



Student Stress Analysis

April Gao, Kelly Wang, Xinyi Lyu, Rohit Chandiramani, Saikripa Mohan

Background

84.4%

University students
have experienced
stress



The University Student Mental Health data was gathered from the Behavioural Research on Aging and Illness in Neuropsychology (BRAIN) Lab at the University of Victoria during the fall of 2020.



This analysis will focus on Stress Levels as the output indicator. Stress indicator, ranging from 1 to 5, are derived from responses to 10 survey questions.



Data Preparation

1

Data Cleaning

- Removed empty rows and NULL values

2

Exploratory Data Analysis

- Visualize distribution via boxplots and histograms
- Correlation Heatmaps
- Forming Hypothesis

3

Data Transformation

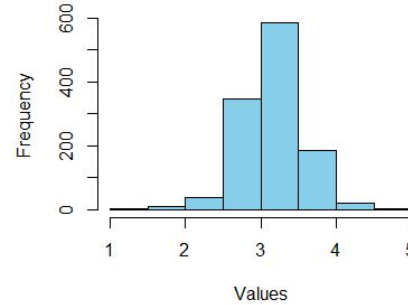
- Aggregate individual question responses to form unified Stress score
- Encode categorical features as factors
- Convert numerical stress score to binary outcome using threshold

3

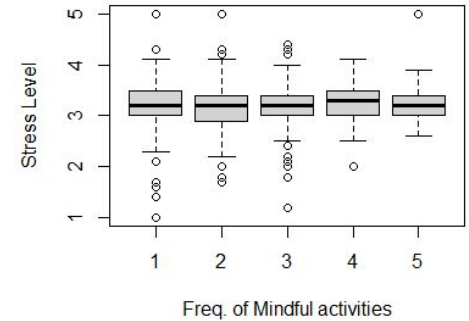
Feature Selection

- Extract relevant columns of interest and remove unwanted features for hypothesis

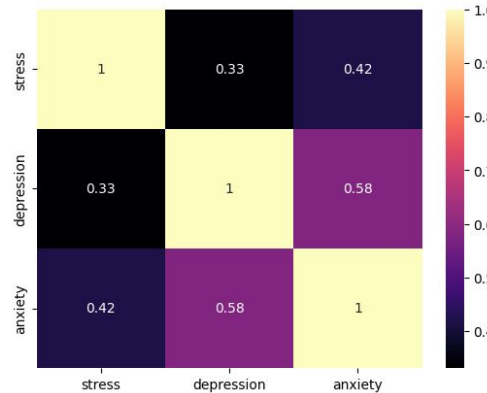
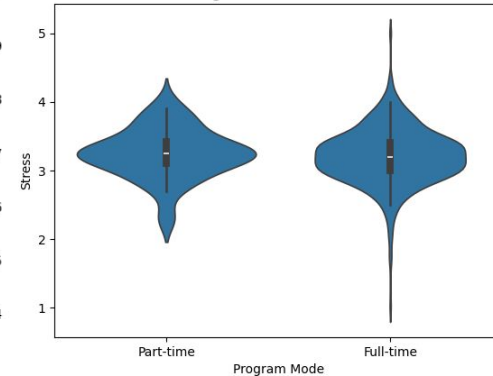
Histogram of Stress Levels



