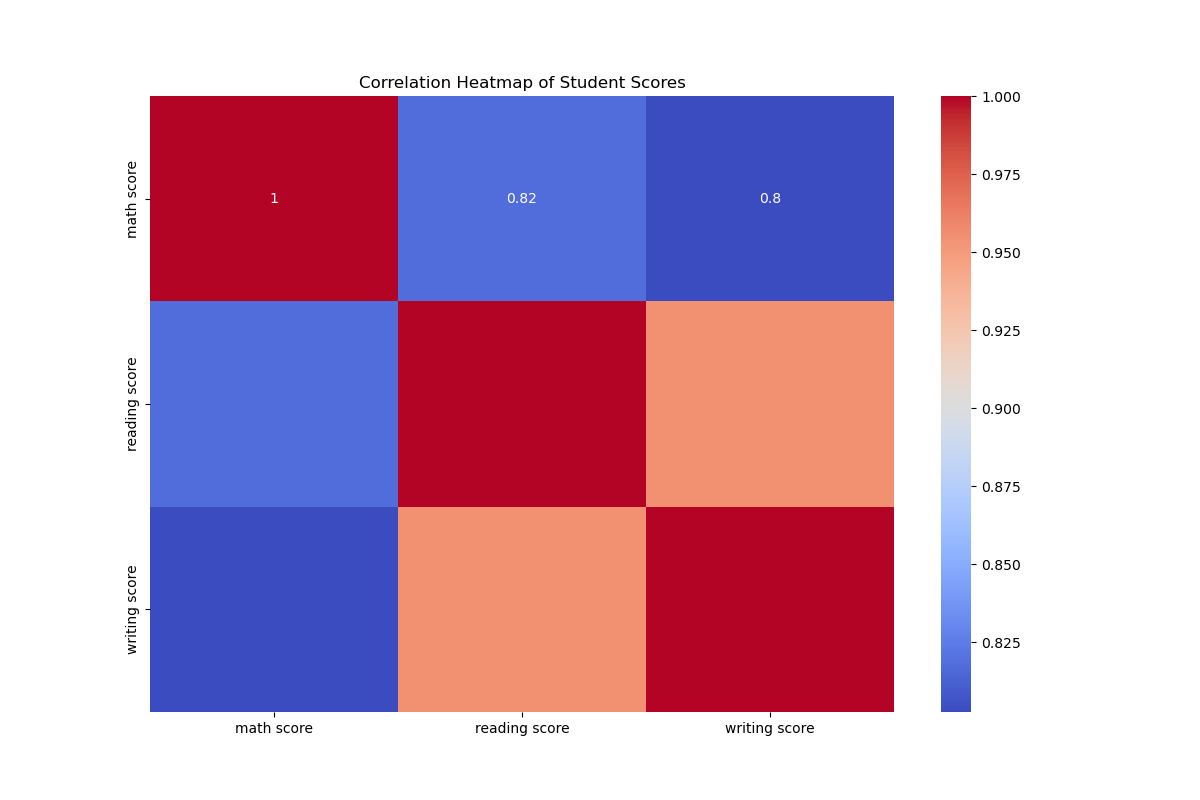
Assignment-1

Sudeepthi Adluru

16353890

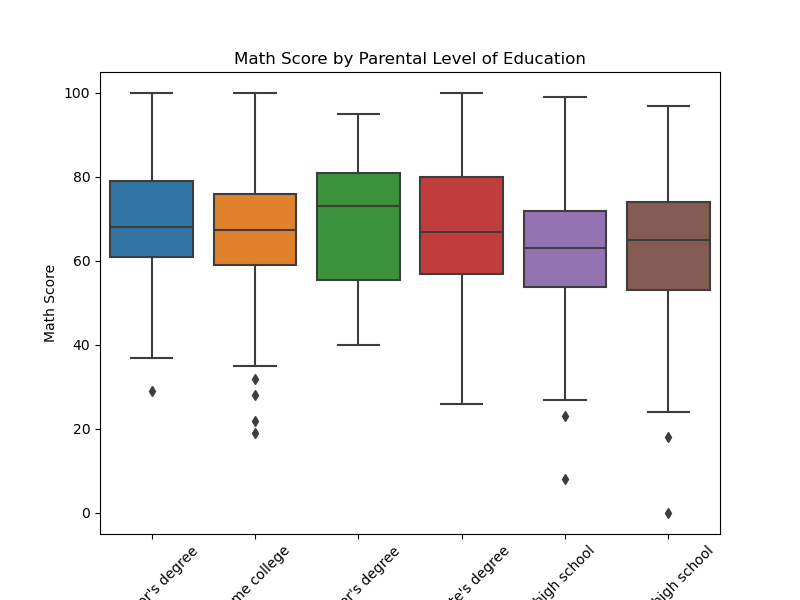
1. This heatmap illustrates the correlation between student scores in three subjects: math, reading, and writing. A correlation measures the strength and direction of a linear relationship between two variables, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation).



* **Math and Reading Scores** have a correlation of **0.82**, indicating a strong positive relationship. This suggests that students who score well in math tend to also score well in reading.
* **Math and Writing Scores** also show a correlation of **0.8**, demonstrating another strong positive relationship.
* **Reading and Writing Scores** are highly correlated, almost perfectly, with a value close to **1**, reflecting that students who excel in reading generally perform very well in writing.

This heatmap shows the close interconnection between performance in these subjects, emphasizing that skills in one area are often reflective of abilities in others.

