HOW I SPEND MY TIME AT AMHERST COLLEGE – A TWO-WEEK STUDY

SAT 231: Calendar Query

DASHA ASIENGA

October 8, 2021

HOW DO I SPEND MY TIME ON A DAILY BASIS?

```
calendar_data <- "STAT231CalendarQuery.ics" %>%
  #use ical package to import into R and create a tibble object
  ical_parse_df() %>%
  as_tibble() %>%
  #create variable for duration of calendar entries, date, and activity
  mutate(
    start_datetime = with_tz(start, tzone = "America/New_York"),
    end_datetime = with_tz(end, tzone = "America/New_York"),
    duration_hours = difftime(end_datetime, start_datetime, units = "hours"),
    date = floor_date(start_datetime, unit = "day"),
    activity = tolower(summary)
  )

#compute total duration of time for each day and activity
activities <- calendar_data %>%
  group_by(date, activity) %>%
  summarize(duration_hours = sum(duration_hours)) %>%
  pivot_wider(names_from = "activity",
              values_from = "duration_hours") %>%
  #clean up variable names and remove spaces
  janitor::clean_names() %>%
  #use lubridate package to enable R to treat the dates as dates and not characters
  mutate(date = lubridate::ymd(date))
```

Questions and Variables of Interest

1. Does how much time I allocate for sleep on each day impact my productivity during the day?

Under this question, I was interested in seeing whether there is a relationship between how much sleep I get and how much time I end up allocating to specific activities in the day, particularly those that I used as a measure of productivity: `work` and `school work`. How is time for the other activities affected based on how much sleep I get?

2. How much of my time do I spend on various activities per day?

Under this question, I was interested in simply exploring how much time I'm spending on different activities in a day. I had 8 variables of interest based on the main activities in my day: `friends_family`, `school_work`, `self_time`, `sleep`, `class`, `work`, `exercise`, and `extra_curricular_activity`. I further created new variables in the wrangling process: `Week`, which tracked whether it was Week 1 or Week 2, and `Weekend`, which tracked whether the data was collected on a weekday versus on a weekend. I was interested in comparing if there was a difference in how I spend my time across those 2 facets.

3. How much time do I allocate outside of class to complete the assignments for each of my four different classes?

Finally, I was interested in exploring how I divided my time outside class to assignments for my four classes this semester. This was an extension of one of the variables of interest in question 2, `school_work`, and I tracked how my time of the day that I was spending on school work was distributed across my 4 courses: data science, computer science, mathematics, and statistics. Do I spend more time on certain classes than others or is my time evenly distributed?

The Data Collection Process

As described above, my variables of interest were defined by my questions of interest. The data collection process in itself was simple. I keyed in the start and end times of every significant activity that I did in a given day in order to ensure the most accurate collection of data, however, most start and end times were rounded to the nearest 5 minutes just because of the constraint of using Google Calendar to collect data. Miscellanous activities were often coded under `self_time` so that could be something to note and transition between activities was often not accounted for. Other than these few drawbacks, the collection of data was seamless and done many times in a single day, often after an activity or two, and the final variables ended up being a great representation of the activities I spend virtually all of my time on here on campus.

Data Visualization 1

Data visualization 1 answers my first question: Does how much time I allocate for sleep on each day impact my productivity during the day?

