**Assignment-1**

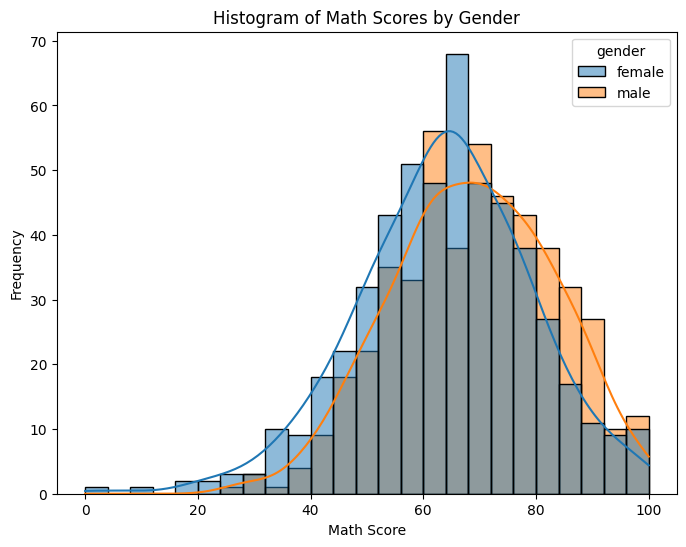
**Abhinav Reddy Ayyadapu**

**16354733**

**Q2.Data Analysis of Student Performance Dataset Results**

**🡪Histogram:**

It is the histogram graph of math scores by gender, though the labels and axes are not fully visible.



* **Distribution:** The histogram shows the distribution of math scores for two groups, likely male and female students. The female scores are represented by the blue bars on the left, while the male scores are represented by the green bars on the right.
* **Frequency:** The height of each bar represents the number of students who scored within that particular range of math scores. For example, the tallest bar for females seems to be between 60 and 70, suggesting that more females scored in that range than any other.
* **Range:** The x-axis represents the range of possible math scores, though the specific values are not visible. It seems to start around 20 and end around 100.

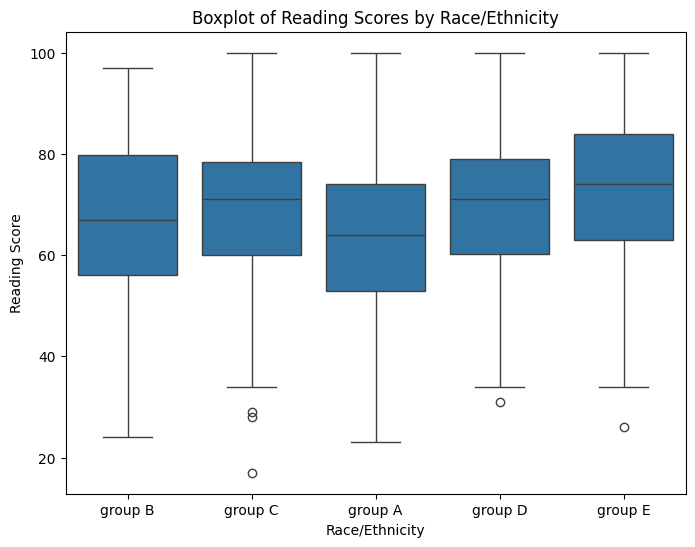
**🡪Box plot:**

It is a boxplot that shows the distribution of reading scores by race/ethnicity.

**Race/ethnicity groups:** The boxplot compares the reading scores of five different groups, labeled "Group A", "Group B", "Group C", "Group D", and "Group E". Unfortunately, the image doesn't provide any information about what these groups represent.

**Distribution:** Each box in the boxplot represents the distribution of reading scores for one of the race/ethnicity groups. The box shows the middle 50% of the scores, with the horizontal line in the middle representing the median score. The whiskers at the top and bottom of the box extend to the highest and lowest scores that are not considered outliers. Outliers are data points that fall outside a certain range based on the interquartile range (IQR). The IQR is calculated by subtracting the first quartile (Q1) from the third quartile (Q3). Any data points that are less than Q1 - 1.5 \* IQR or greater than Q3 + 1.5 \* IQR are considered outliers and are represented by individual circles or squares in the plot.

**Reading scores:** The y-axis represents reading scores, but the specific range of scores is not visible in the image.



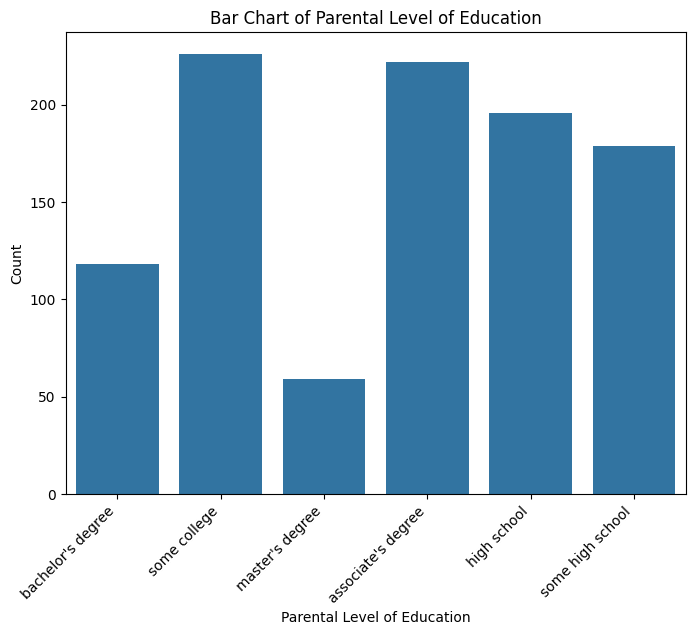
Here are the some of the observations

**Group A**: The median reading score in Group A is the highest among all groups. The box is relatively narrow, indicating that the scores are tightly clustered around the median. There are no outliers.

