1. Introduction

Sleep is essential for both physical and mental well-being. It is a critical factor influencing overall health, including cognitive function, metabolic balance, and cardiovascular performance. Given the global increase in sleep disorders, it is crucial for digital health research and clinical decision-making to thoroughly understand the relationship between lifestyle factors and sleep quality.

Utilizing the Sleep Health and Lifestyle dataset from Kaggle, this project employs data-driven methodologies to explore the connections among demographic, behavioral, and physiological factors and their impact on sleep quality and duration. An interactive R Shiny dashboard is used to present the results. This tool enables users to filter data, visualize key patterns, and derive insights into sleep health at a population level. The objective is to showcase how interactive visual analytics can facilitate digital health interventions, enhance public health monitoring, and support preventive care efforts.

2. Dataset Description

The dataset comprises 374 individuals with 13 variables, encompassing demographic information, lifestyle habits, vital signs, and sleep-related outcomes. The main variables include:

- Sleep Duration (hours): a numerical measure of nightly sleep
- Quality of Sleep (1–10): subjective sleep quality rating
- Stress Level (1–10): perceived stress
- Physical Activity Level (mins/week)
- Daily Steps: average daily walking steps
- Heart Rate (bpm)
- Blood Pressure (split into systolic/diastolic)
- BMI Category: Normal, Overweight, Obese
- Sleep Disorder: None, Sleep Apnea, Insomnia

The dataset is suitable for exploring associations between lifestyle behaviors and sleep outcomes, as well as for demonstrating the value of digital health dashboards in population-level analysis.

3. Methods

3.1 Data Cleaning and Preparation

Data preparation was performed in R using the tidyverse and janitor packages. Key steps included:

- Standardizing column names with clean_names().
- Converting categorical variables (gender, occupation, BMI category, sleep disorder) into factors.
- Splitting the blood pressure field (“126/83”) into systolic and diastolic numeric variables.
- Ensuring correct numeric types for sleep duration, stress level, physical activity, daily steps, and heart rate.
- Removing rows with missing BMI categories only for the BMI-related visualizations.
- A cleaned dataset (sleep_clean.csv) was created for use in the dashboard.

3.2 Dashboard Development

The interactive dashboard was developed in R Shiny, combining reactive filtering with dynamic visualizations. Users can adjust:

- Age range
- Sleep disorder category
- BMI category

All visualizations update automatically based on these filters.

The dashboard includes four analytical plots:

1. Sleep Quality vs. Stress Level: A scatter plot showing the inverse relationship between stress and sleep quality.
2. Sleep Duration by BMI Category: A boxplot comparing sleep duration across Normal, Overweight, and Obese groups.
3. Daily Steps vs. Sleep Quality: A plot visualizing how physical activity correlates with sleep quality.
4. Sleep Duration Distribution: A histogram revealing overall sleep duration patterns in the population.

The dashboard is hosted on shinyapps.io, making it accessible online without requiring local installation.

4. Results

4.1 Stress and Sleep Quality

The scatter plot indicates a clear negative association between stress level and sleep quality. Individuals reporting high stress (7–10) frequently exhibit lower sleep quality scores. This aligns with existing research showing that psychological stress disrupts sleep architecture and reduces restorative sleep.

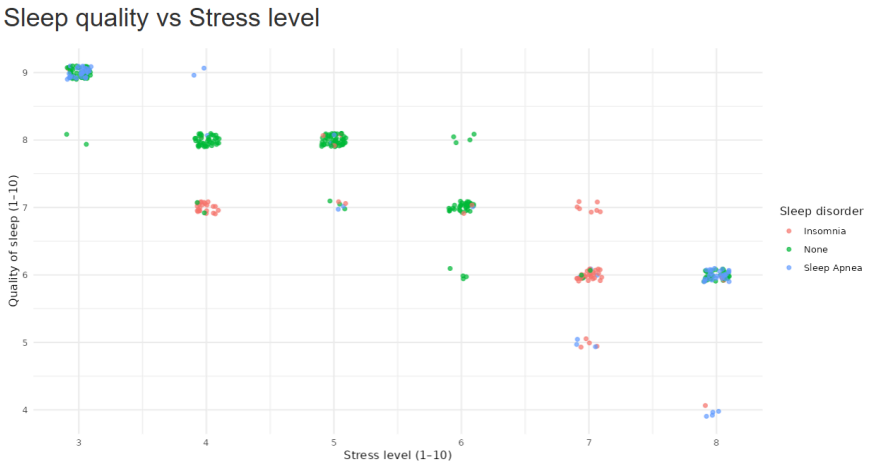

Figure 1: Relationship between stress level and sleep quality across individuals.

4.2 BMI and Sleep Duration

The boxplot reveals that individuals classified as overweight tend to have slightly shorter sleep durations, while the normal and obese groups show broader variation. After removing rows with missing BMI values, the comparison becomes clearer. Although the differences are modest, the visualization highlights how metabolic factors may relate to sleep habits.

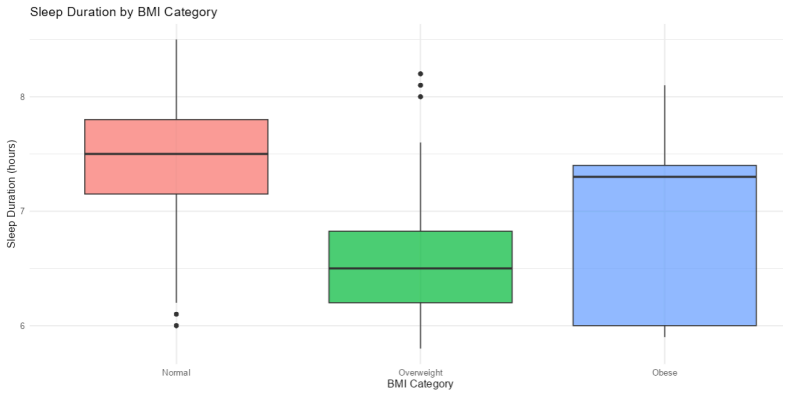

Figure 2: Distribution of sleep duration across BMI categories.

