CECS 450 Data Visualization Group I

Holly Lee, Ivy Ly, Daniel Moreno

Table of Contents

Sleep is Important!

Project Goal

- The goal of this semester project was to **use data visualization** techniques and strategies to present that data in **an effective way.**
- Specifically, we focused on the area of sleep!
- The question we want to answer is:

  What are the factors that contribute to the best quality and duration of sleep?

Datasets

- **Sleep Efficiency Dataset**
  - Found on Kaggle
  - Collected as part of a study conducted a group of artificial intelligence engineering students
  - 15 columns
    - Potential factors: Age, Caffeine consumption, Alcohol consumption
    - Sleep quality assessment: Sleep efficiency, REM sleep percentage, Deep sleep percentage Light sleep percentage
- **Sleep Health and Lifestyle Dataset**
  - Found on Kaggle
  - Collected by a Data Science student
  - 13 columns
    - Potential Factors: BMI Category, Sleep Disorder
    - Sleep quality assessment: Sleep Duration, Quality of Sleep

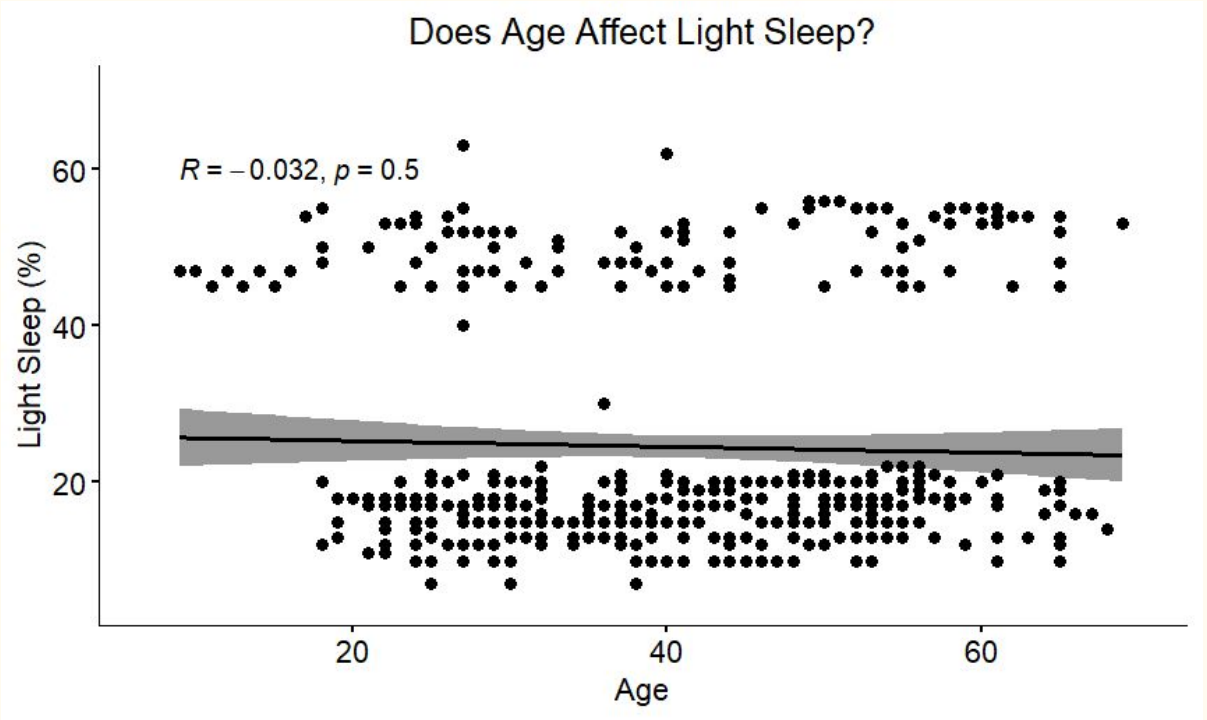
Definitions

- **Sleep Cycle**- when the human body cycles through two phases of sleep: NREM and REM sleep. A person typically goes through 4-6 cycles per night, which starts over every 80- 100 minutes.
- **NREM sleep**- divided in stages (N1, N2, N3) varying in heartbeat, brain waves, eye movements, and breathing activity.
- **REM sleep**- a kind of sleep that occurs at intervals characterized by rapid eye movements, more dreaming, bodily movement, and faster pulse and breathing.
- **Caffeine consumption** - the amount of caffeine consumed in the 24 hours prior to bedtime (in mg)
- **Alcohol consumption** - the amount of alcohol consumed in the 24 hours prior to bedtime (in oz). A standard drink in the United States contain 0.6 oz of pure alcohol. Which means a regular beer (12oz) usually contain about 5% alcohol and a 5 ounces of table wine, typically contains about 12% alcohol.
- **BMI** - Body Mass Index, derived from height and weight of a person, used to determine if the weight is considered “healthy” based on height

