The Story of the Best Parks in Shanghai

Group C: She Jing(1633240), Wang Shiwei(1633236), Zheng Xiaoxi(1633261), Chen Guodong(1633232)

Choose the topic

At the very beginning, we did a lot of research on the website of Shanghai data to collect data. Then the dataset of "Park in Shanghai" captured our attention. In this data set we can find the total area, the location, the green space and the water area of those parks in Shanghai. We can use these data to give some suggestion to people who want to visit parks in Shanghai.

Data collection

We thought that these data were not enough. Chen Guodong recommended a WeChat Subscription named Shanghai Fabu which can provide the data of people flow in many tourist attractions in Shanghai. And most of the parks we found are in the subscription. So we intended to find if there are some relationships between the people flow and the features of each parks. The subscription only offers the current people flow without any historical records, so in order to get more data we decided to record the data of people flow every hour. We showed the data in subscription to Francesca, she suggested us to figure out if it is the current number of people in the park or the current number of people enter the park.

During our collecting proceed, we found that there still many people in parks at closing time. So we concluded that those data of people flow are the current number of people in parks. The process of collecting data seems easy but requires us to check the WeChat subscription per hour. It needs both patience and passion.

We choose 12 parks' data of people flow to build our dataset. And each teammate recorded 4 parks' data every day. The problem was that we found 2 parks' data were not updated every day. In order to solve this problem We tried to call the manager of Shanghai fabu platform. But the manager told us to dial 5257788 which is the number of Shanghai Greening Bureau to find the answer. The Greening Bureau asked us to call 5158693 which is the number of Park Greening Bureau, yet we didn't get any solid answer from them. So we delete 2 parks without full data and choose 10 parks' data to develop our visualization. We eventually collect 13 days of data including 8 workdays, 2 weekends and 3 days in May Day Holiday.

●●●○○ 中国移动 夺 12:05	●●●○○中国移动 夺 12:06 ⑧ ※ ■●	●●●●○○ 中国移动 夺 11:04 ⑧ 券 💷)
く返回 景区实时客流 •••	く返回 景区实时客流 •••	〈 返回 关闭 景区实时客流 •••
上海动物园	上海植物园	上海鲁迅公园
4A级 长宁区 07:30-16:30	4A级 徐汇区 07:00-17:00	3A级 虹口区 06:00-18:00
15926人/22000人	12323 人/34970人	123 人/32000人
数据更新时间 2017-04-29 12:00 当前客流/瞬时最大承载量	数据更新时间 2017-04-29 12:00 当前客流/ 類时最大承载量	数据更新时间 2017-04-29 11:00 当前客流 / 瞬时最大承载量
上海世纪公园 28°C - 15°C 西南风3-4级	上海南翔景区	上海和平公园
4A级 浦东新区 07:00-18:00 晴天	4A级 嘉定区 00:00-23:59	3A级 虹口区 05:00-18:00
5000人/53994人	3122人/8700人	4256 人/18694人
致 描 史 新 时 间 2017-04-29 12:00 当前客意 / 錄时最大承載量		数据更新时间 2017-04-29 11:00 当前客流/
上海大观园	上海明珠湖·西沙湿地景区 27°C - 14°C	上海花米庄行乡村旅游景区
4A级 青浦区 08:00-17:00	四南风3-4级 4A级 崇明区 08:30-16:00 晴天	3A级 奉贤区 07:00-16:30
485人/5271人	1117 人/46185人	1214 人/34050人)
政范史前町1回 2017-04-29 12.00 当前答点/ 與时最大車載量		数据更新时间 2017-04-29 11:00 当前客流 / 順时最大承载量
上海共青森林公园	上海宋庆龄故居纪念馆	上海南社纪念馆
4A级 杨浦区 06:00-17:00	4A级 徐汇区 09:00-16:30	<mark>3A级 金山区</mark> 09:00-16:00
6000入/21000人	101 人/450人	
上海市旅游局 上海市旅游局 合作提供查询平台	上海市旅游局 上海市旅游局 合作提供直询平台	上海市旅游局 上海市旅游局 合作提供查询平台

Figure 1. Subscription of Shanghai Fabu

Analyze data by using different tools

After getting our dataset, we used different tools like Echarts, Tableau, and RAW to do some visualization. We found some interesting types of charts such as bar chart, line chart, radar chart, alluvial diagram, circle packing and so on. While we found that the alluvial diagram couldn't show the trend of people flow. And it is difficult to compare the area of each park in the radar chart. Then we decided to use scatter charts, bar chart, line chart and map to show the features and people flow of different parks.

Following are 2 line charts we used to do some hypothesis. We assumed that there might be some different trends in people flow according to different days like workdays, weekends and holidays.





