

How Can a Wellness Technology Company Play It Smart?

Ankita Gairola

2023-07-30

Introduction

This case study from the Google Data Analytics professional Certificate.

1. Ask

Bellabeat is a high-tech manufacturer of health-focused products for women. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

The aim of this project to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices.

Questions:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

Deliverable:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top high-level content recommendations based on your analysis.

2. Prepare

Data Source : Kaggle and contain 18 csv file

Sample size: 30 Fitbit users

Data Period: 12 March 2016 to 12 May 2016

3. Process

Set working Directory

```
getwd()
```

```
setwd("E:/Capstone projects_Google/Case Study 2")
```

Required Packages

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

```
install.packages("readxl")
```

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

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

```
install.packages("lubridate")
```

```
install.packages("scales")
```

```
install.packages("here")
```

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

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

```
install.packages("RColorBrewer")
```

```
install.packages("Tinytex")
```

```
library("tidyverse")
```

```
## — Attaching core tidyverse packages — tidyverse  
2.0.0 —
```

```
## ✓ dplyr      1.1.2      ✓ readr      2.1.4
```

```
## ✓ forcats   1.0.0      ✓ stringr    1.5.0
```

```
## ✓ ggplot2    3.4.2      ✓ tibble     3.2.1
```

```
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
```

```
## ✓ purrr      1.0.1
```

```
## — Conflicts —
```

```
tidyverse_conflicts() —
```

```
## ✗ dplyr::filter() masks stats::filter()
```

```
## ✗ dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all  
conflicts to become errors
```

```
library("readxl")
```

```
library("skimr")
```

```
library("janitor")
```

```
##
```

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

```
##
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      chisq.test, fisher.test
```

```

library("lubridate")

library("scales")

##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor

library("here")

## here() starts at E:/Capstone projects_Google/Case Study 2

library("ggplot2")

library("dplyr")

library(RColorBrewer)

library(tinytex)

```

Data Importing

```

Daily_Activity <- read.csv("E:/Capstone projects_Google/Case Study 2/Data
Set/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")

Daily_Calories <- read.csv("E:/Capstone projects_Google/Case Study 2/Data
Set/Fitabase Data 4.12.16-5.12.16/dailyCalories_merged.csv")

Daily_Steps <- read.csv("E:/Capstone projects_Google/Case Study 2/Data
Set/Fitabase Data 4.12.16-5.12.16/dailySteps_merged.csv")

Sleep_Day <- read.csv("E:/Capstone projects_Google/Case Study 2/Data
Set/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")

Daily_Intensities <- read.csv("E:/Capstone projects_Google/Case Study 2/Data
Set/Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv")

Weight_Log <- read.csv("E:/Capstone projects_Google/Case Study 2/Data
Set/Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv")

```

Overview of data

```
skim_without_charts(Daily_Activity)
```

Data summary

Name	Daily_Activity
------	----------------

Number of rows 940
Number of columns 15

Column type frequency:

character 1
numeric 14

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDate	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.855407e+09	2.424805e+09	1503960366	2.320127e+09	4.445115e+09	6.962181e+09	8.877689e+09
TotalSteps	0	1	7.637910e+03	5.087150e+03	0	3.789750e+03	7.405500e+03	1.072700e+04	3.601900e+04
TotalDistance	0	1	5.490000e+00	3.920000e+00	0	2.620000e+00	5.240000e+00	7.710000e+00	2.803000e+01
TrackerDistance	0	1	5.480000e+00	3.910000e+00	0	2.620000e+00	5.240000e+00	7.710000e+00	2.803000e+01
LoggedActivitiesDistance	0	1	1.100000e-01	6.200000e-01	0	0.000000e+00	0.000000e+00	0.000000e+00	4.940000e+00
VeryActiveDistance	0	1	1.500000e+00	2.660000e+00	0	0.000000e+00	2.100000e-01	2.050000e+00	2.192000e+01
ModeratelyActiveDistance	0	1	5.700000e-01	8.800000e-01	0	0.000000e+00	2.400000e-01	8.000000e-01	6.480000e+00
LightActiveDistance	0	1	3.340000e+00	2.040000e+00	0	1.950000e+00	3.360000e+00	4.780000e+00	1.071000e+01

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
SedentaryActiveDistance	0	1	0.0000 00e+0 0	1.0000 00e- 02	0	0.0000 00e+0 0	0.0000 00e+0 0	0.0000 00e+0 0	1.1000 00e- 01
VeryActiveMinutes	0	1	2.1160 00e+0 1	3.2840 00e+0 1	0	0.0000 00e+0 0	4.0000 00e+0 0	3.2000 00e+0 1	2.1000 00e+0 2
FairlyActiveMinutes	0	1	1.3560 00e+0 1	1.9990 00e+0 1	0	0.0000 00e+0 0	6.0000 00e+0 0	1.9000 00e+0 1	1.4300 00e+0 2
LightlyActiveMinutes	0	1	1.9281 00e+0 2	1.0917 00e+0 2	0	1.2700 00e+0 2	1.9900 00e+0 2	2.6400 00e+0 2	5.1800 00e+0 2
SedentaryMinutes	0	1	9.9121 00e+0 2	3.0127 00e+0 2	0	7.2975 00e+0 2	1.0575 00e+0 3	1.2295 00e+0 3	1.4400 00e+0 3
Calories	0	1	2.3036 10e+0 3	7.1817 00e+0 2	0	1.8285 00e+0 3	2.1340 00e+0 3	2.7932 50e+0 3	4.9000 00e+0 3

