Assessment of Cardiovascular Fitness Using R Analytics

John Sokol

Stockton University

Department of Natural Sciences & Mathematics

5 December 2017

Introduction:

As society becomes increasingly health conscious, people are taking a greater responsibility for their own health. Millenials are spearheading this health and wellness revolution by implementing smoking abstinence and exercise into their daily lives [1]. The CDC has stated that the national smoking rate in the United States is the lowest it has ever been at 16.8% of adult citizens in 2014, compared to 42.4% of adults in 1965 [1]. According to Global Web Index, almost 4 out of 5 Millenials exercise at least once a week [2]. Although this young adult cohort in the United States embraces a healthy lifestyle, it is difficult to assess level of fitness in an objective manner. However, by employing statistical analysis of 4 key data metrics, the exercising individual may be able to obtain an accurate depiction of their current fitness status.

The objective of this study is to determine the proficiency level of cardiovascular fitness using several metrics that are quantifiable with wearable technology in the two month span of 21 September 2017 to 21 November 2017. The analyzed metrics are values associated with optimal cardiovascular health according to referenced literature and 6 years of personal knowledge accrued participating in Varsity Track and Field at Barnegat High School and NCAA collegiate level competition during undergraduate study at Stockton University. By comparing this data to the standards obtained from the literature, a consensus may be reached regarding overall fitness status using comparisons to standard and personally chosen data values based on fitness goals.

Jayson Resch, Stockton University's Track and Field coach, would frequently refer to the three major influences of fitness level: Exercise/training workload, dietary habits, and sleep. In this study, 2 out of the 3 fitness influences are examined with exercise and sleep. Exercise data organized by week, gym attendance frequency, and the amount of night sleep should affect the 4 health metrics being assessed in this study over a two month span. Prior to 21st September 2017, I ran occasionally, but no more than 5 - 7 miles a week. I did visit my local gym, Retro Fitness, frequently at 3 to 4 times a week, and my amount of nightly sleep was seldom less than 7 hours or greater than 8 hours. After 21st September 2017, I began a training plan that alternated high mileage weeks (10+ miles), then implemented low mileage recovery weeks, while simultanously keeping 3 to 4 visit weekly gym attendence and my aforementioned sleep patterns consistent. The data metrics analyzed are resting heart rate, heart rate variability, Submaximal (SM) VO$_2$ max, and burned active calories. The data was obtained using the Apple Watch Series 2, which exports all of these data points to the 'health' application on the iPhone 6S, which tabulates the data and exports to an Excel file for RStudio analysis.

The hypothesis is my personal fitness data attributes are above average for resting heart rate, heart rate variability, and SM VO$_2$ max, and they will improve over time, indicating my cardiovascular fitness has improved over the two month span. As a result, the data should reveal favorable results suggesting an interpretation of the data that I am healthy from a cardiovascular perspective. If the data does not support expectations, I would evaluate the third influencer of fitness, dietary habits, to explain unmet expectations.

Resting Heart Rate (RHR):
Resting heart rate is defined as the beats per minute (BPM) of an individual that is not sleeping and not subjected to environmental stress. The lowest recorded resting heart rate for a given day is a reliable metric of cardiovascular fitness, as there is little variability that can influence the data. Individuals who exercise

and experience increased cardiovasuclar fitness generally have *lower* RHR measurements than sedenatary individuals Most people fall into the range of 50 to 90 BPM [3].

According to Jensen et al. "Resting heart rate in the range 51 – 80 BPM was associated with about a 40–50% increase in cardiovasuclar moratality risk, a resting heart rate in the range 81 – 90 bpm conferred a twofold increase in risk, and resting heart rates above 90 BPM risk conferred a threefold increase in risk compared to subjects in the lowest heart rate category ($<= 50$ BPM)". Additionally, Jensen states there is a significant correlation between VO_2 max and resting heart rate, as subjects with higher levels of physical fitness are more likely to have lower heart rates [3]. Therefore, I have established 50 BPM as my own benchmark for physical fitness status of 'excellent'.

Heart Rate Variability:
Heart rate variability (HRV) is the measurement of time variation between heartbeats. Simply put, HRV measures how effeiciency your heart can change it's heart rate based on the activity. Individuals who exercise and have excellent cardiovasuclar fitness generally have *higher* HRV measurements than sedenatary individuals. Therefore, this metric is worth analyzing to possiblity observe a relationship between an increased fitness level and higher HRV. Due to significant variability between individuals from day to day or even hour to hour, there are no standards for this health metric [4].

Sub-Maximal VO$_2$ Max:
Sub-Maximal (SM) VO_2 max is the definitive metric of aerobic/cardiovascular fitness for non-olympic athletes that do not have the labratory capabilities of a full-fledged Maximal VO_2 max test. This test measures the amount of oxygen your lungs can breathe from the air and deliver to the muscles under the stress of 85% maximum exercise effort. Heart rate data is compared to the exercise/ workload completed to determine a SM VO_2 max value. A higher SM VO_2 max indicates a higher level of cardiovascular fitness. The unit for VO_2 max is mL/(kg $\cdot$ min) [5].

The Apple Watch Series 2, which was used for SM VO_2 max measurements in this study, does not measure Maximal VO_2 max values. The primary limitation of the Apple Watch is the device has no way to measure real time air consumption. As a result, the Apple Watch relies upon heart rate data, age, weight, gender, and duration of exercise to determine a predicted VO_2 max value using the following formula [5]:

$$SMVO_2Max = 132.853 - (0.0769 * Ibs) - (0.3877 * age) + (6.315 * gender) - (3.2649 * min) - (0.156 * HR)$$

The numeric value for gender is male: 1 or female: 0. Time is equal to the total time elapsed in the workout (min). HR is the highest consistent heart rate value for the workout (BPM).

Burned Daily Calories:
The total amount of calories burned daily is a promising statistic, as it may rationalize an increase or decrease of cardiovascular fitness with appropriate observations over a given period of time. An increased trend of calories burned may indicate an increased workload, suggesting increased cardiovascular fitness. Similar to heart rate variability, there is no literature benchmark for calories burned, as individual calorie burning goals are different and highly specific.

