

Bellabeat Case Study

Brazil

2022-04-07

Install packages

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(plyr)
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## -----
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
```

```
## The following object is masked from 'package:purrr':
##
##   compact
```

```
library(dplyr)
library(tinytex)
```

Importing Dataset

```
activities <- read.csv("dailyActivity_merged.csv")
intensities <- read.csv("dailyIntensities_merged.csv")
calories <- read.csv("dailyCalories_merged.csv")
daily_steps <- read.csv("dailySteps_merged.csv")
weight <- read.csv("weightLogInfo_merged.csv")
sleep <- read.csv("sleepDay_merged.csv")
glimpse(activities)
```

```
## Rows: 940
## Columns: 15
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ TotalSteps <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019~
## $ TotalDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ TrackerDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ FairlyActiveMinutes <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ SedentaryMinutes <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ Calories <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~
```

```
glimpse(intensities)
```

```
## Rows: 940
## Columns: 10
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ SedentaryMinutes <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ FairlyActiveMinutes <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ VeryActiveMinutes <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
```

```
glimpse(calories)
```

```
## Rows: 940
```

```
## Columns: 3
## $ Id      <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~
## $ Calories  <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1786, 1775~
```

```
glimpse(daily_steps)
```

```
## Rows: 940
## Columns: 3
## $ Id      <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~
## $ StepTotal  <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 15506, 1054~
```

Cleaning and Formatting Date

```
sleep$SleepDay=as.POSIXct(sleep$SleepDay, format="%m/%d/%Y", tz=Sys.timezone())
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")

activities$ActivityDate=as.POSIXct(activities$ActivityDate, format="%m/%d/%Y", tz=Sys.timezone())
activities$date <- format(activities$ActivityDate, format = "%m/%d/%y")
```

Summary

```
activities %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes,
         Calories, VeryActiveMinutes) %>%
  summary()
```

```
##      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
## Min.   :   0.00
## 1st Qu.:   0.00
## Median :   4.00
## Mean   : 21.16
## 3rd Qu.: 32.00
## Max.   :210.00
```

```
weight %>%
  select(WeightKg, BMI) %>%
  summary()
```

```
##      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
```

```
sleep %>%
  select(TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
## TotalMinutesAsleep TotalTimeInBed
## Min.   : 58.0      Min.   : 61.0
## 1st Qu.:361.0      1st Qu.:403.0
## Median :433.0      Median :463.0
## Mean   :419.5      Mean   :458.6
## 3rd Qu.:490.0      3rd Qu.:526.0
## Max.   :796.0      Max.   :961.0
```

After going through the data set I noticed the “dailyActivities_merged” contains both “dailyCalories_merge”, and “dailySteps_merged” data set. So, for this analysis, “dailyActivity_merged.csv”, and “sleepDay_merged.csv” were merged.

Merge of data set

