

# Bellabeat Case Study

Vedanta Mohapatra

2022-03-07

```
tinytex::install_tinytex()
```

## Introduction and Background



Figure 1: Bellabeat

Bellabeat, a hightech wellness company that manufactures health-focussed products for women. It is a succesful small company but have the potential to be a big game player.

Their products include:

1. **Bellabeat app:** Provides users with their health data related to their activity, sleep and habits. The data helps users to understand their current habits and make healthy decisions.
2. **Leaf:** A wellness tracker which is connected to the app to track activity, sleep and stess.
3. **Time:** This is a watch that combines a timpiece and smart technology to track user activity, sleep and stress and provides insights on daily wellness.
4. **Spring:** A water bottle that tracks daily water intake and is connected to the app to track hydration levels.
5. **Bellabeat Membership:** A subscription-based membership program for users which offers guidance on having a healthy lifesyle.

Bellabeat has invested in traditional advertising media like radio, billboards, print and television. In additon to this, thye have invested year-round in google search and run ads on youtube. They have good engagement with users on Twitter, Instagram and Facebook.

## Statement of Business Task

Urška Sršen, co-founder of Bellabeat believes analysis of Bellabeat's available consumer data can reveal more oppurtunities for growth. To gain insights on how smart devices are used, Sršen asks us to analyze smart device usage data of non-Bellabeat products and apply the findings to one Bellabeat prodcut.

Key stakeholders

- **Urška Sršen:** Bellabeat's cofounder and Chief Creative Officer
- **Sando Mur:** Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
- **Bellabeat marketing analytics team:** A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

So the best thing to do is

1. Investigate products similar to that of Bellabeat's and gain insights of their smart device usage
2. Find ways on how to apply these trends on Bellabeat products
3. Give recommendations based on the trends

---

## Data preparation

The dataset used in the analysis is FitBit Fitness Tracker Data.

It is a Kaggle data set which This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data which includes actively level, calories burnt, sleep time and weight login info which will help us to examine users habits.

This dataset generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016.

The dataset includes 18 csv (comma seperated values) files.

On further inspection, the datasets were found to be in a combination of long format and wide format.

RoCCC Analysis of dataset

Let's see how the dataset performs according to the following parameters:

- **Reliability:** Not really reliable as datasets consists of only 30 people which is a low sample size. Gender, age, location and occupation of the people are not known.
- **Originality:** Obtained from a third-party source. Not original.
- **Comprehensive:** Dataset has multiple columns covering different areas like intensity, activity, calories etc. Quite comprehensive.
- **Current:** This is a dataset from 2016, 6 years ago. Even though there is no major change in our walking habits, the impact of Covid-19 is ignored.
- **Cited:** The data is well documented and cited. The one who has provided data, Möbius is also mentioned.

There is no bias in the data collected as people consented to applying for the survey.

Files used from the dataset for analysis

1. dailyActivity\_\_merged.csv

2. dailyCalories\_merged.csv
3. dailyIntensities\_merged.csv
4. dailySteps\_merged.csv
5. sleepDay\_merged.csv
6. weightLogInfo\_merged.csv

R Studio is used for the data preparation.

First by installing the various packages and loading them.

```
install.packages("tidyverse", repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vedan/OneDrive/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked  
##
```

```
## The downloaded binary packages are in  
## C:\Users\vedan\AppData\Local\Temp\RtmpmUrClq\downloaded_packages
```

```
install.packages('dplyr', repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vedan/OneDrive/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked  
##
```

```
## The downloaded binary packages are in  
## C:\Users\vedan\AppData\Local\Temp\RtmpmUrClq\downloaded_packages
```

```
install.packages('ggplot2', repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vedan/OneDrive/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked  
##
```

```
## The downloaded binary packages are in  
## C:\Users\vedan\AppData\Local\Temp\RtmpmUrClq\downloaded_packages
```

```
install.packages('tidyr', repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/vedan/OneDrive/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)
```

```
## package 'tidyr' successfully unpacked and MD5 sums checked  
##
```

```
## The downloaded binary packages are in  
## C:\Users\vedan\AppData\Local\Temp\RtmpmUrClq\downloaded_packages
```

```
install.packages('stringr', repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/vedan/OneDrive/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)

## package 'stringr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\vedan\AppData\Local\Temp\RtmpmUrClq\downloaded_packages

library(ggplot2)
library(tidyverse)

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

## v tibble 3.1.6      v dplyr 1.0.8
## v tidyr 1.2.0      v stringr 1.4.0
## v readr 2.1.2      v forcats 0.5.1
## v purrr 0.3.4

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

library(readr)
library(stringr)
library(dplyr)
library(tidyr)
```

---

## Data processing

MS Excel and R Studio are used for Data processing.

Basic cleaning was done on MS Excel removing empty data.

Then we imported the csv files to R Studio using read\_csv

```
dailyActivity_merged <- read_csv("C:/Users/vedan/Desktop/Bellabeat/Fitabase Data 4.12.16-5.12.16/dailyA...

## Rows: 940 Columns: 15
## -- Column specification -----
## Delimiter: ","
## chr (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```

dailyCalories_merged <- read_csv("C:/Users/vedan/Desktop/Bellabeat/Fitabase Data 4.12.16-5.12.16/dailyCalories_merged.csv")

## Rows: 940 Columns: 3
## -- Column specification -----
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, Calories
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

dailySteps_merged <- read_csv("C:/Users/vedan/Desktop/Bellabeat/Fitabase Data 4.12.16-5.12.16/dailySteps_merged.csv")

## Rows: 940 Columns: 3
## -- Column specification -----
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, StepTotal
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

dailyIntensities_merged <- read.csv("C:/Users/vedan/Desktop/Bellabeat/Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv")
sleepDay_merged <- read.csv("C:/Users/vedan/Desktop/Bellabeat/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
weightLogInfo_merged <- read.csv("C:/Users/vedan/Desktop/Bellabeat/Fitabase Data 4.12.16-5.12.16/weightLogInfo_merged.csv")

```

Viewing the data The datasets imported can be viewed using the view function. Also, to avoid loading all record, head function can be used.

```
head(dailyActivity_merged)
```

```

## # A tibble: 6 x 15
##       Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitie~
##       <dbl> <chr>         <dbl>         <dbl>         <dbl>         <dbl>
## 1  1.50e9 4/12/2016         13162          8.5           8.5           0
## 2  1.50e9 4/13/2016         10735          6.97          6.97          0
## 3  1.50e9 4/14/2016         10460          6.74          6.74          0
## 4  1.50e9 4/15/2016          9762          6.28          6.28          0
## 5  1.50e9 4/16/2016         12669          8.16          8.16          0
## 6  1.50e9 4/17/2016          9705          6.48          6.48          0
## # ... with 9 more variables: VeryActiveDistance <dbl>,
## #   ModeratelyActiveDistance <dbl>, LightActiveDistance <dbl>,
## #   SedentaryActiveDistance <dbl>, VeryActiveMinutes <dbl>,
## #   FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## #   SedentaryMinutes <dbl>, Calories <dbl>

```

```
head(dailyCalories_merged)
```

```
## # A tibble: 6 x 3
##       Id ActivityDay Calories
##       <dbl> <chr>         <dbl>
## 1 1503960366 4/12/2016      1985
## 2 1503960366 4/13/2016      1797
## 3 1503960366 4/14/2016      1776
## 4 1503960366 4/15/2016      1745
## 5 1503960366 4/16/2016      1863
## 6 1503960366 4/17/2016      1728
```

```
head(dailySteps_merged)
```

```
## # A tibble: 6 x 3
##       Id ActivityDay StepTotal
##       <dbl> <chr>         <dbl>
## 1 1503960366 4/12/2016      13162
## 2 1503960366 4/13/2016      10735
## 3 1503960366 4/14/2016      10460
## 4 1503960366 4/15/2016       9762
## 5 1503960366 4/16/2016      12669
## 6 1503960366 4/17/2016       9705
```

```
head(dailyIntensities_merged)
```

```
##       Id ActivityDay SedentaryMinutes LightlyActiveMinutes
## 1 1503960366 4/12/2016           728           328
## 2 1503960366 4/13/2016           776           217
## 3 1503960366 4/14/2016          1218           181
## 4 1503960366 4/15/2016           726           209
## 5 1503960366 4/16/2016           773           221
## 6 1503960366 4/17/2016           539           164
## FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
## 1           13           25                0
## 2           19           21                0
## 3           11           30                0
## 4           34           29                0
## 5           10           36                0
## 6           20           38                0
## LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
## 1           6.06           0.55           1.88
## 2           4.71           0.69           1.57
## 3           3.91           0.40           2.44
## 4           2.83           1.26           2.14
## 5           5.04           0.41           2.71
## 6           2.51           0.78           3.19
```

```
head(sleepDay_merged)
```

```
##       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
```

```
head(weightLogInfo_merged)
```

```
##           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 1.462234e+12
## 2           True 1.462320e+12
## 3          False 1.460510e+12
## 4           True 1.461283e+12
## 5           True 1.463098e+12
## 6           True 1.460938e+12
```

