**Bellabeat Case Study:** How smart device fitness data could help unlock new growth and Opportunities.

**Bellabeat** case study is part of the capstone project of the **Google Data Analytics Professional Certificate course** which is completed in September, 2023

**Table of contents**

* Scenario
* Quick Overview
* Background
* Data Analysis Phases
* Ask Phase
* Prepare Phase
* Process Phase
* Analyze Phase
* Share Phase
* Act Phase
* Recommendations

**Scenario**

I am 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.

**Quick Overview**

This is what I put up in my overview

**Background:** An introduction of Bellabeat Company.  
**Data Analysis phase:** The data analysis phase are the Ask, Prepare, Process, Share, and Act.  
**Tools used:** RStudio Programming

Below is the breakdown of my project.

**Background**  
this capstone project is a Bellabeat data analysis study. Bellabeat is a high-tech manufacturer of health-focused products for women and meet different characteristics and team members.

Data Analysis Phases

For this project I followed the data analysis phase which is   
**Ask -> Prepare-> Process-> Analyze -> Share -> Act**

**Step 1: Ask**

In this step, we define the problem and objectives of our case study and its outcome.

**Business Task:**

I have been asked to focus on one of Bellabeat’s products and analyze smart device usage data to gain insight into how consumers are using their smart devices. The insights i discover will then help guide marketing strategy for the company.

**Business Objectives:**

* What are the trends identified?
* How could these trends apply to Bellabeat customers?
* How could these trends help influence Bellabeat marketing strategy?

**Deliverables:**

* A clear summary of the business task
* A description of all data sources used
* Documentation of any cleaning or manipulation of data
* A summary of analysis
* Supporting visualizations and key findings
* High – level content recommendations based on the analysis

**Key stakeholders:**

* **Urška Sršen:** Bellabeat cofounder and chief creative officer.
* **Sando Mur:** Mathematician and Bellabeat’s cofounder key members 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.

**STEP 2: Prepare**

In the Prepare phase, we identify the data being used and its limitations.

**Information on Data Source:**

* Data is publicly available on [Kaggle: FitBit Fitness Tracker Data](https://www.kaggle.com/arashnic/fitbit) and stored in 18 CSV files.
* Generated by respondents from a survey via Amazon Mechanical Turk between 12th March 2016 to 12 May, 2016
* 30 FitBit users consented to the submission of personal tracker data.
* Data collected includes physical activity recorded in minutes, heart rate, sleep monitoring, daily activity and steps.

**Limitations of Data Set:**

* Data is collected 7 years ago in 2016, so users like daily activity, fitness and sleeping habits, diet and food consumption may have changed since then so the data may not be timely or relevant.
* Sample size of 30 FitBit users is not representative of the entire fitness population.
* The data is already collected in a survey so we cannot ascertain its integrity or accuracy.

**Is Data ROCCC?**

A good data source is ROCCC which stands for **R**eliable, **O**riginal, **C**omprehensive, **C**urrent and **C**ited.

* **R**eliable- Low; Not reliable because it contains only 30 respondents.
* **O**riginal – it’s a third party provider(Amazon Mechanical Turk) so it’s very low
* **C**omprehensive- Parameter as the same as most of the products of Bellabeat products so is medium.
* **C**urrent- Data is 7 years old and may not be relevant so its low
* **C**ited- Data is collected form third party hence its UNKNOWN SO ITS Low as well.

Overall, the dataset is considered bad quality data and it is not recommended to produce business recommendations based on this data.

**Data Selection**

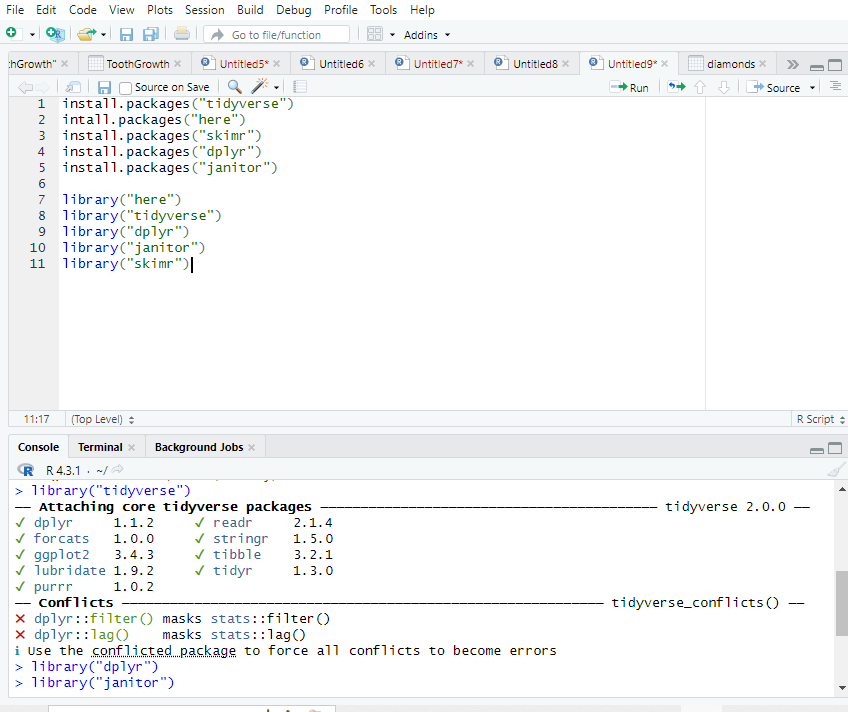
The following file is selected and copied for the analysis dailyActivity\_merged csv

**Tool**

Due to the small sample size data collection and the lack of essential participant characteristics such as gender, age, geography, and lifestyle, this data has limitations.

