

When, Where, and How?

Stat231: Google Calendar Report

Evan Daisy

Due Friday, March 19 by 5:00 PM EST

How do I spend my time?

```
#Reading in the original file
my_calendar0 <- ical_parse_df(file = "/Users/evandaisy/Desktop/Stat Project.ics") %>%
  select(summary, start, end) %>%
  filter(is.na(summary) == FALSE) %>%
  separate(summary, into = c("activity", "company", "location")
            , sep = "[,]"
            , remove = FALSE)
#Generally cleaning the data
my_calendar1 <- my_calendar0 %>%
  #Turning specific friends into "friends"
  mutate(company2 = case_when(str_trim(company) %in% c("alone", "public", "zoom") ~ company,
                              TRUE ~ "friends")) %>%
  #Creating a duration variable in hours
  mutate(duration = (interval(start, end)/ddays(1))*24) %>%
  #Putting super specific things into categories,
  #correcting data entry errors
  mutate(activity = case_when(activity %in% c("cards", "boggle", "secret hitler") ~ "games",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("temptation island", "movie", "prep for bed") ~ "tv",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("eating lunch", "getting lunch") ~ "lunch",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity == "geology" ~ "geo",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity == "bio comps" ~ "comps",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("mario", "phone games") ~ "video games",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("walking + talking", "walking") ~ "walk",
                              TRUE ~ activity)) %>%
  mutate(activity = case_when(activity == "seminar" ~ "bio seminar",
                              TRUE ~ activity)) %>%
  #Taking hour out of the start variable and
  #creating a categorical variable for time of day
  mutate(time = as.numeric(strftime(start, format = "%H"))) %>%
  mutate(time_of_day = case_when(8 <= time & time < 12 ~ "morning", 12 <= time & time <= 16 ~ "afternoon",
```

```
16 < time & time <= 19 ~ "evening", 19 < time & time <= 24 ~ "night",
TRUE ~ "night"))
```

#Cleaning for the activity visualization

```
my_calendar_acts <- my_calendar1 %>%
  mutate(activity = case_when(activity %in% c("tv", "games", "hangin", "video games", "mini golf",
      "in bed on phone", "moving stuff", "music", "nap",
      "napping", "talking") ~ "leisure",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("lunch", "dinner", "breakfast", "waiting for food",
      "late night") ~ "eating",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("check-in", "folding laundry", "shower",
      "testing") ~ "chores",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("route 9 stuff", "orchestra",
      "choir") ~ "extracurriculars",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("frolicking", "walk", "run") ~ "exercise",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("meeting", "research meeting") ~ "research",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("data science", "geo", "psych", "bio") ~ "classes",
      TRUE ~ activity)) %>%
  mutate(activity = case_when(activity %in% c("research", "bio seminar", "comps",
      "grading", "epic stuff") ~ "other work",
      TRUE ~ activity))
```

#Turning activities into waking activities

#(since including sleep makes the visualization look bad)

```
my_calendar_wacts <- my_calendar_acts %>%
  filter(activity != "sleeping")
```

#Creating a calendar for class time only

```
my_calendar_studies <- my_calendar1 %>%
  filter(activity %in% c("data science", "geo", "bio", "psych"))
```

#Turning each calendar into a summary

```
people_cal <- my_calendar1 %>%
  filter(activity != "sleeping") %>%
  group_by(company2, time_of_day) %>%
  summarise(tot_time = sum(duration)) %>%
  rename(company = company2)
```

'summarise()' has grouped output by 'company2'. You can override using the '.groups' argument.

```
activity_cal <- my_calendar_acts %>%
  group_by(activity, time_of_day) %>%
  summarise(tot_time = sum(duration))
```

'summarise()' has grouped output by 'activity'. You can override using the '.groups' argument.

```
wactivity_cal <- my_calendar_wacts %>%
  group_by(activity, time_of_day) %>%
  summarise(tot_time = sum(duration))
```

'summarise()' has grouped output by 'activity'. You can override using the '.groups' argument.

