



A study of Nutrition & Alternative Foods

Team 14

Conor Doyle, Heidi Eren, Julia Luo, Justin Wang

Motivation

- Food nutrition is often difficult to understand
- People on special diets (religious/dietary) often have a had hard time getting their necessary nutrition
- We wanted to find a way to provide alternative foods that contain similar nutritional values to standard foods



The background of the slide is decorated with various colorful, stylized illustrations of fruits and vegetables. These include a blueberry cluster, an orange slice, a green kiwi, a watermelon slice, a lime, a lemon, a strawberry, a banana, a green apple, a cherry, and an orange. The illustrations are scattered around the central text.

17%

of Americans follow some sort
of special diet

<https://www.cdc.gov/nchs/products/databriefs/db389.htm>

Data Source

	name	serving_size	calories	total_fat	saturated_fat	cholesterol	sodium	choline	folate	...	fat	saturated_fatty_acids	monounsaturated_fatty_acids
0	Cornstarch	100 g	381	0.1g	NaN	0	9.00 mg	0.4 mg	0.00 mcg	...	0.05 g	0.009 g	0.016 g
1	Nuts, pecans	100 g	691	72g	6.2g	0	0.00 mg	40.5 mg	22.00 mcg	...	71.97 g	6.180 g	40.801 g
2	Eggplant, raw	100 g	25	0.2g	NaN	0	2.00 mg	6.9 mg	22.00 mcg	...	0.18 g	0.034 g	0.016 g
3	Teff, uncooked	100 g	367	2.4g	0.4g	0	12.00 mg	13.1 mg	0	...	2.38 g	0.449 g	0.589 g
4	Sherbet, orange	100 g	144	2g	1.2g	1mg	46.00 mg	7.7 mg	4.00 mcg	...	2.00 g	1.160 g	0.530 g

- We needed the nutritional facts for a lot of different types of food.
- The dataset above contains more than 8,700 foods
- Includes nutritional facts like protein, vitamins, calories, carbs, etc.
- Foods are all standardized to serving size of 100 grams.

