Is There a Correlation Between Less Than 8 Hours of Sleep and Cognitive Function?

Ethan Choi - echoi090

2025-03-10

Introduction

This project will aim to investigate whether sleeping fewer than eight hours per night has a significant impact on cognitive function. Using an open source dataset provided by Kaggle, this analysis examines cognitive performance metrics such as reaction time and working memory accuracy to determine if a relationship exists.

Load the Dataset

```
# Load the dataset
data <- read.csv("sleep_deprivation_dataset_detailed.csv")
head(data)
```

```
##   Participant_ID Sleep_Hours Sleep_Quality_Score Daytime_Sleepiness
## 1           P1         5.25                  15                 12
## 2           P2         8.70                  12                 14
## 3           P3         7.39                  17                 10
## 4           P4         6.59                  14                  3
## 5           P5         3.94                  20                 12
## 6           P6         3.94                  12                  6
##   Stroop_Task_Reaction_Time N_Back_Accuracy Emotion_Regulation_Score
## 1                      1.60           64.20                       12
## 2                      2.54           65.27                       21
## 3                      3.40           74.28                       35
## 4                      3.54           72.42                       25
## 5                      3.09           99.72                       60
## 6                      2.84           58.80                       46
##   PVT_Reaction_Time Age Gender   BMI Caffeine_Intake Physical_Activity_Level
## 1            365.85  35 Female 30.53               2                       1
## 2            288.95  20   Male 27.28               3                       8
## 3            325.93  18   Male 30.00               1                       2
## 4            276.86  18   Male 34.47               5                       0
## 5            383.45  36   Male 29.70               3                       4
## 6            224.48  28   Male 32.23               3                       6
##   Stress_Level
## 1           33
## 2           37
## 3           32
```

```
## 4            23
## 5            14
## 6            29
```

```r
str(data)
```

```
## 'data.frame':    60 obs. of  14 variables:
##  $ Participant_ID          : chr  "P1" "P2" "P3" "P4" ...
##  $ Sleep_Hours             : num  5.25 8.7 7.39 6.59 3.94 3.94 3.35 8.2 6.61 7.25 ...
##  $ Sleep_Quality_Score     : int  15 12 17 14 20 12 8 14 12 0 ...
##  $ Daytime_Sleepiness      : int  12 14 10 3 12 6 18 21 1 9 ...
##  $ Stroop_Task_Reaction_Time: num  1.6 2.54 3.4 3.54 3.09 2.84 3.16 3.28 1.74 2.61 ...
##  $ N_Back_Accuracy         : num  64.2 65.3 74.3 72.4 99.7 ...
##  $ Emotion_Regulation_Score : int  12 21 35 25 60 46 31 66 38 23 ...
##  $ PVT_Reaction_Time       : num  366 289 326 277 383 ...
##  $ Age                     : int  35 20 18 18 36 28 22 29 20 18 ...
##  $ Gender                  : chr  "Female" "Male" "Male" "Male" ...
##  $ BMI                     : num  30.5 27.3 30 34.5 29.7 ...
##  $ Caffeine_Intake         : int  2 3 1 5 3 3 4 1 3 3 ...
##  $ Physical_Activity_Level : int  1 8 2 0 4 6 5 0 4 4 ...
##  $ Stress_Level            : int  33 37 32 23 14 29 16 4 28 3 ...
```

Data Cleaning, Categorize Data by Less than or More than 8 hours

```r
# Remove missing values
data <- na.omit(data)

# Categorize sleep hours
data$SleepCategory <- ifelse(data$Sleep_Hours < 8, "Less than 8 hours", "8 or more hours")
```

Variable Definitions

- **Stroop Task Reaction Time:** This measures cognitive processing speed. A longer reaction time indicates slower cognitive functions.
- **N-Back Accuracy:** This assesses working memory performance. Higher accuracy indicates better cognitive functioning.
- **Sleep Hours:** Total hours of sleep per night.

