Sleep Health and Lifestyle: A Visual Exploration

2023-10-04

Name:Riddhi Mahesh Dange

CWID: 20012299

Link to dataset: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

Explaination of Topic Chosen and Idea behind it:

Sleep Health and Lifestyle: A Visual Exploration is a data visualization project that explores the relationship between sleep health and lifestyle factors. It uses data from various sources to show how different lifestyle choices, such as diet, exercise, and stress, can affect our sleep. Sleep is essential for good health and well-being. It allows our bodies and minds to rest and repair, and it helps us to function at our best. However, many people do not get enough sleep, and this can have negative consequences for our health. Our lifestyle choices can have a significant impact on our sleep health. For example, regular exercise can help to improve sleep quality, but it is important to avoid exercising too close to bedtime. Stress can also make it difficult to fall asleep and stay asleep. There are a number of different stress management techniques that can help to improve sleep quality. This data visualization project tells a story about sleep health and how to improve it. It is important because it helps people to understand the importance of sleep and make better choices about their lifestyle. The project uses visualization to make the data more engaging and easier to understand.

Process and data selection:

This dataset contains information about the sleep health and lifestyle habits of 400 people. It includes variables 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.

Dataset Columns:

Person ID: An identifier for each individual.

Gender: The gender of the person (Male/Female).

Age: The age of the person in years.

Occupation: The occupation or profession of the person.

Sleep Duration (hours): The number of hours the person sleeps per day.

Quality of Sleep (scale: 1-10): A subjective rating of the quality of sleep, ranging from 1 to 10.

Physical Activity Level (minutes/day): The number of minutes the person engages in physical activity daily.

Stress Level (scale: 1-10): A subjective rating of the stress level experienced by the person, ranging from 1 to 10.

BMI Category: The BMI category of the person (e.g., Underweight, Normal, Overweight).

Blood Pressure (systolic/diastolic): The blood pressure measurement of the person, indicated as systolic pressure over diastolic pressure.

Heart Rate (bpm): The resting heart rate of the person in beats per minute.

Daily Steps: The number of steps the person takes per day.

Sleep Disorder: The presence or absence of a sleep disorder in the person (None, Insomnia, Sleep Apnea).

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(reshape2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:lubridate':
##
##     %--%, union
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
data <- read.csv("/Users/riddhidange/Desktop/Academics/fall2023/FA550/Assignment 2/Sleep_health_and_life
#View(data)
#summary(data)
```

1. Histogram of Sleep Duration

This histogram provides a visual representation of the distribution of sleep durations in the dataset. The x-axis represents different ranges of sleep duration (in hours), and the y-axis represents the frequency or count of individuals falling within each range. The graph shows the distribution of sleep duration in a population. Most people sleep between 6 and 8 hours per night, with the most common sleep duration being 7 hours. A significant number of people sleep less than 6 hours per night, which is the recommended amount of sleep for most adults.The graph is important because it can be used to understand the sleep habits of the population and to identify groups of people who may be at risk for sleep problems. It can also be used to track changes in sleep duration over time

