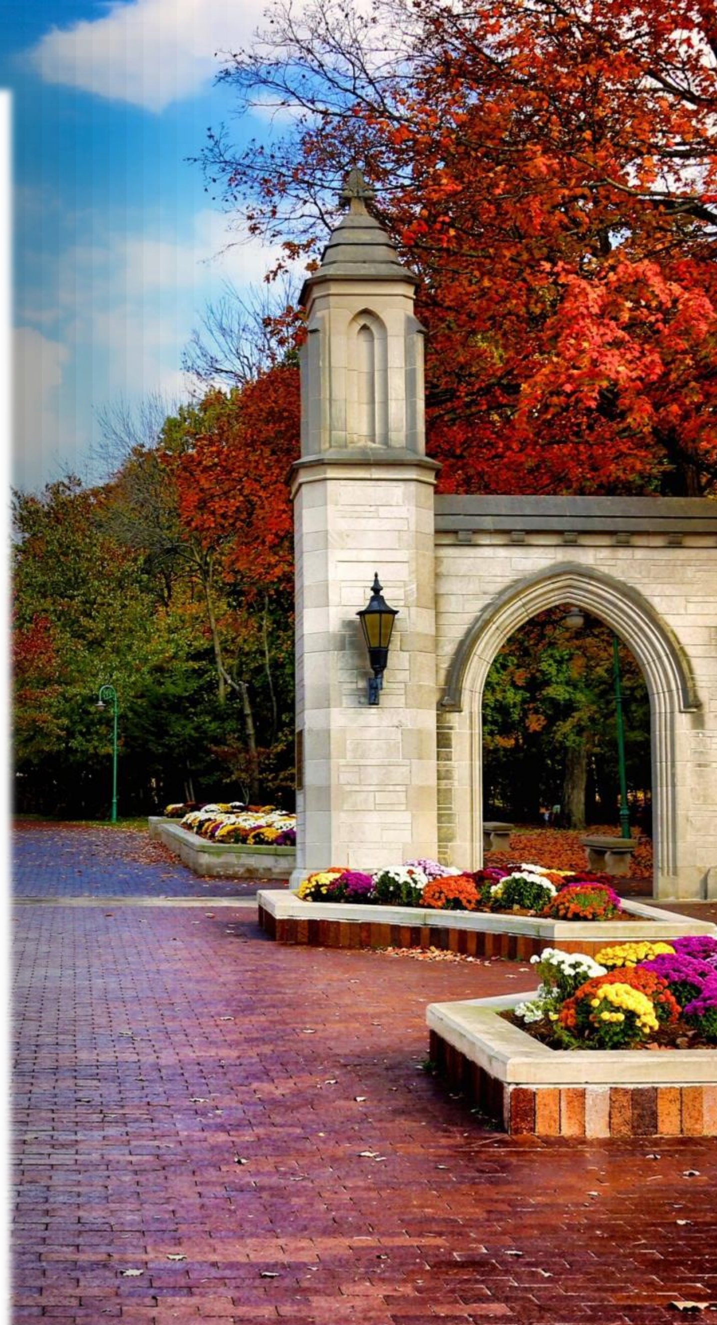
EDA

STAT-S 670

Final Project

Spring 2023



Authored by:

Atharva Shah

Maharshi Gor

Nihaar Nandedkar

Introduction

Sleep is vital for maintaining overall health and well-being. It is a natural state where the body is immobile, eyes are closed, and awareness is temporarily suspended, allowing the immune system, metabolism, and brain function to repair and restore themselves. Sleep efficiency and quality significantly affect physical and mental health, as well as daily performance and productivity. It helps the body recharge, repair, regulate hormones, maintain a healthy weight, control glucose metabolism, and prevent diseases like type 2 diabetes.

Sleep also enhances memory, learning, and cognitive performance, and it is essential for controlling mood and emotions. Sleep efficiency and quality are critical elements that determine how well a person sleeps, measuring how much time is spent sleeping, how often one wakes up during the night, and how rested they feel when they awaken.

Several lifestyle choices, environmental factors, and health conditions can affect sleep quality and quantity, including food, exercise, stress, temperature, light, noise, sleep apnea, restless legs syndrome, and other sleep problems.

The Sleep Efficiency dataset from Kaggle helps gain insights into how exercise, bedtime, caffeine, smoking, and other lifestyle choices impact sleep efficiency, aiming to identify ways to improve sleep for individuals. **By understanding the factors that impact sleep efficiency, individuals can improve their sleep habits and overall health. We aim to explore the fundamental factors that affect the sleep efficiency which proves to be a key factor in explaining sleep health [1].**

Methodology

1. Importing the relevant data into R by importing csv data into the R environment.

2. Cleaning and preparing the data by removing missing values, converting data types, and creating new variables as needed. Missing values in few columns were median imputed, and a threshold of 0.8, that is 80% sleep efficiency, was used as a factor of 1 for good sleep based on recommended research articles (1)

3. Visualizing the data using scatterplots and box plots to explore the relationship between sleep efficiency, sleep quality, and other factors.

4. Fitting linear regression models using R's lm() and loess() functions to quantify the relationship between variables. Separate models can be fit for each variable or a single model for all variables combined.

5. Evaluating model fit using metrics such as R-squared, adjusted R-squared, and mean squared error to determine whether a simple linear model is appropriate or a more complex model is needed.

6. Visualizing model output using ggplot2 by creating scatterplots of the data with the fitted regression line overlaid or plotting the residuals to check for heteroscedasticity or other issues.

7. Fitting alternative models such as polynomial regression or non-linear models using R if a simple linear model is not appropriate.

Dataset

The Sleep Efficiency dataset from Kaggle contains data on 452 instances of sleep and lifestyle factors, including age, gender, bedtime, wakeup time, sleep duration, sleep efficiency, REM sleep, deep sleep, light sleep, awakenings, caffeine consumption, smoking history, alcohol consumption, exercise frequency, and sleep quality. We plot histograms and ECDFs for all the variable, one of which is shown below as an example in figure 1 and figure 2. All the other dataset tendencies are in the appendix reference figures 1 to 19.

Age: This variable measures the age of the participant in years at the time of the sleep observation. It is a continuous variable.

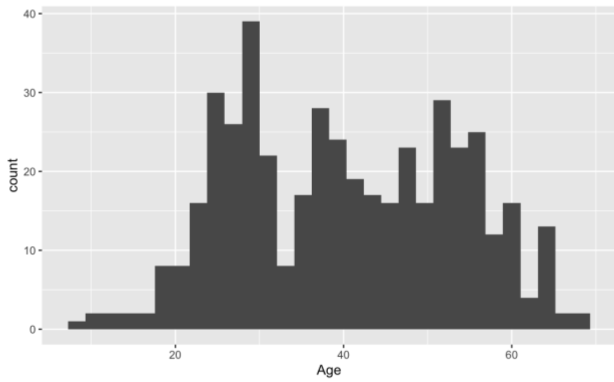


Figure 1: Age distribution histogram

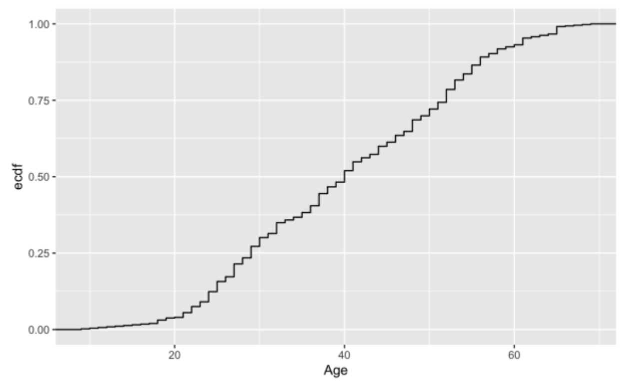


Figure 2: Age distribution ECDF

We see that Age is fairly normally distributed with study specimens ranging from 10 to 70 years. Majority of the sleep study patients are from 20-40 age where we see peak of the count but as such It is scattered throughout the age limits from 10 to 70 evenly as observed from Figure 1 and Figure 2 above.

Gender: This is a categorical variable that indicates the biological sex of the participant. The options are typically male or female. The dataset is balanced in terms of gender and hence no up-sampling or down-sampling is required to better the data characteristics. In total there are 452 study specimens with 227 being male and 225 being females.

Bedtime: This variable measures the time in 24-hour format that the participant went to bed on the night of the sleep observation. It is a continuous variable. The Bedtime available in the data is in 24 hours format and we see that there are only certain bins which is obvious as most common sleep schedule is expected to be in night. We see peak bedtime count at 12 am and the bedtime ranges from 8 pm to 3 am.

Wakeup time: This variable measures the time in 24-hour format that the participant woke up on the morning after the sleep observation. It is a continuous variable. The Bedtime available in the data is in 24 hours format and we see that there are only certain bins which is obvious as most common sleep schedule is expected to be in night. We see peak wakeup time count at 5 am and the wakeup time ranges from 3 am to 12 pm.

Sleep duration: This variable measures the total amount of time that the participant spent asleep on the night of the sleep observation. It is a continuous variable measured in hours. It can be seen that most of the people adhere to 7 hours mark as their optimum sleep duration and hence there is a peak at 7 hours mark with some sleeping for less than 6 hours and some also on 10 hours mark

Sleep efficiency: This variable measures the percentage of time spent asleep compared to the total time spent in bed. It is a continuous variable measured as a fraction. As recommended by the sleep society and one of the most cited journal articles ***-" Desjardins S, Lapierre S, Hudon C, Desgagné A. Factors involved in sleep efficiency: a population-based study of community-dwelling elderly persons. Sleep."*** **[1]**, 80% sleep efficiency should be considered as a healthy and efficient sleep. It is also found to be the same in the above graphs that 80-100% sleep efficiency is the most commonly found pattern among the study patients.

REM sleep: This variable measures the percentage of time spent in rapid eye movement (REM) sleep, which is a stage of sleep associated with dreaming. It is a continuous variable measured as a percentage. REM sleep (Rapid Eye movement) is a component of sleep cycle in which neurons form new connections and development of cognitive abilities enhances. It is stated that 20-40% of REM sleep is considered to boost cognitive awareness and abilities[1]. We also see the same pattern in our dataset also where most of the data ranges from 20-30%.

Deep sleep: This variable measures the percentage of time spent in deep sleep, which is a stage of sleep associated with physical restoration and growth. It is a continuous variable measured as a percentage. Deep sleep is known to be highly associated with the metrics of sleep efficiency and sleep quality. We see that most of the sample data is between 50-70%.

Light sleep: This variable measures the percentage of time spent in light sleep, which is a stage of sleep associated with relaxation and preparation for deeper stages of sleep. It is a continuous variable measured as a percentage. Light sleep percentage is also essential as it eases our body into a deep sleep cycle and it is seen that most of the data points lie between 10-20% while there is another cluster that is between 40-60%.

