香港城市大學
City University of Hong Kong
Innovating into the Future

Group 20: Screen Time & Sleep Quality Study

Group members: Shengpeng RAO, Xiaorui WANG,

Yutong ZHANG, Xueqin SHU, Linlin Xu

Date: November 25, 2025

香港城市大學
City University of Hong Kong
Innovating into the Future

1. Project Overview	Linlin XU
2. Relationship analysis of 80000-dataset	Yutong ZHANG
3. Validation of the 500-dataset	Xueqin SHU
4. Regression analysis on 500-dataset	Shengpeng RAO
5. Multivariate relationship of 80000-dataset	Xiaorui WANG

香港城市大學
City University of Hong Kong
Innovating into the Future

1. Project Overview

Linlin XU

Project Overview

- **Scientific Question:**

The relationship between the total screen time (including for study, entertainment and socializing) students spending on electronic devices each day and their sleep quality.

- **Potential Audiences of interest:**

a) Students;

b) Teachers & Guardians;

c) App Developers.

Project Overview

- **Research Significance：**

For individual students: providing a scientific basis for self-management

For education & guardianship: informing scientific intervention and advocacy

For product design: guiding responsible technology development

Core Value: contribute data-driven insights towards building a more health-conscious digital society

Project Overview

- **Datasets to Use:**

a) Arsalan Jamal (2024). Student Sleep Patterns, Version 1, from https://www.kaggle.com/datasets/arsalanjamal002/student-sleep-patterns/data

b) Aryan Kumar (2025, April), Student Habits and Academic Performance Dataset, from https://www.kaggle.com/datasets/aryan208/student-habits-and-academic-performance-dataset

- **R Packages：**

tidyverse, dplyr, ggplot2, tidyr, reader, purrr, stringr, forcats, lubridate, MASS, moments, pROC, broom, cluster, ClusterR, Rtsne, uwot, gridExtra, corrplot, RColorBrewer, fmsb, reshape2, Matrix.

Project Overview

- **Conclusions：**

a) Screen time has no significant impact on students' sleep quality;

Despite substantial variations in screen time across groups, average sleep duration remained stable at around 7 hours with minimal fluctuation. Both Pearson and Spearman correlation coefficients were close to zero and statistically non-significant.

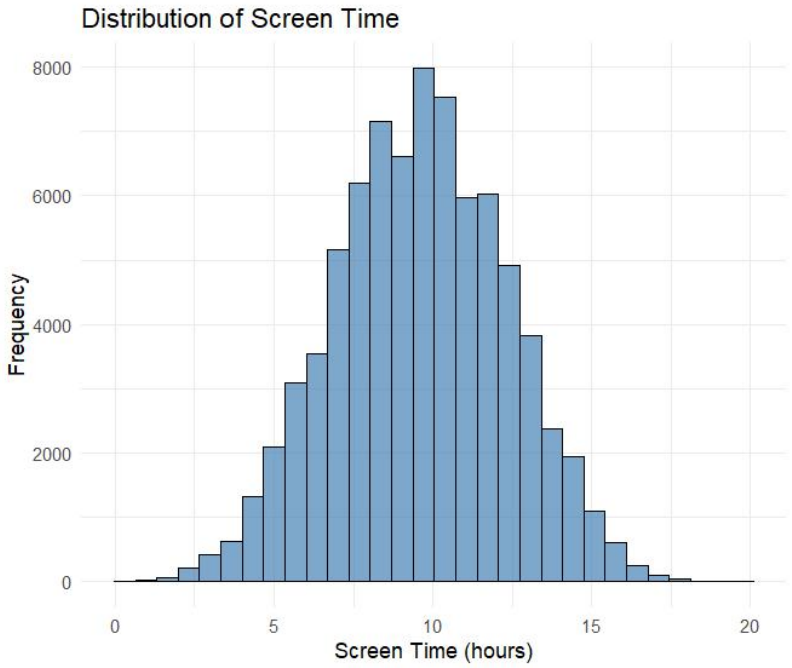
b) Sleep quality demonstrates strong stability and is primarily regulated by the complex interplay of multiple factors.

香港城市大學
City University of Hong Kong
Innovating into the Future

2. Relationship analysis of 80000-dataset

Yutong ZHANG

Data Distribution & Basic Characteristics



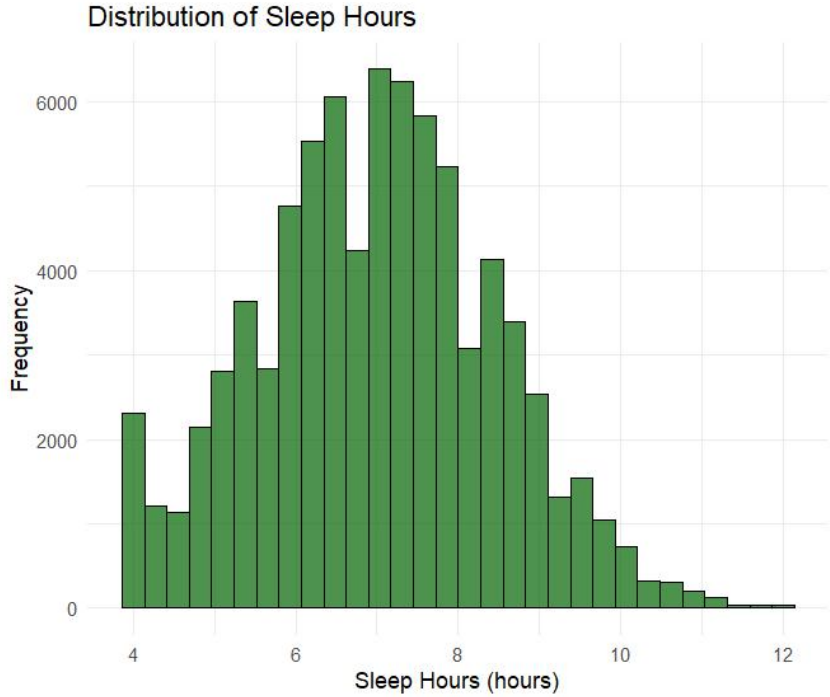


Fig.1 Distribution of Screen Time

Mean: 9.6 hours/day
Range: 0.3 - 19.8 hours
Relatively concentrated distribution
Standard Deviation: 2.7 hours

Fig.2 Distribution of Sleep Hours

Mean: 7.0 hours/day
Range: 4 - 12 hours
Relatively stable distribution Standard Deviation: 1.5 hours

