Google Analytics Capstone: Bellabeats

Sara

2022-08-22

### The following information is from Google Analytics Certificate Program Case Study 2

## Scenario

I am junior data analyst working to a team of marketing analysts at Bellabeat. The stakeholders would like an analysis of data from fitness trackers to find insights about how people are using their fitness trackers. They will use this information to guide a marketing strategy for Bellabeat.

## Company

“Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits.”

## **Products**

* **Bellabeat app:** The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, mindfulness and wellness habits.
* **Leaf:** This wellness tracker can be worn as a bracelet, necklace, or clip; which then connects to the Bellabeat app to track activity, sleep, and stress.
* **Time:** This wellness watch uses smart technology to track user activity, sleep, and stress; and connects to the Bellabeat app to provide you with insights into your daily wellness.
* **Spring:** This is a water bottle that tracks daily water intake. It connects to the Bellabeat app to track your hydration levels.

# **Buisness task**

* **Establish** new growth for the company **by** identifying trends in how fitness trackers data **to** create a marketing strategy.

# Stakeholders

* **Urška Sršen:** Bellabeat’s cofounder and Chief Creative Officer
* **Sando Mur:** Mathematician and Bellabeat’s cofounder

# Data

“This dataset generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.” As per link: <https://www.kaggle.com/datasets/arashnic/fitbit> if you would like more information please refer to link. I did try and find other datasets to make my analysis more accurate. I was unable to find any more fitness tracker data. Although not current this data is reliable. It is a Public dataset so there is any privacy issues if there was I would change/hide or lock user information.

# **Setting up R and loading Data set**

Notes: Installed packages needed in R for analysis

install.packages(c(“rmarkdown”,“skimr”,“tidyverse”,“janitor”,“lubridate”))

library(rmarkdown)  
library(skimr)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.1  
## ✔ readr 2.1.2 ✔ forcats 0.5.2  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(janitor)

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

library(lubridate)

##   
## Attaching package: 'lubridate'  
##   
## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

## **Import Data**

daily\_activity <- read\_csv("dailyActivity\_merged.csv")

## 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.

hr\_seconds <- read\_csv("heartrate\_seconds\_merged.csv")

## Rows: 2483658 Columns: 3  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): Time  
## dbl (2): Id, Value  
##   
## 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.

sleep\_day <- read\_csv("sleepDay\_merged.csv")

## 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.

## **Review Data**

str(daily\_activity)

## 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>

str(hr\_seconds)

## spec\_tbl\_df [2,483,658 × 3] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:2483658] 2.02e+09 2.02e+09 2.02e+09 2.02e+09 2.02e+09 ...  
## $ Time : chr [1:2483658] "4/12/2016 7:21:00 AM" "4/12/2016 7:21:05 AM" "4/12/2016 7:21:10 AM" "4/12/2016 7:21:20 AM" ...  
## $ Value: num [1:2483658] 97 102 105 103 101 95 91 93 94 93 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. Time = col\_character(),  
## .. Value = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(sleep\_day)

## spec\_tbl\_df [413 × 5] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ Id : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...  
## $ SleepDay : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...  
## $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...  
## $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...  
## $ TotalTimeInBed : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. Id = col\_double(),  
## .. SleepDay = col\_character(),  
## .. TotalSleepRecords = col\_double(),  
## .. TotalMinutesAsleep = col\_double(),  
## .. TotalTimeInBed = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

## **Clean Data And Validate**

Ensure column names are clean.

clean\_names(daily\_activity)

