

Exploring the Interplay of Mood, Study Habits, and Productivity

STAT 231: Calendar Query

DAIZY BULUMA

2023-10-11

Introduction

In today's fast paced world, understanding the connection between our emotions, study habits and productivity levels is more crucial than ever. Thus, for my calendar query project I decided to take a closer look at the interplay between mood, study habits and productivity — within the context of my own experiences. This project seeks to answer two questions. The first is to investigate whether there is an association between mood and study habits. The second question explores the influence of various study environments on my overall productivity. Through this project, I hope to uncover hidden patterns that shape my daily routines and empower myself to make informed decisions about how I manage my time and optimize efficiency.

Methods

Data collection

The project entailed both quantitative and categorical variables. For the quantitative variables, I intended to keep track of time (in hours) spent doing personal studies everyday and mood ratings at different times of the day (on a scale of 1-5). Whereas, for the categorical variables I recorded my study locations(Room, Science Center, Frost), time of the day (7.00 pm, 12.00pm) and the day of the week.

To keep track of my mood, I utilized reminders on my phone to prompt me to record my mood every 4 hours. I logged these mood ratings on my google calendar at the corresponding time, ensuring a comprehensive timeline of my daily emotional fluctuations. I created designated time blocks on my google calendar for my personal study sessions and during each study session I initiated a timer to precisely measure the duration of study. Subsequently, I logged the time spent actually studying within the calendar entry. I made sure to record my study location as a description every time I logged the study times. This would help me to track my productivity across the different locations. The day of the week and times of the day were automatically recorded by the google calendar.

Data wrangling

To convert my data into the format needed for my visualizations, I first calculated the duration of each activity on the calendar and stored it in a new variable called `duration_hours`. I then selected the key variables I would need (`activity`, `description`, `duration_hours`, `weekday_label`) and stored them in a new dataset, keeping the original dataset intact. I filtered the new dataset based on activity for each visualization - for example, for the bar chart showing study hours, I filtered to only include activities labeled as “study”. Since mood ratings were stored as text descriptions, I converted them to integers for analysis. For both mood ratings and study hours, I averaged the values across the two week period for each weekday, so Monday values in week 1 were summed with Monday values in week 2 and averaged. I also averaged the mood ratings at different intervals to calculate a daily average mood rating. I grouped the dataset by relevant variables for two of the visualizations. Finally, I collapsed the dataset to create a summary table.

```

suppressWarnings({
# Data import (requires **ical** package)
cal_import <- ical_parse_df("data/cq_calendar.ics")

# Data wrangling
mycal <-
  cal_import %>%
  # Google Calendar event names are in a variable called "summary";
  # "activity" is a more relevant/informative variable name.
  rename(activity = summary) %>%
  mutate(
    # Specify time zone (defaults to UTC otherwise)
    across(c(start, end),
           .fns = with_tz,
           tzzone = "America/New_York"),
    # Compute duration of each activity in hours
    duration_hours = interval(start, end) / hours(1),
    # Examples of getting components of dates/times
    # Note:
    # i. these could be based on either start datetime or end datetime
    # ii. you do NOT need all of these!! so only use what you need
    date = date(start),
    year = year(start),
    month_number = month(start),
    month_label = month(start,
                        label = TRUE,
                        abbr = FALSE),
    weekday_number = wday(start),
    weekday_label = wday(start,
                        label = TRUE,
                        abbr = FALSE),
    hour = hour(start),
    time = hour(start) + minute(start)/60,
    # Convert text to lowercase and remove repeated or leading/trailing
    # spaces to help clean up inconsistent formatting.
    across(c(activity, description),

```

```

        .fns = str_to_lower),
    across(c(activity, description),
        .fns = str_squish)
) %>%
# The first Google Calendar entry is always an empty 1969 event
filter(year != 1969)})

