Data Mining Project: Fitness Trends
BUAN 314 Fall 2024
Dr. Steven Levkoff
Lucia Berni, Natalia Garcia, Pablo Saez

Executive Summary

This report provides an analysis of 45618 observations across multiple FitBit users. The datasets we used track the user's daily activity levels such as minute-level output in physical activity, total number of steps, heart rate, sleep monitoring. The purpose of our analysis is to view health and fitness trends, promote healthy lifestyle choices and explore relationships between the key variables selected.

The methods of analysis include exploratory data analysis using SQL queries and ggplot2 visualizations in R. We found multiple datasets on Kaggle that followed the same users and combined the data. Out of all the variables, we decided to analyze the following in more depth: activity, calories, intensities, total steps, distances, sleep and weight.

The analysis utilized six datasets:

1. Daily Activity: Steps, distance, active/sedentary minutes, and calories.
2. Hourly Calories: Calories burned on an hourly basis.
3. Hourly Intensities: Activity intensities throughout the day.
4. Sleep Data: Total sleep duration, efficiency, and records.
5. Weight Log: Weight, BMI, and body metrics.
6. Heart Rate: Heart rate observations.

Datasets were merged using the common Id column and formatted Date values.

As we wanted to analyze the variables that are actually positively correlated with a healthy and active lifestyle, we chose to analyze the three broader concepts of sleep, activity levels and weight. The key questions we explored included the relationship between step count and calories burned, the allocation of intensity in active minutes throughout the day, relationships between sleep quality, efficiency and duration with calories burned and weight… Our goal is to better understand patterns of physical activity, sleep efficiency, and their impact on fitness outcomes.

Also, through our queries, we gained a closer look at our data which provided us with a clearer understanding of fitness trends. For example we decided to look at the days with most and least activity, days with worse sleep quality and how other factors are impacted by this and so on.

Cleaning the Data

The datasets shared the same User ID column but required some cleaning to ensure consistency and reliability.

1. **Renaming our variables**

To address inconsistencies in column names across datasets, we renamed variables to standardize the data. This step ensured that tables could be merged in an easier way.
For example:

- The SleepDay column in the sleep dataset was renamed to Date.
- The ActivityDate column in the activity dataset was also renamed to Date.

This allowed the datasets to be combined based on the common User Id and Date columns.

2. **Formatting Dates**

All date columns were converted to a consistent format to standardize the data. This process was necessary as the different datasets used varying date formats, which could create issues during merging.

- The dates were reformatted using as.Date() and POSIXct functions in R, ensuring uniformity.

3. **Summarizing data**

A thorough data summary was performed using the summary() function to identify anomalies, unrealistic values, and potential errors. For instance:

- In the sleep dataset, some observations showed individuals sleeping 0 hours, which was quite unrealistic. Additionally, we wanted to analyze the habits of healthy people, so sleeping 0 hours was not something we took into account for our conclusions.
- Summary statistics such as minimum, maximum, and mean values were reviewed to identify logical inconsistencies and set realistic boundaries.

4. **Outliers**

To address outliers, logical boundaries and statistical measures were applied:

- Outliers were detected using the interquartile range method.
- Limits were set for maximum and minimum values in calories, steps and sleep duration among others.

Observations falling outside of these boundaries were removed to ensure the integrity of our analysis.

5. **Missing values**

We did not have any missing values in terms of blanks or lack of data, rather the “missing values” we had represented a lack of activity in itself, or lack of recorded activity.

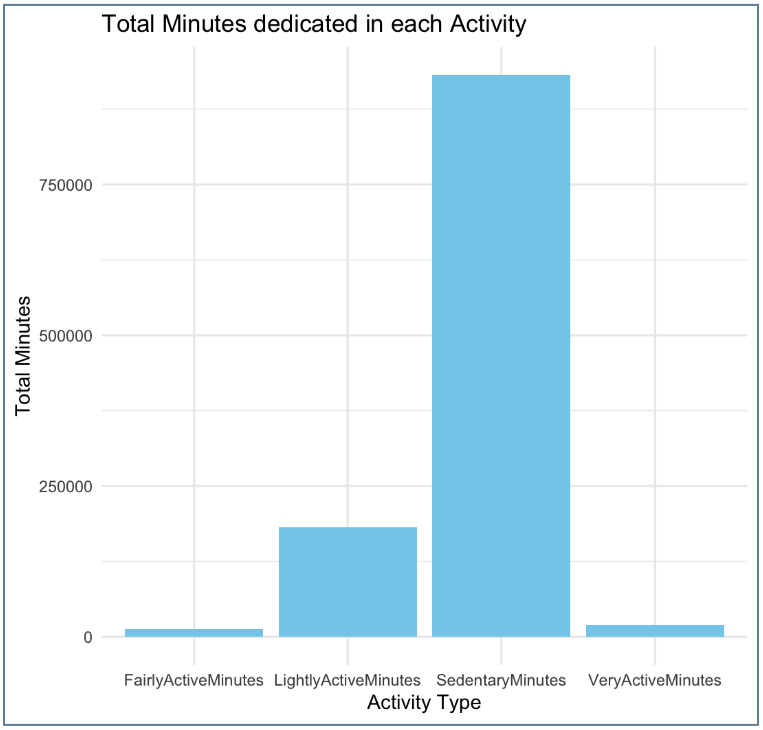
6. **Regression Analysis**

To identify significant variables and relationships to prepare for our analysis, linear regression models were employed on both individual datasets and merged tables. The results revealed which variables had a significant impact on others, allowing us to direct our data visualizations towards more specific ideas.

Analysis

Visualizations:

Visualization 1: Total Minutes dedicated to Each Activity

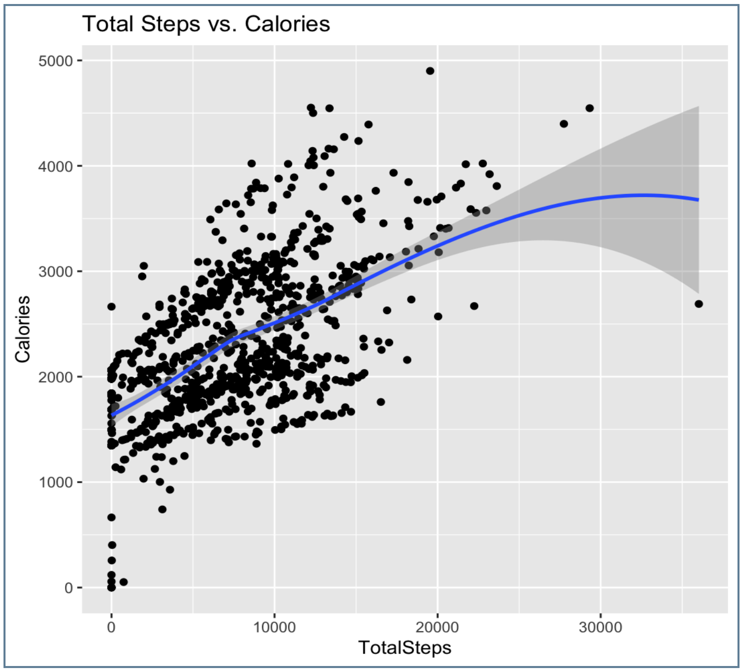



```
activity_distribution <- colSums(activity[, c("VeryActiveMinutes", "FairlyActiveMinutes",
                                              "LightlyActiveMinutes", "SedentaryMinutes")])
activity_distribution_df <- data.frame(
  ActivityType = names(activity_distribution),
  TotalMinutes = as.numeric(activity_distribution)
)

ggplot(activity_distribution_df, aes(x = ActivityType, y = TotalMinutes)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  theme_minimal() +
  labs(title = "Total Minutes dedicated in each Activity",
       x = "Activity Type",
       y = "Total Minutes")
```