```
#1 Histogram
hist(data$`Sleep.Duration`, breaks = 20, main="Histogram of Sleep Duration", xlab="Sleep.Duration (hours
```

Histogram of Sleep Duration

Sleep.Duration (hours)
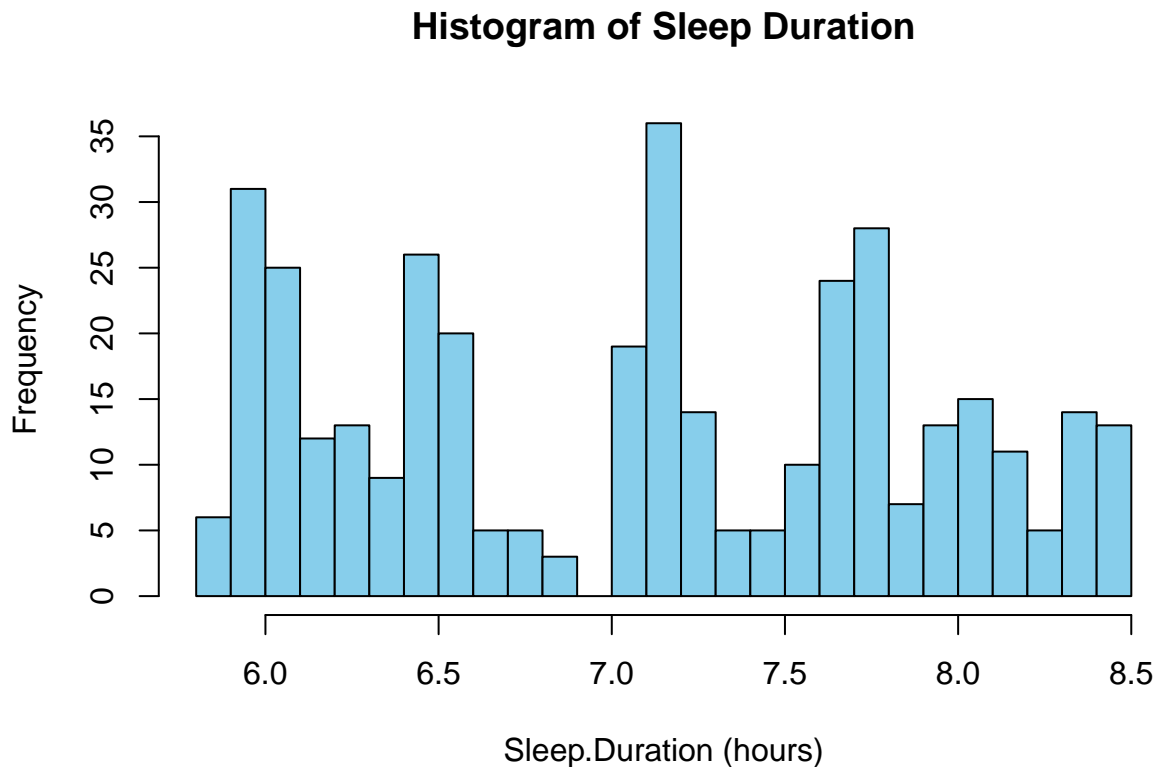
2.Scatterplot of Sleep Duration vs. Quality of sleep

Each point on the scatter plot represents an individual in the dataset. The x-axis shows the sleep duration, and the y-axis shows the quality of sleep andshows the relationship between sleep duration and quality of sleep.The graph shows a positive correlation between sleep duration and quality of sleep, meaning that people who sleep longer tend to have better quality sleep. This is because sleep is essential for our bodies and minds to rest and repair. When we don't get enough sleep, we are more likely to experience fatigue, difficulty concentrating, and irritability.The importance of this graph is that it highlights the importance of getting enough sleep for good health and well-being. Most adults need around 7-8 hours of sleep per night, but some people may need more or less sleep depending on their individual needs.