4.3 Physical Activity and Sleep

Daily steps and physical activity level show a mild positive association with sleep quality. Individuals with higher step counts tend to report better sleep scores, although the trend is not perfectly linear. This suggests that regular physical activity may contribute to improved sleep health.

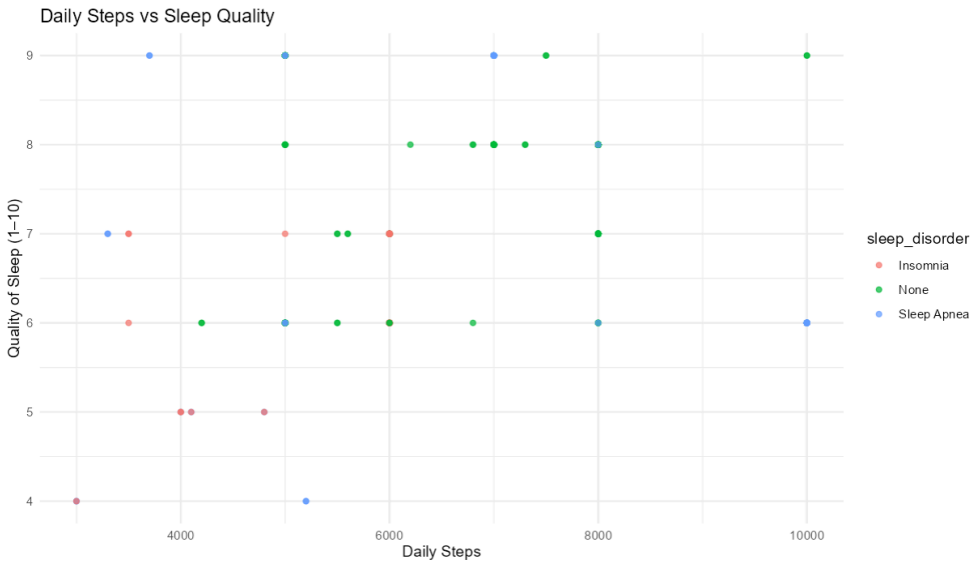

Figure 3: Association between daily steps and sleep quality.

4.4 Sleep Duration Distribution

The histogram indicates that most individuals sleep between 6 and 8 hours, with fewer cases falling below 6 hours or exceeding 9 hours. This distribution reflects typical adult sleep patterns, but also highlights subgroups at risk for short sleep duration.

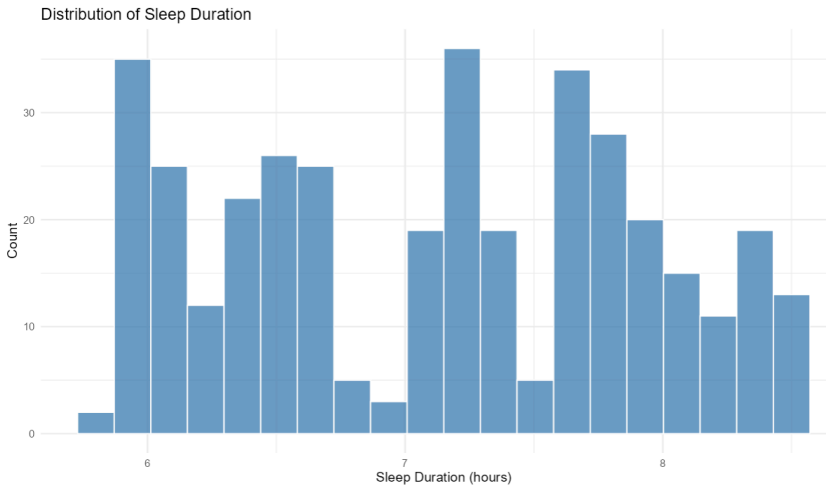

Figure 4: Distribution of sleep duration in the study population.

5. Discussion

The dashboard provides an easy-to-use visualization tool to understand how lifestyle and demographic variables interact to influence the outcome of sleep. The most consistent pattern is a negative correlation between stress and sleep quality, implying that sleep health can be improved through stress management interventions. The positive relationship between sleep quality and physical activity, albeit weak, supports the importance of exercise.

The data has some limitations, such as self-report (e.g., sleep quality, stress), a lack of demographic diversity, and a cross-sectional nature of the data. Additionally, BMI categories may not fully capture body composition differences. Nevertheless, the dashboard shows that interactive analytics can facilitate digital health app usage and facilitate data-driven health insights.

6. Conclusion

This project highlights the importance of integrating data science, visualization, and digital health concepts to examine sleep-related behaviors. The R Shiny dashboard offers an intuitive interface for examining lifestyle determinants of sleep health and can be expanded with predictive models or clinical decision-support features in future work.

7. Links

- **Shiny App:** https://kashfiaanika.shinyapps.io/digitalhealth_sleep/
- **GitHub Repository:** https://github.com/kashfiaanika/DigitalHealth-SleepHealth

Submitted by: Anika Tabassum Kashfia

.