Calender Assignment

Lauren Pelosi

Friday, September 25

```
# Importing the data and preliminary wrangling:
calendar <- calendar %>%
  mutate(start_datetime = with_tz(start, tzone = "America/New_York")
    , end_datetime = with_tz(end, tzone = "America/New_York")
    , date = floor_date(start_datetime, unit = "day"))

data <- calendar %>%
  filter(start>="2020-09-06 10:00:00") %>%
  mutate(subject = case_when(str_detect(summary, "math") ~ "math",
                             (str_detect(summary, "health") ~ "health"),
                              (str_detect(summary, "law")~"law"),
                             (str_detect(summary, "Law") ~ "law"),
                             (str_detect(summary, "data science") ~ "data science"))) %>%
  mutate(type = case_when(str_detect(summary, "homework") ~ "homework",
                                 (str_detect(summary, "recruiting") ~ "recruiting"),
                                 (str_detect(summary, "Jennifer") ~ "recruiting"),
                                 (str_detect(summary, "Reyes") ~ "advising"),
                                 (str_detect(summary, "WBB") ~ "basketball"),
                                 (str_detect(summary, "internship") ~ "internship"),
                                 (str_detect(summary, "ADP") ~ "research"),
                                 (str_detect(summary, "class") ~ "class"),
                                 (str_detect(summary, "exercise") ~ "exercise"),
                                 (str_detect(summary, "OH") ~ "office hours")))

View(data)
```

Questions of Interest and Data Collection

I am interested in three main questions about how I spend my time:

1. Am I spending roughly equal amounts of time on each of my four courses?
2. How is my total screen time divided between school-related activities and non-school related activities?
3. Do I allow my work to affect how much exercise I'm getting in a day?

I would like to be spending roughly equal time on each course, decrease my total screen time, and maintain consistent exercise despite other demands on my time. Better understanding the answers to these questions will help me adjust my behavior for more desirable outcomes.

To answer these questions, I tracked my activities in Google Calendar for two weeks. I recorded for how long I was in class, doing homework, or working in office hours each day, and the subject of that class. I also

recorded when I was in a general academic advising meeting. I've operationalized all these hours as "school" hours. I also recorded other screen-time activities, which include basketball team meetings, working with my research team in the LJST department, working on my internship (which happened to be light in these two weeks), and recruiting for consulting jobs. Recruiting was a dominant demand on my time outside of school and large contributor of screen time, since it involves studying online resources, making zoom calls, and attending virtual events. I call all these activities "non-school" activities. Thirdly, I recorded how much exercise I got each day - the only non-school and non-screen activity I look at.

Am I spending roughly equal amounts of time on each of my four courses?

Here, I use a stacked bar chart to visually evaluate if I spend roughly equal amounts of time on each of my four subjects. The full height of each bar represents the total number of hours I spent on course-related activities in a day, (including, class, homework, and office hours). Each bar further shows the breakdown of this total number of hours into its component parts, or the school subjects I was working on.

```
#Wrangling for the stacked bar chart
#First creating a time variable and choosing relevant variables for this visualization
data <- data %>%
  mutate(time = as.numeric(end-start)/60) %>%
  select(summary, date, subject, type, time, start, end)

class_per_day <- data %>%
  filter(type == "class" | type == "homework" | type == "office hours") %>%
  group_by(date, subject) %>%
  summarize(time = sum(time))
```

```
## `summarise()` regrouping output by 'date' (override with `.groups` argument)
```