Age Related Findings

Age vs. Light Sleep

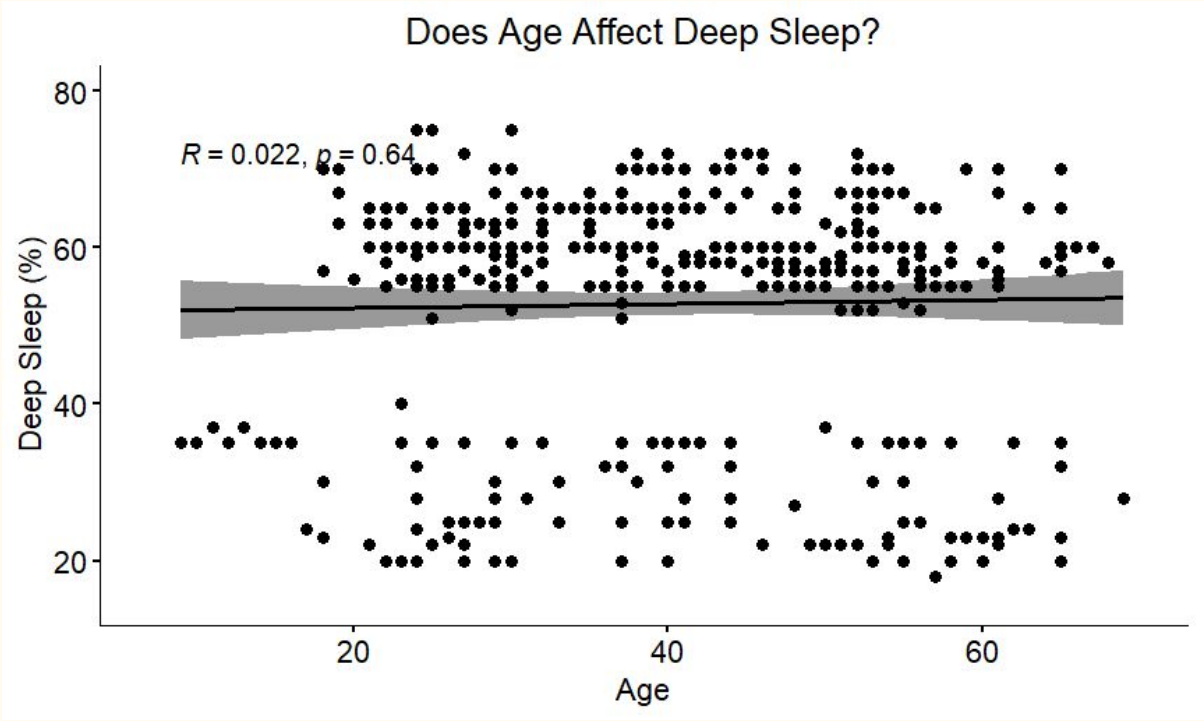
- N1 (light sleep) is the first stage of the sleep cycle when a person falls asleep, which last 1- 7 minutes. The body and brain activity slows down
- On avg., spend ~50%, but graph shows ~24.6%
- H_0 = age doesn't affect light sleep
- H_a = age affects light sleep
- R = -0.032
- P-value = 0.5




```
sleep_data <- sleep_efficiency_data_frame #%>% select(c("Age", "REM.sleep.percentage'
# plot the data as points and calculate
# correlation coefficient with pearson correlation formula
ggscatter(sleep_data, x = "Age", y = "Light.sleep.percentage",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          title = "Does Age Affect Light Sleep?",
          xlab = "Age", ylab = "Light Sleep (%)", xlim = c(7,70), ylim = c(5, 70)) +
  theme(plot.title = element_text(hjust = 0.5))
```

Age vs. Deep Sleep

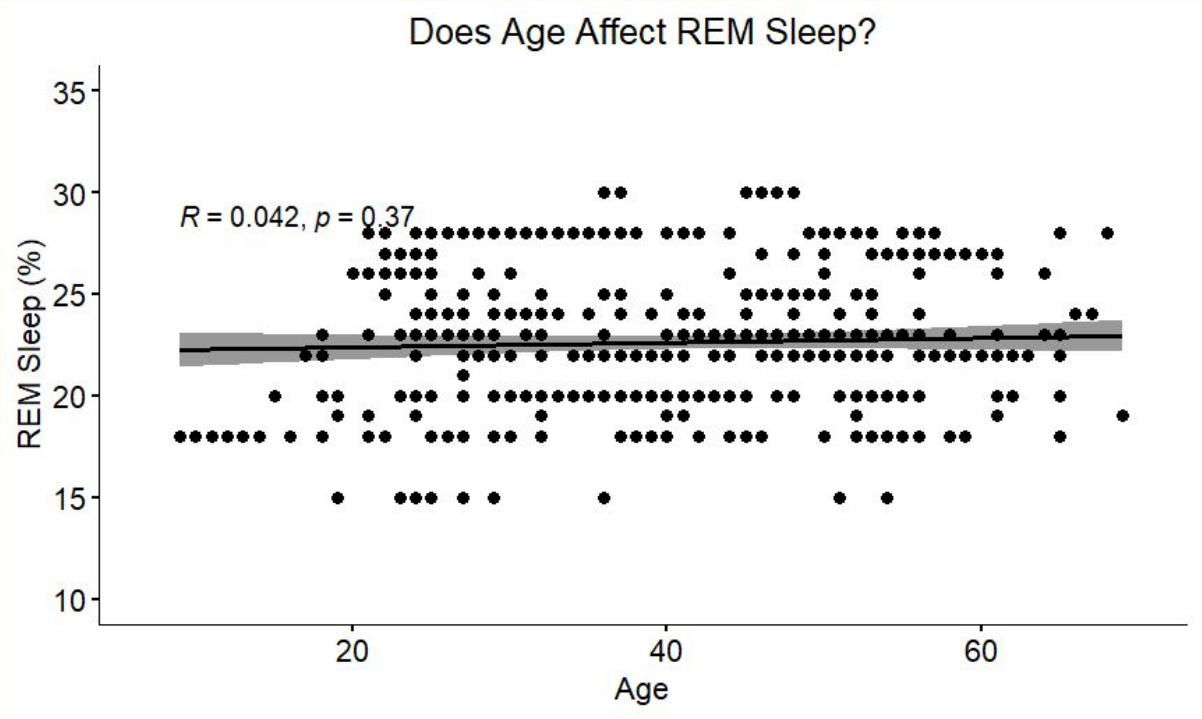
- N3 (deep sleep) is a phase where any environmental noises or activities may fail to wake the sleeping person
- On avg., spend ~15-25%, but graph shows ~52.8%
- H_0 = age doesn't affect deep sleep
- H_a = age affects deep sleep
- R = 0.022
- P-value = 0.64




```
sleep_data <- sleep_efficiency_data_frame #%>% select(c("Age", "REM.sleep.percentage'
#plot the data as points and calculate
#correlation coefficient with pearson correlation formula
ggscatter(sleep_data, x = "Age", y = "Deep.sleep.percentage",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          title = "Does Age Affect Deep Sleep?",
          xlab = "Age", ylab = "Deep Sleep (%)", xlim = c(7,70), ylim = c(15, 80)) +
  theme(plot.title = element_text(hjust = 0.5))
```

