



**Sleep Health & Lifestyle**

**Author**: Thinh Le

**Date**: 10/25/2024

**Class**: Data exploration – Fall 2024

**Instructors**:

* Dr. Clif Baldwin
* Prof. Melissa Laurino

**ABSTRACT**

In our busy daily lives, we often prioritize work, relationships, and fun activities while neglecting self-care. Have you ever wondered why some friends always seem positive and energetic, while you often feel stressed, anxious, and struggle to focus? This could be a sign of insufficient sleep.

This study investigates sleeping behavior using comprehensive data exploration techniques to uncover patterns and trends influencing sleep quality and duration. Many people stay up late to work or study, which can negatively impact their sleep health. The analysis uses data from Kaggle, designed for educational purposes, which includes information on lifestyle habits and sleep patterns.

The R programming language is used to create visually appealing graphs that make the findings easy to understand. These graphs highlight relationships between factors such as age, gender, and sleep duration, as well as the impact of work-related stress on sleep disorders. The aim is to present the data’s story in a clear and accessible way, focusing on visual explanations rather than complex details.

By sharing these insights, the project seeks to raise awareness about the importance of sleep and encourage further research on improving sleep health through better lifestyle choices.

**CONTENTS**

[TABLES iii](#_Toc184373555)

[FIGURES iv](#_Toc184373556)

[1 INTRODUCTION 1](#_Toc184373557)

[1.1 Background 1](#_Toc184373558)

[1.2 Data description 1](#_Toc184373559)

[1.3 Research objective 2](#_Toc184373560)

[1.4 Approach for Unexpected Data Outcomes 2](#_Toc184373561)

[1.5 Hypothesis 2](#_Toc184373562)

[2 DATA EXPLORATION 3](#_Toc184373563)

[2.1 Methodology 3](#_Toc184373564)

[2.1.1 Calculations and statistical analysis method 3](#_Toc184373565)

[3 RESULTS 4](#_Toc184373566)

[3.1 Variations in sleep time per day by age and gender 4](#_Toc184373567)

[3.1.1 Relationships between stress, sleep disorders, and occupation 4](#_Toc184373568)

[3.1.2 The impact of physical activity on sleep 5](#_Toc184373569)

[4 DISCUSSION 6](#_Toc184373570)

[5 CONCLUSION AND FUTURE STUDIES 7](#_Toc184373571)

[6 APPENDIX A: R CODE 8](#_Toc184373572)

[7 REFERENCES 12](#_Toc184373573)

# TABLES

[**Table 1**: Data terms 1](#_Toc183976739)

# FIGURES

[**Figure 1**: Sleep duration across age and gender 4](#_Toc184370524)

[**Figure 2**: Connection between stress and sleep disorders 4](#_Toc184370525)

[**Figure 3**: A count of people and stress levels across occupations 5](#_Toc184370526)

[**Figure 4**: The impact of physical activity on sleep 5](#_Toc184370527)

# INTRODUCTION

## Background

Sleep health is a critical component of overall mental well-being. However, many individuals face challenges in achieving quality sleep due to factors such as stress from work, insufficient physical activity, and the presence of sleep disorders.

Common sleep disorders like insomnia and sleep apnea significantly impact health and quality of life. Insomnia often leads to difficulties in falling or staying asleep, while sleep apnea is characterized by interruptions in breathing during sleep, both of which disrupt restorative rest.

Understanding how lifestyle factors influence sleep is essential for promoting healthier sleep patterns and improving overall health and well-being.

## Data description

The dataset was collected from Kaggle, the world's largest data science community, via the following link: [Kaggle Sleep Health and Lifestyle Dataset](https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset). It was combined from multiple sources by the author and designed for educational purposes. The dataset includes variables related to lifestyle factors, as well as sleep duration and quality, enabling the exploration of how lifestyle choices affect sleep health.