```
activities_and_sleep <-
  merge(x=activities, y=sleep, by = c("Id", "date"))
glimpse(activities_and_sleep)
```

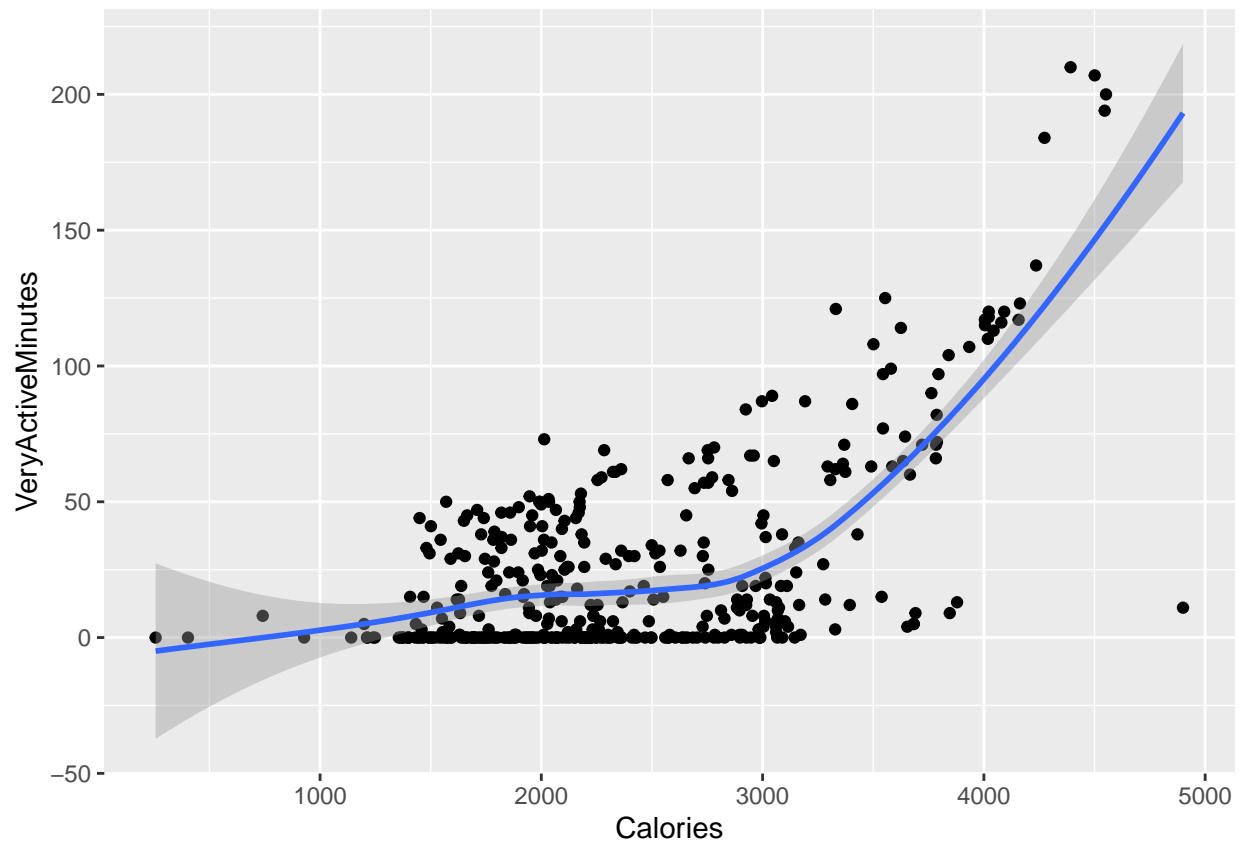
```
## Rows: 413
## Columns: 20
## $ Id                <dbl> 1503960366, 1503960366, 1503960366, 1503960366~
## $ date              <chr> "04/12/16", "04/13/16", "04/15/16", "04/16/16~
## $ ActivityDate      <dtm> 2016-04-12, 2016-04-13, 2016-04-15, 2016-04--
## $ TotalSteps        <int> 13162, 10735, 9762, 12669, 9705, 15506, 10544~
## $ TotalDistance     <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6.3~
## $ TrackerDistance   <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6.3~
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.14, 2.71, 3.19, 3.53, 1.96, 1.3~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 1.26, 0.41, 0.78, 1.32, 0.48, 0.3~
## $ LightActiveDistance <dbl> 6.06, 4.71, 2.83, 5.04, 2.51, 5.03, 4.24, 4.6~
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes  <int> 25, 21, 29, 36, 38, 50, 28, 19, 41, 39, 73, 3~
## $ FairlyActiveMinutes <int> 13, 19, 34, 10, 20, 31, 12, 8, 21, 5, 14, 23,~
## $ LightlyActiveMinutes <int> 328, 217, 209, 221, 164, 264, 205, 211, 262, ~
## $ SedentaryMinutes   <int> 728, 776, 726, 773, 539, 775, 818, 838, 732, ~
## $ Calories           <int> 1985, 1797, 1745, 1863, 1728, 2035, 1786, 177~
## $ SleepDay          <dtm> 2016-04-12, 2016-04-13, 2016-04-15, 2016-04--
## $ TotalSleepRecords  <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361, ~
## $ TotalTimeInBed     <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, ~
```

Visuals

Plots

