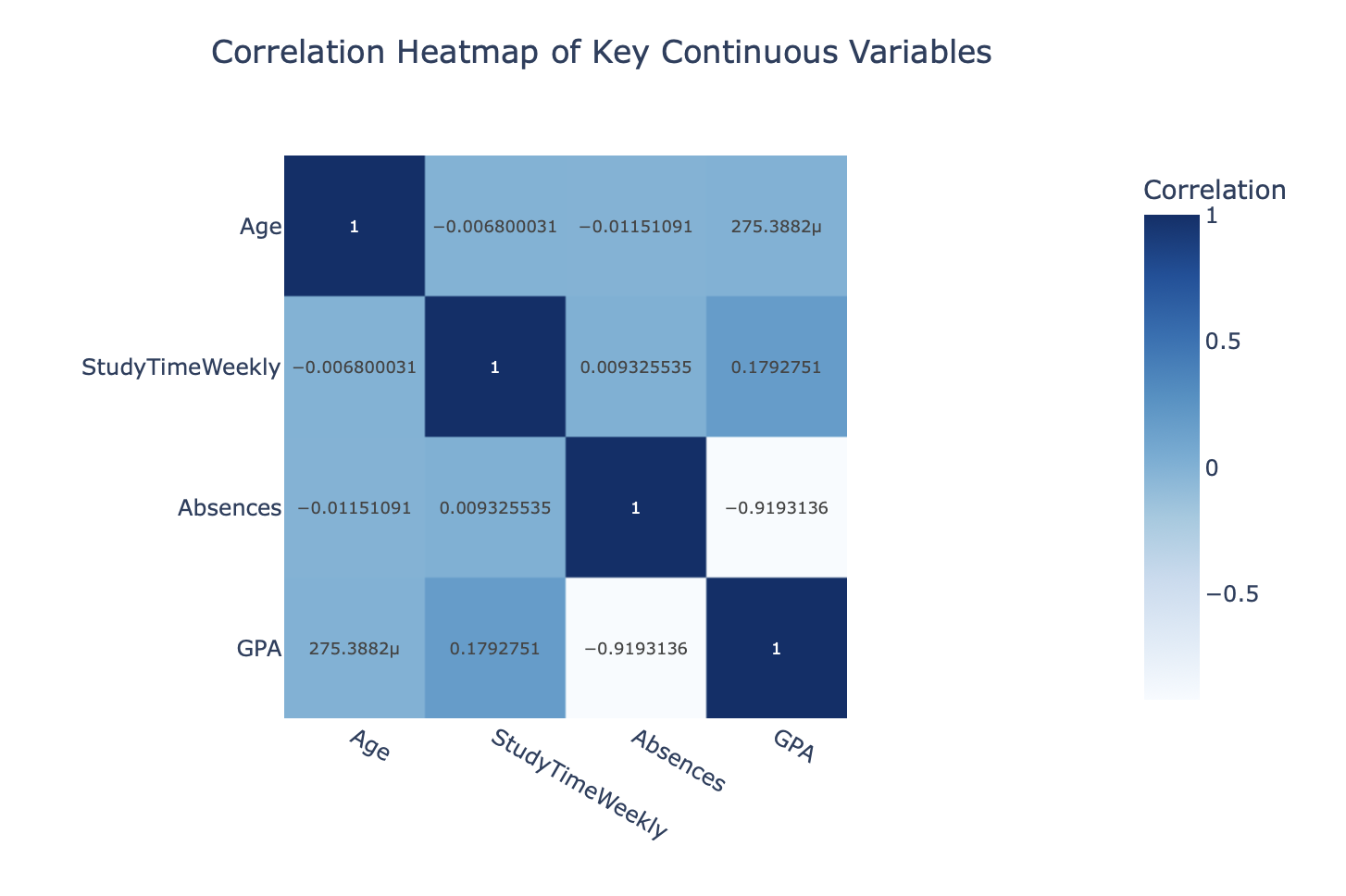
**Link to Published Website:** [**https://ninapatel3.github.io/DS4200-Project/data.html**](https://ninapatel3.github.io/DS4200-Project/data.html)

