BellaBeat Analysis

Kyle Flashman

2023-08-04

This study involves 33 participants who used the Bellabeat Fitness tracker from April 12 through May 12, 2016. The Bellabeat team would like to know how the participants have used the device over the course of the study. We will be looking at the frequency of usage for each user and for each day of the week for all of the participants.

Data Cleaning and Preparation

Three duplicates were removed from the Sleep data set and no 'NA's were found in any of the data sets. Then the timestamps were separated into date and time variables. Also, the day of the week was added as a variable for each of the data sets of interest.

Data Analysis

First, we will be looking at how many times each participant used the device as well as which days were used for all the participants. Next, those who had more than 1000 steps in a day will be looked at. Then the usage of the heartrate, sleep, and weight monitors will be analyzed.

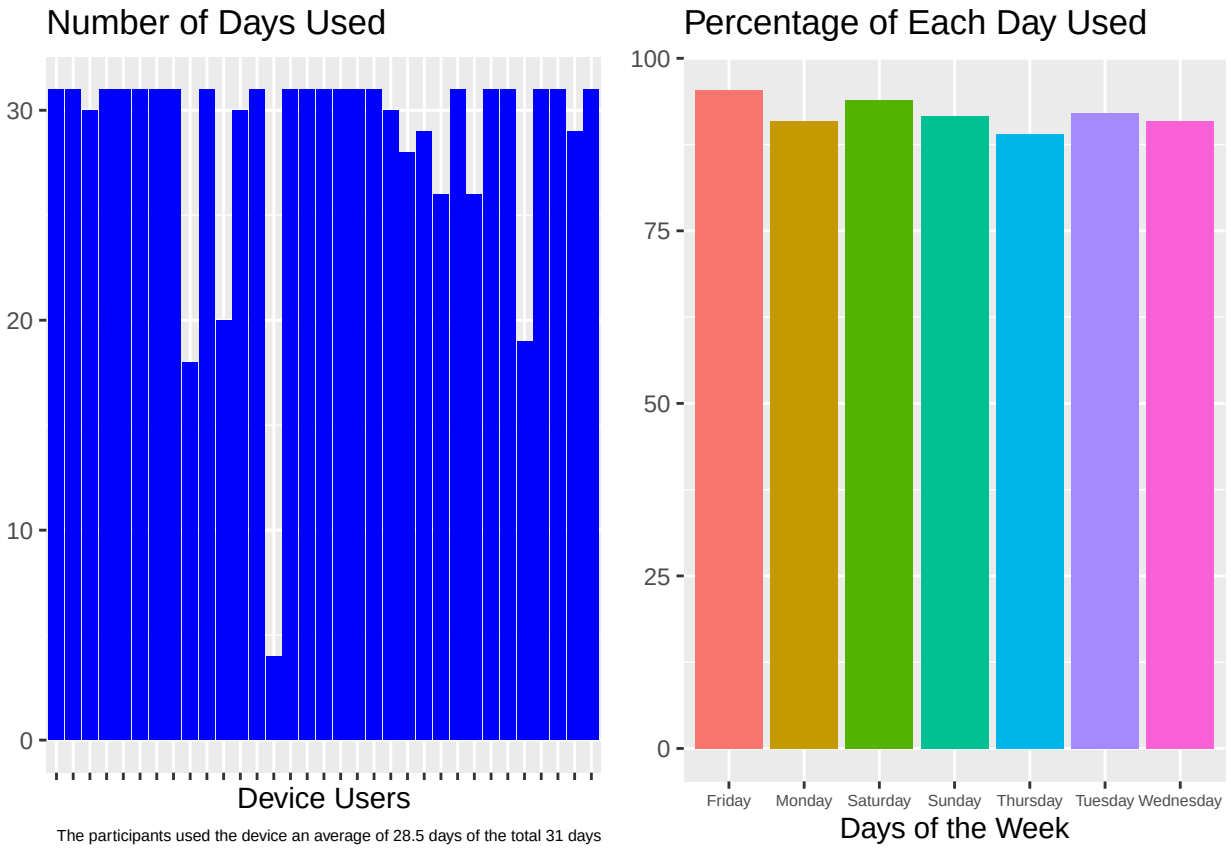
Daily Activities

```
daTotal <- dailyActivity %>% group_by(Id) %>%
  summarise(distinct_points = n_distinct(ActivityDate))
count_days <- c(4, 4, 4, 4, 5, 5, 5) * 33
totalDays <- dailyActivity %>% group_by(Day) %>%
  summarise(total_count = n())
