MA-499 QUALITY OF LIFE RESEARCH PROJECT

Sponsored by Bristol Myers Squibb

Author
Julieta Caroppo, Zoe Hack, Joshua Richter
Stevens Institute of Technology, in collaboration with Bristol Myers Squibb
Spring 2024

Contents

1 Abstract

In this paper, we review the object of our research project and all of the progress that we have made over this semester. We review the history and use of "quality of life" and discuss our approach for measuring it. After considering 17 different data sources, identifying strengths and weaknesses of each, we decided to center our research around fatigue data to measure quality of life. We dig into the data for each of the three subjects with the largest amount of data available to draw conclusions that can be useful in quantifying quality of life. Additionally, we discuss conclusions that we have drawn from this data and examine possible future avenues for research that we did not have the time or scope to consider.

2 Background

2.1 Original Project Proposed

When we selected this project, the project overview was as follows:

> "Benefits of clinical intervention is often measured using quality of life improvement in addition to clinical benefits like event free survival, response etc. Health authorities are keen to understand the impact of treatment on quality of life. However, it is difficult to measure quality of life objectively. Recent studies have shown that some aspects of quality of life (i.e., mobility, sleeping quality, breathing and pulse rate) can be objectively measured using various physiological signal data from wearable. The purpose of this project will be to objectively calculate a quality of life score using vitals data. The expected output from this project should be measured in terms of the algorithm's ability to predict quality of life published in the past clinical studies. The algorithm should have high congruence with subjective quality of life data measured in clinical studies."

All three of us indicated interest in this project because we have a strong interest in data science and some relevant background experience.

2.2 Relevant Data Science Experience

All three of us have had unique opportunities and experiences in the realm of data science which allow us to stay engaged in this project and bring different perspectives to the table. Over the past summer, Joshua completed a data analytics internship with an enterprise software company and has an interest in data science as well as using technology and mathematics to make a positive impact on the world around us. Also this past summer, Julieta took part in a program at Columbia University's Statistics department where she took a course in linear regression and conducted some research on statistical analysis. Additionally, Julieta intends to work in the field of Data Science in the future, so she was very interested in this project and seeing real-world implications of data. For the past two summers, Zoe has interned at a reinsurance company in an actuarial department, working with a lot of different types of real-world data and additionally used different variables to create new measures for trending reinsurance payouts. All three of us were very eager to learn from each others' experiences and perspectives as well as to help create a metric that would have a positive affect on people's lives.

2.3 Delving into Quality of Life

Quality of life is a difficult subject to tackle much less to quantify, so we needed to research the history of "quality of life" itself in order to understand our project and its scope. Unfortunately, quality of life (or "QOL") is a very subjective term used widely among many different industries, so we will have to employ some judgement as well as asking BMS in the future when we work to create our metric.

2.3.1 Quality of Life Literature Review

In attempts to better understand the nuances behind measuring quality of life, a particular article from the Journal Topics in Spinal Cord Injury Rehabilitation became very useful. The article, published in 2014 is a literature review of academic articles concerning and defining the idea and or measure of quality of life. The review had three main objectives: (1) understand the concept of quality of life and how it has changed over time, (2) describe different ways quality of life has been measured and defined, and (3) provide recommendations for researchers to be as consistent and clear as possible when researching QOL [1]. For our purposes, we are mostly interested in some of the surface level information that objective (1) can provide us with and the definitions and metrics from objective (2).

The term "Quality of Life" has been used in the medical field since the 1960s. The first measure for QOL was published in 1984, named Spitzer's QL-Index. It breaks quality of life up into five main components: activity, scoring form (personal care/need or absence of assistance), health, support, and outlook. For each components, there are three options and it is intended that one option is chosen in each category based on what best represents the individual. In 2005, Dijker's Model was published with three major groups of subjective well-being, achievements, and utility. A main difference between Dijker's model and Spitzer's is that the former takes into account societal factors and pressures impacting one's quality of life. Similarly to Dijker's model, the 2011 PROMIS model published by the World Health Organization describes QOL as having three main components: physical health, mental health, and social health. It seems that as time has gone on, models for quality of life have gotten more precise and more accessible than previous ones, becoming easier to understand and hopefully getting more consistent results because there is less room for interpretation.

2.3.2 Canada's Well-Defined Quality of Life Framework

The Canadian government has a well-fleshed-out definition of quality of life, which we thought was helpful to reference throughout our research (Figure 1).

Given that this is an official measure from a government agency (albeit non-U.S.), this is definitely a good resource when computing our own measure. In total, they identify 84 indicators or components for quality of life after much research into the best way to quantify QOL. While there is still room for improvement and more testing, this is definitely a great baseline resource for our studies.

Figure 1: Canadian Quality of Life Framework [2].

2.4 The Use of Wearables Data

As users of wearable devices, we thought that using wearables data to compute quality of life would be a really interesting application of this project. In the past decade, wearable devices have become almost second nature, with approximately 600 million devices in use in 2020 [3]. These types of devices are used at both a consumer level and in the medical field for diagnostic and testing applications. Since these devices are already prevalent in both everyday life (in the form of smartwatches and fitness trackers) and in medical check-ups, trials, and hospital or doctors visits, we figured that by utilizing this type of data, it would be convenient to measure QOL with a data type that is already so widespread and easily accessible. Additionally, this would allow our measure to be easily used for both a commercial audience and in medical diagnoses prognoses.

3 Explorations

3.1 First Steps

In order to figure out a method in building an algorithm to define quality of life, we need data to explore health metrics that contributes to one's quality of life (QoL). We were originally planning to explore wearable data given by BMS. However, since there were some complications regarding that, we as a team decided to find some wearable data sets online that can be useful for our project. Each of us were tasked to find a few data sets, and report back to the team regarding their findings. Eventually, we came across a dataset online dedicated to examining fatigue relating it back to QoL.

3.2 Chosen Data Set

In the goal to measure Quality of Life (QoL) we delved into the intricate relationship between subjective well-being and objective physiological data. Our foundational steps involved a meticulous analysis of the dataset that the team agreed on titled "Continuous multi-sensor wearable data and daily subject-reported fatigue of healthy adults." This dataset is not only a repository of rich, multi-dimensional wearables data but also encapsulates daily fatigue levels reported by the subjects, offering not only an objective view, but also giving us a subjective view of their quality of life.

Understanding that QoL extends beyond mere clinical outcomes such as event-free survival or response rates, health authorities are increasingly prioritizing treatments that enhance patients' overall life experience. Herein lies the challenge: quantifying QoL in a tangible, objective manner. To address this, our project is pioneering an algorithmic approach that translates continuous streams of vitals data into a coherent QoL score. By harnessing the nuances of mobility, sleep quality, respiration, and pulse rate – all gleaned from wearable technologies – we aim to construct a robust model that mirrors the physical and psychological state of an individual.

The dataset we analyzed encompasses 405 recording days from 27 healthy adults, capturing a comprehensive spectrum of physiological signals at one-minute intervals. This includes metrics such as heart rate variability, activity counts, and galvanic skin response, amongst others, reflecting the physiological undercurrents of fatigue – a complex condition affecting physical and mental stamina. Accompanying these objective measures were daily subject-reported outcomes, offering a subjective perspective on fatigue levels. The synthesis of these data streams presents a unique opportunity to develop an algorithm that not only reflects the multifaceted nature of fatigue but also serves as a surrogate marker for QoL.

4 Prior Analysis

4.1 Overview

To build an understanding of the data we chose, each of the three team members performed exploratory statistical analysis on one of the three subjects with the most available data (Subjects 24, 26, and 27 from the dataset's CSV archive). We each had a different approach of what we decided to analyze for each subject, and so each of these three subjects will receive elaboration.

Table 1. List of parameters of the multisensor wearable device (Biovotion Everion), aggregation approach for downsampling, and computed daily summary features

Sensor	Parameter	Description	Unit	Downsampling	Daily features
Accelerometer	ActivityClass	Categorical parameter. Type of physical activity: 0 = undefined, 1 = resting, 9 = other, 10 = biking, 11 = running, 12 = walking	–	Mode	Count per category; One-hot encoding
	ActivityCounts	The activity value indicates the intensity of motion (activity)	–	Sum	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD, Sum
	Steps	Number of steps	–	Sum	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD, Sum
	EnergyExpenditure	Amount of energy a person uses to complete all regular bodily functions, measured in calories	Calories/s	Sum	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD, Sum
Photoplethysmography	HR	Heart Rate	bpm	Quality-weighted average	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
	HRV	Heart rate variability. Indicates the beat to beat variations.	ms	Quality-weighted average	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
	RESP	Respiration rate. Number of breaths a person takes per minute	bpm	Quality-weighted average	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
	BloodPerfusion	Blood perfusion can be measured as the percentage change in blood volume in local tissue resulting from a heartbeat	–	Median	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
	BloodPulseWave	Blood is ejected generating a pulse wave when the heart contracts	–	Median	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
Temperature	SkinTemperature	Skin temperature	°C	Median	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
Galvanic Skin Response	GalvanicSkinResponse	Describes changes in the electrical conductivity of the skin. It is a measure of emotional arousal	kOhm	Mean	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD
Barometer	Barometer	Barometric pressure measures changes of altitude	mbar	Median	Mean, SD, Median, Max., Min., Skewness, Kurtosis, 5th and 95th percentile, FFT, PSD

HR, heart rate; HRV, heart rate variability; RESP, respiratory rate; SD, standard deviation; Max., maximum; Min., minimum; FFT, fast Fourier transform (amplitude and frequency of 2nd to 5th peaks); PSD, power spectral density (amplitude and frequency of 2nd to 5th peaks).

Figure 2: Explanations of data collected from sensors, from data's original study.

As for the data itself, the above figure gives a brief overview of the data recorded from the sensors. Survey data was collected through a series of four questions. A pair of questions asked about mental and physical fatigue respectively on a descriptive scale that can be generalized into a numeric scale from 1 to 5. A third question asked about overall fatigue on a numeric scale from 1 to 10. The fourth question asked about relative fatigue: if the subject was feeling better than, worse than, or similar to what they have recently felt. Sensor data was timestamped every minute of the recording period (even when some or all fields had missing elements), while survey data was, for the most part, recorded once a day.

