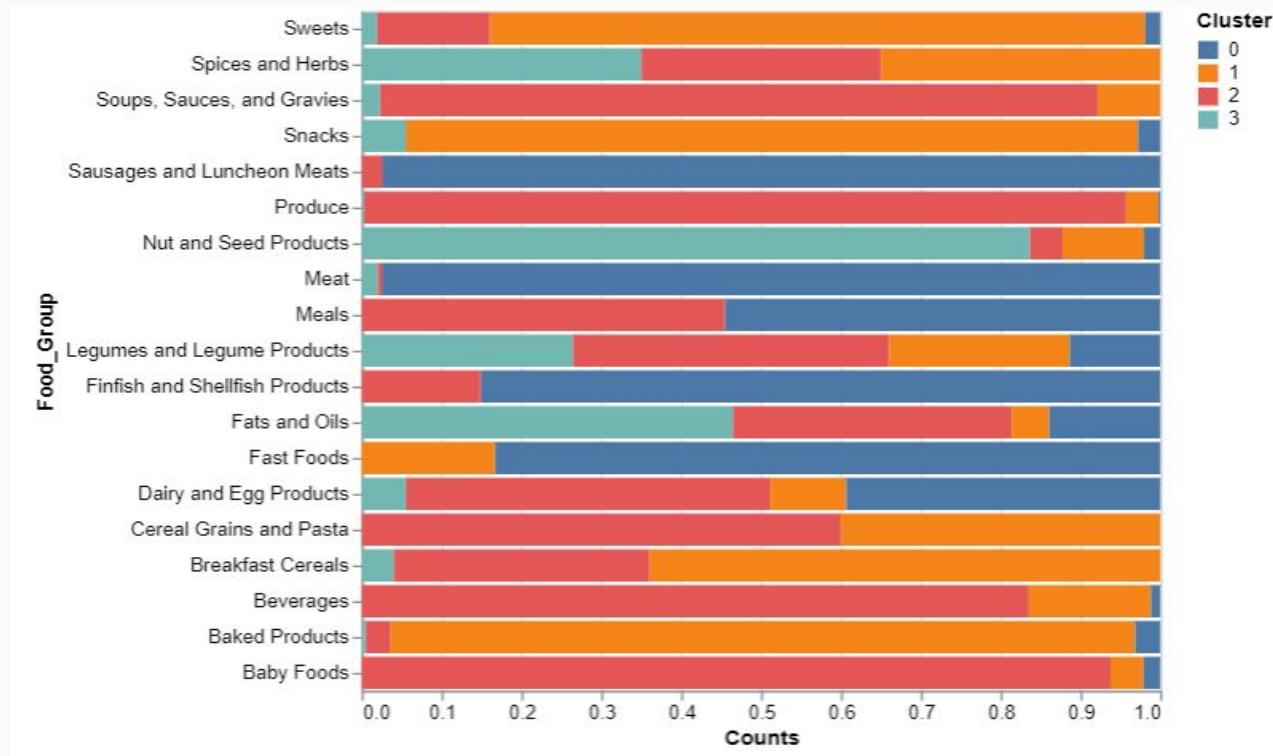
## Question 2:

How do these groups compare  
to grouping by composition?

(Set 1)

# Question 2: How do these groups compare to grouping by composition?

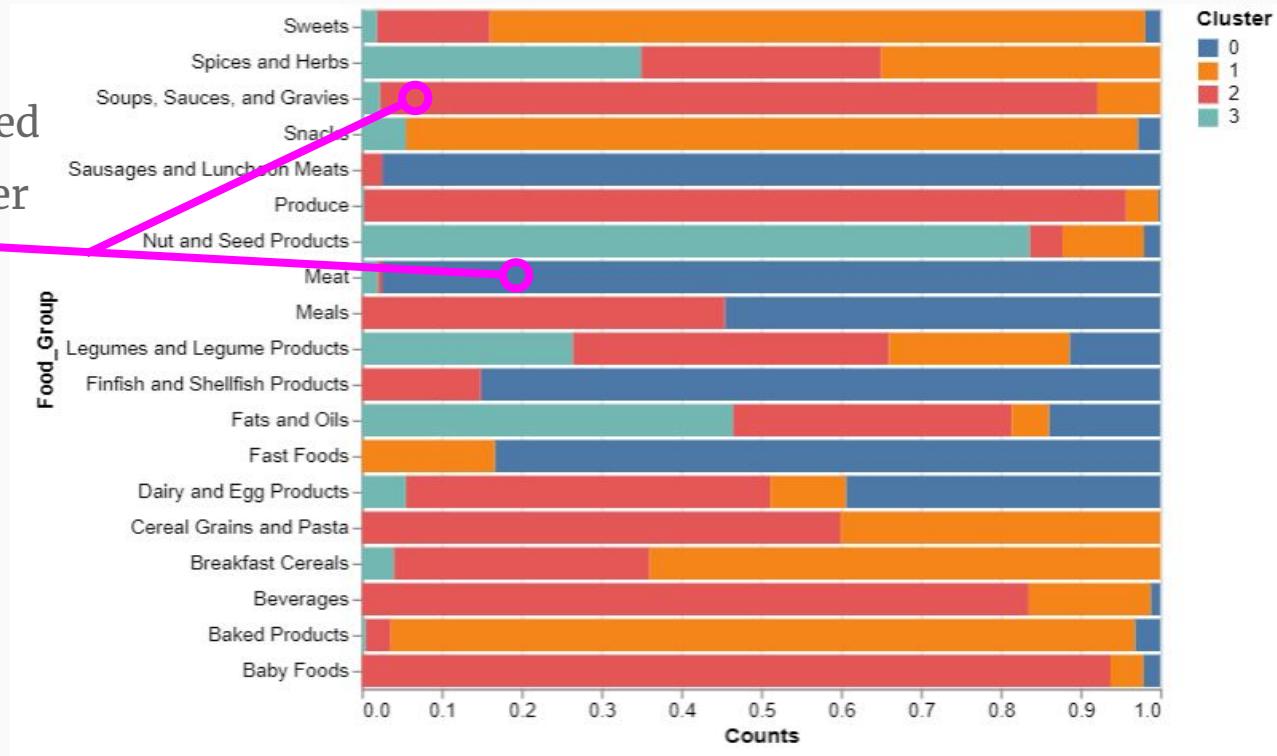
## Distribution of clusters within food groups:



# Question 2: How do these groups compare to grouping by composition?

## Distribution of clusters within food groups:

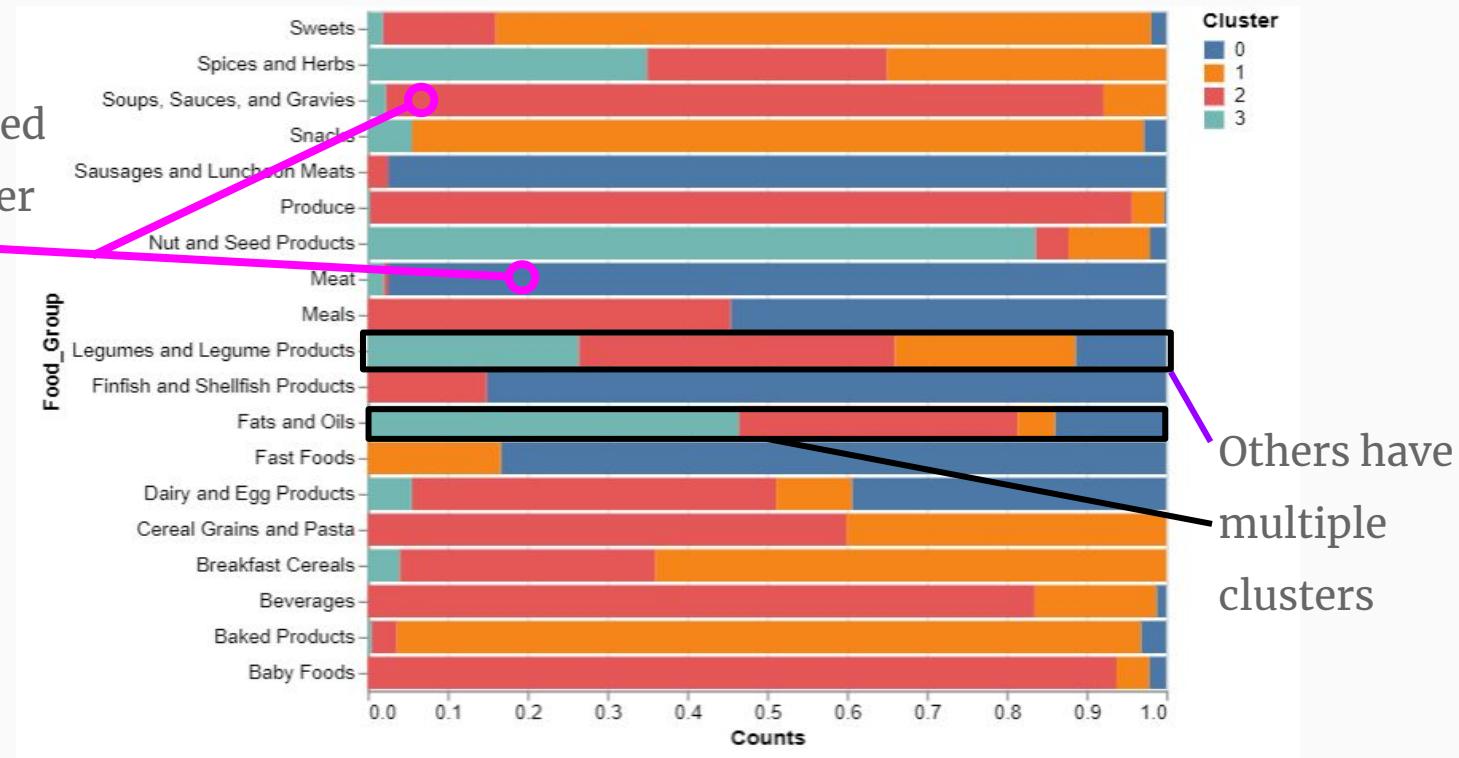
Mostly dominated  
by one cluster per  
category



# Question 2: How do these groups compare to grouping by composition?

Distribution of clusters within food groups:

Mostly dominated  
by one cluster per  
category



Cluster

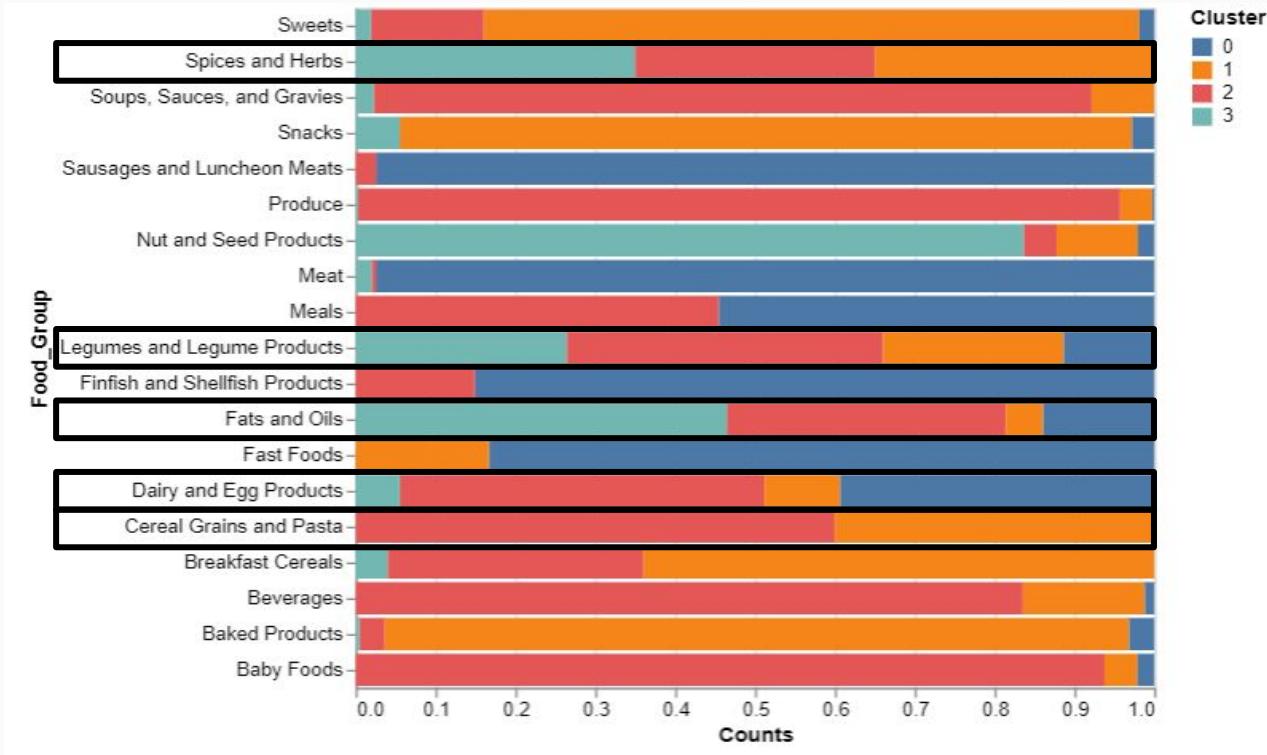
- 0
- 1
- 2
- 3

Others have  
multiple  
clusters

# Question 2: How do these groups compare to grouping by composition?

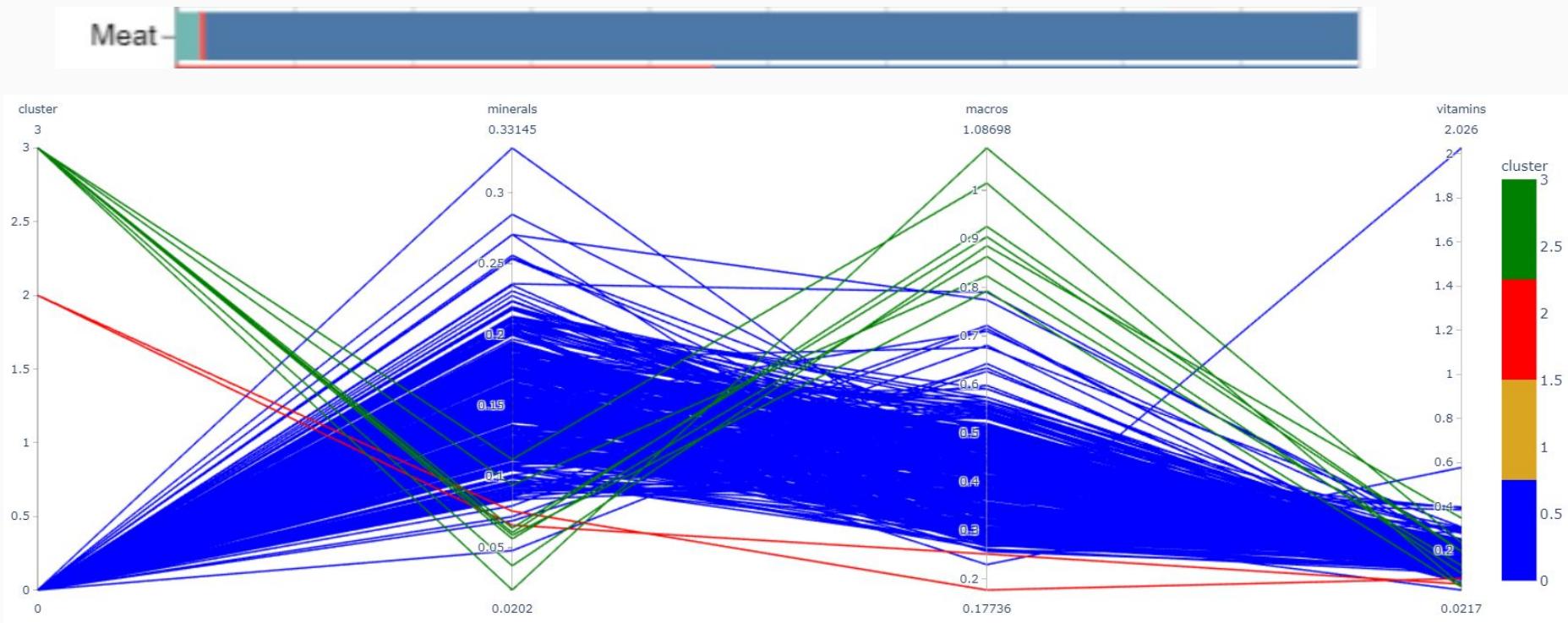
Distribution of clusters within food groups:

