

# Bellabeat Case Study

Santhosh Pagavathsing

11/04/2022

## Business Task

The given task is to analyze smart device usage data and gain insight into how consumers use non-Bellabeat smart devices, in order to help guide future Bellabeat marketing strategies.

## Data Importing and Summarizing

### Data Source Used

The open source FitBit Fitness Tracker Data is collected from Kaggle, posted by user Möbius. It contains data generated by thirty consenting Fitbit users who submitted their personal tracking data. This includes hourly and daily level output for physical activity, steps, heart rate, and sleep monitoring, which gives insight into user habits. This data appears to be credible since it has a usability score of 10, 1032 upvotes, and 30000 downloads.

There are some limitations to this data set such as: not providing unit of measurement for columns (ex. TotalDistance), missing or excess data added (active minutes and sleep data not resulting in a full 24 hours), not enough information of the population that the data was collected from (gender, age, location, etc). These limitations were considered when completing the analysis. The objective is to get an understanding of how people generally use their smart devices and what kind of activity they do.

### Loading Packages

```
library(tidyverse)
library(dplyr)
library(lubridate)
```

### Importing Datasets

```
dailyActivity <- read_csv("/Users/santhosh/Documents/R Studio Projects/Bellabeat/dailyActivity_merged.csv")
dailySleep <- read_csv("/Users/santhosh/Documents/R Studio Projects/Bellabeat/sleepDay_merged.csv")
hourlyIntensities <- read_csv("/Users/santhosh/Documents/R Studio Projects/Bellabeat/hourlyIntensities_merged.csv")
```

### Checking each dataset

```
head(dailyActivity)
```

```
## # A tibble: 6 x 15
##       Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitie~
##       <dbl> <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1  1.50e9 4/12/2016          13162           8.5           8.5           0
## 2  1.50e9 4/13/2016          10735           6.97          6.97          0
## 3  1.50e9 4/14/2016          10460           6.74          6.74          0
## 4  1.50e9 4/15/2016           9762           6.28          6.28          0
## 5  1.50e9 4/16/2016          12669           8.16          8.16          0
## 6  1.50e9 4/17/2016           9705           6.48          6.48          0
## # ... with 9 more variables: VeryActiveDistance <dbl>,
## #   ModeratelyActiveDistance <dbl>, LightActiveDistance <dbl>,
## #   SedentaryActiveDistance <dbl>, VeryActiveMinutes <dbl>,
## #   FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## #   SedentaryMinutes <dbl>, Calories <dbl>
```

```
head(dailySleep)
```

```
## # A tibble: 6 x 5
##       Id SleepDay          TotalSleepReco~ TotalMinutesAsl~ TotalTimeInBed
##       <dbl> <chr>          <dbl>          <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:00:0~          1           327           346
## 2 1503960366 4/13/2016 12:00:0~          2           384           407
## 3 1503960366 4/15/2016 12:00:0~          1           412           442
## 4 1503960366 4/16/2016 12:00:0~          2           340           367
## 5 1503960366 4/17/2016 12:00:0~          1           700           712
## 6 1503960366 4/19/2016 12:00:0~          1           304           320
```

```
head(hourlyIntensities)
```

```
## # 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
```

## Summarizing the data

```
dailyActivity %>%
```

```
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes, TotalSteps, TotalCalories) %>%
  summary()
```

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

Looking at the summary statistics we see that on average, most people spend their time being sedentary (roughly 16 hours) When it comes to activity, most people spend their time being lightly active (roughly 3 hours) rather than being very active or fairly active. We can also see the average number of steps taken per day is 7638 while Calories burned is 2304.