4.2 Subject 24

In analyzing Subject 24's data, one of the first big hurdles to overcome was the sheer amount of data. The survey data had well over 100000 measured timestamps, and so parsing that down to something workable for a human was a priority. While future work with this data will utilize more granular slices, aggregating the data to a daily scale for exploratory purposes was more than sufficient to glean some valuable insights.

```
> summary(daily.means)
    Class             Timestamp     Timestamp.1  ActivityClass    ActivityCounts      Barometer
 Length:139         Min.   : NA   Min.   : NA   Min.   : 1.000   Min.   : 0.7843   Min.   : 906.2
 Class :character   1st Qu.: NA   1st Qu.: NA   1st Qu.: 3.993   1st Qu.: 2.0472   1st Qu.: 972.6
 Mode  :character   Median : NA   Median : NA   Median : 4.603   Median : 2.4343   Median : 980.1
                    Mean   :NaN   Mean   :NaN   Mean   : 4.418   Mean   : 2.5118   Mean   : 981.8
                    3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: 5.049   3rd Qu.: 2.8376   3rd Qu.: 987.0
                    Max.   : NA   Max.   : NA   Max.   :10.200   Max.   :13.0719   Max.   :1035.4
                    NA's   :139   NA's   :139
 BloodPerfusion    BloodPulseWave   EnergyExpenditure GalvanicSkinResponse       HR               HRV
 Min.   :0.0000   Min.   :0.000   Min.   : 292      Min.   :0.4231       Min.   : 44.62   Min.   :17.11
 1st Qu.:0.4985   1st Qu.:1.828   1st Qu.:1652      1st Qu.:1.4252       1st Qu.: 63.55   1st Qu.:45.81
 Median :0.6228   Median :1.982   Median :2616      Median :1.8981       Median : 67.05   Median :50.28
 Mean   :0.6621   Mean   :1.919   Mean   :2904      Mean   :1.8591       Mean   : 68.70   Mean   :49.45
 3rd Qu.:0.7978   3rd Qu.:2.135   3rd Qu.:3383      3rd Qu.:2.2614       3rd Qu.: 70.56   3rd Qu.:55.05
 Max.   :1.6307   Max.   :2.607   Max.   :7396      Max.   :3.8065       Max.   :110.75   Max.   :72.42
 NA's   :1        NA's   :1                         NA's   :7            NA's   :1        NA's   :1
      RESP           Steps          SkinTemperature
 Min.   :10.27   Min.   : 0.0000   Min.   :14.15
 1st Qu.:14.08   1st Qu.: 0.2212   1st Qu.:32.60
 Median :14.50   Median : 8.9424   Median :33.20
 Mean   :14.43   Mean   : 8.0537   Mean   :32.69
 3rd Qu.:15.01   3rd Qu.:11.6945   3rd Qu.:33.76
 Max.   :16.46   Max.   :56.6107   Max.   :35.46
 NA's   :1
```

Figure 3: Summary data generated by R for Subject 24's hourly-aggregated sensor data.

Already, the summary reveals a few insights and flaws to come with this data. The listed NA's point to periods of time where, even over an entire day, no data was recorded for certain sensors (the Timestamp NA's are simply because the summary function is not designed to work with datetime values). Some summary statistics will naturally be more helpful than others; for instance, Activity class is a dubious measure where the sensor attempts to predict what manner of activity the subject is performing, and aggregations based off of it have shown unusual results, such as higher heart rates during what one would assume to be lower exertion activities. That said, several values in particular line up with expectations, such as heart rate ranging from around 40 to 110 beats per minute, with an average around man's average resting heart rate, or steps per minute remaining in a sensible range (though the average and quartiles may appear low, remember that tracking hours for the sensors include resting hours).

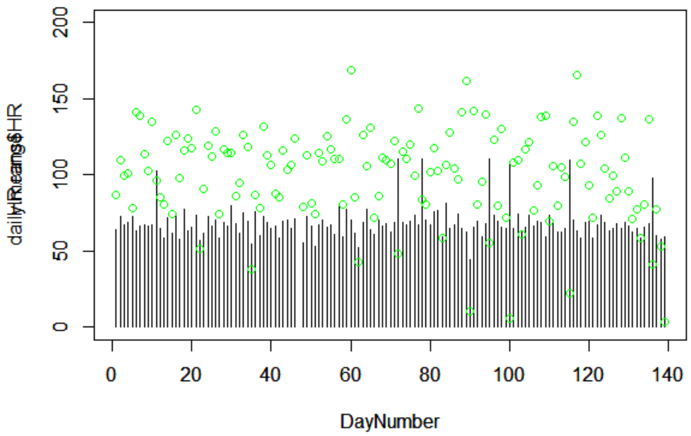

Figure 4: Plot generated by R for Subject 24's mean HR vs. HR range. Black bars are mean HR, green dots are HR range.

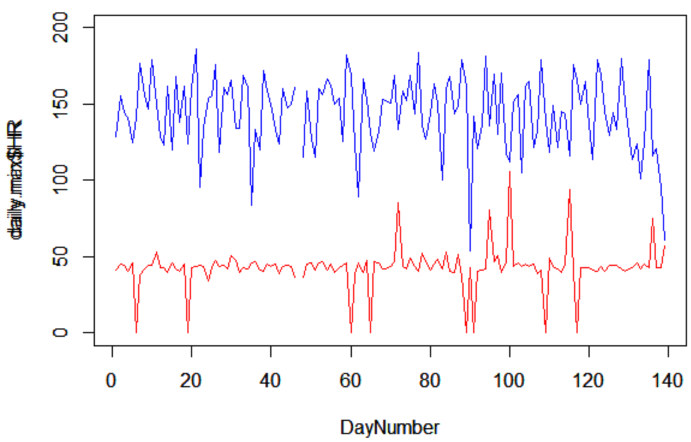

Figure 5: Plot generated by R for Subject 24's min HR vs. max HR.

Though commonly the focal point in statistics, mean on its own rarely tells a full story, especially when working with data prone to outliers such as this. Thus, developing a full understanding of the data through comparing means to maxes, minimums, and ranges was a goal with Subject 24's data. As can be seen in the above figures, while the mean heart rate was relatively stable over time, there was an appreciable amount of variation in the range. Delving further into this through the comparison of minimal and maximal heart rates, this disparity in range usually comes from variation in maximal heart rate, as minimal heart rate is, for the most part, relatively stable. This manner of analysis applies to many elements of this data (heart rate is just the most natural one to grasp the scale of, given it is one of the few sensor measures that man can self-measure).

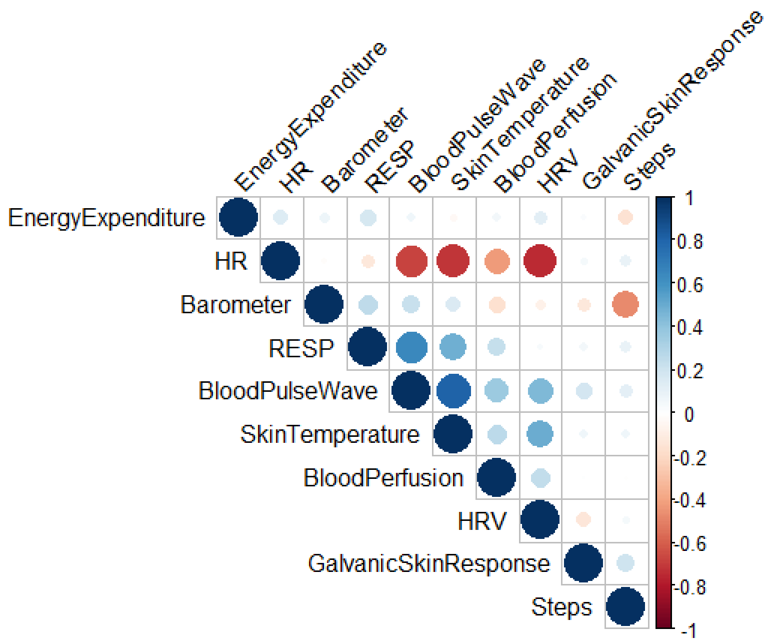

Figure 6: Correlation matrix generated by R for Subject 24's hourly-aggregated sensor data.

Of course, individual measures hardly tell a full story. Finding the correlation between various elements of sensor data is arguably an even more important step than just analyzing individual elements. The above correlation matrix captures the correlation between elements of the sensor data. Blue dots represent positive correlation and red dots negative, with the dot growing larger and darker in color the stronger the correlation. As can be seen, while many elements have weak correlations among each other, notable positive correlations exist among RESP, BloodPulseWave, SkinTemperature, and BloodPerfusion, all of which are related to blood flow in some way. Heart rate is negatively correlated with BloodPulseWave, SkinTemperature, BloodPerfusion, and HRV (heart rate variance), which has some interesting and counter-intuitive implications, such as skin temperature decreasing when heart rate increases (when normally one would associate an active heart rate with physical activity that heats the body up). Also intriguing is the negative correlation between measured barometer pressure and steps taken per minute, though given that it is not a particularly strong correlation, there is always a chance that it is just coincidence, or result of a flaw in sensor design.

On the note of important measurements, strong correlation among aforementioned blood-related measures likely means that when working with data integration, we will be able to trim down some redundancies. While the measures do not act exactly one-to-one with respect to each other, they act similarly enough to be worth the reduction in algorithmic complexity.

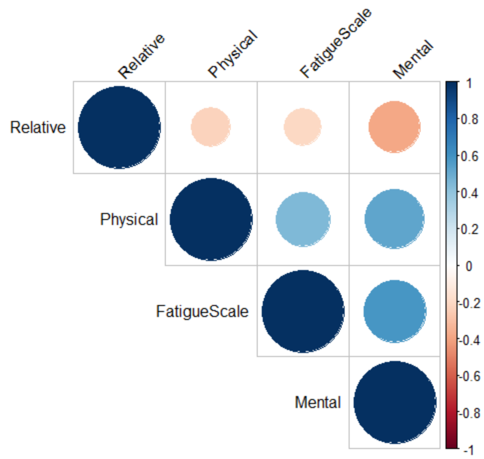

Figure 7: Correlation matrix generated by R for Subject 24's survey data.

