CBSC 185 Final Project

Naka Assoumatine

Import the class data

```
summary(class_data_subset)
```

```
##   StartDate              ID                Q33                Q9_1
##  Length:276         Length:276         Length:276         Length:276
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
glimpse(class_data_subset)
```

```
## Rows: 276
## Columns: 4
## $ StartDate <chr> "9/26/2024 13:52", "9/26/2024 13:52", "9/26/2024 14:03", "9/~
## $ ID        <chr> "100", "100", "100", "10", "17", "5", "11", "16", "4", "8", ~
## $ Q33       <chr> "", "5", "5", "", "6", "2", "13", "14", "7", "6", "13", "6",~
## $ Q9_1      <chr> "", "", "", "", "7", "7", "11", "8", "7.5", "8", "7", "6", "~
```

Interpretation: The initial data consists of four variables: StartDate (survey date), ID (student identifier), Q33 (mood ratings), and Q9_1 (hours of sleep). At this stage, the data is in character format and needs further processing to be analyzed effectively.

Create a subset of the subdata with Id, Survey date mood, and hours of sleep

```
glimpse(class_data_subset)
```

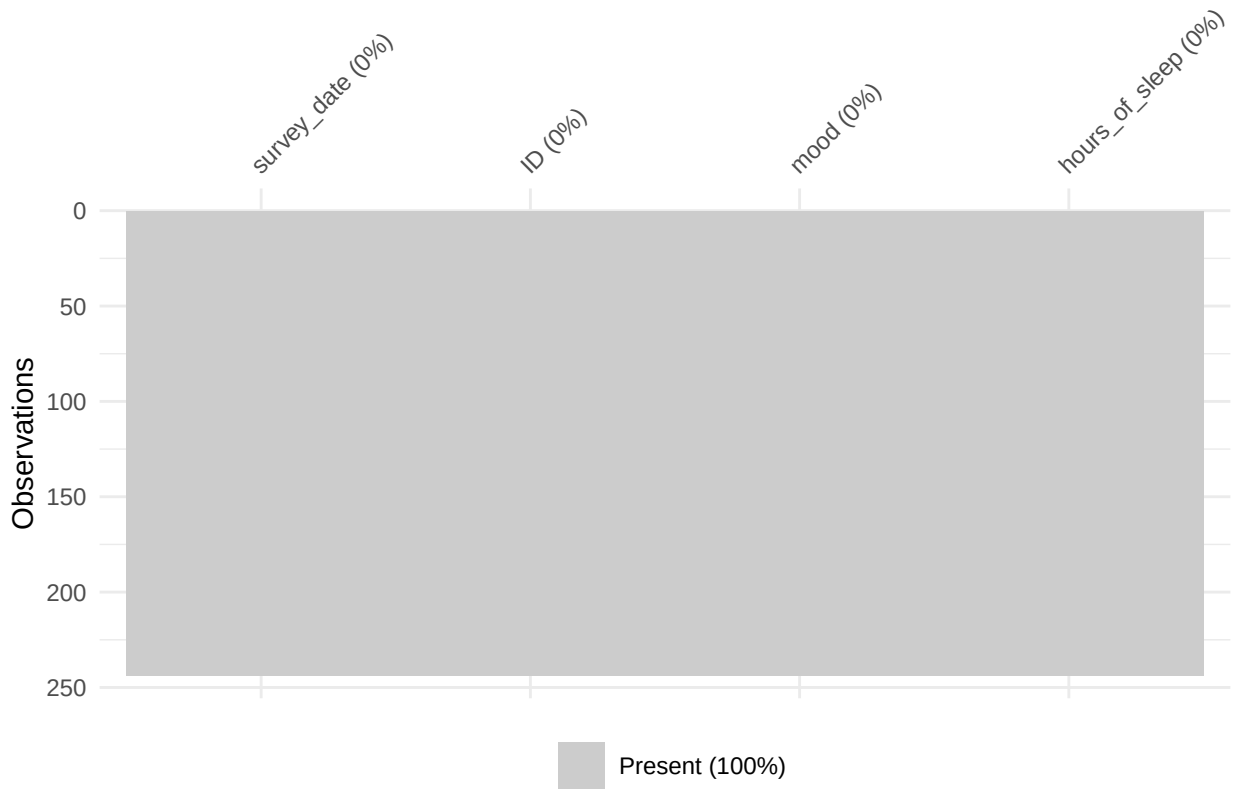
```
## Rows: 273
## Columns: 4
## $ survey_date    <date> 2024-09-26, 2024-09-26, 2024-09-26, 2024-09-26, 2024-0~
## $ ID             <chr> "10", "17", "5", "11", "16", "4", "8", "6", "13", "7", ~
## $ mood           <chr> "", "6", "2", "13", "14", "7", "6", "13", "6", "6", "5"~
## $ hours_of_sleep <chr> "", "7", "7", "11", "8", "7.5", "8", "7", "6", "7", "3"~
```