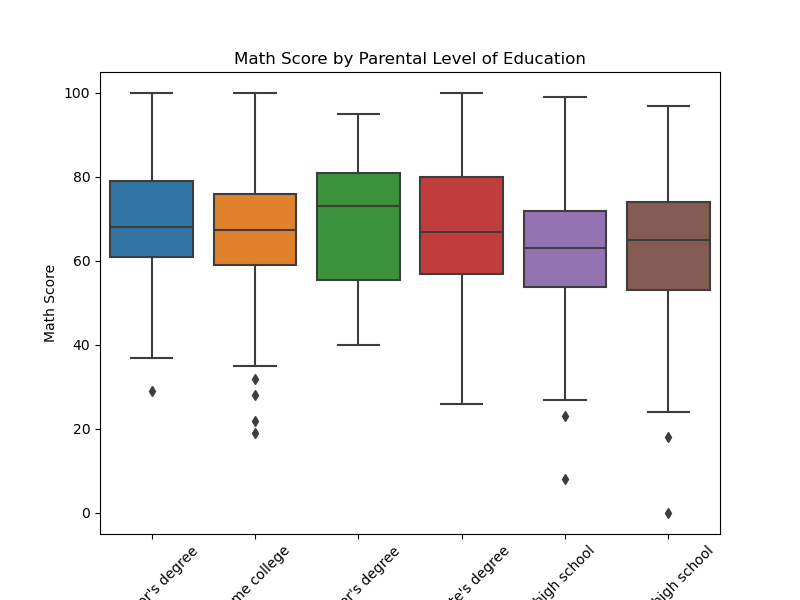
1. This histogram shows the distribution of math scores across all students, along with a kernel density estimate (KDE) to represent the data’s probability density.



* The **math scores** follow a roughly normal distribution, with a majority of students scoring between **40 and 80**.
* The peak is centered around **60-70**, suggesting that the average student score lies in this range.
* There are relatively fewer students scoring below **40** or above **90**, which indicates that most students fall within the middle range of math performance.

This visualization helps in understanding the spread and central tendency of students' math scores.

1. This boxplot compares math scores among students whose parents have different levels of education. Each box represents the spread of scores for a particular parental education level.



* Students whose parents have a **master's degree** or a **bachelor's degree** tend to have higher median math scores compared to those whose parents have only completed **some college** or **high school**.
* The **interquartile range (IQR)**, which is the middle 50% of scores, is wider for some education levels, indicating greater variability in those groups.
* There are some outliers (dots) in groups with lower parental education, suggesting that even though the median score may be lower, a few students still perform exceptionally well.

This plot suggests a possible positive association between parental education level and students’ math performance.

1. **Reading Score vs. Writing Score**

This scatter plot visualizes the relationship between reading and writing scores for students, with the data categorized by gender.



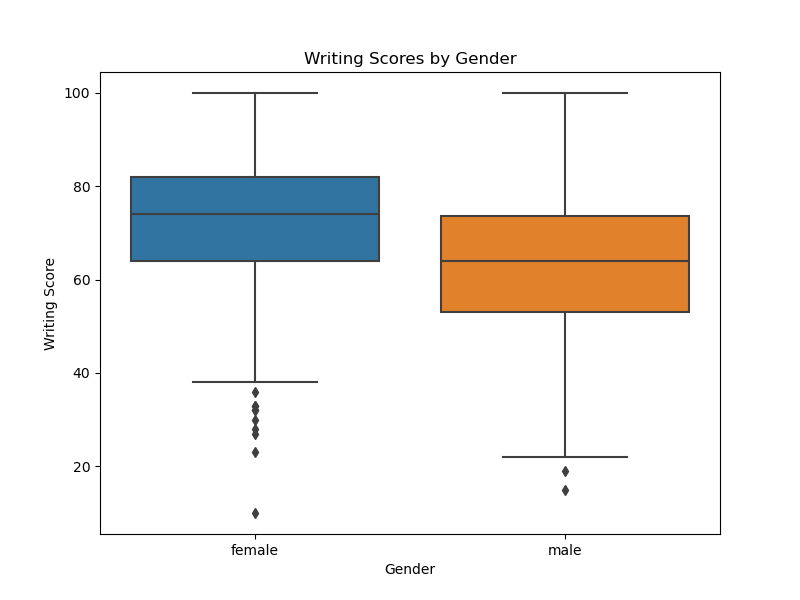
Key points include:

* **Axes**: The x-axis represents reading scores, while the y-axis represents writing scores. Both scores range from approximately 20 to 100.
* **Color Legend**:
  + Blue dots correspond to female students.
  + Orange dots correspond to male students.
* **Observations**:
  + There is a strong positive correlation between reading and writing scores for both genders. As reading scores increase, writing scores also tend to increase.
  + The points form a roughly linear pattern, indicating that students who score well in reading are likely to perform similarly in writing.
  + The data points for both genders largely overlap, indicating that the relationship between reading and writing performance is consistent across genders, with some individual variability.

1. **Writing Scores by Gender**

This box plot compares the distribution of writing scores between male and female students. The key components are:

1. **Axes**:
   1. The x-axis categorizes the data by gender (female and male).
   2. The y-axis represents writing scores, ranging from approximately 20 to 100.
2. **Box Plot Components**:
   1. The boxes represent the interquartile range (IQR), showing the middle 50% of the data. The line inside each box marks the median.
   2. The whiskers extend to show the range of scores, except for outliers, which are represented as individual points outside the whiskers.



* **Observations**:
  + **Female students**:
    - The median writing score is higher for females compared to males.
    - Female students show a slightly broader distribution of scores, with more lower outliers below 40.
  + **Male students**:
    - The writing scores for male students are slightly lower, with the median around 65.
    - There are fewer outliers, and the overall spread of scores is narrower compared to female students.

This analysis indicates that female students tend to have slightly higher writing scores on average, with a broader range of variation.