Fitbit

Karthik

27/03/2022

##Setting Up my Environment

Notes: Setting up my R environment by loading 'tidyverse', 'lubridate', 'ggplot2' and 'readxl' packages

```
install.packages("sqldf",repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/lm/w5zzwdkj53j10r_167k6dy140000gn/T//RtmpchaBhg/downloaded_packages
```

```
library(tidyverse)  #helps wrangle data
```

```
## -- 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   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

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

```
library(lubridate)  #helps wrangle date attributes
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)  #helps visualize data
library(sqldf)
```

```
## Loading required package: gsubfn
## Loading required package: proto
```

```
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 0x0006): Library not loaded: /
##   Referenced from: /Library/Frameworks/R.framework/Versions/4.1/Resources/modules/R_X11.so
##   Reason: tried: '/opt/X11/lib/libSM.6.dylib' (no such file), '/Library/Frameworks/R.framework/Resou
```

```
## Could not load tcltk.  Will use slower R code instead.
```

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

##Importing requires Datasets

Note: Here we are loading various datasets that are collect form the fitbase

```
dailyActivity_merged <- read.csv("Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")
daily_calories <- read.csv("Fitabase Data 4.12.16-5.12.16/dailyCalories_merged.csv")
sleep_day <- read.csv("Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
daily_intensities <- read.csv("Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv")
weight_log <- read.csv("Fitabase Data 4.12.16-5.12.16/weightLogInfo_merged.csv")
```

##Explore the datasets

```
colnames(dailyActivity_merged) #know column names
```

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

```
glimpse(dailyActivity_merged) #This is like a transposed version of print: columns run down the page, a
```

```
## Rows: 940
## Columns: 15
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate             <chr> "04/12/2016", "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, 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, 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~
```

```r
colnames(daily_calories) #know column names
```

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

```r
glimpse(daily_calories) #This is like a transposed version of print: columns run down the page, and dat
```

```
## Rows: 940
## Columns: 3
## $ Id           <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate <chr> "12/4/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16~
## $ Calories     <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1786, 177~
```

```r
colnames(sleep_day) #know column names
```

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

```r
glimpse(sleep_day) #This is like a transposed version of print: columns run down the page, and data run
```

```
## Rows: 413
## Columns: 5
## $ Id                 <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay           <chr> "12/04/2016 12:00 AM", "4/13/2016 12:00:00 AM", "4/~
## $ TotalSleepRecords  <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~
## $ TotalTimeInBed     <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~
```

```r
colnames(daily_intensities) #know column names
```

```
##  [1] "Id"                       "ActivityDate"
##  [3] "SedentaryMinutes"         "LightlyActiveMinutes"
##  [5] "FairlyActiveMinutes"      "VeryActiveMinutes"
##  [7] "SedentaryActiveDistance"  "LightActiveDistance"
##  [9] "ModeratelyActiveDistance" "VeryActiveDistance"
```

```r
glimpse(daily_intensities) #This is like a transposed version of print: columns run down the page, and
```

```
## Rows: 940
## Columns: 10
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate             <chr> "12/4/2016", "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, 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~
```

```
colnames(weight_log) #know column names
```

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

```
glimpse(weight_log) #This is like a transposed version of print: columns run down the page, and data ru
```

```
## Rows: 67
## Columns: 8
## $ Id             <dbl> 1503960366, 1503960366, 1927972279, 2873212765, 2873212~
## $ Date           <chr> "02/05/2016 11:59 PM", "03/05/2016 11:59 PM", "4/13/201~
## $ WeightKg       <dbl> 52.6, 52.6, 133.5, 56.7, 57.3, 72.4, 72.3, 69.7, 70.3, ~
## $ WeightPounds   <dbl> 115.9631, 115.9631, 294.3171, 125.0021, 126.3249, 159.6~
## $ Fat            <int> 22, NA, NA, NA, NA, 25, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ BMI            <dbl> 22.65, 22.65, 47.54, 21.45, 21.69, 27.45, 27.38, 27.25,~
## $ IsManualReport <lgl> TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, ~
## $ LogId          <dbl> 1.46223e+12, 1.46232e+12, 1.46051e+12, 1.46128e+12, 1.4~
```

##Exploring or Analysing the data and Importing sqldf

Note: All the datasets has Id as common field so can use id as primary field for this. It looks like the daily_activity, daily_calories, and daily_intensities have the exact same number of observations.So we should confirm that the values actually match for any given 'ID' number. Lets write the SQL query /syntax to see if there are any values in daily_calories that are in dailyActivity_merged, so created temp data frame

```
dailyActivity_merged2<- dailyActivity_merged %>%
  select(Id,ActivityDate,Calories)

head(dailyActivity_merged2)
```

```
##           Id ActivityDate Calories
## 1 1503960366   04/12/2016     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
```

```
#finding similar elements from 2 tables
sql_check <- sqldf('Select * From dailyActivity_merged2 INTERSECT SELECT * FROM daily_calories')
head(sql_check)
```

```
##           Id ActivityDate Calories
## 1 1503960366    4/13/2016     1797
## 2 1503960366    4/14/2016     1776
## 3 1503960366    4/15/2016     1745
## 4 1503960366    4/16/2016     1863
## 5 1503960366    4/17/2016     1728
## 6 1503960366    4/18/2016     1921
```

```
nrow(sql_check)#number of rows
```

```
## [1] 578
```

Note: From the above codes we can say that since the first six values of daily_activity and daily_calories are same and total observation of the sql query is 940 the values are the same between the dataframes.