**Group B:** The median reading score in Group B is lower than Group A but higher than Groups C, D, and E. The box is wider than Group A.

**🡪Bar plot:**

It is a bar chart that shows the parental level of education for a group of students.



**Parental level of education:** The x-axis shows different levels of parental education, including "high school," "some college," "associate's degree," "bachelor's degree," and "master's degree."

Student count: The y-axis shows the number of students who have parents with each level of education. The numbers are represented by bars of different heights.

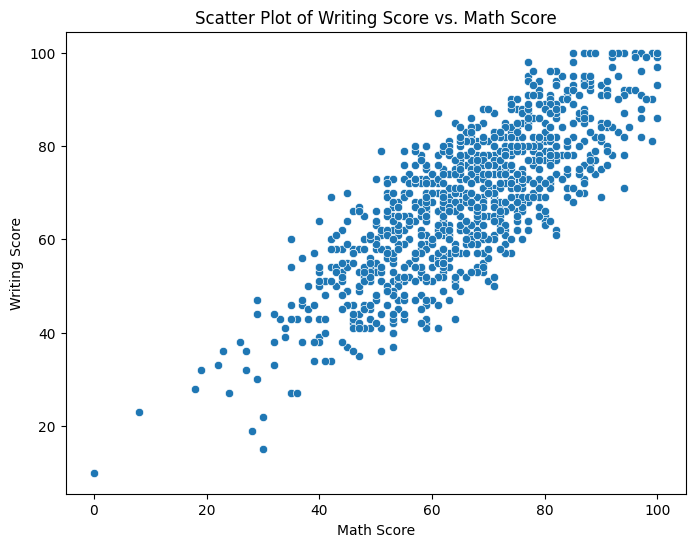
**Observations:**

The most common level of parental education in this group is "some college," followed by "bachelor's degree" and "high school."

There are fewer students whose parents have "associate's degree" or "master's degree."

It is important to note that this is just a single image, and it may not be representative of the general population. Additionally, the image does not provide any information about the students themselves, such as their academic performance or socioeconomic status.

**🡪Scatter plot:**It’s a scatter plot showing the relationship between writing score and math score, but the labels and axes are partially obscured.



**Data points**: Each dot on the graph represents the writing score and math score of an individual student. There appears to be a moderate number of data points, but the exact number is difficult to discern due to overlapping dots.

**Trend:** There is a faint upward trendline visible, suggesting a positive correlation between writing score and math score. This means that students with higher writing scores tend to also have higher math scores, and vice versa. However, the trendline is faint, indicating that the correlation is not very strong.

**Range:** The x-axis likely represents math scores, but the specific range of scores is not visible. The y-axis likely represents writing scores, but the specific range of scores is also not visible.

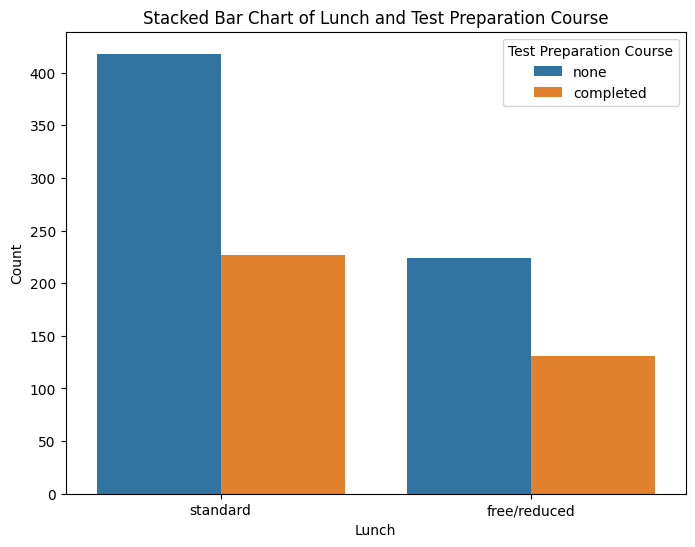
**🡪Stacked bar chart:**

It is a stacked bar chart that shows the lunch participation and test preparation course enrollment for different groups of students, likely standard lunch price students and free/reduced lunch price students.

Lunch participation: Overall, both standard lunch price and free/reduced lunch price students have similar lunch participation rates, with around 220 students participating.

Test preparation course enrollment: There is a significant difference in test preparation course enrollment between the two groups. Standard lunch price students have a much higher enrollment rate, with around 370 students enrolled, compared to only around 30 students from the free/reduced lunch price group.

It is important to note that this chart does not show the total number of students in each group. Therefore, it is impossible to calculate the percentages of students who participate in lunch or enroll in the test preparation course. Additionally, the chart does not provide any information about the reasons for the differences in enrollment rates.



**Cost:** The test preparation course may be offered at a discounted rate for students who qualify for free/reduced lunch, but the discount may not be enough to make it affordable for all families.

**Awareness:** Students from families with higher incomes may be more aware of the test preparation course and its potential benefits.

**Academic support:** Students who qualify for free/reduced lunch may have other academic needs that take priority over enrolling in a test preparation course.

It is important to remember that these are just possible explanations, and more research would be needed to determine the true reasons for the difference in enrollment rates.