Figure 3. Line chart in April 23



Figure 4. Group Work

Develop data visualization



Figure 5. Sketches

We had a clear division of work. Wang Shiwei organized all the dataset and find some related examples of charts. Zheng Xiaoxi worked on Al and RAW to produce some basic charts. Chen Guidong tried to locate every park in the map of Shanghai. And She Jing used Processing to generate a scatter diagram to show the people flow of each park in different time.

PART 1 Scatter Chart

In this part we want to show the people flow of each park in different times. X axis represented parks, and Y axis represented time, the area of each circle represented the people flow per hour in each parks.

We can also compare the data in weekend, workdays and holidays in this diagram through 3 concentric circle with different color. Francesca suggested us to use different color in these 3 concentric circle in order to show the difference between weekend, workdays and holidays.

After finishing our basic scatter charts, we found that the gap between 3 concentric circles is huge. We soon realized that we made a mistake about the radius, we defined the data of people flow as the radius itself, but She Jing and Wang Shiwei said that we should define the people flow as the area of circle, and use the formulas "S= π ×r2" to calculate the radius and generate a new scatter chart. Now the chart is more clear and accurate.







Following are the major formulation and their corresponding code we develop during the coding process.

import processing.pdf.*;

```
size(2360, 2360, PDF, "test.pdf");
background(255);
strokeWeight(1);
stroke(#cbbca6);
noFill();
for (int a = 200; a < 2360; a += 200){
    line(a+200, 0, a+200, 2360);
}
//1. The number of Chenshan</pre>
```

```
fill(29, 32, 136, 99);
```

noStroke();

```
ellipse(400, 2180, r, r); //draw a circul in the first line, represent the people in the park currently at Working day 8:00.
```

PART 2 Bar Chart

The purpose of this part is to help people choose the suitable park, the colorful bar charts show the composition of park and beneath the colorful bar is the grey bar chart which shows the distance from People's Square to each parks, because People's Square can be considered as the center of Shanghai. And tourists and residents can choose to go to the park they like.



Figure 8. Bar chart

We used this colorful stacked bar charts to illustrate the composition of park in figure 5, each bar contains 3 parts. The green part represents the green space in the park, the blue part represents the area of rivers and lakes in the park, and the grey part represents other functions of the park, such as children facilities, BBQ sites and so on.

We sorted every park by size, the first bar on the left is the largest park while the last one is the smallest. Then combined with scatter chart in part 1 we can conclude that the biggest park is not the most popular park. And maybe there are some relations between the distance and the people flow.

Part3 Line Chart

For the purpose of showing the people flow trend of one park in different days. We selected 3 typical parks including the most crowded park(Shanghai Wild Animal Park), the smallest park(Shanghai Heping Park) and the farthest park(the Grand View Garden Area of Shanghai) to

generate 3 line charts. The chart of SH Wild Animal Park shows that the number of people in holidays is larger than weekends and workdays. And there is a peak of people flow between 11pm and 13 pm. While in the chart of the Grand View Garden Area of Shanghai, there are 2 peaks in one day, the first peak is at 10am, and the second peak is at 2pm. The SH Heping Park which can be considered as a weekend park because the largest number of tourists appeared in weekends instead of holidays like the other parks.



Figure 9. The people flow of The Grand View Garden Area of Shanghai



Figure 10. The people flow of SH Peace Park



Figure 11. The people flow of SH Wild Animal Park

Part 4 Map

In order to show the outline of every park and have a better understanding of the surroundings of each park, we located all the parks we chose on the map of Shanghai, and gave an image of details to show more information like how many metro stations are around each park. Through this we can find if there are some relations between the traffic and the people flow. And we also located some infrastructures like residential areas, schools, hospitals around those parks.



Figure 12. Examples of the map

Conclusion

In this study, we find that the data quality provided by the government website is not so good. For example, one of our teammates went to the Shiji Park and find that the park is very crowded without any parking space, while the number of person on the website is small. So some parts of the data may not be accurate. And most of the parks' managers are unwilling to share the data with us. However, we finally got most of the data we need through different ways, and this is an exciting process. If we can get more data, we can predict the people flow in different time and give suggestions to tourists and residents who want to choose and visit some special parks. In that way we can divide people into different places so that they can have a better experience visiting the parks.

Reference

- 1. www.datashanghai.gov.cn/
- 2. <u>baike.baidu.com/</u>
- 3. <u>http://map.baidu.com/</u>
- 4. http://lysh.eastday.com
- 5. http://www.shbinbaipark.com
- 6. <u>www.csnbgsh.cn/</u>
- 7. www.gucunpark.net/
- 8. http://www.sh-daguanyuan.com/
- 9. <u>www.shwzoo.com/</u>
- 10. www.shanghaizoo.cn/