|  |  |
| --- | --- |
| **Person ID** | A unique identifier for each individual |
| **Gender** | The individual's gender (Male/Female) |
| **Age** | The individual's age in years |
| **Occupation** | The occupation or profession of the individual |
| **Sleep Duration** | The number of hours the individual sleeps per day |
| **Quality of Sleep** | A subjective rating of the individual’s sleep quality, ranging from 1 to 10 |
| **Physical Activity Level** | The number of minutes the individual engages in physical activity daily |
| **Stress Level** | A subjective rating of the individual's stress level, ranging from 1 to 10. |
| **Heart Rate** | The individual’s resting heart rate in beats per minute (bpm) |
| **Daily Steps** | The number of steps the individual takes per day |
| **Sleep Disorder** | The presence or absence of a sleep disorder by individual |

**Table 1**: Data terms

## Research objective

The objectives of the data exploration are to:

1. Explore the relationship between sleep length, gender, and age
2. Examine how stress affects sleep disorder, focusing on identifying occupations associated with high stress levels.
3. Analyze impact of exercise on reducing stress and improving sleep quality.
4. Providing general advice for enhancing sleep quality and recommend specific lifestyle changes that could benefit sleep quality for different groups

## Approach for Unexpected Data Outcomes

If the data do not align with the anticipated findings, the analysis will involve reassessing the data collection methods and examining potential outliers. Further research may also be conducted to refine the hypotheses and incorporate additional data if needed.

## Hypothesis

* Null Hypothesis: Age, gender, stress level, and sleep quality and duration show no significant relationship.
* Alternative Hypothesis: There is a significant relationship between age, gender, stress level, and sleep quality and duration.

# DATA EXPLORATION

## Methodology

The analysis will be conducted using RStudio on a Windows 11 platform, which provides a flexible environment for statistical computation, data manipulation, and visualization.

The dataset is pre-cleaned; no additional data cleaning will be required, and no data will be omitted from the study. However, the stress level and quality of sleep will be categorized based on numerical ranges as follows:

|  |  |
| --- | --- |
| Stress Level Categories | Quality of Sleep Categories |
| * Stress level ≤ 2: Very Low * Stress level ≤ 4: Low * Stress level ≤ 6: Moderate * Stress level ≤ 8: High * Stress level ≤ 10: Very High | * Quality of sleep ≤ 2: Very Poor * Quality of sleep ≤ 4: Poor * Quality of sleep ≤ 6: Average * Quality of sleep ≤ 8: Good * Quality of sleep ≤ 10: Excellent |