```
ggplot(data = activities_and_sleep)+  
  geom_point(mapping = aes(x=Calories, y = VeryActiveMinutes))+  
  geom_smooth(mapping = aes(x=Calories, y= VeryActiveMinutes))
```

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

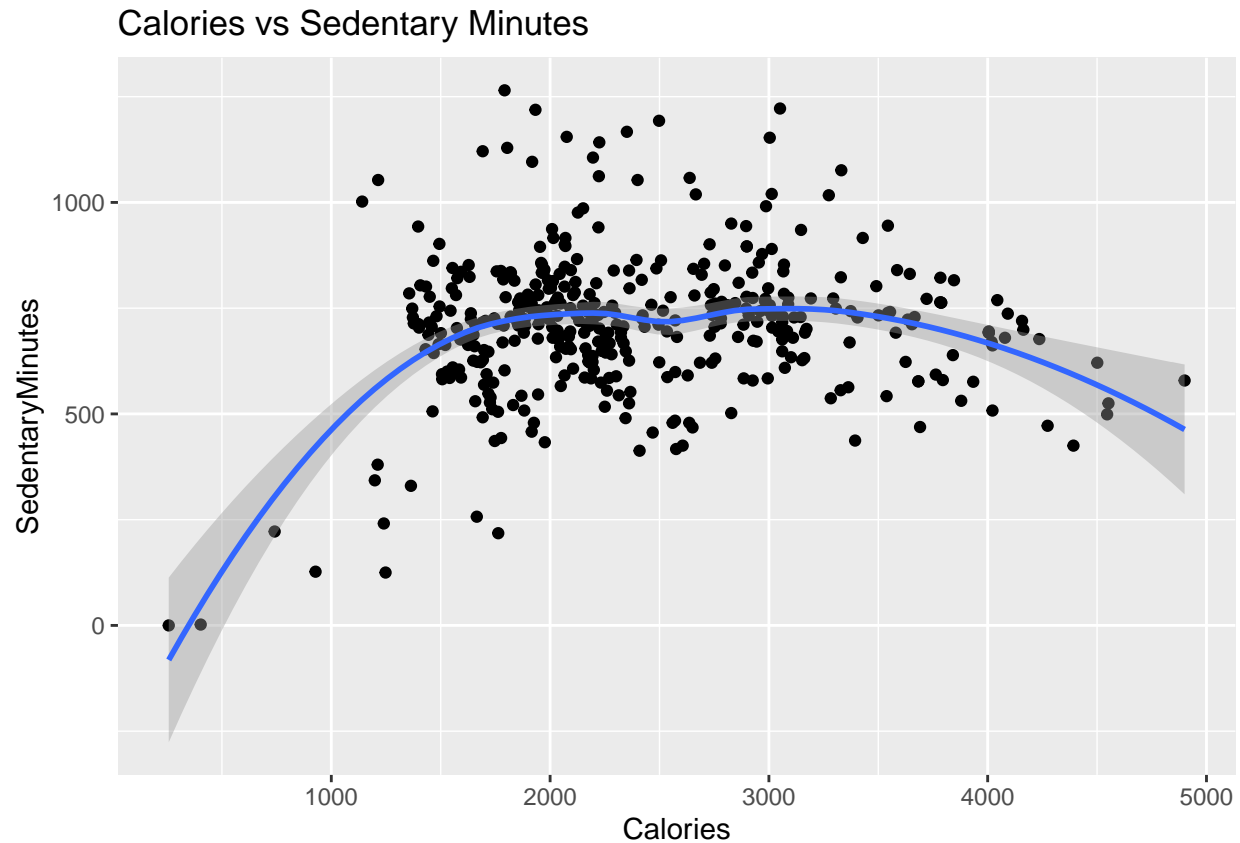


There is a positive correlation between calories and Very active minutes, so women with high calories try to be very active so as to cut down on the calorie level.