Users' time allocation across the following activity levels is shown in this bar chart: Sedentary, Lightly Active, Fairly Active, and Very Active Minutes. The code uses ggplot2 to create a basic bar chart after aggregating the total minutes for each activity category using colSums. The graph shows that, in contrast to other activity categories, users spend the bulk of their time engaging in sedentary behavior. While Fairly Active and Very Active Minutes continue to be notably low, Lightly Active Minutes exhibit a considerable allocation. This pattern highlights the need for more physical activity by reflecting a lifestyle that is primarily idle.

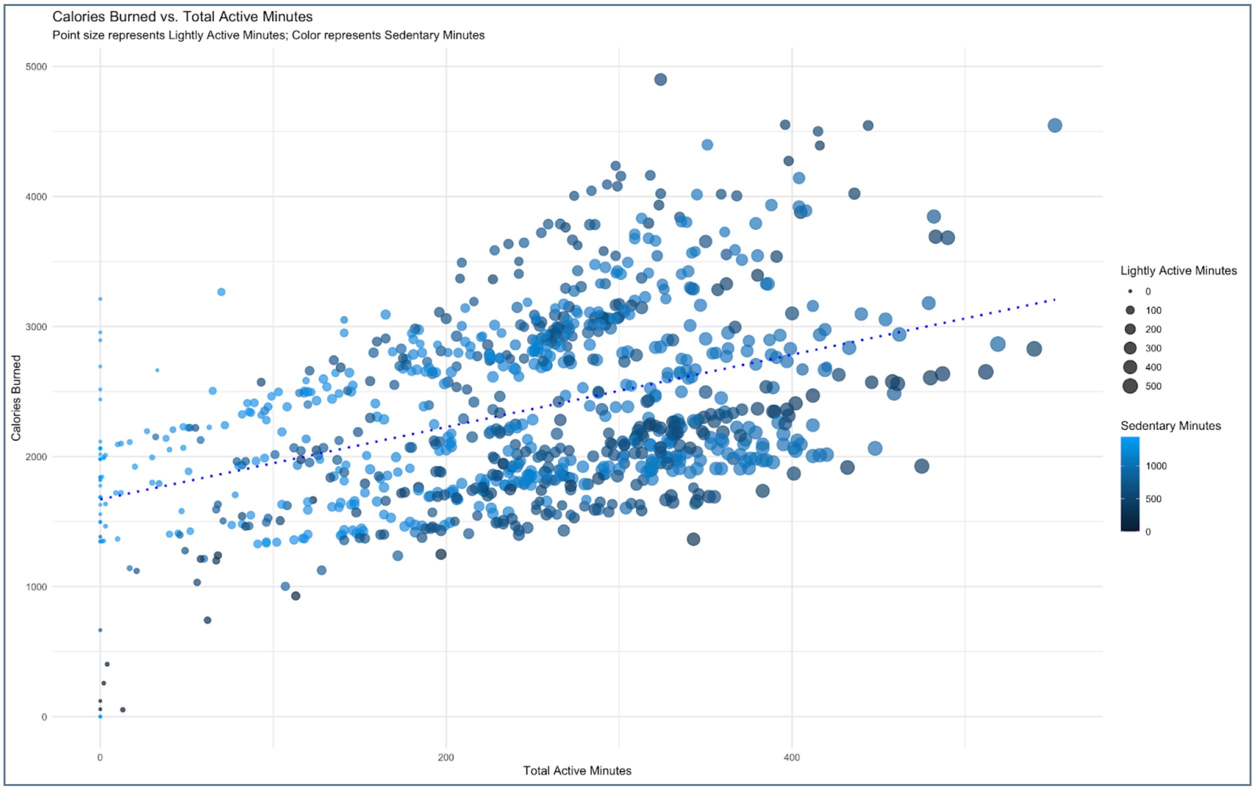
Visualization 2: Total Steps vs Calories burned




```
ggplot(data = activity, aes(x = TotalSteps, y = Calories)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Total Steps vs. Calories")
```

In order to determine whether step count is a reliable indicator of efficient physical activity, this scatter plot illustrates the association between total steps and calories burned. The graph shows a positive connection, showing a clear relationship: calories burnt often rise as the number of steps increases. The non-linear graph, however, indicates that the trend flattens with larger step counts, suggesting that there are diminishing returns in calorie burn after a certain point. A smoothed trend line (geom_smooth) was superimposed on a scatter plot in the ggplot2 display, which successfully draws attention to this pattern.

Visualization 3: Calories Burned vs Total Active Minutes




```
activity$TotalActiveMinutes <- activity$VeryActiveMinutes + activity$FairlyActiveMinutes + activity$LightlyActiveMinutes

ggplot(activity, aes(x = TotalActiveMinutes, y = Calories)) +
  geom_point(aes(color = SedentaryMinutes, size = LightlyActiveMinutes), alpha = 0.7) +
  geom_smooth(method = "lm", color = "blue", se = FALSE, linetype = "dotted") +
  labs(title = "Calories Burned vs. Total Active Minutes",
       subtitle = "Point size represents Lightly Active Minutes; Color represents Sedentary Minutes")
```

