

Bellabeat Fitness Case stud

D L B DIAS

2025-01-05

Introduction

Bellabeat is a high-tech company founded by Urska Srsen and Sando Mur in 2013, specializing in health-focused smart products for women. The company has positioned itself as a leader in wellness technology, offering elegant wearable products that track various health metrics, including activity, sleep, stress, and reproductive health.

Company Background

- **Founders:** Urska Srsen and Sando Mur
- **Founded:** 2013
- **Core Focus:** Manufacturing health-focused smart products for women
- **Mission:** To find trends in the way consumers use firbits in the hopes of improving bellabeats marketing strategies

Products and Features

Bellabeat's main products are designed to track key health metrics, such as:

- **Activity**
- **Sleep**
- **Stress**
- **Reproductive Health**

The products combine advanced technology with an artistic design, aiming to make health tracking both functional and aesthetically pleasing.

Loading Necessary Libraries

```
# Loading necessary libraries  
library(tidyverse) # For data manipulation and visualization
```

```
## -- 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 1.9.3      v tidyr      1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lubridate)    # For working with date and time
library(dplyr)        # For data manipulation functions
library(ggplot2)      # For creating plots
library(tidyr)        # For reshaping data
```

Importing the Data

```
# Importing the data
activity <- read.csv('/Users/bimsaradias/Downloads/FitBit Fitness Tracker Data/mturkfitbit_export_4.12.16-1.csv')
calories <- read.csv('/Users/bimsaradias/Downloads/FitBit Fitness Tracker Data/mturkfitbit_export_4.12.16-1.csv')
intensities <- read.csv('/Users/bimsaradias/Downloads/FitBit Fitness Tracker Data/mturkfitbit_export_4.12.16-1.csv')
sleep <- read.csv('/Users/bimsaradias/Downloads/FitBit Fitness Tracker Data/mturkfitbit_export_4.12.16-1.csv')
weight <- read.csv('/Users/bimsaradias/Downloads/FitBit Fitness Tracker Data/mturkfitbit_export_4.12.16-1.csv')
```

Data Cleaning

```
# Create a list of your data frames for summary analysis
data_list <- list(activity, calories, intensities, sleep, weight)

# Apply summary() to each data frame in the list to see basic statistics
lapply(data_list, summary)
```

```
## [[1]]
##      Id      ActivityDate      TotalSteps      TotalDistance
##  Min.   :1.504e+09  Length:940      Min.    :    0      Min.    : 0.000
##  1st Qu.:2.320e+09  Class :character 1st Qu.: 3790     1st Qu.: 2.620
##  Median :4.445e+09  Mode  :character Median : 7406     Median : 5.245
##  Mean   :4.855e+09                Mean   : 7638     Mean   : 5.490
##  3rd Qu.:6.962e+09                3rd Qu.:10727     3rd Qu.: 7.713
##  Max.   :8.878e+09                Max.   :36019     Max.   :28.030
##  TrackerDistance  LoggedActivitiesDistance  VeryActiveDistance
##  Min.    : 0.000    Min.    :0.0000    Min.    : 0.000
##  1st Qu.: 2.620    1st Qu.:0.0000    1st Qu.: 0.000
##  Median : 5.245    Median :0.0000    Median : 0.210
##  Mean   : 5.475    Mean   :0.1082    Mean   : 1.503
##  3rd Qu.: 7.710    3rd Qu.:0.0000    3rd Qu.: 2.053
##  Max.   :28.030    Max.   :4.9421    Max.   :21.920
##  ModeratelyActiveDistance  LightActiveDistance  SedentaryActiveDistance
##  Min.    :0.0000    Min.    : 0.000    Min.    :0.000000
##  1st Qu.:0.0000    1st Qu.: 1.945    1st Qu.:0.000000
##  Median :0.2400    Median : 3.365    Median :0.000000
```

```

## Mean :0.5675 Mean : 3.341 Mean :0.001606
## 3rd Qu.:0.8000 3rd Qu.: 4.782 3rd Qu.:0.000000
## Max. :6.4800 Max. :10.710 Max. :0.110000
## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## Min. : 0.00 Min. : 0.00 Min. : 0.0 Min. : 0.0
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:127.0 1st Qu.: 729.8
## Median : 4.00 Median : 6.00 Median :199.0 Median :1057.5
## Mean : 21.16 Mean : 13.56 Mean :192.8 Mean : 991.2
## 3rd Qu.: 32.00 3rd Qu.: 19.00 3rd Qu.:264.0 3rd Qu.:1229.5
## Max. :210.00 Max. :143.00 Max. :518.0 Max. :1440.0
## Calories
## Min. : 0
## 1st Qu.:1828
## Median :2134
## Mean :2304
## 3rd Qu.:2793
## Max. :4900
##
## [[2]]
## Id ActivityHour Calories
## Min. :1.504e+09 Length:22099 Min. : 42.00
## 1st Qu.:2.320e+09 Class :character 1st Qu.: 63.00
## Median :4.445e+09 Mode :character Median : 83.00
## Mean :4.848e+09 Mean : 97.39
## 3rd Qu.:6.962e+09 3rd Qu.:108.00
## Max. :8.878e+09 Max. :948.00
##
## [[3]]
## Id ActivityHour TotalIntensity AverageIntensity
## Min. :1.504e+09 Length:22099 Min. : 0.00 Min. :0.0000
## 1st Qu.:2.320e+09 Class :character 1st Qu.: 0.00 1st Qu.:0.0000
## Median :4.445e+09 Mode :character Median : 3.00 Median :0.0500
## Mean :4.848e+09 Mean : 12.04 Mean :0.2006
## 3rd Qu.:6.962e+09 3rd Qu.: 16.00 3rd Qu.:0.2667
## Max. :8.878e+09 Max. :180.00 Max. :3.0000
##
## [[4]]
## Id SleepDay TotalSleepRecords TotalMinutesAsleep
## Min. :1.504e+09 Length:413 Min. :1.000 Min. : 58.0
## 1st Qu.:3.977e+09 Class :character 1st Qu.:1.000 1st Qu.:361.0
## Median :4.703e+09 Mode :character Median :1.000 Median :433.0
## Mean :5.001e+09 Mean :1.119 Mean :419.5
## 3rd Qu.:6.962e+09 3rd Qu.:1.000 3rd Qu.:490.0
## Max. :8.792e+09 Max. :3.000 Max. :796.0
## TotalTimeInBed
## Min. : 61.0
## 1st Qu.:403.0
## Median :463.0
## Mean :458.6
## 3rd Qu.:526.0
## Max. :961.0
##
## [[5]]
## Id Date WeightKg WeightPounds

