

Website Usage Data Visualizations

Introduction

The internet's domination over our generation has left us tethered to our devices almost continuously. This also means that information, perhaps on a subconscious level, can now be made more available. We can see how any one of us changes during the various circumstances of one month. What, how often, and when are we working and playing for instance. Do we work more at home or at work? When are we most active on the internet? There should be patterns here, but how strong are these patterns, generalized over 4 individuals? We can even answer, though weakly, "who may be the laziest of the group?" (that being the one that watches the most streaming content). We can even incorporate weather data and see how that affects our internet usage. Another interesting question is how do non-daily events (exams, holidays, news events, etc) change our usage of the internet for those sets of days? For a company this could be useful in determining when to send a push notification, for a single user this could be helpful in optimizing self productivity by recognizing subconscious patterns.

Our project's goal is to begin to explore these questions over a very small group of people, four to be exact. We are using our own group's online history to investigate the way in which these questioned could potentially be answered and visualized on a much larger scale.

Data

In order to gather data for this project our group formed a shared Google Sheet where we manually entered our web history every day for a month (October 8 - November 8) along with the date, the time of day (morning, afternoon, evening, or night), and our location (home, work, or school). In a Jupyter notebook we broke up our data into four dataframes, one for each of us, and enacted pre-processing. Our data consists of a column for each of the websites, with the rows as a different date and time of day and the counts of each website for that time of day. A sample of that data can be seen below.

	Date	Time	Gmail	Blackboard	Vulture	AVClub	VanityFair	Eater	Pitchfork	YouTube	Twitter	Amazon	Slate	NewYorkTimes	Location
0	2019-10-08	Morning	6	0	0	0	0	0	0	0	0	0	0.0	0	School
1	2019-10-08	Afternoon	6	8	6	4	1	1	5	8	6	0	0.0	2	Home
2	2019-10-08	Evening	2	0	1	0	0	2	0	0	10	3	0.0	0	Home
3	2019-10-08	Night	10	0	4	1	1	1	4	2	1	0	5.0	0	Home
4	2019-10-09	Morning	2	0	6	4	4	1	3	7	4	0	0.0	3	Home

Table 1: Sample raw dataset

We added a day of the week column from our initial date, and then decided to break up our websites into four categories: education, news, social media, and streaming. This was done because the sites we frequent the most vary person to person. We decided to create these categories so that we can make comparisons between the four of us using the same categories. Thus we added a column to each of our datasets for the four categories and an additional column “Total” that sums all of the individual website columns. These manufactured columns can be seen below.

DayofWeek	Social_Media	News	Streaming	Education	Total
1	0	0.0	0	6	6.0
1	6	7.0	8	14	47.0
1	10	0.0	0	2	18.0
1	1	9.0	2	10	29.0
2	4	6.0	7	2	34.0

Table 2: Sample pre-processed dataset

Methodology and Results

Plot One - Grouped Bar Chart

The first visualization we focused on was a grouped bar chart that investigates how time of day effects category. As you can see in the image below, each member of our group has their own bar

chart that shows counts of website views on the y axis and time of day on the x axis. Each of our individual bars is a different website category and our groups are morning, evening, afternoon, and night. We created this bar chart using D3.js and added an interactive feature that allows you to hover over each bar and get the exact count value. When looking at the category feature, we can see that all four of us visit educational sites more than the other three categories, Kei visit news sites less than the three of us, and Steven visits streaming sites more than the three of us. Likewise when examining the time of day feature, it appears that Cairo, Kei, and Zoe are less active at night than the other times of day, and Steven's website usage remains steady throughout all times of the day. Finally, when looking at the visualization as a whole it is evident that Steven's website usage on average is the largest while Kei's average usage is the smallest.