Figure 1: Relationship Between Cardiovascular Fitness & Health Metrics
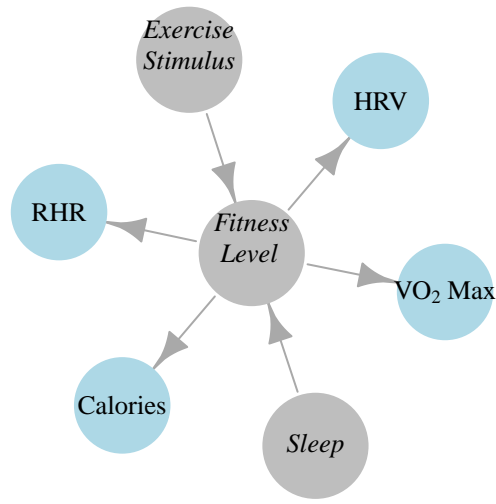

Illustration of study attributes and its relationships. Exercise stimulus, or lack thereof, influences fitness level, which is subsequently assessed by resting heart rate (RHR), Heart rate variability (HRV), sleep, VO_2 Max, and calories burned.


Import & Munge Data:

As previously mentioned, the data, already in Tidy format, was collected via Apple Watch Series 2 and stored in the 'health' app on an iPhone 6S. From there, data was exported to an Excel file and imported into RStudio using the command below:

```r
library(readxl)  # imports Excel file
library(knitr)   # kable function

DE_Cardio <- read_excel("~/Desktop/Data_Exploration_Heart_Project.xlsx", skip = 5)
```

Subsequently, the $ operator assigned each vector based column to separate variables. NA values were removed by calling [!is.na()], then reassigning the NA free vectors to the initial variables.

Influence of Exercise & Sleep on Cardiovasuclar Fitness

Figure 2: Weekly Mileage & Gym Attendance
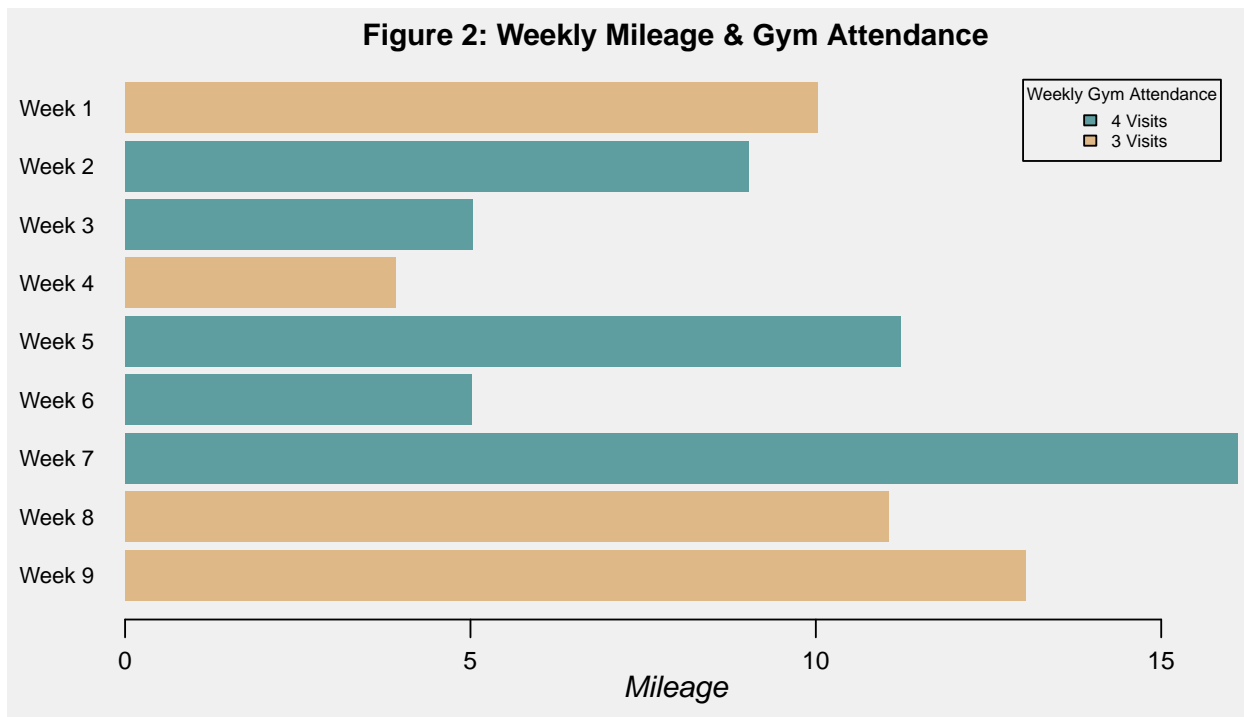
The first step to analyze health metrics is to observe the exercise stimulus influencing fitness level. Figure 2 illustrates the amount of miles ran per week. The training consists of alterating high mileage weeks (10+ miles) with a week or two of recovery (5 miles or less). This training strategy is designed to slowly increase fitness while minimizing chance of injury attributed to excessive miles ran in a small amount of time, which is called overloading. Additionally, weight training at Retro Fitness complements cardiovascular training while strengthing muscles to prevent injury. Each week starts on a Thursday and ends on a Wednesday. For example, week 1 began on Thursday, 9/21/2017 and ended on Wednesday, 9/27/2017. The exception to this rule was week 9, which started on Thursday, 11/16/2017 but ended on Wednesday, 11/21/2017. So the final week of the study was 6 days instead of 7 days.

