Predictive Data Analysis of Sleep Health and Lifestyle

IS 2105 – Business Statistics

Group 11

Ariyarathne U.G.S.T - 22020071

Kasunthara M.A.N - 22020454

Kularatne W.H.R - 22020472

Safran M.Z.M - 22020845

Savindu A.U - 22020901

Table of Content

Introduction

Sleep is a vital component of overall health and well-being, significantly affecting both physical and mental performance. Various factors such as lifestyle habits, stress levels, and cardiovascular health can influence sleep patterns and quality. Understanding these dynamics is essential for promoting better sleep hygiene and preventing sleep-related disorders, which are becoming increasingly prevalent due to modern lifestyle challenges.

This study focuses on analyzing sleep health and its relationship with lifestyle factors using a dataset that includes comprehensive information on sleep duration, quality, physical activity, stress levels, and cardiovascular metrics like blood pressure and heart rate. By examining these variables, the study aims to identify patterns and trends that influence sleep quality and the occurrence of sleep disorders such as insomnia and sleep apnea. This analysis will provide valuable insights into the factors affecting sleep health and help in developing strategies to improve sleep quality and overall well-being.

The objective of this project is to develop a model that can assess the likelihood of sleep disorders based on lifestyle and cardiovascular data. Such a model could be beneficial for individuals and healthcare professionals to identify potential sleep issues early and implement preventive measures.

Data

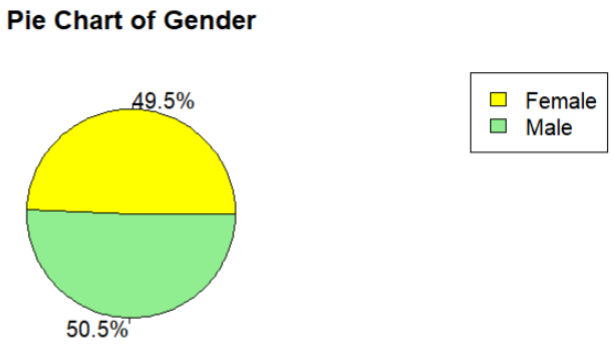
- **Person ID**: A unique identifier assigned to each individual for tracking purposes. (*Qualitative - Nominal*)
- **Gender**: The biological sex of the individual, categorized as Male or Female. (*Qualitative - Nominal*)
- **Age**: The person's age in years, representing their life duration. (*Quantitative - Ratio*)
- **Occupation**: The profession or job the individual is engaged in. (*Qualitative - Nominal*)
- **Sleep Duration (hours)**: The number of hours the individual sleeps per day. (*Quantitative - Ratio*)
- **Quality of Sleep (scale: 1-10)**: A subjective rating where individuals rate their sleep quality from 1 (poor) to 10 (excellent). (*Qualitative - Ordinal*)
- **Physical Activity Level (minutes/day)**: The number of minutes the person spends on physical activities daily. (*Quantitative - Ratio*)
- **Stress Level (scale: 1-10)**: A subjective rating where the person rates their stress level from 1 (low) to 10 (high). (*Qualitative - Ordinal*)
- **BMI Category**: The individual's Body Mass Index classification (e.g., Underweight, Normal, Overweight). (*Qualitative - Ordinal*)
- **Blood Pressure (systolic/diastolic)**: The person's blood pressure measured as systolic over diastolic pressure. (*Quantitative - Ratio*)
- **Heart Rate (bpm)**: The person's resting heart rate in beats per minute. (*Quantitative - Ratio*)
- **Daily Steps**: The number of steps the individual takes each day. (*Quantitative - Ratio*)
- **Sleep Disorder**: The presence or absence of a diagnosed sleep disorder (None, Insomnia, Sleep Apnea). (*Qualitative - Nominal)*

Analysis

Univariate Analysis

- Univariate analysis focuses on a single variable.
- It involves the basic and straightforward examination of individual variables.
- The primary goal is to describe the characteristics of a single variable rather than exploring relationships.
- This analytical approach seeks to identify patterns and trends within individual data points.

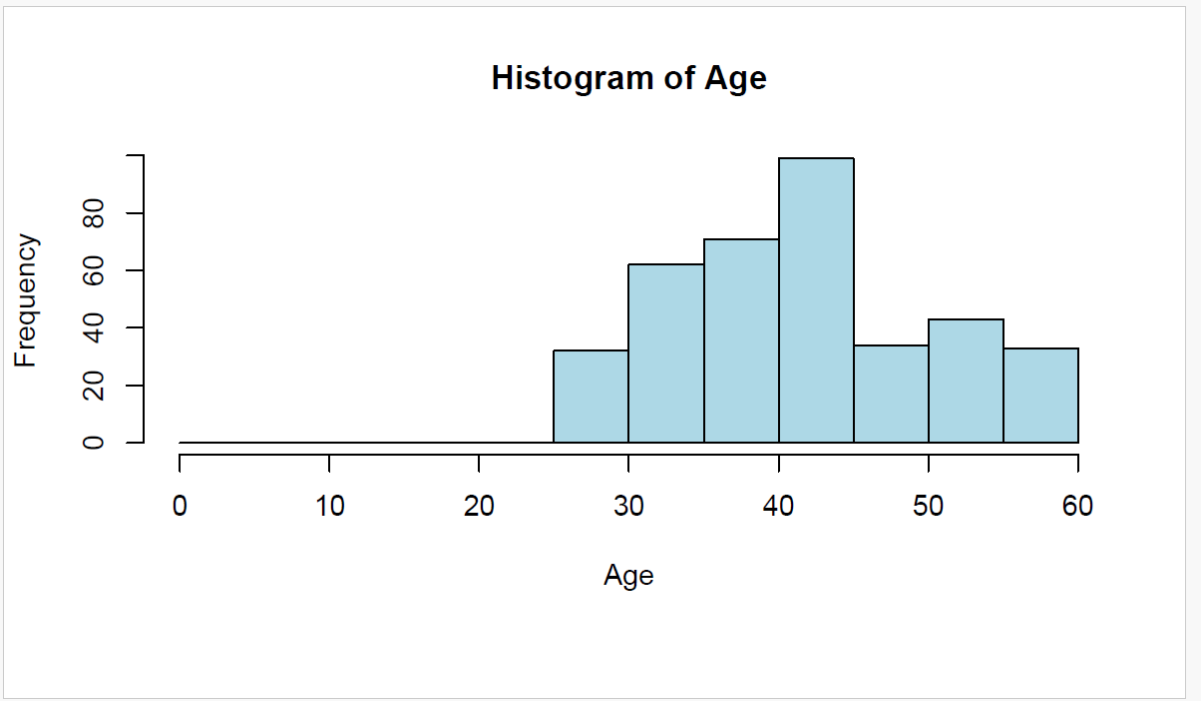
1. Gender




```
> dataset = Sleep_health_and_lifestyle_dataset
> Gender_Freq = table(dataset$Gender)
> Gender_Precentages = prop.table(Gender_Freq)*100
> Colours = c("yellow", "lightgreen")
> pie(Gender_Freq, labels = sprintf("%.1f%%", Gender_Precentages), main = "Gender",
col = Colours )
> legend ("topright", legend =c("Female" , "Male"), fill = Colours)
>
```

The pie chart illustrates the gender distribution in the dataset. It shows that approximately 49.5% of the individuals in the dataset are female, while 50.5% are male. This indicates a relatively balanced representation of both genders in the sample.

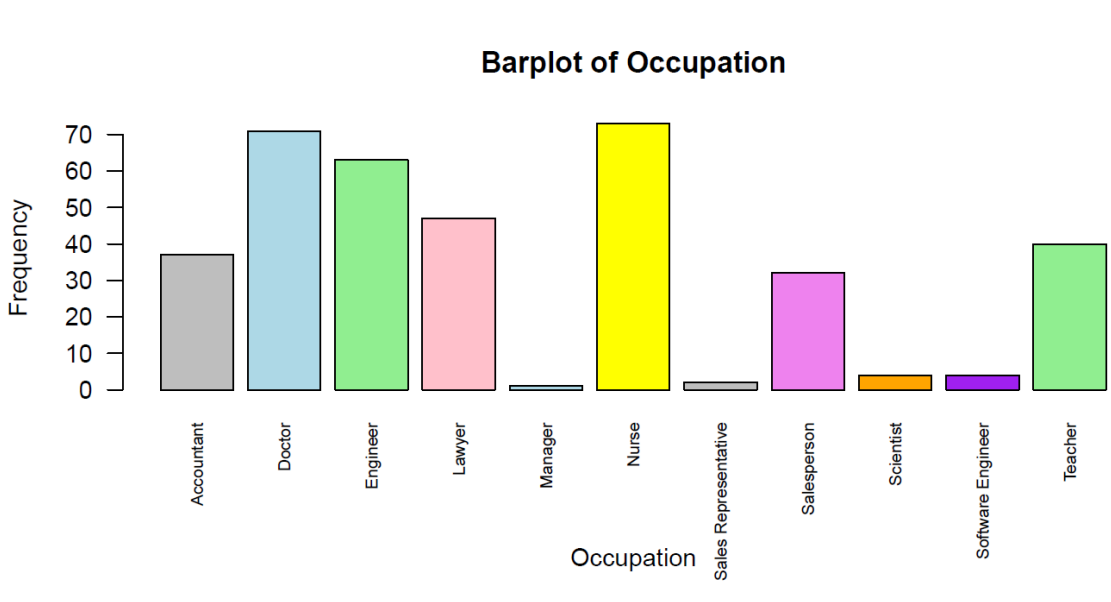
2. Age




```
> hist(dataset$Age,
+      breaks = seq(0, 60, by = 5),
+      col = "lightblue",
+      xlab = "Age",
+      ylab = "Frequency",
+      main = "Histogram of Age")
>
```

The histogram shows the distribution of ages in the dataset. It appears that the majority of individuals in the dataset are between 35 and 45 years old, with a smaller number of individuals in the younger and older age groups. The distribution is generally skewed to the right, indicating that there are a few older individuals in the dataset.

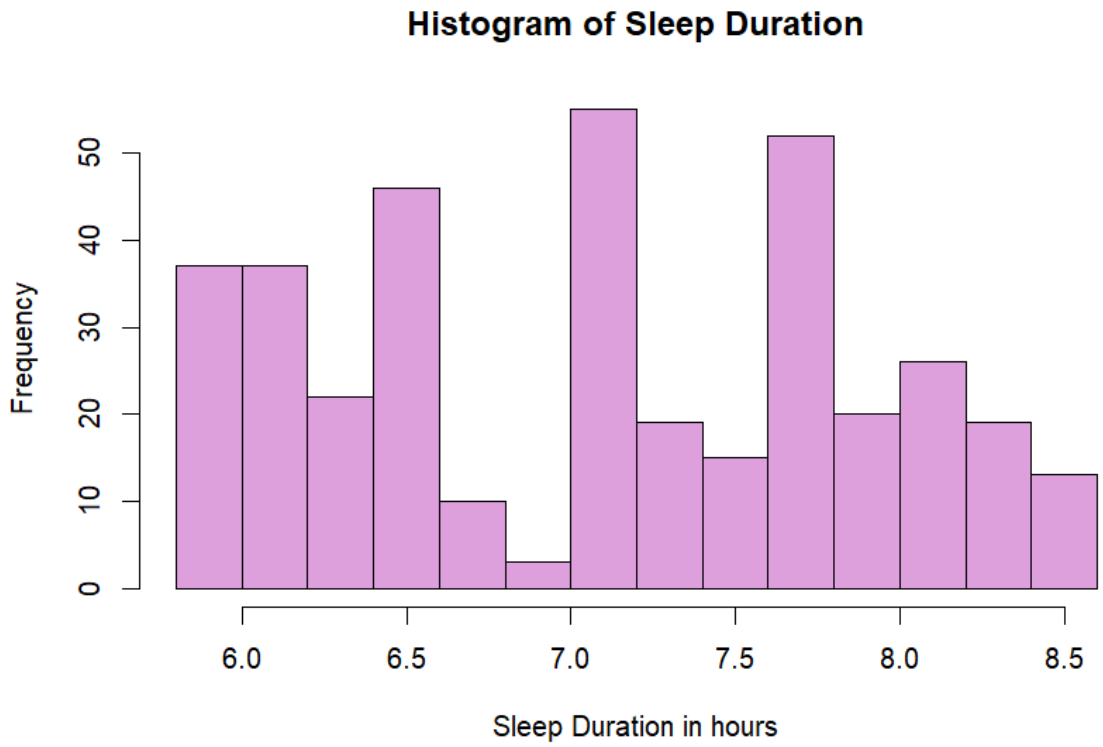
3. Occupation




```
> Occu_freq = table(dataset$Occupation)
> Colour = c("gray", "lightblue", "lightgreen", "pink", "lightblue","yellow","gray","violet","orange",
"purple", "lightgreen")
> par(mar = c(7, 4, 4, 2) + 0.1)
>
> barplot(Occu_freq,
+         names.arg = c("Accountant", "Doctor", "Engineer", "Lawyer", "Manager",
+                       "Nurse", "Sales Representative", "Salesperson",
+                       "Scientist", "Software Engineer", "Teacher"),
+         main = "Barplot of Occupation",
+         ylab = "Frequency",
+         col = Colour,
+         las = 2,            # Vertical label orientation
+         cex.names = 0.7)    # Decrease the font size of the labels
>
> # Adjust the position of the x-axis label (move it lower)
> title(xlab = "Occupation", line = 5)
>
```

The bar plot shows the frequency of different occupations in the dataset. The y-axis represents the frequency or count of individuals in each occupation, while the x-axis lists the different occupations. The plot reveals that the most common occupation in the dataset is Nurse, followed by Doctor and Engineer. The least common occupations are Scientist, Manager, Sales Representative and Software Engineer. The remaining occupations have relatively similar frequencies.