Data Cleaning

- Replaced NaN values with 0
- Normalized scale of nutritional values into the same units (milligrams)
- Dropped unnecessary columns and strings (units)
- Grouped food items by food categories (first word in food item name)

```
# loop through every column in df
from copy import copy

# create dictionary for unit scalar values
scales = {"g": 1000,
          "mg": 1,
          "mcg": .001,
          "iu": .5,
          "": 1}
```

```
for column in df_nutrition.columns:

    # check for incompatible columns
    if column == "name":
        continue

    # fill the NaN values with 0 and initialize list to store new converted values
    df_nutrition = df_nutrition.fillna("0g")
```

```
new_col_values = []

# loop through each row in column
for nutrient in df_nutrition.loc[:,column]:

    # skip non-ints
    if type(nutrient) != int:
        new_col_values.append(nutrient)
        continue

    # index loop through each value which is a string
    for i in range(len(nutrient)):

        # find unit
        unit = ""
        if nutrient[i:] in scales:
            unit = nutrient[i:]
            number = nutrient[:i]
            number = float(number)
            break

    # find conversion value from scales and convert to a new value
    scale = scales[unit]
    new_num = number * scale

    # add converted value to list
    new_col_values.append(new_num)

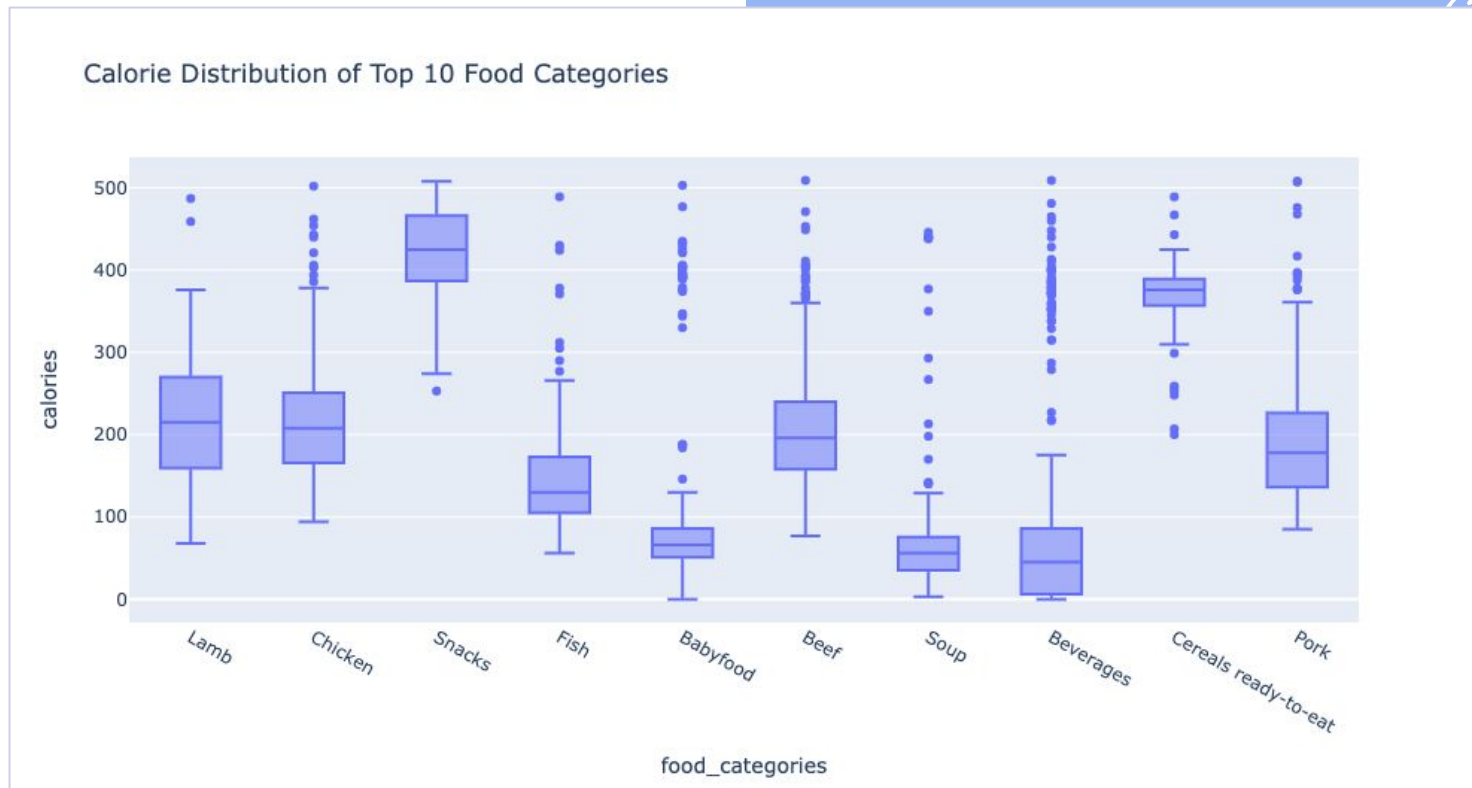
# set converted values
if len(new_col_values) != 0:
    new_values = pd.Series(new_col_values)
    df_nutrition.loc[:,column] = new_values
```

```
# creates a new column "food categories"
df_x_sn['food_categories'] = df_x_sn['name'].apply(lambda x: x.split(',')[0])
```

