CASE_STUDY2

RANA WALEED BHNAM

2025-02-15

Bellabeat// Case study:

1. Summary:

Bellabeat, a high-tech company that manufactures health-focused smart products wants to analyse the usage of one of their products in order to gain insight into how people are already using their smart devices.Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

2. ASK:

2.1 Identify the business task:

The company better target their marketing efforts into their customer's needs based on their usage of their fitness smart devices. With this info then make high-level recommendations for how these trends can inform Bellabeat marketing strategy. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company.

2.2 Questions for the analysis:

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

2.3 The business task:

Given these facts, the business task is defined as searching for user patterns of usage of their smart devices in order to gain insights that would later better orientate marketing decisions.

3. Prepare:

3.1 Dataset used:

The data source used for our case study is FitBit Fitness Tracker Data. This dataset is stored in Kaggle and was made available through Mobius.

3.2 Data Organization and verification:

Available to us are 18 CSV documents. Each document represents different quantitative data tracked by Fitbit. The data is considered long since each row is one time point per subject, so each subject will have data in multiple rows.Every user has a unique ID and different rows since data is tracked by day and time.

I sorted and filtered tables and I was able to verify attributes and observations of each table and relations between tables. Counted sample size (users) of each table and verified time length of analysis 31 days of 33 users. The type of every table is "Microsoft Excel CSV": 1. dailyActivity_merged: Daily Activity, Tracking:

Steps, Distance, Intensities, Calories. 2. dailyCalories_merged: Daily Calories. 3. dailyIntensities_merged: Daily Intensity, Measured in Minutes/Distance, 4. gro. categories. 4. dailySteps_merged: Daily Steps. 5. heartrate_seconds_merged: day and time heartrate logs for just 7 users. 6. hourlyCalories_merged: Hourly Calories burned. 7. hourlyIntensities_merged: Hourly total and average intensity. 8. hourlySteps_merged: Hourly Steps. 9. minuteCaloriesNarrow_merged: Calories burned every minute (Every minute in single row) 10. minuteCaloriesWide_merged: Calories burned (Every minute in single column) 11. minuteIntensitiesNarrow_merged: Intensity counted (Every minute in single row) 12. minuteIntensitiesWide_merged: Intensity counted by minute (Every minute in single column) 13. minuteMETsNarrow_merged: Ratio of the energy you are using in a physical activity compared. 14. minuteSleep_merged: Log Sleep by Minute for 24 users. Value column not specified 15. minuteStepsNarrow_merged: Steps tracked (Every minute in single row) 16. minuteStepsWide_merged: Steps tracked (Every minute in single column) 17. sleepDay_merged: Daily sleep logs, by Total count of sleeps a day & minutes/Time in Bed 18. weightLogInfo_merged: Weight track by day in Kg and Pounds. Calculation of BMI.

4. Process and Analyse:

I will focus my analysis in R due to the accessibility, amount of data and to be able to create data visualization to share my results with stakeholders.

4.1 Installing packages and opening libraries

We will choose the packages that will help us on our analysis and open them. We will use the following packages:

```
library(tidyverse)
```

```
## -- 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   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- 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
```

```
library(tidyr)
library(readr)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(dplyr)
library(lubridate)
library(ggpubr)
library(ggplot2)
```

```r
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

4.2 Importing datasets (Load our CSV. files):

We will upload the datasets that will help us answer our business task. On our analysis we will focus on the following datasets:

Activity_Day

Sleep_Day Sleep_Time

Calories_Day Calories_Hour CaloriesNarrow_Minute CaloriesWide_Minute

Intensities_Day Intensities_Hour IntensitiesNarrow_Minute IntensitiesWide_Minute

Steps_Day Steps_Hour StepsNarrow_Minute StepsWide_Minute

heartrate_Time

METsNarrow_Minute

weight_LogInfo.

```
## Rows: 940 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 413 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 188521 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): date
## dbl (3): Id, value, logId
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
```

```
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 22099 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1325580 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityMinute
## dbl (2): Id, Calories
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 21645 Columns: 62
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityHour
## dbl (61): Id, Calories00, Calories01, Calories02, Calories03, Calories04, Ca...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, Ve...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 22099 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityHour
## dbl (3): Id, TotalIntensity, AverageIntensity
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1325580 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityMinute
## dbl (2): Id, Intensity
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 21645 Columns: 62
## -- Column specification --------------------------------------------------------
```

```
## Delimiter: ","
## chr  (1): ActivityHour
## dbl (61): Id, Intensity00, Intensity01, Intensity02, Intensity03, Intensity0...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 940 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, StepTotal
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 22099 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, StepTotal
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1325580 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityMinute
## dbl (2): Id, Steps
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 21645 Columns: 62
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityHour
## dbl (61): Id, Steps00, Steps01, Steps02, Steps03, Steps04, Steps05, Steps06,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1048575 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Time
## dbl (2): Id, Value
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## Rows: 1325580 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ActivityMinute
## dbl (2): Id, METs
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## Rows: 67 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId
## lgl (1): IsManualReport
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

4.3 Preview our datasets:

- We will preview our selected data frames and check the summary of each column. Take a look at all data.

```
head(Activity_Day)
```

```
## # A tibble: 6 x 15
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
##        <dbl> <chr>             <dbl>         <dbl>           <dbl>
## 1 1503960366 4/12/2016         13162          8.5             8.5
## 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
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

```
head(Sleep_Day)
```

```
## # A tibble: 6 x 5
##           Id SleepDay        TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##        <dbl> <chr>                       <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0~                1                327            346
## 2 1503960366 4/13/2016 12:0~                2                384            407
## 3 1503960366 4/15/2016 12:0~                1                412            442
## 4 1503960366 4/16/2016 12:0~                2                340            367
## 5 1503960366 4/17/2016 12:0~                1                700            712
## 6 1503960366 4/19/2016 12:0~                1                304            320
```

```
head(Sleep_Time)
```

```
## # A tibble: 6 x 4
##           Id date                 value       logId
##        <dbl> <chr>                <dbl>       <dbl>
## 1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
## 2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
## 3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
## 4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
## 5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
## 6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
```

```
head(Intensities_Day)
```

```
## # A tibble: 6 x 10
##         Id ActivityDay SedentaryMinutes LightlyActiveMinutes FairlyActiveMinutes
##      <dbl> <chr>                  <dbl>                <dbl>                <dbl>
## 1   1.50e9 4/12/2016                728                  328                   13
## 2   1.50e9 4/13/2016                776                  217                   19
## 3   1.50e9 4/14/2016               1218                  181                   11
## 4   1.50e9 4/15/2016                726                  209                   34
## 5   1.50e9 4/16/2016                773                  221                   10
## 6   1.50e9 4/17/2016                539                  164                   20
## # i 5 more variables: VeryActiveMinutes <dbl>, SedentaryActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   VeryActiveDistance <dbl>
```

```
head(Intensities_Hour)
```

```
## # A tibble: 6 x 4
##           Id ActivityHour          TotalIntensity AverageIntensity
##        <dbl> <chr>                          <dbl>            <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM             20            0.333
## 2 1503960366 4/12/2016 1:00:00 AM               8            0.133
## 3 1503960366 4/12/2016 2:00:00 AM               7            0.117
## 4 1503960366 4/12/2016 3:00:00 AM               0            0
## 5 1503960366 4/12/2016 4:00:00 AM               0            0
## 6 1503960366 4/12/2016 5:00:00 AM               0            0
```

```
head(IntensitiesNarrow_Minute)
```

```
## # A tibble: 6 x 3
##           Id ActivityMinute        Intensity
##        <dbl> <chr>                     <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM         0
## 2 1503960366 4/12/2016 12:01:00 AM         0
## 3 1503960366 4/12/2016 12:02:00 AM         0
## 4 1503960366 4/12/2016 12:03:00 AM         0
## 5 1503960366 4/12/2016 12:04:00 AM         0
## 6 1503960366 4/12/2016 12:05:00 AM         0
```

```
head(IntensitiesWide_Minute)
```

```
## # A tibble: 6 x 62
##           Id ActivityHour        Intensity00 Intensity01 Intensity02 Intensity03
##        <dbl> <chr>                     <dbl>       <dbl>       <dbl>       <dbl>
## 1 1503960366 4/13/2016 12:00:00~           1           1           0           0
## 2 1503960366 4/13/2016 1:00:00 ~           0           0           0           0
## 3 1503960366 4/13/2016 2:00:00 ~           0           0           0           0
## 4 1503960366 4/13/2016 3:00:00 ~           0           0           0           0
## 5 1503960366 4/13/2016 4:00:00 ~           0           0           0           0
## 6 1503960366 4/13/2016 5:00:00 ~           0           0           0           0
## # i 56 more variables: Intensity04 <dbl>, Intensity05 <dbl>, Intensity06 <dbl>,
## #   Intensity07 <dbl>, Intensity08 <dbl>, Intensity09 <dbl>, Intensity10 <dbl>,
## #   Intensity11 <dbl>, Intensity12 <dbl>, Intensity13 <dbl>, Intensity14 <dbl>,
## #   Intensity15 <dbl>, Intensity16 <dbl>, Intensity17 <dbl>, Intensity18 <dbl>,
## #   Intensity19 <dbl>, Intensity20 <dbl>, Intensity21 <dbl>, Intensity22 <dbl>,
## #   Intensity23 <dbl>, Intensity24 <dbl>, Intensity25 <dbl>, Intensity26 <dbl>,
```

```
## #   Intensity27 <dbl>, Intensity28 <dbl>, Intensity29 <dbl>, ...
```

```
head(Calories_Day)
```

```
## # A tibble: 6 x 3
##           Id ActivityDay Calories
##        <dbl> <chr>          <dbl>
## 1 1503960366 4/12/2016       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
```

```
head(Calories_Hour)
```

```
## # A tibble: 6 x 3
##           Id ActivityHour          Calories
##        <dbl> <chr>                    <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM       81
## 2 1503960366 4/12/2016 1:00:00 AM        61
## 3 1503960366 4/12/2016 2:00:00 AM        59
## 4 1503960366 4/12/2016 3:00:00 AM        47
## 5 1503960366 4/12/2016 4:00:00 AM        48
## 6 1503960366 4/12/2016 5:00:00 AM        48
```

```
head(CaloriesNarrow_Minute)
```

```
## # A tibble: 6 x 3
##           Id ActivityMinute        Calories
##        <dbl> <chr>                    <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM    0.786
## 2 1503960366 4/12/2016 12:01:00 AM    0.786
## 3 1503960366 4/12/2016 12:02:00 AM    0.786
## 4 1503960366 4/12/2016 12:03:00 AM    0.786
## 5 1503960366 4/12/2016 12:04:00 AM    0.786
## 6 1503960366 4/12/2016 12:05:00 AM    0.944
```

```
head(CaloriesWide_Minute)
```

```
## # A tibble: 6 x 62
##           Id ActivityHour Calories00 Calories01 Calories02 Calories03 Calories04
##        <dbl> <chr>              <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 1503960366 4/13/2016 1~      1.89       2.20       0.944      0.944      0.944
## 2 1503960366 4/13/2016 1~      0.786      0.786      0.786      0.786      0.944
## 3 1503960366 4/13/2016 2~      0.786      0.786      0.786      0.786      0.786
## 4 1503960366 4/13/2016 3~      0.786      0.786      0.786      0.786      0.786
## 5 1503960366 4/13/2016 4~      0.786      0.786      0.786      0.786      0.786
## 6 1503960366 4/13/2016 5~      0.786      0.786      0.786      0.786      0.786
## # i 55 more variables: Calories05 <dbl>, Calories06 <dbl>, Calories07 <dbl>,
## #   Calories08 <dbl>, Calories09 <dbl>, Calories10 <dbl>, Calories11 <dbl>,
## #   Calories12 <dbl>, Calories13 <dbl>, Calories14 <dbl>, Calories15 <dbl>,
## #   Calories16 <dbl>, Calories17 <dbl>, Calories18 <dbl>, Calories19 <dbl>,
## #   Calories20 <dbl>, Calories21 <dbl>, Calories22 <dbl>, Calories23 <dbl>,
## #   Calories24 <dbl>, Calories25 <dbl>, Calories26 <dbl>, Calories27 <dbl>,
## #   Calories28 <dbl>, Calories29 <dbl>, Calories30 <dbl>, Calories31 <dbl>, ...
```

```
head(Steps_Day)
```

```
## # A tibble: 6 x 3
##           Id ActivityDay StepTotal
##        <dbl> <chr>           <dbl>
## 1 1503960366 4/12/2016       13162
## 2 1503960366 4/13/2016       10735
## 3 1503960366 4/14/2016       10460
## 4 1503960366 4/15/2016        9762
## 5 1503960366 4/16/2016       12669
## 6 1503960366 4/17/2016        9705
```

```
head(Steps_Hour)
```

```
## # A tibble: 6 x 3
##           Id ActivityHour          StepTotal
##        <dbl> <chr>                     <dbl>
## 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
```

```
head(StepsNarrow_Minute)
```

```
## # A tibble: 6 x 3
##           Id ActivityMinute        Steps
##        <dbl> <chr>                 <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM     0
## 2 1503960366 4/12/2016 12:01:00 AM     0
## 3 1503960366 4/12/2016 12:02:00 AM     0
## 4 1503960366 4/12/2016 12:03:00 AM     0
## 5 1503960366 4/12/2016 12:04:00 AM     0
## 6 1503960366 4/12/2016 12:05:00 AM     0
```

