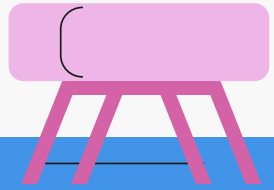
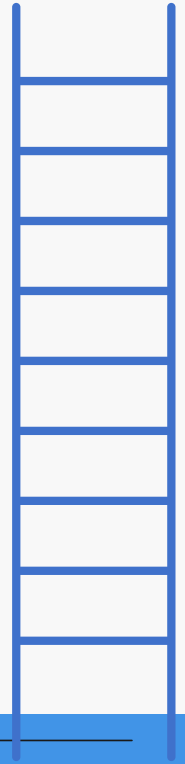
Active Lifestyles Analytics Project

Lucia Berni, Natalia Garcia and Pablo Saez

Table of contents





01

Executive Summary & Cleaning Data

Executive Summary

- Goal: understand the variables effective in leading active lifestyles.
- Study tracking data from FitBit users on their daily activities.
- We found multiple datasets that tracked the same users across 15 variables such as hours of sleep, calories burned, weight, type of activity, sedentary vs. active minutes...
- Our methods include exploratory data analysis, variable correlations, different type of visualizations using ggplot2 and solving data queries for further analysis.

Cleaning the Data

Renaming values

Standardize values and address inconsistencies

Formatting dates

Convert all date columns to a consistent format

Summarizing Data

Delete anomalies such as unrealistic values like people sleeping 0 hours.

Handling outliers

From S3- set realistic boundaries for variables to filter out outliers.

Missing values

No missing values, simply “null” (ex. No activity)

Regression

Selected which variables where most significant using regression analyses on both individual and merged data sets

02

Visualizations

Visualization 1:

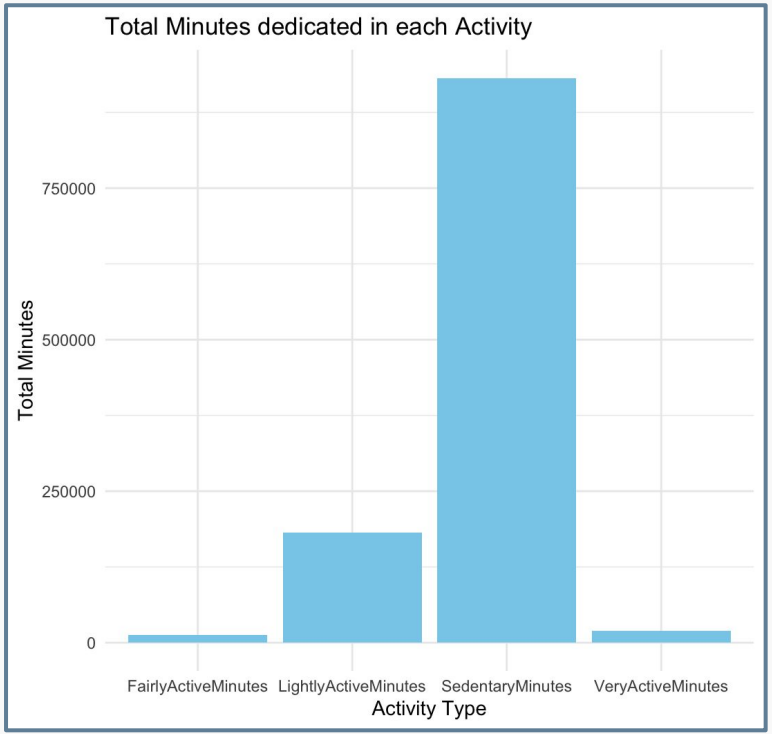
How do users allocate their time across different activity levels?

1. Sedentary Behavior:

   • Users spend the most time in Sedentary Minutes, highlighting a predominantly inactive lifestyle.

2. Light Activity:

   • A noticeable portion of time is spent on Lightly Active Minutes, representing mild physical activity.

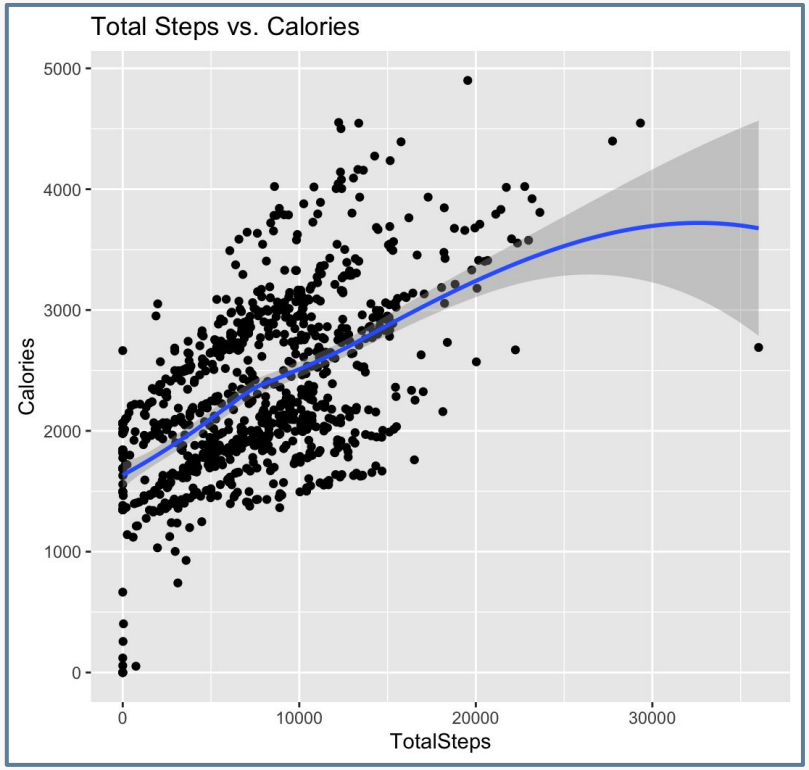


Visualization 2:

