Visualizing Snapchat Usage and Physical Activity Categories through Self-Tracking

Motivation:

I've always been interested in learning more about my daily life through means that take advantage of my professional skillset and ability to use technology to generate knowledge. I discovered the idea of self-tracking through a college class and began collecting different information that would allow me to create data visualizations and learn something new that I had not known before. As graduating senior in college, I had noticed halfway through my freshman year that I was beginning to use social media a lot more than I had in high school. In addition, I was becoming more and more lethargic after a long days' worth of classes, homework, and (often times) fast food. When confronted with the task of developing some kind of data to track about myself, I decided to note my daily Snapchat usage (a social media application) and overall physical activity. Some examples of nagging questions that I wanted answers to were: Who was I Snapchatting the most? At what hour was my step count the highest? What were the overall trends of total daily Snapchats and calories burned?

Methodology:

I began the self-tracking on May 27, 2018 and ended on June 16, 2018. As this was a project for a summer class, I had a limited amount of time to gather the data and make sense of it using visualizations. While I tracked two very different genres of data, the collection process was quite unique for both. First, I differentiated the datasets into "active" and "passive" categories. The active data was something that I had to track by hand every day. I chose to track the number of Snapchats I interacted with daily, where each Snapchat contained relevant metadata that will provided deeper insights into the app usage. The metadata included the following fields: type (sent or received), medium (photo, video, or chat), person's first name, time of Snapchat, and the

place from where we first met before I added them as a friend on Snapchat (high school, college, or work). Every night at 11 PM, I recorded the Snapchats in a Microsoft Excel document. The passive data was something collected automatically using a third-party tracking device. In my case, this device was an Apple Watch, Series 2 Edition. Although the watch maintains sixteen unique identifiers, I incorporated the following three into my visuals: number of steps walked, exercise minutes, and calories burned. After the tracking period was over, I was able to view the passive dataset by using Apple's native iOS Health application that allowed me to export the data into an XML format, which I ultimately manipulated via programming. Examples of both raw datasets are listed below:

Active Data

Sender	Receiver	Date	TimeSent	Medium	Affiliation
Sumanyu	Alec	5/27/2018	0:15:00	Photo	College
Sumanyu	Navneeth	5/27/2018	3:44:00	Photo	HighSchool
Chase	Sumanyu	5/27/2018	5:44:00	Photo	Work
Alec	Sumanyu	5/27/2018	6:58:00	Photo	College
Sumanyu	Stephanie	5/27/2018	8:55:00	Photo	HighSchool
Stephanie	Sumanyu	5/27/2018	9:01:00	Photo	HighSchool

Passive Data


```
<Record type="HKQuantityTypeIdentifierDistanceWalkingRunning" sourceName="Sumanyu's Apple Watch" sourceVersion="3.1.3" device="&lt;&lt;HKDevice: 0x28317
bd90&gt;, name:Apple Watch, manufacturer:Apple, model:Watch, hardware:Watch2,4, software:3.1.3&gt;" unit="mi" creationDate="2017-04-10 22:31:51 -0400" st
artDate="2017-04-10 22:26:12 -0400" endDate="2017-04-10 22:27:13 -0400" value="0.0381317"/>
 <Record type="HKQuantityTypeIdentifierDistanceWalkingRunning" sourceName="Sumanyu's Apple Watch" sourceVersion="3.1.3" device="&lt;&lt;HKDevice: 0x28317
bd40&gt;, name:Apple Watch, manufacturer:Apple, model:Watch, hardware:Watch2,4, software:3.1.3&gt;" unit="mi" creationDate="2017-04-10 22:31:51 -0400" st
artDate="2017-04-10 22:27:13 -0400" endDate="2017-04-10 22:28:15 -0400" value="0.0149206"/>
```


