Requiem for a Balanced Life: Analysis of My Behavioural Patterns as a PhD Student*

Christina Wei

April 3, 2023

Various studies have shown that PhD students are having trouble with their work-life balance and coping with the stress of their workloads. To understand my own time management skills, I analyzed my timesheet data to understand my behavioural patterns over the first two semesters of my PhD journey. I found that I am spending about 40 hours a week, with most of my time spent on research, course work, and teach assistant related activities. There is evidence of improper work-life balance based on my tendency to work in late evenings as well as on weekends. To strive for better balance, I have outlined several suggestions to purposefully plan for each day, as well as created prediction models to estimate and plan for upcoming projects.

Table of contents

---

*Code and data supporting this analysis is available at: https://github.com/christina-wei/PhD-Time-Analysis.git

1

1 Introduction

Pursing a PhD degree can be an intellectually fulfilling journey to be at the forefront of academic knowledge. However, this can also be a difficult journey, as many students are struggling to cope with the stress of the work. Based on a study conducted on more than 3,500 PhD students in Belgium, the authors found that 50% ofthe students experienced psychological distress during their PhD, with more than 30% were at risk of developing a psychiatric disorder, especially depression (Levecque et al. 2017). Another study shows that around 70% of students do not perceive themselves having good work-balance balance, with 90% of them reporting an imbalance in their personal life (Balachandran 2021). Reading these cautionary tales, it makes me question myself on whether I am striking a good work-life balance in my PhD journey.

While I feel like I have a good handle on my stress level and actively managing my work-life balance through time management, I want to quantitatively assess whether I am achieving these goals. Since I have the habit of tracking my time spent on academic work since starting as a PhD student at the University of Toronto, I decided to analyze my timesheet data to answer the question on whether I have a good work-life balance in my PhD journey. For this analysis, I am using the amount of time I spent on academic work as the estimand for proportion of work in my life, while deducing the non-work related aspects as the time not tracked as academic work.

This analysis has provided me with various insights into my behavioural patterns as a PhD student, as well as areas of improvements to strive for better work-life balance. Over the last 8 months, I spent the majority of my time on research, source work and teaching assistant related activities, working on average about 6 hours a day. I prefer to work in sessions around 2 hours in length, focusing on 2-3 topics each day. As for work-life balance, there are definitely areas of improvement as I tend to work in late evenings (past 9pm), as well as significant amount of time on weekends. To help myself manage my time better, I built a couple of prediction models to estimate the number of hours or the completion dates for projects based on past behavioural patterns.

In the remainder of the paper, will first discuss in the Data section the characteristics of the timesheet data and the data cleaning procedure applied to it, as well as limitations of this data source. Next, the Results section presents trends and findings in behaviour patterns related to individual sessions, daily efforts, as well as trend over time. The following Model section outlines the prediction models used to calibrate and estimate the number of hours available or the completion dates for a given set of project parameters. Lastly, the paper wraps up with a Discussion section sharing my reflections adjusting from full-time job to a PhD student, getting in the flow of work, concrete suggestions to improve work-life balance, and future work to improve this analysis.

2 Data

2.1 Source & Data Cleaning

The timesheet data used for this analysis is based on task tracking from August 15, 2022 to March 13, 2023. I used the an mobile application Toggl Track[1] on my phone to keep track of my academic activities every day. There are in total 581 time entries tracked for the given time period.

The following information is tracked for each time entry:

- `Description` of the activity that is being tracked

- `Project` where the activity belongs to, such as class code (e.g. INF3104) or research project (e.g. Lit Review). Sometimes this value is not populated in the timesheet

- `Start date`, `start time`, `end date`, and `end time` of the tracked activity

The raw timesheet data is exported directly from the Toggl Track application. I applied data cleaning routine to include additional information on the type of work, as well as the effort duration in hours for each entry. The `type` of work is defined as:

---

[1]https://toggl.com/track/

Table 1: Sample of the timesheet data

project	description	start_date	start_time	end_date	end_time	duration	type	effort_hours
INF2169	paper 1	2023-03-10	21:32:56	2023-03-10	22:37:28	01:04:32	Course Work	1.1
INF3104	paper 3	2023-03-12	18:08:08	2023-03-12	20:38:04	02:29:56	Course Work	2.5
INF2169	paper 1	2023-03-12	21:28:36	2023-03-13	00:30:24	03:01:48	Course Work	3.0
Talks	AI as Financial Infrastructure	2023-03-13	10:00:00	2023-03-13	11:30:00	01:30:00	Learning	1.5
INF2169	class	2023-03-13	11:45:00	2023-03-13	13:00:00	01:15:00	Course Work	1.2
INF3104	class	2023-03-13	13:10:00	2023-03-13	16:10:00	03:00:00	Course Work	3.0

- Course Work - time spent on course related activities, like attending classes, working on assignments

- TA - time spent on teaching assistant obligations, like marking assignments and

- Research - activities related to research, including reading papers, thematic analysis, and writing

- Admin - administrative tasks like applying for grants, filling in timesheets

- Learning - learning activities outside of coursework, like workshops, writing help, orientations

- Other Activities - catches miscellaneous activities like lab meetings, updating personal websites

Data was cleaned and analyzed using the open source statistically programming language R (R Core Team 2022), using functionalities from `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2022), `readr` (Wickham, Hester, and Bryan 2022), `tibble` (Müller and Wickham 2022), `here` (Müller 2020), `kableExtra` (Zhu 2021) and `knitr` (Xie 2014).

Please see Table 1 for a sample of the timesheet data.

2.2 Data Limitations

As the timesheet data depends on my action to manually tracked my time accurately, there is definitely the potential for *human errors* in the data. Specifically, there may be missing data that I forgot to enter into the application. Also, some data are entered retrospectively after the actual activity happened, which could impact the accuracy of time tracked. Also, there are potential issues of tracking time to the wrong project or description that's not reflective of the actual activity.

