

Capstone Project Report: Quantified Self

INTRODUCTION

The Quantified Self movement, coined by *Wired's* Gary Wolf and Kevin Kelly in 2007, incorporates wearable consumer technology in order to monitor personal metrics such as heart rate, sleep patterns, and physical activity. The hope is those that monitor personal metrics will be inspired to improve their metrics by improving their health habits.

For six years, I've used various FitBit devices and become more active. Several months ago, I started using a FitBit Flex 2 and began tracking my sleep as well as collecting more granular data about my activity.

With this data, I will answer the following questions:

- Are there relationships between my own activity, sleep patterns, and sleep quality?
- What are my sleep and activity patterns?
- What steps can I take to improve my health patterns?

DATA

FitBit allows data exports of up to 31 days. I exported data for sleep during October, November, December, January, and February. I also exported data for sleep during the same months. I then had 10 files:

- fitbit_activities_october_2017.csv
- fitbit_activities_november_2017.csv
- fitbit_activities_december_2017.csv
- fitbit_activities_january_2018.csv
- fitbit_activities_february_2018.csv
- fitbit_sleep_october_2017.csv
- fitbit_sleep_november_2017.csv
- fitbit_sleep_december_2017.csv
- fitbit_sleep_january_2018.csv
- fitbit_sleep_february_2018.csv

Each of the “activities” files contain the following ten variables:

1. Date
2. Calories burned
3. Steps
4. Distance
5. Floors
6. Minutes sedentary
7. Minutes lightly active
8. Minutes fairly active
9. Minutes very active
10. Activity calories

Each of the “sleep” files contain the following nine variables:

1. Start time
2. End time

3. Minutes asleep
4. Minutes awake
5. Number of awakenings
6. Time in bed
7. Minutes REM sleep
8. Minutes light sleep
9. Minutes deep sleep

DATA WRANGLING

The first step I took in wrangling my data was to cut down on the number of files I was working with. I combined all the activity data into an activity data set and all the sleep data into a sleep data set.

I then renamed each of the columns in the two data sets to remove spaces and capital letters.

The data was not properly sorted in chronological order, so each set needed to be sorted according to date and sleep start time.

Several columns in each of the data sets contained missing values as my particular FitBit device is not capable of recording certain metrics. I removed the following columns with missing values from the data sets:

1. Floors
2. Minutes REM sleep
3. Minutes light sleep
4. Minutes deep sleep

I discovered the number of rows, observations, in each of my data sets differed. In order to cleanly combine the activity and sleep data sets, I needed to determine why the number of observations differed. Upon examining the data, I discovered daytime naps were recorded independent of overnight sleep and decided to eliminate daytime naps from the data set by removing any “minutes asleep” observations less than 240.

I also discovered a day in October on which no sleep was recorded because I opted to charge my device overnight rather than track my sleep. In order to replace those missing values, I went into the csv file and manually added sleep values for that night by calculating the average of each variable for the night before and the night after the night of missing values.

Once I ensured the number of rows in each data set was equal and that the data sets were sorted in the same order, I combined the activity and sleep data sets.

I then added a weekday column so I could eventually explore trends in my activity and sleep data according to the day of the week.

I was also interested in plotting with a histogram my sleep start and sleep end hours, so I created columns for each of those values.

DATA LIMITATIONS

Before performing analysis on my data and drawing conclusions from that analysis, I needed to acknowledge the limitations of the data.

My sleep data can be misleading in that wearable devices measure sleep according to movement and lying still is recorded as sleep when, in reality, sleep may not be occurring. In addition, my device does not measure sleep stages - REM vs. light vs deep sleep - as other devices do.

Because my device can only detect arm movement recorded by the accelerometer, it does not record other activities such as stationary bike rides, strength training, or yoga, which all can affect calories burned as well as sleep.

This data set also does not account for additional lifestyle factors that affect sleep and activity such as travel, stress, illness, and nutrition.

