Yelp Business Trajectories

An visualization of popularity of businesses on Yelp

Haoqin He	haoqin.he@nyu.edu	hh1536
Jiawei Tang	jt875@nyu.edu	jt875
Kenan Mao	kenan.mao@nyu.edu	km3734
Haonan Wang	hw1375@nyu.edu	hw1375

Project page (on Github): https://github.com/NYU-CS6313-Fall16/Yelp-Business-Trajectories-7
Video: https://youtu.be/nl9HSEH0tJo
Working demo: https://nyu-cs6313-fall16.github.io/Yelp-Business-Trajectories-7/

What is the problem you want to solve and who has this problem?

There are a lot of businesses are Yelp and their popularity changes over time. Sometimes a lot of customers are writing reviews for a certain business and sometimes less. We'd like to understand the reasons behind this kind of change.

One assumption is that the type of food a restaurant serves will affect its popularity since cuisine trend itself is changing overtime. The other assumption is that reviews of Yelp Elite Users have impact on business popularity. We've observed that some Yelp users will be awarded as Elites due to their contribution to the community. Most of them write high-quality reviews and have many followers. We assume review of this kind of user will certainly affect decision-making of normal users thus influence the popularity of business.

First, the answer of this problem can be helpful to business managers, since they certainly want to improve the popularity of their business. Besides, the answer can also be helpful for Yelp or similar companies so that they can understand whether this kind of product itself is affecting and able to improve the popularity of business and how they can achieve that.

What are the driving analytical questions you want to be able to answer with your visualization?

Make a list of questions you want to be able to answer with your data analysis and visualization. Make a list starting with each question followed by a more detailed description of what the question means. This should be expressed in the domain language, not in abstract terms (for instance I prefer something like "find customers with a similar behavior" than "cluster the data points according to their similarity")

- What's the popularity of each main food type over time?
    - Before discussion of relation between food type and business, we need to know the trend of popularity of each food type. We'd like to understand how does popularity food type change over year? Is it increasing or decreasing?
- Does popularity of a restaurant follows the trend of its food type?
    - Since we need to understand the relation between food type and restaurant, we need to compare the trend of food type with the trend of popularity. If the two trend looks similar, it means the restaurant is following the trend of its food type.
- Do reviews of Yelp Elite Users affect popularity of a business?
    - This question aims at discussing the impact of Yelp Elite Users on business. First we need to understand the trend of a popularity of a business. Then we need to look into the reviews of Yelp Elite Users. Does popularity decreases after a bad review? Or does popularity increases after a good review?

What does your data look like? Where does it come from? What real-world phenomena does it capture?

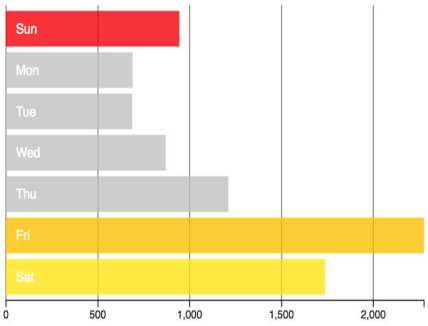
name	type	description	range/category
type	Categorical	Food types. Aggregated from original category of business in the dataset	American and British, Comfort Food, Halal, Indian, French, Italian, Japanese and Korean, Spanish, Chinese, Other
year	quantitative	year	2003 - 2016
rank	ordinal	Popularity rank of each food type in each year. Ranked according to yearly popularity.	1 - 11 (11 as non-exist)
month	Quantitative	month	1-12
Business popularity	Quantitative	Popularity of business in a month. Calculated as count of reviews on that business in a month.	Positive numbers

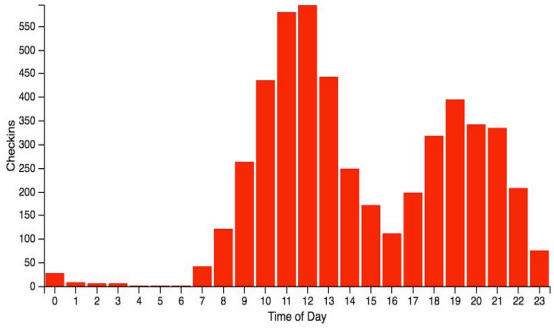
Food type popularity	Quantitative	Popularity of a food type in a month. Calculated as count of reviews on business serving that food type and divided by number of these businesses.	Positive numbers

What have others done to solve this or related problems?