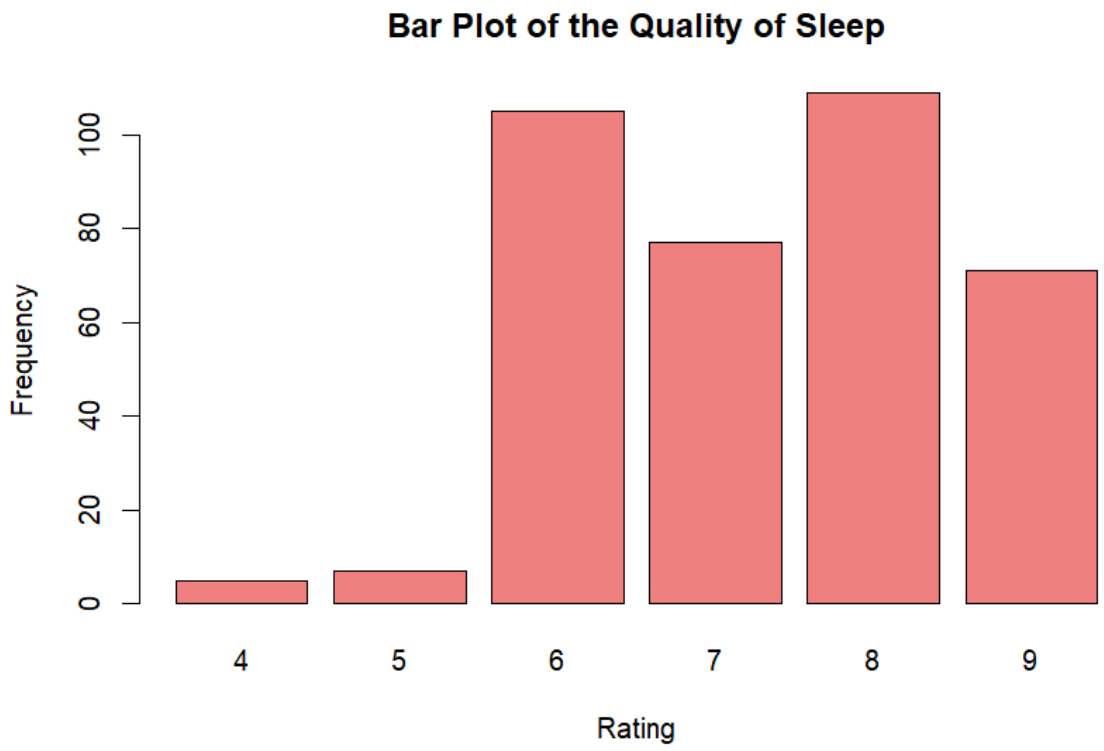
4. Sleep Duration




```
> hist(`Sleep Duration`,
+      breaks = 10,
+      main = "Histogram of Sleep Duration",
+      xlab = "Sleep Duration in hours",
+      ylab = "Frequency",
+      col = "plum",
+      border = "black")
```

The histogram shows sleep durations ranging from about 6 to 8.5 hours, with most individuals sleeping around 7 to 7.5 hours. These durations are the most common, as seen by the tallest bars. While the data is spread out, there are clear peaks at 6, 7, and 7.5 hours, indicating some variation in sleep habits. Fewer people reported sleeping around 6.5 or 8.5 hours. Overall, the distribution suggests that although sleep durations vary, most people in the group tend to sleep between 7 and 7.5 hours.

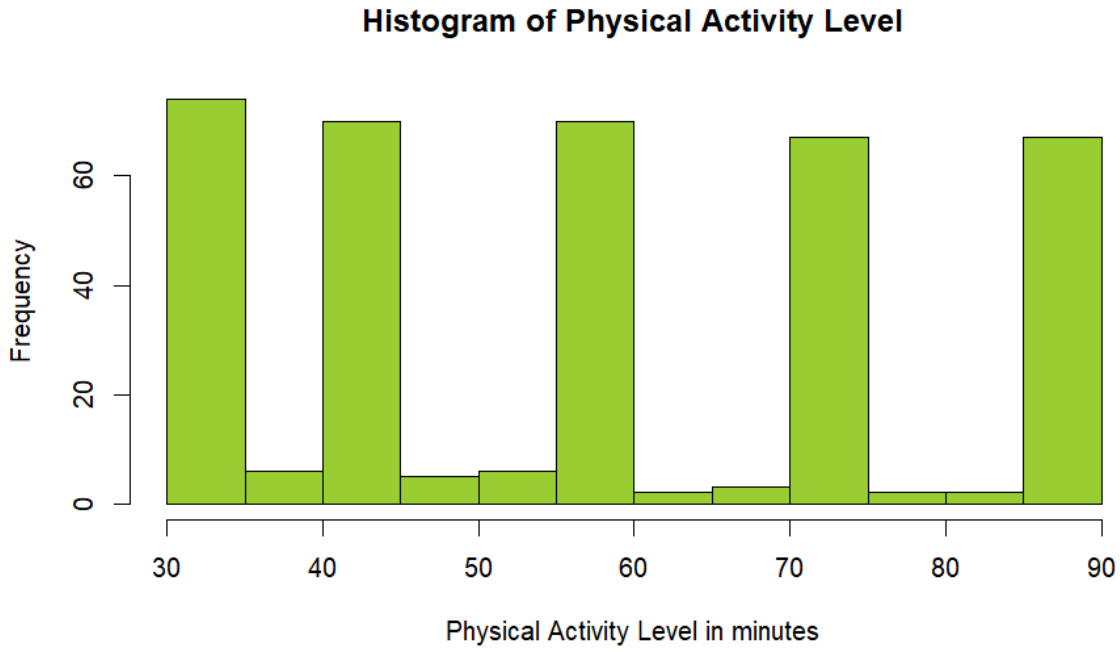
5. Quality of Sleep




```
> categories <- c("4","5","6","7","8","9")
> barplot(table(`Quality of Sleep`),
+         names.arg = categories,
+         col = "lightcoral",
+         main = "Bar Plot of the Quality of Sleep",
+         xlab = "Rating",
+         ylab = "Frequency")
```

Most people rated their sleep quality as either a 6 or 8, suggesting that overall, sleep was generally moderate to good among the group. Fewer individuals gave ratings of 4 or 5, indicating that very poor sleep was less common. While ratings of 7 and 9 were less frequent than 6 and 8, they still accounted for a significant portion of responses. Overall, the distribution implies that most individuals experienced moderate to good sleep quality, with very few reporting very poor sleep, highlighting a generally positive sleep experience within the group surveyed.

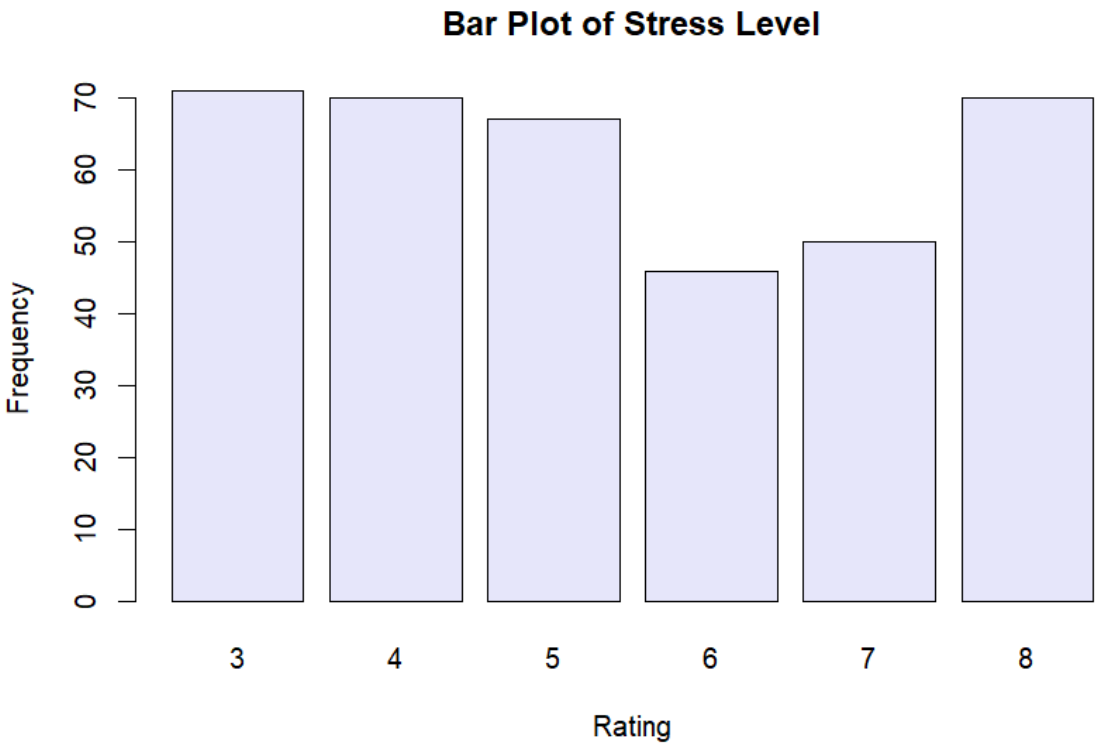
6. Physical Activity level




```
> hist(`Physical Activity Level`,
+      breaks = 10,
+      main = "Histogram of Physical Activity Level",
+      xlab = "Physical Activity Level in minutes",
+      ylab = "Frequency",
+      col = "yellowgreen",
+      border = "black")
```

The histogram shows the distribution of physical activity levels among participants. The x-axis represents the physical activity level in minutes, while the y-axis represents the frequency of participants with that level. The results indicate that most participants are moderately active, with a peak frequency occurring between 30 and 40 minutes of daily physical activity. There is a significant drop in frequency between 40 and 50 minutes, followed by a slight increase between 50 and 60 minutes. The frequency then decreases again between 60 and 70 minutes, before reaching another peak between 70 and 80 minutes. Overall, the data suggests that a majority of participants are moderately active, with a smaller proportion being more active or less active.

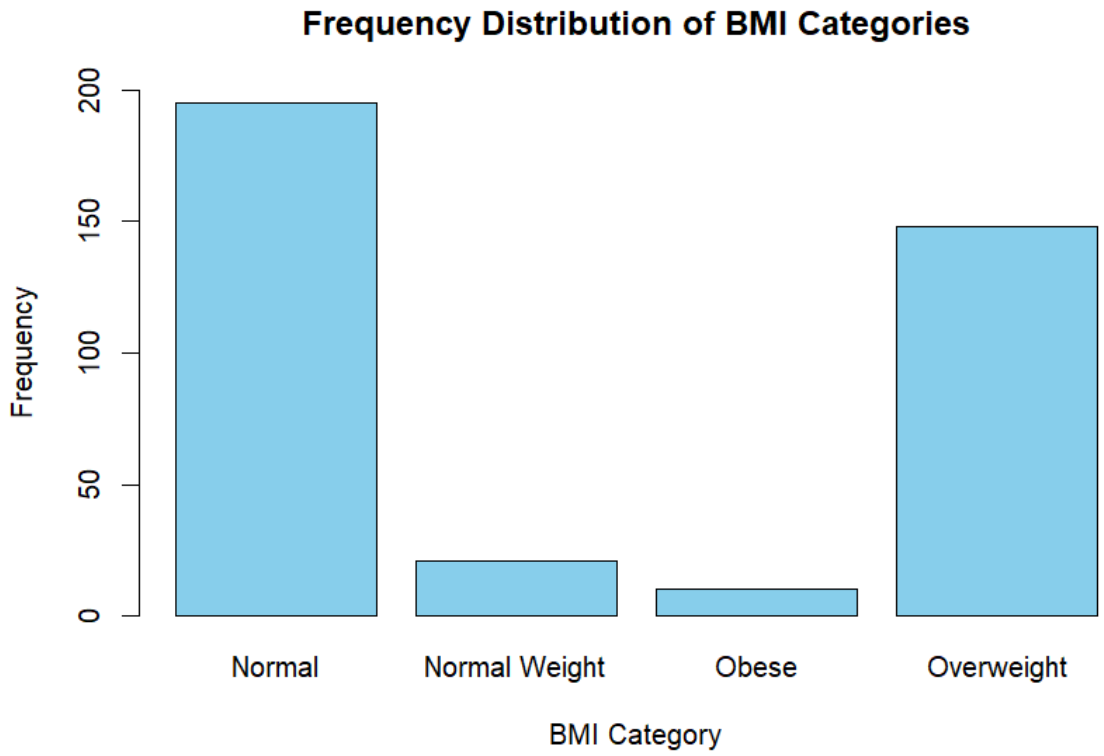
7. Stress level




```
> barplot(table(`Stress Level`),
+         col = "lavender",
+         main = "Bar Plot of Stress Level",
+         xlab = "Rating",
+         ylab = "Frequency")
```

This bar plot illustrates the distribution of stress level ratings, ranging from 3 to 8, based on the frequency of individuals reporting each level of stress. Ratings 3, 4, and 8 are the most common, each with approximately 70 individuals, indicating a bimodal distribution where both low (3, 4) and high (8) stress levels are frequently reported. In contrast, fewer individuals report moderate stress levels, with a significant drop at rating 6, which has around 40 individuals. Ratings 5 and 7 show slightly higher frequencies but still fall below the peaks seen at the extremes. This suggests that while many individuals experience either low or high stress, moderate stress levels are less commonly reported in this dataset.

