

Diary of a [Data Science Student]: Coronatine

Stat231: Google Calendar Project

Xander Schwartz

May 4, 2020

Contents

Overview	2
How do I spend my time?	2
Details	2
What to do	2
Tips	3
Diary of a Data Science Student: Cabin Fever During COVID-19	4
Calendar Creation	4
Wrangling	4
Graphs	6
Insights	17
Reflections	18
References	19

Overview

This project will be undertaken individually. While you can discuss general ideas with your peers, please do not share code directly with each other. You can use other resources (including course materials, the internet, me, ...).

How do I spend my time?

One challenge of remote learning that many students have voiced is the difficulty of structuring our time in the absence of the usual structures in place when on campus.

How *are* we spending our time now? In this project, we'll track how we're spending our time using Google Calendar, then import the calendar data into R to analyze and provide insights into how we're spending our time.

The project aims to give you the opportunity to “play the whole [Data Science] game”, including proposing a question of interest that can be addressed with data and partaking in *data collection*, as well as wrangling, visualizing, analyzing, and reflecting on the information gathered. I hope an ancillary benefit is that it provides you with insights into how you're spending your time, and ideas for how better to manage your time. (One tip: create a “weekly plan” at the beginning of each week, and document any changes as the week goes on. See the “Video: Google Calendar Tips” on our Moodle course page.)

Feel free to extend upon this basic question “how do I spend my time?” or explore a variation of it. For instance, some other ideas include:

- Document “intended time” doing things (e.g. studying, sleeping) versus “actual time” doing those things, and compare results
- If you already use Google Calendar as a way to keep track of your schedule, you could compare how your time was spent at the beginning of the semester (on campus) to how your time is being spent at the end of the semester (since remote learning/COVID-19 isolation)
- Feeling artistic? Feel free to go the *data art* route, and create visualizations using other mediums outside of R. For examples: Dear Data (<http://www.dear-data.com/theproject>) and Environmental Graphiti (<http://www.environmentalgraphiti.org/series>).

Details

What to do

1. Identify 2-4 primary questions of interest to you about how you spend your time.
2. Track your time in your Google Calendar application for approximately 7-10 days (a longer amount of time would be ideal, but we're pressed for time here!). Fill in blocks of time and mark an entry with the activity you were performing: sleeping, studying, eating, exercising, socializing, etc. How you fill in and categorize your blocks of time should depend on what your questions of interest are.
3. Export your calendar data to `.ics` file format (a universal calendar format used by several email and calendar programs). This should take less than 5 minutes: <https://support.google.com/calendar/answer/37111?hl=en>.
4. Import the `.ics` file into R as a dataframe using the `ical` package (code to get you started is given below).

5. Create relevant summaries and visualizations, wrangling the data as necessary along the way. (Note: MUCH, MUCH LESS wrangling is expected here than in the Clearinghouse Challenge. The `ical` package imports the calendar data into a nicely formatted dataframe, with variables for summary of the event, time start, time end, and description.) Wrangling should *not* take a substantial amount of time.
6. What insights can you glean from how you're spending your time? Write a 3-5 paragraph report introducing your questions of interest, explaining what you found, and reflecting on the answers to your questions posed.
7. Finally, write a short reflection (1-2 paragraphs) on the process. 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 legal and ethical responsibilities do you have?

Tips

- Avoid creating calendar entries with a “repeat” schedule. Unfortunately, calendar entries based on “repeat” schedules only appear once (the first occurrence) in the R dataframe. (Or, if you do use repeat events – you would need to write an algorithm to code the recurring events.)
- The color coding of events is also lost when exporting the data into the `.ics` file, so don't rely on color-coding your calendar to give you information.
- There is a `googlecalendar` package under development that connects with Google's Calendar API that you may come across while searching online. Although you are free to explore that API if you want, it is not necessary nor expected that you utilize it for this project. The steps for importing your Google Calendar data into R are given below. (I have not explored the package or API myself; it could be a dead end.)

Diary of a Data Science Student: Cabin Fever During COVID-19

For your final submission, please delete the above instructions, and include all of your (reproducible) code and written responses below.

Then, intersperse your written report and any summary tables or visualizations here.

Calendar Creation

```
my_calendar <- ical_parse_df(file = "aschwartz23@amherst.edu.ics") %>%
mutate(start_datetime = with_tz(start, tzzone = "America/New_York")
, end_datetime = with_tz(end, tzzone = "America/New_York")
, length_sec = end_datetime - start_datetime
, length_min = length_sec/60
, length_hrs = length_min/60
, date = floor_date(start_datetime, unit = "day"))

may <- ical_parse_df(file = "may.ics") %>%
mutate(start_datetime = with_tz(start, tzzone = "America/New_York")
, end_datetime = with_tz(end, tzzone = "America/New_York")
, length_sec = end_datetime - start_datetime
, length_min = length_sec/60
, length_hrs = length_min/60
, date = floor_date(start_datetime, unit = "day"))
```

