Bellabeat Case Study

Sumia Tahir

2022-03-29

Background

Bellabeat is a high-tech manufacturer of health-focused products for women. Bellabeat has several products to offer. The Bellabeat app provides users about data related to their health (sleep, activity, menstrual cycle, etc.) and connects to Bellabeat's smart products. Bellabeat's Leaf tracker collects data on sleep, activity and stress, and can be worn as a bracelet, necklace or clip. The Bellabeat's smart watch - Time - also collects user activity, sleep and stress. Bellabeat also offers a smart water bottle, called Spring, that tracks a user's intake of water. Bellabeat would like to get some insights that would help them increase their customer base. This case study will look at some data from the FitBit Fitness Tracker to help understand customer trends in usage of other smart fitness devices and extend these findings to guide Bellabeat's marketing strategy. Bellabeat's Leaf tracker most closely resembles the FitBit; therefore, the results of the analysis will be used to make recommendations for the marketing of the Leaf tracker.

Business Task: Identify trends in the daily usage of the FitBit Fitness Tracker and use these insights to recommend marketing strategies for Bellabeat's Leaf tracker.

Description of the Data

The dataset contains observations from 33 Fitbit users on their physical activity, heart rate and sleep. The dataset was available on Kaggle through Mobius and open for public use. It was collected by a survey via Amazon Mechanical Turk. The survey responses spanned two months (03.02.2016-05.12.2016) but the available data on Kaggle only shows observations for one month. The data consists of 18 Excel files that contain both long and wide data. This file contains metadata of this dataset: https://www.fitabase.com/media/1930/fitabasedatadictionary102320.pdf

Limitations of the Dataset

For our case study, this dataset is not reliable. A major limitation of this dataset is that the gender of the users is unknown. Bellabeat's products are for women; therefore, not all of the data findings may align with usage trends in women. Also, there may be some bias in that perhaps users who used the FitBit more or were more active tended to respond to the survey. The data was made available through a second party (Mobius) so it is not original. The data is from 6 years ago so it is not current. Another limitation of the dataset is the relatively small sample size. There is data from only 33 users. The observations only span one month so that is another limitation because the season may affect activity, sleep and thereby, usage trends. Therefore, this dataset is not comprehensive. It is also not cited. Therefore, this is not a good datasource.

Data Cleaning

Loading the libraries

```
## Installing packages into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
## also installing the dependency 'vctrs'

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

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     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()

##
## Attaching package: 'lubridate'

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

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)

## here() starts at /cloud/project

##
## Attaching package: 'janitor'

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

Importing the data

We will import 2 files, the 'sleep' and 'dailyActivity' logs.

```r
dailysleep <- read.csv("sleepDay_merged.csv")
dailyactivity <- read.csv("dailyActivity_merged.csv")
head(dailysleep)
```

```
##           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(dailyactivity)
```

```
##           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
```

Cleaning & Formatting the Data

Let's check the columns of the datasets. We can add a 'Date' column to both datasets so that the format is consistent. We can also add a column for 'TotalActiveMinutes' that combines the time of the three intensities of exercise (Very Active, Fairly Active, Lightly Active). We can also add a column called 'TotalDeviceTime' (in hours) that sums the sedentary and active times to check device usage. We can also check for missing values using the 'skim_without_charts' function.

```
dailysleep$Date <- as.Date(dailysleep$SleepDay,format="%m/%d/%Y")
dailyactivity$Date <- as.Date(dailyactivity$ActivityDate,format="%m/%d/%Y")
dailyactivity <- mutate(dailyactivity, TotalActiveMinutes = VeryActiveMinutes + FairlyActiveMinutes + Li
dailyactivity <- mutate(dailyactivity, TotalDeviceTime = ((TotalActiveMinutes + SedentaryMinutes)/60))
n_distinct(dailyactivity$Id)
```

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

```
glimpse(dailyactivity)
```

```
## Rows: 940
## Columns: 18
## $ 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~
## $ Date                     <date> 2016-04-12, 2016-04-13, 2016-04-14, 2016-04-~
## $ TotalActiveMinutes       <int> 366, 257, 222, 272, 267, 222, 291, 345, 245, ~
## $ TotalDeviceTime          <dbl> 18.23333, 17.21667, 24.00000, 16.63333, 17.33~
```

```
skim_without_charts(dailyactivity)
```

Table 1: Data summary

Name	dailyactivity
Number of rows	940
Number of columns	18
Column type frequency:	
character	1
Date	1
numeric	16
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDate	0	1	8	9	0	31	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
Date	0	1	2016-04-12	2016-05-12	2016-04-26	31

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.855407e+09	2.424805e+09	1.50396e+09	2.320127e+09	4.445115e+09	6.962181e+09	8.877689e+09
TotalSteps	0	1	7.637910e+03	5.087150e+03	0.00000e+00	3.789750e+03	7.405500e+03	1.072700e+04	3.601900e+04
TotalDistance	0	1	5.490000e+00	3.920000e+00	0.00000e+00	2.620000e+00	5.240000e+00	7.710000e+00	2.803000e+01
TrackerDistance	0	1	5.480000e+00	3.910000e+00	0.00000e+00	2.620000e+00	5.240000e+00	7.710000e+00	2.803000e+01
LoggedActivitiesDistance	0	1	1.100000e-01	6.200000e-01	0.00000e+00	0.000000e+00	0.000000e+00	0.000000e+00	4.940000e+00

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
VeryActiveDistance	0	1	1.500000e+00	2.060000e+00	0.000000e+00	0.000000e+00	2.100000e-01	2.050000e+00	2.192000e+01
ModeratelyActiveDistance	0	1	5.700000e-01	8.800000e-01	0.000000e+00	0.000000e+00	2.400000e-01	8.000000e-01	6.480000e+00
LightActiveDistance	0	1	3.340000e+00	2.040000e+00	0.000000e+00	1.950000e+00	3.360000e+00	4.780000e+00	1.071000e+01
SedentaryActiveDistance	0	1	0.000000e+00	1.000000e-02	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.100000e-01
VeryActiveMinutes	0	1	2.116000e+01	3.284000e+01	0.000000e+00	0.000000e+00	4.000000e+00	3.200000e+01	2.100000e+02
FairlyActiveMinutes	0	1	1.356000e+01	1.999000e+01	0.000000e+00	0.000000e+00	6.000000e+00	1.900000e+01	1.430000e+02
LightlyActiveMinutes	0	1	1.928100e+02	1.091700e+02	0.000000e+00	1.270000e+02	1.990000e+02	2.640000e+02	5.180000e+02
SedentaryMinutes	0	1	9.912100e+02	3.012700e+02	0.000000e+00	7.297500e+02	1.057500e+03	1.229500e+03	1.440000e+03
Calories	0	1	2.303610e+03	7.181700e+02	0.000000e+00	1.828500e+03	2.134000e+03	2.793250e+03	4.900000e+03
TotalActiveMinutes	0	1	2.275400e+02	1.217800e+02	0.000000e+00	1.467500e+02	2.470000e+02	3.172500e+02	5.520000e+02
TotalDeviceTime	0	1	2.031000e+01	4.430000e+00	3.000000e-02	1.650000e+01	2.400000e+01	2.400000e+01	2.400000e+01

```
n_distinct(dailysleep$Id)
```

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

```
glimpse(dailysleep)
```

```
## Rows: 413
## Columns: 6
## $ 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~
## $ Date               <date> 2016-04-12, 2016-04-13, 2016-04-15, 2016-04-16, 20~
```

```
skim_without_charts(dailysleep)
```

Table 5: Data summary

Name	dailysleep
Number of rows	413
Number of columns	6
Column type frequency:	
character	1
Date	1
numeric	4
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
SleepDay	0	1	20	21	0	31	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
Date	0	1	2016-04-12	2016-05-12	2016-04-27	31

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	5.000979e+09	2.060360e+09	1503960366	3977333714	4702921684	6962181067	8792009665
TotalSleepRecords	0	1	1.120000e+00	3.450000e-01	1	1	1	1	3
TotalMinutesAsleep	0	1	4.194700e+02	1.183440e+02	58	361	433	490	796
TotalTimeInBed	0	1	4.586400e+02	1.271000e+02	61	403	463	526	961

The dailyactivity dataset contains 940 observations and 18 columns from 33 distinct users. The dailysleep dataset contains 413 observations and 6 columns from 24 distinct users. We can see that there are no missing values. The ‘Date’ column is now consistent over both datasets (this will help when we merge the datasets). The column names are clear and consistent so there is no need to change them.

Let’s check some basic summary statistics:

```
dailyactivity %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes,
         TotalActiveMinutes,
         TotalDeviceTime) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes TotalActiveMinutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :  0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:146.8
##  Median : 7406   Median : 5.245   Median :1057.5   Median :247.0
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   :227.5
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.:317.2
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :552.0
##  TotalDeviceTime
##  Min.   : 0.03333
##  1st Qu.:16.49583
##  Median :24.00000
##  Mean   :20.31255
##  3rd Qu.:24.00000
##  Max.   :24.00000
```