Stress Level by Mindful Activities

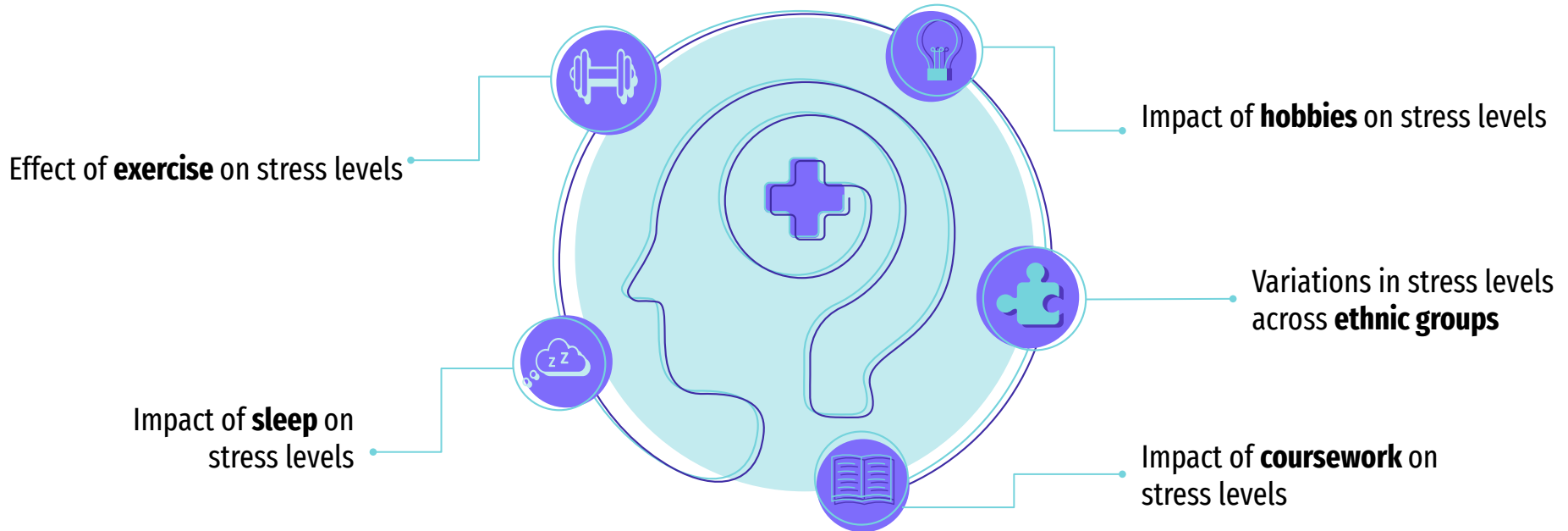


Program Mode vs Stress



Project Goal

- Identify major characteristics and lifestyles of students experiencing stress



Question 1: Exercise impact on stress



Frequency/Intensity of exercise

Frequency ranged from 0-28 for

- Strenuous
- Moderate
- Mild

Data Preparation

- Excluded responses that are greater than 10
- Created total exercise column summing all 3 intensity levels
- Introduced binary variable: 'exercise' indicating presence or absence of any exercise
- Introduced binary stress indicator for above or below average

| | Strenuous.exercise | Moderate.exercise | Mild.exercise | stress |
|--------------------|--------------------|-------------------|---------------|--------------|
| Strenuous.exercise | 1.00000000 | 0.3106549 | 0.119978677 | -0.087982888 |
| Moderate.exercise | 0.31065490 | 1.00000000 | 0.361519804 | 0.018389001 |
| Mild.exercise | 0.11997868 | 0.3615198 | 1.000000000 | 0.003705613 |
| stress | -0.08798289 | 0.0183890 | 0.003705613 | 1.000000000 |

| | total_exercise | stress |
|----------------|----------------|-------------|
| total_exercise | 1.00000000 | -0.02595188 |
| stress | -0.02595188 | 1.00000000 |

Test Performed: ANOVA

Hypothesis: H0: There is no difference in numeric stress levels between students who exercise and those who do not.

Result: p-value equal to 0.685, we fail to reject the null hypothesis

Test Performed: Logistic Regression

Hypothesis: H0: Exercise frequency does not affect the odds of students having stress levels above average.

Result: Each one-unit increase in Strenuous exercise is associated with an 11% decrease in the odds of experiencing stress levels above average

| | Estimate | Std. Error | z value | Pr(> z) |
|--------------------|----------|------------|---------|--------------|
| (Intercept) | 0.35724 | 0.10718 | 3.333 | 0.000859 *** |
| Strenuous.exercise | -0.11645 | 0.03128 | -3.723 | 0.000197 *** |
| Moderate.exercise | 0.04769 | 0.02983 | 1.599 | 0.109842 |
| Mild.exercise | -0.01415 | 0.02529 | -0.560 | 0.575666 |

