

# Bellabeat Case Study

Alan Roebuck

2023-09-30

This case study was performed as a part of the Google Analytics career certificate course. All code necessary to reproduce visual and statistical analysis associated with this project are available via Github ([https://github.com/jar3559/Bellabeat\\_Case\\_Study\\_Google\\_Analytics](https://github.com/jar3559/Bellabeat_Case_Study_Google_Analytics)).

## 1. Case Study Scenario

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 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.

- **Bellabeat Products**

- *Bellabeat app*: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products
- *Leaf*: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
- *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
- *Spring*: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

## 2. Assignment

You are asked to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices and to select one Bellabeat product to apply these insights to in your presentation. These question will guide your analysis:

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?

## 3. Data Availability and Description

### 3.1 Data Source

To complete this project, data for Fitbit users at various time resolutions was obtained from Kaggle (<https://www.kaggle.com/datasets/arashnic/fitbit>), which was originally sourced from Furberg et al., 2016, doi: 10.5281/zenodo.53894 (<https://doi.org/10.5281/zenodo.53894>).

### 3.2 Data Files

The dataset consist of the following files:

File Name	Description
dailyActivity_merged.csv	Daily calories, steps, and intensities (active minutes)
dailyCalories_merged.csv	Total daily calories burned
dailyIntensities_merged.csv	Total daily minutes for which users were active or sedentary
dailySteps_merged.csv	Total daily steps
heartrate_seconds_merged.csv	User heart rate beats per minute at second intervals
hourlyCalories_merged.csv	Hourly calories
hourlyIntensities_merged.csv	Number of active and sedentary minutes by hour
hourlySteps_merged.csv	Number of hourly steps
minuteCaloriesNarrow_merged.csv	Calories burned by minute in long frame format
minuteCaloriesWide_merged.csv	Calories burned by minute in wide frame format
minuteIntensitiesNarrow_merged.csv	Active and sedentary minutes in long format
minuteIntensitiesWide_merged.csv	Active and sedentary minutes in wide format
MInuteMETsNarrow_merged.csv	Metabolic equivalents (MET) per minute
minuteSleep_merged.csv	Each individual minute logged as sleep
minuteStepsNarrow_merged.csv	Steps logged each minute long format
minuteStepsWide_merged.csv	Steps logged each minute wide format
sleepDay_merged.csv	Total number of minutes slept per day
weightLogInfo_merged.csv	Information related to user weight

### 3.3 Initial Assessment of Data

Broadly, it cannot be said with confidence this data is a good representation of the Bellabeat client base, which targets middle age women. For instance, this dataset provides no information on race, sex, age, nor geographic information. Thus, we can only make small inferences regarding the broader fitness community rather than specifically those specifically related to women.

Furthermore, this data is quite limited, with only 33 participants over a 31 day period. Thus, we would most likely consider this data a preliminary effort to better inform the design of a larger study rather than being directly used to inform major decisions.

## 4. Questions and Hypotheses

Here, a set of 5 questions and corresponding hypothesis are presented. Questions 1-3 ask about general fitness patterns for users who participated in the study while questions 4 and 5 are directly linked with how users use the fitness watches and products with implications marketing strategy.

**Q1:** When are people most active throughout the day and does that couple with energy expenditure? How does this compare among people of different activity level?

- *H1:* It is hypothesized that all users of all activity levels will be most active in the late evening, likely due to uptick in workouts at the end of the day or due to general after work/ after school activities that will occur

**Q2:** Are people more active during the typical workweek or on weekends and are there differences based on peoples activity level?

- *H2:* It is hypothesized that users of all activity levels will see a general increase in activity during the weekends as this is when most people have free time.

**Q3:** Do more physically active users get more hours of sleep at night?

- *H3:* It is hypothesized that sedentary users sleep more compared to active users

**Q4:** Is fitness level an indicator how many days users consistently wear a fitness watch (e.g., motivation to regularly use product)?

- *H4:* It is hypothesized that more active people are more motivated to consistently wear their fitness watch on a daily basis.

**Q5:** Is fitness level an indicator of how often wear a fitness watch to monitor sleep activity (e.g. motivation to use a product feature)?

- *H5* It is hypothesized that more active users are also more motivated to use their watch to monitor sleep activity and will more consistently wear their devices to bed.