```

Results

Part 1 : Study Hours and Mood

```

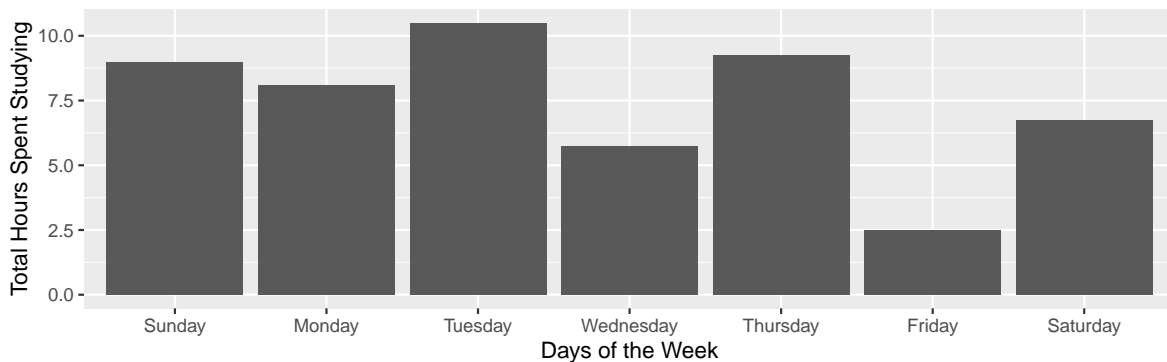
# Use `select` function to keep the variables/columns needed
# Use the `filter` function to only keep observations where `activity == "actual studying"
# Store the resulting data frame as a new data frame called `mycal1`
mycal1 <- mycal %>%
  select(activity,description,duration_hours,weekday_label) %>%
  filter(activity == "actual studying" )

# Plot the bar chart
p <- ggplot(
  data = mycal1,
  aes(x = weekday_label, y = duration_hours )) +
  geom_col()+
  labs( title = "Total Hours Spent Studying On Each Day Of The Week",
        subtitle = "September 17, 2023 to September 30, 2023",
        y = "Total Hours Spent Studying", x = "Days of the Week" )
p

```

Total Hours Spent Studying On Each Day Of The Week

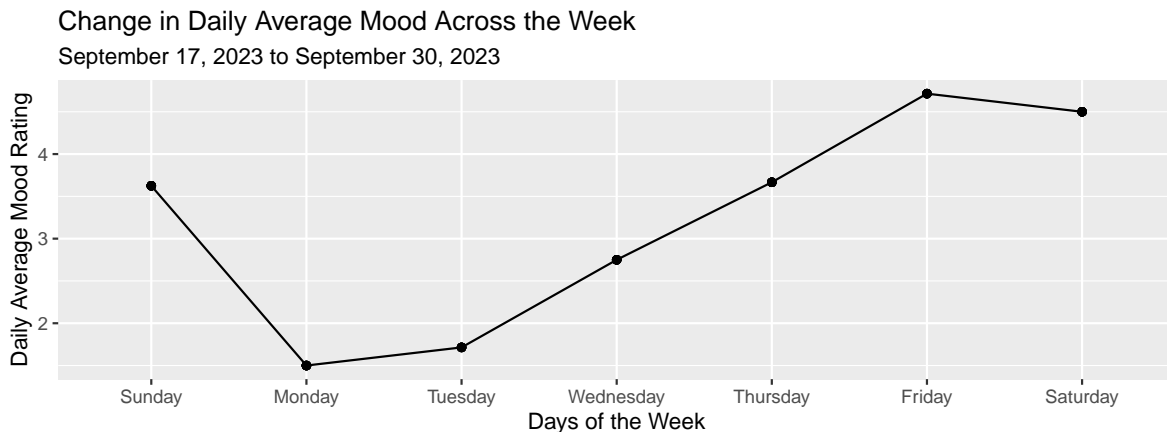
September 17, 2023 to September 30, 2023



```
# Use `select` function to keep the variables/columns needed
# Use the `filter` function to only keep observations where `activity == "mood"`
# Group by weekday_label in order to find the average mood of each day of the week
# and store the result in a new variable `mean_mood`
# Store the resulting data frame as a new data frame called `mycal2`
mycal2 <- mycal %>%
  select(activity,description,weekday_label) %>%
  filter(activity == "mood" ) %>%
# convert from char to int
  mutate(mood_int = as.integer(description)) %>%
  group_by(weekday_label) %>%
  mutate(mean_mood = mean(mood_int, na.rm = TRUE)
         ,GROUP = 1)

# Assign `GROUP` to group in order for `geom_line` function to work
# Plot the line graph
p2 <- ggplot(
  data = mycal2,
  aes(weekday_label, mean_mood, group = GROUP)) +
  geom_line()+
  geom_point()+
  labs( title = "Change in Daily Average Mood Across the Week",
        subtitle = "September 17, 2023 to September 30, 2023",
        y = "Daily Average Mood Rating" ,
```

```
x = "Days of the Week" )
p2
```



The bar chart shows that I study most on Tuesday and least on Friday. There is not much variation among the days of the week apart from Wednesday and Friday which are significantly less than the other days. I then decided to visualize the trend of mood ratings across the week to see if there is any similarity with how the daily study hours changed. The line graph shows that my mood is lowest on Monday and it increases progressively as the week goes by. However, since the trend is not similar to the bar chart, there is not much information to conclude any association between hours spent studying and daily average mood. Thus, I plotted the two variables with an intent of looking for an association.

