

# Be-Healthy

Pooja Rathee

2022-04-08

**First we install required packages and go through them**

```
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(ggplot2)
library(scales)

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##   discard

## The following object is masked from 'package:readr':
##
##   col_factor

#Importing the required data

Activity <- read.csv("data/dailyActivity_merged.csv")
Calories<- read.csv("data/hourlyCalories_merged.csv")
Intensities <- read.csv("data/hourlyIntensities_merged.csv")
sleepDay <- read.csv("data/sleepDay_merged.csv")
weight <- read.csv("data/weightLogInfo_merged.csv")
```

**A quick look in the following columns**

```
colnames(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(Calories)
```

```
## [1] "Id" "ActivityHour" "Calories"
```

```
colnames(Intensities)
```

```
## [1] "Id" "ActivityHour" "TotalIntensity" "AverageIntensity"
```

```
colnames(sleepDay)
```

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

```
colnames(weight)
```

```
## [1] "Id" "Date" "WeightKg" "WeightPounds"
## [5] "Fat" "BMI" "IsManualReport" "LogId"
```

## Run the Head function

```
head(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

```
head(Calories)
```

##	Id	ActivityHour	Calories
## 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

```
head(Intensities)
```

##	Id	ActivityHour	TotalIntensity	AverageIntensity
## 1	1503960366	4/12/2016 12:00:00 AM	20	0.333333
## 2	1503960366	4/12/2016 1:00:00 AM	8	0.133333
## 3	1503960366	4/12/2016 2:00:00 AM	7	0.116667
## 4	1503960366	4/12/2016 3:00:00 AM	0	0.000000
## 5	1503960366	4/12/2016 4:00:00 AM	0	0.000000
## 6	1503960366	4/12/2016 5:00:00 AM	0	0.000000

```
head(sleepDay)
```

##	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep
## 1	1503960366	4/12/2016 12:00:00 AM	1	327
## 2	1503960366	4/13/2016 12:00:00 AM	2	384
## 3	1503960366	4/15/2016 12:00:00 AM	1	412
## 4	1503960366	4/16/2016 12:00:00 AM	2	340
## 5	1503960366	4/17/2016 12:00:00 AM	1	700
## 6	1503960366	4/19/2016 12:00:00 AM	1	304
##	TotalTimeInBed			
## 1	346			
## 2	407			
## 3	442			
## 4	367			
## 5	712			
## 6	320			

```
head(weight)
```

```
##           Id           Date WeightKg WeightPounds Fat   BMI
## 1 1503960366 5/2/2016 11:59:59 PM    52.6    115.9631  22 22.65
## 2 1503960366 5/3/2016 11:59:59 PM    52.6    115.9631  NA 22.65
## 3 1927972279 4/13/2016 1:08:52 AM   133.5    294.3171  NA 47.54
## 4 2873212765 4/21/2016 11:59:59 PM    56.7    125.0021  NA 21.45
## 5 2873212765 5/12/2016 11:59:59 PM    57.3    126.3249  NA 21.69
## 6 4319703577 4/17/2016 11:59:59 PM    72.4    159.6147  25 27.45
##   IsManualReport      LogId
## 1             True 1.462234e+12
## 2             True 1.462320e+12
## 3            False 1.460510e+12
## 4             True 1.461283e+12
## 5             True 1.463098e+12
## 6             True 1.460938e+12
```

Take a look at the following data and format as date and time

```
# intensities
Intensities$ActivityHour=as.POSIXct(Intensities$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.time)
Intensities$time <- format(Intensities$ActivityHour, format = "%H:%M:%S")
Intensities$date <- format(Intensities$ActivityHour, format = "%m/%d/%y")

# calories
Calories$ActivityHour=as.POSIXct(Calories$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
Calories$time <- format(Calories$ActivityHour, format = "%H:%M:%S")
Calories$date <- format(Calories$ActivityHour, format = "%m/%d/%y")