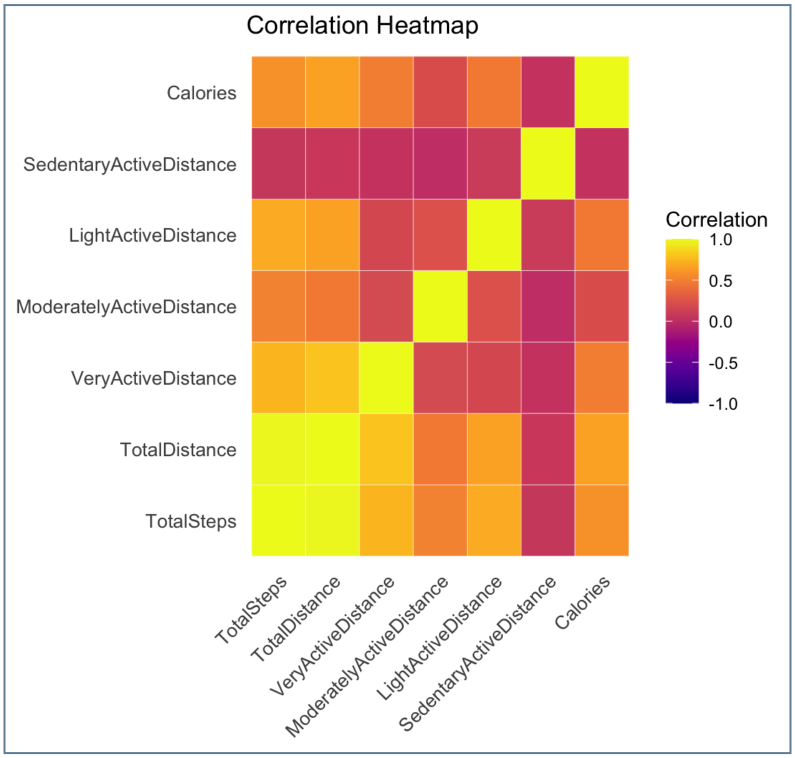
With the color gradient signifying Sedentary Minutes and the point size denoting Lightly Active Minutes, this scatter figure illustrates the correlation between Total Active Minutes and Calories Burned. The graph shows that there is a positive link between the number of calories burned and the total active minutes. However, if their total activity time is high enough, even people with higher sedentary minutes can burn a considerable number of calories. A trend line highlights the growing pattern in the ggplot2-created image, which sheds light on how active minutes affect caloric expenditure.

Visualization 4: Calories Burned Vs Total Steps


```r
ggplot(activity, aes(x = TotalSteps, y = Calories)) +
  geom_point(aes(color = TotalDistance, size = VeryActiveMinutes), alpha = 0.7) +
  geom_smooth(method = "lm", color = "red", se = FALSE, linetype = "dashed") +
  labs(title = "Calories Burned vs. Total Steps",
       subtitle = "Point size represents Very Active Minutes; Color represents Total Distance")
```

With two more dimensions—the bubble size denotes Very Active Minutes, and the color gradient denotes Total Distance Traveled—this bubble plot shows the correlation between Total Steps and Calories Burned. The graphic demonstrates a definite upward trend: those who walk more often typically burn more calories. The darker hues, which stand for farther traveled, emphasize how important movement intensity is for burning calories. Larger bubbles also show that, even after adjusting for the overall number of steps, people who participate in high-intensity activities (more Very Active Minutes) burn more calories. Using ggplot2, this graph skillfully integrates several activity measures into a single, perceptive display.

Visualization 5: Correlation Heatmap of Activity Metrics

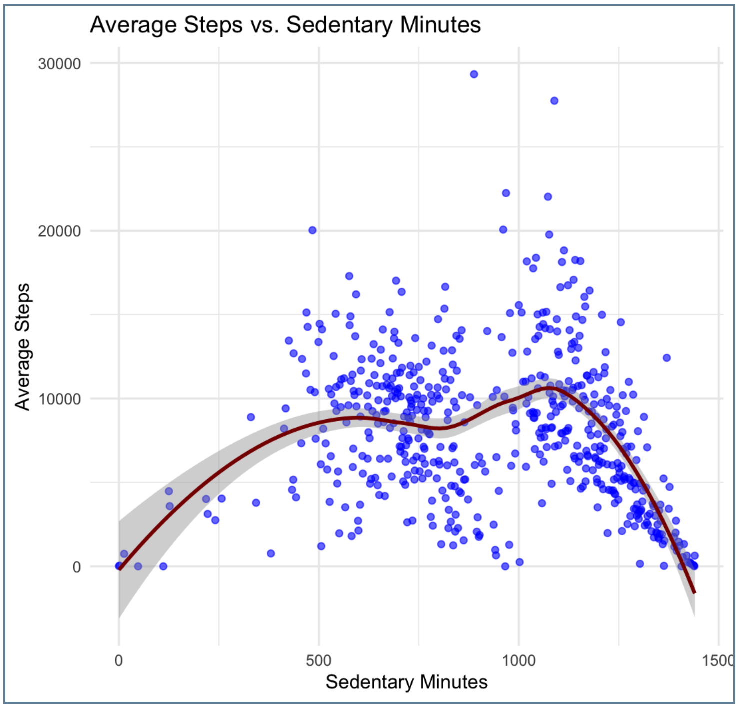



```r
correlation_data <- activity[, c("TotalSteps", "TotalDistance", "VeryActiveDistance",
                                 "ModeratelyActiveDistance", "LightActiveDistance",
                                 "SedentaryActiveDistance", "Calories")]
correlation_matrix <- cor(correlation_data, use = "complete.obs")
correlation_long <- melt(correlation_matrix)

ggplot(correlation_long, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_viridis(name = "Correlation", option = "plasma") +
  labs(title = "Correlation Heatmap")
```

The relationships between calories, steps, and other distance measurements are shown in this heatmap. Higher levels of physical activity directly enhance calorie burn, as seen by the strong positive association between calories and Very Active Distance, Total Steps, and Total Distance. Conversely, the low impact of Sedentary Active Distance is highlighted by its modest relationships with other factors. Geom_tile() in ggplot2 was used for visualization, while cor() was used to compute correlations in order to construct the heatmap. Overall, the graph highlights the large contribution of intense activity (more steps and active movement) to calorie expenditure.

