Bellabeat Case Study

Zane Urbane

2024-06-22

Data Preparation

1. Load the packages

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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 errors
```

2. Load the data

```
setwd("/cloud/project/input_bellabeat")

# Daily data
daily_activity <- read.csv("/cloud/project/input_bellabeat/dailyActivity_merged.csv")
daily_sleep <- read.csv("sleepDay_merged.csv")

# Hourly data
hourly_intensities <- read.csv("/cloud/project/input_bellabeat/hourlyIntensities_merged.csv")
hourly_steps <- read.csv("/cloud/project/input_bellabeat/hourlySteps_merged.csv")
hourly_calories <- read.csv("/cloud/project/input_bellabeat/hourlyCalories_merged.csv")

# Check for NA

sum(is.na(daily_activity))
```

```
## [1] 0
```

```
sum(is.na(daily_sleep))
```

```
## [1] 0
```

```
sum(is.na(hourly_intensities))
```

```
## [1] 0
```

```
sum(is.na(hourly_steps))
```

```
## [1] 0
```

```
sum(is.na(hourly_calories))
```

```
## [1] 0
```

```
# Check for duplicates

sum(duplicated(daily_activity))
```

```
## [1] 0
```

```
sum(duplicated(daily_sleep))
```

```
## [1] 3
```

```
sum(duplicated(hourly_intensities))
```

```
## [1] 0
```

```
sum(duplicated(hourly_steps))
```

```
## [1] 0
```

```
sum(duplicated(hourly_calories))
```

```
## [1] 0
```

```
# Manage the duplicates in daily_sleep dataframe

daily_sleep <- daily_sleep[!duplicated(daily_sleep), ]

sum(duplicated(daily_sleep)) #check result
```

```
## [1] 0
```

3. Explore and clean the data

```
# View the daily data frames
# Check how many unique participants are there in each dataframe

n_distinct(daily_activity$Id)
```

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

```
n_distinct(daily_sleep$Id)
```

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

```
# There are more participants in daily activity than in daily sleep records.
# Overview of the data sets - daily_activity

head(daily_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
```

```
glimpse(daily_activity)
```

```
## 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, 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~
```

```
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
```

```
# Column ActivityDate is recognized as a character (chr) type, so it needs to be changed to a date. And
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate, format = "%m/%d/%Y")
daily_activity$Id <- as.character(daily_activity$Id)

# In this case columns about Distance is not revelant and won't be used in analysis.
daily_activity <- daily_activity[,-c(4:10)]

# Add a column with weekdays in daily_activity dataframe
daily_activity$Weekday <- weekdays(daily_activity$ActivityDate)

head(daily_activity) #checking
```

```
##           Id ActivityDate TotalSteps VeryActiveMinutes FairlyActiveMinutes
## 1 1503960366   2016-04-12      13162                25                  13
## 2 1503960366   2016-04-13      10735                21                  19
## 3 1503960366   2016-04-14      10460                30                  11
## 4 1503960366   2016-04-15       9762                29                  34
## 5 1503960366   2016-04-16      12669                36                  10
## 6 1503960366   2016-04-17       9705                38                  20
##   LightlyActiveMinutes SedentaryMinutes Calories   Weekday
## 1                  328              728     1985   Tuesday
## 2                  217              776     1797 Wednesday
## 3                  181             1218     1776  Thursday
## 4                  209              726     1745    Friday
## 5                  221              773     1863  Saturday
## 6                  164              539     1728    Sunday
```

```
# Overview of the data sets - daily_sleep

head(daily_sleep)
```

```
##           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
```

```
glimpse(daily_sleep)
```

```
## Rows: 410
## Columns: 5
## $ Id                 <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay           <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~
## $ 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~
```

```
summary(daily_sleep)
```

```
##        Id               SleepDay         TotalSleepRecords TotalMinutesAsleep
##  Min.   :1.504e+09   Length:410         Min.   :1.00      Min.   : 58.0
##  1st Qu.:3.977e+09   Class :character   1st Qu.:1.00      1st Qu.:361.0
##  Median :4.703e+09   Mode  :character   Median :1.00      Median :432.5
##  Mean   :4.995e+09                      Mean   :1.12      Mean   :419.2
##  3rd Qu.:6.962e+09                      3rd Qu.:1.00      3rd Qu.:490.0
##  Max.   :8.792e+09                      Max.   :3.00      Max.   :796.0
##  TotalTimeInBed
##  Min.   : 61.0
##  1st Qu.:403.8
##  Median :463.0
##  Mean   :458.5
##  3rd Qu.:526.0
##  Max.   :961.0
```

```
# Column SleepDay is recognized as a character (chr) type so it needs to be changed to date. And Id col
daily_sleep$SleepDay <- as.Date(daily_sleep$SleepDay, format = "%m/%d/%Y")
daily_sleep$Id <- as.character(daily_sleep$Id)

head(daily_sleep) #checking
```

```
##           Id   SleepDay TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 2016-04-12                 1                327            346
## 2 1503960366 2016-04-13                 2                384            407
## 3 1503960366 2016-04-15                 1                412            442
## 4 1503960366 2016-04-16                 2                340            367
## 5 1503960366 2016-04-17                 1                700            712
## 6 1503960366 2016-04-19                 1                304            320
```

```r
# Merge daily activities data with sleep data for analysis together.
# In daily activity data there are 940 rows, but in sleep data only 410 rows.
# I will use left join to keep all data and explore it later.

comb_daily_data <- left_join(daily_activity, daily_sleep, by= c("Id","ActivityDate" = "SleepDay"))

# Rename ActivityDate column as Date

colnames(comb_daily_data)[colnames(comb_daily_data) == "ActivityDate"] <- "Date"

head(comb_daily_data)
```

```
##           Id       Date TotalSteps VeryActiveMinutes FairlyActiveMinutes
## 1 1503960366 2016-04-12      13162                25                  13
## 2 1503960366 2016-04-13      10735                21                  19
## 3 1503960366 2016-04-14      10460                30                  11
## 4 1503960366 2016-04-15       9762                29                  34
## 5 1503960366 2016-04-16      12669                36                  10
## 6 1503960366 2016-04-17       9705                38                  20
##   LightlyActiveMinutes SedentaryMinutes Calories   Weekday TotalSleepRecords
## 1                  328              728     1985   Tuesday                 1
## 2                  217              776     1797 Wednesday                 2
## 3                  181             1218     1776  Thursday                NA
## 4                  209              726     1745    Friday                 1
## 5                  221              773     1863  Saturday                 2
## 6                  164              539     1728    Sunday                 1
##   TotalMinutesAsleep TotalTimeInBed
## 1                327            346
## 2                384            407
## 3                 NA             NA
## 4                412            442
## 5                340            367
## 6                700            712
```

```r
glimpse(comb_daily_data) #checking
```

```
## Rows: 940
## Columns: 12
## $ Id                   <chr> "1503960366", "1503960366", "1503960366", "150396~
## $ Date                 <date> 2016-04-12, 2016-04-13, 2016-04-14, 2016-04-15, ~
## $ TotalSteps           <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 15~
## $ VeryActiveMinutes    <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 41, 3~
## $ FairlyActiveMinutes  <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21, 5,~
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, 211,~
## $ SedentaryMinutes     <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818, 83~
## $ Calories             <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1~
## $ Weekday              <chr> "Tuesday", "Wednesday", "Thursday", "Friday", "Sa~
## $ TotalSleepRecords    <int> 1, 2, NA, 1, 2, 1, NA, 1, 1, 1, NA, 1, 1, 1, 1, N~
## $ TotalMinutesAsleep   <int> 327, 384, NA, 412, 340, 700, NA, 304, 360, 325, N~
## $ TotalTimeInBed       <int> 346, 407, NA, 442, 367, 712, NA, 320, 377, 364, N~
```

```r
# Check for NA, duplicates and count unique participants

sum(is.na(comb_daily_data))
```

```
## [1] 1590
```

```
sum(duplicated(comb_daily_data))
```

```
## [1] 0
```

```
n_distinct(comb_daily_data$Id)
```

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

```
# There is 2650 NA values.
# NA is only in merged data in sleep column. Replace it with zero for calculations.