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

# sleep
sleepDay$SleepDay=as.POSIXct(sleepDay$SleepDay, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
sleepDay$date <- format(sleepDay$SleepDay, format = "%m/%d/%y")

# weight
weight$date <- format(weight$date, format = "%m/%d/%y")
weight$time <- format(weight$date, format = "%H:%M:%S")
```

Identify all the observations with respect to their Id's

```
n_distinct(Activity$Id)
```

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

```
n_distinct(Calories$Id)
```

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

```
n_distinct(Intensities$Id)
```

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

```
n_distinct(sleepDay$Id)
```

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

```
n_distinct(weight$Id)
```

```
## [1] 8
```

Take a look at the summary of the following files

```
# activity
```

```
Activity %>%  
  select(TotalSteps,  
         TotalDistance,  
         SedentaryMinutes, Calories) %>%  
  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
```

```
# explore num of active minutes per category
```

```
Activity %>%  
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%  
  summary()
```

```
##      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
```

```
Calories %>%  
  select(Calories) %>%  
  summary()
```

```
##      Calories
## Min.   : 42.00
## 1st Qu.: 63.00
## Median : 83.00
## Mean   : 97.39
## 3rd Qu.:108.00
## Max.   :948.00
```

```
# sleep
sleepDay %>%
  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
```

```
# weight
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
```

## Merging the data of sleepDay and Activity by Id and date

```
merged_data <- merge(sleepDay, Activity, FUN=c("Id", "date"))
head(merged_data)
```

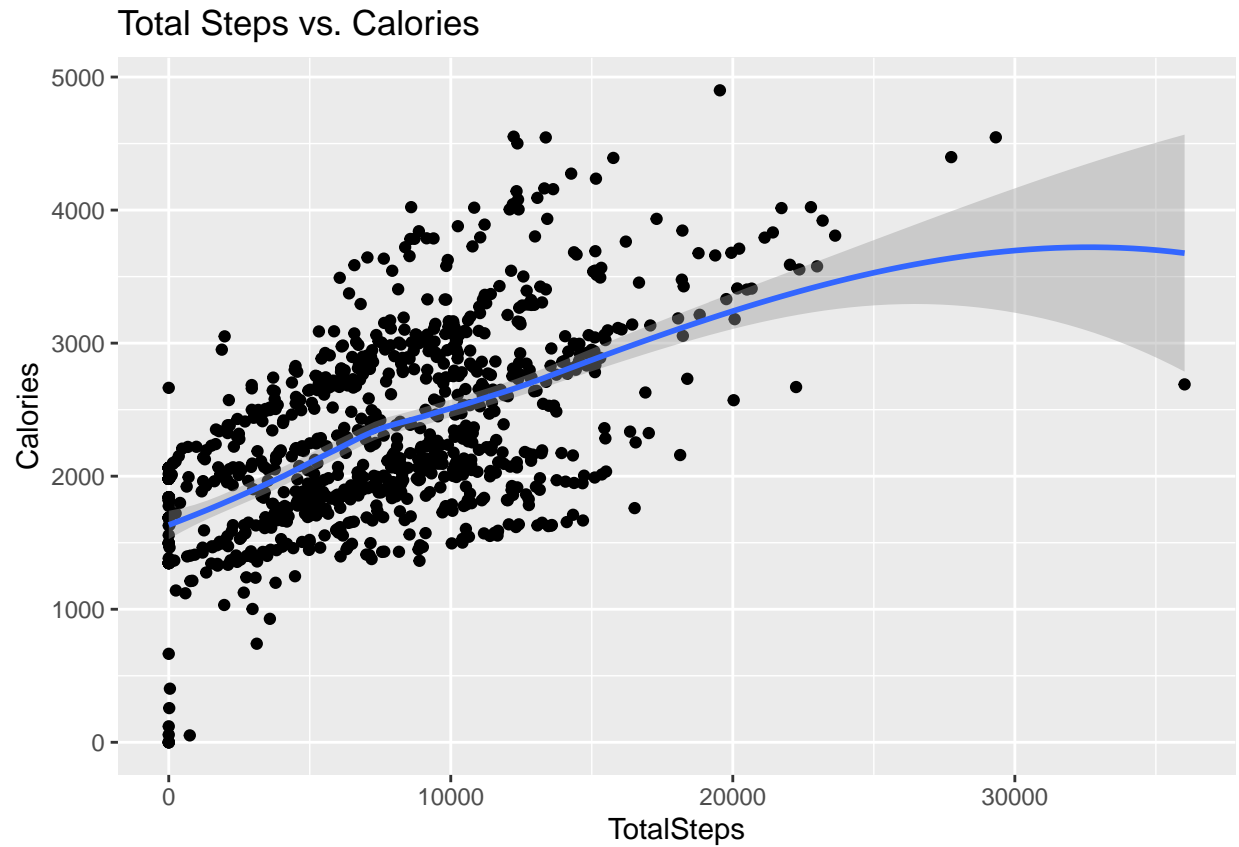
```
##      Id      date SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 04/12/16 2016-04-12             1                327
## 2 1503960366 04/13/16 2016-04-13             2                384
## 3 1503960366 04/15/16 2016-04-15             1                412
## 4 1503960366 04/16/16 2016-04-16             2                340
## 5 1503960366 04/17/16 2016-04-17             1                700
## 6 1503960366 04/19/16 2016-04-19             1                304
## TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1             346   2016-04-12      13162           8.50           8.50
## 2             407   2016-04-13      10735           6.97           6.97
```