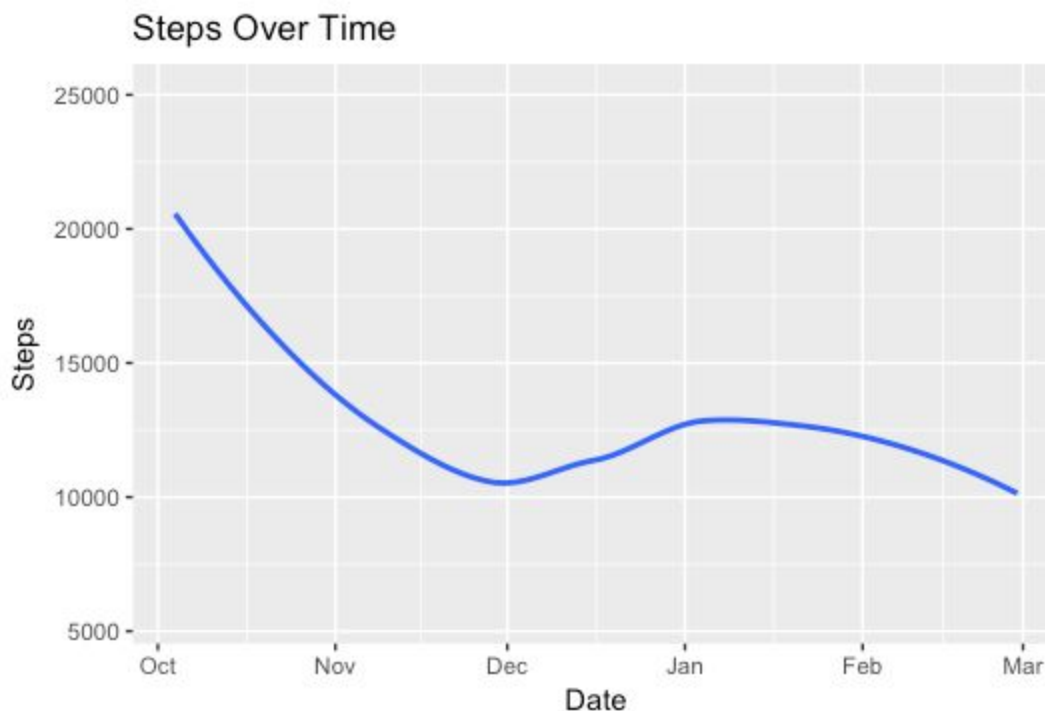
Because of these limitations, I might consider using other devices in the future with more precision to record sleep such as the SleepScore Max Sleep Improvement System. I might also consider recording other factors with a limited number of values such as a 1-5 stress scale or noting with a binary value when I'm sleeping away from home.

ANALYSIS OF ACTIVITY DATA

The mean number of daily steps during the time period (October - February) studied is 12,855, which is 29% above the commonly cited 10,000 steps a day recommendation.

Steps for this time period range from 2,018 on Saturday, December 2, 2017 to 60,681 on Sunday, November 5, 2017, the day of New York City Marathon.

I plotted the steps over this time period with a smooth line graph and observe steps decline from October through November, pick back up in December and decline again from January through February, in line with my race training patterns.

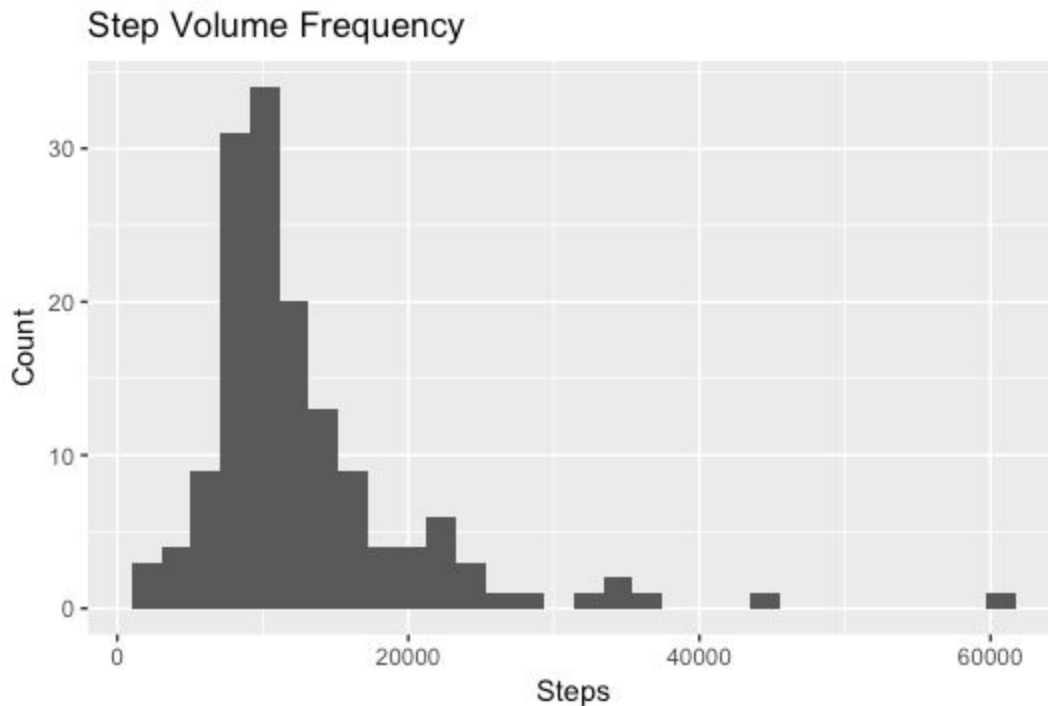


Sunday has the highest number of mean steps for the time period studied. However, taking over 60,000 steps on Sunday, November 5 created an outlier and skewed the mean steps on Sundays. Because of that, I looked at the median steps for each day of the week. The day of the week with highest median steps is Wednesday, which makes sense as during the time period studied, I usually participated in a group run on Wednesday mornings.