The selected *granularity* of data used in timesheet tracking is also a potential limitation. There are aspects that are not currently captured which may be useful for analysis. For example, data can be enriched by tracking productivity, location, and more details on the activities for each session. Some of these attributes are not tracked because they do not have a field on the

application to be entered. But even if the additional fields are available on the app, there is also the consideration to balance the ease of performing data tracking vs. the amount of data to be tracked. If the act of tracking data becomes cumbersome to do, I may stop tracking time all together to avoid the hassle, which results in more loss in data than not having more data attributes.

There is also an interesting limitation on the *consistency* of project categorization over time as more knowledge is gained on how to efficiently organize projects. For example, my one-on-one meetings with my supervisor were not categorized into any projects initially because I wasn't sure where to put them. After a few months, I created the project "Research-General" to capture general activities related to research, and have tagged future one-on-one meetings into this category. For this particular situation, the data cleaning routine moved the untagged 1:1 meetings into the "Research-General" project. However, there may be other activities with different projects over time that were not adjusted.

3 Results

3.1 Type of Work

From August 15, 2022 to March 13, 2023, I have tracked 1034 hours over 581 sessions. Looking at Figure 1, most of my time are spent on Research (398 hours, 38%) and Course Work (343 hours, 33%) type of work. Performing teaching assistant duties takes about 12% of my time with 124 hours. The remaining 17% of my time are pretty equally split across Learning, Admin, and Other Activities.

To understand better the time spent on Research and Course Work, let's drill down to the project level details. Looking at Research time specifically, Figure 2a shows that a significant amount of time is spent on working on the literature review paper (Lit Review) at 253 hours, which is the first research paper that I worked on as the lead author. The marginalization in financial industry (Finance CA Marginalization) is a workshop paper I authored, where I spent 21 hours working on the manuscript. The projects on older adults' perception on conversational agents (OAPerception), industry digital design marginalization (Industry DDN) and perceptions of conversational agents with disfluency (Disfluency perception) do not have complete data, as work done before August 15, 2022 was not tracked. The "Other" category shows that I spent 21 hours in the tracked period on general research activities, such as reading literature to keep up to date with research.

Looking at time spent on Course Work in Figure 2b , there are two courses that are taking significant amount of time: INF2241 Critical Making at 142 hours and INF3104 Data Science Foundations at 103 hours. Both of these courses are technical in nature, as the course work requires both programming skills as well as written communication skills to summarize findings. It is worth noting that INF3104 and INF2169 are courses that I am currently taking, and there will be more time spent on them that will not reflected in this analysis.
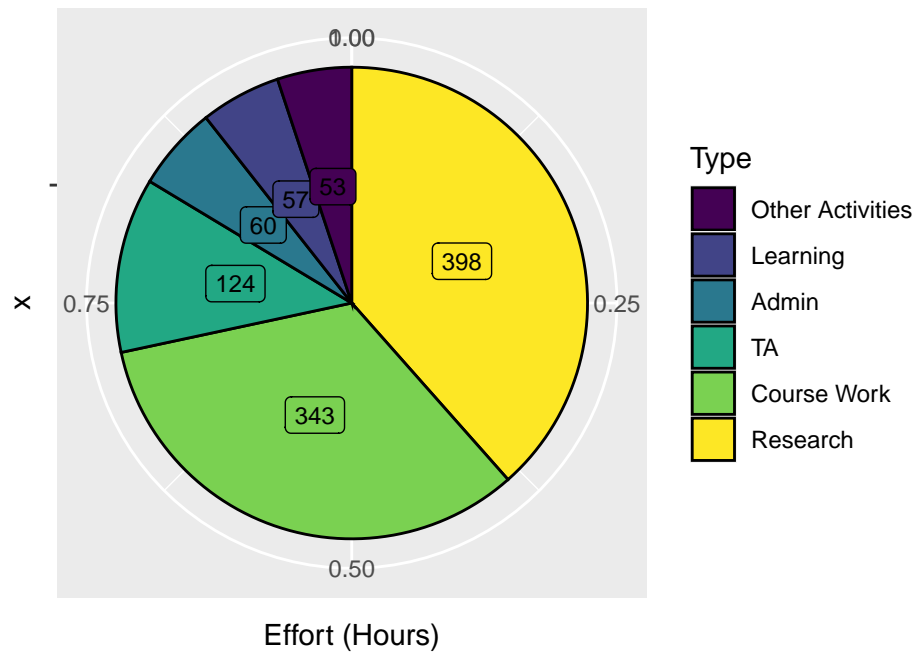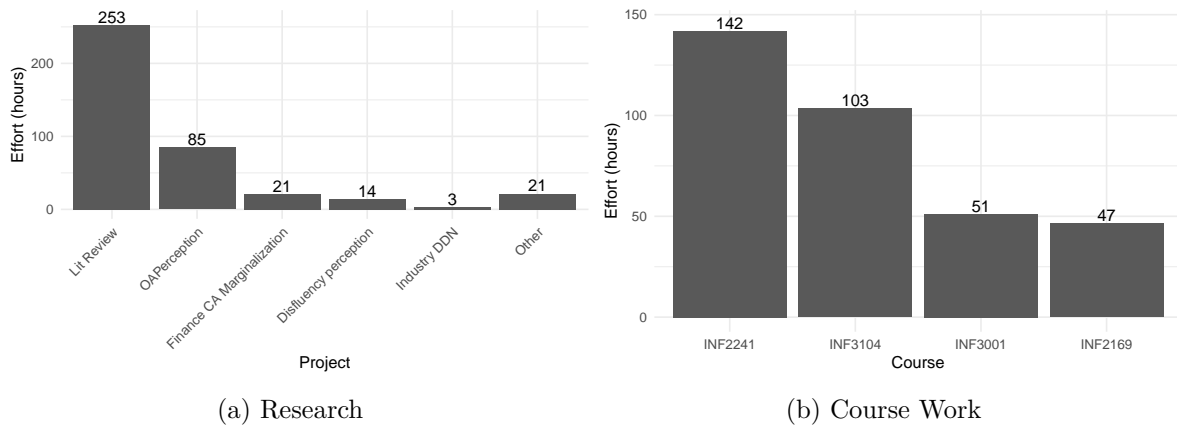
Figure 1: Amount of time spent on each type of activity