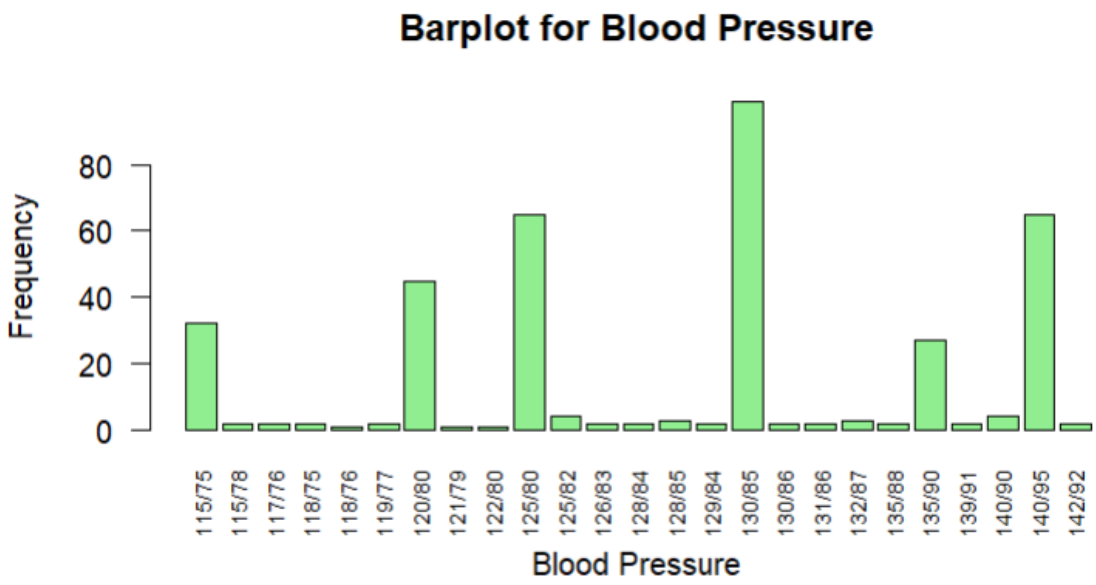
8. BMI Category




```
> bmi_frequency <- table(sleep_data$`BMI Category`)
>
> barplot(bmi_frequency,
+         main = "Frequency Distribution of BMI Categories",
+         xlab = "BMI Category",
+         ylab = "Frequency",
+         col = "skyblue",
+         ylim = c(0, max(bmi_frequency) + 5))
```

The bar graph reveals that the majority of individuals (195) fall into the "Normal" BMI category, followed by a significant number (148) classified as "Overweight." In contrast, the "Normal Weight" and "Obese" groups are much smaller, with 21 and 10 individuals, respectively. This suggests that while a large portion of the population maintains a healthy BMI, there is a notable prevalence of overweight individuals, which may indicate a need for better weight management. The relatively low counts in the "Normal Weight" and "Obese" categories suggest that extremes of BMI are less common in this dataset. Overall, the data highlights the importance of addressing the overweight population to improve health outcomes

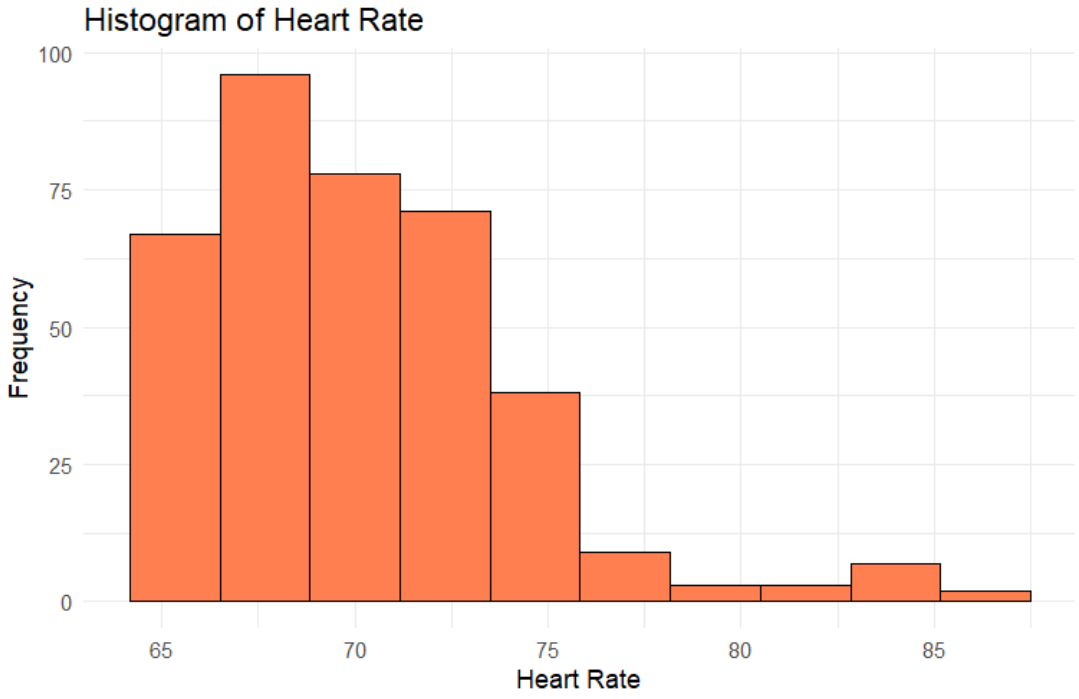
9. Blood Pressure




```
> bp = table(dataset$`Blood Pressure`)
> par(mar = c(7, 4, 4, 2) + 0.1)
> barplot(bp, main = "Barplot for Blood Pressure", xlab = "Blood Pressure", ylab = "Frequency", col = "ligh
tgreen", las = 2, cex.names = 0.7)
>
```

This barplot illustrates the distribution of blood pressure measurements from the dataset. The x-axis shows various systolic/diastolic blood pressure combinations, and the y-axis represents the frequency (number of occurrences) of these measurements in the dataset. The blood pressure readings range from lower values like 115/75 to higher values like 142/92, covering a broad spectrum of systolic and diastolic pressure combinations. Some blood pressure readings occur much more frequently than others. Such as 130/86 has the highest frequency, reaching close to 80 occurrences and 120/80 and 125/80 also show relatively higher frequencies compared to the other values. Many blood pressure values, such as 117/77, 128/85, and 140/95, occur less frequently, showing a nearly flat distribution at the bottom of the y-axis.

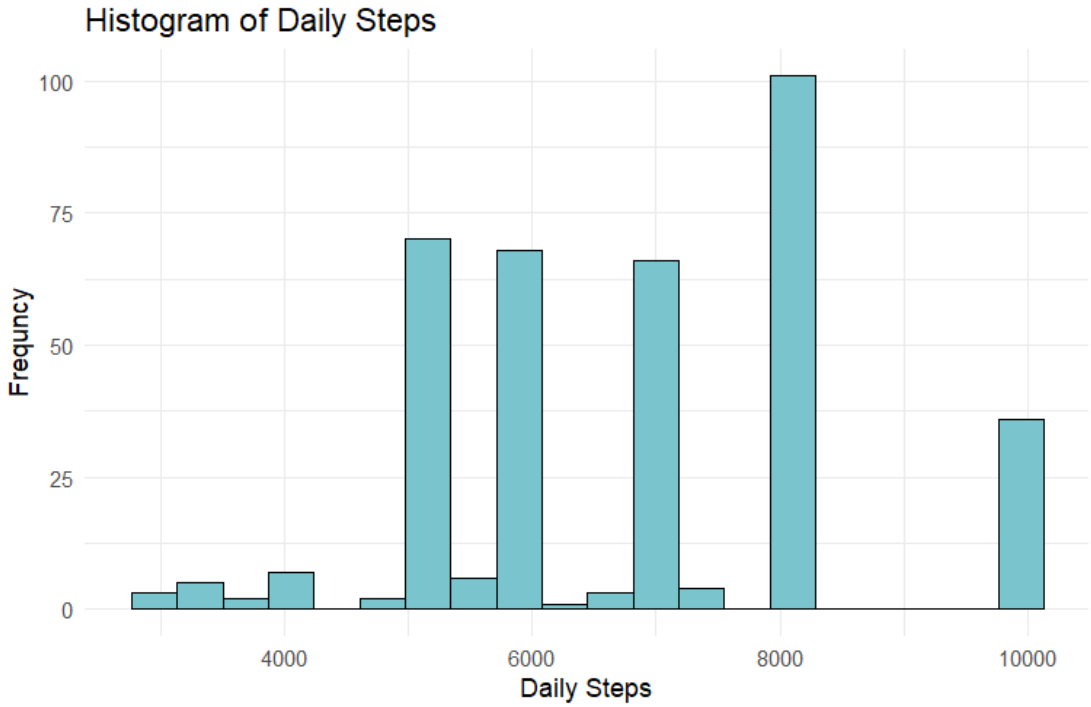
10.Heart Rate




```
ggplot(SleepData, aes(x = `Heart Rate`)) +
    geom_histogram(bins = 10, fill = "coral", color = "black") +
    labs(title = "Histogram of Heart Rate",
         x = "Heart Rate",
         y = "Frequency") +
    theme_minimal()
```

This Histogram illustrates the distribution of heart rates within the sampled population. The x-axis represents heart rate values ranging from 60 to 85, while the y-axis shows the frequency of occurrences, ranging from 0 to 100. The majority of heart rates fall between 65 and 75 beats per minute, with a peak around 68-69. This distribution is unimodal and skewed to the right, indicating that there are fewer individuals with heart rates above 75 beats per minute. This pattern suggests a generally healthy population with a relatively low risk of heart disease.

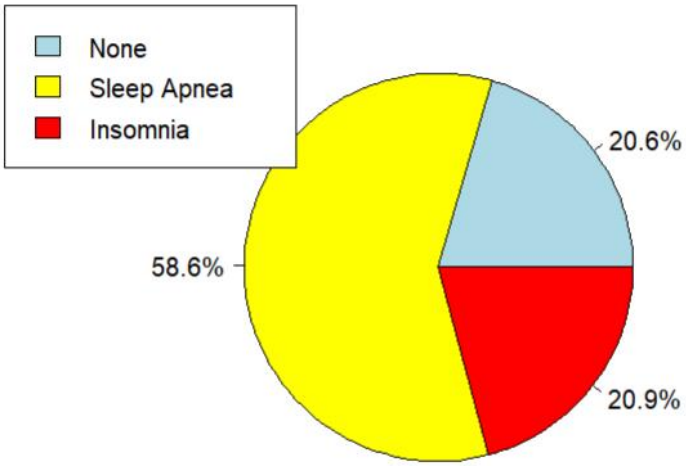
11. Daily Steps




```
ggplot(SleepData, aes(x = `Daily Steps`)) +
    geom_histogram(bins = 10, fill = "cadetblue3",color = "black") +
    labs(title = "Histogram of Daily Steps",
         x = "Daily Steps",
         y = "Frequncy") +
    theme_minimal()
```

This bar graph shows the frequency of daily steps taken in intervals of 2000, ranging from 0 to 10000 steps. The y-axis, labeled “frequency,” ranges from 0 to 100 in increments of 20, while the x-axis, labeled “Daily Steps,” lists the step intervals. Observations from the graph indicate that the 4000-6000 steps range has the highest frequency, suggesting that most individuals in the sample take between 4000 and 6000 steps daily. This implies moderate activity levels, which might be below the recommended daily step count for optimal health. The 0-2000 and 8000-10000 steps ranges have lower frequencies, indicating fewer individuals fall into these categories, suggesting that very low and very high activity levels are less common in the sample. In conclusion, these insights can help public health officials understand physical activity patterns and design interventions to encourage higher daily step counts, promoting better overall health.

12.Sleep Disorder




```
> dataset = Sleep_health_and_lifestyle_dataset
> sleepdisorder = table(dataset$Sleep.Disorder)
> disorder_Presentages = prop.table(sleepdisorder)*100
> Colours = c("lightblue", "yellow", "red")
> pie(sleepdisorder, labels = sprintf("%.1f%%", disorder_Presentages), Main = "Sleep Disorder", col = Colours)
```

```
> legend("topleft", legend = c("None", "Sleep Apnea", "Insomnia"), fill = Colours)
```

The pie chart illustrates the distribution of sleep disorders within a surveyed population. The largest segment, representing 58.6% of the population, indicates that the majority do not suffer from any sleep disorder. However, 20.9% of the population is affected by sleep apnea, and 20.6% suffers from insomnia, highlighting that these two disorders are relatively common. This distribution suggests that while more than half of the population has healthy sleep patterns, a significant portion still faces sleep-related issues. The presence of sleep disorders in over 40% of the population underscores the importance of addressing these conditions through medical intervention, lifestyle changes, and public health initiatives to improve overall health and quality of life.

Bivariate Analysis

✓ Bivariate analysis focuses on the relationship between two variables.

✓ It involves examining how one variable influences or correlates with another.

