Bellabeat Case Study

Tricia Smith

2023-07-06

Bellabeat Analysis Report

Introduction

Bellabeat is a trendy wellness brand for women. They manufacture wearable devices and other products that are capable of monitoring biometric and lifestyle data to help women get acquainted with their bodies and make health conscious choices.

Case Study Questions

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?

Buisness Task

Identify trends to help Bellabeat to grow and recommend ideas for Bellabeat's marketing strategy improvements based on the trends detected in smart device usage.

The Data Used For This Analysis

I used the FitBit Fitness Tracker dataset that was made available on Kaggle, to see what I used click here.

Install packages and load librarys

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

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

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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 error
## Loading required package: gsubfn
##
## Loading required package: proto
```

```
## Warning in fun(libname, pkgname): couldn't connect to display ":0"
```

```
## Loading required package: RSQLite
```

Create Data Frames From Imported Data

```
## Rows: 21645 Columns: 62
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityHour
## dbl (61): Id, Intensity00, Intensity01, Intensity02, Intensity03, Intensity0...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 188521 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): date
## dbl (3): Id, value, logId
##
## i Use `spec()` to retrieve the full column specification for this data.
```

```
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 21645 Columns: 62
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityHour
## dbl (61): Id, Steps00, Steps01, Steps02, Steps03, Steps04, Steps05, Steps06,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 413 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 67 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId
## lgl (1): IsManualReport
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, Ve...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
```

```
## dbl (2): Id, StepTotal
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 2483658 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Time
## dbl (2): Id, Value
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 22099 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 22099 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityHour
## dbl (3): Id, TotalIntensity, AverageIntensity
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 22099 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, StepTotal
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 21645 Columns: 62
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityHour
## dbl (61): Id, Calories00, Calories01, Calories02, Calories03, Calories04, Ca...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Cleaning Data Frames With Janitor

Cleaning up column names

```
## # A tibble: 21,645 x 62
##            id activity_hour      intensity00 intensity01 intensity02 intensity03
##         <dbl> <chr>                    <dbl>       <dbl>       <dbl>       <dbl>
##  1 1503960366 4/13/2016 12:00:0~           1           1           0           0
##  2 1503960366 4/13/2016 1:00:00~           0           0           0           0
##  3 1503960366 4/13/2016 2:00:00~           0           0           0           0
```

```
##  4 1503960366 4/13/2016 3:00:00~           0           0           0           0
##  5 1503960366 4/13/2016 4:00:00~           0           0           0           0
##  6 1503960366 4/13/2016 5:00:00~           0           0           0           0
##  7 1503960366 4/13/2016 6:00:00~           0           0           0           0
##  8 1503960366 4/13/2016 7:00:00~           0           0           0           0
##  9 1503960366 4/13/2016 8:00:00~           0           0           0           0
## 10 1503960366 4/13/2016 9:00:00~           0           1           1           1
## # i 21,635 more rows
## # i 56 more variables: intensity04 <dbl>, intensity05 <dbl>, intensity06 <dbl>,
## #   intensity07 <dbl>, intensity08 <dbl>, intensity09 <dbl>, intensity10 <dbl>,
## #   intensity11 <dbl>, intensity12 <dbl>, intensity13 <dbl>, intensity14 <dbl>,
## #   intensity15 <dbl>, intensity16 <dbl>, intensity17 <dbl>, intensity18 <dbl>,
## #   intensity19 <dbl>, intensity20 <dbl>, intensity21 <dbl>, intensity22 <dbl>,
## #   intensity23 <dbl>, intensity24 <dbl>, intensity25 <dbl>, ...
```

```
## # A tibble: 188,521 x 4
##            id date                 value      log_id
##         <dbl> <chr>                <dbl>       <dbl>
##  1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
##  2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
##  3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
##  4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
##  5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
##  6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
##  7 1503960366 4/12/2016 2:53:30 AM     1 11380564589
##  8 1503960366 4/12/2016 2:54:30 AM     2 11380564589
##  9 1503960366 4/12/2016 2:55:30 AM     2 11380564589
## 10 1503960366 4/12/2016 2:56:30 AM     2 11380564589
## # i 188,511 more rows
```

```
## # A tibble: 21,645 x 62
##         id activity_hour steps00 steps01 steps02 steps03 steps04 steps05 steps06
##      <dbl> <chr>           <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1  1.50e9 4/13/2016 12~       4      16       0       0       0       9       0
##  2  1.50e9 4/13/2016 1:~       0       0       0       0       0       0       0
##  3  1.50e9 4/13/2016 2:~       0       0       0       0       0       0       0
##  4  1.50e9 4/13/2016 3:~       0       0       0       0       0       0       0
##  5  1.50e9 4/13/2016 4:~       0       0       0       0       0       0       0
##  6  1.50e9 4/13/2016 5:~       0       0       0       0       0       0       0
##  7  1.50e9 4/13/2016 6:~       0       0       0       0       0       0       0
##  8  1.50e9 4/13/2016 7:~       0       0       0       0       0       0       0
##  9  1.50e9 4/13/2016 8:~       0       0       0       0       0       0       0
## 10  1.50e9 4/13/2016 9:~       0      14      10      31      37      17      25
## # i 21,635 more rows
## # i 53 more variables: steps07 <dbl>, steps08 <dbl>, steps09 <dbl>,
## #   steps10 <dbl>, steps11 <dbl>, steps12 <dbl>, steps13 <dbl>, steps14 <dbl>,
## #   steps15 <dbl>, steps16 <dbl>, steps17 <dbl>, steps18 <dbl>, steps19 <dbl>,
## #   steps20 <dbl>, steps21 <dbl>, steps22 <dbl>, steps23 <dbl>, steps24 <dbl>,
## #   steps25 <dbl>, steps26 <dbl>, steps27 <dbl>, steps28 <dbl>, steps29 <dbl>,
## #   steps30 <dbl>, steps31 <dbl>, steps32 <dbl>, steps33 <dbl>, ...
```

```
## # A tibble: 413 x 5
##          id sleep_day total_sleep_records total_minutes_asleep total_time_in_bed
##       <dbl> <chr>                   <dbl>                <dbl>              <dbl>
##  1   1.50e9 4/12/201~                   1                  327                346
```

```
##  2   1.50e9 4/13/201~                  2                 384                 407
##  3   1.50e9 4/15/201~                  1                 412                 442
##  4   1.50e9 4/16/201~                  2                 340                 367
##  5   1.50e9 4/17/201~                  1                 700                 712
##  6   1.50e9 4/19/201~                  1                 304                 320
##  7   1.50e9 4/20/201~                  1                 360                 377
##  8   1.50e9 4/21/201~                  1                 325                 364
##  9   1.50e9 4/23/201~                  1                 361                 384
## 10   1.50e9 4/24/201~                  1                 430                 449
## # i 403 more rows
```

```
## # A tibble: 67 x 8
##            id date  weight_kg weight_pounds   fat   bmi is_manual_report  log_id
##         <dbl> <chr>     <dbl>         <dbl> <dbl> <dbl> <lgl>              <dbl>
##  1 1503960366 5/2/~      52.6          116.    22  22.6 TRUE             1.46e12
##  2 1503960366 5/3/~      52.6          116.    NA  22.6 TRUE             1.46e12
##  3 1927972279 4/13~     134.           294.    NA  47.5 FALSE            1.46e12
##  4 2873212765 4/21~      56.7          125.    NA  21.5 TRUE             1.46e12
##  5 2873212765 5/12~      57.3          126.    NA  21.7 TRUE             1.46e12
##  6 4319703577 4/17~      72.4          160.    25  27.5 TRUE             1.46e12
##  7 4319703577 5/4/~      72.3          159.    NA  27.4 TRUE             1.46e12
##  8 4558609924 4/18~      69.7          154.    NA  27.2 TRUE             1.46e12
##  9 4558609924 4/25~      70.3          155.    NA  27.5 TRUE             1.46e12
## 10 4558609924 5/1/~      69.9          154.    NA  27.3 TRUE             1.46e12
## # i 57 more rows
```

```
## # A tibble: 940 x 15
##            id activity_date total_steps total_distance tracker_distance
##         <dbl> <chr>               <dbl>          <dbl>            <dbl>
##  1 1503960366 4/12/2016           13162           8.5              8.5
##  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
##  7 1503960366 4/18/2016           13019           8.59             8.59
##  8 1503960366 4/19/2016           15506           9.88             9.88
##  9 1503960366 4/20/2016           10544           6.68             6.68
## 10 1503960366 4/21/2016            9819           6.34             6.34
## # i 930 more rows
## # i 10 more variables: logged_activities_distance <dbl>,
## #   very_active_distance <dbl>, moderately_active_distance <dbl>,
## #   light_active_distance <dbl>, sedentary_active_distance <dbl>,
## #   very_active_minutes <dbl>, fairly_active_minutes <dbl>,
## #   lightly_active_minutes <dbl>, sedentary_minutes <dbl>, calories <dbl>
```

```
## # A tibble: 940 x 3
##            id activity_day calories
##         <dbl> <chr>           <dbl>
##  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
```

```
##  7 1503960366 4/18/2016        1921
##  8 1503960366 4/19/2016        2035
##  9 1503960366 4/20/2016        1786
## 10 1503960366 4/21/2016        1775
## # i 930 more rows
```

```
## # A tibble: 940 x 10
##            id activity_day sedentary_minutes lightly_active_minutes
##         <dbl> <chr>                    <dbl>                  <dbl>
##  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
##  7 1503960366 4/18/2016                 1149                    233
##  8 1503960366 4/19/2016                  775                    264
##  9 1503960366 4/20/2016                  818                    205
## 10 1503960366 4/21/2016                  838                    211
## # i 930 more rows
## # i 6 more variables: fairly_active_minutes <dbl>, very_active_minutes <dbl>,
## #   sedentary_active_distance <dbl>, light_active_distance <dbl>,
## #   moderately_active_distance <dbl>, very_active_distance <dbl>
```

```
## # A tibble: 940 x 3
##            id activity_day step_total
##         <dbl> <chr>             <dbl>
##  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
##  7 1503960366 4/18/2016         13019
##  8 1503960366 4/19/2016         15506
##  9 1503960366 4/20/2016         10544
## 10 1503960366 4/21/2016          9819
## # i 930 more rows
```

```
## # A tibble: 2,483,658 x 3
##            id time                  value
##         <dbl> <chr>                 <dbl>
##  1 2022484408 4/12/2016 7:21:00 AM    97
##  2 2022484408 4/12/2016 7:21:05 AM   102
##  3 2022484408 4/12/2016 7:21:10 AM   105
##  4 2022484408 4/12/2016 7:21:20 AM   103
##  5 2022484408 4/12/2016 7:21:25 AM   101
##  6 2022484408 4/12/2016 7:22:05 AM    95
##  7 2022484408 4/12/2016 7:22:10 AM    91
##  8 2022484408 4/12/2016 7:22:15 AM    93
##  9 2022484408 4/12/2016 7:22:20 AM    94
## 10 2022484408 4/12/2016 7:22:25 AM    93
## # i 2,483,648 more rows
```

```
## # A tibble: 22,099 x 3
##            id activity_hour         calories
```

```
##         <dbl> <chr>                    <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM       81
##  2 1503960366 4/12/2016 1:00:00 AM        61
##  3 1503960366 4/12/2016 2:00:00 AM        59
##  4 1503960366 4/12/2016 3:00:00 AM        47
##  5 1503960366 4/12/2016 4:00:00 AM        48
##  6 1503960366 4/12/2016 5:00:00 AM        48
##  7 1503960366 4/12/2016 6:00:00 AM        48
##  8 1503960366 4/12/2016 7:00:00 AM        47
##  9 1503960366 4/12/2016 8:00:00 AM        68
## 10 1503960366 4/12/2016 9:00:00 AM       141
## # i 22,089 more rows
```

```
## # A tibble: 22,099 x 4
##            id activity_hour         total_intensity average_intensity
##         <dbl> <chr>                          <dbl>             <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM             20             0.333
##  2 1503960366 4/12/2016 1:00:00 AM               8             0.133
##  3 1503960366 4/12/2016 2:00:00 AM               7             0.117
##  4 1503960366 4/12/2016 3:00:00 AM               0             0
##  5 1503960366 4/12/2016 4:00:00 AM               0             0
##  6 1503960366 4/12/2016 5:00:00 AM               0             0
##  7 1503960366 4/12/2016 6:00:00 AM               0             0
##  8 1503960366 4/12/2016 7:00:00 AM               0             0
##  9 1503960366 4/12/2016 8:00:00 AM              13             0.217
## 10 1503960366 4/12/2016 9:00:00 AM              30             0.5
## # i 22,089 more rows
```

```
## # A tibble: 22,099 x 3
##            id activity_hour         step_total
##         <dbl> <chr>                     <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM       373
##  2 1503960366 4/12/2016 1:00:00 AM        160
##  3 1503960366 4/12/2016 2:00:00 AM        151
##  4 1503960366 4/12/2016 3:00:00 AM          0
##  5 1503960366 4/12/2016 4:00:00 AM          0
##  6 1503960366 4/12/2016 5:00:00 AM          0
##  7 1503960366 4/12/2016 6:00:00 AM          0
##  8 1503960366 4/12/2016 7:00:00 AM          0
##  9 1503960366 4/12/2016 8:00:00 AM        250
## 10 1503960366 4/12/2016 9:00:00 AM       1864
## # i 22,089 more rows
```

```
## # A tibble: 21,645 x 62
##          id activity_hour calories00 calories01 calories02 calories03 calories04
##       <dbl> <chr>              <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
##  1   1.50e9 4/13/2016 12~      1.89       2.20       0.944      0.944      0.944
##  2   1.50e9 4/13/2016 1:~      0.786      0.786      0.786      0.786      0.944
##  3   1.50e9 4/13/2016 2:~      0.786      0.786      0.786      0.786      0.786
##  4   1.50e9 4/13/2016 3:~      0.786      0.786      0.786      0.786      0.786
##  5   1.50e9 4/13/2016 4:~      0.786      0.786      0.786      0.786      0.786
##  6   1.50e9 4/13/2016 5:~      0.786      0.786      0.786      0.786      0.786
##  7   1.50e9 4/13/2016 6:~      0.786      0.786      0.786      0.786      0.786
##  8   1.50e9 4/13/2016 7:~      0.786      0.786      0.786      0.786      0.786
##  9   1.50e9 4/13/2016 8:~      0.944      0.786      0.786      0.786      0.786
```

```
## 10   1.50e9 4/13/2016 9:~      0.944      2.20       2.04       2.52       2.67
## # i 21,635 more rows
## # i 55 more variables: calories05 <dbl>, calories06 <dbl>, calories07 <dbl>,
## #   calories08 <dbl>, calories09 <dbl>, calories10 <dbl>, calories11 <dbl>,
## #   calories12 <dbl>, calories13 <dbl>, calories14 <dbl>, calories15 <dbl>,
## #   calories16 <dbl>, calories17 <dbl>, calories18 <dbl>, calories19 <dbl>,
## #   calories20 <dbl>, calories21 <dbl>, calories22 <dbl>, calories23 <dbl>,
## #   calories24 <dbl>, calories25 <dbl>, calories26 <dbl>, calories27 <dbl>, ...
```