Age vs. REM Sleep

- A kind of sleep that occurs at intervals characterized by rapid eye movements, more dreaming, bodily movement, and faster pulse and breathing.
- On avg., spend ~25%, but graph shows ~22.6%
- H_0 = age doesn't affect REM sleep
- H_a = age affects REM sleep
- R = 0.042
- P-value = 0.37

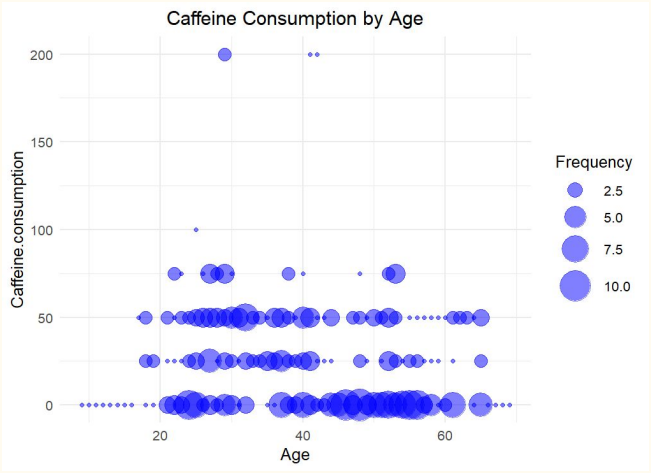


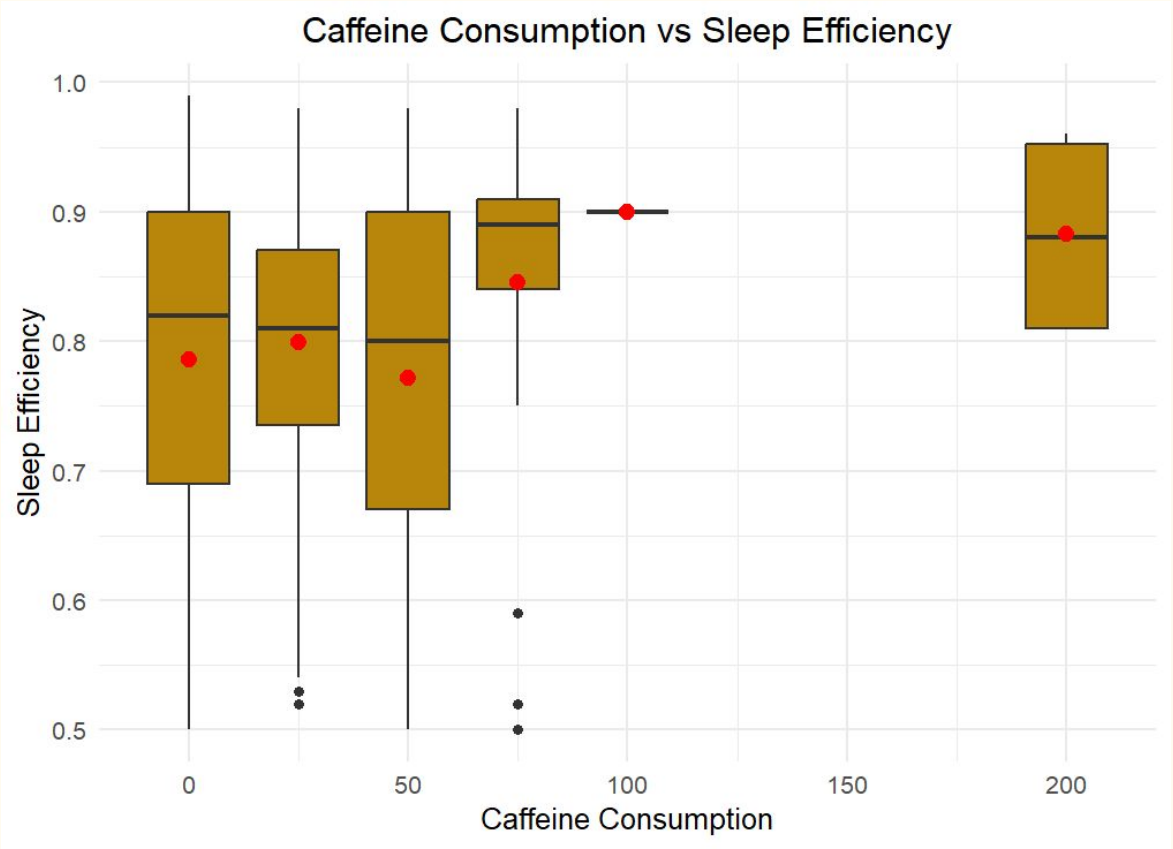

```
#read file
sleep_data <- sleep_efficiency_data_frame #%>% select(c("Age", "REM.sleep.percentage"
# plot the data as points and calculate
# correlation coefficient with pearson correlation formula
ggscatter(sleep_data, x = "Age", y = "REM.sleep.percentage",
          title = "Does Age Affect REM Sleep?",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Age", ylab = "REM Sleep (%)", xlim = c(7,70), ylim = c(10, 35)) +
  theme(plot.title = element_text(hjust = 0.5))
```

Findings regarding Substance consumption & Occupation

Does Caffeine Affect Sleep?

