Calendar Query Individual Project

STAT 231: Calendar Query

Dhyey Mavani, Class of 2025

Last updated December 4, 2021

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

```
# 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)))
```

Introduction, Data Collection Process and Nomenclature

As a college freshman, I was excited to explore how I am using my time when school is in progress. Hence, I used this project to get a sense of how I am managing my time and am I meeting my expectations. Specifically, I wanted to explore my schedule's time divide between the activities related to mental health, physical heath and studies in general. Thus, I collected the data and framed the following questions.

I used Google calendar for all of my time-keeping and I updated it as time progressed. I entered the specific time slots at different times throughout the day in the Google calendar on my iPhone for convenience. 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, the following are the variables which I used:

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 and compare intended time doing activities like **career related work, sleep, sports, meditation and procrastination** with actual time doing those activities, and compare results.

Question 2

Document and compare the actual time spent in performing activities like **career related work, total well being and sleep**

Question 3

Document and compare the actual time spent in **procrastination and total well being**

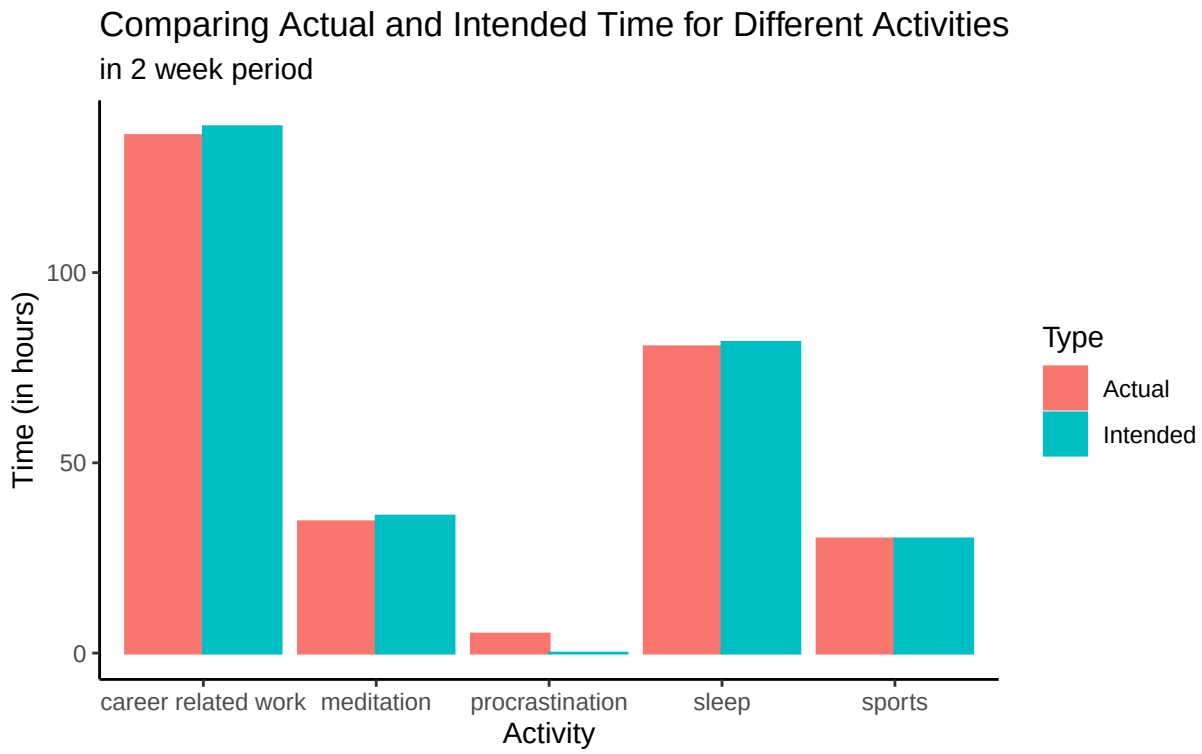
Data Visualization 1

This data visualization is in form of histogram with time on y-axis and activity type on x-axis. We have two types of times: the actual and intended time spent performing an activity respectively. I used a color scheme to distinguish actual and intended time duration as mentioned in the legend labelled as type. I have stacked the bars relating different types of times of same activity to aid us in comparing the actual and intended times with less visual or statistical 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
# storing the intended times spent for performing 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 %>%
  rename("Actual" = actual_duration_min,
         "Intended" = intended_duration_min) %>%
  pivot_longer(cols = Actual:Intended,
               names_to = "Type",
               values_to = "Time")