Remove Empty Rows From Data Frames (value for "which" not specified, so code defaults to c("rows", "cols") effectively handling both at once to ensure no empty data fields)

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 21,645 x 62
##            Id ActivityHour       Intensity00 Intensity01 Intensity02 Intensity03
##         <dbl> <chr>                    <dbl>       <dbl>       <dbl>       <dbl>
##  1 1503960366 4/13/2016 12:00:0~           1           1           0           0
##  2 1503960366 4/13/2016 1:00:00~           0           0           0           0
##  3 1503960366 4/13/2016 2:00:00~           0           0           0           0
##  4 1503960366 4/13/2016 3:00:00~           0           0           0           0
##  5 1503960366 4/13/2016 4:00:00~           0           0           0           0
##  6 1503960366 4/13/2016 5:00:00~           0           0           0           0
##  7 1503960366 4/13/2016 6:00:00~           0           0           0           0
##  8 1503960366 4/13/2016 7:00:00~           0           0           0           0
##  9 1503960366 4/13/2016 8:00:00~           0           0           0           0
## 10 1503960366 4/13/2016 9:00:00~           0           1           1           1
## # i 21,635 more rows
## # i 56 more variables: Intensity04 <dbl>, Intensity05 <dbl>, Intensity06 <dbl>,
## #   Intensity07 <dbl>, Intensity08 <dbl>, Intensity09 <dbl>, Intensity10 <dbl>,
## #   Intensity11 <dbl>, Intensity12 <dbl>, Intensity13 <dbl>, Intensity14 <dbl>,
## #   Intensity15 <dbl>, Intensity16 <dbl>, Intensity17 <dbl>, Intensity18 <dbl>,
## #   Intensity19 <dbl>, Intensity20 <dbl>, Intensity21 <dbl>, Intensity22 <dbl>,
## #   Intensity23 <dbl>, Intensity24 <dbl>, Intensity25 <dbl>, ...
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 188,521 x 4
##            Id date                  value       logId
##         <dbl> <chr>                 <dbl>       <dbl>
##  1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
##  2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
##  3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
##  4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
##  5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
##  6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
##  7 1503960366 4/12/2016 2:53:30 AM     1 11380564589
##  8 1503960366 4/12/2016 2:54:30 AM     2 11380564589
##  9 1503960366 4/12/2016 2:55:30 AM     2 11380564589
## 10 1503960366 4/12/2016 2:56:30 AM     2 11380564589
## # i 188,511 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 21,645 x 62
##          Id ActivityHour Steps00 Steps01 Steps02 Steps03 Steps04 Steps05 Steps06
##       <dbl> <chr>          <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
```

```
##  1   1.50e9 4/13/2016 1~       4      16       0       0       0       9       0
##  2   1.50e9 4/13/2016 1~       0       0       0       0       0       0       0
##  3   1.50e9 4/13/2016 2~       0       0       0       0       0       0       0
##  4   1.50e9 4/13/2016 3~       0       0       0       0       0       0       0
##  5   1.50e9 4/13/2016 4~       0       0       0       0       0       0       0
##  6   1.50e9 4/13/2016 5~       0       0       0       0       0       0       0
##  7   1.50e9 4/13/2016 6~       0       0       0       0       0       0       0
##  8   1.50e9 4/13/2016 7~       0       0       0       0       0       0       0
##  9   1.50e9 4/13/2016 8~       0       0       0       0       0       0       0
## 10   1.50e9 4/13/2016 9~       0      14      10      31      37      17      25
## # i 21,635 more rows
## # i 53 more variables: Steps07 <dbl>, Steps08 <dbl>, Steps09 <dbl>,
## #   Steps10 <dbl>, Steps11 <dbl>, Steps12 <dbl>, Steps13 <dbl>, Steps14 <dbl>,
## #   Steps15 <dbl>, Steps16 <dbl>, Steps17 <dbl>, Steps18 <dbl>, Steps19 <dbl>,
## #   Steps20 <dbl>, Steps21 <dbl>, Steps22 <dbl>, Steps23 <dbl>, Steps24 <dbl>,
## #   Steps25 <dbl>, Steps26 <dbl>, Steps27 <dbl>, Steps28 <dbl>, Steps29 <dbl>,
## #   Steps30 <dbl>, Steps31 <dbl>, Steps32 <dbl>, Steps33 <dbl>, ...