In this analysis, the datasets we will use is daily activity, daily calories, daily intensities, daily steps, and heart rate by seconds, minute METs, daily sleep, and weight log information. Because of the numerous packages and data visualization options available, R Studio was utilized to accomplish this analysis.

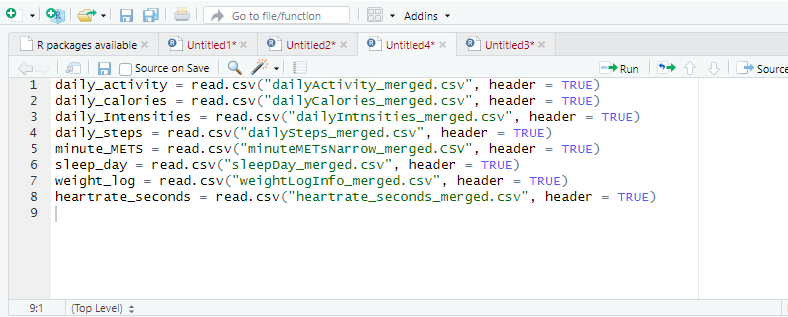
**Data loading**



**Importing data**

The datasets in the CSV file were opened using excel for proper formatting of the time/date where necessary. Such as, the time and/or date formatting were changed from “custom” to “time” and/or “short date” as needed. The files then were imported to R Studio and the data frames were created.

I used the environment pane to import all the csv files I will be using for this case study.

