Kevin Lee [ksl66]
Yulu Cao [yc758]

Data Overview

Data Source

We had three sources of data:

1) Map data (topojson) of Japan from NIAESVIC
2) Demographic data of Kyoto from e-Stat
3) Kyoto restaurant Tabelog (Japanese Yelp) review data from Kaggle

Data Preprocessing + Filtering

A fair amount of preprocessing was required to obtain all the data we needed to create our visualization. The variables we were interested in were the names and cuisines of restaurants, the average ratings of each restaurant, the average price of a meal at each restaurant, the location of the restaurant, the population of different wards (regions) in Kyoto, and the map data of Kyoto.

First, we used a Python script to import the restaurant dataset as a .csv file and converted this to a pandas dataframe. We then found that there were 97 unique cuisines among the restaurants in the dataset, so we decided to sort them into large categories and ultimately narrowed the number of cuisines down to 8. We also modified the price column of the restaurant reviews dataset using Regex to transform strings of price ranges (e.g., ¥4000~¥4999) into numbers and further into number-form averages of these ranges. The dataset was then exported to a .json file with the restaurant name, cuisine, price, rating, and location (coordinates) columns included.

Next, we edited the topojson file we had obtained of Japan by using an external tool (https://mapshaper.org/) which helped us to 'dissolve' the topojson and to simplify it into attributes we were interested in. The original topojson we had of Japan had far too many small regions that we felt was unnecessary and visually confusing, so we decided to edit the topojson so that we would just be left with Japan's wards (subdivisions of cities in Japan).
Our map data was then loaded into our visualization script, where we reduced the scope of the map down from all of Japan to just the 11 wards of Kyoto, achieved by filtering on the ward name attribute (CSS_NAME) that was already in the topojson.

We then took our demographics dataset and added a new attribute to each entry of the dataset. This attribute was the Japanese name for the wards. By adding this new attribute to the demographic dataset, we were able to connect it to the Japan map topojson, whose ward attribute values were written in Japanese. In doing this, we were able to assign each ward in the topojson, attributes from the demographic dataset such as Japanese/foreigner population per ward.

Lastly, we assigned wards to each of the datapoints in the restaurant dataset. Each restaurant had a set of coordinates (longitude and latitude) that we were able to use to connect to the wards geodata provided by the Japan map topojson. By using d3.geoContains(), we were able to determine which ward a restaurant's coordinates belonged to. Assigning a ward attribute to each of the restaurants later enabled us to calculate things such as 'average rating of restaurants in X ward', or 'number of Korean restaurants in Y ward', which was important for our project.

Design Rationale

Marks and Visual Channels

Our visualization's marks are circles, where each circle represents a single restaurant. While we did consider using icons instead of circles, we decided to rule the idea out because we felt as though they would hardly be visible given the number of overlapping and crowded restaurants (~900 restaurants), and the relatively small size of the map.

The most visible channel that we employed is the positioning of the circles on the map. Restaurants were plotted on the map in accordance with their longitude / latitude values. We felt as though it was important to visually represent our data on a map because the goal of our visualization is to help people who may want to establish a restaurant in Kyoto to find a strategic location to do so. By having a map for the visualization, our users would be able to notice spatial trends in data, such as where there are clusters of restaurants of a certain cuisine. We zoomed into the map of Kyoto and eliminated as many of the unnecessary surrounding regions as possible so that we could make the relevant parts of the map on the screen as large as possible (easier to spot spatial trends, more space for overcrowded data points).

The colors of our circles were another major visual channel. As explained earlier, we decided to use color instead of icons to represent different cuisines. The original restaurant dataset started off with over 90 cuisines, and we felt the need to reduce this down to around 7 or 8 different cuisines. We were aware that there was a tradeoff between having too many colors and having too few cuisines to show. Given the goal of this visualization, it would be somewhat ridiculous to narrow down the cuisines too much (e.g., all of Japanese food being classified as just one cuisine) but having too many colors would also undermine the ability of the user to interpret the visualization without having to constantly refer back to the color legend. In the end, we chose to go without 8 instead of our original desire to go with just 5 unique cuisines. This was because we realized that most of our users would probably not be seeing all the cuisines on the map at once – most users would likely only have 1 or 2 cuisines in the filter at once (e.g., someone looking to open a French restaurant probably would not be using the 'Japanese (Noodles)' filter), which meant that color overload would not be as big of a concern as we first thought.
To make colors as easily distinguishable from each other as possible, we used an online tool to help us pick distinct colors (https://mokole.com/palette.html).

It is also important to briefly discuss the channels we *didn't* use. We chose not to vary the size or color hue of the circles because we thought that these would not be largely noticeable given that there were almost 900 data points (restaurants) presented on our map. The amount of overlap between circles would make it quite hard to see any differences in size, and we did not want to make circles huge in the first place because we also wanted the map underneath the circles to be visible and clickable.