### Calculations and statistical analysis method

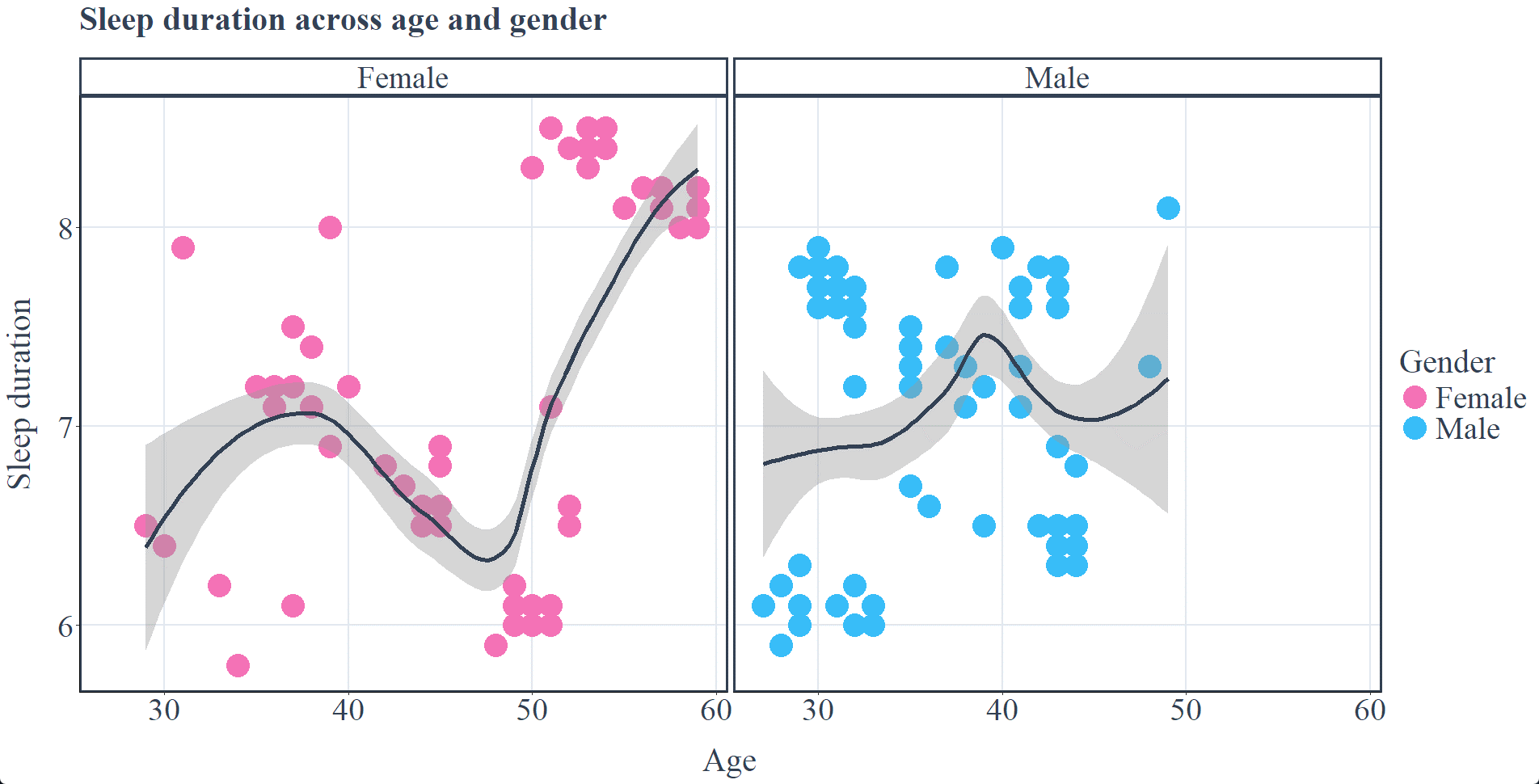
The analysis will explore relationships and patterns among key variables, including:

* *Sleep Duration by Age and Gender*: Analyzes variations in sleep length by age and gender.
* *Stress Level, Sleep Disorder and Occupation*: Analyzes the connection between stress levels, sleep disorders, and the occupations most associated with high stress.
* *Physical Activity and Sleep Quality*: Examines how activity time affects sleep quality.

These analyses may involve calculations and visualization techniques to illustrate trends and potential interactions between variables. The methods are designed to be transparent and straightforward, ensuring that anyone without a background in the subject area can follow the thought process and interpret findings.

