Calendar Query Individual Project

STAT 231: Calendar Query

Dhyey Mavani, Class of 2025

Last updated September 25, 2021

How do I spend most precious and game-changing resource of the world: TIME?

PROVIDED CODE CHUNK:

```
# Data import and preliminary wrangling
calendar_data <- "FinalDhyeyCal.ics" %>%
  # Use ical package to import into R
  ical_parse_df() %>%
  # Convert to "tibble" data frame format
  as_tibble() %>%
  mutate(
    # Use lubridate package to wrangle dates and times
    start_datetime = with_tz(start, tzone = "America/New_York"),
    end_datetime = with_tz(end, tzone = "America/New_York"),
    duration_min = difftime(end_datetime, start_datetime, units = "mins"),
    date = floor_date(start_datetime, unit = "day"),
    # Convert calendar entry to all lowercase and rename
    activity = tolower(summary))

# Compute total duration of time for each day & activity
#Creating new data set named activities
#sums up the time of each activity in entire day
activities <- calendar_data %>%
  group_by(date, activity) %>%
  summarize(duration_min = sum(as.numeric(duration_min)))
```

Describe your question(s) here. Briefly describe your data collection process, including how you defined variables of interest.

I imported my 14-day calendar data from 09-12-2021 to 09-25-2021 with Google calendar in `ical` format.

In my calendar query project:

1. `activity` represents career_related_work, sleep, meditation and procrastination as the name of activities.
2. `duration_min` represents the time duration in minutes.
3. `actual_duration_min` represents total time spent doing that activity over the course of 14-days of data collection.
4. `intended_duration_min` represents the intended total time spent doing that activity over the course of 14-days.
5. `Total_well_being` sums up the voluntary self-care in terms of meditation and sports for better analysis of below posed questions.

Question 1

Document intended time doing things **(career_related_work, sleep, sports, meditation, procrastination)** versus actual time doing those things, and compare results.

Question 2

Document time spent in **(career_related_work) vs (total_well_being = sports + meditation) vs sleep**

Question 3

Document time spent in **procrastination vs (total_well_being = sports + meditation)**

Describe what information is conveyed through data visualization #1 (that you'll create below) here.

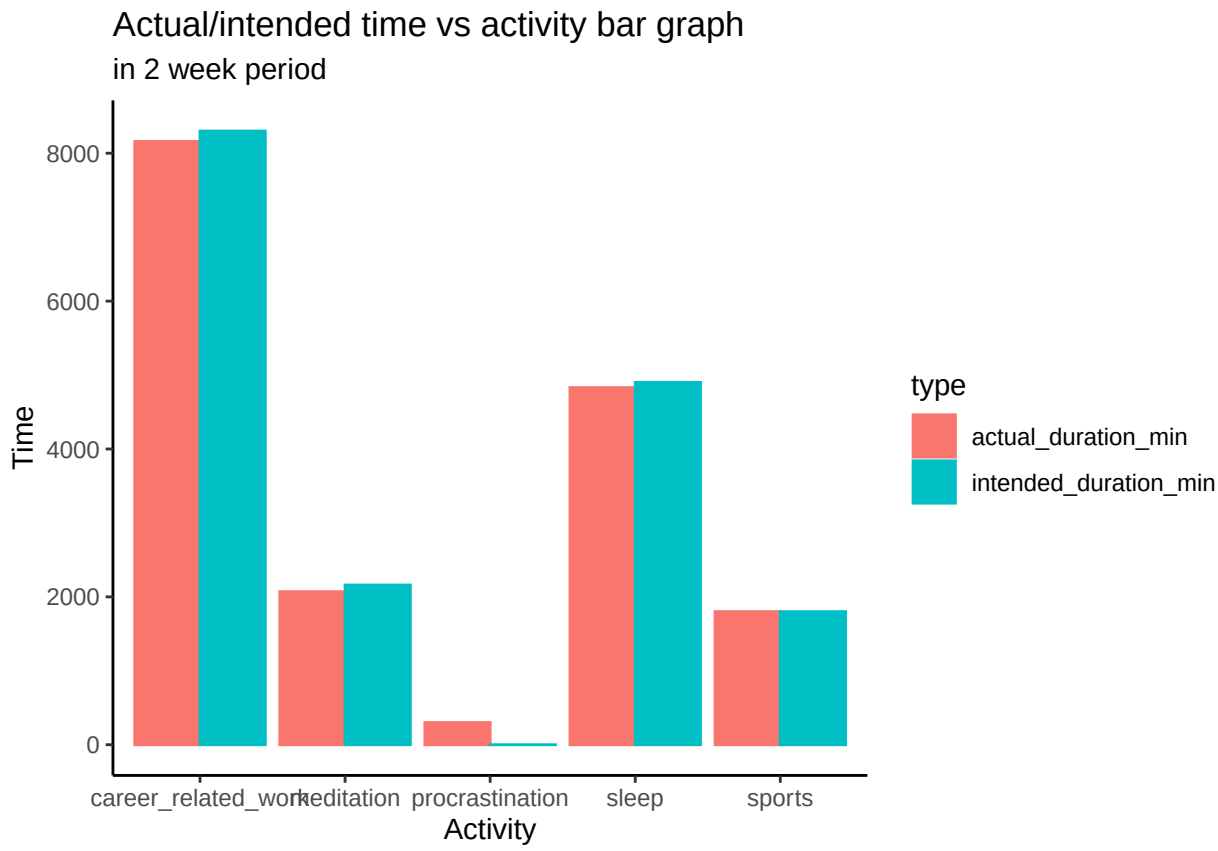
1. (bar graph, activity type on x-axis, time (both actual and intended) on y-axis, different colors for actual and intended)

