

COMP7507
Visualization and Visual Analytics
Course Project (2019-20)



Group Number: No.11

Instructor: Dr. Loretta YK Choi

2019.11.30

catalog

1 Visualizations	- 3 -
2 Explanation	- 3 -
2.1 Region	- 3 -
2.2 Political Camp	- 4 -
2.3 Age Distribution	- 4 -
2.4 Gender and Education Background	- 6 -
2.5 Investigation from HKU	- 7 -
2.6 Career Experience.....	- 8 -
2.7 Google Trends.....	- 10 -
2.8 Investigation	- 11 -
3 Difficulties.....	- 12 -
3.1 Some data are hard to collect.....	- 12 -
3.2 Some features are hard to quantify	- 12 -
3.3 Some data is difficult to find a suitable display chart.....	- 13 -
4 Methods & Justification of the choices	- 13 -
4.1 Choice for Website	- 13 -
4.2 Choice for Google Charts	- 13 -
5 Improvements	- 13 -
5.1 Display of Party Grouping	- 14 -
5.2 Social Network Mapping.....	- 14 -
5.3 Display of the standpoints about social issues	- 14 -
5.4 Display of Past Election Data	- 14 -
6 Limitations	- 15 -
6.1 There are limited ways to present.....	- 15 -
6.2 The visualization can be used in a limited way.....	- 15 -
7 Tasks and division of labor.....	- 16 -

1 Visualizations

Our visualizations show the approximately complete information from different aspects of 20 candidates in 5 regions based on web pages, including basic personal information, education background, career experience, political camp, cognition degree, etc.

In the visualization, we respectively used the home page and five pages to show the general statistics of these 20 candidates and the comparisons between candidates in the same region.

A detailed explanation of this visualization is given in the following section. All the codes of our project are stored on Github [<https://github.com/jimzhang828/visualization-project>] and our data are stored in Google Sheets. We also provide a website demo on Github Pages [<https://jimzhang828.github.io/visualization-project/>], which reads data directly from the Google Sheets, so that the visualization on the website can be updated dynamically.

2 Explanation

Our objective is to provide Hong Kong citizens with enough information about candidates, which can be divided into the following three aspects.

1. Information about the candidates themselves.
2. Comparison of different candidates.
3. Information about other citizens' attitudes towards them.

In order to achieve this objective, we subdivide the above three kinds of information into different specific sections. So finally, our visualization is designed to show region, age, political camp, gender, educational background, career experience, the results of the poll and the search index these eight aspects. The following is a detailed explanation of these eight parts.

2.1 Region

In addition to the home page, we use five different pages to show the information of candidates in Hong Kong Island, Kowloon West, Kowloon East, New Territories East and New Territories West.

You can click the names of these regions to enter the corresponding page and browse the information display for candidates from this region as figure 1 shows.

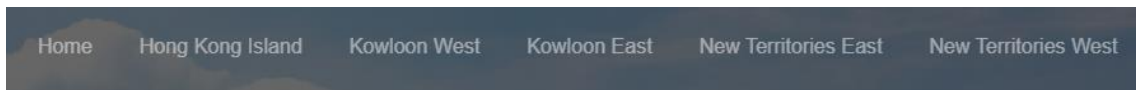


Figure 1. Pages of 5 regions

2.2 Political Camp

We integrate their political camps into other parts of visualization, mainly using colors to distinguish different political camps. Blue represents pro-establishment camp, yellow represents pro-democracy camp, and gray represents independent camp.

We can see this feature in age and investigation results visualization.

2.3 Age Distribution

The first section on our home page shows the age distribution of all candidates.

Each person's name is used as a label, and the age is indicated by the length of the bar. The longer the bar is, the older the person is.

The colors of the bars represent the political camp to which they belong. Blue represents pro-establishment camp, yellow represents pro-democracy camp, and gray represents independent camp.

And when you place the cursor over these bars, the labels of name and age appear. Also, you can sort the bars by ascending, descending, and name order as figure 2, figure 3, figure 4 show.