Variable Relationship & Group Analysis

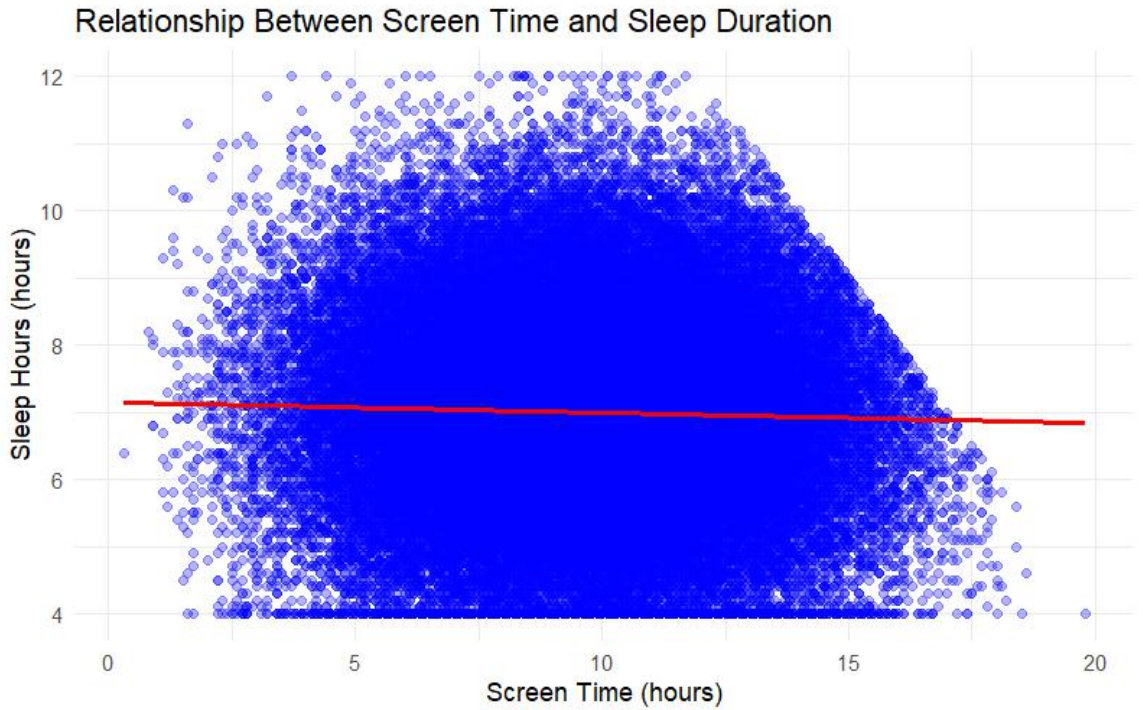


Fig.3 Relationship Between Screen Time and Sleep Duration

Pearson Correlation Coefficient: -0.029
p-value: < 0.001
Conculusion: Very weak negative correlation

Variable Relationship & Group Analysis

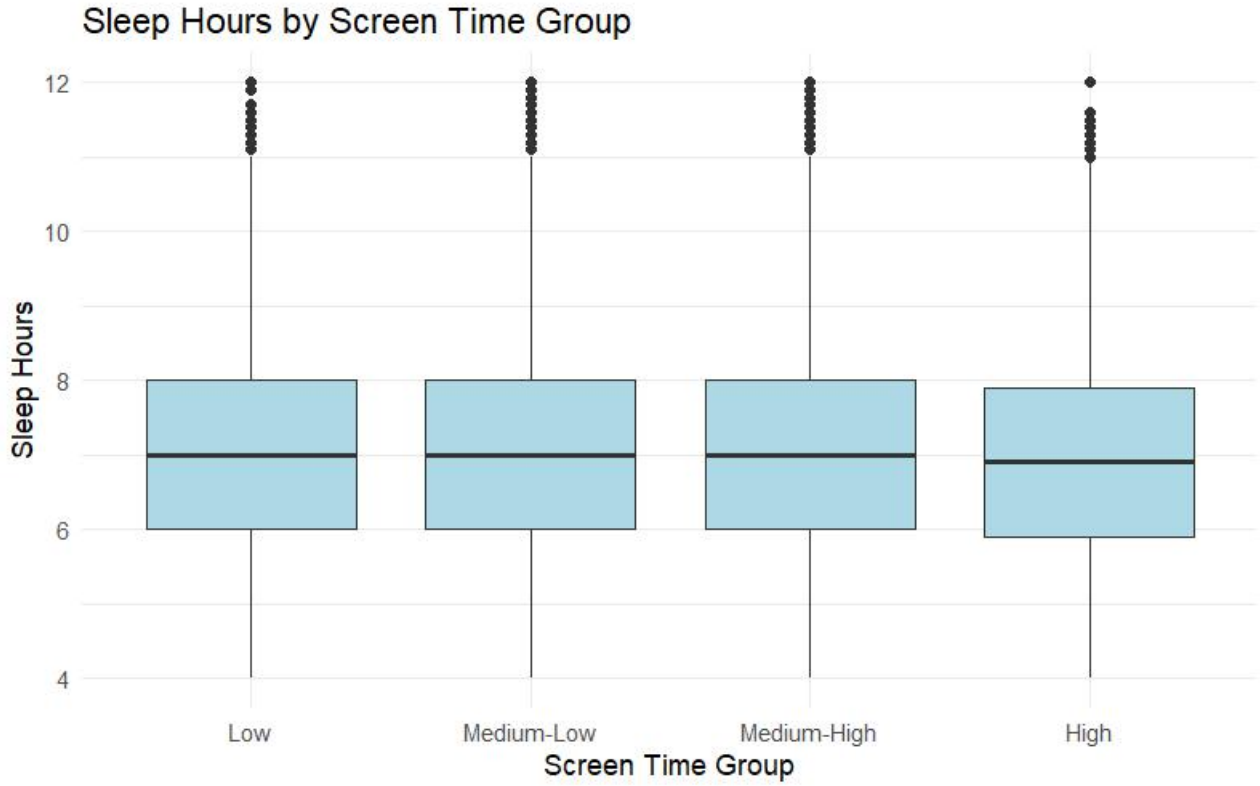


Fig.4 Sleep Hours by Screen Time Group

Conclusion: The average sleep durations across different screen-time groups are highly similar.

Research Conclusions & Implications

- **Negligible Practical Impact**

Only a statistically significant but practically very weak association exists between screen time and sleep duration.

- **Sleep Duration Shows Remarkable Stability**

Even as screen time more than doubled from the lowest to the highest usage group (6.1h → 13.1h),average sleep duration remained anchored around 7 hours, with a negligible maximum fluctuation of just 0.09 hours (≈ 5 minutes).

- **Other Factors Are Likely Dominant**

Sleep duration appears to be primarily governed by factors other than screen time (e.g., circadian rhythms, academic stress, daily routines).

香港城市大學
City University of Hong Kong
Innovating into the Future

3. Validation of the 500-dataset

Xueqin SHU

Validation of the 500-dataset

The dataset contains 500 rows (students) and 14 columns. Each row represents a unique student, and the columns describe their attributes and sleep-related information.

- **I. check and delete the NA（ missing counts is 0)**
- **II. check and delete outliers**

Rules: Screen Time > 20 hours/day; Study Hours > 20 hours/day; Sleep Quality < 1 或 > 10(each category has 0 outliers)

- **III. combine study time and screen time**

Study_Hours: Average number of hours spent studying per day (float).

Screen_Time: Average number of hours spent on screens (excluding studying) per day (float).