totalDays$total <- count_days
dayOrder <- factor(totalDays$Day, levels= c("Sunday", "Monday",
                                            "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
totalDays <- totalDays[order(dayOrder),]

daStepsCount <- dailyActivity[dailyActivity$TotalSteps > 1000,] %>%
  group_by(Id) %>% summarise(total_count = n())
stepsByDay <- dailyActivity[dailyActivity$TotalSteps > 1000,] %>%
  group_by(Day) %>% summarise(total_count = n())
stepsByDay$total <- count_days
stepsByDay <- stepsByDay[order(dayOrder),]

p1 <- ggplot(daTotal) +
```

```
  geom_col(aes(x = daTotal$Id, y = daTotal$distinct_points), fill = "blue") +
  theme(axis.text.x = element_blank(), axis.title.y = element_blank(), plot.caption = element_text(size
  labs(x = "Device Users", title = "Number of Days Used",
       caption = glue("The participants used the device an average of {round(mean(daTotal$distinct_poin
p2 <- ggplot(totalDays) +
  geom_col(aes(x = totalDays$Day, y = totalDays$total_count/totalDays$total*100, fill = totalDays$Day))
  theme(legend.position = "none", axis.text.x = element_text(size = 6),
        axis.title.y = element_blank()) +
  labs(x = "Days of the Week", title = "Percentage of Each Day Used")
grid.arrange(p1, p2, nrow = 1)
```