Awakenings: This variable measures the number of times the participant woke up during the night of the sleep observation. It is a discrete variable. The number of awakenings is seen from as less as no awakenings or 4 awakenings for the study patients. It has 14 missing values which are identified and removed as it is a discrete variable and no imputation would have been ideal to perform.

Caffeine consumption: This variable measures the amount of caffeine consumed by the participant on the day of the sleep observation. It is a continuous variable measured in milligrams. Caffeine consumption is the amount of caffeine consumed by the sleep participants in mg/day unit. A single cup of coffee would usually have 25 mg of caffeine and we see that generally study participants usually don't drink coffee at all or have 2 cups of coffee on the daily basis. Recommended limit of caffeine intake is around 30 mg and we can see all the data points largely following that limit.

Smoking history: This variable is a factor variable that indicates whether the participant is a current or former smoker or has never smoked. There are nearly 300 instances of participants with no smoking history. Although we don't have the data about if they did smoke previously or how prior to the study did they quit but this can also provide valuable insights for the research problem. This only shows active smoking status and we encounter about 7 missing values which are ignored through some variable manipulation.

Alcohol consumption: This variable is a binary variable that indicates the amount of alcohol the participant consumes alcohol regularly. We can see that most of the participants have no alcohol consumption whereas it is evenly distributed between 30 ml to 150 ml of alcohol levels. We see 9 missing values in the dataset which are handled as we handle missing values in smoking status that we explore later.

Exercise frequency: This variable measures the number of times per week the participant engages in exercise. It is a discrete variable. We observe that exercise frequency ranges from 0 times a week to 5 times a week. There are 11 missing values which are median imputed to better maintain the dataset tendencies.

Sleep quality: This variable is a factor variable that indicates the subjective quality of the participant's sleep. Sleep quality is measured using a subjective rating scale from 1 to 5, where 1 represents "very poor" and 5 represents "very good." Sleep quality is subjective measure of sleep but thanks to modern technologies, brain waves frequency can be translated into a 5 category datapoint.

Overall, the Sleep Efficiency dataset provides a comprehensive set of variables that can be used to explore the relationships between sleep efficiency, lifestyle choices, and other factors that impact sleep quality. The variables are diverse, and the dataset contains a sufficient number of observations to allow for robust statistical analysis.

Data exploration and outliers handling

After initial preprocessing of the data which included converting few of the categorical variables to suitable formats i.e., bedtime and wakeup time to suitable numerical hours and handling the missing values by either median imputation or by ignoring certain binary variables, we check for the outliers through boxplots and statistical thresholds and treat them accordingly. One of the boxplots among the many plotted is shown here for the sake of simplicity and space saving.

Deep sleep percentage amongst many has few outliers on the lower tail as seen from the figure. We choose to ignore every outlier because of 2 reasons. Mainly we wanted to see whether extreme lifestyle guarantees poor sleep cycles and secondly, we also wanted to explore what are inter-correlation of the variables in such extreme data points.

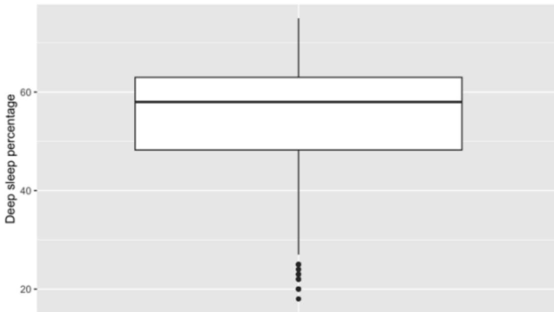


Figure 3: Deep sleep percentages boxplot

We start by exploring the heatmap of the correlation matrix that is in Figure 4. A heatmap is a useful tool for understanding the correlation between variables in a dataset. By visualizing the correlation matrix as a heatmap, we can quickly see which variables are strongly correlated, which variables are weakly or negatively correlated, and which variables may be redundant or highly correlated with other variables in the dataset. This information can guide further analysis and help identify which variables may be most useful for predicting outcomes or building models.

Sleep quality and sleep efficiency have a moderate positive correlation of 0.56. ***This indicates that individuals who report higher sleep quality are more likely to have higher sleep efficiency, and vice versa***. The positive correlation between sleep quality and sleep efficiency (Ref. figure 20) suggests that individuals who rate their sleep quality higher are likely to spend more of their time in bed actually asleep, rather than tossing and turning or being awake. This could be due to a variety of factors, such as having a regular sleep schedule, minimizing exposure to bright screens before bed, and practicing relaxation techniques to reduce anxiety and stress. Conversely, individuals who have lower sleep quality may be more likely to experience disrupted sleep, resulting in lower sleep efficiency. ***The other interesting thing to note is that sleep quality and sleep efficiency is highly negatively correlated with awakenings and deep sleep/ light sleep percentage*** (Ref. figure 26).

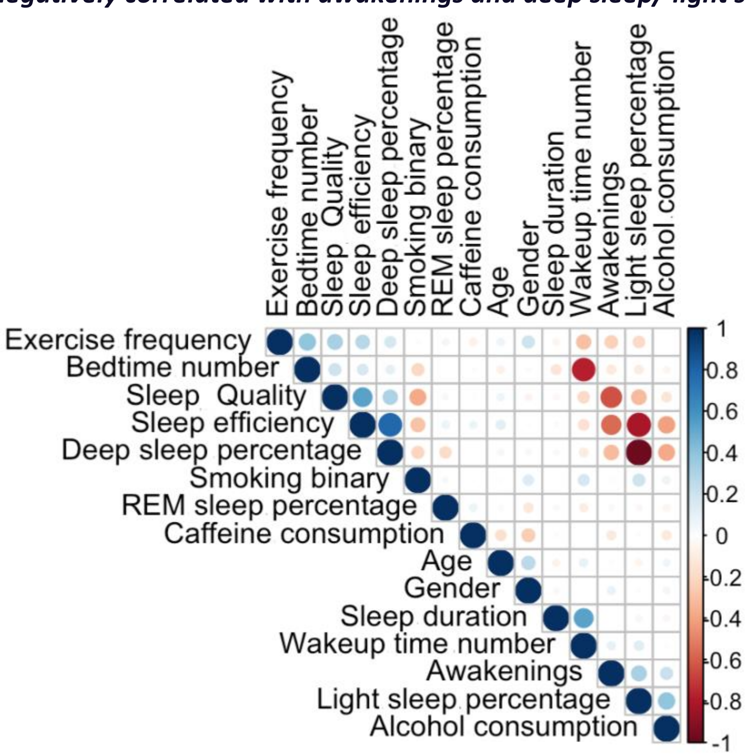


Figure 4: Correlation plot for all the features

There is an upward trend which indicates a positive correlation between the features and it can be seen that people with higher sleep quality tend to have a higher sleep efficiency as well with the cluster of points higher than ones in lower sleep quality bins (Ref. figure 20).

As seen by the correlation heatmap, it can be seen visually that less awakenings are ***better for sleep efficiency and after 2 awakenings, it seems that sleep efficiency stagnates and does not depend on awakenings altogether***. We can draw an important conclusion that less awakenings are a crucial factor for sleep.

We also see that ***sleep quality and efficiency are not related to age, gender and sleep duration*** as they have barely any correlation Figure 5. We also investigate it through plotting scatter plot among these variables that is against sleep efficiency and sleep duration, faceted out by age bins of 20 and color coded by gender. After **converting sleep efficiency to a binary variable by 80% threshold** [1] value as discussed in the report earlier, a mosaic plot (Ref. figure 25) has been plotted which also confirms that sleep efficiency and gender are actually independent. Even though lines do not match perfectly but it isn't that far off to draw alternative conclusions concretely. We move on to explore if sleep cycle percentages have any effect on the sleep efficiency and quality. Essentially Light sleep + REM sleep percentage + Deep sleep percentage = 100% as during our sleep it alternates between these three stages. (Ref. figure 21,22,23)

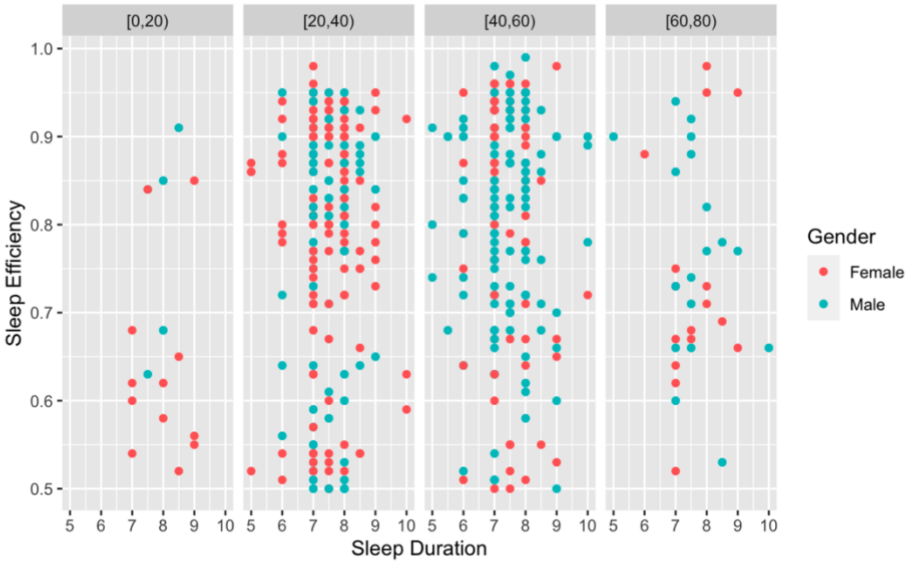


Figure 5: Sleep efficiency vs Sleep duration faceted out by age

All the graphs show that ***only deep sleep/light sleep has effects on sleep efficiency and not REM sleep on both of the genders.*** We also see that deep sleep or light sleep has no effects on sleep quality as well for both of the genders. (Ref. figure 21,22,23)

Other lifestyle factors that can impact awakenings and sleep quality include exercise habits, stress levels, and bedtime routines. Regular exercise has been shown to improve sleep quality and reduce the frequency of awakenings. Stress management techniques such as meditation or relaxation exercises can also help reduce stress levels and improve sleep. Establishing a consistent bedtime routine, such as avoiding screens before bed and winding down with a calming activity like reading, can also improve sleep quality and reduce the frequency of awakenings.

In addition to lifestyle factors, environmental factors such as noise and light levels can also impact sleep quality and awakenings. Creating a comfortable sleep environment with minimal noise and light can help promote better sleep quality and reduce the frequency of awakenings.