Question 1: Exercise impact on stress

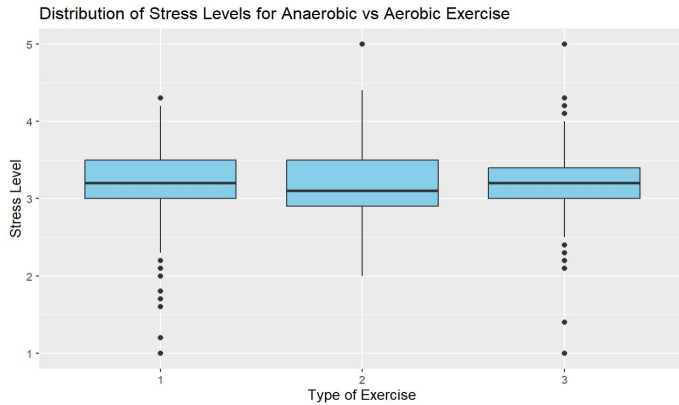


Type of exercise

- 1: Anaerobic
- 2: Aerobic
- 3: Both

Data Preparation

- Utilized both binary stress and numeric stress indicator



Test Performed: ANOVA

Hypothesis: H0: There is no difference in stress levels between students with different types of exercise.

Result: p-value equal to 0.297, we fail to reject the null hypothesis.

Test Performed: Chi Square

Hypothesis: H0: There is no association between exercise type and whether a student is above average stress level or not.

Result: p-value equal to 0.118, we fail to reject the null hypothesis.

Question 2: Variations in stress levels across ethnic groups

Test Performed: ANOVA

Hypothesis:

H_0 : mean stress levels across ethnic groups are equal

H_a : difference in stress levels between at least 2 ethnic groups

Result: P-value = 0.005, we reject the null hypothesis at the significance level of 0.05 and conclude that a significant difference exists between at least 2 ethnic groups.

| Ethnicity | Sample size |
|--------------------|-------------|
| Caucasian | 748 |
| Asian | 287 |
| Other | 108 |
| Black | 20 |
| Hispanic or Latino | 19 |
| Indigenous | 6 |
| Metis | 4 |

Question 2: Variations in stress levels across ethnic groups

Test Performed: Two Sample t-test

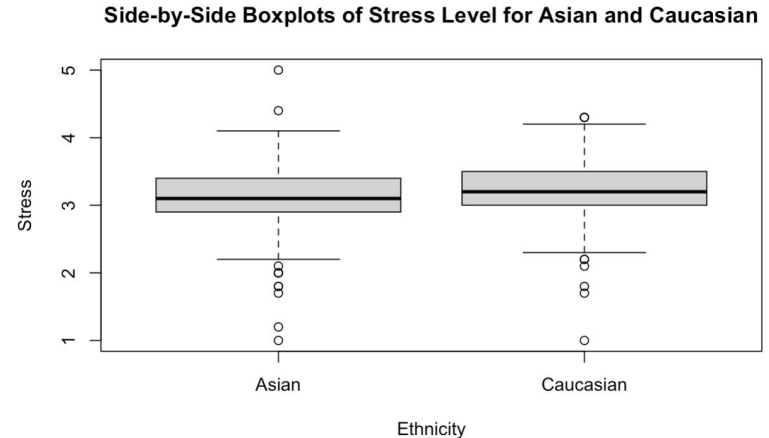
Hypothesis:

H_0 : mean stress levels between Caucasians and Asians are equal

H_a : difference in stress levels between the 2 ethnic groups

Result: P-value = 0.003, we reject the null hypothesis at the significance level of 0.05 and conclude that a significant difference exists between the two ethnic groups.

95% confidence level is (0.03, 0.15), suggesting that the mean stress level of Caucasians is expected to be between 0.15 and 0.03 higher than the mean stress level of Asians.



Question 3: Impact of coursework on student stress

Test Performed: ANOVA & Bootstrapped ANOVA

Hypothesis: H0: There is no difference in stress levels between students from different programs of study

Result: p-value equal to 0.2, Bootstrap CI : (-1.84, 2.38), we fail to reject the null hypothesis.

Test Performed: ANOVA & Bootstrapped ANOVA

Hypothesis: H0: There is no difference in stress levels between students from different academic years of study

Result: p-value equal to 0.21, Bootstrap CI : (-2.39, 3.26), we fail to reject the null hypothesis.

| Program of study | Sample size |
|------------------------|-------------|
| Sciences | 486 |
| Social Sciences | 294 |
| Engineering | 114 |
| Business | 113 |
| Humanities | 87 |
| Fine Arts | 36 |
| Education | 33 |
| Social Development | 29 |

