

N-OF-1: BETTER LIVING THROUGH SELF-EXPERIMENTATION

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri - Columbia

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
MATTHEW WAAGE
Dr. Edward Mirielli, Thesis Supervisor

MAY 2023

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

N-OF-1: BETTER LIVING THROUGH SELF-EXPERIMENTATION

presented by Matthew Waage,

a candidate for the degree of Master of Science in Data Science and Analytics,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Edward Mirielli

Professor Ilker Ersoy

Professor Timothy Haithcoat

DEDICATION

I'd like to thank my parents for giving me the foundation to build a good life, for giving me advice whenever I needed it, and for helping me in every way they can. I'd also like to thank my brother for being someone I can bounce ideas off of and giving me interesting things to think about.

ACKNOWLEDGEMENTS

I'd like to thank my committee members; Dr. Mirielli, Dr. Ersoy, and Dr. Haithcoat, and also Dr. Scott and all of the other great professors I had for furthering my Data Science knowledge and skills, and giving me the foundation for an entirely new career.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF TABLES	vi
LIST OF ILLUSTRATIONS	viii
ABSTRACT	ix
1 INTRODUCTION	1
2 LITERATURE REVIEW	5
2.1 N-of-1	6
2.2 Self-Experimentation and the Quantified Self	7
2.3 Wearables and Health Informatics	9
2.4 Time and Activity Tracking	9
2.5 Gratitude Journaling	10
2.6 Mood	11
2.7 Personalized Machine Learning of Depressed Mood Using Wearables	11
3 MATERIALS AND METHODS	13
3.1 Participant	13
3.2 Study and Experimental Procedures	14
3.3 Data	15
3.3.1 <i>Mood Data</i>	17
3.3.2 <i>Time Tracking and Activity Data</i>	31
3.3.3 <i>Fitbit Data</i>	52
3.3.4 <i>Vital Signs Data</i>	57
3.3.5 <i>TempoDisc Data</i>	61

3.4	Data Carpentry	61
3.5	Statistical Analyses	64
3.6	Machine Learning	71
3.6.1	<i>Personalized ML Feature Importance</i>	73
4	RESULTS	77
5	DISCUSSION	79
	BIBLIOGRAPHY	83
	APPENDIX	87

LIST OF TABLES

3.1	Mood Form Table Description	19
3.2	Mood Form Missing Data	22
3.3	Mood Form Table	23
3.4	Mood Form Statistics	23
3.5	Mood Form Table Full	23
3.6	Mood Form Statistics Full	23
3.7	Mood Form Percentages	23
3.8	Experiment 2 Phases	30
3.9	Activities Description	33
3.10	aTimeLogger Table	35
3.11	Time Data Table	35
3.12	Time Statistics	35
3.13	Activity Statistics	39
3.14	Activity Statistics Standard Deviation	40
3.15	Activity Statistics Total	40
3.16	Fitbit Description	53
3.17	Fitbit Data Table	53
3.18	Fitbit Statistics	53
3.19	Vitals Data Table	58
3.20	Vitals Statistics	59
3.21	TempoDisc Description	61
3.22	TempoDisc Data Table	62
3.23	TempoDisc Statistics	62
3.24	ANOVARM Data Table	69
3.25	ANOVARM Statistics	69

3.26	ANOVARM Results	69
3.27	ANCOVA Results	70
3.28	Mann-Whitney U Results	70
3.29	Machine Learning Results	73
1	PHQ-9/GAD-7 Questions and Scale	87

LIST OF ILLUSTRATIONS

3.1	Study Timeline	14
3.2	Experiment 2 Timeline	15
3.3	Overall Data Flowchart	15
3.4	Mood Data Flow	16
3.5	Mood Over Time	24
3.6	Mood Histogram	24
3.7	Mood Timeline Heatmap	25
3.8	Pain Over Time	26
3.9	Pain Timeline Heatmap	27
3.10	Weight Over Time	28
3.11	Weight Histogram	28
3.12	Weight Timeline Heatmap	29
3.13	PHQ-9/GAD-7 Data Flow	30
3.14	aTimeLogger Process	31
3.15	Time Data Flow	34
3.16	Activity Timeline	37
3.17	Sleeping Timeline	41
3.18	Sleeping Histogram	42
3.19	Sleeping Timeline Barchart	43
3.20	Sleeping Timeline Scatter	43
3.21	Sleeping Durations Timeline Heatmap	44
3.22	Working Counts Timeline Heatmap	45
3.23	Working Timeline	46
3.24	Working Histogram	47
3.25	Internet Timeline	48

3.26	Gym Durations Timeline Heatmap	49
3.27	Gym Timeline Scatter	50
3.28	Internet Timeline Scatter	50
3.29	Journaling Durations Timeline Heatmap	51
3.30	Step Count Over Time	54
3.31	Step Count Histogram	54
3.32	Step Count Timeline Heatmap	55
3.33	Heart Rate Histogram	56
3.34	Resting Heart Rate Over Time	56
3.35	Heart Rate Mean Over Time.png	56
3.36	Vitals Data Flow	57
3.37	Vitals Description	58
3.38	Blood Pressure Systolic and Blood Pressure Diastolic Over Time	59
3.39	Mean Arterial Pressure Over Time	60
3.40	Vitals Heart Rate Over Time	60
3.41	Respiration Rate Over Time	60
3.42	Ambient Temperature Over Time	62
3.43	Mood vs. Gym Durations Seconds	64
3.44	Step Count vs. Gym Durations Seconds	65
3.45	Step Count vs. Internet Durations Seconds	65
3.46	Mood Data Correlation Heatmap	66
3.47	Mood vs Time 7 Day Rolling Average Correlation Heatmap	67
3.48	Timeline of Mann-Whitney U Groups	70
3.49	Machine Learning Flow	72
3.50	SHAP Feature Importance	74
3.51	SHAP Correlation Tired Effect on Mood	75
3.52	SHAP Correlation Loneliness Effect on Mood	76
3.53	SHAP Correlation Gym Durations Seconds 7 Day Rolling Mean Effect on Mood	76
1	Fitbit API Data Flow Part 1	88
2	Fitbit API Data Flow Part 2	89

N-OF-1: BETTER LIVING THROUGH SELF-EXPERIMENTATION

Matthew Waage

Dr. Edward Mirielli, Thesis Supervisor

ABSTRACT

This project's aim was to create a platform for personalized health data analysis, testing, and prediction, making it easier for ordinary people who are interested in N-of-1 trials to do their own self-experiments and take control of their mental and physical health. In these studies a single subject is observed and different interventions are systematically evaluated on them over time. These are typically longitudinal, occurring over weeks or months, with several rounds of treatments and evaluations in the form of a number of AB assignments. In these studies wearable technology, trackers, apps, sensors, and other IoT devices may be used to record information about the subject multiple times per day or week, if not constantly.

In this study a singular self-experimenter collected data on themselves from several different sources such as mood questionnaires and a Fitbit wearable, among others. This data from the various sources was merged so that a variety of statistical methods could be performed. A few different modes of experimenting went into this study. One experiment tested the claim that spending 15 minutes per day writing in a gratitude journal had an effect on the subject's mood. This was achieved through a BABABA crossover phase design study, with each of the three phases being 28 days, for a total of 84 days in the experiment. The tests done for this experiment were the more traditional ANOVARM and ANCOVA, which were used to discover whether the intervention (B) phases were significantly different from the baseline (A), with relation to the subject's mood. Another test compared the claim that there was a difference in the subject's mood between the two groups of pre-experiment and during the experimental phase, through a Mann-Whitney U test. The last part of the study was a more complex machine learning (ML) pipeline that sought to predict the subject's mood based on over 3 years of daily collected data. The ML pipeline ingested the data, created several different ML models such as random forests and support vector machines, and compared which model was best at predicting the subject's mood. Feature importance was extracted from the best model through SHapley Additive exPlanations (SHAP), where the weight of the various feature effects on the target, in this case the subject's mood, was obtained. This notified the subject which behaviors had an effect on their mood. These different modes of experimenting were then compared, to see which was easier to implement or understand for future self-experimenters.

Keywords:

N-of-1 trials, self-experimentation, self-tracking, quantified self, personal informatics, single-case experimental design, machine learning, wearables,

Chapter 1

INTRODUCTION

This study fuses several different experiments together. While all within the data science field in general, involving longitudinal N-of-1 trials, and using overlapping datasets, these experiments examine the data in different ways. There was an AB crossover phase design experiment that sought to discover if gratitude journaling had an effect on the subject's mood, a study that sought to discover if the act of experimenting had an effect on the subject's mood by separating the periods of before and after the experiment began, and a machine learning study that sought to discover the features that contributed most to the subject's mood. At the end these different modes of experimenting and analysis were compared. All of these studies are viewed through the lens of the Quantified Self.

The Quantified Self is a fairly recent 'movement,' or community of ordinary people that record and analyze numerous aspects or facets of their lives to try and find any relationships that may be present so that they may be able to understand and improve themselves. These are typically conducted in an individualized manner throughout their daily life, this type of study being a subset of single-case experiments or N-of-1 trials.

In N-of-1 trials a single subject or patient is observed and different treatments or interventions are systematically evaluated on them over time. Typically, these are longitudinal studies occurring over weeks or months, with several rounds, sequences, or phases of treatments and evaluations in the form of a number of AB or BA assignments (at times with washout periods, wherein no interventions take place for a time, to return to a baseline state).

In the first part of this study the intervention was gratitude journaling, with the aim to find if this had an effect on the subject's mood, gratitude journaling being the habit of recording and reflecting on things that one is grateful for on a regular basis. Did the subject's mood improve on days they spent time gratitude journaling? In this experiment the intervention was spending 15 minutes per day journaling and thinking about things to be grateful for. This experiment was of the

BABABA variety, with each of the 3 phases of this experiment lasting 28 days for a total of 84 days. The second part of the study was conducted by performing a basic test to discover if there was a difference between the periods of before and after this experiment began. Does the simple act of running an experiment affect one's mood?

The main part of the study was where over three years worth of daily data on the subject was collected to find if there were any relationships between any of those features and mood using a machine learning pipeline. The aim was not only to predict mood scores, but also to identify the variables that contributed the most to those mood scores. The machine learning pipeline ignores some of the difficulty of setting up the experimental design with the AB phases, and simply compares whether any of the features collected were related to mood by a best-fit model. The ML pipeline ingested the data, created several different machine learning models such as random forests, support vector machines, and linear regressors, and compared which one was best at predicting the subject's mood through the mean absolute percentage error (MAPE). Additionally, feature importance was extracted from this best model through SHapley Additive exPlanations (SHAP), where the weight of the various feature's effects on the target, the subject's mood, was obtained, giving a way for the subject to change whichever behaviors may be affecting their mood the most. All of these different modes of self-experimentation were then compared, to see which was easier to implement or understand for future experimenters.

All of these studies use overlapping datasets. Typically in these studies wearable technology, trackers, apps, sensors and other Internet of Things (IoT) devices are used to record information about the subject multiple times per day or week, if not constantly. In this study the subject collected data on themselves from several different sources such as; mood questionnaires, time and activity tracking through a phone app, a Fitbit wearable wristwatch, a blood pressure machine, and a body temperature monitor, among others. All of the data from these various sources was then collated and combined so that the various statistical or machine learning methods could be performed.

Overall, this project's aim was to create a platform for personal health data analysis, testing, and prediction, making it easier for ordinary people who are interested in these types of N-of-1 trials to do their own self-experiments and take control of their mental and physical health, improving their lives. As this is an N-of-1 or single-subject experiment, any conclusions that can be drawn will only be applicable to that single subject, but the method can be applied to any other self-experimenter. The goal was to find clinically significant behaviors that affected the subject's mood so that they could change those behaviors or features.

This study draws on a lot of previous research in the areas of N-of-1 trials, single-case experimental

design, self-experimentation, self-tracking, quantified self, machine learning, and wearables. One previous study in particular is the central component for much of the machine learning section, where the authors, Shah et al., created personalized machine learning models for many individuals with depression [1].

There are other studies that have tried to create platforms for N-of-1 trials as well, generally in more specific or niche domains. Sleep is a common domain for self-experimenters to study and there have been several platforms created and studies exploring how to make it easier for them to do this such as the SleepCoacher app [2] which was analyzed and then improved upon with the help of other studies by the same author [3], [4].

A more general purpose system is the Physiqua platform [5], which was made so self-experimenters could collect data from several different sources and find trends between their wearables and questionnaires. Other more recent apps in this same vein have been created for similar purposes such as QuantifyMe [6] and Self-E [7]. The StudyMe [8] and StudyU [9] apps are for both self-experimenters and researchers, making it easier for the latter to create and share more valid and reliable experiments that the former may sign up for. This gives the researchers more subjects for their experiments and the self-experimenters more validity to their self-studying.

A general overview of N-of-1 studies [10] gives the reader instructions on how to perform these types of trials successfully. There are also extensive manuals for how to implement them, the most correct experimental design given a study's particular attributes, especially if they are of a psychological or medical nature, the statistics necessary to gain meaningful insights about the subjects, and any ethical concerns that should be considered [11], [12].

This was a multi-faceted but very interesting and worthwhile study. Studies of this nature are able to make it so that anyone has the tools to improve themselves and their lives. Some of the tools developed in this process are in the relatively new domain or industry of personalized medicine, which combines all of these different data sources for people to gain insights about their health and wellness, without the need for advanced statistical or computer programming knowledge. There is room for growth and advancement to make these tools even more available and understandable for the average person. The insights gained from the data and through these tools and techniques is only applicable to the individual self-experimenter, but the possibility is there in the future for experimenters to join a cohort and combine their data so that their results are more generalizable. As is always a concern when dealing with human subjects we should consider how the act of studying them and the tools used to do so is as ethical as possible. Due to this study being a self-experiment, the subject could stop at any time and for any reason.

Building on all of this previous research, our aim in this study was to answer several related questions:

- (1) For a single-subject self-experimenter, does gratitude journaling, compared to no intervention, improve their mood?
- (2) Does the mere act of self-experimenting affect the subject's mood?
- (3) Can a personalized machine learning model find the features that affect the subject's mood the most?
- (4) Are any of these modes of self-experimenting more or less difficult to perform or are the outcomes more or less difficult to understand than the others?

This study overall was just the latest iteration in a newer kind of personalized data science and analytics that uses wearables and questionnaires to help individual people, but there is much more to be explored. There is a wide-open future for more studies of this micro-sized scope which not only have the potential for incredible outcomes for individual people, and implications for not only health policies and hospital procedures, but also for interesting and previously unknown scientific interactions using massive amounts of data.

Chapter 2

LITERATURE REVIEW

This project is the latest at the intersection of several different domains of study, including psychology and psychiatry, medicine, personal informatics, and data science. Much of the previous research in these separate domains is essentially siloed from each other, making it difficult for advances in one to be adopted in another. This project attempted the interdisciplinary problem of trying to combine some of the compelling findings of these various domains and studies together into a single platform for personal health data analysis, testing, and prediction, making it easier for ordinary people who are interested in N-of-1 trials to do their own self-experiments so that they can take control of their mental and physical health and try to improve their lives.

There are several sub-components that, when in combination, improve upon each other and together create a more cohesive unit that can be used to improve people's lives in meaningful ways. In the areas of psychology, psychiatry, and medicine, we have rigorous research on best practices for N-of-1 or single-case experiments, the statistics necessary to verify that there was a worthwhile intervention, and that the subject was not adversely affected by the experiment. In psychology and psychiatry there is much research on what affects a subject's mood, and how gratitude journaling may improve one's life satisfaction. Wearable technology and collected personal health informatics can inform us about a patient's health and wellbeing. The tools and analytics provided through a data science framework can be used to improve on or automate much of the data collection and work that goes into all of these separate domains.

The initial lens through which this project was first conceived is from the 'Quantified Self' viewpoint, wherein typically non-academic data enthusiasts, biohackers, and coders, collect information about themselves to try and improve their health and wellness. These personal- or citizen-scientists want to understand and discover themselves through numbers, "self-knowledge through numbers," being a common phrase in the community. The motivation and drive for these self-experimenters

is one of self-improvement. The phrase “What gets measured, gets managed,” attributed to Peter Drucker, also touches on the ethos of the movement.

2.1 N-of-1

The main component and backbone of this project relies heavily on previous research into N-of-1 studies, also known as single-case experimental designs (SCED). Double-blind Random Control Trials (RCT) are typically seen as the gold standard in experimental design, but they have their limitations as well, such as catering to the average, not to each individual. At times a more personalized approach is needed, such as for a rare or hard to solve case, or when a more patient-centered type of care is desired.

There is a long history of N-of-1 trials or SCEDs, but early on they were almost entirely conducted in the domains of psychology and psychiatry. Starting in the 1980s, the medical field adopted and adapted these types of experiments for their own uses [11], [12]. These N-of-1 experiments involve a single subject or patient in the experiment wherein the patient is observed, treated, and evaluated throughout the study, potentially over weeks or months [10], [13]. The patient acts as their own control, systematically alternating between receiving the medication or intervention, and not, for some extended period of time. This is typically done in a crossover or cross-phase design such as in an ABAB fashion, with the A phases being the baseline of no intervention given and the B phases of the experiment being the intervention round [11], [12]. In some experiments, such as those involving medications or other drugs, a washout period is suggested or required, where the patient is given time for the medication to completely leave their system so that they can return to a more baseline state, but in more behavioral studies this may or may not be feasible, as the subject can not unlearn the techniques given. There are several different variations on these N-of-1 experiments with differing complexities and uses, such as ABBA or its inverse BAAB which may be beneficial to determine if there is a time-dependent effect of the different phases. There are also studies that vary the intervention such as ABACAD and so forth, where the C and D sequences may be the same type of intervention but of a different magnitude, such as an increasing dose of some medication, to find the amount that works best for the patient, or a completely different intervention, such as gratitude journaling or meditation.

These N-of-1 experiments can be tough to design and manage, and can often be time consuming, but the outcomes for the individual patient can be unmatched. The patients are often very involved in the data collection and decision making processes and because of this and the nature of why they

enrolled in the study to begin with, are often more committed to sticking to the schedule or program when real progress is able to be made towards improving their health [14].

N-of-1 studies do have their drawbacks and can not be used in every situation though. With RCTs there is a certain amount of external validity, if the results of an RCT study show that there is a significant and useful effect from medication, there can be a reasonable assurance that it will work on a wider population. This cannot be said of N-of-1 studies where we can not generalize the results of a single person to groups of other people. There are ways in which meta-analyses can be conducted by combining the results from multiple similar N-of-1 studies together to ascertain if there is a more generalizable effect, but these are much more difficult than starting with a traditional RCT in the first place [11], [12]. Additionally, which statistical methods necessary to use on the various N-of-1 designs can be difficult to determine as well. Whether to use simple visual analysis, a parametric, non-parametric, time series analysis, or traditional inferential statistic all depends on the data, and cannot be determined solely on the use of an N-of-1 trial itself [11], [12].

N-of-1s, while complicated, have been thoroughly investigated and refined over time. The gaps in how to design them and how to interpret any results are closing, but there are areas of research closely knit to them that still have future work to be done. There are many considerations to make when thinking about and actually conducting an N-of-1 experiment, but one of the most notable reasons is that a single person can potentially administer them all by, on, and for themselves.

2.2 Self-Experimentation and the Quantified Self

In the last few years there has been a growing interest and body of work in self-experiments, studies where the researcher and subject are one in the same [15], [16]. In these studies the researcher has found that either the data is not there and must be created themselves, or that whatever they are studying is too intensive and unreasonable to ask for another person to subject themselves to [17], [18]. Most often, the researcher-subject is interested in something affecting themselves, and design a study around treating or improving whatever that may be. These self-experiments can be in any area of life that can be measured, where some metric can be used to record a difference in some state of being they have been or are currently in.

Sleep and how to improve it or increase the time spent in it is perhaps the most common type of self-experiment, where the self-researcher changes some behavior or medication to see how it affects them [2], [3]. There are other domains that can be explored as well though, from fitness and nutrition, to mindfulness, meditation, mood and general wellbeing, to personal finance and productivity. They

can even extend to certain diseases and disorders like Irritable Bowel Syndrome (IBS), Multiple Sclerosis (MS), or chronic pain. Generally, the lower the barrier of entry the experiment has, with the greatest personally assumed or analyzed effect on the individual, the more likely that the self-experimenter will stick to and find any results from their self-study. Apps, platforms, and wearables for analyzing each of these specific domains and problems that affect people have been created and are in common use, such as Fitbit, Apple Health, MyFitnessPal, TummyTrials, and many more, but a more generalizable platform, one that is more open-ended and self-created is sometimes needed for these self-researchers [4].

Not every platform for self-experimentation is as easily used in the way that the self-researcher wants and some are more restrictive in their design than others. There have been some platforms that have attempted to decouple and make extensible the independent variables from the dependent ones, with easily modifiable metrics, so that the experimenter can do exactly what they want and study what they want on themselves, in other words a general-purpose system. There are several platforms that follow this design, each iteration being built on the previous one and each has generally made it easier for the self-researcher to experiment on some previously hard to collect or manage metric so that some new effect can be studied. Physiqua [5], QuantifyMe [6], and Self-E [7] are some such platforms that have been created to collect and manage, then combine some number of separate sets of data together from various sources, and ultimately analyze in some way the different interventions that the self-experimenter has implemented on themselves. The latest, Self-E, is especially interesting because of the Bayesian techniques used to analyze the data. Other platforms have been created as well and go even further than these, making it easier for researchers to collect and manage multiple N-of-1 self-experiments so that meta-analyses can be conducted more easily. The StudyU [9] and StudyMe [8] apps have been created to do just that and are just as general-purpose as the others.

These researchers and their apps are often adjacent to or directly involved with the Quantified Self, a community of people who collect and analyze data in a number of ways to try and improve whatever they want about themselves. Even before the smartphone, cheap sensors, and other technologies, there have been people who have collected data about themselves to see if they could improve something that was ailing them, or just for its own sake [19]. Since advances in wearables and the Internet of Things (IoT) has come into the mainstream, and the general population can see their usefulness, the Quantified Self has seen some increase in notoriety as well [20], [21]. There are many news stories written on the topic, and even some journal articles have been published referencing the community and a number of the self-experiments, apps, devices, and analyses that these people use, created, or have conducted in its name [22]–[30]. Most of these self-experimenters, apps, and

platforms, make use of or connect to wearables and other sensors to collect the data that they need to be analyzed.

2.3 Wearables and Health Informatics

Wearables, smartphones, smartwatches, sensors, and the Internet of Things (IoT) which these all fall under, are all used in health- and personal-informatics, and human-computer interaction. The goal is generally of self-tracking, self-monitoring, and self-reflection. The tools that people use to collect this personally relevant information are often made so that their data can be easily shared or downloaded for further analysis. These tools help people gain self-knowledge about their health, behaviors, habits, and thoughts.

Smartphones, smartwatches, and wearable fitness trackers are perhaps the most often used devices in these types of studies and by individuals who are motivated to accomplish some of their goals or are interested in trying to understand their behaviors. Fitbit, Apple Watch, and the Samsung Galaxy wristwatch, are just some of the recognizable consumer-grade brands that can be used for this purpose. Combined with their paired smartphone apps like the Fitbit app, Apple's Health app, or Samsung's MyFitnessPal, there is a lot of utility that can be had for the individual to explore, analyze, and visualize what behaviors or metrics they are trying to change.

Again, we find that fitness related objectives are a popular goal. People are most often interested in improving their sleep, in increasing their time spent walking or exercising, or in changing their diet and how much or what types of things they are eating everyday. There are many benefits to these apps and platforms such as social interaction and accountability, and people tend to stick to their goals if they have partners or interactions with people doing the same [26]. Fitness tracking is just one of the areas that people are interested in improving though, time-tracking and general activity data is another area of interest.

2.4 Time and Activity Tracking

Time and activity tracking has been a frequent topic of discussion in popular-psychology, self-help, business, and even ancient ethics, but as far as modern scientific research is concerned it mainly resides in the domain of industrial-organizational psychology, essentially exploring job-performance, or how to get the most work out of the employee for the least amount of time. This is not the only possibility for time-management related research though, academic performance, differences in

personality measures, and psychological wellbeing are other areas of interest that have been explored as well [31].

Wellbeing and distress are closely linked, and time-management skills and tools can be used to reduce stress or increase general life-satisfaction. One's personality can play a large role in this however, all of the Big Five personality traits are related significantly with one's time-management skills [31]. Whether one is a 'type A' personality or has a high or low self-esteem is also correlated with their time-management skills.

What has been found overall is that there is a moderate effect on time-management and job performance, academic achievement, and general wellbeing. In other words, people who spend some time structuring their time, such as through a daily or weekly planner, or specifically set aside time for studying or other goal-oriented tasks, generally perform better at work and school or have higher life satisfaction overall. All in all, research has found that time-management works, and how one chooses to spend their time has large effects on their life-satisfaction. General wellbeing and life-satisfaction are often overlooked areas of research but there is at least one domain, positive psychology, that seeks to improve them.

2.5 Gratitude Journaling

Gratitude research in general seeks to alleviate feelings of depression and despair, and improve general life-satisfaction, putting the human at the forefront of the study, the individual person and their wellbeing at the center of attention [32]. Gratitude has several meanings even within psychological research, it can mean a thing that a person is, their perspective on life, one of their traits, or a thing that a person does, something they practice.

Gratitude journaling is the act of writing, usually on a regular basis, about all of the things, people, and events one feels grateful for in their life, what they cherish and hold dear or close to their heart. These can range from being grateful that one has food on their table, to having a good friend to spend time with, or simply reflecting on smelling a nice candle that day.

Journaling is just one tool and area of research within positive psychology but there are other ways that a person can practice gratitude. Other common interventions include writing a gratitude letter or doing a short Three Good Things (TGT) exercise. Letters typically consist of writing a personal note thanking an individual for something previously unsaid, though not always actually delivered to this recipient, whereas the TGT exercise is simply putting a limit on the gratitude journaling to three good things that happened that day or week. Ultimately the goal for much of

positive psychology and gratitude research is to try and figure out how to make people happier. Studies on mood and happiness are some of the staples of social-psychological research.

2.6 Mood

Happiness, depression and other mood disorders, and subjective-wellbeing in general have always been areas of interest to psychologists and psychiatrists and this is a trend that will continue with this study. Depression affects millions of people around the world, about 16 million Americans and over 300 million people worldwide [1]. About 1 in 4 people will experience and meet criteria for a mental disorder at some point in their life [33], which has massive effects on not only the individuals afflicted but on society in general. Depression's impact on the world's economy alone is over \$200 billion per year [1], and subclinical levels of stress, anxiety, mental anguish, and distress affect many people around the world and throughout their lives as well [33].

Mental health apps are a common tool used to fight depression and for keeping track of one's mood over time. Around two-thirds of teens and young adults have used an app to track some aspect of their health, including their mental health, and a little less than half of adults have done the same. There are hundreds of thousands of health apps available to consumers with more than 10,000 of those being for mental health alone [33]. Mood is often tracked easily enough on a 5 or 7 point Likert scale, where the user is regularly notified or reminded to fill out the form, usually once per day, such as in one popular example, the Daylio app [34] .

Many of these domains and topics, including mood research, wearables and health informatics, and especially N-of-1 research, are combined in what could be described as a landmark study, with the further allure, and complication, of machine learning.

2.7 Personalized Machine Learning of Depressed Mood Using Wearables

The latest study combining many of these domains, subdomains, and topics, is from a paper by Shah et al. [1] In this memorable study the researchers created personalized machine learning models for each of the participants with depression, to try and explain which behaviors and other indicators predict their individual or personalized mixture of depressed mood. Shah et al. collected fitness and nutrition data from wearables, Ecological Momentary Assessment (EMA) and mood questionnaire data, neurocognitive and Electroencephalography (EEG) data, and then created a machine learning

pipeline that not only predicted each of the participant’s individual moods, but also determined which of those variables or features had the biggest impact on their depressed mood. Much of the study presented here then was a continuation of their work, creating a platform available for others to use their own data and create their own machine learning models to predict their own mood or any feature.

These many related domains, topics, and experiments give an overview of and explain any gaps that may exist from previous interdisciplinary research. While some of the domains have extensive, thorough, and rigorous work done already, there is still much to be done to productionize and make freely available the results of all this work so that it can help improve people’s lives.

Chapter 3

MATERIALS AND METHODS

First we will reiterate the research questions, then describe the participant, the overall study procedures and general data acquisition, and how each of the various data sources were collected, processed, and ultimately able to be aggregated, including any relevant or interesting charts, graphs, and statistics as needed. After the data was put into a more manageable form we will then describe the more advanced statistical analyses done and the machine learning pipeline which enabled the feature importance charts. Here we will restate and keep the research questions in mind:

- (1) For a single-subject self-experimenter, does gratitude journaling, compared to no intervention, improve their mood?
- (2) Does the mere act of self-experimenting affect the subject's mood?
- (3) Can a personalized machine learning model find the features that affect the subject's mood the most?
- (4) Are any of these modes of self-experimenting more or less difficult to perform or are the outcomes more or less difficult to understand than the others?

3.1 Participant

A single self-experimenter was the lone participant of this study, the author and subject being one and the same. The subject was a graduate student at the University of Missouri, male, and 31 years of age at the end of this longitudinal project. The subject gave consent to participate in the study, agreeing to the interventions and information collection measures, and acknowledged that they were allowed to stop at any time and for any reason. IRB approval was not necessary for this study because of the benign autoethnographic or idiographic nature of the study. None of the data was confidential or needed to be de-identified or anonymized, an impossible task for an N-of-1 study,

but despite this, some of the data collected was of a more private, personal, or sensitive nature and may have been obscured or otherwise protected in some way so that it would be more suitable for a public release, including some feature names being modified.

3.2 Study and Experimental Procedures

The participant took part in 2 different longitudinal studies, the longest running experiment beginning on 01/01/2020 and carrying on through 03/10/2023, which will serve as the end of data collection, though it is ongoing (Figure 3.1). In this first experiment the self-experimenter tracked their mood through a daily questionnaire, several fitness metrics through their Fitbit wearable such as steps walked per day, and also their minute-by-minute time and activity data through the aTimeLogger app [35] on their Android smartphone. This first study overlapped with and was ongoing as the second experiment began, which was an interventional BABABA crossover phase design study, which started on 11/13/2022 and ended on 02/04/2023, for three 28 day phases totalling 84 days (Figure 3.2). This was the study wherein the subject spent 10-15 minutes per day gratitude journaling, and in addition started tracking several new metrics as well. The subject spent time gratitude journaling in an app, Presently [36] on the B days of the study and received no intervention on the A days. The journaling could have just as easily been done on a piece of paper or in a physical journal, but the possibility for further analysis, such as by creating sentiment scores from the exported notes, is there for future studies.

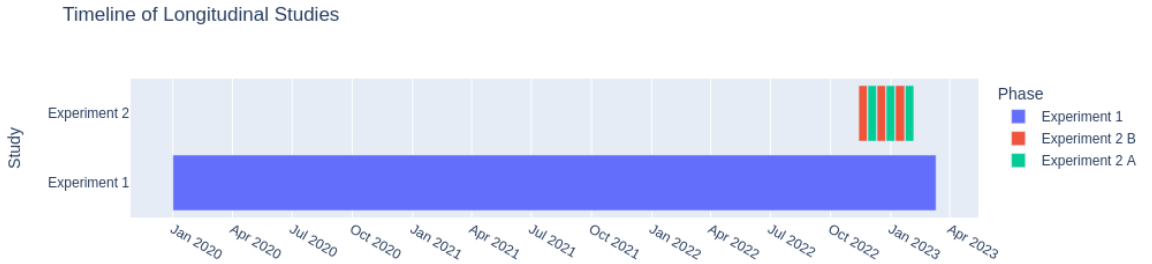


Figure 3.1: Study Timeline

Over 3 years of the authors own self-tracked data was analyzed, with some of the features examined having been collected from the beginning of the entire longitudinal study such as mood, pain, weight, Fitbit data, and time and activity data, among others. For the crossover phase design experiment that came later, a new set of features were collected such as blood pressure and temperature, in

addition to the features from the first experiment during their overlap.



Figure 3.2: Experiment 2 Timeline

3.3 Data

As with most end-to-end data science and analysis projects, much of the time for this project was spent on data carpentry, data engineering, and on some basic data operations, configuring and working with several different devices, smartphone apps, and in this case on personal Fitbit and Google accounts and applications.

