

# Does the day of the week affect my sleeping and eating habits?

Stat231: Google Calendar Report

Majd Rouhana

Due Friday, September 25 by 5:00 PM EST

## How do I spend my time?

```
# reading in the data
path <- "/home/class22/mrouhana22/git/majd_STAT231/projects"
filename <- "STAT231_c_pl4hgkq7kd4cehd0u83k85aja4@group.calendar.google.com.ics"

# mutating the data to include certain variables
my_calendar0 <- ical_parse_df(file = paste0(path,"/",filename)) %>%
  mutate(start_datetime = with_tz(start, tzzone = "America/New_York")
    , end_datetime = with_tz(end, tzzone = "America/New_York")
    , length_hour = end_datetime - start_datetime
    , date = floor_date(start_datetime, unit = "day")
    , Day = weekdays(as.Date(date)))

# putting `Day` in order of the days of the week
my_calendar0$Day <- factor(my_calendar0$Day,
  levels = c("Sunday", "Monday",
    "Tuesday", "Wednesday",
    "Thursday", "Friday",
    "Saturday"))

# creating and wrangling a dataset where I only analyze my sleeping patterns
# I filtered the data to only include the summaries about sleeping
# I then found the average hrs of sleep per night
sleepCalendar <- my_calendar0 %>%
  filter(summary == "Sleep") %>%
  group_by(Day) %>%
  summarize(meanSlp = round(mean(length_hour), digits = 2))

## `summarise()` ungrouping output (override with `.groups` argument)

# creating and wrangling a dataset where I only analyze my eating patterns
# I first created the `eat` variable so I could later filter out all summaries that
# weren't related to consuming calories.
# I created the `numDay` variable so that I could later divide the counts of
# the times I consumed by the number of times I logged for each day.
# I subsetted the data to only look at the variables I'm interested in.
# I sliced the data to remove duplicate rows.
eatCalendar <- my_calendar0 %>%
  mutate(eat = ifelse((summary == "Coffee" | summary == "Lunch" |
    summary == "Dinner" | summary == "Snack"),
```

```

      "Eating", "Not Eating"),
    numDay = ifelse((Day == "Friday" | Day == "Saturday"), 3, 2)) %>%
  filter(eat == "Eating") %>%
  subset(select = c(eat, Day, numDay)) %>%
  group_by(Day) %>%
  summarize(meanEat = round(n() / numDay, digits = 2)) %>%
  arrange(meanEat) %>%
  slice(from = 1, to = 12)

```

```
## `summarise()` regrouping output by 'Day' (override with `.groups` argument)
```

```

# creating a dataset to use for my table
# merges both the sleep and eating pattern calendars
mergedCalendar <- merge(sleepCalendar, eatCalendar)

```

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

I am interested in analyzing if the day of the week affects my sleeping and/or eating habits. In particular, I was interested to see if there's an overlap between the days I sleep less and the days I eat more. I defined my variables of interest as the hours I slept each night and the amount of times I consumed calories a day (could be anything with calories, including drinks, snacks or even gum; I only recorded the times I sat down to consume, not the amount of items consumed). I collected this data by using my Google calendar; at the end of every day, I would try to log the hours I slept the previous night and the amount of times I consumed calories that day. For the times I consumed calories, I specified whether it was coffee, lunch, dinner or a snack (I ended up not chewing any gum or having any sugary drinks outside an occasional meal). For the times I consumed calories, I did not specify the amount of time it took me to consume, as I wasn't interested in that aspect of data. As a result, I left the duration as Google Calendar's default of one hour.

