

## Goal:

**How are different types of cuisine different/similar to each other in the ingredients they use?**

Japanese, Jamaican, Brazilian, and more. Each country has its distinct way of cooking and distinct tastes. What is the defining taste and/or ingredient for each culture? What makes them unique? What are ingredients used universally across different cultures?

## Design Process and thought process:

### Dataset analysis:

Dataset: [Recipe Ingredients Dataset | Kaggle](#)

The dataset I acquired contains recipe information. Each recipe has an ID, the associated cuisine, and a list of ingredients used in the recipe. This makes it hard to categorize taste, but very easy to associate ingredients and the frequencies they appear in a certain cuisine.

The data set is processed in Python(Pandas) to make the information more organized. The recipe data set is converted into a matrix, with ingredients as rows and cuisine as columns. Each data point is the (number of times ingredient x appeared in cuisine y)/(total number of cuisine y recipes).

Original(train) & after processed(wo\_outliers\_percent.csv)

```
{
  "id": 10259,
  "cuisine": "greek",
  "ingredients": [
    "romaine lettuce",
    "black olives",
    "grape tomatoes",
    "garlic",
    "pepper",
    "purple onion",
    "seasoning",
    "garbanzo beans",
    "feta cheese crumbles"
  ]
}
```

```
ingredients,brazilian,british,cajun_creole,chinese,filip
(    oz.) tomato sauce,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0
(    oz.) tomato paste,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0
(10 oz.) frozen chopped spinach,0.1,0.1,0.1,0.1,0.1,0.1,
"(10 oz.) frozen chopped spinach, thawed and squeezed dry
(14 oz.) sweetened condensed milk,0.1,0.1,0.1,0.1,0.1,0.
(14.5 oz.) diced tomatoes,0.1,0.1,0.1,0.1,0.1,0.1,0.
(15 oz.) refried beans,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0
1% low-fat buttermilk,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.
1% low-fat chocolate milk,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.
```

## Initial Sketches:

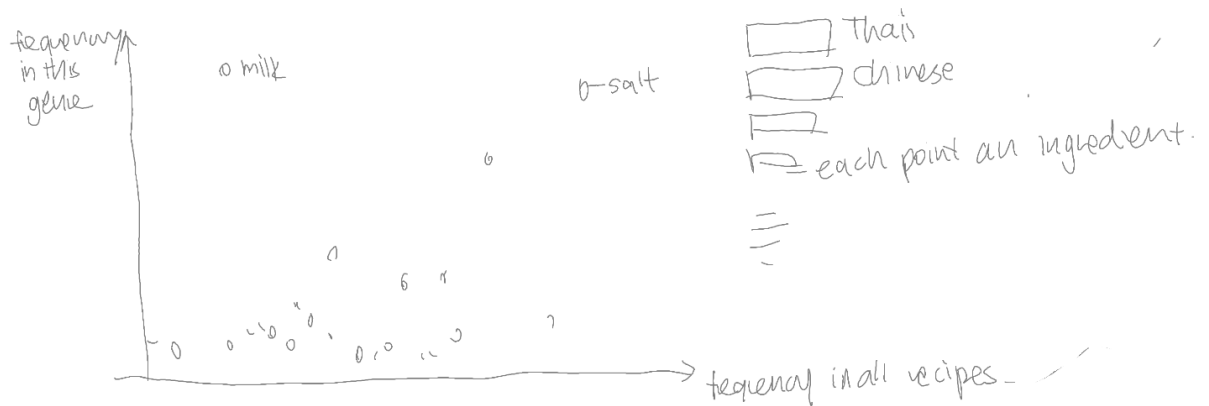
I've thought about three potential ways to visualize the data:

- 1.) **Scatterplot** - for each ingredient graph the frequency it appears in all recipes over the frequency it appears per genre. Have tabs that you can switch genres.

**Interaction:** brushing and selecting tabs on the side.

**Pros:** visualizes how some ingredients may be more common in one cuisine over the average of all recipes

**Cons:** Hard to see the relationship between different cuisines. No significant meaning behind the brushing interaction.

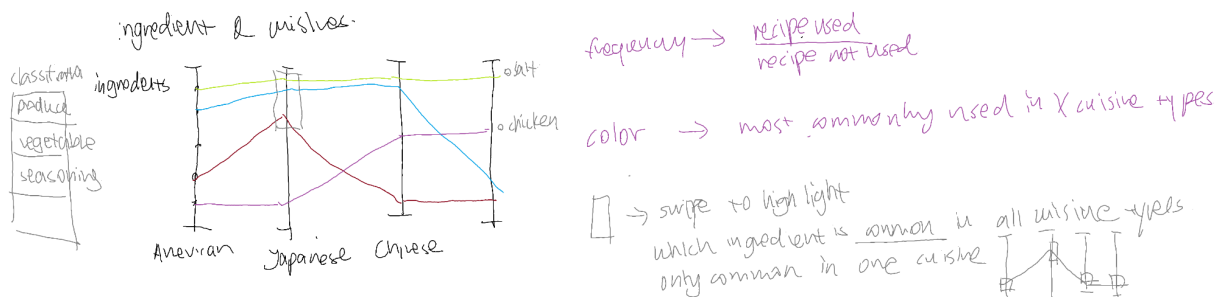


- 2.) **Parallel Coordinates:** Represent each cuisine as an axis, and have all the ingredients as lines spanning across the axis. Brush to select and view certain ingredients.