```
head(StepsWide_Minute)
```

```
## # A tibble: 6 x 62
##          Id ActivityHour Steps00 Steps01 Steps02 Steps03 Steps04 Steps05 Steps06
##       <dbl> <chr>           <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1    1.50e9 4/13/2016 1~       4      16       0       0       0       9       0
## 2    1.50e9 4/13/2016 1~       0       0       0       0       0       0       0
## 3    1.50e9 4/13/2016 2~       0       0       0       0       0       0       0
## 4    1.50e9 4/13/2016 3~       0       0       0       0       0       0       0
## 5    1.50e9 4/13/2016 4~       0       0       0       0       0       0       0
## 6    1.50e9 4/13/2016 5~       0       0       0       0       0       0       0
## # i 53 more variables: Steps07 <dbl>, Steps08 <dbl>, Steps09 <dbl>,
## #   Steps10 <dbl>, Steps11 <dbl>, Steps12 <dbl>, Steps13 <dbl>, Steps14 <dbl>,
## #   Steps15 <dbl>, Steps16 <dbl>, Steps17 <dbl>, Steps18 <dbl>, Steps19 <dbl>,
## #   Steps20 <dbl>, Steps21 <dbl>, Steps22 <dbl>, Steps23 <dbl>, Steps24 <dbl>,
## #   Steps25 <dbl>, Steps26 <dbl>, Steps27 <dbl>, Steps28 <dbl>, Steps29 <dbl>,
## #   Steps30 <dbl>, Steps31 <dbl>, Steps32 <dbl>, Steps33 <dbl>, Steps34 <dbl>,
## #   Steps35 <dbl>, Steps36 <dbl>, Steps37 <dbl>, Steps38 <dbl>, ...
```

```
head(METsNarrow_Minute)
```

```
## # A tibble: 6 x 3
##           Id ActivityMinute         METs
##        <dbl> <chr>                 <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM    10
## 2 1503960366 4/12/2016 12:01:00 AM    10
## 3 1503960366 4/12/2016 12:02:00 AM    10
## 4 1503960366 4/12/2016 12:03:00 AM    10
## 5 1503960366 4/12/2016 12:04:00 AM    10
## 6 1503960366 4/12/2016 12:05:00 AM    12
```

```
head(weight_LogInfo)
```

```
## # A tibble: 6 x 8
##           Id Date       WeightKg WeightPounds   Fat   BMI IsManualReport   LogId
##        <dbl> <chr>         <dbl>        <dbl> <dbl> <dbl> <lgl>            <dbl>
## 1 1503960366 5/2/2016 ~     52.6         116.    22  22.6 TRUE           1.46e12
## 2 1503960366 5/3/2016 ~     52.6         116.    NA  22.6 TRUE           1.46e12
## 3 1927972279 4/13/2016~    134.          294.    NA  47.5 FALSE          1.46e12
## 4 2873212765 4/21/2016~     56.7         125.    NA  21.5 TRUE           1.46e12
## 5 2873212765 5/12/2016~     57.3         126.    NA  21.7 TRUE           1.46e12
## 6 4319703577 4/17/2016~     72.4         160.    25  27.5 TRUE           1.46e12
```

- Identify all the columns names in all data.

```
colnames(Activity_Day)
```

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

```
colnames(Sleep_Day)
```

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

```
colnames(Sleep_Time)
```

```
## [1] "Id"    "date"  "value" "logId"
```

```
colnames(Intensities_Day)
```

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

```
colnames(Intensities_Hour)
```

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

```
colnames(IntensitiesNarrow_Minute)
```

```
## [1] "Id"             "ActivityMinute" "Intensity"
```

```
colnames(IntensitiesWide_Minute)
```

```
##  [1] "Id"           "ActivityHour" "Intensity00"  "Intensity01"  "Intensity02"
##  [6] "Intensity03"  "Intensity04"  "Intensity05"  "Intensity06"  "Intensity07"
## [11] "Intensity08"  "Intensity09"  "Intensity10"  "Intensity11"  "Intensity12"
## [16] "Intensity13"  "Intensity14"  "Intensity15"  "Intensity16"  "Intensity17"
## [21] "Intensity18"  "Intensity19"  "Intensity20"  "Intensity21"  "Intensity22"
## [26] "Intensity23"  "Intensity24"  "Intensity25"  "Intensity26"  "Intensity27"
## [31] "Intensity28"  "Intensity29"  "Intensity30"  "Intensity31"  "Intensity32"
## [36] "Intensity33"  "Intensity34"  "Intensity35"  "Intensity36"  "Intensity37"
## [41] "Intensity38"  "Intensity39"  "Intensity40"  "Intensity41"  "Intensity42"
## [46] "Intensity43"  "Intensity44"  "Intensity45"  "Intensity46"  "Intensity47"
## [51] "Intensity48"  "Intensity49"  "Intensity50"  "Intensity51"  "Intensity52"
## [56] "Intensity53"  "Intensity54"  "Intensity55"  "Intensity56"  "Intensity57"
## [61] "Intensity58"  "Intensity59"
```

```
colnames(Calories_Day)
```

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

```
colnames(Calories_Hour)
```

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

```
colnames(CaloriesNarrow_Minute)
```

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

```
colnames(CaloriesWide_Minute)
```

```
##  [1] "Id"           "ActivityHour" "Calories00"   "Calories01"   "Calories02"
##  [6] "Calories03"   "Calories04"   "Calories05"   "Calories06"   "Calories07"
## [11] "Calories08"   "Calories09"   "Calories10"   "Calories11"   "Calories12"
## [16] "Calories13"   "Calories14"   "Calories15"   "Calories16"   "Calories17"
## [21] "Calories18"   "Calories19"   "Calories20"   "Calories21"   "Calories22"
## [26] "Calories23"   "Calories24"   "Calories25"   "Calories26"   "Calories27"
## [31] "Calories28"   "Calories29"   "Calories30"   "Calories31"   "Calories32"
## [36] "Calories33"   "Calories34"   "Calories35"   "Calories36"   "Calories37"
## [41] "Calories38"   "Calories39"   "Calories40"   "Calories41"   "Calories42"
## [46] "Calories43"   "Calories44"   "Calories45"   "Calories46"   "Calories47"
## [51] "Calories48"   "Calories49"   "Calories50"   "Calories51"   "Calories52"
## [56] "Calories53"   "Calories54"   "Calories55"   "Calories56"   "Calories57"
## [61] "Calories58"   "Calories59"
```

```
colnames(Steps_Day)
```

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

```
colnames(Steps_Hour)
```

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

```
colnames(StepsNarrow_Minute)
```

```
## [1] "Id"             "ActivityMinute" "Steps"
```

```r
colnames(StepsWide_Minute)
```

```
##  [1] "Id"           "ActivityHour" "Steps00"      "Steps01"      "Steps02"
##  [6] "Steps03"      "Steps04"      "Steps05"      "Steps06"      "Steps07"
## [11] "Steps08"      "Steps09"      "Steps10"      "Steps11"      "Steps12"
## [16] "Steps13"      "Steps14"      "Steps15"      "Steps16"      "Steps17"
## [21] "Steps18"      "Steps19"      "Steps20"      "Steps21"      "Steps22"
## [26] "Steps23"      "Steps24"      "Steps25"      "Steps26"      "Steps27"
## [31] "Steps28"      "Steps29"      "Steps30"      "Steps31"      "Steps32"
## [36] "Steps33"      "Steps34"      "Steps35"      "Steps36"      "Steps37"
## [41] "Steps38"      "Steps39"      "Steps40"      "Steps41"      "Steps42"
## [46] "Steps43"      "Steps44"      "Steps45"      "Steps46"      "Steps47"
## [51] "Steps48"      "Steps49"      "Steps50"      "Steps51"      "Steps52"
## [56] "Steps53"      "Steps54"      "Steps55"      "Steps56"      "Steps57"
## [61] "Steps58"      "Steps59"
```

```r
colnames(METsNarrow_Minute)
```

```
## [1] "Id"             "ActivityMinute" "METs"
```

```r
colnames(weight_LogInfo)
```

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

- Identify all the structure in all data.

```r
str(Activity_Day)
```

```
## spc_tbl_ [940 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                       : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate             : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps               : num [1:940] 13162 10735 10460 9762 12669 ...
##  $ TotalDistance            : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance          : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance       : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance      : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes        : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes      : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes     : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes         : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                 : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
```

```
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Sleep_Day)
```

```
## spc_tbl_ [413 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:0
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Sleep_Time)
```

```
## spc_tbl_ [188,521 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id   : num [1:188521] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ date : chr [1:188521] "4/12/2016 2:47:30 AM" "4/12/2016 2:48:30 AM" "4/12/2016 2:49:30 AM" "4/12/1
##  $ value: num [1:188521] 3 2 1 1 1 1 1 2 2 2 ...
##  $ logId: num [1:188521] 1.14e+10 1.14e+10 1.14e+10 1.14e+10 1.14e+10 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   date = col_character(),
##   ..   value = col_double(),
##   ..   logId = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Intensities_Day)
```

```
## spc_tbl_ [940 x 10] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                       : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDay              : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ SedentaryMinutes         : num [1:940] 728 776 1218 726 773 ...
##  $ LightlyActiveMinutes     : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ FairlyActiveMinutes      : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ VeryActiveMinutes        : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ LightActiveDistance      : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ VeryActiveDistance       : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
```

```
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDay = col_character(),
##   ..   SedentaryMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   VeryActiveDistance = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Intensities_Hour)
```

```
## spc_tbl_ [22,099 x 4] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id              : num [1:22099] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour    : chr [1:22099] "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM" "4/12/2016 2:00:00
##  $ TotalIntensity  : num [1:22099] 20 8 7 0 0 0 0 0 13 30 ...
##  $ AverageIntensity: num [1:22099] 0.333 0.133 0.117 0 0 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityHour = col_character(),
##   ..   TotalIntensity = col_double(),
##   ..   AverageIntensity = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(IntensitiesNarrow_Minute)
```

```
## spc_tbl_ [1,325,580 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id            : num [1:1325580] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityMinute: chr [1:1325580] "4/12/2016 12:00:00 AM" "4/12/2016 12:01:00 AM" "4/12/2016 12:02:
##  $ Intensity     : num [1:1325580] 0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityMinute = col_character(),
##   ..   Intensity = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(IntensitiesWide_Minute)
```

```
## spc_tbl_ [21,645 x 62] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id          : num [1:21645] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr [1:21645] "4/13/2016 12:00:00 AM" "4/13/2016 1:00:00 AM" "4/13/2016 2:00:00 AM"
##  $ Intensity00 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity01 : num [1:21645] 1 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity02 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity03 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity04 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity05 : num [1:21645] 1 0 0 0 0 0 0 0 0 1 ...
```

```
##  $ Intensity06 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity07 : num [1:21645] 1 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity08 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity09 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity10 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity11 : num [1:21645] 0 0 0 1 0 0 0 0 1 1 ...
##  $ Intensity12 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity13 : num [1:21645] 0 0 0 0 0 0 0 0 1 1 ...
##  $ Intensity14 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity15 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity16 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity17 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity18 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity19 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity20 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity21 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity22 : num [1:21645] 0 0 0 0 0 0 0 0 1 0 ...
##  $ Intensity23 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity24 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity25 : num [1:21645] 1 0 0 0 0 0 0 1 0 0 ...
##  $ Intensity26 : num [1:21645] 1 0 0 0 0 0 0 1 0 0 ...
##  $ Intensity27 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity28 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity29 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity30 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity31 : num [1:21645] 0 0 0 0 0 0 0 1 0 0 ...
##  $ Intensity32 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity33 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity34 : num [1:21645] 0 0 0 1 0 0 0 0 0 0 ...
##  $ Intensity35 : num [1:21645] 0 0 0 1 0 0 0 0 0 0 ...
##  $ Intensity36 : num [1:21645] 0 0 0 1 0 0 0 0 1 1 ...
##  $ Intensity37 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity38 : num [1:21645] 0 0 0 0 0 0 0 0 1 1 ...
##  $ Intensity39 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity40 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity41 : num [1:21645] 0 0 0 0 0 0 0 1 0 0 ...
##  $ Intensity42 : num [1:21645] 0 0 0 0 0 0 0 1 1 0 ...
##  $ Intensity43 : num [1:21645] 0 0 0 0 0 0 0 0 1 0 ...
##  $ Intensity44 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity45 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity46 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity47 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity48 : num [1:21645] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Intensity49 : num [1:21645] 0 0 0 0 0 0 0 0 1 0 ...
##  $ Intensity50 : num [1:21645] 0 0 0 0 0 0 0 1 0 0 ...
##  $ Intensity51 : num [1:21645] 1 0 0 0 0 0 0 1 1 1 ...
##  $ Intensity52 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity53 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity54 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity55 : num [1:21645] 1 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity56 : num [1:21645] 0 0 0 0 0 0 0 0 1 0 ...
##  $ Intensity57 : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Intensity58 : num [1:21645] 0 0 0 0 0 0 0 0 1 0 ...
##  $ Intensity59 : num [1:21645] 0 0 0 0 0 0 0 1 1 1 ...
```