```
dailyActivity_merged3<- dailyActivity_merged %>%
  select(Id, ActivityDate, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, VeryActiveMinute

head(dailyActivity_merged3)
```

```
##           Id ActivityDate SedentaryMinutes LightlyActiveMinutes
## 1 1503960366   04/12/2016              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
##   FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
## 1                  13                25                       0
## 2                  19                21                       0
## 3                  11                30                       0
## 4                  34                29                       0
## 5                  10                36                       0
## 6                  20                38                       0
##   LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
## 1                6.06                     0.55               1.88
## 2                4.71                     0.69               1.57
## 3                3.91                     0.40               2.44
## 4                2.83                     1.26               2.14
## 5                5.04                     0.41               2.71
## 6                2.51                     0.78               3.19
```

```
sql_check2 <- sqldf('Select * from dailyActivity_merged3 INTERSECT select * from daily_intensities')
head(sql_check2)
```

```
##           Id ActivityDate SedentaryMinutes LightlyActiveMinutes
## 1 1503960366    4/13/2016              776                  217
## 2 1503960366    4/14/2016             1218                  181
## 3 1503960366    4/15/2016              726                  209
## 4 1503960366    4/16/2016              773                  221
## 5 1503960366    4/17/2016              539                  164
## 6 1503960366    4/18/2016             1149                  233
##   FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
## 1                  19                21                       0
## 2                  11                30                       0
## 3                  34                29                       0
## 4                  10                36                       0
## 5                  20                38                       0
## 6                  16                42                       0
##   LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
## 1                4.71                     0.69               1.57
## 2                3.91                     0.40               2.44
## 3                2.83                     1.26               2.14
## 4                5.04                     0.41               2.71
## 5                2.51                     0.78               3.19
## 6                4.71                     0.64               3.25
```

```
nrow(sql_check2)
```

```
## [1] 578
```

##Analysing the datas from the based on assumption made

###Checking if the data in dailyActivity_merged is greater than that in sleep_day and weight_log

```
n_distinct(dailyActivity_merged$Id)
```

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

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

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

```
n_distinct(weight_log$Id)
```

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

As per the understanding its seen that the data in dailyActivity_merged is more

###Number of Observation in Each dataframe

```
nrow(dailyActivity_merged)
```

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

```
nrow(sleep_day)
```

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

```
nrow(weight_log)
```

```
## [1] 67
```

###Getting the summary of Tables

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

```
##    TotalSteps    TotalDistance    SedentaryMinutes VeryActiveMinutes
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :  0.00  
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:  0.00  
##  Median : 7406   Median : 5.245   Median :1057.5   Median :  4.00  
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   : 21.16  
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.: 32.00  
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :210.00  
```

```
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  
```

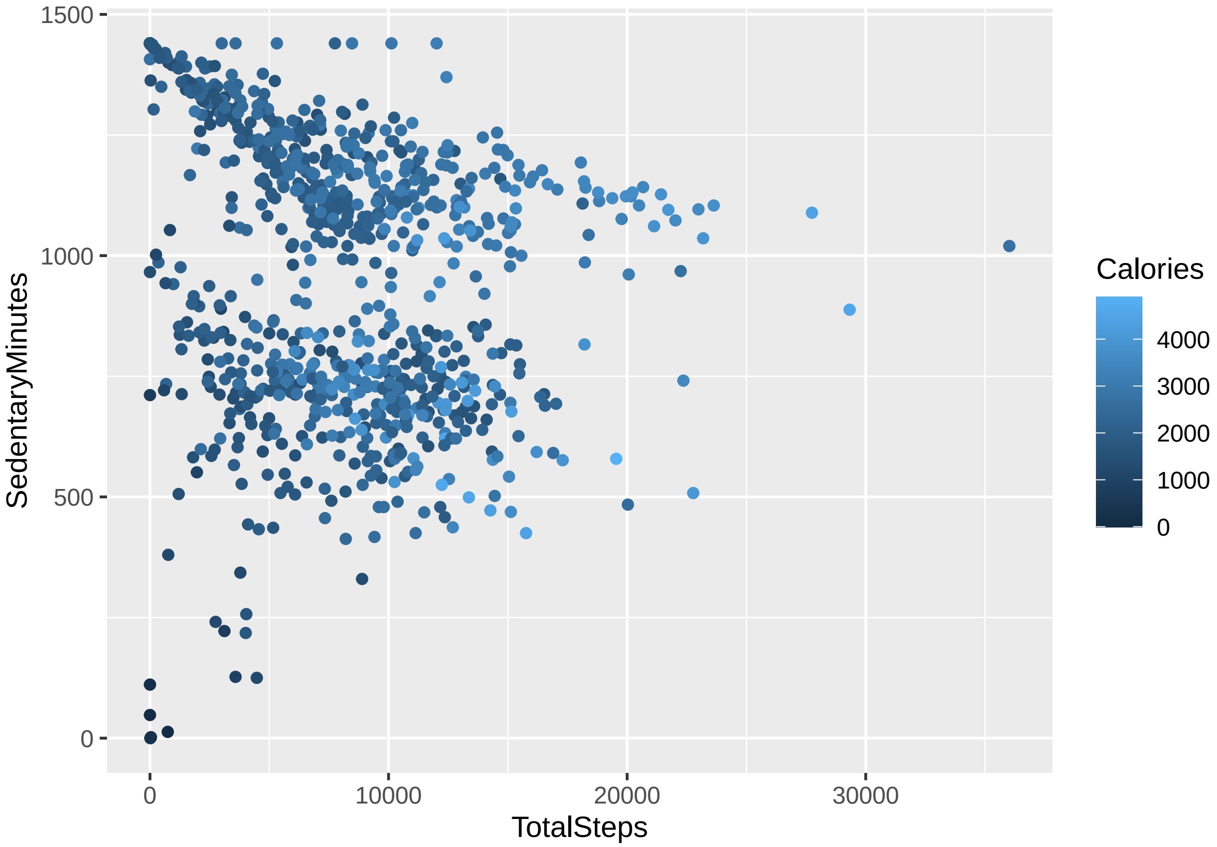
```
weight_log %>%
  select(WeightPounds,
         BMI) %>%
  summary()
```

