

INTERPRETATIONS

V1 - Gender box-plots

The boxplots show that female students tend to score lower in math but slightly higher in reading compared to males. The spread of math scores for females is larger than males, suggesting more variability in their performance. Males have a slightly tighter range in math scores. Both math and reading scores for females show more lower-score outliers. This may suggest that while many female students perform around the middle to higher range, there is a small group who are scoring much lower than the majority. It can also be seen up the upper range that across all the students, at least a few have gotten 100s, and only one female got a 0.

V2 - Test prep impact box plots

The boxplot shows that students who did not complete the test preparation course (“none”) have more variability in their math scores compared to those who completed it. The “none” group also has more lower outliers, showing that several students in this group performed poorly. In contrast, the median math score for students who completed the test prep course is slightly higher, suggesting better math scores with test preparation. The quartiles for the completed group are also slightly higher overall, meaning their scores tend to be more concentrated at higher values. However, both groups include students who scored 100, so high achievement is possible regardless of test preparation. In general, completing the test prep course appears to be associated with slightly better and more consistent math performance, but the difference is not dramatic.

V3 - Lunch Type and Overall Average Performance

The bar chart shows that students who receive free or reduced lunch have a lower overall average test score compared to students with standard lunch. The difference between the two bars is noticeable, maybe suggesting a gap in academic performance between the groups. This may indicate that socioeconomic factors, which can be associated with lunch status, could be related to academic scores. However, we know correlation doesn't equal causation, so lunch type itself is likely not the direct factor of the difference, and there may be other underlying factors influencing the correlation, such as access to resources, parental education, or additional academic support. I think it could be beneficial to extend this analysis by breaking the data down

by individual subjects (math, reading, and writing) instead of only looking at the overall average. That would allow us to see whether the gap is consistent across all subjects or driven more strongly by one area.

V4 - Subject Correlations (Heatmap)

The heatmap shows correlations of 1.00 along the diagonal, which is expected since each subject should be perfectly correlated with itself. Looking at the relationships between different subjects, math and writing have a strong positive correlation of 0.80, and math and reading have a correlation of 0.82. The strongest relationship is between reading and writing, with a very high correlation of 0.95. This makes sense as reading and writing are subjects that usually go hand-in-hand, while math requires specific knowledge and skills. In terms of scores, this suggests that students who perform well in reading are very likely to also perform well in writing. All three subjects are strongly positively related, pointing to a consistent pattern of overall academic performance. In terms of the visualization, the heatmap includes redundant values because correlation matrices are symmetric, so each correlation appears twice. Additionally, as we learned in lectures, color alone is not the most precise/easiest way to interpret numeric differences, since small changes in shade can be hard to see.

V5- Math vs Reading with Regression Lines

The scatter plot shows a strong positive relationship between reading and math scores for both groups, meaning that as reading scores increase, math scores tend to increase as well. In the lower score ranges (bottom left corner of the plot), students who did not complete the test preparation are more prevalent than those who did complete it. The regression line for the “completed” group appears to cut off more abruptly at the lower end, suggesting fewer very low-performing students in that group. However, the “none” group may be more spread out because it has almost twice as many data points (642) as the other group (358). With a larger sample size, you’re more likely to see extreme values and wider variability simply due to having more observations. Both regression lines have a similar upward slope, showing that reading and math performance move together similarly for both groups, and test prep does not significantly change the relationship between the two subjects.