Interpretation: After initial data cleaning, we've filtered out placeholder IDs (value "100") and renamed variables to be more intuitive and descriptive. The survey_date variable has been converted to a proper date format, stripping out the time component. This cleaned dataset gives us a foundation for investigating how sleep duration relates to student mood throughout the semester.

Visualize missing data

Interpretation: The visualization reveals that approximately 11% of the hours_of_sleep data is missing, while the other variables (survey_date, ID, and mood) are complete. This missing data pattern suggests that participants consistently reported their mood but sometimes failed to record their sleep duration. Since we're investigating the relationship between sleep and mood, we'll need to address these missing values before analysis.

Visualize missing data after removing missing data



Interpretation: After removing observations with missing values, our dataset is now complete with no missing values. While this ensures clean analysis, it's important to note that removing data points could potentially introduce bias if the missing sleep data isn't randomly distributed. However, with a strong remaining sample size, we can proceed with our analysis.

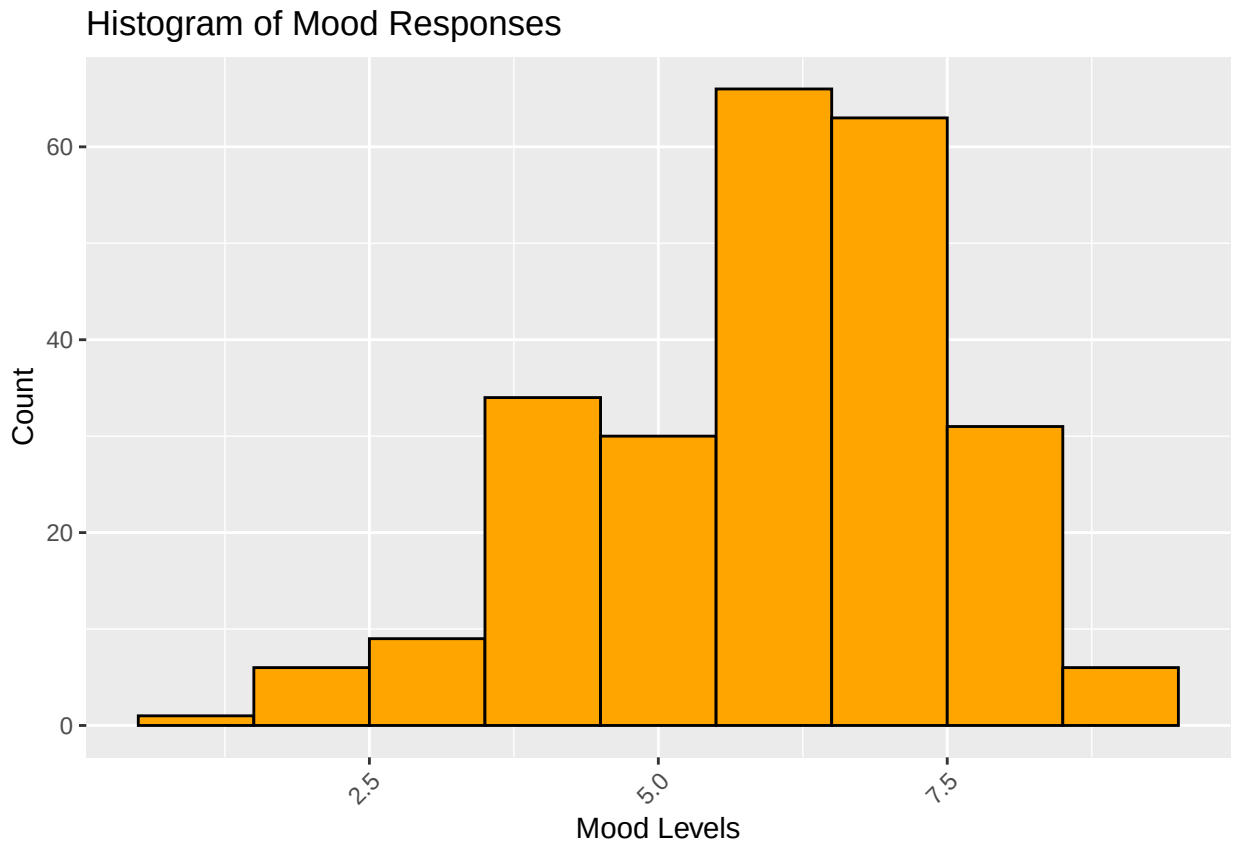
Boxplot hours of sleep

Interpretation: The initial boxplot reveals a significant outlier in sleep hours - a value around 530 hours which is clearly unrealistic. This is likely a data entry error where a student may have entered "5:30" (5.5 hours) incorrectly as "530". This extreme outlier is distorting our visualization and needs to be corrected before we can properly assess the sleep duration distribution.

Boxplot hours of sleep after hardcoding the outlier 530

Interpretation: After correcting the outlier (converting 530 to 5.5 hours), the boxplot now shows a more realistic distribution of sleep durations. Most students report sleeping between approximately 5 to 10 hours per night, with a median around 7 hours. There are still a few outliers on both the lower and upper ends, but these appear to be within plausible ranges (e.g., some students reporting very little sleep before exams or extended sleep on weekends). This distribution aligns with typical sleep patterns for college students.

