HW 1

Question 1:

1.a: The 5 W's

Who, What, Where, Why, When

This data is interpreting sleep schedules from college students to determine whether or not poor sleep schedules leads to negative internalized and externalized behaviors such as aggression and depression -respectfully. This data can tell us about the longevity of sleep habits in college students and how preventative measures can be taken in order to prevent long-lasting sleep disorders. Within this data men seem to be more at risk of sleep-related disorders, however the significance in relation to woman is not that noticeable.

1.b: Variable Types

```
#installing package
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```r
library(readxl)

df <- read.csv("sleep.data.cleaned.csv")

#creating categories
categories <- c(
  "id", "sex", "age", "race", "ethnicity",
  "totalepworth", "anxiety_clinical"
)
# creating table with categories
table(categories)
```

```
categories
             age anxiety_clinical        ethnicity               id
               1                1                1                1
            race              sex     totalepworth
               1                1                1
```

```r
# data frame with the variable info
variable_info <- data.frame(
  variable = c("id", "sex", "age", "race", "ethnicity", "totalepworth", "anxiety_clinical"),
  variable_type = c("Categorical (ID)", "Categorical", "Quantitative (Continuous)",
                    "Categorical", "Categorical", "Quantitative (Discrete)", "Categorical (Bi
  Units = c("None", "None", "Years", "None", "None", "Score (0-24)", "None (e.g., Yes/No)")
)

# Print the table
print(variable_info)
```

```
          variable             variable_type               Units
1               id          Categorical (ID)                None
2              sex               Categorical                None
3              age Quantitative (Continuous)               Years
4             race               Categorical                None
5        ethnicity               Categorical                None
6     totalepworth   Quantitative (Discrete)        Score (0-24)
7 anxiety_clinical      Categorical (Binary) None (e.g., Yes/No)
```

Question 2

2a: Filtering the Data

```
# Flitering Data
#| echo: true`

library(dplyr)

# filter for females
filtered_only_females <- filter(df, sex == "f")

#filter for males
filtered_only_males <- filter(df, sex == "m")
```

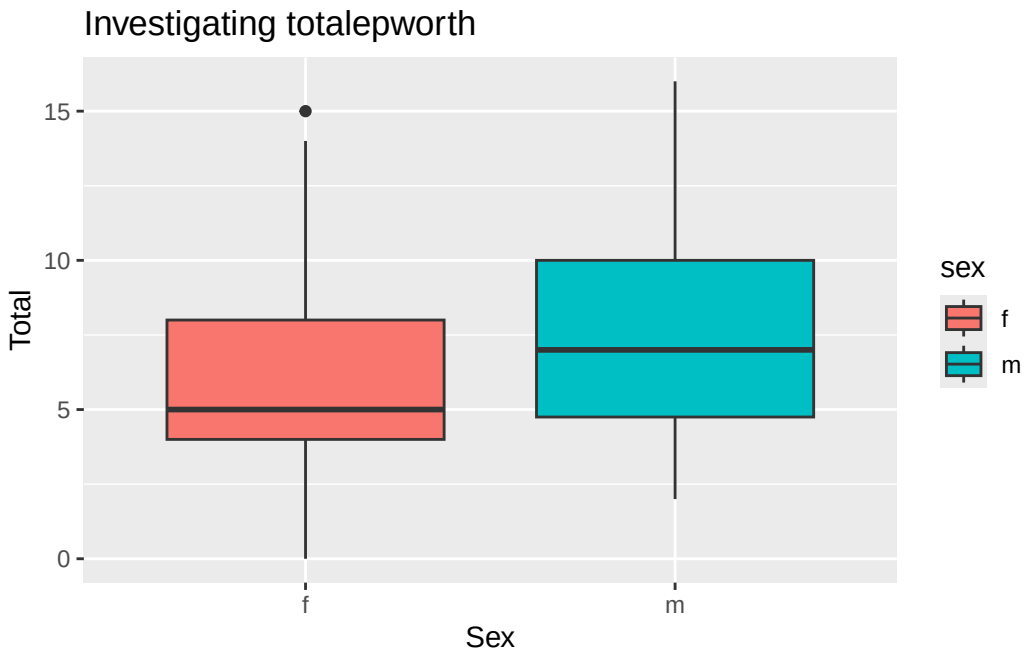
2b: Investigating totalepworth‘

Think:

Because the data in **totalepworth** is numeric I think that the approriate graph for this designation of data would be a box-plot. It shows the appropriate outliers and describes an accurate spread of data in a way that would visually enhance the understanding of the data. Since the participants represent the larger portion of college students at large, having a box plot would visually show where each gender lies in terms of mean, spread, and outliers. Giving an accurate representation of visual data applicable to a larger pool of people. The variable here shows the likeliness of dozing off in daytime situations in comparison to gender. I expect the graph of male EP to be skewed to the right in opposition to the female graph. I think this because the article states that males have a tendency to experience higher sleep problems, and this would be shown on a box plot with a skew to the right in comparision to the females graph.

```
# Investigating Think-Show-Tell
#think

#Show:
library(dplyr)

#as a factor:
df$sex <- as.factor(df$sex)

library(ggplot2)

ggplot(df, aes(x = sex, y = totalepworth, fill = sex)) +
  geom_boxplot() +
  labs(
    title = "Investigating totalepworth",
```

```
    x = "Sex",
    y = "Total" )
```



Tell:

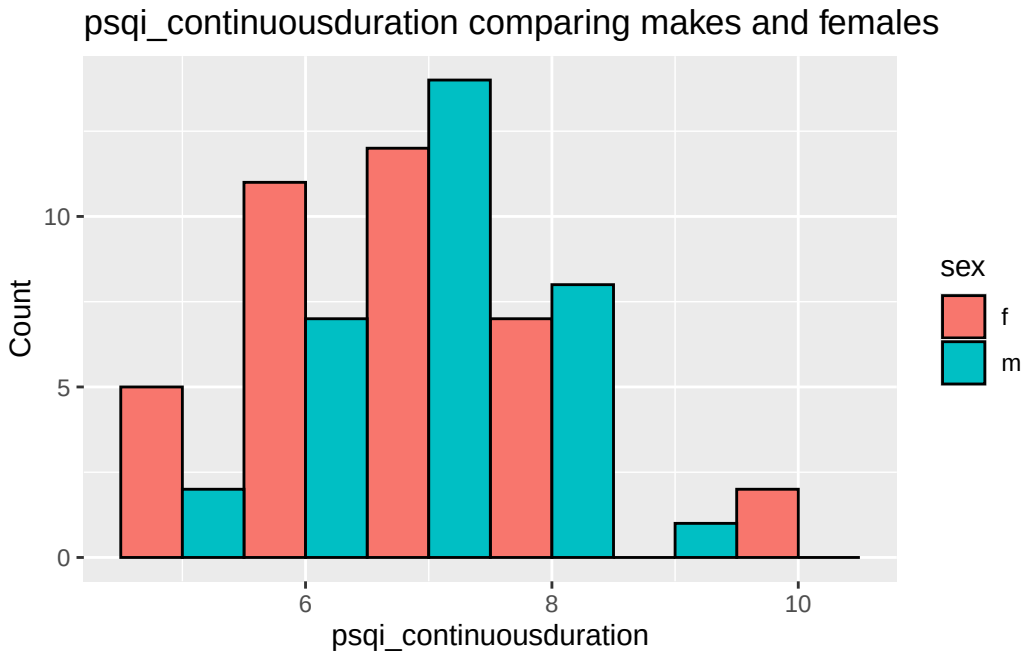
The findings have shown that my initial expectation of the graphs were correct. The male boxplot is more right-leaning compared to the female box-plot because of a higher mean, and overall spread of data.

2.c: Investigating PSQI

Think: I think that a histogram of this data will show us that there is a higher number of females with more sleep per night because the article suggested that feamles on average have a better sleep cycle and adhere to a sleep schedule more often over men. I predict that the male portion of the graph will skew more towards the left (less sleep hours) and the females skew more towards the right (more sleep hours) in correlation with the conclusions of the article that men develop more sleep related issues, indicating that maybe they get on average less sleep than women do.

```
#Show:
library(dplyr)

#as a factor:
df$sex <- as.factor(df$sex)
```

```r
library(ggplot2)

ggplot(df, aes(x = psqi_continuousduration, fill = sex)) +
  geom_histogram(binwidth = 1, position = "dodge", color = "black") +
  labs(
    title = "psqi_continuousduration comparing makes and females",
    x = "psqi_continuousduration",
    y = "Count"
  )
```



Tell: My initial interpretation was actually incorrect. Females sleep on average an hour less per night than males do, however there is a slight outlier at 10 hours with more women sleeping in the 10 hour range than men. But if you look at the middle, 7 hours, more men are sleeping an everage of 7 hours than women. However this contradicts the bins all the way to the left at 5 hours, with significantly more women sleeping at an average of 5 hours than men. This graph tells us that on average men sleep more than woman do. And because of the outlier, men also have a more consistent average sleeping time compared to women.