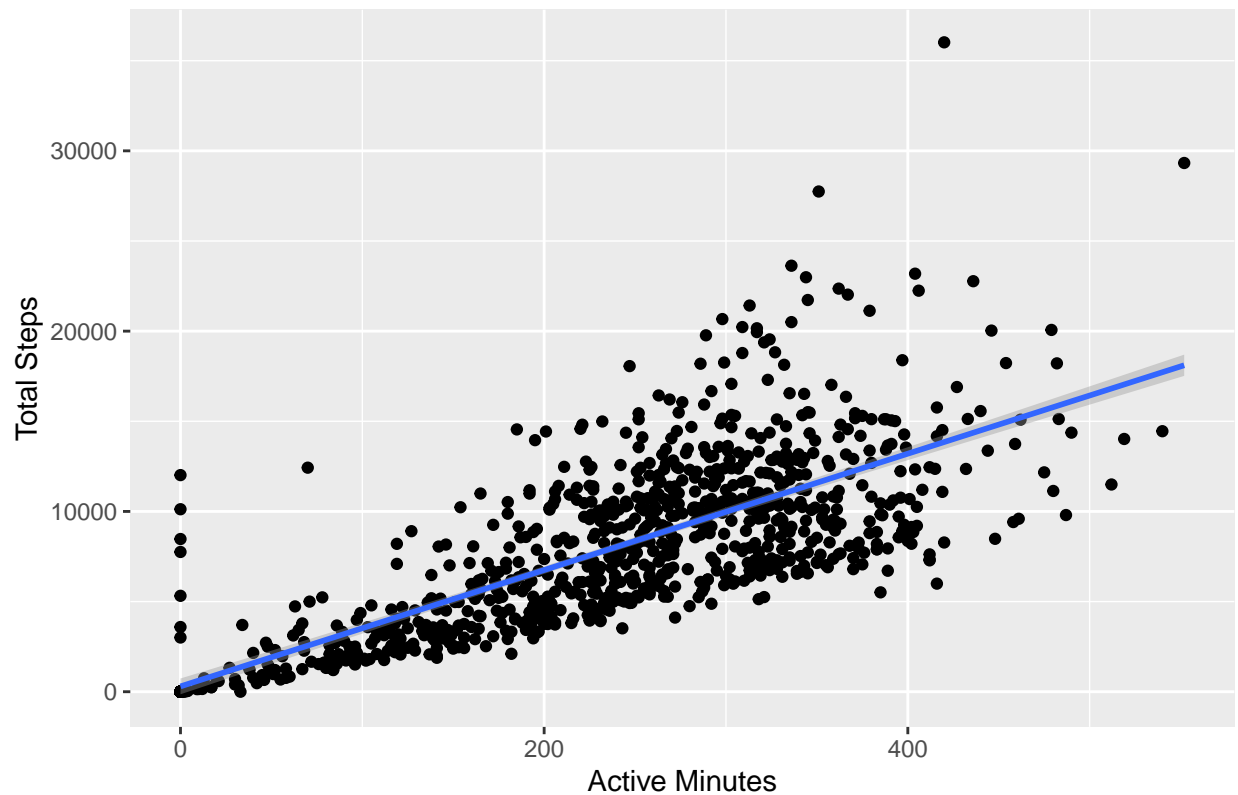
## Analysis

### Total Steps VS. Active Minutes

```
new_dailyActivity <- dailyActivity %>%
  mutate(ActiveMinutes = VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes) %>%
  mutate(AllMinutes = ActiveMinutes + SedentaryMinutes)

new_dailyActivity %>%
  ggplot(aes(x = ActiveMinutes, y = TotalSteps)) +
  geom_point() + geom_smooth(method = "lm") +
  labs(title="Total Steps VS. Active Minutes", y= "Total Steps", x = "Active Minutes")
```