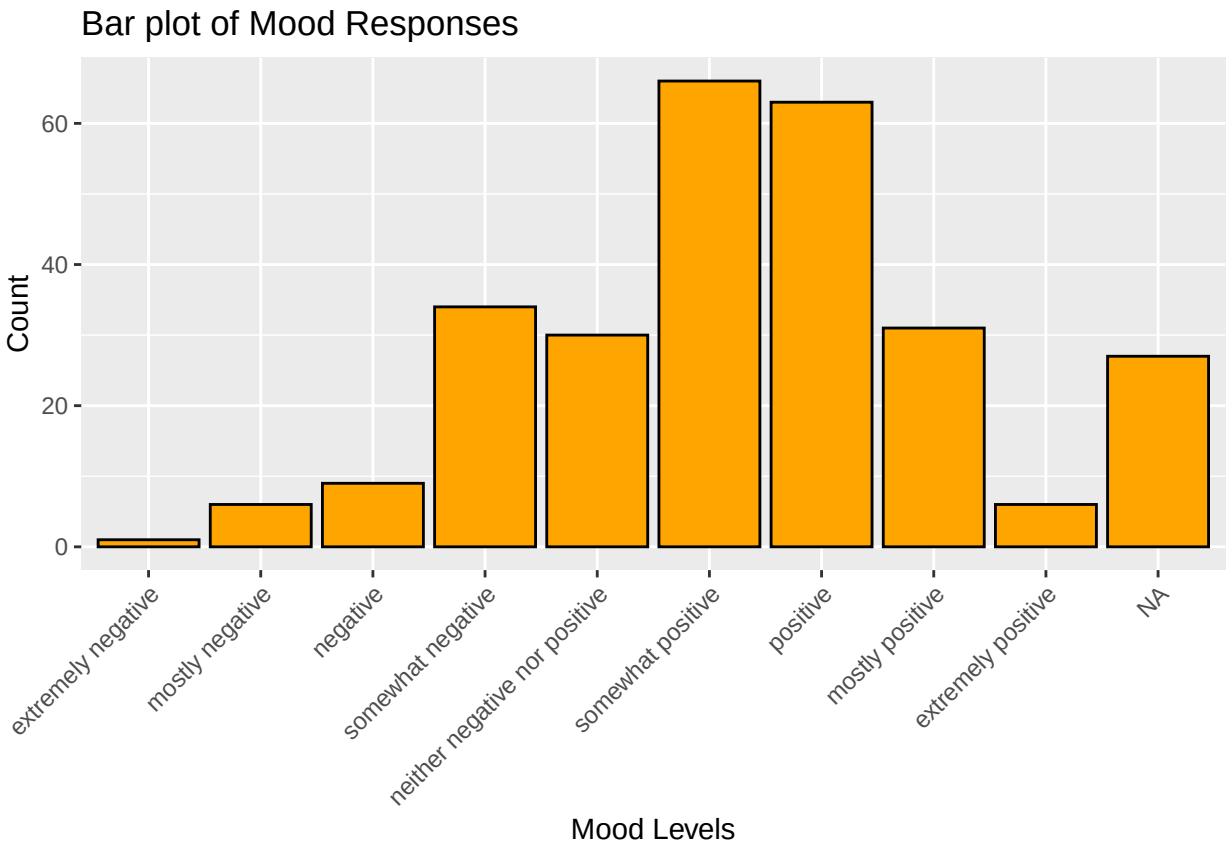
Histogram of mood



Interpretation: The histogram of mood values shows that student responses tend to cluster toward the positive end of the scale. There's a notable peak around values 6-7, indicating that most students report feeling "somewhat positive" to "positive" on the mood scale. There are relatively fewer responses at the extreme ends (very negative or extremely positive). This right-skewed distribution is common in subjective well-being scales, as people generally tend to report positive rather than negative emotional states when asked about their mood.