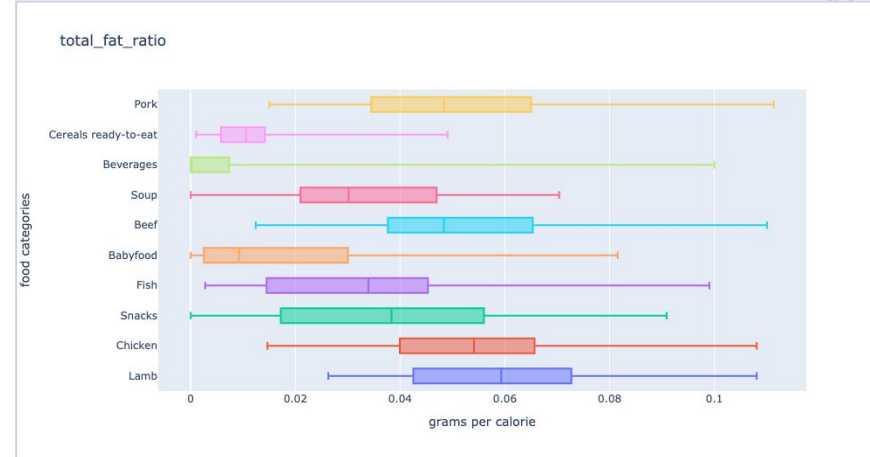
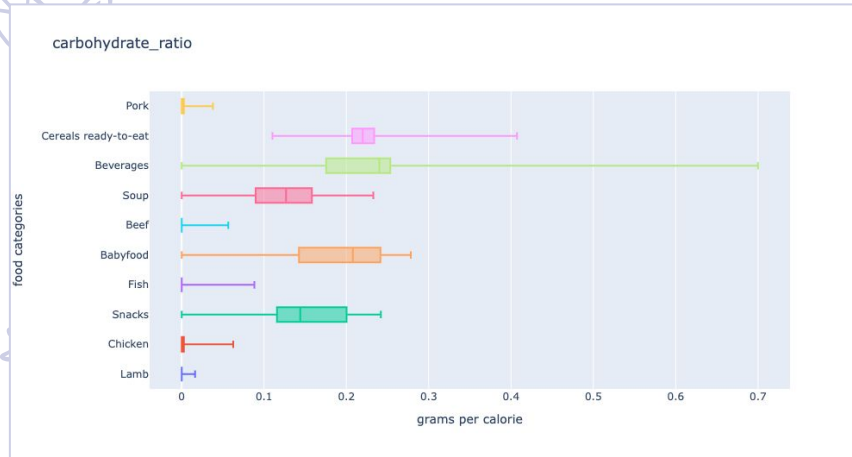
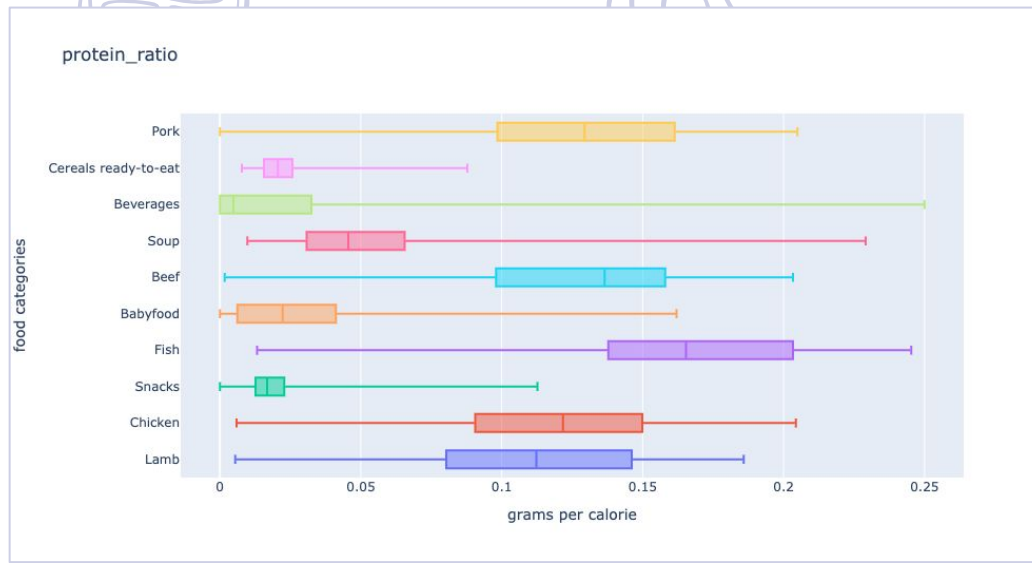
	food_categories	name	serving_size	calories	total_fat	saturated_fat	cholesterol	sodium	choline	folate	...	fat	saturated_fatty_acids
0	Cornstarch	Cornstarch	100000.0	381	100.0	0.0	1.1	9.0	0.4	0.000	...	50.0	9.0
1	Nuts	Nuts, pecans	100000.0	691	72000.0	6200.0	1.1	0.0	40.5	0.022	...	71970.0	6180.0
2	Eggplant	Eggplant, raw	100000.0	25	200.0	0.0	1.1	2.0	6.9	0.022	...	180.0	34.0
3	Teff	Teff, uncooked	100000.0	367	2400.0	400.0	1.1	12.0	13.1	22.000	...	2380.0	449.0
4	Sherbet	Sherbet, orange	100000.0	144	2000.0	1200.0	1.0	46.0	7.7	0.004	...	2000.0	1160.0
...
8784	Beef	Beef, raw, all grades, trimmed to 0" fat, sepa...	100000.0	125	3500.0	1400.0	62.0	54.0	64.5	0.004	...	3500.0	1353.0
8785	Lamb	Lamb, cooked, separable lean only, composite o...	100000.0	206	8900.0	3900.0	109.0	50.0	64.5	0.000	...	8860.0	3860.0

Cleaned data!