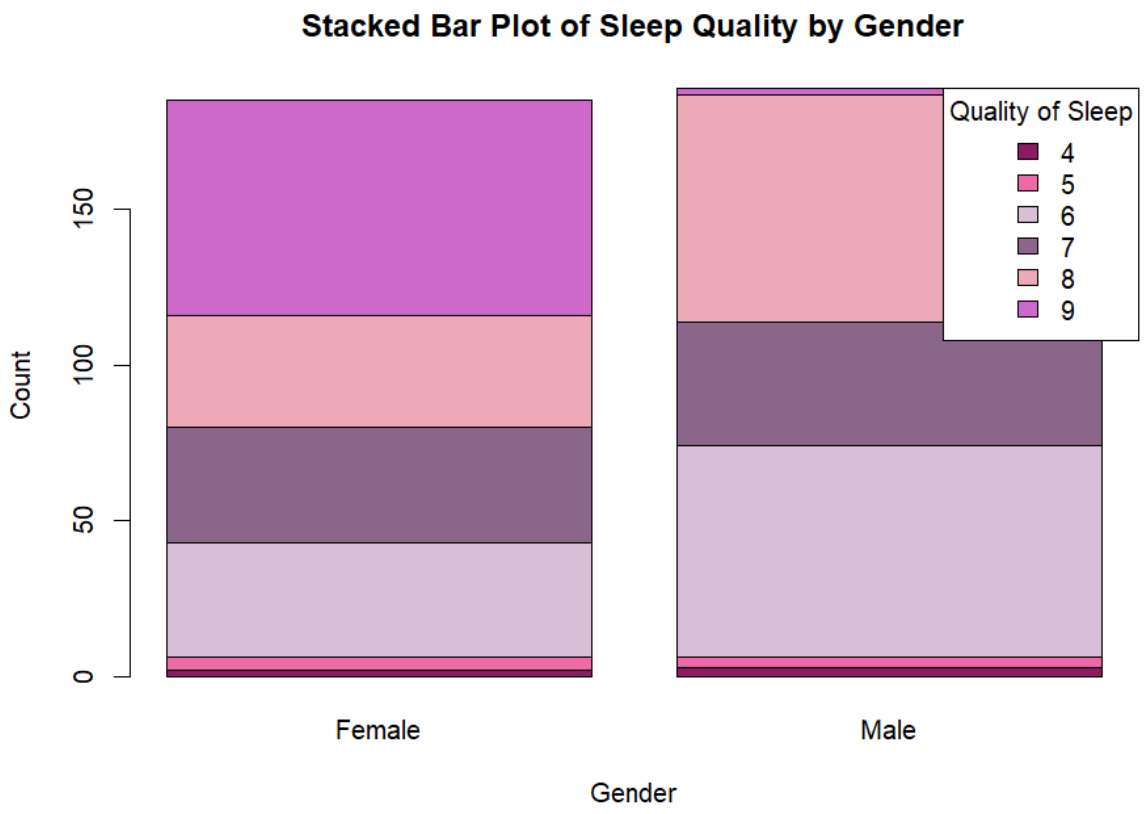
✓ The primary goal is to identify associations, patterns, or relationships between the two variables.

✓ Common techniques include scatter plots, correlation analysis, cross-tabulation, and regression analysis.

✓ This approach helps in understanding the nature (positive, negative, or no relationship) and strength of the interaction between the two variables.

✓ Bivariate analysis is often a step toward deeper multivariate analysis.

- **Gender vs Quality of Sleep**



```
> barplot(table(`Quality of Sleep`,Gender),
+         main = "Stacked Bar Plot of Sleep Quality by Gender",
+         col = c("maroon4", "hotpink2", "thistle","plum4","pink2","orchid3"),
+         xlab = "Gender",
+         ylab = "Count",
+         beside = FALSE)
> legend("topright",
+        title = "Quality of Sleep",
+        legend = sort(unique(`Quality of Sleep`)),
+        fill = c("maroon4", "hotpink2", "thistle", "plum4", "pink2", "orchid3"))
```

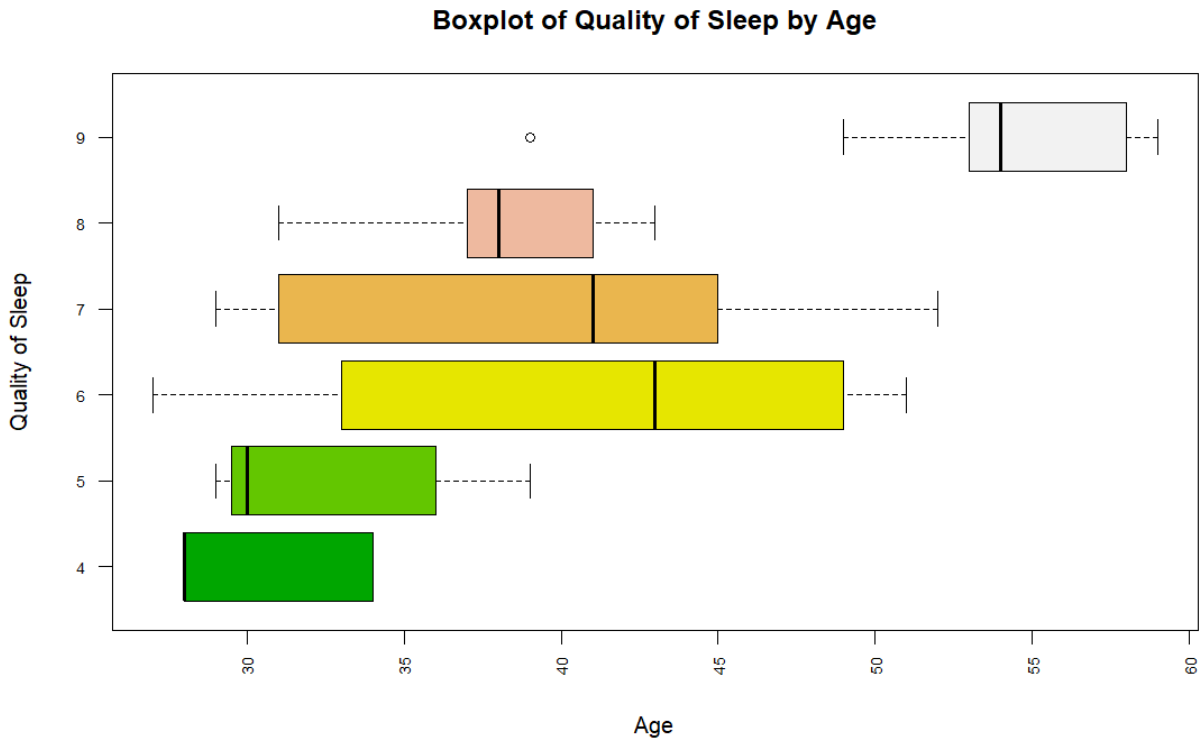
Key findings:

- Both females and males have a similar number of participants with sleep quality scores of 7 and 8. These are the most common scores for both genders.
- Females have a higher proportion of participants with sleep quality scores of 6 and 9. This suggests that females tend to report slightly better sleep quality than males.
- Males have a higher proportion of participants with sleep quality scores of 4 and 5. This suggests that males are more likely to report lower sleep quality than females.

Interpretation:

When it comes to sleep quality, females and males have similar experiences overall. However, females are more likely to report better sleep than males, while males are more likely to report poorer sleep.

- **Age vs Quality of Sleep**




```
> boxplot(Age ~ `Quality of Sleep`,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Age",
+         xlab = "Age",
+         ylab = "Quality of Sleep",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
```

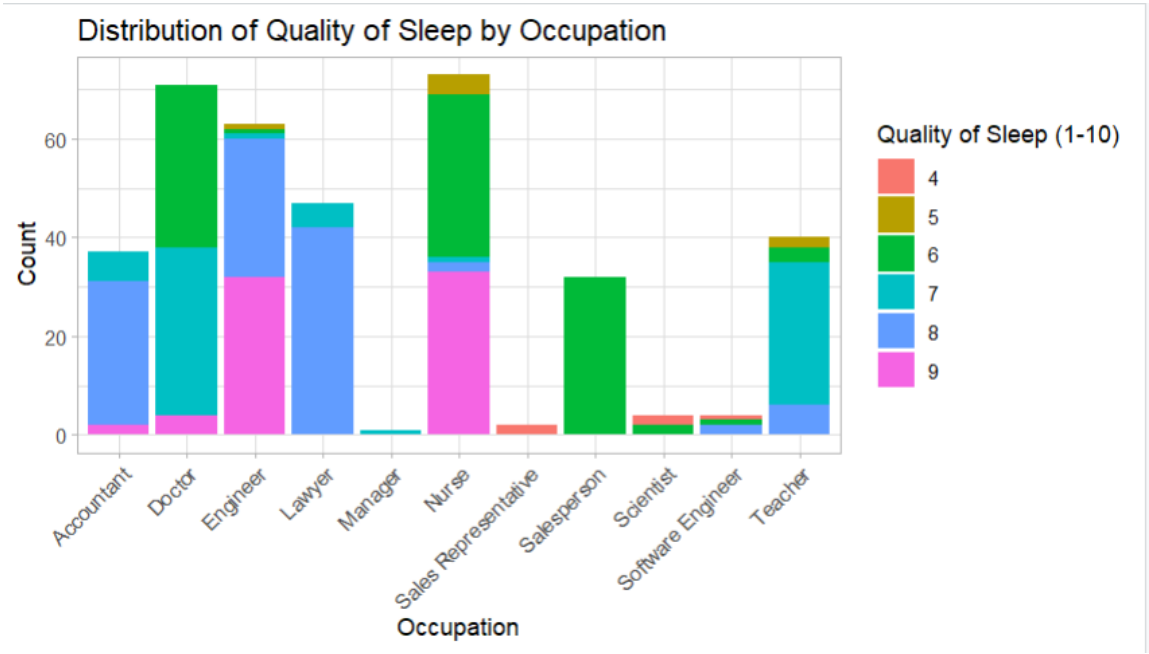
Key findings:

- The median quality of sleep tends to decrease as people get older. This is indicated by the fact that the line inside the box (which represents the median) is lower for older age groups.
- There is a wider range of sleep quality scores for younger people. This is shown by the longer whiskers and outliers on the left side of the boxplot for younger age groups.
- The range of sleep quality scores becomes narrower as people get older. This is shown by the shorter whiskers and fewer outliers on the right side of the boxplot for older age groups.

Interpretation:

As people get older, they tend to sleep less well. However, there are some young people who sleep poorly and some older people who sleep well. Overall, though, younger people tend to have a wider range of sleep quality experiences than older people.

- **Occupation vs Quality of Sleep**




```
> count_data <- dataset %>%
+     group_by(Occupation, `Quality of Sleep`) %>%
+     tally() %>%
+     ungroup()
>
> ggplot(count_data, aes(x = Occupation, y = n, fill = factor(`Quality of Sleep`))) +
+     geom_bar(stat = "identity") +
+     theme_light() +
+     labs(title = "Distribution of Quality of Sleep by Occupation",
+          x = "Occupation",
+          y = "Count",
+          fill = "Quality of Sleep (1-10)") +
+     theme(axis.text.x = element_text(angle = 45, hjust = 1))
>
```

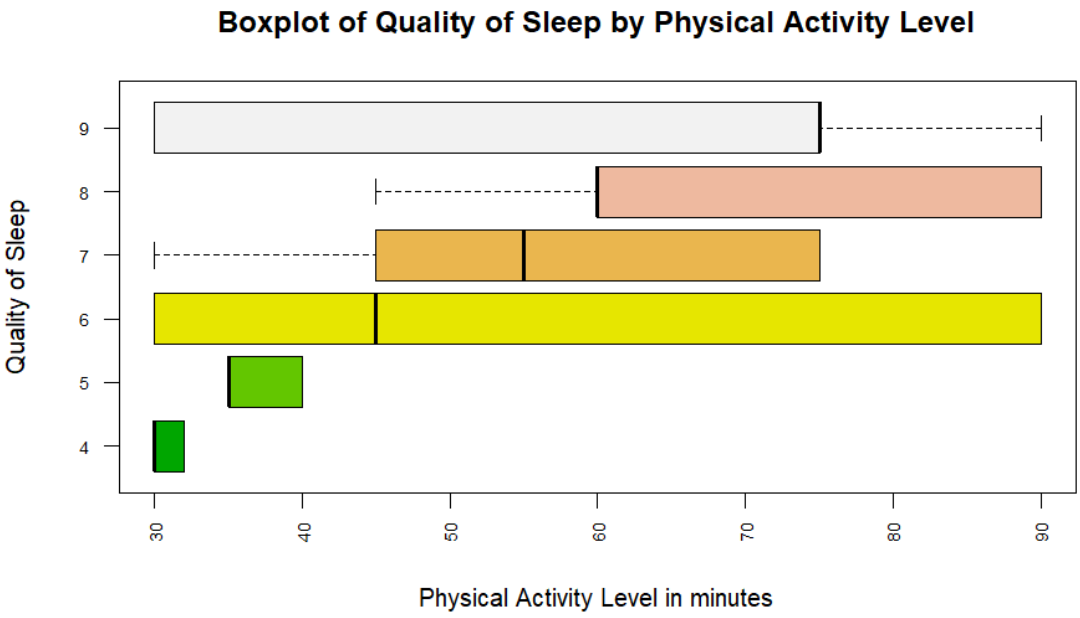
Key Observations:

- Most common sleep quality: The most common sleep quality rating across all occupations is 7, followed by 8 and 6.
- Occupation-specific patterns: Some occupations exhibit distinct patterns in sleep quality distribution. For example, nurses have a higher proportion of individuals reporting sleep quality ratings of 8 and 9, while salespeople have a higher proportion of individuals reporting sleep quality ratings of 4 and 5.
- Variability: There is variability in sleep quality within each occupation. This is evident in the multiple colors representing different sleep quality ratings for each occupation.

Interpretation**:**

Accountants and teachers show a relatively wide range of sleep quality ratings, indicating that sleep quality varies significantly among individuals in these professions. Furthermore, Engineers and scientists have a relatively high proportion of individuals reporting sleep quality ratings of 6 and 7. Moreover, Managers have a relatively high proportion of individuals reporting sleep quality ratings of 7 and 8. Overall, the chart suggests that there are variations in sleep quality among individuals in different occupations. However, further analysis would be needed to identify the underlying factors contributing to these differences.

