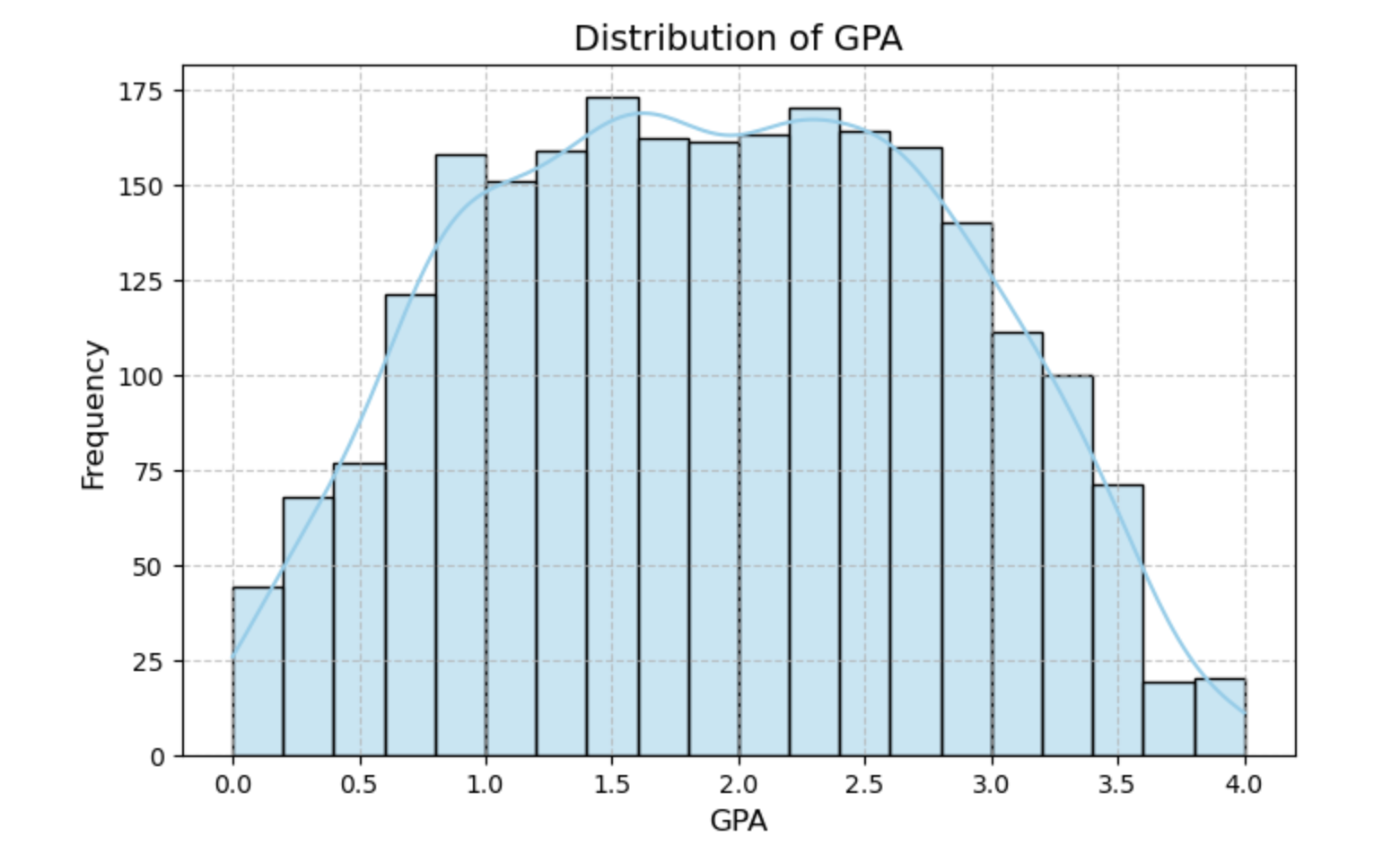
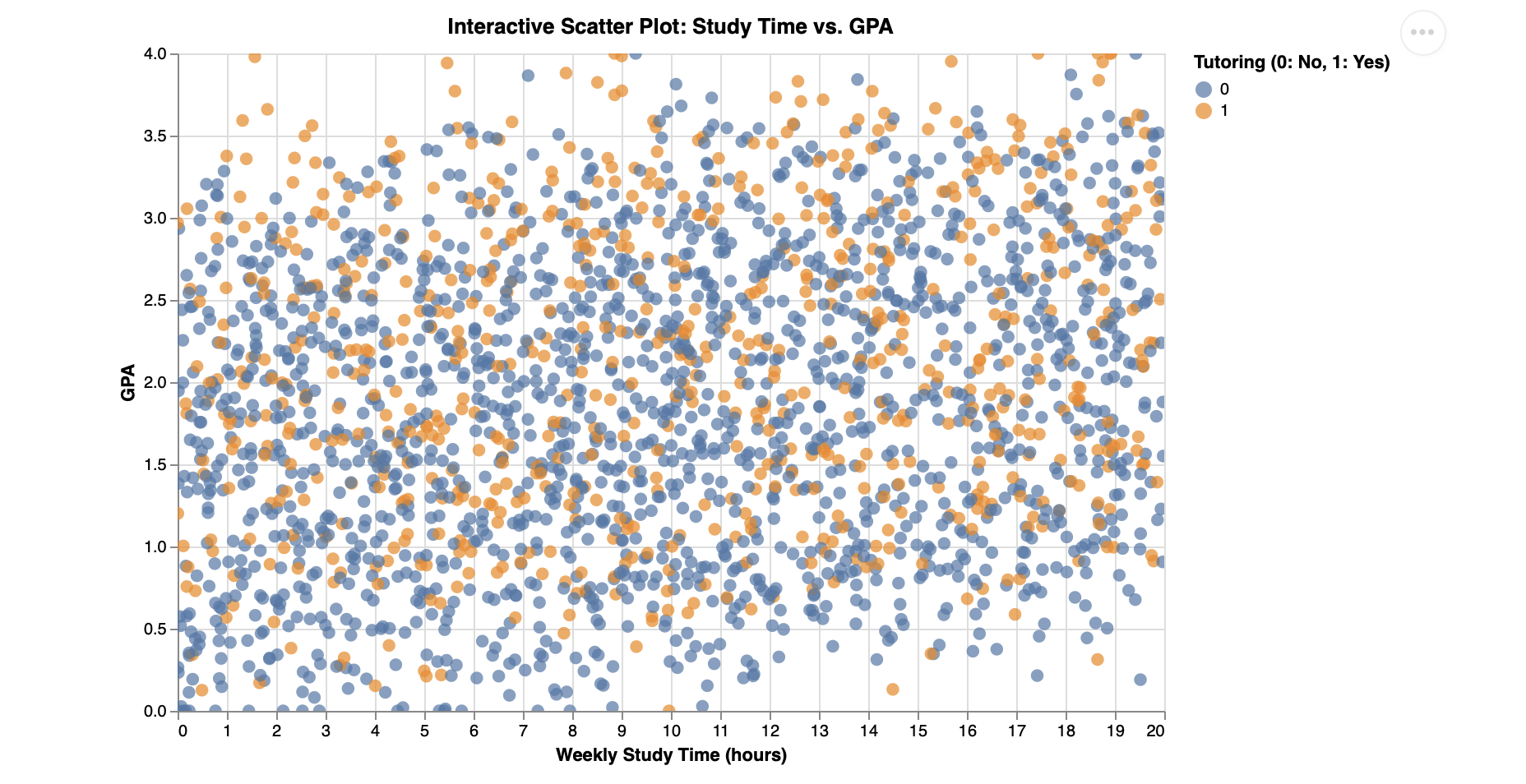
**Link to Published Website:** [**https://ninapatel3.github.io/DS4200-Project/data.html**](https://ninapatel3.github.io/DS4200-Project/data.html)