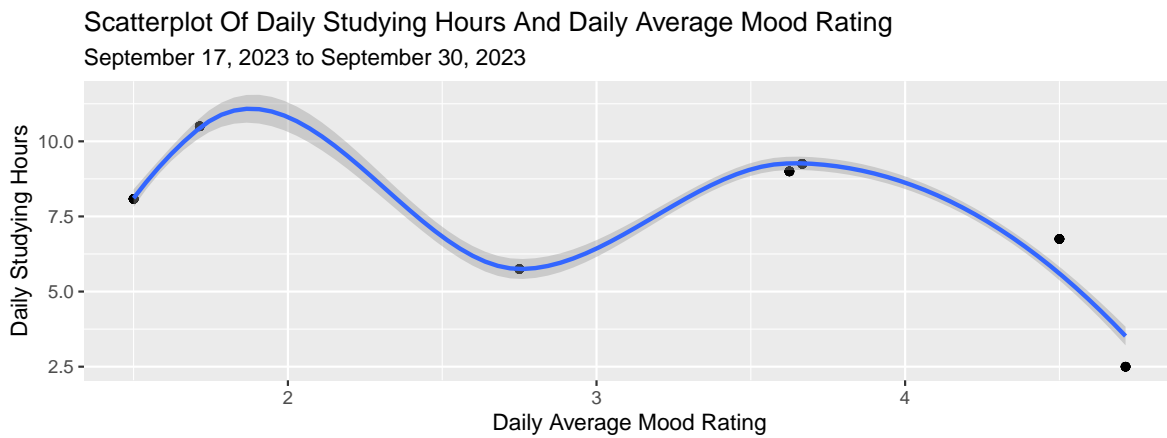
```
# Use `suppressWarnings` function to prevent display of warning messages in pdf
# Use `select` function to keep the variables/columns needed
# Use the `filter` function to only keep observations where
# `activity == "mood"` or `activity == "actual studying"`
suppressWarnings(
  mycal4 <- mycal %>%
    select(activity,description,weekday_label,duration_hours) %>%
    filter(activity == "mood" | activity == "actual studying" ) %>%
  # Store values of `duration_hours` in `study_hours` for observations
  # where `activity == "actual studying"`
  mutate(mood_int = as.integer(description),
    study_hours = case_when(activity == "actual studying" ~ duration_hours)) %>%
```

```

group_by(weekday_label) %>%
mutate(mean_mood = mean(mood_int, na.rm = TRUE),
       daily_hours = sum(study_hours, na.rm = TRUE)))

# Plot the graph
p4 <- ggplot(
  data = mycal4,
  aes(y = daily_hours, x = mean_mood)) +
  geom_point()+
  geom_smooth()+
  labs( title = "Scatterplot Of Daily Studying Hours And Daily Average Mood Rating",
        subtitle = "September 17, 2023 to September 30, 2023",
        y = "Daily Studying Hours" ,
        x = "Daily Average Mood Rating" )
p4

```



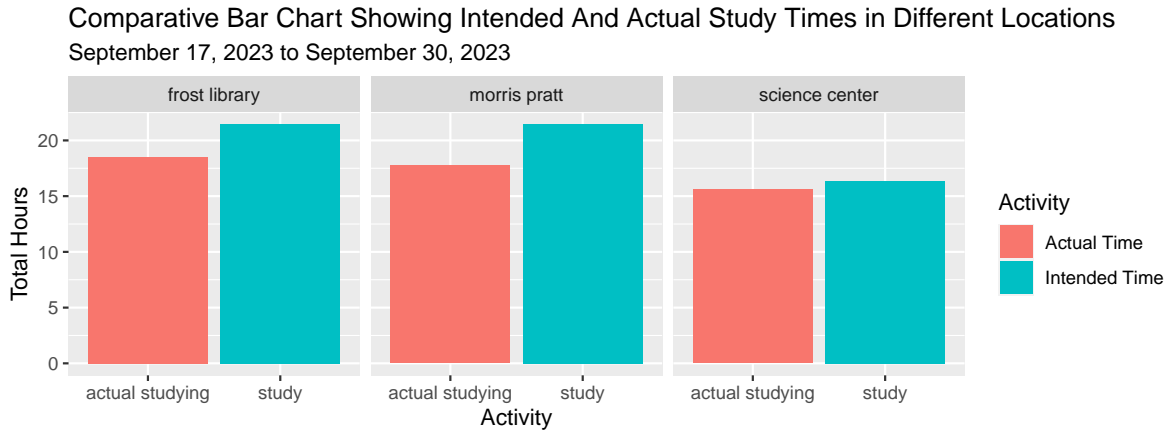
From the scatterplot, we see that there is no linear association between Daily Studying Hours and Average Mood Ratings. However, it is important to note that the data collected over this time period is not sufficient to draw conclusions because there are very few points on the plot.

