STAT231: Google Calendar Report

Sean Wei

Due Friday, September 25 by 5:00 PM EST

Contents

0.1 Importing Data

```
# Import Calendar Data
library(lubridate)
library(ical)

path <- "/Users/seanwei/Desktop/STAT231-swei1999/calendar"
filename <- "seanlonewei@gmail.com.ics"

my_calendar <- ical_parse_df(file=paste0(path,"/",filename)) %>%
  mutate(start_datetime = with_tz(start, tzone = "America/New_York"),
         end_datetime = with_tz(end, tzone = "America/New_York"),
         length_sec = end_datetime - start_datetime,
         date = floor_date(start_datetime, unit = "day"))

# Initial Wrangling - Filter for Necessary Features & Dates, Convert Time to
# Hours (b/c the Time Was Imported as Seconds), and Add Time Ranges/Specific Days
my_calendar <- my_calendar %>%
  mutate(
    length_hour = as.numeric(round(length_sec * 0.000277778, digits = 2)),
    end_hour = hour(end_datetime),
    time_range = case_when(end_hour > 0 & end_hour < 12 ~ "Morning",
                           end_hour >= 12 & end_hour < 16 ~ "Afternoon",
                           end_hour >= 16 & end_hour < 20 ~ "Evening",
                           TRUE ~ "Night"),
    day = weekdays(date)
  ) %>%
  filter(date >= "2020-09-07") %>%
  select(-c(uid, description, last.modified, status, length_sec))
```

For this assignment, the questions I hoped to answer were:

1) Generally, how do I spend my days?

2) During what parts of the day do I do specific activities?

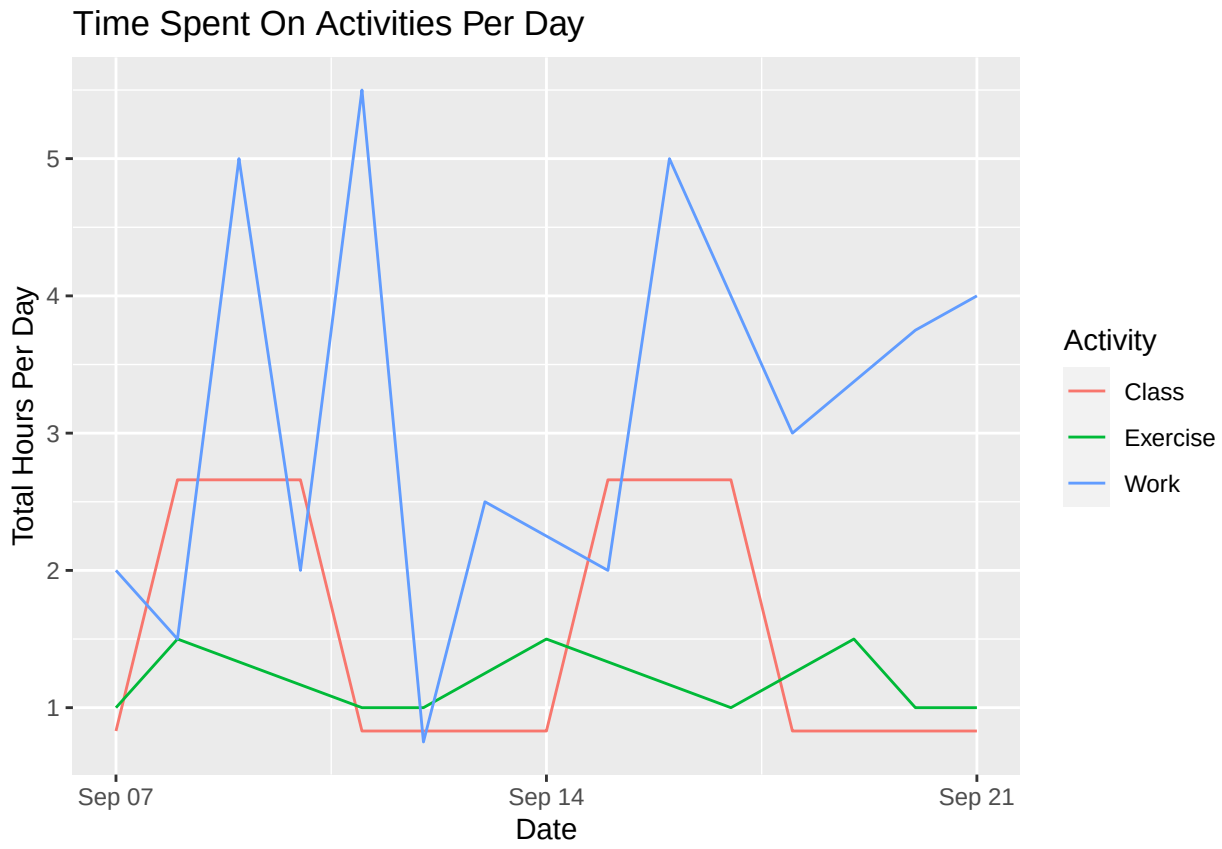
3) How much time do I spend on work outside of class?