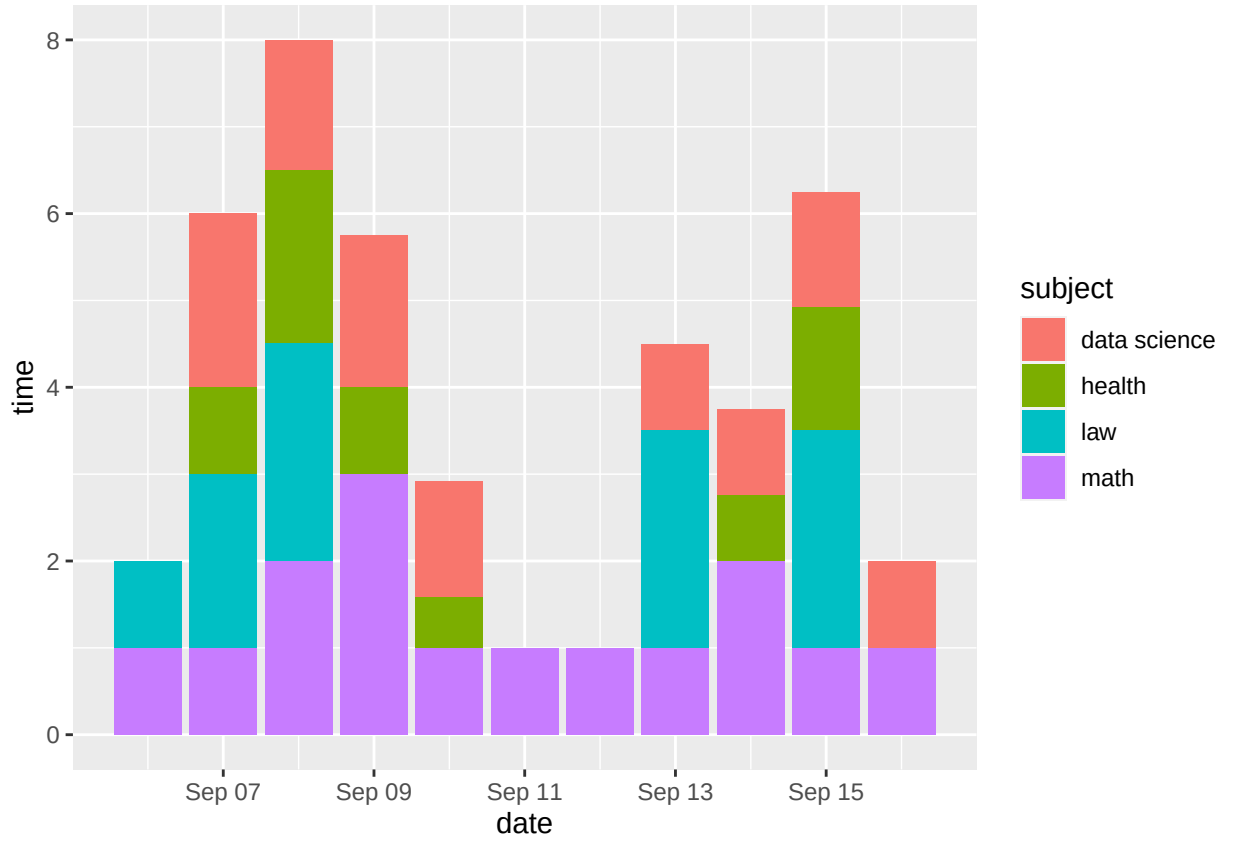
```
#Table to go with plot:
table_Q1 <- class_per_day %>%
  pivot_wider(id_cols = date, names_from = subject, values_from = time) %>%
  filter(date != "2020-06-22") %>%
  select(date, health, math, law, "data science")

table_Q1[is.na(table_Q1)] <- 0

print(table_Q1)
```

```
## # A tibble: 11 x 5
## # Groups:   date [11]
##    date                health  math   law `data science`
##    <dttm>               <dbl> <dbl> <dbl>          <dbl>
##  1 2020-09-06 00:00:00  0         1   1              0
##  2 2020-09-07 00:00:00  1         1   2              2
##  3 2020-09-08 00:00:00  2         2   2.5            1.5
##  4 2020-09-09 00:00:00  1         3   0              1.75
##  5 2020-09-10 00:00:00  0.583     1   0              1.33
##  6 2020-09-11 00:00:00  0         1   0              0
##  7 2020-09-12 00:00:00  0         1   0              0
##  8 2020-09-13 00:00:00  0         1   2.5            1
##  9 2020-09-14 00:00:00  0.75      2   0              1
## 10 2020-09-15 00:00:00  1.42      1   2.5            1.33
## 11 2020-09-16 00:00:00  0         1   0              1
```

```r
#Stacked bar chart
ggplot(class_per_day, aes(fill = subject, y = time, x = date)) +
    geom_bar(position="stack", stat="identity")
```



There are a couple significant takeaways from the stacked bar chart. Responsive to my research question, the chart tells me that my time is not split evenly between my courses. I spend the most time working on my math course - at least an hour on the course each day - which is about double the time I spend working on my Health Economics course, which I call "health." This could be in part because I was preparing for a math exam over this time period.

