Bellabeat Smart Device

Chinonso

2024-01-21

Setting up my environment

Setting up my R environment by loading: tidyverse, ggplot2, dplyr, skimr, janitor, here

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

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

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

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

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

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

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

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(skimr)
```

```r
library(dplyr)
library(janitor)
```

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

Loading my dataset

```r
daily_activity <- read.csv("dailyActivity_merged.csv")
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
```

```r
sleep_day <- read.csv("sleepDay_merged.csv")
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
```

Familiarizing with the columns names to understand them

```
colnames(daily_activity)
```

```
##  [1] "Id"                       "ActivityDate"
##  [3] "TotalSteps"               "TotalDistance"
##  [5] "TrackerDistance"          "LoggedActivitiesDistance"
##  [7] "VeryActiveDistance"       "ModeratelyActiveDistance"
##  [9] "LightActiveDistance"      "SedentaryActiveDistance"
## [11] "VeryActiveMinutes"        "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes"     "SedentaryMinutes"
## [15] "Calories"
```

```
colnames(sleep_day)
```

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

Understanding some summary statistics

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

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

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

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

Checking how many observations in each dataframe

```
nrow(daily_activity)
```

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

```
nrow(sleep_day)
```

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

Renaming the column names

```
daily_activity<- rename_with(daily_activity, tolower)
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
```

```
sleep_day <- rename_with (sleep_day, tolower)
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
```

Checking for quick summary of daily_activity dataframe.

```
daily_activity %>%
  select(totalsteps,
         totaldistance,
         sedentaryminutes)%>%
  summary()
```

```
##    totalsteps    totaldistance    sedentaryminutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8
##  Median : 7406   Median : 5.245   Median :1057.5
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2
```

```
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5
##  Max.   :36019   Max.   :28.030   Max.   :1440.0
```

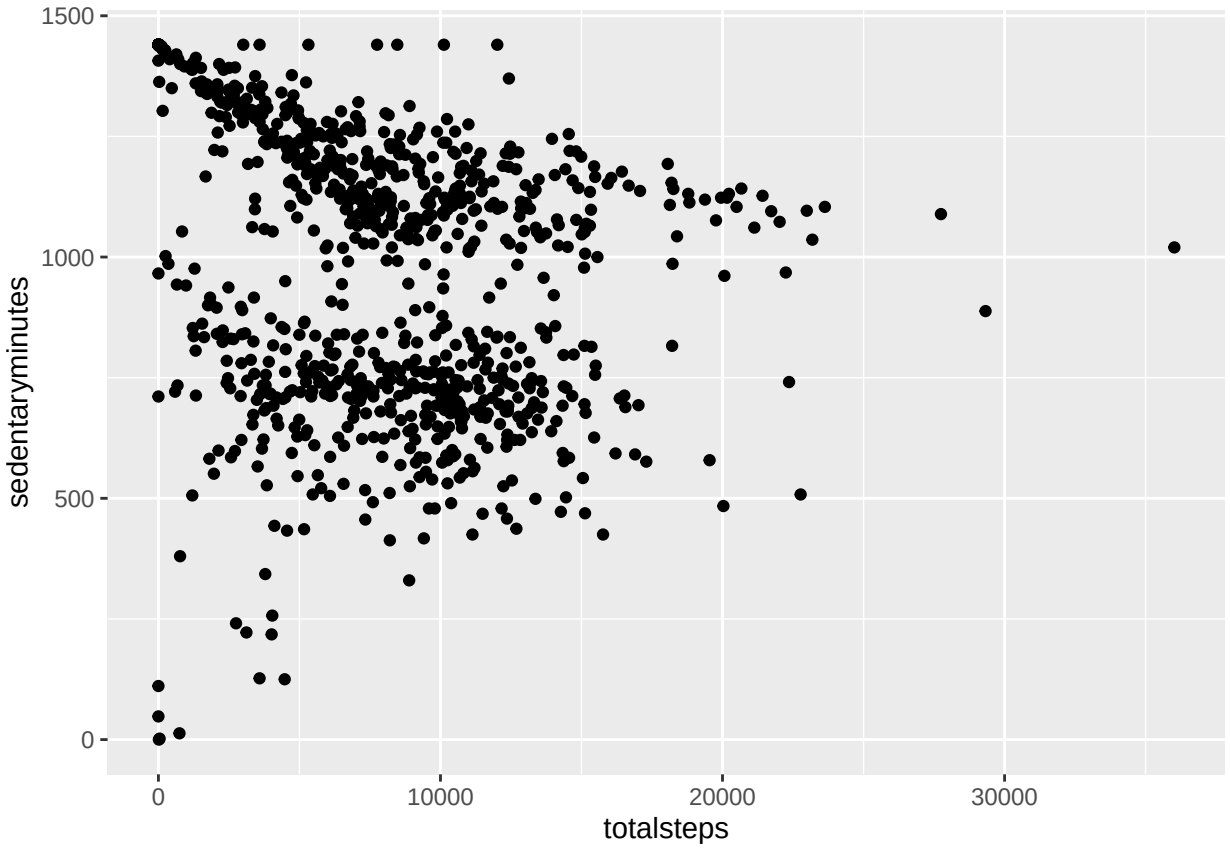
```
sleep_day %>%
  select(totalsleeprecords,
         totalminutesasleep,
         totaltimeinbed) %>%
  summary()
```