## # A tibble: 940 × 15  
## id activity…¹ total…² total…³ track…⁴ logge…⁵ very\_…⁶ moder…⁷ light…⁸  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1503960366 4/12/2016 13162 8.5 8.5 0 1.88 0.550 6.06  
## 2 1503960366 4/13/2016 10735 6.97 6.97 0 1.57 0.690 4.71  
## 3 1503960366 4/14/2016 10460 6.74 6.74 0 2.44 0.400 3.91  
## 4 1503960366 4/15/2016 9762 6.28 6.28 0 2.14 1.26 2.83  
## 5 1503960366 4/16/2016 12669 8.16 8.16 0 2.71 0.410 5.04  
## 6 1503960366 4/17/2016 9705 6.48 6.48 0 3.19 0.780 2.51  
## 7 1503960366 4/18/2016 13019 8.59 8.59 0 3.25 0.640 4.71  
## 8 1503960366 4/19/2016 15506 9.88 9.88 0 3.53 1.32 5.03  
## 9 1503960366 4/20/2016 10544 6.68 6.68 0 1.96 0.480 4.24  
## 10 1503960366 4/21/2016 9819 6.34 6.34 0 1.34 0.350 4.65  
## # … with 930 more rows, 6 more variables: sedentary\_active\_distance <dbl>,  
## # very\_active\_minutes <dbl>, fairly\_active\_minutes <dbl>,  
## # lightly\_active\_minutes <dbl>, sedentary\_minutes <dbl>, calories <dbl>, and  
## # abbreviated variable names ¹​activity\_date, ²​total\_steps, ³​total\_distance,  
## # ⁴​tracker\_distance, ⁵​logged\_activities\_distance, ⁶​very\_active\_distance,  
## # ⁷​moderately\_active\_distance, ⁸​light\_active\_distance

clean\_names(hr\_seconds)

## # A tibble: 2,483,658 × 3  
## id time value  
## <dbl> <chr> <dbl>  
## 1 2022484408 4/12/2016 7:21:00 AM 97  
## 2 2022484408 4/12/2016 7:21:05 AM 102  
## 3 2022484408 4/12/2016 7:21:10 AM 105  
## 4 2022484408 4/12/2016 7:21:20 AM 103  
## 5 2022484408 4/12/2016 7:21:25 AM 101  
## 6 2022484408 4/12/2016 7:22:05 AM 95  
## 7 2022484408 4/12/2016 7:22:10 AM 91  
## 8 2022484408 4/12/2016 7:22:15 AM 93  
## 9 2022484408 4/12/2016 7:22:20 AM 94  
## 10 2022484408 4/12/2016 7:22:25 AM 93  
## # … with 2,483,648 more rows

clean\_names(sleep\_day)

## # A tibble: 413 × 5  
## id sleep\_day total\_sleep\_records total\_minutes\_…¹ total…²  
## <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 1503960366 4/12/2016 12:00:00 AM 1 327 346  
## 2 1503960366 4/13/2016 12:00:00 AM 2 384 407  
## 3 1503960366 4/15/2016 12:00:00 AM 1 412 442  
## 4 1503960366 4/16/2016 12:00:00 AM 2 340 367  
## 5 1503960366 4/17/2016 12:00:00 AM 1 700 712  
## 6 1503960366 4/19/2016 12:00:00 AM 1 304 320  
## 7 1503960366 4/20/2016 12:00:00 AM 1 360 377  
## 8 1503960366 4/21/2016 12:00:00 AM 1 325 364  
## 9 1503960366 4/23/2016 12:00:00 AM 1 361 384  
## 10 1503960366 4/24/2016 12:00:00 AM 1 430 449  
## # … with 403 more rows, and abbreviated variable names ¹​total\_minutes\_asleep,  
## # ²​total\_time\_in\_bed

### **Find Missing Values**

is.null(daily\_activity)

## [1] FALSE

is.null(hr\_seconds)

## [1] FALSE

is.null(sleep\_day)

## [1] FALSE

### **Find Duplicate Data**

sum(duplicated(daily\_activity))

## [1] 0

sum(duplicated(sleep\_day))

## [1] 3

sum(duplicated(hr\_seconds))

## [1] 0

3 Duplicates found in sleep\_day data

### **Remove Duplicate Data**

sleep\_day<- unique(sleep\_day)

### **Recheck for Duplicates**

sum(duplicated(sleep\_day))

## [1] 0

### **Check Length(number of Users) for data**

length(unique(daily\_activity$Id))

## [1] 33

length(unique(hr\_seconds$Id))

## [1] 14

length(unique(sleep\_day$Id))

