CUISINE VISUALIZER

PROCESS BOOK

COM-480 | Data Visualization | EPFL 2022

Venugopal Bhargav & Andreas Aarrestad

Introduction

Our purpose

Food undisputedly plays a huge role in our lives. It is our literal driving force in our day to day lives and we simply cannot go without food. It is only natural that we want to enjoy our food, seeing as it is such a big part of our lives.

One of the many ways we can enjoy our food is to add some variety to the things we consume and stimulate our palettes with different flavors. Another way to enjoy your food is to make it yourself. That way you get the satisfaction of knowing what you put in your body and also enjoy the taste. It is essential that you are aware if you are eating healthy, with all the nutrients that your body requires.

These are what we based our visualization project on. We wanted to allow people to explore new dishes and expand their palettes. Our intention was to show them that there are several dishes similar to the ones they already enjoy, from places they would have never expected. For those who enjoy making their own meals, we decided to give them a helping hand by suggesting the “easiest” cuisines from which they can choose to make the dishes if they have a busy lifestyle and want a simpler meal to cook up. Finally, we wanted to make sure users knew what they were consuming and to show them how cuisines compare in terms of nutritional value. That way they would know how often or seldom they should be eating dishes from said cuisines.

Audience

Our project is aimed toward regular food enthusiasts, travelers and health enthusiasts. We’ve envisioned the structure of our website as a chain of visualizations where each has a unique facet targeted towards a specific use case and target audience. This makes it easy for each user to decide a cuisine and recipe by using a visualization that matches their own prioritizations and preferences.

Overview

First, we want to introduce the user to all of the available cuisines. We chose to do this through a map to give a brief overview of the origins of the cuisines and their relation with other cuisines. The physical origin of each cuisine would be indicated by a circle or a symbol on the map. The user can then interact with these to discover similar cuisines.

Our next section is specifically tailored towards those looking for cuisines which recipes are usually fast to make. This can be those with limited time to cook or those who just don’t find it worthwhile to spend too much time in the kitchen. The data will be presented in a “stacked area” diagram inside an hourglass and users will be able to hover to get more information on the cuisine they choose to hover on.

Finally, we have a visualization tailored towards those seeking the ability to weigh the nutritional values of cuisines to their preferred flavor composition. This is done through a scatter plot which would plot one of the six flavor attributes with the cuisine's closeness to recommended nutritional values.

The Data

The dataset

The dataset we chose to work on is Yummly-28K, a dataset consisting of approximately 28 thousand recipes gathered from the Yummly personalized recipe recommendation website. Each recipe consists of an image of the dish, its ingredients, its nutritional values and various additional data. Though this dataset has been primarily used for image recognition, we found that with the information that it provides, we would be able to do a lot more than just that.

While excluding the images for each dish, the dataset is merely available as a folder of 28 thousand files where each file is a 1000-line JSON dictionary consisting of extensive information related to the dish's cuisine, ingredients, macronutrients, micronutrients and their respective units of measurement. We would then need to extract the relevant attributes from the JSON files and aggregate them in a CSV file.

Overall, our aim with this dataset is to produce educational and multi-faceted insights and visualizations on various cuisines and dishes by clustering and ranking them by their ingredients, flavor and nutritional values which in turn would be used to provide a useful and interactive way of receiving suggestions for new cuisines and dishes.

Exploratory Data Analysis

We could have used some kind of Regex patterns to re-extract and clean up the ingredient attribute from the combined string as mentioned earlier, but trying to get the pattern right would most likely prove unnecessarily time consuming and result in poor performance. Instead, we chose to use a web scraper to retrieve the ingredients from the website directly as the quantity, unit and instruction for each ingredient are located in separate HTML tags.