```

```
## Min.      :1.504e+09   Length:67      Min.      : 52.60   Min.      :116.0
## 1st Qu.:6.962e+09   Class :character 1st Qu.: 61.40   1st Qu.:135.4
## Median :6.962e+09   Mode  :character Median : 62.50   Median :137.8
## Mean    :7.009e+09           Mean    : 72.04   Mean    :158.8
## 3rd Qu.:8.878e+09           3rd Qu.: 85.05   3rd Qu.:187.5
## Max.    :8.878e+09           Max.    :133.50   Max.    :294.3
##
##      Fat      BMI      IsManualReport      LogId
## Min.      :22.00   Min.      :21.45   Length:67      Min.      :1.460e+12
## 1st Qu.:22.75   1st Qu.:23.96   Class :character 1st Qu.:1.461e+12
## Median :23.50   Median :24.39   Mode  :character Median :1.462e+12
## Mean    :23.50   Mean    :25.19           Mean    :1.462e+12
## 3rd Qu.:24.25   3rd Qu.:25.56           3rd Qu.:1.462e+12
## Max.    :25.00   Max.    :47.54           Max.    :1.463e+12
## NA's      :65
```

```
# Check if there are any missing values across all datasets
any(is.na(data_list))
```

```
## [1] FALSE
```

Data Preprocessing

```
# Convert time columns to proper date-time format
# intensities
intensities$ActivityHour = as.POSIXct(intensities$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.tz)
intensities$time <- format(intensities$ActivityHour, format = "%H:%M:%S")
intensities$date <- format(intensities$ActivityHour, format = "%m/%d/%y")

# calories
calories$ActivityHour = as.POSIXct(calories$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
calories$time <- format(calories$ActivityHour, format = "%H:%M:%S")
calories$date <- format(calories$ActivityHour, format = "%m/%d/%y")

# activity
activity$ActivityDate = as.POSIXct(activity$ActivityDate, format="%m/%d/%Y", tz=Sys.timezone())
activity$date <- format(activity$ActivityDate, format = "%Y-%m-%d")

# sleep
sleep$SleepDay=as.POSIXct(sleep$SleepDay, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())
colnames(sleep)[colnames(sleep) == "SleepDay"] <- "date"
View(sleep)
```

Exploratory Data Analysis

```
# Show a summary of key columns in activity
activity %>%
  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
```

```
# Explore the number of active minutes per category
```

```
activity %>%
```

```
  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
```

```
# Show a summary of calories data
```

```
calories %>%
```

```
  select(Calories) %>%
  summary()
```

```
##      Calories
##  Min.       : 42.00
## 1st Qu.: 63.00
## Median : 83.00
## Mean   : 97.39
## 3rd Qu.:108.00
## Max.    :948.00
```

```
# Show a summary of sleep data
```

```
sleep %>%
```

```
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##      TotalSleepRecords      TotalMinutesAsleep      TotalTimeInBed
##  Min.       :1.000      Min.       : 58.0      Min.       : 61.0
## 1st Qu.:1.000      1st Qu.:361.0      1st Qu.:403.0
## Median :1.000      Median :433.0      Median :463.0
## Mean   :1.119      Mean   :419.5      Mean    :458.6
## 3rd Qu.:1.000      3rd Qu.:490.0      3rd Qu.:526.0
## Max.    :3.000      Max.    :796.0      Max.     :961.0
```

```
# Show a summary of weight data
```

```
weight %>%
```

```
  select(WeightKg, BMI) %>%
  summary()
```

```
##      WeightKg      BMI
## Min.   : 52.60   Min.   :21.45
## 1st Qu.: 61.40   1st Qu.:23.96
## Median : 62.50   Median :24.39
## Mean   : 72.04   Mean    :25.19
## 3rd Qu.: 85.05   3rd Qu.:25.56
## Max.   :133.50   Max.    :47.54
```

Merging Data

```
# Merge the 'sleep' and 'activity' datasets based on the 'Id' and 'date' columns
merged_data <- merge(sleep, activity, by=c('Id', 'date'))
head(merged_data)
```