There is not quite as much to say about Subject 24's survey data. Relative fatigue was interpreted as a value of -1 for relatively worse, 0 for relatively the same, and 1 for relatively better, and so the correlations are self-explanatory for the most part. If relative fatigue is negative, then the fatigue measures are larger because the subject is feeling worse. Unsurprisingly, physical and mental fatigue are quite strongly correlated, which will be worth keeping in mind when developing a quality of life metric. A slight challenge in working with relative fatigue in particular is that it is somewhat of a delayed measurement, and so there may be value in more directly comparing it to differences in fatigue as opposed to just a singular fatigue value.

4.3 Subject 26

For subject 26, there were originally 81575 records, making it the third largest dataset out of all of the subjects. After parsing through the data by deleting empty records, the following depicts the data summary for subject 26.

```
   SubjectID     DateTime           Timezone           PROquestion        PROanswer_value PROanswer_choice
 Min.   :26   Length:228         Length:228         Length:228         Min.   :1.000   Length:228
 1st Qu.:26   Class :character   Class :character   Class :character   1st Qu.:3.000   Class :character
 Median :26   Mode  :character   Mode  :character   Mode  :character   Median :6.000   Mode  :character
 Mean   :26                                                            Mean   :5.053
 3rd Qu.:26                                                            3rd Qu.:7.000
 Max.   :26                                                            Max.   :9.000
                                                                       NA's   :171
  Timestamp          ActivityClass    ActivityCounts       Barometer      BloodPerfusion    BloodPulseWave   EnergyExpenditure
 Length:81575       Min.   : 1.000   Min.   : 0.7843   Min.   : 901.9   Min.   :0.0000   Min.   :0.000   Min.   :   40.09
 Class :character   1st Qu.: 1.000   1st Qu.: 0.7843   1st Qu.: 981.5   1st Qu.:0.3300   1st Qu.:1.780   1st Qu.: 1256.06
 Mode  :character   Median : 1.000   Median : 0.7843   Median : 985.4   Median :0.4800   Median :1.980   Median : 1381.00
                    Mean   : 3.992   Mean   : 1.7882   Mean   : 985.2   Mean   :0.5299   Mean   :2.021   Mean   : 1769.07
                    3rd Qu.: 9.000   3rd Qu.: 1.1765   3rd Qu.: 989.1   3rd Qu.:0.6800   3rd Qu.:2.250   3rd Qu.: 1911.90
                    Max.   :12.000   Max.   :73.3333   Max.   :1008.5   Max.   :2.5500   Max.   :4.920   Max.   :15609.66
                    NA's   :2077                       NA's   :34       NA's   :332      NA's   :332
 GalvanicSkinResponse       HR               HRV              RESP           Steps          SkinTemperature
 Min.   :0.131         Min.   : 36.00   Min.   :  0.00   Min.   : 0.00   Min.   :  0.000   Min.   :24.86
 1st Qu.:0.937         1st Qu.: 52.63   1st Qu.: 43.53   1st Qu.:14.25   1st Qu.:  0.000   1st Qu.:34.88
 Median :1.827         Median : 63.18   Median : 54.73   Median :15.55   Median :  0.000   Median :35.56
 Mean   :1.999         Mean   : 64.96   Mean   : 55.96   Mean   :15.71   Mean   :  4.276   Mean   :35.51
 3rd Qu.:2.883         3rd Qu.: 72.35   3rd Qu.: 66.82   3rd Qu.:16.87   3rd Qu.:  0.000   3rd Qu.:36.25
 Max.   :5.000         Max.   :179.00   Max.   :137.50   Max.   :39.00   Max.   :175.000   Max.   :51.65
 NA's   :10465         NA's   :332      NA's   :4588     NA's   :1515
```

Figure 8: Subject 26 Data Summary.

Looking at the above summary, it is important to note that there are still some fields with a value of 'NA' because in this data refinement, only records with all fields being blank or 'NA' were removed.

Below are some visualizations of different variables measured for subject 26.

Figure 9: Distribution of Subject 26's collected heart rate data.

The distribution of Heart Rate data makes sense, having a skew right which would indicate bouts of exercise and higher exertions of energy.

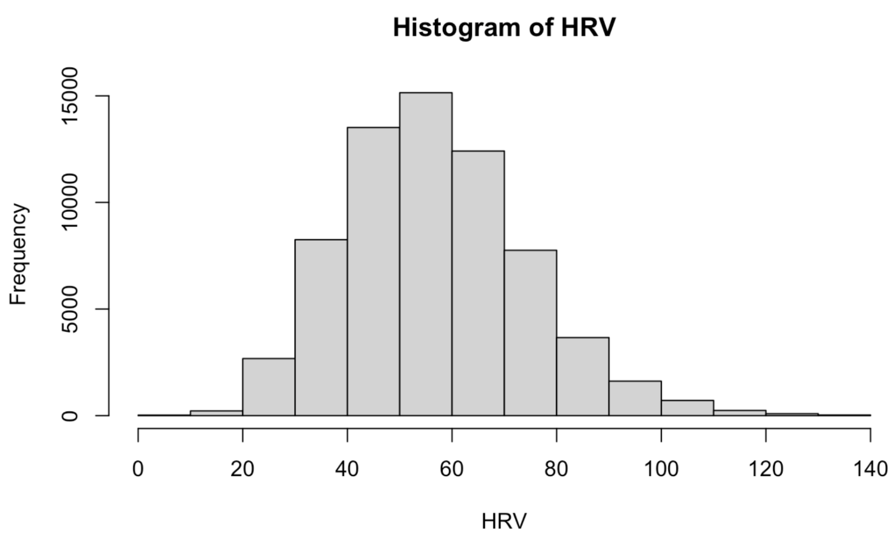

Figure 10: Distribution of Subject 26's collected heart rate variation data.

The distribution of HRV, or Heart Rate Variation, seems to be a relatively normal distribution with a mean of around 50, something that can be useful to identify unlikely or unexpected variations of heartrate.

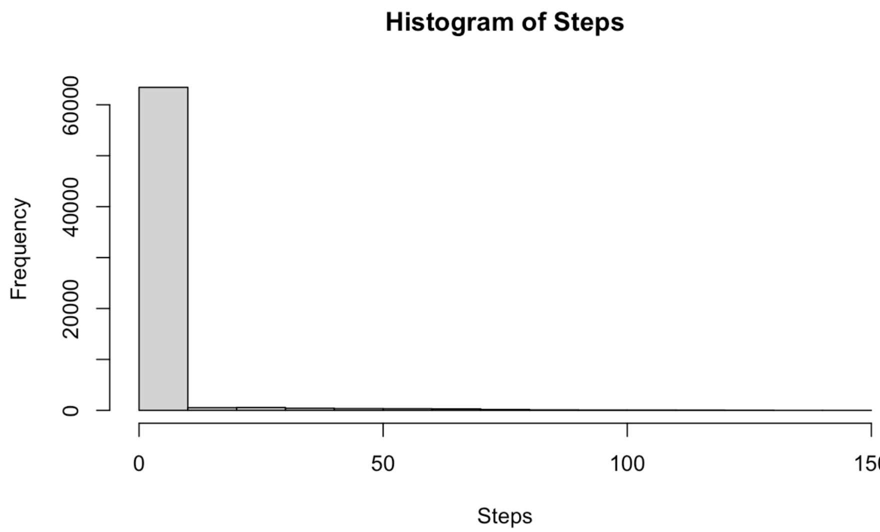

Figure 11: Distribution of Subject 26's collected step data.

At first look, the distribution of steps seems to not make much sense, having most of the fields having zero steps. However, this is most likely due to the fact that steps weren't taken at every timestamp, so steps may not be a helpful metric in our measure of quality of life. This also helps us to know that at least for now, our measure for QOL may not have any immediate use for commercial applications since the count of steps are what commercial wearable devices typically center around calculating.

What is arguably more important than understanding individual variables is understanding how different variables interact with each other. The following correlation matrix can help to understand some of these interactions.

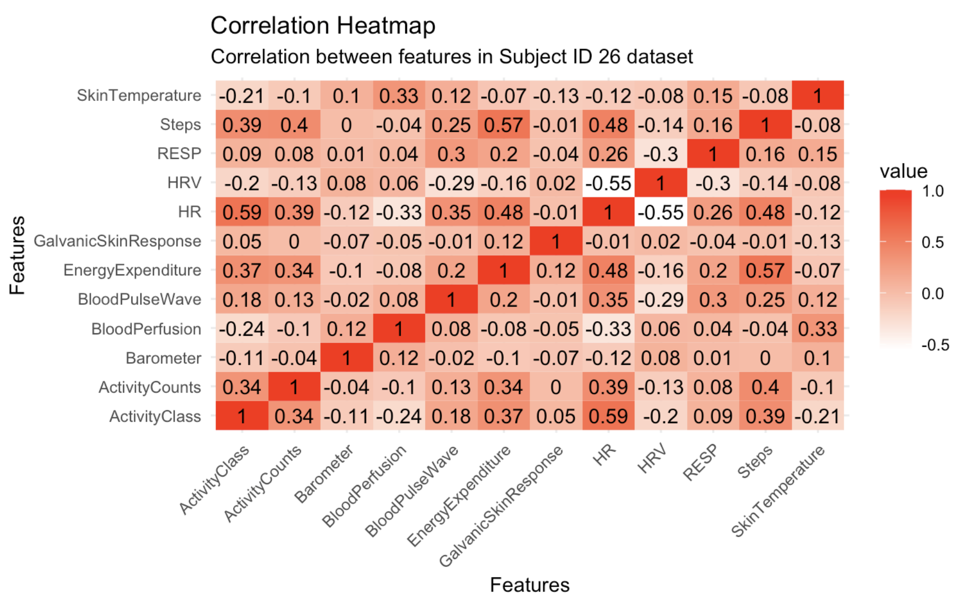

Figure 12: Correlation Matrix for Subject 26

According to the correlation matrix heatmap, we can see that the pairs of variables with the highest correlations are Heart Rate and Activity class with a positive correlation of 0.59, Steps and Energy Expenditure with a positive correlation of 0.57, and Heart Rate and Heart Rate Variation with a negative correlation of -0.55. The first two of these make sense intuitively: the higher numeric values for activity class correspond to more intense activity, which would definitely cause higher heart rate during exercise, so this positive relatively high correlation definitely makes sense. For steps and energy expenditure, even though as seen above there is not much useful data on steps, it also intuitively makes sense that more movement would cause higher amounts of energy expenditure. For the relationship between HR and HRV, a correlation of -0.55 indicates that higher values of heart rate cause the measure of heart rate variation to decrease. This relationship can be potentially useful in our future investigations for a quality of life measure depending on if we can quantify either heart rate or heart rate variation to a specific measure in quality of life since we know the relationship between HR and HRV.

