Analysis of FitBit Fitness Track Data for Bellabeat

Yaxin Guan

2022/3/3

Data Source

FitBit Fitness Track Data

Load necessary packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last

## The following object is masked from 'package:purrr':
##
##     transpose
```

Loading CSV files

The data are from April 12th, 2016 to May 12th, 2016.

```
# Filepath <- "User/Capstone-Project/"
daily_activity <- fread(paste0(Filepath, "dailyActivity_merged.csv"))
```

```
sleep_day <- fread(paste0(Filepath, "sleepDay_merged.csv"))
```

```
weight_log <- fread(paste0(Filepath, "weightLogInfo_merged.csv"))
```

```
hourly_intensities <- fread(paste0(Filepath, "hourlyIntensities_merged.csv"))
hourly_calories <- fread(paste0(Filepath, "hourlyCalories_merged.csv"))
hourly_steps <- fread(paste0(Filepath, "hourlySteps_merged.csv"))
```

Explore data

```
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1: 1503960366    4/12/2016      13162          8.50            8.50
## 2: 1503960366    4/13/2016      10735          6.97            6.97
## 3: 1503960366    4/14/2016      10460          6.74            6.74
## 4: 1503960366    4/15/2016       9762          6.28            6.28
## 5: 1503960366    4/16/2016      12669          8.16            8.16
## 6: 1503960366    4/17/2016       9705          6.48            6.48
##    LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1:                        0              1.88                      0.55
## 2:                        0              1.57                      0.69
## 3:                        0              2.44                      0.40
## 4:                        0              2.14                      1.26
## 5:                        0              2.71                      0.41
## 6:                        0              3.19                      0.78
##    LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1:                6.06                       0                25
## 2:                4.71                       0                21
## 3:                3.91                       0                30
## 4:                2.83                       0                29
## 5:                5.04                       0                36
## 6:                2.51                       0                38
##    FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
```

```
## 1:                  13                 328              728     1985
## 2:                  19                 217              776     1797
## 3:                  11                 181             1218     1776
## 4:                  34                 209              726     1745
## 5:                  10                 221              773     1863
## 6:                  20                 164              539     1728
```

```
colnames(daily_activity)
```

```
##  [1] "Id"                       "ActivityDate"
##  [3] "TotalSteps"               "TotalDistance"
##  [5] "TrackerDistance"          "LoggedActivitiesDistance"
##  [7] "VeryActiveDistance"       "ModeratelyActiveDistance"
##  [9] "LightActiveDistance"      "SedentaryActiveDistance"
## [11] "VeryActiveMinutes"        "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes"     "SedentaryMinutes"
## [15] "Calories"
```

```
head(sleep_day)
```

```
##            Id              SleepDay TotalSleepRecords TotalMinutesAsleep
## 1: 1503960366 4/12/2016 12:00:00 AM                 1                327
## 2: 1503960366 4/13/2016 12:00:00 AM                 2                384
## 3: 1503960366 4/15/2016 12:00:00 AM                 1                412
## 4: 1503960366 4/16/2016 12:00:00 AM                 2                340
## 5: 1503960366 4/17/2016 12:00:00 AM                 1                700
## 6: 1503960366 4/19/2016 12:00:00 AM                 1                304
##    TotalTimeInBed
## 1:            346
## 2:            407
## 3:            442
## 4:            367
## 5:            712
## 6:            320
```

```
colnames(sleep_day)
```

```
## [1] "Id"                 "SleepDay"           "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed"
```

```
head(weight_log)
```

```
##            Id                  Date WeightKg WeightPounds Fat   BMI
## 1: 1503960366  5/2/2016 11:59:59 PM     52.6     115.9631  22 22.65
## 2: 1503960366  5/3/2016 11:59:59 PM     52.6     115.9631  NA 22.65
## 3: 1927972279  4/13/2016 1:08:52 AM    133.5     294.3171  NA 47.54
## 4: 2873212765 4/21/2016 11:59:59 PM     56.7     125.0021  NA 21.45
## 5: 2873212765 5/12/2016 11:59:59 PM     57.3     126.3249  NA 21.69
## 6: 4319703577 4/17/2016 11:59:59 PM     72.4     159.6147  25 27.45
##    IsManualReport         LogId
## 1:           TRUE 1462233599000
```

```
## 2:           TRUE 1462319999000
## 3:          FALSE 1460509732000
## 4:           TRUE 1461283199000
## 5:           TRUE 1463097599000
## 6:           TRUE 1460937599000
```

```
colnames(weight_log)
```

```
## [1] "Id"             "Date"           "WeightKg"       "WeightPounds"
## [5] "Fat"            "BMI"            "IsManualReport" "LogId"
```