Another interesting finding, which I wasn't looking for directly, is that the total time I spend on school work varies greatly from day to day, with a maximum of 8 hours and a minimum of 1. I think this is because, some days, recruiting events demand a lot of my time. In day to day life, I hadn't noticed this disparity because both school work and recruiting events involve screen time, and feel somewhat the same. The next visualization will break down school and non-school screen time.

How is my total screen time divided between school-related activities and non-school related activities?

To answer my next question, I use a stacked area plot to show how my screen time has accumulated over time, as well as visually demonstrate the division of that screen time between school and non-school activities.

```r
data2 <- data %>%
  select(date, type, time) %>%
  mutate(category = case_when(str_detect(type, "class") ~ "school",
```

```
                              (str_detect(type, "homework") ~ "school"),
                              (str_detect(type, "office hours") ~ "school"),
                               (str_detect(type, "advising") ~ "school"),
                              (str_detect(type, "recruiting") ~ "non_school"),
                              (str_detect(type, "internship") ~ "non_school"),
                              (str_detect(type, "research") ~ "non_school"),
                             (str_detect(type, "basketball") ~ "non_school"))) %>%
  filter(category != "NA") %>%
  filter(date <= "2020-09-16")

data3 <- data2 %>%
  group_by(date, category) %>%
  summarize(time = sum(time))
```

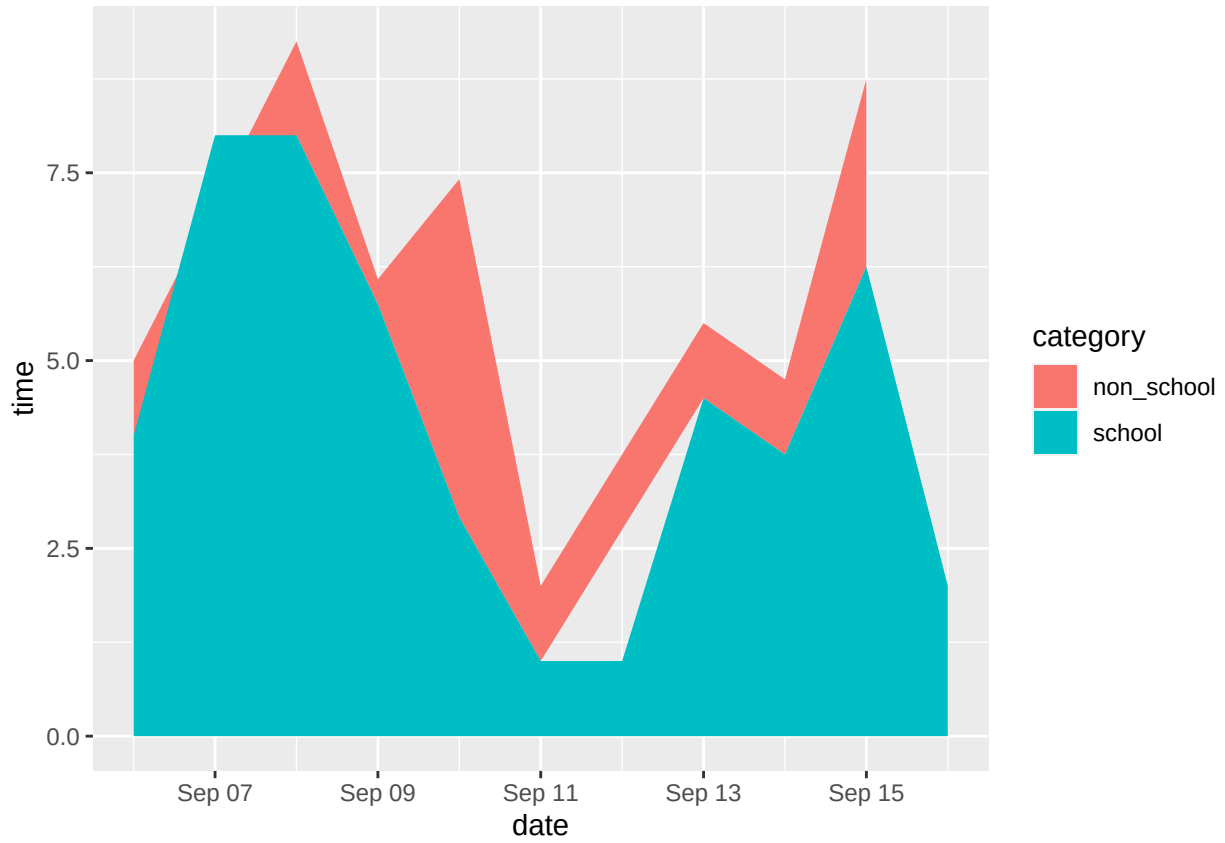
```
## `summarise()` regrouping output by 'date' (override with `.groups` argument)
```

```
#Table to go with plot
table_Q2 <- data3 %>%
  pivot_wider(id_cols = date, names_from = category, values_from = time)
table_Q2[is.na(table_Q2)] <- 0
table_Q2 <- table_Q2 %>%
  mutate(screen_time = (non_school + school))

print(table_Q2)
```

```
## # A tibble: 11 x 4
## # Groups:   date [11]
##    date                non_school school screen_time
##    <dttm>                   <dbl>  <dbl>       <dbl>
##  1 2020-09-06 00:00:00      1       4           5
##  2 2020-09-07 00:00:00      0       8           8
##  3 2020-09-08 00:00:00      1.25    8           9.25
##  4 2020-09-09 00:00:00      0.333   5.75        6.08
##  5 2020-09-10 00:00:00      4.5     2.92        7.42
##  6 2020-09-11 00:00:00      1       1           2
##  7 2020-09-12 00:00:00      0       1           1
##  8 2020-09-13 00:00:00      1       4.5         5.5
##  9 2020-09-14 00:00:00      1       3.75        4.75
## 10 2020-09-15 00:00:00      2.5     6.25        8.75
## 11 2020-09-16 00:00:00      0       2           2
```

```
#Creating  the plot
ggplot(data3, aes(x = date, y = time, fill = category)) +
    geom_area()
```