4.3.1 FatiguePROs for Subject 26

The FatiguePROs dataset is an objective measure of fatigue corresponding to the physical measured metrics described above. The first question asked respondees to rate their fatigue for the day on a scale of 1 to 10, with 10 being "the worst tiredness you can imagine." The third and fourth questions (our other questions of interest) have the same answer choices: Never, Sometimes, Often, Regularly, Always. Question 3 asks how often the respondee felt physically exhausted that day and question 4 asks how often they felt mentally exhausted that day.

A linear regression was performed to see how physical and mental fatigue impact overall fatigue for subject 26. In order to do this, the answer choices "Never, Sometimes, Often, Regularly, Always" corresponded to "0, 1, 2, 3, 4".

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.689612077
R Square	0.475564817
Adjusted R Square	0.456141292
Standard Error	1.510208855
Observations	57

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	111.6826428	55.84132141	24.48396004	2.70196E-08
Residual	54	123.1594624	2.280730786		
Total	56	234.8421053			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.873888064	0.499238096	3.753495734	0.000428058	0.872975945	2.874800183	0.872975945	2.874800183
PE_Num	1.035511263	0.272477977	3.800348476	0.000368987	0.489225811	1.581796715	0.489225811	1.581796715
ME_Num	0.921344117	0.212018519	4.345583221	6.1745E-05	0.49627258	1.346415654	0.49627258	1.346415654

Figure 13: Linear regression of physical and mental exhaustion against overall fatigue for subject 26.

The results of this linear regression indicate that for subject 26, physical exhaustion impacts overall fatigue more than mental exhaustion does, indicating that Fatigue Score = 1.874 + 1.036(Physical exhaustion) + 0.921(Mental Exhaustion). The p-values less than 0.05 for each of the variables and intercept indicate that there is statistically significant evidence that fatigue is made up of these weights of mental and physical exhaustion. Understanding how mental and physical exhaustion may impact fatigue can help us understand how these measure can also impact quality of life.

4.4 Subject 27

When beginning the data analysis process for subject 27, there are two types of data sets to approach and analyze. One was a CSV file of Subject 27's health vitals along with the timestamps. The dimensions of this CSV file had 313583 rows, and 13 columns (12 being each of the health parameters, and one being the time stamps). The other was fatiguePROs_27, which contained subjective data of answers to survey questions being asked (as mentioned in the Overview).

```
  SubjectID    DateTime           Timezone           PROquestion        PROanswer_value PROanswer_choice
 Min.   :27   Length:252         Length:252         Length:252         Min.   :1       Length:252
 1st Qu.:27   Class :character   Class :character   Class :character   1st Qu.:1       Class :character
 Median :27   Mode  :character   Mode  :character   Mode  :character   Median :2       Mode  :character
 Mean   :27                                                            Mean   :2
 3rd Qu.:27                                                            3rd Qu.:3
 Max.   :27                                                            Max.   :5
                                                                       NA's   :189
  Timestamp         ActivityClass    ActivityCounts     Barometer      BloodPerfusion   BloodPulseWave   EnergyExpenditure
 Length:313583      Min.   : 0.00    Min.   : 0.78    Min.   : 901.3   Min.   :0.00     Min.   :0.00     Min.   :   22.18
 Class :character   1st Qu.: 1.00    1st Qu.: 0.78    1st Qu.: 965.1   1st Qu.:0.24     1st Qu.:2.74     1st Qu.: 1331.00
 Mode  :character   Median : 1.00    Median : 0.78    Median : 968.8   Median :0.35     Median :3.06     Median : 1350.42
                    Mean   : 3.46    Mean   : 1.83    Mean   : 973.3   Mean   :0.37     Mean   :3.07     Mean   : 1526.66
                    3rd Qu.: 9.00    3rd Qu.: 1.96    3rd Qu.: 979.4   3rd Qu.:0.47     3rd Qu.:3.41     3rd Qu.: 1376.73
                    Max.   :12.00    Max.   :38.82    Max.   :1010.8   Max.   :2.18     Max.   :5.00     Max.   :11405.41
                    NA's   :281938   NA's   :271185   NA's   :271304   NA's   :271378   NA's   :271401   NA's   :271185
 GalvanicSkinResponse       HR              HRV              RESP             Steps        SkinTemperature
 Min.   :0.33         Min.   :  0.00   Min.   : 11.00   Min.   : 0.00    Min.   :  0.00   Min.   :25.47
 1st Qu.:1.67         1st Qu.: 64.99   1st Qu.: 36.57   1st Qu.:14.84    1st Qu.:  0.00   1st Qu.:34.25
 Median :2.77         Median : 71.92   Median : 43.87   Median :16.30    Median :  0.00   Median :34.94
 Mean   :2.90         Mean   : 72.88   Mean   : 46.61   Mean   :16.79    Mean   :  4.01   Mean   :34.94
 3rd Qu.:4.23         3rd Qu.: 79.10   3rd Qu.: 53.16   3rd Qu.:18.09    3rd Qu.:  0.00   3rd Qu.:35.63
 Max.   :5.00         Max.   :154.84   Max.   :134.70   Max.   :39.00    Max.   :138.00   Max.   :46.28
 NA's   :304998       NA's   :271378   NA's   :276477   NA's   :271908   NA's   :271185   NA's   :271185
```

Figure 14: 5 number summary of Subject 27.

To begin our data analysis on the sensor data, we produce a 5-Number Summary. Subject 27's activity profile, as indicated above, suggests a predominantly sedentary or light activity lifestyle, with occasional bursts of more intense exercise or movement. The ActivityClass parameter, which categorizes the type of physical activity, shows that the subject spent most of their time engaged in low-intensity activities, with a median class of 3.46 indicating activities such as resting. ActivityCounts, capturing the intensity of motion, peaked at 38.82 but had a median of only 1.83, reinforcing the predominance of sedentary behavior. Cardiovascular data, encompassing heart rate (HR) and heart rate variability (HRV), reflects a normal resting heart rate with a median of 71.92 beats per minute (bpm), yet exhibits a broad range that reaches up to 154.84 bpm. The variability in HRV, stretching to a maximum of 134.70 milliseconds (ms), suggests fluctuating stress or activity levels. However, the presence of numerous missing data points for HR and HRV (over 270,000 NA's each) calls into question the completeness of the cardiovascular dataset.

Respiratory patterns, as observed through the respiration rate (RESP), hover within a normal range, averaging around 16.30 bpm. However, like the cardiovascular data, a significant number of missing data entries (over 271,000 NA's) could hinder the interpretation of respiratory trends.

The steps recorded are minimal, with a median count of zero, which could be attributed to sensor error or the subject's inactive periods. This finding is intriguing and somewhat contradictory to the activity data, possibly pointing to sensor inaccuracies or non-wear time. Similarly, skin temperature measurements ranged between 25.47°C to 46.28°C, with a median of 34.94°C, which aligns with normal skin temperature variation but also suffers from a high number of missing values (over 271,000 NA's).

Galvanic skin response, a proxy for emotional arousal, shows a median of 2.77 with values stretching up to 5, indicating diverse states of arousal or stress throughout the data collection period. Again, the substantial missing data (over 304,000 NA's) may limit insights into the subject's stress patterns.

Lastly, the barometric pressure readings and energy expenditure metrics suggest that Subject 27 experienced a wide range of environmental conditions and fluctuating activity levels, as demonstrated by the considerable range in calories burned—from a mere 22.18 to a staggering 11,405.41 calories. The accuracy and implications of these extremes would need careful validation considering potential outliers or errors.

In conclusion, Subject 27's recorded health parameters sketch a picture of a subject with fluctuating activity levels, typical physiological responses, and potential periods of stress or emotional arousal. Nonetheless, the interpretation of these patterns is significantly challenged by the extensive missing data across several health parameters, highlighting the importance of addressing these gaps for robust analysis.