```
head(hourly_intensities)
```

```
##            Id          ActivityHour TotalIntensity AverageIntensity
## 1: 1503960366 4/12/2016 12:00:00 AM             20         0.333333
## 2: 1503960366  4/12/2016 1:00:00 AM              8         0.133333
## 3: 1503960366  4/12/2016 2:00:00 AM              7         0.116667
## 4: 1503960366  4/12/2016 3:00:00 AM              0         0.000000
## 5: 1503960366  4/12/2016 4:00:00 AM              0         0.000000
## 6: 1503960366  4/12/2016 5:00:00 AM              0         0.000000
```

```
colnames(hourly_intensities)
```

```
## [1] "Id"               "ActivityHour"     "TotalIntensity"   "AverageIntensity"
```

```
head(hourly_calories)
```

```
##            Id          ActivityHour Calories
## 1: 1503960366 4/12/2016 12:00:00 AM       81
## 2: 1503960366  4/12/2016 1:00:00 AM       61
## 3: 1503960366  4/12/2016 2:00:00 AM       59
## 4: 1503960366  4/12/2016 3:00:00 AM       47
## 5: 1503960366  4/12/2016 4:00:00 AM       48
## 6: 1503960366  4/12/2016 5:00:00 AM       48
```

```
colnames(hourly_calories)
```

```
## [1] "Id"           "ActivityHour" "Calories"
```

```
head(hourly_steps)
```

```
##            Id          ActivityHour StepTotal
## 1: 1503960366 4/12/2016 12:00:00 AM       373
## 2: 1503960366  4/12/2016 1:00:00 AM       160
## 3: 1503960366  4/12/2016 2:00:00 AM       151
## 4: 1503960366  4/12/2016 3:00:00 AM         0
## 5: 1503960366  4/12/2016 4:00:00 AM         0
## 6: 1503960366  4/12/2016 5:00:00 AM         0
```

```
colnames(hourly_steps)
```

```
## [1] "Id"           "ActivityHour" "StepTotal"
```

Energy expenditure formula is from: **HSS** Calories/minutes = 0.0175 x MET (of activity) x body weight (in kg) Calories will burn even when sitting quietly, so calories burn when the hourly step is 0.

```
n_distinct(daily_activity$Id)
```

```
## [1] 33
```

```
n_distinct(sleep_day$Id)
```

```
## [1] 24
```

```
n_distinct(weight_log$Id)
```

```
## [1] 8
```

It looks like there may be more participants in the daily activity dataset than the sleep dataset.

Number of observations

```
nrow(daily_activity)
```

```
## [1] 940
```

```
nrow(sleep_day)
```

```
## [1] 413
```

```
nrow(weight_log)
```

```
## [1] 67
```

```
nrow(hourly_calories)
```

```
## [1] 22099
```

```
nrow(hourly_intensities)
```

```
## [1] 22099
```

```
nrow(hourly_steps)
```

```
## [1] 22099
```

Summaries

For the daily activity data frame:

```
daily_activity %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8
##  Median : 7406   Median : 5.245   Median :1057.5
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5
##  Max.   :36019   Max.   :28.030   Max.   :1440.0
```

For the sleep data frame:

```
sleep_day %>%
  select(TotalSleepRecords,
  TotalMinutesAsleep,
  TotalTimeInBed) %>%
  summary()
```

```
##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :1.000     Min.   : 58.0      Min.   : 61.0
##  1st Qu.:1.000     1st Qu.:361.0      1st Qu.:403.0
##  Median :1.000     Median :433.0      Median :463.0
##  Mean   :1.119     Mean   :419.5      Mean   :458.6
##  3rd Qu.:1.000     3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :3.000     Max.   :796.0      Max.   :961.0
```

For the weight log data frame:

```
weight_log %>%
  select(WeightKg,WeightPounds, BMI) %>%
  summary()
```

```
##     WeightKg       WeightPounds        BMI
##  Min.   : 52.60   Min.   :116.0   Min.   :21.45
##  1st Qu.: 61.40   1st Qu.:135.4   1st Qu.:23.96
##  Median : 62.50   Median :137.8   Median :24.39
##  Mean   : 72.04   Mean   :158.8   Mean   :25.19
##  3rd Qu.: 85.05   3rd Qu.:187.5   3rd Qu.:25.56
##  Max.   :133.50   Max.   :294.3   Max.   :47.54
```

For the hourly intensities data frame:

```
hourly_intensities %>%
  select(TotalIntensity, 'AverageIntensity(in sec)' = AverageIntensity) %>%
  summary()
```

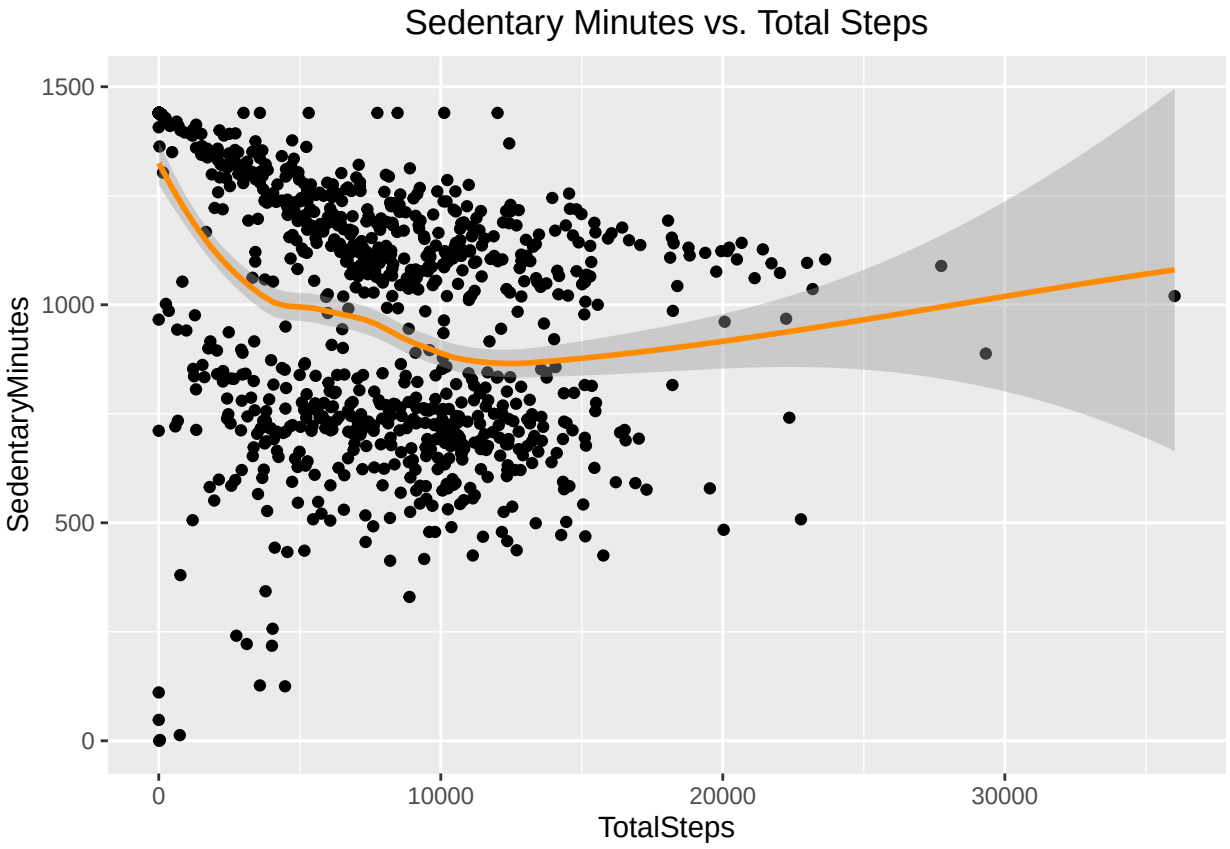
```
##  TotalIntensity   AverageIntensity(in sec)
##  Min.   :  0.00   Min.   :0.0000
##  1st Qu.:  0.00   1st Qu.:0.0000
##  Median :  3.00   Median :0.0500
##  Mean   : 12.04   Mean   :0.2006
##  3rd Qu.: 16.00   3rd Qu.:0.2667
##  Max.   :180.00   Max.   :3.0000
```

Data Visualization

```
ggplot(data=daily_activity, aes(x=TotalSteps, y=SedentaryMinutes)) + geom_point() + geom_smooth(color =
  theme(plot.title = element_text(hjust = 0.5))
```

Sedentary Minutes vs. Total Steps

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
# ggsave("sedentaryminutes_totalsteps.jpg")
```