Food Groups  
that are  
dominated  
by multiple  
clusters

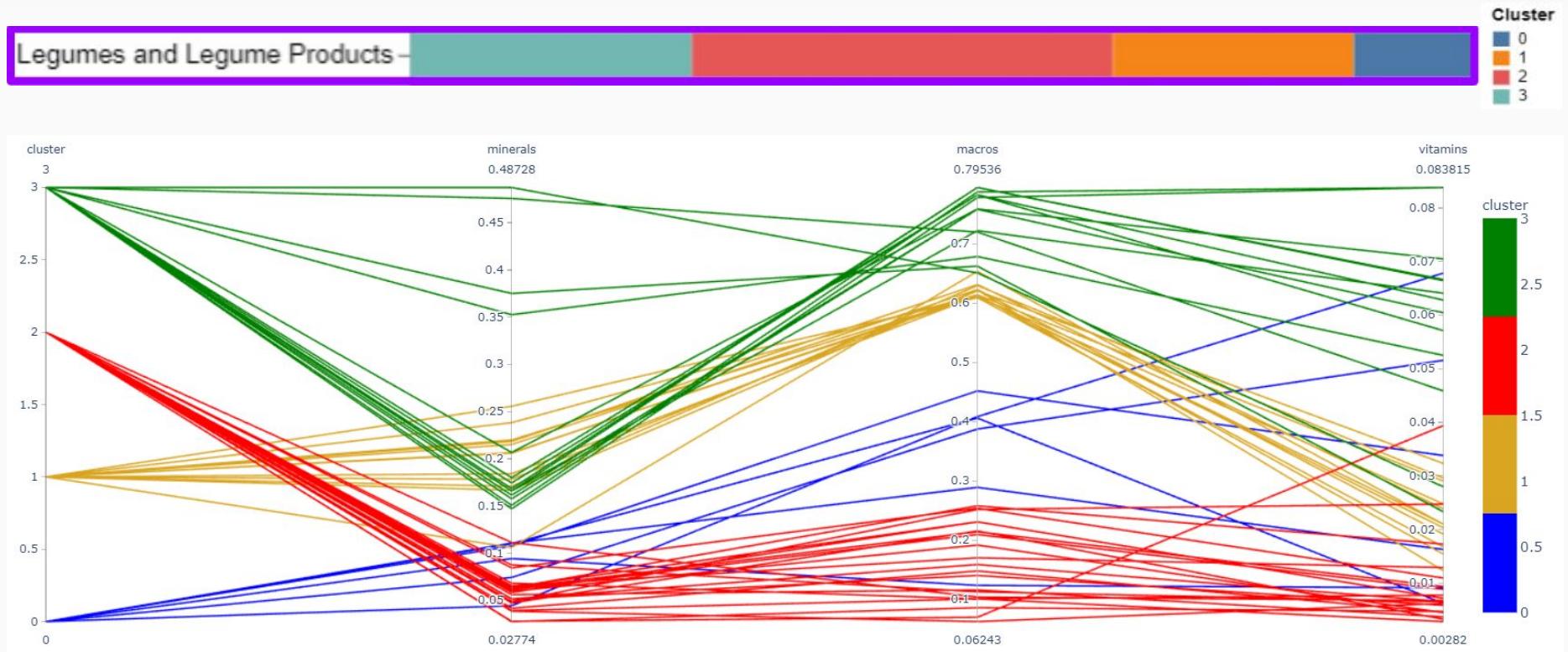


# Question 2: How do these groups compare to grouping by composition?

Parallel coordinates of food group with single dominated cluster:

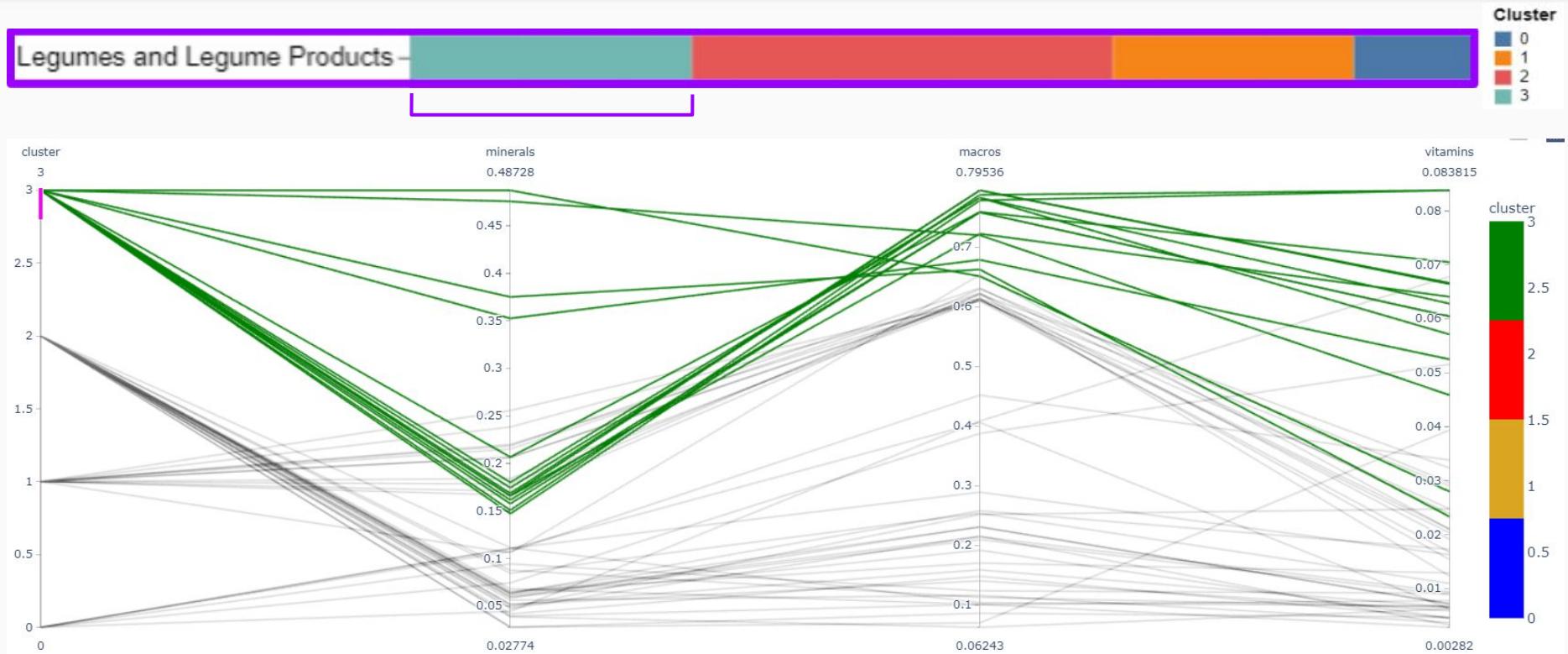


## Question 2: How do these groups compare to grouping by composition?

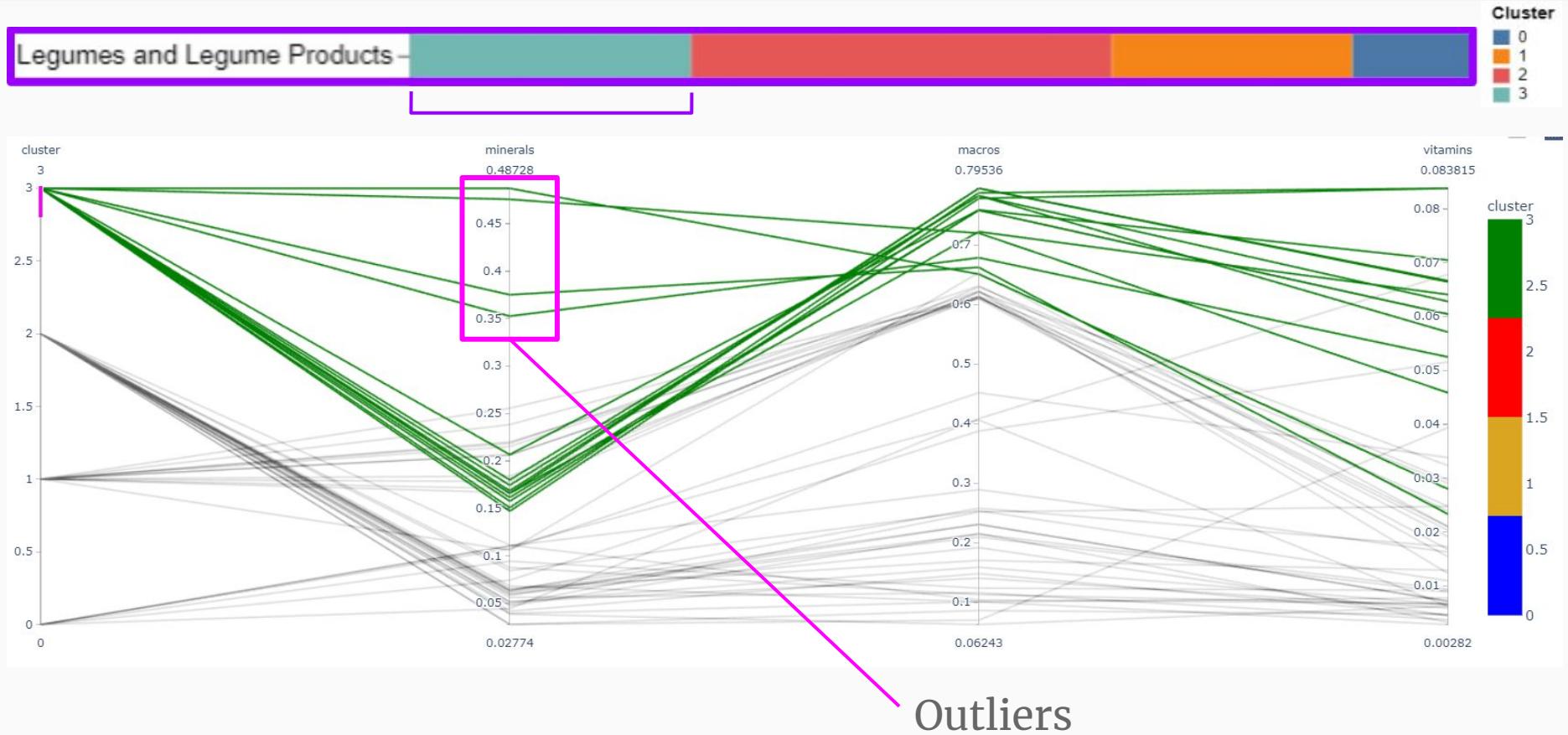


Outliers for food groups dominated with multiple clusters:

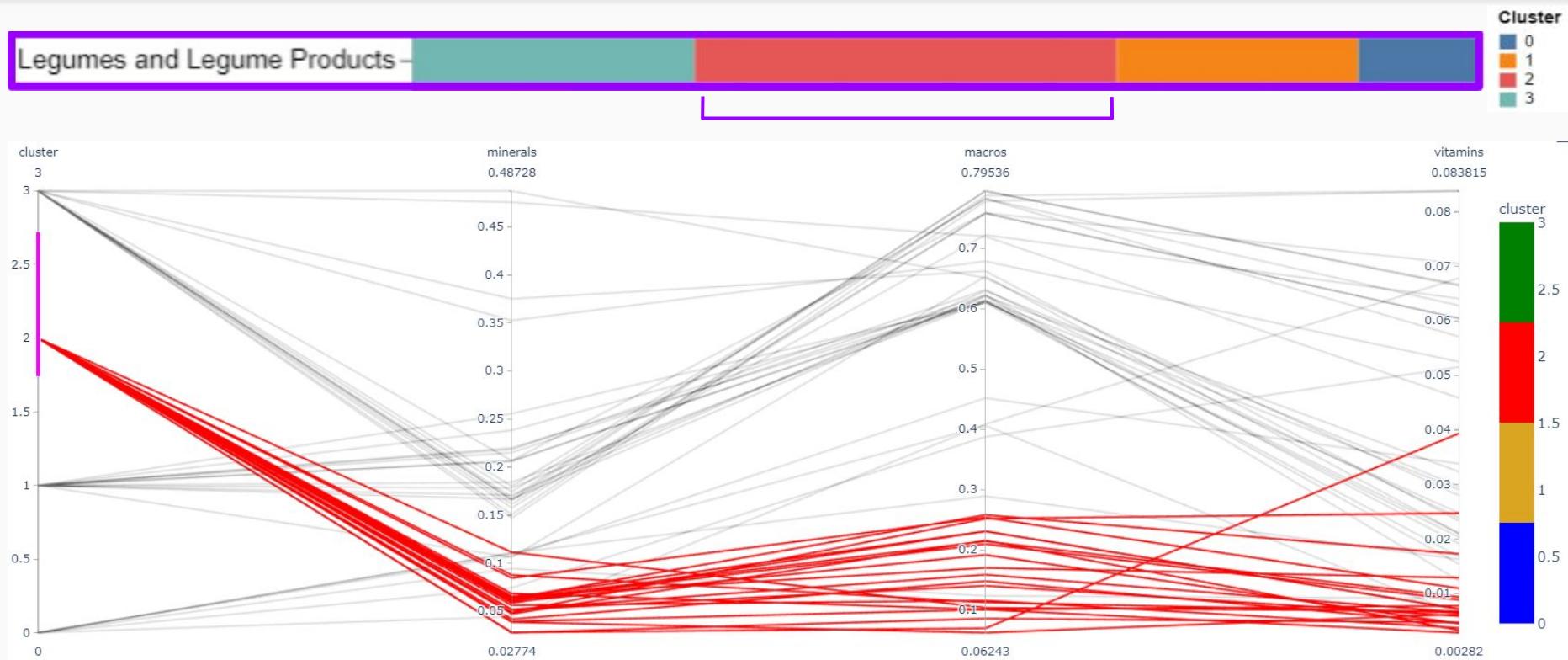
## Question 2: How do these groups compare to grouping by composition?



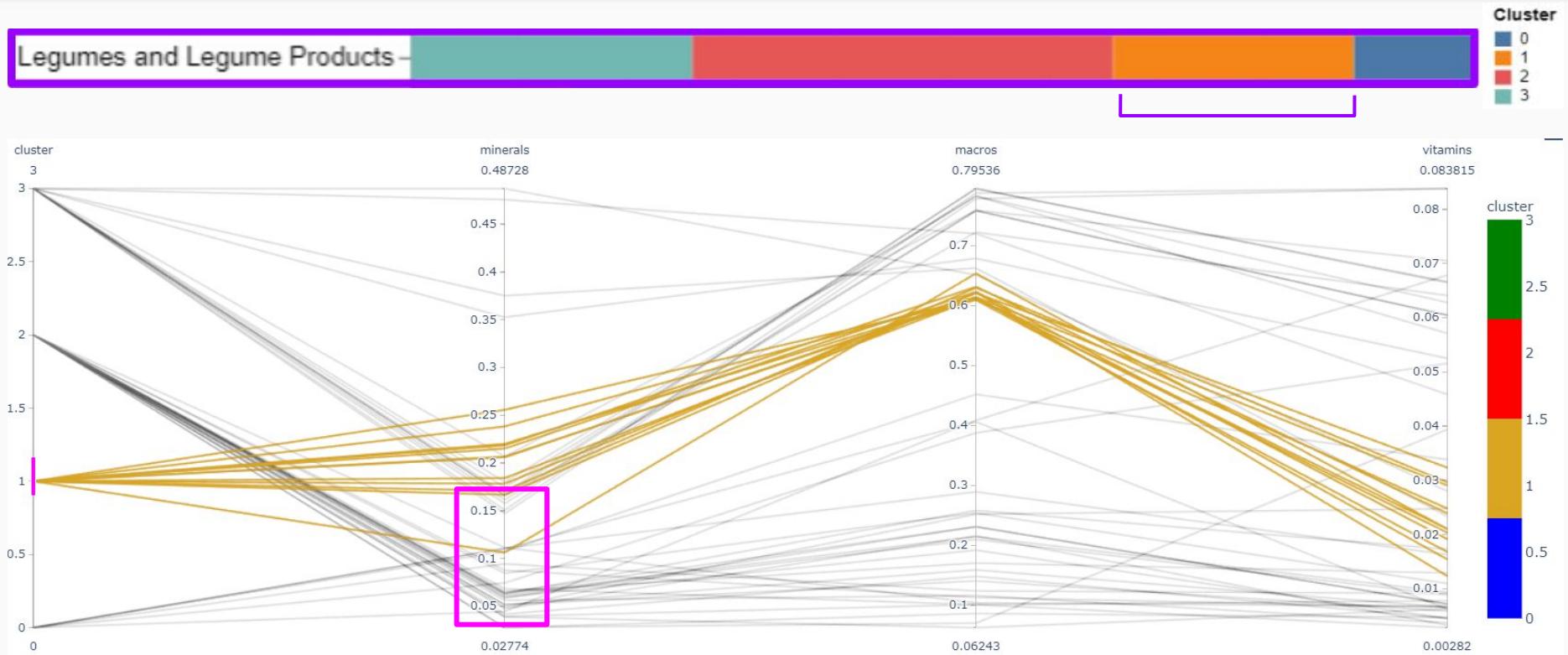
## Question 2: How do these groups compare to grouping by composition?



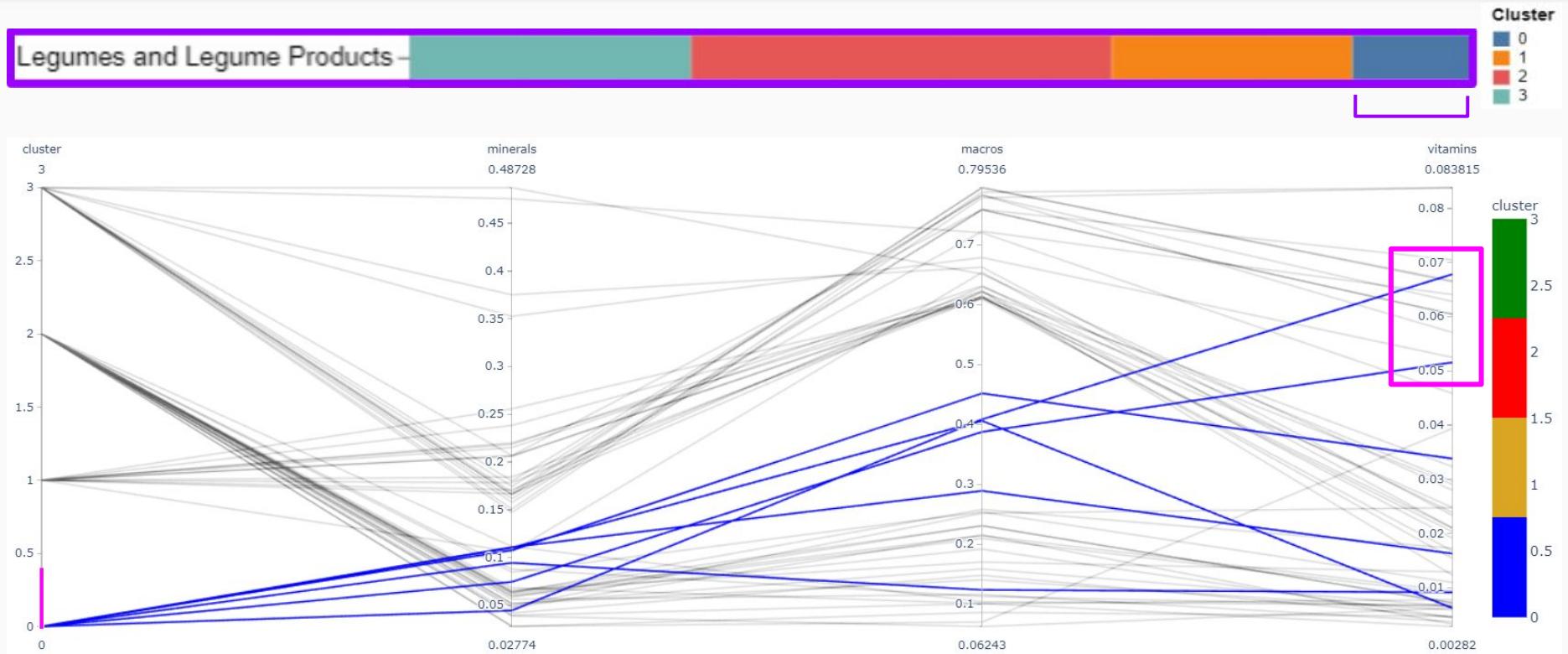
Question 2: How do these groups compare to grouping by composition?



