

# BellaBeat Data Analysis Case Study using R

Salami Bayonle Lukman

20/02/2022

## Background

Bellabeat a high-tech company that manufactures health-focused smart products for women. Urška Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

## Task

Analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. Select one Bellabeat product to apply these insights to in your presentation.

## Ask

Questions for analysis

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat customers?
- How could these trends help influence Bellabeat marketing strategy?

## Business Task

Analyze smart device usage data of one Bellabeat product in order to gain insight into how consumers use non-Bellabeat smart devices and to unlock new opportunities for growth.

## Prepare Data

The data in this analysis is a publicly available data from the Fitbit Fitness Tracker Data set found at <https://www.kaggle.com/arashnic/fitbit>

## Step 1: Import your data

```
activity <- read.csv("dailyActivity_merged.csv")
calories <- read.csv("dailyCalories_merged.csv")
sleep <- read.csv("sleepDay_merged.csv")
steps <- read.csv("dailySteps_merged.csv")
intensity <- read.csv("dailyIntensities_merged.csv")
```

**Step 2:** To get at a sample of the data and also preview all the column names: replace the syntax “activity” with the other datasets.

```
head(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(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(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(intensity)
```

```
##           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(sleep)
```

```
## 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
```

### Step 3: Import R library

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.5 v purrr 0.3.4
## v tibble 3.1.1 v dplyr 1.0.6
## v tidyr 1.1.3 v stringr 1.4.0
## v readr 1.4.0 v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

```
library(ggplot2)
library(dplyr)
library(janitor)
```

```
##
```

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

```
## The following objects are masked from 'package:stats':
```

```
##
##      chisq.test, fisher.test
```

```
library(ggrepel)
library(ggpubr)
library(tidyr)
library(lubridate)
```

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

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

#### Step 4: Process and Transform data

```
activity_change <- activity %>%
  mutate(date_new = as_date(mdy(ActivityDate)))
head(activity_change)
```

```
##           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  date_new
## 1                13                328                728    1985 2016-04-12
## 2                19                217                776    1797 2016-04-13
## 3                11                181               1218    1776 2016-04-14
## 4                34                209                726    1745 2016-04-15
## 5                10                221                773    1863 2016-04-16
## 6                20                164                539    1728 2016-04-17
```

```
steps_change <- steps %>%
  mutate(date_new = as_date(mdy(ActivityDay)))
head(steps_change)
```

```
##           Id ActivityDay StepTotal  date_new
## 1 1503960366 4/12/2016      13162 2016-04-12
```

```
## 2 1503960366 4/13/2016 10735 2016-04-13
## 3 1503960366 4/14/2016 10460 2016-04-14
## 4 1503960366 4/15/2016 9762 2016-04-15
## 5 1503960366 4/16/2016 12669 2016-04-16
## 6 1503960366 4/17/2016 9705 2016-04-17
```

```
sleep_change <- sleep %>%
  separate(SleepDay, into=c("Date", "Time"), sep = " ") %>%
  select(Id, Date, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  group_by(Id, Date) %>%
  mutate(date_new = as_date(mdy(Date)))
```

```
## 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(sleep_change)
```

```
## # A tibble: 6 x 6
## # Groups:   Id, Date [6]
##       Id Date      TotalSleepReco~ TotalMinutesAsle~ TotalTimeInBed date_new
##       <dbl> <chr>          <int>          <int>          <int> <date>
## 1 1.50e9 4/12/2~           1           327           346 2016-04-12
## 2 1.50e9 4/13/2~           2           384           407 2016-04-13
## 3 1.50e9 4/15/2~           1           412           442 2016-04-15
## 4 1.50e9 4/16/2~           2           340           367 2016-04-16
## 5 1.50e9 4/17/2~           1           700           712 2016-04-17
## 6 1.50e9 4/19/2~           1           304           320 2016-04-19
```

## Step 5: Analyze and Share you data

- To get the relationship between total number of steps walked and calories burned we merge 'activity\_change' data with 'steps\_change' data