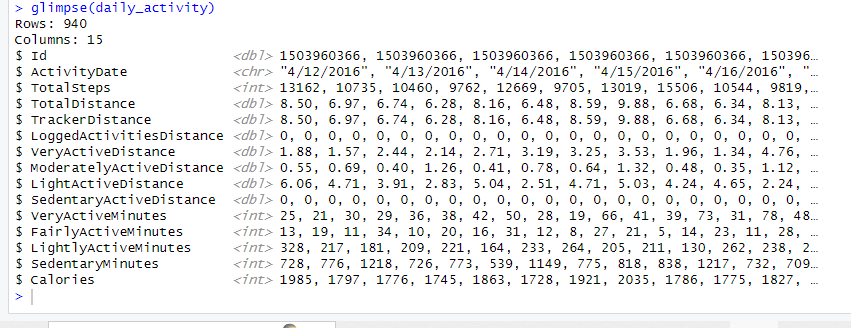


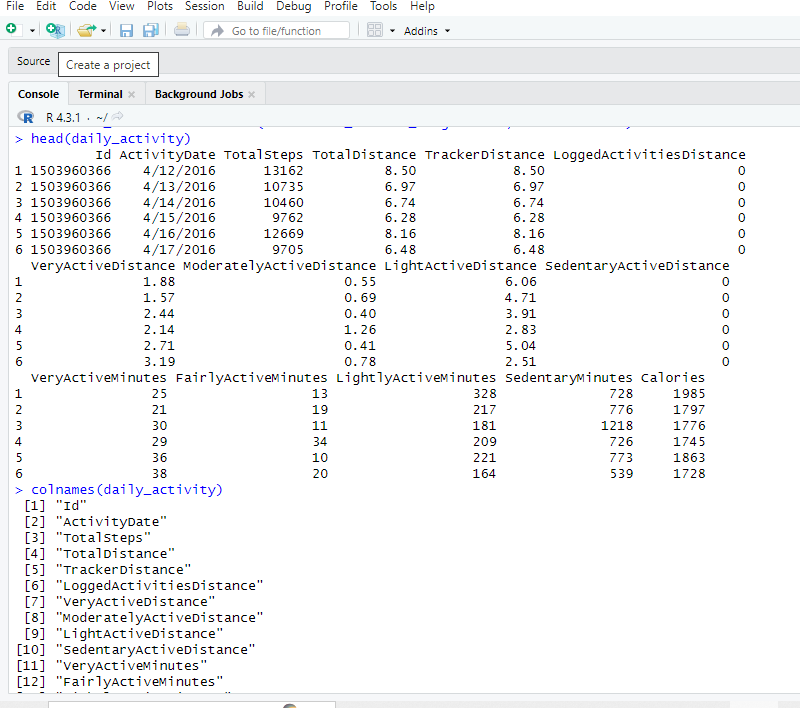
**Step 3: Process**

**Viewing data frames**

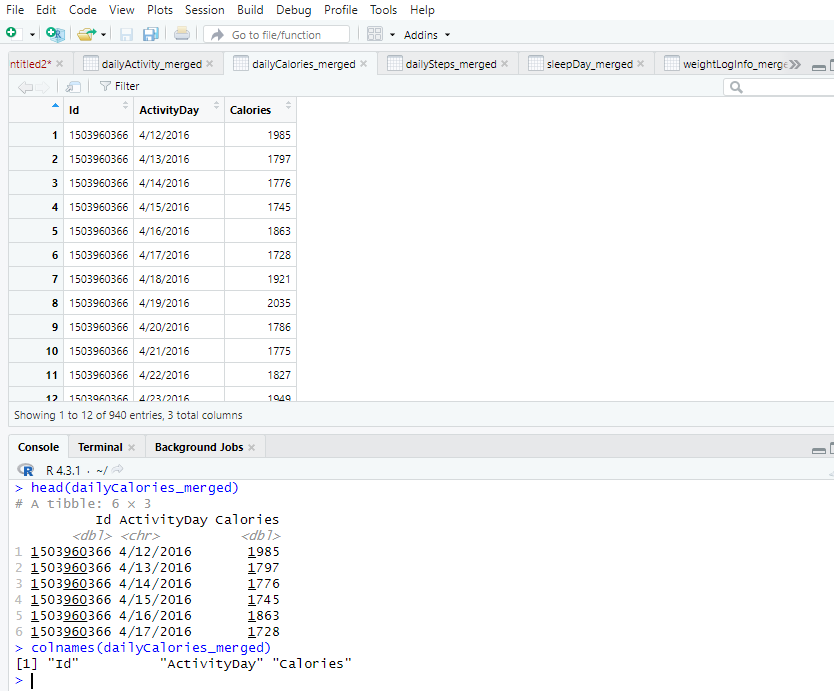
The head function is used to confirm that the data frames were imported appropriately. To investigate the data frames and uncover similarities, the colnames and glimpse functions were utilized.