```
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityHour = col_character(),
##   ..   Intensity00 = col_double(),
##   ..   Intensity01 = col_double(),
##   ..   Intensity02 = col_double(),
##   ..   Intensity03 = col_double(),
##   ..   Intensity04 = col_double(),
##   ..   Intensity05 = col_double(),
##   ..   Intensity06 = col_double(),
##   ..   Intensity07 = col_double(),
##   ..   Intensity08 = col_double(),
##   ..   Intensity09 = col_double(),
##   ..   Intensity10 = col_double(),
##   ..   Intensity11 = col_double(),
##   ..   Intensity12 = col_double(),
##   ..   Intensity13 = col_double(),
##   ..   Intensity14 = col_double(),
##   ..   Intensity15 = col_double(),
##   ..   Intensity16 = col_double(),
##   ..   Intensity17 = col_double(),
##   ..   Intensity18 = col_double(),
##   ..   Intensity19 = col_double(),
##   ..   Intensity20 = col_double(),
##   ..   Intensity21 = col_double(),
##   ..   Intensity22 = col_double(),
##   ..   Intensity23 = col_double(),
##   ..   Intensity24 = col_double(),
##   ..   Intensity25 = col_double(),
##   ..   Intensity26 = col_double(),
##   ..   Intensity27 = col_double(),
##   ..   Intensity28 = col_double(),
##   ..   Intensity29 = col_double(),
##   ..   Intensity30 = col_double(),
##   ..   Intensity31 = col_double(),
##   ..   Intensity32 = col_double(),
##   ..   Intensity33 = col_double(),
##   ..   Intensity34 = col_double(),
##   ..   Intensity35 = col_double(),
##   ..   Intensity36 = col_double(),
##   ..   Intensity37 = col_double(),
##   ..   Intensity38 = col_double(),
##   ..   Intensity39 = col_double(),
##   ..   Intensity40 = col_double(),
##   ..   Intensity41 = col_double(),
##   ..   Intensity42 = col_double(),
##   ..   Intensity43 = col_double(),
##   ..   Intensity44 = col_double(),
##   ..   Intensity45 = col_double(),
##   ..   Intensity46 = col_double(),
##   ..   Intensity47 = col_double(),
##   ..   Intensity48 = col_double(),
##   ..   Intensity49 = col_double(),
```

```
##   ..   Intensity50 = col_double(),
##   ..   Intensity51 = col_double(),
##   ..   Intensity52 = col_double(),
##   ..   Intensity53 = col_double(),
##   ..   Intensity54 = col_double(),
##   ..   Intensity55 = col_double(),
##   ..   Intensity56 = col_double(),
##   ..   Intensity57 = col_double(),
##   ..   Intensity58 = col_double(),
##   ..   Intensity59 = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Calories_Day)
```

```
## spc_tbl_ [940 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id         : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDay: chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ Calories   : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDay = col_character(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Calories_Hour)
```

```
## spc_tbl_ [22,099 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id          : num [1:22099] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr [1:22099] "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM" "4/12/2016 2:00:00 AM"
##  $ Calories    : num [1:22099] 81 61 59 47 48 48 48 47 68 141 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityHour = col_character(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(CaloriesNarrow_Minute)
```

```
## spc_tbl_ [1,325,580 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id            : num [1:1325580] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityMinute: chr [1:1325580] "4/12/2016 12:00:00 AM" "4/12/2016 12:01:00 AM" "4/12/2016 12:02:
##  $ Calories      : num [1:1325580] 0.786 0.786 0.786 0.786 0.786 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityMinute = col_character(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
str(CaloriesWide_Minute)
```

```
## spc_tbl_ [21,645 x 62] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id          : num [1:21645] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr [1:21645] "4/13/2016 12:00:00 AM" "4/13/2016 1:00:00 AM" "4/13/2016 2:00:00 AM"
##  $ Calories00  : num [1:21645] 1.888 0.786 0.786 0.786 0.786 ...
##  $ Calories01  : num [1:21645] 2.202 0.786 0.786 0.786 0.786 ...
##  $ Calories02  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories03  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories04  : num [1:21645] 0.944 0.944 0.786 0.786 0.786 ...
##  $ Calories05  : num [1:21645] 2.045 0.944 0.786 0.786 0.786 ...
##  $ Calories06  : num [1:21645] 0.944 0.944 0.786 0.786 0.786 ...
##  $ Calories07  : num [1:21645] 2.202 0.786 0.786 0.786 0.786 ...
##  $ Calories08  : num [1:21645] 0.944 0.944 0.786 0.786 0.786 ...
##  $ Calories09  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories10  : num [1:21645] 0.944 0.944 0.786 0.786 0.786 ...
##  $ Calories11  : num [1:21645] 0.786 0.786 0.786 2.045 0.786 ...
##  $ Calories12  : num [1:21645] 0.786 0.944 0.786 0.944 0.786 ...
##  $ Calories13  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories14  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories15  : num [1:21645] 0.944 0.786 0.786 0.944 0.786 ...
##  $ Calories16  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories17  : num [1:21645] 0.786 0.786 0.786 0.944 0.786 ...
##  $ Calories18  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories19  : num [1:21645] 0.786 0.786 0.786 0.786 0.944 ...
##  $ Calories20  : num [1:21645] 1.888 0.786 0.786 0.786 0.786 ...
##  $ Calories21  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories22  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories23  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories24  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories25  : num [1:21645] 2.045 0.786 0.786 0.786 0.786 ...
##  $ Calories26  : num [1:21645] 2.359 0.786 0.786 0.786 0.786 ...
##  $ Calories27  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories28  : num [1:21645] 2.045 0.786 0.786 0.786 0.786 ...
##  $ Calories29  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories30  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories31  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories32  : num [1:21645] 2.045 0.786 0.786 0.786 0.786 ...
##  $ Calories33  : num [1:21645] 1.888 0.786 0.786 0.944 0.786 ...
##  $ Calories34  : num [1:21645] 0.944 0.786 0.786 2.045 0.786 ...
##  $ Calories35  : num [1:21645] 0.786 0.786 0.786 2.045 0.786 ...
##  $ Calories36  : num [1:21645] 0.786 0.786 0.786 1.888 0.786 ...
##  $ Calories37  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories38  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories39  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories40  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories41  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories42  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories43  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories44  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories45  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories46  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories47  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories48  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
```

```
##  $ Calories49  : num [1:21645] 0.786 0.786 0.786 0.786 0.786 ...
##  $ Calories50  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories51  : num [1:21645] 2.045 0.786 0.786 0.786 0.786 ...
##  $ Calories52  : num [1:21645] 2.045 0.786 0.786 0.786 0.786 ...
##  $ Calories53  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories54  : num [1:21645] 2.359 0.786 0.786 0.786 0.786 ...
##  $ Calories55  : num [1:21645] 1.888 0.786 0.786 0.786 0.786 ...
##  $ Calories56  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories57  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories58  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  $ Calories59  : num [1:21645] 0.944 0.786 0.786 0.786 0.786 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityHour = col_character(),
##   ..   Calories00 = col_double(),
##   ..   Calories01 = col_double(),
##   ..   Calories02 = col_double(),
##   ..   Calories03 = col_double(),
##   ..   Calories04 = col_double(),
##   ..   Calories05 = col_double(),
##   ..   Calories06 = col_double(),
##   ..   Calories07 = col_double(),
##   ..   Calories08 = col_double(),
##   ..   Calories09 = col_double(),
##   ..   Calories10 = col_double(),
##   ..   Calories11 = col_double(),
##   ..   Calories12 = col_double(),
##   ..   Calories13 = col_double(),
##   ..   Calories14 = col_double(),
##   ..   Calories15 = col_double(),
##   ..   Calories16 = col_double(),
##   ..   Calories17 = col_double(),
##   ..   Calories18 = col_double(),
##   ..   Calories19 = col_double(),
##   ..   Calories20 = col_double(),
##   ..   Calories21 = col_double(),
##   ..   Calories22 = col_double(),
##   ..   Calories23 = col_double(),
##   ..   Calories24 = col_double(),
##   ..   Calories25 = col_double(),
##   ..   Calories26 = col_double(),
##   ..   Calories27 = col_double(),
##   ..   Calories28 = col_double(),
##   ..   Calories29 = col_double(),
##   ..   Calories30 = col_double(),
##   ..   Calories31 = col_double(),
##   ..   Calories32 = col_double(),
##   ..   Calories33 = col_double(),
##   ..   Calories34 = col_double(),
##   ..   Calories35 = col_double(),
##   ..   Calories36 = col_double(),
##   ..   Calories37 = col_double(),
##   ..   Calories38 = col_double(),
```

```
##   ..   Calories39 = col_double(),
##   ..   Calories40 = col_double(),
##   ..   Calories41 = col_double(),
##   ..   Calories42 = col_double(),
##   ..   Calories43 = col_double(),
##   ..   Calories44 = col_double(),
##   ..   Calories45 = col_double(),
##   ..   Calories46 = col_double(),
##   ..   Calories47 = col_double(),
##   ..   Calories48 = col_double(),
##   ..   Calories49 = col_double(),
##   ..   Calories50 = col_double(),
##   ..   Calories51 = col_double(),
##   ..   Calories52 = col_double(),
##   ..   Calories53 = col_double(),
##   ..   Calories54 = col_double(),
##   ..   Calories55 = col_double(),
##   ..   Calories56 = col_double(),
##   ..   Calories57 = col_double(),
##   ..   Calories58 = col_double(),
##   ..   Calories59 = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Steps_Day)
```

```
## spc_tbl_ [940 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id         : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDay: chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ StepTotal  : num [1:940] 13162 10735 10460 9762 12669 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDay = col_character(),
##   ..   StepTotal = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Steps_Hour)
```

```
## spc_tbl_ [22,099 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id          : num [1:22099] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr [1:22099] "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM" "4/12/2016 2:00:00 AM"
##  $ StepTotal   : num [1:22099] 373 160 151 0 0 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityHour = col_character(),
##   ..   StepTotal = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(StepsNarrow_Minute)
```

```
## spc_tbl_ [1,325,580 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id            : num [1:1325580] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
```

```
##  $ ActivityMinute: chr [1:1325580] "4/12/2016 12:00:00 AM" "4/12/2016 12:01:00 AM" "4/12/2016 12:02:0
##  $ Steps         : num [1:1325580] 0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityMinute = col_character(),
##   ..   Steps = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(StepsWide_Minute)
```

```
## spc_tbl_ [21,645 x 62] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id          : num [1:21645] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr [1:21645] "4/13/2016 12:00:00 AM" "4/13/2016 1:00:00 AM" "4/13/2016 2:00:00 AM"
##  $ Steps00     : num [1:21645] 4 0 0 0 0 0 0 0 0 0 ...
##  $ Steps01     : num [1:21645] 16 0 0 0 0 0 0 0 0 14 ...
##  $ Steps02     : num [1:21645] 0 0 0 0 0 0 0 0 0 10 ...
##  $ Steps03     : num [1:21645] 0 0 0 0 0 0 0 0 0 31 ...
##  $ Steps04     : num [1:21645] 0 0 0 0 0 0 0 0 0 37 ...
##  $ Steps05     : num [1:21645] 9 0 0 0 0 0 0 0 0 17 ...
##  $ Steps06     : num [1:21645] 0 0 0 0 0 0 0 0 0 25 ...
##  $ Steps07     : num [1:21645] 17 0 0 0 0 0 0 0 0 12 ...
##  $ Steps08     : num [1:21645] 0 0 0 0 0 0 0 0 0 6 ...
##  $ Steps09     : num [1:21645] 0 0 0 0 0 0 0 0 0 30 ...
##  $ Steps10     : num [1:21645] 0 0 0 0 0 0 0 0 0 7 ...
##  $ Steps11     : num [1:21645] 0 0 0 10 0 0 0 0 6 109 ...
##  $ Steps12     : num [1:21645] 0 0 0 0 0 0 0 0 0 140 ...
##  $ Steps13     : num [1:21645] 0 0 0 0 0 0 0 0 19 145 ...
##  $ Steps14     : num [1:21645] 0 0 0 0 0 0 0 0 0 152 ...
##  $ Steps15     : num [1:21645] 0 0 0 0 0 0 0 0 0 117 ...
##  $ Steps16     : num [1:21645] 0 0 0 0 0 0 0 0 0 20 ...
##  $ Steps17     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps18     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps19     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps20     : num [1:21645] 6 0 0 0 0 0 0 0 0 0 ...
##  $ Steps21     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps22     : num [1:21645] 0 0 0 0 0 0 0 0 32 0 ...
##  $ Steps23     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps24     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps25     : num [1:21645] 11 0 0 0 0 0 0 26 0 0 ...
##  $ Steps26     : num [1:21645] 21 0 0 0 0 0 0 11 0 0 ...
##  $ Steps27     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps28     : num [1:21645] 8 0 0 0 0 0 0 0 0 0 ...
##  $ Steps29     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps30     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps31     : num [1:21645] 0 0 0 0 0 0 0 9 0 0 ...
##  $ Steps32     : num [1:21645] 8 0 0 0 0 0 0 0 0 0 ...
##  $ Steps33     : num [1:21645] 6 0 0 0 0 0 0 0 0 0 ...
##  $ Steps34     : num [1:21645] 0 0 0 11 0 0 0 0 0 0 ...
##  $ Steps35     : num [1:21645] 0 0 0 9 0 0 0 0 0 0 ...
##  $ Steps36     : num [1:21645] 0 0 0 6 0 0 0 0 45 21 ...
##  $ Steps37     : num [1:21645] 0 0 0 0 0 0 0 0 0 39 ...
##  $ Steps38     : num [1:21645] 0 0 0 0 0 0 0 0 7 84 ...
##  $ Steps39     : num [1:21645] 0 0 0 0 0 0 0 0 0 117 ...
```

