ANALYSIS OF SLEEP & FATIGUE LEVELS

BACKGROUND:

Hello, I'm Johnny Nguyen, a Computer Science major at Rutgers University. As we all know, college students lead hectic lives, constantly juggling assignments, projects, and exam preparations, leaving us with little time for sleep. This has led me to experience fatigue and grumpiness upon waking up, resulting in a dependence on coffee to function normally like most college students. However, I've observed that some people can function well on just 3-6 hours of sleep, while others who sleep for 10+ hours still feel tired. This has made me curious about the factors that influence fatigue levels. Does the duration of sleep impact fatigue levels? Can activities such as using the phone before bedtime or eating breakfast affect our energy levels?

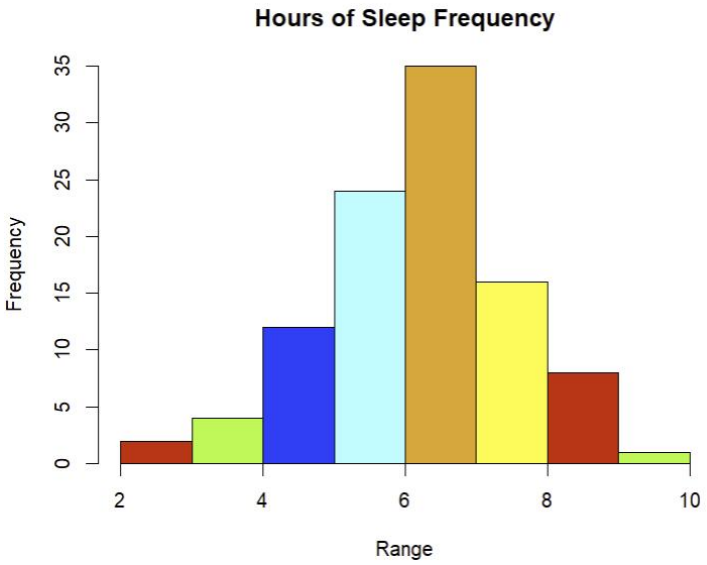
DATASET: Sleep Study | Kaggle

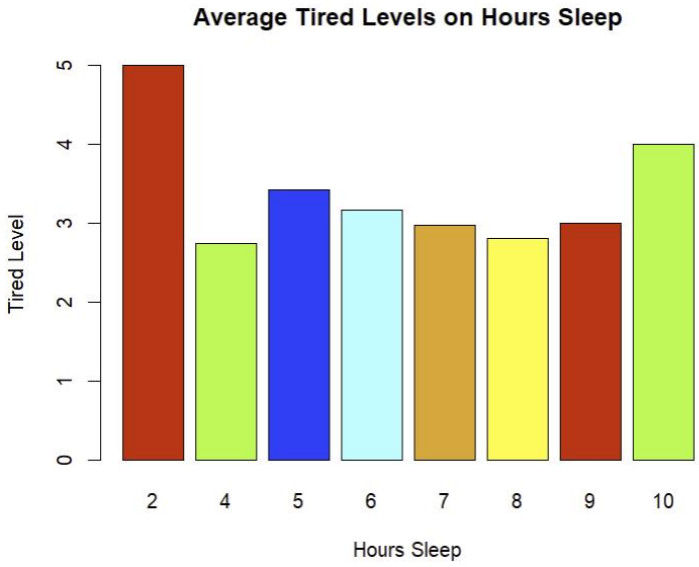
After extensive research on datasets related to sleep and fatigue levels, I came across a sleep study dataset on Kaggle that appears to meet my criteria. Its credibility cannot be ascertained as the author did not disclose the data source. Additionally, the sample size is relatively small, comprising just over 100 participants within the U.S. Despite the limitations of the dataset, I believe there may still be hidden patterns or relationships within the data that are worth exploring. It's also important to note that due to the uncertainties surrounding the dataset's credibility and the possibility of hidden variables, any results obtained should be interpreted with caution and not presented as definitive facts.