## [1] 24

## **Fix any inconsistencies, manipulate data to find relevant information and combine data**

### **Change dates to standard format**

daily\_activity$ActivityDate = as.POSIXct(daily\_activity$ActivityDate,tz=Sys.timezone(),format="%m/%d/%Y")  
  
hr\_seconds$Time = as.POSIXct(hr\_seconds$Time,tz=Sys.timezone(),format="%m/%d/%Y")  
  
sleep\_day$SleepDay = as.POSIXct(sleep\_day$SleepDay ,tx=Sys.timezone(), format = "%m/%d/%Y ")  
  
hr\_seconds$Time = format(hr\_seconds$Time,format='%Y/%m/%d')

### **Rename SleepDay to Activity Date and Combine**

sleep\_day$minNotAsleep = (sleep\_day$TotalTimeInBed - sleep\_day$TotalMinutesAsleep)  
  
sleep\_day$hoursSleep = (sleep\_day$TotalMinutesAsleep / 60)   
sleep\_day$hoursSleep = round(sleep\_day$hoursSleep,digits = 1)

### **Rename the Time column to ActiveDate to be consistent with other data**

sleep\_day <- rename(sleep\_day,ActivityDate=SleepDay)  
  
daily\_activity\_1 <- merge(daily\_activity,sleep\_day, by=c("Id","ActivityDate"))

### **Find the daily maximum,minimum and average heart rate and round average to the nearest whole number**

hr\_seconds\_sum <- hr\_seconds %>%  
 group\_by(Id,Time) %>%  
 summarize(MaxHeartRate = max(Value,na.rm=T),  
 MinHeartRate = min(Value,na.rm=T),  
 AvgHeartRate = mean(Value,na.rm=T)) %>%  
 arrange(Id,Time)

## `summarise()` has grouped output by 'Id'. You can override using the `.groups`  
## argument.

hr\_seconds\_sum$AvgHeartRate = round(hr\_seconds\_sum$AvgHeartRate,digits = 0)

### **Rename the Time column to ActiveDate to be consistent with other data**

hr\_seconds\_sum <- rename(hr\_seconds\_sum,ActivityDate=Time)  
  
daily\_activity\_all <- merge(daily\_activity\_1,hr\_seconds\_sum, by=c("Id","ActivityDate"))

### **Calculate toalta active minutes and combine data**

daily\_activity\_all <- daily\_activity\_all %>%  
 mutate(TotalActiveMinutes = rowSums(across(c(VeryActiveMinutes,FairlyActiveMinutes,LightlyActiveMinutes))))  
  
View(daily\_activity\_all)

Now all my data has been cleaned and double checked.

# **Analysis**

### **Review Summaries**

daily\_activity\_all %>%  
select(TotalSteps, TotalDistance,TotalActiveMinutes,hoursSleep,minNotAsleep,MinHeartRate,MaxHeartRate,AvgHeartRate,SedentaryMinutes) %>%  
 summary()

## TotalSteps TotalDistance TotalActiveMinutes hoursSleep   
## Min. : 42 Min. : 0.030 Min. : 4.0 Min. : 1.000   
## 1st Qu.: 5245 1st Qu.: 3.680 1st Qu.:217.0 1st Qu.: 6.500   
## Median : 9148 Median : 6.380 Median :283.0 Median : 7.300   
## Mean : 8522 Mean : 6.021 Mean :273.2 Mean : 7.106   
## 3rd Qu.:11135 3rd Qu.: 7.920 3rd Qu.:333.0 3rd Qu.: 8.200   
## Max. :22770 Max. :17.540 Max. :540.0 Max. :12.900   
## minNotAsleep MinHeartRate MaxHeartRate AvgHeartRate   
## Min. : 0.00 Min. :36.00 Min. : 80.0 Min. : 59.00   
## 1st Qu.:16.00 1st Qu.:45.00 1st Qu.:123.0 1st Qu.: 67.00   
## Median :24.00 Median :51.00 Median :133.0 Median : 73.00   
## Mean :28.55 Mean :49.97 Mean :135.5 Mean : 73.67   
## 3rd Qu.:39.00 3rd Qu.:54.00 3rd Qu.:149.0 3rd Qu.: 79.00   
## Max. :94.00 Max. :70.00 Max. :199.0 Max. :105.00   
## SedentaryMinutes  
## Min. : 2   
## 1st Qu.: 621   
## Median : 707   
## Mean : 702   
## 3rd Qu.: 781   
## Max. :1222