- **IV.I check the original data type(sleep quality: integer; total screen time: numeric)**
- **IV.II check the range of values for sleep quality(they are 1-10）**

• V.I descriptive statistics of total screen time

Standard deviation:3.55
Skewness: 0.03 (0=symmetric)
Kurtosis:1.99(3=normal distribution)

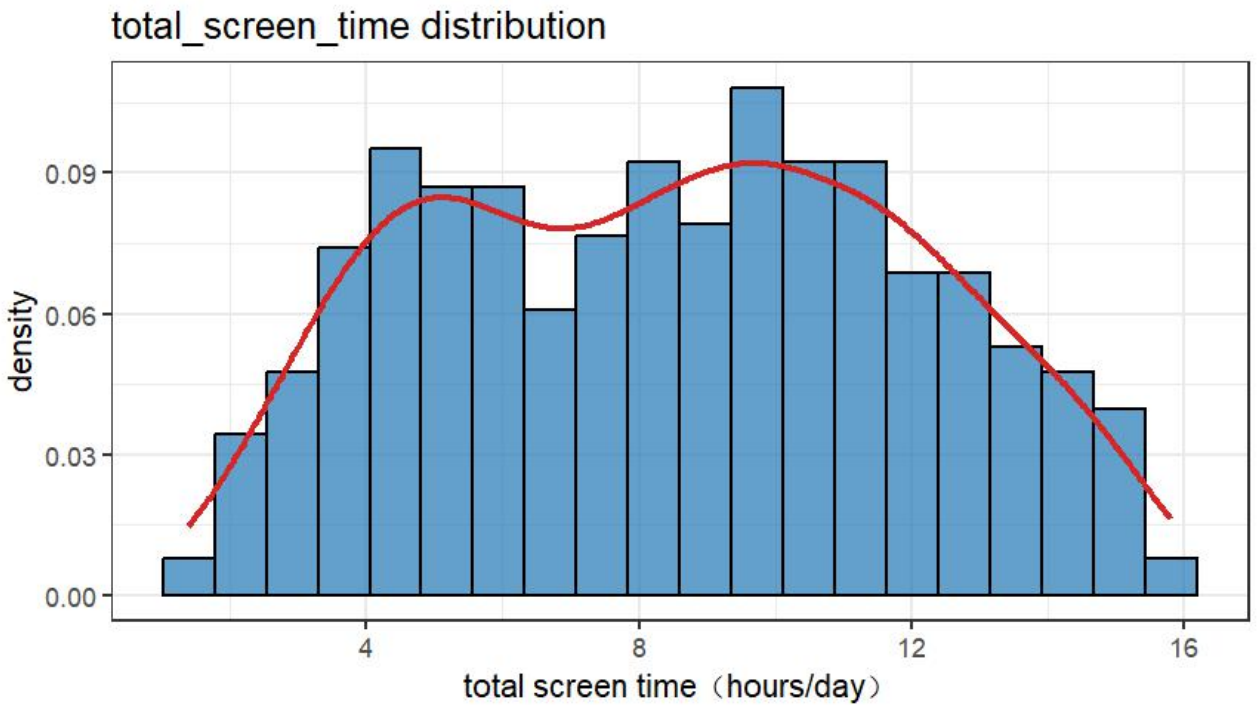


Fig.5 Total Screen Time Distribution

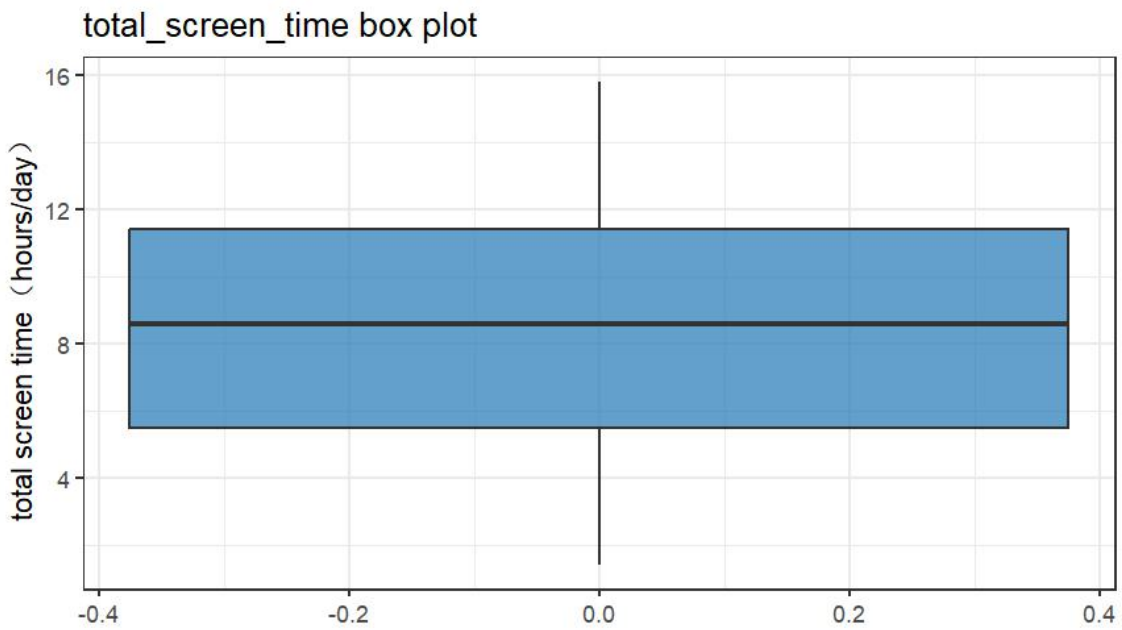


Fig.6 Total Screen Time Box Plot

- **V.II descriptive statistics of Sleep Quality**

Mean:5.362
Standard deviation:2.97
Skewness:0.04(0=symmetric)
Kurtosis:1.73(3=normal distribution)