## 5. Data Preparation, Cleaning, and Aggregation

### 5.1 Retrieve Daily Activity, Sleep, and Weight Data

1. First, daily activity, sleep and weight data were retrieved from the files 'dailyActivity\_merged.csv', 'sleepDay\_merged.csv', and 'weightLogInfo\_merged.csv'.
2. The total active and sedentary minutes were totaled as a potential indicator of how much time during the day a fitness watch was worn. This is assessed in more detail in the following section
3. A BMI class (World Health Organization (<https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations>)) was assigned as follows:

BMI Score	BMI Class
Under 18.5	Underweight
18.5 - 25	Normal
25 - 30	Overweight
Over 30	Obese

### 5.2 Assessment of Total Active and Sedentary Minutes as Indicators of Time a Watch is Worn

Above, the total daily active and sedentary minutes were summed to represent the cumulative time a fitness watch was worn throughout the day. However, after skimming the data, it was observed that there are a number of instances in which zero steps were logged. This is likely due to the fact that a Fitbit will continue to record data even when a watch is not worn. For instance, when a watch is not worn, Fitbit will continue to log minutes as "sedentary".

The table below shows that there were 77 data entries in which the *total wear time* would indicate some use of the watch throughout the day, however the watch recorded a total step value of '0'. This is an indication that some users were not actively wearing their watch a despite having some data recorded. There were 15 participants with at least 1 entry and 4 participants with greater than 10 daily entries that meet this criteria.

**This indicates that the Total Wear Time calculated based on the cumulative active and sedentary minutes is not reliable** and an alternative means to assess how consistently users wear their fitness watch is needed.

```
## # A tibble: 15 × 2
##       Id active_nonuse
##   <dbl>     <int>
## 1 1503960366         1
## 2 1844505072        10
## 3 1927972279        14
## 4 4020332650        14
## 5 4057192912         1
## 6 4702921684         1
## 7 5577150313         2
## 8 6117666160         5
## 9 6290855005         5
## 10 6775888955         9
## 11 7007744171         2
## 12 7086361926         1
## 13 8253242879         1
## 14 8583815059         1
## 15 8792009665        10
```

```
## # A tibble: 1 × 1
##   total_active_nonuse
##   <int>
## 1         77
```

## 5.3 Aggregate Data By Individual User to Determine Average Daily Activities

The data was aggregated by individual user to describe their **average daily activity**. The total number of days in which a watch was used was determined based on the number of daily steps taken rather than the total active and sedentary minutes, which was deemed unreliable above. However, an additional challenge with regards to assigning activity based on the number of daily steps is that there are some instances in which the step count is unreasonably low (e.g., less than 100) assuming a healthy sampling population. According to the Mayo Clinic, the average American walks 3,000-4,000 steps per day. Thus, a very loose approach is taken to say a person must have walked at least 500 steps in order for the watch to be credited as a daily use.

Once filtering to include only records above the 500 step threshold, the number of daily entries for each user were summed and the active daily use was arbitrarily defined as:

Number of Days Counted	Daily Use Metric
30+ days	High Daily Use
20-29 days	Moderate Daily Use
Fewer than 20 days	Low Daily Use

Second, the average daily steps for each user was determined (where daily steps > 500) and an activity score was assigned based on recommendations from the 10,000 Steps Organization

(<https://www.10000steps.org.au/articles/healthy-lifestyles/counting-steps/>). Activity levels were assigned as follows:

Number of Daily Steps	Activity Level
Fewer than 5000	Sedentary
5,000 - 10,000	Lightly Active
Over 10,000	Active

Third, the average number of hours of sleep, total time in bed, and total time awake in bed (time in bed - hours of sleep) were calculated.

Fourth, a table containing the individual user and their BMI was made.

Finally, all aggregated data just described were merged

```
## # A tibble: 6 × 14
##       Id days_used active_use   avg_steps sd_steps activity_level BMI_class
##       <dbl>   <int> <chr>         <dbl>   <dbl> <chr>         <chr>
## 1 1503960366     30 High Daily U... 12521.   2099. Active       Normal
## 2 1624580081     31 High Daily U...  5744.   6177. Lightly Active Not Repo...
## 3 1644430081     30 High Daily U...  7283.   4325. Lightly Active Not Repo...
## 4 1844505072     17 Low Daily Use    4690.   1823. Sedentary     Not Repo...
## 5 1927972279     13 Low Daily Use    2115.    966. Sedentary     Obese
## 6 2022484408     31 High Daily U... 11371.   2807. Active       Not Repo...
## # i 7 more variables: sleep_days_used <dbl>, avg_sleep_hr <dbl>,
## #   sd_sleep_hr <dbl>, avg_time_bed_hr <dbl>, sd_time_bed_hr <dbl>,
## #   avg_awake_hr <dbl>, sd_awake_hr <dbl>
```

## 5.4 Retrieve Hourly Calorie and Steps Data

Hourly calorie and hourly steps data were retrieved from the files 'hourlyCalories\_merged.csv' and 'hourlySteps\_merged.csv'. In both cases, data is reshaped from 'long' to 'wide' format with column names renamed based on hourly intervals.