day_of_week mean_steps		day_of_week median_steps	
<chr>	<dbl>	<chr>	<dbl>
1 Monday	9160	1 Monday	9621
2 Tuesday	10181	2 Tuesday	9643
3 Friday	11441	3 Friday	9826
4 Wednesday	13652	4 Saturday	11223
5 Thursday	13656	5 Sunday	13054
6 Saturday	15247	6 Thursday	13463
7 Sunday	16608	7 Wednesday	14978

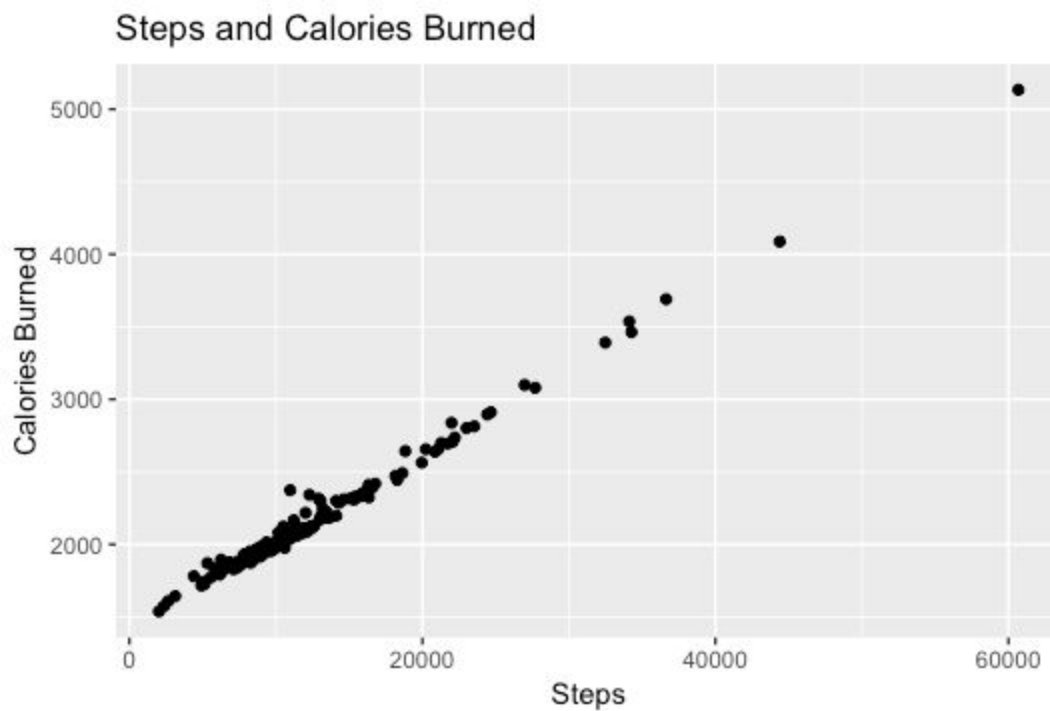
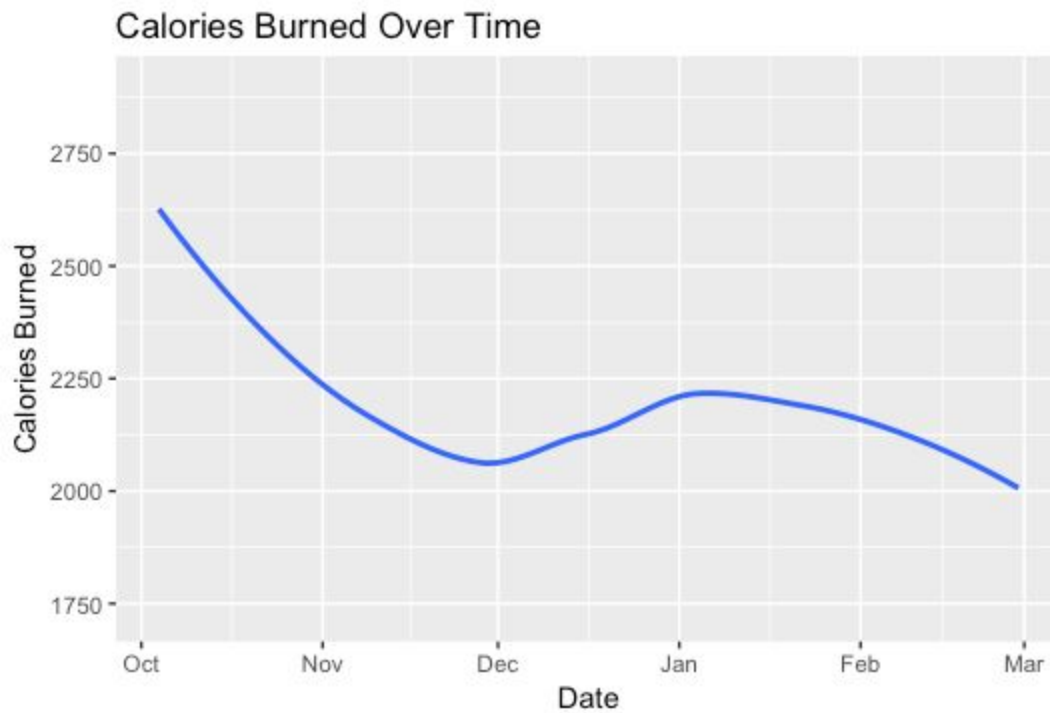
Monday had the lowest number of mean steps, 9,160, as well as the lowest number of median steps, 9,621, for the time period studied. This makes sense as Mondays sometimes follow running races on Sundays and Mondays are spent recovering and staying off my feet. In addition, I will often take a strength training class at the gym after work on Mondays, physical activity that is unrecorded in the step data. With all this in mind, I might think about making an effort to take walks on Monday mornings before work when I'm feeling up to it.

