

Life in a Bubble: A 2-Week Categorical Time Analysis

Stat231: Google Calendar Report

Jack Dove

Due Friday, September 25 by 5:00 PM EST

How do I spend my time?

Calendar Data Importation and Wrangling

```
#Loading: 7 Calendars, with subjectively-assigned productivity scores.

#Baseball
my_calendar1 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/Baseball@group.calendar.google.com",
, end_datetime = with_tz(end, tzzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
  mutate(calendar = "Baseball") %>%
  mutate(productivity = 5*as.numeric(length_min)/60)

#Classes (Combined with Office Hours and TA Hours)
my_calendar2 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/Classes@group.calendar.google.com",
, end_datetime = with_tz(end, tzzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
  mutate(calendar = "Classes + Office Hours") %>%
  mutate(productivity = 9*as.numeric(length_min)/60)

#Homework
my_calendar3 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/Homework@group.calendar.google.com",
, end_datetime = with_tz(end, tzzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
  mutate(calendar = "Homework") %>%
  mutate(productivity = 10*as.numeric(length_min)/60)

#Meals (Combined with Miscellaneous into Extracurriculars)
my_calendar4 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/Meals@group.calendar.google.com",
, end_datetime = with_tz(end, tzzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
  mutate(calendar = "Extracurriculars") %>%
  mutate(productivity = 2*as.numeric(length_min)/60)

#Miscellaneous (Combined with Meals into Extracurriculars)
my_calendar5 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/Miscellaneous@group.calendar.google.com",
, end_datetime = with_tz(end, tzzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
```

```

mutate(calendar = "Extracurriculars") %>%
mutate(productivity = 2*as.numeric(length_min)/60)

#Office Hours (Combined with Classes and TA Hours)
my_calendar6 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/Office_Hours@group.calendar.google.com",
, end_datetime = with_tz(end, tzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
mutate(calendar = "Classes + Office Hours") %>%
mutate(productivity = 9*as.numeric(length_min)/60)

#TA Hours (Combined with Classes and Office Hours)
my_calendar7 <- ical_parse_df(file = "/Users/jackdove/Desktop/calendars/TA_Hours@group.calendar.google.com",
, end_datetime = with_tz(end, tzone = "America/New_York") , length_min = end_datetime - start_datetime
, date = floor_date(start_datetime, unit = "day")) %>%
mutate(calendar = "Classes + Office Hours") %>%
mutate(productivity = 9*as.numeric(length_min)/60)

#Joining Calendars, adding time lengths by hour, general cleaning
calendardata <- my_calendar1 %>%
  full_join(my_calendar2) %>%
  full_join(my_calendar3) %>%
  full_join(my_calendar4) %>%
  full_join(my_calendar5) %>%
  full_join(my_calendar6) %>%
  full_join(my_calendar7) %>%
  mutate(weekday = weekdays(date, abbreviate=F)) %>%
  mutate(weekdayid = wday(date)) %>%
  select(-c(uid, description, last.modified, status)) %>%
  arrange(start) %>%
  mutate(length_hours = as.numeric(length_min/60)) %>%
  mutate(summary = tolower(summary))

```

Questions of Interest + Data Collection

This project provided me with blocked out time to do what I've never had the motivation to do: track and analyze my time. For two weeks, I maintained consistent notation of all of my non-sleeping time uses.

I had two main questions I wanted to answer: 1) What activities am I spending the most time on? 2) How does my productivity compare by weekday?

I coded each activity by category using seven total sub-calendars (combined into four categories) on Google Calendars, and assigned each of the four groups an hourly productivity score.

- A) Baseball: Score = 5, any athletics-related activities, such as a workout or practice.
- B) Classes + Office Hours: Score = 9, time spent in either setting for any of my four classes.
- C) Extracurriculars: Score = 2, a wide category including meals, video games, COVID-19 tests, club meetings, and more.
- D) Homework: Score = 10, hours spent completing class assignments.

After some preliminary wrangling, I had hour-by-hour data as well as productivity scores to work with. In pursuit of answers to my questions, I created two visualizations, as well as an informative data table.

Visualization #1: Categorical Time Plot

To chart my time usage by date for each category, I used a stacked area chart, which used color to distinguish categories and width to compare time spent by day. Additionally, the chart displayed how much time I spent awake each day, illuminating any potential sleep time trends.

Regardless of date, this visualization immediately points to the large amounts of time spent on extracurriculars and homework. While homework time usage remains consistently high, as the two weeks progress, I spend increasingly more time on baseball and classes + office hours than on extracurriculars.