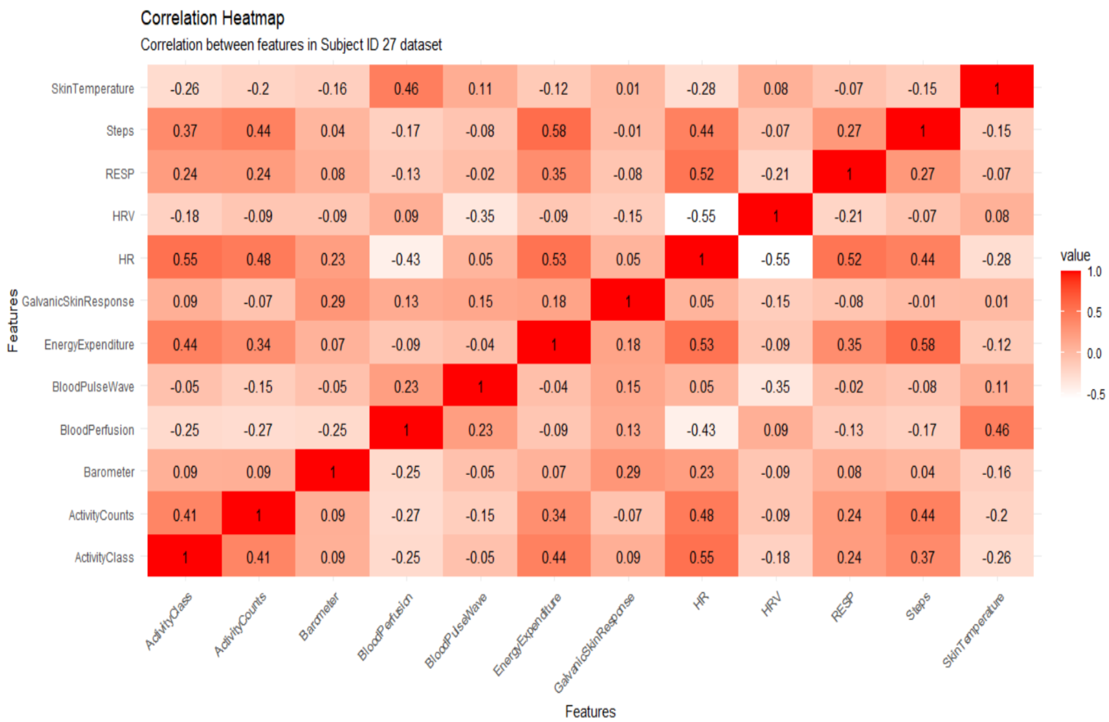

Figure 15: Heat map of health vitals generated by R for Subject 27's survey data.

In the figure above we can see the analysis of the correlation heatmap has yielded several critical insights into the interrelationships among the variables. Bright red cells in the heatmap signal strong positive correlations, with values nearing 1 suggesting a direct relationship where an increase in one variable is mirrored by an increase in another. In the heatmap the highest correlation we can see is between "Steps" and "EnergyExpenditure," which stands at 0.58, either underscoring the direct contribution of step count to activity counts or activity class.

Contrastingly, the heatmap lacks deep blue cells, which would indicate strong negative correlations, where one variable's increase would correspond to another's decrease. The most negative correlations we see HRV and HR, having a value -0.55. This means that, generally, as the heart rate increases, the variability between heartbeats tends to decrease, and vice versa.

The moderate negative correlation of -0.55 between Heart Rate (HR) and Heart Rate Variability (HRV) as captured on Subject 27 suggests an inverse relationship where an increase in heart rate typically corresponds to a decrease in the variability between heartbeats. This correlation is physiologically plausible and reflects the body's expected responses to various states of activity and rest. During periods of relaxation or low physical activity, the heart rate tends to be lower and the heart's rhythm more variable, indicating a responsive autonomic nervous system and good cardiovascular health. Conversely, when the subject engages in physical activity or experiences stress, the heart rate increases and the rhythm becomes more consistent, leading to a reduction in HRV.

Such wearable device data offers valuable insights into a subject's cardiovascular system's adaptability and overall health. For instance, a consistently high heart rate coupled with low HRV could signal chronic stress or potential overtraining, which may have adverse effects on the subject's quality of life. In the broader context of health monitoring and wellness, understanding the relationship between HR and HRV can be instrumental in assessing the impact of lifestyle on cardiac health and stress levels. When integrating these measurements into an algorithm designed to gauge Quality of Life, they serve as critical biomarkers reflecting the interplay between a person's physiological state, their daily activities, and perceived well-being.

In terms of variable relationships, some, like "HR" (Heart Rate) and "BloodPerfusion," exhibit moderate positive correlations with various other variables, suggesting their potential as indicators of overall activity level or physiological response. These correlations are particularly important when considering data usage for predictive modeling. In such scenarios, highly correlated variables may introduce redundancy, and one might be omitted to simplify the model. However, this step must be approached with caution to prevent the loss of significant information.

In summary, the heatmap indicates a diverse range of relationships, with certain variables, especially those associated with physical activity, showing positive correlations.

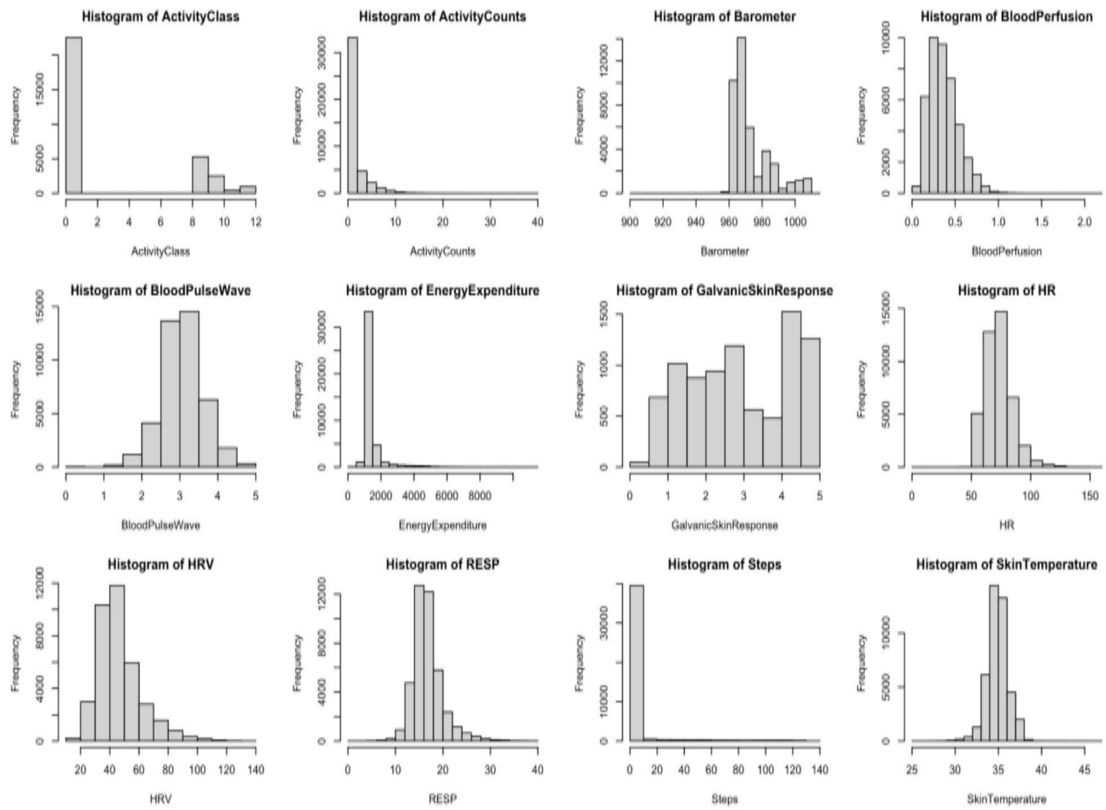

Figure 16: Histograms for sensor variables generated by R for Subject 27's survey data.

Now, let us go more in depth with Subject 27's sensor data variables. Specifically, let us explore the distribution shapes, central tendencies, and variability.
From the produced histograms we can examine some notable shapes of the variables.

Histogram of ActivityClass: The histogram shows a right-skewed distribution, with the majority of the data concentrated in the lower activity classes. This suggests that Subject 27 spent most of their time engaged in sedentary or light activities, with very few instances of higher-intensity activities, or their wearable devices was not active.

Histogram of ActivityCounts: This histogram also leans towards the lower end, with a sharp peak near zero, which indicates that the subject had many periods of low activity. The long tail to the right suggests occasional periods of higher activity, but these are infrequent.

Histogram of Barometer: The barometric pressure readings show a multi-modal distribution with several peaks, suggesting there were distinct ranges within which barometric pressure commonly fell. This could be due to changes in temperature or the subject moving between different altitudes.

Histogram of BloodPerfusion: Blood perfusion has a distribution that peaks at lower values and then tails off, which indicates that most of the time, the blood perfusion measured was on the lower end of the scale, with fewer high-perfusion events.

Histogram of BloodPulseWave: This histogram is roughly symmetrical and bell-shaped, suggesting a normal distribution around a central value. This indicates a consistent pattern in the pulse wave measurements from the subject's blood ejections.

Histogram of EnergyExpenditure: There is a high frequency of lower values with a right-skewed distribution, showing that Subject 27 typically expended low amounts of energy, with occasional spikes representing periods of high energy expenditure.

Histogram of GalvanicSkinResponse: The distribution is somewhat uniform with slight peaks, suggesting that the subject experienced a variety of arousal levels fairly evenly, without a clear central tendency.

Histogram of HR (Heart Rate): The heart rate histogram appears right-skewed, with most of the data concentrated at lower heart rates and fewer occurrences of higher rates, which could indicate that the subject was mostly at rest with occasional periods of activity.

Histogram of HRV (Heart Rate Variability): This histogram shows a right-skewed distribution, suggesting that the subject generally had lower variability in their heart rate, with fewer instances of high variability.

Histogram of RESP (Respiration Rate): The respiration rate histogram is right-skewed, similar to heart rate, suggesting that the subject's respiration rate was generally low, with infrequent periods of higher rates.

Histogram of Steps: This histogram is extremely skewed to the right, with almost all the data concentrated at zero. This indicates that the subject took very few steps most of the time, which is consistent with the sedentary activity class and count data.

Histogram of SkinTemperature: The skin temperature histogram is roughly normally distributed with a slight right skew, suggesting that while there's a central temperature around which most readings are clustered, there are occasional periods of higher temperatures.

In summary, Subject 27's data shows a predominance of sedentary behavior with sporadic bouts of activity, generally low heart rate and variability, and minimal steps taken throughout the monitoring period. The physiological responses measured by skin temperature, blood perfusion, and galvanic skin response show a range of values with varying degrees of centralization and spread, reflecting the subject's diverse reactions to their environment and conditions.

Shifting our focus to the subjective data, there are four questions that were asked. Subject 27, answered these questions daily, but only for the months of February, March, August, and September. To interpret the trend of these questions, there are some time series diagrams created in the figure below depicting it. Only the month of March will be displayed for the focused question in the figure below.

