Capstone - Bellabeat

Keerthana

2025-08-15

```r
install.packages("tidyverse")
install.packages("janitor")
install.packages("lubridate")
```

```r
library(tidyverse)
library(janitor)
library(dplyr)
library(lubridate)
library(ggplot2)
```

Preparing data

```r
## Checking if the files are loaded in path
files <- list.files(pattern = "*.csv", full.names = TRUE)
files
```

```
## [1] "./dailyActivity_merged.csv"     "./heartrate_seconds_merged.csv"
## [3] "./hourlyIntensities_merged.csv" "./sleepDay_merged.csv"
## [5] "./weightLogInfo_merged.csv"
```

```r
## Reading csv files
dailyActivity_merged <- read.csv("dailyActivity_merged.csv")
sleepDay_merged <- read.csv("sleepDay_merged.csv")
weightLogInfo_merged <- read.csv("weightLogInfo_merged.csv")
hourlyIntensities_merged <- read.csv("hourlyIntensities_merged.csv")
```

Processing data

```r
## Cleaning up dates
dailyActivity_merged <- dailyActivity_merged %>%
  mutate(ActivityDate = mdy(ActivityDate))

sleepDay_merged <- sleepDay_merged %>%
  mutate(SleepDay = mdy_hms(SleepDay)) %>%   # includes time part
  mutate(SleepDay = as.Date(SleepDay))    # keeping only the date part
```

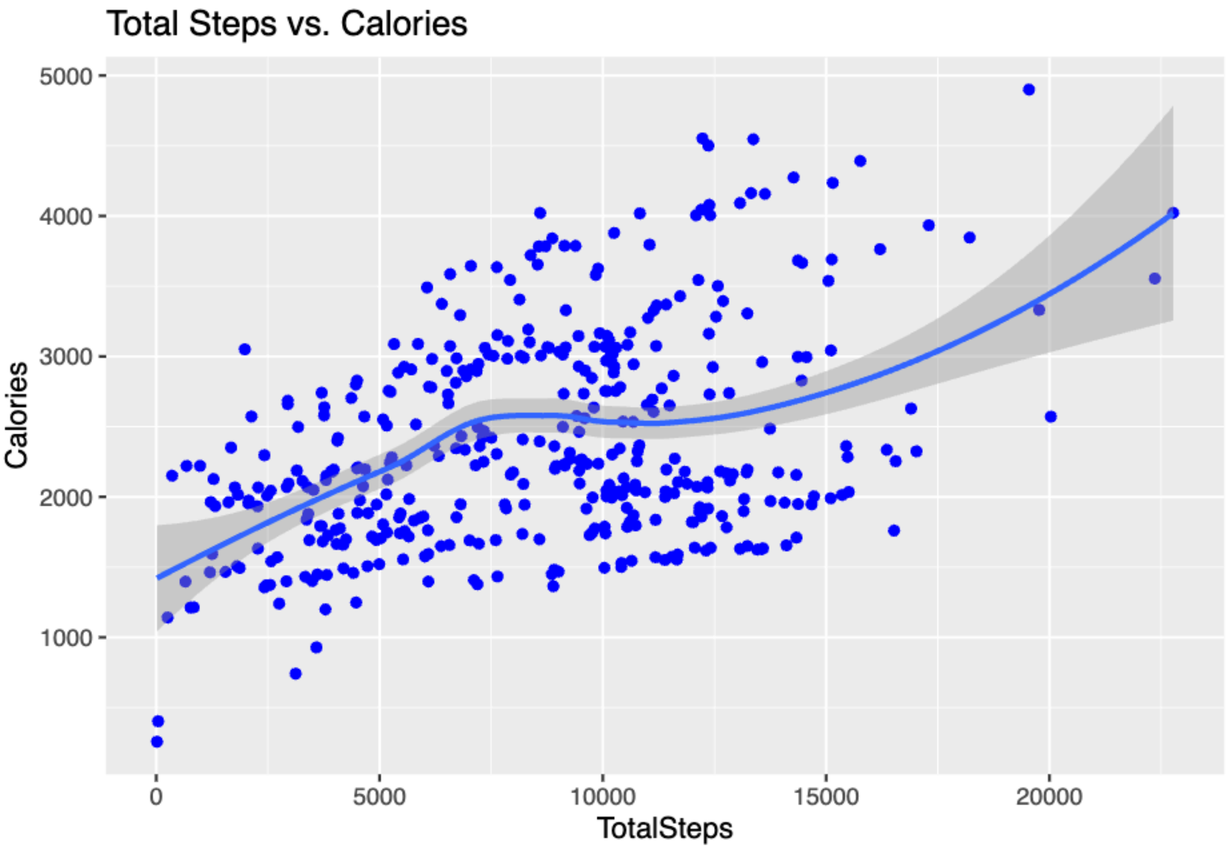
```r
## Joining Activity Data and Sleep Data
activity_sleep <- dailyActivity_merged %>%
  inner_join(sleepDay_merged,
             by = c("Id" = "Id", "ActivityDate" = "SleepDay"))

head(activity_sleep)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-15       9762          6.28            6.28
## 4 1503960366   2016-04-16      12669          8.16            8.16
## 5 1503960366   2016-04-17       9705          6.48            6.48
## 6 1503960366   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
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1                 1                327            346
## 2                 2                384            407
## 3                 1                412            442
## 4                 2                340            367
## 5                 1                700            712
## 6                 1                304            320
```