Summary Statistics

```r
# Get summary statistics for indicated variables
summary(data$Sleep_Hours)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.120   4.107   5.690   5.806   7.285   8.820
```

```r
summary(data$Stroop_Task_Reaction_Time)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.600   2.595   3.260   3.245   3.965   4.490
```

```r
summary(data$N_Back_Accuracy)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   50.90   64.62   74.27   75.01   85.60   99.73
```

Correlation Analysis

```r
# Compute correlation between sleep hours and cognitive metrics
cor_test_reaction <- cor(data$Sleep_Hours, data$Stroop_Task_Reaction_Time)
cor_test_accuracy <- cor(data$Sleep_Hours, data$N_Back_Accuracy)
print(paste("Correlation between Sleep Hours and Reaction Time:", round(cor_test_reaction, 3)))
```

```
## [1] "Correlation between Sleep Hours and Reaction Time: 0.056"
```

```r
print(paste("Correlation between Sleep Hours and N-Back Accuracy:", round(cor_test_accuracy, 3)))
```

```
## [1] "Correlation between Sleep Hours and N-Back Accuracy: -0.11"
```
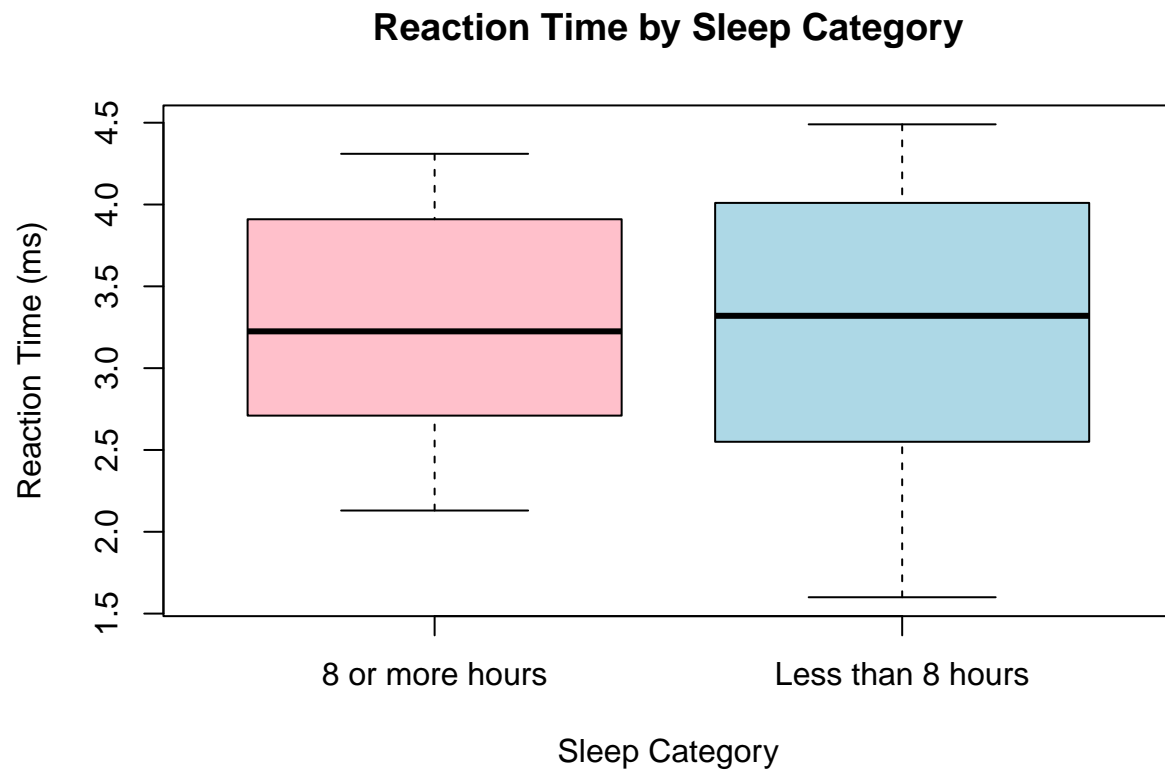
Correlation Statistic Interpretation:

A correlation of 0.056 between Sleep Hours and Reaction Time suggests a weak positive relationship, meaning sleep duration has some impact on reaction speed. A correlation of -0.11 between Sleep Hours and N-Back Accuracy indicates a weak negative relationship, suggesting that those who sleep less might have slightly better accuracy (this isn't logical, so we are going to do further analysis), but the effect is small and can't be guaranteed.

