# STAT 231 Project 1: Calendar Query

## How do I spend my time?

## Contents

## Summary

Our first project is an **individual project** that will allow you to interrogate something of interest to you about how you spend your time. As students with schedules packed with classes, activities, sports, performances, and more, it can be useful to examine where your time really goes from day to day. Often our expectations don't match our reality.

This project will give you an opportunity to work through the full cycle of the scientific process. I hope, as an added benefit, you gain useful insights into how you spend your time, and perhaps make adjustments if desired!

Thank you to Albert Kim (Smith College) and Johanna Hardin (Pomona College) for the original Google Calendar assignment (see the set up for their courses here and here). They were inspired by an episode of the *Not So Standard Deviations* podcast titled "Compromised Shoe Situation", in which the hosts (Roger Peng and Hilary Parker) discuss a data science design challenge on getting to work on time.

**Timeline**

| Activity | Date |
| --- | --- |
| Proposal due | Tuesday, September 6 by 10pm |
| Revised proposal due | Thursday, September 8 by 10pm |
| Begin data collection | Monday, September 12 or upon approval |
| Export, import, wrangle, analyze | Iterate throughout project period |
| Submit Calendar Query | Friday, September 30 by 10pm |

# Details

## 1. Pose 2 to 4 questions of interest

Identify *2 to 4 primary questions* of interest to you about how you spend your time. I hope your questions address something meaningful to you, but you should address questions you feel comfortable sharing with me.

One goal of this project is to give you practice working with time and date data, so your questions **must** include activities with measurable start and end times (even if only approximate!).

### Things to consider

*Your report will be private (only shared with me) by default.*

- Many students choose to make their work public as part of a portfolio and/or professional website to show off their data science skills. If you think you might go this route, make sure you are addressing questions you are comfortable sharing publicly.

*Your questions should be able to be answered with only 14 days worth of data.*

- If you happen to have additional data available via time-tracking or calendar apps, you can incorporate more!

*Your questions should be feasible to measure and not be too burdensome to track.*

- I've had students interested in tracking how long they spent distracted, for example, but the thing with getting distracted is that you usually don't realize it's happening, making it hard to track when the distraction period starts and ends!
- Try to focus on events that take up more continuous periods of times rather than in short bursts (unless you have other ways of collecting/tracking those "short burst" types of data).
- You might want to test out data collection before you submit your proposal to help you get a sense of what's feasible.

*You may measure additional things related to the activities you are tracking.*

- For example, students who track their sleeping and waking times might additionally record their mood each morning or whether they had caffeine close to bedtime.

*Feel free to expand upon the basic question of "How do I spend my time?" or explore a variation of it.* Some other ideas to help you get started include:

- document *intended time* doing things (e.g. studying, sleeping) versus *actual time* doing those things, and compare results
- document time spent on each course, and/or time spent on different parts of a course (e.g. in class, reading, homework)
- document time spent on school vs. work vs. leisure vs. rest, . . .
- if you already use a calendar app as a way to keep track of your schedule, you might compare how your time was spent last year at this time versus how your time is being spent this semester

**Project proposal process**

Your project proposal will be submitted as a **new issue** in your private GitHub repo. It should take the format provided in the README and include your corresponding answers to the following questions.

1. List 2 to 4 primary questions you are interested in investigating about how you spend your time. If you have supplemental questions, you may also list them here, but please indicate them in a secondary list.

2. At a minimum, you will use the start time and end time of your recorded activities to calculate the durations of those activities (e.g., "hours spent in class"). Identify what variables you will be measuring, what type of variables they are (quantitative or categorical), and their corresponding units or levels. Replace the examples provided below with your own data.

3. How do you plan to collect your data? Describe what tool(s) you will use (Google calendar, toggl, a spreadsheet?) **and** what process you will use to collect and record data (e.g., "I will use the Notes app to list what time it is when I lay down to sleep and then again when I wake up, and I will add the block to my calendar each day while I'm eating breakfast at Val", "I will start a timer entry when I enter Val and stop it when I leave", "When I sit down to study, I will create a block of time on my calendar to indicate how long I plan to study. When I am done studying, I will create a second block with how long I actually spent studying.")

4. Either describe in detail and/or sketch and upload two (or more) visualizations that will address one or more of your questions of interest. If describing, make sure to discuss the type of plot, scale or coordinate system, visual cues, etc.

5. Either describe in detail and/or sketch and upload a summary table or results table that will address one or more of your questions of interest. If describing, make sure to discuss the rows, columns, and types of values you would include in your table.

I would encourage you to answer the questions within the README first, then copy and paste the questions and answers into the body of the new issue.

You will have flexibility to edit/modify your questions, visualizations, and table, but we must always start with a plan! The more detail you have early on, the easy it is to carry out as you go.

I will review your project proposals and you will have the opportunity to revise and resubmit. The turnaround time will be quite short to give you ample time to collect and then analyze data.

## 2. Collect data for 14 days (or more!)

Use a time-tracking or calendar app to track your activities of interest for **at least 14 days** (the goal is to start data collection by Monday, September 12).