Analysing data

```r
# Steps vs Calories burnt
ggplot(activity_sleep, aes(x=TotalSteps, y=Calories)) +
  geom_point(color="blue") +
  geom_smooth() +
  labs(title="Total Steps vs. Calories")
```

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

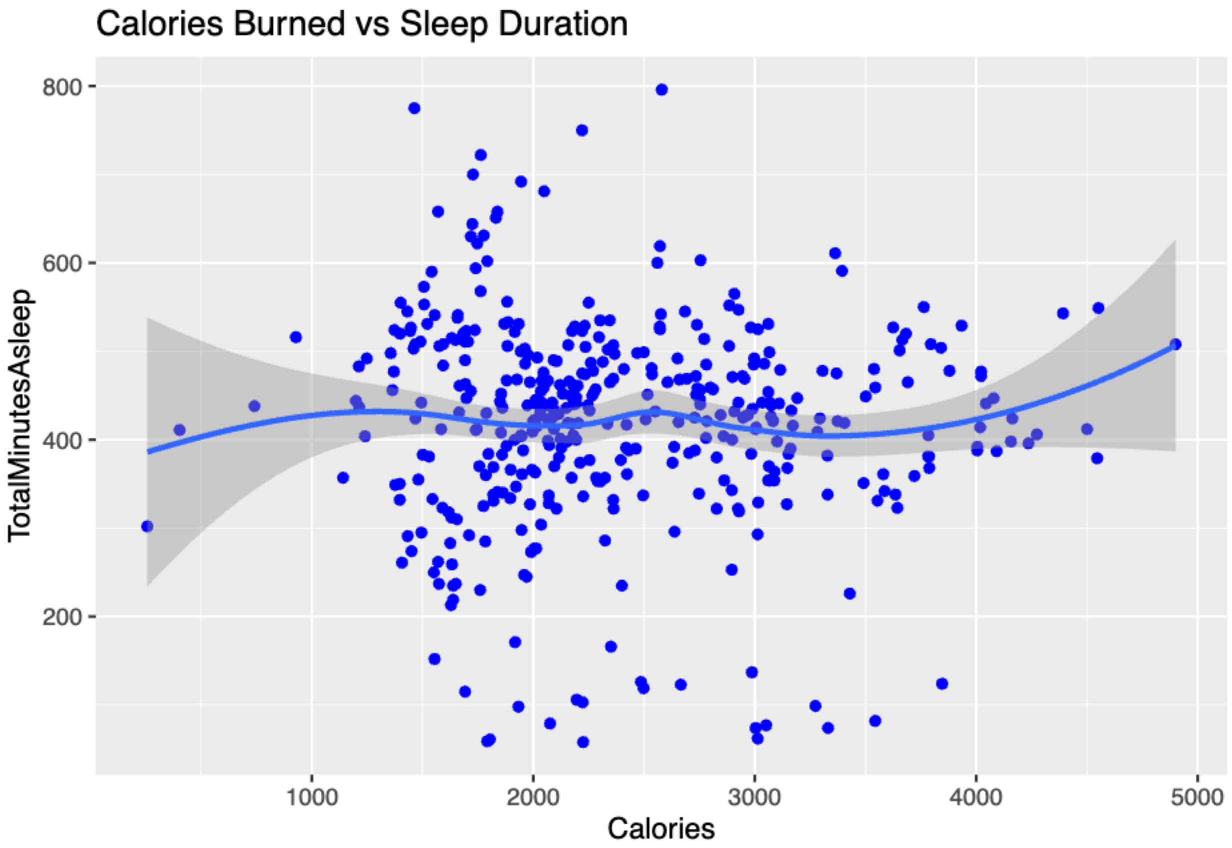


Observations

- The graph shows that the higher the number of steps taken by the user, the higher the calories burnt.
- There is a positive correlation between the two.