**Visualization 1:**

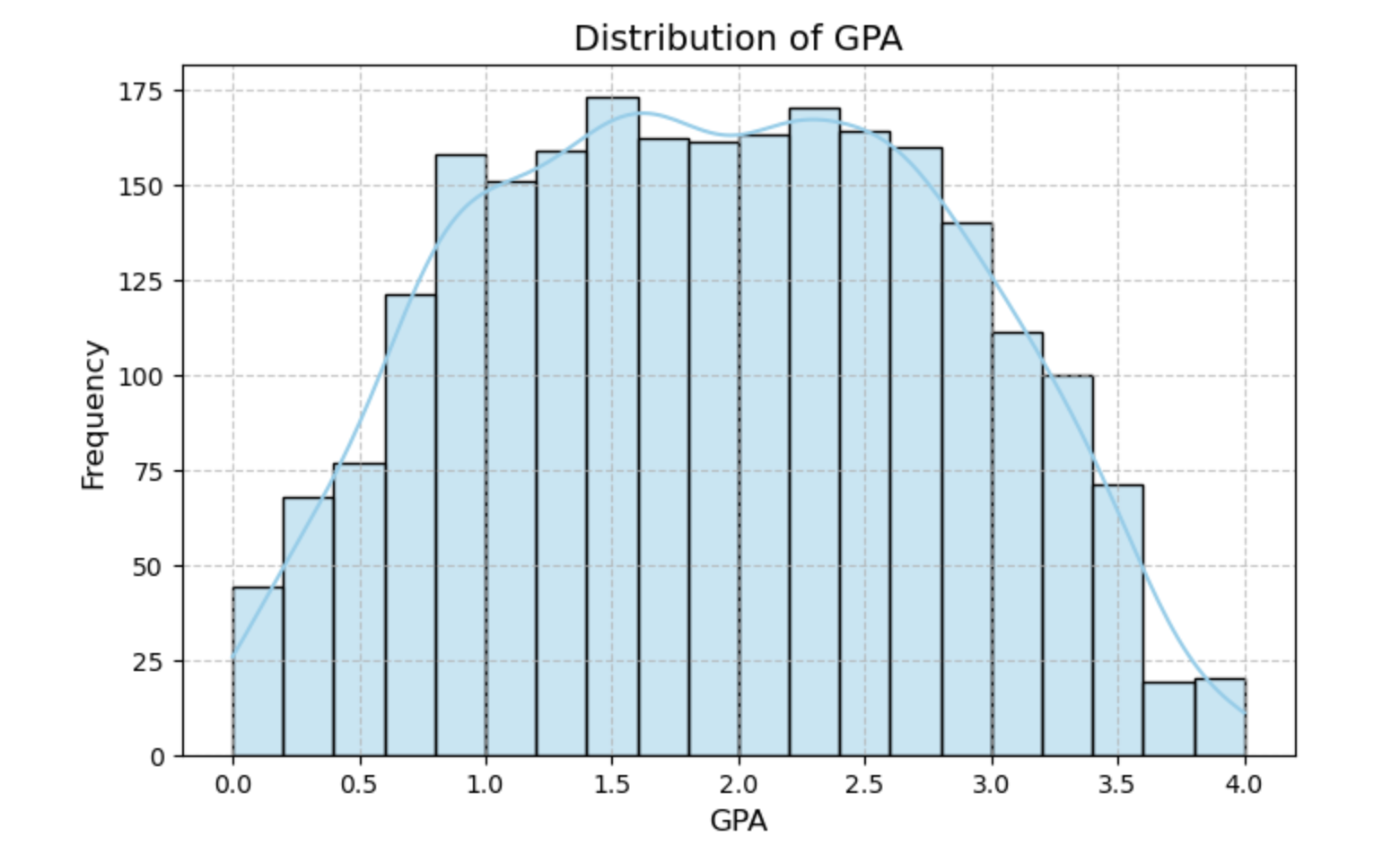
****

**Takeaway:** We begin with a correlation heatmap that illustrates the strength and direction of linear relationships between key continuous variables: Age, Study Time, Absences, and GPA. The most notable finding is a strong negative correlation between Absences and GPA (-0.92), indicating that students who miss more school tend to have significantly lower academic performance.There is also a weak positive correlation between Study Time and GPA (0.18), suggesting that while studying more may help, it has a relatively small effect compared to attendance. Meanwhile, Age shows virtually no correlation with any of the other variables, highlighting that age alone is not a major factor in academic performance within this dataset. Ultimately, the heatmap underscores that school attendance is a much stronger predictor of GPA than either age or weekly study time. We continue to explore these relationships through other visuals.

