BELLABEAT MARKETING ANALYSIS

Macarena L. Fernandez Carro

31/12/23

CONTENTS

[**INTRODUCTION 3**](#_gyxip46w61b7)

[**TASK 3**](#_6eqlxnhpxx4)

[**DATA SOURCES 4**](#_lz2ouhw1rjtx)

[FITBIT FITNESS TRACKER DATA 4](#_kvda4wk1kmtg)

[**ANALYSIS 4**](#_uo9ylqrkq5if)

[**GENERAL OVERVIEW OF USER USAGE 5**](#_fwdfeny58tmw)

[**POPULAR FEATURES 5**](#_788ovadjb07o)

[**USER ENGAGEMENT 6**](#_9nqcee5escrs)

[**ACTIVITY PATTERNS DEPENDING ON THE DAY OF THE WEEK 7**](#_jyfdmjextb3g)

[**ACTIVITY INTENSITY BY DAY OF THE WEEK 7**](#_ae9g54t55xy0)

[**ACTIVITY METRIC BY DAY OF THE WEEK 8**](#_o2h5czi8g2gf)

[**PEAK TIMES FOR INTENSE ACTIVITIES 8**](#_a0zokwe51gfu)

[**CORRELATIONS 9**](#_boumnjb81z5)

[CORRELATIONS BASED AT THE MINUTE LEVEL 10](#_n3vdt2p6cpsd)

[**SLEEP ANALYSIS 11**](#_pnql19oq687k)

[USERS’ SLEEP PATTERNS 11](#_quw90dk5sy64)

[MOST PREVALENT SLEEP TIMES 12](#_clceqm90r03a)

[SLEEP DURATION BY USER 13](#_ms1gx5w45nsp)

[SLEEP STAGES 14](#_gsxvu4mxvhl2)

[**CORRELATION BETWEEN METs AND MEAN HEART RATE 15**](#_lsfimvz4bpmg)

[**CORRELATION BETWEEN STEPS AND HEART RATE 15**](#_qpssfq6eng38)

[**RELATIONSHIP BETWEEN INTENSITY LEVEL AND HEART RATE 16**](#_qcgpadkiofpp)

[**BMI AND WEIGHT DATA ANALYSIS 18**](#_b397kjoek30p)

[BMI AND ACTIVITY 18](#_i243skw4q2h9)

[BMI AND HEART RATE 19](#_35y61x8t972q)

[**KEY INSIGHTS 20**](#_bkeqc3bvsez1)

[DATA LIMITATIONS 21](#_r6fspls1bsmc)

[**RECOMMENDATIONS 21**](#_tjmzkln1y3wl)

# INTRODUCTION

Bellabeat has been a cutting edge high-tech company producing health-focused smart products for females since 2013. By 2016, Bellabeat had opened offices all over the world, allowing women to empower and inspire themselves with knowledge about their own health and habits.

The co-founders Urška Sršen (CCO) and Sando Mur are determined to increase the success of the company and believe that non-Bellabeat data analysis can provide insights into further growth opportunities.

Key stakeholders:

* Urška Sršen: Bellabeat co founder and Chief Creative Officer
* Sando Mur: Mathematician and Bellabeat cofounder; key member of the Bellabeat executive team
* Bellabeat marketing analytics team

# TASK

To analyze FitBit fitness tracker data in depth and gain insights into how their consumers are using the app in order to discover trends to develop a Bellabeat marketing strategy.

The main questions are:

* What are some trends in smart device usage?
* How could these trends apply to Bellabeat customers?
* How could these treds help influence Bellabeat marketing strategy?

# DATA SOURCES

## FITBIT FITNESS TRACKER DATA

The data was obtained from kaggle under a CC0: Public Domain License. The repository belongs to MÖBIUS. It was collected from an Amazon Mechanical Turk distributed survey during December 2016. The survey was completed by 30 FitBit users.

The data consists of 18 cvs files, one for each measurement. The data is organized by export session ID (Column A) and timestamp (Column B).

*LIMITATIONS*

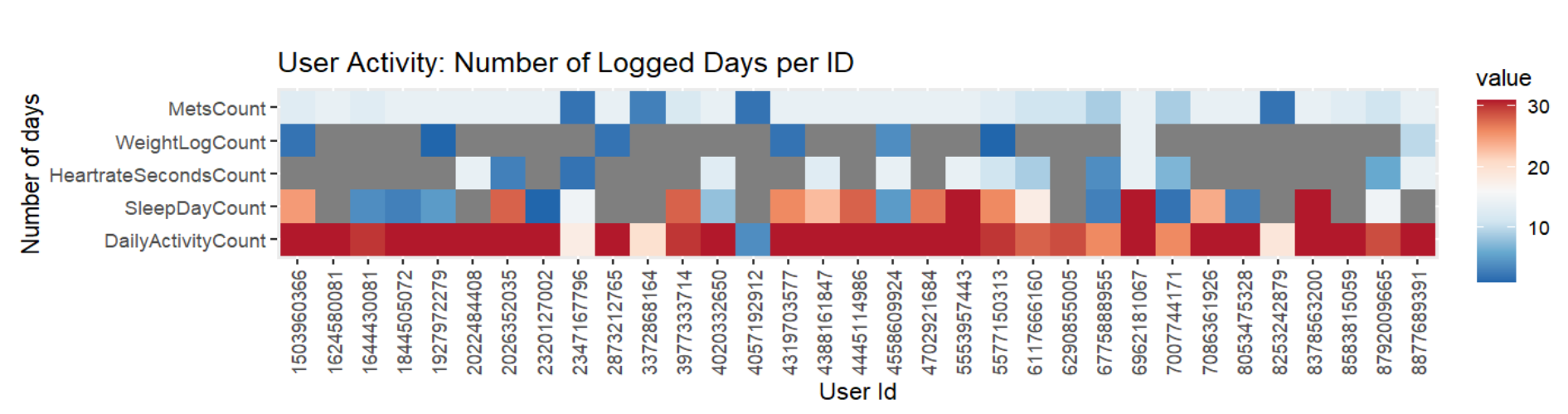
* The age of the data. Consumer habits might have changed over time and the data could be outdated
* Sample size. 30 users is not representative of the entire female population
* Data collection. Survey answers are not as robust and could put into question the validity and accuracy of the data
* Output variation. The data output varies between different FitBit devices, user preferences and user behavior.

After performing a ROCCC analysis of the dataset, this is considered to be of LOW quality and any recommendations based on this data should be made with caution.

# ANALYSIS

### GENERAL OVERVIEW OF USER USAGE

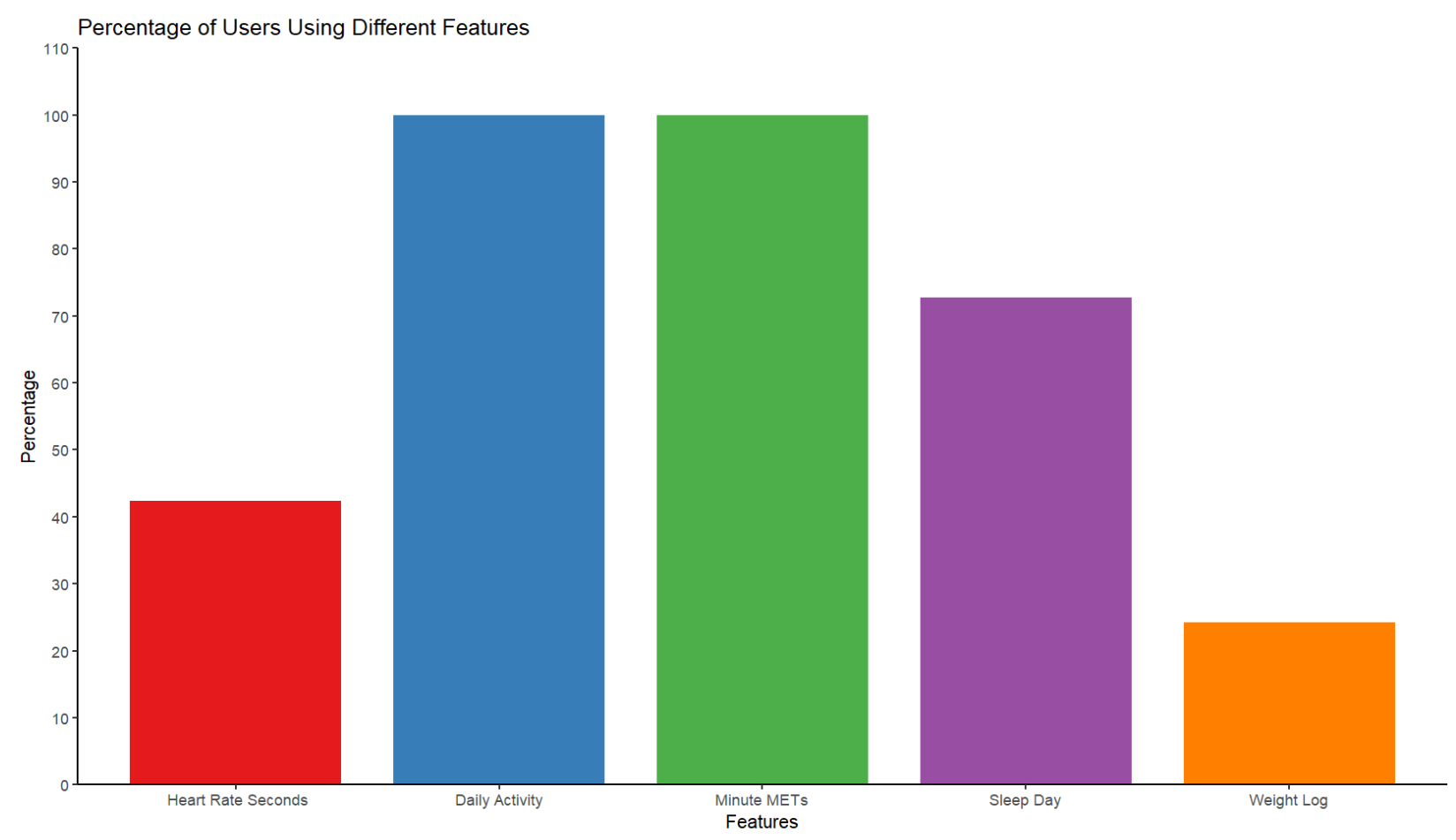
The first step of the data analysis was to see the input values for each user in each category.



This analysis provides us with a concise overview of user activities and the user engagement with particular features of the app. The above heatmap represents an overview of user activities, considering the total number of days each user engaged with the features. This way, we can showcase that some users are more proactive than others in utilizing certain features.

