title: " How does sleep duration affect Anxiety level in male and female students?"

author: “Ifreador Aziagbe”

date: “2024-10-23”

Project 1

output: word\_document

Introduction

How does sleep duration affect Anxiety levels in male and female students? To respond to this question, I will analyze the dataset that can acknowledge the understanding of the mechanism, which is the ’ Student Mental Health Survey’. The dataset I used comes from Kaggle and contains twenty-one column variables such as sleep duration, gender identity, Student age, feeling anxiety, degree level, etc. For this analysis, I will focus on “sleep duration” and “anxiety levels” for “male” and “female” students to determine if there is a difference in how sleep affects anxiety across genders.”

Reference: <https://www.kaggle.com/datasets/abdullahashfaqvirk/student-mental-health-survey>

Data analysis  
   
In this analysis, I will start by cleaning the dataset to remove any missing or incomplete data, so that the results are accurate. Then, I will perform exploratory data analysis (EDA) to get a better understanding of the data. I will look at the overall patterns of sleep duration and anxiety levels for both male and female students. I will calculate key statistics, like the average (mean) and highest (maximum) values for sleep duration and anxiety. To visualize the data, I will use scatter plots to see how sleep duration affects anxiety levels for males and females. Additionally, I will create histograms to show the distribution of anxiety levels and sleep duration for each gender group.

# 1 library and load the dataset

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)  
library(dplyr)  
library(readr)  
Student\_data <- read.csv("C:/Users/YAO GEDEON/Downloads/Project 1data 101/MentalHealthSurvey new.csv")  
  
# clean of Dataset   
  
str(Student\_data)

## 'data.frame': 87 obs. of 21 variables:  
## $ gender : chr "Female" "Female" "Male" "Male" ...  
## $ age : int 17 17 17 18 18 18 18 18 18 18 ...  
## $ university : chr "PU" "PU" "PU" "PU" ...  
## $ degree\_level : chr "Undergraduate" "Undergraduate" "Undergraduate" "Undergraduate" ...  
## $ degree\_major : chr "Computer Science" "Computer Science" "Information Technology" "Data Science" ...  
## $ academic\_year : chr "1st year" "1st year" "1st year" "1st year" ...  
## $ cgpa : chr "3.5-4.0" "3.5-4.0" "0.0-0.0" "2.5-3.0" ...  
## $ residential\_status : chr "On-Campus" "On-Campus" "Off-Campus" "Off-Campus" ...  
## $ campus\_discrimination : chr "No" "No" "No" "No" ...  
## $ sports\_engagement : chr "No Sports" "No Sports" "7+ times" "4-6 times" ...  
## $ average\_sleep : chr "4-6 hrs" "4-6 hrs" "7-8 hrs" "4-6 hrs" ...  
## $ study\_satisfaction : int 4 4 3 4 5 5 3 1 2 4 ...  
## $ academic\_workload : int 5 4 5 3 3 4 3 4 4 3 ...  
## $ academic\_pressure : int 5 4 5 4 1 5 4 5 4 3 ...  
## $ financial\_concerns : int 5 5 1 4 1 2 5 5 5 2 ...  
## $ social\_relationships : int 1 3 3 4 3 3 2 2 1 3 ...  
## $ depression : int 5 4 4 3 1 4 4 5 5 3 ...  
## $ anxiety : int 5 4 4 4 2 3 3 5 4 3 ...  
## $ isolation : int 5 5 4 4 2 4 4 5 5 3 ...  
## $ future\_insecurity : int 3 2 5 2 1 4 5 5 5 4 ...  
## $ stress\_relief\_activities: chr "Sleep" "Religious Activities" "Outdoor Activities" "Sports and Fitness, Creative Outlets, Online Entertainment, Outdoor Activities" ...

head(Student\_data)

