Ingredient Interactions: An Exploration of Recipes

Lucy Li and Katherine Song

Goals

The goal of our interface is to allow a general audience interested in food to explore how ingredients in dishes connect and distinguish different cuisines around the world and how they are used with each other. Insights that we expect our audience to be able to learn after interacting with our interface include what popular ingredients in different cuisines are, what cuisines are similar to each other, and how ingredients are used in conjunction with each other in a cross-cultural context.

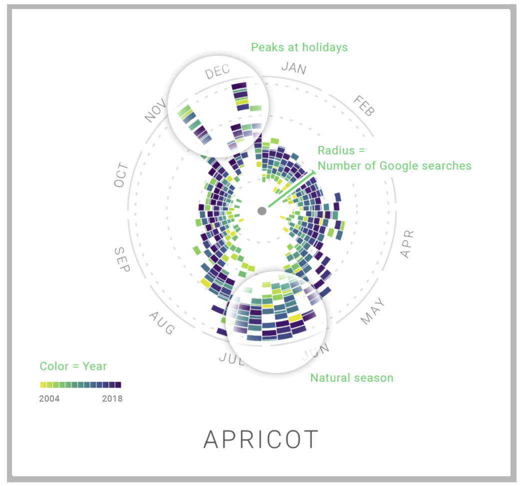


Figure 1: "The Rhythm of Food"

Related Work

We draw inspiration from several pieces of related research and visualizations. Food is central to many cultural norms and practices, and as such, it is the topic of many visualizations made for general audiences. For example, Google News Labs' Rhythm of Food visualizations explore seasonal trends in food-related searches through a series of interactive "year clock" radial charts (Figure 1)[1]. In addition to exploring search trends, visualizations are also useful in helping hopeful chefs understand what ingredient pairings are common and/or complementary. Some examples of these that particularly motivated us include David McCandless and Willow Tyrer's "Taste Buds" piece has 3-level "trees" of ingredients that shows common ingredient pairings from a database of 1000 online recipes[2]. For example, *cress* is one of several nodes of the *greens & salad* tree, and *chicken*, *eggs*, *mayo*, *pink fish*, and *potato* surround the *cress* node as suggested pairings (Figure 2). Similarly, Michael Moyer and Jan Willem Tulp created an

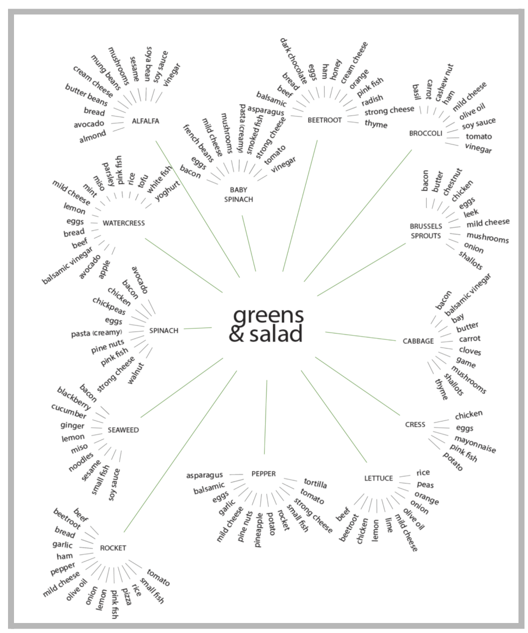


Figure 2: "Taste Buds" flavor tree

[1] Google News Lab and Truth & Beauty, "The Rhythm of Food," http://rhythm-of-food.net.

[2] D. McCandless and W. Tyrer, "Taste Buds: Complementary Flavours," Information is Beautiful, 2009 https://www.informationisbeautiful.net/visualizations/taste-buds/.

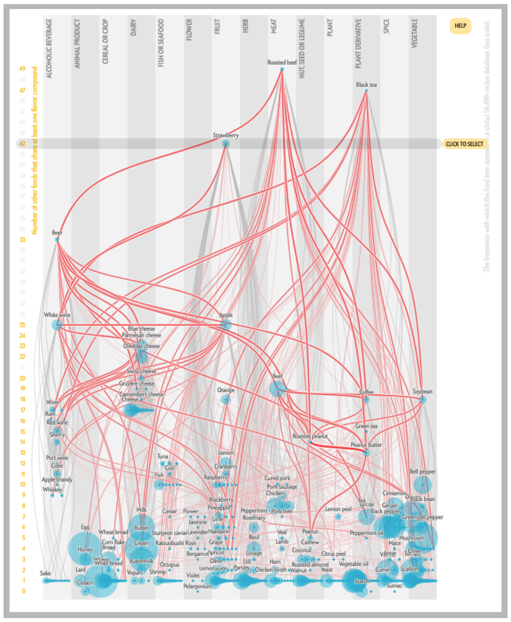


Figure 3: "The Flavor Connection" interactive map

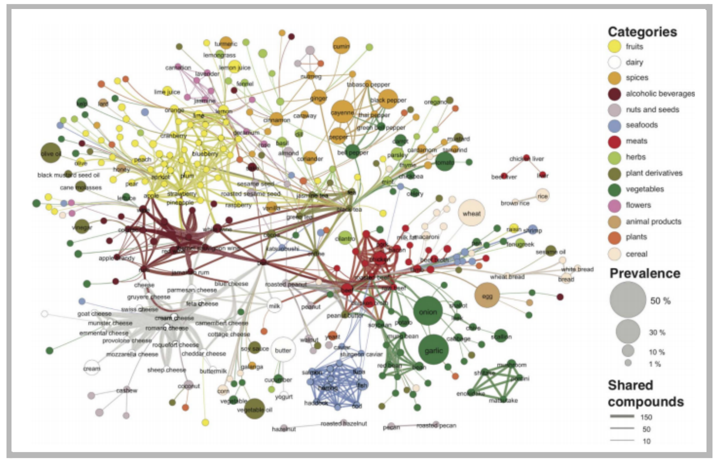


Figure 4: "Flavor Network"

interactive flavor map with ingredients as nodes and connections between them indicating pairings based on the similarities of the ingredients' flavor compounds (Figure 3)[3]. We found the ability to click on nodes and hide others to better explore pairings in a complex network particularly noteworthy. Additionally, they also mentioned that pairings based on flavor compounds were followed to different extents in different cuisines, which also inspired us to support cross-cultural explorations in our own visualization. The cuisine-ingredient link was explored in more detail in Y. Ahn et al.'s paper on the principles of food pairing[4], which offered a variety of bar charts, scatter plots, Venn diagrams, and linked node structures to highlight cross-cultural similarities and differences. One of their figures (Figure 4 here) is a flavor network showing how different ingredients are related through flavor compounds. While a bit overwhelming without interactivity, we appreciated that a viewer could quickly start to identify particular clusters of food that were related.