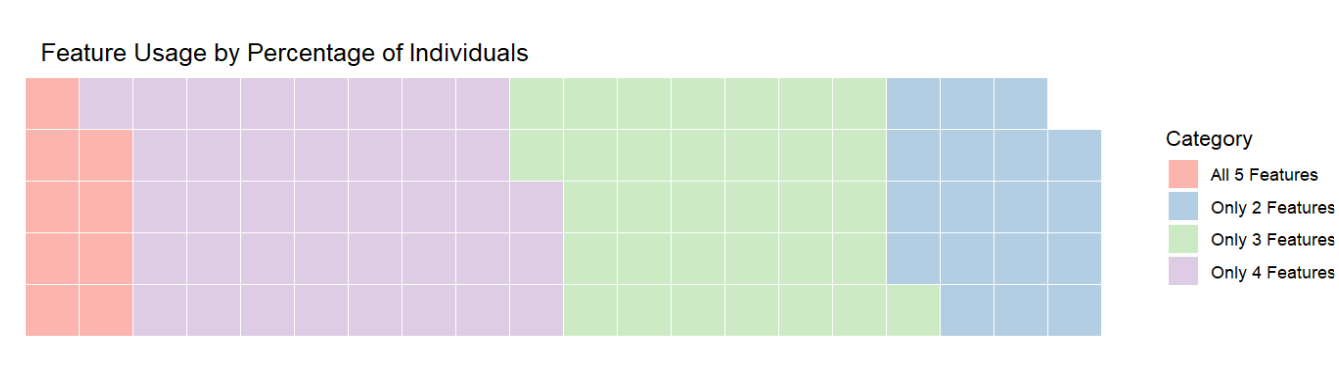
The results show that some users are more proactive than others in engaging with the app and that logging in weight values was the least popular feature. This was the least popular feature of the app. Following this functionality, logging in heart rate values and sleep values were the next least popular functions. The data suggests that the primary focus of the users was to monitor other activities like intense activity, calories burned, sedentary activities among others.

### POPULAR FEATURES



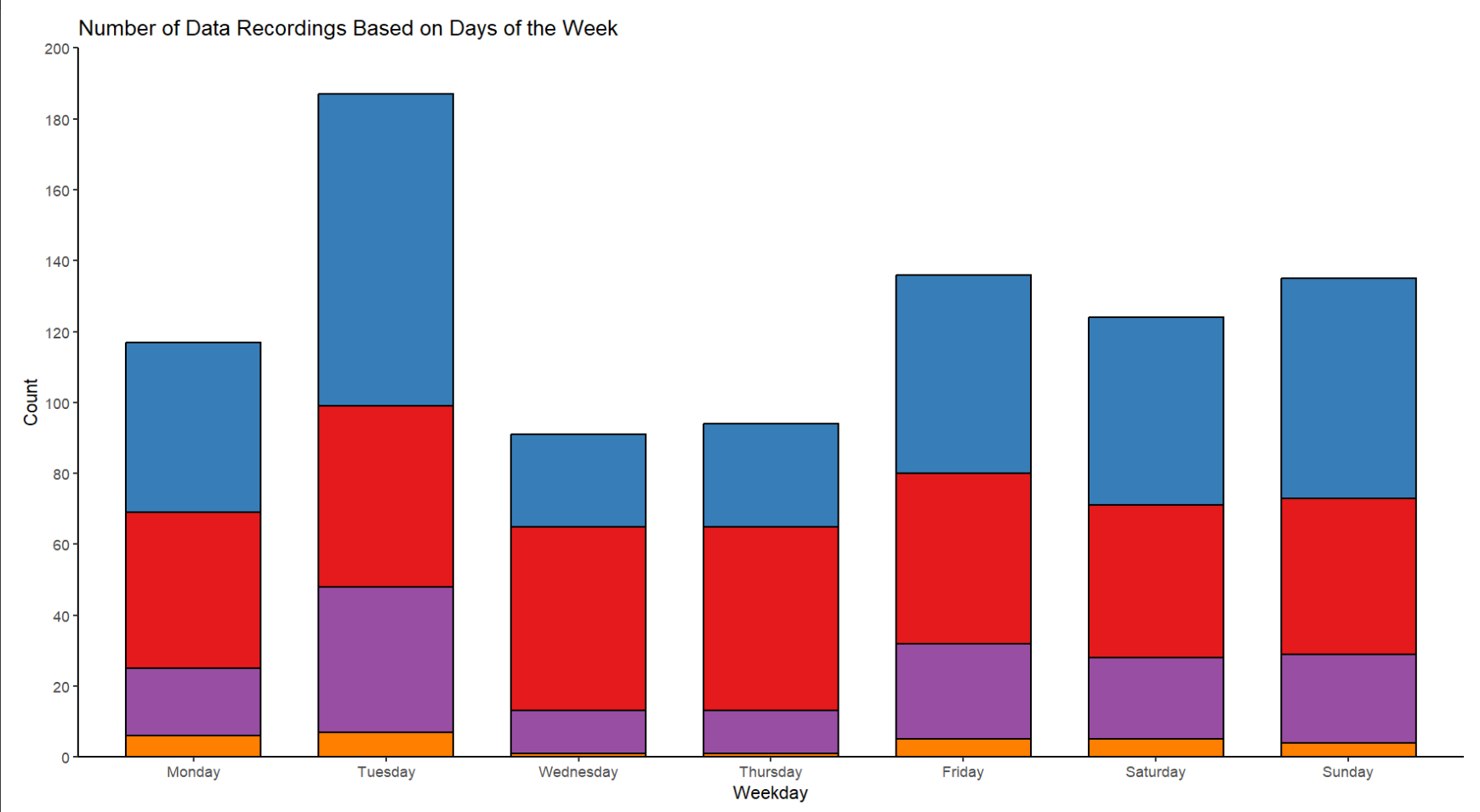
Based on the bar chart above, we can observe that the activity and MET features were used by 100% of the users. Furthermore, 72.7% of users logged data using the sleep tracking feature. Less than 50% of the users recorded their heart rates and only 24.2% of the users engaged with the weight log.

### USER ENGAGEMENT



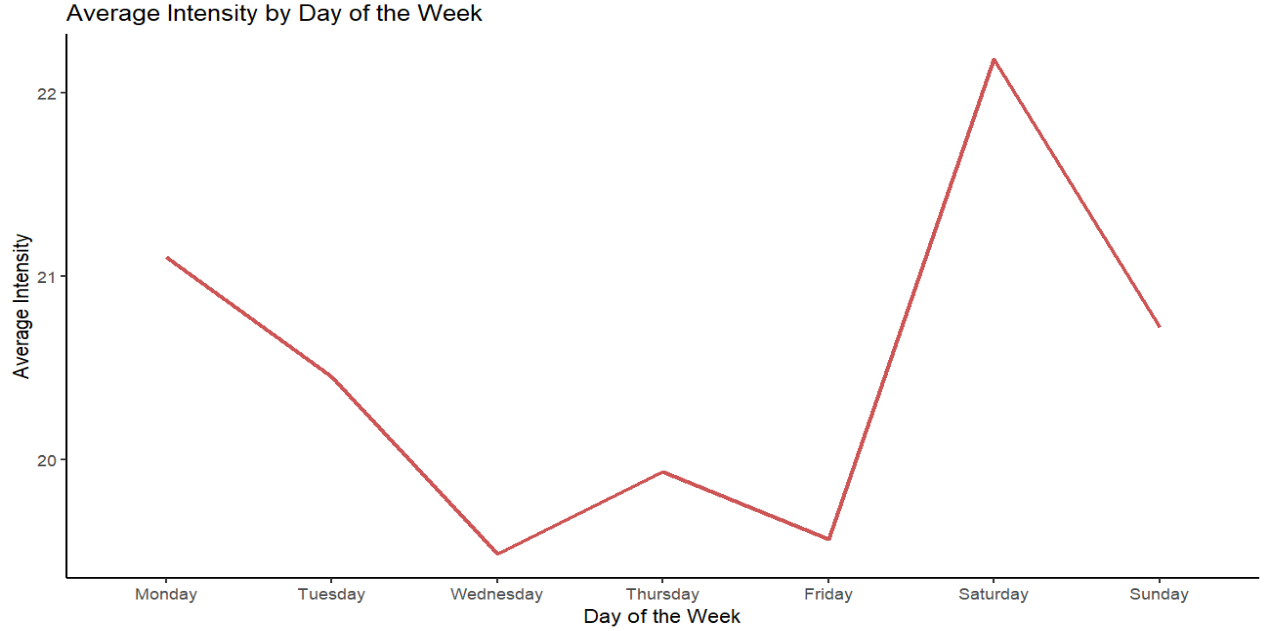
The waffle chart above allows us to visualize how users engaged with the different features of the app as a whole. It is apparent that less than 10% of users used all 5 features, followed by 28% using only 2 features. The most popular option, as showcased by approximately 40% of users, was to engage with 4 features of the app.

### ACTIVITY PATTERNS DEPENDING ON THE DAY OF THE WEEK



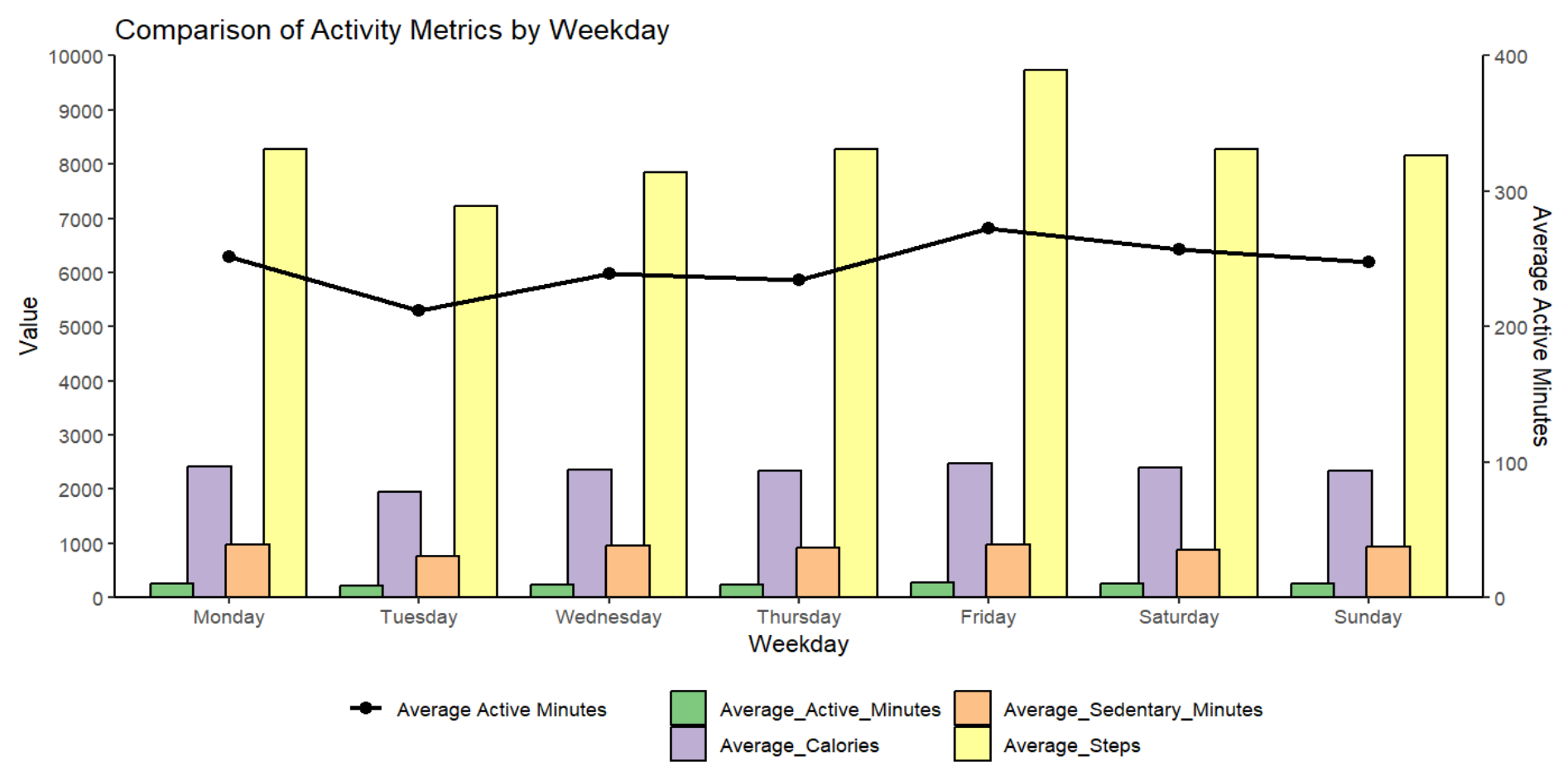
Based on the data shown on the stack bar graph, we can appreciate that users tend to log a higher amount of data on Tuesdays with Wednesday and Tuesday being the least popular days.