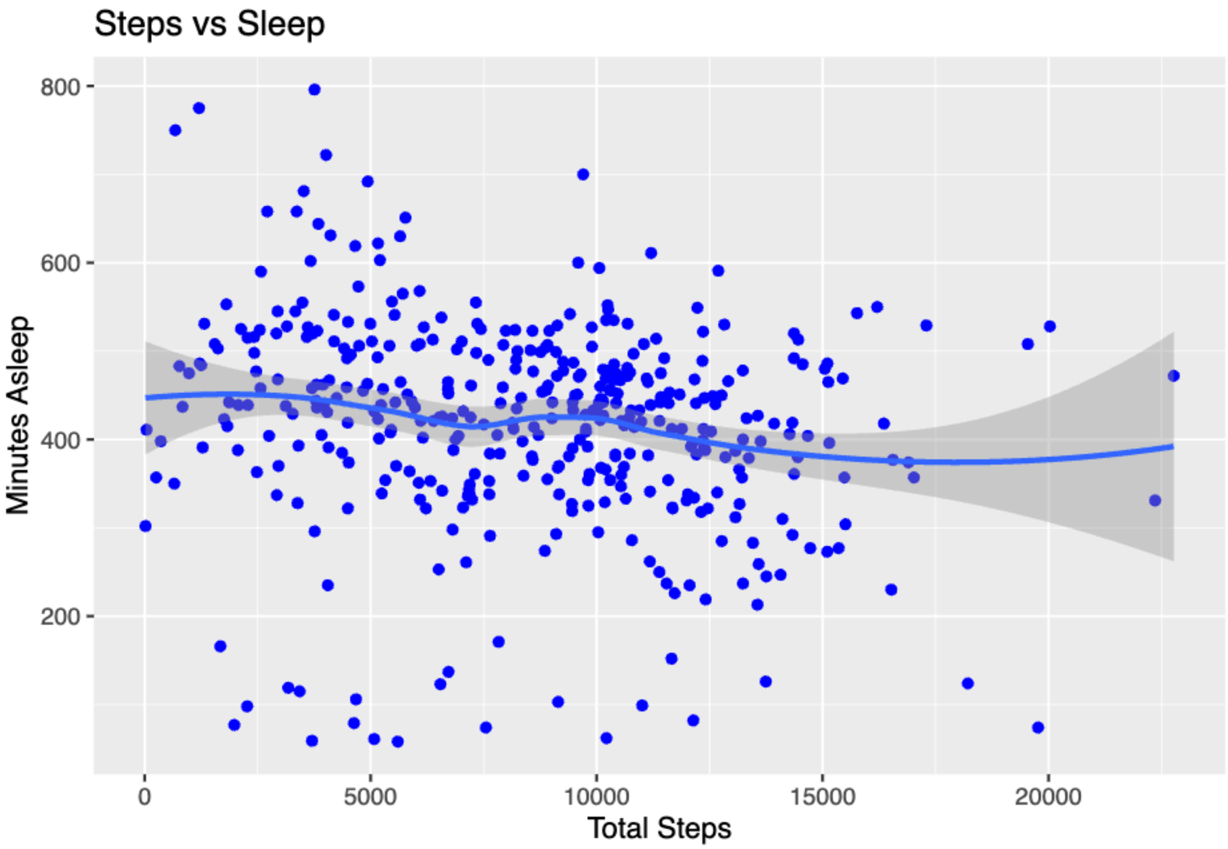
```r
# Calories Burned vs Sleep Duration
ggplot(activity_sleep, aes(x = Calories, y = TotalMinutesAsleep)) +
  geom_point(color = "blue") +
  geom_smooth() +
  labs(title = "Calories Burned vs Sleep Duration")
```

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



```r
# Steps vs Sleep
ggplot(activity_sleep, aes(x = TotalSteps, y = TotalMinutesAsleep)) +
  geom_point(color = "blue") +
  geom_smooth() +
  labs(title = "Steps vs Sleep", x = "Total Steps", y = "Minutes Asleep")
```