```
##Calories vs Sedentary Minutes
```

```
ggplot(data = activities_and_sleep)+  
  geom_point(mapping = aes(x=Calories, y =SedentaryMinutes))+  
  geom_smooth(mapping = aes(x=Calories, y= SedentaryMinutes)) + labs(title = "Calories vs Sedentary Min
```

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

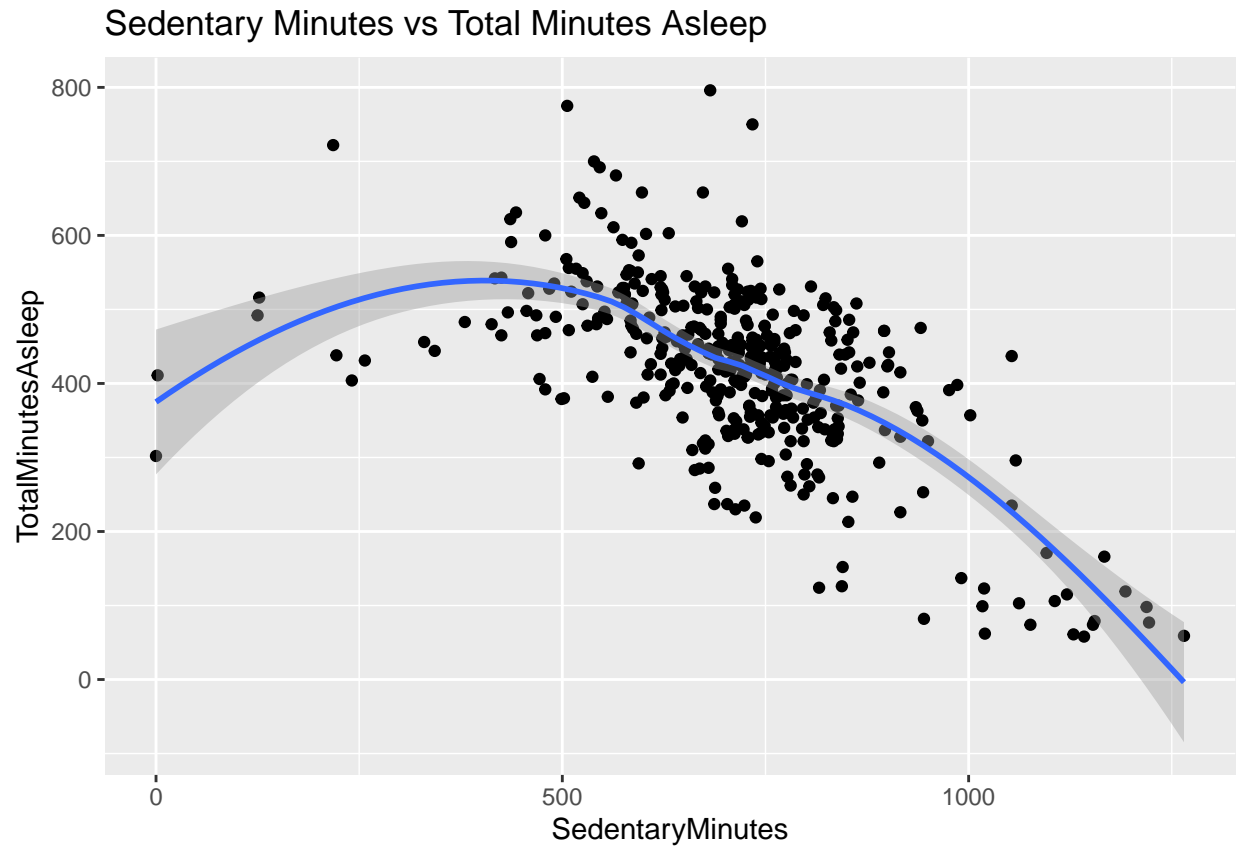


The plot shows that majority of women with sedentary minutes of ≥ 500 mins have calorie level of over 2000kcal, which is not healthy, as the standard calorie level for women is 2000kcal

Sedentary Minutes vs Total Minutes Asleep