`skim_without_charts(Daily_Calories)`

Data summary

Name Daily_Calories
Number of rows 940
Number of columns 3

Column type frequency:

character 1
numeric 2

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDay	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
---------------	-----------	---------------	------	----	----	-----	-----	-----	------

skim_v variable	n_mi ssing	comple te_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.8554 07e+09	2.4248 05e+09	15039 60366	232012 7002.0	44451 14986	6.9621 81e+09	88776 89391
Calorie s	0	1	2.3036 10e+03	7.1817 00e+02	0	1828.5	2134	2.7932 50e+03	4900
skim_without_charts(Daily_Intensities)									

Data summary

Name Daily_Intensities
Number of rows 940
Number of columns 10

Column type frequency:

character 1
numeric 9

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDay	0	1	8	9	0	31	0

Variable type: numeric

skim_variab le	n_m issin g	compl ete_ra te	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.8554 07e+0 9	2.4248 05e+0 9	1503 9603 66	2.3201 27e+0 9	4.4451 15e+0 9	6.9621 81e+0 9	8.8776 89e+0 9
SedentaryM inutes	0	1	9.9121 00e+0 2	3.0127 00e+0 2	0	7.2975 00e+0 2	1.0575 00e+0 3	1.2295 00e+0 3	1.4400 00e+0 3
LightlyActiv eMinutes	0	1	1.9281 00e+0 2	1.0917 00e+0 2	0	1.2700 00e+0 2	1.9900 00e+0 2	2.6400 00e+0 2	5.1800 00e+0 2
FairlyActive Minutes	0	1	1.3560 00e+0 1	1.9990 00e+0 1	0	0.0000 00e+0 0	6.0000 00e+0 0	1.9000 00e+0 1	1.4300 00e+0 2
VeryActive Minutes	0	1	2.1160 00e+0	3.2840 00e+0	0	0.0000 00e+0	4.0000 00e+0	3.2000 00e+0	2.1000 00e+0

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
			1	1		0	0	1	2
SedentaryActiveDistance	0	1	0.0000 00e+00	1.0000 00e-02	0	0.0000 00e+00	0.0000 00e+00	0.0000 00e+00	1.1000 00e-01
LightActiveDistance	0	1	3.3400 00e+00	2.0400 00e+00	0	1.9500 00e+00	3.3600 00e+00	4.7800 00e+00	1.0710 00e+00
ModeratelyActiveDistance	0	1	5.7000 00e-01	8.8000 00e-01	0	0.0000 00e+00	2.4000 00e-01	8.0000 00e-01	6.4800 00e+00
VeryActiveDistance	0	1	1.5000 00e+00	2.6600 00e+00	0	0.0000 00e+00	2.1000 00e-01	2.0500 00e+00	2.1920 00e+00

`skim_without_charts(Daily_Steps)`

Data summary

Name Daily_Steps
Number of rows 940
Number of columns 3

Column type frequency:

character 1
numeric 2

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDay	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.8554 07e+09	2.4248 05e+09	15039 60366	2.3201 27e+09	444511 4986.0	69621 81067	88776 89391
StepTotal	0	1	7.6379 10e+03	5.0871 50e+03	0	3.7897 50e+03	7405.5	10727	36019

`skim_without_charts(Sleep_Day)`

Data summary

Name Sleep_Day
Number of rows 413
Number of columns 5

Column type frequency:

character 1
numeric 4

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
SleepDay	0	1	20	21	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	5.000979e+09	2.06036e+09	1503960366	3977333714	4702921684	6962181067	8792009665
TotalSleepRecords	0	1	1.120000e+00	3.50000e-01	1	1	1	1	3
TotalMinutesAsleep	0	1	4.194700e+02	1.18340e+02	58	361	433	490	796
TotalTimeInBed	0	1	4.586400e+02	1.27100e+02	61	403	463	526	961

[skim_without_charts](#)(Weight_Log)

Data summary

Name Weight_Log
Number of rows 940
Number of columns 10

Column type frequency:

character 1
numeric 9

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ActivityDay	0	1	8	9	0	31	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
Id	0	1	4.855407e+09	2.424805e+09	1503966	2.320127e+09	4.445115e+09	6.962181e+09	8.877689e+09
SedentaryMinutes	0	1	9.912100e+02	3.012700e+02	0	7.297500e+02	1.057500e+03	1.229500e+03	1.440000e+03
LightlyActiveMinutes	0	1	1.928100e+02	1.091700e+02	0	1.270000e+02	1.990000e+02	2.640000e+02	5.180000e+02
FairlyActiveMinutes	0	1	1.356000e+01	1.999000e+01	0	0.000000e+00	6.000000e+00	1.900000e+01	1.430000e+02
VeryActiveMinutes	0	1	2.116000e+01	3.284000e+01	0	0.000000e+00	4.000000e+00	3.200000e+01	2.100000e+02
SedentaryActiveDistance	0	1	0.000000e+00	1.000000e-02	0	0.000000e+00	0.000000e+00	0.000000e+00	1.100000e+01
LightActiveDistance	0	1	3.340000e+00	2.040000e+00	0	1.950000e+00	3.360000e+00	4.780000e+00	1.071000e+01
ModeratelyActiveDistance	0	1	5.700000e-01	8.800000e-01	0	0.000000e+00	2.400000e-01	8.000000e-01	6.480000e+00
VeryActiveDistance	0	1	1.500000e+00	2.660000e+00	0	0.000000e+00	2.100000e-01	2.050000e+00	2.192000e+01

colnames(Daily_Activity)

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

```
colnames(Daily_Calories)
```

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

```
colnames(Daily_Intensities)
```

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

```
colnames(Daily_Steps)
```

```
## [1] "Id"          "ActivityDay" "StepTotal"
```

```
colnames(Sleep_Day)
```

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

```
colnames(Weight_Log)
```

