Unit 2 Homework

Dataset: *student_performance.csv*

The attached dataset contains information about a group of college students and their academic performance. It includes the following variables:

student_id: A unique identifier for each student (integer).

age: The age of the student (integer).

gender: The gender of the student (character: "Male" or "Female").

major: The student's chosen major (character: "Science", "Engineering", "Business", "Arts", or "Other").

gpa: The student's grade point average on a 4.0 scale (numeric).

sat_math: The student's SAT math score, ranging from 200 to 800 (integer).

sat_verbal: The student's SAT verbal score, ranging from 200 to 800 (integer).

study_hours: The average number of hours the student spends studying per week (numeric).

extracurricular: The number of extracurricular activities the student participates in (integer).

part_time_job: Whether the student has a part-time job (logical: TRUE or FALSE).

Homework Questions: Calculate the answers to the following questions and use that information to answer the Unit 2 homework quiz.

Q1: Calculate the mean, median, and mode of the students' ages. Interpret the results.

Q2: Compute the range, variance, and standard deviation of the students' GPAs. Explain the meaning of each measure.

Q3: Determine the skewness and kurtosis of the study_hours variable. Discuss the implications of the results.

Q4: Tell R-Studio to ignore any missing values and re-calculate the answers for Q1 – Q3. Discuss the implications of this change.

Q5: Calculate the standardized scores (z-scores) for the sat_math and sat_verbal variables. Interpret the meaning of a z-score.

Q6: Create a histogram of the students' ages and describe the distribution.

Q7: Generate a scatterplot of gpa against study_hours. Discuss any visible patterns or relationships.

Q8: Compare the mean GPAs of students with and without part-time jobs using an appropriate plot.