To answer these questions, I kept track of the start and end times of my classes, doing homework, and exercising from September 7th, 2020 to September 21st, 2020. The main variables of interest in the `my_calendar` dataset include: `summary` (the type of activity that was done), `date` (calendar date), `length_hour` (how many hours I did for one instance of an activity), `time_range` (whether the activity took place during the morning, afternoon, evening, or night), and `day` (day of the week). Both the `summary` and `date` variables came from importing the calendar data. The `length_hour` variable was calculated by subtracting the start and end times. Since that result was in seconds, it had to be converted to hours by getting multiplied by 0.000277778. The `time_range` variable was generated by assigning a pre-calculated time range based on how early/late the activity ended. Lastly, the `day` variable was generated by the R function `weekdays()`, which extracts the weekday from `date`.

0.2 Visualization #1

In the visualization below, my goal was to see how much time I allocated to class, homework, and exercise throughout these two weeks. To accomplish this, the dataset had to be altered so that there was one total number of hours for each activity in a day, as there were days where I had multiple classes or did work multiple times. The visualization is a line plot of the total number of hours for each activity over the two weeks. From the legend, the viewer can gather that the red line represents class time, the green line represents exercise, and the blue line represents work completed outside of class.

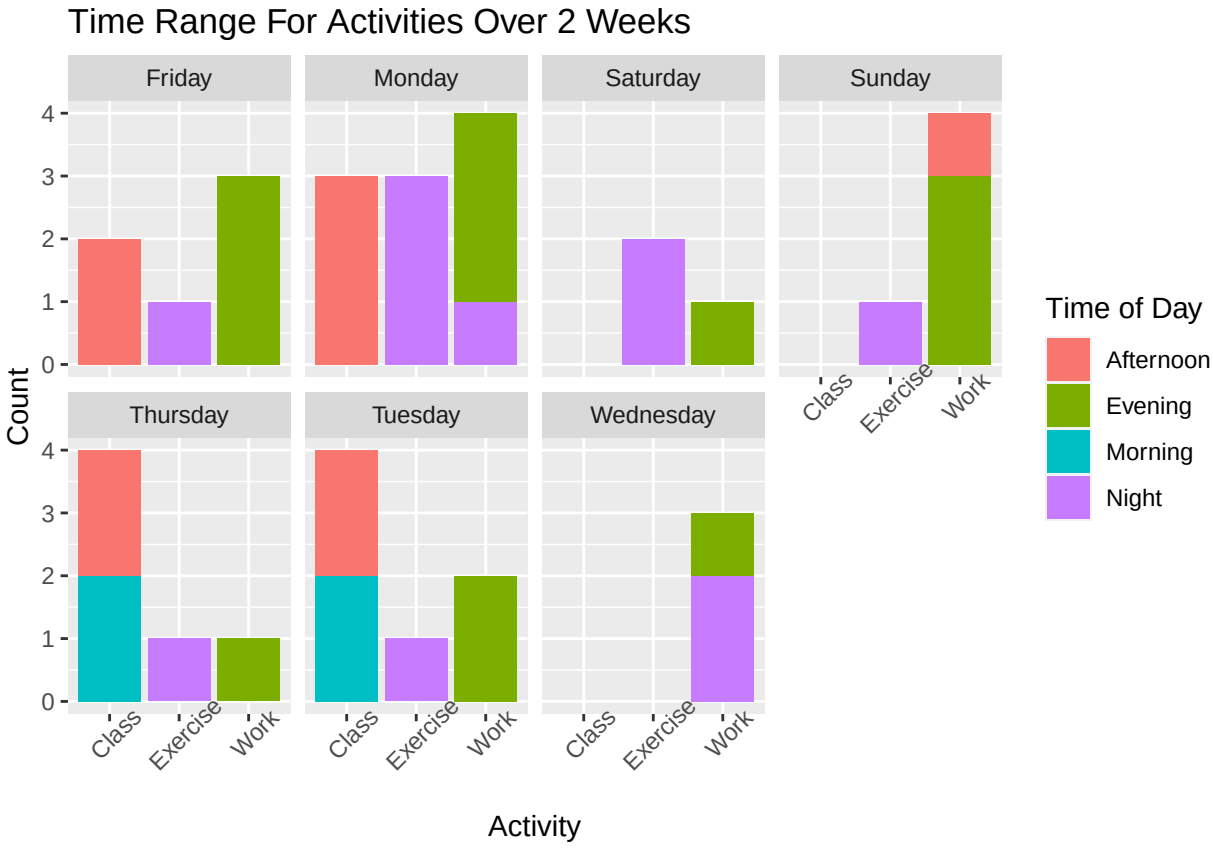
```
# Get Total Hours of Each Activity per Day
total_hours <- my_calendar %>%
  group_by(summary, date) %>%
  summarise(total_hours = sum(length_hour))

# Plotting the Visual
 ggplot(total_hours, aes(x = date, y = total_hours, color = summary)) +
  geom_line() +
  labs(x = "Date", y = "Total Hours Per Day", color = "Activity") +
  ggtitle("Time Spent On Activities Per Day")
```