```
## 3      442    2016-04-15      9762      6.28      6.28
## 4      367    2016-04-16     12669      8.16      8.16
## 5      712    2016-04-17      9705      6.48      6.48
## 6      320    2016-04-19     15506      9.88      9.88
##      LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1              0              1.88              0.55
## 2              0              1.57              0.69
## 3              0              2.14              1.26
## 4              0              2.71              0.41
## 5              0              3.19              0.78
## 6              0              3.53              1.32
##      LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1              6.06              0              25
## 2              4.71              0              21
## 3              2.83              0              29
## 4              5.04              0              36
## 5              2.51              0              38
## 6              5.03              0              50
##      FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1              13              328              728      1985
## 2              19              217              776      1797
## 3              34              209              726      1745
## 4              10              221              773      1863
## 5              20              164              539      1728
## 6              31              264              775      2035
```

## Understanding some summary statistics by visualization

```
ggplot(data=Activity, aes(x=TotalSteps, y=Calories)) +
  geom_point() + geom_smooth() + labs(title="Total Steps vs. Calories")
```

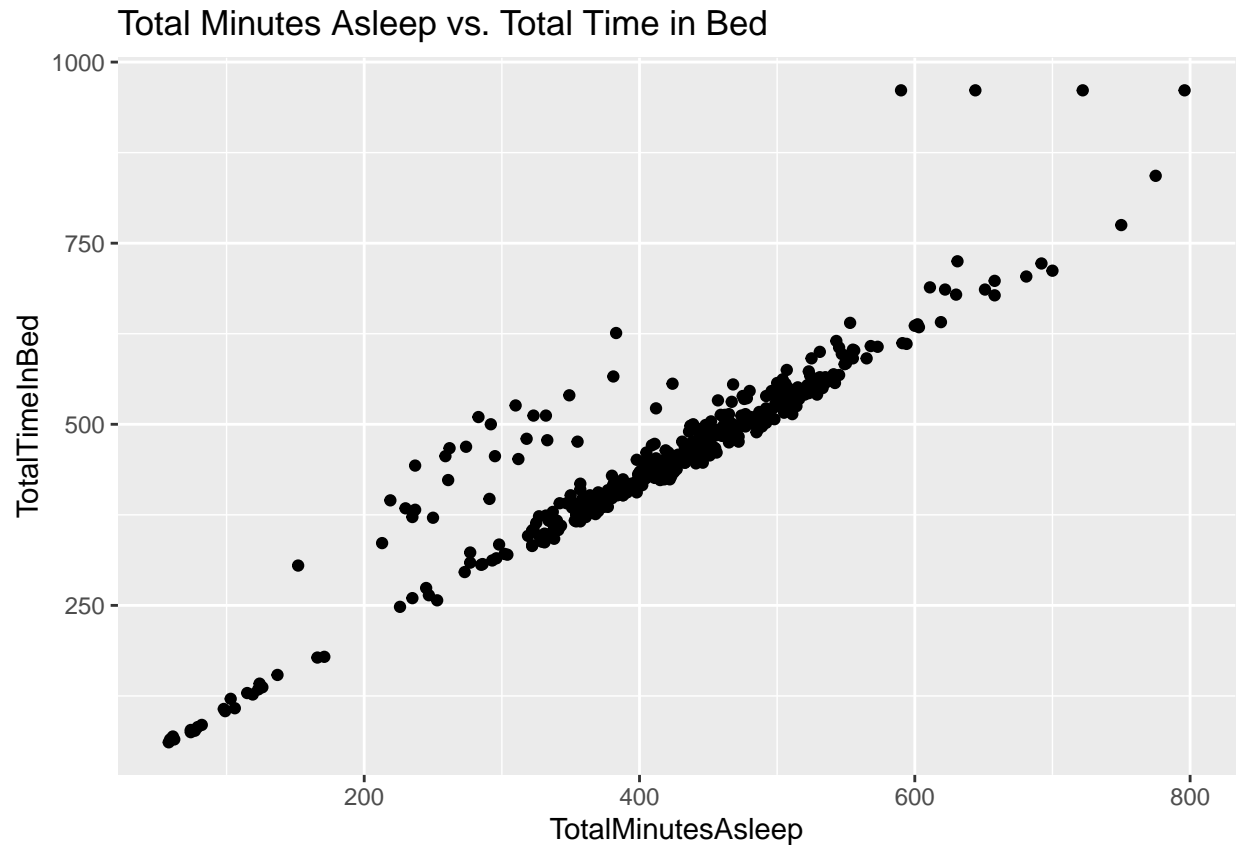
```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



# There is a positive correlation between Total Steps and Calories, which is obvious - the more active we are, the more calories we burn. # Let's Visualize the sleepDay dataframe

```
ggplot(data=sleepDay, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) +  
  geom_point()+ labs(title="Total Minutes Asleep vs. Total Time in Bed")
```





# The relationship between Total Minutes Asleep and Total Time in Bed looks linear. So if the Bellabeat users want to improve their sleep, we should consider using notification to go to sleep.