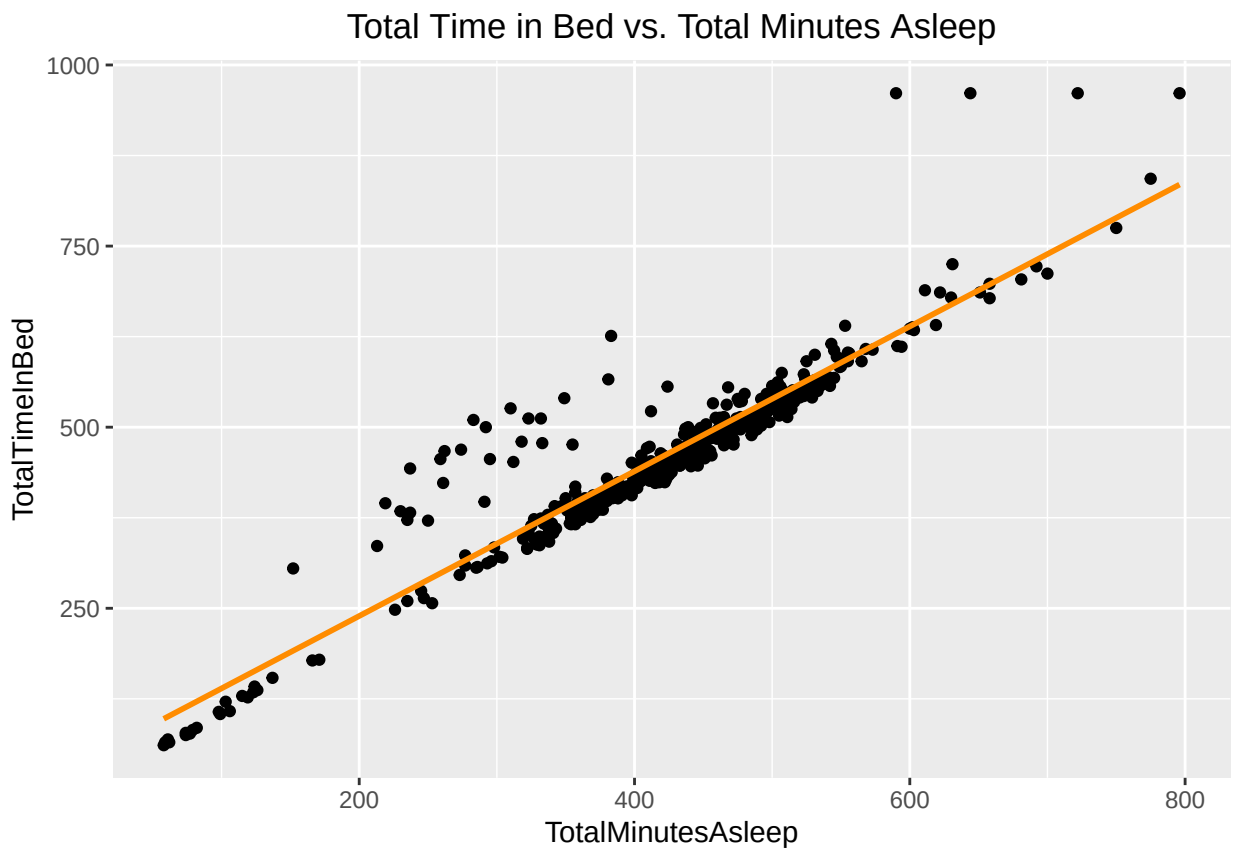
```
cor.test(daily_activity$TotalSteps, daily_activity$SedentaryMinutes,
         alternative = "less")
```

```
##
##  Pearson's product-moment correlation
##
## data:  daily_activity$TotalSteps and daily_activity$SedentaryMinutes
## t = -10.615, df = 938, p-value < 2.2e-16
## alternative hypothesis: true correlation is less than 0
## 95 percent confidence interval:
##  -1.0000000 -0.2786998
## sample estimates:
##        cor
## -0.3274835
```

Although the trend is not obvious in the graph, the correlation is negative.

```
ggplot(data=sleep_day, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) + geom_point() +
  geom_smooth(aes(group = 1),method = "lm", formula = y ~ x,se = FALSE, color = "darkorange") + labs(ti
  theme(plot.title = element_text(hjust = 0.5))
```

Total Time in Bed vs. Total Minutes Asleep




```
# ggsave("Totaltimeinbed_totalasleep.jpg")
```

```
cor.test(sleep_day$TotalMinutesAsleep, sleep_day$TotalTimeInBed,
         alternative = "greater")
```

```
##
##  Pearson's product-moment correlation
##
## data:  sleep_day$TotalMinutesAsleep and sleep_day$TotalTimeInBed
## t = 51.483, df = 411, p-value < 2.2e-16
## alternative hypothesis: true correlation is greater than 0
## 95 percent confidence interval:
##  0.9186882 1.0000000
## sample estimates:
##       cor
## 0.9304575
```

Positive and strong correlation (close to 1) as expected.

```
#combined_data <- merge(sleep_day, daily_activity, by = "Id", allow.cartesian=TRUE)
combined_data <- right_join(sleep_day, daily_activity, by = "Id")
```