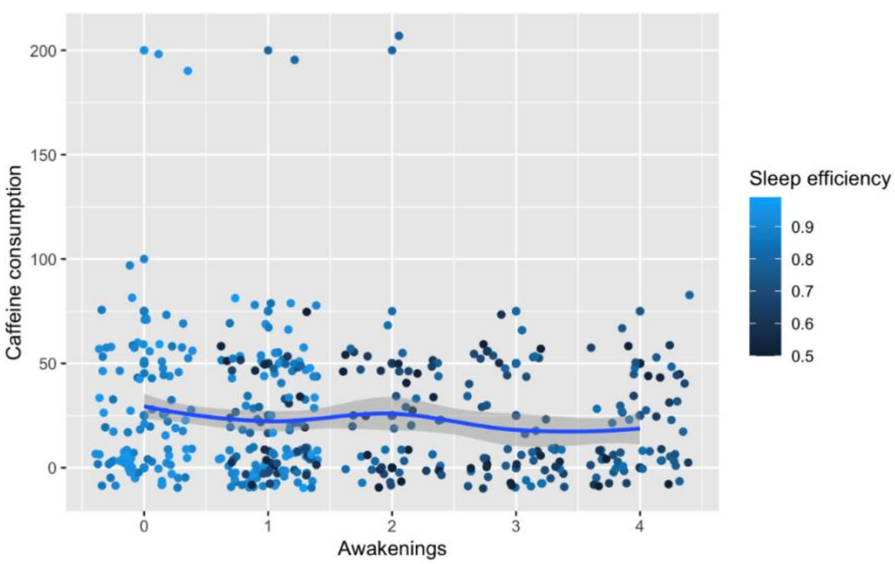


Figure 6: Sleep efficiency vs Sleep duration faceted out by age

Caffeine is a stimulant that can interfere with the natural sleep cycle by blocking the action of adenosine, a chemical in the brain that promotes sleep. As a result, consuming caffeine close to bedtime can make it more difficult to fall asleep, reduce the amount of deep sleep, and increase the number of awakenings during the night. (Figure 6).

In terms of sleep efficiency, caffeine consumption can decrease the percentage of time spent in bed actually asleep. This can be particularly problematic for individuals who consume large amounts of caffeine or who consume it later in the day. To promote better sleep quality and sleep efficiency, it is generally recommended to limit caffeine consumption or avoid it altogether, especially in the hours leading up to bedtime. We see that

there is no relation between caffeine and awakenings at all but just awakenings and sleep efficiency. Sleep efficiency is obvious as it is measured that way but sleep quality also shows decline with higher number of awakenings. (Ref. figure 27)

We also see that ***smoking and caffeine (both neuro-stimulants) together also have very little effects on the awakenings as well.*** Smoking and caffeine can have different effects on different individuals and that their impact on sleep and awakenings can depend on various factors, such as the amount consumed, timing of consumption, and individual tolerance levels. (Ref. figure 28)

While some studies have suggested that smoking and caffeine can contribute to awakenings during the night, other studies have found no significant relationship between smoking or caffeine consumption and sleep disruptions. One possible explanation for the lack of effect could be that individuals may develop a tolerance to the effects of caffeine or smoking over time, reducing their impact on sleep quality and awakenings.

We also wanted to see whether there is any relation between when you go to sleep vs sleep efficiency or quality. We plot a scatter plot of bedtime vs sleep efficiency. We observe that ***there is no relation with the time you sleep with either sleep efficiency or sleep quality.*** It just goes on to show that sleep efficiency and sleep quality go hand in hand generally and it is not upon when you sleep. It is due to the fact that when you sleep doesn't have any significance on how well you sleep. (Ref. figure 29)

We also see the effects of exercise on sleep quality and sleep efficiency. Regular exercise can have a positive impact on sleep quality and efficiency, with some studies suggesting that it can lead to improvements in both parameters. Exercise has been shown to increase the amount of time spent in deep sleep, which is important for restorative processes in the body, and can also help to reduce the time it takes to fall asleep and the number of awakenings during the night.

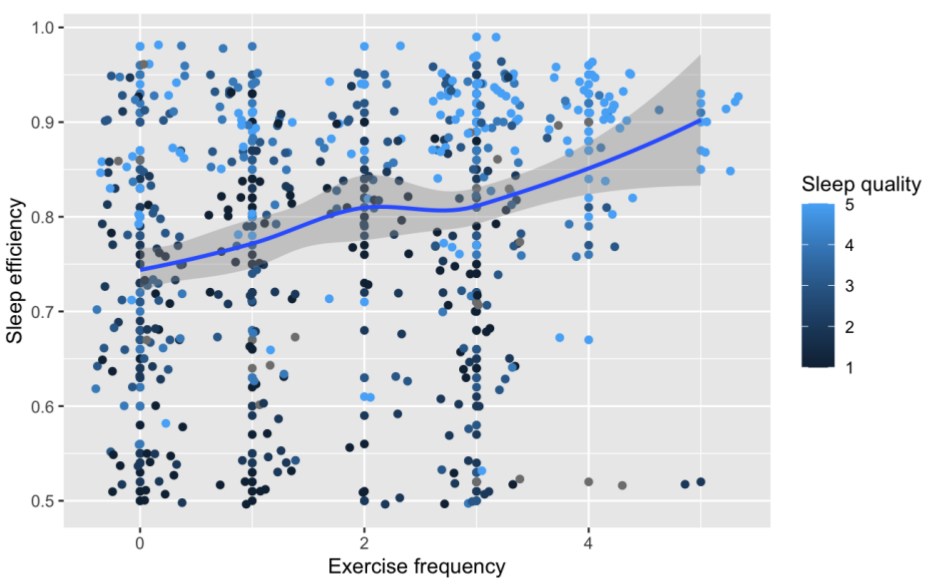


Figure 7: Sleep efficiency vs Exercise frequency

The frequency of exercise can also play a role in its effects on sleep quality and efficiency. Some studies suggest that exercising at least three to four times a week can lead to improvements in sleep, while others have found that daily exercise may be even more beneficial. However, it is worth noting that the timing of exercise can also be important, with some studies suggesting that exercising too close to bedtime may actually have a negative impact on sleep quality and make it more difficult to fall asleep. (Figure 7).

We also see that mosaic plot (Figure 8) also showcases these dependencies and it is obvious that ***people who exercise more frequently are more health conscious and generally tend to sleep better and adopt healthier lifestyle choices***. These people also in general do not engage in unhealthy habits as smoking which is shown in the plot below. (Figure 9).

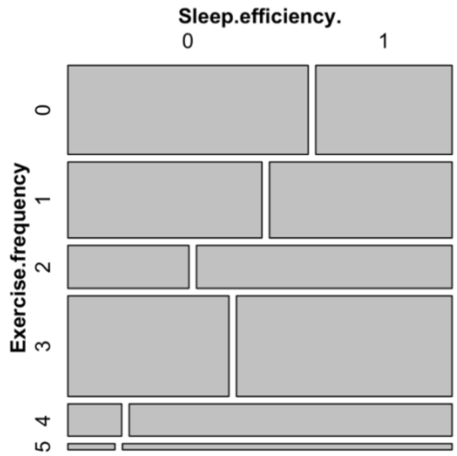


Figure 8: Mosaic plot Sleep efficiency and Exercise frequency

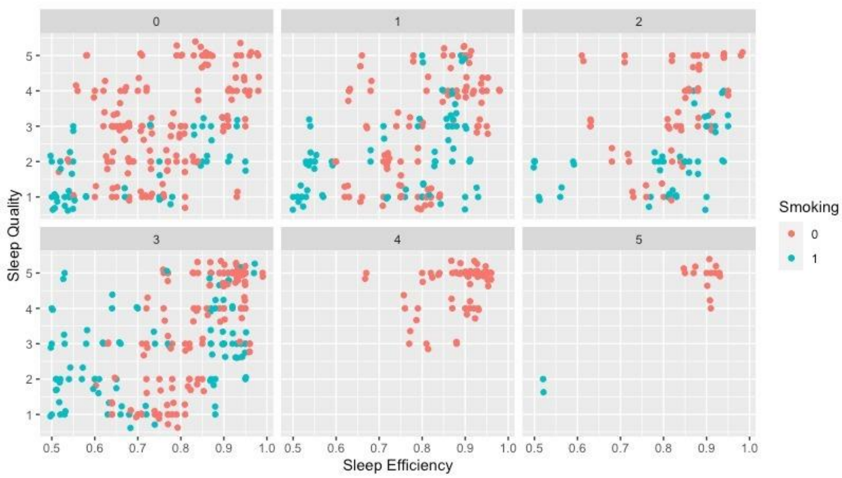


Figure 9: Sleep efficiency vs Sleep quality faceted out by exercise frequency

As we see that as exercise frequency increases, the cluster of black dots shift to right top corner indicating higher sleep efficiency and sleep quality and in general do not smoke (Figure 9). Even ***alcohol consumption doesn't show any concrete correlation***. Although lower alcohol consumption does tend to increase the sleep quality but no concrete statements can be made. (Ref. figure 30)

Even though, alcohol consumption can have a significant impact generally on both the quality and efficiency of sleep. While alcohol is a sedative and can make it easier to fall asleep initially, it can also disrupt the natural sleep cycle and lead to poorer overall sleep quality. Only smoking affects significantly to both the variables

Feature selection and dimensionality reduction

After exhausting all the correlations, we move on to the next stage of feature selection and dimensionality reduction. We see that **deep sleep/light sleep percentage, awakenings, exercise frequency and smoking** have significant impact on overall sleep health. We aim to boil it down two principal components across these features for all 452 samples. We see that loading of PC1 and PC2
as follows for top 4 features only as per PCA. (Table 1)

Features	PC1	PC2
Deep sleep percentage	-0.71	-0.04
Awakenings	0.20	0.001
Exercise frequency	-0.12	0.006
Sleep quality	-0.021	0.047

Table 1: PCA loadings for top 4 features

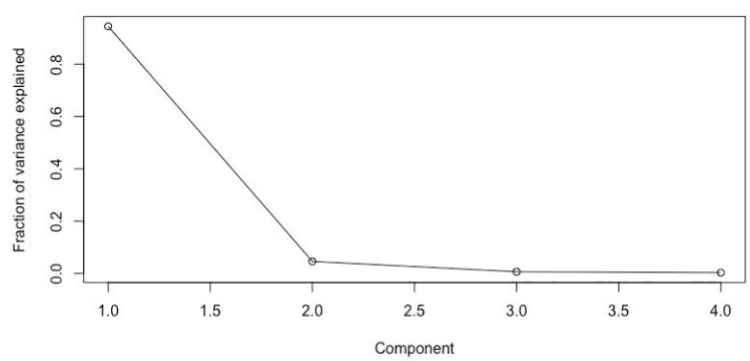


Figure 10: Scree plot for PCA

We transform all the 452 samples with these PC rotations and plot a scree plot which shows that one PC explains 99% of the variance explained. The said scree plot is given in Figure 10.