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

Data Cleaning

Data Information checking

```
str(Daily_Activity)
```

```
## 'data.frame': 940 obs. of 15 variables:
## $ Id : num 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09
## ...
## $ ActivityDate : chr "4/12/2016" "4/13/2016" "4/14/2016"
## "4/15/2016" ...
## $ TotalSteps : int 13162 10735 10460 9762 12669 9705 13019
## 15506 10544 9819 ...
## $ TotalDistance : num 8.5 6.97 6.74 6.28 8.16 ...
## $ TrackerDistance : num 8.5 6.97 6.74 6.28 8.16 ...
## $ LoggedActivitiesDistance: num 0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveDistance : num 1.88 1.57 2.44 2.14 2.71 ...
## $ ModeratelyActiveDistance: num 0.55 0.69 0.4 1.26 0.41 ...
## $ LightActiveDistance : num 6.06 4.71 3.91 2.83 5.04 ...
## $ SedentaryActiveDistance : num 0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveMinutes : int 25 21 30 29 36 38 42 50 28 19 ...
## $ FairlyActiveMinutes : int 13 19 11 34 10 20 16 31 12 8 ...
## $ LightlyActiveMinutes : int 328 217 181 209 221 164 233 264 205 211
## ...
## $ SedentaryMinutes : int 728 776 1218 726 773 539 1149 775 818
```

```

838 ...
## $ Calories          : int  1985 1797 1776 1745 1863 1728 1921 2035
1786 1775 ...

str(Daily_Calories)

## 'data.frame':      940 obs. of  3 variables:
## $ Id              : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ Calories       : int  1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...

str(Daily_Intensities)

## 'data.frame':      940 obs. of  10 variables:
## $ Id              : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09
...
## $ ActivityDay      : chr  "4/12/2016" "4/13/2016" "4/14/2016"
"4/15/2016" ...
## $ SedentaryMinutes  : int  728 776 1218 726 773 539 1149 775 818
838 ...
## $ LightlyActiveMinutes : int  328 217 181 209 221 164 233 264 205 211
...
## $ FairlyActiveMinutes : int  13 19 11 34 10 20 16 31 12 8 ...
## $ VeryActiveMinutes   : int  25 21 30 29 36 38 42 50 28 19 ...
## $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
## $ LightActiveDistance   : num  6.06 4.71 3.91 2.83 5.04 ...
## $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
## $ VeryActiveDistance    : num  1.88 1.57 2.44 2.14 2.71 ...

str(Daily_Steps)

## 'data.frame':      940 obs. of  3 variables:
## $ Id              : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDay: chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ StepTotal      : int  13162 10735 10460 9762 12669 9705 13019 15506 10544
9819 ...

str(Sleep_Day)

## 'data.frame':      413 obs. of  5 variables:
## $ Id              : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ SleepDay        : chr  "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00
AM" "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...
## $ TotalSleepRecords : int  1 2 1 2 1 1 1 1 1 1 ...
## $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
## $ TotalTimeInBed     : int  346 407 442 367 712 320 377 364 384 449 ...

str(Weight_Log)

## 'data.frame':      940 obs. of  10 variables:
## $ Id              : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09
...
## $ ActivityDay      : chr  "4/12/2016" "4/13/2016" "4/14/2016"

```

```
"4/15/2016" ...
## $ SedentaryMinutes      : int  728 776 1218 726 773 539 1149 775 818
838 ...
## $ LightlyActiveMinutes  : int  328 217 181 209 221 164 233 264 205 211
...
## $ FairlyActiveMinutes   : int  13 19 11 34 10 20 16 31 12 8 ...
## $ VeryActiveMinutes     : int  25 21 30 29 36 38 42 50 28 19 ...
## $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
## $ LightActiveDistance    : num  6.06 4.71 3.91 2.83 5.04 ...
## $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
## $ VeryActiveDistance     : num  1.88 1.57 2.44 2.14 2.71 ...
```

Any Missing value

```
which(is.na(Daily_Activity))
```

```
## integer(0)
```

```
which(is.na(Daily_Calories))
```

```
## integer(0)
```

```
which(is.na(Daily_Intensities))
```

```
## integer(0)
```

```
which(is.na(Daily_Steps))
```

```
## integer(0)
```

```
which(is.na(Sleep_Day))
```

```
## integer(0)
```

```
which(is.na(Weight_Log))
```

```
## integer(0)
```

Number of unique participants

```
n_distinct(Daily_Activity$Id)
```

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

```
n_distinct(Daily_Calories$Id)
```

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

```
n_distinct(Daily_Intensities$Id)
```

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

```
n_distinct(Daily_Steps$Id)
```

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

```
n_distinct(Sleep_Day$Id)
```

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

```
n_distinct(Weight_Log$Id)
```

