Project 1: Calendar Query

Due by 10pm ET Friday October 1

How do I spend my time?

As we adjust to being back in-person, having full class schedules, actually having to physically move from one space to another, it can be useful to examine where the heck our time goes from day to day.

In this **individual project**, you will undertake the full data analysis cycle to examine aspects of how you spend your time. I hope, as an added benefit, you gain useful insights and perhaps make adjustments if needed.

Project details

1. Pose a question of interest

Identify **2-4 primary questions** of interest to you about how you spend your time. We may be sharing results with peers, so be sure that you identify questions **you feel comfortable sharing with your peers and instructor**.

Feel free to expand upon the basic question of "How do I spend my time?" or explore a variation of it. Some other ideas 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, etc.)
- Document time spent on school vs. work vs. leisure vs. rest, etc.
- If you already use Google Calendar as a way to keep track of your schedule, you could compare how your time was spent last year at this time versus how your time is being spent this semester.

2. Collect data

For 14 days (or more!), track your time in a calendar application. I recommend Google calendar, but other electronic calendars are acceptable as long as you can download an `ics` file format, a universal calendar format used by several email and calendar programs.

- If you already use a calendar app, you should create a new calendar within the app dedicated to this activity. That way, your pre-existing calendars will be kept private.
- Have another way of tracking time you'd rather use? Let's talk! We'll see if it's possible to incorporate the data.
- The color coding of events is lost when exporting the data into the `ics` file, so don't rely on color-coding your calendar to give you information for this assignment.

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.

Over the 14 days (or more!), regularly export your calendar data to the `ics` file format. This should take less than 5 minutes (e.g., see instructions for Google calendar export).

You'll want to iterate between collecting data and wrangling (next step) to identify problems with data collection as early as possible.

3. Import, wrangle, and analyze!

Import the `ics` file into R as a dataframe using the **`ical`** package (code to get you started is given below). You'll also need **`tidyverse`** and **`lubridate`** packages to accomplish this task.

- The code below can be used to import the `ics` file into R as a dataframe. Be sure to update the capitalized parts of the code with the relevant path to where your `ics` file is located and your email address.

```
my_calendar <- ical_parse_df(file = "PATH-TO-ICS/YOUR-EMAIL-ADDRESS.ics") %>%
  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"))
```

Create **two visualizations and one table**, wrangling the data as necessary along the way. Remember, you should be iterating between collecting data and wrangling to catch any problems as early as possible.

We will use part of one class to work on wrangling our calendar data, but to take full advantage of the class time, it would be helpful if you attempted some wrangling prior to that class.

4. Draw conclusions and communicate results.

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

Your report should follow the provided template.

5. Reflect

Finally, write a one-page reflection on this experience. Particular questions to reflect on are included in the template provided.

Timeline

Activity	Timeline
Proposal due	Thursday, September 9 10pm ET
Begin data collection	Monday, September 13 or upon approval
Export and wrangle data	Do this regularly throughout the 2-week period
Finish data collection	Sunday, September 26 (or 2 weeks from approval)
Final report due	Friday, October 1

References

Thank you to Albert Kim (Smith College) and Johanna Hardin (Pomona College) for the Google Calendar assignment. 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.

Code rubric (50%)

