

FitBit Fitness Tracker_Bellabeat Case Study

A.A RAFID RAIHAN

2024-06-01

Ask

This project was undertaken to analyze smart device usage data to gain insights into how consumers use non-Bellabeat smart devices.

The stakeholder provided some datasets from smart devices. The questions asked by the stakeholder were:

- What are the trends in smart device usage?
- How can these trends be applied to Bellabeat's customers?
- How can these trends help influence Bellabeat's marketing strategy?

Key stakeholders included:

- Urška Sršen: Bellabeat's co-founder and Chief Creative Officer
- Sando Mur: Mathematician, co-founder of Bellabeat, and key member of the executive team
- Marketing analytics teammates: Responsible for collecting, analyzing, and reporting data that guides Bellabeat's marketing strategy

Preaper

The data used for this case study is titled "FitBit Fitness Tracker Data" share by Mobius on Kaggle and has been licensed to use by public domain. The data is said to have 30 eligible participants who responded to a survey via Amazon Mechanical Turk between 3/12/2016-5/12/2016 and consented to have their personal tracker data shared. The output data consists of 29 CSV files, in long format, split between the date ranges of 3/12/2016-4/11/2016 and 4/12/2016-5/12/2016 and further broken down into data sets by activity type (sleep, weight, daily activity, hourly, minute). The data is organized by using a unique Id per participant to track their logs by day and hour.

Process

Install Library and Packages

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'  
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'  
## (as 'lib' is unspecified)
```

```

install.packages("tidyr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

install.packages("skimr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

install.packages("lubridate")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

install.packages("janitor")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

library(tidyverse)

## -- 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(dplyr)
library(tidyr)
library(skimr)
library(janitor)

##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

library(lubridate)

```

Load Dataset

```

## 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.

```

```
## 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.
```

Review Data using head(), colnames() dan glimpse()

```
head(SleepDay)
```

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

```
head(DailyActivity)
```

```
## # A tibble: 6 x 15
##       Id ActivityDate TotalSteps TotalDistance TrackerDistance
##       <dbl> <chr>          <dbl>          <dbl>          <dbl>
## 1 1503960366 4/12/2016      13162          8.5            8.5
## 2 1503960366 4/13/2016      10735          6.97           6.97
## 3 1503960366 4/14/2016      10460          6.74           6.74
## 4 1503960366 4/15/2016       9762          6.28           6.28
## 5 1503960366 4/16/2016      12669          8.16           8.16
## 6 1503960366 4/17/2016       9705          6.48           6.48
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

```
glimpse(SleepDay)
```

```
## Rows: 413
## Columns: 5
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay    <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~
## $ TotalSleepRecords <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <dbl> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~
## $ TotalTimeInBed   <dbl> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~
```

```
glimpse(DailyActivity)
```

```
## Rows: 940
## Columns: 15
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
```

```
## $ TotalSteps          <dbl> 13162, 10735, 10460, 9762, 12669, 9705, 13019~
## $ TotalDistance       <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ TrackerDistance     <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance  <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ LightActiveDistance  <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes    <dbl> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ FairlyActiveMinutes  <dbl> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ LightlyActiveMinutes <dbl> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ SedentaryMinutes     <dbl> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ Calories             <dbl> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~
```

```
colnames(SleepDay)
```

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

```
colnames(DailyActivity)
```

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

Process

Cleansing and transformation data

Clean column

```
SleepDay <- SleepDay %>%
  rename(Date = SleepDay) %>%
  clean_names()

DailyActivity <- DailyActivity %>%
  rename(Date = ActivityDate) %>%
  clean_names()
```

Clean column date

```
SleepDay$date <- as.Date(SleepDay$date, format = "%m/%d/%Y")
DailyActivity$date <- as.Date(DailyActivity$date, format= "%m/%d/%Y")
```

