BellaBeat

George.B

2022-11-20

summary

1.The Business Project

- The task is to look out for trend on how people use the company smart devices and these insights can be used by BellaBeat.

Working with Rstudio.

Settings the Rstudio and loading all necessary package to work with and then importing the BellaBeat dataset into my Rstudio.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
```

```
## Warning: package 'readr' was built under R version 4.2.2
```

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

```
library(dplyr)
library(tidyr)
library(here)
```

```
## Warning: package 'here' was built under R version 4.2.2
```

```
## here() starts at C:/Users/pc/Documents
```

```
library(ggplot2)
library(colorspace)
library(readr)
```

importing the dataset.

```
library(readr)
dailyActivity_merged <- read_csv("C:/Users/pc/Desktop/Fitabase Data 4.12.16-5.12.16/dailyActivity_merge
```

```
## 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.
```

```
View(dailyActivity_merged)
sleepDay_merged <- read_csv("C:/Users/pc/Desktop/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
```

```
## 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.
```

```
View(sleepDay_merged)
weightLogInfo_merged <- read_csv("C:/Users/pc/Desktop/Fitabase Data 4.12.16-5.12.16/weightLogInfo_merge
```

```
## 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.
```

```
View(weightLogInfo_merged)
```

Let do some quick summary check of the data columns.

```
str(dailyActivity_merged)
```

```
## 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>
```

```
colnames(dailyActivity_merged)
```

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

```
str(sleepDay_merged)
```

```
## 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:(
##  $ 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>
```

```
colnames(sleepDay_merged)
```

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

```
str(weightLogInfo_merged)
```

```
## 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>
```

```
colnames(weightLogInfo_merged)
```

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

Descriptive Analyses will now beging.

selecting and cleaning the data to wrangle. In these case i will be selecting some vital columns from thr
dailyActivity_merge dataset to work with.

```
dailyActivity_merged2 <- dailyActivity_merged %>%
  select(TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, Calories) %>%
  drop_na() %>%
  summarise(TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, Calories)
```

visual representation

```
dailyActive_merge_pattern <- dailyActivity_merged2 %>%
  group_by(TotalDistance, Calories) %>%
  drop_na() %>%
  summarise(TotalDistance, Calories) %>%
  arrange(TotalDistance, Calories)
```

```
## 'summarise()' has grouped output by 'TotalDistance', 'Calories'. You can
## override using the '.groups' argument.
```

```
ggplot(data=dailyActive_merge_pattern)+
  geom_point(mapping=aes(x=TotalDistance, y=Calories, fill=Calories))+
  geom_smooth(mapping=aes(x=TotalDistance, y=Calories, fill=Calories))
```

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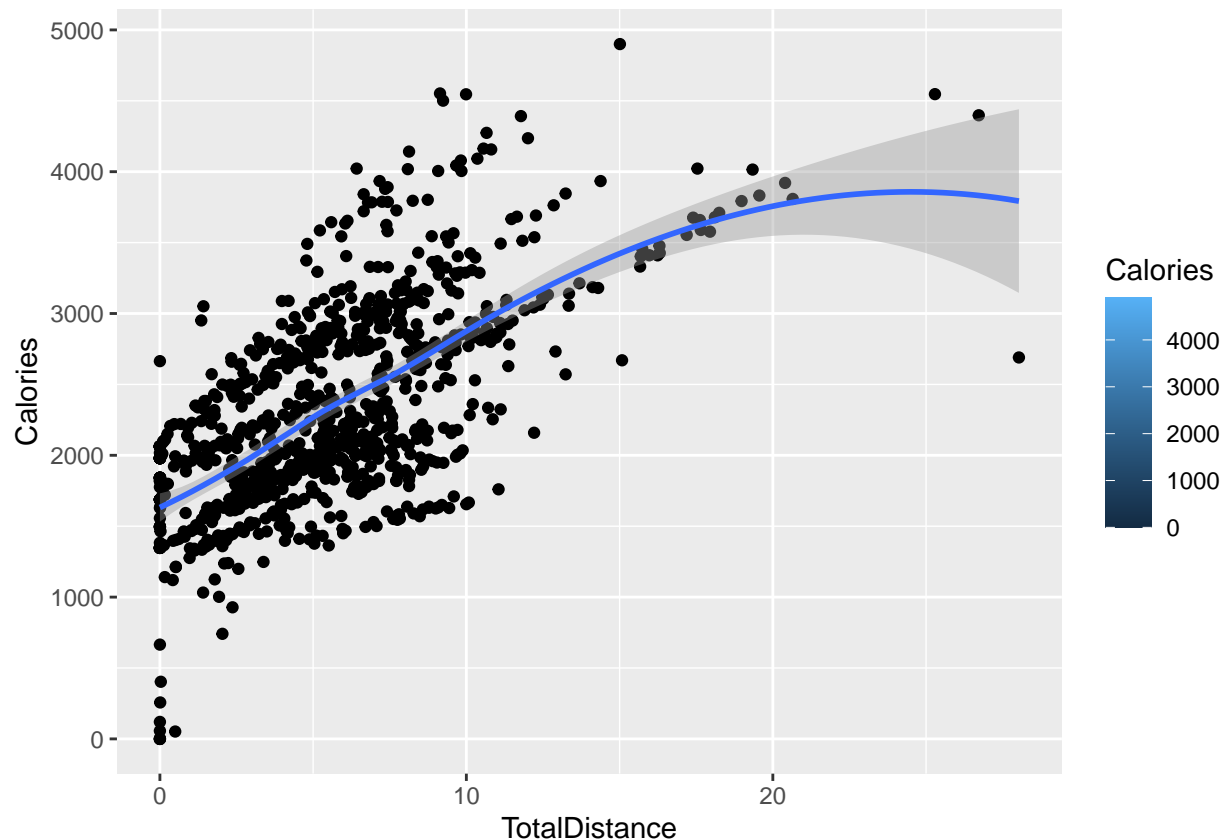

The graph about shows that the more distance covered by a user the more calories they burn out of their body, and the lesser distance covered the smaller the calories the user burn out of they body.

active records

here i am analysing the very active distance covered by each user against the time they use in covering that distance.