- **Physical Activity Level vs Quality of Sleep**




```
> boxplot(`Physical Activity Level` ~ `Quality of Sleep`,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Physical Activity Level",
+         xlab = "Physical Activity Level in minutes",
+         ylab = "Quality of Sleep",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
```

Key findings:

- As physical activity level increases, the quality of sleep generally improves. This is indicated by the fact that the median quality of sleep (the line inside the box) increases as physical activity level increases.
- There is a wide range of sleep quality scores for people who take fewer steps. This is shown by the longer whiskers and outliers on the left side of the boxplot.
- The range of sleep quality scores becomes narrower as physical activity level increases. This is shown by the shorter whiskers and fewer outliers on the right side of the boxplot.

Interpretation:

The more physically active people are, the better they tend to sleep. However, there are some people who sleep well even if they are not very active, and there are some people who sleep poorly even if they are very active. Overall, though, being more physically

active is associated with better sleep quality.

- **Stress Level vs Quality of Sleep**


```
> boxplot(`Stress Level` ~ `Quality of Sleep`,
+         data = dataset,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Stress Level",
+         xlab = "Quality of Sleep",
+         ylab = "Stress Level ",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
>
```

Key Findings:

- High Stress Levels (9): Individuals with the highest stress levels show significantly lower sleep quality (around 3-4 on the Quality of Sleep scale). This suggests a strong negative correlation between stress and sleep quality.
- Moderate Stress Levels (7-8): These groups display greater variability in sleep quality,

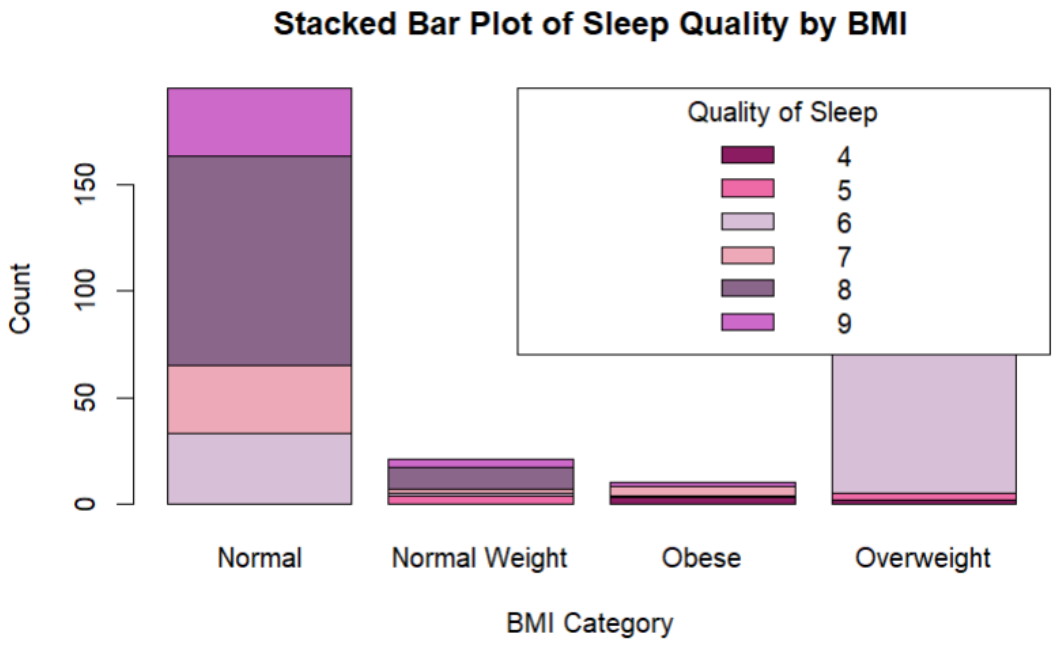
with scores ranging from 4 to 7. Despite the stress, many individuals still maintain moderate to high sleep quality, though some experience poor sleep.

- Outliers: In the moderate-stress group (Stress Level 8), there is an outlier with extremely low sleep quality, highlighting that even individuals with moderate stress can experience poor sleep in some cases.

Interpretation:

The boxplot suggests that increased stress generally leads to poorer sleep quality, but this relationship is not uniform. While higher stress levels correlate with lower sleep quality, moderate stress can result in a wide range of sleep experiences, from poor to good. This variability indicates that other factors, such as individual coping mechanisms or lifestyle choices, may also influence the impact of stress on sleep. The presence of outliers shows that stress does not uniformly dictate sleep outcomes for everyone.

- **BMI Category vs Quality of Sleep**



```
> barplot(table(`Quality.of.Sleep`, `BMI.Category`),
+ main="Stacked Bar Plot of Sleep Quality by BMI",
+ col= c("maroon4", "hotpink2", "thistle", "pink2", "plum4", "orchid3"),
+ xlab = "BMI Category",
+ ylab = "Count",
+ beside = FALSE)
```

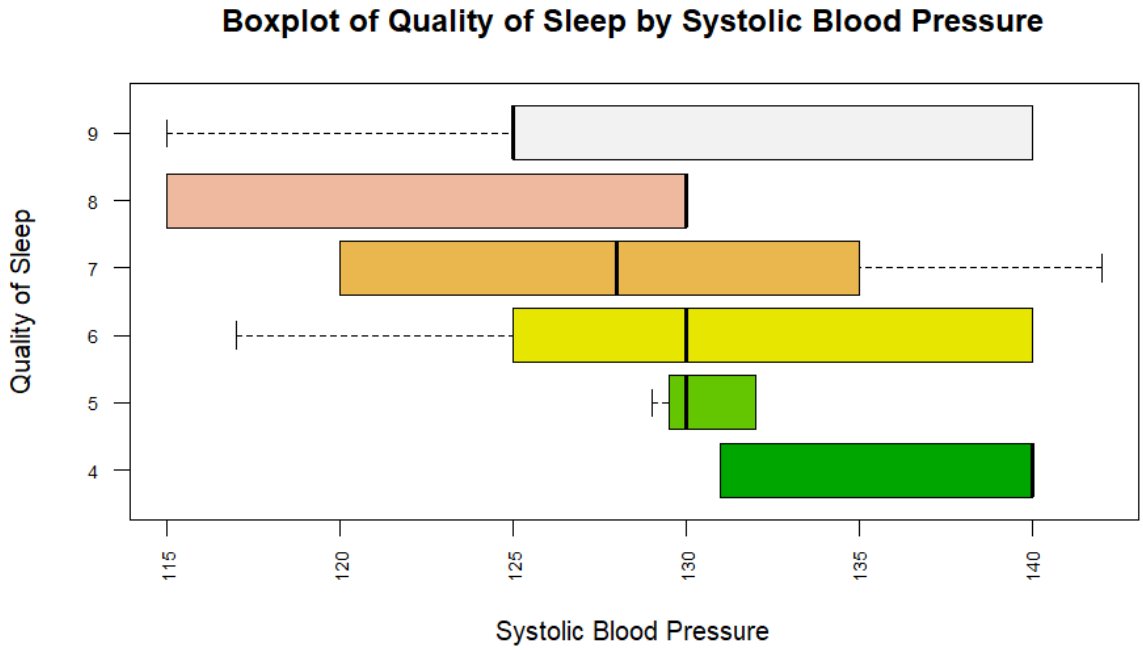
Key Findings:

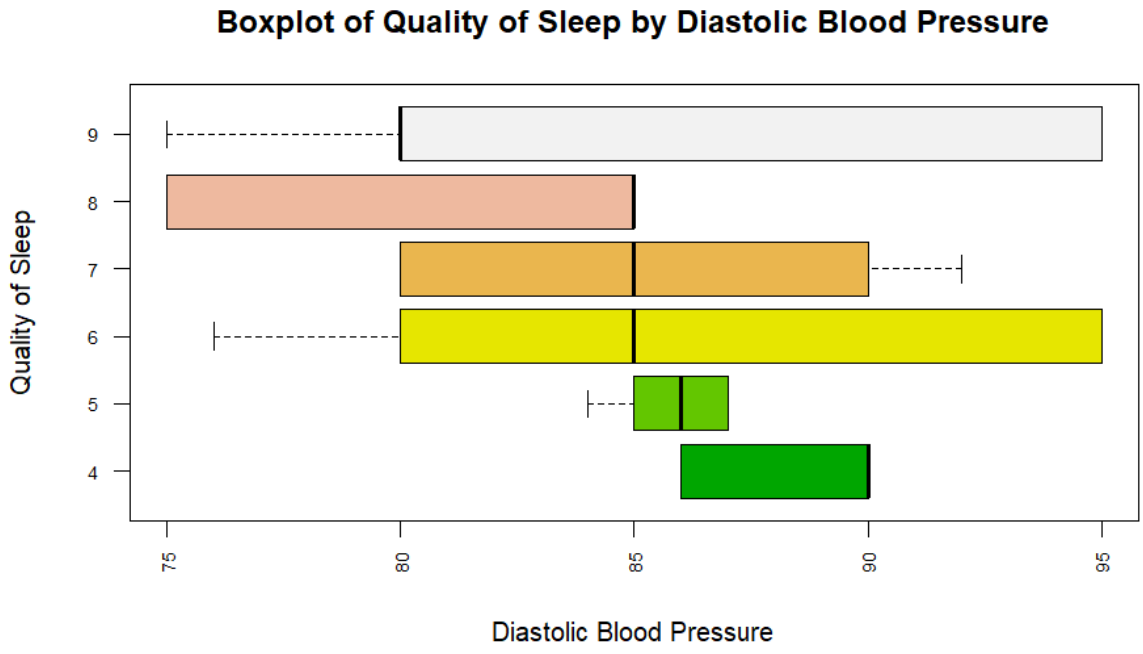
- Highest Population in Normal BMI: The majority of individuals belong to the Normal BMI category, with the Overweight group following closely behind, while Obese and Normal Weight categories have significantly fewer participants.
- Sleep Quality Varies Across All BMI Groups: Sleep quality scores (ranging from 4 to 9) are spread across all BMI categories, with the Normal and Overweight groups reporting the most balanced range of sleep quality.
- Moderate Sleep Quality is Common: The most frequent sleep quality ratings are between 6 and 7, particularly in the Normal and Overweight categories, indicating that most people experience moderate sleep quality.

Interpretation:

The data suggests that BMI does not have a clear or strong influence on sleep quality, as individuals across all BMI categories report a wide range of sleep experiences. While the Normal and Overweight groups have a higher number of individuals, with many reporting moderate sleep quality scores, the smaller Obese and Normal Weight groups also display varied sleep quality. Overall, there is no consistent pattern indicating that higher or lower BMI corresponds directly to better or worse sleep quality.

- **Blood Pressure vs Quality of Sleep**





```
> systolic_bp <- as.numeric(gsub("/.*", "", `Blood Pressure`))
> diastolic_bp <- as.numeric(gsub(".*/", "", `Blood Pressure`))
>
> # Create separate boxplots for systolic and diastolic blood pressure
> boxplot(systolic_bp ~ `Quality of Sleep`,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Systolic Blood Pressure",
+         xlab = "Systolic Blood Pressure",
+         ylab = "Quality of Sleep",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
>
> boxplot(diastolic_bp ~ `Quality of Sleep`,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Diastolic Blood Pressure",
+         xlab = "Diastolic Blood Pressure",
+         ylab = "Quality of Sleep",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
```

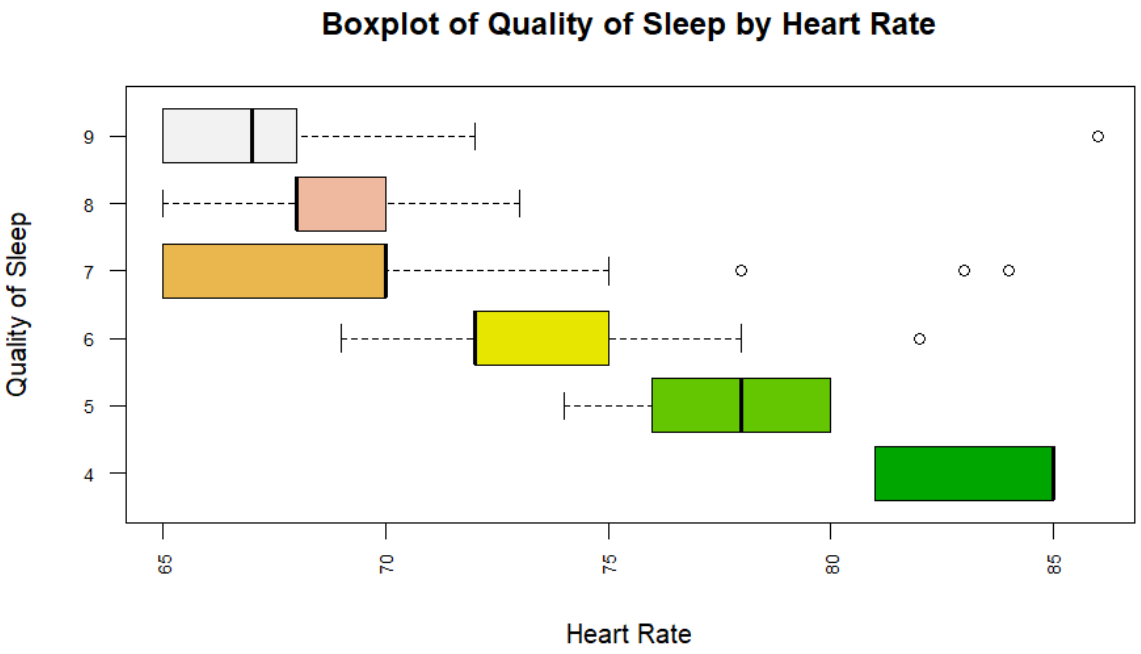
Key Findings:

- Both systolic and diastolic blood pressure are associated with sleep quality. As either systolic or diastolic blood pressure increases, the overall quality of sleep tends to decrease.
- The relationship between blood pressure and sleep quality is not linear. While higher blood pressure is generally associated with lower sleep quality, there is a wide range of sleep quality scores for people with both high and low blood pressure.
- Diastolic blood pressure seems to have a slightly stronger relationship with sleep quality than systolic blood pressure. This is suggested by the more consistent pattern of decreasing sleep quality as diastolic blood pressure increases.