comb_daily_data[is.na(comb_daily_data)] <- 0

head(comb_daily_data)
```

```
##           Id       Date TotalSteps VeryActiveMinutes FairlyActiveMinutes
## 1 1503960366 2016-04-12      13162                25                  13
## 2 1503960366 2016-04-13      10735                21                  19
## 3 1503960366 2016-04-14      10460                30                  11
## 4 1503960366 2016-04-15       9762                29                  34
## 5 1503960366 2016-04-16      12669                36                  10
## 6 1503960366 2016-04-17       9705                38                  20
##   LightlyActiveMinutes SedentaryMinutes Calories   Weekday TotalSleepRecords
## 1                  328              728     1985   Tuesday                 1
## 2                  217              776     1797 Wednesday                 2
## 3                  181             1218     1776  Thursday                 0
## 4                  209              726     1745    Friday                 1
## 5                  221              773     1863  Saturday                 2
## 6                  164              539     1728    Sunday                 1
##   TotalMinutesAsleep TotalTimeInBed
## 1                327            346
## 2                384            407
## 3                  0              0
## 4                412            442
## 5                340            367
## 6                700            712
```

```
glimpse(comb_daily_data) #checking
```

```
## Rows: 940
## Columns: 12
## $ Id                   <chr> "1503960366", "1503960366", "1503960366", "150396~
## $ Date                 <date> 2016-04-12, 2016-04-13, 2016-04-14, 2016-04-15, ~
## $ TotalSteps           <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 15~
## $ VeryActiveMinutes    <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 41, 3~
## $ FairlyActiveMinutes  <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21, 5,~
## $ LightlyActiveMinutes <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, 211,~
## $ SedentaryMinutes     <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818, 83~
## $ Calories             <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1~
## $ Weekday              <chr> "Tuesday", "Wednesday", "Thursday", "Friday", "Sa~
## $ TotalSleepRecords    <dbl> 1, 2, 0, 1, 2, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1~
## $ TotalMinutesAsleep   <dbl> 327, 384, 0, 412, 340, 700, 0, 304, 360, 325, 0, ~
## $ TotalTimeInBed       <dbl> 346, 407, 0, 442, 367, 712, 0, 320, 377, 364, 0, ~
```

```
# daily_sleep columns changed data type to dbl, I will change it back to int.

comb_daily_data$TotalSleepRecords <- as.integer(comb_daily_data$TotalSleepRecords)
comb_daily_data$TotalMinutesAsleep <- as.integer(comb_daily_data$TotalMinutesAsleep)
comb_daily_data$TotalTimeInBed <- as.integer(comb_daily_data$TotalTimeInBed)

head(comb_daily_data) #checking
```

```
##           Id       Date TotalSteps VeryActiveMinutes FairlyActiveMinutes
## 1 1503960366 2016-04-12      13162                25                  13
## 2 1503960366 2016-04-13      10735                21                  19
## 3 1503960366 2016-04-14      10460                30                  11
## 4 1503960366 2016-04-15       9762                29                  34
## 5 1503960366 2016-04-16      12669                36                  10
## 6 1503960366 2016-04-17       9705                38                  20
##   LightlyActiveMinutes SedentaryMinutes Calories   Weekday TotalSleepRecords
## 1                  328              728     1985   Tuesday                 1
## 2                  217              776     1797 Wednesday                 2
## 3                  181             1218     1776  Thursday                 0
## 4                  209              726     1745    Friday                 1
## 5                  221              773     1863  Saturday                 2
## 6                  164              539     1728    Sunday                 1
##   TotalMinutesAsleep TotalTimeInBed
## 1                327            346
## 2                384            407
## 3                  0              0
## 4                412            442
## 5                340            367
## 6                700            712
```

```
# Check for NA, duplicates and count unique participants

sum(is.na(comb_daily_data))
```

```
## [1] 0
```

```
sum(duplicated(comb_daily_data))
```

```
## [1] 0
```

```
n_distinct(comb_daily_data$Id)
```

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

```
# View the hourly data frames
# Check how many unique participants are there in each dataframe

n_distinct(hourly_calories$Id)
```

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

```
n_distinct(hourly_intensities$Id)
```

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

```
n_distinct(hourly_steps$Id)
```

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

```r
# Overview of the data sets - hourly_calories

head(hourly_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
```

```r
glimpse(hourly_calories)
```

```
## Rows: 22,099
## Columns: 3
## $ Id           <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityHour <chr> "4/12/2016 12:00:00 AM", "4/12/2016 1:00:00 AM", "4/12/20~
## $ Calories     <int> 81, 61, 59, 47, 48, 48, 48, 47, 68, 141, 99, 76, 73, 66, ~
```

```r
summary(hourly_calories)
```

```
##        Id            ActivityHour          Calories
##  Min.   :1.504e+09   Length:22099       Min.   : 42.00
##  1st Qu.:2.320e+09   Class :character   1st Qu.: 63.00
##  Median :4.445e+09   Mode  :character   Median : 83.00
##  Mean   :4.848e+09                      Mean   : 97.39
##  3rd Qu.:6.962e+09                      3rd Qu.:108.00
##  Max.   :8.878e+09                      Max.   :948.00
```

```r
# Overview of the data sets - hourly_intensities

head(hourly_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
```

```r
glimpse(hourly_intensities)
```

```
## Rows: 22,099
## Columns: 4
## $ Id               <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 15039~
## $ ActivityHour     <chr> "4/12/2016 12:00:00 AM", "4/12/2016 1:00:00 AM", "4/1~
## $ TotalIntensity   <int> 20, 8, 7, 0, 0, 0, 0, 0, 13, 30, 29, 12, 11, 6, 36, 5~
## $ AverageIntensity <dbl> 0.333333, 0.133333, 0.116667, 0.000000, 0.000000, 0.0~
```

```r
summary(hourly_intensities)
```

```
##        Id            ActivityHour       TotalIntensity   AverageIntensity
##  Min.   :1.504e+09   Length:22099       Min.   :  0.00   Min.   :0.0000
##  1st Qu.:2.320e+09   Class :character   1st Qu.:  0.00   1st Qu.:0.0000
##  Median :4.445e+09   Mode  :character   Median :  3.00   Median :0.0500
```

```
##  Mean   :4.848e+09                     Mean   : 12.04   Mean   :0.2006
##  3rd Qu.:6.962e+09                     3rd Qu.: 16.00   3rd Qu.:0.2667
##  Max.   :8.878e+09                     Max.   :180.00   Max.   :3.0000
```

```
# Overview of the data sets - hourly_steps

head(hourly_steps)
```

```
##           Id          ActivityHour StepTotal
## 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
```

```
glimpse(hourly_steps)
```

```
## Rows: 22,099
## Columns: 3
## $ Id           <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityHour <chr> "4/12/2016 12:00:00 AM", "4/12/2016 1:00:00 AM", "4/12/20~
## $ StepTotal    <int> 373, 160, 151, 0, 0, 0, 0, 0, 250, 1864, 676, 360, 253, 2~
```

```
summary(hourly_steps)
```

```
##        Id            ActivityHour         StepTotal
##  Min.   :1.504e+09   Length:22099       Min.   :    0.0
##  1st Qu.:2.320e+09   Class :character   1st Qu.:    0.0
##  Median :4.445e+09   Mode  :character   Median :   40.0
##  Mean   :4.848e+09                      Mean   :  320.2
##  3rd Qu.:6.962e+09                      3rd Qu.:  357.0
##  Max.   :8.878e+09                      Max.   :10554.0
```

```
# Column ActivityHour is recognized as a character (chr) type, so it needs to be changed to a date in e
hourly_calories$ActivityHour <- as.POSIXct(hourly_calories$ActivityHour, format = "%m/%d/%Y %I:%M:%S %p
hourly_calories$Id <- as.character(hourly_calories$Id)
hourly_intensities$ActivityHour <- as.POSIXct(hourly_intensities$ActivityHour, format = "%m/%d/%Y %I:%M
hourly_intensities$Id <- as.character(hourly_intensities$Id)
hourly_steps$ActivityHour <- as.POSIXct(hourly_steps$ActivityHour, format = "%m/%d/%Y %I:%M:%S %p")
hourly_steps$Id <- as.character(hourly_steps$Id)

head(hourly_calories) #checking
```

```
##           Id        ActivityHour Calories
## 1 1503960366 2016-04-12 00:00:00       81
## 2 1503960366 2016-04-12 01:00:00       61
## 3 1503960366 2016-04-12 02:00:00       59
## 4 1503960366 2016-04-12 03:00:00       47
## 5 1503960366 2016-04-12 04:00:00       48
## 6 1503960366 2016-04-12 05:00:00       48
```