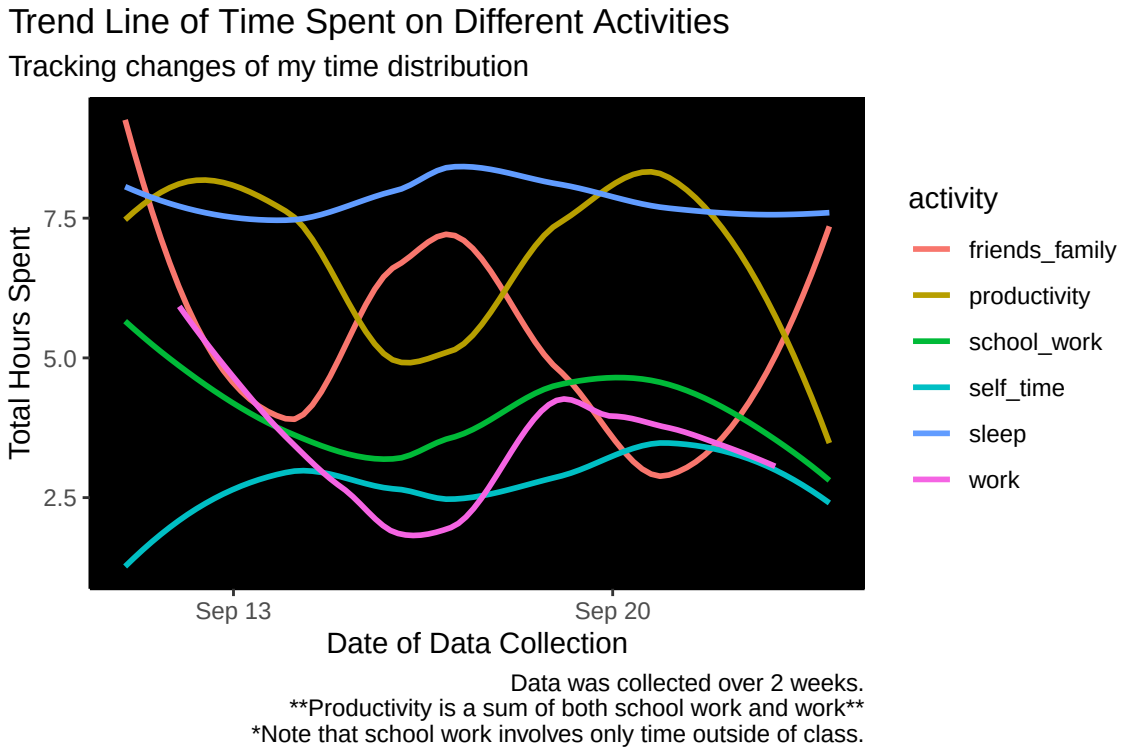
There are two related plots in this visualization. The first one shows the date on the x-axis and the time spent on some of the relevant activities, differentiated by color, on the y-axis. The second plot is an extension of the first plot to show the specific relationship between sleep and productivity, a newly defined variables as a sum of school work and work.

The first plot shows that the activity I spend the most time on is sleep and the amount of sleep I got over the two-week period didn't change so much. This makes sense because I'm very strict about my sleep schedule and this can be seen by the range of the x-axis in the second plot. The amount of time spent on the other activities displayed here, however, fluctuated a lot. This may be influenced by whether it was a weekday or a weekend, but this will be visualized in the second visualization.