## value for "which" not specified, defaulting to c("rows", "cols")

## # A tibble: 413 x 5
##            Id SleepDay       TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##         <dbl> <chr>                      <dbl>              <dbl>          <dbl>
##  1 1503960366 4/12/2016 12:~                1                327            346
##  2 1503960366 4/13/2016 12:~                2                384            407
##  3 1503960366 4/15/2016 12:~                1                412            442
##  4 1503960366 4/16/2016 12:~                2                340            367
##  5 1503960366 4/17/2016 12:~                1                700            712
##  6 1503960366 4/19/2016 12:~                1                304            320
##  7 1503960366 4/20/2016 12:~                1                360            377
##  8 1503960366 4/21/2016 12:~                1                325            364
##  9 1503960366 4/23/2016 12:~                1                361            384
## 10 1503960366 4/24/2016 12:~                1                430            449
## # i 403 more rows

## value for "which" not specified, defaulting to c("rows", "cols")

## # A tibble: 67 x 8
##            Id Date      WeightKg WeightPounds   Fat   BMI IsManualReport   LogId
##         <dbl> <chr>        <dbl>        <dbl> <dbl> <dbl> <lgl>            <dbl>
##  1 1503960366 5/2/2016~     52.6         116.    22  22.6 TRUE           1.46e12
##  2 1503960366 5/3/2016~     52.6         116.    NA  22.6 TRUE           1.46e12
##  3 1927972279 4/13/201~    134.          294.    NA  47.5 FALSE          1.46e12
##  4 2873212765 4/21/201~     56.7         125.    NA  21.5 TRUE           1.46e12
##  5 2873212765 5/12/201~     57.3         126.    NA  21.7 TRUE           1.46e12
##  6 4319703577 4/17/201~     72.4         160.    25  27.5 TRUE           1.46e12
##  7 4319703577 5/4/2016~     72.3         159.    NA  27.4 TRUE           1.46e12
##  8 4558609924 4/18/201~     69.7         154.    NA  27.2 TRUE           1.46e12
##  9 4558609924 4/25/201~     70.3         155.    NA  27.5 TRUE           1.46e12
## 10 4558609924 5/1/2016~     69.9         154.    NA  27.3 TRUE           1.46e12
## # i 57 more rows

## value for "which" not specified, defaulting to c("rows", "cols")

## # A tibble: 940 x 15
##            Id ActivityDate TotalSteps TotalDistance TrackerDistance
##         <dbl> <chr>             <dbl>         <dbl>           <dbl>
```

```
##  1 1503960366 4/12/2016         13162          8.5            8.5
##  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
##  7 1503960366 4/18/2016         13019          8.59           8.59
##  8 1503960366 4/19/2016         15506          9.88           9.88
##  9 1503960366 4/20/2016         10544          6.68           6.68
## 10 1503960366 4/21/2016          9819          6.34           6.34
## # i 930 more rows
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 940 x 3
##            Id ActivityDay Calories
##         <dbl> <chr>          <dbl>
##  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
##  7 1503960366 4/18/2016       1921
##  8 1503960366 4/19/2016       2035
##  9 1503960366 4/20/2016       1786
## 10 1503960366 4/21/2016       1775
## # i 930 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 940 x 10
##         Id ActivityDay SedentaryMinutes LightlyActiveMinutes FairlyActiveMinutes
##      <dbl> <chr>                  <dbl>                <dbl>                <dbl>
##  1  1.50e9 4/12/2016                728                  328                   13
##  2  1.50e9 4/13/2016                776                  217                   19
##  3  1.50e9 4/14/2016               1218                  181                   11
##  4  1.50e9 4/15/2016                726                  209                   34
##  5  1.50e9 4/16/2016                773                  221                   10
##  6  1.50e9 4/17/2016                539                  164                   20
##  7  1.50e9 4/18/2016               1149                  233                   16
##  8  1.50e9 4/19/2016                775                  264                   31
##  9  1.50e9 4/20/2016                818                  205                   12
## 10  1.50e9 4/21/2016                838                  211                    8
## # i 930 more rows
## # i 5 more variables: VeryActiveMinutes <dbl>, SedentaryActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   VeryActiveDistance <dbl>
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 940 x 3
```

```
##            Id ActivityDay StepTotal
##         <dbl> <chr>           <dbl>
##  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
##  7 1503960366 4/18/2016       13019
##  8 1503960366 4/19/2016       15506
##  9 1503960366 4/20/2016       10544
## 10 1503960366 4/21/2016        9819
## # i 930 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 2,483,658 x 3
##            Id Time                  Value
##         <dbl> <chr>                 <dbl>
##  1 2022484408 4/12/2016 7:21:00 AM    97
##  2 2022484408 4/12/2016 7:21:05 AM   102
##  3 2022484408 4/12/2016 7:21:10 AM   105
##  4 2022484408 4/12/2016 7:21:20 AM   103
##  5 2022484408 4/12/2016 7:21:25 AM   101
##  6 2022484408 4/12/2016 7:22:05 AM    95
##  7 2022484408 4/12/2016 7:22:10 AM    91
##  8 2022484408 4/12/2016 7:22:15 AM    93
##  9 2022484408 4/12/2016 7:22:20 AM    94
## 10 2022484408 4/12/2016 7:22:25 AM    93
## # i 2,483,648 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 22,099 x 3
##            Id ActivityHour          Calories
##         <dbl> <chr>                    <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM       81
##  2 1503960366 4/12/2016 1:00:00 AM        61
##  3 1503960366 4/12/2016 2:00:00 AM        59
##  4 1503960366 4/12/2016 3:00:00 AM        47
##  5 1503960366 4/12/2016 4:00:00 AM        48
##  6 1503960366 4/12/2016 5:00:00 AM        48
##  7 1503960366 4/12/2016 6:00:00 AM        48
##  8 1503960366 4/12/2016 7:00:00 AM        47
##  9 1503960366 4/12/2016 8:00:00 AM        68
## 10 1503960366 4/12/2016 9:00:00 AM       141
## # i 22,089 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 22,099 x 4
##            Id ActivityHour          TotalIntensity AverageIntensity
##         <dbl> <chr>                          <dbl>            <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM             20            0.333
##  2 1503960366 4/12/2016 1:00:00 AM               8            0.133
##  3 1503960366 4/12/2016 2:00:00 AM               7            0.117
##  4 1503960366 4/12/2016 3:00:00 AM               0            0
```

```
##  5 1503960366 4/12/2016 4:00:00 AM              0            0
##  6 1503960366 4/12/2016 5:00:00 AM              0            0
##  7 1503960366 4/12/2016 6:00:00 AM              0            0
##  8 1503960366 4/12/2016 7:00:00 AM              0            0
##  9 1503960366 4/12/2016 8:00:00 AM             13            0.217
## 10 1503960366 4/12/2016 9:00:00 AM             30            0.5
## # i 22,089 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 22,099 x 3
##            Id ActivityHour          StepTotal
##         <dbl> <chr>                     <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM       373
##  2 1503960366 4/12/2016 1:00:00 AM        160
##  3 1503960366 4/12/2016 2:00:00 AM        151
##  4 1503960366 4/12/2016 3:00:00 AM          0
##  5 1503960366 4/12/2016 4:00:00 AM          0
##  6 1503960366 4/12/2016 5:00:00 AM          0
##  7 1503960366 4/12/2016 6:00:00 AM          0
##  8 1503960366 4/12/2016 7:00:00 AM          0
##  9 1503960366 4/12/2016 8:00:00 AM        250
## 10 1503960366 4/12/2016 9:00:00 AM       1864
## # i 22,089 more rows
```