**daily\_activity:**

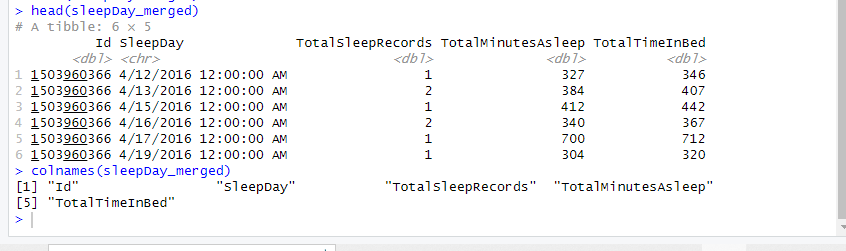




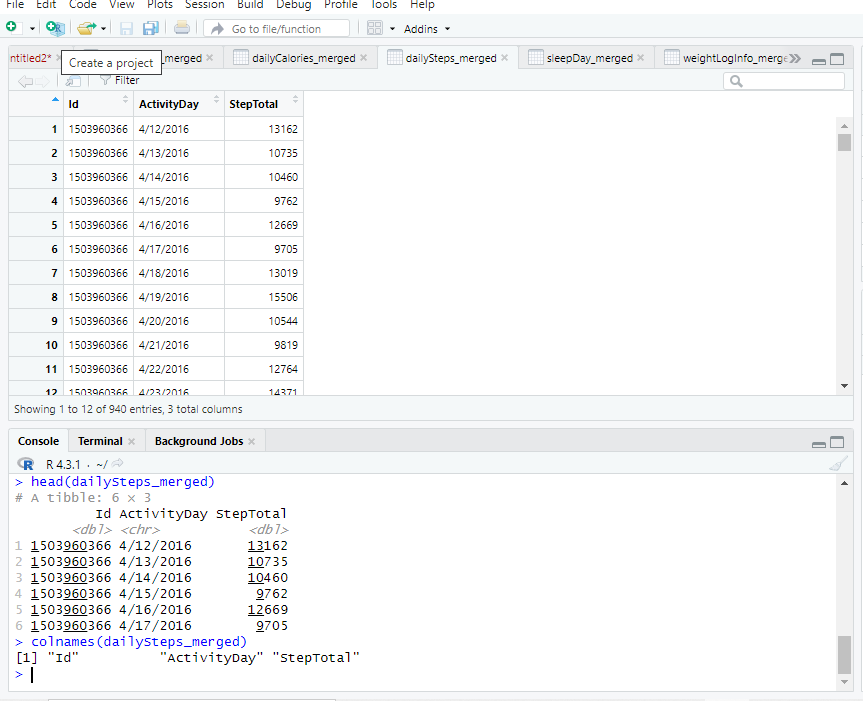
**Daily\_calories**



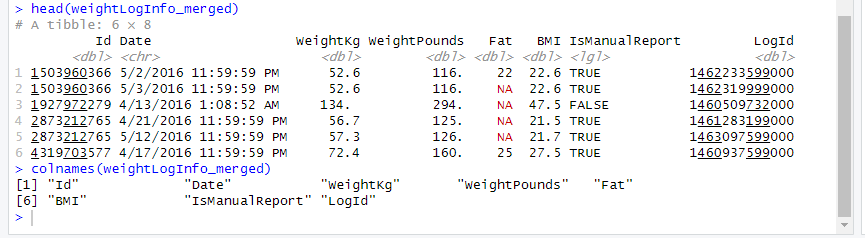
**Sleep\_day**



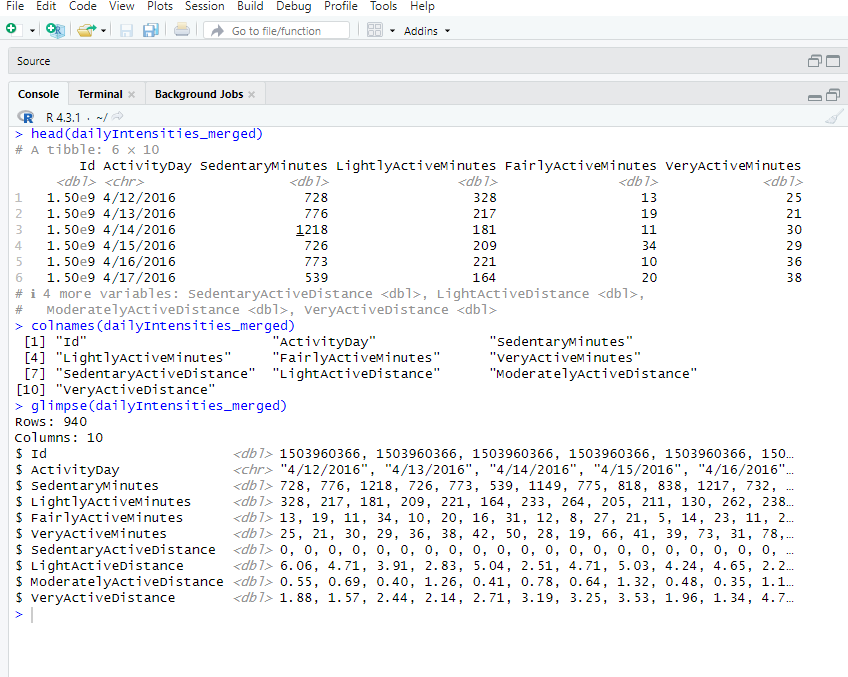
**daily\_steps**



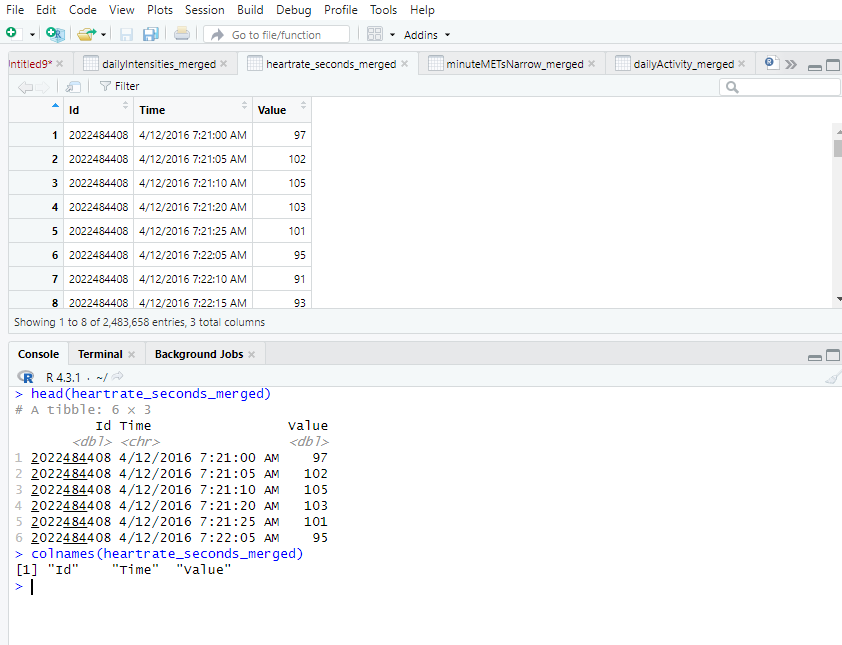
**Weight\_log**



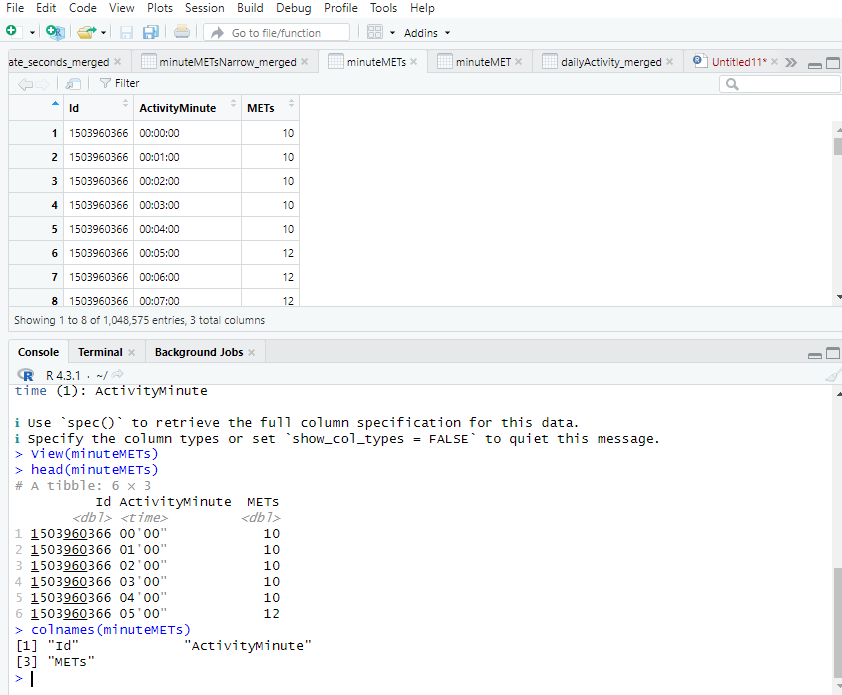
**daily­\_intensities**

****

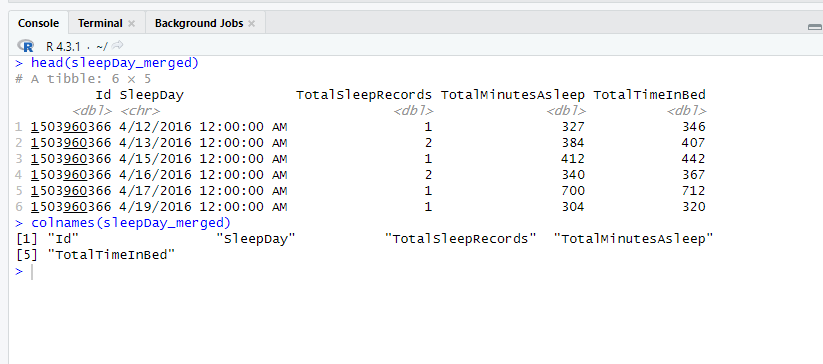
**heartrate\_seconds:**

****

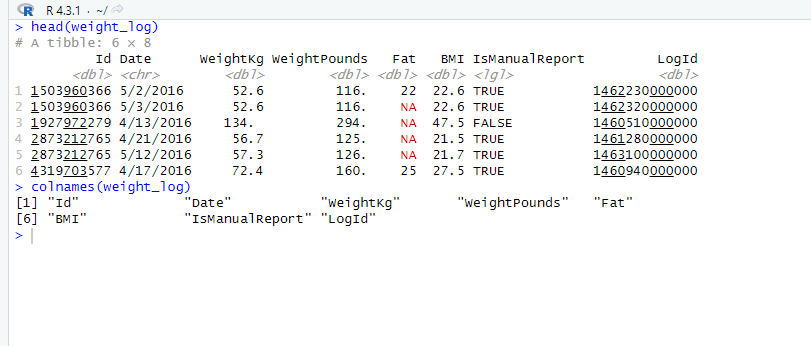
**minute\_METs:**

****

**sleep\_day:**

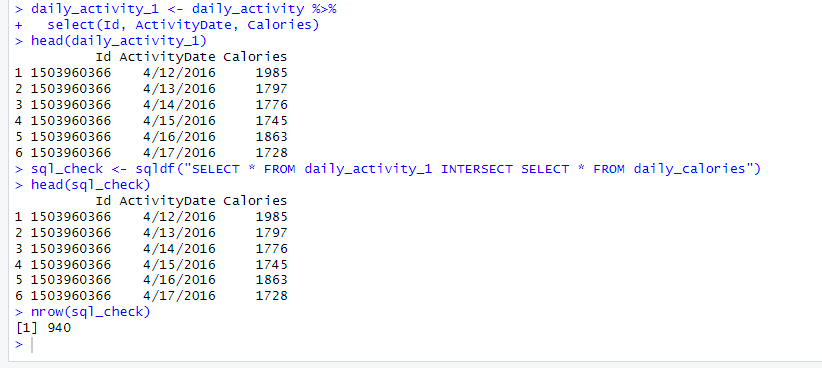
****

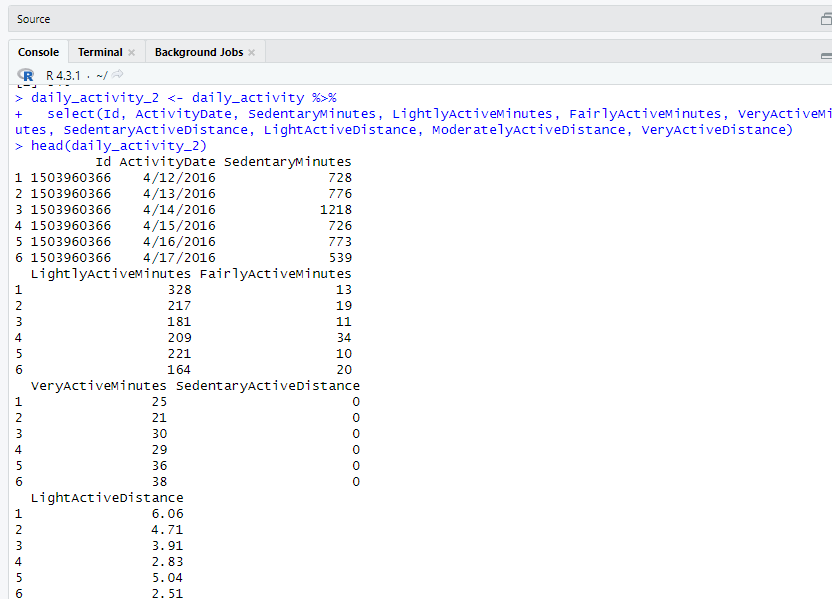
**Weight\_log**

****

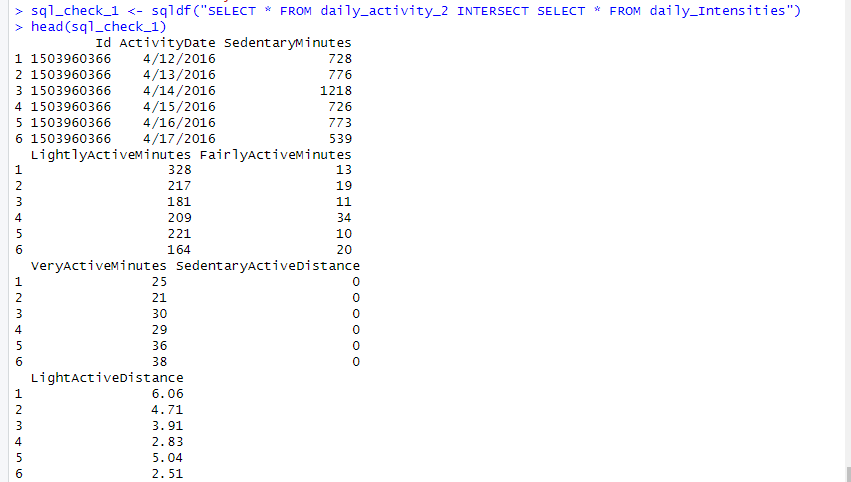
**Removing data frames**

There are eight data frames including id fields will all be merged. But calories, intensities and steps appear to be included in the daily activity data frame. The number of observations must be the same and the observations must match each ID number to utilize the daily activity frame instead of daily calories, daily intensities, and daily steps.The sqldf package, an R package for running SQL statements on R data frames, is loaded to determine if the values of daily\_calories, daily\_intensities, and daily\_steps are present in daily activity using SQL syntax. However, because the number of columns in both data frames must be the same, a temporary data frame containing the critical columns is needed to be created first.