## gender age university degree\_level degree\_major academic\_year  
## 1 Female 17 PU Undergraduate Computer Science 1st year  
## 2 Female 17 PU Undergraduate Computer Science 1st year  
## 3 Male 17 PU Undergraduate Information Technology 1st year  
## 4 Male 18 PU Undergraduate Data Science 1st year  
## 5 Male 18 PU Undergraduate Data Science 1st year  
## 6 Male 18 PU Undergraduate Data Science 1st year  
## cgpa residential\_status campus\_discrimination sports\_engagement  
## 1 3.5-4.0 On-Campus No No Sports  
## 2 3.5-4.0 On-Campus No No Sports  
## 3 0.0-0.0 Off-Campus No 7+ times  
## 4 2.5-3.0 Off-Campus No 4-6 times  
## 5 3.5-4.0 Off-Campus No No Sports  
## 6 1.5-2.0 On-Campus No No Sports  
## average\_sleep study\_satisfaction academic\_workload academic\_pressure  
## 1 4-6 hrs 4 5 5  
## 2 4-6 hrs 4 4 4  
## 3 7-8 hrs 3 5 5  
## 4 4-6 hrs 4 3 4  
## 5 7-8 hrs 5 3 1  
## 6 4-6 hrs 5 4 5  
## financial\_concerns social\_relationships depression anxiety isolation  
## 1 5 1 5 5 5  
## 2 5 3 4 4 5  
## 3 1 3 4 4 4  
## 4 4 4 3 4 4  
## 5 1 3 1 2 2  
## 6 2 3 4 3 4  
## future\_insecurity  
## 1 3  
## 2 2  
## 3 5  
## 4 2  
## 5 1  
## 6 4  
## stress\_relief\_activities  
## 1 Sleep  
## 2 Religious Activities  
## 3 Outdoor Activities  
## 4 Sports and Fitness, Creative Outlets, Online Entertainment, Outdoor Activities  
## 5 Sleep  
## 6 Religious Activities

tail(Student\_data)

## gender age university degree\_level degree\_major academic\_year  
## 82 Female 22 COMSATS Undergraduate Software Engineering 4th year  
## 83 Male 22 PU Undergraduate Data Science 2nd year  
## 84 Male 23 COMSATS Undergraduate Computer Science 3rd year  
## 85 Male 23 PU Undergraduate Data Science 2nd year  
## 86 Male 26 PU Postgraduate Data Science 1st year  
## 87 Male 26 KUST Undergraduate Data Science 4th year  
## cgpa residential\_status campus\_discrimination sports\_engagement  
## 82 3.5-4.0 Off-Campus Yes No Sports  
## 83 2.5-3.0 Off-Campus No No Sports  
## 84 2.5-3.0 Off-Campus No No Sports  
## 85 2.5-3.0 Off-Campus Yes 1-3 times  
## 86 2.5-3.0 On-Campus Yes 1-3 times  
## 87 3.5-4.0 Off-Campus Yes 1-3 times  
## average\_sleep study\_satisfaction academic\_workload academic\_pressure  
## 82 2-4 hrs 5 5 5  
## 83 4-6 hrs 3 3 3  
## 84 4-6 hrs 3 5 3  
## 85 7-8 hrs 3 3 4  
## 86 7-8 hrs 4 4 4  
## 87 7-8 hrs 5 5 1  
## financial\_concerns social\_relationships depression anxiety isolation  
## 82 5 1 5 5 5  
## 83 3 2 3 4 4  
## 84 5 1 5 5 5  
## 85 1 4 3 3 2  
## 86 5 2 5 4 4  
## 87 5 1 5 5 3  
## future\_insecurity  
## 82 5  
## 83 5  
## 84 5  
## 85 3  
## 86 5  
## 87 5  
## stress\_relief\_activities  
## 82 Religious Activities, Sleep  
## 83 Sleep  
## 84 Sports and Fitness  
## 85 Religious Activities, Online Entertainment, Outdoor Activities  
## 86 Social Connections, Online Entertainment  
## 87 Creative Outlets, Social Connections, Online Entertainment

sum(is.na(Student\_data))

## [1] 0

colSums(is.na(Student\_data))

## gender age university   
## 0 0 0   
## degree\_level degree\_major academic\_year   
## 0 0 0   
## cgpa residential\_status campus\_discrimination   
## 0 0 0   
## sports\_engagement average\_sleep study\_satisfaction   
## 0 0 0   
## academic\_workload academic\_pressure financial\_concerns   
## 0 0 0   
## social\_relationships depression anxiety   
## 0 0 0   
## isolation future\_insecurity stress\_relief\_activities   
## 0 0 0

library(tidyverse) library(ggplot2) library(dplyr) library(readr) Student\_data <- read.csv(“C:/Users/YAO GEDEON/Downloads/Project 1data 101/MentalHealthSurvey new.csv”) Student\_data # clean of Dataset

str(Student\_data) head(Student\_data) tail(Student\_data) sum(is.na(Student\_data)) colSums(is.na(Student\_data))

# 2--Calculate the average sleep duration by gender  
Student\_data <- Student\_data %>%  
 mutate(  
 average\_sleep\_numeric = case\_when(  
 average\_sleep == "4-6 hrs" ~ 5.5,  
 average\_sleep == "7-8 hrs" ~ 7.5,  
 average\_sleep == "2-4 hrs" ~ 2.5,   
 TRUE ~ NA\_real\_   
 )  
 )  
  
