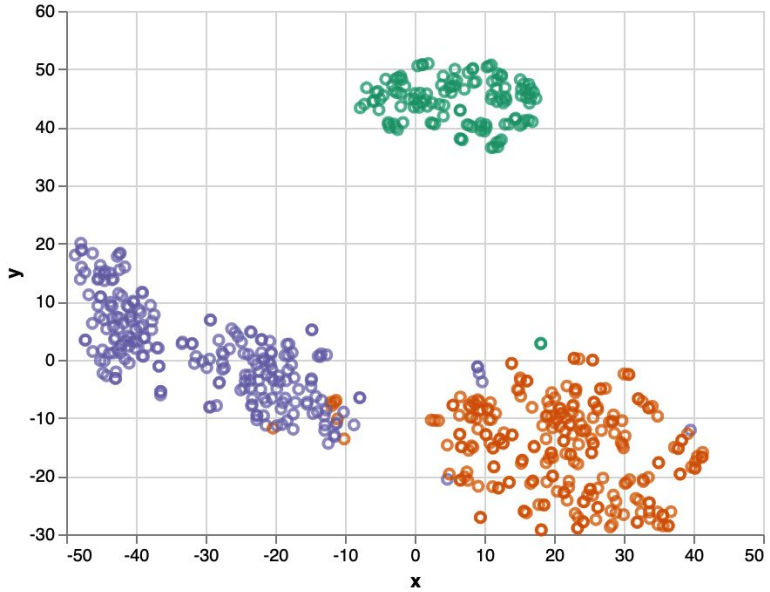
USDA Food Composition

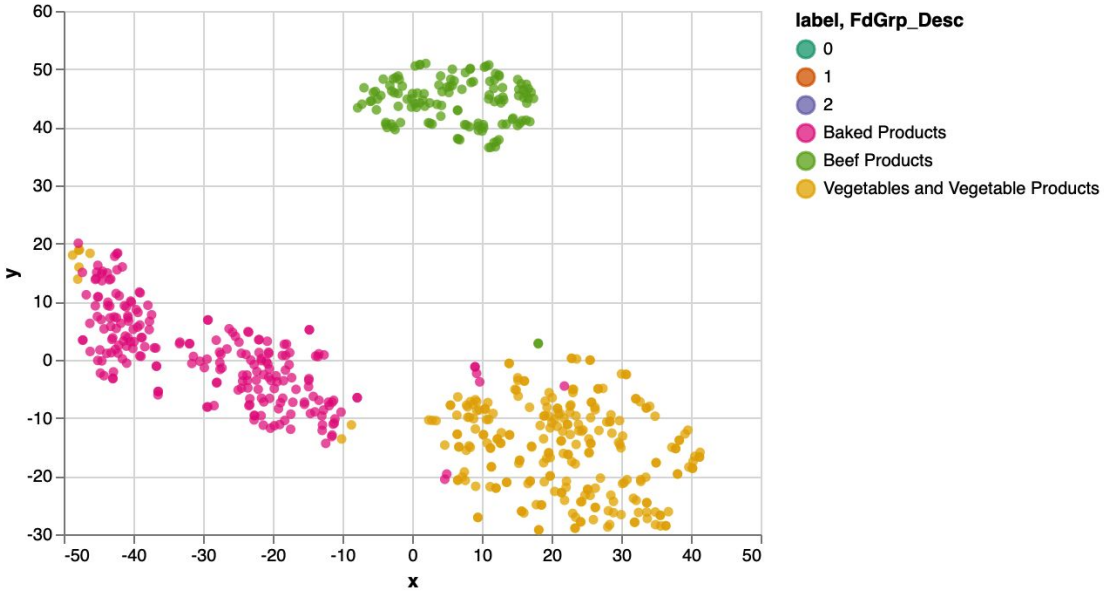
Sunny Yoo
Haohan Zeng
Lynn Li

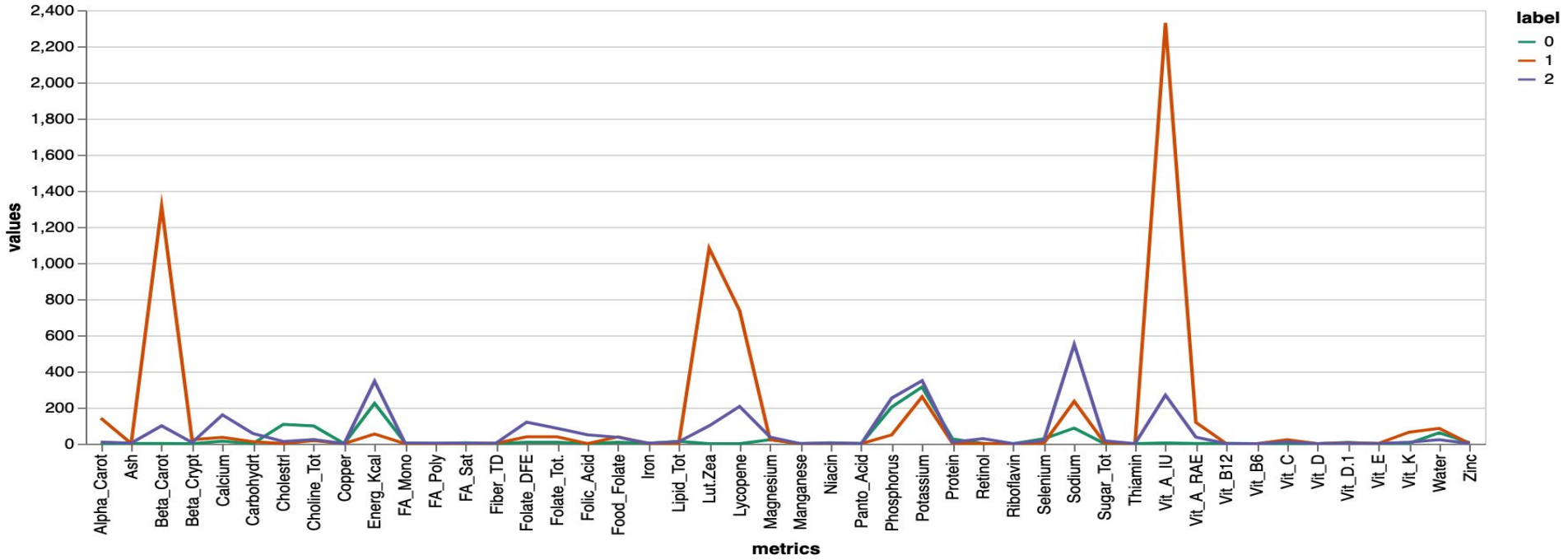
Background

1. We are analyzing the USDA Food Composition data set which include multi-dimensional attributes
2. It includes the statistics of all nutrients, macro-nutritions, vitamins, mineral for each type of food
3. We wanted to avoid having too much dataset and create confusing clusters. Therefore, we reduced the data into three groups. These three groups have most number of data examples and they are all distinct food group to each others in terms of macro-nutrients, vitamins, and minerals which could potentially make better clusters.
    a. Vegetables and Vegetable Products
    b. Baked Products
    c. Beef Products
4. We use TSNE (t-distributed stochastic neighbor embedding) for data dimension reduction and k-means clustering algorithm