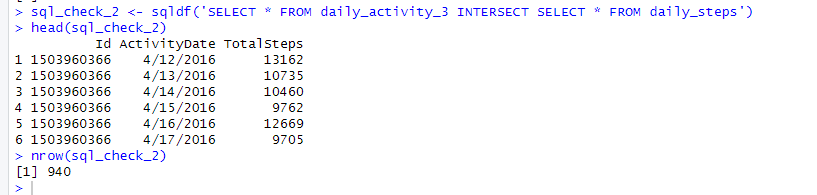


****

****

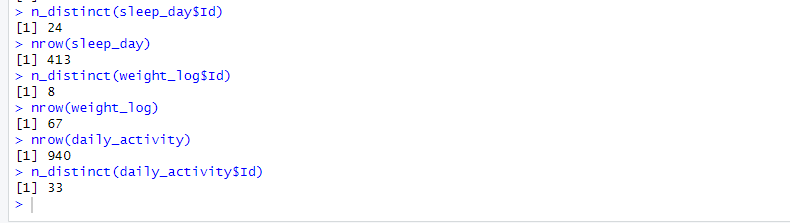
****

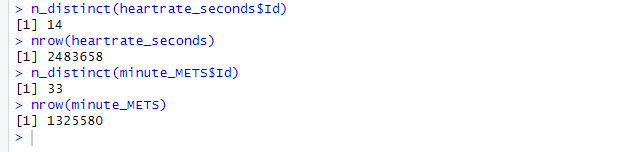
****

****The result of the head function of the temporary data frames matches that of the results of the original data frames. And the results of the SQL date frames also match that of the temporary data frames. The number of the observations of each SQL data frame is 940. In conclusion the datasets of daily\_calories, daily\_intensities and that of daily\_steps are all in daily\_activity. These three data frames will have to be removed from the analysis for simplicity.

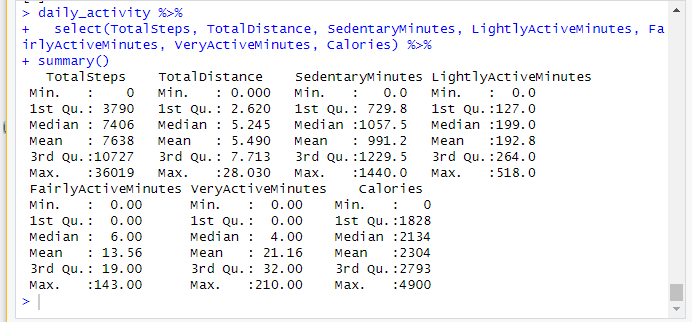
**Step4: Analyze**

Summarizing the data:   
the n\_distict() and nrow() functions are used to determine the number of unique vales and the number of rows in a data frame respectively.



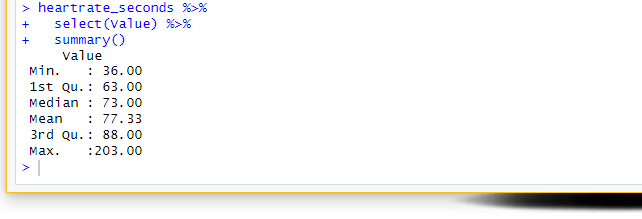
The weight log data frames contains a very low number of participants based on the n\_distint() function which makes it difficult for reliable recommendations and conclusions to be done. The summary() function is used to pull key statistics about the data frames.

**daily\_activity**



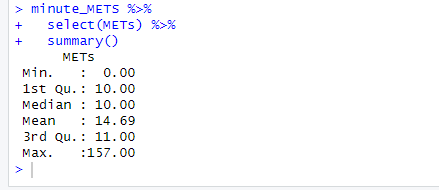
This summary shows the average user is taking 7638 steps a day, missing the recommended 10,000 steps for health by the **Centers for Disease Control and** **Prevention**. On average, users are getting 21.16 minutes of very active activity a day, this equates to 148.12 minutes a week. The **Centers for Disease Control and Prevention** recommends 75 minutes of vigorous activity a week, so in this case it’s either the Fitbit user is doing too much or is achieving extra health benefits. Also the participants are having an average of 991.2 minutes of sedentary time a day this is too much and lead to health issues because the body functions best upright. Scientist have proven that 40 minutes of moderate activity a day will balance out the effects of sitting up to 10 hours a day.in conclusion the Fitbit users are burning more calories than the usual.