**Interaction:** Brushing data on the axis

**Pros:** Probably the best option among the three at explore how an ingredient is associated with different cuisines. And also you have the most control over how you filter your data ("which ingredients appear commonly in Russian and Chinese cuisine, but not in Japanese cuisine").

**Cons:** (Very)Difficult to implement. With a dense dataset it can start to look messy

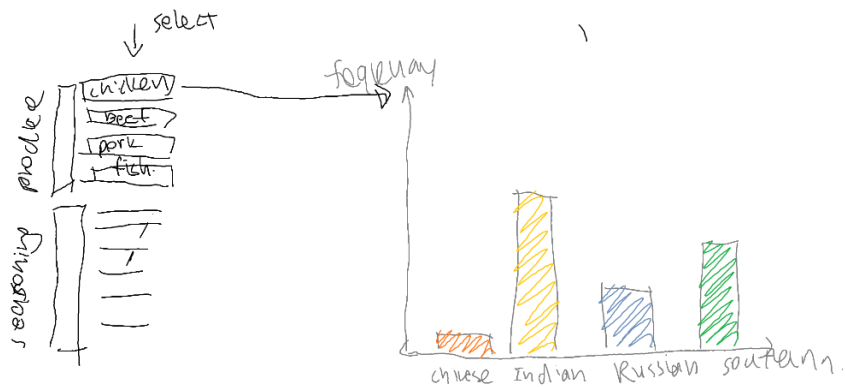


- 3.) **Interactive Bar Chart Display:** Select the ingredients from a list (preferably organized by their food type: produce, vegetables, etc). Display the corresponding percentage on a bar chart corresponding to different cuisines.

**Interaction:** Selecting ingredients

**Pros:** Easy to implement, easily view the distribution of this ingredient in different cuisines.

**Cons:** not very interactive, hard to interact when there's a lot of ingredients. And also missing information on the classification of ingredients into food categories.



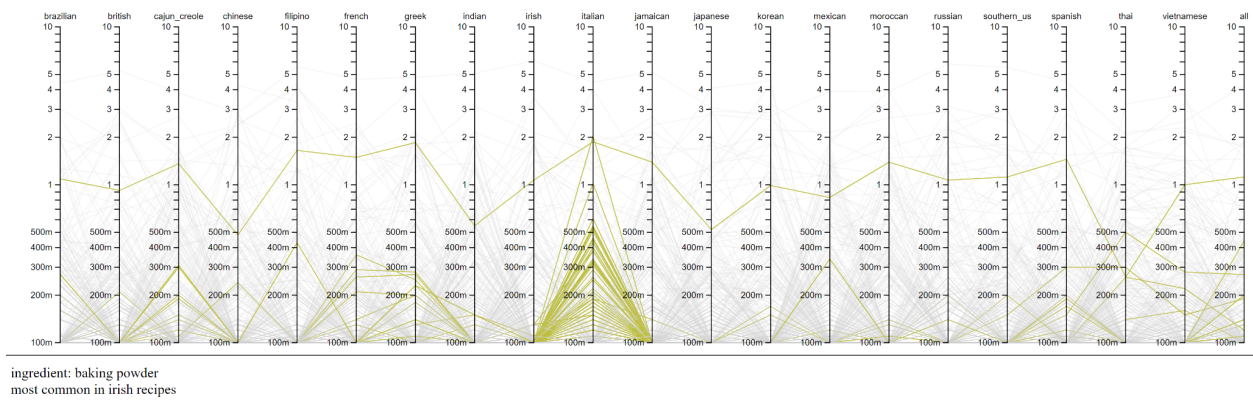
After careful consideration, I chose the parallel coordinate. I think this is the best way to represent this data and gives the user the most amount of agency to explore the dataset (and perhaps explore what they're interested in).

# Development Process

Overall every step along the process took longer than I expected. I spent a fair amount of time parsing and formatting the data to fit the parallel coordinates, but the majority of time was spent on debugging and learning D3. It was a lot more challenging to program brushable parallel coordinates than I expected. I ended up not being able to complete the brushable parallel coordinates on time. Instead, I programmed it such that you can click on the different axis and it will display the ingredients that have the highest representation in that cuisine.

Each axis is the cuisine, and the numeric values is the frequency of that ingredient's appearance in that cuisines's recipes.

The hope was that clicking on each cuisine axis allows you to see the ingredient that appears most frequent in that cuisine, but also highlight their distribution in other cuisines. Thus you can achieve a similar exploration as brushable parallel coordinates -- you find some ingredients only appear in very specific cuisines, while others appear commonly in all cuisines.



## Improvements/unimplemented features

I think the brushing would have added a lot to this project, and it's a shame that I wasn't able to implement it. Other than that, it will be helpful to have the ability to highlight individual ingredients and display their frequency on different axis upon hovering (Cumin: appears in 4.5% of indian cuisines, 0.2 percent in french cuisines, etc).