Figure 3: Amount of Nightly Sleep
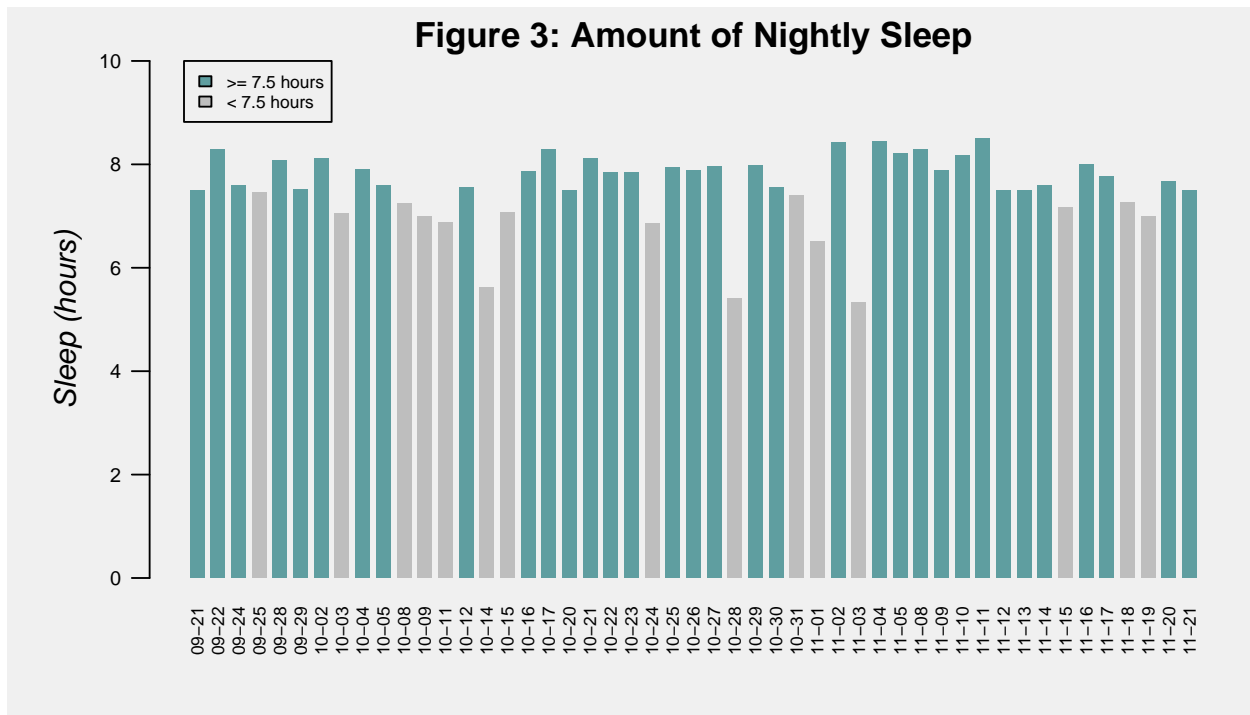
Adequate amount of sleep on a consistent basis is extremely important for optimum athletic performance and physical fitness. Sleep facilitates numerous biological functions in the body such as carbohydrate metabolism, food breakdown, and protein synthesis, all of which directly influence cardiovascular performance for better or worse [6]. This consequently affects the body's ability to adapt to cardiovascular stimuli and increase fitness. Therefore, sleep is analyzed to see if sleep habits agree with fitness changes over the two month period. A positive correlation should exist between greater amounts of nightly sleep and increased cardiovascular fitness, as well as lower amounts of nightly sleep and worsened cardiovascular fitness.

Every athlete's sleep requirements are different. However, from personal experience, my minimum sleep benchmark for optimal athletic performance is 7.5 hours of sleep per night. The frequency of nightly sleep greater than or equal to 7.5 hours and sleep variability is observed in this study. This data will be used to determine if sleep was adequate and consistent enough to affect my cardiovascular fitness level.

My personal nightly sleep duration goal is 7.5 hours to optimize my fitness level. Some people need more, while others need less. To put this value into perspective, 5 Time NFL Super Bowl Champion Tom Brady (QB) of the New England Patriots consistently gets 8.5 - 9 hours of sleep a night [8]. However, my fellow teammate during undergraduate NCAA Track and Field competition only slept for an average of 6 hours per night, and he consistently performed at a high level.

```
length(which(sleep_hrs >= 7.5)) / 49 # nights which sleep >= 7.5 hrs: 34
```

[1] 0.6938776

There is a 69% frequency rate of sleep duration greater than or equal to 7.5 hours. This is a promising statistic. However, we want to observe the distribution of the sleep dataset to ensure the absence of excess outliers.

```
var(sleep_hrs) %>% sqrt() # calculates standard deviation
```

[1] 0.7067476

The standard deviation of ~ 0.71 is close to 0, which indicates the sleep dataset does not have excessive variability. As a result, it is determined that the sleep dataset is consistent and reliable, indicating sleep should help me maximize exericse/training benefits to increase cardiovascular fitness.

Visualization of Cardiovascular Fitness:

Table 1: Personal Statistics: 9/21/2017 - 11/21/2017

Statistics	Min	1st Quartile	Median	Average	3rd Quartile	Max	Std Deviation
Resting Heart Rate (BPM)	44.00	49.250	51.0	51.110	53.000	58.00	2.54
Heart Rate Variability (ms)	19.00	45.750	57.5	63.920	81.500	137.00	27.04
SM VO2 Max ((mL/(kg · min))	49.71	50.410	51.0	50.800	51.120	51.81	0.56
Active Calories (kcal)	92.40	337.500	443.0	477.500	635.000	909.00	184.97
Sleep (hrs)	5.33	7.267	7.6	7.541	7.983	8.50	0.71