(a) Research

(b) Course Work

Figure 2: Break down time spent for Research and Course Work

Table 2: Summary statistics for each session

Avg Session Time	SD Session Time	Min Session Time	Max Session Time
1.8	1.1	0.2	10.7

3.2 Session Patterns

The average time spent on each session is 1.8 hours, with a standard deviation of 1.1 hours. The minimum amount of time spent on a session is ~10 minutes on TA marking activities, while the maximum amount of time spent in one session is 10.7 hours working on paper 1 for INF3104. (see Table 2 for summary statistics)

Looking at the **length of time** spent per session, Figure 3a shows that the typical length of sessions are less than 3 hours (87%), with most of them within the 1 to 2 hours range (41%). It is rare to have sessions over 4 hours long (<5%). Breaking down the sessions by type of work, we see in Figure 3b that the medians of the session for each type are not significantly different, and are within the range of 1-2 hours per session. Course work and other activities have some large outliers in data that are over 6 hours long.
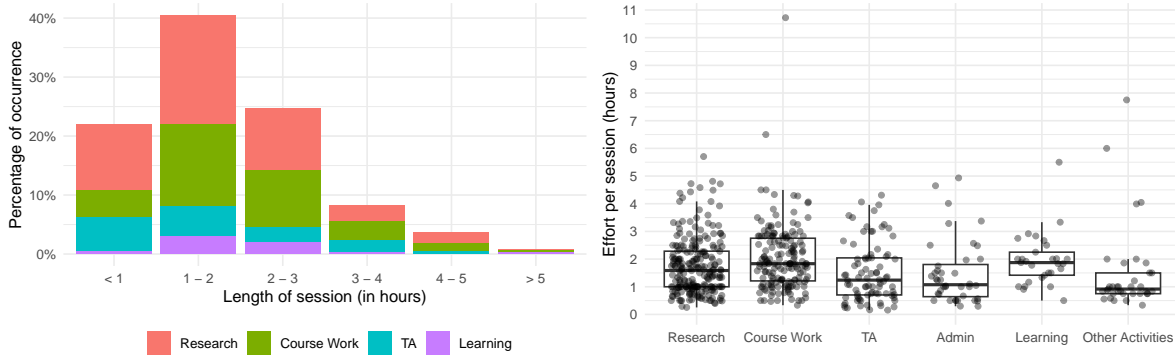
(a) Percentage sessions for different lengths of time   (b) Distribution of session length by type

Figure 3: Visualize sessions by length of time

Let's take a look at the outlier sessions (>4 hours) for each type of work (Table 3):

- *Research* - there are 11 outlier sessions, clustered mostly between 4-5 hours, with the maximum length at 5.7 hours

- *Course Work* - there are 9 outlier sessions, with a couple of large outliers at 10.7 hours and 6.5 hours. All the outlier sessions are related to courses INF2241 and INF3104

- *TA* - there are 2 outlier sessions, both are related to assignment mark with length around 4 hours

Table 3: Summary statistics for sessions that are over 4 hours long

Type	Count	Avg Effort	Max Effort	Tasks
Research	11	4.6	5.7	data analysis, paper writing, paper review
Course Work	9	5.2	10.7	assignments from INF2241, INF3104
TA	2	4.2	4.3	CCT419 (UX and board games) marking
Admin	3	4.5	4.9	NSERC application, SRI scholarship
Learning	1	5.5	5.5	Critical Computing - qualitative analysis workshop
Other Activities	3	5.9	7.8	Data Science Hackathon, update resume

- *Admin* - there are 3 outlier sessions, all of them related to scholarship applications

- *Learning* - the 1 outlier is the Critical Computing workshop that lasted 5.5 hours

- *Other Activities* - there are 3 outlier sessions, two of them (6 and 7.5 hours) are related to the Data Science Hackathon that I participated as a mentor for full days. The other outlier is spending 4 hours updating my resume

To understand the patterns for **hours of the day** that I am working, Figure 4 visualizes the percentage of sessions that are occurring during each hour of the day. Overall, it looks most of the working time are between 8am and midnight, with two periods of time that have the most frequent number of work sessions: between 11am to 4pm, and between 9pm and midnight. The earliest I started working is around 8am, and the latest that I finished work is around 4am. As for hours of the day by different types of work, it looks like Course Work takes a larger proportion of sessions in the early afternoons (12pm - 3pm), and Research takes a larger proportion in late evenings (9pm to midnight).

3.3 Daily Patterns

The average time spent on work each day is 5.7 hours, with a standard deviation of 2.5 hours. The minimum amount of time that I worked in a day was 0.5 hours, while the maximum amount of time working in one day is 12.5 hours. On average I worked on 2 different projects each day. The minimum number of projects I worked on in a day is 1, while the maximum number of projects I worked in a day is 6 (see Table 4 for summary statistics).

3.3.1 Hours Worked Per Day

Looking at the distribution of the hours worked per day Figure 5a, most of the values are centered around the mean of 5.7 hours between 4 - 8 hours per day. There are some extra long days in the dataset, with close to 10 days that I have worked for over 10 hours per day. There are usually right before a specific deadline for submitting a grant application or a large
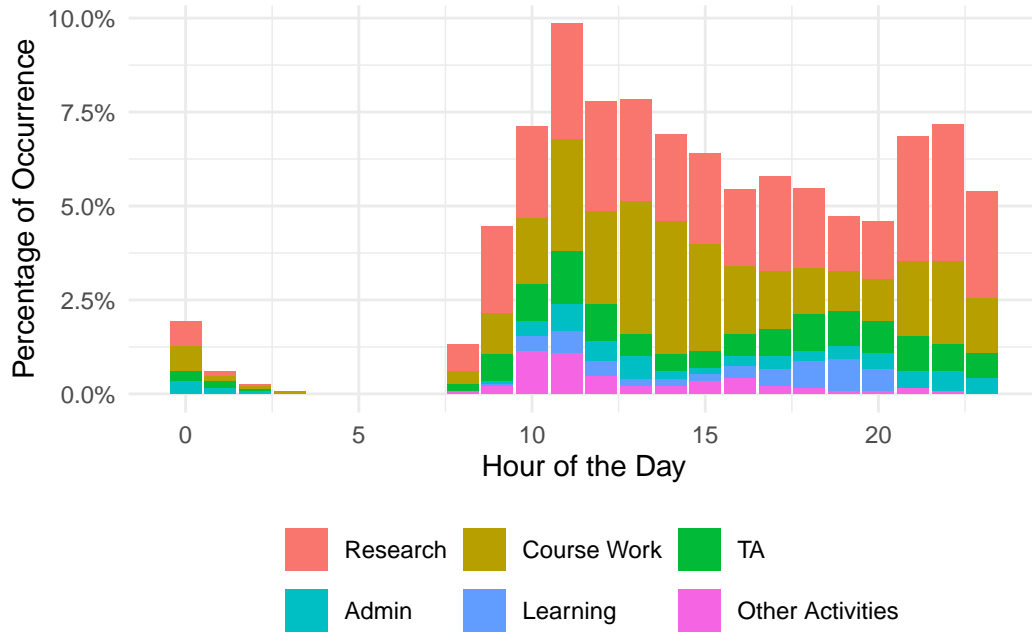
Figure 4: Percentage of sessions occurring during each hour of the day

