Calendar Assignment: How Do I Spend My Time?

Stat231: Google Calendar Report

Angelica Kim

Due Friday, September 25 by 5:00 PM EST

How do I spend my time?

```
library(tidyverse)
```

```
## -- Attaching packages ------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ---------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ical)
library(xtable)

path <- "/Users/angelica/Library/Mobile Documents/com~apple~CloudDocs/Fall 2020/STAT 231 Data Science/ha
filename <- "haikim20@amherst.edu.ics"
my_calendar <- ical_parse_df(file = paste0(path,"/",filename)) %>%
  mutate(start_datetime = with_tz(start, tzone = "America/New_York"),
         end_datetime = with_tz(end, tzone = "America/New_York"),
         length_sec = end_datetime - start_datetime,
         date = floor_date(start_datetime, unit = "day"))
```

```r
my_calendar2 <- my_calendar %>%
  filter(date >= "2020-09-01") %>%
  #convert the activity names to lower case letters
  mutate(summary2 = tolower(summary),
         #convert time length from seconds to hours
         length_hr = length_sec/3600,
         #create a day variable based on dates
         day = case_when(day(date)%%7 == 0 ~ 'Monday',
                         day(date)%%7 == 1 ~ 'Tuesday',
                         day(date)%%7 == 2 ~ 'Wednesday',
                         day(date)%%7 == 3 ~ 'Thursday',
                         day(date)%%7 == 4 ~ 'Friday',
                         day(date)%%7 == 5 ~ 'Saturday',
                         day(date)%%7 == 6 ~ 'Sunday')) %>%
    #remove unwanted variables
    select(-c(uid, description, last.modified, status, start, end, length_sec, summary))
```

Data sets for bar plot

```r
#amount of time I actually have
free <- my_calendar2 %>%
  filter(summary2 == "sleep" | summary2 == "class" ) %>%
  group_by(date, day) %>%
  summarize(freetime = 24-sum(length_hr))
```

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

```r
#time I spent for studying
study <- my_calendar2 %>%
  filter(summary2 == "study") %>%
  group_by(date, day) %>%
  summarize(studytime = sum(length_hr))
```

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

```r
#create a vector with days in the order of Monday through Sunday
x <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")

#data set that shows average time I have, average studying time, ratio of average studying time to aver
my_calendar3 <- free %>%
  left_join(study) %>%
  #sorts the day variable in the order of Monday through Sunday
  mutate(day = factor(day, levels = x),
         studytime = case_when(is.na(studytime)~0, TRUE~as.numeric(studytime))) %>%
  group_by(day) %>%
  summarize(free = mean(freetime),
            study = mean(studytime)) %>%
  arrange(day)
```

```
## Joining, by = c("date", "day")
```

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

```
#convert the data set into a narrow format
my_calendar4 <- my_calendar3 %>%
  gather(type, length_hr, free, study)
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

Data sets for boxplot

```
my_calendar6 <- free %>%
  left_join(study) %>%
  mutate(studytime = case_when(is.na(studytime)~0, TRUE~as.numeric(studytime)),
         percentStudy = as.numeric(studytime)/as.numeric(freetime))
```

```
## Joining, by = c("date", "day")
```

```
meal <- my_calendar2 %>%
  #create a new column to distinguish whether I cooked at home or ate out
  separate(summary2, into = c("activity1", "activity2")
           , sep = "[()]"
           , remove = FALSE) %>%
  #fix a misspelled row
  mutate(activity2 = case_when(startsWith(activity2, "o") ~ "outside",
                   TRUE ~ activity2)) %>%
  #include only the rows related to eating activity
  filter(activity1 %in% c("breakfast", "lunch", "dinner")) %>%
  #reorder columns
  select(date, day, activity1, activity2) %>%
  #merge with 'study' data frame that contains studying time
  left_join(my_calendar6)
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 56 rows [1, 12, 15,
## 16, 22, 23, 24, 25, 26, 28, 30, 31, 35, 41, 50, 52, 53, 98, 102, 104, ...].

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 174 rows [2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 17, 18, 19, 20, 21, 27, 29, 32, ...].

