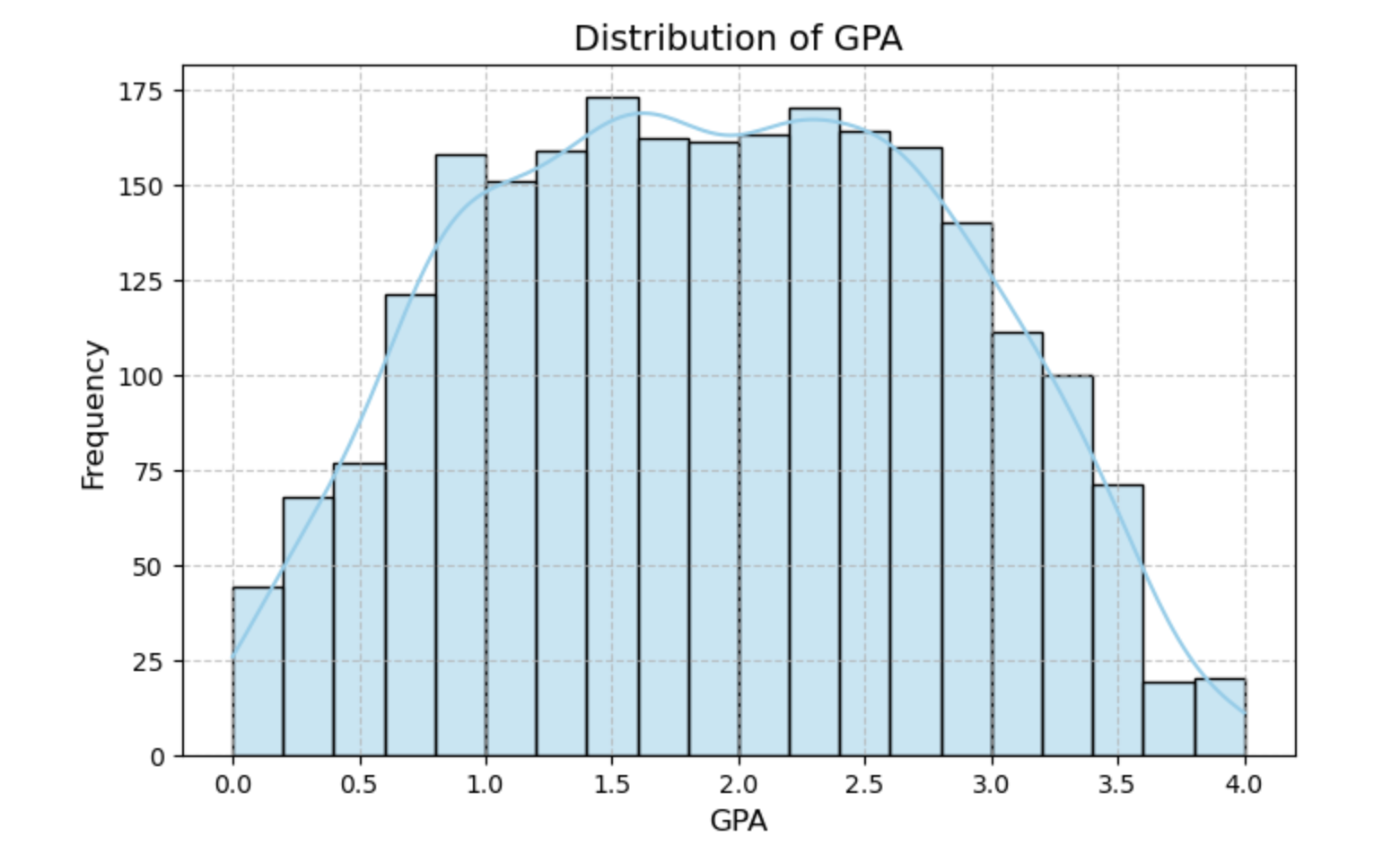
**Link to Published Website:** [**https://ninapatel3.github.io/DS4200-Project/data.html**](https://ninapatel3.github.io/DS4200-Project/data.html)