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

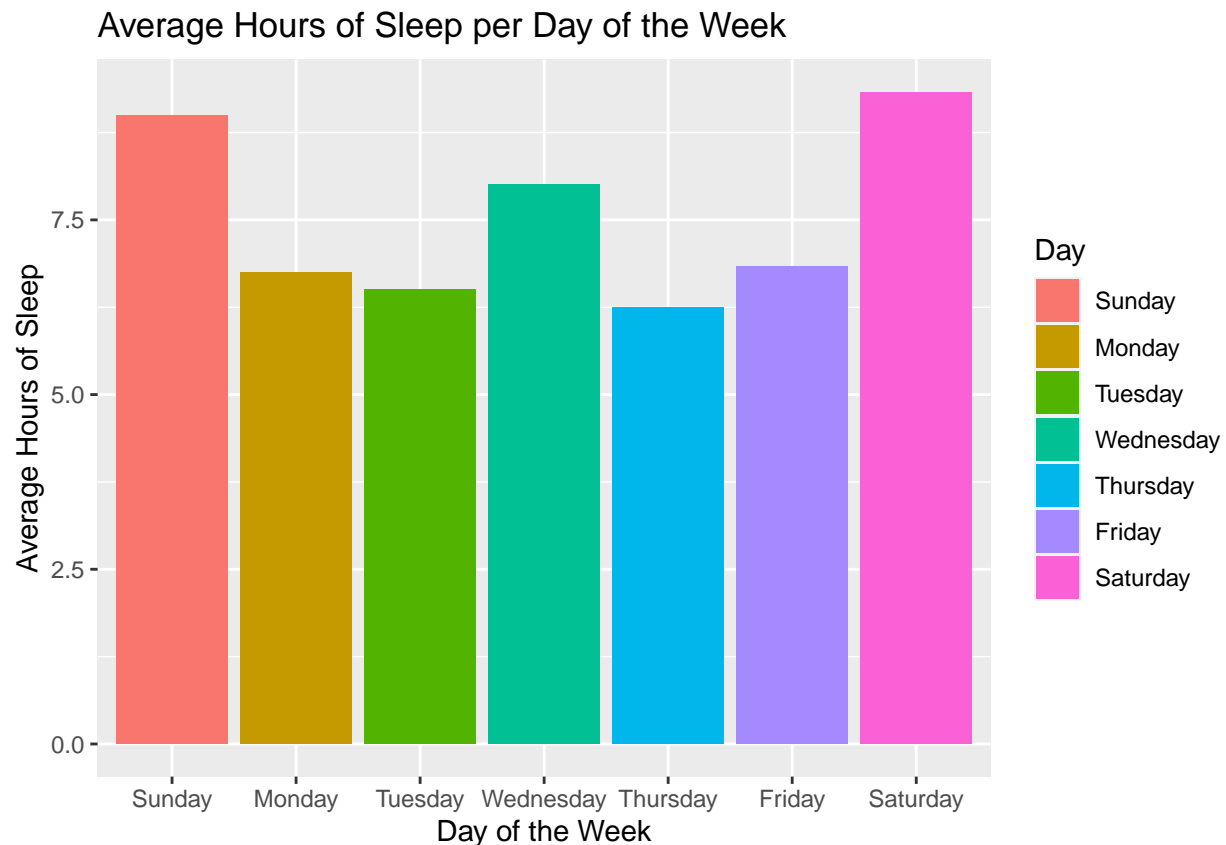
My first data visualization is a bar chart that analyzes the average hours of sleep I get per day of the week. From the chart, it appears that I get the most sleep on the weekends, while the least sleep on weekdays. In particular, I got the most sleep on Saturday, and the least sleep on Thursday. This definitely makes sense to me; I typically am able to sleep in on Saturdays as I usually have nothing to do in the morning, and I sleep less on Thursday nights because I always have a lot of work due for Friday.

```

# creating a bar chart that analyzes my avg. hrs. of sleep per day of the week
ggplot(data = sleepCalendar,
  aes(x = Day, y = meanSlp, fill = Day)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Hours of Sleep per Day of the Week",
    y = "Average Hours of Sleep",
    x = "Day of the Week")