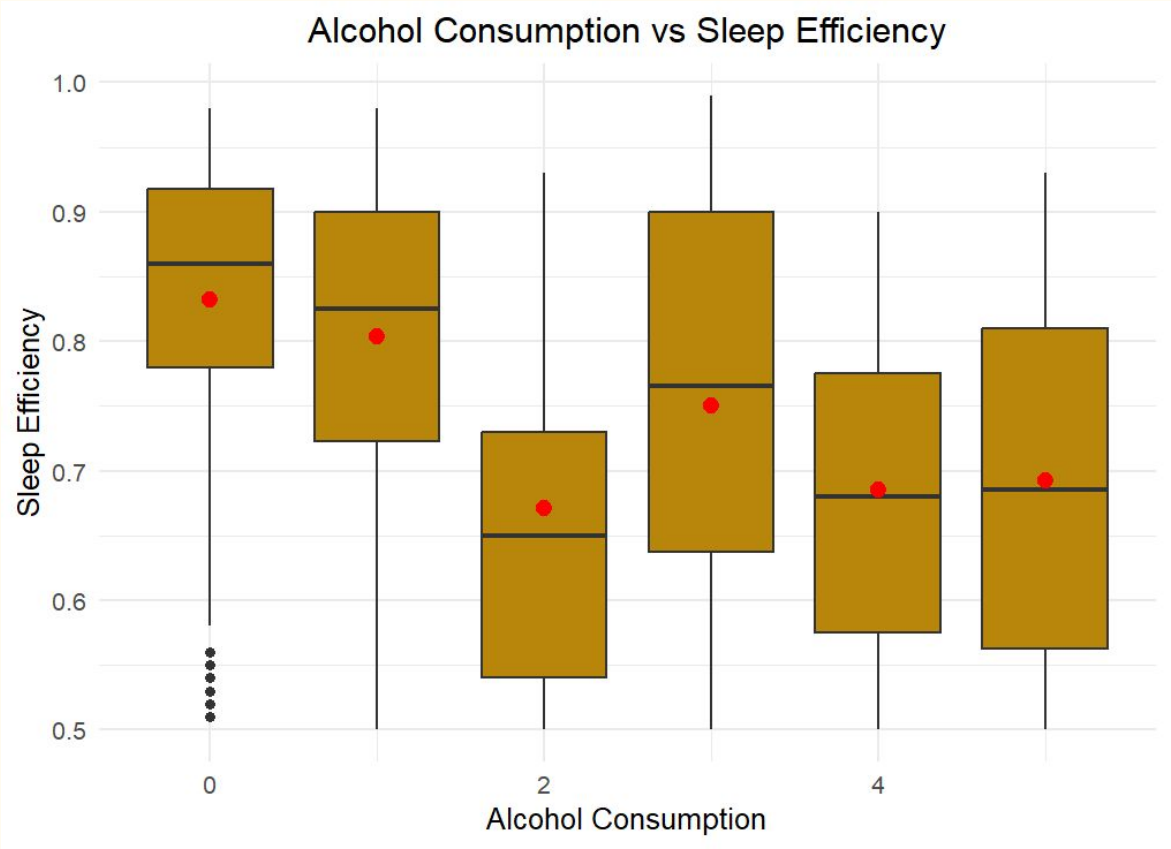
- Median decreased slightly at 25 mg and 50 mg caffeine consumption
- Median goes up to 0.87 - 0.9 when people consume more caffeine



