

# Case Study (Bellabeat Fitness Company)

By Israel Yinka

2022-10-21

## Scenario

I am a junior data analyst on the marketing analyst team at Bellabeat, a high-tech manufacturer of feminine health products. A successful little firm, Bellabeat has the ability to expand and take over the global market for smart gadgets. Urka Sren, co-founder and chief creative officer of Bellabeat, believes that examining fitness data from smart devices might provide the company with new development opportunities. To better understand how people engage with their smart devices, I've been asked to focus on one of Bellabeat's products and analyze data from linked devices. The information I uncover will then have an impact on the company's marketing strategy.

## Bellabeat Products

**Bellabeat App:** Access to health data on one's exercise, sleep, stress, menstrual cycle, and mindfulness practices is made available to users. Users may utilize this knowledge to understand their habits and make healthy decisions. The Bellabeat app is compatible with their collection of smart wellness products.

**Leaf:** The classic wellness tracker from Bellabeat is offered as a clip, necklace, or bracelet. The Leaf tracker connects to the Bellabeat app to monitor activity, sleep, and stress.

**Time:** This health watch has the typical look of a wristwatch while tracking user activity, sleep, and stress. The Bellabeat app is connected to the Time watch in order to provide you insights into your daily wellness.

**Spring:** This water bottle makes use of a daily water intake measurement method.

## Ask

In order to understand important (successful and failed) consumer trends in the worldwide market for smart devices, the business task is to examine non-Bellabeat smart device user data. The results of this study will also have an impact on the next Bellabeat marketing plans and assist identifying how to apply these trends to Bellabeat clients. Such insights are applied to the Bellabeat App and to future products in order to optimize profits and growth for the company and to capitalize on Bellabeat's rapidly growing user base in the smart device/tech-wellness sector. Urka Sren, Sando Mur (the other co-founder and a significant member of the Bellabeat executive team), the Bellabeat executive team, the Bellabeat Marketing Analytics Team, and the Bellabeat stakeholders are the key Bellabeat players.

## Prepare Data

**Data Source:** FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius).

**Data Not Reliable:** Since just 30 carefully selected individuals, as opposed to the 30+ million FitBit users overall, are included in this data, it is skewed. More specifically, I would have preferred at least a year's worth of data because the data was only gathered over the course of two months, which is insufficient time to discover reliable and accurate trends.

**Data Not Original:** The data set was created by survey respondents who responded to a distributed survey made possible by Amazon Mechanical Turk. It would have been better if FitBit had directly delivered the information.

**Data Not Comprehensive:** I would prefer to have information covering a year's duration of time; the data was only acquired during a two-month period, which is insufficient. Furthermore, it is hard to tell whether the people were picked at random or with prejudice.

**Data Not Current:** Since the data were collected six years ago, they do not represent current trends in data gathering.

### Importing Dataset:

```
daily_activity <- read.csv("dailyActivity_merged.csv")
sleep_daily <- read.csv("sleepDay_merged.csv")
daily_calories <- read.csv("dailyCalories_merged.csv")
```

## Process Data

**Verifying Data:** I had to make sure that the data sets had loaded correctly and without any mistakes.

```
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
```

```
head(sleep_daily)
```

```
##           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(daily_calories)
```

```
##      Id ActivityDay Calories
## 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
```

```
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"
```

```
colnames(sleep_daily)
```

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

```
colnames(daily_calories)
```

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

```
View(daily_activity) View(sleep_daily) View(daily_calories)
```

I learned through looking over the data sets that some people didn't want to track their sleep, and some didn't want to do it for the full month. This will makes conducting a full analysis much more difficult than expected. Additionally, several days had zero statistics recorded that had to be deleted.

```
daily_activity_new <- daily_activity %>%
  filter(TotalSteps !=0)
```

```
sleep_daily_new <- sleep_daily %>%
  filter(TotalSleepRecords !=0)
```

```
daily_calories_new <- daily_calories %>%
  filter (Calories !=0)
```

