CapstoneProject

Nguyen Nguyen

1/28/2022

## STEP 1: ASK

#### 1.0 Background

You are a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. You have been asked to focus on one of Bellabeat’s products and analyze smart device data (ie. FitBit fitness tracker usage data) to gain insight into how consumers are using their smart devices. The insights you discover will then help guide marketing strategy for the company. You will present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat’s marketing strategy.

#### 1.2 Business Task:

Analyze FitBit Fitness Tracker Data to gain insights into how consumers are using the FitBit app and discover trends and insights for Bellabeat marketing strategy.

#### 1.3 Business Objectives:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

### 1.4 Deliverables:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of analysis
5. Supporting visualizations and key findings
6. Your top high-level content recommendations based on your analysis

#### 1.5 Key Stakeholders:

Urška Sršen: Bellabeat’s cofounder and Chief Creative Officer

Sando Mur: Mathematician and Bellabeat’s cofounder; key member of the Bellabeat executive team

Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat’s marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat’’s mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

## STEP 2: PREPARE

#### 2.1 Information on Data Source:

1. The data is publicly available on Kaggle: FitBit Fitness Tracker Data and stored in 18 csv files.
2. Respondents from a dispersed survey conducted via Amazon Mechanical Turk between March 12 and May 12, 2016 generated this data.
3. The Data is made up of 30 FitBit users who agreed to have their personal tracker data shared.

#### 2.2 Limitations of Data Set:

1. Data was collected in 2016. Data may not be recent or meaningful because users’ daily activities, fitness and sleeping habits, diet and food consumption may have modified since then.
2. Sample size of 30 unidentifed gender FitBit users is not representative of the entire population. Considered that Bellabeat product is tailored for female users, the data may not be the most accurate
3. As data is acquired through a survey, hence unable to determine the integrity or accuracy of data.

#### 2.3 Is Data ROCCC?

A good data source is ROCCC which stands for Reliable, Original, Comprehensive, Current, and Cited.

Reliable - LOW - This poll is unreliable since just 30 people responded.

Original - LOW - Provided by a third party (Amazon Mechanical Turk)

Comprehensive - MED - Most of Bellabeat’s products have criteria that match those of.

Current - LOW - Data is 6 years old and no longer relevant.

Cited - LOW - Data was obtained from a third party, hence the source is unclear.

Overall, the dataset is of poor quality, and it is not recommended that business recommendations be based on it.

#### 2.4 Data Selection:

The “dailyActivity\_merged.csv” file is chosen and copied for analysis.

## STEP 3: PROCESS

We are using R for data preparation and processing.

#### 3.1 Preparing the Environment:

Import the library that we will use to clean, process, and visulaize the data

library(tidyverse)  
library(readr)  
library(dplyr)  
library(tibble)  
library(psych)  
library(ggplot2)  
library(tidyverse)  
library(RColorBrewer)

#### 3.2 Importing data set

Reading in the file to be used for analysis.

# importing the csv file that I will be using  
dailyActivity\_merged <- read\_csv("/Users/nguyennguyen/Downloads/Fitabase Data 4.12.16-5.12.16/dailyActivity\_merged.csv")

#### 3.3 Data cleaning and manipulation

Previewing the data using the glimpse function to familiralize with the data

glimpse(dailyActivity\_merged)

## 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…

Find out if there is any missing data in the dailyActivity\_merged

# obtain the number of missing data points per column  
missing\_values\_count = sum(is.na(dailyActivity\_merged))  
missing\_values\_count

## [1] 0

Finding out the basic structure of daily\_activity:

