Bellabeat Case Study

Emerson Ceccarelli

5/1/2022

Bellabeat Time Analysis

1.1 Intro & Background

I am taking the role of a junior data analyst working on the marketing analyst team at Bellabeat, a fictional 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. I have been asked to focus on one of Bellabeat's products, Time, and analyze smart device data to gain insight into how consumers are using their smart devices. The insights discovered will then help guide marketing strategy for the company.

Here is a short discription of the product:

> Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

1.2 About the 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. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates.

Sršen knows that an analysis of Bellabeat's available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

2.1 About the Data

I have been encouraged to use the public data set *FitBit Fitness Tracker Data* (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

2.2 Importing the Data

As mentioned above this is a public data set. The useful files were imported to RStudio from a local hard drive.

```r
library(readr)
stepsDay <- read_csv("~/Desktop/Google CaCert - BellaBeat/Fitabase Data 4.12.16-5.12.16/dailySteps_merge

library(readr)
heartrateSeconds <- read_csv("~/Desktop/Google CaCert - BellaBeat/Fitabase Data 4.12.16-5.12.16/heartra

library(readr)
sleepDay <- read_csv("~/Desktop/Google CaCert - BellaBeat/Fitabase Data 4.12.16-5.12.16/sleepDay_merged

#Narrow form to be comparable with stepsDay
library(readr)
stepsMinute <- read_csv("~/Desktop/Google CaCert - BellaBeat/Fitabase Data 4.12.16-5.12.16/minuteStepsN
```

Preliminary analysis reveals some insights and limitations of the data. First it is clear that there is an abundance of information, however the actual sample size and duration is relatively small. Furthermore, there are no demographics given which could have been useful for analysis since Bellabeat is specifically marketed to women. Overall this data is rather clean with some strange naming consistencies. I will be looking at the daily charts for steps and sleep. I have chosen not to include the heartrate data in the analysis as it is ultimately inefficient to analyze since the tracking should happen automatically and does not inherently increase user interaction nor is there an appropriate amount of room to advance technology without considerable cost to the consumer.

2.3 Cleaning the Data

The following lines are eliminating duplicate or missing entries.

```r
stepsDay %>%
  drop_na() %>%
  distinct()

heartrateSeconds %>%
  drop_na() %>%
  distinct()

sleepDay %>%
  drop_na() %>%
  distinct()

stepsMinute %>%
```

```
  drop_na() %>%
  distinct()
```

The "stepsMinute" file had a massive amount of empty data since the device would still record 0 steps per minute. So the next line of code removes empty entries.

```
stepsMinute <- filter(stepsMinute, Steps != 0)
```


3.1 Analysis

Next is looking at the data for "sleepDay" and "stepsDay" to look at their usage in the form of a "geom_count" to plot the information in order to see which type of data is collected more often and by more users.

```
plot_sleepDay <- ggplot(data = sleepDay) +
  geom_count(mapping = aes(x = Id/1000000000, y = TotalTimeInBed), show.legend = FALSE) +
  xlim(1503960366/1000000000, 8877689391/1000000000) +
  labs(x ="Individuals", y = "Total Time in Bed", title = "Sleep Tracking per Day") +
  labs(caption = "This chart aims to visualize the general usage of tracking sleep patterns")

plot_stepsDay <- ggplot(data = stepsDay) +
  geom_count(mapping = aes(x = Id/1000000000, y = StepTotal), show.legend = FALSE) +
  xlim(1503960366/1000000000, 8877689391/1000000000) +
  labs(x = "Individuals", y = "Total Steps", title = "Step Tracking per Day") +
  labs(caption = "This chart aims to visualize the gerneral usage of traking steps per day")
```

Now its time to look at the plots and figure out which type of data is more dense.
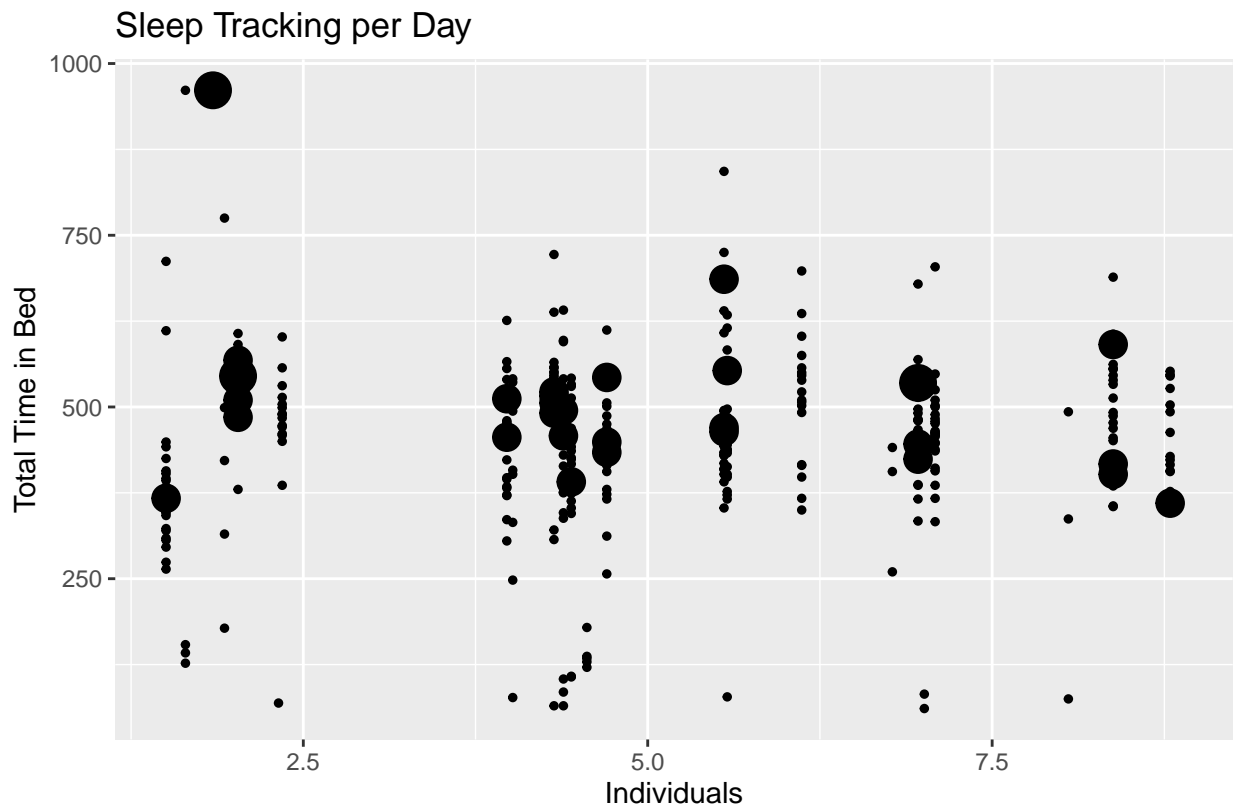
Sleep Tracking per Day