I recommend using your Amherst Google calendar or the toggl track app (you can sign up via Google with your Amherst email address). Other calendars or apps are acceptable as long as you can download your data as a **CSV** file or as a calender in **ICS** file format, a universal calendar format used by several email and calendar programs.

- Questions related to other sources of data or file types should be supplemental since I cannot guarantee that I (or the SDS Fellows) can provide support.

*If using a calendar app*, fill in blocks of time on your calendar, 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.

*If using a time-tracking app*, note the activity and start/end the timer when the activity starts/ends in real-time. Your app should also have the option to add entries manually for activities you can't track in real-time (approximate start/end times are fine!).

**Lessons learned the hard way**

*General*

- If you already use a calendar app or time-tracking app, you should create a new calendar(s) (or workspace within the toggl app) dedicated to this project. That way, your pre-existing information will be kept private.

- Be as consistent as possible about how you use calendar fields and how you type your entries.

- Export and wrangle your data early and often. You'll want to iterate between collecting and wrangling your data to identify problems with data collection as early as possible.

- If it is meaningful to track that something *didn't occur* on a particular day, you might consider adding a 1 second (or other very short) entry for the non-event. There is a way to add in 0-minute entries after the fact within R but it might be easier to turn a 1-second entry into a 0 instead.

*Specific to Google calender*

- The color coding of Google calendar events is lost when exporting the data into the ICS file, so *do not* rely on color-coding your calendar to give you information for this assignment.

- If you try to add an activity as a repeated event on Google calendar, it will not actually export with repeated entries. You should manually add each activity and all its repetitions separately.

## 3. Export, import, wrangle, and analyze!

Over the 14 days (or more!) of data collection, regularly export your data to either a CSV or an ICS file format and work on wrangling and analyzing it.

**Export to CSV or ICS**

It should take under 5 minutes to export your data.

- Google calendar: see instructions for exporting one or more Google calendars.

- toggl track app: log in via a browser, go the **Reports** section, select the **Detailed** view, verify you have the correct time frame for data collection (you can choose a specific date range), then click the download button in the top right to export the data as a CSV.
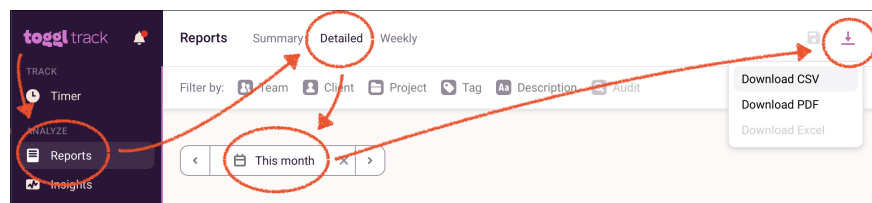


**Import into R**

Import the CSV or ICS file into R. Sample code will be provided for importing ICS files with the **ical** package.

**Wrangle**

Use the **tidyverse** and **lubridate** packages to wrangle the data. You should not resort to base R loops/indexing/extraction or any packages outside of what is introduced in class for wrangling your data unless absolutely necessary (please run it by me or Google first if you think you can't accomplish something using **tidyverse** or **tidyverse**-friendly packages).

Limited sample code will be provided to get you started. We will use part of one class to work on wrangling our data, but to take full advantage of the class time, it would be helpful if you attempted some wrangling on your own first.

**Analyze**

Create **at least two visualizations** using **ggplot2** and **at least one publication-quality table** using `kable()` and additional functions available from **kableExtra** (e.g., `kable_styling()`). The table should be a *summary table* or a *table of results*.

- A *summary table* is a table of relevant summary statistics (e.g., measures of center, measures of spread, proportions, regression coefficients) that are informative or useful in addressing your question(s) of interest.

- A *table of results* is a table summarizing the results of statistical inference (e.g., confidence intervals, hypothesis tests, regression modeling) that includes relevant estimates (e.g., effect sizes, regression

coefficients, confidence intervals) and any corresponding p-values.

- Here is a resource to create awesome PDF tables with **kableExtra** (or awesome HTML tables for later projects!).

## 4. Draw conclusions and communicate results.

What insights can you glean from how you're spending your time? Write a brief report introducing your questions of interest, explaining what you found, and reflecting on the answers to your questions posed.

Your report should include the following:

- Describe the questions of interest: what are they and why are they important or useful to you?
- Describe your data collection process, including how you defined variables of interest (include the levels/categories of categorical variables and the units for quantitative variables).
- In connection with the questions you posed, describe what information is conveyed through each visualization or table (avoid describing visualization choices). What did you learn about how you spent your time? What are the big take-aways for you? If relevant, how does this affect how you might spend your time going forward?

## 5. Reflect

Finally, write a brief reflection on this experience. Particular questions to reflect on include. . .

- How much time and what quality and/or type of data do you think you'd need to collect in order to satisfactorily answer your question(s) of interest? Would it be hard to collect that data? Why or why not?
- What were two (or more) main difficulties in the data collection and analysis process that you encountered?
- What lessons will you carry forward into future data collection and/or analysis projects?