Sleep Health and Lifestyle Dataset

Code

Student Name: Yeung Lai

Dataset :

https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset (https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset)

Introduction

This project will use “Sleep Health and Lifestyle” Dataset as original data to analysis. The dataset is covering a various variable that related to sleep and daily habits. It contains all details such as gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders. With 374 test subjects in total and sleep patterns,13 variables in the whole dataset offers a detailed and nuanced view of individuals’ lives and their relationship with sleep.

Sleep health is a fundamental aspect of our well-being, influencing both physical and mental functions. Poor sleep can disrupt our daily lives and hinder our ability to focus. Furthermore, the quality of sleep, whether good or bad, may also impact our physical condition like blood pressure and heart rate. By studying this data, we can understand the relationships between lifestyle and sleep health.

The dataset enables us to recognize patterns and relationships between these two aspects. For instance, examining how different occupations or stress levels correlate with sleep duration and quality provides valuable insights. The dataset provides a wide-ranging scope for unveiling the relation between lifestyle choices and sleep health.

Hide

```
# Yeung Lai M23W0573
library(GGally)
library(ggplot2)
library(tidyverse)
library(readr)
library(tidyverse)
library(viridisLite)

sleep_health_and_lifestyle_dataset <- read_csv("/Users/laiyeung/Documents/KCG/Data An
lysis/Sleep_health_and_lifestyle_dataset.csv")
```

```
Rows: 374 Columns: 13── Column specification ────────────────────────────────────────
────────────────────
Delimiter: ","
chr (5): Gender, Occupation, BMI Category, Blood Pressure, Sleep Disorder
dbl (8): Person ID, Age, Sleep Duration, Quality of Sleep, Physical Activity Le...
ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
View(sleep_health_and_lifestyle_dataset)
```

Research Question and Hypotheses

The project will ask the following research question and hypotheses:

Research Question

• Does the level of physical activity have a significant impact on the quality of sleep?

Hypotheses

• H0: There is no significant correlation between the levels of daily physical activity and sleep quality.

• H1: There is a significant positive correlation between the levels of daily physical activity and sleep quality.

Visualization 1

To study our research question, we need to know the relation between Sleep Duration and Physical Activity. Scatter Plot can help to visualize the relationship, allow us to observe if there is any discernible pattern or trend between these two variables.