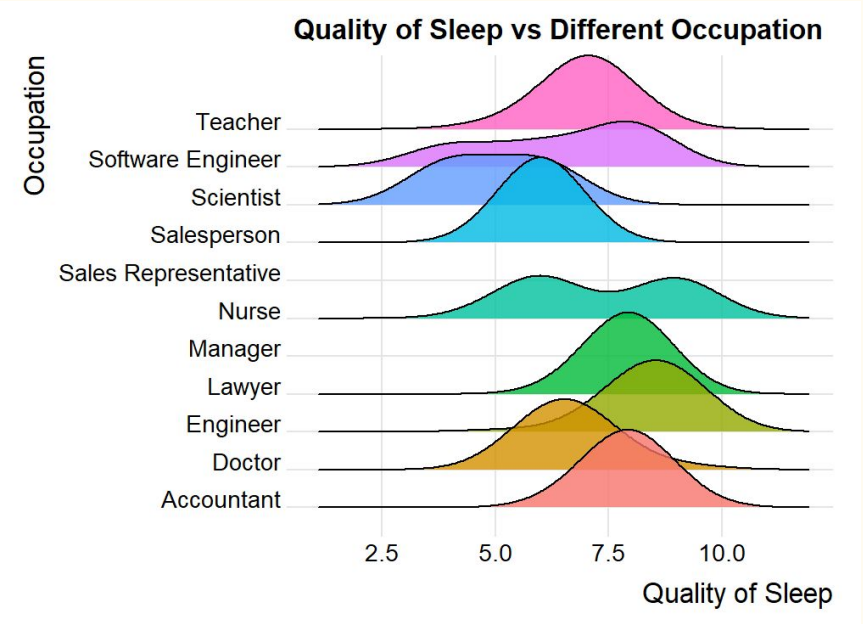


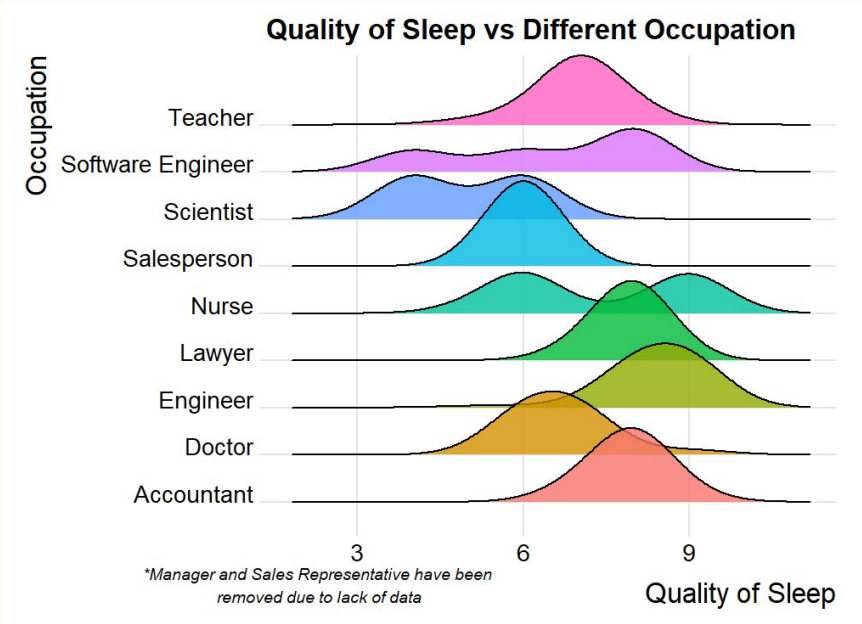
Does Alcohol Affect Sleep?

- Sleep efficiency decreases as alcohol consumption increases
    - 0.86 at 0 oz alcohol consumption
    - 0.65 at 2 oz alcohol consumption
    - 0.68 at 5 oz alcohol consumption



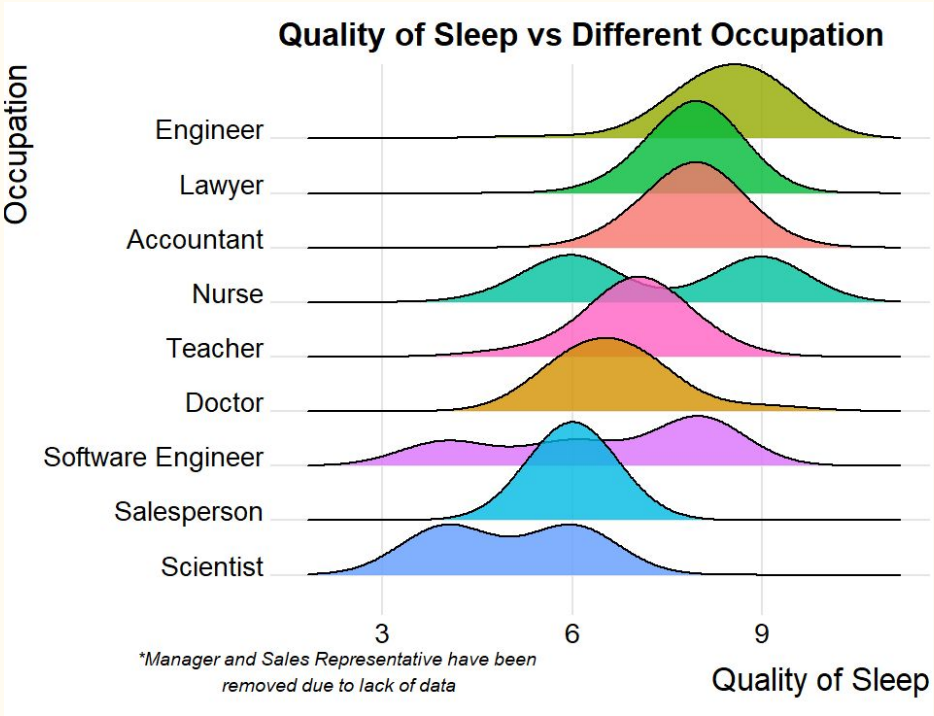
Does Occupation Affect Sleep? (preprocessing)

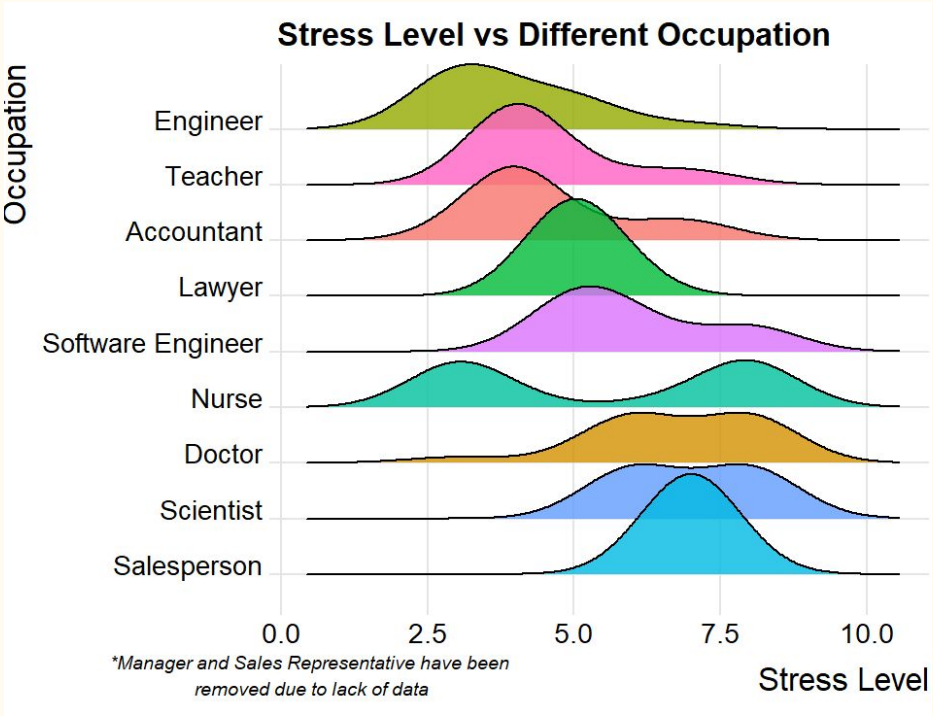





```
# omit manager and sales Representative because theres not enough data
job_compare_qs <- job_compare_qs %>% filter(Occupation != "Manager" & Occupation != "Sales Representative")
```

