Bellabeats_project

2024-09-26

```r
# Load package
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(tinytex)
```

DAILY ACTIVITY ANALYSIS

```r
# Create data frame for daily activity
daily_activity <- read.csv("~/Desktop/Google Analytics/Google Analytics Project/Bellabeat_data/1dailyAct

# Take a look at the daily_activity data.
head(daily_activity)
```

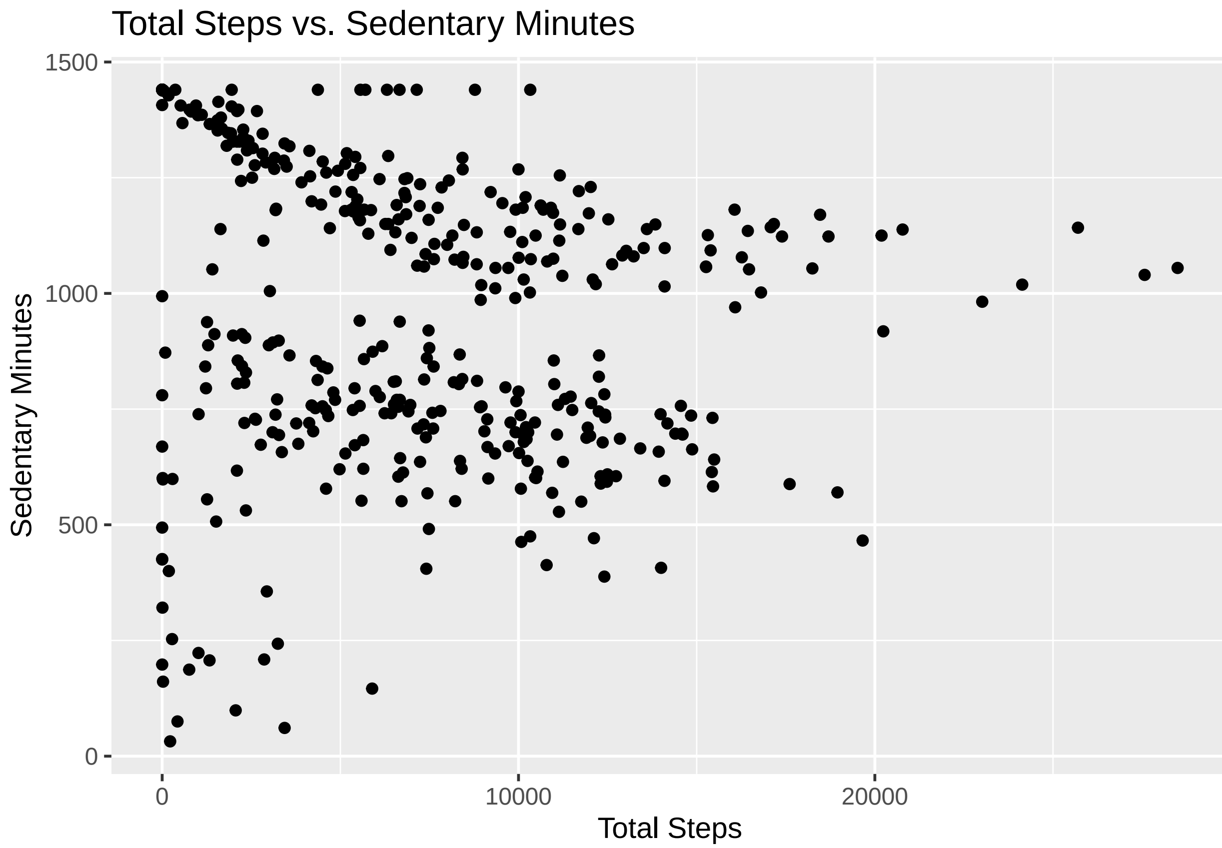
```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    3/25/2016      11004          7.11            7.11
## 2 1503960366    3/26/2016      17609         11.55           11.55
## 3 1503960366    3/27/2016      12736          8.53            8.53
## 4 1503960366    3/28/2016      13231          8.93            8.93
## 5 1503960366    3/29/2016      12041          7.85            7.85
## 6 1503960366    3/30/2016      10970          7.16            7.16
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               2.57                     0.46
## 2                        0               6.92                     0.73
## 3                        0               4.66                     0.16
## 4                        0               3.19                     0.79
## 5                        0               2.16                     1.09
## 6                        0               2.36                     0.51
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                4.07                       0                33
## 2                3.91                       0                89
## 3                3.71                       0                56
```

```
## 4                4.95                       0               39
## 5                4.61                       0               28
## 6                4.29                       0               30
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  12                  205              804     1819
## 2                  17                  274              588     2154
## 3                   5                  268              605     1944
## 4                  20                  224             1080     1932
## 5                  28                  243              763     1886
## 6                  13                  223             1174     1820
```

```r
# Summary Statistics:
daily_activity %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes
##  Min.   :    0   Min.   : 0.000   Min.   :  32.0  
##  1st Qu.: 1988   1st Qu.: 1.410   1st Qu.: 728.0  
##  Median : 5986   Median : 4.090   Median :1057.0  
##  Mean   : 6547   Mean   : 4.664   Mean   : 995.3  
##  3rd Qu.:10198   3rd Qu.: 7.160   3rd Qu.:1285.0  
##  Max.   :28497   Max.   :27.530   Max.   :1440.0
```

```r
# Plot of Daily Activity
ggplot(data = daily_activity, aes(x = TotalSteps, y = SedentaryMinutes)) +
  geom_point() +
  ggtitle("Total Steps vs. Sedentary Minutes") +
  xlab("Total Steps") +
  ylab("Sedentary Minutes")
```