While there was a high volume of steps in a small number of days during the 120-day time period studied, I took between 9,000 - 11,000 steps on most days.



The mean number of estimated calories burned during this time period is 2,192. Calories burned for this time period range from 1,538 on Saturday, December 2, 2017 to 5,134 on Sunday, November 5, 2017, the day of New York City Marathon.

Calories burned follows a similar trend to steps. And these two values are directly related as Fitbit estimates calories burned based on user weight and recorded activity.

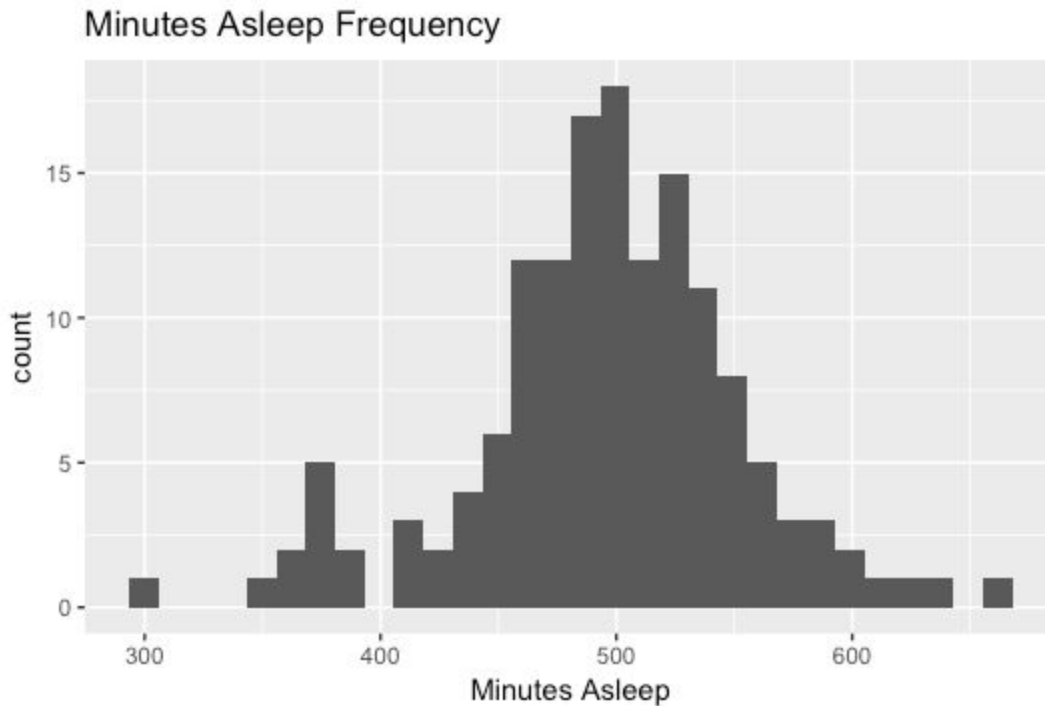


ANALYSIS OF SLEEP DATA

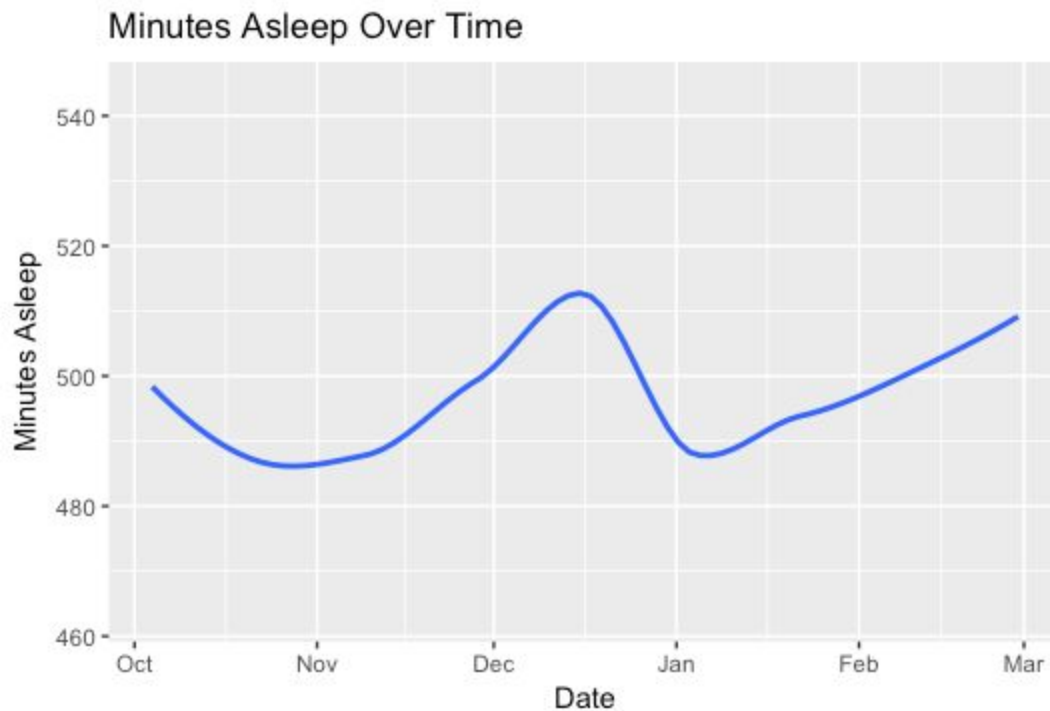
The mean number of time asleep during the time period (October - February) studied is 8 hours, 16 minutes, 10% above the commonly cited recommended 7-8 hours of sleep per night recommendation. Since my device is worn on my wrist and perceives time laying completely still as sleep, my actual sleep time may, in reality, be more within the recommended 7-8 hours per night. However, I still might consider that the amount of time I spend sleeping is slightly in excess.