Still, there are some variable-length columns that would be more beneficial if they were constant-length. We therefore decide to perform some simplifications of the dataset:

- The keys of the 'flavors' dictionary is distributed to six new columns replacing the 'flavors' column as it is either empty or contains the six same keys for each recipe
- The 'cuisine' variable length array is replaced by the first item in the array. We justify this as only a small minority of the dishes' are categorized as multi-cuisine

After the initial preprocessing, performing dataset simplifications and reordering the order of the data, we are left with the following format for each recipe:

id	08923
name	Easy Ravioli Lasagna
cuisine	Italian
n_servings	8
time_s	3300
piquancy	0
sweetness	0.1667
sourness	0.1667
saltiness	0.8333
meatiness	0.3333
bitterness	0.8333
energy_kcal	2881.53
fat_g	35.78
protein_g	60.53
carbs_g	31.41
vitamin_e_g	0
vitamin_c_g	0.01
vitamin_b12_g	0
vitamin_d_iu	7.02
vitamin_k_g	0
vitamin_a_iu	1537.29
vitamin_9_g	0
vitamin_b6_g	0
fatty_acids_g	16.33
cholesterol_g	0.12
sodium_g	1.27
fiber_g	5.71
sugar_g	9.17
calcium_g	0.36
iron_g	0
magnesium_g	0.07
zinc_g	0
potassium_g	0.73
url	http://www.yummly.com/recipe/Easy-Ravioli-Lasa...
ingredients	["italian sausage", "marinara sauce", "crushed...

Figure 1: Data object after all the necessary preprocessing

As seen below, it's evident that certain cuisines like the American, Italian and Mexican cuisines are heavily overrepresented. We countered this a bit by normalizing the weights of each cuisine during data analysis and working with average values.

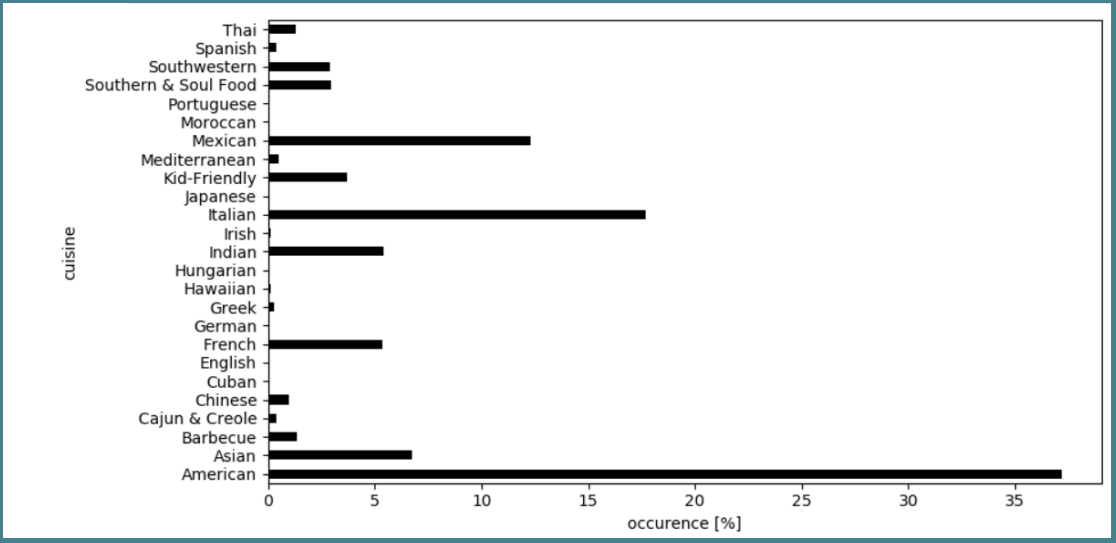

Figure 2: Bar plot of the occurrence of each cuisine in the dataset

Design

The World Map

We found it important that users could visualize how similar a certain cuisine is to another. It was also important for them to discover similar cuisines from regions they would never have expected. As such, we decided that the map with the connecting lines was the best way to visualize the actual distance between the origins of the cuisines so as to give the surprise

factor. We also wanted to utilize the entire screen width in order to amplify this effect. In addition to that, we wanted to show the extent of similarity in a very intuitive way – the thickness of the connecting lines. We found that the map with connecting lines was quite informative on its own. However, we decided to value-add by showing some simple information about the hovered cuisine using the tooltip. You can see below our initial idea of our sketch and the final execution.