```
dailyActivity_merged_veryactive <- dailyActivity_merged %>%
  group_by(VeryActiveDistance, VeryActiveMinutes) %>%
  drop_na() %>%
  summarise(VeryActiveDistance, VeryActiveMinutes) %>%
  arrange(VeryActiveDistance, VeryActiveMinutes)
```

```
## 'summarise()' has grouped output by 'VeryActiveDistance', 'VeryActiveMinutes'.
## You can override using the '.groups' argument.
```

```
ggplot(data=dailyActivity_merged_veryactive)+
  geom_point(mapping=aes(x=VeryActiveDistance, y=VeryActiveMinutes))
```
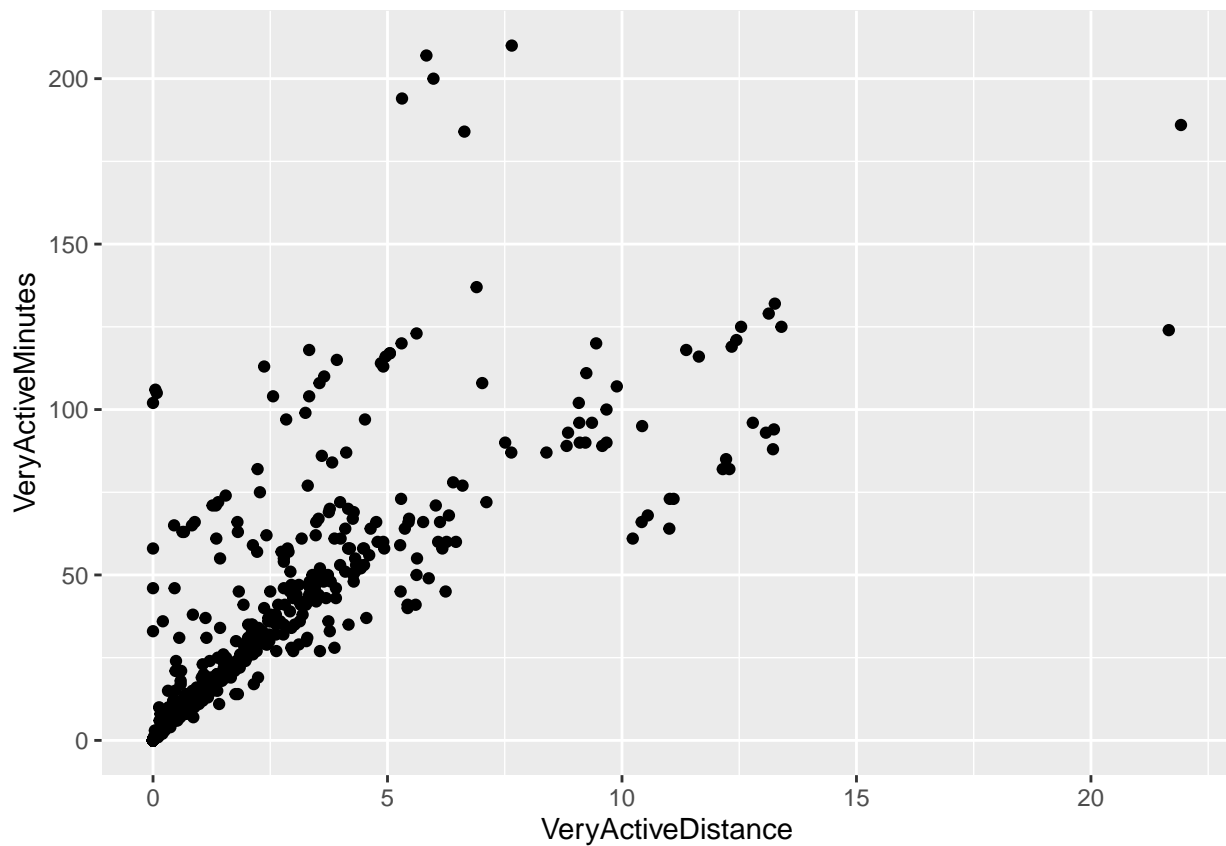