### Data Integrity

The selected datasets are checked and each of them have a common field Id. To maintain data integrity, the following piece of code involving length and unique function is run to check distinct values of id.

```
length(unique(dailyActivity_merged$Id))
```

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

```
length(unique(dailyCalories_merged$Id))
```

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

```
length(unique(dailySteps_merged$Id))
```

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

```
length(unique(dailyIntensities_merged$Id))
```

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

```
length(unique(sleepDay_merged$Id))
```

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

```
length(unique(weightLogInfo_merged$Id))
```

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

We can see the data in `sleepDay_merged` and `weightLogInfo_merged` have only 24 and 8 unique IDs in comparison to 33 of other tables. These missing values can effect the analysis.

---

## Data Analysis

The analysis tool used again is R Studio.

Observed trends The following hypothesis can be made with the data available among various tables

1. Relation between total distance and calories.
2. Relation between total distance and total steps.
3. Relation between Activity minutes and calories.
4. Relation between Time spent in bed and calories.
5. Relation between Activity minutes and BMI.
6. Relation between sleeping minutes and Non-active minutes.

Organising and Formatting the data . Now the tables `sleepDay_merged` as well as `weightLogInfo_merged` has Date recorded in Date + Time format. So using the separate function, we have created a new dataframe “`sleepDay_new`” where the Sleepday column is divided into two separate columns: `ActivityDay` and `Time`.

```
sleepDay_new <- sleepDay_merged %>% separate(SleepDay, c("ActivityDay", "Time"), sep=" ")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 413 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
head(sleepDay_new)
```

##	Id	ActivityDay	Time	TotalSleepRecords	TotalMinutesAsleep
## 1	1503960366	4/12/2016	12:00:00	1	327
## 2	1503960366	4/13/2016	12:00:00	2	384
## 3	1503960366	4/15/2016	12:00:00	1	412
## 4	1503960366	4/16/2016	12:00:00	2	340
## 5	1503960366	4/17/2016	12:00:00	1	700



```
## 6 1503960366 4/19/2016 12:00:00 1 304
## TotalTimeInBed
## 1 346
## 2 407
## 3 442
## 4 367
## 5 712
## 6 320
```

Similarly, a new dataframe “weightLogInfo\_new” is also created where the date column is divided into two separate columns: ActivityDate and Time

```
weightLogInfo_new <- weightLogInfo_merged %>% separate(Date,c("ActivityDate", "Time"), sep=" ")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 67 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
head(weightLogInfo_new)
```

```
##      Id ActivityDate      Time WeightKg WeightPounds Fat   BMI
## 1 1503960366 5/2/2016 11:59:59    52.6    115.9631 22 22.65
## 2 1503960366 5/3/2016 11:59:59    52.6    115.9631 NA 22.65
## 3 1927972279 4/13/2016 1:08:52   133.5    294.3171 NA 47.54
## 4 2873212765 4/21/2016 11:59:59    56.7    125.0021 NA 21.45
## 5 2873212765 5/12/2016 11:59:59    57.3    126.3249 NA 21.69
## 6 4319703577 4/17/2016 11:59:59    72.4    159.6147 25 27.45
##   IsManualReport      LogId
## 1           True 1.462234e+12
## 2           True 1.462320e+12
## 3          False 1.460510e+12
## 4           True 1.461283e+12
## 5           True 1.463098e+12
## 6           True 1.460938e+12
```

It can be seen the columns are divided in the tables.

## Data Aggregation

To analyze Time spent in bed and calories, the tables dailyActivity\_merged and sleepDay\_new can be merged with Id and ActivityDay column to form a new data frame “sleep\_calories”.

```
sleep_calories <-merge.data.frame(dailyCalories_merged, sleepDay_new, by = c("Id", "ActivityDay"))
head(sleep_calories)
```

```
##      Id ActivityDay Calories      Time TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016    1985 12:00:00                1              327
## 2 1503960366 4/13/2016    1797 12:00:00                2              384
## 3 1503960366 4/15/2016    1745 12:00:00                1              412
## 4 1503960366 4/16/2016    1863 12:00:00                2              340
## 5 1503960366 4/17/2016    1728 12:00:00                1              700
## 6 1503960366 4/19/2016    2035 12:00:00                1              304
```

```
## TotalTimeInBed
## 1 346
## 2 407
## 3 442
## 4 367
## 5 712
## 6 320
```

Also, for the Activity minutes and BMI, the tables dailyActivity\_merged and weightLogInfo\_new can be merged to give the new data frame “activity\_weight”.

```
activity_weight <-merge.data.frame(dailyActivity_merged, weightLogInfo_new, by = c("Id", "ActivityDate"),
head(activity_weight)
```

```
##      Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366 5/2/2016 14727 9.71 9.71
## 2 1503960366 5/3/2016 15103 9.66 9.66
## 3 1927972279 4/13/2016 356 0.25 0.25
## 4 2873212765 4/21/2016 8859 5.98 5.98
## 5 2873212765 5/12/2016 7566 5.11 5.11
## 6 4319703577 4/17/2016 29 0.02 0.02
## LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1 0 3.21 0.57
## 2 0 3.73 1.05
## 3 0 0.00 0.00
## 4 0 0.13 0.37
## 5 0 0.00 0.00
## 6 0 0.00 0.00
## LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1 5.92 0.00 41
## 2 4.88 0.00 50
## 3 0.25 0.00 0
## 4 5.47 0.01 2
## 5 5.11 0.00 0
## 6 0.02 0.00 0
## FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories Time
## 1 15 277 798 2004 11:59:59
## 2 24 254 816 1990 11:59:59
## 3 0 32 986 2151 1:08:52
## 4 10 371 1057 1970 11:59:59
## 5 0 268 720 1431 11:59:59
## 6 0 3 1363 1464 11:59:59
## WeightKg WeightPounds Fat BMI IsManualReport LogId
## 1 52.6 115.9631 22 22.65 True 1.462234e+12
## 2 52.6 115.9631 NA 22.65 True 1.462320e+12
## 3 133.5 294.3171 NA 47.54 False 1.460510e+12
## 4 56.7 125.0021 NA 21.45 True 1.461283e+12
## 5 57.3 126.3249 NA 21.69 True 1.463098e+12
## 6 72.4 159.6147 25 27.45 True 1.460938e+12
```

Also to see the relation of sleep quality we need to merge sleepDay\_new and dailyIntensities\_merged

```
sleeprelation <- merge.data.frame(dailyIntensities_merged, sleepDay_new, by=c("Id", "ActivityDay"))
```

