

# Tracking Time

## STAT 231: Calendar Query

Gloria Wu

2025-02-27

### Introduction

The primary question of interest for this project is: is there a relationship between the amount of time I spend sleeping each night and the amount of time I spend on various activities (studying, exercising, “rotting”, knitting, and reading) the following day? Additionally, I am interested if there is a relationship between the time that I spend on various activities throughout the day. This question is important to me because I deeply believe in the importance of having a good night of sleep in order to maximize productivity and efficiency the next day. Since coming to college, I have made my sleep schedule a priority, and I hope to see if any fluctuations in my sleep negatively impacts my workflow the following day.

### Methods

For this project, I used self-collected data over the period of 18 days: from 2.7.25 to 2.24.25. Specific methods of data collection and data wrangling are described below. Finally, I generated two scatterplots and a table to visualize the relationship between my variables of interest.

The variables of interest for this project are:

**sleep:** time spent sleeping. This includes time spent sleeping at night, as well as nap time, if any, during the day.

**exercise:** time spent on general exercise (warm-up, cool-down, workouts, and long walks).

**math, cosc, econ, psyc, stat:** time spent working on MATH 220, COSC 112, ECON 111, PSYC 100, and STAT 231, respectively. This includes any individual time spent studying and time spent in office hours, but not time spent in the actual class. For **cosc**, this also includes

individual time spent self-learning computer science.

**knit**: time spent knitting.

**read**: time spent reading.

**rot**: time spent on any form of social media (Instagram, Snapchat, YouTube, etc.), with the exception of watching YouTube while knitting.

## Data Collection

Data was collected via Apple watch/health (**sleep** was automatically tracked) and the Notes app. The start and end times of each activity, with the exception of **sleep**, were written down. Then, data was transferred from Notes and Apple health to Google Calendar. Then, calendar data was periodically exported as an .ics file to R Studio.

## Data Wrangling

To wrangle my data, I calculated the duration of each activity across each separate day. Because I wanted to compare sleep data with performance the following date, I made sure to set the date of each activity to the date at the end of the event. Then, I created four sub-datasets: (1) sleep data only, (2) all activity data without sleep, (3) knit data only, and (4) study data only. Then, I joined the sleep data (1) with the data about other activities (2), in order to map the amount of time spent on other activities relative to how much sleep I got that night. I also joined the knit data (3) with the study data (4), in order to compare the amount of time spent knitting versus studying on a day-to-day basis. Lastly, to investigate the relationship between rot times and study times, I widened the dataset, removed unnecessary variables, and generated a correlation table. For the correlation table, I also created a new variable, **study**, which summed all of my study times by date, in order to see my overall study habits.

```
cal_import <- ical_parse_df("data/wu-calendar-data.ics")

mycal <- cal_import |>
  rename(activity = summary) |>
  mutate(
    # specify time zone
    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),
  # set sleep as the end date to compare with following day's data
  date = date(end),
  year = year(start),
  year_end = year(end),
  weekday_number = wday(start),
  hour = hour(start),
  # 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)
) |>
filter(year_end != 1969)

```

Warning: There was 1 warning in `mutate()`.

i In argument: `across(c(start, end), .fns = with\_tz, tzone = "America/New\_York")`.

Caused by warning:

! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.

Supply arguments directly to `.fns` through an anonymous function instead.

```

# Previously
across(a:b, mean, na.rm = TRUE)

# Now
across(a:b, \(x) mean(x, na.rm = TRUE))

```

```

cal <- mycal |>
  select(activity, duration_hours, date) |>
  # sum the amount of time for each activity by date
  group_by(activity, date) |>
  summarize(duration = sum(duration_hours))

```

```

# create new dataset for sleep values
sleep_data <- cal |>
  filter(activity == "sleep")

# filter out sleep values
scat <- cal |>
  filter(activity != "sleep") |>
  # join with sleep values to compare individual activities with sleep
  inner_join(sleep_data, by = "date")

# create new dataset with only knit values
knit_data <- cal |>
  filter(activity == "knit")

# only keep study values
knit_study <- cal |>
  filter(activity == "math" | activity == "csc" | activity == "econ" |
          activity == "psyc" | activity == "stat") |>
  # join with knit values to compare study times with knit
  inner_join(knit_data, by = "date")

cal_wider <- cal |>
  # organize data by activity, change NA's to 0
  pivot_wider(names_from = "activity", values_from = "duration",
              values_fill = 0)

# generate correlation table for rot vs. study
cal_cor <- cal_wider |>
  select(- date, - sleep, - knit, - exercise, - read) |>
  mutate(study = math + csc + econ + psyc + stat) |>
  # reorder variables
  select(rot, study, math, csc, econ, psyc, stat) |>
  cor()