Time asleep for this time period ranges from 4 hours, 54 minutes on Thursday, November 9, 2017, the night before an early flight to start a vacation, to 10 hours, 56 minutes on Saturday, December 30, 2017, a day on which an early nap bled into sleep time.

A histogram of minutes asleep is almost normally distributed with most nights of sleep consisting of around 500 minutes (8 hours, 20 minutes).



I plotted the minutes asleep over this time period with a smooth line graph and observe minutes asleep peaks mid-December and then declines until early January, when it begins to climb through the remainder of January and February.



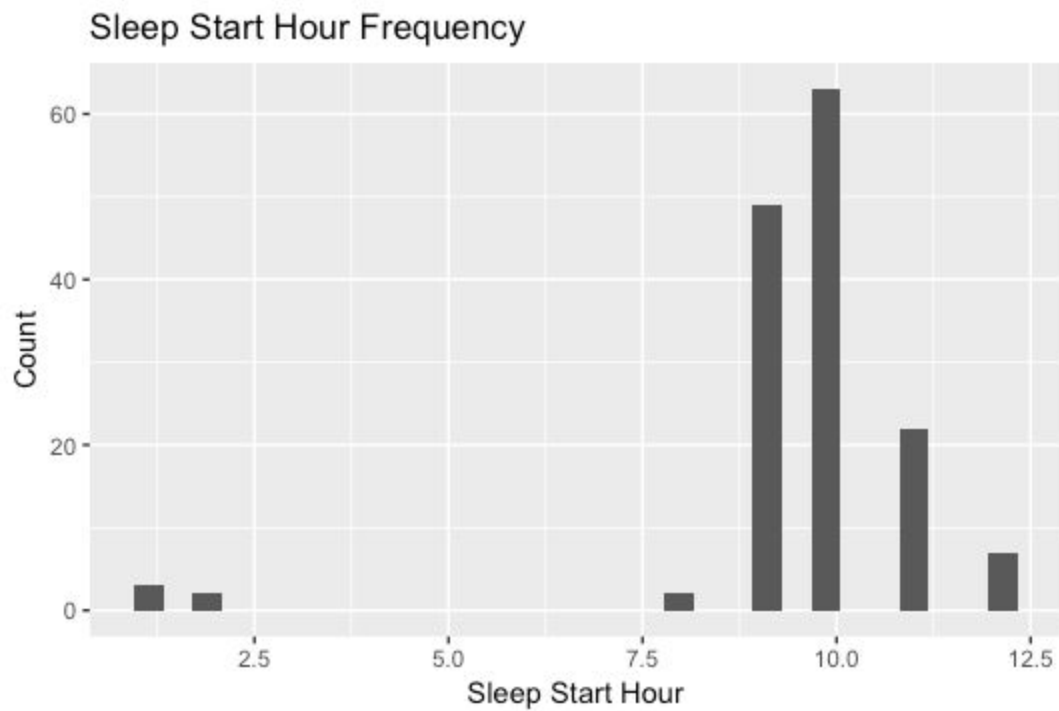
Saturday has the highest number of mean minutes asleep, 517, for the time period studied. This makes sense as I typically sleep in on Sunday mornings. Tuesday has the lowest number of mean minutes asleep, 470. During October through December, I participated in group runs at 6:30 am on Wednesday mornings and the early wake-up time likely interfered with the sleep I might record if I was not participating in a 6:30 am activity. The mean amount of sleep occurring on Tuesdays, however, is still 7 hours, 50 minutes and well within the daily recommended 7-8 hours.

