YOUR TITLE HERE

STAT 231: Calendar Query

Sebastian Montesinos

Last updated March 7, 2022

Introduction

The first question I addressed is how the amount of time I spend on each class changes over the course of a week. One of my goals this semester is to have a structured and planned week where I dedicate certain days to certain classes. This allows me to easily plan when I will do work and not get overloaded on any particular day. I aim to concentrate my data science work towards the end of the week - specifically, around Thursday, Friday, and Saturday. I aim to do my oceanography and religion work around the start of the week, particularly Monday and Tuesday respectively. Finally, I aim to scatter my thesis work throughout the week.

The second question I address is how the amount of time I spend on work in a day relates to the amount of personal time I take during a day. One might expect that since the more time I spend on one activity the less time there is for the other, there will be a simple negative correlation. However, I have noticed that I sometimes give myself more time for recreational activities when I have done a lot of work as a 'reward', and spend more time on independent academic projects or looking for jobs when I do not spend a lot of time on work in a day since I feel guilty if I am not productive in a day. Therefore, it could also be the case that as I spent more time on schoolwork I also take more personal time, making the two positively correlated. I was also interested in how this relationship might change on weekdays vs weekends, since I try to give myself more of a break during the weekends.

Data collection

I coded my data into three broad categories: 'work', for anything related to school (ie. class and homework), 'personal' for anything related to recreation (ie. gaming, seeing friends) and 'extracurricular' for any work outside of class (ie. job applications, clubs, independent projects). I decided to code the work category as four subcategories corresponding to my four classes: data science, oceanography, religion, and my thesis. I decided not to further divide the other two categories since my questions did not require doing so. So, in the end, I had six ways I marked off time on my calendar: 4 for my classes, 1 for personal time, and 1 for extracurricular. The units for all of these categories were in time, specifically minutes spent on each activity, which I used google calendar to code in.

```
# Data import and preliminary wrangling
calendar_data <- "SMontesinosCalendarQuery.ics" %>%
  ## Use ical package to import into R
  ical_parse_df() %>%
  ## Convert to "tibble" data frame format
  as_tibble() %>%
  ## calendar event descriptions 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)
    start_datetime = with_tz(start, tzone = "America/New_York"),
    end_datetime = with_tz(end, tzone = "America/New_York"),
    ## Compute duration of each activity in hours
    ## Feel free to use minutes instead
    duration = interval(start_datetime, end_datetime) / hours(1),
    ## Convert text to lower case and trim spaces to help clean up
    ## potential inconsistencies in formatting
    activity = str_to_lower(activity),
    ## separate date from time
    date = floor_date(start_datetime, unit = "day"),
    ## Examples of ways to parse dates, times (keep only what you need!)
    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) %>%
  ## Turning the date variable into a date type
  mutate(date = ymd(date)) %>%
  ## Including only dates after I started collecting for the project
  filter(year >= 2022 & month >= 2 & day >= 17 | year >= 2022 & month >= 3) %>%
  ##Removing trailing whitespace
  mutate(activity = str_trim(activity)) %>%
  ##Replacing spaces with underscores in activity names
  mutate(activity = str_replace(activity, " ", "_")) %>%
  ##Creating another column that records whether the observation was made on the weekday or weekend
  mutate(week_status = case_when(day_of_week == "Sat" | day_of_week == "Sun" ~ "Weekend",
                                 TRUE ~ "Weekday"))
```

To address my first question, I intend to create a line graph that shows the average amount of time I spend on work on each day of the week, by the specific class I am working on. This graph will have the day of the week on the x axis and the time I spend on work on the y axis. There will be four lines, each corresponding to my four courses. The peaks of these lines

```
# Preparing dataset for first visualization
# Computing total duration for each activity per day of the week
activities_total <- calendar_data %>%
  group_by(date, day_of_week, activity) %>%
  summarize(duration = sum(duration)) %>%
#Pivoting wider to get rows as days and activities as columns
  pivot_wider(names_from = activity, values_from = duration)

#Filling in columns with NAs with '0' for 0 time spent
activities_total[is.na(activities_total)] = 0

#Calculating the mean time spent for each activity by day of the week
activities_average <- activities_total %>%
  group_by(day_of_week) %>%
```

```
  summarise(DS_Average = mean(data_science),
            O_Average = mean(oceanography),
            R_Average = mean(religion),
            T_Average = mean(thesis)) %>%
  pivot_longer(-day_of_week,
               names_to = "Activity",
               values_to = "Duration")
```

```
# Preparing dataset for second visualization


activities_comparison <- calendar_data %>%
#Selecting relevant variables
  select(day, activity, duration, week_status) %>%
# Adding a unique row identifier so I can pivot
  mutate(row = row_number()) %>%
# Pivoting wider to get each activity by day
  pivot_wider(names_from = activity, values_from = duration) %>%
# Dropping row identifier
  select(-row)
#replacing NA values with Os
activities_comparison[is.na(activities_comparison)] = 0

# Grouping by day and summing the activity duration into categories I want to compare
activities_comparison2 <- activities_comparison %>%
  group_by(day, week_status) %>%
# Calculating total school work time, personal time, and extracurricular time
  summarise(work_time = sum(thesis, oceanography, religion, data_science), personal_time = sum(personal
#
```

