INFO 5100 Project 3
Team: Cynthia Mahoney, Carolyn Morikawa, Emily Lin

Topic: Visualizing the Presence of Vegetarian and Vegan Restaurants in the U.S.

INTRODUCTION

Around 5% of adults in the U.S. consider themselves vegetarians. Vegetarianism, veganism, flexitarianism (eating mostly plant-based foods while allowing meat and other animal products in moderation) have all been on the rise. If you're a plant-eating foodie living in America, where should you go to find a vibrant food scene? In this project, we sought to explore the distribution of vegetarian and vegan restaurants in the U.S. Which state offers the most number of vegetarian and vegan restaurants per million residents? Which cuisine and diet is the most popular amongst the existing veg/vegan restaurants? How do the prices of the most popular cuisines and diets compare to each other?

DATA PROCESSING

Vegetarian and Vegan Dataset

We sourced our vegetarian and vegan restaurant data from a dataset found on Kaggle. This dataset is a smaller sample of a larger dataset from Datafiniti's Business Database and thus, not all vegetarian and vegan restaurants are represented by our visualization. The dataset originally consisted of 47 total columns but we kept only 21 of them to start – we chose to drop columns that we knew we were not interested in due to their irrelevance to our hypothesis and questions (i.e. "twitter", "websites", "sourceURLs", etc.). Furthermore, the original dataset started with 10K rows because each row represented a single menu item from a restaurant. However, we were only focusing on restaurants holistically, and not granularly by menu item. Thus, we further filtered down the dataset until only unique restaurant ids showed up. This converted our 10K row dataset into a 200+ row dataset, significantly reducing the size of our files.

One of the variables that we were super interested in using is price. However, all the prices for the restaurants are associated with each individual menu item which meant that dropping the additional rows from the original dataset caused us to lose a lot of data about the restaurant pricing. In order to use pricing as a variable, we created an average price calculator that took in all the menu item prices – there were two columns "menus.amountMax" and "menus.amountMin" but we chose to look at maximum menu price because of the assumption that the upper bound price is more important to a greater number of people than minimum price. Our average price calculator took the max price of each menu item for each unique restaurant, averaged them together to find the overall average maximum price of the restaurant, and spat out a clean price dataset that we used later to build our double lollipop/barbell chart that shows difference in average maximum price for each cuisine by vegetarian diet and vegan diet.

Additionally, since our project interest focused heavily on cuisine and diet type (in this dataset's case, it came down to either vegan, vegetarian, or restaurants that served both), we had to clean the columns called "categories" and "cuisines" which had all the relevant information in that regard. Both of these variables were quite tricky because each one was a string value that was formatted and structured very differently from the next. For the column "category", each restaurant had a long string of information consisting of diet type, cuisine, type of food, etc. We ran a Python script to parse out dietary information (vegan, vegetarian, both, etc.) from the categories column and to consolidate different verbiage for one cuisine type into one consistent variable; for instance, "Thai,Vegetarian" and "Thai,Asian/Pacific,Vegetarian" were all converted to a single "Thai" cuisine type in order to make the dataset more readable and easier to use. After cleaning the dataset, we counted the number of each cuisine and diet type into a new dataset that we read into our index.html file to build our bar charts.

US Population Dataset

In order to determine the color of the state using the ratio of vegan/vegetarian restaurants per 1 million residents, we used a U.S. Census dataset for the US Population in 2019. Overall, this dataset was already pretty usable. We only kept the columns "NAME", "abbreviation", "population_estimate2019", and "population_rank2019" which are mostly likely to come in handy later on when we build out the map visualization and associated interactions. Only the 50 U.S. states were kept and no U.S. territory was included in order to make sure our map is of a meaningful size and scale based on the restaurant data that we already had.

DESIGN RATIONALE

U.S. Map

The goal of the U.S. map is to visualize the geographic distribution of vegan and vegetarian restaurants. We used orange circle marks to represent individual restaurants. We used horizontal and vertical aligned positions to convey the geographic locations (latitude and longitude) of the restaurants. All circle marks have the same radius which scales according to the map's zoom level (e.g., circles grow smaller as the map is more zoomed in to allow for differentiation of restaurants in close geographic proximity). Upon hovering, a circle turns red and its radius increases to clearly demonstrate to the user which circle they're selecting as a text box appears to share more information about the selected restaurant.