Wrangling

```
my_calendar <- my_calendar%>%filter(str_detect(start, "2020-04"))
may <- may%>%filter(str_detect(start, "2020-05"))

my_calendar <- my_calendar%>%mutate(useOfTime= ifelse(str_detect(my_calendar$summary,
"Sleep"),"Sleep",
"Work"),"School Work",
ifelse(str_detect(my_calendar$summary,
"work"),"School Work",
ifelse(str_detect(my_calendar$summary,
"Science"),"School Work",
ifelse(str_detect(my_calendar$summary,
"Facetime"),"Social",
ifelse(str_detect(my_calendar$summary,
"Running"),"Athletic",
ifelse(str_detect(my_calendar$summary,
"Tiger"),"TV",
ifelse(str_detect(my_calendar$summary,
"West World"),"TV",
ifelse(str_detect(my_calendar$summary,
"Dinner"),"Food",
ifelse(str_detect(my_calendar$summary,
"Breakfast"),"Food",
ifelse(str_detect(my_calendar$summary,
```

```

"Snack"), "Food",
  ifelse(str_detect(my_calendar$summary,
    "Social"), "Social",
    ifelse(str_detect(my_calendar$summary,
      "Squash"), "Athletic",
      ifelse(str_detect(my_calendar$summary,
        "Nap"), "Sleep",
        ifelse(str_detect(my_calendar$summary,
          "Pysch"), "School Work",
          ifelse(str_detect(my_calendar$summary,
            "Nap"), "Sleep",
            ifelse(str_detect(my_calendar$summary,
              "Minutes"), "TV", "Other")))))))))))))))

may <- may%>%mutate(useOfTime= ifelse(str_detect(may$summary,
  "Sleep"), "Sleep",
  ifelse(str_detect(may$summary,
    "Work"), "School Work",
    ifelse(str_detect(may$summary,
      "work"), "School Work",
      ifelse(str_detect(may$summary,
        "Science"), "School Work",
        ifelse(str_detect(may$summary,
          "Facetime"), "Social",
          ifelse(str_detect(may$summary,
            "Running"), "Athletic",
            ifelse(str_detect(may$summary,
              "Tiger"), "TV",
              ifelse(str_detect(may$summary,
                "West World"), "TV",
                ifelse(str_detect(may$summary,
                  "Dinner"), "Food",
                  ifelse(str_detect(may$summary,
                    "Breakfast"), "Food",
                    ifelse(str_detect(may$summary,
                      "Snack"), "Food",
                      ifelse(str_detect(may$summary,
                        "Social"), "Social",
                        ifelse(str_detect(may$summary,
                          "Squash"), "Athletic",
                          ifelse(str_detect(may$summary,
                            "Nap"), "Sleep",
                            ifelse(str_detect(may$summary,
                              "Pysch"), "School Work",
                              ifelse(str_detect(may$summary,
                                "Nap"), "Sleep",
                                ifelse(str_detect(may$summary,
                                  "Minutes"), "TV", "Other")))))))))))))))

```

Graphs

```
may <- may%>% filter(!str_detect(start,"2020-05-14"))
my_calendar <- my_calendar%>% filter(!str_detect(start,"2020-04-16"))

g <- ggplot(my_calendar, aes (x = useOfTime, y = length_sec)) +
  geom_col() + ylab("Minutes") + xlab("Use of Time at Home")
m <- ggplot(may, aes (x = useOfTime, y = length_min)) + geom_col()+
  ylab("Minutes") + xlab("Use of Time at Amherst")

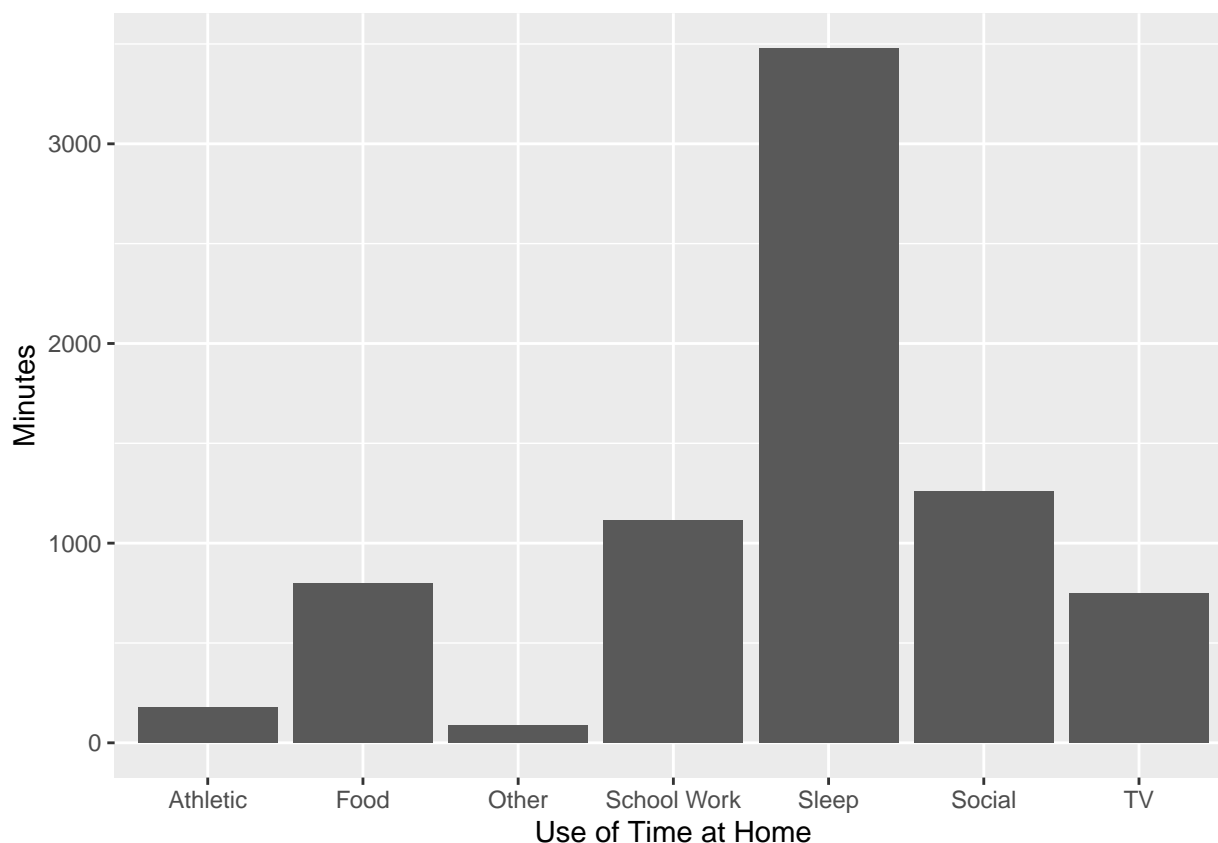
fullTypeAtHome <- ggplot(my_calendar, aes (x = summary, y = length_sec)) +
  geom_col()+ theme(axis.text.x = element_text(angle = 90, hjust = 1))+ ylab("Minutes") +
  ylab("Minutes") + xlab("Use of Time at Home")

fullTypeAtAmherst <- ggplot(may, aes (x = summary, y = length_min)) + geom_col()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+ ylab("Minutes") +
  ylab("Minutes") + xlab("Use of Time at Amherst")
```

Description: These charts show how I have used my time at Amherst and at Home. The use of time is broken into larger categories, such as TV time or School Work.