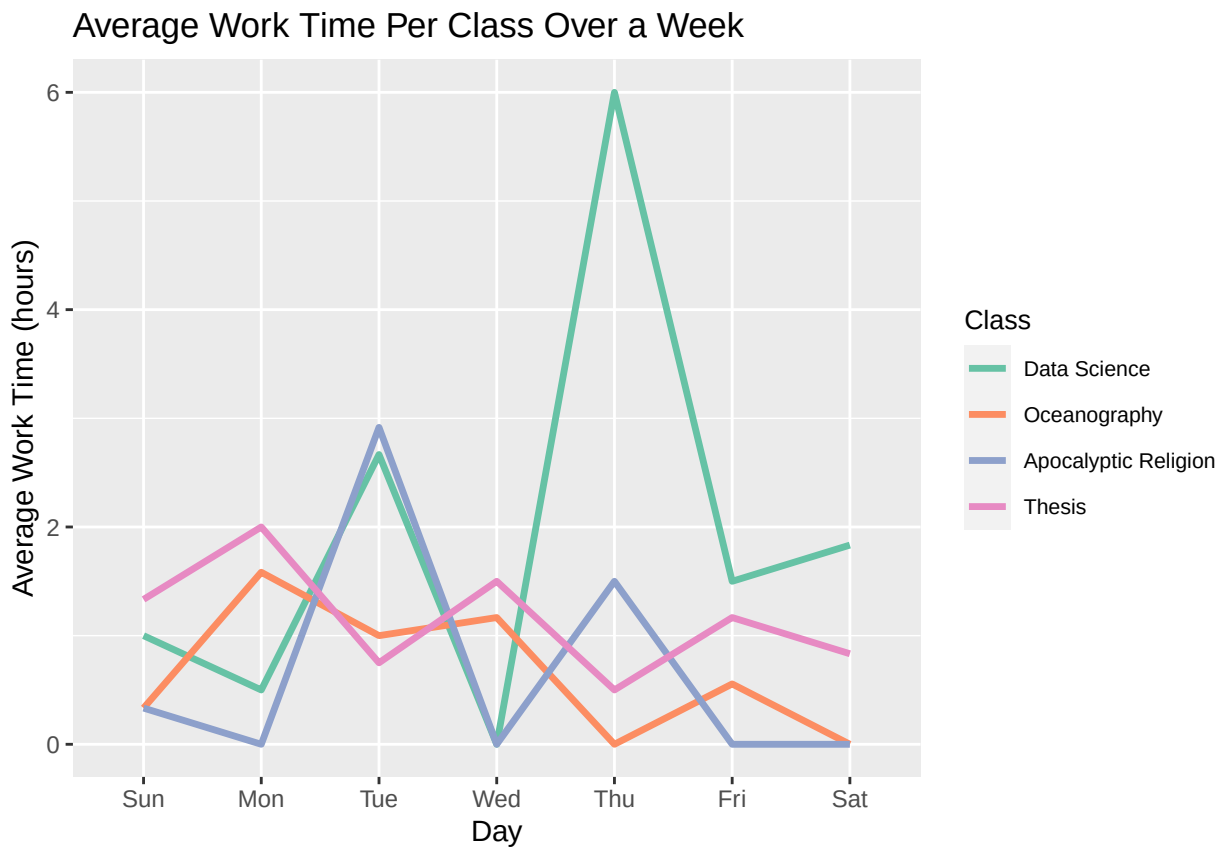
```
# Preparing dataset for table


week_comparison <- activities_comparison2 %>%
  group_by(week_status) %>%
  summarise(mean_work = mean(work_time),
            median_work = median(work_time),
            sd_work = sd(work_time),
            mean_personal = mean(personal_time),
            median_personal = median(personal_time),
            sd_personal = sd(personal_time),
            mean_extracurricular = mean(extracurricular_time),
            median_extracurricular = median(extracurricular_time),
            sd_extracurricular = sd(extracurricular_time)) %>%
  t() %>%
  janitor::row_to_names(1)
```