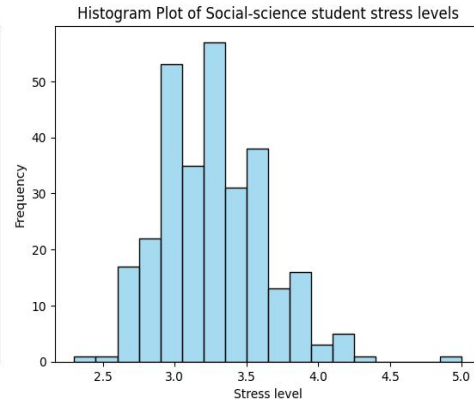
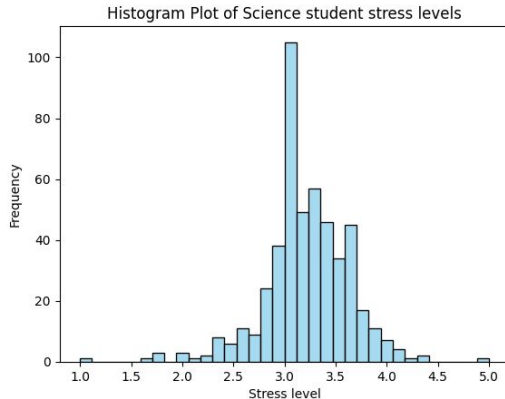
| Year | Sample size |
|-------------|-------------|
| Freshman | 325 |
| Sophomore | 320 |
| Junior | 287 |
| Senior | 217 |
| Victory Lap | 43 |

Question 3: Impact of coursework on student stress

Test Performed: Bootstrapped 2-sample t-test

Hypothesis: H0: There is no difference in proportion of stressed students studying Sciences and Social sciences

Result: Bootstrap CI : (0.1601, 0.3237), we reject the null hypothesis at the significance level of 0.05 and conclude that the proportion of stressed students studying Sciences and Social sciences is not the same



Test Performed: ANOVA (with interaction)

Hypothesis: H0: There is no difference in stress levels between students from different programs and years taken together

Result: Program:Year_credits P-value = 0.04, hence we Reject the null hypothesis that the means of different programs, across academic years interacting together are not same

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|----------------------|------|--------|---------|---------|------------|
| Program | 7 | 1.71 | 0.2443 | 1.433 | 0.18786 |
| Year_credits | 4 | 1.06 | 0.2638 | 1.548 | 0.18601 |
| Program:Year_credits | 28 | 8.89 | 0.3176 | 1.864 | 0.00426 ** |
| Residuals | 1152 | 196.30 | 0.1704 | | |

Question 4: Impact of sleep on student stress



Parameter – Hours_sleep

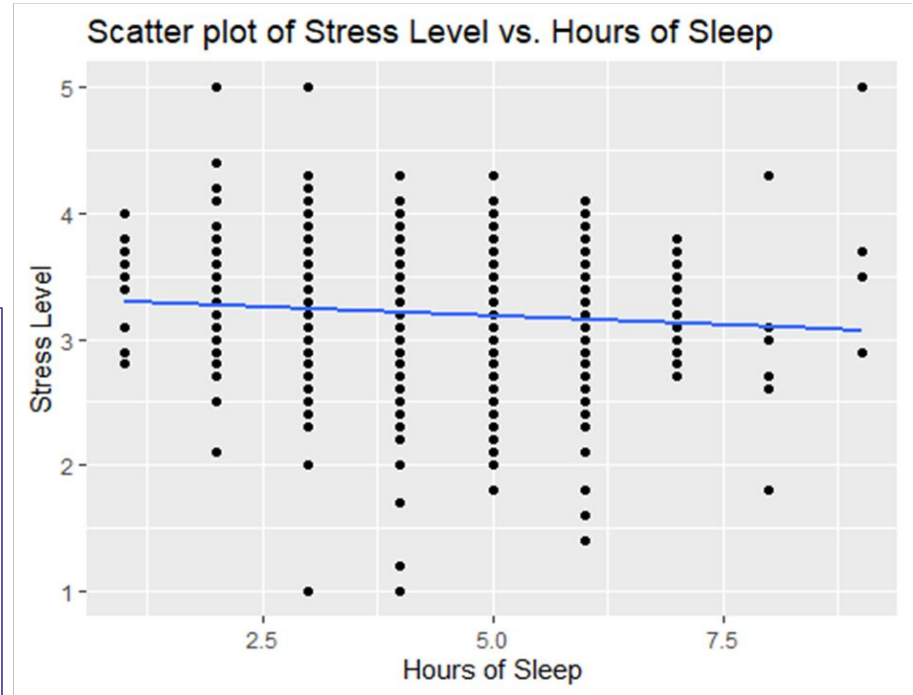
- number of sleeping hours on an average night
- scaled from 1 to 9

Test Performed: Pearson's Correlation

Hypothesis: H0: There is no correlation between hours of sleep and students' stress level. (correlation coefficient = 0)

Result: $r = -0.088$, $p\text{-value} = 0.0024 < 0.05$, we reject the null hypothesis.