**Design Idea:**

A heatmap was chosen to provide a quick and intuitive overview of relationships between multiple continuous variables. The color gradient from deep blue (strong negative correlation) to bright yellow (strong positive correlation) makes it easy to spot both high and low values at a glance. Numeric correlation values are included directly in each cell to add clarity. Organizing the variables symmetrically along both axes allows for cross-checking relationships without confusion. This design is especially useful for identifying which variables may be worth exploring further in deeper statistical analyses or modeling.

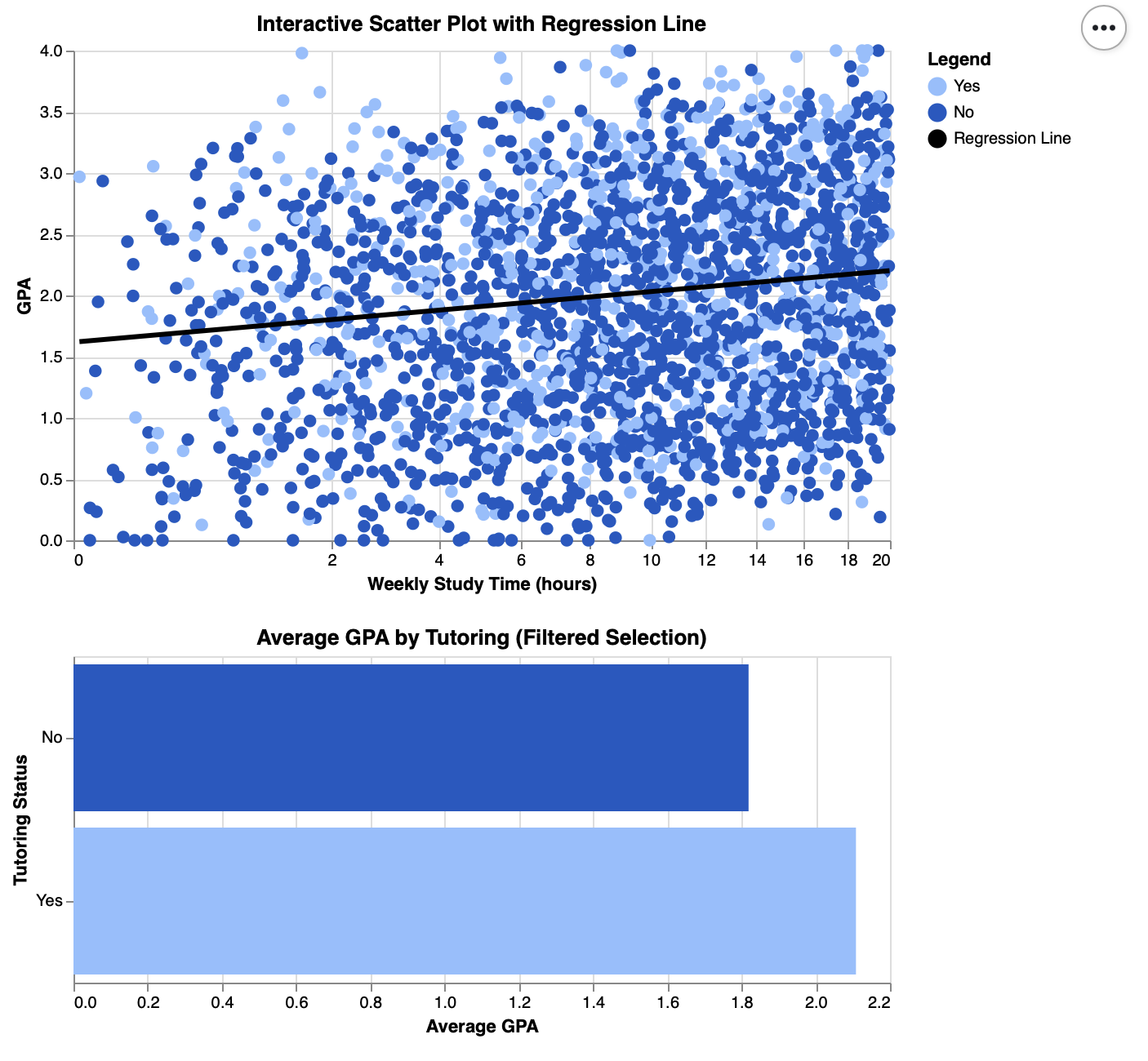
**Visualization 2:**

****

**Takeaway:** We see that the curve follows a roughly bell-shaped pattern, with most GPAs clustering between 1.5 and 3.0, and fewer students falling at the very low or very high ends of the scale. The shape of the distribution suggests a relatively normal academic spread, where most students perform near the average and extreme outcomes are less common. This pattern indicates that the dataset captures a typical section of student performance, making it a reliable foundation as we continue to draw conclusions in our analysis.

**Design Idea:** We chose a histogram to observe the shape of GPA distribution as we felt it would effectively display the frequency of GPA values across the dataset. Each bar represents a range of GPA values, allowing for easy identification of where most data points lie.