```
##   WeightPounds        BMI       
##  Min.   :116.0   Min.   :21.45  
##  1st Qu.:135.4   1st Qu.:23.96  
##  Median :137.8   Median :24.39  
##  Mean   :158.8   Mean   :25.19  
##  3rd Qu.:187.5   3rd Qu.:25.56  
##  Max.   :294.3   Max.   :47.54  
```

##Visualisation or Plotting the exploration

I would like to start with the relationship between steps taken in a da and sedentary(people were inactive) minutes

```
ggplot(data=dailyActivity_merged, aes(x=TotalSteps, y=SedentaryMinutes, color = Calories)) + geom_point
```

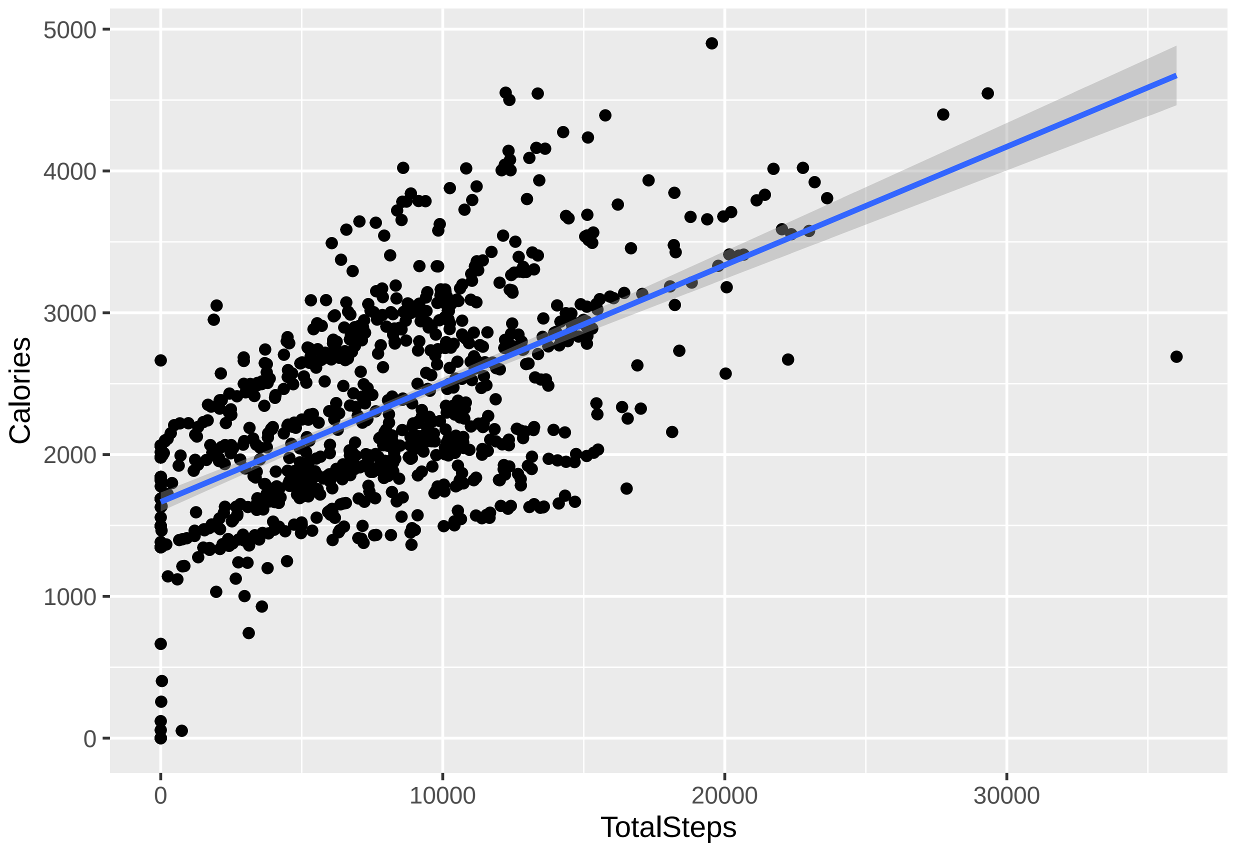


Form thus we could see that total as one doesnt move when inactive

Now I will plot the graph between calories and total steps to see the relationship between them.