Cekc result transformation

```
head(SleepDay)
```

```
## # A tibble: 6 x 5
```

```
##           id date          total_sleep_records total_minutes_asleep total_time_in_bed
##          <dbl> <date>                <dbl>                <dbl>                <dbl>
## 1  1.50e9 2016-04-12                    1                    327                    346
## 2  1.50e9 2016-04-13                    2                    384                    407
## 3  1.50e9 2016-04-15                    1                    412                    442
## 4  1.50e9 2016-04-16                    2                    340                    367
## 5  1.50e9 2016-04-17                    1                    700                    712
## 6  1.50e9 2016-04-19                    1                    304                    320
```

```
head(DailyActivity)
```

```
## # A tibble: 6 x 15
##           id date          total_steps total_distance tracker_distance
##          <dbl> <date>                <dbl>                <dbl>                <dbl>
## 1 1503960366 2016-04-12          13162                8.5                8.5
## 2 1503960366 2016-04-13          10735                6.97               6.97
## 3 1503960366 2016-04-14          10460                6.74               6.74
## 4 1503960366 2016-04-15           9762                6.28               6.28
## 5 1503960366 2016-04-16          12669                8.16               8.16
## 6 1503960366 2016-04-17           9705                6.48               6.48
## # i 10 more variables: logged_activities_distance <dbl>,
## #   very_active_distance <dbl>, moderately_active_distance <dbl>,
## #   light_active_distance <dbl>, sedentary_active_distance <dbl>,
## #   very_active_minutes <dbl>, fairly_active_minutes <dbl>,
## #   lightly_active_minutes <dbl>, sedentary_minutes <dbl>, calories <dbl>
```

Analyst

How many unique users are there each dataset?

```
n_distinct(SleepDay$id)
```

```
## [1] 24
```

```
n_distinct(DailyActivity$id)
```

```
## [1] 33
```

How many observations are there in each dataframe?

```
nrow(SleepDay)
```

```
## [1] 413
```

```
nrow(DailyActivity)
```

```
## [1] 940
```

sleep patterns

```
summary(SleepDay)
```

```
##           id          date          total_sleep_records
##  Min.   :1.504e+09   Min.   :2016-04-12   Min.   :1.000
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19   1st Qu.:1.000
```

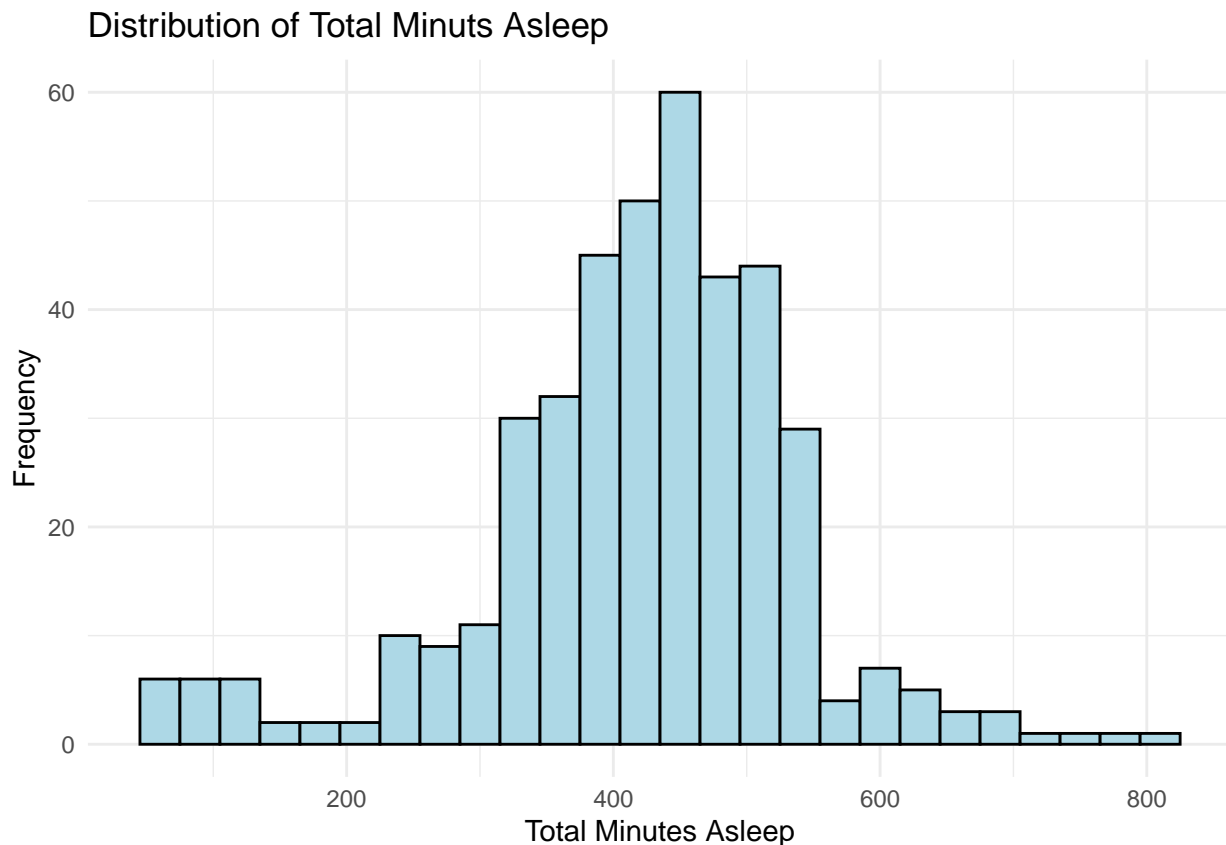
```
## Median :4.703e+09   Median :2016-04-27   Median :1.000
## Mean   :5.001e+09   Mean   :2016-04-26   Mean   :1.119
## 3rd Qu.:6.962e+09   3rd Qu.:2016-05-04   3rd Qu.:1.000
## Max.   :8.792e+09   Max.   :2016-05-12   Max.   :3.000
## total_minutes_asleep total_time_in_bed
## Min.    : 58.0      Min.    : 61.0
## 1st Qu.:361.0      1st Qu.:403.0
## Median :433.0      Median :463.0
## Mean    :419.5      Mean    :458.6
## 3rd Qu.:490.0      3rd Qu.:526.0
## Max.    :796.0      Max.    :961.0
```