- http://people.ischool.berkeley.edu/~sayantan.satpati/yelp/#portfolioModal1

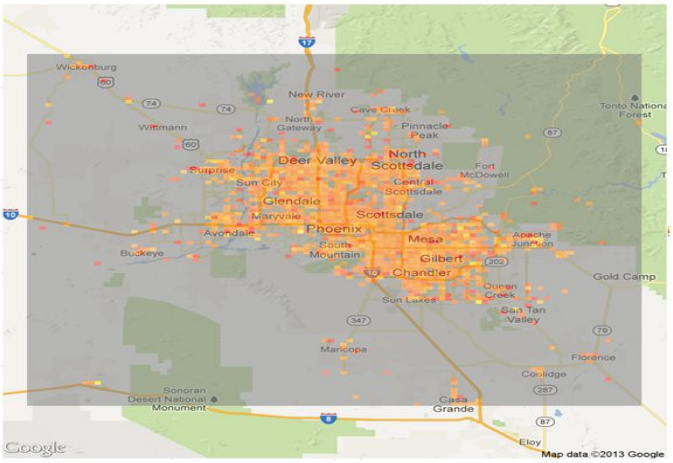
  How are my customer checkins looking like? What are popular days and hours? How can I plan my resources for effective customer service?





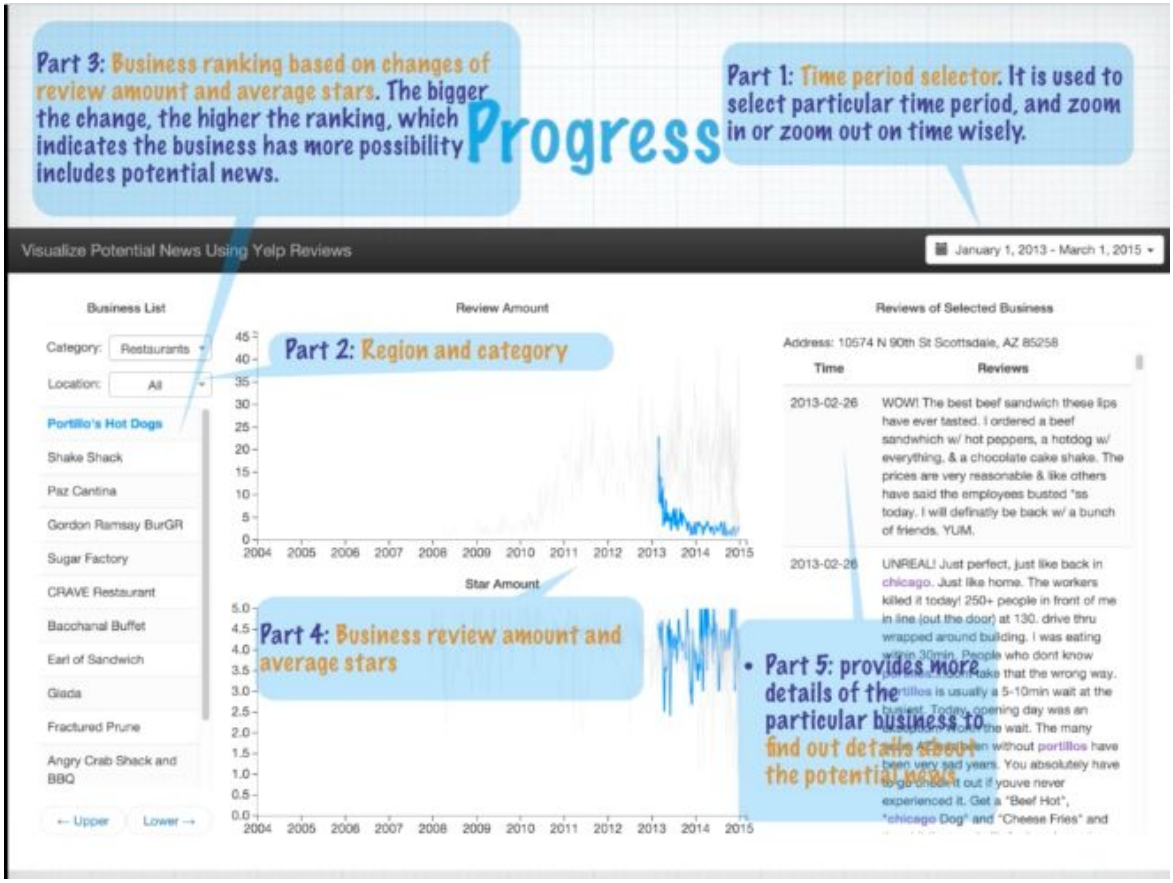
This project visualizes popular days and hours for restaurant to better manage their service, while we care about time trend of popularity of restaurants in a longer period.