```
## value for "which" not specified, defaulting to c("rows", "cols")
```

```
## # A tibble: 21,645 x 62
##           Id ActivityHour Calories00 Calories01 Calories02 Calories03 Calories04
##        <dbl> <chr>              <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
##  1    1.50e9 4/13/2016 1~      1.89       2.20       0.944      0.944      0.944
##  2    1.50e9 4/13/2016 1~      0.786      0.786      0.786      0.786      0.944
##  3    1.50e9 4/13/2016 2~      0.786      0.786      0.786      0.786      0.786
##  4    1.50e9 4/13/2016 3~      0.786      0.786      0.786      0.786      0.786
##  5    1.50e9 4/13/2016 4~      0.786      0.786      0.786      0.786      0.786
##  6    1.50e9 4/13/2016 5~      0.786      0.786      0.786      0.786      0.786
##  7    1.50e9 4/13/2016 6~      0.786      0.786      0.786      0.786      0.786
##  8    1.50e9 4/13/2016 7~      0.786      0.786      0.786      0.786      0.786
##  9    1.50e9 4/13/2016 8~      0.944      0.786      0.786      0.786      0.786
## 10    1.50e9 4/13/2016 9~      0.944      2.20       2.04       2.52       2.67
## # i 21,635 more rows
## # i 55 more variables: Calories05 <dbl>, Calories06 <dbl>, Calories07 <dbl>,
## #   Calories08 <dbl>, Calories09 <dbl>, Calories10 <dbl>, Calories11 <dbl>,
## #   Calories12 <dbl>, Calories13 <dbl>, Calories14 <dbl>, Calories15 <dbl>,
## #   Calories16 <dbl>, Calories17 <dbl>, Calories18 <dbl>, Calories19 <dbl>,
## #   Calories20 <dbl>, Calories21 <dbl>, Calories22 <dbl>, Calories23 <dbl>,
## #   Calories24 <dbl>, Calories25 <dbl>, Calories26 <dbl>, Calories27 <dbl>, ...
```

Remove Any duplicate Rows of Data

```
## # A tibble: 21,645 x 62
##            Id ActivityHour       Intensity00 Intensity01 Intensity02 Intensity03
##         <dbl> <chr>                    <dbl>       <dbl>       <dbl>       <dbl>
##  1 1503960366 4/13/2016 12:00:0~           1           1           0           0
##  2 1503960366 4/13/2016 1:00:00~           0           0           0           0
##  3 1503960366 4/13/2016 2:00:00~           0           0           0           0
```

```
##  4 1503960366 4/13/2016 3:00:00~           0           0           0           0
##  5 1503960366 4/13/2016 4:00:00~           0           0           0           0
##  6 1503960366 4/13/2016 5:00:00~           0           0           0           0
##  7 1503960366 4/13/2016 6:00:00~           0           0           0           0
##  8 1503960366 4/13/2016 7:00:00~           0           0           0           0
##  9 1503960366 4/13/2016 8:00:00~           0           0           0           0
## 10 1503960366 4/13/2016 9:00:00~           0           1           1           1
## # i 21,635 more rows
## # i 56 more variables: Intensity04 <dbl>, Intensity05 <dbl>, Intensity06 <dbl>,
## #   Intensity07 <dbl>, Intensity08 <dbl>, Intensity09 <dbl>, Intensity10 <dbl>,
## #   Intensity11 <dbl>, Intensity12 <dbl>, Intensity13 <dbl>, Intensity14 <dbl>,
## #   Intensity15 <dbl>, Intensity16 <dbl>, Intensity17 <dbl>, Intensity18 <dbl>,
## #   Intensity19 <dbl>, Intensity20 <dbl>, Intensity21 <dbl>, Intensity22 <dbl>,
## #   Intensity23 <dbl>, Intensity24 <dbl>, Intensity25 <dbl>, ...
```

```
## # A tibble: 187,978 x 4
##           Id date                 value       logId
##        <dbl> <chr>                <dbl>       <dbl>
##  1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
##  2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
##  3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
##  4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
##  5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
##  6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
##  7 1503960366 4/12/2016 2:53:30 AM     1 11380564589
##  8 1503960366 4/12/2016 2:54:30 AM     2 11380564589
##  9 1503960366 4/12/2016 2:55:30 AM     2 11380564589
## 10 1503960366 4/12/2016 2:56:30 AM     2 11380564589
## # i 187,968 more rows
```

```
## # A tibble: 21,645 x 62
##          Id ActivityHour Steps00 Steps01 Steps02 Steps03 Steps04 Steps05 Steps06
##       <dbl> <chr>          <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1   1.50e9 4/13/2016 1~       4      16       0       0       0       9       0
##  2   1.50e9 4/13/2016 1~       0       0       0       0       0       0       0
##  3   1.50e9 4/13/2016 2~       0       0       0       0       0       0       0
##  4   1.50e9 4/13/2016 3~       0       0       0       0       0       0       0
##  5   1.50e9 4/13/2016 4~       0       0       0       0       0       0       0
##  6   1.50e9 4/13/2016 5~       0       0       0       0       0       0       0
##  7   1.50e9 4/13/2016 6~       0       0       0       0       0       0       0
##  8   1.50e9 4/13/2016 7~       0       0       0       0       0       0       0
##  9   1.50e9 4/13/2016 8~       0       0       0       0       0       0       0
## 10   1.50e9 4/13/2016 9~       0      14      10      31      37      17      25
## # i 21,635 more rows
## # i 53 more variables: Steps07 <dbl>, Steps08 <dbl>, Steps09 <dbl>,
## #   Steps10 <dbl>, Steps11 <dbl>, Steps12 <dbl>, Steps13 <dbl>, Steps14 <dbl>,
## #   Steps15 <dbl>, Steps16 <dbl>, Steps17 <dbl>, Steps18 <dbl>, Steps19 <dbl>,
## #   Steps20 <dbl>, Steps21 <dbl>, Steps22 <dbl>, Steps23 <dbl>, Steps24 <dbl>,
## #   Steps25 <dbl>, Steps26 <dbl>, Steps27 <dbl>, Steps28 <dbl>, Steps29 <dbl>,
## #   Steps30 <dbl>, Steps31 <dbl>, Steps32 <dbl>, Steps33 <dbl>, ...
```

```
## # A tibble: 410 x 5
##           Id SleepDay       TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##        <dbl> <chr>                      <dbl>              <dbl>          <dbl>
##  1 1503960366 4/12/2016 12:~                1                327            346
```

```
##  2 1503960366 4/13/2016 12:~                 2                384            407
##  3 1503960366 4/15/2016 12:~                 1                412            442
##  4 1503960366 4/16/2016 12:~                 2                340            367
##  5 1503960366 4/17/2016 12:~                 1                700            712
##  6 1503960366 4/19/2016 12:~                 1                304            320
##  7 1503960366 4/20/2016 12:~                 1                360            377
##  8 1503960366 4/21/2016 12:~                 1                325            364
##  9 1503960366 4/23/2016 12:~                 1                361            384
## 10 1503960366 4/24/2016 12:~                 1                430            449
## # i 400 more rows
```

```
## # A tibble: 67 x 8
##            Id Date      WeightKg WeightPounds   Fat   BMI IsManualReport   LogId
##         <dbl> <chr>        <dbl>        <dbl> <dbl> <dbl> <lgl>            <dbl>
##  1 1503960366 5/2/2016~     52.6         116.    22  22.6 TRUE           1.46e12
##  2 1503960366 5/3/2016~     52.6         116.    NA  22.6 TRUE           1.46e12
##  3 1927972279 4/13/201~    134.          294.    NA  47.5 FALSE          1.46e12
##  4 2873212765 4/21/201~     56.7         125.    NA  21.5 TRUE           1.46e12
##  5 2873212765 5/12/201~     57.3         126.    NA  21.7 TRUE           1.46e12
##  6 4319703577 4/17/201~     72.4         160.    25  27.5 TRUE           1.46e12
##  7 4319703577 5/4/2016~     72.3         159.    NA  27.4 TRUE           1.46e12
##  8 4558609924 4/18/201~     69.7         154.    NA  27.2 TRUE           1.46e12
##  9 4558609924 4/25/201~     70.3         155.    NA  27.5 TRUE           1.46e12
## 10 4558609924 5/1/2016~     69.9         154.    NA  27.3 TRUE           1.46e12
## # i 57 more rows
```