Visualization

Boxplot for Sleep Category vs Stroop Task Reaction Time

```r
boxplot(Stroop_Task_Reaction_Time ~ SleepCategory, data=data,
        main="Reaction Time by Sleep Category",
        xlab="Sleep Category", ylab="Reaction Time (ms)",
        col=c("pink", "lightblue"))
```
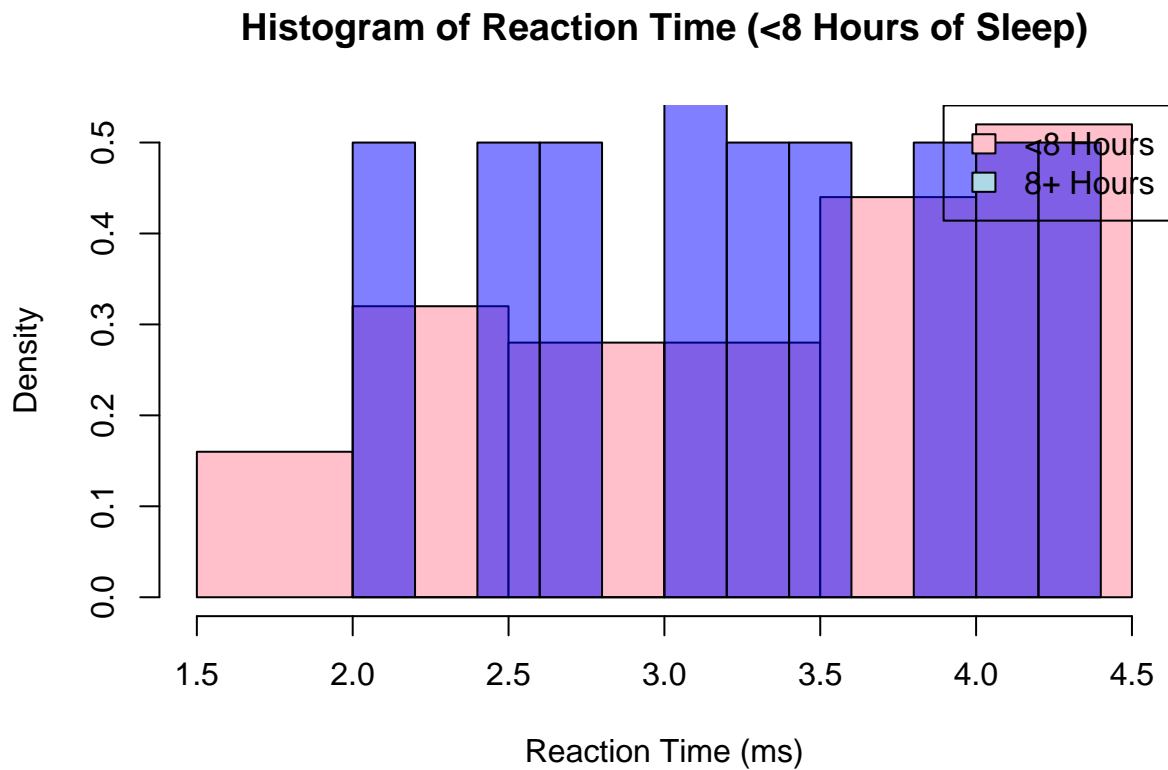
Reaction Time by Sleep Category

Boxplot Interpretation:

The "Less than 8 hours" boxplot indicates slightly longer reaction times, but not by too much, just as predicted by our correlation statistic.

Histogram for Reaction Time

```r
hist(data$Stroop_Task_Reaction_Time[data$Sleep_Hours < 8],
     main="Histogram of Reaction Time (<8 Hours of Sleep)",
     xlab="Reaction Time (ms)", col="pink", breaks=10, probability=TRUE)
hist(data$Stroop_Task_Reaction_Time[data$Sleep_Hours >= 8],
     col=rgb(0,0,1,0.5), breaks=10, probability=TRUE, add=TRUE)
legend("topright", legend=c("<8 Hours", "8+ Hours"), fill=c("pink", "lightblue"))
```

Histogram of Reaction Time (<8 Hours of Sleep)

Histogram Interpretation:

This histogram overlaps both sleep categories, giving more insight on the relationship between sleep hours and reaction time. Again, there is a slight but noticeable difference.