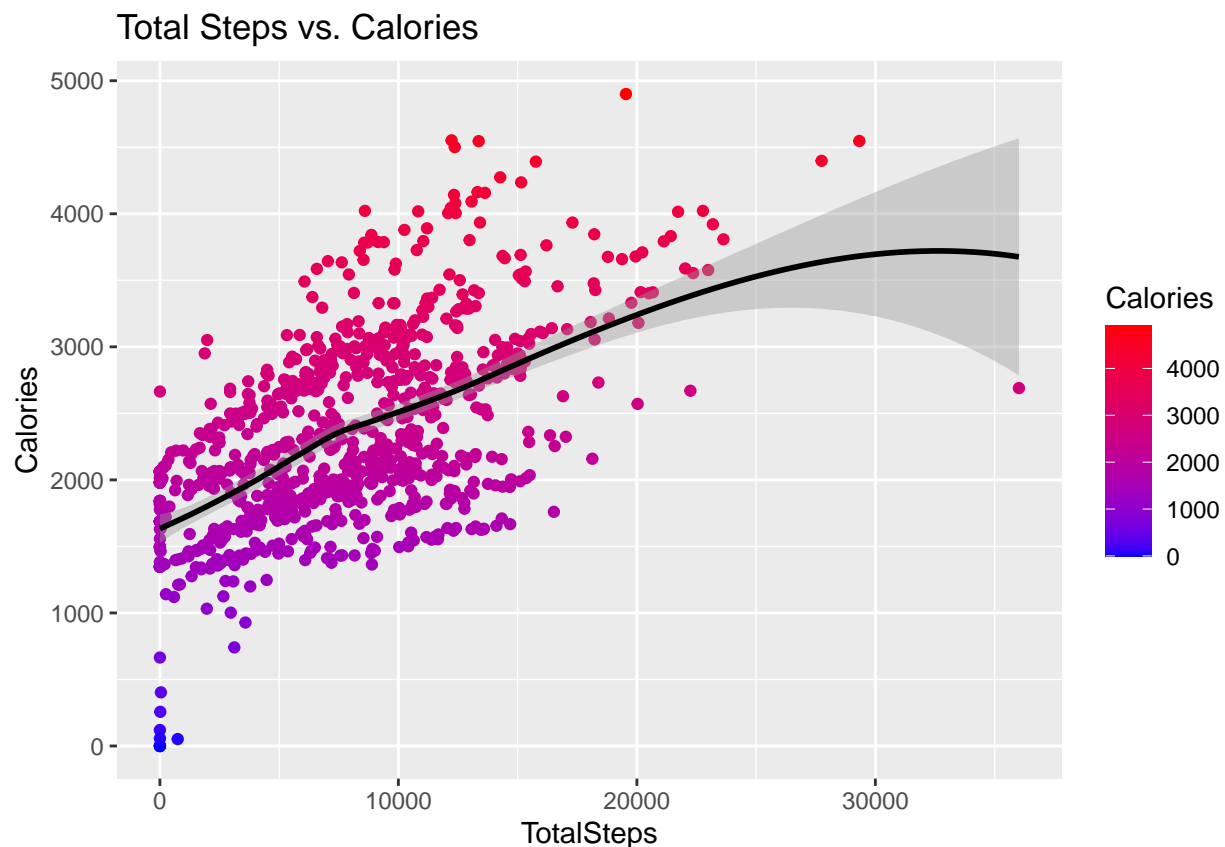
```
##      Id      date TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 2016-04-12              1                327             346
## 2 1503960366 2016-04-13              2                384             407
## 3 1503960366 2016-04-15              1                412             442
## 4 1503960366 2016-04-16              2                340             367
## 5 1503960366 2016-04-17              1                700             712
## 6 1503960366 2016-04-19              1                304             320
##      ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 2016-04-12      13162          8.50          8.50
## 2 2016-04-13      10735          6.97          6.97
## 3 2016-04-15       9762          6.28          6.28
## 4 2016-04-16      12669          8.16          8.16
## 5 2016-04-17       9705          6.48          6.48
## 6 2016-04-19      15506          9.88          9.88
##      LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1              0              1.88              0.55
## 2              0              1.57              0.69
## 3              0              2.14              1.26
## 4              0              2.71              0.41
## 5              0              3.19              0.78
## 6              0              3.53              1.32
##      LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1              6.06              0              25
## 2              4.71              0              21
## 3              2.83              0              29
## 4              5.04              0              36
## 5              2.51              0              38
## 6              5.03              0              50
##      FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1              13              328              728      1985
## 2              19              217              776      1797
## 3              34              209              726      1745
## 4              10              221              773      1863
## 5              20              164              539      1728
## 6              31              264              775      2035
```

Visualizations

Total Steps vs. Calories

```
# Visualization: Total Steps vs. Calories
ggplot(data=activity, aes(x=TotalSteps, y=Calories, color=Calories)) +
  geom_point() +
  geom_smooth(color='black') +
  labs(title="Total Steps vs. Calories") +
  scale_color_gradient(low = "blue", high = "red") # Custom gradient from blue to red
```

