



**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 spend most of our time on work, relationships, and fun activities, but we often forget to care for ourselves. Have you ever wondered why some friends always seem positive and energetic, while you often feel stressed, anxious, and struggle to focus? This might be a sign that you’re not getting enough sleep.

This study does not focus on why sleep is important. Instead, it investigates the sleeping behavior of individuals through comprehensive data exploration techniques, aiming to identify patterns and trends that influence sleep quality and duration. The goal is to present the story behind the data in a way that allows people without prior knowledge to easily understand it through graphs and analysis, without the need for detailed explanations.

**CONTENTS**

[TABLES i](#_Toc183976772)

[FIGURES i](#_Toc183976773)

[1 INTRODUCTION 1](#_Toc183976774)

[1.1 Data description 1](#_Toc183976775)

[1.2 Research objective 1](#_Toc183976776)

[1.3 Approach for Unexpected Data Outcomes 1](#_Toc183976777)

[1.4 Hypothesis 1](#_Toc183976778)

[2 DATA EXPLORATION 1](#_Toc183976779)

[2.1 Methodology 1](#_Toc183976780)

[2.1.1 Data acquisition 1](#_Toc183976781)

[2.1.2 Materials and software utilized 1](#_Toc183976782)

[2.1.3 Data cleaning and inclusion criteria 1](#_Toc183976783)

[2.1.4 Calculations and statistical analysis method 1](#_Toc183976784)

[2.2 Results 1](#_Toc183976785)

[2.2.1 Variations in sleep length by age and gender 1](#_Toc183976786)

[2.2.2 Relationships between stress, sleep disorders, and occupation 1](#_Toc183976787)

[2.2.3 How doing physical activities affects sleep quality 1](#_Toc183976788)

[2.3 Discussion 1](#_Toc183976789)

[2.4 Conclusion and future studies 1](#_Toc183976790)

[3 APPENDIX A: R CODE FOR DATA ANALYSIS AND GRAPH CREATION 1](#_Toc183976791)

[4 REFERENCES **Error! Bookmark not defined.**](#_Toc183976792)

# TABLES

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

# FIGURES

[**Figure 1**: Sleep duration across age and gender 1](#_Toc183976740)

[**Figure 2**: Connection between stress and sleep disorders 1](#_Toc183976741)

[**Figure 3**: Stress levels by occupation 1](#_Toc183976742)

[**Figure 4**: How physical activity time affects sleep quality 1](#_Toc183976743)

# INTRODUCTION

## Data description

|  |  |
| --- | --- |
| **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. |
| **BMI Category** | The BMI category of the person (e.g., Underweight, Normal, Overweight). |
| **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   * **None**: No sleep problems. * **Insomnia**: The individual experiences difficulty falling asleep or staying asleep. * **Sleep Apnea**: Breathing stops briefly during sleep, often causing snoring and tiredness. |

**Table 1**: Data terms

## Research objective

The objectives of the data exploration are to:

1. Explore the relationship between sleep quality and other factors, like gender, age, and physical activity level.
2. To examine the impact of exercise on reducing stress and improving sleep quality.
3. 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

This section describes the methodology for obtaining and processing data, including calculations, statistical analyses, and a discussion of the results.

## Methodology

### Data acquisition

The dataset was collected from Kaggle via the following link: [Kaggle Sleep Health and Lifestyle Dataset](https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset). The dataset contains 374 rows and includes variables capturing various aspects of individuals' lifestyle choices and health metrics.