Relation between sedentary minutes(non-active minutes) and BMI can be found. Notice that there is no Total active minutes, but only lightly, fairly and very active minutes, whose sum is actually total active minutes. So a column is added to the dataframe by summing the VeryActiveMinutes, FairlyActiveMinutes and LightlyActiveMinutes.

```
activity_weight$TotalActiveMinutes <- activity_weight$VeryActiveMinutes+activity_weight$FairlyActiveMinutes+activity_weight$LightlyActiveMinutes
head(activity_weight)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    5/2/2016    14727         9.71         9.71
## 2 1503960366    5/3/2016    15103         9.66         9.66
## 3 1927972279    4/13/2016     356         0.25         0.25
## 4 2873212765    4/21/2016    8859         5.98         5.98
## 5 2873212765    5/12/2016    7566         5.11         5.11
## 6 4319703577    4/17/2016     29         0.02         0.02
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0              3.21              0.57
## 2                        0              3.73              1.05
## 3                        0              0.00              0.00
## 4                        0              0.13              0.37
## 5                        0              0.00              0.00
## 6                        0              0.00              0.00
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                5.92                0.00                41
## 2                4.88                0.00                50
## 3                0.25                0.00                 0
## 4                5.47                0.01                 2
## 5                5.11                0.00                 0
## 6                0.02                0.00                 0
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories      Time
## 1                 15                 277              798    2004 11:59:59
## 2                 24                 254              816    1990 11:59:59
## 3                  0                  32              986    2151  1:08:52
## 4                 10                 371             1057    1970 11:59:59
## 5                  0                 268              720    1431 11:59:59
## 6                  0                  3             1363    1464 11:59:59
##   WeightKg WeightPounds Fat   BMI IsManualReport      LogId
## 1     52.6    115.9631  22 22.65          True 1.462234e+12
## 2     52.6    115.9631  NA 22.65          True 1.462320e+12
## 3    133.5    294.3171  NA 47.54         False 1.460510e+12
## 4     56.7    125.0021  NA 21.45          True 1.461283e+12
## 5     57.3    126.3249  NA 21.69          True 1.463098e+12
## 6     72.4    159.6147  25 27.45          True 1.460938e+12
##   TotalActiveMinutes
## 1                 333
## 2                 328
## 3                  32
## 4                 383
## 5                 268
## 6                   3
```

The data has been analysed. All possible trends and relations have been investigated.

---

## Sharing the data

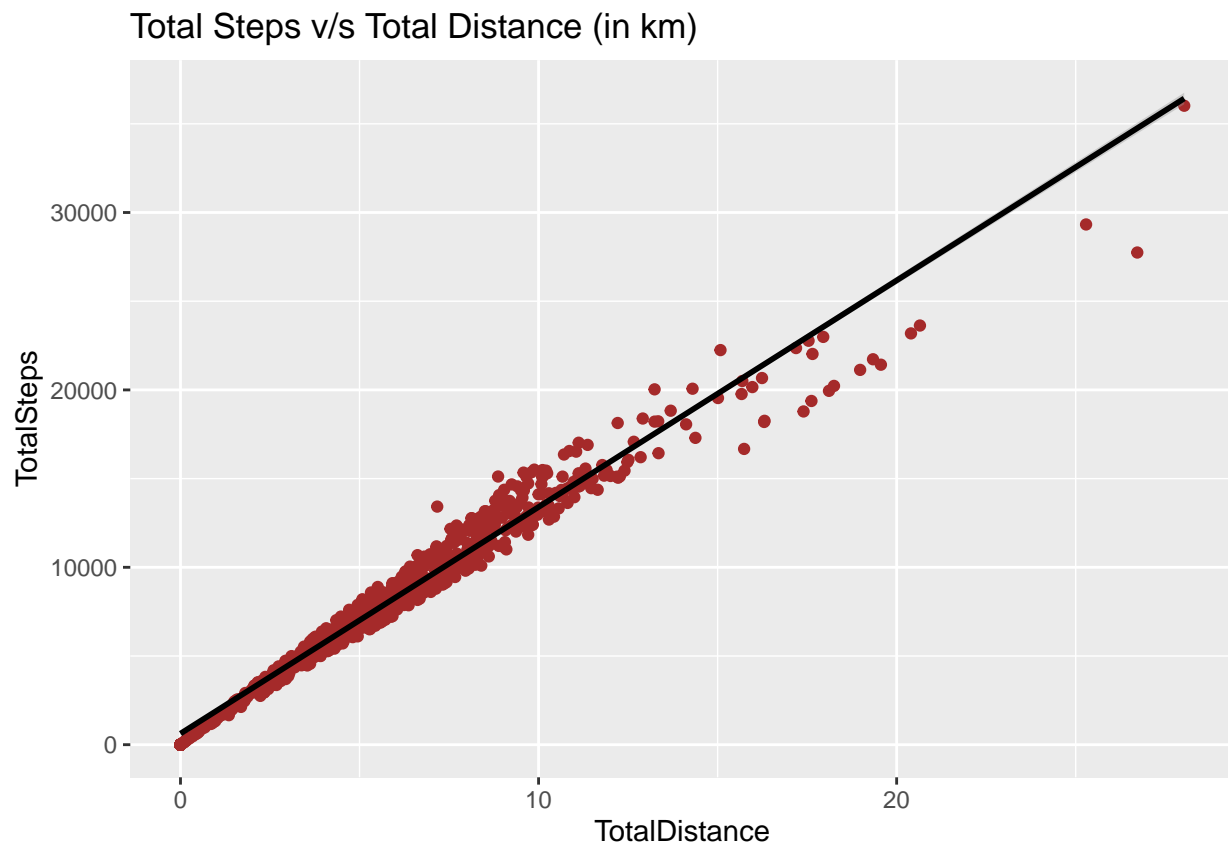
For visualising the data, R Studio is used.

ggplot is used for visualising data.

Total Steps v/s Total Distance

```
ggplot(data=dailyActivity_merged)+  
  geom_point(mapping=aes(x=TotalDistance, y=TotalSteps), color="brown")+  
  geom_smooth(mapping=aes(x=TotalDistance, y=TotalSteps), color="black",method="lm")+  
  ggtitle("Total Steps v/s Total Distance (in km)")
```

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

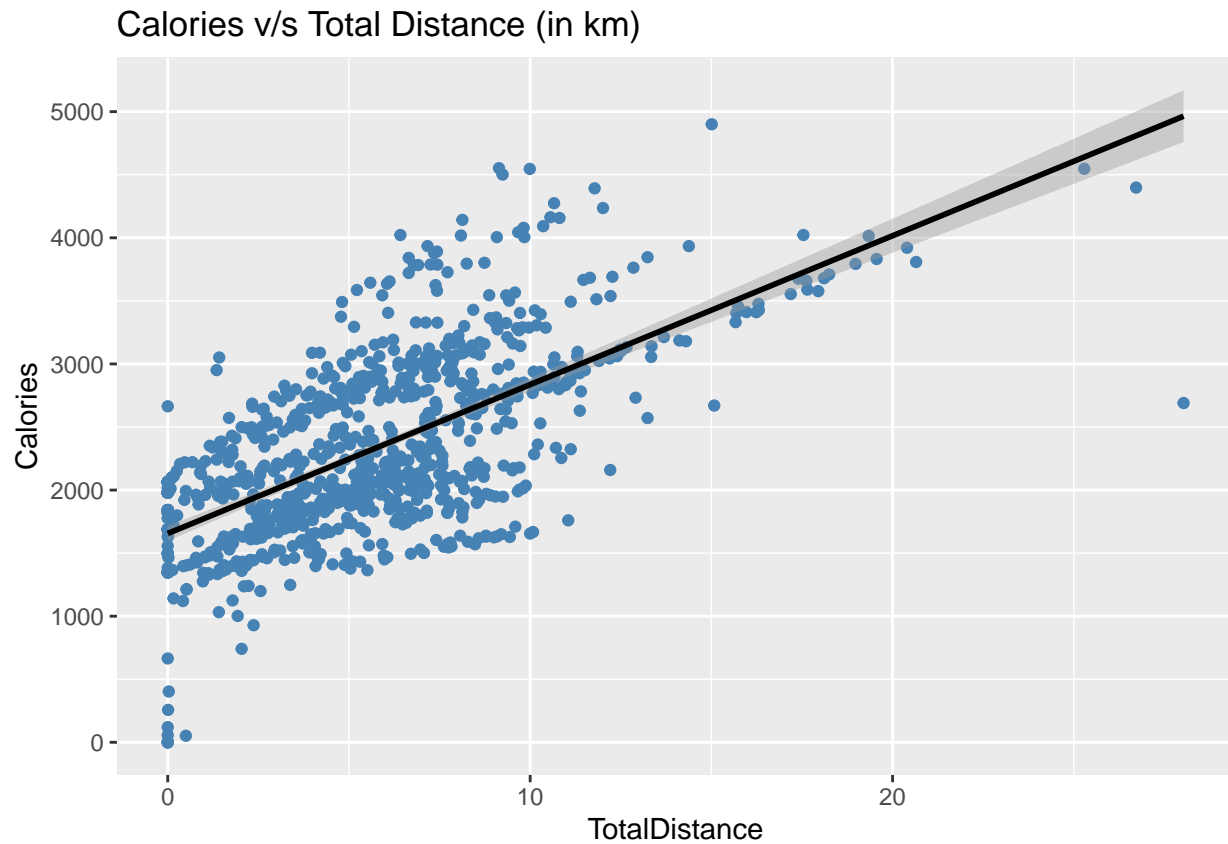


We can see the total steps and total distance have a direct positive relation, i.e if you covering more distance you are walking more and more the number of steps.