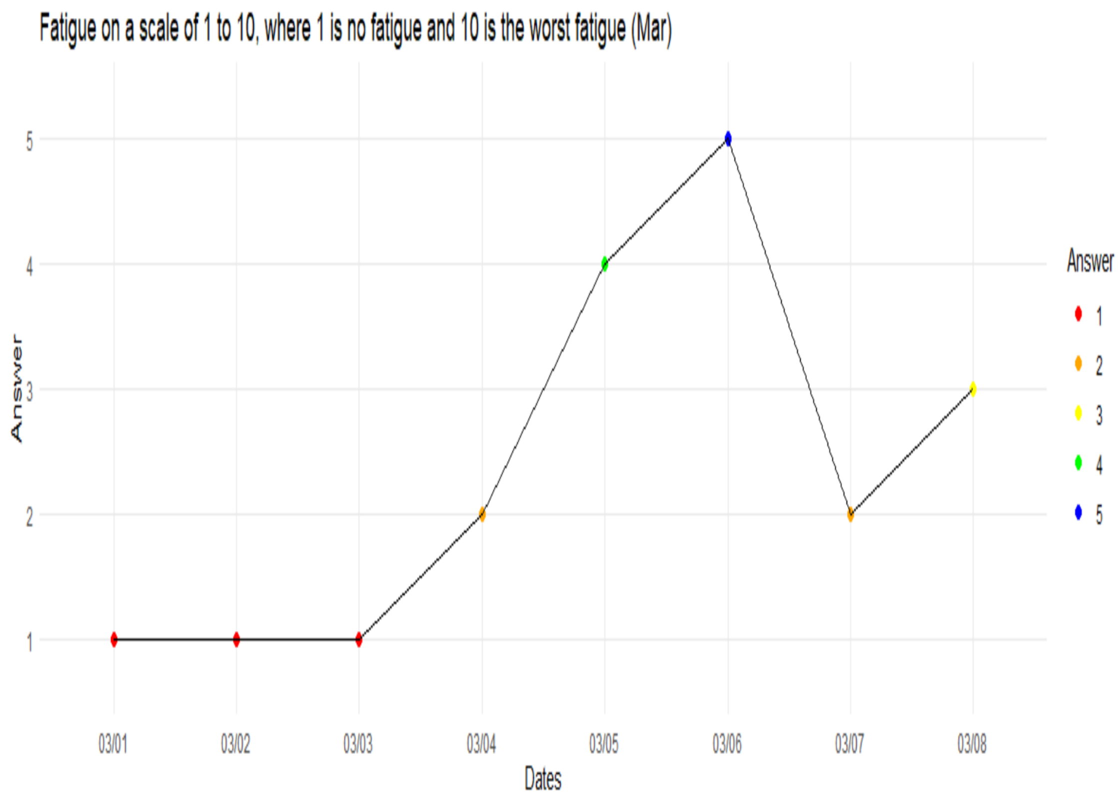

Figure 17: Histograms for sensor variables generated by R for Subject 27's survey data.

The time series created above illustrates Subject 27's self-reported fatigue levels over a span of eight days in March (since there was missingness among the four months). The graph plots fatigue on a scale of 1 to 10, with 1 indicating no fatigue and 10 representing the worst fatigue.

Starting from the 1st of March, Subject 27 reported minimal fatigue, with a score of 1, depicted in red dots. This low level of fatigue continues consistently on the 2nd of March. On the 3rd of March, there is no data point, which might indicate that the subject did not report their fatigue level on this day, or there could have been an issue with data collection.

On the 4th of March, there is a noticeable increase in fatigue levels, with the subject reporting a level of 2, shown in orange. This upward trend in fatigue continues sharply as we move to the 5th of March, where the fatigue level more than doubles to a score of 5, marked in blue, indicating a significant increase in the subject's perceived fatigue.

However, the 6th of March shows a dramatic decline in fatigue levels back down to a score of 2. This suggests some form of recovery or rest may have taken place, or a possible respite from the activities or factors causing the fatigue.

The graph shows another increase on the 7th of March, with fatigue levels rising again, this time to a score of 3, represented in yellow. Finally, the series ends on the 8th of March with the fatigue level dropping slightly to a score of 2.

Overall, the graph indicates fluctuations in the subject's fatigue levels, with notable peaks and troughs. The data suggests variability in the subject's daily experiences, which could be due to several factors such as physical activity, sleep quality, stress, and other health-related issues. The graph serves as a visual tool to quickly identify days with unusually high or low fatigue, which can be a starting point for further investigation into the causes of these fluctuations.

5 New Findings

5.1 Overview

This semester's chunk of the project was largely focused on exploring possible metrics of quality of life, as well as digging more into the self-reported fatigue scores. Additionally, in contrast to the single-subject per team member focus we took in the first semester, we focused on applying methodologies to multiple subjects, as to make sure that methodologies could be more broadly applied to our population rather than being overly tuned for one specific subject.

5.2 Average Days

Defining quality of life is far from trivial for many reasons; among them is the fact that everybody has a different baseline for their health an activity. What one person may call an extraordinarily strenuous day, another may call below-average. Thus, creating a singular baseline for quality of life based off of absolute variation from some defined objective data metrics would result in a model with no manner of adaptability. Rather, it would be wise to detect data that differs significantly from a subject's personal baseline.

While experimenting with trying to devise individual metrics that may correlate with regards to quality of life for a singular subject, it was noticed that, arranging the 24 hours of the day as an axis against sensor data, a clear pattern consistent with our preconceptions emerged. As an example, heartrate would be stable and relatively low during sleeping hours, would increase in morning hours, spiking around the tenth hour, decreasing midday, sometimes spiking again around the thirteenth or fourteenth hour, and usually spiking one last time in the evening before leveling out during sleeping hours.

We would proceed to apply this concept to every subject's data. This would lend credence to our preconceptions of an observable daily pattern of measurements. To create each day in the life of a subject, we took the maximal heart rate measured during a particular hour.

Note that heartrate, like all measurements, was taken every minute. Prior discussions explain our reasoning for this metric, but in summary we felt it was more informative and less liable to being impacted by device failure than something like average heart rate over an hour. For comparison, we also took total steps per hour and created "average days" in a similar fashion.

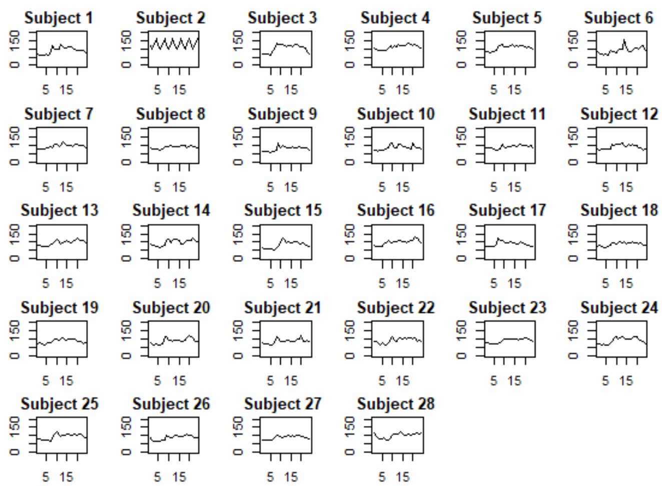

Figure 18: Plot generated by R for every subject's "average day" heartrate. Note that some subjects have minimal data that ends up breaking this methodology (most notably, Subject 2).

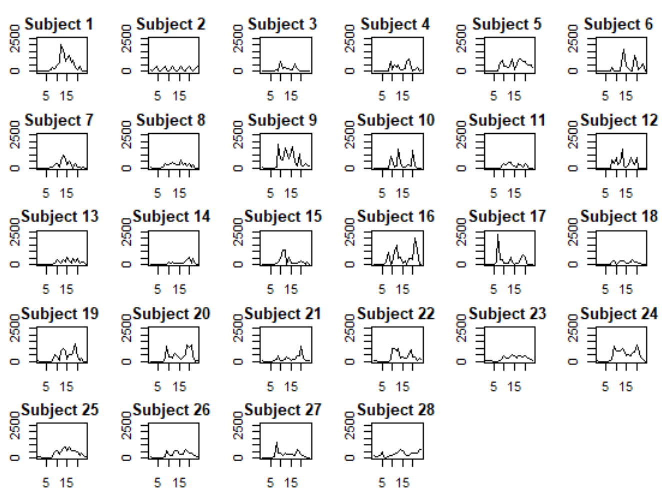

Figure 19: Plot generated by R for every subject's "average day" steps/hour. Note that some subjects have minimal data that ends up breaking this methodology (most notably, Subject 2).

With knowledge of each subject's "average day", so to speak, a natural followup is to compare it to other days. Aggregation can create patterns that make us assume the typical day functions differently than in reality. For example, if a subject is active either at 9, 10, or 11 in the morning on most days, but which one of those hours they are active is

inconsistent, an aggregation might imply a plateau of continuous activity, or two miniature spikes in a three hour period.
The natural course of progression would be to followup by examining if and how a connection exists between a day's deviation from the average day and that day's self-reported fatigue. Due to time limitations, study of this correlation by this section's author was not possible, but if a correlation were to exist between a day's "non-averageness" and a subject's self-reported fatigue, that would obviously be a very potent measure. If a correlation were not to exist, that would also be a very interesting result, as it could further imply mankind's lacklustre ability to gauge their personal fatigue in a meaningful way.