```
head(hourly_intensities)
```

```
##           Id        ActivityHour TotalIntensity AverageIntensity
## 1 1503960366 2016-04-12 00:00:00             20         0.333333
## 2 1503960366 2016-04-12 01:00:00              8         0.133333
## 3 1503960366 2016-04-12 02:00:00              7         0.116667
```

```
## 4 1503960366 2016-04-12 03:00:00              0         0.000000
## 5 1503960366 2016-04-12 04:00:00              0         0.000000
## 6 1503960366 2016-04-12 05:00:00              0         0.000000
```

```
head(hourly_steps)
```

```
##           Id        ActivityHour StepTotal
## 1 1503960366 2016-04-12 00:00:00       373
## 2 1503960366 2016-04-12 01:00:00       160
## 3 1503960366 2016-04-12 02:00:00       151
## 4 1503960366 2016-04-12 03:00:00         0
## 5 1503960366 2016-04-12 04:00:00         0
## 6 1503960366 2016-04-12 05:00:00         0
```

```
summary(hourly_calories)
```

```
##       Id             ActivityHour                     Calories
##  Length:22099       Min.   :2016-04-12 00:00:00.00   Min.   : 42.00
##  Class :character   1st Qu.:2016-04-19 01:00:00.00   1st Qu.: 63.00
##  Mode  :character   Median :2016-04-26 06:00:00.00   Median : 83.00
##                     Mean   :2016-04-26 11:46:42.58   Mean   : 97.39
##                     3rd Qu.:2016-05-03 19:00:00.00   3rd Qu.:108.00
##                     Max.   :2016-05-12 15:00:00.00   Max.   :948.00
```

```
# Merge hourly data together. In all 3 data sets ar 22099 rows, and columns: Id and ActivityHour

merge_1 <- merge(hourly_calories, hourly_intensities, by = c("Id", "ActivityHour"))
comb_hourly_data <- merge(merge_1, hourly_steps, by = c("Id", "ActivityHour"))

# Rename ActivityDate column as Date

colnames(comb_hourly_data)[colnames(comb_hourly_data) == "ActivityDate"] <- "Date"

head(comb_hourly_data)
```

```
##           Id        ActivityHour Calories TotalIntensity AverageIntensity
## 1 1503960366 2016-04-12 00:00:00       81             20         0.333333
## 2 1503960366 2016-04-12 01:00:00       61              8         0.133333
## 3 1503960366 2016-04-12 02:00:00       59              7         0.116667
## 4 1503960366 2016-04-12 03:00:00       47              0         0.000000
## 5 1503960366 2016-04-12 04:00:00       48              0         0.000000
## 6 1503960366 2016-04-12 05:00:00       48              0         0.000000
##   StepTotal
## 1       373
## 2       160
## 3       151
## 4         0
## 5         0
## 6         0
```

```
glimpse(comb_hourly_data)
```

```
## Rows: 22,099
## Columns: 6
## $ Id              <chr> "1503960366", "1503960366", "1503960366", "1503960366~
## $ ActivityHour    <dttm> 2016-04-12 00:00:00, 2016-04-12 01:00:00, 2016-04-12~
## $ Calories        <int> 81, 61, 59, 47, 48, 48, 48, 47, 68, 141, 99, 76, 73, ~
```

```
## $ TotalIntensity   <int> 20, 8, 7, 0, 0, 0, 0, 0, 13, 30, 29, 12, 11, 6, 36, 5~
## $ AverageIntensity <dbl> 0.333333, 0.133333, 0.116667, 0.000000, 0.000000, 0.0~
## $ StepTotal        <int> 373, 160, 151, 0, 0, 0, 0, 0, 250, 1864, 676, 360, 25~
```

```
summary(comb_hourly_data) #checking
```

```
##       Id              ActivityHour                      Calories
##  Length:22099       Min.   :2016-04-12 00:00:00.00   Min.   : 42.00
##  Class :character   1st Qu.:2016-04-19 01:00:00.00   1st Qu.: 63.00
##  Mode  :character   Median :2016-04-26 06:00:00.00   Median : 83.00
##                     Mean   :2016-04-26 11:46:42.58   Mean   : 97.39
##                     3rd Qu.:2016-05-03 19:00:00.00   3rd Qu.:108.00
##                     Max.   :2016-05-12 15:00:00.00   Max.   :948.00
##  TotalIntensity   AverageIntensity   StepTotal
##  Min.   :  0.00   Min.   :0.0000   Min.   :    0.0
##  1st Qu.:  0.00   1st Qu.:0.0000   1st Qu.:    0.0
##  Median :  3.00   Median :0.0500   Median :   40.0
##  Mean   : 12.04   Mean   :0.2006   Mean   :  320.2
##  3rd Qu.: 16.00   3rd Qu.:0.2667   3rd Qu.:  357.0
##  Max.   :180.00   Max.   :3.0000   Max.   :10554.0
```

```
# Check for NA, duplicates and count unique participiants

sum(is.na(comb_hourly_data))
```

```
## [1] 0
```

```
sum(duplicated(comb_hourly_data))
```

```
## [1] 0
```

```
n_distinct(comb_hourly_data$Id)
```

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

```
# Add a column with weekdays in comb_hourly_data dataframe
comb_hourly_data$Weekday <- weekdays(comb_hourly_data$ActivityHour)

head(comb_hourly_data) #checking
```

```
##           Id        ActivityHour Calories TotalIntensity AverageIntensity
## 1 1503960366 2016-04-12 00:00:00       81              20         0.333333
## 2 1503960366 2016-04-12 01:00:00       61               8         0.133333
## 3 1503960366 2016-04-12 02:00:00       59               7         0.116667
## 4 1503960366 2016-04-12 03:00:00       47               0         0.000000
## 5 1503960366 2016-04-12 04:00:00       48               0         0.000000
## 6 1503960366 2016-04-12 05:00:00       48               0         0.000000
##   StepTotal Weekday
## 1       373 Tuesday
## 2       160 Tuesday
## 3       151 Tuesday
## 4         0 Tuesday
## 5         0 Tuesday
## 6         0 Tuesday
```

4.Explore daily merge data.

```
# I would like to understand how Fitbit users used the tracker.
# Did them used it all day long or only when excersising.
```

```r
# There is 1440 min in a day so I will check thas data in time columns sum up or not.

comb_daily_data <- mutate(comb_daily_data, TotalMinutes = VeryActiveMinutes + FairlyActiveMinutes + Ligh