In general, defining and exploring the relationships between cuisine type and ingredients are tasks that are of interest to amateur cooks and researchers alike[5], with applications including labeling online recipes, developing intuition for cooking foods from an unfamiliar cuisine, and generating new fusion recipes. Most notably, on his FlowingData website, Nathan Yau used the

---

[3] M. Moyer and J. Willem Tulp, "The Flavor Connection," Scientific American vol. 309 no. 3, 2013 https://www.scientificamerican.com/article/flavor-connection-taste-map-interactive/.

[4] Y. Ahn et al., "Flavor Network and the Principles of Food Pairing," Sci Rep 1, 196, 2011 https://www.nature.com/articles/srep00196.

[5] H. Su et al., "Automatic Recipe Cuisine Classification by Ingredients," Ubicomp '14 Adjunct, 2014 https://dl.acm.org/doi/pdf/10.1145/2638728.2641335.

same database that we're planning on using to create 3 visualizations [6]. The first two are sets of bar charts that visualize the 5 most used ingredients and the 5 most cuisine-specific ingredients (Figure 5). We borrowed this format of presentation and the concept of "cuisine-specific ingredients." We did identify a few points where we wanted to incorporate interactivity and color-coding to allow viewers to explore cross-country/continental similarities more easily, as we will demonstrate in the next section. The last visualization on FlowingData is an interactive chart that allows the user to select a cuisine to generate a scatter plot showing how often various ingredients are used in that cuisine compared to other cuisines.

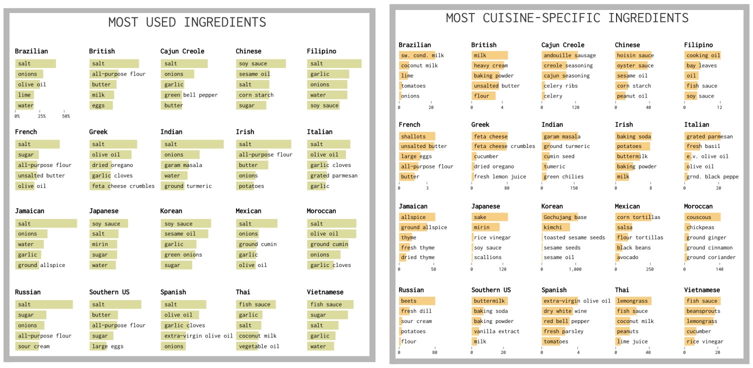


Figure 5: FlowingData Most Common Ingredients

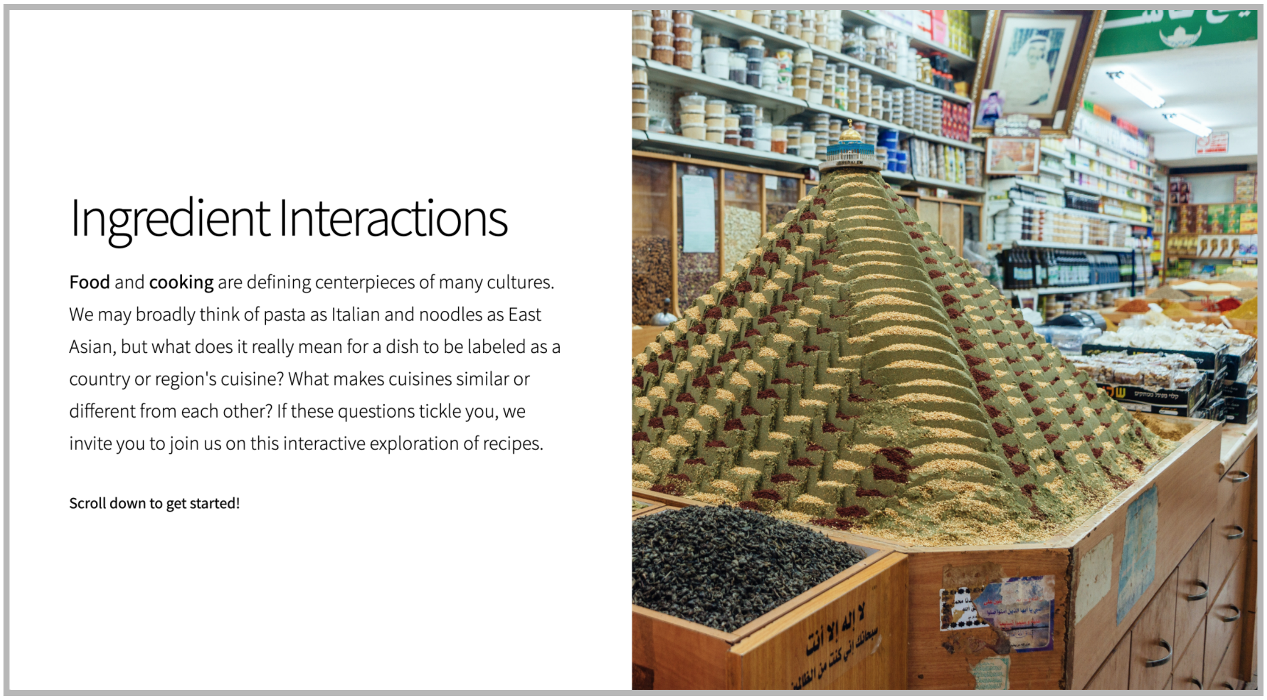
Visualization Description/Screenshots

Our final product is a website (https://lucy3.github.io/ingredient_viz/) containing 4 interactive charts with interspersed text to guide a viewer through an exploration of our dataset. We suggest viewing the website in Google Chrome, and it is possible that it may load slowly the first time. Annotated screenshots with elaboration of some design decisions are below:

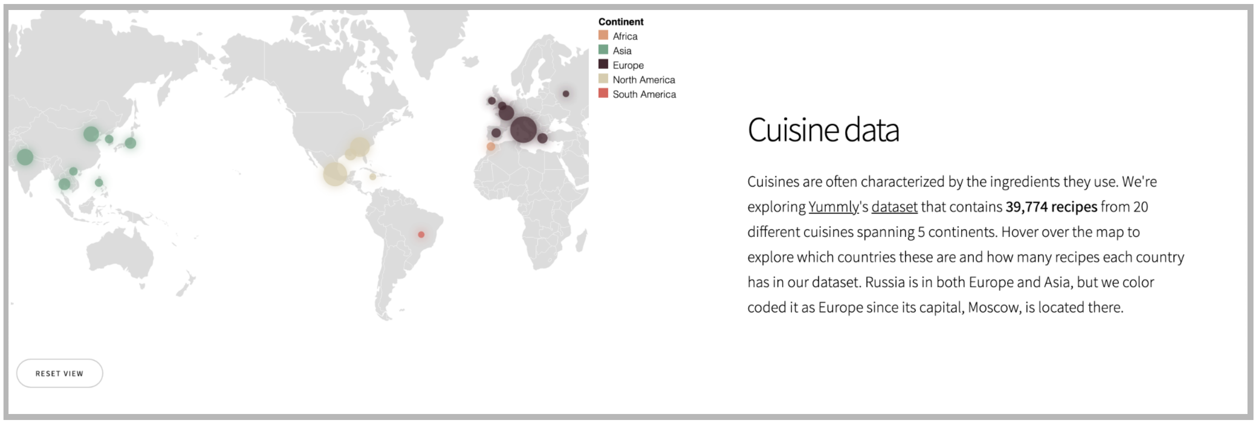
First, we introduce the viewer to our topic: food. The first image is a personal photo from a spice market in Jerusalem.

---

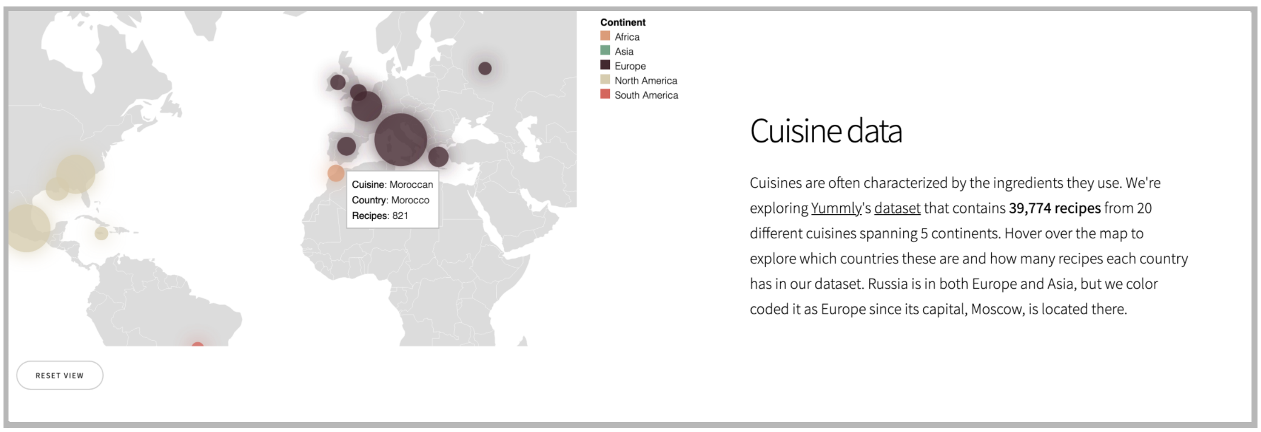
[6] N. Yau, "Cuisine Ingredients," FlowingData blog, 2007 https://flowingdata.com/2018/09/18/cuisine-ingredients/.



We then introduce our dataset. We do this by presenting an interactive map with text to allow the viewer to see what cuisines our dataset contains and how many recipes come from each cuisine. The location of each bubble is centered over country capitals, which is an imperfect simplification of a cuisine's "center" but seemed to be the least distracting way to present this information. Originally, we colored whole countries by their continent, but based on feedback from our usability study, this made Russia and Brazil loom distractingly large despite not contributing many recipes in our dataset, and it also made it difficult to understand geographic groupings. Based on their respective Wikipedia articles, for "Cajun-Creole" cuisine, we centered the bubble over New Orleans, Louisiana, and for "Southern US" cuisine, we centered the bubble over Atlanta, Georgia. The size of a bubble is proportional to the number of recipes from that cuisine in our dataset. We introduce our color scheme here that geographically groups cuisines by continent. We selected these colors to still be distinguishable by color-blind people, which we tested on color-checker.com and with 1 color-blind user. While we do not intend to imply that all European cuisines, for example, are alike, because of the moderately large number of cuisines in our dataset, we felt that it would be helpful to use create colored groups to remind viewers of geographical proximity as they explore the rest of the charts.



The viewer may zoom into the map and hover over a bubble to view the corresponding cuisine name, country name, and number of recipes. A "Reset View" button resets the view of the map.



We then provide an opportunity for our viewers to explore external resources to learn more about each of the cuisines in our dataset. We present a gallery of "typical" dishes from each cuisine from their respective Wikipedia pages. Hovering over each image brings up a short list of popular dishes, and clicking on the image takes the viewer to the cuisine's Wikipedia page.

An Overview

Often, cuisines close to each other geographically share common ingredients and dishes, but historical connections such as immigration, colonialism, and trade have led to additional waves of influence. **Hover** over each picture of a dish to see the cuisine it is from. **Click** on the picture to read more about the history of each cuisine and representative dishes from the cuisine's Wikipedia page.

