**Case Study 2：Bellabeat – trend discovery in smart fitness devices usage**

**Liew Sharon**

***Business Tasks***

Analyze how consumers are using their smart fitness devices.

***Data used***

The data used in this project is the FitBit Fitness Tracker Data, which is a public dataset on Kaggle. The dataset contains personal fitness tracker from thirty Fitbit users within a month with their consent.

The dataset included:

Daily Activity Data : included total steps, total distance, total calories, intensities within a day.

Hourly Activity Data : included total steps, total distance, total calories, intensities within an hour.

Minutely Activity Data：included total steps, total distance, total calories, intensities within a minute.

Others：included heart rate, sleep time and weight.

***Roadmap***

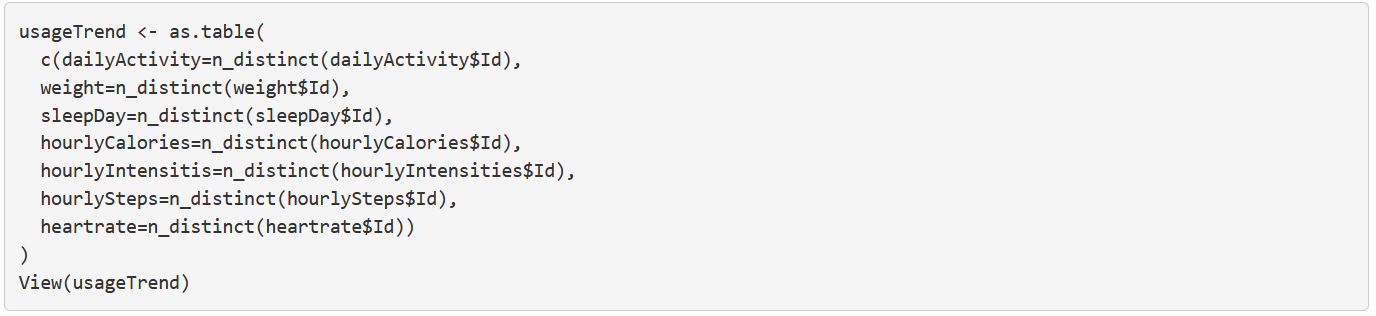
*Prepare*

1. Check and import the data

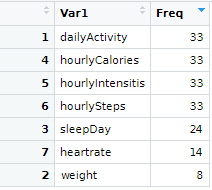
As the dataset is too large, I checked every dataset by *str()* function in R. I choose daily and hourly activity data, heart rate, sleep time and weight data to analyze how users using their smart devices. I use read\_csv() function to import the data.

1. Preliminary trend analysis

I checked the distinct user ID in each dataset, which can represent the popularity of each functions in smart devices.



From the results, the popularity of the functions can be sorted by following orders: **activity > sleep day > heart rate > weight**. The results showed a trend that most of the users choose a smart devices due to its activity recorded functions. It also comes to a new question: what kinds of user will use different functions?



1. Data Cleaning

After checking the summary of each dataset, all information needed is organized and completed.

*Process*

1. Prediction

After checking the dataset, I made some predictions for the usage trends:

H1：User who are trying to lose weight will have higher daily activity intensity.

H2: User who are having more daily activities will have a better sleep quality.

H3: User who are having more daily activities will tend to use different functions.

1. Selection

There are lots of information in dataset. Consider the information value, I choose to analyze the trends by following aspects:

Activity: Use *TotalSteps*, *TotalDistances* and *Calories* to represent the overall activity status of users.

Intensity: Use *VeryActiveMinutes* and *SedentaryMinutes* to represent the detailed intensity status of users.

Sleep Quality: Use *TotalMinutesAsleep* and *SleepOnsetLatency*(subtraction between in-bed duration and asleep duration) to represent the sleep quality of users.

Weight: Consider the differences in gender, age and height among users, use *BMI* to represent users’ weight.