```
dailysleep %>%
  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
```

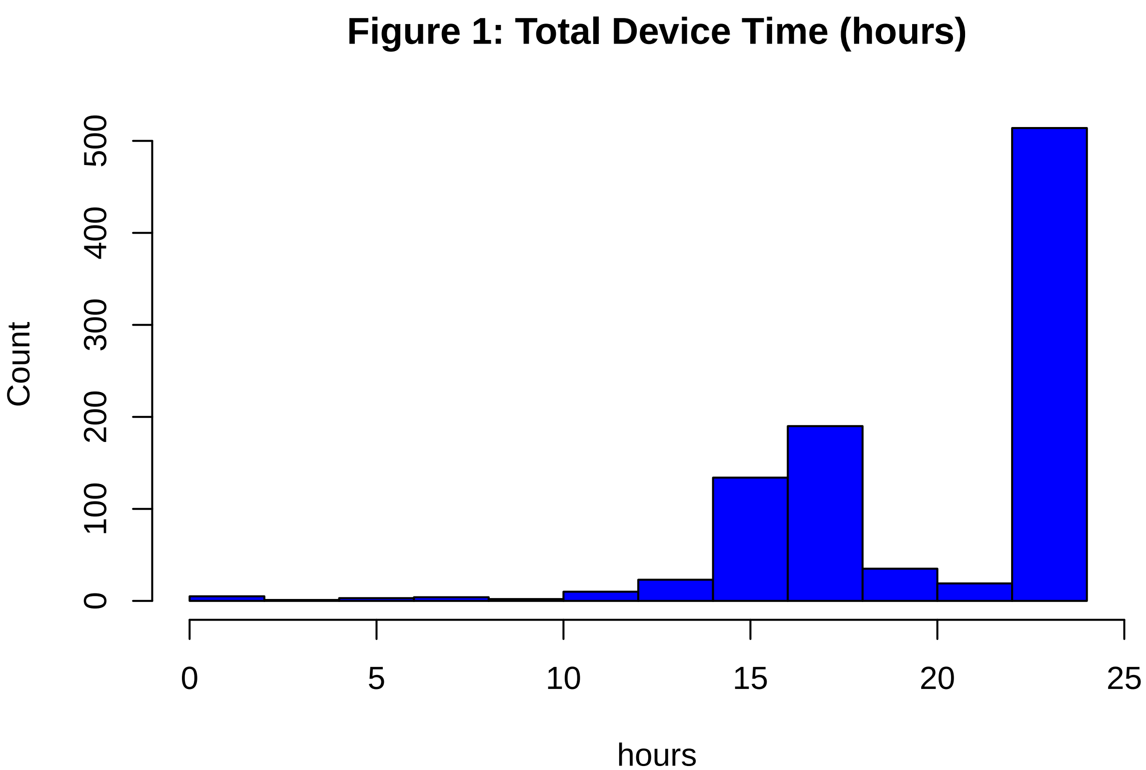
On average, the FitBit data is collected over 20 hours a day. The statistics for SedentaryMinutes do not match up with TotalTimeInBed. It is not clear if these are measuring two different things. Another reason could be that the sedentaryminutes include observations from 33 users whereas totaltimeinbed only has data from 24 users. On average, users slept 420 minutes (7 hours) a day. The average total distance is 5.5 kilometers.

Analyzing the data

Daily Activity dataset

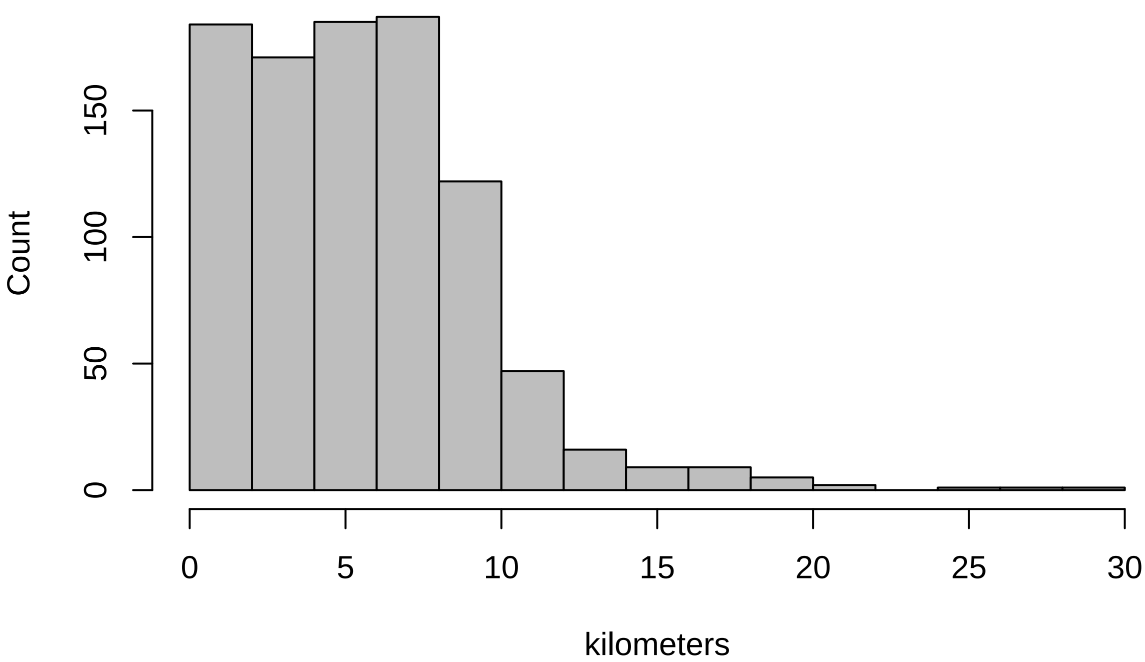
We can look at some trends in the 'dailyactivity' dataset by generating some plots. First, we will make some histograms.

```
hist(dailyactivity$TotalDeviceTime, main='Figure 1: Total Device Time (hours)', xlab='hours', ylab='Cou
```




```
hist(dailyactivity$TotalDistance, main='Figure 2: Total Distance (km)', xlab='kilometers', ylab='Count'
```

Figure 2: Total Distance (km)




```
percent_active <- dailyactivity$TotalActiveMinutes/(dailyactivity$TotalActiveMinutes + dailyactivity$Sec
hist(percent_active, main='Figure 3: Percent Active', xlab='%', ylab='Count', col='green')
```

Figure 3: Percent Active

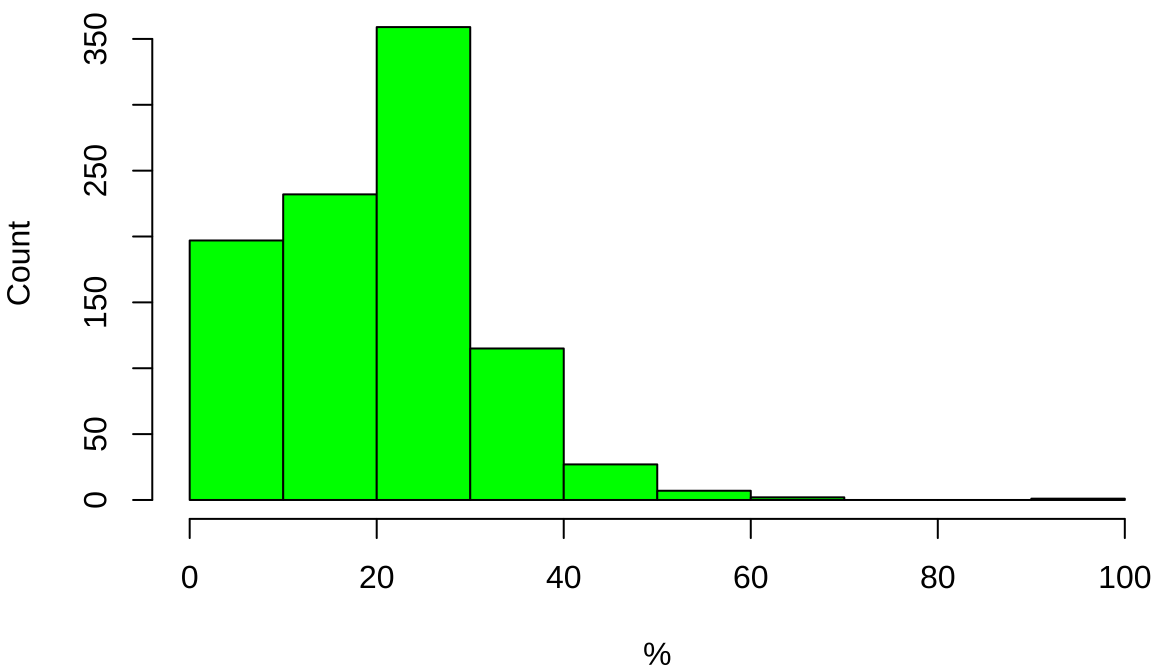


Figure 1 shows that on most days, we have data for 24 hours. However, a significant number of days have data recorded for only 15-17 hours. This indicates usage of the device. It would be interesting to know if device usage varies with day of the week, and if there are any differences between usage between men and women.

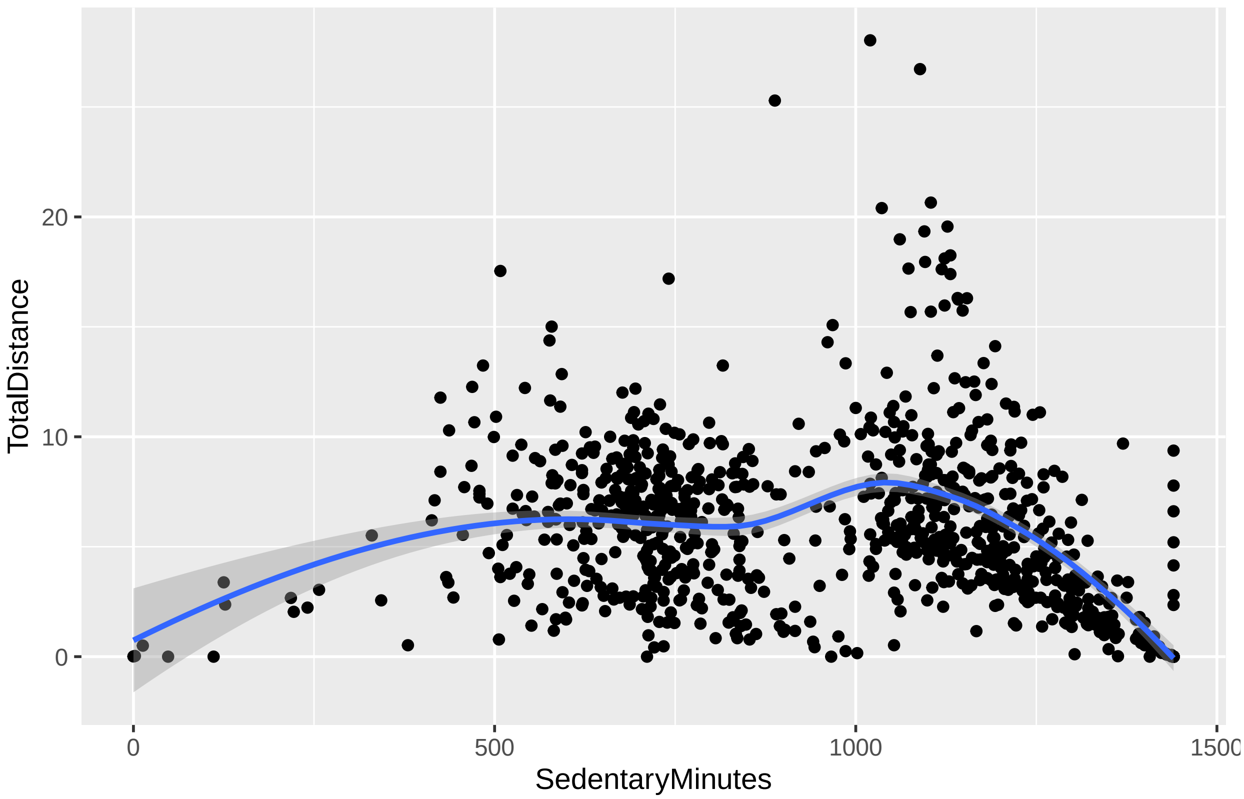
Figure 2 shows that most users recorded a total distance of under 10 km per day. Figure 3 shows that most users were 'active' for 20-30% of the time the device was used.

We can also generate some plots to see the relationship between the variables.