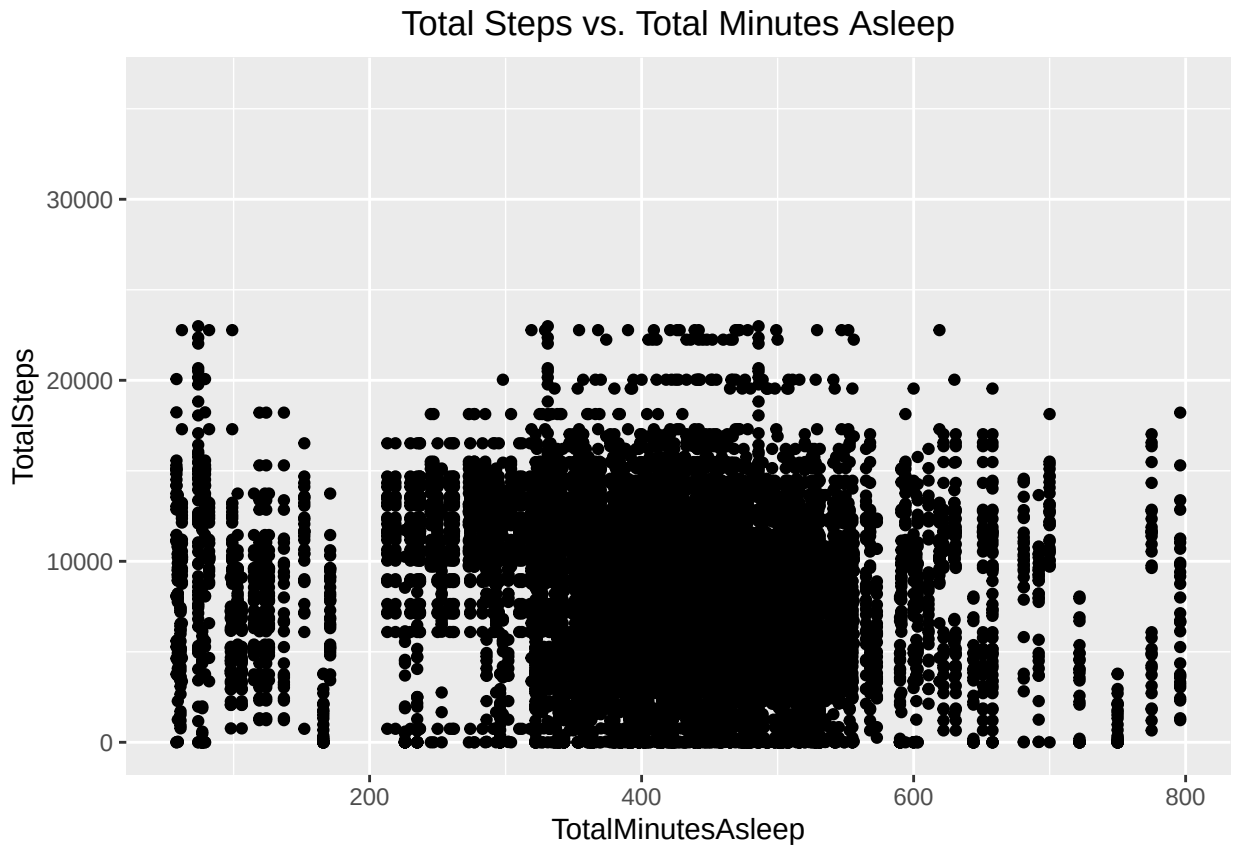
```
n_distinct(combined_data$Id)
```

```
## [1] 33
```

```
ggplot(data = combined_data, aes(x = TotalMinutesAsleep, y = TotalSteps)) +
  geom_point() + labs(title = "Total Steps vs. Total Minutes Asleep") +
  theme(plot.title = element_text(hjust = 0.5))
```

Total Steps vs. Total Minutes Asleep

```
## Warning: Removed 227 rows containing missing values (geom_point).
```




```
cor.test(combined_data$TotalMinutesAsleep, combined_data$TotalSteps)
```