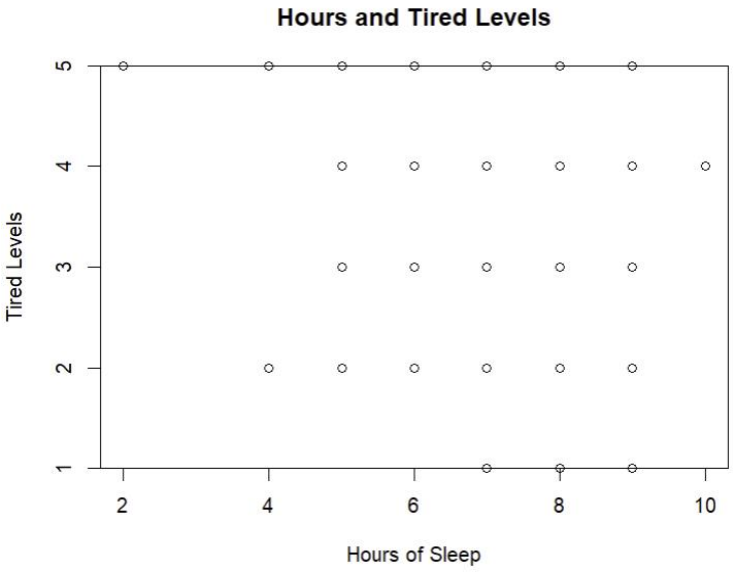
Findings

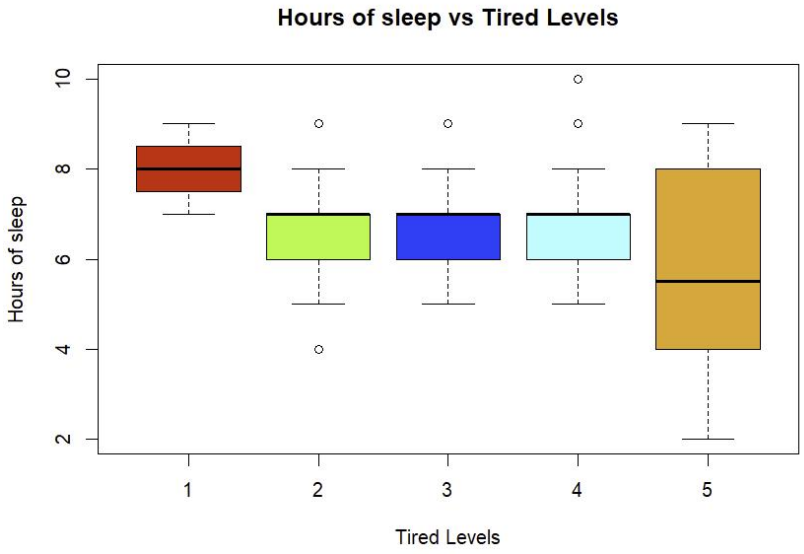
(1 being not tired, 5 being very tired)
(Hours ranges from 2-10)

