

Tasty Tastes

Different ideas for the topic:

Settling down the topic was quite a challenge for our group. At first we logged onto the Shanghai Open Data Platform (<http://www.datashanghai.gov.cn/>). There are a lot of accessible data, however, the five of our team members could not find a theme that we are all interested in.

Then we checked for different sources of data, and developed four different ideas: Shanghai night snacks; popular dishes of hot pots; most popular milk-tea beverages; sweetness of Shanghai dishes. All of them were related to food, since we thought that food was a topic related to everybody's daily life.

However, choosing from the four topics was a huge pain. We have to consider about the accessibility of different data, and it seemed that none of them was easy to get access to. So we decided to split up the tasks, which meant that each of us took one topic and tried to get access to the source and see if we can get some data. We sent e-mails to Haidilao, a famous hotpot chain store in China, to see if we can get the data of their sales records. We also sent emails to Eleme, Meituan, and other food delivery platforms to ask if we could get the data of their food orders and delivery routes at night, so that we could get some ideas about who in Shanghai usually order food at late night and what do they eat. Replies came very quick, yet all of them were disappointing. It seems that all the data we asked were their "commercial secrets", so they were not able to help us.

At this step, we encountered the first obstacle of our project: It seemed that the data for the topic that we chose could not be reached. That leaves only one option for us: to change our topic, and to find reachable data that we are interested in.

After some struggle, we decided to focus on the current situation of food delivery service. Instead of the food orders and delivery routes, we slightly altered our focus point to the quality and quantity of the whole food delivery service in Shanghai.

Specifying the topic:

As we've decided to focus on the food delivery service, we listed several

aspects that we consider to be relatively important for our topic:

- ✧ The quality of the shops' environment
- ✧ The addresses of the shops
- ✧ The types of food the shops sell
- ✧ The sales number of each shop
- ✧ The rate of the shops given by the users
- ✧ The time of open hour of the shops

Then our task changes into fetching data from websites. Our team member Shuhao used a crawler called Bazhuayu (a software robot downloaded from <http://www.bazhuayu.com/> to automatically fetch data from Eleme's websites. The input was the approximate location, and the output was the accessible information of each shop near this location.

At the very beginning, we chose Yangpu District and other two districts near Yangpu, since this area is most familiar to us. Then Francesca suggested that we could also put Lujiazui into our dataset, which is the financial center of Shanghai. That area is very near Yangpu, just across the Huangpu River, adding it into our dataset might bring us more interesting patterns. So the four districts became our input.

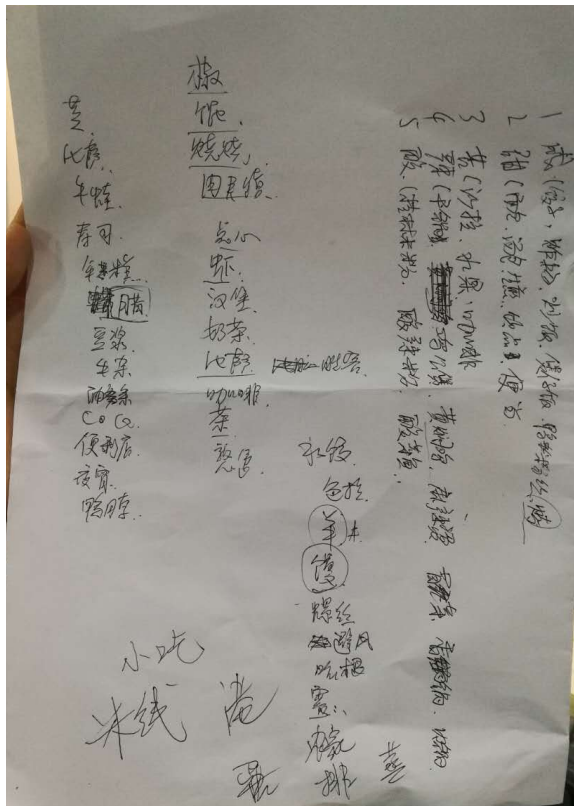
The original raw data was rather dirty. There were a few shops in Suzhou and other provinces included in the dataset, also, the addresses of the shops were written in different formats, which demanded challenge for us to organize and put them into order.

Processing and visualizing the data:

Now we have such a large dataset. How to process the data and highlight the significant entries? How to categorize different data? This is another troublesome task.

We discussed several ways to categorize the food. One is that we categorize the food by their appearing formats, like rice, dumplings, hotpots, noodles, fried things... But according to this standard, we have tons of categories which was difficult to visualize, not to mention discovering any trend. So we changed a way of thinking, and found that according to traditional Chinese pharmacology, there are five flavors: sweetness, sourness, bitterness, peppery hotness, and saltiness

(<http://www.zysj.com.cn/lilunshuji/zhongyaojibenlilunzhishi140/263-7-1.html>). This could be a possible way of categorizing our data.