```
View(dailyActivity_merged_veryactive)
View(dailyActivity_merged_veryactive)
```

Sleepday analyses

time to wrangle the second dataset on my workspace. firstly i want to select, sort and summarise the particular columns to analyse.

```
sleepday_01 <- sleepDay_merged %>%
  separate(SleepDay, into = c('Date', 'Time'), sep= ' ')
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 413 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
print(sleepday_01)
```

```
## # A tibble: 413 x 6
##            Id Date      Time     TotalSleepRecords TotalMinutesAsleep TotalTim~1
##         <dbl> <chr>     <chr>                <dbl>              <dbl>      <dbl>
##  1 1503960366 4/12/2016 12:00:00                 1                327        346
##  2 1503960366 4/13/2016 12:00:00                 2                384        407
##  3 1503960366 4/15/2016 12:00:00                 1                412        442
##  4 1503960366 4/16/2016 12:00:00                 2                340        367
##  5 1503960366 4/17/2016 12:00:00                 1                700        712
##  6 1503960366 4/19/2016 12:00:00                 1                304        320
##  7 1503960366 4/20/2016 12:00:00                 1                360        377
##  8 1503960366 4/21/2016 12:00:00                 1                325        364
##  9 1503960366 4/23/2016 12:00:00                 1                361        384
## 10 1503960366 4/24/2016 12:00:00                 1                430        449
## # ... with 403 more rows, and abbreviated variable name 1: TotalTimeInBed
```

```
total_s_t <- sleepday_01%>%
  group_by(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  drop_na() %>%
  summarise(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed)
```

```
## 'summarise()' has grouped output by 'TotalSleepRecords', 'TotalMinutesAsleep',
## 'TotalTimeInBed'. You can override using the '.groups' argument.
```

```
print(total_s_t)
```

```
## # A tibble: 413 x 3
## # Groups:   TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed [407]
##    TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##                <dbl>              <dbl>          <dbl>
##  1                 1                 58             61
##  2                 1                 59             65
##  3                 1                 61             69
##  4                 1                 62             65
##  5                 1                 74             75
##  6                 1                 74             78
##  7                 1                 77             77
##  8                 1                 79             82
##  9                 1                 82             85
## 10                 1                 98            107
## # ... with 403 more rows
```