Hours vs Tired Levels

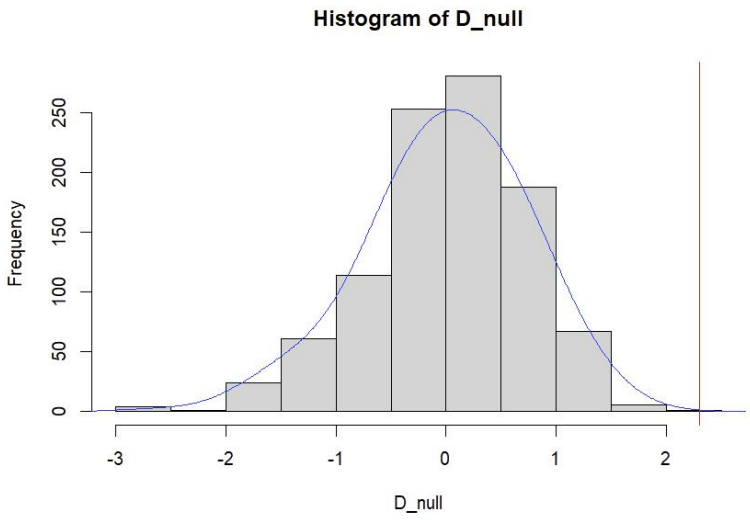


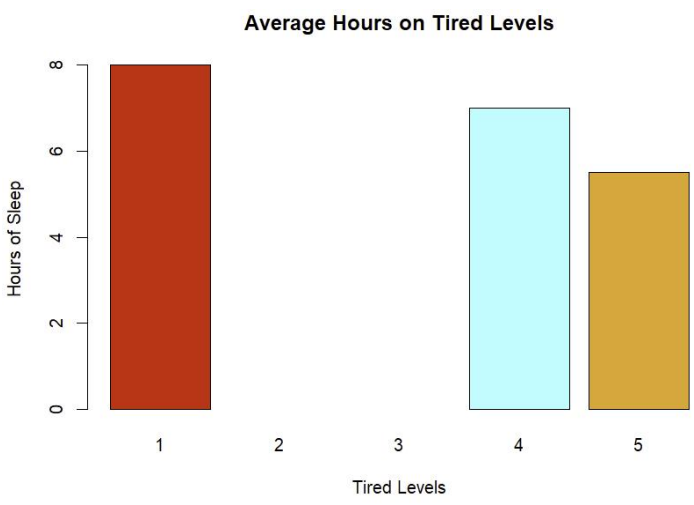






The plots indicate that fatigue levels are at their lowest when individuals get 7-9 hours of sleep. Some individuals may experience fatigue despite sleeping for 7-9 hours, potentially due to hidden variables. Durations such as 2, 4, and 10 hours have very small sample sizes, which makes it difficult to draw reliable conclusions from them. To test this observation, we will conduct a permutation test comparing the length of sleep and fatigue levels.



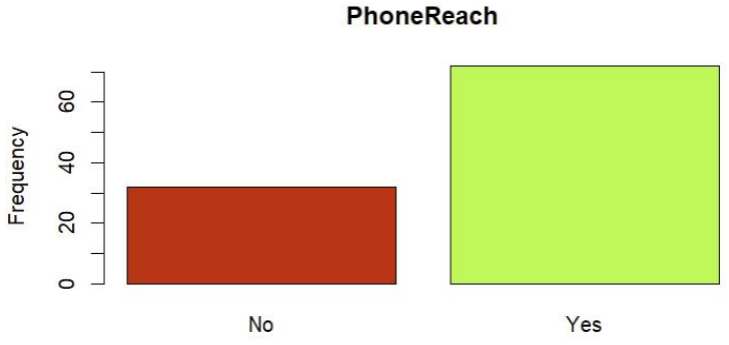


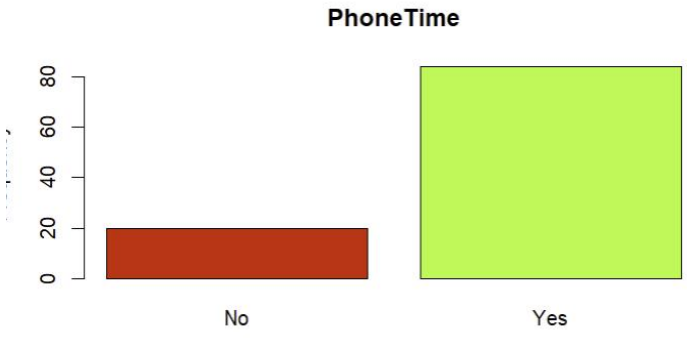

```
> PermutationTestSecond::Permutation(sleep, "Tired", "Hours", 1000, "5", "1")
[1] 0
```