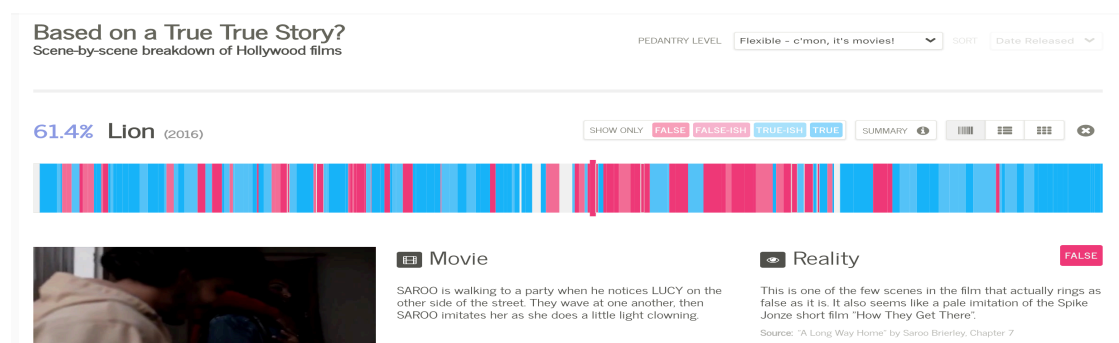


We made a list about which dish belongs to which flavor. Normal dishes like steamed rice, dumplings, fried things and many other are considered to be “salty”, since salt is the most common Chinese seasoning. For sweetness, bakeries, soup dumplings, beverages and some other are included. Peppery hotness contains all kinds of food that is famous for its spiciness, such as hot-spicy pot, spicy soup and so on. Sour food includes Guilin rice noodles, hot and sour rice noodles, Fish with Pickled

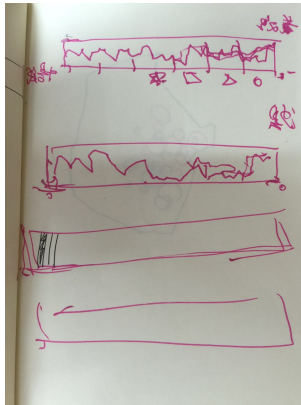
Cabbage and some other food like that. As for bitterness, we considered bitterness related to freshness, so we included salad and coffee in this category.

With our five new categorizations, we can start to play with different visualization tools.

At first we referred to <http://www.informationisbeautiful.net/visualizations/based-on-a-true-true-story/>, and thought that we could do a similar kind of visualization as the following: What we referred to:



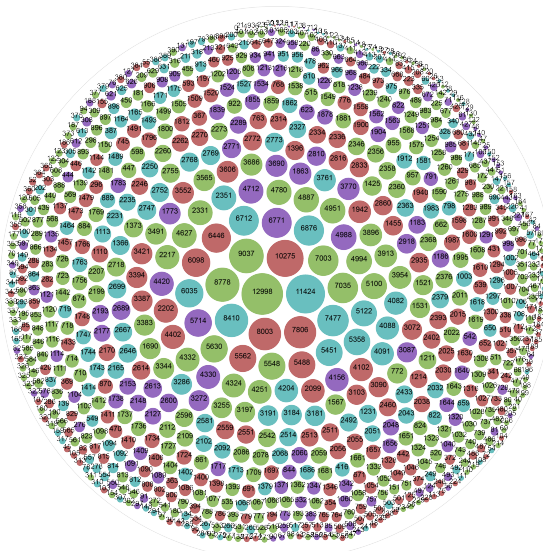
What we would like to do:



We intended to show flavor by color and sales number by the length of each bar. However, Francesca reminded us that with the data that we have, we might not be able to get a bar as tidy and beautiful as the example we looked at. Also, the dataset that we have is so big that if we do a similar kind of visualization, we might get a super long bar that is impossible to see details unless with the support of vector.

She suggested that we should try different visualization forms first, and after we played with several visualizations, we can finalize our ideal version.

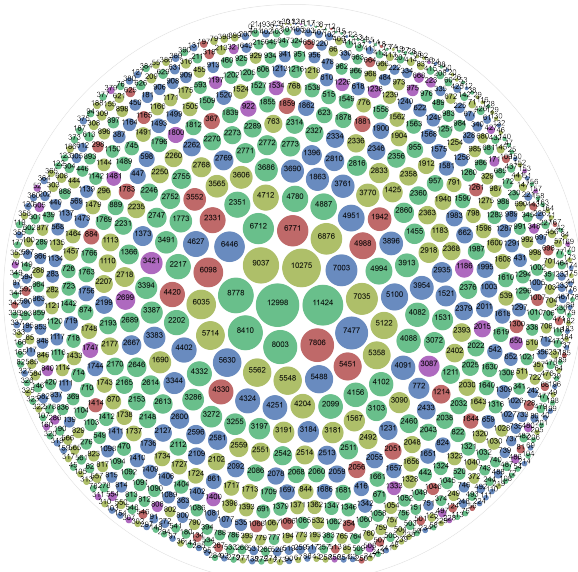
1. Circle Packing



Graph 1.1

In this visualization (Graph 1.2), the colors represent the type (flavor: salty, sweet, spicy, bitter and sour) of the food that each shop mainly sells.