Table 4: Summary statistics for each day on number of sessions and effort

Statistics	Hours Worked	Number of Projects
Mean	5.7	2
SD	2.5	1
Min	0.5	1
Max	12.5	6

assignment. Assessing the trend for hours worked on each day of the week in Figure 5b, it looks like I spend slightly more time working from Monday to Thursday with around 6 - 7 hours a day, a slightly less time working on weekends from Friday to Sunday with around 4 - 5 hours a day.
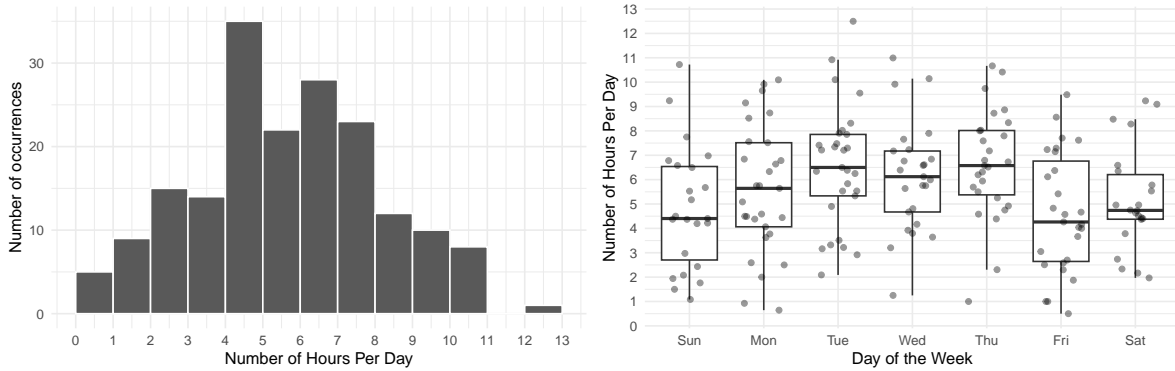

(a) Distribution of hours worked per day (b) Hours worked per day for each day of the week

Figure 5: Hours worked each day

3.3.2 Number of Projects Per Day

It will also be interesting to understand how many **different projects** I am working on each day. Analyzing my timesheet overall (Figure 6a), I am usually working on between 1 to 3 projects per day, with the maximum of 6 projects per day. Drilling down into the data, there are some differences in the number of projects based on the day of the week (Figure 6b). It looks like Saturdays and Sundays are mostly focused on 1 or 2 projects, while on weekdays I work on a higher number of projects per day, mostly between 2 to 3 projects.

As a student who is currently in the course work portion of my PhD, my schedule is driven by the courses I am taking per semester. Therefore, I analyzed the data based for each semester to see if there are differences in the number of projects per day. Comparing the graphs between Fall 2022 (Figure 7a) and Winter 2023 (Figure 7b) semesters, there are noticeable differences during the weekdays between the two. In Fall 2022, Thursdays stand out as an outlier with a higher number of projects per day, with many of them working on 6 projects. In Winter 2023, Mondays are pretty consistent at 3 projects a day, while Fridays have a wider distribution of values. Weekends do not seem different between the two semesters, working on 1-2 projects per day.
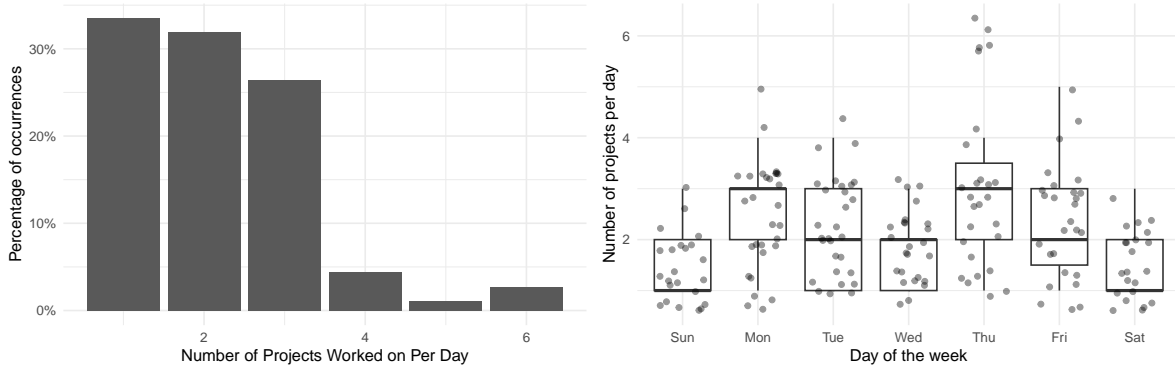
(a) Distribution of number of projects per day   (b) Number of projects per day for each day of the week