```
ggplot(data=dailyActivity_merged, aes(x=TotalSteps, y = Calories))+ geom_point() + stat_smooth(method=lı
```

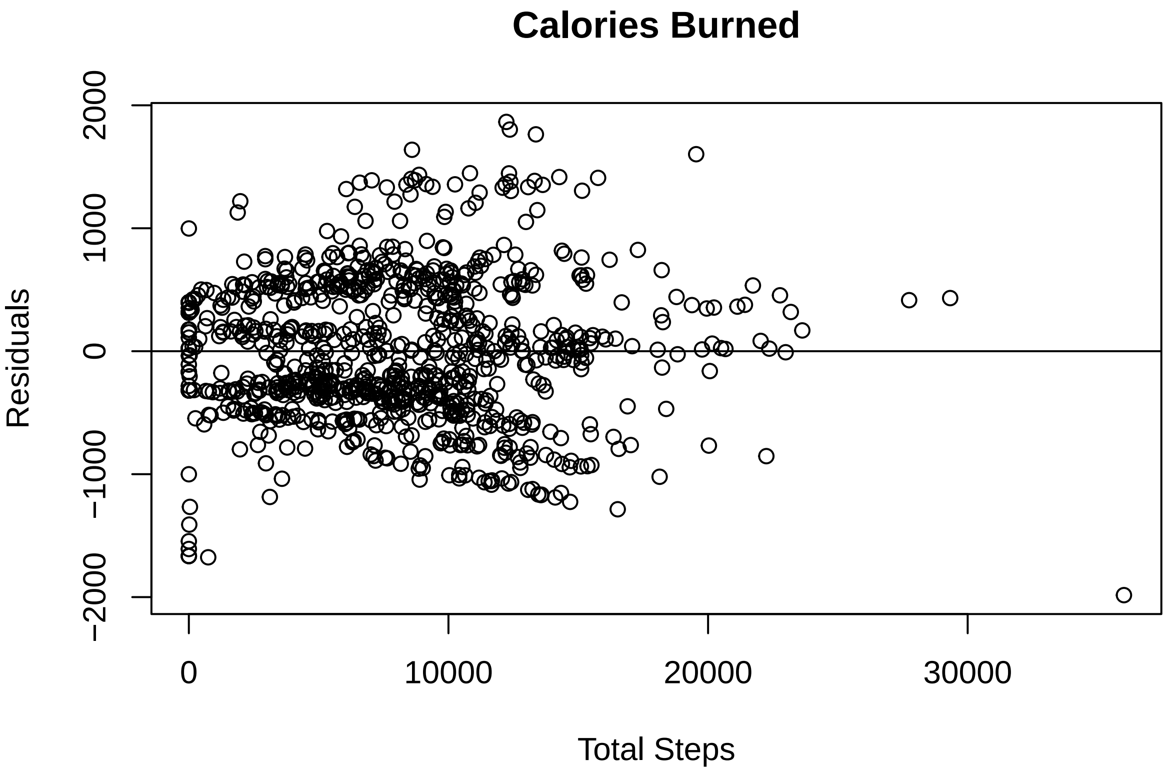
```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



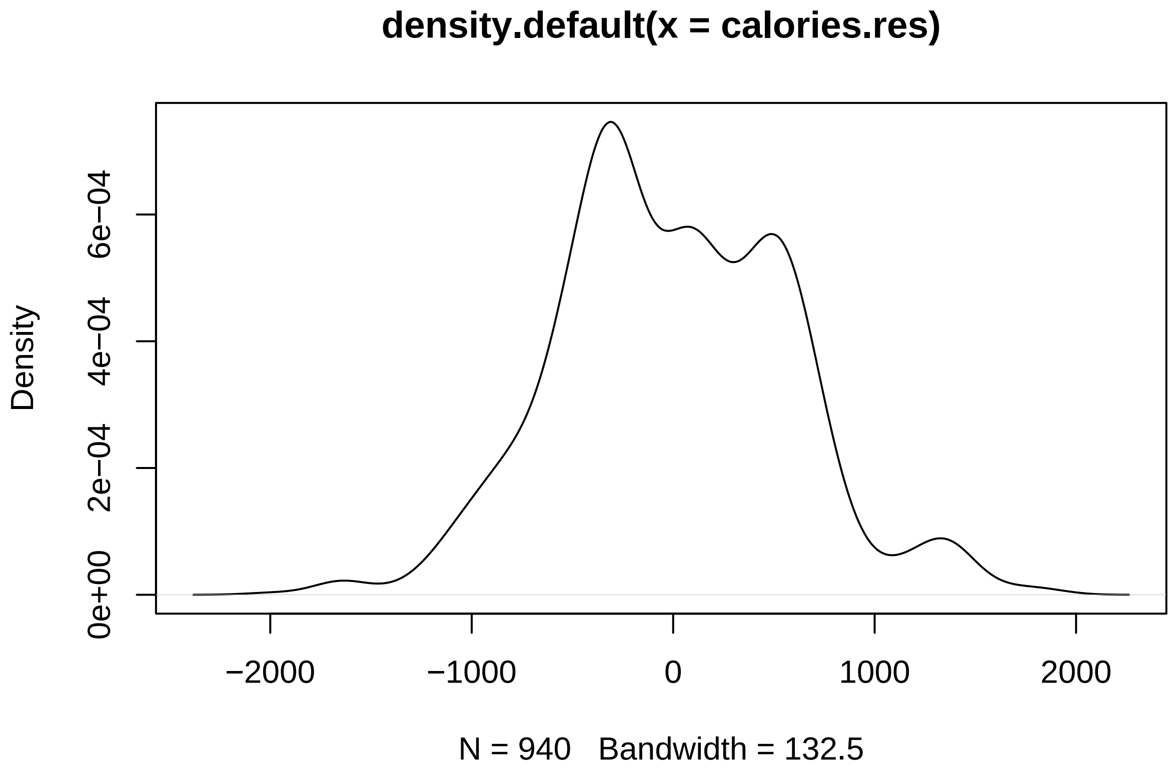
This shows the difference between estimated and actual calories

```
calories.lm <- lm(Calories ~ TotalSteps, data = dailyActivity_merged)
calories.res <- resid(calories.lm)

plot(dailyActivity_merged$TotalSteps, calories.res, ylab="Residuals",
     xlab = "Total Steps", main = "Calories Burned")
abline(0,0)
```