- https://www.kaggle.com/c/yelp-recruiting/visualization/1154



A heatmap of average business ratings broken up by small areas, yellow squares are 1-star average areas, red squares are 5-star average areas.
Our project has the same module to present the rating distribution over New York. This projects can be a good example for the color using and map matching.

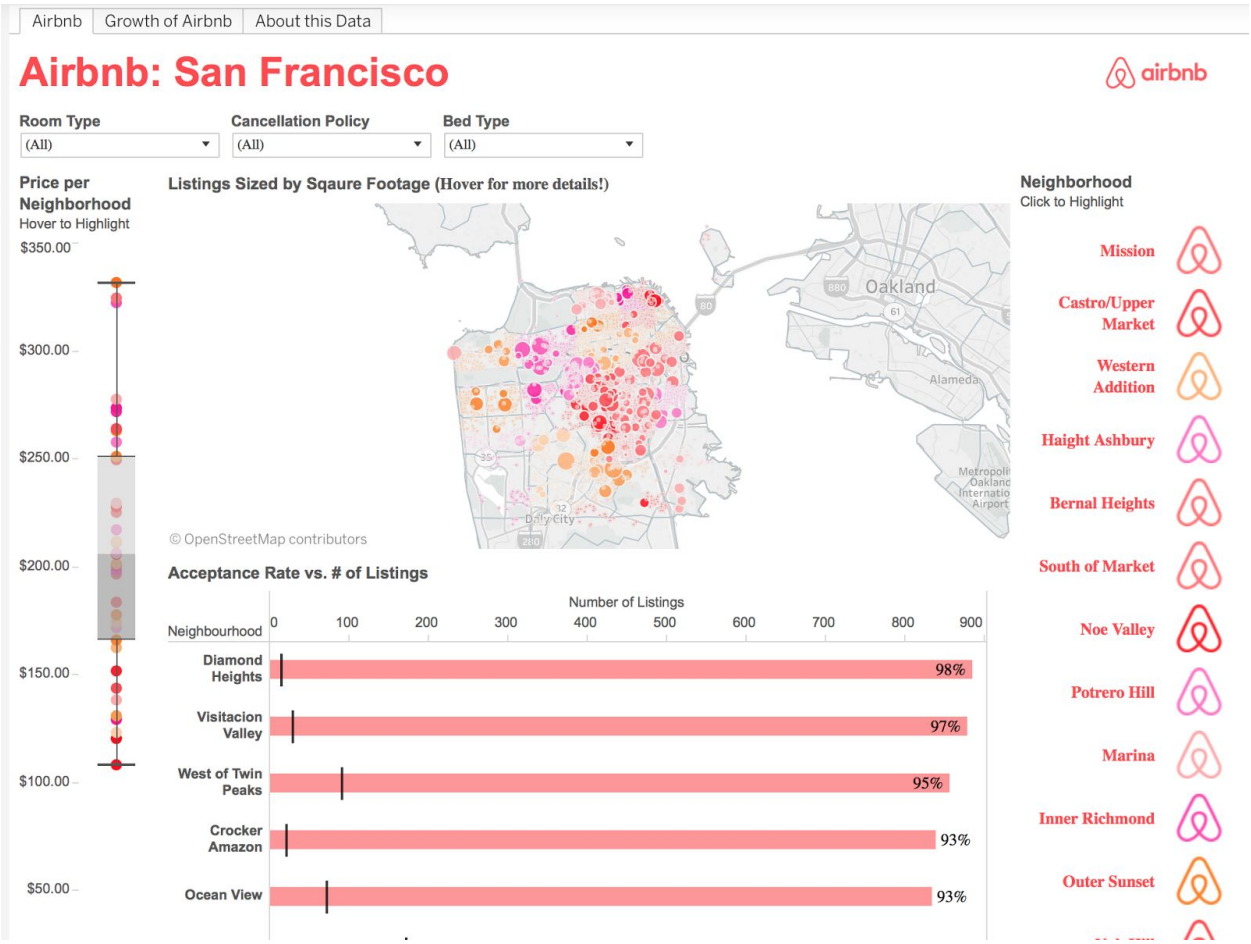
- https://docs.google.com/document/d/1dnX4LjAm40v2nLBm2bWV9WUH-JQNcY0ZRZHnIZ4CNic/edit



This project mainly use two line chart to visualize the yelp business information. One is how the review amount and the other is how average stars through different years. These two are both useful since they indicates the popularity of the yelp. However, it may aggregate too much, we can add create some graph that show more details.

