CS/INFO 3300 - Project 2: Final Writeup

Theodore Carrel (tjc233) , Heather Zhu (yz927) , & Eric Sun (ejs336)

Apr 11, 2019

A) Datasets

The data for the restaurant week and the inspection violations were both taken from Kaggle.
NYC Restaurant Week 2018 dataset
NYC Restaurant Inspections dataset

The geojson data for the NYC map was taken from data.betaNYC website
http://data.beta.nyc/dataset/pediacities-nyc-neighborhoods/resource/35dd04fb-81b3-479b-a074-a27a37888ce7?fbclid=IwAR0tWE1gTEeaAWqyTF9Dm6s36in6dzFVrfORgCZver5T1l__6xSpnKTq8dQ

The restaurant week dataset was from 2018. It has 348 restaurants, each with a ton of variables including reviews for ambience, food, service, etc., latitude and longitudinal coordinates, name, address, and phone number of the location, and more. Of those variables we primarily used the name, location (address and coordinates), phone number, and all the review information including the number of stars it got from each review.

The inspection dataset is from 2010-2017 and contains a ton of information about most of the restaurants in NYC. Each violation included a lot of variables such as date, code, grade, score, address/name/cuisine type of the location, and whether the violation was critical or not. We focused primarily on the date, description and type of violation for each restaurant.

The geojson dataset contains the features of NYC counties. We used this to draw the map.

Data Processing

Data preprocessing was heavily involved in this assignment. The first thing we focused on was identifying how many violations in the violation dataset was actually necessary. Because our restaurant week dataset was only a small subset of all the restaurants in NYC, we had to remove all non-relevant violations in the dataset which significantly improved our processing time. We did this in python using pandas.

While doing this, we also counted how many violations each restaurant had and sorted them so that they would be organized by date. Finally, we also used matplotlib to make a histogram of the number of violations for restaurants so we had an idea for how useful a slider for the map for violations would be.

B) Visual Design Rationale

Because our dataset involved a map and the details of each point was focused around individual restaurants, we used mouseover and clicks to display more specific information about each circle. In order to display this deeper information, we decided to divide our visualization into three parts: the overall map view, a more specific line graph over time displaying the violations, and a text overflow that allows the user to read the violations each restaurant got. Our end goal was to allow a user to look at these restaurants and see when/what kind of violations they got, and to potentially use this data as a way to help decide whether they would want to go to that location or not.

To display the restaurants, we used circles as a mark and the color as a channel to indicate the number of violations they got. The double use of having the color as a channel as well as the slider being a histogram allows the user to understand how many restaurants they'd be filtering by using the slider, and helps them identify restaurants with more/less violations. Users can also zoom and drag on the map, which allows them to identify targets.

We chose to include the entire NYC map instead of just focusing on Manhattan despite most of the restaurants being there because we felt that having the other counties helped the user visualize the map. Also, not all of the restaurants were in NYC, and we wanted to include them all. In order to help the user, upon loading the webpage the map automatically zooms in on Manhattan, though the user can zoom out if they want to.

A key was created in the top left corner of the map to explain the color scheme to the users. Restaurants that did not have any violations were all colored the same shade of blue. A gradient red scale was developed for use with restaurants with at least one violation. The scale gets progressively darker as the number of violations increase. This was done so that users can quickly distinguish on the map which restaurants had zero violations, which had a small amount, and which had greater amounts.

In displaying the violations, we used a line graph, with the circles indicating the month the violation occurred and the color indicating whether a critical violation occurred or not (red for critical, black for non). This gives the user a visual scale of how many violations a restaurant has had and when they got it. We chose to add this inclusion because it drastically helps the user in organizing how a restaurant is performing in terms of number of violations.

To allow for further inspection of detail, we added an overflow text at the bottom that allows the user to explore the ratings of the restaurant, its location, and all the violations that it has received. Upon clicking a circle in the line graph, the document will automatically scroll to the violations that occurred in that month. This decision was made to allow the user to have a clear connection between the visual element of the number of violations a restaurant got to the actual description of the violations. The text is also either red or black to indicate critical or non-critical. We chose to do this because we felt that users would be more interested to see critical violations.

For the text, we decided to display the ratings as stars filled in for how good their overall review was for each category. We decided to use stars instead of just the number value or a graph because the stars indicated the ratings better.