Figure 3: Initial Sketch of the Map with connecting lines

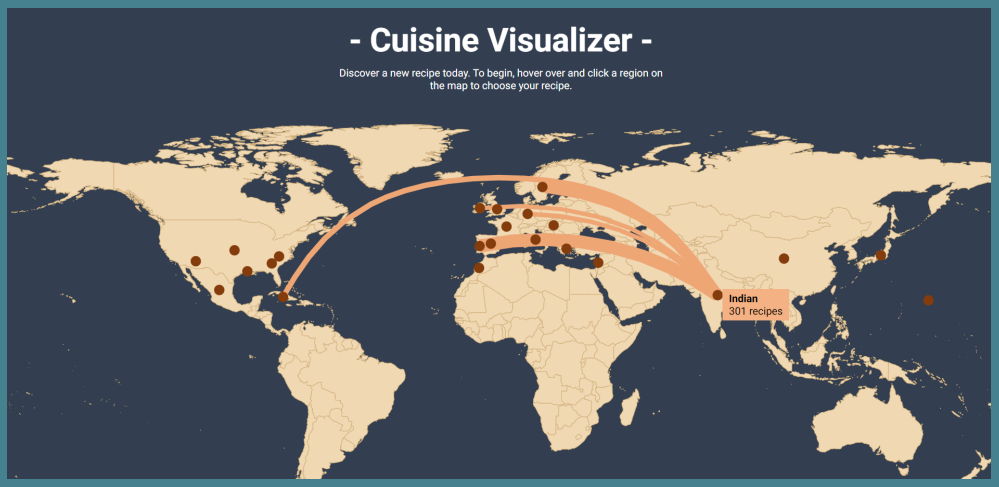

Figure 4: Final implementation of the map with connecting lines

The Hourglass

Again, we wanted our visualization to be very intuitive. At a first glance, we wanted the user to know that this visualization had something to do with time. And so, we decided to go with an hourglass shape. We also wanted them to be able to recognize which cuisines were most time-consuming easily. As such, we went with a “stacked area chart” inside the hourglass where the area at the top would represent the most time-consuming cuisine, while the one at the bottom would represent the least. In the spirit of recommending “easy” cuisines, we wanted to provide additional information that would be relevant to the complexity of the cuisine. As such, we wanted to show the user the average number of ingredients that each

cuisine uses upon hovering. You can see below the progression from the sketch to the final visualization.