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

4. Analyze

Summary of Data

```
summary(Daily_Activity)
```

```
##           Id           ActivityDate       TotalSteps     TotalDistance
##  Min.      :1.504e+09   Length:940         Min.       :    0   Min.       : 0.000
## 1st Qu.:2.320e+09   Class :character 1st Qu.: 3790   1st Qu.: 2.620
## Median :4.445e+09   Mode  :character Median : 7406   Median : 5.245
## Mean      :4.855e+09                Mean      : 7638   Mean      : 5.490
## 3rd Qu.:6.962e+09                3rd Qu.:10727   3rd Qu.: 7.713
## Max.      :8.878e+09                Max.       :36019   Max.      :28.030
## TrackerDistance  LoggedActivitiesDistance  VeryActiveDistance
##  Min.       : 0.000   Min.       :0.0000   Min.       : 0.000
## 1st Qu.: 2.620   1st Qu.:0.0000   1st Qu.: 0.000
## Median : 5.245   Median :0.0000   Median : 0.210
## Mean      : 5.475   Mean      :0.1082   Mean      : 1.503
## 3rd Qu.: 7.710   3rd Qu.:0.0000   3rd Qu.: 2.053
## Max.      :28.030   Max.      :4.9421   Max.      :21.920
## ModeratelyActiveDistance  LightActiveDistance  SedentaryActiveDistance
##  Min.       :0.0000   Min.       : 0.000   Min.       :0.000000
## 1st Qu.:0.0000   1st Qu.: 1.945   1st Qu.:0.000000
## Median :0.2400   Median : 3.365   Median :0.000000
## Mean      :0.5675   Mean      : 3.341   Mean      :0.001606
## 3rd Qu.:0.8000   3rd Qu.: 4.782   3rd Qu.:0.000000
## Max.      :6.4800   Max.      :10.710   Max.      :0.110000
## VeryActiveMinutes  FairlyActiveMinutes  LightlyActiveMinutes
## SedentaryMinutes
##  Min.       : 0.00   Min.       : 0.00   Min.       : 0.0   Min.       : 0.0
## 1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.:127.0   1st Qu.: 729.8
## Median : 4.00   Median : 6.00   Median :199.0   Median :1057.5
## Mean      :21.16   Mean      :13.56   Mean      :192.8   Mean      : 991.2
## 3rd Qu.:32.00   3rd Qu.:19.00   3rd Qu.:264.0   3rd Qu.:1229.5
## Max.      :210.00   Max.      :143.00   Max.      :518.0   Max.      :1440.0
##           Calories
##  Min.       :    0
## 1st Qu.:1828
## Median :2134
## Mean      :2304
## 3rd Qu.:2793
## Max.      :4900
```

```
summary(Daily_Calories)
```

```
##           Id           ActivityDay       Calories
##  Min.      :1.504e+09   Length:940         Min.       :    0
```

##	1st Qu.:2.320e+09	Class :character	1st Qu.:1828
##	Median :4.445e+09	Mode :character	Median :2134
##	Mean :4.855e+09		Mean :2304
##	3rd Qu.:6.962e+09		3rd Qu.:2793
##	Max. :8.878e+09		Max. :4900

summary(Daily_Intensities)

##	Id	ActivityDay	SedentaryMinutes	
LightlyActiveMinutes				
##	Min. :1.504e+09	Length:940	Min. : 0.0	Min. : 0.0
##	1st Qu.:2.320e+09	Class :character	1st Qu.: 729.8	1st Qu.:127.0
##	Median :4.445e+09	Mode :character	Median :1057.5	Median :199.0
##	Mean :4.855e+09		Mean : 991.2	Mean :192.8
##	3rd Qu.:6.962e+09		3rd Qu.:1229.5	3rd Qu.:264.0
##	Max. :8.878e+09		Max. :1440.0	Max. :518.0
##	FairlyActiveMinutes	VeryActiveMinutes	SedentaryActiveDistance	
##	Min. : 0.00	Min. : 0.00	Min. :0.000000	
##	1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.:0.000000	
##	Median : 6.00	Median : 4.00	Median :0.000000	
##	Mean : 13.56	Mean : 21.16	Mean :0.001606	
##	3rd Qu.: 19.00	3rd Qu.: 32.00	3rd Qu.:0.000000	
##	Max. :143.00	Max. :210.00	Max. :0.110000	
##	LightActiveDistance	ModeratelyActiveDistance	VeryActiveDistance	
##	Min. : 0.000	Min. :0.0000	Min. : 0.000	
##	1st Qu.: 1.945	1st Qu.:0.0000	1st Qu.: 0.000	
##	Median : 3.365	Median :0.2400	Median : 0.210	
##	Mean : 3.341	Mean :0.5675	Mean : 1.503	
##	3rd Qu.: 4.782	3rd Qu.:0.8000	3rd Qu.: 2.053	
##	Max. :10.710	Max. :6.4800	Max. :21.920	

summary(Daily_Steps)

##	Id	ActivityDay	StepTotal
##	Min. :1.504e+09	Length:940	Min. : 0
##	1st Qu.:2.320e+09	Class :character	1st Qu.: 3790
##	Median :4.445e+09	Mode :character	Median : 7406
##	Mean :4.855e+09		Mean : 7638
##	3rd Qu.:6.962e+09		3rd Qu.:10727
##	Max. :8.878e+09		Max. :36019

summary(Sleep_Day)

##	Id	SleepDay	TotalSleepRecords	
TotalMinutesAsleep				
##	Min. :1.504e+09	Length:413	Min. :1.000	Min. : 58.0
##	1st Qu.:3.977e+09	Class :character	1st Qu.:1.000	1st Qu.:361.0
##	Median :4.703e+09	Mode :character	Median :1.000	Median :433.0
##	Mean :5.001e+09		Mean :1.119	Mean :419.5
##	3rd Qu.:6.962e+09		3rd Qu.:1.000	3rd Qu.:490.0
##	Max. :8.792e+09		Max. :3.000	Max. :796.0
##	TotalTimeInBed			

```
## Min. : 61.0
## 1st Qu.:403.0
## Median :463.0
## Mean :458.6
## 3rd Qu.:526.0
## Max. :961.0
```