## Joining, by = c("date", "day")
```

Data sets for the table

```
study2 <- my_calendar3 %>%
  mutate(avgpercentStudy = as.numeric(study)/as.numeric(free)) %>%
  select(day, avgpercentStudy)

my_calendar5 <- my_calendar2 %>%
```

```
  filter(summary2 == "sleep") %>%
  group_by(day) %>%
  mutate(day = factor(day, levels = x)) %>%
  summarize(sleeptime = as.numeric(mean(length_hr))) %>%
  #split sleeping time into three levels
 # mutate(sleep_level = cut(sleeptime, c(0, 6, 8, 20), labels=c('Less', 'Normal', 'More')))
  left_join(study2)
```

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

```
## Joining, by = "day"
```

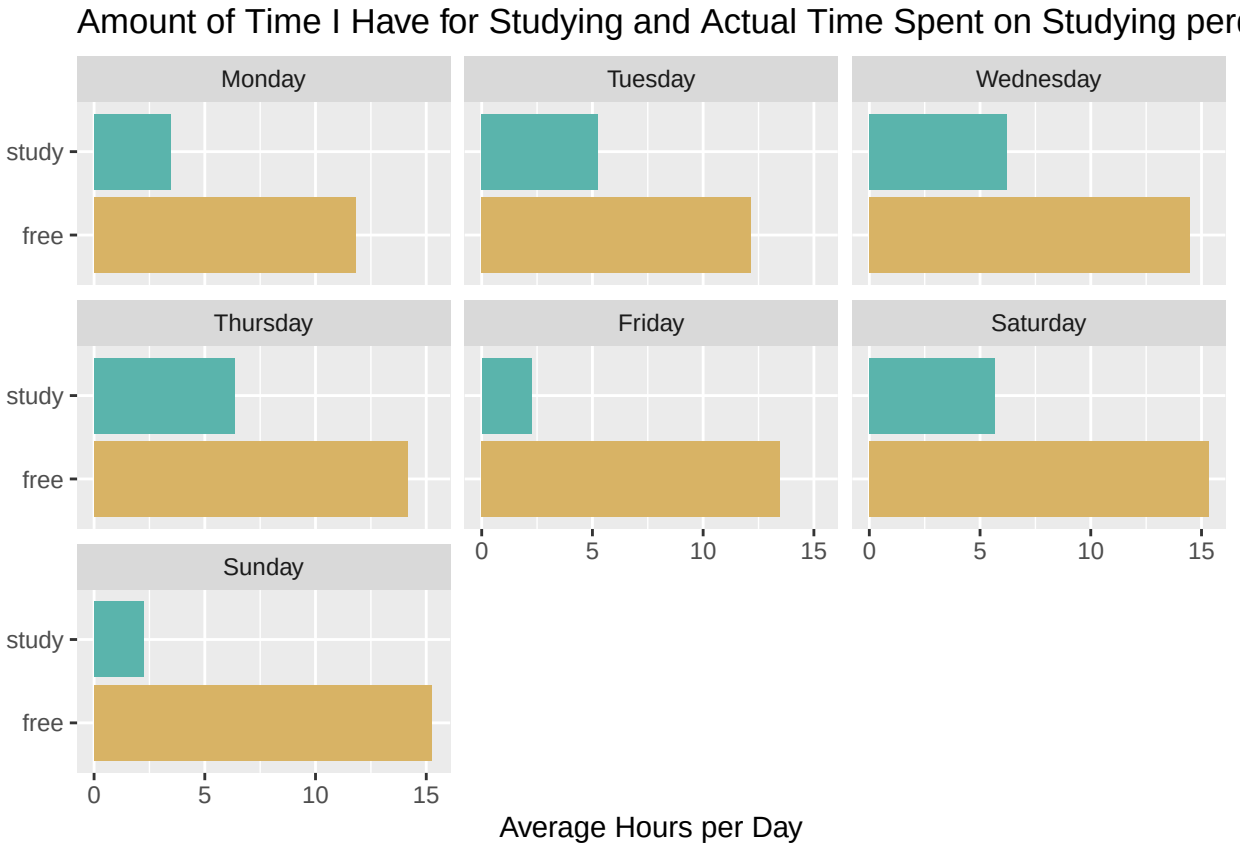
My questions of interest are following: 1. How much do I procrastinate?–difference in time I have for studying and the actual time I spend on it. 2. How does my sleep schedule affect productivity? 3. How much time do I spend on cooking and eating meals? Does spending time on cooking affect my studying time?

I used Google Calendar to explore how I spend my time for 32 days. To address the questions of interest above, I only kept track of a few activities that take up most of my time during the day–studying, class, eating and sleeping. At the end of approximately one month period, I collected my data for wrangling and analysis.