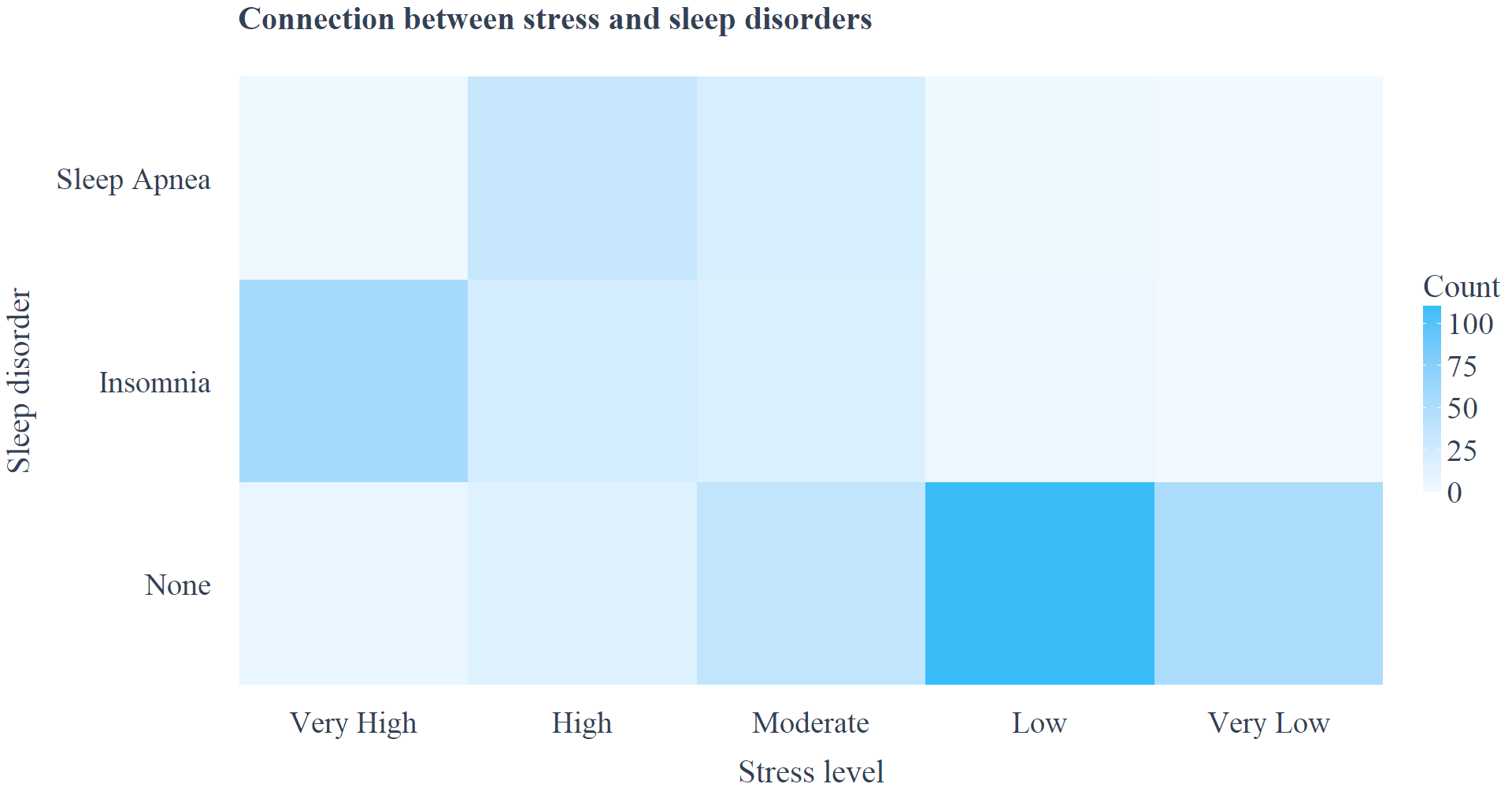
# RESULTS

## Variations in sleep time per day by age and gender

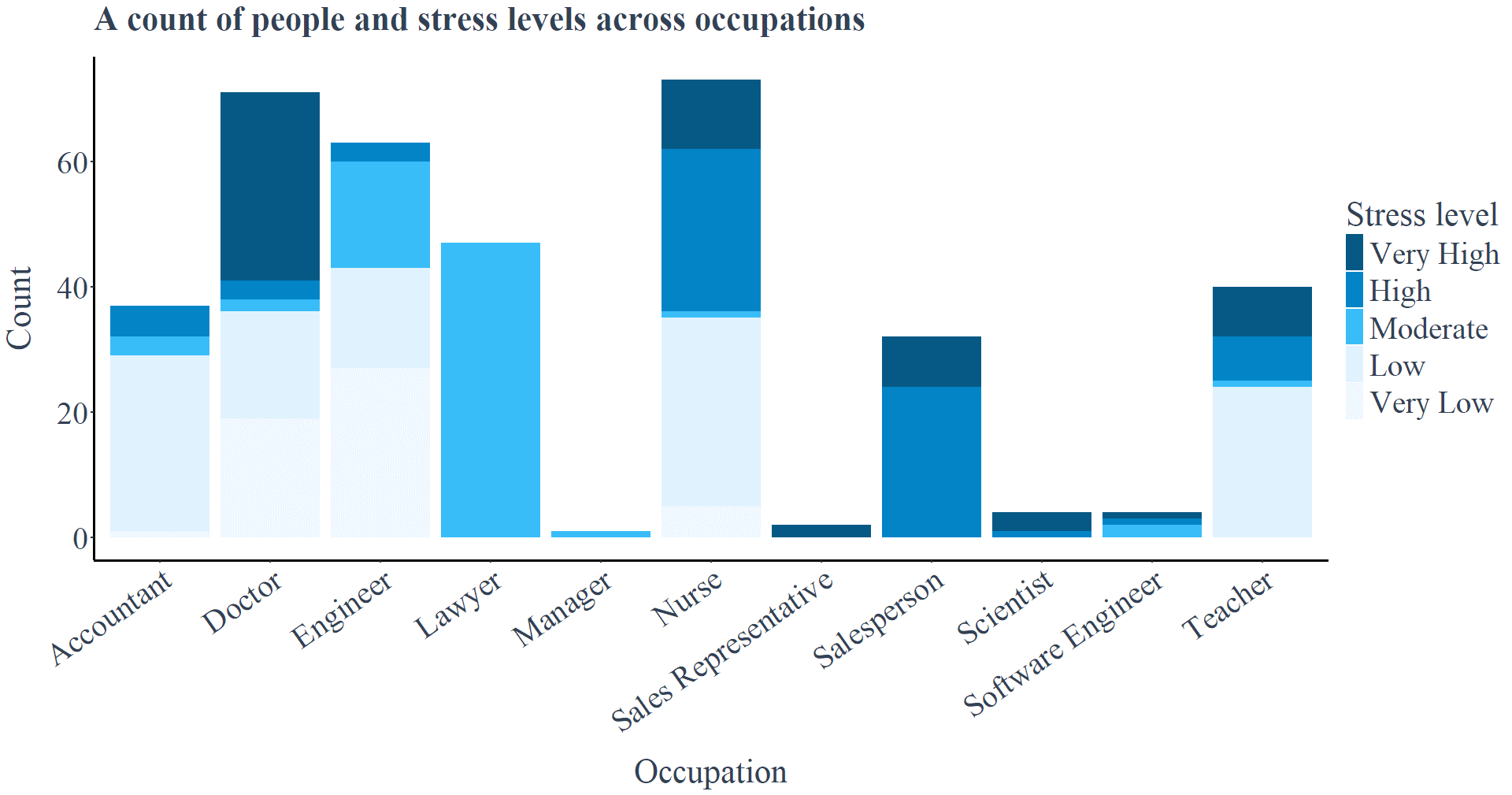


**Figure 1**: Sleep duration across age and gender

### Relationships between stress, sleep disorders, and occupation

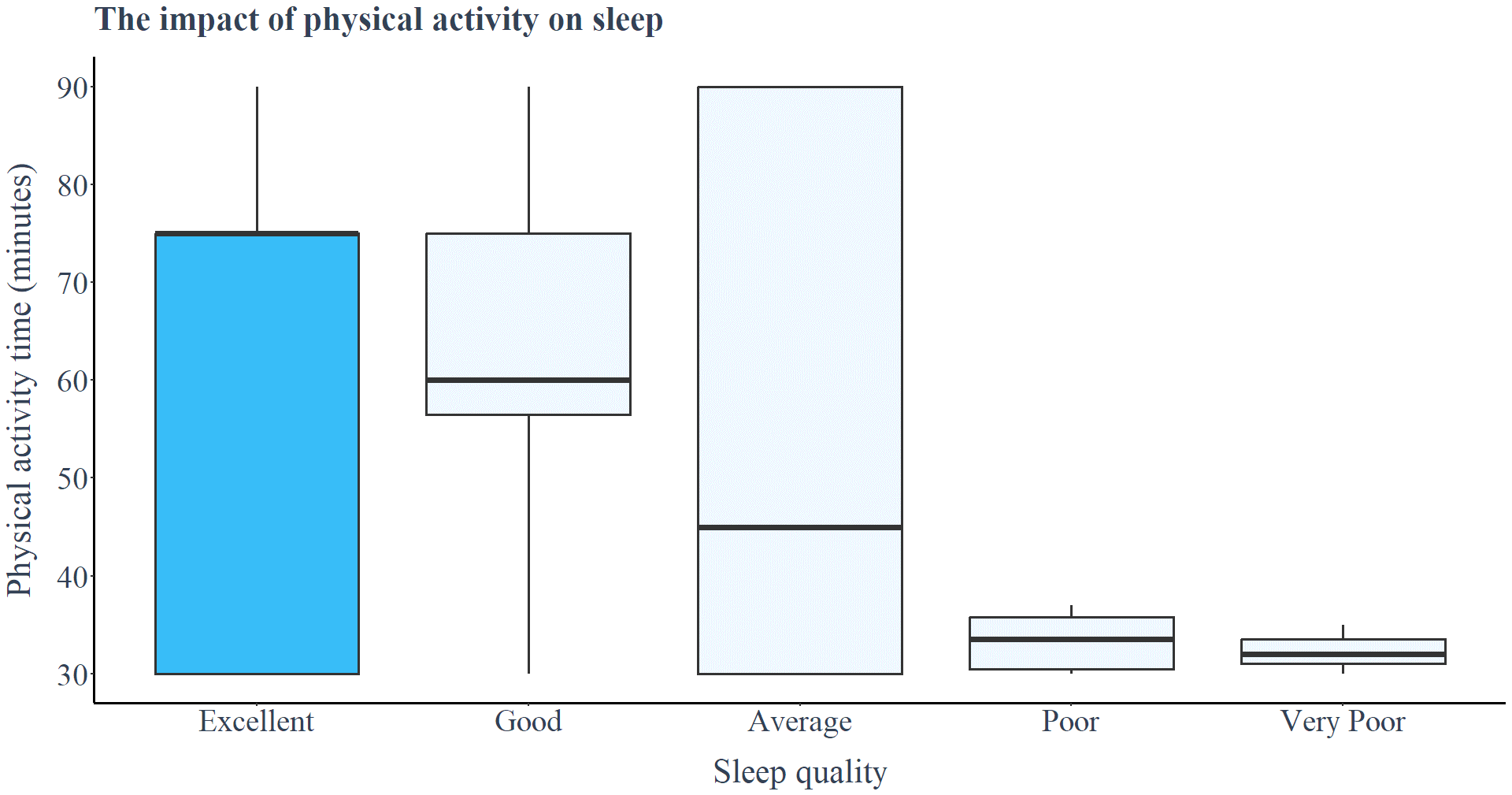


**Figure 2**: Connection between stress and sleep disorders



**Figure 3**: A count of people and stress levels across occupations

### The impact of physical activity on sleep



**Figure 4**: The impact of physical activity on sleep

# DISCUSSION

Both men and women sleep less than 7.5 hours between ages 25 and 50 (**Figure 1**). However, women’s sleep gradually increases after age 50, eventually exceeding 8 hours. For men, no specific data was available for ages above 50.