All nutritions





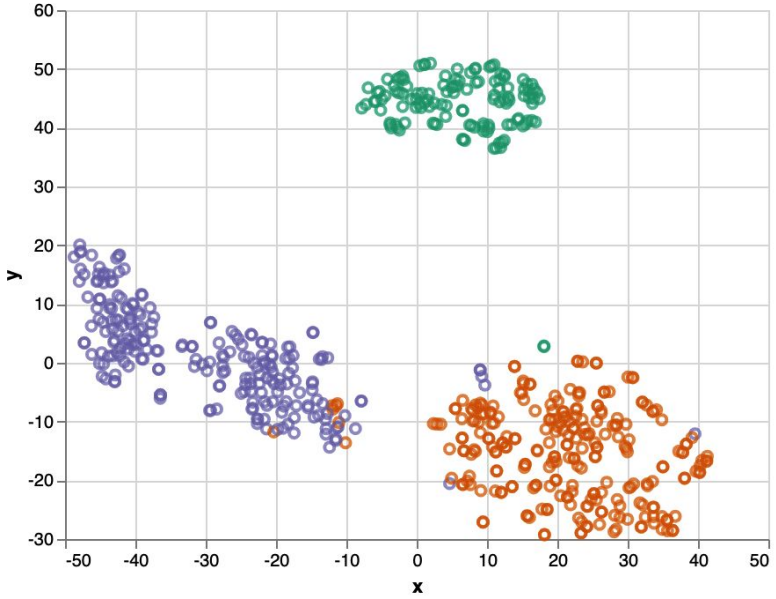


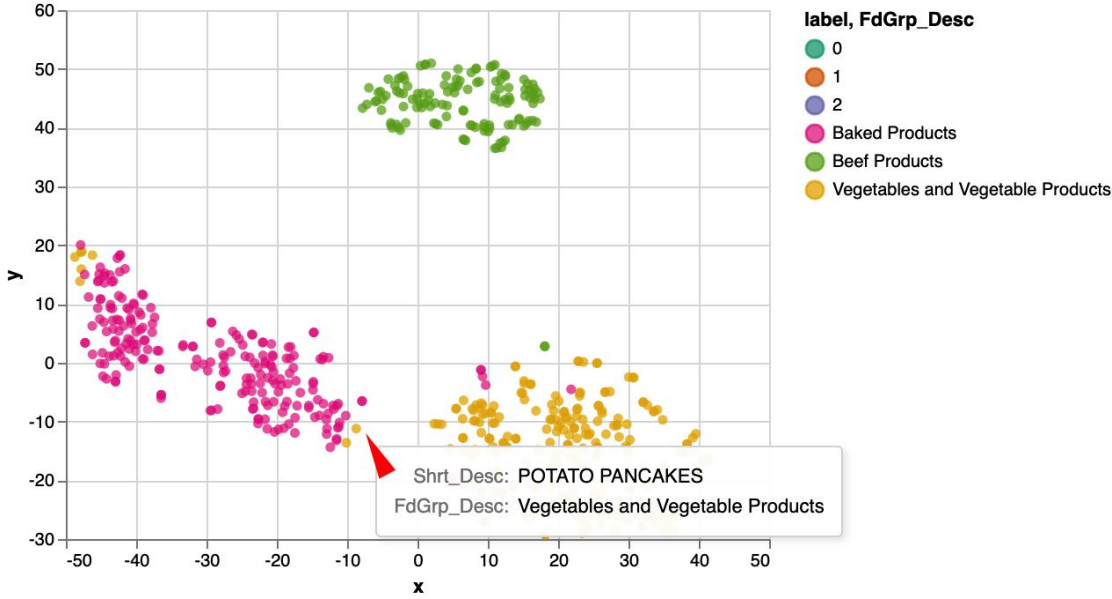
Cluster	High	Low
0	Cholesterol	Vitamin C, Vitamin K, sodium, calcium
1	alpha-carot, Beta-Carot , Lut-Zea and vitamin A, water, vitamin K	Energy-kcal, phosphorous, potassium
2	Sodium and Energy-kcal	water

- Food groups are really distinct to each other
- There are still outliers involved in different clusters such as
    - canned bread
    - fried tortillas
    - potato pancake

  Even they are considered as vegetable product, it fall into the based product cluster
- Several baking leavening agent, such as baking soda, are closer to vegetable
- Beef tripe is also an interesting outlier. Even it’s a beef product, it is closer to vegetable cluster as well.