1. number of rows and columns
2. name of columns
3. type of values

# show structure of data   
str(dailyActivity\_merged)

## spec\_tbl\_df [940 × 15] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ ActivityDate : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...  
## $ TotalSteps : num [1:940] 13162 10735 10460 9762 12669 ...  
## $ TotalDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...  
## $ TrackerDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...  
## $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ VeryActiveDistance : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...  
## $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...  
## $ LightActiveDistance : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...  
## $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ VeryActiveMinutes : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...  
## $ FairlyActiveMinutes : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...  
## $ LightlyActiveMinutes : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...  
## $ SedentaryMinutes : num [1:940] 728 776 1218 726 773 ...  
## $ Calories : num [1:940] 1985 1797 1776 1745 1863 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. ActivityDate = col\_character(),  
## .. TotalSteps = col\_double(),  
## .. TotalDistance = col\_double(),  
## .. TrackerDistance = col\_double(),  
## .. LoggedActivitiesDistance = col\_double(),  
## .. VeryActiveDistance = col\_double(),  
## .. ModeratelyActiveDistance = col\_double(),  
## .. LightActiveDistance = col\_double(),  
## .. SedentaryActiveDistance = col\_double(),  
## .. VeryActiveMinutes = col\_double(),  
## .. FairlyActiveMinutes = col\_double(),  
## .. LightlyActiveMinutes = col\_double(),  
## .. SedentaryMinutes = col\_double(),  
## .. Calories = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

Note that the data frame has 940 rows and 15 columns and all of the columns are num type except for the ActivityDate column, which is chr type.

Check for the unique ID values in the table

# count distinct value of "Id"  
unique\_id <- n\_distinct(dailyActivity\_merged$Id)  
print(paste("Number of unique ID: ", unique\_id))

## [1] "Number of unique ID: 33"

Create a copy of the data frame andadding three columns: DayOfWeek, TotalExerciseMinutes, and TotalExerciseHours to manipulate and make further calculation for analysis.

# create a copy of the table to manipulate  
# adding three columns: DayOfWeek, TotalExerciseMinutes, and TotalExerciseHours  
# fill the columns with NaN  
copy\_daily <- dailyActivity\_merged %>%  
 add\_column(DayOfWeek = NaN, .after = "ActivityDate") %>%  
 add\_column(TotalExcerciseMinutes = NaN, .after = "SedentaryMinutes") %>%  
 add\_column(TotalExcerciseHours = NaN, .after = "TotalExcerciseMinutes")   
  
# verify that the columns had been added successfully  
str(copy\_daily)

## spec\_tbl\_df [940 × 18] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ ActivityDate : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...  
## $ DayOfWeek : num [1:940] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ...  
## $ TotalSteps : num [1:940] 13162 10735 10460 9762 12669 ...  
## $ TotalDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...  
## $ TrackerDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...  
## $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ VeryActiveDistance : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...  
## $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...  
## $ LightActiveDistance : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...  
## $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...  
## $ VeryActiveMinutes : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...  
## $ FairlyActiveMinutes : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...  
## $ LightlyActiveMinutes : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...  
## $ SedentaryMinutes : num [1:940] 728 776 1218 726 773 ...  
## $ TotalExcerciseMinutes : num [1:940] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ...  
## $ TotalExcerciseHours : num [1:940] NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ...  
## $ Calories : num [1:940] 1985 1797 1776 1745 1863 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. ActivityDate = col\_character(),  
## .. TotalSteps = col\_double(),  
## .. TotalDistance = col\_double(),  
## .. TrackerDistance = col\_double(),  
## .. LoggedActivitiesDistance = col\_double(),  
## .. VeryActiveDistance = col\_double(),  
## .. ModeratelyActiveDistance = col\_double(),  
## .. LightActiveDistance = col\_double(),  
## .. SedentaryActiveDistance = col\_double(),  
## .. VeryActiveMinutes = col\_double(),  
## .. FairlyActiveMinutes = col\_double(),  
## .. LightlyActiveMinutes = col\_double(),  
## .. SedentaryMinutes = col\_double(),  
## .. Calories = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

The following data manipulation is performed:

1. For further analysis, populate column DayOfTheWeek by separating the date into days of the week.
2. Populate column TotalExcerciseMinutes by summing up VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes and SedentaryMinutes.
3. Populate column TotalHours by dividing the TotalExcerciseMinutes by 60.