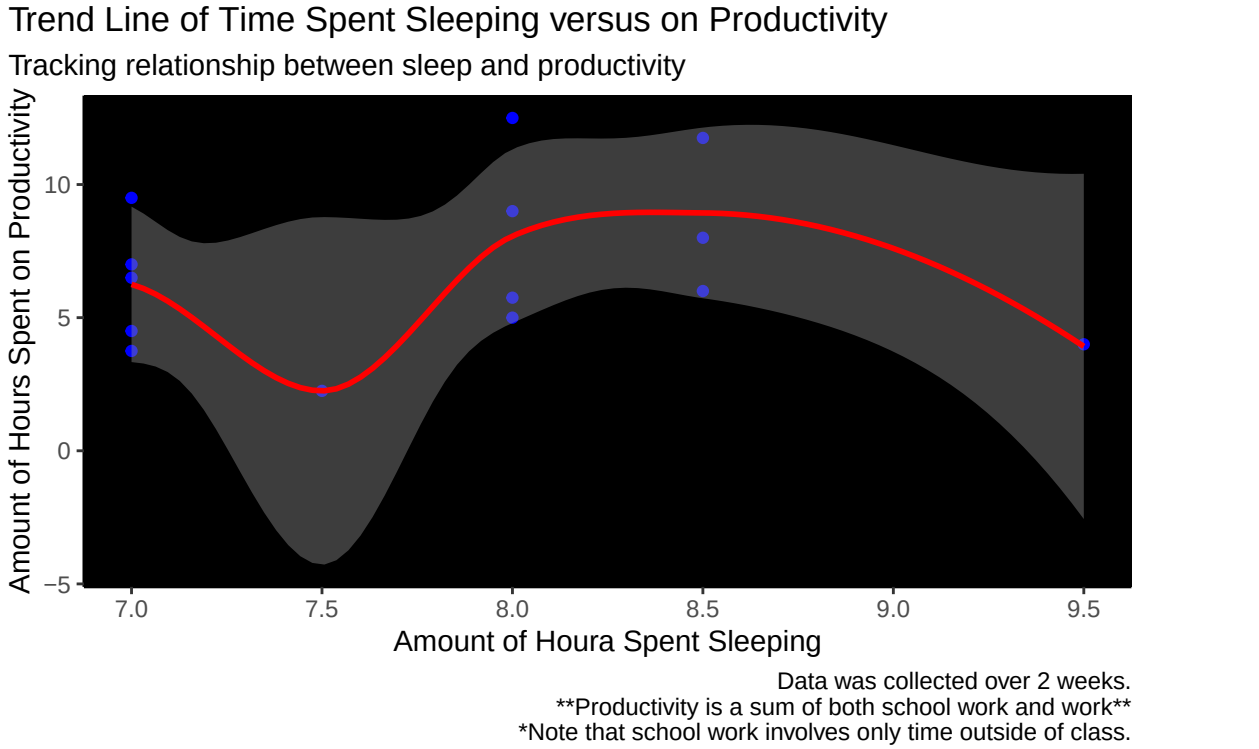
The second plot shows a notable dip in productivity at the 7.5 hours of sleep mark and a steady trend for all hours of sleep after that.

```
#wide data set to show the relationship between time spent on sleep and on different
#activities as well as create a new variable `productivity`.
activities_productivity <- activities %>%
  mutate(productivity = sum(school_work, work, na.rm = TRUE)) %>%
  select(date, friends_family, school_work, sleep, work, self_time, productivity)
```

```r
#long data set to show the relationship between time spent on sleep and productivity.
activities_productivity_long <- activities_productivity %>%
  pivot_longer(cols = -date,
               names_to = "activity",
               values_to = "duration_hours")


#Trend Line of Time Spent on Different Activities over the two-week period
sleep_activities <- ggplot(activities_productivity_long,
                           mapping = aes(x = date,
                                         y = duration_hours,
                                         color = activity)) +
                    #specifies type of line
                    geom_smooth(method = "loess", se = FALSE) +
                    #informative and stylistic features of the plot
                    labs(title = "Trend Line of Time Spent on Different Activities",
                         subtitle = "Tracking changes of my time distribution",
                         caption = "Data was collected over 2 weeks.
                         **Productivity is a sum of both school work and work**
                         *Note that school work involves only time outside of class.",
                         x = "Date of Data Collection",
                         y = "Total Hours Spent") +
                    theme(legend.position = "right",
                          plot.title.position = "plot",
                          panel.background = element_rect("black"))

sleep_activities
```



```r
#Smooth Line of Time Spent Sleeping versus Time Spent on Productive Activities
sleep_productivity <- ggplot(activities_productivity, mapping = aes(x = sleep,
                                                                    y = productivity)) +
                      geom_point(color = "blue") +
                      #specifies type of line
                      geom_smooth(method = "loess", color = "red") +
                      #informative and stylistic features of the plot
                      labs(title = "Trend Line of Time Spent Sleeping versus on Productivity",
                           subtitle = "Tracking relationship between sleep and productivity",
                           caption = "Data was collected over 2 weeks.
                           **Productivity is a sum of both school work and work**
                           *Note that school work involves only time outside of class.",
                           x = "Amount of Houra Spent Sleeping",
                           y = "Amount of Hours Spent on Productivity") +
                      theme(plot.title.position = "plot",
                            panel.background = element_rect("black"))

sleep_productivity
```



Data Visualization 2

Data visualization 2 answers my second question: How much of my time do I spend on various activities per day?