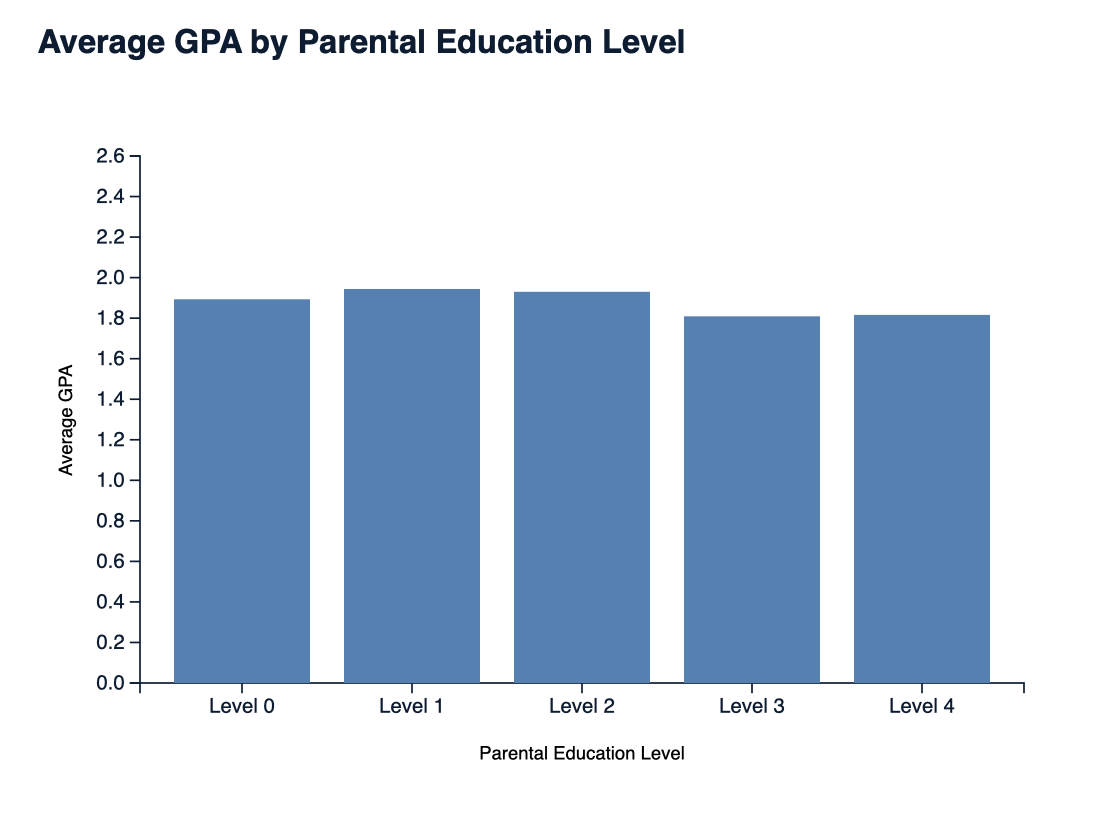
**Visualization 3:**

****

**Takeaway:** This interactive scatter plot explores the relationship between students’ weekly study time and their GPA, with tutoring status distinguished by color (blue for non-tutored, light blue for tutored). The black regression line shows a very weak positive linear relationship, indicating that students who study more hours per week tend to have slightly higher GPAs, though the effect appears minimal. Students at all study time levels display a wide spread of GPA outcomes, emphasizing that study time alone does not strongly predict academic performance. The scatter plot is linked to the bar chart below, which dynamically updates to show the average GPA by tutoring status for the filtered group. While students with tutoring appear to have slightly higher average GPAs, the significant overlap between tutored and non-tutored students suggests that tutoring is only one of many factors influencing academic success.

**Design Idea:** We chose a scatter plot to effectively visualize the distribution of students based on two continuous variables: weekly study time and GPA. By color-coding points based on tutoring status (orange for “Yes,” blue for “No”), the chart incorporates a third variable without overwhelming the viewer. To enhance interpretation, we included a regression line to show the overall relationship between study time and GPA. The interactive features, such as tooltips on hover and a brush selection tool, allow users to explore specific clusters and ranges in greater detail. Even axis scaling and clear labeling improve readability, especially in identifying patterns or the absence thereof. Overall, this interactive scatter plot provides a dynamic and intuitive way to examine how study habits and tutoring relate to academic performance.