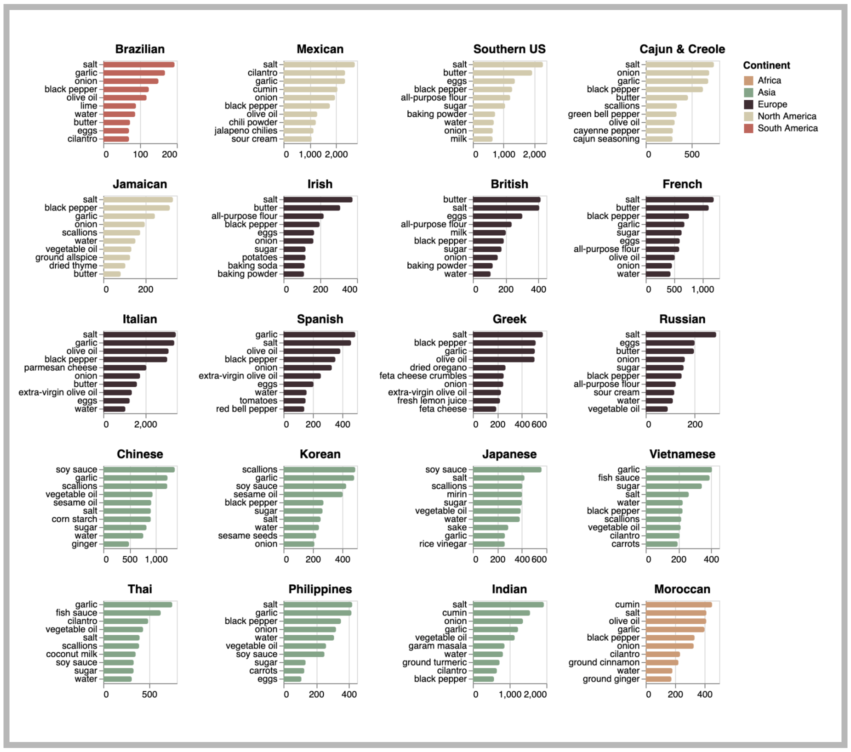
The page for Cajun & Creole leads to the page for Lousiana Creole cuisine, so click here if you want to learn about Cajun cuisine.

Next, we present our second chart, which is a series of barcharts showing the 10 most popular ingredients for each cuisine based on the recipes in our dataset. We provide a couple of questions in our text to help our viewer begin to explore. We made the bars relatively thin to allow viewers with smaller computer screens to view the whole chart without scrolling.

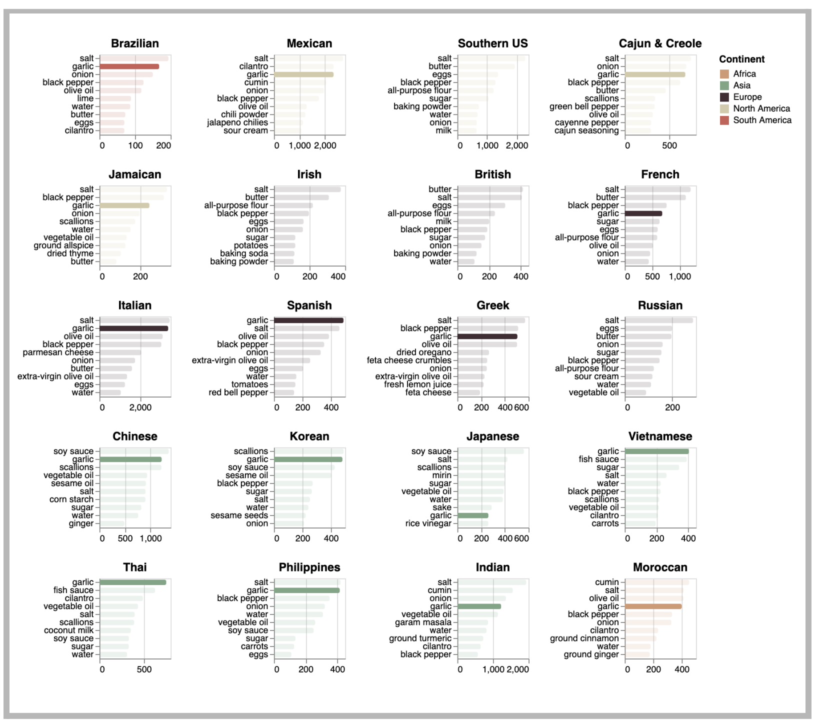
Popular Ingredients

First, we'll take a look at what ingredients tend to be **most popular** in each cuisine. Below, you'll see a series of bar charts showing the top 10 ingredients in each cuisine and their frequencies, with cuisines colored and grouped by continent.

Hover over bars to see which popular ingredients are shared across cuisines. Which cuisines tend to use a lot of **butter**? Which cuisines use a lot of **soy sauce**? Are any foundational ingredients uniquely popular to only one cuisine?



Hovering over a particular ingredient in a particular cuisine highlights that ingredient in other cuisines by reducing the opacity of all other bars. Below is an example of hovering over "garlic" in Brazilian cuisine:



We felt that one natural thing to wonder while exploring the above bar charts is how similar/different cuisines are based on their ingredients. The bar charts above start to give some semblance of this, but we wanted to create a 2D scatterplot to help viewers more easily find similarities. We represent each cuisine as two coordinates (as described in the Data section). We use the generic labels of “Cuisine Dimension 1” and “Cuisine Dimension 2” in an attempt to not overwhelm the general viewer with technical jargon or unnecessary complexity. From our usability testing, we realized we needed some basic description of this “computational magic,” so we provide this along with links to more thorough explanations of what we did before presenting our 2D scatterplot:

From these charts, we get some sense that some cuisines may share many ingredients, while some do not. Are cuisines within the same continent more similar to each other than cuisines across continents? Maybe there is a better way to see if this is the case...

Cuisine Similarity

We want to know how cuisines compare to each other. However, there are over **65 thousand unique ingredients** in our dataset, and we can't just go through each and every one and eyeball if it's shared or not between two cuisines. That would take forever!

Instead, we take the long list of ingredients for each cuisine and do some computational magic to squash it down ...