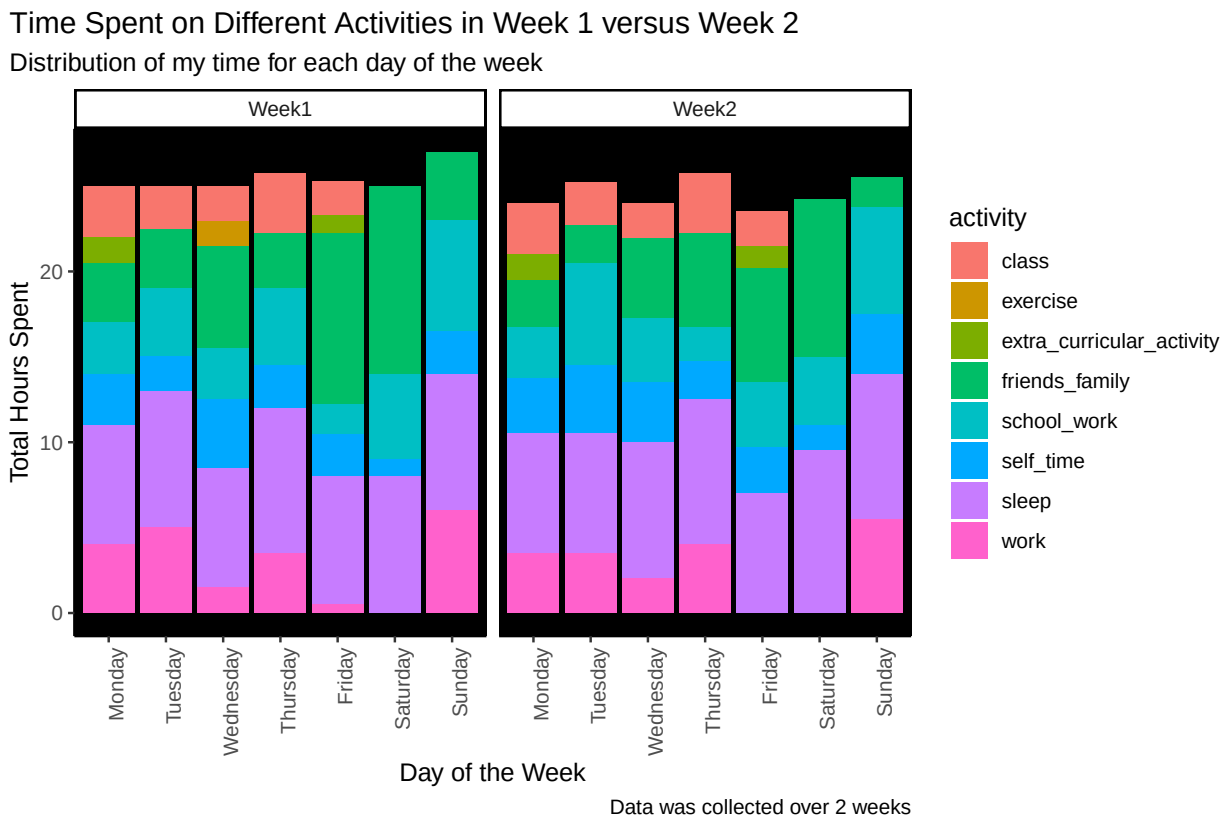
In this visualization, I faceted a stacked bar plot depicting how much time I spent on different activities in 2 ways: whether it was Week 1 or Week 2 and whether it was a weekday or weekend. These plots enhanced the information portrayed by labeling the axis based on what day of the week it was as that can carry a lot more information than arbitrary dates.

