Lewis Arnsten

The main Netflix data file I was interested in was “ViewingActivity.csv”. After loading this file into a pandas dataframe I began creating my visualization of the amount of time spent on Netflix every day for a year. This visualization would allow me to test the hypothesis that days of heavy Netflix use would be grouped together in the same week. To create this visualization, I first summed the durations of all episodes watched each day for a year. I then created a 52 by 7 by 3 array (an image) where the R value of each pixel would be determined by the time spent watching Netflix on the corresponding day. I made white squares indicate very little time spent on Netflix and each darker shade of red indicate more time spent on Netflix. This yielded a satisfyingly striking visualization.

I then calculated the start and end times of every show that was watched for the year. For each week, I found how long the user spent watching Netflix with less than a five-minute break between the end of one episode and the start of the next. The combined time of these episodes watched back-to-back I called the binge time. After finding a week's worth of binge sessions I sorted and plotted the top five longest binge sessions (if there were that many) in a bar graph.

I spent a long time trying to figure out how to link these two visualizations together. Though I have almost no web development experience, I decided that a website would be the best way to connect them. Thus, I created a website that would allow users to click a link corresponding to a week in my visualization and see the week’s binge statistics.