## Question 2: How do these groups compare to grouping by composition?



## Question 2: How do these groups compare to grouping by composition?



# Question 2: How do these groups compare to grouping by composition?

## List of outliers

ID	cluster	group	description	macros	minerals	vitamins	
63	1090	1	Dairy and Egg Products	MILK,DRY,WHL,W/ ADDED VITAMIN D	0.991497	0.467553	1.060659
65	1092	1	Dairy and Egg Products	MILK,DRY,NONFAT,INST,W/ ADDED VITA & VITAMIN D	0.940169	0.607528	1.259574
78	1113	1	Dairy and Egg Products	WHEY,ACID,DRIED	0.877777	0.950164	0.235243
94	1137	3	Dairy and Egg Products	EGG,YOLK,DRIED	1.059231	0.399149	1.615144
98	1154	1	Dairy and Egg Products	MILK,DRY,NONFAT,REG,W/ ADDED VITA & VITAMIN D	0.951012	0.601015	1.247678
125	2003	3	Spices and Herbs	SPICES,BASIL,DRIED	0.794407	3.078479	1.511919
140	2047	3	Spices and Herbs	SALT,TABLE	0.000000	1.016506	0.000000
206	4128	0	Fats and Oils	VEG OIL SPRD,UNSPEC OILS,APPROX 37% FAT,W/ SALT	0.494067	0.020425	0.769056
1284	16104	0	Legumes and Legume Products	BACON,MEATLESS	0.593612	0.137084	0.251900
1287	16115	3	Legumes and Legume Products	SOY FLOUR,FULL-FAT,RAW	1.046393	1.464479	0.170304
1288	16117	3	Legumes and Legume Products	SOY FLOUR,DEFATTED	0.946248	1.500091	0.103903
1709	20099	1	Cereal Grains and Pasta	MACARONI,DRY,ENR	0.914199	0.225000	0.026808
1711	20109	1	Cereal Grains and Pasta	NOODLES,EGG,DRY,ENRICHED	0.929916	0.254631	0.086108
1811	42178	0	Fats and Oils	MAYONNAISE,MADE WITH TOFU	0.483890	0.138013	0.094306

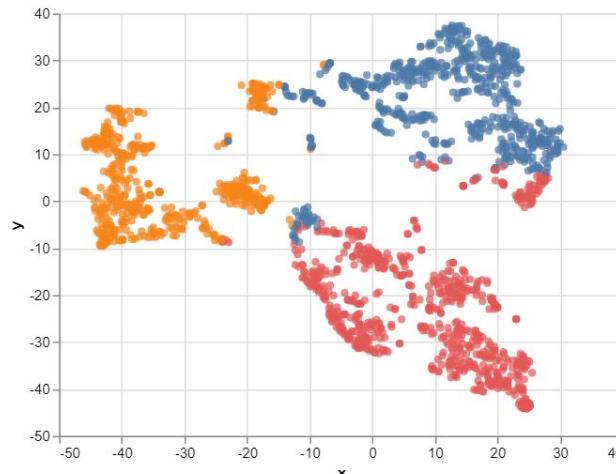
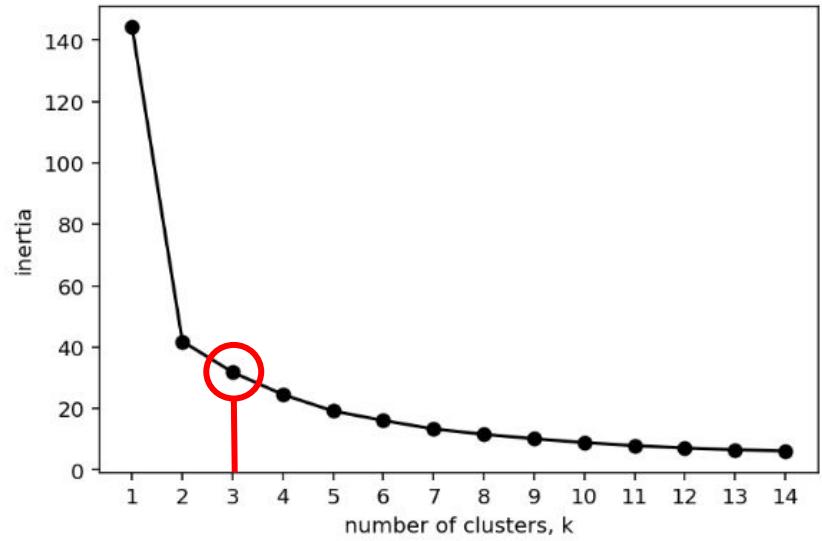
# Question 3:

How does this apply to nutrient subsets?

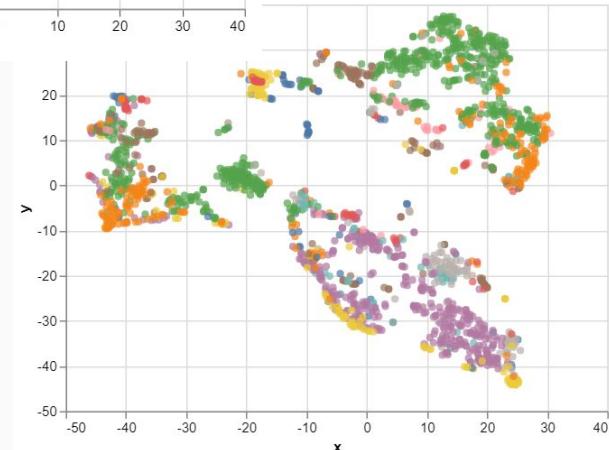
(Set 2)

# Question 3: How does this apply to nutrients?

## Macronutrient Subset:



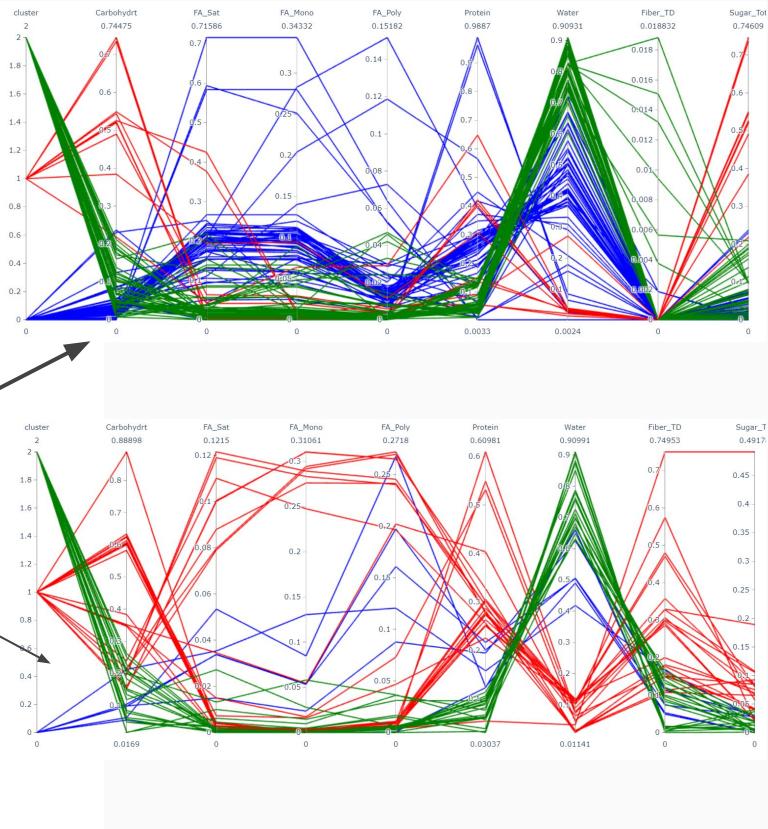
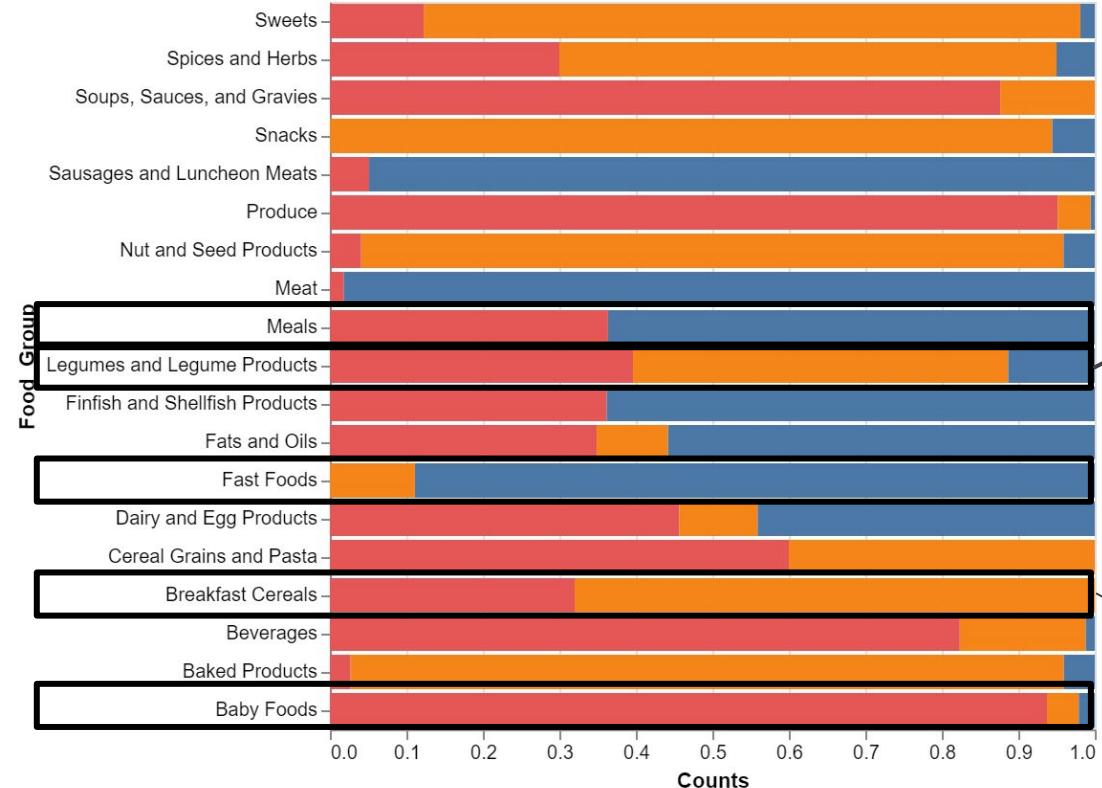
Food labels



High  
Dimensional  
K-Means  
Clustering

# Question 3: How does this apply to nutrients?

## Macronutrient Subset Outliers:

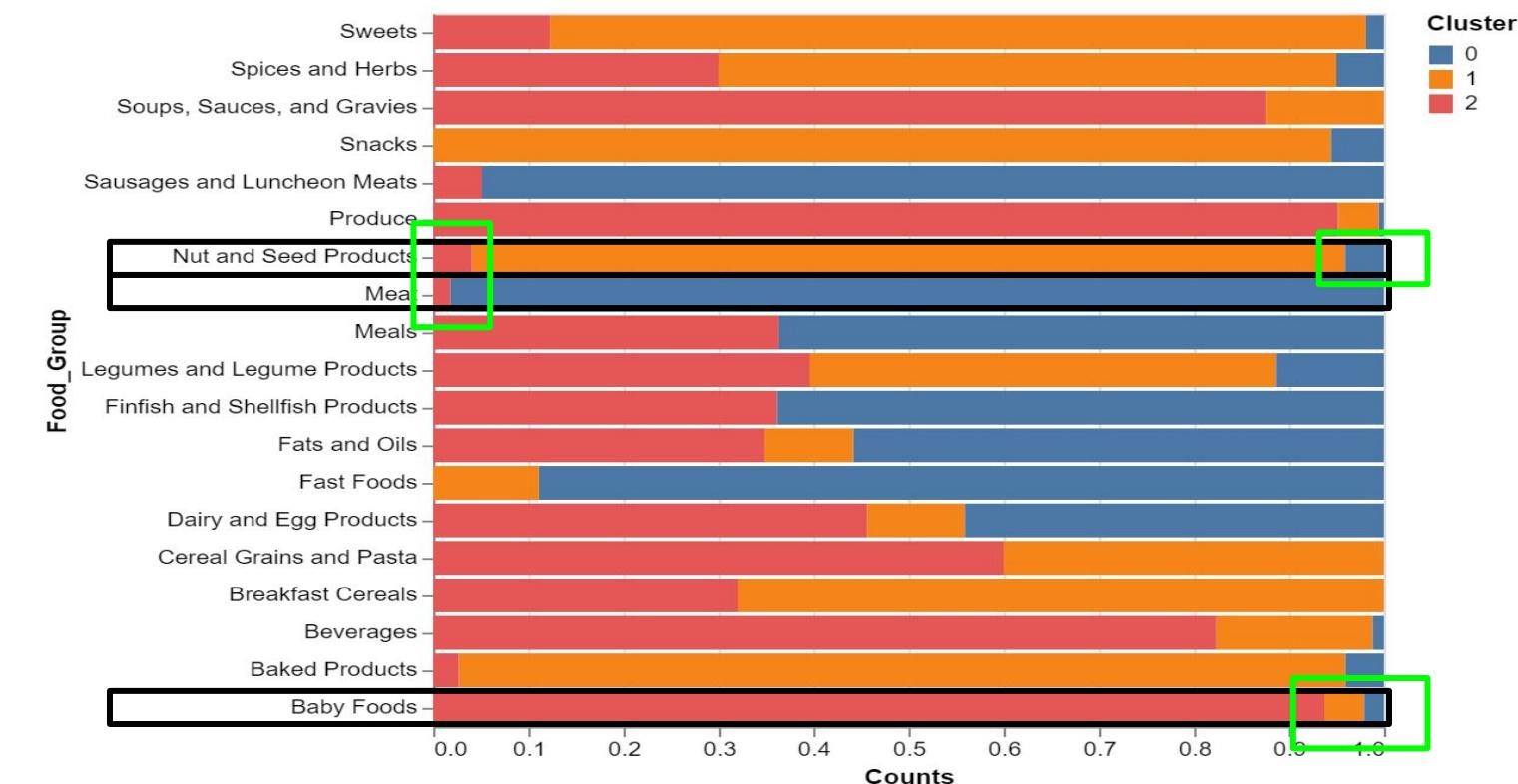


## Question 3: How does this apply to nutrients?

70	2	Dairy and Egg Products	MILK,CHOC,FLUID,COMM,WHL,W/ ADDED VIT A & VITA...
71	2	Dairy and Egg Products	MILK,CHOC,FLUID,COMM,RED FAT
73	2	Dairy and Egg Products	MILK,CHOC BEV,HOT COCOA,HOMEMADE
78	1	Dairy and Egg Products	WHEY,ACID,DRIED
79	1	Dairy and Egg Products	WHEY,SWEET,DRIED
92	0	Dairy and Egg Products	EGG,WHOLE,DRIED
93	0	Dairy and Egg Products	EGG,WHITE,DRIED,PDR,STABILIZED,GLUCOSE RED
94	0	Dairy and Egg Products	EGG,YOLK,DRIED
96	0	Dairy and Egg Products	BUTTER,WITHOUT SALT
124	0	Dairy and Egg Products	EGG,WHITE,DRIED,STABILIZED,GLUCOSE RED
1264	1	Legumes and Legume Products	CAROB FLOUR
1281	1	Legumes and Legume Products	PEANUT BUTTER,SMOOTH STYLE,W/ SALT
1299	1	Legumes and Legume Products	USDA CMDTY,PNUT BUTTER,SMOOTH
1304	1	Legumes and Legume Products	PEANUT BUTTER,SMOOTH STYLE,WO/SALT

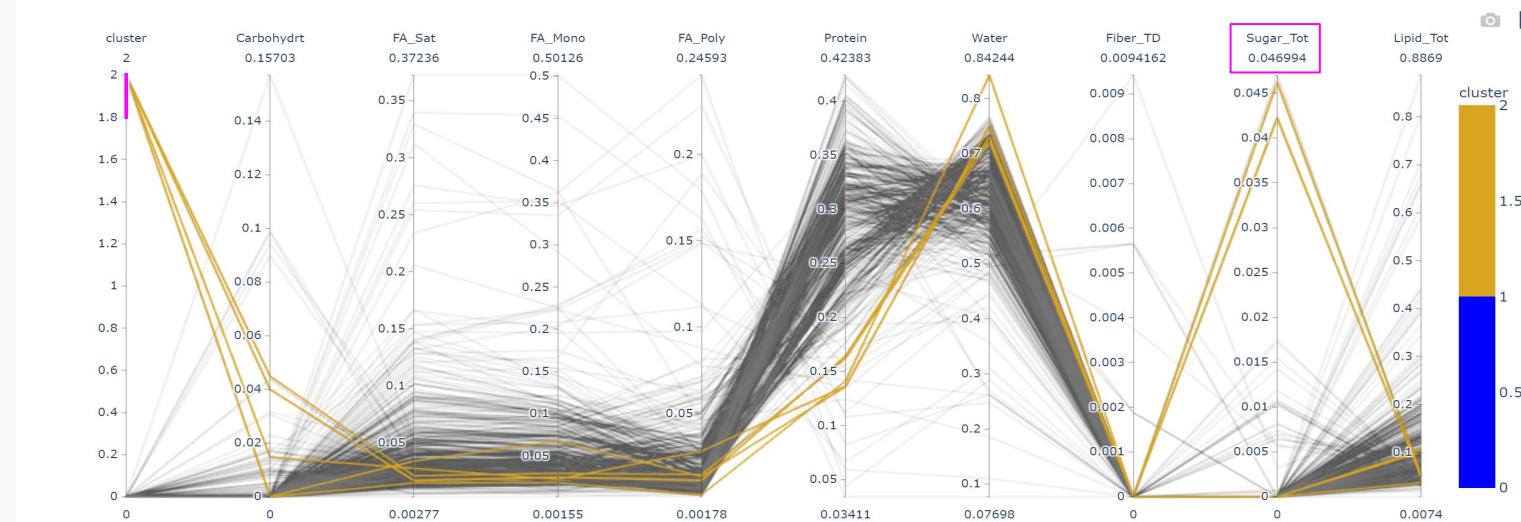
# Question 3: How does this apply to nutrients?