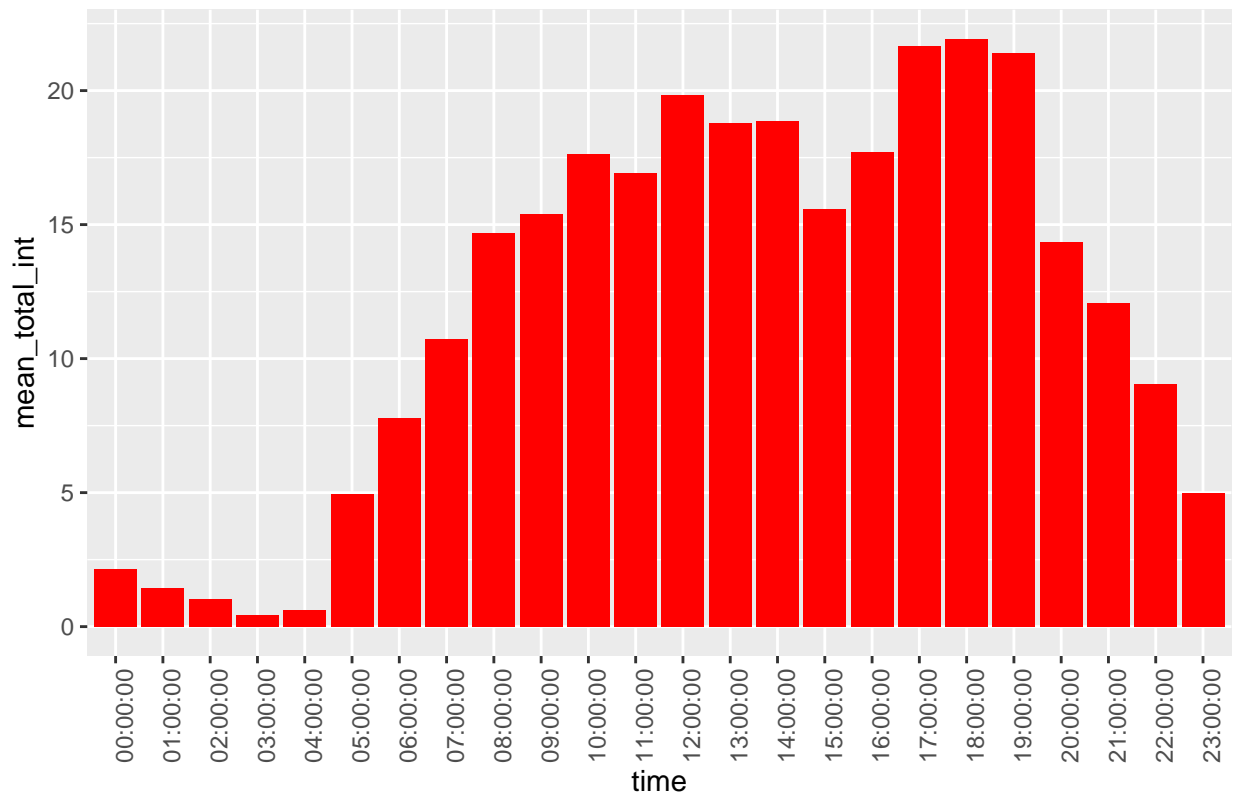
## Again, intensities

```
int_new <- Intensities %>%
  group_by(time) %>%
  drop_na() %>%
  summarise(mean_total_int = mean(TotalIntensity))

ggplot(data=int_new, aes(x=time, y=mean_total_int)) + geom_histogram(stat = "identity", fill='red') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title="Average