Does Occupation Affect Sleep?






```
# used fct_reorder to reorder the rows by median in the graph
job_compare_qs %>% mutate(class = fct_reorder(Occupation, Quality.of.Sleep, .fun = 'median')) %>%
  ggplot(aes(x = Quality.of.Sleep, y = reorder(Occupation, Quality.of.Sleep), fill = Occupation)) +
```

```
.desc = TRUE
```

Health Related Findings

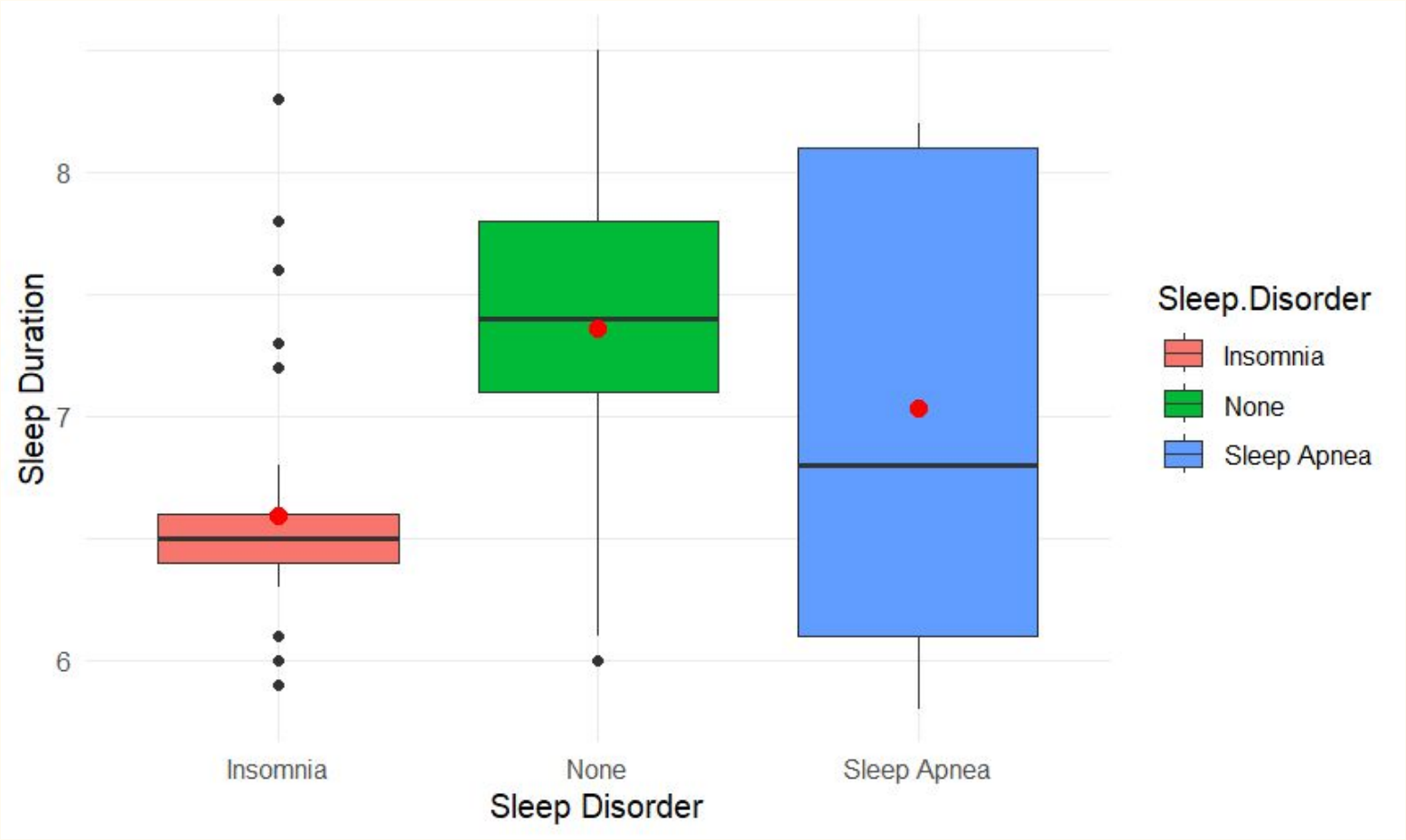
How do Sleep Disorders affect Sleep Duration

Sleep Disorders present in the data:

- Insomnia - difficulty to fall asleep and/or difficulty to stay awake
- Normal - no sleep disorder
- Sleep Apnea - when a person pauses in breathing during sleep
- Code:

```
# Graph to find correlation between sleep disorder and sleep duration-----------
sleep <- lifestyle %>% select(c("Sleep.Disorder","Sleep.Duration"))
ggplot(sleep, aes(x = Sleep.Disorder, y = Sleep.Duration)) +
  geom_boxplot(aes(group = Sleep.Disorder, fill = Sleep.Disorder)) +
  labs(x = "Sleep Disorder", y = "Sleep Duration") +
  stat_summary(fun.y = mean, geom = "point", shape = 20, size = 4, color = "red",
               fill = "red") +
  theme_minimal()
```