Figure 5: Initial sketch of the Hourglass visualization

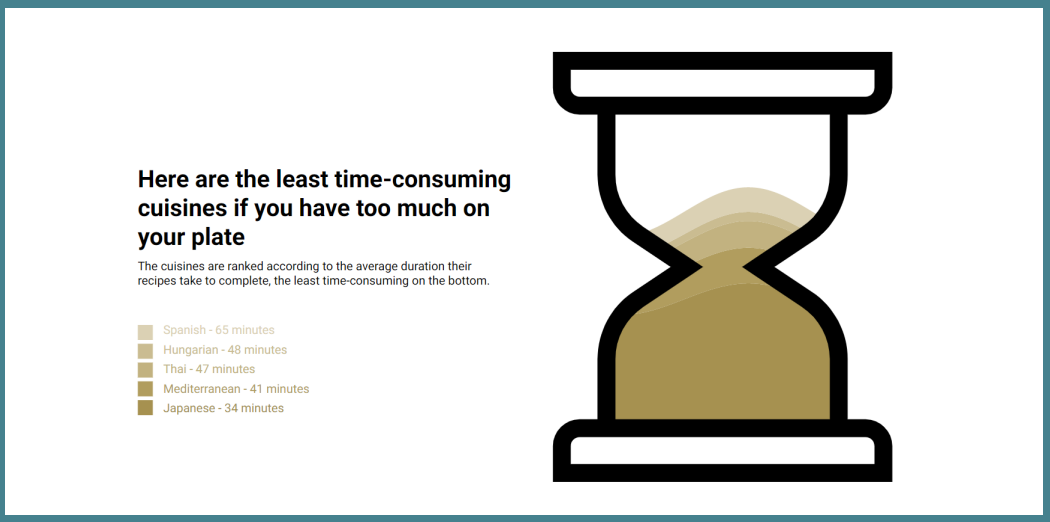

Figure 6: Final implementation of the Hourglass visualization

The Scatter Plot

For our final visualization, we wanted users to have an overview of how various cuisines rank in terms of their preferred flavor profile and also in terms of how much they deviate from recommended nutritional values (amount of sugar, amount of sodium etc). We decided that instead of overcomplicating things, a simple scatter plot would have the desired effect on the user. Furthermore, we wanted the users to be able to compare how the various cuisines differ when it comes to flavor profiles. As such, we provided a dropdown menu with the 6 flavor components which users can choose from. Upon changing, they can see the animations showing how cuisines differ between the various flavor profiles. Users can then also hover over the various scatter points and a tooltip will show which cuisine the point represents. You can observe the progress from our sketch to our final version below.