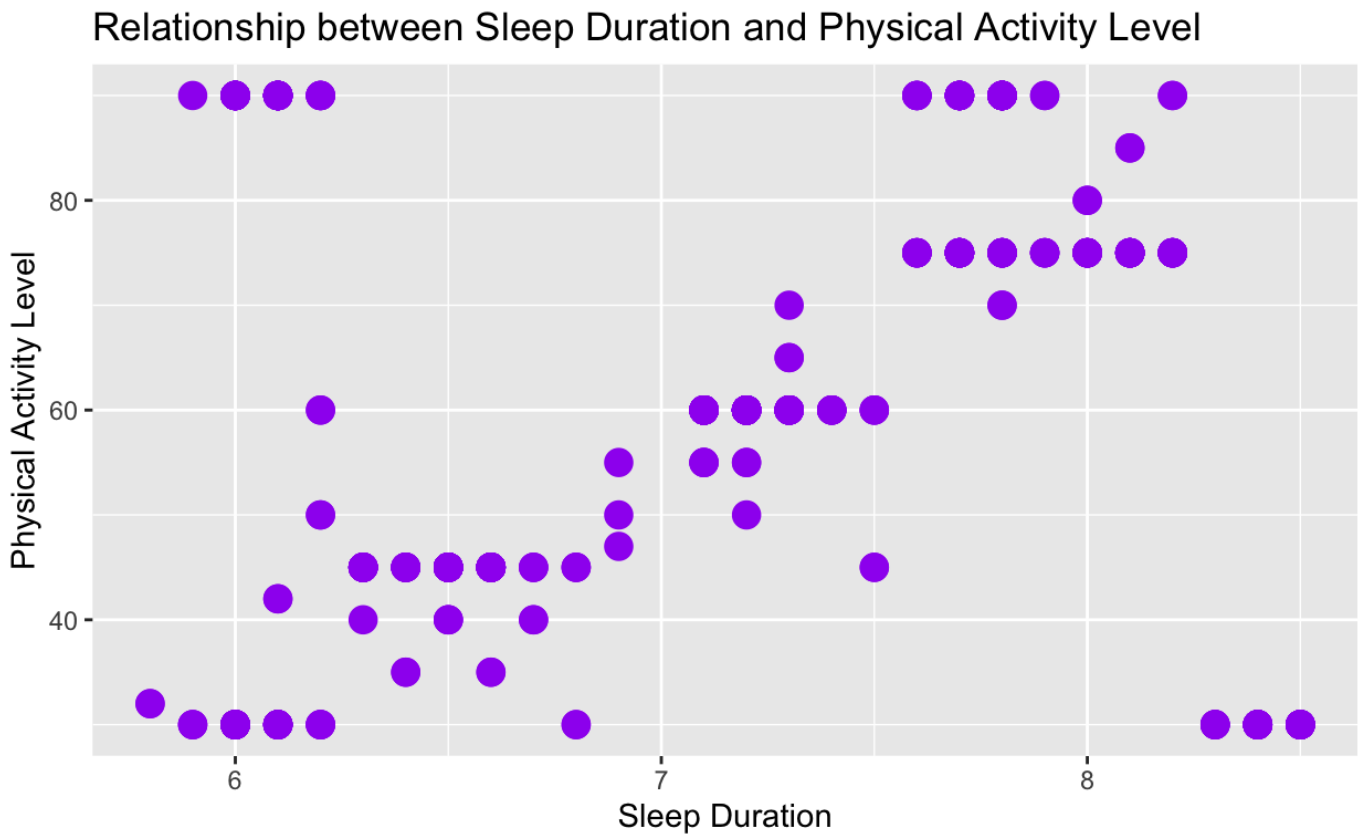
Hide

```
#CORRELATION Analysis
cor_matrix <- cor(sleep_health_and_lifestyle_dataset[, c("Sleep Duration", "Physical
Activity Level", "Quality of Sleep")])
print(cor_matrix)
```

```
                       Sleep Duration Physical Activity Level Quality of Sleep
Sleep Duration              1.0000000               0.2123603        0.8832130
Physical Activity Level     0.2123603               1.0000000        0.1928965
Quality of Sleep            0.8832130               0.1928965        1.0000000
```

Hide

```
#Scatter plot
sleep_health_and_lifestyle_dataset%>%
  ggplot(aes(`Sleep Duration`,`Physical Activity Level`))+
  geom_point(col="purple",size=4)+
  labs(x = "Sleep Duration", y = "Physical Activity Level",
       title = "Relationship between Sleep Duration and Physical Activity Level")
```