```

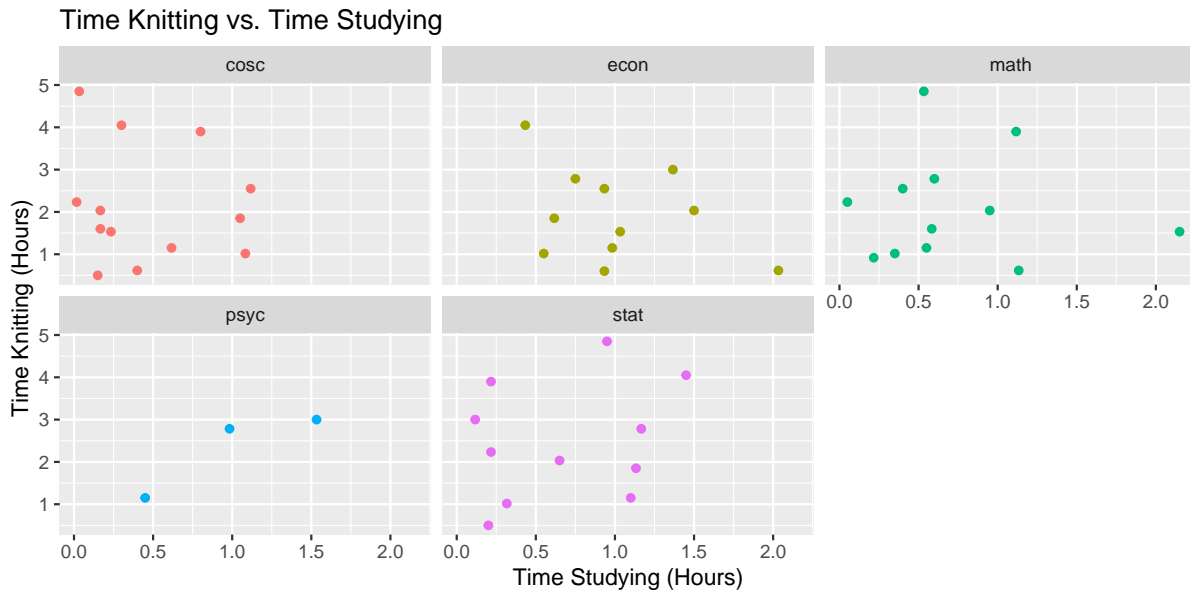
## Results

```
scat |>
  ggplot(aes(x = duration.x, y = duration.y, color = activity.x)) +
  geom_point(show.legend = FALSE) +
  facet_wrap(~ activity.x) +
  labs(title = "Time Sleeping vs. Time Spent During the Day",
       x = "Time Spent During the Day (Hours)",
       y = "Time Sleeping (Hours)")
```



First, to see if there was a relationship between the amount of time spent sleeping and the amount of time spent on various activities the next day, I created a faceted scatterplot comparing sleep times with times spent on different activities. There seems to be a relatively weak, negative relationship between sleep times and knit times, sleep times and time spent on psychology, and sleep times and time spent on statistics. There seems to be a relatively weak, positive relationship between sleep times and time spent on economics. For the rest of the variables (*cosc*, *exercise*, *math*, *read*, *rot*, *stat*), there doesn't seem to be a strong relationship between them and sleep times.

```
knit_study |>
  ggplot(aes(x = duration.x, y = duration.y, color = activity.x)) +
  geom_point(show.legend = FALSE) +
  facet_wrap(~ activity.x) +
  labs(title = "Time Knitting vs. Time Studying",
       x = "Time Studying (Hours)",
       y = "Time Knitting (Hours)")
```