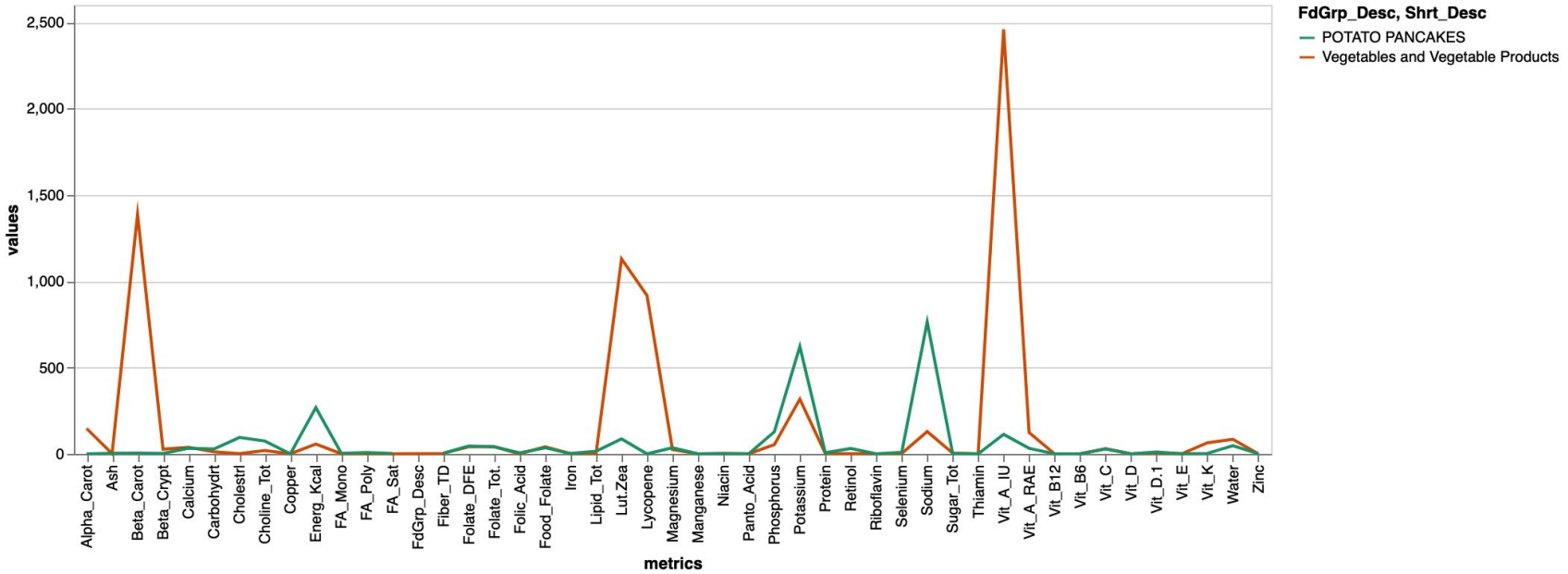
All nutritions





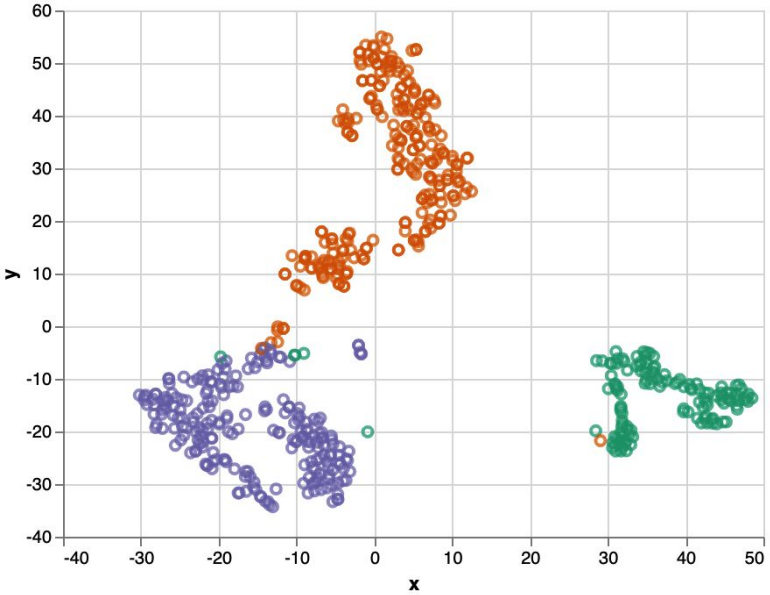
Outlier - Potato Pancakes (Vegetable Product)

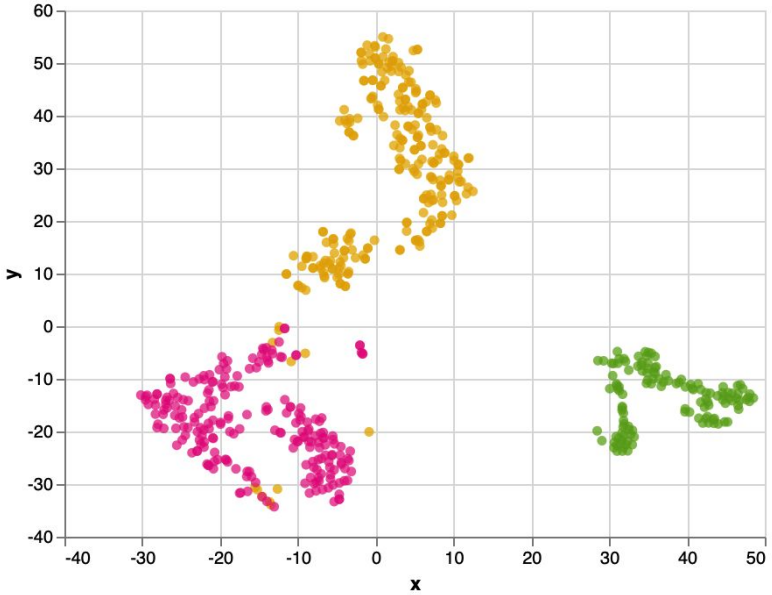
```
food_to_analyze = "POTATO PANCAKES"
create_parallel_cord(all_col)+show_individual_cord(all_col,food_to_analyze)
```

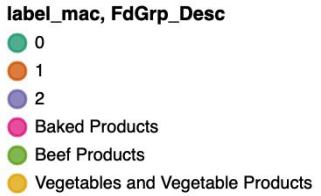


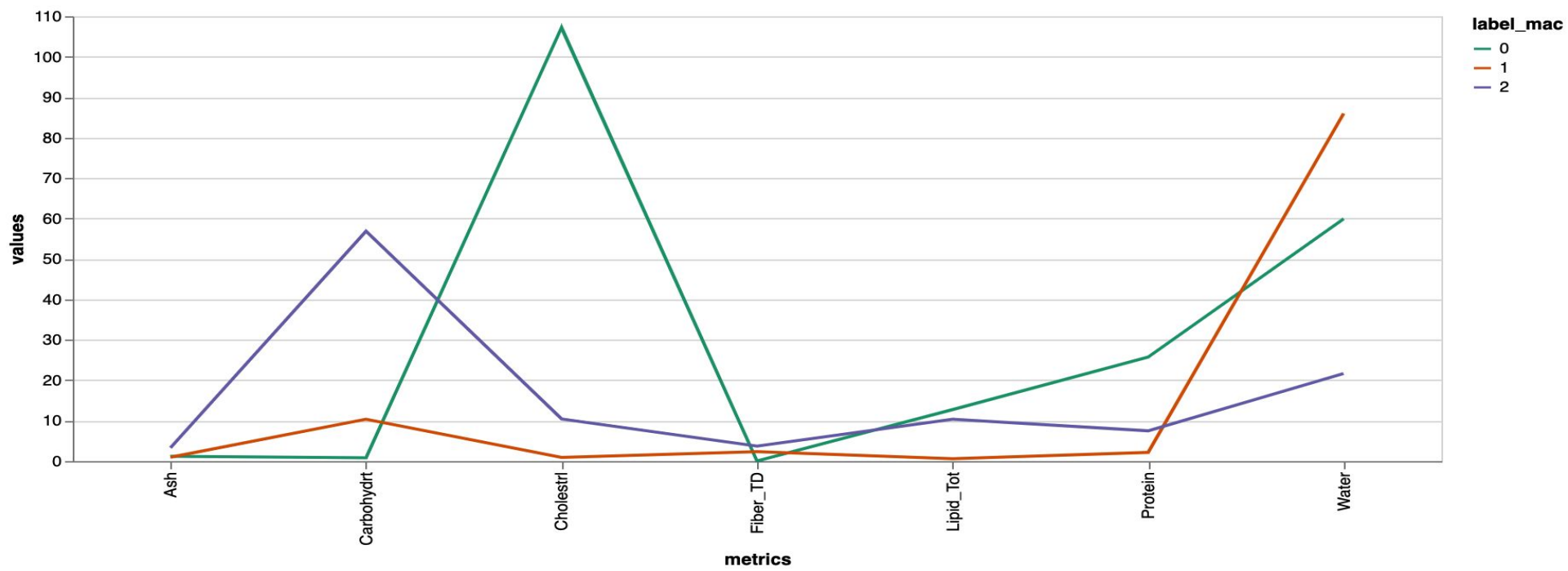
Potato pancake is a vegetable product but seems to be closer to baked product. The reason for this outlier is that potato pancakes have low level of vitamin A, Lut-Zea (Lutein and Zeaxanthin) and Beta_Carot (beta-Carotene) while those three components are the most important characteristics that make vegetable product clusters. The lack of those three components make potato pancake an outlier from its own food group (vegetable).