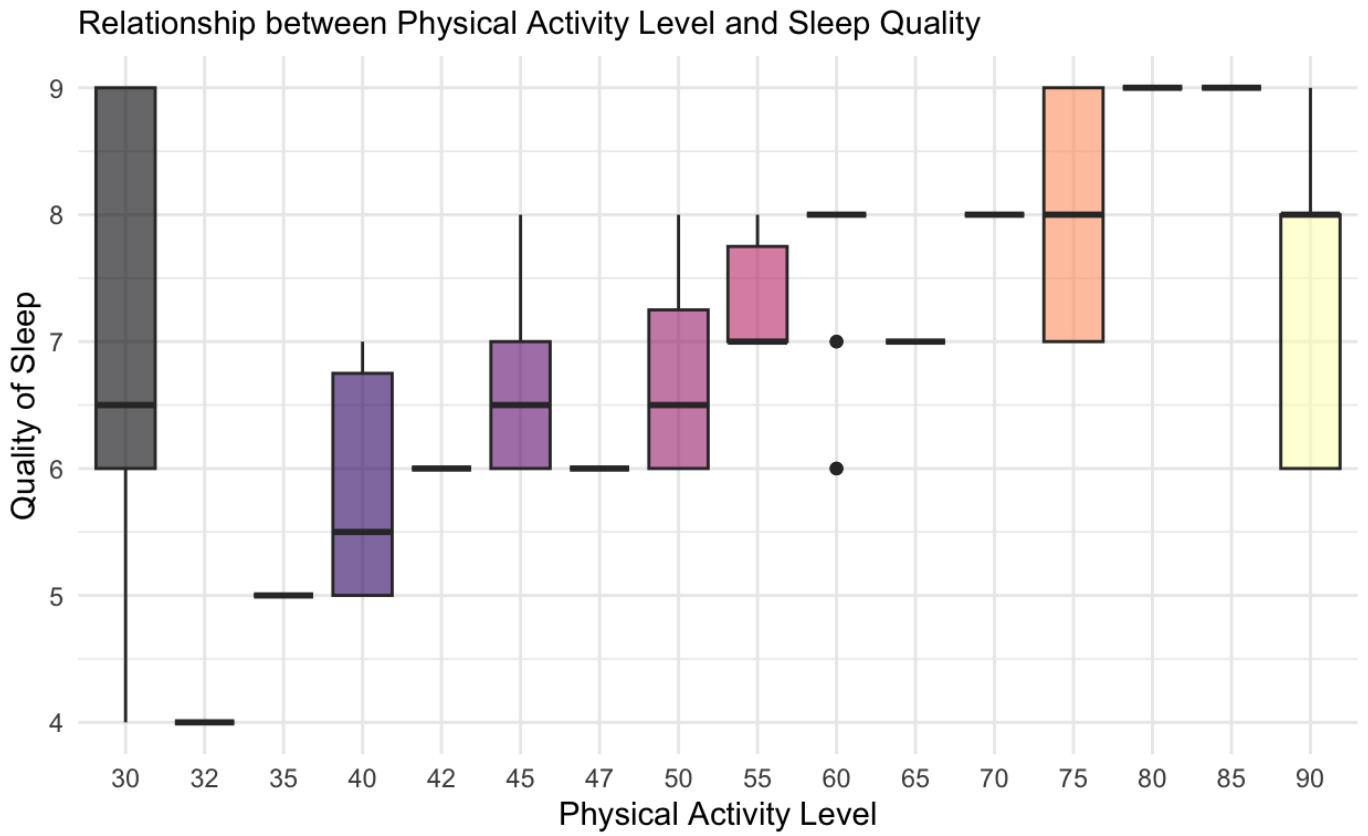
Scatter plot with title “Relationship between Sleep Duration and Physical Activity Level”. The x-axis represents sleep duration with the number of hours the person sleeps per day, the y-axis show physical activity level with minutes of the person engages in physical activity daily. Each point on the plot stand for an observation of the dataset with the correlation matrix. In the plot show a trend line with a positive slope, as physical activity level increases the sleep duration will also correlativity increase.

Visualization 2

Box plot provides a clear comparison of the rating of sleep quality across duration of physical activity.

Hide

```
# Box plot
ggplot(sleep_health_and_lifestyle_dataset, aes(x = as.factor(`Physical Activity Level
`), y = `Quality of Sleep`, fill = as.factor(`Physical Activity Level`))) +
  geom_boxplot() +
  scale_fill_viridis(discrete = TRUE, alpha = 0.6, option = "A") +
  theme_minimal() +  # You can change the theme as per your preference
  theme(
    legend.position = "none",
    plot.title = element_text(size = 11)
  ) +
  ggtitle("Relationship between Physical Activity Level and Sleep Quality") +
  xlab("Physical Activity Level") +
  ylab("Quality of Sleep")
```



In the boxplot titled “Relationship between Physical Activity Level and Sleep Quality,” the x-axis represents the physical activity level, measured in the number of minutes a person engages in physical activity daily. The y-axis represents the rating of sleep quality, ranging from 1 to 10. Upon observation, it is evident that having over 30 minutes of daily physical activity correlates with a sleep quality rating around 6-7. Within the range of 40-60 minutes of daily physical activity, there is an observable increasing trend in the sleep quality rating. However, engaging in over 90 minutes of physical activity results in a decrease in the sleep quality rating.