Previous studies in New Zealand show that older adults (65+) are more likely to have long sleep durations, while younger adults (<65) are more prone to short sleep durations [1]. This pattern suggests that men’s sleep duration may also increase with age, similar to women.

People with low or very low stress levels usually report no sleep disorders (**Figure 2**). Previous studies confirm that stress is the strongest cause of insomnia, followed by sleep apnea. Women are more at risk of insomnia, while men are more prone to sleep apnea [2]. Besides, certain high-stress occupations, like doctors, nurses, salespeople, scientists, software engineers, and teachers are significantly associated with sleep disorders due to elevated stress levels​ (**Figure 3**).

The role of physical activity in enhancing sleep quality is very important. Engaging in higher levels of physical activity is consistently linked to excellent sleep quality (**Figure 4**). Previous study by Schultchen et al. found a strong two-way relationship between physical activity, stress, and emotions. In addition, Physical activity not only improves sleep but also supports mental and physical health​ [3].

# CONCLUSION AND FUTURE STUDIES

Our analysis highlights the significant impact of lifestyle factors on sleep duration and quality. We explored patterns of sleep duration by gender and age, uncovering insights into how different demographics experience sleep. Jobs with high stress levels can lead to more sleep problems, showing the need to manage stress for better sleep.

Exercising regularly is a simple way to lower stress and improve both physical and mental health. A good tip for better health and sleep is to exercise more, avoid staying up late to work or study, and wake up early instead. This can help you stay healthier, sleep better, and be more productive.