```
ggplot(data = activities_and_sleep)+
  geom_point(mapping = aes(x=SedentaryMinutes, y = TotalMinutesAsleep))+
  geom_smooth(mapping = aes(x=SedentaryMinutes, y= TotalMinutesAsleep))+ labs(title = "Sedentary Minutes vs Total Minutes Asleep")
```

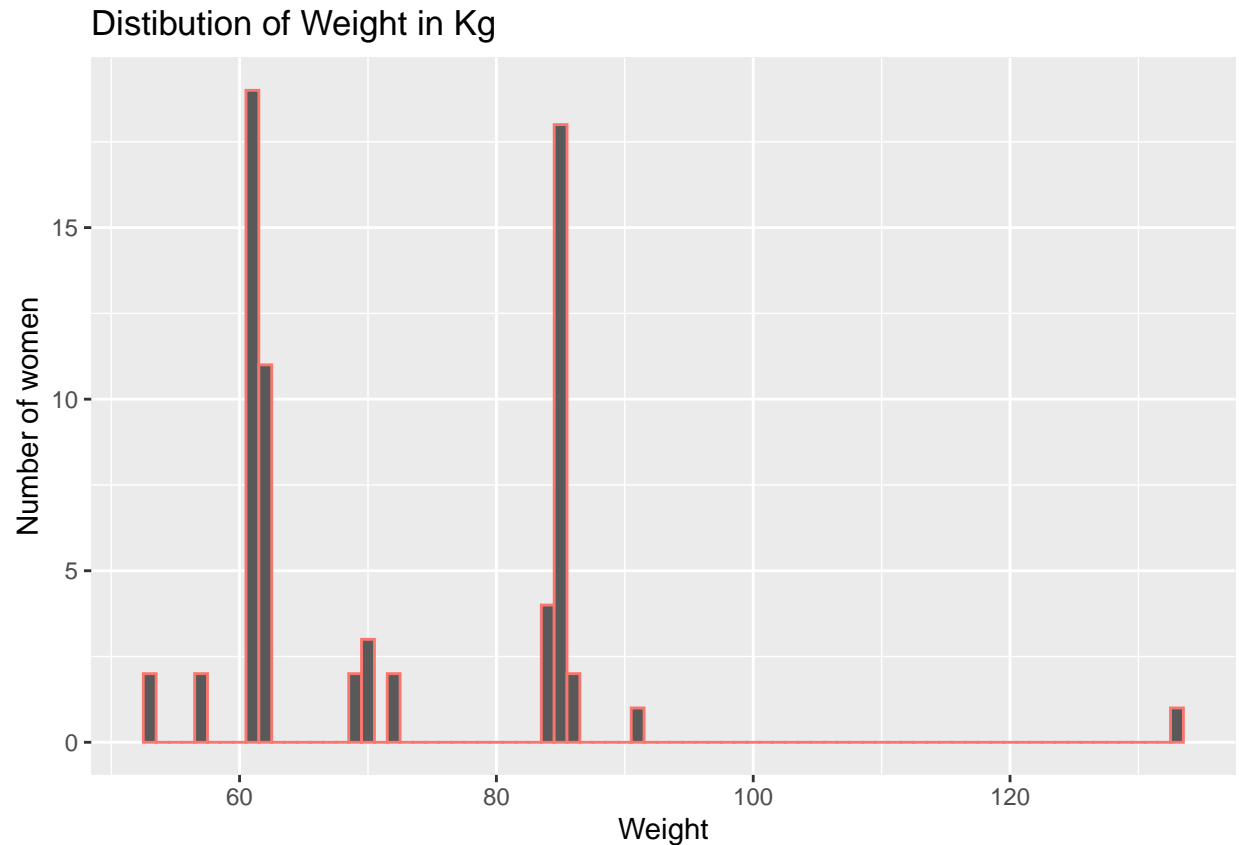
```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



There is a negative correlation between the Total Minutes Asleep and the Sedentary Minutes, seems that women who have more sedentary Minutes, spend less minutes Asleep.

Histogram of weight per Kg

```
qplot(x=WeightKg,
      data = weight, color = ('blue'),
      binwidth = 1) + theme(legend.position = "None") + labs(x="Weight", y="Number of women", title = "D")
```



The diagram above shows that majority of Bellabeat women are within the weight bracket of 60kg and 80kg respectively

Summary and Recommendations

Bellabeat was founded in 2013 by Sandro Mur and Urska Srsen, and since then driven at their goal of helping women manage their health and wellness, helping women to live in harmony with themselves. Based on these Bellabeat has created the first health tracker made specifically for women. Its with this products and services the data for this analysis were gathered.

Target Audience

All women both working full-time, part-time, and from home, irrespective of their profession or age.

Trends

An average sleep time of over 6hr is very healthy.

Most Bellabeat women have average sedentary minutes 991 (16.5h) which is above the recommended average of 551min (9.2h).

Most Bellabeat women have an average calorie of 2300kcal, this is above the normal of 2000kcal for women generally.

Recommendations

Bellabeat app could be designed to notify Bellabeat women when they have had sedentary minutes above normal, and recommend a short or light exercise or, and stretch which may aid in reducing risk factor for cardio-metabolic health. Also calorie levels above normal must be notified to Bellabeat women, and recommendations on specific food that may help balance the calorie levels and aid health and wellness in Bella women.

Thank you