### ACTIVITY INTENSITY BY DAY OF THE WEEK



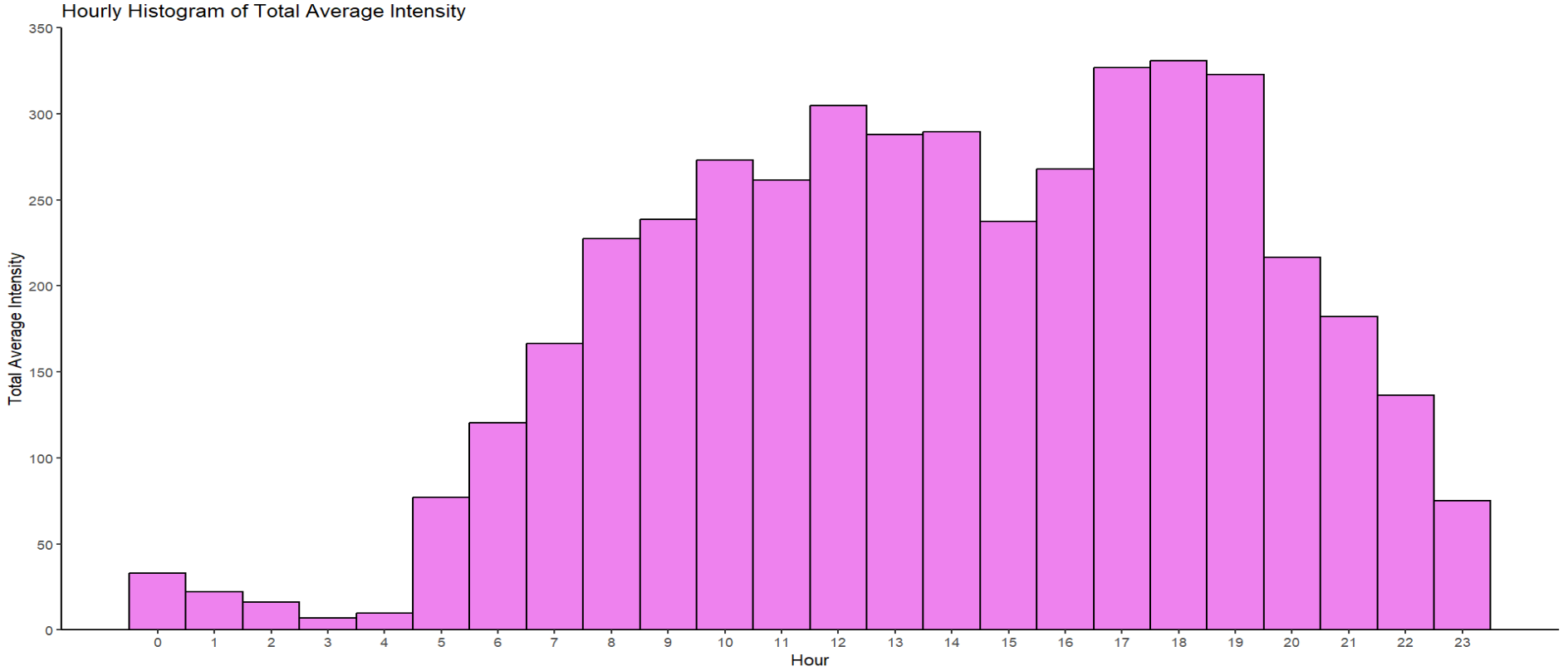
This time series graph allows us to see the pattern of users engaging in intense activities depending on the weekday. As we can observe, users then engage in more intense activities on Saturdays while on Wednesdays and Fridays the activities are less intense.

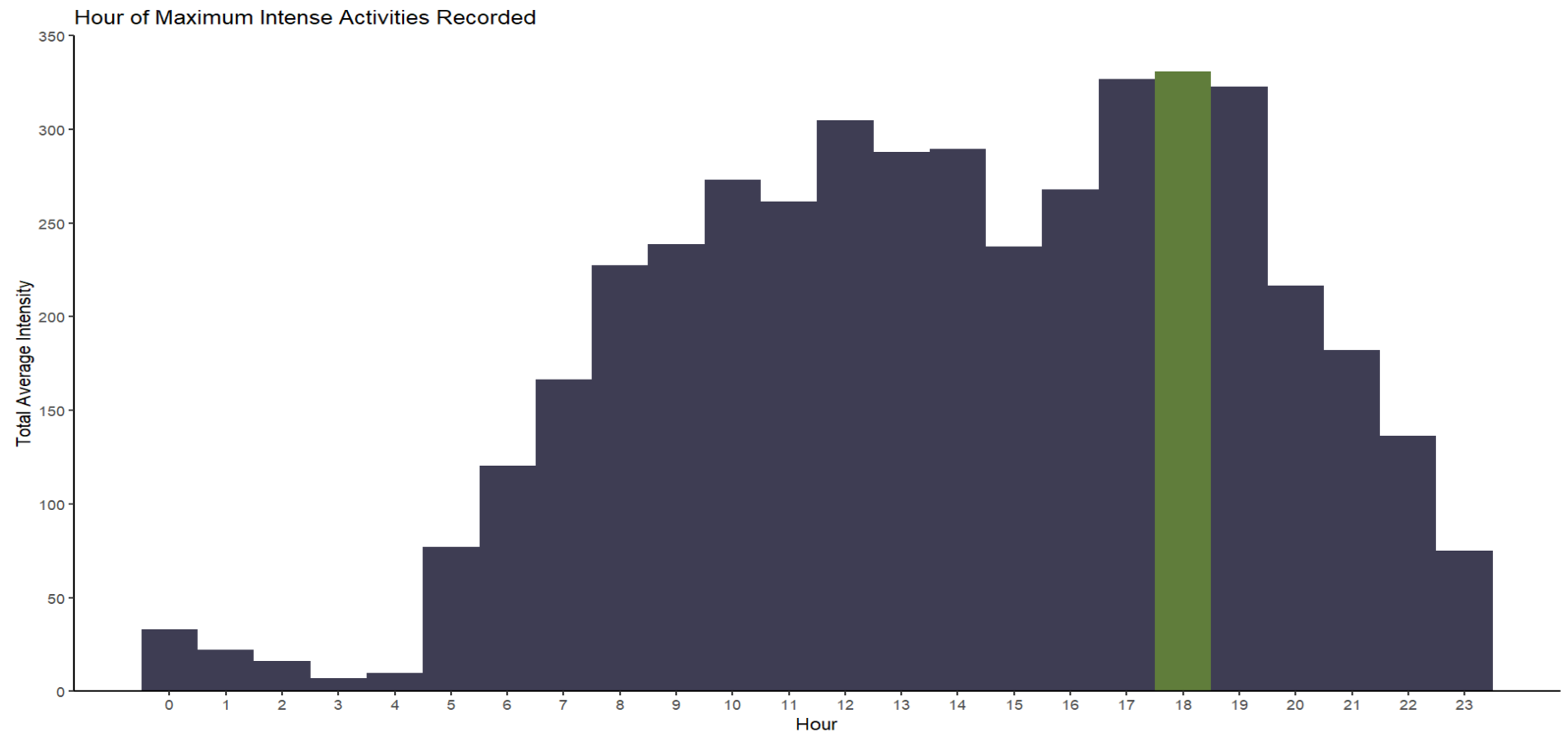
### ACTIVITY METRIC BY DAY OF THE WEEK



The combination chart above reveals a notable upward trend in various activity metrics—active minutes, average steps taken, and sedentary minutes—specifically on Fridays followed by Mondays.It also exhibits a downward trend in those metrics on Tuesdays.