To create the final visualizations, I used R, RStudio, and various graphing and data manipulation packages. Before writing the code that made the visuals, I invested many hours into sketching the outlines and visions for the early stages on graph paper. This allowed me to see concrete examples of what I envisioned the results looking like and an easier means of modifying portions without having to redraw entire sketches. Instead of using drag-and-drop graphic-design tools such as Canva or Easely, where images are prebuilt and must be

concatenated together to form a coherent visual, my method of developing the visualizations allowed me to customize my visuals without relying on premade shapes and figures. As such, I cleaned the raw data using the dplyr library and visualized it using the ggplot2 library. In the below images, I first selected data that involved calories, filtered it by the dates of my collection period, and performed mathematical aggregations that resulted in a data frame containing the information to be visualized.

```
calsDF <- df %>%
  filter(type == 'HKQuantityTypeIdentifierActiveEnergyBurned') %>%
  filter(as.Date(date) > as.Date('2018-05-26')) %>%
  filter(as.Date(date) < as.Date('2018-06-17')) %>%
  group_by(date,dayofweek) %>%
  summarize(cals=sum(value)) %>%
  mutate(Week=ifelse(as.Date(date) > as.Date("2018-06-09"), 3,
    ifelse(as.Date(date) < as.Date("2018-06-03"), 1, 2))) %>%
  print (n=100)
```

	date	dayofweek	calories	Week
	<chr>	<ord>	<dbl>	<dbl>
1	2018-05-27	Sunday	728.	1
2	2018-05-28	Monday	1159.	1
3	2018-05-29	Tuesday	659.	1
4	2018-05-30	Wednesday	676.	1
5	2018-05-31	Thursday	540.	1
6	2018-06-01	Friday	404.	1
7	2018-06-02	Saturday	716.	1

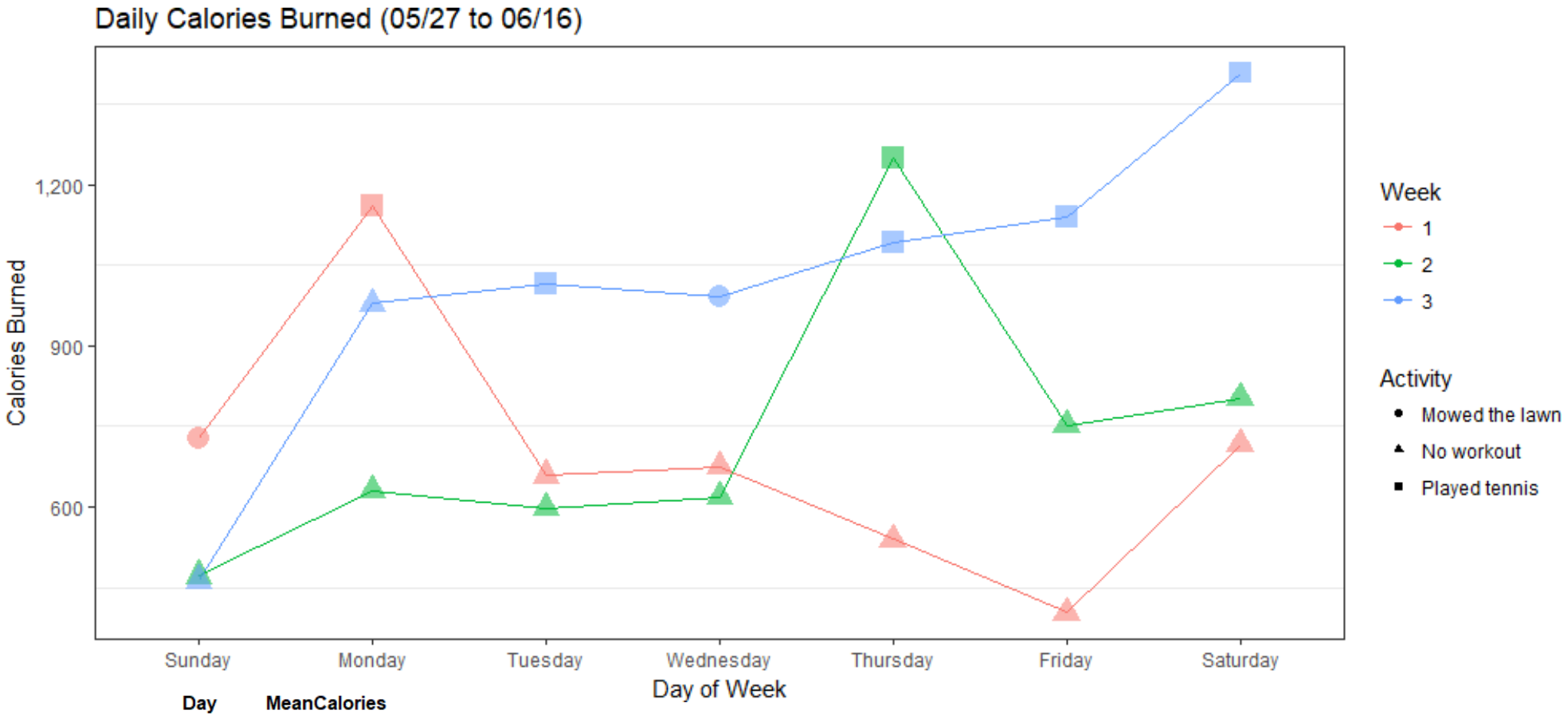
The data frame was then converted to a ggplot2 object by adding dots, lines, colors, and labels. Sample code corresponding to the plot is listed below:

```
ggplot(calsDF, aes(x=dayofweek, y=cals, color=Week, group=Week)) +
  geom_line() +
  geom_point(aes(shape=Activity, size = 2, alpha=0.5)) +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_brewer() +
  theme_bw() +
  theme(panel.grid.major = element_blank()) +
  guides(size=FALSE, alpha=FALSE) +
  labs(title = 'Daily Calories Burned (05/27 to 06/16)',
    x = 'Day of Week', y = 'Calories Burned')
```