```
##  $ Steps40     : num [1:21645] 0 0 0 0 0 0 0 0 0 22 ...
##  $ Steps41     : num [1:21645] 0 0 0 0 0 0 0 0 28 0 0 ...
##  $ Steps42     : num [1:21645] 0 0 0 0 0 0 0 0 7 31 0 ...
##  $ Steps43     : num [1:21645] 0 0 0 0 0 0 0 0 0 20 0 ...
##  $ Steps44     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps45     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 122 ...
##  $ Steps46     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 125 ...
##  $ Steps47     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 91 ...
##  $ Steps48     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 73 ...
##  $ Steps49     : num [1:21645] 0 0 0 0 0 0 0 0 0 19 0 ...
##  $ Steps50     : num [1:21645] 0 0 0 0 0 0 0 0 16 0 0 ...
##  $ Steps51     : num [1:21645] 9 0 0 0 0 0 0 0 13 21 8 ...
##  $ Steps52     : num [1:21645] 8 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps53     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps54     : num [1:21645] 20 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps55     : num [1:21645] 1 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps56     : num [1:21645] 0 0 0 0 0 0 0 0 0 31 0 ...
##  $ Steps57     : num [1:21645] 0 0 0 0 0 0 0 0 0 0 0 ...
##  $ Steps58     : num [1:21645] 0 0 0 0 0 0 0 0 0 42 0 ...
##  $ Steps59     : num [1:21645] 0 0 0 0 0 0 0 0 16 2 105 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityHour = col_character(),
##   ..   Steps00 = col_double(),
##   ..   Steps01 = col_double(),
##   ..   Steps02 = col_double(),
##   ..   Steps03 = col_double(),
##   ..   Steps04 = col_double(),
##   ..   Steps05 = col_double(),
##   ..   Steps06 = col_double(),
##   ..   Steps07 = col_double(),
##   ..   Steps08 = col_double(),
##   ..   Steps09 = col_double(),
##   ..   Steps10 = col_double(),
##   ..   Steps11 = col_double(),
##   ..   Steps12 = col_double(),
##   ..   Steps13 = col_double(),
##   ..   Steps14 = col_double(),
##   ..   Steps15 = col_double(),
##   ..   Steps16 = col_double(),
##   ..   Steps17 = col_double(),
##   ..   Steps18 = col_double(),
##   ..   Steps19 = col_double(),
##   ..   Steps20 = col_double(),
##   ..   Steps21 = col_double(),
##   ..   Steps22 = col_double(),
##   ..   Steps23 = col_double(),
##   ..   Steps24 = col_double(),
##   ..   Steps25 = col_double(),
##   ..   Steps26 = col_double(),
##   ..   Steps27 = col_double(),
##   ..   Steps28 = col_double(),
##   ..   Steps29 = col_double(),
```

```
##   ..   Steps30 = col_double(),
##   ..   Steps31 = col_double(),
##   ..   Steps32 = col_double(),
##   ..   Steps33 = col_double(),
##   ..   Steps34 = col_double(),
##   ..   Steps35 = col_double(),
##   ..   Steps36 = col_double(),
##   ..   Steps37 = col_double(),
##   ..   Steps38 = col_double(),
##   ..   Steps39 = col_double(),
##   ..   Steps40 = col_double(),
##   ..   Steps41 = col_double(),
##   ..   Steps42 = col_double(),
##   ..   Steps43 = col_double(),
##   ..   Steps44 = col_double(),
##   ..   Steps45 = col_double(),
##   ..   Steps46 = col_double(),
##   ..   Steps47 = col_double(),
##   ..   Steps48 = col_double(),
##   ..   Steps49 = col_double(),
##   ..   Steps50 = col_double(),
##   ..   Steps51 = col_double(),
##   ..   Steps52 = col_double(),
##   ..   Steps53 = col_double(),
##   ..   Steps54 = col_double(),
##   ..   Steps55 = col_double(),
##   ..   Steps56 = col_double(),
##   ..   Steps57 = col_double(),
##   ..   Steps58 = col_double(),
##   ..   Steps59 = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(METsNarrow_Minute)
```

```
## spc_tbl_ [1,325,580 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id            : num [1:1325580] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityMinute: chr [1:1325580] "4/12/2016 12:00:00 AM" "4/12/2016 12:01:00 AM" "4/12/2016 12:02:0
##  $ METs          : num [1:1325580] 10 10 10 10 10 12 12 12 12 12 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityMinute = col_character(),
##   ..   METs = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(weight_LogInfo)
```

```
## spc_tbl_ [67 x 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id            : num [1:67] 1.50e+09 1.50e+09 1.93e+09 2.87e+09 2.87e+09 ...
##  $ Date          : chr [1:67] "5/2/2016 11:59:59 PM" "5/3/2016 11:59:59 PM" "4/13/2016 1:08:52 AM" "4
##  $ WeightKg      : num [1:67] 52.6 52.6 133.5 56.7 57.3 ...
##  $ WeightPounds  : num [1:67] 116 116 294 125 126 ...
##  $ Fat           : num [1:67] 22 NA NA NA NA 25 NA NA NA NA ...
```

```
##  $ BMI           : num [1:67] 22.6 22.6 47.5 21.5 21.7 ...
##  $ IsManualReport: logi [1:67] TRUE TRUE FALSE TRUE TRUE TRUE ...
##  $ LogId         : num [1:67] 1.46e+12 1.46e+12 1.46e+12 1.46e+12 1.46e+12 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   Date = col_character(),
##   ..   WeightKg = col_double(),
##   ..   WeightPounds = col_double(),
##   ..   Fat = col_double(),
##   ..   BMI = col_double(),
##   ..   IsManualReport = col_logical(),
##   ..   LogId = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

4.4 Understanding some summary statistics

- Count the number of unique participants that are there in each dataframe.

```
n_distinct(Activity_Day$Id)
```

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

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

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

```
n_distinct(Sleep_Time$Id)
```

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

```
n_distinct(Intensities_Day$Id)
```

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

```
n_distinct(Intensities_Hour$Id)
```

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

```
n_distinct(IntensitiesNarrow_Minute$Id)
```

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

```
n_distinct(IntensitiesWide_Minute$Id)
```

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

```
n_distinct(Calories_Day$Id)
```

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

```
n_distinct(Calories_Hour$Id)
```

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

```
n_distinct(CaloriesNarrow_Minute$Id)
```

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

```
n_distinct(CaloriesWide_Minute$Id)
```

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

```
n_distinct(Steps_Day$Id)
```

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

```
n_distinct(Steps_Hour$Id)
```

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

```
n_distinct(StepsNarrow_Minute$Id)
```

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

```
n_distinct(StepsWide_Minute$Id)
```

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

```
n_distinct(METsNarrow_Minute$Id)
```

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

```
n_distinct(weight_LogInfo$Id)
```

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

- The number of observations that are in each dataframe.

```
nrow(Activity_Day)
```

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

```
nrow(Sleep_Day)
```

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

```
nrow(Sleep_Time)
```

```
## [1] 188521
```

```
nrow(Intensities_Day)
```

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

```
nrow(Intensities_Hour)
```

```
## [1] 22099
```

```
nrow(IntensitiesNarrow_Minute)
```

```
## [1] 1325580
```

```
nrow(IntensitiesWide_Minute)
```

```
## [1] 21645
```

```
nrow(Calories_Day)
```

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

```
nrow(Calories_Hour)
```

```
## [1] 22099
```

```
nrow(CaloriesNarrow_Minute)
```

```
## [1] 1325580
```

```
nrow(CaloriesWide_Minute)
```

```
## [1] 21645
```

```
nrow(Steps_Day)
```

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

```
nrow(Steps_Hour)
```

```
## [1] 22099
```

```
nrow(StepsNarrow_Minute)
```

```
## [1] 1325580
```

```
nrow(StepsWide_Minute)
```

```
## [1] 21645
```

```
nrow(METsNarrow_Minute)
```

```
## [1] 1325580
```

```
nrow(weight_LogInfo)
```

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

4.5 Cleaning and formatting:

Now that we got to know more about our data structures we will process them to look for any errors and inconsistencies.

4.5.1 Veryfying number of users

- Firstly, we want to make sure how many unique users are per data frame. Even though 30 is the minimal sample size.

```
library(skimr)
n_unique(Activity_Day$Id)
```

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

```
n_unique(Sleep_Day$Id)
```

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

```
n_unique(Sleep_Time$Id)
```

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

```
n_unique(Intensities_Day$Id)
```

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

```
n_unique(Intensities_Hour$Id)
```

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

```
n_unique(IntensitiesNarrow_Minute$Id)
```

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

```
n_unique(IntensitiesWide_Minute$Id)
```

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

```
n_unique(Calories_Day$Id)
```

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

```
n_unique(Calories_Hour$Id)
```

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

```
n_unique(CaloriesNarrow_Minute$Id)
```

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

```
n_unique(CaloriesWide_Minute$Id)
```

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

```
n_unique(Steps_Day$Id)
```

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

```
n_unique(Steps_Hour$Id)
```

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

```
n_unique(StepsNarrow_Minute$Id)
```

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

```
n_unique(StepsWide_Minute$Id)
```

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

```
n_unique(METsNarrow_Minute$Id)
```

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

```
n_unique(weight_LogInfo$Id)
```

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

- Duplicates, We will now look for any duplicates:

```
sum(duplicated(Activity_Day))
```

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

```
sum(duplicated(Sleep_Day))
```

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

```
sum(duplicated(Sleep_Time))
```

```
## [1] 543
```

```
sum(duplicated(Intensities_Day))
```

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

```
sum(duplicated(Intensities_Hour))
```

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

```r
sum(duplicated(Calories_Day))
```

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

```r
sum(duplicated(Calories_Hour))
```

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

```r
sum(duplicated(Steps_Day))
```

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

```r
sum(duplicated(Steps_Hour))
```

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

```r
sum(duplicated(weight_LogInfo))
```

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

- Remove duplicates and N/A

```r
Activity_Day <- Activity_Day %>%
  distinct() %>%
  drop_na()

Sleep_Day <- Sleep_Day %>%
  distinct() %>%
  drop_na()

Sleep_Time <- Sleep_Time %>%
  distinct() %>%
  drop_na()

Intensities_Day <- Intensities_Day %>%
  distinct() %>%
  drop_na()

Intensities_Hour <- Intensities_Hour %>%
  distinct() %>%
  drop_na()

Calories_Day <- Calories_Day %>%
  distinct() %>%
  drop_na()

Calories_Hour <- Calories_Hour %>%
  distinct() %>%
  drop_na()

Steps_Day <- Steps_Day %>%
  distinct() %>%
  drop_na()

Steps_Hour <- Steps_Hour %>%
  distinct() %>%
  drop_na()
```

```
weight_LogInfo <- weight_LogInfo %>%
  distinct() %>%
  drop_na()

sum(duplicated(Activity_Day))
```

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

```
sum(duplicated(Sleep_Day))
```

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

```
sum(duplicated(Sleep_Time))
```

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

```
sum(duplicated(Intensities_Day))
```

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

```
sum(duplicated(Intensities_Hour))
```

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

```
sum(duplicated(Calories_Day))
```

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

```
sum(duplicated(Calories_Hour))
```

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

```
sum(duplicated(Steps_Day))
```

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

```
sum(duplicated(Steps_Hour))
```

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

```
sum(duplicated(weight_LogInfo))
```

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

- Let's have some quick summary statistics of the data sets:

```
## activity
library(dplyr)
Activity_Day %>%
  select(TotalSteps, TotalDistance, SedentaryMinutes, Calories) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes    Calories
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :   0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:1828
##  Median : 7406   Median : 5.245   Median :1057.5   Median :2134
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   :2304
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.:2793
##  Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :4900
```