From this summary, consumers are meeting an average of 8522 steps a day. Havard Health suggests that there is benefit for meeting a 7500 step goal a day. (link)<https://www.health.harvard.edu/staying-healthy/how-many-steps-should-i-take-each-day>. The cdc recommends getting more the 7 hours or more of sleep a night the average showen is 7 (link)<https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html> There is relevant information about the highest daily heart rate but the the maxium is bordering out of normal range and the highest average heart rate is also above normal range.

## \*\* PLot Data

Review the correlation between daily steps taken and calories burned.

ggplot(daily\_activity\_all,aes(x=TotalSteps,y=Calories,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Step Totals and Calories Burned")

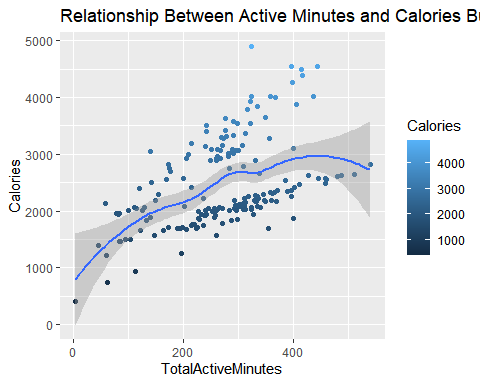
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



Correlation Identified, review if correlation is the same between with total activity minutes and calories burned.

ggplot(daily\_activity\_all,aes(x=TotalActiveMinutes,y=Calories,color=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Active Minutes and Calories Burned")

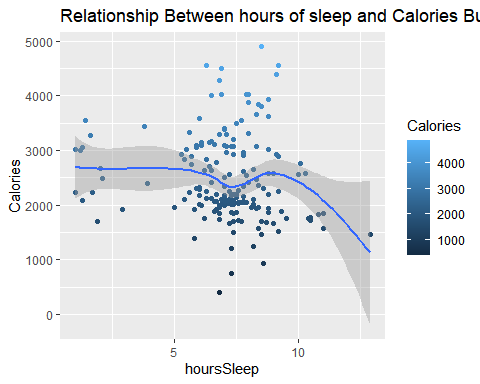
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



Same correlation identified. Analysis any correlation with Calories burned and hours of sleep.

ggplot(daily\_activity\_all,aes(x=hoursSleep,y=Calories,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between hours of sleep and Calories Burned")

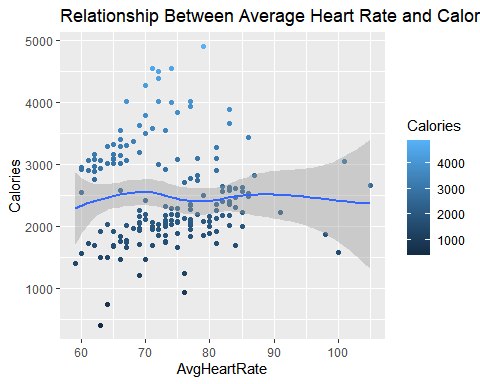
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



Identify if there is any correlation between average heart rate, lowest heart rate and highest heart rate; with calories burned

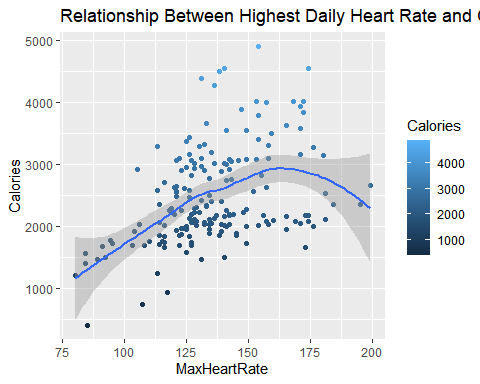
ggplot(daily\_activity\_all,aes(x=AvgHeartRate,y=Calories,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Average Heart Rate and Calories Burned")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