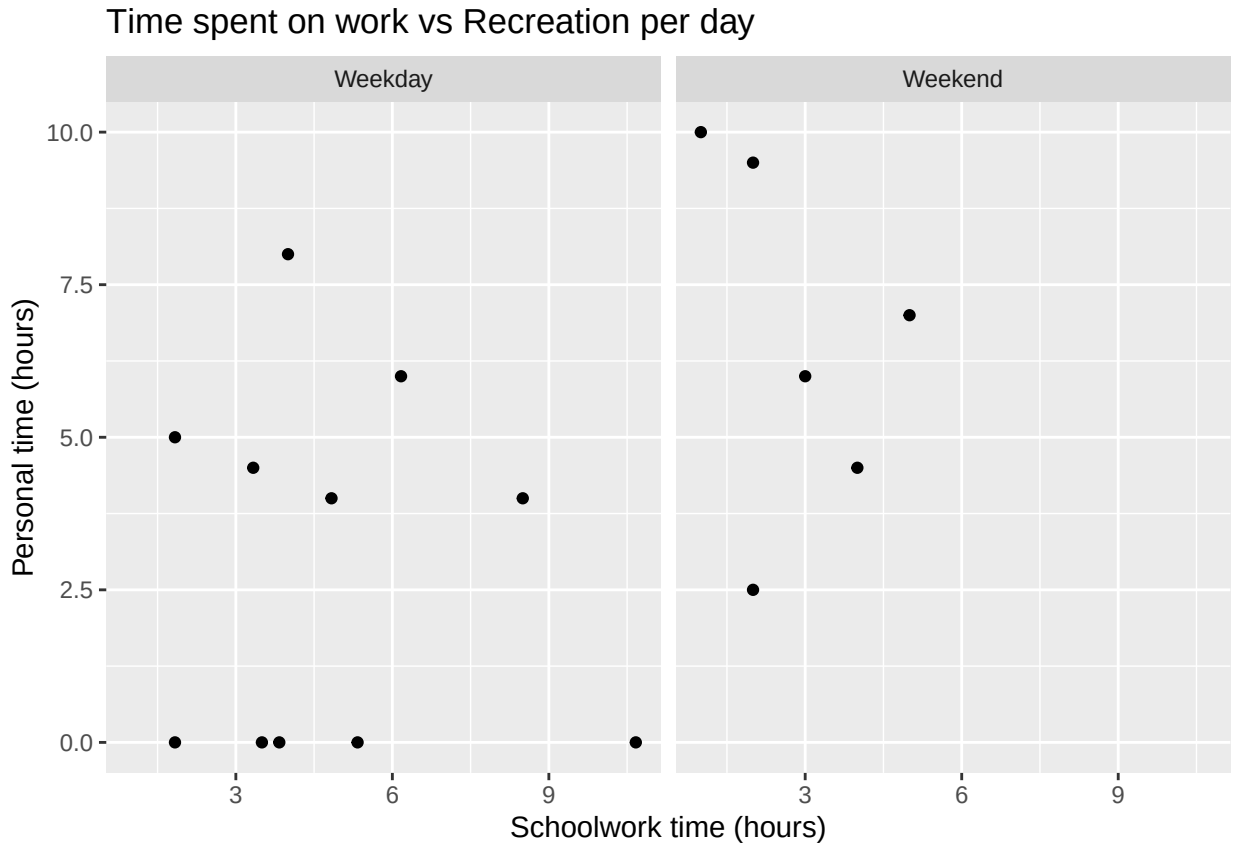
Results

```r
# Code for first data visualization
# Be sure to provide meaningful title and axes labels and
# resize figure appropriately
# Only code for your first visualization should be here (no or very minimal wrangling code)
# Remove all these comments!
p <- ggplot(data = activities_average,
            aes(x=day_of_week,
                y = Duration,
                color = Activity,
                group = Activity)) +
  geom_line(size = 1.2, alpha = 1) +
  scale_color_brewer(type = "qual",
                     palette = 7,
                     labels = c("Data Science",
                                "Oceanography",
                                "Apocalyptic Religion",
                                "Thesis")) +
  labs(DS_Average = "Data Science",
       title = "Average Work Time Per Class Over a Week",
       x = "Day",
       y = "Average Work Time (hours)",
       color = "Class") +
  theme(legend.text = element_text(size = 8)) +
  theme(legend.title = element_text(size = 10))

p
```




```
#Plotting total work time against personal time per day
n <- ggplot(data = activities_comparison2, aes(x= work_time, y = personal_time)) +
#Creating 1 plot for weekends and one for weekdays
  facet_wrap(~week_status, nrow = 1) +
#Using points to represent each day
  geom_point() +
#Labeling the graph
  labs(title = "Time spent on work vs Recreation per day",
       y = "Personal time (hours)",
       x = "Schoolwork time (hours)")
n
```




```
# Code for table
# Only code for your table should be here (no or very minimal wrangling code)

week_comparison <- week_comparison %>%
  kable(booktabs = TRUE)
week_comparison
```

	Weekday	Weekend
mean_work	4.893939	2.833333
median_work	4.0	2.5
sd_work	2.705214	1.471960
mean_personal	2.863636	6.583333
median_personal	4.0	6.5
sd_personal	2.950347	2.888194
mean_extracurricular	0.7272727	1.0000000
median_extracurricular	0.0	0.5
sd_extracurricular	1.009050	1.264911

Conclusions

Reflection

I think I did not have quite enough data to completely answer some of my important questions. For instance, my first question was about how the time I spend on each class changed throughout the course of an average week. However, I only had two weeks worth of data to average across. These two weeks are not perfectly representative of my average week. For instance, I have spent a lot more time on my thesis in the past and happened to have less work for it in the last two weeks since I am between writing sections. To get a representative sample of my typical week I think collecting data for at least 1-2 months would be ideal. Collecting this data would not be too difficult since I would just need to continue to track my time for a full month.