Figure 6: Numbers of projects worked on per day
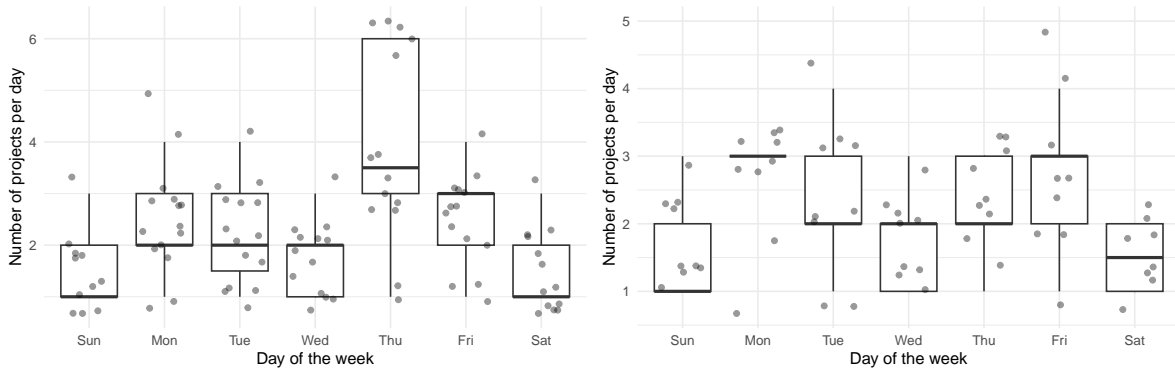
(a) Fall 2022 (Sep - Dec)   (b) Winter 2023 (Jan - Mar)

Figure 7: Numbers of projects worked on per day for each semester

3.4 Trend Over Time

Analyzing my weekly effort over time (Figure 8), I can see a slight upward trend over time as I am spending more time overall on academic matters. Most of the increase comes from efforts spent on research related work in 2023. Also, there are three different semesters that are captured here: Summer 2022, Fall 2022, and Winter 2023. Each semester seems to have a different pattern. In Summer 2022, I was spending about 20 hours a week mostly focused on research projects. In Fall 2023, the majority of my time are spent on course work and TA related activities. In this semester, Winter 2023, most of my time are spent on research and course work activities. There is also a noticeable decrease in time in December, where I took a break over the holidays.
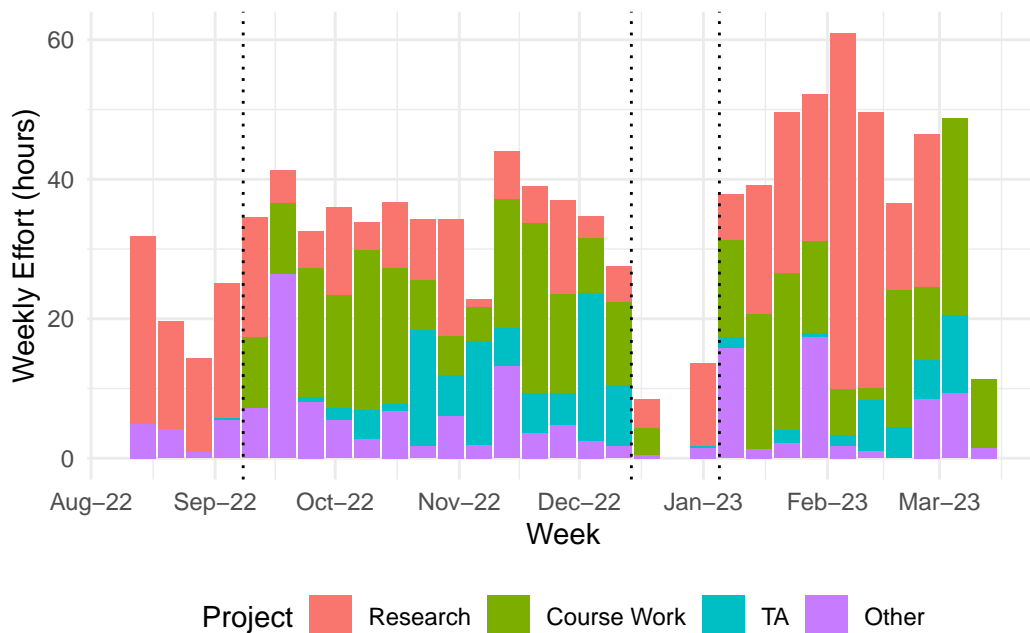

Figure 8: Weekly effort over time

3.4.1 Course Work & TA

As both course work and TA are structured based on each semester, their trends are analyzed together in this section. Looking at the trend over time for course work for both Fall 2022 and Winter 2023 semester in Figure 9a, it looks like there is a pattern where there are more work at the beginning and the end of the semester, with some down time in the middle. As for TA work (Figure 9b), while the workload is mostly steady at a few hours a week, there are noticeable peaks when a large amount of time is needed for fast turnaround on deliverables.
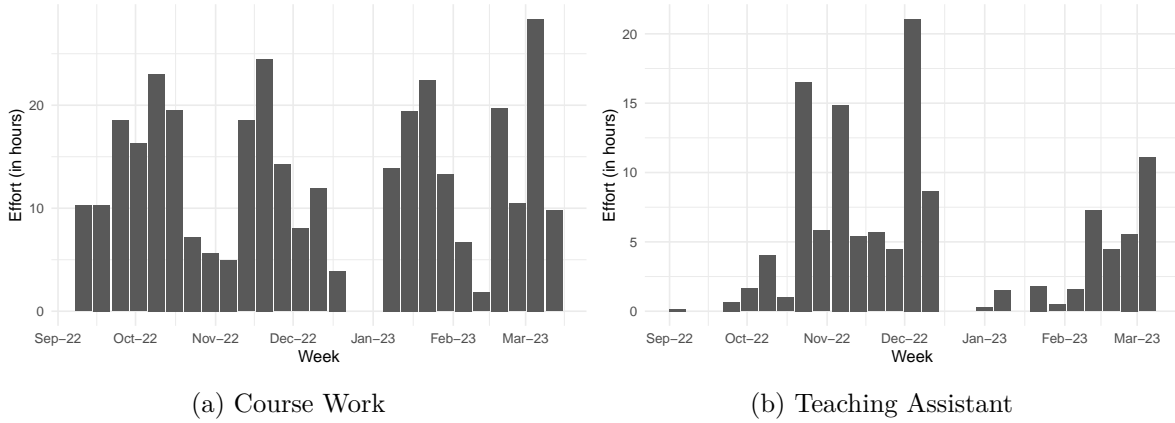
(a) Course Work

(b) Teaching Assistant

Figure 9: Weekly effort for differe type of work

Table 5: Summary statistics for weekly effort on research related activities

Mean	SD	Min	Max
14.2	11.3	1.1	51

3.4.2 Research

Since August 2022, I have been spending on average of 14.2 hours per week on research related activities. There is high variation (standard deviation = 11.3 hours) in the weekly effort on research, with minimum of 1.1 hours and maximum of 51 hours (see Table 5). The variation is correlated with the submissions deadlines for conferences and journals, as the number of hours spent on research increases dramatically in the weeks leading up to the deadline (Figure 10). One good example is the time spent on the systematic literature review paper (Lit Review), where there are a couple of weeks with over 40 hours before the deadline.