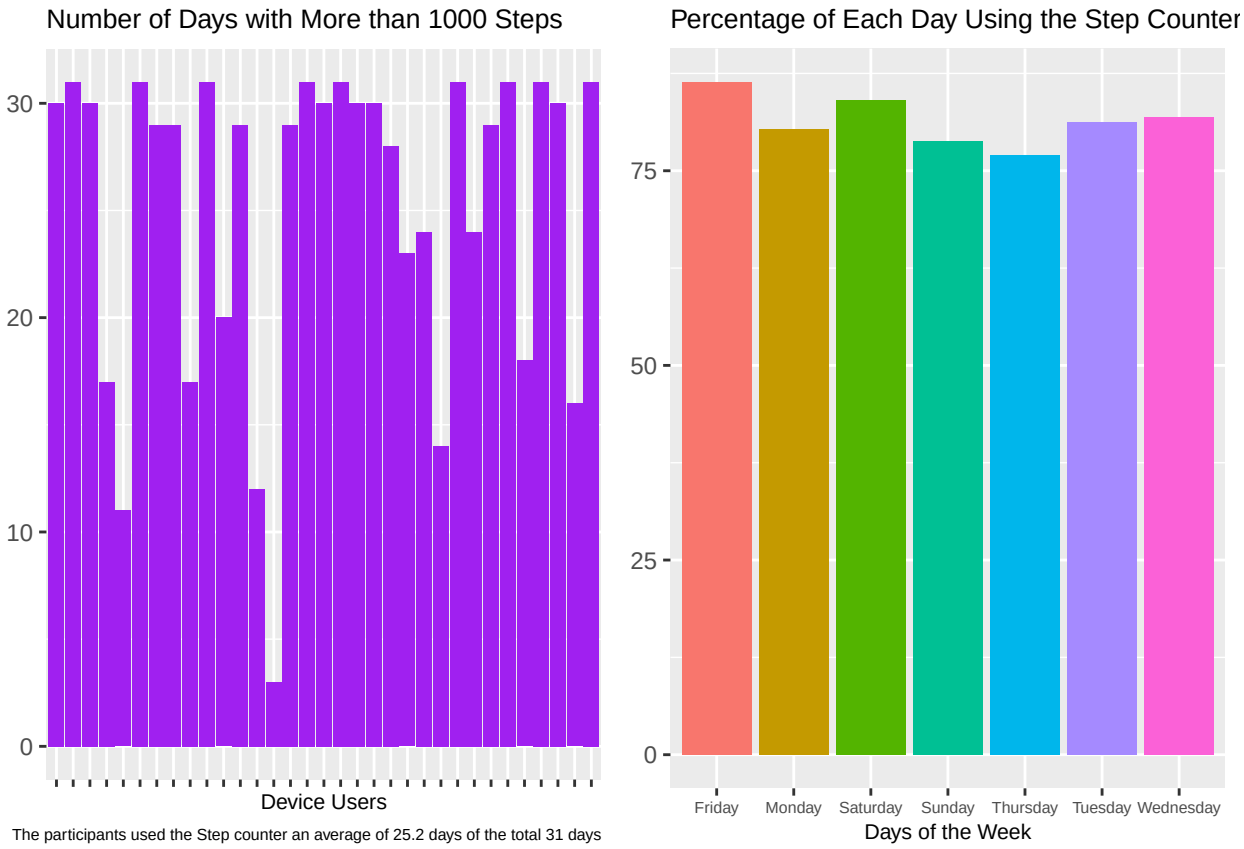


23 of the 33 participants used the device everyday of the study with an average of 28.5 days. Friday was the most popular day to use at 95% followed by Saturday and Tuesday. All days were above 89%.

Step Counter

```
p21 <- ggplot(daStepsCount) +
  geom_col(aes(x = daStepsCount$Id, y = daStepsCount$total_count), fill = "purple") +
  theme(axis.text.x = element_blank(), axis.title.y = element_blank(),
        title = element_text(size = 8))+
  labs(x = "Device Users", title = "Number of Days with More than 1000 Steps",
       caption = glue("The participants used the Step counter an average of {round(mean(daStepsCount$to
p22 <- ggplot(stepsByDay) +
  geom_col(aes(x = stepsByDay$Day, y = stepsByDay$total_count/stepsByDay$total*100, fill = stepsByDay$Da
```

```
  theme(legend.position = "none", axis.title.y = element_blank(), axis.text.x = element_text(size = 6),
        title = element_text(size = 8))+
  labs(x = "Days of the Week", title = "Percentage of Each Day Using the Step Counter")
grid.arrange(p21, p22, nrow = 1)
```