Interpretation:

People with lower blood pressure, both systolic and diastolic, tend to sleep better than those with higher blood pressure. However, there are exceptions to this general trend, and other factors may also influence sleep quality. It's important to maintain healthy blood pressure levels to support overall well-being, including sleep.

- **Heart Rate vs Quality of Sleep**




```
> boxplot(`Heart Rate` ~ `Quality of Sleep`,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Heart Rate",
+         xlab = "Heart Rate",
+         ylab = "Quality of Sleep",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
```

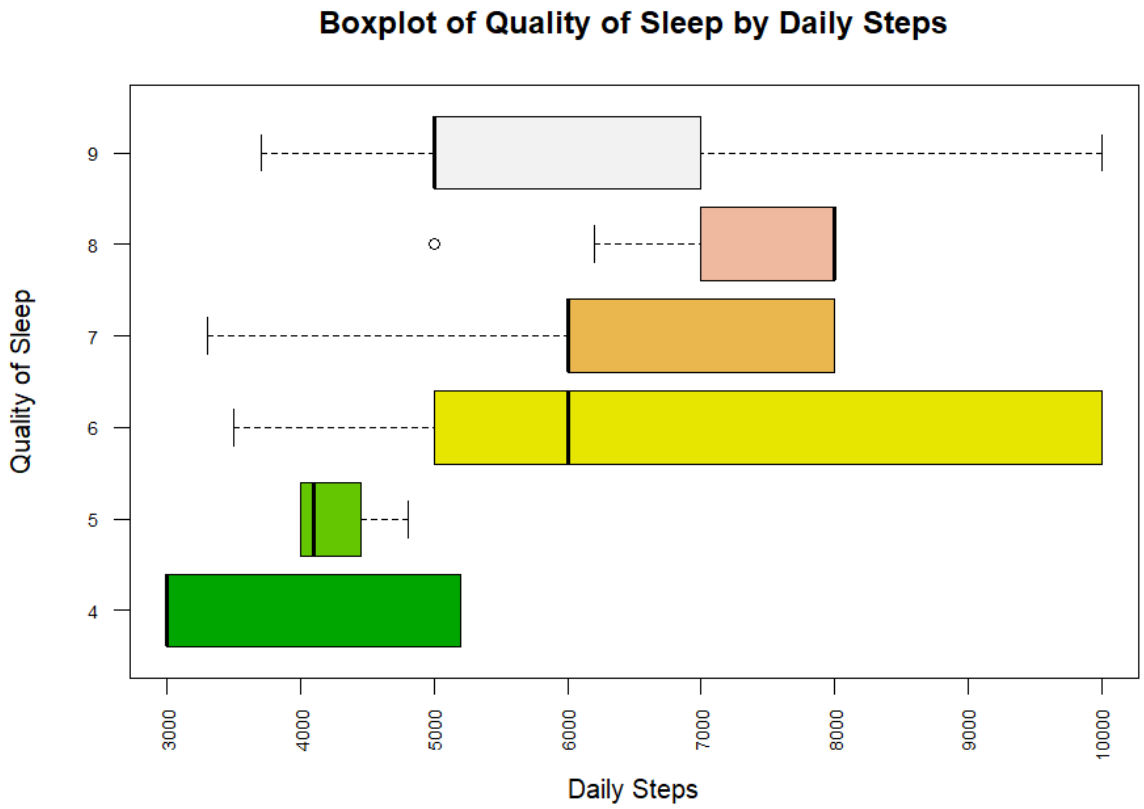
Key findings:

- As heart rate increases, the quality of sleep tends to decrease. This is indicated by the fact that the median quality of sleep (the line inside the box) decreases as heart rate increases.
- There is a wide range of sleep quality scores for people with lower heart rates. This is shown by the longer whiskers and outliers on the left side of the boxplot.
- The range of sleep quality scores becomes narrower as heart rate increases. This is shown by the shorter whiskers and fewer outliers on the right side of the boxplot.

Interpretation:

People with lower heart rates tend to sleep better than people with higher heart rates. However, there are some people with low heart rates who sleep poorly, and there are some people with high heart rates who sleep well. Overall, though, a lower heart rate is associated with better sleep quality.

- **Daily Steps vs Quality of Sleep**




```
> boxplot(`Daily Steps` ~ `Quality of Sleep`,
+         col = terrain.colors(6),
+         main = "Boxplot of Quality of Sleep by Daily Steps",
+         xlab = "Daily Steps",
+         ylab = "Quality of Sleep",
+         horizontal = TRUE,
+         las = 2,
+         cex.axis = 0.7)
```

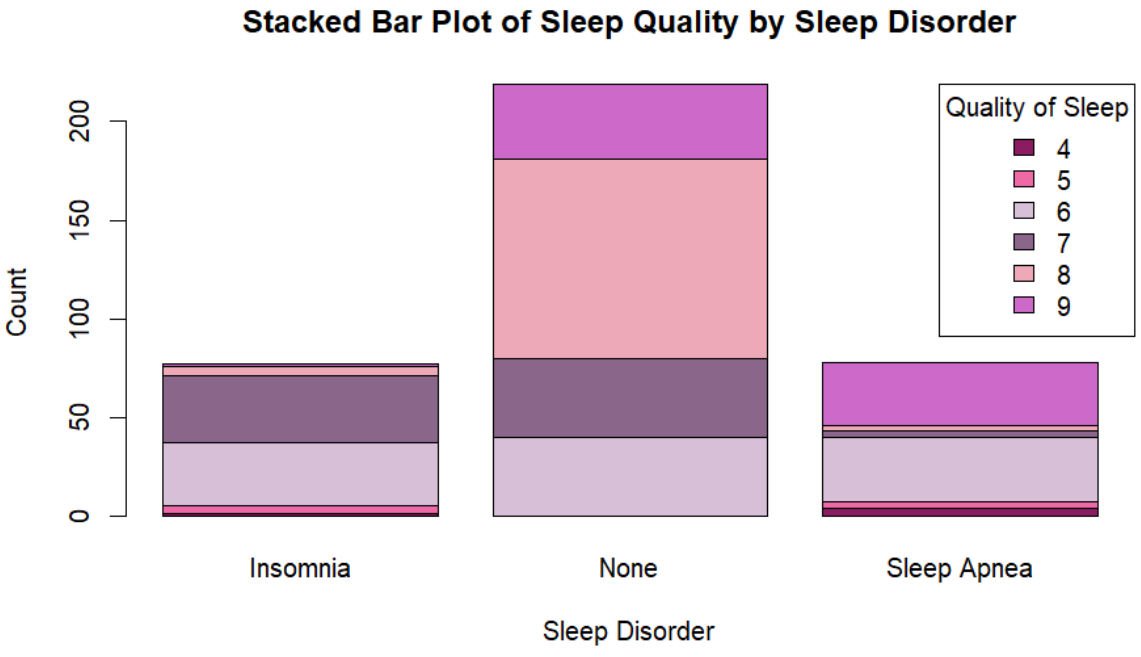
Key findings:

- As the number of daily steps increases, the quality of sleep generally improves. This is indicated by the fact that the median quality of sleep (the line inside the box) increases as the number of daily steps increases.
- There is a wide range of sleep quality scores for people who take fewer steps. This is shown by the longer whiskers and outliers on the left side of the boxplot.
- The range of sleep quality scores becomes narrower as the number of daily steps increases. This is shown by the shorter whiskers and fewer outliers on the right side of the boxplot.

Interpretation:

The more steps people take each day, the better they tend to sleep. However, there are some people who sleep well even if they don't take many steps, and there are some people who sleep poorly even if they take a lot of steps. Overall, though, taking more steps is associated with better sleep quality.

- **Sleep Disorder vs Quality of Sleep**



```
> barplot(table(`Quality of Sleep`,`Sleep Disorder`),
+         main = "Stacked Bar Plot of Sleep Quality by Sleep Disorder",
+         col = c("maroon4", "hotpink2", "thistle","plum4","pink2","orchid3"),
+         xlab = "Sleep Disorder",
+         ylab = "Count",
+         beside = FALSE)
> legend("topright",
+        title = "Quality of Sleep",
+        legend = sort(unique(`Quality of Sleep`)),
+        fill = c("maroon4", "hotpink2", "thistle", "plum4", "pink2", "orchid3"))
```

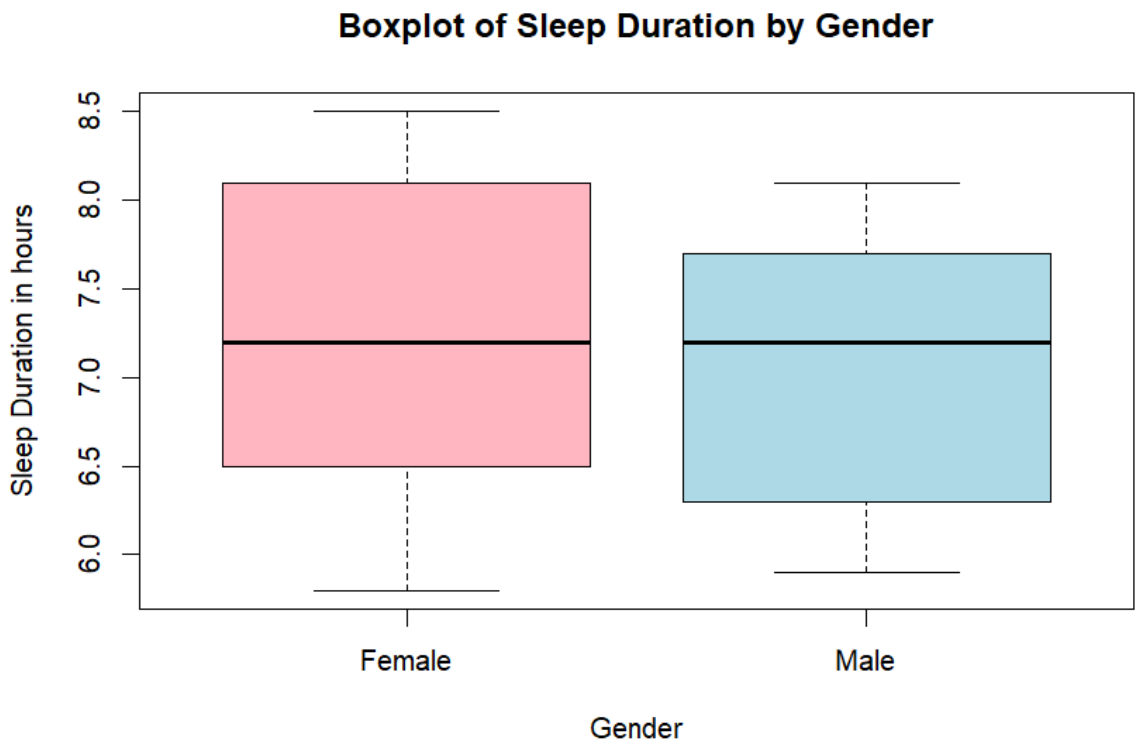
Key findings:

- People with insomnia tend to have lower sleep quality scores. This is shown by the larger proportion of darker shades (representing lower scores) in the insomnia bar.
- People with sleep apnea also tend to have lower sleep quality scores than those without any sleep disorder. However, their sleep quality is not as low as those with insomnia.
- People without any sleep disorder generally report higher sleep quality scores. This is shown by the larger proportion of lighter shades (representing higher scores) in the "None" bar.

Interpretation:

People who have insomnia or sleep apnea tend to sleep worse than people who don't have any sleep problems. People with insomnia have the most trouble sleeping, while people with sleep apnea have some difficulties but not as severe as those with insomnia. People without any sleep disorder generally sleep the best.

- **Gender vs Sleep Duration**




```
> boxplot(`Sleep Duration` ~ Gender,
+         col = c("lightpink", "lightblue"),
+         main = "Boxplot of Sleep Duration by Gender",
+         xlab = "Gender",
+         ylab = "Sleep Duration in hours")
```

Key insights:

Females:

- The median sleep duration for females is around 7.5 hours. This means that half of the females sleep more than 7.5 hours, and the other half sleep less.
- Most women's sleep duration ranges from about 6 to 8.5 hours. The box shows where most women fall in terms of sleep, and the whiskers extending from the box indicate the lowest and highest values.