This data visualization is in form of grouped bar graph with time on y-axis and activity names on x-axis. We have two types of times: `actual_duration_min` and `intended_duration_min` representing the actual time and intended time respectively. I used a color scheme to distinguish actual and intended time durations as mentioned in the legend of visualization. I have stacked the bars relating different types of times of same activity so that we can compare between the actual and intended times without any difficulty.

```
# Code for data wrangling to create an appropriate data set for visualization #1

# Creating a new data set to group the times just by activity
# Summarizing the total time consumed per activity over the data period.
time_activities <- activities %>%
  group_by(activity) %>%
  summarise("actual_duration_min" = sum(duration_min))

# Hardcoding a new data frame which stores the intended times for each activity
intended_time_activities <- data.frame(
  activity = c("career_related_work",
               "meditation",
               "procrastination",
               "sleep",
               "sports"),
  intended_duration_min = c(8300, 2160, 0, 4900, 1800))

# Joining both of the above tables by activity
# so as to have actual and intended times in same table
activities_full <- time_activities %>%
  inner_join(intended_time_activities, by = "activity")

# Dividing each activity's time field into two parts:
# actual and intended, so as to make the visualization
# Used pivot_longer to accomplish this task.
# Used "type" to store "actual" and "intended" types of times
# Used "time" to store values
activities_full_longer <- activities_full %>%
  pivot_longer(cols = actual_duration_min:intended_duration_min,
               names_to = "type",
               values_to = "time"
  )
# Data wrangling for visualization 1 ended.
```

```
# Code for creating data visualization #1

# Using `ggplot` to layout the data set and specifications
# Used `position = "dodge"`to join the actual & intended bars under activity
# Used labs to add labels to different components of the generated graph
  ggplot(data = activities_full_longer,
         mapping = aes(x = activity,
                       y = time,
```

```
                      color = type,
                      fill = type)) +
geom_col(position = "dodge")+
  labs(title = "Actual/intended time vs activity bar graph",
       subtitle = "in 2 week period",
       y = "Time",
       x = "Activity")
```