We did still recognize the fact that there may be too few channels to help accomplish the effects we wanted. To make it easier to spot trends when all cuisines were selected in the filters, we implemented code such that the circles of the most abundant restaurants in each region would be raised. For example, if there are a large number of 'Japanese (Noodles)' in region/ward 'X', then of the circles in region X, those representing 'Japanese (Noodles)' would be raised relatively higher than the circles representing other cuisines. When having multiple cuisines selected in the filters, this would make it such that it would be easier to spot which cuisines dominated by number in a particular region (although domination by abundance does not necessarily mean that they are more popular – this was addressed by the rating filters).

<u>Interactivity</u>
We had three different filters that all worked together to help users navigate through the visualization to address questions they may have had when looking for strategic places to place their restaurant. The visualization was divided largely into two parts; the map on the right and the filters on the left. As such, upon entering the page, all users would be able to clearly see all the filters that are available to them.
The text below both the rating and price slider ("showing restaurants with rating/price at or above: X") even when first loading into the page would help users to immediately understand that these sliders can be used to manipulate what would be shown on the map.
Furthermore, the cuisines filter being color coded makes it stand out on an otherwise colorless page - this makes the color codes easy to see, and in conjunction with the fact that the cuisine filter is positioned quite close to the map, there is no real use for a separate color legend - the cuisine filter acts as both the filter and the legend.

Users would be able to adjust all these filters to immediately see colorful results on the map - because the map and the page in general is white, similar to the cuisine filter being eye-catching due to the stark contrast, the same is true for the colored circles that appear on the map. We felt that the real time change of colors on the map when using the sliders or the cuisine filter would make it very obvious for users as to what is possible. Through the provided filters, users could look into fairly specific questions such as “in Kyoto, what area has the highest density of Western-cuisine restaurants that score at least a 4.0 in ratings and are moderately expensive?”

Mouseover function was implemented with regards to the map’s regions and the restaurant circles. By hovering over individual restaurants, users are able to see textual information such as the restaurant’s name, average meal price, and rating. By hovering over regions, users are also able to see a textual summary, such as the region name, average rating of restaurants in the region (taking active filters into account), average price of restaurants in the region (again, taking active filters into account), and the region’s population-to-restaurant ratio. Through these interactions, we felt that users would be able to gain a high-level overview through the region summaries, and then would be able to dive further into lower-level, restaurant-by-restaurant exploration once they identified a specific region/ward that they were interested in.

The text for the mouseovers was positioned towards the bottom right of the map, close to where most of the circles on the map were concentrated. This was also an intentional design choice; we wanted to make it such that users looking at the map would not have to move their eyes too much to go from looking at where they were mousing over to the summary statistics / text. This is also the reason why the entire visualization is crowded towards the center of the page in general; we wanted to make the project reasonably compact so that many things would be visible at once. The position of the summary text for regions was placed slightly higher on the page than the summary text for restaurants to create a distinction between what the summary text was addressing (individual restaurants vs. regions) .

We did debate between making the region mouseover into an on-click interaction instead - however, we felt as though it was not intuitive for users to click on regions without any external guidance (no real signals indicating that clicking a region would yield textual summary), and thus decided to opt for the mouseover instead, which users could easily figure out simply by moving their mouse over the map once.

Story

<u>Intended User</u>

As mentioned throughout this report, our visualization was created with a certain user group in mind, this being people who are thinking about setting up a restaurant in Kyoto. These people would use the visualization as a decision-helping tool, allowing them to explore trends with spatial layout of potential competing restaurants, filterable by price/rating/cuisine.

<u>Insights</u>

Because of the nature of the visualization being exploratory and radically different depending on the user’s individual circumstances (e.g., their planned restaurant’s cuisine, their planned price range, etc.), there are no definitive insights that we drew from the visualization we created. However, while playing around with the visualization by ourselves, we did notice some general trends.

Firstly, many of the restaurants in our dataset are focused around the Shimogyo-ku and Nakagyo-ku wards. These wards also have the highest density of high rating and high priced restaurants. What we found interesting, however, was that the high rating restaurants were not necessarily the ones that were highly priced. In fact, we found that when the rating slider was set moderately high, the remaining restaurants would often have an average meal price that was in the lower half of the price range. Meanwhile, the ratings of the restaurants that had high prices were quite middling, with just one or two restaurants out of ~10 also clearly possessing a high rating. While we cannot infer any causation, we do think there is somewhat of an inverse correlation to be observed between rating and price; one reason that could potentially explain this being that places that are priced highly are more likely to be judged under standards disproportionately harsh for how relatively high the price (and quality of food) is.

Japanese seafood restaurants and Western restaurants tend to be the most expensive of places, while there is a much more even distribution of cuisines when it comes to highly rated restaurants.

Izakayas are clearly the most abundant of the restaurant categories, but tend to filter out rather quickly when raising the sliders for either rating or for price. This could imply that many izakayas are simply seen as places to grab a 'just-good-enough' meal / drink at reasonable prices.

Team Contributions

Kevin
- Visualization brainstorming + design + searching for datasets (~4 hours)
- Pre-processing + Sliders/Filters + Visualization (~8 hours)
- Report: Data Overview + Design Rationale (~2 hours)

Yulu
- Visualization brainstorming + design + searching for datasets (~4 hours)
- Filters + Visualization (~8 hours)
- Report: Design Rationale + Story (~2 hours)