Males:

- The median sleep duration for males is slightly lower, at about 7.25 hours. Again, half

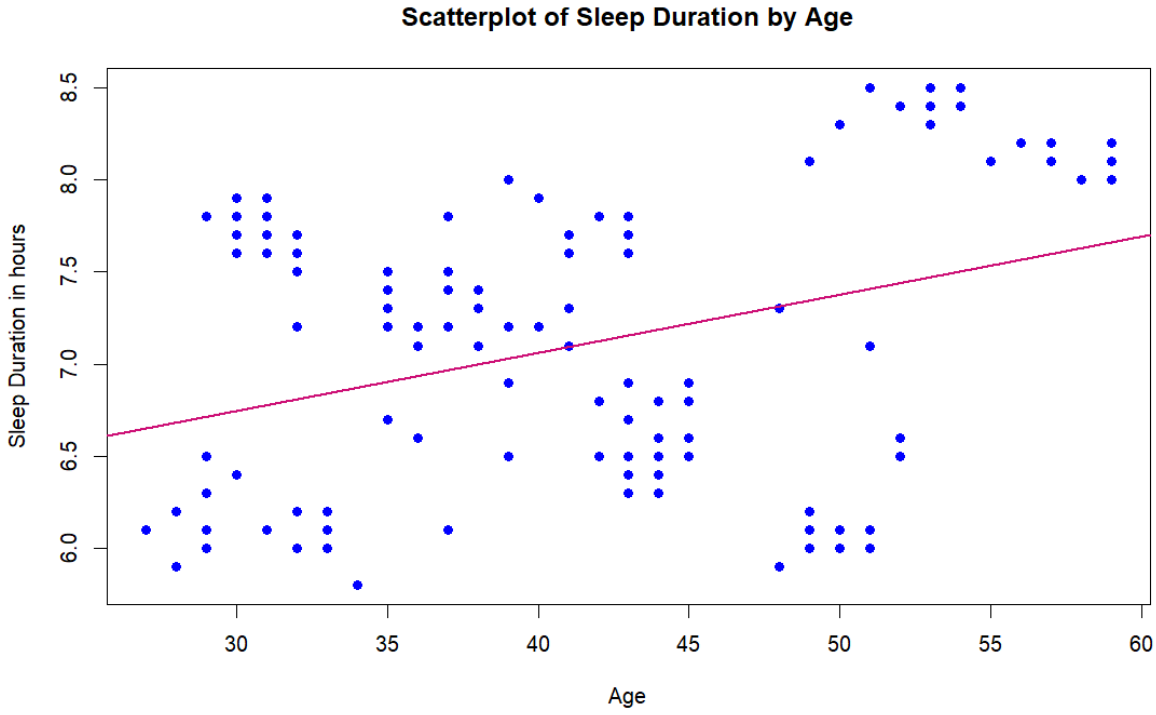
of the males sleep more than this amount, and half sleep less.

- The range for males is similar to that of females, but on average, they tend to sleep a little less.

Interpretation:

Females seem to get a bit more sleep on average compared to males. However, the difference isn't huge, and the range of sleep times overlaps for both genders. So, while there might be a slight tendency for women to sleep more, many men and women have similar sleep patterns.

- **Age vs Sleep Duration**



```
plot(Age, `Sleep Duration`,
     main = "Scatterplot of Sleep Duration by Age",
     xlab = "Age",
     ylab = "Sleep Duration in hours",
     pch = 16,
     col = "blue")
model <- lm(`Sleep Duration` ~ Age)


abline(model, col = "deeppink3", lwd = 2)
```

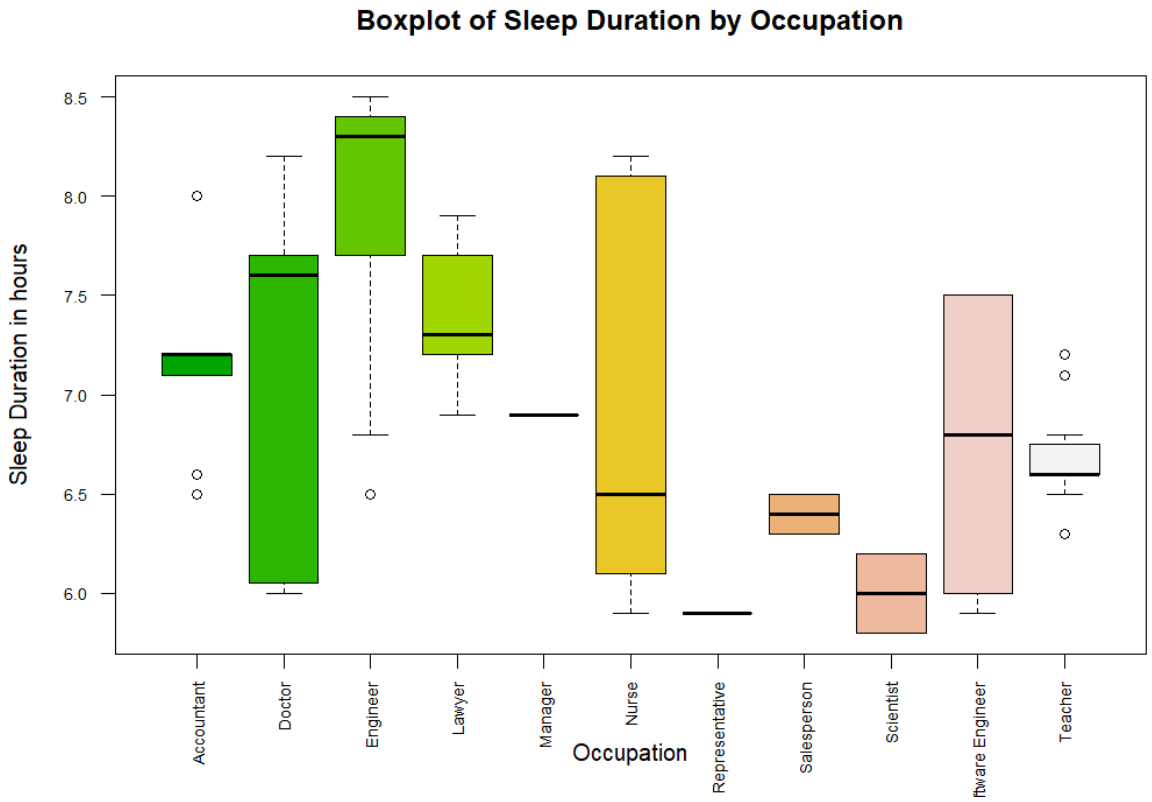
Key Findings:

- Younger adults (in their 30s and 40s): There's a lot of variation in sleep. Some younger people sleep around 6 hours, while others sleep closer to 8 hours. However, many younger adults seem to sleep less than 7 hours.
- Middle-aged and older adults (in their 50s and 60s): People in this age group seem to get more consistent sleep, with a larger number clustering around 7.5 to 8 hours. There are fewer people in this age group who sleep less than 6.5 hours.
- General trend: While there's no strict rule, there's a tendency for people to sleep a bit longer as they age. Younger people tend to have more variability in how much they sleep, while older people tend to have a more regular sleep duration.

Interpretation:

People's sleep habits vary a lot, but one interesting pattern is that younger adults (especially those in their 30s and 40s) often sleep less and have more irregular sleep patterns. As people get older, they seem to settle into a more regular sleep schedule, with many sleeping around 8 hours. So, it seems that older people tend to sleep slightly more than younger adults, though everyone is different, and the amount of sleep needed can vary widely from person to person.

- **Occupation vs Sleep Duration**




```
> boxplot(`Sleep Duration` ~ Occupation,
+         col = terrain.colors(11),
+         main = "Boxplot of Sleep Duration by Occupation",
+         xlab = "Occupation",
+         ylab = "Sleep Duration in hours",
+         las = 2,
+         cex.axis = 0.7)
```

Key Insights:

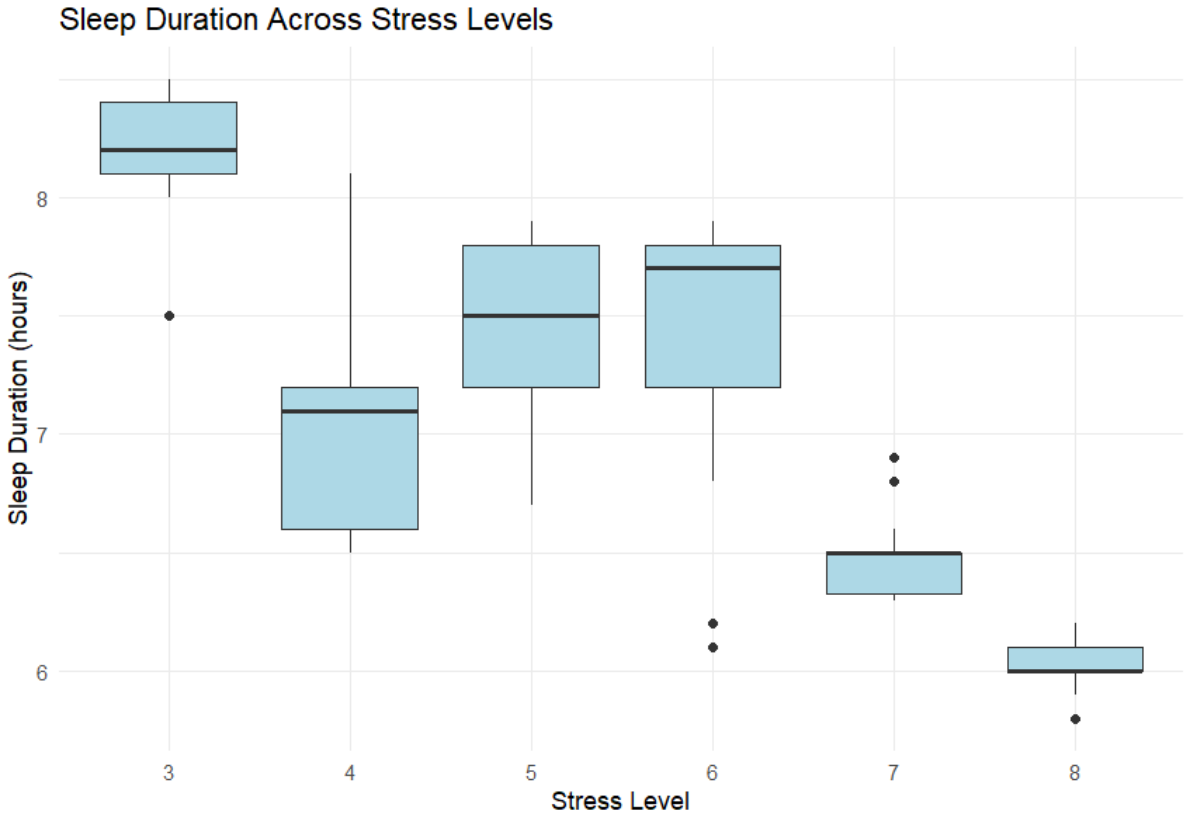
- Doctors: The sleep duration among doctors is highly variable, with some sleeping as little as 6 hours and others sleeping more than 8 hours. The median sleep duration is around 7.5 hours, indicating that half of the doctors sleep more and half sleep less than this amount.

- Engineers: Engineers tend to have more consistent sleep patterns, typically getting between 7.5 to 8 hours of sleep. The narrow range suggests that their sleep habits are more uniform compared to other professions.
- Nurses: Sleep patterns among nurses vary widely, with some getting over 8 hours of sleep while others get as little as 6 hours. This variation reflects the irregularity of their work schedules, likely due to shifts.
- Software Engineers and Scientists: These professions appear to have shorter sleep durations, with a median around 6.5 to 7 hours. This suggests that many individuals in these jobs may not get as much rest as those in other professions.
- Teachers, Accountants, and Salespersons: Individuals in these professions tend to sleep less, with typical sleep durations between 6.5 to 7 hours. Their median sleep duration is on the lower side compared to doctors and engineers.

Interpretation:

The boxplot highlights significant differences in sleep duration based on occupation. Professions like doctors and engineers, despite the variability in doctors' sleep, generally show higher median sleep durations. In contrast, software engineers, scientists, teachers, accountants, and salespersons tend to sleep less, with more individuals in these professions getting fewer than 7 hours of sleep on average. This suggests that certain professions may face more demanding schedules or workloads, potentially leading to shorter or more inconsistent sleep patterns. Overall, while there is variability within each group, certain occupational trends in sleep duration are evident.

- **Stress Level vs Sleep Duration**




```
>   ggplot(sleep_data, aes(x=as.factor(`Stress Level`),
+                          y=`Sleep Duration`)) +
+   geom_boxplot(fill="lightblue") +
+   labs(title="Sleep Duration Across Stress Levels",
+        x="Stress Level",
+        y="Sleep Duration (hours)") +
+   theme_minimal()
```

Key findings:

- The box plot shows the relationship between these two variables. There seems to be a negative correlation between stress level and sleep duration. As stress levels increase, the median sleep duration tends to decrease.
- The box plots show that there is variability in sleep duration within each stress level. The length of the boxes indicates the spread of data, while the whiskers represent the

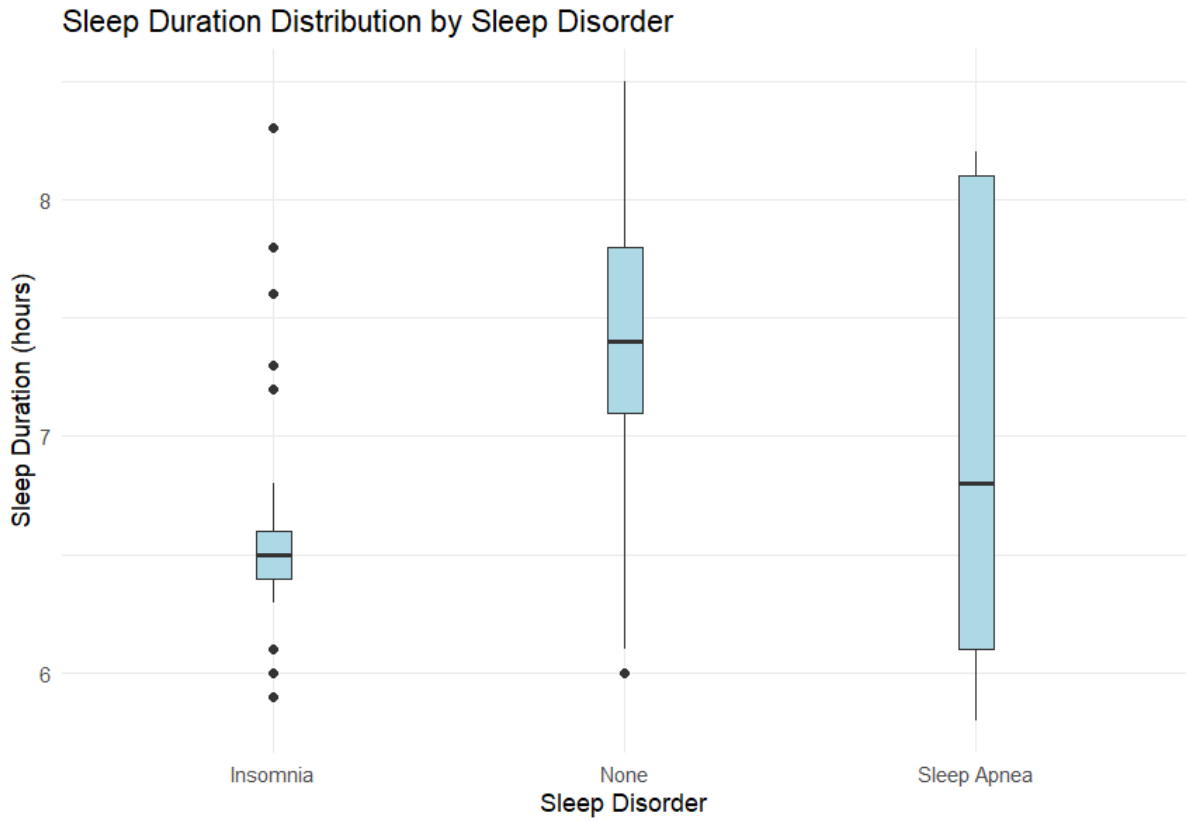
range of values excluding outliers.