### Materials and software utilized

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

### Data cleaning and inclusion criteria

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

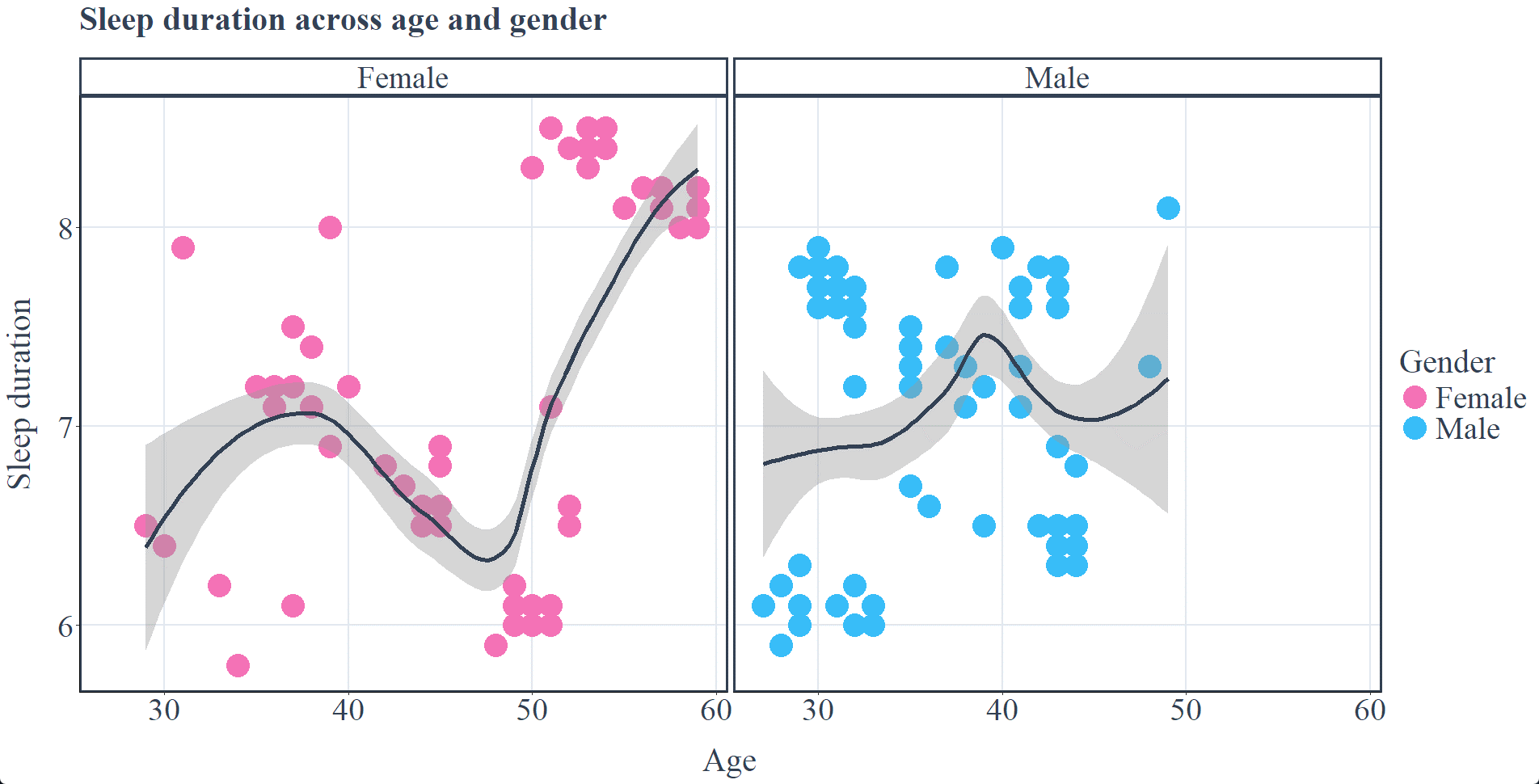
The analysis will explore relationships and patterns among key variables, including:

* *Sleep Duration by Age and Gender*: Analyzes variations in sleep length by age and gender.
* *Stress Level, Sleep Disorder and Occupation*: Explores the relationships between stress, sleep disorders, and occupation.
* *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