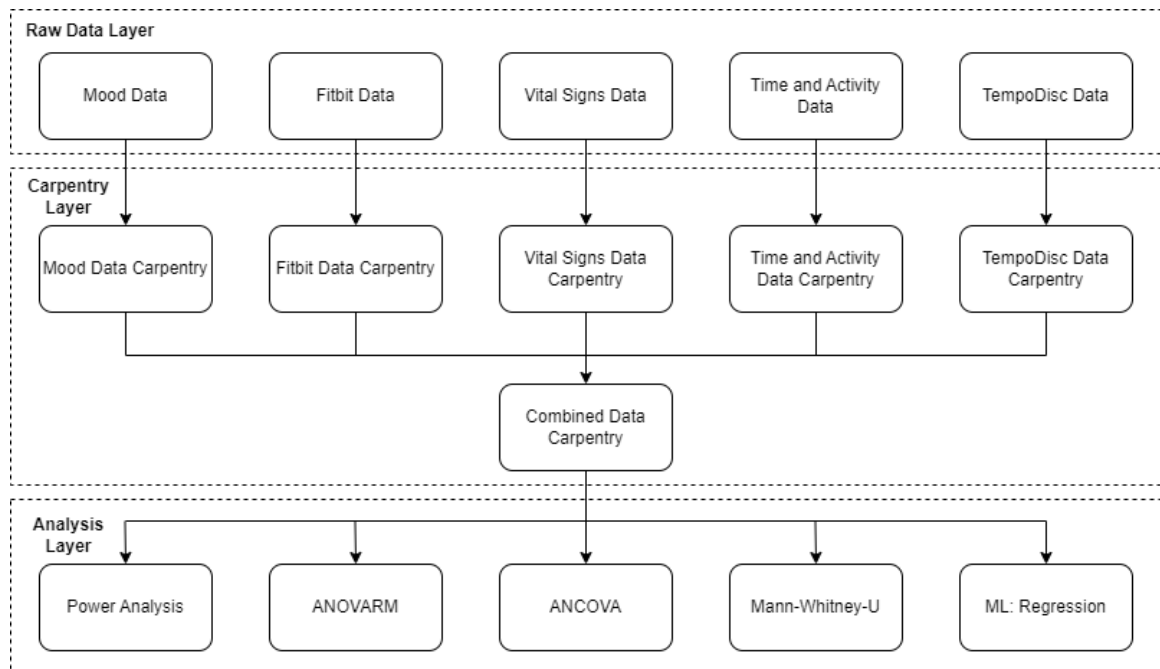


Figure 3.3: Overall Data Flowchart

Here (Figure 3.3), we can see the general overview and path taken for each of the individual data sources, how they were preprocessed with basic massaging or munging, and how they were eventually merged together so that further analyses could take place. Each of the top layer boxes represent the raw data sources, generally in the form of a Google Sheet or Comma Separated Value (CSV) file, or in some way collected through an API. In the carpentry layer that raw data was then transformed so that it could all be combined together and prepared for further analyses or used to create charts and graphs of several different varieties, depending on the need. In the final layer we have the list of analyses that will be explained in more detail later.

Much of the raw data was collected through Google Forms and Google Sheets. A Form is a simple customizable questionnaire and a Sheet is an online spreadsheet editor, similar to Microsoft Excel. First we created the Google Forms and then linked those to their own Google Sheets within our Google account which allowed for the answers from the Form to automatically update a Sheet. Once the data was in the Sheet then we could easily download it. A shortcut was created to these various Forms on our Android phone's home screen so that we could, with the tap of an icon, access our customized Form so that this entire data ingestion process could begin. The appeal for this general workflow, using the mood data as an example (Figure 3.4), from tapping the icon on the phone screen, to filling out a Form, to that data automatically being sent to a Sheet, to then downloading that data through the Google API with a custom-made program, is that it is an easily customizable and extensible way for the self-experimenter to create new datasets based on their individualized wants, needs, goals, or other objectives.

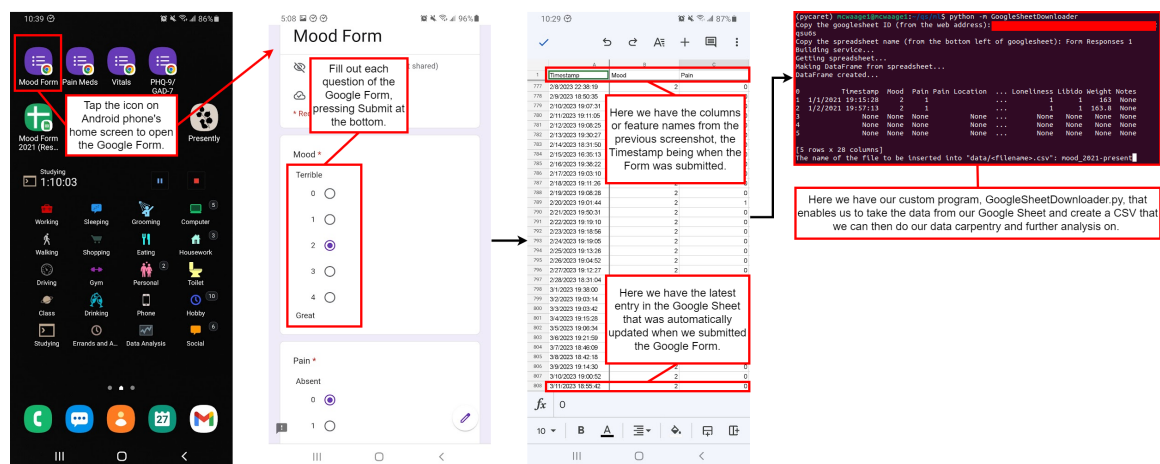


Figure 3.4: Mood Data Flow

To even begin accessing our data through the Google API, first we had to create a Google cloud app, which we did by following a couple of different guides (<https://developers.google.com/sheets/api/quickstart/python>, https://github.com/markwk/qs_ledger/blob/master/google_sheets/google_sheet_get_data_example.ipynb) [37], [38]. At the end of this process we were given a Google API key and token which, when placed in the correct files and folders, allowed us to use our custom program `GoogleSheetDownloader.py`, and easily download our data from the Forms and Sheets in our Google account.

One thing to note is that when this project began the author had no experience programming or any data science knowledge or education, so a no code approach was the only avenue towards any basic data analysis. The author used the builtin Google Charts within Google Sheets to visually analyze the trends in their mood, weight, and other variables. Any future self-experimenters recreating this process or using these tools for their own purposes could stop at that stage, not needing to go any further or go through with obtaining a Google API key and downloading the resultant data, meaning they would not need to have any coding or programming knowledge and still be able to gain insights and knowledge about themselves.

Due to the nature of real-world longitudinal studies, those occurring outside of the clinical setting, there are many opportunities for confounds, outliers, and missing or incorrectly input data to arise, even for the self-experimenting data-enthusiast. Any one of the various data sources may have had missing or incorrect data each day. From time to time, daily in most cases, the data was reviewed for any errors and modified if any were caught. Despite this, there may still be missing or erroneously coded or attributed data. As this is a real-world study, activities of daily living had the potential to interrupt segments of data collection in various ways, such as devices malfunctioning or breaking, data uploads failing, work, sleep, or other time commitments and obligations breaking the flow of data collection, or simply getting distracted while filling out a Form. Each of the data sources and their confounds, potential afflictions that they experienced, how those problems were dealt with, and how they were ultimately transformed and joined together for further analysis will now be explored in more detail.

3.3.1 *Mood Data*

The heart of this study and the entire collection of datasets was the Mood Data. The subject's mood and its improvement being the main motivation behind the entire study, it was often the main target for any examination, statistical analyses, or machine learning models. First we will cover how

the Mood Data was created and any rationalizations for why the system and questions were set up in the way they were, any problems that affected this data source, and then we will continue on to how it was transformed for further use, including any exploratory visualizations, tables, and other statistical analyses as needed.

Referring back to Figure 3.4, we find that the flow for the user and their Mood Data being creating each day is; (1) tapping on a Google Form icon on their phone’s home screen, (2) filling out the Mood Form, which (3) automatically updates a Google Sheet, which (4) is then able to be downloaded through the Google API with our custom `GoogleSheetDownloader.py`. At the end of this process we have created a CSV file that can then be sent on to the Data Carpentry layer (Figure 3.3).

Encased within the Mood Data item we actually have three separate mood and anxiety questionnaires used altogether in this study; (1) a modified version of the Hamilton Depression Rating Scale (HAM-D or HDRS), which turned into our custom Mood Form, (2) the Patient Health Questionnaire (PHQ-9), and (3) its companion questionnaire, the General Anxiety Disorder 7 (GAD-7) [39]–[41]. The modified HAM-D was created for the longest running Experiment 1 (Figure 3.1), and used in both experiments, while both the PHQ-9 and GAD-7 were created for Experiment 2 (Figure 3.2). The self-experimenter’s flow for the PHQ-9/GAD-7 Form that began with Experiment 2 was identical to the first, only the icon being tapped and the Form being filled out representing a different dataset.

The HAM-D, what would become our main Mood Form, was modified to fit our needs and went through one revision since its creation. The Mood Form is separated into 2 different versions, (1) the period of 01/01/2020 to 12/31/2020, and (2), 01/01/2021 until the present. When this revision took place many previously tracked variables were modified and the scale or range of options on others was changed. When the Form was first created in 2020 there were 39 variables being tracked, including the Timestamp from when the Form was submitted. Many of these items were either too difficult to keep track of or just uninteresting and irrelevant and were discarded, while for others items the values on their scales were too difficult to differentiate between and their ranges were rescaled. With the revised Mood Form starting in 2021, we reduced that number of variables being tracked to 28. Here we have a list of all of the items present in the 2020 and revised versions of the Mood Form (Table 3.1). Some of the items in the revised Form were dropped and a couple were added as well with any blank areas showing where that item is not present in that Mood Form version.

For most of the Mood Form items, the subject tried to quantify their current feelings based on each of the item’s names such as, Mood, Pain, Tired, etc. In the original 2020 Mood Form the main Mood item was on a 0-10 scale, from Terrible to Great, with the default being a 7. For the

	Mood Form Item	2020 Measures	Revised Measures	Description
1	Timestamp	Datetime	Datetime	The time the Mood Form was submitted, created automatically
2	Mood	0-10	0-4	How the subject felt their mood was that day
3	Pain	0-10	0-10	How much pain the subject felt at the time of the Form submission
4	Pain Location		String Field	Where their pain was located, if any
5	Self-Image	0-10		How the subject thought about themselves as a person
6	Energy	0-10		How energetic or excited the subject thought they were
7	Concentration	0-10		How well the subject was able to focus on tasks
8	Clearheaded	0-10		How clear or foggy the subject felt
9	Productiveness	0-10		How much the subject felt they accomplished that day
10	Tired	0-4	0-2	How sleepy or exhausted the subject felt
11	Sick	0-4	0-2	How sick the subject felt
12	Anxiousness	0-4	0-2	How worried the subject felt they were, or if they were ruminating excessively
13	Depressed	0-4	0-2	How down or low the subject felt
14	Suicidal	0-4	0-2	Whether the subject thought about suicide
15	Self-Harm	0-4	0-2	Whether the subject thought about or performed any self-harm actions
16	Thoughts of Death	0-4	0-2	Whether the subject had any thoughts about death such as images of them dying
17	Thoughts of Worthlessness	0-4	0-2	Whether the subject had any thoughts of having no value
18	Thoughts of Hopelessness	0-4	0-2	Whether the subject had any thoughts of no hope for the future
19	Repetitive Thoughts	0-4	0-2	How much the subject had overwhelming or repeated thoughts throughout the day
20	Brain Sludge	0-4	0-2	Whether the subject felt their brain or head was heavy or weighing them down
21	Sluggish	0-4	0-2	How much the subject felt they were physically slowed down, like trudging through water
22	Restless	0-4	0-2	Whether the subject felt they got little to no sleep or if they couldn't sit still
23	Existential Crisis	0-4	0-2	Whether the subject felt they were having a moment of crisis or loss of purpose
24	Weird		0-2	How different or strange the subject felt that was not encompassed by the other measures
25	Loss of Interest	0-4	0-2	Whether the subject felt they didn't enjoy things they previously had
26	Self Isolation	0-4	0-2	Whether the subject was purposefully spending their time alone
27	Paranoia	0-4	0-2	Whether the subject felt persecuted or that they may have been imagining things that were not true
28	Elevated	0-4	0-2	Whether the subject felt they were above their normal mood, talking or moving faster than normal
29	Irritable	0-4	0-2	Whether the subject felt they were getting annoyed more easily by other people or things
30	Anger	0-4		Whether the subject felt they were getting angered more easily by other people or things
31	Loneliness	0-4	0-2	How much the subject felt they were lonely or without friends or other relationships
32	Libido	0-4	0-2	How much the subject felt their libido was
33	Notes	String Field	String Field	A freeform string field that the subject could write in to explain any abnormal events affecting them
34	Shower	0-1		Whether the subject showered that day
35	Brush teeth	0-1		Whether the subject brushed their teeth that day
36	Floss	0-1		Whether the subject flossed that day
37	Shaved/Trimmed	0-1		Whether the subject shaved or trimmed their hair that day
38	Haircut	0-1		Whether the subject had a haircut that day
39	Trimmed Nails	0-1		Whether the subject trimmed their nails that day
40	Vitamins	0-1		Whether the subject had their vitamins that day
41	Weight	Float Field	Float Field	The subjects weight in pounds that day

Table 3.1: Mood Form Table Description

This table lists each of the items in each of the Mood Form versions, their scales or range of values used, and a more detailed description of what is meant by each item.

revision this was converted to a 0-4 scale with a neutral mood of 2. This was done because it was difficult to differentiate the meaningfulness between each of the numbers on such a broad scale. An example of this in practice would be if the subject felt that they were doing better than an average day, they would have given the Mood item a score of 8 out of 10 in the original Mood form or a 3 out of 4 in the updated version. Pain was the same under both versions with a 0-10 scale, from ‘Absent’ to ‘Incapacitating’ as this is the standard used in most hospital and research settings. Items 5-9 were discarded with the revision because they were too difficult to quantify. Items 10-32 were converted from a 0-4 scale (Absent, Mild, Moderate, Severe, Incapacitating) to a 0-2 scale (Absent, Mild/Moderate, Severe/Incapacitating), because of the difficulty differentiating between each of the values. Items 34-40 were dropped in the revised version simply because they were uninteresting. Irritable and Angry were combined into a single variable Irritable/Angry with the revision. Pain location was added in the revision, which is a string field where the subject could describe where their pain was located, if present. Weird was another variable that was added with the revision in 2021 as well, which is a catchall for any feelings that could not easily be described but were still worth quantifying. Notes were a freeform string field where the subject could write anything that may have affected the data collection that day. Finally, the Weight item was a float field that the subject could input with their weight in pounds for that day. Weight was included with the Mood Data instead of the Fitbit Data or somewhere else because they were both collected daily and it was simply easier to add it to the Form here.

As mentioned previously, there is much room for confounds, outliers, and missing or incorrect data in longitudinal studies of this nature. The Mood Form was meant to be completed once per day for the entire 3 year period and with that length of time comes errors of many different varieties. Generally, the Form was completed at around 7 p.m. each day, as that was when an alarm was set to occur, but there were days when it was filled out earlier, later, or even the next day if a major error occurred, forcing the subject to complete it retroactively based on what the subject thought they had felt the previous day.

One additional piece of complexity with time-series data in general, that affects this dataset as well, is time-leakage. As this dataset was indexed on the date, the data points in this dataset represented their respective variables for their entire day, from 00:00:00 to 23:59:59, not since last entry. If the questionnaire was completed at 2 p.m. on a day, that meant that even if the subject felt some change and had a different Mood or Pain level later in the day, either better or worse, the value that was actually recorded was the only one that counted towards any future analyses. This means that some of the other features continued to accrue the rest of the day like Step Count and all

of the time-based features, making it so that some portion of the features were in the future hours of our target variable's collection time.

Confounds arise in psychological research frequently as what is being studied, such as subjective Mood, may have latent unaccounted for variables that are affecting them. Real-world non-clinical research is also affected by confounds to an even greater degree as not everything the subject experiences can be controlled for. For this study there were several known confounds that were not or could not be measured and many that were unknown. Many things in a subject's life throughout a longitudinal study can change, especially the longer it persists. A couple possible confounds perhaps most closely related to the Mood dataset is that the subject started taking creatine supplements and several different vitamins throughout the study. These have the potential for affecting some of the other variables collected but were not recorded. Another possible confound unaccounted for could be changes in diet or water intake but others are unknown. Any of the time related confounds such as different life events, social life changes, or work and sleep schedules will be covered in the Time and Activity Data section.

Confounds were not the only potential problems affecting this data though, missing data being another avenue for errors. Missing data was actually not as much of a problem as expected and for the 2020 Mood Data the only variable that had any significant data missing was the Weight variable, with nearly 25% of the days not having a weight input. This would happen from time to time because the subject's routine may have changed or they simply forgot to weigh themselves on some days. With the revised Mood Data that number was reduced to 13.13% though after the carpentry was performed a Weight was created for any missing days based on the previous day's value. A more in depth look at the missing values can be seen with Table 3.2.

With the revision in 2021 much of the data in 2020 had to be converted to the new scale. The values in item 2, Mood, were assigned new values with 5 and below on the 2020 scale being converted to a 0, any 6 values being converted to a 1, any 7 values being converted to 2, any 8 values being converted to 3, and any 9 or 10 values being converted to a 4. Items 10-32 were rescaled as well with any 0 values staying the same at 0 on the new scale, any 1 or 2 values being converted to 1, and any 3 or 4 values being converted to 2 on the new scale. In other words they were rescaled from the 0-4 scale with Absent, Mild, Moderate, Severe, Incapacitating, to a 0-2 scale of Absent, Mild or Moderate, and Severe or Incapacitating.

With the many different items from the different versions of the Mood Form all on the same scale, further analysis could take place. With Table 3.3 and Table 3.4, we are given an overview of the data after resampling was performed, which created a Date value for any of the missing days

		2020	Revised	Combined	Combined
	Mood Form Item	% Missing Values	% Missing Values	% Missing (Before Carpentry)	% Missing (After Carpentry)
1	Timestamp	1.87	0	0.6	0
2	Mood	1.87	0	0.6	0
3	Pain	1.87	0	0.6	0
4	Pain Location		95	96.59	96.59
5	Self-Image	1.87			
6	Energy	1.87			
7	Concentration	1.87			
8	Clearheaded	1.87			
9	Productiveness	1.87			
10	Tired	1.87	0	0.6	0
11	Sick	1.87	0	0.6	0
12	Anxiousness	1.87	0	0.6	0
13	Depressed	1.87	0	0.6	0
14	Suicidal	1.87	0	0.6	0
15	Self-Harm	1.87	0	0.6	0
16	Thoughts of Death	1.87	0	0.6	0
17	Thoughts of Worthlessness	1.87	0	0.6	0
18	Thoughts of Hopelessness	1.87	0	0.6	0
19	Repetitive Thoughts	1.87	0	0.6	0
20	Brain Sludge	1.87	0	0.6	0
21	Sluggish	1.87	0	0.6	0
22	Restless	1.87	0	0.6	0
23	Existential Crisis	1.87	0	0.6	0
24	Weird		0	31.86	0
25	Loss of Interest	1.87	0	0.6	0
26	Self Isolation	1.87	0	0.6	0
27	Paranoia	1.87	0	0.6	0
28	Elevated	1.87	0	0.6	0
29	Irritable	1.87	0	0.6	0
30	Anger	1.87			
31	Loneliness	1.87	0	0.6	0
32	Libido	1.87	0	0.6	0
33	Notes	98.39	99.25	98.98	98.98
34	Shower	2.96			
35	Brush teeth	2.96			
36	Floss	2.96			
37	Shaved/Trimmed	2.96			
38	Haircut	2.96			
39	Trimmed Nails	2.96			
40	Vitamins	2.96			
41	Weight	24.46	13.13	16.87	0

Table 3.2: Mood Form Missing Data

throughout the 3 year data collection, but before any carpentry like filling in those now missing values was actually performed. Table 3.5, and Table 3.6 describe the full Mood Data set after all carpentry was performed.

Date	Mood	Pain	Pain Location	Tired	Sick	Anxiousness	Depressed	Suicidal	Self-Harm	Thoughts of Death	Repetitive Thoughts	Thoughts of Worthlessness	Thoughts of Hopelessness	Brain Sludge	Sluggish	Restless	Existential Crisis	Weird	Loss of Interest	Self Isolation	Paranoia	Elevated	Irritable/Angry	Loneliness	Libido	Notes	Weight
2019-12-24	2	1		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
2019-12-25	1	0		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2019-12-26	1	1		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2019-12-27	1	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2019-12-28	1	0		1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2023-03-07	2	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2023-03-08	2	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	168.8
2023-03-09	2	0		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2023-03-10	2	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	166.6
2023-03-11	2	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	

Table 3.3: Mood Form Table

	Mood	Pain	Tired	Sick	Anxiousness	Depressed	Suicidal	Self-Harm	Thoughts of Death	Repetitive Thoughts	Thoughts of Worthlessness	Thoughts of Hopelessness	Brain Sludge	Sluggish	Restless	Existential Crisis	Weird	Loss of Interest	Self Isolation	Paranoia	Elevated	Irritable/Angry	Loneliness	Libido	Weight	
count	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	1167	800	1167	1167	1167	1167	1167	1167	1167	1167	976
mean	2.118	0.201	0.278	0.025	0.009	0.009	0.001	0.001	0.016	0.154	0.015	0.007	0.018	0.011	0.081	0.003	0.003	0.016	0.002	0.001	0.008	0.008	0.367	1.007	165.100	
std	0.451	0.404	0.446	0.196	0.109	0.105	0.029	0.029	0.127	0.364	0.127	0.092	0.133	0.105	0.272	0.058	0.050	0.127	0.041	0.029	0.088	0.097	0.494	0.083	4.822	
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	155.8
25%	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	161.8
50%	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	163.8
75%	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	167.8
max	3	3	1	1	2	2	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	2	2	2	178.4

Table 3.4: Mood Form Statistics

Date	Mood	Pain	Tired	Sick	Anxiousness	Depressed	Suicidal	Self-harm	Thoughts of Death	Repetitive Thoughts	Thoughts of Worthlessness	Thoughts of Hopelessness	Brain Sludge	Sluggish	Restless	Existential Crisis	Weird	Loss of Interest	Self Isolation	Paranoia	Elevated	Irritable/Angry	Loneliness	Libido	Weight	
2020-01-01	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	159.4	
2020-01-02	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	160.4
2020-01-03	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	158
2020-01-04	1	2	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	159.4
2020-01-05	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	158.2
2023-03-06	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	169.6
2023-03-07	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	169.6
2023-03-08	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	168.8
2023-03-09	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	168.8
2023-03-10	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	166.6

Table 3.5: Mood Form Table Full

	Mood	Pain	Tired	Sick	Anxiousness	Depressed	Suicidal	Self-Harm	Thoughts of Death	Repetitive Thoughts	Thoughts of Worthlessness	Thoughts of Hopelessness	Brain Sludge	Sluggish	Restless	Existential Crisis	Weird	Loss of Interest	Self Isolation	Paranoia	Elevated	Irritable/Angry	Loneliness	Libido	Weight	Weird	
count	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165
mean	2.123	0.197	0.275	0.025	0.009	0.009	0.001	0.001	0.015	0.155	0.015	0.007	0.018	0.011	0.081	0.003	0.003	0.016	0.002	0.001	0.008	0.008	0.008	0.367	1.007	164.947	0.002
std	0.446	0.462	0.447	0.196	0.109	0.105	0.029	0.029	0.123	0.364	0.127	0.092	0.133	0.105	0.272	0.059	0.059	0.127	0.041	0.029	0.088	0.097	0.494	0.083	4.778	0.041	
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	155.8	
25%	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	161.8	
50%	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	163.8	
75%	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	167.2	
max	3	3	1	1	2	2	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	2	2	2	178.4	1

Table 3.6: Mood Form Statistics Full

Mood	Pain	Tired	Sick	Annoyance	Depressed	Suicidal	Self-harm	Thoughts of Death	Repetitive Thoughts	Thoughts of Worthlessness	Thoughts of Hopelessness	Brain Sludge	Sluggish	Restless	Existential Crisis	Weird	Loss of Interest	Self Isolation	Paranoia	Elevated	Irritability	Loneliness	Libido
0	0.00%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
1	3.02%	1.000%	1.000%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2	7.69%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%	2.00%
3	16.48%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%	3.00%
4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
5	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
6	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
8	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
9	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 3.7: Mood Form Percentages

Now we will show a few of the more interesting trends with the full Mood dataset. Mood of course being the main motivator, we will examine it first. Figure 3.5 and Figure 3.7 show us the general trends for the subject’s mood for the entire 3 year period of the study. We can see how over time there were periods of high, low, or neutral moods. Despite the highest score possible of 4 for Mood, Great, not being recorded, we can see with Figure 3.6 that there were more Good than Bad and Terrible days. Overall 79.66% of the Mood scores were 2, Neutral, which on the 2020 scale would have been a 7 out of 10. A score of Good, accounted for 16.48% of the days, and the scores of Bad and Terrible were 3.52% and 0.34% respectively. Table 3.7 gives us a more in depth look at the distributions for each of the Mood Form items.

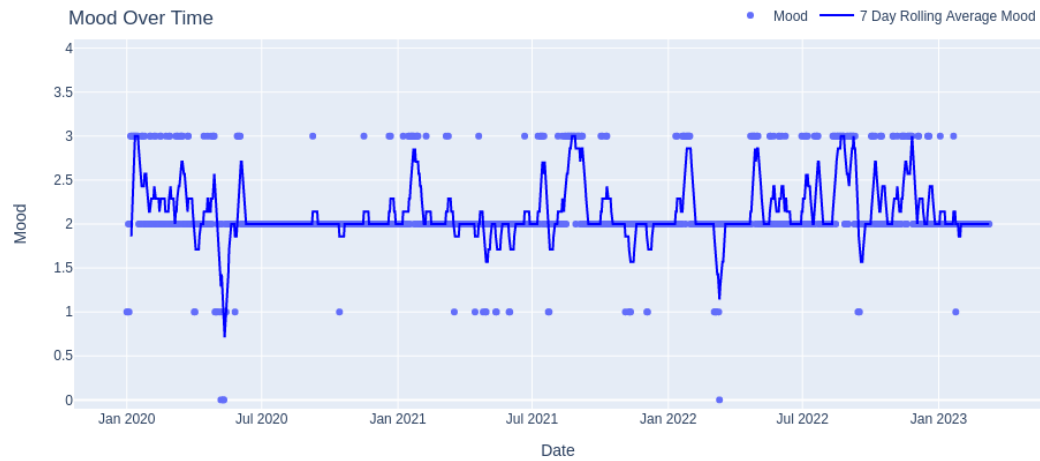


Figure 3.5: Mood Over Time

We can see the trend over time for the subject's Mood, with periods of Good, Neutral, and Bad or Terrible moods.

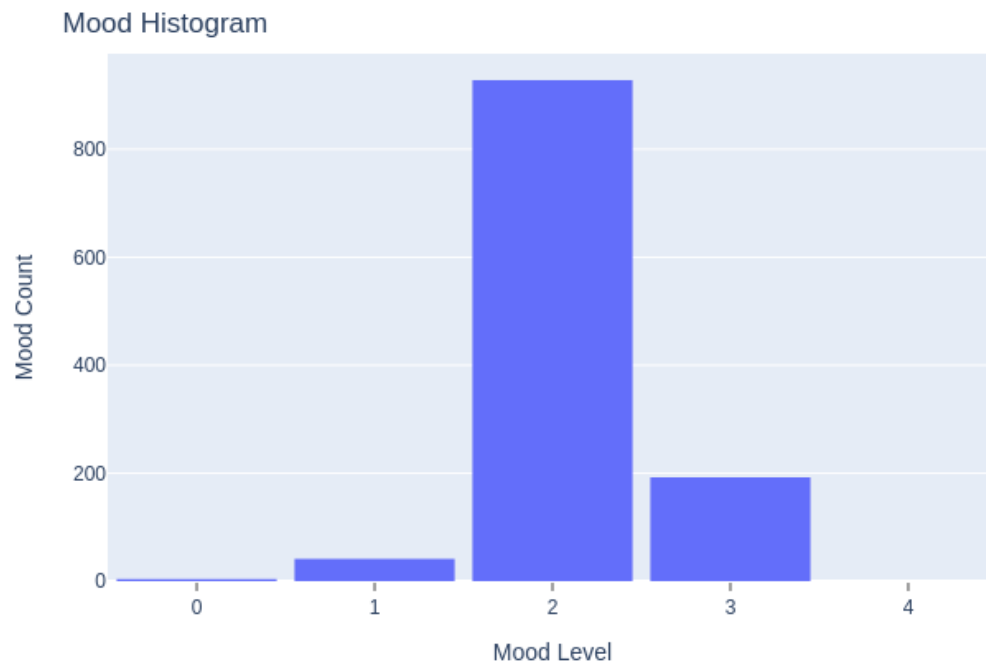


Figure 3.6: Mood Histogram

We can see the subject had more Good than Bad or Terrible values for their Mood, but for the most part remained neutral.

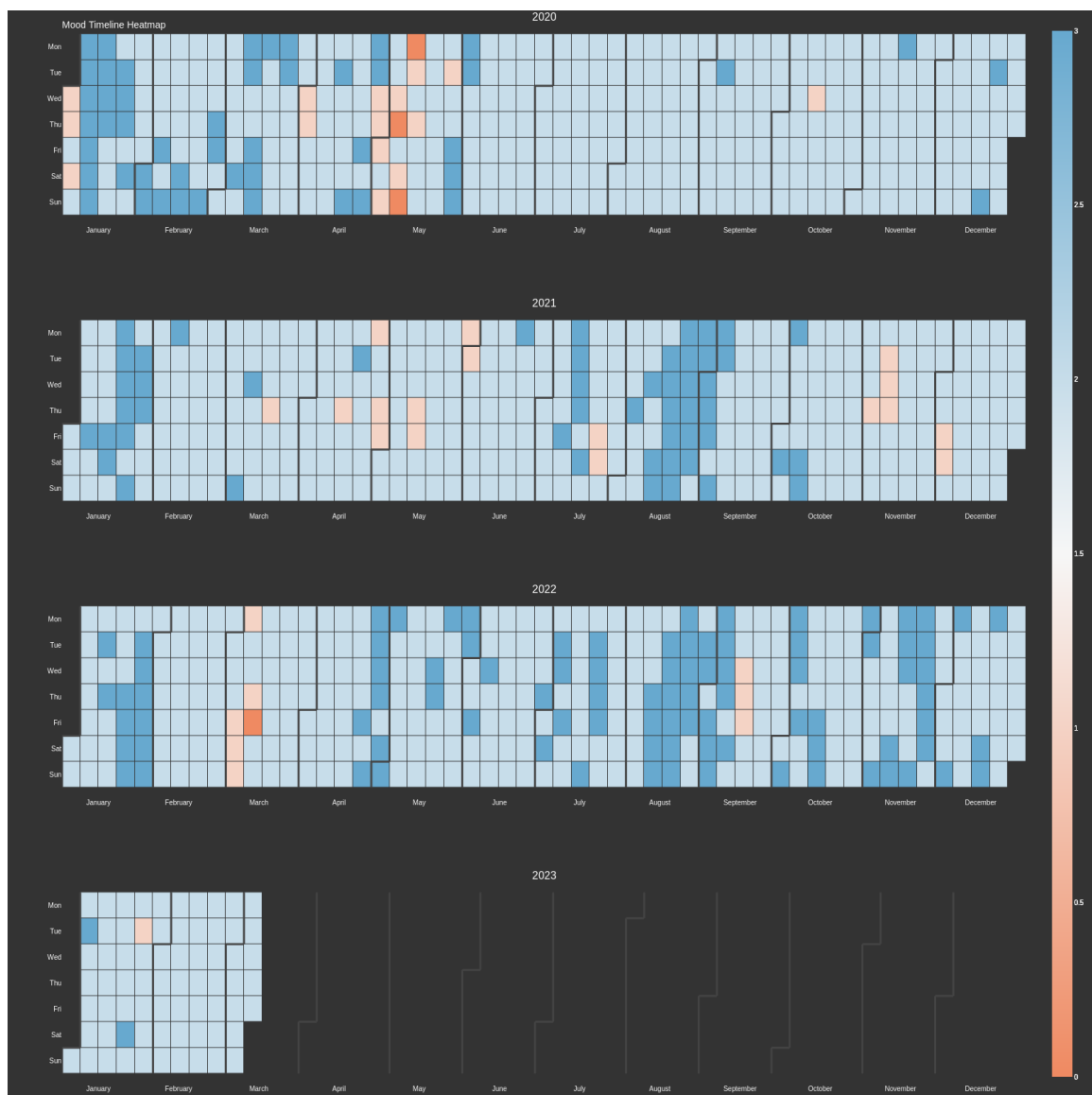


Figure 3.7: Mood Timeline Heatmap

For each of the year-long periods of the Mood item data we can see the value for the subject's mood for each day, the day of the week, and each month as well, with bands of Good or Bad moods, though mostly remaining neutral throughout the 3 year period.