```
ggplot(dailyactivity, aes(x=SedentaryMinutes, y=TotalDistance)) + geom_point() + geom_smooth() + labs(t
```

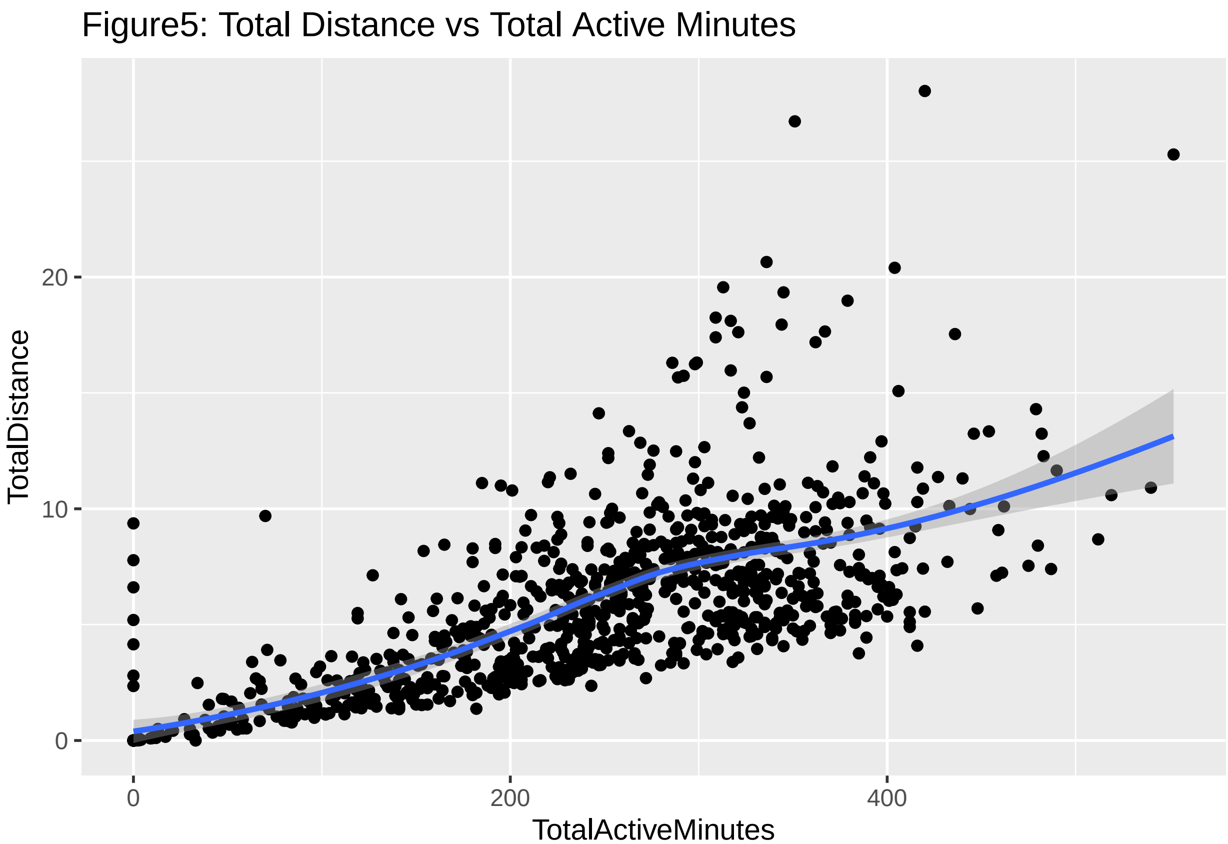
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figure 4: Total Distance vs Sedentary Minutes



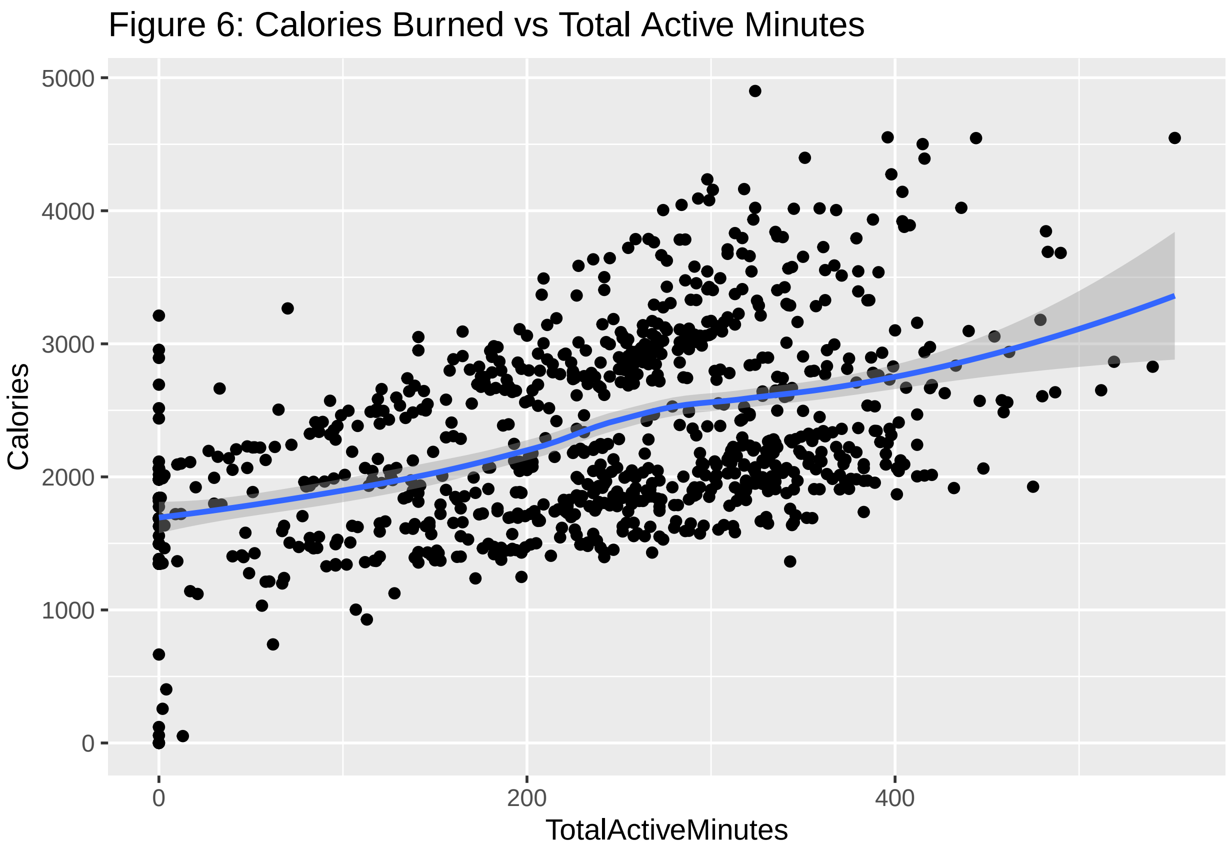

```
ggplot(dailyactivity, aes(x=TotalActiveMinutes, y=TotalDistance)) + geom_point() + geom_smooth() + labs
```

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