Fig.7 Sleep_Quality distribution histogram

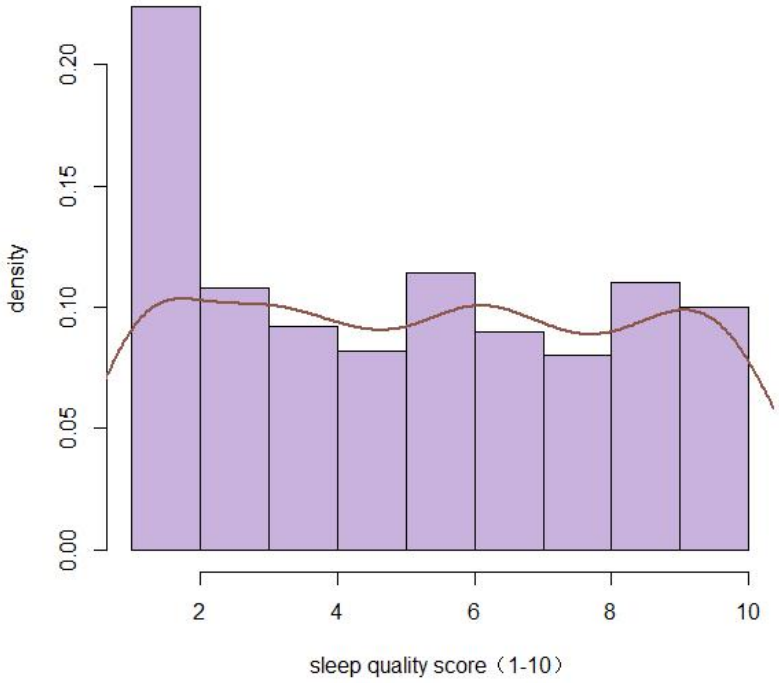


Fig.8 Sleep_Quality score proportion

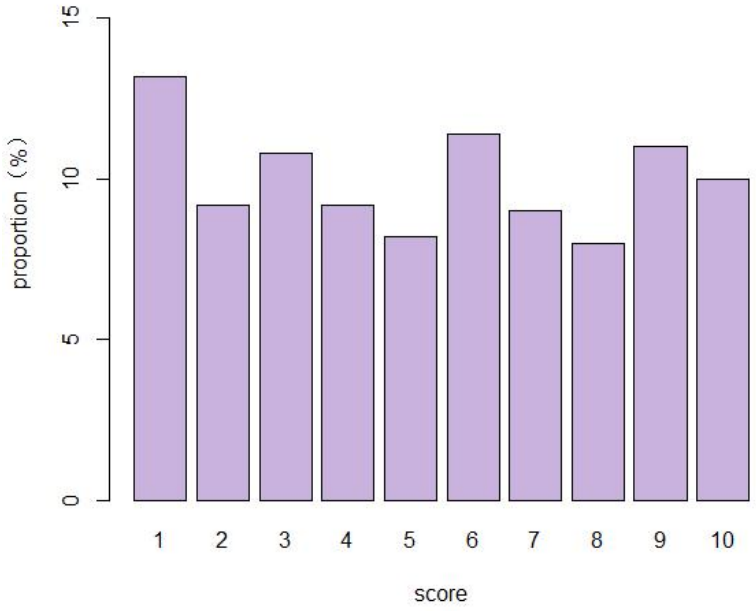
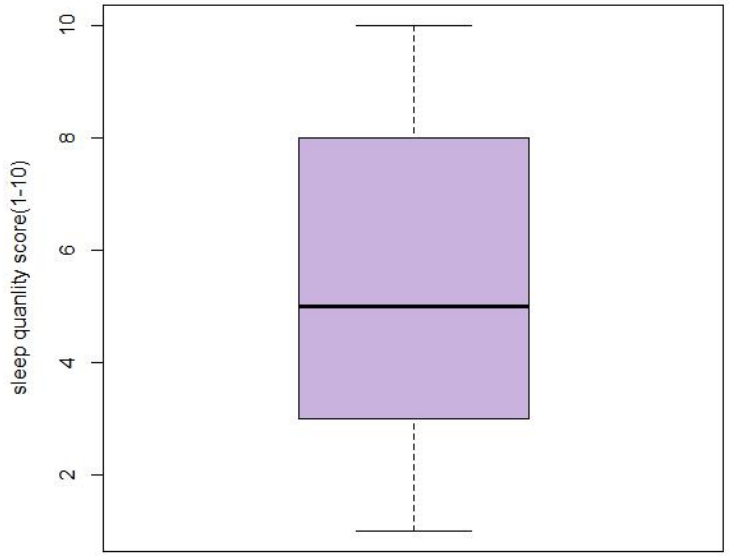
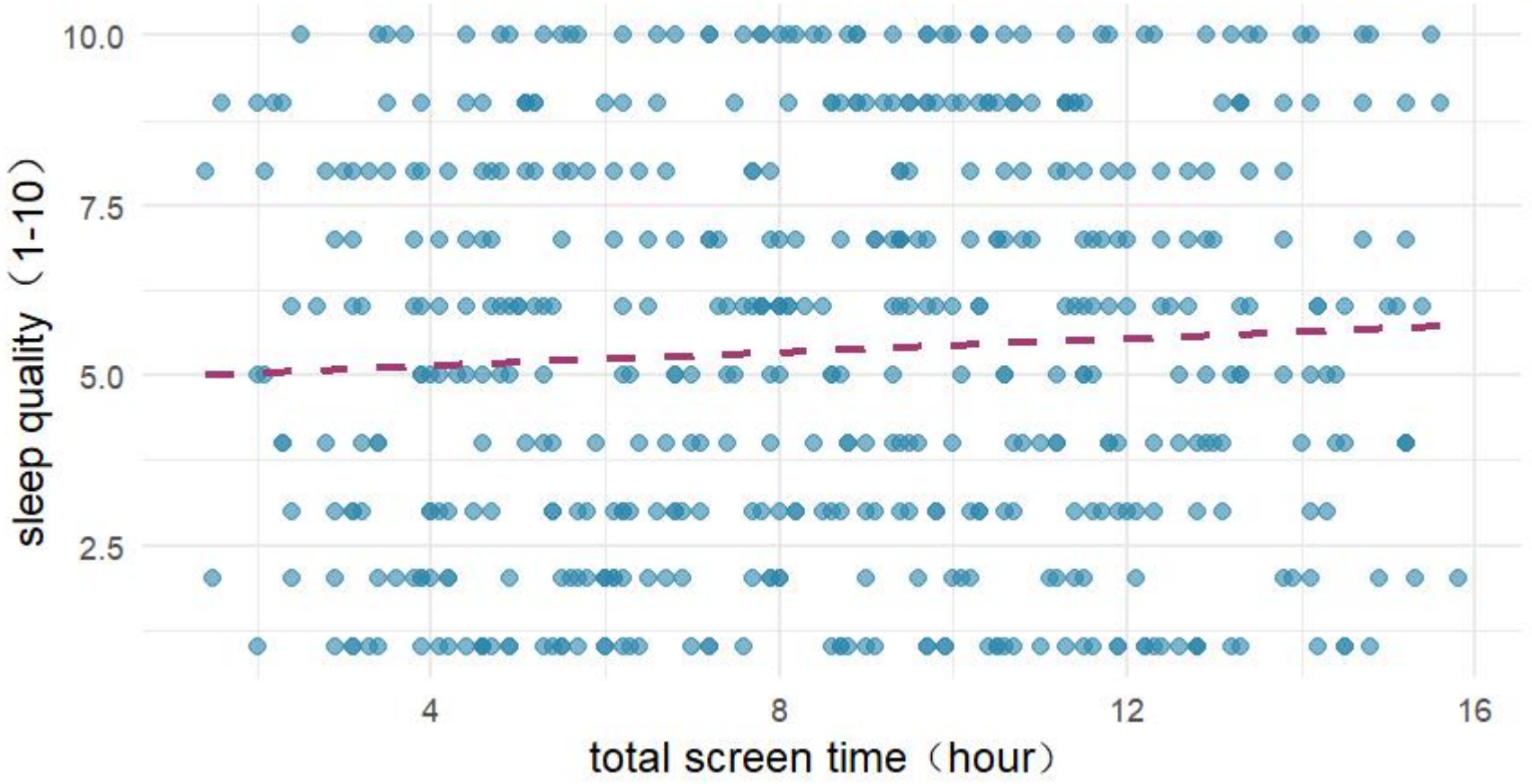