**heartrate\_seconds:**



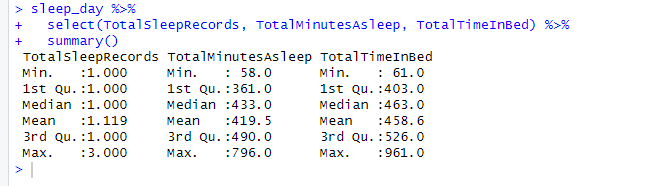
Despite the number of users in the heart rate data frame, the average heartrate is 77.33 beats per minute (bpm) which is considered as a normal beat for most people and not the entire population. This is because resting heart rates between different people can vary by as much as 70 bpm. However changes in resting heartrate over days can be a sign of infection, menstrual cycle effects, or other acute triggers. This calls for regular monitoring.

**minute\_METs:**



The summary of minute METs shows the average user has a MET of 14.69. A MET is the ratio of your working metabolic rate relative to your resting metabolic rate. One MET is the energy you use when at rest. This means an activity with a MET of four would require a person to exert four times the energy they do when they are sitting. Therefore, a user averaging 14.69 MET throughout the day is considerably high, which leads to the assumption that the Fitbit is not calculating this data point correctly. Due to this, the minute MET data frame will not be used further in this analysis.

**sleep\_day:**

******

The summary of the sleep data frame displays the average user sleeps once per day for 419.5 minutes, or roughly 7 hours. This falls within the **Centers for Disease Control and** **Prevention** recommendations for adults in order to get the proper amount of rest. The average participant is spending 458.6 minutes or 7.64 hours in bed .This means the typical user is spending 38.6 minutes awake in bed. According to **Health Central**, people should not spend more than 1 hour in bed awake. This is to prevent a mental link being formed between being awake and being in bed which can lead to insomnia.

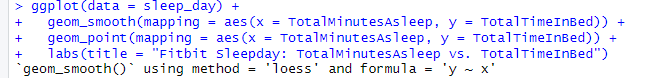
**Weight\_log:**

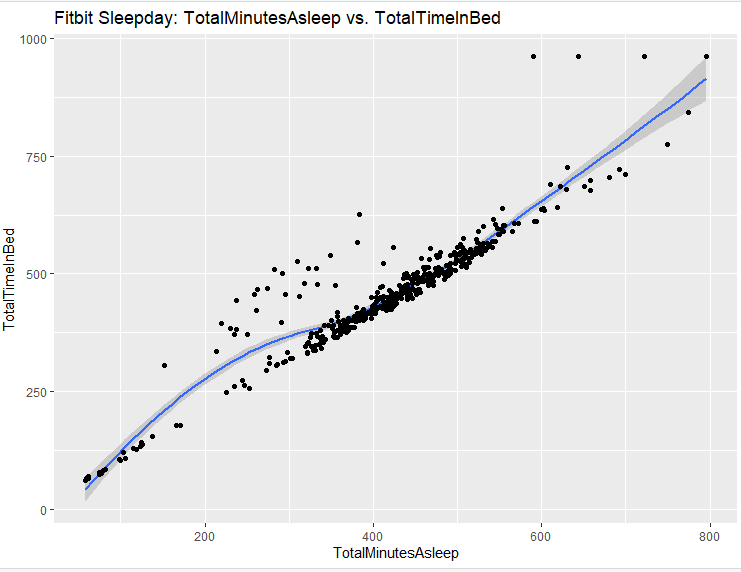
******

While this data frame has a low number of participants, the average BMI is 25.19. This is considered an overweight BMI. However, BMI can be a screening tool and does not diagnose the body fatness or health of an individual.

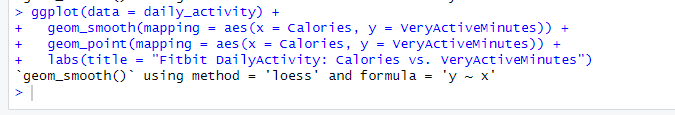
**Step 5: Share**

Ggplot2() function of R Studio was used to create data visualizations that depict patterns and trends found in this data frames.



**** Figure 1: The Relationship between Total Minutes Asleep and Total Tim In Bed

**Figure 1** shows a positive relationship between total minutes asleep and the total time spent in bed.



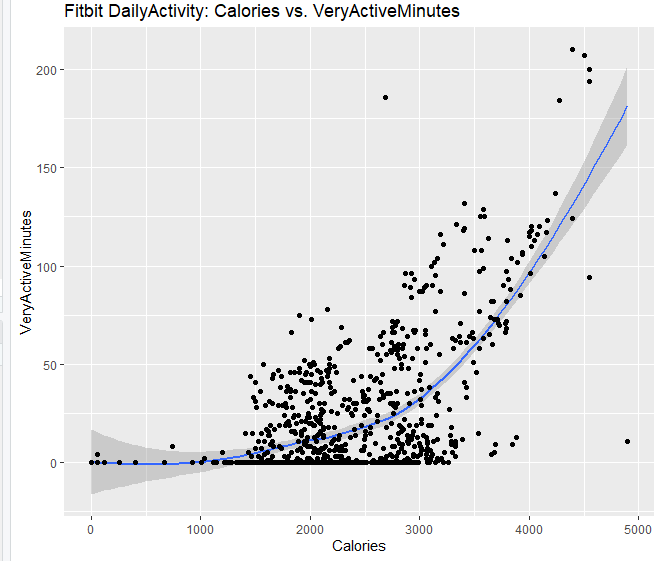
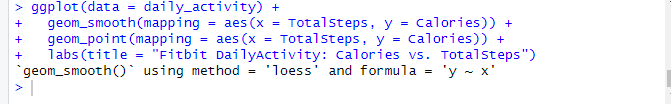


Figure 2: The Relationship between VeryActiveMinutes and Calories burned

Figure 2 shows a positive relationship between very active minutes and total daily calories burned. This means that the more vigorous physical activity the participant did, the more calories they burned.