The first faceted plot shows that with the exception of not having class on weekends and a few differences, how I spent my time was largely similar across the 2 weeks and didn't differ drastically, suggesting some sort of routine in place.

The second faceted plot shows that I generally have more activities on weekdays than on weekends and the distribution of those activities is different. For example, I spend a lot of time with my friends on Saturday compared to the rest of the week.

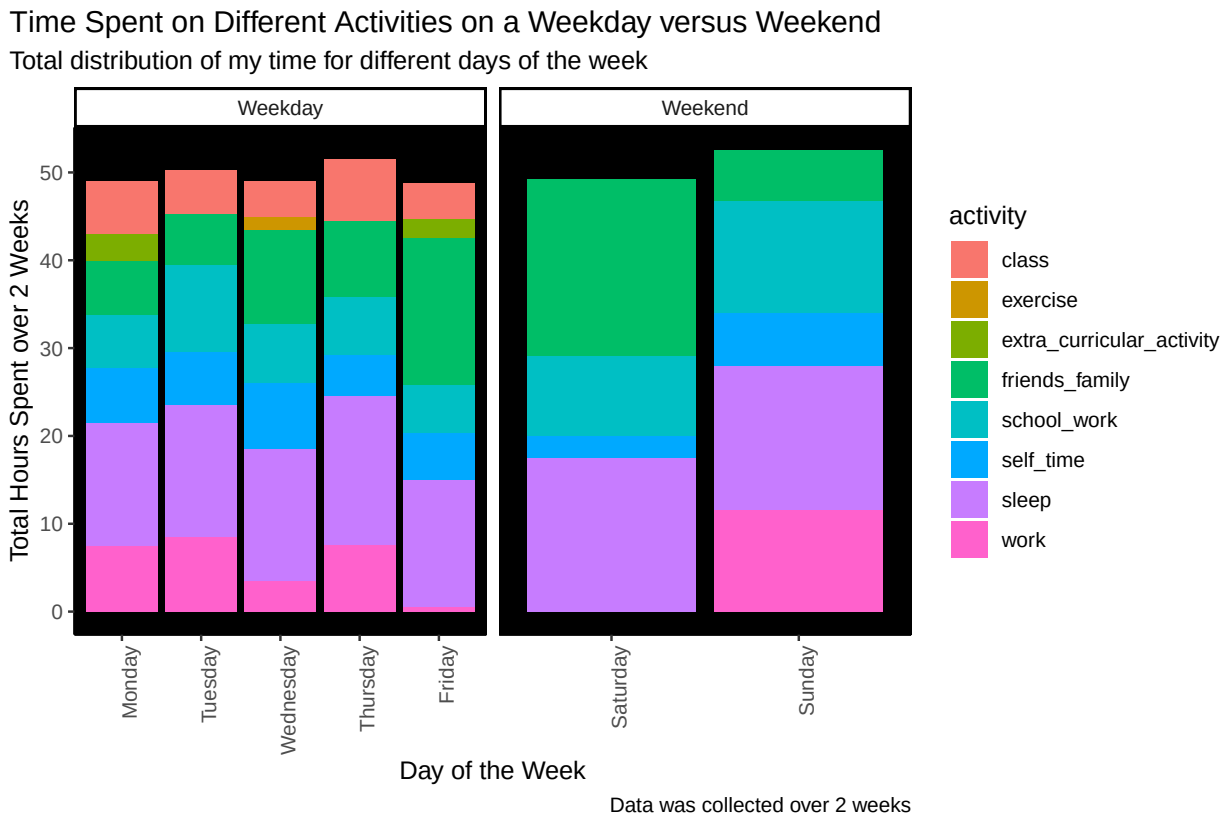
```
#clean, renamed data set
activities_days_renamed <- activities %>%
  #select activities of interest
  select(date, friends_family, school_work, self_time, sleep,
         work, exercise, class, extra_curricular_activity) %>%
  #create 3 columns: date, activity, and productivity
  pivot_longer(cols = -date,
               names_to = "activity",
               values_to = "duration_hours") %>%
  #create columns from the dates in the data set
  pivot_wider(names_from = "date",
              values_from = "duration_hours") %>%
  #rename dates to allow easier manipulation and creation of new variables
  rename("Week1.Saturday" = "2021-09-11",
         "Week1.Sunday" = "2021-09-12",
         "Week1.Monday" = "2021-09-13",
         "Week1.Tuesday" = "2021-09-14",
         "Week1.Wednesday" = "2021-09-15",
         "Week1.Thursday" = "2021-09-16",
         "Week1.Friday" = "2021-09-17",
         "Week2.Saturday" = "2021-09-18",
         "Week2.Sunday" = "2021-09-19",
         "Week2.Monday" = "2021-09-20",
         "Week2.Tuesday" = "2021-09-21",
         "Week2.Wednesday" = "2021-09-22",
         "Week2.Thursday" = "2021-09-23",
         "Week2.Friday" = "2021-09-24")  %>%
  #assign renamed dates to the 'Day` variable
  pivot_longer(cols = -activity,
               names_to = "Day",
               values_to = "Duration_hours") %>%
  #create new variables -- Day (overwritten) and Week -- separated by '.' in the names
  separate(col = Day, into = c("Week", "Day"), sep="\\.") %>%
  #declare the `Day` variable as categorical and assign values
  mutate(Day = factor(Day, levels = c("Monday", "Tuesday", "Wednesday",
                                      "Thursday", "Friday", "Saturday",
                                      "Sunday")),
         #manually define the `Weekend` variable as weekday or weekend
         Weekend = fct_collapse(Day,