Fig.9 Sleep_Quality box plot



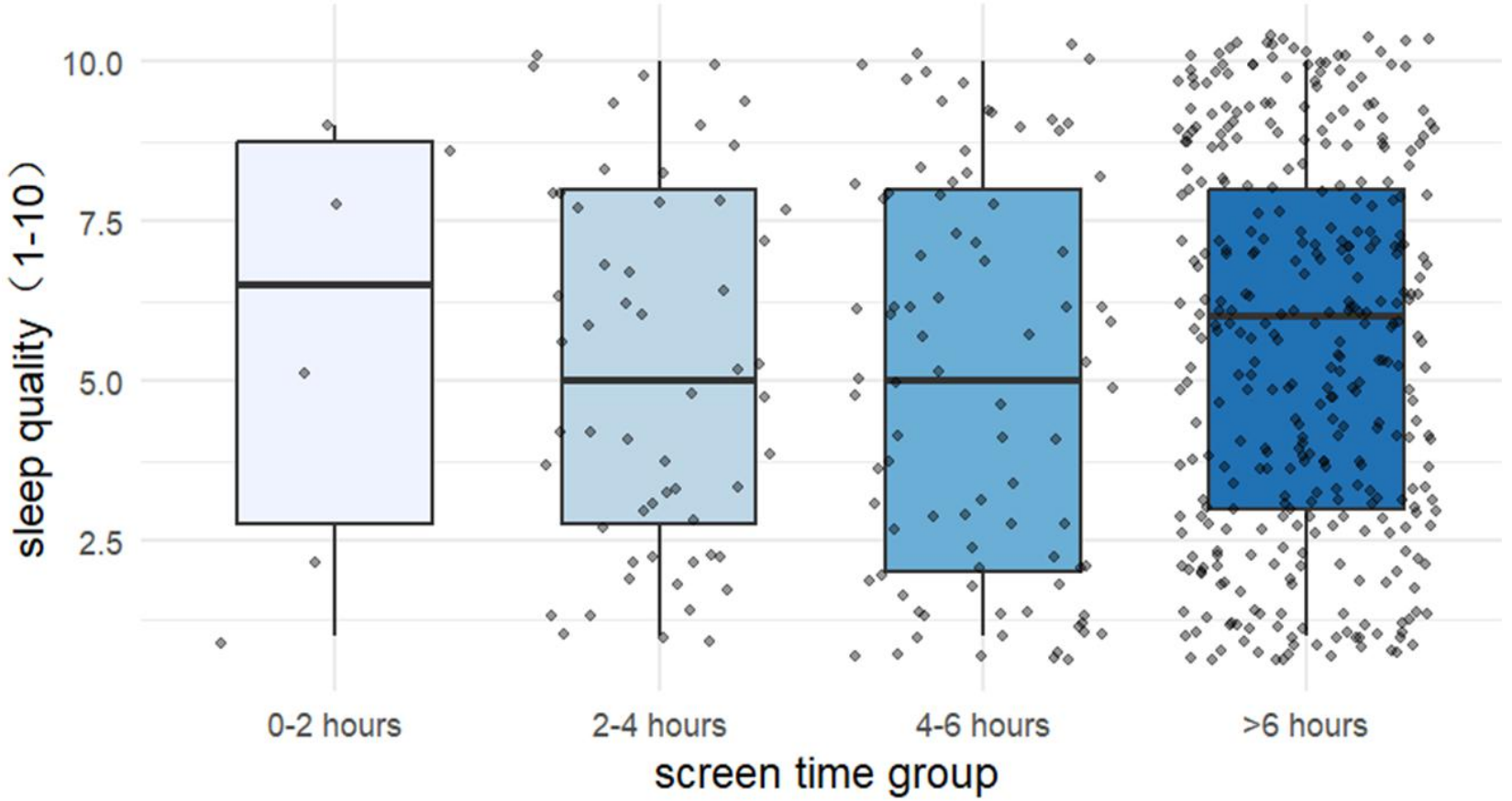
- **VI.I draw scatter plot**

Fig.10 the relationship between total screen time and sleep quality



- **VI.II draw box plot**

Fig.11 Distribution of sleep quality across different total screen time gr



- **VII.I Pearson**

correlation coefficient：-0.0268 (linear relationship)

p value：0.5505(p≥0.05), not significant

Correlation intensity： weak

- **VII.II Spearman**

correlation coefficient：-0.0262 (monotonic relationship)

p value：0.5585(p≥0.05), not significant

Correlation intensity： weak

Core conclusion:

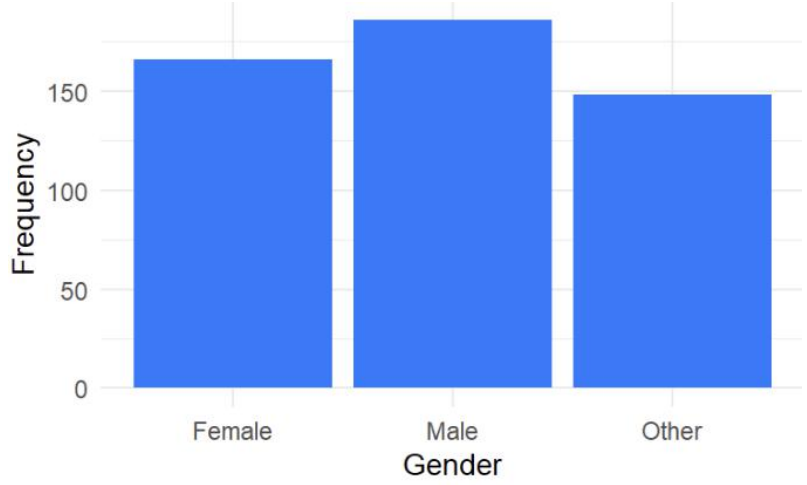
No significant correlation was detected between total screen time and sleep quality（p≥0.05）

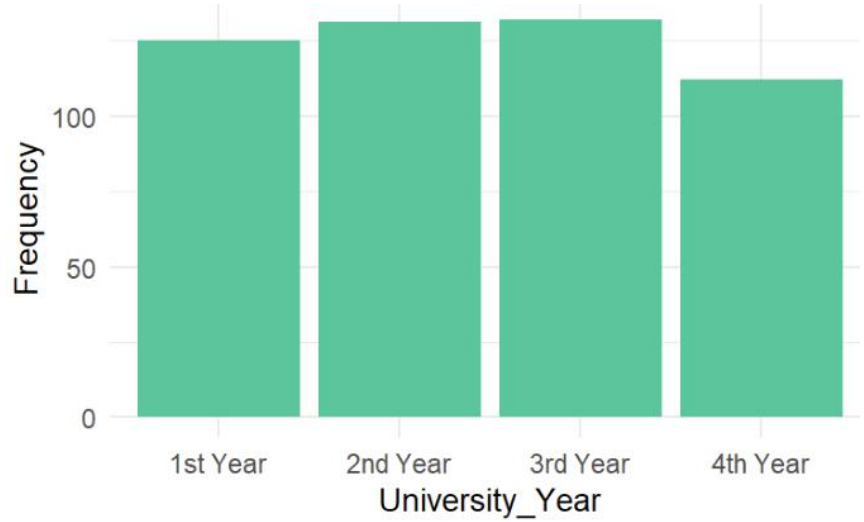
香港城市大學
City University of Hong Kong
Innovating into the Future