Total Intensity vs. Time")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Average Total Intensity vs. Time

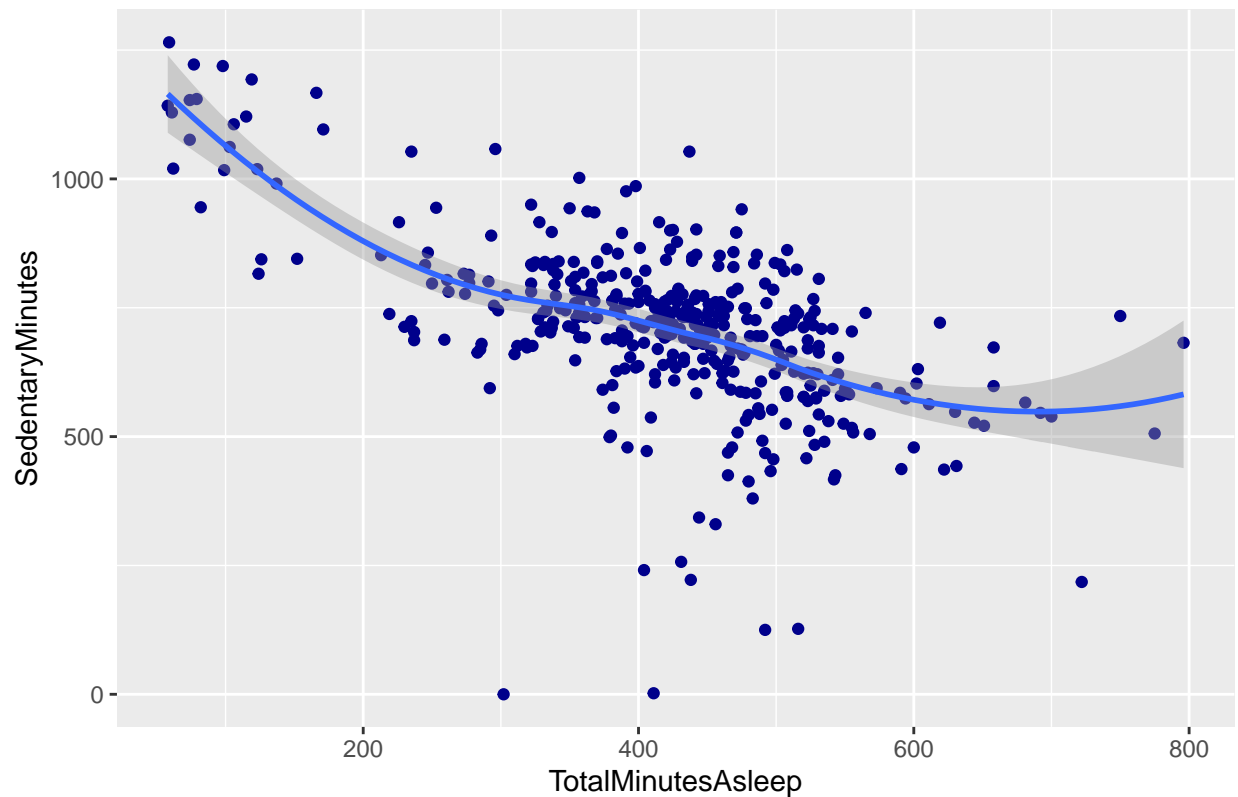


# By this graph,we found out that people are more active between 5 am and 10pm. # Then, visualize the merged data