Macro-nutritions









Cluster	High	Low
0	cholesterol, protein	carbohydrate, fiber
1	water	cholesterol, protein, lipid_tot
2	carbohydrate, ash, fiber	water

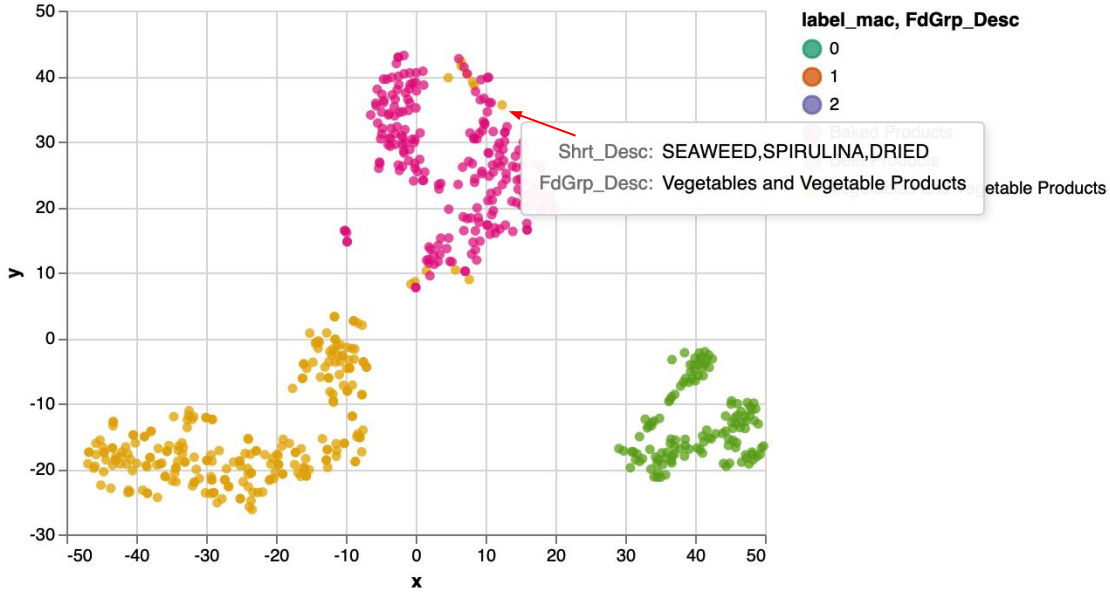
- Macronutrients are very well distributed for each products with distinctive pattern
- One assumption for this distinct cluster is that macronutrients tend to clearly exist more in certain product. As an example, food that falls into beef product should have more proteins than food that falls in to vegetable product.

- There are some vegetables' macronutrients are closer to baked product, such as
  - oven-heated potato
  - egg custard pie

  which makes sense because these products are rich in starch
- However there are some surprising outlier, such as
  - sun dried tomato
  - onion flake
  - dried seaweed

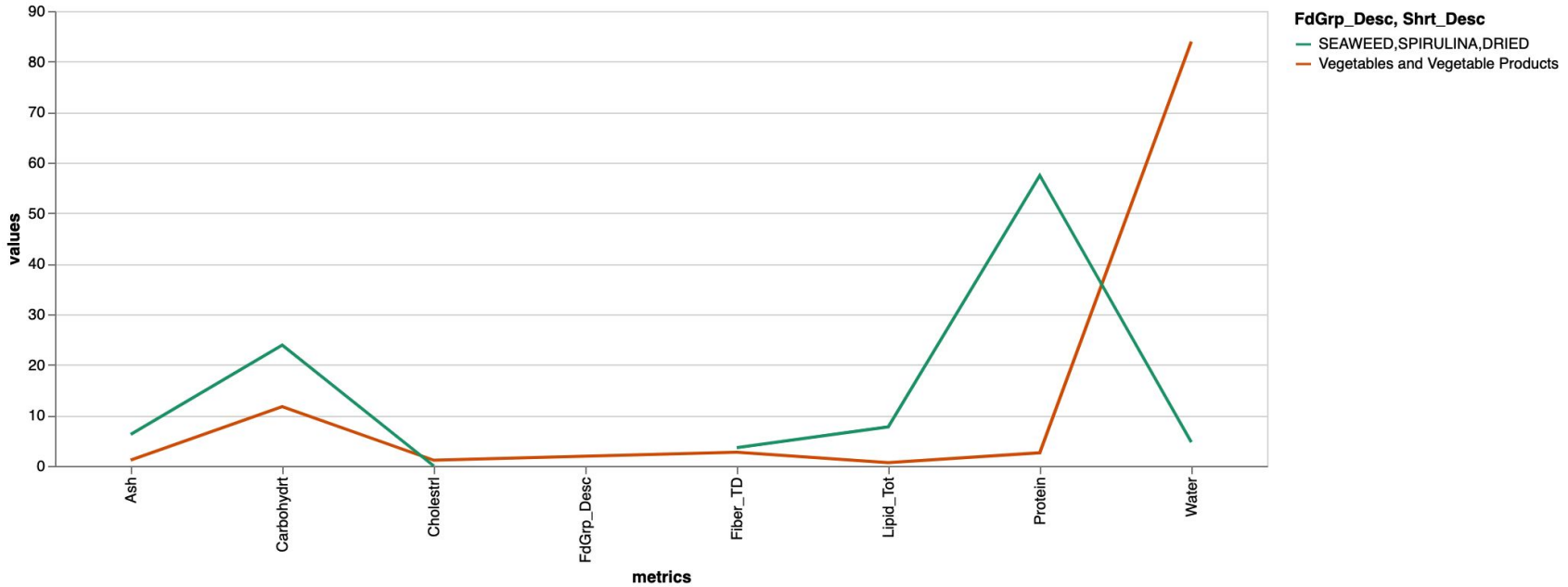
  Which indicates that dehydrated vegetable is closer to baked product in terms of macro-nutrition level . These outliers are also cluster together within the baked product cluster