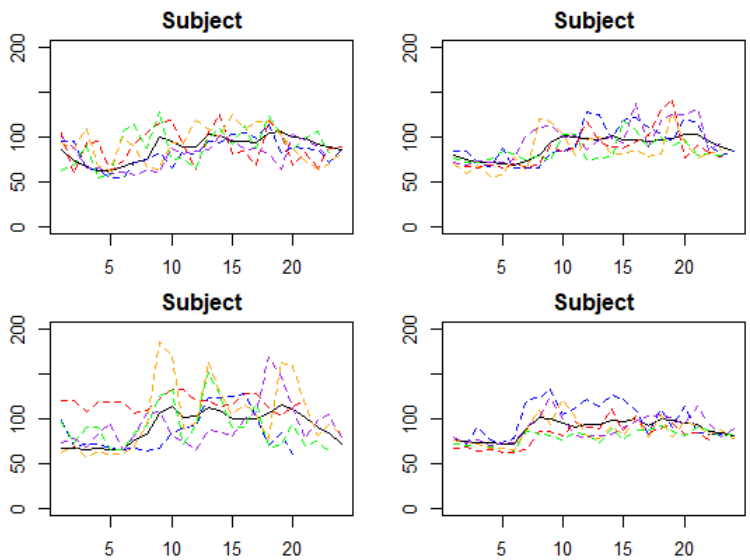

Figure 20: Plot generated by R for four subjects' "average day" and several randomly chosen days. The solid black lines represent their "average day" while the dashed colored lines each represent a different random day.

5.3 Feature Importance and Predicting Survey Answers

Another aspect we observed was the subjective data. We wanted to see how well we can predict one's answers on how they are feeling and to see if one actually knows how they are doing based on their wearable data. This is important when studying quality of life because it allows us to understand the correlation between physiological metrics and self-reported well-being, thereby offering insights into how accurately individuals can self-assess their health status. This can lead to better personalized health interventions and improved monitoring of daily health fluctuations.

First, we planned an approach on how we make a machine learning model to predict each question one subject (here we will use Subject 27) had to answer. The first step was to select which health metrics that would be good to add to a model based on their importance and impact. In order to do that, we clean the data and format the dates to ensure compatibility between the two datasets. We parsed the timestamps for both datasets, applying the necessary adjustments to align the date formats. We then applied this parsing to merge the datasets on the Date column. To summarize each objective feature in the wearable dataset, we chose the median for most variables, as it is a reliable summary statistic. For the steps feature, we used the sum, as it is more appropriate for cumulative measures.

Focusing on Question 1, which asked subjects to describe their fatigue on a scale of 1 to 10, we leveraged the fact that the answers were already in numerical format. This allowed us to directly analyze the relationship between fatigue levels and the wearable data features. Our primary goal was to determine which features were driving fatigue.

We utilized a Random Forest Regression model to identify the most important features in predicting fatigue. This model was chosen due to its robustness and ability to handle complex interactions between features. Once we identified the significant features, we quantified their importance and displayed them. The features found to be most influential included Heart Rate (HR), Steps, BloodPulseWave, Barometer, and Energy Expenditure.

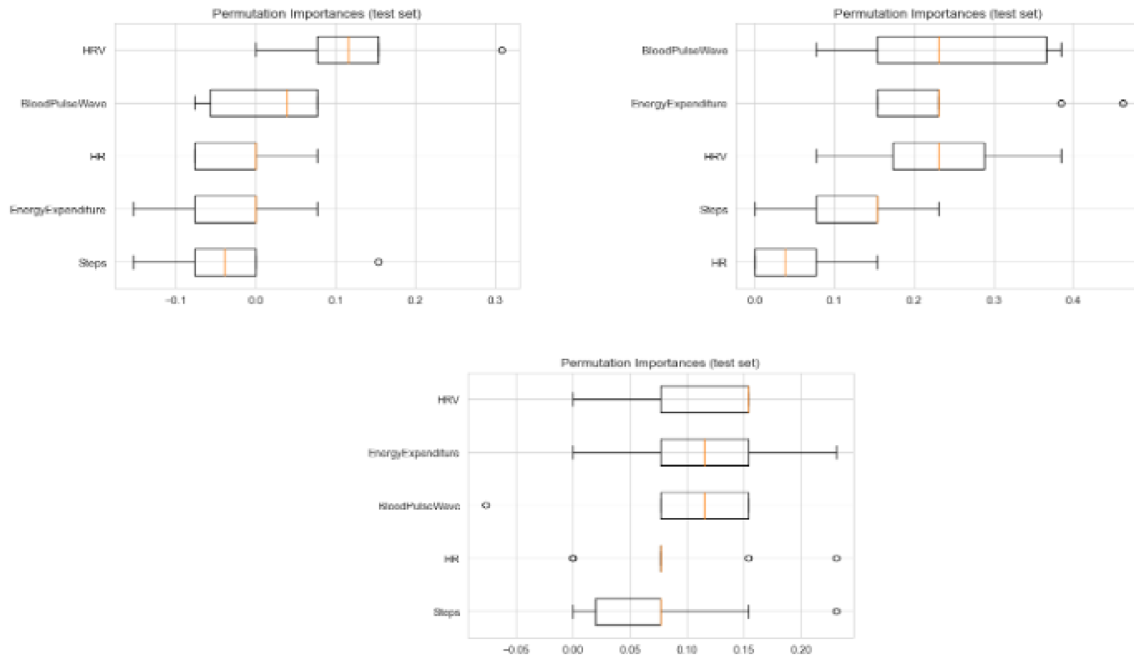

Figure 21: Permutation Importance measures the decrease in model score when a single feature value is randomly shuffled. This process breaks the relationship between the feature and the target, thus the drop in model score provides a measure of how much the model depends on the feature.

Having identified these key features, we proceeded to use them as predictors in models for Questions 2, 3, and 4. These questions addressed different aspects of the subjects' well-being, including physical and mental exhaustion and overall feeling compared to the previous day. We prepare for these models by removing Question 1 from the dataset and categorized the remaining data by each question. Our goal was to create individual models for each question using the same input features, thereby assessing how well these features could predict responses to similar questions.

Since the target variables for Questions 2, 3, and 4 were categorical, we created mappings for encoding the responses. For instance, for Question 2 (feeling better, worse, or the same), we encoded the responses as 0 (Worse), 1 (Same), and 2 (Better). For Questions 3 and 4, which assessed physical and mental exhaustion respectively, we used similar ordinal encodings, where 0 represented 'Never' and the highest value represented 'Always'.

We split the data into training and testing sets with an 80/20 split. The Random Forest Regression model revealed that physiological measures related to heart function (HR and HRV), physical activity levels (Steps and Energy Expenditure), and environmental factors (Barometer) were critical in assessing fatigue. The importance scores indicated that HRV and HR were consistently significant predictors across different models, suggesting their strong association with subjective well-being responses. Interestingly, the feature 'Steps' showed minimal predictive power, indicating its lower relevance in the context of

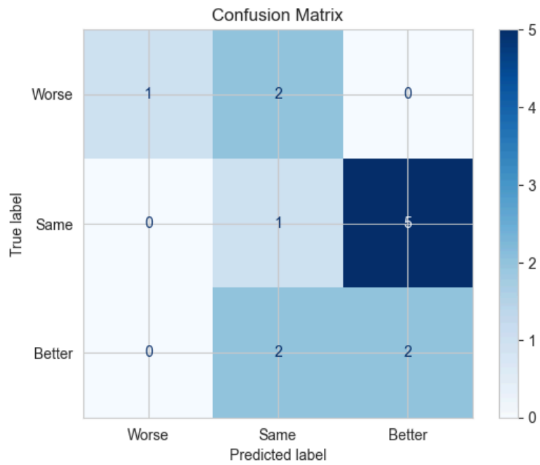

Figure 22: From the confusion matrices, it is clear that for Question 2, the model predicts the subject is doing better, when in reality they are feeling the same. There was also the case where the model predicted the subject should be feeling the same, when the participant was actually feeling better or worse. This can be due to other factors that were not included in the fatiguePROs dataset that affected the subject.

subjective well-being assessments. For our three models' predictive accuracy, we have:

- Question 2 (feeling better, worse, or the same): The model's accuracy was relatively low, highlighting the challenge in predicting overall mood changes based on objective measures.
- Question 3 (physical exhaustion): The model achieved an accuracy of approximately 77%, indicating a stronger relationship between physical exhaustion levels and the selected features.
- Question 4 (mental exhaustion): The model showed moderate predictive capability with an accuracy of around 61%, reflecting some degree of predictability for mental fatigue.

We further evaluate the models' performance, we analyzed the confusion matrices for each question:

For Figure 22, the confusion matrix of question 3 demonstrated better performance, with the model accurately predicting physical exhaustion levels. This suggests a strong relationship between the selected features and physical fatigue. For Figure 23, The confusion matrix of question 4 indicated moderate performance in predicting mental exhaustion, with some misclassifications. This showcases the complexity of predicting mental states based on physiological data alone.

Our findings underscore the importance of feature selection in predicting subjective well-being based on wearable data. While certain physiological metrics, such as HRV and HR, were reliable predictors of fatigue, others, like Steps, were less informative. These insights highlight both the potential and limitations of using wearable technology to monitor and predict various aspects of health-related quality of life (QoL).The model's low

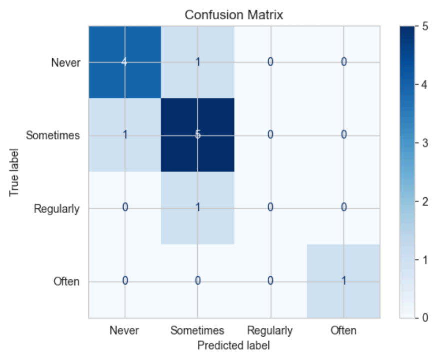

Figure 23: Confusion Matrix for Question 3

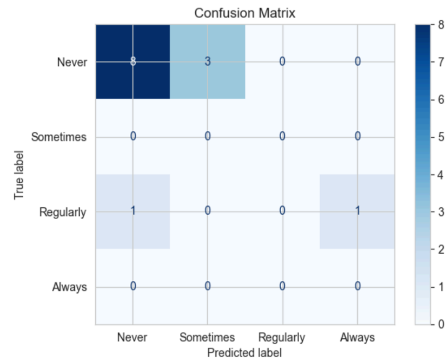

Figure 24: Confusion Matrix for Question 4

accuracy for Question 2, which asked whether the subject felt better, worse, or the same compared to the previous day, can be attributed to several factors. One key reason is the inherent subjectivity and complexity in self-assessing overall well-being. Unlike specific physical or mental exhaustion levels, overall well-being is influenced by a myriad of factors, including psychological, social, and environmental elements, which may not be fully captured by physiological metrics alone.

