Bellabeat Analysis

2022-06-27

Bellabeat

Company Background

Bellabeat is a small manufacturer of smart health devices marketed for women. Cofounder and Chief Creative Officer, Urška Sršen, would like an analysis of Bellabeat device data. Urška Sršen, along with cofounder Sando Mur and the Bellabeat marketing analytics team, hope to gain insight into an effective marketing strategy as a result of this analysis.

The Task at Hand

I hope to identify trends for both Bellabeat and other similar products and help develop marketing based on these trends.

The Data

The data that I used is a CSV formatted Kaggle data set titled: FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): It can be located at this website https://www.kaggle.com/datasets/arashnic/fitbit. The data set contains personal fitness tracker from thirty FitBit users that consented to the submission of their data. Their data includes minute-level output for physical activity, heart rate, and sleep monitoring. It also includes information about daily activity, steps, and heart rate. There are some limitations in this data due to only thirty users submitting their data.

Processing and Analyzing

I first started by installing the cleaning and analysis packages I needed and loaded their libraries.

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("lubridate")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("here")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("skimr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("janitor")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library("tidyverse")
```

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

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     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("here")
```

```
## here() starts at /cloud/project
```

```
library("skimr")
library("janitor")
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Then, I loaded the CSV files.

```
daily_activity <- read.csv("dailyActivity_merged.csv")
daily_calories <- read.csv("dailyCalories_merged.csv")
daily_intensities <- read.csv("dailyIntensities_merged.csv")
daily_steps <- read.csv("dailySteps_merged.csv")
sleep_day <- read.csv("sleepDay_merged.csv")
weight_log <- read.csv("weightLogInfo_merged.csv")
```

Then, I took a look at the data frames.

```
glimpse(daily_activity)
```

```
## Rows: 940
## Columns: 15
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate             <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ TotalSteps               <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019~
## $ TotalDistance            <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ TrackerDistance          <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
```

```
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance       <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ LightActiveDistance      <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ SedentaryActiveDistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes        <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ FairlyActiveMinutes      <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ LightlyActiveMinutes     <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ SedentaryMinutes         <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ Calories                 <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~
```

```
glimpse(daily_calories)
```

```
## Rows: 940
## Columns: 3
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~
## $ Calories    <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1786, 1775~
```

```
glimpse(daily_intensities)
```

```
## Rows: 940
## Columns: 10
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDay              <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ SedentaryMinutes         <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ LightlyActiveMinutes     <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ FairlyActiveMinutes      <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ VeryActiveMinutes        <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ SedentaryActiveDistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ LightActiveDistance      <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ VeryActiveDistance       <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
```

```
glimpse(daily_steps)
```

```
## Rows: 940
## Columns: 3
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~
## $ StepTotal   <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 15506, 1054~
```

```
glimpse(sleep_day)
```

```
## Rows: 413
## Columns: 5
## $ Id                 <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay           <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~
## $ TotalSleepRecords  <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~
## $ TotalTimeInBed     <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~
```

```
glimpse(weight_log)
```

```
## Rows: 67
## Columns: 8
## $ Id             <dbl> 1503960366, 1503960366, 1927972279, 2873212765, 2873212~
## $ Date           <chr> "5/2/2016 11:59:59 PM", "5/3/2016 11:59:59 PM", "4/13/2~
```

```
## $ WeightKg       <dbl> 52.6, 52.6, 133.5, 56.7, 57.3, 72.4, 72.3, 69.7, 70.3, ~
## $ WeightPounds   <dbl> 115.9631, 115.9631, 294.3171, 125.0021, 126.3249, 159.6~
## $ Fat            <int> 22, NA, NA, NA, NA, 25, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ BMI            <dbl> 22.65, 22.65, 47.54, 21.45, 21.69, 27.45, 27.38, 27.25,~
## $ IsManualReport <chr> "True", "True", "False", "True", "True", "True", "True"~
## $ LogId          <dbl> 1.462234e+12, 1.462320e+12, 1.460510e+12, 1.461283e+12,~
```

Then, I previewed the column names.

```
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(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
```

```
head(daily_intensities)
```

```
##           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(daily_steps)
```

```
##           Id ActivityDay StepTotal
## 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(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
```

```
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 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
```

I took another look at the column names for summary statistics and merging purposes.

```
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(daily_calories)
```

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

```
colnames(daily_intensities)
```

```
##  [1] "Id"                       "ActivityDay"
##  [3] "SedentaryMinutes"         "LightlyActiveMinutes"
##  [5] "FairlyActiveMinutes"      "VeryActiveMinutes"
##  [7] "SedentaryActiveDistance"  "LightActiveDistance"
##  [9] "ModeratelyActiveDistance" "VeryActiveDistance"
```

```
colnames(daily_steps)
```

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

```
colnames(sleep_day)
```

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

```
colnames(weight_log)
```

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

Then, I did calculations for a summary.

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

```
##    TotalSteps    TotalDistance    SedentaryMinutes    Calories
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :   0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:1828
##  Median : 7406   Median : 5.245   Median :1057.5   Median :2134
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   :2304
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.:2793
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :4900
```