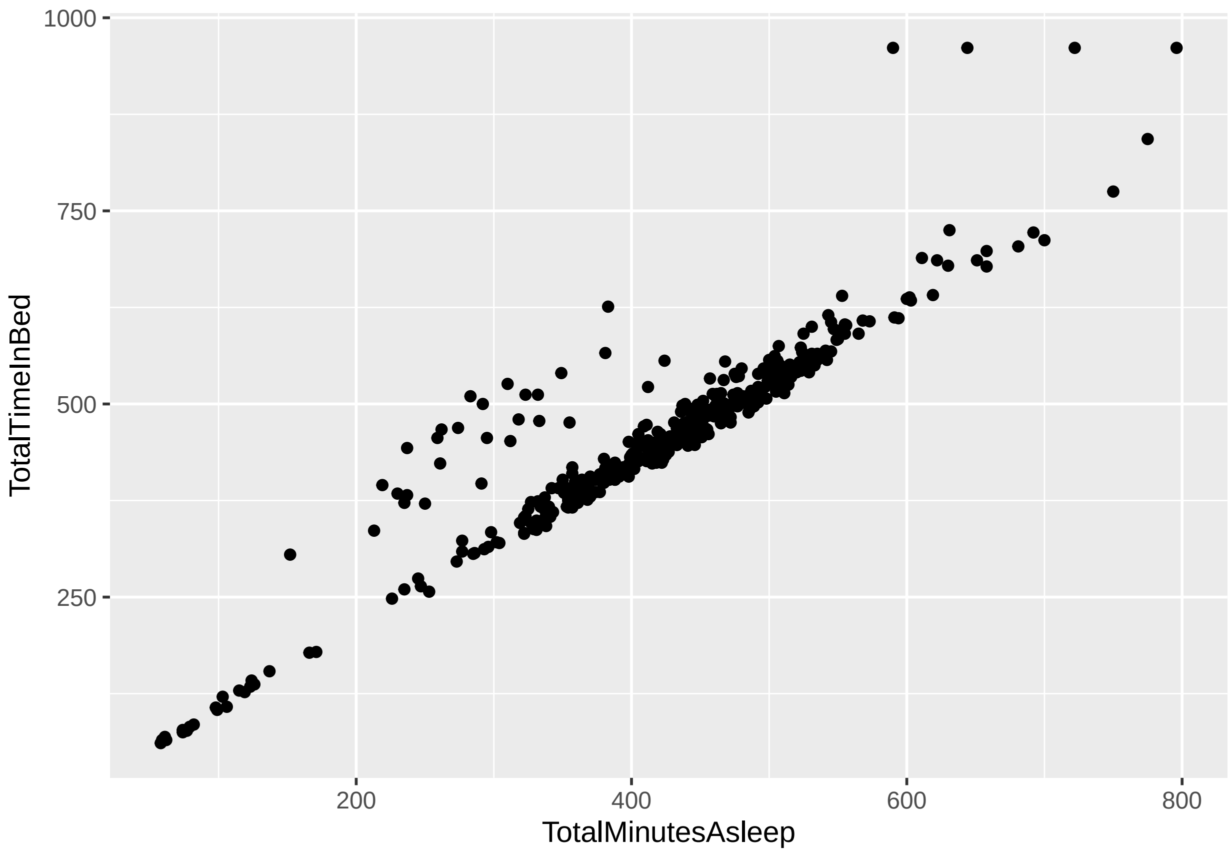


```
#This looks messy so we will plot using desity plot
plot(density(calories.res))
```



Let's look at our sleep data, we should see a practically 1:1 trend from the amount of time slept and the total time someone spends in bed.

