Bellabeat Project

Fatima

2023-12-14

Introduction

Welcome to the Bellabeat data analysis case study! Bellabeat is a high-tech manufacturer of health-focused products for women, and meet different characters and team members. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. You have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights you discover will then help guide marketing strategy for the company. You will present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat's marketing strategy.

In order to answer the key business questions for the case study, you will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act.

ASK

A clear statement of the business task: We are trying to see in what ways customers generally use their smartphone devices to see if any of those usages can be applied to Bellabeat to extend its functionality. Stakeholders: CCO/Co-founder.

PREPARE

A description of all data sources used

1. Daily activity for 31 participants for x days was recorded for total steps, total distance, very active distance, minutes, moderately active distance, minutes, light active distance, minutes, sedentary minutes, and calories.
2. Hourly calories burned by each participant
3. Hourly steps taken by each participant.
4. Minute-wise steps taken by each participant.
5. Sleep routine in the day.
6. Heart rate per second.

PROCESS

To begin the process of cleaning data for further data anLalysis, the following steps were utilized,

Installing and loading common packages and libraries

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

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

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

```
library("skimr")
library("here")
```

```
## here() starts at /Users/fatimaperwaizkhan/Documents/Google Analytics/Capstone
```

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.3     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.4     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/ conflicted package ]8;; to force all conflicts to become errors
```

```
library(rmarkdown)
library(dplyr)
library(ggplot2)
library(readxl)
```

Loading your CSV files

Here we'll create a dataframes for each of the CSV files in the folder.

```
daily_activity <- read.csv("~/Documents/Google Analytics/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")
sedentary_min <- daily_activity$SedentaryMinutes
hourly_calories <- read.csv("~/Documents/Google Analytics/Fitabase Data 4.12.16-5.12.16/hourlyCalories_merged.cs
v")
step_minute <- read.csv("~/Documents/Google Analytics/Fitabase Data 4.12.16-5.12.16/minuteStepsNarrow_merged.cs
v")
hourly_step <- read.csv("~/Documents/Google Analytics/Fitabase Data 4.12.16-5.12.16/hourlySteps_merged.csv")
sleep_day <- read.csv("~/Documents/Google Analytics/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
second_heartrate <- read.csv("~/Documents/Google Analytics/Fitabase Data 4.12.16-5.12.16/heartrate_seconds_merge
d.csv")
```

Exploring key tables

Taking a look at the daily_activity data.

```
skim_without_charts(daily_activity)
```

Data summary

Name	daily_activity
Number of rows	940
Number of columns	15
_______________________	
Column type frequency:	
character	1
numeric	14
________________________	
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: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	
Id	0	1	4.855407e+09	2.424805e+09	1503960366	2.320127e+09	4.445115e+09	6.962181e+09	8.877
TotalSteps	0	1	7.637910e+03	5.087150e+03	0	3.789750e+03	7.405500e+03	1.072700e+04	3.601
TotalDistance	0	1	5.490000e+00	3.920000e+00	0	2.620000e+00	5.240000e+00	7.710000e+00	2.803
TrackerDistance	0	1	5.480000e+00	3.910000e+00	0	2.620000e+00	5.240000e+00	7.710000e+00	2.803
LoggedActivitiesDistance	0	1	1.100000e-01	6.200000e-01	0	0.000000e+00	0.000000e+00	0.000000e+00	4.940
VeryActiveDistance	0	1	1.500000e+00	2.660000e+00	0	0.000000e+00	2.100000e-01	2.050000e+00	2.192
ModeratelyActiveDistance	0	1	5.700000e-01	8.800000e-01	0	0.000000e+00	2.400000e-01	8.000000e-01	6.480
LightActiveDistance	0	1	3.340000e+00	2.040000e+00	0	1.950000e+00	3.360000e+00	4.780000e+00	1.071

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
SedentaryActiveDistance	0	1	0.000000e+00	1.000000e-02	0	0.000000e+00	0.000000e+00	0.000000e+00	1.10C
VeryActiveMinutes	0	1	2.116000e+01	3.284000e+01	0	0.000000e+00	4.000000e+00	3.200000e+01	2.100
FairlyActiveMinutes	0	1	1.356000e+01	1.999000e+01	0	0.000000e+00	6.000000e+00	1.900000e+01	1.430
LightlyActiveMinutes	0	1	1.928100e+02	1.091700e+02	0	1.270000e+02	1.990000e+02	2.640000e+02	5.180
SedentaryMinutes	0	1	9.912100e+02	3.012700e+02	0	7.297500e+02	1.057500e+03	1.229500e+03	1.440
Calories	0	1	2.303610e+03	7.181700e+02	0	1.828500e+03	2.134000e+03	2.793250e+03	4.900

```
head(daily_activity)
```

	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance
	<dbl>	<chr>	<int>	<dbl>	<dbl>	<dbl>
1	1503960366	4/12/2016	13162	8.50	8.50	0
2	1503960366	4/13/2016	10735	6.97	6.97	0
3	1503960366	4/14/2016	10460	6.74	6.74	0
4	1503960366	4/15/2016	9762	6.28	6.28	0
5	1503960366	4/16/2016	12669	8.16	8.16	0
6	1503960366	4/17/2016	9705	6.48	6.48	0

6 rows | 1-7 of 16 columns

```
glimpse(hourly_calories)
```

```
## Rows: 22,099
## Columns: 3
## $ Id           <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150396036…
## $ ActivityHour <chr> "4/12/2016 12:00:00 AM", "4/12/2016 1:00:00 AM", "4/12/20…
## $ Calories     <int> 81, 61, 59, 47, 48, 48, 48, 47, 68, 141, 99, 76, 73, 66, …
```

```
str(hourly_calories)
```

```
## 'data.frame':    22099 obs. of  3 variables:
##  $ Id          : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr  "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM" "4/12/2016 2:00:00 AM" "4/12/2016 3:00:00
AM" ...
##  $ Calories    : int  81 61 59 47 48 48 48 47 68 141 ...
```

Identifying all the columns in the daily_activity data.

```
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"
```

Taking a look at the sleep_day data.

```
head(sleep_day)
```

	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
	<dbl>	<chr>	<int>	<int>	<int>
1	1503960366	4/12/2016 12:00:00 AM	1	327	346
2	1503960366	4/13/2016 12:00:00 AM	2	384	407
3	1503960366	4/15/2016 12:00:00 AM	1	412	442

	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
	<dbl>	<chr>	<int>	<int>	<int>
4	1503960366	4/16/2016 12:00:00 AM	2	340	367
5	1503960366	4/17/2016 12:00:00 AM	1	700	712
6	1503960366	4/19/2016 12:00:00 AM	1	304	320

6 rows

Grouping and Filtering Data

Deleting zero-entry rows (Removed rows that show no activity throughout the day, technically not possible)

```
daily_activity %>% group_by(Id) %>% drop_na() %>% summarize(mean_LightActiveDistance = mean(LightActiveDistance))
```

Id	mean_LightActiveDistance
<dbl>	<dbl>
1503960366	4.1529032
1624580081	2.6067742
1644430081	3.6090000
1844505072	1.6474193
1927972279	0.5070968
2022484408	4.9425806
2026352035	3.4361290
2320127002	2.9803226
2347167796	4.2216667
2873212765	4.1435484

1-10 of 33 rows Previous **1** 2 3 4 Next

```
daily_activity_clean<- daily_activity %>% filter(LightActiveDistance != "0" )
daily_activity_clean <- daily_activity_clean %>% filter(TotalSteps>9)
```

Understanding some summary statistics

How many unique participants are there in each dataframe? It looks like there may be more participants in the daily activity dataset than the sleep dataset.