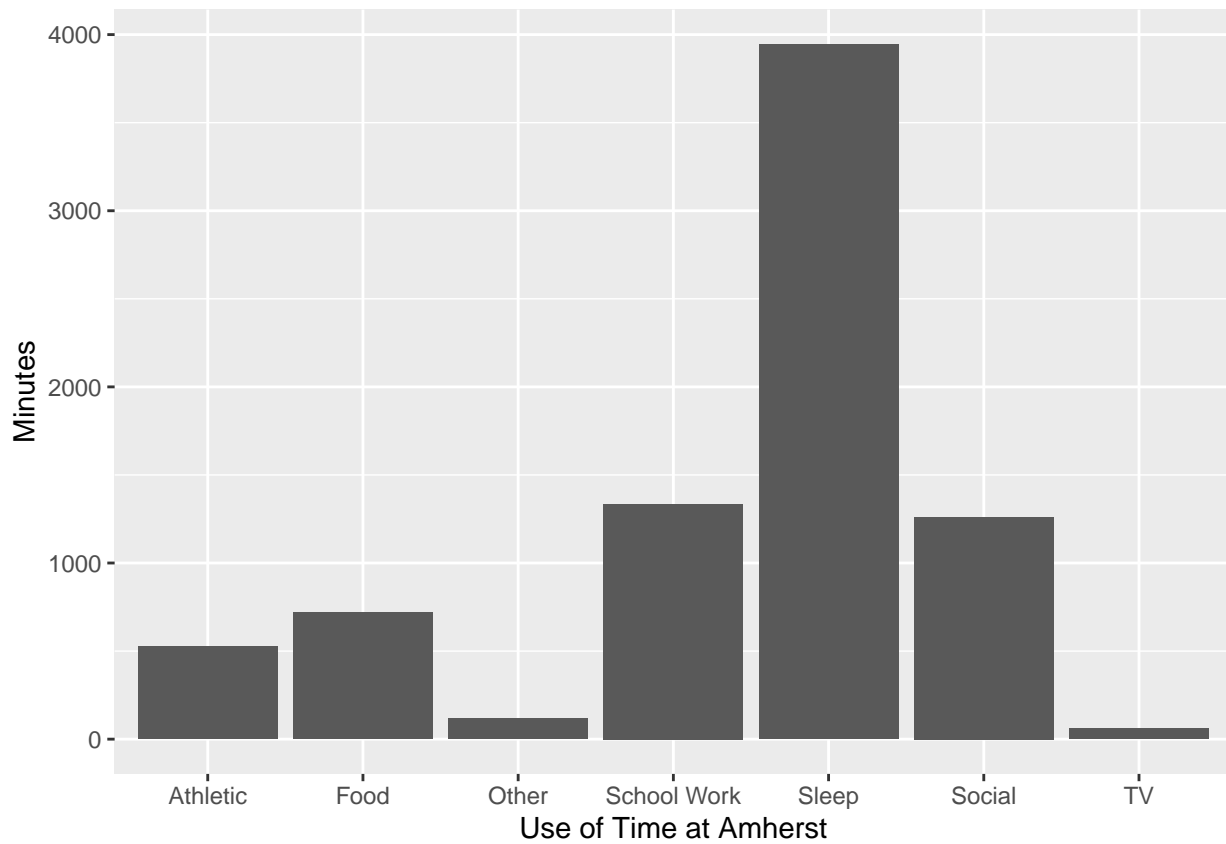
g

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



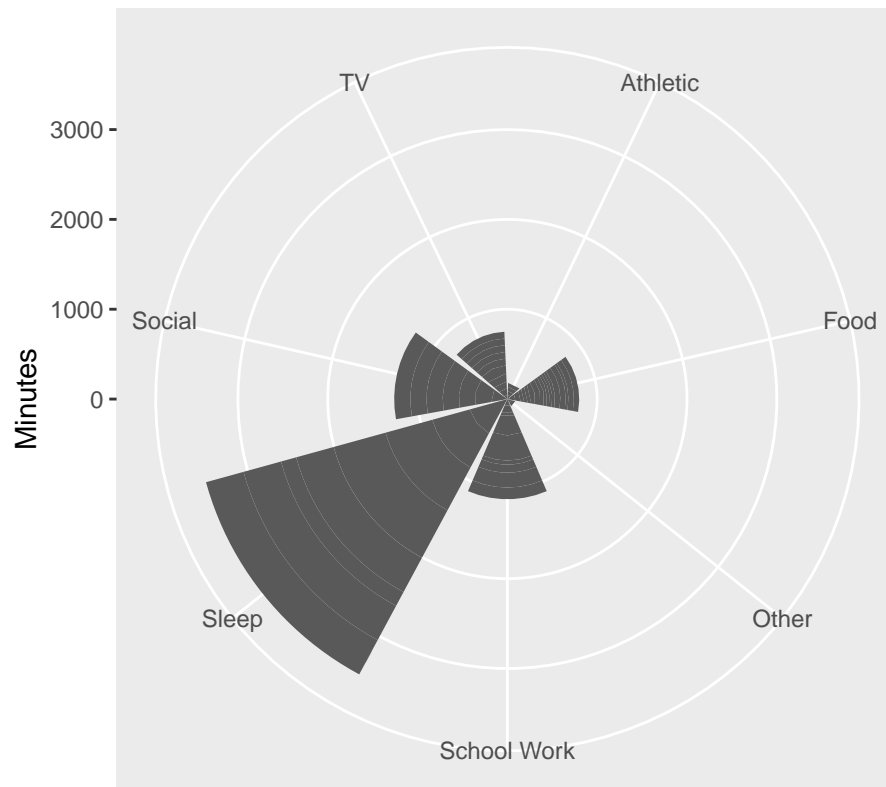
```
m
```

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



```
g+  
coord_polar("x", start=0)
```