I also divided the date and time column into two independent columns in the sleep data set.

```
sleep_daily_new_2 <- sleep_daily_new %>%
  separate(SleepDay, c("Date", "Time"), " ")
```

```
## 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, ...].
```

```
View(sleep_daily_new)
```

```
n_distinct(daily_activity_new$Id)
```

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

```
n_distinct(daily_calories_new$Id)
```

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

```
n_distinct(sleep_daily_new_2$Id)
```

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

Only 24 individuals entered their sleep data. Furthermore, although the data source claims that the sample size consists of 30 persons, the daily activity data actually lists 33 individuals.

```
nrow(daily_activity_new)
```

```
## [1] 863
```

```
nrow(daily_calories_new)
```

```
## [1] 936
```

```
nrow(sleep_daily_new_2)
```

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

```
nrow(unique(daily_activity_new))
```

```
## [1] 863
```

```
nrow(unique(daily_calories_new))
```

```
## [1] 936
```

```
nrow(unique(sleep_daily_new_2))
```

```
## [1] 410
```

I created unique rows for the data sets to further tidy it up.

```
sleep_daily_final <- unique(sleep_daily_new_2)
daily_activity_final <- unique(daily_activity_new)
daily_calories_final <- unique(daily_calories_new)
```

## Analyse Data

Each data collection had to just contain the columns I would use for my more in-depth analysis.

```
daily_activity_final_2 <- daily_activity_final %>%
  select(Id, ActivityDate, TotalSteps, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes, Calories)
sleep_daily_final_2 <- sleep_daily_final %>%
  select(Id, Date, TotalMinutesAsleep, TotalTimeInBed)
```

```

mean(daily_activity_final_2$TotalSteps)

## [1] 8319.393
mean(daily_activity_final_2$VeryActiveMinutes)

## [1] 23.01506
mean(daily_activity_final_2$FairlyActiveMinutes)

## [1] 14.7752
mean(daily_activity_final_2$LightlyActiveMinutes)

## [1] 210.0162
mean(daily_activity_final_2$SedentaryMinutes)

## [1] 955.7532
mean(daily_activity_final_2$Calories)

## [1] 2361.295
mean(sleep_daily_final_2$TotalMinutesAsleep)

## [1] 419.1732
mean(sleep_daily_final_2$TotalTimeInBed)

## [1] 458.4829

```

**Trends:** \* On average, a person takes 8053 total steps each day. \* On average, people spend 23.02 minutes being Very Active, 14.78 minutes being Fairly Active, 210 minutes being Lightly Active, and 955.6 minutes being Sedentary. \* Compared to the 419.2 minutes spent sleeping, people spent 458.5 minutes on average in bed. The individuals got somewhat less sleep than the minimum number of 7 hours recommended for adults, or an average of 955.8 minutes each day, or nearly 16 hours per day.

**Questions:** \* In what ways may Bellabeat promote regularity in user login and data tracking? \* How would Bellabeat help the average user reach the necessary daily Total Steps goal? \* How would Bellabeat help users get a better night's sleep?

## Share

Using graphics to convey key thoughts and discoveries:

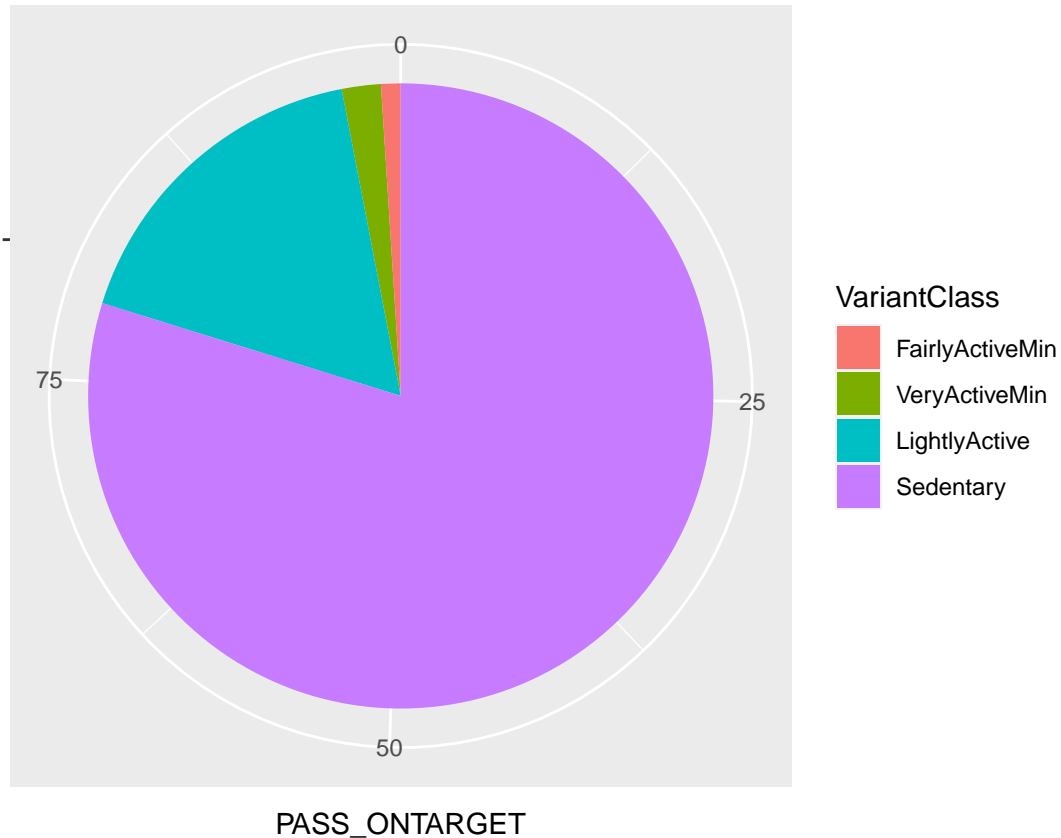
```

VeryActiveMin <- sum(daily_activity_final_2$VeryActiveMinutes)
FairlyActiveMin <- sum(daily_activity_final_2$FairlyActiveMinutes)
LightlyActiveMin <- sum(daily_activity_final_2$LightlyActiveMinutes)
SedentaryMin <- sum(daily_activity_final_2$SedentaryMinutes)
TotalMin <- VeryActiveMin + FairlyActiveMin + LightlyActiveMin + SedentaryMin

dataset <- tribble(
  ~ VariantClass, ~ PASS_ONTARGET,
  "VeryActiveMin", 2,
  "FairlyActiveMin", 1,
  "LightlyActive", 17,
  "Sedentary", 79,
) %>%
  mutate(VariantClass = fct_reorder(VariantClass, PASS_ONTARGET))