```
##  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
```

Visualizing my data

plotting a few explorations

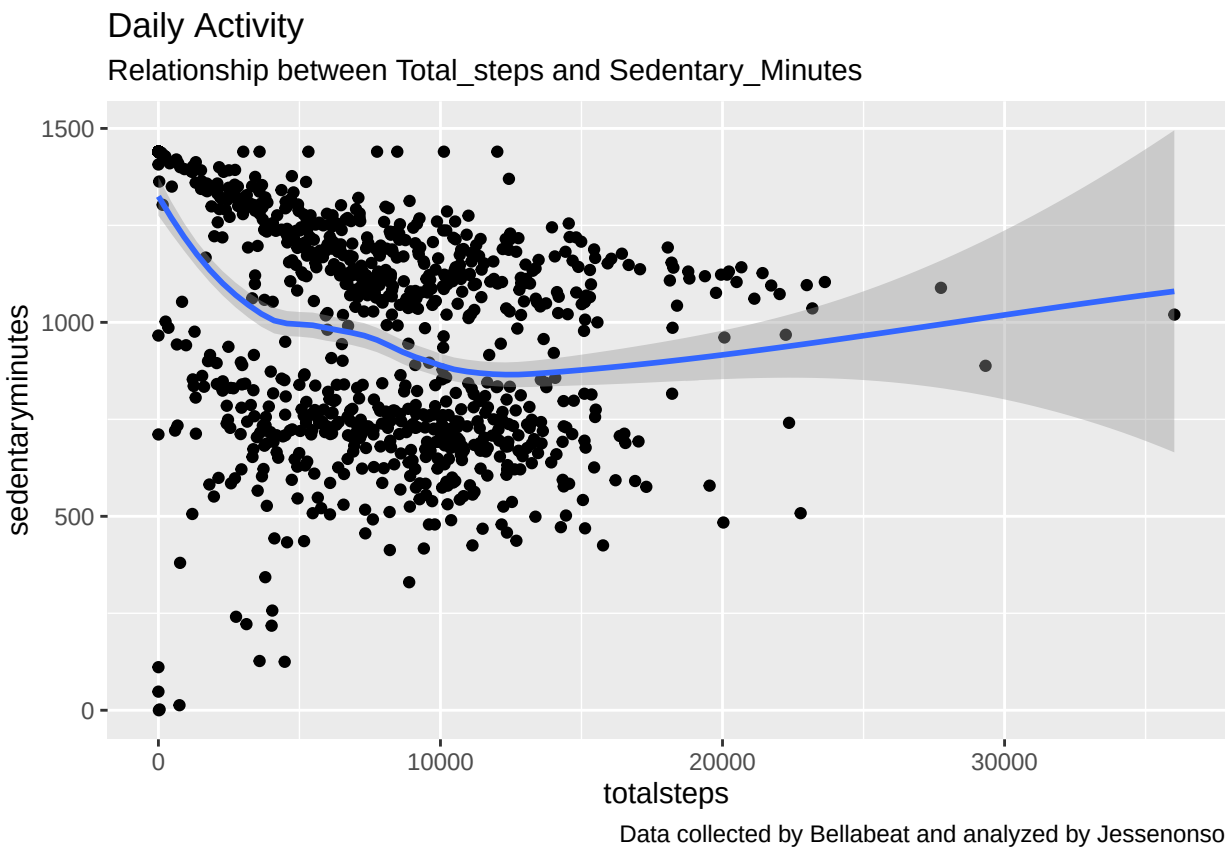
```
ggplot(data=daily_activity) +
  geom_point(mapping = aes(x=totalsteps,
                           y= sedentaryminutes))
```