day_of_week	minutes_asleep
<chr>	<dbl>
1 Tuesday	470
2 Thursday	479
3 Wednesday	488
4 Sunday	500
5 Monday	504
6 Friday	515
7 Saturday	517

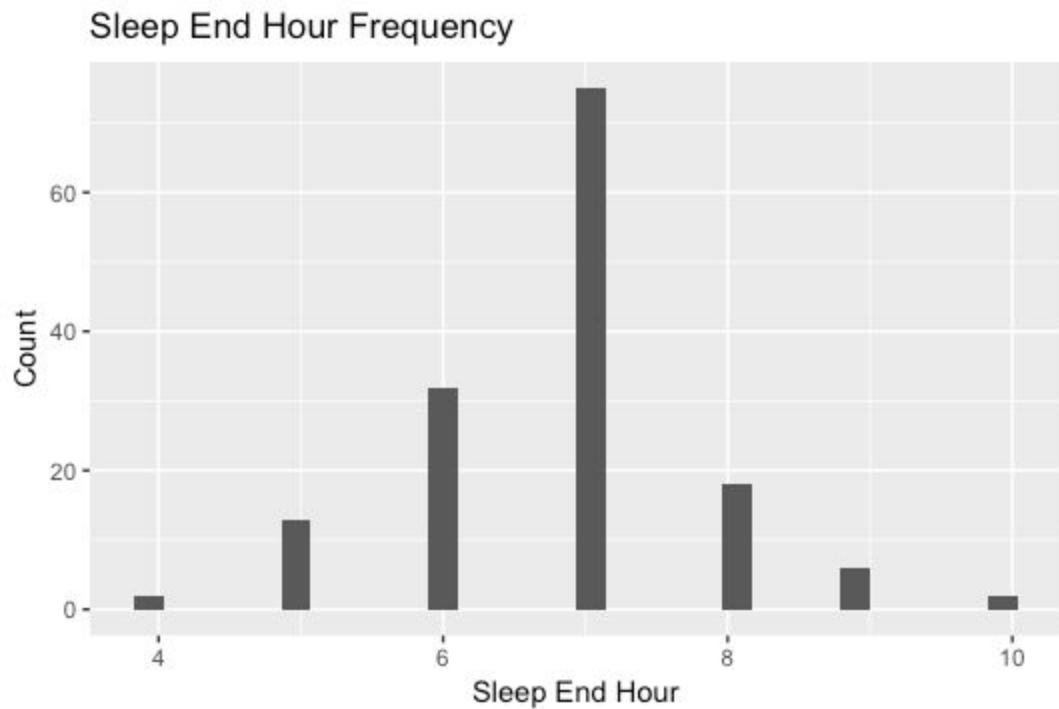
According to the mean sleep start hour and mean sleep end hour, I typically fall asleep in the 9pm hour and wake up in the 6pm hour. This fluctuates according to the day of the week. Tuesdays have the earliest mean sleep end hour as well as the earliest sleep start hour.

day_of_week	sleep_start_hour	day_of_week	sleep_end_hour
<chr>	<dbl>	<chr>	<dbl>
1 Tuesday	9.29	1 Tuesday	6.24
2 Sunday	9.33	2 Thursday	6.57
3 Wednesday	9.50	3 Wednesday	6.73
4 Monday	9.52	4 Sunday	6.86
5 Saturday	9.57	5 Friday	6.95
6 Friday	9.86	6 Monday	6.95
7 Thursday	10.1	7 Saturday	7.38

A histogram of sleep start hour indicates sleep start most frequently begins before 10pm.