Visualization 6: Average Steps Vs. Sedentary Minutes



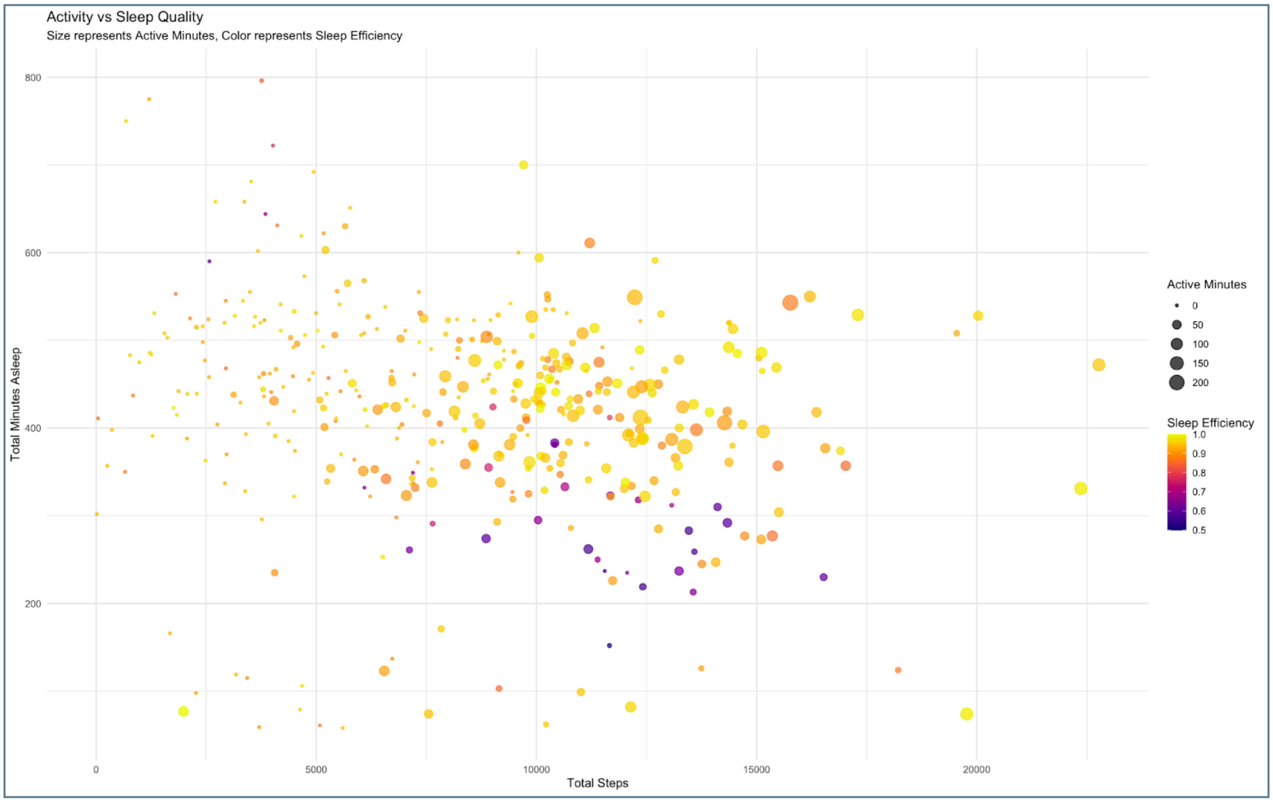

```
avg_sedentary_steps <- activity %>%
  group_by(SedentaryMinutes) %>%
  summarise(AverageSteps = mean(TotalSteps, na.rm = TRUE))

ggplot(avg_sedentary_steps, aes(x = SedentaryMinutes, y = AverageSteps)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "loess", color = "darkred") +
  labs(title = "Average Steps vs. Sedentary Minutes")
```

This scatter plot illustrates the connection between Sedentary Minutes and Average Steps to determine how extended periods of inactivity influence physical activity levels. The trend, depicted using geom_smooth() in ggplot2, reveals a distinct peak activity range where individuals with moderate sedentary minutes attain the greatest step counts. Nevertheless, average steps drop dramatically once sedentary minutes go beyond 700, signaling diminished physical involvement with extended inactivity. The analysis indicates that keeping a balance between rest and movement aids in maximizing physical activity levels. Promoting breaks to reduce excessive sedentary behavior could greatly enhance overall activity patterns.

Merges- Activity and Sleep

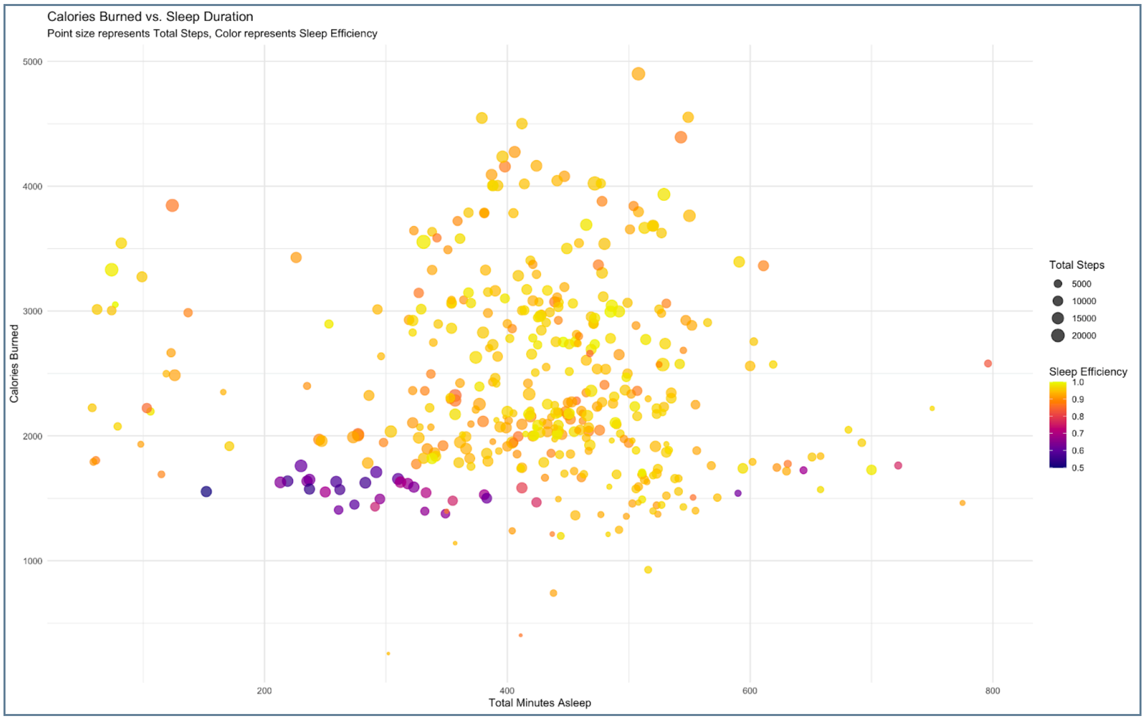
Visualization 7: Activity vs. Sleep Quality




```
merged_data <- merge(activity, sleep, by.x = c("Id", "ActivityDate"), by.y = c("Id", "SleepDay"))
ggplot(merged_data, aes(x = TotalSteps, y = TotalMinutesAsleep)) +
  geom_point(aes(color = SleepEfficiency, size = VeryActiveMinutes), alpha = 0.7) +
  scale_color_viridis_c(option = "plasma") +
  labs(title = "Activity vs Sleep Quality",
       subtitle = "Point size represents Active Minutes, Color represents Sleep Efficiency")
```

This scatter plot examines the connection between Total Steps and Total Minutes Asleep, with point size denoting Active Minutes and color representing Sleep Efficiency. The trend line, produced with geom_smooth() in ggplot2, reveals a slight positive correlation, indicating that greater physical activity might have a modest effect on sleep duration. Nonetheless, the plot suggests no strong correlation, as sleep duration fluctuates independently of the number of steps taken. Importantly, elevated sleep efficiency (shown by warmer colors) is observed with moderate levels of activity, supporting the concept that a balance of physical exercise and rest enhances sleep quality.