```
#2 Scatterplot
plot(data$`Sleep.Duration`, data$`Quality.of.Sleep`, main="Scatter Plot of Sleep Duration vs. Quality o
```

**Scatter Plot of Sleep Duration vs. Quality of Sleep**
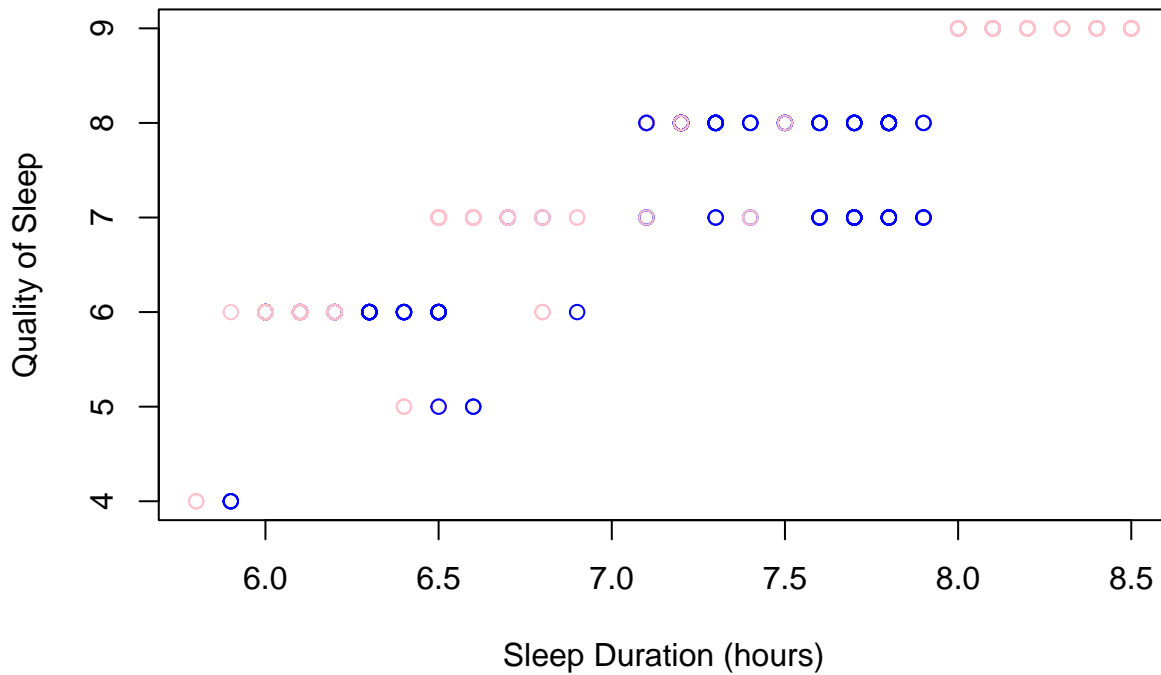
3. Bar graph of Average. Physical Activity Level by BMI Category

The bar chart compares the average physical activity level (in minutes per day) across different BMI categories. Each bar represents a BMI category, and the height of the bar indicates the average physical activity level for individuals in that category. It helps you analyze how physical activity levels vary among different BMI groups. The graph shows that people with a normal BMI tend to have the highest average physical activity level, followed by people who are overweight and people who are underweight. People who are obese have the lowest average physical activity level.This is an important graph because it shows the relationship between physical activity and BMI. BMI is a measure of body fatness, and it is calculated using a person's height and weight. Physical activity is any bodily movement that enhances or maintains physical fitness and overall health and wellness. People with a higher BMI are more likely to be physically inactive. This is because it can be more difficult for people with a higher BMI to engage in physical activity.

```
#3 Bar Graph
ggplot(data, aes(x = `BMI.Category`, y = `Physical.Activity.Level`)) +
  geom_bar(stat = "summary", fun = "mean", fill = "yellow") +
  labs(title = "Average Physical Activity Level by BMI Category",
       x = "BMI Category",
       y = "Average Physical Activity Level (minutes/day)")
```

Average Physical Activity Level by BMI Category
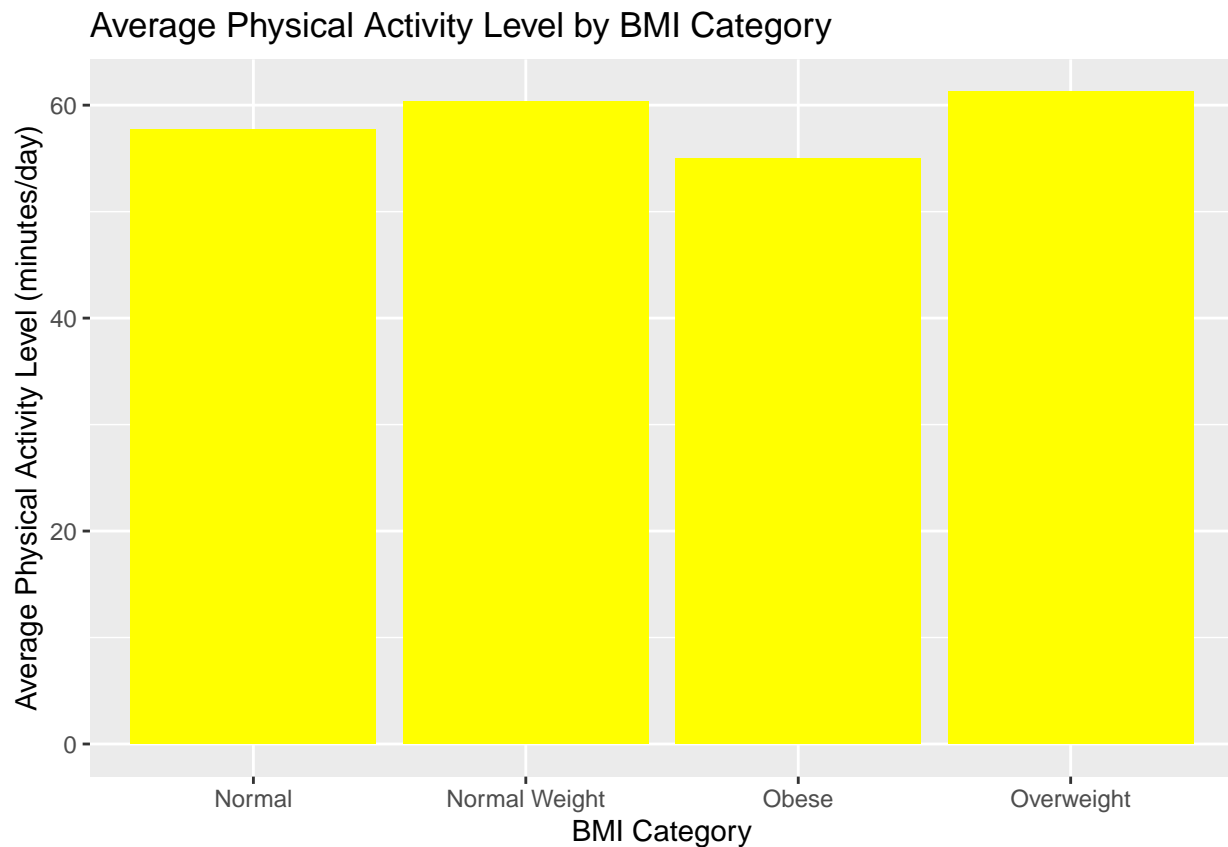
4.

Blood Pressure over Age

This line chart shows how blood pressure (both systolic and diastolic) changes with age. Each point on the line represents the average blood pressure measurements for individuals in specific age groups. It helps identify trends and patterns in blood pressure across different age ranges. The graph shows that blood pressure tends to increase with age.