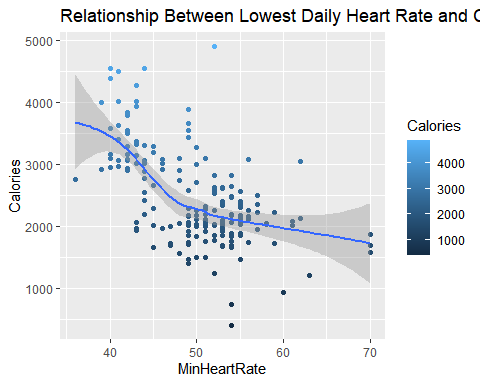
ggplot(daily\_activity\_all,aes(x=MaxHeartRate,y=Calories,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Highest Daily Heart Rate and Calories Burned")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



ggplot(daily\_activity\_all,aes(x=MinHeartRate,y=Calories,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Lowest Daily Heart Rate and Calories Burned")

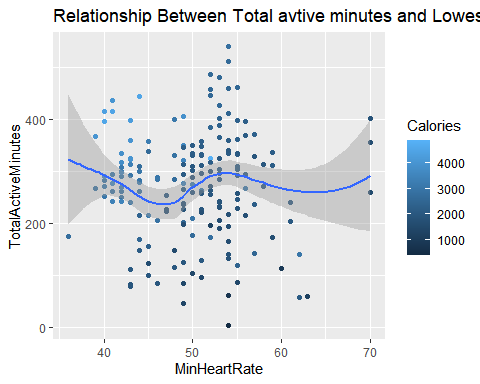
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

 From these graphs there is clearly a positive relationship between the amount of steps/active minutes and how many calories are burned. It appears as though sleep does not correlate with the amount of calories burned except in the case where hours asleep are greater than 10 then there seems to be less calories burned. It also appears as though there is a positive correlation for the amount of calories burned with highest daily heart rate until about 162 beats per minutes, after that there is a negative correlation. There seems to be a correlation between lowest daily heart rate and calories burned where the lower the lowest the daily heart rate is the more calories burnt.

Identify if there is a correlation between lowest daily heart rate and activity minutes

ggplot(daily\_activity\_all,aes(x=MinHeartRate,y=TotalActiveMinutes,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Total avtive minutes and Lowest Daily Heart Rate")

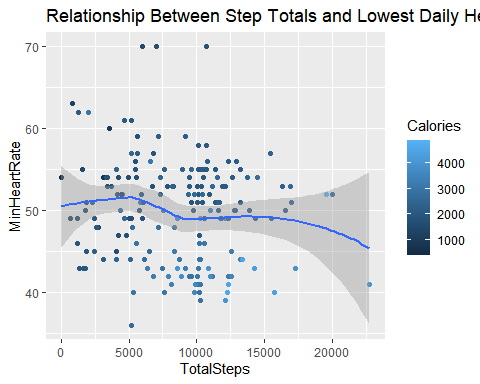
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

 Correlation not seen.

View relationship between Total steps and lowest daily heart rate

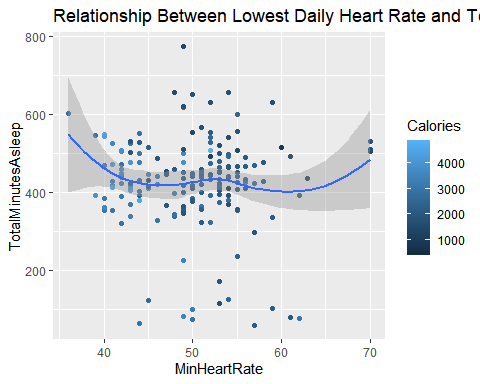
ggplot(daily\_activity\_all,aes(x=TotalSteps,y=MinHeartRate,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Step Totals and Lowest Daily HeartRrate")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

 Identify if there is a correlation between sleep and lowest average heart rate

ggplot(daily\_activity\_all,aes(x=MinHeartRate,y=TotalMinutesAsleep,colour=Calories))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(title = "Relationship Between Lowest Daily Heart Rate and Total Minutes Asleep")