A scatter plot for the said PCA gives very convincing results as it can be seen that a line threshold can be fitted with the given PCA (PCA loadings are in (Ref. figure 31)). We show ***the linear regression fit*** in the graph below for sleep efficiency with **5 highly correlated variables obtained from data explorations above i.e., Sleep quality, Deep sleep percentage, Exercise frequency, Smoking status, Awakenings (Figure 11) and threshold of 0.8 for sleep efficiency as discussed above.**

Linear Model Fit

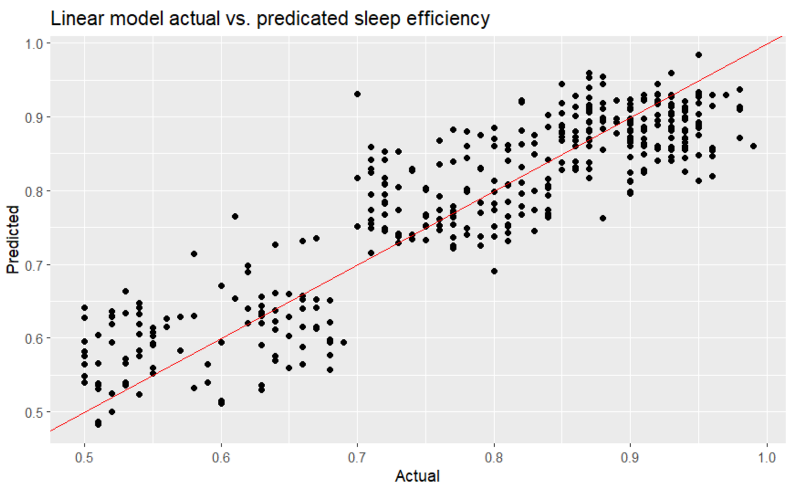


Figure 11: Linear fit model

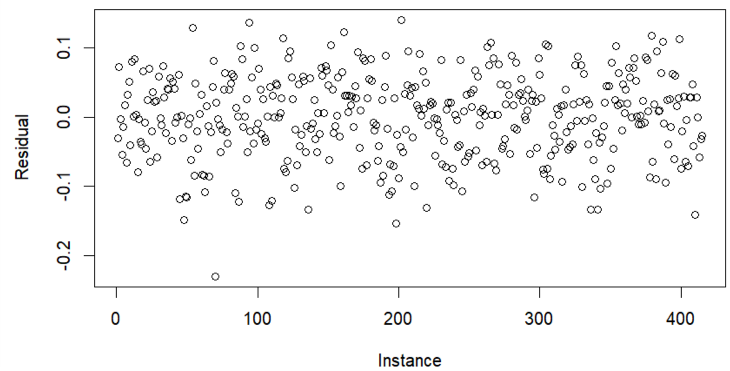


Figure 12: Residuals for the linear fit

Since most of the points fall along the AB line and don't deviate much from it, we can say that the model does a good job at predicting the sleep efficiency percentage. This can also be proved from the residuals plot which shows a straight line centered close to 0. which implies that the errors are normally distributed and unbiased, further supporting the model's effectiveness. Therefore, based on the alignment of predicted and actual values, as well as the residuals plot, we can conclude that the model does a good job at predicting the sleep efficiency percentage. We have also explored the linear fit with PCA components (Ref. figure 32,33) but it didn't yield decent results. Results are in table 2.

Logistic Model

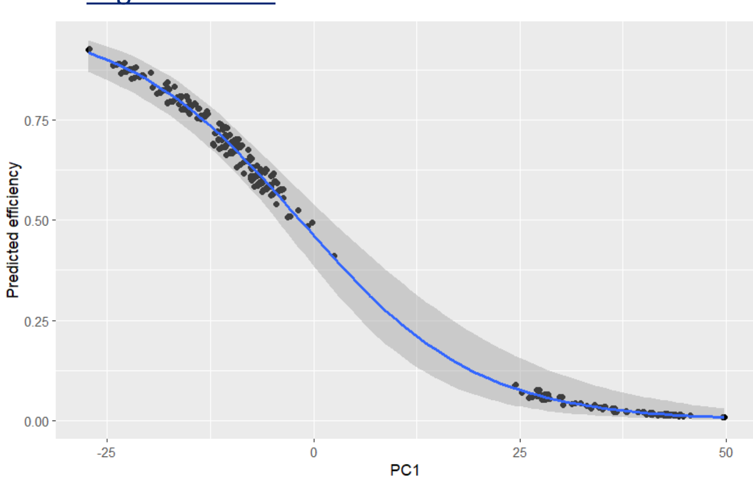


Figure 13: Logistic model fitted with PCA

Logistic Model does a good job at predicting if an individual's sleep efficiency is **above the 80% threshold or not.** From the plots of the model fit for PC1 we see that, the likelihood of higher sleep efficiency decreases with increase in PC1 and the most directed vectors for PC1. We can conclude that the logistic model is effective in predicting sleep efficiency above or below the 80% threshold, and PC1 is a significant predictor of sleep efficiency. We have also explored the logistic regression fit with raw variables (Ref. figure 34,35)

Model	Linear model without PCA	Linear model with PCA	Logistic model with PCA
Model results	R-squared = 0.79	R-squared = 0.66	Accuracy-79.55

Table 2: Model results with different result parameters

Limitations

As with any dataset obtained from an open-source website, the sleep efficiency dataset used in this analysis also has some limitations which are as follows –

1. As the dataset used in this analysis is generated based on the results of a controlled study conducted on a small set of people, the number of instances in the dataset is very low which in turn could produce false trends and correlations and thus affect the final model performance.

2. Wearable devices used in this study to track variables like sleep duration, light sleep percentage, deep sleep percentage, REM sleep percentage and awakenings aren't always fully accurate and could thus generate incorrect data for some instances.

3. Although it is well known that caffeine, alcohol, and smoking can affect your sleep, their effect on your sleep also varies on various other factors like how much caffeine/alcohol one consumes or how regularly one smokes. The dataset we have used contains the frequency of consumption rather than amount of consumption for alcohol and contains a binary indicator for smoking status rather than a count of how many times one smokes.

4. Caffeine, Alcohol and Smoking are more likely to disturb the sleep cycle when consumed closer to bedtime as compared to any other point in the day. As our dataset does not contain any information for when the person consumes either of these 3 it could create misleading observations or trends in the analysis.

5. Although exercise plays a big role in regulating your sleep cycle and thus improving your sleep quality and efficiency, not all forms of exercise have the same impact on your sleep cycle. The dataset only tracks the exercise frequency of the person over the week and not the intensity of the exercise.

6. There are multiple methods that can be used to measure the sleep quality for an individual and hence our analysis only holds for the method that was used in the study that generated this dataset.

Despite the limitations of this dataset, since the distribution of the data is mostly normal in the case of continuous variables and not imbalanced in the case of categorical variables, the models generated from this dataset perform well and do a good job at capturing the general trend.

Conclusion

From the results of the exploratory data analysis, principal component analysis and model fitting, the following conclusions can be drawn –

1. Sleep quality seems to be a good indicator of sleep efficiency as people with higher sleep quality generally have a higher sleep efficiency. A person's age and gender seem to be ineffective in determining their sleep efficiency.

2. To a certain extent, a person's habits like exercise frequency, alcohol/caffeine consumption and mainly smoking can influence their sleep quality and thus impact their sleep efficiency.

3. Although it is commonly believed that sleeping early and waking up early is better for an individual, the results from this study show that time of sleeping and time of waking up have absolutely no impact on your sleep efficiency.

4. Sleeping longer is not important but rather how much of your sleep is spent in light sleep, deep sleep, or REM sleep. From our analysis, it is evident that having a higher percentage of deep sleep and a lower percentage of light sleep in your overall sleep results in a higher sleep efficiency. REM sleep on its own was found to have no impact on sleep efficiency.

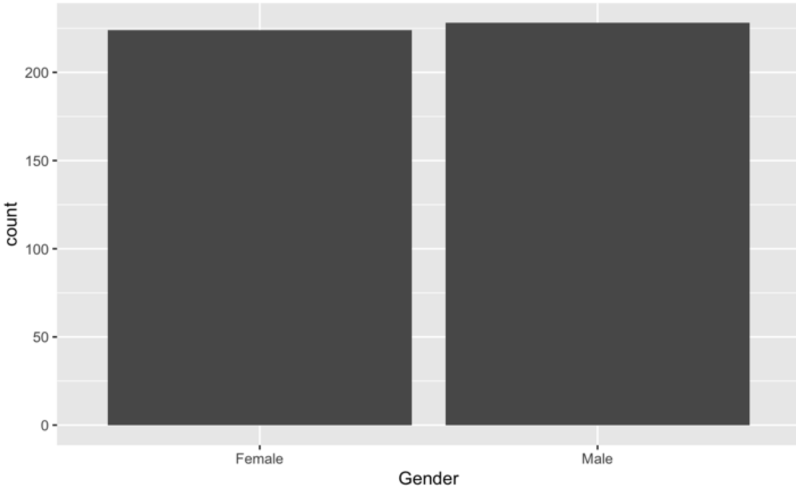
5. An individual's total awakenings at night greatly impacts their sleep quality and sleep efficiency. A person's lifestyle choices like exercise frequency, and whether they consume caffeine or alcohol, or if the person smokes or not can impact their total awakenings.

6. On fitting a linear model to predict the sleep efficiency as a function of the interaction between an individual's deep sleep percentage, sleep quality, awakenings, and lifestyle choices, we observe that the model does a fairly good job of at predicting an individual's sleep efficiency as the RMSE of the fitted model is 0.06 and the R-squared value is 78.9

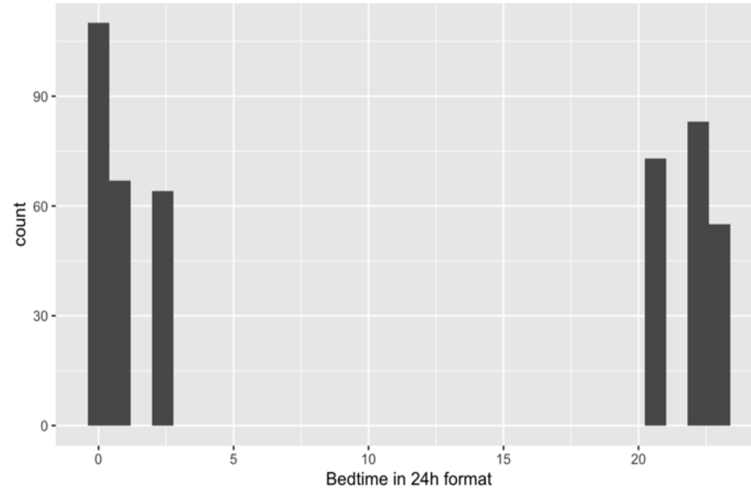
7. On fitting a logistic regression model to predict whether a person's sleep efficiency is above the threshold of 80% or not as a function of the person's deep sleep percentage, awakenings, lifestyle choices, and sleep quality we get a model with an accuracy of 88.43%.