**Visualization 1:**

****

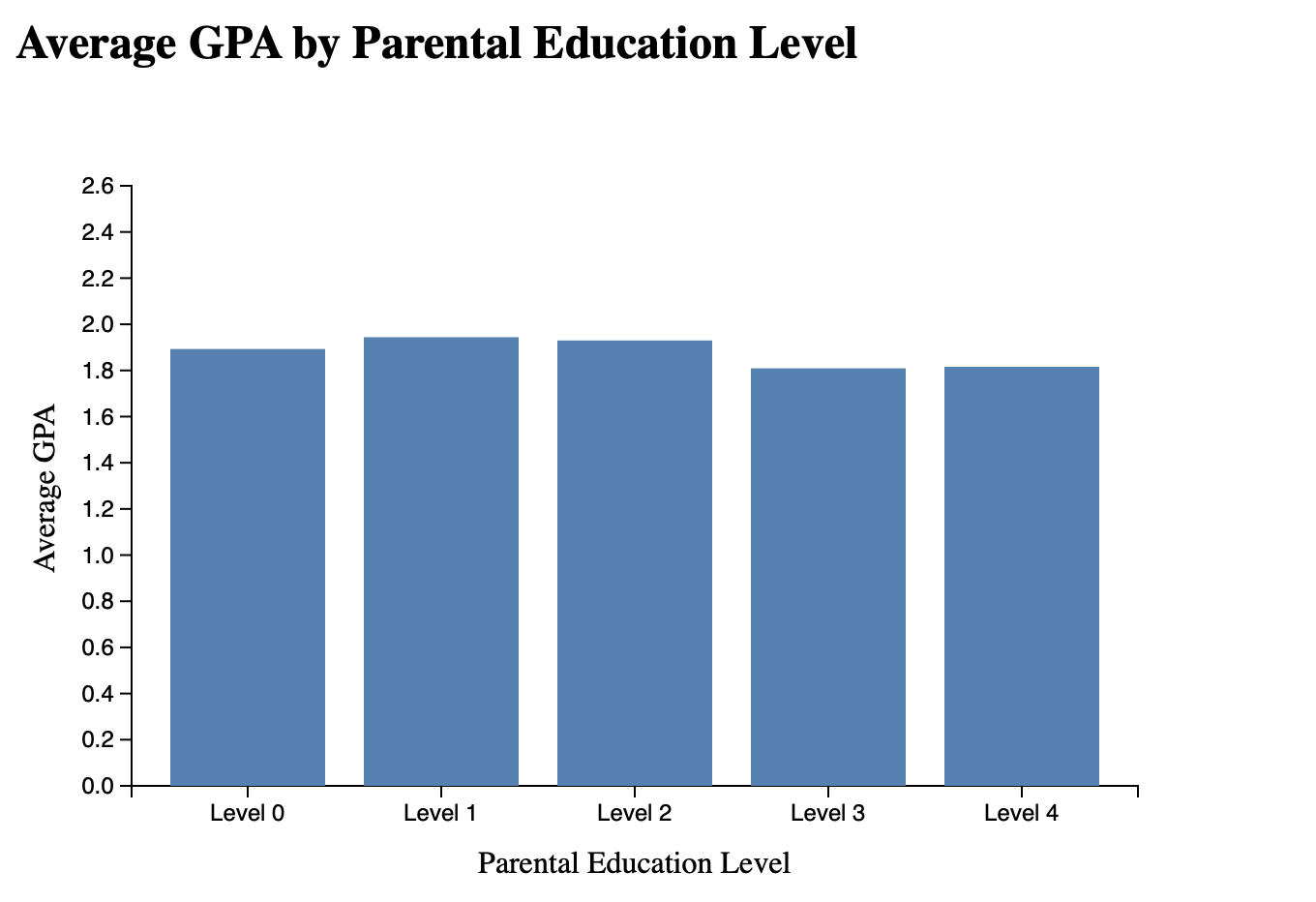
**Design Idea:** We chose to start with a histogram to observe the shape of GPA distribution as we felt it would effectively display the frequency of GPA values across the dataset. Each bar represents a range of GPA values, allowing for easy identification of where most data points lie.