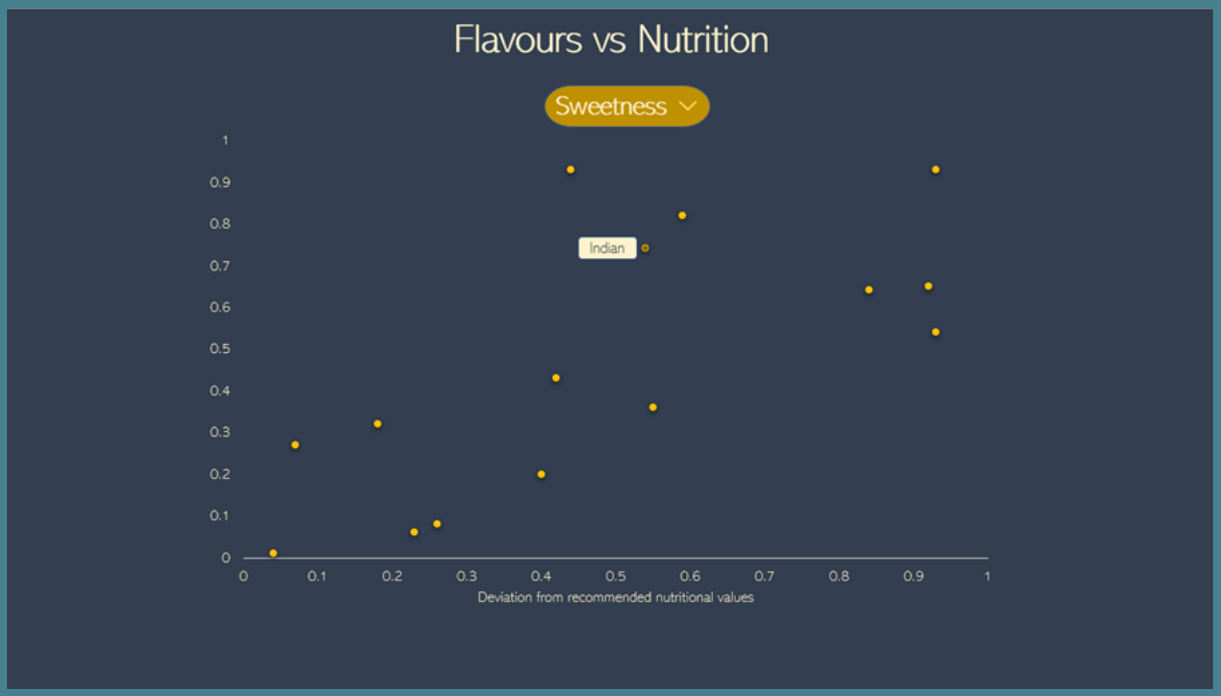

Figure 7: Initial sketch of the Flavor vs Nutrition Scatterplot

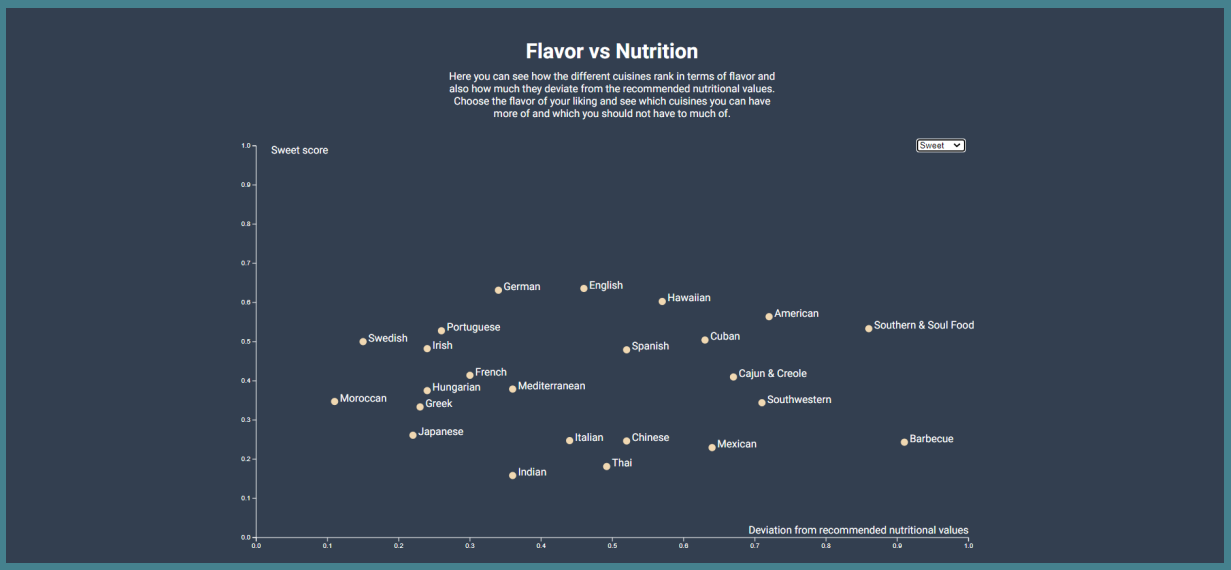

Figure 8: Final implementation of the Flavor vs Nutrition Scatterplot

The Dropdown Menu

After providing all this information for the user, we wanted to give users actual dish suggestions so that they could then go find recipes for it or try it out the next time they go out for a meal. We found this to be a good way to have some closure to the whole experience of the website and only decided on it once we started working on the other visualizations and felt something was missing. Below is the version of the dropdown menu, which is not completely finished because of lack of time. We wanted to also include a spider plot consisting of the taste profile of the recipe and nutritional information also. Ideally, we would also have a link to the website, an image of the recipe and the instructions for the recipe but this would require us to web scrape as the dataset did not include this.