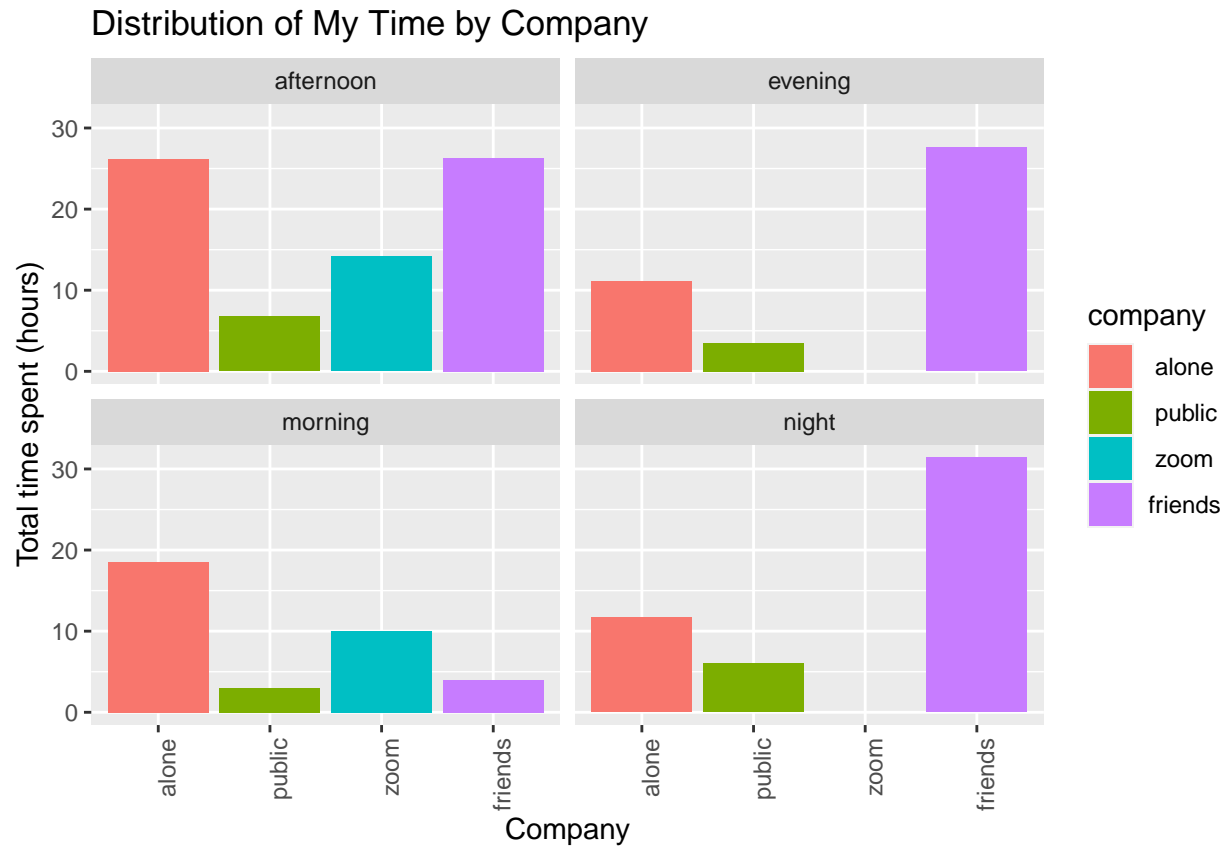
```
class_cal <- my_calendar_studies %>%
  filter(activity != "sleeping") %>%
  group_by(activity, time_of_day) %>%
  summarise(tot_time = sum(duration))
```

'summarise()' has grouped output by 'activity'. You can override using the '.groups' argument.

My very broad questions coming in were who do I spend most of my time with, where do I spend most of my time, and what do I spend most of my time doing, and how does this vary based on time of day? I collected data for every significant period of time (loosely defined as 15 minutes) during which I did primarily one thing, and for this observation I recorded where I was (could be a building like “val” or a room like “common room”), who I was with (could be a specific name, “alone”, “zoom”, “public”, or the name of a group, like “room group”), and what I was doing (could be the name of a class like “geo” or just the most appropriate name I could come up with, like “hangin”). I ended up collapsing each of these categorical variables into more broad categories, such as “friends” or “leisure”, and I decided to focus my limited number of visualizations on company, class time, and activities, because those for the location data weren’t as interesting.