- Total Minutes Asleep: The average user sleep time is about 419 minutes (about 7 hours).
- Total Time in Bed: The average time spent in bed is about 458 minutes (about 7.6 hours).

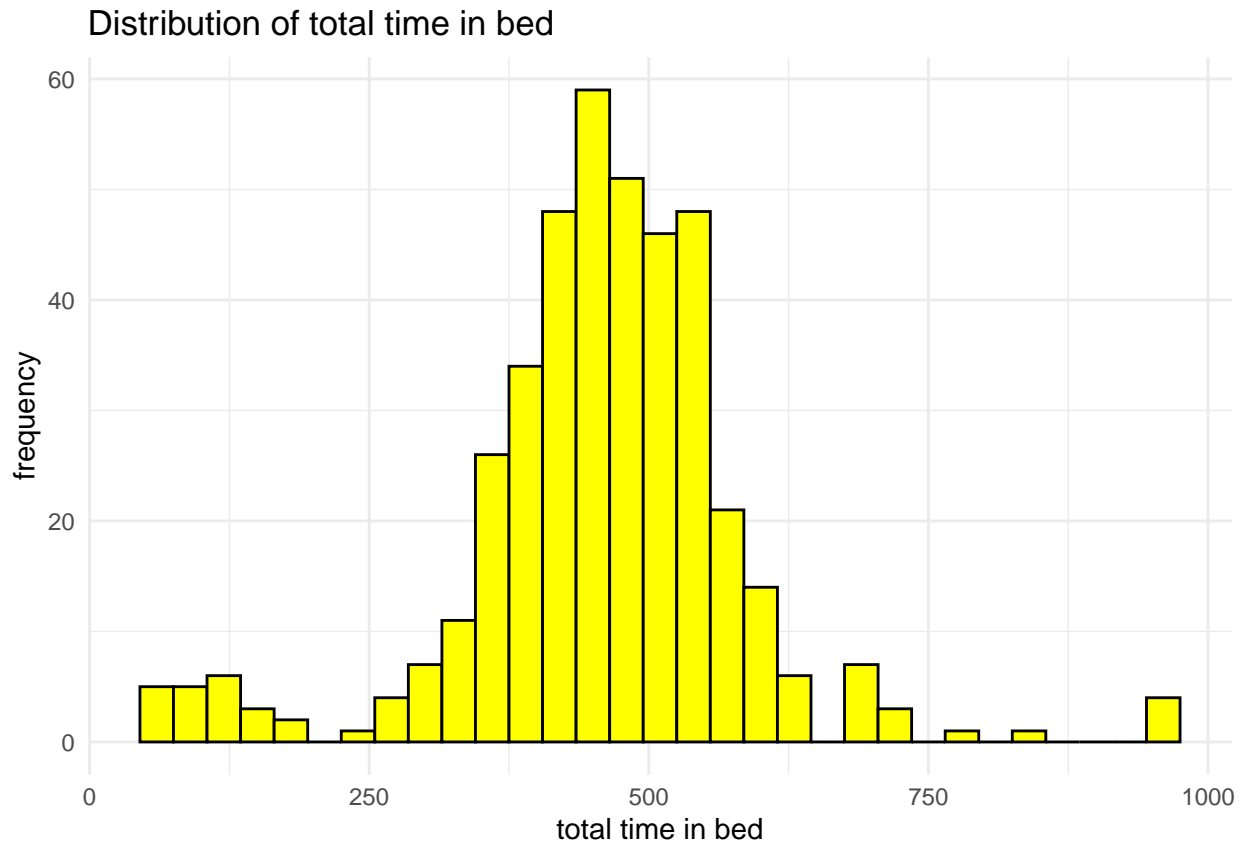
View distribution on sleep patterns

- Visualization of sleep patterns will give an idea of how users manage their sleep time.

```
ggplot(SleepDay) +
  geom_histogram(aes(x=total_minutes_asleep), binwidth = 30 , color = "black", fill = "lightblue") +
  labs(title = "Distribution of Total Minuts Asleep", x = "Total Minutes Asleep", y = "Frequency") +
  theme_minimal()
```



```
ggplot(SleepDay) +
  geom_histogram(aes(x = total_time_in_bed), binwidth = 30, color = "black", fill = "yellow") +
  labs(title = "Distribution of total time in bed", x = "total time in bed", y = "frequency") +
  theme_minimal()
```



- Distribution of Total Minutes Asleep: Most users sleep around 300-400 minutes per night.
- Distribution of Total Time in Bed: Time spent in bed ranges from 300-500 minutes for most users.

Daily Activities

`summary(DailyActivity)`

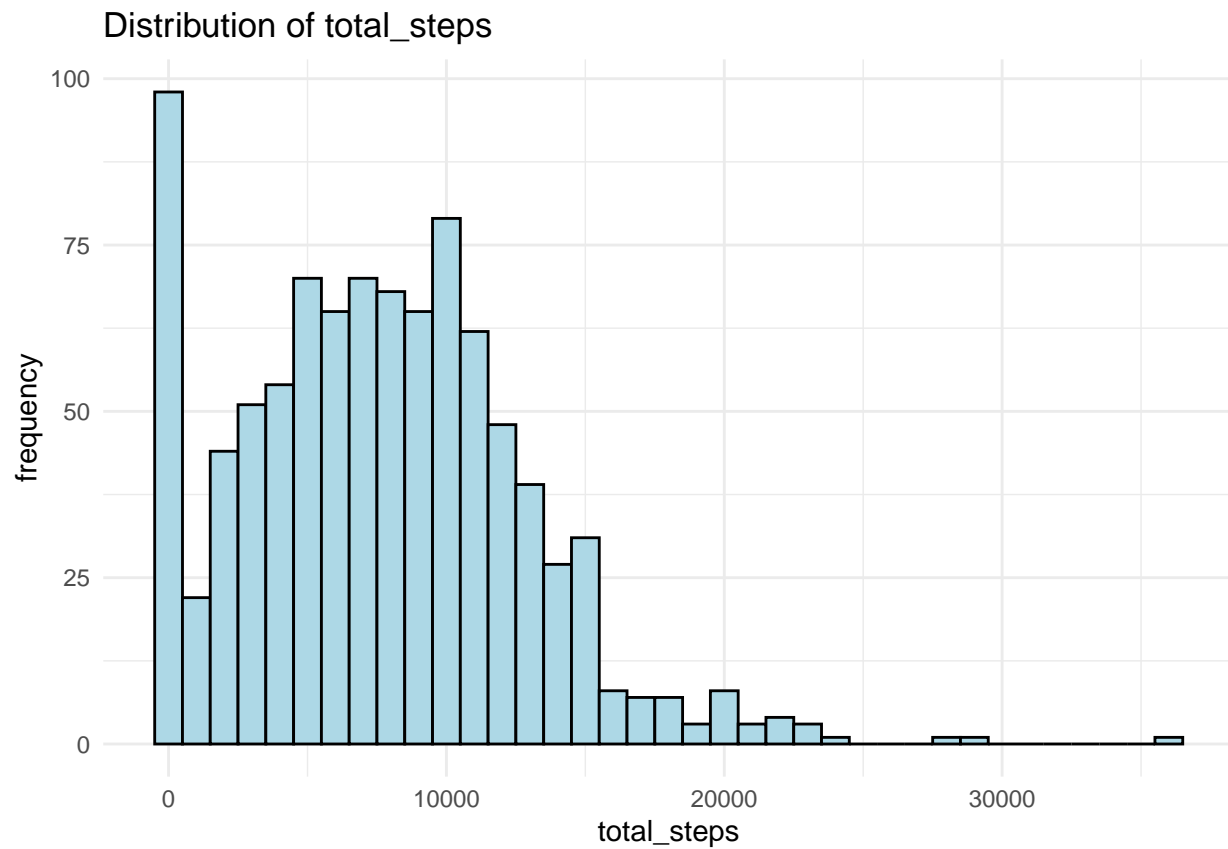
```