RStudio was used to munge the data collection. The summary() function displayed the basic statistics of the data, while standard deviation was computed manually. The tables were formatted using the Knitr libary with the included kable function.
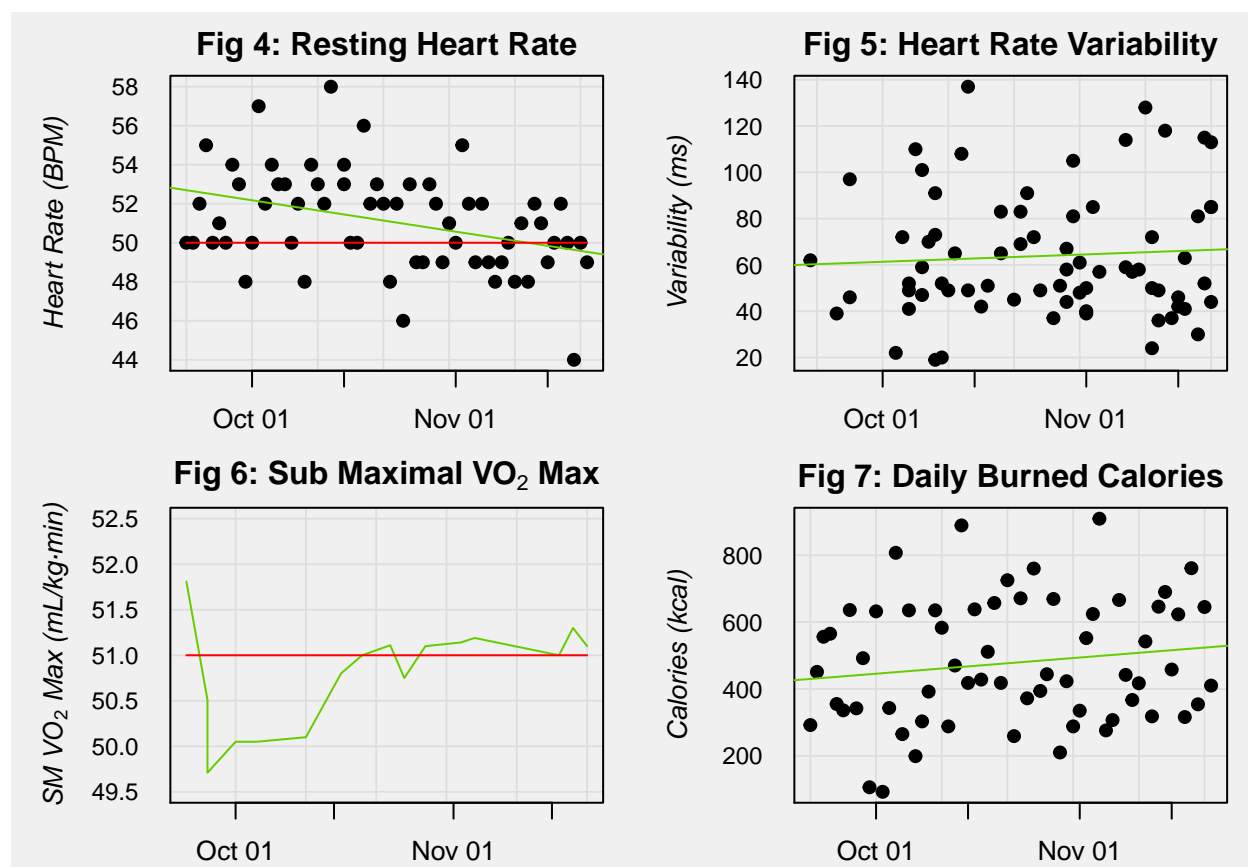
Red lines denote literature values for ideal fitness level; green lines denote personal data trends.

Assessment of Cardiovascular Fitness:

Resting Heart Rate (Figure 4):
According to the green linear regression line, my resting heart rate experienced a steady decline over the two month period. Linear regression began at approximately 53 BPM and lowered to even less than 50 BPM, the benchmark value for a healthy heart according to literature references. To quantify this data, the following R commands were employed:

```
bpm_less_equal_50 <- length(which(rhr_bpm <= 50))        # 29 times RHR <= 50
bpm_percentage <- bpm_less_equal_50 / length(rhr_bpm)    # frequency of RHR <= 50: 29/62
df_bpm <- data.frame(bpm_less_equal_50,length(rhr_bpm),bpm_percentage)
kable(df_bpm, caption = "RHR Quantification: 9/21/2017 - 11/21/2017",
col.names = c("Total RHR: <= 50 BPM", "Total Recorded RHR", "Percent Total RHR: <= 50"))
```

Table 2: RHR Quantification: 9/21/2017 - 11/21/2017

Total RHR: <= 50 BPM	Total Recorded RHR	Percent Total RHR: <= 50
29	62	0.4677419

Approximately 47% of total RHR values were less than or equal to 50 BPM. That is fairly close to half of the recorded data. However, this statistic does not tell us how the data points change over the two month period. For the RHR trend to be analyzed, the first and second half of the recording period of 9/21/2017 to 10/21/2017 and 10/21/2017 to 11/21/2017, respectively, was observed:

```
first_start <- as.Date("2017-09-21")        # start of first half
half_start <- as.Date("2017-10-21")         # end of first half

rhr_frame_frst <- data.frame(rhr_date,rhr_bpm)  # categorizes RHR: 09/21/2017-10/21/2017
rhr_frame_frst <- subset(rhr_frame_frst,rhr_frame_frst$rhr_date <= half_start)
rhr_subset_frst <- length(which(rhr_frame_frst$rhr_bpm <= 50))
subset_percentage_frst <- rhr_subset_frst / length(rhr_frame_frst$rhr_bpm) # (10/31) ~ 32%

df_sub_frst <- data.frame(rhr_subset_frst,length(rhr_frame_frst$rhr_bpm),subset_percentage_frst)
kable(df_sub_frst,caption = "RHR Quantification: 09/21/2017 - 10/21/2017",
col.names = c("RHR: <= 50 BPM:", "Recorded RHR","Percent RHR: <= 50:"))
```

Table 3: RHR Quantification: 09/21/2017 - 10/21/2017

RHR: <= 50 BPM:	Recorded RHR	Percent RHR: <= 50:
10	31	0.3225806

Table 4: RHR Quantification: 10/21/2017 - 11/21/2017

RHR: <= 50 BPM:	Recorded RHR	Percent RHR: <= 50:
19	32	0.59375

32% of the resting heart rate values in the first half of the study were <= 50. In contrast, 59% of the resting heart rate values less than or equal to 50 were recorded during the second half (10/21/2017 - 11/21/2017). This trend accounts for an 84% increase of RHR <= 50 frequency. Additionally, the correlation coefficient is -0.37, suggesting a negatively linear relationship. This data indicates my resting heart rate experienced a decrease during this time period in comparison to the first half of the data collection period, a relationship that agrees with the green linear regression line illustrated in figure 4. To put an RHR of less than 50 BPM into context, Usain Bolt has recorded an average resting heart rate of 33 BPM [9].