                                Weekday = c("Monday", "Tuesday", "Wednesday",
                                            "Thursday", "Friday"),
                                Weekend = c("Saturday", "Sunday")))


#Stacked Bar Plot Faceted by Week
week <- ggplot(data = activities_days_renamed, mapping = aes(x = Day,
                                                             y = Duration_hours,
```

```
                                                    fill = activity)) +
        geom_col() +
        #orient the values on the x-axis vertically for better readability
        theme(axis.text.x = element_text(angle = 90, hjust =1)) +
        #compare the stacked bar plots by Week 1 versus Week 2
        facet_wrap(~Week) +
        #informative and stylistic features
        labs(title = "Time Spent on Different Activities in Week 1 versus Week 2",
             subtitle = "Distribution of my time for each day of the week",
             caption = "Data was collected over 2 weeks",
             x = "Day of the Week",
             y = "Total Hours Spent") +
        theme(legend.position = "right",
              plot.title.position = "plot",
              panel.background = element_rect("black"))
week
```




```
#Stacked Bar Plot Faceted by Weekday versus Weekend
weekend <- ggplot(data = activities_days_renamed, mapping = aes(x = Day,
                                                                y = Duration_hours,
                                                                fill = activity)) +
           geom_col() +
           #orient the values on the x-axis vertically for better readability
           theme(axis.text.x = element_text(angle = 90, hjust =1)) +
           #compare the stacked bar plots by whether it was a weekday or weekend
```

```
            facet_wrap(~Weekend, scales = "free_x") +
            #informative and stylistic features
            labs(title = "Time Spent on Different Activities on a Weekday versus Weekend ",
                 subtitle = "Total distribution of my time for different days of the week",
                 caption = "Data was collected over 2 weeks",
                 x = "Day of the Week",
                 y = "Total Hours Spent over 2 Weeks") +
            theme(legend.position = "right",
                  plot.title.position = "plot",
                  panel.background = element_rect("black"))


weekend
```