Boxplot for Sleep Category vs N-Back Accuracy

```
boxplot(N_Back_Accuracy ~ SleepCategory, data=data,
        main="N-Back Accuracy by Sleep Category",
        xlab="Sleep Category", ylab="N-Back Accuracy (%)",
        col=c("lightcyan", "lavender"))
```

N–Back Accuracy by Sleep Category
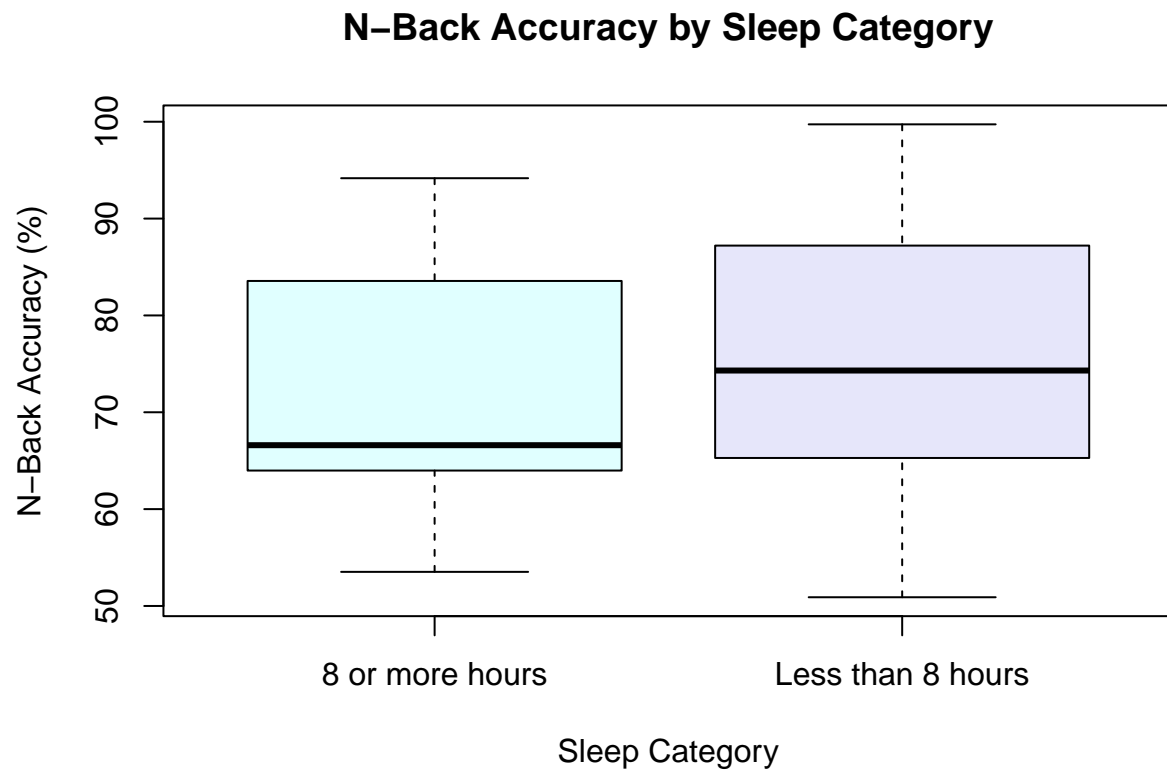
Boxplot Interpretation:

Higher N-Back accuracy scores indicate better cognitive performance. A higher median in the "Less than 8 hours" category does not support our hypothesis that insufficient sleep impairs cognitive function. In fact, it disproves it and isn't logical either.

