Where Do My Hours Go?

STAT 231: Calendar Query

May Mishima

2025-10-09

Practicing Academic Integrity

If you worked with others or used resources outside of provided course material, such as Generative AI, (anything besides our textbook, course materials in the repo, labs, R help menu) to help complete this assignment, please acknowledge them below using a bulleted list. Remember that according to our class framework, if you engage with Generative AI tools, the tool and date of use must be listed here, and then a complete record of the chat history must be appended to the assignment.

I acknowledge the following individuals with whom I worked on this assignment:

Name(s) and how they assisted you

-

I used the following sources to help complete this assignment:

Source(s) and where used in the assignment

- https://ggplot2.tidyverse.org/reference/geom_bar.html to figure out geom_col()

Introduction

As a student-athlete, I've often wondered how I truly spend my time. My overarching question for this project was, "How do I spend my time throughout the week?". More specifically, I wanted to explore how my hours were distributed for each activity and how the day of the week had any impact on the distribution. One important concept I wanted to examine was work–life balance. As a student-athlete, I often find myself caught up in academic and athletic responsibilities, sometimes sacrificing my downtime in pursuit of success in both areas. Lastly, I was interested in where I spent most of my time and which locations played the biggest roles in my daily routine. By analyzing this data, I hope to better understand my habits, reflect on how I allocate my time, and find ways to become more efficient while maintaining a healthier balance between work and rest.