```
ggplot(dailyactivity, aes(x=TotalActiveMinutes, y=Calories)) + geom_point() + geom_smooth() + labs(titl
```

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



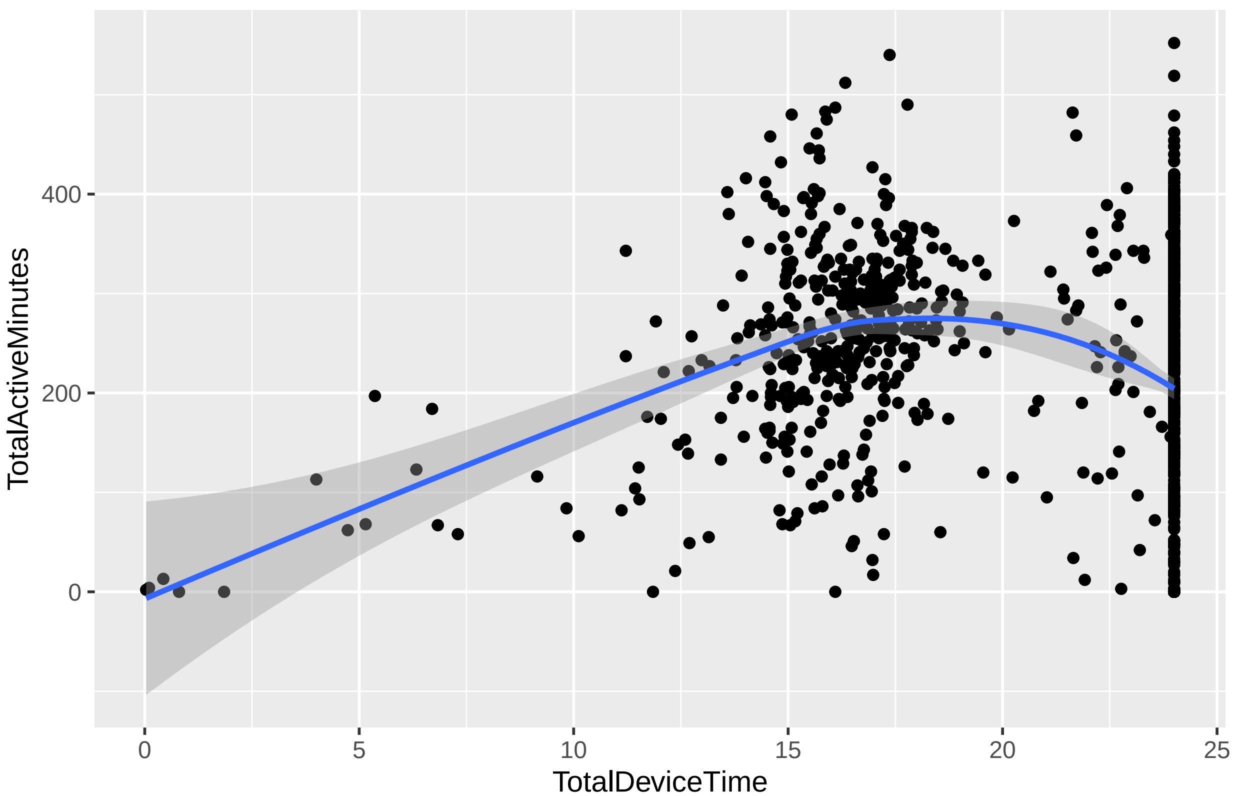

```
ggplot(dailyactivity, aes(x=VeryActiveMinutes, y=Calories)) + geom_point() + geom_smooth() + labs(title=
```

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


```
ggplot(dailyactivity, aes(x=TotalDeviceTime, y=TotalActiveMinutes)) + geom_point() + geom_smooth() + lab
```

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

Figure 8: Total Active Minutes vs Total Device Time




```
ggplot(dailyactivity, aes(x=SedentaryMinutes, y=TotalDeviceTime)) + geom_point() + geom_smooth() + labs
```

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

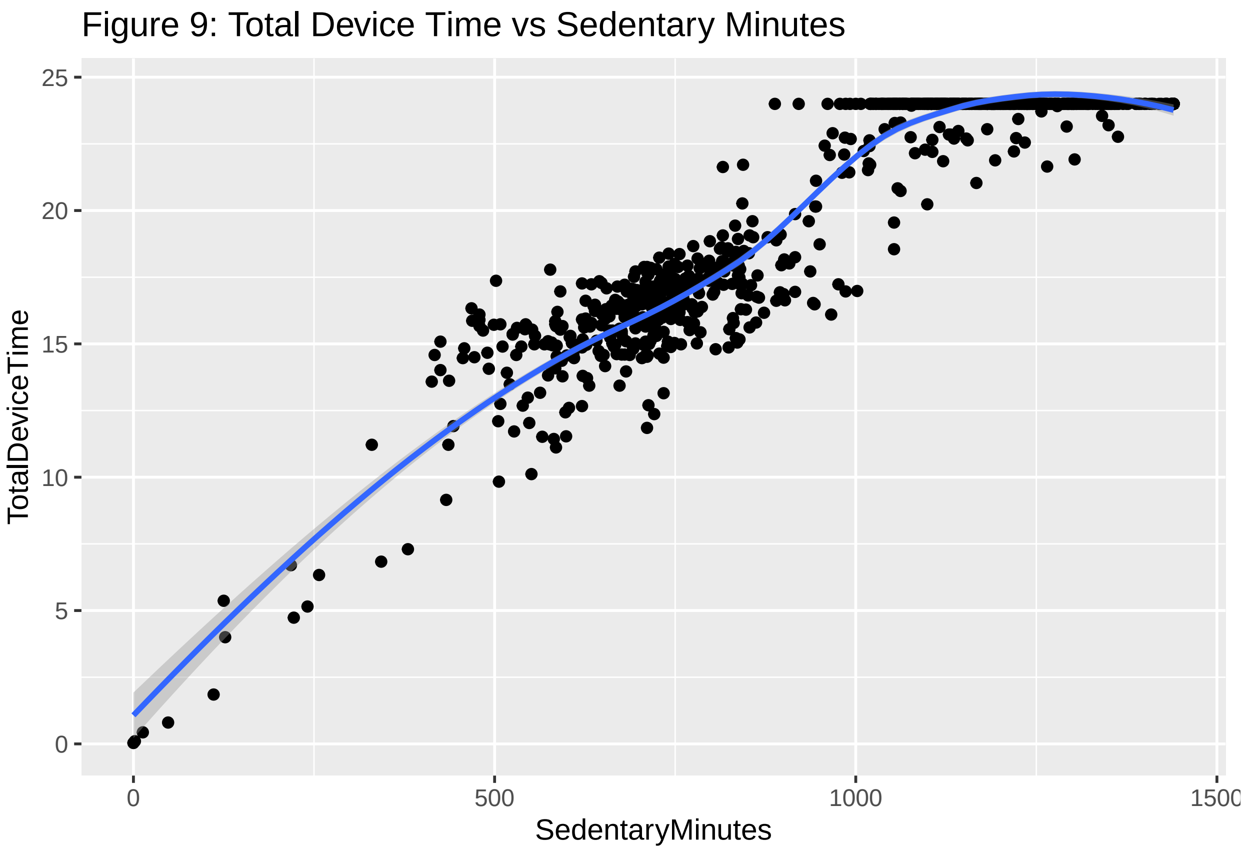


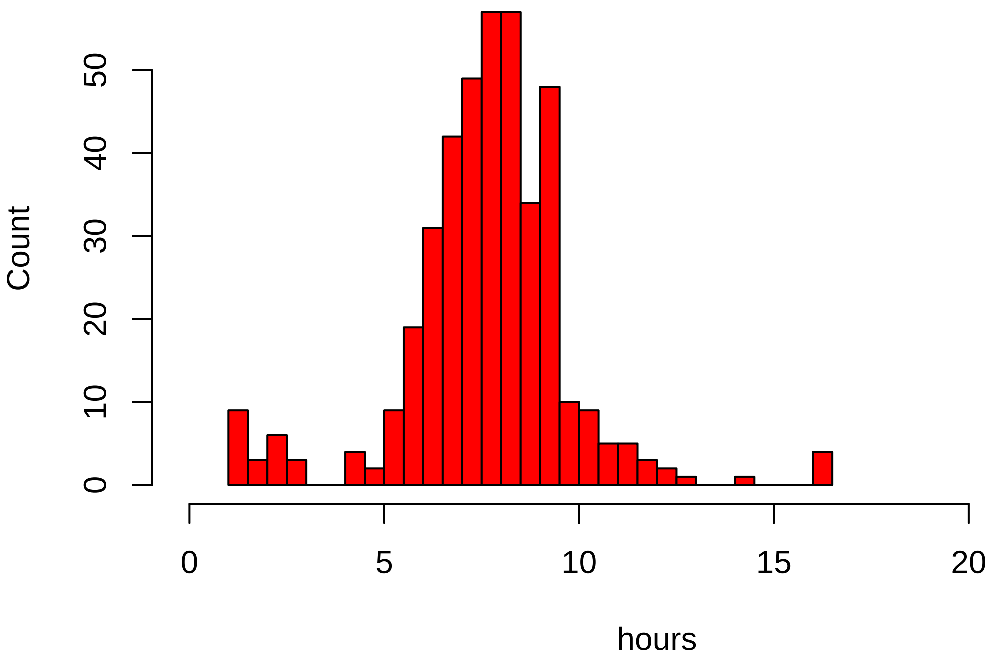
Figure 4 shows that there is not a clear relationship between total distance and sedentary minutes. Figure 5 shows that total distance directly correlates with total active minutes, which makes sense. Figure 6 shows that the number of calories burned also directly correlates with total active minutes, as expected. Figure 7 shows that calories burned increase at a faster rate as very active minutes increases. This can be a motivating trend. Figure 8 shows the relationship between total active minutes and total device time. It seems to show that as total device time increases so does total number of active minutes, which makes sense because if more time is recorded, it'll likely include more active minutes. Figure 9 shows a very clear relationship between total device time and sedentary minutes. As sedentary minutes increases, so does the total device time. Perhaps people who are more sedentary tend to use the device for longer times. Usually, sedentary time is more than active time. However, the wearability and comfort of the device can also be evaluated to see if there is any issue with wearing the device during intense activity.

Sleep dataset

We can generate some histograms to check the distribution of total time in bed and total time asleep.

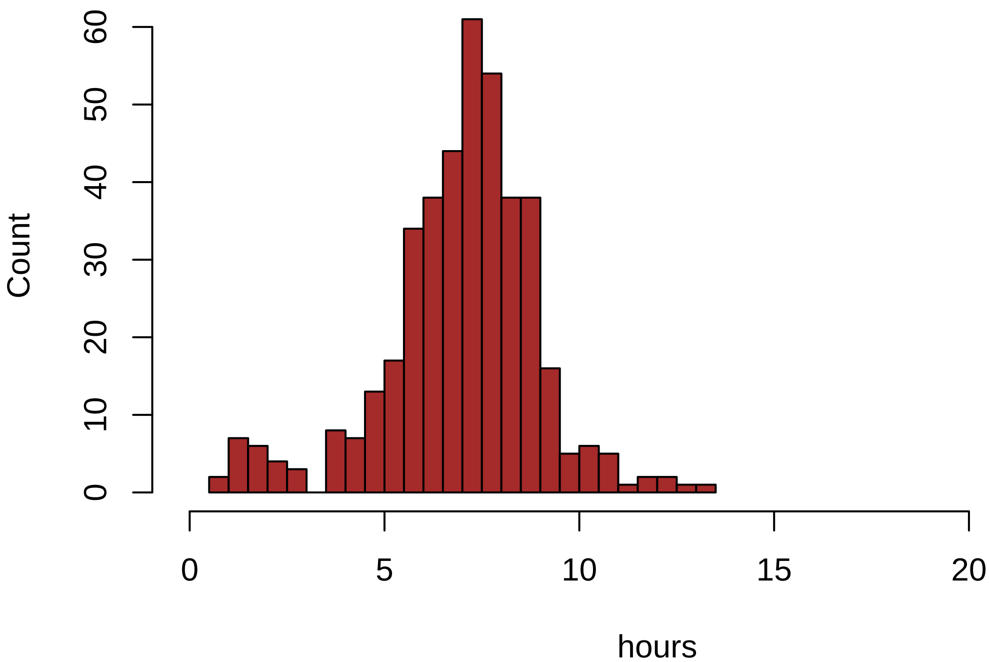
```
hist(dailysleep$TotalTimeInBed/60, main='Figure 10: Total Time in Bed', xlab='hours', ylab='Count', brea
```

Figure 10: Total Time in Bed