Density plot for N-Back Accuracy

```
range_nback <- range(data$N_Back_Accuracy)
range_density <- range(density(data$N_Back_Accuracy[data$Sleep_Hours < 8])$y,
                       density(data$N_Back_Accuracy[data$Sleep_Hours >= 8])$y)
plot(density(data$N_Back_Accuracy[data$Sleep_Hours < 8]),
     main="Density Plot of N-Back Accuracy", xlab="N-Back Accuracy (%)", col="pink", lwd=2, xlim=range_
lines(density(data$N_Back_Accuracy[data$Sleep_Hours >= 8]), col="lightblue", lwd=2)
legend("topright", legend=c("<8 Hours", "8+ Hours"), col=c("pink", "lightblue"), lwd=2)
```

Density Plot of N−Back Accuracy
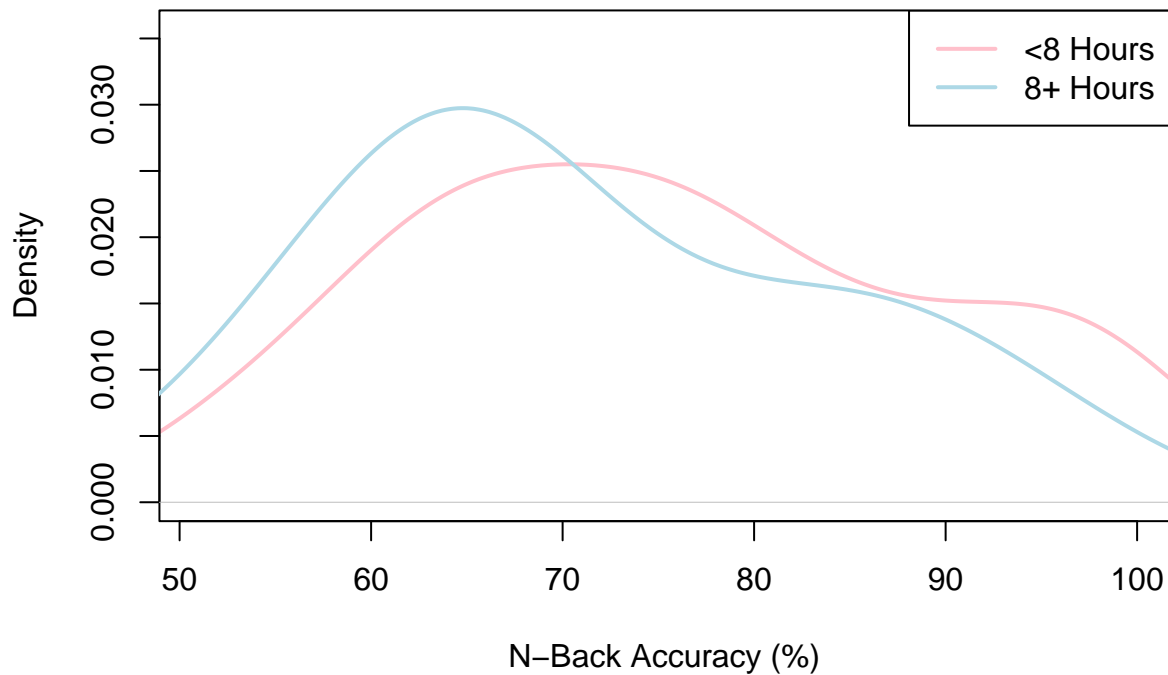
Density Graph Interpretation:

This density plot better shows a distribution of cognitive accuracy scores. If the red curve (less than 8 hours) or blue curve (more than 8 hours) skews lower, it suggests worse cognitive performance for sleep-deprived individuals and well-rested indivuduals, respectively. Since both seem to have similar skewness with variability, a definite relationship cannot be determined. This means that other factors may be at play.