```

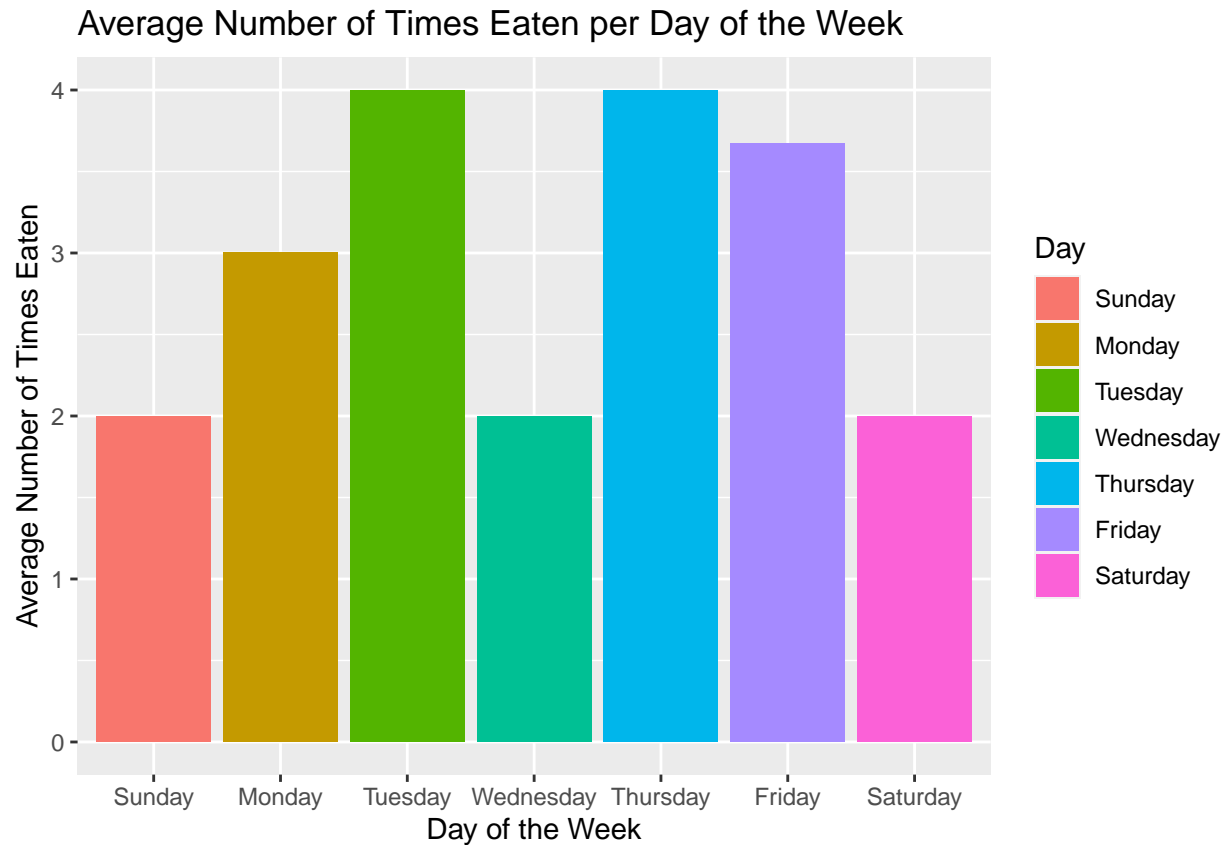
```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



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

My second data visualization is a bar chart that analyzes the average number of times I eat/consume calories per day of the week. From the chart, it appears that I consume calories the most number of times on Tuesdays and Thursdays, while the least on Saturday and Sunday. This seems to somewhat overlap with my sleep patterns, indicating a possible inverse relationship. The days I slept the least (Tuesday and Thursday) seem to correspond with the days that I consumed calories the most number of times, and the days I slept the most (Saturday and Sunday) correspond with the days I consumed the least amount of times.

```
# creating a bar chart that analyzes my avg. # of times eaten per day of the week
ggplot(data = eatCalendar,
  aes(x = Day, y = meanEat, fill = Day)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Number of Times Eaten per Day of the Week",
    y = "Average Number of Times Eaten",
    x = "Day of the Week")
```



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

The table below combines the data shown in both charts above to make it easier to read and compare numerically. From the table, we can see more easily that there seems to be a present trend; on days I sleep less I tend to consume more times, and on days I sleep more I tend to consume less.