```
## # A tibble: 940 x 15
##            Id ActivityDate TotalSteps TotalDistance TrackerDistance
##         <dbl> <chr>              <dbl>         <dbl>           <dbl>
##  1 1503960366 4/12/2016         13162          8.5             8.5
##  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
##  7 1503960366 4/18/2016         13019          8.59            8.59
##  8 1503960366 4/19/2016         15506          9.88            9.88
##  9 1503960366 4/20/2016         10544          6.68            6.68
## 10 1503960366 4/21/2016          9819          6.34            6.34
## # i 930 more rows
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

```
## # A tibble: 940 x 3
##            Id ActivityDay Calories
##         <dbl> <chr>          <dbl>
##  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
```

```
##  7 1503960366 4/18/2016       1921
##  8 1503960366 4/19/2016       2035
##  9 1503960366 4/20/2016       1786
## 10 1503960366 4/21/2016       1775
## # i 930 more rows
```

```
## # A tibble: 940 x 10
##         Id ActivityDay SedentaryMinutes LightlyActiveMinutes FairlyActiveMinutes
##      <dbl> <chr>                  <dbl>                <dbl>                <dbl>
##  1  1.50e9 4/12/2016                728                  328                   13
##  2  1.50e9 4/13/2016                776                  217                   19
##  3  1.50e9 4/14/2016               1218                  181                   11
##  4  1.50e9 4/15/2016                726                  209                   34
##  5  1.50e9 4/16/2016                773                  221                   10
##  6  1.50e9 4/17/2016                539                  164                   20
##  7  1.50e9 4/18/2016               1149                  233                   16
##  8  1.50e9 4/19/2016                775                  264                   31
##  9  1.50e9 4/20/2016                818                  205                   12
## 10  1.50e9 4/21/2016                838                  211                    8
## # i 930 more rows
## # i 5 more variables: VeryActiveMinutes <dbl>, SedentaryActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   VeryActiveDistance <dbl>
```

```
## # A tibble: 940 x 3
##            Id ActivityDay StepTotal
##         <dbl> <chr>           <dbl>
##  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
##  7 1503960366 4/18/2016       13019
##  8 1503960366 4/19/2016       15506
##  9 1503960366 4/20/2016       10544
## 10 1503960366 4/21/2016        9819
## # i 930 more rows
```

```
## # A tibble: 2,483,658 x 3
##            Id Time                 Value
##         <dbl> <chr>                <dbl>
##  1 2022484408 4/12/2016 7:21:00 AM    97
##  2 2022484408 4/12/2016 7:21:05 AM   102
##  3 2022484408 4/12/2016 7:21:10 AM   105
##  4 2022484408 4/12/2016 7:21:20 AM   103
##  5 2022484408 4/12/2016 7:21:25 AM   101
##  6 2022484408 4/12/2016 7:22:05 AM    95
##  7 2022484408 4/12/2016 7:22:10 AM    91
##  8 2022484408 4/12/2016 7:22:15 AM    93
##  9 2022484408 4/12/2016 7:22:20 AM    94
## 10 2022484408 4/12/2016 7:22:25 AM    93
## # i 2,483,648 more rows
```

```
## # A tibble: 22,099 x 3
##            Id ActivityHour          Calories
```

```
##         <dbl> <chr>                    <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM       81
##  2 1503960366 4/12/2016 1:00:00 AM        61
##  3 1503960366 4/12/2016 2:00:00 AM        59
##  4 1503960366 4/12/2016 3:00:00 AM        47
##  5 1503960366 4/12/2016 4:00:00 AM        48
##  6 1503960366 4/12/2016 5:00:00 AM        48
##  7 1503960366 4/12/2016 6:00:00 AM        48
##  8 1503960366 4/12/2016 7:00:00 AM        47
##  9 1503960366 4/12/2016 8:00:00 AM        68
## 10 1503960366 4/12/2016 9:00:00 AM       141
## # i 22,089 more rows
```

```
## # A tibble: 22,099 x 4
##            Id ActivityHour          TotalIntensity AverageIntensity
##         <dbl> <chr>                          <dbl>            <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM             20            0.333
##  2 1503960366 4/12/2016 1:00:00 AM               8            0.133
##  3 1503960366 4/12/2016 2:00:00 AM               7            0.117
##  4 1503960366 4/12/2016 3:00:00 AM               0            0
##  5 1503960366 4/12/2016 4:00:00 AM               0            0
##  6 1503960366 4/12/2016 5:00:00 AM               0            0
##  7 1503960366 4/12/2016 6:00:00 AM               0            0
##  8 1503960366 4/12/2016 7:00:00 AM               0            0
##  9 1503960366 4/12/2016 8:00:00 AM              13            0.217
## 10 1503960366 4/12/2016 9:00:00 AM              30            0.5
## # i 22,089 more rows
```

```
## # A tibble: 22,099 x 3
##            Id ActivityHour          StepTotal
##         <dbl> <chr>                     <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM       373
##  2 1503960366 4/12/2016 1:00:00 AM        160
##  3 1503960366 4/12/2016 2:00:00 AM        151
##  4 1503960366 4/12/2016 3:00:00 AM          0
##  5 1503960366 4/12/2016 4:00:00 AM          0
##  6 1503960366 4/12/2016 5:00:00 AM          0
##  7 1503960366 4/12/2016 6:00:00 AM          0
##  8 1503960366 4/12/2016 7:00:00 AM          0
##  9 1503960366 4/12/2016 8:00:00 AM        250
## 10 1503960366 4/12/2016 9:00:00 AM       1864
## # i 22,089 more rows
```

```
## # A tibble: 21,645 x 62
##           Id ActivityHour Calories00 Calories01 Calories02 Calories03 Calories04
##        <dbl> <chr>             <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
##  1    1.50e9 4/13/2016 1~      1.89       2.20       0.944      0.944      0.944
##  2    1.50e9 4/13/2016 1~      0.786      0.786      0.786      0.786      0.944
##  3    1.50e9 4/13/2016 2~      0.786      0.786      0.786      0.786      0.786
##  4    1.50e9 4/13/2016 3~      0.786      0.786      0.786      0.786      0.786
##  5    1.50e9 4/13/2016 4~      0.786      0.786      0.786      0.786      0.786
##  6    1.50e9 4/13/2016 5~      0.786      0.786      0.786      0.786      0.786
##  7    1.50e9 4/13/2016 6~      0.786      0.786      0.786      0.786      0.786
##  8    1.50e9 4/13/2016 7~      0.786      0.786      0.786      0.786      0.786
##  9    1.50e9 4/13/2016 8~      0.944      0.786      0.786      0.786      0.786
```

```
## 10    1.50e9 4/13/2016 9~      0.944      2.20       2.04       2.52       2.67
## # i 21,635 more rows
## # i 55 more variables: Calories05 <dbl>, Calories06 <dbl>, Calories07 <dbl>,
## #   Calories08 <dbl>, Calories09 <dbl>, Calories10 <dbl>, Calories11 <dbl>,
## #   Calories12 <dbl>, Calories13 <dbl>, Calories14 <dbl>, Calories15 <dbl>,
## #   Calories16 <dbl>, Calories17 <dbl>, Calories18 <dbl>, Calories19 <dbl>,
## #   Calories20 <dbl>, Calories21 <dbl>, Calories22 <dbl>, Calories23 <dbl>,
## #   Calories24 <dbl>, Calories25 <dbl>, Calories26 <dbl>, Calories27 <dbl>, ...
```

Now that the data is clean we can begin a preliminary analysis of our data.

The first things we will look for are any trends that may be apparent in our data.

To do this we will begin to visualize some hypothesis that we may have about existing trends.

First I created a new table to be able to see how many unique user ID's there were in each data set.

Then exploring the table we can view how many user's had data reported for each of our data sets.There really are not very many users so our data definitely has some room for improvement there. In the meantime we can analyze what we do have available.

Let's take a quick look at some of our table summaries to get a broader idea of what we are looking at.

Summary Data

```
##    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
```

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

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

```
##  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
```

```
##     WeightKg           BMI
##  Min.   : 52.60   Min.   :21.45
```

```
##  1st Qu.: 61.40   1st Qu.:23.96
##  Median : 62.50   Median :24.39
##  Mean   : 72.04   Mean   :25.19
##  3rd Qu.: 85.05   3rd Qu.:25.56
##  Max.   :133.50   Max.   :47.54
```

This gives us some interesting assumptions that we can more deeply analyze in our data below.

Merging Two Datasets To Compare Sleep To Activity Levels

Prior to beginning the merge, taking a glance at my sleep_per_day and daily_activity tables one thing is evident. The activity table houses it strings in a date format while the sleep table contains date values, as such a simple merge will be ineffectual. We must first transform our data types to match. If we do not we can not compare the date values which will result in us having no data when our query is run, as we need to be able to filter our query based on the Ids and the date of occurrence. This prevents us having multiple entries for the same user on the same days.

To begin we change the string data to a Date data type in the daily_activity table.

```
daily_activity$ActivityDate=as.Date(daily_activity$ActivityDate,format="%m/%d/%Y")
```

Since the date column of our sleep_per_day is formatted as if it was date time, but is actually saved as a date we must change it to the correct format so we can compare it.

```
sleep_per_day$SleepDay = as.Date(sleep_per_day$SleepDay, format = "%m/%d/%Y")
```

Now that the date has been created we can run our merge code.