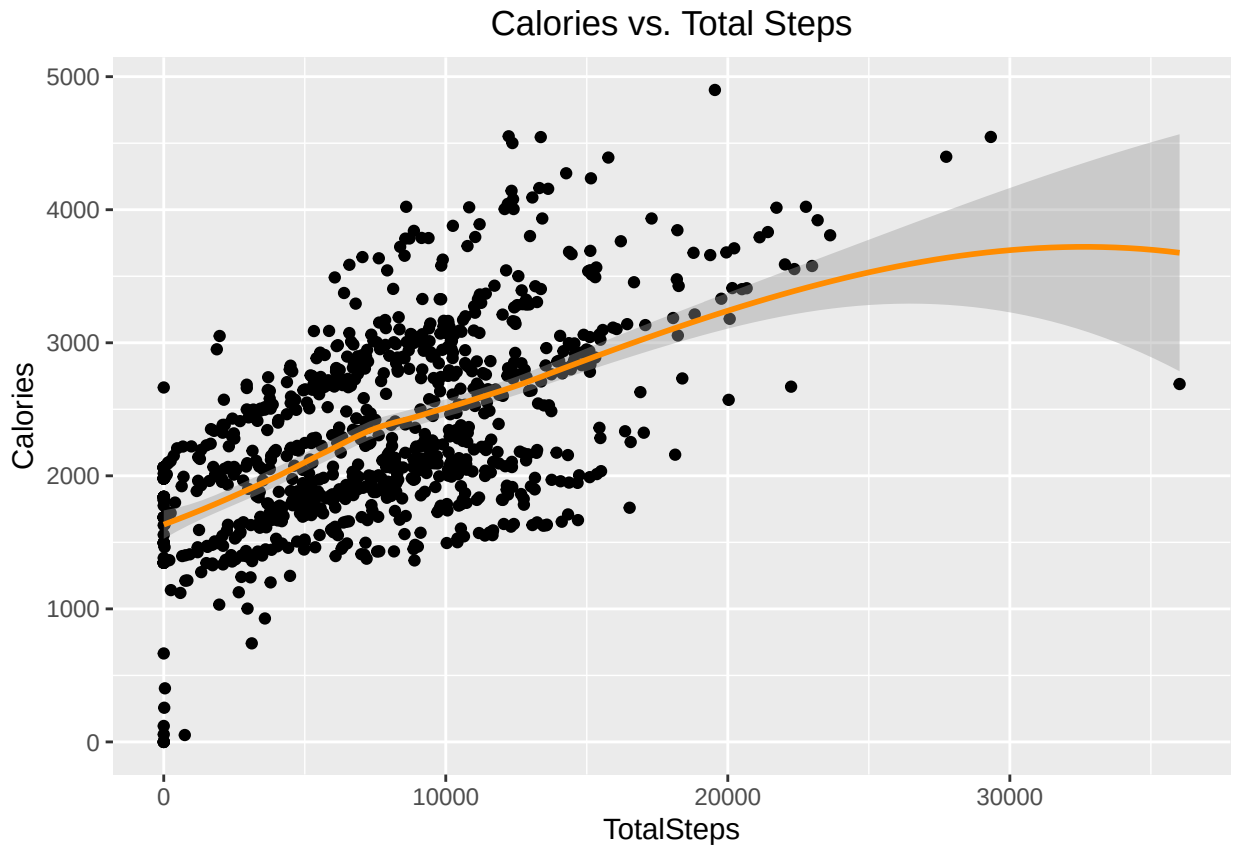
```
##
##  Pearson's product-moment correlation
##
## data:  combined_data$TotalMinutesAsleep and combined_data$TotalSteps
## t = -11.044, df = 12439, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.11591302 -0.08110962
## sample estimates:
##         cor
## -0.09854146
```

The correlation is negative, but correlation does not mean the causation. More data and investigation (increase of sample size, controlled study, etc.) are needed to show that total steps can keep people feel energetic and decrease the sleep times.

```
ggplot(data = daily_activity, aes(x = TotalSteps, y = Calories)) + geom_point() +
  geom_smooth(method = "loess",color = "darkorange") + labs(title = "Calories vs. Total Steps") + theme
```

Calories vs. Total Steps

```
## `geom_smooth()` using formula 'y ~ x'
```




```
# ggsave("calories_totalsteps.jpg")
```

```
cor.test(daily_activity$TotalSteps, daily_activity$Calories,
         alternative = "greater")
```

```
##
##  Pearson's product-moment correlation
##
## data:  daily_activity$TotalSteps and daily_activity$Calories
## t = 22.472, df = 938, p-value < 2.2e-16
## alternative hypothesis: true correlation is greater than 0
## 95 percent confidence interval:
##  0.5555268 1.0000000
## sample estimates:
##       cor
## 0.5915681
```

The correlation is positive, which matches the trend in graph. The more time a person spends on walking, the more calories one will burn.

```
hourly_intensities$Date <- format(as.Date(hourly_intensities$ActivityHour,
                                  format = "%m/%d/%Y"), format = "%m/%d/%Y") # Date

hourly_intensities$ActivityHour <- mdy_hms(hourly_intensities$ActivityHour,
                                           tz = Sys.timezone())

hourly_intensities$Time <- format(hourly_intensities$ActivityHour,
                                  format = "%H:%M:%S")

hourly_intensities$day_of_week <- format(as.Date(hourly_intensities$ActivityHour), "%A")
```

```
hourly_intensities$day_of_week <- ordered(hourly_intensities$day_of_week, levels=c("Sunday", "Monday",
```