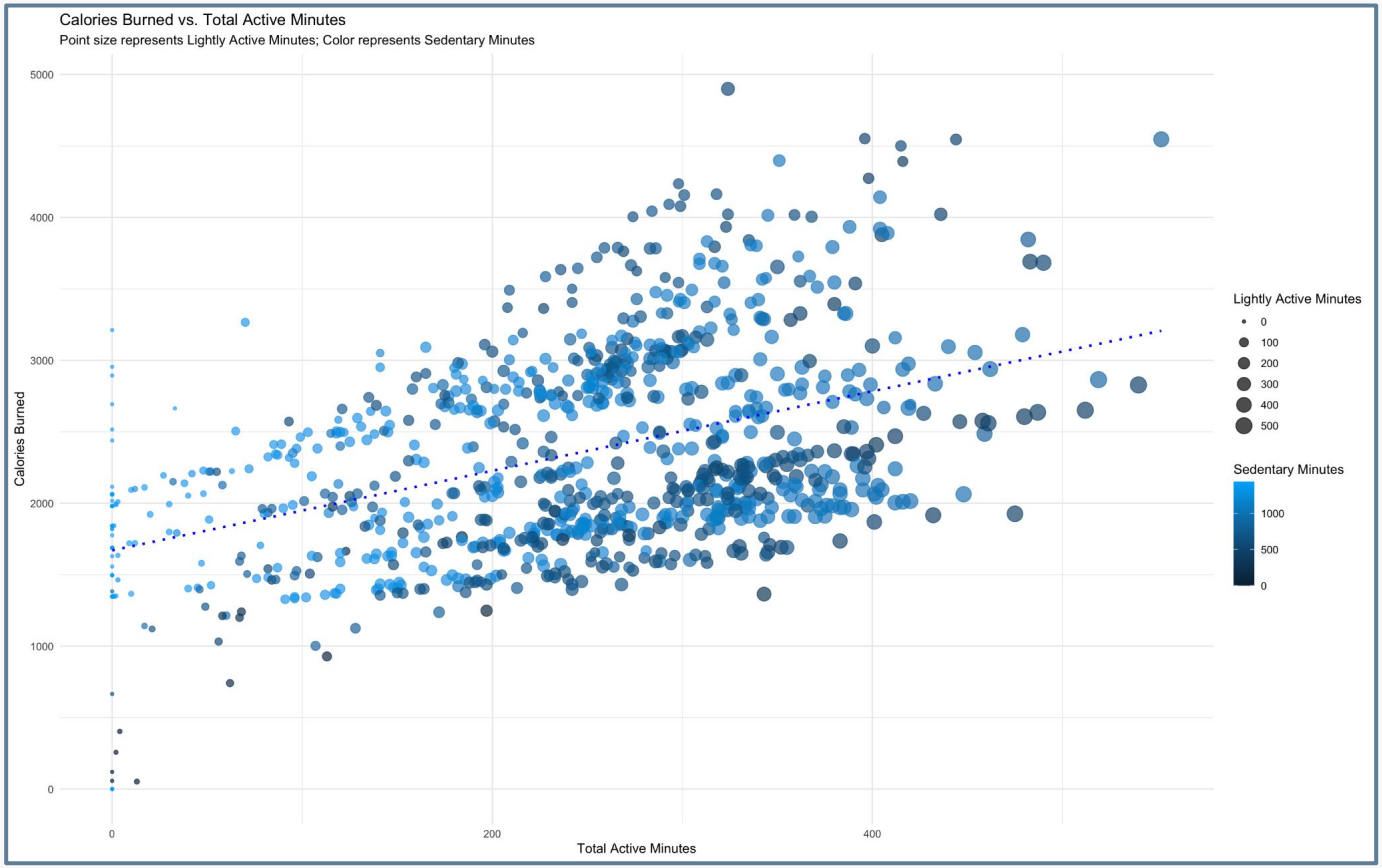
Is there any correlation between the Total Steps and the Calories Burned?

- Can we measure effective physical activity through steps?

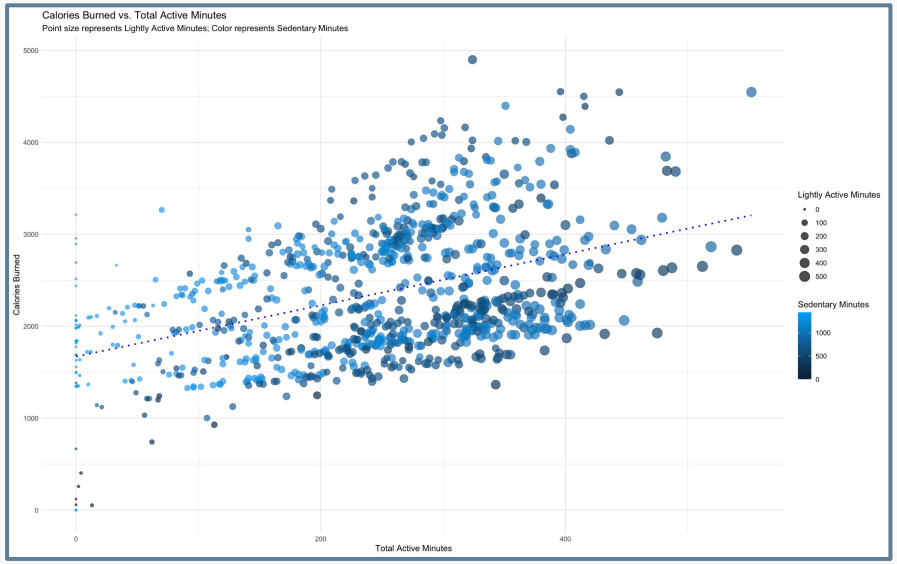
1. Positive Correlation:
   - As the number of steps increases, the calories burned also generally increase, indicating a direct relationship.
2. Non-linear Trend:
   - The trend flattens at higher step counts, suggesting diminishing returns in calorie burn beyond a certain activity level.