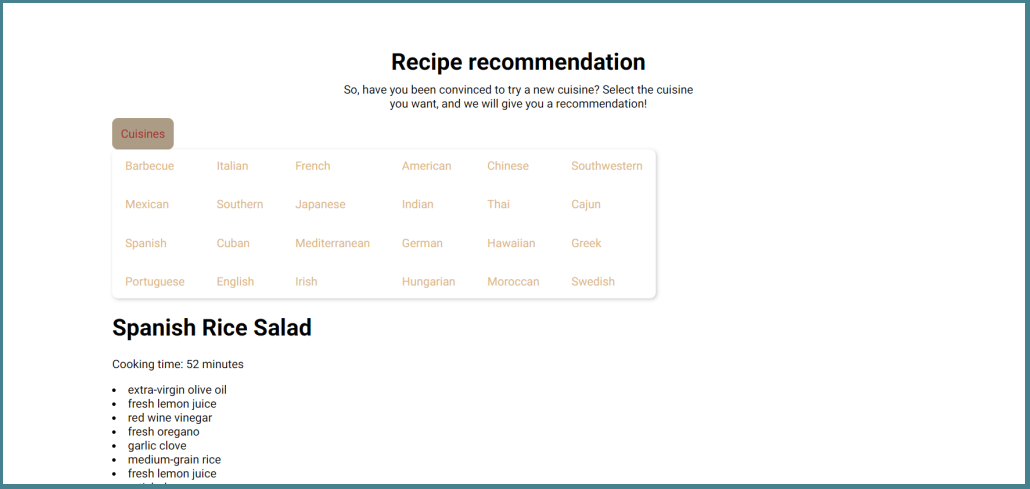

Figure 9: Final implementation of the recipe recommender

Challenges

The main challenge about the project was the fact that we were working in a group of 2 instead of the suggested 3. This led to each of us having a bigger workload and as such more time and energy consumed.

As for the project itself, we first struggled with finding a suitable dataset. We had many interesting ideas for the data visualization but we could not find suitable datasets despite spending a significant portion of 2 days searching. We then decided that we needed to change our approach to finding a dataset first and then building ideas off what we have. This was a lot more manageable and we found a dataset a few hours after.

The dataset that we found was very extensive with lots of information. However, a lot of the information we wanted to use was formatted in a very inconvenient way which would require heavy pre-processing in order to be useful. As such, we also spent a significant amount of time thinking of more time-efficient ways to extract the necessary data from the dataset. We still had to preprocess the data a lot and also figure out how to do web-scraping to make our job easier. There was also a very skewed representation of the cuisines. Some cuisines had thousands of recipes while others had less than ten.

Another challenge was narrowing down on the visualizations that we wanted to focus on. We had many ideas and possibilities with the dataset that we had found but limitations in time meant that we had to pick and choose the most important ones. We also had to consider the feasibility in terms of difficulty of execution for each case before deciding.

For the technical aspect of the project, it was difficult to work with d3.js because it involved lots of trial and error. It also required lots of learning by doing, which is very time consuming. Figuring out how to parse the data from CSV files using d3 was particularly challenging. It was also particularly challenging to correlate the flavor similarities between cuisines and tie that to the thickness of the connecting lines.

Reflection & Improvements

Overall, we are satisfied with how our website turned out. The countless possibilities did leave us wanting to accomplish more but we did the best we could given the time we had.

We could have definitely added more improvements to the website. Firstly, we could have made the page mobile friendly, seeing as websites are now made for smaller screens first before the big screens.

We would also have liked to provide more interactivity for our visualizations:

- Adding the option to choose what parameter is used to calculate the closeness between cuisines by a drop down menu or a wheel (Map)
- Giving users the option of sorting according to time or the ingredients instead of only sorting it by time required (Hourglass)
- Adding functionality that would allow the user to recognize which cuisine a point represents and easily distinguish them (Scatterplot)
- Plotting the dishes of a specific cuisine instead of the averaged values of cuisines to provide a better idea of how dishes in specific rank in nutrition and flavor (Scatterplot)

Peer Assessment

Both members were involved in the decision-making process of the majority of the aspects in this project like: the dataset, the purpose, the visualizations and data to use and the design of the website.

Andreas Aarrestad did the data cleaning, the website skeleton, the map visualization, the scatter-plot visualization, the hourglass visualization & part of the cuisine recommender and parts of all the reports.

Venugopal Bhargav worked on the sketches for the visualization, the dropdown menu section, the website screencast and majority of all the reports.