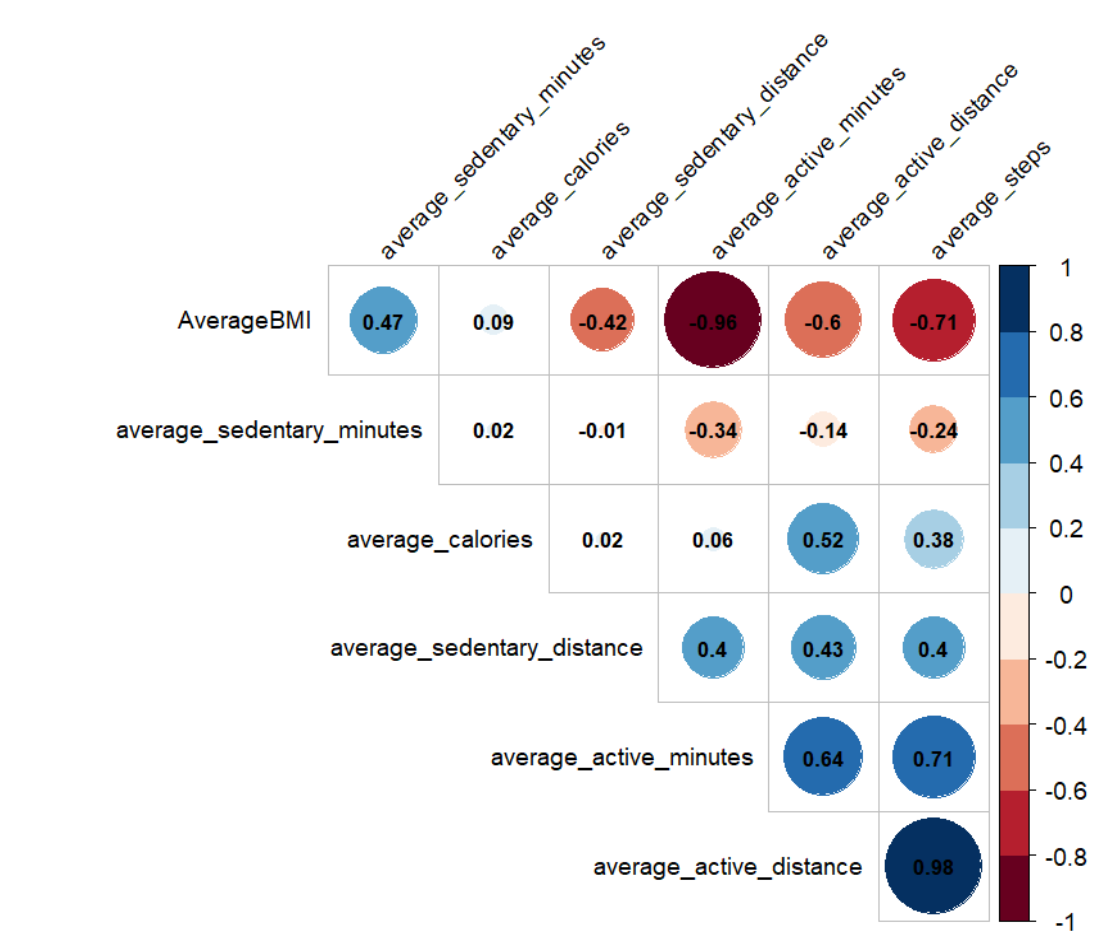
### PEAK TIMES FOR INTENSE ACTIVITIES





These histograms reveal a clear pattern in more intense activities recorded in the afternoon/late afternoon, with the maximum total average intensity recorded at 18:00. However, a very interesting second pic is observed at around 12:00. This pattern might indicate that some users experience mid afternoon fatigue or like to wind up after work.

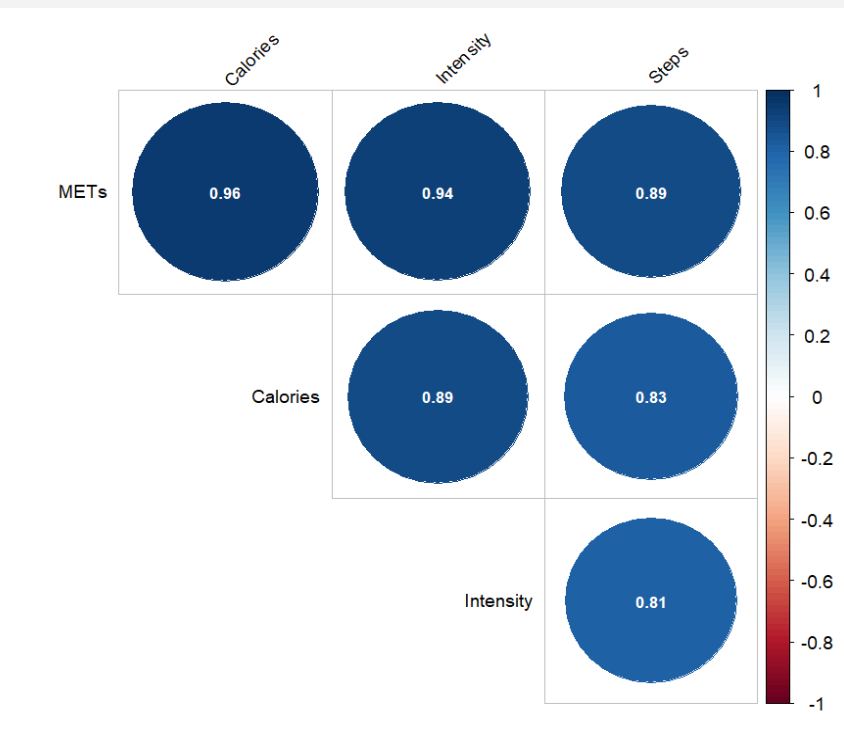
### CORRELATIONS



The correlogram reveals notable patterns in the relationships between various variables. Some variables exhibit strong correlations, while others demonstrate only moderate, slight, or almost no correlations. Specifically, we observe a strong negative correlation of -0.96 between average active minutes and average BMI. Additionally, there exists a moderate negative correlation of -0.71 between average steps and average BMI. On the other hand, average calories and average active distance display a moderate positive correlation with a coefficient of 0.52

#### CORRELATIONS BASED AT THE MINUTE LEVEL

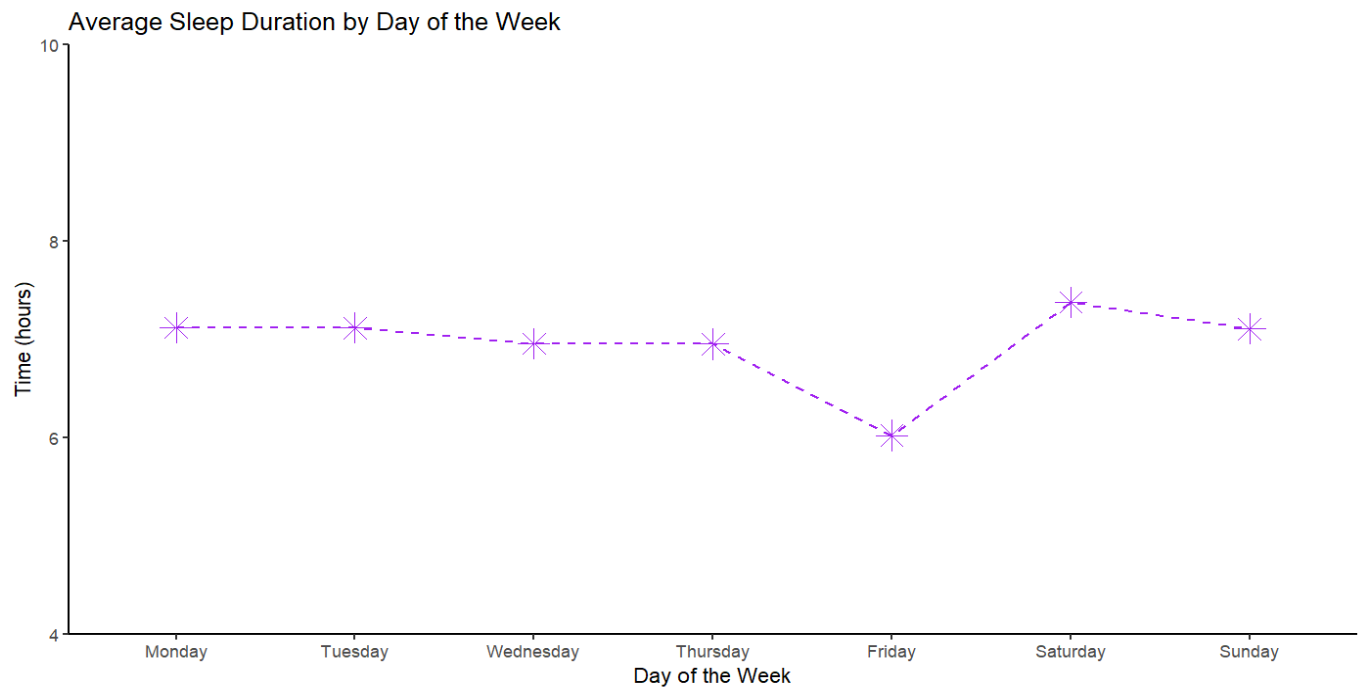
As our earlier correlation analysis was based on daily averages, which may not provide a comprehensive insight,we will explore correlations between various datasets recorded at the minute level for each individual. This in-depth analysis will provide a more detailed understanding of the relationships between various variables, enabling us to create a more precise and comprehensive picture of their correlations.



The correlogram suggest notable patterns in the relationships among various variables. We see a strong correlations between Intensity and Steps(0.81),Intensity and METs(0.94),Intensity and Calories(0.89), Steps and Calories(0.83),Steps and METs(0.89),METs and Calories(0.96). This findings may imply that higher intensity levels are associated with increased steps, METs( metabolic equivalent), and calorie loss.