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

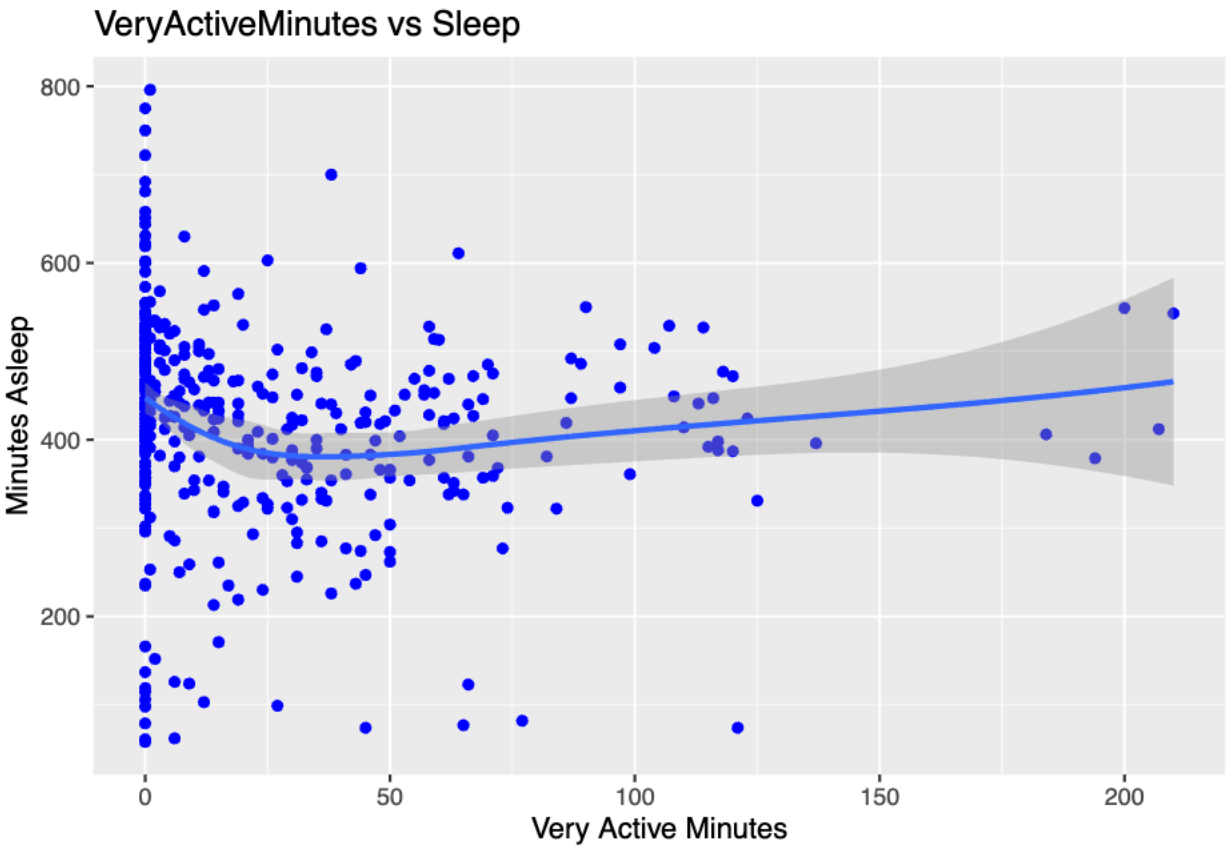


Observations

- The graphs above show that there is no strong correlation between
  - Calories burnt and time slept
  - Steps walked and time slept
- The blue regression line is almost flat which means
  - Burning more calories will not increase sleep time
  - Walking more steps will not increase sleep time

```
# VeryActive vs sleep time
ggplot(activity_sleep, aes(x = VeryActiveMinutes, y = TotalMinutesAsleep)) +
  geom_point(color = "blue") +
  geom_smooth() +
  labs(title = "VeryActiveMinutes vs Sleep", x = "Very Active Minutes", y = "Minutes Asleep")
```