8. On performing Principal Component Analysis, we observe that PC1 and PC2 together explain ~96.4% of the variation in the data. By using these Principal Components to create a linear model without interaction for predicting sleep efficiency, we obtain an RMSE of 0.079 and an R-squared value of 65.5. Using the Principal Components to create a logistic regression model we get a model with an accuracy of 79.55%. According to PCA biplot, we can see that most significant factor to determine sleep efficiency was Deep sleep percentage, Awakenings and Exercise frequency which determine overall sleep health

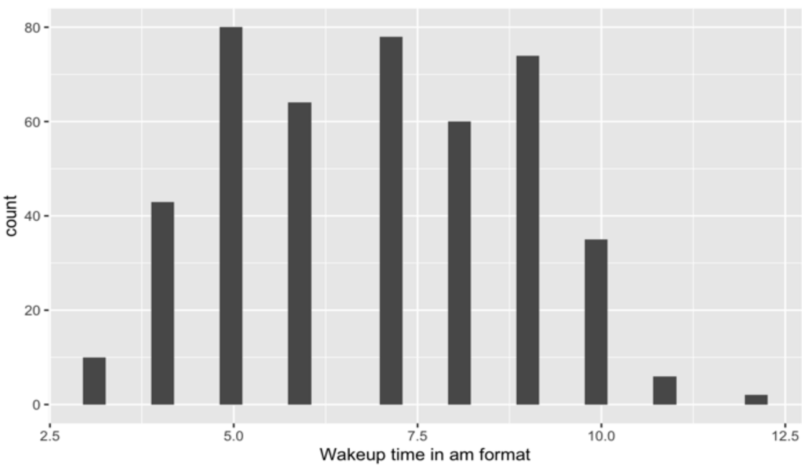
Appendix



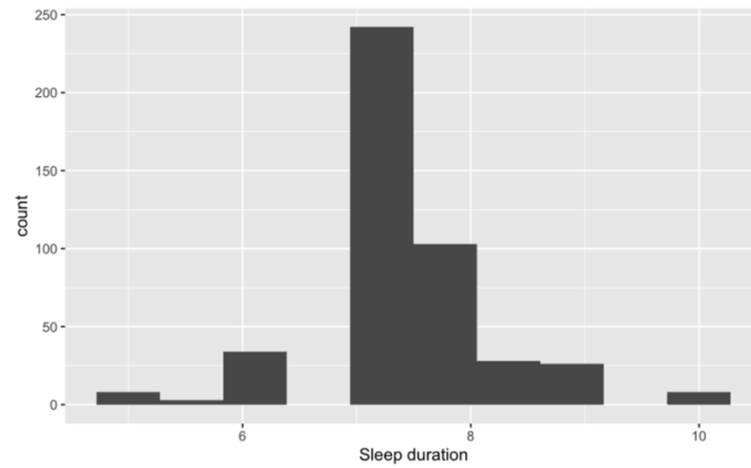
Ref. Figure 1: Gender histogram



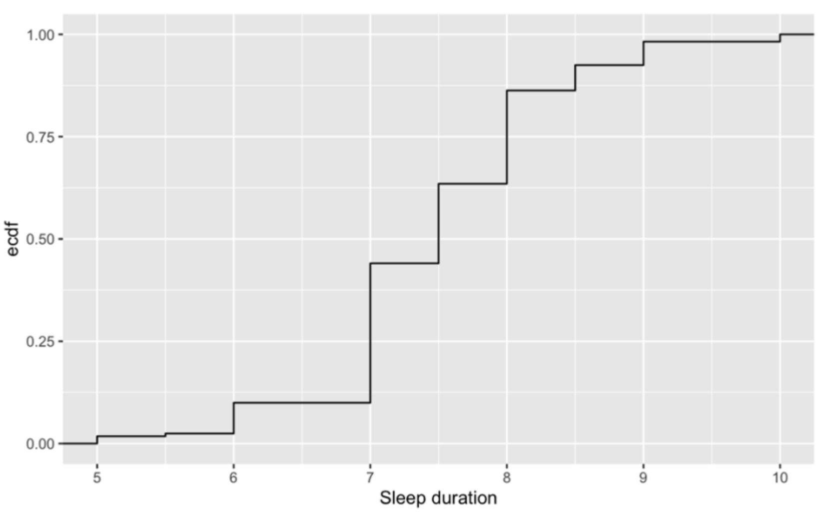
Ref. Figure 2: Bedtime hours histogram



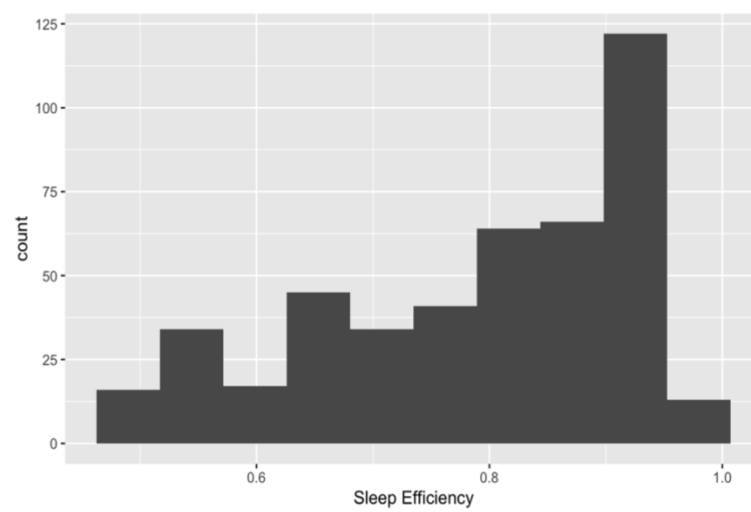
Ref. Figure 3: Wakeup time hour histogram



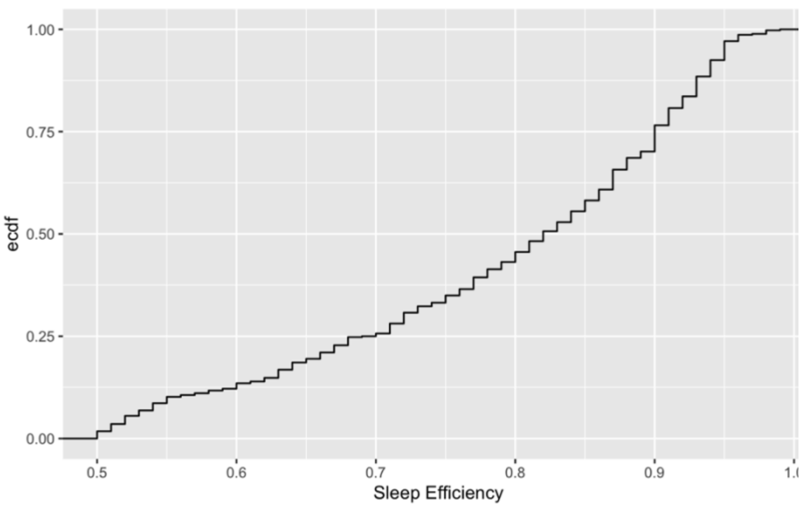
Ref. Figure 4: Sleep duration histogram



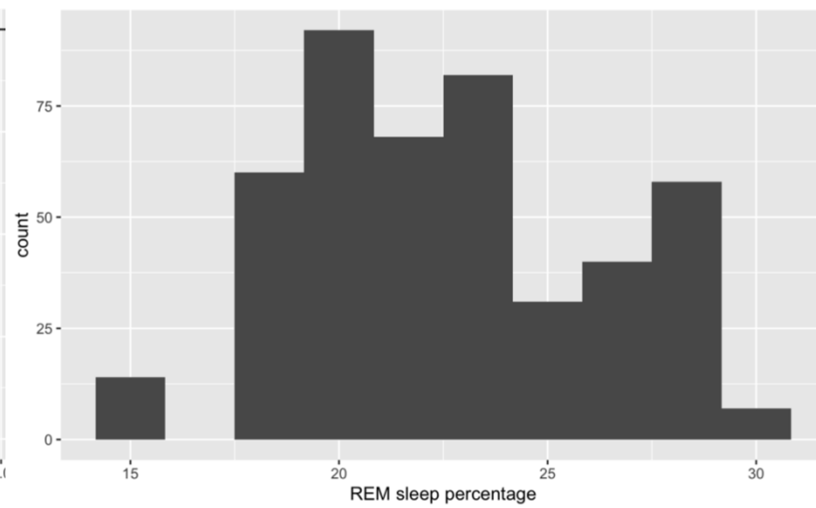
Ref. Figure 5: Sleep duration ECDF



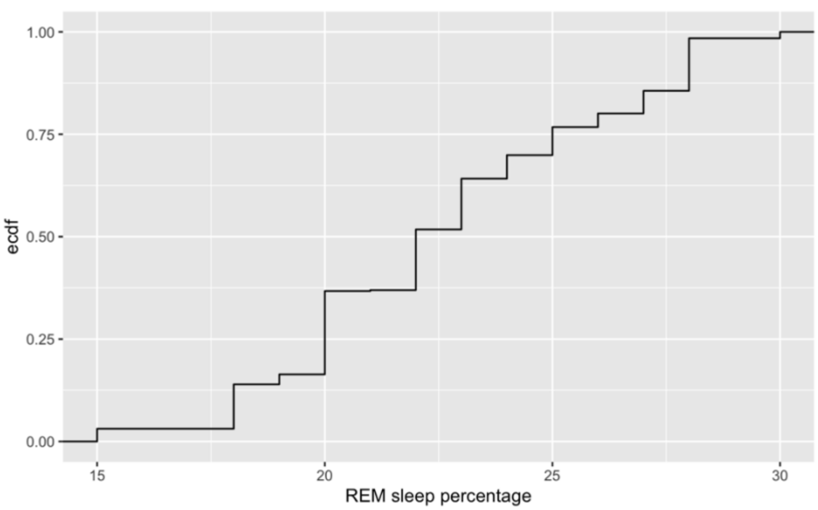
Ref. Figure 6: Sleep efficiency histogram