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

```
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0
##  Median :  4.00    Median :  6.00      Median :199.0
##  Mean   : 21.16    Mean   : 13.56      Mean   :192.8
##  3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0
##  Max.   :210.00    Max.   :143.00      Max.   :518.0
```

```
## sleep
Sleep_Day %>%
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :1.00      Min.   : 58.0      Min.   : 61.0
##  1st Qu.:1.00      1st Qu.:361.0      1st Qu.:403.8
##  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
```

```
## calories
Calories_Day %>%
  select(Calories) %>%
  summary()
```

```
##     Calories
##  Min.   :   0
##  1st Qu.:1828
##  Median :2134
##  Mean   :2304
##  3rd Qu.:2793
##  Max.   :4900
```

```
CaloriesWide_Minute %>%
  select(Calories10, Calories20, Calories30, Calories40 ) %>%
  summary()
```

```
##    Calories10        Calories20        Calories30        Calories40
##  Min.   : 0.7027   Min.   : 0.7027   Min.   : 0.7027   Min.   : 0.7027
##  1st Qu.: 0.9357   1st Qu.: 0.9357   1st Qu.: 0.9357   1st Qu.: 0.9357
##  Median : 1.2176   Median : 1.2176   Median : 1.2176   Median : 1.2176
##  Mean   : 1.6212   Mean   : 1.6151   Mean   : 1.6238   Mean   : 1.6261
##  3rd Qu.: 1.4327   3rd Qu.: 1.4327   3rd Qu.: 1.4327   3rd Qu.: 1.4327
##  Max.   :17.4391   Max.   :16.3037   Max.   :17.3472   Max.   :19.7499
```

```
##  Intensities
Intensities_Day %>%
  select(SedentaryActiveDistance, LightActiveDistance, ModeratelyActiveDistance, VeryActiveDistance) %>%
  summary()
```

```
##  SedentaryActiveDistance LightActiveDistance ModeratelyActiveDistance
##  Min.   :0.000000        Min.   : 0.000      Min.   :0.0000
```

```
##  1st Qu.:0.000000        1st Qu.: 1.945      1st Qu.:0.0000
##  Median :0.000000        Median : 3.365      Median :0.2400
##  Mean   :0.001606        Mean   : 3.341      Mean   :0.5675
##  3rd Qu.:0.000000        3rd Qu.: 4.782      3rd Qu.:0.8000
##  Max.   :0.110000        Max.   :10.710      Max.   :6.4800
##  VeryActiveDistance
##  Min.   : 0.000
##  1st Qu.: 0.000
##  Median : 0.210
##  Mean   : 1.503
##  3rd Qu.: 2.053
##  Max.   :21.920
```

```
Intensities_Hour %>%
  select(AverageIntensity) %>%
  summary()
```

```
##  AverageIntensity
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0500
##  Mean   :0.2006
##  3rd Qu.:0.2667
##  Max.   :3.0000
```

```
IntensitiesWide_Minute %>%
  select(Intensity10, Intensity20, Intensity30, Intensity40, Intensity50) %>%
  summary()
```

```
##   Intensity10      Intensity20      Intensity30      Intensity40
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.000
##  Mean   :0.2014   Mean   :0.1998   Mean   :0.2016   Mean   :0.203
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :3.0000   Max.   :3.0000   Max.   :3.0000   Max.   :3.000
##   Intensity50
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean   :0.201
##  3rd Qu.:0.000
##  Max.   :3.000
```

```
## Steps
StepsWide_Minute %>%
  select(Steps10, Steps20, Steps30, Steps40, Steps50) %>%
  summary()
```

```
##     Steps10           Steps20           Steps30           Steps40
##  Min.   :  0.000   Min.   :  0.000   Min.   :  0.000   Min.   :  0.000
##  1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.:  0.000
##  Median :  0.000   Median :  0.000   Median :  0.000   Median :  0.000
##  Mean   :  5.343   Mean   :  5.296   Mean   :  5.395   Mean   :  5.381
##  3rd Qu.:  0.000   3rd Qu.:  0.000   3rd Qu.:  0.000   3rd Qu.:  0.000
##  Max.   :180.000   Max.   :179.000   Max.   :181.000   Max.   :184.000
```

```
##     Steps50
##  Min.   :  0.000
##  1st Qu.:  0.000
##  Median :  0.000
##  Mean   :  5.329
##  3rd Qu.:  0.000
##  Max.   :182.000
```

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

```
##     WeightKg          BMI
##  Min.   :52.60   Min.   :22.65
##  1st Qu.:57.55   1st Qu.:23.85
##  Median :62.50   Median :25.05
##  Mean   :62.50   Mean   :25.05
##  3rd Qu.:67.45   3rd Qu.:26.25
##  Max.   :72.40   Max.   :27.45
```

4.5.2 Clean and rename columns:

- We want to ensure that column names are using right syntax and same format in all datasets since we will merge them later on. We are changing the format of all columns to lower case.

```
library(janitor)
library(dplyr)
clean_names(Activity_Day)
```

```
## # A tibble: 940 x 15
##            id activity_date total_steps total_distance tracker_distance
##         <dbl> <chr>               <dbl>          <dbl>            <dbl>
##  1 1503960366 4/12/2016           13162           8.5              8.5
##  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
##  7 1503960366 4/18/2016           13019           8.59             8.59
##  8 1503960366 4/19/2016           15506           9.88             9.88
##  9 1503960366 4/20/2016           10544           6.68             6.68
## 10 1503960366 4/21/2016            9819           6.34             6.34
## # i 930 more rows
## # i 10 more variables: logged_activities_distance <dbl>,
## #   very_active_distance <dbl>, moderately_active_distance <dbl>,
## #   light_active_distance <dbl>, sedentary_active_distance <dbl>,
## #   very_active_minutes <dbl>, fairly_active_minutes <dbl>,
## #   lightly_active_minutes <dbl>, sedentary_minutes <dbl>, calories <dbl>
```

```
Activity_Day <- rename_with(Activity_Day, tolower)

clean_names(Sleep_Day)
```

```
## # A tibble: 410 x 5
##          id sleep_day total_sleep_records total_minutes_asleep total_time_in_bed
##       <dbl> <chr>                   <dbl>                <dbl>              <dbl>
```

```
##  1   1.50e9 4/12/201~                  1                 327                 346
##  2   1.50e9 4/13/201~                  2                 384                 407
##  3   1.50e9 4/15/201~                  1                 412                 442
##  4   1.50e9 4/16/201~                  2                 340                 367
##  5   1.50e9 4/17/201~                  1                 700                 712
##  6   1.50e9 4/19/201~                  1                 304                 320
##  7   1.50e9 4/20/201~                  1                 360                 377
##  8   1.50e9 4/21/201~                  1                 325                 364
##  9   1.50e9 4/23/201~                  1                 361                 384
## 10   1.50e9 4/24/201~                  1                 430                 449
## # i 400 more rows
```

```
Sleep_Day <- rename_with(Sleep_Day, tolower)

clean_names(Sleep_Time)
```

```
## # A tibble: 187,978 x 4
##            id date                  value      log_id
##         <dbl> <chr>                 <dbl>       <dbl>
##  1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
##  2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
##  3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
##  4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
##  5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
##  6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
##  7 1503960366 4/12/2016 2:53:30 AM     1 11380564589
##  8 1503960366 4/12/2016 2:54:30 AM     2 11380564589
##  9 1503960366 4/12/2016 2:55:30 AM     2 11380564589
## 10 1503960366 4/12/2016 2:56:30 AM     2 11380564589
## # i 187,968 more rows
```

```
Sleep_Time <- rename_with(Sleep_Time, tolower)

clean_names(Intensities_Day)
```

```
## # A tibble: 940 x 10
##            id activity_day sedentary_minutes lightly_active_minutes
##         <dbl> <chr>                    <dbl>                  <dbl>
##  1 1503960366 4/12/2016                  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
##  7 1503960366 4/18/2016                 1149                    233
##  8 1503960366 4/19/2016                  775                    264
##  9 1503960366 4/20/2016                  818                    205
## 10 1503960366 4/21/2016                  838                    211
## # i 930 more rows
## # i 6 more variables: fairly_active_minutes <dbl>, very_active_minutes <dbl>,
## #   sedentary_active_distance <dbl>, light_active_distance <dbl>,
## #   moderately_active_distance <dbl>, very_active_distance <dbl>
```

```
Intensities_Day <- rename_with(Intensities_Day, tolower)

clean_names(Intensities_Hour)
```

```
## # A tibble: 22,099 x 4
##            id activity_hour         total_intensity average_intensity
##         <dbl> <chr>                           <dbl>             <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM              20             0.333
##  2 1503960366 4/12/2016 1:00:00 AM                8             0.133
##  3 1503960366 4/12/2016 2:00:00 AM                7             0.117
##  4 1503960366 4/12/2016 3:00:00 AM                0             0
##  5 1503960366 4/12/2016 4:00:00 AM                0             0
##  6 1503960366 4/12/2016 5:00:00 AM                0             0
##  7 1503960366 4/12/2016 6:00:00 AM                0             0
##  8 1503960366 4/12/2016 7:00:00 AM                0             0
##  9 1503960366 4/12/2016 8:00:00 AM               13             0.217
## 10 1503960366 4/12/2016 9:00:00 AM               30             0.5
## # i 22,089 more rows
```

```
Intensities_Hour <- rename_with(Intensities_Hour, tolower)

clean_names(Calories_Day)
```

```
## # A tibble: 940 x 3
##            id activity_day calories
##         <dbl> <chr>           <dbl>
##  1 1503960366 4/12/2016        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
##  7 1503960366 4/18/2016        1921
##  8 1503960366 4/19/2016        2035
##  9 1503960366 4/20/2016        1786
## 10 1503960366 4/21/2016        1775
## # i 930 more rows
```

```
Calories_Day <- rename_with(Calories_Day, tolower)

clean_names(Calories_Hour)
```

```
## # A tibble: 22,099 x 3
##            id activity_hour         calories
##         <dbl> <chr>                   <dbl>
##  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
##  7 1503960366 4/12/2016 6:00:00 AM        48
##  8 1503960366 4/12/2016 7:00:00 AM        47
##  9 1503960366 4/12/2016 8:00:00 AM        68
## 10 1503960366 4/12/2016 9:00:00 AM       141
## # i 22,089 more rows
```

```
Calories_Hour <- rename_with(Calories_Hour, tolower)

clean_names(Steps_Day)
```

```
## # A tibble: 940 x 3
##            id activity_day step_total
##         <dbl> <chr>             <dbl>
##  1 1503960366 4/12/2016         13162
##  2 1503960366 4/13/2016         10735
##  3 1503960366 4/14/2016         10460
##  4 1503960366 4/15/2016          9762
##  5 1503960366 4/16/2016         12669
##  6 1503960366 4/17/2016          9705
##  7 1503960366 4/18/2016         13019
##  8 1503960366 4/19/2016         15506
##  9 1503960366 4/20/2016         10544
## 10 1503960366 4/21/2016          9819
## # i 930 more rows
```

```
Steps_Day <- rename_with(Steps_Day, tolower)

clean_names(Steps_Hour)
```

```
## # A tibble: 22,099 x 3
##            id activity_hour         step_total
##         <dbl> <chr>                      <dbl>
##  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
##  7 1503960366 4/12/2016 6:00:00 AM           0
##  8 1503960366 4/12/2016 7:00:00 AM           0
##  9 1503960366 4/12/2016 8:00:00 AM         250
## 10 1503960366 4/12/2016 9:00:00 AM        1864
## # i 22,089 more rows
```

```
Steps_Hour <- rename_with(Steps_Hour, tolower)

clean_names(heartrate_Time)
```

```
## # A tibble: 1,048,575 x 3
##            id time           value
##         <dbl> <chr>          <dbl>
##  1 2022484408 4/12/2016 7:21    97
##  2 2022484408 4/12/2016 7:21   102
##  3 2022484408 4/12/2016 7:21   105
##  4 2022484408 4/12/2016 7:21   103
##  5 2022484408 4/12/2016 7:21   101
##  6 2022484408 4/12/2016 7:22    95
##  7 2022484408 4/12/2016 7:22    91
##  8 2022484408 4/12/2016 7:22    93
##  9 2022484408 4/12/2016 7:22    94
## 10 2022484408 4/12/2016 7:22    93
## # i 1,048,565 more rows
```

```
heartrate_Time <- rename_with(heartrate_Time, tolower)

clean_names(weight_LogInfo)
```

```
## # A tibble: 2 x 8
##           id date   weight_kg weight_pounds   fat   bmi is_manual_report  log_id
##        <dbl> <chr>      <dbl>         <dbl> <dbl> <dbl> <lgl>              <dbl>
## 1 1503960366 5/2/2~      52.6          116.    22  22.6 TRUE             1.46e12
## 2 4319703577 4/17/~      72.4          160.    25  27.5 TRUE             1.46e12
```

```
weight_LogInfo <- rename_with(weight_LogInfo, tolower)
head(weight_LogInfo)
```

```
## # A tibble: 2 x 8
##           id date       weightkg weightpounds   fat   bmi ismanualreport   logid
##        <dbl> <chr>         <dbl>        <dbl> <dbl> <dbl> <lgl>            <dbl>
## 1 1503960366 5/2/2016 ~     52.6         116.    22  22.6 TRUE           1.46e12
## 2 4319703577 4/17/2016~     72.4         160.    25  27.5 TRUE           1.46e12
```

- We will check our cleaned datasets:

```
head(Activity_Day)
```

```
## # A tibble: 6 x 15
##           id activitydate totalsteps totaldistance trackerdistance
##        <dbl> <chr>             <dbl>         <dbl>           <dbl>
## 1 1503960366 4/12/2016         13162          8.5             8.5
## 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
## # i 10 more variables: loggedactivitiesdistance <dbl>,
## #   veryactivedistance <dbl>, moderatelyactivedistance <dbl>,
## #   lightactivedistance <dbl>, sedentaryactivedistance <dbl>,
## #   veryactiveminutes <dbl>, fairlyactiveminutes <dbl>,
## #   lightlyactiveminutes <dbl>, sedentaryminutes <dbl>, calories <dbl>
```

```
head(Sleep_Day)
```

```
## # A tibble: 6 x 5
##           id sleepday        totalsleeprecords totalminutesasleep totaltimeinbed
##        <dbl> <chr>                       <dbl>              <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0~                 1                327            346
## 2 1503960366 4/13/2016 12:0~                 2                384            407
## 3 1503960366 4/15/2016 12:0~                 1                412            442
## 4 1503960366 4/16/2016 12:0~                 2                340            367
## 5 1503960366 4/17/2016 12:0~                 1                700            712
## 6 1503960366 4/19/2016 12:0~                 1                304            320
```

```
head(Sleep_Time)
```

```
## # A tibble: 6 x 4
##           id date                  value       logid
##        <dbl> <chr>                 <dbl>       <dbl>
## 1 1503960366 4/12/2016 2:47:30 AM     3 11380564589
## 2 1503960366 4/12/2016 2:48:30 AM     2 11380564589
## 3 1503960366 4/12/2016 2:49:30 AM     1 11380564589
## 4 1503960366 4/12/2016 2:50:30 AM     1 11380564589
## 5 1503960366 4/12/2016 2:51:30 AM     1 11380564589
## 6 1503960366 4/12/2016 2:52:30 AM     1 11380564589
```

- Consistency of date and time columns: Now that we have verified our column names and change them to lower case, we will focus on cleaning date-time format for Activity_Day and Sleep_Day since we will merge both data frames. Since we can disregard the time on Sleep_Day data frame we are using as_date instead as as_datetime.

```
Activity_Day <- Activity_Day %>%
  rename(date = activitydate) %>%
  mutate(date = as_date(date, format = "%m/%d/%Y"))

Sleep_Day <- Sleep_Day %>%
  rename(date = sleepday) %>%
  mutate(date = as_date(date, format = "%m/%d/%Y %I:%M:%S %p"))
head(Sleep_Day)
```

```
## # A tibble: 6 x 5
##           id date       totalsleeprecords totalminutesasleep totaltimeinbed
##        <dbl> <date>                 <dbl>              <dbl>           <dbl>
## 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
```

5. Analyze and Share:

We will analyze trends of the users of FitBit and determine if that can help us on BellaBeat's marketing strategy.