**Visualization 2:**

****

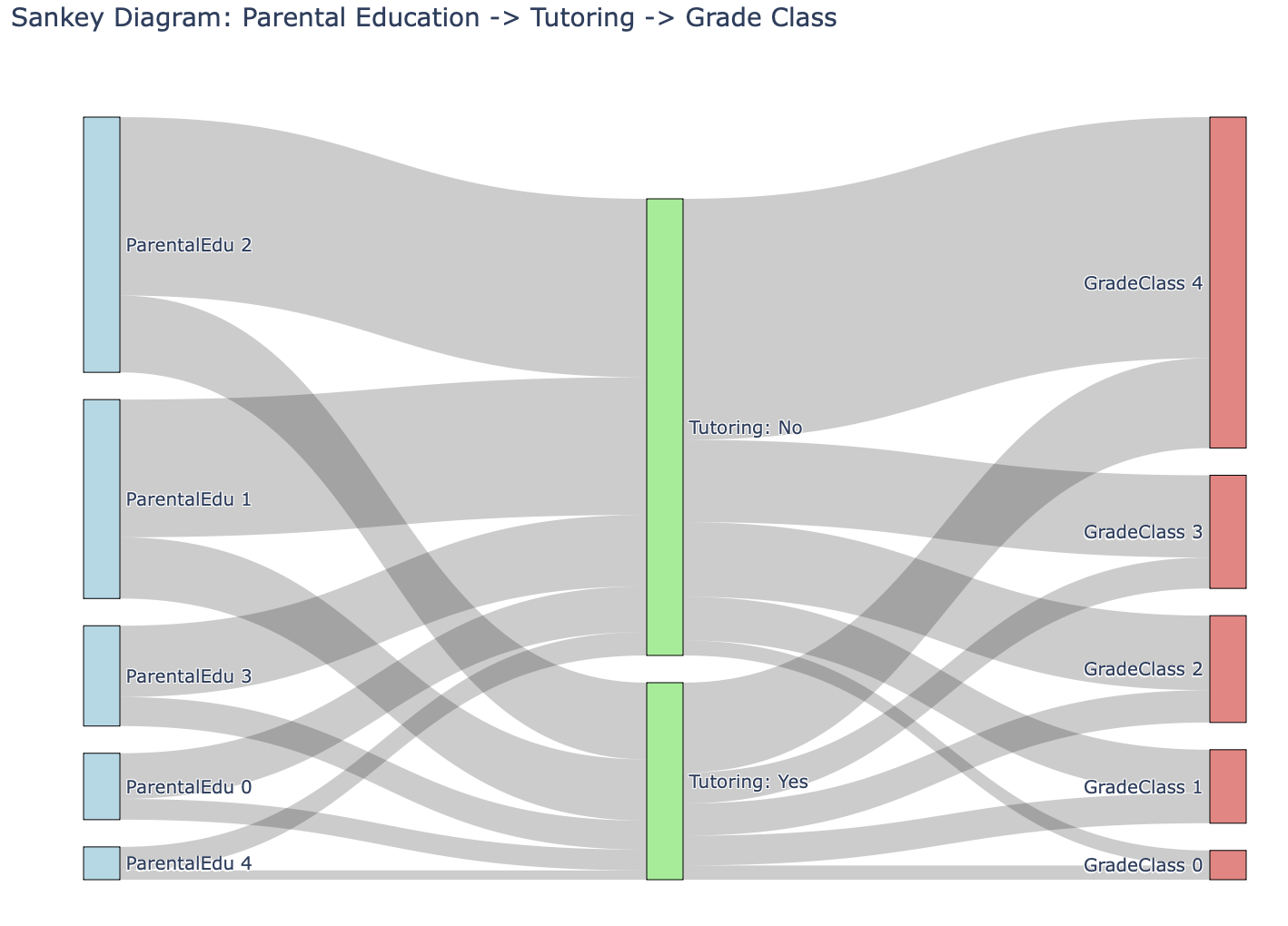
**Design Idea:** We chose to create a scatter plot to effectively illustrate the individual distribution of students based on two continuous variables: study time and GPA. By color-coding the points based on tutoring status (orange for “Yes,” blue for “No”), the scatter plot easily adds a third variable without crowding the graph.The interactive element of this graph allows users to hover over points for more detail. The even scaling and clean labeling of axes help make the patterns, or in this case the lack thereof, more readable. Overall this interactive scatter plot is a good tool for highlighting multiple variables and dispersion in the data.