There is a significant but weak negative correlation between the two variables.



Base Model: $\beta_0 + \beta_1 \text{Hours_Sleep}$

Question 4: Impact of sleep on student stress

Test Performed: Wald Test

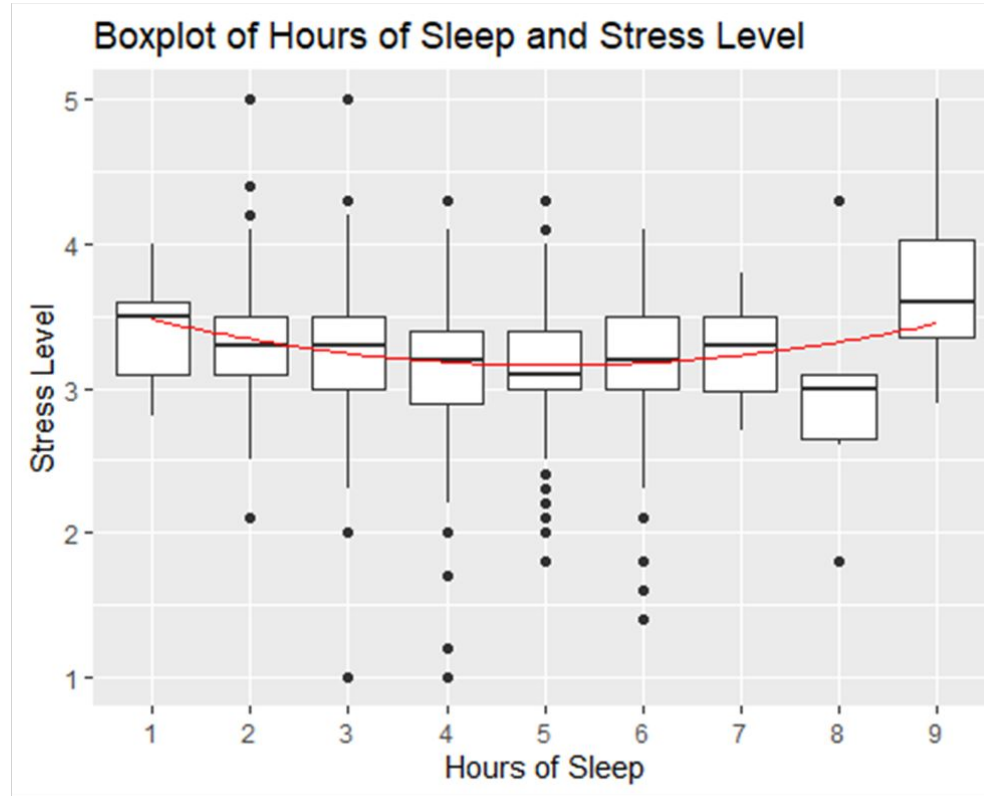
Polynomial Model:

$$\beta_0 + \beta_1 \text{Hours_Sleep} + \beta_2 \text{Hours_Sleep}^2$$

Hypothesis: $H_0: \beta_2 = 0$

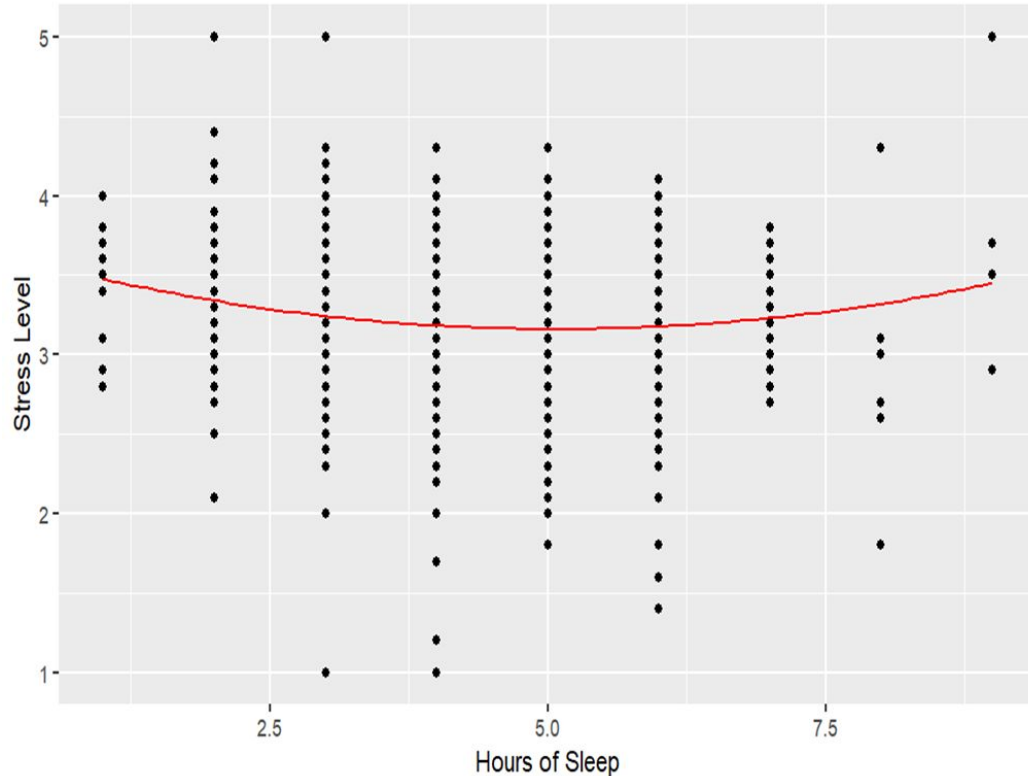
Result: p-value = 3.014e-06, we reject the null hypothesis. Thus, we can conclude that the polynomial regression model with the quadratic term provides a better fit to the data than the simple linear regression model.