For the state geometric shapes making up the map, we used a multi-hue, sequential, and colorblind-safe blue ColorBrewer scale. We used color saturation and luminance as the primary channels to depict density of veg/vegan restaurants for each state. Density was calculated by dividing each state's population in millions by its number of restaurants. We chose per million instead of per hundred thousands or a smaller denominator in order to avoid meaninglessly small ratios such as 5.1E+7 restaurants. A quantile scale was applied for even distribution of states into

four density levels. County outlines subtly but clearly show up in grey after the map surpasses a certain zoom threshold, allowing the user to get more granular position information.

Double Lollipop/Barbell Chart

The goal of the double lollipop chart is to visualize the difference between the average maximum prices of vegan and vegetarian restaurants by cuisine. We used blue and orange circle marks in order to visualize the pricing of the vegetarian and vegan restaurants, respectively. We used blue and orange here as channels to represent vegetarian and vegan restaurants in order to effectively show the difference between the pricing per cuisine. We chose to use these colors here (and consistently throughout our visualizations) because we want to be more inclusive of various forms of color-blindness; blue and orange are contrasting colors that are more visually accessible compared to many other different combinations of colors. Furthermore, another mark we used is a vertical line to connect the two vegetarian and vegan circle points. This helps the viewer connect the difference between two points, especially if they are separated by a larger price difference which may result in a more unclear vertical alignment to the cuisine type on the x-axis. The vertical position of the circle plots is used as a channel to visualize price and price difference (y axis is price). The horizontal position of the circle plots is used as a channel to show the cuisine type the circle plots are representative of. We used a log scale for the price y-axis because there is one price value significantly larger than others – the log scale helps show the difference between the other points for each cuisine more clearly.

Bar Charts

We decided to use a standard bar chart to reflect the data for the dietary filters. We wanted to make sure that the visual design of the dietary barchart matched the cuisine bar chart.The marks are the orange rectangles to indicate the counts for the number of restaurants associated with a given dietary restriction such as vegan, vegetarian or both. The channels of the dietary bar chart of each bar are horizontal aligned position and varying vertical aligned length. We decided to include a linear scale as well as a scaleBand in the creation of our bar chart to accurately reflect the count data associated with each value.

INTERACTIVE ELEMENTS

In an effort to showcase the skills we learned this semester, we decided to incorporate a variety of interactive elements into our final project. The first element of interactivity corresponds directly to the map visualization. We provided a zoom in function that allows the user to view more precisely the location of each restaurant. We wanted to include this feature to demonstrate a higher level of understanding from the course, but also to give the user more flexibility and to amplify explicit details on the map. This allows the user to view precisely the location of each restaurant in accordance with their current location. Furthermore, when the user zooms, the color

scheme stays consistent with the visualization. The zoom feature also allows users who have poorer eyesight to deep dive into the map and navigate through the U.S.

The second element of interactivity we explored was located on the map as well. We decided to include a mouseover feature for the corresponding data points on the map. This text box includes restaurant details about each point such as type of cuisine, state and city. We decided to position the textbook in the left hand corner of the screen because users typically read from left to right so we thought this would be very intuitive for the user. Additionally, when the user hovers over a dot on the map, the dot changes color from orange to navy, indicating that the dot has been selected. We created this feature because some of the dots overlap a little bit and we wanted to make sure that the text box and additional details correspond with the correct data point on the map.

The third element of interactivity we included in our visualization was a filtering option. A user can search for restaurants related to their dietary restrictions as well as their preferred cuisine type. Once the user selects a filter, the corresponding points related to that filter will populate the map. In addition, we included an “All Dietary” or “All Cuisine” button which would display all points on the map. This would be an easy way for users to revert back to the static map view indicating all restaurants! We decided to include this filter to allow the user more flexibility in their restaurant search instead of trying to sift through all the points on the map. Furthermore, once a filter is selected the color will change to a lighter blue to indicate that the filter has been selected. We incorporated this interactive feature to clearly inform the user what value they selected and which points on the map correspond to that filter.