Age distribution



Figure 2. Ascending sorting

Age distribution



Figure 3. Descending sorting

Age distribution



Figure 4. sorting by name

Insight:

1. By sorting, we can see that people who are in Pro-establishment tend to be older than people who are in pro-democracy on average.

2.4 Gender and Education Background

This is the second section displayed on the home page.

We used a pie chart to show the gender and education background distribution of the candidates as figure 5 shows. And when you put the cursor over the pie chart, it shows the names of these people.

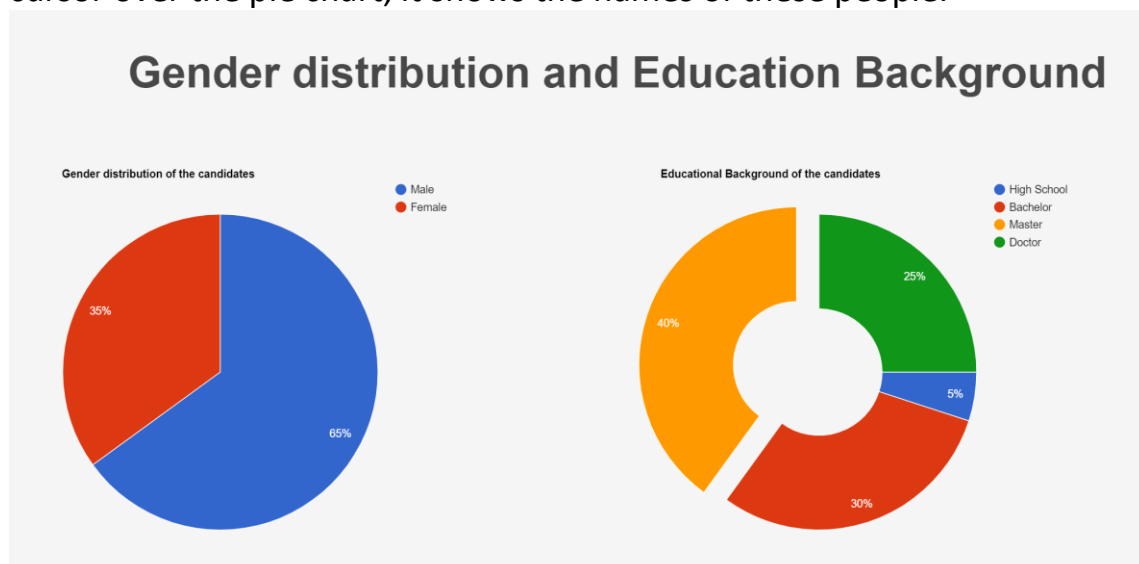


Figure 5. Gender distribution

From the gender pie-chart we can see that 65% of the candidates are male and 35% are female.

From the education background pie-chart we can see that the candidates have a high average degree, and the majority have a master's degree.

2.5 Investigation from HKU

This is the third section displayed on the home page.

We use bubble chart to show the results of investigation in HKU. Each bubble represents a candidate, and the size of the bubble represents their age. The size value of the bubble is normalized, so the difference in shape has been enlarged therefore, make it easy to compare.

Also, the colors of the bubbles represent the political camp to which they belong. Blue represents pro-establishment camp, yellow represents pro-democracy camp, and gray represents independent camp.

The range of abscissa is from 28% to 57%, represents their support rate. The range of ordinate is from 55% to 100%, represents their recognition rate.

When you place the cursor on each bubble, you can see the labels for each person, including name, support rate, recognition rate, camp and age as figure 7 shows.