Table

Finally, the table that I created answers my third question: How much time do I allocate outside of class to complete the assignments for each of my four different classes?

This table includes summary statistics of how much time outside of class I allocated to each of my 4 courses over the two-week period. I collected the data at a time when I was not so busy in some of my classes so this summary may not be generalized to how much time I actually allocate to my different classes throughout the semester at Amherst college. Nevertheless, this table depicts how much time I was spending on my courses over the 2 weeks and includes the total time spent, the average time spent per week (obtained by dividing

the total time by 2), the average and median times spent on the course per data entry, and the standard deviation. It provides a lot of useful information on what classes might be taking most and least of my time.

```
#school work data set for table
activities_schoolwork <- activities %>%
  rename(computer_science = computer_science_school_work,
         mathematics = calculus_school_work,
         data_science = data_science_school_work,
         statistics = statistics_school_work) %>%
  #select variables of interest -- classes enrolled in
  select(computer_science, data_science, mathematics, statistics) %>%
  rename("Computer Science" = computer_science,
         "Data Science" = data_science,
         "Statistics" = statistics,
         "Mathematics" = mathematics) %>%
  #place the different courses under a variable `homework` and document duration
  pivot_longer(cols = - date,
               names_to = "homework",
               values_to = "duration") %>%
  #get rid of NA values
  drop_na() %>%
  #perform operations grouped by which class it was
  group_by(homework) %>%
  #obtain summary statistics
  summarise(
    Entries = n(),
    "Total" = sum(duration),
    "Average per Week" = sum(duration)/2,
    "Average per Entry" = mean(duration),
    "Median per Entry" = median(duration),
    "Standard Deviation" = sd(duration)
  ) %>%
  #arrange in descending order by which classes most time was spent on
  arrange(desc(Total)) %>%
  rename("Class Assignment" = homework) %>%
  #format it to be pdf-presentable
  kable(booktabs = TRUE, digits = 1, caption = "") %>%
  #add header that spans above the 7 columns in the table
  add_header_above(c("Time Outside of Class Spent on Different Courses" = 7),
                   color = "white", background = "black", bold = TRUE) %>%
  #add informative footnote
  add_footnote("The data was collected over 2 weeks") %>%
  #scale down the table to fit the width of the page, hold the table's position below
  #the R chunk, and stripe the table as a stylistic feature
  kable_styling(latex_options=c("scale_down", "hold_position", "striped"))


activities_schoolwork
```

Conclusion

This calendar query project led to a lot of insight on how I typically spend my time at Amherst college and I was able to draw a lot of conclusions.

Table 1:

Time Outside of Class Spent on Different Courses						
Class Assignment	Entries	Total	Average per Week	Average per Entry	Median per Entry	Standard Deviation
Data Science	8	23.2 hours	11.6 hours	2.9 hours	2.8 hours	0.8
Mathematics	10	15.0 hours	7.5 hours	1.5 hours	1.5 hours	0.4
Statistics	7	12.8 hours	6.4 hours	1.8 hours	1.0 hours	1.3
Computer Science	7	5.0 hours	2.5 hours	0.7 hours	0.5 hours	0.5

[a] The data was collected over 2 weeks

My first question was primarily curious about the relationship between how much time I spend on sleep and how much time I spend on other activities. Since I only collected data over 14 days and because there isn't much variability in how much time I spend sleeping during the week, it's difficult to see if there is any relationship between how much time I sleep and how much time I spend on other activities, including how it is directly related to how productive I am during the day. Nevertheless, it's interesting to see the trend lines of how much time I was allocating to sleep and other activities over the entire 2 weeks and how that stayed consistent or changed on certain dates. Later, I was able to see the influence of it being a weekday versus a weekend in my second visualization. From this visualization, I was able to see that sleep was the activity I spent most time on in a given day.

My second question was primarily curious about the general distribution of my time over different activities by week and by whether it was a weekday or weekend. This visualization showed that a lot of the distribution of my time was constant over the two weeks, which was an indicator of the routine that I have when I'm at school. Interestingly, time in class takes up a small portion of my day and a huge portion of my time is spent on doing work and school work during the weekday. It also highlighted how the distribution changes depending on whether it is a weekday or a weekend – for example, I tended to sleep more and spend more time with friends on the weekend and I tended to have more work during the weekdays.