Student\_data$average\_sleep <- as.numeric(as.character(Student\_data$average\_sleep))

## Warning: NAs introduced by coercion

Student\_data$anxiety\_level <- as.numeric(as.character(Student\_data$anxiety))  
  
# Check if there are any NA values generated by this conversion  
sum(is.na(Student\_data$average\_sleep\_numeric))

## [1] 0

sum(is.na(Student\_data$anxiety))

## [1] 0

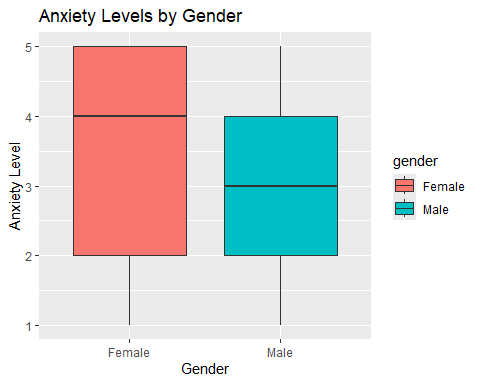
avg\_sleep <- Student\_data %>%  
 group\_by(gender) %>%  
 summarize(mean\_sleep = mean(average\_sleep\_numeric, na.rm = TRUE))  
avg\_sleep

## # A tibble: 2 × 2  
## gender mean\_sleep  
## <chr> <dbl>  
## 1 Female 5.71  
## 2 Male 6.20

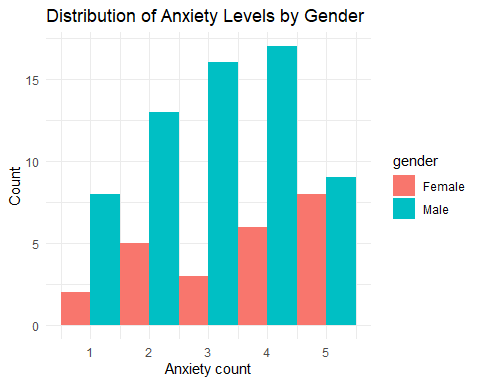
# 3 Find the maximum anxiety level by gender  
# What is the maximum anxiety level reported by male and female students?  
# Convert anxiety to numeric and filter out NA values  
Student\_data <- Student\_data %>%  
 mutate(anxiety = as.numeric(as.character(anxiety))) %>%  
 filter(!is.na(anxiety))  
# Calculate the maximum anxiety level by gender  
max\_anxiety <- Student\_data %>%  
 group\_by(gender) %>%  
 summarize(max\_anxiety = max(anxiety, na.rm = TRUE))  
max\_anxiety

## # A tibble: 2 × 2  
## gender max\_anxiety  
## <chr> <dbl>  
## 1 Female 5  
## 2 Male 5

# 4--Create a boxplot of Anxiety Levels by Gende  
ggplot(Student\_data, aes(x = gender, y = anxiety, fill = gender)) +  
 geom\_boxplot() +  
 labs(  
 title = "Anxiety Levels by Gender",  
 x = "Gender",  
 y = "Anxiety Level"  
 ) +  
 theme\_gray()



# 5--What is the distribution of anxiety levels among male and female students?  
# Histogram of anxiety levels by gender  
ggplot(Student\_data, aes(x = anxiety, fill = gender)) +  
 geom\_histogram(binwidth = 1, position = "dodge") +  
 labs(title = "Distribution of Anxiety Levels by Gender",   
 x = "Anxiety count",   
 y = "Count") +  
 theme\_minimal()



Conclusion   
   
   
 In this analysis, we found that sleep duration has a significant effect on anxiety levels in male and female students. Our results showed that students who sleep longer tend to report lower anxiety levels. The scatter plot confirmed this relationship, indicating that getting enough sleep can help reduce anxiety. We also noticed differences between genders, suggesting that boys and girls may experience anxiety and sleep differently . Like we can see ,male have more anxiety than female while male students males have longer sleep durations than female .  
   
 Future Directions

These findings are important because they suggest that schools should encourage good sleep habits to help students manage their anxiety. also, if male students sleep more but still have higher anxiety, it might be due to poor sleep quality , high stress levels,biological diferrences For future research, it would be useful to look at how the quality of sleep affects anxiety and to conduct studies over a longer period to see if changes in sleep patterns have lasting effects on anxiety levels. By exploring these areas, we can find better ways to support student mental health.

Reference: <https://www.kaggle.com/datasets/abdullahashfaqvirk/student-mental-health-survey>

I also use google and <https://stackoverflow.com/> which help me to Calculate the average sleep duration by gender (~)