```
#4 Line Graph
bp_age_df <- data.frame(
  age = data$Age,
  systolic_bp = data$`Blood.Pressure`)

ggplot(bp_age_df, aes(x = age, y = systolic_bp)) +
  geom_line() +
  labs(title = "Blood Pressure Over Age", x = "Age", y = "Systolic Blood Pressure (mmHg)")
```

Blood Pressure Over Age
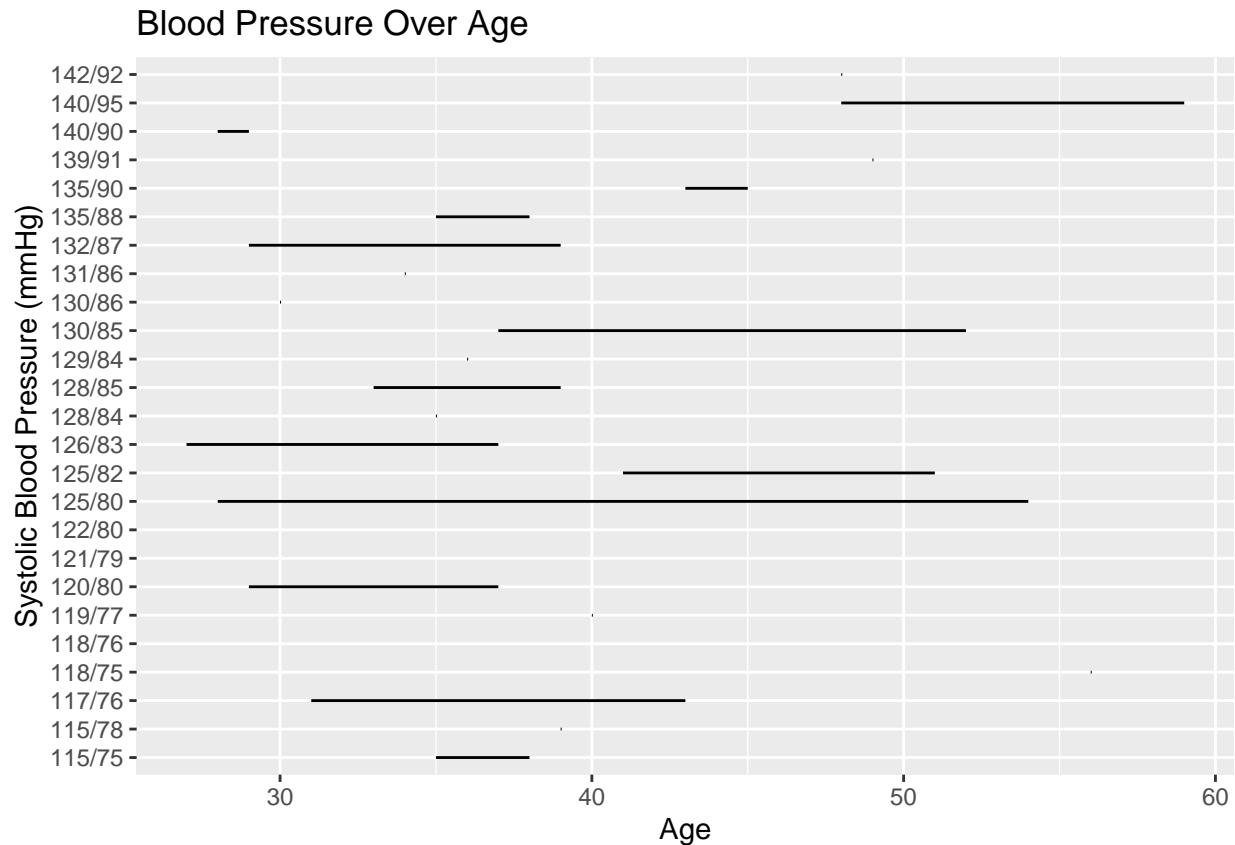
5. Box plot of Stress Levels by Gender

The box plot compares the distribution of stress levels between male and female individuals. It provides a visual summary of the minimum, first quartile, median, third quartile, and maximum stress levels for each gender. Box plots are useful for comparing distributions and identifying potential gender-related differences in stress levels. The box plot shows that females tend to have higher stress levels than males. The median stress level for females is 6, while the median stress level for males is 5. The IQR is also slightly wider for females, suggesting that there is more variation in stress levels among females than among males.

