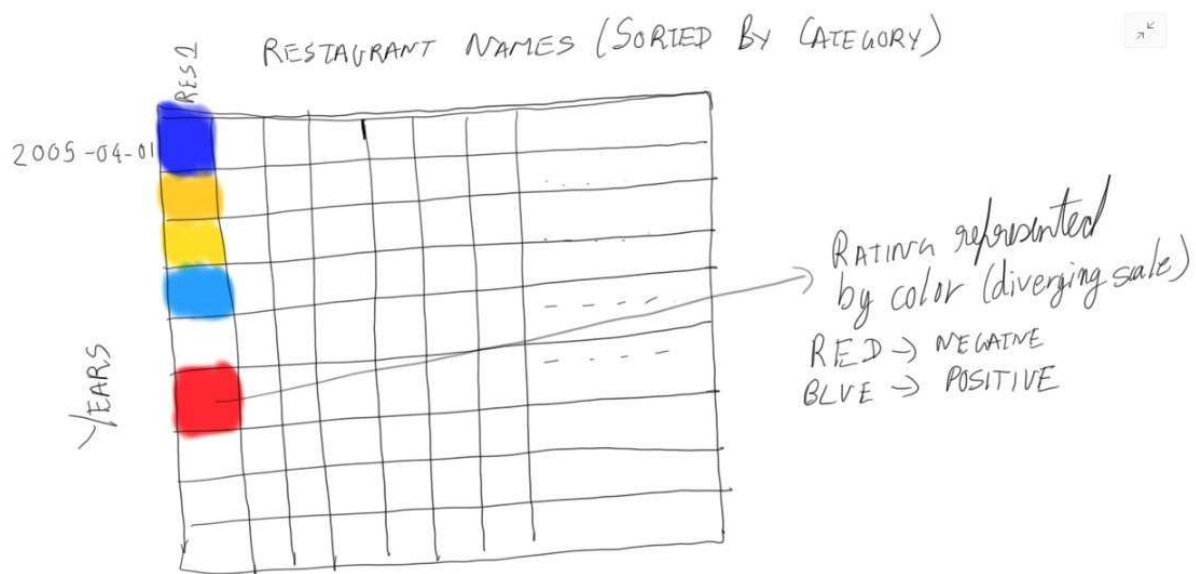INFORMATION VISUALIZATION FALL 2017
FINAL PROJECT
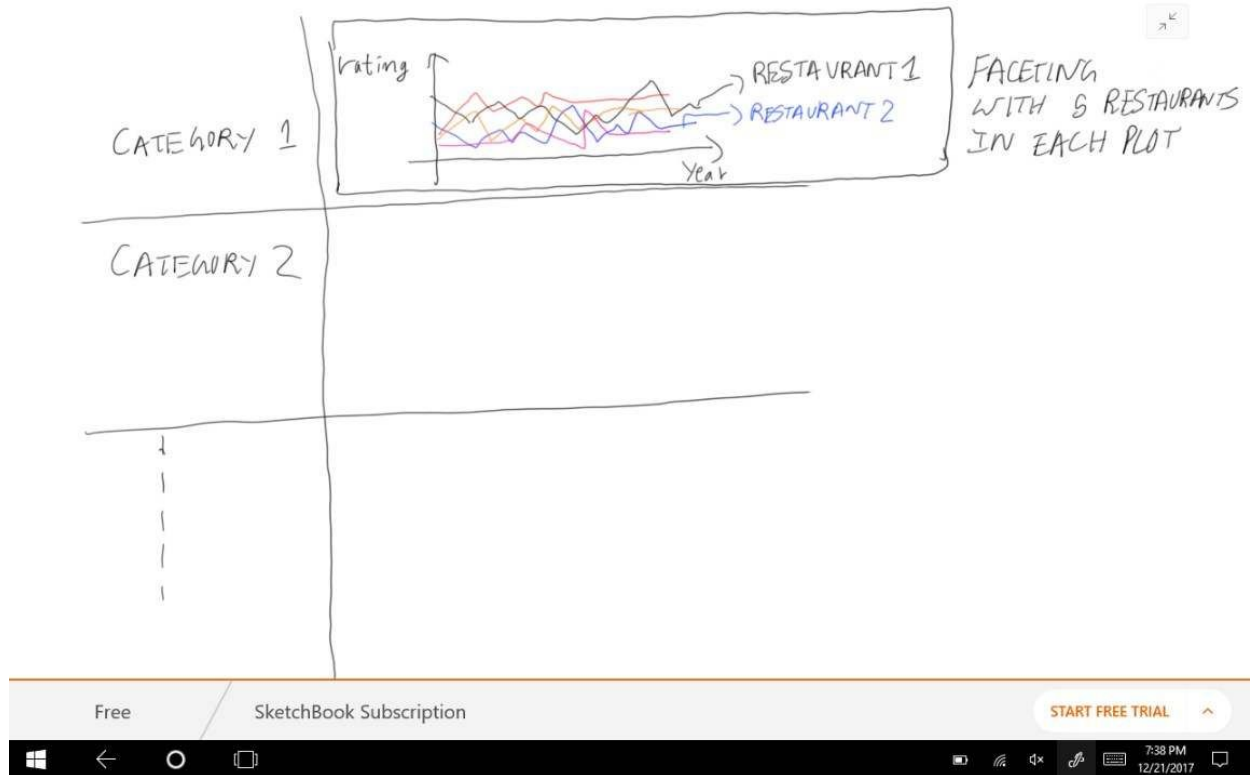
Santhoshkrishnachaitanya Chelikavada
N11408956/sc5612