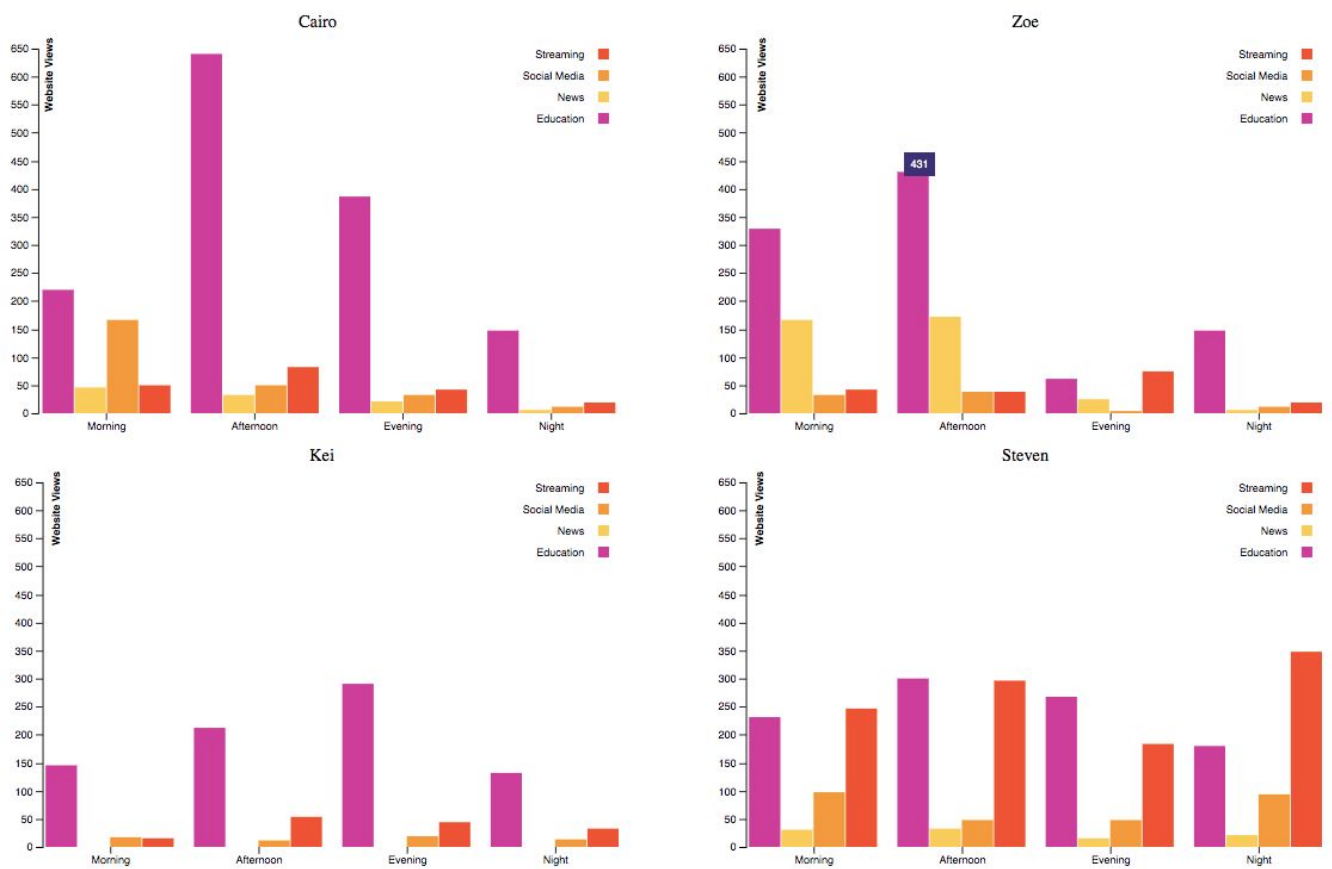


Figure 1: Grouped bar chart of website views by time of day and category of site

Plot Two - Pie Chart/Donut Chart

The second visualization we chose to use were donut pie charts. The pie charts represent the overall month's average web usage for each group member split into 4 different categories,

streaming, education, social media, and news. The general use of this visualization being the comparison, first between each group member, and second between each category within each group member's individual graph. A pie chart is the best visualization tool to express these comparisons because it is a part-to-whole analysis, each category of internet use compared to the overall amount of internet use. It clearly expresses the difference in each piece quickly, and is a straightforward graph that can be understood by an audience without any statistical background. The benefit of using a donut pie chart versus a classic pie chart is the white space in the middle. The white space draws the eye of the audience or viewer, bringing attention to the comparison we are attempting to express.