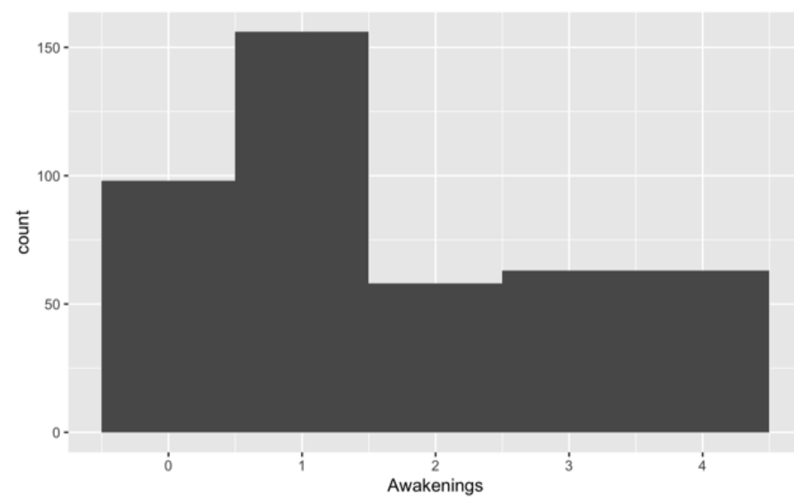
Ref. Figure 7: Sleep Efficiency ECDF



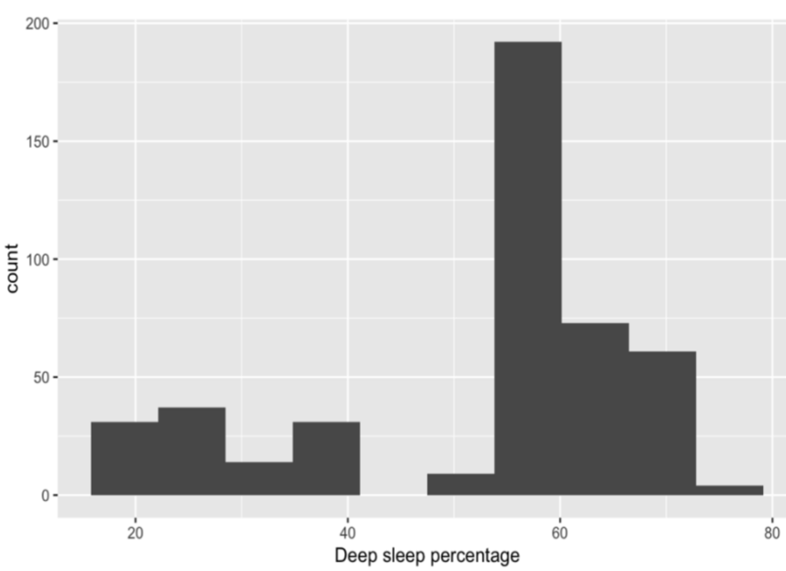
Ref. Figure 8: REM sleep percentage histogram



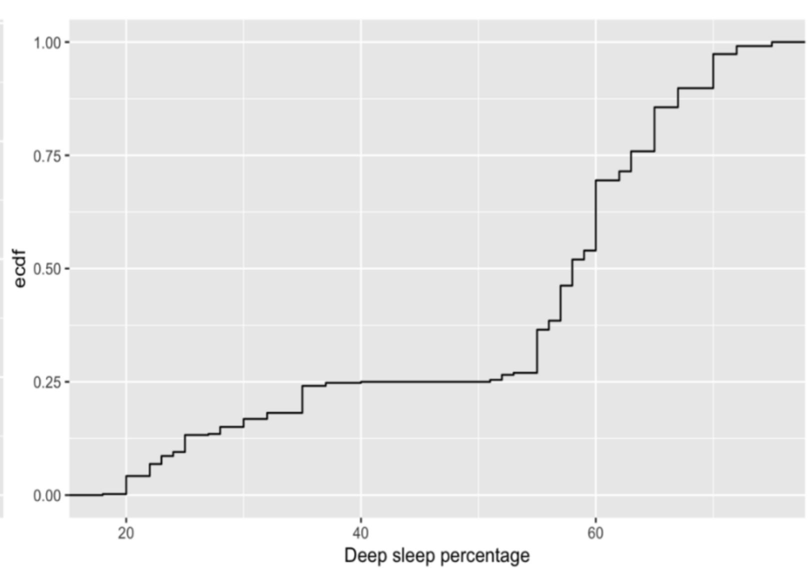
Ref. Figure 9: REM sleep percentage ECDF



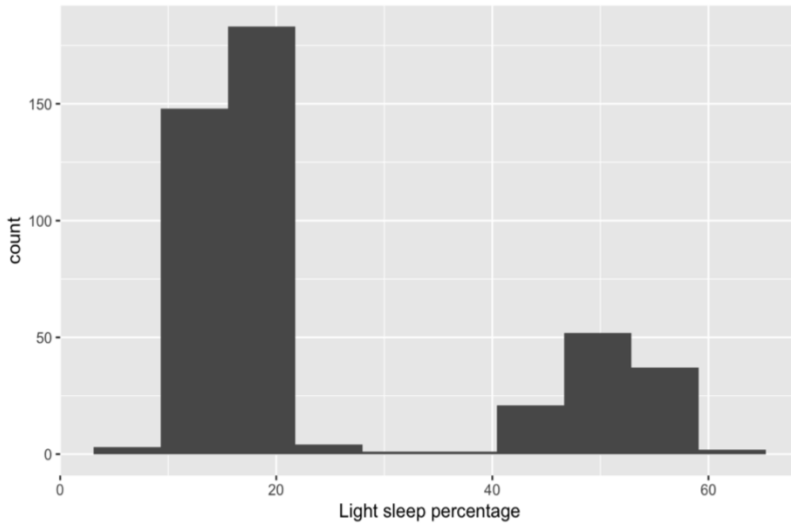
Ref. Figure 10: Awakenings histogram



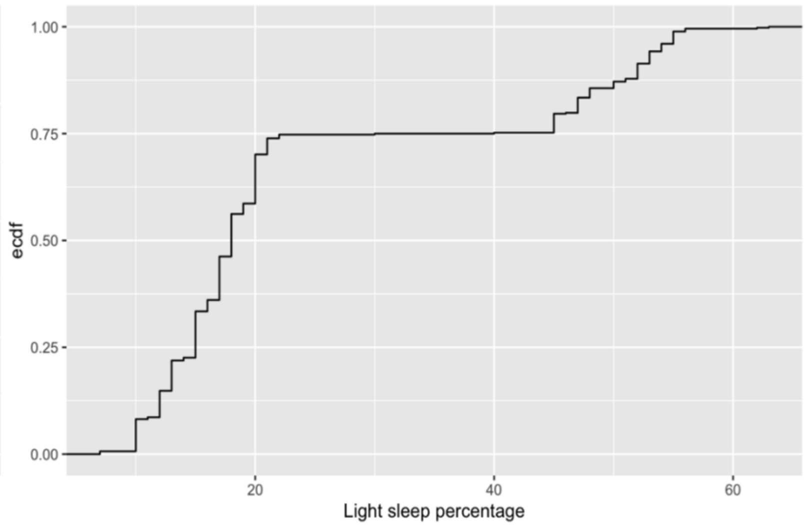
Ref. Figure 11: Deep sleep percentage histogram



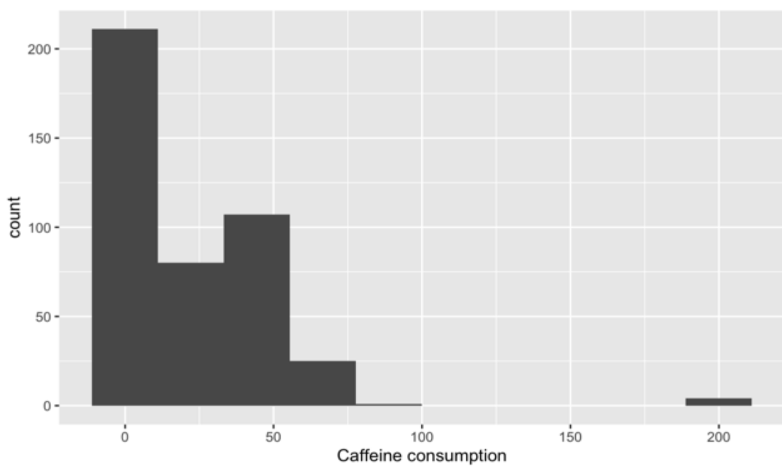
Ref. Figure 12: Deep sleep percentage ECDF



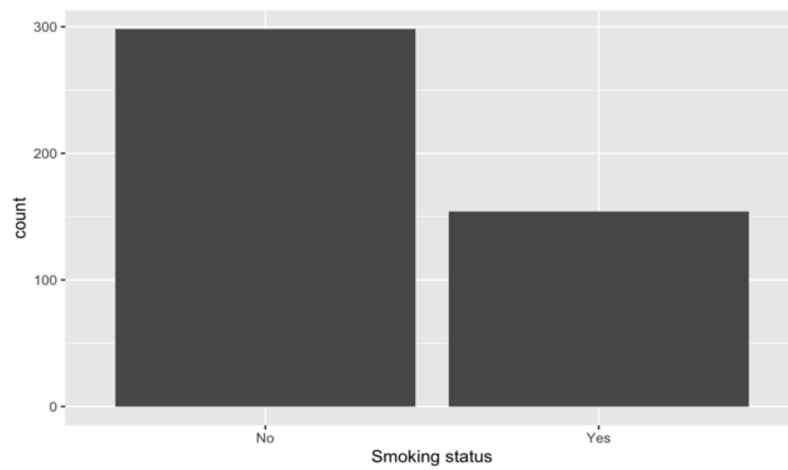
Ref. Figure 13: Light sleep percentage histogram



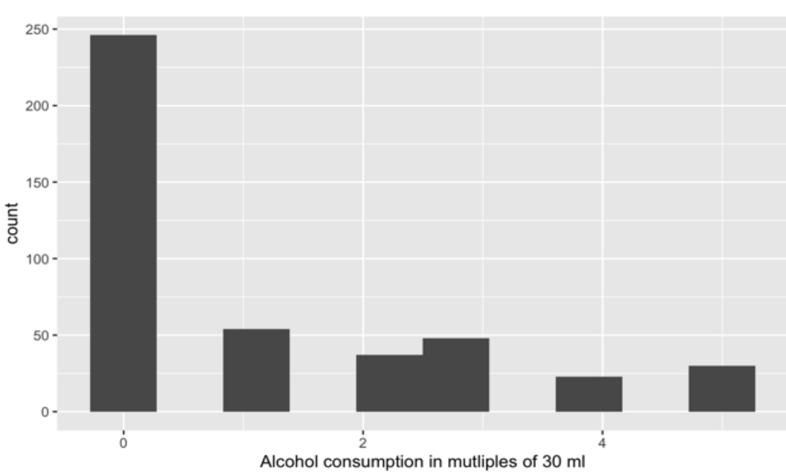
Ref. Figure 14: Light sleep percentage ECDF