Data collection

For this project, I tracked my daily activities over a two-week period using Google Calendar. Each calendar entry included the activity name, location, and a brief description of what the activity involved. I categorized each activity into one of five categories: academics, athletics, daily living, leisure, or work. I also recorded the duration of each activity in hours. The location variable identified where each activity took place, and the description provided additional context about the task. Together, these variables allowed me to analyze how my time was distributed across different activities, days, and environments.

```r
# Data import and preliminary wrangling
calendar_data <- "MyCalendar.ics" |>
  ## Use calendar package to import into R
  calendar::ic_read() |>
  ## Convert to "tibble" data frame format
  as_tibble() |> # calendar event descriptions are in variable called "summary"
  ## "activity" is a more relevant/informative variable name
  rename(activity = SUMMARY,
    location = LOCATION,
    description = DESCRIPTION) |>
  mutate(
    ## Specify time zone (defaults to UTC otherwise)
    start_datetime = with_tz(DTSTART, tzone = "America/New_York"),
    end_datetime = with_tz(DTEND, tzone = "America/New_York"),
    ## Compute duration of each activity in hours
    duration = interval(start_datetime, end_datetime) / hours(1),
    ## Convert text to lower case & trim spaces to clean up
    activity = str_to_lower(activity),
    activity = str_trim(activity),
    ## Separate date from time; Examples of ways to parse dates, times
    date = floor_date(start_datetime, unit = "day"),
    year = year(date),
    month = month(date, label = FALSE),
    day = day(date),
    day_of_week = wday(date, label = TRUE), day_of_year = yday(date)) |>
    ## Remove spurious year (added to every Google calendar)
  filter(year != 1969) |>
```

```
## Keep only the variables I want
select(activity, description, location, start_datetime, end_datetime,
    duration, date, year, month, day, day_of_week, day_of_year)

#fix inconsistencies with spelling
str_detect(calendar_data$location, "Greenway")
```

```
  [1] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
 [13] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
 [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
 [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
 [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[109]  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE
[121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[205] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE
[217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
str_detect(calendar_data$location, "Hit")
```

```
  [1] FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE
 [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
 [25]  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
 [37] FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE
 [49]  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
 [61] FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
 [73] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
```

```
 [85]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
 [97] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
[109] FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE
[121]  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
[133] FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
[145] FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
[157]  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE
[169]  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE
[181] FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE
[193]  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE
[205]  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[217] FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE
```

```r
  calendar_data <- calendar_data |>
    mutate(location = case_when(
      str_detect(location, "Greenway") ~ "Greenways Residence Halls",
      str_detect(location, "Hit") ~ "Hitchcock Hall",
      TRUE ~ location))

#wrangling for boxplot
#categorize activities into Work and Leisure
calendar_data <- calendar_data |>
  mutate(category = case_when(
    activity %in% c("academics", "work") ~ "Work",
    str_detect(description, "softball") ~ "Work",
    activity == "leisure" |
      str_detect(description, "LateNight|run|Squash") ~ "Leisure"
  ))
daily_totals <- calendar_data |>
  group_by(date, category) |>
  summarize(total_duration = sum(duration)) |>
   filter(category %in% c("Work", "Leisure"))
#wrangling of location_sum data for table
location_sum <- calendar_data |>
  group_by(location) |>
  summarize(
    avg_hours = mean(duration),
```

```
    total_hours = sum(duration)
  ) |>
  mutate(
percent_of_total = 100 * total_hours / sum(total_hours)) |> #create % variable
  arrange(desc(percent_of_total))
location_sum1 <- location_sum |>
filter(percent_of_total >=1) |>
  #cleaning up row names
  rename(
    "Average Hours" = avg_hours,
    "Total Hours" = total_hours,
    "Percent of Total" = percent_of_total,
    "Location" = location
  )
```

```
glimpse(calendar_data)
```

```
Rows: 224
Columns: 13
$ activity       <chr> "athletics", "dailyliving", "leisure", "dailyliving", "~
$ description    <chr> "Softball_Lift", "Lunch", "Nap", "Breakfast", "GetReady~
$ location       <chr> "Alumni Gym", "Science Center", "Hitchcock Hall", "Vale~
$ start_datetime <dttm> 2025-09-17 07:45:00, 2025-09-24 13:00:00, 2025-09-24 1~
$ end_datetime   <dttm> 2025-09-17 08:45:00, 2025-09-24 13:20:00, 2025-09-24 1~
$ duration       <dbl> 1.0000000, 0.3333333, 2.5000000, 0.5000000, 0.5000000, ~
$ date           <dttm> 2025-09-17, 2025-09-24, 2025-09-24, 2025-09-25, 2025-0~
$ year           <dbl> 2025, 2025, 2025, 2025, 2025, 2025, 2025, 2025, 2025, 2~
$ month          <dbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9~
$ day            <int> 17, 24, 24, 25, 17, 17, 17, 17, 17, 25, 25, 25, 25, 25,~
$ day_of_week    <ord> Wed, Wed, Wed, Thu, Wed, Wed, Wed, Wed, Wed, Thu, Thu, ~
$ day_of_year    <dbl> 260, 267, 267, 268, 260, 260, 260, 260, 260, 268, 268, ~
$ category       <chr> NA, NA, "Leisure", NA, NA, NA, "Leisure", NA, "Leisure"~
```

In this analysis, I use two visualizations and a summary table to explore the patterns in how I spend my time. I use colored bar graphs that are separated by activity and days of the week to determine how my total hours are distributed across activities. I also use a colored boxplot to

compare the total hours per day that I spend on work activities vs. leisure. Finally, I include a summary table that reports average hours, total hours, and percent of total time I spend at each location.

Results

The first figure captures how the total hours I spend each day of the week on certain activities. I created a bar graph using geom_col() to show the total hours instead of the frequency of each activity. It appears that I spend the most hours on daily living (sleeping, eating, getting ready, cleaning) with Sunday having the most hours of daily living which makes sense because I have the most free time and the time to prepare myself for the week. I also noticed that I spend the most time on academics on Wednesday which also makes sense since I have the most amount of classes as well as time to work on homework.

```
#bar graph of total hours spent based on activity split by day of the week
ggplot(calendar_data, aes(x = day_of_week, y = duration, fill = activity)) +
  geom_col(show.legend = FALSE) + #col to add total hours in y
  facet_wrap(~ activity, scales = "free_y") + #diff scale for each activity
  labs(title = "Total Hours Spent per Activity by Day of the Week",
    x = "Day of the Week",
    y = "Total Hours")
```