Visualization 3:



Visualization 3:



1. **Positive Correlation**:

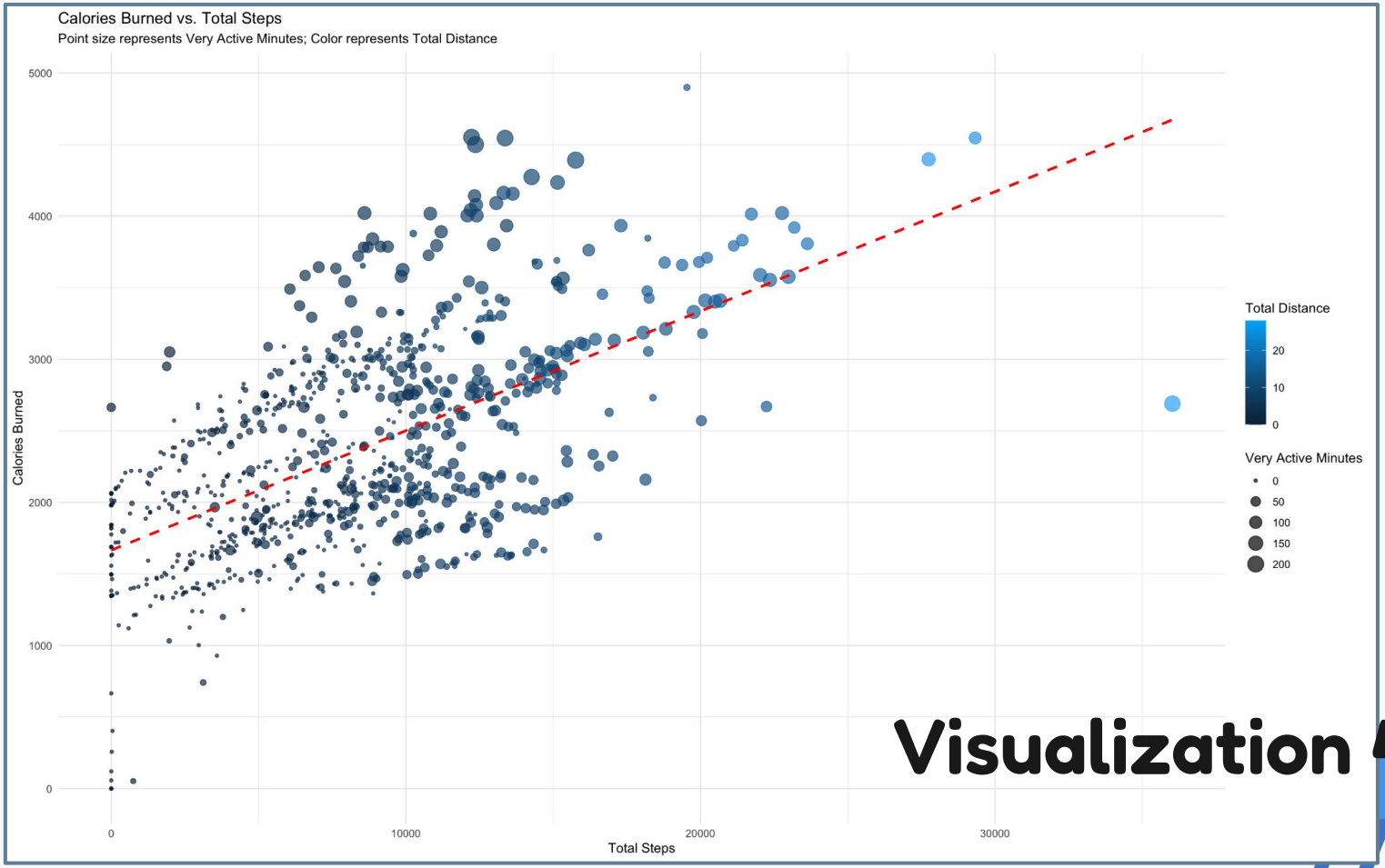
   • Calories burned generally increase with total active minutes, as indicated by the upward trend in the dotted regression line.

2. **Impact of Sedentary Minutes**:

   • Darker points (representing higher sedentary minutes) tend to cluster toward the lower active minutes and calorie ranges, suggesting a sedentary lifestyle limits calorie burn.

3. **Role of Lightly Active Minutes**:

   • Larger points (representing higher lightly active minutes) appear more evenly distributed, indicating light activity contributes to overall calorie expenditure, but not as much as more intense activity.