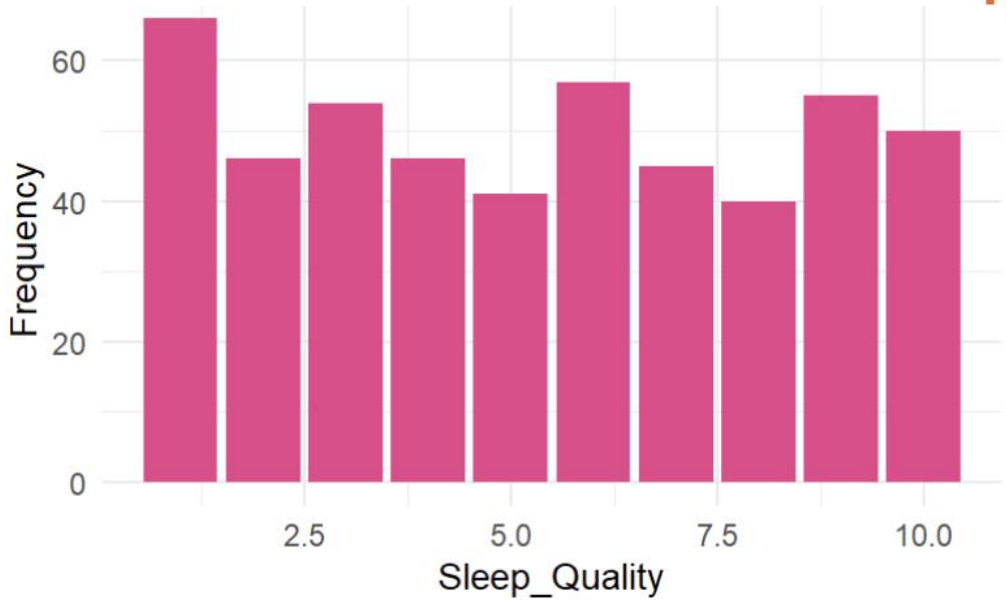
4. Regression analysis on 500-dataset

Shengpeng RAO

Categorical Data Visualization & Transformation







1. Details on transformation
2. The reason for transformation
3. The reason for model selection

Logistic Regression & Interpretation

```
Call:
glm(formula = SQ_binary ~ Study_Hours + Gender_Female, family = binomial,
    data = Sleep_n)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.26101    0.20302  -1.286   0.1986
Study_Hours   -0.05404    0.02780  -1.944   0.0519 .
Gender_Female -0.37700    0.20864  -1.807   0.0708 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 635.59  on 499  degrees of freedom
Residual deviance: 629.06  on 497  degrees of freedom
AIC: 635.06

Number of Fisher Scoring iterations: 4
```