Future studies could explore:

* **Technology Use**: Investigate how daily screen time on phones, computers, and social media apps affects sleep, especially due to blue light and mental stimulation.
* **Alcohol and Substances**: Examine how alcohol and other substances impact sleep, focusing on timing and consumption levels.
* **Eating Habits**: Analyze late-night eating and meal choices to determine their influence on sleep quality and duration.

# APPENDIX A: R CODE

library(dplyr) # Work with data frames

library(ggplot2) # Plot graphs

library(treemapify) # Plot treemap

library(tidyr) # Create tidy data

# Create colors

slate\_200 <- "#e2e8f0"

slate\_700 <- "#334155"

pink\_400 <- "#f472b6"

sky\_800 <- "#075985"

sky\_600 <- "#0284c7"

sky\_400 <- "#38bdf8"

sky\_100 <- "#e0f2fe"

sky\_50 <- "#f0f9ff"

normal\_text\_size <- 30

title\_text\_size <- 32

margin\_size <- 20

# Reads the data

df <- read.csv('sleep-health-and-lifestyle.csv')

df <- df %>%

mutate(Stress.Level = case\_when(

Stress.Level <= 2 ~ "Very Low",

Stress.Level <= 4 ~ "Low",

Stress.Level <= 6 ~ "Moderate",

Stress.Level <= 8 ~ "High",

Stress.Level <= 10 ~ "Very High"

)) %>%

mutate(Quality.of.Sleep = case\_when(

Quality.of.Sleep <= 2 ~ "Very Poor",

Quality.of.Sleep <= 4 ~ "Poor",

Quality.of.Sleep <= 6 ~ "Average",

Quality.of.Sleep <= 8 ~ "Good",

Quality.of.Sleep <= 10 ~ "Excellent"

))

df$Stress.Level <- factor(

df$Stress.Level,

levels = c("Very High", "High", "Moderate", "Low", "Very Low")

)

df$Quality.of.Sleep <- factor(

df$Quality.of.Sleep,

levels = c("Excellent", "Good", "Average", "Poor", "Very Poor")

)

df$Sleep.Disorder <- factor(

df$Sleep.Disorder,

levels = c("None", "Insomnia", "Sleep Apnea")

)

# Sleep duration across age and gender

ggplot(df,

aes(x = Age, y = Sleep.Duration, color = Gender)) +

geom\_point(size = 10) +

geom\_smooth(color = slate\_700, size = 2) +

facet\_wrap(~ Gender) +

scale\_colour\_manual(values = c(

"Female" = pink\_400,

"Male" = sky\_400

)) +

labs(title = "Sleep duration across age and gender",

x = "Age",

y = "Sleep duration") +

theme\_classic() +

theme(

text = element\_text(color = slate\_700, family = "serif"),

plot.title = element\_text(size = title\_text\_size, face = "bold",

margin = margin(b = margin\_size)),

panel.border = element\_rect(color = slate\_700, fill = NA, linewidth = 2),

panel.grid.major = element\_line(color = slate\_200, linewidth = 1),

strip.background = element\_rect(color = slate\_700, fill = NA, linewidth = 2),

strip.text = element\_text(size = normal\_text\_size, color = slate\_700),

legend.key.height = unit(1.25, "cm"),

legend.text = element\_text(size = normal\_text\_size),

legend.title = element\_text(size = title\_text\_size),

axis.text = element\_text(size = normal\_text\_size, color = slate\_700),

axis.title = element\_text(size = title\_text\_size),

axis.title.x = element\_text(margin = margin(t = margin\_size)),

axis.title.y = element\_text(margin = margin(r = margin\_size))

)

# Connection between stress and sleep disorders

stress\_level\_and\_sleep\_disorder <- df %>%

count(Stress.Level, Sleep.Disorder) %>%

complete(Stress.Level, Sleep.Disorder, fill = list(n = 0))