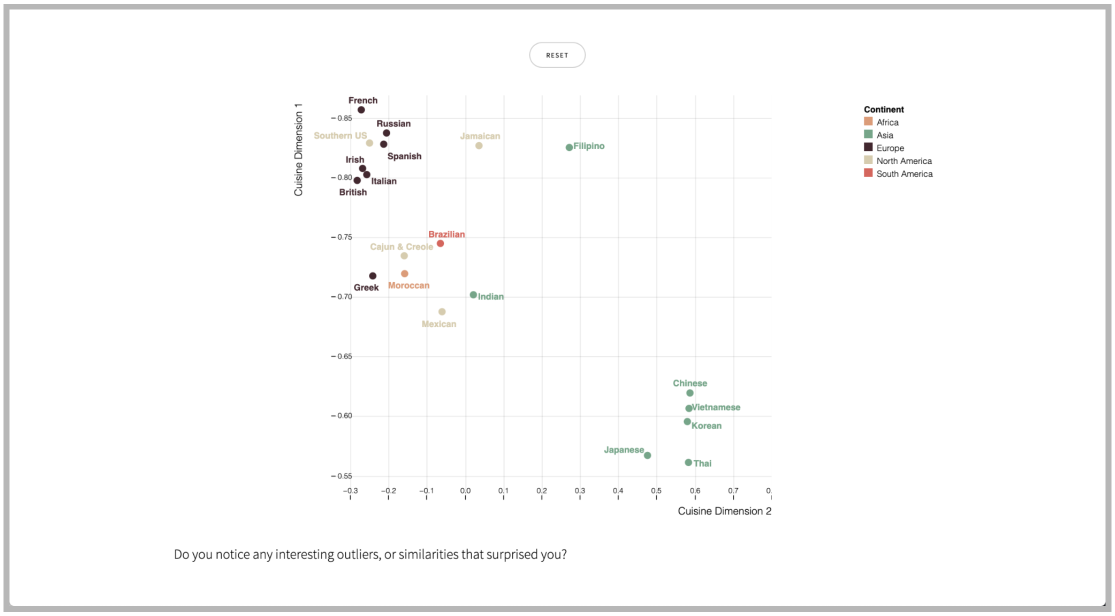
Jamaican: 332 salt, 173 onions, 153 water, 139 garlic, 128 ground allspice, 125 pepper, 111 scallions, 105 dried thyme, 103 black pepper, 101 garlic cloves, 96 vegetable oil, 89 ground black pepper, 80 sugar, 75 ground cinnamon, 72 thyme, 72 brown sugar, 71 coconut milk, 67 soy sauce, 66 olive oil, 66 curry powder ...

↓

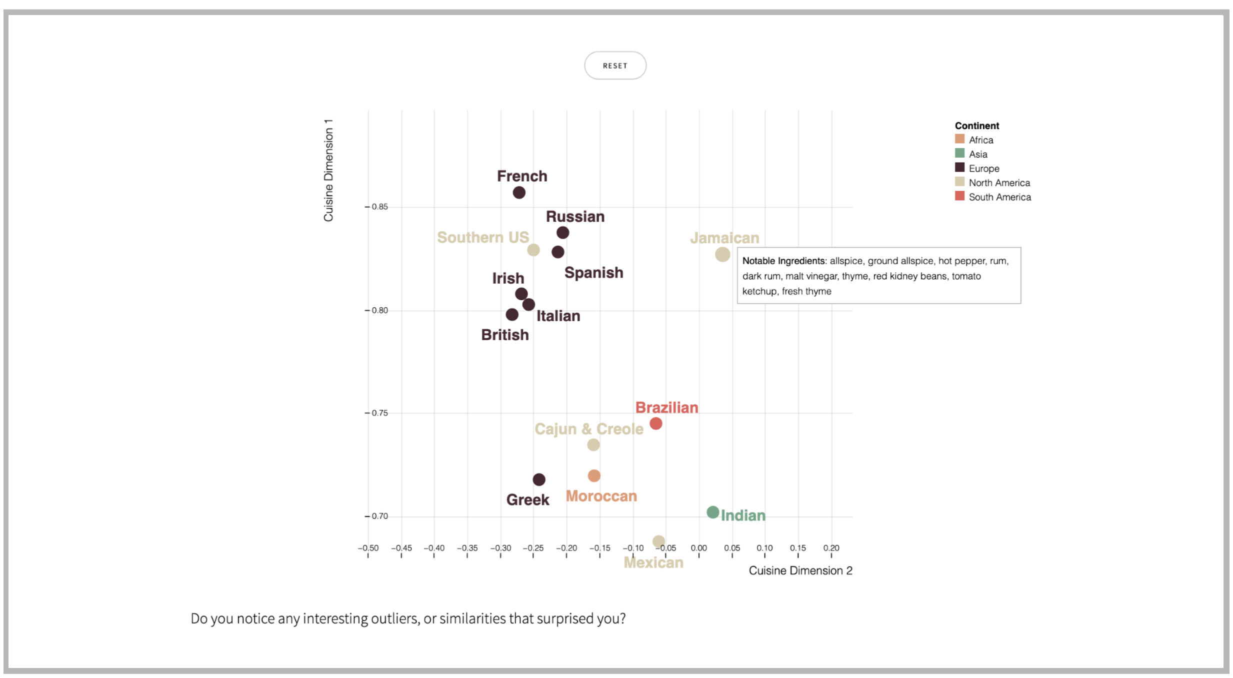
Jamaican: 0.83, 0.04.

... all the way to just **two numbers** (which we'll call "Cuisine Dimension 1" and "Cuisine Dimension 2"), so we can look at how much cuisines relate to each other on a 2D chart. In the scatter plot below, cuisines are located based on the values of these two numbers. **Cuisines that are closer to each other are more similar**.

Hover over each cuisine's point to view more information about where it is from and some of its **notable ingredients**. A notable ingredient is one that is often used in one cuisine but not in any other cuisine. These are extracted using a calculation similar to what was in done in Flowing Data's visualization.



We use a Voronoi-based adjustment to allow labels for each cuisine to be placed in a way that without overlapping with one another. We also enable zooming and panning to allow viewers to get a better look at neighborhoods of cuisines. The "Reset" button at the top resets the view of the scatterplot. Hovering over each point triggers a tooltip that shows the cuisine's "notable ingredients," which are those that are popular in that cuisine but not others. Below is an example of zooming in on and hovering over the "Jamaican" point:



For consistency and to allow viewers to get a sense of geographical similarities and differences, we use the color scheme we established in our prior map.

