# **Motivation:** We want to understand what stresses out engineering students. Engineering programs are tough with long study hours, heavy workloads and high-pressure exams. These demands can make students feel overwhelmed and stressed, affecting how well they do in school and how they feel overall.

# We believe that by looking at things like how well students sleep, how much they study, if they get headaches, and how much they do outside of school, we can figure out what stresses them the most. This will help us find patterns and understand what causes the most stress. If we can figure this out, we might find ways to help students deal with stress better. For example, if we find that sleep or study load is a big problem, schools could help students manage their time better or encourage healthier habits. This could lead to better grades, better health, and less burnout for students.

**Data Description:** For our project, we collected data about what might cause stress for engineering students. We’ll look at things like sleep quality, headaches, how students feel about their academic performance, their study load, and how much they get involved in extracurricular activities. Every variable is rated from 1-5.

• Sleep Quality (1 = poor, 5 = excellent) : categorical

• Headaches (how often they happen in a week) : continuous

• Academic Performance (1 = poor, 5 = excellent) : categorical

• Study Load (1 = very light, 5 = very heavy) : categorical

• Extracurricular Activities (How many times a week) : continuous

• Stress Levels (1 = very low, 5 = very high) : categorical

Each of these measures shows a different part of student life that could relate to stress, and we believe these will help us understand what factors make students feel more stressed.

**Exploratory Data Analysis (EDA):** The EDA includes looking at the distributions, correlations, and patterns in these variables to see how each might relate to stress levels.

1. Useful variables : we expect that all variables mentioned ( sleep quality, headaches, academic performance,

study load , and extracurricular activities) will be useful for analyzing stress levels. Each factors represent

a potential stressor and we believe that sleep quality and study load might be particularly strong indicator of

stress.

R code: summary(stress\_data)

Interpretation: The summary statistics help us understand the data.

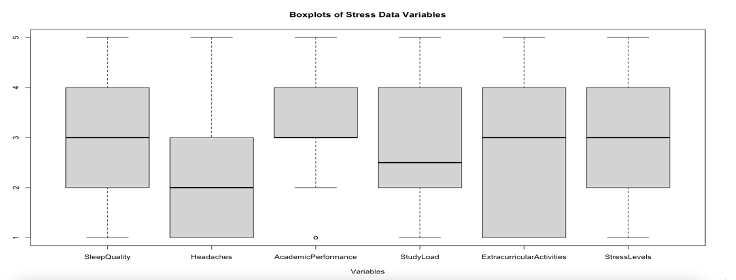
- Sleep Quality, Academic Performance, Study Load, and Stress Levels have values from 1 to 5, showing a mix of responses.

- Headaches and Extracurricular Activities also show a range of answers, meaning students experience different levels. Based on what we see, Sleep Quality and Study Load could be the most important factors related to stress levels.

2. High-order Terms or Transformations: At this stage, we don’t think we will need to use higher-order terms or transformations since the data is already on scale 1-5. However, if non-linear relationships are identified during exploratory data analysis (EDA), we may consider applying transformations to better capture non-linear effects.

3. Outliers: We used boxplots to spot any outliers in the data. We’ll check if these outliers are errors or just show different stress levels in students.

R code: boxplot(stress\_data)



The boxplot shows that most variables don’t have outliers, so the data looks normal. Only Academic Performance has one outlier which is a low score that’s different from the others. The other variables, like Sleep Quality, Headaches, Study Load, and Stress Levels, look pretty consistent without any outliers.

4. Assumptions: For this analysis, we assume that each factor is independent, there’s a linear relationship between factors and stress, and that residuals are normally distributed. If these assumptions are not met, we may need to adjust our methods.