Conclusions:

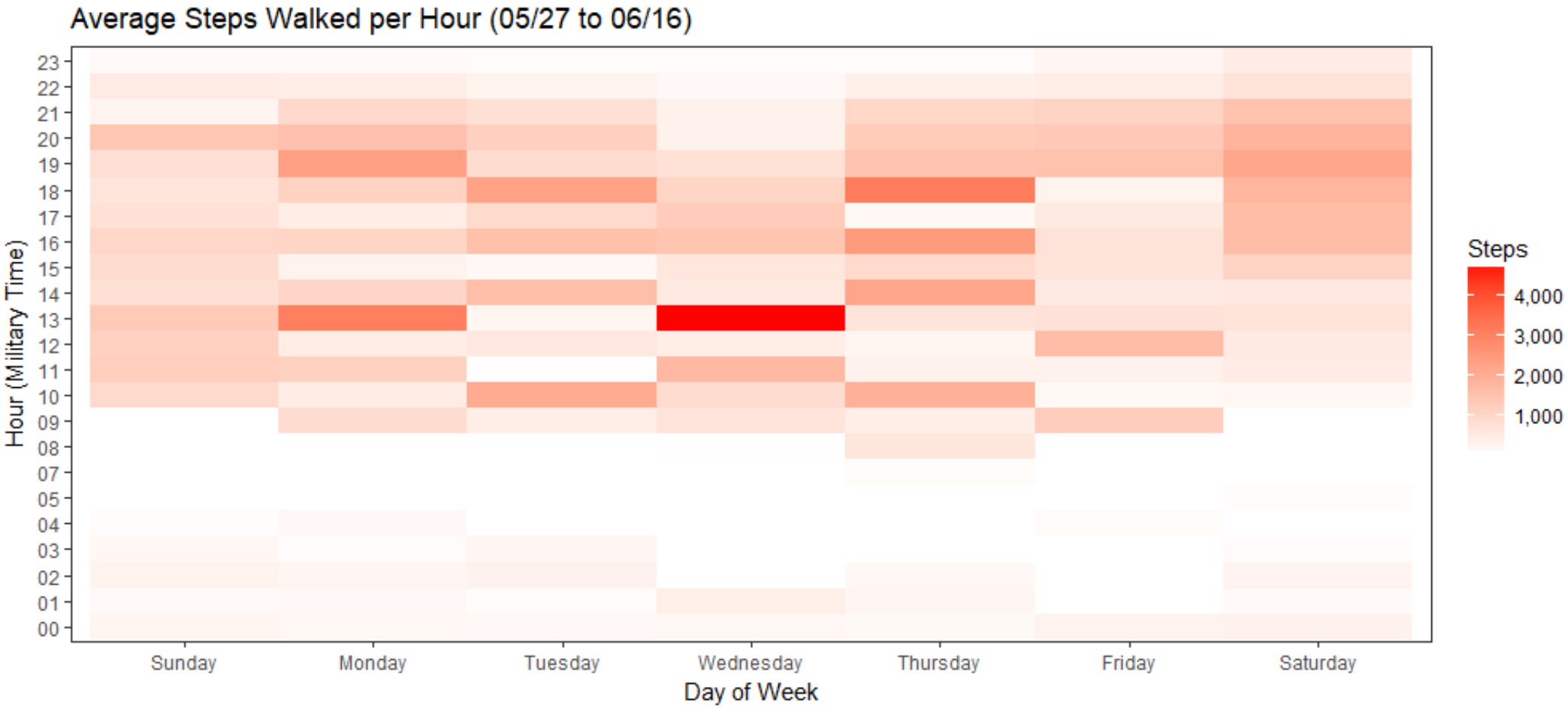
Instead of attempting to develop a relationship between the two datasets as I tried doing in my initial passthrough, I focused on answering questions I had about each collection individually. I discovered that I was physically active the most on Saturdays, usually playing tennis as a form of exercise. I used Sundays to rest and recover from the entire week's worth of activities and classes. On Mondays and Wednesdays at 1 PM, I walked the most amount of steps due to the large distance I had to cover between classes. Finally, I got the most sleep on Wednesdays and Sundays, as I was not active between the early morning hours of those days. Regarding Snapchat usage, I kept in touch with my high school friends the most, using photos instead of chats or videos to maintain conversations. Although I attempted to reduce the app usage over three weeks, the opposite happened because of matching schedules between my friends and I. Additionally, I used Snapchat the most during weekdays as weekends were less active because of familial commitments, completion of homework, and physical activity.