The other important aspect of our donut pie charts is the chosen color pallet. The sharp contrasting hues are always a good tool to use when demonstrating separate, non-continuous, and unconnected categories. The colors make it blatantly obvious to the viewer the difference between each category. We did not use a progression of luminance or saturation to avoid the insinuation of continuity or dependence between each category.

The biggest limitation to a pie chart is that it expresses a quick, basic understanding of the difference between categories but does not provide numerical, or exact differences. It allows only for a rudimentary visual comprehension. In order to compensate for this limitation, we added a component to give the percentage of each category that shows up when the cursor is above said category. This not only helps accommodate for the constrictions of a pie chart, but gives the visualization another dimension to draw the audience in.

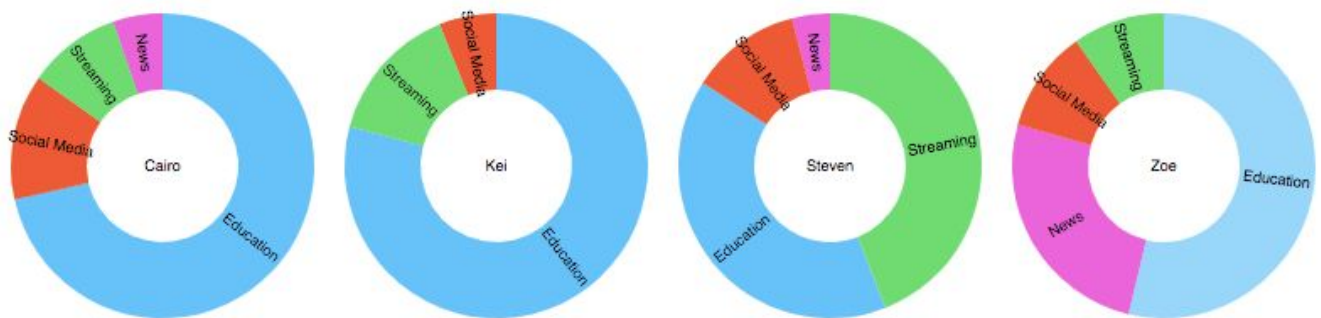


Figure 2: Donut chart of individual group members website usage by category

Plot Three - Line Chart/Stream Graph

The primary purpose of the third visualization is to compare each member's website usage by days. For this purpose, we decided to use the line chart to visualize our data because it allows the

user to easily compare each member's number of clicks on the same day. Each of four lines in the chart represents each member in our group and shows the total number of clicks per day on the y-axis. To make a comparison easier, we implemented an interactive feature where the user can see the absolute value of clicks for each person on the same day by hovering over a line in D3.js (shown in the image below).

As the chart below shows, each of us has a different web usage pattern. For example, the number of websites Steven browses is more than the average in most of the days while Kei's line usually comes below the average. However, the maximum number of clicks in a day is 80 made by Cairo, who sometimes shows spikes in the chart. When Cairo marks the highest in a day, the numbers reach more than 70, whereas the maximum number of clicks made by Zoe and Kei is around 60. Zoe's line seems to constantly jiggle. It's appearing that when she browses more websites than usual, she sees fewer websites in the next day.