head(comb_daily_data)
```

```
##           Id       Date TotalSteps VeryActiveMinutes FairlyActiveMinutes
## 1 1503960366 2016-04-12      13162                25                  13
## 2 1503960366 2016-04-13      10735                21                  19
## 3 1503960366 2016-04-14      10460                30                  11
## 4 1503960366 2016-04-15       9762                29                  34
## 5 1503960366 2016-04-16      12669                36                  10
## 6 1503960366 2016-04-17       9705                38                  20
##   LightlyActiveMinutes SedentaryMinutes Calories   Weekday TotalSleepRecords
## 1                  328              728     1985   Tuesday                 1
## 2                  217              776     1797 Wednesday                 2
## 3                  181             1218     1776  Thursday                 0
## 4                  209              726     1745    Friday                 1
## 5                  221              773     1863  Saturday                 2
## 6                  164              539     1728    Sunday                 1
##   TotalMinutesAsleep TotalTimeInBed TotalMinutes
## 1                327            346         1440
## 2                384            407         1440
## 3                  0              0         1440
## 4                412            442         1440
## 5                340            367         1407
## 6                700            712         1473
```

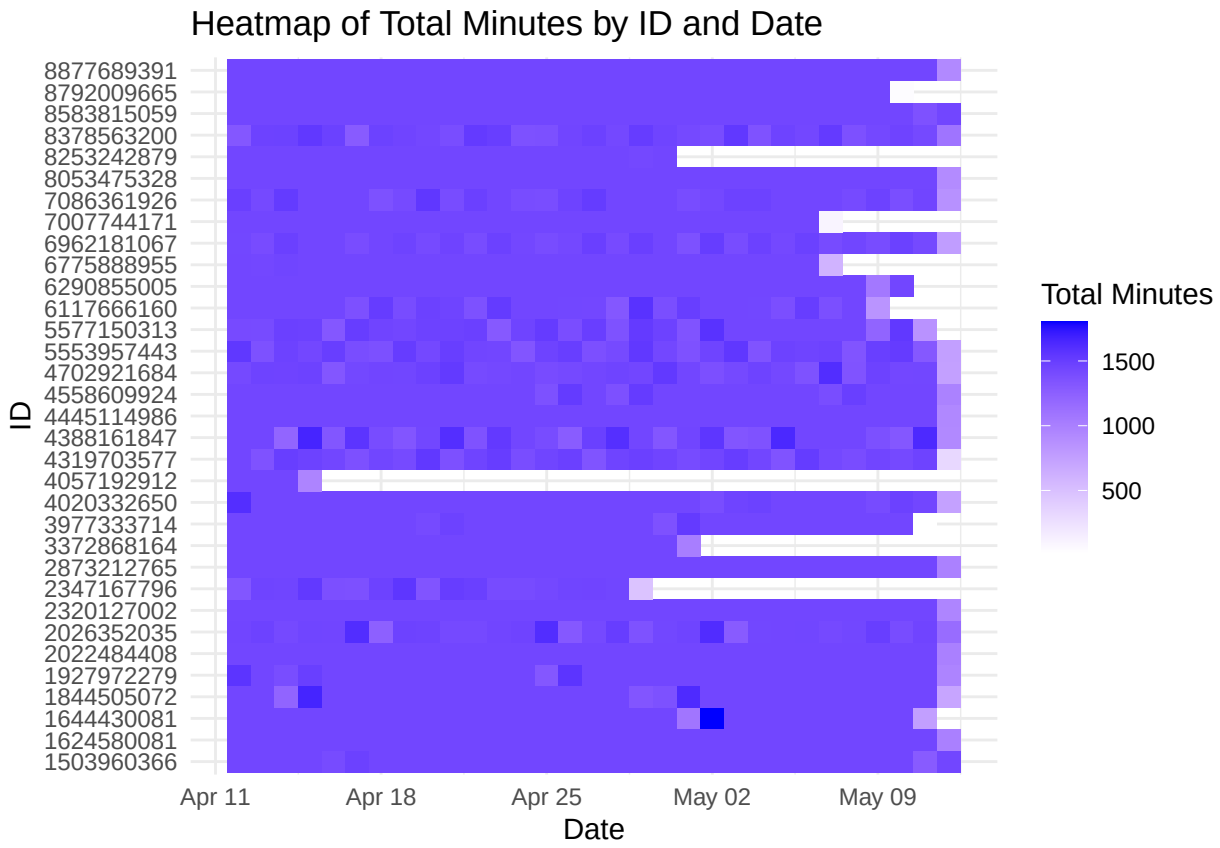
```r
summary(comb_daily_data$TotalMinutes)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      26    1440    1440    1419    1440    1799
```

```r
# Check precise how much of records ar 1440 min a day
count_1440_minutes <- sum(comb_daily_data$TotalMinutes == 1440)
print(count_1440_minutes)
```

```
## [1] 604
```

```r
# 604 of 940 records include all day.

ggplot(comb_daily_data, aes(x = Date, y = Id, fill = TotalMinutes)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Heatmap of Total Minutes by ID and Date",
       x = "Date", y = "ID", fill = "Total Minutes") +
  theme_minimal()
```




```
# There are users that have not used Fitbit tracker for all the period.
# It looks like that some minuts don't go to right date.
# That could be because the sleep data is daily and could refer to 2 dates.
# From heat map looks like last date could be incompleat.
# Lets look at average minutes per day
Id_activity <- comb_daily_data %>%
  group_by(Id) %>%
  summarise(
    total_dates = n_distinct(Date),
    total_minutes = sum(TotalMinutes)
  ) %>%
  mutate(average_total_minutes = total_minutes / total_dates)%>%
  arrange(desc(average_total_minutes))

print(Id_activity)
```

```
## # A tibble: 33 x 4
##    Id         total_dates total_minutes average_total_minutes
##    <chr>            <int>         <int>                 <dbl>
##  1 8253242879          19         27343                 1439.
##  2 8583815059          31         44559                 1437.
##  3 1503960366          31         44485                 1435
##  4 2026352035          31         44393                 1432.
##  5 1927972279          31         44290                 1429.
##  6 8378563200          31         44240                 1427.
##  7 6290855005          29         41384                 1427.
##  8 1624580081          31         44197                 1426.
```

```
##  9 2022484408          31         44196                 1426.
## 10 2873212765          31         44188                 1425.
## # i 23 more rows
```

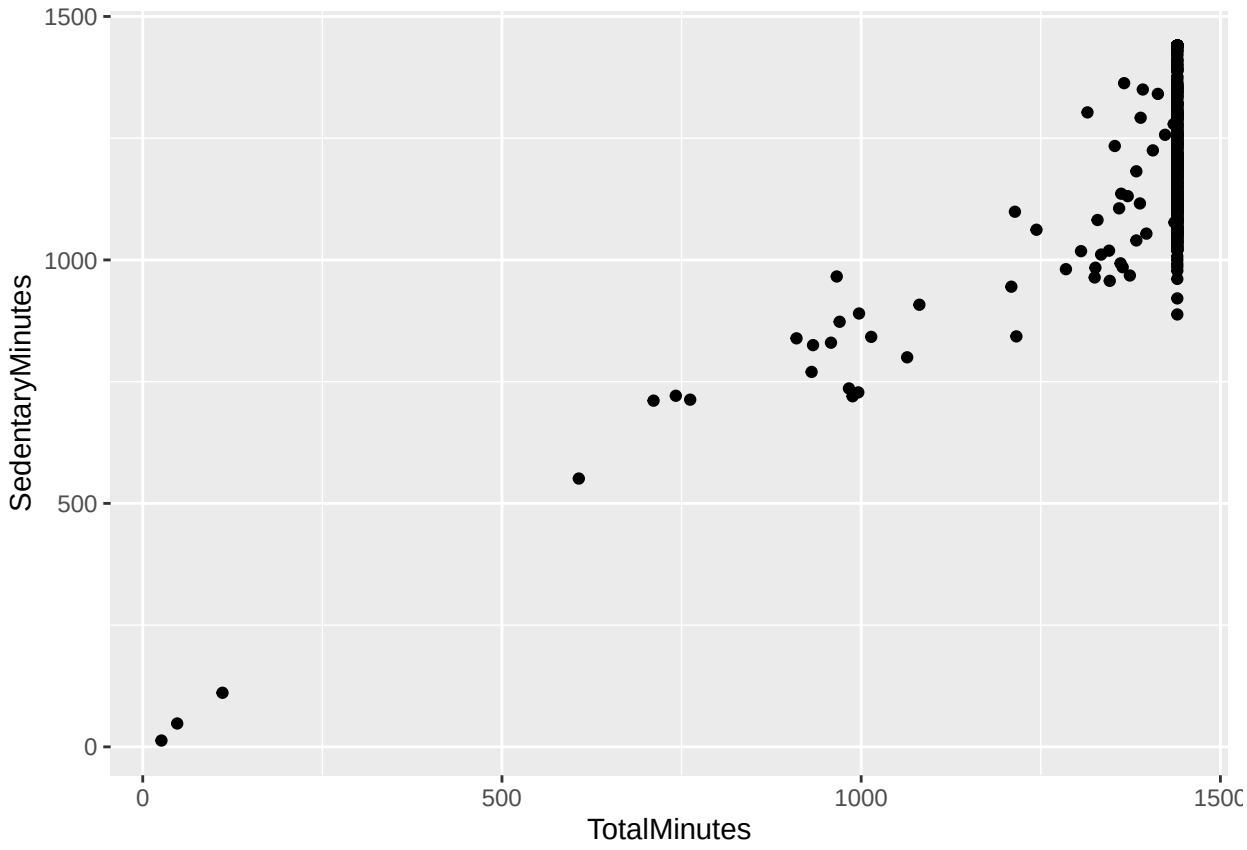
```
summary(Id_activity)
```

```
##       Id             total_dates    total_minutes   average_total_minutes
##  Length:33          Min.   : 4.00   Min.   : 5290   Min.   :1322
##  Class :character   1st Qu.:29.00   1st Qu.:41384   1st Qu.:1416
##  Mode  :character   Median :31.00   Median :43922   Median :1423
##                     Mean   :28.48   Mean   :40412   Mean   :1416
##                     3rd Qu.:31.00   3rd Qu.:44183   3rd Qu.:1426
##                     Max.   :31.00   Max.   :44559   Max.   :1439
```

```
# Average minutes for day is from 1322 to 1439 so sleeping time accuracy could be the cases.