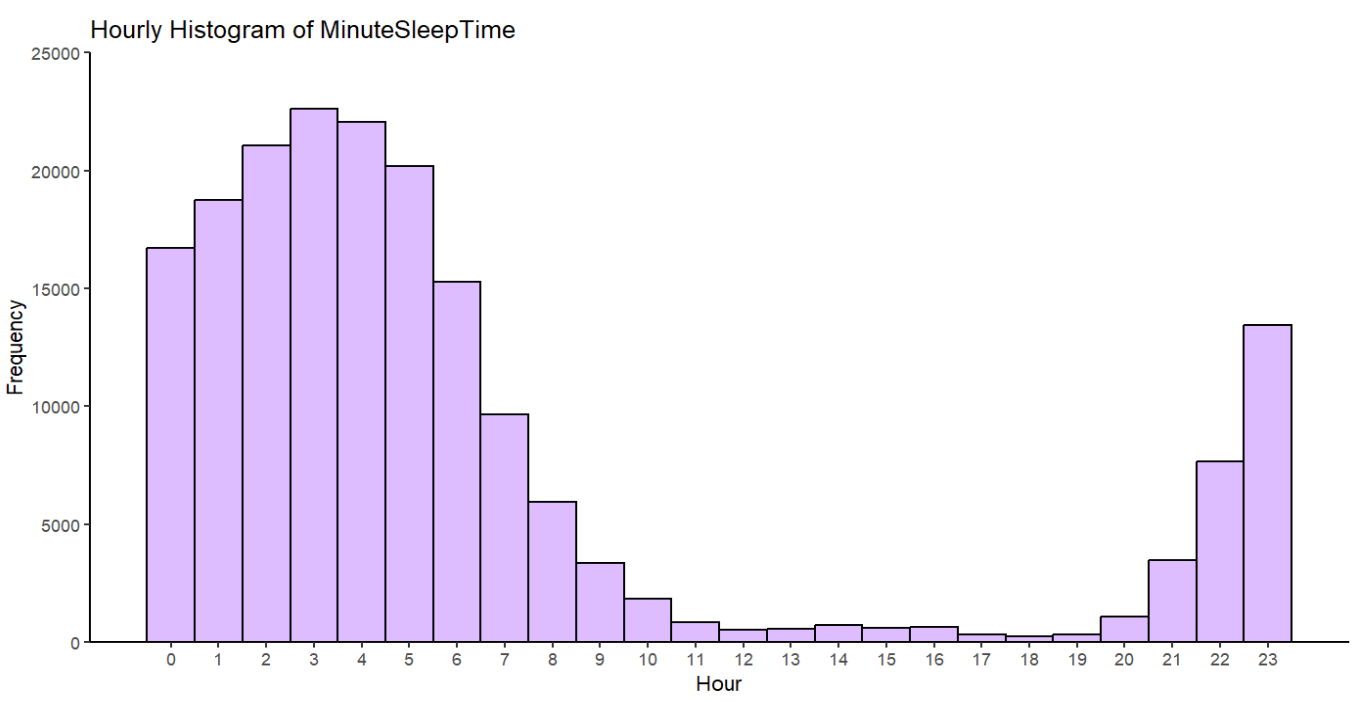
### SLEEP ANALYSIS

#### USERS’ SLEEP PATTERNS



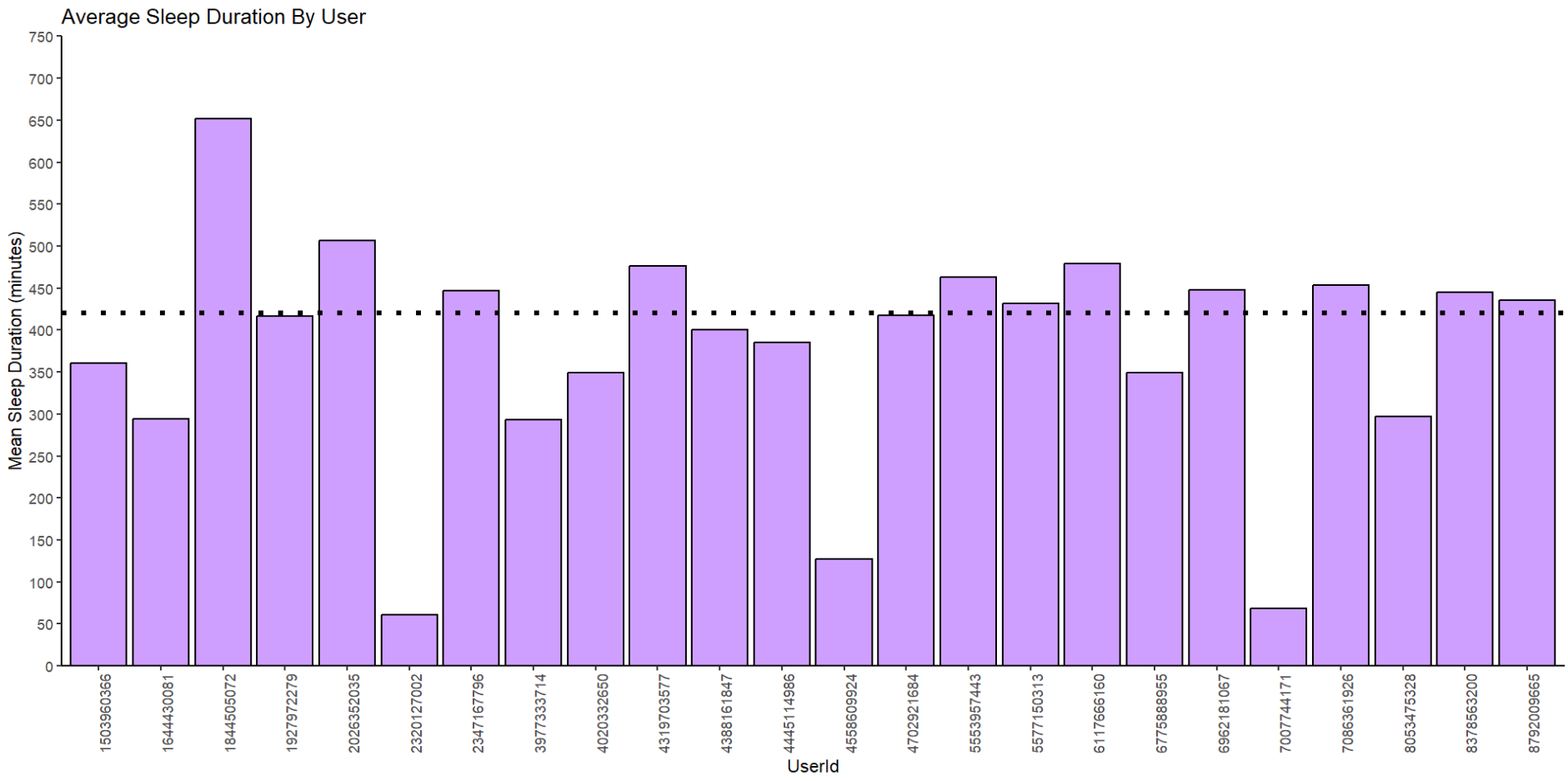
The line graph above suggests that even though user’s sleep more on Saturdays, they sleep much less on Friday nights.

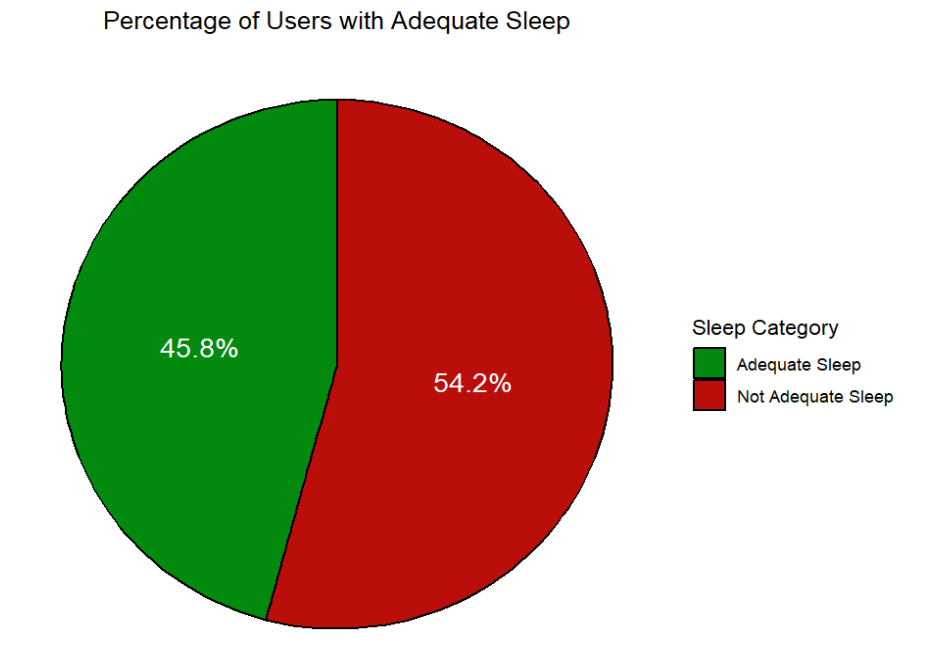
#### MOST PREVALENT SLEEP TIMES



The histogram indicates that the most popular time for sleeping is from 11:00 PM to 7:00 AM, with a concentration of data particularly between 1:00 AM and 6:00 AM.

#### SLEEP DURATION BY USER

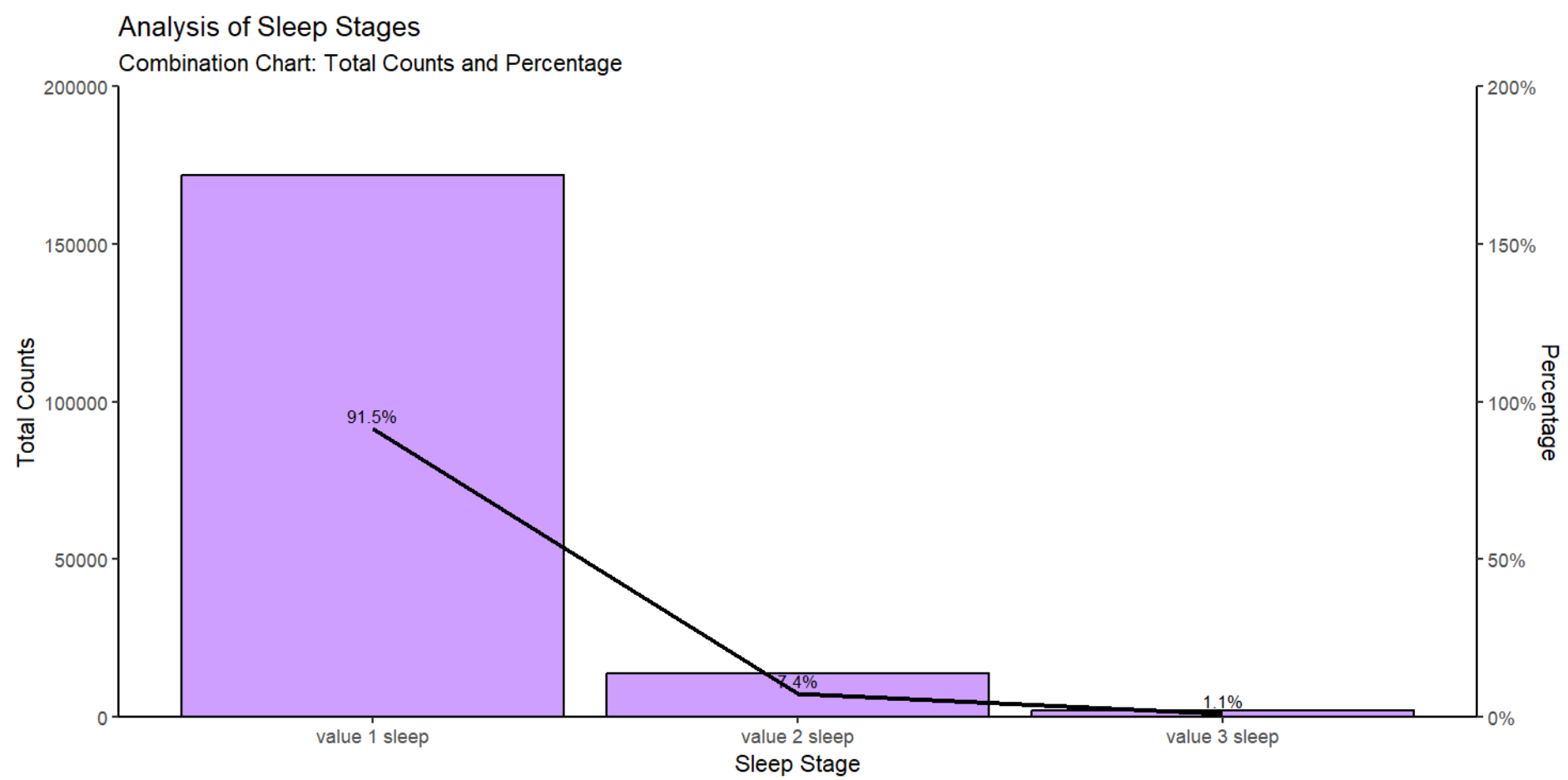




For the purpose of the analysis, 7 hours of sleep (420 minutes) is considered as the standard minimum, considering the absence of individual-specific details such as age and sleep pathologies.

The bar chart shows that some users have a very low average if recorded sleep time. This highlights the potential lack of compliance with this feature. As seen from the pie chart, out of the 24 individuals who recorded their daily sleep duration, 45.8% maintained an average sleep duration aligning with the recommended 7 hours.