The fourth interactive element we implemented is directly related to the filters and the corresponding bar charts to identify the count of each dietary restriction or cuisine. When the user selects a given feature, the corresponding bar on the bar chart is highlighted by changing colors from orange to blue. This indicates to the user additional information regarding that dietary restriction as well as the cuisine. The user will click on for example thai, and the bar on the cuisine bar chart will change to navy indicating the number of restaurants there are of the thai variety. We included the number of restaurants associated with that cuisine or dietary restriction to give a more accurate representation to the user instead of them having to make assumptions by the y axis. We chose this interactive element because we thought it would be interesting and unique to provide some sort of pop-up feature for the user!

The final element of interactive we implemented is a hover over feature on our lollipop graph. When a circle is hovered over, the circle size increases as well as the black outline. Not only does the circle size increase, a text box also pops up to indicate what each point represents. The mouseover text box includes information related to the diet, type of cuisine as well as the average max/min price for each of the top 10 cuisines. When creating this interactive feature, we

thought about just including the average price, but we wanted to give the user more information to inform their decisions about a given restaurant/cuisine.

Overall, all of these elements of interactive make our visualization more accessible, interesting and advanced for the user!

THE STORY

We successfully answered our first question by elucidating the states that offer the most number of veg/vegan restaurants per million residents: Arizona, California, Hawaii, Illinois, Maine, New Hampshire, and New York. So, if one is looking for a vibrant green-eating scene, chances are these would be good states to explore! Interestingly, the relative density of veg/vegetarian restaurants in a state does not seem to linearly correlate with the state's population since California, Illinois, and New York are among the top 6 most populous states while Maine, New Hampshire, and Hawaii are among the 11 least populous states. This brings us to a follow-up question which would need to be explored through additional visualizations and statistical testing: Is there any relationship between the number of veg/vegan restaurants in a state and its population?

We also successfully elucidated insights for the second and third questions we sought out to answer: Among the top 10 most popular cuisines, Indian and American cuisines are by far the two most popular for existing veg/vegan restaurants. This makes sense if we consider that India is ranked the top "vegetarian country" in the world with a third of its total population being vegetarian. As expected, there are significantly more vegetarian restaurants than vegan restaurants. Surprisingly, for a majority of the 10 cuisines, vegetarian restaurants were more expensive than vegan restaurants. This was surprising to us because we had assumed that vegan restaurants would have higher price points because veganism is a more restrictive diet than vegetarianism. We would be curious to do this vegetarian-vegan price comparison with a larger dataset of restaurants.

TEAM CONTRIBUTIONS

Cynthia:

- Created and implemented color scale for states based on restaurant-to-popultaion ratio
- Created map legend
- Implemented zoom and pan functionality for dots on the map
- Wrote introduction and findings sections of the write up
- Contributed to design rationale portion of the write up
- Total time spent on project: 13
    - Write up: 2 hours

    - Meetings: 4 hours
    - Coding: 7 hours (4 - map, 1 - map legend, 1 - dots zoom, 1 - formatting/misc.)

Emily:

- Cleaned Vegetarian and Vegan restaurant dataset; wrote Python scripts to pre-process restaurant data and to calculate average price for lollipop chart
- Cleaned US population data
- Implemented mouseover details for restaurant plots on U.S. map
- Created top 10 cuisines distribution bar chart
- Created double lollipop/bar chart, its legend, and mouseover interactions
- Implemented bar chart bar highlighting for both dietary and cuisine filters
- Helped with CSS styling (annotations, bar labels, positioning, sizing, etc.)
- Wrote data processing section of the write up
- Contributed to design rationale portion of the write up
- Total Time spent on project: 15
    - Write up: 1 hour
    - Meetings: 4 hours
    - Coding: 10 hours (3 - data processing, 6 - visualizations, 1 - stylistic details/wrap-up)

Carolyn:

- Created dietary filter to filter corresponding points on map
- Created cuisine filter to filter corresponding points on map
- Created json for the dictionary related to the count of dietary filters
- Created & styled Dietary bar chart
- Assisted in styling for the cuisine bar chart
- Helped with overall CSS styling
- Wrote Interactive section of the write up
- Contributed to design rational portion of the write up
- Total Time spent on project: 14
    - Write up: 1 hour
    - Meetings: 4 hours
    - Coding: 9
    - I spent the longest time adding in the filters and making sure that the filters corresponded to the points on the map! I also spent a while trying to style the graphs and overall look of our data visualizations

REFERENCES

ColorBrewer
Datafiniti's Business Database
Gallup Poll
Vegetarian and Vegan Restaurants Kaggle Data
US Population 2019