5.1 Plotting a few explorations

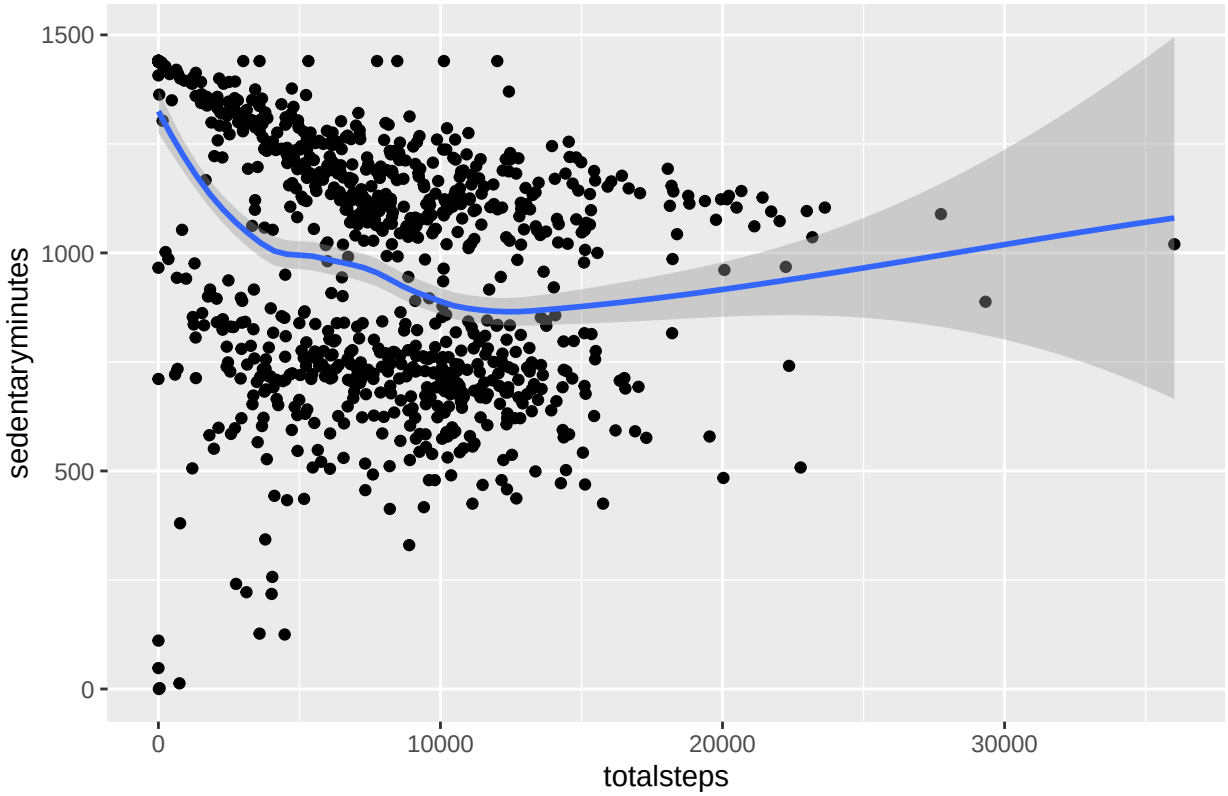
5.1.1 For Daily Activity:

- the relationship between steps taken in a day and sedentary minutes.

```
library(ggplot2)
ggplot(data=Activity_Day, aes(x=totalsteps, y=sedentaryminutes)) + geom_point() + geom_smooth() + labs(
```

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

Fig.1: Total Steps vs. Sedentary Minutes



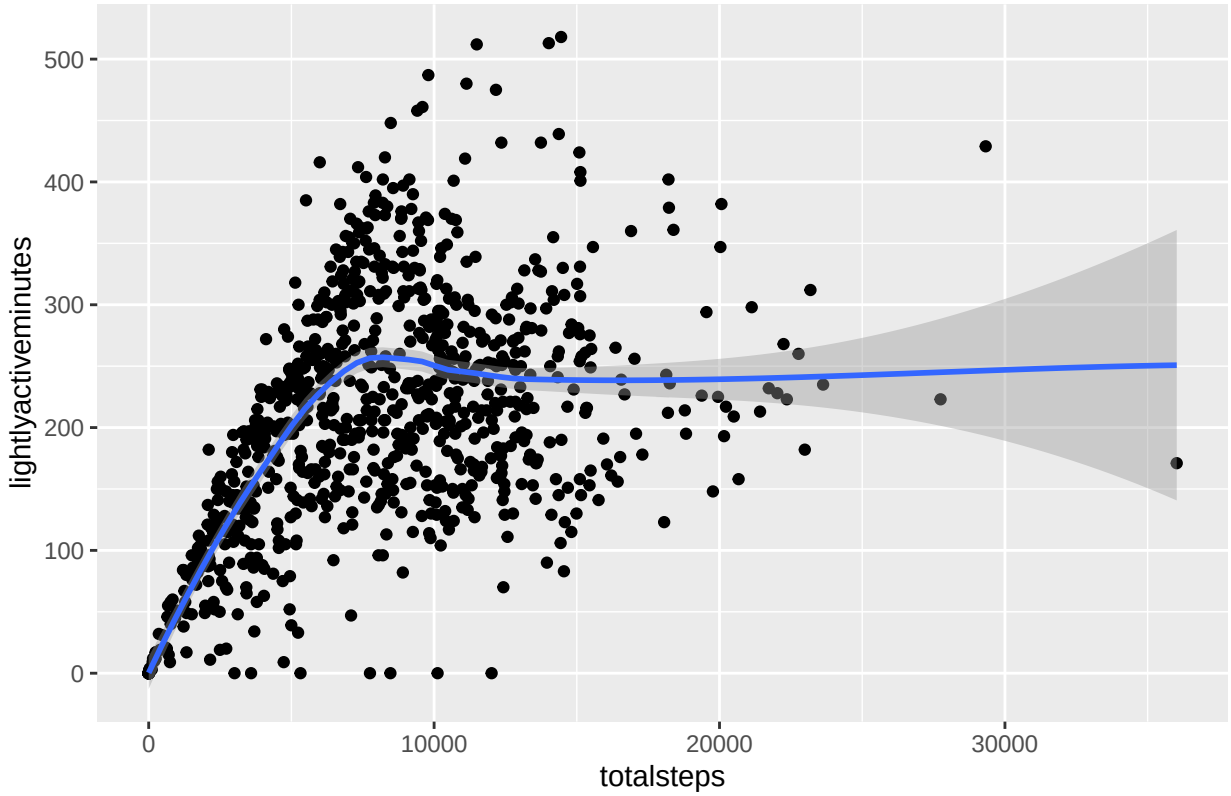
From Fig.1, there is no correlation between Activity_Day level based on steps and the Sedentary Minutes.

- the relationship between steps taken in a day and lightly active at all minutes.

```
ggplot(data=Activity_Day, aes(x=totalsteps, y=lightlyactiveminutes)) + geom_point() + geom_smooth() + la
```

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

Fig.2: Total Steps vs. Lightly−active at all minutes

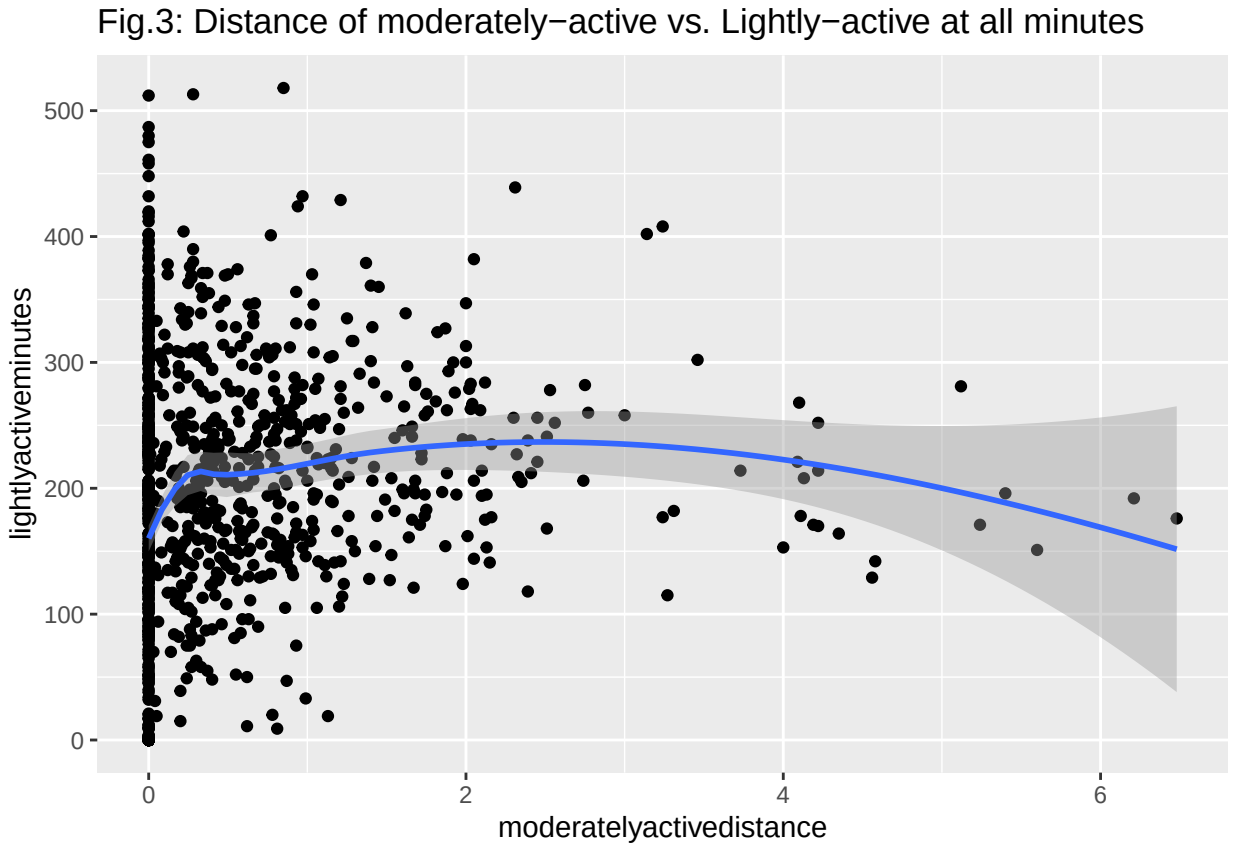


From Fig.2, we can see a positive correlation between steps and lightly active at all minutes burned.

- the relationship between the distance of moderately active and lightly active at all minutes.

```
ggplot(data=Activity_Day, aes(x=moderatelyactivedistance, y=lightlyactiveminutes)) + geom_point() + geo
```

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



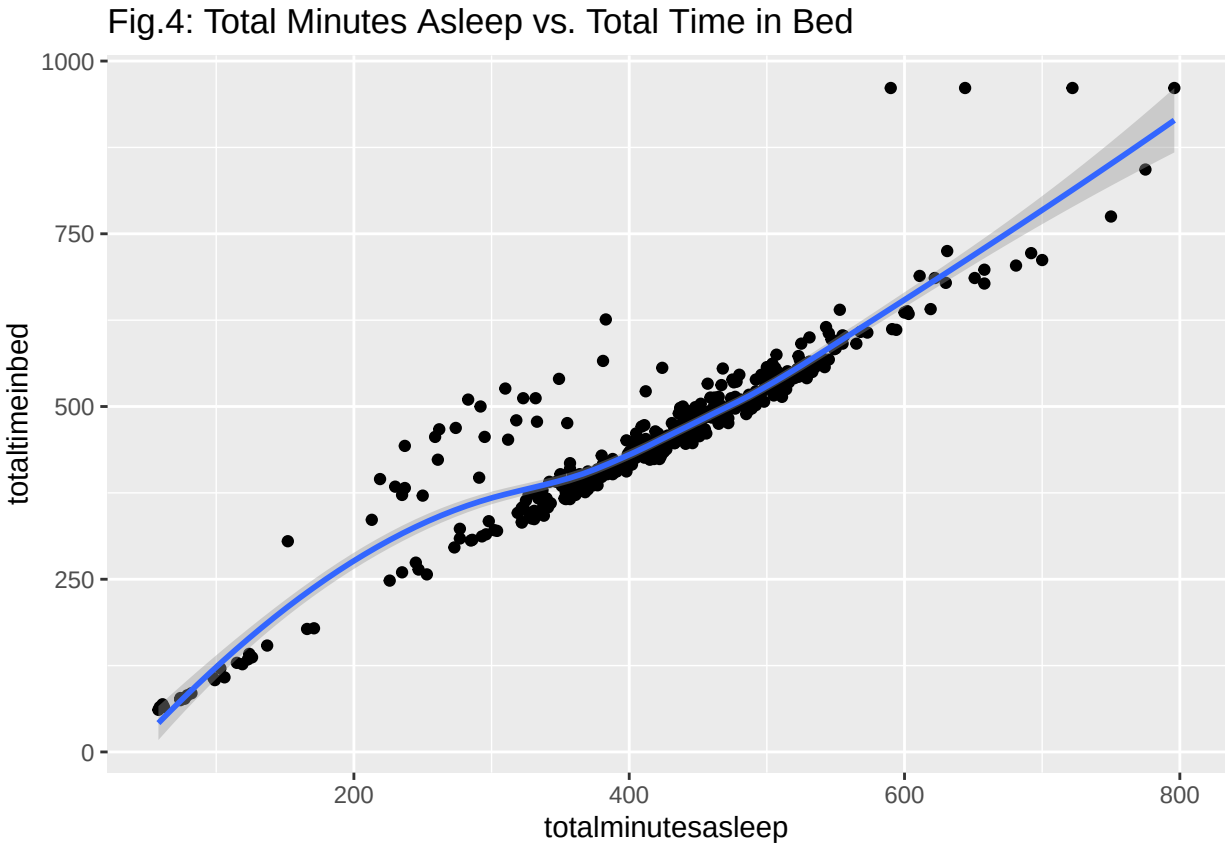
From Fig.3, there is no correlation between the Distance of moderately-active and Lightly-active at all minutes.