# How many days partipics were involved
id_count <- Id_activity %>%
  group_by(total_dates) %>%
  summarise(id_count = n_distinct(Id))
print(id_count)
```

```
## # A tibble: 9 x 2
##   total_dates id_count
##         <int>    <int>
## 1           4        1
## 2          18        1
## 3          19        1
## 4          20        1
## 5          26        2
## 6          28        1
## 7          29        2
## 8          30        3
## 9          31       21
```

```
# And I want to see how looks those records, that don't have daily sleep records.

comb_data_no_sleep <- comb_daily_data[comb_daily_data$TotalSleepRecords == 0,]
ggplot(data=comb_data_no_sleep, aes(x=TotalMinutes, y=SedentaryMinutes)) + geom_point()
```




```
count_1440_minutes_no_sleep <- sum(comb_data_no_sleep$TotalMinutes == 1440)
print(count_1440_minutes_no_sleep)
```

```
## [1] 478
```

```
# 478 from 530 records are 1440 min whitout sleep time.
# Looks like sleeping time is added to SedentaryMinutes.
# Maybe users needed to add it manually.
```

```
#For daily activities analysis I will use only data that includes sleeping data.
daily_analysis_data <- comb_daily_data %>%
  filter(TotalSleepRecords > 0)
#Add colomn with some activity time
daily_analysis_data <- mutate(daily_analysis_data, SomeActivity = VeryActiveMinutes + FairlyActiveMinut
#Add colomn with laying in bed time
daily_analysis_data <- mutate(daily_analysis_data, LayingMinutes = TotalTimeInBed - TotalMinutesAsleep)

head(daily_analysis_data)
```

```
##           Id       Date TotalSteps VeryActiveMinutes FairlyActiveMinutes
## 1 1503960366 2016-04-12      13162                25                   13
## 2 1503960366 2016-04-13      10735                21                   19
## 3 1503960366 2016-04-15       9762                29                   34
## 4 1503960366 2016-04-16      12669                36                   10
## 5 1503960366 2016-04-17       9705                38                   20
## 6 1503960366 2016-04-19      15506                50                   31
##   LightlyActiveMinutes SedentaryMinutes Calories   Weekday TotalSleepRecords
## 1                  328              728     1985   Tuesday                 1
```

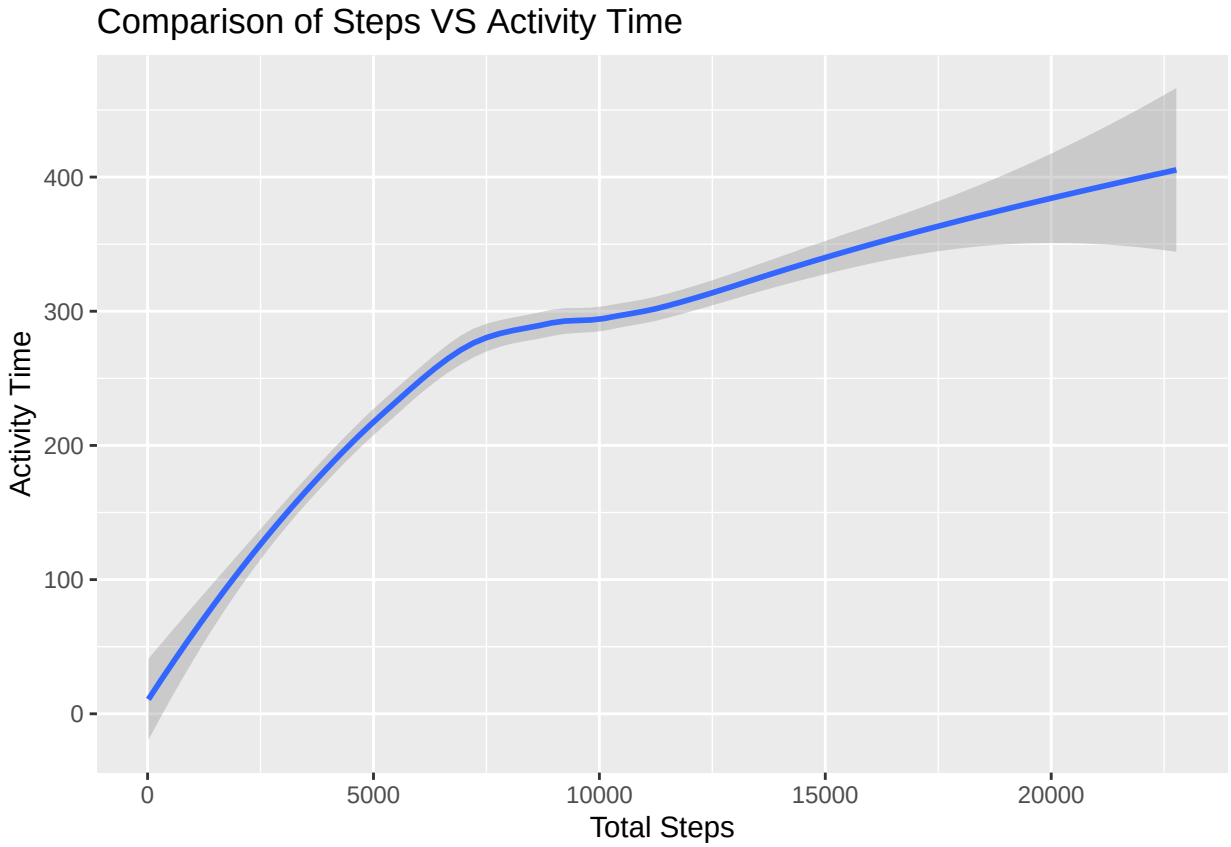
```
## 2                  217              776     1797 Wednesday                 2
## 3                  209              726     1745    Friday                 1
## 4                  221              773     1863  Saturday                 2
## 5                  164              539     1728    Sunday                 1
## 6                  264              775     2035   Tuesday                 1
##   TotalMinutesAsleep TotalTimeInBed TotalMinutes SomeActivity LayingMinutes
## 1                327            346         1440          366            19
## 2                384            407         1440          257            23
## 3                412            442         1440          272            30
## 4                340            367         1407          267            27
## 5                700            712         1473          222            12
## 6                304            320         1440          345            16
```