Another interesting trend was with the Pain item. Here (Figure 3.8, Figure 3.9) we can see how the subject’s pain began to subside in May of 2021, and as we will see in the Time and Activity Data section, this trend coincides with when the subject began to go to the gym on a more regular basis. The majority of Pain scores were 0, or Absent with 82.92%, with the next most having a score of 1 with 14.51%, 2 with 2.49%, and 3 with 0.09% of the total scores. We can see how the subject did not experience any Pain above a 3, but for almost the entire first half of data collection they had much more regular or chronic low-grade Pain. Even visually inspecting the data like this can give the self-experimenter insights about themselves and is one form of analysis often used in N-of-1 studies similar to this [11], [12].

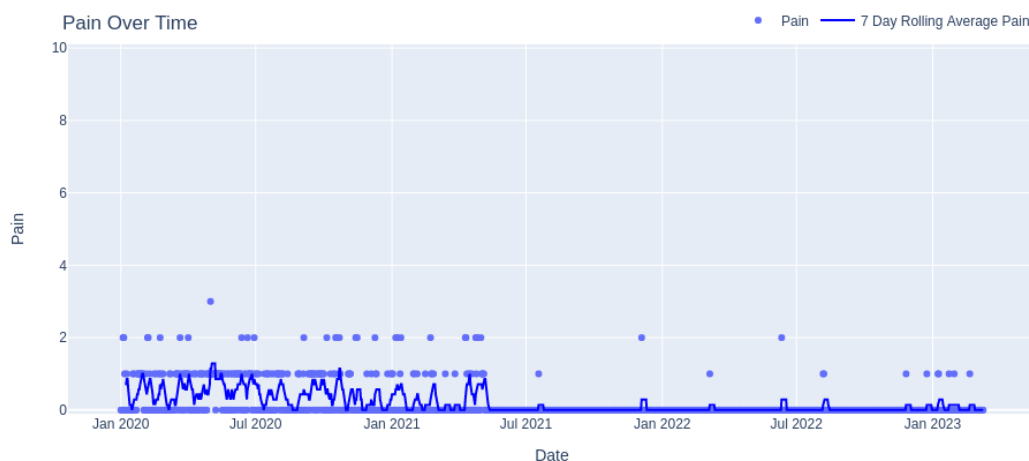


Figure 3.8: Pain Over Time

We can see there was a drop-off in Pain starting in May of 2021 when the subject started spending more time in the gym.

One more variable worth inspecting here is the Weight item. With Figure 3.10, Figure 3.12, and Figure 3.11 we can see how there was a lot of variation in the subject’s Weight. This also coincides with the increase in time spent at the gym similar to the Pain item.

Once each of the datasets are explored in more detail then we will examine each of their relationships with the subject’s Mood and the other items in the Mood Form in more detail. The Mood Form was not the only mood questionnaire though, we also created a new Google Form for the start of Experiment 2, a combined Patient Health Questionnaire (PHQ-9) and General Anxiety Disorder 7 (GAD-7) [40], [41].

The PHQ-9 and GAD-7 are typical questionnaires used in hospitals and other psychological

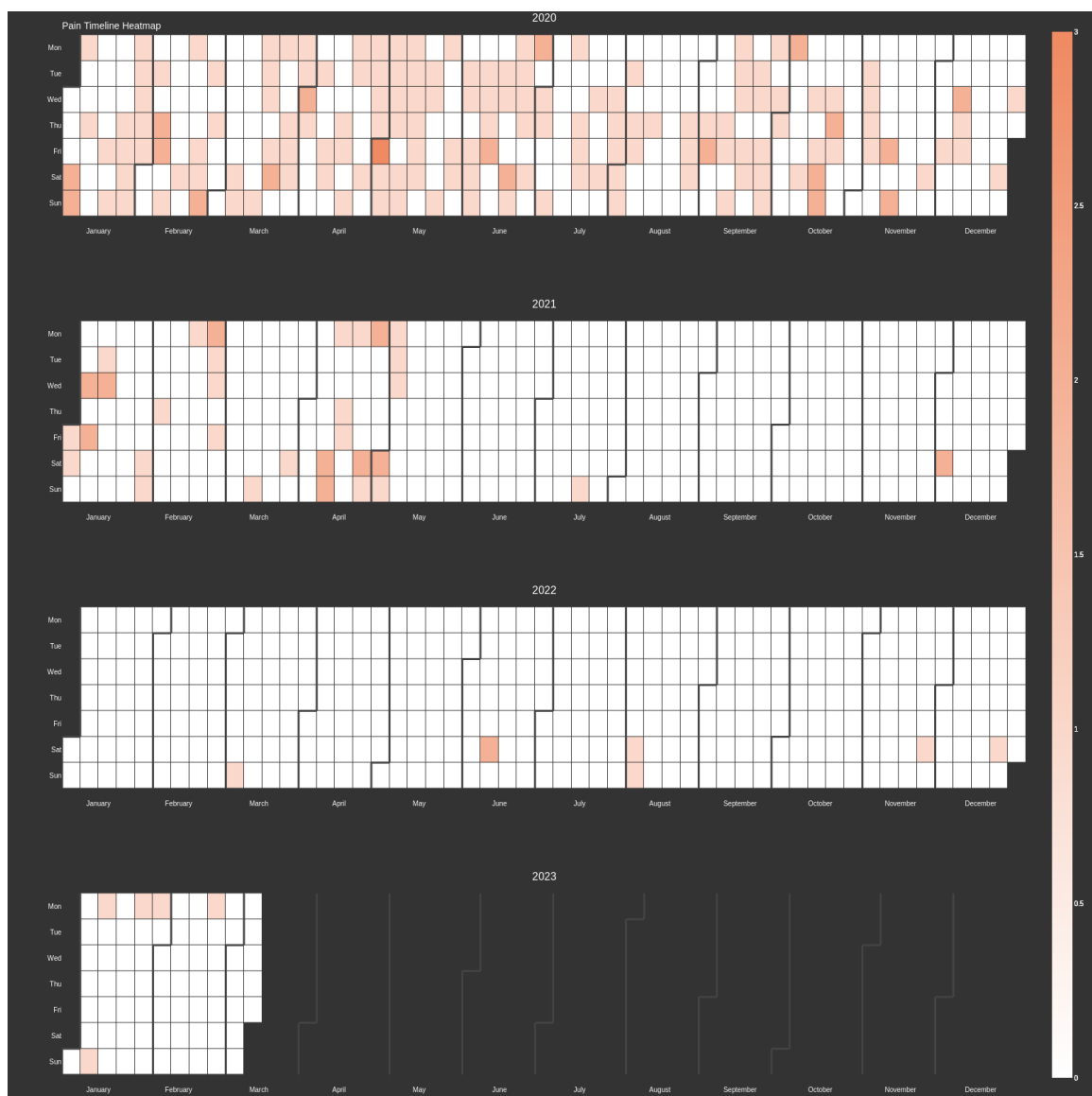


Figure 3.9: Pain Timeline Heatmap

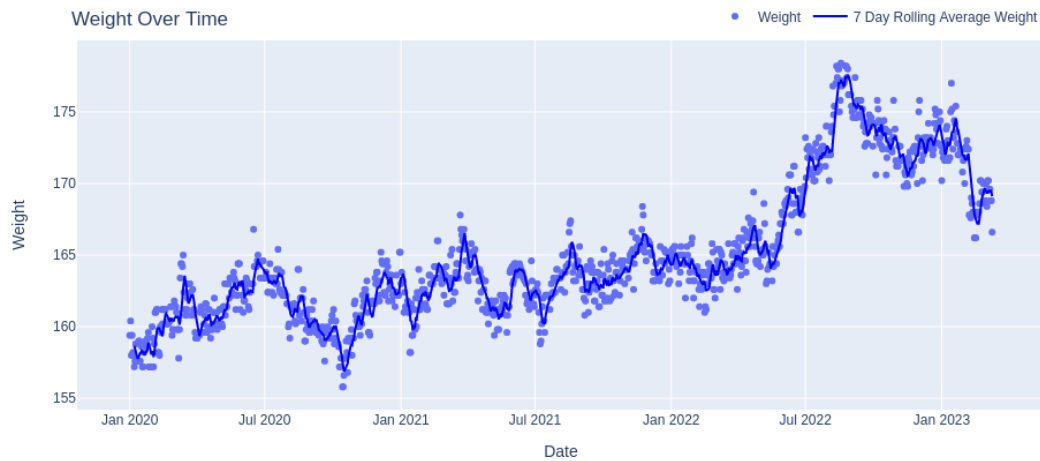


Figure 3.10: Weight Over Time

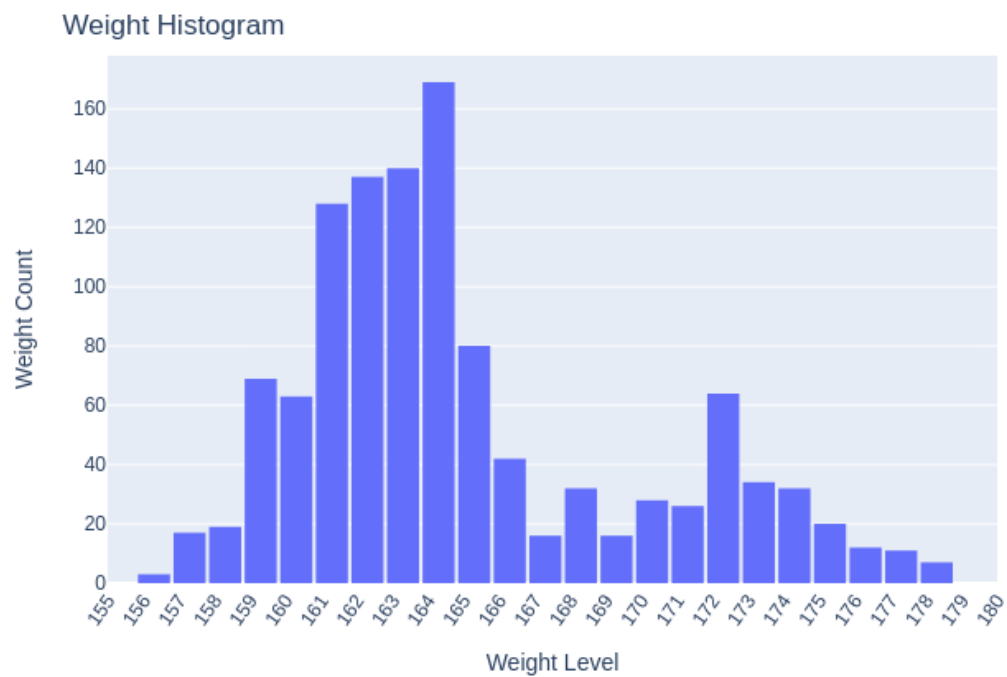


Figure 3.11: Weight Histogram

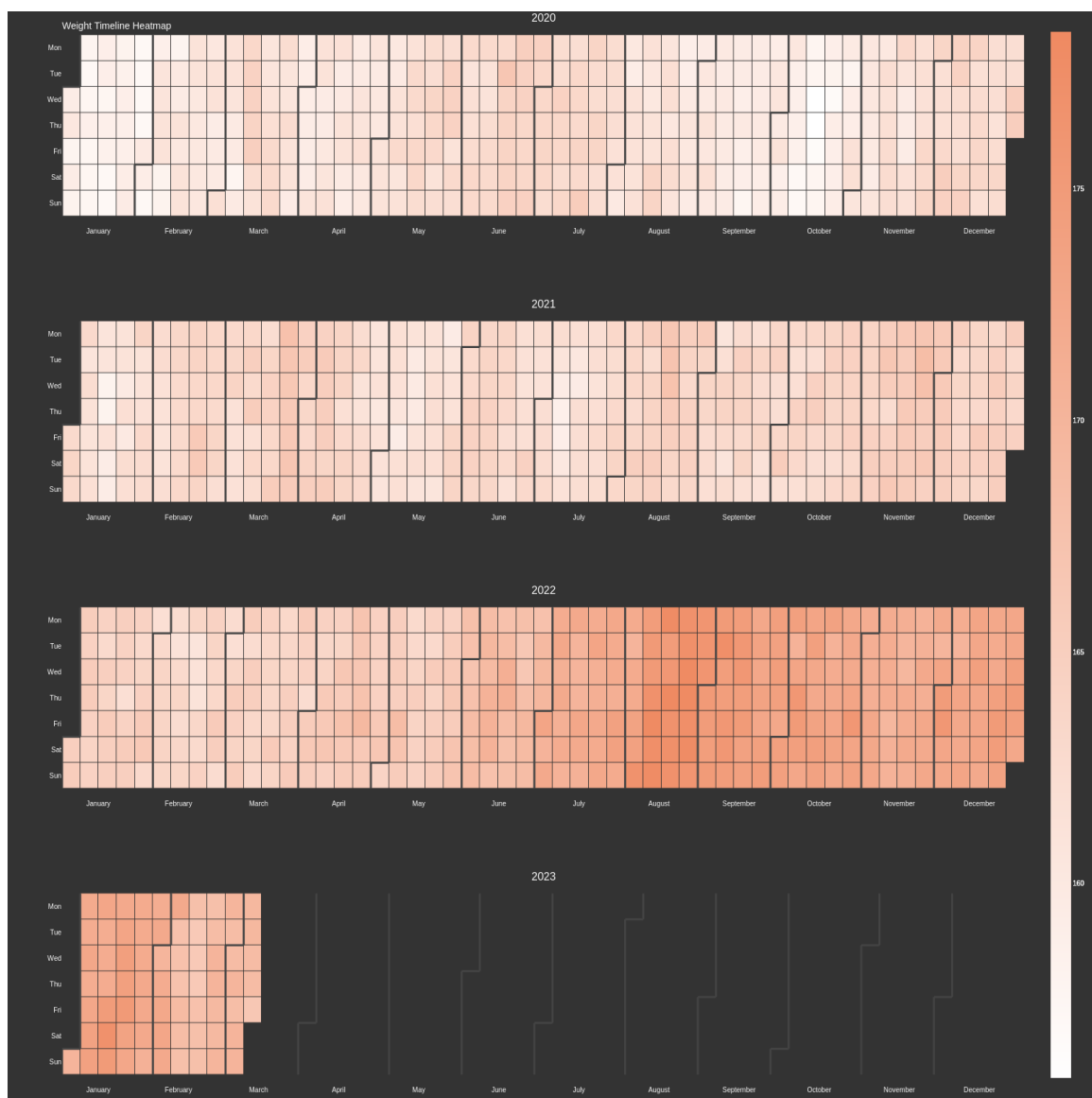


Figure 3.12: Weight Timeline Heatmap

research settings which aim to quickly discern the level of a patient’s depression and anxiety, if any. They are often used by medical professionals to decide whether the patient should be admitted for a psychiatric event or if they are in a crisis of some sort. These questionnaires were used in this study for a more peer-reviewed, valid, and reliable measure compared to our custom Mood Form questions. There are more in depth and robust versions of these measures but as the subject was already collecting many other measures in their daily life they wanted to keep this one brief. This combined PHQ-9/GAD-7 Google Form had the same data flow as the Mood form (Figure 3.13) but was completed only once every 2 weeks instead of daily. The PHQ-9/GAD-7 data collection started with Experiment 2 on 11/13/2022. With Table 3.8 we can see each day this Form was completed and the start of each of the phases.

Date	Day of Experiment	Start of Phase
11/13/2022	1	B1
11/27/2022	15	A1
12/11/2022	29	B2
12/25/2022	43	A2
1/8/2023	57	B3
1/22/2023	71	A3
2/5/2023	85	

Table 3.8: Experiment 2 Phases

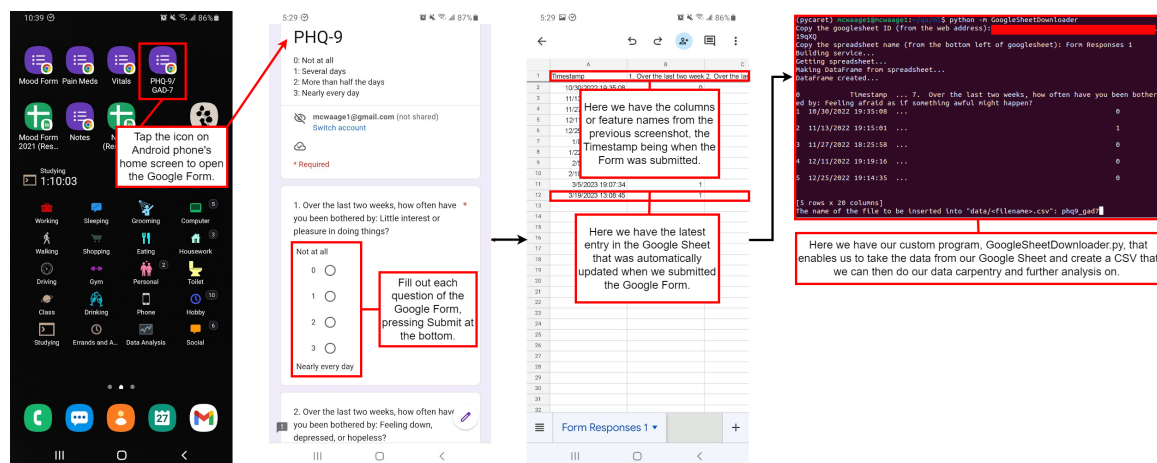


Figure 3.13: PHQ-9/GAD-7 Data Flow

Unlike the subjects in the Shah et al. paper [1], this study did not require or draw its subjects from a clinically depressed population, and we therefore did not find any relationships for these items over the period studied. If the subject were to score high enough on these measures then consideration for stopping the experiment altogether could be made.

Overall the Mood dataset was extremely thorough and details much about the subject, and for

any future self-experimenters there is much that can be learned collecting even one of these measures regularly and creating basic visualizations or statistical analyses about them, but the next section may reveal even more insights.

3.3.2 Time Tracking and Activity Data

While the Mood data may be the heart of the study and main motivator for analysis, the Time Tracking and Activity Data is where some of the most revealing insights and opportunities for change and improvement for each of the variables like Mood comes from. First we will cover how this data was created, how the system was set up, descriptions of each of the activities recorded, any problems that may have affected this data source, and then we will continue on to how it was transformed for further use, including any visualizations, tables, and other statistical analyses as needed.

The Time Tracking and Activity data was collected for Experiment 1 during the entire 3 year period (Figure 3.1), through the aTimeLogger app [35] which is available on the Android Play Store or Apple App Store. As is recommended by Daskalova et al. [4], a period of prototyping and exploration before the main experiment begins was conducted. During this period, the last few weeks of 2019, the subject tried several different apps and methods for the purpose of time tracking and after some refining, revision of the activities, and better understanding of the dataflow, found that the aTimeLogger app filled their needs for data collection the best. With Experiment 1 started, we can see the flow for the time tracking and activity data source now. First we will view how the data was actually recorded throughout the day and a more detailed description of each of the activities.

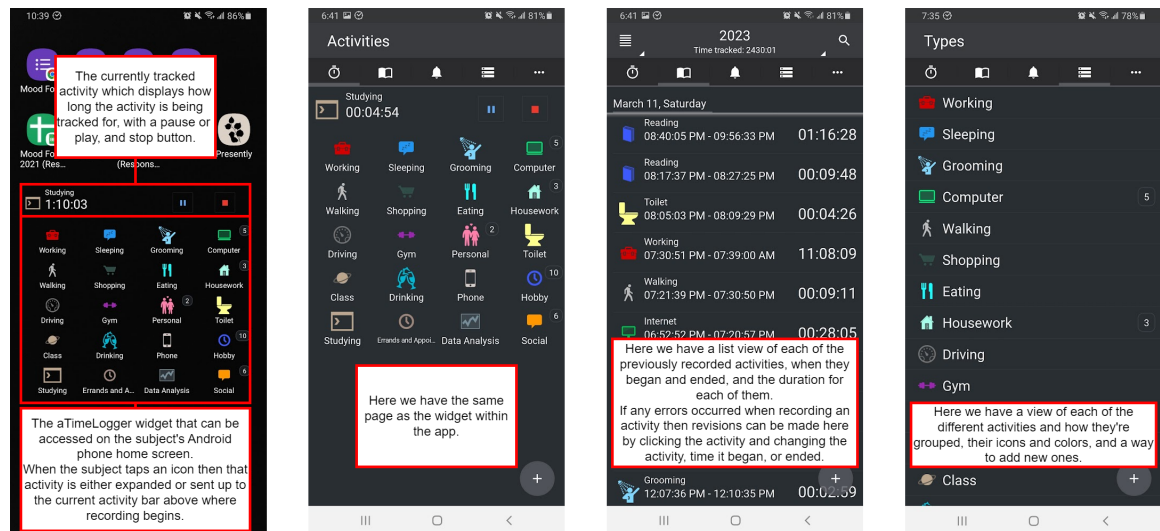


Figure 3.14: aTimeLogger Process

Here (Figure 3.14), we have the flow for how aTimeLogger was used throughout the day. The widget on the subject’s Android smartphone home screen was pressed as each activity such as Work, Sleep, or Walking, occurred. The subject pressed the icon when the activity started and tapped the stop button as they ended. If any errors occurred, such as incorrectly selecting the wrong activity, then revisions or deletions could be made in the app. Periodic review of the list of activities was performed to ensure that missing or incorrect activities were reduced to a minimum. For organizational purposes a file structure was set up within the app and several groups were created, such as the Hobbies and Social categories, where more specific activities of that nature were grouped together. With Figure 3.9, we will now describe each of the activities in more detail.

The activities were originally based on the American Time Use Survey [42], but were modified and made more specific to the subject’s personal activities that they experienced, performed, or took part in. Multiple activities could be selected through the app but this was reduced to a minimum by the subject. The option was there for the subject to select Eating and Drinking during Friends Time, but only two were typically selected as a self-imposed maximum number of activities at any one moment.

Despite the activities being a block of time, there were instances where the subject would briefly stop performing the activity such as getting something from another room while the selected activity in the app continued running, an example being getting up to grab some food while Internet or Movies/TV continued to record. Only when the activity was completely physically stopped, and a new activity began, was the activity also stopped in the app. Now we will describe the flow from the everyday recording to data that could be used for further analysis.

With Figure 3.15, we can see how this data source had a similar data acquisition flow to the Mood data from before with; (1) tapping on aTimeLogger widget icons on their phone’s home screen as activities throughout the day occurred, to (2) sending that report to a Google Sheet, to (3) downloading that data through the Google API with our custom `GoogleSheetDownloader.py`. At the end of this process we have created a CSV file that can then be sent on to the Data Carpentry layer (Figure 3.3).

Within the Time and Activity Data Carpentry step is our custom preprocessing program, `ActivityStats.ipynb`, that converts the activity data created by the aTimeLogger app into a form that can be merged with all of the other data on their linked Date columns. Table 3.10 and Table 3.11 below shows what that transformation from one to the other looks like.

Our custom program creates a CSV file that is indexed on the Date like the Mood data. This program finds each of the activities for each day and counts how frequent they occurred during

	Activity	Activity Description
1	Work	Time spent at work or on the job
2	Sleeping	Time spent trying to sleep or nap
3	Grooming	Time spent in the shower, brushing teeth, or on other personal care activities
4	Movies/TV	Time spent watching TV, streaming movies, or in a movie theater
5	Games	Time spent playing videogames
6	Internet	Time spent browsing the internet or generally on the computer
7	Programming	Time spent on personal programming projects
8	GameDev	Time spent on personal game development projects
9	Walking	Time spent walking from one location to another or as exercise
10	Shopping	Time spent shopping for clothes or food in person
11	Eating	Time spent eating
12	Cleaning	Time spent cleaning, including vacuuming, doing the laundry, or scraping snow off the car
13	Dishes	Time spent cleaning the dishes
14	Cooking	Time spent cooking
15	Driving	Time spent driving
16	Gym	Time spent in the gym
17	Personal	Time spent on personal activities
18	Toilet	Time spent in the bathroom
19	Class	Time spent in class or school functions
20	Drinking	Time spent drinking alcohol
21	Phone	Time spent making calls (or on the phone when no other activity applied)
22	Drumming	Time spent playing the drums
23	Writing	Time spent creative writing
24	Reading	Time spent reading books or articles
25	Keyboard	Time spent playing the keyboard
26	Drawing	Time spent drawing
27	Swimming/Tanning	Time spent at the pool or outside in the Sun
28	Journaling	Time spent gratitude journaling
29	Traveling	Time spent flying or traveling
30	Biking	Time spent riding the bike
31	Mycology	Time spent on mycology
32	Studying	Time spent studying, including homework, reading articles, and programming related to classwork
33	Errands and Appointments	Time spent on errands and appointments such as going to the bank
34	Data Analysis	Time spent on data analysis (before programming took over)
35	Dating	Time spent on dating and romantic relationships
36	Family Time	Time spent with family
37	Friends Time	Time spent with friends
38	Concert/Event	Time spent at concerts or other events
39	Dancing	Time spent dancing
40	Club/Social Group	Time spent at social clubs, groups, or volunteering

Table 3.9: Activities Description

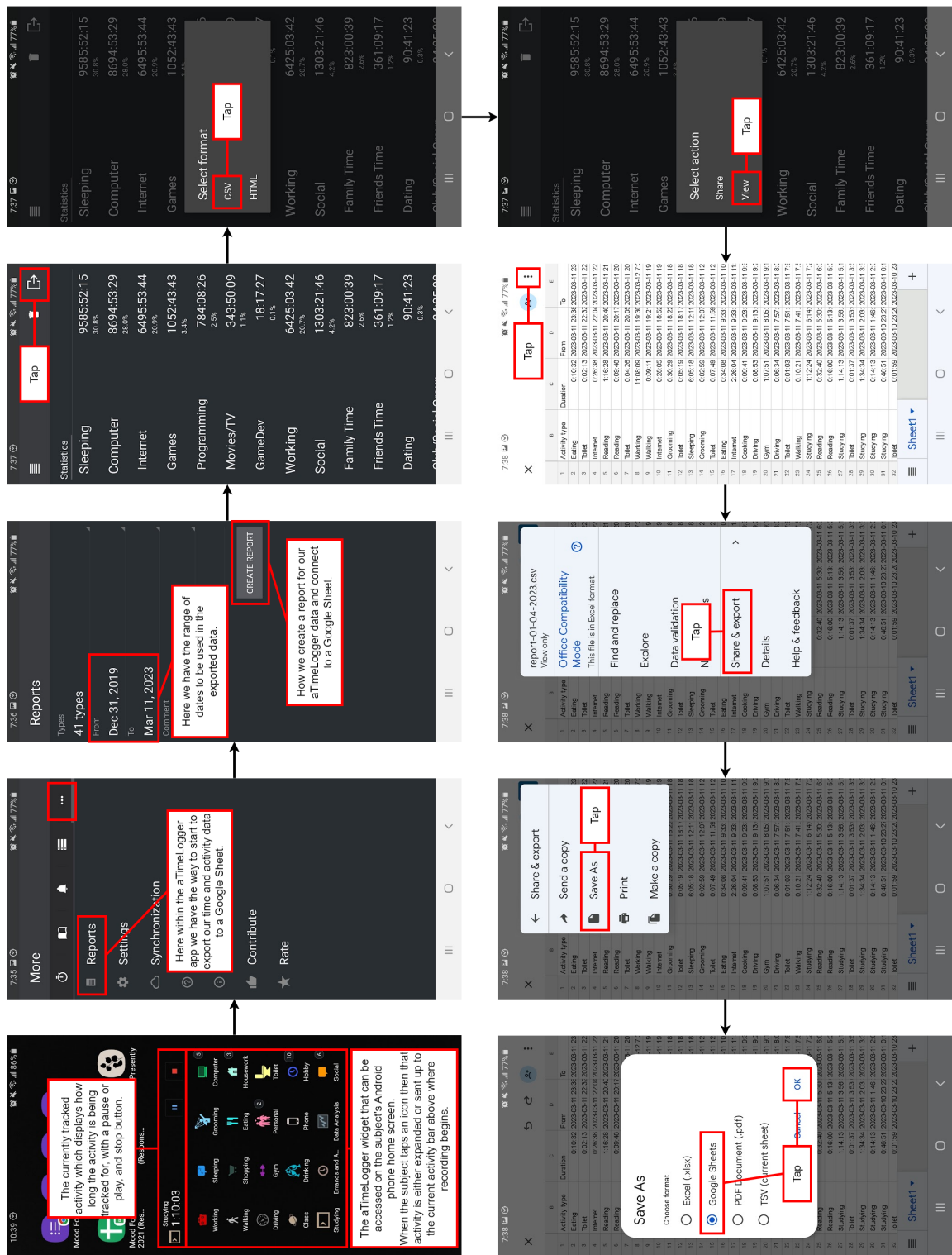


Figure 3.15: Time Data Flow

	Group	Activity type	Duration	From	To	Comment
1		Toilet	0:17:55	2023-03-12 23:56:28	2023-03-13 0:14:23	
2		Studying	0:19:21	2023-03-12 23:35:58	2023-03-12 23:55:19	
3		Eating	0:11:26	2023-03-12 23:14:19	2023-03-12 23:25:45	
4		Toilet	0:02:06	2023-03-12 22:38:49	2023-03-12 22:40:55	
5		Studying	1:00:01	2023-03-12 21:35:43	2023-03-12 22:35:44	
...	...	...	...	...	...	...
6	Computer	Internet	3:08:53	2019-12-31 8:40:09	2019-12-31 11:49:02	
7	Housework	Cooking	0:13:30	2019-12-31 8:20:10	2019-12-31 8:33:40	
8	Computer	Internet	2:26:52	2019-12-31 5:52:35	2019-12-31 8:19:27	
9		Toilet	0:01:21	2019-12-31 5:48:13	2019-12-31 5:49:34	
10		Sleeping	4:17:50	2019-12-31 1:29:56	2019-12-31 5:47:46	

Table 3.10: aTimeLogger Table

Date	sleeping_durations_seconds	sleeping_counts	sleeping_durations	sleeping_durations_seconds_7_day_rolling_mean	internet_durations_seconds	internet_counts	internet_durations	internet_durations_seconds_7_day_rolling_mean	working_durations_seconds	working_counts	working_durations	working_durations_seconds_7_day_rolling_mean
2020-01-01	20201	2	7:17:41	0	37528	7	10:20:38	0	0	0	0	0
2020-01-02	65405	2	15:23:25	0	29718	3	8:15:18	0	0	0	0	0
2020-01-03	41486	2	11:31:06	0	39231	4	9:43:51	0	2241	1	5:37:21	0
2020-01-04	15951	1	4:18:21	0	38116	2	1:03:36	0	43764	2	12:42:44	0
2020-01-05	16834	1	4:40:34	0	36763	5	10:12:43	26571	28089	1	7:47:49	0
...	...	...	...	...	...	...	...	...	...	...	...	...
2023-03-02	31331	1	8:42:11	26734	50453	5	14:00:53	30353	0	0	0:00:00	18852
2023-03-03	52740	2	14:38:00	28965	5121	2	1:25:21	27428	1947	1	4:24:07	18838
2023-03-04	55274	1	7:51:14	29364	12100	4	3:22:00	26819	44525	2	12:19:13	18800
2023-03-05	23667	1	6:17:47	30249	14425	3	4:00:25	26922	44809	2	12:26:49	18990
2023-03-06	24738	1	6:52:18	30649	19198	4	5:19:58	25045	28145	1	7:49:05	19022

Table 3.11: Time Data Table

	sleeping_durations_seconds	sleeping_counts	sleeping_durations_seconds_7_day_rolling_mean	internet_durations_seconds	internet_counts	internet_durations_seconds_7_day_rolling_mean	working_durations_seconds	working_counts	working_durations_seconds_7_day_rolling_mean
count	1166	1166	1166	1166	1166	1166	1166	1166	1166
mean	29557.0223	1.388507719	29405.39194	20202.2916	3.934819897	19925.45455	19839.42024	0.925389348	19748.40909
std	9767.187219	0.5703744749	3130.126646	12155.45622	1.83598889	6217.34171	17858.84017	0.7929456012	5881.710501
min	0	0	0	0	0	0	0	0	0
25%	22991.25	1	27970.75	10549	3	15313.75	0	0	17240.25
50%	27718	1	29580	18063.5	4	19465.5	16587.5	1	19167.5
75%	35983.25	2	31168.75	27679.5	5	24296.75	32404	2	23290.5
max	62138	4	36714	62929	12	37602	72596	3	38024

Table 3.12: Time Statistics

that day, how much time was spent in seconds on that activity per day, the 7 day rolling average for each activity, and a more human readable time format of that activity that is used for several charts and graphs. As we can see in Table 3.9 there are 40 activities total while this program creates four columns per activity for each day, thus creating a table 161 wide including the Date column. Table 3.12 shows a snippet of that processed table's general statistical distributions

As with the Mood data there are many chances for confounds, outliers, and missing or incorrect data to arise. Many of the time-related confounds may actually be represented with this data though, such as any changes in life like social life changes, or work and sleep schedules transitions. The differences in the various activities and how they change over time really do represent those different life changes. Something else to note is that due to the nature of time-tracking data, there are only so many minutes or seconds in a day, and despite some of the days having more time tracked than there is actual time in the day, due to the possibility of multiple activities being tracked simultaneously, there is a Poisson distribution in effect for much of the Time and Activity Data.