## Macronutrient Subset Outliers:



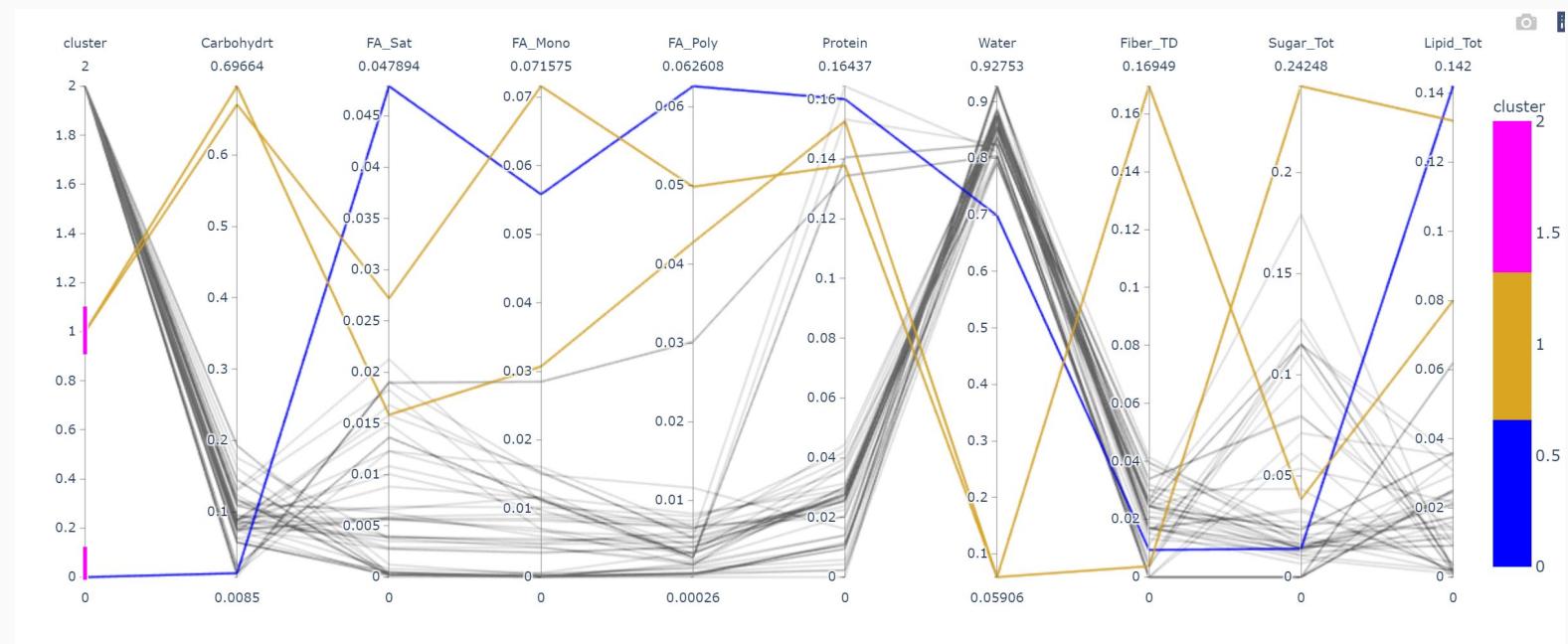
# Macros Outliers

ID		description	cluster	group	Carbohydrt	FA_Sat	FA_Mono	FA_Poly	Protein	Water	Fiber_TD	Sugar_Tot	Lipid_Tot
677	10871	PORK CURED,HAM & H2O PRODUCT,WHL,BNLESS,LN,HTD...	2	Meat	0.044909	0.021282	0.029371	0.014738	0.162150	0.721822	0.0	0.046192	0.0546
678	10872	PORK CURED,HAM & H2O PRODUCT,WHL,BNLESS,LN,UNHTD	2	Meat	0.040008	0.016728	0.024388	0.011153	0.164369	0.731231	0.0	0.042285	0.0486
687	10905	PORK CURED,HAM & H2O PRODUCT,WHL,BNLESS,LN & F...	2	Meat	0.040008	0.017213	0.025069	0.011431	0.164136	0.730330	0.0	0.042184	0.0499
692	10922	PORK CURED,HAM & H2O PRODUCT,WHL,BNLESS,LN & F...	2	Meat	0.044909	0.021282	0.029371	0.014738	0.162150	0.721822	0.0	0.046192	0.0546



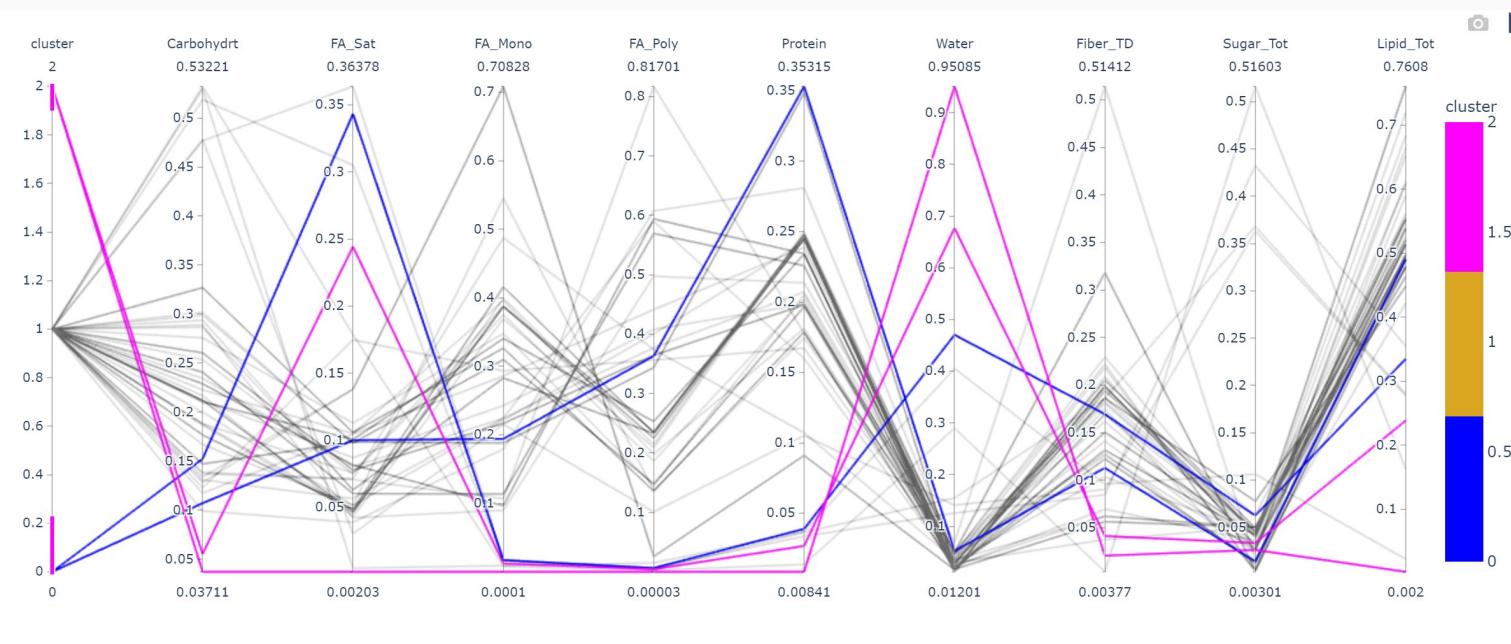
# Macros Outliers

ID		description	cluster	group	Carbohydrt	FA_Sat	FA_Mono	FA_Poly	Protein	Water	Fiber_TD	Sugar_Tot	Lipid_Tot
151	3017	Babyfood, meat, turkey sticks, junior	0	Baby Foods	0.014003	0.047894	0.055814	0.062608	0.160047	0.698699	0.009416	0.014028	0.142
179	3189	BABYFOOD,CRL,OATMEAL,DRY	1	Baby Foods	0.696639	0.015838	0.030709	0.042778	0.152687	0.059059	0.169492	0.038477	0.080
180	3213	BABYFOOD,COOKIES	1	Baby Foods	0.671134	0.027190	0.071575	0.049844	0.137850	0.059059	0.003766	0.242485	0.132



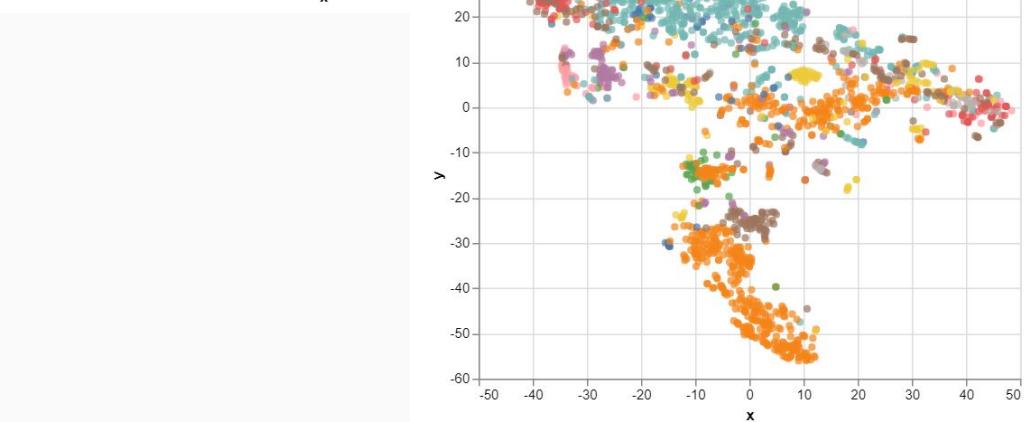
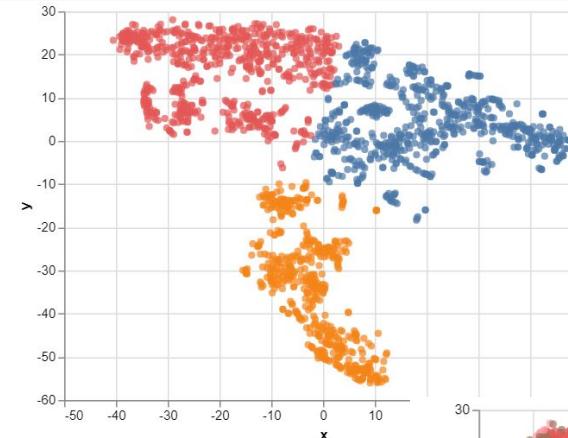
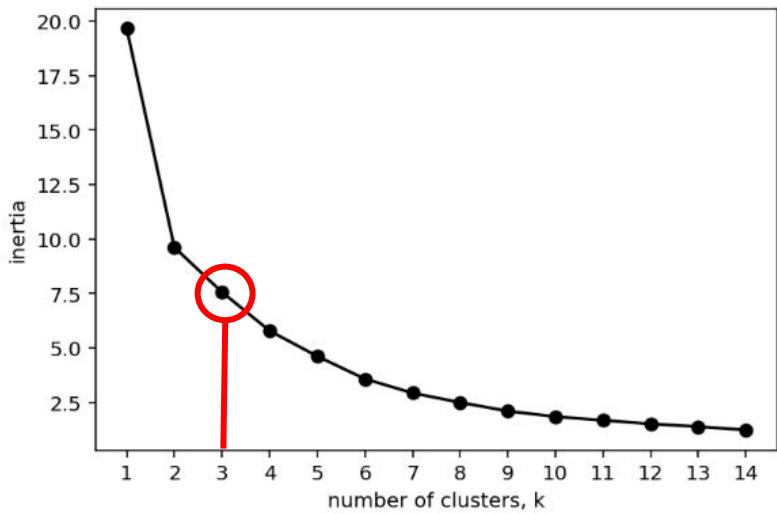
# Macros Outliers

ID		description	cluster	group	Carbohydrt	FA_Sat	FA_Mono	FA_Poly	Protein	Water	Fiber_TD	Sugar_Tot	Lipid_Tot
957	12014	PUMPKIN&SQUASH SD KRNLS, DRIED	0	Nut and Seed Products	0.107121	0.100101	0.194076	0.363284	0.353154	0.052352	0.112994	0.014028	0.4905
972	12104	COCONUT MEAT,RAW	0	Nut and Seed Products	0.152330	0.343318	0.017027	0.006339	0.038902	0.470370	0.169492	0.062425	0.3349
975	12117	COCONUT MILK,RAW (LIQ EXPRESSED FROM GRATED ME...	2	Nut and Seed Products	0.055411	0.244385	0.012116	0.004520	0.026752	0.676877	0.041431	0.033467	0.2384
976	12119	COCONUT H2O (LIQ FROM COCONUTS)	2	Nut and Seed Products	0.037107	0.002035	0.000096	0.000035	0.008411	0.950851	0.020716	0.026152	0.0020



# Question 3: How does this apply to nutrients?

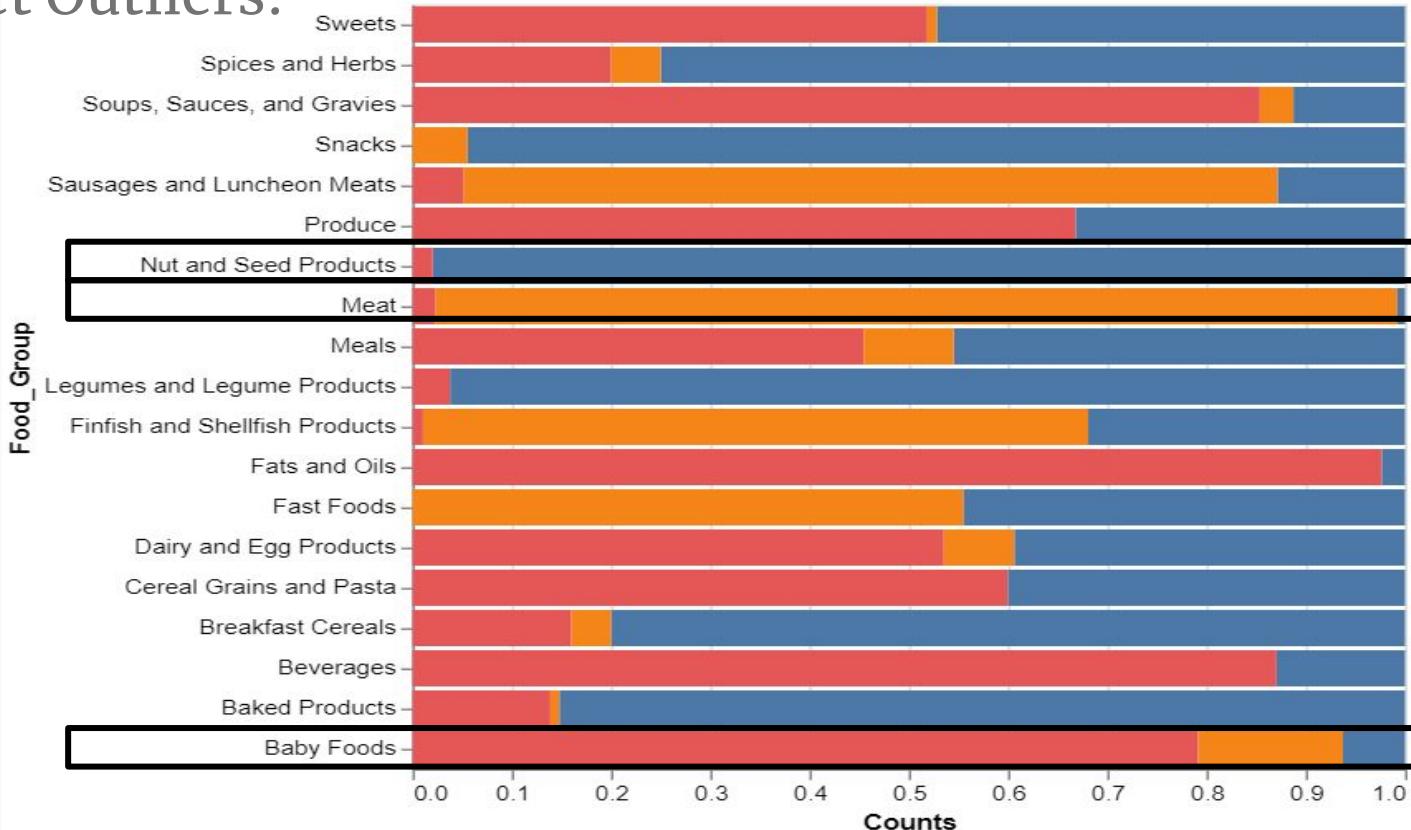
Mineral Subset:



Low  
Dimensional  
K-Means  
Clustering

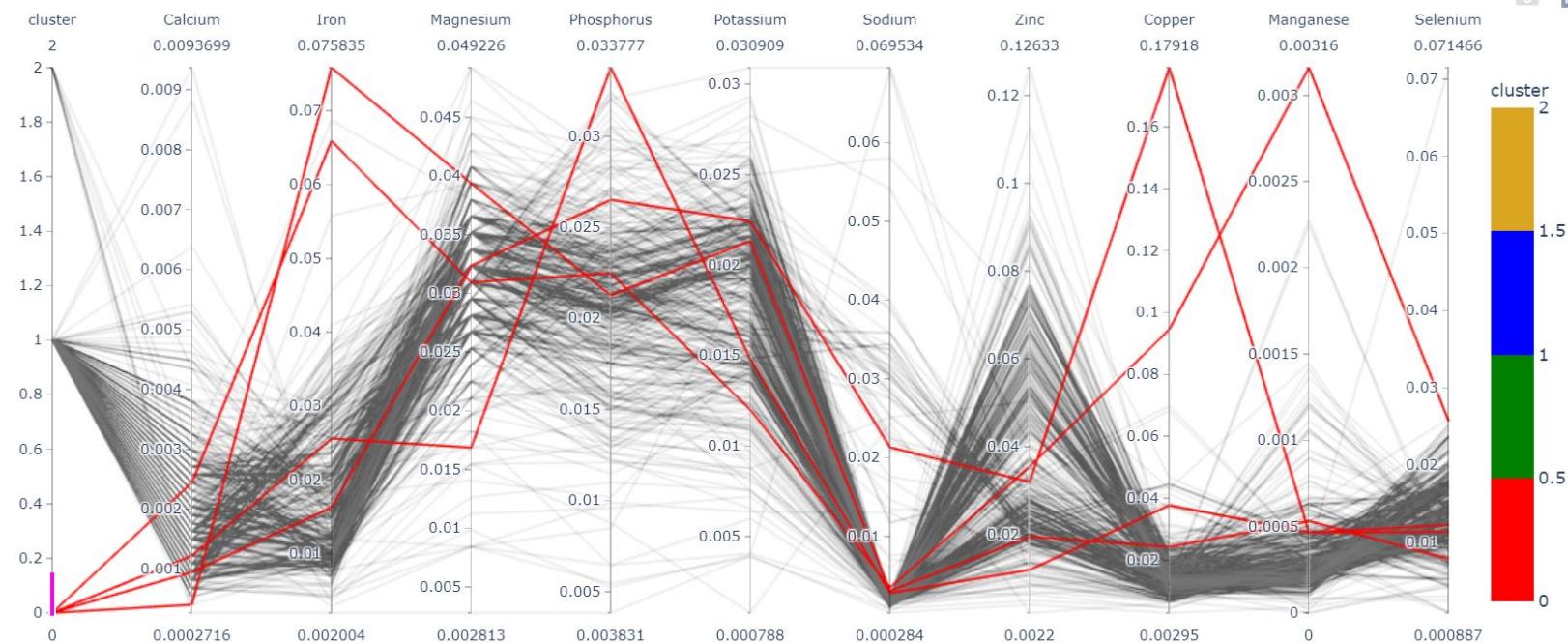
# Mineral Outliers

## Mineral Subset Outliers:



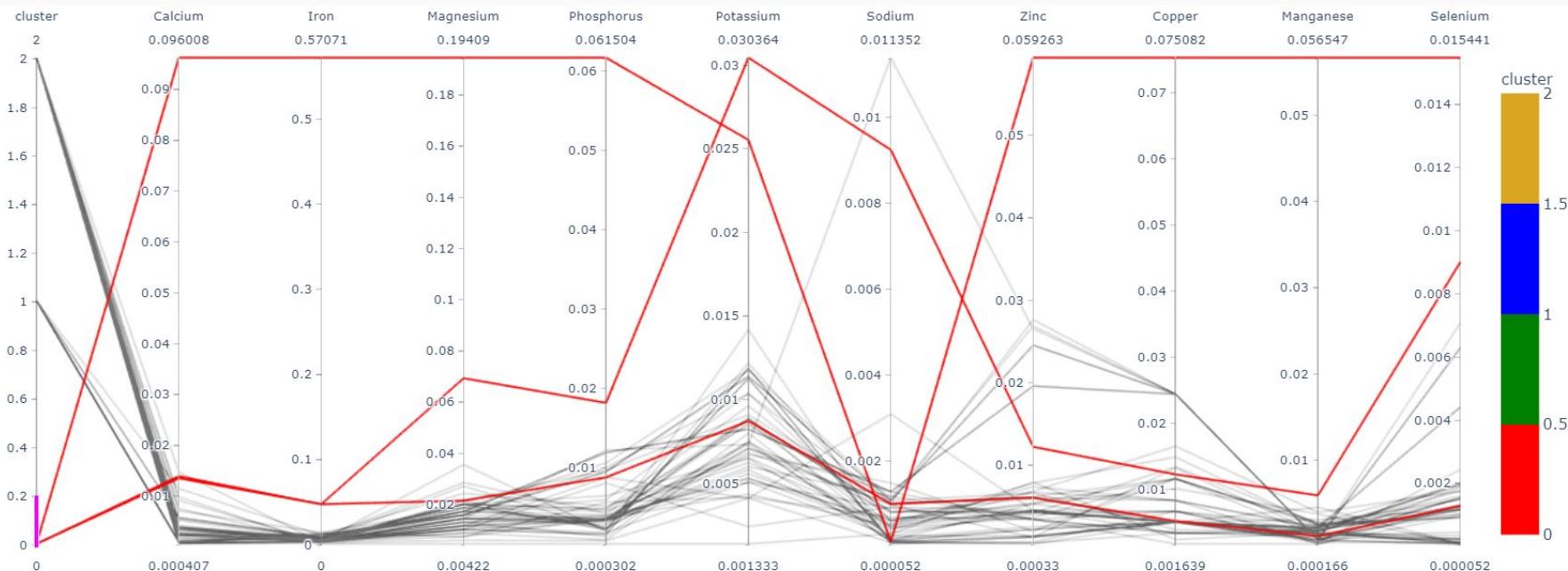
# Mineral Outliers

	description	cluster	group	Calcium	Iron	Magnesium	Phosphorus	Potassium	Sodium	Zinc	Copper	Manganese	Selenium
TURKEY,WHL,GIBLETS,RAW		0	Meat	0.002444	0.065924	0.030942	0.022484	0.012000	0.003509	0.035514	0.094590	0.003160	0.025717
PORK,CURED,HAM,RUMP,BONE-IN,LN & FAT,HTD,RSTD		0	Meat	0.000951	0.016258	0.032349	0.026517	0.022424	0.021312	0.031996	0.179180	0.000466	0.012259
BEEF,VAR MEATS&BY-PRODUCTS,BRAIN,CKD,SIMMRD		0	Meat	0.001222	0.025612	0.016878	0.033777	0.014788	0.002787	0.011985	0.037705	0.000466	0.011372



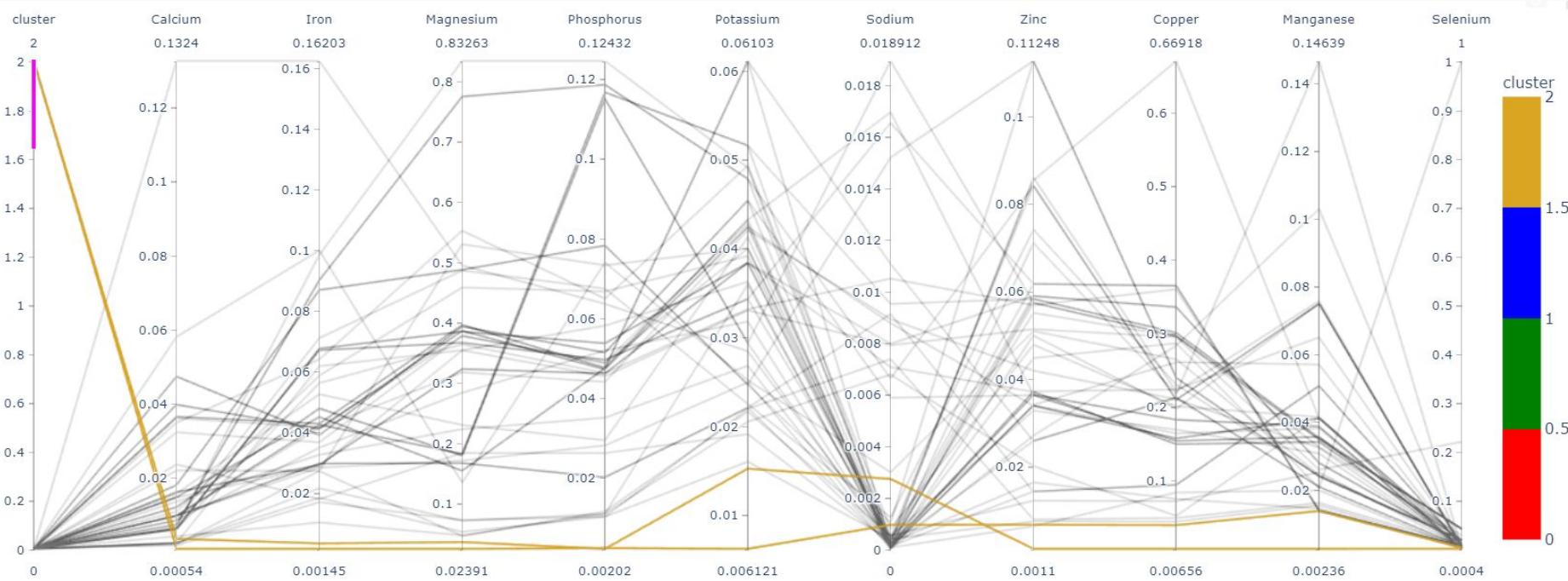
# Mineral Outliers

	description	cluster	group	Calcium	Iron	Magnesium	Phosphorus	Potassium	Sodium	Zinc	Copper	Manganese	Selenium
BABYFOOD,CRL OATMEAL,DRY		0	Baby Foods	0.096008	0.570713	0.194093	0.061504	0.025455	0.000103	0.059263	0.075082	0.056547	0.015441
BABYFOOD COOKIES		0	Baby Foods	0.013715	0.046548	0.068917	0.018048	0.030364	0.009211	0.012095	0.012131	0.005788	0.008972
BABYFOOD,YOGURT,WHL MILK,W/ FRUIT,MULTIGRAIN C...		0	Baby Foods	0.013308	0.046548	0.021097	0.008671	0.008667	0.000980	0.005937	0.005082	0.001064	0.001252



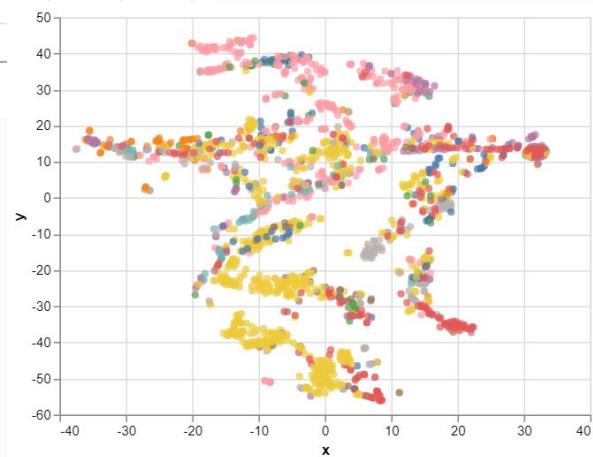
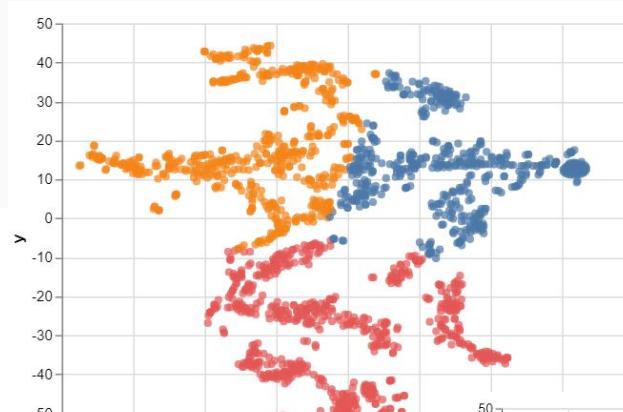
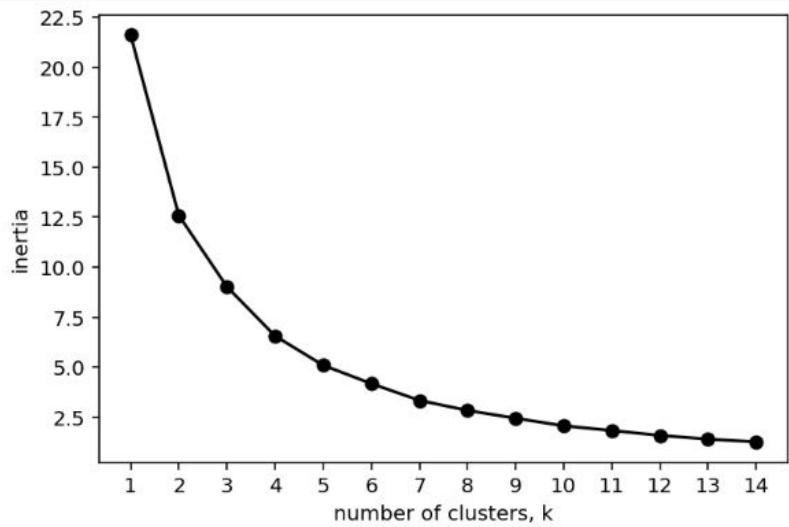
# Mineral Outliers`

	description	cluster	group	Calcium	Iron	Magnesium	Phosphorus	Potassium	Sodium	Zinc	Copper	Manganese	Selenium
NUTS,COCNT CRM,CND,SWTND		2	Nut and Seed Products	0.000543	0.001448	0.023910	0.002218	0.006121	0.000929	0.006597	0.038689	0.013555	0.002869
COCONUT H2O (LIQ FROM COCONUTS)		2	Nut and Seed Products	0.003259	0.003229	0.035162	0.002017	0.015152	0.002709	0.001100	0.006557	0.002362	0.000522



# Question 3: How does this apply to nutrient?

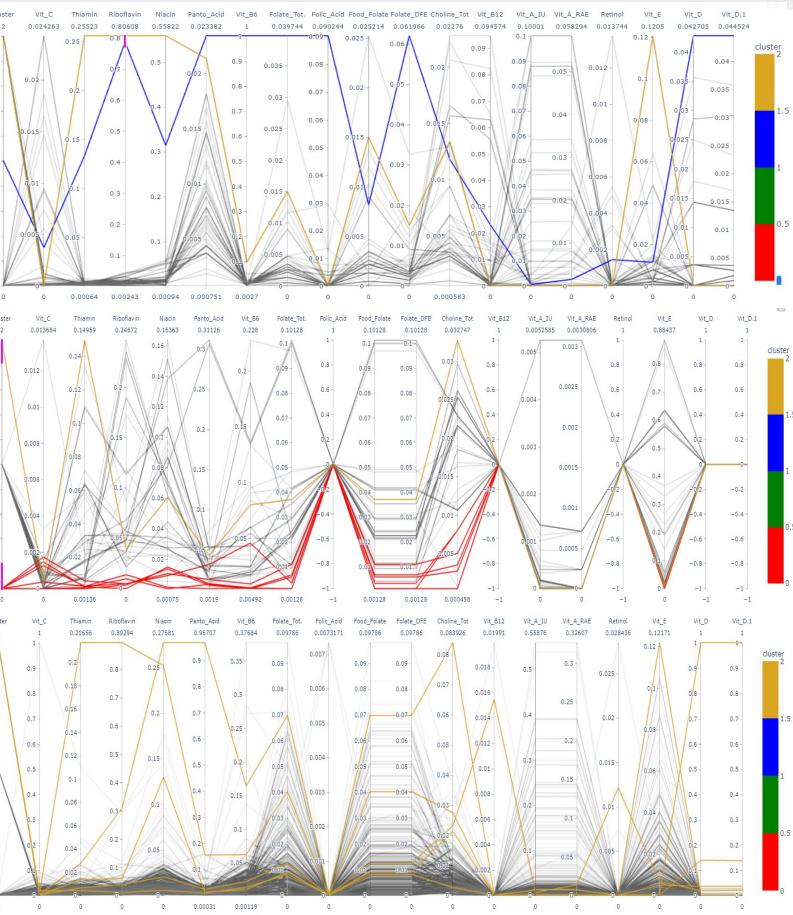
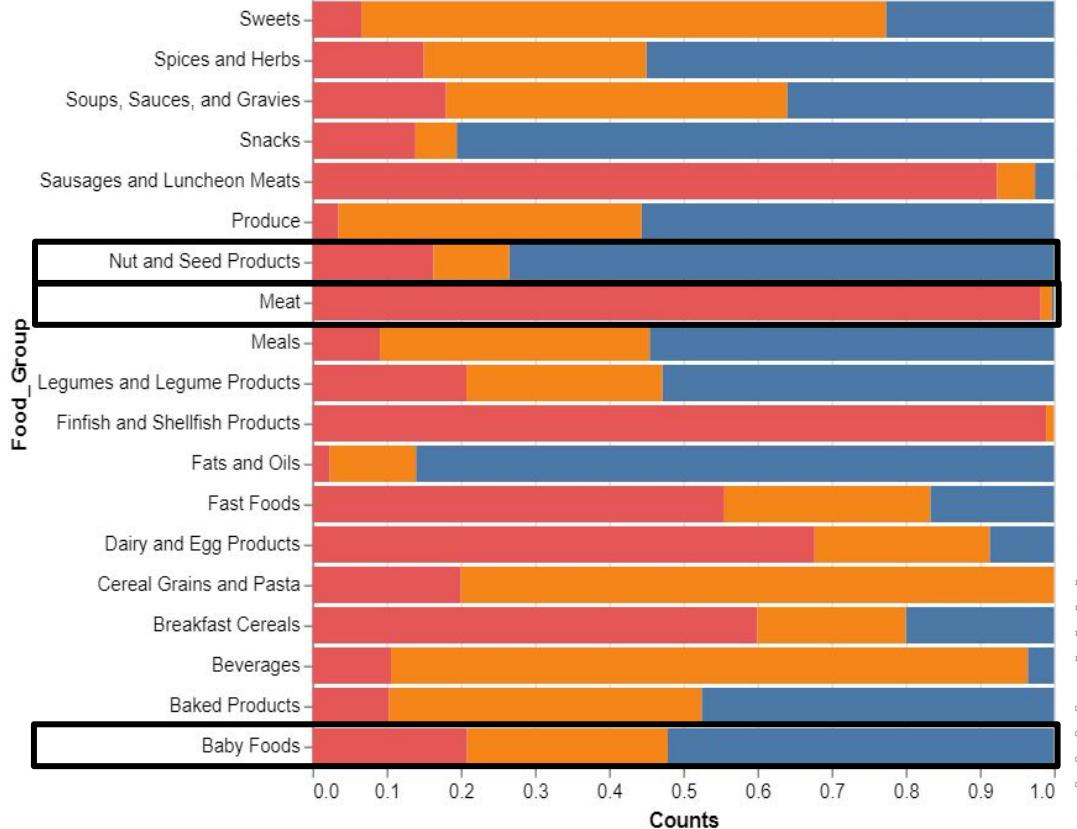
## Vitamin Subset:



Low  
Dimensional  
K-Means  
Clustering

# Vitamins Outliers

## Vitamin Subset Outliers:



# Vitamins Outliers

	description	cluster	group	Vit_C	Thiamin	Riboflavin	Niacin	Panto_Acid	Vit_B6	Folate_Tot.	Folic_Acid	Food_Folate	Folate_DFE	Choline_Tot	Vit_B12
BABYFOOD,CRL,OATMEAL,DRY		2	Baby Foods	0.000000	0.255232	0.806083	0.558219	0.021305	0.09544	0.014957	0.000000	0.014957	0.014957	0.013357	0.000000
BABYFOOD,COOKIES		1	Baby Foods	0.003684	0.132757	0.785645	0.313446	0.023382	1.00000	0.039744	0.090244	0.008120	0.061966	0.011775	0.022897
COCONUT MILK,RAW (LIQ EXPRESSED FROM GRATED ME...		0	Nut and Seed Products	0.001474	0.002366	0.000000	0.014920	0.008089	0.005594	0.006838	0.0	0.006838	0.006838	0.003537	0.0
COCONUT H2O (LIQ FROM COCONUTS)		0	Nut and Seed Products	0.001263	0.002730	0.013869	0.001571	0.001901	0.005425	0.001282	0.0	0.001282	0.001282	0.000458	0.0
COCONUT MEAT,DRIED (DESICCATED),SWTND,SHREDDED		0	Nut and Seed Products	0.000368	0.002821	0.004866	0.009306	0.031913	0.045940	0.003419	0.0	0.003419	0.003419	0.008031	0.0
SEEDS,FLAXSEED		2	Nut and Seed Products	0.000316	0.149591	0.039173	0.060467	0.043538	0.080183	0.037179	0.0	0.037179	0.037179	0.032747	0.0

# Extended Analysis: Exploring Superfoods

(Set 3)

# Extended Analysis: Exploring Superfoods

## Superfoods

How do foods we consider ‘super’ or extra healthy compare to their clusters and labels?

# Extended Analysis: Exploring Superfoods

## Superfoods

How do foods we consider ‘super’ or extra healthy compare to their clusters and labels?

Figs, Flax Seed, Ginger Root, Kale, Salmon



Ginger  
Figs



Ginger  
Flax Seeds



Ginger



Curly kale

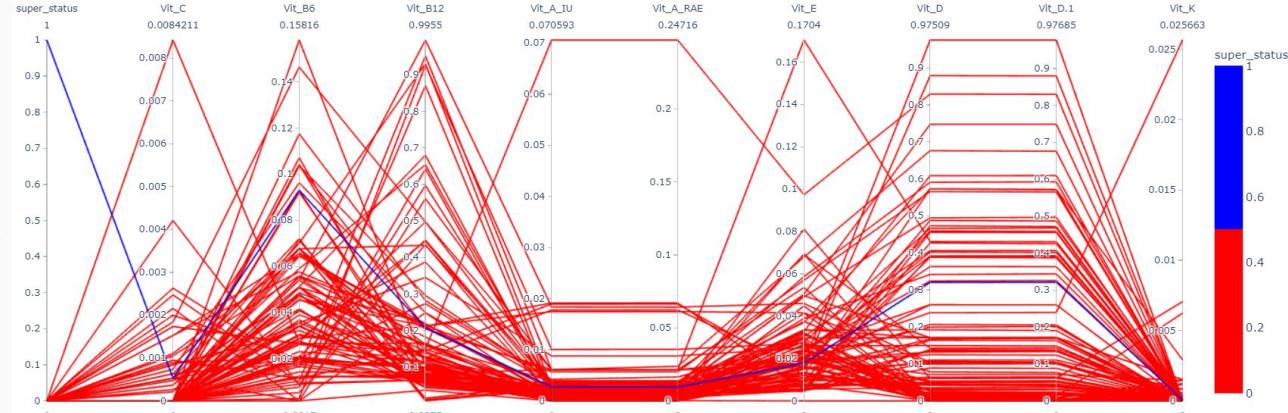


Salmon

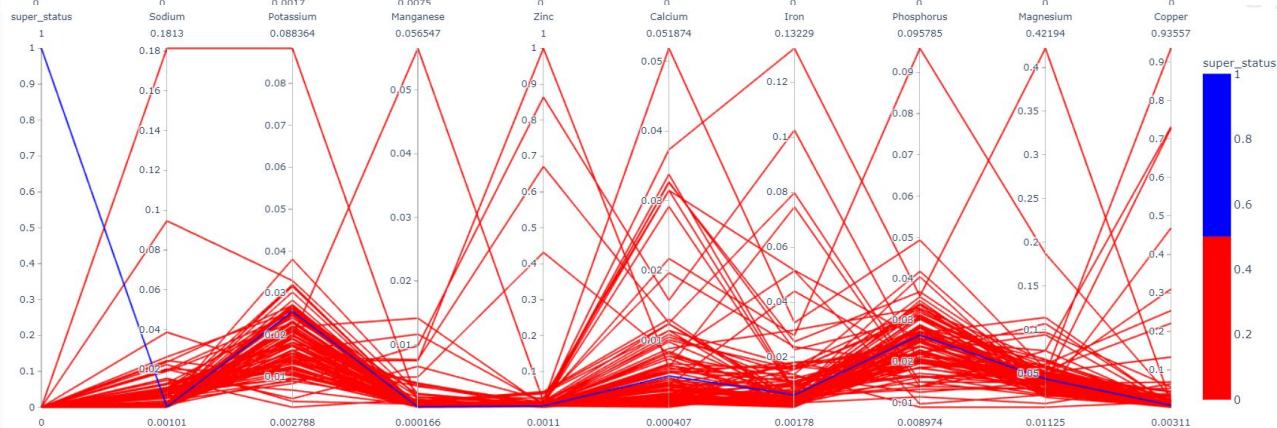
# Extended Analysis: Exploring Superfoods

## Superfoods- Salmon v.s. Finfish and Shellfish Group

Vitamins



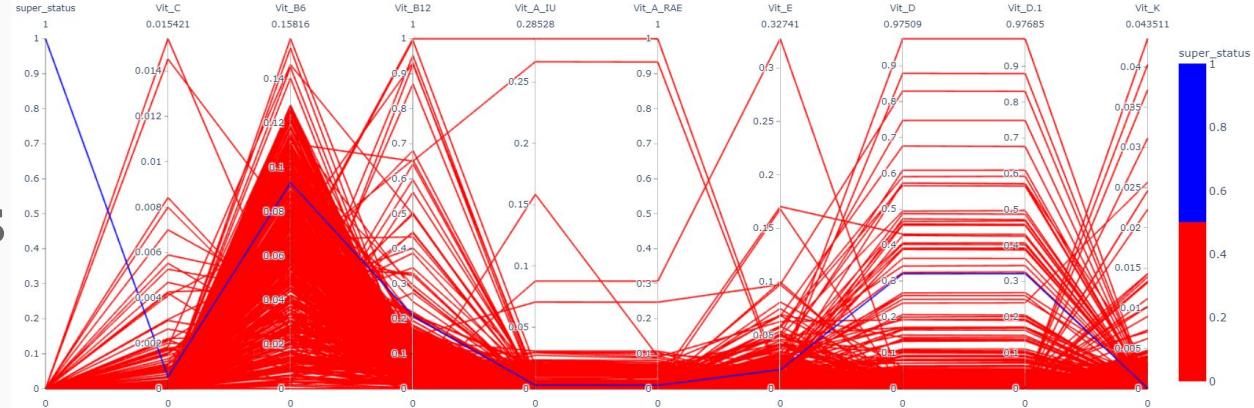
Minerals



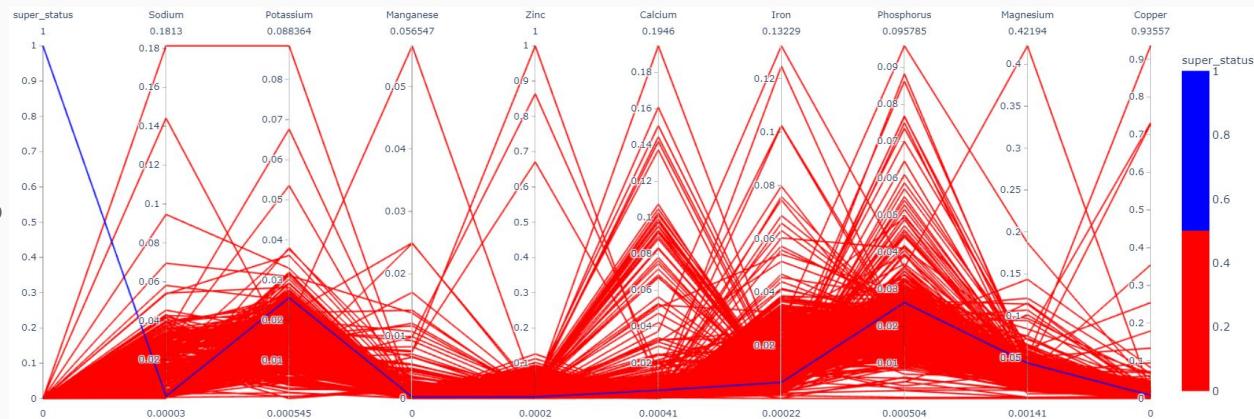
# Extended Analysis: Exploring Superfoods

## Superfoods- Salmon v.s. K-Means Data Cluster

Vitamins



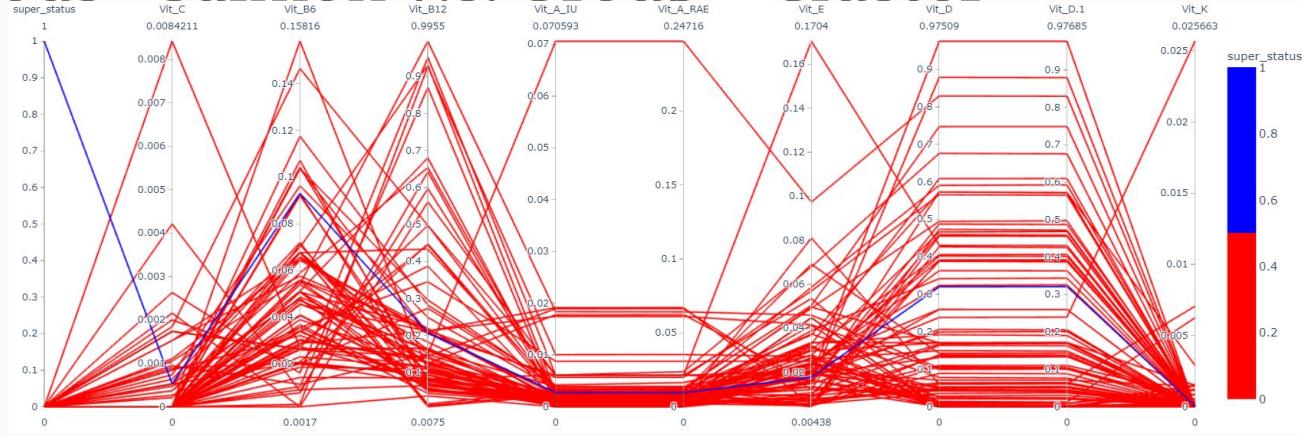
Minerals



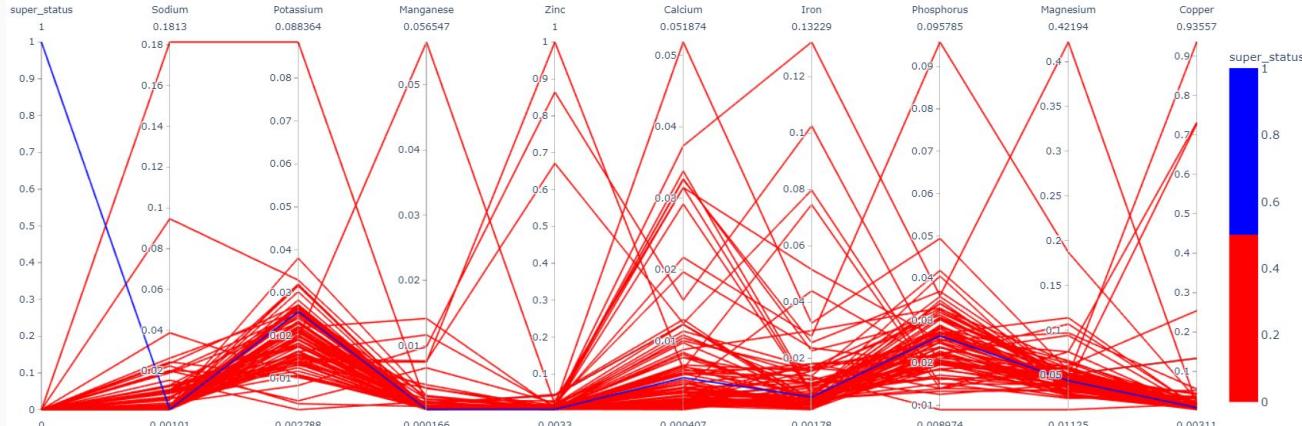
# Extended Analysis: Exploring Superfoods

## Superfoods- Salmon v.s. Group $\cap$ Cluster

Vitamins

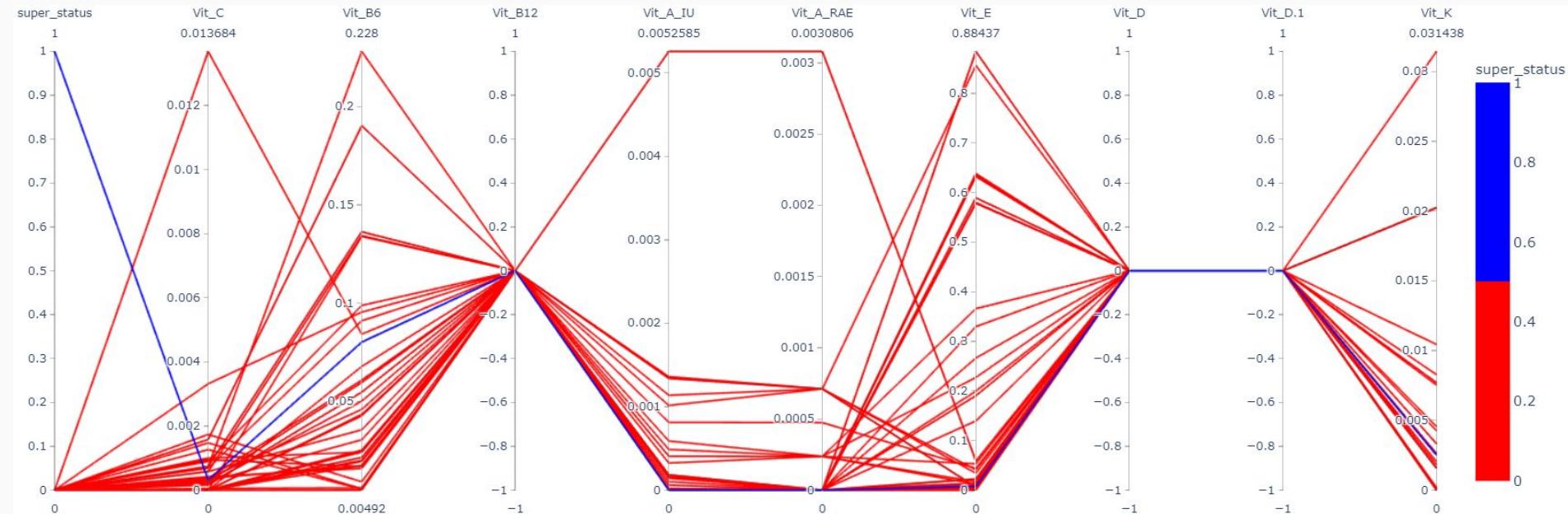


Minerals



# Extended Analysis: Exploring Superfoods

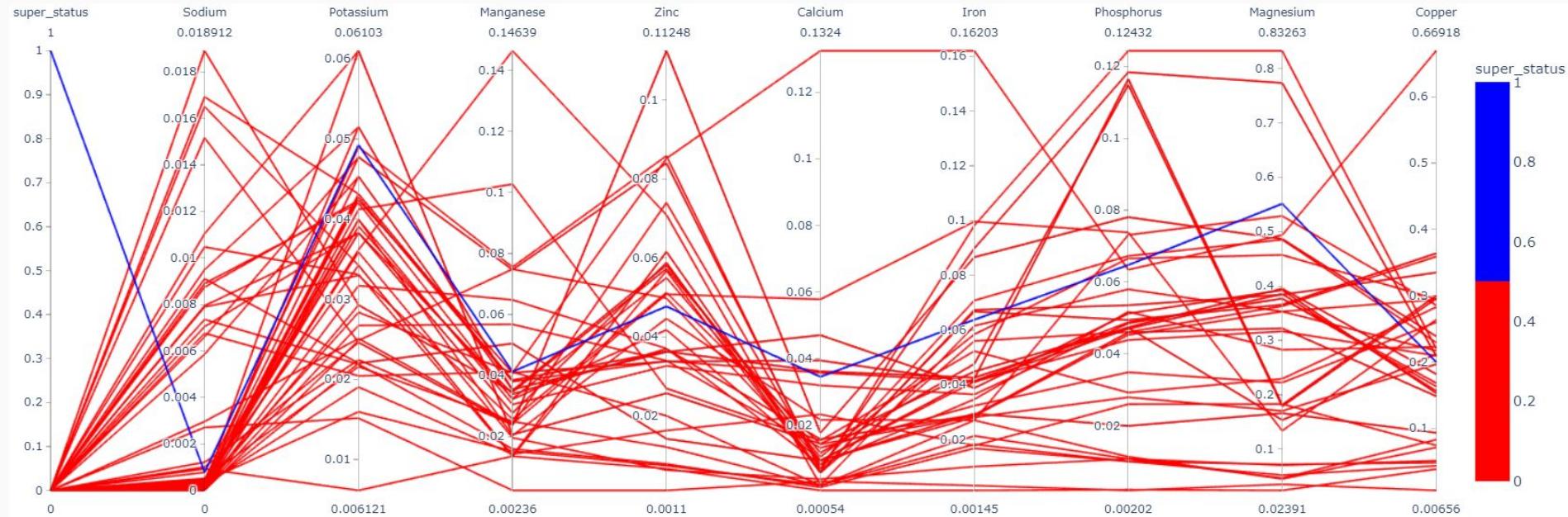
## Superfoods- Flax Seed v.s. Nuts and Seeds Group