Now that viewers have spent some time thinking about cross-cuisine comparisons, we want to slightly shift to exploring cross-cultural connections at the ingredient level. For this, we created a network, where nodes are ingredients and edges are their co-occurrence (as described in the Data section). The size of a node corresponds to the popularity of that ingredient, and nodes are colored based on whether an ingredient is used broadly (colored gray) or mostly in one continent (colored with that continent's color). The length of edges between nodes is determined using a force-directed graph algorithm, which treats each edge as a mechanical spring and allows for clearer identification of node communities, which consist of ingredients that tightly co-occur with each other and each others' neighbors.

Again, we provide a few "starter" questions to help viewers begin to explore this network. These questions emphasize the idea that this network acts as a landscape of ingredients, where different regions correspond to coherent collections of ingredients and dishes. For example, there are two clusters colored green (meaning they are ingredients used in Asian cuisine), rather than one, because Indian ingredients and East/Southeast Asian ingredients do not co-occur as much. Note that we use text bolding to emphasize actions the user can take for each of our interactive visualizations.

A Network of Ingredients

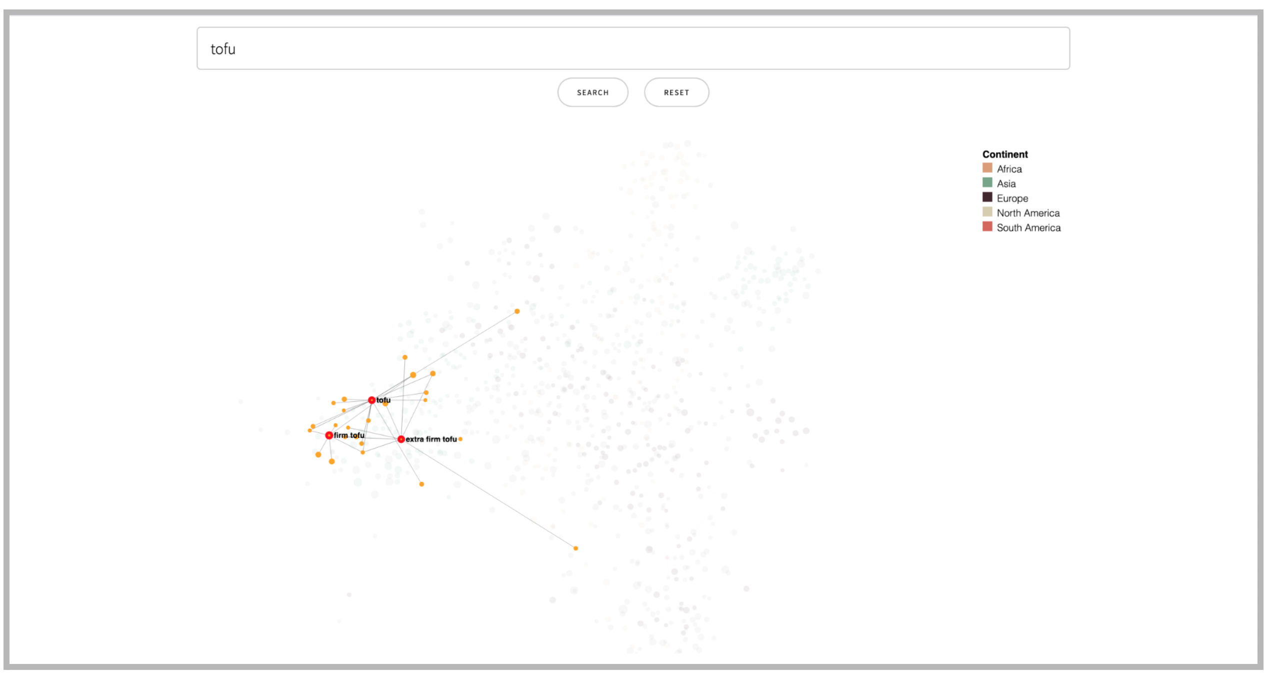
Let's zoom in even further on the ingredients used in these cuisines. We know that some classic pairs of flavors go very well together, like tomato and basil. What are other common pairings of ingredients? Are pairings divided based on regional cuisine differences?

We create a network of popular ingredients, drawing edges between the most significantly co-occurring pairs. Nodes are colored by continent, if at least **75% of their recipes** are in that continent. Gray nodes are ingredients that are used across continents.

Food for thought: what regions of the network seem to correspond to **Asian cuisines**? Can you find the community of nodes that corresponds to **dessert** ingredients?

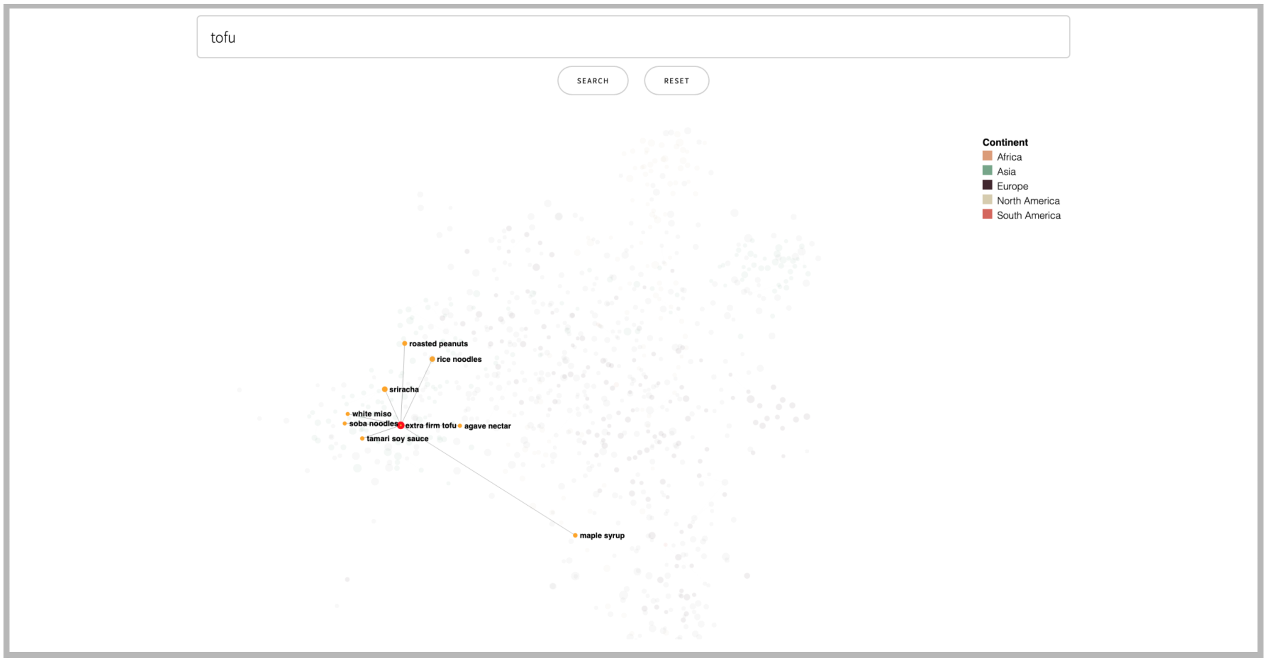
You can **zoom** for a closer look, **hover** over nodes to see their neighbors, **click** on nodes to lock in the view of their and their neighbors' labels, and **click** again to release.