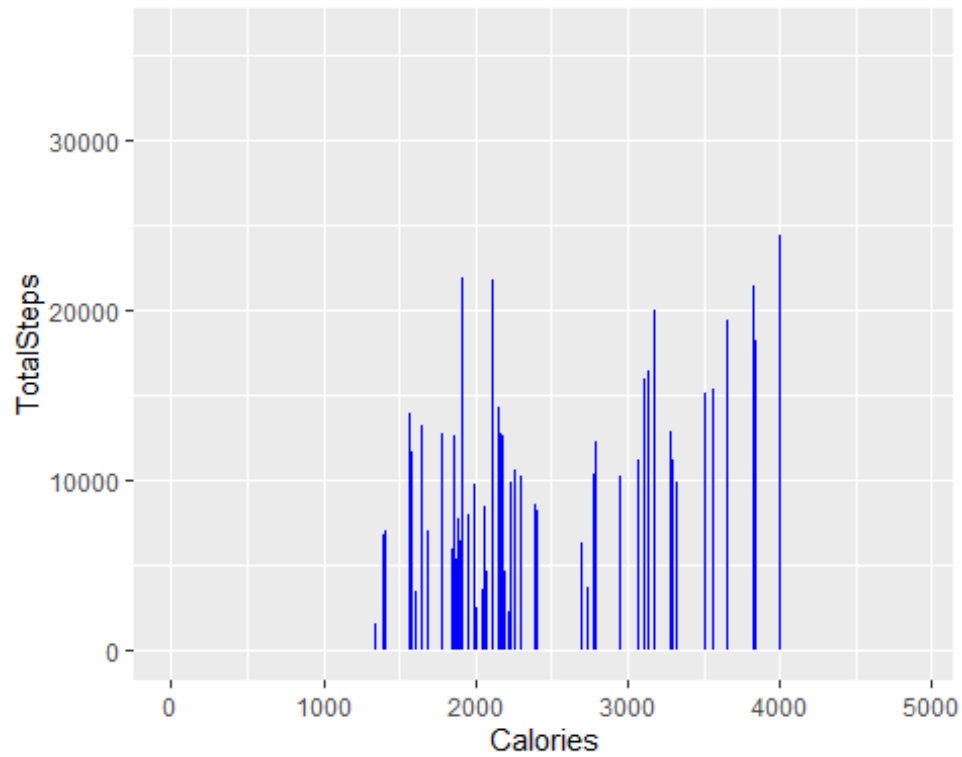
`summary(Weight_Log)`

```
##      Id      ActivityDay      SedentaryMinutes
LightlyActiveMinutes
## Min. :1.504e+09 Length:940      Min. : 0.0      Min. : 0.0
## 1st Qu.:2.320e+09 Class :character 1st Qu.: 729.8 1st Qu.:127.0
## Median :4.445e+09 Mode  :character Median :1057.5 Median :199.0
## Mean :4.855e+09      Mean : 991.2 Mean :192.8
## 3rd Qu.:6.962e+09      3rd Qu.:1229.5 3rd Qu.:264.0
## Max. :8.878e+09      Max. :1440.0 Max. :518.0
## FairlyActiveMinutes VeryActiveMinutes SedentaryActiveDistance
## Min. : 0.00      Min. : 0.00      Min. :0.000000
## 1st Qu.: 0.00      1st Qu.: 0.00      1st Qu.:0.000000
## Median : 6.00      Median : 4.00      Median :0.000000
## Mean : 13.56      Mean : 21.16      Mean :0.001606
## 3rd Qu.: 19.00      3rd Qu.: 32.00      3rd Qu.:0.000000
## Max. :143.00      Max. :210.00      Max. :0.110000
## LightActiveDistance ModeratelyActiveDistance VeryActiveDistance
## Min. : 0.000      Min. :0.0000      Min. : 0.000
## 1st Qu.: 1.945      1st Qu.:0.0000      1st Qu.: 0.000
## Median : 3.365      Median :0.2400      Median : 0.210
## Mean : 3.341      Mean :0.5675      Mean : 1.503
## 3rd Qu.: 4.782      3rd Qu.:0.8000      3rd Qu.: 2.053
## Max. :10.710      Max. :6.4800      Max. :21.920
```

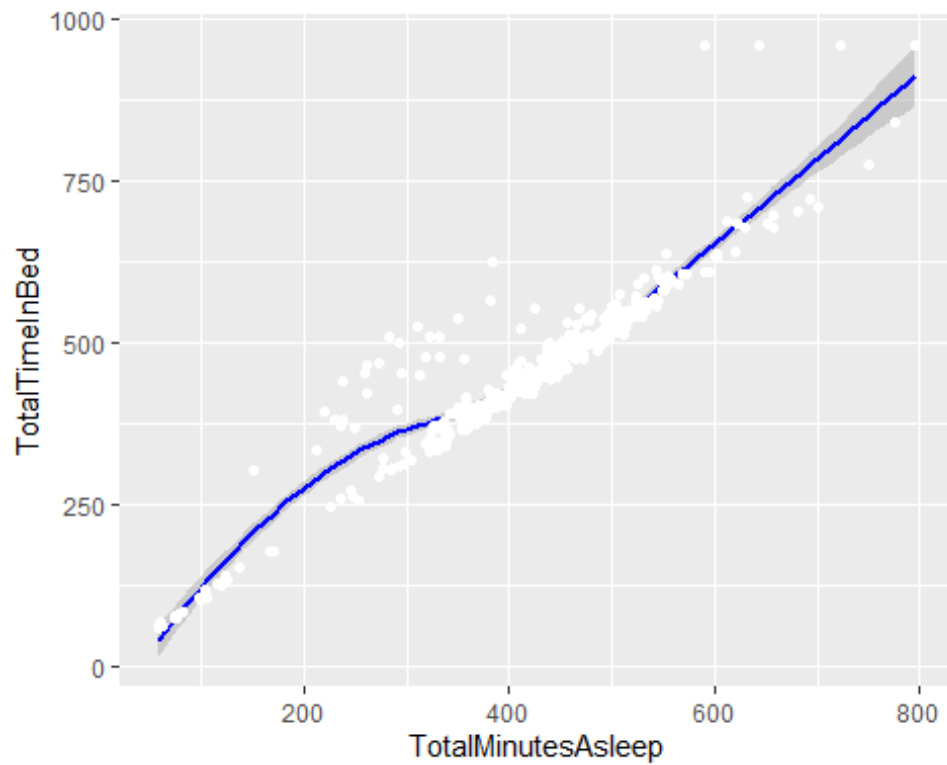
Visualization

Finding relationship between steps and calories from Daily_Activity data with Graphs

```
ggplot(data = Daily_Activity) +
  geom_col(mapping = aes(x = Calories, y = TotalSteps), fill = "blue")
```