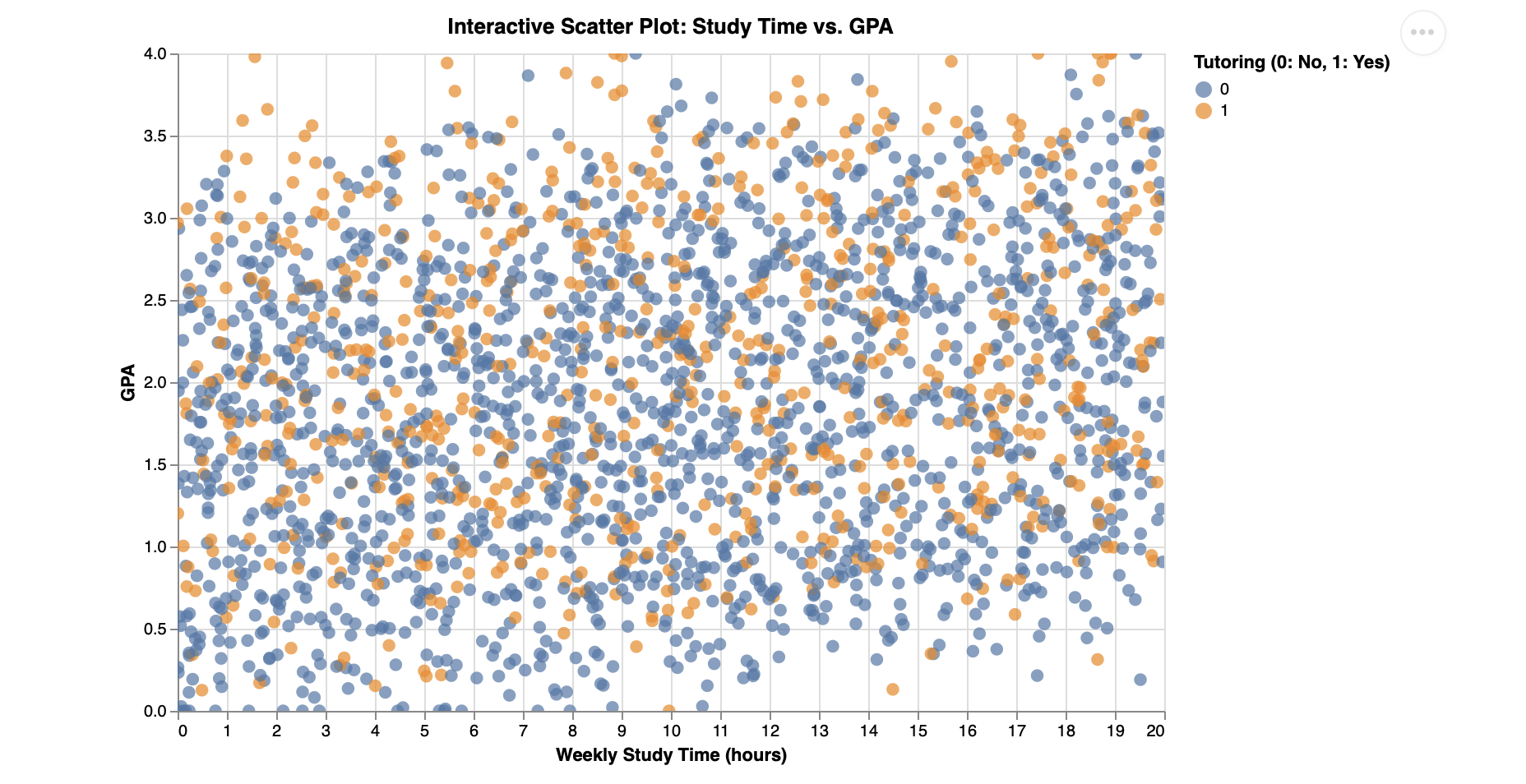
**Visualization 1:**

****

**Takeaway:** To begin we started with a histogram displaying the distribution of GPA to get an overview of academic performance among the students in the dataset. The distribution appears to be mainly normal, and most students have GPAs centered between 1.5 and 3.0, with fewer students at the very low or very high ends of the spectrum. This bell-shaped curve suggests that the dataset represents a typical academic population, where a majority of students perform around the average, and extreme outcomes are relatively rare. Understanding that this dataset represents a well rounded range of the population helps us to make overall conclusions about the data.