##          id          date      total_steps  total_distance
##  Min.   :1.504e+09  Min.   :2016-04-12  Min.    :    0  Min.    : 0.000
## 1st Qu.:2.320e+09  1st Qu.:2016-04-19  1st Qu.: 3790  1st Qu.: 2.620
## Median :4.445e+09  Median :2016-04-26  Median : 7406  Median : 5.245
## Mean   :4.855e+09  Mean   :2016-04-26  Mean    : 7638  Mean    : 5.490
## 3rd Qu.:6.962e+09  3rd Qu.:2016-05-04  3rd Qu.:10727  3rd Qu.: 7.713
## Max.   :8.878e+09  Max.   :2016-05-12  Max.    :36019  Max.    :28.030
## tracker_distance logged_activities_distance very_active_distance
##  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
## moderately_active_distance light_active_distance sedentary_active_distance
##  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
## very_active_minutes fairly_active_minutes lightly_active_minutes
## 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
## sedentary_minutes    calories
## Min.    : 0.0          Min.    : 0
## 1st Qu.: 729.8         1st Qu.:1828
## Median :1057.5         Median :2134
## Mean    : 991.2         Mean    :2304
## 3rd Qu.:1229.5         3rd Qu.:2793
## Max.    :1440.0        Max.    :4900
```