One of the starkest visuals presented is the trend line of total activity. By the end of the two weeks, I hit a record low of less than ten active hours, which is four less than the active-hour maximum value at 15 hours. Looking back, some of my new-semester energy may have dwindled by September 20th.

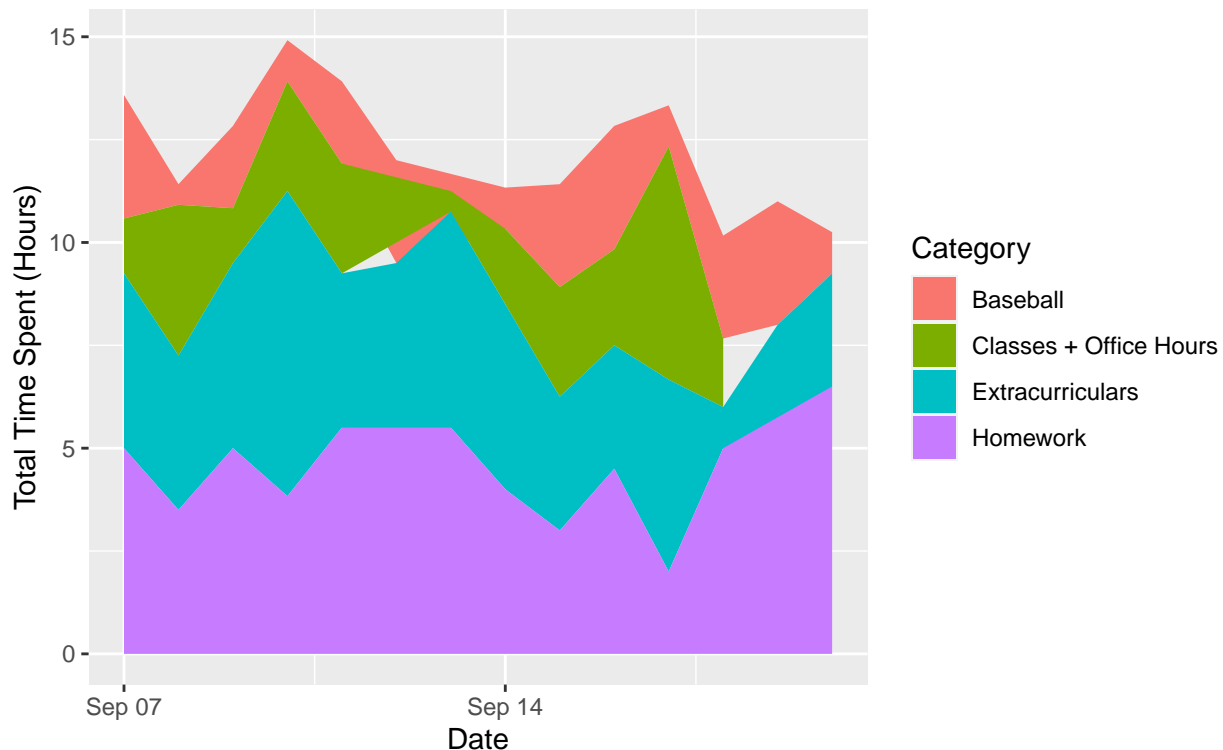
Overall, this chart points to a lot of extracurricular and homework time spent, as well as an increasing amount of sleep and non-active time allotted.

```
#Finding total time spent by category and date for chart 1
calendarchart1 <- calendardata %>%
  group_by(calendar, date) %>%
  rename(Category = calendar) %>%
  summarize(Total = sum(length_hours)) %>%
  arrange(desc(Total))

#Area chart by category of daily total time spent
ggplot(calendarchart1, aes(x=date, y = Total, group=Category, fill=Category)) +
  geom_area() +
  labs(x= "Date", y= "Total Time Spent (Hours)", title = "Time Spent by Category (Sept. 7-20)",
       subtitle="Each area represents a category's daily portion, adding up to total active hours.")
```

Time Spent by Category (Sept. 7–20)

Each area represents a category's daily portion, adding up to total active hours.



Visualization #2: Productivity Chart

While time spent is important, I wanted to see just how productive I am during each day of the week. After assigning each category a productivity score (see questions of interest + data collection for levels), I took the average of each weekday's productivity scores by category, which included data from two of each weekday, displayed in the bar chart below.

The bar chart tells stories both of weekday output comparisons and trends in productivity type as the week progresses. At midweek, my productivity is relatively balanced, yet by the end of the week (Sunday), my productivity is mostly coming from homework. Saturday is clearly a day off, as I only score 27 points of productivity (for reference, 10 hours of homework is 100 points).

