Calendar Query Individual Project

STAT 231: Calendar Query

Dhyey Mavani, Class of 2025

Last updated September 24, 2021

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

```r
# 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 dataset 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 data with Google calendar in `ical` format. The below mentioned questions have variables defined according to the name of activity which represents the time spent doing that activity. I defined an additional variable called `Total_well_being` which 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, sleeping, 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. As we have two types of times namely `actual_duration_min` and `intended_duration_min`, I am using a color scheme to distinguish them as mentioned in the legend of visualization, which is titled *"type"*. I have stacked the bars relating different types of times with the same activity so that user 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
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.
```

```r
# 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 graph generated
  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")
```


Actual/intended time vs activity bar graph
in 2 week period
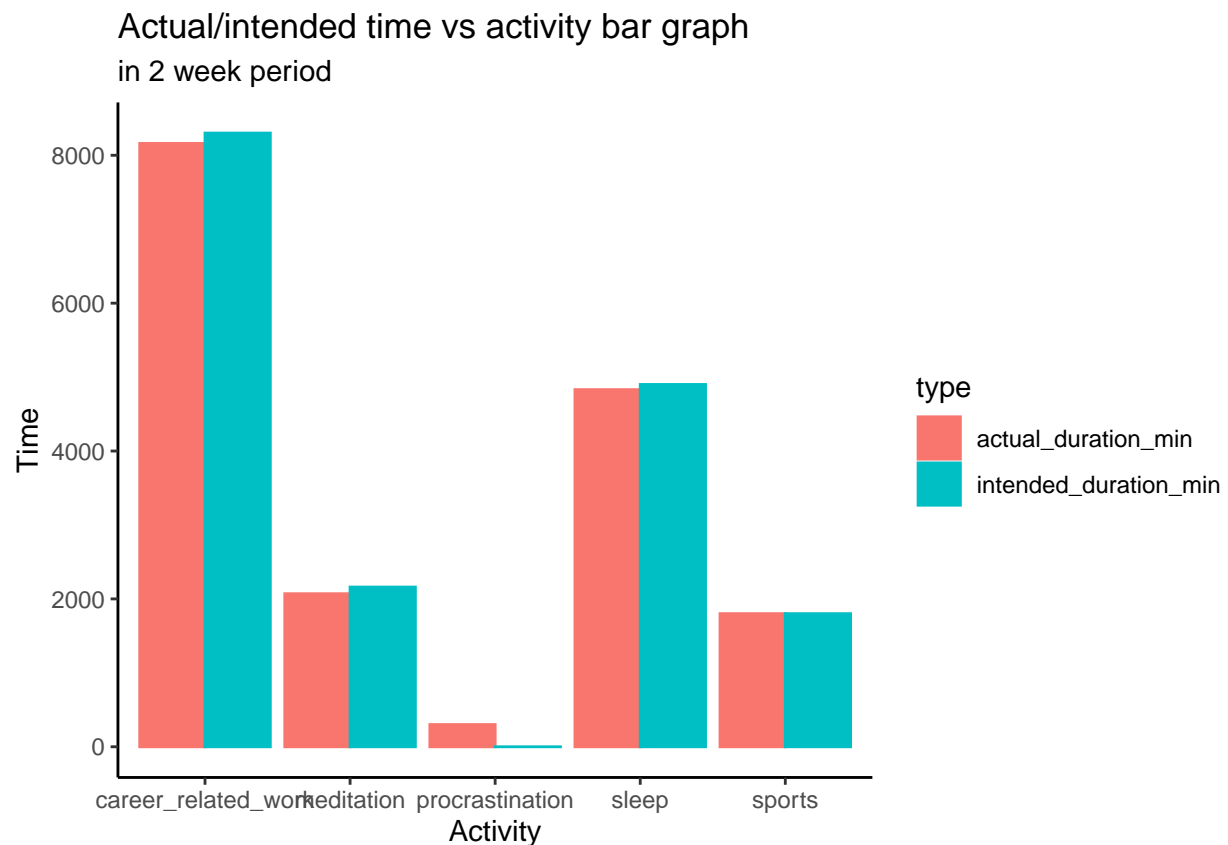
Describe what information is conveyed through data visualization #2 (that you'll create below) here.

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 activities on x-axis and time on y-axis. I combined the activities sports and meditation into one overarching theme called `Total_well_being` so as to allow a more big picture comparision in order to answer the questions posed by me. In this graph we plot the actual times doing each activity in hours. Moreover, as a visual cue I am using the shades of blue color to signify the quantity of hours spent, so that user can have better comparison between the far placed bars which can be decieving by eye for some.