```
daily_calories %>%
  select(Calories) %>%
```

```
  summary()
```

```
##     Calories
##  Min.   :   0
##  1st Qu.:1828
##  Median :2134
##  Mean   :2304
##  3rd Qu.:2793
##  Max.   :4900
```

```
daily_intensities %>%
  select(SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, VeryActiveMinutes) %>%
  summary()
```

```
##  SedentaryMinutes LightlyActiveMinutes FairlyActiveMinutes VeryActiveMinutes
##  Min.   :   0.0   Min.   :  0.0        Min.   :  0.00      Min.   :  0.00
##  1st Qu.: 729.8   1st Qu.:127.0        1st Qu.:  0.00      1st Qu.:  0.00
##  Median :1057.5   Median :199.0        Median :  6.00      Median :  4.00
##  Mean   : 991.2   Mean   :192.8        Mean   : 13.56      Mean   : 21.16
##  3rd Qu.:1229.5   3rd Qu.:264.0        3rd Qu.: 19.00      3rd Qu.: 32.00
##  Max.   :1440.0   Max.   :518.0        Max.   :143.00      Max.   :210.00
```

```
daily_steps %>%
  select(StepTotal) %>%
  summary()
```

```
##    StepTotal
##  Min.   :    0
##  1st Qu.: 3790
##  Median : 7406
##  Mean   : 7638
##  3rd Qu.:10727
##  Max.   :36019
```

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

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

```
weight_log %>%
  select(WeightPounds) %>%
  summary()
```

```
##   WeightPounds
##  Min.   :116.0
##  1st Qu.:135.4
##  Median :137.8
##  Mean   :158.8
##  3rd Qu.:187.5
##  Max.   :294.3
```

I took a look at comprehensive data frames.

```
skim_without_charts(daily_activity) %>%
  summary()
```

Table 1: Data summary

Name	daily_activity
Number of rows	940
Number of columns	15
_____	
Column type frequency:	
character	1
numeric	14
_____	
Group variables	None

```
skim_without_charts(daily_calories) %>%
  summary()
```

Table 2: Data summary

Name	daily_calories
Number of rows	940
Number of columns	3
_____	
Column type frequency:	
character	1
numeric	2
_____	
Group variables	None

```
skim_without_charts(daily_intensities) %>%
  summary()
```

Table 3: Data summary

Name	daily_intensities
Number of rows	940
Number of columns	10
_____	
Column type frequency:	
character	1
numeric	9
_____	
Group variables	None

```
skim_without_charts(daily_steps) %>%
  summary()
```

Table 4: Data summary

Name	daily_steps
Number of rows	940
Number of columns	3
Column type frequency:	
character	1
numeric	2
Group variables	None

```
skim_without_charts(sleep_day) %>%
  summary()
```

Table 5: Data summary

Name	sleep_day
Number of rows	413
Number of columns	5
Column type frequency:	
character	1
numeric	4
Group variables	None

```
skim_without_charts(weight_log) %>%
  summary()
```

Table 6: Data summary

Name	weight_log
Number of rows	67
Number of columns	8
Column type frequency:	
character	2
numeric	6
Group variables	None

I wanted to see how many unique participants there were to get some conclusions.

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

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

```
n_distinct(daily_calories$Id)
```

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

```
n_distinct(daily_intensities$Id)
```

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

```
n_distinct(daily_steps$Id)
```

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

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

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

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

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

Then I wanted to see the number of observations.

```
nrow(daily_activity)
```

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

```
nrow(daily_calories)
```

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

```
nrow(daily_intensities)
```

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

```
nrow(daily_steps)
```

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

```
nrow(sleep_day)
```

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

```
nrow(weight_log)
```

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

I separated the date and time columns in the sleep and weight logs to more easily merge the data sets.

```
sleep_day_sep <- separate(sleep_day,SleepDay,into=c('date','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, ...].
```

```
weight_log_sep <- separate(weight_log,Date,into=c('date','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, ...].
```

I renamed the "ActivityDay" column to "date" in the daily_activity table to make merging easier.

```
colnames(daily_activity)[2] <- "date"
```

I then merged the sleep_activity table with daily_activity and weight_activity with daily_activity.

```
sleep_activity <- merge(daily_activity, sleep_day_sep, by=c("Id", "date"))
weight_activity <- merge(daily_activity, weight_log_sep, by=c("Id", "date"))
```

I wanted to see if any of the participants were dropped as a result of the merge.