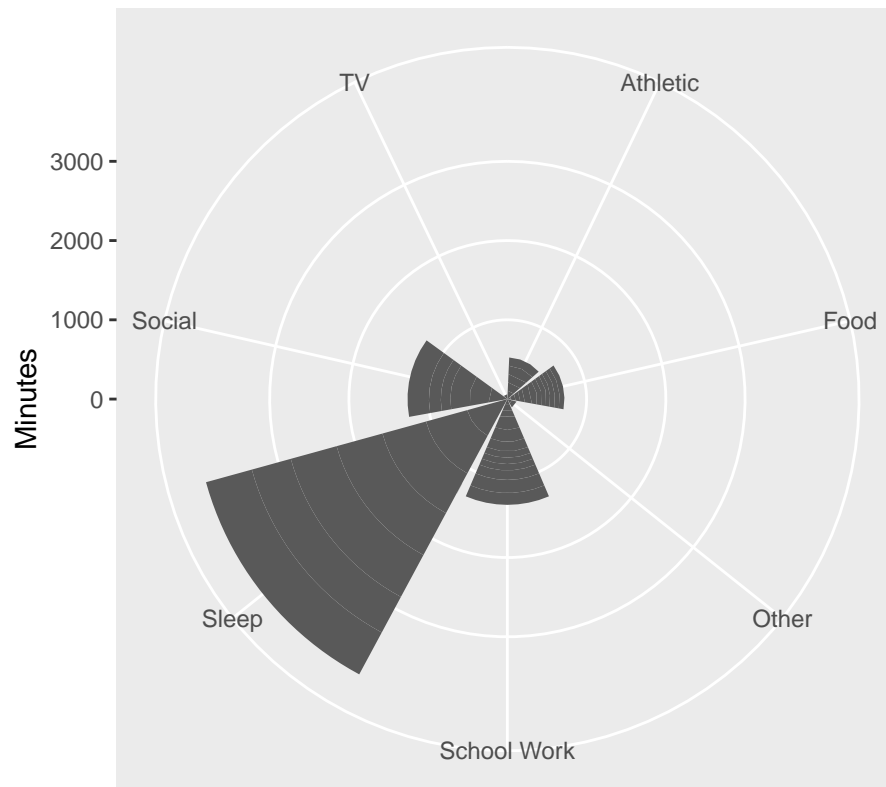
```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



Use of Time at Home