## Total Steps VS. Active Minutes

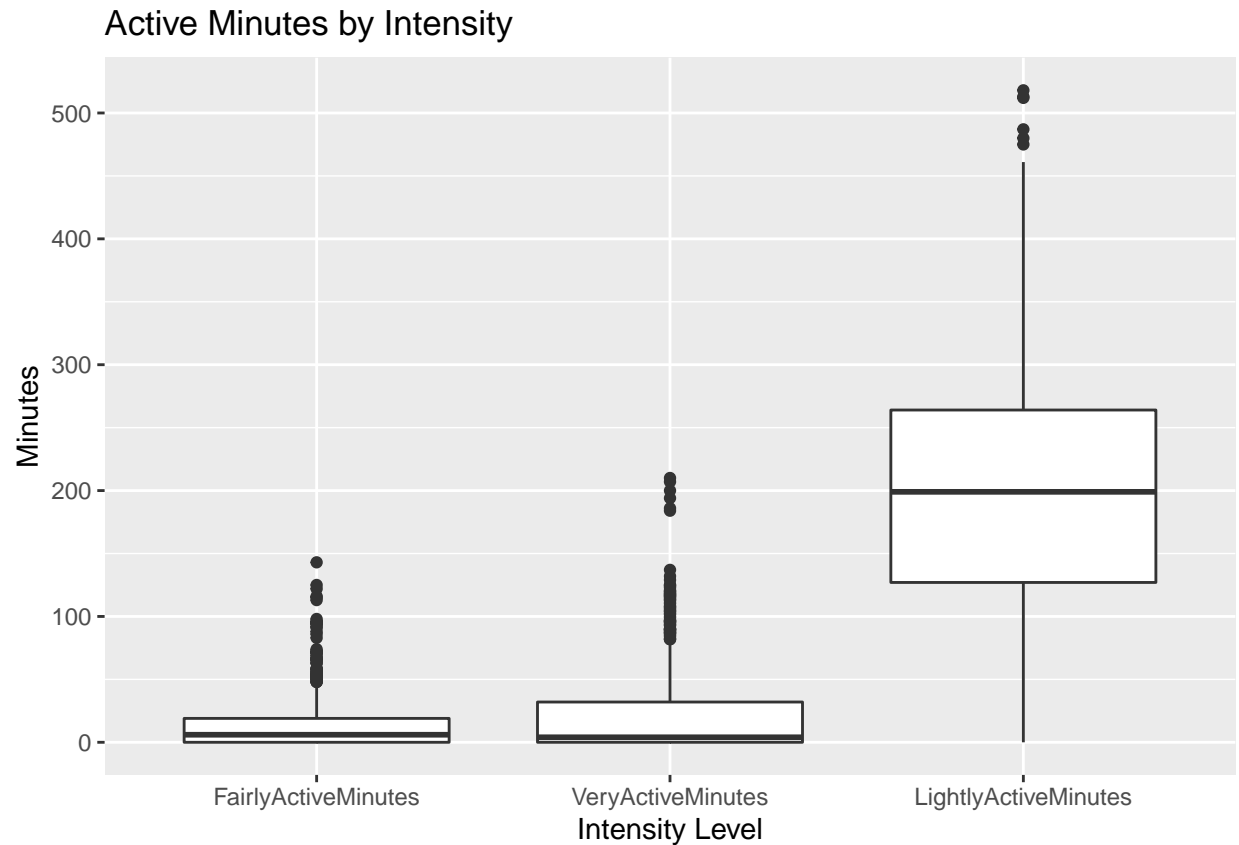


This graph compares the total minutes a person spent in a day (Very active minutes, fairly active minutes, and lightly active minutes) in relation to the number of steps they took in that day. Here we can see that there is a correlation between the active minutes and total steps. This can mean that most people favor spending their active time running, jogging, or walking.

## Active Minutes by Intensity

```
Narrow_Activity <- dailyActivity %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%
  pivot_longer(VeryActiveMinutes:LightlyActiveMinutes, names_to = "IntensityType", values_to = "Minutes")
  arrange(IntensityType)

ggplot(Narrow_Activity, aes(x = reorder(IntensityType, Minutes, no.rm = TRUE), y = Minutes)) + geom_boxplot()
labs(title="Active Minutes by Intensity", y= "Minutes", x = "Intensity Level")
```



The graph visualizes the spread of each activity level by category from fairly active, to lightly active. As noted while getting the summary statistics, we can see that on average most minutes are spent doing light activities.