- There are a few outliers present in the data, especially in the higher stress level groups. These are individuals who report either exceptionally high or low sleep duration despite their stress levels.

Interpretation:

Overall, the graph suggests that higher stress levels are associated with shorter sleep duration. However, individual experiences may vary, and other factors like physical activity, diet, and sleep environment can also influence sleep duration.

- **Sleep Disorder vs Sleep Duration**




```
> ggplot(sleep_data, aes(x = `Sleep Disorder`, y = `Sleep Duration`)) +
+   geom_boxplot(width = 0.1, fill = "lightblue") +
+   labs(title = "Sleep Duration Distribution by Sleep Disorder",
+        x = "Sleep Disorder",
+        y = "Sleep Duration (hours)") +
+   theme_minimal()
```

Key findings:

- Sleep apnea: Individuals with sleep apnea tend to have the lowest sleep duration, as indicated by the median and the overall distribution of the box plot.
- Insomnia: Individuals with insomnia also exhibit lower sleep duration compared to those without any sleep disorder.
- No sleep disorder: Individuals without any sleep disorder tend to have the highest sleep duration, with a median around 7.5 hours.
- Variability: There is variability in sleep duration within each sleep disorder group. This is evident in the length of the boxes and the presence of outliers.

Interpretation:

Overall, the graph suggests that sleep disorders, particularly sleep apnea, are associated with shorter sleep duration. Individuals with insomnia may also experience some sleep disruption, leading to lower sleep duration compared to those without any sleep disorder. However, individual experiences may vary, and other factors like stress, physical activity, and lifestyle can also influence sleep duration.

Conclusion

The report provides a detailed analysis of the relationship between sleep health and various lifestyle factors, offering valuable insights into how different elements of daily life impact both the quality and duration of sleep. It reveals that individuals who engage in regular physical activity tend to experience better sleep quality and longer sleep durations. Conversely, those with higher stress levels are more likely to suffer from poor sleep, indicating the importance of stress management in maintaining healthy sleep patterns.

Cardiovascular health also plays a significant role in sleep, as people with lower heart rates and healthier blood pressure levels report better overall sleep quality. The report highlights the variation in sleep patterns across different professions, with engineers and doctors generally experiencing better and more consistent sleep than those in other fields. This could be attributed to more structured work schedules, while professions with more irregular hours, like nurses, exhibit greater variability in sleep duration.

Additionally, the study underscores the impact of sleep disorders, such as sleep apnea and insomnia, on sleep quality and duration. Those diagnosed with any of these conditions show notably lower sleep quality compared to individuals without any sleep disorders. In conclusion, the report emphasizes the importance of a balanced lifestyle including physical activity, stress management, and maintaining cardiovascular health as crucial for improving sleep quality and preventing sleep related disorders. These findings can inform strategies to promote better sleep hygiene and overall well-being.

Appendices

Univariate Analysis :

1. **Gender:**

   dataset <- Sleep_health_and_lifestyle_dataset Gender_Freq <- table(dataset$Gender) Gender_Precentages <- prop.table(Gender_Freq) * 100 Colours <- c("yellow", "lightgreen") pie(Gender_Freq, labels = sprintf("%.1f%%", Gender_Precentages), main = "Gender", col = Colours) legend("topright", legend = c("Female", "Male"), fill = Colours)

2. **Age:**

   hist(dataset$Age, breaks = seq(0, 60, by = 5), col = "lightblue", xlab = "Age", ylab = "Frequency", main = "Histogram of Age")

3. **Occupation:**

   barplot(table(dataset$Occupation), names.arg = c("Accountant", "Doctor", "Engineer", "Lawyer", "Manager", "Nurse", "Sales Representative", "Salesperson", "Scientist", "Software Engineer", "Teacher"), main = "Barplot of Occupation", ylab = "Frequency", col = c("gray", "lightblue", "lightgreen", "pink", "lightblue", "yellow", "gray", "violet", "orange", "purple", "lightgreen"), las = 2, cex.names = 0.7) title(xlab = "Occupation", line = 5)

4. **Sleep Duration:**

   hist(`Sleep Duration`, breaks = 10, main = "Histogram of Sleep Duration", xlab = "Sleep Duration in hours", ylab = "Frequency", col = "plum", border = "black")

5. **Quality of Sleep**:

   categories <- c("4","5","6","7","8","9")

   barplot(table(`Quality of Sleep`), names.arg = categories, col = "lightcoral", main = "Bar Plot of the Quality of Sleep", xlab = "Rating", ylab = "Frequency")

6. **Physical Activity Level:**

   hist(`Physical Activity Level`, breaks = 10, main = "Histogram of Physical Activity Level", xlab = "Physical Activity Level in minutes", ylab = "Frequency", col = "yellowgreen", border = "black")

7. **Stress Level**:

   barplot(table(`Stress Level`), col = "lavender", main = "Bar Plot of Stress Level", xlab = "Rating", ylab = "Frequency")

8. **BMI Categories :**

barplot(bmi_frequency, main = “Frequency Distribution of BMI Categories”, xlab = “BMI Category”, ylab = “Frequency”, col = “skyblue”, ylim = c(0, max(bmi_frequency) + 5))

9. **Blood Pressure :**

barplot(bp, main = “Barplot for Blood Pressure”, xlab = “Blood Pressure”, ylab = “Frequency”, col = “lightgreen”, las = 2, cex.names = 0.7)

10. **Heart Rate** :

ggplot(SleepData, aes(x = `Heart Rate`)) + geom_histogram(bins = 10, fill = "coral", color = "black") + labs(title = "Histogram of Heart Rate", x = "Heart Rate", y = "Frequency") + theme_minimal()

11. **Daily Steps :**

ggplot(SleepData, aes(x = `Daily Steps`)) + geom_histogram(bins = 10, fill = "cadetblue3",color = "black") + labs(title = "Histogram of Daily Steps", x = "Daily Steps", y = "Frequency") + theme_minimal()

12. **Sleep Disorder :**

Dataset = Sleep_health_and_lifestyle_dataset

Sleepdisorder = table(dataset$Sleep.Disorder)

Colours = c(“lightblue”, “yellow”, “red”)

pie(Sleepdisorder, labels = sprintf(“%.1f%%”, disorder_Presentages), Main = “Sleep Disorder”, col = Colours)

legend(“topleft”, legend = c(“None”, “Sleep Apnea”, Insomnia), fill = Colours)

Bivariate Analysis :

1. **Sleep Quality by Gender:**

barplot(table(`Quality of Sleep`,Gender), main = "Stacked Bar Plot of Sleep Quality by Gender", col = c("maroon4", "hotpink2", "thistle","plum4","pink2","orchid3"), xlab = "Gender", ylab = "Count", beside = FALSE)

legend("topright", title = "Quality of Sleep", legend = sort(unique(`Quality of Sleep`)), fill = c("maroon4", "hotpink2", "thistle", "plum4", "pink2", "orchid3"))

2. **Quality of Sleep by Age**:

boxplot(Age ~ `Quality of Sleep`, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Age", xlab = "Age", ylab = "Quality of Sleep", horizontal = TRUE, las = 2, cex.axis = 0.7)

3. **Quality of Sleep by Occupation:**

ggplot(dataset %>% group_by(Occupation, `Quality of Sleep`) %>% tally() %>% ungroup(), aes(x = Occupation, y = n, fill = factor(`Quality of Sleep`))) + geom_bar(stat = "identity") + theme_light() + labs(title = "Distribution of Quality of Sleep by Occupation", x = "Occupation", y = "Count", fill = "Quality of sleep (1-10)") + theme(axis.text.x = element_text(angle = 45, hjust = 1))

4. **Quality of Sleep by Physical Activity Level:**

boxplot(`Physical Activity Level` ~ `Quality of Sleep`, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Physical Activity Level", xlab = "Physical Activity Level in minutes", ylab = "Quality of Sleep", horizontal = TRUE, las = 2, cex.axis = 0.7)

5. Quality of Sleep by Stress Level:

boxplot(`Stress Level` ~ `Quality of Sleep`, data = dataset, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Stress Level", xlab = "Quality of Sleep", ylab = "Stress Level", horizontal = TRUE, las = 2, cex.axis = 0.7)

6. **Sleep Quality by BMI:**

barplot(table(`Quality of Sleep`, `BMI.Category`), main="Stacked Bar Plot of Sleep Quality by BMI", col=c("maroon4", "hotpink2", "thistle", "pink2", "plum4", "orchid3"), xlab="BMI Category", ylab="Count", beside=FALSE)

7. Quality of Sleep by Systolic Blood Pressure:

systolic_bp <- as.numeric(gsub("/.*", "", `Blood Pressure`))

diastolic_bp <- as.numeric(gsub(".*/", "", `Blood Pressure`))

boxplot(systolic_bp ~ `Quality of Sleep`, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Systolic Blood Pressure", xlab = "Systolic Blood Pressure", ylab = "Quality of Sleep", horizontal = TRUE, las = 2, .axis = 0.7)

boxplot(diastolic_bp ~ `Quality of Sleep`, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Diastolic Blood Pressure", xlab = "Diastolic Blood Pressure", ylab = "Quality of Sleep", horizontal = TRUE, las = 2, cex.axis = 0.7)

8. Quality of Sleep by Heart Rate:

boxplot(`Heart Rate` ~ `Quality of Sleep`, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Heart Rate", xlab = "Heart Rate", ylab = "Quality of Sleep", horizontal = TRUE, las = 2, cex.axis = 0.7)

9. Quality of Sleep by Daily Steps:

boxplot(`Daily Steps` ~ `Quality of Sleep`, col = terrain.colors(6), main = "Boxplot of Quality of Sleep by Daily Steps", xlab = "Daily Steps", ylab = "Quality of Sleep", horizontal = TRUE, las = 2, cex.axis = 0.7)

10. Sleep Quality by Sleep Disorder:

barplot(table(`Quality of Sleep`,`Sleep Disorder`), main = "Stacked Bar Plot of Sleep Quality by Sleep Disorder", col = c("maroon4", "hotpink2", "thistle","plum4","pink2","orchid3"), xlab = "Sleep Disorder", ylab = "Count", beside = FALSE)

legend("topright", title = "Quality of Sleep", legend = sort(unique(`Quality of Sleep`)), fill = c("maroon4", "hotpink2", "thistle", "plum4", "pink2", "orchid3"))

11. Sleep Duration by Gender:

boxplot(`Sleep Duration` ~ Gender, col = c("lightpink", "lightblue"), main = "Boxplot of Sleep Duration by Gender", xlab = "Gender", ylab = "Sleep Duration in hours")

12. Sleep Duration by Age:

plot(Age, `Sleep Duration`, main = "Scatterplot of Sleep Duration by Age", xlab = "Age", ylab = "Sleep Duration in hours", pch = 16, col = "blue")

model <- lm(`Sleep Duration` ~ Age)

abline(model, col = "deeppink3", lwd = 2)

13. Sleep Duration by Occupation:

boxplot(`Sleep Duration` ~ Occupation, col = terrain.colors(11), main = "Boxplot of Sleep Duration by Occupation", xlab = "Occupation", ylab = "Sleep Duration in hours", las = 2, cex.axis = 0.7)

14. Sleep Duration by Stress Levels :

gglot(sleep_data, aes(x=as.tactor(`Stress Level`), y = `Sleep Duration`)) + geom_boxplot(fill = “lightblue”) + labs(title = “Sleep Duration Across Stress Levels”, x=”Stress Level”, y=”Sleep Duration (hours)”) + theme_minimal()

15. Sleep Disorder by Sleep Duration :

ggplot(sleep_data, aes(x = `Sleep Disorder`, y = Sleep Duration)) + goem_boxplot(width = 0.1, fill = “lightblue”) + labs(title = “Sleep Duration Distribution by Sleep Disorder”, x = “Sleep Disorder”, y = “Sleep Duration (hours)”) + theme_minimal()

References

Dataset:

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

Guide:

https://www.w3schools.com/R/r_data_frames.asp

https://www.tutorialspoint.com/r/index.htm

Rstudio Software:

https://posit.co/download/rstudio-desktop/