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



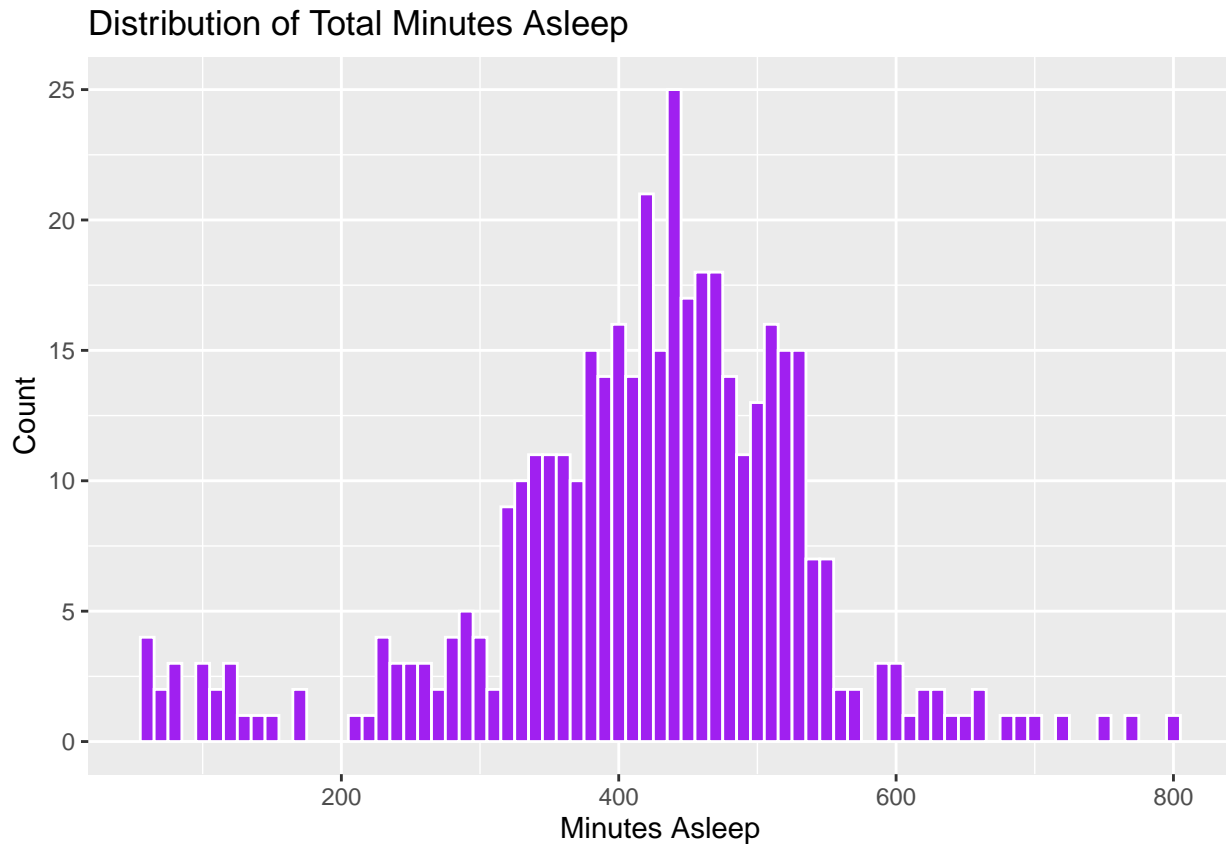
The scatter plot “Total Steps vs. Calories” reveals a **positive relationship** between the number of steps taken and calories burned, indicating that individuals who take more steps generally burn more calories.

Key observations include: 1. **Color Gradient:** Points are colored to show calorie levels, with **higher calorie values in red** and **lower values in blue**, emphasizing the trend. 2. **Trend Line:** The black fitted curve suggests that the relationship between steps and calories is **non-linear**, with calories increasing at a **decreasing rate** as steps increase. 3. **Confidence Interval:** The shaded region around the curve reflects the range of uncertainty in the trend, providing a visual indication of variability in the data.

This analysis highlights the consistent but complex relationship between physical activity (steps) and energy expenditure (calories).

Distribution of Total Minutes Asleep

```
# Histogram: Distribution of Total Minutes Asleep
ggplot(data=sleep, aes(x=TotalMinutesAsleep)) +
  geom_histogram(binwidth=10, fill="purple", color="white") +
  labs(title="Distribution of Total Minutes Asleep", x="Minutes Asleep", y="Count")
```



Key Observations:

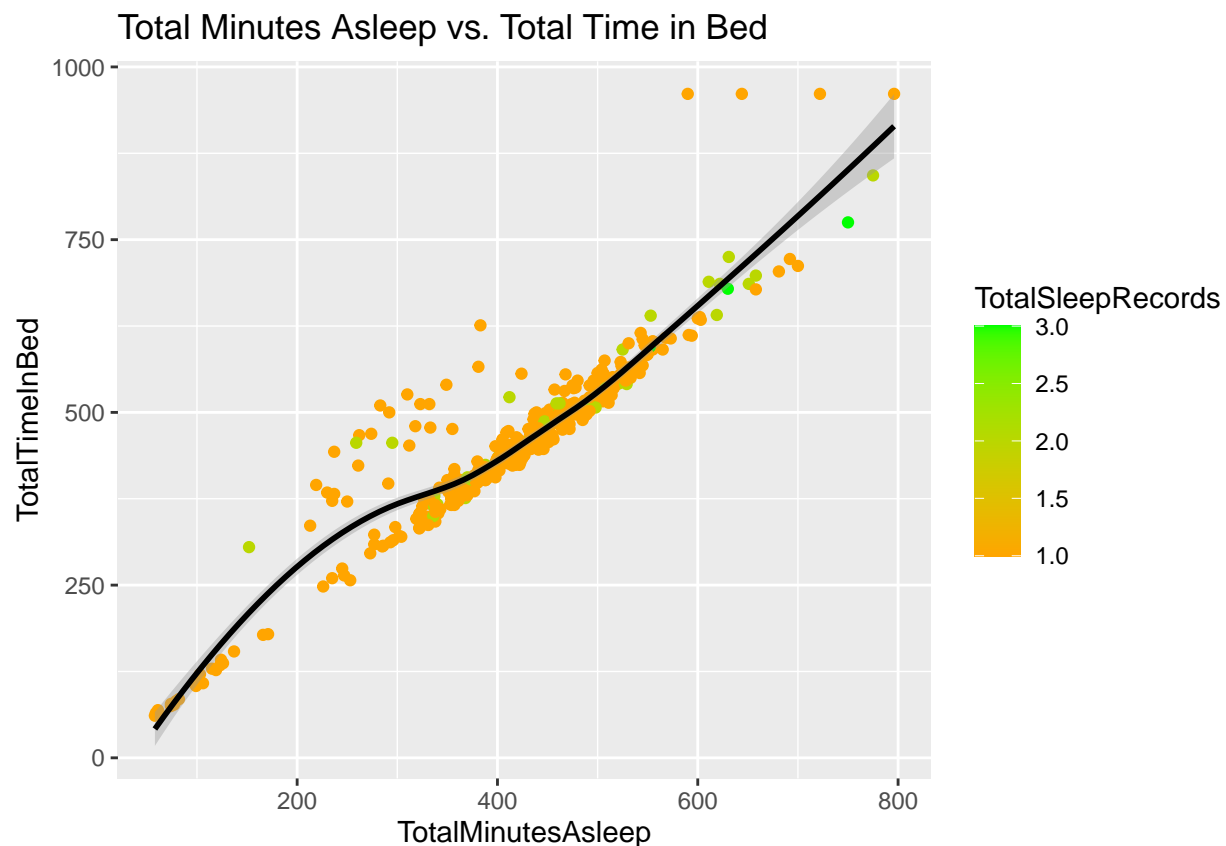
1. **Shape:** The distribution appears to be **bell-shaped or normal**, with the majority of sleep durations clustering around the center and fewer occurrences of very short or very long durations.
2. **Center:** The **most frequent sleep duration** is approximately **400 minutes** (6 hours and 40 minutes), indicating the typical sleep time for the group.
3. **Spread:** Sleep durations range from around **100 to 800 minutes**, showcasing a **wide variation** in sleep patterns across individuals.
4. **Symmetry:** The distribution is **relatively symmetrical**, suggesting that shorter and longer sleep durations are distributed fairly evenly around the central value.

These observations suggest that most individuals in the sample have a consistent sleep pattern, with a typical duration of about 6 hours and 40 minutes. However, there is variability in the data, with some individuals experiencing significantly shorter or longer sleep durations.

Total Minutes Asleep vs. Total Time in Bed with custom color gradient

```
# Scatter plot: Total Minutes Asleep vs. Total Time in Bed with custom color gradient
ggplot(data=sleep, aes(x=TotalMinutesAsleep, y=TotalTimeInBed, colour=TotalSleepRecords)) +
  geom_point() +
  geom_smooth(color='black') +
  labs(title="Total Minutes Asleep vs. Total Time in Bed") +
  scale_color_gradient(low = "orange", high = "green")
```

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



The scatter plot “Total Minutes Asleep vs. Total Time in Bed” reveals a positive relationship between the total minutes asleep and the total time spent in bed, indicating that individuals who spend more time in bed generally sleep longer.