There were days in the 3 year period of data collection where some of the activities became outliers. One example being a several day period where the subject spent time Driving across a few states and then back, resulting in a massive amount of Driving time relative to the rest of the dataset. This and other points may be outliers, but they were not discarded.

Missing or incorrect data was harder to identify because the time data was created constantly and if a period of time passed and a review did not occur or had not caught an anomaly then there was typically no way to know which activity was missing or incorrectly attributed. There were some instances where it was obvious however, such as a long period of Grooming when the subject normally was Sleeping. In cases like these then the subject could guesstimate that they accidentally misattributed the activities, and a correction was made for that instance in-app. True missing data, such as an activity that was not tracked at all for a period, was much more difficult to identify and replace and for the most part was left missing unless the event was memorable in some way. Some activities were more inconsistently tracked as well such as Eating, which may or may not have been tracked if the subject quickly ate something small like a snack, unless it was especially filling or caloric dense.

We will look at a few of the more interesting time related charts and graphs for this data source now. First we will look at our custom timeline chart that shows each activity's per-second tracked data for the entire 3 year period.

Originally created as an interactive dashboard where the user was able to hover over any of the bars for more specific information about each activity, the Figure 3.16 timeline chart shows the

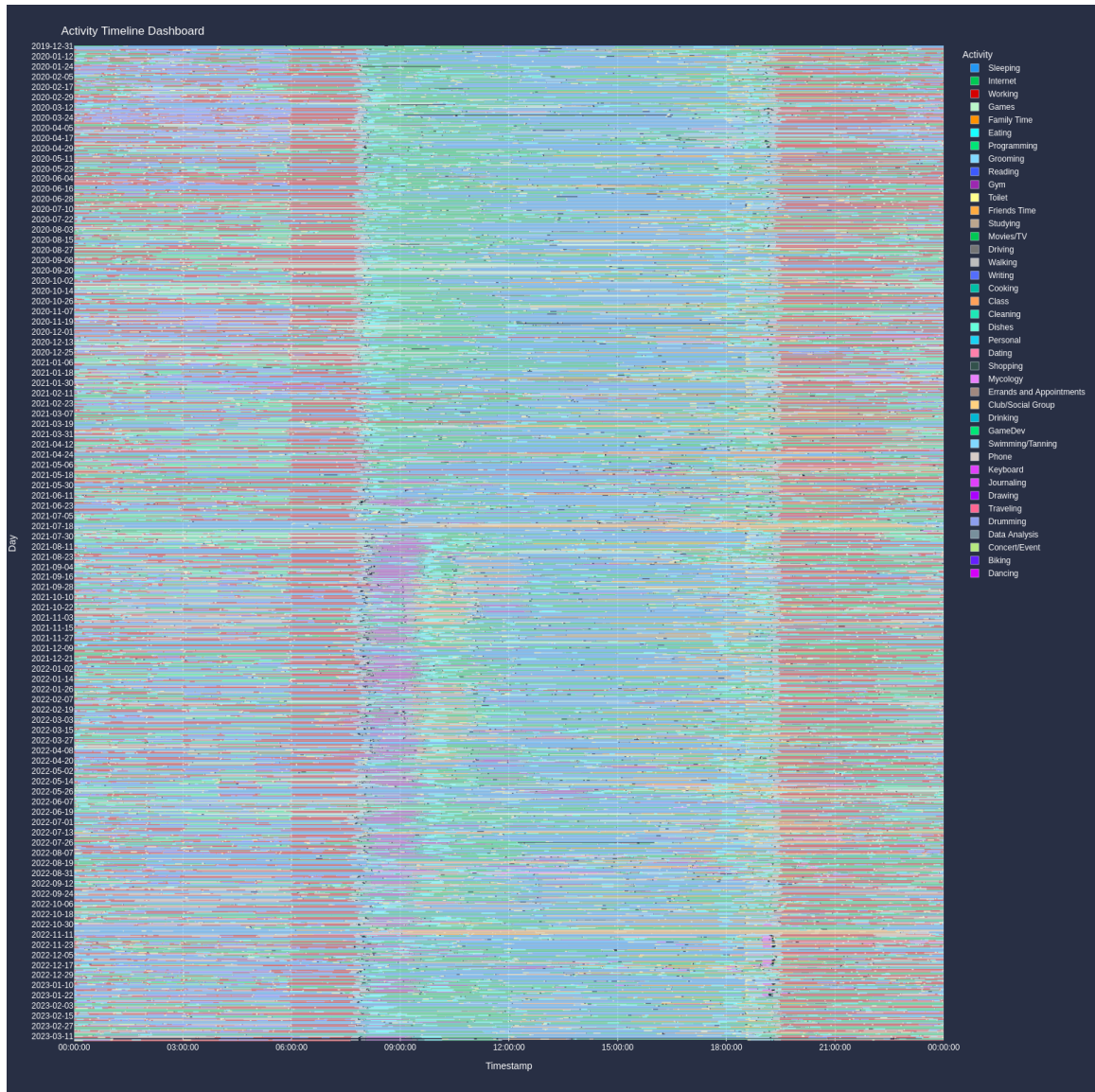


Figure 3.16: Activity Timeline

Here we have a timeline of the subject's life for the entire 3 year period. On the Y-axis we have each Date and on the X-axis we have the within-day hourly and minute data. Each horizontal colored bar represents an instance of an activity for each day. We can see the general trend for the subject to wake up at around 6 p.m., go to work at 7:30 p.m. and work through the night until 8 a.m. In May of 2021 we can see a change with the added time spent going to the Gym after work for about an hour each day.

subject's entire 3 year period of their personal per-second time tracked data. This is not only a timeline chart on one time axis, it is on two time axes, more clearly showing the shifting nature of activities through the days. The Y-axis, in descending order from the oldest to most recent date, represents each day in the time dataset, while on the X-axis we have each particular activity represented by a horizontal colored bar which shows that particular instance's start and end times, from 00:00:00 to 23:59:59. With Table 3.13 we can see a detailed statistical description of each of the activities for the entire 3 year period which includes statistics like the total number of instances of that activity, how many times on average that activity occurred per day, the percentage of the time that activity occurred, and the minimum, maximum, and quartile distributions as well, among several others. Table 3.14 shows us a more human readable average time spent on each activity plus or minus the standard deviation.

Table 3.15 shows us the total time tracked in the dataset, the total number of instances of all of the activities, the total average number of activities per day and the standard deviation activities per day, as well as the minimum, maximum, and quartile ranges of the activities per day. We will note here that the subject had on average about 27 activities throughout the day, plus or minus 7. This statistic shows how frequently the subject would take their phone out and go through the process described in Figure 3.14. In this process the subject would get out their phone and tap one of the activities on their home screen to start the new instance of activity tracking and, generally at the same moment, stop another activity, on average about 27 times per day for the entire 3 year period.

As the Figure 3.16 chart and descriptive statistics are perhaps only truly interesting for deep inspection, analysis, and introspection specific to the self-experimenter, with their unique domain knowledge of themselves, we will move on to more specific activity timelines and their related charts and statistics. There were 3 activities that took up the lion's share of time in the subject's life for the 3 year period which we will examine more in depth now.

Sleeping is a large part of everyone's lives, and for this subject that was no different. The subject spent an average of about 7 hours per day Sleeping for a total of about 34% of their life for the duration of the data collection. One irregularity of note is that the subject was typically a night shift worker for the entire 3 year period and thus had the opposite schedule from the average person, waking up at about 6 or 6:30 p.m. and Sleeping anywhere from 10:00 a.m. to about noon. The abrupt cutoffs are visible in Figure 3.17 when they were woken up from an alarm, however, on off days their sleep could be completely flipped as they may have shifted to dayshift for a short period.

Originally an interactive dashboard as well, the longer upper tail seen in Figure 3.18 is responsible for bringing the average up somewhat, with a minor bimodal distribution around the 60,000 second

	index	Sleeping	Internet	Working	Games	Family Time	Eating	Programming	Grooming	Reading	Gym
	total_num_instance_activity	1351	4312	551	355	219	3533	736	2364	1415	386
	mean_num_activity_per_day	1.16	3.69	0.47	0.3	0.19	3.02	0.63	2.02	1.21	0.33
	activity_percent	34.215	23.182	22.995	3.755	2.936	2.860	2.801	2.239	1.959	1.497
	mean_activity_start	9:33:52	12:10:07	9:58:04	9:06:27	15:24:50	12:14:13	6:29:55	14:00:12	6:22:43	9:53:45
	median_activity_start	10:41:17	11:31:13	7:56:00	6:45:48	15:36:39	11:41:24	3:20:01	14:23:26	3:32:34	8:23:01
	mean_activity_end	15:28:26	13:34:56	15:55:29	11:44:21	19:10:19	12:27:42	7:29:29	14:16:07	6:45:41	10:58:58
	median_activity_end	17:09:36	13:01:19	15:57:14	10:06:52	19:50:21	11:53:33	4:37:25	14:43:24	3:57:16	9:28:20
	sum_activity_dhms	399:15:10:38	270:18:26:03	268:13:54:43	43:20:43:43	34:07:00:39	33:09:37:59	32:17:08:26	26:03:46:16	22:21:00:36	17:11:30:44
	mean_activity_durations_hms	7:05:58	1:30:25	11:41:55	2:57:56	3:45:29	0:13:37	1:04:00	0:15:56	0:23:17	1:05:13
	std_activity_durations_hms	2:53:40	1:19:51	1:47:24	2:06:40	2:08:35	0:09:06	0:42:11	0:12:46	0:16:23	0:06:53
	min_activity_durations_hms	0:00:01	0:00:01	0:40:52	0:05:14	0:02:02	0:00:00	0:04:08	0:00:01	0:00:01	0:22:26
	25_quant_activity_durations_hms	5:30:56	0:32:08	12:11:01	1:26:56	2:11:47	0:06:54	0:34:43	0:03:58	0:13:38	1:01:48
	median_activity_durations_hms	6:57:03	1:08:05	12:20:17	2:22:39	3:56:30	0:11:26	0:51:32	0:11:30	0:21:03	1:03:51
	75_quant_activity_durations_hms	8:42:35	2:03:42	12:26:14	4:06:59	5:06:38	0:18:22	1:27:03	0:26:33	0:27:08	1:07:18
	max_activity_durations_hms	19:51:44	11:04:28	13:36:37	13:46:25	10:01:43	1:20:42	4:54:31	1:36:36	3:04:51	1:28:35
	index	Toilet	Friends Time	Studying	Movies/TV	Driving	Walking	Writing	Cooking	Class	Cleaning
	total_num_instance_activity	8238	64	359	206	1564	1402	274	1051	107	326
	mean_num_activity_per_day	7.05	0.05	0.31	0.18	1.34	1.2	0.23	0.9	0.09	0.28
	activity_percent	1.431	1.288	1.287	1.227	0.990	0.971	0.702	0.582	0.537	0.411
	mean_activity_start	12:13:13	12:23:03	7:50:48	10:45:58	13:29:15	13:18:55	7:26:03	12:16:42	10:39:05	16:01:48
	median_activity_start	11:45:35	13:51:09	3:28:25	9:19:42	13:02:38	10:48:18	4:08:06	9:57:35	9:26:42	17:00:12
	mean_activity_end	12:16:08	17:08:10	8:47:18	12:15:39	13:39:53	13:30:32	8:07:37	12:26:01	12:03:29	16:22:56
	median_activity_end	11:48:42	20:41:20	4:54:27	10:46:15	13:13:44	11:02:24	4:53:07	10:08:45	10:50:55	17:24:40
	sum_activity_dhms	16:17:07:50	15:01:09:17	15:00:44:12	14:07:50:09	11:13:25:17	11:08:17:26	8:04:44:01	6:19:16:21	6:06:30:59	4:19:10:28
	mean_activity_durations_hms	0:02:55	5:38:35	1:00:17	1:40:09	0:10:39	0:11:39	0:43:05	0:09:19	1:24:24	0:21:12
	std_activity_durations_hms	0:03:00	3:59:45	0:39:46	1:08:52	0:10:11	0:07:58	0:28:13	0:05:14	0:41:30	0:22:59
	min_activity_durations_hms	0:00:15	0:07:55	0:07:01	0:12:01	0:01:20	0:00:01	0:01:18	0:00:00	0:18:47	0:01:28
	25_quant_activity_durations_hms	0:01:09	2:26:26	0:30:40	0:55:28	0:08:01	0:09:41	0:22:11	0:05:18	1:10:32	0:07:58
	median_activity_durations_hms	0:01:42	4:43:55	0:52:27	1:30:00	0:09:15	0:10:19	0:37:17	0:08:07	1:21:09	0:13:22
	75_quant_activity_durations_hms	0:03:07	8:08:42	1:18:42	2:03:21	0:11:30	0:11:30	0:58:42	0:12:38	1:26:46	0:22:36
	max_activity_durations_hms	0:49:11	16:17:30	3:43:27	8:05:13	3:29:33	2:05:18	3:31:12	0:42:12	6:06:00	3:03:05
	index	Dishes	Personal	Dating	Shopping	Mycology	Errands and Appointments	Club/Social Group	Drinking	GameDev	Swimming/Tanning
	total_num_instance_activity	136	688	29	133	57	88	11	52	9	15
	mean_num_activity_per_day	0.12	0.59	0.02	0.11	0.05	0.08	0.01	0.04	0.01	0.01
	activity_percent	0.386	0.357	0.324	0.199	0.112	0.103	0.086	0.082	0.065	0.062
	mean_activity_start	13:02:11	11:15:55	14:09:16	17:14:18	14:43:26	12:52:55	14:04:51	19:03:04	13:14:27	14:50:17
	median_activity_start	12:19:10	11:03:30	17:00:36	18:38:52	15:15:37	13:37:52	12:59:33	20:01:48	17:45:20	14:57:22
	mean_activity_end	13:49:59	11:24:36	16:49:18	17:39:29	15:16:23	13:12:36	16:16:13	19:28:37	15:04:11	15:59:49
	median_activity_end	13:15:40	11:12:32	18:28:49	19:05:56	15:33:31	13:51:22	15:34:34	20:24:46	19:50:16	15:58:58
	sum_activity_dhms	4:12:19:09	4:03:58:46	3:18:41:23	2:07:47:51	1:07:18:24	1:04:51:57	1:00:05:03	0:22:59:54	0:18:17:27	0:17:23:07
	mean_activity_durations_hms	0:47:47	0:08:43	3:07:38	0:25:10	0:32:57	0:19:41	2:11:22	0:26:32	2:01:56	1:09:32
	std_activity_durations_hms	0:24:27	0:06:30	2:46:36	0:15:47	0:42:55	0:25:24	0:29:51	0:32:21	0:39:10	0:13:44
	min_activity_durations_hms	0:05:16	0:02:07	0:36:02	0:01:40	0:02:08	0:01:05	1:18:26	0:00:00	1:02:45	0:35:00
	25_quant_activity_durations_hms	0:29:42	0:04:44	1:18:15	0:14:21	0:06:30	0:03:18	1:53:45	0:06:50	1:55:42	1:02:58
	median_activity_durations_hms	0:46:18	0:06:22	2:06:47	0:21:12	0:11:21	0:05:32	2:19:41	0:15:00	1:58:02	1:11:00
	75_quant_activity_durations_hms	1:02:30	0:10:29	4:00:07	0:31:39	0:44:36	0:31:37	2:27:26	0:31:02	2:36:56	1:19:34
	max_activity_durations_hms	2:00:02	0:53:15	13:38:59	1:45:58	2:42:57	2:16:31	2:52:01	2:35:47	2:47:19	1:26:00
	index	Phone	Keyboard	Journaling	Drawing	Traveling	Drumming	Data Analysis	Concert/Event	Biking	Dancing
	total_num_instance_activity	45	6	45	3	2	16	10	2	1	1
	mean_num_activity_per_day	0.04	0.01	0.04	0	0	0.01	0.01	0	0	0
	activity_percent	0.054	0.047	0.038	0.024	0.024	0.024	0.021	0.015	0.007	0.001
	mean_activity_start	12:16:43	14:38:34	18:31:52	9:33:29	7:37:20	18:16:28	5:36:05	15:49:32	13:57:33	23:24:14
	median_activity_start	12:05:19	18:23:11	18:59:58	5:23:35	7:37:20	19:35:06	4:32:25	15:49:32	13:57:33	23:24:14
	mean_activity_end	12:36:13	16:30:39	18:46:11	11:50:24	10:59:36	18:41:13	6:11:20	17:54:20	15:48:24	23:40:02
	median_activity_end	12:08:19	20:45:13	19:14:36	6:55:33	10:59:36	20:05:10	5:01:14	17:54:20	15:48:24	23:40:02
	sum_activity_dhms	0:15:16:37	0:13:04:38	0:10:44:04	0:06:50:43	0:06:44:31	0:06:35:57	0:05:52:27	0:04:09:36	0:01:50:51	0:00:15:48
	mean_activity_durations_hms	0:20:22	2:10:46	0:14:19	2:16:54	3:22:16	0:24:45	0:35:15	2:04:48	1:50:51	0:15:48
	std_activity_durations_hms	0:20:32	0:29:51	0:01:48	0:49:41	0:07:01	0:05:41	0:16:10	1:56:05	0:00:00	0:00:00
	min_activity_durations_hms	0:00:35	1:44:13	0:10:03	1:31:58	3:17:18	0:13:37	0:17:53	0:42:43	1:50:51	0:15:48
	25_quant_activity_durations_hms	0:03:23	1:52:56	0:14:54	1:50:14	3:19:47	0:20:33	0:23:12	1:23:46	1:50:51	0:15:48
	median_activity_durations_hms	0:12:19	2:00:22	0:15:03	2:08:30	3:22:16	0:24:50	0:30:46	2:04:48	1:50:51	0:15:48
	75_quant_activity_durations_hms	0:33:37	2:17:46	0:15:10	2:39:22	3:24:44	0:29:34	0:40:38	2:45:50	1:50:51	0:15:48
	max_activity_durations_hms	1:32:13	3:05:38	0:16:15	3:10:15	3:27:13	0:33:06	1:07:32	3:26:53	1:50:51	0:15:48

Table 3.13: Activity Statistics

Activity	mean_std_durations
Sleeping	07:05:58 ± 02:53:40
Internet	01:30:25 ± 01:19:51
Working	11:41:55 ± 01:47:24
Games	02:57:56 ± 02:06:40
Family Time	03:45:29 ± 02:08:35
Eating	00:13:37 ± 00:09:06
Programming	01:04:00 ± 00:42:11
Grooming	00:15:56 ± 00:12:46
Reading	00:23:17 ± 00:16:23
Gym	01:05:13 ± 00:06:53
Toilet	00:02:55 ± 00:03:00
Friends Time	05:38:35 ± 03:59:45
Studying	01:00:17 ± 00:39:46
Movies/TV	01:40:09 ± 01:08:52
Driving	00:10:39 ± 00:10:11
Walking	00:11:39 ± 00:07:58
Writing	00:43:05 ± 00:28:13
Cooking	00:09:19 ± 00:05:14
Class	01:24:24 ± 00:41:30
Cleaning	00:21:12 ± 00:22:59
Dishes	00:47:47 ± 00:24:27
Personal	00:08:43 ± 00:06:30
Dating	03:07:38 ± 02:46:36
Shopping	00:25:10 ± 00:15:47
Mycology	00:32:57 ± 00:42:55
Errands and Appointments	00:19:41 ± 00:25:24
Club/Social Group	02:11:22 ± 00:29:51
Drinking	00:26:32 ± 00:32:21
GameDev	02:01:56 ± 00:39:10
Swimming/Tanning	01:09:32 ± 00:13:44
Phone	00:20:22 ± 00:20:32
Keyboard	02:10:46 ± 00:29:51
Journaling	00:14:19 ± 00:01:48
Drawing	02:16:54 ± 00:49:41
Traveling	03:22:16 ± 00:07:01
Drumming	00:24:45 ± 00:05:41
Data Analysis	00:35:15 ± 00:16:10
Concert/Event	02:04:48 ± 01:56:05
Biking	01:50:51 ± 00:00:00
Dancing	00:15:48 ± 00:00:00

Table 3.14: Activity Statistics Standard Deviation

index	activity_df_totals
total_durations_seconds_in_df	110900640
total_durations_dhms_in_df	1283:13:44:00
total_days_in_df	1168
num_unique_activities_in_df	40
total_instances_of_activities_in_df	30162
total_activities_per_day_mean	26.94
total_activities_per_day_std	6.98
total_activities_per_day_min	1
total_activities_per_day_25	22
total_activities_per_day_median	27
total_activities_per_day_75	32
total_activities_per_day_max	54

Table 3.15: Activity Statistics Total

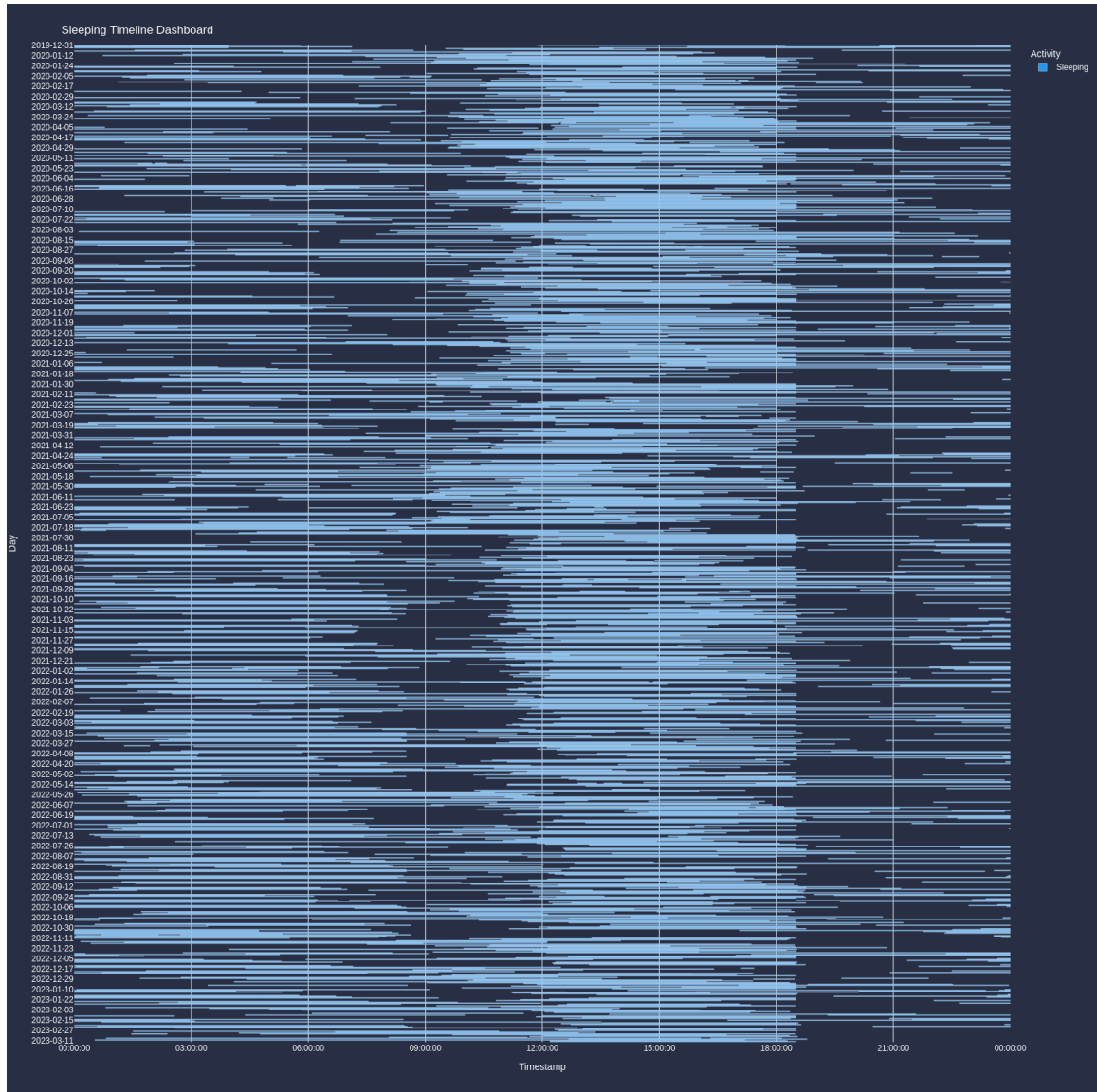


Figure 3.17: Sleeping Timeline

Here we can see how the subject generally woke up at about 6 p.m. for work but had more variability on when they went to sleep depending on if they were working or not.

or about 16 and a half hour mark, each 30,000 seconds being a little over 8 hours. This histogram shows a semi-normal distribution with an abrupt cleft which can be explained by the subject's work schedule, when their alarm was likely to sound and interrupt their sleep.

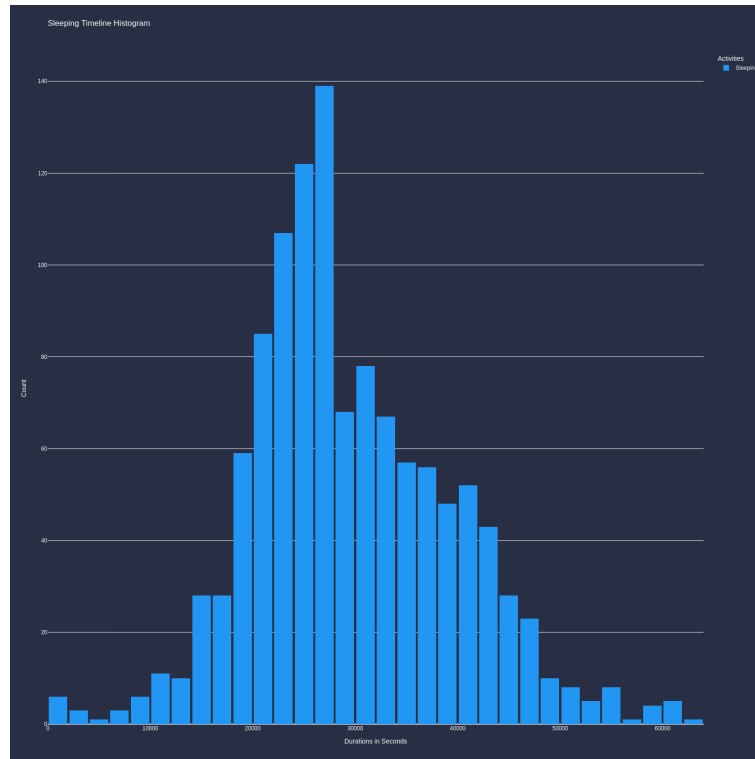


Figure 3.18: Sleeping Histogram

Figure 3.19 and Figure 3.20 show us a couple different views of the Sleeping data, how it is distributed, and how it changed over time. A slight increase in the time spent Sleeping can be seen from the beginning of the data collection to the end, which may be attributed to the slight decrease in Working over time.

Figure 3.21 shows us an interactive Github style heatmap dashboard, each Z value, represented by each box's color, shows us which days were spent Sleeping more or less than average, bands of the brighter blues in this case visible during the week as Working on the weekend, Friday, Saturday, and Sunday nights, became the norm in August of 2021.