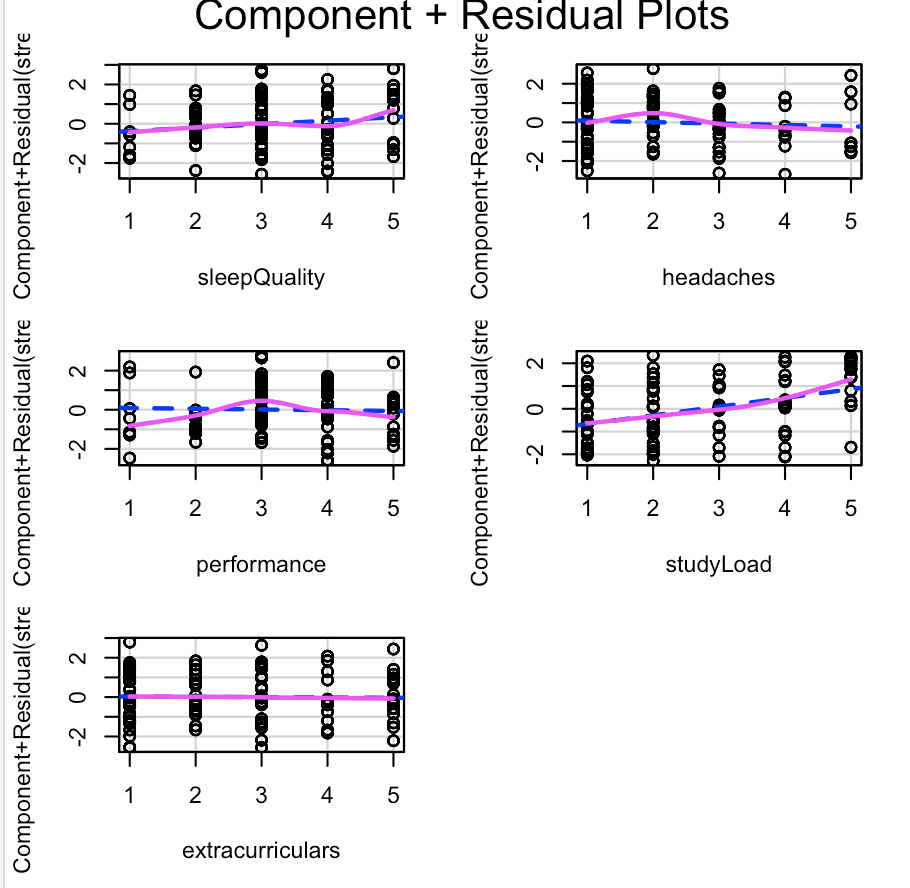
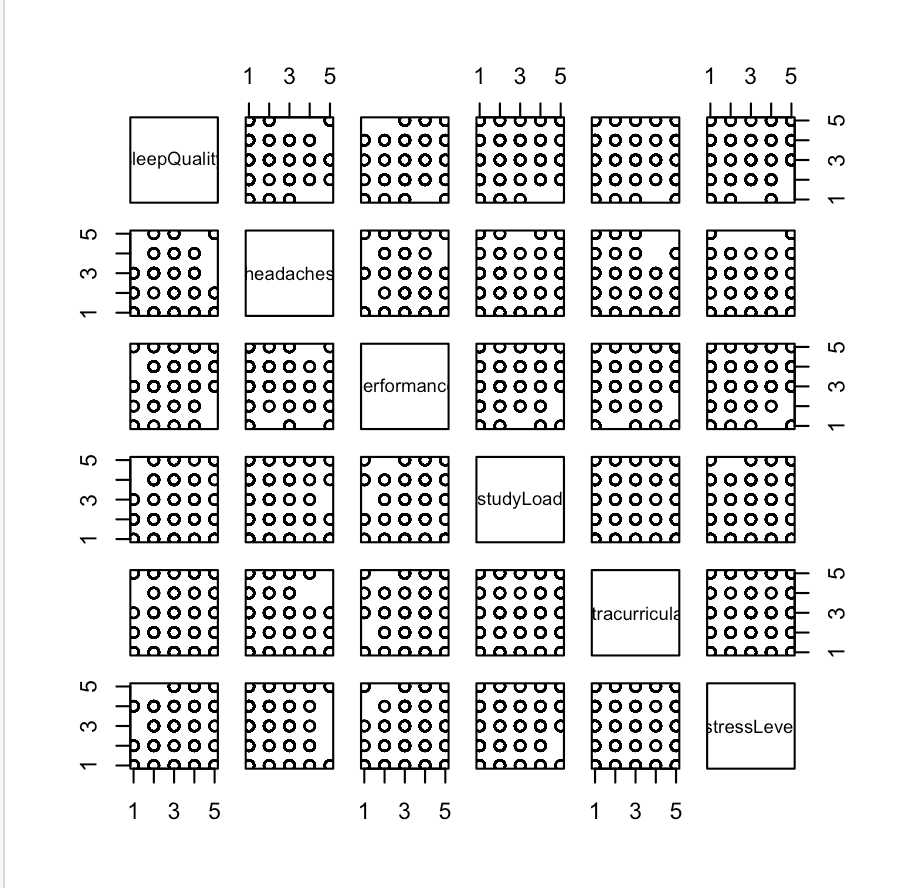
We can test our assumptions of linearity with R using basic scatter plots with our full model, as well as component and residual plots.

R Code:

pairs(stress\_data)

full\_model <- lm(stressLevel ~ ., data = stress\_data)

crPlots(full\_model)



The component and residual plots show how each factor relates to stress levels.The pink lines for sleep quality, study load, and extracurriculars are mostly straight, meaning the linearity assumption looks good for these factors. However, headaches and performance show some curves,suggesting the relationship might not be fully linear for these variables.

5. Collinearity: We checked for collinearity in our dataset by calculating the correlation matrix. A high correlation of 0.98 was found between SleepQuality and SleepQuality\_sq, showing potential collinearity. Most other variables showed low correlations with the main predictors. To avoid collinearity issues, we may need to consider removing one of the highly correlated variables.

R Code:

# Calculate the correlation matrix

cor(stress\_data)

# Pairs plot to visualize relationships

pairs(stress\_data)

6. Interaction term: We may include combinations of variables to see if together they affect stress in a stronger way. For example, combining study load with sleep quality may show a bigger impact on stress.In this model, we added interactions between study load and sleep quality, and between headaches and extracurricular activities.

Code:

# Model with multiple interaction terms

model\_multiple\_interactions <- lm(StressLevels ~ StudyLoad \* SleepQuality + Headaches \* ExtracurricularActivities, data=stress\_data)

# Summary of the model

summary(model\_multiple\_interactions)

The output shows that the interaction between headaches and extracurricular activities has a strong effect on stress (p < 0.001). The interaction between study load and sleep quality has a smaller effect. These interactions let us see how some factors together can influence stress more than each one alone.

7.Training / Testing split : We’ll split the data before modeling, using only the training data for EDA.

**Future Direction:** Afterwe finish exploring the data, we will use linear regression to figure out which factors have the biggest impact on stress. This will help us see which things make students the most stressed. For example, if sleep quality is the biggest factor, schools could encourage students to sleep better. If study load is the biggest stressor, schools could think about reducing it.

***We, the project teams members, certify that below is an accurate account of the percentage of effort contributed by each team member in the project and report.***

| Project Team Member | Percentage of total effort |
| --- | --- |
| Alexandra Ramlogan Salgado | 50% |
| Munchootsorn Wangsriviroj | 50% |

Work Cited

Bhansali, Samyak. “Student stress factors.” *Student stress factors*, Kaggle, 2024, https://www.kaggle.com/datasets/samyakb/student-stress-factors/data. Accessed 24 October 2024.