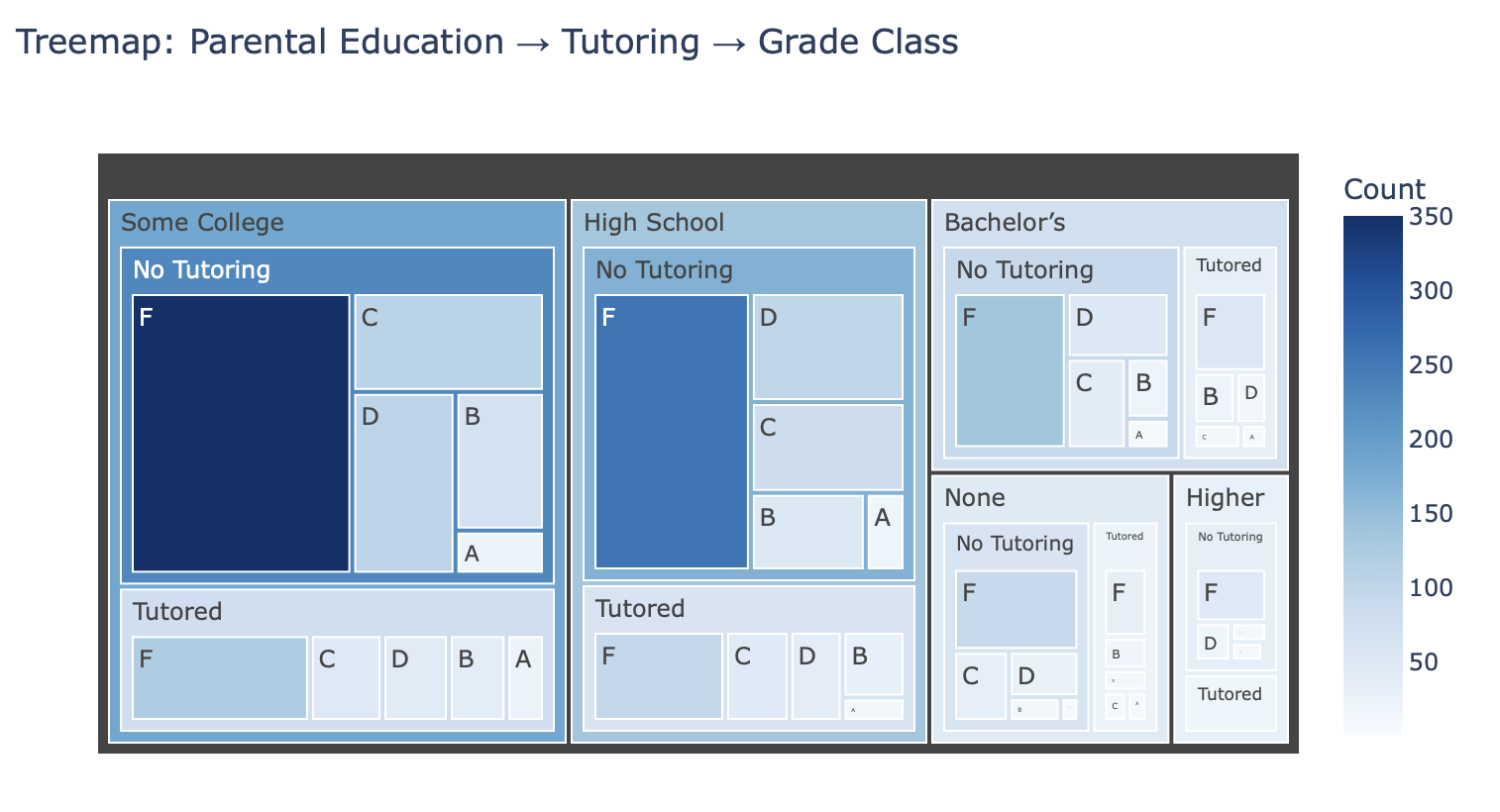
**Visualization 4:**

****

**Takeaway:** We next wanted to explore the average GPA of students based on their parents’ level of education, ranging from no formal education (Level 0) to graduate-level education (Level 4), and we did this with a bar chart. Surprisingly, the differences in GPA across these categories are minimal. Students whose parents completed higher education do not show significantly higher GPAs compared to those whose parents had no high school education. In fact, students whose parents completed only high school (Level 1) or some college (Level 2) show slightly higher average GPAs than those at the highest education levels. This finding challenges common assumptions about the direct impact of parental education on academic outcomes and suggests that other factors may play a more influential role in student performance.

**Design Idea:** A bar chart was chosen because they are effective in comparing average values across distinct, non-continuous categories. Each bar represents a different parental education level, making it easy to see how GPA varies across the spectrum. The y-axis is scaled to highlight subtle differences, and the consistent blue coloring ensures the visual remains simple and easy to interpret. Overall we chose this design to draw quick comparisons and prompt further investigation into why higher parental education does not necessarily correlate with higher student GPA.