```
summary(daily_analysis_data)
```

```
##       Id                 Date              TotalSteps    VeryActiveMinutes
##  Length:410         Min.   :2016-04-12   Min.   :   17   Min.   :  0.00
##  Class :character   1st Qu.:2016-04-19   1st Qu.: 5189   1st Qu.:  0.00
##  Mode  :character   Median :2016-04-27   Median : 8913   Median :  9.00
##                     Mean   :2016-04-26   Mean   : 8515   Mean   : 25.05
##                     3rd Qu.:2016-05-04   3rd Qu.:11370   3rd Qu.: 38.00
##                     Max.   :2016-05-12   Max.   :22770   Max.   :210.00
##  FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes    Calories
##  Min.   :  0.00      Min.   :  2.0        Min.   :   0.0   Min.   : 257
##  1st Qu.:  0.00      1st Qu.:158.0        1st Qu.: 631.2   1st Qu.:1841
##  Median : 11.00      Median :208.0        Median : 717.0   Median :2207
##  Mean   : 17.92      Mean   :216.5        Mean   : 712.1   Mean   :2389
##  3rd Qu.: 26.75      3rd Qu.:263.0        3rd Qu.: 782.8   3rd Qu.:2920
##  Max.   :143.00      Max.   :518.0        Max.   :1265.0   Max.   :4900
##    Weekday          TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Length:410         Min.   :1.00      Min.   : 58.0      Min.   : 61.0
##  Class :character   1st Qu.:1.00      1st Qu.:361.0      1st Qu.:403.8
##  Mode  :character   Median :1.00      Median :432.5      Median :463.0
##                     Mean   :1.12      Mean   :419.2      Mean   :458.5
##                     3rd Qu.:1.00      3rd Qu.:490.0      3rd Qu.:526.0
##                     Max.   :3.00      Max.   :796.0      Max.   :961.0
##   TotalMinutes   SomeActivity   LayingMinutes
##  Min.   : 323   Min.   :  2.0   Min.   :  0.00
##  1st Qu.:1426   1st Qu.:206.5   1st Qu.: 17.00
##  Median :1440   Median :263.5   Median : 25.50
##  Mean   :1430   Mean   :259.5   Mean   : 39.31
##  3rd Qu.:1466   3rd Qu.:315.5   3rd Qu.: 40.00
##  Max.   :1799   Max.   :540.0   Max.   :371.00
```

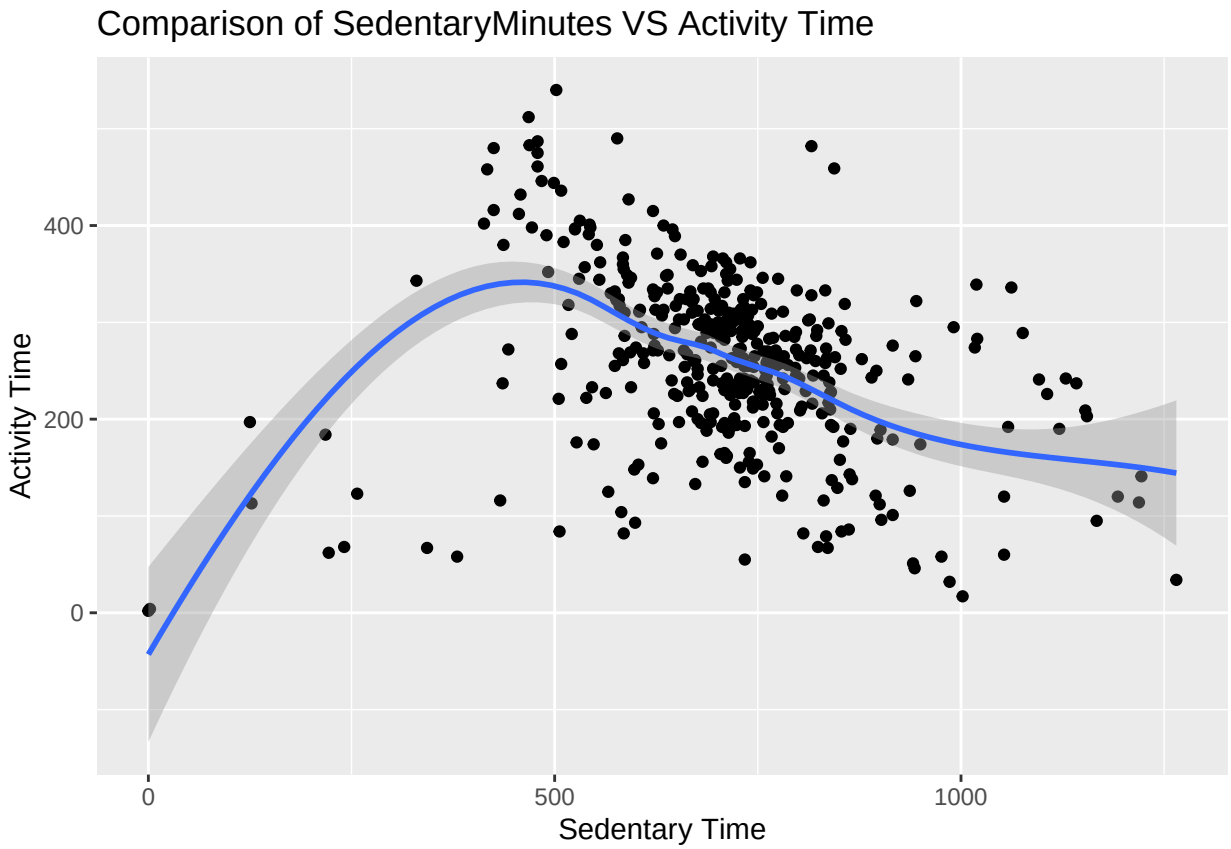
```
daily_analysis_data <- mutate(daily_analysis_data, SedetaryBedMinutes = LayingMinutes + SedentaryMinutes

#There is still days whit SedentaryMinutes 1265.0, that means sleeping time is less than 3 hours. And T

ggplot(data=daily_analysis_data, aes(x=TotalSteps, y=SomeActivity)) + geom_smooth(method = "loess") +
  labs(title = "Comparison of Steps VS Activity Time", x = "Total Steps", y = "Activity Time")
```

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

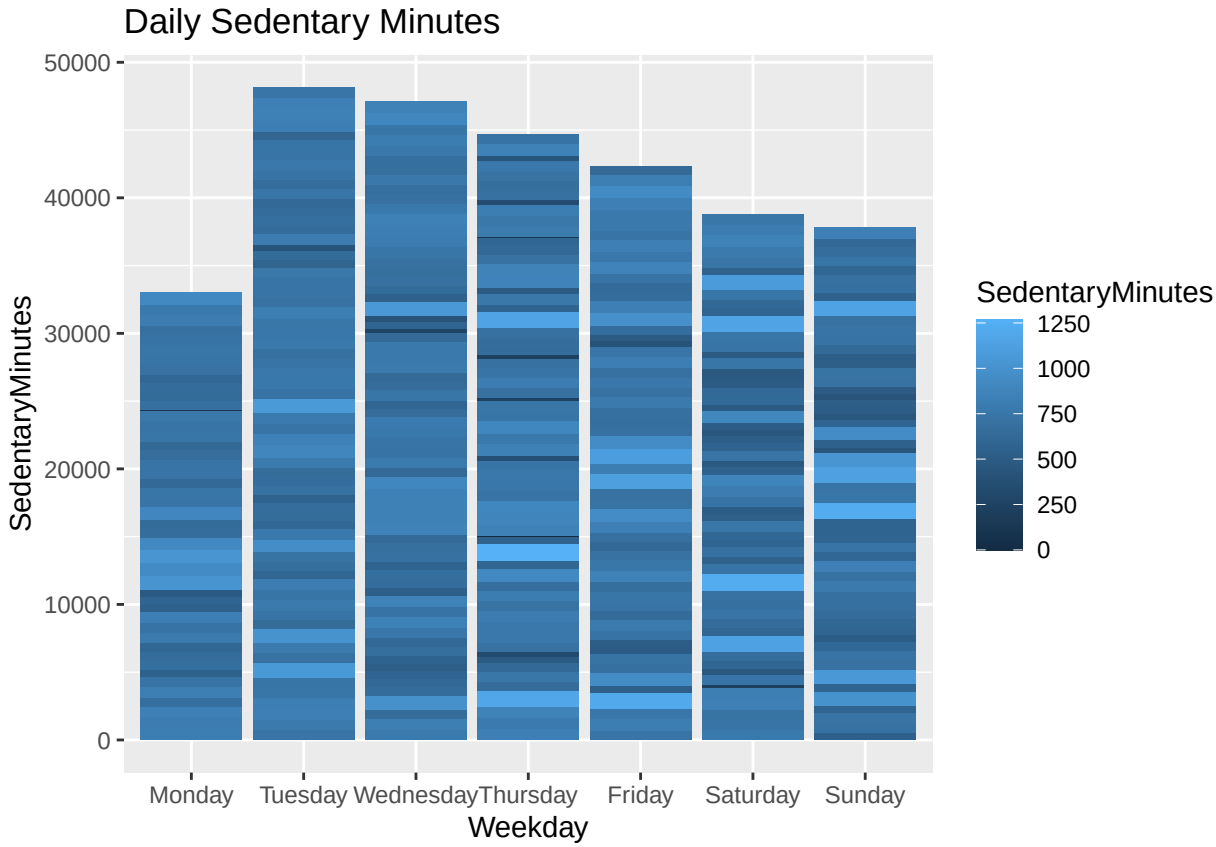