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.651364   0.089868  40.630 < 2e-16 ***
## Hours_sleep   -0.192839   0.041683  -4.626 4.13e-06 ***
## I(Hours_sleep^2) 0.018947   0.004688   4.042 5.65e-05 ***
```



Question 4: Impact of sleep on student stress

Scatterplot of Stress Level vs. Hours of Sleep with Polynomial Regression Model



Result: Adjusted R-squared: 0.01951

The model only explains about 1.95% of the variability in stress level

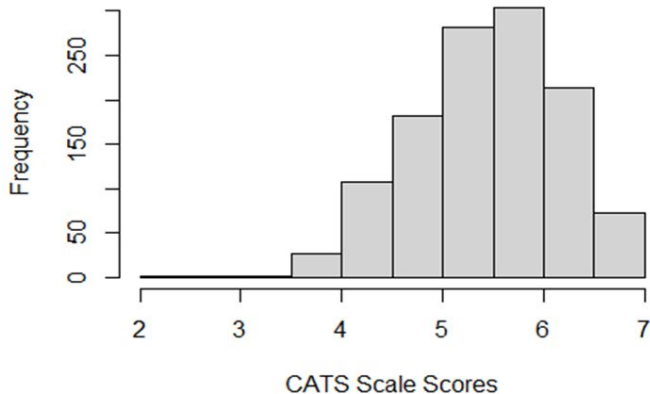
Question 4: Impact of sleep on student stress



Parameter – sleep (Attitudes)

- range from 1 to 7
- higher score indicates more favorable attitudes about sleep.

Distribution of CATS Scale Scores



Test Performed: Pearson's Correlation

Hypothesis: H0: There is no correlation between attitudes towards sleeping and students' stress level. (correlation coefficient = 0)

Result: $r = 0.1247027$, $p\text{-value} = 1.576e-05$, we reject the null hypothesis.

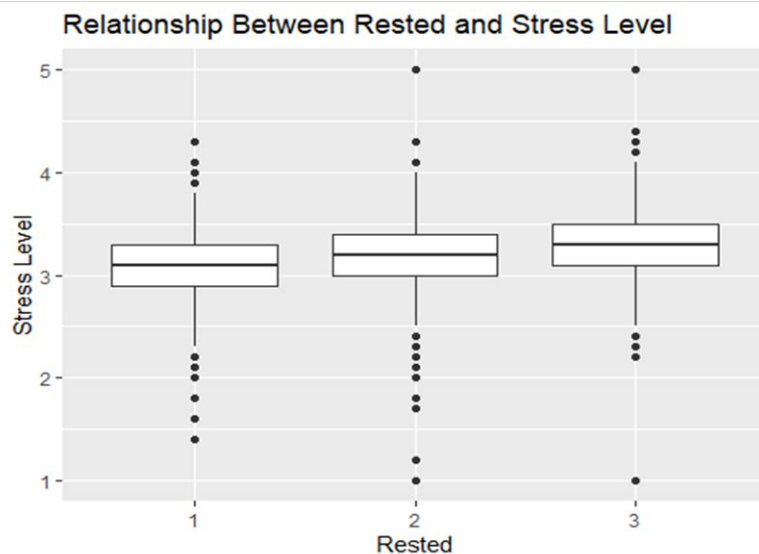
There is a significant but weak positive correlation between the two variables.

Question 4: Impact of sleep on student stress



Parameter - Rested

- encoded answers to whether participants feel well-rested
- 1 = 'yes', 2 = 'somewhat', 3 = 'no'



Test Performed: ANOVA

Hypothesis: H0: Rested variable has no effect on stress levels. $\mu_{yes} = \mu_{somewhat} = \mu_{no}$

Result: p-value = 2.2e-08, we reject the null hypothesis. The effect of "Rested" on stress levels is significant overall.

Test Performed: Pairwise t-test with Bonferroni adjust

Hypothesis: H0: There is no difference in mean stress levels between each pair of groups being compared.

Result: reject the null hypothesis

| | | |
|-------------|---------|----------|
| ## | Yes | Somewhat |
| ## Somewhat | 0.0015 | - |
| ## No | 1.1e-08 | 0.0012 |

Question 4: Impact of sleep on student stress

Test Performed: Multiple Linear Regression

Model:

$$\beta_0 + \beta_1 \text{Hours_Sleep} + \beta_2 \text{Hours_Sleep}^2 + \beta_3 \text{sleep} + \beta_4 \text{Rested}$$

Hypothesis: $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

Result: p-value = $3.582e-11$, we reject the null hypothesis. Thus, we can conclude that stress levels are affected by the time, attitudes, and quality of sleeping.

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.020347   0.155316  19.446 < 2e-16 ***
## Hours_sleep  -0.133839   0.042800  -3.127  0.00181 **
## I(Hours_sleep^2) 0.015093   0.004732   3.190  0.00146 **
## sleep         0.053225   0.018063   2.947  0.00328 **
## Rested        0.078854   0.018767   4.202  2.85e-05 ***
```

Result: Adjusted R-squared: 0.04186 (> 0.02)

More variability in stress level can be explained by the integrated model.

Question 5: Impact of hobbies on student stress

Test Performed: Two Sample t-test

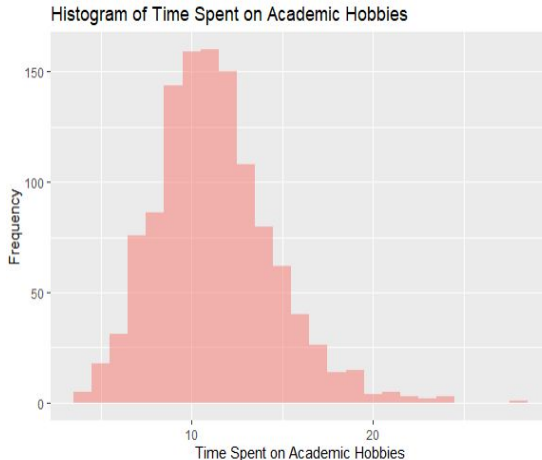
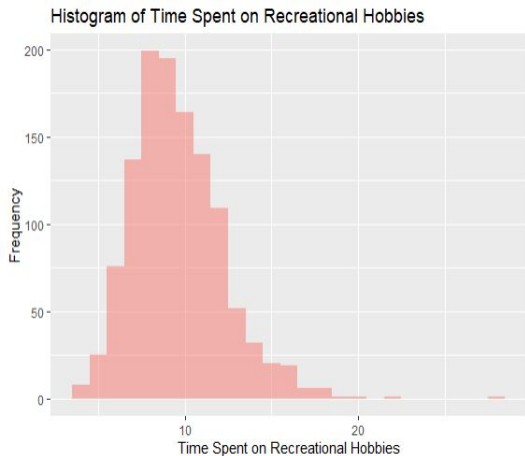
Hypothesis:

H_0 : Time spent on hobbies are same between students who are stressed and students who are not stressed.