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



As sleep hour and sedentary minutes are similar we could merge the to set by ID field

```
combined_sleep_day_data <- merge(sleep_day,dailyActivity_merged,by="Id")
head(combined_sleep_day_data)
```

```
##           Id            SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 12/04/2016 12:00 AM                 1                327
## 2 1503960366 12/04/2016 12:00 AM                 1                327
## 3 1503960366 12/04/2016 12:00 AM                 1                327
## 4 1503960366 12/04/2016 12:00 AM                 1                327
## 5 1503960366 12/04/2016 12:00 AM                 1                327
## 6 1503960366 12/04/2016 12:00 AM                 1                327
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1            346   05/07/2016      11992          7.71            7.71
## 2            346   05/06/2016      12159          8.03            8.03
## 3            346   05/01/2016      10602          6.81            6.81
## 4            346    4/30/2016      14673          9.25            9.25
## 5            346   04/12/2016      13162          8.50            8.50
## 6            346    4/13/2016      10735          6.97            6.97
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               2.46                     2.12
## 2                        0               1.97                     0.25
## 3                        0               2.29                     1.60
## 4                        0               3.56                     1.42
## 5                        0               1.88                     0.55
## 6                        0               1.57                     0.69
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                3.13                       0                37
## 2                5.81                       0                24
## 3                2.92                       0                33
## 4                4.27                       0                52
## 5                6.06                       0                25
## 6                4.71                       0                21
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  46                  175              833     1821
## 2                   6                  289              754     1896
## 3                  35                  246              730     1820
## 4                  34                  217              712     1947
## 5                  13                  328              728     1985
## 6                  19                  217              776     1797
```

```
n_distinct(combined_sleep_day_data$Id)
```

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

```
combined_sleep_day_data2 <- merge(sleep_day,dailyActivity_merged,by="Id",all=TRUE)
head(combined_sleep_day_data2)
```

```
##           Id            SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 12/04/2016 12:00 AM                 1                327
## 2 1503960366 12/04/2016 12:00 AM                 1                327
## 3 1503960366 12/04/2016 12:00 AM                 1                327
## 4 1503960366 12/04/2016 12:00 AM                 1                327
## 5 1503960366 12/04/2016 12:00 AM                 1                327
## 6 1503960366 12/04/2016 12:00 AM                 1                327
##   TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1            346   05/07/2016      11992          7.71            7.71
## 2            346   05/06/2016      12159          8.03            8.03
## 3            346   05/01/2016      10602          6.81            6.81
## 4            346    4/30/2016      14673          9.25            9.25
## 5            346   04/12/2016      13162          8.50            8.50
## 6            346    4/13/2016      10735          6.97            6.97
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               2.46                     2.12
## 2                        0               1.97                     0.25
## 3                        0               2.29                     1.60
## 4                        0               3.56                     1.42
## 5                        0               1.88                     0.55
## 6                        0               1.57                     0.69
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                3.13                       0                37
## 2                5.81                       0                24
## 3                2.92                       0                33
## 4                4.27                       0                52
## 5                6.06                       0                25
## 6                4.71                       0                21
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  46                  175              833     1821
## 2                   6                  289              754     1896
## 3                     35                    246               730      1820
## 4                     34                    217               712      1947
## 5                     13                    328               728      1985
## 6                     19                    217               776      1797
```

```
n_distinct(combined_sleep_day_data2$Id)
```

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

###Sedentary time VS Sleep Time

```
sedentary.lm <- lm(SedentaryMinutes ~ TotalTimeInBed, data = combined_sleep_day_data)
sedentary.lm
```