From the stacked area plot, I again see two significant takeaways. The first is that, in response to my research question, my screen time is predominately dedicated to schoolwork. When schoolwork is reduced, non-schoolwork time increases to maintain above 5 hours of screen time most days. Exceptions happened on Sept 11 and Sept 12, for which total screen time was very low (on Sept 12, I had 0 hours of non-school screen time).

Another takeaway is that my hypothesis to explain fluctuations in time spent on schoolwork (from my first research question) is not fully correct. While sometimes low schoolwork hours correlated with high non-schoolwork hours, low time spent on schoolwork cannot always be explained by high time spent on non-schoolwork.

Next I will look at the last activity I recorded, exercise, and how time spent exercising correlates with screen time.

Do I allow my work to affect how much exercise I'm getting in a day?

```
tot_col <- data %>%
  filter(type != "exercise") %>%
  select(date, type, time) %>%
  group_by(date) %>%
  summarize(total = sum(time))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
exc_col <- data %>%
  select(date, type, time) %>%
  group_by(date) %>%
  filter(type == "exercise") %>%
  summarize(exercise = sum(time))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
sixth <- data.frame("2020-09-06", "5.000000", "0")
names(sixth) <- c("date", "total", "exercise")
seventh <- data.frame("2020-09-07", "8.000000", "0")
names(seventh) <- c("date", "total", "exercise")
thirteenth <- data.frame("2020-09-13", "5.500000", "0")
names(thirteenth) <- c("date", "total", "exercise")

table <- inner_join(tot_col, exc_col)
```

```
## Joining, by = "date"
```

```r
table <- rbind(table, sixth, seventh, thirteenth)

table$row_num <- seq.int(nrow(table))
table_Q3 <- table %>%
  select(date, total, exercise) %>%
  rename(non_exercise = total) %>%
  arrange(exercise) %>%
  mutate(non_exc_type = case_when
    (non_exercise > 8 ~ "high",
    (non_exercise > 4 ~ "medium"),
    (non_exercise > 0 ~ "low"))) %>%
  mutate(exc_type = case_when
    (exercise == "0" ~ "none",
    (exercise == "0.5" ~ "low"),
    (exercise == "0.75" ~ "medium"),
    (exercise == "1") ~ "high"))

print(table_Q3)
```

```
## # A tibble: 11 x 5
##    date                non_exercise     exercise non_exc_type exc_type
##    <dttm>              <chr>            <chr>    <chr>        <chr>
##  1 2020-09-06 00:00:00 5.000000         0        medium       none
##  2 2020-09-07 00:00:00 8.000000         0        high         none
##  3 2020-09-13 00:00:00 5.500000         0        medium       none
##  4 2020-09-14 00:00:00 4.75             0.5      medium       low
##  5 2020-09-15 00:00:00 8.75             0.5      high         low
##  6 2020-09-16 00:00:00 2                0.5      low          low
##  7 2020-09-09 00:00:00 6.08333333333333 0.75     medium       medium
##  8 2020-09-10 00:00:00 7.41666666666667 0.75     medium       medium
##  9 2020-09-11 00:00:00 2                0.75     low          medium
## 10 2020-09-08 00:00:00 9.25             1        high         high
## 11 2020-09-12 00:00:00 1                1        low          high
```