```
# More active minutes more steps.

ggplot(data=daily_analysis_data, aes(x=SedentaryMinutes, y=SomeActivity)) +
  geom_point()+
  geom_smooth(method = "loess") +
  labs(title = "Comparison of SedentaryMinutes VS Activity Time", x = "Sedentary Time", y = "Activity Ti
```

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



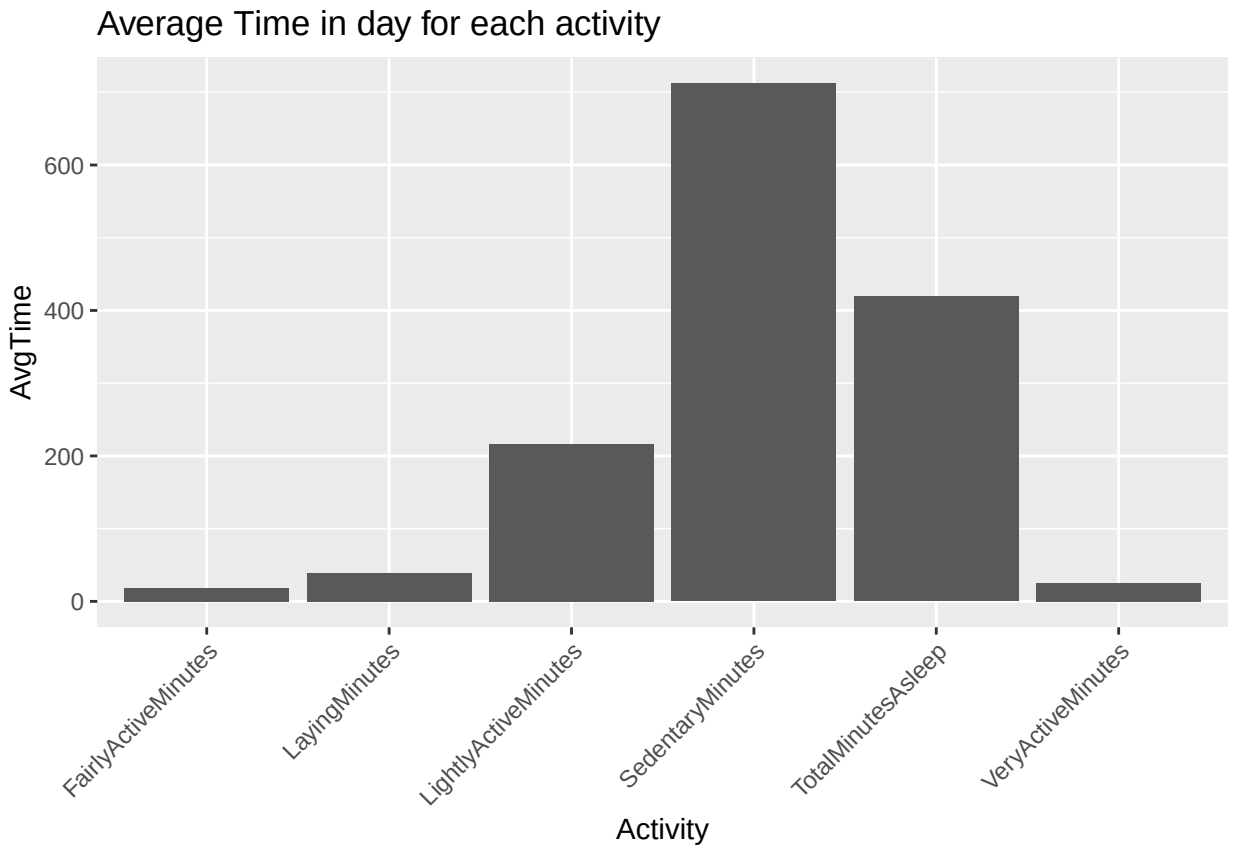

```
daily_analysis_data$Weekday <- factor(daily_analysis_data$Weekday, levels = c("Monday", "Tuesday", "Wedn

ggplot(data=daily_analysis_data, aes(x=Weekday, y=SedentaryMinutes, fill=SedentaryMinutes))+
  geom_bar(stat="identity")+
  labs(title="Daily Sedentary Minutes")
```




```
#Lets look at daily time distribution

sums_act <- colSums(daily_analysis_data[, c("VeryActiveMinutes", "FairlyActiveMinutes", "LightlyActiveMi

total_sum_act <- sum(sums_act)

percentages_sum_act <- (sums_act / total_sum_act) * 100

avg_act <- colMeans(daily_analysis_data[, c("VeryActiveMinutes", "FairlyActiveMinutes", "LightlyActiveMi

sum_daily_activites <- data.frame(
  Activity = names(sums_act),
  TotalTime = as.numeric(sums_act),
  PercentageOfTotal = percentages_sum_act,
  AvgTime = as.numeric(avg_act)
)

head(sum_daily_activites) #checking
```

```
##                                  Activity TotalTime PercentageOfTotal   AvgTime
## VeryActiveMinutes       VeryActiveMinutes     10269          1.751376  25.04634
## FairlyActiveMinutes   FairlyActiveMinutes      7349          1.253370  17.92439
## LightlyActiveMinutes LightlyActiveMinutes     88782         15.141752 216.54146
## SedentaryMinutes         SedentaryMinutes    291961         49.793891 712.10000
## TotalMinutesAsleep     TotalMinutesAsleep    171861         29.310859 419.17317
## LayingMinutes               LayingMinutes     16117          2.748751  39.30976
```

```
ggplot(data=sum_daily_activites, aes(x=Activity, y=AvgTime))+
  geom_bar(stat="identity")+
  labs(title="Average Time in day for each activity")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
#Most of the time is spent as sedentary minutes.

#Explore activities by weekdays

avg_by_weekday <- daily_analysis_data %>%
  group_by(Weekday) %>%
  summarise(
    AvgVeryActiveMinutes = mean(VeryActiveMinutes, na.rm = TRUE),
    AvgFairlyActiveMinutes = mean(FairlyActiveMinutes, na.rm = TRUE),
    AvgLightlyActiveMinutes = mean(LightlyActiveMinutes, na.rm = TRUE),
    AvgSedentaryMinutes = mean(SedentaryMinutes, na.rm = TRUE),
    AvgTotalMinutesAsleep = mean(TotalMinutesAsleep, na.rm = TRUE),
    AvgLayingMinutes = mean(LayingMinutes, na.rm = TRUE)
  )