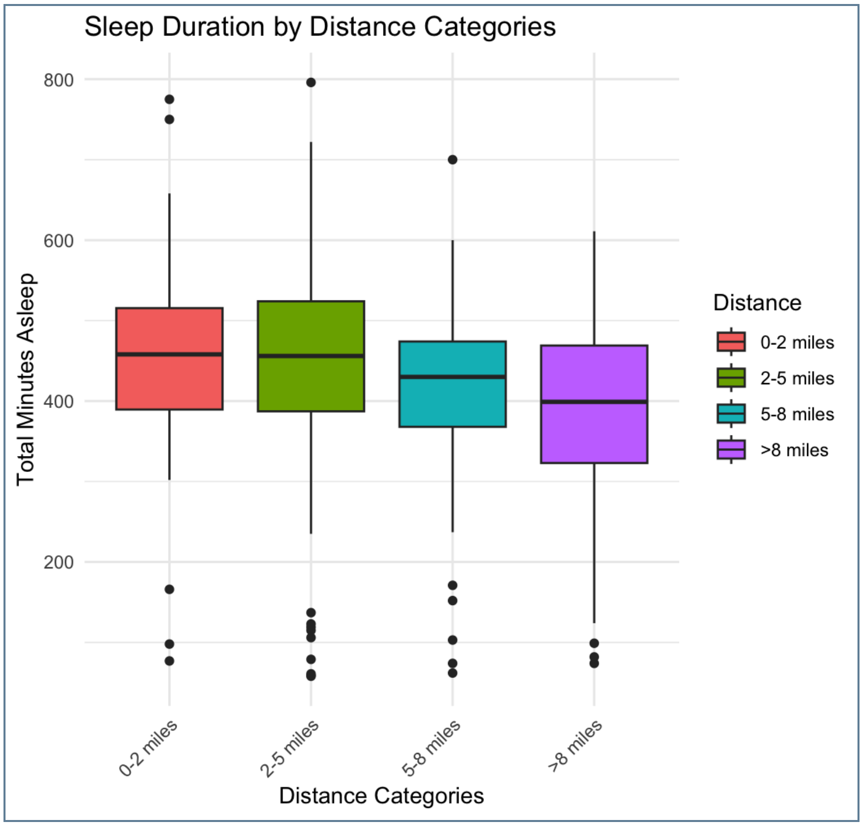
Visualization 8: Calories Burned Vs. Sleep Duration




```
ggplot(merged_data, aes(x = TotalMinutesAsleep, y = Calories)) +
  geom_point(aes(color = SleepEfficiency, size = TotalSteps), alpha = 0.7) +
  scale_color_viridis_c(option = "plasma") +
  labs(title = "Calories Burned vs Sleep Duration",
       subtitle = "Point size represents Total Steps, Color represents Sleep Efficiency")
```

This scatter plot examines the connection between Calories Burned and Total Minutes Asleep, with the size of the points depicting Total Steps and the color representing Sleep Efficiency. The trend line, generated using geom_smooth() in ggplot2, indicates a minor positive correlation, implying that people who sleep longer often burn more calories. Nevertheless, the variability in the data underscores that calorie expenditure is affected by several factors beyond sleep duration alone. Importantly, points with greater sleep efficiency (yellow shades) group around moderate sleep lengths, highlighting the significance of sleep quality over mere duration for better physical results.

Visualization 9: Sleep Duration by Distance Categories

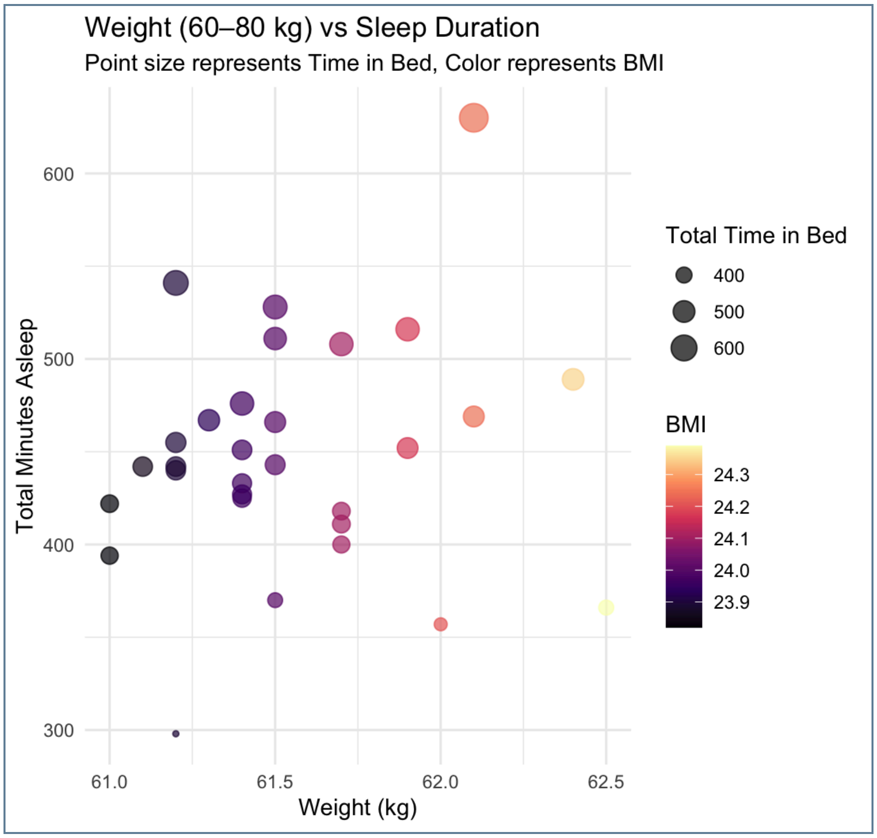



```
merged_data <- merged_data %>%
  mutate(DistanceCategory = cut(TotalDistance, breaks = c(0, 2, 5, 8, Inf),
                               labels = c("0-2 miles", "2-5 miles", "5-8 miles", ">8 miles")))

ggplot(merged_data, aes(x = DistanceCategory, y = TotalMinutesAsleep)) +
  geom_boxplot(aes(fill = DistanceCategory)) +
  labs(title = "Sleep Duration by Distance Categories")
```

This box plot illustrates Total Minutes Asleep among four distance categories (0-2, 2-5, 5-8, >8 miles). The plot indicates that sleep duration stays fairly stable regardless of the distance covered during the day, implying a minimal effect of distance on sleep duration. Nevertheless, there is increased variability in sleep duration for individuals who traveled shorter distances (0-2 miles), as demonstrated by the broader interquartile range and the existence of outliers. This variability could be linked to other lifestyle factors, such as stress or extended periods of inactivity. The plot effectively emphasizes that while activity plays a role, sleep duration might be influenced by more extensive lifestyle trends.