```
ggplot(data=sleepday_01)+
  geom_smooth(mapping=aes(x=TotalTimeInBed, y=TotalSleepRecords, group=1, color='TotalMinutesAsleep'))+
  geom_smooth(mapping=aes(x=TotalMinutesAsleep, y=TotalSleepRecords, group=2, color='TotalTimeInBed'))+
  labs(title = "BellaBeat Device: SleepingTime Vs. SleepRecords", subtitle = "Sleeping Pattern")
```

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

BellaBeat Device: SleepingTime Vs. SleepRecords
Sleeping Pattern
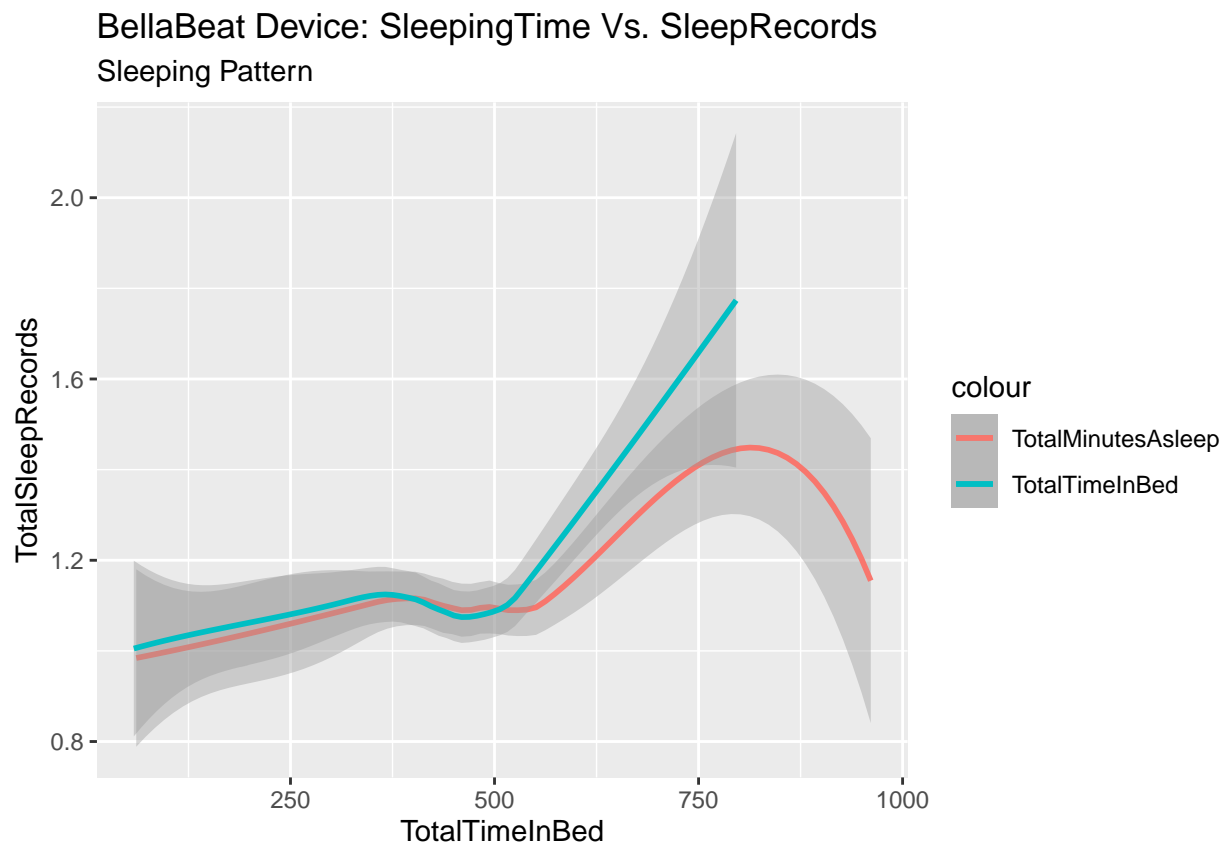
From the above presenation i analyse and visualised the relationship totalminutesasleep and totaltimeinbed by plottin the totalsleeprecords. accord to the graphical represantation, user spend more time sleeping in bed.

weight analyses

here i will be analysing the weight of the users against their body mass index(BMI)

```
weightactivity <- weightLogInfo_merged %>%
  select(WeightPounds, BMI, IsManualReport) %>%
  group_by(WeightPounds, BMI, IsManualReport) %>%
  drop_na() %>%
  summarise(WeightPounds, BMI, IsManualReport) %>%
  arrange(WeightPounds, BMI, IsManualReport)
```

```
## 'summarise()' has grouped output by 'WeightPounds', 'BMI', 'IsManualReport'.
## You can override using the '.groups' argument.
```