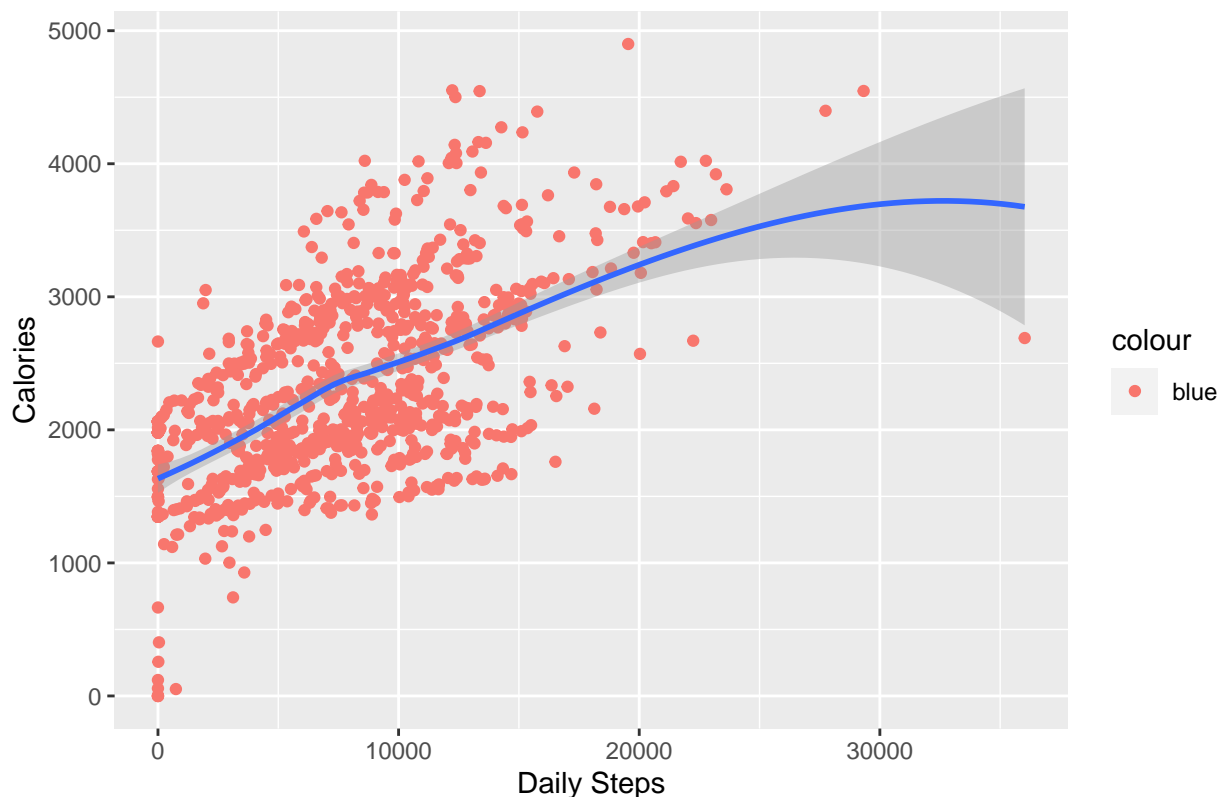
```
merge_activity_steps <- merge(activity_change, steps_change, all = TRUE) %>% group_by (Id, date_new)
head(merge_activity_steps)
```

```
## # A tibble: 6 x 18
## # Groups:   Id, date_new [6]
##       Id date_new ActivityDate TotalSteps TotalDistance TrackerDistance
##       <dbl> <date>      <chr>          <int>          <dbl>          <dbl>
## 1 1503960366 2016-04-12 4/12/2016      13162          8.5           8.5
## 2 1503960366 2016-04-13 4/13/2016      10735          6.97          6.97
## 3 1503960366 2016-04-14 4/14/2016      10460          6.74          6.74
## 4 1503960366 2016-04-15 4/15/2016       9762          6.28          6.28
## 5 1503960366 2016-04-16 4/16/2016      12669          8.16          8.16
## 6 1503960366 2016-04-17 4/17/2016       9705          6.48          6.48
## # ... with 12 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <int>, FairlyActiveMinutes <int>,
## #   LightlyActiveMinutes <int>, SedentaryMinutes <int>, Calories <int>,
## #   ActivityDay <chr>, StepTotal <int>
```

```
ggplot(data = merge_activity_steps) + geom_point(mapping = aes(x=TotalSteps,y=Calories, color="blue"))
```