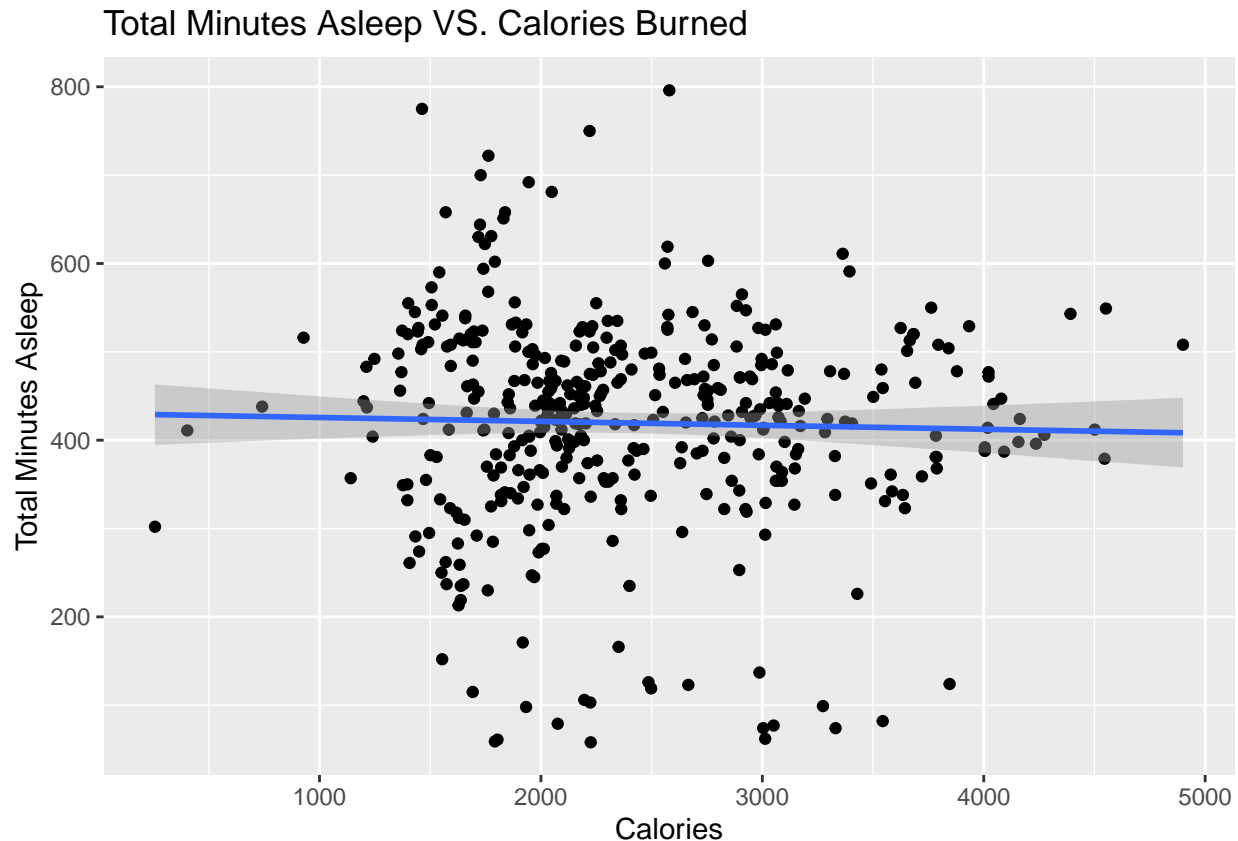
### Total Minutes Asleep VS Calories Burned

```
dateSleep <- dailySleep %>%
  rename(dateRecord = SleepDay) %>%
  mutate(dateRecord = as.Date(mdy_hms(dateRecord)))

dateActivity <- new_dailyActivity %>%
  rename(dateRecord = ActivityDate) %>%
  mutate(dateRecord = as.Date(mdy(dateRecord)))