```

```
ggplot(dataset, aes(x = "", y = PASS_ONTARGET, fill = VariantClass)) +
  geom_col(width = 1) +
  coord_polar(theta = "y") +
  xlab(NULL)
```



```
my_plot_a <- ggplot(data=daily_activity_final_2) +
  geom_point(mapping=aes(x=TotalSteps, y=Calories), color="red") +
  geom_smooth(mapping=aes(x=TotalSteps, y=Calories)) +
  labs(title="The Relationship Between Total Steps and Calories Burned", x="Total Steps", y="Calories Burned")

combined_data <- merge(daily_activity_final_2, sleep_daily_final_2, by="Id")

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

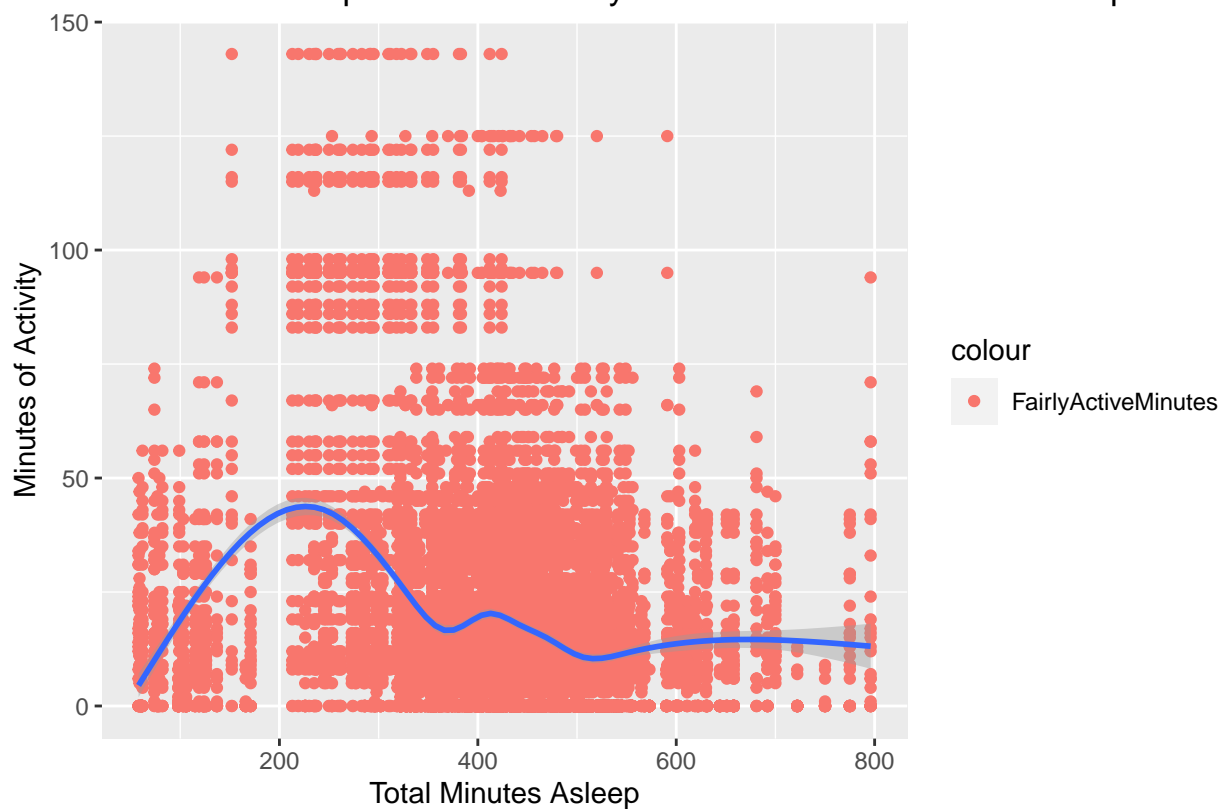


```
my_plot_b <- ggplot(data=combined_data) +
  geom_point(mapping=aes(x=TotalMinutesAsleep, y=FairlyActiveMinutes, color="FairlyActiveMinutes")) +
  geom_smooth(mapping=aes(x=TotalMinutesAsleep, y=FairlyActiveMinutes, regLineColor="blue")) +
  labs(title="The Relationship Between Activity Levels and Total Minutes Asleep", x="Total Minutes Asleep")
```

```
## Warning: Ignoring unknown aesthetics: regLineColor
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