Sketches

Sketch 1 - Matrix Visualization

Sketch 2 - Line Graph with Faceting

CATEGORY 1

CATEGORY 2

rating

RESTAURANT 1
RESTAURANT 2

Year

FACETING
WITH 5 RESTAURANTS
IN EACH PLOT

Discussion

Sketch 1 -
One of the biggest advantages that sketch 1 presents is the fact that a restaurant, year pair can be easily identified because position is used as a channel and the rating can be estimated using the color hue. Other than that, the fact there is control over the ordering of the years and the restaurants makes it possible to categorize the restaurants under their respective categories. It's also great that one plot (the matrix visualization plot) is enough to represent all the information and answer the questions posed. It is also very easy to compare ratings of restaurants over time. This visualization also allows for the distribution of ratings over time for categories to be understood, because the ordered list of categories will let the viewer know "how much" of the graph represents a category. For example, the first 35 columns would represent the American food category and the viewer can estimate distribution of ratings by looking at values in this 35 column block.

A possible disadvantage is the fact that when all the restaurants are put up on the matrix visualization plot, it could be hard to discern one restaurant,year pair from its neighboring pairs. However, this problem was handled by using a circle within the rectangles representing a restaurant, year pair and varying it's area and color to represent the rating that the restaurant has. Another disadvantage is that it is hard to compare values between restaurants that are too far apart.

Sketch 2 -
The advantage with sketch 2 was that it allowed for comparisons between ratings of restaurants that maybe far apart (4 places away) within the same category which is more than what sketch 1 could offer. The other advantage was that because only 5 restaurants were compared at the same time, it would be easier infer the ratings of a restaurant without channels interfering.

The biggest disadvantage is it will be hard to keep track of where a restaurant is on the visualization. Though faceting with small multiples solves the problem of the line graph having too many lines, it makes it hard to find out where a restaurant is. One has to look through every single line graph to find out where a restaurant is. Moreover, comparing one multiple of the small multiples visualization with another multiple is hard - that is it is very hard to compare the ratings between a restaurant in one of the multiples with a restaurant in a different multiple, over time.