Bar plot of mood

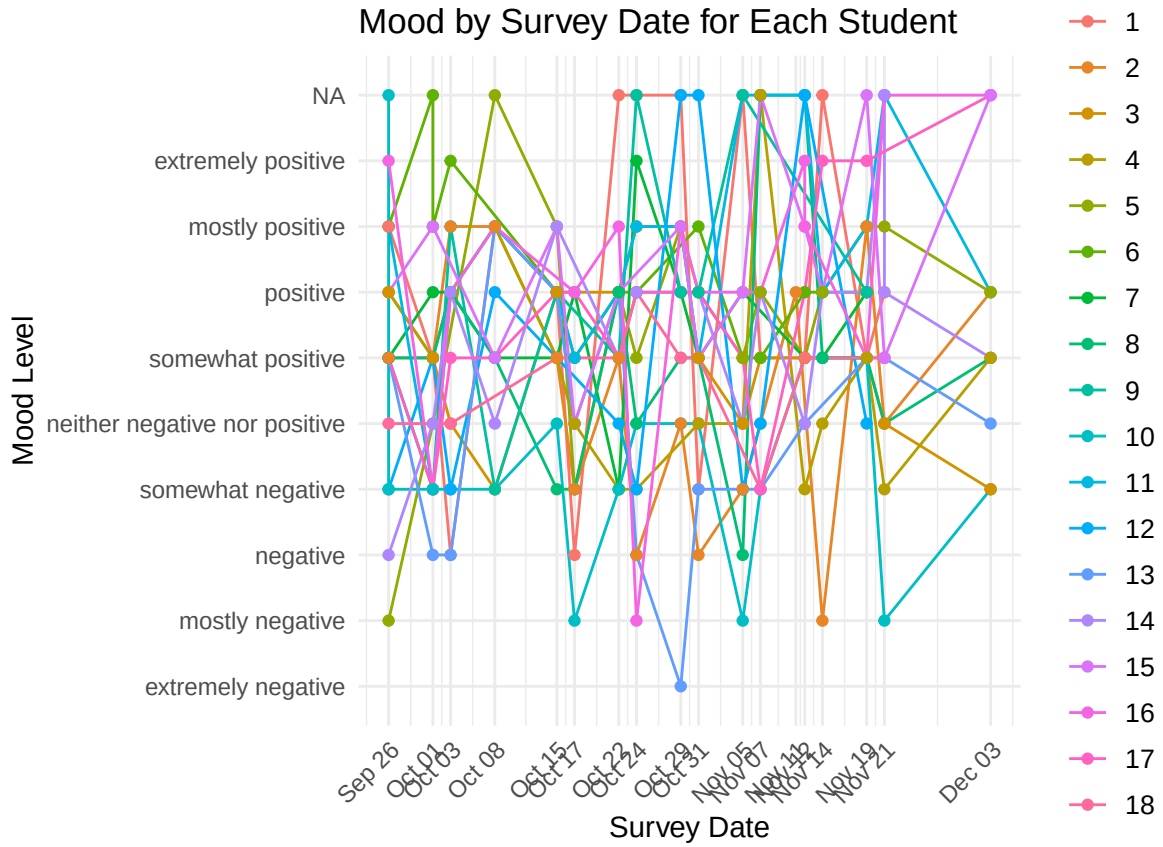
To allow the reader to understand what the mood values are, I will create a bar plot with the values of the mood levels on the x-axis. This can be seen on the graph below:



Interpretation: This bar plot presents the same data as the histogram but with labeled mood categories, providing better context for interpretation. We can now clearly see that "positive" and "somewhat positive" are the most frequently reported mood states among students. Interestingly, there are very few reports of "extremely negative" or "mostly negative" moods, which suggests that severe negative emotional states were uncommon during the study period. The prevalence of positive mood states may reflect either genuine well-being among the student population or potential response bias where students are reluctant to report strong negative emotions.