```
n_distinct(daily_activity_clean$Id)
```

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

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

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

How many observations are there in each dataframe?

```
nrow(daily_activity_clean)
```

```
## [1] 853
```

```
nrow(sleep_day)
```

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

##ANALYZE What are some quick summary statistics we'd want to know about each data frame?

For the daily activity dataframe:

```
daily_activity_clean %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes
##  Min.   :   16   Min.   : 0.010   Min.   :   0.0
##  1st Qu.: 4933   1st Qu.: 3.380   1st Qu.: 720.0
##  Median : 8059   Median : 5.600   Median :1019.0
##  Mean   : 8358   Mean   : 6.005   Mean   : 950.1
##  3rd Qu.:11101   3rd Qu.: 7.920   3rd Qu.:1185.0
##  Max.   :36019   Max.   :28.030   Max.   :1438.0
```

For the sleep dataframe:

```
sleep_day %>%
  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
```

What does this tell us about how this sample of people's activities?

mean steps = 8358

mean distance = 6.005

mean Sedentary mins = 950.1

mean sleeps = 1.119

mean mins asleep = 419.5

mean time in bed = 458.6

Plotting a few explorations

What's the relationship between steps taken in a day and sedentary minutes? How could this help inform the customer segments that we can market to? E.g. position this more as a way to get started in walking more? Or to measure steps that you're already taking?

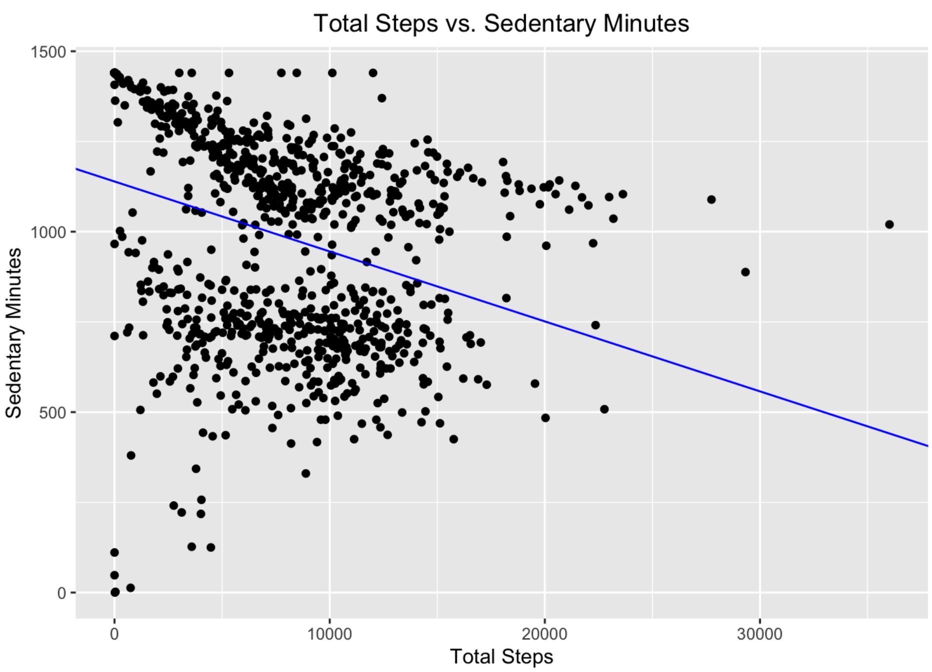
We should target audience who are taking less than 11,000 steps a day. This is because 75% of the population takes less than 11101 steps a day.

Using Linear Regression to Find the Relationship between Sedentary Minutes and Total Steps.

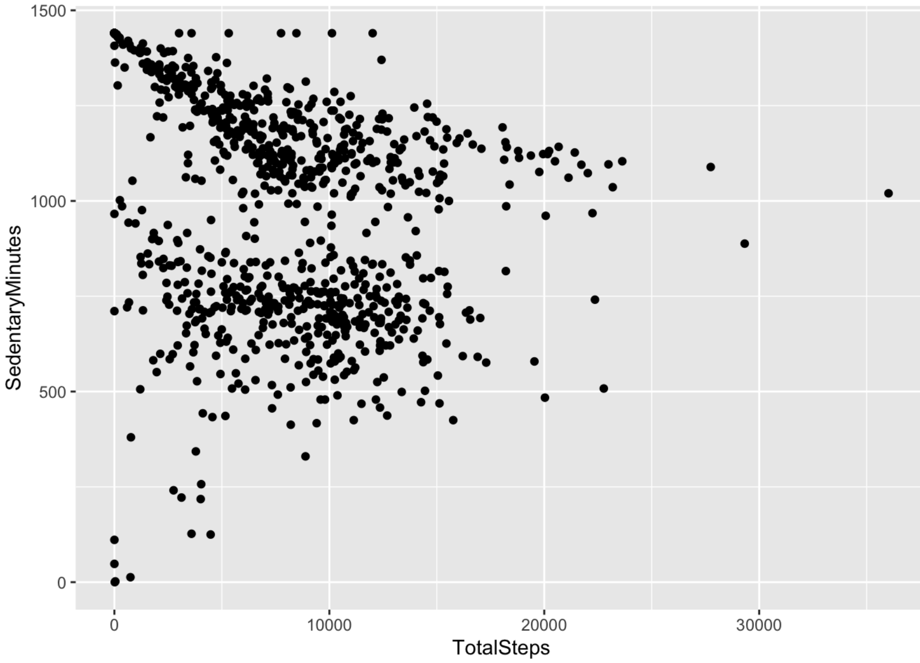
```
TotalStep_mod<-lm(data=daily_activity, SedentaryMinutes ~ TotalSteps ) #SIGNIFICANT
summary(TotalStep_mod)
```

```
##
## Call:
## lm(formula = SedentaryMinutes ~ TotalSteps, data = daily_activity)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1139.0  -237.9   104.4   242.7   579.2
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.139e+03  1.676e+01   67.97   <2e-16 ***
## TotalSteps  -1.939e-02  1.827e-03  -10.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 284.8 on 938 degrees of freedom
## Multiple R-squared:  0.1072, Adjusted R-squared:  0.1063
## F-statistic: 112.7 on 1 and 938 DF,  p-value: < 2.2e-16
```