In the future, I'd like to incorporate more categories of data that my watch collects and create different visualizations to generate new knowledge about my daily life. To replicate this experiment, I would advise others to be diligent while recording their Snapchat usage. Often times, after an exhausting day of classes and homework, I would be too tired to enter data by hand. To combat this, I instead maintained a note on my phone that recorded names and times of Snapchats so that I could transfer that data later on. This helped me stay motivated during the collection process and ensured that I did not spend too many long stretches doing the boring job of data entry. Developing an automated way to export the necessary Snapchat metadata would be a longer-term goal. This would save time and allow for more attention to be paid to details in the visualization-design phase.

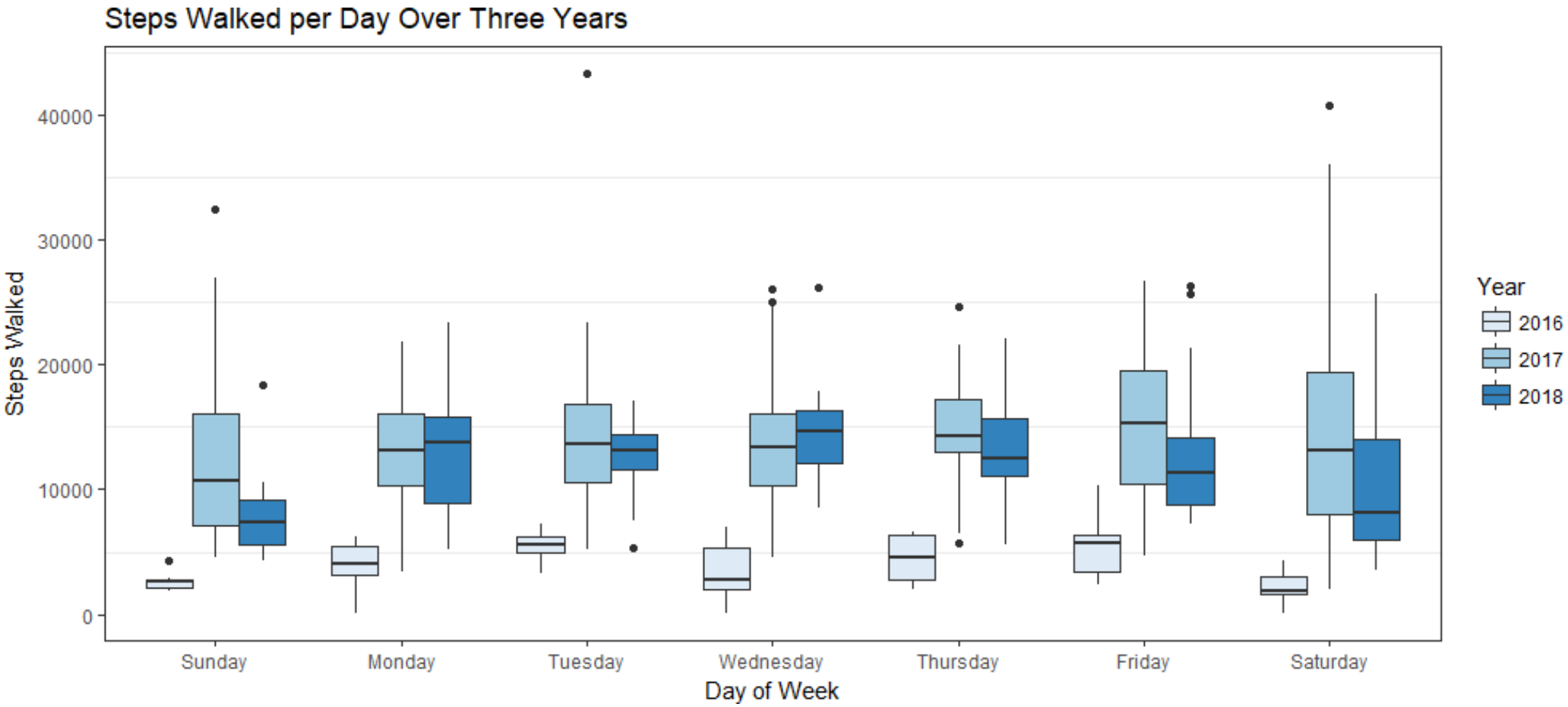


Day	MeanCalories
Sunday	537.1662
Monday	922.8833
Tuesday	757.3980
Wednesday	762.4510
Thursday	960.8413
Friday	765.5627
Saturday	975.0900