# Populate the values for the DayOfWeek column  
copy\_daily$DayOfWeek = weekdays(as.Date(dailyActivity\_merged$ActivityDate, "%m/%d/%Y"))  
  
  
# Populate the TotalExcerciseMinutes column by adding up all of the minutes of the corresponding day  
copy\_daily$TotalExcerciseMinutes = copy\_daily$VeryActiveMinutes + copy\_daily$FairlyActiveMinutes + copy\_daily$LightlyActiveMinutes + copy\_daily$SedentaryMinutes  
  
  
# Populate the TotalExcerciseHours column by dividing the TotalExcerciseMinutes by 60  
copy\_daily$TotalExcerciseHours = round(copy\_daily$TotalExcerciseMinutes/60, digits=0)  
  
# verify that the columns have been populated successfully  
glimpse(copy\_daily)

## Rows: 940  
## Columns: 18  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036…  
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/…  
## $ DayOfWeek <chr> "Tuesday", "Wednesday", "Thursday", "Friday",…  
## $ 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…  
## $ TotalExcerciseMinutes <dbl> 1094, 1033, 1440, 998, 1040, 761, 1440, 1120,…  
## $ TotalExcerciseHours <dbl> 18, 17, 24, 17, 17, 13, 24, 19, 18, 18, 24, 1…  
## $ Calories <dbl> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203…

The data cleaning and manipulation process has been completed. As a result, the data is now ready for analysis.

## STEP 4: ANALYZE

#### 4.1 Perform calculations

Pulling the statistics of copy\_daily for analysis:

1. count - number of rows
2. mean (average)
3. std (standard deviation)
4. min
5. max