```
m +  
coord_polar("x", start=0)
```

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

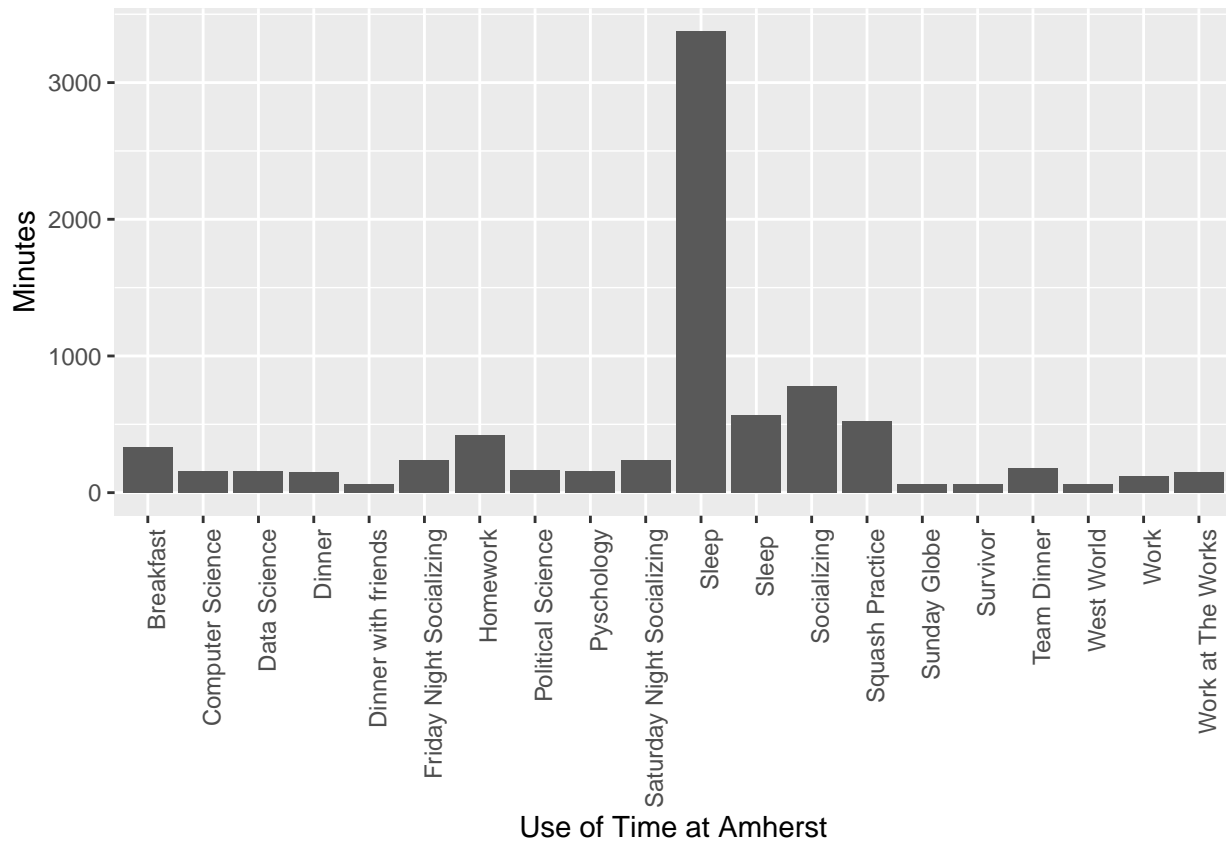



Use of Time at Amherst

Description: This chart analyzes time spent in the same way as the graphs above, but without breaking them into broader categories.

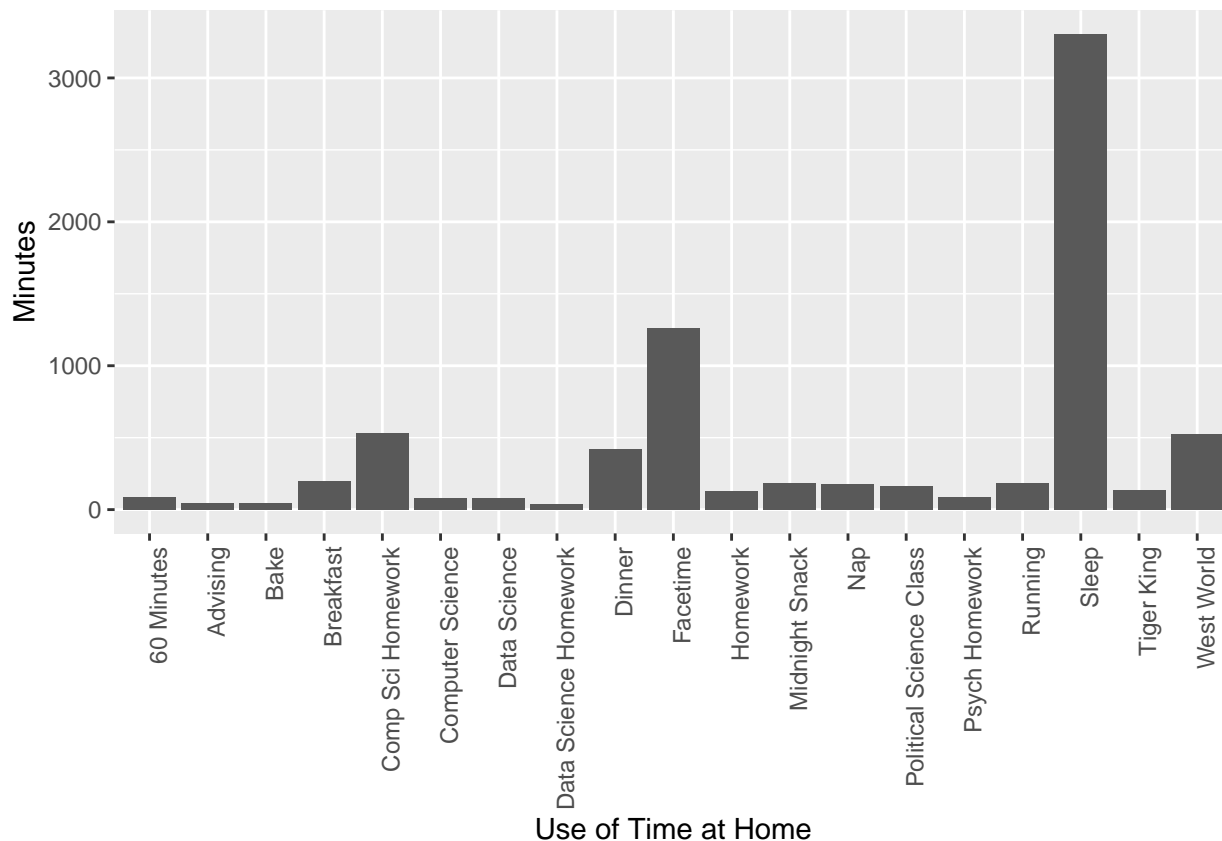
```
fullTypeAtAmherst
```

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



```
fullTypeAtHome
```

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



```
my_calendar <- my_calendar %>% mutate(Type = "Home")%>%
mutate("timeToUse" = length_sec)
```

```
may <- may %>% mutate(Type = "Amherst")%>%
mutate("timeToUse" = length_min)
```

```
fullCal <- full_join(my_calendar, may)
```

```
## Joining, by = c("uid", "summary", "start", "end", "description",
## "last.modified", "status", "start_datetime", "end_datetime", "length_sec",
## "length_min", "length_hrs", "date", "useOfTime", "Type", "timeToUse")
```

```
## Warning: Column `uid` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `summary` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `description` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `length_sec` has different attributes on LHS and RHS of join
```

```
## Warning: Column `length_min` has different attributes on LHS and RHS of join
```

```
## Warning: Column `length_hrs` has different attributes on LHS and RHS of join
```

```
## Warning: Column `timeToUse` has different attributes on LHS and RHS of join
```

```
fullCal <- mutate(fullCal, dayOfWeek = strptime(start,'%A'))

fullCal <- mutate(fullCal, partofWeek =ifelse(str_detect(fullCal$dayOfWeek,
  "Sunday"),"Weekend",
  ifelse(str_detect(fullCal$dayOfWeek,
  "Sat"),"Weekend","Weekday")))

group <- fullCal$useOfTime
value <- fullCal$timeToUse
data <- data.frame(group,value)

fullGG <- ggplot(fullCal, aes (x = useOfTime, y = timeToUse,
  fill = Type, alpha = partofWeek)) +
  geom_col() + theme(axis.text.x = element_text(angle = 90,
  hjust = 1))+

  ylab("Minutes")+ xlab("Use Of Time") +
  guides(fill=guide_legend(title="Location"))+
  guides(alpha=guide_legend(title="Part Of Week"))+
  scale_alpha_manual(values = c(0.65, 1))

dayofWeekGG <- ggplot(fullCal, aes (x = useOfTime, y = timeToUse,
  fill = dayOfWeek)) + geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ylab("Minutes")+ xlab("Use Of Time") +
  guides(fill=guide_legend(title="Day of the Week"))