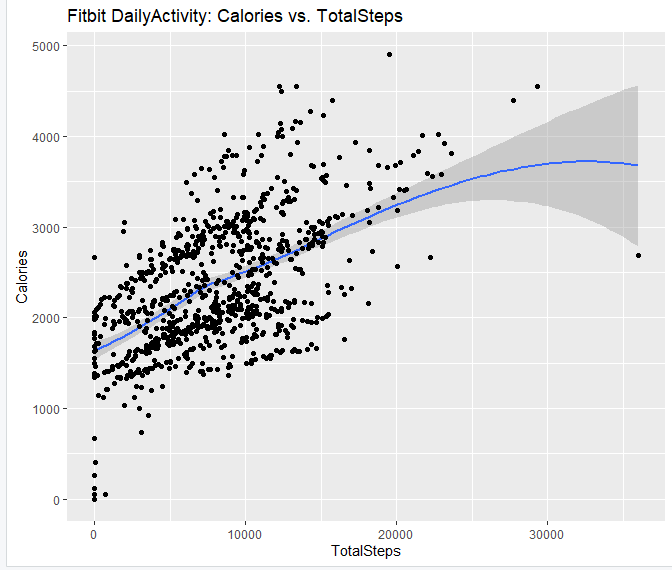
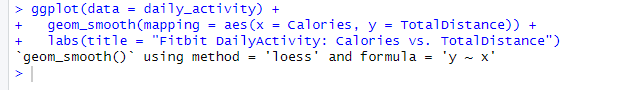


Figure 3: The Relationship between TotalSteps and Calories burned

Figure 3 shows a positive relationship between total daily steps taken and total calories burned. This means the more steps the Fitbit users took, the more calories they burned.



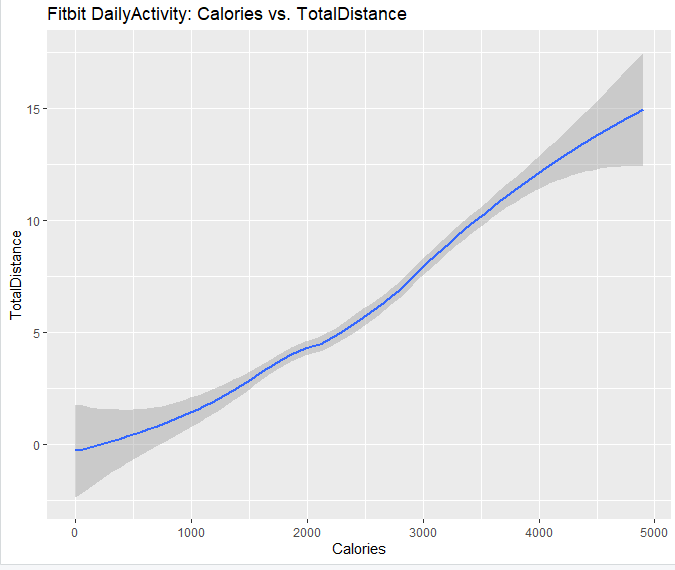


Figure 4: The Relationship between TotalDistance and Calories burned.

Figure 4 shows a positive relationship between total distance and total daily calories burned. As the participants moved a higher distance the more calories they burn.

**Step 6: Act  
Conclusion**after analyzing the datasets and the trends from the visualizations, we can conclude that non-active people have the tendency of having an unhealthy lifestyle. Some of the most important relations we found in the analysis are that the more vigorous physical activity the user does the more calories they burn. Also, the more steps the user takes, the more calories they burn. Lastly, the more distance a user moves, the more calories they burn. The relationship between the various activities we identified from the analysis of the participants who use smart devices to measure their daily activity can help Bellabeat make data-driven decisions for future products or functionality.

**Recommendations for Bellabeat app:**

* Create weekly fitness and wellness challenges to encourage use.
* Bellabeat can include a feature in the Bellabeat app that alerts users who have a high amount of sedentary minutes.
* Enable users to add friends and view each other’s activity.
* If the user’s resting heart rate differs considerably from the typical one enable alarm alerts.
* Bellabeat can add a timely notification in Leaf/Time to encourage users to get up and walk around more frequently in order to decrease their inactive minutes.
* Recommend users to get at least 10,000 steps each day and enable an alert to motivate the user to meet this goal.
* If a user has spent an hour awake in bed, enable alerts to encourage movement.
* Enable notifications to encourage activity if a user has been sedentary for an extended period of time.
* Encourage users to enter into weight and height to track BMI

**Recommendations for Bellabeat membership:**

* Offer a discount for the smart device when a user refers a friend
* There should be a month free trial on the subscription
* There should a discount on membership for all users.
* You can as well partner with some health and fitness companies to blend the exercise and health talk for users.
* There should be a rule which allows core members to pay in installments.

**Work Cited**

[About Adult BMI | Healthy Weight, Nutrition, and Physical Activity | CDC](https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html)  
[About Adult BMI | Healthy Weight, Nutrition, and Physical Activity | CDC](https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html)  
[The dangers of sitting: why sitting is the new smoking - Better Health Channel](https://www.betterhealth.vic.gov.au/health/healthyliving/the-dangers-of-sitting)  
[What Are METs, and How Are They Calculated? (healthline.com)](https://www.healthline.com/health/what-are-mets#calculation)