Another useful information is that they can let the user see the review details if they want. This design reserved the information that dataset provided but won't let the overview seem too messy.

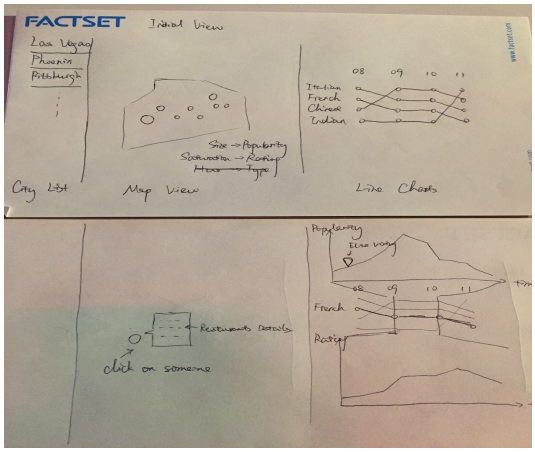
- https://public.tableau.com/en-us/s/gallery/airbnb-prices-san-francisco



This project is analyzing publicly available information about Airbnb's listings of San Francisco. Color is used to visualize the type and price of the residential house. It can be very helpful of our project, for instance, visualizing our food type and price.

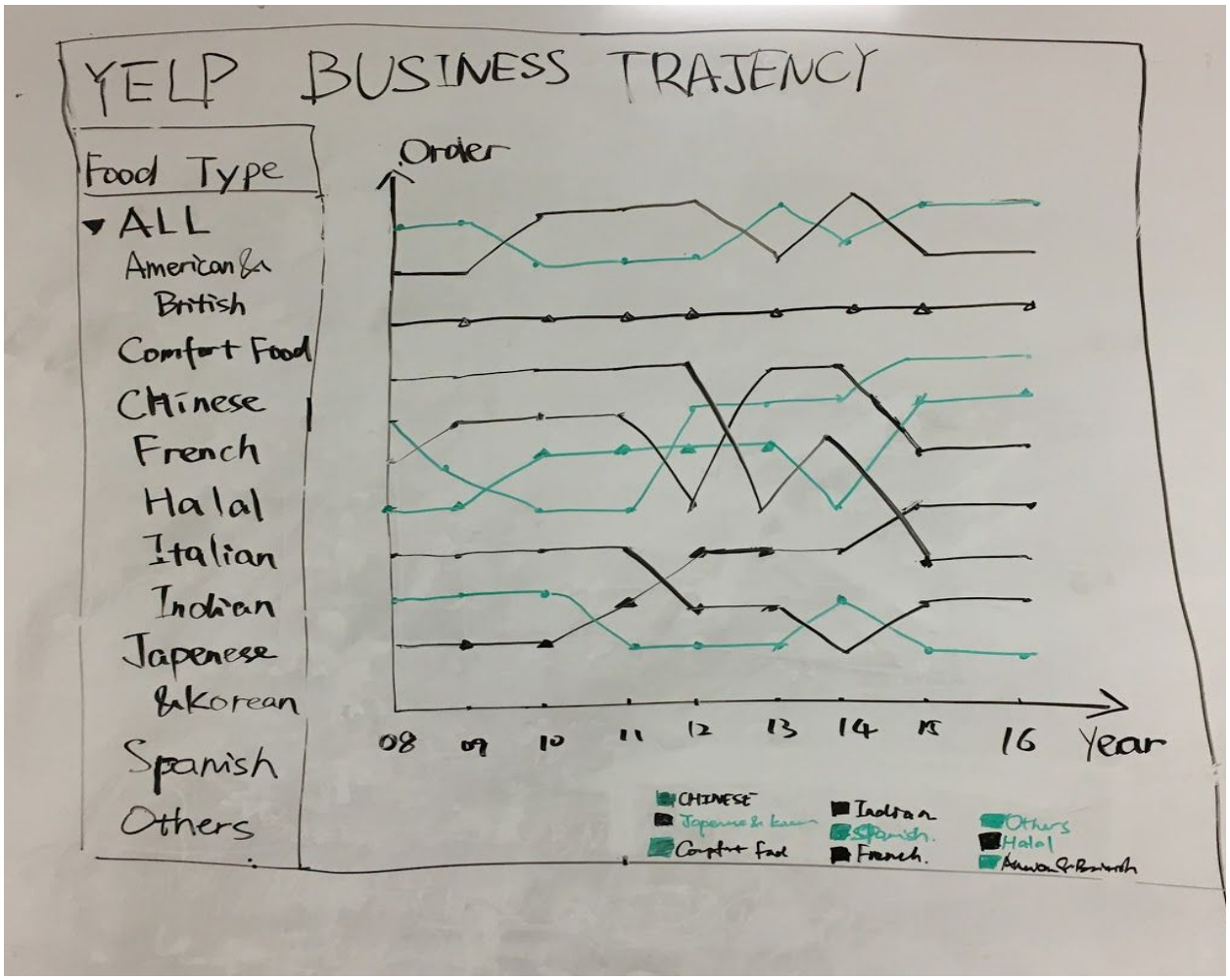
Design Iterations

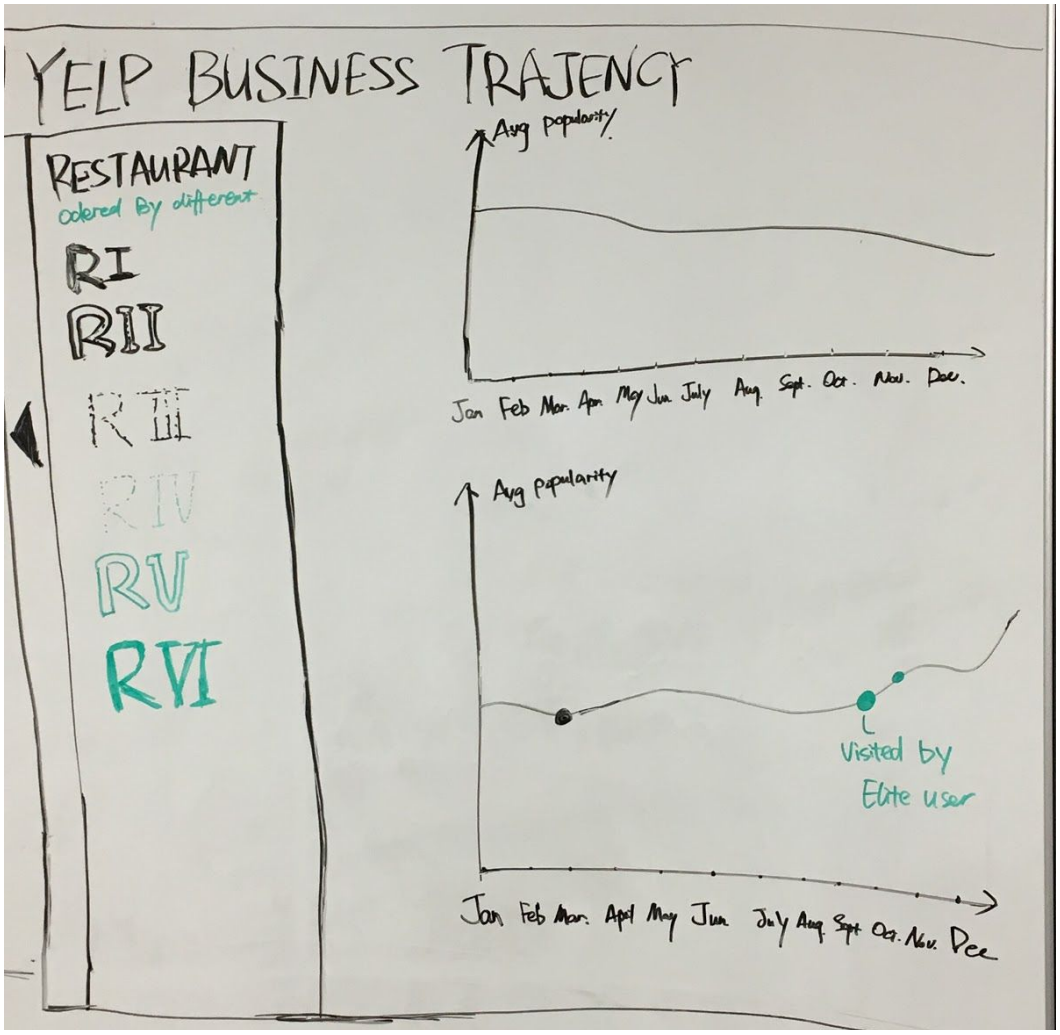
First mockup:



The page is two way directed. First is overview shows the bar of cities, map of the selected city with its restaurants represented by different color saturation (rating) and size (popularity). On the right is the rank of each food type through year.
Second is detail of each restaurant with how its popularity changes over selected year with elite visiting in that year represented by colored dot, and the changes of the food type it belongs to to compare.
The disadvantage of this mockup is that it introduces too much redundant information like location, price range and rating.

Second mockup:

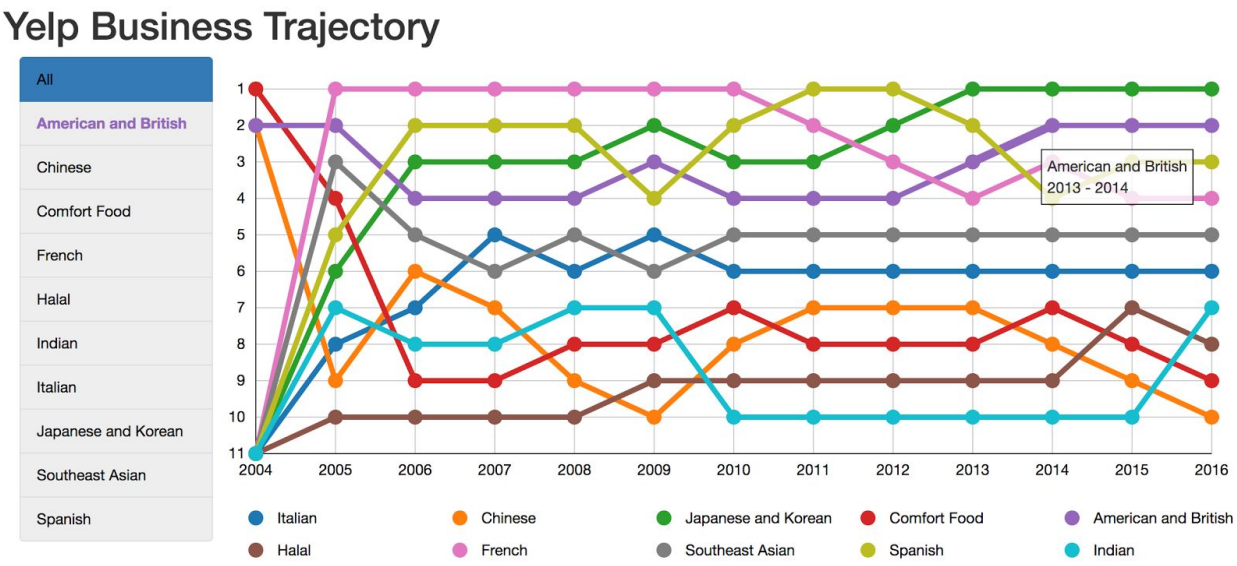




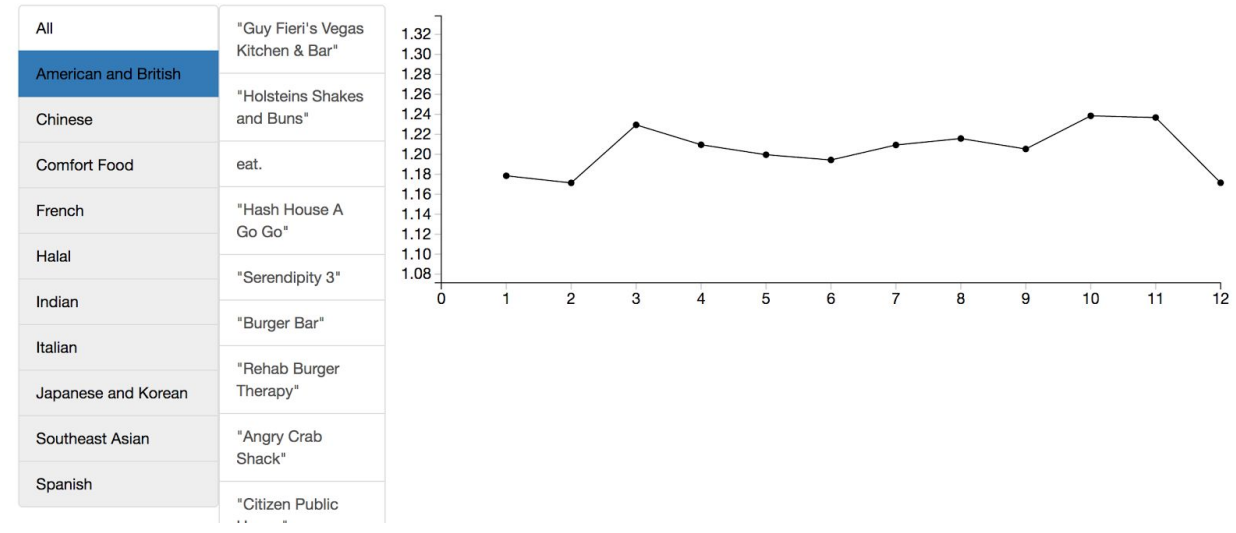
In second mockup we removed the map visualization since we thought that map has little relation to our questions . Instead we set the rank of the food types through years as the overview combined with the food type bars.
In detail page, we still show the single restaurant's popularity through selected year and the food type's changes it belongs to. In addition, we add a list of abnormal restaurants ordered by the degree of its abnormality to let people see why they are abnormal.