Outlier - Dried Seaweed (Vegetable Product)



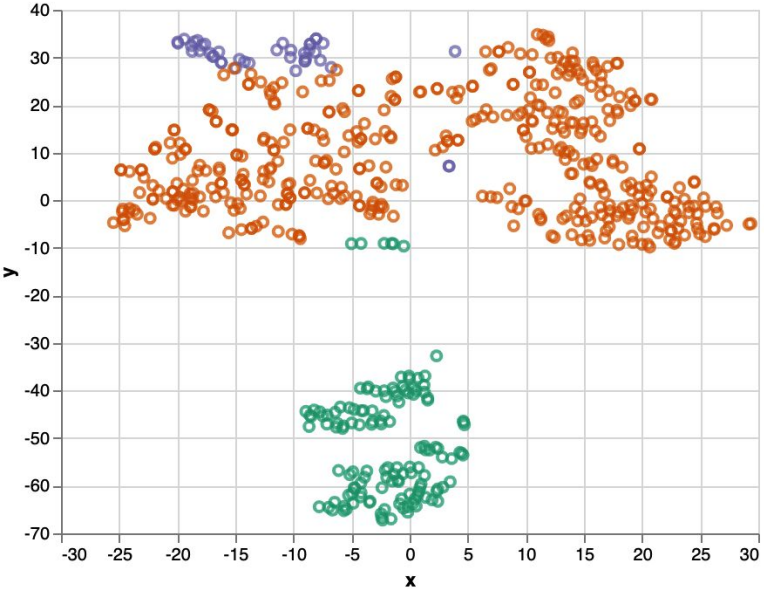
Outlier - Dried Seaweed

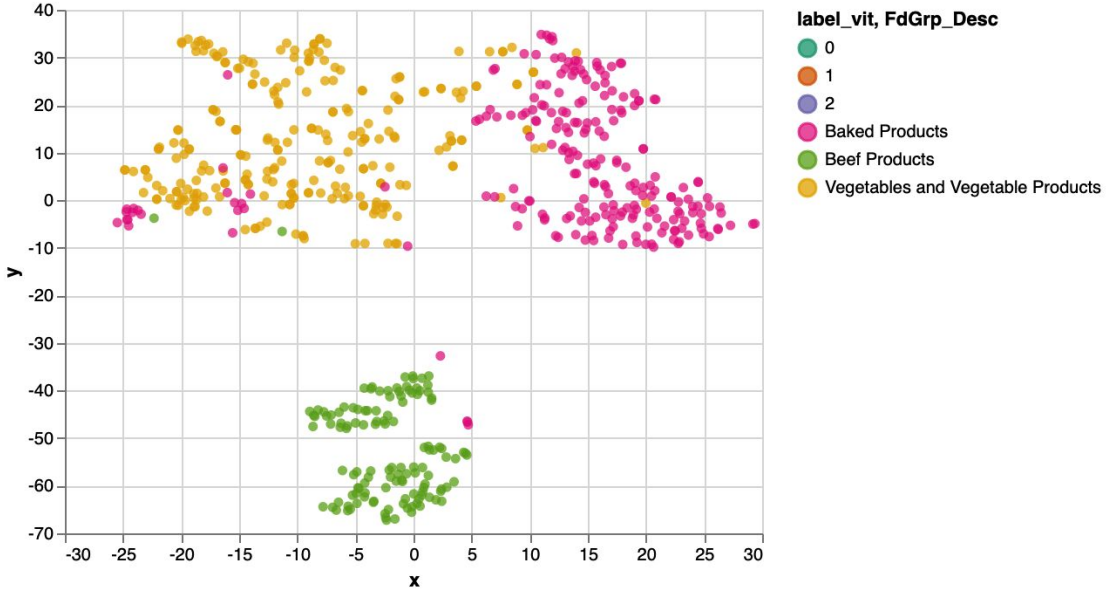
```
food_to_analyze = "SEAWEED,SPIRULINA,DRIED"
create_parallel_cord(macronu_col)+ show_individual_cord(macronu_col,food_to_analyze)
```

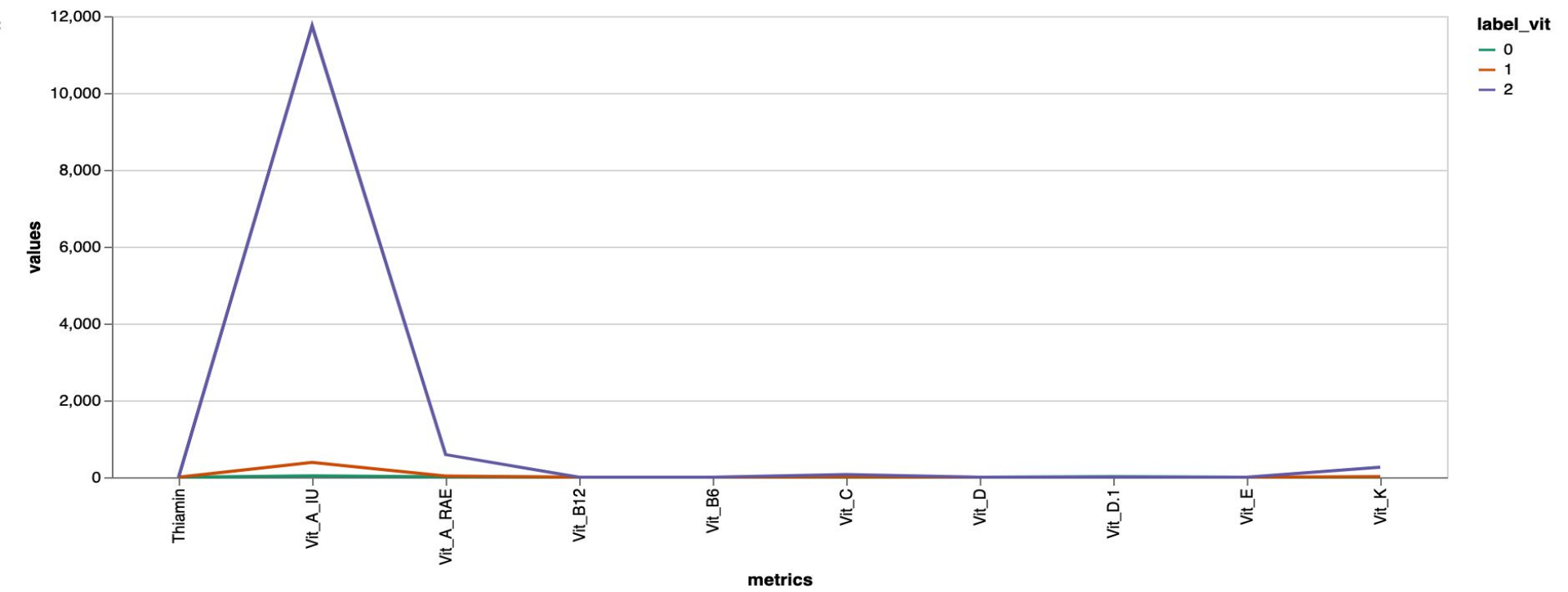


Dried Seaweed is a vegetable product but has been classified closer to baked product. The reason for this outlier is that macronutrients-cluster for vegetable heavily depends on the amount of water. However, we can notice that seaweed have low level of water, which makes seaweed an outlier from its own food group. In addition, dried seaweed has a much higher level of carbohydrate and protein.