**Visualization 3:**



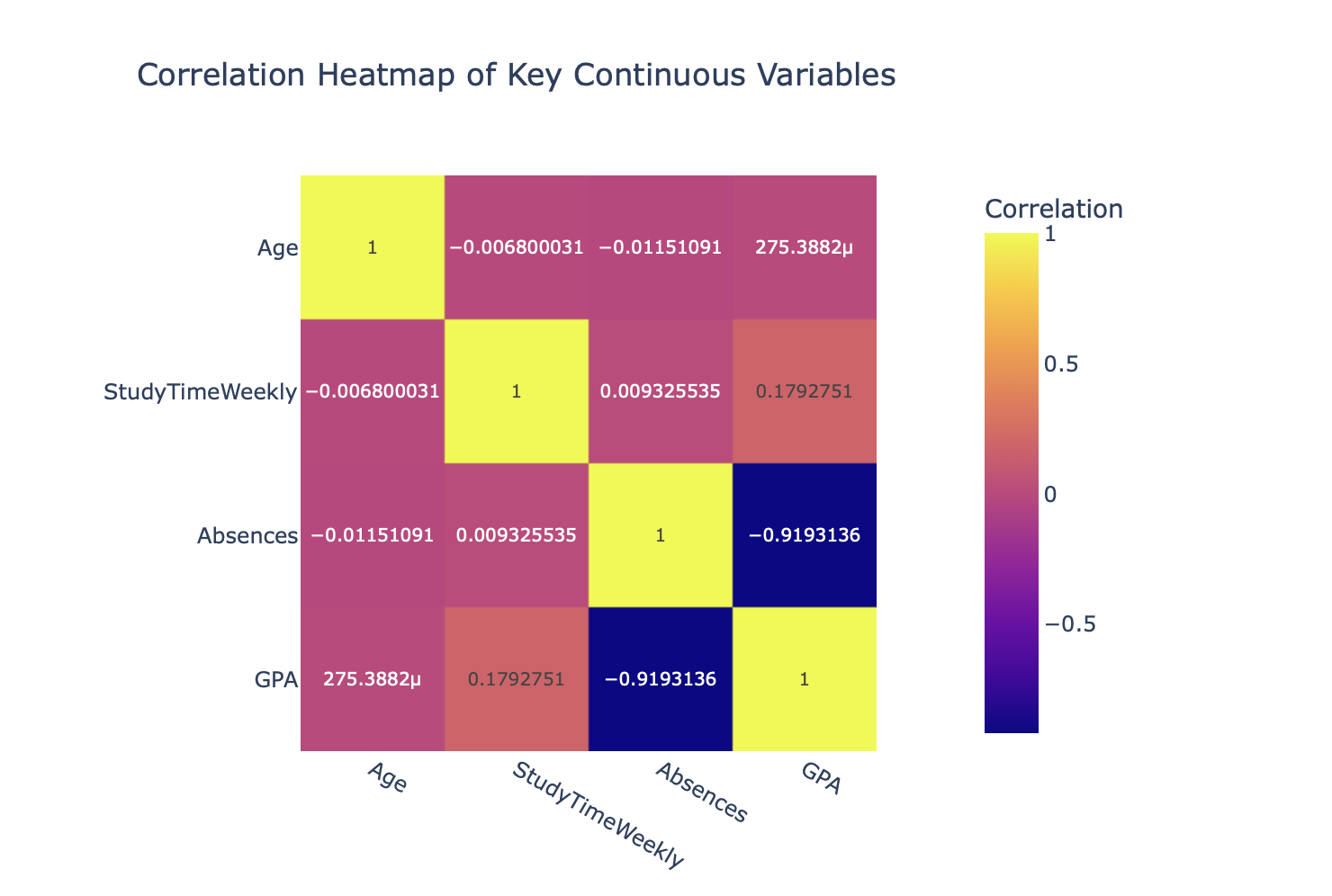
**Design Idea:** A bar chart was chosen because they are effective in comparing average values across distinct, non-continuous categories. Each bar represents a different parental education level, making it easy to see how GPA varies across the spectrum. The y-axis is scaled to highlight subtle differences, 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 4:**

****

**Design Idea:** A Sankey diagram was selected because it visually captures the flow of students through a series of categorical stages, allowing viewers to trace how parental education connects with grade class. The width of each band represents the number of students at each transition, emphasizing volume and making patterns in the data more intuitive. By splitting the middle node into “Tutoring: Yes” and “Tutoring: No,” the diagram adds an important intermediary step that helps explain differences in academic results. Colors are used strategically to distinguish source, intermediate, and outcome categories while keeping the visualization readable and structured. This design helps communicate complex relationships in a format that is both complex and understandable.