Analysis for daily activity

TotalSteps:

- Minimum: 0 (some days with no recorded steps, potentially very inactive or data missing).
- Median: 7,406 (half of the participants take at least this many steps daily).
- Maximum: 36,019 (highly active individuals). # TotalDistance:
- Median: 5.25 miles (average daily distance is close to 5 miles, but some individuals are covering significantly more).
- Maximum: 28.03 miles (potential outliers or very active days) # SedentaryMinutes:
- Median: 1,057.5 minutes (around 17.6 hours spent sedentary daily on average).
- Maximum: 1,440 minutes (entire day sedentary, likely indicating inactive or non-usage days).

What these statistics tell us

- Most participants are relatively active, but there is a significant portion of the day that is sedentary.
- The scatter plot of TotalSteps vs. SedentaryMinutes shows how activity correlates with sedentary behavior. Since sedentary minutes are high in most observations, there's likely an inverse relationship between steps taken and sedentary minutes.
- Participants with fewer steps tend to have higher sedentary minutes, which is expected because if they aren't walking much,they're likely spending more time inactive.

Marketing

- Bellabeat could use this insight to target users both with higher and lower sedentary behavior.
- Market to less active users, motivate and encourage active small increases in daily movement and take breaks from sitting
- Market to more active users to improve their daily routines, focusing on increasing their steps further, adding advanced tracking features, promote fitness challenges/competitions.