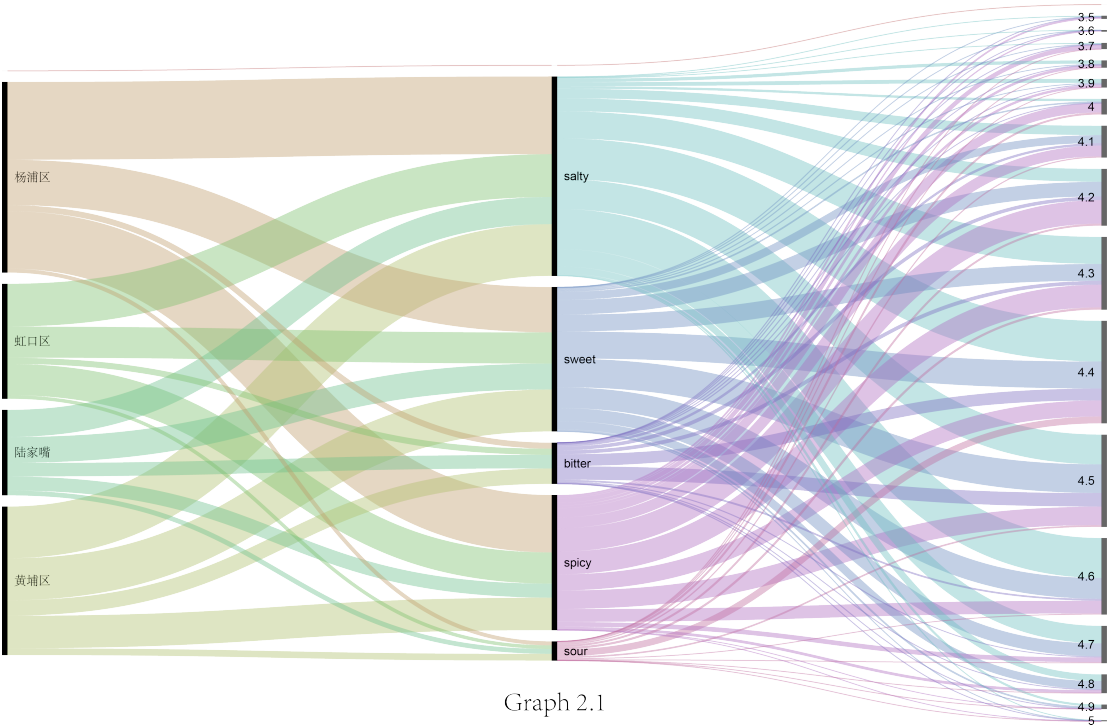
In this visualization (Graph 1.1), each circle represents a shop of food delivery. Different colors represent different districts. The size of the circles represents the sales number.



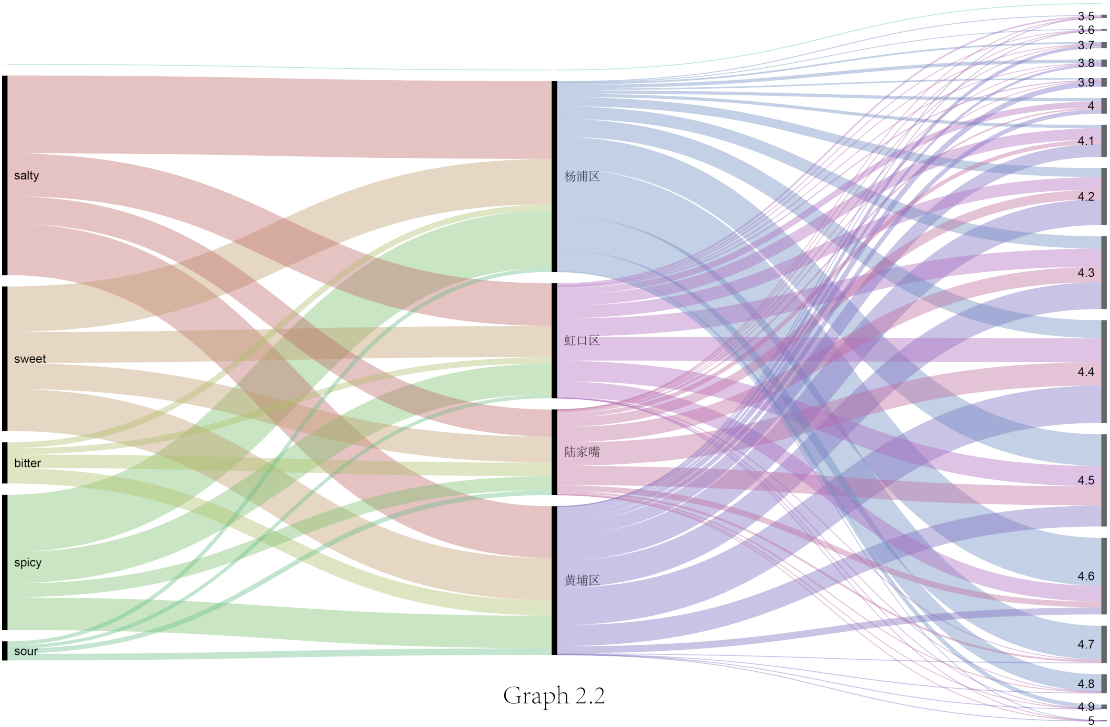
Graph 1.2

It is not clear for us to see any kinds of pattern from circle packing graphs, but we can spot the best-selling shops in no time.