1. The basis for model construction
2. Coefficient interpretation
3. Deviance interpretation

Evaluation of Prediction Ability & Conclusion

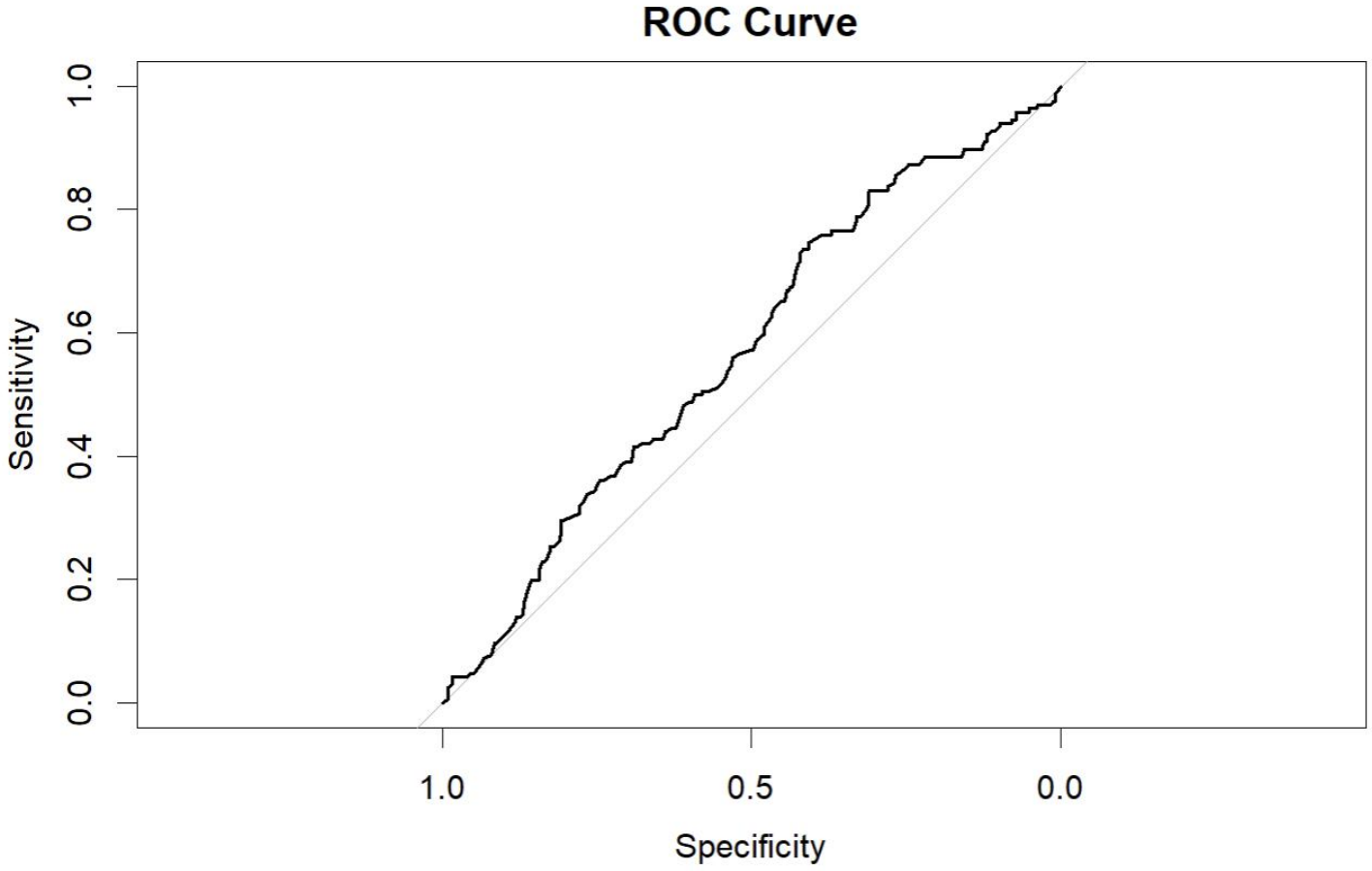


Fig.15 ROC Curve

1. Evaluation
2. Conclusion
3. Research shortage & envisage

香港城市大學
City University of Hong Kong
Innovating into the Future

5. Multivariate relationship of 80000-dataset

Xiaorui WANG

Multivariate Relationship

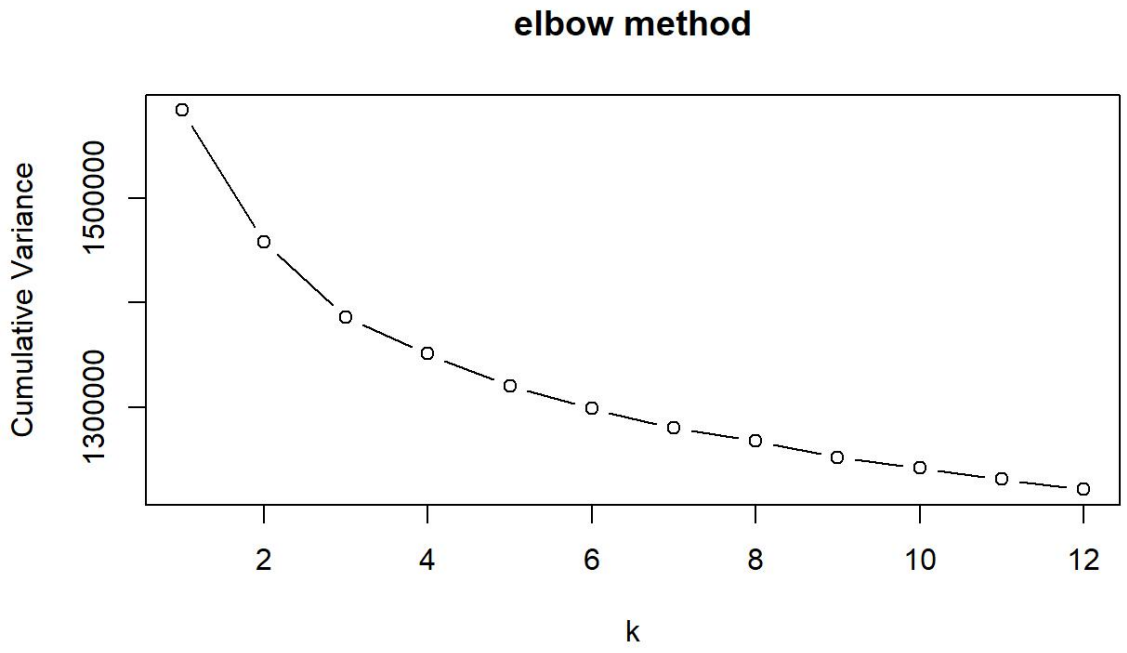


Fig.16 Elbow method scatter diagram

Use elbow method to get the best k value

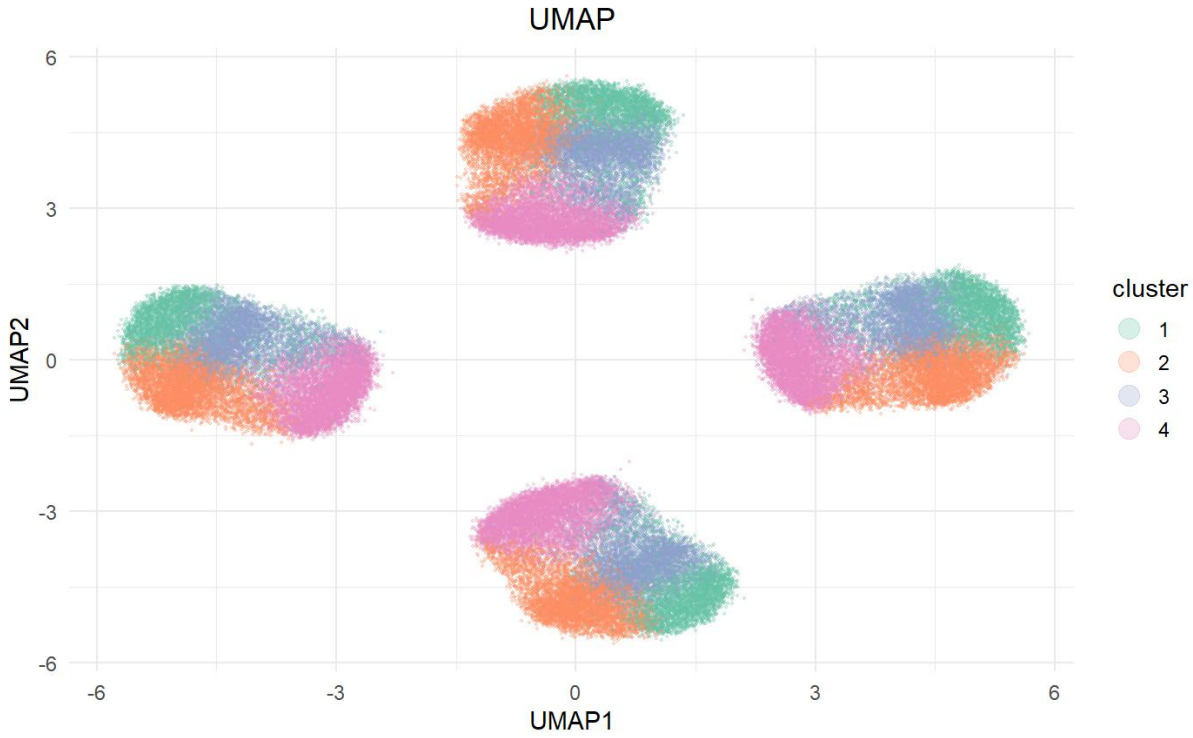
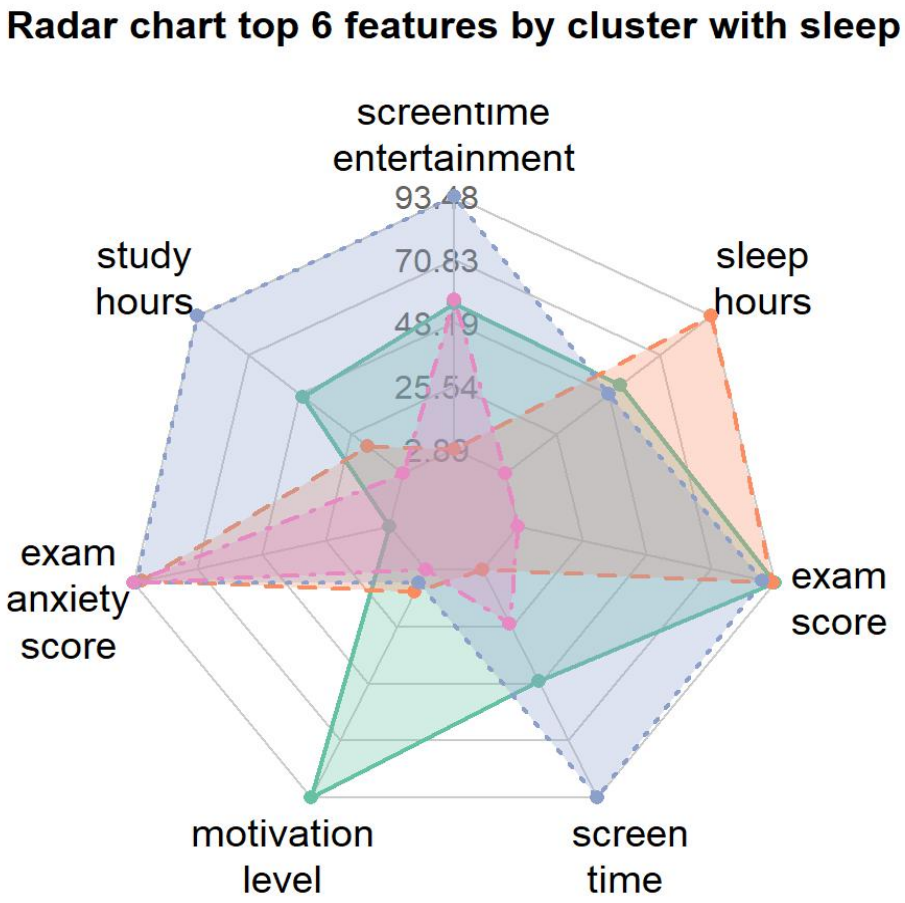


Fig.17 Umap Clustering Plot

Use umap clustering method based on the k value

Multivariate Relationship



- **Cluster 1: Hard-working**
- **Cluster 2: Good sleeping**
- **Cluster 3: Talent**
- **Cluster 4: Playful**