# getting the general statistics of the table  
format(describeBy(copy\_daily), scientific=FALSE)

## vars n mean sd  
## Id 1 940 4855407369.332978249 2424805475.657959938  
## ActivityDate\* 2 940 15.410638298 8.913708533  
## DayOfWeek\* 3 940 4.169148936 2.015938968  
## TotalSteps 4 940 7637.910638298 5087.150741753  
## TotalDistance 5 940 5.489702122 3.924605909  
## TrackerDistance 6 940 5.475351058 3.907275943  
## LoggedActivitiesDistance 7 940 0.108170940 0.619896518  
## VeryActiveDistance 8 940 1.502680851 2.658941165  
## ModeratelyActiveDistance 9 940 0.567542551 0.883580319  
## LightActiveDistance 10 940 3.340819149 2.040655388  
## SedentaryActiveDistance 11 940 0.001606383 0.007346176  
## VeryActiveMinutes 12 940 21.164893617 32.844803057  
## FairlyActiveMinutes 13 940 13.564893617 19.987403954  
## LightlyActiveMinutes 14 940 192.812765957 109.174699751  
## SedentaryMinutes 15 940 991.210638298 301.267436790  
## TotalExcerciseMinutes 16 940 1218.753191489 265.931767055  
## TotalExcerciseHours 17 940 20.313829787 4.437282772  
## Calories 18 940 2303.609574468 718.166862134  
## median trimmed mad  
## Id 4445114986.000 4777480921.4188833 3586057951.152600  
## ActivityDate\* 15.000 15.3018617 11.860800  
## DayOfWeek\* 4.000 4.2114362 2.965200  
## TotalSteps 7405.500 7388.2579787 5109.780900  
## TotalDistance 5.245 5.2027394 3.795456  
## TrackerDistance 5.245 5.1948404 3.788043  
## LoggedActivitiesDistance 0.000 0.0000000 0.000000  
## VeryActiveDistance 0.210 0.8845878 0.311346  
## ModeratelyActiveDistance 0.240 0.3758910 0.355824  
## LightActiveDistance 3.365 3.3155186 2.105292  
## SedentaryActiveDistance 0.000 0.0000000 0.000000  
## VeryActiveMinutes 4.000 14.0039894 5.930400  
## FairlyActiveMinutes 6.000 9.3603723 8.895600  
## LightlyActiveMinutes 199.000 193.5492021 102.299400  
## SedentaryMinutes 1057.500 998.6289894 386.958600  
## TotalExcerciseMinutes 1440.000 1251.0398936 0.000000  
## TotalExcerciseHours 24.000 20.8444149 0.000000  
## Calories 2134.000 2261.1023936 692.374200  
## min max range  
## Id 1503960366 8877689391.000000 7373729025.000000  
## ActivityDate\* 1 31.000000 30.000000  
## DayOfWeek\* 1 7.000000 6.000000  
## TotalSteps 0 36019.000000 36019.000000  
## TotalDistance 0 28.030001 28.030001  
## TrackerDistance 0 28.030001 28.030001  
## LoggedActivitiesDistance 0 4.942142 4.942142  
## VeryActiveDistance 0 21.920000 21.920000  
## ModeratelyActiveDistance 0 6.480000 6.480000  
## LightActiveDistance 0 10.710000 10.710000  
## SedentaryActiveDistance 0 0.110000 0.110000  
## VeryActiveMinutes 0 210.000000 210.000000  
## FairlyActiveMinutes 0 143.000000 143.000000  
## LightlyActiveMinutes 0 518.000000 518.000000  
## SedentaryMinutes 0 1440.000000 1440.000000  
## TotalExcerciseMinutes 2 1440.000000 1438.000000  
## TotalExcerciseHours 0 24.000000 24.000000  
## Calories 0 4900.000000 4900.000000  
## skew kurtosis se  
## Id 0.17656001 -1.2763188 79088434.1732696146  
## ActivityDate\* 0.09398421 -1.1796855 0.2907331155  
## DayOfWeek\* -0.13166986 -1.2602995 0.0657526791  
## TotalSteps 0.65081271 1.1476866 165.9245620350  
## TotalDistance 1.12268169 3.0771574 0.1280065305  
## TrackerDistance 1.13093128 3.1673533 0.1274412893  
## LoggedActivitiesDistance 6.27735641 40.9765157 0.0202187951  
## VeryActiveDistance 2.98661453 11.8097523 0.0867250983  
## ModeratelyActiveDistance 2.76235564 10.0376887 0.0288192123  
## LightActiveDistance 0.18166624 -0.1917035 0.0665588399  
## SedentaryActiveDistance 8.56250402 98.3785375 0.0002396059  
## VeryActiveMinutes 2.16920300 5.7224172 1.0712793544  
## FairlyActiveMinutes 2.47158430 7.9236091 0.6519172353  
## LightlyActiveMinutes -0.03780838 -0.3701857 3.5608860758  
## SedentaryMinutes -0.29355887 -0.6737466 9.8262603258  
## TotalExcerciseMinutes -1.01550318 1.1107889 8.6737378583  
## TotalExcerciseHours -1.01908037 1.1328578 0.1447282060  
## Calories 0.42110319 0.6076430 23.4240202653

**Interpreting statistical findings:**

Users logged 7,638 steps on average, which is insufficient. According to the CDC, an adult should aim for at least 10,000 steps per day to benefit from general health, weight loss, and fitness improvement.

Sedentary users account for the vast majority, logging an average of 991 minutes or 20 hours per day, accounting for 81% of overall average minutes.

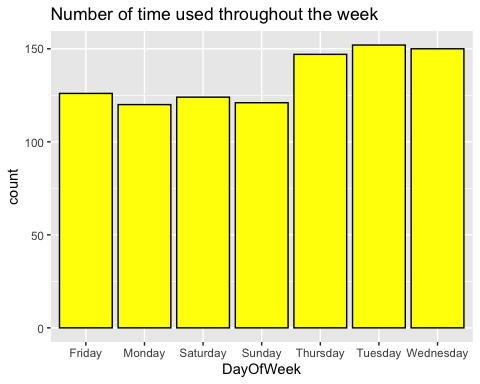
It is worth noting that the average number of calories burned is 2,303. Could not go into detail because calories burned are affected by a variety of factors like age, mass, everyday chores, exercise, hormones, and everyday calorie consumption.