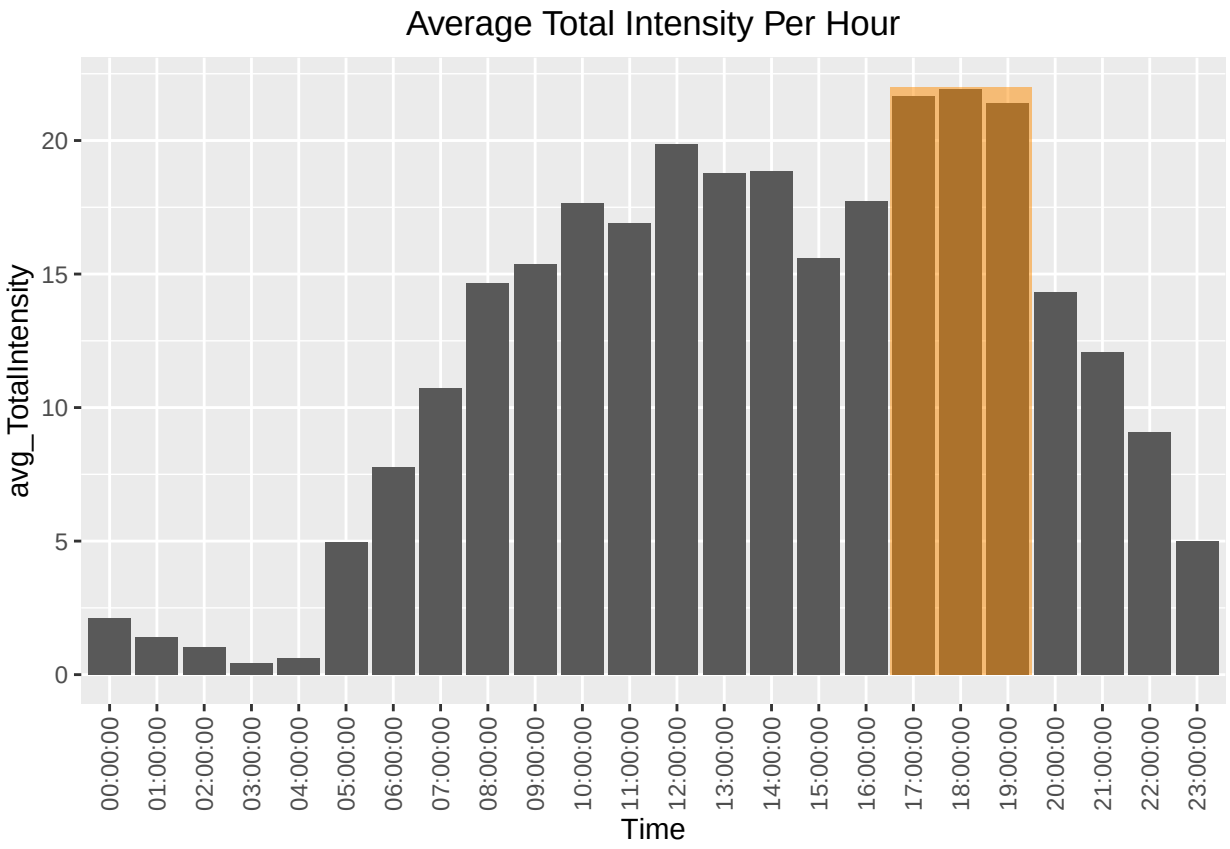
```
extract_data <- hourly_intensities[, c(3,6)]
plot_data <- extract_data %>%
  group_by(Time) %>%
  summarise(avg_TotalIntensity = mean(TotalIntensity))

extract_data2 <- hourly_intensities[, c(3,7)]
plot_data2 <- extract_data2 %>%
  group_by(day_of_week) %>%
  summarise(avg_TotalIntensity = mean(TotalIntensity))
```

Time of Intensities

```
ggplot(data = plot_data, aes(x = Time, y = avg_TotalIntensity)) +
  geom_bar(stat = "identity") + labs(title = "Average Total Intensity Per Hour") +
  annotate("rect", xmin = 17.5, ymin = 0, xmax = 20.5, ymax = 22,
           fill = "darkorange", alpha = 0.5)+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5),
        plot.title = element_text(hjust = 0.5))
```