## 5.5 Clean Hourly Data

The hourly data was cleaned by removing data where the total number of daily steps was less than 500.

```
## [1] "Hourly Steps"
```

```

##          Id          Date    active_use activity_level hr01 hr02 hr03 hr04 hr05
## 1 1503960366 2016-04-12 High Daily Use      Active  373  160  151   0   0
## 2 1503960366 2016-04-13 High Daily Use      Active  144   0   0   36   0
## 3 1503960366 2016-04-14 High Daily Use      Active   81   58   0   0   0
## 4 1503960366 2016-04-15 High Daily Use      Active   83   75  124   0   0
## 5 1503960366 2016-04-16 High Daily Use      Active  459   0   63   6   6
## 6 1503960366 2016-04-17 High Daily Use      Active   0   0   0   8   0
##   hr06 hr07 hr08 hr09 hr10 hr11 hr12 hr13 hr14 hr15 hr16 hr17 hr18 hr19 hr20
## 1    0    0    0  250 1864  676  360  253  221 1166 2063  344  489 1386  558
## 2    0    0  126  275 1679  295  524  486  302 1577  493  299  695 1343  258
## 3    0    0    0    0  902 1106   50  501 1779 1228  139  408   37  322   11
## 4   26    0    7   55   21  154  425  270  373    0    0    0 1774  509  519
## 5   12    0  101  175 1181 1295 2043 1513 3616  580    0   15  380   96  146
## 6    0    0    0    0    0    0   84  260  402  252  191   13 2613 1616  754
##   hr21 hr22 hr23 hr24
## 1 1733  684   89  338
## 2    0    0  437 1766
## 3  806 2670  362    0
## 4   77  280 2145 2768
## 5  385  463   42   92
## 6 2955  487   70    0

```

```
## [1] "Hourly Calories"
```

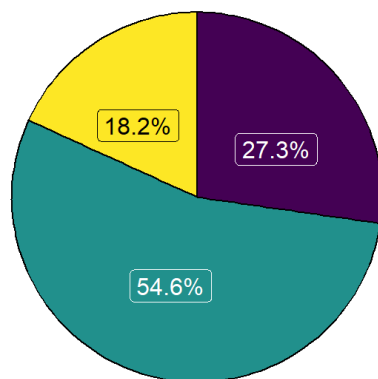
```

##          Id          Date    active_use activity_level hr01 hr02 hr03 hr04 hr05
## 1 1503960366 2016-04-12 High Daily Use      Active   81   61   59   47   48
## 2 1503960366 2016-04-13 High Daily Use      Active   69   48   47   53   47
## 3 1503960366 2016-04-14 High Daily Use      Active   56   54   47   47   47
## 4 1503960366 2016-04-15 High Daily Use      Active   60   58   59   47   47
## 5 1503960366 2016-04-16 High Daily Use      Active   77   48   52   48   49
## 6 1503960366 2016-04-17 High Daily Use      Active   47   48   47   49   48
##   hr06 hr07 hr08 hr09 hr10 hr11 hr12 hr13 hr14 hr15 hr16 hr17 hr18 hr19 hr20
## 1   48   48   47   68  141   99   76   73   66  110  151   76   83  124  104
## 2   47   48   59   69  136   77   84   82   72  128   79   71   87  118   73
## 3   47   47   47   47   93  109   64   85  137  113   66   88   61   72   55
## 4   49   47   49   54   50   63   84   74   67   47   47   47  140   87   97
## 5   49   48   54   62  124  110  147  119  214   82   51   54   79   65   67
## 6   48   48   48   48   48   47   63   76   83   69   73   57  174  136   95
##   hr21 hr22 hr23 hr24
## 1  132  100   65   81
## 2   47   48   74  135
## 3   91  174   79   50
## 4   64   75  151  182
## 5   81   75   55   56
## 6  191   79   58   50

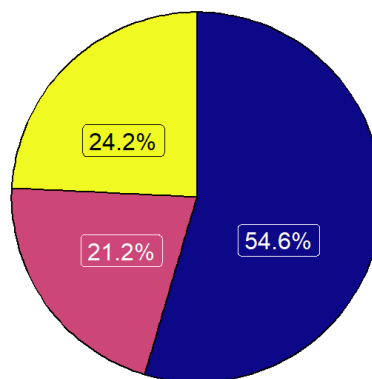
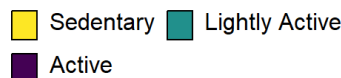
```

## 5.6 Preliminary Assessment of Data Population Represented in Study

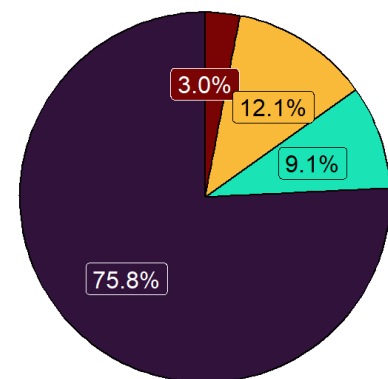
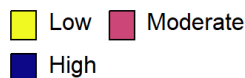
- Here, a series of pie charts are used to display general information about fitness watch users in this study. The proportion of 'Lightly Active' users represents more than 50% of the data set and is overrepresented compared to 'Active' and 'Sedentary' users.
- Similarly, the number of users who more consistently wear their fitness watch is >50% and is also over represented compared to those who only moderately wore or were inconsistent in their daily fitness watch usage.
- Finally, of the 33 participants in the study, most did not provide weight information rendering use of the BMI or weight related metrics as unreliable for effective use in this study.