Additional Analysis: Stress and Lifestyle Factors

```
# Summary statistics for stress and physical activity
summary(data$Stress_Level)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    8.75   17.50   17.87   26.25   40.00
```

```
summary(data$Physical_Activity_Level)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   4.000   4.117   6.000  10.000
```

```
# Compute correlation between stress/activity levels and cognitive metrics
cor_test_stress <- cor(data$Stress_Level, data$N_Back_Accuracy)
cor_test_activity <- cor(data$Physical_Activity_Level, data$N_Back_Accuracy)
print(paste("Correlation between Stress Level and N-Back Accuracy:", round(cor_test_stress, 3)))
```

```
## [1] "Correlation between Stress Level and N-Back Accuracy: -0.123"
```

```
print(paste("Correlation between Physical Activity Level and N-Back Accuracy:", round(cor_test_activity
```

```
## [1] "Correlation between Physical Activity Level and N-Back Accuracy: -0.178"
```

Correlation Interpretation:

A correlation of -0.123 between Stress Level and N-Back Accuracy suggests that higher stress is weakly associated with lower cognitive accuracy. A correlation of -0.178 between Physical Activity Level and N-Back Accuracy suggests that individuals with higher activity levels might perform slightly better in cognitive tasks
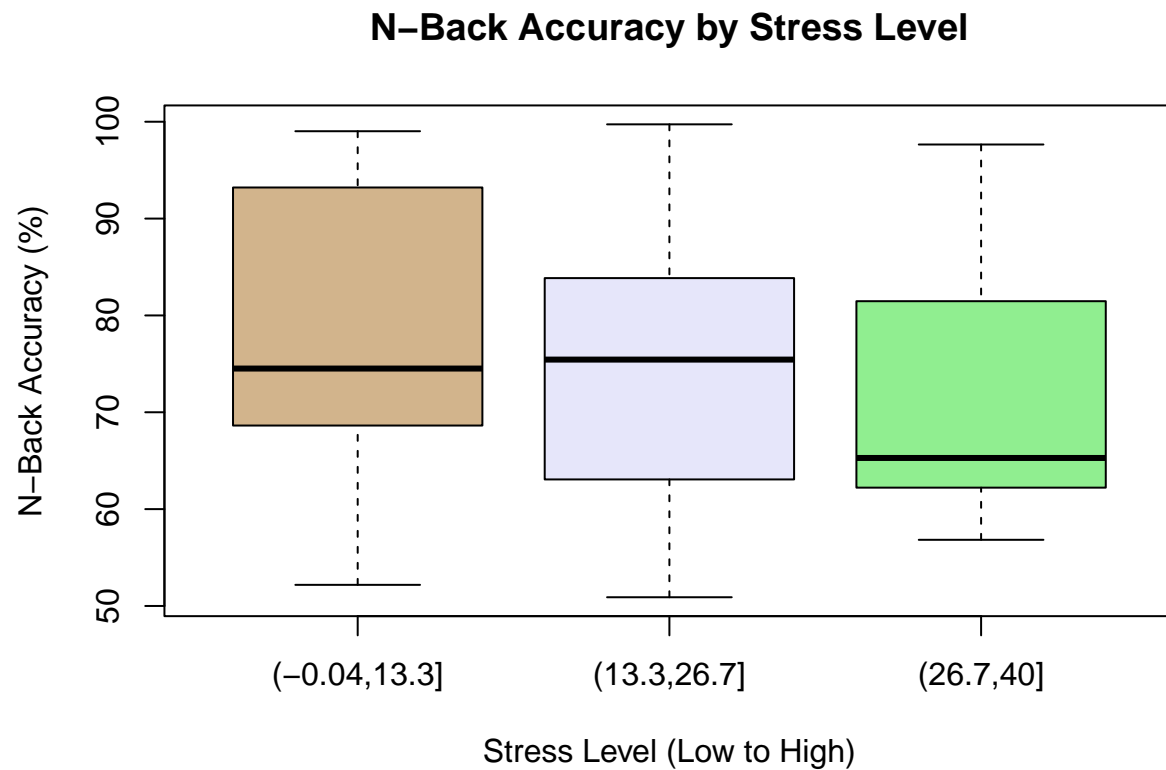
Visualization of Stress and Activity Levels

Boxplot for Stress Level vs N-Back Accuracy

```
boxplot(N_Back_Accuracy ~ cut(data$Stress_Level, breaks=3), data=data,
        main="N-Back Accuracy by Stress Level",
        xlab="Stress Level (Low to High)", ylab="N-Back Accuracy (%)",
        col=c("tan", "lavender", "lightgreen"))
```

N−Back Accuracy by Stress Level

Boxplot Interpretation:

Higher stress levels indicate poor cognitive performance.