Visualization 4:

Visualization 4:

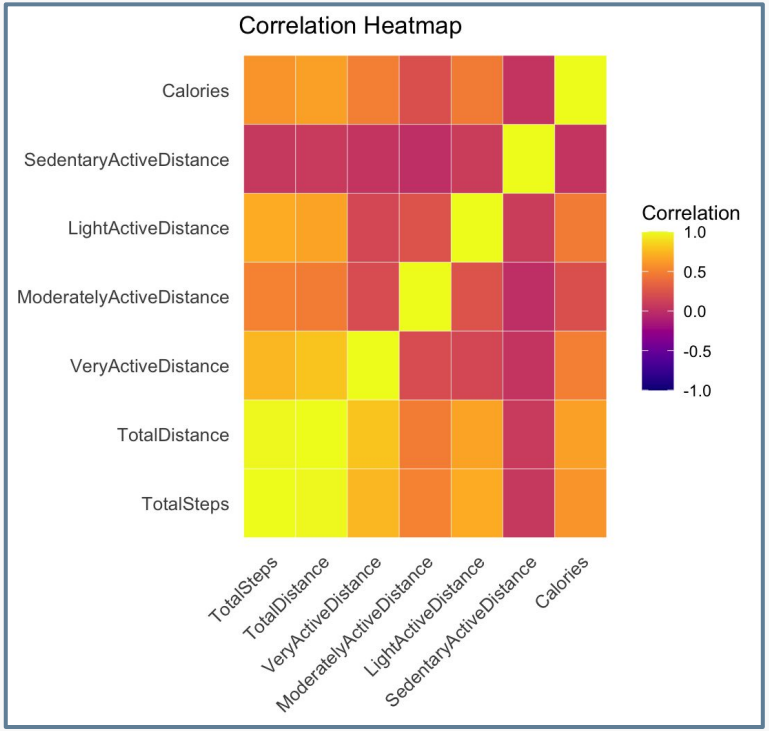
-Darker shades indicate higher total distance traveled.
-This shows that individuals who cover more distance tend to burn more calories, reinforcing the role of movement intensity in calorie expenditure.
-Larger bubbles indicate more time spent in high-intensity activity.
- This highlights that individuals with higher very active minutes typically achieve greater calorie burn, even when controlling for total steps.

Visualization 5:

-Calories is strongly positively correlated with variables like Very Active Distance, Total Distance, and Total Steps, indicating that increased physical activity directly impacts calorie burn.

-Sedentary Active Distance has weaker correlations with most other variables, suggesting that it has less influence on calorie burn compared to active distances.

-Very Active Distance and **Total Steps** show a strong correlation, as higher steps often include very active movements.



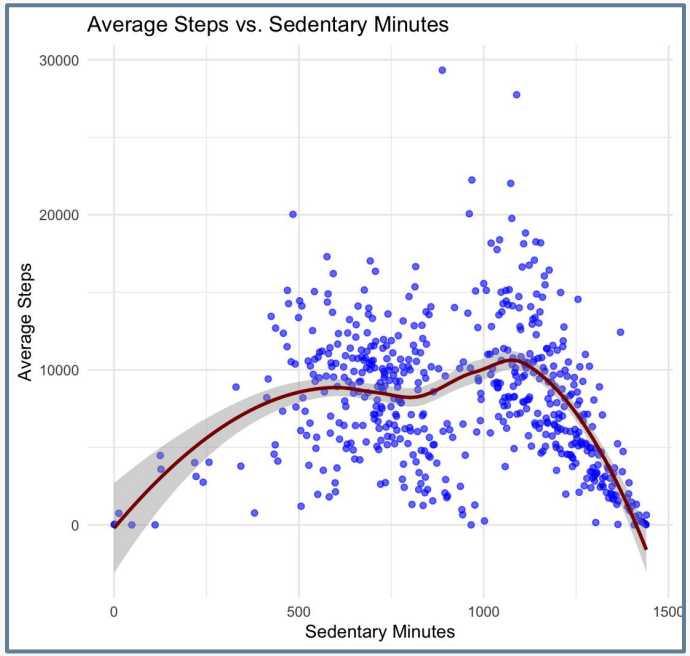
Visualization 6:

Encouraging users to limit prolonged sedentary behavior while incorporating active breaks may improve overall activity levels, as reflected in step counts.

Peak Activity: The highest step counts are observed for users with sedentary minutes in the moderate range, suggesting an optimal balance of activity and rest.

Decline Beyond Peak: As sedentary minutes increase beyond 700, average steps drop sharply, indicating less physical engagement for users with prolonged inactivity.