```
print(weightactivity)
```

```
## # A tibble: 67 x 3
```

```
## # Groups:   WeightPounds, BMI, IsManualReport [36]
##    WeightPounds   BMI IsManualReport
##           <dbl> <dbl> <lgl>
## 1          116.  22.6 TRUE
## 2          116.  22.6 TRUE
## 3          125.  21.5 TRUE
## 4          126.  21.7 TRUE
## 5          134.  23.8 TRUE
## 6          134.  23.8 TRUE
## 7          135.  23.9 TRUE
## 8          135.  23.9 TRUE
## 9          135.  23.9 TRUE
## 10         135.  23.9 TRUE
## # ... with 57 more rows
```

```
ggplot(data=weightactivity)+
  geom_smooth(mapping = aes(x=WeightPounds, y=BMI, linetype=IsManualReport))+
  labs(title = "Bella Device: WeightPOunds Vs. BMI", subtitle = "Manual Report on Users")
```

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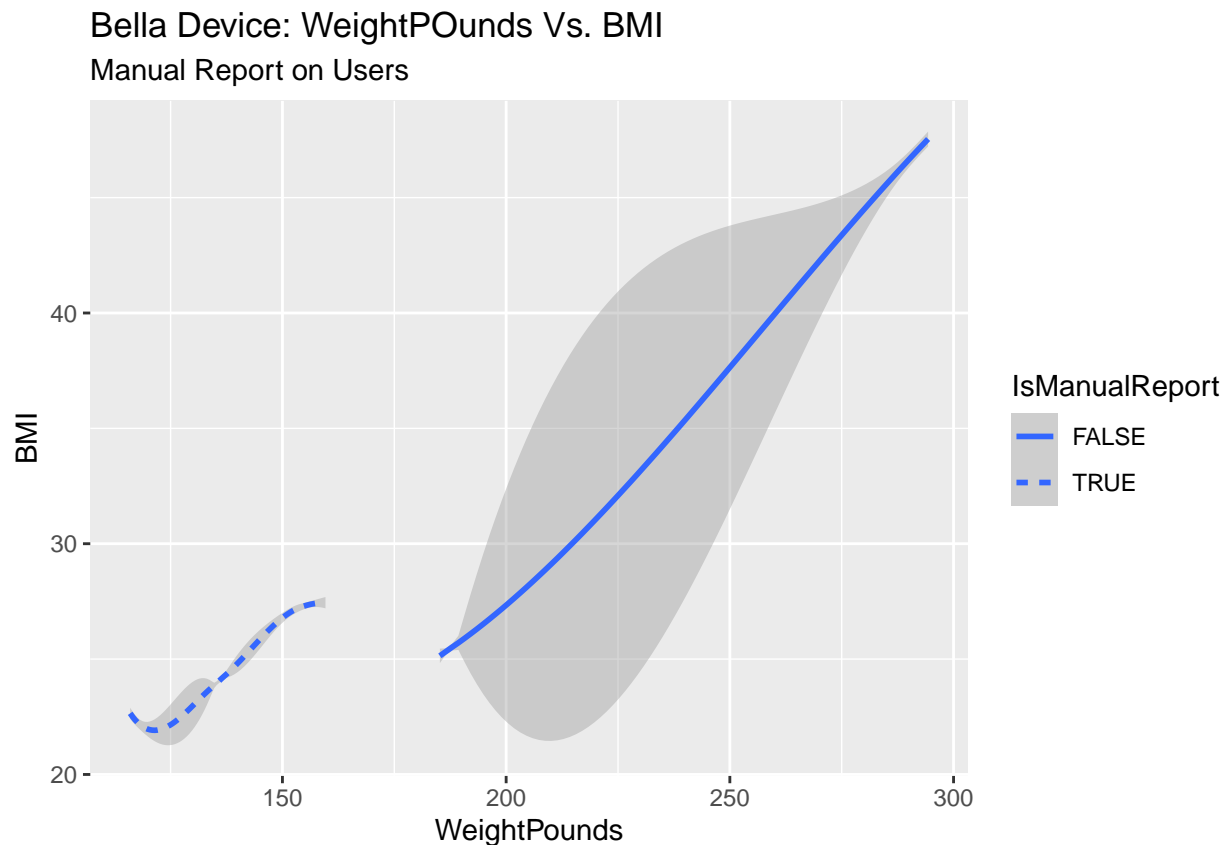

this plot showcase a correlation between users weight and their body mass, The TRUE report conforms to
the users weighting below 170poundsand having a body mass of less 30, While the FALSE report shows
that user who weigh higher most surely have a higher body mass.