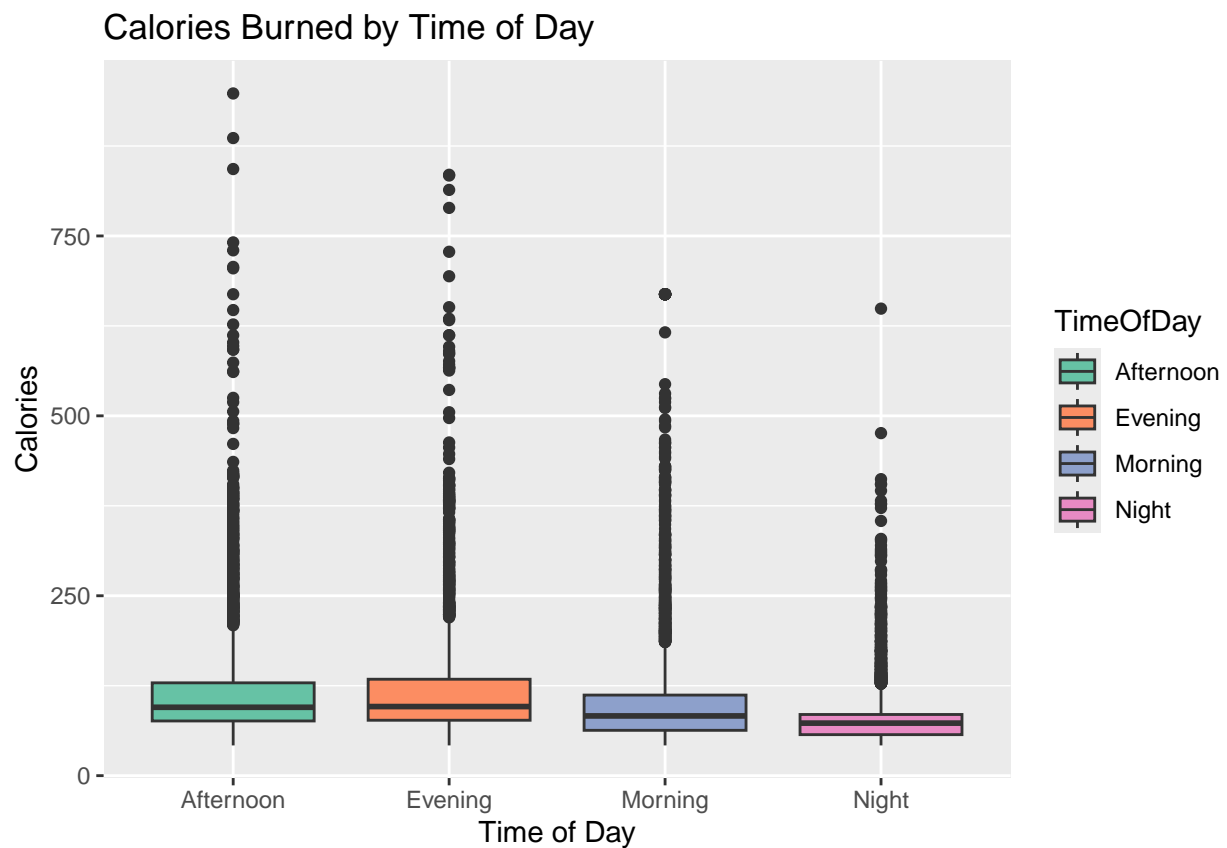
- Key Observations:
1. **Positive Correlation:** The overall trend shows a strong positive relationship between the total minutes asleep and the total time spent in bed.
 2. **Curve:** The black fitted curve suggests that the relationship is **non-linear**, with additional time in bed leading to smaller increases in sleep duration as the time in bed gets longer.
 3. **Color Coding:** Points are color-coded based on “TotalSleepRecords,” with higher values (green) appearing more frequently towards the upper end, indicating longer sleep durations and more time in bed for individuals with more sleep records.
 4. **Data Spread:** Variability around the trend line highlights individual differences in sleep patterns and other influencing factors.

Data Transformation

```
# Create a new column 'TimeOfDay' for classification of calories by time of day
calories <- calories %>%
  mutate(TimeOfDay = case_when(
    hour(ActivityHour) >= 5 & hour(ActivityHour) < 12 ~ "Morning",
    hour(ActivityHour) >= 12 & hour(ActivityHour) < 17 ~ "Afternoon",
    hour(ActivityHour) >= 17 & hour(ActivityHour) < 21 ~ "Evening",
    TRUE ~ "Night"
  ))
```

Calories Burned by Time of Day

```
# Boxplot: Calories Burned by Time of Day
ggplot(data=calories, aes(x=factor(TimeOfDay), y=Calories, fill=TimeOfDay)) +
  geom_boxplot() +
  labs(title="Calories Burned by Time of Day", x="Time of Day", y="Calories") +
  scale_fill_brewer(palette="Set2")
```



The box plot “Calories Burned by Time of Day” illustrates the distribution of calories burned during different periods: Afternoon, Evening, Morning, and Night.

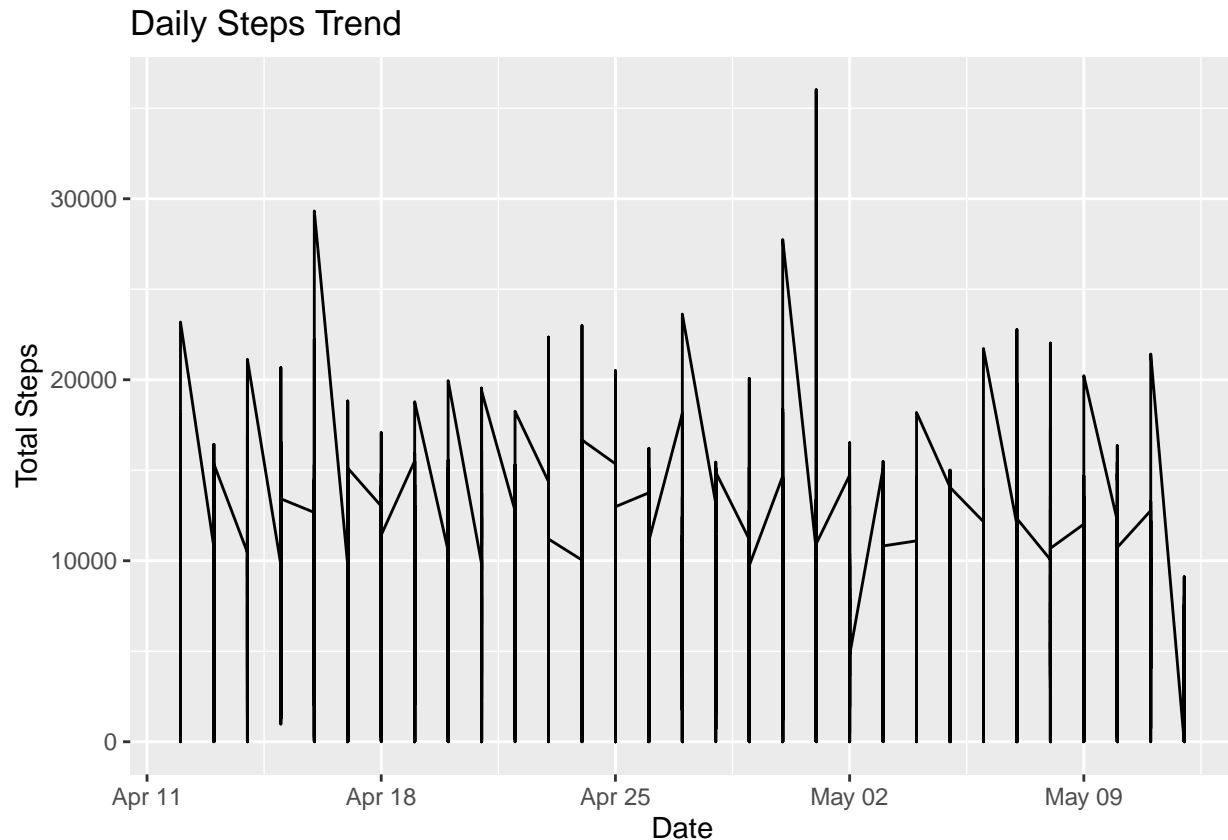
Key Observations: 1. **Median:** The Evening has the highest median calories burned, while the Afternoon has the lowest.