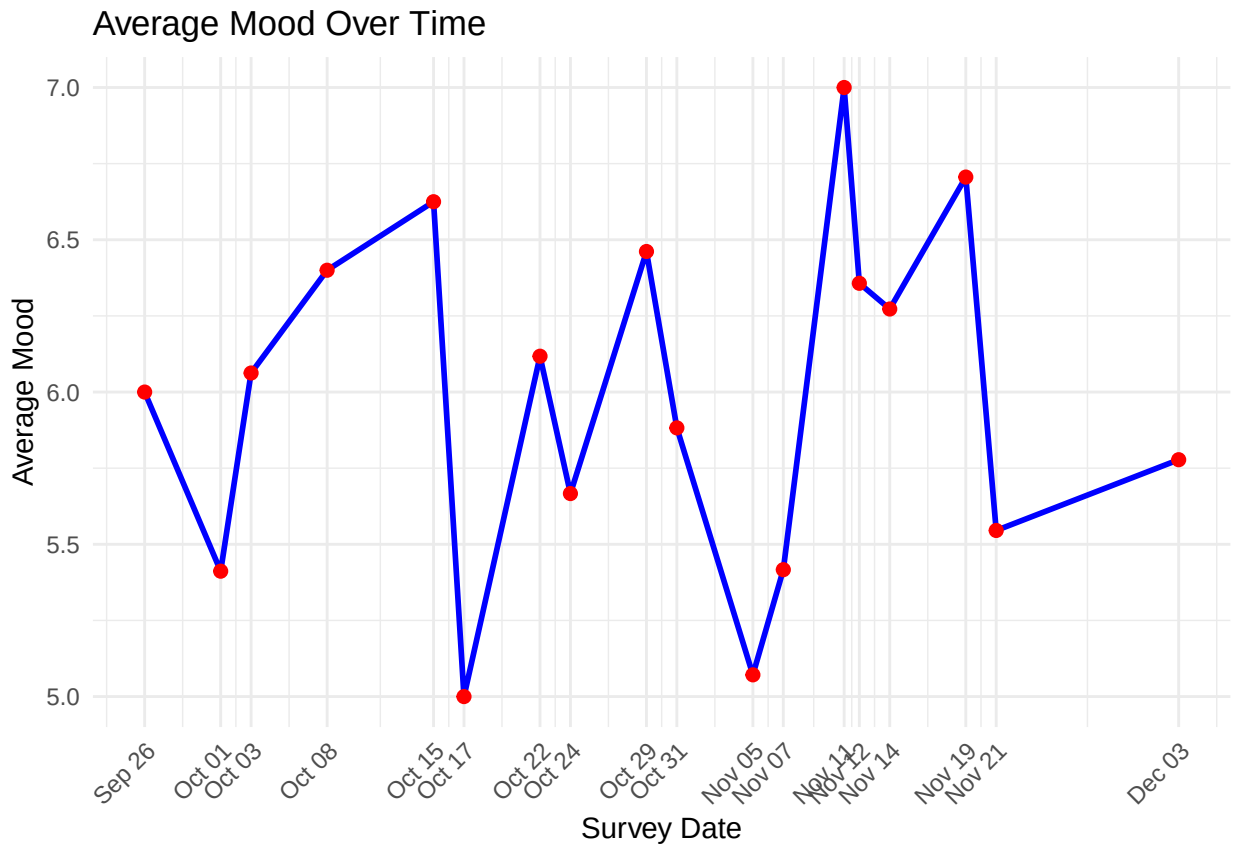
Mood by Survey Date for Each Student

This will give an insight of the trend of the mood for all students



Interpretation: This visualization tracks individual students' mood fluctuations throughout the semester (September through December). We can observe significant variability both between students and within individual students over time. Some students display relatively stable mood patterns, while others show considerable fluctuations. There appears to be no universal pattern across all students, highlighting the highly individualized nature of emotional experiences. This suggests that while we may find general trends between sleep and mood in our population, personal factors likely play a significant role in determining individual emotional states, which is why incorporating student ID as a random effect in our regression model will be important.

Overall average mood overtime



Interpretation: The average mood across all students shows moderate fluctuation throughout the semester, typically ranging between 5.0 and 7.0 on our 9-point scale (corresponding to "neither negative nor positive" to "positive"). While there are some noticeable ups and downs, there's no clear seasonal pattern or significant trend over time. This suggests that, on average, the collective mood of the class remained relatively stable throughout the academic period. Interestingly, we don't see the expected decline in mood during traditional high-stress periods (like midterms or finals week), which might indicate that other factors, like adequate sleep, could be buffering against academic stress.