Boxplot for Physical Activity Level vs N-Back Accuracy and Stroop Task Reaction Time

```r
boxplot(N_Back_Accuracy ~ cut(data$Physical_Activity_Level, breaks=3), data=data,
        main="N-Back Accuracy by Physical Activity Level",
        xlab="Physical Activity Level (Low to High)", ylab="N-Back Accuracy (%)",
        col=c("gray", "white", "navy"))
```

N−Back Accuracy by Physical Activity Level

N−Back Accuracy (%)

Physical Activity Level (Low to High)

(−0.01,3.33] (3.33,6.67] (6.67,10]
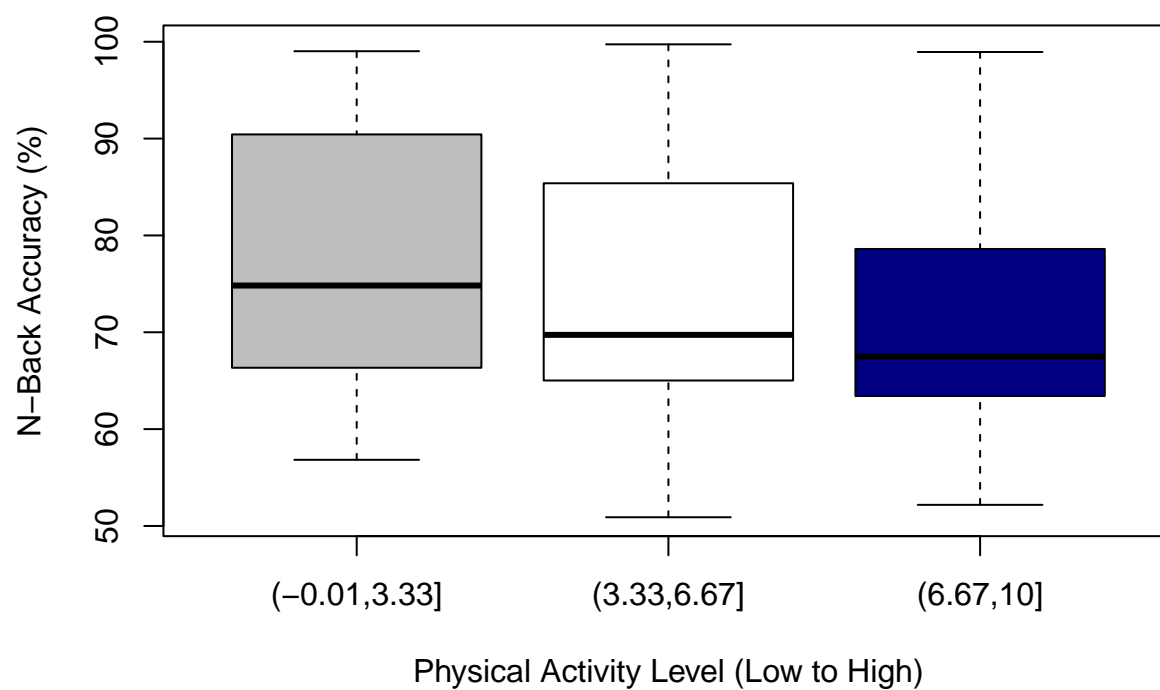
```r
boxplot(Stroop_Task_Reaction_Time ~ cut(data$Physical_Activity_Level, breaks=3), data=data,
        main="Reaction Time by Physical Activity Level",
        xlab="Physical Activity Level (Low to High)", ylab="Reaction Time (ms)",
        col=c("gray", "white", "navy"))
```