2. **Spread:** The Evening shows the widest spread (IQR) of calories burned, indicating greater variability, whereas the Afternoon has a narrower range.

3. **Outliers:** Outliers are present, particularly in the Evening and Night categories, representing unusually high or low calorie counts.

Daily Steps Trend over time

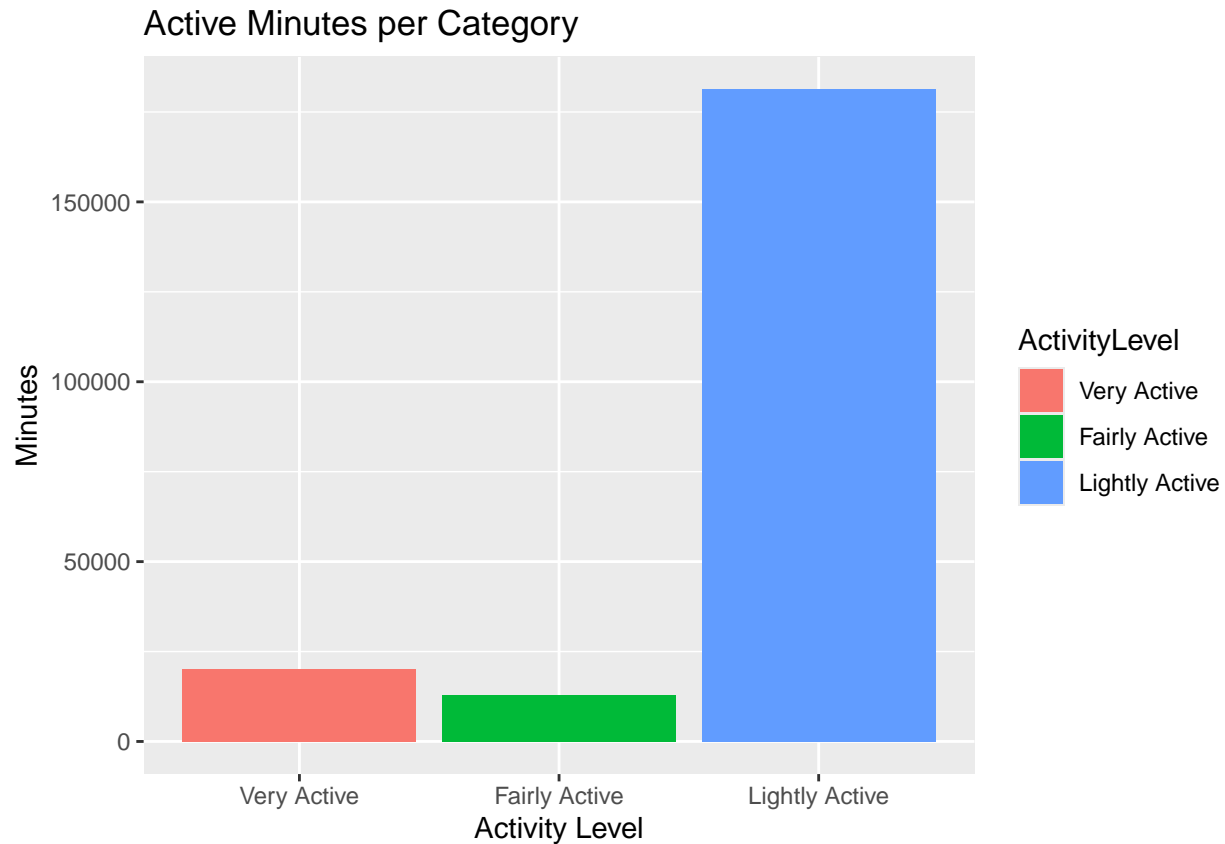
```
# Line plot: Daily Steps Trend over time
ggplot(data=activity, aes(x=ActivityDate, y=TotalSteps)) +
  geom_line() +
  labs(title="Daily Steps Trend", x="Date", y="Total Steps")
```



Activity Levels Visualization

```
# Reshaping the data for activity levels
activity %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%
  gather(key="ActivityLevel", value="Minutes") %>%
```

```
mutate(ActivityLevel = factor(ActivityLevel,
                              levels = c("VeryActiveMinutes", "FairlyActiveMinutes", "LightlyActiveMinutes"),
                              labels = c("Very Active", "Fairly Active", "Lightly Active")))) %>%
ggplot(aes(x=ActivityLevel, y=Minutes, fill=ActivityLevel)) +
geom_bar(stat="identity") +
labs(title="Active Minutes per Category", x="Activity Level", y="Minutes")
```