ggplot(stress\_level\_and\_sleep\_disorder,

aes(x = Stress.Level, y = Sleep.Disorder, fill = n)) +

geom\_tile() +

scale\_fill\_gradient(low = sky\_50, high = sky\_400) +

labs(title = "Connection between stress and sleep disorders",

x = "Stress level",

y = "Sleep disorder",

fill = "Count") +

theme\_classic() +

theme(

text = element\_text(color = slate\_700, family = "serif"),

plot.title = element\_text(size = title\_text\_size,

color = slate\_700,

face = "bold",

margin = margin(b = margin\_size, l = margin\_size)),

legend.key.height = unit(1.25, "cm"),

legend.text = element\_text(size = normal\_text\_size),

legend.title = element\_text(size = title\_text\_size),

axis.line = element\_blank(),

axis.ticks = element\_blank(),

axis.text = element\_text(size = normal\_text\_size, color = slate\_700),

axis.title = element\_text(size = title\_text\_size),

axis.title.x = element\_text(margin = margin(t = margin\_size)),

axis.title.y = element\_text(margin = margin(r = margin\_size))

)

# Stress levels by occupation

ggplot(df,

aes(x = Occupation, fill = Stress.Level)) +

geom\_bar() +

scale\_fill\_manual(

values = c("Very High" = sky\_800,

"High" = sky\_600,

"Moderate" = sky\_400,

"Low" = sky\_100,

"Very Low" = sky\_50)

) +

labs(title = "A count of people and stress levels across occupations",

x = "Occupation",

y = "Count",

fill = "Stress level") +

theme\_classic() +

theme(

text = element\_text(color = slate\_700, family = "serif"),

plot.title = element\_text(size = title\_text\_size,

color = slate\_700,

face = "bold",

margin = margin(b = margin\_size)),

legend.key.height = unit(1.25, "cm"),

legend.text = element\_text(size = normal\_text\_size),

legend.title = element\_text(size = title\_text\_size),

axis.line = element\_line(linewidth = 1.25),

axis.ticks = element\_line(linewidth = 1),

axis.text = element\_text(size = normal\_text\_size, color = slate\_700),

axis.text.x = element\_text(angle = 35, hjust = 1),

axis.title = element\_text(size = title\_text\_size),

axis.title.x = element\_text(margin = margin(t = margin\_size)),

axis.title.y = element\_text(margin = margin(r = margin\_size))

)

# Phyical activity time & sleep quality

ggplot(df,

aes(x = Quality.of.Sleep,

y = Physical.Activity.Level,

fill = ifelse(

Quality.of.Sleep == "Excellent",

"Highlighted",

"Normal"

))) +

geom\_boxplot(size = 1.25) +

scale\_y\_continuous(breaks = seq(30, 90, 10)) +

scale\_fill\_manual(

values = c("Highlighted" = "#38bdf8", "Normal" = "#f0f9ff")

) +

labs(title = "The impact of physical activity on sleep",

x = "Sleep quality",

y = "Physical activity time (minutes)") +

theme\_classic() +

theme(

legend.position = "none",

text = element\_text(color = slate\_700, family = "serif"),

plot.title = element\_text(size = title\_text\_size,

color = slate\_700,

face = "bold",

margin = margin(b = margin\_size)),

axis.line = element\_line(linewidth = 1.25),

axis.ticks = element\_line(linewidth = 1),

axis.text = element\_text(size = normal\_text\_size, color = slate\_700),

axis.title = element\_text(size = title\_text\_size),

axis.title.x = element\_text(margin = margin(t = margin\_size)),

axis.title.y = element\_text(margin = margin(r = margin\_size))

)

# REFERENCES

|  |  |
| --- | --- |
| [1] | Rosemary Gibson, Tasnima Akter, Courtney Jones, Andy Towers, "Characteristics of Atypical Sleep Durations Among Older Compared to Younger Adults: Evidence from the New Zealand Health Survey," *The Journals of Gerontology: Series A, Volume 78, Issue 10,* p. Pages 1908–1918, October 2023. |
| [2] | R. M. Merrill, "Mental Health Conditions According to Stress and Sleep Disorders," *International Journal of Environmental Research and Public Health,* vol. 19, no. 13, p. 7957, 2022. |
| [3] | Dana Schultchen, Julia Reichenberger, Theresa Mittl, Tabea R M Weh, Joshua M Smyth, Jens Blechert, Olga Pollatos, "Bidirectional relationship of stress and affect with physical activity and healthy eating," *Br J Health Psychol,* vol. 24, no. 2, pp. 315-333, 2019. |