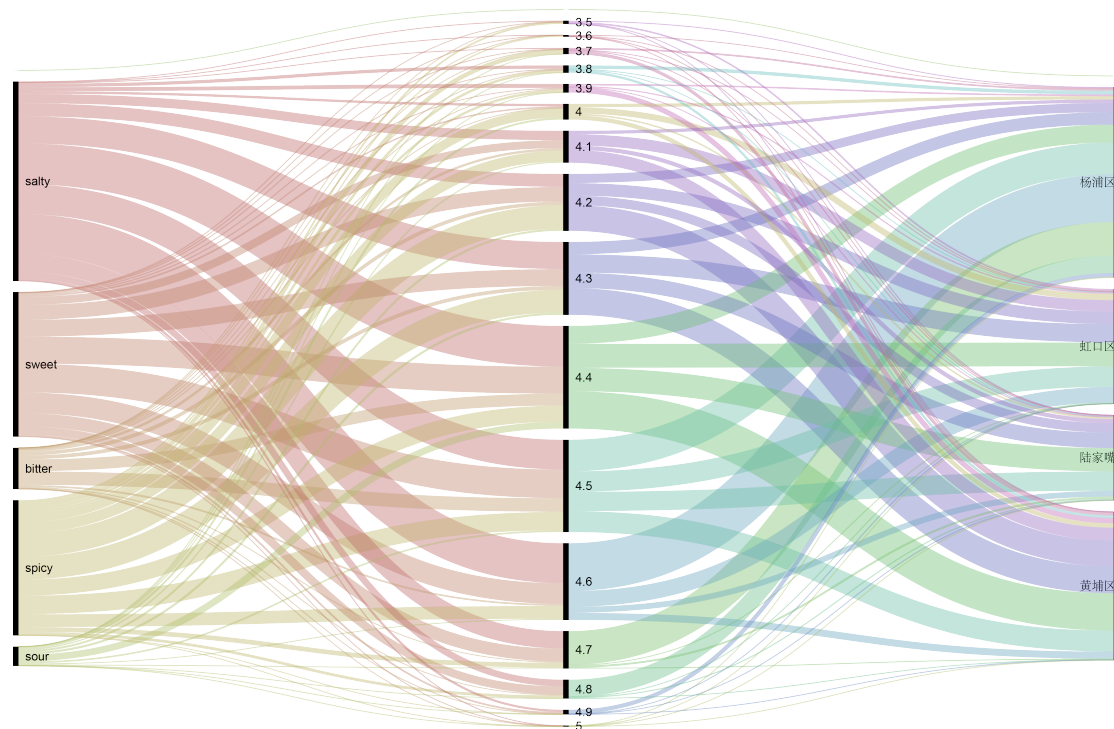
2. Alluvial Diagram



Graph 2.1



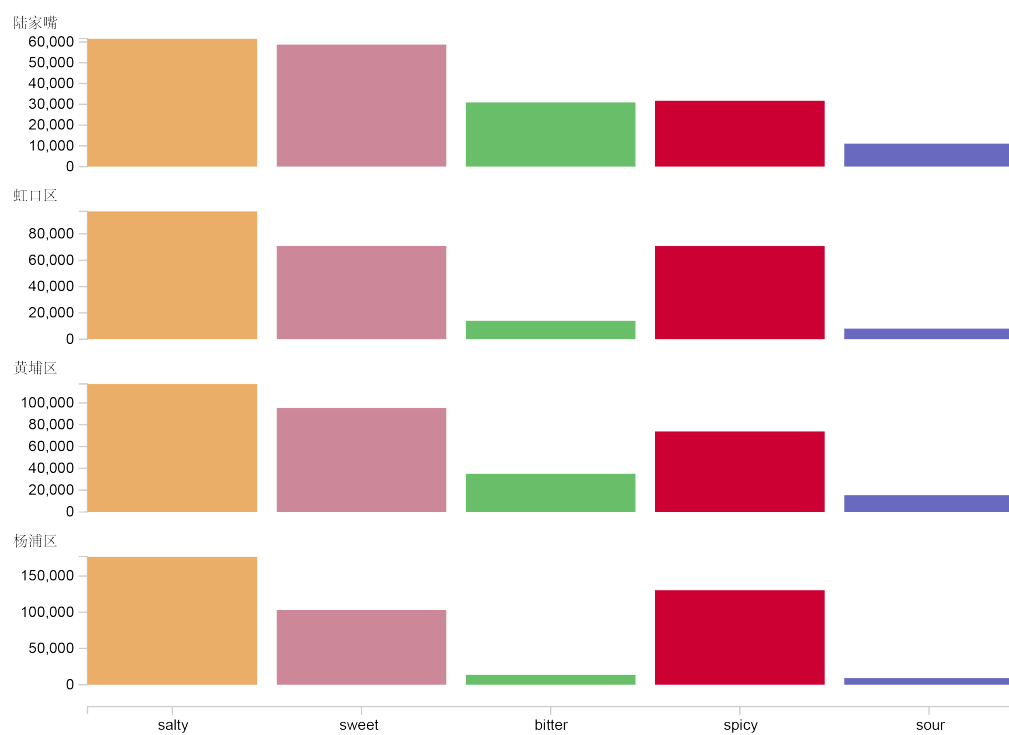
Graph 2.2



Graph 2.3

From Graph 2.1, 2.2 and 2.3, we can see some trend: like which district is typically fond of a certain kind of flavor, and which district has higher rated shops, also, we can see the whole amount of shops by the width of the stripes.