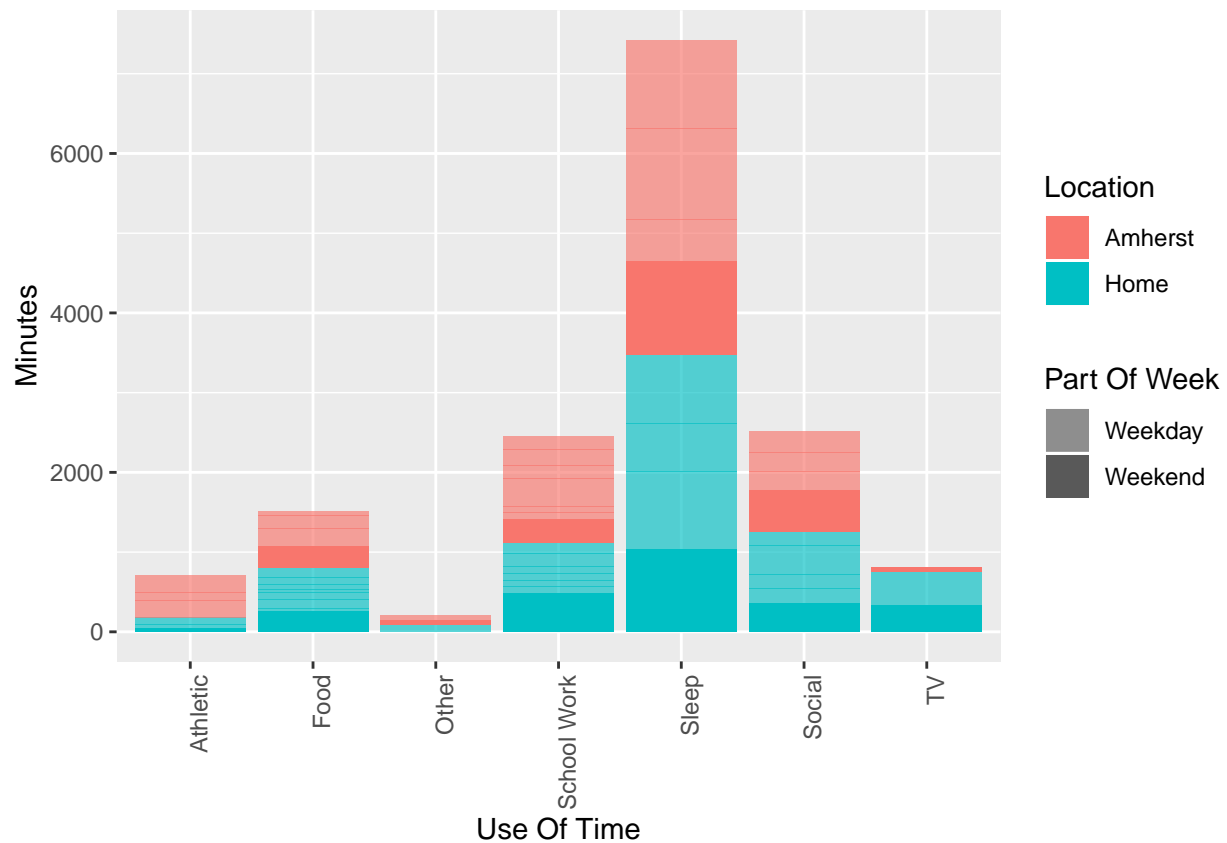
categories<-ggplot(fullCal, aes (fill = useOfTime, y = timeToUse,
  x = dayOfWeek)) + geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ylab("Minutes")+ xlab("Use Of Time") +
  guides(fill=guide_legend(title="Day of the Week"))

weekStatus <- ggplot(fullCal, aes (x = useOfTime, y = timeToUse,
  fill = partofWeek)) + geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ylab("Minutes")+ xlab("Use Of Time") +
  guides(fill=guide_legend(title="Type of Day"))
```

Description: This chart shows how I spent my time over the two periods (broken down into categories), but with the alpha level changing depending on whether it was at home or Amherst and weekday or weekend.