```
#anova(TotalStep_mod)
ggplot(data = daily_activity) +
 geom_point(mapping = aes(x=TotalSteps, y=SedentaryMinutes))+
  geom_abline(slope = coef(TotalStep_mod)[["TotalSteps"]],
              intercept = coef(TotalStep_mod)[["(Intercept)"]], color = "blue")+
  labs(title = "Total Steps vs. Sedentary Minutes",
       y = "Sedentary Minutes", x = "Total Steps") +
          theme(plot.title = element_text(hjust = 0.5))
```




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

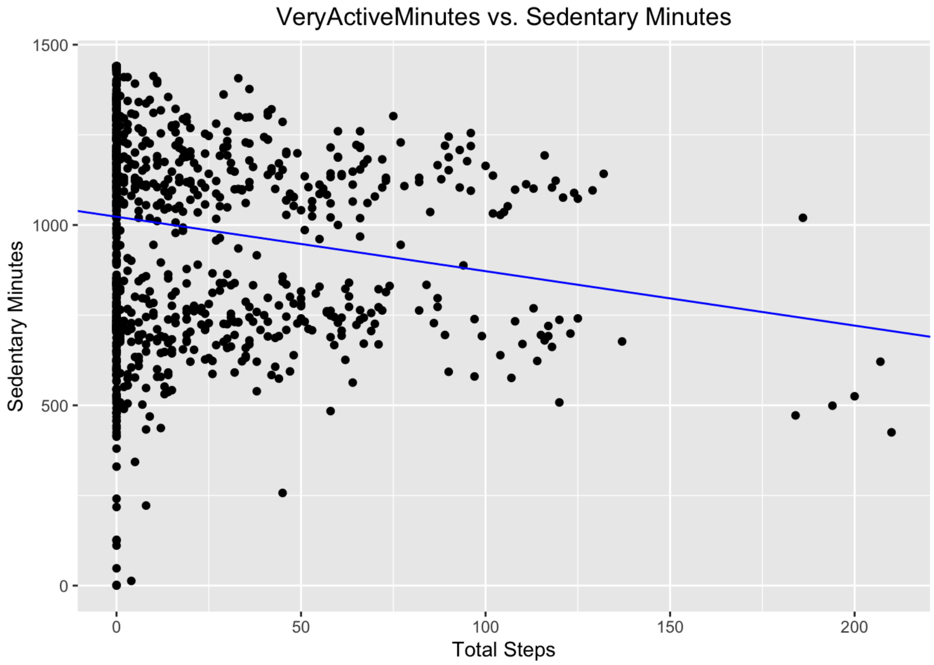


Using Linear Regression to Find the Relationship between Sedentary Minutes and Very Active Minutes.

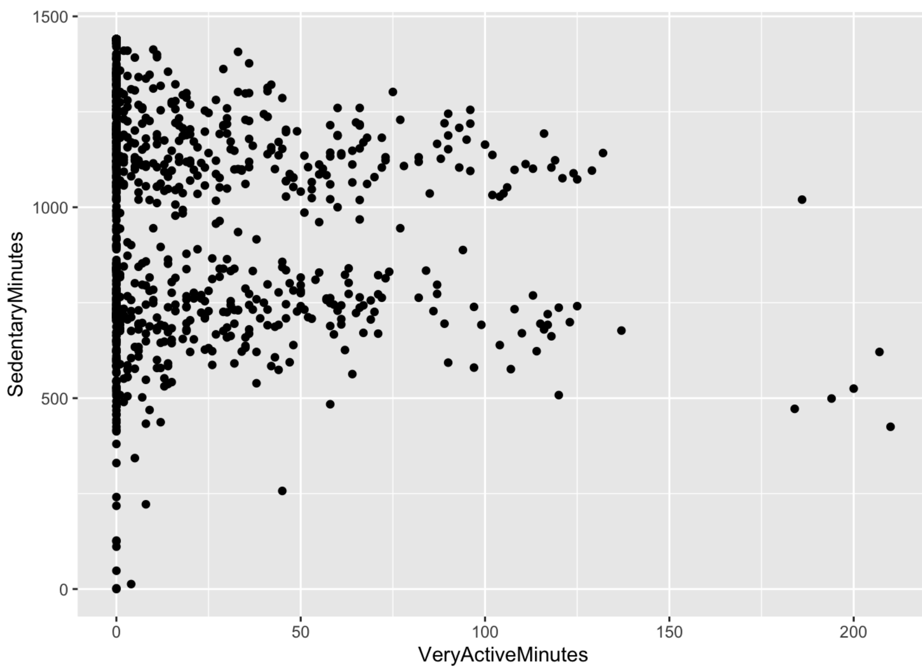
```
ActiveMin_mod<-lm(data=daily_activity, SedentaryMinutes ~ VeryActiveMinutes ) #SIGNIFICANT
summary(ActiveMin_mod)
```

```
##
## Call:
## lm(formula = SedentaryMinutes ~ VeryActiveMinutes, data = daily_activity)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1023.18  -251.30    68.76   248.58   433.67
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1023.1789    11.5381  88.679  < 2e-16 ***
## VeryActiveMinutes   -1.5104     0.2954  -5.113 3.84e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 297.3 on 938 degrees of freedom
## Multiple R-squared:  0.02712,    Adjusted R-squared:  0.02608
## F-statistic: 26.14 on 1 and 938 DF,  p-value: 3.841e-07
```

```
ggplot(data = daily_activity) +
 geom_point(mapping = aes(x=VeryActiveMinutes, y=SedentaryMinutes))+
  geom_abline(slope = coef(ActiveMin_mod)[["VeryActiveMinutes"]],
              intercept = coef(ActiveMin_mod)[["(Intercept)"]], color = "blue")+
  labs(title = "VeryActiveMinutes vs. Sedentary Minutes",
       y = "Sedentary Minutes", x = "Total Steps") +
          theme(plot.title = element_text(hjust = 0.5))
```



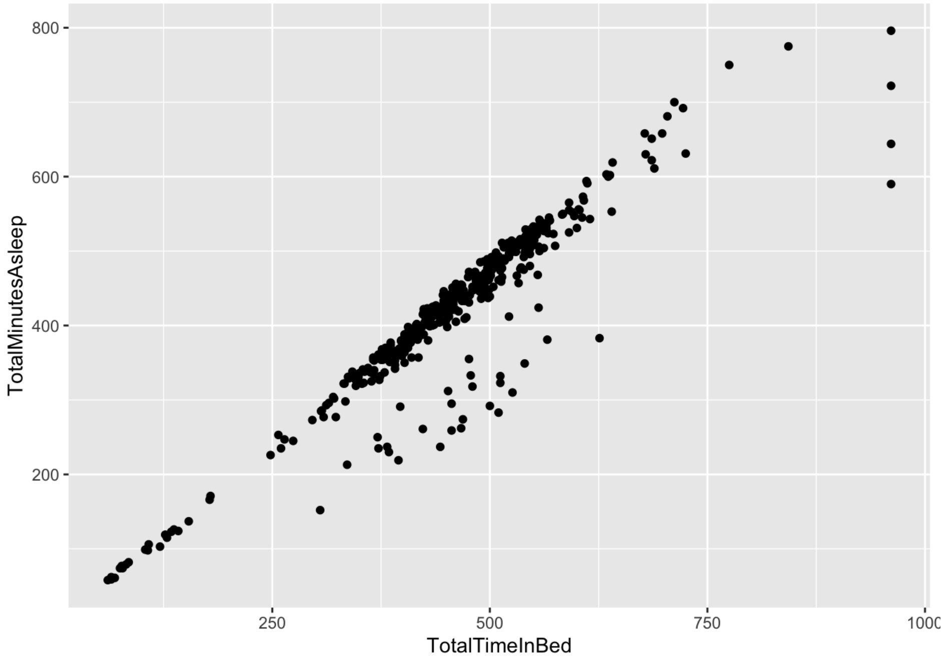

```
ggplot(data=daily_activity, aes(x=VeryActiveMinutes, y=SedentaryMinutes)) + geom_point()
```



Relationship between minutes asleep and time in bed

What's the relationship between minutes asleep and time in bed? You might expect it to be almost completely linear - are there any unexpected trends? Some people spend more time in bed but not necessarily sleeping.