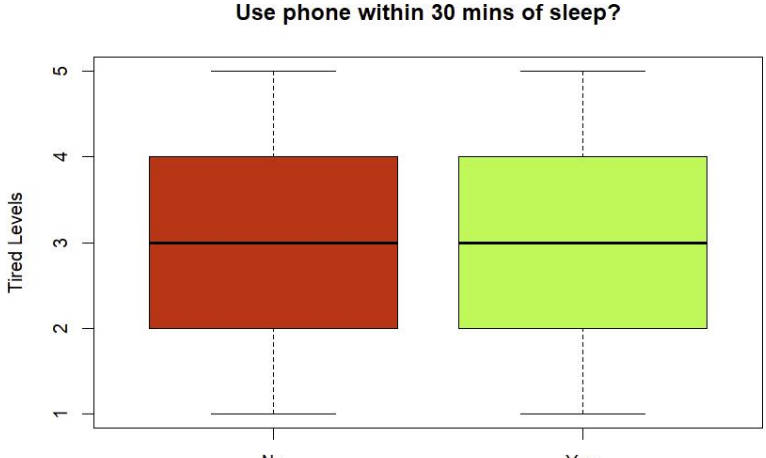
Null hypothesis: Mean sleep hours of people that are not tired is the same as those that are tired.
Alternative hypothesis: Mean sleep hours of people that are not tired is higher than mean sleep of tired people.

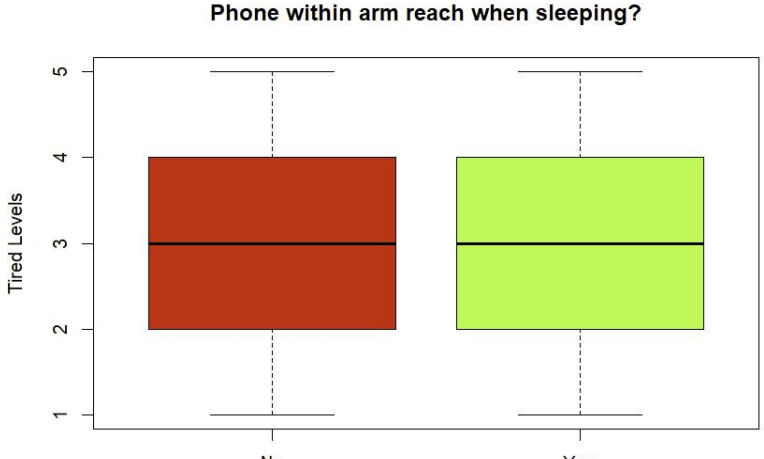
The p-value of the permutation test is 0, indicating that we should reject null hypothesis. Therefore, we can conclude that feeling less tired is associated with longer sleep periods.