## STEP 5: SHARE

We are producing visuals and presenting our conclusions based on our analysis in this step.

#### 5.1 Data Visualisation and Findings

# displaying the bar graph to see how often the users use the app throughout the week  
ggplot(data=copy\_daily, mapping= aes(x=DayOfWeek)) +  
 geom\_bar(color="black", fill="yellow" ) +   
 labs(title = "Number of time used throughout the week")

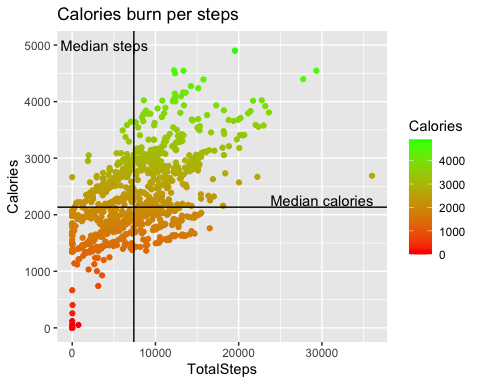


**Frequency of usage across the week**

We’re looking at the regularity of FitBit app usage in terms of days of the week in this graph.

1. We noticed that users prefer or remember (providing them the impression of benefit that they have forgotten) to record their activities on the app throughout the weekdays of Tuesday through Friday.
2. It is worth noting that the frequency reduced on Friday and will continue over the weekend and Monday.

# extract the median calories and steps from above table  
med\_calories <- 2134  
med\_steps <- 7405  
  
# scatter plot to graph Calories vs TotalSteps  
# also plot the median calories and median steps  
ggplot(data=copy\_daily,mapping=aes(x=TotalSteps, y=Calories,color=Calories)) + geom\_point() + scale\_colour\_gradient(low="red", high="green") +  
 geom\_hline(yintercept = med\_calories)+   
 geom\_vline(xintercept = med\_steps) +  
 labs(title = "Calories burn per steps")+  
 annotate("text",x=3850, y=5000,label="Median steps")+  
 annotate("text",x=30000, y=2250,label="Median calories")



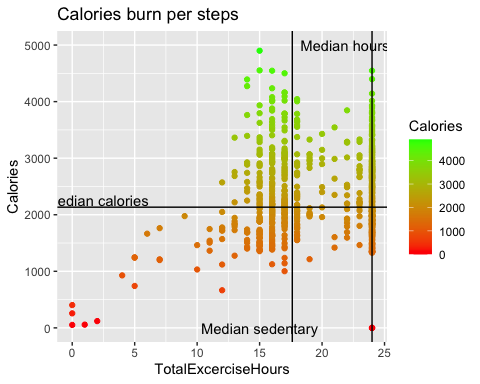
**Every step you take burns calories.**

We discovered from the scatter plot:

1. It is a positive relationship.
2. We discovered that the intensity of calories burned increases when users are between 0 and 14,000 steps, with the rate of calories burned decreasing after 14,000 steps.
3. There were a few outliers:

* No steps taken, and no to little calories were burned.
* One observation of more than 35,000 steps with less than 3,000 calories burnt.
* Outliers could happened due to the result of natural variance in data, changes in user activity, or errors in data collecting (ie. miscalculations, data contamination or human error).