C) Interactive Elements

The interactive elements we have are: circles on the map can be clicked on to display more information, mousing over a restaurant displays its name and violations, the map can be zoomed and dragged around, clicking and dragging on the violation slider allows users to filter restaurants, on the line graph circles can be clicked on and the document will automatically scroll to that part of the page, and mousing over the review stars tells the user how many of each star the restaurant got.

For the circles, we choose to make them expand in size and make the cursor look like it can be clicked. Also, at the beginning of the page load, the bottom text encourages the user to click on a circle. We know that it's not a perfect system but we believe that the visual cues we give are enough to suggest to the user to try clicking on a circle.

Our violation slider has the text "violation slider" above it, which encourages the user to try clicking on it. Also, mousing over it causes the cursor to change format. The user can utilize the brush feature to select a portion of the histogram, creating boundaries for a filter. The map automatically updates and only displays the restaurants whose number of violations fit within the selected boundary. The user can drag the selected area to a different part of the histogram, or click outside of the selected area to reset the feature and display all restaurants.

When a user clicks on a restaurant circle, the line graph appears, and with it text that encourages the user to click on a node in the graph. This text disappears to help keep the page stylized, but we hope that the user is able to see it before it disappears. Upon mousing over the nodes, the user is also told that red nodes is critical and black is non-critical, so this allows for communication of a "legend" without actually having to add another legend. This idea of red and black being critical/non-critical also translates to the text as well, so we don't have to explain that again.

One thing that we had a hard time encouraging the user to do was to mouse over the stars. In the end, we decided that because the information of the number of stars isn't as relevant as the violations (or at least, wasn't the focus of our visualization), we left it as something that a curious user might attempt and be able to find more information about. We also left it because we believe that even without the details of how many stars each restaurant got the overall ratings would be enough to convey the quality of the restaurant.

D) Story

We made our visualization to tell the story of some of the best restaurants in New York City and their quality. However, instead of focusing on reviews and critics, we decided to focus on the violations they've had and the things that might be wrong with their food. Through our visualization we hope to allow the user to gain some insight into what might be wrong with how some of these restaurants prepare their food and perhaps help them choose where they want to eat.

Some surprising things we found was just how many of these restaurants had a lot of violations. It was a little shocking to see that so many of these restaurants, most of which had greater than 4 stars in every category, had 20 or more violations. One restaurant (Obao) had 80 violations but still pretty good reviews, which was pretty incredible.

The insights we attempted to convey with our visualization was how the top restaurants performed with their safety rules and how many violations they committed.

Team Contributions

Eric

- All data preprocessing
    - Removing all non-relevant violations
    - Calculating the number of violations each restaurant got
    - Sorting the violations by date
- Display restaurants on the map
- Zoom/drag, scaling zoom with circles
- Tooltip mouseovers
- Bottom text overflow
    - Creating the stars/gradient needed to fill part of the star
    - Formatting/displaying the violation text
    - On click of line graph scrolling to the corresponding date in the text

The parts of the work I accomplished that took the most amount of time was the data preprocessing to sort and remove violations, finding out how to make the star symbols/how to display only part of them with gradients, making the click of the line graph correspond to the correct date in the violation text, and general style tweaks with formatting the html elements together.

Theo

- Implemented the slider
    - Filtered out results based on the restaurant's number of violations

- Developed the color scale based on number of violations
    - Assigned each data point the necessary shade
- Created the legend for the map

The part of the project I spent the most time on was implementing the slider. There was some difficulty creating the histogram and making sure the brush feature was working correctly with filtering the data. Despite a couple small issues everything was able to be resolved and the feature works great for showing certain restaurants based on a user-defined boundary on the number of violations received.

Heather

- Set up NYC map
    - Showing all neighborhoods with clear borders
    - Implemented the hover over interaction to show each neighborhood's name
    - Adjusted, zoomed, and rotated the map to set the map's default position to exhibit circles better
- Implemented the line graph for each circle after click
    - Calculated the number of violation occurrences for each restaurant within each month
    - Created the scales corresponding to the maximum value of violations in each month and the earliest + latest time the violations occurred for each restaurant
    - Drew the trend chart (lines, circles)
    - Transition animation
- General html formatting

The part of the project I spent most time on was making the line graph. For each circle clicked, I extracted the month for each violation record and then counted the number of violations in every month. After that, I reformatted the time information I got earlier, and put it into d3 timescale function. Collecting and processing data for each restaurant after clicks and playing around with d3 time scale were the tricky parts.