```
hist(dailysleep$TotalMinutesAsleep/60, main='Figure 11: Total Time Asleep', xlab='hours', ylab='Count',
```

Figure 11: Total Time Asleep




```
ggplot(dailysleep, aes(x=TotalTimeInBed, y=TotalMinutesAsleep)) + geom_point() + geom_smooth()+ labs(ti
```

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

Figure 12: Total Time Asleep vs Total Time in Bed

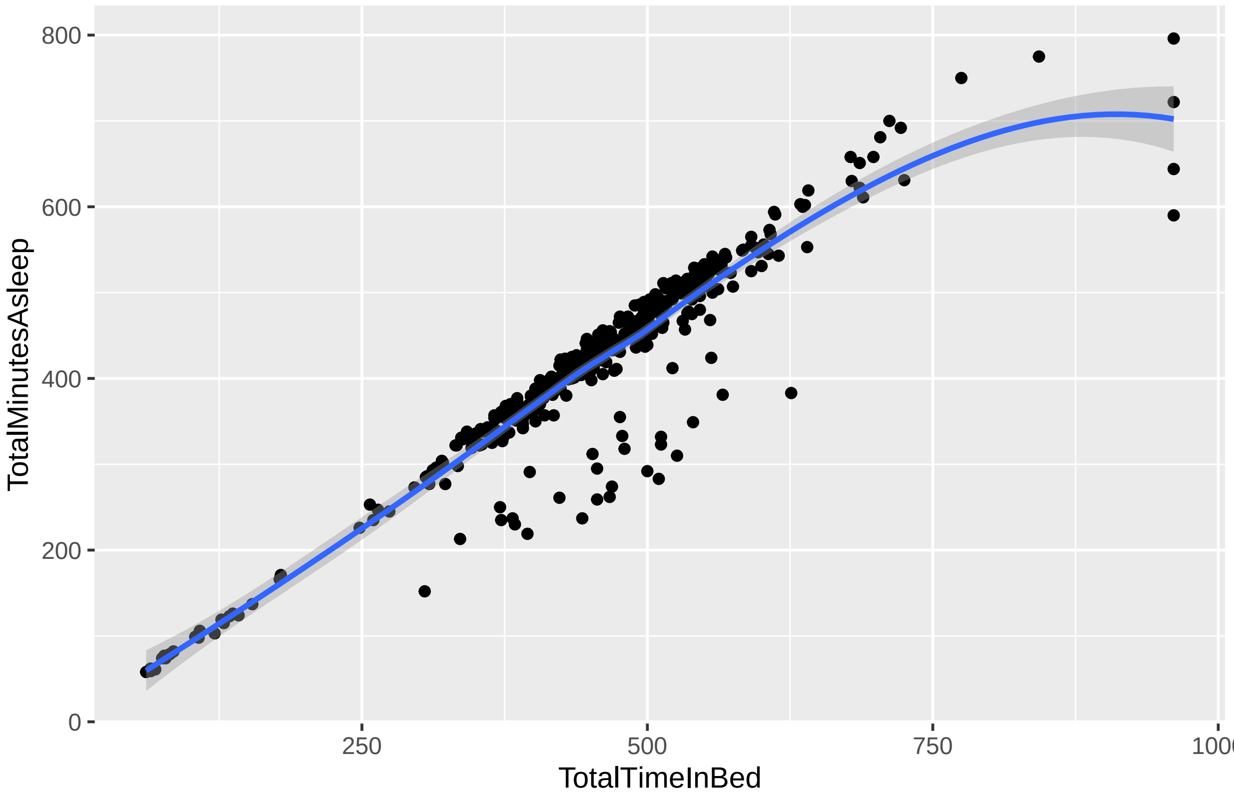


Figure 10 shows that most users spent about 7-9 hours in bed. Figure 11 shows that the total asleep has a very similar distribution to total time in bed, which is expected. Figure 12 confirms this correlation between the two.

Merging datasets

We can merge the daily activity and sleep datasets to see more trends. We can call this merged dataset 'sleep_activity'.

```
sleep_activity <- merge(dailysleep, dailyactivity, by=c("Id", "Date"))
head(sleep_activity)
```

```
##           Id       Date              SleepDay TotalSleepRecords
## 1 1503960366 2016-04-12 4/12/2016 12:00:00 AM                 1
## 2 1503960366 2016-04-13 4/13/2016 12:00:00 AM                 2
## 3 1503960366 2016-04-15 4/15/2016 12:00:00 AM                 1
## 4 1503960366 2016-04-16 4/16/2016 12:00:00 AM                 2
## 5 1503960366 2016-04-17 4/17/2016 12:00:00 AM                 1
## 6 1503960366 2016-04-19 4/19/2016 12:00:00 AM                 1
##   TotalMinutesAsleep TotalTimeInBed ActivityDate TotalSteps TotalDistance
## 1                327            346    4/12/2016      13162          8.50
## 2                384            407    4/13/2016      10735          6.97
## 3                412            442    4/15/2016       9762          6.28
## 4                340            367    4/16/2016      12669          8.16
## 5                700            712    4/17/2016       9705          6.48
## 6                304            320    4/19/2016      15506          9.88
##   TrackerDistance LoggedActivitiesDistance VeryActiveDistance
## 1            8.50                        0               1.88
```