```
n_distinct(sleep_activity$Id)
```

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

```
n_distinct(weight_activity$Id)
```

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

It ends up being the same number of participants as the sleep and weight logs, which makes sense.

Analysis with Visualizations

First, I installed the visualization package and loaded the library.

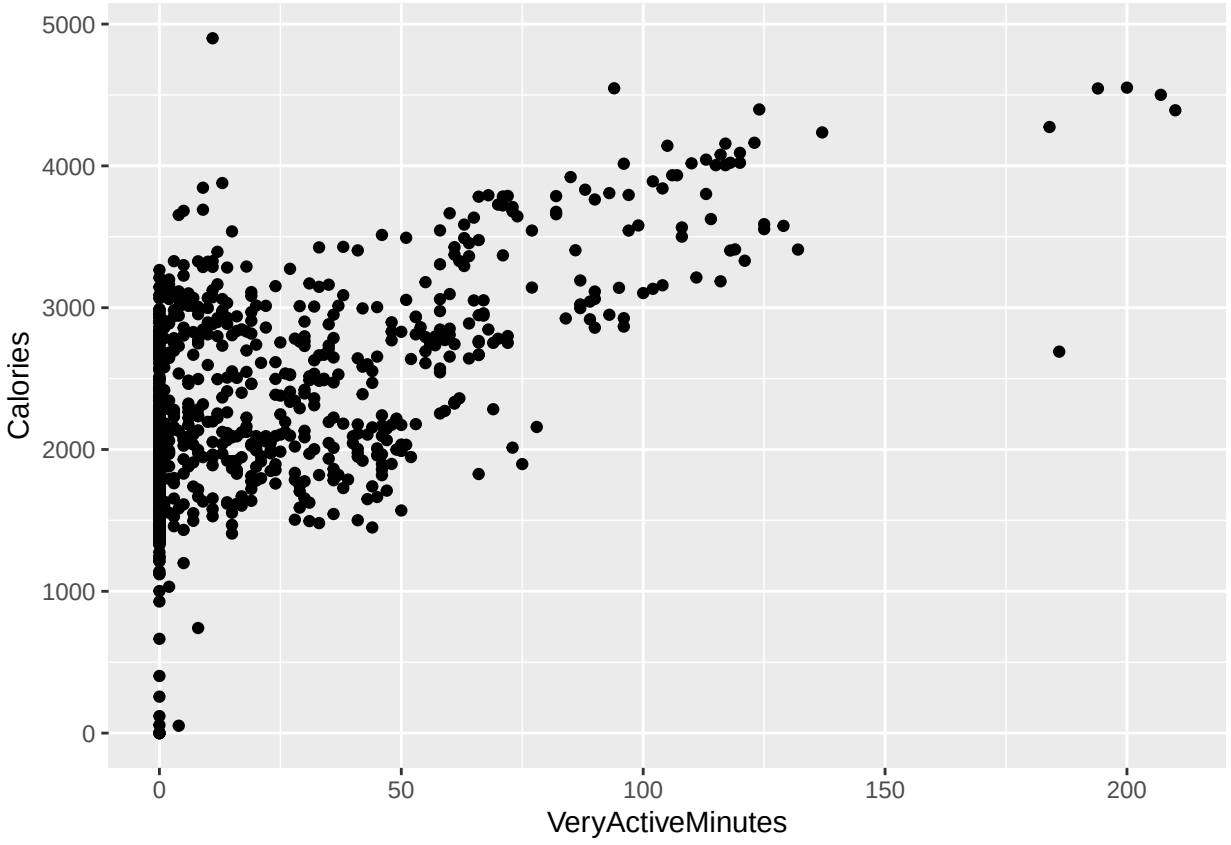
```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