Total Hours Spent per Activity by Day of the Week
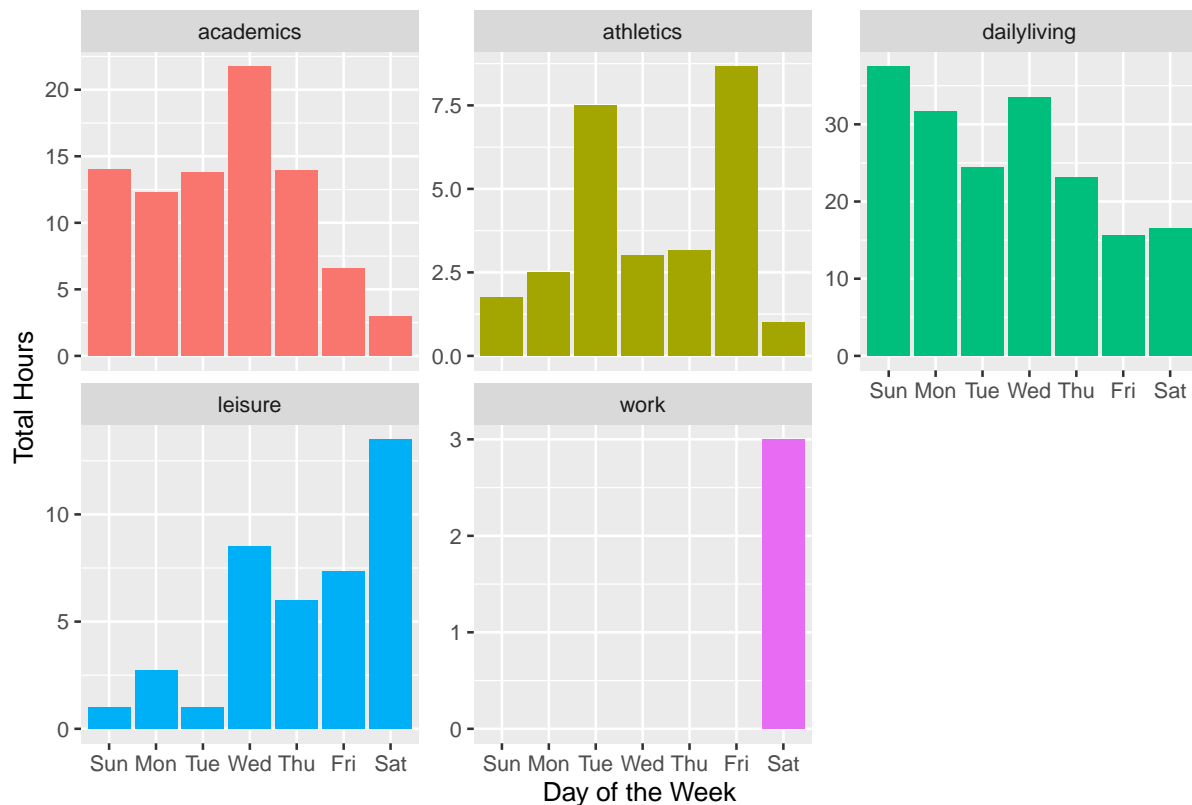
The second figure shows my work-leisure balance on a daily basis. I grouped "leisure" to include activities such as calling my family, napping, playing games, attending parties, and more. On the other hand, I grouped "work"as any time I spent in classes or studying, and any softball commitments that I had. Looking at the boxplot, I see that my work leisure balance is not great as I spend almost twice the amount I relax on doing work. This makes sense as a student-athlete because I do not get much down time to myself and most of the time I don't have softball or schoolwork, I spend on daily living activities such as eating and sleeping.

```
#boxplot for work-leisure balance for each day
ggplot(daily_totals, aes(x = category, y = total_duration, fill = category)) +
  geom_boxplot() +
  labs(title = "My Work-Leisure Balance on a Daily Basis",
       x = "Category",
       y = "Duration per Day (hours)",
       fill = "Category")
```