- Total Steps: The average user's daily steps are about 7638 steps.
- Total Distance: The average distance traveled by users is about 5.5 miles.
- Very Active Minutes: The average very active time is about 21 minutes.
- Lightly Active Minutes: The average light activity time is about 193 minutes.
- Sedentary Minutes: The average sedentary time is about 991 minutes (about 16.5 hours).

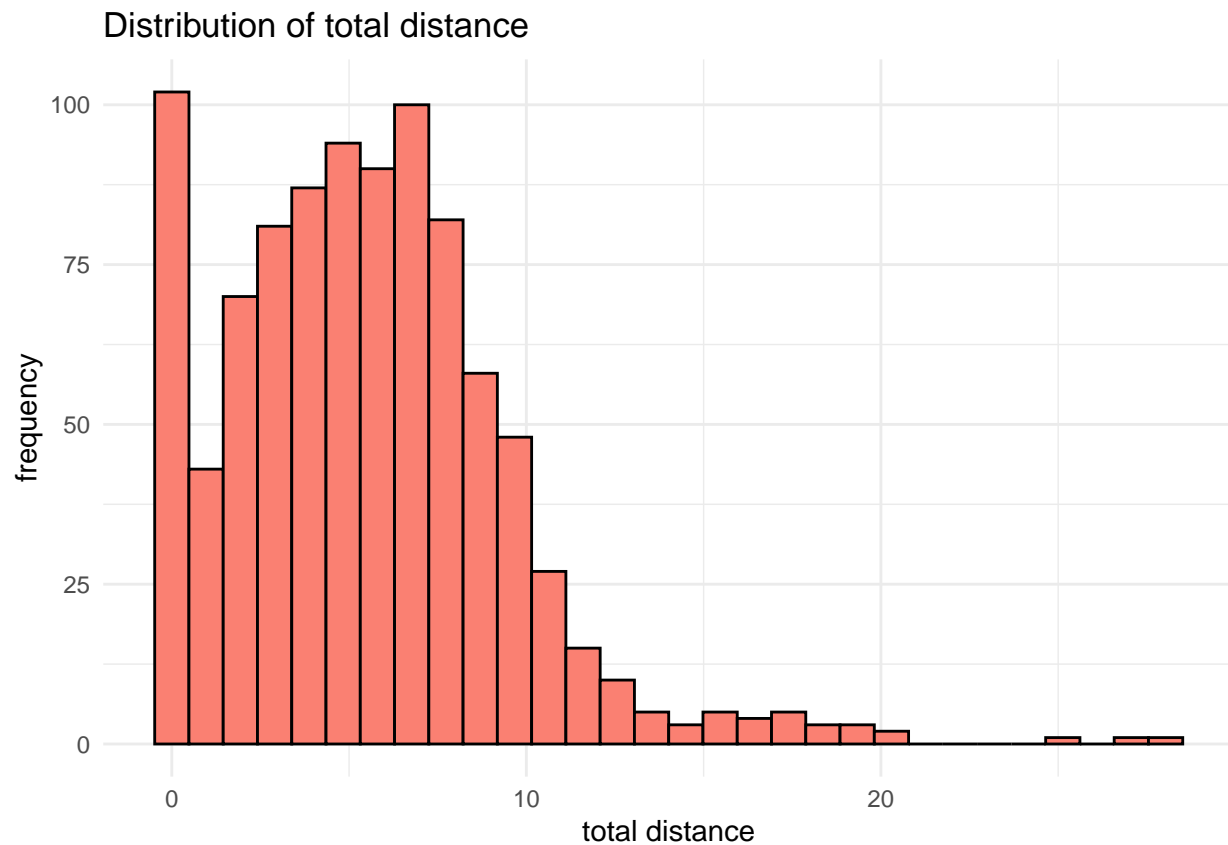
view distribution on daily activity

```
ggplot(DailyActivity) +
  geom_histogram(aes(total_steps), binwidth = 1000, color = "black", fill = "lightblue") +
  labs(title = "Distribution of total_steps", x = "total_steps", y= "frequency") +
  theme_minimal()
```