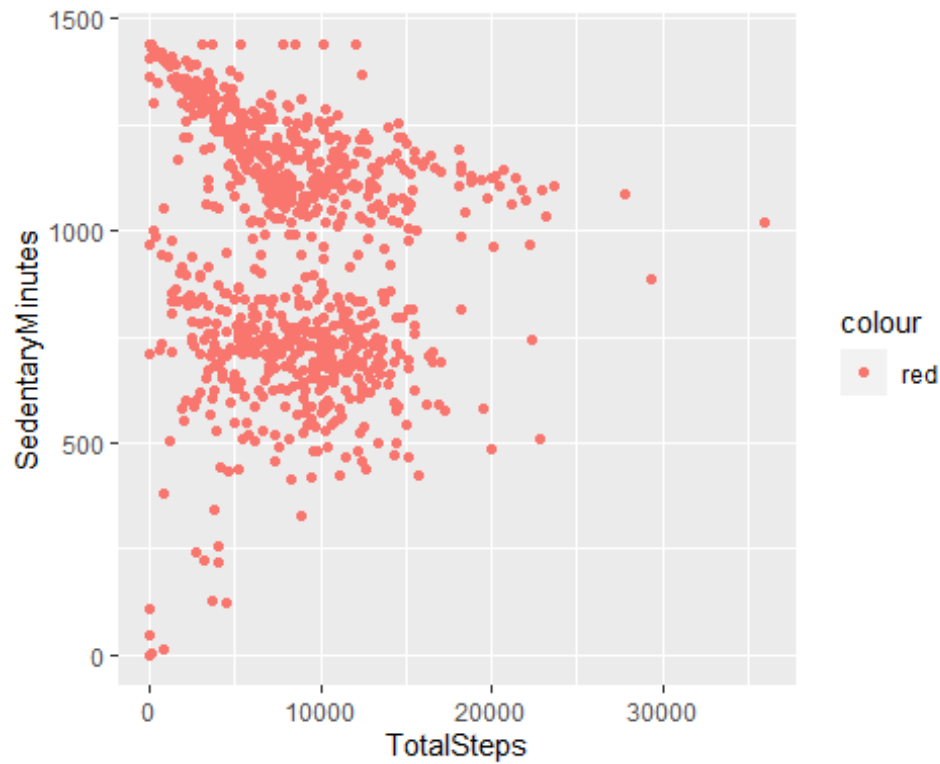


```
ggplot(data = Sleep_Day) + geom_smooth(mapping= aes(x=TotalMinutesAsleep,
y=TotalTimeInBed),color = "blue") + geom_point(mapping=
aes(x=TotalMinutesAsleep, y=TotalTimeInBed),color= "white")
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



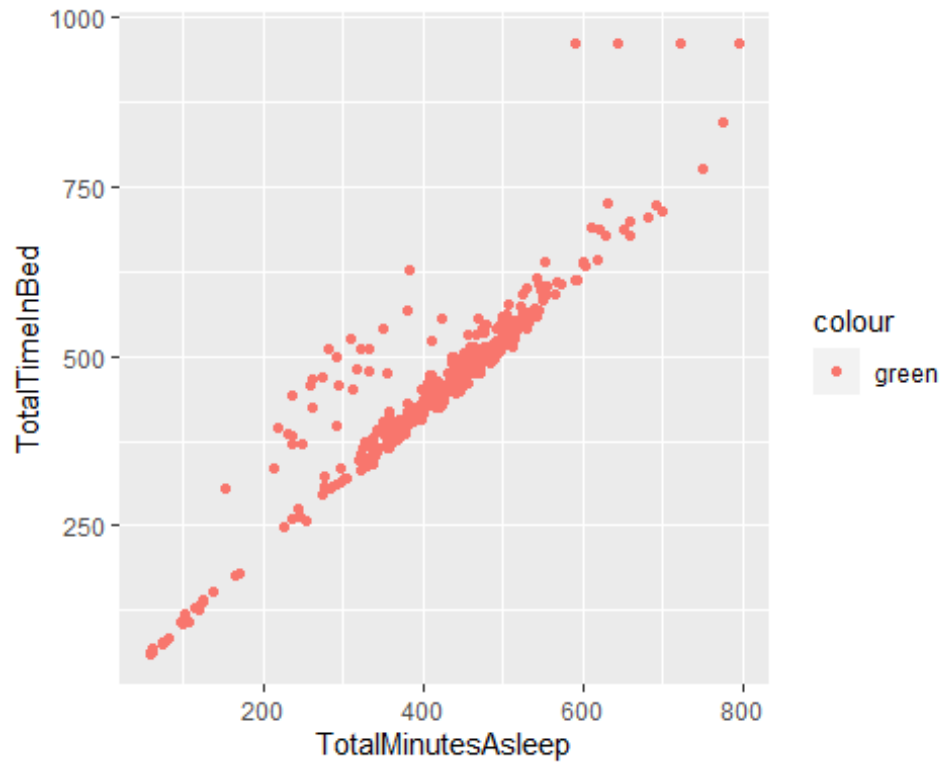
Relationship between steps and sedentary minutes in Daily_Activity with Graphs