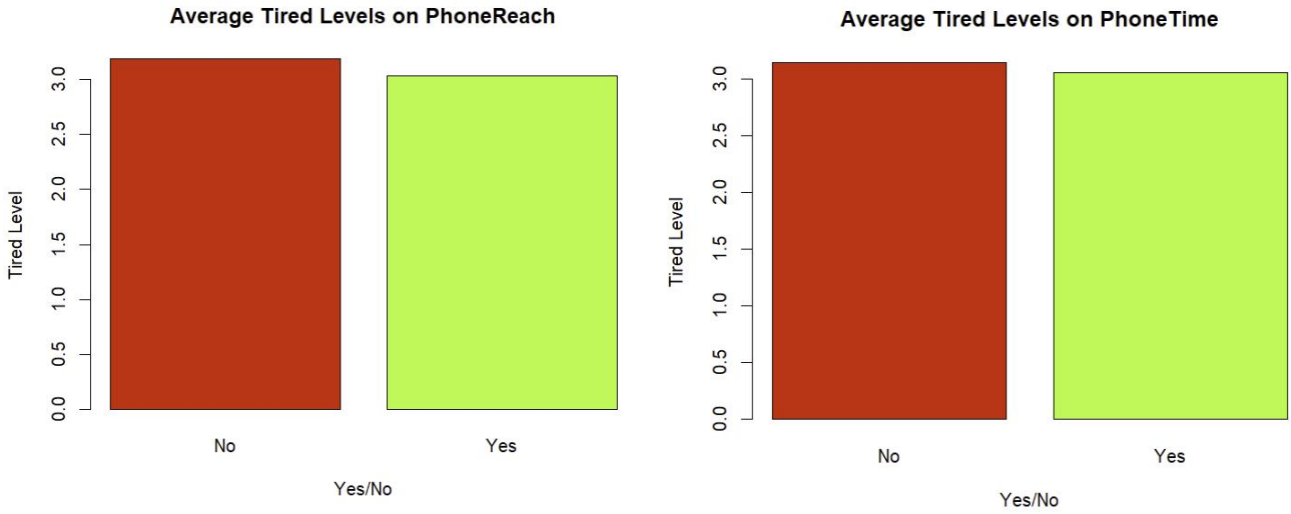
Phone usage vs Tired Levels



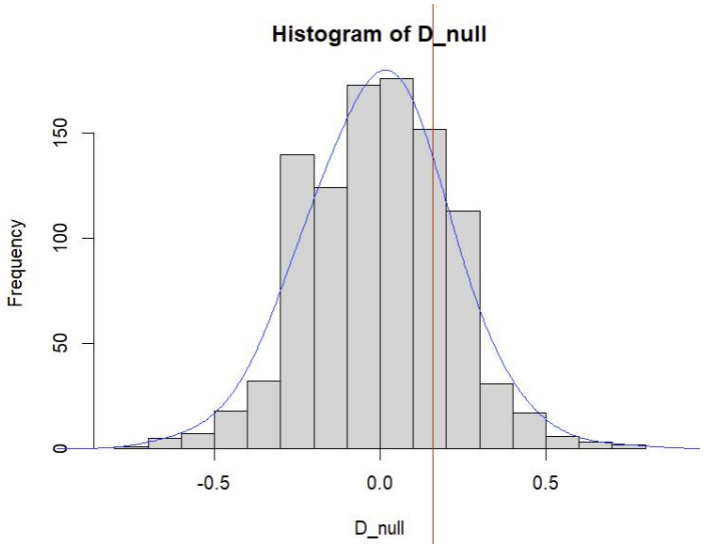








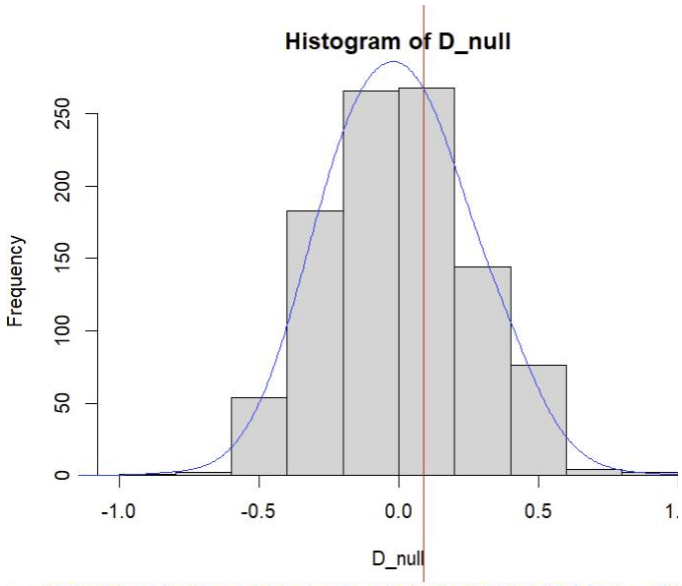
Even though most people use their phones after sleep, there seems to be no correlation between sleep fatigue levels and phone usage. It's possible that the dataset may not have sufficient data to reveal such a relationship. Let's test it.




```
> PermutationTestSecond::Permutation(sleep, "PhoneReach", "Tired", 1000, "No", "Yes")
[1] 0.172
> |
```

Null hypothesis: Mean tired levels of people who sleep with their phone next to them is the same as those that don't

Alternative hypothesis: Mean tired levels of people that sleep with their phone next to them are higher than mean tired levels of people that don't sleep with their phone next to them.