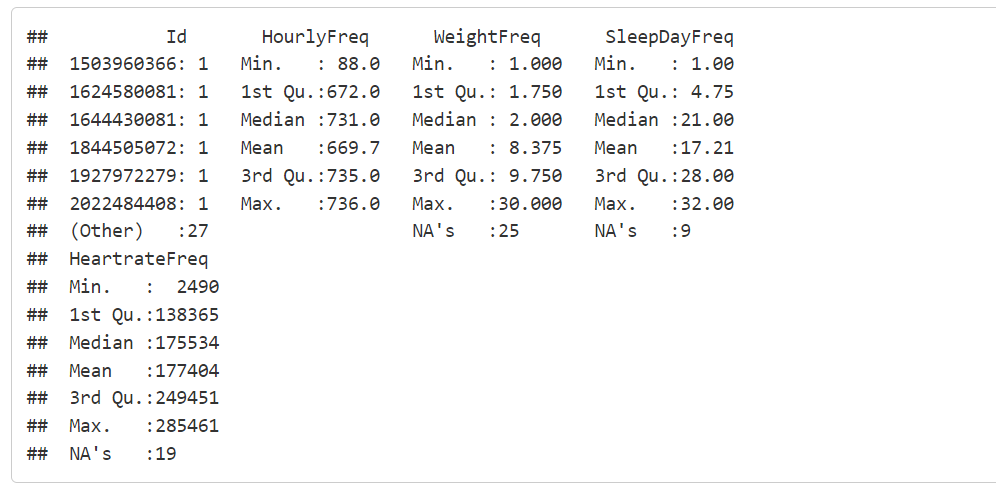
Frequency: Count the frequencies of function usage among *hourlyActivity, heartrate, weight and sleep Day* datasets.

1. Calculation

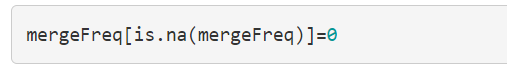
There are two calculation need to be carried out:

1) The usage frequency in each functions;



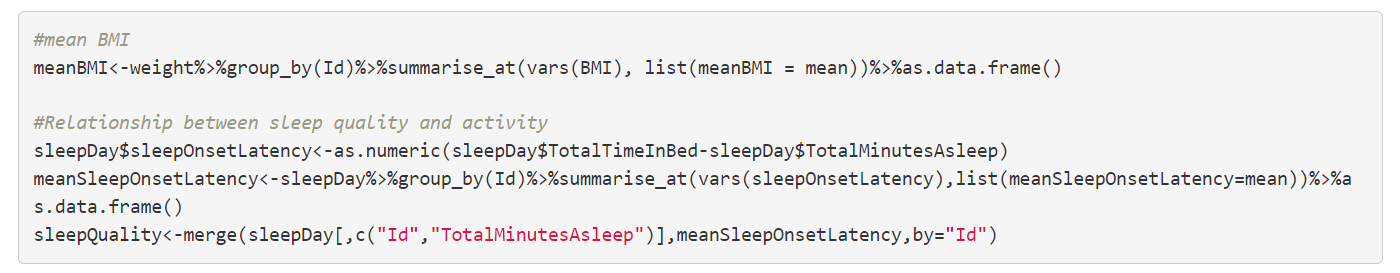


From the summary, I mentioned that the usage frequency of weight recorded function are extremely low. Among 8 users who using this functions, their using frequencies are only 8 days per month in average.



For further analysis, I use 0 to fill the missing value.

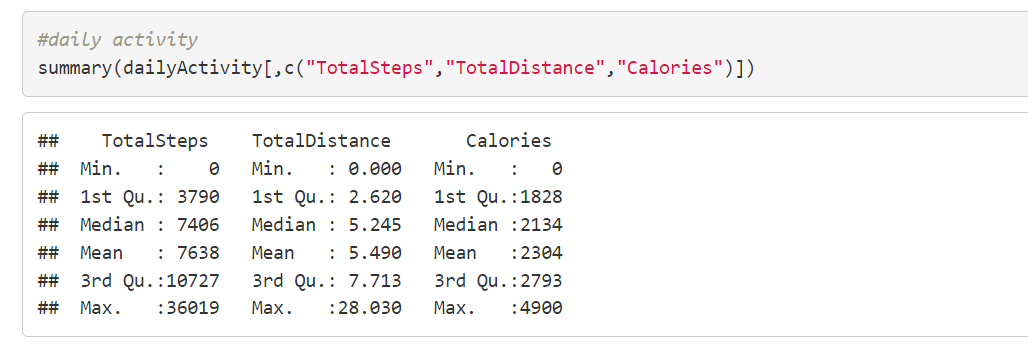
2) The calculation of mean weight and sleep quality for each user.