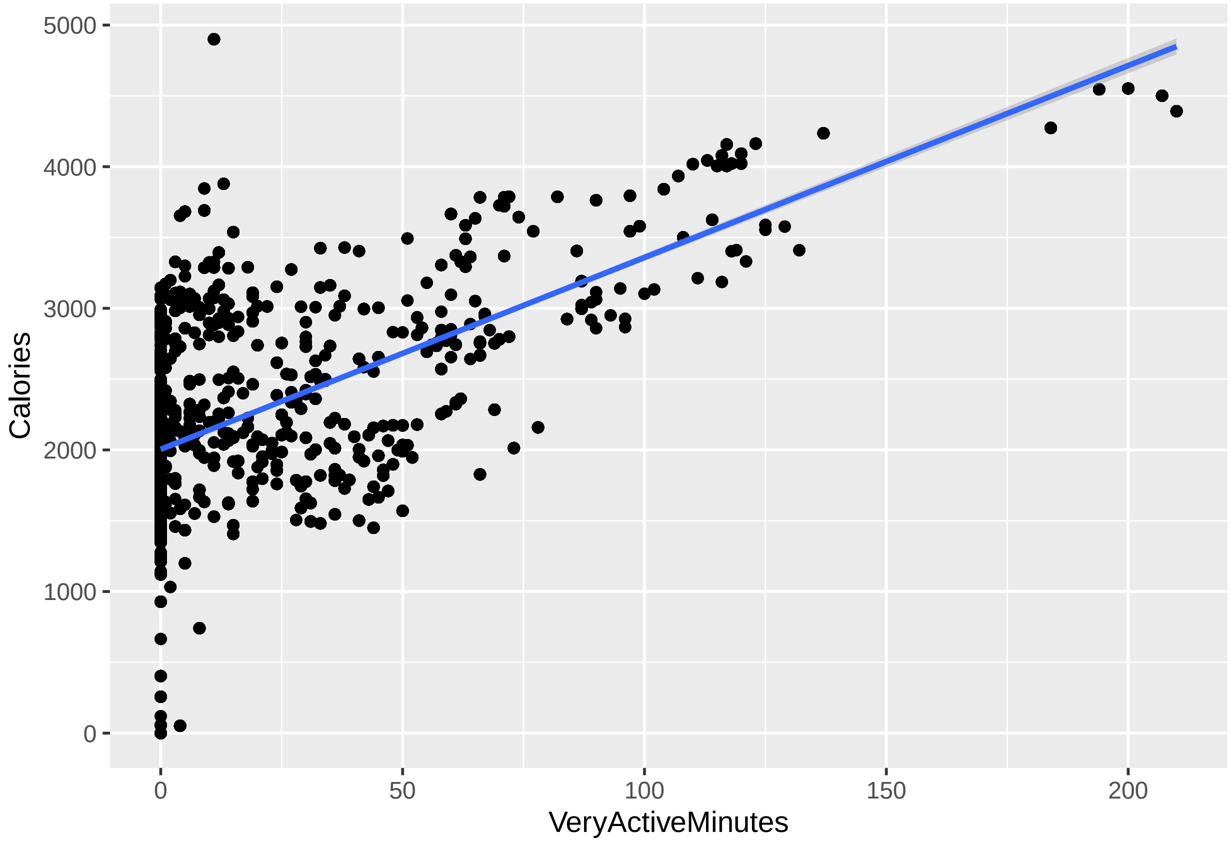
```
##
## Call:
## lm(formula = SedentaryMinutes ~ TotalTimeInBed, data = combined_sleep_day_data)
##
## Coefficients:
##    (Intercept)  TotalTimeInBed
##       921.9598         -0.2678
```

```
#And now a pearson correlation coefficient:
cor(combined_sleep_day_data$TotalTimeInBed,combined_sleep_day_data$SedentaryMinutes, method = "pearson"
```

```
## [1] -0.128011
```

```
ggplot(data = combined_sleep_day_data, aes(x=VeryActiveMinutes, y=Calories)) + geom_point() + stat_smoo
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



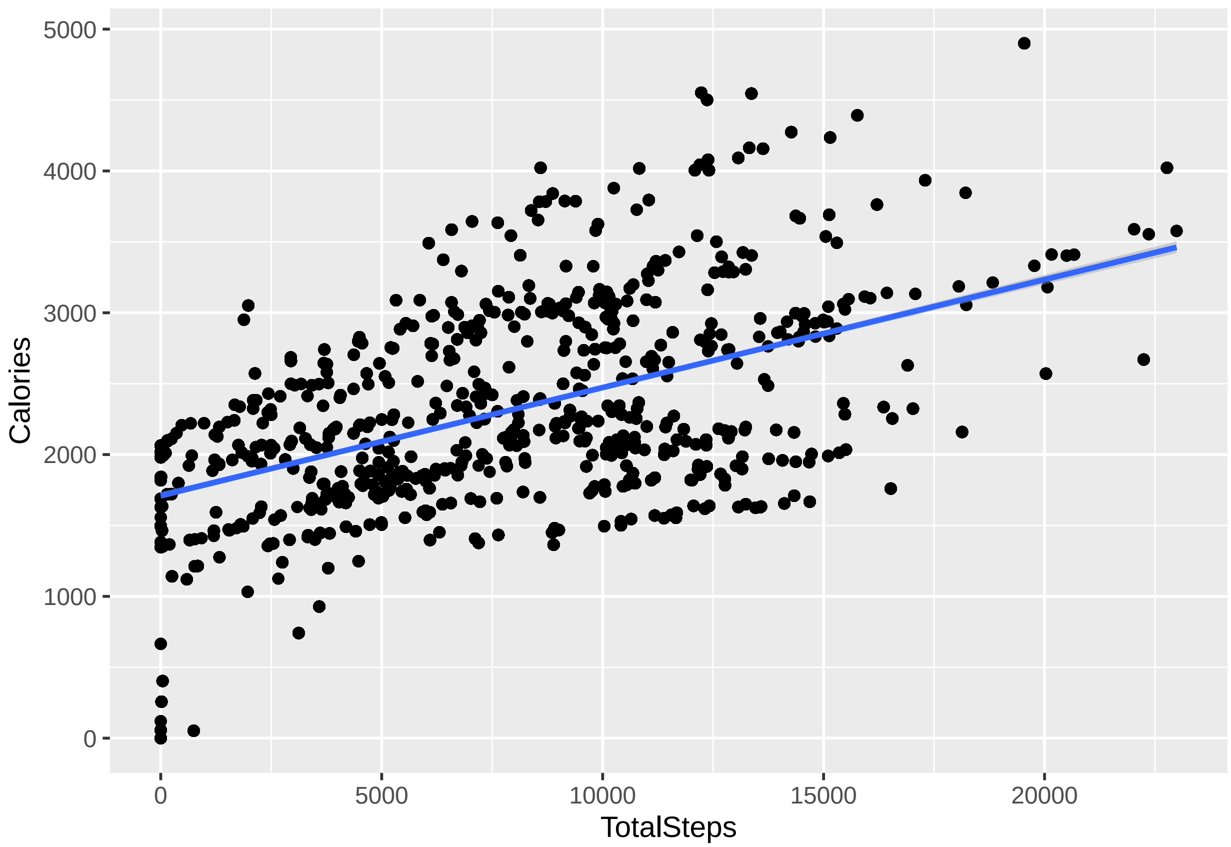
Lets check correlation between total steps taken and calories