Based on the graph in Figure 10, there are usually only 1 or 2 research projects that I'm working on at a given time. As a research project finishes, then I can take on new projects for research. This rolling schedule can be seen with the OA Perception project as the main focus in August and September. After this project is finished, the next major project, Lit Review is started in October and continue to be the main focus until end of February. It is also interesting to observe the differences between being the lead author on Lit Review where there are high peaks of writing time before submission versus a contributing researcher on OA Perception where it is a steady amount of work before the deadline.
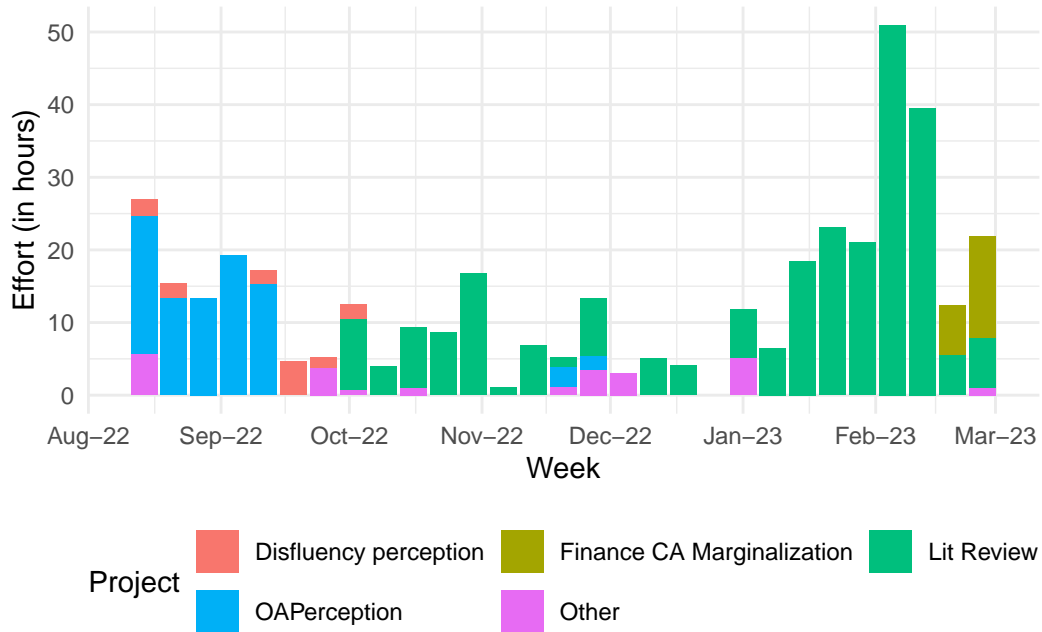
Figure 10: Weekly effort for differe type of work

4 Model

There are two prediction models created to help me with time management: one is to predict the number of hours available working on a deliverable based on a given targeted delivery date; the other is based on the estimated effort in hours to project the estimated completion date. Also, a calibration routine is used to control the model parameters used in calibration. These models assume that weekly patterns of hours spent working per day in the past is representative of future behavioural working patterns.

4.1 Calibrate Model Parameters

The model parameter used in the prediction models are based on the average hours spent working for each day of the week. There are some optional inputs that users can use to filter timesheet data: using start date and end date to select a specific period, or to specific the type of work (e.g. Research) to be used in calibration. The timesheet data is then used to compute the average number of hours per day, which is then summarized by calculating the average hours for each day of the week.

Here is an example of invoking the calibration routine based on all available timesheet data with the output model parameters in Table 6.

Table 6: Sample of model parameter output with no filtering criteria

wkday	avg_effort
Sun	4.8
Mon	5.6
Tue	6.5
Wed	6.1
Thu	6.6
Fri	4.6
Sat	5.2

Table 7: Sample of model parameter output filtered from Jan 9 to Apr 1, 2023 and Research

wkday	avg_effort
Sun	4.0
Mon	2.8
Tue	3.7
Wed	4.6
Thu	3.8
Fri	2.1
Sat	3.7

```
calibrate_model()
```

Here is another example of invoking the calibration routine based on timesheet data from January 9 to April 1, 2023, using only entries related to Research type of work. The corresponding model parameters are shown in Table 7.

```
calibrate_model(as.Date("2023-01-09"), as.Date("2023-04-01"), "Research")
```

4.2 Predict Hours Available Between Dates

This function is used to estimate the number of hours I have available to work on a project if I have a target completion date in mind. As such, the input parameters for this function are defined as the following:

- `model_param` specifies the calibrated number of hours worked for each day of the week (e.g. Table 6)

- `start_date(optional)` specifies the start date of the project. If not specified, it will default to today's date

- `end_date` specifies the target completion date of the project

- `pct_effort(optional)` specifies the multiplier to be applied to the calibrate hours in model parameters. For example, specifying 1.2 here will scale up the calibrated hours by 120%.

- `include_weekends(optional)` specifies whether the estimated hours should include Saturdays and Sundays. If not specified this value is default to be TRUE to include weekends in prediction.