# extract the median hours and sedentary hours from above table  
med\_hours <- 24  
med\_sedentary <- 1057/60  
  
# scatter plot to graph Calories vs TotalExcerciseHours  
# also plot the median calories, median total hours, and median sedentary hour  
ggplot(data=copy\_daily,mapping=aes(x=TotalExcerciseHours, y=Calories,color=Calories)) +  
 geom\_point() + scale\_colour\_gradient(low="red", high="green") +  
 geom\_hline(yintercept = med\_calories)+   
 geom\_vline(xintercept = med\_hours) +  
 geom\_vline(xintercept = med\_sedentary) +  
 labs(title = "Calories burn per steps") +  
 annotate("text",x=2, y=2250,label="Median calories")+  
 annotate("text",x=15, y=0,label="Median sedentary")+  
 annotate("text",x=22, y=5000,label="Median hours ")



**Calories burned for each tracked hour**

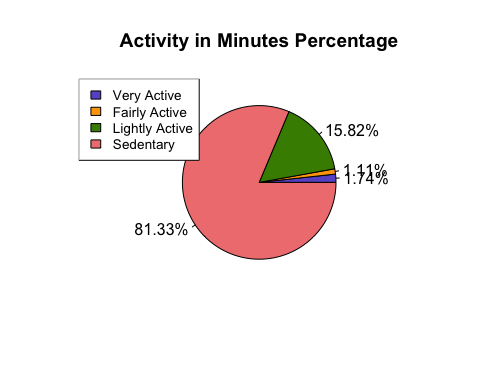
The scatter plot reveals:

A weak positive association in which the number of hours logged does not correspond to the number of calories burned. This is mostly attributable to the average sedentary hours plotted between 16 and 17 hours.

We can detect a few outliers once more:

* Outliers with the same zero value
* A strange red dot at 24 hours with no calories burned, which could be due to the same causes as stated above.

# calculating total of individual minutes column  
very\_active\_mins = sum(copy\_daily$VeryActiveMinutes)  
fairly\_active\_mins = sum(copy\_daily$FairlyActiveMinutes)  
lightly\_active\_mins = sum(copy\_daily$LightlyActiveMinutes)  
sedentary\_mins = sum(copy\_daily$SedentaryMinutes)  
  
# pie chart to display the minutes columns  
value=c(very\_active\_mins,fairly\_active\_mins,lightly\_active\_mins ,sedentary\_mins)  
pie\_labels <- paste0(round(100 \* value/sum(value), 2), "%")  
pie(value, labels = pie\_labels,col=c("slateblue", "orange", "chartreuse4", "lightcoral"),main = "Activity in Minutes Percentage")  
  
# legend of the pie chart   
legend("topleft", inset = -.0001, legend = c("Very Active","Fairly Active","Lightly Active" ,"Sedentary"), fill = c("slateblue", "orange", "chartreuse4", "lightcoral"),cex = 0.85)



**Minutes of Activity as a Percentage**

As evidenced by the pie chart:

1. Sedentary minutes account for the largest chunk, accounting for 81.3 percent.
2. This means that users are logging everyday activities such as daily commutes, idle movements (going from one location to another), or doing errands using the FitBit app.
3. According to the modest percentage of fairly active activity (1.1 percent) and very active activity (1.7 percent), the app is rarely used to track fitness. This is quite discouraging because the FitBit app was designed to promote fitness.

## STEP 6: ACT

In the final stage, we will present our findings and make suggestions based on our study.

In this section, we review our business questions and discuss our high-level business recommendations.

**1. What are the discovered trends?**

* The FitBit app is used by the majority of users (81.3 percent) to record sedentary behaviors rather than health practices.
* Users prefer to record their activities throughout the week as opposed to the weekends, possibly because they spend more time outside during the week and stay indoors on weekends.

**2. What implications do these patterns have for Bellabeat customers?**

* Both companies create products aimed at giving women with information on their health, behaviors, and fitness, as well as helping them to understand their present styles and make good choices. These prevalent health and fitness trends can very well be applied to Bellabeat clients.

**3. How can these tendencies affect Bellabeat’s marketing strategy?**

* The Bellabeat marketing team can encourage customers by teaching and preparing them with knowledge about fitness benefits, recommending different forms of exercise, and providing calories consumption and burn summary data on the Bellabeat app.
* On weekends, the Bellabeat app can send notifications to motivate users to workout.