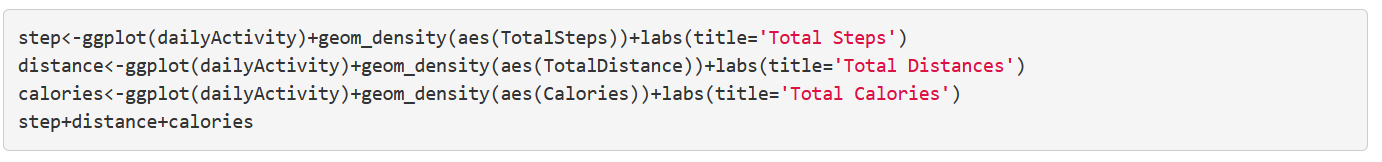
*Analysis and visualization*

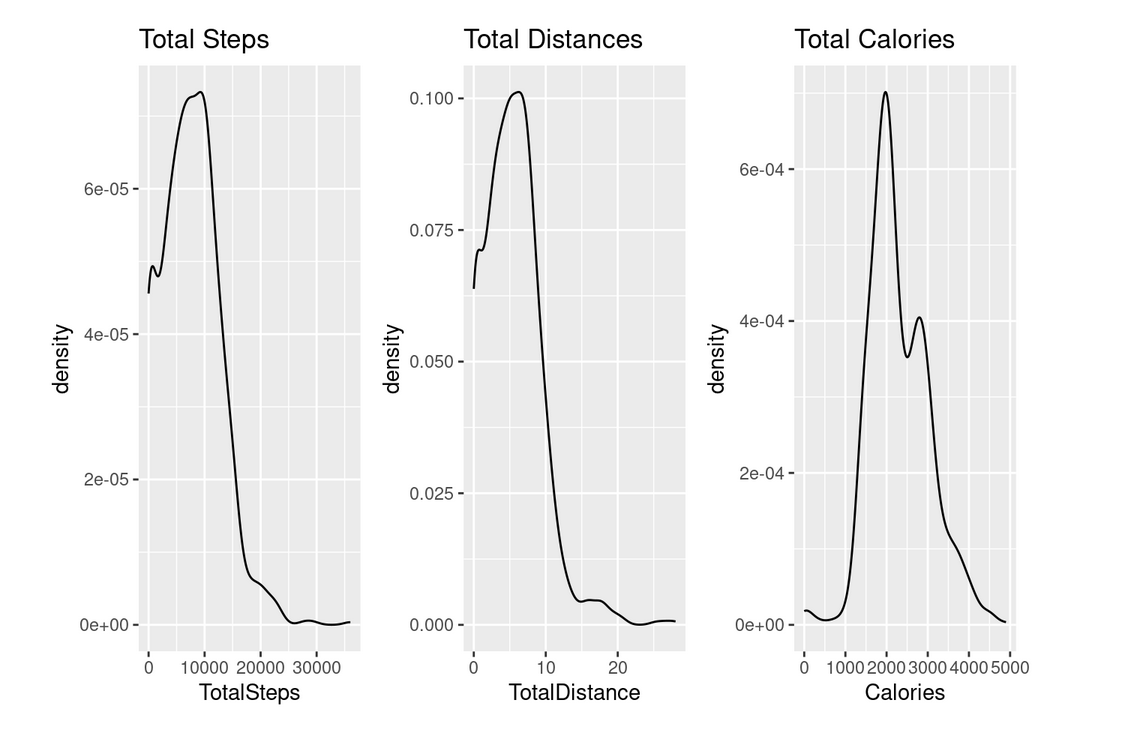
1. Descriptive Analysis

The analysis started at descriptive analysis in order to find the user characteristics and behaviors.