Investigation from HKU

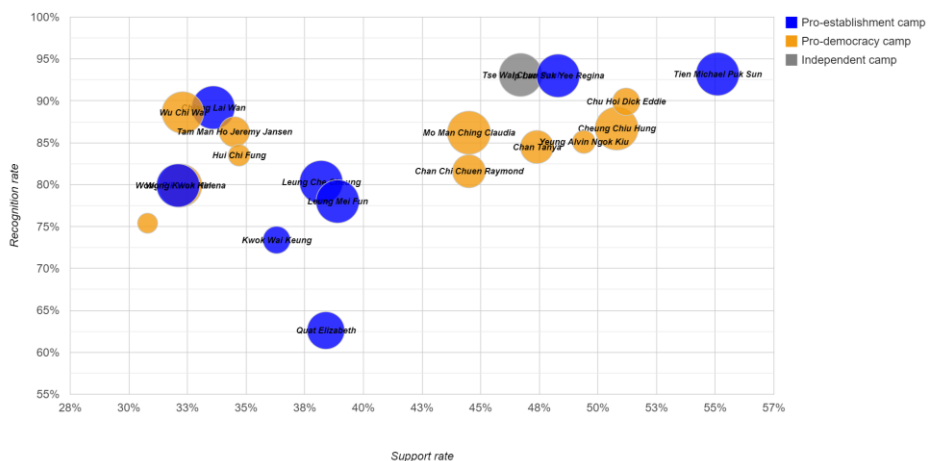


Figure 6. Investigation from HKU

The investigation results of four candidates in the same region also showed in the region pages.

Insights:

1. The average support and recognition rate is 43% and 85% respectively. And the candidate who has the highest support rate is from Pro-establishment. The Pro-democracy has an average higher support. We can get this information easily by seeing the number of the yellow bubble beyond average is more. On the other side the blue bubble is only two. This can give us one indication of the recent election of Hong Kong's district council and predict legislative council election.

2. There is no point at the bottom right of the chart. It's coincided that situation that if candidate doesn't show themselves out, they won't have a good support rate. Though some candidate makes others know him, they only have a low support rate.

3. Age sometimes means more experiences and abilities, which should bring a higher support rate. Also, more experiences and activities by the ages will make candidate more recognizable. But the age is irrelevant with the support rate and recognition rate. E.g. Chueng is only 38 years old but has a high support rate.

4. This graph can also help you find the overall state of your favorite candidate.

2.6 Career Experience

The first section on our region pages shows the work experience of the four candidates in the same region.

In order to make the citizens better understand the past of the candidates, we use Gantt chart to show the main experience of the four candidates as figure 8 shows. This work experience includes their experience as a member of the legislative council and other work experience.