Visualizations



Macronutrients



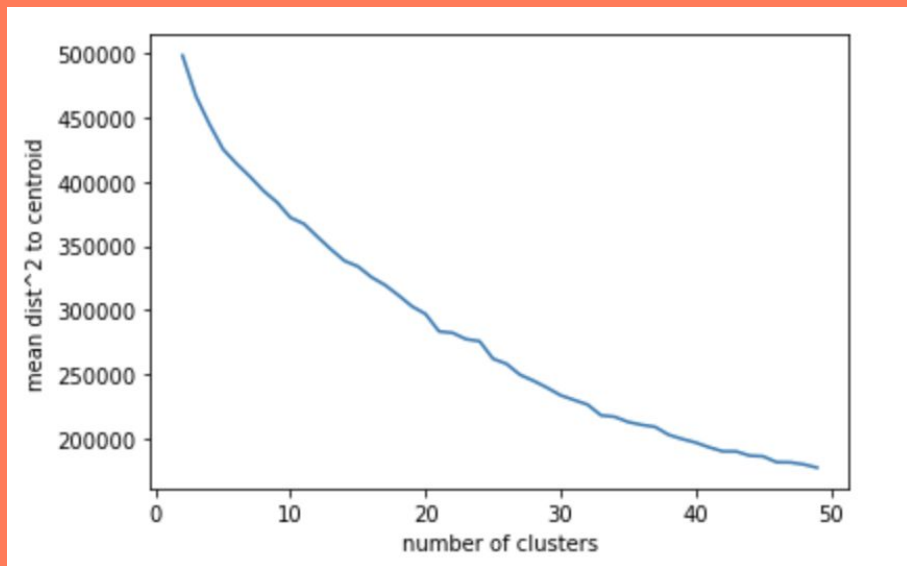
Machine Learning - Clustering

- After PCA and Scale Normalization, we clustered foods with similar nutritional value
- Goal:
 - Observe categorical variance within each cluster
 - Different food categories within each cluster
 - Given a food, recommend alternative foods in same cluster



How many clusters to use?