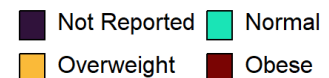
Activity Level



Average Days Used



BMI Class



## 6. Results

*Question 1: When are people most active throughout the day and does that couple with energy expenditure? How does this compare among people of different activity level?*

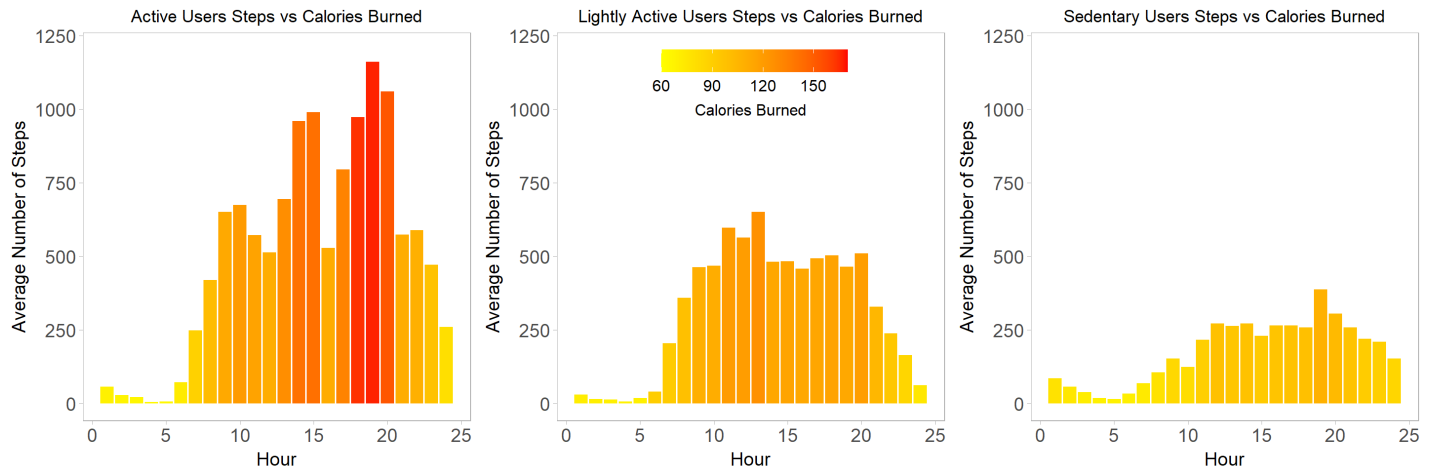
**Major Finding:** The hypothesis that people will be most active in the evening is partially confirmed and is based primarily on a users activity level.

**Key Points:**

- Active Users were most active with the greatest energy expenditure during the 6-8 pm hours, likely driven by desire for physical activity after the conclusion of the work day.
- Active users also see a spike in activity around the 12-2 pm hours, likely driven by mid-day lunchtime walks.
- Lightly active users also saw a small uptick around the 12-2 pm hours suggesting some small efforts for mid-day lunchtime walks
- Sedentary users averaged <500 steps per hour with a small up tick around 7 pm that may be driven by dinnertime efforts.



- These users also displayed slightly higher number of steps in early hours indicating these users may have a later bed time schedule.