Final Visualization

- Overview
    - From overview we can see the rank and its change of each food type over the year. Each line with a separate color hue represents a food type. Each dot represent a rank in a year.

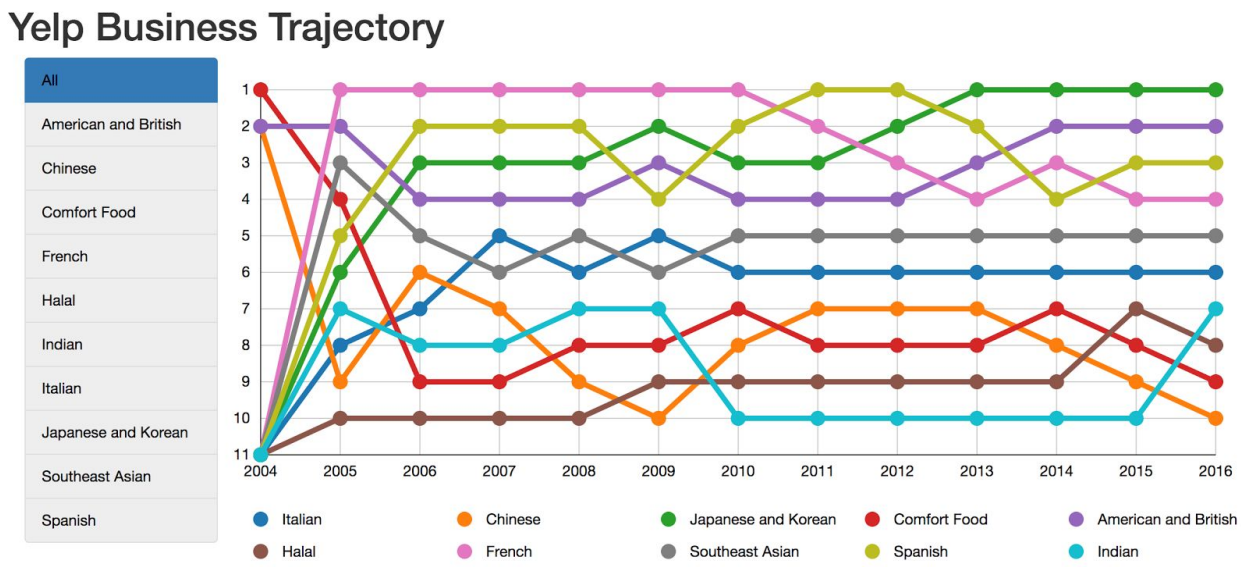


- Food type popularity trend
    - This view will be shown after user choose a time segment of a certain food type in the previous view. The trend of popularity of selected food type in that year is shown with a line chart. Also a list of restaurants who don't follow trend of this food type will be shown.