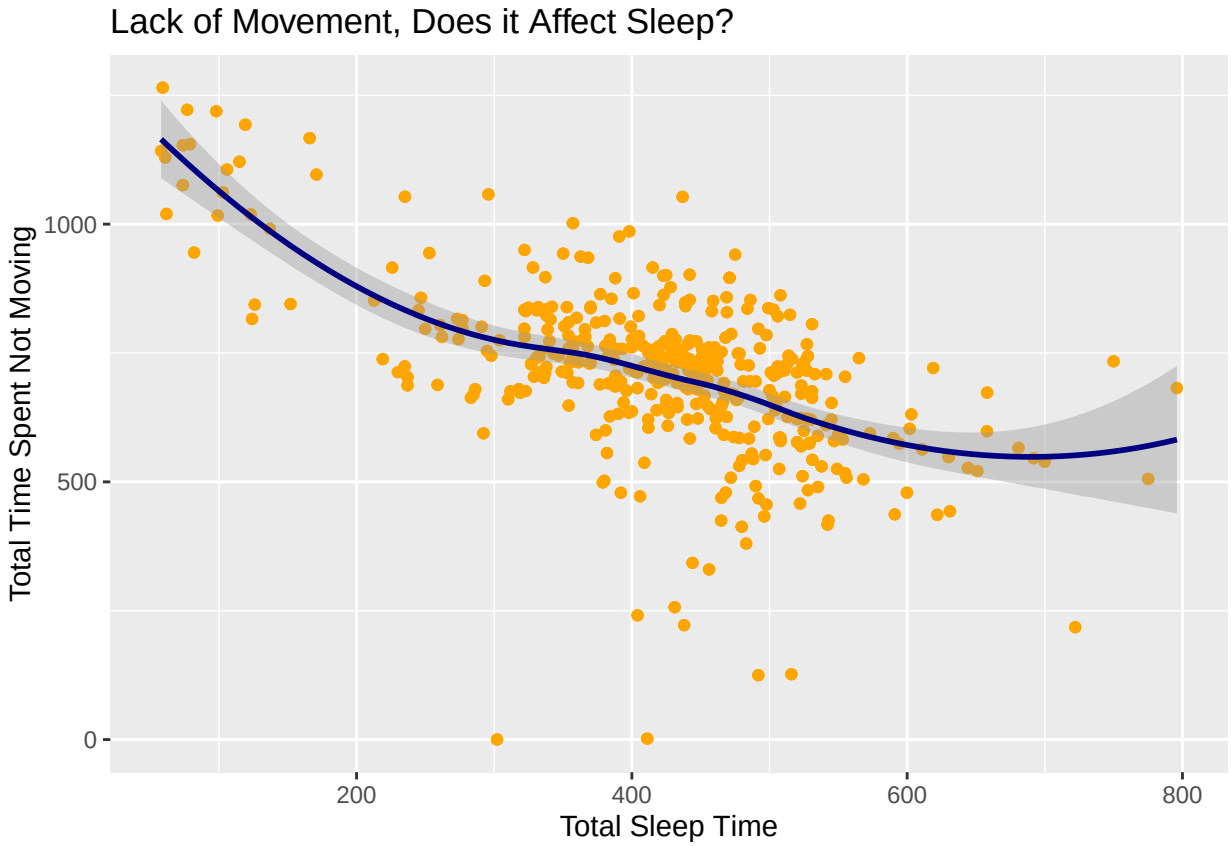
```
sleep_vs_activity = sqldf("
                          SELECT *
                          FROM daily_activity activity JOIN sleep_per_day sleep
                          ON activity.Id == sleep.Id AND activity.ActivityDate == sleep.SleepDay")
```

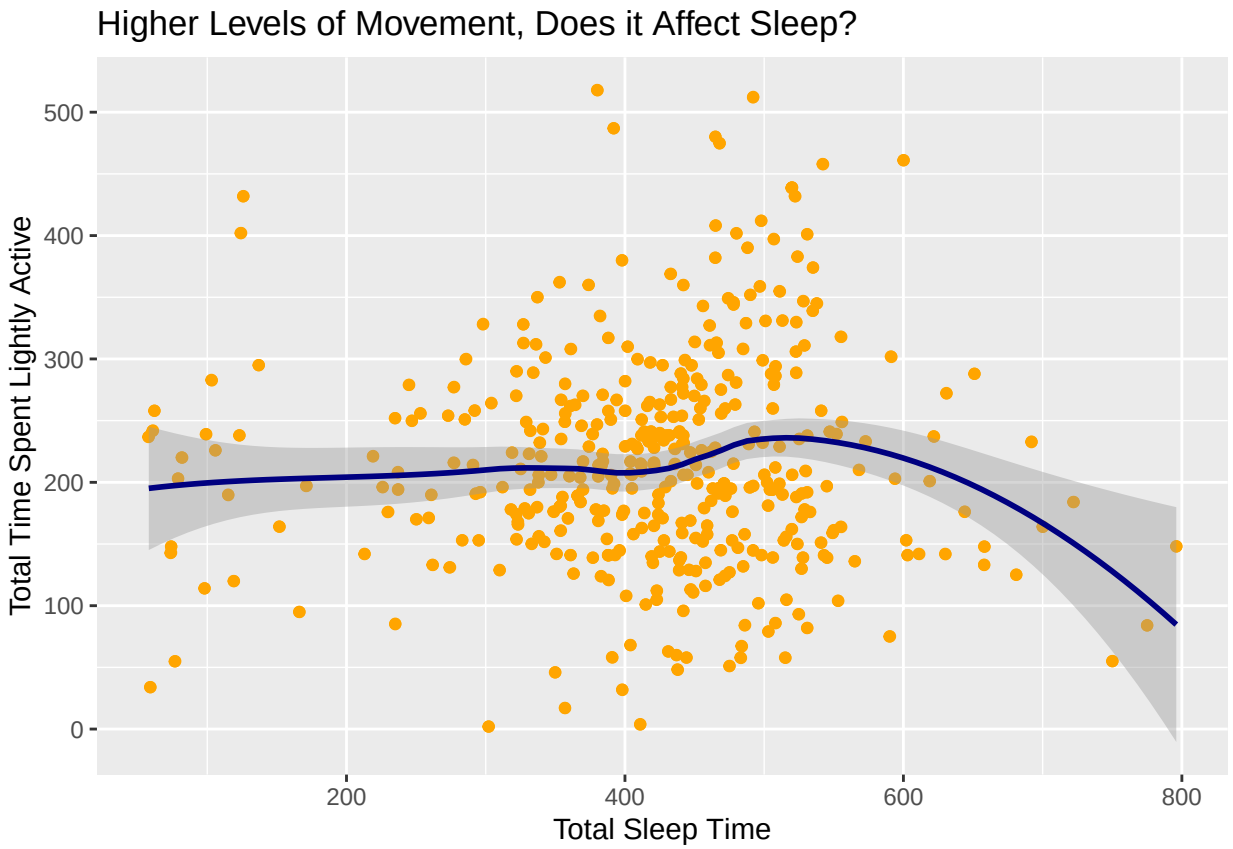
Now we have a table that very clearly has only one daily entry per user, as we should not have multiples as it renders our data inaccurate. Then when looking at our table all looks good except the 2 Id columns. To fix this without altering the data I just ran the code below.

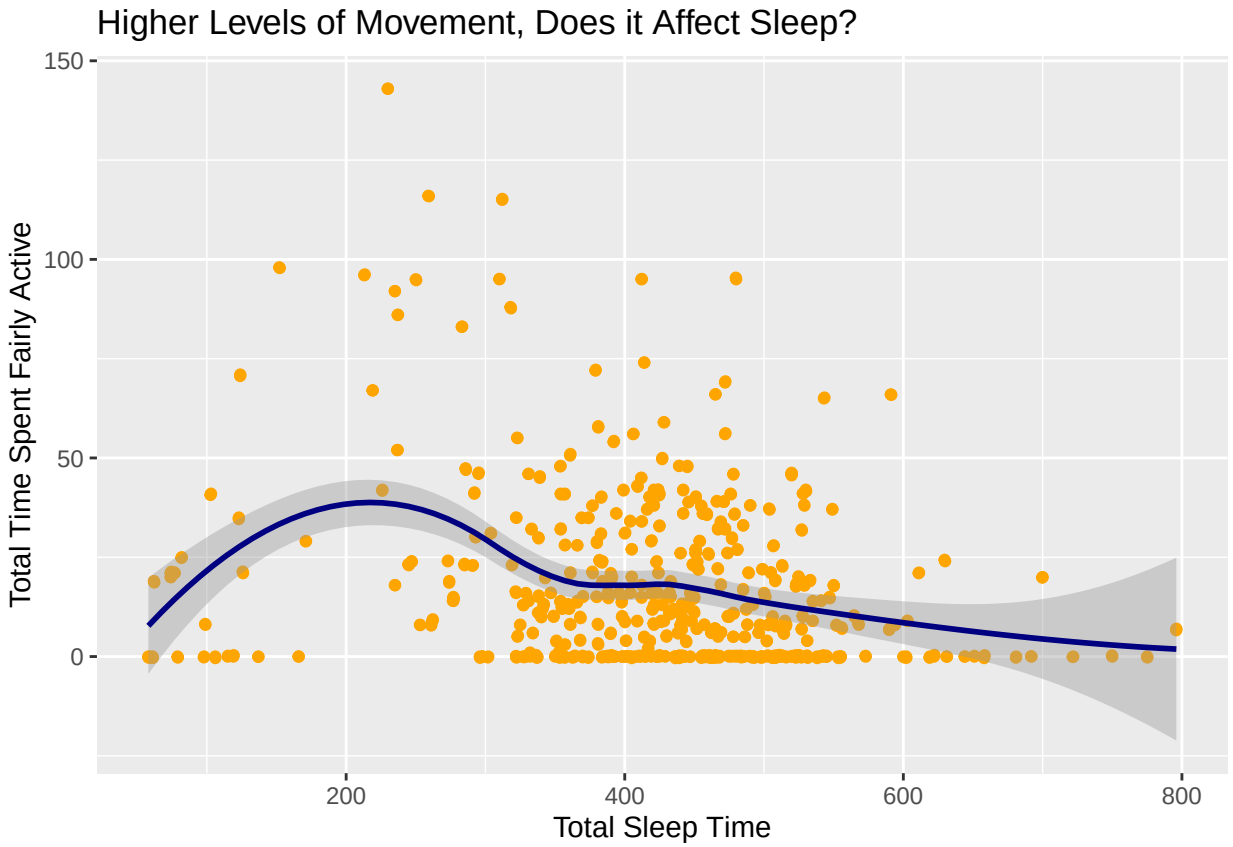
```
sleep_vs_activity = clean_names(sleep_vs_activity)
```