```
ggplot(data=daily_activity) +
  geom_point(mapping = aes(x=totalsteps,
                           y= sedentaryminutes)) +
```

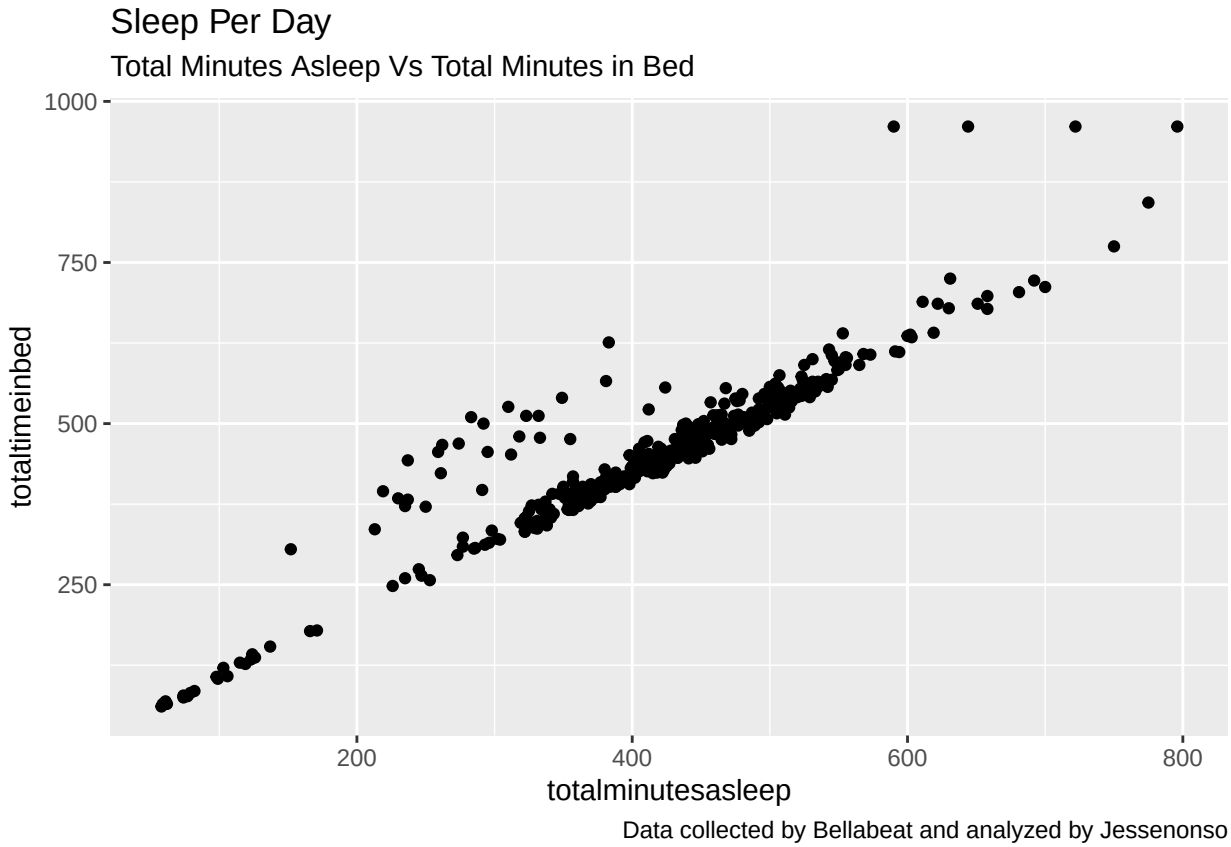
```
  geom_smooth(mapping = aes(x=totalsteps,
                            y= sedentaryminutes)) +
  labs(title = "Daily Activity",
       subtitle = "Relationship between Total_steps and Sedentary_Minutes",
       caption = "Data collected by Bellabeat and analyzed by Jessenonso")
```

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



Plotting the relationship between minutes in bed and time asleep.