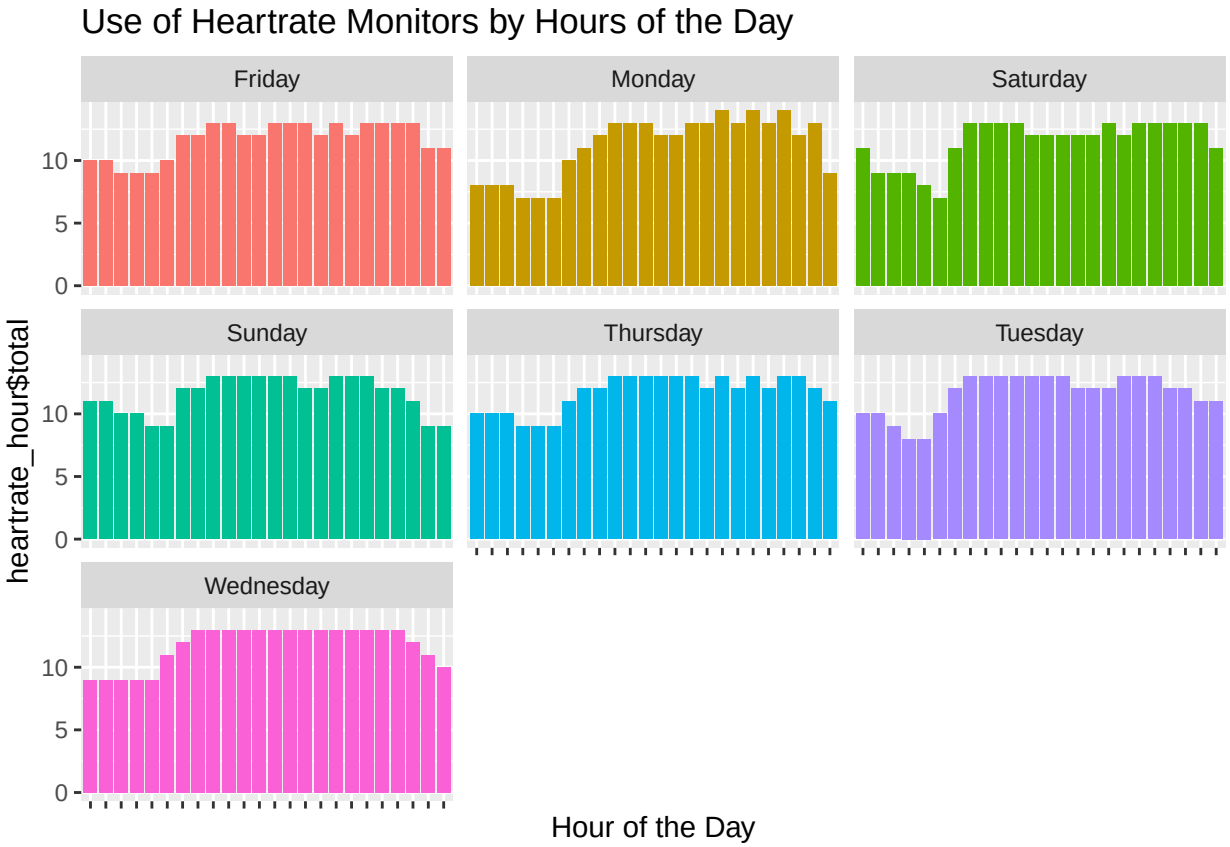


I am considering the participants with less than 1000 steps in a day as not using the step counter for that given day. And as shown, there were an average of 25.2 days using the step counter over the 31 days of the study. Friday and Saturday were the clear favorite days to use the step counter at 86% and 84% respectively. Thursday was the least popular day with 77% using the step counter.

Heartrate Monitor

```
heartrate_hour <- heartrate %>% group_by(Day, Hour) %>%
  summarise(total = n_distinct(Id, Day, Hour), .groups = 'keep')

ggplot(heartrate_hour) +
  geom_col(aes(x = heartrate_hour$Hour, y = heartrate_hour$total, fill = heartrate_hour$Day)) +
  facet_wrap(~heartrate_hour$Day) +
  theme(legend.position = "none", axis.text.x = element_blank()) +
  labs(x = "Hour of the Day", title = "Use of Heartrate Monitors by Hours of the Day")
```



Each day has roughly the same shape for using the heartrate monitor when granulated to the hour of the day. Less usage while sleeping during the nights and more usage throughout the middle of the days. Fridays and Sundays were the most popular days to monitor heartrate.

Sleep

```
totalSleep <- sleep %>% group_by(Id) %>% summarise(distinct_points = n_distinct(SleepDay))
sleep_merged <- merge(x = daTotal, y = totalSleep, by = "Id", all.x = T)
sleep_merged[is.na(sleep_merged$distinct_points.y),]$distinct_points.y <- 0

sleepByDay <- sleep %>% group_by(Day) %>% summarise(countDays = n())
sleepByDay$totalDays <- totalDays$total

p31 <- ggplot(sleep_merged) +
  geom_col(aes(x = sleep_merged$Id, y = sleep_merged$distinct_points.y), fill = "brown") +
  theme(axis.text.x = element_blank(), axis.title.y = element_blank()) +
  labs(x = "Device Users", title = "Logged Sleep Days",
       caption = glue("The participants logged their sleep an average of {round(mean(sleep_merged$distin

p32 <- ggplot(sleepByDay) +
  geom_col(aes(x = sleepByDay$Day, y = sleepByDay$countDays/sleepByDay$totalDays*100, fill = sleepByDay$
  theme(legend.position = "none", axis.title.y = element_blank(), axis.text.x = element_text(size = 6))
  labs(x = "Days of the Week", title = "Percentage of Sleep Logged by Day")
```