```
ggplot(data=sleep_day, aes(x=TotalTimeInBed, y=TotalMinutesAsleep)) + geom_point()
```



What could these trends tell you about how to help market this product? Or areas where you might want to explore further? We can explore what do people do when they are in bed but not sleeping. Are they using phone or reading?

Merging the datasets for daily_activity and sleep together

Merging the clean data set for daily_activity with the corresponding values from sleep_activity. The merging was done based on ID and date matching.

```
new_sleep <- as.Date(sleep_day$SleepDay, "%m/%d/%y")
#new_date <- as.Date(daily_activity$ActivityDay, "%m/%d/%y")
sleep_day <- mutate(sleep_day, Date_only = new_sleep)
daily_activity_clean <- mutate(daily_activity_clean, new_date = as.Date(daily_activity_clean$ActivityDate, "%m/%
d/%y"))
combined_data <- left_join(sleep_day, daily_activity_clean, by= c("Id" = "Id", "Date_only" = "new_date"))
n_distinct(combined_data$Id)
```

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

Merging the data set for hourly steps and hourly calories based on ID and date matching and exploring their relationship.

```
hourly_combined <- left_join(hourly_calories, hourly_step, by= c("Id" = "Id", "ActivityHour" = "ActivityHour"))
str(hourly_combined)
```

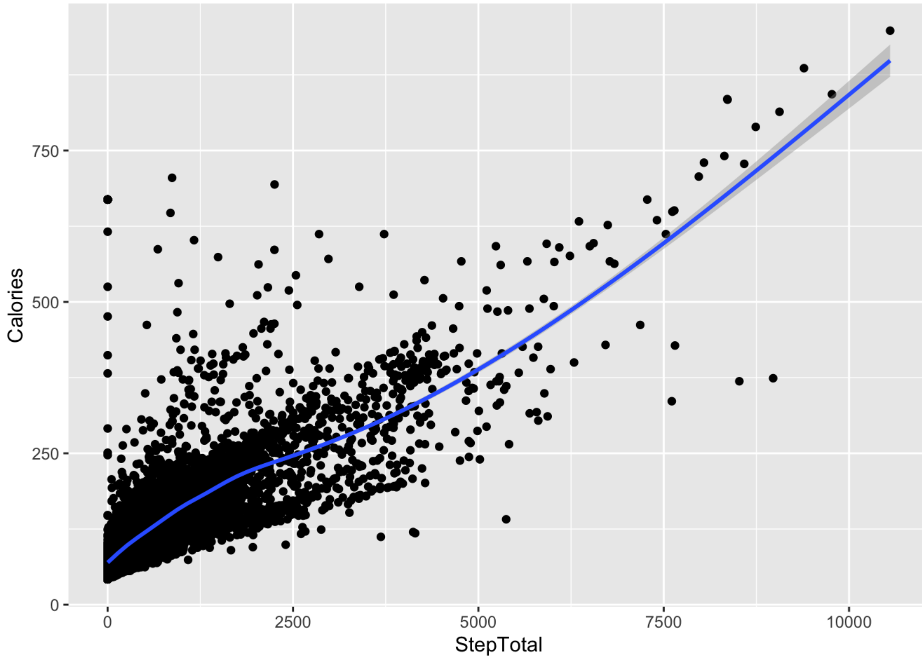
```
## 'data.frame':    22099 obs. of  4 variables:
##  $ Id          : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr  "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM" "4/12/2016 2:00:00 AM" "4/12/2016 3:00:00
AM" ...
##  $ Calories    : int  81 61 59 47 48 48 48 47 68 141 ...
##  $ StepTotal   : int  373 160 151 0 0 0 0 0 250 1864 ...
```

```
glimpse(hourly_combined)
```

```
## Rows: 22,099
## Columns: 4
## $ Id           <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150396036…
## $ ActivityHour <chr> "4/12/2016 12:00:00 AM", "4/12/2016 1:00:00 AM", "4/12/20…
## $ Calories     <int> 81, 61, 59, 47, 48, 48, 48, 47, 68, 141, 99, 76, 73, 66, …
## $ StepTotal    <int> 373, 160, 151, 0, 0, 0, 0, 0, 250, 1864, 676, 360, 253, 2…
```

```
ggplot(data=hourly_combined, aes(x=StepTotal, y=Calories)) + geom_point() + geom_smooth()
```

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



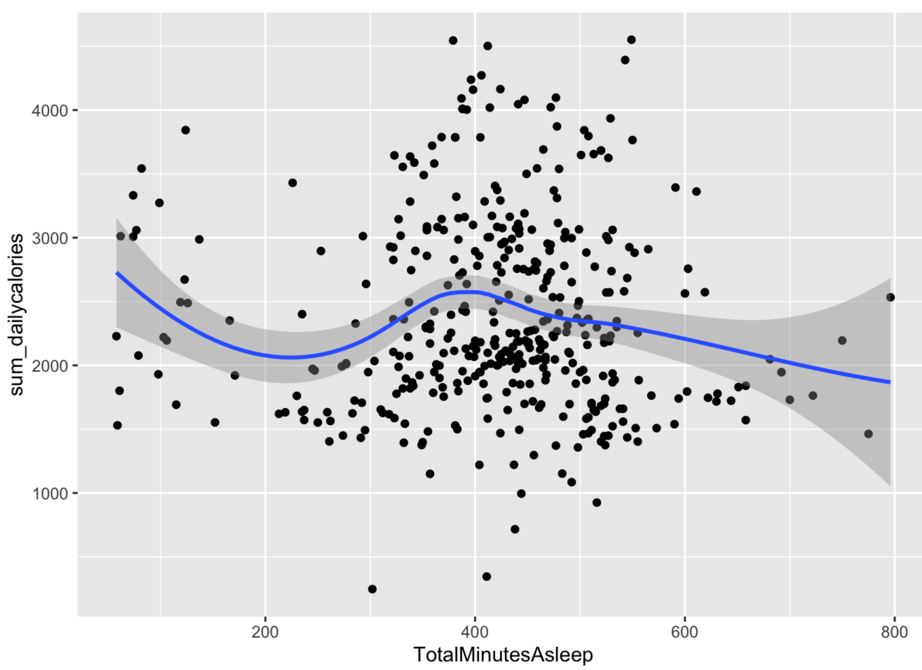
Further merging the two merged datasets using left join.

```
hourly_combined <- mutate(hourly_combined, Date_only = as.Date(hourly_combined$ActivityHour, "%m/%d/%y"))
sum_daily <- hourly_combined %>% group_by(Id, Date_only) %>% summarize(sum_dailycalories = sum(Calories), sum_dai
lysteps = sum(StepTotal))
```

```
## `summarise()` has grouped output by 'Id'. You can override using the `.groups`
## argument.
```

```
hourly_combined2 <- left_join(combined_data,sum_daily, by= c("Id" = "Id", "Date_only" = "Date_only"))
ggplot(data=hourly_combined2, aes(x=TotalMinutesAsleep, y=sum_dailycalories)) + geom_point() + geom_smooth()
```

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



Average sleep by each participant