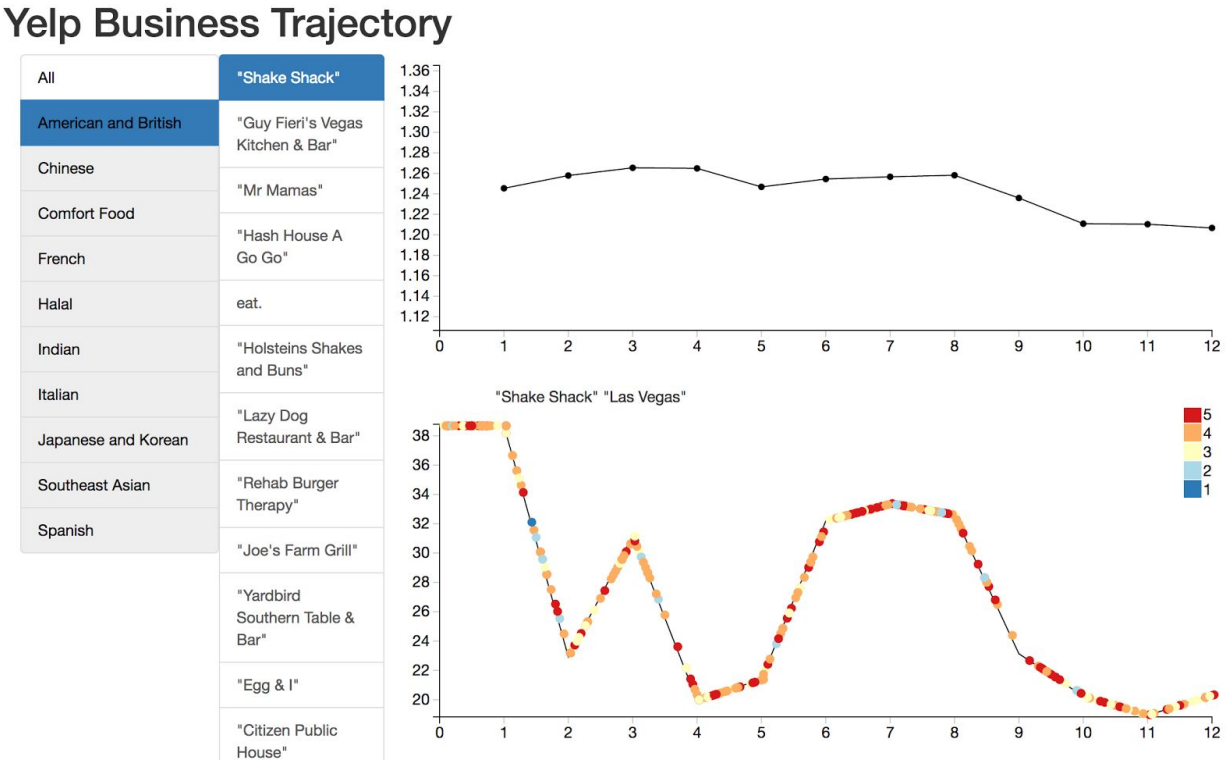
- Business popularity trend
    - This view will be shown after user choose a business in the previous view. The new line chart is showing the trend of popularity of that business. Each dot represents a visit of Yelp Elite User on that business. The horizontal location of the dot represents the time of that visit. The color of the dot represents the rating in that review.

Findings



From the overview, we can see that at the very first view years, since Yelp is still a startup, there's no clear pattern for ranks of food types due to incomplete information in the product. Afterwards, French food, Spanish food, Japanese and Korean food and American and British food forms first tier with high ranking. Among them, Spanish food grows significantly since 2009 and then drops back in 2014. Meanwhile, Japanese and Korean food is growing step by step

with American and British food while French food is dropping these years. For other food types, Italian food and Southeast Asian food are fighting in second tier.



As we can see from this line chart of popularity of Shake Shack in a certain year, the trend doesn't quite follows the trend of the food type (American and British). After several bad reviews (blue dots) from Elite users in February, the popularity of this business drops down. Then it uprise again avoiding bad reviews. However in April the popularity drops again due to several bad reviews. This kind of phenomenon also happens in August and September.

This kind of situation also happens on this Indian restaurant. As we can see, after March the popularity of Indian food is actually rising while the popularity of this restaurant is dropping due to a bad review from Elite users. Again after July the popularity of Indian food is dropping while the popularity of this business is uprising due to a medium rating from Elite user.

Limitations and Future Works

- In our visualization, user cannot see details of a review like the content of that review, which could be helpful if user want to observe why there's such a good or bad review influencing the popularity of a restaurant.
- In our visualization, whether a restaurant follows the trend of food type is not shown in a good form. Although we have a list of 20 abnormal business, it would be better if we show richer visualization on all business in that food type.