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

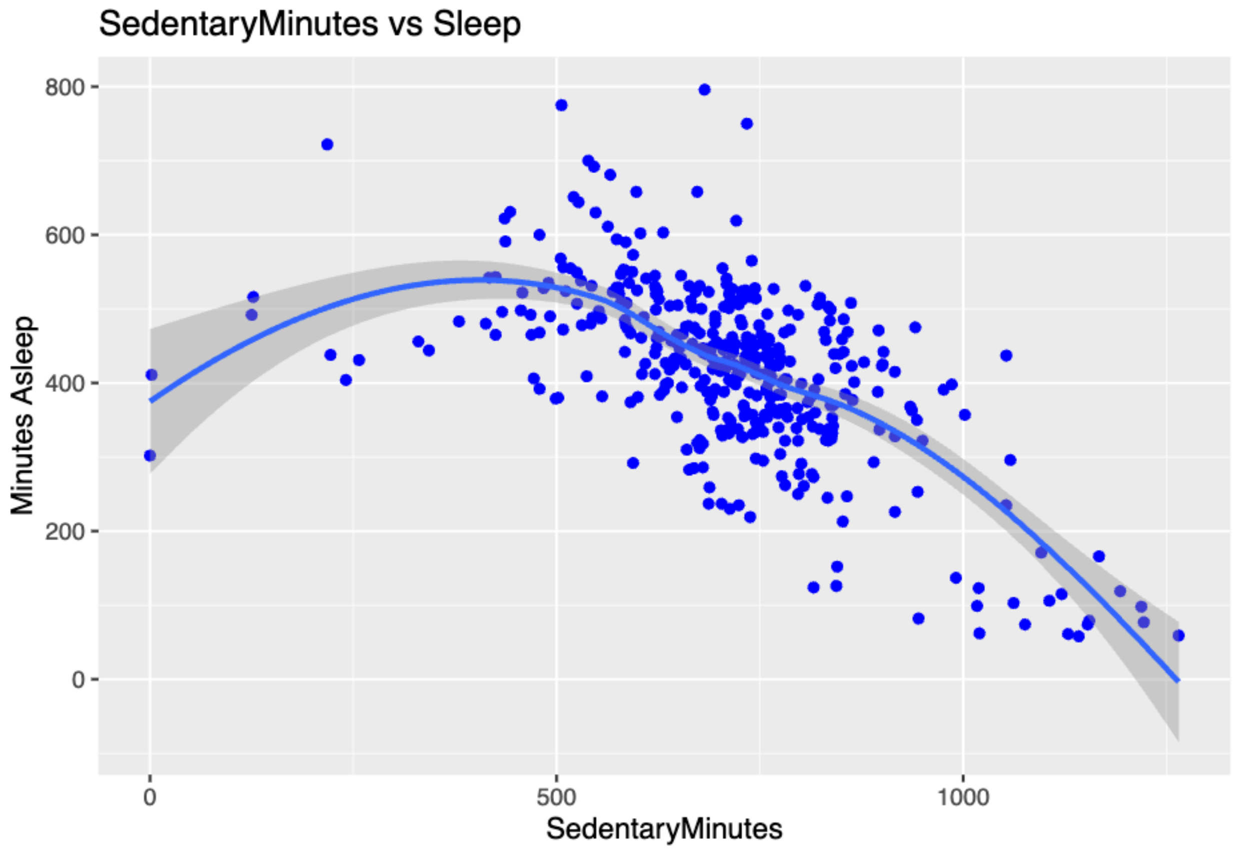


Observations

- Being active might be loosely linked to a little more sleep, but the relationship is weak.
- It is not a guarantee that if a person is more active, they will have long spans of sleep.

```
#Sedentary Active vs sleep time
ggplot(activity_sleep, aes(x = SedentaryMinutes, y = TotalMinutesAsleep)) +
  geom_point(color = "blue") +
  geom_smooth() +
  labs(title = "SedentaryMinutes vs Sleep", x = "SedentaryMinutes", y = "Minutes Asleep")
```

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



Observations * The graph shows that the more sedentary the user is, the amount of time he sleeps is less. * The user can be notified/reminded to actively spend minutes, for better sleep via Bellabeat gadgets.

```
# Summaries
dailyActivity_merged %>%
  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
```