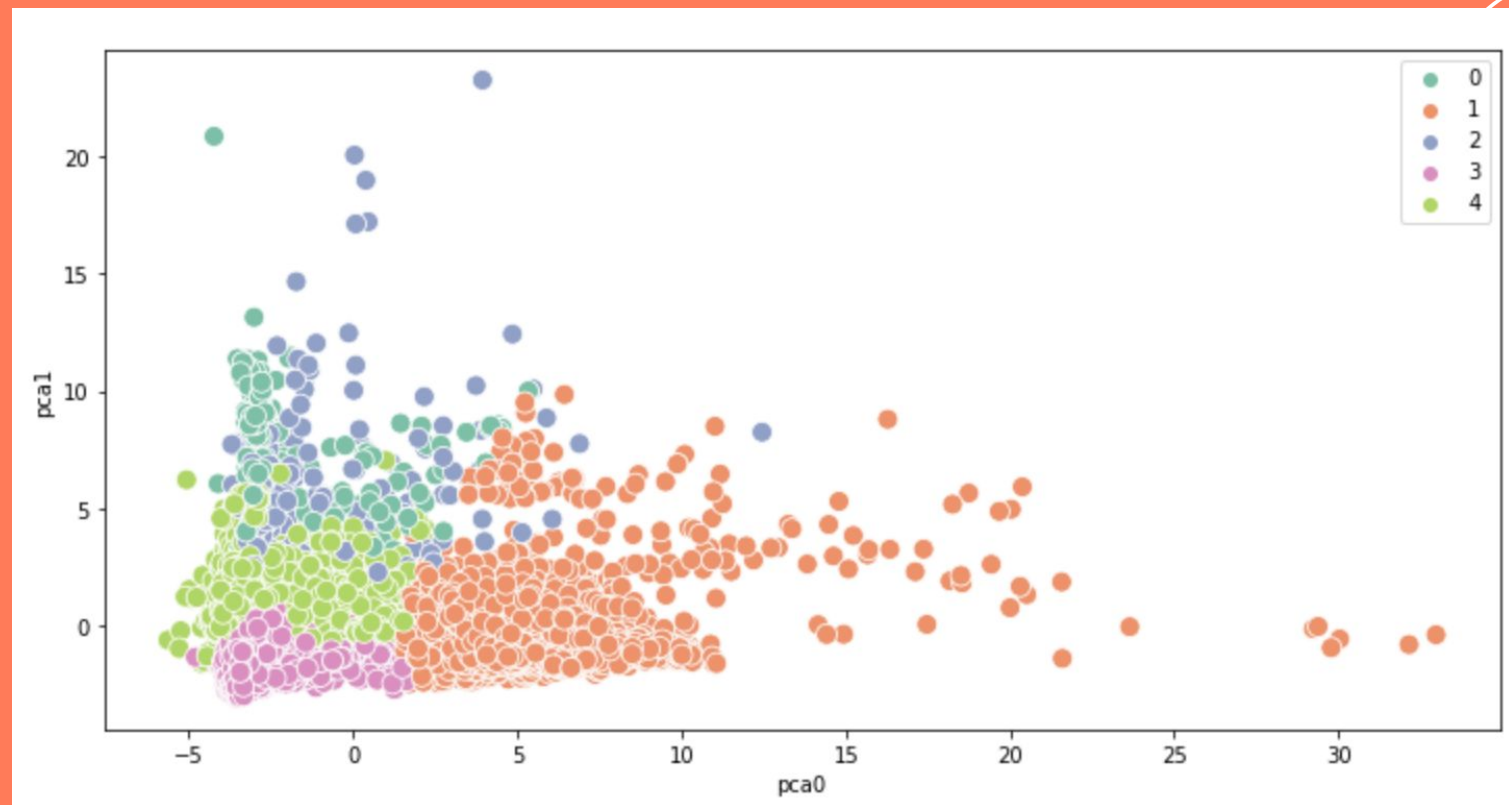
Number of clusters vs mean distance to centroids



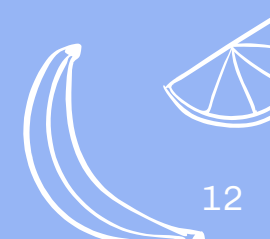
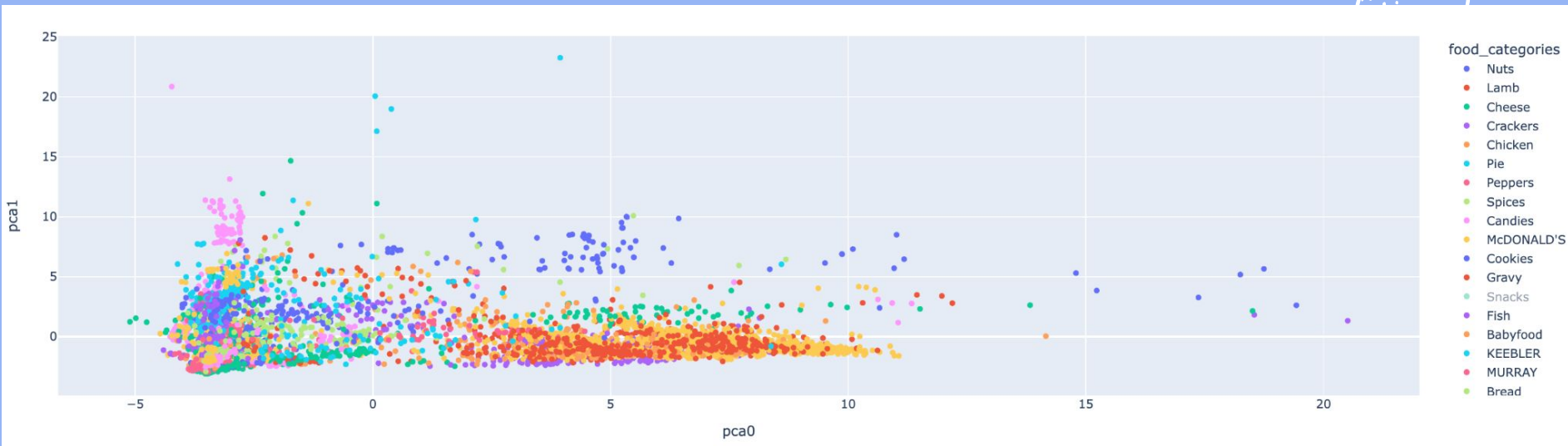
No distinct elbow =
no distinctly good number
of clusters to use