It was very clear that we needed to support interactivity to make exploring this large network manageable. Our most important feature is the search bar. Viewers may search for an ingredient, and nodes containing that search term in the name are highlighted and labeled:

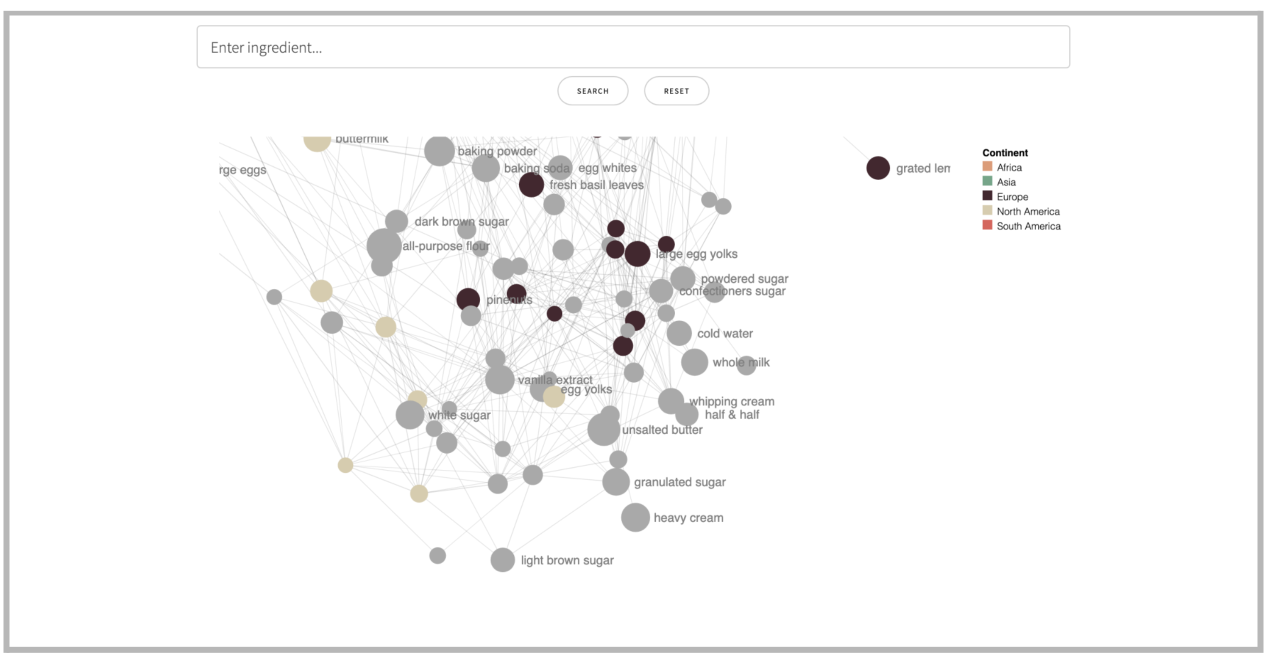


By clicking on a node (either before or after searching for it), that node is highlighted along with its connections. The highlighting and labeling is "locked" until the selected node is re-clicked on or the

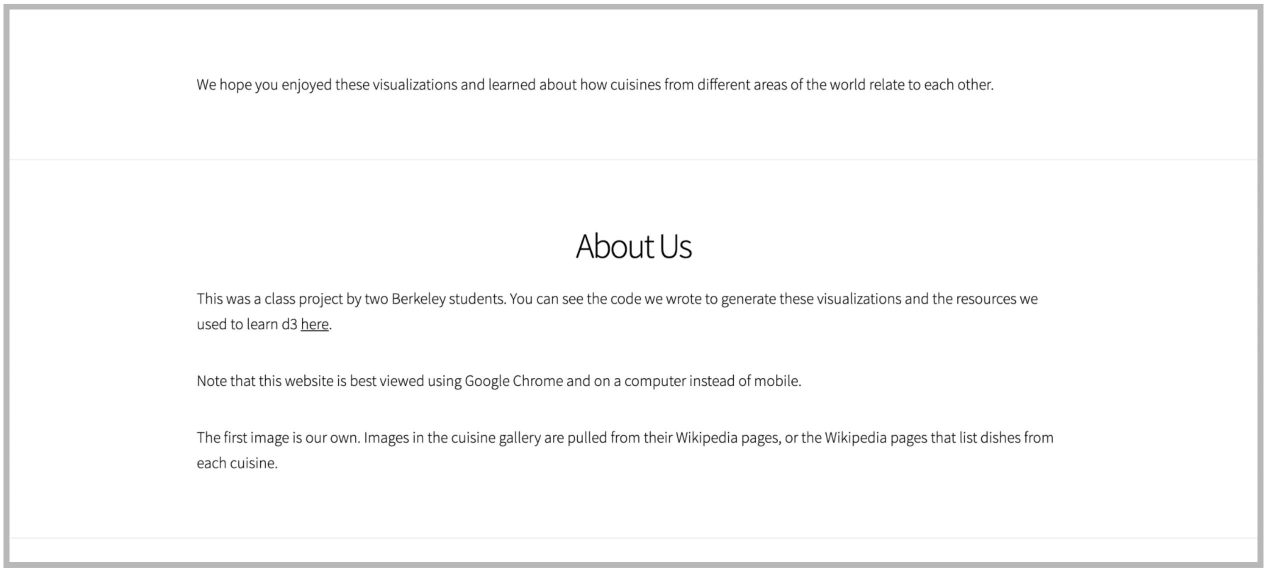
“Reset” button is pressed. Initially, hovering over other nodes unlocked this, but our usability tests suggested that that made it too easy to “lose” a selection.