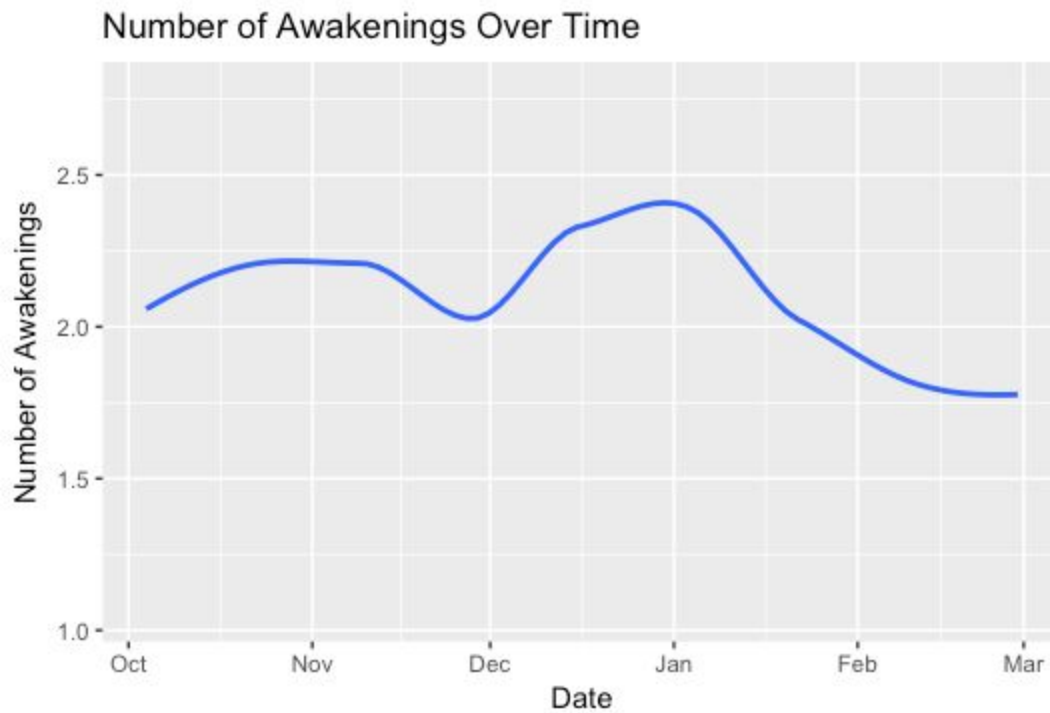


A histogram of sleep end hour is almost normally distributed around the 7am hour.

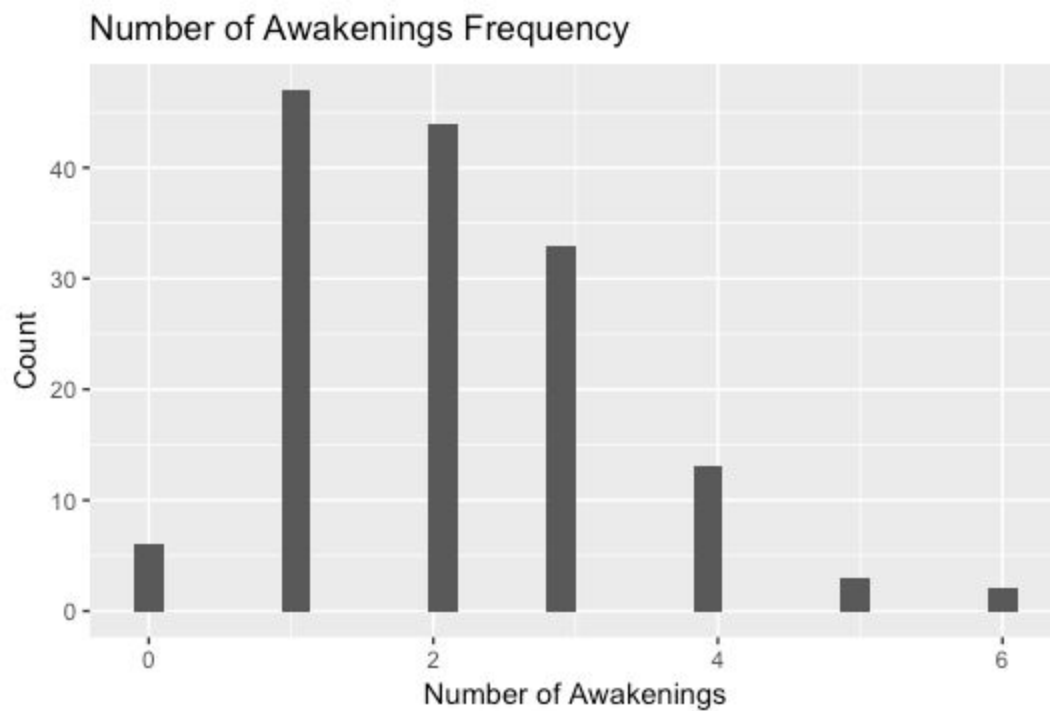


While my device does not explicitly record sleep quality, I correlate a lower number of awakenings with a higher sleep quality. During the time period studied, I averaged 2.1 times awake per night.