I chose Sketch 1, because I felt it answered all the questions that are to be answered and it did so better than Sketch 2. For one, all the data is specified within one visualization whereas with Sketch 2, there are multiple plots one has to look at. I also felt Sketch 1 does a much better job at presenting trends within categories as a block than Sketch 2 does, as one has to somehow perceive trends in Sketch 2 by looking at all the different multiples that are not placed together which makes the task less intuitive. Sketch 1 on the other hand allows the viewer to look at the trends within a block of columns, making it easier to perceive trends within categories, intuitively.

Moreover, the data had some missing data points. This meant that the line graph for all restaurants would not be uniform and plotting the data would have meant that lines would have had empty breaks in the middle. With the matrix visualization, breaks in the data don't break the visualization like breaks do with line graphs, as if a rectangle corresponding to a restaurant, year pair is empty, one can understand that there is no data point to represent that pair. With Sketch 2, this would have looked awkward and would have spoiled the visualization entirely.

Implementation

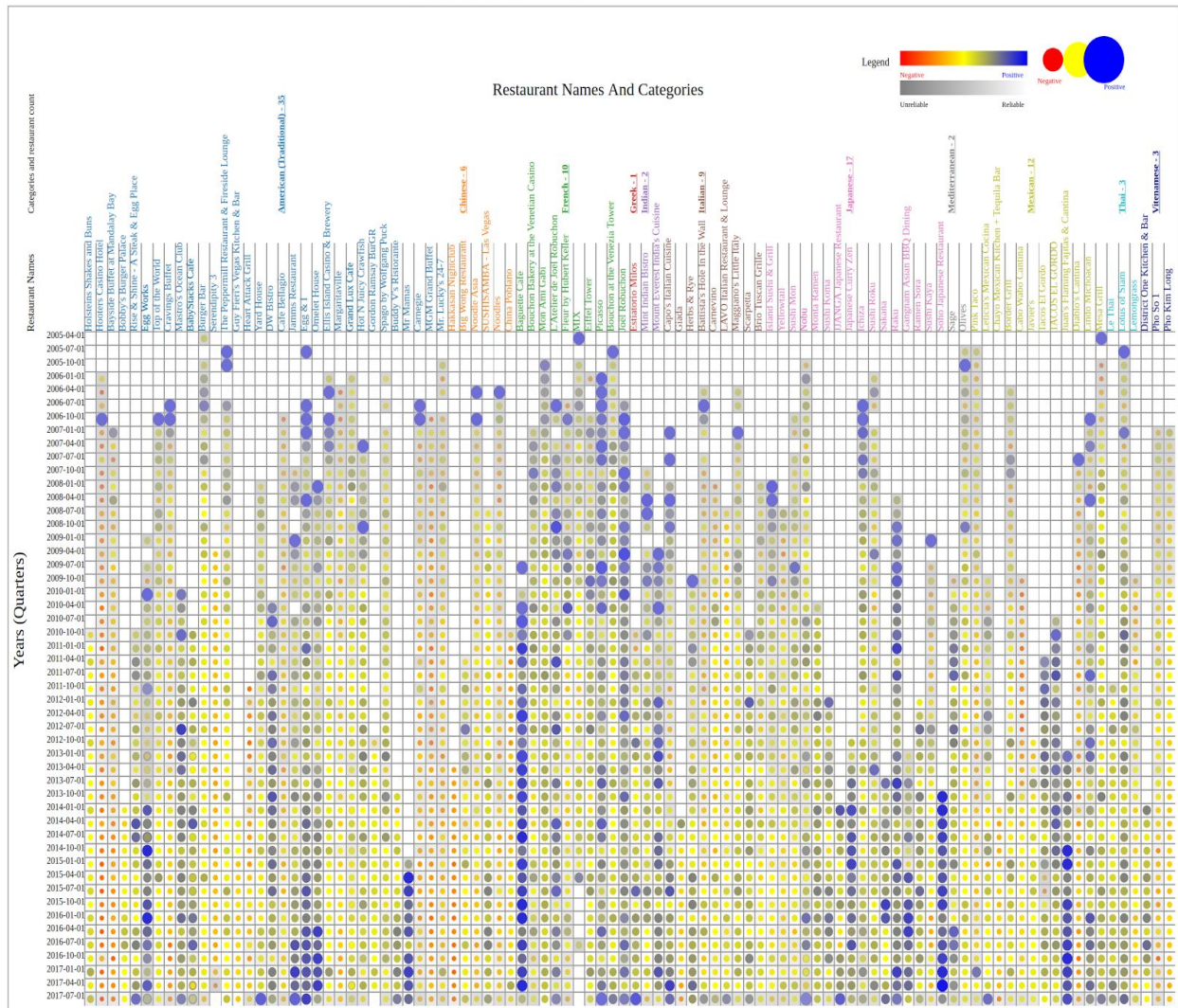**The GitHub repository containing the code can be found here**
All the pictures can be found in this Google Photos album
Please click on the individual pictures to be redirected to a better looking, higher resolution version of the photo.