# stylizing names for better graph readability
activities_full_longer$activity[c(1,2)] <- c("career related work", "career related work")
```

```
# Code for creating data visualization #1
# Setting the height and width of the figure in the golden ratio
# 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/60,
```

```
                      color = Type,
                      fill = Type)) +
geom_col(position = "dodge")+
  labs(title = "Comparing Actual and Intended Time for Different Activities",
       subtitle = "in 2 week period",
       y = "Time (in hours)",
       x = "Activity")
```



We can see that my actual time doing the activities was always greater than the intended time doing that activity except the case of procrastination. Moreover, my actual and intended times for doing each activity were really close. As far as the extremes were concerned, we can notice that I spent most time doing career related work. Not surprisingly, I spent least time doing procrastination.

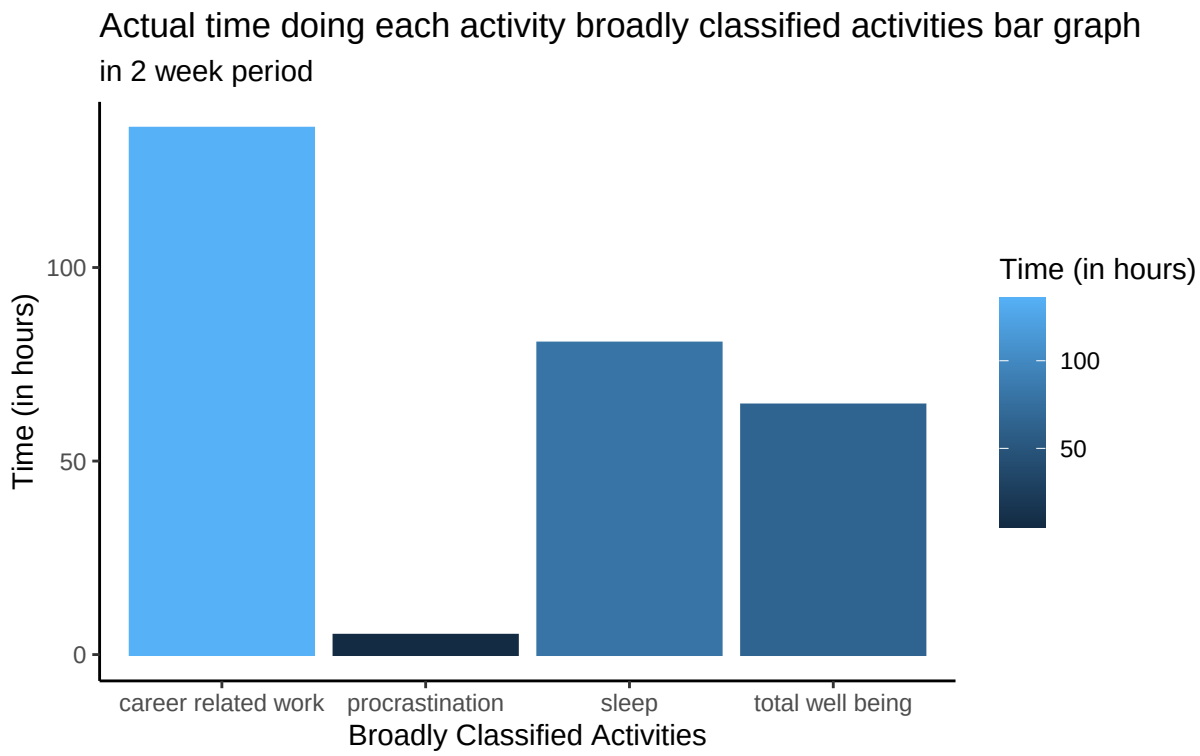
In addition to this, we can see that I spent almost double the time I spent sleeping in doing career related work. Finally, according to the figure below, I spent almost equal time meditating and playing sports. This indicates that I am managing my mental and physical well being in an acceptable manner.

Data Visualization 2

This data visualization is in the form of bar chart with updated activities on x-axis and actual time (in hours) on y-axis. This facilitated a big-picture comparison in order to answer the questions posed. In this graph I plotted the actual times I spent performing each activity in hours. Moreover, I used the shades of blue color as a visual cue to quantify hours spent. This is intentionally implemented so that any user can have better comparison between the bar pairs.