```
hourly_combined2<-mutate(hourly_combined2, sleep_hours = TotalMinutesAsleep/60)
mean_sleep_perperson <- hourly_combined2 %>% drop_na() %>%  group_by(Id) %>% summarize(mean_sleep = mean(sleep_ho
urs))
mean(mean_sleep_perperson$mean_sleep)
```

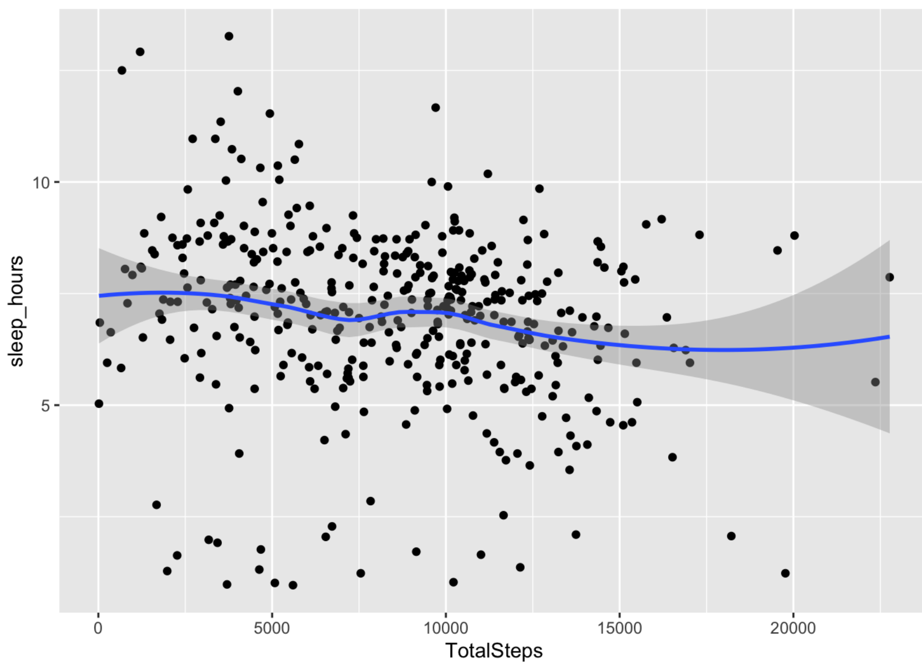
```
## [1] 6.294125
```

Relationship between Sleep hours and Total Steps/Calories.

Both steps and calories are inversely related to sleep hours.

```
ggplot(data=hourly_combined2, aes(x=TotalSteps, y=sleep_hours)) + geom_point() + geom_smooth()
```

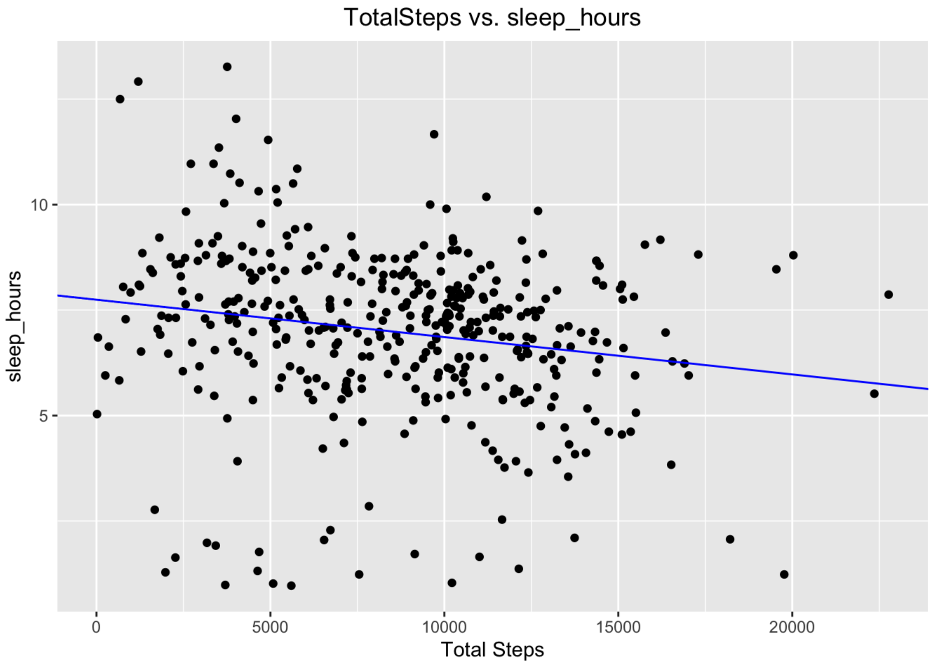
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
steps_mod<-lm(data=hourly_combined2, sleep_hours ~ TotalSteps ) #SIGNIFICANT
summary(steps_mod)
```

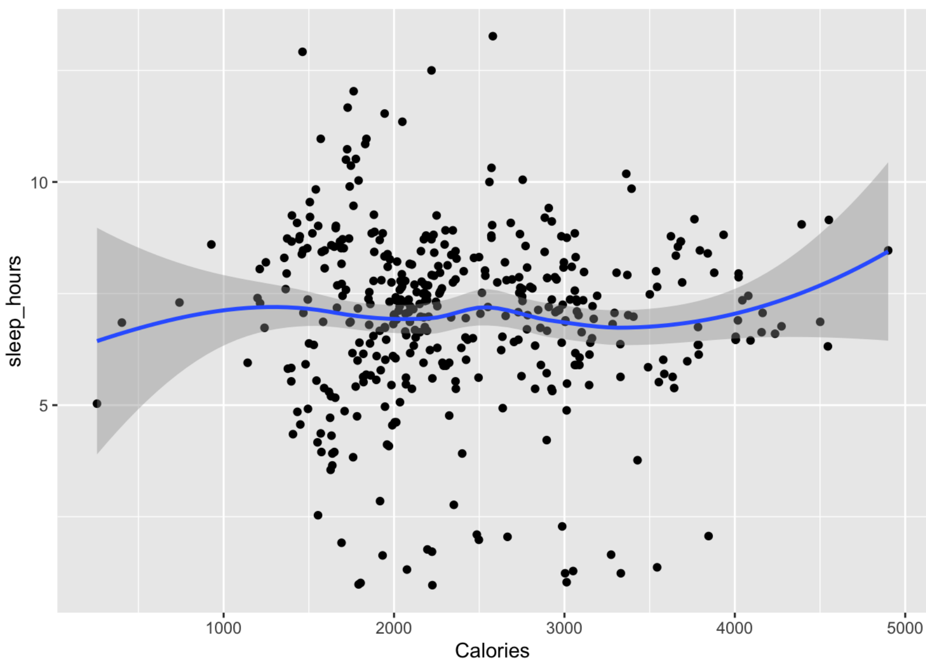
```
## 
## Call:
## lm(formula = sleep_hours ~ TotalSteps, data = hourly_combined2)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -6.4369 -0.9444  0.2144  1.1696  5.8514 
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  7.748e+00  2.184e-01  35.485  < 2e-16 ***
## TotalSteps  -8.867e-05  2.299e-05  -3.856 0.000134 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.94 on 411 degrees of freedom
## Multiple R-squared:  0.03492,    Adjusted R-squared:  0.03257 
## F-statistic: 14.87 on 1 and 411 DF,  p-value: 0.0001336
```

```
ggplot(data = hourly_combined2) +
 geom_point(mapping = aes(x=TotalSteps, y=sleep_hours))+
  geom_abline(slope = coef(steps_mod)[["TotalSteps"]],
              intercept = coef(steps_mod)[["(Intercept)"]], color = "blue")+
  labs(title = "TotalSteps vs. sleep_hours",
       y = "sleep_hours", x = "Total Steps") +
          theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(data=hourly_combined2, aes(x=Calories, y=sleep_hours)) + geom_point() + geom_smooth()
```