```
# creating a table to see a relationship between the days
# of the week and my sleep & calorie consumption patterns
knitr::kable(mergedCalendar,
  col.names = c(
    "Day of the Week",
    "Average Hours of Sleep",
    "Average Number of Times Eaten"))
```

Day of the Week	Average Hours of Sleep	Average Number of Times Eaten
Friday	6.83 hours	3.67
Monday	6.75 hours	3.00
Saturday	9.33 hours	2.00
Sunday	9.00 hours	2.00
Thursday	6.25 hours	4.00
Tuesday	6.50 hours	4.00
Wednesday	8.00 hours	2.00

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

From both data visualizations and the table, it appears that on days I sleep less I tend to consume more times a day, and on days I sleep more I tend to consume less. Also, I tended to sleep more on the weekends,

and sleep less on weekdays (particularly Tuesday and Thursday, most likely due to a higher workload). This tells me that sleeping well is important to staying healthy, losing weight and resisting the temptation to have a late-night snack. It also tells me that time management goes hand-in-hand with my sleep. If I want to sleep well, I need to make sure to get my work done on time so I'm not sacrificing sleep to finish it all in one night. In addition, on days I slept less I tended to drink coffee. While I did not represent the meal categories in either of my visualizations or my table, it is something I noticed while looking at my Google calendar. This also indicates that I might be somewhat dependent on coffee when I don't get a good night's rest. Drinking coffee also adds to my "number of times I consumed calories" variable, which could be a possible indicator to why the number of times I consumed calories a day is higher on days I sleep less. Ultimately, my dataset is relatively small, so it's difficult to simply confirm any of my previous claims as fact. If we wanted more conclusive evidence, we must collect a larger set of data, conduct statistical tests for significance, and then repeat the process multiple times.

## Reflection

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?

This assignment was pretty interesting to me! While I'm not surprised by some of the results I found, I learned a lot through wrangling and working with data I collected myself.

One aspect of data collection that proved to be difficult was measuring the exact amount of time I slept. This is because it's hard to tell exactly when you fall asleep. To work around this, I estimated/rounded to the nearest 30 minutes of what I thought I slept each night. As a result, some of my estimates may be too high, while others may be too low. Another hurdle I had to overcome was my own human forgetfulness. For some days, I would forget to log my data. Because of this, my memory of each variable of interest may not have been as accurate as if I had just done it. Thus, I may have under- or overestimated the amount of times I consumed calories or the amount of hours I slept. This implies that data collection and analysis may not always be perfect.

In the future, I need to keep in mind that there's always room for error in my projects and in others'. Every data collection process or analysis project that I have been involved with or will be involved with in the future has possibility for error, so I must always keep in mind to address this somewhere within my studies. Withholding this information may deceive those who are not as experienced in the field of statistics due to their lack of knowledge of the data collection process.

To effectively answer my question of interest, I would need at least 30 weeks of data. This is because a sample is considered "large enough" when there are at least 30 observations. Since I'm interested in each day of the week, I need to collect 30 weeks of data to achieve this. The data would be easy to collect, but the more data I collect the more possibilities of error there are for me to make. With more things I need to log, the chance of me mistyping, inaccurately logging data or forgetting to log data rises.

As someone who provides data, I expect that my data is accurately obtained and logged when given. On most websites, my given data is logged by a computer, which eliminates the possibility of human error besides my own. However, I do expect my given data to adhere to each website's own privacy policy and terms of service. Since I am agreeing to them, I am consenting to websites using and selling my data. If a website had no such terms, I would certainly object to the usage of my own data.

As someone who analyzes others' data, I have an ethical responsibility to adhere to the rules I set and inform to others. I must make sure before I proceed in any study that the data I've been given by others has been given voluntarily and with consent to be used in the setting I outline. Using someone's own data without their consent is unethical and an infringement on their rights.