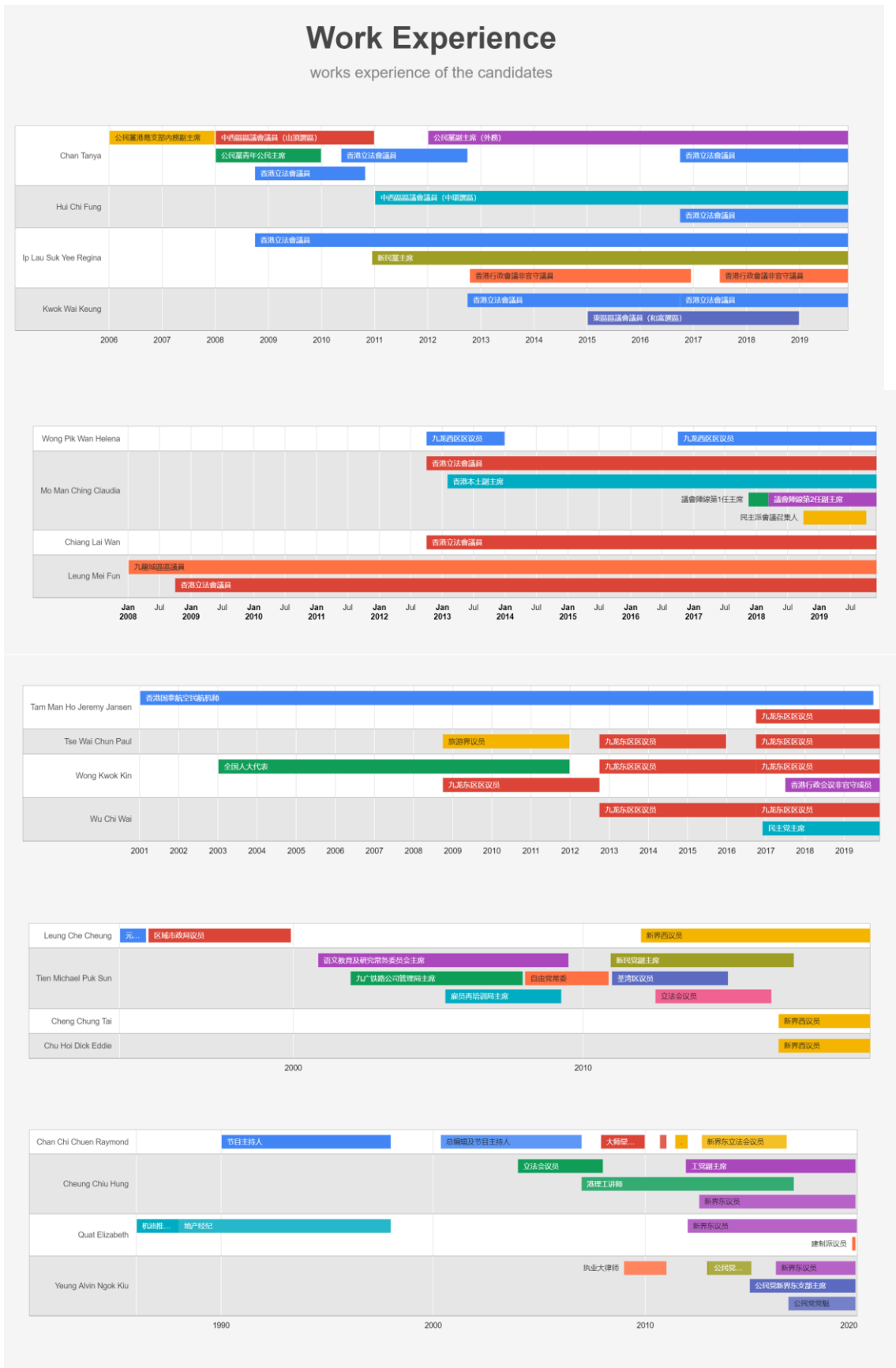


Figure 7. Work experience of candidates

2.7 Google Trends

The second section on our region pages shows the search index of the four candidates in the same region.

We collected the search index of these four candidates on Google Trends for each month in 2019 and plotted it with a trend chart using different colored curves to represent different candidates to show their search popularity on Google for HK citizens. The general pattern is shown below.