5.1.2 For Daily Sleep:

- the relationship between minutes asleep and time in bed.

```
ggplot(data=Sleep_Day, aes(x=totalminutesasleep, y=totaltimeinbed)) + geom_point() + geom_smooth() + la
```

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



From Fig.4, we can see a positive correlation between asleep minutes and the time in Bed burned.

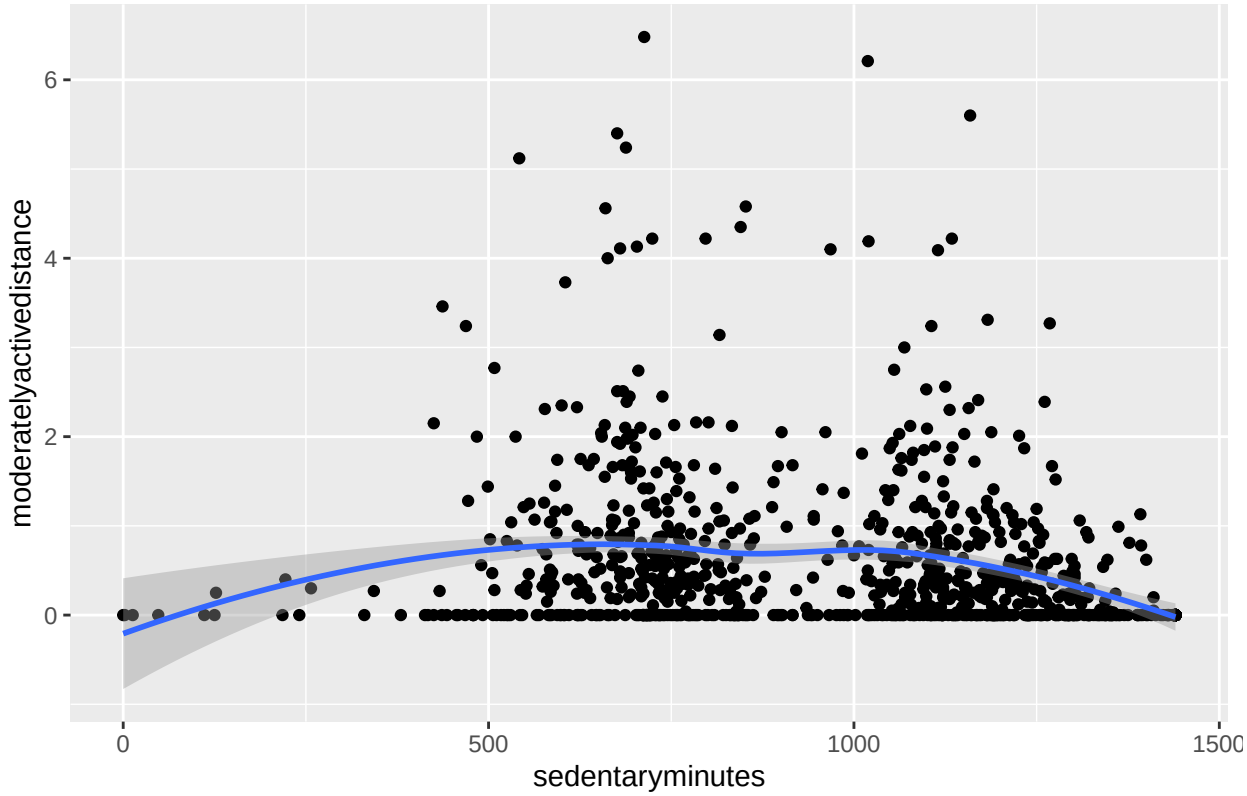
5.1.3 For Daily Intensities:

- the relationship between sedentary minutes vs. the distance of the moderately active.

```
ggplot(data= Intensities_Day, aes(x=sedentaryminutes , y=moderatelyactivedistance )) + geom_point() + g
```

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

Fig.5: Sedentary Minutes vs. Distance of moderately−active



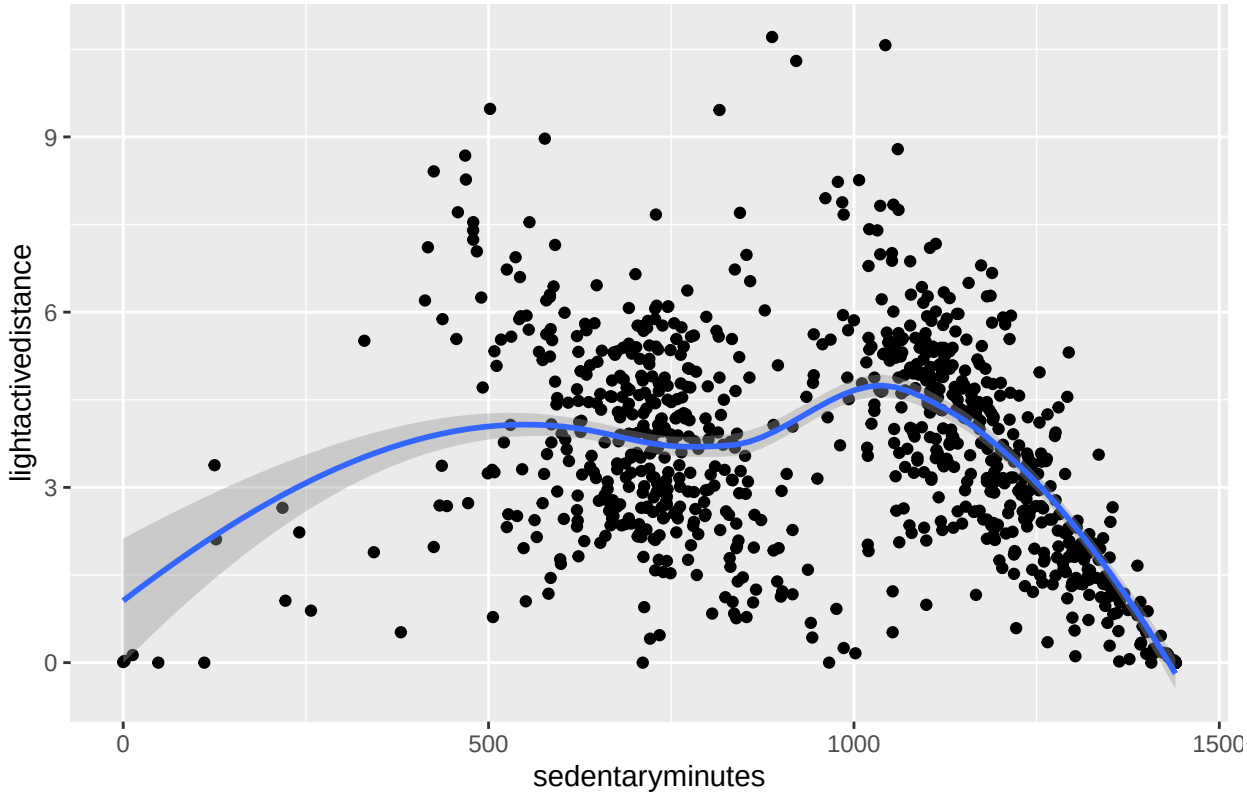
While Fig.5, we can see a little positive correlation between sedentary minutes and the distance of the moderately active.

- the relationship between sedentary minutes and the lightly active at all minutes.

```
ggplot(data= Intensities_Day, aes(x=sedentaryminutes , y=lightactivedistance )) + geom_point() + geom_sm
```

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

Fig.6: Sedentary Minutes vs. the Distance of Lightly−active

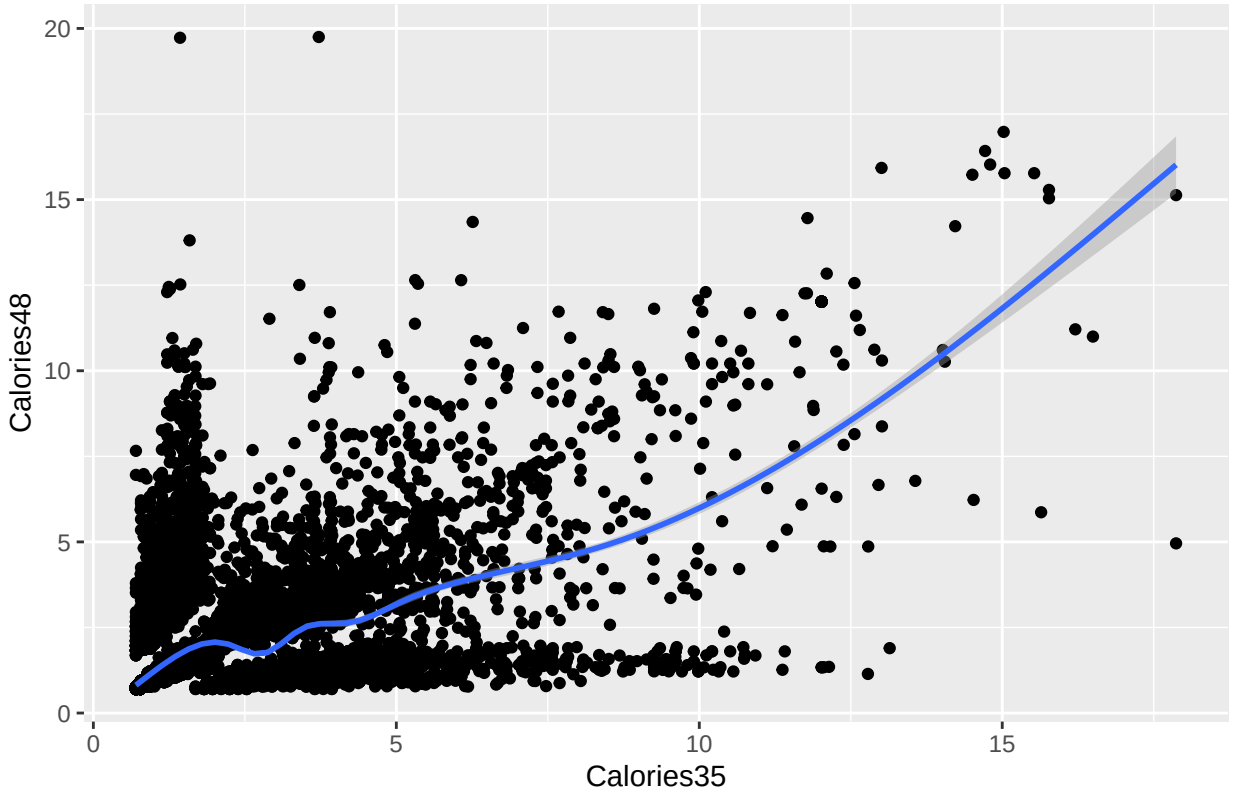


5.1.4 For Hourly Calories_Wide: - the relationship between Calories35 and Calories48.

```
#library(ggplot2)
ggplot(data= CaloriesWide_Minute, aes( x=Calories35, y=Calories48 )) + geom_point() + geom_smooth() + la
```

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

Fig.7: Calorie N.35 vs. Calorie N.48



From Fig.7, we can see a positive correlation between Calorie N.35 and Calorie N.48.