Additionally, individuals' perceptions of their well-being can fluctuate based on external circumstances and personal biases, making it challenging to predict accurately. The variability in the relative importance of features across different models also highlights the complexity of predicting subjective states. The underlying structure of the data and the nature of the questions significantly influence the predictive power of different features. For instance, HRV emerged as a consistently important feature, especially in Models 2 and 4, indicating its strong association with the target variables related to physical and mental exhaustion. In contrast, BloodPulseWave was particularly influential in Model 3 but less so in others, suggesting it might be more relevant for specific aspects of fatigue. Steps, on the other hand, appeared to be the least useful predictor, as indicated by its low or negative importance values across all models.

5.4 Finding Relationships Between Health Metrics and Fatigue Survey Data

The first, yet most important, step for this portion of the analysis was to aggregate the wearable health metric data so it became more meaningful to us. In its raw form, the metric data provides readings at the minute level for the different measures per subject as we have previously mentioned. Comparatively, the Fatigue PROs subjective survey data have one entry per day per question (questions about fatigue score on a scale of 1-10, fatigue compared to yesterday, mental exhaustion, and physical exhaustion). We determined that they best way the aggregate the wearable health metric data was to look at the measures both by hour and by day to be able to have a broader picture of each subject's metrics average metrics for the entire day as well as at each specific hour of the day. Additionally, the aggregation by hour also allowed us to see how a subject's metrics fluctuated throughout the day with less granularity than the raw data.

While aggregating the data, it was important to define new and redefine some old measures from the data. One measure that was completely omitted was the variable "HRV" or Heart Rate Variation. There was no information available about how the original HRV value was measured in the by minute data, so we decided that it would have no meaning or use to use it in our aggregation if we were unable to explain its meaning. In each of our aggregations (by hour and by day), we updated the Heart Rate, or HR, measure by including two new variables: avgHR and maxHR taking the average and maximum of the heartrate variables, respectively, over each aggregation interval. Similarly, over each aggregation interval, the variable Steps was redefined as sumSteps as the sum of steps in each time interval. A new measure "RHR," Resting Heart Rate, was created during aggregation as a means to construct a baseline for each aggregation time period that could be later used in calculations of variation. The variable was calculating by taking the average heart rate over each interval where the number of steps was equal to zero, or equivalently, when the subject was not moving or "at rest." FIGURES FOR SHOWING HOW THIS AGGREGATED DATA IS MORE HELPFUL THAN THE RAW DATA.

Because there are many different health metric variables, there are many different possible ways to explore the relationship between the wearable data and the self-reported fatigue survey responses. To start this portion of data exploration, we selected a few variables from the health metric data that we thought would intuitively impact levels of fatigue. The main two variables we identified were Heart Rate (and its variation) and number of Steps. The rationale behind this was that if a person was more stressed or mentally exhausted, they could potentially have a higher heart rate. Additionally, on days that a person takes a higher amount of steps, it is likely that either that day or the following day, they will have elevated levels of physical exhaustion also impacting their fatigue survey results. LINEAR REGRESSION INFORMATION + GRAPHS FROM POSTER WITH EXPLANATION

As briefly mentioned previously, another avenue for connecting the health metric data with subjective fatigue data is by looking at if today's health data will impact tomorrow's fatigue scores, or if yesterday's fatigue scores impact today's health data. This can more easily be referred to by lagging the data. + REGRESSIONS AND PLOTS RUN

The final component of our research blends a few approaches together: using both the idea of a person's "average day" and seeing if there is any pattern between deviation from the average day and level of fatigue, or similar for the next day's fatigue using lagging data again. A few different potential measures for average day were considered, but for this portion of the study, WHICH AGGRIGATION DID WE USE + GRAPHICS

6 Open Questions

6.1 Future Possibilities

While we ultimately did not quite form a predicative model for quality of life, our research still contains very valuable insights for a future team with more time, resources, and expertise to pick up and continue working. This section of the report will cover our aims as well as lingering questions and concerns.

One key area for future exploration is the enhancement of data collection and integration. Our study primarily relied on physiological data from wearable devices and subjective survey responses. Future research could benefit from incorporating additional data sources, such as psychological assessments, environmental conditions, and social factors. By integrating a wider range of data, it may be possible to capture a more comprehensive picture of an individual's quality of life and improve the accuracy of predictive models.

Future research can also delve into developing real-time monitoring and feedback systems, which represents another exciting possibility. With advancements in wearable technology and data processing, providing users with immediate insights into their health status is becoming increasingly feasible. Future work could explore implementing real-time analytics and feedback mechanisms that offer timely interventions and support. Such systems could help users proactively manage their health and respond to changes in their well-being more effectively.

Lastly, a significant challenge that can be tackled in predicting quality of life is the inherent subjectivity and variability in individuals' self-assessments. Future research could investigate methods to better account for these factors, such as incorporating personalized baselines and adaptive algorithms that adjust predictions based on individual differences. Additionally, exploring techniques for quantifying and mitigating the impact of external influences, such as stress or environmental changes, could enhance the robustness of predictive models. Specifically, considering the findings from our models like for the question 2 model: given the low accuracy for predicting whether subjects felt better, worse, or the same compared to the previous day, future models could benefit from including more contextual data, such as sleep quality, stress levels, and major life events, which might significantly influence overall well-being. Basically, the model can tell us what is the major factor from yesterday causing one's well-being today.

6.1.1 Predicative Model

Many of the ideals for a predicative model we have outlined in our prior report still hold true. As stated prior, the goal of this model would be for it to use vitals data (not unlike the sensors from the study we have focused on) to produce a quality of life score for a given patient. Integrated data alongside an objectively-driven quality of life metric would be the backbone for this model, and it would also aim to utilize machine learning and feature engineering to further the model. Techniques such as deep learning, ensemble learning, and reinforcement learning could be explored to handle the complexity and variability of QoL data. Also since not every dataset is perfect, implementing techniques to handle the imbalance, such as Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning, could improve model performance for categories with fewer data points.

For a future project, there are several important fundamental details we are keeping in mind. One of the most natural questions of the model will be how far out it can predict. One would assume that the further into the future the model can predict quality of life, the better, but that is not necessarily the case. If a model is tuned to predict far in the future off of a relatively thin slice of current vitals data, what happens if vital signs shift between the date the prediction is made and the target date? Any predictions made would become effectively worthless if the model is not designed to consider these variations (and considering every possible shift in vitals would lead to a muddled and unreadable prediction). Of course, a model tuned to predict the extreme immediate future would also not be of much practical use: predicting a patient's quality of life in the next day likely would be of little value.

Another important question will be the general applicability of the model. As mentioned, the data set focuses on healthy adults averaging around 40 years of age, and so the model will likely end up tuned to predict quality of life for healthy adults averaging around 40 years of age. This will be reflected in both the vital signs used to build the model and in the quality of life metric devised, and so generalizing it to a broader population will likely be a nontrivial task. The data set has no information on what age each particular subject is, so constructing estimations based on outlier participants is not an option. Ideally, in the long term, future studies of broader groups would allow a broader model to be created, but as is, understanding that this model will have a certain degree of inaccuracy for the broader population is important.

The actual output is also worth discussion. A prediction yielding a single quality of life value with no prescribed room for error is simply irresponsible, especially with the many aforementioned caveats of the data the model will be built from. Ideally, the model will predict a range of values for quality of life, and there would be several ways to accomplish this. Directly generating a variance for the prediction and using that to create a standard distribution of quality of life is a straightforward option that comes to mind, though that may be biased at the top and bottom of a hypothetical quality of life scale (depending on how the scale is defined). Another option could be developing a model that runs predictions several times, but with slight variance introduced in initial factors (either pre-defined or seeded with semi-random numbers), and producing a quartile distribution for the metric (an idea running in parity with aforementioned notions of building a model to handle the natural slight variation in human vitals). Ultimately, it is a balancing act of how precise to make the model versus how much the model can actually be applied in a real situation, and discovering that balance is the key to a model that is both statistically rigorous and practically useful.

6.1.2 Questions of Quality of Life

This phase of the project opened up several new possible ways to look at a quality of life metric. Clearly, a single measurement is not enough to accurately predict it, and figuring out a compounded measurement is an extraordinarily challenging task. This is further convoluted by the fact that our studies have shown that people tend to be very poor at reading themselves and at accurately knowing their own quality of life; there is a very large overhead in this problem in the form of human error and opinion. Individual humans all have different baselines for their health and wellness, and so a "one-size fits all" approach to a model will almost certainly either not work at all or run into extreme issues with accuracy and precision.

This is not all to say that a metric for quality of life does not or cannot exist. Rather, it is entirely possible that this metric would have to be tailor-made for each individual this model would aim to track, likely based off of similar statistical inquiries to our study. Perhaps then that is what a future study should seek to solidify: not a universal quality of life metric and predictive model, but a process to find someone's personal quality of life metric and predicative model.

Though we largely developed our ideas on quality of life through the lens of heart rate and steps taken, there are almost certainly other factors one could use as predictive signs, both among those taken that we have overlooked, and among factors that were not measured. Environmental temperature, for example, likely has some manner of impact on one's perceived quality of life. In general, future research would benefit from examination of more environmental factors than what this dataset covered.

6.2 Conclusion

Our studies have laid a strong foundation for future work with quality of life and for creating more fully-featured predictive models. Though we encountered many challenges and eventually ran into issues of scope, our work is no less important and foundational for tackling a question as broad and challenging as that of one's quality of life.

References

[1] M. Post, "Definitions of quality of life: What has happened and how to move on," *Topics in Spinal Cord Injury Rehabilitation*, vol. 20, no. 3, pp. 167–180, 2014.

[2] S. C. Government of Canada. (2022) Quality of life hub. [Online]. Available: www160.statcan.gc.ca/index-eng.htm

[3] V. Vijayan, "Review of wearable devices and data collection considerations for connected health," *Sensors (Basel)*, vol. 21, no. 16, pp. 55–89, 2021.