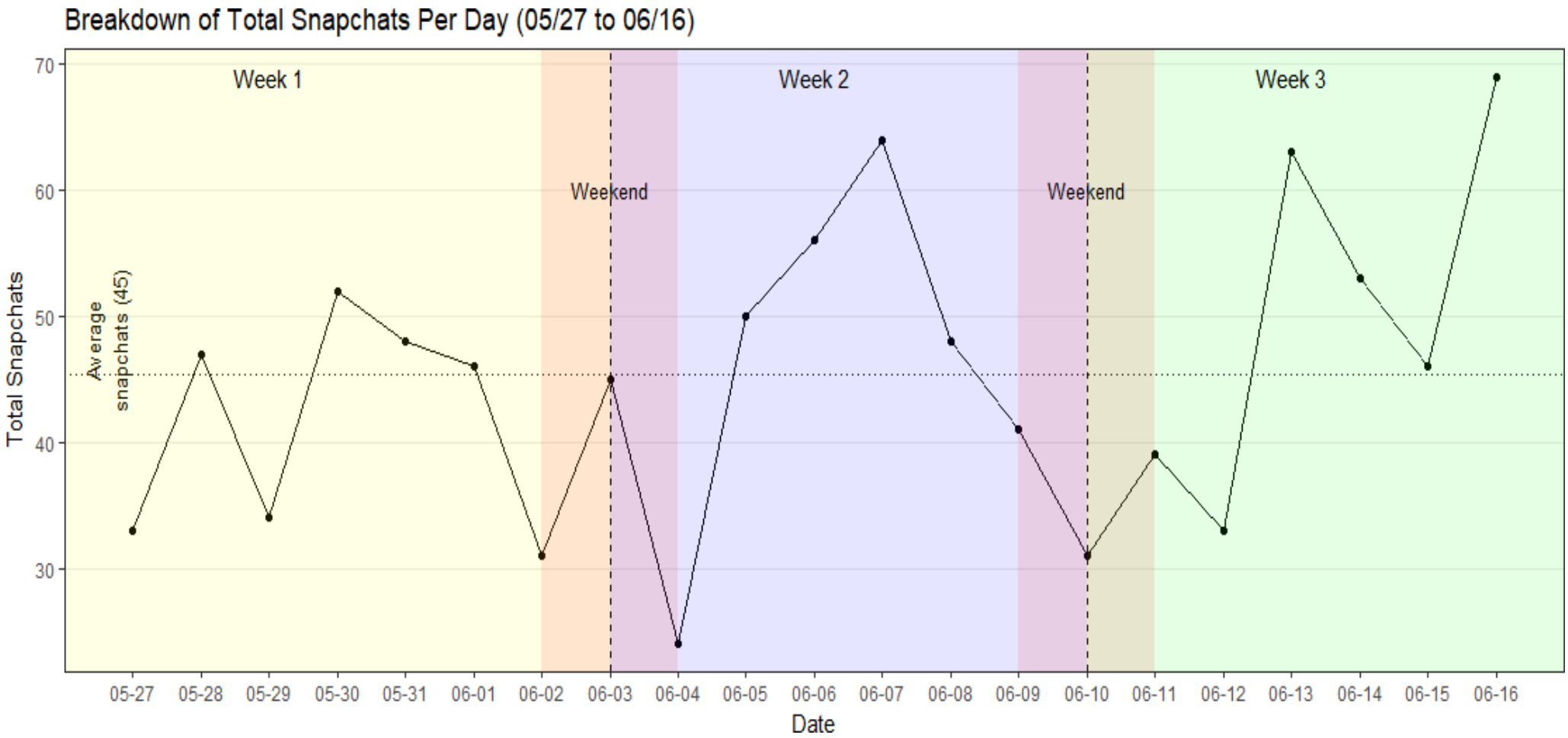
This line plot displays the number of calories I burned on a daily basis over three weeks. The colors represent the progression of calories across each week. Various shapes such as circles, triangles, and squares are used to indicate the type of activity I performed on that specific day. Every time I played tennis, I was able to burn more than 1000 calories. A table at the bottom left aggregates the daily calories and displays the average burned for the given day of the week. Mondays, Thursdays, and Saturdays were when I burned the most calories. Sundays were clear rest days because it had the lowest average calorie count. I was more active in the third week of data collection than the first two.