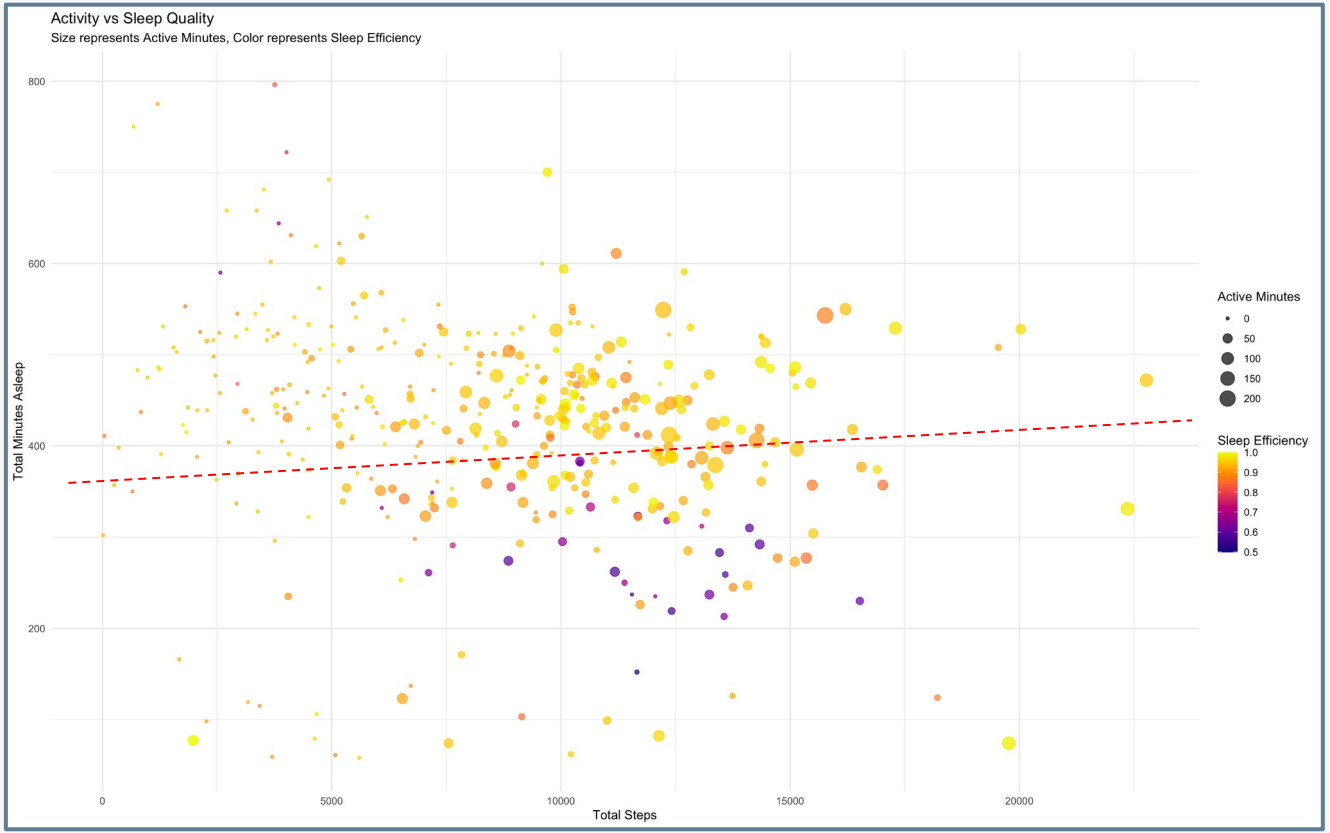
Activity Patterns: This trend implies that users who avoid prolonged sedentary behavior while maintaining moderate rest tend to be more physically active overall.



MERGES

Activity & Sleep

Visualization 7:

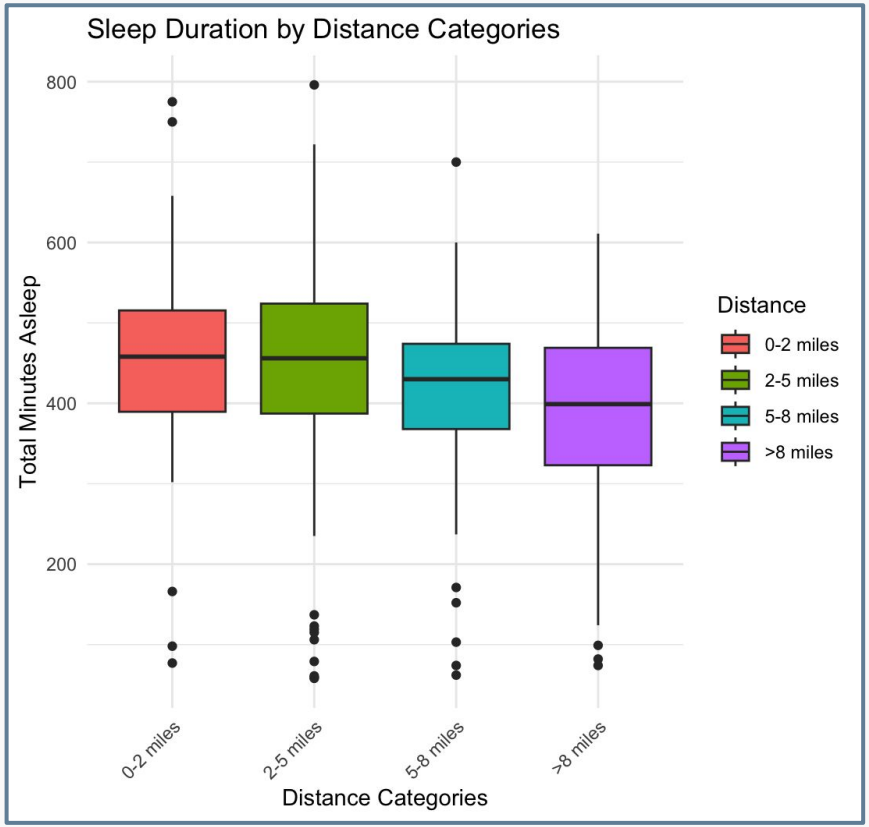


Visualization 8:

Visualization 9:

Conclusions:

- Sleep duration appears relatively independent of distance traveled during the day.
- Variability in sleep duration is higher among those covering shorter distances, potentially influenced by other lifestyle factors like stress or inactivity.