Calories v/s Total Distance

```
ggplot(data=dailyActivity_merged)+
  geom_point(mapping=aes(x=TotalDistance, y= Calories), color="steelblue")+
  geom_smooth(mapping=aes(x=TotalDistance, y=Calories), color="black", method="lm")+
  ggtitle("Calories v/s Total Distance (in km)")
```

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

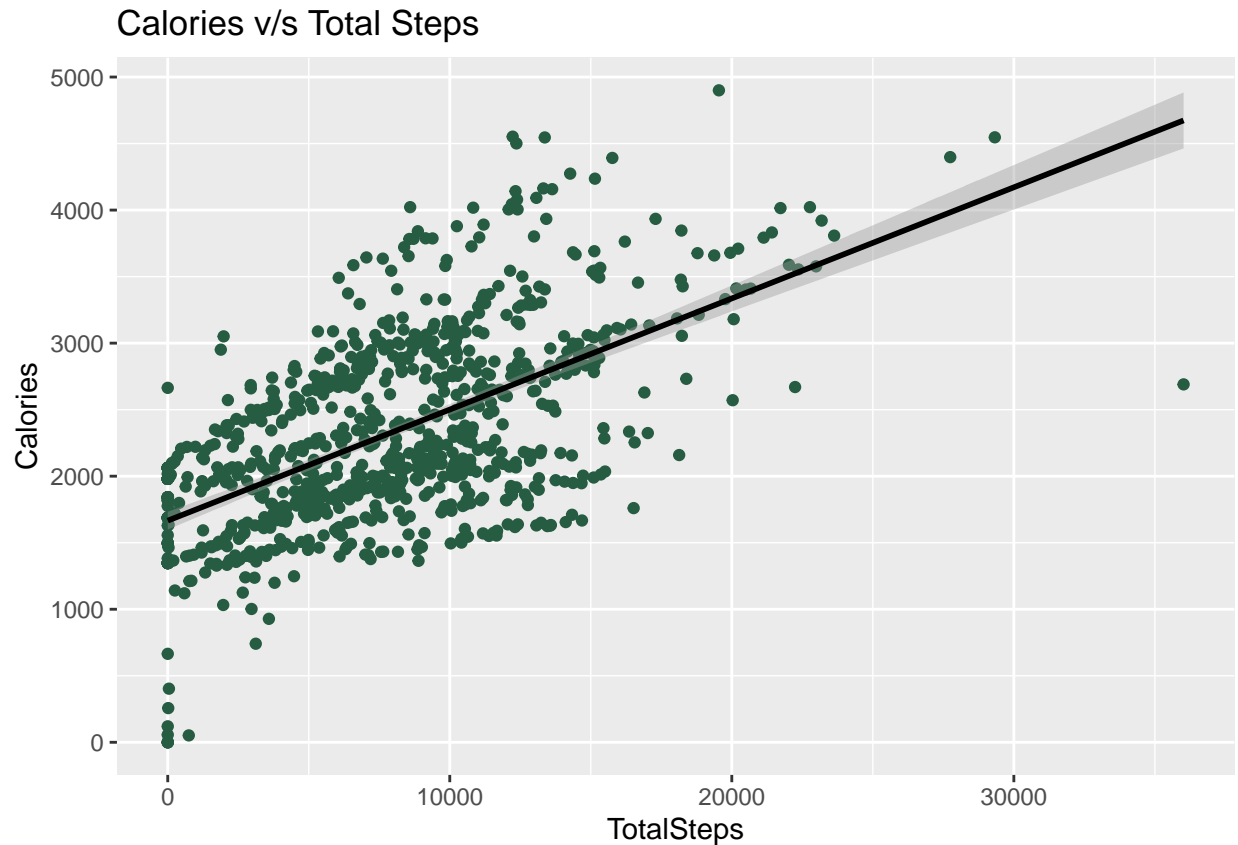


Here also a direct positive relation is found, i.e if you are covering more distance, higher the number of calories you will burn.

Calories v/s Total Steps

```
ggplot(data=dailyActivity_merged)+
  geom_point(mapping=aes(x=TotalSteps, y=Calories), color="#265c41")+
  geom_smooth(mapping=aes(x=TotalSteps, y=Calories), color="black", method="lm")+
  ggtitle("Calories v/s Total Steps")
```

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



A direct positive relation is found here too i.e more steps you take, higher the number of calories you will burn.

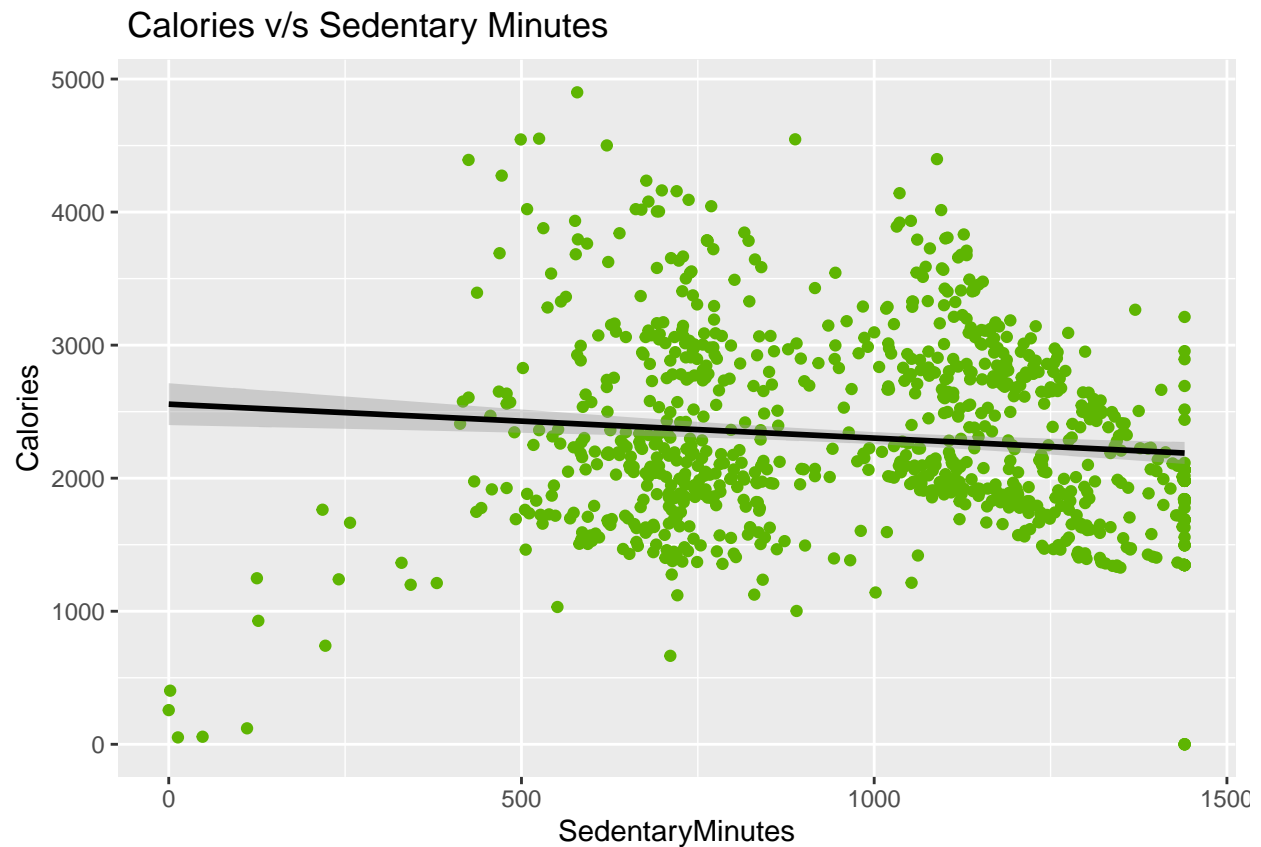
This was expected as we have already seen the direct relation between total steps and total distance.