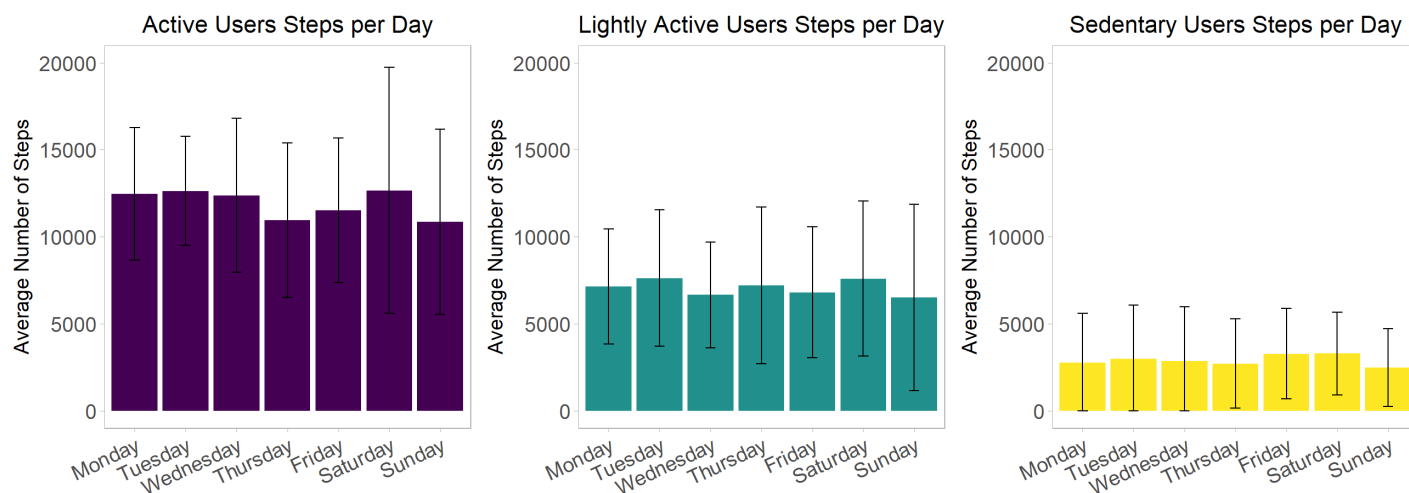


**Question 2:** Are people more active during the typical workweek or on weekends and are there differences based on peoples activity level?

**Major Finding:** People are equally active across the entire week rejecting the hypothesis that people would be more active on the weekends. This is true regardless of activity level.

**Key Points:**

- There was no significant difference in the number of steps taken for each day of the week regardless of activity level.
  - This is observed visually in the figure below noting the relatively stable average number of steps taken each day of the week for each activity level and the strong overlapping error bars
  - This observation was confirmed statistically with a non-parametric Kruskal-Wallis test for each activity level which highlight the number of steps taken each day of the week is not significantly different ( $p > 0.05$ )



```
## [1] "Kruskal Wallis Test for Highly Active Users to determine if there are significant differences in the number of steps taken throughout each day of the week"
```

```
## # A tibble: 1 × 6
##   .y.          n statistic    df      p method
## * <chr>      <int>      <dbl> <int> <dbl> <chr>
## 1 TotalSteps  260        10.0     6 0.123 Kruskal-Wallis
```

```
## [1] "Kruskal Wallis Test for Lightly Active Users to determine if there are significant differences in the number of steps taken throughout each day of the week"
```

```
## # A tibble: 1 × 6
##   .y.          n statistic    df      p method
## * <chr>      <int>      <dbl> <int> <dbl> <chr>
## 1 TotalSteps  501         7.77     6 0.255 Kruskal-Wallis
```

```
## [1] "Kruskal Wallis Test for Sedentary Users to determine if there are significant differences in the number of steps taken throughout each day of the week"
```

```
## # A tibble: 1 × 6
##   .y.          n statistic    df      p method
## * <chr>      <int>      <dbl> <int> <dbl> <chr>
## 1 TotalSteps  179         2.47     6 0.871 Kruskal-Wallis
```