Visualization 10: Weight vs Sleep Duration

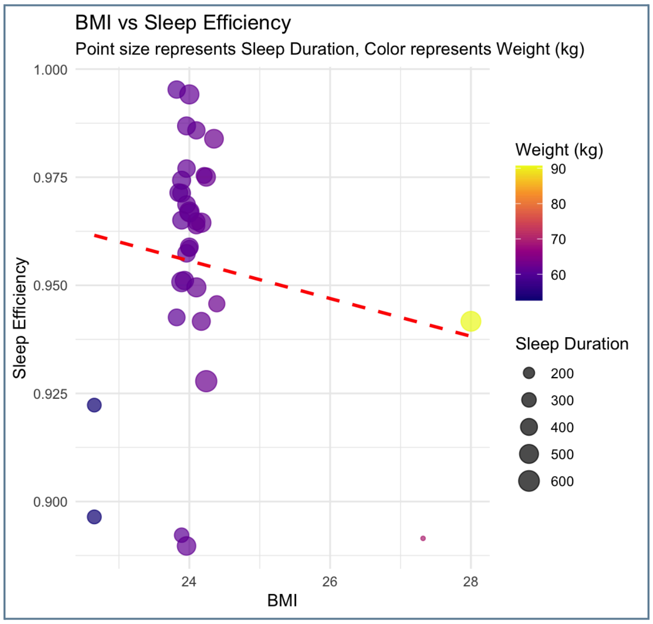



```r
filtered_data <- merged_sleep_weight %>%
  filter(WeightKg >= 55 & WeightKg <= 65)

ggplot(filtered_data, aes(x = WeightKg, y = TotalMinutesAsleep)) +
  geom_point(aes(color = BMI, size = TotalTimeInBed), alpha = 0.7) +
  scale_color_viridis_c(option = "magma") +
  labs(title = "Weight (60-80 kg) vs Sleep Duration")
```

This scatter plot examines the relationship between Weight (60–80 kg) and Total Minutes Asleep, with the point size representing Total Time in Bed and color indicating BMI. The graph reveals that within this weight range, sleep duration remains relatively stable, showing minimal variation with weight or BMI. Instead, Total Time in Bed appears to have a stronger influence on sleep duration, as larger points cluster toward higher sleep values. The visualization was created using ggplot2 specifically leveraging the geom_point() function with aesthetics for size and color to represent multiple variables.

Visualization 11: BMI vs Sleep Efficiency




```
filtered_data <- merged_sleep_weight %>%
  filter(SleepEfficiency >= 0.5 & SleepEfficiency <= 1, BMI >= 15 & BMI <= 40)

ggplot(filtered_data, aes(x = BMI, y = SleepEfficiency)) +
  geom_point(aes(color = WeightKg, size = TotalMinutesAsleep), alpha = 0.7) +
  geom_smooth(method = "lm", color = "red", linetype = "dashed") +
  labs(title = "BMI vs Sleep Efficiency")
```

This scatter plot illustrates the connection between BMI and Sleep Efficiency, where the size of the points signifies Sleep Duration and the color reflects Weight (kg). The slight decline, represented with geom_smooth() in ggplot2, implies that individuals with elevated BMI may experience slightly lower sleep efficiency. Nevertheless, the correlation is weak, as fluctuations in sleep efficiency seem minor and affected by other elements such as lifestyle or stress. This graph underscores that BMI by itself does not considerably affect sleep efficiency, highlighting the necessity to investigate other factors like diet, stress, or daily habits for enhancing sleep quality.

Queries

Query 1: Ranking Highest Steps

Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance
1624580081	2016-05-01	36019	28.03	28.03	0	21.92
8877689391	2016-04-16	29326	25.29	25.29	0	13.24
8877689391	2016-04-30	27745	26.72	26.72	0	21.66
8877689391	2016-04-27	23629	20.65	20.65	0	13.07
8877689391	2016-04-12	23186	20.40	20.40	0	12.22

ModeratelyActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes
4.19	1.91	0.02	186	63
1.21	10.71	0.00	94	29
0.08	4.93	0.00	124	4
0.44	7.10	0.00	93	8
0.34	7.82	0.00	85	7

LightlyActiveMinutes	SedentaryMinutes	Calories	TotalActiveMinutes	DayOfWeek	StepCategory	ActiveMinutes
171	1020	2690	420	Sunday	15k+	420
429	888	4547	552	Saturday	15k+	552
223	1089	4398	351	Saturday	15k+	351
235	1104	3808	336	Wednesday	15k+	336
312	1036	3921	404	Tuesday	15k+	404

This query ranks the days with the highest number of steps, emphasizing activity trends throughout the week. Sunday and Saturday stand out as the most active days, featuring step counts that surpass 15,000 steps, indicating that users tend to be more physically active on weekends. Conversely, activity during the weekdays (for instance, Wednesday and Tuesday) demonstrates lower yet still notable step counts, probably impacted by work schedules. The information highlights a trend of heightened physical activity on weekends, potentially owing to greater leisure time availability.

Query 2: Ranking slower steps

LightlyActiveMinutes	SedentaryMinutes	Calories	TotalActiveMinutes	DayOfWeek
0	1440	0	0	Thursday
0	1440	1347	0	Sunday
0	1440	1347	0	Monday
0	1440	1347	0	Tuesday
0	1440	1348	0	Monday

This query pinpoints the days that exhibit the fewest steps taken, showcasing a lack of physical activity on Thursday, Sunday, Monday, and Tuesday. These entries reflect 1440 minutes of inactivity (the entire 24 hours), zero calories expended, and no minutes of activity, signifying total immobility on these days. The occurrence of several Mondays points to a consistent pattern of inactivity at the week's start, potentially stemming from exhaustion or job-related habits. This underscores the importance of promoting increased physical activity during weekdays to offset sedentary behavior and enhance health results.

Query 3, 4: Ranking Highest/Lowest Distance

DayOfWeek
Sunday
Saturday
Saturday
Wednesday
Tuesday

Highest Distance

DayOfWeek
Thursday
Sunday
Monday
Tuesday
Wednesday
Monday

Lowest Distance

Highest Distance (Query 3):

The days where the longest distances are recorded are mainly weekends (Sunday and Saturday), next are Wednesday and Tuesday. This implies that users tend to be more engaged during weekends and mid-week, likely because of more leisure time or organized exercise schedules.