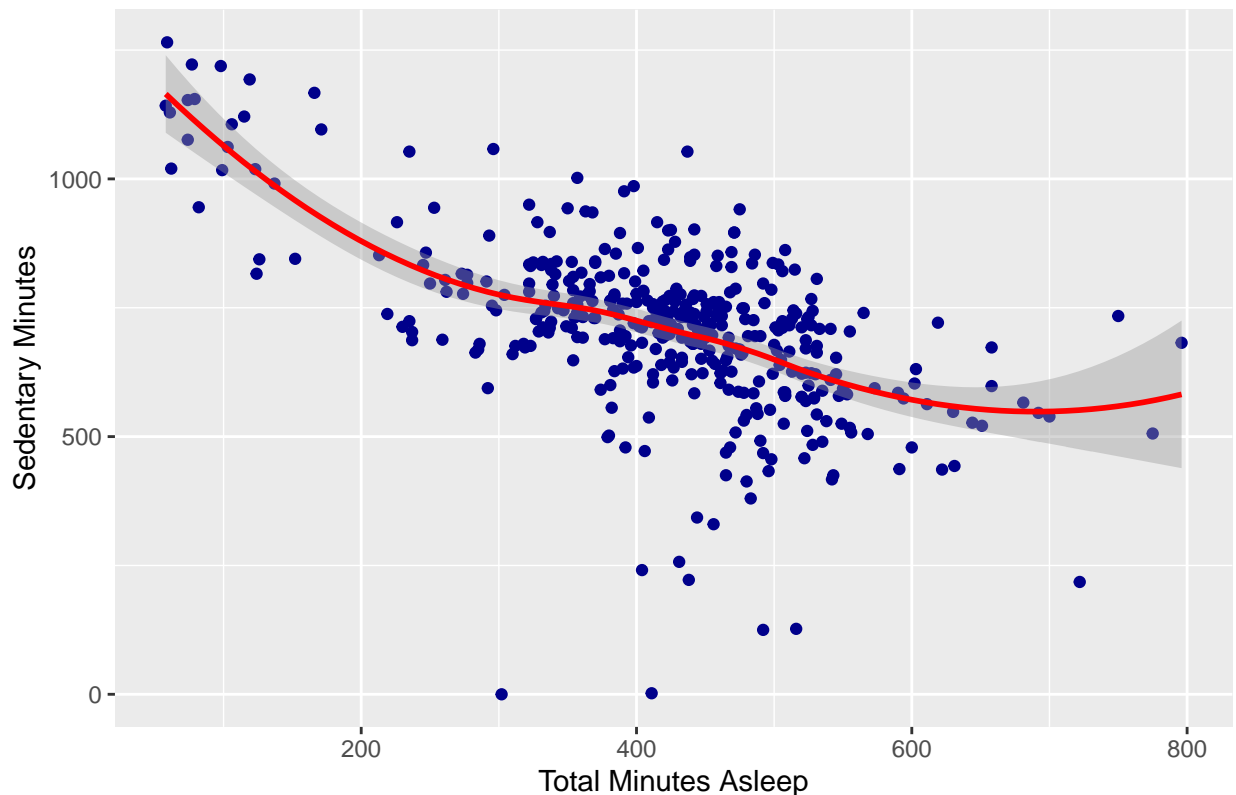


Sleep vs Sedentary Minutes

```
# Scatter plot: Minutes Asleep vs. Sedentary Minutes
ggplot(data=merged_data, aes(x=TotalMinutesAsleep, y=SedentaryMinutes)) +
  geom_point(color='darkblue') + geom_smooth(color='red') +
  labs(title="Minutes Asleep vs. Sedentary Minutes", x="Total Minutes Asleep", y="Sedentary Minutes")
```

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

Minutes Asleep vs. Sedentary Minutes



Key Observations: 1. **Negative Correlation:** The overall trend suggests a negative correlation between total minutes asleep and sedentary minutes. As the total minutes asleep increase, the sedentary minutes tend to decrease.

2. **Curve:** The red curve indicates that the relationship might not be perfectly linear. It suggests a curved pattern, where the decrease in sedentary minutes slows down as sleep duration increases.
3. **Data Spread:** There is a considerable amount of scatter around the curve. This indicates that there is a fair amount of variability in sedentary minutes even for individuals with similar sleep durations.

Recommendations for Bellabeat's Marketing Strategy:

-Focus on Key Features: Center marketing efforts around the features that users value most, such as step tracking, calorie counting, sleep tracking, and activity monitoring.

-Promote Health and Wellness Benefits: Position Bellabeat products as essential tools for improving health and well-being, highlighting benefits like increased physical activity, better sleep, and reduced sedentary behavior.

-Incorporate Gamification for Motivation: Integrate gamification elements in the Bellabeat app, such as challenges, rewards, and leaderboards, to motivate users and foster consistent engagement.

-Create a Personalized Experience: Utilize user data to offer tailored recommendations and feedback, including personalized goals, activity suggestions, and insights into sleep patterns.

By implementing these recommendations, Bellabeat can effectively leverage the insights from Fitbit data to enhance its marketing strategies and improve the overall health and well-being of its users.