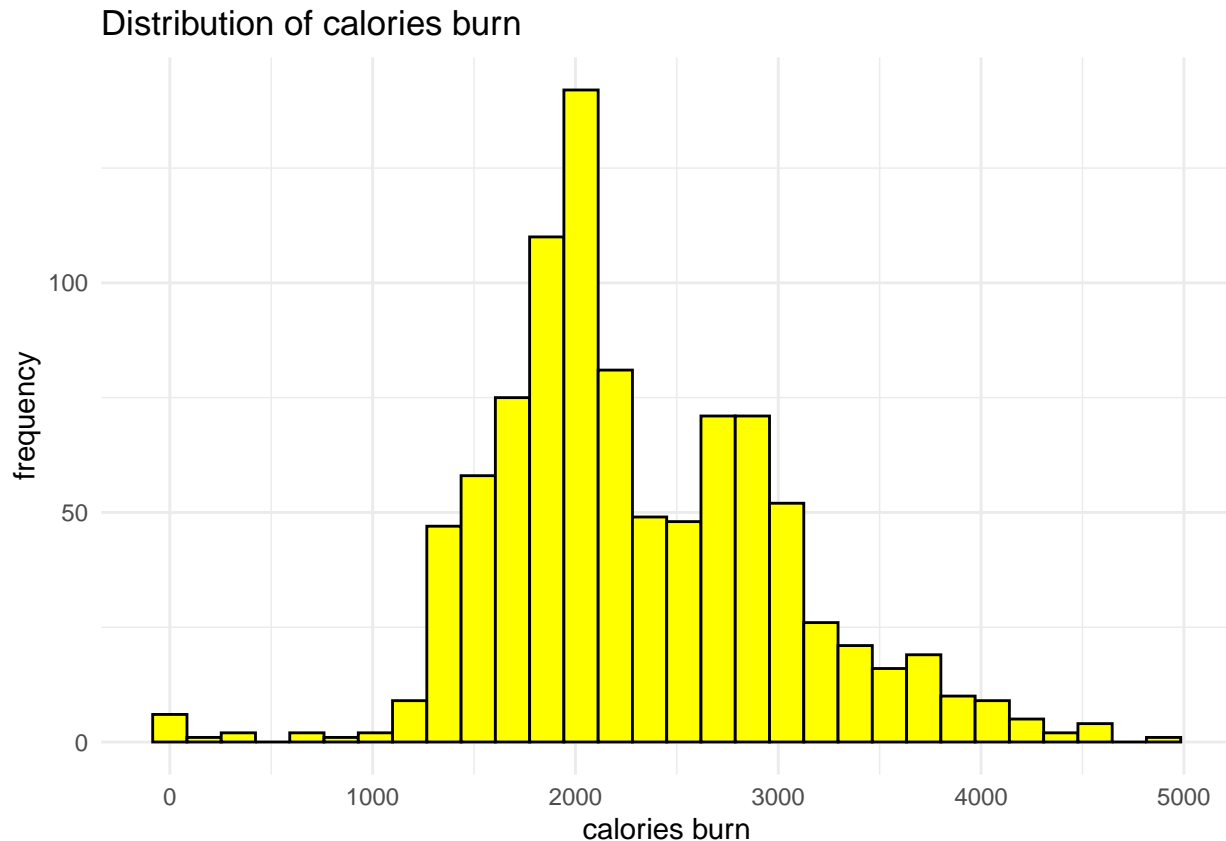
```
ggplot(DailyActivity) +  
  geom_histogram(aes(total_distance), color = "black", fill = "salmon") +  
  labs(title = "Distribution of total distance", x="total distance", y = "frequency") +  
  theme_minimal()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggplot(DailyActivity) +  
  geom_histogram(aes(calories), color = "black", fill = "yellow") +  
  labs(title = "Distribution of calories burn", x = "calories burn", y = "frequency") +  
  theme_minimal()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



- Total Steps Distribution: Most users walk around 5000-10000 steps per day.
- Total Distance Distribution: Most users travel about 4-8 miles per day.
- Calories Burned Distribution: Calories burned daily range from 1500-3000 calories for most users.

Hypothesis

The use of hypothesis testing methods to evaluate some assumptions or statements about the data

Hypothesis 1: Is the Average User Sleep Time Sufficient?

- Null Hypothesis (H0): The average user sleep time is 8 hours (480 minutes) per night.
- Alternative Hypothesis (H1): The average user sleep time is less than 8 hours (480 minutes) per night.