3. Bar Chart

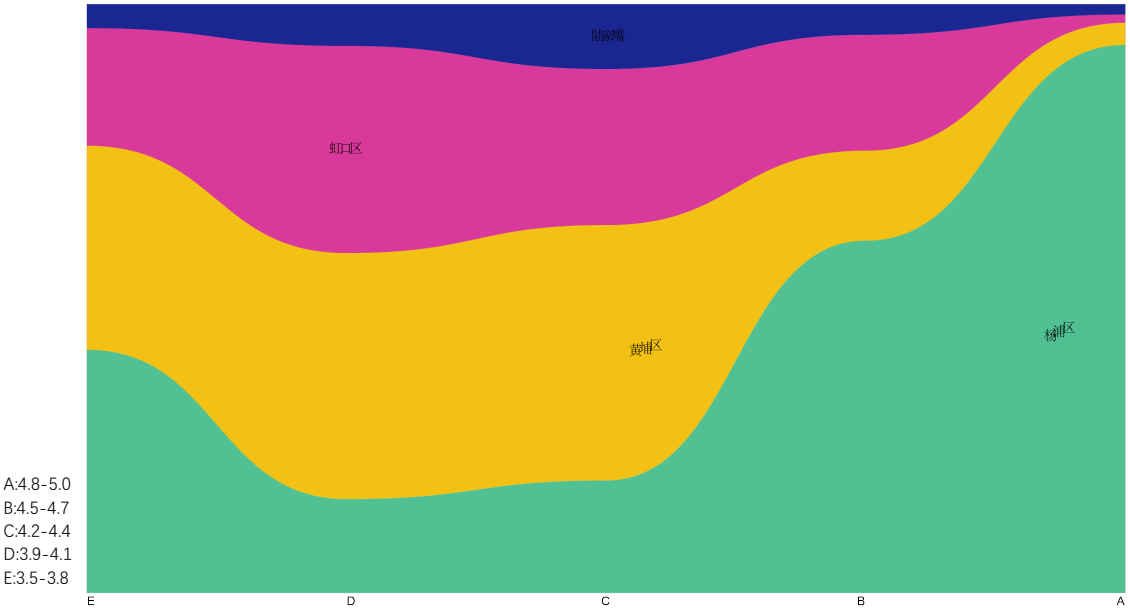


Graph 3

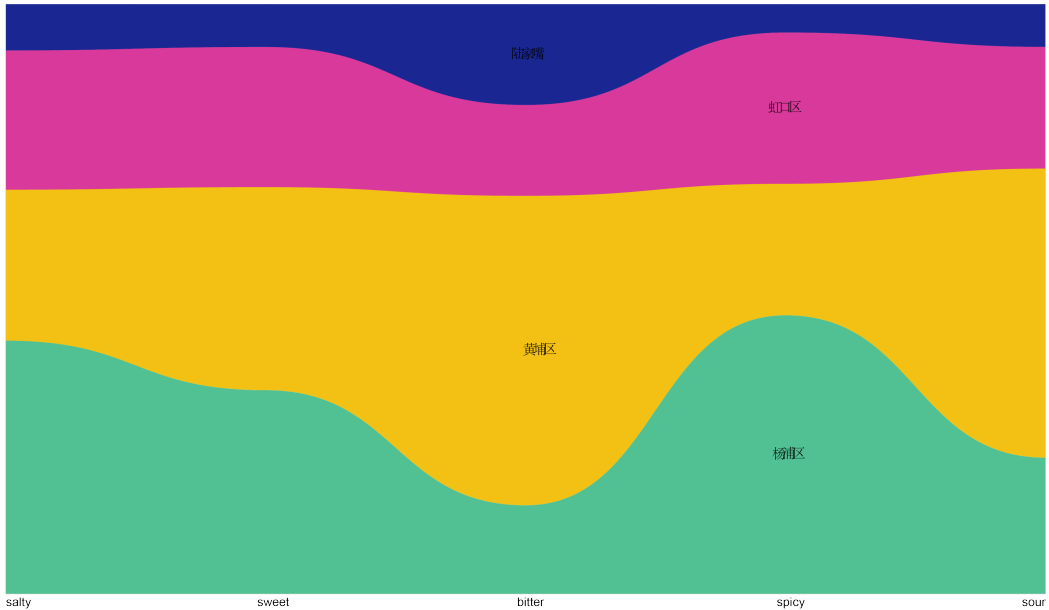
The preferences of flavors of each district are clearly showed in Graph 3, each district uses different scale and uses salty flavor as comparison , so that the portion of each flavor could be told at a glance.

