SBN8: Boston Local Food Festival Data Vizualization. DS 4200 S20

Dan Krasnonosenkikh, **Christian Yiu**, **Jacob Hoffman**, **Tian Xia**

Service-Learning Course Project as part of [DS 4200: Information Presentation and Visualization](), taught by [Prof. Dylan Cashman](), [Data Visualization @ Khoury](), [Northeastern University]().

Motivation

The following visualizations were created in collaboration with SBN in order help show off what SBN has been able to accomplish so far to potential sponsors. It is important for SBN to of acquire funding from organizations other than the State of Massachusetts in order for their organization to be more financially flexible. Our hope is that our visualization will be able to further highlight the successes of the food festival, especially for the further advancement of their mission statement.

One of the main things that SBN emphasizes is their diversity in their demographics. It is also a key feature that they are proud of wish to show off. As a result, our main focus will be to navigate the

user through the our visualization to highlight their diversity.

Visualization

Include the interactive visualization as part of this page. Static example follows.

■Festival Attendee
■Selected

■Festival Attendee
■Selected

Attendees Race Distribution

Attendees Age Distribution

Attendees Income Distribution

Demo Video

Embedded MP4 demo video using the HTML5 <video> tag. For example, this screen recording Prof. Cody Dunne made of Mike Bostock's flexible transitions in D3 slide:

Visualization explanation

Final visualization screenshots (PNG images), design justifications, UI walk-through, and linked presentation slides.

Included in Data Analysis section, see below.

Data Analysis

Summary of data, data types, and data preprocessing.

Expectation: Data sources match the problem statement and are appropriate. These descriptions should be very explicit so someone could read your page and properly reproduce your results.

We heavily utilized the zip-code data to locate attendees, and to act as bins for multiple attendees from the same area. Additionally, we

use the income level (categorical), race (categorical), and favorite experiences (activity/cuisines) from the food festival (qualitative, non-categorical). The data is sourced from surveys handed out during the food festival and was then given to us in an excel file. Some preparation/preprocessing in excel was required to make managing the data go smoothly.

Data Preprocessing Steps:
1. Normalization of columns that exhibited a One-to-Many relationship
   - Several columns contained data which violated First Normal Form (1NF), which states that each table cell should only have one discrete piece of data
   - Solution was to create new tables for each column that displayed this One-to-Many relationship and have those tables reference each attendee ID (foreign key).
2. Addition of longitude and latitude data
   - In addition to the zip-code data that was provided for each attendee, we also added the corresponding longitude and latitude coordinates for each zip-code in order to gain a more accurate location mapping in our final vizualization.
   - There are numerous online tools available for mapping US zip-codes to specific geographic coordinates, the addition of the longitude and latitude columns was done manually with the help of these tools.

Task Analysis

Summary of task table.

Expectation: Clearly describes domain tasks, processes, goals and abstract tasks for domain problems.

There are three main tasks that our vizualization hopes to accomplish.

1. Demographic Analysis
    - Emphasizing the wide array of demographic backgrounds across all the food festival attendees is an important point of upsell to potential future sponsors. As a result, our vizualization should provide an interactive, user-friendly way to display the different categories of demographic information available to festival organizers (i.e. Race, Income levels, Attendee point of origin, etc.)

Design Process

Sketches and design choices to justify final visualization.

Expectation: Evidence of iterative improvement. Logical discussion of design choices grounded in theory from course. Discusses feedback from usability testing.

From our original conversations with SBN, we knew we wanted to focus on showing the demographics of attendees. Looking through the data, the best options to show this were the zip-code location, race, age, and income data.

The zipcode location lent itself well to a map-based visualization. For the other three sets of data, we were originally unsure of how we wanted to implement them and after some thought settled on bar charts as they were set into a few categories that lent themselves well to bar chart formats. We then implemented brushing and one-way linking to enable the user to better see the demographics of the attendees coming from different locations. [Add more about design choices/theory]

Conclusion

Short summary of work completed and areas for improvement/future-work.

Expectation: Meaningfully wraps up project and has good future directions.

Our final visualization shows the demogrpahics of attendees from a particular zipcode. Utilizing a map with brushing that links to three bar charts with those demographics we provide a way of better understanding the festival attendees.

If we were to continue work on this project, we would further improve usability. We would also consider implementing a visualization to show more about the attendees experience at the festival. A better understanding of the demographics and their preferences would hopefully provide many opportunites to glean new insights.

Acknowledgments

List here where any code, packages/libraries, text, images, designs, etc. that you leverage come from.

- [All US zip codes with their corresponding latitude and longitude coordinates.](#) from United States Census Bureau
- [In-class d3 tutorial](#)> by Dylan Cashman

  >
- [D3: Data-Driven Documents](#) by Mike Bostock.
- [Pure CSS responsive "Fork me on GitHub" ribbon](#) by Chris Heilmann.