Statistical tests Parametric

Hide

```
# Perform Pearson correlation
correlation_result <- cor.test(sleep_health_and_lifestyle_dataset$`Physical Activity
Level`, sleep_health_and_lifestyle_dataset$`Quality of Sleep`, method = "pearson")

# Print the result
print(correlation_result)
```

```
    Pearson's product-moment correlation

data:  sleep_health_and_lifestyle_dataset$`Physical Activity Level` and sleep_health_
and_lifestyle_dataset$`Quality of Sleep`
t = 3.7917, df = 372, p-value = 0.0001745
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.09331529 0.28865656
sample estimates:
      cor
0.1928965
```

For the result, the correlation coefficient of approximately 0.193 between the variables “Physical Activity Level” and “Quality of Sleep.” The t-statistic is 3.7917, and the corresponding p-value is 0.0001745, which is below the commonly accepted significance level of 0.05. The 95 percent confidence interval for the correlation coefficient is between 0.0933 and 0.2887.

The low p-value suggests the rejection of the null hypothesis, providing evidence in support of a significant positive correlation between daily physical activity levels and sleep quality.

Statistical tests Non-Parametric

Hide

```
# Perform the Mann Whitney U test
mwu_test_result <- wilcox.test(sleep_health_and_lifestyle_dataset$`Physical Activity
Level`,sleep_health_and_lifestyle_dataset$`Quality of Sleep`)

# Print the results
print(mwu_test_result)
```

```

    Wilcoxon rank sum test with continuity correction

data:  sleep_health_and_lifestyle_dataset$`Physical Activity Level` and sleep_health_
and_lifestyle_dataset$`Quality of Sleep`
W = 139876, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

For the Mann-Whitney U test, we used a test to compare two groups: Physical Activity Level and Quality of Sleep. The test strongly suggests that there’s a significant difference in Physical Activity Level between groups with different Quality of Sleep. The test statistic W=139876 indicate the different between the groups. The p-value <2.2e-16, which is very close to zero and much smaller than 0.05, indicates that we reject the null hypothesis, suggesting a significant difference in Physical Activity Level between groups with different Quality of Sleep. The alternative hypothesis “true location shift is not equal to 0” suggests that there might be a real difference.

In summary, in the parametric test, Pearson’s correlation, and the non-parametric test, Mann Whitney U test, both tests provide evidence to support the relationship between “Physical Activity Level” and “Quality of Sleep.” Pearson’s correlation suggests a positive correlation between those two variables, while the Mann-Whitney U test indicates differences in “Physical Activity Level” among groups with different “Quality of Sleep.” Both results support the hypothesis that there is a significant positive correlation between the levels of daily physical activity and sleep quality.

Synopsis

In this project, we use the “Sleep Health and Lifestyle” dataset to investigate a specific research question and formulate hypotheses. Our focus was on exploring the relationship between physical activity and sleep quality. By practice statistical tests and visualizations using R, we successfully validated our hypotheses.

Our research findings indicate that an increase in physical activity minutes correlates positively with sleep duration. However, it was observed that excessively long physical activity durations could also have an adverse impact on sleep quality.

The “Sleep Health and Lifestyle” dataset, with its various features, provided a rich source of information. Further exploration of this dataset has more potential to reveal additional information or allow for the expansion of our study.

Citations:

https://www.kaggle.com/code/natecekay/an-analysis-of-health-and-lifestyle-metrics (https://www.kaggle.com/code/natecekay/an-analysis-of-health-and-lifestyle-metrics)

https://www.data-to-viz.com/ (https://www.data-to-viz.com/)

https://mathsathome.com/understand-and-compare-box-plots/ (https://mathsathome.com/understand-and-compare-box-plots/)

https://medium.com/@bcholig/understanding-sleep-health-and-lifestyle-exploring-the-interplay-of-sleep-duration-physical-2dfa7819f094 (https://medium.com/@bcholig/understanding-sleep-health-and-lifestyle-exploring-the-interplay-of-sleep-duration-physical-2dfa7819f094)