Heart Rate Variability (Figure 5):
This metric is difficult to analyze due to such high standard deviation. Compared with the average HRV of 62.44 ms, the standard deviation is 26.60, indicating that 68% of the data is between plus or minus 26.60 ms of 62.44 ms, a high variation. Figure 5 illustrates this point, as the green linear regression line is nearly

horizontal with a slope of 0.10. To quantify this further, the correlation coefficient is 0.071. Heart rate variability can be influenced by several environmental factors that attribute to this variability. Although health journals cite HRV as an excellent metric of cardiovascular health status, the two month data collection window is not enough time for variability to rescind from the dataset. Consequently, HRV is not taken into consideration for health analysis [4].

SM VO$_2$ Max (Figure 6):
The concept of VO$_2$ max is likely the most obscure metric out of the 4 assessed in this project; this data point must be put into context by comparison. Norms, or standards of VO$_2$ max is based on empirical evidence that is generalized in Golding et al, The Y's Way to Physical Exercise (1986). These values can be used as a comparative tool to get a qualitative gauge of fitness level. However, these literature values are for maximal VO$_2$ max, not sub-maximal. Consequently, this comparison should be taken with a grain of salt, only to be used as generalized gauge of fitness level. Surreptitiously, to ascertain a general gauge of fitness is the objective of this project. Standards of Maximal VO$_2$ max levels of various training levels cateogrized by gender and age are the following [7]:

```
kable(vo2data_frame_male,caption = "Male VO~2~ Max Norms (mL/(kg·min)):",
col.names = c("Performance","Age 13 - 19","20 - 29","30 - 39","40 - 49","50 - 59","60 +"))
```

Table 5: Male VO$_2$ Max Norms (mL/(kg · min)):

Performance	Age 13 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 +
Superior	> 55.9	>52.4	> 49.4	> 48.0	> 45.3	> 44.2
Excellent	51.0 - 55.9	46.5 - 52.4	45.0 - 49.4	43.8 - 48	41.0 - 45.3	36.5 - 44.2
Good	45.2 - 50.9	42.5 - 46.4	41.0 - 44.9	39.0 - 43.7	35.8- 40.9	32.3 - 36.4
Average	38.4 - 45.1	36.5 - 42.4	35.5 - 40.9	33.6 - 38.9	31.0 - 35.7	26.1 - 32.2
Poor	35.0 - 38.3	33.0 - 36.4	31.5 - 35.4	30.2 - 33.5	26.1 - 30.9	20.5 - 26.0
Very Poor	< 35	< 33.0	< 31.5	< 30.2	< 26.1	< 20.5

Table 6: Female VO$_2$ Max Norms (mL/(kg · min)):

Performance	Age 13 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 +
Superior	> 41.9	> 41.0	> 40.0	> 36.9	> 35.7	> 31.4
Excellent	39.0 - 41.9	37.0 - 41.0	35.7 - 41.0	32.9 - 36.9	31.5 - 35.7	30.3 - 31.4
Good	35.0 - 38.9	33.0 - 36.9	33.0 - 36.9	29.0 - 32.8	27.0 - 31.4	24.5 - 30.2
Average	31.0 - 34.9	29.0 - 32.9	29.0 - 32.9	24.5 - 28.9	22.8 - 26.9	20.2 - 24.4
Poor	25.0 - 30.9	23.6 - 28.9	23.6 - 28.9	21.0 - 24.4	20.2 - 22.7	17.5 - 20.1
Very Poor	< 25.0	< 23.6	< 22.8	< 21.0	< 20.2	< 17.5

The green line plot in figure 6 illustrates a sharp decrease then gradual increase of SM VO$_2$ max. From the end of September to the middle of October,SM VO$_2$ max increased in value from 50.05 to 51.11 mL/(kg · min), a jump likely attributed to positive physiological response to the training stimulus. The highest recorded SM VO$_2$ max since data collection began on 21 September 2017 is 51.3 mL/(kg · min). According to these table values, a male age 23 with a 51.3 VO2 max would have a qualitative assessment of 'Excellent'. The 51.3 data point is SM VO$_2$ max, so my maximal VO$_2$ max is likely higher than 51.3, so there is a high level of confidence with this assessment. To put this number in persective, Lance Armstrong, the former American road cyclist, recorded a maximal VO$_2$ max of 84.0. To reiterate, the SM Vo$_2$ max value must be interpreted lightly, but relative to the two month data collection period, the data suggests an increase in aerobic capacity.

Daily Burned Calories (Figure 7):
The amount of daily calories burned is telling of aerobic effort exerted on any given day. Theoretically, increased cardiovascular activity (running, weight lifting), increases active calories burned, thereby improving

8

cardiovascular fitness. These aspects are interrelated, so burned calories was collected to observe the presence of this relationship. According to the calories dataset, there was a minor increase in regards to the green linear regression line for daily calories burned; the correlation coefficient is 0.152. This trend agrees with the theoretical rationale mentioned previously. However, a larger dataset should be used to ensure that the increased trend is correlated with improved fitness.

Project Shortcomings:
The obvious shortcoming of this project is the small data collection duration of two months (9/21/2017 to 11/21/2017). If a larger time period was available to collect data, the end result may have been different regarding the trends of the data. The choosen metrics can also be put into question. Sub-Maximal VO_2 max is not equivalent to Maximal VO_2 max, so comparison of the collected SM VO_2 and the literature Maximal VO_2 is not ideal. Additionally, according to numerous resources, heart rate variability is an excellent indiactor of health, but the variability of HRV made analysis not feasible. There was not enough time to implement statistical tests, as those would strengthen the conclusions made about the data. Finally, I would like to investigate a quantitative value measuring aerobic fitness that takes into consideration several reliable health metrics, and normalizes a score from 0 to 100. 0 being a couch potato and 100 being a professional triathlete.