Lastly, I was curious about how I was spending time outside of class between my 4 courses. The summary table is arranged in descending order by which classes I spent the most time on to which classes I spent the least time on. It displays that I spend the most time in my "school work" activity time on Data Science, followed by Mathematics, Statistics, and Computer Science respectively. It highlighted that I should consider spending more time on my Computer Science class than I currently do.

Overall, it was really interesting and insightful to see how I spend my time at Amherst college, and perhaps which areas I should allocate more time for.

REFLECTION FROM THE CALENDAR QUERY PROJECT

As described in the report above, collecting data myself about how I spend my day, wrangling that data, and finally constructing beautiful and neat visualizations led to a lot of insightful information. The process of data collection itself was quite simple with the exception of a few difficulties in gathering accurate data. I often found that there were times when I wasn't doing anything substantive, which reflects why my stacked bar plots are not all the same height, even though they should be because there are only 24 hours in a given day. In moments where I was doing a lot of 5 - 10 minute miscellaneous activities over an extended period of time, I would key that block in under `self_time`. However, short miscellaneous activities in between activities of interest were either unaccounted for or miss-attributed for under another activity. Lastly, I multitask a lot – for example, I study and do homework with friends at times so that time of my day is accounted for under 2 variables: `friends_family` and `school_work`. Failing to include one of the variables would have ended up being a misrepresentation of how much time I was actually allocating to different activities, so I had to compromise and settle for having some days that are longer than 24 hours from a data analysis sense.

When it came to wrangling the data, my main challenge was coding new variables `Week` for whether it was Week 1 or Week 2, and `Weekend` for whether it was a weekday or a weekend. This process was manageable, but to do so, I had to rename all the dates according to what day of the week that data corresponds to and which week that date belonged to so that I could split the variables by name. I only had 14 data points in this data set and renaming all 14 dates was a time-consuming task. I wonder if there is a simpler way to accomplish this because that would have been a major hurdle had I had more data points to work with than I did, let alone a big data set.

These reflections have a lot of implications for future data collection and analysis projects. For one, it highlights how, more often than not, our data is related and not independent. In my example, time spent studying was also time spent with friends, and this can have implications when we're looking at large data sets that also have the potential to make a huge difference in today's world. It also highlights how understanding computational analysis is very crucial and having more research done into data analysis and ways to make data wrangling easier can save a lot of the time and energy that can then go into understanding what the data is telling us.

In order to fully answer my questions of interest in this Calendar Query Project, particularly question 1 and question 3, I would need at least a whole semester's worth of data in order to have something substantial to analyze, and more data than that for other bigger questions of interest. The hurdle with this is that data collection can be a hard and time-consuming process. For this project, I had to constantly keep track of my day to day, hour to hour. It was only two weeks long, but at times it would feel like a daunting task and I was glad to finally be done with keeping track of every minute of my day and get my hands on the analysis part of the project after the data collection period. Doing this for a whole semester seems incredibly difficult, and it can be harder to collect data depending on our questions of interest. For example, recording precise lab measurements for an entire year can become difficult, but it is necessary so as to have the most accurate analysis we can afford, and use that to inform our questions.

I chose my variables in the most general setting but some of this information can become private or I could have just as easily chosen a variable that could have reflected something more personal. Even in the variables I chose, knowing how I spend my time and what activities I do at a certain time can prove evil in the wrong hands and could be used as a mechanism to stalk or perform other harmful infringements of my rights. Simply put, any data in the wrong hands is dangerous. Therefore, as someone who provides data to Facebook, Google, etc., I expect that my information be treated with respect and that my privacy be protected. In hand with that, as someone who analyzes other people's data, I have the ethical responsibility to treat that data with respect. This involves not only respecting the privacy of the people who provide the data in the data sets that I use, but also using their data to facilitate good, not evil.

In the end, I really enjoyed this project and I'm excited to undertake other data science projects in the near future. I was surprised by what I was able to produce in the end as well as how much I learned in the process and I am definitely one step closer towards being a data scientist.