### Calories v/s Activity minutes

#### 1. Calories v/s Sedentary Minutes

```
ggplot(data=dailyActivity_merged)+
  geom_point(mapping=aes(x=SedentaryMinutes, y= Calories), color="#5eb502")+
  geom_smooth(mapping=aes(x=SedentaryMinutes, y= Calories), color="black",method="lm")+
  ggtitle(" Calories v/s Sedentary Minutes")
```

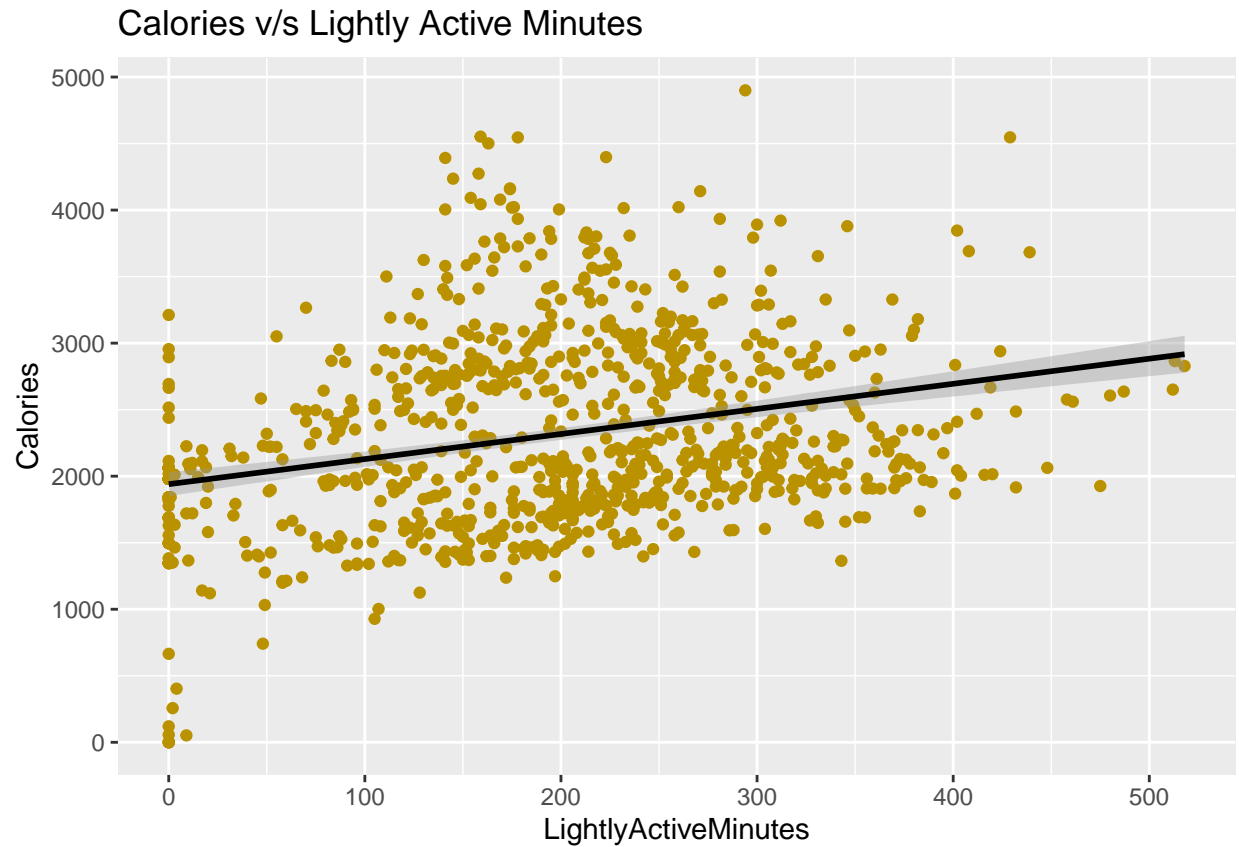
```
## 'geom_smooth()' using formula 'y ~ x'
```



We can see there is a small decrease in calories burnt more the amount of sedentary/ non-active minutes. 2.  
Calories v/s Lightly Active Minutes

```
ggplot(data=dailyActivity_merged)+
  geom_point(mapping=aes(x=LightlyActiveMinutes, y= Calories), color="#ba9200")+
  geom_smooth(mapping=aes(x=LightlyActiveMinutes, y= Calories), color="black",method="lm")+
  ggtitle("Calories v/s Lightly Active Minutes")
```

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

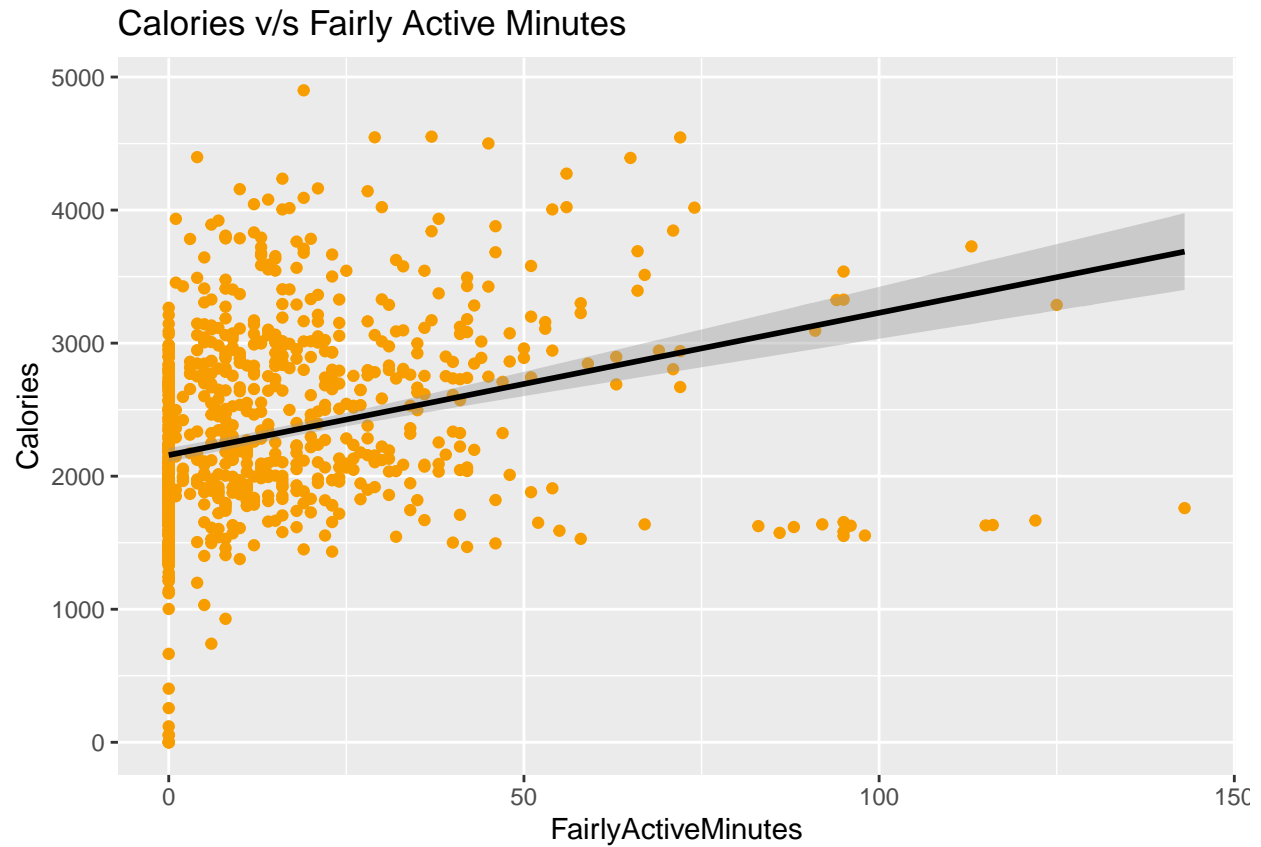


### 3. Calories v/s Fairly Active Minutes

```
ggplot(data=dailyActivity_merged)+  
  geom_point(mapping=aes(x=FairlyActiveMinutes, y= Calories), color="#f79d00")+  
  geom_smooth(mapping=aes(x=FairlyActiveMinutes, y= Calories), color="black",method="lm")+  
  ggtitle("Calories v/s Fairly Active Minutes")
```