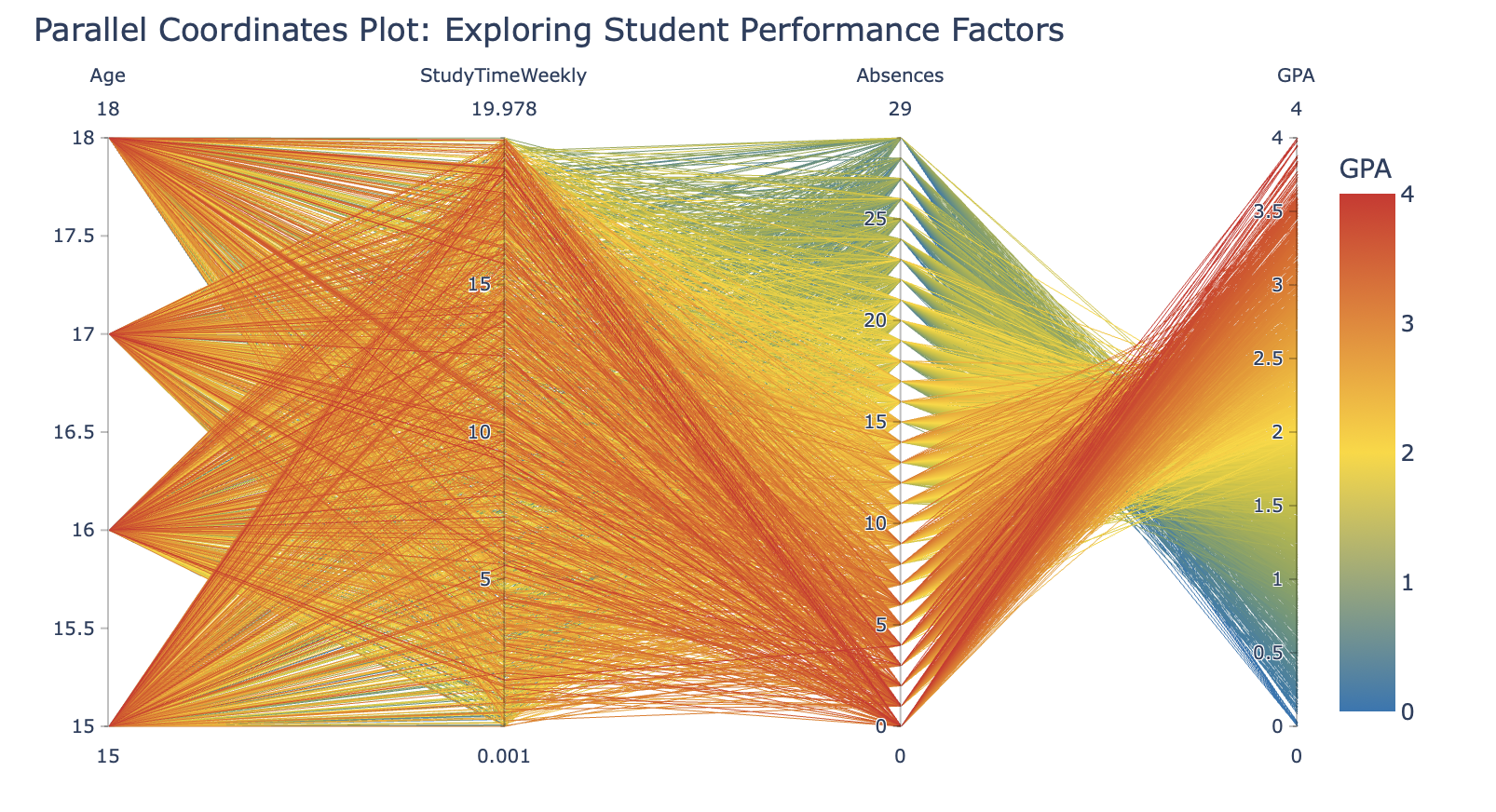
**Visualization 5:**

****

**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 6:**

****

**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.