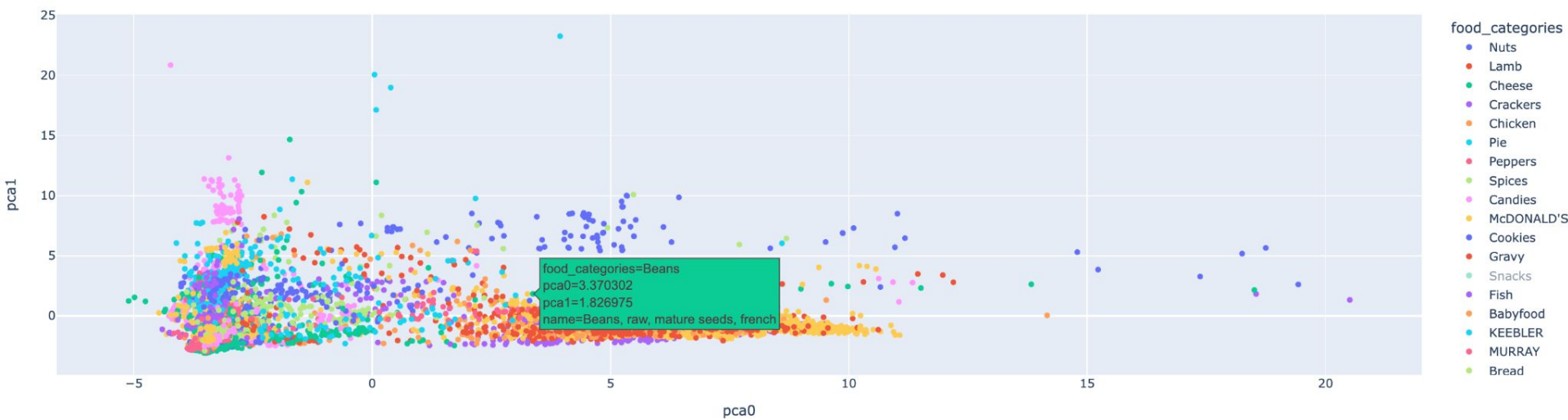
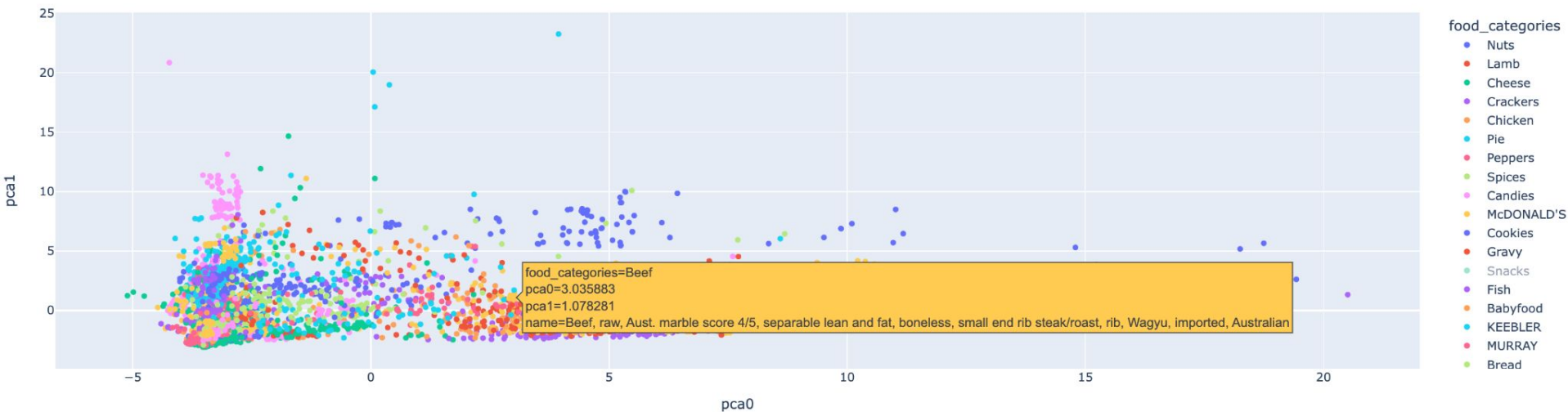
Determining how many clusters to use for k-means cluster plot