```
#5 Boxplot
boxplot(data$`Stress.Level` ~ data$Gender, col=c("blue", "pink"), xlab="Gender", ylab="Stress Level", ma
```

Box Plot of Stress Levels by Gender
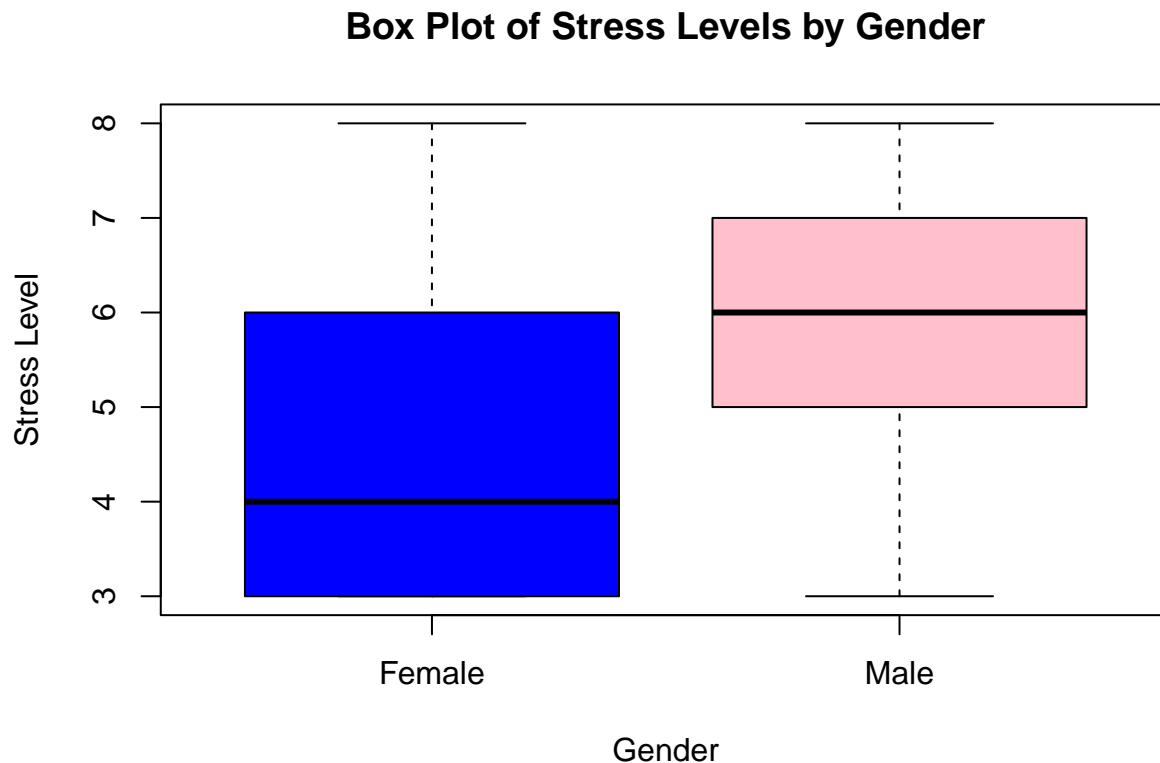
6. Pie chart of Sleep Disorder Distribution

The pie chart illustrates the distribution of sleep disorders in the dataset. Each slice represents a specific sleep disorder category (None, Insomnia, Sleep Apnea), and the size of the slice corresponds to the proportion of individuals with that sleep disorder. It gives a quick overview of the prevalence of different sleep disorders in the dataset and shows the distribution of sleep disorders among the general population. The pie chart shows that insomnia is the most common sleep disorder, affecting 10-15% of adults. Sleep apnea is the second most common sleep disorder, affecting 2-4% of adults.