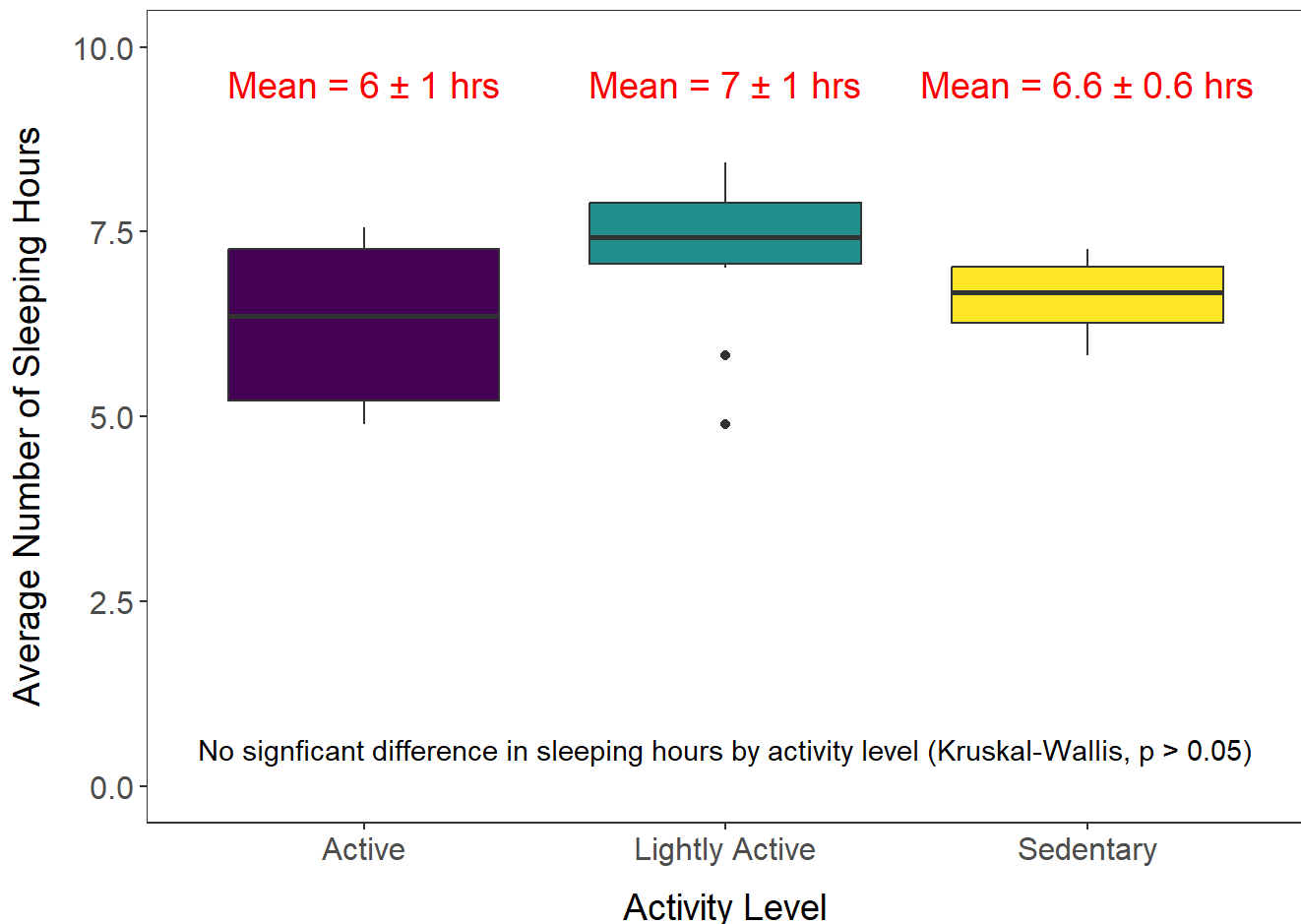
### Question 3: Do more physically active users get more hours of sleep at night?

**Major Finding:** There was no difference in how many hours of sleep a user received based on their activity level rejecting the hypothesis that sedentary users would likely sleep more than more active users.

Key Points:

- The average hours of sleep is provided in the table below and displayed visually in the figure below. While lightly active users had a slightly higher number of hours of sleep, this was not significantly different (Kruskal-Wallis,  $p > 0.05$ ).
  - Note this analysis removed 2 outliers that indicated  $< 3$  hours of sleep and 1 outlier that indicated  $> 10$  hours of sleep in one night.

```
## # A tibble: 3 × 4
##   activity_level median_sleep avg_sleep sd_sleep
##   <chr>          <dbl>     <dbl>   <dbl>
## 1 Active          6.36       6.26    1.18
## 2 Lightly Active  7.43       7.19    1.07
## 3 Sedentary       6.68       6.61    0.629
```



```
## [1] "Kruskal Wallis Test to determine if there is a difference in the average numbe of sleepi
ng hours by activity level"
```