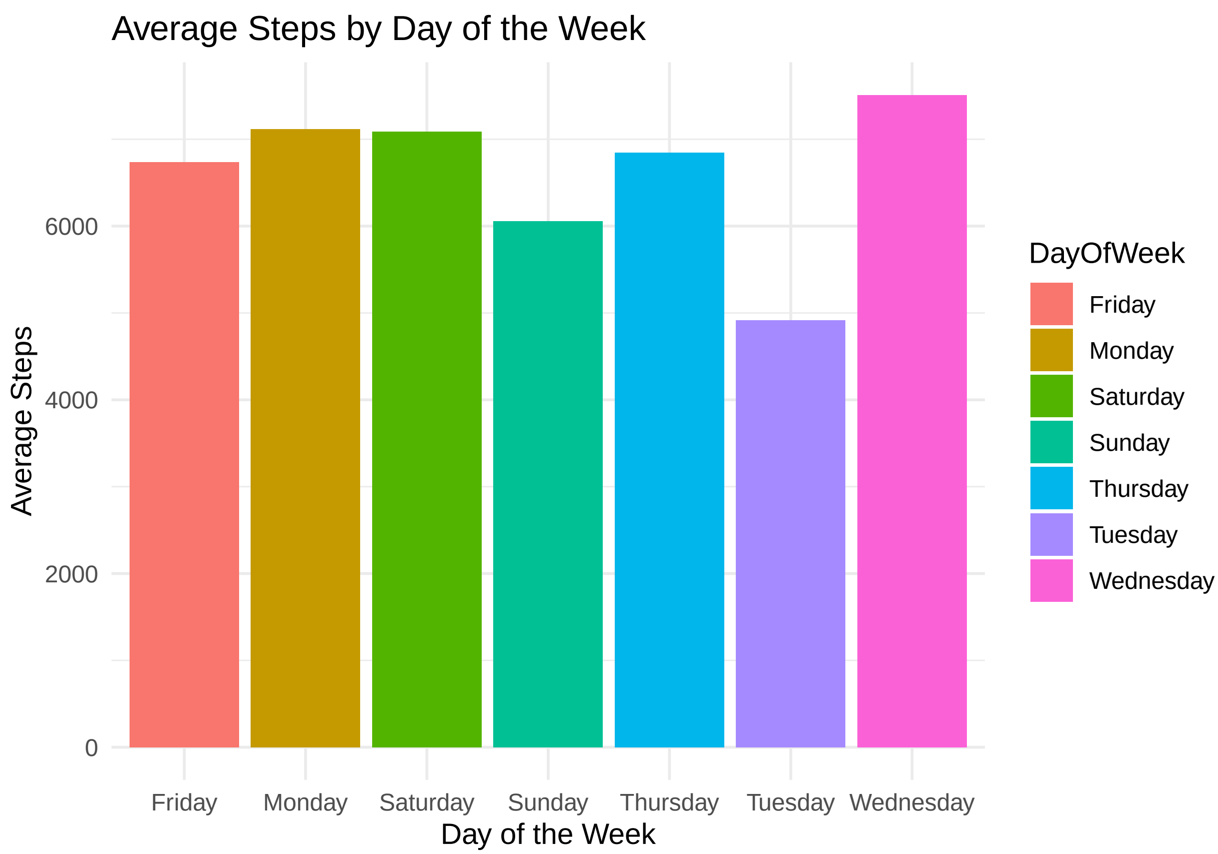
```r
# Convert Activity Date to Date and extract the day of the week
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate, format = "%m/%d/%Y")
daily_activity$DayOfWeek <- weekdays(daily_activity$ActivityDate)
```

```r
# Summarize total steps by day of the week
activity_by_day <- daily_activity %>%
  group_by(DayOfWeek) %>%
  summarise(AvgSteps = mean(TotalSteps), AvgSedentaryMinutes = mean(SedentaryMinutes))
```

```r
# View the summarized data
activity_by_day
```

```
## # A tibble: 7 x 3
##   DayOfWeek AvgSteps AvgSedentaryMinutes
##   <chr>        <dbl>               <dbl>
## 1 Friday       6738.               1055.
## 2 Monday       7119.               1033.
## 3 Saturday     7090.                977.
## 4 Sunday       6058.               1016.
## 5 Thursday     6847.               1055.
## 6 Tuesday      4915.                850.
## 7 Wednesday    7511.               1011.
```

```r
# Bar chart for average steps by day of the week
ggplot(activity_by_day, aes(x = DayOfWeek, y = AvgSteps, fill = DayOfWeek)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Steps by Day of the Week", x = "Day of the Week", y = "Average Steps") +
  theme_minimal()
```



Analysis of steps per day

- Tuesday and Saturday have the highest average steps (8,153 on Saturday, 8,125 on Tuesday).
- Sunday has the lowest average steps (6,933).
- Other weekdays (Monday, Wednesday, Thursday, Friday) have more moderate activity levels, generally around 7,400–7,800 steps.