SleepActivity_merged <- merge(dateSleep, dateActivity, by=c("Id", "dateRecord"))

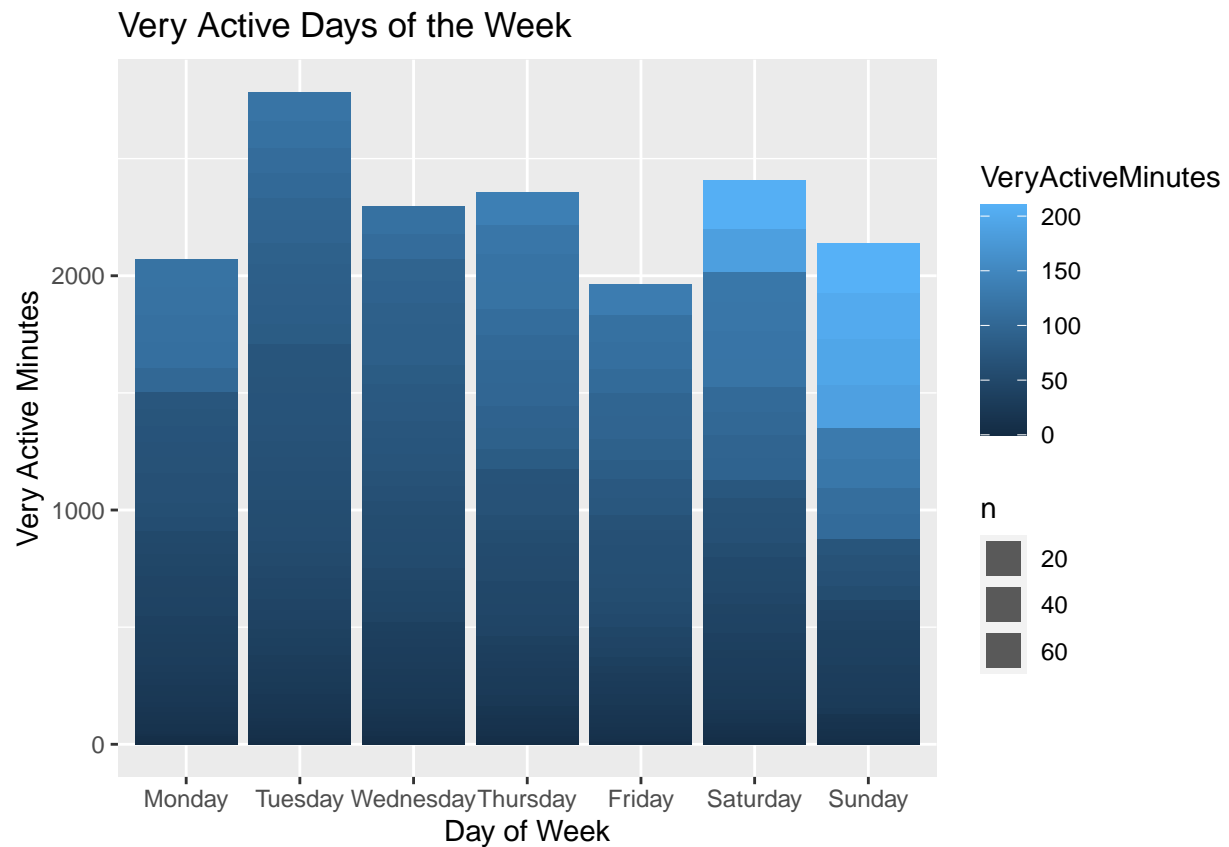
SleepActivity_merged %>%
  ggplot(aes(x = Calories, y = TotalMinutesAsleep)) +
  geom_point() + geom_smooth(method = "lm") +
  labs(title="Total Minutes Asleep VS. Calories Burned", y = "Total Minutes Asleep", x = "Calories")
```