To estimate the number of hours, the algorithm is projecting the list of dates that are available between the start date and the end date, calculating the number of days available for each day of the week (e.g. 3 Mondays, 2 Tuesdays ..), multiplying them by the average number of hours for each day of the week, then scaling them up by the `pct_effort` parameter to generate the final estimate.

$$total\_hours = \sum_{i=1}^{N} \beta_i * n_i * \alpha \qquad (1)$$

The parameters in Equation 1 are:

- i = day of the week (e.g. 1 = Mon, 2 = Tues ..)

- N = 1 to 7 if include_weekends = TRUE; 1 to 5 if include_weekends = FALSE

- β_i = average hours for the day of the week

- n_i = number of the day of the week (e.g. 3 Mondays)

- α = pct_effort multiplier value for effort scaling

For example, the following code snippet estimates the number of hours I have available between now and April 17 based all the timesheet data available. Based on the prediction model, I have 106 hours available for work, and it also generates a detailed daily schedule for my review (sample schedule in Table 8).

```
model_param <- calibrate_model()
predict_hours_available(model_param, end_date = as.Date("2023-04-17"))
```

Table 8: Sample output of detailed daily schedule from predict hours available algorithm

date	wkday	hours
2023-04-03	Mon	5.6
2023-04-04	Tue	6.5
2023-04-05	Wed	6.1
2023-04-06	Thu	6.6
2023-04-07	Fri	4.6
2023-04-08	Sat	5.2

4.3 Predict Completion Date

On the other hand, I can also predict the estimated completion date if I have an estimated number of effort for a project. The algorithm of this prediction is similar to predicting the hours available between two date by using the average number of hours worked for each day of the week as the baseline. Specifically, the input parameters for this function are defined as the following:

- `model_param` specifies the calibrated number of hours worked for each day of the week (e.g. Table 6)

- `start_date(optional)` specifies the start date of the project. If not specified, it will default to today's date

- `effort_hours` specifies the estimated number of hours for the project

- `pct_effort(optional)` specifies the multiplier to be applied to the calibrate hours in model parameters. For example, specifying 1.2 here will scale up the calibrated hours by 120%.

- `include_weekends(optional)` specifies whether the estimated hours should include Saturdays and Sundays. If not specified this value is default to be TRUE to include weekends in prediction.

To estimate the completion date, the algorithm generates a sequence of daily dates from start date based on criteria of whether to include weekend dates, then populating each date by the estimated hours for each day of the week multiplied by the `pct_effort` parameter. When the cumulative sum of the daily efforts exceeds the estimate number of hours for the project, then the routine is stopped with the final date in the sequence returned as the estimated date of completion.

As an example, I would like to use this prediction model to help me estimate the date that I will be completing the Final Paper for the Data Science Foundations course. I would like to use this semester's timesheet related to course work to estimate my behavioural patterns. I

Table 9: Sample output of detailed daily schedule from predict completion date algorithm

date	wkday	hours	cumulative
2023-03-31	Fri	1.6	1.6
2023-04-03	Mon	2.0	3.6
2023-04-04	Tue	2.2	5.9
2023-04-05	Wed	2.1	8.0
2023-04-06	Thu	1.8	9.8
2023-04-07	Fri	1.6	11.4

am estimating the remaining effort on the paper is 10 hours, and I can spend about 50% of my total efforts working on this. Lastly, I do not want to work on weekends. The code snippet below demonstrates the translation of my requirements into parameters for calibration and prediction models. The model predicts that I will be able to complete the assignment on April 7, 2023 with the detailed schedule generated in Table 9.

```
model_param <- calibrate_model(
  from_date = as.Date("2023-01-09"),
  to_date = as.Date("2023-03-30"),
  work_type = "Course Work")

predict_completion_date(
  model_param,
  effort_hours = 10,
  pct_effort = 0.5,
  include_weekends = FALSE)
```

5 Discussion

The first

5.1 Adjustment Going from Industry into Academia

Since I was working full time for the last 15 years, I am quite familiar with the workday structure of 8 hour working days from Monday to Friday. Based on my previous schedule at working full time, I was spending on average about 10 hours a day, which is about 50 hours per week. While the work can be tiring sometimes, it was generally manageable and I was able to find a balance between work and life.

Comparing between the workload between full time work and as a PhD student, I feel that it is harder as a PhD student to balance the work with personal life. This is evident in the timesheet analysis that working on weekends is a common occurrence (Figure 5b). Based on my timesheet data, I am surprised to see that most weeks I am spending slightly less than 40 hours a week (Figure 8), which time-wise is less than the ~50 hours a week that I did in my full-time job. If I am quantitatively working less hours as a PhD student, why does it feel more tiring? There are a few potential contributing factors to this: level of familiarity, type of work, and mixture of cognitive demands in tasks.

First, I am still learning the skills needed in order to excel in my PhD journey. As I have not done academic research previously, I need to learn foundational knowledge associated with it, as well as applying them to the projects that I am working on. There are many aspects of new learning associated with research: how to think about a research problem, how to synthesize findings, how to write a paper suitable for a conference to name a few. Compare this to my full time job, I am very familiar with leading a product development team to implement new applications in the financial industry. When faced with new problems, I can derive based on past experiences on different ways to approach and resolve them. The new learnings associated with being a PhD student feels unfamiliar, therefore it feels more tiring compare to previous work-related activities.

Secondly, there are fundamentally different types of work done as an academic researcher vs. industry practitioner. Academic research is more of a creative process, to seek out domains that have not been explored before and to generate new macro knowledge fo fill these gaps. The process can be unstructured and lacking a sense of achievement as it is not a linear progress. Industry practitioners, specifically those on the implementation side, are generally applying existing knowledge to specific use cases for their applications. There is also knowledge creation in the process, but it is usually at the micro level to adapt general knowledge towards specific applications. There are usually concrete deliverables defined and it is easier to have a sense of achievement based on this. The creative mindset of academia vs. execute mindset of industry may also be a factor for feeling that PhD work is harder, as it is difficult to grasp a sense of concrete achievement.

Lastly, there are different mixtures of tasks that I work on during the course of a day between PhD and full-time work. Based on this timesheet analysis, most of my time as a PhD student are spent on research, course work and and teaching assistant obligations. All of these are mentally demanding tasks that require my full attention. On the other hand, full-time jobs have some activities requiring lower mental engagement scattered throughout the day, like attending general information meetings (e.g. town halls), daily stand-up meetings, and reviewing customer feedback. These activities give me a mental break from highly cognitive work, reducing the feeling of tiredness.