Vitamins


y
x
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-30
-25
-20
-15
-10
-5
0
5
10
15
20
25
30


label_vit, FdGrp_Desc
0
1
2
Baked Products
Beef Products
Vegetables and Vegetable Products
y
x
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-30
-25
-20
-15
-10
-5
0
5
10
15
20
25
30

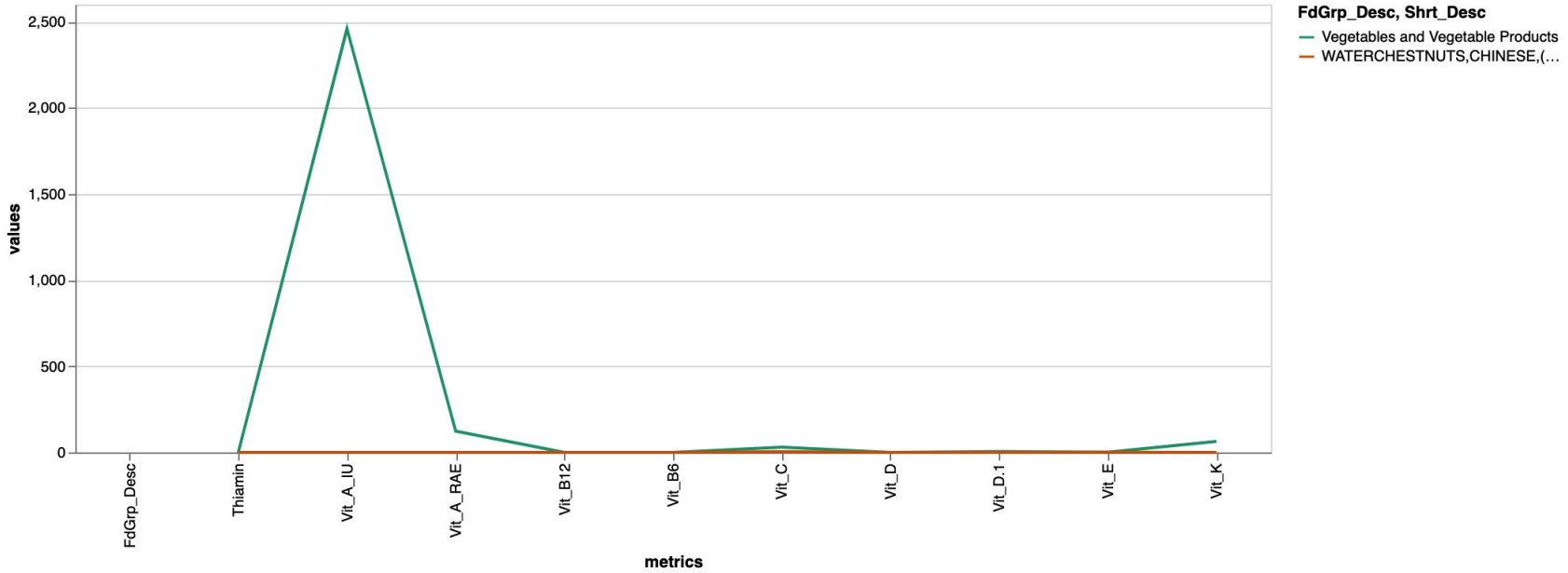


Cluster	High	Low
0	Relatively rich in Vitamin A	Vitamin C, Vitamin D,Vitamin K
1	Nothing significant	Vitamin A, Vitamin C, Vitamin D, Vitamin K
2	Vitamin A, Vitamin K	Vitamin C, Vitamin D

- Baked products and vegetable products have more overlaps
- Beef products have distinct clusters compared to baked and vegetable products
- Outliers: Chinese water chestnut, raw bamboo

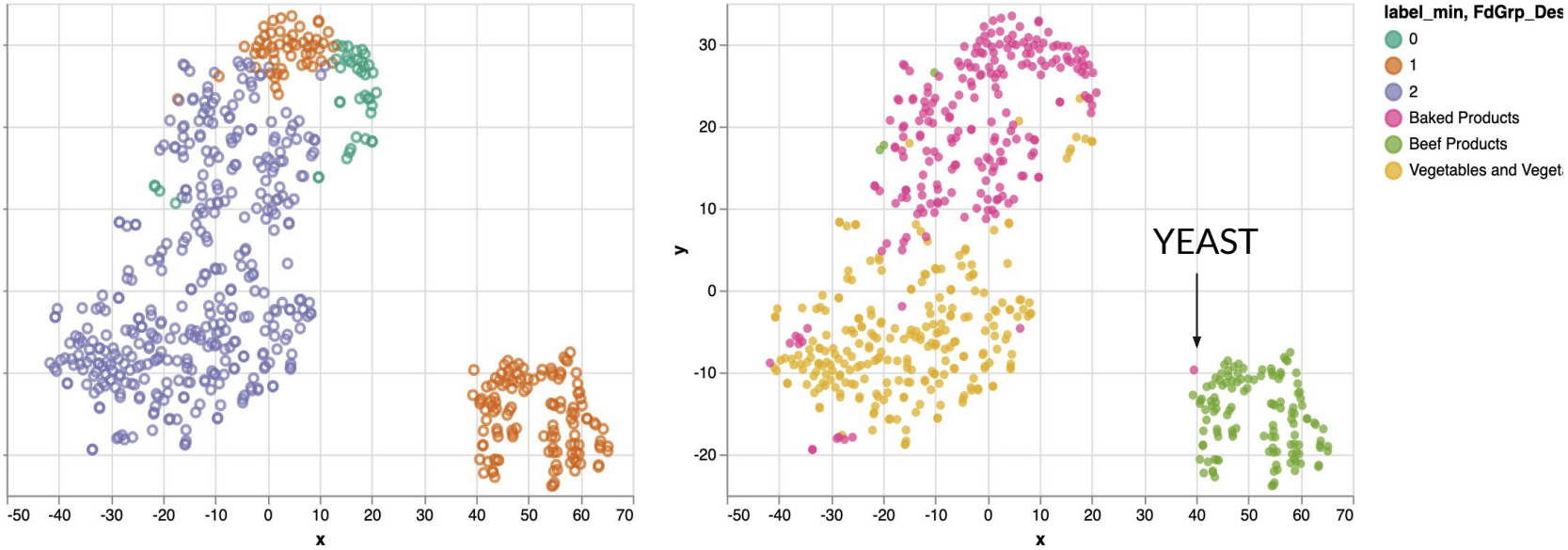
Outlier - Chinese Water Chestnut (Vegetable)