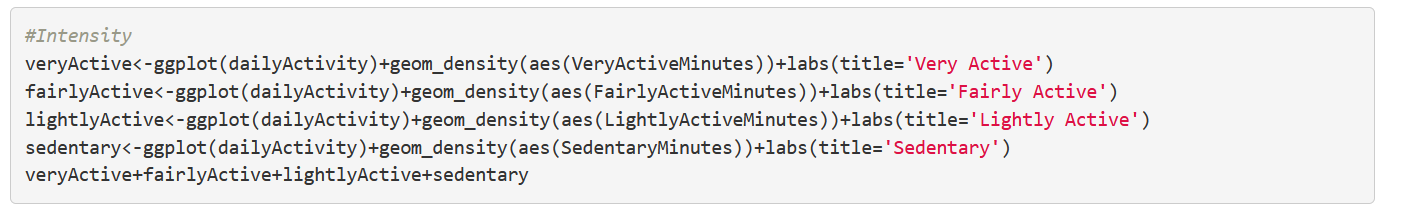


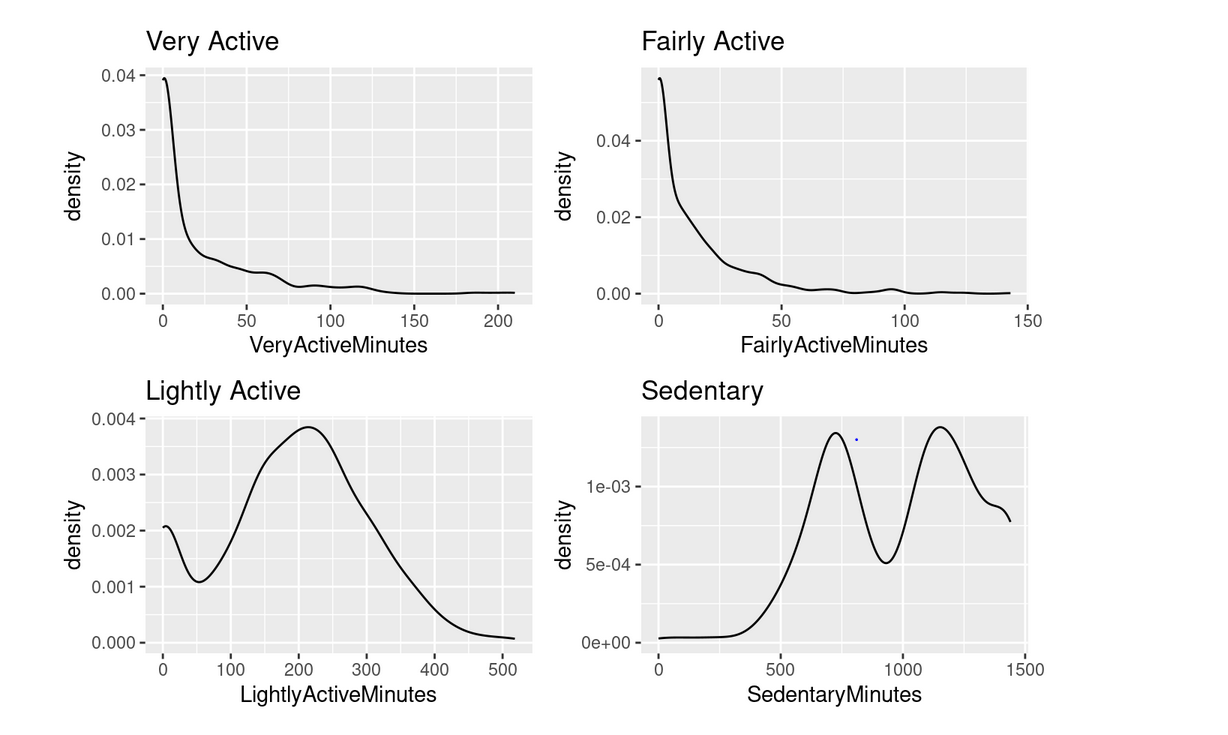
The Fitbit user averagely took 7638 steps for 5.5km and burnt 2304 calories per day.