Regression model

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mood ~ hours_of_sleep + survey_date + (1 | ID)
##    Data: class_data_subset
##
##      AIC      BIC   logLik deviance df.resid
##    885.0    902.5   -437.5    875.0      238
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.03318 -0.48463  0.08805  0.60110  2.13365
##
## Random effects:
```

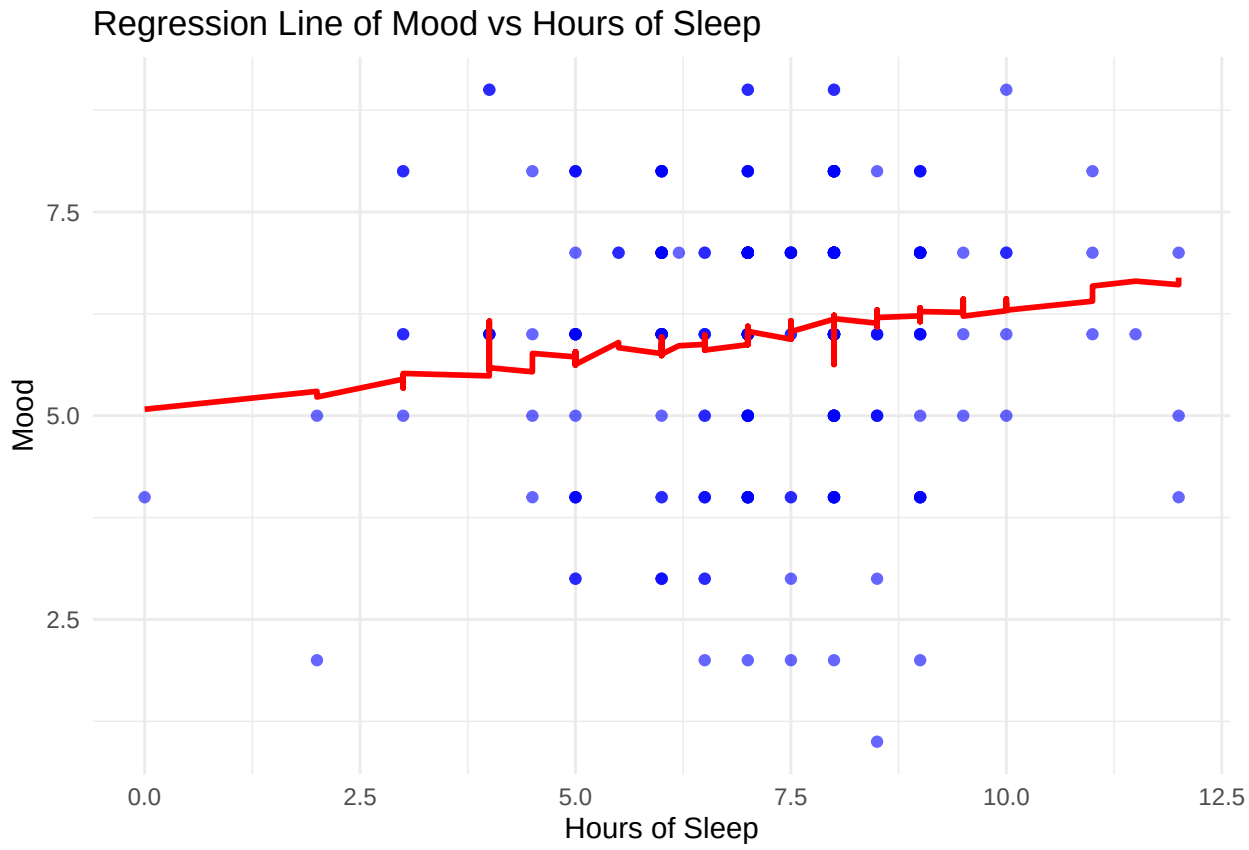
```
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 0.5598   0.7482
##  Residual             1.9062   1.3807
## Number of obs: 243, groups:  ID, 18
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)    -61.526821  95.614159  -0.643
## hours_of_sleep   0.133239   0.060425   2.205
## survey_date      0.003325   0.004776   0.696
##
## Correlation of Fixed Effects:
##             (Intr) hrs_f_
## hours_f_slp  0.030
## survey_date -1.000 -0.034