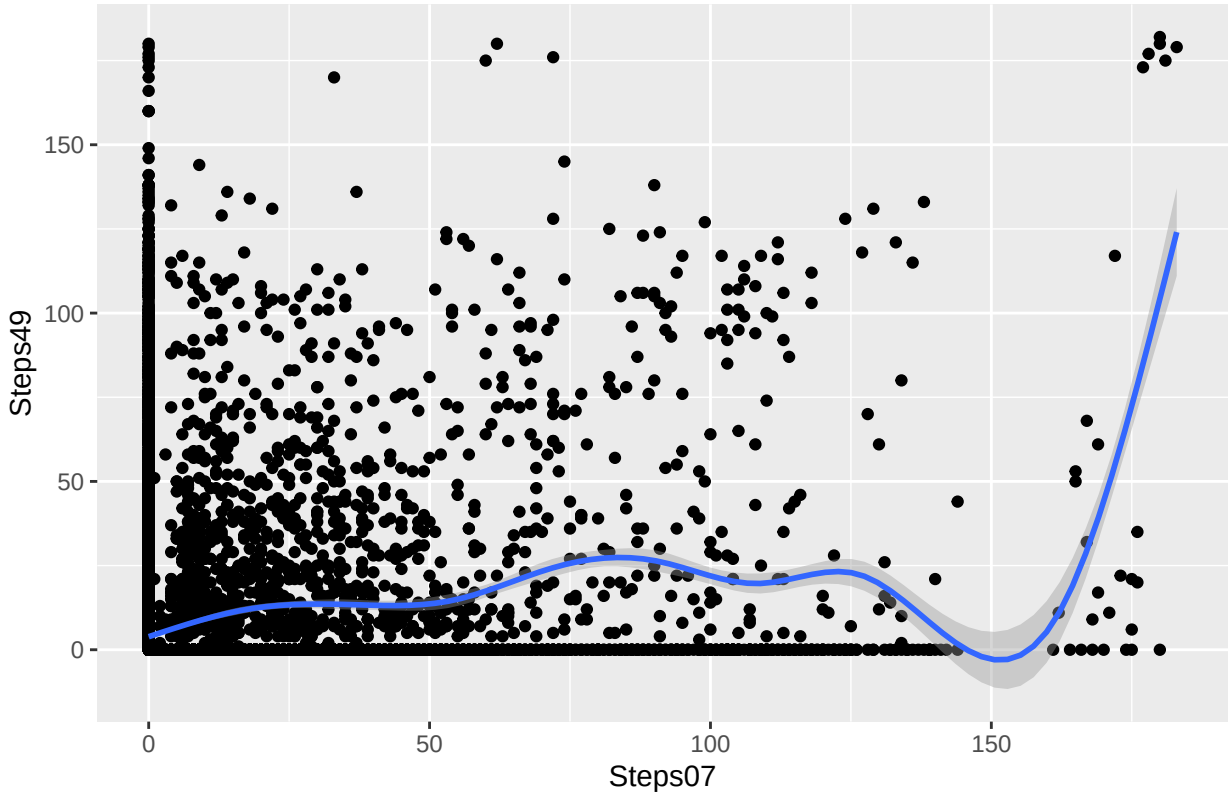
5.1.5 For Hourly Steps_Wide:

- the relationship between Steps07 and Steps49.

```
ggplot(data= StepsWide_Minute, aes( x=Steps07, y=Steps49 )) + geom_point() + geom_smooth() + labs(title=
```

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

Fig.8: Steps N.7 vs. Steps N.49



From Fig.8, we can see a little correlation between dteps N.35 and steps N.48 specialist after 150 degree.

5.2 Merging these datasets together:

- Before beginning to visualize the data, We will merge Activity_Day and Sleep_Day to see any correlation between variables by using "ID" & "DATE" as their primary keys. also, we will check the merge Activity_Day and Calories_Day to see correlation between variables by using "id","calories", then combine and print the result of merges:

```
merge_1 <- merge(Activity_Day, Sleep_Day, by = c("id", "date"))
merge_2 <- merge(Activity_Day, Calories_Day, by = c("id","calories"))

combined_data <- merge(merge_1, merge_2, by="id")
head(combined_data)
```

```
##           id     date.x totalsteps.x totaldistance.x trackerdistance.x
## 1 1503960366 2016-04-12        13162             8.5               8.5
## 2 1503960366 2016-04-12        13162             8.5               8.5
## 3 1503960366 2016-04-12        13162             8.5               8.5
## 4 1503960366 2016-04-12        13162             8.5               8.5
## 5 1503960366 2016-04-12        13162             8.5               8.5
## 6 1503960366 2016-04-12        13162             8.5               8.5
##   loggedactivitiesdistance.x veryactivedistance.x moderatelyactivedistance.x
## 1                          0                1.88                       0.55
## 2                          0                1.88                       0.55
## 3                          0                1.88                       0.55
## 4                          0                1.88                       0.55
## 5                          0                1.88                       0.55
## 6                          0                1.88                       0.55
```

```
##   lightactivedistance.x sedentaryactivedistance.x veryactiveminutes.x
## 1                  6.06                         0                  25
## 2                  6.06                         0                  25
## 3                  6.06                         0                  25
## 4                  6.06                         0                  25
## 5                  6.06                         0                  25
## 6                  6.06                         0                  25
##   fairlyactiveminutes.x lightlyactiveminutes.x sedentaryminutes.x calories.x
## 1                    13                   328                728       1985
## 2                    13                   328                728       1985
## 3                    13                   328                728       1985
## 4                    13                   328                728       1985
## 5                    13                   328                728       1985
## 6                    13                   328                728       1985
##   totalsleeprecords totalminutesasleep totaltimeinbed calories.y     date.y
## 1                 1                327            346       1783 2016-05-11
## 2                 1                327            346       1859 2016-05-10
## 3                 1                327            346       1786 2016-04-20
## 4                 1                327            346       1788 2016-04-24
## 5                 1                327            346       1797 2016-04-13
## 6                 1                327            346       1985 2016-04-12
##   totalsteps.y totaldistance.y trackerdistance.y loggedactivitiesdistance.y
## 1        12770            8.13             8.13                           0
## 2        12207            7.77             7.77                           0
## 3        10544            6.68             6.68                           0
## 4        10039            6.41             6.41                           0
## 5        10735            6.97             6.97                           0
## 6        13162            8.50             8.50                           0
##   veryactivedistance.y moderatelyactivedistance.y lightactivedistance.y
## 1                 2.56                       1.01                  4.55
## 2                 3.35                       1.16                  3.26
## 3                 1.96                       0.48                  4.24
## 4                 2.92                       0.21                  3.28
## 5                 1.57                       0.69                  4.71
## 6                 1.88                       0.55                  6.06
##   sedentaryactivedistance.y veryactiveminutes.y fairlyactiveminutes.y
## 1                         0                  36                    23
## 2                         0                  46                    31
## 3                         0                  28                    12
## 4                         0                  39                     5
## 5                         0                  21                    19
## 6                         0                  25                    13
##   lightlyactiveminutes.y sedentaryminutes.y activityday
## 1                    251               669   5/11/2016
## 2                    214               746   5/10/2016
## 3                    205               818   4/20/2016
## 4                    238               709   4/24/2016
## 5                    217               776   4/13/2016
## 6                    328               728   4/12/2016
```

- Take a look at how many participants are in this data set.

```
library(dplyr)
n_distinct(combined_data$Id)
```

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

- Find and view the Average_Day depending on the merge that is grouped by ID with the mean of totalsteps and calories.

```
Average_Day <- merge_1 %>%
  group_by(id) %>%
  summarise(mean_Steps_Day = mean(totalsteps), mean_calories_Day = mean(calories))
head(Average_Day)
```

```
## # A tibble: 6 x 3
##           id mean_Steps_Day mean_calories_Day
##        <dbl>          <dbl>             <dbl>
## 1 1503960366         12406.             1872.
## 2 1644430081          7968.             2978.
## 3 1844505072          3477              1676.
## 4 1927972279          1490              2316.
## 5 2026352035          5619.             1541.
## 6 2320127002          5079              1804
```

- Combine and print the merge result of Sleep_Time and Activity_Day.

```
Sleep_Time1 <- Sleep_Time %>%
  rename(date_time = date) %>%
  mutate(date_time = as.POSIXct(date_time, format ="%m/%d/%Y %I:%M:%S %p" , tz=Sys.timezone()))
head(Sleep_Time1)
```

```
## # A tibble: 6 x 4
##           id date_time           value       logid
##        <dbl> <dttm>              <dbl>       <dbl>
## 1 1503960366 2016-04-12 02:47:30     3 11380564589
## 2 1503960366 2016-04-12 02:48:30     2 11380564589
## 3 1503960366 2016-04-12 02:49:30     1 11380564589
## 4 1503960366 2016-04-12 02:50:30     1 11380564589
## 5 1503960366 2016-04-12 02:51:30     1 11380564589
## 6 1503960366 2016-04-12 02:52:30     1 11380564589
```

5.2.1 Hourly steps throughout the day

Getting deeper into our analysis we want to know when exactly are users more active in a day. We will use the hourly_steps data frame and separate date_time column.

```
Steps_Hour1 <- Steps_Hour %>%
  separate(activityhour, into = c("date", "time"), sep=" ")
```

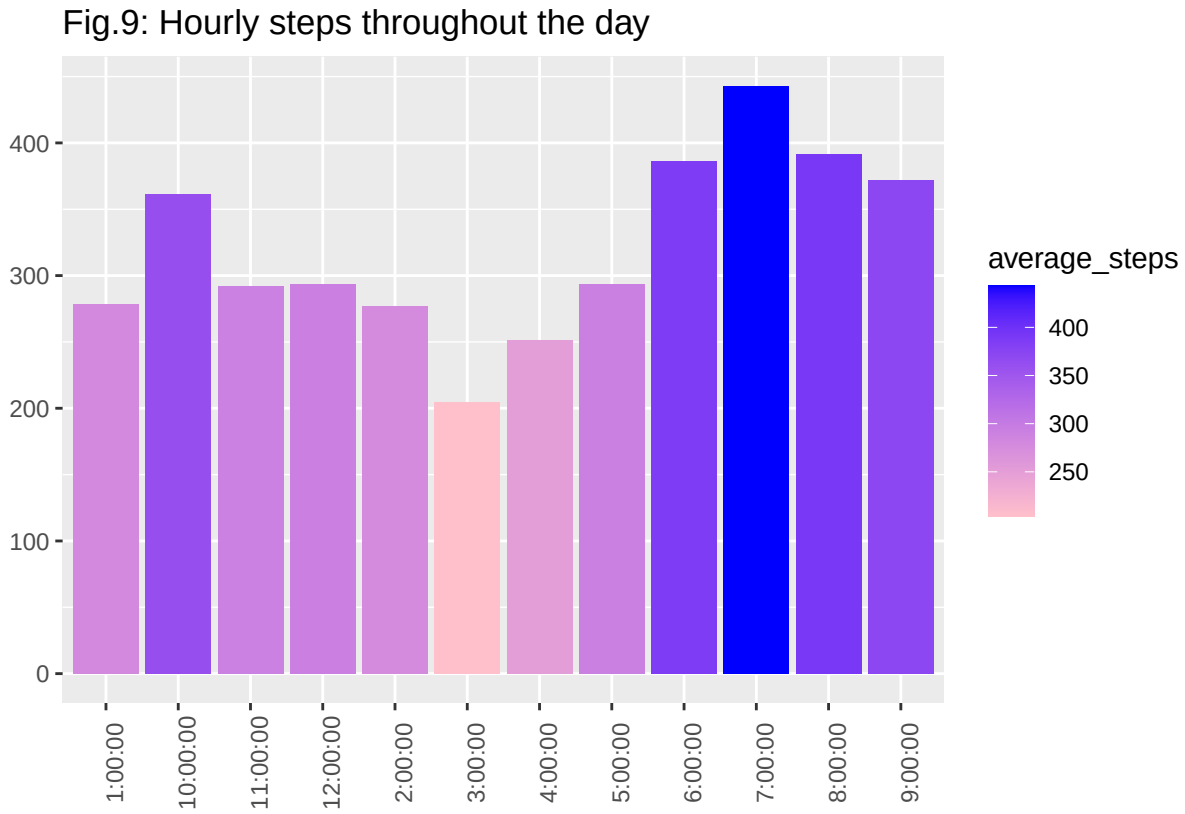
```
## Warning: Expected 2 pieces. Additional pieces discarded in 22099 rows [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
head(Steps_Hour1)
```

```
## # A tibble: 6 x 4
##           id date      time     steptotal
##        <dbl> <chr>     <chr>        <dbl>
## 1 1503960366 4/12/2016 12:00:00       373
## 2 1503960366 4/12/2016 1:00:00        160
## 3 1503960366 4/12/2016 2:00:00        151
## 4 1503960366 4/12/2016 3:00:00          0
## 5 1503960366 4/12/2016 4:00:00          0
```

```
## 6 1503960366 4/12/2016 5:00:00          0
```

```
Steps_Hour1 %>%
group_by(time) %>%
summarize(average_steps = mean(steptotal)) %>%
ggplot() +
geom_col(mapping = aes(x=time, y = average_steps, fill = average_steps)) +
labs(title = "Fig.9: Hourly steps throughout the day", x="", y="") +
scale_fill_gradient(low = "pink", high = "blue") +
theme(axis.text.x = element_text(angle = 90))
```



We can see that users are more active between 8am and 7pm. Walking more steps during lunch time from 12pm to 2pm and evenings from 5pm and 7pm.

5. Conclusion (Act):

As we already know, collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

After analyzing FitBit Fitness Tracker Data, I found some insights that would help influence Bellabeat marketing strategy:

- The relationship between Total Minutes Asleep and Total Time in Bed looks linear. So if the Bellabeat users want to improve their sleep, we should consider using notification to go to sleep.
- we can see a positive correlation between Calorie N.35 and Calorie N.48.
- we can see a positive correlation between steps and lightly active at all minutes burned.
- After visualizing Hourly steps throughout the day, I found out that people are more active between 8am and 7pm. Walking more steps during lunch time from 12pm to 2pm. And most activity happens

between 5 pm and 7 pm - I suppose, that people go to a gym or for a walk after finishing work. We can use this time in the Bellabeat app to remind and motivate users to go for a run or walk.