```
fullGG
```

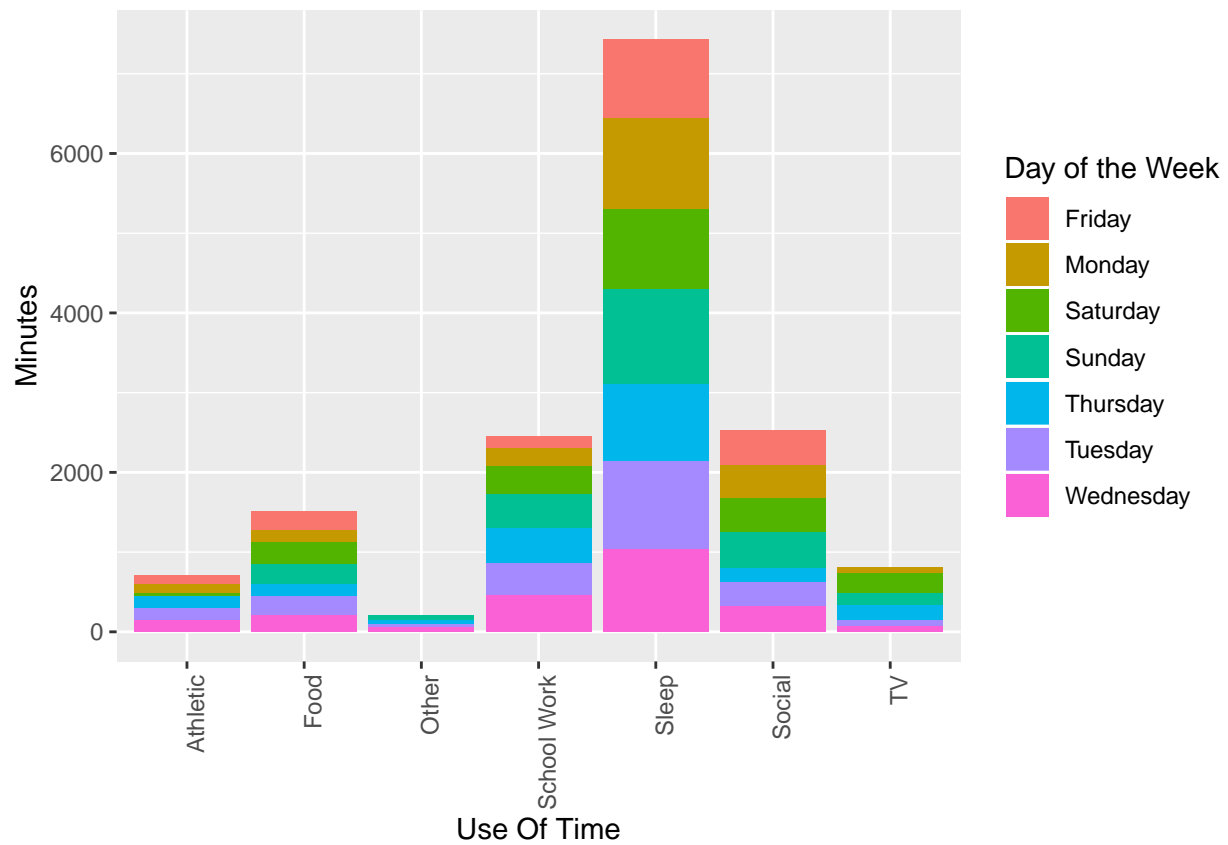
```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



Description: This chart looks at time spent (broken down into categories) broken down by the day of the week as well as the same chart, but with days of week on the x axis and the fill being the category.