Figure 3.22 shows us that change in work schedules, any days where the count of Working for that day was 1 meant that the subject either was starting work that night, typically at 7:30 p.m. or leaving it in the morning at about 8 a.m. A day that had 2 or 3 counts of Working meant that the subject worked both in the morning, starting the new day at 00:00:00 by continuing their shift from the previous night, until the end of their shift at 8 a.m, and then Working again sometime later

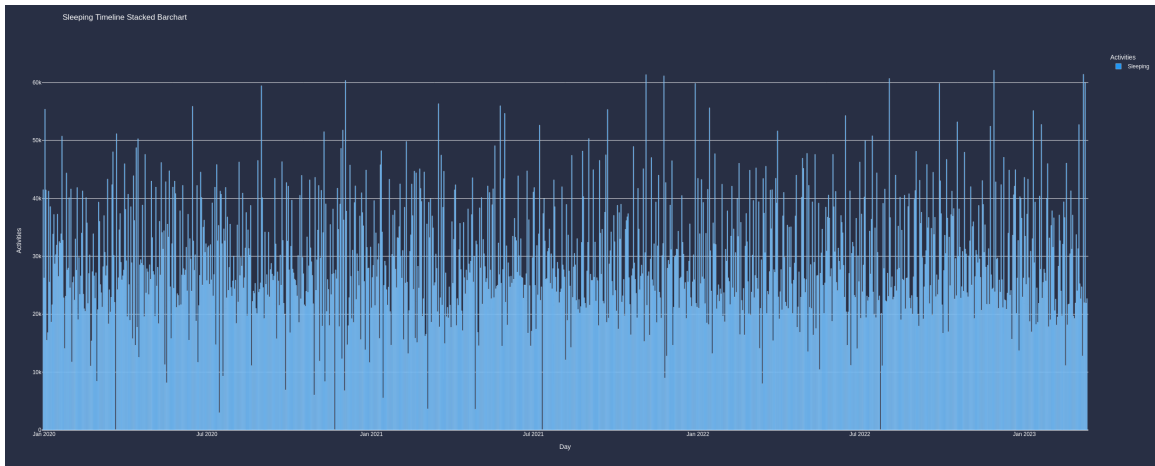


Figure 3.19: Sleeping Timeline Barchart

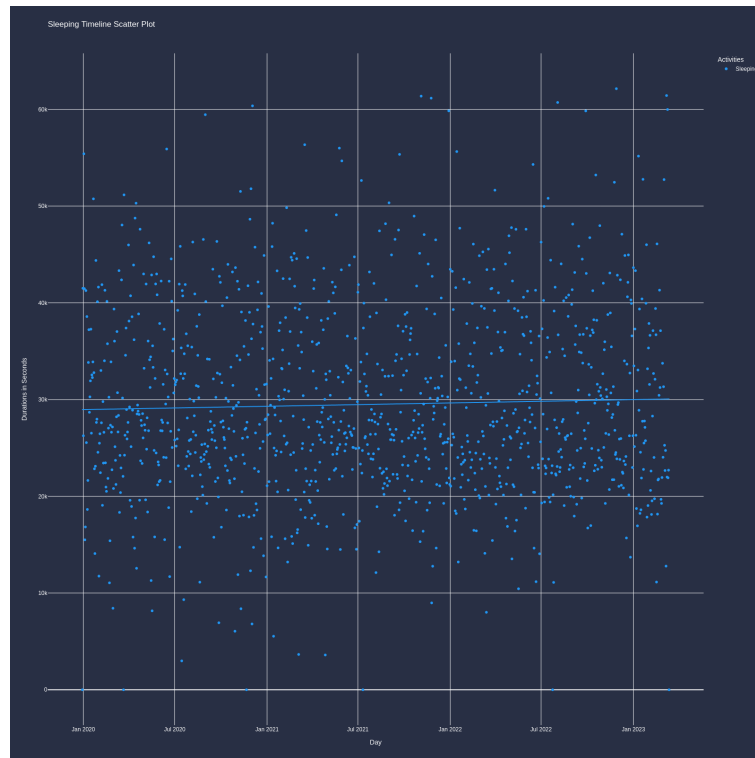


Figure 3.20: Sleeping Timeline Scatter

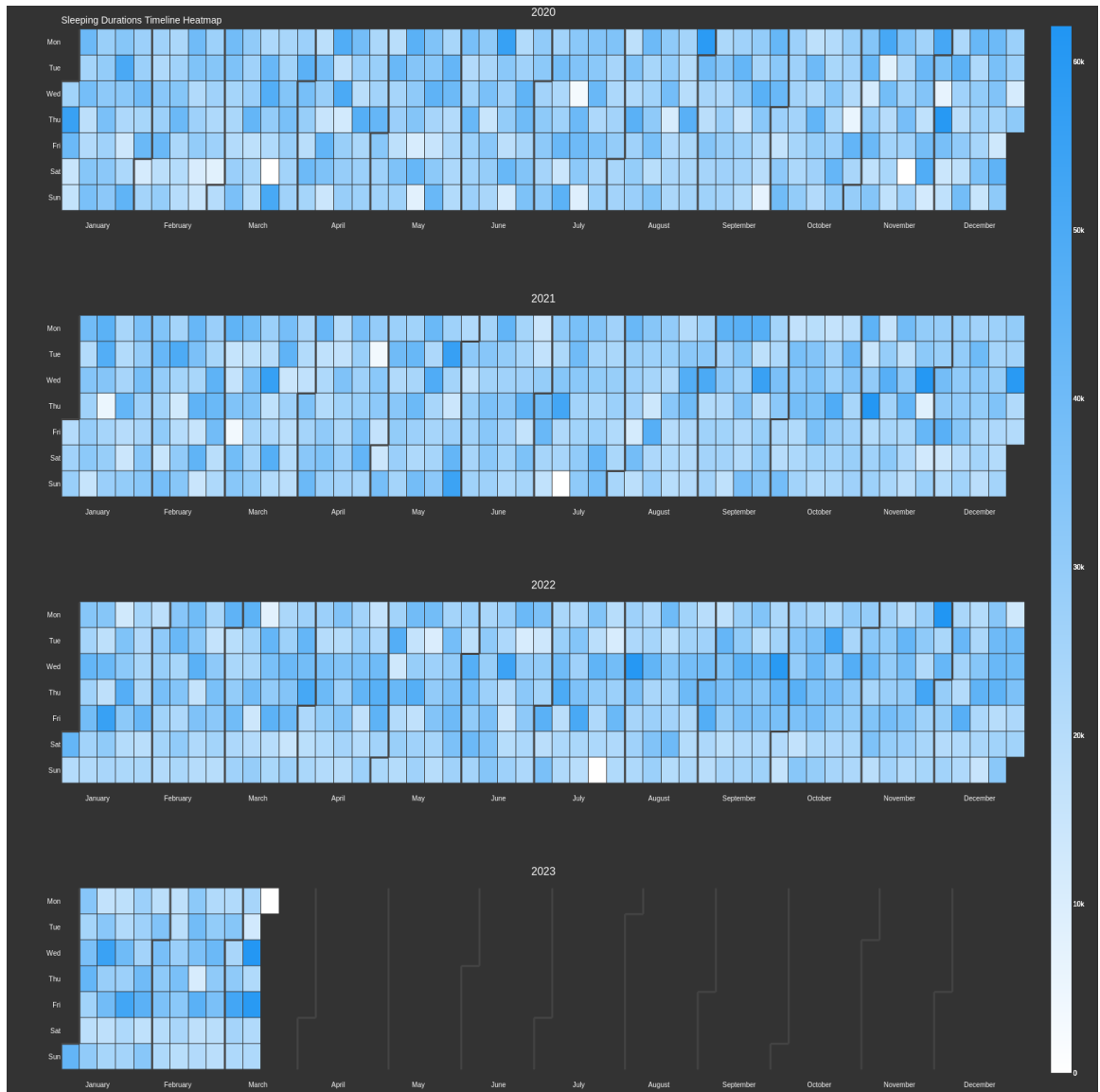


Figure 3.21: Sleeping Durations Timeline Heatmap
 We can see bands of brighter blue during the week starting in 2021 when the subject switched their work schedule to weekends.

that day or night. Figure 3.23 and Figure 3.24 give us the Working schedules in a different view of those distributions as well. Overall, the subject spent less than half the days Working, a little less than 23% of their life during the 3 year period, but for about 12 hours at a time for a work week of typically 36 hours. Together, Sleeping and Working accounted for about 57% of their time, but there was at least one other activity that took a large chunk out of their time as well.

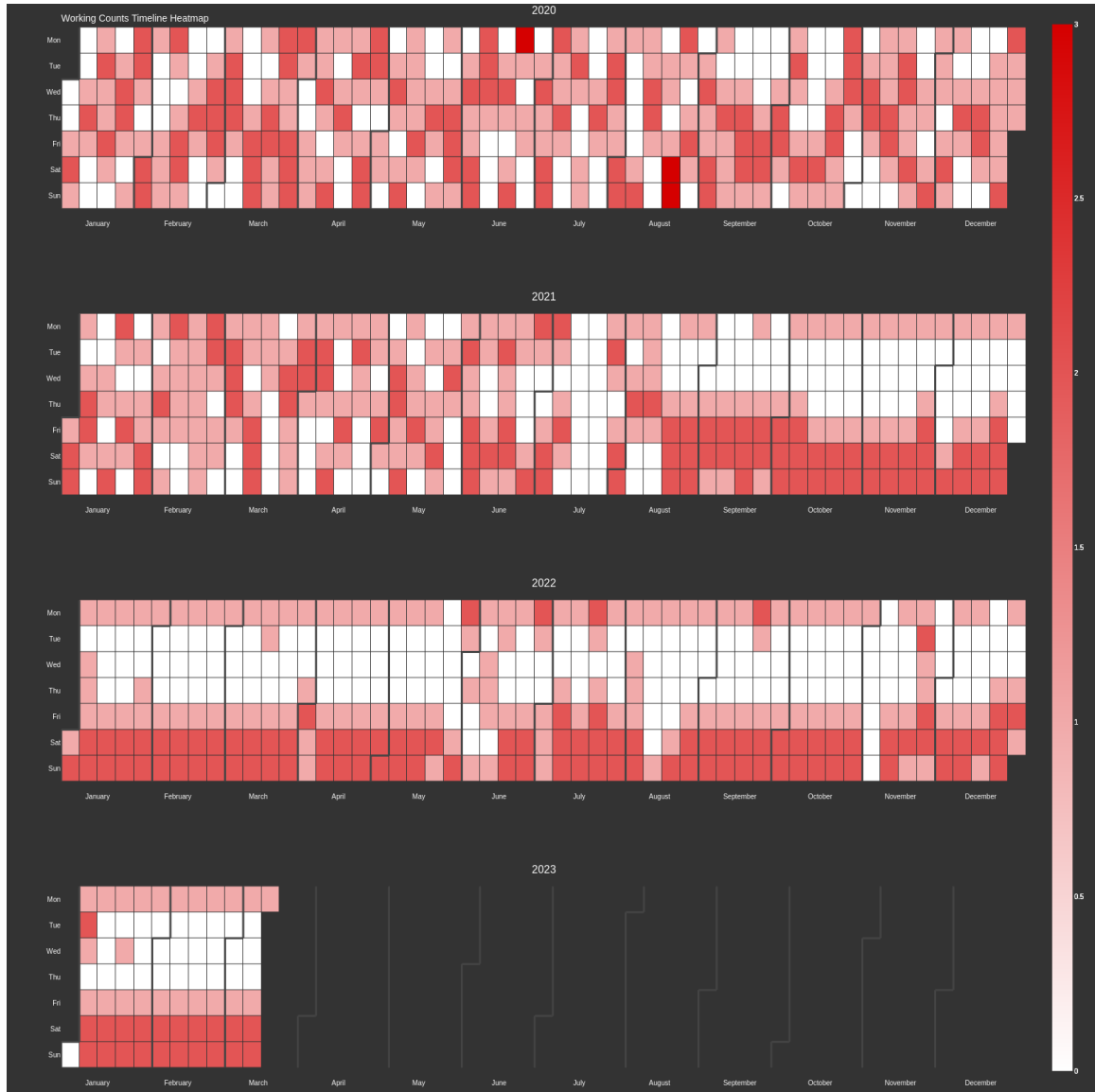


Figure 3.22: Working Counts Timeline Heatmap

Internet and computer usage took up a little more than 23% of the subject's total time, and overall, Sleeping, Working, and Internet took up about 80% of their time, for better or worse. This corresponds to about 19 hours in the day being accounted for with just these 3 activities, leaving all of the other 37 activities being tracked about 5 hours of the day to compete for their share. With

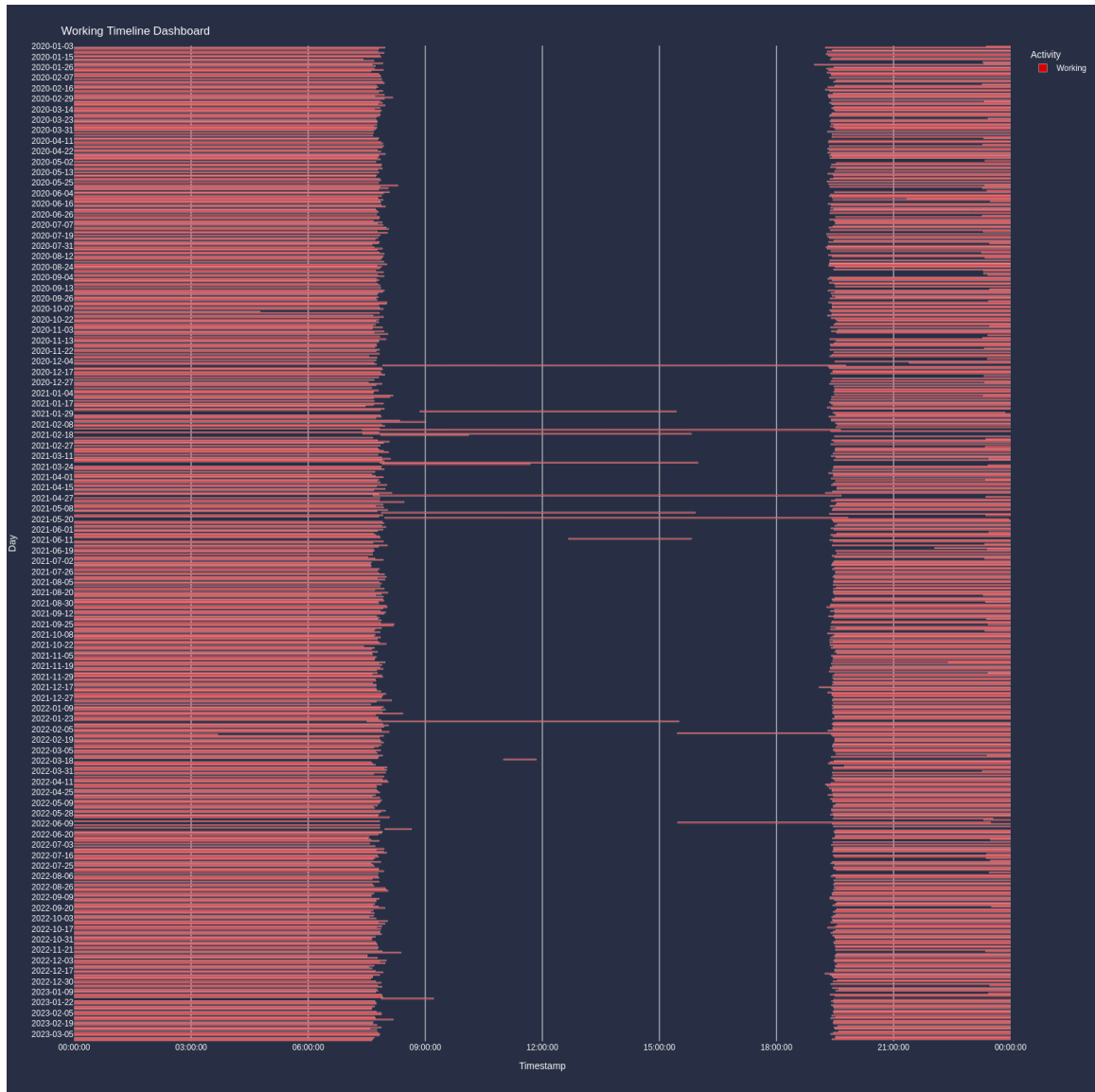


Figure 3.23: Working Timeline

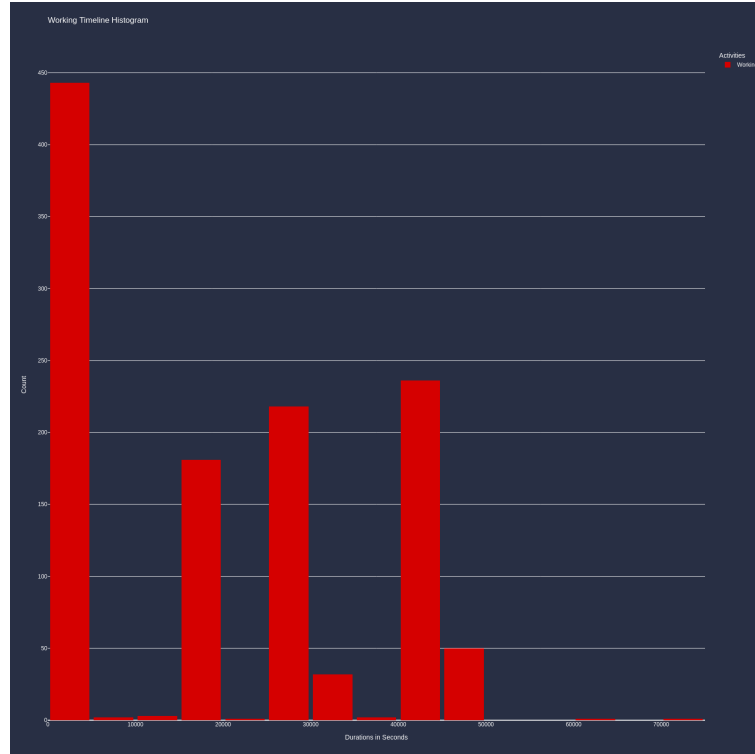


Figure 3.24: Working Histogram

some activities being able to be tracked simultaneously, this is not exactly true though, the Working activity especially susceptible to overlaps such as with Programming, Studying, or Reading. We can see changes in the subject's time-usage however, with an increase in Gym time starting in about May of 2021, where new sections of time were carved out for the increase in the new time-commitment. Figure 3.26 and Figure 3.27 shows that increase in Gym and the corresponding decrease in Internet can be seen with Figure 3.28

Before we move on to the next section, first we will examine the activity that the subject spent time on for Experiment 2, gratitude journaling. The Journaling activity was collected in the same way as the rest of the time tracking data like Working and Sleeping, in the aTimeLogger app. This was typically done for about 10 to 15 minutes per day, for each of the 3 separate 14 day B phases of the experiment, the only notable difference specific to this activity being that the subject also recorded their thoughts in an app, Presently [36], as that time was being recorded. The subject typically did this Journaling at about 7 p.m. for this experiment, right before completing the Mood and PHQ-9/GAD-7 Forms. With this data in a file and able to be easily exported, similar to our other pipelines through our `GoogleSheetDownloader.py`, there is a future for further analysis such as sentiment analysis that could be explored. One thing of note here too is that there is one day in

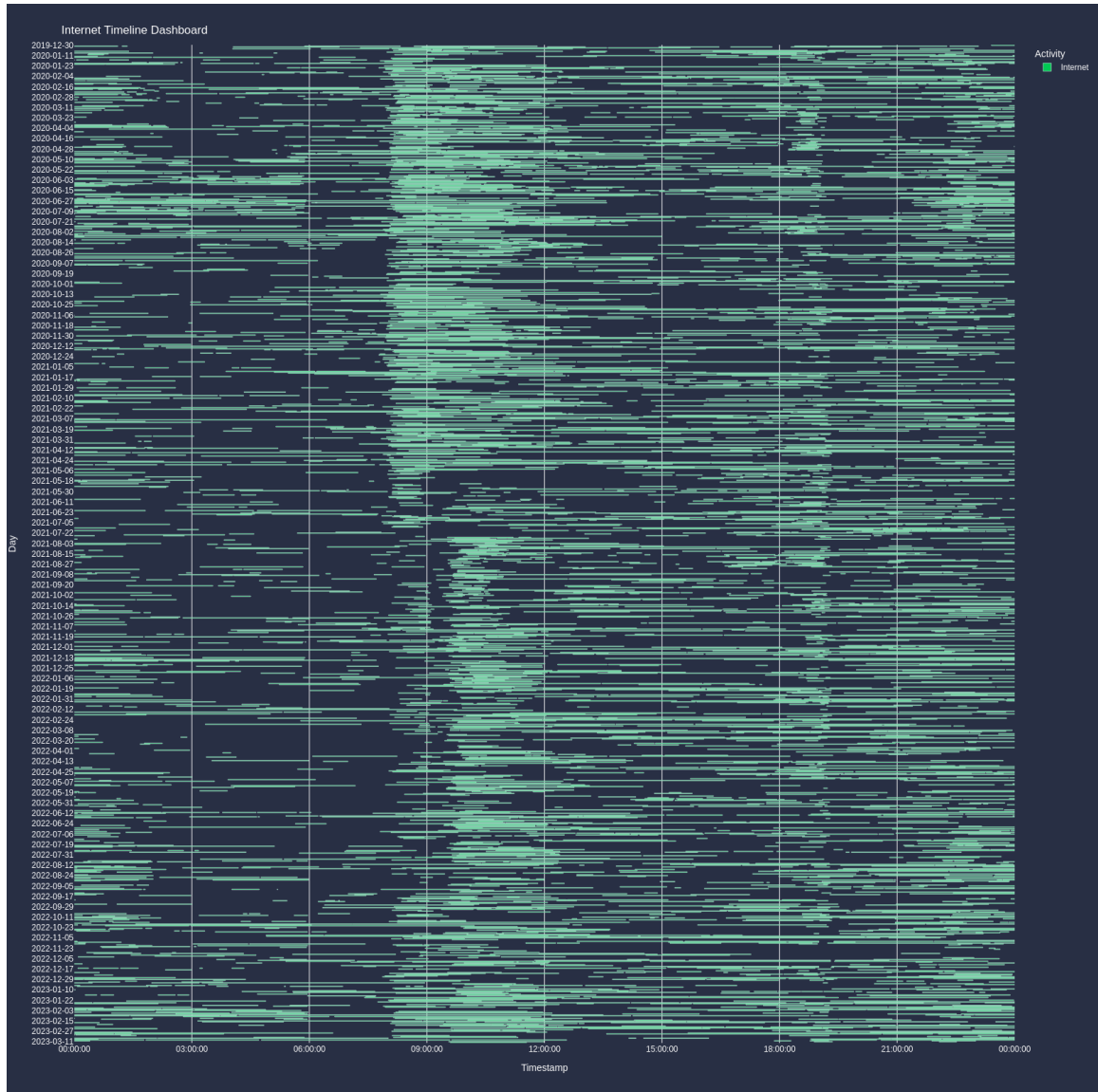


Figure 3.25: Internet Timeline

The shift to more time spent on Internet can be seen beginning in May of 2021, where a chunk of Internet is cut out from when the subject typically arrived home from Working, and instead went to the Gym.

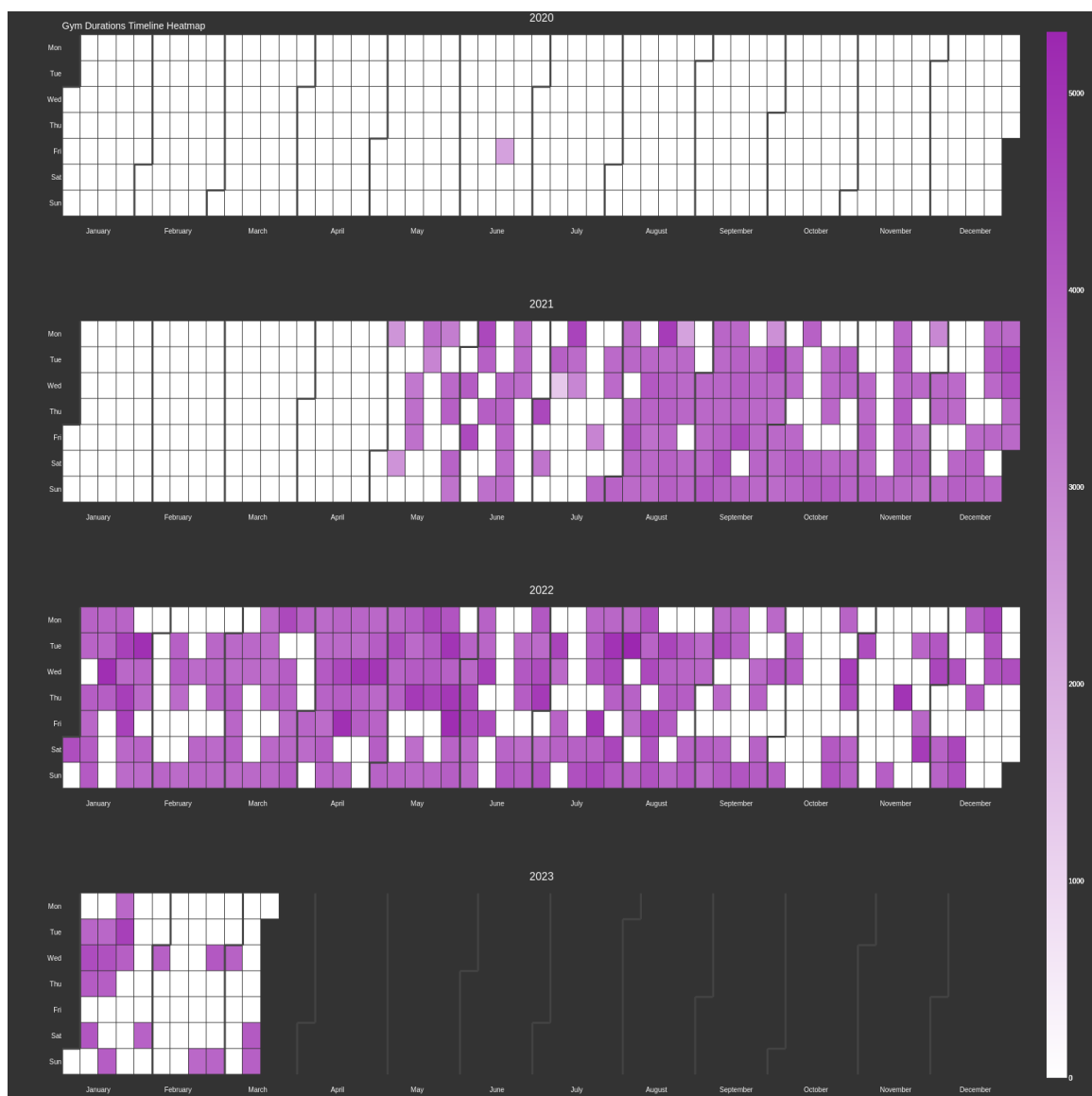


Figure 3.26: Gym Durations Timeline Heatmap

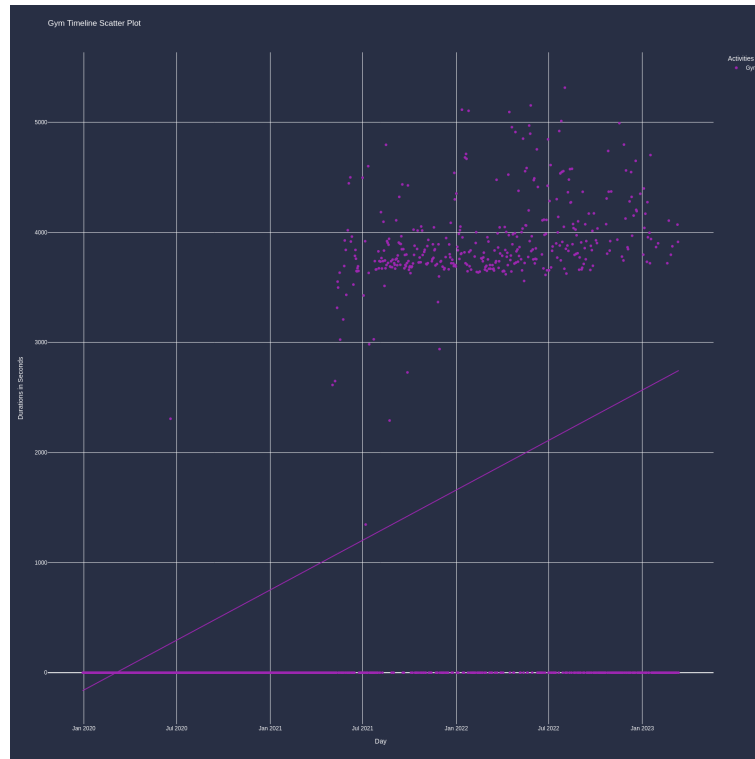


Figure 3.27: Gym Timeline Scatter

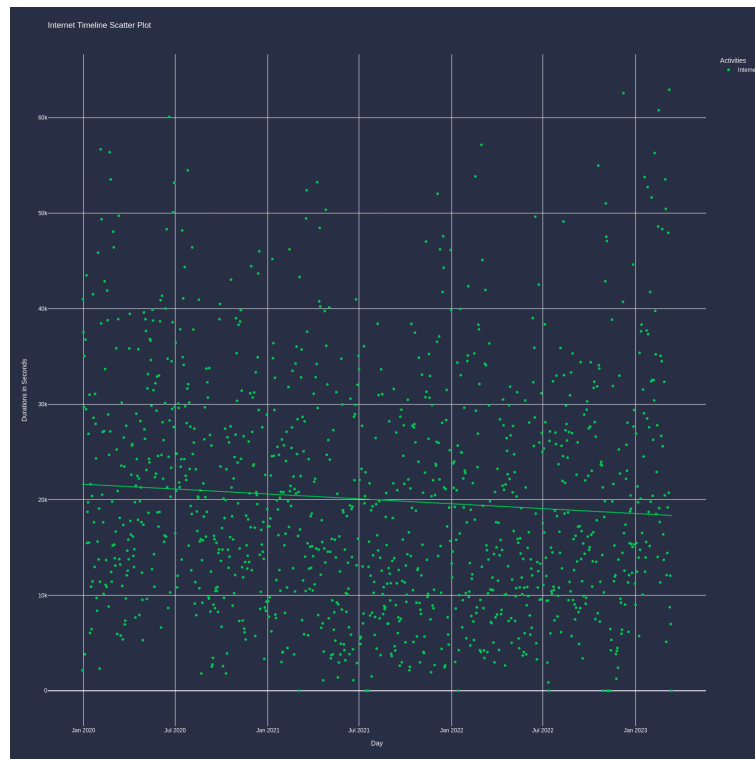


Figure 3.28: Internet Timeline Scatter

the 3rd B phase of the experiment where the subject missed one day of Journaling (Figure 3.29). This was always a possibility and why such a large number of phases and total days in the experiment was considered and ultimately undertaken.

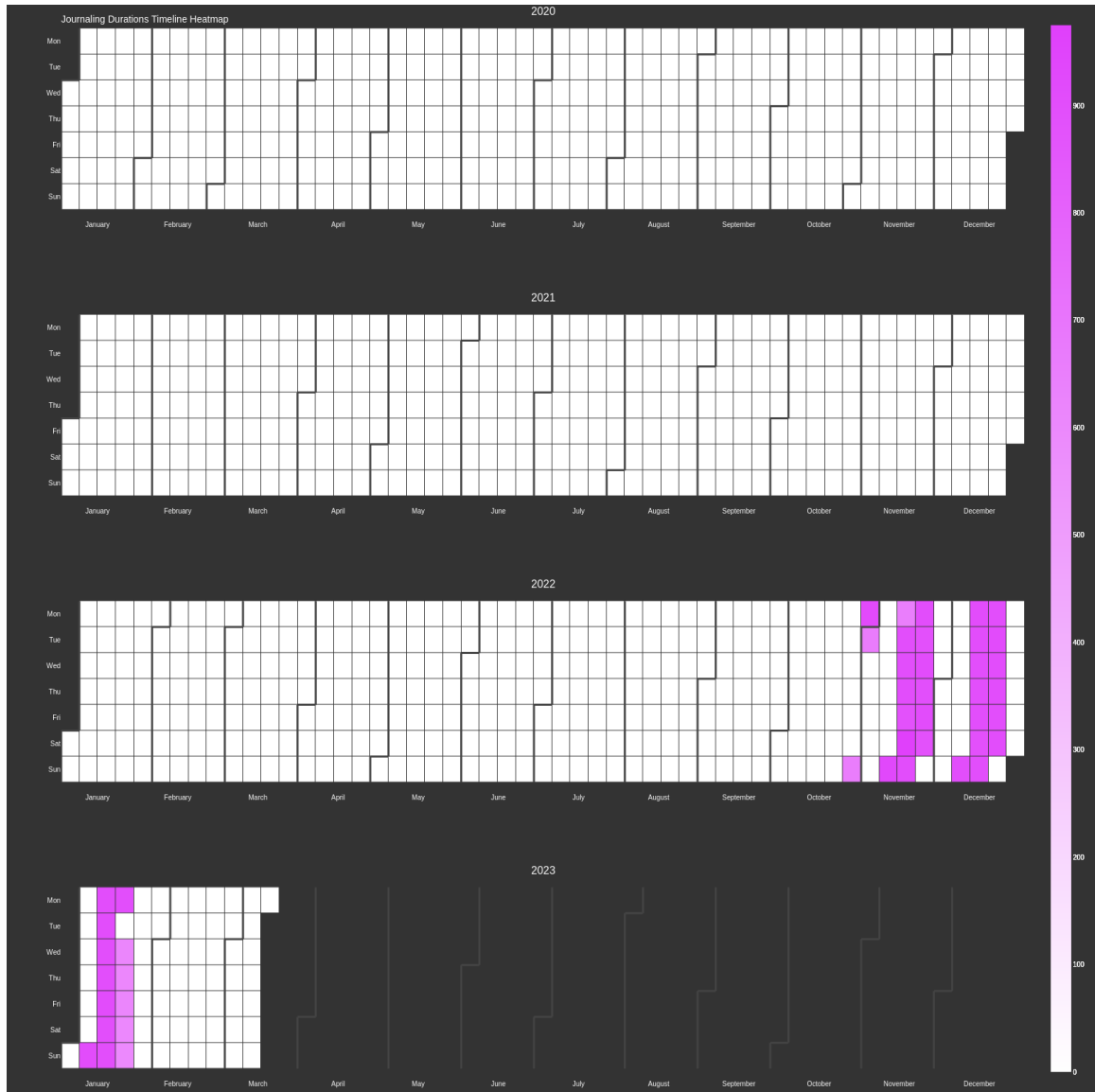


Figure 3.29: Journaling Durations Timeline Heatmap

Overall, these detailed and interactive visual explorations of how one spends their time can be used to gain great insight into one's life, and the corresponding descriptive statistics can help drive introspection even further. A future self-experimenter may find that they are spending too much time on things that they wish were not and less on things they wish they were, fomenting real change and improvement. The Time Tracking and Activity Data may be where some of the most interesting insights are gained, but the next datasource may have implications for one's health and fitness.

3.3.3 *Fitbit Data*

For the Fitbit Data source we will first cover how this data was created and how the data collection system was structured, descriptions of each of the different features recorded, any problems that may have affected this data source, and then we will continue on to how it was transformed for further use, including any visualizations, tables, and other statistical analyses as needed.

Within the Fitbit Data item in the Raw Data layer seen in Figure 3.3 we have a fairly simple data flow compared to the rest of the datasets, consisting of (1) simply wearing the Fitbit watch throughout each day, and (2) downloading the subject's data through a custom program which (3) creates a CSV file that can be joined with the rest of the data. To actually be able to access the subject's data though, they had to request and be granted a Fitbit API key. Beginning and following along with the tutorial on the Fitbit developer's document page (dev.fitbit.com/apps), the subject went through the process of obtaining an API key. Once that was complete, the data download was as simple as specifying the date range of data requested within our custom `fitbit_downloader.ipynb`, typically done in several smaller chunks that were then combined. The Fitbit data collection started for Experiment 1 with the subject wearing the watch throughout every day, typically only taking it off to charge for half an hour to an hour on most days throughout the 3 year period, even wearing it while Sleeping. Once the data was done being collected, queried from the API, and downloaded into a CSV, we were able to send it to the Data Carpentry layer.