## MODEL INFO:
## Observations: 243
## Dependent Variable: mood
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = 885.04, BIC = 902.50
## Pseudo-R² (fixed effects) = 0.02
## Pseudo-R² (total) = 0.24
##
## FIXED EFFECTS:
## ----------------------------------------------------------------
##                          Est.    S.E.   t val.     d.f.      p
## ------------------- -------- ------- -------- -------- ------
## (Intercept)           -61.53   95.61    -0.64   227.39   0.52
## hours_of_sleep          0.13    0.06     2.21   240.30   0.03
## survey_date             0.00    0.00     0.70   227.41   0.49
## ----------------------------------------------------------------
##
## p values calculated using Satterthwaite d.f.
##
## RANDOM EFFECTS:
## -------------------------------------
##   Group      Parameter    Std. Dev.
## ---------- ------------- -----------
##     ID      (Intercept)     0.75
##  Residual                   1.38
## -------------------------------------
##
## Grouping variables:
## -------------------------
##  Group   # groups   ICC
## ------- ---------- ------
##   ID        18      0.23
## -------------------------
```

Interpretation: Our linear mixed effects model examines how hours of sleep and survey date predict student mood while accounting for individual differences:

1. **Fixed Effects**:
    - **Hours of Sleep**: There is a statistically significant positive relationship between sleep duration and mood (= 0.133, p = 0.03). For each additional hour of sleep, mood increases by approximately 0.13 points on our 9-point scale. While this effect is modest in magnitude, it's statistically significant and supports our hypothesis that more sleep is associated with more positive mood states.
    - **Survey Date**: The survey date does not have a significant effect on mood (p = 0.49), suggesting that mood doesn't systematically change over the course of the semester when controlling for sleep duration and individual differences.
2. **Random Effects**:
    - The random effect for student ID accounts for individual differences in baseline mood. The variance of 0.56 indicates moderate variability between students.
    - The ICC (Intraclass Correlation Coefficient) of 0.23 suggests that approximately 23% of the variation in mood can be attributed to individual differences between students.
3. **Model Fit**:
    - The model explains about 24% of the total variance in mood (Pseudo-R^2 = 0.24), which is a moderate level of explanatory power for behavioral data.
    - The fixed effects alone (sleep hours and survey date) account for only 2% of the variance (Pseudo-R^2 fixed effects = 0.02), highlighting the importance of accounting for individual differences in this type of analysis.

Overall, these results support our hypothesis that more sleep is associated with better mood, though the relationship is relatively modest. Individual differences play a substantial role in determining mood states.

Regression line of mood vs hours of sleep



Interpretation: This visualization illustrates the relationship between hours of sleep and mood, with the red regression line showing the model's predicted values. The blue points represent the actual observations, which show considerable variability around the regression line. The positive slope of the regression line confirms our statistical finding: as sleep duration increases, mood tends to improve. However, the wide scatter of points around the line highlights that sleep is just one of many factors influencing mood - individual differences, daily events, and other unmeasured variables likely account for much of the observed variation. Nevertheless, the clear upward trend supports the importance of adequate sleep for emotional well-being among students.

Overall Conclusions

Based on our analysis of 243 observations from 18 students, we found evidence supporting our hypothesis that sleep duration positively affects mood. The key findings include:

1. Sleep duration has a small but statistically significant positive effect on mood (= 0.133, p = 0.03), with each additional hour of sleep associated with a 0.133-point increase on our 9-point mood scale.
2. Individual differences account for approximately 23% of the variation in mood (ICC = 0.23), highlighting the importance of considering personal factors when studying emotional states.
3. There was no systematic change in mood over the course of the semester, suggesting that temporal factors played a minimal role compared to sleep and individual differences.

4. The overall model explained about 24% of the variance in mood, indicating that while sleep is an important factor, other unmeasured variables also substantially influence emotional states.

These findings align with existing research on the relationship between sleep and psychological well-being, supporting the recommendation that adequate sleep is an important component of maintaining positive emotional states for college students. Future research might explore additional factors that interact with sleep to influence mood, such as academic workload, physical activity, or social support.