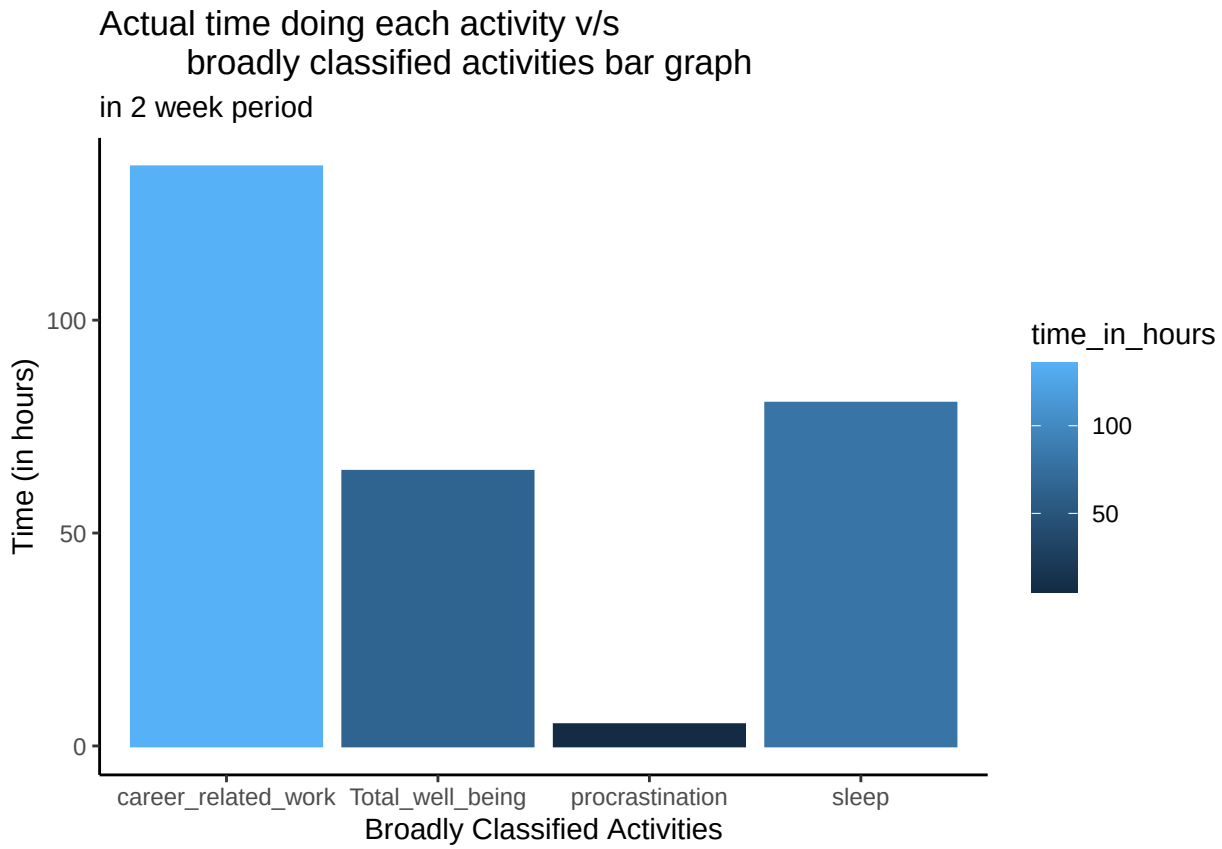
Describe what information is conveyed through data visualization #2 (that you'll create below) here.

2. (Bar chart, Y & color indicating actual time in hours, X indicating modified overarching activities)

This data visualization is in the form of bar chart with newly formed activities on x-axis and actual time in hours on y-axis. I have combined the activities sports and meditation into one overarching theme: `Total_well_being`. This facilitates a big-picture comparison in order to answer the questions posed by me as part of the calendar query assignment. In this graph we plot the actual times doing each activity in hours. Moreover, I am using the shades of blue color as a visual cue to quantify hours spent. This is intentionally implemented so that any average user can have better comparison between the far placed bars which can be deceiving by eye for some at first sight.

```
# Code for data wrangling to create an appropriate data set for visualization #2

# Using `fct_collapse` I made a new data set which has a row named
# `Total_well_being` which I define as `sports + meditation`
time_activities_new <- time_activities %>%
  mutate("new_activities" = fct_collapse(
    activity, "Total_well_being" = c("sports", "meditation"))
        )

# Again summarizing total time per activity and dropping possible NAs.
new_activities_dataset <- time_activities_new %>% group_by(new_activities) %>%
  summarize("time" = sum(actual_duration_min)) %>%
  drop_na()

# redefining time in hours for better relatability
hourly_new_activities_dataset <- new_activities_dataset %>%
  mutate(time_in_hours = time/60)
```

```
# Code for creating data visualization #2

# Using `ggplot` to layout the data set and specifications
# Used labs to add labels to different components of graph generated
ggplot(data = hourly_new_activities_dataset,
         mapping = aes(x = new_activities,
                       y = time_in_hours,
                       color = time_in_hours,
                       fill = time_in_hours)) +
  geom_col()+
    labs(title = "Actual time doing each activity v/s
         broadly classified activities bar graph",
         subtitle = "in 2 week period",
         y = "Time (in hours)",
         x = "Broadly Classified Activities")
```


Actual time doing each activity v/s
broadly classified activities bar graph
in 2 week period
Time (in hours)
100
50
0
career_related_work
Total_well_being
procrastination
sleep
Broadly Classified Activities
time_in_hours
100
50

Describe what information is conveyed through the table (that you'll create below) here.

The rows would be time spent doing that activity day-wise and the average time row additionally.

The columns will indicate different activities so that I can fill up the respective day-wise times on cells

```
# Code for table
final_table <- calendar_data %>%
  group_by(date, activity) %>%
  summarise(duration_min = sum(as.numeric(duration_min))) %>%
  mutate(avg_duration_min = mean(as.numeric(duration_min)))
```

To conclude, briefly summarize what you found in response to the questions posed here.