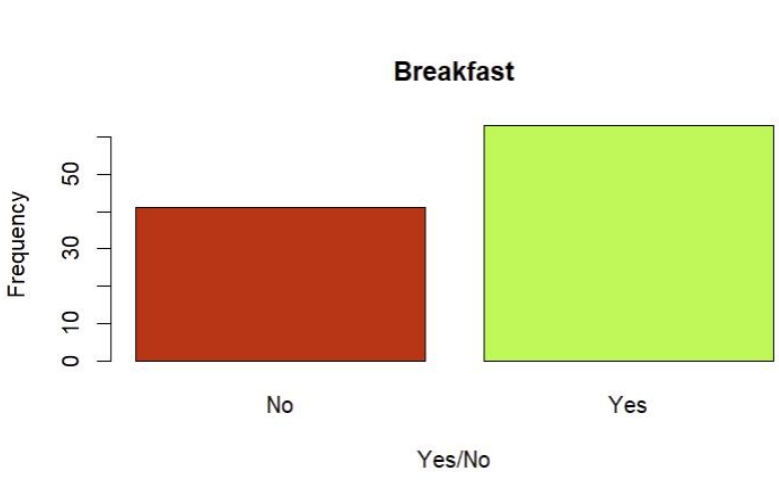


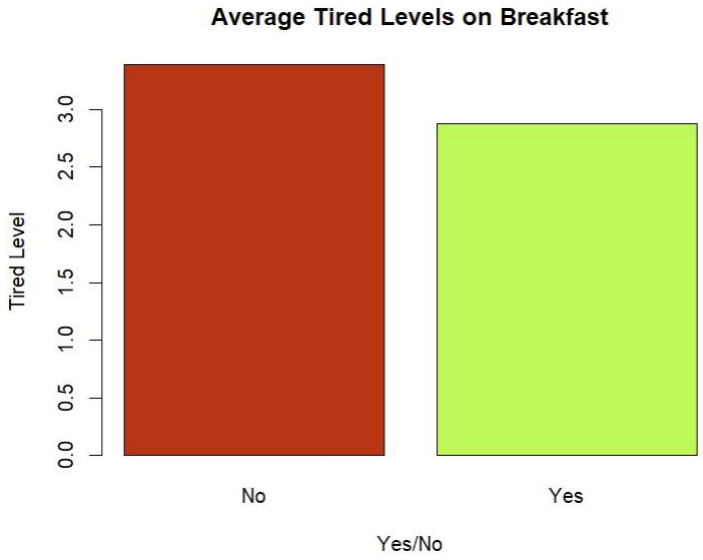

```
> PermutationTestSecond::Permutation(sleep, "PhoneTime", "Tired", 1000, "No", "Yes")
[1] 0.387
```

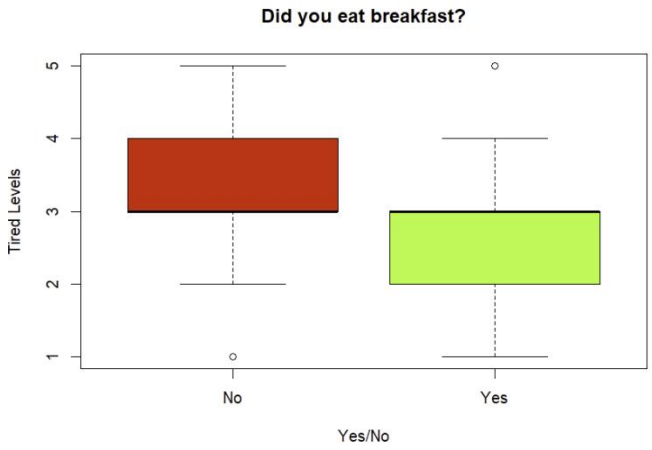
Null hypothesis: Mean tired levels of people who use their phone 30 minutes before bed is the same as those that don't.
Alternative hypothesis: Mean tired levels of people who use their phone 30 minutes before bed is higher than mean tired levels of people that don't use their phone within 30 minutes.