```
ggplot(data=merged_data, aes(x=TotalMinutesAsleep, y=SedentaryMinutes)) +
  geom_point(color='darkblue') + geom_smooth() +
  labs(title="Minutes Asleep vs. Sedentary Minutes")
```

## 'geom\_smooth()' using method = 'loess' and formula 'y ~ x'

Minutes Asleep vs. Sedentary Minutes



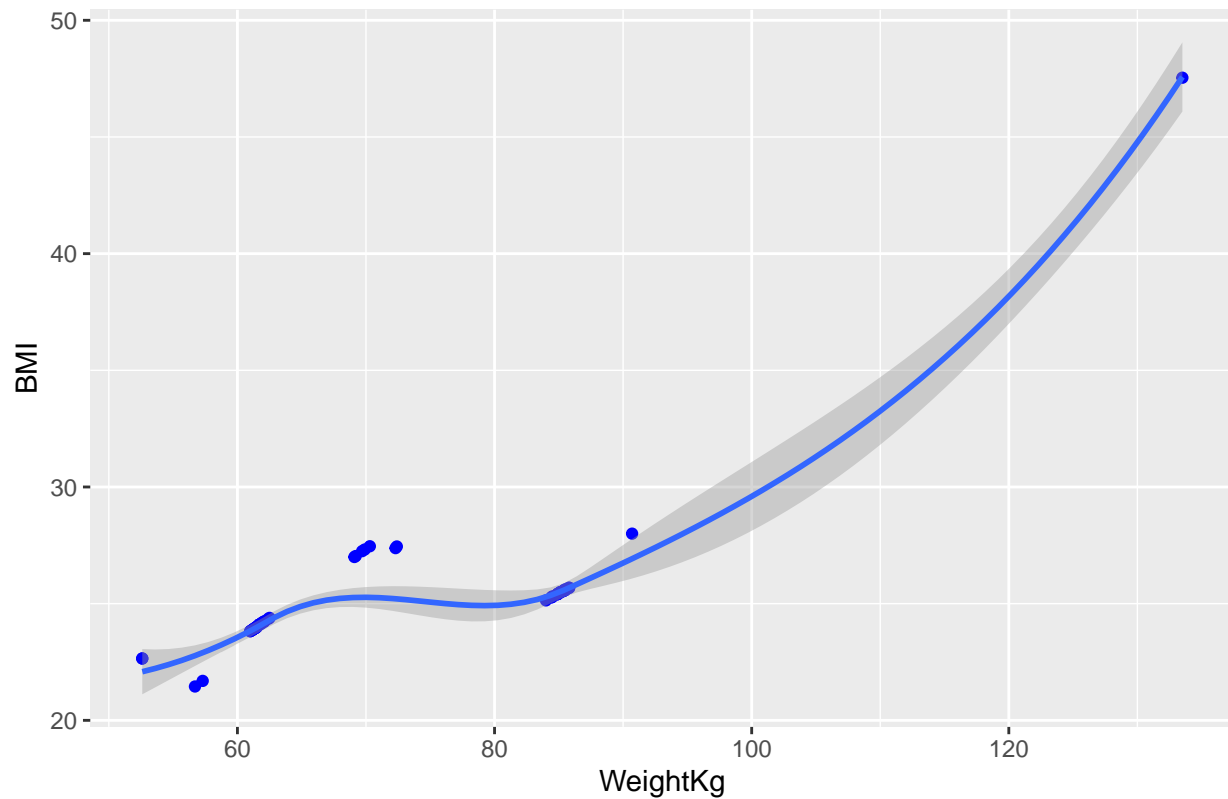
Here we can clearly see the negative relationship between Sedentary Minutes and Sleep time.

As an idea: if Bellabeat users want to improve their sleep, Bellabeat app can recommend reducing sedentary time. # For the weight dataframe