```
## 2           6.97                        0              1.57
## 3           6.28                        0              2.14
## 4           8.16                        0              2.71
## 5           6.48                        0              3.19
## 6           9.88                        0              3.53
##   ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
## 1                     0.55                6.06                       0
## 2                     0.69                4.71                       0
## 3                     1.26                2.83                       0
## 4                     0.41                5.04                       0
## 5                     0.78                2.51                       0
## 6                     1.32                5.03                       0
##   VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## 1                25                  13                  328              728
## 2                21                  19                  217              776
## 3                29                  34                  209              726
## 4                36                  10                  221              773
## 5                38                  20                  164              539
## 6                50                  31                  264              775
##   Calories TotalActiveMinutes TotalDeviceTime
## 1     1985                366        18.23333
## 2     1797                257        17.21667
## 3     1745                272        16.63333
## 4     1863                267        17.33333
## 5     1728                222        12.68333
## 6     2035                345        18.66667
```

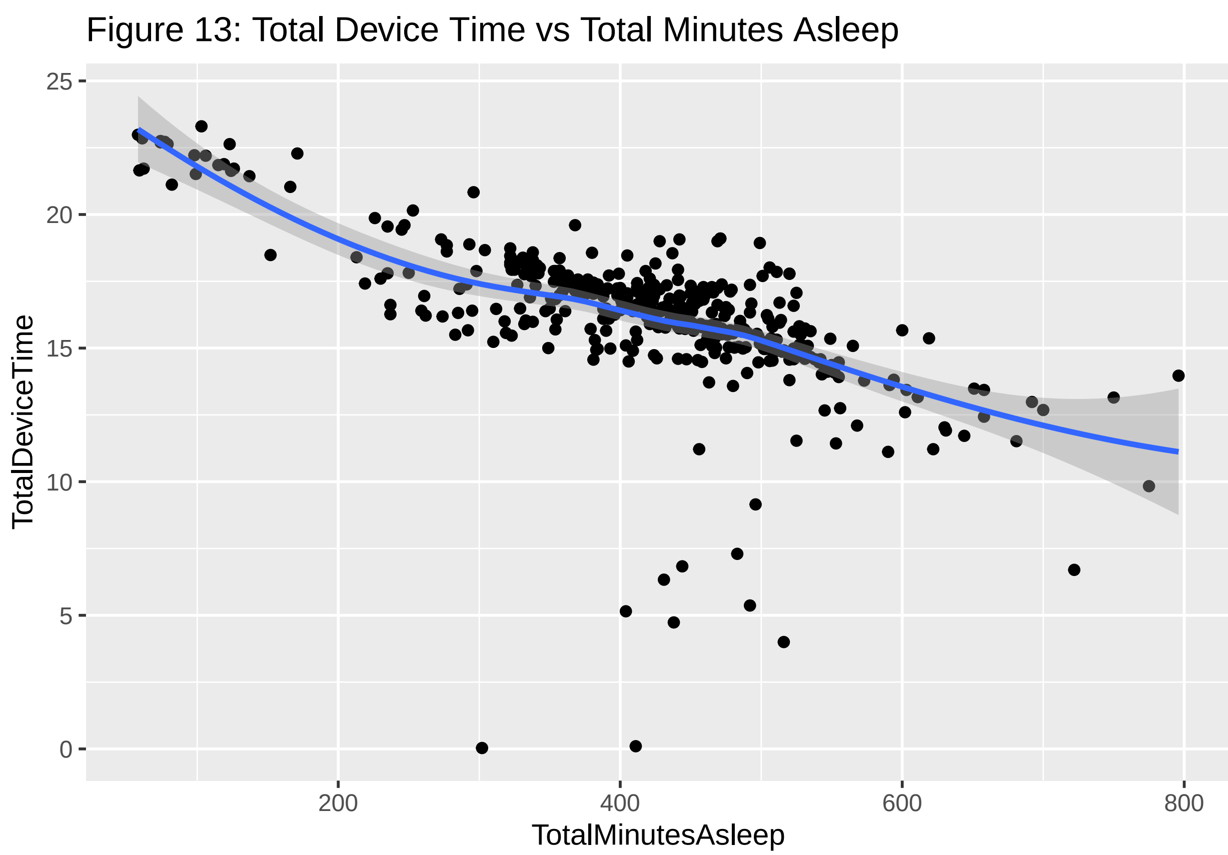
```
n_distinct(sleep_activity$Id)
```

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

Since we did an inner join, the combined dataset only has data from 24 distinct users, as was the case with the sleep dataset.

```
ggplot(sleep_activity, aes(TotalMinutesAsleep, TotalDeviceTime)) + geom_point() + geom_smooth() + labs(
```

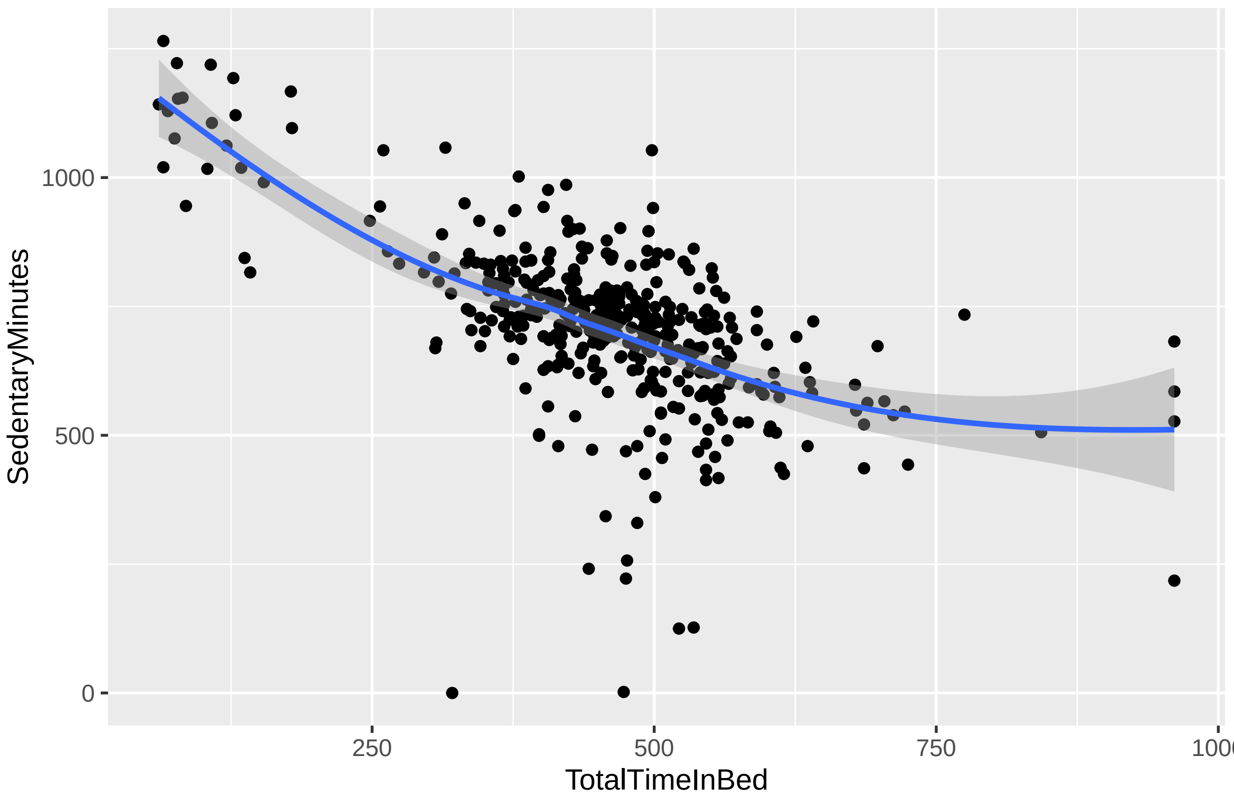
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(sleep_activity, aes(TotalTimeInBed, SedentaryMinutes)) + geom_point() + geom_smooth()+labs(title=
```

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

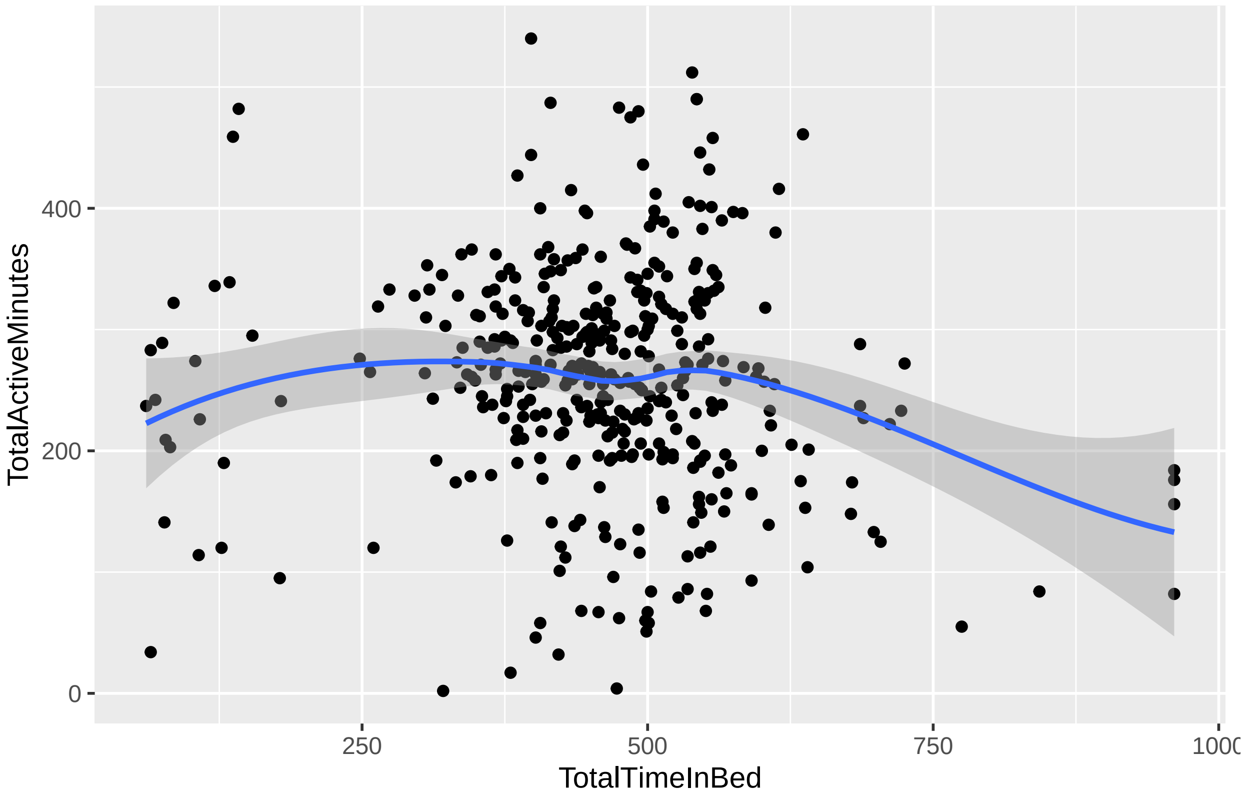
Figure 14: Sedentary Minutes vs Total Time in Bed




```
ggplot(sleep_activity, aes(TotalTimeInBed, TotalActiveMinutes)) + geom_point() + geom_smooth() + labs(t
```