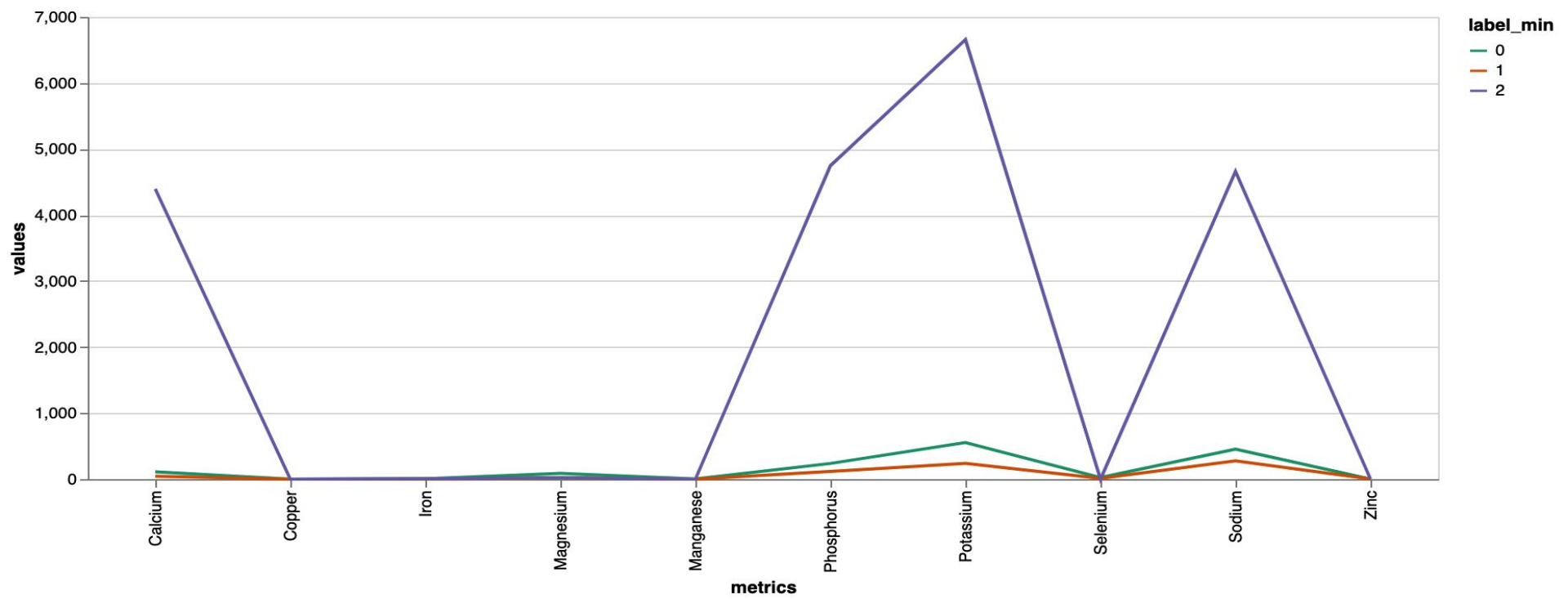
```
vit_food_to_analyze = "WATERCHESTNUTS,CHINESE,(MATAI),RAW"
create_parallel_cord(vitamin_col)+show_individual_cord(vitamin_col,vit_food_to_analyze)
```



Chinese Water Chestnut is vegetable product but has been identified closer baked product. The reason for this outlier is that the vitamin-clusters for vegetable product depends heavily on Vitamin A, Vitamin K, and Vitamin C. However, Chinese Water Chestnut lack these most important components. Therefore, it has been identified as an outlier from its own food group.

Minerals


label_min, FdGrp_Des
0
1
2
Baked Products
Beef Products
Vegetables and Veget
YEAST
x
y
-50 -40 -30 -20 -10 0 10 20 30 40 50 60 70
30 20 10 0 -10 -20

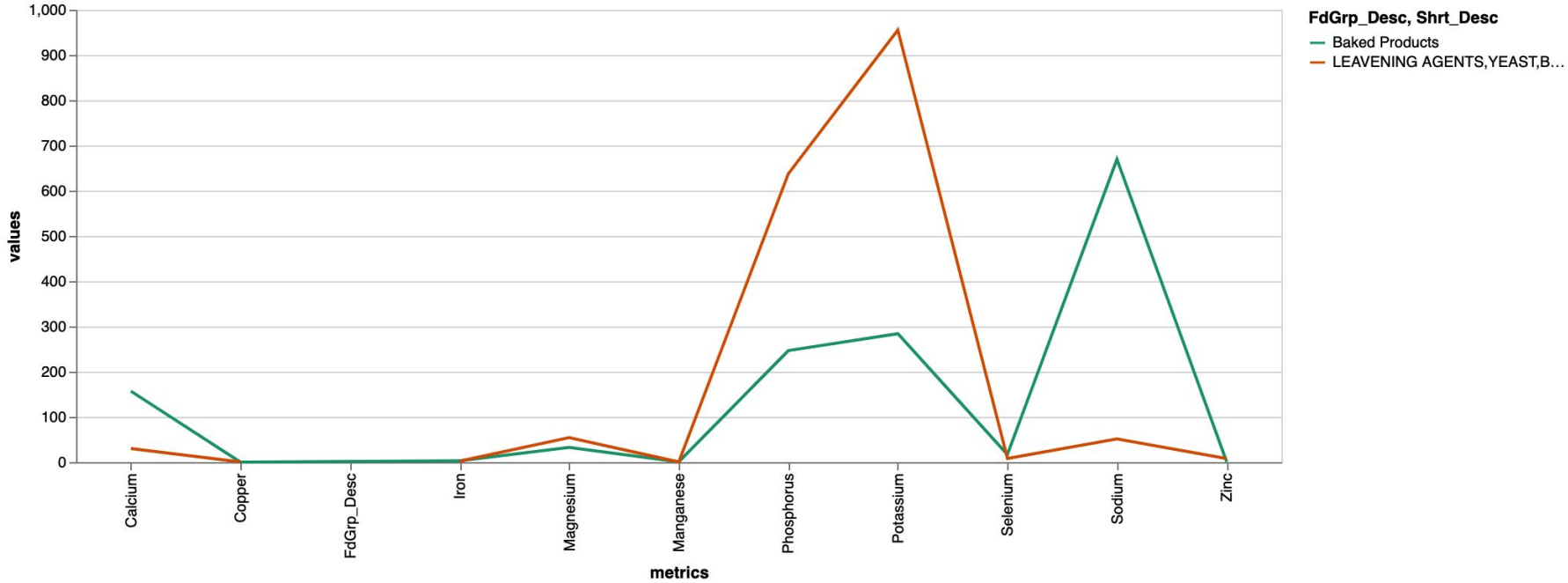


Cluster	High	Low
0	Relatively high in Sodium, Phosphorus	Relatively low in Calcium
1		Sodium, Calcium, Potassium, Phosphorus
2	high in Sodium, Phosphorus, Potassium	