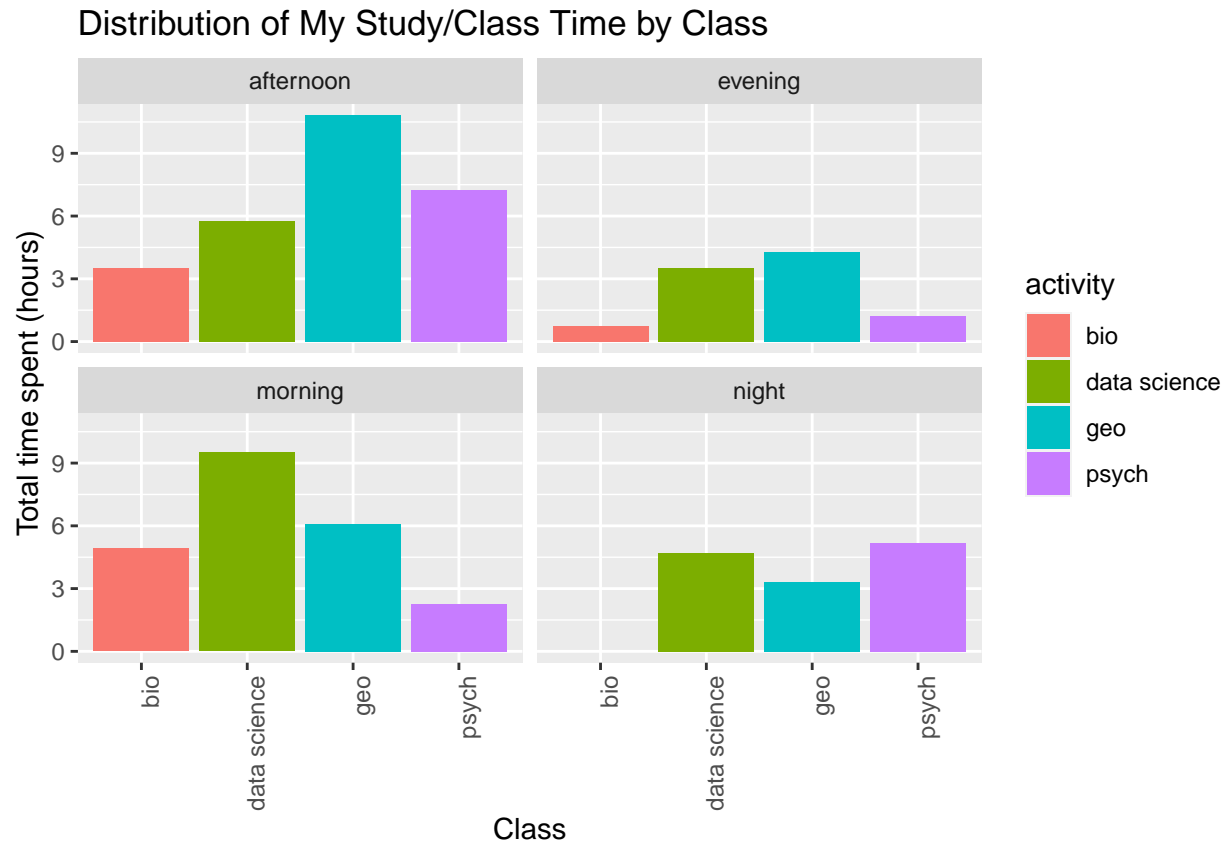
This first visualization describes the distribution of who I spend my time with. I was surprised to find that even during the pandemic, I still spend most of my time with friends, especially later in the day.

```
ggplot(people_cal, mapping = aes(x = company, y = tot_time, fill = company)) +
  geom_col() +
  labs(title = "Distribution of My Time by Company") + xlab("Company") + ylab("Total time spent (hours)") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  facet_wrap("time_of_day")
```



My second visualization depicts the distribution by class of the time I spend on classes. I was interested to see the ways in which this distribution changed over the course of a typical day, with more data science in the morning, more geology in the afternoon, and more psych at night.

```
ggplot(class_cal, mapping = aes(x = activity, y = tot_time, fill = activity)) + geom_col() +
  labs(title = "Distribution of My Study/Class Time by Class") +
  xlab("Class") + ylab("Total time spent (hours)") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) + facet_wrap("time_of_day")
```



This table shows the fractions of my time in the morning, afternoon, evening, and night respectively that I spend on each broad category of activity. During the morning I spend the majority of my time on classes, in the afternoon I spend the majority of my time on class and other work, and then in the evening and night leisure takes up a more significant portion of my time.

Note: “chores” includes activities like doing laundry and showering, whereas “other work” is more other academic work, such as grading and math research.

```
#Finding the denominators for the fractions
wactivity_totals <- my_calendar_wacts %>%
  group_by(time_of_day) %>%
  summarise(total = sum(duration))

#Turning waking activities into a summary and widening
wactivity_cal <- my_calendar_wacts %>%
  group_by(activity, time_of_day) %>%
  summarise(tot_time = sum(duration)) %>%
  pivot_wider(names_from = time_of_day, values_from = tot_time) %>%
  mutate(evening = case_when(is.na(evening) == TRUE ~ 0, TRUE ~ evening)) %>%
  mutate(morning = case_when(is.na(morning) == TRUE ~ 0, TRUE ~ morning)) %>%
  rename(Activity = activity) %>%
  rename(Afternoon = afternoon) %>%
  rename(Morning = morning) %>%
  rename(Evening = evening) %>%
  rename(Night = night)
```