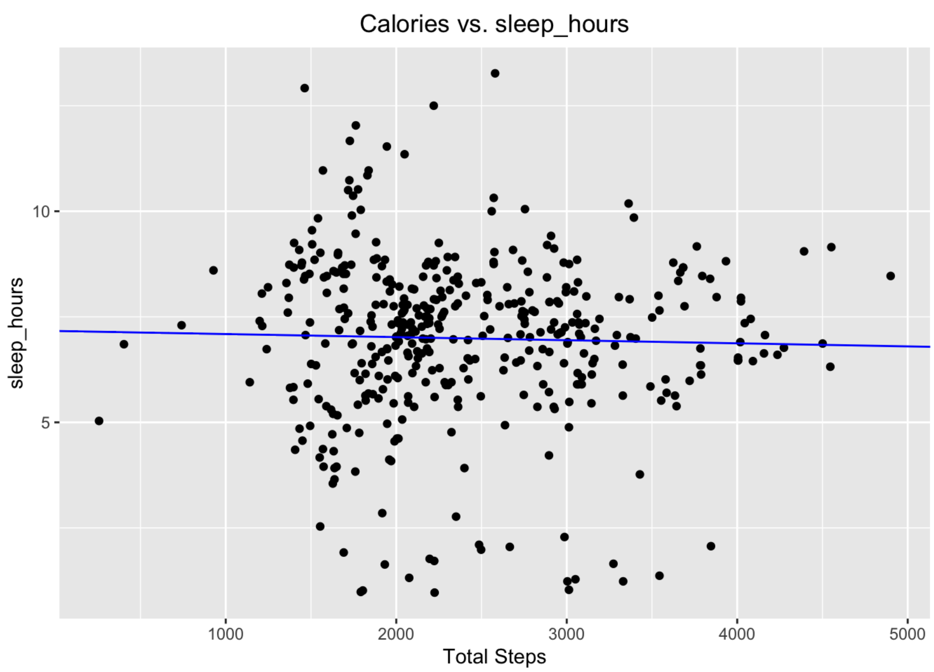
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
calories_mod<-lm(data=hourly_combined2, sleep_hours ~ Calories ) #SIGNIFICANT
summary(calories_mod)
```

```
##
## Call:
## lm(formula = sleep_hours ~ Calories, data = hourly_combined2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0525 -0.9280  0.2049  1.1530  6.2890
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.168e+00  3.207e-01  22.351   <2e-16 ***
## Calories    -7.375e-05  1.275e-04  -0.579    0.563
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.974 on 411 degrees of freedom
## Multiple R-squared:  0.0008137,  Adjusted R-squared:  -0.001617
## F-statistic: 0.3347 on 1 and 411 DF,  p-value: 0.5632
```

```
ggplot(data = hourly_combined2) +
 geom_point(mapping = aes(x=Calories, y=sleep_hours))+
  geom_abline(slope = coef(calories_mod)[["Calories"]],
              intercept = coef(calories_mod)[["(Intercept)"]], color = "blue")+
  labs(title = "Calories vs. sleep_hours",
       y = "sleep_hours", x = "Total Steps") +
          theme(plot.title = element_text(hjust = 0.5))
```



SHARE

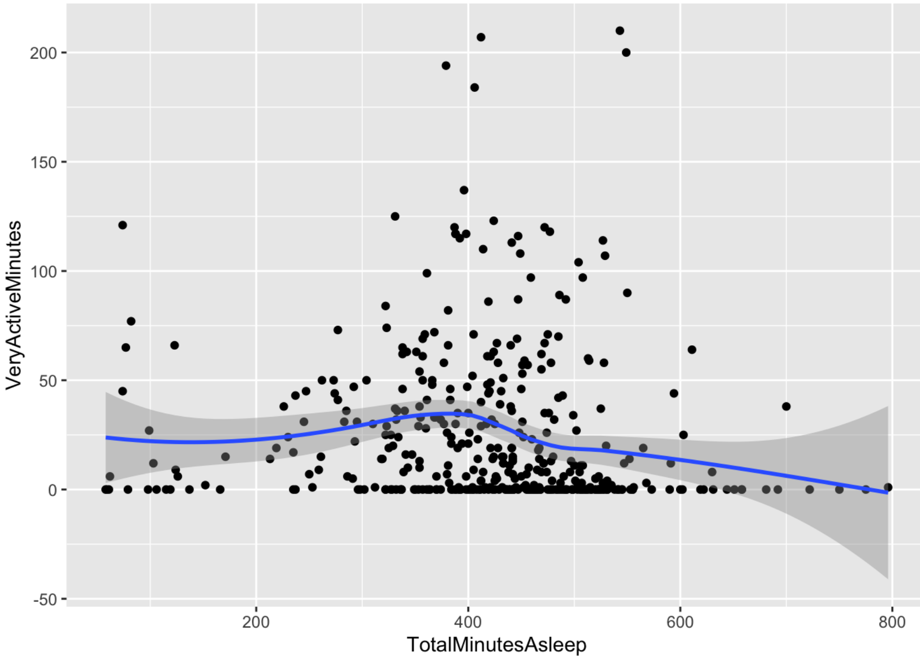
What is relationship between sleep minutes and going to the gym (Very Active Minutes)?

It seems like people who sleep mean minutes = 419.5 are more likely to be very active in the day. People on extreme sleep habits are likely to have lesser active minutes.

Do participants who sleep more also take more steps or fewer steps per day? no apparent relationship.

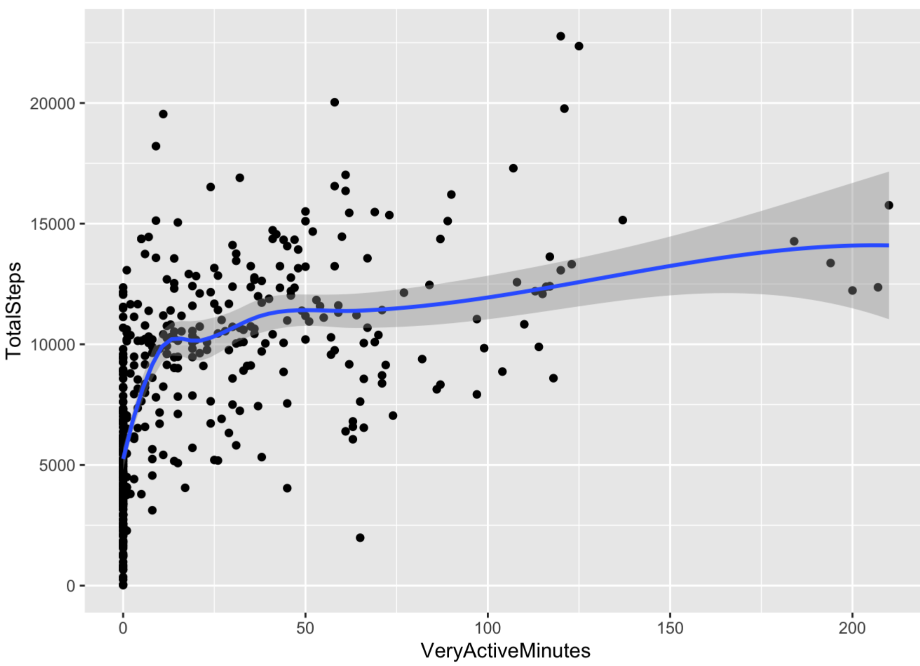
```
ggplot(data=combined_data, aes(x=TotalMinutesAsleep, y=VeryActiveMinutes)) + geom_point() + geom_smooth()
```

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



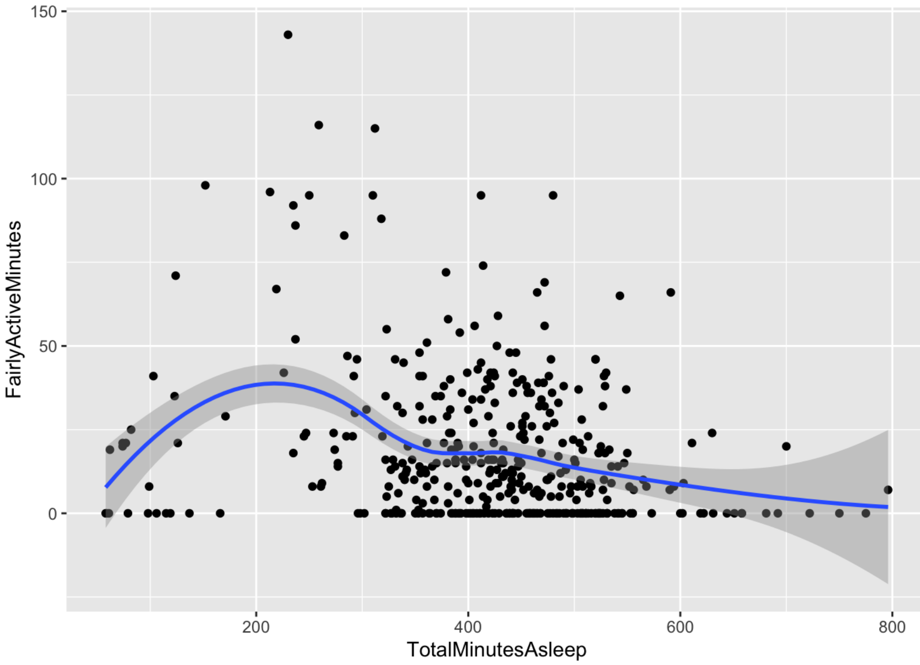