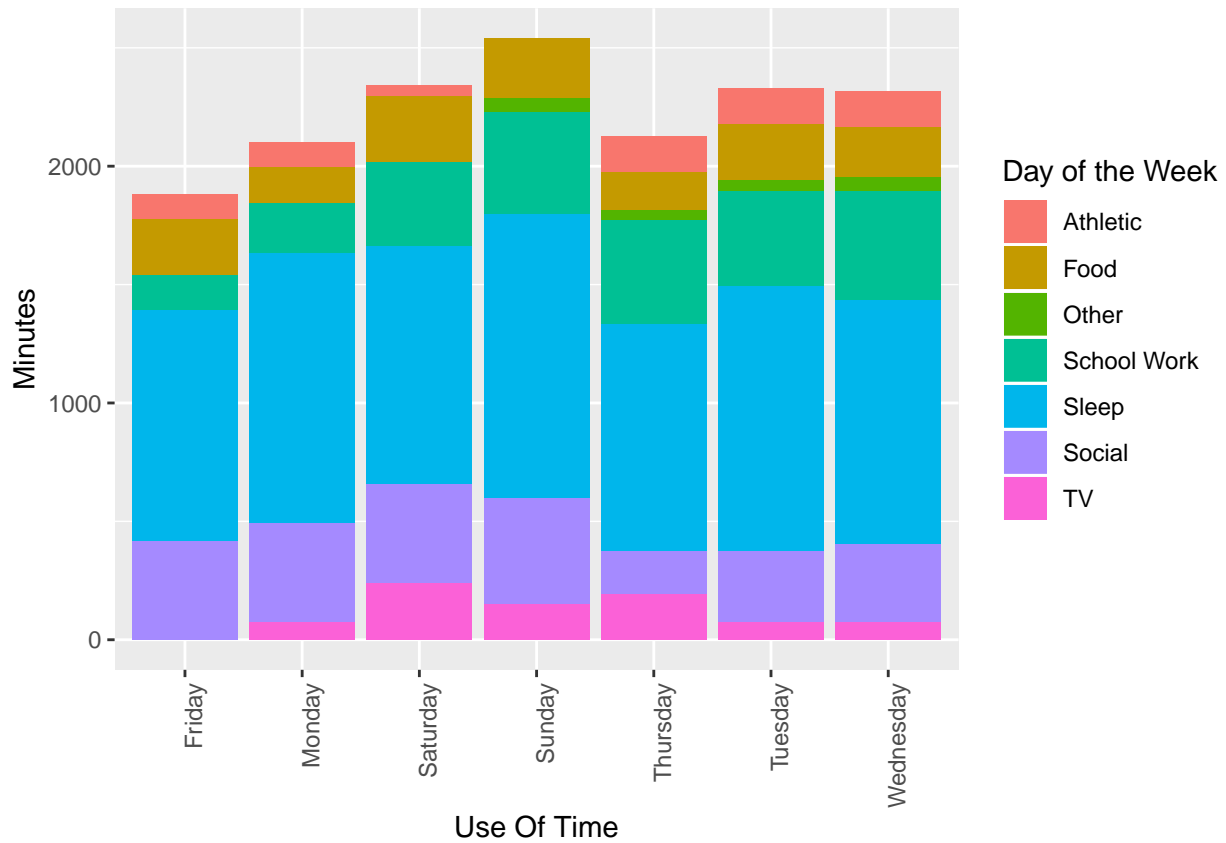
```
dayOfWeekGG
```

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



```
categories
```

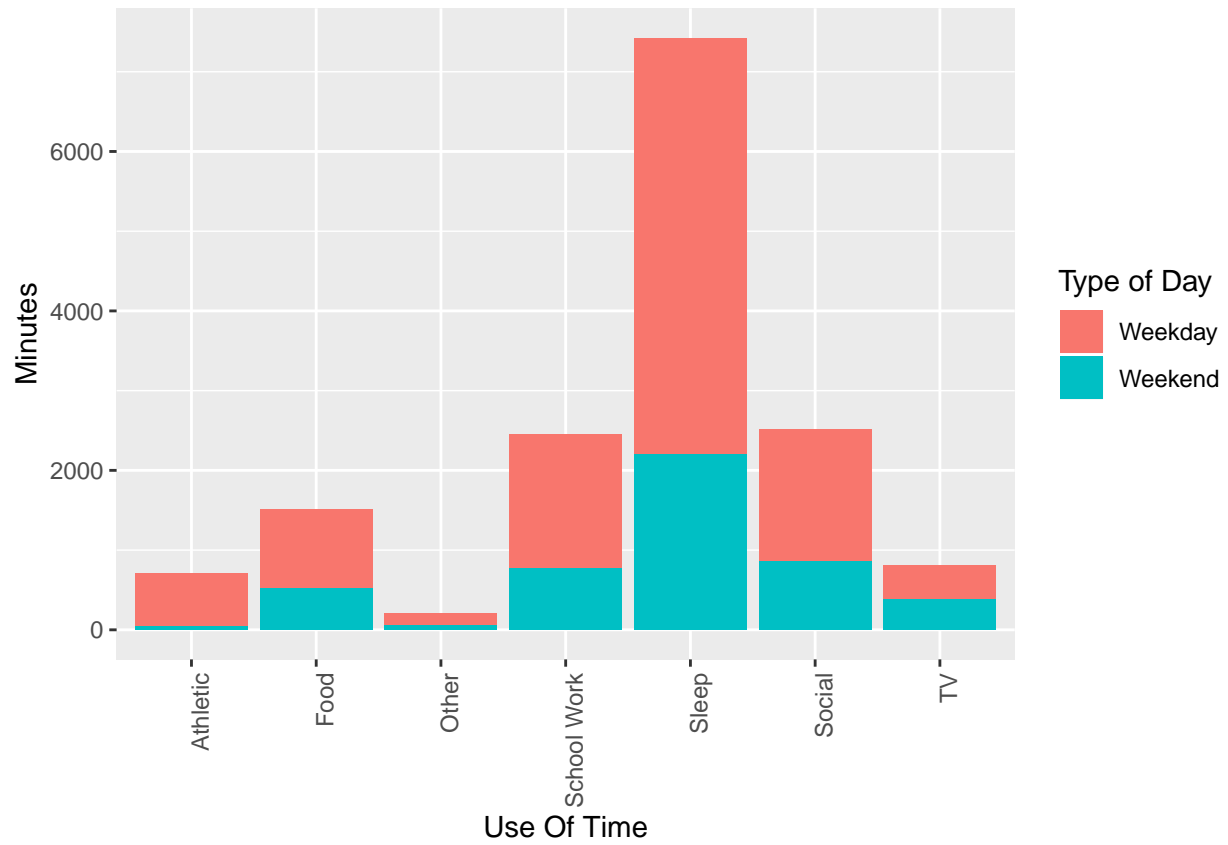
```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



Description: This chart looks at time spent (broken down into categories) broken down by whether the time was spent on a weekday or weekend. It's important to remember that weekdays should be a large percentage (5:2 ratio).

```
weekStatus
```

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



Insights

Write a 3-5 paragraph report introducing your questions of interest, explaining what you found, and reflecting on the answers to your questions posed.

Question of Interest: My question of interest for this project was trying to see how the quarantine has changed the way I spend my time. Am I still using my time in generally the same ways, or am I spending a lot more time on non-productive things? To measure this, I tracked my time for a week at home as well as creating a mock calendar for how I would have spent a week at Amherst. Before creating any ways of answering this question, I broke my time down into seven categories that covered most of my time spent in both locations: Athletic, Food, Other, School Work, Sleep, Social and TV. I then attempted to answer the question by looking at a variety of ways to measure time spent.

Findings: I went into this project hoping to look at the changes in how I spent my time at school and at home. Overall, I was surprised by the lack of overall change in my time spent at Amherst and at Home. The most notable difference I found was that I spend a lot more time watching TV at home and a lot less time doing something athletic compared to Amherst (this is mostly due to no longer having a 90 minute daily team practice). Many of my other pursuits remained relatively steady, such as how much time was spent on school work or sleep.

Reflection on Answer: Despite the breakdowns of how I spent my time seeming very similar, this breakdown of how I've spent my time would seem very different if we only looked at the data without the broader categories. While at school, I may have spent my time in the dorm with friends, I am now spending that time FaceTiming or my time at Val is now replaced by family dinners. More than anything, this project serves as a reminder of that while my home life is not too far off from my time at school, it is an artificial replication.

Reflections

What expectations do you have when you give your data (e.g. to Facebook, Google, MapMyRun, etc.)? As someone who analyzes others' data: What legal and ethical responsibilities do you have?

Reflection: When giving someone else my data, I think I'm pretty cynical. I expect almost anything connected to the internet is being used by some company to profit. I'm not sure how much that individually affects me (beyond targeted ads perhaps), but I don't think it is private. If I were creating legal and ethical regulations or acting as the data analyst, I believe that the companies should be explicit in how they are using our data. I don't necessarily believe that companies must stop using user data, but rather that the user must have easy access to how their data is being used. Similarly, the company should have an easy opt out function as well, if the user does not want their data to be used. From doing this project, I think I am now more acutely aware of what the data can look like. While my life is pretty uninteresting in quarantine, I can imagine that at some point for me or other people that there could be items on my calendar that I wouldn't want to risk being made public (intentionally or accidentally), so perhaps I would be more wary of using certain programs/applications in which my data could be collected.

References