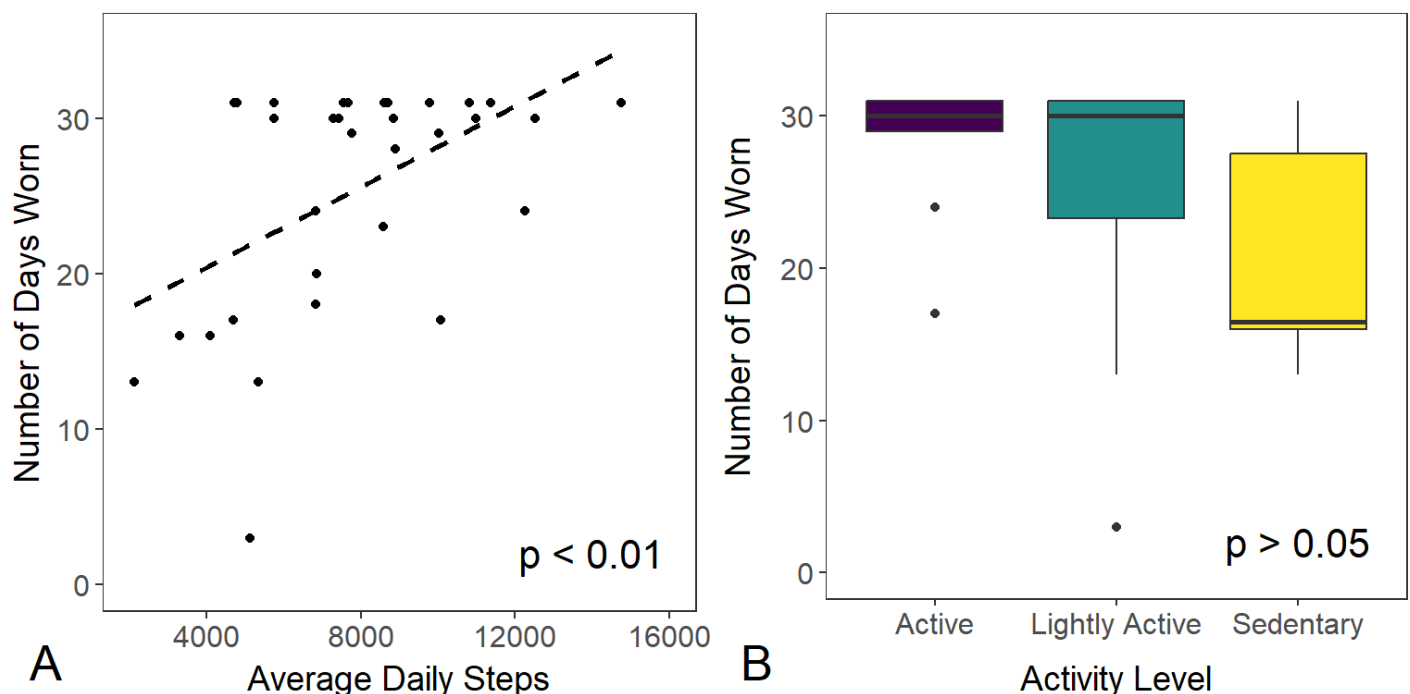
```
## # A tibble: 1 × 6
##   .y.          n statistic    df      p method
## * <chr>      <int>    <dbl> <int> <dbl> <chr>
## 1 avg_sleep_hr    20      3.30     2 0.192 Kruskal-Wallis
```

**Question 4:** *Is fitness level an indicator how many days users wear a fitness watch within a month (e.g., motivation to regularly use product)?*

**Major Findings:** Users who take more steps wear their watch more consistently indicating that more active users are likely to show higher motivation and need for purchasing a fitness watch.

Key Points:

- There was a statistically significant linkage between the number number of days a user wore their watch during the study period and the average number of steps the user took during the study period. (Spearman rank correlation,  $p < 0.01$ ).
- Broadly, the number of days the fitness watch was worn during the study period was higher for active and lightly active users, although we could not statically confirm this when breaking down specifically by group (Kruskal-Wallis,  $p > 0.05$ ).



```
## [1] "Spearman Rank Correlation for the number of days the watch was used vs the average number of daily steps during the study period"
```

```
##
## Call:
## lm(formula = rank(days_used) ~ rank(avg_steps), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.409  -6.420  -2.894   6.157  15.712
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.6165     3.0446   3.159  0.00352 **
## rank(avg_steps)  0.4343     0.1563   2.780  0.00917 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.547 on 31 degrees of freedom
## Multiple R-squared:  0.1995, Adjusted R-squared:  0.1737
## F-statistic: 7.726 on 1 and 31 DF,  p-value: 0.009167
```

```
## [1] "Kruskal-Wallis test to determine if number of days worn during the study period is users of different activity level"
```

```
## # A tibble: 1 × 6
##   .y.          n statistic    df    p method
## * <chr>    <int>    <dbl> <int> <dbl> <chr>
## 1 days_used    33      2.47     2 0.291 Kruskal-Wallis
```