One interesting discovery is that people in Lujiazui orders bitter food more than any other district. We assume that since Lujiazui is the financial center of Shanghai, those white collars need more coffee and other fresh food to keep themselves in the best condition.

4. Stream Graph



Graph 4.1

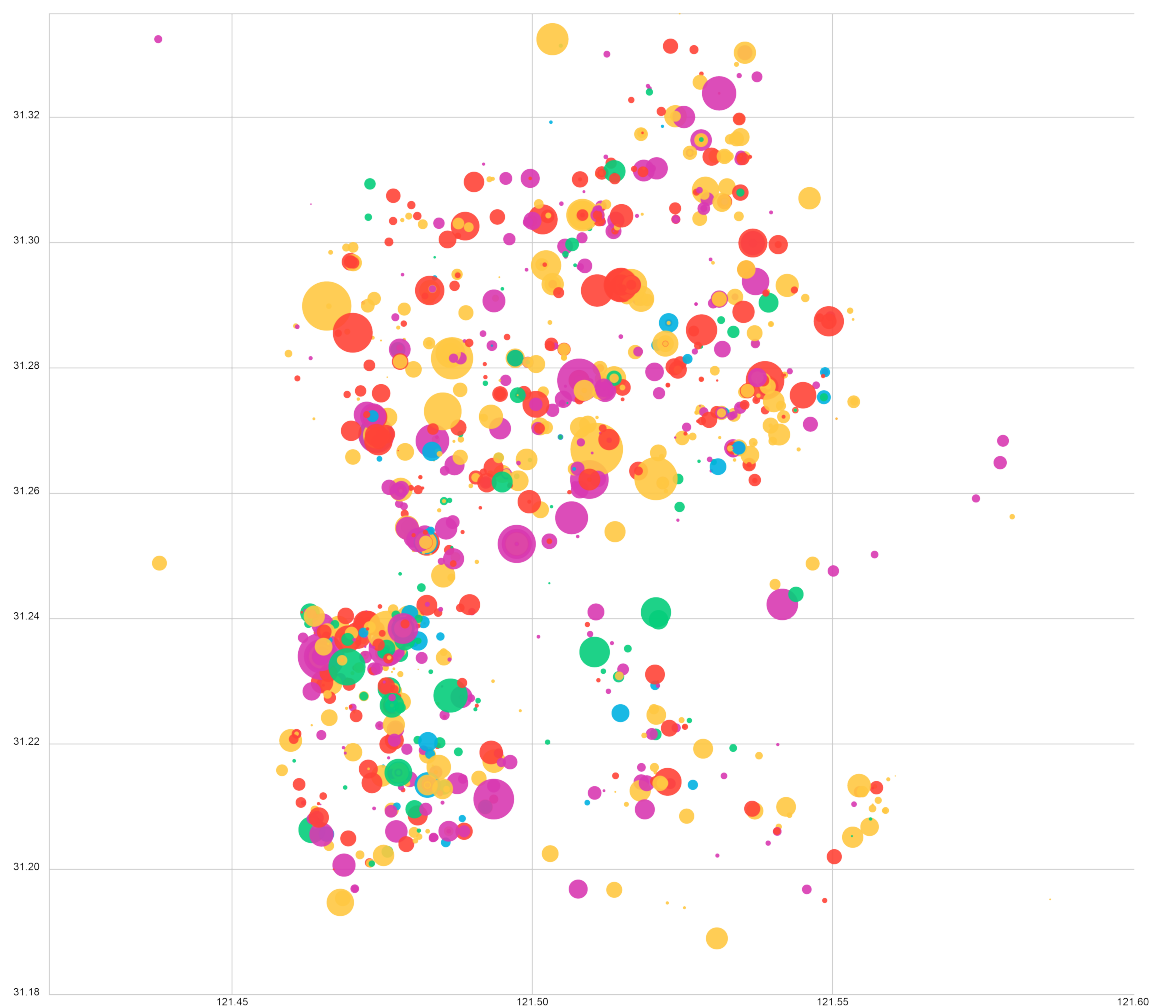


Graph 4.2

Graph 4.1 and 4.2 show very obviously the patterns. According to graph 1, Yangpu District has the most best-rated shops. Also, most of the shops in our dataset is located in Yangpu. As for graph 2, clearly that Lujiazui District has a special love for bitterness, and Yangpu District has a preference for spiciness.