```
mu <- 480
t_test_sleep <- t.test(SleepDay$total_minutes_asleep, mu=mu, alternative = "less")
print(t_test_sleep)
```

```
##
## One Sample t-test
##
## data: SleepDay$total_minutes_asleep
## t = -10.395, df = 412, p-value < 2.2e-16
## alternative hypothesis: true mean is less than 480
## 95 percent confidence interval:
##      -Inf 429.0675
## sample estimates:
```

```
## mean of x
## 419.4673
```

Hypothesis 2: Do Active Users Burn More Calories?

- Null Hypothesis (H0): There is no difference in average calories burned between highly active and non-highly active users.
- Alternative Hypothesis (H1): The average calories burned by highly active users is higher than non-highly active users.

Now let's test the second hypothesis using a two-sample t-test. We will compare

- very active users (VeryActiveMinutes > 30 minutes per day)
- users who are not very active (VeryActiveMinutes ≤ 30 minutes per day).

```
active_users <- DailyActivity %>%
  filter(very_active_minutes > 30) %>%
  pull(calories)

inactive_users <- DailyActivity %>%
  filter(very_active_minutes <= 30) %>%
  pull(calories)

t_test_activity <- t.test(active_users, inactive_users, alternative="greater")
print(t_test_activity)
```

```
##
## Welch Two Sample t-test
##
## data: active_users and inactive_users
## t = 14.666, df = 358.09, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 681.8428 Inf
## sample estimates:
## mean of x mean of y
## 2873.239 2105.016
```

From the hypothesis test results, it says that

- Hypothesis 1: Is the Average User Sleep Time Sufficient?

The alternative hypothesis is acceptable because the p-value is very small (well below 0.05), we reject the null hypothesis. This means that the average user sleep time is indeed significantly less than 8 hours per night.

- Hypothesis 2: Do Active Users Burn More Calories?

The alternative hypothesis is acceptable because of the 95 percent confidence interval value, with that high confidence value we reject the null hypothesis.

Recommendations

1. Product Development:

- More Detailed Sleep Monitoring Features: Bellabeat can develop more detailed features to monitor sleep quality and provide improvement suggestions.
- Daily Activity Integration: Expand the device's ability to track different types of physical activity and provide recommendations accordingly.

2. Marketing Strategy:

- Education Campaign: Conduct educational campaigns that inform customers about the importance of physical activity and adequate sleep.
- Offer Personalization: Offer personalized products and services based on users' activity data and sleep patterns.