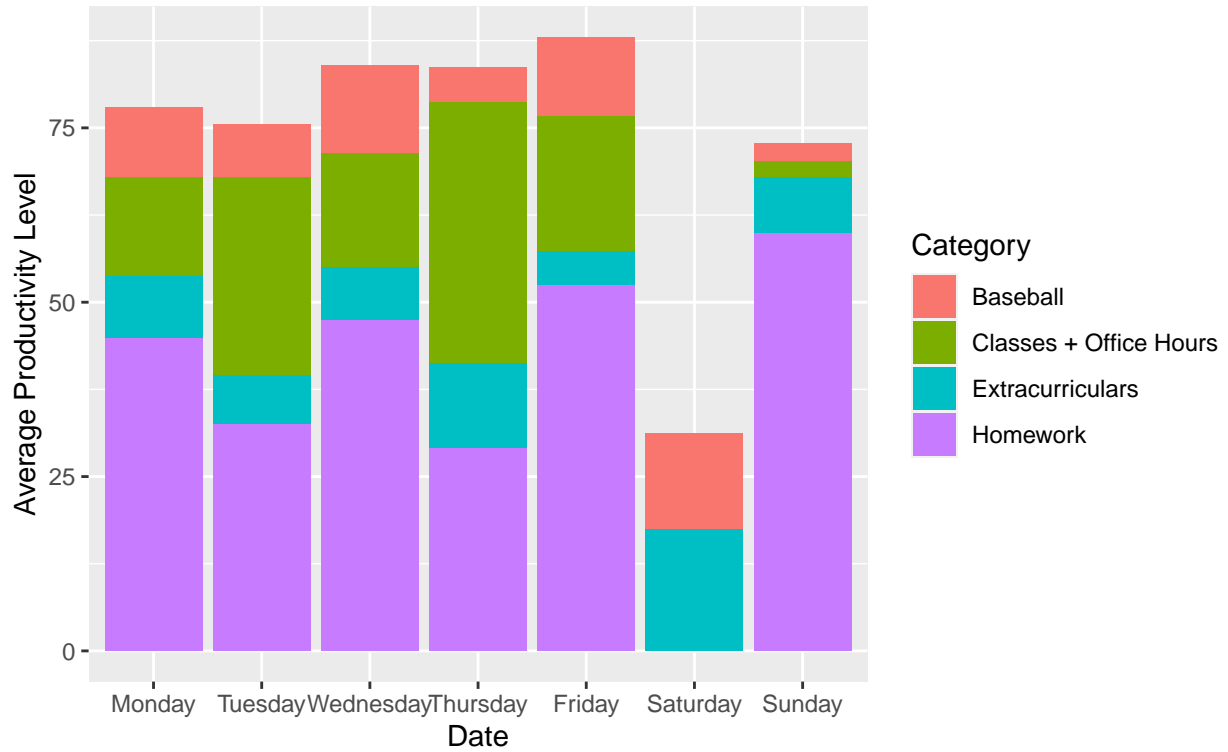
Total productivity follows a steady rise through Friday, resetting on Saturday and rising again on Sunday. This chart is also very consistent with my schedule, as the thick M/W/F baseball bars represent my typical practice times, and the thick Tues./Thurs. classes + office hours bars display my loaded academic schedule on those days. My main takeaways from this chart are that my productivity increases as the week continues, it resets on Saturday, and most of my productivity comes during homework sessions.

```
#Finding productivity by category and date for chart 2
calendarchart2 <- calendardata %>%
  group_by(calendar, weekday) %>%
  rename(Category = calendar) %>%
  summarize(weekdayid = mean(weekdayid), Productivity = sum(productivity)) %>%
  arrange(desc(Productivity))

# Bar chart of productivity by weekday and category
ggplot(calendarchart2, aes(x=factor(weekday, c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")), y = "Average Productivity Level", title = "Productivity Levels by Weekday", subtitle="Each category was assigned a productivity score weighted by time spent. "))
  geom_col() +
  labs(x= "Date", y = "Average Productivity Level", title = "Productivity Levels by Weekday",
       subtitle="Each category was assigned a productivity score weighted by time spent. ")
```

Productivity Levels by Weekday

Each category was assigned a productivity score weighted by time spent.



Data Table: Category Summary Statistics and Metrics

The data table below provides context to the visualizations above. With the productivity scores included, the table displays which categories I spent the most time in, the number of times I underwent an activity in a category, and the average time spent per activity, as well as the daily time per day metric for each category.