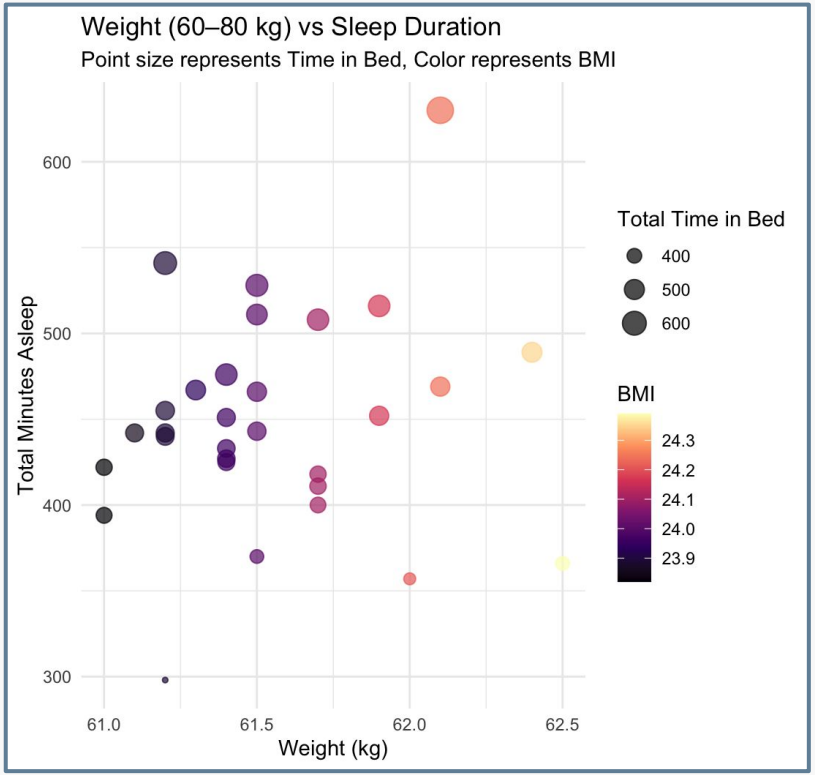
MERGES

Sleep & Weight

Visualization 10:

Conclusions:

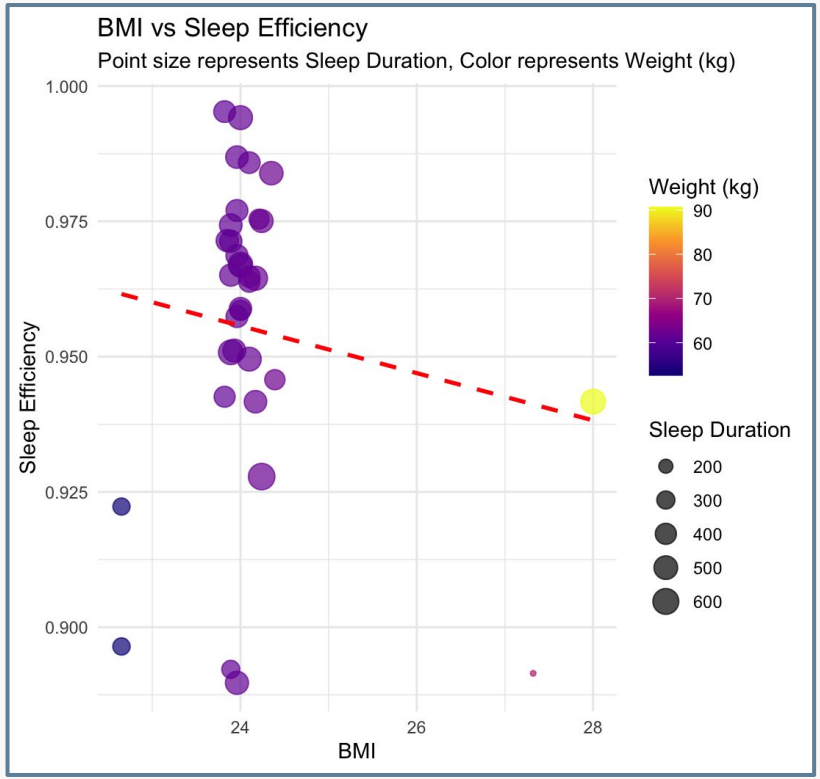
- Within this weight range, sleep duration appears relatively stable and is influenced more by time in bed than by weight or BMI.
- Suggests that factors beyond weight and BMI, such as sleep hygiene or daily routines, might better explain variations in sleep duration.



Visualization 11:

Conclusions:

- Individuals with higher BMI tend to have slightly reduced sleep efficiency, but the effect is minimal and likely influenced by additional factors.
- Sleep duration does not appear to correlate strongly with BMI or sleep efficiency.
- Highlights the need to explore lifestyle, diet, or stress levels as potential contributors to sleep efficiency.




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

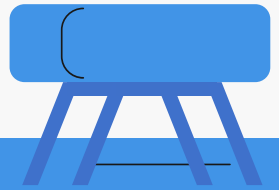
ModeratelyActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes
4.19	1.91	0.02	186	63
1.21	10.71	0.00	94	29
0.08	4.93	0.00	124	4
0.44	7.10	0.00	93	8
0.34	7.82	0.00	85	7

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

Query 2:

Ranking lowest 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



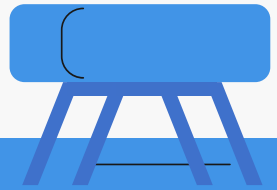
Query 3 & 4:

Ranking highest distance:

DayOfWeek
Sunday
Saturday
Saturday
Wednesday
Tuesday

Ranking lowest distance:

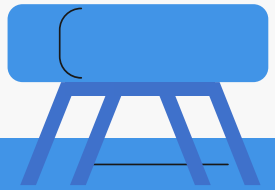
DayOfWeek
Thursday
Sunday
Monday
Tuesday
Wednesday
Monday



Query 5, 6, 7:

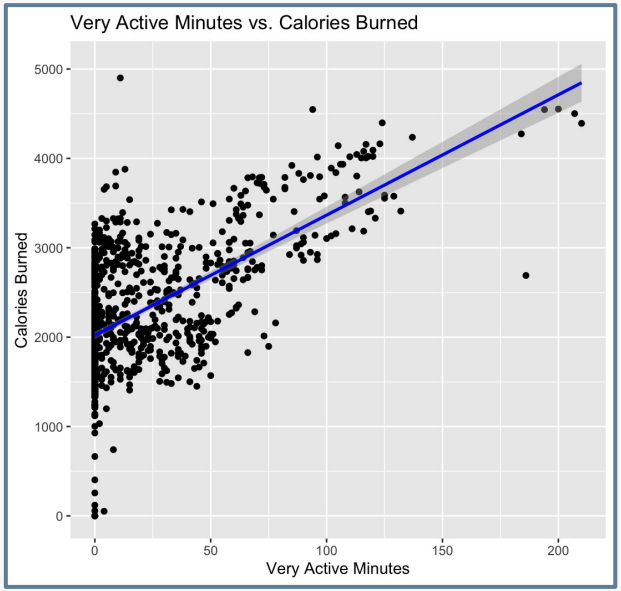
```
# Query to get relevant data
activity_data <- sqldf("SELECT TotalSteps, VeryActiveMinutes, LightlyActiveMinutes, Calories
                        FROM activity")
```

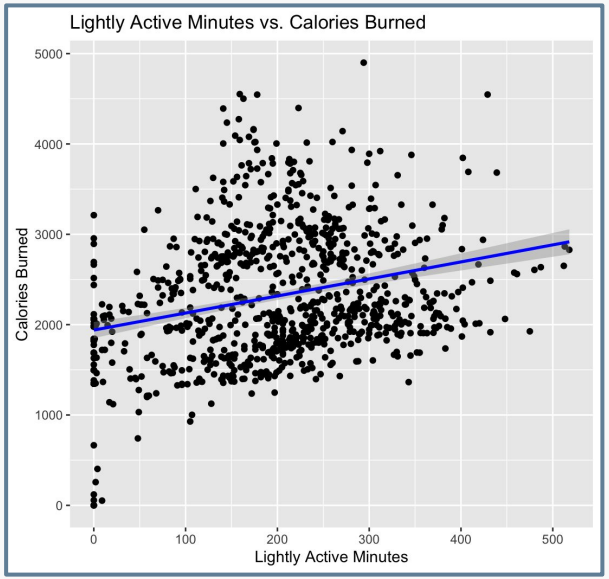
```
# Print correlations
print(paste("Correlation between Total Steps and Calories:", cor_steps_calories))
] "Correlation between Total Steps and Calories: 0.591568086245336"
print(paste("Correlation between Very Active Minutes and Calories:", cor_veryactive_calories))
] "Correlation between Very Active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



Query 5, 6, 7:


Total Steps vs. Calories Burned
Calories Burned
Total Steps
0
1000
2000
3000
4000
5000
0
10000
20000
30000

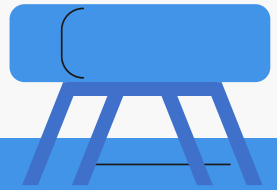

Very Active Minutes vs. Calories Burned
Calories Burned
Very Active Minutes
0
1000
2000
3000
4000
5000
0
50
100
150
200


Lightly Active Minutes vs. Calories Burned
Calories Burned
Lightly Active Minutes
0
1000
2000
3000
4000
5000
0
100
200
300
400
500

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

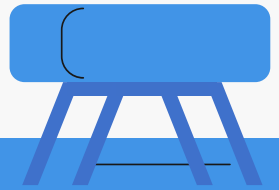


Query 9 & 10:

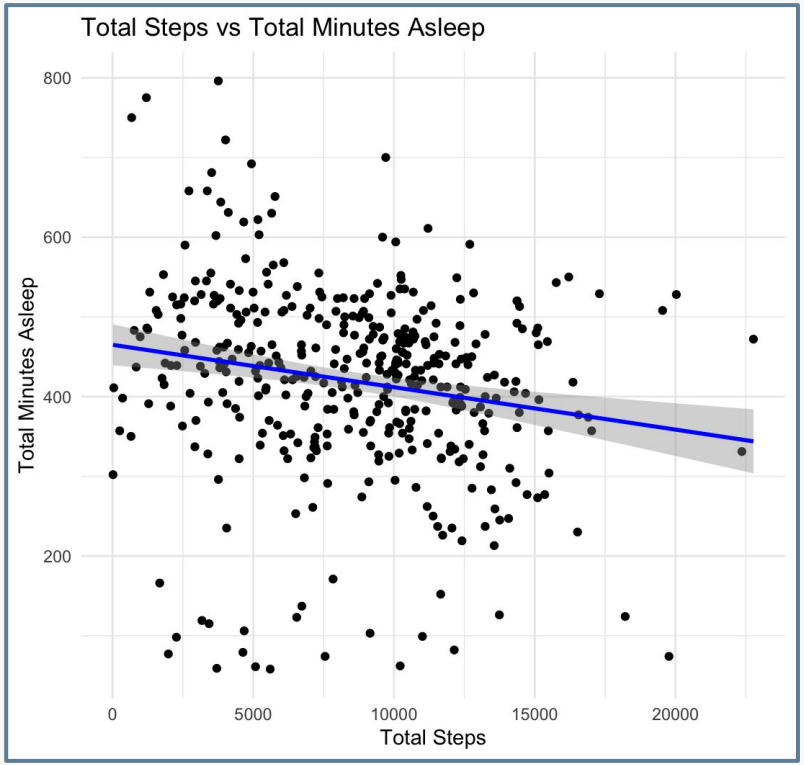
Activity & Sleep Quality:

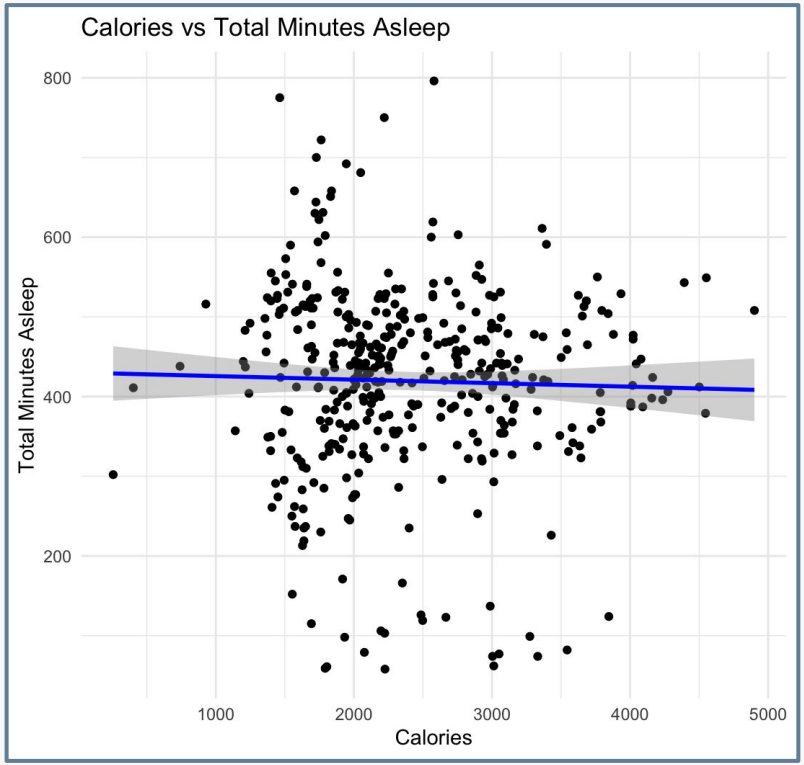
TotalSteps	TotalActiveMinutes	Calories	TotalMinutesAsleep	TotalTimeInBed
13162	366	1985	327	346
10735	257	1797	384	407
9762	272	1745	412	442
12669	267	1863	340	367
9705	222	1728	700	712
15506	345	2035	304	320