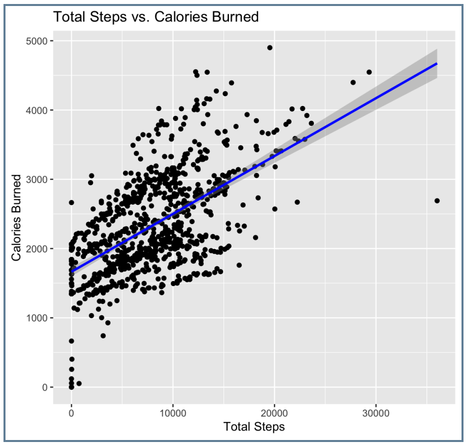
Lowest Distance (Query 4):

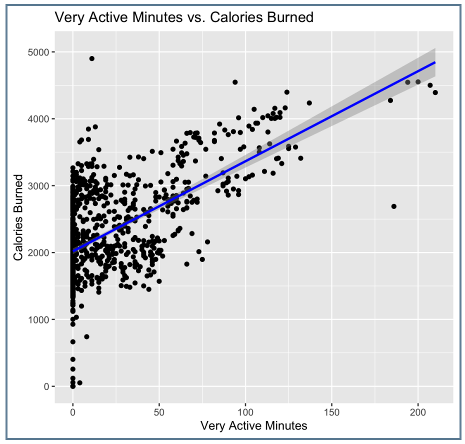
The least distances are noted on Thursday, Sunday, Monday, and Tuesday, with Monday listed two times. This shows diminished activity levels at the week's beginning, possibly attributed to tiredness or job responsibilities, and underscores irregularity in physical activity throughout the week.

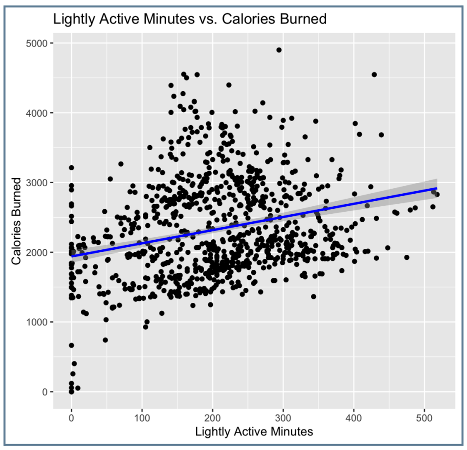
Query 5,6,7

```
# Query to get relevant data
activity_data <- sqldf("SELECT TotalSteps, VeryActiveMinutes, LightlyActiveMinutes, Calories
                        FROM activity")
```

```
# Print correlations
print(paste("Correlation between Total Steps and Calories:", cor_steps_calories))
] "Correlation between Total Steps and Calories: 0.591568086245336"
print(paste("Correlation between Very Active Minutes and Calories:", cor_veryactive_calories))
] "Correlation between Very Active Minutes and Calories: 0.615838268270337"
print(paste("Correlation between Lightly Active Minutes and Calories:", cor_lightactive_calories))
] "Correlation between Lightly Active Minutes and Calories: 0.286717534017549"
```







Total Steps vs. Calories: 0. 591 – a moderate positive relationship, suggesting that more steps typically result in a higher number of calories burned.

Very Active Minutes vs. Calories: 0. 615 – a stronger relationship, underscoring the significance of vigorous activity.

Lightly Active Minutes vs. Calories: 0. 286 – a weaker relationship, indicating minimal calorie expenditure from light activity.

The visual representations above (scatter plots) illustrate these ideas:

Total Steps vs. Calories Burned – an upward trend with some dispersion.

Very Active Minutes vs. Calories Burned – a distinct upward trajectory, highlighting the effects of vigorous activity's contribution.

Lightly Active Minutes vs. Calories Burned – a less pronounced slope, depicting smaller calorie impacts.

This examination shows that vigorous activity (Very Active Minutes) has the most considerable effect on calorie expenditure, followed by Total Steps, while Lightly Active Minutes have a minimal effect.

Query 8: Distance by Activity Intensity

ActivityDate	TotalVeryActiveDistance	TotalModeratelyActiveDistance	TotalLightActiveDistance
2016-04-12	60.27	11.42	112.53
2016-04-13	43.78	13.86	103.65
2016-04-14	49.82	16.82	117.76
2016-04-15	34.84	13.33	124.32
2016-04-16	63.80	22.68	110.42
2016-04-17	36.65	15.92	90.31

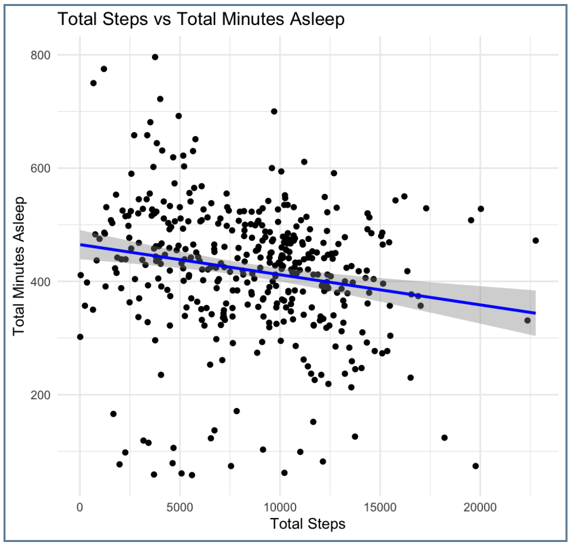
This analysis examines the distance traveled during various levels of activity (Very Active, Moderately Active, and Lightly Active) on particular dates. The information shows that Lightly Active Distance consistently accounts for the largest portion of total distances, with figures surpassing 100 units on each day. Very Active Distance fluctuates widely, reaching a maximum of 63. 80 on 2016-04-16, whereas Moderately Active Distance remains the least overall.

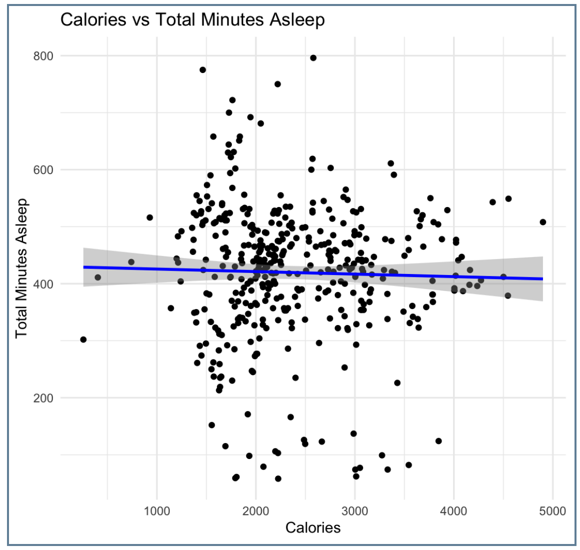
This trend suggests that the majority of users participate in light physical activities, whereas high-intensity movements (Very Active Distance) occur less often. The results emphasize the necessity of promoting more vigorous physical activity to attain improved fitness results.