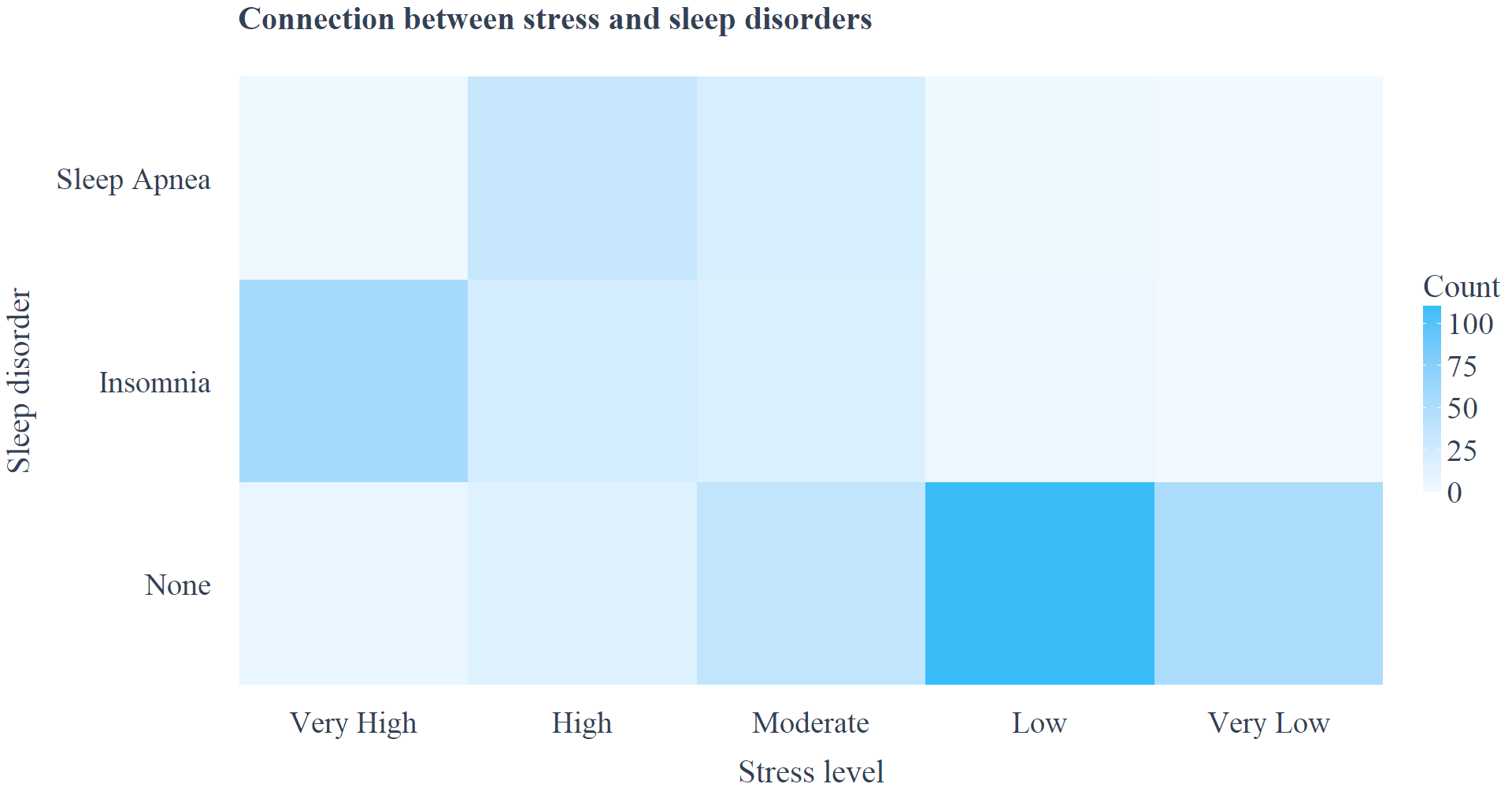
### Variations in sleep length by age and gender