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

## Total Daily Steps vs. Total Calories Burned



- To get the relationship between total number of Daily steps and total number of daily sleep we merge 'activity\_change' data with 'sleep\_change' data

```
merge_activity_sleep <- merge(sleep_change, activity_change, all = TRUE) %>%
  group_by (Id, date_new)
head(merge_activity_sleep)
```

```
## # A tibble: 6 x 20
## # Groups:   Id, date_new [6]
##       Id date_new Date TotalSleepReco~ TotalMinutesAsle~ TotalTimeInBed
##       <dbl> <date>   <chr>         <int>             <int>             <int>
## 1  1.50e9 2016-04-12 4/12/2~           1             327             346
## 2  1.50e9 2016-04-13 4/13/2~           2             384             407
## 3  1.50e9 2016-04-14 <NA>           NA              NA              NA
## 4  1.50e9 2016-04-15 4/15/2~           1             412             442
## 5  1.50e9 2016-04-16 4/16/2~           2             340             367
## 6  1.50e9 2016-04-17 4/17/2~           1             700             712
## # ... with 14 more variables: ActivityDate <chr>, TotalSteps <int>,
## # TotalDistance <dbl>, TrackerDistance <dbl>, LoggedActivitiesDistance <dbl>,
## # VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## # LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## # VeryActiveMinutes <int>, FairlyActiveMinutes <int>,
## # LightlyActiveMinutes <int>, SedentaryMinutes <int>, Calories <int>
```

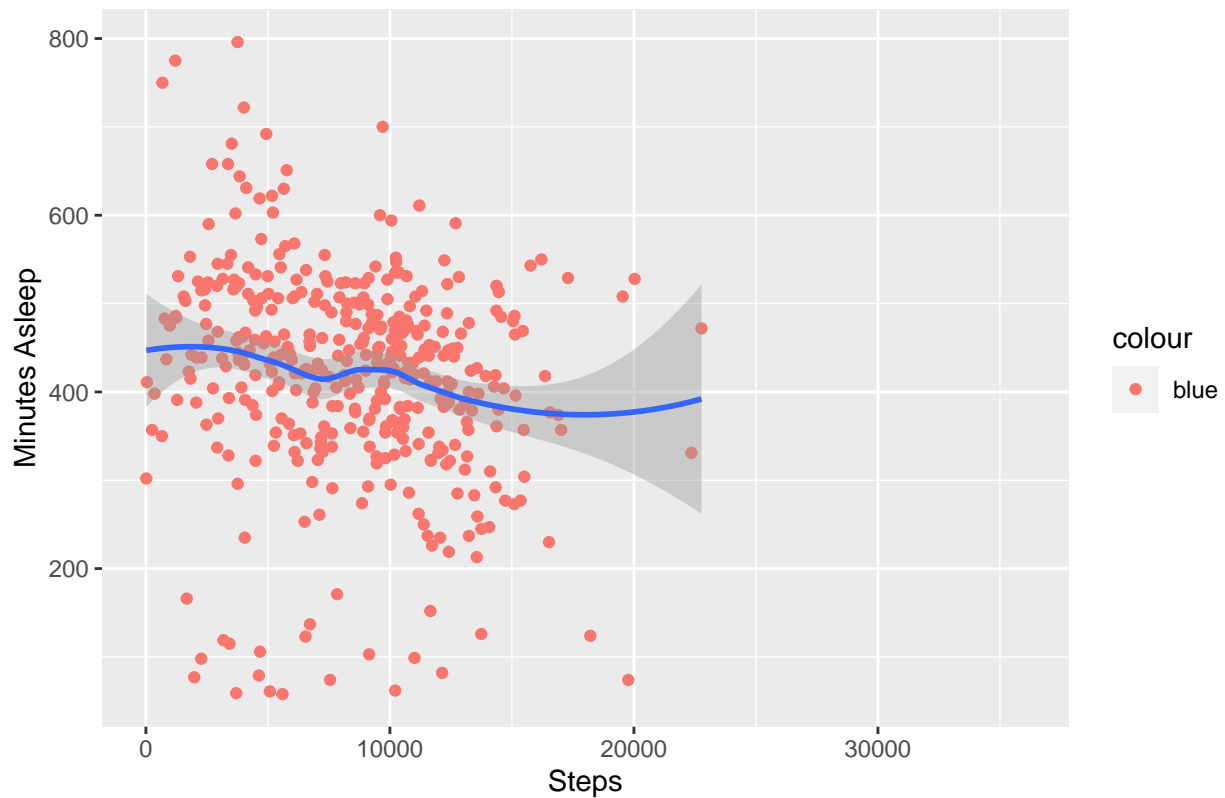
```
ggplot(data = merge_activity_sleep) + geom_point(mapping = aes(x=TotalSteps, y=TotalMinutesAsleep, color=
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 530 rows containing missing values (geom_point).
```