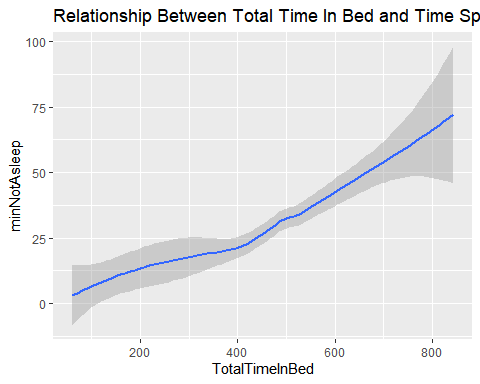
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

 No relationship noted.

Identify correlations between minutes in bed

ggplot(data=daily\_activity\_all) +  
 geom\_smooth(mapping=aes(x=TotalTimeInBed,y=minNotAsleep)) +  
 labs(title="Relationship Between Total Time In Bed and Time Spent Awake In Bed")

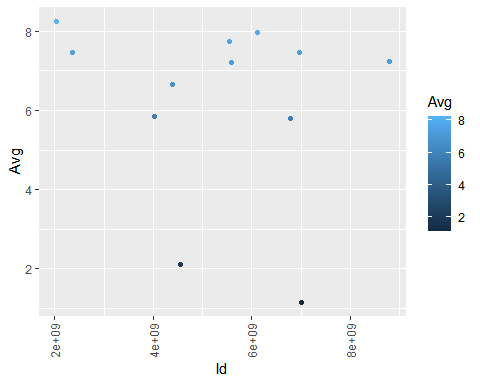
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



Correlation identified between minutes awake in bed and total time in bed.

Graph to determine the amount of hours of sleep individuals averaged.

daily\_activity\_all %>%  
 group\_by(Id) %>%  
 summarize(Avg = mean(hoursSleep,na.rm=T)) %>%  
 ggplot() +  
 geom\_point(aes(x=Id, y=Avg,color = Avg)) +  
 theme(axis.text.x=element\_text(angle = 90, vjust = 0.5))



Identified that many consumers are not getting the recommended amount of sleep on average.

# **Summary of Analysis**

* Data summary at the beginning of my analysis found some useful information. If this information was calculated on an individual basis it could really give consumers insightful information about their health and ways to improve their health. There was an obvious connection between total steps - and calories burned, as well as active minutes and calories burned. There was not a correlation between hours of sleep and calories burned except when total hours slept was over ten hours then there was a negative correlation.
* There was a positive correlation between highest daily heart rate and calories burned until about 162 beats per minute then there was a negative correlation.
* There was a negative correlation between lowest daily heart rate where the higher the lowest daily heart rate was the less calories were burned. There did not seem to be any correlation between total active minutes and lowest daily heart rate, lowest daily heart rate and minutes asleep or between total steps and lowest daily heart rate
* There was a positive correlation between minutes in bed awake and total minutes in bed. There, the longer people were awake in bed, the longer total time they spent in bed. the same correlation between minutes in bed awake and minutes in bed asleep. It seems the longer time spent in bed not asleep, the more sleep was needed.
* Although my summary earlier suggested consumers averaged about 7 hours of sleep, individually some consumers were averaging less than 7.

# **Suggestions Based On Analysis**

Suggested marketing campaign targeted towards using consumers individual data to help make suggestions to improve their health. Example: ”you burn more calories on average when you get this much sleep”, “If you get your heart rate up to 162 bpm for this long you will be in peak range for calorie burning, for health benefits you should get at least 7500 calories a day.” The message is that if people purchase and wear the Bellabeat activity tracker and wear them consistently Bella beat can use their data to help make suggestions to try and help improve your overall health.”

# **Problems with the dataset/analysis**

* The data set was not large enough to come to any valid conclusions about the population as a whole. The data did not include information about the individuals such as gender, height, ethnicity, ect..
* The data set did not include how often people were not wearing their activity tracker, likely daily missing data; such as when their devices are being charged.
* Tracker accuracy unknown, so there could be a discrepancy compared to other tracking methods such as: steps, calories burnt, sleep, ect. was being tracked accurately.