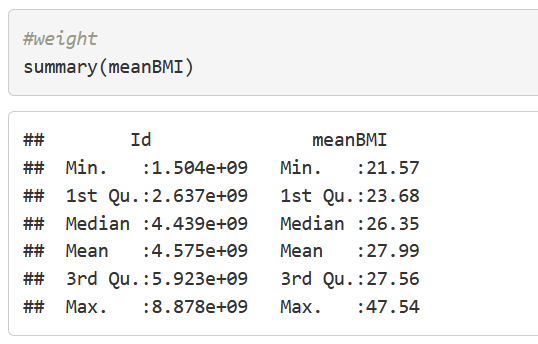
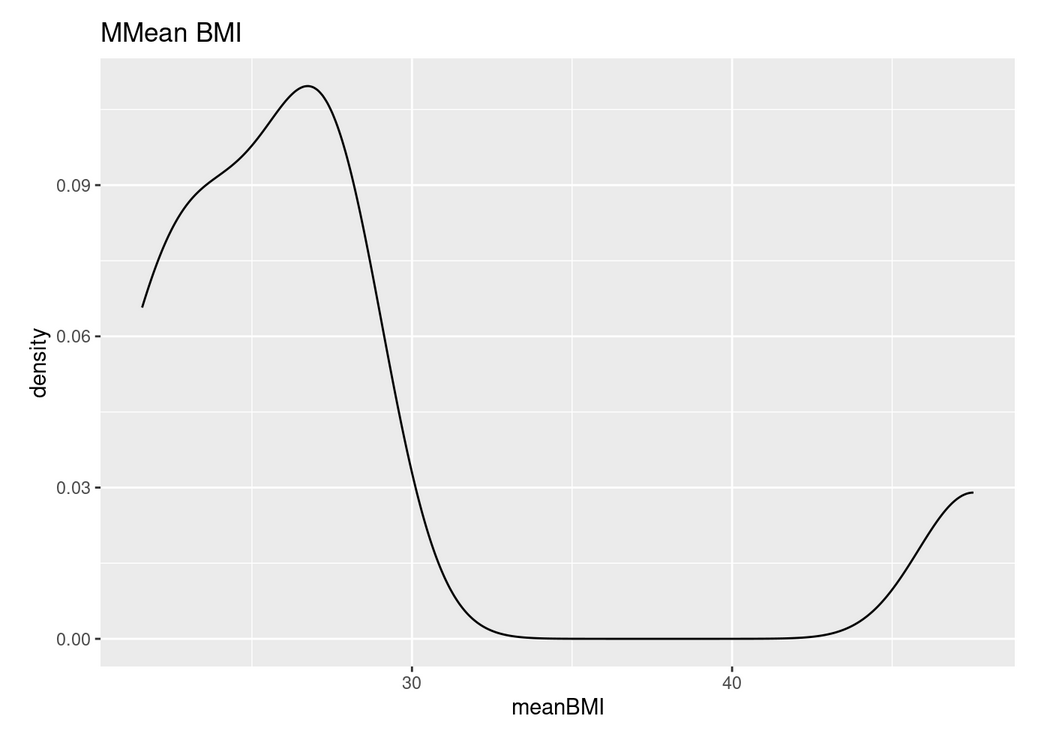


From the density graph, the distributions are reasonable. Most users took around 10000 steps and 5km per day. Fitbit users have a healthy habit as the average American only walks 3,000 to 4,000 steps a day. Most users burnt around 2000 calories per day, it is also a big number.