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



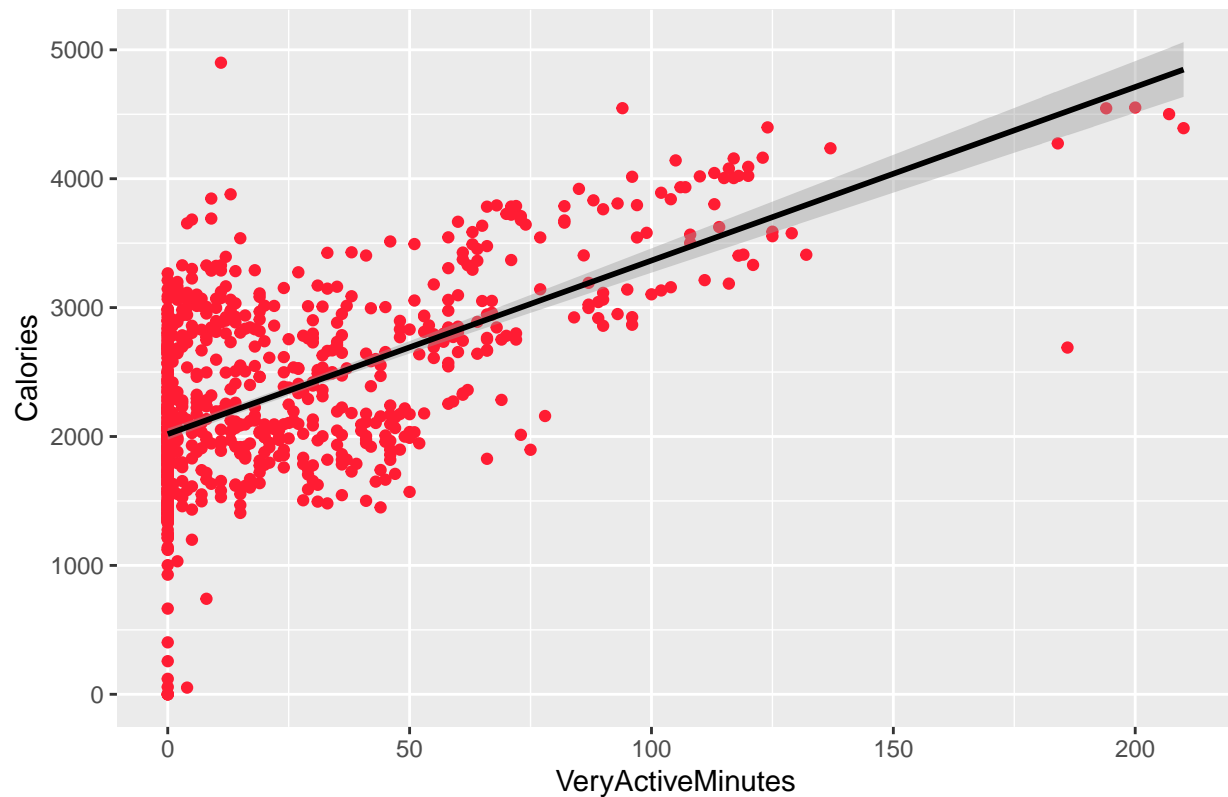


#### 4. Calories v/s Very Active Minutes

```
ggplot(data=dailyActivity_merged)+
  geom_point(mapping=aes(x=VeryActiveMinutes, y= Calories), color="#ff1c33")+
  geom_smooth(mapping=aes(x=VeryActiveMinutes, y= Calories), color="black", method="lm")+
  ggtitle("Calories v/s Very Active Minutes")
```

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

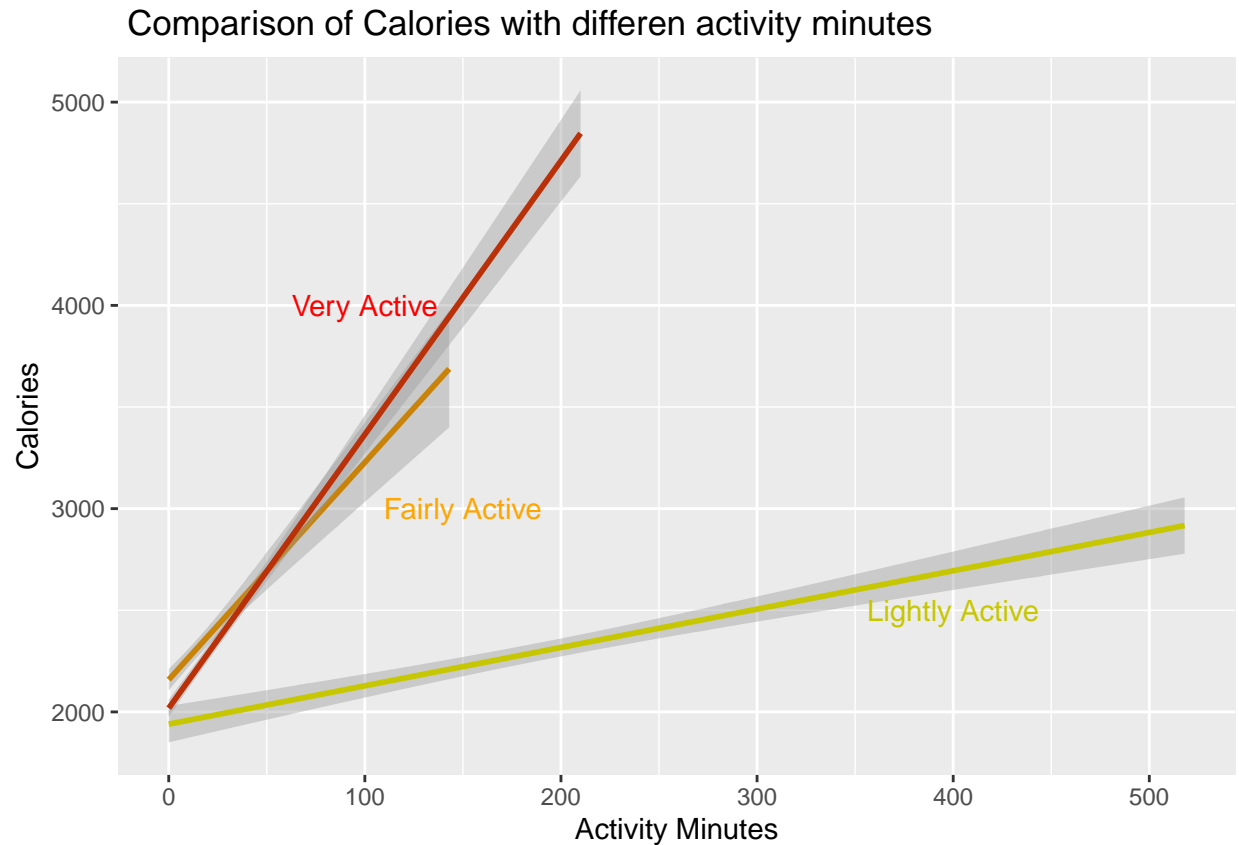
Calories v/s Very Active Minutes



5.

```
ggplot(data=dailyActivity_merged)+
  geom_smooth(mapping=aes(x=LightlyActiveMinutes, y= Calories), color="#c7c700",method="lm")+
  geom_smooth(mapping=aes(x=FairlyActiveMinutes, y= Calories), color="#c98308",method="lm")+
  geom_smooth(mapping=aes(x=VeryActiveMinutes, y= Calories), color="#ba3009", method="lm")+
  labs(y="Calories", x="Activity Minutes")+
  annotate(geom="text", x=400, y=2500, label="Lightly Active",color="#c7c700")+
  annotate(geom="text", x=150, y=3000, label="Fairly Active", color="Orange" )+
  annotate(geom="text", x=100, y=4000, label="Very Active", color="red")+
  ggtitle(" Comparison of Calories with differen activity minutes")
```

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



Calories burnt has a positive relation with Activity minutes.