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

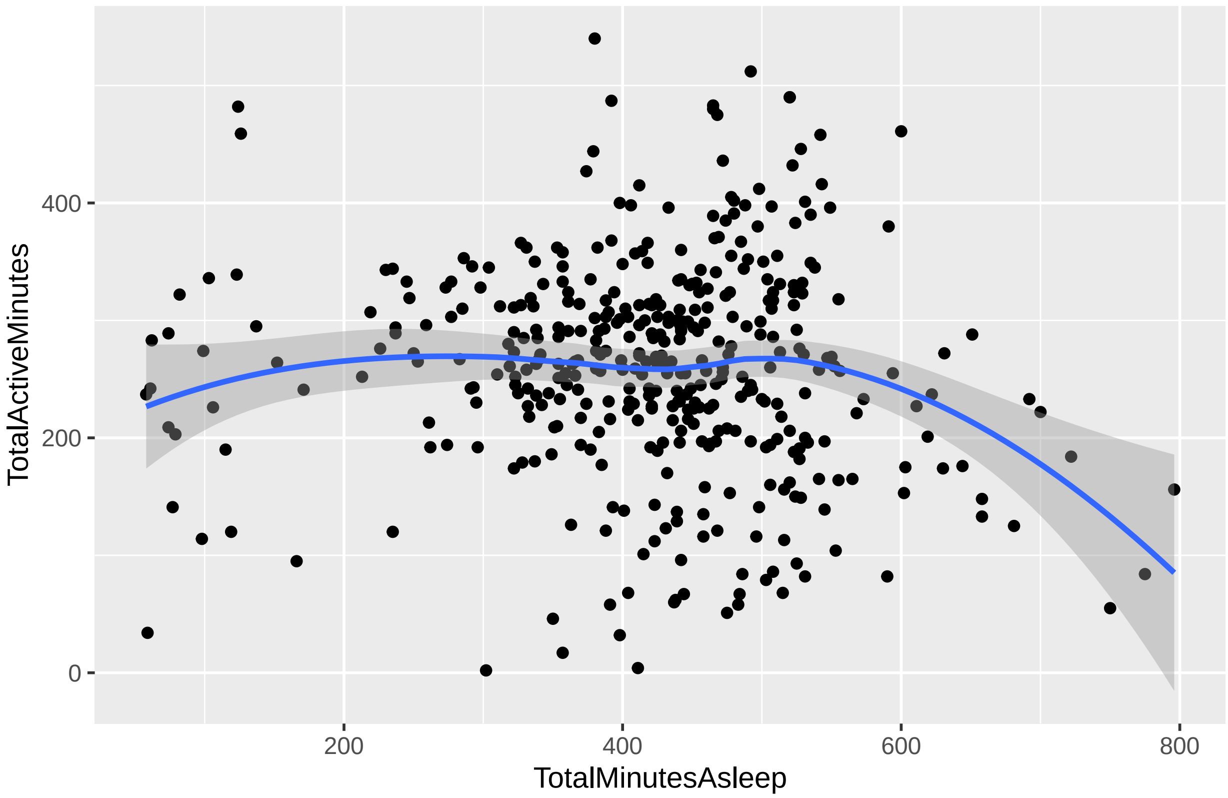
Figure 15: Total Active Minutes vs Total Time In Bed




```
ggplot(sleep_activity, aes(TotalMinutesAsleep, TotalActiveMinutes)) + geom_point() + geom_smooth() + la
```

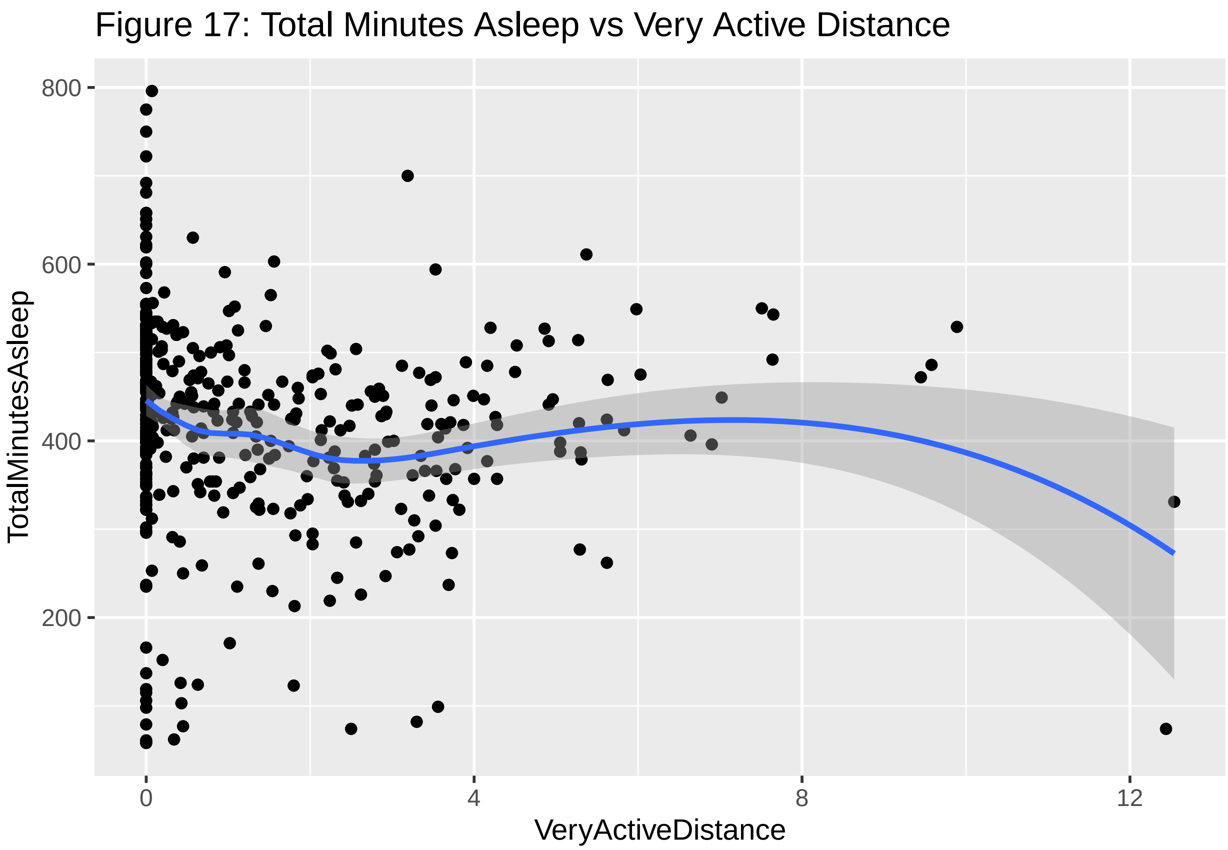
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figure 16: Total Active Minutes vs Total Minutes Asleep




```
ggplot(sleep_activity, aes(VeryActiveDistance, TotalMinutesAsleep)) + geom_point() + geom_smooth() + lal
```

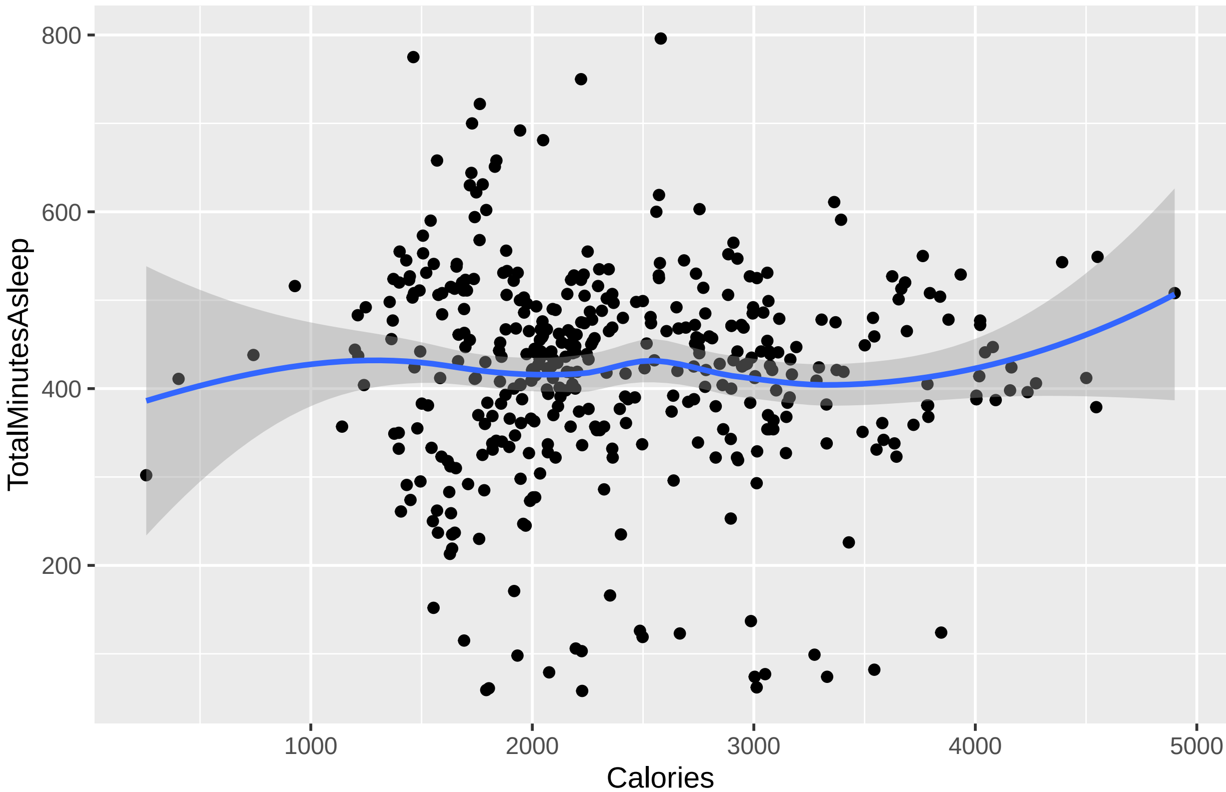
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(sleep_activity, aes(Calories, TotalMinutesAsleep)) + geom_point() + geom_smooth()+labs(title='Fig
```

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

Figure 18: Total Minutes Asleep vs Calories burned