Result: P-value = 0.1802, we fail to reject the null hypothesis at the significance level of 0.05 and conclude that the time spent on hobbies by both the groups are same. 95% confidence level is (-1.022, 6.14), suggesting that the interval contains 0 and thus, hobbies on the whole might not impact stress levels.

Hobbies included are:

1. athletics, such as varsity sports or intramurals
2. partying or going out to bars and clubs
3. playing games, such as video games or board games, with friends.
4. watching online recreational content such as on Netflix or Youtube
5. participating in academic organizations and research
6. Studying
7. attending office hours
8. utilizing educational resources, such as the library and online tools.



Question 5: Impact of hobbies on student stress

Test Performed: Two Sample t-test on individual hobbies.

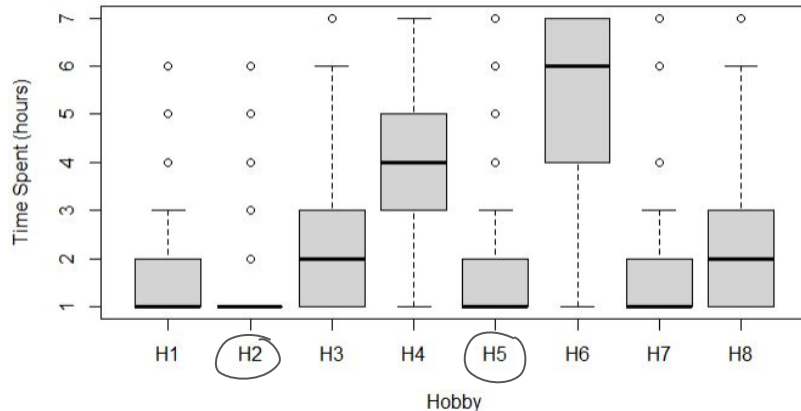
Null Hypothesis: Time spent on this particular hobby are same between students who are stressed and students who are not stressed.

Significant Result: P-value = 0.026, we reject the null hypothesis at the significance level of 0.05 and conclude that the time spent on **hobby 3** by both the groups are not the same. Students who are less stressed are spending more time playing games with friends.

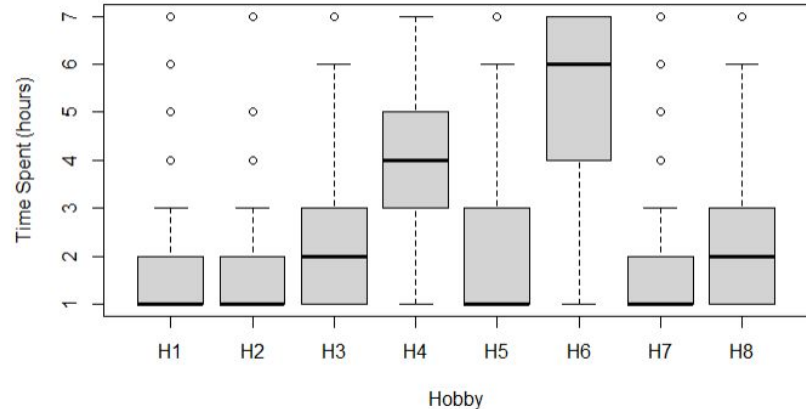
Hobbies included are:

1. athletics, such as varsity sports or intramurals
2. partying or going out to bars and clubs
3. playing games, such as video games or board games, with friends.
4. watching online recreational content such as on Netflix or Youtube
5. participating in academic organizations and research
6. Studying
7. attending office hours
8. utilizing educational resources, such as the library and online tools.

Comparison of Time Spent on Hobbies among stressed students



Comparison of Time Spent on Hobbies among not stressed students



Conclusion



Exercise intensity, frequency, and type show minimal to no impact on stress levels



Stress levels differs across ethnic groups



The nature of coursework shows impact on stress levels when seen as an interaction of multiple features



Time, Attitudes, and Quality of sleeping have impact on stress levels



Time spent on hobbies do not differ with stress levels

Questions ?



Thank you!