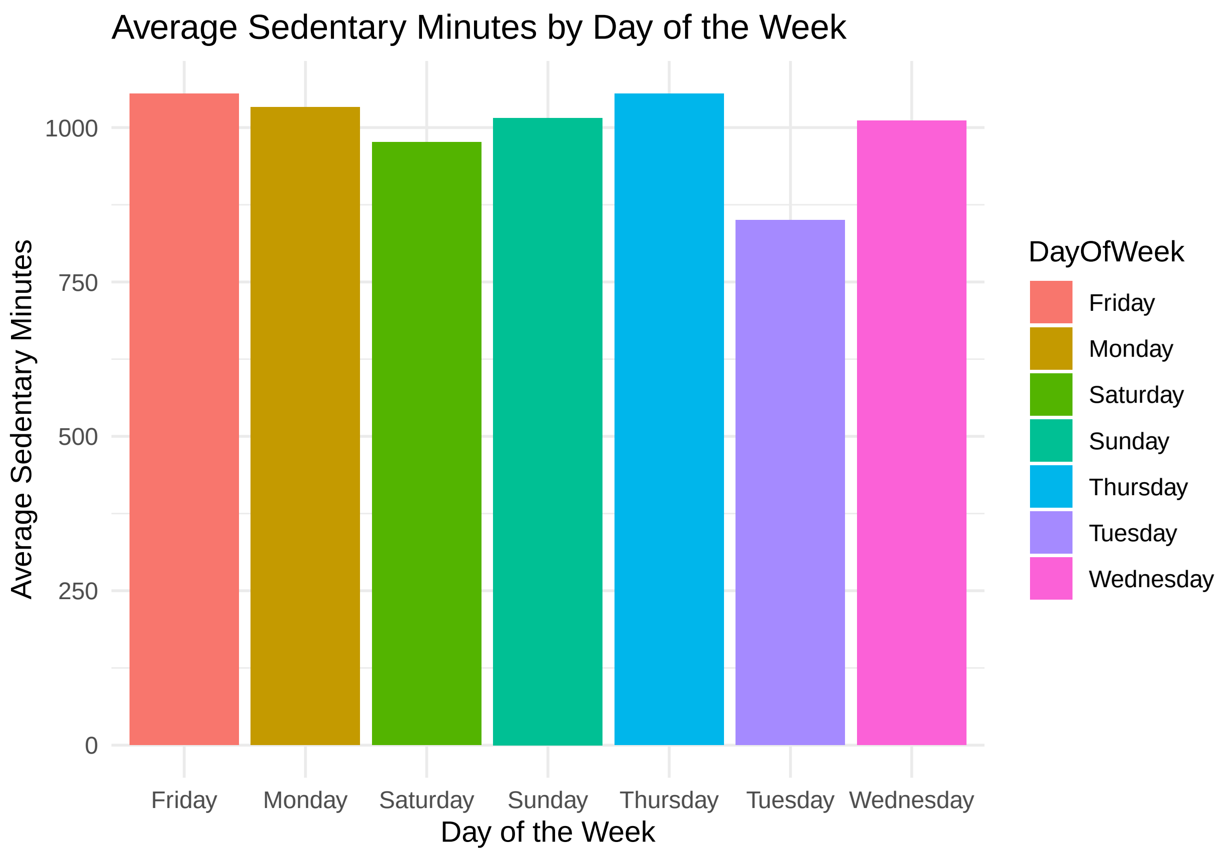
What these statistics tell us

- Saturday is the most active day of the week. This suggests that users might be using their time off to engage in physical activities, like walking or exercising.
- Sunday, however, shows a drop in steps, which could suggest that users tend to rest or be more sedentary before the start of the workweek.
- Tuesday is unexpectedly high, indicating that users are maintaining a good activity level at the beginning of the week.

Marketing

- Activity levels are slightly lower but still consistent from Monday to Friday. This suggests that users are maintaining moderate physical activity on workdays, which Bellabeat could encourage even more with targeted messaging.

```
# Bar chart for average sedentary minutes by day of the week
ggplot(activity_by_day, aes(x = DayOfWeek, y = AvgSedentaryMinutes, fill = DayOfWeek)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Sedentary Minutes by Day of the Week", x = "Day of the Week", y = "Average Seden
  theme_minimal()
```



Analysis of sedentary minutes per day

- Users are sedentary for long periods on weekdays, especially Monday and Tuesday (both with over 1,000 sedentary minutes).
- Sedentary minutes drop a bit on the weekends, especially Saturday.