Conclusion:

Four metrics of cardiovascular health between 21 September 2017 to 21 November 2017 were successfully imported from Excel and munged with R for data analysis. Resting heart rate proved to be a reliable data point, and the data suggests an overall RHR decrease over the two month period. Heart rate variability was an unreliable statistic that had a high standard deviation compared with the average, and a linear regression slope of 0.10, indicating no relationship is present. SM VO_2 max experienced a slight increase of 49.71 to 51.3 over the two months, suggesting aerobic capacity improved. The increase of total amount of daily burned calories agrees with increased overall cardiovascular workload over these two months. Finally, sleep patterns remained fairly consistent notwithstanding three days of sleep < 6 hours. Coalescing the conclusions of these metrics, a statement of confidence can be made that my cardiovasuclar fitness has increased between 21 September 2017 and 21 November 2017. However, more research is necessary to determine a concrete assessment of fitness.

References:

[1]. Trends in Current Cigarette Smoking Among High School Students and Adults, United States, 1965–2014. (2016, March 30). Retrieved November 19, 2017, from https://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm

[2]. Valentine, O. (2017, August 09). 76% of Millennials Exercise At Least Once a Week. Retrieved November 19, 2017, from http://blog.globalwebindex.net/chart-of-the-day/76-percent-of-millennials-exercise-at-least-once-a-week/

[3]. Jensen et al. (2013, June 15). Elevated resting heart rate, physical fitness and all-cause mortality: a 16-year follow-up in the Copenhagen Male Study. Retrieved November 19, 2017, from http://heart.bmj.com/content/99/12/882

[4]. Stauss, H. M. (2003, November 01). Heart rate variability. Retrieved November 19, 2017, from http://ajpregu.physiology.org/content/285/5/R927.short

[5]. Golding et al, The Y's Way to Physical Exercise (1986). Retrieved November 19, 2017.

[6]. Hanson, S., PhD. (n.d.). Sleep and Athletes. Retrieved November 19, 2017, from https://www.gssiweb.org/en/sports-science-exchange/Article/sse-167-sleep-and-athletes

[7]. VO2 Max Table of Values. (n.d.). Retrieved November 19, 2017, from https://www.runandtricoach.com/articles/vo2-max-table

[8]. Sleep habits of the world's most elite athletes: Tom Brady. (2017, September 11). Retrieved November 19, 2017, from https://www.fatiguescience.com/blog/sleephabitsoftombrady/

[9]. Murugappan, R. (2016, November 26). Your resting heart. Retrieved November 30, 2017, from http://www.star2.com/living/viewpoints/2016/11/27/your-resting-heart/