Food clusters graph



Food clusters Interactive plot





Alternative Food System - Sample 1

Inputted Food: Ground Beef

Alternatives:

- Avocados
- Fish Liver
- Turkey
- Pork

```
find_n_closest("Beef, raw, ground, grass-fed", 50)
```

```
{'Pork, cooked, lean and fat, low sodium, ham, cured': 0.017843684757858694,  
'USDA Commodity, canned, pork': 0.043957304262286535,  
'Seal, meat (Alaska Native), ringed': 0.048658630566079254,  
'Potatoes, prepared, frozen, o'brien': 0.058090040892777084,  
'Alcoholic beverage, 34 proof, coffee with cream, liqueur': 0.06700533320136434,  
'Pie, prepared from recipe, cherry': 0.07005294024222644,  
'Avocados, all commercial varieties, raw': 0.07400692014851402,  
'Turkey, bone removed, with skin, cooked, smoked, wing': 0.07424252201464111,  
'Fish, liver (Alaska Native), northern, pike': 0.08538866340520772,  
'Pork, roasted, extra lean and regular, low sodium, boneless, ham, cured': 0.08760493495472903,  
'Turkey, bone removed, with skin, cooked, smoked, drumstick': 0.10789165851415682,  
'Veal, raw, pancreas, variety meats and by-products': 0.10800607184535513,  
'LOMA LINDA Redi-Burger, unprepared, canned': 0.11045709109762991,  
'Canada Goose, raw, skinless, breast meat': 0.11364885931938903,  
'USDA Commodity, canned, luncheon meat': 0.11740733828884181,  
'Avocados, California, raw': 0.12884990891191894,
```

Alternative Food System - Sample 2

Inputted Food: Peanuts

```
find_n_closest("Peanuts, raw, all types", 50)
```

Alternatives:

- Pistachio
- Poppy Seeds
- Goat Cheese
- Sunflower Seeds
- Lamb Liver

```
'Nuts, raw, pistachio nuts': 2.0800703948798422,  
'Seeds, flaxseed': 2.0905638858830353,  
'Spices, poppy seed': 2.117423075407111,  
'Seeds, paste, sesame butter': 2.1179571867348153,  
'Lamb, soaked and fried, cooked, liver, imported, New Zealand': 2.1285297375837056,  
'Nuts, without salt added, with peanuts, dry roasted, mixed nuts': 2.1477356610811413,  
'Nuts, raw, cashew nuts': 2.1498811046720023,  
'Seeds, dried, whole, sesame seeds': 2.1894384593926834,  
'Nuts, lightly salted, without peanuts, oil roasted, mixed nuts': 2.3645200434329974,  
'Seeds, roasted and toasted, whole, sesame seeds': 2.3674066017153255,  
'Nuts, without salt added, plain, cashew butter': 2.468518141341181,  
'Cheese, hard type, goat': 2.5390124447800115,  
'Seeds, without salt, toasted, sunflower seed kernels': 2.550753067723989,  
'Wheat germ, crude': 2.5737125839980384,  
'Nuts, with salt added, with peanuts, dry roasted, mixed nuts': 2.629742759599611,  
'Nuts, without salt added, oil roasted, cashew nuts': 2.7058581599937024,  
'Nuts, with salt added, oil roasted, cashew nuts': 2.707430673222643,  
'Nuts, without salt added, without peanuts, oil roasted, mixed nuts': 2.7352047630241025,  
'Seeds, dried, watermelon seed kernels': 2.753660241289099,  
'Seeds, with salt added, toasted, sunflower seed kernels': 2.820803706969269,
```

Takeaways

- Our plans changed a lot: originally wanted to explore nutrition and implement a health score, but it wasn't a realistic goal
- Nutrition is very complex: foods we assumed were very different others were actually similar, and vice versa
- Limitations: our recommendation system doesn't consider whether a food is healthy



Citations

- <https://www.kaggle.com/datasets/trolukovich/nutritional-values-for-common-foods-and-products>
- <https://www.healthline.com/health/food-nutrition/six-essential-nutrients#minerals>
- <https://my.clevelandclinic.org/health/articles/4182-fat-and-calories>
- <https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label>