What these statistics tell us

- Users are more likely to remain sedentary on weekdays, potentially due to work or school-related activities.
- Sedentary minutes are lower on Saturday, indicating users might be more active on this day.

Marketing

- Bellabeat can promote reminders to move or have short workouts throughout the weekdays, and take breaks from sitting

- Bellabeat can encourage users to build on weekend trends and remain active throughout the weekend, such as motivating them to keep up their good habits into Sunday. They can promote wellness achievements or personalized notifications congratulating users for their high activity on Saturdays.

Summary

- This project focuses on analyzing smart device usage trends, specifically daily activity and sedentary behavior, to provide marketing insights for Bellabeat, a company producing health-focused devices.
- By examining data from users, the analysis reveals key trends, such as higher activity levels on Saturday and Tuesday and increased sedentary time during the weekdays, especially on Monday and Tuesday. Tuesday has both high activity and sedentary levels.
- Based on these insights, Bellabeat could tailor their marketing strategies by:
  - Encourage Sunday activity through app challenges and reminders.
  - Promote midweek activity breaks to reduce sedentary behavior.
  - Leverage Saturday's activity highs by offering achievements and fitness goals.
- These insights help Bellabeat target both inactive and active users, promoting personalized features to boost engagement and overall wellness.

SLEEP DATA ANALYSIS

```
# Create dataframe for the sleep data.
sleep_day <- read.csv("~/Desktop/Google Analytics/Google Analytics Project/Bellabeat_data/sleepDay_merg
```

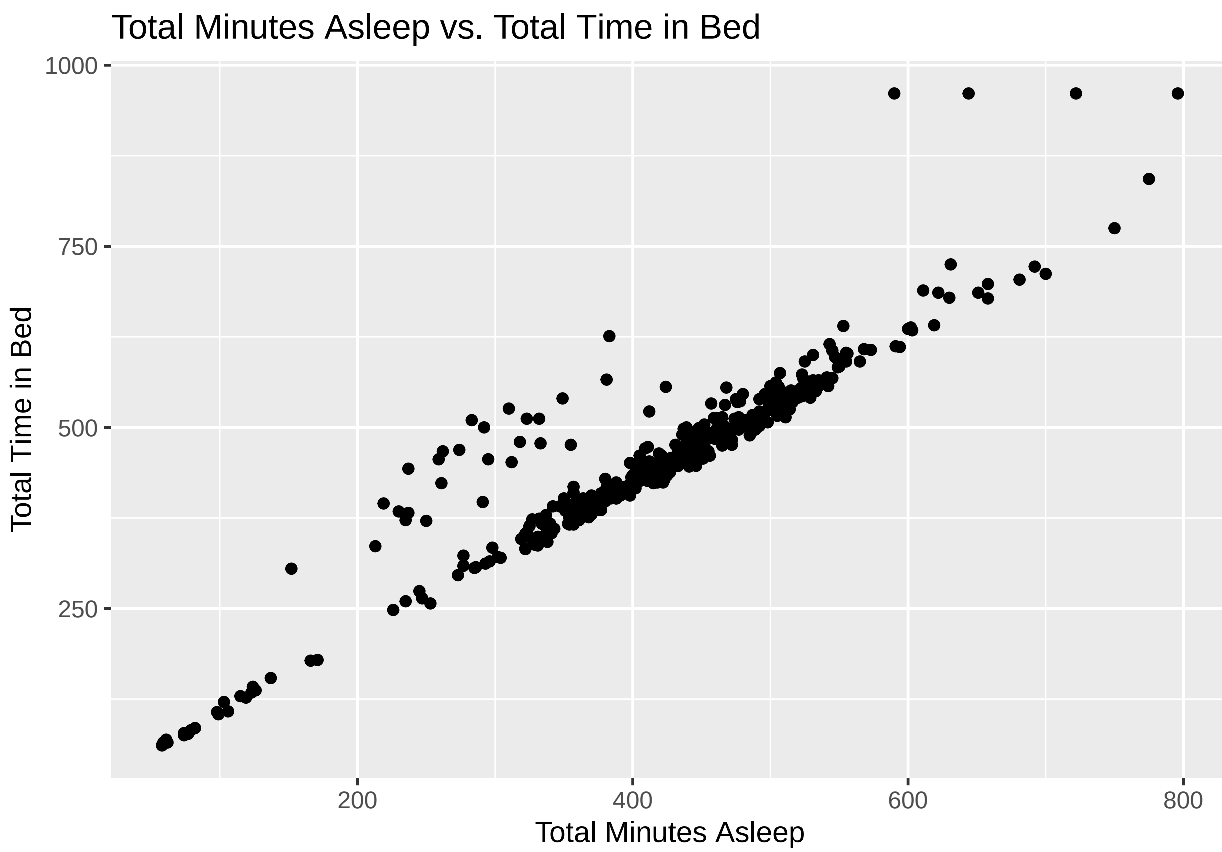
```
# Take a look at the sleep_day data.
head(sleep_day)
```

```
##           Id              SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                 1                327
## 2 1503960366 4/13/2016 12:00:00 AM                 2                384
## 3 1503960366 4/15/2016 12:00:00 AM                 1                412
## 4 1503960366 4/16/2016 12:00:00 AM                 2                340
## 5 1503960366 4/17/2016 12:00:00 AM                 1                700
## 6 1503960366 4/19/2016 12:00:00 AM                 1                304
##   TotalTimeInBed
## 1            346
## 2            407
## 3            442
## 4            367
## 5            712
## 6            320
```

```
# Summary Statistics:
sleep_day %>%
  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
```

```
# Plot of sleep vs time in bed
ggplot(data=sleep_day, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) +
  geom_point() +
  ggtitle("Total Minutes Asleep vs. Total Time in Bed") +
  xlab("Total Minutes Asleep") +
  ylab("Total Time in Bed")
```