How do Sleep Disorders affect Sleep Duration



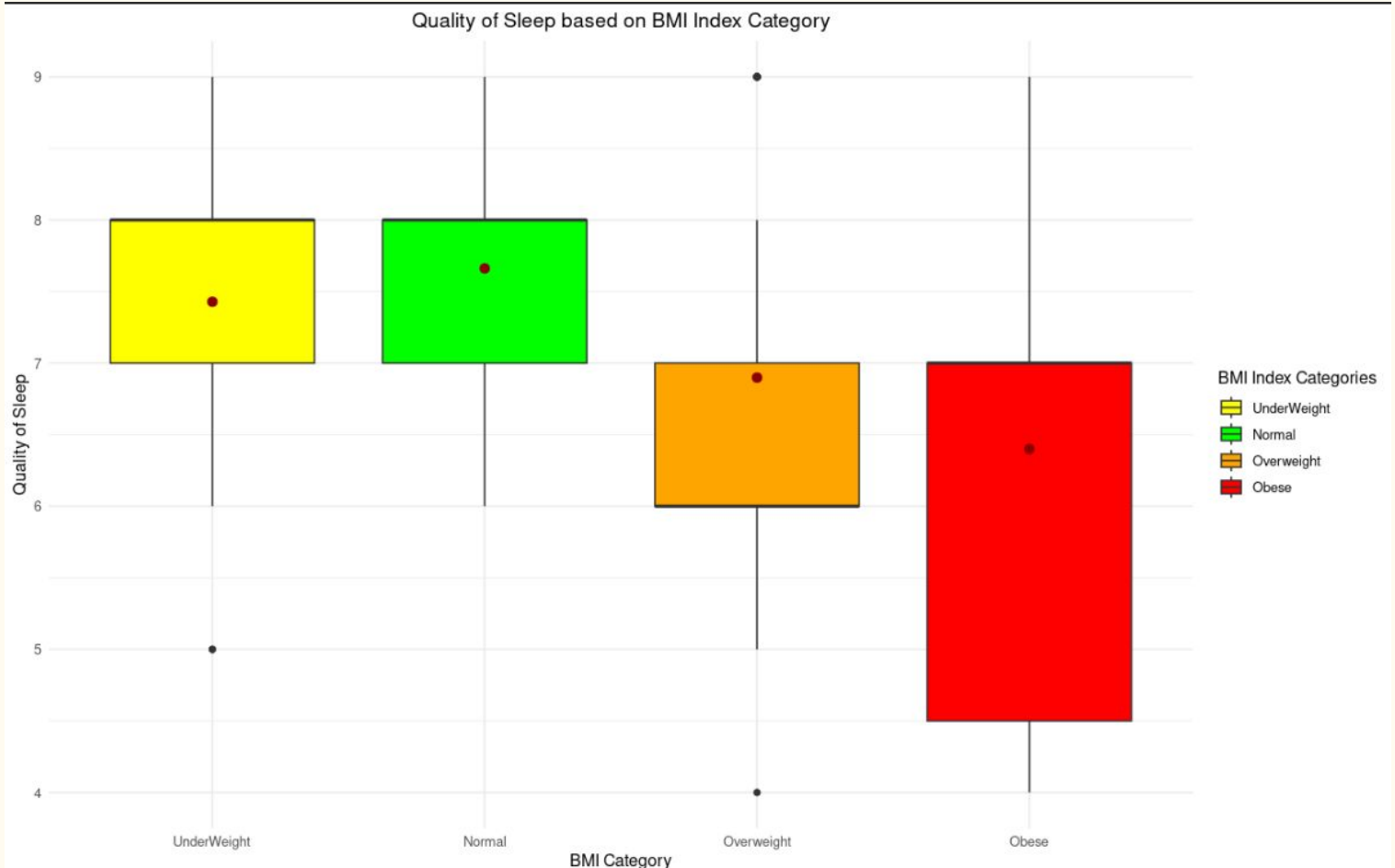
Does BMI affect Sleep Quality?

Ranges of BMIs:

- Underweight: < 18.5 (less than 18.5)
- Normal: 18.6 - 24.9
- Overweight: 25 - 29.9
- Obese: 30 < (more than 30)
- Code:

```
# Graph to find correlation between BMI and sleep quality----------------------
sleep <- lifestyle %>% select(c("BMI.Category","Quality.of.Sleep"))
sleep$BMI.Category <- factor(sleep$BMI.Category,
                             levels=c("Normal Weight", "Normal", "Overweight", "Obese"))
levels(sleep$BMI.Category) <- c("UnderWeight", "Normal", "Overweight", "Obese")
ggplot(sleep, aes(x = BMI.Category, y = Quality.of.Sleep, fill = interaction(BMI.Category))) +
  geom_boxplot() +
  scale_fill_manual(values = c("yellow","green","orange","red"))+
  labs(x = "BMI Category", y = "Quality of Sleep") +
  stat_summary(fun.y = mean, geom = "point", shape = 20, size = 4, color = "red4",
               fill = "black") +
  theme_minimal() +
  ggtitle("Quality of Sleep based on BMI Index Category") +
  theme(plot.title = element_text(hjust = 0.5)) +
  guides(fill = guide_legend(title="BMI Index Categories"))
```

Does BMI affect Sleep Quality?