# Vitamins

# Extended Analysis: Exploring Superfoods

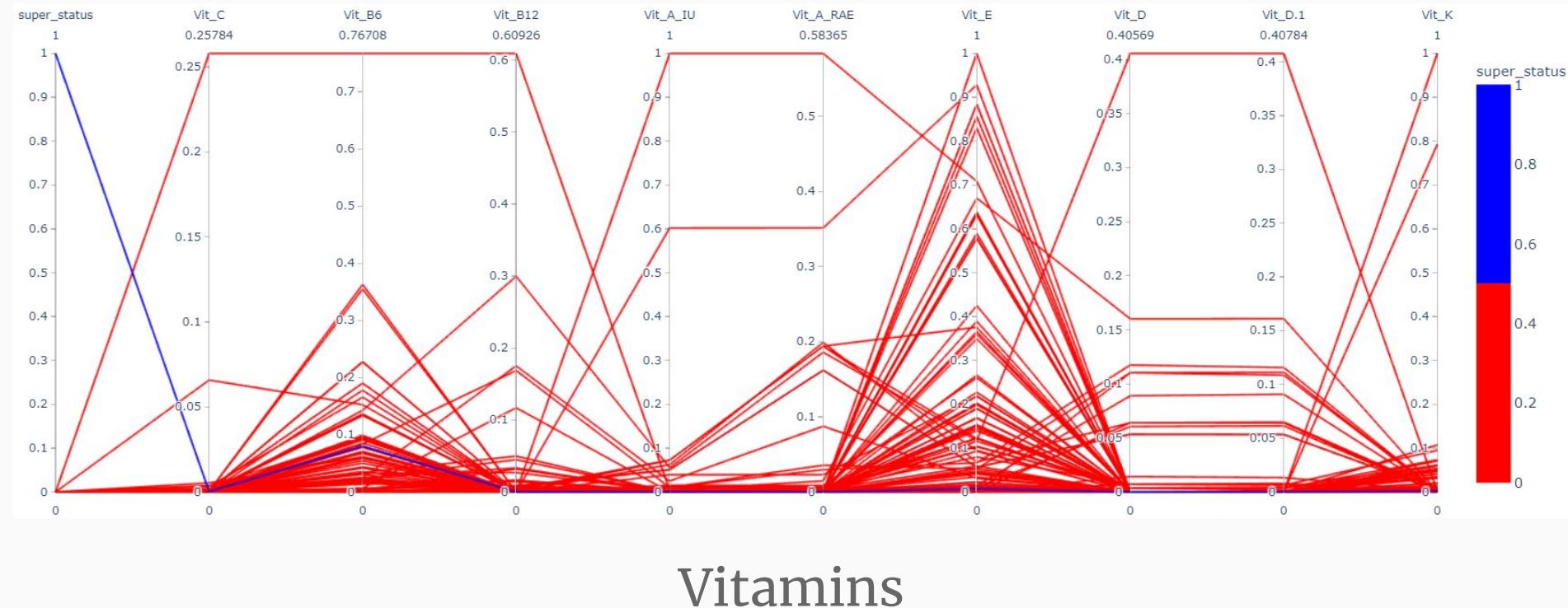
## Superfoods- Flax Seed v.s. Nuts and Seeds Group



## Minerals

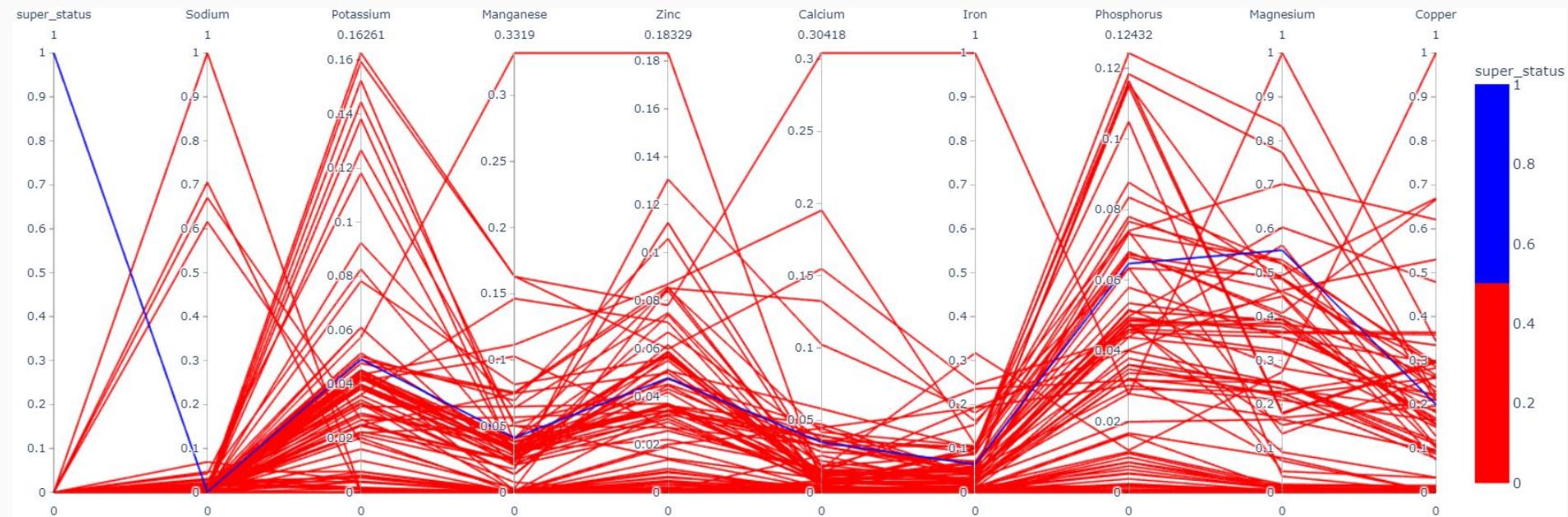
# Extended Analysis: Exploring Superfoods

## Superfoods- Flax Seed v.s. K-Means Data Cluster



# Extended Analysis: Exploring Superfoods

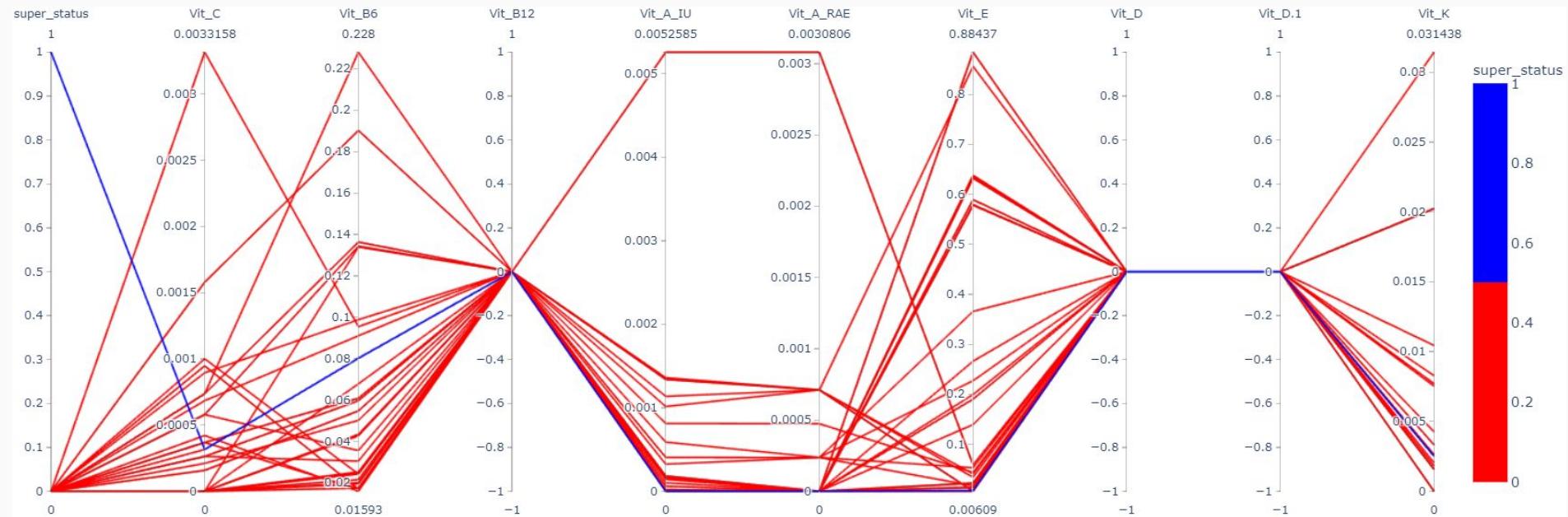
## Superfoods- Flax Seed v.s. K-Means Data Cluster



Minerals

# Extended Analysis: Exploring Superfoods

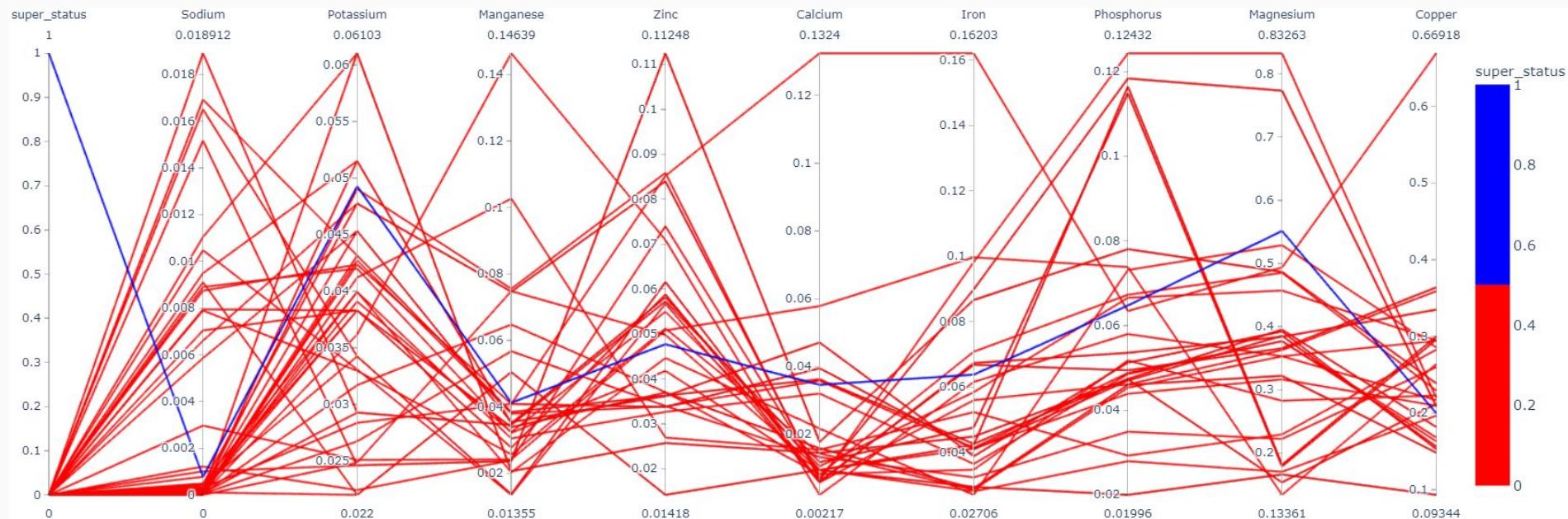
## Superfoods- Flax Seed v.s. Group ∩ Cluster



Vitamins

# Extended Analysis: Exploring Superfoods

## Superfoods- Flax Seed v.s. Group $\cap$ Cluster



## Minerals

# Findings and Implications

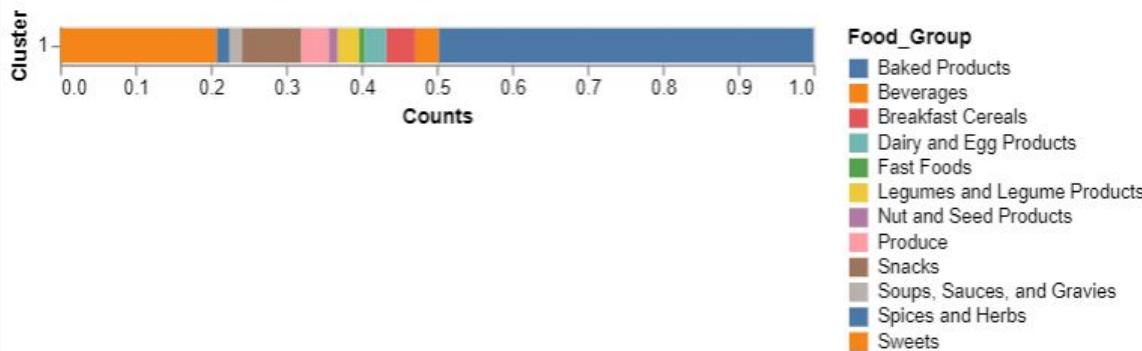
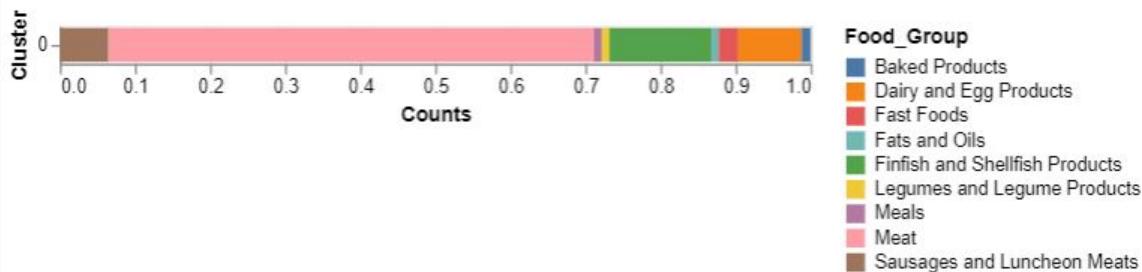
# Findings and Implications

- Food labels that consist of foods with distinct compositions
  - Dairy and Egg, Legumes, Spices and Herbs, Fats and Oils, Cereal and Grains and Pasta
- Macronutrient, vitamins, and minerals
  - Same food outliers- Oatmeal, coconut, pork, etc.
- “Superfoods” were not special according to vitamins and minerals
  - Nutrition facts did better compared to label than cluster
    - Might be “super” given context of labels

Thank you!

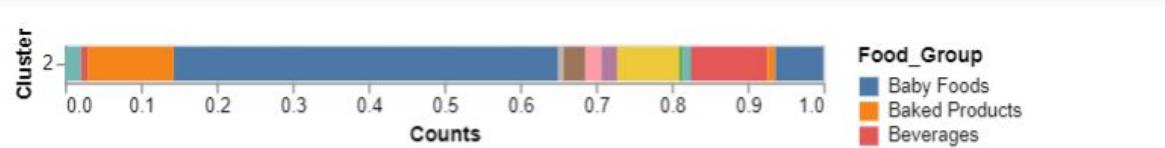
# Question 2: How do these groups compare to grouping by composition?