```
# 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
hourly_new_activities_dataset <- new_activities_dataset %>%
  mutate(time_in_hours = time/60)
```

```
# Code for creating data visualization #2

# stylizing names for better graph readability
hourly_new_activities_dataset$new_activities <-
  c("career related work","total well being","procrastination","sleep")

# 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 broadly classified activities bar graph",
         subtitle = "in 2 week period",
         y = "Time (in hours)",
         x = "Broadly Classified Activities",
         color = "Time (in hours)",
         fill = "Time (in hours)")
```



We can conclude that I spent comparable times in performing activities which come under total well being and sleeping. Moreover, time spent by me in procrastination when added to the same in activities leading to total well being can prove beneficial for me by allowing the time spent by me sleeping to balance out better with time spent in doing activities leading to total well being. Time spent in ascending order is: procrastination $<$ total well being $<$ sleep $<$ career related work

Finally, we can see that I spent almost double the time I spent sleeping in doing career related work.

Tabular Visualization

In the table, rows indicate day-wise time spent doing that activity and the average time row at last. The columns indicate different activities, so that we can look up the respective day-wise times devoted to certain activities in the cells.

We can conclude that the time devoted in sleeping for me is perfectly consistent statistically because the mean is same as the value for any specific day. Also, we can say that I am more consistent in meditation and sports compared to career related work and procrastination. To lend support to this argument, we can look at the standard deviation row in the table. It shows that my standard deviation in time spent doing career related work was the highest and the same while sleeping was the lowest.

Moreover, we can see the average time spent row for a big-picture estimate of time spent doing each activity per day. On an average, the descending order of time spent doing the given activity each day is: career related work > sleep > meditation > sports > procrastination.

```
# Modifying activities data set by removing all NAs
# Pivoting wider to move activity to column names and duration_min to cells
initial_table <- activities %>%
  drop_na() %>%
  pivot_wider(id_cols = NULL,
              names_from = activity,
              values_from = duration_min)

# Changing the column of dates to row names
intermediate_table <- column_to_rownames(initial_table, "date")

# Creating a new data table with a new row containing activity-wise mean and sd
mean_row <- summarize_all(intermediate_table, mean)
stddev_row <- summarize_all(intermediate_table, sd)
max_row <- summarize_all(intermediate_table, max)
min_row <- summarize_all(intermediate_table, min)

# Renaming the row
rownames(mean_row) <- "Average Time over 14 days"
rownames(stddev_row) <- "Standard Deviation over 14 days"
rownames(max_row) <- "Maximum Time over 14 days"
rownames(min_row) <- "Minimum Time over 14 days"