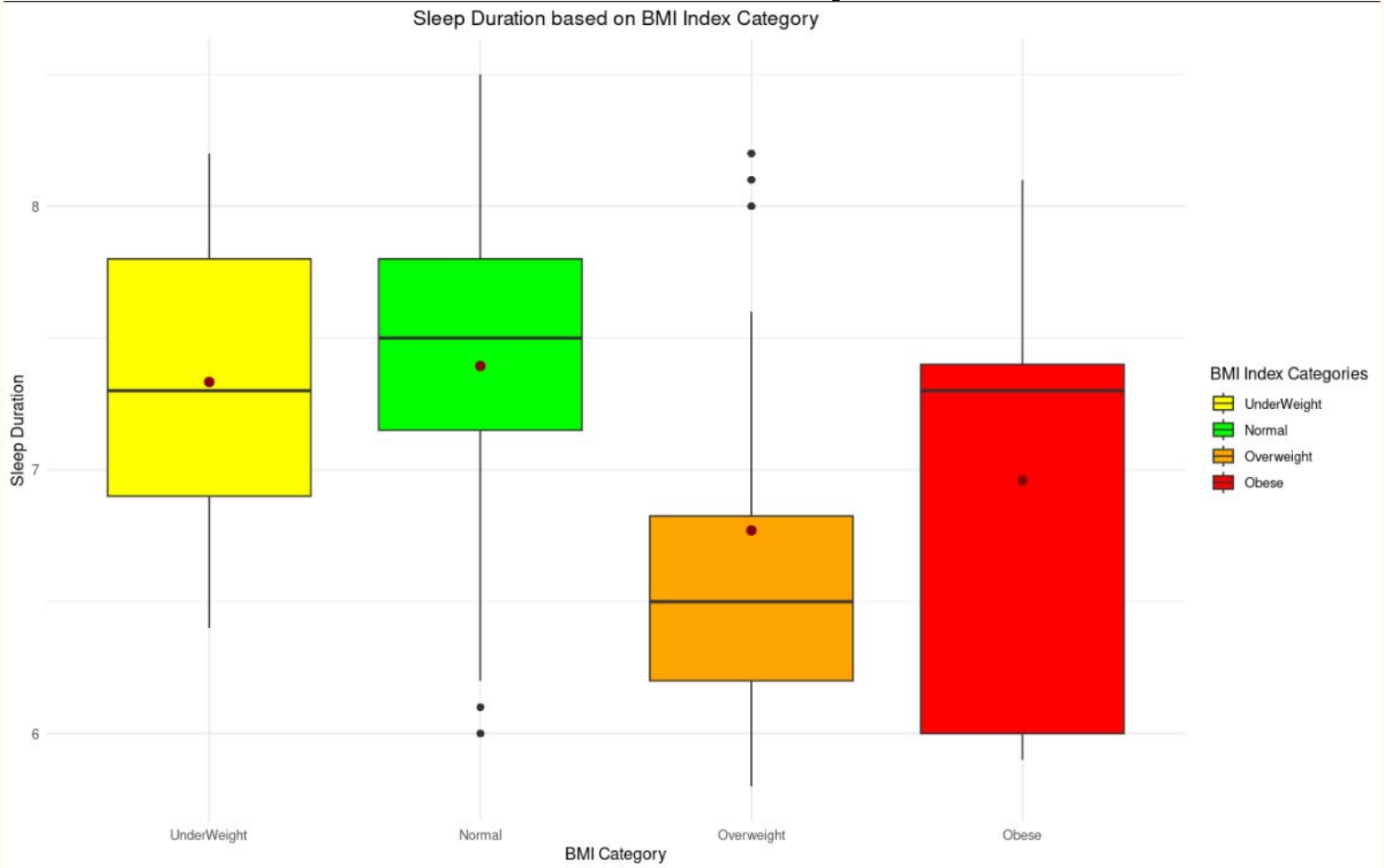
Does BMI affect Sleep Duration?

Ranges of BMIs:

- Underweight: < 18.5 (less than 18.5)
- Normal: 18.6 - 24.9
- Overweight: 25 - 29.9
- Obese: 30 < (more than 30)
- Code:

```
# Graph to find correlation between BMI and sleep duration----------------------
sleep <- lifestyle %>% select(c("BMI.Category","Sleep.Duration"))
sleep$BMI.Category <- factor(sleep$BMI.Category ,
                      levels = c("Normal Weight", "Normal", "Overweight", "Obese"))
levels(sleep$BMI.Category) <- c("UnderWeight", "Normal", "Overweight", "Obese")
ggplot(sleep, aes(x = BMI.Category, y = Sleep.Duration, fill = interaction(BMI.Category))) +
  geom_boxplot() +
  scale_fill_manual(values = c("yellow","green","orange","red")) +
  labs(x = "BMI Category", y = "Sleep Duration") +
  stat_summary(fun.y = mean, geom = "point", shape = 20, size = 4, col = "red4",
               fill = "black") +
  theme_minimal() +
  ggtitle("Sleep Duration based on BMI Index Category") +
  theme(plot.title = element_text(hjust = 0.5)) +
  guides(fill = guide_legend(title="BMI Index Categories"))
```

Does BMI affect Sleep Duration?



Concluding Thoughts

Conclusion

- Back to our initial question: What are the factors that contribute to the best quality and duration of sleep?
  - BMI and overall good health plays a key role in having better sleep quality and being able to sleep the right amount
  - Drinking **little to no alcohol** also seems to have a positive effect on the amount of sleep as well as the quality of that sleep
  - Get a less stressful job like engineer, accountant
  - **AVOID BEING A SOFTWARE ENGINEER** (or scientist)

- What did **NOT** have an effect on better sleep/longer sleep duration?
  - Age doesn't affect a person's sleep cycle. Sleep cycle can vary from person to person and from night to night based on a wide range of factors.
  - Surprisingly drinking caffeine before sleeping did not have much of an impact on the efficiency of sleep

Thank You!