Viewers may also zoom and pan to manually explore the network. If the zoom level exceeds a certain threshold, the labels of ~25% of the most popular ingredients show up. We initially had all the labels show up, but partially because of the density of our graph, this created a very unpleasant viewing experience. Below is an example of zooming into the bottom arm of the network, which appears to be a dessert corner:



We close our interface with final text:



Data

The dataset that we used for this project is a collection of 39774 Yummly.com recipes from 20 different cuisines. We found this dataset on Kaggle (https://www.kaggle.com/kaggle/recipe-ingredients-dataset). Each row in the dataset describes a recipe and contains a list of ingredients in that recipe along with a tag identifying the cuisine that the recipe comes from. There are no amounts listed for ingredients -- just ingredient names.

For our interactive map, we supplemented the cuisine dataset by manually incorporating information on country name, capital with latitude/longitude coordinates (from https://simplemaps.com/data/world-cities), and continent for each cuisine. We also counted the total number of recipes per cuisine.

For our bar charts, we totaled the number of recipes containing each ingredient and sorted them to show the top ten most frequent ones for each cuisine. We combined a couple of ingredients that were extremely similar (e.g. unsalted butter/salted butter/butter, pepper/black pepper/ground black pepper, extra large eggs/large eggs/eggs, large garlic cloves/garlic cloves/garlic, scallions/green onions).

For the scatterplot, we created ingredient count vectors for each cuisine and transformed them using TF-IDF and singular value decomposition to assign each cuisine an x and y coordinate along two dimensions. Notable ingredients were calculated by dividing the probability of an ingredient in the target cuisine with its probability in all other cuisines.

For the ingredient network, we filtered to ingredients that occur at least 50 times in our dataset. To eliminate the dominance of extremely common ingredients like “salt” and “vegetable oil”, we used normalized pointwise mutual information (NPMI) instead of raw co-occurrence frequency to assign weights to network edges. Our network only shows edges where NPMI > 0.25.

Tools

Data transformations were done using Python in a Jupyter notebook. All prototyping was done in Observable. We used vega-lite for the bar charts and d3 for the other 3 charts. The Github repository containing the code for data transformations and the Observable notebook we used to create our visualization are linked in the “Links” section below. Our web page uses a pre-made HTML/CSS/Javascript template, where the gallery of dishes from our cuisines is a pre-styled class from the template.

Results

We tested a prototype of our website in the form of an Observable notebook with 3 users. Our full usability testing writeup is linked in the Links section below, but we summarize our procedures and key findings here.

We conducted a 30-minute interview over Zoom followed by a 4-question Google Forms questionnaire for each user. All 3 users were able to easily complete a list of basic “tasks” that we presented to gauge the readability and informativity of our charts. Different users gravitated towards different insights, which encouraged us in concluding that our visualization did indeed support exploration.

A lack of legends and explanations in our prototype made our use of color particularly confusing. Our prototype’s explanations were also sometimes confusing in their level of technical detail, which is why we rewrote all explanations in our final web page. Our explanations were also edited to include more segues and questions to guide the user. Some of the tasks we used for our user study became exploration prompts on our web page’s narrative for each visualization. Users also did not always read the text in Observable in each notebook, and we hope that our web page format encourages this more.

We redesigned our map after our user study so that instead of coloring each country in our dataset with a solid color, we colored circles that are sized according to the number of recipes instead. We think that this makes it easier for viewers to quickly and more accurately understand where the recipes in our dataset come from. We also added tooltips to the scatterplot to facilitate the viewing of notable ingredients, especially on smaller screens but also for viewers more inclined to skip through text (originally, this information was shown above the plot, mostly because we hadn’t figured out how to show tooltips in d3 before our user study). For our network, we increased the readability of the network by only labeling nodes if they are clicked on or if they are large enough and zoomed in on. We also fixed several performance bugs. One of the difficulties of the network visualization is that it contains a lot of data (which supports exploration of the dataset) but still needs to be consumable and manageable; we tried to manage this in our final website by supporting ways to search for nodes and selectively disclose information.

Links

Visualization website: https://lucy3.github.io/ingredient_viz/

Full usability testing writeup:
https://docs.google.com/document/d/1ZQIFLL0EODiefUhfR57TFE0lWwr6kEwbHIvwdWOyER0/edit

Observable notebook: https://observablehq.com/d/6ff8d3fcb0277a42

Github repository for data transformations: https://github.com/lucy3/ingredient_viz/settings

Contributions

Task	**Lucy**	**Katherine**
Searching for dataset	50%	50%
Data transformations	80% - TF-IDF & SVD, NPMI transformations for notable ingredients and cuisine comparisons	20% - Manual combination of some dataset ingredients (e.g. garlic and garlic cloves), Voronoi-based transformations of TF-IDF coordinates to make scatterplot labels not overlap
Visualization prototyping	30% - Bar charts, performance enhancements for force-directed diagram, initial wireframes	70% - Map, scatterplot, adding interactivity to force-directed diagram (enabling hover, click, search, and reset functions), enabling tooltips
User testing	50% - planning, user recruitment, interviewing/note taking	50% - planning, user recruitment, interviewing/note taking
Website making	60% - Main text, descriptions of algorithms, adding images	40% - Layout, intro text/image
Documentation (user testing + final writeups)	40%	60%