0.3 Visualization #2

The aim of the second visualization is to see what time of day I tend to do each activity. Below, the viewer can see a series of bar graphs that show the number of times I did a specific activity over these two weeks. Each graph represents a separate day of the week, and each bar itself is colored based on what time of day the activity occurred. For example, the Friday graph conveys I had class two times in the afternoon in total on Fridays, and the Thursday graph conveys that I had class four times in total on Thursdays, twice in the morning and twice in the afternoon.

```
 # Plotting the Visual
ggplot(my_calendar, aes(x = summary)) +
  geom_bar(aes(fill = time_range)) +
  facet_wrap(~day, nrow = 2, ncol = 4) +
  labs(x = "Activity", y = "Count", fill = "Time of Day") +
  ggtitle("Time Range For Activities Over 2 Weeks") +
  theme(axis.text.x = element_text(angle = 45))
```



0.4 Table

Lastly, I used a table to try to see the amount of time I spent on work outside of class. To do this, the data first had to be manipulated so that I had the total hours of homework and class time for each day. From there, the homework to class time ratio was calculated for each day. The table is designed such that each row represents a specific day and the columns show the date, day of the week, homework hours, class time hours, and the work to class ratio. The table is also grouped by the week.

```
library(kableExtra)

# Filter for Wanted Data & Change Data to Wide Format to
# Get Specific Hours of Work and Class per Day
calendar_table <- my_calendar %>%
  filter(summary != "Exercise") %>%
  group_by(date, summary) %>%
  summarise(total_hours = sum(length_hour)) %>%
  pivot_wider(id_cols = date, names_from = summary, values_from = total_hours)

# Change All NA Values to 0
calendar_table[is.na(calendar_table)] = 0

# Extra Wrangling - Create Ratio of Homework to Class Time, Adding Day,
# Capitalizing Date, and Selecting Order of Columns
calendar_table <- calendar_table %>%
  mutate(
    `Work to Class Ratio` =
      case_when(Work == 0 ~ 0,
                Class == 0 ~ Work,
                TRUE ~ round(Work/Class, 2)),
    Day = weekdays(date)
  ) %>%
  rename(Date = date) %>%
  select(Day, Work, Class, `Work to Class Ratio`)

# Removes Last Observation b/c it's a Monday (Start of a 3rd Week)
calendar_table <- calendar_table[-nrow(calendar_table),]

# Create Table Using Kable
kable(calendar_table, booktabs = TRUE, linesep = "") %>%
  kable_styling(latex_options = "HOLD_position") %>%
  row_spec(0, bold = TRUE) %>%
  pack_rows("Week 1", 1, 7) %>%
  pack_rows("Week 2", 8, 13) %>%
  footnote(general = "There is no Saturday observation in week 2 because there were no recordings of cl
           threeparttable = TRUE)
```

Date	Day	Work	Class	Work to Class Ratio
Week 1				
2020-09-07	Monday	2.00	0.83	2.41
2020-09-08	Tuesday	1.50	2.66	0.56
2020-09-09	Wednesday	5.00	0.00	5.00
2020-09-10	Thursday	2.00	2.66	0.75
2020-09-11	Friday	5.50	0.83	6.63
2020-09-12	Saturday	0.75	0.00	0.75
2020-09-13	Sunday	2.50	0.00	2.50
Week 2				
2020-09-14	Monday	0.00	0.83	0.00
2020-09-15	Tuesday	2.00	2.66	0.75
2020-09-16	Wednesday	5.00	0.00	5.00
2020-09-17	Thursday	0.00	2.66	0.00
2020-09-18	Friday	3.00	0.83	3.61
2020-09-20	Sunday	3.75	0.00	3.75

Note:
There is no Saturday observation in week 2 because there were no recordings of class or work on that day.

0.5 Summary of Visuals

0.6 Reflection