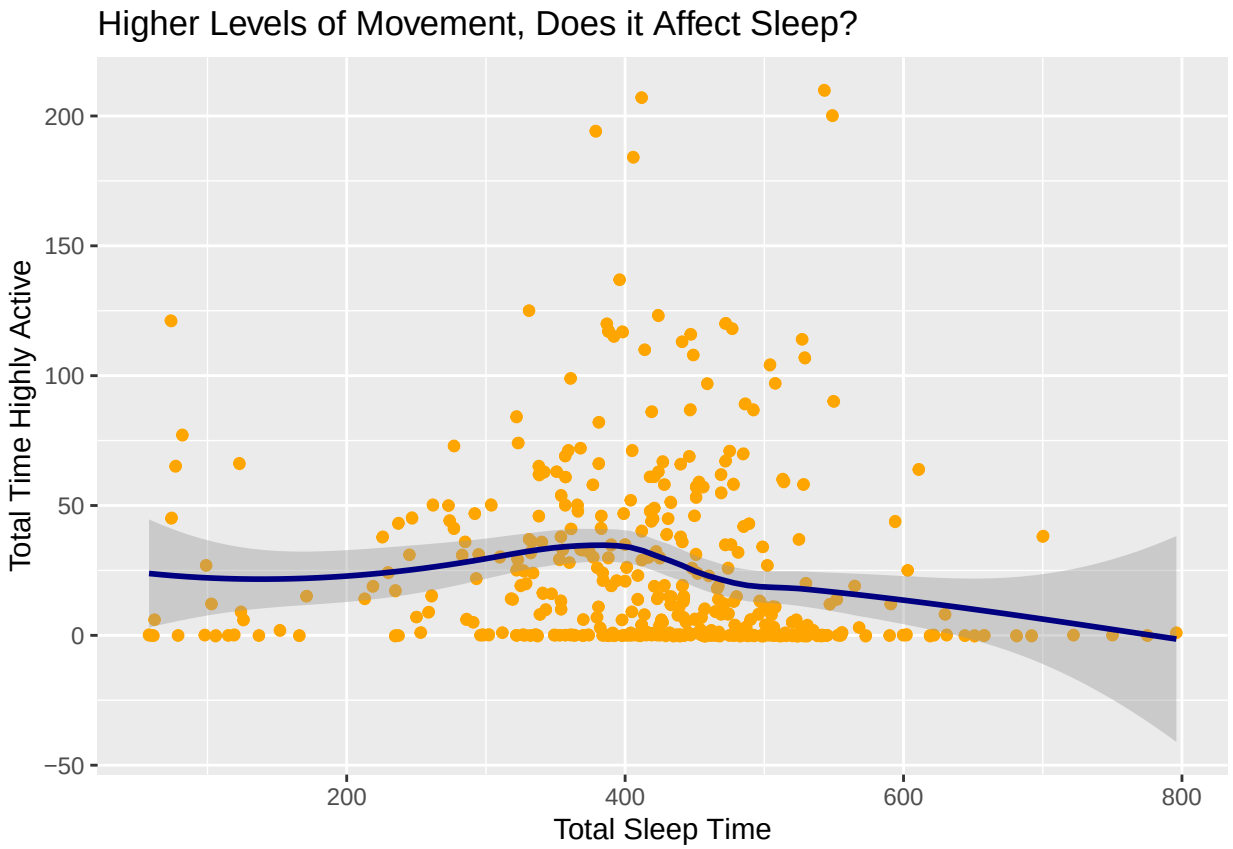
This gives all columns unique names so none of our data is duplicated.

Compare Sleep to Activity Level Visuals




Higher Levels of Movement, Does it Affect Sleep?
Total Time Spent Lightly Active
0
100
200
300
400
500
200
400
600
800
Total Sleep Time


Higher Levels of Movement, Does it Affect Sleep?
Total Time Spent Fairly Active
0
50
100
150
200
400
600
800
Total Sleep Time



To dig into these charts further I ran some R code to calculate the correlation coefficient.

```
cor(sleep_vs_activity$sedentary_minutes, sleep_vs_activity$total_minutes_asleep)
```

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

```
cor(sleep_vs_activity$lightly_active_minutes, sleep_vs_activity$total_minutes_asleep)
```

```
## [1] 0.03291407
```

```
cor(sleep_vs_activity$fairly_active_minutes, sleep_vs_activity$total_minutes_asleep)
```

```
## [1] -0.2445346
```

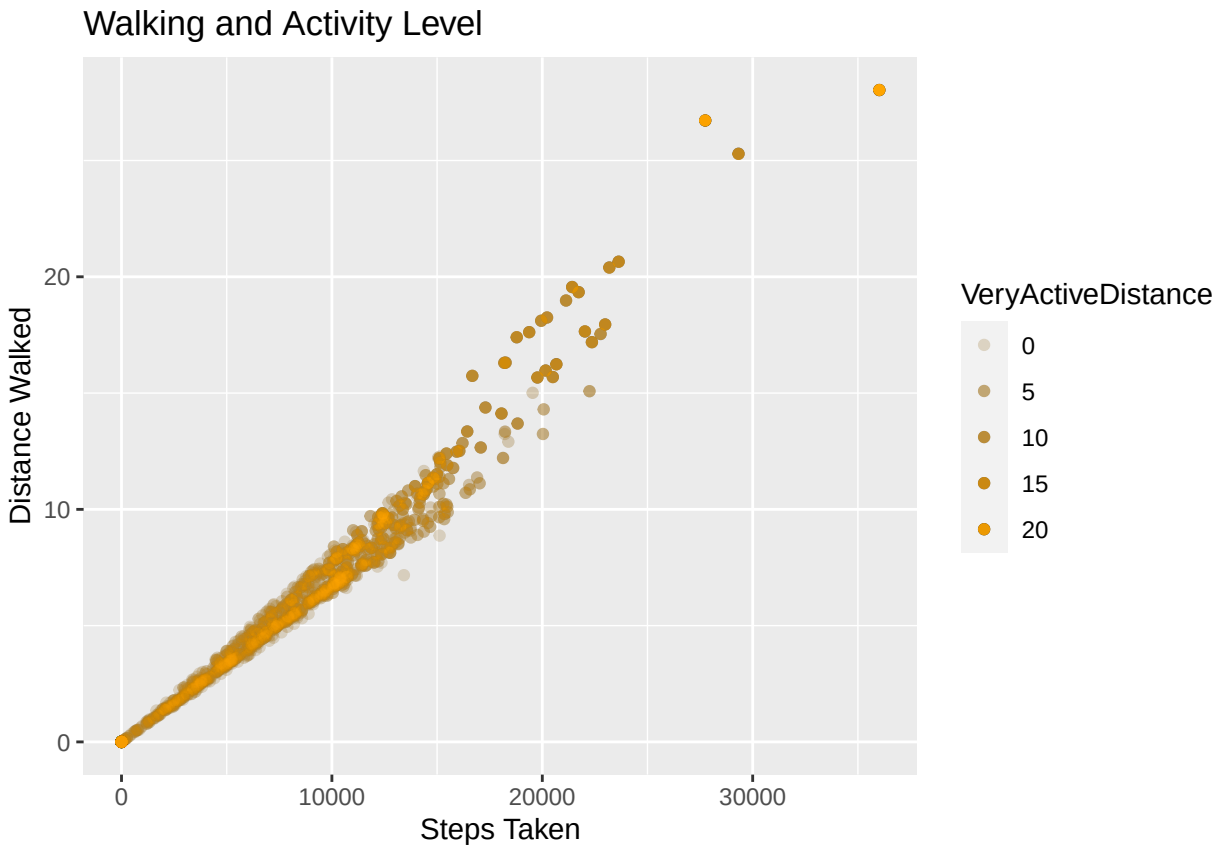
```
cor(sleep_vs_activity$very_active_minutes, sleep_vs_activity$total_minutes_asleep)
```

```
## [1] -0.09043628
```

The verdict shows a reasonable significant correlation of lack of movement having an impact on sleep times. The correlation coefficient of our sedentary users comparative to their total time asleep came in at -0.599. This points to being more active does indeed equal more sleep. However as the users activity increases appear to get slightly more sleep. Hence getting our less active users more active may be beneficial.