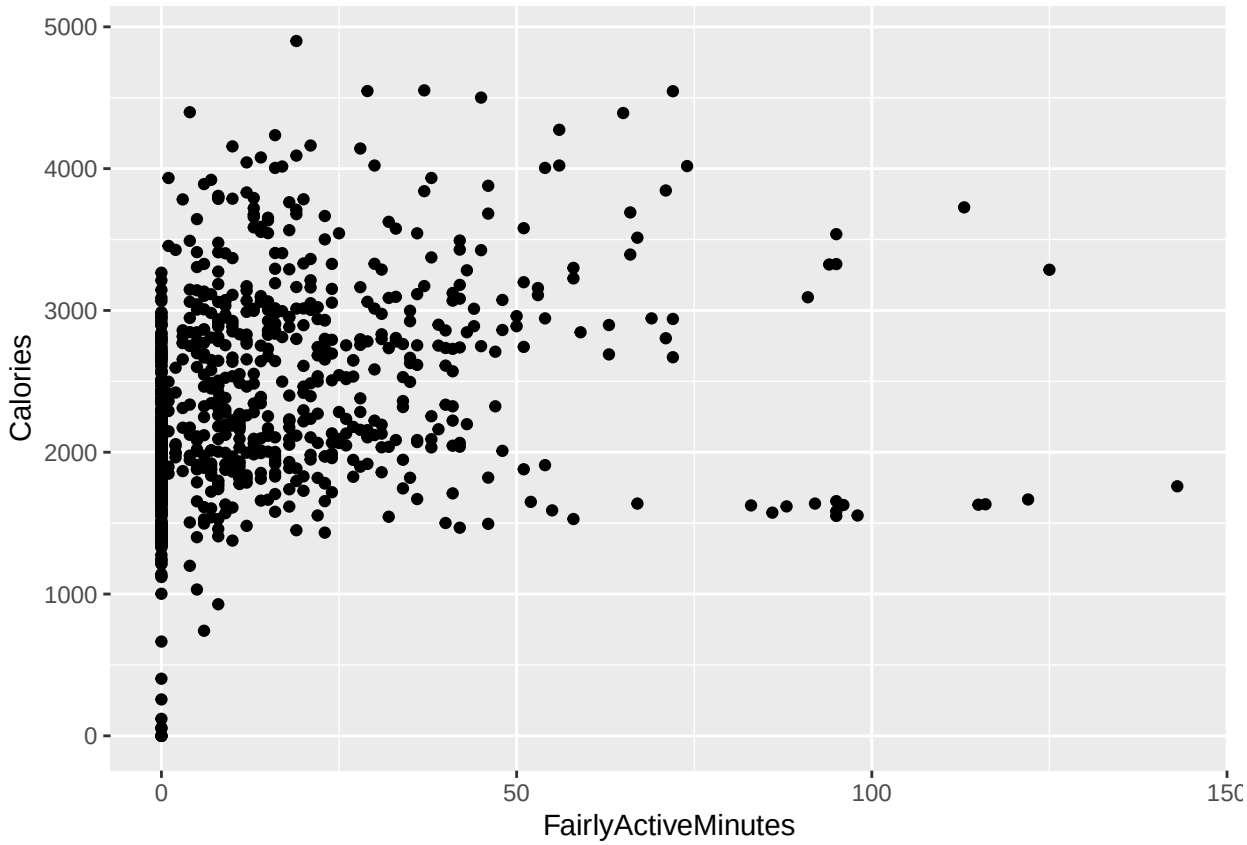
```
library("ggplot2")
```

I plotted the daily activities and calories burned to make sure that the data made sense. I knew that the more activity, especially "Very Active" activity, should result in more calories burned. I also did one plot comparing the "Very Active Minutes" with "Total Steps".

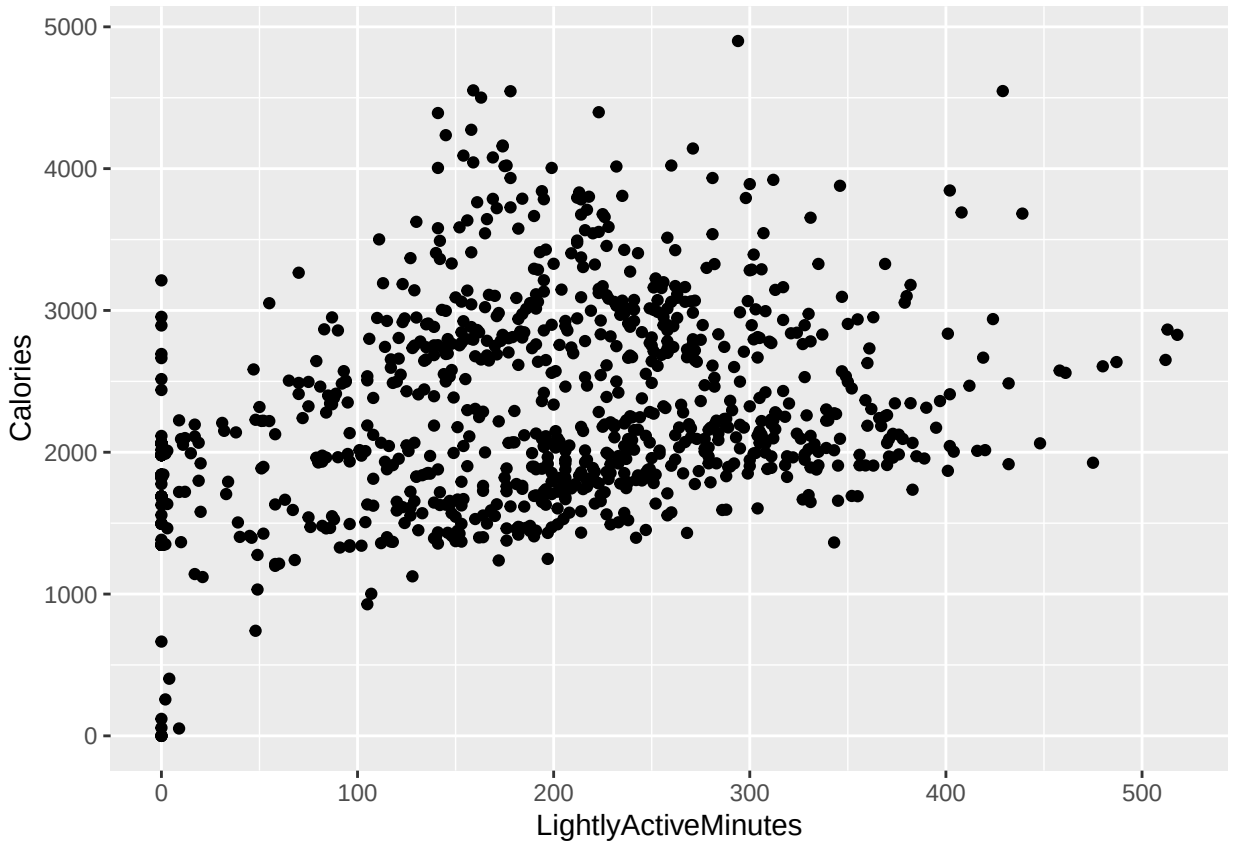
```
ggplot(data=daily_activity, aes(x=VeryActiveMinutes, y=Calories)) + geom_point()
```