Arranged in descending order total time, extracurriculars took up 65 hours over the course of two weeks, with 73 total events. Meanwhile, homework and baseball training show high mean durations and productivity scores, combining to take up 5.6 hours per day. While this table doesn't tell many complicated stories, it helps provide useful standards to compare visualization values to.

```
#Calculating and organizing key statistics by category (for data table)
calendartable <- calendardata %>%
  group_by(calendar) %>%
  rename(Category = calendar) %>%
  summarize(Total = sum(length_hours), 'Mean Duration' = mean(length_hours), 'Time per Day' = Total/14,
  arrange(desc(Total))

#Data Table: Key measures of categories
knitr::kable(calendartable, caption = "September 7th-20th Time Usage (Hours)", digits = round(2), label
```

Table 1: September 7th-20th Time Usage (Hours)

Category	Total	Mean Duration	Time per Day	Number of Events	Hourly Productivity
Extracurriculars	65.58	0.90	4.68	73	2
Homework	53.33	1.78	3.81	30	10
Classes + Office Hours	26.33	0.98	1.88	27	9
Baseball	25.00	1.56	1.79	16	5

Conclusion

I asked two key questions pre-analysis: 1) What activities am I spending the most time on? 2) How does my productivity compare by weekday?

Right off the bat, upon a quick look at my first visualization, I can tell I spend most of my time completing homework and participating in extracurriculars. My visualization brings the nuance of the date to that question, displaying a negative trend in active time spent as the two weeks progressed. My key thoughts, upon this first analysis, are that I spend a bit too much time on extracurriculars, and should look to be more consistent with my sleep schedule if I want to raise my hours on the baseball field and in class/office hours.

As far as productivity goes, according to my second visualization, I am increasingly productive as the week goes on, but lose steam by Saturday, saving lots of homework for Sundays. Again, I'd like to be more consistent in the future, and monitoring my sleep might help productivity as well.

Getting meals and taking COVID-19 tests are necessary, but the data table shows that I participated in about 5 extracurriculars per day. The visualizations answered both of my questions, and the data table confirmed what needs to change to up time output and productivity: I should pursue consistent sleep and weekday balance if I'd like to make beneficial changes to my schedule.

Reflection

My main reflection from this assignment is how enlightening this data can be; I've previously been a great scheduler, but a poor post-week reflector of how I spent my time. I'm really glad I could do this as a class assignment.

The hardest part of the data collection process was the needed consistency of my data entry. As remedial as typing "breakfast" into Google Calendars is, I decided that, for two weeks, I'd make sure to track all of my active hours. The analysis of the data was most challenging in the wrangling and cleaning phase. I used the `view()` function and its corresponding sorters to identify category and spelling errors, allowing me to re-import fixed calendars with the same code. The internet provided me with several helpful techniques, including how to yield weekday names from dates.

These two hurdles, collection consistency and cleaning difficulty, were significant, but will allow me to have a better understanding of data science as a whole. While in the past I've been able to quickly design plots with clean data, I gained an appreciation for how much time is spent in those two phases of projects, even getting a new, joyful feeling when I was finally able to apply plots and a data table to my data. In the future, I will look to collect more personal data for self-projects and to allot more time for the wrangling phase of statistics and data science projects.

To truly answer my questions, I'd need a full year of data. If I was looking to get precise answers, I'd want to see any seasonal differences in productivity, as well as whether my location plays a factor in time output and productivity. That collection process would be harder, as I'd be tracking more qualities of each activity than I did during these two weeks of September 2020. While I don't think I'm going to continue this study, I'm going to apply some of my conclusions from it to my schedule, as well as saving this collection process as a potential schedule issue diagnosis tool.

When I sign up for applications that require my personal data, I feel comfortable trading my data for their service. For example, JP Morgan intakes all credit card purchases made by its clients, turning that information into an economic forecaster which helps them make investments in different industries. While some might view that as profiting off of someone's private data, I view that as trading my valuable information for a valuable service. However, I draw the line at where a company's decision making is directly contingent on personal data. For example, a bank declining a customer's loan request based on their race would violate my standards for data use. Even though a particular race might historically display greater default rates, certain data must be encrypted in order to ensure that every debtor and creditor receives a clean slate, which I'd like to have when I look to obtain my first mortgage loan.

However, if I am working with someone else's data, I have legal expectations I must follow, as well as ethics to stand by. For example, I'm working on building a pseudo-gambling web scraping application, where my friends can place theoretical wagers on games. If that money were real, me causing them harm through their data (for example, posting their losses on social media and suggesting they have a gambling problem) would be unethical. I hope to work with data in environments where the person inputting data and the analyst of that data are on the same page expectations-wise.