Google Trends

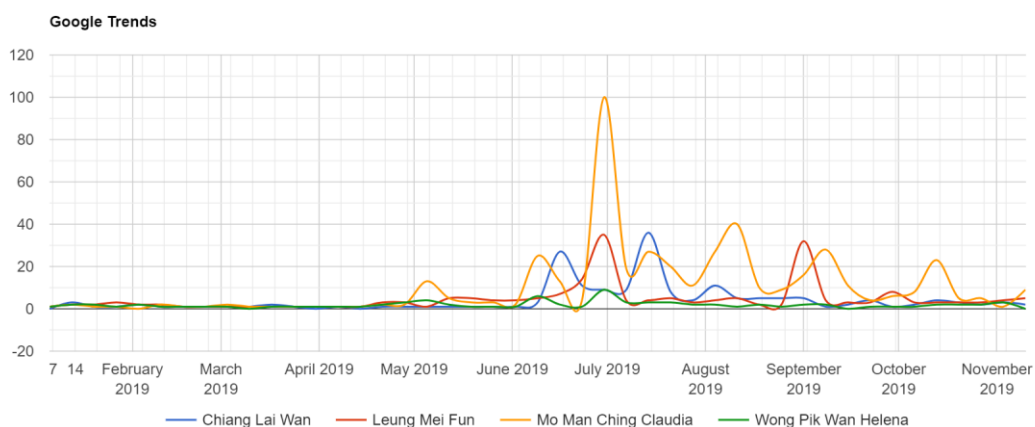
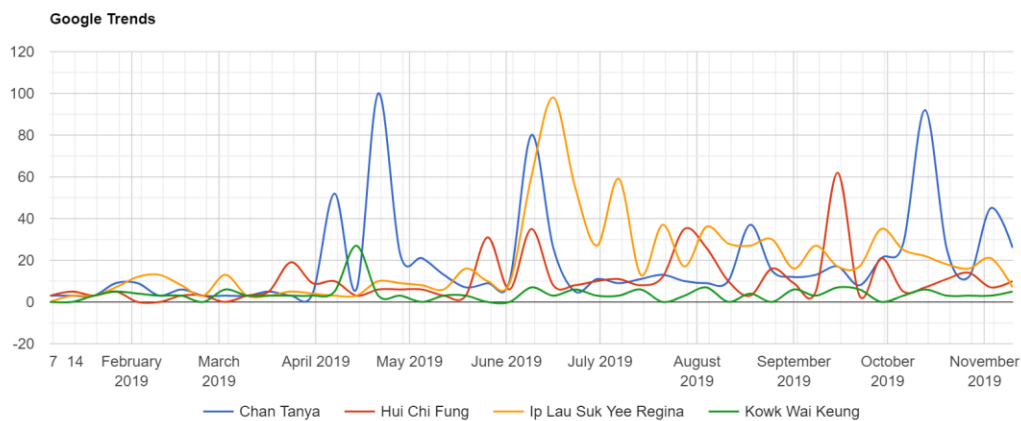


Figure 8. Google Trends of candidates

Insight:

1. Users can easily find out the most popular candidate from this chart. And according to the color, users can compare two candidates. It is obvious that this one is less popular compare to the others.

2.8 Investigation

The third section on our region pages shows the support and recognition rate of four candidates.

At the detailed investigation graph. The users can directly compare the support rate of candidate and find out the most popular candidate and his camp in a second. The graph is showed below:



Figure 9. Investigation of candidates

Insight:

1.The candidate with the highest support rate may have a high chance to win the election and financial professionals can predict the political trend in this district according to this chart.

3 Difficulties

During the whole process of designing the visualization, we encountered many difficulties, including difficulties in data collection, data processing and visualization implement. These difficulties are described in detail below.

3.1 Some data are hard to collect

In order to present as complete a picture as possible for HK citizens from the candidates themselves, the comparisons between the candidates and the attitudes of citizens these three aspects, we would like to list more information that needs to be collected at the beginning.

For example, in terms of the candidates themselves, in addition to the information now on display, we have also considered collecting their marital status, criminal records and so on. However, the channels for data collection are limited, and we spent a lot of time collecting and sorting data to obtain the above eight aspects of data.

Even if we finally decide to use only the data from the above eight areas, there are cases where the data is ambiguous or missing. We need to combine information from different sources to verify the correctness of the current data. And the method such as inference or mean value is used to fill in the data without definite value.

3.2 Some features are hard to quantify

After data collection, we found that most of the data we got were textual data, such as education background, party affiliation, work experience and so on.

For text data, it is difficult to visualize and compare. In order to visualize the data, we need to quantify it first, such as using classification methods to transform it into statistical data. For example, the candidates' education background is divided into four categories: high school, bachelor's degree, master's degree and doctor's degree. Then we can visualize it.

For the data that cannot be quantified, we need to choose the most appropriate display method to achieve better visualization effect.

3.3 Some data is difficult to find a suitable display chart

Since some data volumes are small, visualization can be difficult. For example, there are only 20 age data in total. If we use bar charts or pie charts, the information that can show is very limited.

Therefore, we added political camps and names to the visualization of age distribution and added sorting function to make the visualization express more effective information.