```
ggplot(data=combined_data, aes(x=VeryActiveMinutes, y=TotalSteps)) + geom_point() + geom_smooth()
```

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



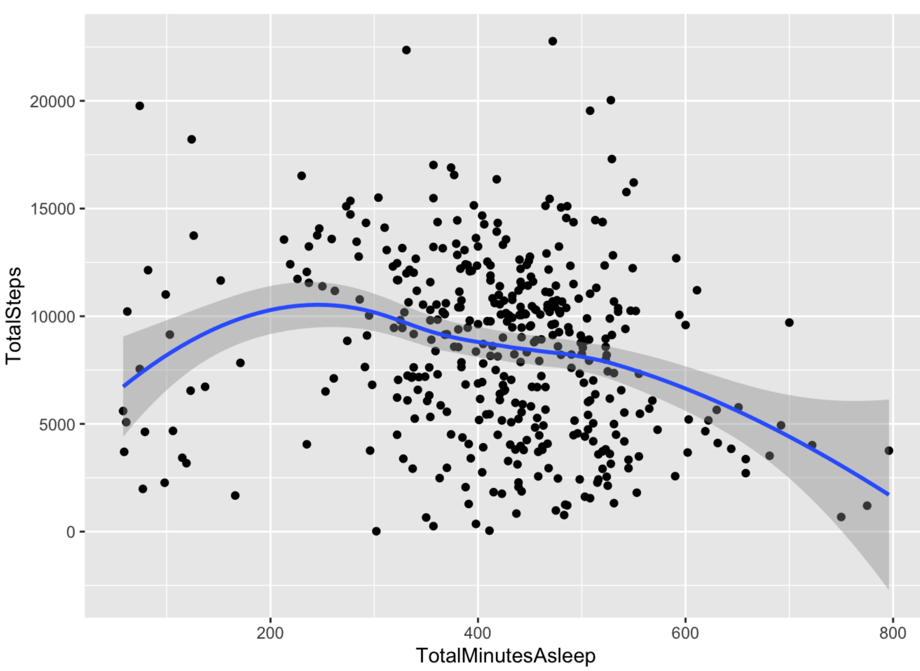

```
ggplot(data=combined_data, aes(x=TotalMinutesAsleep, y=FairlyActiveMinutes)) + geom_point() + geom_smooth()
```

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



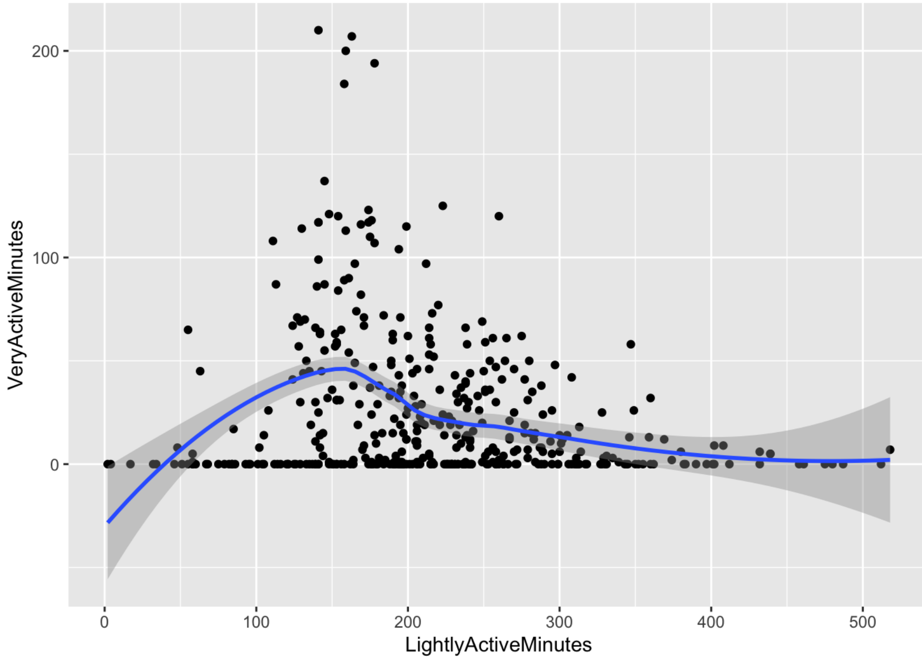

```r
ggplot(data=combined_data, aes(x=TotalMinutesAsleep, y=TotalSteps)) + geom_point()+ geom_smooth()
```

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