- There still exist lots of overlaps between vegetable products and baked products
  - Using mineral to group foods into baked products and vegetables will be difficult
- Outliers include yeast, which has similar mineral level as beef products
- Beef chuck also has similar mineral level as baked product

Outlier - Yeast (Baked Product)

```
food_to_analyze = "LEAVENING AGENTS,YEAST,BAKER'S,ACTIVE DRY"
create_parallel_cord(mineral_col) + show_individual_cord(mineral_col,food_to_analyze)
```



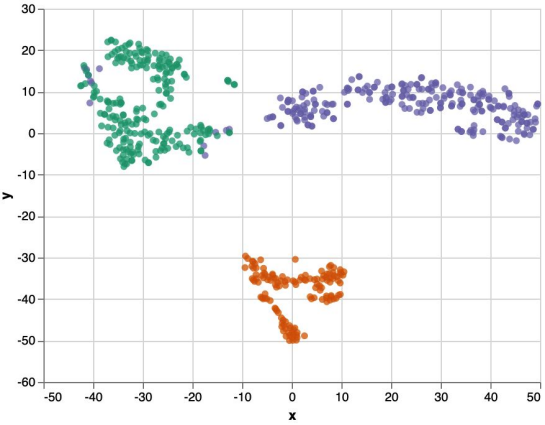
Yeast is baked product but has been classified closer to beef product mineral-cluster. The reason for this outlier is that the mineral-cluster for baked product heavily depends on the amount of Calcium, Potassium, Phosphorus, and Sodium. However, Yeast have low level of Calcium and Sodium while having too high level of Phosphorus and Potassium, therefore, it has been classified as an outlier.

Our Question

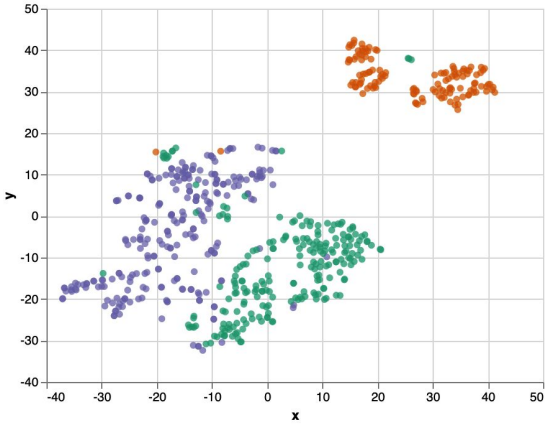
Which food characteristics (macronutrients, vitamins, minerals) show the least overlapping?

Our Answer

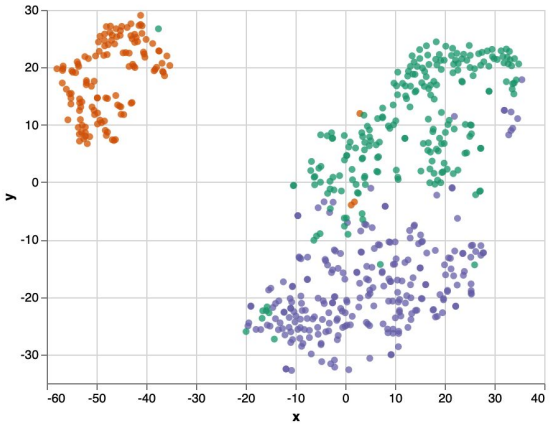
Out of all three food groups, macro-nutritions shows the most distinctive distribution while vitamins and minerals showed more overlapping. However, each three characteristics does show distinct distributions. Therefore, when we try to group food according to their components, we should consider all of those characteristics .



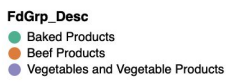
macro-nutritions



vitamin

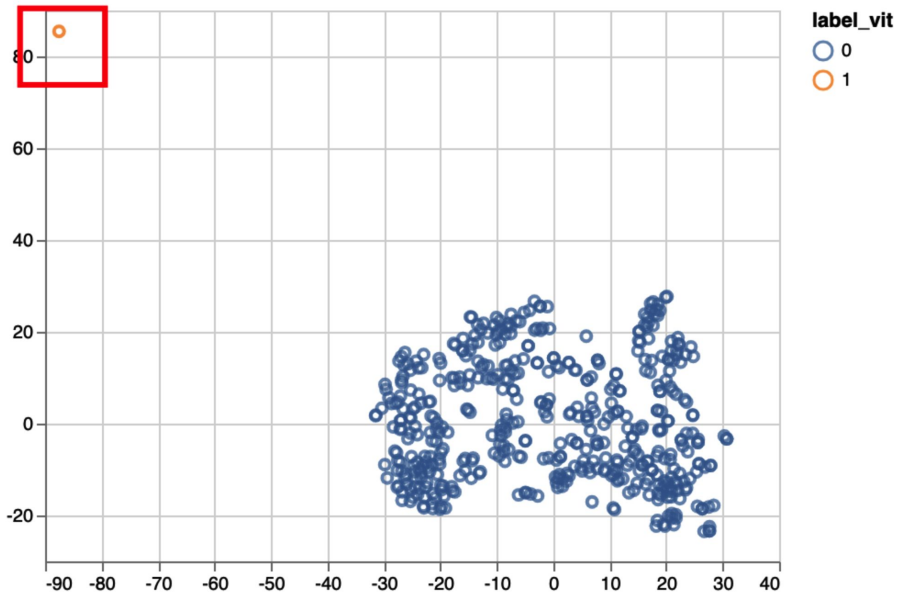


mineral



Summary

From all the charts above, we can see the clusters generated by K-Means, which group the data with their mathematical distance are very similar to the clusters grouped by their actual category. However, K-means clustering is sensitive to outliers and it is easy to generate imbalanced clustering result.



Summary

Despite the distinct pattern, there are indeed some overlaps and outliers. However we can confidently say that majority of the food in their own food group has its very unique nutrition pattern. What's interesting about outliers is that, it can provide some guidance for people when they choose their meal. For example, for dried product such as dried pepper, sun dried tomato, dried seaweed, and certain mushroom such as shitake mushrooms, most people consuming them thinking they are vegetables. However in terms of nutrition, they are closer to the baked product group. People have this knowledge in mind would make more informed decisions.