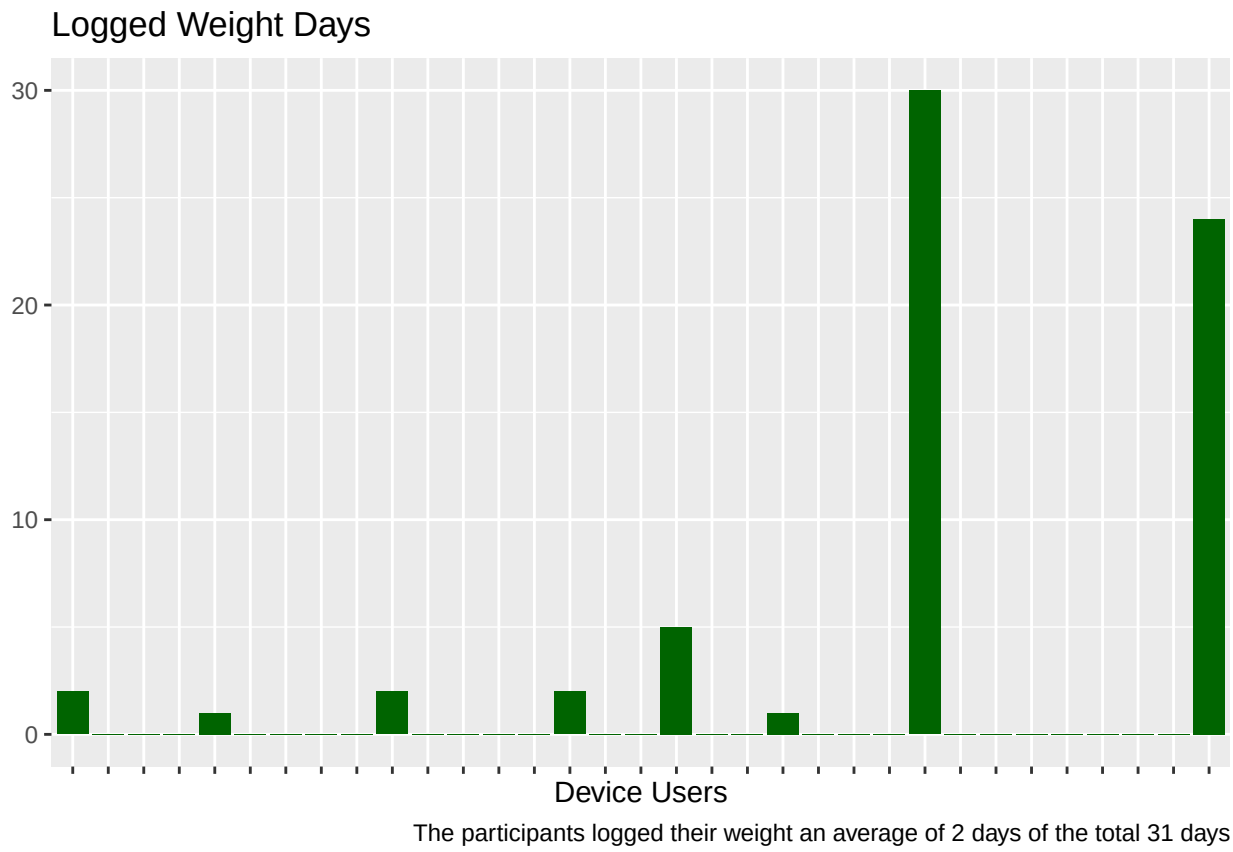
Not as many participants logged their sleep compared to the other features of the device. Only an average

of 12.4 days were logged. Also, unlike the other features Wednesdays and Tuesdays were the most popular days to log their sleep and Saturday and Sunday the least popular.

Weight

```
totalWeight <- Weight %>% group_by(Id) %>% summarise(distinct_points = n_distinct(Date))
weight_merged <- merge(x = daTotal, y = totalWeight, by = "Id", all.x = T)
weight_merged[is.na(weight_merged$distinct_points.y),]$distinct_points.y <- 0

ggplot(weight_merged) +
  geom_col(aes(x = weight_merged$Id, y = weight_merged$distinct_points.y), fill = "dark green") +
  theme(axis.text.x = element_blank(), axis.title.y = element_blank())+
  labs(x = "Device Users", title = "Logged Weight Days",
       caption = glue("The participants logged their weight an average of {round(mean(weight_merged$dis
```



Logging weight was the least popular feature of the fitness device with only an average of 2 days were logged.

Conclusion

All in all, since sleep and weight were the least popular features of the device it may help to bring up reminders to do so as this will better help them manage these important aspects of one's health. Also including a notification of how many steps they took on the previous day so they can be more aware of this in future days.