2d: Comparing

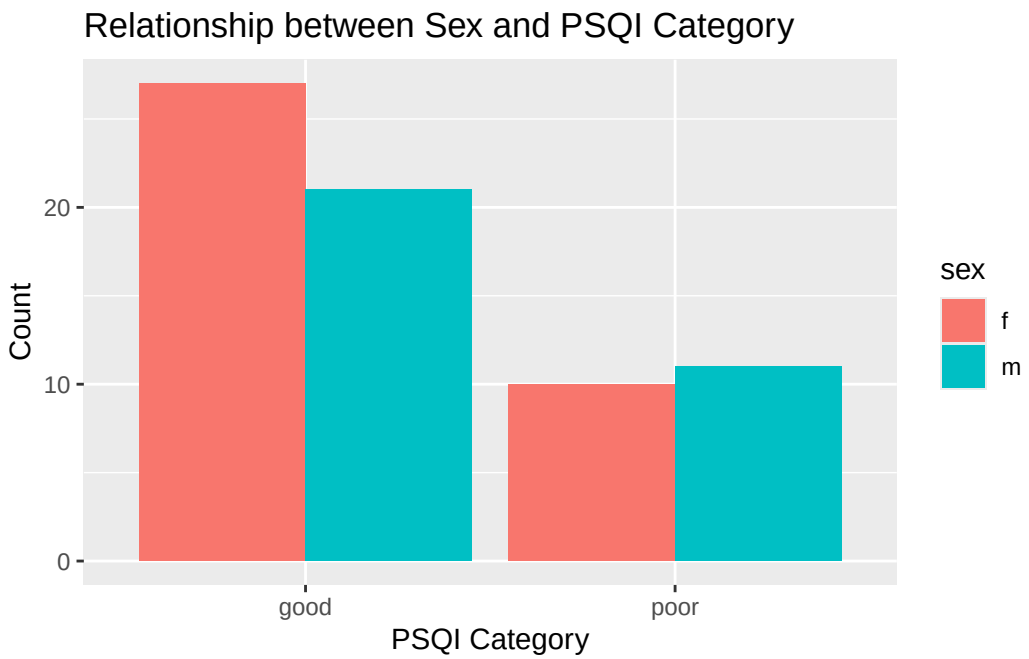
These two graphs suggest that male students are more likely to experience daytime sleepiness, even though they report getting more sleep at night on average. The right-skewed distribution in the Ep graph for males indicates that a portion of them experience significantly higher

sleepiness during the day. This, when viewed alongside the PSQI duration graph, suggests that the sleep males are getting may be of lower quality — potentially due to sleep disruptions, poor sleep habits, or insufficient REM sleep. As college students lifestyle choices like stress, screen time, and irregular schedules could be major factors. This finding highlights that sleep quantity doesn't guarantee feeling well-rested and that improving sleep quality may be beneficial for males to reduce daytime drowsiness and improve their day to day lives.

Question 3

Think: This data is telling us what type of sleep males vs females are getting at night. If the sleep is "good" or "poor". I predict that people will mainly be sleeping well, because participants with sleep-related issues were excluded from the study in the beginning. I also predict that females will report having better sleep because they have previously reported feeling less drowsy than men during the day, suggesting that their overall sleep is good rather than bad and they report having less disruptions in the night. because of this the graph will be skewed to the right with females reporting higher in the "good category" than males do.

```
library(ggplot2)
library(dplyr)
library(tidyr)


library(ggplot2)

ggplot(df, aes(x = psqi_category, fill = sex)) +
  geom_bar(position = "dodge") +
  labs(
    title = "Relationship between Sex and PSQI Category",
    x = "PSQI Category",
    y = "Count"
  )
```



Tell: I was slightly off with my interpretation of the data. I predicted the correct outcome just not in the right area. When I was making my observation I did not take into account the fact that 0-5 means "good" and anything reported above a 5 is considered poor. So with my previous analysis still the same I will tweak the findings to show a graph skewed to the left showing the proper interpretation of data. More females are reporting better sleep and I think this is related to the fact that females are not drowsy during the day, they report having less drwosy episodes and this graph here shows the exact correlation between good sleep and poor day-time activities related to daytime drowsiness.

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

```
[1] 2
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).