In both phone usages, the p-value is greater than 5 percent. Therefore, we do not reject null hypothesis - there is no difference in fatigue level whether we use our phones or not.

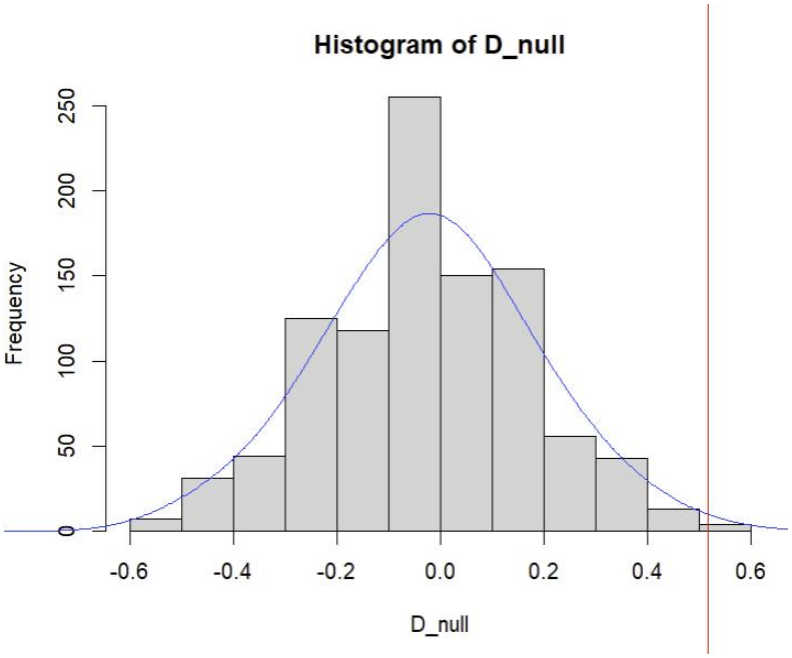
Breakfast vs Tired Levels







As you can see, individuals who eat breakfast appear to experience less fatigue than those who skip breakfast.