```
ggplot(data=sleep_day) +
  geom_point(mapping = aes(x= totalminutesasleep,
                           totaltimeinbed)) +
  labs(title = "Sleep Per Day",
       subtitle = "Total Minutes Asleep Vs Total Minutes in Bed",
       caption = "Data collected by Bellabeat and analyzed by Jessenonso")
```



Merging two datasets together

```
combined_data<- merge(daily_activity, sleep_day, by= "id")
head(combined_data)
```

```
##           id activitydate totalsteps totaldistance trackerdistance
## 1 1503960366     5/7/2016      11992          7.71            7.71
## 2 1503960366     5/7/2016      11992          7.71            7.71
## 3 1503960366     5/7/2016      11992          7.71            7.71
## 4 1503960366     5/7/2016      11992          7.71            7.71
## 5 1503960366     5/7/2016      11992          7.71            7.71
## 6 1503960366     5/7/2016      11992          7.71            7.71
##   loggedactivitiesdistance veryactivedistance moderatelyactivedistance
## 1                        0               2.46                     2.12
## 2                        0               2.46                     2.12
## 3                        0               2.46                     2.12
## 4                        0               2.46                     2.12
## 5                        0               2.46                     2.12
## 6                        0               2.46                     2.12
##   lightactivedistance sedentaryactivedistance veryactiveminutes
## 1                3.13                       0                37
## 2                3.13                       0                37
## 3                3.13                       0                37
## 4                3.13                       0                37
## 5                3.13                       0                37
## 6                3.13                       0                37
##   fairlyactiveminutes lightlyactiveminutes sedentaryminutes calories
## 1                  46                  175              833     1821
```