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

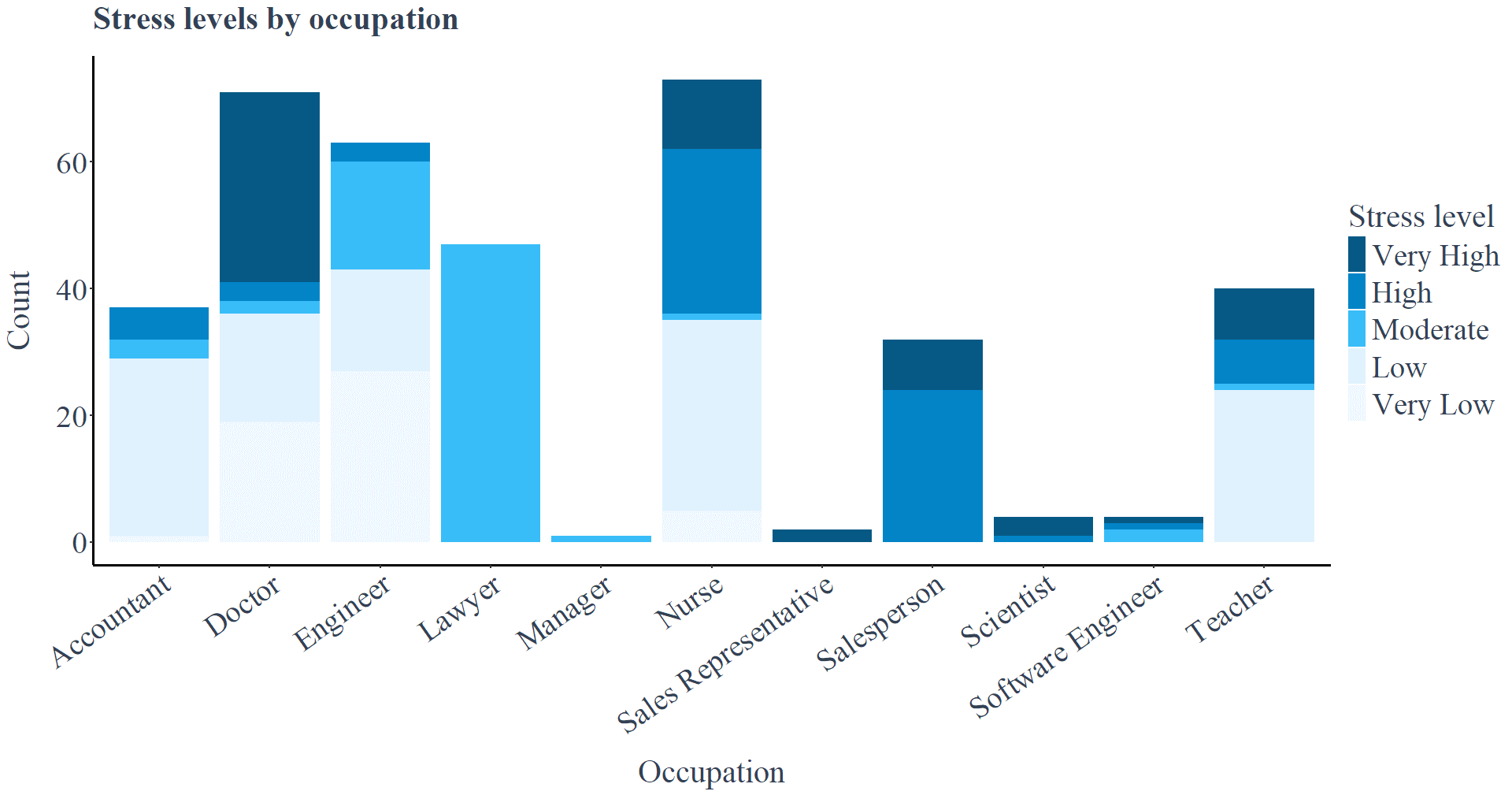
The graphs show sleep duration by age for females and males.