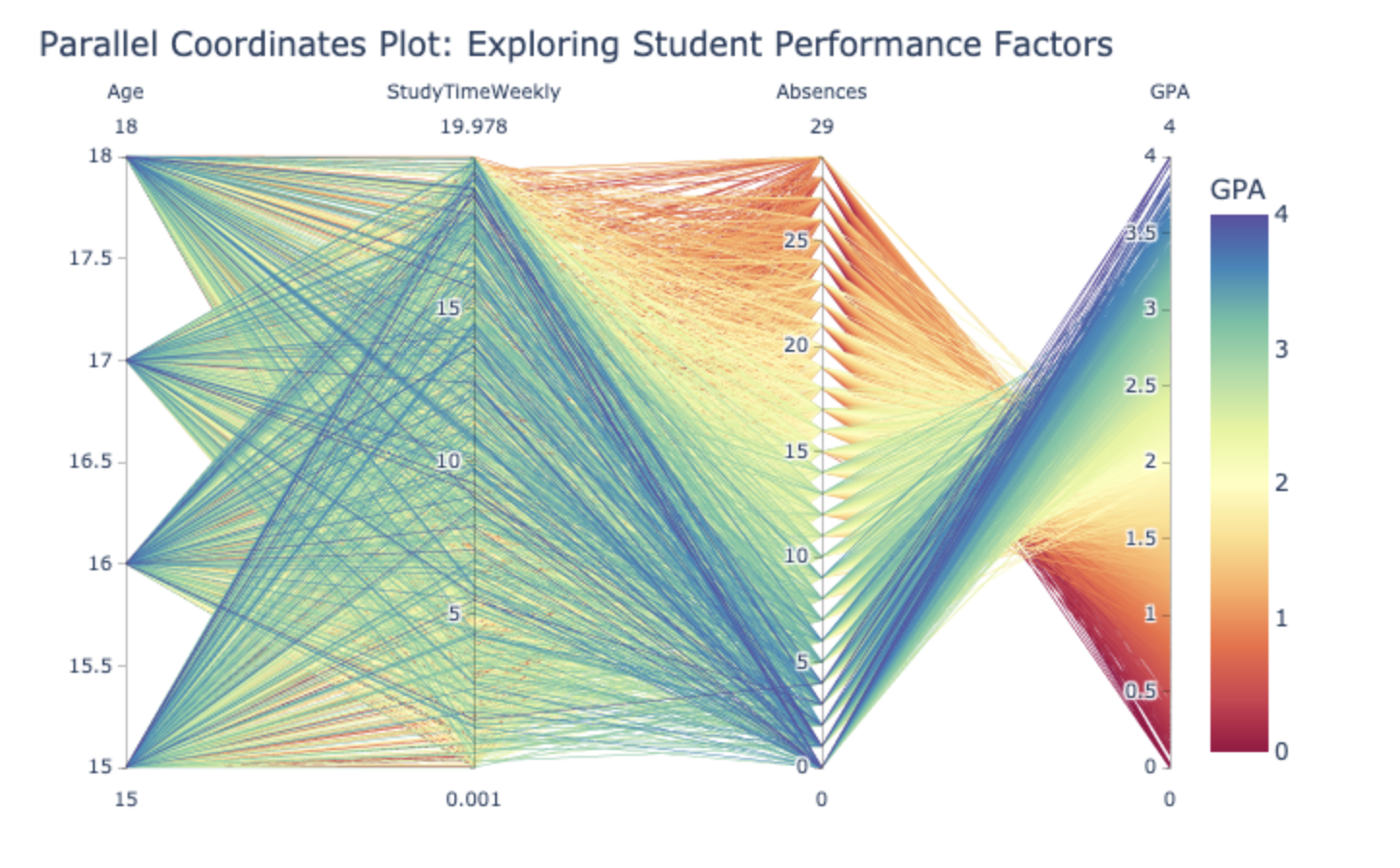
**Visualization 5:**

****

**Takeaway:** The treemap visualization breaks down students by parental education level, tutoring status, and final grade classification. A major takeaway is the dominance of "F" grades among students who did not receive tutoring, especially from parents with "Some College" or "High School" education levels. This suggests that tutoring may play a role in mitigating poor academic outcomes regardless of parental education. Although "F" grades still appear among tutored students, their distribution is more balanced across other grade levels (A-D), highlighting a potential benefit from academic support. Notably, students with parents holding Bachelor's or Higher degrees still show instances of lower performance, indicating that parental education alone may not be sufficient to ensure academic success.

**Design Idea:** A treemap was selected because it clearly communicates the breakdown of student performance across multiple categorical variables, parental education level, tutoring status, and grade classification, in a compact, hierarchical format. Each block’s size reflects the number of students in that category, allowing patterns in academic performance to stand out visually. By organizing the treemap into nested layers, viewers can easily follow how parental education flows into tutoring access and ultimately into academic outcomes. The use of color reinforces the distribution of students by count, helping draw attention to areas with a high concentration of "F" grades. This design makes it easier to compare the impact of tutoring across different education backgrounds while maintaining readability and structure. It effectively balances complexity and clarity, making nuanced relationships between categories more accessible to viewers.

**Visualization 6:**

****

**Takeaway**: This parallel coordinate plot visualizes how multiple continuous variables, Age, Study Time, Absences, and GPA, interact at the individual student level. Each line represents a student, with their values shown across vertical axes. The lines are color-coded by GPA using a gradient scale, allowing patterns to emerge visually across multiple dimensions. A clear trend is the inverse relationship between absences and GPA: students with few absences (near 0) often have higher GPAs, while those with more absences tend to fall into the lower GPA range. There is also a modest trend of higher study time aligning with slightly better GPA outcomes, though the variation is less pronounced. As observed in previous plots, age shows minimal correlation with GPA. This plot reinforces earlier findings that school attendance plays a significant role in academic performance, while also helping identify clusters of students with similar academic profiles.

**Design Idea**: A parallel coordinate plot was selected to display multi-dimensional relationships between continuous variables for each student. Each line represents a student, and the line’s path across each vertical axis shows their value for each variable. By coloring the lines based on GPA using a gradient, the visualization allows viewers to detect patterns across variables and see which traits are commonly associated with higher or lower academic performance. This design is especially useful for identifying complex interactions and trends that wouldn’t be visible in simple two-variable plots.