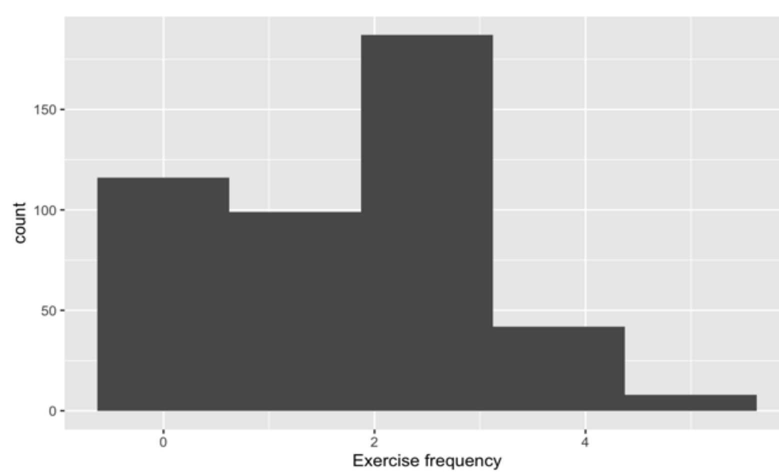
Ref. Figure 15: Caffeine consumption histogram



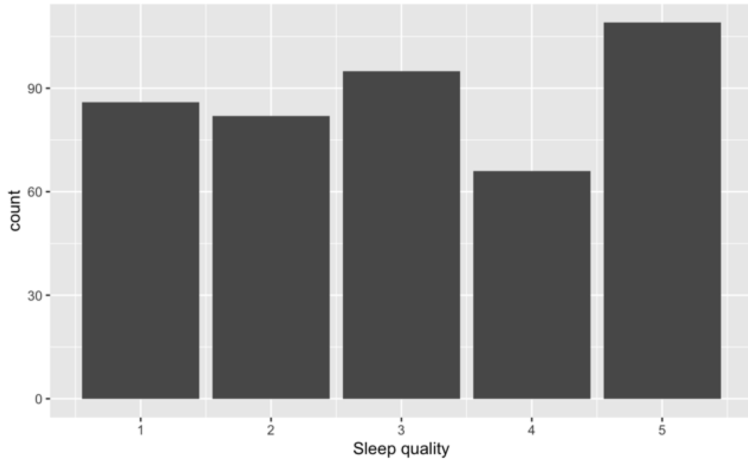
Ref. Figure 16: Smoking status histogram



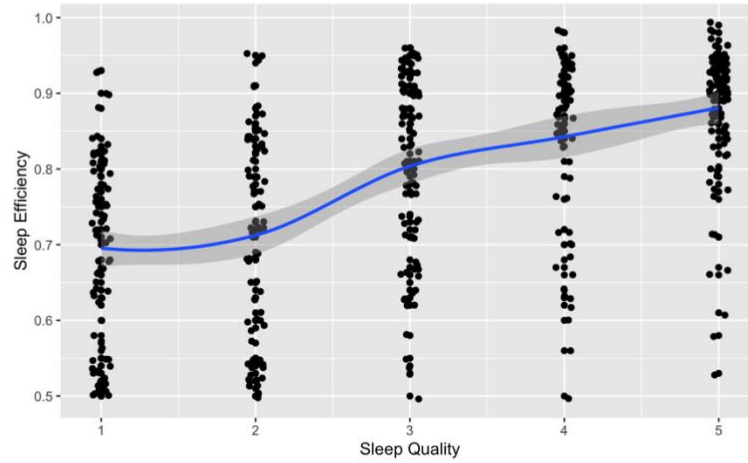
Ref. Figure 17: Alcohol consumption histogram



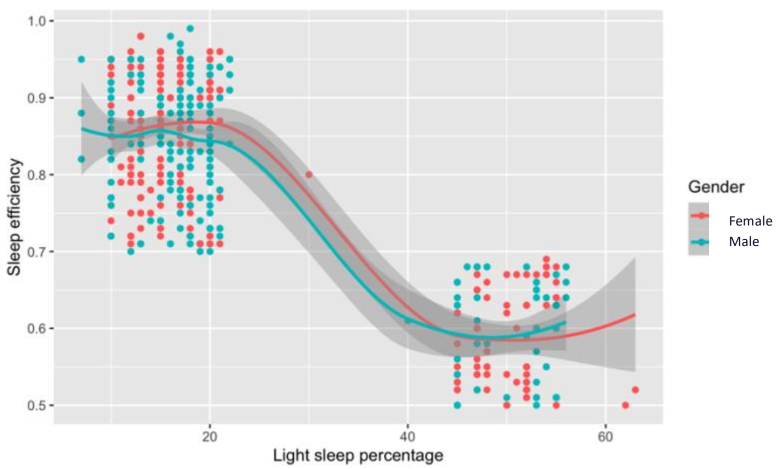
Ref. Figure 18: Exercise frequency histogram



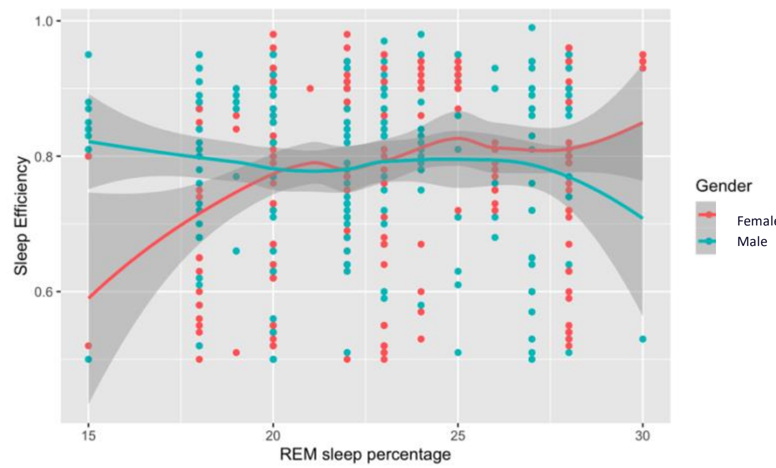
Ref. Figure 19: sleep quality histogram



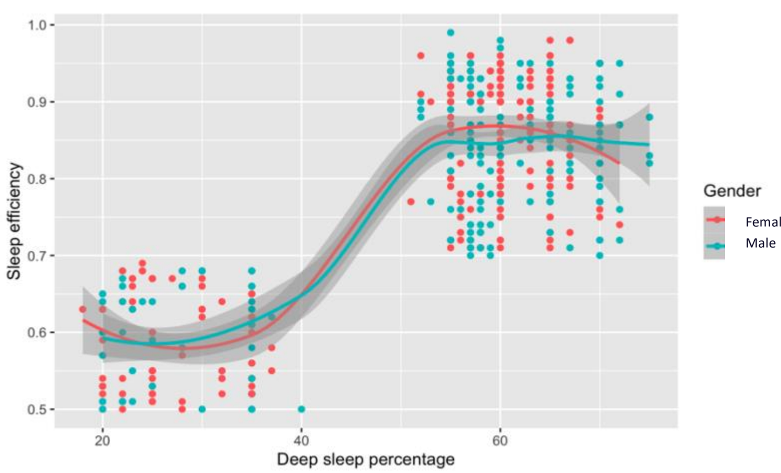
Ref. Figure 20: Sleep Quality vs Sleep Efficiency



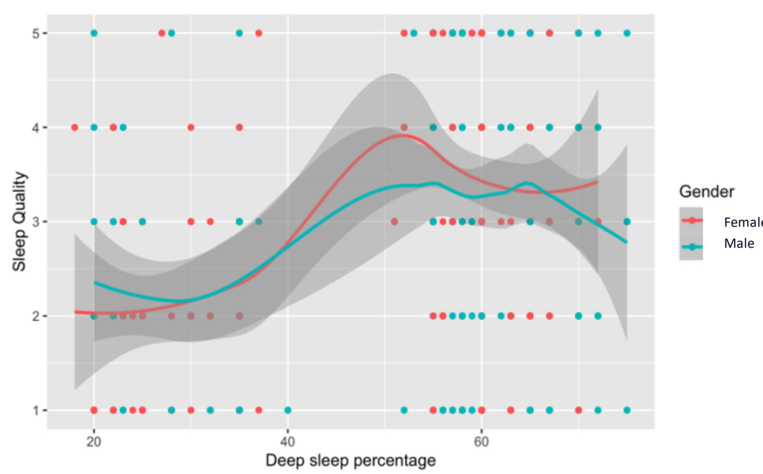
Ref. Figure 21: Light sleep percentage vs Sleep Efficiency



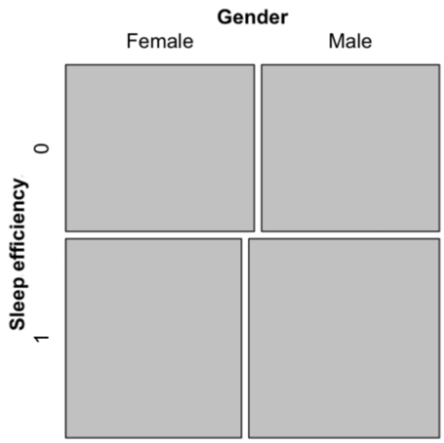
Ref. Figure 22: REM sleep percentage vs Sleep Efficiency



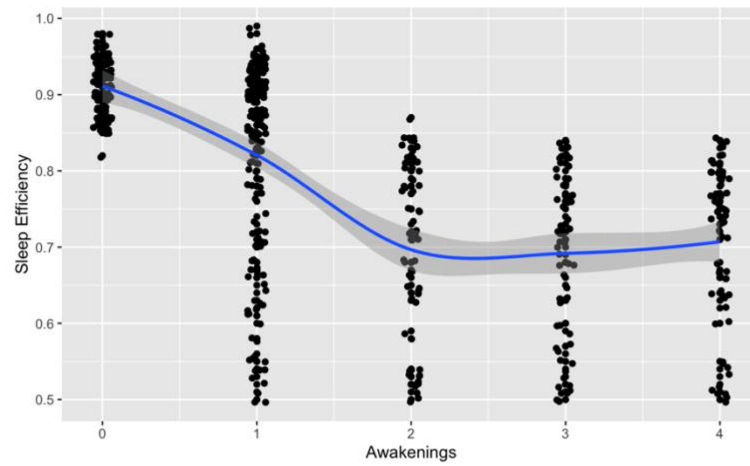
Ref. Figure 23: Deep sleep percentage vs Sleep Efficiency