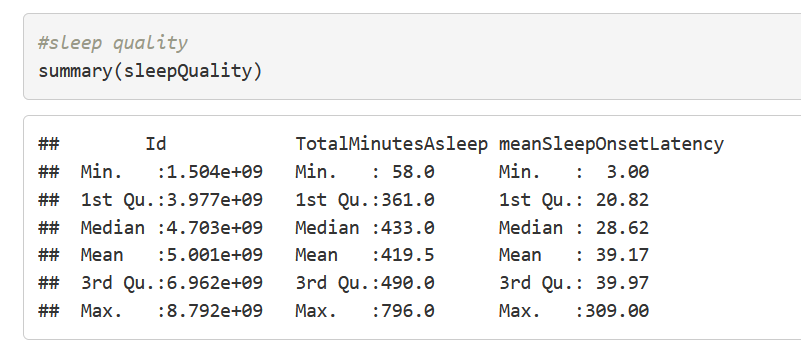
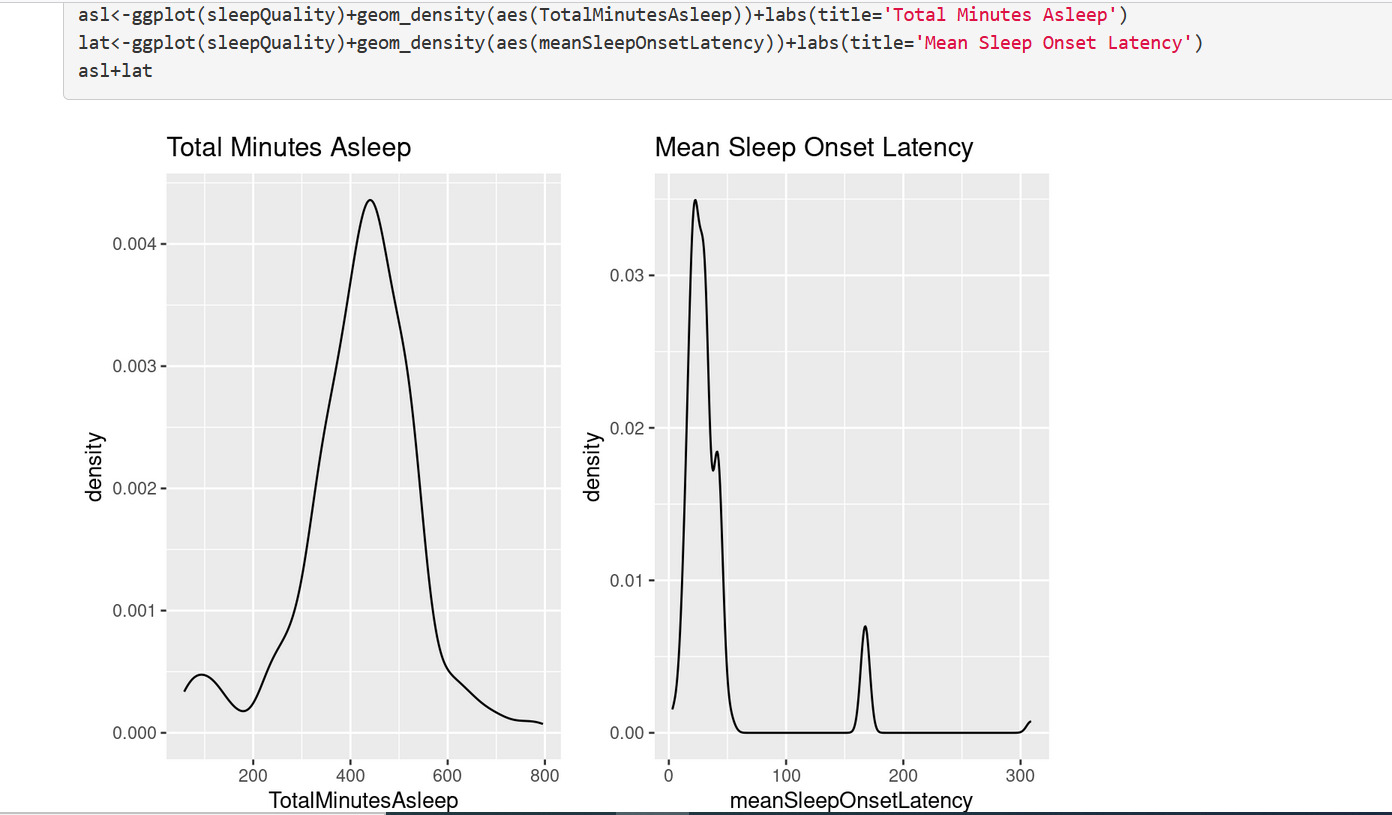