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



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

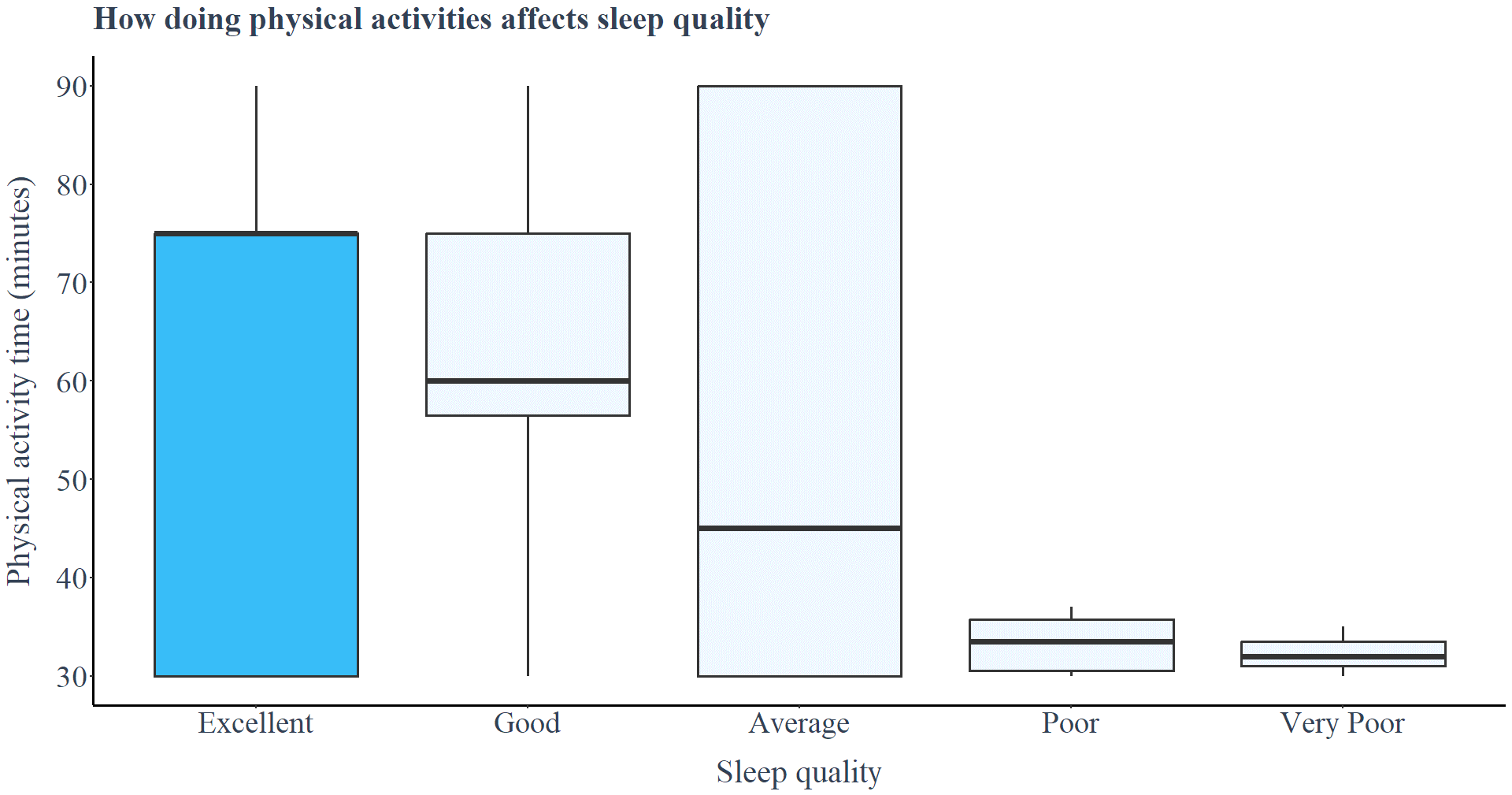
This heatmap highlights the impact of stress on sleep disorders.



**Figure 3**: Stress levels by occupation

This stacked bar chart shows the number of people and stress levels across occupations.

### The impact of physical activity on sleep



**Figure 4**: How physical activity time affects sleep quality

## Discussion

Sleep length changes with age and differs between men and women. In **Figure 1**, both men and women sleep less than 7.5 hours between ages 25 and 50. However, women’s sleep gradually increases after age 50, eventually exceeding 8 hours. For men, no specific data was available for older ages.

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

**Figure 2** shows a clear link between stress and sleep disorders. People with low or very low stress levels usually report no sleep disorders. On the other hand, high stress levels are linked to sleep apnea, and very high stress levels often cause insomnia. Previous studies have confirmed 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]. Certain jobs, like doctors, nurses, salespeople, scientists, software engineers, and teachers, are associated with very high stress levels, which can worsen sleep problems​(**Figure 3**).

People who exercise regularly tend to sleep better (Figure 4). It also helps reduce stress and negative emotions, boosting positive feelings. A study by Schultchen et al. found a strong two-way relationship between physical activity, stress, and emotions. Physical activity not only improves sleep but also supports mental and physical health​ [3].

## Conclusion and future studies

Age, gender, stress, and physical activity all play key roles in sleep quality. Understanding these factors can help individuals make better choices to improve their sleep and overall health.

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 FOR DATA ANALYSIS AND GRAPH CREATION

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 <- 29

title\_text\_size <- 31

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" = "#075985",

"High" = "#0284c7",

"Moderate" = "#38bdf8",

"Low" = "#e0f2fe",

"Very Low" = "#f0f9ff")

) +

labs(title = "Stress levels by occupation",

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 = "How doing physical activities affects sleep quality",

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. |