5.2 Getting in the Flow

There are some evidence in the data for work sessions where I have gotten into the flow and worked for a prolonged period of time. It is interesting to observe some patterns on the type of work in the sessions where I have worked for over 4 hours (Table 3). Specifically, work that are technical in nature as well as paper writing and review are most common in longer sessions.

Some of the longer sessions came from working on technical assignments from INF2241 Critical Making and INF3104 Data Science Foundations courses. The longest session recorded at 10.7 hours was working on the first assignment analyzing COVID clinics across Toronto. Also, some technical aspects of research are also longer in the length of work, such as thematic analysis for coding and data analysis of literature reviews. Based on my personal experience, I sometimes experience a deep sense of involvement working on technical problems, where I am following a particular train of thought and do not want to be disturbed until it can be traced to a satisfactory conclusion.

Another aspect for getting into the flow for longer sessions are related to paper writing and paper reviews. Especially for paper writing, it is interesting that many times I have experienced writer's block where I am forcing the words to come out, but sometimes the words are flowing freely with fully formed thoughts and well crafted sentences. I usually do not know which state of paper writing I will be engaged in until about half an hour into the writing process. Paper reviews on the other hand usually have better flow, where I prefer to dedicate several hours to read through the whole manuscript reviewing for cohesiveness overall as well as sentence level revisions.

5.3 Improving Work-Life Balance

While overall I am satisfied with my efforts over the last 8 months as a PhD student and feel like I am progressing well in my studies, this analysis shows me some areas of improvement to be more efficient in my time management as well as achieving better work-life balance.

First, I can be more efficient with my time by defining a couple of topics to focus on at the beginning of each day, and engage in work sessions that are no longer than 2 hours in length. This is in line with my behavioural patterns, as I tend to have study sessions around 2 hours length (Table 2), and focus on 1 to 3 projects per day (Table 4) . The caveat to this is that if I get into the flow of the work, then the work sessions can be longer as not to disturb the deep engagement. Also, time planning should take into account of the scheduled activities of the day (e.g. attending classes) to plan for topics and work sessions around these fixed time slots.

Second, the analysis shows that I have been spending a significant amount of time working on weekends (Figure 5b) as well as working late into the evenings (Figure 4). This pattern is not ideal for work-life balance, as there are no defined breaks within my schedule for work-life balance. Also, evenings and weekends are usually prime times for social interactions with

family and friends. The tendency to work during these times will have an impact on this important aspect of my life. To improve upon this, I should plan my working hours to be between 9am and 8pm Mondays to Fridays, leaving the evenings and weekends to decompress and to engage in leisure or social activities. This will require better time management skills but it is a worthy goal to aim for.

Lastly, an analysis of my time allocations periodically can be an useful way to reveal interesting patterns in my behaviour and offer insights for continuous improvement. As student life can be different semester by semester as demonstrated in Figure 7, it will be useful to perform this analysis at the end of each semester to discover consistencies or changes in trends as well as to calibrate the prediction models used for estimating hours and completion dates. This is needed to detect potential regime changes in the data. For example, after I complete the course work portion of my PhD and move into the preparing for qualifying exam phase, my behaviours may be changing dramatically. The prediction model should only be calibrated with data that reflects my future behavioural patterns.

Based on my current study patterns, I am proposing the following daily schedule (Table 10) to be followed on weekdays. This will achieve 37.5 hours of work time each week with each session between 1.5 and 2 hours and adequate breaks in between them, without needing to work in late evenings or weekends.

Table 10: Proposed Schedule for Weekdays

Time	Duration (hours)	Activity
09:00 - 11:00	2	Study session #1
11:00 - 11:30	0.5	Break
11:30 - 13:00	1.5	Study session #2
13:00 - 14:00	1	Lunch
14:00 - 16:00	2	Study session #3
16:00 - 17:00	1	Break
17:00 - 19:00	2	Study session #4

5.4 Future Work

While this analysis has provided many interesting insights I can use to manage my time as PhD student, there are some improvements I can made to this analysis that can make it more insightful. First, I can incorporate social engagement data into the analysis to have a quantitative measure of work-life balance. This can be done by incorporating my Google calendar details where I plan for social activities. Also, it will be interesting to incorporate the activities data from my FitBit into the analysis, such as physical activities and sleep schedules. Both of these sources will bring in additional perspectives to analyze my time from a more holistic point of view rather than just focused on academic activities.

The prediction models can also be improved by building more nuances into their algorithms. One simple improvement to the estimation is to account for statutory holidays and planned vacations in the model. Also, as additional sources of data related to social and physical activities are incorporated into the analysis, the model can be enhanced to propose a schedule suitable to achieving academic excellence and maintaining a healthy work-life balance.

6 Conclusion

Pursing a PhD can be an exciting journey filled with self-improvement and new learnings, but it can be a stressful period of time, causing improperly balance between work and life. Having proper time management skills and reflecting on the allocation time can be used to mitigate against the stress and strive for a better work-life balance. By analyzing into my own timesheet data tracking activities since starting my PhD, it has provided me with a number of actionalble insights into my own behavioural patterns and areas of improvements to incorporate into my daily practices. I am looking forward to leverage the prediction models to plan for upcoming work, as well as performing this analysis every semester to continuously improve my time management skills.

Appendix

An interface was built using Shiny to expose the model for calibrating and predicting number of hours available or completion date. Please see Figure 11 for a snapshot of the application.

Figure 11: Screenshot of the application for prediction models

Reference

Balachandran, Srija. 2021. "PhD Work-Life Balance – a Challenging Path." AgRefine. https://agrefine.eu/2021/11/02/phd-work-life-balance-a-challenging-path/.

Levecque, Katia, Frederik Anseel, Alain De Beuckelaer, Johan Van der Heyden, and Lydia Gisle. 2017. "Work Organization and Mental Health Problems in PhD Students." *Research Policy* 46 (4): 868–79.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2022. *Tibble: Simple Data Frames.* https://CRAN.R-project.org/package=tibble.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.