From the descriptive analysis of intensity, most users took around 200 minutes for low intensity activities, less than 50 minutes for both fair intensity activities and high intensity activities, which is a reasonable distribution. For the sedentary, it is interesting that most users sit for around 750 minutes and 1200 minutes but it showed a dramatic decrease between them. I think it represent different types of population, for those who sedentary for 750 minutes, they may be a student or non-office worker, who no need to stay at a place continuously except of sleeping hour. For those who sedentary for 1200 minutes are the office workers, who need to stay at office for a whole day.

The mean BMI of users is 28, and most users’ BMI are between 20-29, distributed in healthy weight range and slightly overweight range. However, only 8 users are recording their weight, it can’t represent the distribution of all users.

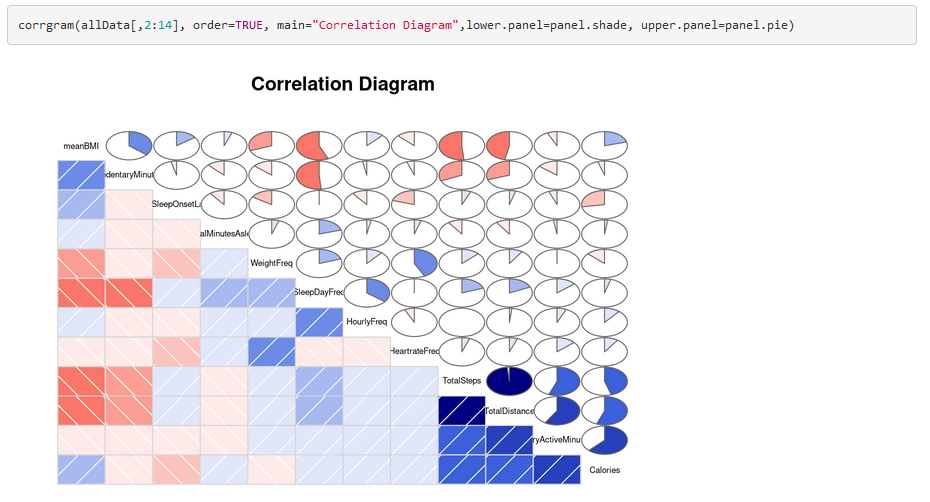
 

For sleep quality, most users slept around 400-500 minutes per day, which is 6-8 hours per day. Most of them can fall asleep after 30 minutes they went to bed, but some of them may have an insomnia problem as they need to take about 2 hours to fall asleep.

1. Relationship between weight, sleep quality and activity



As there are too many variables, a correlation diagram is drawn to easier visualize the correlation relationships.



There are few strong correlation relationships can be observed:

1. *BMI* is positively correlated with *sedentary minutes*, as sedentary habit will increase weight.
2. *BMI* is negatively correlated with *total steps and distances*, as more daily activities help reduce weight.
3. *BMI* is negatively correlated with *sleep day and weight recorded frequency*, as user who in healthier weight range will tend to use more functions of smart devices.
4. *Total steps* is positively correlated with *total distance, intensity and calories*, which is very reasonable.
5. *Sleep day recorded frequency* and *hourly activity recorded frequency* have a positive correlation. It may mainly due to users who use smart devices for sleep records will automatically record as hourly activities too,
6. *Weight recorded frequency* and *heartrate recorded frequency* have a positive correlation. It may represent users who are actively in losing weight as they will measure their weight frequently and also check their heartrate while doing exercise.

***Conclusion***

The mainly findings from Fitbit data analysis:

1. The popularity of function is following by the order: activity > sleep day > heart rate > weight.
2. User activities shown a reasonable relationship: More activities burnt more calories and had healthier weight.
3. Users who are trying lose weight will mention their weight and the heartrate function is useful for them.
4. Users who use smart devices will overall having a high daily activities.