The Fitbit Data Carpentry was done in much the same way as the Mood Data, resampling the data on the Date in case any dates inside the period downloaded were missing, such as if the patient was not wearing the Fitbit for entire days at a time. Any of those missing values were replaced with the mean value for that variable.

This dataset included fewer variables than the rest of the datasets so far, these being the per day Step Count, Number of Floors walked up, Resting Heart Rate, and the general descriptive statistics for Heart Rate for each day, which can be seen with Table 3.16. The descriptive statistic Heart Rate features in this data set, all except for Resting Heart Rate, were engineered based on the intra-minute Heart Rate data captured by the Fitbit. This meant that for each day the minimum, maximum, and so on Heart rate were found from the 1440 minutes in each day. There were several other options available as additional data points through the Fitbit API that the subject's watch was recording, such as the subject's sleep beginning and end times, similar to the Sleeping activity from the Time Tracking and Activity Data, but these were deemed irrelevant, unnecessary, or already accounted for in the other datasets.

	Fitbit Item	Description
1	Step Count	The steps walked per day
2	Resting Heart Rate	The resting heart rate for the subject per day
3	Number of Floors	The number of floors walked up per day
4	Heart Rate Count	The number of heart rate's captured per minute in the day
5	Heart Rate Mean	The mean heart captured rate per day
6	Heart Rate STD	The standard deviation heart rate captured per day
7	Heart Rate Minimum	The minimum heart rate captured per day
8	Heart Rate 25%	The 25% quartile heart rate captured per day
9	Heart Rate Median	The median heart rate captured per day
10	Heart Rate 75%	The 75% quartile heart rate captured per day
11	Heart Rate Maximum	The maximum heart rate captured per day

Table 3.16: Fitbit Description

There were of course the usual outliers, and missing or incorrect data. As the Fitbit had to be worn for any data collection to occur at all, there were hours or days with missing data, including a week-long period when the subject had to replace a broken watch. The Fitbit was not perfect even if being worn correctly though, there were times when the device would erroneously record steps when the subject was doing other activities like Drumming or going to the Gym. There were several outliers as well, such as one day in 2021 when the subject spent most of the day walking with friends, greatly inflating the maximum Step Count and Number of Floors compared to other days, these were not discarded however.

With the Carpentry Layer complete and the Fitbit Data ready to be merged with the other datasets, we will show a snippet and basic statistical description of this dataset with Table 3.17 and Table 3.18 and then a few of the more interesting trends. As one of the main selling points of the Fitbit is as a pedometer, we will look at Steps Count first with Figure 3.30, Figure 3.31, and Figure 3.32 giving a good overview of this feature.

Date	Step Count	Resting Heart Rate	Number of Floors	Heart Rate Count	Heart Rate Mean	Heart Rate STD	Heart Rate Minimum	Heart Rate 25%	Heart Rate Median	Heart Rate 75%	Heart Rate Maximum
2020-01-01	1041	56	0	1349	61.69	7.50	46	57	60	65	98
2020-01-02	321	58	0	1421	61.50	7.57	39	56	61	66	102
2020-01-03	1659	57	1	1298	63.01	10.60	48	54	62	70	135
2020-01-04	8407	58	8	1390	66.99	10.91	53	60	64	71	126
2020-01-05	4520	57	2	1407	64.26	9.23	51	58	62	68	118
...	...	...	...	...	...	...	...	...	...	...	...
2023-03-06	5008	54	3	1414	62.90	13.04	45	52	58	74	115
2023-03-07	4960	55	8	1419	71.59	9.81	50	66	70	75	125
2023-03-08	358	53	0	1407	57.00	10.92	43	49	52	63	103
2023-03-09	749	54	0	1440	62.08	8.75	48	56	60	67	104
2023-03-10	3203	53	1	1390	60.64	17.66	44	48	52	68	120

Table 3.17: Fitbit Data Table

	Step Count	Resting Heart Rate	Number of Floors	Heart Rate Count	Heart Rate Mean	Heart Rate STD	Heart Rate Minimum	Heart Rate 25%	Heart Rate Median	Heart Rate 75%	Heart Rate Maximum
count	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165	1165
mean	6102	55	4	1364.99	64.97	13.38	46.32	55.35	61.29	70.78	126.41
std	3480	2	9	194.20	4.44	3.16	3.35	3.57	4.82	6.59	17.02
min	0	49	0	0	51.42	4.71	33	47	51	53	78
25%	3720	53	1	1386	62.08	11.13	45	53	58	66	114
50%	5765	54	3	1401	64.52	13.31	46	55	61	70	125
75%	8513	56	5	1411	67.58	15.61	48	57	64	75	139
max	27999	63	270	1440	89.88	24.61	68	78	89	101	211

Table 3.18: Fitbit Statistics

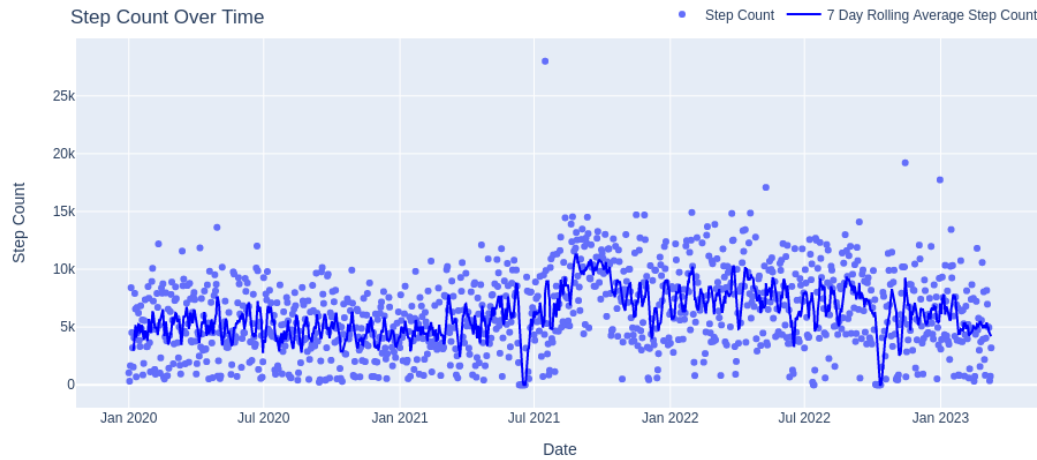


Figure 3.30: Step Count Over Time

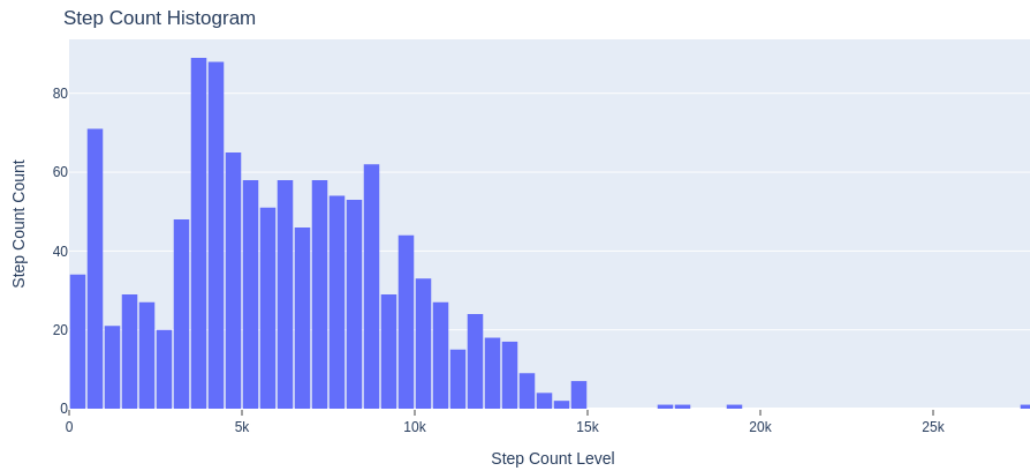


Figure 3.31: Step Count Histogram

Figure 3.30 gives us a look at how the subject's number of steps per day changed over time. We can see at least one major outlier, but there are also some interesting changes in the trends such as a bump in the average number of steps around August of 2021, when the subject started Walking to Class for their first semester of graduate school. We can see that this trend lessened somewhat but never returned back to its previous average, with much fewer near-zero Step Count days as time went on. Figure 3.31 shows us how the Step Count feature was distributed, with a minor bimodal section near zero, either because the Fitbit was not recording, was not being worn, or because the subject just was not actually moving very much that day. The subject had an average of about 6102

steps walked per day.

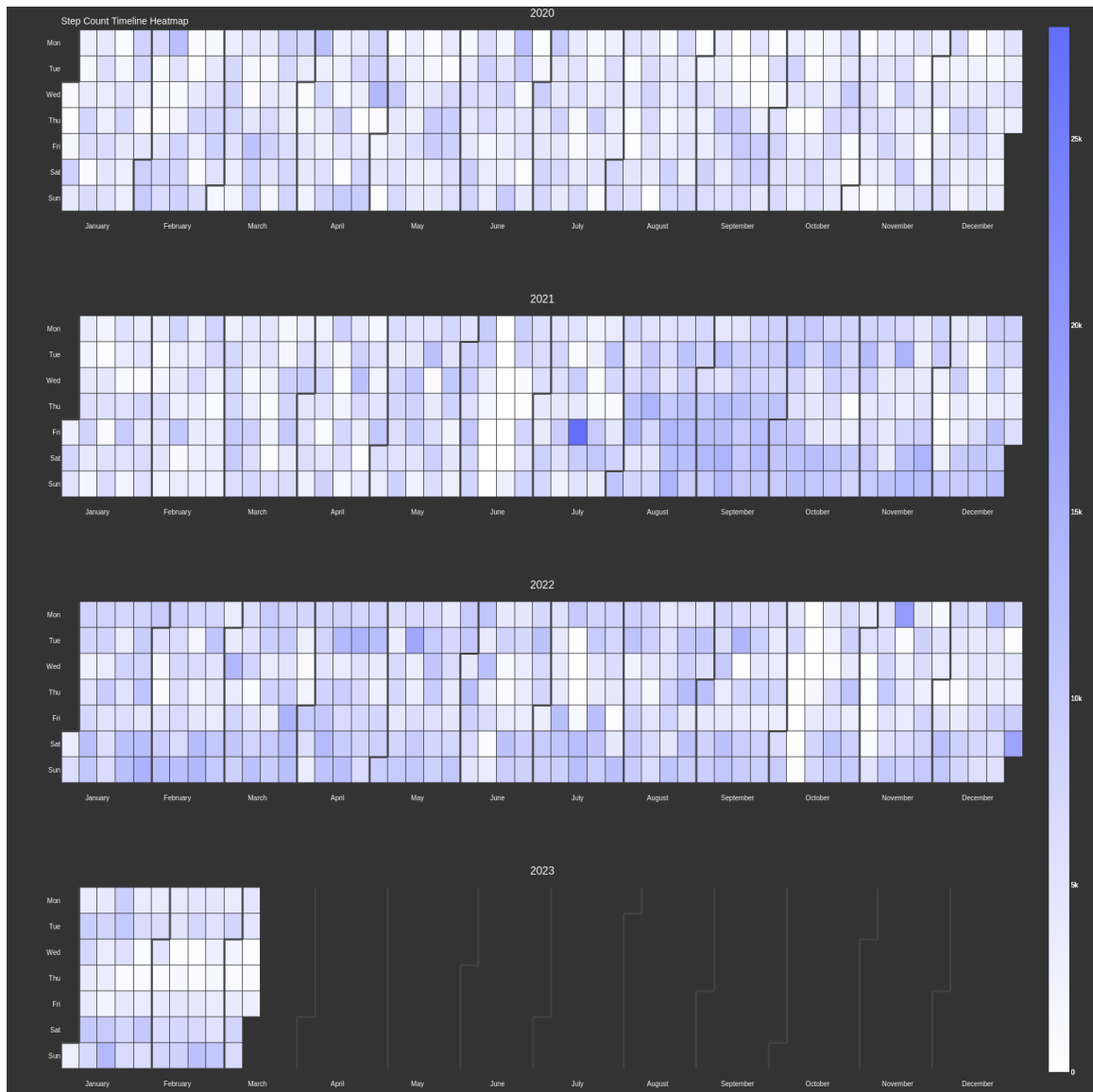


Figure 3.32: Step Count Timeline Heatmap

This gives us yet another view of the Step Count data, this time showing some hot and cold zones of increased or decreased activity like the differences in the week vs. the weekend in 2022, when the subject shifted their work schedule (Figure 3.22).

With Figure 3.33 we are shown the distribution for the intra-minute data for the subjects heart rate, with 1,706,400 per-minute heart rate values in this dataset. We will leave this section with Figure 3.34 and With Figure 3.35.

Overall, these visual explorations of one's fitness metrics can be used to spur oneself into more action. The self-experimenter may ask "Am I as active as I should be?" or "Am I doing everything I can for my heart health?" The Fitbit Data may offer insights to lifestyle changes that one can make,

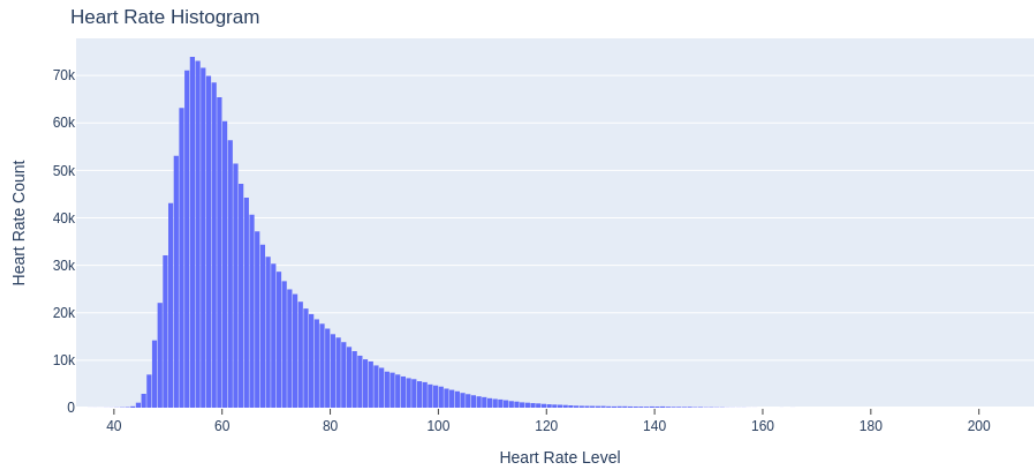


Figure 3.33: Heart Rate Histogram

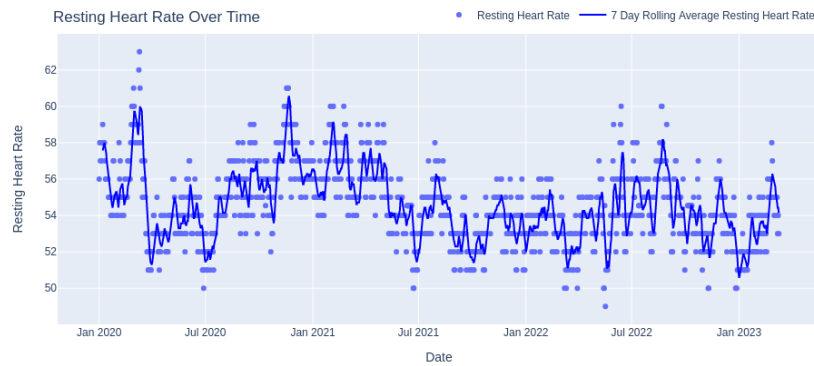


Figure 3.34: Resting Heart Rate Over Time

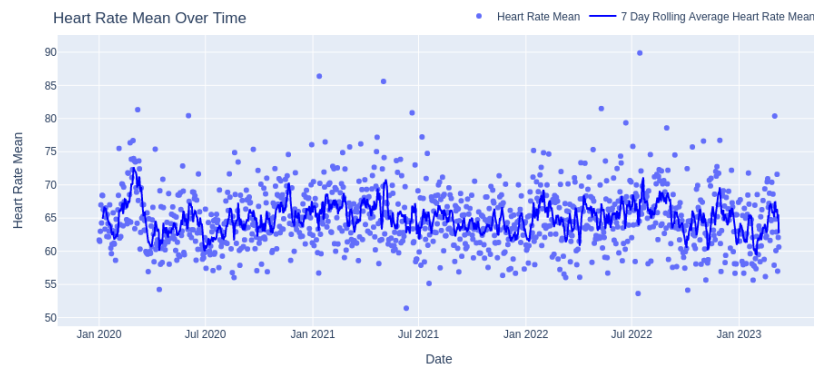


Figure 3.35: Heart Rate Mean Over Time.png

but the next datasource explores questions like the latter with several new health metrics.

3.3.4 Vital Signs Data

For the Vital Signs Data source (Figure 3.3) we will first cover how this data was created and collected, descriptions of each of the different features recorded, any problems that may have affected this data source, and then we will continue on to how it was transformed, including any visualizations, tables, and other statistical analyses as needed.

This data was collected using nearly the same process as the Mood Data with the addition of (1) the subject self-administering their own vital signs through a blood-pressure monitor and thermometer once per day, but after that the processes were identical, by inputting that data through (2) a Google Form, which (3) automatically uploaded to a Google Sheet, and was able to (4) be downloaded through our `GoogleSheetDownloader.py` as seen in Figure 3.36.

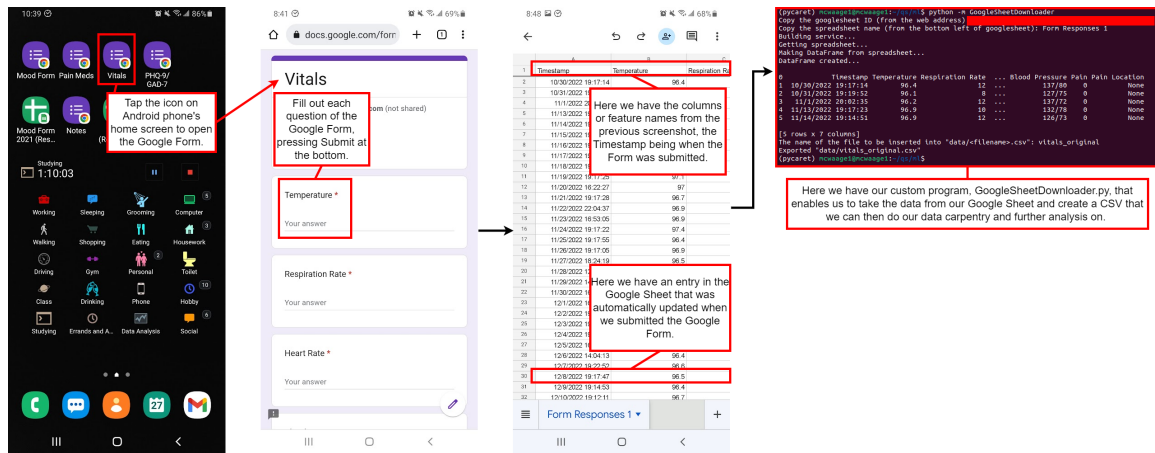


Figure 3.36: Vitals Data Flow

This data collection began with the start of Experiment 2 on 11/13/2022 and was recorded every day after that for each of the different phases and beyond (Figure 3.1). The collection times varied some but were generally completed at about 7 p.m. like the other Forms, after Journaling but before completing the Mood and PHQ-9/GAD-7 Form. After the data was finished being collected then the carpentry layer could begin. A description of each of the fields in the Vitals Form can be seen with Table 3.37. For the Vital Signs Data Carpentry we did some feature engineering and created a new variable, Mean Arterial Pressure, which is the weighted average of the Systolic and Diastolic blood pressures.

The Vitals Data Carpentry had the same steps as the Mood and Fitbit Data, resampling the data on the Date in case any dates inside the period requested were missing, such as if the subject

	Vitals Item	Description
1	Temperature	The temperature of the subject, recorded by mouth with a thermometer (°F)
2	Respiration Rate	The number of respirations counted within a minute
3	Vitals Heart Rate	The heart rate displayed on the blood pressure monitor
4	Blood Pressure Systolic	The subject's systolic blood pressure as taken by the blood pressure monitor
5	Blood Pressure Diastolic	The subject's diastolic blood pressure as taken by the blood pressure monitor
6	Mean Arterial Pressure	The weighted average of the systolic and diastolic blood pressures

Figure 3.37: Vitals Description

did not take their vitals for a day. Any of those missing values were replaced with the mean value for that variable. One thing to note here is that for the period of Experiment 2, the subject generally took their vitals at about 7 p.m., and because they had a variable Sleeping schedule, most of the time going into work at 7:30 p.m., on those work days they took their vital signs soon after waking up and on non-working days they may have been taken after being awake for most of the day.

Date	Temperature	Respiration Rate	Vitals Heart Rate	Blood Pressure Systolic	Blood Pressure Diastolic	Mean Arterial Pressure
2022-11-13	96.90	10.00	58.00	132.00	78.00	96.00
2022-11-14	96.90	12.00	76.00	126.00	73.00	90.67
2022-11-15	96.80	14.00	57.00	135.00	74.00	94.33
2022-11-16	97.10	14.00	63.00	141.00	69.00	93.00
2022-11-17	96.80	14.00	59.00	131.00	74.00	93.00
...	...	...	...	...	...	...
2023-03-08	96.60	12.00	74.00	130.00	76.00	94.00
2023-03-09	96.66	10.78	61.25	131.62	72.71	92.34
2023-03-10	96.66	10.78	61.25	131.62	72.71	92.34
2023-03-11	96.20	10.00	64.00	131.00	71.00	91.00
2023-03-12	96.80	10.00	59.00	136.00	75.00	95.33

Table 3.19: Vitals Data Table

Each of these statistics in Table 3.20 has the opportunity to give some insight for the self-experimenter, the blood pressure measures especially having great potential for their long term health and wellness. A normal or good blood pressure is usually around 120/80, meaning 120 Systolic and 80 Diastolic. As the subject's Systolic is averaging around 130, they may have an elevated risk of hypertension which is worth investigating further [43]. We will view a few of the more interesting trends with this dataset now, starting with the Systolic and Diastolic Blood Pressures. The Systolic and Diastolic blood pressures seem to be trending upwards, which is worth noting for the self-experimenter, but is there a reason behind this? Figure 3.39, Figure 3.40, and Figure 3.41 also seem to show a general trend upwards for each of their respective features as well.

	Temperature	Respiration Rate	Vitals Heart Rate	Blood Pressure Systolic	Blood Pressure Diastolic	Mean Arterial Pressure
count	120.00	120.00	120.00	120.00	120.00	120.00
mean	96.64	10.84	60.64	130.80	72.37	91.85
std	0.44	2.24	8.51	5.37	4.22	3.75
min	95.80	7.00	46.00	112.00	60.00	82.00
25%	96.30	9.00	55.00	128.00	70.00	89.67
50%	96.60	10.00	59.00	131.00	72.71	92.33
75%	96.90	12.00	63.25	134.00	75.00	94.33
max	98.00	18.00	88.00	143.00	83.00	99.33

Table 3.20: Vitals Statistics



Figure 3.38: Blood Pressure Systolic and Blood Pressure Diastolic Over Time

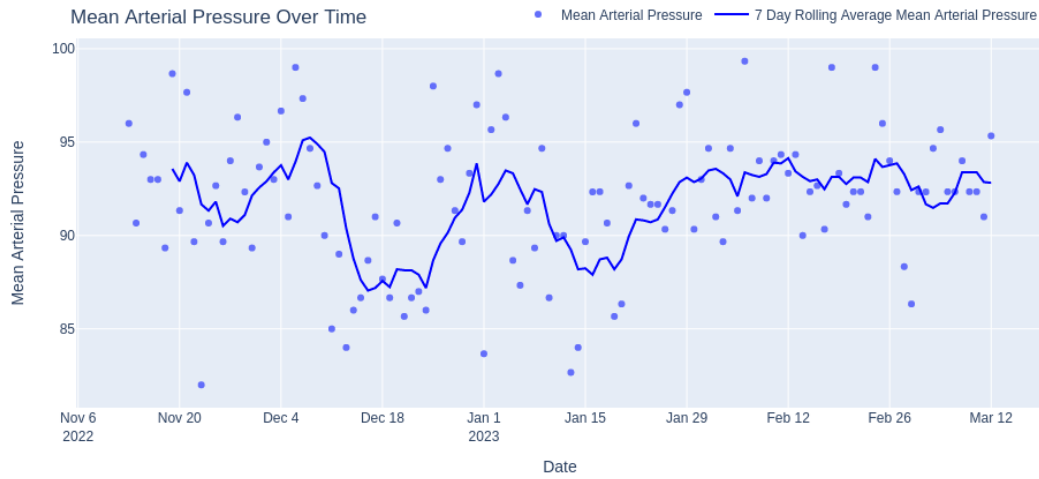


Figure 3.39: Mean Arterial Pressure Over Time

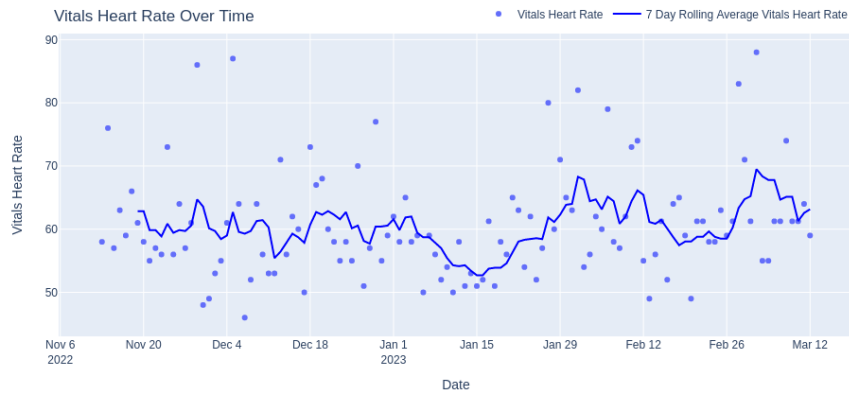


Figure 3.40: Vitals Heart Rate Over Time

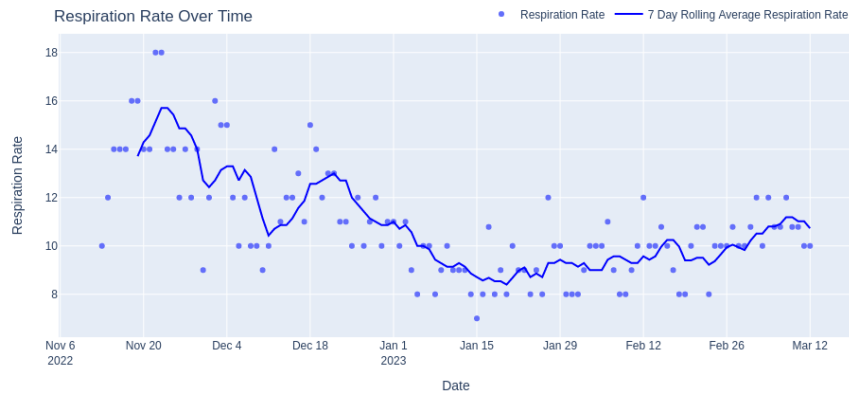


Figure 3.41: Respiration Rate Over Time

The Vital Signs Data seems to offer a lot in the way of future analysis and probes the question of if the subject is doing everything they can to be in the best health possible. There is one final dataset to briefly explore before we move on to the statistical analyses and machine learning sections.

3.3.5 *TempoDisc Data*

In this section we will briefly describe the TempoDisc Data source, how the data was gathered, descriptions of each of the different features recorded, and the problems that affected this dataset, including a few tables, and other statistical analyses as needed.

The TempoDisc [44] is a small wearable device that began being used with the start of Experiment 2 (Figure 3.2). With it, the subject could collect the Ambient Temperature, Humidity, Dew Point, and Pressure around them as it monitored each of these metrics per minute. The device was fitted on the subject’s keychain which they took with them everywhere. After a few days of data collection exporting the data through their smartphone was required for the data on the device to not be overwritten. Once the data was exported from the subject’s smartphone in CSV format, it was collected and combined in the `tempodisc_carpentry.ipynb` program. A description, snippet, and statistical description of the data can be seen with Table 3.21, Table 3.22, Table 3.23.

	TempoDisc Item	Description
1	Ambient Temperature	The ambient temperature (°F)
2	Ambient Humidity	The ambient relative humidity (%RH)
3	Ambient Dewpoint	The ambient dewpoint (°F)
4	Ambient Pressure	The ambient pressure (hPA)

Table 3.21: TempoDisc Description

Unfortunately, in the last weeks of Experiment 2 the device started malfunctioning and would not sync with the subject’s smartphone, making it impossible for data collection to continue. Ultimately because of this, the more inconsistent uploading and loss of data, and the data collection being somewhat erratic, probably due to being inside the subject’s pocket for most of the day, this dataset was ultimately discarded. A visual inspection of the per-minute Ambient Temperature can be seen with Figure 3.42.

3.4 Data Carpentry

From our Overall Data Flowchart (Figure 3.3), we can see that the individual and combined data carpentry items were a significant portion of the work done for this project. We already discussed

Date	Ambient Temperature	Ambient Humidity	Ambient Dewpoint	Ambient Pressure
2022-11-01 01:44:00+00:00	70	41.5	45.7	992.3
2022-11-01 01:45:00+00:00	70	41.6	45.8	992.2
2022-11-01 01:46:00+00:00	70	41.5	45.7	992.3
2022-11-01 01:47:00+00:00	69.8	41.5	45.5	992.2
2022-11-01 01:48:00+00:00	69.8	41.6	45.6	992.3
...	...	...	...	...
2023-01-07 16:30:00+00:00	83.1	17	34.1	999.3
2023-01-07 16:31:00+00:00	83.3	17.4	34.9	999.4
2023-01-07 16:32:00+00:00	83.3	18.2	36	999.3
2023-01-07 16:33:00+00:00	83.5	18.5	36.6	999.3
2023-01-07 16:34:00+00:00	83.7	18.8	37.1	999.3

Table 3.22: TempoDisc Data Table

	Ambient Temperature	Ambient Humidity	Ambient Dewpoint	Ambient Pressure
count	87948	87948	87948	87948
mean	73.07	26.69	35.75	980.93
std	6.93	8.40	8.30	41.59
min	35.8	2.8	-15.8	757.7
25%	68.9	21.4	30.9	984.3
50%	70.2	26.3	35.3	993.2
75%	75.4	31.3	40.5	997.6
max	91.9	66.1	68.5	1011.9

Table 3.23: TempoDisc Statistics

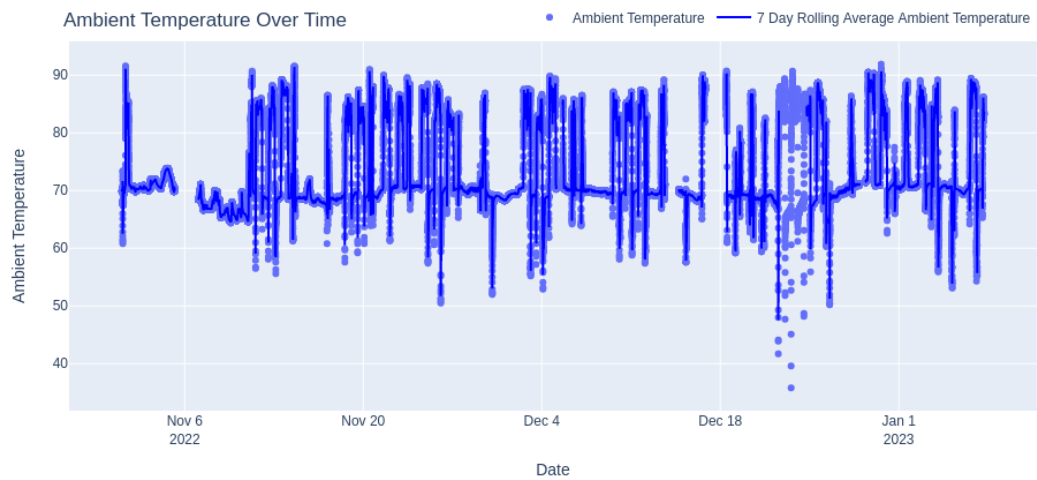


Figure 3.42: Ambient Temperature Over Time

some of the carpentry that went into each of the individual data sources, but we will examine that and the combined data carpentry in more detail now.