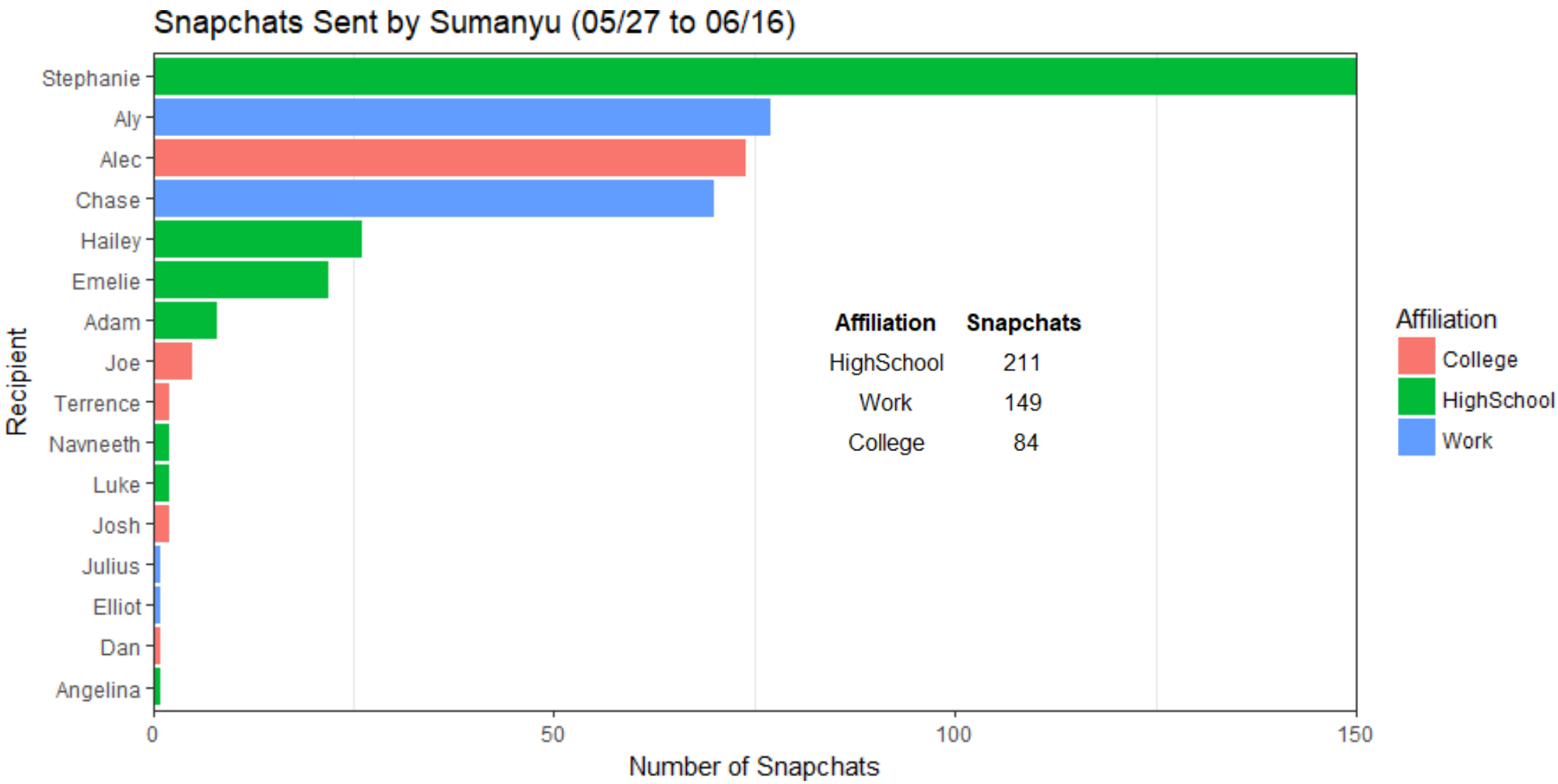
This heatmap uses shades of colors to indicate the frequency of steps walked on average per hour on a daily basis. Each day is split into hourly chunks as marked on the y-axis. The darkest shade, red, means that I walked an average of almost 5000 steps between 1 and 2 PM on Wednesdays. This time during the week, I am always walking the most distance because I have two classes located on opposite sides of my college campus. The heatmap also conveys an additional piece of information: my sleep schedule. On the bottom half of the map, white boxes indicate hours where I was asleep due to no walking data recorded. During early Thurday mornings, on average, I did not sleep as much. However, Wednesdays were the days where I got the most sleep.