Query 9,10: Activity and Sleep Quality

TotalSteps	TotalActiveMinutes	Calories	TotalMinutesAsleep	TotalTimeInBed
13162	366	1985	327	346
10735	257	1797	384	407
9762	272	1745	412	442
12669	267	1863	340	367
9705	222	1728	700	712
15506	345	2035	304	320

```
> # Print correlations
> print(paste("Correlation between Total Steps and Minutes Asleep:", cor_steps_sleep))
[1] "Correlation between Total Steps and Minutes Asleep: -0.18686649892546"
> print(paste("Correlation between Total Steps and Time in Bed:", cor_steps_bed))
[1] "Correlation between Total Steps and Time in Bed: -0.164059712512068"
> print(paste("Correlation between Active Minutes and Minutes Asleep:", cor_active_sleep))
[1] "Correlation between Active Minutes and Minutes Asleep: -0.0637605952066786"
> print(paste("Correlation between Active Minutes and Time in Bed:", cor_active_bed))
[1] "Correlation between Active Minutes and Time in Bed: -0.0933415426575205"
> print(paste("Correlation between Calories and Minutes Asleep:", cor_calories_sleep))
[1] "Correlation between Calories and Minutes Asleep: -0.02852571334282"
> print(paste("Correlation between Calories and Time in Bed:", cor_calories_bed))
[1] "Correlation between Calories and Time in Bed: -0.132507095796556"
```





Total Steps vs. Minutes Asleep: -0. 18 – a weak negative correlation, suggesting that increased step counts do not necessarily enhance sleep duration.

Active Minutes vs. Minutes Asleep: -0. 06 – minimal correlation, indicating little to no effect of active minutes on sleep duration.

Calories vs. Minutes Asleep: -0. 28 – a slightly stronger negative relationship, indicating that higher calorie expenditures may not correspond with longer sleep.

The scatter plots visually validate these findings, displaying scattered trends with minimal linear connections. This analysis indicates that physical activity alone is not strongly linked to sleep quality, and other elements like lifestyle or stress may have a more substantial impact on sleep duration and efficiency.

Query 11:

```
> head(sedentary_sleep_analysis)
  SedentaryMinutes TotalMinutesAsleep TotalTimeInBed
1              728                327            346
2              776                384            407
3              726                412            442
4              773                340            367
5              539                700            712
6              775                304            320
>
```

```
> print(paste("Correlation between Sedentary Minutes and Total Minutes Asleep:", cor_sedentary_sleep))
[1] "Correlation between Sedentary Minutes and Total Minutes Asleep: -0.59939400560339"
> print(paste("Correlation between Sedentary Minutes and Total Time in Bed:", cor_sedentary_bed))
[1] "Correlation between Sedentary Minutes and Total Time in Bed: -0.618713452873604"
```

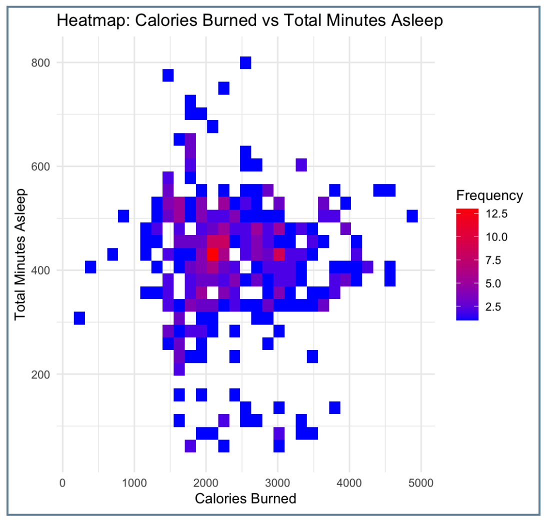
Sedentary Minutes vs. Total Minutes Asleep: -0. 59 – a moderate negative correlation, signifying that increased sedentary behavior is related to less sleep duration.

Sedentary Minutes vs. Total Time in Bed: -0. 61 – a stronger negative association, indicating that higher sedentary minutes result in reduced total time spent in bed.

The data suggests a possible connection between extended inactivity and diminished sleep duration, implying that those with more sedentary minutes might encounter inferior sleep quality. These results stress the significance of integrating physical activity throughout the day to counteract the effects of extended sedentary behavior.

Query 12:

```
> head(calories_sleep_analysis)
  Calories TotalMinutesAsleep TotalTimeInBed
1     1985                327            346
2     1797                384            407
3     1745                412            442
4     1863                340            367
5     1728                700            712
6     2035                304            320
```



This query examines the connection between Calories Burned and Total Minutes Asleep utilizing the calories_sleep_analysis dataset. The table offers a glimpse of the data, indicating that calories burned fluctuate between 1728–2035, while the duration of sleep spans from 304–700 minutes.

The heatmap illustrates the connection between the two variables, where elevated frequencies (marked in red) concentrate around moderate calorie burn figures and mid-range sleep durations. This suggests an absence of a distinct strong correlation between calories burned and sleep duration. The dispersed pattern implies that calorie expenditure alone is not a major predictor of sleep length, highlighting the impact of other factors such as activity type, stress, or lifestyle habits.

Conclusion

This project has given us many insights towards the world of fitness trends, health and the factors that are important to lead an effective, healthy and active lifestyle.

Our main takeaway is that balance is crucial. Our analyses show that balance is key to maintaining optimal health and well-being. We were shocked to find that the majority of users allocated their time to sedentary activities, which highlights the need to encourage healthy lifestyle choices. Light physical activities account for the majority of user movements, while high-intensity activities are not common among users. Activity levels tend to peak during the week day with users walking larger distances and having a higher step count, likely given time for leisure. Following this logic, Mondays show the lowest activity levels due to the beginning of the labour week and tiredness from the weekend.

Moderate exercise paired with enough breaks from sedentary behavior is key in effective calorie burn. In fact, we found that people with 700 sedentary minutes per day average 62.5% lower activity levels than those with 500 minutes or less. In fact, most of our correlations suggest the idea that the benefits of physical activity show diminishing returns past a certain point, highlighting the importance of staying away from extremes, a common fitness practice. It is important to encourage regular active breaks to boost overall physical engagement and improve energy expenditure.

In terms of sleep efficiency, we also found that individual variability in sleep patterns can be due to lifestyle elements such as stress, anxiety and stimulation given no large correlation with physical activity metrics. However, there are some implications given sleep duration has 20% more variability in individuals who walk less than two miles daily. Sleep is one of the fundamental pillars in overall health and energy expenditure throughout the day. It is crucial to have a solid and consistent sleep routine, prioritizing a decent time and amount of hours for optimal efficiency.

All in all, we have found that to achieve the best outcomes in fitness, setting balanced goals is the solution. Moderate activity levels of around 30-60 active minutes per day yield the best balance across engagement, energy expenditure, calories and active minutes.