## The Relationship Between Activity Levels and Total Minutes Asleep



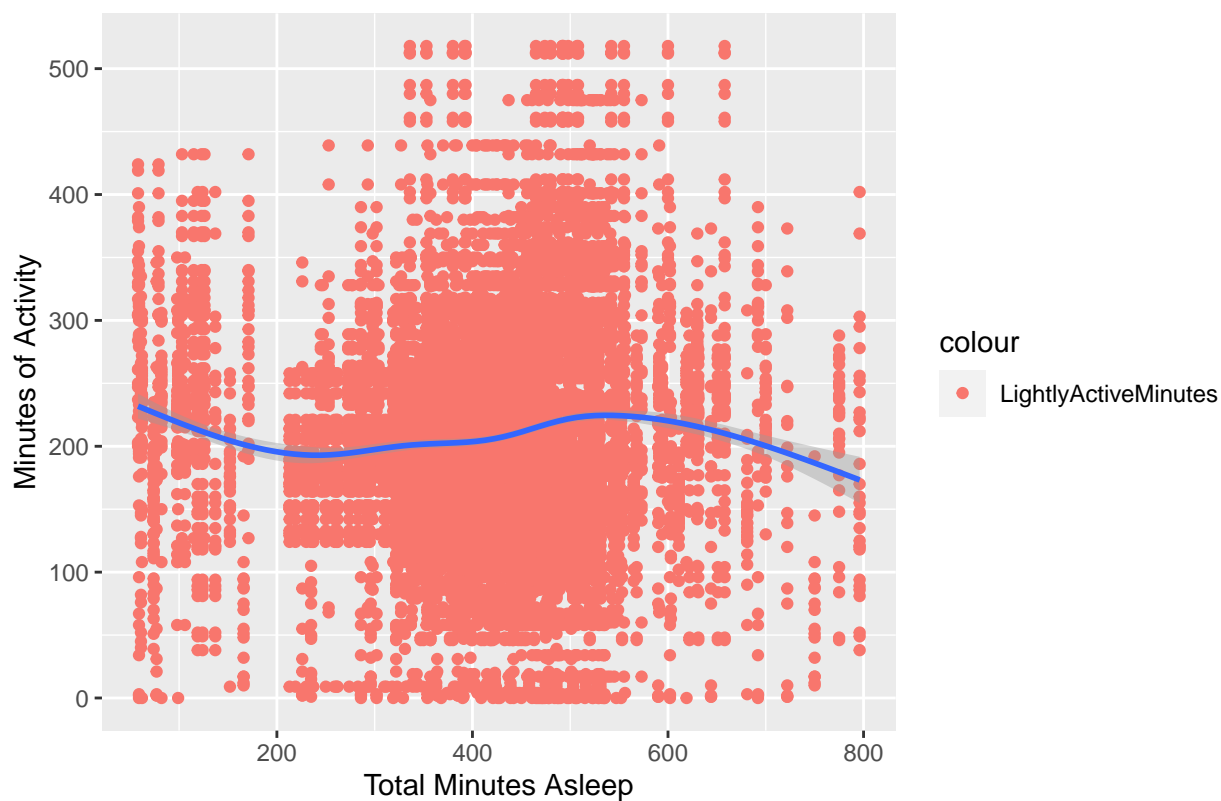
```
my_plot_c <- ggplot(data=combined_data) +
  geom_point(mapping=aes(x=TotalMinutesAsleep, y=LightlyActiveMinutes, color="LightlyActiveMinutes")) +
  geom_smooth(mapping=aes(x=TotalMinutesAsleep, y=LightlyActiveMinutes, regLineColor="blue")) +
  labs(title="The Relationship Between Activity Levels and Total Minutes Asleep", x="Total Minutes Asleep")
```

```
## Warning: Ignoring unknown aesthetics: regLineColor
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



## The Relationship Between Activity Levels and Total Minutes Asleep

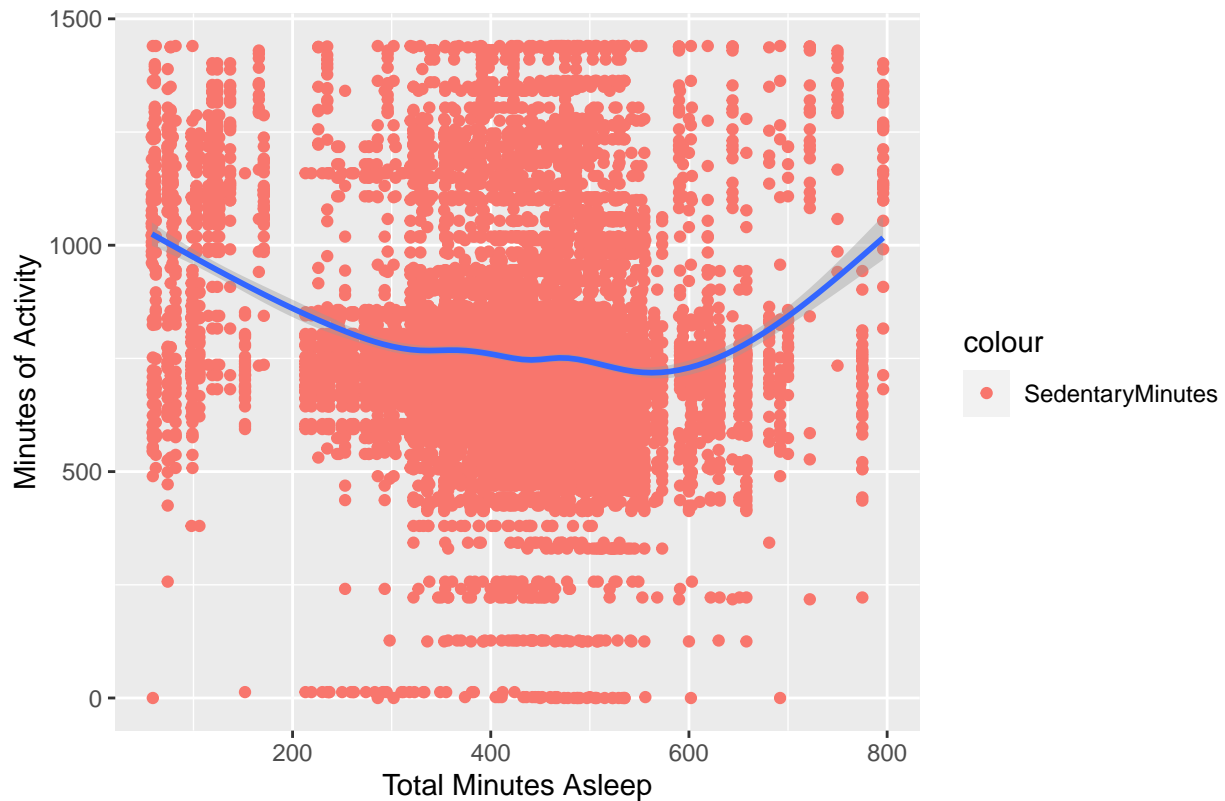