Appendix: R Code

```r
library(dplyr)    # primarily for the pipe (%>%) operator
library(knitr)    # kable function
library(igraph)   # network graphs

# creates network graph of 2 gray and 5 blue circles,
# adds edges (connecting lines) and configures auto layout
g <- make_empty_graph()
g <- g %>% add_vertices(2, color = "gray", size = 72) %>%
  add_vertices(5, color = "lightblue", size=50)
g <- g %>% add_edges(c(2,1))
g <- g %>% add_edges(c(1,3,1,4, 1,5, 1,6, 1,7))
coords <- layout.auto

# establishes fonts, line width, arrow sizes,
plot(g, vertex.label= c(expression(italic("Fitness\n Level")),
expression(italic("Exercise\nStimulus")),"RHR","HRV",expression("VO"[2]*" Max"),
"Calories","Sleep"),layout = coords, vertex.frame.color = NA,
vertex.label.cex = 0.78,edge.width = 1,margin = -0.15, vertex.label.color = "black",
edge.arrow.width = 1, edge.arrow.size = 1)

# caption and center placement
title("Figure 1: Relationship Between Cardiovascular Fitness & Health Metrics",
cex.main = 0.95, adj = 0.7)

# mileage and gym barplot
mileage_dates <- DE_Cardio$`Mileage Total (Weeks)`
mileage_dates <- mileage_dates[!is.na(mileage_dates)]

mileage_totals <- DE_Cardio$`Mileage Total (Mileage)`
mileage_totals <- mileage_totals[!is.na(mileage_totals)]

par(bg="#F0F0F0", mar=c(2.8,3,1.5,0.5), mgp=c(1.5,0.6,-0.05))
barplot(rev(mileage_totals), names.arg = rev(mileage_dates), las = 1, cex.axis = 0.8,
cex.names = 0.75, border = NA, space = 0.15, xlab = expression(italic("Mileage")),
main = "Figure 2: Total Weekly Mileage & Gym Attendance",horiz = TRUE,
col = c(rep("#DEB887",2), rep("#5F9EA0",3),"#DEB887",rep("#5F9EA0",2),"#DEB887"))
category <- c("4 Visits","3 Visits")          # legend config
color <- c("#5F9EA0","#DEB887")
legend(13,10.4, category,cex = 0.57,fill = color, title = "Weekly Gym Attendance")

# imports Excel file
library(readxl)
```

```r
DE_Cardio <- read_excel("~/Desktop/Data_Exploration_Heart_Project.xlsx", skip = 5)

# assigns dates of RHR to rhr_date; removes NA values; converts to class 'Date'
rhr_date <- DE_Cardio$`RHR Date (Day/Month)`
rhr_date <- rhr_date[!is.na(rhr_date)] %>% as.Date()


rhr_bpm <- DE_Cardio$`Resting Heart Rate (BPM)`    # assigns RHR values
rhr_bpm <- rhr_bpm[!is.na(rhr_bpm)]                # removes NA values
rhr_line <- c(rep(50,57))                          # vector for red line in RHR plot
bpm_less_equal_50 <- length(which(rhr_bpm <= 50))  # BPM <= 50
rhr_sd <- var(rhr_bpm) %>% sqrt()                  # RHR std deviation

# assigns dates of HRV to hrv_date; removes NA values; converts to class 'Date'
hrv_date <- DE_Cardio$`HRV Date (Day/Month)`
hrv_date <- hrv_date[!is.na(hrv_date)] %>% as.Date()


hrv_ms <- DE_Cardio$`Heart Rate Variability (ms)`  # assigns HRV values
hrv_ms <- hrv_ms[!is.na(hrv_ms)]                   # removes NA values
hrv_sd <- var(hrv_ms) %>% sqrt()                   # HRV standard deviation

# assigns dates of VO2 to vo2_date; removes NA values; converts to class 'Date'
vo2_date <- DE_Cardio$`Vo2 Max Date (Day/Month)`
vo2_date <- vo2_date[!is.na(vo2_date)] %>% as.Date()


vo2_mL <- DE_Cardio$`Vo2 Max (mL/Kg min)`          # assigns VO2 values
vo2_mL <- vo2_mL[!is.na(vo2_mL)]                   # removes NA values

vo2_line <- c(rep(51.0,16))                        # vector for red line VO2 line graph
vo2_sd <- var(vo2_mL) %>% sqrt()                   # VO2 std deviation

# assigns dates of Active Calories to rhr_date; removes NA values; converts to class 'Date'
act_cal_date <- DE_Cardio$`Kilocalorie Date (Day/Month)`
act_cal_date <- act_cal_date[!is.na(act_cal_date)] %>% as.Date()

act_cal_kilo <- DE_Cardio$`Kilocalorie Daily Total (kcal)`  # assigns active calorie values
act_cal_kilo <- act_cal_kilo[!is.na(act_cal_kilo)]          # removes NA values
act_cal_kilo_sd <- var(act_cal_kilo) %>% sqrt()             # active calorie std deviation

# assigns dates of sleep to sleep_date; removes NA values; converts to class 'Date'
sleep_date <- DE_Cardio$`Sleep Date (Day/Month)`
sleep_date <- sleep_date[!is.na(sleep_date)]
sleep_date_format <- format(sleep_date, format = "%m-%d")

sleep_hrs <- DE_Cardio$`Amount of Nightly Sleep (Hours)`  # assigns sleep values
sleep_hrs <- sleep_hrs[!is.na(sleep_hrs)]                 # removes NA values
sleep_sd <- var(sleep_hrs) %>% sqrt()                    # sleep std deviation

# basic statistics values from summary() function; std deviation added manually
statistics <- c("Resting Heart Rate (BPM)","Heart Rate Variability (ms)",
"SM VO2 Max ((mL/(kg·min))","Active Calories (kcal)", "Sleep (hrs)")
min_ <- c(44.00, 19.00, 49.71,92.40,5.33)
first_qu <- c(49.25, 45.75,50.41,337.50,7.267)
median_ <- c(51.00, 57.50,51.00,443.00,7.60)
mean_ <- c(51.11,63.92, 50.80,477.50,7.541)
third_qu <- c(53.00, 81.50, 51.12,635.00,7.983)
```

```r
max_ <- c(58.00,137.00, 51.81,909.00,8.50)
std_dev <- c(rhr_sd,hrv_sd,vo2_sd,act_cal_kilo_sd, sleep_sd) %>% round(2)

# creates data frame of above data; tabulates using kable function
stats_df <- data.frame(statistics,min_,first_qu,median_,mean_,third_qu,max_,std_dev)
kable(stats_df,caption = "Personal Statistics: 9/21/2017 - 11/21/2017",
col.names = c("Statistics","Min","1st Quartile","Median","Average","3rd Quartile","Max",
"Std Deviation"))

# creates a 2 X 2 vector for 4 graphs
# establishes size parameters, horizontal y label, and gray background
par(mfrow=c(2,2), mar=c(2.3,5,2,1), las=1, bg="#F0F0F0")

# resting heart rate
plot(rhr_date,rhr_bpm, xlab = "",
ylab = expression(italic("Heart Rate (BPM)")),        # y axis label
main = "Fig 2: Daily Lowest Heart Rate", pch=19,      # figure title; black datapoint dots