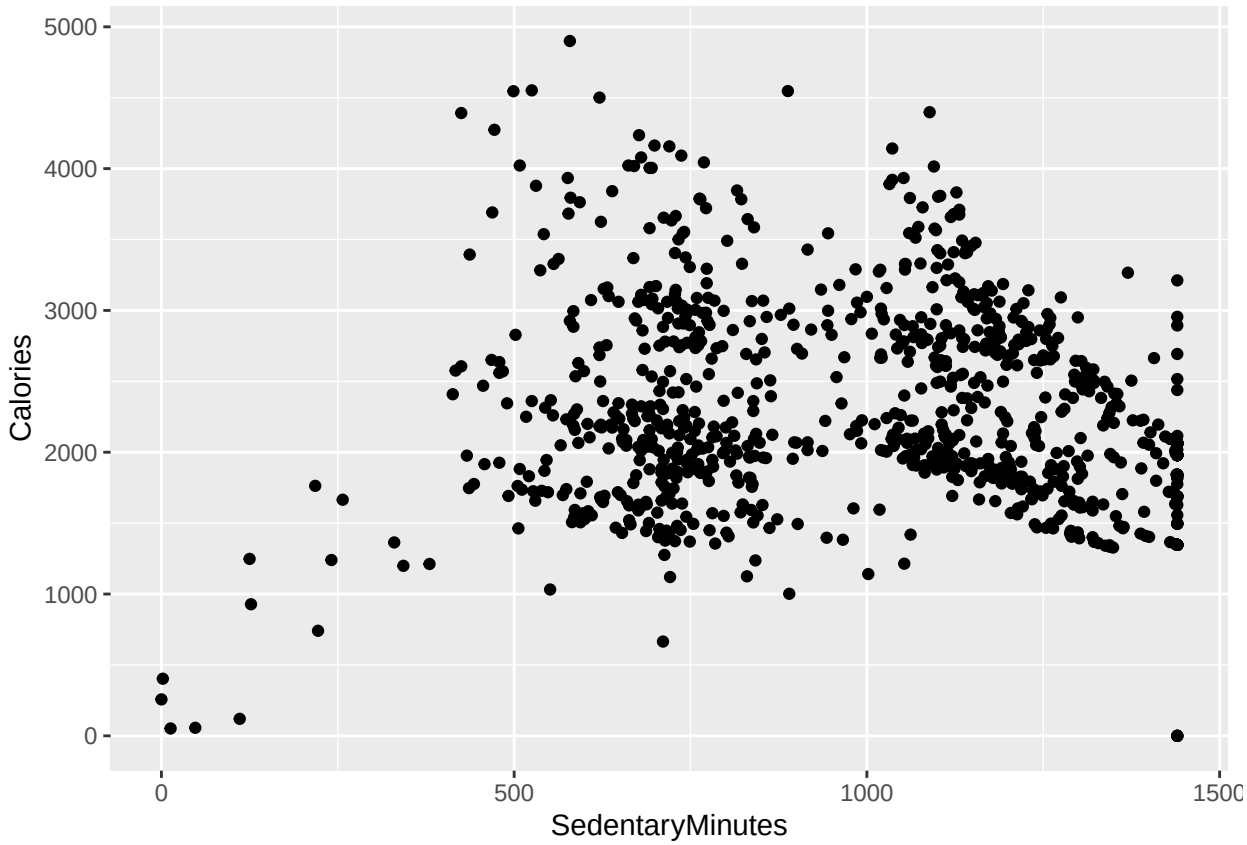

```
ggplot(data=daily_activity, aes(x=FairlyActiveMinutes, y=Calories)) + geom_point()
```




```
ggplot(data=daily_activity, aes(x=LightlyActiveMinutes, y=Calories)) + geom_point()
```

