**Names of Team Members (max. 3 members):**

* **Yun Lin**
* **Rongjia Jing**

### *Notes:*

### *First, please re-save this document on your computer, RENAMING the file to contain your last names (e.g.students Sachar & Lemberg would name their document as Tempate\_1\_Sachar\_Lemberg.)*

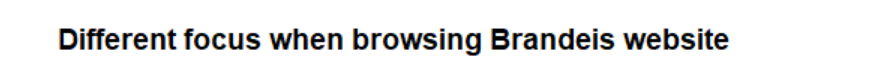
### *Point values of each part are shown below; 10 points will be allocated for the quality of your business writing (organization, clarity, grammar, etc.).*

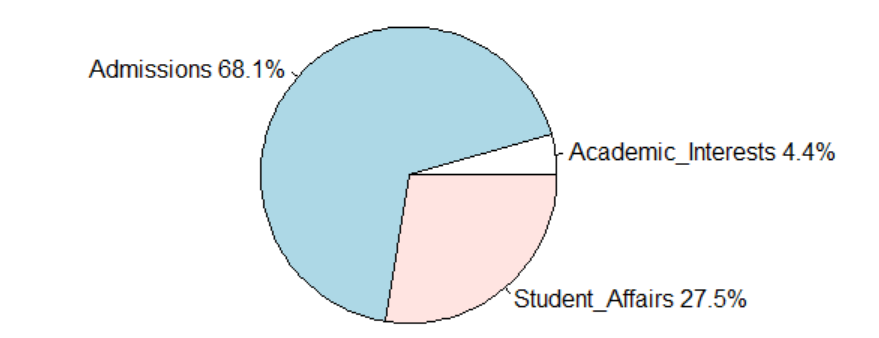
### *All team members will receive the same grade. It is up to the team to ensure that all members deserve the same grade.*

### *Type or paste your responses into the boxes below. The boxes will expand to fit your answers.*

A. Seven visualizations (20 points)

**1.**



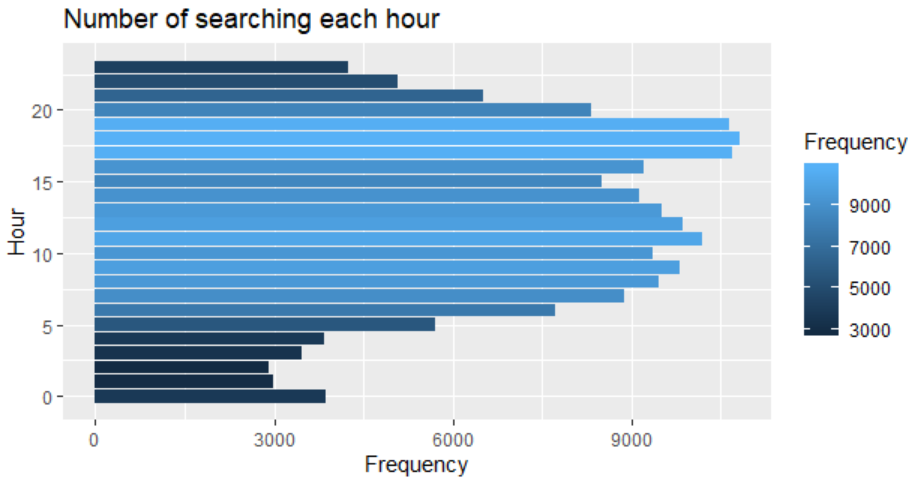


**2.**

图片包含 屏幕截图

描述已自动生成

**3.**



**4.**

Boxplot of average time on page among people of different focus

图片包含 屏幕截图

描述已自动生成

**5.**

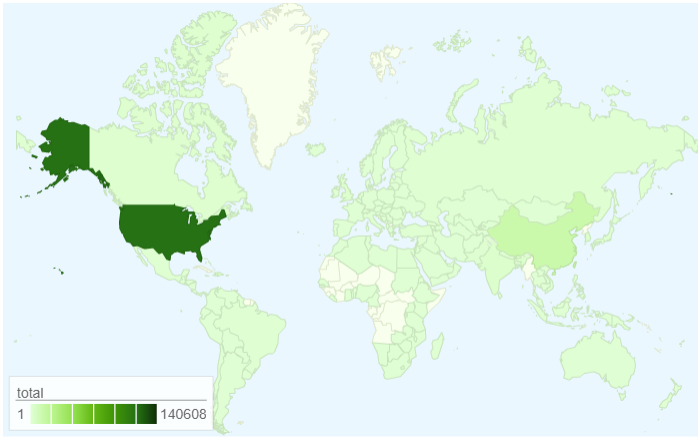
Word cloud of “pagetitle”