```r
ggplot(data=combined_data, aes(x=LightlyActiveMinutes, y=VeryActiveMinutes)) + geom_point()+ geom_smooth()
```

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

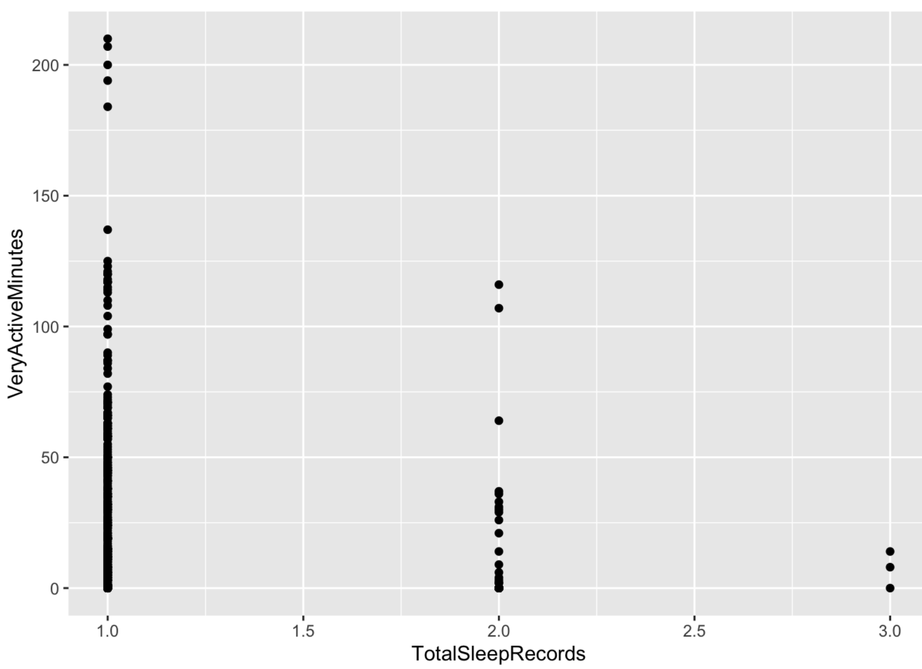


How many minutes of the Sedetary minutes are spend not-sleeping.

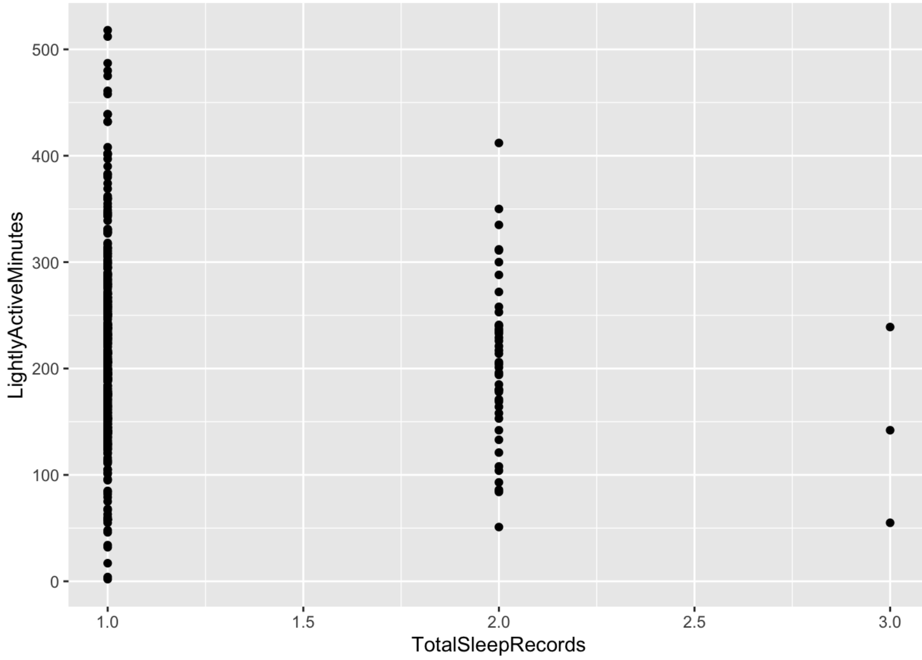
```
combined_data <- mutate(combined_data, InactivityOutsideSleep = SedentaryMinutes - TotalMinutesAsleep)
```

People who take a nap and stay active? People who sleep only once in the day are more likely to stay more active.

```
ggplot(data=combined_data, aes(x=TotalSleepRecords, y=VeryActiveMinutes)) + geom_point()
```

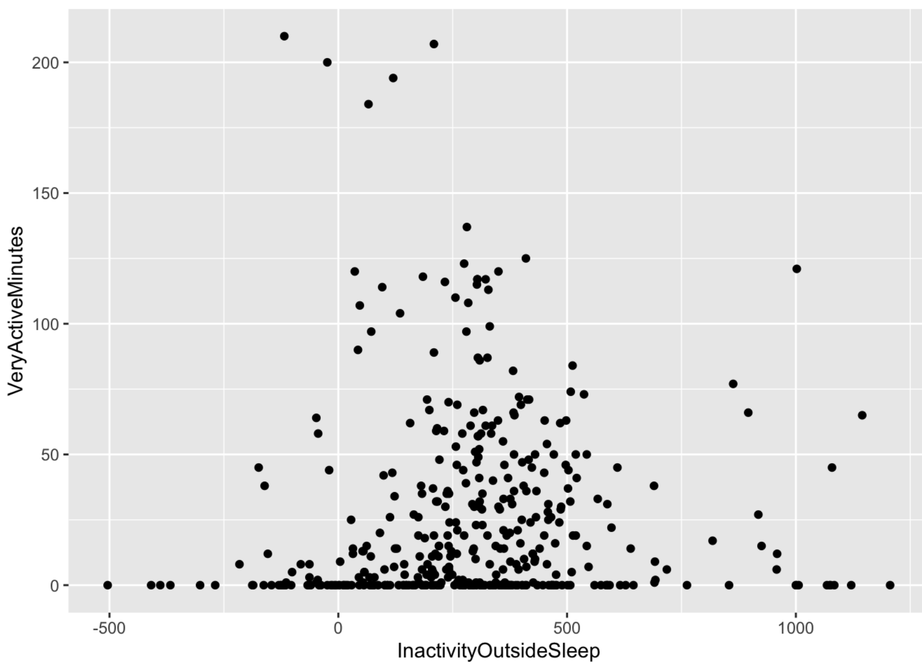



```
ggplot(data=combined_data, aes(x=TotalSleepRecords, y=LightlyActiveMinutes)) + geom_point()
```



Inactivity outside sleep and Active Minutes

```
ggplot(data=combined_data, aes(x=InactivityOutsideSleep, y=VeryActiveMinutes)) + geom_point()
```



CONCLUSION

Following are the important findings from the analysis,

1. While the mean minutes asleep for a participant is 419.5, the mean time people spend in their beds is 458.6. This means participants like to do some of their leisure activities in the bed.
2. While the total mean steps for participants is 8358, this mean for any new marketing campaign associated with steps, we should target audience who are taking less than 11,000 steps a day. This is because 75% of the population takes less than 11101 steps a day.

4.The Relationship between Sedentary Minutes and Total Steps shows a significant inverse relationship. This is intuitive that people spending more time in sedentary acts are less likely to take steps.

5. Running a linear regression model shows that the Relationship between Sedentary Minutes and Very Active Minutes is inverse. This means the more active people are they spend less Sedentary Minutes.
6. The relationship between asleep minutes and the time spent in the bed shows that some people spend more time in bed but not necessarily sleeping.

7. The relationship between total very active minutes and total steps doesn't show a constant linear relationship, which shows that people who non-running activities such as weight lifting to for very active exercise.
8. The average seen for the participants is 6.294125 hours a day, which is close to the recommended sleep of 7 hours a day.
9. Most participants wore their smart watches every hour from 11th April to 12th May.

The above findings can be applied to Bellabeat's Marketing Strategy?

1. Given that the the total mean steps for participants is 8358, the marketing campaign should be targeted to audience who are taking less than 11,000 steps a day. This is because 75% of the population takes less than 11101 steps a day. Bellabeat can send people a reminder to at least complete 10,000 steps a day.
2. The relationship between total steps and sleep hours show an inverse relationship. This could mean that over walking could affect people's sleep. Bellabeat could suggest a walking range for people for better sleep.