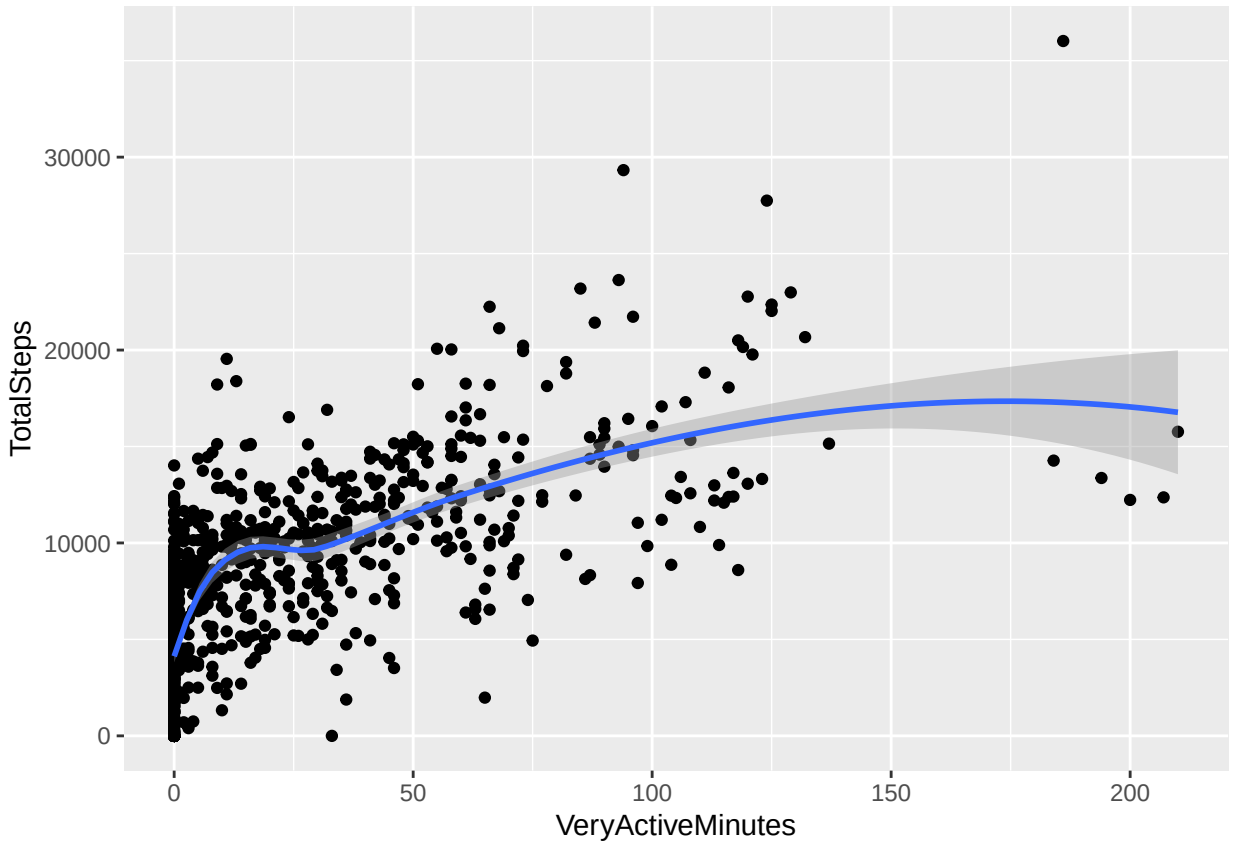



```
ggplot(data=daily_activity, aes(x=SedentaryMinutes, y=Calories)) + geom_point()
```




```
ggplot(data=daily_activity, aes(x=VeryActiveMinutes, y=TotalSteps)) + geom_point() + geom_smooth()
```

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



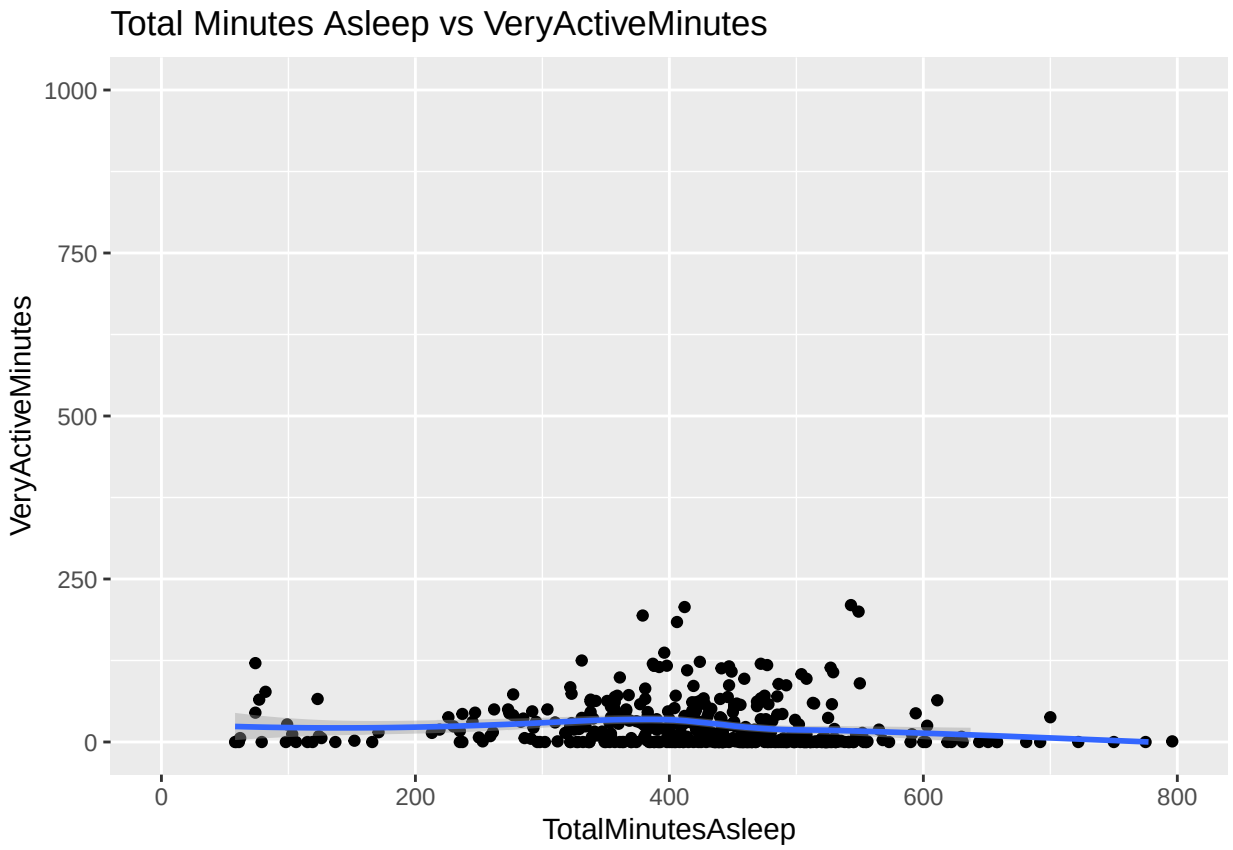
As you can see, the results are what one might expect.