This chart aims to visualize the general usage of tracking sleep patterns

Step Tracking per Day
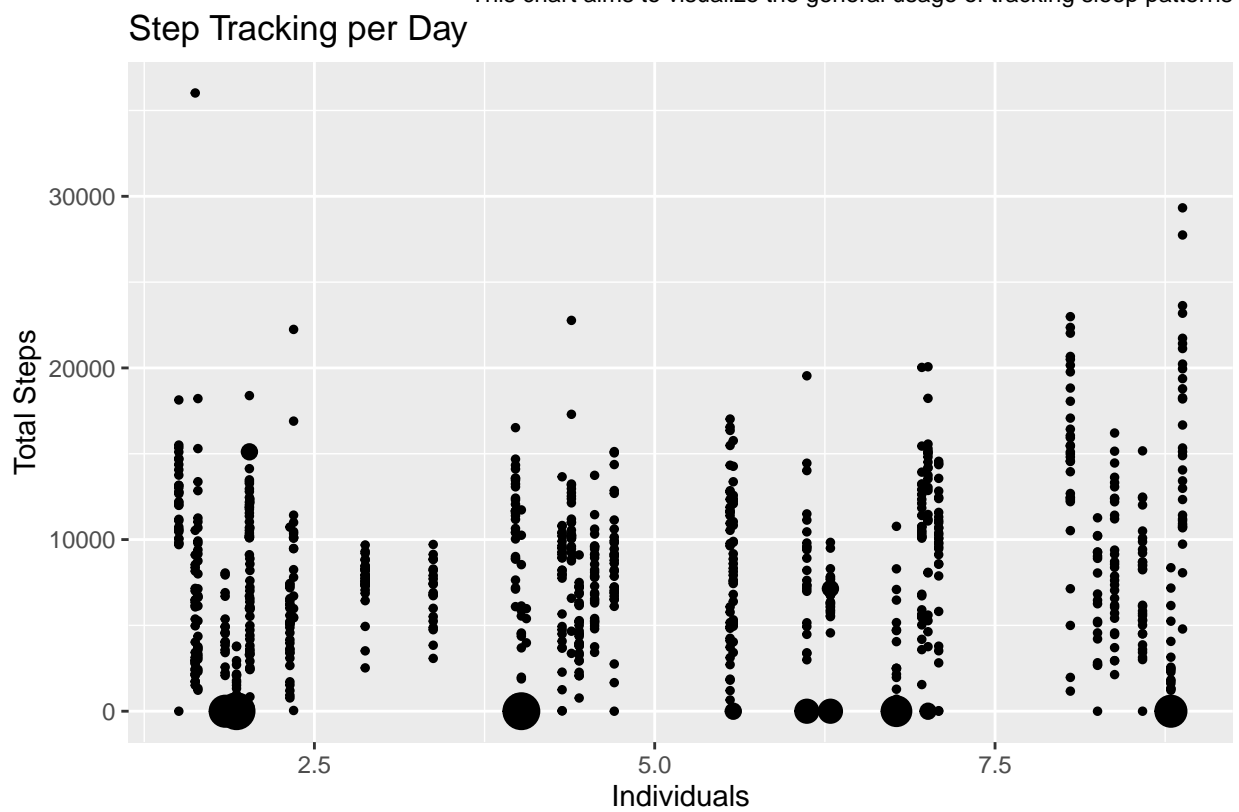
This chart aims to visualize the gerneral usage of traking steps per day

4.1 Conclusion & Recomendations

Based on the charts above, it is difficult to say which data type is more dense but there are some factors to consider. While the sleep data does have higher amounts of pooled information, the steps data has more individuals entering data. For that reason, I would suggest further developing resources to track consumers' walking data over sleeping data. This particular advancement would target more consumers overall for similarly dense pools of data.