head(avg_by_weekday)
```

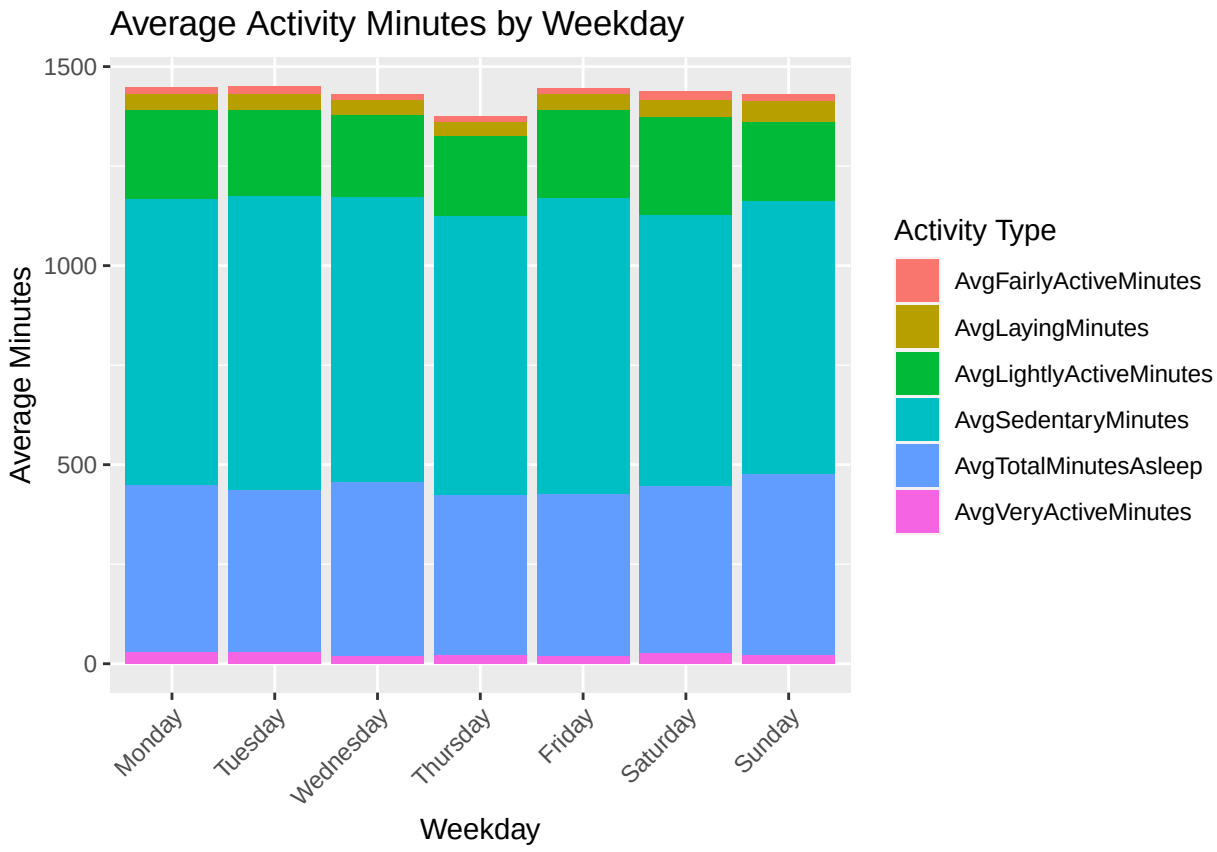
```
## # A tibble: 6 x 7
##   Weekday   AvgVeryActiveMinutes AvgFairlyActiveMinutes AvgLightlyActiveMinutes
##   <fct>                    <dbl>                  <dbl>                   <dbl>
## 1 Monday                    30.7                   19.1                    222.
## 2 Tuesday                   30.6                   20.0                    217.
```

```
## 3 Wednesday                 21.3                  16.7                  208.
## 4 Thursday                  22.9                  15.9                  203.
## 5 Friday                    21.2                  14.6                  223.
## 6 Saturday                  27.6                  22.7                  247.
## # i 3 more variables: AvgSedentaryMinutes <dbl>, AvgTotalMinutesAsleep <dbl>,
## #   AvgLayingMinutes <dbl>
```

```
summary(avg_by_weekday)
```

```
##       Weekday  AvgVeryActiveMinutes AvgFairlyActiveMinutes
##  Monday   :1   Min.   :21.16        Min.   :14.58
##  Tuesday  :1   1st Qu.:21.74        1st Qu.:16.30
##  Wednesday:1   Median :22.86        Median :16.76
##  Thursday :1   Mean   :25.20        Mean   :17.97
##  Friday   :1   3rd Qu.:29.09        3rd Qu.:19.57
##  Saturday :1   Max.   :30.72        Max.   :22.72
##  Sunday   :1
##  AvgLightlyActiveMinutes AvgSedentaryMinutes AvgTotalMinutesAsleep
##  Min.   :200.0           Min.   :680.4       Min.   :401.3
##  1st Qu.:205.5           1st Qu.:693.0       1st Qu.:405.0
##  Median :216.6           Median :714.5       Median :419.1
##  Mean   :217.1           Mean   :711.8       Mean   :419.6
##  3rd Qu.:222.5           3rd Qu.:729.2       3rd Qu.:427.1
##  Max.   :246.8           Max.   :743.1       Max.   :452.7
##
##  AvgLayingMinutes
##  Min.   :33.58
##  1st Qu.:36.60
##  Median :38.75
##  Mean   :39.53
##  3rd Qu.:40.20
##  Max.   :50.76
##
```

```
avg_by_weekday_long <- avg_by_weekday %>%
  pivot_longer(cols = starts_with("Avg"),
               names_to = "ActivityType",
               values_to = "Minutes")

ggplot(avg_by_weekday_long, aes(x = Weekday, y = Minutes, fill = ActivityType)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Activity Minutes by Weekday",
       x = "Weekday",
       y = "Average Minutes",
       fill = "Activity Type") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



5.

Export files

```
# Preparing and exporting files for analysis and presentation in Tableau

# Only data I will need and in long format

exp_daily_act <- daily_analysis_data %>%
  select(Id, Date, Weekday, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinut
  pivot_longer(
    cols = c(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes, TotalMinute
    names_to = "ActivityType",
    values_to = "Minutes"
  )

head(exp_daily_act)
```

```
## # A tibble: 6 x 5
##   Id         Date       Weekday ActivityType         Minutes
##   <chr>      <date>     <fct>   <chr>                  <int>
## 1 1503960366 2016-04-12 Tuesday VeryActiveMinutes         25
## 2 1503960366 2016-04-12 Tuesday FairlyActiveMinutes       13
## 3 1503960366 2016-04-12 Tuesday LightlyActiveMinutes     328
## 4 1503960366 2016-04-12 Tuesday SedentaryMinutes         728
## 5 1503960366 2016-04-12 Tuesday TotalMinutesAsleep       327
## 6 1503960366 2016-04-12 Tuesday LayingMinutes             19
```

```
summary(exp_daily_act) #checking
```

```
##       Id                 Date                Weekday    ActivityType
```

```
##  Length:2460        Min.   :2016-04-12   Monday   :276   Length:2460
##  Class :character   1st Qu.:2016-04-19   Tuesday  :390   Class :character
##  Mode  :character   Median :2016-04-27   Wednesday:396   Mode  :character
##                     Mean   :2016-04-26   Thursday :384
##                     3rd Qu.:2016-05-04   Friday   :342
##                     Max.   :2016-05-12   Saturday :342
##                                          Sunday   :330
##     Minutes
##  Min.   :   0.0
##  1st Qu.:  18.0
##  Median : 112.5
##  Mean   : 238.3
##  3rd Qu.: 428.2
##  Max.   :1265.0
##
```

```
write.csv(exp_daily_act, file = "/cloud/project/output_bellabeat/daily_data.csv",  row.names = FALSE)

# Export the Hourly data
write.csv(comb_hourly_data, file= "/cloud/project/output_bellabeat/comb_hourly_data.csv", row.names = F

#For summary
exp_summary_act <- daily_analysis_data %>%
  select(Id, Date, Weekday, SomeActivity, TotalMinutesAsleep, SedetaryBedMinutes) %>%
  pivot_longer(
    cols = c(SomeActivity, TotalMinutesAsleep, SedetaryBedMinutes),
    names_to = "ActivityType",
    values_to = "Minutes"
  )

head(exp_summary_act)
```

```
## # A tibble: 6 x 5
##   Id         Date       Weekday   ActivityType       Minutes
##   <chr>      <date>     <fct>     <chr>                <int>
## 1 1503960366 2016-04-12 Tuesday   SomeActivity           366
## 2 1503960366 2016-04-12 Tuesday   TotalMinutesAsleep     327
## 3 1503960366 2016-04-12 Tuesday   SedetaryBedMinutes     747
## 4 1503960366 2016-04-13 Wednesday SomeActivity           257
## 5 1503960366 2016-04-13 Wednesday TotalMinutesAsleep     384
## 6 1503960366 2016-04-13 Wednesday SedetaryBedMinutes     799
```

```
summary(exp_summary_act) #checking
```

```
##       Id                 Date                Weekday    ActivityType
##  Length:1230        Min.   :2016-04-12   Monday   :138   Length:1230
##  Class :character   1st Qu.:2016-04-19   Tuesday  :195   Class :character
##  Mode  :character   Median :2016-04-27   Wednesday:198   Mode  :character
##                     Mean   :2016-04-26   Thursday :192
##                     3rd Qu.:2016-05-04   Friday   :171
##                     Max.   :2016-05-12   Saturday :171
##                                          Sunday   :165
##     Minutes
##  Min.   :   2.0
##  1st Qu.: 291.0
```

```
##  Median : 433.0
##  Mean   : 476.7
##  3rd Qu.: 675.8
##  Max.   :1271.0
##
```

```
# Export the Hourly data
write.csv(exp_summary_act, file= "/cloud/project/output_bellabeat/summary_act_data.csv", row.names = FA
```