Restaurants And Their Rating Over Time

Matrix Visualization Representation

Santhoshkrisnachaitanya Chelikavada (sc5612)

Challenges and changes
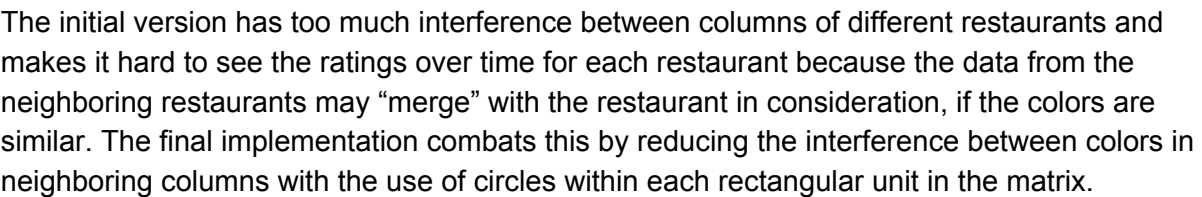
The final implementation uses circles within each year,restaurant pair's rectangle and depicts the rating using the radius and color - this is a departure from Sketch 1 which uses just the area of the rectangle and depicts rating with color hue. The reason I made this change was to reduce the interference between colors in different columns.
This was the initial version that I had built -

Restaurants And Their Rating Over Time

Matrix Visualization Representation

The initial version has too much interference between columns of different restaurants and makes it hard to see the ratings over time for each restaurant because the data from the neighboring restaurants may "merge" with the restaurant in consideration, if the colors are similar. The final implementation combats this by reducing the interference between colors in neighboring columns with the use of circles within each rectangular unit in the matrix.

After using circles, the implementation looked like this -

Restaurants And Their Rating Over Time

Matrix Visualization Representation

Santhoshkrisnachaitanya Chelikavada (sc5612)