The table tells me that exercise is likely uncorrelated with work done that day. My hypothesis that a high amount of time spent on other activates might detract from exercise does not seem to be supported; “high” non_exercise is paired with both “low” and “none” exercise on different days. Similarly, “low” non_exercise is paired with “low”, “medium”, and “high” exercise on different days. An alternate hypothesis could be that there are some days of high motivation and low motivation: “high” non_exercise is paired with “high” exercise on one day, while “low” non_exercise" is paired with “low” exercise on another day.

#Findings

After investigating my three research questions, I’ve found three key takeaways:

1. Of the time I spend on coursework, I spend a disproportionately large amount on my math course.
2. While my screen time is very high overall, it is more often dedicated to my courses rather than to other activities, like recruiting, my internship, and my research project.
3. While I know my high screen time overall has caused the amount of time I spend exercising each day to be lower than at other times in my life, variations in screen time from day to day to not seem to dictate how much exercise I get in a particular day.

I can now apply this information. I can focus on ways to be more efficient in my math work in order to cut down on my total screen time, since academic work is the biggest driver of my screen time and math is the biggest driver of the time I spend on coursework.

#Reflections

One main area of difficulty I had in the data collection process was identifying exactly how long each activity lasted. For example, I might have paused activity 1 to briefly do activity 2. I tried to correct for this by recording the dominant activity during the time, activity 1, and adding a small amount of time on the next occasion I worked on activity 2 for a significant period. While this method of data collection captured major trends, it did cause inaccuracies in the data. In future projects, I could practice better diligence to stay occupied with just one activity at a time. This will improve both my focus and my data collection. However, I think such minor inaccuracies will come up in any project that requires logging my activities, and I should consider this when giving weight to the data and the answers it provides me, in this project and in projects like this.

To more fully answer the questions I have posed, I think I would need to record data for at least an entire academic year, though I could break when school breaks. To accurately asses whether I let other activities interfere with time I set aside to exercise, I would need to look at periods of time in which I’m engaged in different types of activities. Two semesters, involving two sets of four courses, would help me better understand if my time spent exercising is resistant to all types of time demands or just some (STEM vs. humanities courses, courses I find challenging vs. courses I find less challenging, etc.). This would improve the quality of the data I use to answer my other two questions as well: I could consider if I evenly divide my time between another set of four courses, and evaluate how different courses demand screen time. The diligence to collect this data would be challenging, and a significant bias in the. data would arise if I were off-campus in one semester (as I am now), and on-campus in another.

I know that this data is not my own now that I have given it to Google; it will be used to get targeted ads in front of me, and likely track my behavior in ways of which I am not aware. If I were to do a similar project with others’ data, it would feel best to me if they understood fully what I was tracking and why I was tracking it, and they had given their informed consent. In that project, I would have the responsibility to accurately represent their data and draw conclusions that I know are not misleading. While I believe I’ve presented my data accurately here, I would feel the stakes were higher if someone else were to be making choices based on conclusions I’ve presented to them.