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

# Using `fct_collapse` I made a new dataset 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
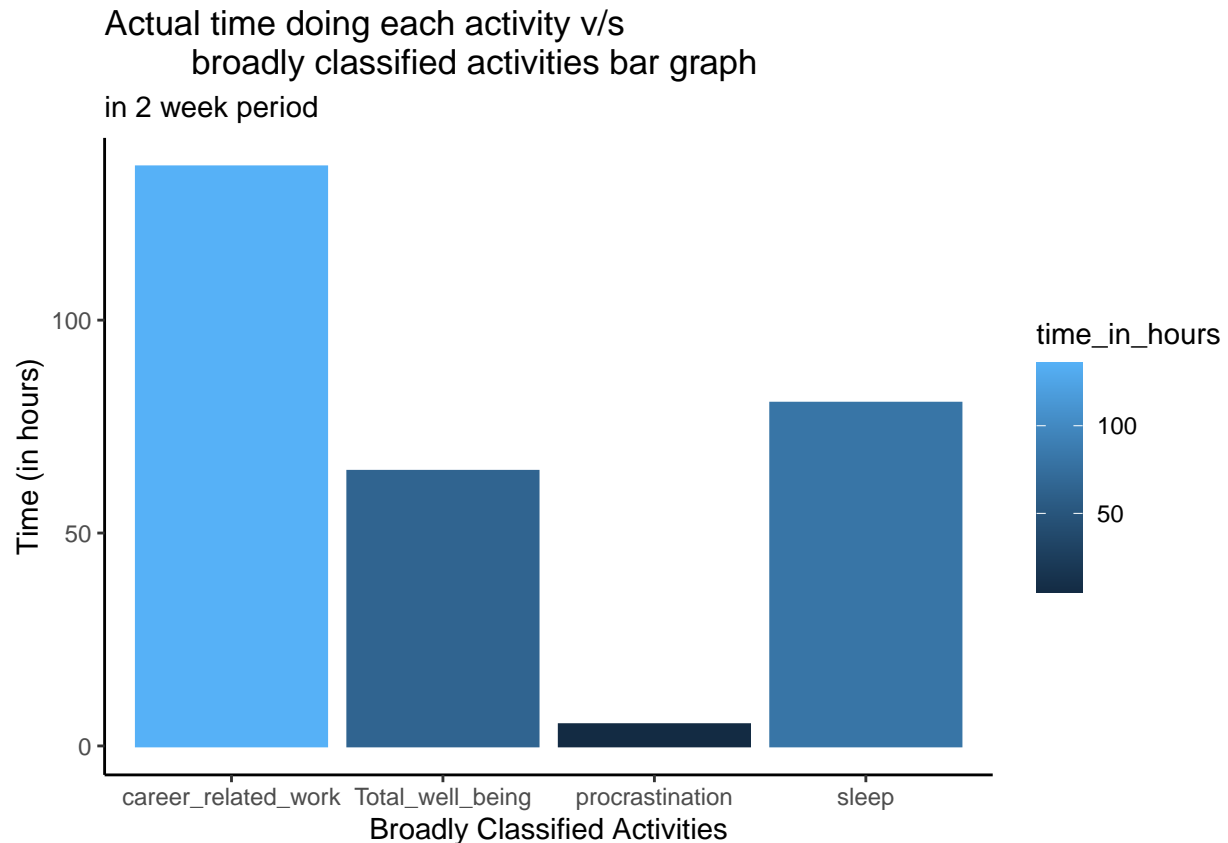
Describe what information is conveyed through the table (that you'll create below) here.

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
calendar_data
```

A tibble: 135 x 12

uid summary start end description
1 1969-12-31 19:00:00 1969-12-31 19:00:00
2 5279ehovvajkb~ sleep 2021-09-12 00:00:00 2021-09-12 05:45:00 ""
3 3303864B-1AB2~ career_re~ 2021-09-20 21:00:00 2021-09-20 23:00:00 ""
4 3jrjc9b5icdck~ sleep 2021-09-13 00:00:00 2021-09-13 05:45:00 ""
5 3acldjnvrspil~ sleep 2021-09-15 00:00:00 2021-09-15 05:45:00 ""
6 552ji0et5aa1c~ sleep 2021-09-14 00:00:00 2021-09-14 05:45:00 ""
7 04cmogrcfbrtg~ sleep 2021-09-16 00:00:00 2021-09-16 05:45:00 ""
8 47sghpbd99n0f~ sleep 2021-09-17 00:00:00 2021-09-17 05:45:00 ""
9 4q46qjb7i6upv~ sleep 2021-09-18 00:00:00 2021-09-18 05:45:00 ""
10 0jsvndalib60v~ sleep 2021-09-19 00:00:00 2021-09-19 05:45:00 ""

... with 125 more rows, and 7 more variables: last.modified , # status , start_datetime , end_datetime , # duration_min , date , activity

```
activities
```

A tibble: 71 x 3

Groups: date [15]

date activity duration_min 1 1969-12-31 00:00:00 0 2 2021-09-12 00:00:00 career_related_work 585 3 2021-09-12 00:00:00 meditation 150 4 2021-09-12 00:00:00 procrastination 30 5 2021-09-12 00:00:00 sleep 345 6 2021-09-12 00:00:00 sports 180 7 2021-09-13 00:00:00 career_related_work 600 8 2021-09-13 00:00:00 meditation 150 9 2021-09-13 00:00:00 procrastination 15 10 2021-09-13 00:00:00 sleep 345 # ... with 61 more rows

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.

yada yada yada

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

yada yada yada

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

yada yada yada

REFLECTION ON THE CALENDAR QUERY PROJECT