```
ggplot(sleep_activity, aes(VeryActiveMinutes, TotalMinutesAsleep)) + geom_point() + geom_smooth()+labs(
```

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

Figure 19: Total Minutes Asleep vs Very Active Minutes

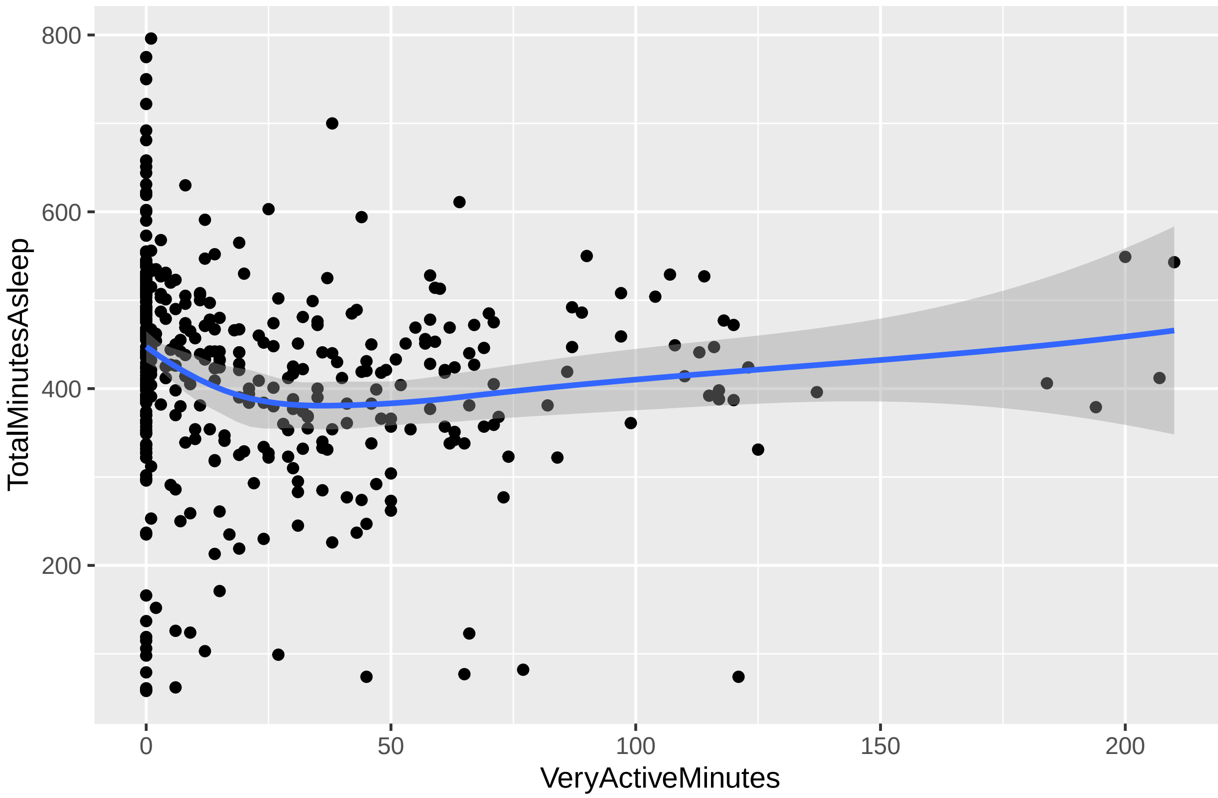


Figure 13 shows that as total minutes asleep increases, the total device time decreases. So people who sleep longer tend to wear their device less. This seems to be the opposite trend of what we saw in the sedentary minutes vs device time plot. Perhaps sedentary minutes and total time asleep are not measuring the same thing. This is what Figure 14 is also implying, that sedentary minutes are not the same as time in bed. Figures 15-19 shows that there is not much correlation between sleep and activity.

Correlation Coefficients

```
cor(sleep_activity$SedentaryMinutes, sleep_activity$TotalMinutesAsleep)
```

```
## [1] -0.599394
```

```
cor(sleep_activity$TotalActiveMinutes, sleep_activity$TotalMinutesAsleep)
```

```
## [1] -0.0637606
```

```
cor(sleep_activity$Calories, sleep_activity$TotalMinutesAsleep)
```

```
## [1] -0.02852571
```

```
cor(sleep_activity$TotalDistance, sleep_activity$TotalMinutesAsleep)
```

```
## [1] -0.1721427
```

```
cor(sleep_activity$Calories, sleep_activity$VeryActiveMinutes)
```

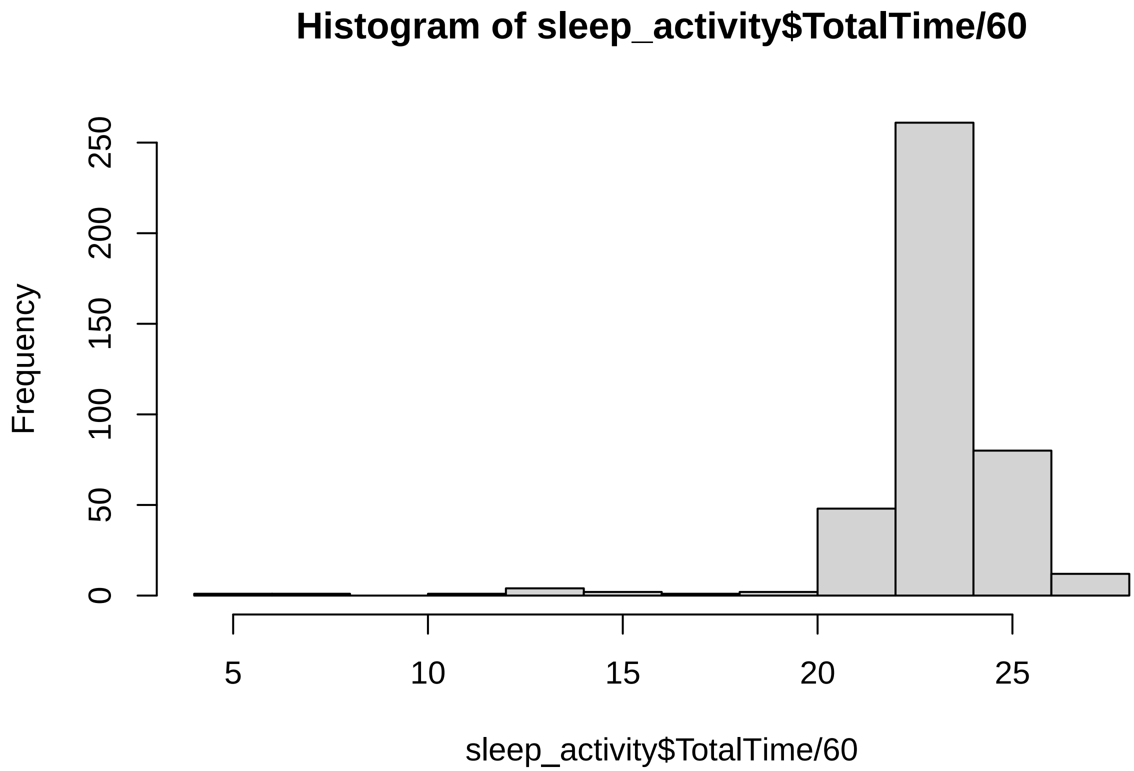
```
## [1] 0.6104889
```

```
cor(sleep_activity$TotalTimeInBed, sleep_activity$TotalMinutesAsleep)
```

```
## [1] 0.9304575
```

Let's look further into the sedentary vs total minutes.

```
sleep_activity <- mutate(sleep_activity, TotalTime = TotalActiveMinutes + SedentaryMinutes +
                              TotalMinutesAsleep)
hist(sleep_activity$TotalTime/60)
```



Conclusions

The goal of this analysis was to recommend marketing strategies to Bellabeat to increase the sales of its products. We used data from FitBit to find trends that may help in marketing of Bellabeat products. However, one of the main limitations of this dataset was that it included observations from both male and female while Bellabeat only makes products for women. To increase sales of products, we can target those individuals who we think will be more likely to use such devices, or we can show the benefit of these devices to motivate even those individuals who may not initially think about buying this product. This analysis showed that most users of FitBit had average health habits, such as sleeping 7 hours a day or covering about 5km a day. So a generally healthy audience would be likely to be interested in these products. We also saw that those who were most sedentary seemed to use this more. But another explanation may be that some active people took the device off during sedentary periods. This dataset had a lot of limitations including lack of explanation of how the variables were collected; therefore, it's hard to make definitive interpretations. To motivate individuals, we saw that calories burned had a strong correlation with very active minutes. Therefore, individuals can be motivated to engage in higher intensity activities. We also saw that the sleep time directly correlated with total time in bed. Therefore, individuals can be reminded to head to bed earlier to encourage better sleep habits. Further studies could collect data on the online activity of users of these devices to see which websites/social media is more often used. Then ads on those sites can be increased.