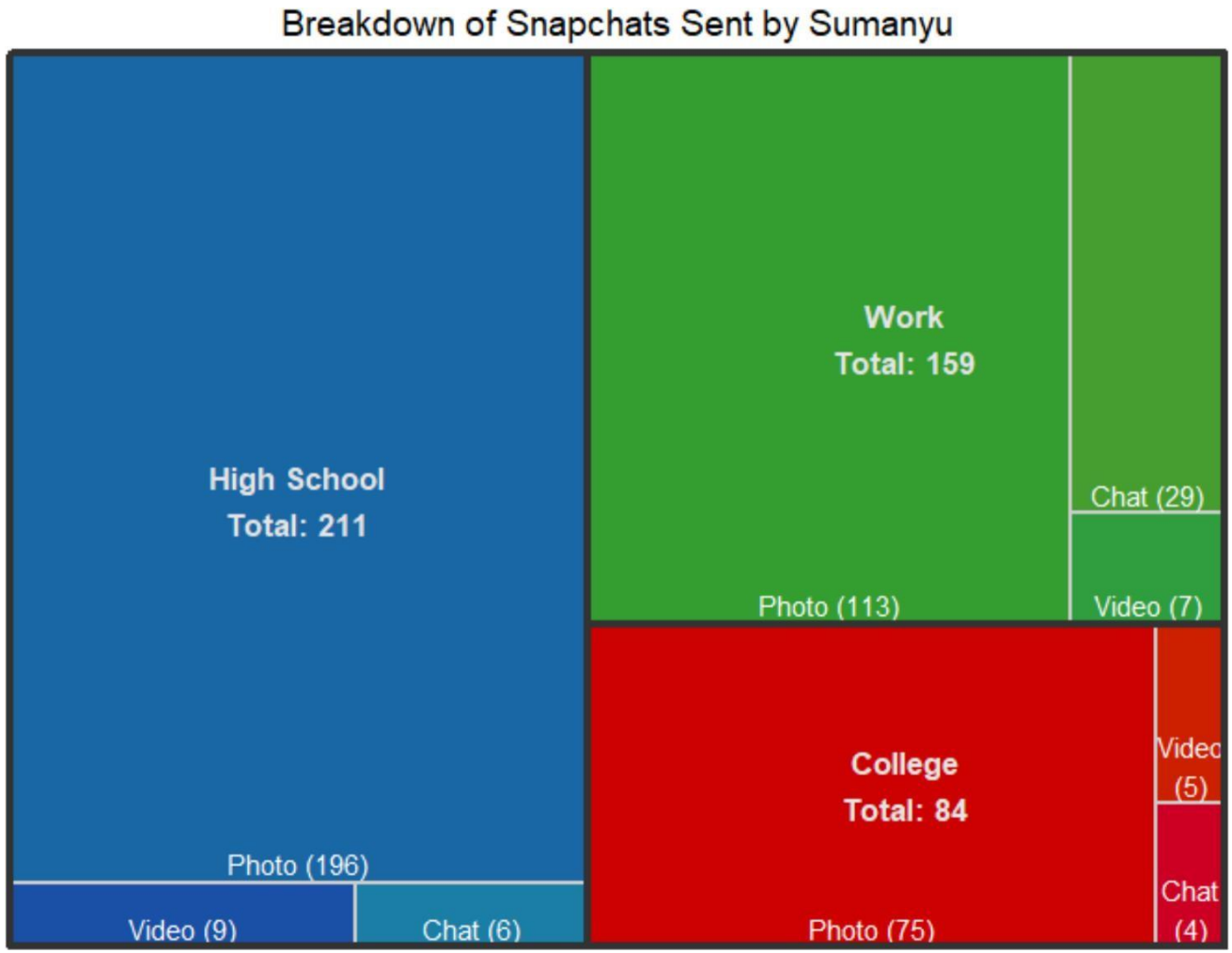
This graph is a group of box plots that indicate steps walked by day over the past three years. Various shades of blue are used to represent the years that the data has been collected. The year 2016 had the least amount of data because of one reason: I got my iPhone in November of that year. Hence, the data visualized is all from November and December. The black dots are outliers that occur for each day. For example, there was a Tuesday in 2017 that I walked a little over almost 50000 steps. Walking around the entire city of San Francisco will elevate those numbers for sure! Finally, the black bars in the middle of the box are median values.



This line plot displays a trend line for the total number of snapchats sent and received per day over the three-week data collection period. The three time periods are differentiated by color in the background and weekends are shaded in light red. A dotted line runs through the center of the plot to display the average amount of daily Snapchats. Most of the days were around the average mark, while June 4th and June 16th were outlying values. Those days were included to account for statistical inaccuracies. A noticeable increase from week one to week three is visible. This was because the majority of my friends returned home from vacation and had more time to use Snapchat.



This bar plot displays a breakdown of the Snapchats I have sent to my friends. During the data collection phase, I had a field that recorded the affiliation of the friend and myself. In other words, it represents the relationship we formed before becoming "friends" on the app. On the y-axis, the names of the people who received my snaps are present (16 in total). I keep in touch with my friends from high school the most even though they are all in college now. Finally, there is a summary table located in the middle of the graph that shows aggregate information of total Snapchats I've sent to friends.



This treemap categorizes the type of Snapchats I sent to my affiliations. The types include photo, video, and chat. The affiliations are broken down into high school, college, and work. I kept in touch with my high school friends the most and used photos to communicate the most. The reason Snapchat is popular is because of the quick sharing of photos, so it makes sense that it was the most popular type for me. Interestingly, chats were highest in the work category. I used them the most in this affiliation because my conversations with friends would require more text, which the chat feature worked the best for.