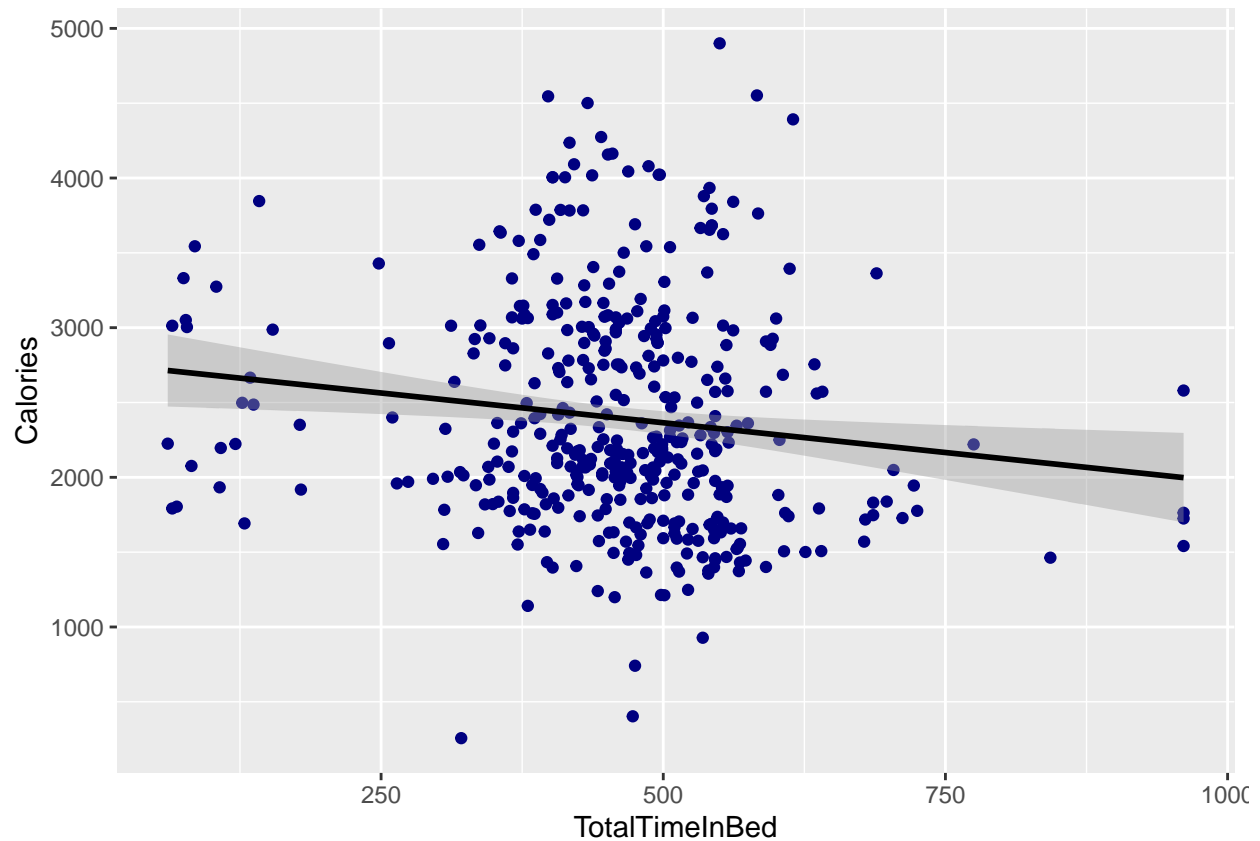
We can see a trend here.

The value of the slope increase with increasing intensity of activity minutes. That means more calories are burnt when you are very active than when you are fairly active which in turn is more than lightly active for the same amount of time.

Calories v/s Total time in bed

```
ggplot(data=sleep_calories)+
  geom_point(mapping=aes(x=TotalTimeInBed, y=Calories), color="navyblue")+
  geom_smooth(mapping=aes(x=TotalTimeInBed, y=Calories), color="black", method="lm")
```

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

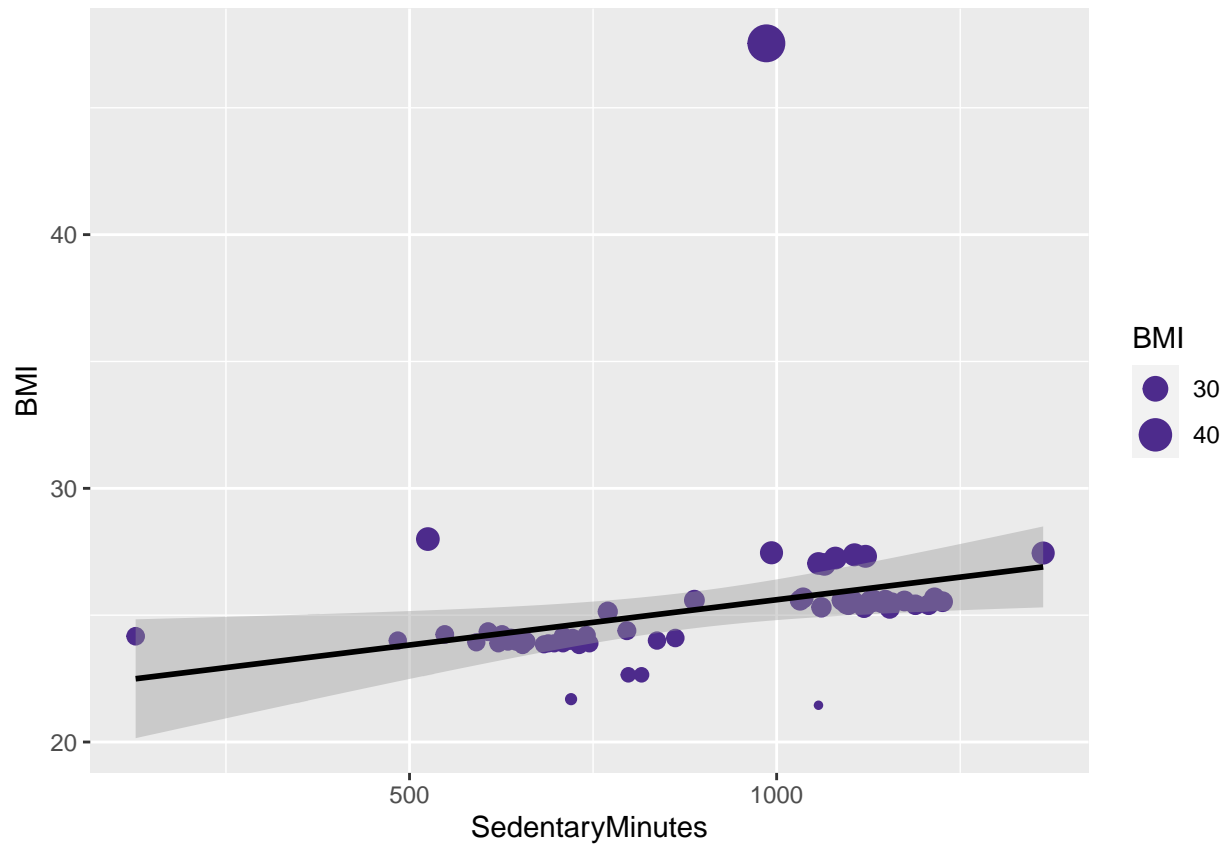


It can be observed, more time you spend in bed, lesser calories you burn.