The Mood Form data was separated into the 2020 and Revised versions. The 2020 Mood Form data had 13 columns dropped from it and any rows that were completely blank were removed. It was then resampled on the Date columns, inserting any missing dates. The items in this dataset were then rescaled to the revised Mood Form data scales with the Mood changing from 5 to 0, 6 to 1, 7 to 2, 8 to 3, and 9 or 10 to 4. The majority of the rest of the items were rescaled from 0 to 0, 1 and 2 to 1, and from 3 and 4 to 2. Two of the items, Irritable and Angry were converted into a single column Irritable/Angry. The Revised Mood Form had any completely blank rows removed and was resampled on the Date columns for any missing dates as well. Any missing values were imputed with the median value for that item except for the Weight item which was replaced with the previous day's value. There were no additional features created solely from this dataset. A total of 26 columns including the Date were in the Mood Form Data.

For the Time and Activity Data, the aTimeLogger data was transformed into a wider format indexed on the Date, with the durations in seconds of each activity per day, the counts of the instances of each activity per day, and the 7 day rolling average durations in seconds of each activity per day, for a total of 121 columns including the Date.

The Fitbit Data included the Step Counts per day, Resting Heart Rate per day, Number of Floors per day, and the general statistical descriptions of the subject's heart rate rate for each day like the minimum and maximum, for a total of 12 columns in this dataset including the Date. Any missing values for the Step Count and Resting Heart rate were replaced with the mean values for those columns.

The Vitals Data included the subject's Temperature, Respiration Rate, Heart Rate, Systolic and Diastolic Blood Pressure, and Mean Arterial Pressure which was created from the Systolic Blood Pressure multiplied by one-third plus the Diastolic Blood Pressure multiplied by two-thirds for 7 columns in this dataset including the Date.

Once all of the individual data carpentry was done then they could be sent to the combined carpentry item and joined or merged together on the shared Date column. From there some new features were created like the previous Date's lagged or shifted value for each feature and some time series features like the Month, Day of the Month, Weekday, and Week of the Year.

Overall we were left with 2 final datasets to work with, the data for Experiment 1 and Experiment 2 respectively. After all carpentry was complete, the dataset for Experiment 1 had 1165 rows for each of the days in the experiment, from 01/01/2020 to 03/10/2023 and 317 columns for each of the

Mood Form, Time and Activity, and Fitbit datasets. The dataset for Experiment 2 had a subset of the rows from Experiment 1, from 11/13/22 to 02/04/2023, for 84 rows or days, but included all of the columns and even expanded those with the Vitals Data as well, bringing the total to 330. With these 2 datasets we were able to answer all of our questions.

Now we will take a look at some of the relationships between the many variables visible with these datasets with Figure 3.43, Figure 3.44, and Figure 3.45. While one can examine each of the features in the entire dataset against one another in this manner, a correlation heatmap reveals the bigger picture more succinctly such as with Figure 3.46 and Figure 3.47.

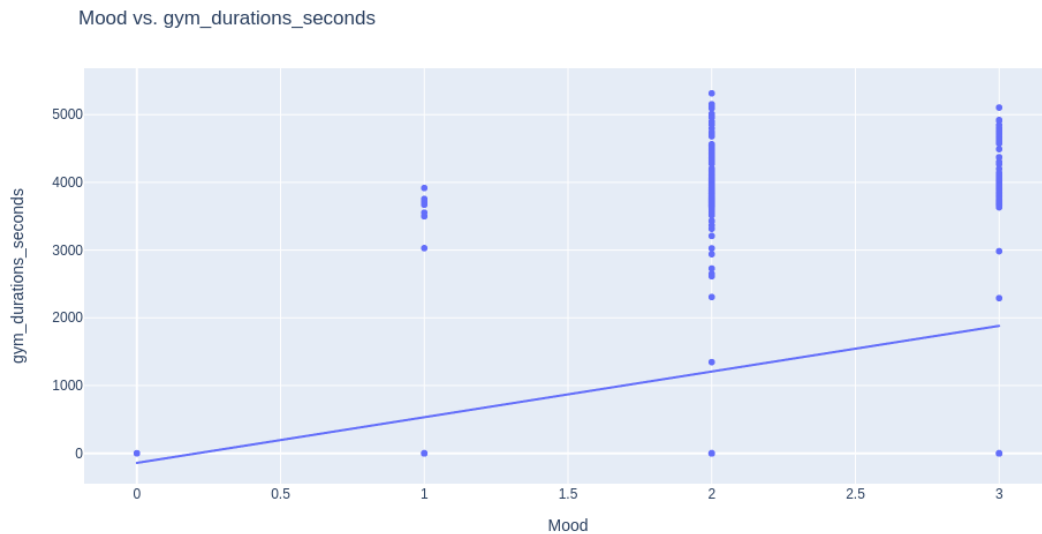


Figure 3.43: Mood vs. Gym Durations Seconds

This shows that as the subject spent more time in the Gym, their Mood improved, though the answer to whether or not one caused the other will be explored later.

With all of the data collected and merged together, the various problems with each dataset accounted for, and many visual and statistical explorations of the wide dataset, some of the more advanced statistical analyses and the machine learning pipeline can begin, and our questions may be answered.

3.5 Statistical Analyses

There were several different statistical analyses that were performed after the data was put into their final forms. With these tests and methods we were able to answer a couple of our research questions, namely:

- (1) For a single-subject self-experimenter, does gratitude journaling, compared to no intervention,



Figure 3.44: Step Count vs. Gym Durations Seconds
 This shows us the fairly obvious connection between steps and time spent in the Gym, as one increases, the other does as well.

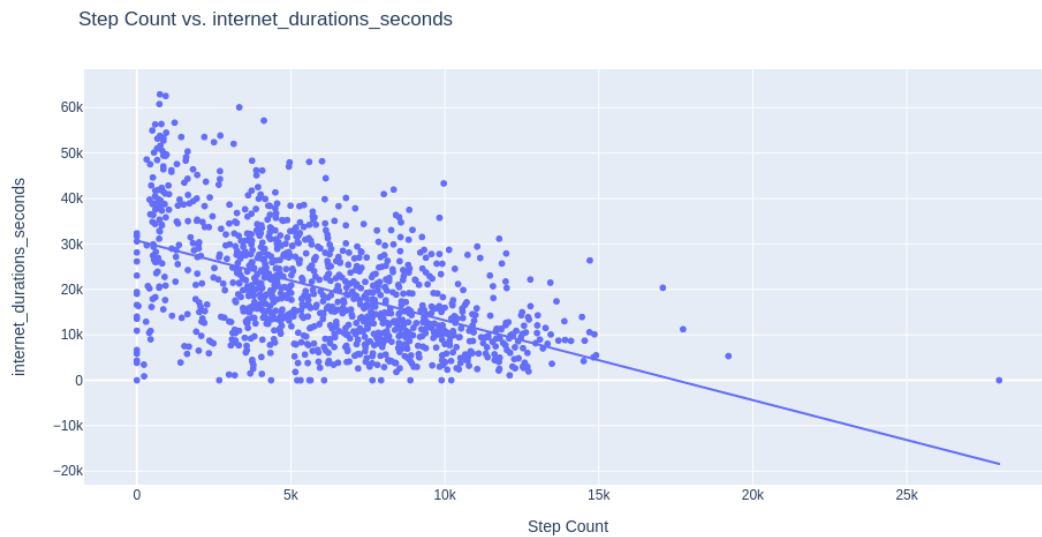


Figure 3.45: Step Count vs. Internet Durations Seconds
 This shows us the opposite, though still obvious conclusion, that as the subject spent more time on the computer, they spent less time walking.

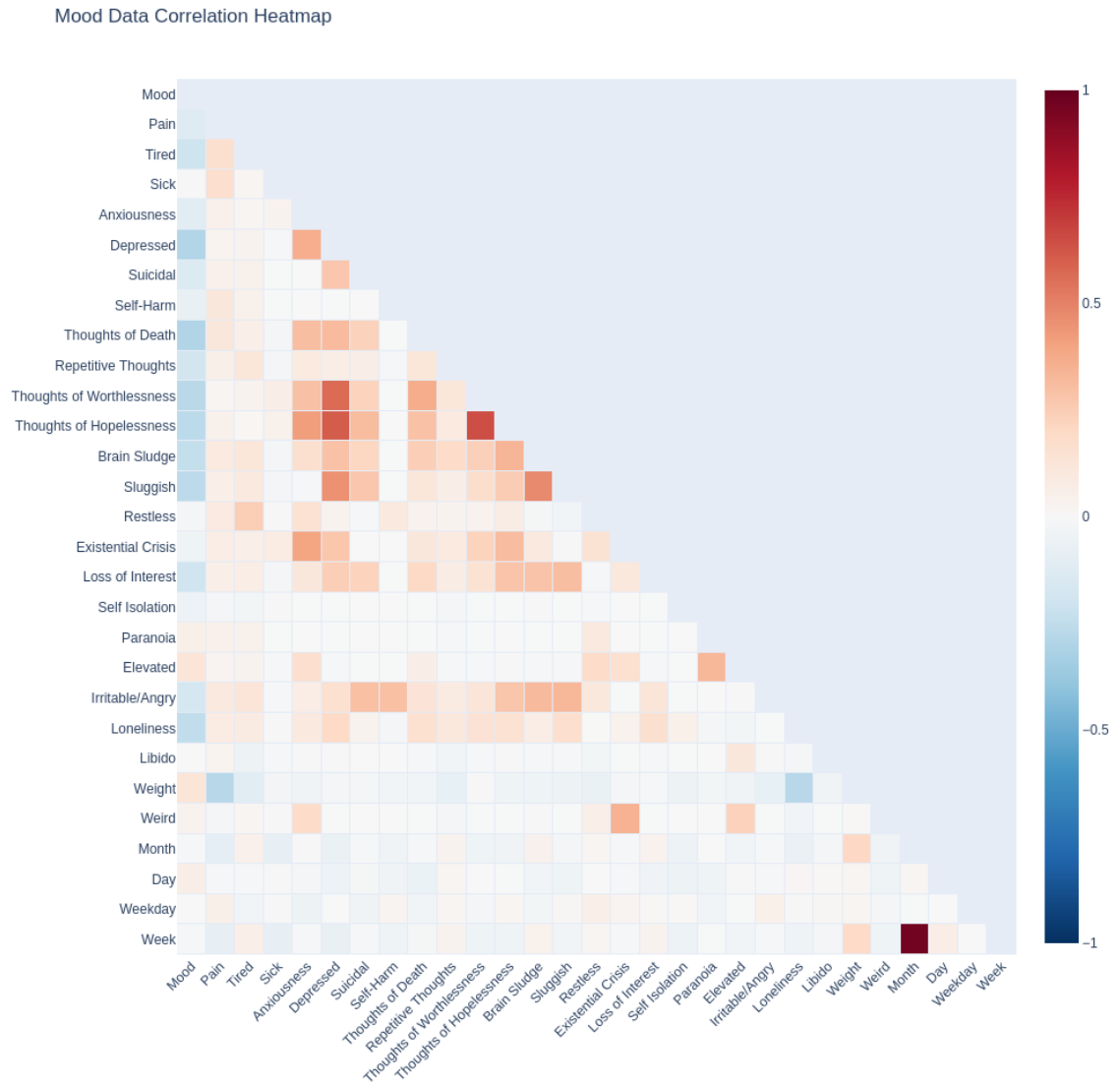


Figure 3.46: Mood Data Correlation Heatmap

This shows us the subject's Mood Data variables correlated against one another, with some not so obvious relationships such as Weight being related to an improved Mood and a decreased Pain level, the latent variable in this case likely being time spent in the Gym, though we can not say which causes the other with just this information.

Mood vs Time 7 Day Rolling Average Correlation Heatmap

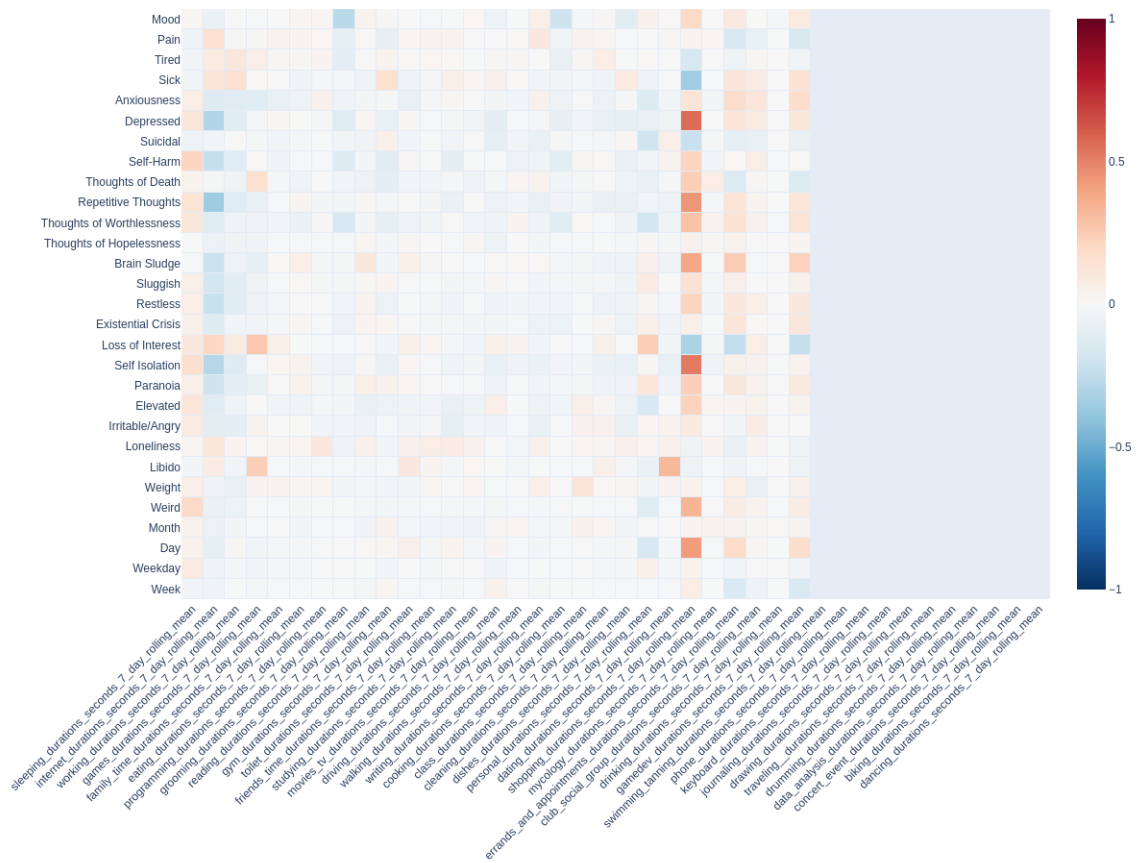


Figure 3.47: Mood vs Time 7 Day Rolling Average Correlation Heatmap

This reveals some of the more interesting connections between the datasets such as how there is a relationship between the subject's time spent on the Internet and most of the Mood Data items, like its positive relationship with Pain, meaning as one increases so does the other.

improve their mood? and,

- (2) Does the mere act of self-experimenting affect the subject's mood?

Which leaves the other questions to be answered later. So far we have seen how visual analysis can be used effectively for the subject and other self-experimenters to gain insights about their mental health, time usage, fitness and physical health, but more advanced statistics were necessary to answer these questions.

Power analysis was the first statistical test to be conducted. This measures whether we were able to conclude that the other tests performed have actually answered our questions, or if they were more likely due to random chance. We performed this test on the dataset for Experiment 2 to make sure that we would have enough data before undertaking the experiment. We used the `power_rm_anova` function through the `pingouin` Python package to calculate the power needed for the experiment. We used this test because we would go on to use Analysis of Variance Repeated Measures (ANOVARM) and Analysis of Covariance (ANCOVA) tests which can be used for within-subject repeated measures such as this. The power test required 4 of 5 measures to discern the fifth, in this case the fifth being the power. We were interested in reaching the typical power of 0.80 and found a set of parameters that fulfilled that goal. These parameters included the typical alpha of 0.05, eta_squared of 0.1, an epsilon of 1 because of there being only 2 phases being repeated, and then because we wanted to complete the experiment in a reasonable amount of time we found that 3 repeated measures, or phases, with 28 samples, or days, per measure was enough to reach that threshold. With these parameters we found that Experiment 2 had a power of over 0.97. With the power analysis complete we could move on to the ANOVARM, measuring whether there was a significant difference between the different phases of the experiment for the subject's Mood.

The ANOVARM compared the separate means of the subject's Mood for the 3 BA crossover phases in Experiment 2 (Figure 3.2). In this experiment the subject spent 10-15 minutes per day gratitude journaling, asking the question of if this activity had any affect on their Mood. According to our results (Table 3.26) with the p-value below our alpha's threshold of 0.05, we can say that we reject the null hypothesis, answering this question and finding that yes; spending time gratitude journaling, compared to no intervention, does improve the subject's Mood. Some things to keep in mind with this result however is that as this is a self-study, it is possible there is bias involved. The subject may have wanted the outcome and this may have skewed their Mood, affecting the results. In other types of studies a double blind may have been put in place which would make biases like this impossible, so that the researcher could not affect the results in this way, but as the subject and researcher are one and the same this was not possible with this type of experiment. In addition to

Day of Phase	Phase 1		Phase 2		Phase 3	
	Mood B	Mood A	Mood B	Mood A	Mood B	Mood A
1	3	2	2	2	2	2
2	3	2	2	2	2	2
3	3	2	2	2	2	1
4	3	2	2	2	2	2
5	2	2	2	2	2	2
6	2	2	2	2	2	2
7	2	2	3	2	2	2
8	3	3	3	2	2	2
9	3	3	3	2	2	2
10	3	2	2	3	2	2
11	3	2	2	2	2	2
12	3	2	2	2	2	2
13	3	2	2	2	2	2
14	3	2	2	2	3	2

Table 3.24: ANOVARM Data Table

	Phase 1		Phase 2		Phase 3	
	Mood B	Mood A	Mood B	Mood A	Mood B	Mood A
count	14	14	14	14	14	14
mean	2.79	2.14	2.21	2.07	2.07	1.93
std	0.43	0.36	0.43	0.27	0.27	0.27
min	2	2	2	2	2	1
25%	3	2	2	2	2	2
50%	3	2	2	2	2	2
75%	3	2	2	2	2	2
max	3	3	3	3	3	2

Table 3.25: ANOVARM Statistics

Source	SS	DF	MS	F	p-unc	ng2	eps
Within	1.58	2	0.79	14.529	0.00006	0.39	0.95
Error	1.42	26	0.05				

Table 3.26: ANOVARM Results

the ANOVARM we performed an ANCOVA test as well, with the subject's Mood as the dependent variable and the number of times they spent Journaling per day as a covariate we found the same results, that the number times the subject spent Journaling was again significantly related to their Mood with an p-value of less than 0.05 (Table 3.27).

Source	SS	DF	F	p-unc	np2
Phases	0.2976	1	1.8194	0.1811	0.0220
journaling_counts	1.9286	1	11.7898	0.0009	0.1271
Residual	13.25	81			

Table 3.27: ANCOVA Results

With the first question answered, we can move on to the next. The second question asked whether the mere act of self-experimenting had an effect on the subject's mood. By separating the periods of before and after Experiment 2 started on 11/13/2022 (Figure 3.48), and comparing these 2 groups, we could uncover that answer. The Mann-Whitney U test is used to find whether two samples are equal or not, and because our samples were of different sizes, 1047 for pre-experiment and 118 after respectively, and as the Mood feature is ordinal, on a 0 to 4 Likert scale, we met its assumptions. With the assumptions met we could calculate the test, which we found were not significant with a p-value above the alpha of 0.05, meaning we failed to reject the null hypothesis that there was a difference between the 2 groups, as shown in Table 3.28. This means that the mere act of self-experimenting had no effect on the subject's Mood.

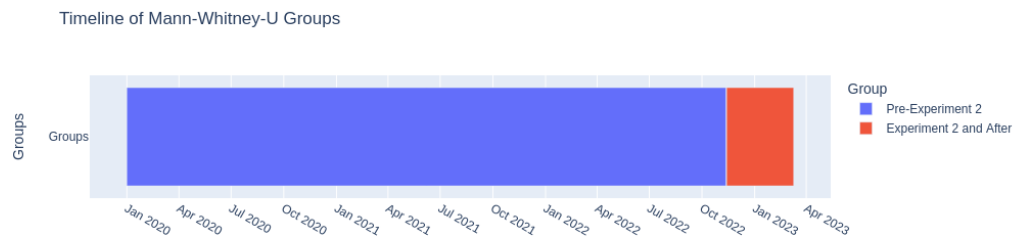


Figure 3.48: Timeline of Mann-Whitney U Groups

U-val	alternative	p-val	RBC	CLES
60850.5	two-sided	0.7038	0.0149	0.4925

Table 3.28: Mann-Whitney U Results

We have answered a couple of our questions now through statistical analysis, but traditional

methods like these are not the only way to discern if various features have an effect on another, in the next section we will explore how machine learning can answer our questions as well.

3.6 Machine Learning

In this section we will be exploring our third question which asks:

(3) Can a personalized machine learning model find the features that affect the subject's mood the most?

We explored this question with a Machine Learning (ML) pipeline to find the best model. Ten-fold cross-validation (CV) was used at each step to find the best model possible, the evaluation for what denoted the best model in our case being the one with the lowest mean absolute percentage error (MAPE) score. This meant the best model was the one that predicted the most correct amount of Mood values. One thing to note here is that as this is a personalized ML approach, what works for one person may not work for another. If another self-experimenter were to collect the same metrics with their own personal data, in the same manner as our subject, and input it into the same pipeline, they may end up with a very different model being chosen to predict their mood. However, an individualized approach for the self-experimenter to be able to find the features that most affected their mood, and in the future intervene on these top features, was the goal.

With Figure 3.49 we can see the flow for our ML pipeline. We used the `pycaret` [45] Python package to perform this pipeline, which is a wrapper around many other ML packages and tools like `scikit-learn`, `XGBoost`, and `LightGBM`. The workflow with this package was; (1) setup, (2) comparing a wide range of models which allowed us to (3) choose the best based on an ensemble blending method, then (4) tune the best model, (5) evaluate the model, and (6) interpret the model. With this package the workflow is reproducible for other self-experimenters to create their own personalized models to predict their own target metrics like their mood.

Before the setup stage was the preprocessing stage which we performed in each of the various data source's carpentry stages. In these carpentry stages we typically imputed any missing values for any of the features with the mean value for that feature, and also engineered new features such as our minimum and maximum Heart Rate for each day. Within the ML pipeline setup stage are where the processes of scaling, training, and evaluation were all performed, the scaling done for our data being the maximum absolute scaler, which scales each feature with the maximum absolute value found during training, converting the feature onto a scale from approximately -1 to 1.

The next stage involved comparing a wide array of models, with everything from linear regression

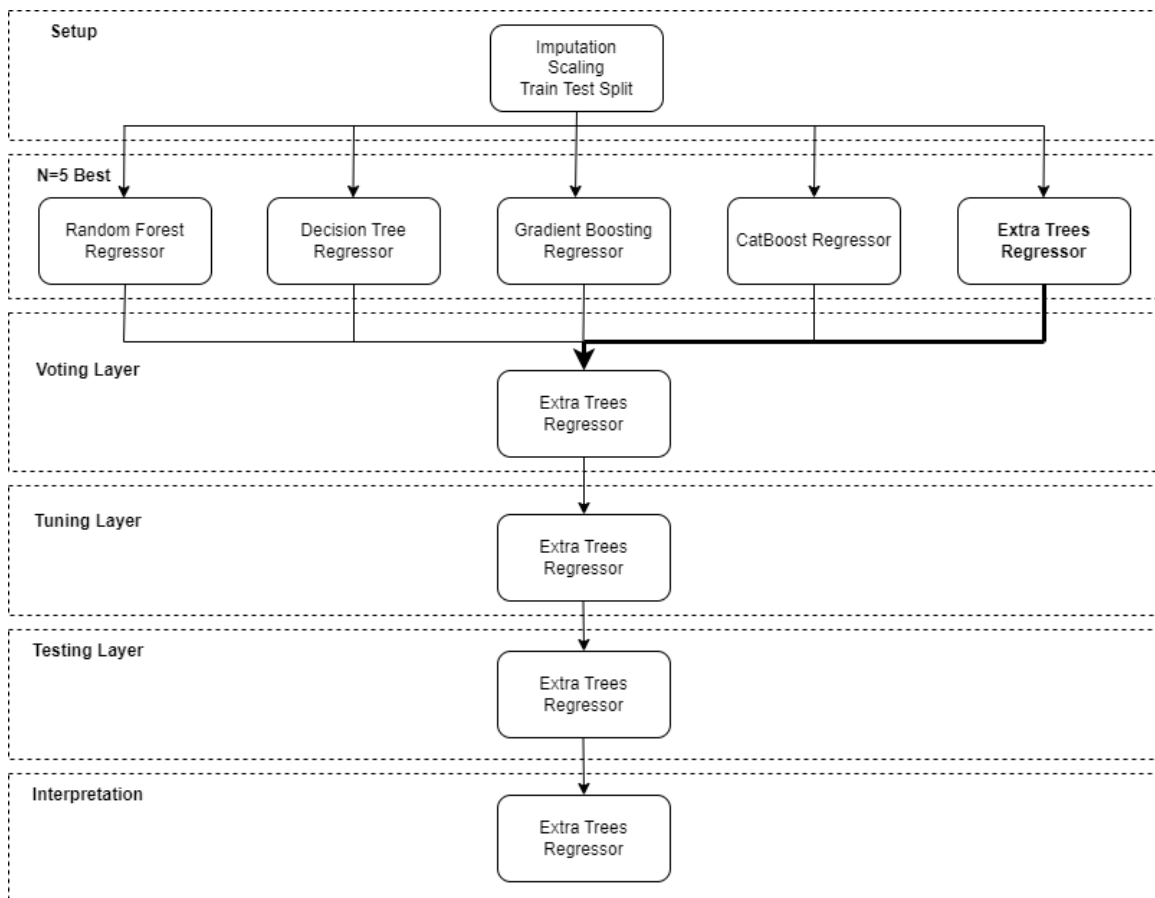


Figure 3.49: Machine Learning Flow

and random forests to extreme gradient boosting regressors, 24 in total. For each of the cross-validation (CV) folds during each of the stages, the values for each fold were randomly selected from the dwindling portion of the training set after each fold, leaving 10% of the values of the training set for each of the folds. The 5 best performing of these models, based on the lowest average MAPE score during these 10 CV folds, were carried over to the next stage.

In the voting stage a voting regressor was deployed to blend the best models. Blending in this manner is a composite ensemble learning method, the idea behind this method being to combine several different ML regressors and return their average predicted values. Performing this step can help a set of different ML models balance out each of their individual weaknesses [1]. Once the best model was chosen it was sent to the tuning layer.

In the tuning layer the chosen model's hyperparameters were iteratively changed so that the best combination of hyperparameters were chosen, maximizing the model's performance, in our case the model with the lowest average MAPE during the 10 fold CV. The output of this layer was a scoring grid with the cross-validated scores of the best model. The search algorithm can be random, Bayesian, or several others, but in our case was the scikit-learn random grid search which varied some of the hyperparameters like the number of estimators and leaves.

In the testing layer we evaluated our chosen model on our holdout test set created in the setup phase which gave us several different scoring metrics to base how well our model actually performed on unseen data. These metrics include our MAPE score among several others, Table 3.29 showing us these metrics.

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Extra Trees Regressor	0.2075	0.1161	0.3408	0.3578	0.1097	0.1018

Table 3.29: Machine Learning Results

With our somewhat satisfactory results, despite a low R2 which can be explained by the fact that we are dealing with a human subject, we can move on to the interpretation of the model, finding the features that most affected the subject's Mood.

3.6.1 *Personalized ML Feature Importance*

One integration with the `pycaret` package is another Python package, `shap` or SHapley Additive exPlanations (SHAP) [46]. This package gives us a different way of explaining the feature importance for any fitted ML model. SHAP gives us a method to explain how individual predictions are made

by a ML model, deconstructing a prediction into a sum of the contributions from each of the model's features. For each row in the data, the contribution from each feature towards the model's prediction will vary, depending on the values of the features for that particular row. SHAP finds this weighted average marginal contribution towards the model's score among all possible permutations. This essentially gives us how influential each feature is towards the prediction of the model and in which direction.

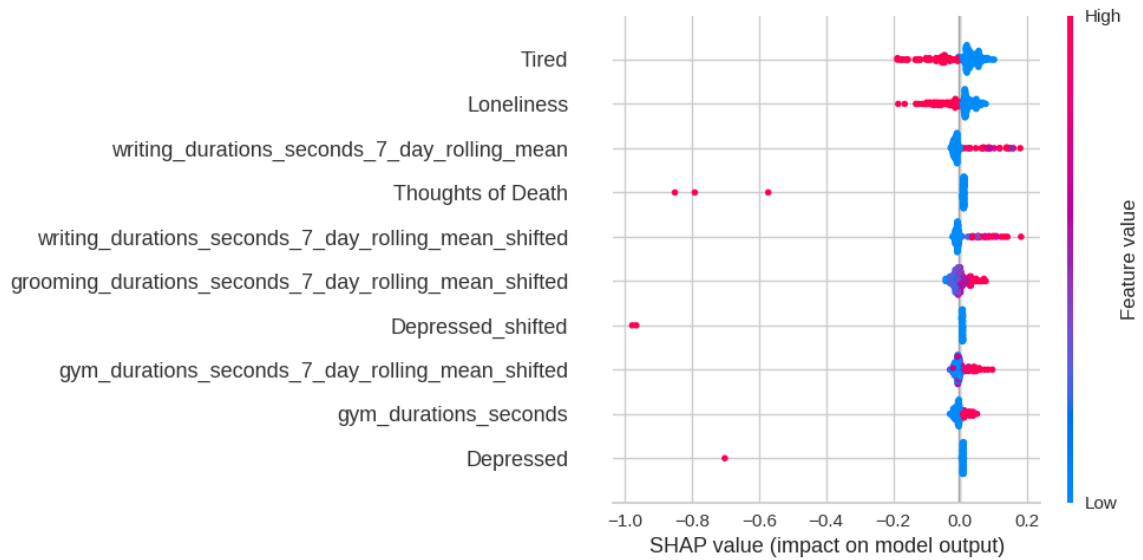


Figure 3.50: SHAP Feature Importance

This plot shows the features that contribute the most to the subject's Mood. On the Y-axis we have the features that were the most important and on the X-axis we have how much those features have contributed to a lower or higher Mood according to their SHAP value. We found that the feature Tired was the most important feature in predicting a Mood score, the higher a Tired value (to the right) was associated with a corresponding decrease in the feature value, which in this case is Mood. Essentially, as the subject's Tired increased, their Mood decreased. We get just the opposite relationship with writing_durations_seconds_7_day_rolling_mean which finds that as time spent writing increases so does Mood. Some features actually contribute more to predicting a lower mood, but were more sparse so were weighted lower than the features that have more occurrences to draw from.

Another way to view the relationship between Mood and the other features is with a scatterplot, in the first two cases, Figure 3.51 and Figure 3.52, some jitter was added since the original non-scaled values were on a scale from 0, 1, or 2 possible. Since SHAP values represent a feature's responsibility for a change in the model output, the plot below represents the change in predicted

Mood as Tired changes. Vertical variability at a single value of Tired represents interaction effects with other features. To help reveal those interactions we can color by another feature, which the package does automatically, finding the next most influential feature, in this case picking `internet_durations_seconds_7_day_rolling_shifted` as the second-order interaction effect. Another way to explain this is that if there is a trend upwards then it has a positive relationship with the target, Mood, and if it is trending downwards then it has a negative relationship.