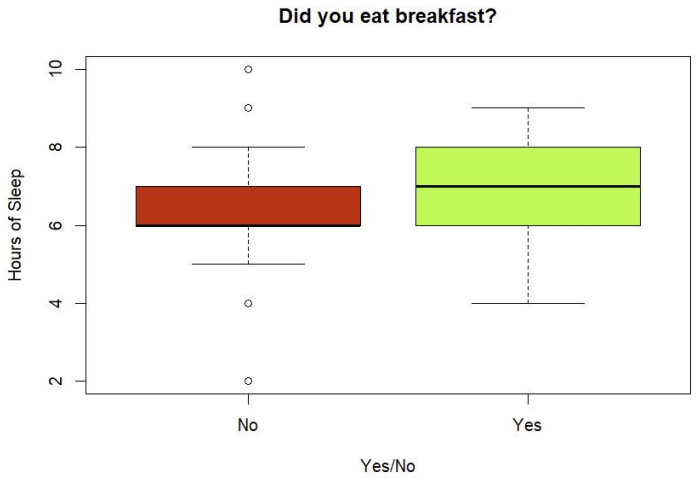


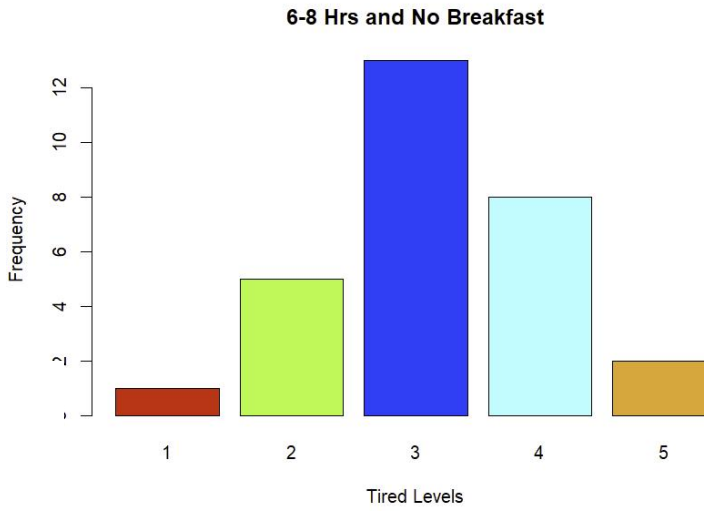

```
> PermutationTestSecond::Permutation(sleep, "Breakfast", "Tired", 1000, "Yes", "No")
[1] 0.003
```

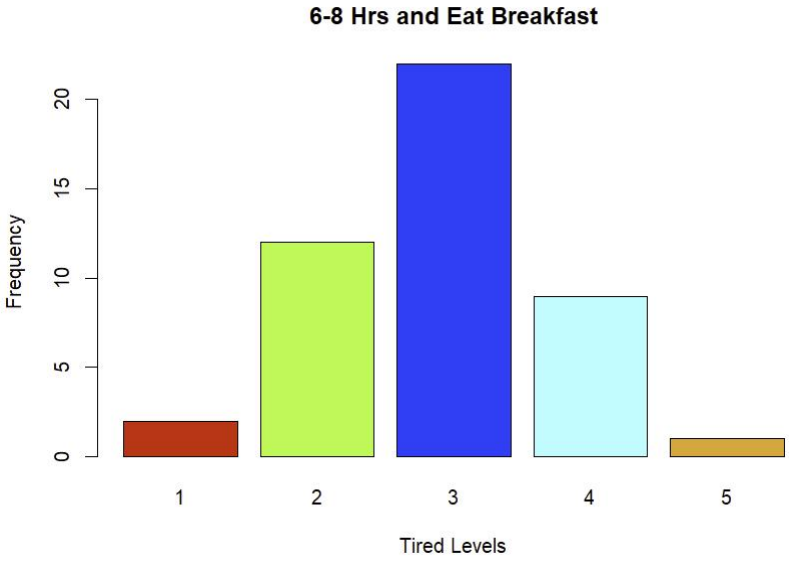
Null hypothesis: Mean tired levels of people who skip breakfast is the same as those that eat breakfast.
Alternative hypothesis: Mean tired levels of people who skip breakfast is higher than mean tired levels of people that eat breakfast.

The p-value is less than 5 percent so we reject null hypothesis. Therefore, we can conclude that eating breakfast is associated with feeling less tired, based on the available data.

Extra data:







Note: These are just extra data that I find worth observing. I will not be conducting a permutation test for this data due to the sample size being too small after subsetting it.

You can see that people that skip breakfast and get 6-8 hrs of sleep are more tired than those that eat breakfast. Also worth noting that people that sleep for more hours also chose to eat breakfast. Is this a correlation?

Conclusion:

From observing the plots:

- 6-8 hours of sleep is optimal for less fatigue
- Phone usages does not impact fatigue levels
- Eating breakfast is optimal for less fatigue
- Around 8 hours of sleep and eating breakfast is optimal for less fatigue.

Thoughts:

It goes to show that sample size is crucial for data analysis. When the sample size is too small, the data can appear to be random or even nonsensical. I often find myself having to exclude certain data or limit the depth of my analysis to avoid drawing false conclusions. It's not surprising that there are very few people who sleep for less than 4 hours. However, if the sample size were larger, we might be able to gain a better understanding of the relationship between these outliers.

As sleep is a complex topic, it is important to acknowledge that the data collected from the survey may be subject to bias, as various factors such as age, daily activities, and individual honesty could impact results. Overall, the experiment produced interesting results and was a worthwhile endeavor.