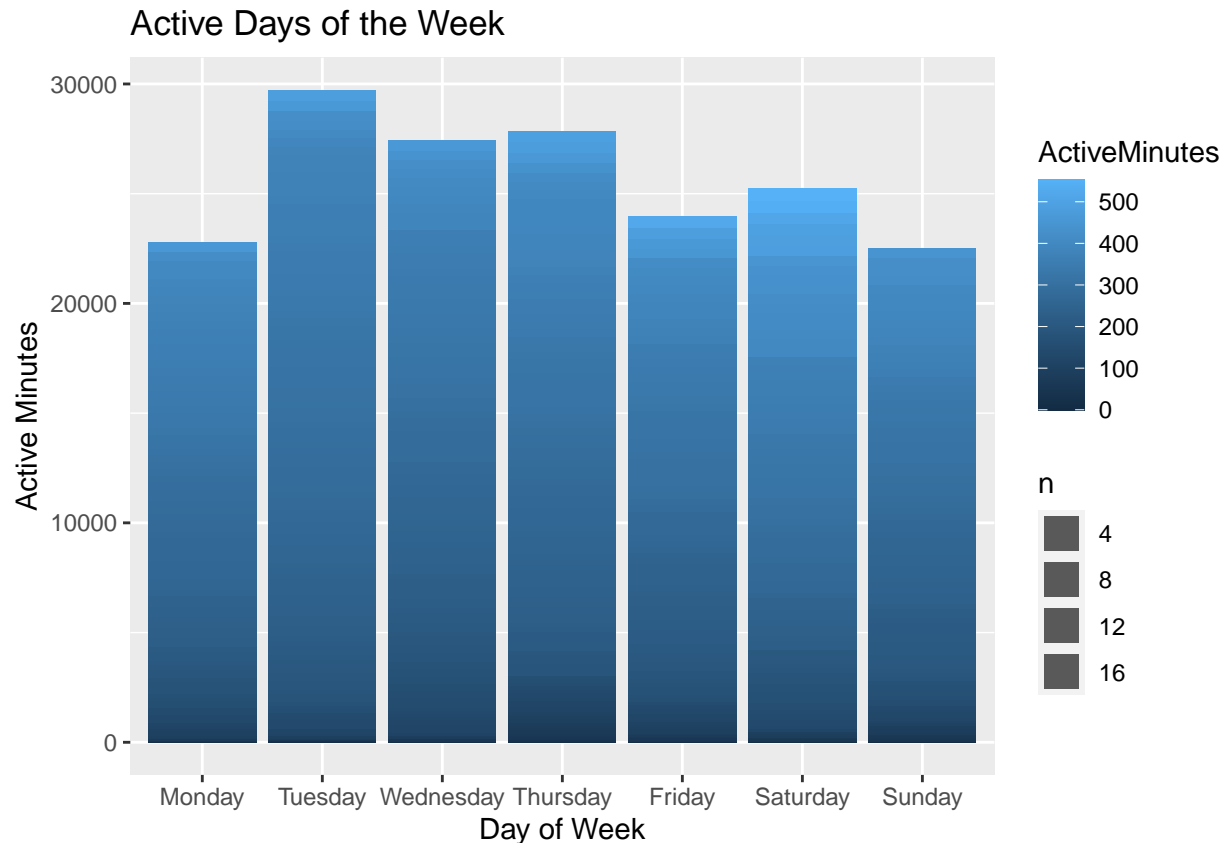
This graph compares the amount of sleep a person got for a day with the number of calories they burned. To get this, first we needed to merge the sleep data with the activity data. This was done by performing an inner join between the two data sets by each person's Id and the Date that the data was recorded. To my surprise there was no correlation between sleep and calories burned. This was also checked with activity and step data as well, there was no correlation there either. This is surprising since it would make sense that if a person were to be more active that would make them more tired and make them fall asleep for longer periods of time. However, this shows that each person is different with their own sleeping patterns and schedules, and that doesn't necessarily rely on activity levels.