Because there seemed to be the greatest relationship between sleep times and time spent knitting, I decided to investigate if there was a relationship between the time spent knitting each day and the time spent studying on each subject. Thus, I created another faceted scatterplot, this time comparing time spent knitting and time spent studying per day, separated by subject. There seems to be a moderately strong, positive relationship between time spent knitting and time spent on psychology; however, there are only 3 data points and therefore hard to draw a conclusion. For the remainder of the subjects, there doesn't seem to be any noticeable relationships between time spent knitting and studying; however, generally, the overall trend seems to be that the more time I spend knitting, the less time I spend on other subjects.

```
kable(cal_cor, digits = 2, booktabs = TRUE, align = c("c"), format = "latex",
      longtable = FALSE,
      caption = "Correlations between Rot and Study Times")
```

Table 1: Correlations between Rot and Study Times

	rot	study	math	cosc	econ	psyc	stat
rot	1.00	0.05	0.19	-0.13	0.15	-0.07	-0.10
study	0.05	1.00	0.51	0.37	0.76	0.37	0.31
math	0.19	0.51	1.00	0.11	0.34	-0.18	-0.24
cosc	-0.13	0.37	0.11	1.00	0.12	-0.29	-0.02
econ	0.15	0.76	0.34	0.12	1.00	0.29	-0.09
psyc	-0.07	0.37	-0.18	-0.29	0.29	1.00	0.16
stat	-0.10	0.31	-0.24	-0.02	-0.09	0.16	1.00

Finally, I was interested in how the time I spent “rotting” (or essentially wasting away) affected my productivity that day. Thus, I created a correlation table between my rot times and study times (while the time I spend studying may not indicate productivity, it’s the closest measure that I have). This table shows that there is a weak, negative relationship between the time I spend rotting and studying computer science ( $r = -0.13$ ), psychology ( $r = -0.07$ ) and statistics ( $r = -0.10$ ), and a weak, positive relationship between the time I spend rotting and studying math ( $r = 0.19$ ) and economics ( $r = 0.15$ ). Interestingly, there is a positive (albeit very weak) relationship between the time I spend rotting and my overall study time every day ( $r = 0.05$ ).

## Conclusions

The biggest takeaway that I have from this project is that there is a negative relationship between the amount of time I spend sleeping and the amount of time spent knitting. This makes sense because generally, if I spend a long time knitting (or on any hobby in general), it would happen towards the end of the day, when I usually sleep. Perhaps going forward, I will be more mindful of the time I spend on hobbies, so as to not cut into my sleep time. Surprisingly, there is not any strong (either positive or negative) relationships between sleep times and study times, or knit/rot times and study times.

## Reflection

The biggest obstacle in collecting data was consistency; because the data collection was on such a short time interval, I didn't really get into the habit of writing down start and end times. This means that some of my time intervals may not be correct, especially for the amount of time I spent rotting each day. Additionally, I found that many of my activities happened over short periods of time (1-5 minutes), especially for rotting. Because they happened for such a short time (such as briefly opening the app and checking messages), I didn't feel the need to write down the start and end time. While this difference seems small, it may have built up to a significant amount over the 2.5 weeks.

For future data collection, I would be more careful to find a collection method that is more reliable. In the instance of "rotting," I could collect overall screen time data, rather than relying on my own ability to write down start and end times. Additionally, I think I would need a much longer time interval (such as a month) to answer my question of interests; because I had two exams (mathematics and economics) at the end of the data collection period, I likely spent more time on those two classes than other classes, which wouldn't be representative of my normal studying habits. I don't think it would be difficult to collect that data, although I would change my method of collecting rot time, as described above.

As someone who provides data, I would like for clear and accessible information on where my data is being stored and sold. For instance, I have heard that location apps such as Life360 sell data to insurance companies, which would then change how much people get charged for insurance; information like this would influence my decision to download certain apps. As someone who analyzes others' data, at the very least, I have a responsibility to maintain anonymity and confidentiality for all data, as well as giving credit where due. As a data collector, I have a responsibility to not encroach on others' privacy (for instance, I may be comfortable with sharing my "rot" time, but others may not be).



## Sources

<https://stackoverflow.com/questions/45845231/how-to-prevent-a-kable-from-splitting-between-pages> - How to print a kable without it being split between two pages

<https://stackoverflow.com/questions/35618260/remove-legend-ggplot-2-2> - How to remove a legend from a graph generated by ggplot

<https://stackoverflow.com/questions/8161836/how-do-i-replace-na-values-with-zeros-in-an-r-dataframe> - How to replace NA values with 0