Reaction Time by Physical Activity Level
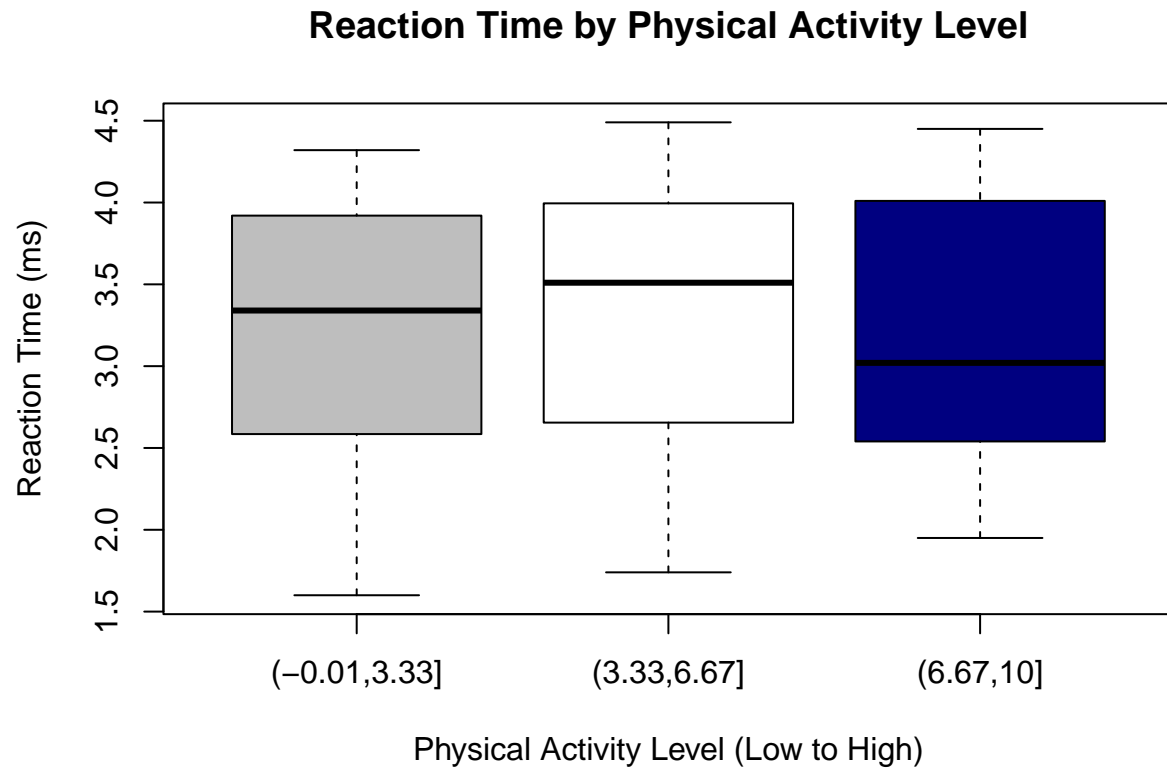
Boxplots Interpretation:

Higher physical activity levels don't contribute to a worsened cognitive performance, but they significantly improve reaction times.

Summary

Reaction Time: The boxplot suggests that individuals sleeping fewer than 8 hours tend to have longer reaction times, implying slower cognitive processing.

Cognitive Accuracy: The N-Back Accuracy boxplot does not explicitly show that individuals with less than 8 hours of sleep tend to perform worse on memory recall tasks. In addition, our density graoh is varied, showing that sleep can or cannot affect N-Back Accuracy. This means that other factors must be at play for cognitive accuracy, rather than just sleep hours

Stress and Cognitive Performance: Higher stress levels indicate lower N-Back Accuracy scores, suggesting that stress negatively affects cognitive performance and memory.

Physical Activity and Cognitive Performance: Higher activity levels are associated with better reaction times, implying that exercise contributes positively to cognitive processing speeds.

Conclusion

This analysis concludes that while sleep duration influences cognitive performance, other factors such as stress levels and physical activity also contribute significantly. Future research should consider these additional lifestyle factors when evaluating cognitive function.

Sources   https://www.nceas.ucsb.edu/sites/default/files/2020-04/colorPaletteCheatsheet.pdf

https://sqlpad.io/tutorial/boxplots-r-comprehensive-guide/

https://www.datacamp.com/doc/r/histograms-and-density

https://dataanalytics.org.uk/plot-two-overlapping-histograms-on-one-chart-in-r/ "'