#### SLEEP STAGES

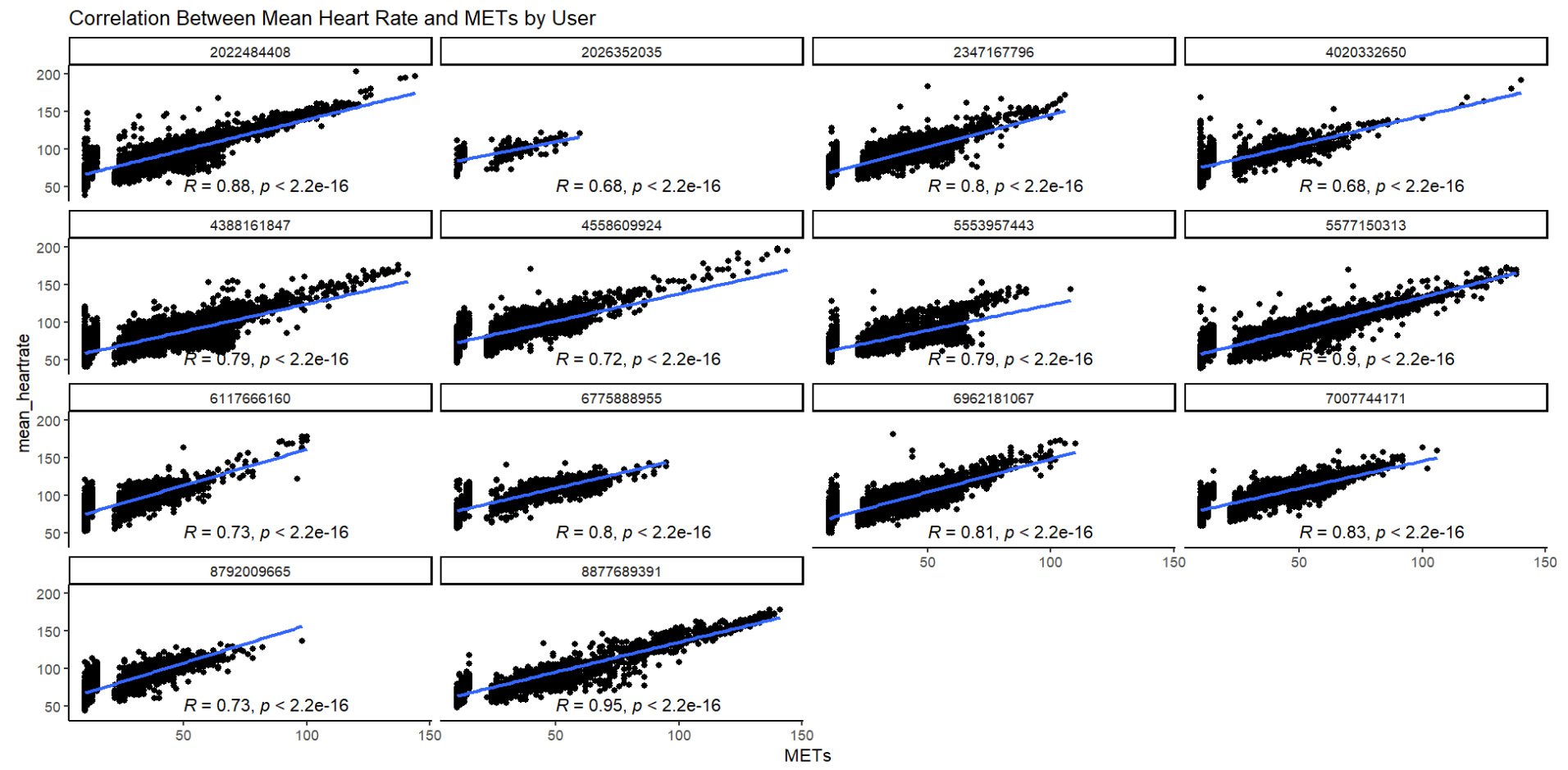


The categorization of values is determined based on the value scores found in the minute\_sleep table. Upon analyzing and subsequently visualizing the total sleep counts in minutes for each sleep category, we observed that sleep with a value score of 1 occurred 91.5% of the time, followed by value 2 sleep at 7.4%, and value 3 sleep at 1.1%.

The Fitbit website provides insights into typical sleep stages, indicating that light sleep generally constitutes about 50 to 60 percent or more of one's nightly sleep, deep sleep ranges from 10 to 25 percent (depending on age), and REM sleep makes up approximately 20 to 25 percent of the entire night's sleep.

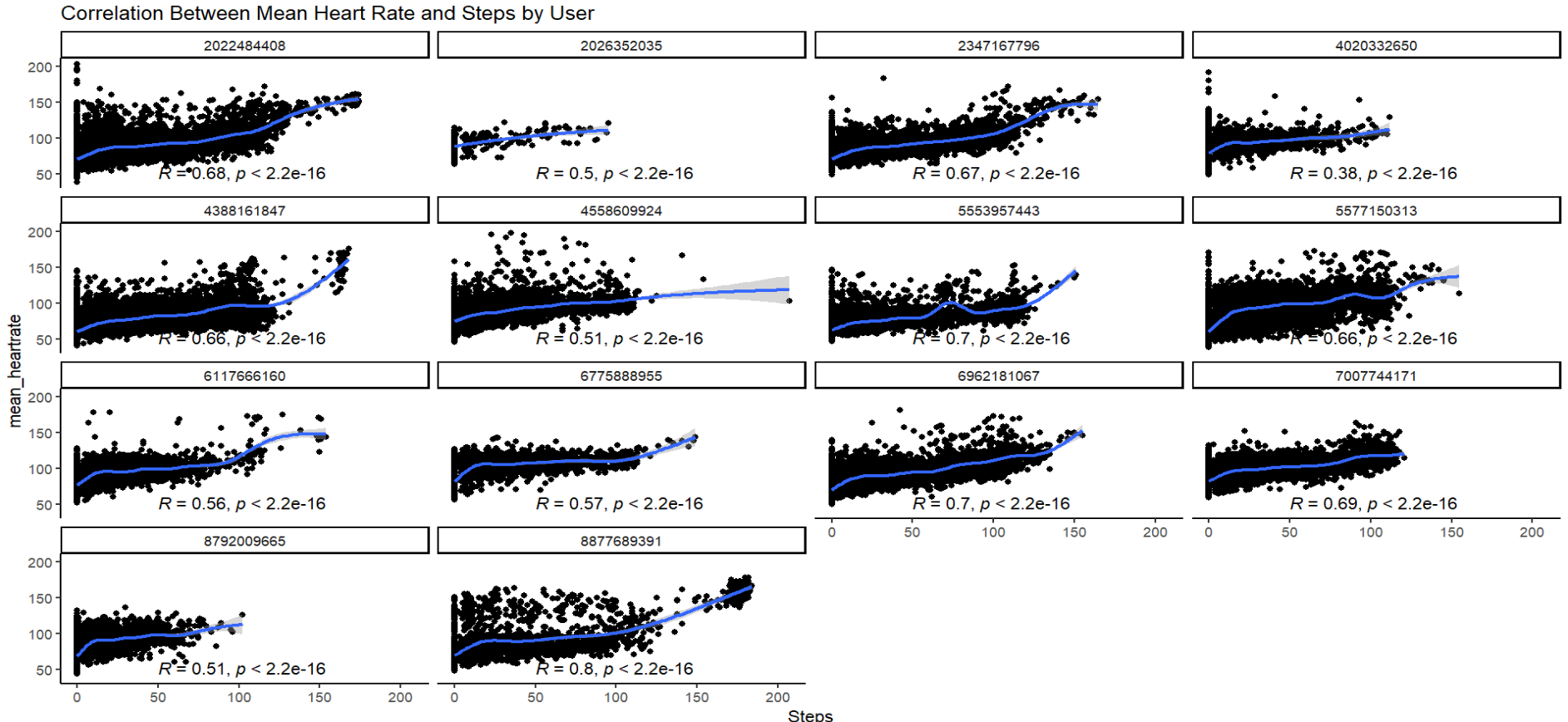
However, due to lack of detailed information on what these value scores precisely correspond to and their relationship with specific sleep stages (such as light sleep, deep sleep, or REM) or other factors like heart rate, drawing meaningful insights becomes challenging. The lack of clarity regarding the nature of these categories makes it difficult to draw any conclusions from our findings, and additional context is needed to interpret the significance of the observed sleep patterns.

### CORRELATION BETWEEN METs AND MEAN HEART RATE



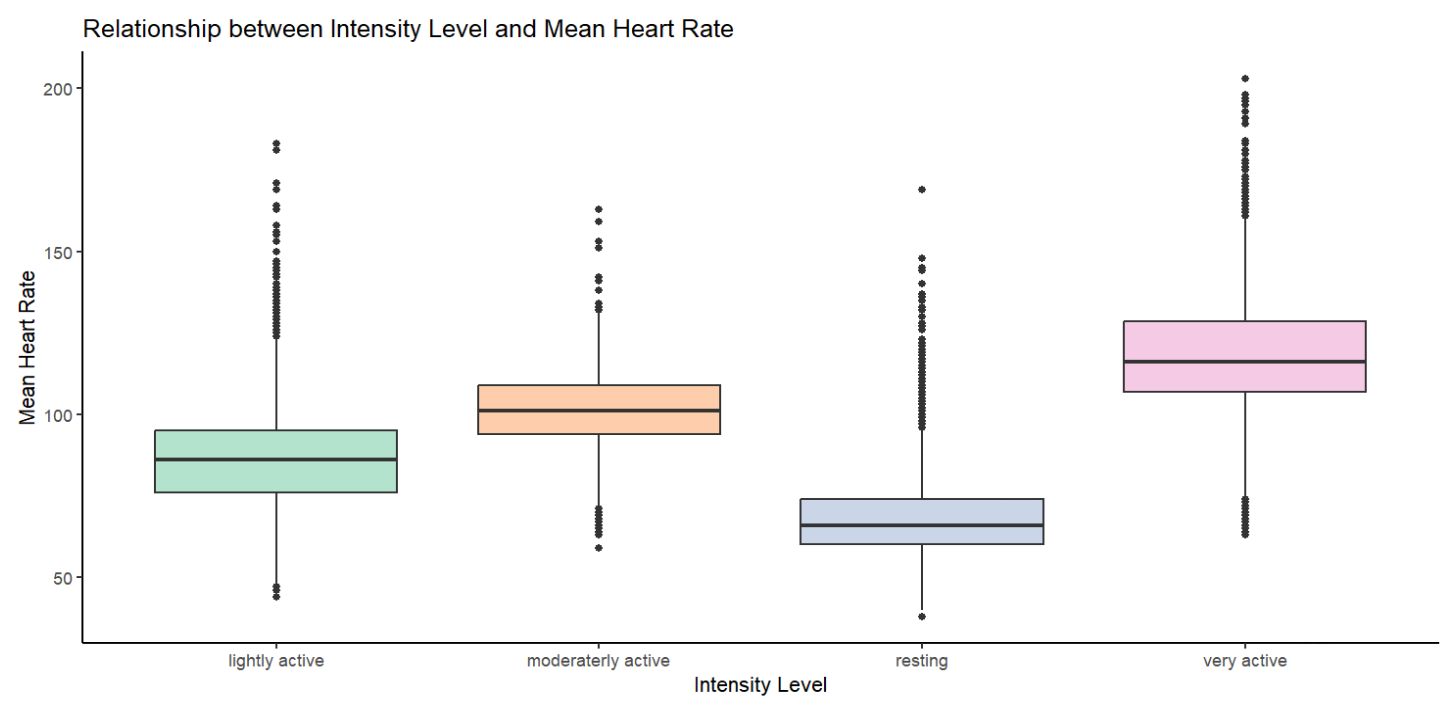
These results showcase a consistent positive linear relationship for each individual. The total Pearson Correlation Coefficient value of 0.7503 further indicates a strong positive relationship between the analyzed variables.