```
my_plot_d <- ggplot(data=combined_data) +
  geom_point(mapping=aes(x=TotalMinutesAsleep, y=SedentaryMinutes, color="SedentaryMinutes")) +
  geom_smooth(mapping=aes(x=TotalMinutesAsleep, y=SedentaryMinutes, regLineColor="blue")) +
  labs(title="The Relationship Between Activity Levels and Total Minutes Asleep", x="Total Minutes Asleep")
```

```
## Warning: Ignoring unknown aesthetics: regLineColor
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

## The Relationship Between Activity Levels and Total Minutes Asleep



### Act

#### Recommendation:

- Encourage consumers to utilize a wifi-connected scale rather than manually entering their weights in order to collect more data for a thorough study.
- Organize campaigns encouraging users to engage in short, active workouts throughout the week and lengthier workouts on the weekends, particularly on Sunday when we observe the lowest step counts and highest inactive minutes.
- A point-based reward system can be used in conjunction with a healthy educational campaign. Users who complete the whole week's workout will earn Bellabeat points redeemable for merchandise or subscriptions.
- When a user has been sitting still for a long time, a gadget like the Leaf health tracker can beat or vibrate to remind them to move about. In a similar manner, it can alert the user to the need for sleep after detecting an extended period of time spent awake in bed.