图片包含 文字

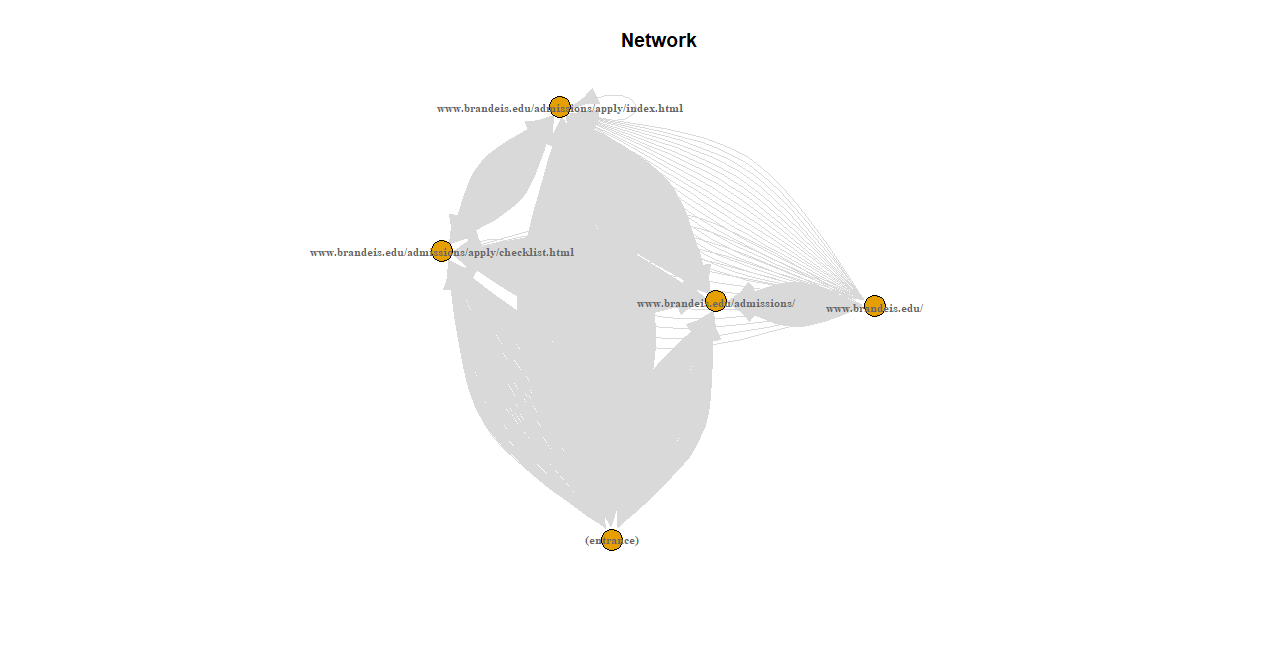
描述已自动生成

**6.**

<http://127.0.0.1:16807/custom/googleVis/GeoMapID19e46b147d81.html>



**7.**



B. Explanation of Choices

1. For each visualization that you created, briefly describe your thought process: why did you select the variables (attributes and metrics) that you chose? Why did you select the particular type of visualization for these variables? What questions are you trying to answer with this visualization? (20 points):
2. At first, we try to have a basic understanding of people’s purpose of visit. So, variable “Focus” (describing different focus when browsing Brandeis website) was selected. And pie chart labeled with different focus’ names and percentage was used because pie chart can more intuitively present the proportion of each part.
3. Variable “Avgtimeonpage” was selected to visualize visitors’ average time on page and scatterplot was used because it can clearly show the percentage of each category “newvistors” and compare the frequency difference between each number of visitors.
4. Trying to find out at what time people usually visit Brandeis’s website, “Datehour” (describing the time people visit the website) and bar plot was selected since it can clearly show the difference between each group.
5. Then the group was separated according to variable “Focus” to “academicinterest”, “admissions” and “student affairs”, and we can see the statistic difference of these three group of people of their average time on pages.
6. To find out the frequency of key word existing in the “pagetitle”, a word cloud is drawn to show the time of occurrence with different colors and size. The darker the color, the bigger frequency of word ouucrence.
7. And a map was drawn to show times of visiting Brandeis’s website in each country all over the world. “Country” (describing the country each visit was from) was chosen. The darker the color, the more visits, which gives a more intuitive visualization.
8. Finally we want to figure out the pathway of visitors, including the entrance, linkage page and the website visits most. So a network diagram is drawn to show the path of Brandeis web visitors and the websites they visit.

C. Interpretation

1. What insights about Brandeis web traffic in November/December does your visualization reveal? Explain how specific visualizations clearly provide these insights. (20 points):

First, we can draw some conclusions on the visitors’ behavior according to the bar chart and scatterplot. As is vividly demonstrated, the majority of people visit Brandeis websites during 5pm to 7pm or around 11am and they almost spend fewer than 100 seconds on page on average.

Second, we can get some information about the characteristic of visitors by map plot. As the color on the map shows, most of the people visited Brandeis websites were from North America, and the second largest amount were from China.

Third, we can also get the visitors’ focus information through the pie chart. Visitors’ focus are mostly “admission” and only a few people focused on “academic interests”. The pie chart clearly shows that 68% of the visitors focused on “admission”, while 27.5% focused on “student affairs” and only 4.4% focused on “academic affairs”.