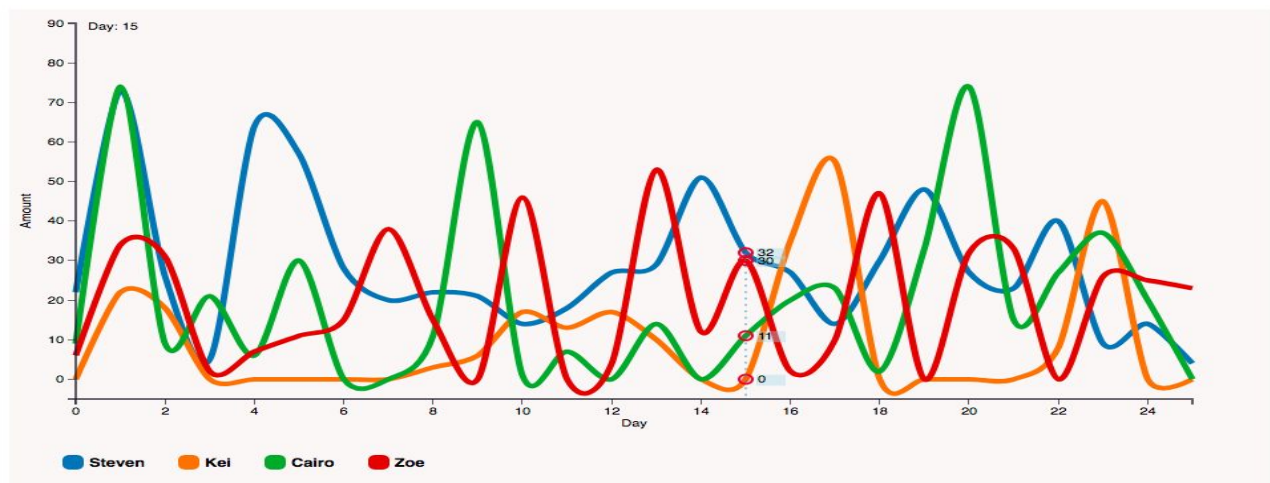


Figure 3: Line chart of website counts over time

Though line chart allows us to compare each member's website usage per day easily, it's hard to compare the trend of the web usage of each member. Thus, we added the stream graph to compare each member's web usage as the whole. As you can see from the graph below, Steven constantly uses the web on a daily basis. It's interesting to note that Zoe and Kei seem to show a similar trend, meaning that when Zoe visits more websites than the previous day, Kei also tends to visit more websites than the last day. If we focus on the difference between the web usage in the weekdays and weekends, almost all of us visit fewer websites in every weekend than we do in weekdays.

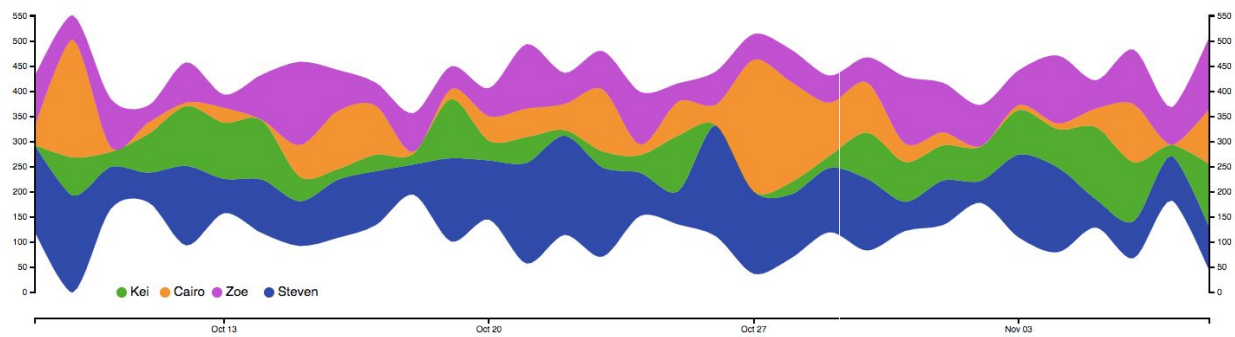


Figure 4: Streamgraph of website counts over time