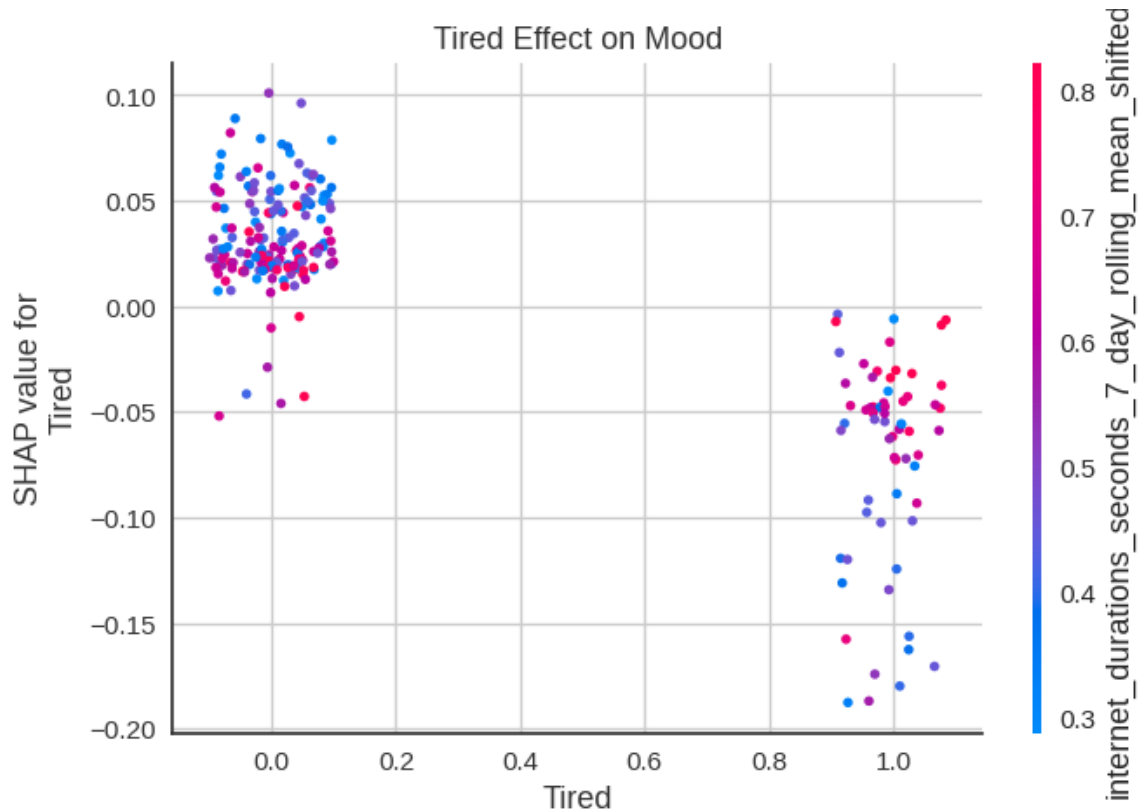


Figure 3.51: SHAP Correlation Tired Effect on Mood

We have found several features that had an effect on the subject's Mood. Examining these features further on by repeating this process, such as by targeting Tired on a subsequent ML pipeline, may give the subject the ability to iteratively find the features that affect their Mood the most, drilling down to what the essential cause for their own Great or Terrible Mood actually is. Overall we have found that statistical analyses and ML can both be used to answer our questions and find which activities, behaviors, or other tracked features affect the subject's Mood.

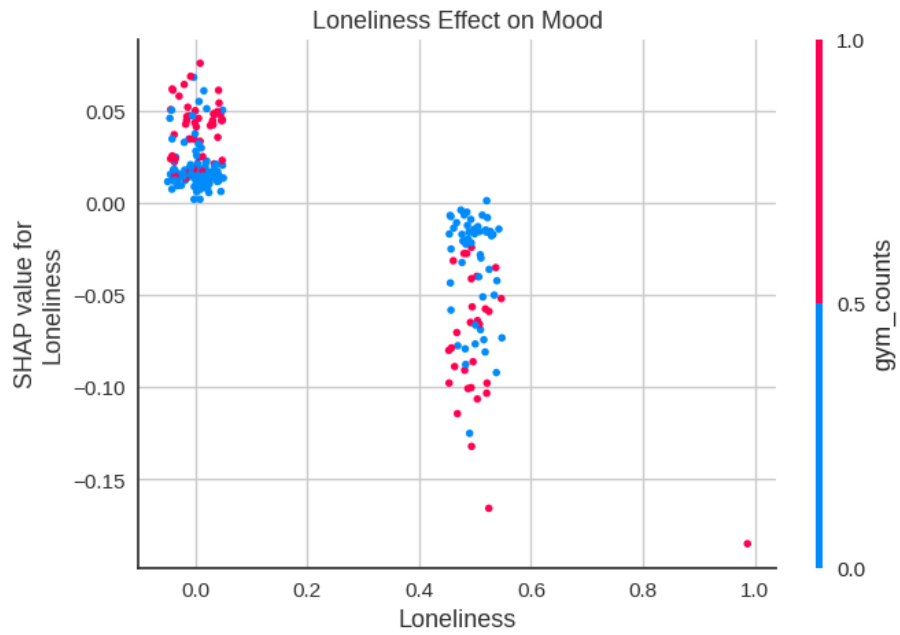


Figure 3.52: SHAP Correlation Loneliness Effect on Mood

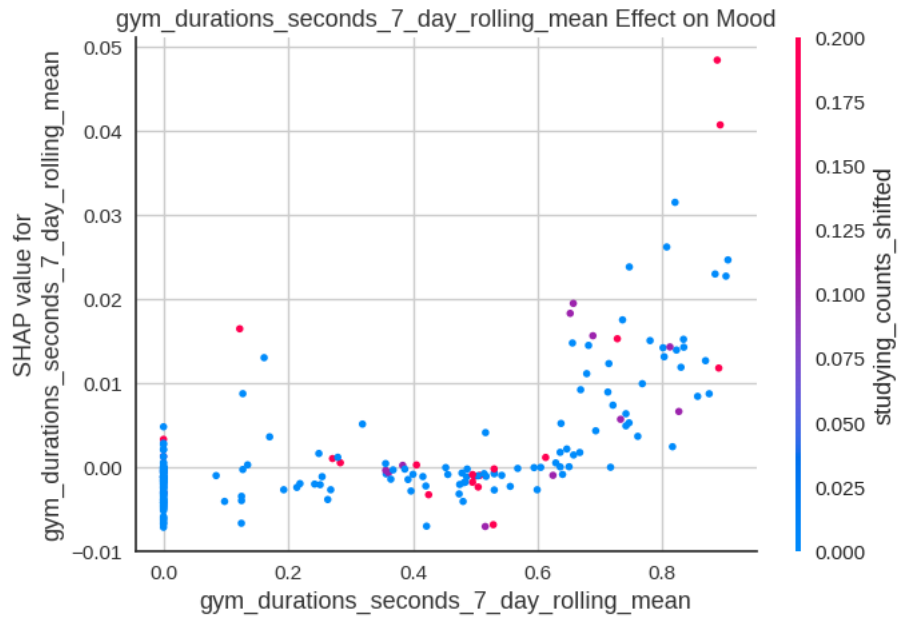


Figure 3.53: SHAP Correlation Gym Durations Seconds 7 Day Rolling Mean Effect on Mood
 Though the SHAP values are much lower for this chart, meaning that the feature has much less of an effect on the target of Mood, this figure shows that as time spent in the Gym increases, Mood does as well, though not entirely linearly.

Chapter 4

RESULTS

Many of our questions have been answered through the subject’s N-of-1 self-experiments. The single-subject performed two different longitudinal experiments, the first over a 3 year period, and a second of 84 days that overlapped with the first. The subject collected data from 5 different sources; Mood Data, Time Tracking and Activity Data, Fitbit Data, Vitals Data, and TempoDisc Data, though ultimately discarding the last. They used Google Forms, Sheets, and their API, a Fitbit wearable smartwatch and their API, several different Android smartphone apps, and many Python libraries and tools to extract and ingest, transform, combine, and collate these various datasets together to answer their questions.

For our first question: (1) For a single-subject self-experimenter, does gratitude journaling, compared to no intervention, improve their mood? We have found that yes, gratitude journaling does improve the subject’s Mood. The method for finding this result was through an ANOVARM and ANCOVA, wherein a p-value of less than 0.05 was reported for both of the tests, thus allowing us to reject the null hypothesis (Table 3.26, Table 3.27). The data used in this experiment was collected over three 28 day crossover phases for a total of 84 days in this experiment (Figure 3.2). During the B sequences of this BABABA study, alternating every 14 days, the subject spent about 15 minutes per day gratitude journaling, and subsequently not intervening at all on the A sequences.

For our second question: (2) Does the mere act of self-experimenting affect the subject’s mood? It was found that no, there was no difference in the subject’s mood for the periods of before and after the second experiment began. This result was found through a Mann-Whitney U test wherein the p-value was reported to be greater than the alpha of 0.05, meaning we fail to reject the null hypothesis (Table 3.28).

For our third question: (3) Can a personalized machine learning model find the features that affect the subject’s mood the most? We found that yes, we can find the most important features

affecting the subject's mood, Tired being the feature that affects the subject's Mood the greatest. This was achieved through a personalized ML pipeline to predict the subject's individual Mood rating (Table 3.29, Figure 3.50). For this 3 year longitudinal study a total of 317 features were collected on the subject across the domains of mental health, fitness and physical activity, and lifestyle and time-tracking, including things like sleep, work, and internet usage. These features were acquired from mood questionnaires, smartwatches, and smartphone apps, and once the study was complete all of this data was able to be collated together for each day of the entire 3 year period. Through this ML pipeline we were able to find the features that determined the subject's Mood, which may be used as targets for creating new models or as possible topics for interventions in the future, getting to the root causes affecting the subject's Mood.

In addition to these questions we also visually explored many relationships within and between the various datasets, finding many insights on the changing nature of the subject's Mood, Pain, Weight, and various time-tracked activities. We also explored these and other features and how they may have influenced each other with timelines, histograms, scatterplots, and heatmaps, showing trends and changes in schedules over time.

But what of our fourth question?: (4) Are any of these modes of self-experimenting more or less difficult to perform or are the outcomes more or less difficult to understand than the others? The different modes of self-experimenting, namely (A) conducting an AB phase design self-experiment versus (B) collecting data of any sort as an individual and running it through a ML pipeline, both come with their own unique challenges and ways to interpret the data. We will discuss this and other problems and insights found in the next section.

Chapter 5

DISCUSSION

If a self-experimenter wished to find out if a specific variable or feature were affecting them, such as a new medication and its effect on their depression, then a more traditional statistical test has a much more interpretable result. Even if the self-experimenter had little to no statistical knowledge they could conduct an experiment such as this, performing these statistical tests, and likely interpret what was or was not a significant result. One advantage of these tests is the relatively short period of turnaround for finding out whether what they were doing had any result or not. This cannot be said for the ML method, a much greater amount of data being needed to discern an acceptable result and predict with any accuracy. The advantage to the ML route though is in the "throwing science at the wall" method, collecting whatever possible and seeing what results can be found. One other advantage to this method is the ability to visually analyze any changes over time for the wide variety of data and make course corrections of a sort. Perhaps one target is not seeing noticeable changes, even through various interventions, but the added value from knowing what is affecting them steers them in directions they would not have known about before data collection. For our fourth question we find that the data collection is easiest for the ML method, but the interpretation is harder, while it is just the opposite for the statistical analyses. Conducting a regimented experiment in this manner, especially sticking to a fixed schedule during normal life while collecting data at nearly the same time every day for weeks or months, is cumbersome. The insights gained from regular, less intrusive or time-consuming data collection however, allows for more frequent introspection. There is a certain economy of scale to self-experimentation data collection as well, in that it gets easier with the more extensible methods, such as simply adding a new item to a questionnaire.

This study drew heavily from Daskalova et al. [2]–[4], [7] throughout the setup of the self-experimentation portion, and Shah et al. [1] for the ML portion, wherein the researchers created personalized ML models for each of the N-of-1 participants with depression in their study. They

comprehensively took into account various factors that contribute to depression and went even further to find the features that best predicted each individual’s depression. Much of this study then is a continuation of their work on personalized medicine and analytics, the goal being to create a more streamlined way for self-experimenters to actually perform these kinds of analyses and experiments.

In this study however we collected a much different set of features, the more unique among them being the Time Tracking and Activity Data. With this data source, the Mood Data, Fitbit Data, and Vital Signs Data, we found several different ways for insights to be gained. Any one of these sources is able to be taken up at will, wherever the self-experimenter’s interests may lie, each being relatively low barriers of entry for data collection and analysis individually. The testing and deeper analysis is a bigger hurdle. There were two different experiments, the first being the main 3 year longitudinal study. In this study the subject used all of the data except the Vitals Data, and tried to predict their Mood through an ML pipeline, further seeking the features that most affected it, ultimately finding that the feature Tired was the main cause. In experiment 2, which was only 84 days long total in comparison, the subject added the Vitals Data into data collection mix and spent alternating blocks of 14 days gratitude journaling and no intervention at all, finding through the ANOVARM and ANCOVA tests that it was significantly related to Mood. It was also found through a Mann-Whitney U test that experimenting had no effect on the subject’s Mood, the groups in this case being of before and after Experiment 2 began.

One limitation of this study, and one that affects N-of-1 studies in general, is the problem of having little to no external validity. While this is a significant problem in Random Control Trials (RCT), this is an acceptable loss here, as the goal is clinical significance for the individual, a more personalized treatment. One way to solve this problem though, is by pooling a group or cohort of many different N-of-1 trials, all experimenting in the same way and testing the same intervention as a meta-analysis [9], [11], [12]. One additional caveat that comes with longitudinal real-world studies of this type is how the activities of daily living have the potential to interrupt segments of data collection in numerous ways, such as devices malfunctioning or breaking, data uploads failing, or work, sleep, and other time commitments and obligations breaking the regular flow of data collection.

Despite this, there are wide implications for personalized analytics and healthcare through self-experimentation. One of the unique contributions that this study makes is as an avenue for future self-experimenters, even those with low or no coding or statistics knowledge, to be able to perform their own analyses or studies. As this project was designed to be target-agnostic, they may go on to experiment in any number of ways, finding or intervening on the features that affect them the most personally. Possible features or targets for data collection and self-experimentation are endless,

including; medications, blood sugar levels for diabetes management, symptom tracking for IBS and other chronic diseases or illnesses, food and nutrition data like calories or macronutrients, caffeine or water consumption, fitness and exercise, blood panels, gut and mouth bacteria panels, genetic data, finances, weather, seasons, or sunlight, computer and internet usage, or reading, writing, or music listening habits. If it can be quantified, it can be experimented on. Perhaps the most potential for benefit is with a more high risk or vulnerable patient population like children and infants, who can benefit from detailed data collection, but do not have the capabilities to do it themselves. A parent or caretaker could track the child’s weight, food consumption, temperature, or sleep times, and know when something is wrong. Tracking of this sort opens up the opportunity for one being informed that they are likely going to be sick or have a medical episode before even they know it, based on their daily average temperature, blood pressure, or respirations results.

One of the greatest benefits to the methods presented in this study is as a general life-reviewer and improver, making introspection and self-knowledge easier. Here, through a data-driven approach, we present methods for self-experimenters that seek personalized treatments of any quantifiable feature, to be able to easily collect data on themselves from a wide variety of sources, conduct experiments, test hypotheses, and develop models predicting those features. On the more technical side the work presented here gives people a set of tools for plotting and statistics related to longitudinal time-tracking data, making it easier for people to visually analyze and reveal the underlying statistical distributions of their task or activity based data.

There is also much opportunity for improvement with new integrations or standards. A standardized or default approach similar to other electronic health records that are HIPAA compliant should be one goal for a platform for self-experimenters. The StudyMe [8], StudyU [9], and Self-E [7] apps, being the latest at achieving that goal, streamlining the processes of data acquisition, and sharing the information from N-of-1 trials. What these lack though are the statistical or ML analyses being integrated within the app for the self-experimenter to be able to understand themselves. The best app would be one that would make it easier for self-experimenters to perform their own studies, tracking whatever they wanted, and discovering within the app through the statistical or ML methods which features were responsible for their target, while also being able to share that data with researchers or other self-experimenters so that their data can be pooled together so that new and more generalizable insights can be gained.

Conflict of Interest/Ethics Declaration

The author declares no competing interests.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Code Availability

Analytics code is available upon request from the author.

BIBLIOGRAPHY

- [1] R. V. Shah et al., “Personalized machine learning of depressed mood using wearables,” *Transl Psychiatry*, vol. 11, no. 1, Art. no. 1, Jun. 2021, doi: 10.1038/s41398-021-01445-0.
- [2] N. Daskalova et al., “SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Tokyo Japan: ACM, Oct. 2016, pp. 347–358. doi: 10.1145/2984511.2984534.
- [3] N. Daskalova et al., “SleepBandits: Guided Flexible Self-Experiments for Sleep,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–13. doi: 10.1145/3313831.3376584.
- [4] N. Daskalova, K. Desingh, A. Papoutsaki, D. Schulze, H. Sha, and J. Huang, “Lessons Learned from Two Cohorts of Personal Informatics Self- Experiments,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–22, Sep. 2017, doi: 10.1145/3130911.
- [5] F. J. Blaauw et al., “Let’s get Physiqua! – An intuitive and generic method to combine sensor technology with ecological momentary assessments,” *Journal of Biomedical Informatics*, vol. 63, pp. 141–149, Oct. 2016, doi: 10.1016/j.jbi.2016.08.001. [6] S. Taylor, A. Sano, C. Ferguson, A. Mohan, and R. W. Picard, “QuantifyMe: An Open-Source Automated Single-Case Experimental Design Platform,” *Sensors (Basel)*, vol. 18, no. 4, p. E1097, Apr. 2018, doi: 10.3390/s18041097.
- [7] N. Daskalova et al., “Self-E: Smartphone-Supported Guidance for Customizable Self - Experimentation,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–13. doi: 10.1145/3411764.3445100.
- [8] A. M. Zenner, E. Böttinger, and S. Konigorski, “StudyMe: A New Mobile App for User-Centric N-of-1 Trials,” p. 22. [9] S. Konigorski et al., “StudyU: A Platform for Designing and Conducting Innovative Digital N-of-1 Trials,” *J Med Internet Res*, vol. 24, no. 7, p. e35884, Jul. 2022, doi: 10.2196/35884.
- [10] K. W. Davidson, M. Silverstein, K. Cheung, R. A. Paluch, and L. H. Epstein, “Personalized (N-of-1) Trials: A Primer,” *JAMA Pediatr*, vol. 175, no. 4, pp. 404–409, Apr. 2021, doi: 10.1001/jamapediatrics.2020.5801.
- [11] K. Rl et al., “Design and Implementation of N-of-1 Trials: A User’s Guide,” p. 94, Feb. 2014.
- [12] J. Nikles and G. Mitchell, Eds., *The Essential Guide to N-of-1 Trials in Health*. Dordrecht: Springer Netherlands, 2015. doi: 10.1007/978-94-017-7200-6.
- [13] F. Naughton, “A starter kit for undertaking n-of-1 trials,” vol. 1, no. 5, p. 10.

- [14] C. J. Nikles, A. M. Clavarino, and C. B. Del Mar, “Using n-of-1 trials as a clinical tool to improve prescribing,” *Br J Gen Pract*, vol. 55, no. 512, pp. 175–180, Mar. 2005.
- [15] E. J. Daza, “Causal Analysis of Self-tracked Time Series Data Using a Counterfactual Framework for N-of-1 Trials*,” *Methods Inf Med*, vol. 57, no. Suppl 1, pp. e10–e21, Feb. 2018, doi: 10.3414/ME16-02-0044.
- [16] E. J. Daza, “Person as Population: A Longitudinal View of Single-Subject Causal Inference for Analyzing Self-Tracked Health Data.” *arXiv*, Jan. 22, 2019. Accessed: Sep. 05, 2022. [Online]. Available: <http://arxiv.org/abs/1901.03423>
- [17] “Do it yourself? When the researcher becomes the subject.” <https://www.science.org/content/article/do-it-yourself-when-researcher-becomes-subject> (accessed Dec. 12, 2022).
- [18] L. Vesnic-Alujevic, M. Breitegger, and . Guimarães Pereira, ““Do-It-Yourself” Healthcare? Quality of Health and Healthcare Through Wearable Sensors,” *Sci Eng Ethics*, vol. 24, no. 3, pp. 887–904, Jun. 2018, doi: 10.1007/s11948-016-9771-4.
- [19] “The Personal Analytics of My Life—Stephen Wolfram Writings.” <https://writings.stephenwolfram.com/2012/03/the-personal-analytics-of-my-life/> (accessed Dec. 12, 2022).
- [20] “Homepage,” Quantified Self. <https://quantifiedself.com/> (accessed Dec. 12, 2022).
- [21] “Home - Open Humans.” <https://www.openhumans.org/> (accessed Dec. 12, 2022).
- [22] M. Swan, “Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0,” *Journal of Sensor and Actuator Networks*, vol. 1, no. 3, Art. no. 3, Dec. 2012, doi: 10.3390/jsan1030217.
- [23] M. Swan, “The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery,” *Big Data*, vol. 1, no. 2, pp. 85–99, Jun. 2013, doi: 10.1089/big.2012.0002.
- [24] G. Appelboom, M. LoPresti, J.-Y. Reginster, E. Sander Connolly, and E. P. L. Dumont, “The quantified patient: a patient participatory culture,” *Current Medical Research and Opinion*, vol. 30, no. 12, pp. 2585–2587, Dec. 2014, doi: 10.1185/03007995.2014.954032.
- [25] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz, “Understanding quantified-selfers’ practices in collecting and exploring personal data,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, in CHI ’14. New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 1143–1152. doi: 10.1145/2556288.2557372.
- [26] J. N. Gilmore, “Everywear: The quantified self and wearable fitness technologies,” *New Media & Society*, vol. 18, no. 11, pp. 2524–2539, Dec. 2016, doi: 10.1177/1461444815588768.

- [27] A. D. Williams, “Harnessing the quantified self movement for optimal mental health and wellbeing,” in Proceedings of the first Workshop on Lifelogging Tools and Applications, in LTA ’16. New York, NY, USA: Association for Computing Machinery, Oct. 2016, p. 37. doi: 10.1145/2983576.2983585.
- [28] P. Abend and M. Fuchs, “Introduction. The Quantified Self and Statistical Bodies,” undefined, 2016, Accessed: Sep. 05, 2022. [Online]. Available: <https://www.semanticscholar.org/paper/The-quantified-self-archive%3A-documenting-lives-data-Trace-Zhang/9a1c96a4af09a5d34e4a9a209ef56242482d9277>
- [29] T. Sharon, “Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare,” *Philos. Technol.*, vol. 30, no. 1, pp. 93–121, Mar. 2017, doi: 10.1007/s13347-016-0215-5.
- [30] D.-H. Shin and F. Biocca, “Health experience model of personal informatics: The case of a quantified self,” *Computers in Human Behavior*, vol. 69, pp. 62–74, Apr. 2017, doi: 10.1016/j.chb.2016.12.019.
- [31] B. Aeon, A. Faber, and A. Panaccio, “Does time management work? A meta-analysis,” *PLOS ONE*, vol. 16, no. 1, p. e0245066, Jan. 2021, doi: 10.1371/journal.pone.0245066.
- [32] L. Jans-Beken Phd et al., “Gratitude and health: An updated review,” *The Journal of Positive Psychology*, pp. 1–40, Aug. 2019, doi: 10.1080/17439760.2019.1651888.
- [33] S. M. Schueller, M. Neary, J. Lai, and D. A. Epstein, “Understanding People’s Use of and Perspectives on Mood-Tracking Apps: Interview Study,” *JMIR Ment Health*, vol. 8, no. 8, p. e29368, Aug. 2021, doi: 10.2196/29368.
- [34] “Daylio - Journal, Diary and Mood Tracker,” Daylio. <https://daylio.net/> (accessed Mar. 19, 2023).
- [35] “aTimeLogger - Time Tracker - Apps on Google Play.” <https://play.google.com/store/apps/details?id=com.aloggers.atimeloggerapp&hl=en> (accessed Mar. 24, 2023).
- [36] “Presently: A Gratitude Journal - Apps on Google Play.” https://play.google.com/store/apps/details?id=journal.gratitude.com.gratitudejournal&hl=en_US (accessed Mar. 24, 2023).
- [37] “Python quickstart — Google Sheets,” Google Developers. <https://developers.google.com/sheets/api/quickstart/python> (accessed Mar. 26, 2023).
- [38] M. Koester, “Quantified Self (QS) Ledger.” Mar. 25, 2023. Accessed: Mar. 26, 2023. [Online]. Available: https://github.com/markwk/qs_ledger/blob/218e9d8abd52135f8c1de797a8ec390621bf6bc2/google_sheets/google_sheet_get_data_example.ipynb

- [39] M. Hamilton, “A Rating Scale for Depression,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 23, no. 1, pp. 56–62, Feb. 1960, doi: 10.1136/jnnp.23.1.56.
- [40] K. K. S. Rl, and W. Jb, “The PHQ-9: validity of a brief depression severity measure,” *Journal of general internal medicine*, vol. 16, no. 9, Sep. 2001, doi: 10.1046/j.1525-1497.2001.016009606.x.
- [41] S. Rl, K. K. W. Jb, and L. B, “A brief measure for assessing generalized anxiety disorder: the GAD-7,” *Archives of internal medicine*, vol. 166, no. 10, May 2006, doi: 10.1001/archinte.166.10.1092.
- [42] “ATUS home : U.S. Bureau of Labor Statistics.” <https://www.bls.gov/tus/> (accessed Apr. 01, 2023).
- [43] “High Blood Pressure Symptoms and Causes - cdc.gov.” <https://www.cdc.gov/bloodpressure/about.htm> (accessed Apr. 05, 2023).
- [44] “Innovators in Bluetooth Monitoring.” <https://bluemaestro.com/home> (accessed Apr. 05, 2023).
- [45] “PyCaret — pycaret 3.0.0 documentation.” <https://pycaret.readthedocs.io/en/stable/index.html#pycaret> (accessed Apr. 07, 2023).
- [46] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Apr. 07, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

APPENDIX

PHQ-9 Questions

1. Over the last two weeks, how often have you been bothered by: Little interest or pleasure in doing things?
2. Over the last two weeks, how often have you been bothered by: Feeling down, depressed, or hopeless?
3. Over the last two weeks, how often have you been bothered by: Trouble falling asleep, or sleeping too much?
4. Over the last two weeks, how often have you been bothered by: Feeling tired or having little energy?
5. Over the last two weeks, how often have you been bothered by: Poor appetite or overeating?
6. Over the last two weeks, how often have you been bothered by: Feeling bad about yourself—or that you are a failure or have let yourself or your family down?
7. Over the last two weeks, how often have you been bothered by: Trouble concentrating on things, such as reading the newspaper or watching television?
8. Over the last two weeks, how often have you been bothered by: Moving or speaking so slowly that other people could have noticed? Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual?
9. Over the last two weeks, how often have you been bothered by: Thoughts that you would be better off dead or of hurting yourself in some way?

GAD-7 Questions

1. Over the last two weeks, how often have you been bothered by: Feeling nervous, anxious, or on edge?
2. Over the last two weeks, how often have you been bothered by: Not being able to stop or control worrying?
3. Over the last two weeks, how often have you been bothered by: Worrying too much about different things?
4. Over the last two weeks, how often have you been bothered by: Trouble relaxing?
5. Over the last two weeks, how often have you been bothered by: Being so restless that it is hard to sit still?
6. Over the last two weeks, how often have you been bothered by: Becoming easily annoyed or irritable?
7. Over the last two weeks, how often have you been bothered by: Feeling afraid as if something awful might happen?

Scale

- 0 Not at all
- 1 Several days
- 2 More than half the days
- 3 Nearly every day

Table 1: PHQ-9/GAD-7 Questions and Scale

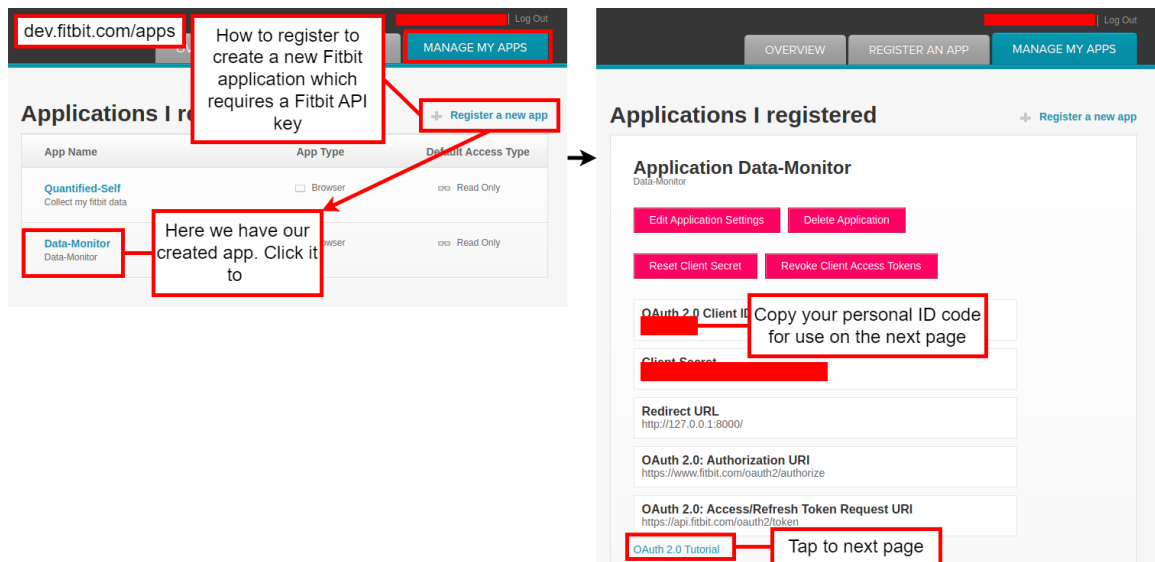


Figure 1: Fitbit API Data Flow Part 1

[Getting Started](#)
[Build](#)
[Manage](#)
[Community](#)

Fitbit OAuth 2.0 Tutorial

This interactive tutorial shows how to access user data from the Fitbit Web API using the OAuth 2.0 protocol. For more details see the [Developer Guide](#).

App Settings

First we need some settings for a registered application on the Fitbit Developer Site.

Client ID: Insert copied ID from previous page

Application Type: [Select](#)

Application Name:

Redirect URI:

Click to generate cryptographic key

PKCE Code Verifier: SUBMIT

PKCE Code Challenge:

Getting an Access Token

Accessing user data requires an Access Token. Here we will step through the process of asking a user to authorize your app and retrieving the token.

Step 1: Generate PKCE and State Values

For optimal security Fitbit recommends using the Proof of Key for Code Exchange (PKCE) and State aspects of the OAuth protocol. This entails having your app generate three unique values for each authorization request. Click the Generate buttons below to generate sample values.

PKCE Code Verifier: SUBMIT

PKCE Code Challenge:

Step 2: Display Authorization Page

The next step is to display Fitbit's Authorization web page to the user. The URL for this should include the PKCE Code Challenge, State and the requested permission scopes.

Scopes: The permission scopes to request access to. Only request what your app needs.

☒ activity	☒ cardio_fitness	☒ electrocardiogram	☒ heartrate
☒ location	☒ nutrition	☒ oxygen_saturation	☒ profile
☒ respiratory_rate	☒ settings	☒ sleep	☒ social
☒ temperature	☒ weight		

Click link and copy web address

Authorization URL:

Display this URL in a browser to ask the user to authorize your app.

Step 3: Handle the Redirect

After the user clicks Allow, your app will be redirected to the Redirect URL, specified in Application Settings. This is an endpoint on your app that will receive the authorization code and state. The URL will have two added query parameters: `code` and `state`.

Tip: If your app hasn't implemented the redirect endpoint, you will see an error or empty page in your browser. The URL should still contain the needed query params. For this tutorial, you can simply copy and paste the URL below and we will extract the params.

Authorization Code:

Step 4: Get Tokens

Now your app must exchange the authorization code for a pair of Access and Request Tokens. Here we will construct the HTTP request the app should make to the Token endpoint. This is also where the PKCE Code Verifier comes into play, which helps ensure tokens are delivered only to the app that requested them.

Server apps must pass an Authorization header with their Basic Token, which is their Client ID and Secret formatted like so: "Basic " + base64encode(client_id + ":" + client_secret).

Authorization:

Click

SUBMIT REQUEST

Response:

Step 5: Check Scopes

Finally the app should check which of the requested scopes the user actually granted. This is noted in the `scopes` field of the previous step's response. If the user did not grant a certain scope your app should not attempt to access the API endpoints that require it.

Access User Data

With an Access Token your app may start requesting user data.

Access Token

Access Token:

Fitbit API key that can be used to download the user's personal data

API Endpoint URL

The API endpoint to fetch user data from. The user must have granted permission for the relevant scope.

Example call using curl:

```
curl -i -H "Authorization: Basic " -X GET https://api.fitbit.com/1/user/:profile.json
```

SUBMIT REQUEST

Click to begin the authorization of the newly created Fitbit API key

Response:

Figure 2: Fitbit API Data Flow Part 2