```
#6 Pie Chart
sleep_disorder_counts <- table(data$`Sleep.Disorder`)
pie(sleep_disorder_counts, labels = names(sleep_disorder_counts), main="Sleep Disorder Distribution", co
```

Sleep Disorder Distribution
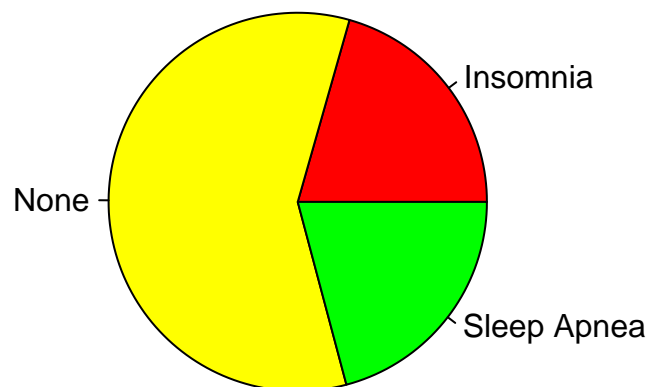
7. Bar graph of Average Sleep duration by Gender

This bar graph depicts the average sleep duration for each gender category. Each bar represents a gender group, and the height of the bar corresponds to the average sleep duration in hours. This visualization allows for a quick comparison of average sleep duration between genders. The graph shows that women tend to sleep slightly longer than men. The average sleep duration for women is 7.0 hours, while the average sleep duration for men is 6.8 hours.

```
#7 Bar Graph
gender_means <- data %>%
  group_by(Gender) %>%
  summarise(mean_sleep_duration = mean(Sleep.Duration))