Fig.18 Radar chart of 4 cluster

Multivariate Relationship

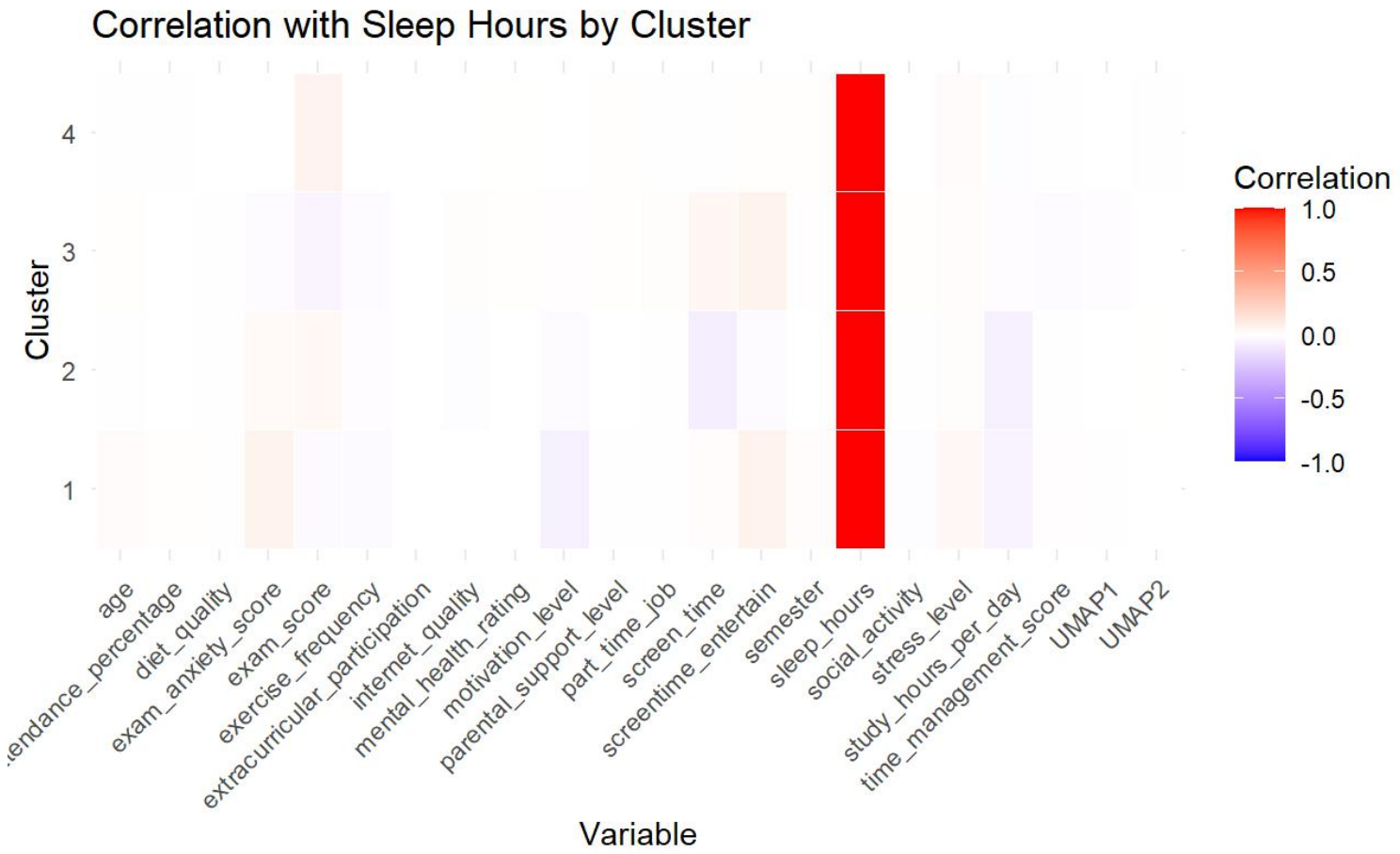


Fig.19 Heatmap of sleep vs. others by clusters

Modeling by Cluster

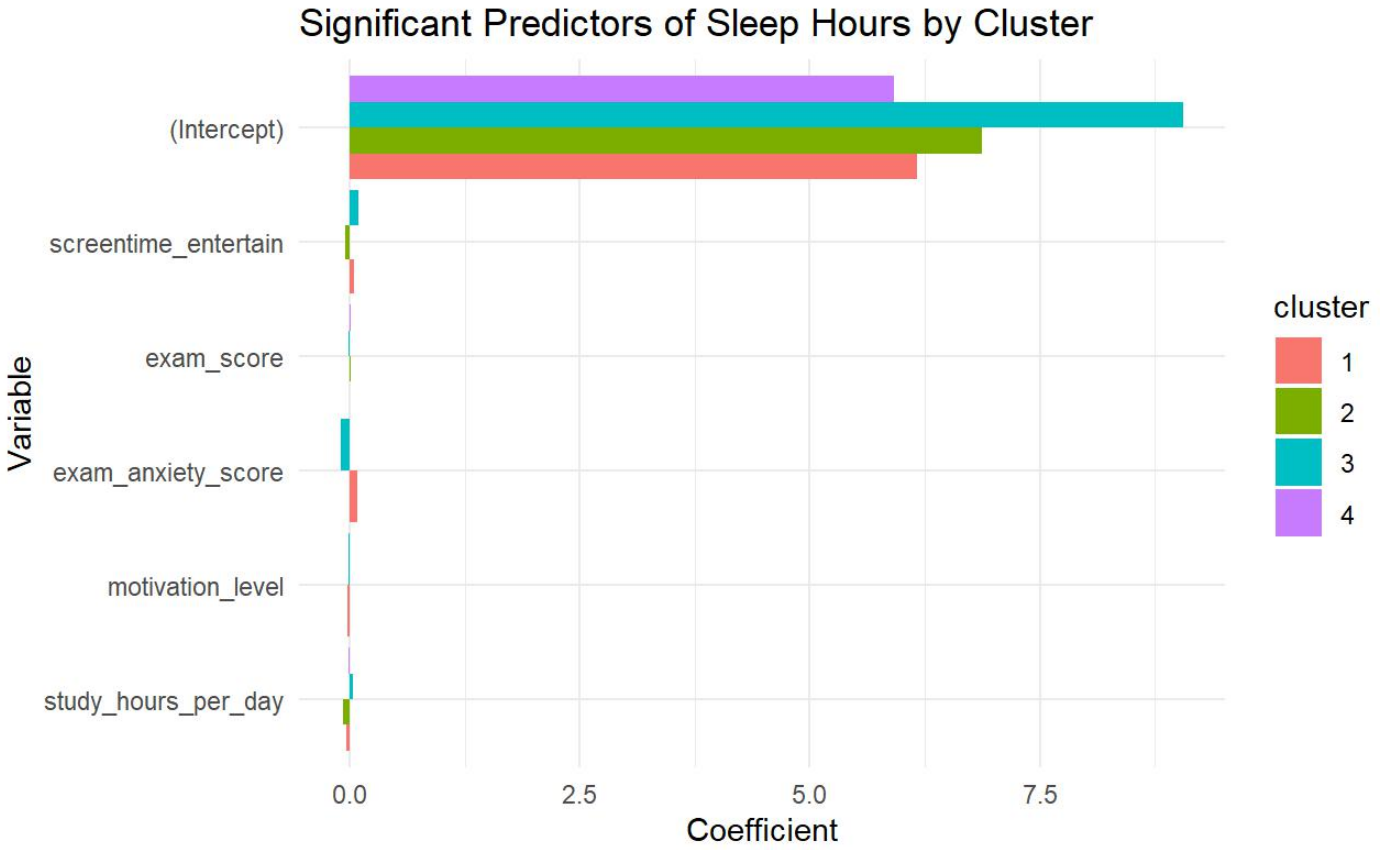


Fig.20 Sleep predict model

The results showed that the influence of the variables was relatively weak, indicating that the driving factors of sleep behavior vary across groups.

Github link

https://github.com/Archmi7/sleep-project/tree/main

Thank You !

Q&A