### Total number of steps vs. Total daily sleep time



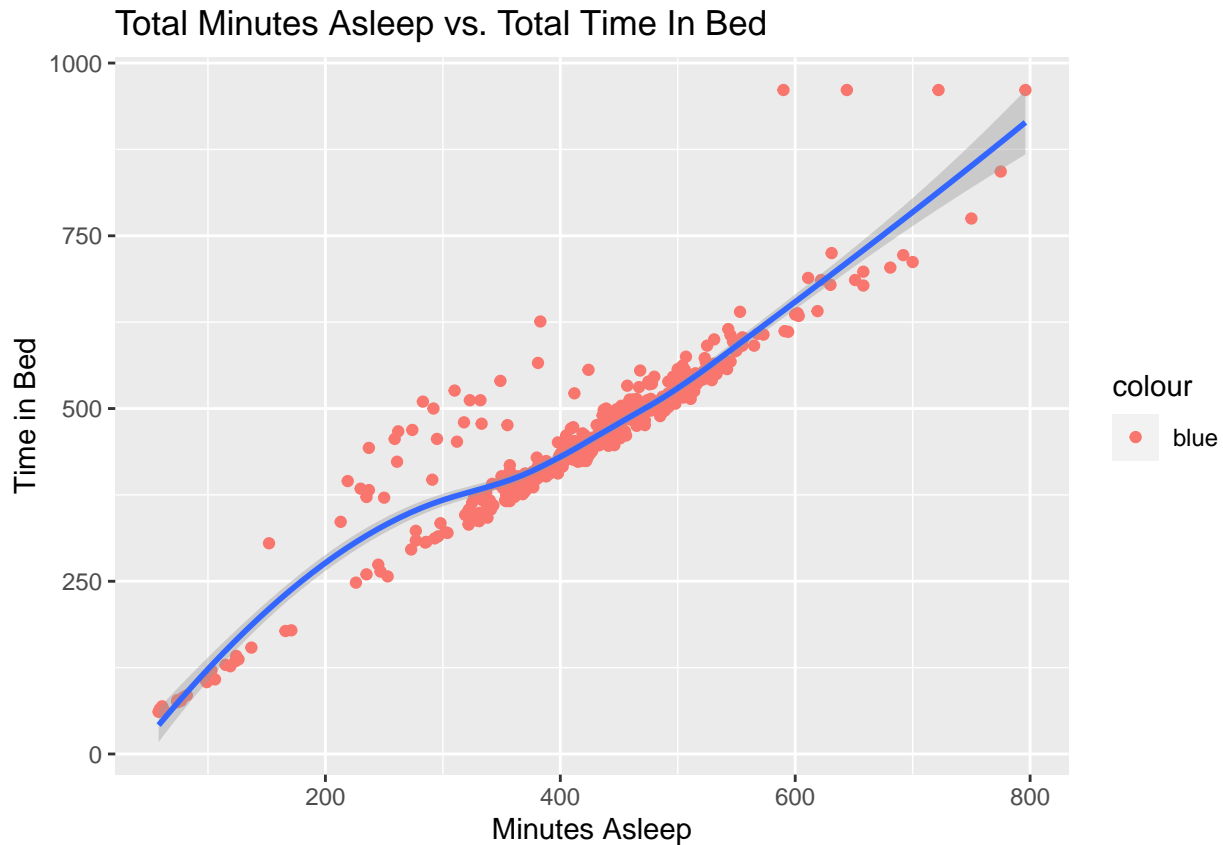
Relationship between Total Minutes Asleep and Total Time in Bed

```
ggplot(data = merge_activity_sleep) + geom_point(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed, colour = "blue"))
```

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

```
## Warning: Removed 530 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 530 rows containing missing values (geom_point).
```



- Average daily Steps

```
mean(activity_change$TotalSteps)
```

```
## [1] 7637.911
```

- Average Minutes Asleep

```
mean(sleep_change$TotalMinutesAsleep)
```

```
## [1] 419.4673
```

## CONCLUSIONS

This data gave us great insights on how these smart devices are being used and how helpful they can be. From my analysis it can be deduced that:

- There exist a Positive Correlation between number of steps and calories burned.
- No correlation between number of steps and sleep time.
- There exist a strong positive correlation between minutes asleep and time spent in bed.
- The average daily steps is approximately 7638 steps.
- Then average minutes asleep is about 7 hours.

## RECOMMENDATIONS

- Bellabeat Spring can alert the user after every 2000 steps to drink water through Bellabeat App. This also acts as a notification to be more active.
- Bellabeat App can notify users about sleep time. It should be able to be preprogrammed for a specific sleep time.



Thanks to Google and Couseira for contribution to Data analytics learning community.