As mentioned above, the selection of display chart for text data also brings us difficulties. The candidate's work history, for example, makes it difficult to choose a visual chart. After many discussions, we found that the Gantt Chart could be used to show the time and duration of the candidates' work experience, and four candidates could be compared on the same chart. Therefore, we finally chose Gantt Chart for the visual display of this part.

4 Methods & Justification of the choices

We used Website and Google Chart to design the visualization tool.

4.1 Choice for Website

At first, we consider different method from the class, including Tableau, D3 and so on. However, we think a good visualization tool should have a good sharing ability and easy to access for users. A website is definitely a good choice, because nearly everyone can open a website by their smartphone or electronic equipment.

4.2 Choice for Google Charts

Google Charts is an extremely popular and complete tool that can visualize data on our website. Besides, Google Charts provided us with lots of various diagrams that we need, and the documents are complete and are frequently maintained by Google. Also, it provides a lot of API for ingesting chart data from other sources like Google Sheets that we adopt.

5 Improvements

There were other things we wanted to show when we designed

visualization. For various reasons we were not able to implement these in the visualization design. The details are as follows.

5.1 Display of Party Grouping

We collected the party groupings represented by the candidates. We planned to use a tree to show where they are in each party, whether they hold important positions, etc. But such data are hard to come by, and we can only get a rough framework of the organization from the party's website. Visualizations based on this data do not present valid information.

If appropriate channels can be found to obtain such data, the party affiliation of the candidates is also a part worthy of attention by Hong Kong citizens.

5.2 Social Network Mapping

If we have more time and energy, we can get the social information between candidates from Facebook or other social medias and draw it into a social network graph.

This social network graph can show whether they are close to each other, whether they are well connected, or whether there are small groups. It will also be an effective message for the Hong Kong citizens.

5.3 Display of the standpoints about social issues

If conditions permit, we can use crawlers to get the candidates' opinions and positions on recent hot issues in social networks or news interviews. The frequency of their speech can reflect their concern about social issues, and their standpoints can also be reflected in their speech.

This part of the visual display can let the public know more about a candidate and help them to infer how the candidate will represent the public in the future.

5.4 Display of Past Election Data

Not enough information is on display about Hong Kong citizens' attitudes towards candidates in our visualization. We only chose the results of the poll in HKU for presentation.

Actually, if we can find enough data about the results of the past Legislative Council election, we can use these data to show the past performance of some candidates because many of them have run for the councilor before.

6 Limitations

Reviewing our visualization, there are two main limitations as below.

6.1 There are limited ways to present

Due to the limitation of the data we collected, we can only choose some visualization methods like we presented now to show the information we obtain.

In fact, there are many other good presentation methods, such as tree diagram, network diagram, geography diagram, thermodynamic diagram, etc. But we have no chance to apply to our visualization.

Also, there are many dynamic interactions that enhance the usability of visualization. We haven't been able to apply these interaction methods.

6.2 The visualization can be used in a limited way

Our visualization only used regions for categorization. We compare the all the candidates on the home page or compare the four candidates in the same region on the region pages. So the visualization can be used in a limited way.

If we can compare candidates under more different categories, the users can have an alternative perspective for these information. For example, in addition to dividing them by their region, we can also divide them by their party groupings and compare them. We can also divide them by age, education background even by their level of activity and compare them.

In this way, we can have many different ways to use and understand this visualization.

7 Tasks and division of labor

We divided the work of the project into the following seven tasks, The completion of each person is shown in the table 1.

Every team member was involved in collecting data, analyzing data, implementing the code and writing the final reports. And Zhang Wenjun and Wang Yao together completed the construction of the website framework. Zuo Qing and Zhang Wenjun chose the plot together. Wang Yao and Zuo Qing finished the test together.

Tasks	Members
Collect data	ALL
Analyze data	ALL
Website framework	ZHANG WENJUN, WANG YAO
Choose plots	ZUO QING, ZHANG WENJUN
implement	ALL
testing	WANG YAO, ZUO QING
Write final reports	ALL

Table 1. Tasks and division of labor