```
## 2                  46                 175              833     1821
## 3                  46                 175              833     1821
## 4                  46                 175              833     1821
## 5                  46                 175              833     1821
## 6                  46                 175              833     1821
##                 sleepday totalsleeprecords totalminutesasleep totaltimeinbed
## 1 4/12/2016 12:00:00 AM                 1                327            346
## 2 4/13/2016 12:00:00 AM                 2                384            407
## 3 4/15/2016 12:00:00 AM                 1                412            442
## 4 4/16/2016 12:00:00 AM                 2                340            367
## 5 4/17/2016 12:00:00 AM                 1                700            712
## 6 4/19/2016 12:00:00 AM                 1                304            320
```

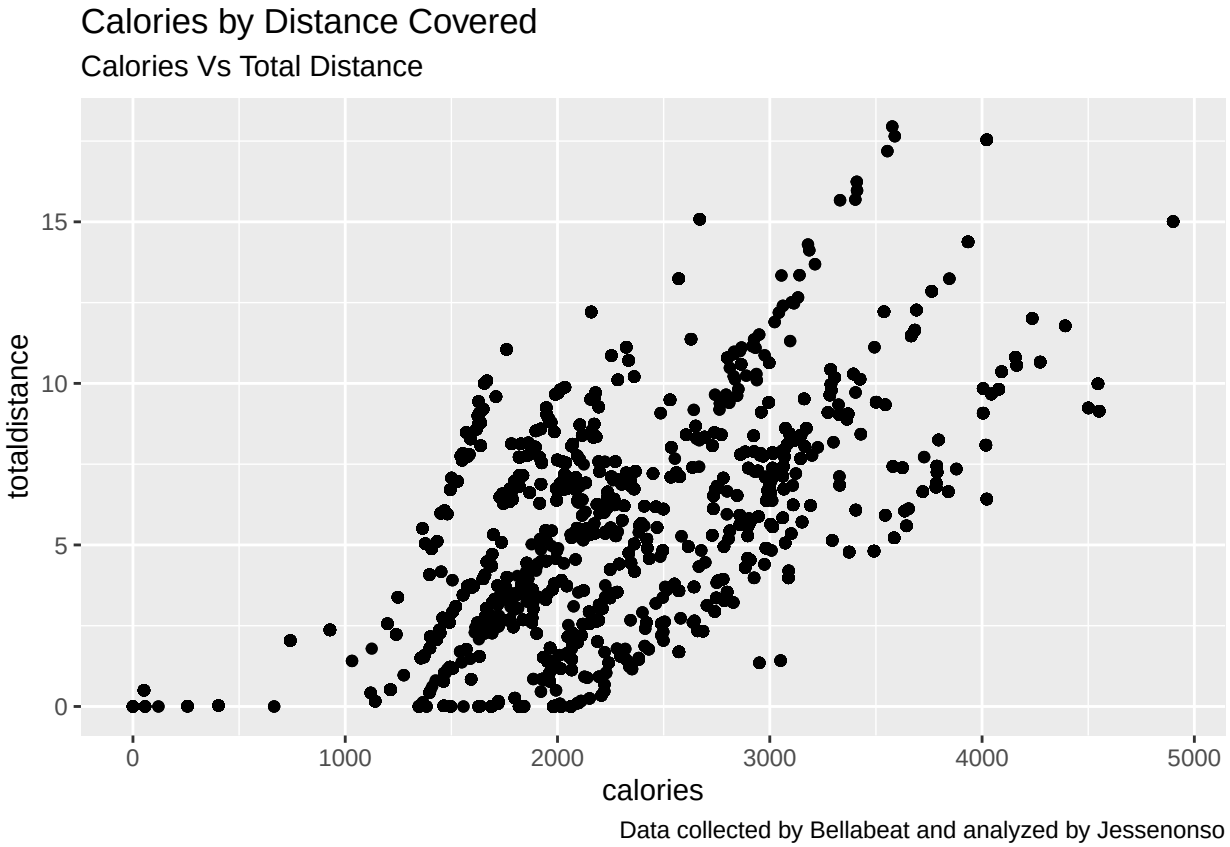
Checking how many participants are in this data set

```
n_distinct(combined_data$id)
```

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

Visualizing the combined dataset to gain more insights.

```
ggplot(data=combined_data) +
  geom_point(mapping = aes(x= calories,
                           y= totaldistance)) +
  labs(title = "Calories by Distance Covered",
       subtitle = "Calories Vs Total Distance",
       caption = "Data collected by Bellabeat and analyzed by Jessenonso")
```



Conclusions and Recommendation

1. The plot between total time in bed and the total minutes asleep shows that the higher the time spent in the bed, the higher the total minutes asleep. People who wish to get more sleep time are encouraged to spend more time in bed.
2. The plot between calories and total distance shows that the higher the total distance covered, the higher the calories burnt. In other words, people who are interested in burning more calories are encouraged to take more steps to cover more distance.
3. There is no strong correlation between total-steps and sedentary-minutes. This means that the total steps covered in the day slightly relates sedentary time. Fewer total-steps leads to more sedentary steps. You're encouraged to take note of your steps taken and your sedentary minutes.

Thank you.