I plotted number of awakenings over time with a smooth line graph and observe number of awakenings peaks in late December and has declined since.



During the time period studied, I woke up between 0 and 6 times per night, but 1-2 times most frequently.



ANALYSIS OF ACTIVITY + SLEEP DATA

In order to determine what relationships, if any, exist between my activity and sleep data, I made a correlation matrix.

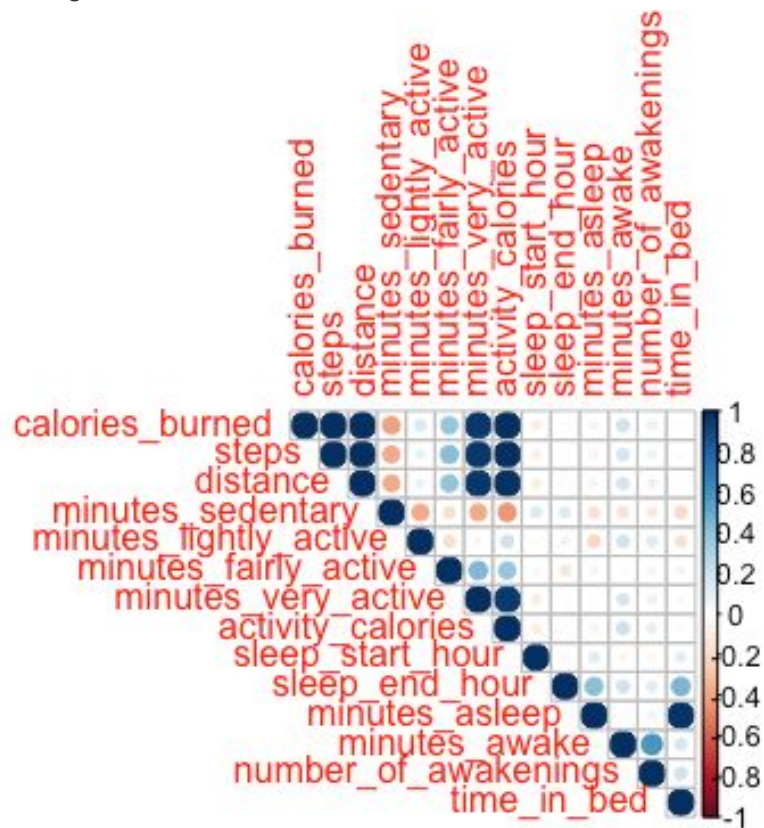
Not surprisingly, calories burned, steps, minutes fairly active, minutes very active, and activity calories are highly correlated. In addition, minutes sedentary are negatively correlated to calories burned, steps, distance, all types of active minutes, and activity calories.

Some other obvious high correlations exist between number of awakenings and minutes awake, time in bed and sleep end hour, as well as minutes asleep and time in bed.

Some light correlations exist between minutes awake during sleep hours and calories burned, steps, distance, minutes lightly active, minutes very active, and activity calories. These correlations, however, are not strong enough to draw any sound conclusions.

There does not appear to be any significant correlation between steps and time asleep. In other words, high levels of activity do not necessarily cause me to sleep any more or less.

There does not appear to be any significant correlation between steps and number of awakenings or minutes very active and number of awakenings.



KEY TAKEAWAYS

- On average, I take close to 13,000 steps a day, but my step habits are inconsistent and fluctuate according to the day of the week as well as the time of the year
- On average, I sleep 8 hours and 16 minutes a night, but time asleep also fluctuates slightly according to the day of the week as well as the time of the year
- I will typically fall asleep between 9:30 pm and 10:00 pm, awaken two times per night, and finish sleeping between 6:30 am and 7:00 am
- There does not appear to be any significant correlations between my activity and sleep habits

RECOMMENDATIONS

- Strive for slightly more steps on Mondays, Tuesdays, and Fridays for more consistent step volume throughout the week
- Strive for steps in small spurts throughout the workday when step counts tend to drop
- Use the Fitbit app to strive for an hourly goal of 250 steps per hour during waking hours
- Establish a consistent bedtime, ideally around 10:30 pm (since sleep end time is usually between 6:00 am and 7:30 am)
- Take steps to avoid sleeping more than the recommended 7-8 hours per night
- Note energy levels when sleep is reduced
 - Am I sleeping 8+ hours per night because I need to or because I allow myself to do so?
 - Can I sleep less, have more waking hours, and not feel fatigued?
- Continue to track the variables studied to observe changes in habits during one calendar year