In addition, the “wordcloud” diagram and network graph further demonstrate the visitors’ preference and searching purpose. According to the “wordcloud” diagram and network plot, we can know the word “apply” has the biggest number of frequency ; and “service”, “student”, “financial” and “academic” follow.

The majority of visitors entered directly into [www.brandeis.edu/admission/apply/index.html](http://www.brandeis.edu/admission/apply/index.html) from entrance and turn to [www.brandeis.edu/admission/apply/checklist.html](http://www.brandeis.edu/admission/apply/checklist.html), or [www.brandeis.edu/admissions/](http://www.brandeis.edu/admissions/). Most important, [www.brandeis.edu/admissions/](http://www.brandeis.edu/admissions/) acts as the linkage of other websites, as it connects the rest four websites. Therefore we can safely say the visitors are interested in “admission” affair most because they start at the “apply” linkage to other related websites.

1. What additional questions do your visualizations raise for further investigation? (10 points):

One question is why student enter the website by linkage “[www.brandeis.edu/admission/apply/index.html](http://www.brandeis.edu/admission/apply/index.html)” rather than “[www.brandeis.edu/admissions/](http://www.brandeis.edu/admissions/)”, because the latter one is the homepage of the website. And if people enter it by the first link, it seems like they are very goal-oriented to apply the school, and the wordcloud shows the same result as well. So we can speculate that Brandeis admission center’s adverting works very well and we can have further investigation on how does Brandeis advertise on other websites.

D. Improvement

1. Assuming we can always improve, select one of the graphs to take a fresh look and to think of better way of displaying the information. Critically evaluate the choice of visualization for the current graph and comment on ways in which this graphing approach fall short? Explain your thinking. Change visualization by replacing the initial display type with another. Very briefly comment on what you did and the ways in which it improves or detracts from the original display. Would users find the new display more or less helpful? Explain. (20 points)

The current scatterplot of visitors and average time on page still has space to improve. It has separated the group according to visiting times and show the distribution of each group’s average time on page. However, it cannot shows the trend of change between different groups.

So we improve the scatterplot to a new plot including the histogram and density line, and we limited the x axis maximum to 150 cause if there is no limit, the graph will show that over 80% of the average seconds on page was around 0 second. In the graph, we can see the change of average time on page when the group number increase. Even though the majority of people spend approximately 0 second on the page, we still see a small group of visitors spend 15000 seconds, and some groups of people spend 5000 seconds on the page. So the new graph displays information that way more helpful for users.