yaxt="n",panel.first=grid(NULL, NULL,                 # background grid
col="#DEDEDE", lty="solid", lwd=0.9))                 # removes y axis; replaced by axis()
axis(2, tick=FALSE, cex.axis=0.9, asp=1/2)            # green linear regression line
abline(lm(rhr_bpm ~ rhr_date), col="#66CD00")         # reestablishes y axis; no tick marks
lines(rhr_date,rhr_line,col = "red")                  # red benchmark line


# heart rate variability
plot(hrv_date,hrv_ms , xlab = "",
ylab = expression(italic("Variability (ms)")),
main = "Fig 3: Heart Rate Variability", pch=19, yaxt="n",
panel.first=grid(NULL, NULL, col="#DEDEDE", lty="solid", lwd=0.9))
axis(2, tick=FALSE, cex.axis=0.9, asp=1/2)
abline(lm(hrv_ms ~ hrv_date), col="#66CD00")


# VO2 max
plot(vo2_date, vo2_mL, xlab = "",
ylab = expression(italic("SM VO"[2]*" Max (mL/kg·min)")),
main = expression(bold("Fig 4: Sub Maximal VO"[2]*" Max")),
ylim = c(49.5,52.5),las=1,
type="n", yaxt="n", panel.first=grid(NULL, NULL,
col="#DEDEDE", lty="solid", lwd=0.9))
axis(2, tick=FALSE, cex.axis=0.9, asp=1/2)
lines(vo2_date,vo2_mL, col="#66CD00")          # green trend line
lines(vo2_date,vo2_line,col="red")             # benchmark red line


# active calories
plot(act_cal_date,act_cal_kilo, xlab = "",
ylab = expression(italic("Calories (kcal)")),
yaxt="n", main = "Fig 5: Daily Burned Calories", pch=19,
panel.first=grid(NULL, NULL, col="#DEDEDE", lty="solid", lwd=0.9))
axis(2, tick=FALSE, cex.axis=0.9, asp=1/2)
abline(lm(act_cal_kilo ~ act_cal_date), col="#66CD00")

# Sleep barplot; formats x and y axis, font sizes, label orientations
# for loop determines color (#5F9EA0) when sleep observation >= 7.5 hours
fill_colors <- c()
for (i in 1:length(sleep_date_format)) {
```

```r
  if (sleep_hrs[i] >= 7.5){
    fill_colors <- c(fill_colors, "#5F9EA0")
  } else {
    fill_colors <- c(fill_colors, "gray")
  }
}
par(bg="#F0F0F0", mar=c(4,4,1.5,0.5), mgp=c(2.5,0.8,0))
barplot(sleep_hrs,names.arg = sleep_date_format,las = 2, cex.axis = 0.7,
cex.names = 0.6, ylim = c(0,10), border = NA, space = 0.4,
xlab = expression(italic("Calender Progression: September 2017 - November 2017")),
ylab = "", main = "Figure 6: Recorded Amount of Nightly Sleep",
col = fill_colors)

title(ylab = expression(italic("Sleep (hours)")),mgp = c(1.7,1,0)) # main barpot title
category <- c(">= 7.5 hours","< 7.5 hours")                        # legend config
color <- c("#5F9EA0","gray")
legend(0,10, category,cex = 0.57,fill = color)
```

```r
bpm_less_equal_50 <- length(which(rhr_bpm <= 50))        # 29 times RHR <= 50
bpm_percentage <- bpm_less_equal_50 / length(rhr_bpm)    # frequency of RHR <= 50: 29/62
df_bpm <- data.frame(bpm_less_equal_50,length(rhr_bpm),bpm_percentage)
kable(df_bpm, caption = "RHR Quantification: 9/21/2017 - 11/21/2017",
col.names = c("Total RHR: <= 50 BPM", "Total Recorded RHR", "Percent Total RHR: <= 50"))
```

```r
first_start <- as.Date("2017-09-21")              # start of first half
half_start <- as.Date("2017-10-21")               # end of first half

rhr_frame_frst <- data.frame(rhr_date,rhr_bpm)  # categorizes RHR: 09/21/2017-10/21/2017
rhr_frame_frst <- subset(rhr_frame_frst,rhr_frame_frst$rhr_date <= half_start)
rhr_subset_frst <- length(which(rhr_frame_frst$rhr_bpm <= 50))
subset_percentage_frst <- rhr_subset_frst / length(rhr_frame_frst$rhr_bpm) # (10/31) ~ 32%

df_sub_frst <- data.frame(rhr_subset_frst,length(rhr_frame_frst$rhr_bpm),
subset_percentage_frst)
kable(df_sub_frst,caption = "RHR Quantification: 09/21/2017 - 10/21/2017",
col.names = c("RHR: <= 50 BPM:", "Recorded RHR","Percent RHR: <= 50:"))
```

```r
# male VO2 max data frame and kable
age13_19 <- c("> 55.9", " 51.0 - 55.9", "45.2 - 50.9"," 38.4 - 45.1","35.0 - 38.3","< 35")
age20_29 <- c(">52.4","46.5 - 52.4","42.5 - 46.4","36.5 - 42.4","33.0 - 36.4","< 33.0")
age30_39 <- c("> 49.4","45.0 - 49.4","41.0 - 44.9","35.5 - 40.9","31.5 - 35.4","< 31.5")
age40_49 <- c("> 48.0","43.8 - 48","39.0 - 43.7","33.6 - 38.9"," 30.2 - 33.5","< 30.2")
age50_59 <- c("> 45.3","41.0 - 45.3", "35.8- 40.9","31.0 - 35.7","26.1 - 30.9","< 26.1")
age_60 <- c("> 44.2","36.5 - 44.2","32.3 - 36.4","26.1 - 32.2","20.5 - 26.0","< 20.5")
Performance <- c("Superior","Excellent","Good","Average","Poor","Very Poor")
vo2data_frame_male <- data.frame(Performance,age13_19,age20_29,age30_39,age40_49,
age50_59,age_60)
```

```r
library(knitr)
kable(vo2data_frame_male,caption = "Male VO~2~ Max Norms (mL/(kg·min)):",
col.names = c("Performance","Age 13 - 19","20 - 29","30 - 39","40 - 49","50 - 59","60 +"))
# Same kable format used for the female table

# female VO2 max data frame and kable
agef13_19 <- c("> 41.9", "39.0 - 41.9", "35.0 - 38.9"," 31.0 - 34.9","25.0 - 30.9","< 25.0")
```

```r
agef20_29 <- c("> 41.0","37.0 - 41.0","33.0 - 36.9","29.0 - 32.9","23.6 - 28.9","< 23.6")
agef30_39 <- c("> 40.0","35.7 - 41.0","33.0 - 36.9","29.0 - 32.9","23.6 - 28.9","< 22.8")
agef40_49 <- c("> 36.9","32.9 - 36.9","29.0 - 32.8","24.5 - 28.9", "21.0 - 24.4","< 21.0")
agef50_59 <- c("> 35.7","31.5 - 35.7","27.0 - 31.4","22.8 - 26.9","20.2 - 22.7","< 20.2")
agef_60 <- c("> 31.4","30.3 - 31.4","24.5 - 30.2","20.2 - 24.4","17.5 - 20.1","< 17.5")
Performance <- c("Superior","Excellent","Good","Average","Poor","Very Poor")
vo2data_frame_female <- data.frame(Performance,agef13_19,agef20_29,agef30_39,agef40_49,
agef50_59,agef_60)

kable(vo2data_frame_female,caption = "Female VO~2~ Max Norms (mL/(kg·min)):",
col.names = c("Performance","Age 13 - 19","20 - 29","30 - 39","40 - 49","50 - 59", "60 +"))
```

```r
length(which(sleep_hrs >= 7.5)) / 49 # nights which sleep >= 7.5 hrs: 34
```

```r
var(sleep_hrs) %>% sqrt() # calculates sleep std deviation
```