My Work–Leisure Balance on a Daily Basis

Duration per Day (hours)

Category
- Leisure
- Work

Category: Leisure, Work
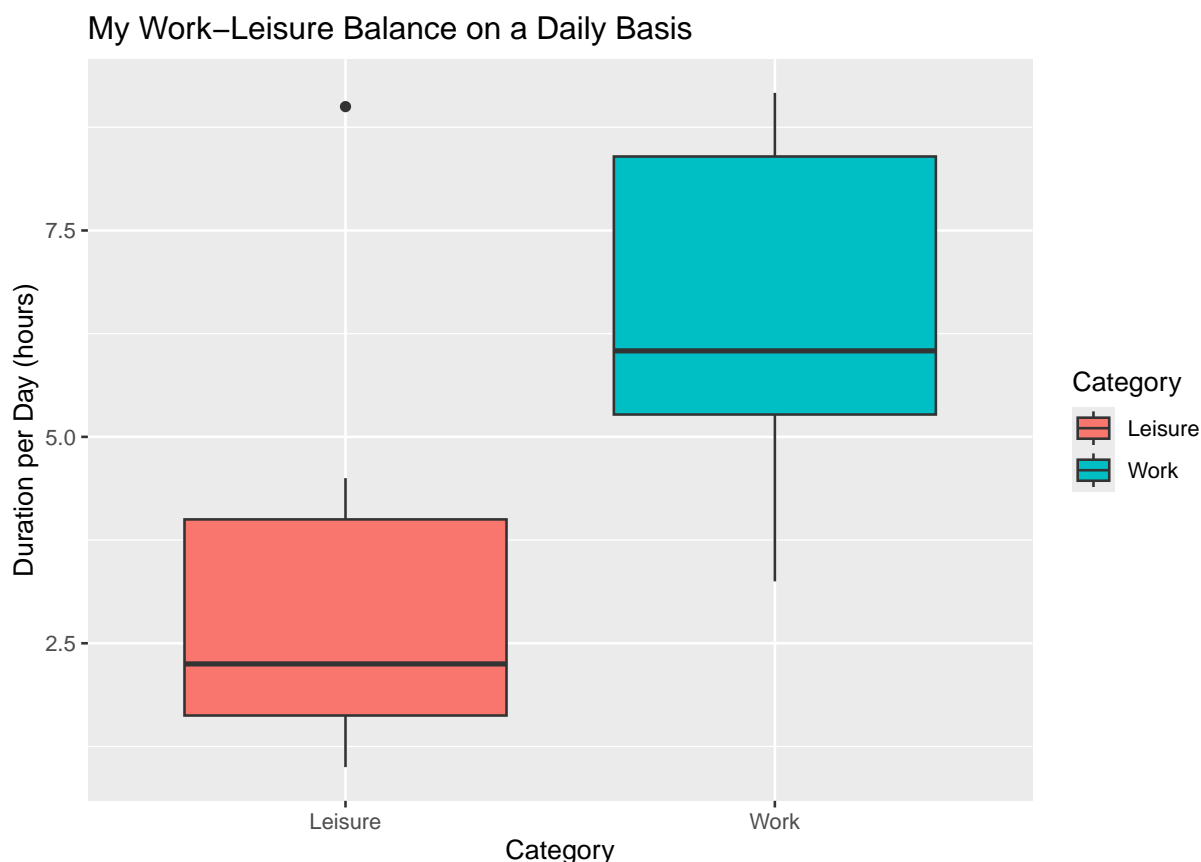
The table shows where I spend my time the most. The three summary statistics I chose were the average amount of hours, the total hours, and percentage of total hours that I spent at each location. I only presented the locations that was 1% of the total or more for organization sake, but I found that I spent the most hours in my dorm or at the dining hall. I found it interesting that just because my mean duration at a location was high does not mean that it was a high proportion of where I spent my time. I also understand that some of the averages are not very accurate because I spent a long time in them but did not go that often.

```
knitr::kable(location_sum1, digits = 3, booktabs = TRUE)
```