Since the activity minutes and sleep minutes had the most participants, I decided to see if there was some relationship between the two to help Bellabeat get an idea on how they might want to market their products.

```
ggplot(data = sleep_activity, aes(x = TotalMinutesAsleep, y = VeryActiveMinutes)) +
  geom_point() + geom_smooth() + xlim(0,800) +
  ylim(0,1000) + labs(title = "Total Minutes Asleep vs VeryActiveMinutes")
```

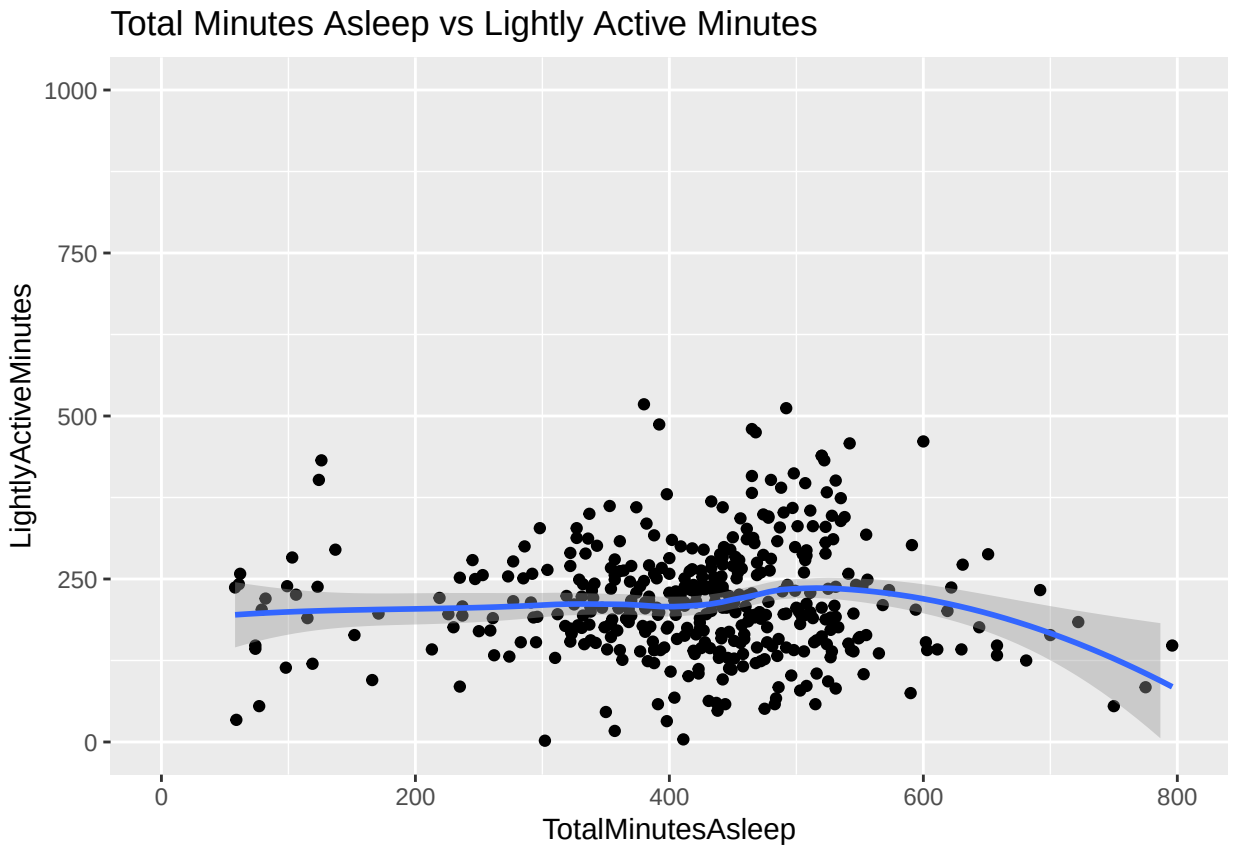
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 2 rows containing missing values (geom_smooth).
```



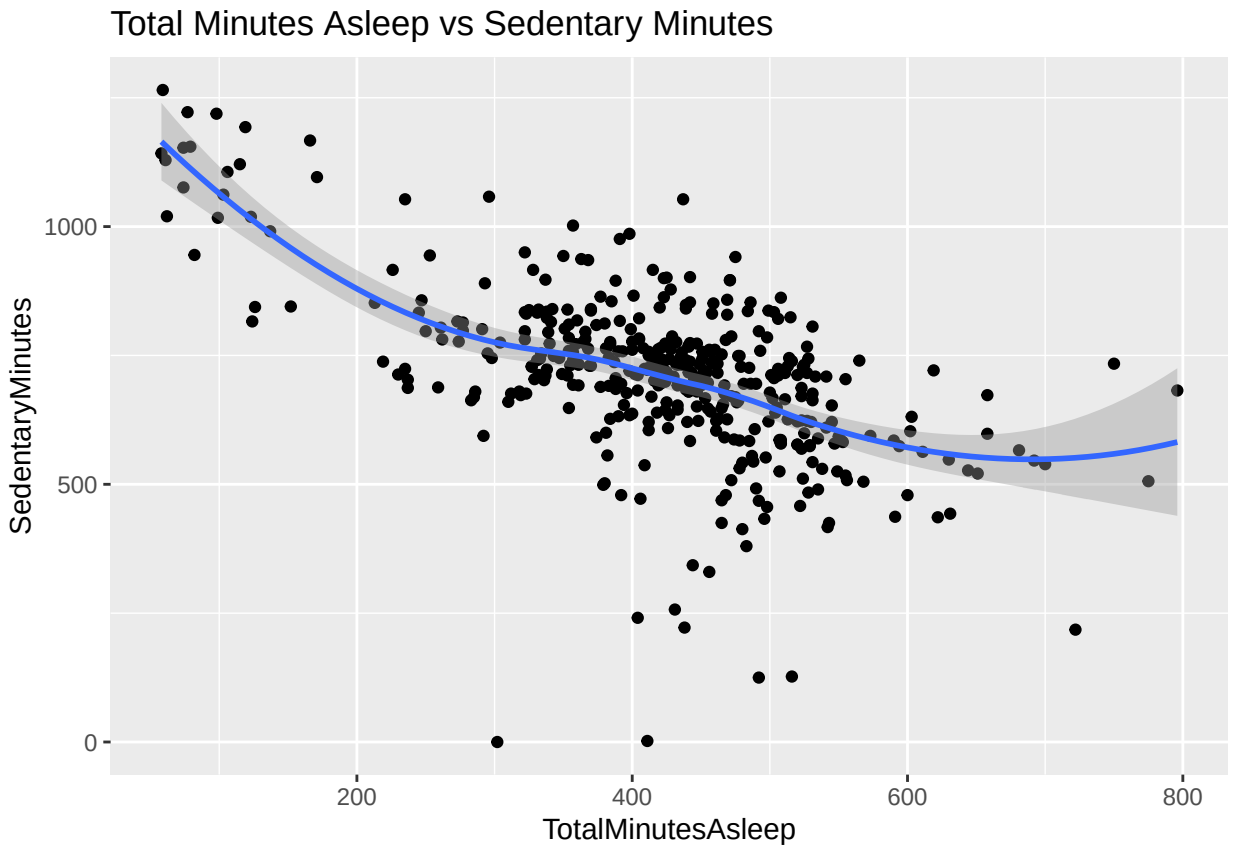

```
ggplot(data = sleep_activity) + (mapping=aes(x = TotalMinutesAsleep, y = LightlyActiveMinutes)) +
  geom_point() + geom_smooth() + xlim(0,800) +
  ylim(0,1000) + labs(title = "Total Minutes Asleep vs Lightly Active Minutes")
```

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




```
ggplot(data = sleep_activity, aes(x = TotalMinutesAsleep, y = SedentaryMinutes)) +
  geom_point() + geom_smooth() + labs(title = "Total Minutes Asleep vs Sedentary Minutes")
```

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



It seems in general that there is a negative correlation with sleep and activity. However, the increase in activity didn't lead to a positive correlation. Instead it appeared to level out where the more active someone was didn't necessarily mean the more time someone spent asleep.

Conclusion and Recommendations

The analysis indicates that the participants used their devices most commonly for tracking activity, steps, and sleep. I would encourage Bellabeat to focus their marketing on activity because that appeared to be the primary use case. In addition, it would help promote a positive image for women to focus on their health and fitness and general well-being as opposed to things like weight and diet. To extend the analysis, I would like to be able to examine data sets that focus on the demographics of the users (gender, age groups, etc.) and the usage of Bellabeat's products themselves, rather than data from a competitor's product.