

# Time Spent on Athletics and Academics

SAT 231: Calendar Query

Sofia Guttman

2024-02-28

## Introduction

1. Time spent in class versus doing homework/studying This question is important because for every hour spent in class, I want to be spending 2.5-3 hours doing supplemental studying, reading, and assignments. I want to make sure my time is being spent and allocated properly.
2. Time spent on academics versus athletics This question is important because now that I am in the middle of lacrosse season, my time is very valuable. I want to maintain my time for studies because my time for athletics is not very flexible. I want to see how many hours in the day are left over that I can use for my studies.

## Data collection

I took note of every time I would study, do an assignment, go to class, go to a workout, go to practice, or go to office hours over the course of 14 days. The variables are hours (quantitative) and day of the week (qualitative).

```
# Load required libraries
library(ical)
library(tidyverse)
library(lubridate)

# Data import using the correct Google Calendar URL
cal_import <- ical_parse_df("https://calendar.google.com/calendar/ical/c_e5fb39631c03a80b1
```

```

# 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,
      tzone = "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

```
# Group by weekday_label and calculate total time spent in each activity
mycal_summary <- mycal_pivot %>%
  group_by(weekday_label) %>%
  summarise(
    Class = sum(class, na.rm = TRUE),      # Corrected from 'Class' to 'class'
    Practice = sum(practice, na.rm = TRUE),
    Academics = sum(Academics, na.rm = TRUE)
  ) %>%
  ungroup()

# Arrange the weekdays in the correct order
weekdays_order <- c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")
mycal_summary$weekday_label <- factor(mycal_summary$weekday_label, levels = weekdays_order)

# Display the table
mycal_summary
```

One table I will use will have the day of the week in one column, time spent in athletics in the next column, time spent in the classroom in the next column, and time spent doing homework/studying in the next column.

```
library(ggplot2)

# Filter data for 'class' activity
class_data <- mycal_filtered %>%
  filter(activity == "class")

# Create a scatter plot
ggplot(class_data, aes(x = duration_hours, y = duration_hours)) +
  geom_point() + # Add points
```

```
labs(x = "Time spent in Academics", y = "Time spent in Class") + # Label axes
ggtitle("Time Spent in Class vs. Academics") + # Add title
theme_minimal() # Apply a minimal theme
```

I will use a scatter plot to display the first question. The y axis will be time spent in class and the x axis will be time spent doing homework/studying.

```
library(ggplot2)

# Filter data for "Academics" and "Practice" activities
filtered_activities <- mycal_filtered %>%
  filter(activity %in% c("Academics", "practice"))

# Create a bar graph
ggplot(filtered_activities, aes(x = weekday_label, y = duration_hours, fill = activity)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Day of the Week", y = "Hours") +
  ggtitle("Hours Spent on Academics and Practice by Day of the Week") +
  scale_fill_manual(values = c("Academics" = "blue", "practice" = "red")) +
  theme_minimal()
```

I will use bar graphs to display the second question. The x axis will be the different days of the week with two bars in each day. One of academics and one for athletics. The y axis will represent time.

## Conclusions

Through this project, although I am not sure that my data is displayed 100% correctly through the code, I learned a lot. I learned that a lot of my time that I dedicate to doing work is after practice. While logging this information, I realized that I am very tired while doing my homework. Although I dedicate ample time to my studies, the quality is not as high as I would prefer due to the time of the day that I am working.

## Reflection

In gathering accurate data, it was difficult to be accurate because during my studies, I would take short five minute breaks. Therefore the time is not 100% accurate. Additionally, another issue is many of my days are structured the same. Therefore, there is not a lot of variation in the data between days except for Fridays and the weekend.

In the future, I would like to test the time of day and number of tasks I complete. I feel that I can make better use of my time in the morning to work when I am less tired. I would perhaps also like to repeat this project over the course of a full month and gather more intense data.

I have an ethical responsibility to be truthful in order to accurately represent my time and dedications. If I were to lie, then I would have less to learn about myself and my habits.