5. Scatter Plot (Final viz)

As we knew about the quality and quantity of our dataset, we thought that we could do a map of the shops and see the scattering of our dataset. So we reprocessed our dataset, we put the addresses of the shops into an online tool to transform into latitude and longitude (<http://www.gpsppg.com/latitude-and-longitude.htm>), then we have a new series of “code number” to visualize the exact locations of each shop.

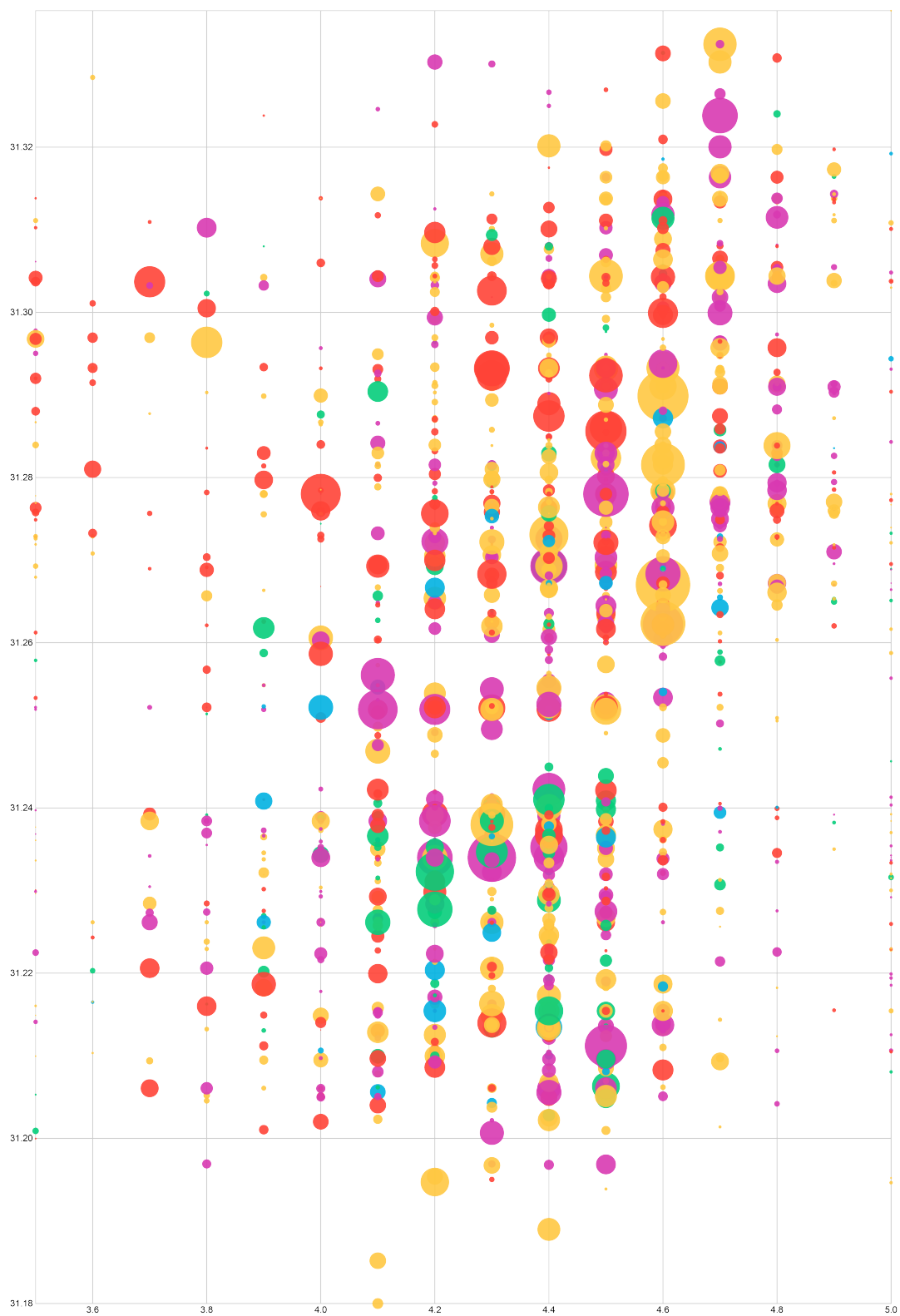


Graph 5.1

Graph 5.1 used latitude and longitude as horizontal and vertical axis to make

a map of the shops (You can even see the outline of the Huangpu River). The size of the circles represents their sales number, the bigger the circle is, the more orders that shop has sold in the last month. The colors represent the flavors (yellow: salty, purple: sweet, red: spicy, green: bitter, blue: sour). From the map, we can also tell that most food are scattered all over the places that we chose, but big-selling bitter food (big green dots) only shows up in places near Lujiazui area (including Huangpu District). With the knowledge from previous graphs, we know that most bitter orders are from Lujiazui, so we could possibly assume that most orders that the big-selling bitter food shops in Huangpu District received are from Lujiazui. They have to deliver food orders across the Huangpu River. Maybe this phenomenon was because the rent was too expensive for the shops to open right at Lujiazui area.