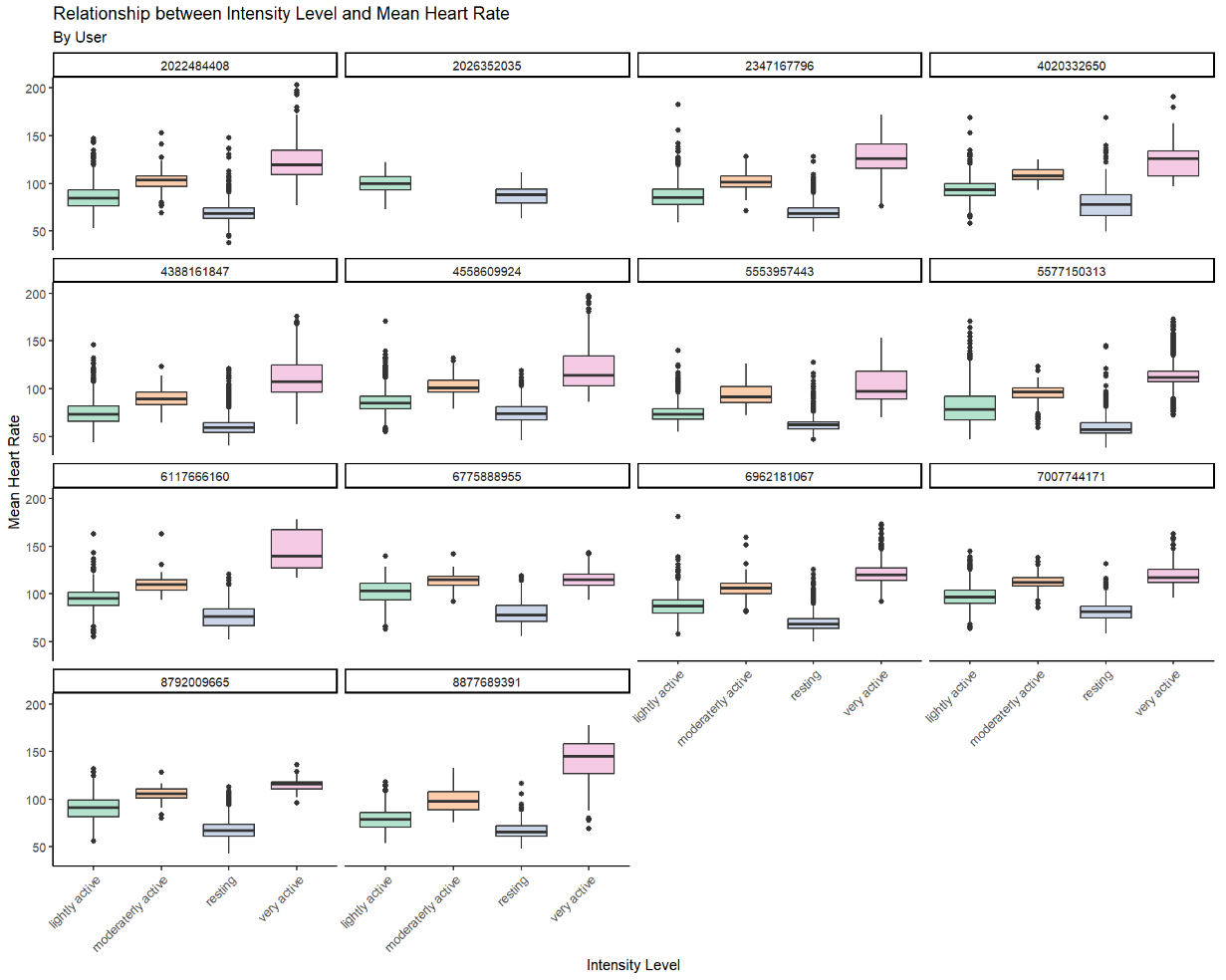
### CORRELATION BETWEEN STEPS AND HEART RATE



In analyzing the scatter plot for each individual and examining the relationship between the number of steps and the mean heart rate, a discernible pattern emerged. Notably, as the number of steps increased, there was a consistent and observable slight upward trend in the heart rate. This is further corroborated by a Person Correlation Coefficient of 0.6016.

### RELATIONSHIP BETWEEN INTENSITY LEVEL AND HEART RATE

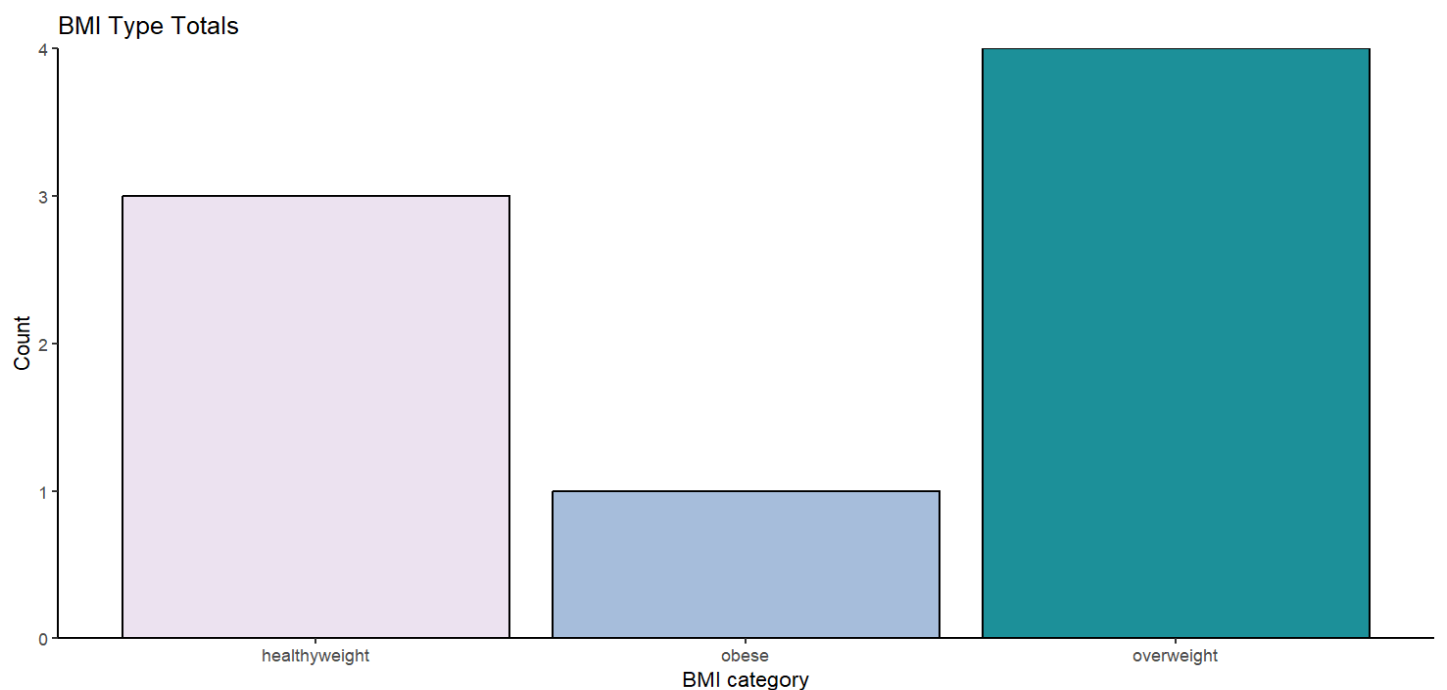




While comparing the intensity level with the mean heart rate for each individual, a consistent pattern was revealed. The median of resting heart rate was found to be lower than when individuals were lightly, moderately, or highly active. Furthermore, depending on the level of intensity, the heart rate exhibited changes, indicating that the higher the intensity level, the higher the heart rate. This pattern was particularly pronounced during highly intensive activities, where significantly higher heart rates were observed.

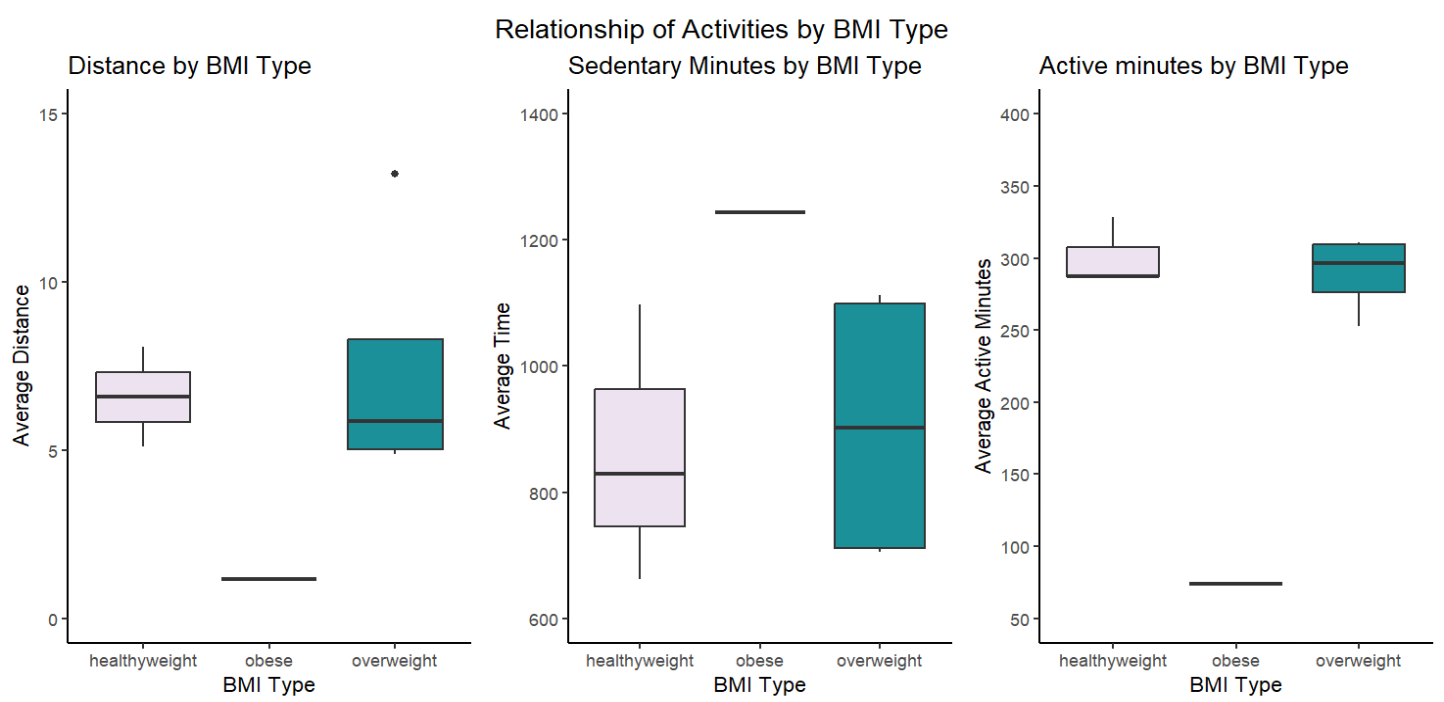
*For all the relationships that we analyzed above, it's important to emphasize that correlation does not imply causation, and further investigation would be needed to establish any causal relationship between these variables.*