What I found was that, once I used circles within rectangular units within the matrix I could now use the area of the rectangle to also represent whether the data was reliable or not. If the number of ratings were lower than 40 (the average number of ratings), I colored the rectangular units within the matrix grey to represent how "reliable it is". If it had one rating, it would be of a darker shade than a cell with 20 ratings is.
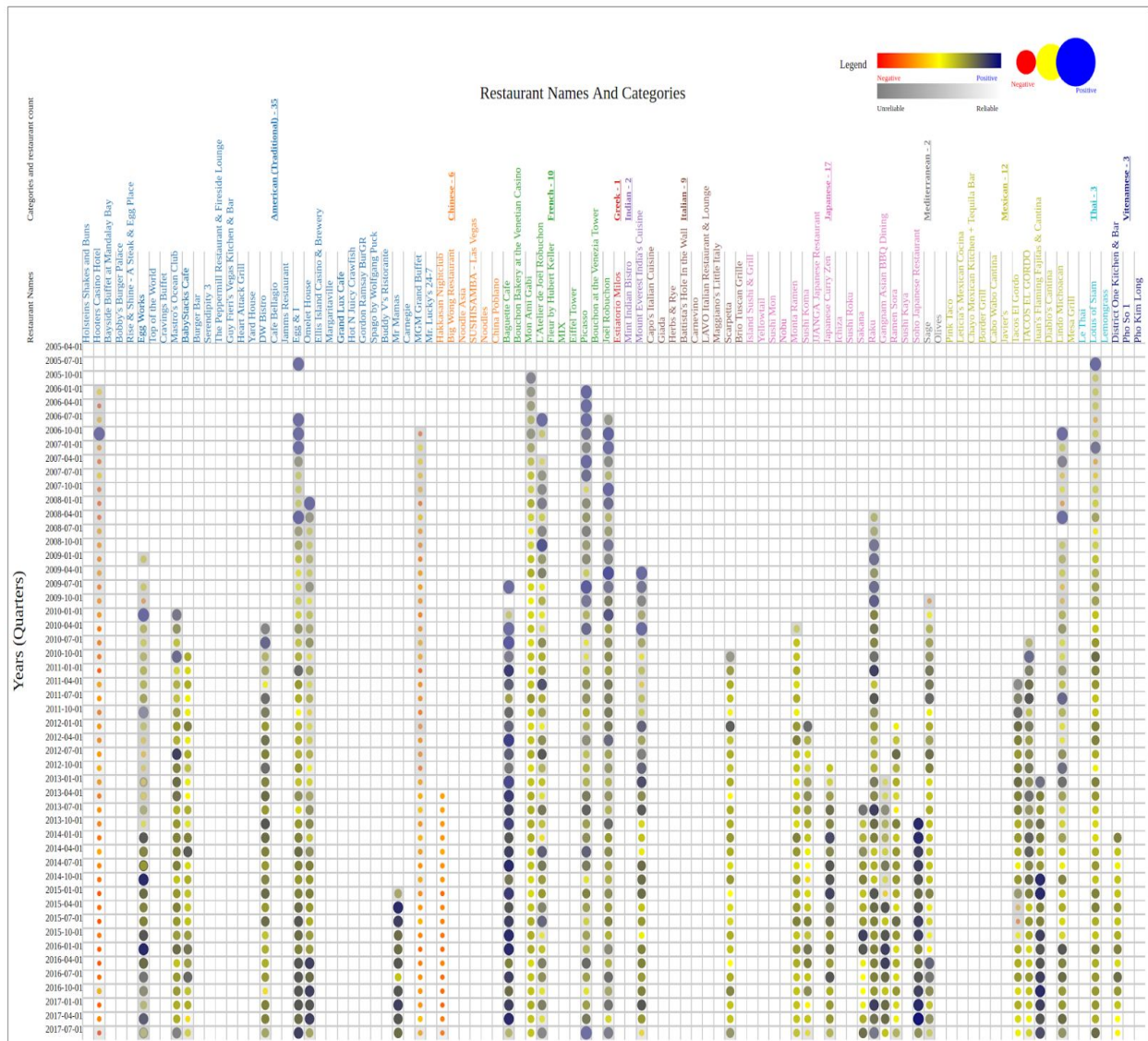
A big challenge I faced was that it was hard to use the json data as is to generate the visualization. So I had to write multiple functions that converted the json data to multiple formats to build the different components of the visualization.

Data Analysis Question 1: "Are there restaurants that have consistently negative reviews?"
The picture for this question can be found [here](#)

Restaurants And Their Rating Over Time

Matrix Visualization Representation

The above visualization displays a restaurant's data only if it has consistently had a positive rating for at least 60% of the time or has had a negative rating for at least 40% of the time.

Data Analysis Question 2: Are there restaurants for which in some time periods their rating drops considerably?