Part 2 : Study Environments and Productivity

```
# Use `select` function to keep the variables/columns needed
# Use the `filter` function to only keep observations where
# `activity == "study"` or `activity == "actual studying"`
# group by location and activity because we need separate plots for each location
# find total hours spent studying on each location and store in `t_study`
mycal3 <- mycal %>%
  select(activity,duration_hours,description) %>%
  filter(activity == "study"| activity == "actual studying")%>%
  rename(location = description) %>%
  group_by(location,activity) %>%
  summarize(t_study = sum(duration_hours))

# Plot the bar chart
# Fill by activity to have different colors for the two activities
# Use location in `facet_wrap` function to have separate plots for the locations
p3 <- ggplot(
  data = mycal3,
  aes(x = activity, y =t_study, fill = activity)) +
  geom_bar(stat = "identity")+
  facet_wrap(~location)+
  labs( title = "Comparative Bar Chart Showing Intended And Actual Study Times in Different Locations",
        subtitle = "September 17, 2023 to September 30, 2023",
        x = "Activity" ,
        y = "Total Hours") +
# Rename `study` to `Intended Time` and `actual studying` to `Actual Time`
  scale_fill_discrete(name = "Activity",
                      breaks=c("actual studying", "study"),
                      labels=c("Actual Time", "Intended Time"))
```

p3



The bar chart shows that I study in the Science Center the least amount of hours compared to the other location. In all three locations, my actual study time is less than the duration I intended to study. However, on the summary table below, we see that my productivity is highest in the Science Center (95.4%) while I am least productive in my dorm room (82.6%).

Summary Table

```
# Use `pivot_wider` function on mycal3 dataset to make the two activities columns
# Calculate the productivity((`actual studying`/intended study) * 100 )) at each location
# Summarize the information and use `kable` function to make a table
mycal3 %>%
  pivot_wider(names_from = activity, values_from = t_study) %>%
  summarize(study, `actual studying`, prod = (`actual studying`/study) * 100 ) %>%
  kable(booktabs = TRUE, digits = 2, col.names = c("Location", "Actual Study Time", "Intende
```

Location	Actual Study Time	Intended Study Time	Productivity(%)
frost library	21.50	18.50	86.05
morris pratt	21.50	17.75	82.56
science center	16.33	15.58	95.41

Conclusions

One of the most remarkable takeaways from this study is the unexpected finding that I am most productive in the Science Center. This discovery challenges my previous tendencies, as I typically refrained from studying in this location. Moving forward, based on this newfound insight, I intend to incorporate more study sessions in the Science Center into my routine.

Moreover, the study suggests that my moods, as they fluctuate, do not appear to significantly influence the number of hours I allocate to personal studies. This shows that I can maintain a consistent level of productivity irrespective of mood variations. This is also important as it implies that I can still make substantial progress even on days when my mood is less than optimal.

Reflection

During the data collection and analysis process for this project, I encountered a few challenges. Firstly, maintaining consistency in mood ratings posed a challenge. While I set reminders to record my mood at four-hour intervals, there were instances when external factors, such as distractions or time constraints, interfered with timely mood assessments. This inconsistency may have introduced bias into the data, impacting the validity of mood-related conclusions. Secondly, accurately measuring the time spent on personal studies was difficult. Although I used timers during study sessions, there were instances where I unintentionally paused or forgot to stop the timer, leading to overestimations or underestimations of study durations. Such inaccuracies could affect the assessment of the relationship between study time and productivity.

These challenges carry important lessons for future analysis projects. They emphasize the need for careful planning and better data collection procedures to ensure accuracy. Furthermore, they highlight that although larger datasets can offer more insights, they can also pose challenges in terms of time and effort. Regarding the amount of data required to answer my research questions, a minimum of one month period would likely yield more generalizable insights. However, collecting a this amount of data might be challenging due to the time and effort required, especially for variables that require frequent entries like mood ratings.

As someone who provides data to platforms like Facebook, Google, or MapMyRun, my expectations include data privacy and transparency. I expect that my personal data will be handled responsibly, kept confidential, and used only for the intended purpose. Additionally, I expect transparency regarding data usage and the option to control and consent to how my data is utilized. When analyzing others' data, my ethical responsibilities include safeguarding individuals' privacy and ensuring data is used in a way that respects consent and confidentiality. Moreover, I should aim to minimize bias and ensure the results are presented honestly and without misrepresentation