```
> # 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"
```



Query 9 & 10:

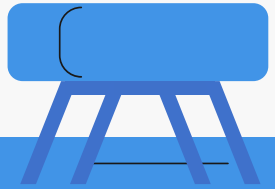

Total Steps vs Total Minutes Asleep
Total Minutes Asleep
800
600
400
200
0
5000
10000
15000
20000
Total Steps


Calories vs Total Minutes Asleep
Total Minutes Asleep
800
600
400
200
1000
2000
3000
4000
5000
Calories

Query 11:

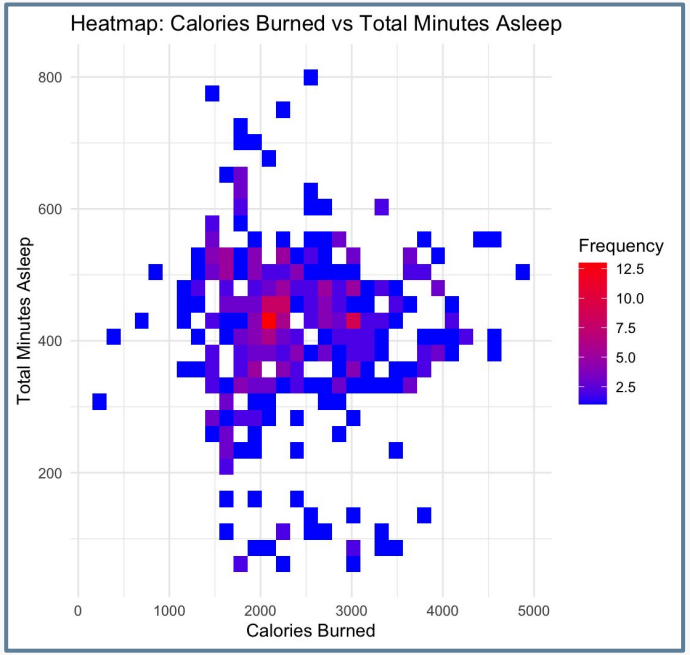
```
> head(sedentary_sleep_analysis)
  SedentaryMinutes TotalMinutesAsleep TotalTimeInBed
1              728                327            346
2              776                384            407
3              726                412            442
4              773                340            367
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"
```



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
```



Conclusion

Final remarks

- Balance matters
- Moderate exercise, combined with breaks from sedentary behavior, significantly improves calorie burn and overall activity levels.
- Users with sedentary minutes below 500 averaged around 8,000 steps daily, while those with over 700 sedentary minutes saw a sharp decline of approximately **62.5%**, averaging fewer than 3,000 steps.
- There are diminishing returns towards high physical activity levels and sleep.
- Impact of sedentary behavior.
- Encouraging regular active breaks can boost overall physical engagement and improve energy expenditure
- Individual variability
- Sleep and efficiency affected by more than physical activity; lifestyle elements.
- Below 2 miles, 20% higher variability in sleep duration.
- Set optimal goals
- Moderate activity levels of around 30-60 active minutes per day yielded the best balance

Thanks!