‘summarise()’ has grouped output by ‘activity’. You can override using the ‘.groups’ argument.

```
#Finding percents using new (wide) columns
wactivity_percents <- wactivity_cal%>%
  mutate(Evening = Evening / 42.25000) %>%
  mutate(Morning = Morning / 35.48333) %>%
  mutate(Afternoon = Afternoon / 73.46667) %>%
  mutate(Night = Night / 49.11667)
kable(wactivity_percents, digits = 2)
```

Activity	Afternoon	Evening	Morning	Night
chores	0.02	0.01	0.11	0.01
classes	0.37	0.23	0.64	0.27
eating	0.13	0.31	0.08	0.08
exercise	0.10	0.05	0.11	0.11
extracurriculars	0.01	0.00	0.00	0.09
leisure	0.11	0.33	0.04	0.43
other work	0.26	0.06	0.01	0.01

```
# if you want to make your table's appearance nicer, check out:
# the xtable package (https://cran.r-project.org/web/packages/xtable/vignettes/xtableGallery.pdf); or
# the kable function in the knitr package (https://bookdown.org/yihui/rmarkdown-cookbook/kable.html)
```

Assuming this two week period is representative, I found that I spend most of my time on classes in the morning and afternoon, with data science making up most of my class time in the morning, geology making up most of my class time in the afternoon, and psych being a more frequent focus at night. I found that I still spend the most waking time with my friends (as opposed to alone, on Zoom, or in public), especially in the evening and night. Lastly, I found that I spend most of my time on classwork in the morning and academic work in the afternoon (but not all classwork), and focus on leisure more in the evening and night.

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?

As I was collecting my data, I found it difficult to accurately record when I started doing each activity and when I stopped. It would have taken more discipline than I have to record the exact minute I started each activity when I started it and the exact moment I stopped when I stopped, and even if I had done so, I wouldn't be accounting for the time spent collecting this data (since if I did I would need to account for the time spent collecting data on collecting the data and the cycle would continue until it ended up being all I ever did).

With this in mind, I chose to focus on an arbitrary period of time that I considered "significant" and only record instances when I spent that amount of time or more on an activity. The choice of what to record was also heavily influenced by my own judgement along the way however, because sometimes I would need to record what I had been doing after the fact, and especially in these instances I was likely to leave out activities that I didn't consider memorable or important, such as being on my phone. I never recorded data on going to the bathroom for example, even though I imagine that would make up a decent amount of my time in the long run.

In future data collection projects, I will try to set more achievable goals, such as recording all the time I spend on a certain class. This narrows the focus to an extent that I think would make it more manageable to accurately record time spent and would allow the project to take up less of my time and mental energy. For future data analysis projects, I will keep in mind researcher effects like the tendency to neglect or under-report instances deemed "unimportant", which can heavily skew the data.

In order to truly answer the questions of interest, even if I were to say "who do I spend my time with during this semester?", I would need to have completely accurate data about every interaction I had and who I was with for every minute of every day for the entire semester. I think having the kind of data I had for a whole semester is achievable (I did kind of get into the habit of it after a while) but having it be completely accurate (for the reasons addressed in paragraph 1) would be next to impossible. I would also need more blind data collection for this analysis to be accurate, because I noticed myself spending my time differently because I knew I was recording data (spending more time on a single thing to make things simpler for example, and not jumping around as much). I think to truly answer the questions of interest I would need to have a camera implanted in my eye without my knowledge that recorded this data without bias and with the utmost precision.

When I share data unintentionally (online for example), I expect that it be sent directly into algorithms to target ads or something and not accessed by an individual. Growing up in the 21st century I have come to expect that computers know almost everything about me, but my expectation is that my data remain digital and does not make contact with a human recipient in any kind of understandable form. This might not be the case, but it would be my hope. When I provide data to human researchers, I expect that the data will be recorded anonymously and not be attributed to me in any way, especially if it includes very personal information.

As someone who analyzes the data of others, I would give them the same courtesy I expect from human data analysts. I would anonymize data that is not anonymous, but I would hope that I wouldn't have access to

non-anonymous data in the first place. If I knew the identities of the participants in a study for any reason, I would not disclose these names or any amount of information that would be sufficient to connect them to a specific case in the data.

I think it is also my responsibility as a data analyst to be careful with the findings that I report. It would be unethical to exaggerate my findings or extrapolate in my reporting in a way that cannot be done with the data at hand, because this data analysis could affect people's lives, for example if companies, governments, or individuals use it to justify their actions. If my data analysis could be in any way used to justify discrimination or violence, I would need to take care to report it in such a way that it cannot be used to do so. Objectivity is especially important in such cases, where a finding could have a substantial impact on people's lives.