Question 1: Document intended time doing things (career_related_work, sleeping, sports, meditation, procrastination) versus actual time doing those things, and compare results.

From the data visualization #1 we can see that for each of the activities there are two bars placed near each other in a color coded fashion to represent the actual and intended times for each activity in terms of height of the bars measured along the y-axis. We can infer from the visualization #1 that I am doing a decent job of fulfilling my expectations in terms of career_related_work, sports, sleep and meditation. However, I am doing 330 mins of procrastination over the period of 2-weeks, which is way too high according to my expectation of 0 mins per week. Even if I devote that time to sleeping then also I will get 20+ mins of extra sleep everyday, which will in turn make my sleep to 6 hours each day. Moreover, we can see that the amount of time I spend doing meditation is comparable to that of sports, which is surprising to me because when I am meditating it seems that time passes too fast.

Question 2: Document time spent in (career_related_work) vs (total_well_being = sports + meditation) vs sleep

From the data visualization #2 we can see the actual time spent for each of the 3 big-picture terms I used namely `career_related_work`, `sleep`, and `Total_well_being`. Just for clarification, here by `Total_well_being` I mean total time spent in `sports` and `meditation`. We can notice that my bar for `Total_well_being` is comparable to that of `sleep`, but I have spent more time in the involuntary well being (sleep) than voluntary well being (meditation + sports), which is again expected. But, the interesting thing is there is a small chunk of time I WASTE in `procrastination`. If I devote that time to either meditation or sports or any other type of well_being which I can count under `Total_well_being`, then I would achieve a more proper balance between the voluntary and involuntary sides of well being. Moreover, I also noticed that as the semester progresses, I would need to increase my work week from near about 80 hours to 100 or 120 hours so as to accommodate aspects of career other than the regular academics. After the analysis of the plot, I can tell that I can juggle between `meditation`, `sleep` and `sports` to accommodate some more time for `career_related_work` so as to achieve a perfect work life balance where I can maintain both mental and physical peace.

Question 3: Document time spent in procrastination vs (total_well_being = sports + meditation)

From data visualization #2 we can tell that I am doing a decent job in turning towards the side of voluntary `total_well_being` without getting distracted to do `procrastination`. But, honestly I still think that I can further reduce the time I spend procrastinating in order to gain more mental and/or physical strength via `sports` and/or `meditation`. I don't think that `procrastination` is taking a major amount of time out of my schedule when in comes to the `career_related_work`, but I strongly believe that it is more productive to improve on actual skills and strengths which will be handy in life ahead.

REFLECTION ON THE CALENDAR QUERY PROJECT

In the data collection process, I sometimes was not able to add the exact time intervals in which I performed a given activity. Moreover, I have to enter every specifications in the exact name of variables every time, which was annoying at times and really vague for some times where I was doing 2 or more of the listed activities at the same time. In data analysis I faced problem in wrangling and rechecking the calendar for spelling errors in activities because in that case I will get more variables which would be cumbersome to handle as we will be very prone to manipulate data at the cost of accuracy. Also, as I am a pretty consistent person as you can tell, I just added some in class hours in repetitions on the basis of daily and weekly on selected days, which didn't show up when I imported the data, so I had to manually enter each time.

For future projects, I think I should try to enter each time slot individually as I complete the activity, so as tpo avoid the last time wrangling hassles. Moreover, I would highly consider reflecting upon my percentage involvement in different activities when I am doing them simultaneously and divide the time accordingly so as to minimize the sources of error.

I think as a college student in the semester system, the life changes drastically as the semester progresses. So, I would rather prefer to collect data over an entire semester so as to get a better sense of my time allocations and priorities while including different situations which I encounter like tournaments and midterms. I believe that it would be difficult especially to be consistent and accurate at the smae time while manually entering the data for such a long sample time without help of a second person in managing the accuracy of data by eliminating the bias.

As a client who provides data to the companies namely Facebook, Google and MapMyRun, I expect that my data is in a safe space where no one other than me can access, edit or delete it. This comes down to data privacy. I expect these companies to store my data in a very confidential manner such as an encryption, which is random and renderable only to the person who provides that data. The companies which made encryption, store data and provided functionality to render data should not be able to either access or modify the data without the consent of the user who provided the data.

As someone who analyses the data generated by others, I have a ethical responsibility to keep all the aspects of data confidential. Also, I have the responsibility to not alter the original data while performing my analysis because I do not have the copyright or legal ownership to that data without the written legal consent of the person to whom the data belongs.