Code	Excellent	Good	Satisfactory	Poor	Unacceptable
Functionality (30%)	*25-30 points*: The code is completely functional and responds correctly producing the correct outputs.	*18-24 points*: The program is mostly functional and responds correctly producing the correct outputs in most cases. There are minor problems with the program implementation.	*11-17 points*: The code is marginally functional with numerous errors and/or incomplete code sections.	*6-10 points*: The code is minimally functional with significant portions of the code missing or incomplete.	*0-5 points*: The code is not functional, producing no correct outputs, or was not attempted.
Readability (10%)	*9-10 points*: The code is extremely well organized, properly formatted, and easy to follow.	*7-8 points*: The code is reasonably easy to read. There are minor formatting problems.	*5-6 points*: The code readable only with significant effort. There is little to no proper formatting.	*1-4 points*: The code is poorly organized and difficult to read. There is little to no consistency in formatting.	*0 points*: The code is readable only by the author or someone extremely knowledgeable with its layout and purpose.
Documentation (10%)	*9-10 points*: The code is extremely well documented. Comments are completely consistent with the associated code. There are no spelling errors.	*7-8 points*: The code is reasonably well documented. There are minor formatting omissions that would have improved users understanding of code purpose. There may be limited spelling errors.	*5-6 points*: The code is marginally documented. There are significant portions of the code that are not documented or documented incorrectly. There are significant spelling errors that detract from the documentation.	*1-4 points*: The code is poorly documented. There are minimal comments and/or the comments are incorrect.	*0 points*: The code is not documented.

Content rubric (50%)

Code	Excellent	Good	Satisfactory	Poor	Unacceptable
Visualizations (15%)	*13-15 points*: The visualizations are relevant to the question(s) of interest and use effective visual cues. All graphical variable types used are well-suited for the type and scale of data they represent, and there are clear labels and legends.	*10-12 points*: The visualizations are relevant to the question(s) of interest and use effective visual cues. All graphical variable types used are well-suited for the type and scale of data they represent. Labeling and legends are not completely clear.	*7-9 points*: The visualizations are relevant to the question(s) of interest, but visual cues neither enhance nor detract from the main message. Most graphical variable types used are well-suited for the type and scale of data they represent. Labeling and legends not completely clear.	*4-6 points*: The visualizations are relevant to the question(s) of interest, but visual cues are ineffective or not utilized. Some graphical variable types used are wellsuited for the type and scale of data they represent. Unclear labeling and/or legends.	*0-3 points*: There are no visualizations, or the visualizations are not relevant to the question(s) of interest.
Table (10%)	*9-10 points*: The table displays summary statistics relevant to the question(s) of interest, is easy to read, has clear labels, and is nicely formatted. Significant figures are consistent and appropriate.	*7-8 points*: The table displays summary statistics relevant to the question(s) of interest and is easy to read. Some labels are not clear. Significant figures are consistent and appropriate.	*5-6 points*: The table displays summary statistics relevant to the question(s) of interest. Some labels are not clear. Some significant figures are inconsistent and/or inappropriate.	*2-4 points*: The table displays summary statistics relevant to the question(s) of interest. Labeling is absent or incorrect. Format is hard to read. Some significant figures are inconsistent and/or inappropriate.	*0-1 point*: There is no table, or the table is not relevant to the question(s) of interest.
Written summary (15%)	*13-15 points*: Written report is organized appropriately, and addresses all relevant questions correctly. Grammar and spelling are virtually error-free.	*10-12 points*: Written report is organized appropriately, and addresses most relevant questions correctly. There are few errors in grammar or spelling.	*7-9 points*: Written report is organized appropriately, and addresses some relevant questions correctly. There are few errors in grammar or spelling.	*4-6 points*: Written report is poorly organized with missing components, and few relevant questions are addressed correctly. There are few errors in grammar or spelling.	*0-3 points*: Visualizations and table were not accompanied by written explanations and/or language used impedes meaning because of errors in usage.
Reflection (10%)	*9-10 points*: Reflection addresses all questions thoughtfully and thoroughly. Length is appropriate. Grammar and spelling are virtually error-free.	*7-8 points*: Reflection addresses most questions thoughtfully and thoroughly. Length is appropriate. There are few errors in grammar or spelling.	*5-6 points*: Reflection addresses some questions thoughtfully. Length is appropriate. There are few errors in grammar or spelling.	*2-4 points*: Reflection fails to address most questions. Length is much longer or much shorter than 1 page. There are few errors in grammar or spelling.	*0-1 point*: No reflection is included, or does not address any questions.