The more time in bed, the more amount of minutes asleep.

If more time in bed but less minutes asleep, then users are not getting adequate sleep.

```r
# Calculate the average amount of sleep (in minutes)
average_sleep_minutes <- mean(sleep_day$TotalMinutesAsleep)
average_sleep_hours <- average_sleep_minutes / 60
average_sleep_minutes
```

```
## [1] 419.4673
```

```r
average_sleep_hours
```

```
## [1] 6.991122
```

Users are getting an average sleep of 419 minutes or 7 hours per night.

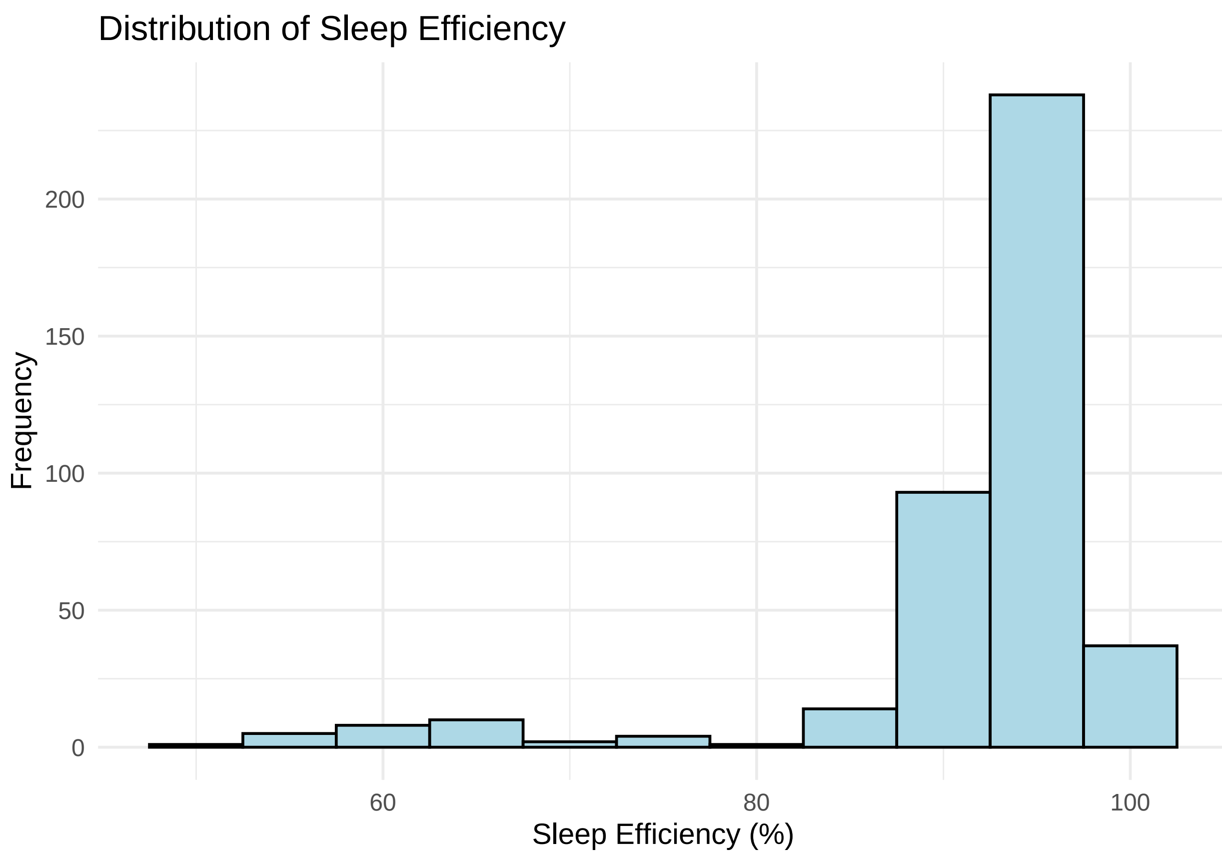
```r
# Create a sleep efficiency column
sleep_day <- sleep_day %>%
  mutate(SleepEfficiency = (TotalMinutesAsleep / TotalTimeInBed) * 100)

summary(sleep_day$SleepEfficiency)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   49.84   91.22   94.31   91.68   96.07  100.00
```

Summary Statistics of Sleep Efficiency

- Minimum Sleep Efficiency: 49.84% (some users are spending significant time in bed but not sleeping well).
- 1st Quartile, 91.22% (25% of users have sleep efficiency below this value).
- Median: 94.31% (half of the users have sleep efficiency of 94.31% or higher).
- Mean: 91.68% (average sleep efficiency across all users).
- 3rd Quartile, 96.07% (75% of users have sleep efficiency below this value).
- Maximum: 100%, (some users are spending all their time in bed sleeping).

```r
# Histogram of sleep efficiency
ggplot(sleep_day, aes(x = SleepEfficiency)) +
  geom_histogram(binwidth = 5, fill = "lightblue", color = "black") +
  labs(title = "Distribution of Sleep Efficiency",
       x = "Sleep Efficiency (%)",
       y = "Frequency") +
  theme_minimal()
```



Summary statistics of sleep efficiency histogram and what it tells us

- Most users have high sleep efficiency, clustering between 91% and 100%, this suggests that the majority of users sleep efficiently; spending most of their time in bed asleep.
- A few users have lower sleep efficiency (below 85%), which could indicate issues with sleep quality or disruptions during the night.

Marketing and Summary

- High sleep efficiency suggests that Bellabeat's users might already have a solid foundation for sleep, but the company could offer features to optimize sleep routines or help users improve their efficiency even further.
- For users with low sleep efficiency, Bellabeat could market sleep improvement features, such as personalized recommendations to reduce awake time in bed (reduced screen time, red light before bed, practising good bed time habits).
- Future analyses could explore the relationship between physical activity and sleep efficiency, potentially highlighting how Bellabeat's devices can help users improve their overall wellness by optimizing both activity and sleep patterns.