Location	Average Hours	Total Hours	Percent of Total
Hitchcock Hall	2.015	181.333	53.570
Valentine Dining Hall	0.754	31.667	9.355
Greenways Residence Halls	1.750	22.750	6.721
Frost Library	1.513	19.667	5.810

Location	Average Hours	Total Hours	Percent of Total
Webster Hall	0.974	12.667	3.742
Softball Field	2.896	11.583	3.422
Alumni Gym	1.000	9.000	2.659
Science Center	1.024	7.167	2.117
Converse Hall	0.875	5.250	1.551
Beneski Museum of Natural History	1.104	4.417	1.305

Conclusions

Through the analysis of my data, I gained a much clearer understanding of how I spend my time as a student-athlete during the fall season and where my daily balance truly lies. The results showed that a large portion of my time is devoted to commitments—both academic and athletic—as well as essential daily activities such as sleeping, getting ready, and eating. I also noticed that on days with fewer athletic commitments, I was able to dedicate more time to studying and catching up on schoolwork. My biggest takeaway was that my work–leisure balance is not as healthy as I had hoped. I tend to spend significantly more time working than relaxing, and the boxplot made it clear that I need to prioritize time for myself. Seeing this breakdown visually helped me recognize how structured and packed my schedule is. Moving forward, I want to be more intentional about scheduling personal time to rest and recover, ensuring a healthier balance between productivity and well-being. I also realized that I often study in the same places, and changing up my study locations could help me stay more motivated and focused throughout the week.

Reflection

The biggest challenge I faced during data collection and analysis was maintaining consistency. When I began analyzing my data, I noticed several spelling errors and extra spaces before or after words. At first, I thought I could simply fix these issues directly in my Google Calendar and re-export the data, but I quickly realized there were too many inconsistencies to correct manually. Although I was relieved to learn there were ways to clean and standardize the data, this process made my data wrangling much more tedious.

Another difficulty I encountered was how to group my data effectively. When I first started collecting information, I tried to format my activity, description, and location fields in a way that would make analysis easier later on. However, I soon realized this approach limited the level of detail I could include. For example, how should I classify an activity like getting Chick-fil-A and going for a drive with a friend—was it daily living or leisure? In the future, I would like to create more specific activity categories or include more descriptive notes. For instance, I could indicate whether I was studying, doing homework, or reading for class, and whether I was with friends or alone. To avoid consistency issues, I could also create a simple key or coding guide in my notes so that I use the same naming conventions for each activity.

Although I only collected two weeks of data for this project, I believe it provides a fairly accurate representation of my life at school during the fall season. The data would likely look different if I were in-season or at home for the summer, but overall, it captures my general habits well. Some limitations still exist; for example, the average hours spent in certain locations were based on only one or two events, which made those averages less meaningful. Like any dataset, mine would benefit from a larger sample size, but I still believe these two weeks give strong insight into how I manage my time.

From an ethical standpoint, as someone who provides data, I expect my information to be securely stored and only shared with my consent. A common problem is that users often overlook long and complicated terms of service agreements that request permission to access their data. There needs to be a more transparent and accessible way for users to understand what data is being collected and how it is used.

As someone who analyzes data, I recognize my responsibility to protect the privacy of others. With the increasing use of generative AI, it is especially important not to upload others' data into these systems without full transparency and consent. Additionally, as an analyst, I am responsible for presenting data honestly and clearly. The way I choose to visualize and interpret data shapes how others perceive it, so it is my ethical duty to ensure my work represents the data accurately and without distortion.