```
lm(Calories ~ VeryActiveMinutes, data = combined_sleep_day_data)
```

```
##
## Call:
## lm(formula = Calories ~ VeryActiveMinutes, data = combined_sleep_day_data)
##
## Coefficients:
##       (Intercept)  VeryActiveMinutes
##           2004.36              13.55
```

```
ggplot(data = combined_sleep_day_data, aes(x=TotalSteps, y=Calories)) + geom_point() +stat_smooth(metho
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



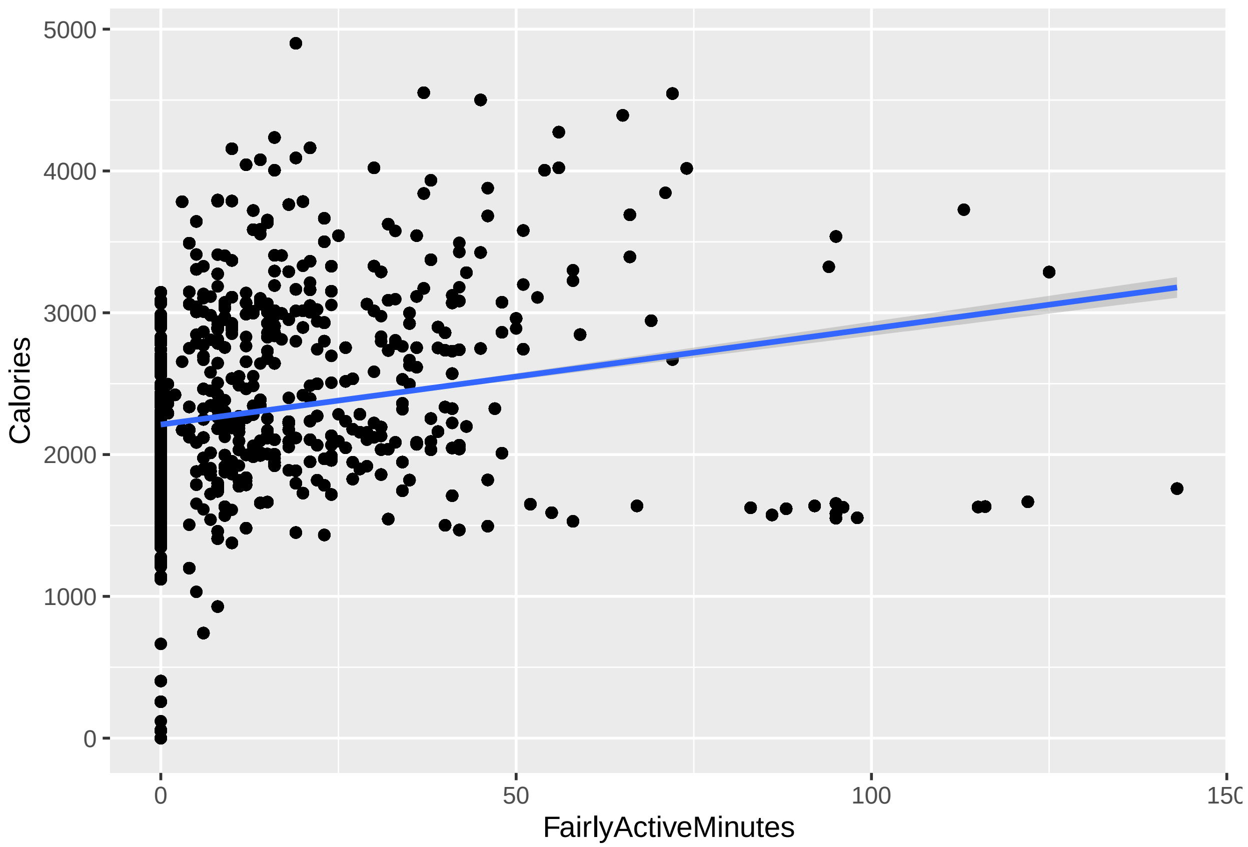
Correlation between fairlyactivemiutes taken and calories.

```
lm(Calories ~ TotalSteps, data = combined_sleep_day_data)
```

```
##
## Call:
## lm(formula = Calories ~ TotalSteps, data = combined_sleep_day_data)
##
## Coefficients:
## (Intercept)   TotalSteps
##   1.711e+03    7.616e-02
```

```
ggplot(data = combined_sleep_day_data, aes(x=FairlyActiveMinutes, y=Calories)) + geom_point() + stat_sm
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
lm(Calories ~ FairlyActiveMinutes, data = combined_sleep_day_data)
```

```
##
## Call:
## lm(formula = Calories ~ FairlyActiveMinutes, data = combined_sleep_day_data)
##
## Coefficients:
##         (Intercept)  FairlyActiveMinutes
##             2211.85                 6.76
```

##Conclusion I prepossessed, explored, analysed and visualized the fitbit users dataset quite deeply, and gave some marketing strategy above.

###Final Marketing Strategy

I would focus on the fact that simply collecting more data from different competitors one could see more trends.

Also the best relationship was in between veryactiveminutes and calories so the people who are very active tend to burn the most calorie this can be a good marketing strategy.

We could also add the features that would automatically measures the calories intake based on the food and beverages that was consumed and show how much of the calories intake today was not composated by workout