```
ggplot(data = Daily_Activity, aes(x=TotalSteps, y=SedentaryMinutes, colour = "red")) + geom_point()
```

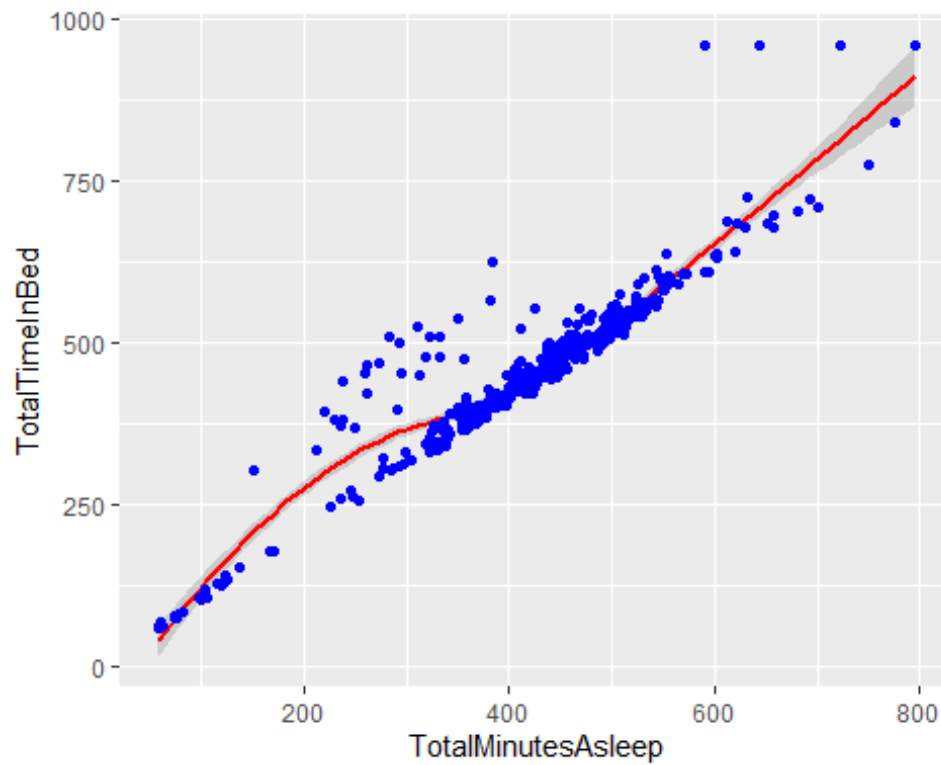


Finding relationship between total time in bed and total minutes asleep from Sleep_Day Data

```
ggplot(data = Sleep_Day, aes(x = TotalMinutesAsleep, y = TotalTimeInBed, color = "green")) + geom_point()
```

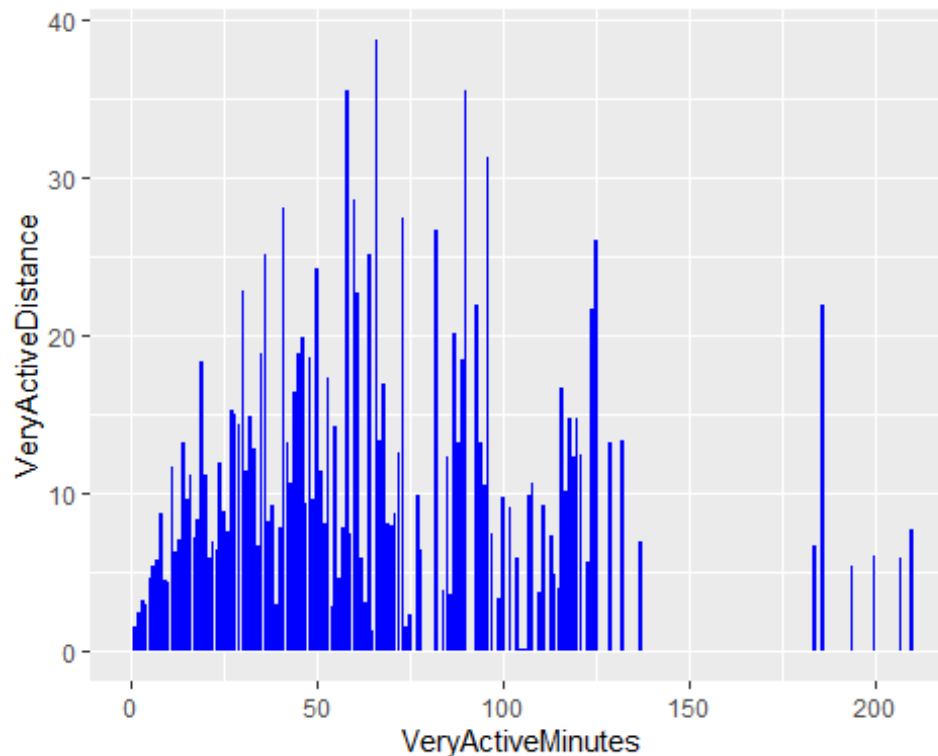


```
ggplot(data = Sleep_Day) + geom_smooth(mapping= aes(x=TotalMinutesAsleep,
y=TotalTimeInBed),color = "red") + geom_point(mapping=
aes(x=TotalMinutesAsleep, y=TotalTimeInBed),color= "blue")
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Finding relationship between Very Active Minutes and Very Active Distance from Weight_Log data with Graphs

```
ggplot(data = Weight_Log) +  
  geom_col(mapping = aes(x = VeryActiveMinutes, y = VeryActiveDistance), fill  
= "blue")
```



Merging Daily Activity and Sleep Day dataset

```
DailyActivity_SleepDay <- merge(Daily_Activity, Sleep_Day, all = TRUE)
```

```
view(DailyActivity_SleepDay)
```

```
colnames(DailyActivity_SleepDay)
```

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

Number of unique participants in combined dataset

```
n_distinct(DailyActivity_SleepDay$Id)
```