We also want to see the relationship between the flavors of the shops and the rates they received, as well as the relationship between the location and the rates. For the former purpose, we used longitude as the vertical axis and scores and the horizontal axis, and then we have a rather scattered graph of shops sorted by their rates. The colors of the circles show their flavors. On the other hand, for the latter purpose, we changed the colors of the circles so that each color represents a district.



Graph 5.2



Graph 5.3

Graph 5.2 and 5.3 could be looked at together.

From the graphs, we can see that the best-selling shops are mostly rated at 4.3-4.7, and shops in Yangpu District are higher rated. Also, two shops in Yangpu

are rated low (at 3.7 and 3.8), but they still have rather big sales numbers. One sells spicy food and the other sells salty food.

The full-score shops all have little sales number, this made us rather curious: Are they newly-opened shops or do they just cope with a few orders a day to ensure the quality of their food?

Entity shops investigation:

1. A 梦 Eamonn's



This store is the best-selling shop of food delivery.

According to online information, they have their own delivery team and fine packaging. Our team ordered the second best-selling item of that shop (showed on the left).

The man of delivery made three deep bows and wished us a good appetite.

The packaging was really careful, our team member noticed that their chopsticks were of good quality and even had peppermints inside.

Even though the taste was just so-so, we still gave a 5-star rate, for its service, and also for its attitude.



2. G-name Korean food

This store rates rather low in our dataset, but still have a considerable sales number. Therefore, we decided to take a close look at the dishes and comments of this shop.



韩式年糕拉面



泡菜五花肉石锅拌饭



金枪鱼蔬菜沙拉

Some of the comments said that the pictures online were not the same as the real food. He called the shop, and the seller said that the real food should be considered as final.

As we can see from the pictures, their packaging was not so satisfying and their food looked unclean.

3. Y-name salad

This store belongs to the bitterness flavor. Its rate is rather low, as well as its sales number. Is it really that bad? We searched this store online.



澳洲西冷牛肉色拉

推荐酱料：蜂蜜芥末酱 主料：牛肉，西兰花，羽衣甘蓝，藜麦，紫甘蓝，芝麻菜，绿色生菜，黄瓜，胡萝卜，金针菇，南瓜子，混合坚果 卡路里：376kcal

¥ 52起

选规格

商品评价 (24)

★★★★★ 2017-05-09 12:05



★★★★★ 2017-05-07 18:19



★★★★★ 2017-05-04 15:25



不好吃，底下还有咸菜

◀ 1 2 3 4 ... 8 ▶

The online pictures looked nice, however, the pictures of the real food looked like this:



Also, the comments online showed that the taste was rather disappointing. No wonder it doesn't get good sales performances.

4. S-name rice

This store is one of the two "strange" shops in Yangpu District. Its rate was rather low, but still had a big sales number.


★★★★★ (1390) 月售4189单

所有商品 | 评价 | 商家资质

全部(3549)	满意(3123)	不满意(426)	有图(36)
----------	----------	----------	--------





The real pictures made people lose their appetite. Maybe the big sales number is because of its price.

Reflections:

The whole process from collecting to visualizing data and then the investigation was interesting as well as challenging to every one of our group members. We don't have any technical background, before this class, we know little about this area. Following Francesca's classes, we began to get some ideas about how to fetch data and how to use tools to visualize our data.

The most challenging part for our group is to decide which part of our dataset we really want to visualize. As we have fetched a lot of information about different food shops, it became hard for us to abandon some of the information. This process is called “数据降维” in Chinese. The categorizing was challenging, too. We had to decide which method was the most scientific one. And we went through some articles (<http://baike.baidu.com/item/%E5%8F%A3%E5%91%B3>, <http://baike.baidu.com/item/%E4%BA%94%E5%91%B3/1269566>, <http://www.zysj.com.cn/lilunshuji/zhongyaojibenlilunzhishi140/263-7-1.html>). According to these categorizations and our own knowledge, we went through our dataset and categorized our data manually. This was rather challenging for us, and we wonder if there could be a better way to categorize the flavors.

The final outcome of our project was higher than our expectation. We were amazed that data could be this beautiful. Some of the visualizations really gave us a surprise. At first we were only interested in the type of food that people in different district preferred, but the results showed a lot more. With our visualization and dataset, it became much clearer that which food shop offered good and popular food. When people want to order food from Eleme, they can refer to our data visualization and make the right choices. Thanks to Francesca, we also met some people who are really passionate in this field, which encouraged all of us a lot.

Team members:

Ma Shuhao

Yang Lixing

Wu Tong

Cai Zhelin

Tai Lunyue