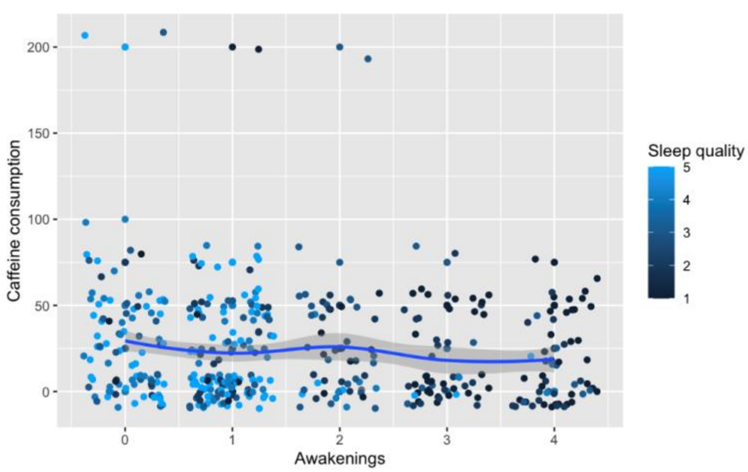
Ref. Figure 24: Deep sleep percentage vs Sleep Quality



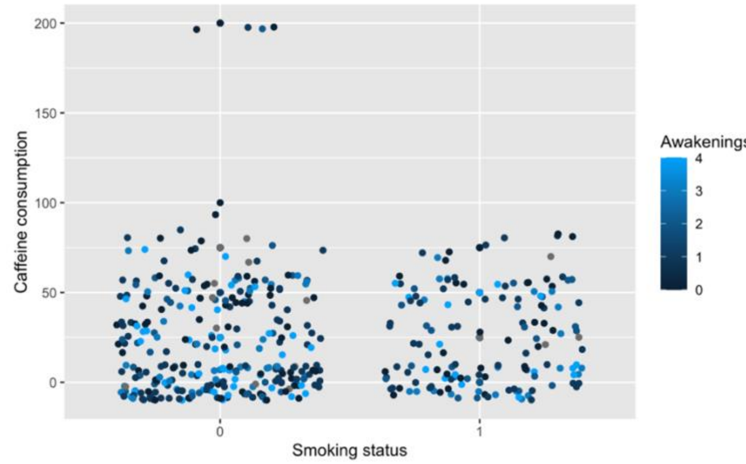
Ref. Figure 25: Gender vs Sleep Efficiency mosaic plot



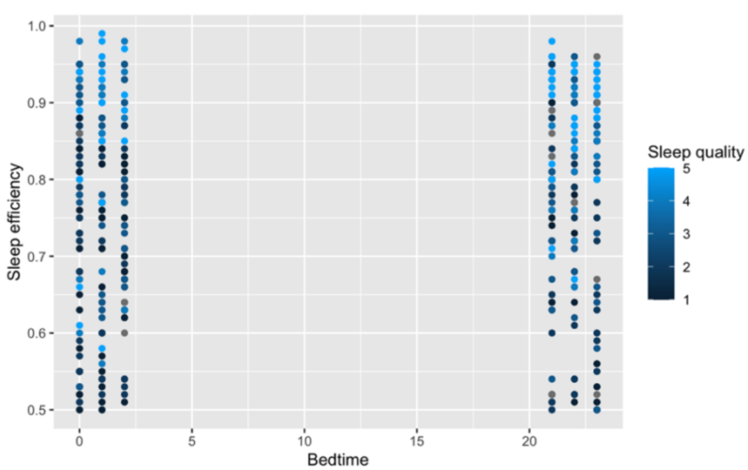
Ref. Figure 26: Awakenings vs Sleep Efficiency



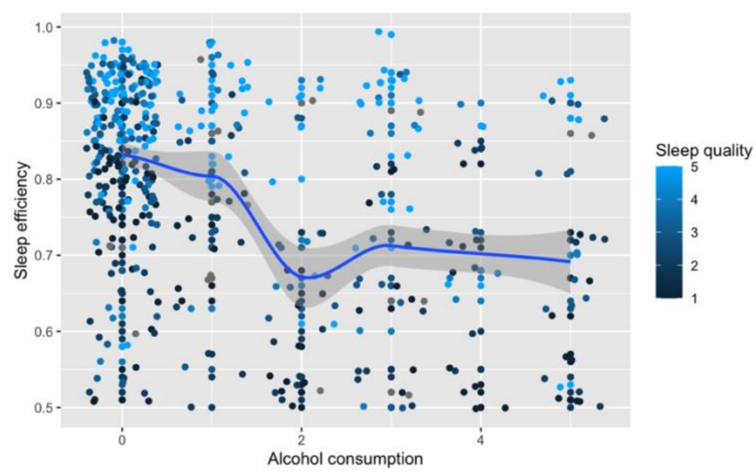
Ref. Figure 27: Awakenings vs Caffeine consumption



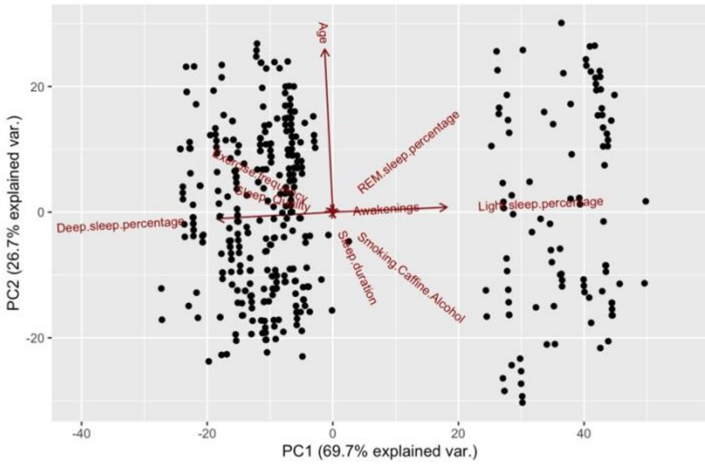
Ref. Figure 28: Smoking status vs Caffeine consumption



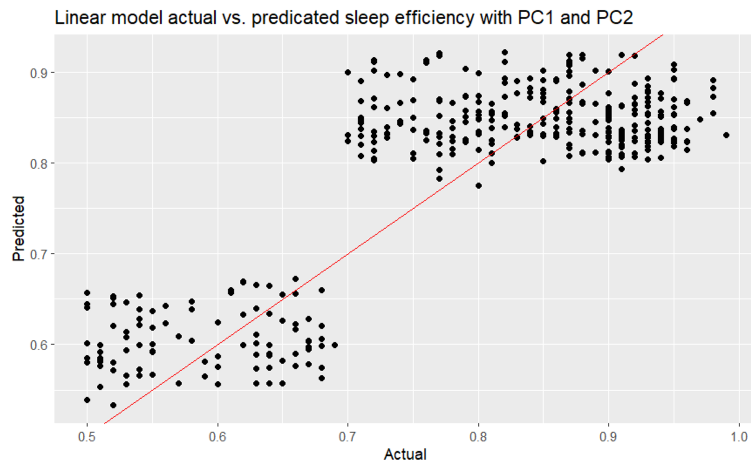
Ref. Figure 29: Bedtime vs Sleep Efficiency



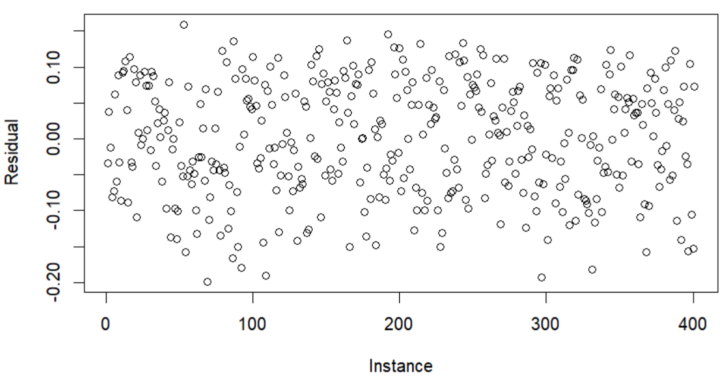
Ref. Figure 30: Alcohol consumption vs Sleep Efficiency



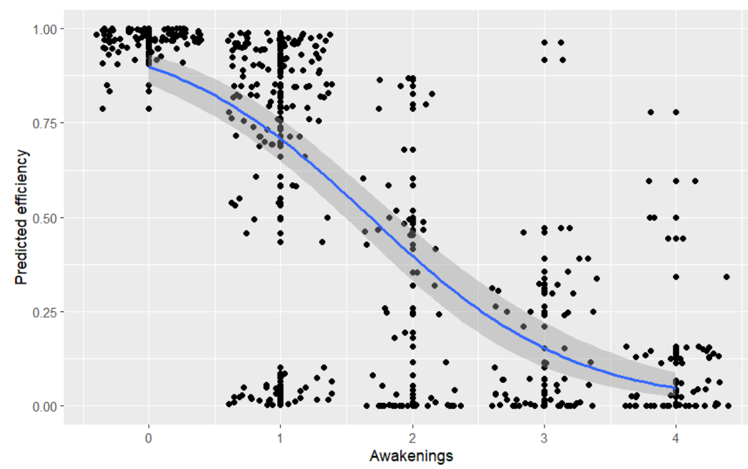
Ref. Figure31: PCA biplot



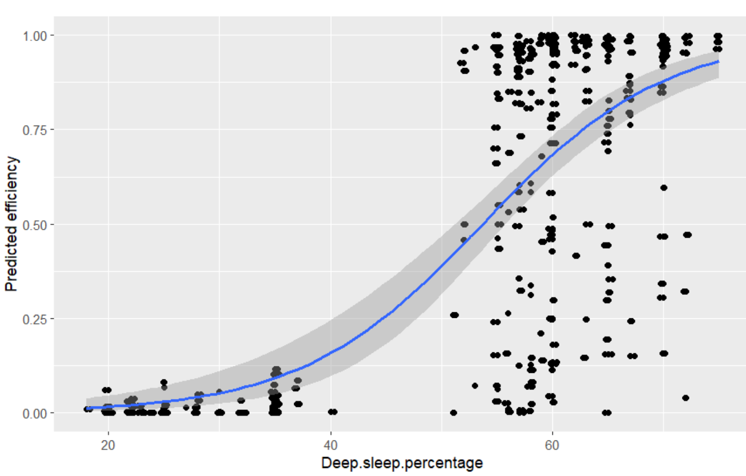
Ref. Figure 32: Linear model with PC1 and PC2



Ref. Figure 33: Residuals Linear model with PC1 and PC2



Ref. Figure 34: Awakenings vs Predicted Efficiency (Logistic Regression)



Ref. Figure 35: Deep Sleep percentage vs Predicted Efficiency (Logistic Regression)