Distribution of food groups within clusters:



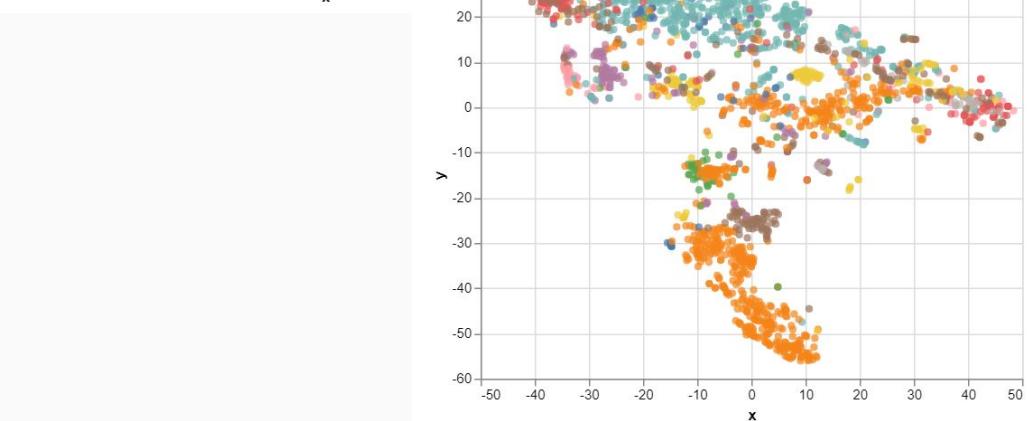
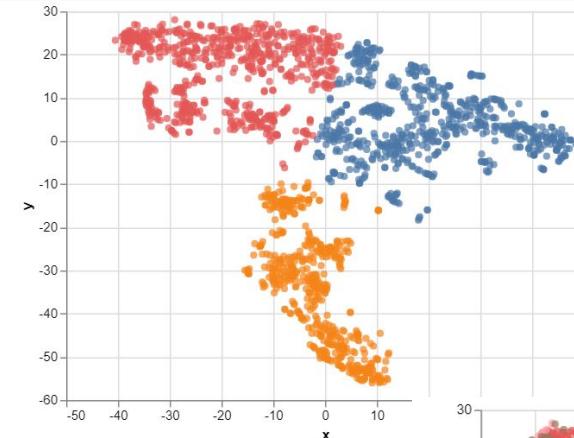
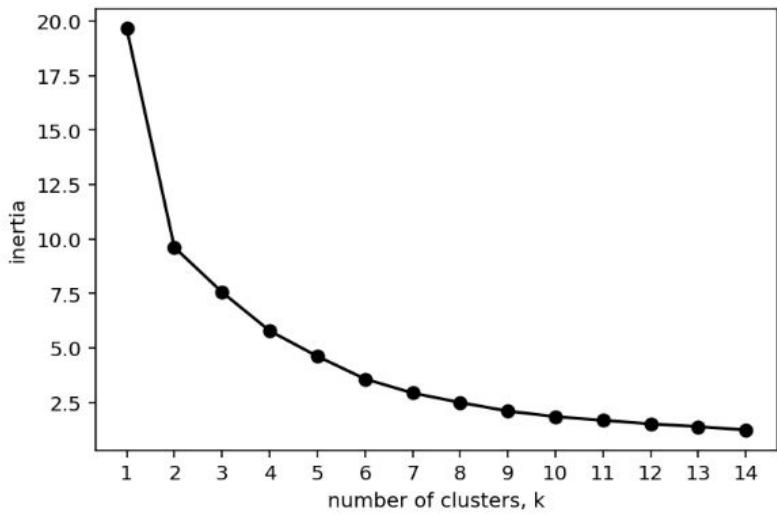
# Question 2: How do these groups compare to grouping by composition?

Distribution of food groups within clusters:

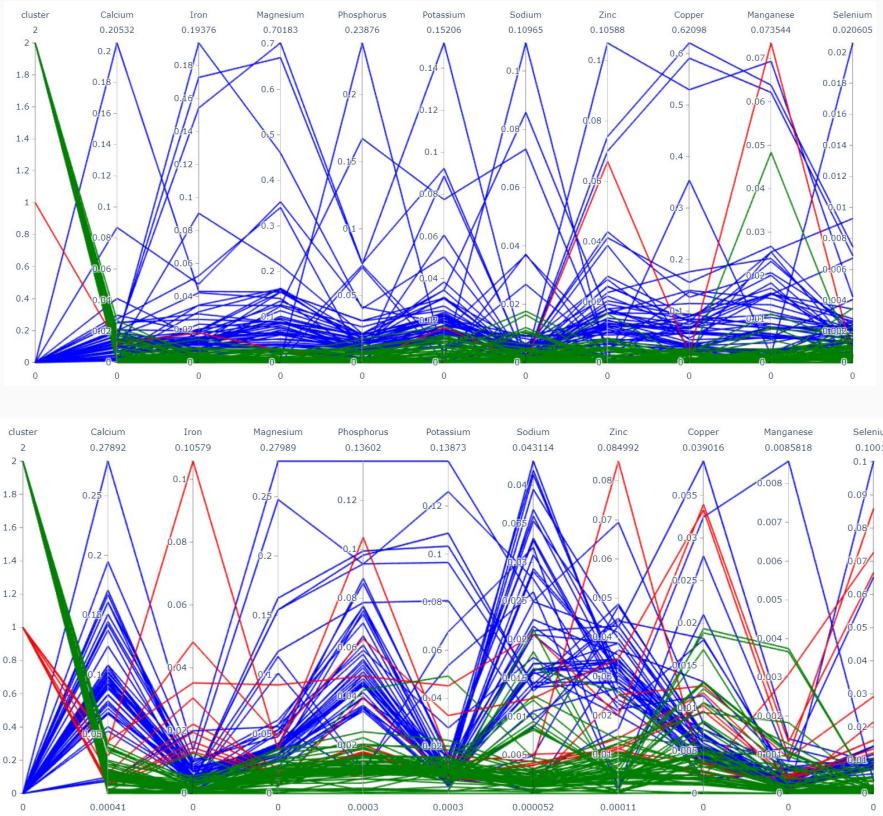
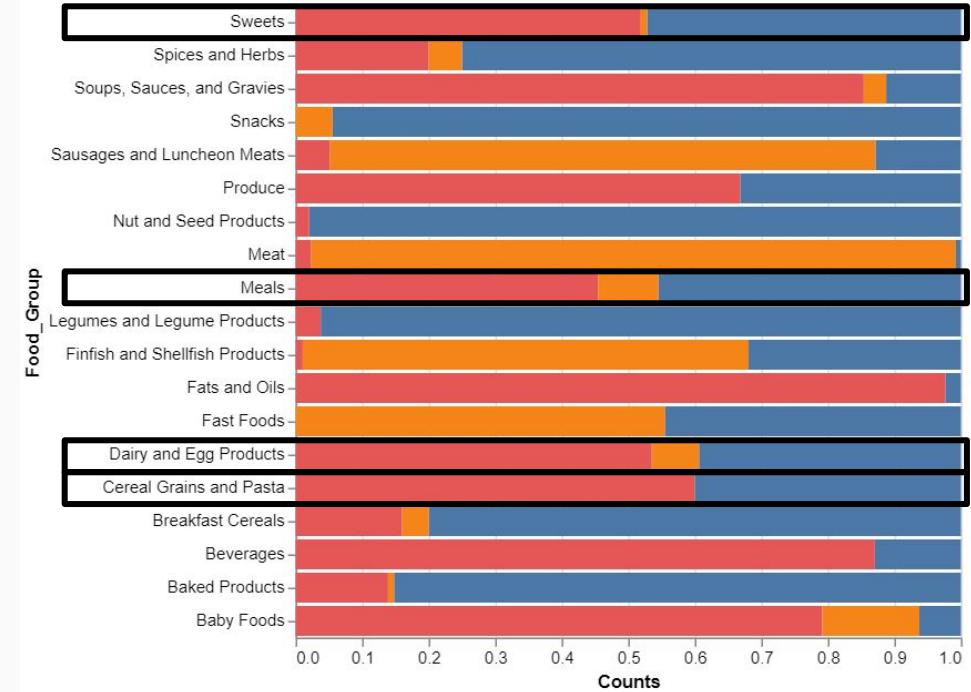


# Question 3: How does this apply to nutrients?

Mineral Subset:



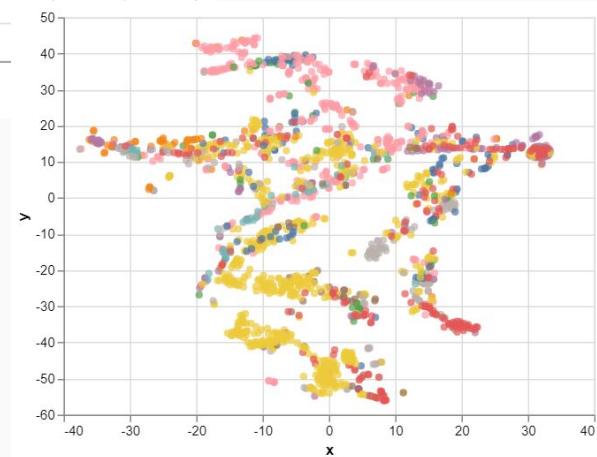
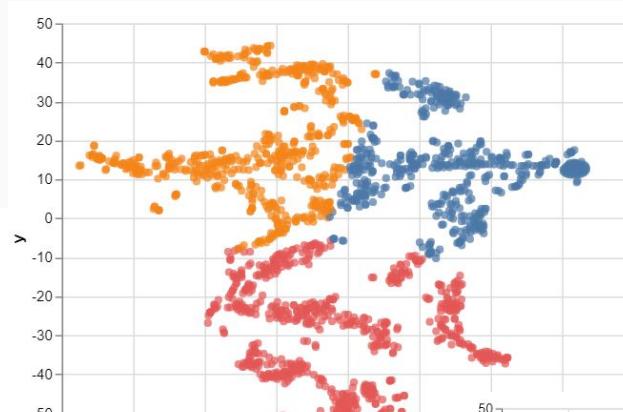
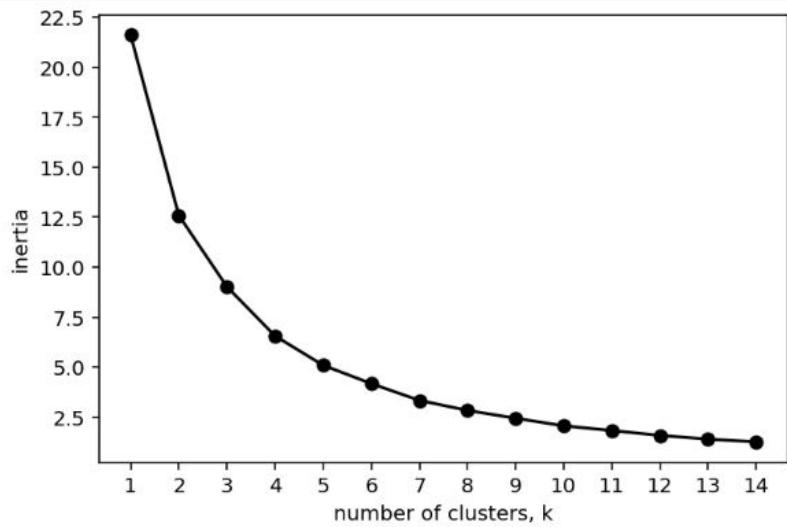
Low  
Dimensional  
K-Means  
Clustering



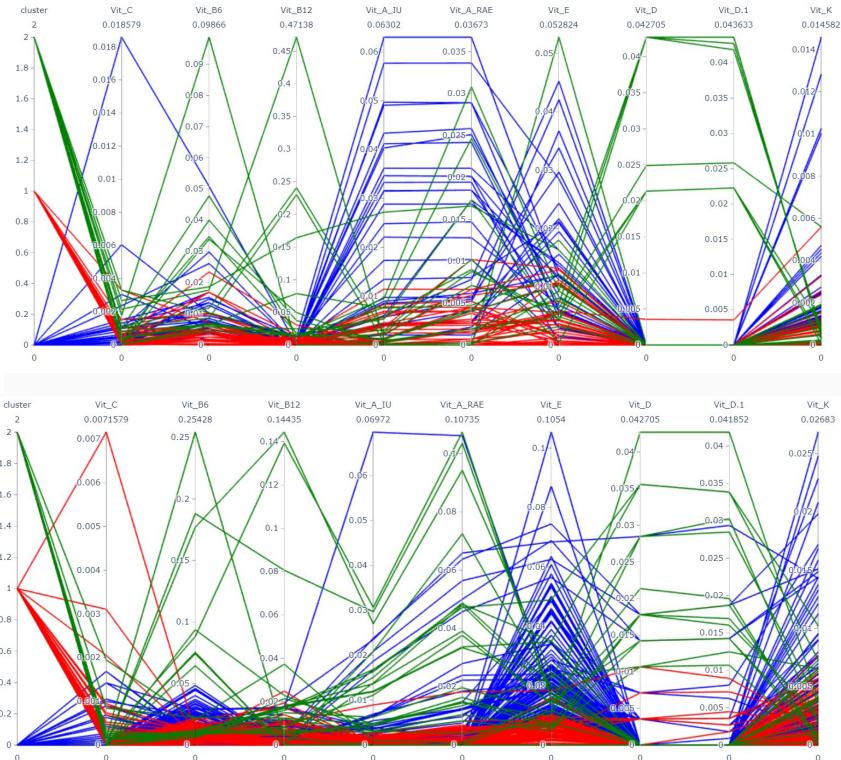
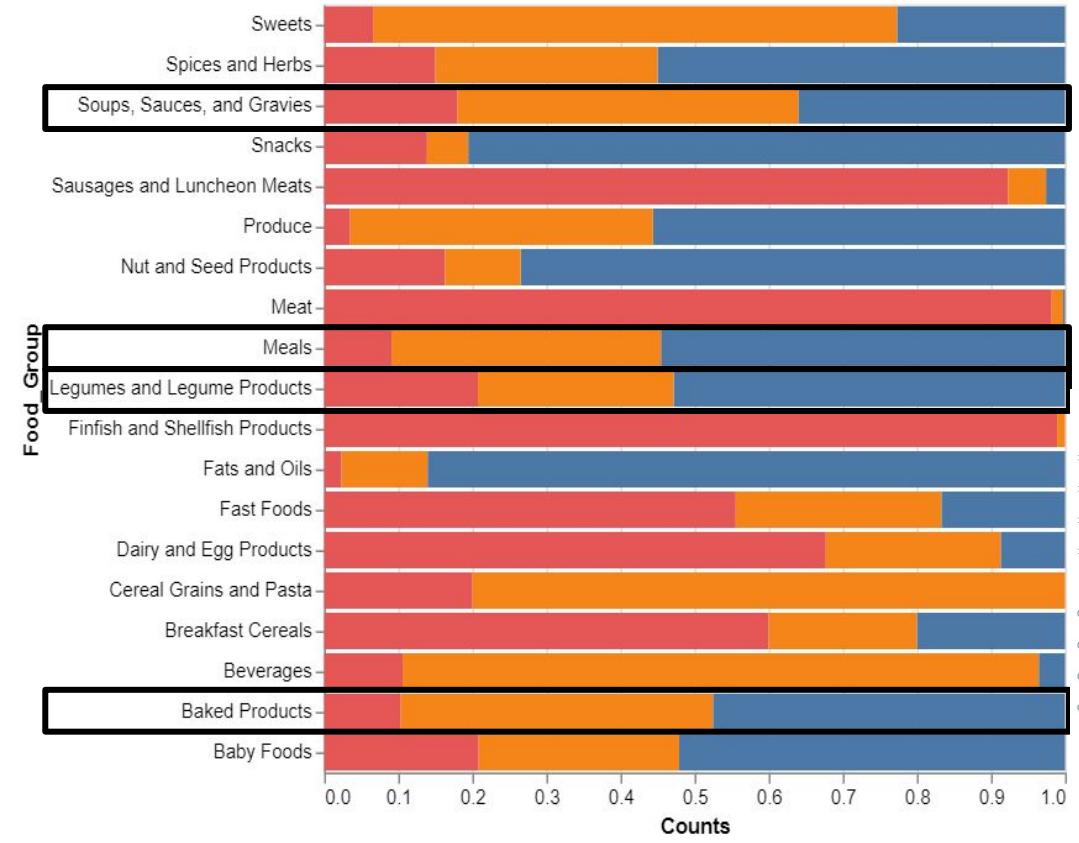
cluster	group	description
78	0 Dairy and Egg Products	WHEY,ACID,DRIED
79	0 Dairy and Egg Products	WHEY,SWEET,DRIED
92	1 Dairy and Egg Products	EGG,WHOLE,DRIED
94	1 Dairy and Egg Products	EGG,YOLK,DRIED
124	0 Dairy and Egg Products	EGG,WHITE,DRIED,STABILIZED,GLUCOSE RED
1605	0 Sweets	BAKING CHOC,UNSWTND,SQUARES
1640	0 Sweets	CANDIES,SESAME CRUNCH
1647	0 Sweets	COCOA,DRY PDR,UNSWTND
1648	0 Sweets	COCOA,DRY PDR,UNSWTND,PROC W/ALKALI
1658	0 Sweets	PUDDINGS,VANILLA,DRY MIX,INST
1673	0 Sweets	DESSERTS,PUDD,COCNT CRM,DRY MIX,INST
1679	1 Sweets	SUGARS,MAPLE
1829	0 Dairy and Egg Products	ENSURE PLUS,LIQ NUTR
1833	0 Sweets	PUDDINGS,CHOC FLAVOR,LO CAL,INST,DRY MIX
1835	0 Sweets	PUDDINGS,ALL FLAVORS XCPT CHOC,LO CAL,INST,DRY...

# Question 3: How does this apply to nutrient?

## Vitamin Subset:



Low  
Dimensional  
K-Means  
Clustering



347	0	Soups, Sauces, and Gravies	SOUP,VEG BF,CND,COND
351	2	Soups, Sauces, and Gravies	SOUP,ONION,DRY,MIX
361	2	Soups, Sauces, and Gravies	SAUCE,HOMEMADE,WHITE,THIN
362	2	Soups, Sauces, and Gravies	SAUCE,HOMEMADE,WHITE,MED
363	2	Soups, Sauces, and Gravies	SAUCE,HOMEMADE,WHITE,THICK
1266	0	Legumes and Legume Products	CHILI WITH BEANS,CANNED
1274	0	Legumes and Legume Products	PEAS,SPLIT,MATURE SEEDS,RAW
1298	2	Legumes and Legume Products	SOYMILK,ORIGINAL & VANILLA,W/ ADDED CA,VITAMIN...
1414	1	Baked Products	CAKE,CHERRY FUDGE W/CHOC FRSTNG
1517	2	Baked Products	PANCAKES PLN,FRZ,RTH (INCLUDES BUTTERMILK)
1531	0	Baked Products	PIE,PUMPKIN,COMMLY PREP
1563	2	Baked Products	WAFFLES,PLN,FRZ,READY -TO-HEAT,TSTD
1572	2	Baked Products	WAFFLE,BTTRMLK,FRZ,RTH,TSTD
1732	0	Meals	EGG ROLLS,PORK,REFR,HTD
1805	1	Meals	RESTAURANT,FAMILY STYLE,FRIED MOZZARELLA STKS