# Binding them together to create a final table
# Using kable to stylize the final table
final_summarized_table <- rbind(make.row.names = T,
                                intermediate_table,
                                mean_row,
                                stddev_row,
                                max_row,
                                min_row) %>%
  kableExtra::kable(digits = 0, format = "latex", booktabs = TRUE,
                    col.names = c("Time for Career Related Work (hours)",
                                  "Time for Meditation (hours)",
                                  "Time for Procrastination (hours)",
                                  "Time for Sleep (hours)",
                                  "Time for Sports (hours)")) %>%
  kableExtra::row_spec(row = c(15,16,17,18), bold = TRUE) %>%
  kableExtra::kable_styling(latex_options = c("striped", "scale_down"))
```

	Time for Career Related Work (hours)	Time for Meditation (hours)	Time for Procrastination (hours)	Time for Sleep (hours)	Time for Sports (hours)
2021-09-12	585	150	30	345	180
2021-09-13	600	150	15	345	120
2021-09-14	630	150	30	345	120
2021-09-15	600	150	15	345	120
2021-09-16	630	150	30	345	120
2021-09-17	600	150	15	345	120
2021-09-18	600	120	15	345	120
2021-09-19	585	150	30	345	180
2021-09-20	615	150	15	345	120
2021-09-21	630	150	30	345	120
2021-09-22	480	150	15	345	120
2021-09-23	630	150	30	345	120
2021-09-24	480	150	15	345	120
2021-09-25	495	150	15	345	120
Average Time over 14 days	**583**	**148**	**21**	**345**	**129**
Standard Deviation over 14 days	**55**	**8**	**8**	**0**	**22**
Maximum Time over 14 days	**630**	**150**	**30**	**345**	**180**
Minimum Time over 14 days	**480**	**120**	**15**	**345**	**120**

```
# Printing final table
final_summarized_table
```

Answering Posed Questions

Question 1: Document and compare intended time doing activities like career related work, sleep, sports, meditation and procrastination with actual time doing those activities, and compare results.

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 5.5 hours of procrastination over the period of 2-weeks, which is way too high according to my intended value of 0 hours. If I devote that time to sleep, then I can mange to get more than 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 spent doing meditation is comparable to that of sports, which is surprising to me because when I am doing meditation it seems that time passes too fast.

Question 2: Document and compare the actual time spent in performing activities like career related work, total well being and sleep

We can notice that bar height for total well being is comparable to that of sleep. But, on a closer look we can conclude that I have spent more time in the involuntary well being (sleep) than voluntary well being (total well being), which is again expected. The interesting thing is there is a small chunk of time I waste in procrastination, which can be considered as a piece taking away some time from my well being. If I devote that time to any type of voluntary 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 and compare the actual time spent in procrastination and total well being

From data visualization 2 we can tell that I am doing a decent job in spending time in activities which count under 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. From the big picture, 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 further reduce procrastination.

Reflection on the Calendar Query Project

Write your one-page reflection here in paragraph form. In particular, address:

- What difficulties in the data collection and analysis process did you encounter? Identify two of your main hurdles in gathering accurate data.
- What implications does that have for future data collection and/or analysis projects?
- How much data do you think you'd need to collect in order to answer your question(s) of interest? Would it be hard to collect that data? Why or why not?
- As someone who provides data, what expectations do you have when you give your data (e.g. to Facebook, Google, MapMyRun, etc.)?
- As someone who analyzes others' data, what ethical responsibilities do you have?

Answer:

Sometimes during the data collection process I was not able to add the exact time intervals in which I performed a given activity. Moreover, I found the need to enter every specification every time very intimidating and really vague especially when I was doing 2 or more of the listed activities at the same time. In data analysis, I faced problems in wrangling and rechecking the calendar for spelling errors in activities to avoid unwanted splitting of variable data. 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 everything again, which made things even worse.

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

As far as the amount of data collected for my calendar query as a college student in the semester system is concerned, I think my life changes significantly as the semester progresses. So, I would rather prefer to collect data over an entire semester in order 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 to be consistent and accurate at the same 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.

In terms of Data Ethics, 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 using technologies such as encryption, which is random and accessible only to the person to whom the encrypted data belongs. The companies which make encryption, store data, and provide 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.

Apart from my expectations from these companies, 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 of that data without the written legally-authorized consent of the person to whom the data belongs.