Among the variables in the original data set, I am focusing on the summary, which is converted to lower case letters for the ease of analysis, date of an activity, its starting time and date, ending time and date and length of time converted to hours. I also added the day variable which expresses day of the week of each date.

Visualization 1

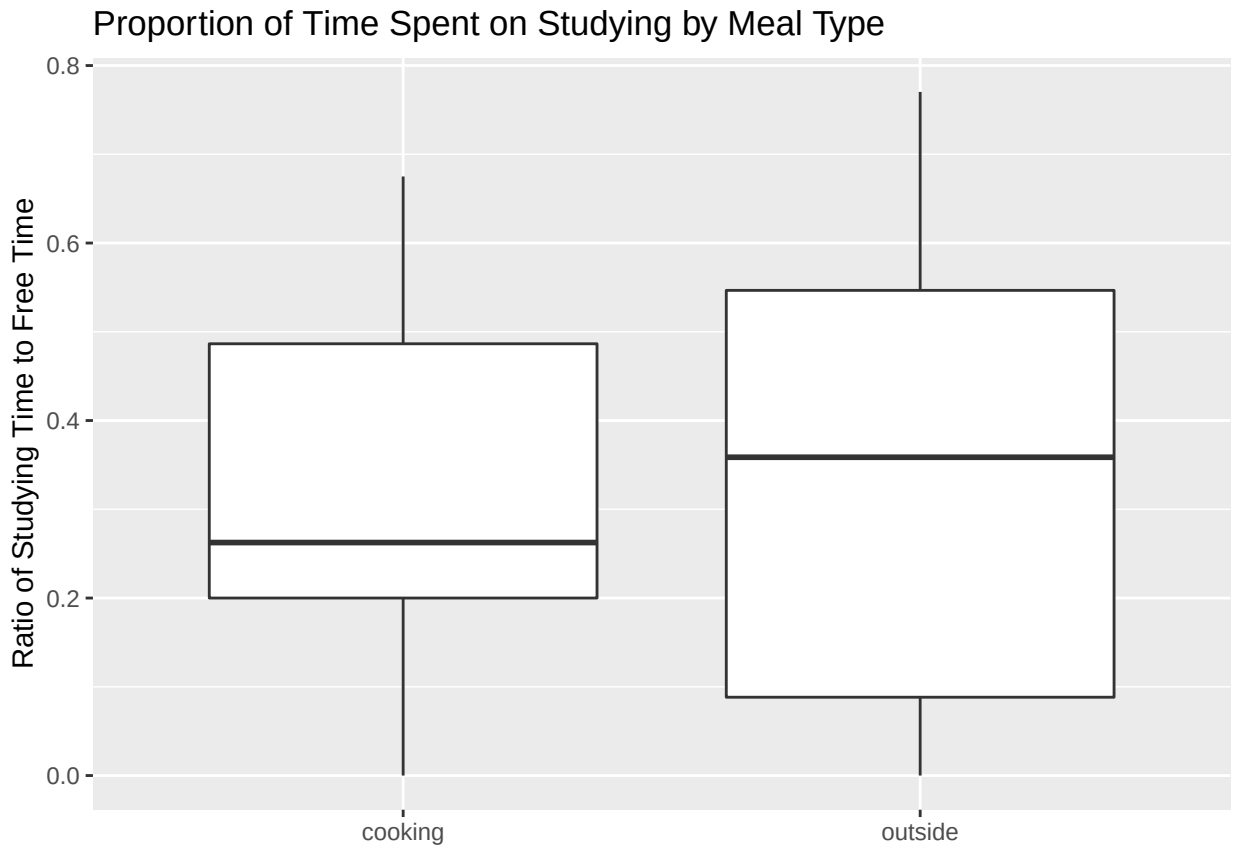
```
ggplot(my_calendar4, aes(x=type, y=length_hr, fill = type)) +
  geom_bar(stat = "identity") +
  facet_wrap(~day) +
  scale_fill_manual(values = c("free" = "#d8b365",
                               "study" = "#5ab4ac")) +
  coord_flip() +
  theme(legend.position="none") +
  xlab(NULL) +
  ylab("Average Hours per Day") +
  ggtitle("Amount of Time I Have for Studying and Actual Time Spent on Studying perday")
```



The bar plot compares the average amount of time I have for studying and the average amount of time I allot on studying over the week. I tend to spend more time studying on Wednesdays and Thursdays. This is presumably because all my classes have problem sets due on Fridays and I spend more time working on them especially on Thursdays.

Visualization 2