The picture for this question can be found here

Restaurants And Their Rating Over Time

Matrix Visualization Representation
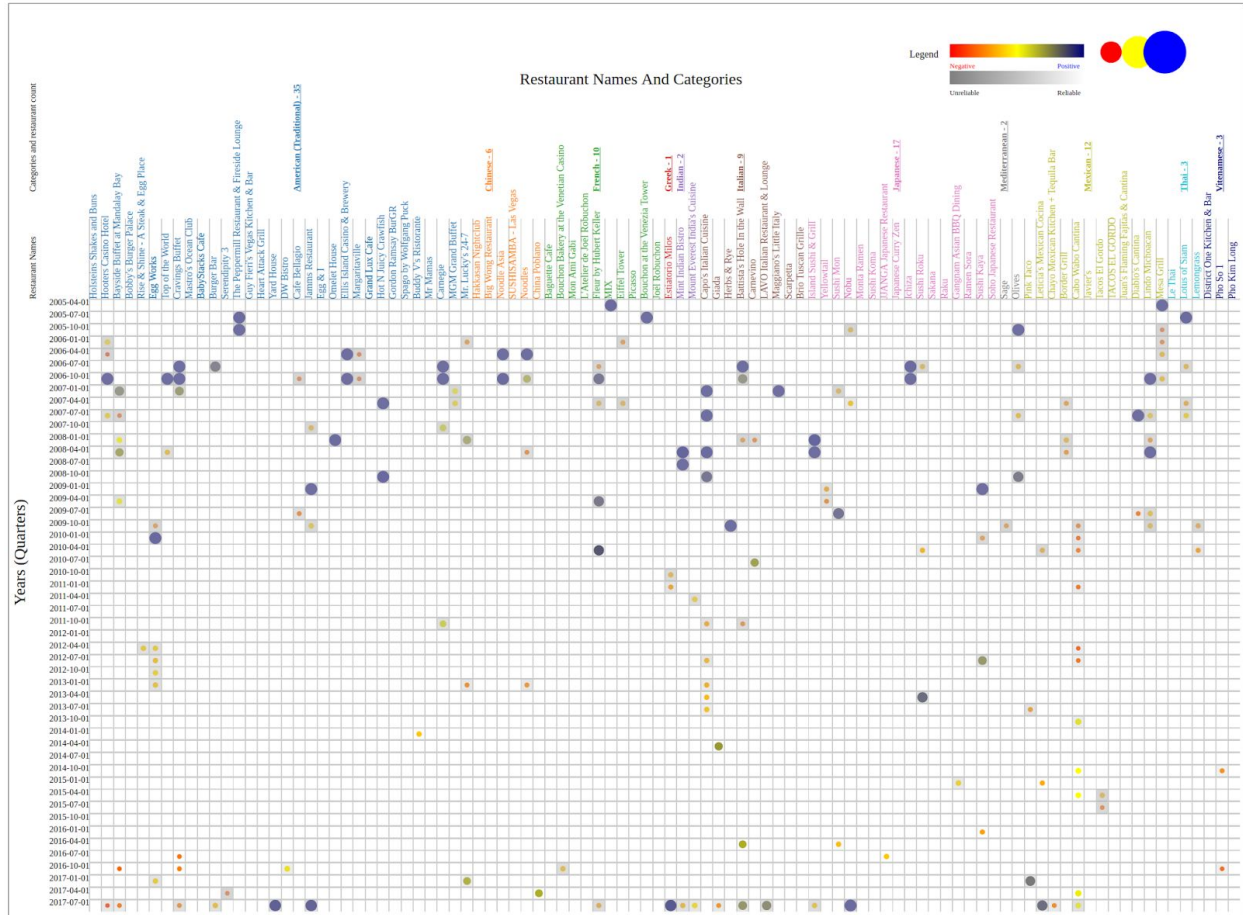
The above visualization displays a restaurant's data for a particular quarter, if that rating exceeds or goes below that restaurant's average rating by 0.75.