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

User activities in minutes through Pie Chart

```
Sedentary <- sum(DailyActivity_SleepDay$SedentaryMinutes)
Active <- sum(DailyActivity_SleepDay$VeryActiveMinutes)
Fairly <- sum(DailyActivity_SleepDay$FairlyActiveMinutes)
Lightly <- sum(DailyActivity_SleepDay$LightlyActiveMinutes)

x <- c(Sedentary,Lightly,Fairly,Active)
labels <- c("Sedentary", "Lightly Active", "Fairly Active", "Very Active")

piepercent <- round(100*x/sum(x),1)

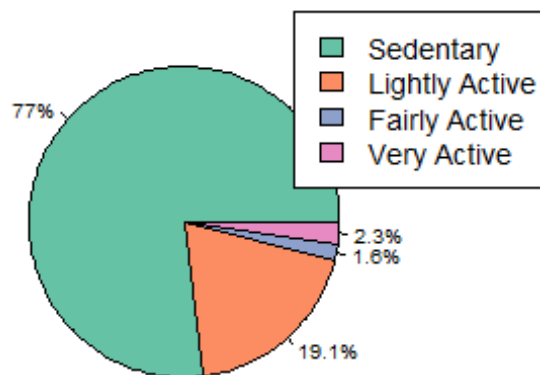
labels <- paste0(piepercent,"%")

colour = brewer.pal(length(piepercent), "Set2")

pie(x,labels = labels, main = "Active Minutes Pie Chart", col = colour, cex =
0.65)

legend("topright",c("Sedentary","Lightly Active", "Fairly Active", "Very
Active"), cex = 0.9, fill = colour)
```

Active Minutes Pie Chart



Add a new column in the combined dataset converting 'char' into 'Date'.

```
DailyActivity_SleepDay$ActivityDate2 <-
mdy(DailyActivity_SleepDay$ActivityDate)

colnames(DailyActivity_SleepDay)
```

```
## [1] "Id" "ActivityDate"
## [3] "TotalSteps" "TotalDistance"
## [5] "TrackerDistance" "LoggedActivitiesDistance"
## [7] "VeryActiveDistance" "ModeratelyActiveDistance"
## [9] "LightActiveDistance" "SedentaryActiveDistance"
## [11] "VeryActiveMinutes" "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes" "SedentaryMinutes"
## [15] "Calories" "SleepDay"
## [17] "TotalSleepRecords" "TotalMinutesAsleep"
## [19] "TotalTimeInBed" "ActivityDate2"
```

convert activity date into days of week.

```
DailyActivity_SleepDay$weekday <- wday(DailyActivity_SleepDay$ActivityDate2,
label = TRUE)
```

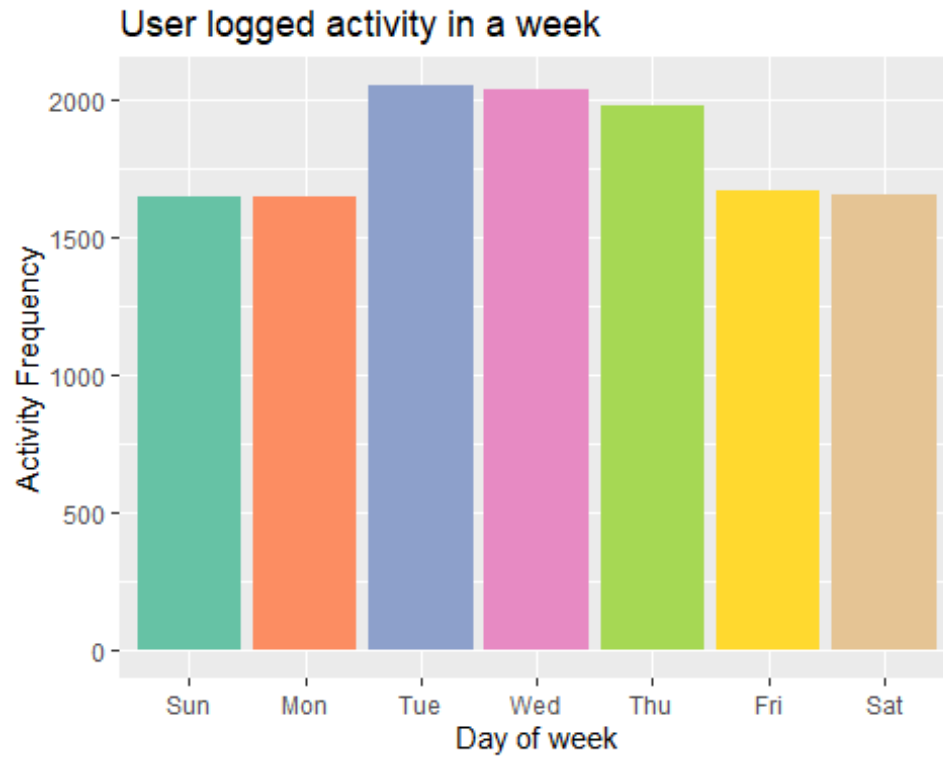
```
colnames(DailyActivity_SleepDay)
```

```
## [1] "Id" "ActivityDate"
## [3] "TotalSteps" "TotalDistance"
## [5] "TrackerDistance" "LoggedActivitiesDistance"
## [7] "VeryActiveDistance" "ModeratelyActiveDistance"
## [9] "LightActiveDistance" "SedentaryActiveDistance"
## [11] "VeryActiveMinutes" "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes" "SedentaryMinutes"
## [15] "Calories" "SleepDay"
## [17] "TotalSleepRecords" "TotalMinutesAsleep"
## [19] "TotalTimeInBed" "ActivityDate2"
## [21] "weekday"
```

```
head(DailyActivity_SleepDay$weekday)
```

```
## [1] Sat Sat Sat Sat Sat Sat
## Levels: Sun < Mon < Tue < Wed < Thu < Fri < Sat
```

```
ggplot(data = DailyActivity_SleepDay, aes(x = weekday)) + geom_bar(fill =
brewer.pal(7, "Set2")) + labs(title = "User logged activity in a week", x =
"Day of week", y = "Activity Frequency")
```



5. Share

1. Positive correlation between total steps and calories burnt.
2. Negative correlation between total steps and sedentary minutes.
3. Positive correlation between total minutes in bed and total minutes asleep.
4. Daily activity high in mid of the week.