ggplot(gender_means, aes(x = Gender, y = mean_sleep_duration, fill = Gender)) +
  geom_col() +
  labs(title = "Average Sleep Duration by Gender", x = "Gender", y = "Sleep Duration (hours)") +
  scale_fill_manual(values = c("pink", "blue"))
```
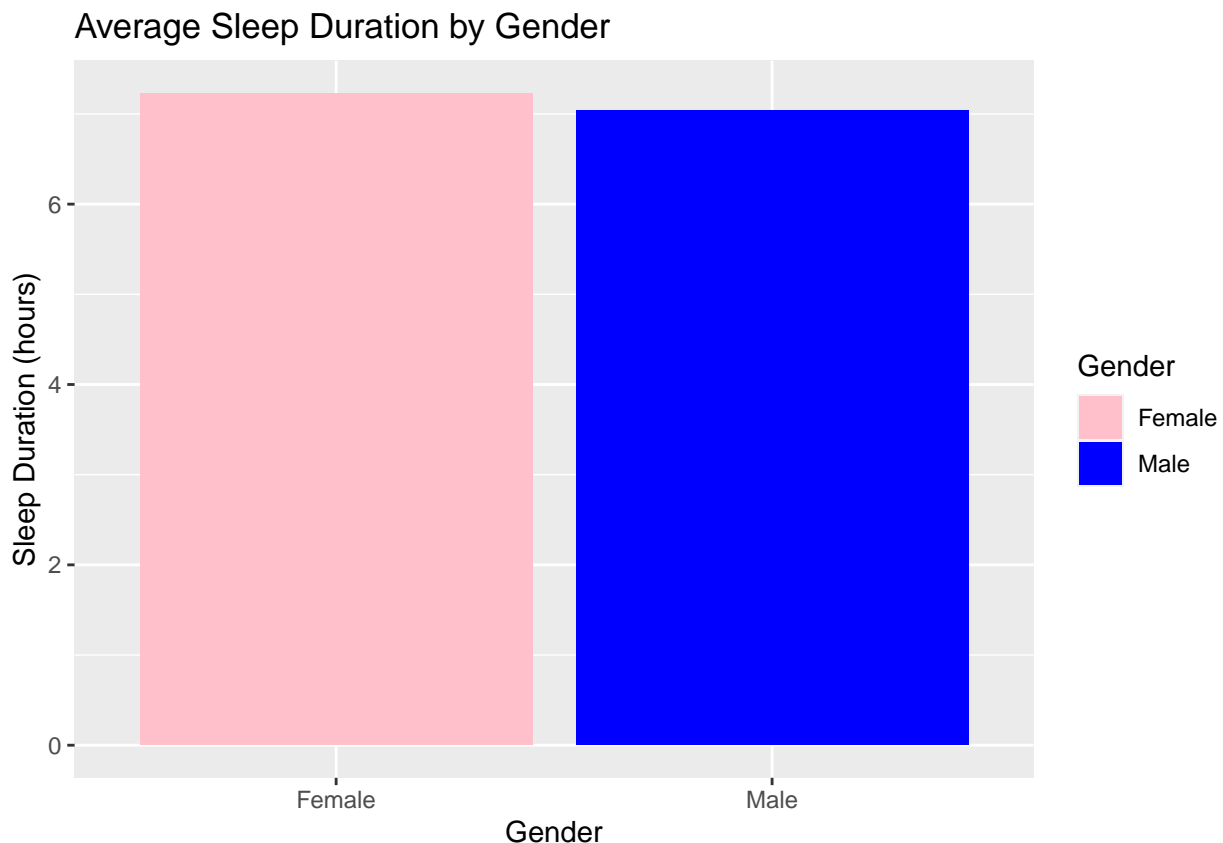

8. Sleep Duration by Occupation

The violin chart provides insights into the distribution of sleep durations among different occupations. Each violin plot represents a different occupation, and the width of the plot indicates the density of sleep durations. Wider sections represent more common durations, while narrower sections represent less common ones. This allows for a visual understanding of the variation in sleep duration across different occupations. The graph shows that workers in some occupations tend to get less sleep than others. For example, workers in the production and healthcare support occupations get an average of 6.7 hours of sleep per night, while workers in the education and computer and mathematical occupations get an average of 7.1 hours of sleep per night.

```
#8 Violin Chart
ggplot(data, aes(x = `Occupation`, y = `Sleep.Duration`)) +
  geom_violin() +
  labs(title = "Sleep Duration by Occupation", x = "Occupation", y = "Sleep Duration (hours)")
```

Warning: Groups with fewer than two data points have been dropped.

Sleep Duration by Occupation

Conclusion:

In this study, the Sleep Health and Lifestyle: A Visual Exploration revealed insights into complex connections among sleep patterns, lifestyle choices, and health for 400 individuals across 13 variables. The sleep duration histogram reflected diverse habits, and the scatter plot hinted at intricate dynamics in sleep duration versus quality correlation. Examining lifestyle factors, the bar chart comparing physical activity across BMI categories indicated variations for targeted interventions. The box plot on stress levels by gender emphasized potential gender-related differences, guiding tailored stress management. The cardiovascular health analysis stressed age-specific trends through a line chart on blood pressure changes over age. Additional insights from gender-specific and occupational analyses highlighted potential sleep pattern differences. The study acknowledges limitations; the dataset is a snapshot, and observed correlations don't imply causation. Future research with longitudinal data and more variables could enhance understanding. In conclusion, this exploration deepens our understanding of the interplay between sleep, lifestyle, and health.