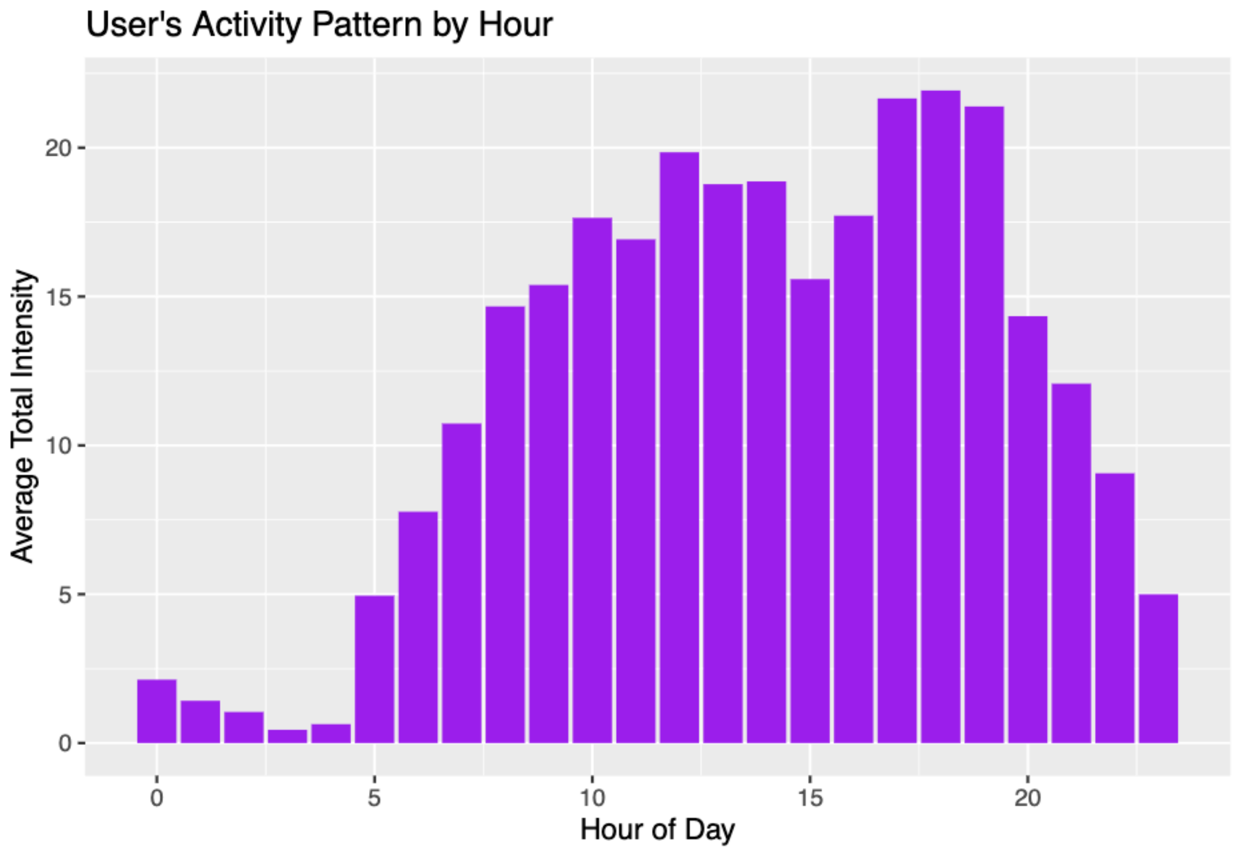
```
sleepDay_merged %>%
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##  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
```

```
weightLogInfo_merged %>%
  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
```

```
# Preferred time to be active
activity_hourly <- hourlyIntensities_merged %>%
  mutate(Hour = hour(mdy_hms(ActivityHour))) # Picks up only hours

hourly_pattern <- activity_hourly %>%
  group_by(Hour) %>%
  summarise(AverageTotalIntensity = mean(TotalIntensity, na.rm = TRUE))

ggplot(hourly_pattern, aes(x = Hour, y = AverageTotalIntensity)) +
  geom_col(fill = "purple") +
  labs(title = "User's Activity Pattern by Hour",
       x = "Hour of Day",
       y = "Average Total Intensity")
```



Observations

- It is seen from the data that the users are mostly Lightly Active.
- It can be seen that people mostly prefer to go for nap 1 time a day.
- The average time period people tend to sleep is around 7 hours, per day.
- It can be seen from the bar chart that the average Total intensity is high between 5pm-7pm. This might be the time slot that Bellabeat gadgets can focus on notifying the user's workout time.

Summary:

- The brand Bellabeat can think of linking the data with their other products(or apps) - so that users can track their diet plans, connecting the Activity tracking.
- Users can be informed about the arriving sleep slots/window. This is because, few users spend more time in bed to fall asleep.
- The time frame 5pm-7pm can be used as a window by Bellabeat to notify users of the activity slot.
- Though people who stay active are not directly connected to high sleep, people who are sedentary get low sleeping time. Bellabeat can use this data to remind the people with sedentary people to stay active.

Data source: FitBit Fitness Tracker Data from Kaggle