### BMI AND WEIGHT DATA ANALYSIS



Of the 8 users who recorded their weight data, 4 fall into the overweight category, 3 belong to the healthy weight group, and 1 individual is categorized as obese, exhibiting an unusually high BMI. Assessing the reliability of this data requires a double check.

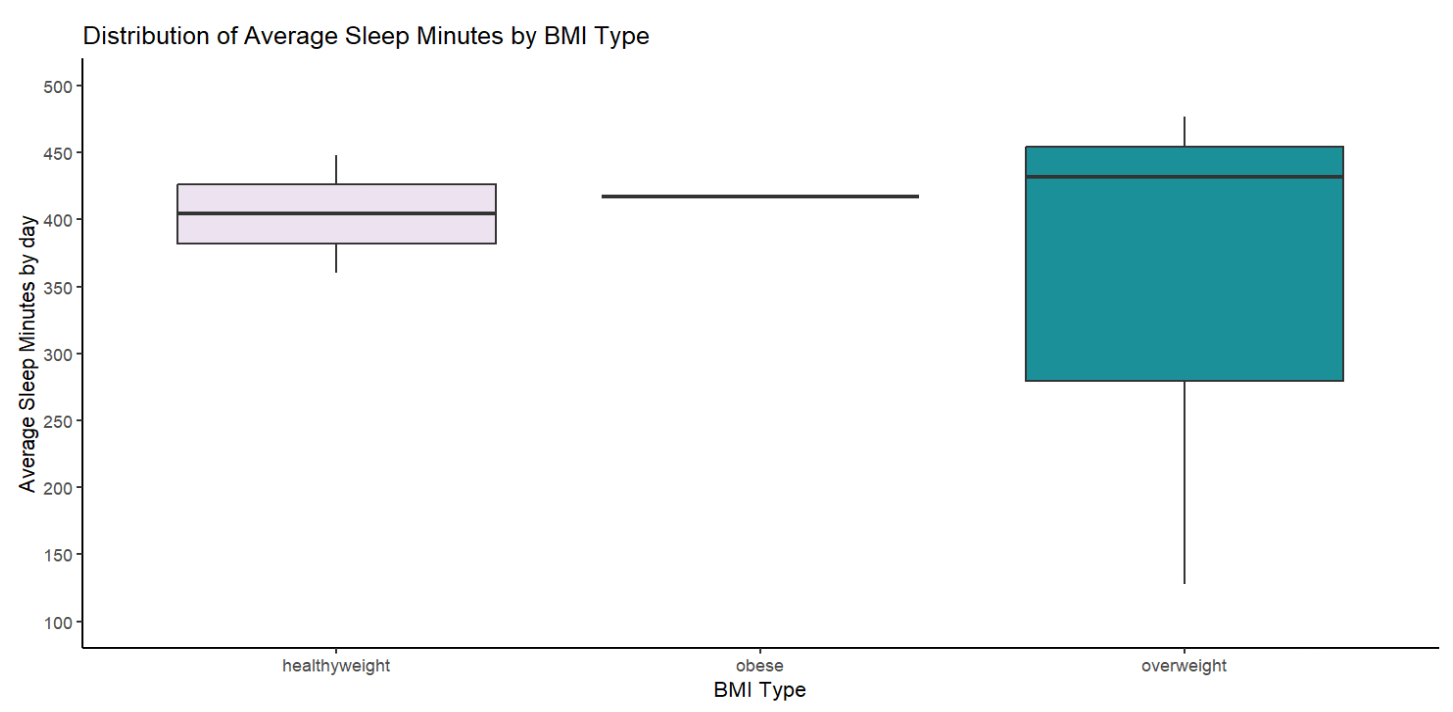
#### BMI AND ACTIVITY



The boxplots above showcase notable distinctions in average active distance, sedentary distance, and active minutes between individuals classified as overweight and those with a healthy weight. Overweight individuals demonstrate a lower median value of average active distance, with data concentrated closer to the first quartile. In contrast, their healthy weight counterparts exhibit a higher median value, indicating a concentration of higher values that extends slightly toward the upper quartile.

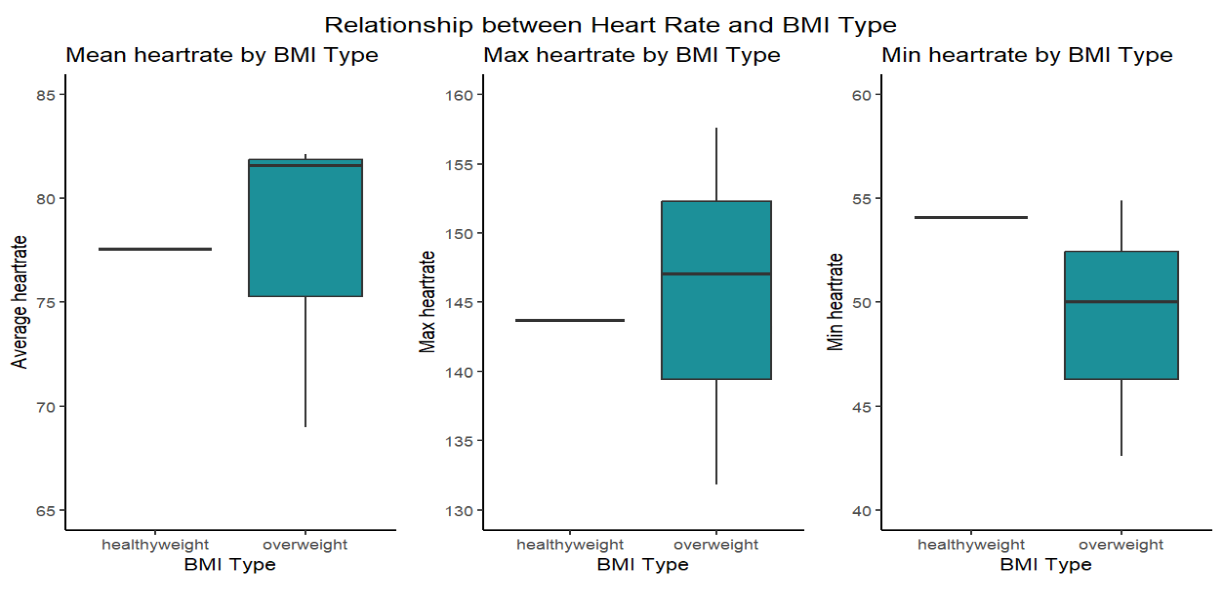
Conversely, overweight individuals display a higher median value of average sedentary distance compared to healthy weight individuals. The data for the healthy weight group is concentrated in the lower quartile, suggesting less variability in sedentary distance among these individuals.

The median values for average active minutes among overweight individuals are nearly identical to those for healthy weight individuals.It's important to note that insights from the obese group are completely excluded due to extremely limited sample size which equals only one individual.



In analyzing the boxplot that compares the average sleep minutes per day among different BMI groups, it is apparent that the overweight group exhibits a higher median sleep duration compared to the healthy weight group. Furthermore, the data for the overweight group is more concentrated in the upper quartile, indicating a prevalence of higher sleep duration values within this group. However, it is important to note that the presence of a very limited number of samples restricts us from drawing definitive insights. So, we should proceed with caution in interpreting these observations as a true representation of the overall trend.

#### BMI AND HEART RATE



In our effort to explore the correlation between BMI type and average heart rate, we merged two datasets, namely 'average\_BMI\_individual' and 'heartrate\_day\_average' for each individual. Unfortunately, only four individuals had recorded both sets of information, with only one person falling into the healthy weight group, while the remaining three individuals were classified as overweight. Despite the limited sample size, we proceeded to create boxplots illustrating the relationship between the average of all recorded heart rates, the average of maximum heart rates, and the average of minimum heart rates for each group.

The boxplots indicate that the overweight group tends to have a higher average heart rate compared to the healthy weight group. Additionally, the median value for the average maximum heart rate is higher in the overweight group compared to the healthy weight group. Conversely, the median for the average minimum heart rate is higher for individuals in the healthy weight category. Given the small sample size, drawing definitive comparisons may be unreliable, and therefore, these insights should be interpreted cautiously.

### KEY INSIGHTS

* Activity and METs tracking are the most popular features whereas weight tracking is the least popular feature.
* A majority of users engage with multiple features.
* Users log more intense activities on Saturdays, with variations of activity intensity throughout the week.,
* Users perform more intense activities between 12:00 - 13:00 and 17:00 - 18:00.
* Users tend to sleep more on Saturdays, most commonly between 23:00 - 7:00.
* 45.8% of users maintain, on average, a sleep duration greater than the recommended 7 hours.
* High intensity levels are associated with higher value of steps, METs, and calories.
* A strong negative correlation is observed between average active minutes and BMI, indicating the a higher BMI is associated with decreased activity minutes.
* Weight data must be taken with precaution as very few users engage with this feature.

#### DATA LIMITATIONS

The minute sleep data records timestamps with seconds in the :30 format, while other minute datasets use the standard :00 format for seconds. Due to this difference in timestamp formatting, direct comparisons with other minute-level dataset (mean heart rate per minute), couldn't be conducted. It's crucial to have an universal and standardized recording format for performing analysis across different datasets.

# RECOMMENDATIONS

* Bellabeat should encourage users to utilize all features to comprehensively track their health.
* Bellabeat should incorporate personalized recommendations by enabling users to input additional personal data. This data could then be used to tailor insights and suggestions, helping users better manage their overall wellness based on individual characteristics.
* Bellabeat should integrate features to set a personal goal and develop better personalized notification systems to keep users updated about their progress.