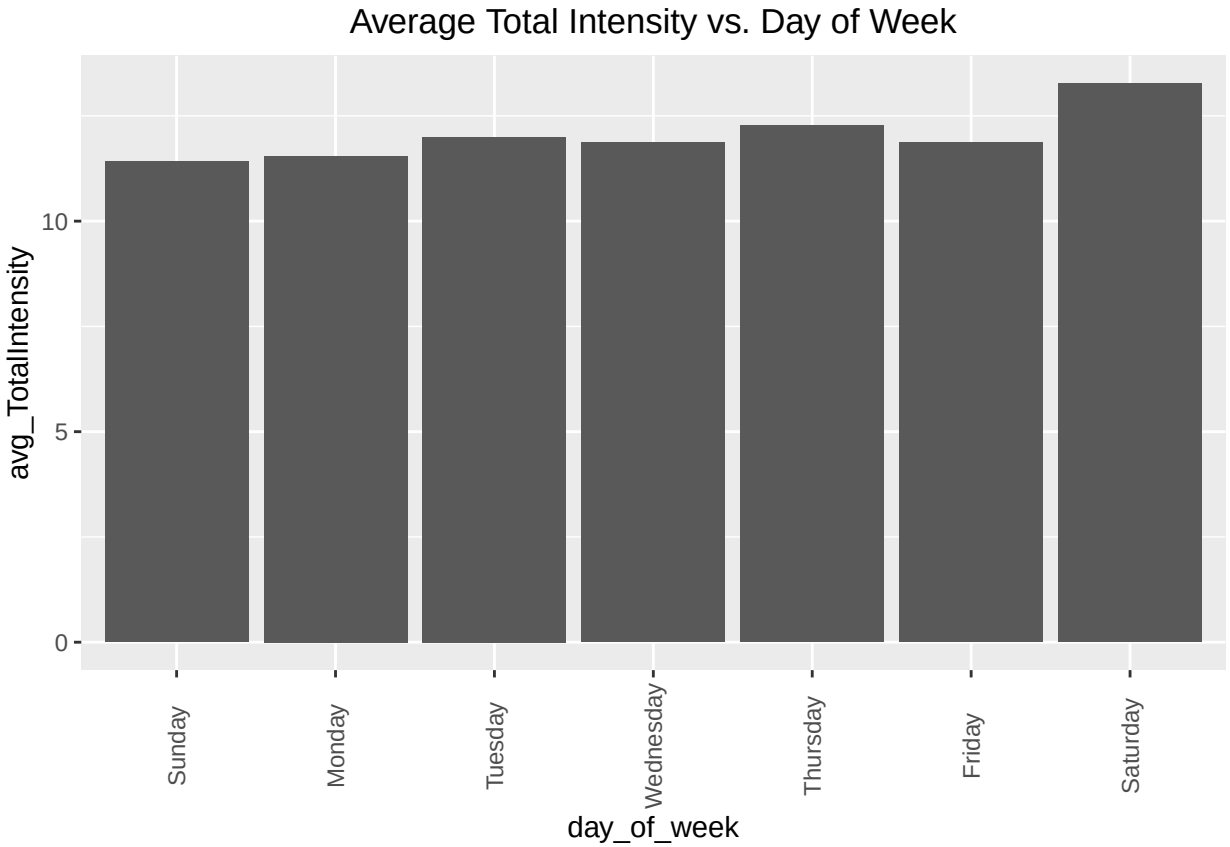
Average of Total Intensity vs. Time




```
#ggsave("time_total_intensity.jpg")
```

```
ggplot(data = plot_data2, aes(x = day_of_week, y = avg_TotalIntensity)) +
  geom_bar(stat = "identity") + labs(title = "Average Total Intensity vs. Day of Week") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5),
        plot.title = element_text(hjust = 0.5))
```

Average of Total Intensity vs. Days of Week



Conclusion

The total steps, calories burned, total minutes asleep, total time in bed, hourly intensity, and sedentary minutes are six key elements that people recorded with their smart devices and used in this analysis. The correlation between total steps and calories burned is positive (0.5915681). The correlation between total minutes asleep and total time in bed burned is strong and positive (0.9304575). Moreover, the hourly intensity shows that the participants (8 unique Id in this case) have more intensities at 5:00 PM to 7:00 PM. It is reasonable because most people are off work at that time, yet increasing the sample size will help defining the trend better. The data from the dailyActivity_merged.csv file shows the mean sedentary minutes of the participant is 991.2 (16.52 hr).

Based on the trends above, the Leaf (classic wellness tracker) of Bellabeat is good products for keeping track of activity and sleep. Bellabeat's Leaf can keep track of activity like walking and calories burned. The Leaf can remind the user of ones sleep goal to maintain a good sleep habit. The Leaf is able to keep track of light sleep and deep sleep. It also has goal for steps and active hours for exercise to achieve wellness goal. Leaf has the inactivity alert feature, which it will remind the user to move more or less often, by vibrating consecutively when the user has been inactive. The user needs to do a certain number of steps to not have the Leaf reminds. It will be better if Bellabeat can extend on this feature, such as reminding the user to stand up when ones sedentary time is long. Since there is job position requires the person to stay in the same location most of time and cannot satisfy the requirement of steps, it will help the user to burn more calories even in the work environment when the user stands up instead of sitting.

These trends could help influence Bellabeat marketing strategy by allowing Bellbeat to advertise its products are capable to do the same as other fitness trackers and go beyond them. With Bellbeat's app, the user can share ones good or bad sleep that measures by the Leaf on social networks and message app. Bellabeat's app also keeps track of hydration and reproductive health, offers meditation and breathing exercise, calculates

sensitivity to stress based on all the data that the app tracks, and has the alarm function. Bellabeat mentions it is the "only health tracker made for women" in its website. It can also introduce its product as an excellent gift for a woman and worthy investment to maintain wellness.