BMI v/s Non-Activity Minutes

```
ggplot(data=activity_weight)+
  geom_point(mapping=aes(x=SedentaryMinutes, y=BMI, size=BMI), color="#4d2b8c")+
  geom_smooth(mapping=aes(x=SedentaryMinutes, y=BMI), color="Black", method="lm")
```

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

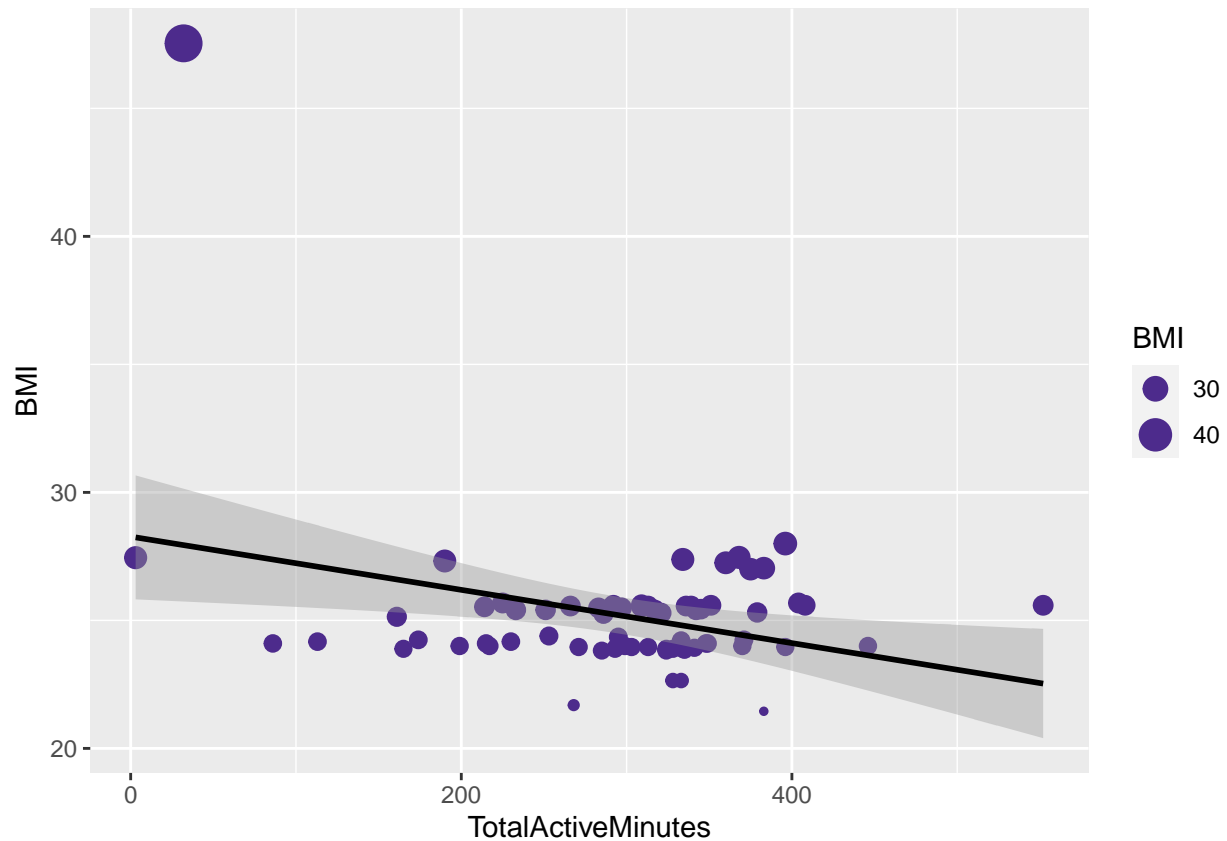


A person having a larger BMI has more non-activity minutes.

BMI v/s Total-Activity Minutes

```
ggplot(data=activity_weight)+
  geom_point(mapping=aes(x=TotalActiveMinutes, y=BMI, size=BMI), color="#4d2b8c")+
  geom_smooth(mapping=aes(x=TotalActiveMinutes, y=BMI), color="Black", method="lm")
```

## 'geom\_smooth()' using formula 'y ~ x'

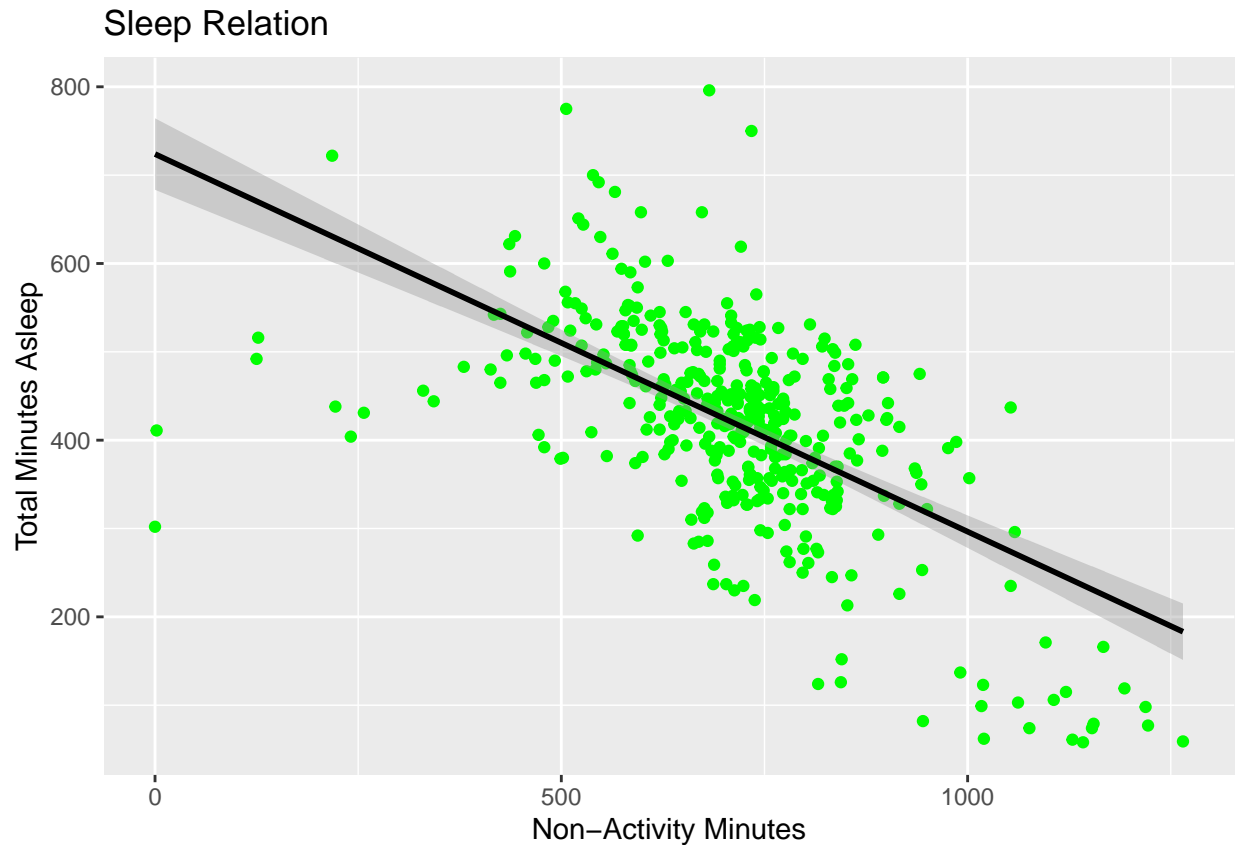


A person having a larger BMI has less activity minutes.

#### Sleep Relation

```
ggplot(data=sleeprelation)+
  geom_point(mapping=aes(x=SedentaryMinutes, y=TotalMinutesAsleep), color = "green")+
  geom_smooth(mapping=aes(x=SedentaryMinutes, y=TotalMinutesAsleep), color="Black", method="lm")+
  labs(y="Total Minutes Asleep", x="Non-Activity Minutes")+
  ggtitle("Sleep Relation")
```

## 'geom\_smooth()' using formula 'y ~ x'



There is a negative relation which indicates the higher non-active minutes can lead to lesser sleep minutes, which indicates deterioration in sleeping level.

(NOTE: The above two results have a very small sample size)

## Acting on the data

### Trends and Relations

From the above analysis we can establish the following relations.

1. Great distance travelled means more steps taken and higher amount of calories burnt.
2. Greater active minutes leads to higher amount of calories burnt.
3. People who are not really active have an unhealthy lifestyle with lesser sleep amount and high BMI.
4. People who burn low calories tend to spend more time on bed.

### Recommendations to expand business

1. Target members should be people who have lower active minutes and higher sedentary minutes.
2. Timely alerts and reminders can be given to motivate them.
3. Can introduce a reward system on the app to encourage users to recommend to friends/family members.

4. A free trial for 2 weeks/ 4 weeks can be given to test the products.
5. Can give personalised feedback according to each one's BMI and activity levels.

#### Recommendations for Bellabeat products

1. The trackers i.e Leaf and Time should be advertised more which should include its benefits.
2. With very minimal to no data on hydration levels, production of Spring can be reduced.
3. For memberships, referrals from person to person can be given an initial discount.