### Most Active Days of the Week

```
dailyActivity %>% select(ActivityDate, VeryActiveMinutes) %>%
  mutate(ActivityDate = weekdays(as.Date(mdy(ActivityDate)))) %>%
  mutate(ActivityDate = factor(ActivityDate, levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))) %>%
  ggplot(aes(ActivityDate, VeryActiveMinutes, fill=VeryActiveMinutes)) +
  geom_bar(stat="sum") +
  labs(title="Very Active Days of the Week", y = "Very Active Minutes", x = "Day of Week")
```



```
new_dailyActivity %>% select(ActivityDate, ActiveMinutes) %>%
  mutate(ActivityDate = weekdays(as.Date(mdy(ActivityDate)))) %>%
  mutate(ActivityDate = factor(ActivityDate, levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))) %>%
  ggplot(aes(ActivityDate, ActiveMinutes, fill=ActiveMinutes)) +
  geom_bar(stat="sum") +
  labs(title="Active Days of the Week", y = "Active Minutes", x = "Day of Week")
```

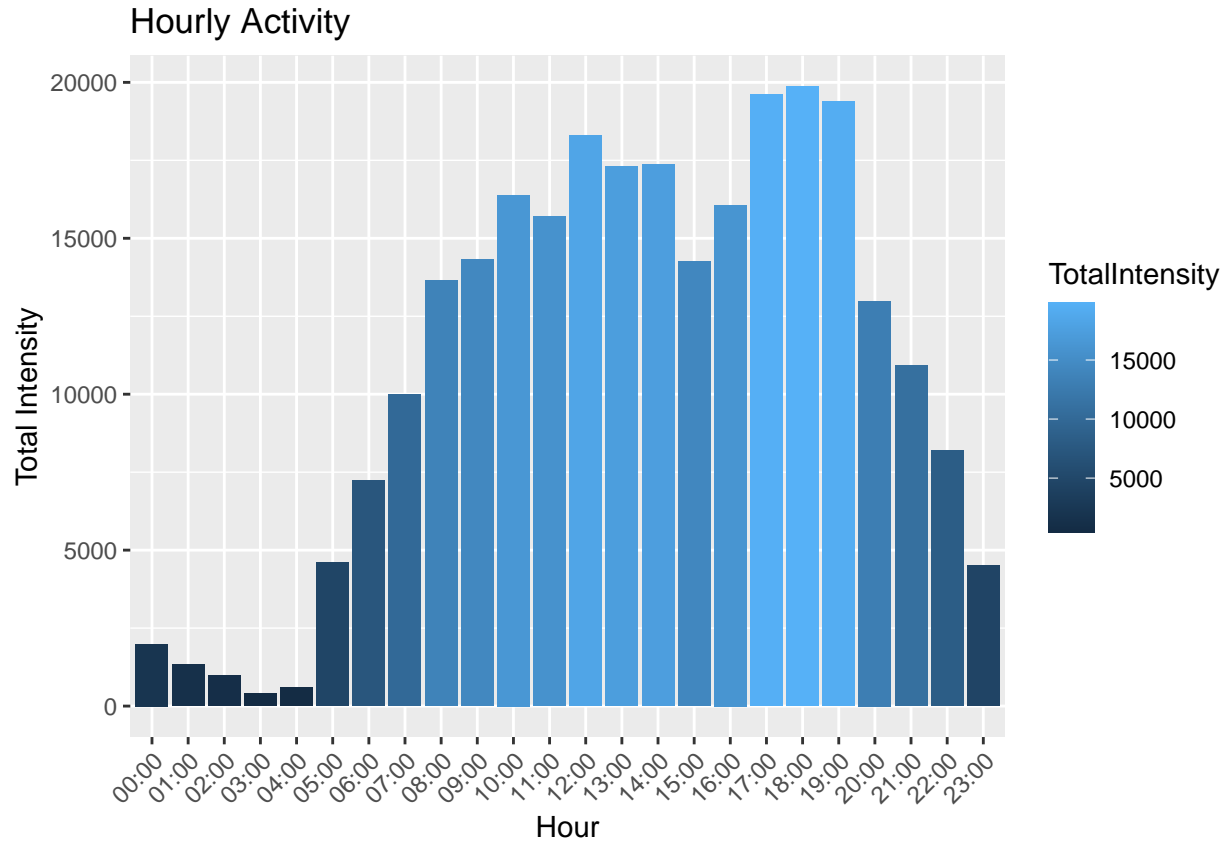


These graphs show which days of the week people are the most active. The first graph shows very active minutes while the second focuses on total active minutes. For the first graph we see that on Tuesday, people spend more time being very active compared to other days of the week, and Friday they spend the least. This also shows that, by the lighter blue chunks, people spend longer periods of time being very active on Sundays more than other days. Looking at the second graph we see that Tuesday is the most active day while Monday and Sunday are the least.

## Hourly Activity

```
hourlyIntensities %>%
  mutate(ActivityHour = format(as.POSIXct(mdy_hms(ActivityHour)), format = "%H:%M")) %>%
  select(ActivityHour, TotalIntensity) %>%
  group_by(ActivityHour) %>%
  summarise(TotalIntensity = sum(TotalIntensity)) %>%
  ggplot(aes(ActivityHour, TotalIntensity, fill=TotalIntensity)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 45, hjust=1)) +
  labs(title="Hourly Activity", y = "Total Intensity", x = "Hour")
```





This graph shows the average activity in a day per hour. It is clear that 5-7pm is when the most activity happens since these are the times with highest intensities. We can also see that 12-4am is the time where the least activity happens with the lowest average intensities.

## Recomendations

### Personalized Messages

A good way to get people more active and interacting with their Bellabeat device is to send them personalized messages. This can be notifications that

- encourage the user during or after high activity intensities
- compare their current performance to a previous time; whether they improved or they should keep going to beat their last record to increase consistency and intensity
- compare their performance to the average of users in a similar demographic such as age range or location.
- motivate users to get active when they've spent too much time being sedentary

### Advertisements

The data shows that the most active time of the day is 5-7pm. This is typically when the work day ends so it can be said that most people are most active after work. The target audience for advertisements should include working women who, like the data suggests, spend most of their day being sedentary. The

advertisements should motivate customers to start being active during their free time, and to purchase a Bellabeat device to easily help them with their active lifestyle.

The more data collected the better it is to understand customers and give them guidance. Showcasing the convenience, ease, and style of wearing a Bellabeat product makes customers believe that they can easily wear the device and always have their access to their personalized statistics and messages.

The data also suggests that on average, users typically spend their active time running, jogging, or walking. The advertisements should also show how a Bellabeat product would be a great hassle free addition to a person's daily running gear and how it can push them to go further.