```
ggplot(data = meal, mapping = aes(x = activity2, y = percentStudy)) +
  geom_boxplot() +
  xlab(NULL) +
  ylab("Ratio of Studying Time to Free Time") +
  ggtitle("Proportion of Time Spent on Studying by Meal Type")
```




```
mosaic::favstats(percentStudy~activity2, data = meal)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
##   activity2 min         Q1    median        Q3       max      mean        sd  n
## 1   cooking   0 0.20000000 0.2625000 0.4864865 0.6750000 0.3287966 0.1819324 26
## 2   outside   0 0.08823529 0.3586207 0.5466687 0.7702703 0.3395368 0.2410846 30
##   missing
## 1       0
## 2       0
```

In this boxplot, I plotted the ratio of studying time to time I actually have over two meal types—whether I cooked at home or ate out. It appears that I tend to allot more time on studying when I eat outside than when I cook. The median ratio of studying time for eating outside is 0.358 and 0.263 for cooking. This result is quite intuitive as it takes more time to cook and prep for meals than to eat outside and I would have less time to study if I cook.

Table

I created this table to explore how my sleep schedule affects productivity. The table contains average sleeping time and average ratio of studying time to time I actually have over the week. I expected that sleeping more would enhance my productivity. However, the table shows that I studied the most when I slept the least. It

is also notable that I sleep the most and study the least on Fridays. Again, this is possibly because I have most of my homework due on Fridays; I dedicated more time to sleeping than any other days of the week after submitting all the homework.

```
x <- xtable(my_calendar5)
x
```

% latex table generated in R 4.0.2 by xtable 1.8-4 package % Thu Oct 8 01:55:54 2020

	day	sleeptime	avgpercentStudy
1	Monday	6.17	0.29
2	Tuesday	6.88	0.43
3	Wednesday	6.75	0.43
4	Thursday	5.75	0.45
5	Friday	8.88	0.17
6	Saturday	5.79	0.37
7	Sunday	8.75	0.15

Answering my questions of interest

1. How much do I procrastinate?–difference in time I have for studying and the actual time I spend on it. I spend way less time on studying than time I have for studying. I spend time on studying the most on Wednesdays and Thursdays.

2. How much time do I spend on cooking and eating meals? Does spending time on cooking affect my studying time? I spend more time on on studying when I eat outside than when I cook for my own meals.

3. How does my sleep schedule affect productivity? I tend to study more when I sleep less. On Thursdays, I sleep the least, 5.75 hours on average, and study the most, dedicating 45% of my time to studying.

Reflection

This project provided valuable insight into how I spend my time. Realistically, however, it was hard to keep a detailed track of how I spend time. When collecting the data, I focused only on a few activities—classes, sleeping, eating, and studying, so the data would not provide accurate representation of how I spend time every day. Also, the data recognizes the length of time spent on a schedule that lasts overnight as if it is spent only on the start date. When calculating the amount of time I actually have on a certain day, I had to either perform a very complicated data wrangling or assign the length of time spent manually to each day in the data collection process. For future projects, I would use a different code for reading the data set into R so that it assigns the length of time spent on an overnight activity to each day automatically.

I collected the data set for 32 days, but it was not sufficient to establish the relationship between how I spent time on two different activities. I would have to collect the data for at least six months to discover meaningful relationships.

As someone who provides data, I expect that my data is treated with ethical considerations. I expect that my data is kept confidential and anonymous so that it cannot be traced to its provider. In order to ensure privacy of my data, I hope that data usage and any information gathered in the data analysis procedure are clearly communicated to me.

As someone who analyzes others' data, I think it is imperative to be aware of ethical implications of my analysis, given the privacy and sensitivity of the data. It is important to communicate the use of data and obtain informed consent from the data provider. Data collectors must constrain unrestricted, unauthorized access to data. Also, they must not let their preconceptions or personal opinions interfere with the data collection process.