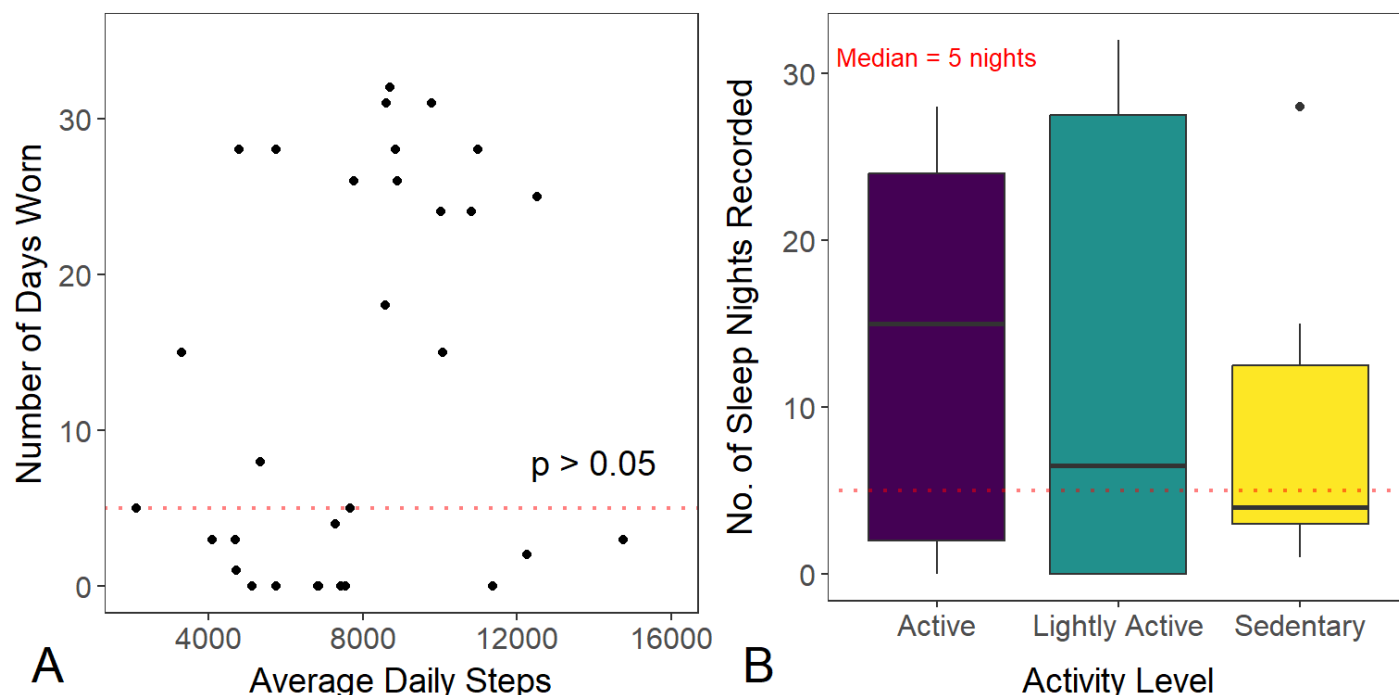
**Question 5:** *Is fitness level an indicator of how often users sleep with their watch?*

**Major Findings:** Patterns in which fitness watches were used to monitor sleep activity were not linked with fitness level, but rather indicate as a whole that monitoring sleep activity is not a high priority for users.

Key Points:

- The median number of nights that users wore their fitness watch during the overnight hours was 5 (out of 31).

- Activity level was not an indicator of how consistent a person wears their fitness watch to bed, but rather there was a broad variation among all groups indicating varying levels of desire among all users in whether or not to use the watch while they sleep.



```
## [1] "Spearman Rank Correlation for the number of sleep days vs the average number of daily steps during the study period"
```

```
##
## Call:
## lm(formula = rank(sleep_days_used) ~ rank(avg_steps), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.1578 -10.0264  0.7238  7.3158  15.2106
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    13.6449     3.3891   4.026 0.00034 ***
## rank(avg_steps)  0.1974     0.1739   1.135 0.26520
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.514 on 31 degrees of freedom
## Multiple R-squared:  0.03988,    Adjusted R-squared:  0.008905
## F-statistic: 1.288 on 1 and 31 DF,  p-value: 0.2652
```

## 7. Conclusions

1. At a high level, this analysis shows some general patterns in fitness behavior. For example, highly active users are more active during lunch and after work hours, however; there is no indication among fitness groups that people would be more active over the weekends as was hypothesized.
2. While it was hypothesized that sedentary users would likely sleep more, there was no difference in sleep activity among users of different activity levels.
3. On average, people who take more daily steps are likely to wear their fitness watch more consistently, however, this was not the case for wearing the fitness watches at night time, which had no predictable trends.

## 8. Recommendations

1. It is highly recommended first and foremost that Bellabeat conduct their own study that would allow them to collect targeted data at specifically links to questions and needs relevant to their client base. This current dataset does not provide insight on sex, race, age, or geographic location and is limited to 33 participants. So we can not conclusively say this data set is represented of their clients and interpretations much be exercised with caution.
2. This analysis clearly indicates that highly active people are more motivated to use fitness watches, and likely support any kind of fitness products. While it may be tempting from a marketing perspective to target people who would like to lose weight, it would seem that targeting people who are already keen on fitness is a better approach.
3. There may be opportunity to encourage less active people to be come more active through notifications on the app (e.g., reminders to get up and move) or providing some kind of incentives become active. People love competition, even if only with themselves. Encouraging people to become healthy will help incentivize them to use their current watches more consistently and more likely to purchase another watch again in the future, or other fitness products.
4. This analyse clearly indicates that sleep monitoring is not a consistently favorable feature, rather instead a very low use feature. This is likely due to fitness watches being bulky and potentially uncomfortable for people when trying to sleep. This highlights potential opportunity for Bellabeat to market their “Leaf” product as a potential alternative to monitor wellness activities, such as sleep, without the need for something large and intrusive on their wrist.