```
ggplot(data=weight, aes(x=WeightKg, y=BMI)) + geom_point(color = 'Blue')+ geom_smooth() +  
  labs(title ="Relationship between Weightkg and BMI Intake")
```

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

Relationship between Weightkg and BMI Intake



A positive correlation obtained between them. # plotting the graph for estimating calories in relation to weightkg

```
merged_2 <- merge(weight, Activity, FUN=c("Id","Date"))
head(merged_2)
```

##	Id	Date	WeightKg	WeightPounds	Fat	BMI
## 1	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 2	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 3	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 4	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 5	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65
## 6	1503960366	5/2/2016 11:59:59 PM	52.6	115.9631	22	22.65

##	IsManualReport	LogId	time	ActivityDate	TotalSteps
## 1	True	1.462234e+12	5/2/2016 11:59:59 PM	2016-04-16	12669
## 2	True	1.462234e+12	5/2/2016 11:59:59 PM	2016-04-18	13019
## 3	True	1.462234e+12	5/2/2016 11:59:59 PM	2016-04-15	9762
## 4	True	1.462234e+12	5/2/2016 11:59:59 PM	2016-05-08	10060
## 5	True	1.462234e+12	5/2/2016 11:59:59 PM	2016-04-17	9705
## 6	True	1.462234e+12	5/2/2016 11:59:59 PM	2016-04-19	15506

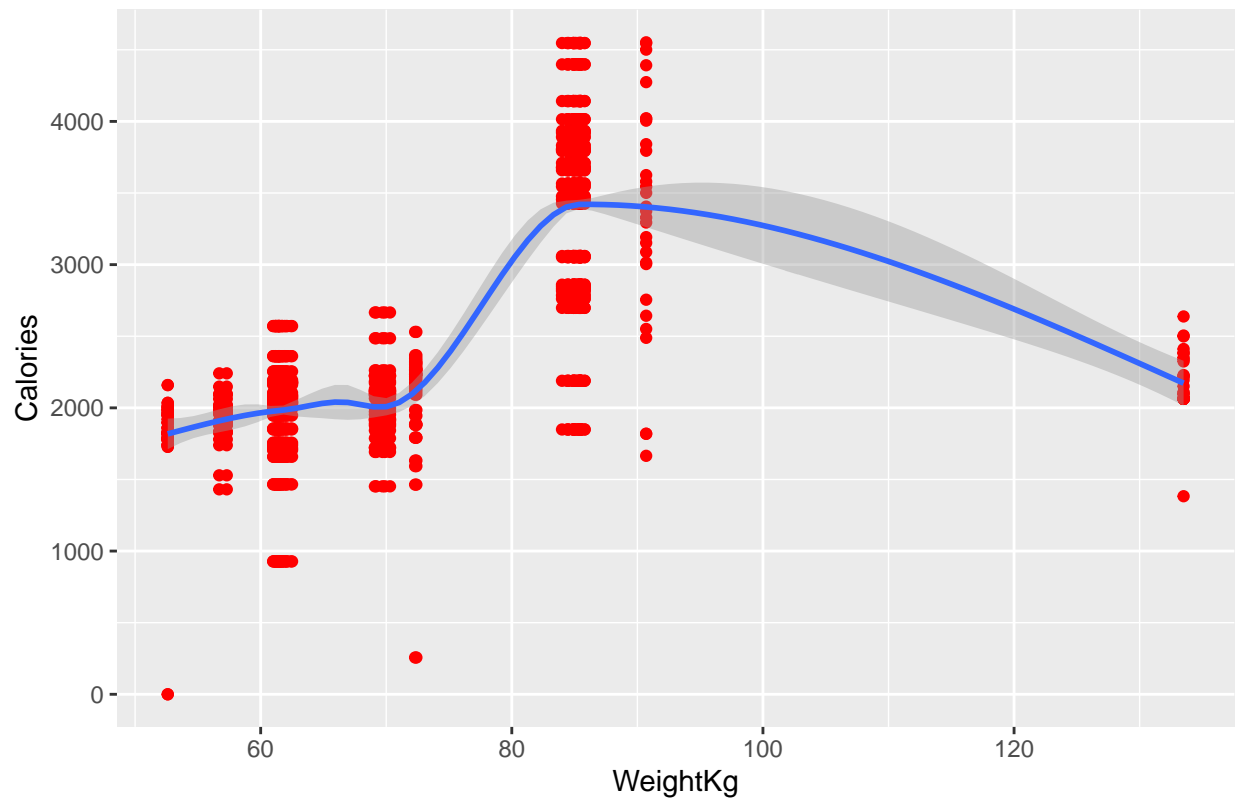
##	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance
## 1	8.16	8.16	0	2.71
## 2	8.59	8.59	0	3.25
## 3	6.28	6.28	0	2.14
## 4	6.58	6.58	0	3.53
## 5	6.48	6.48	0	3.19

```
## 6          9.88          9.88          0          3.53
##  ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
## 1          0.41          5.04          0
## 2          0.64          4.71          0
## 3          1.26          2.83          0
## 4          0.32          2.73          0
## 5          0.78          2.51          0
## 6          1.32          5.03          0
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## 1          36          10          221          773
## 2          42          16          233          1149
## 3          29          34          209          726
## 4          44          8          203          574
## 5          38          20          164          539
## 6          50          31          264          775
##  Calories      date
## 1      1863 04/16/16
## 2      1921 04/18/16
## 3      1745 04/15/16
## 4      1740 05/08/16
## 5      1728 04/17/16
## 6      2035 04/19/16
```

```
ggplot(data=merged_2, aes(x=WeightKg, y=Calories)) + geom_point(color = 'red') + geom_smooth()+
  labs(title = "Relationship between WeightKg and Calories")
```

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

Relationship between WeightKg and Calories



People of 80-100 weightkg are more active than other groups.