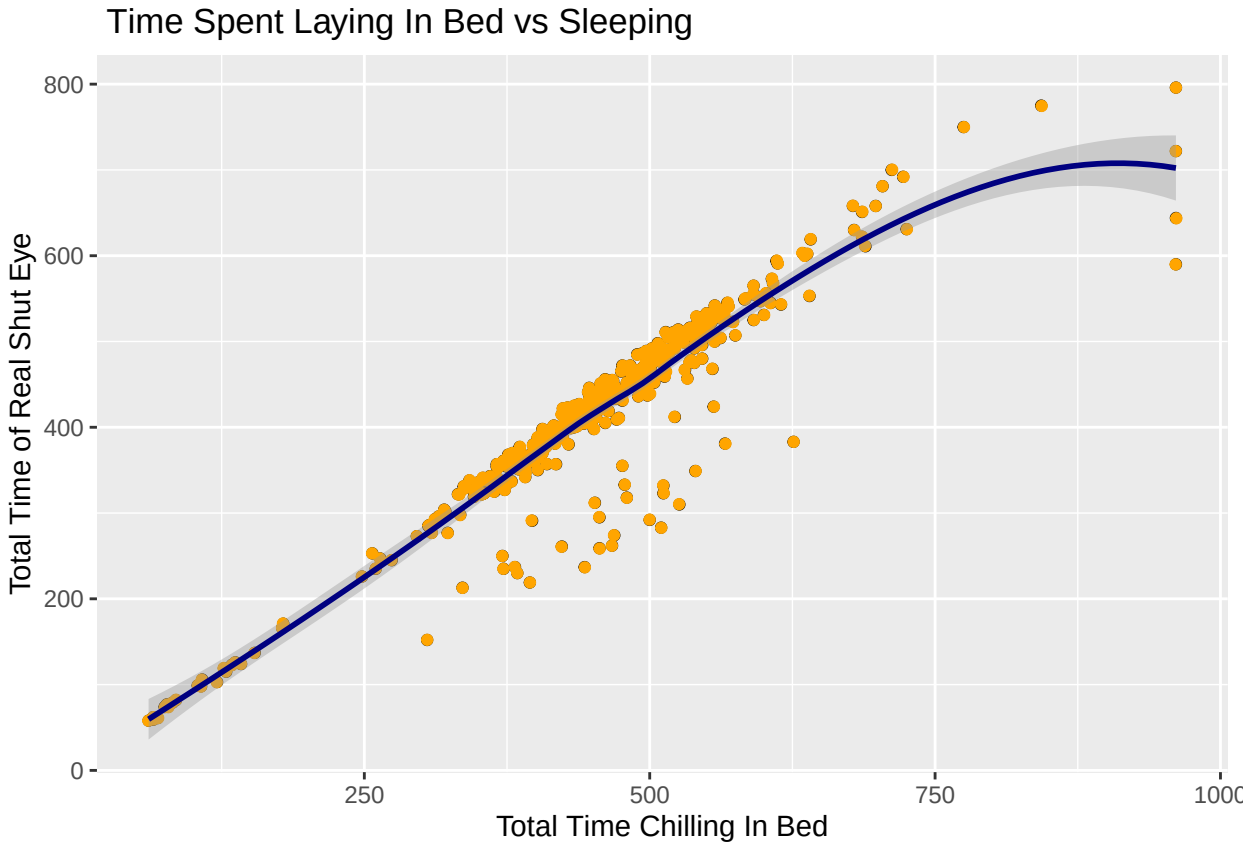
Comparing Distance and Very Active Distance

```
ggplot(data = daily_activity, aes(x = TotalSteps, y = TotalDistance, alpha = VeryActiveDistance)) +
  geom_point(fill = "orange") +
  geom_jitter(color = "orange") +
  labs(title = "Walking and Activity Level") +
  labs(x = "Steps Taken", y = "Distance Walked")
```



Taking a look at this chart we notice some interesting data. People that take more steps are not necessarily more active overall. In this chart the lighter orange points indicate users that had more Very Active Distances reported, this can mean a few things, they could be running or exercising in some other manner that is being picked up by the wearable.

Visualizing Time Spent In Bed Vs. Time Actually Asleep



Calculating the correlation coefficient

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

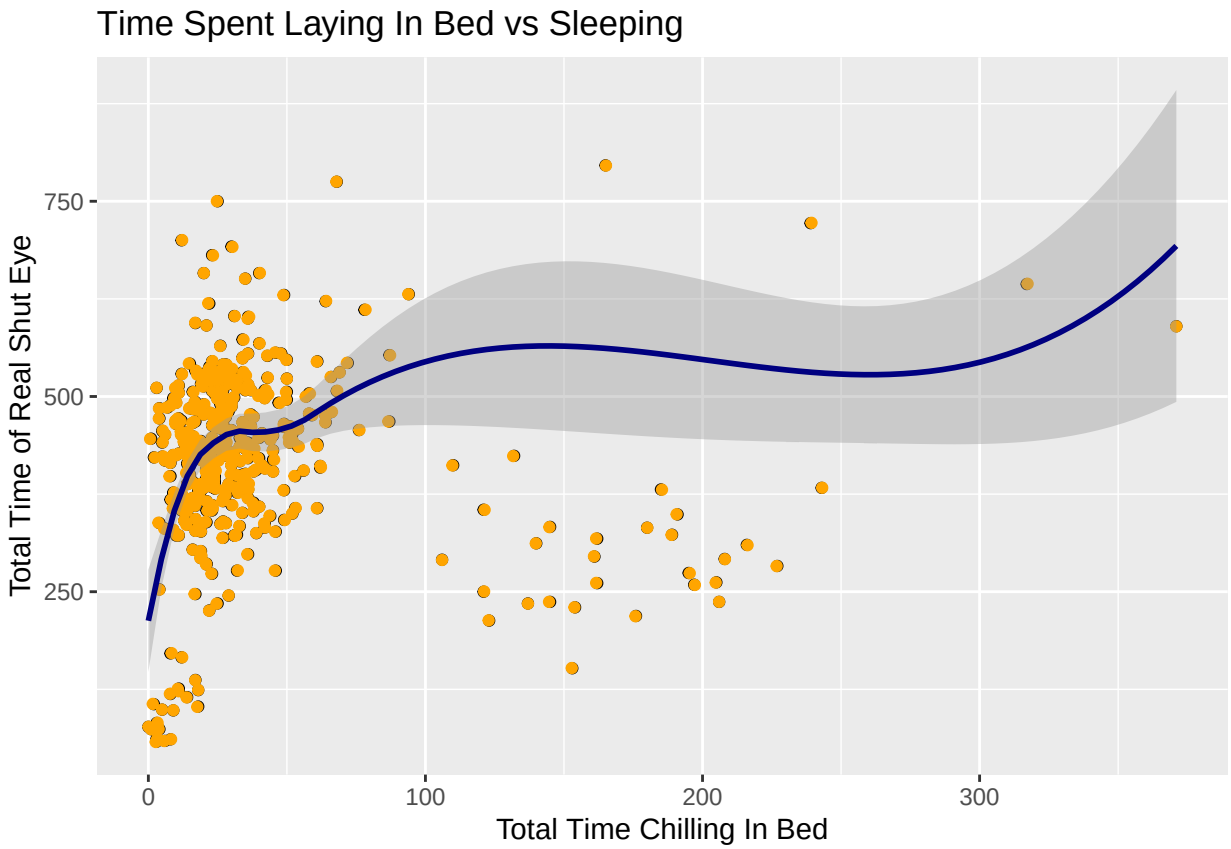
```
## [1] 0.9304575
```

Based on the above calculation, it makes perfect sense that the more you sleep the more time you spend in bed. The question now is how much time is spent in bed idly while not sleeping? We can take a further look at that.

First we will add a new column to our sleep_per_day table.

```
sleep_per_day$ChillinInBed = sleep_per_day$TotalTimeInBed - sleep_per_day$TotalMinutesAsleep
```

Then we'll add our new column into a similar visual.



Let's take a look at the coreelational coefficient for this as well.

```
cor(sleep_per_day$ChillinInBed, sleep_per_day$TotalMinutesAsleep)
```

```
## [1] -0.001761677
```

So this gives us a result of -0.001, fair to say that there is very little relationship between these 2 variables. One could ascertain that some just like to hang in there beds scrolling tik tok, or that may just be me.

Summary

After combing over the data set, I came up with some recommendations for the Bellabeat marketing team's strategy based on trends I saw in the smart device usage.

- According to our summary data the average number of steps per day is 7,638. This is lower than what is recommended. Most wearables aim for a target move goal of 10,000-12,000 steps per day. Bellabeat could add a feature to remind people to get up and move or walk around at regular intervals to help ensure the wearers are hitting their marks.
- Most participants are only lightly active. Bellabeat could try to hype up these users by sending notifications that they are approaching more active movement detected and could offer some type of incentive/positive reinforcement measures, to get these users more active.
- Bellabeat can also suggest bedtimes based on a persons sleep trands, and alert the user based on their sleeping schedules. By helping people stay on top of their sleep schedule, the data shows they will sleep more when not being sedentary.
- Bellabeat can also encourage users to a variety of different workouts and exercise plans. As our data showed, some people do not need to walk very far to get their activity levels up so we could work on encouraging other types of exercises our audience may be interested in.

Though this data set was rather limited, we did get some good insights from it. As more information is collected further trends could be explored. For now though, the Bellabeat team can use this information and hopefully drive a good marketing strategy to help drive the company's growth.