**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.

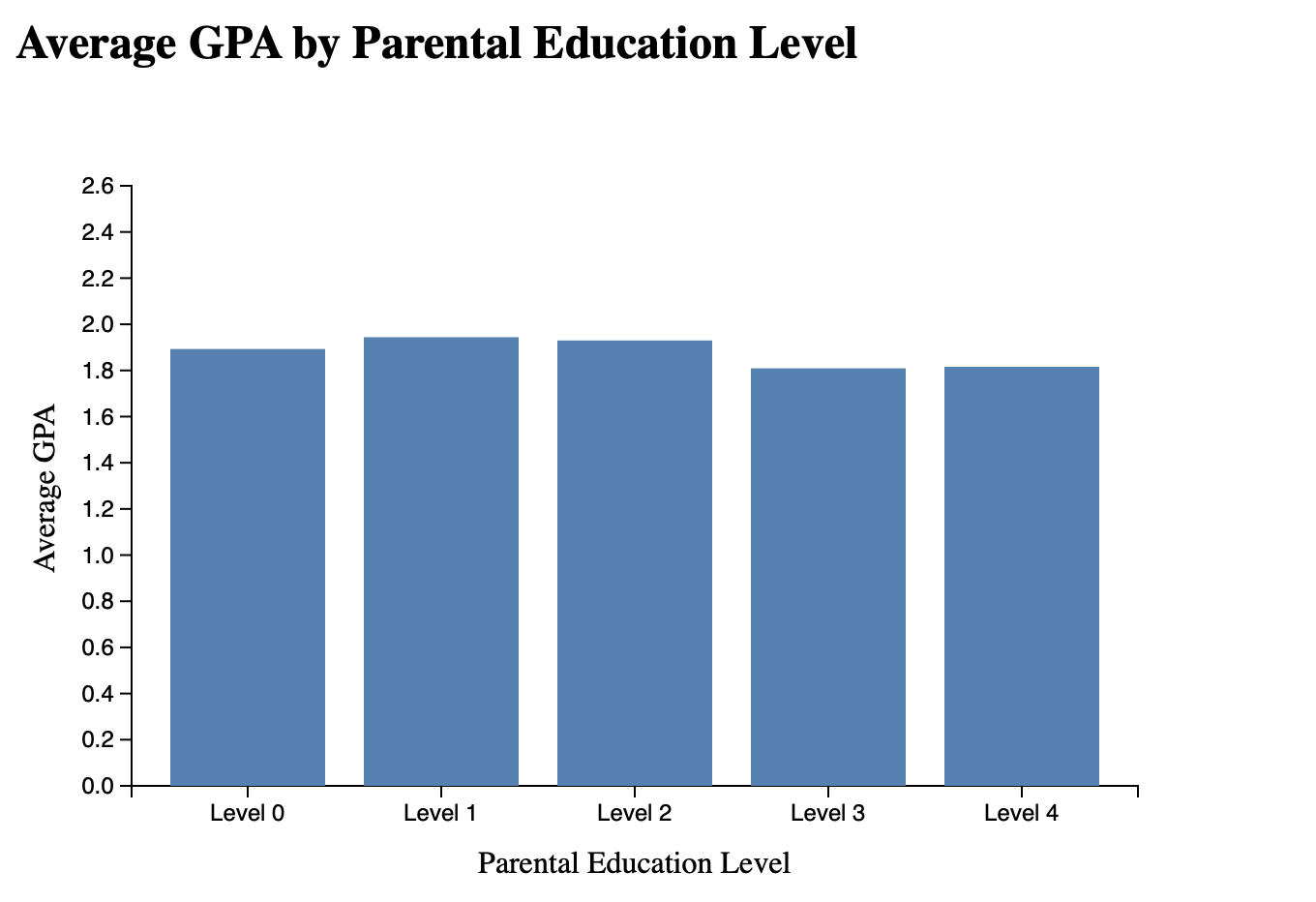
**Visualization 2:**

****

**Takeaway:** This scatter plot examines the relationship between students' weekly study time and their GPA, with tutoring status represented through color (blue representing having been tutored). Overall, the data reveals a very weak linear relationship between study time and GPA. Students across all study time levels display a wide range of GPAs, from low to high, suggesting that simply studying more hours does not necessarily predict stronger academic performance. Additionally, while there appears to be a slight concentration of higher GPAs among students with tutoring (orange), the overlap between tutored and non-tutored students remains visually substantial. This indicated to us that while tutoring may play a role in performance, it is likely not the sole determining factor.

**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 tol for highlighting multiple variables and dispersion in the data.

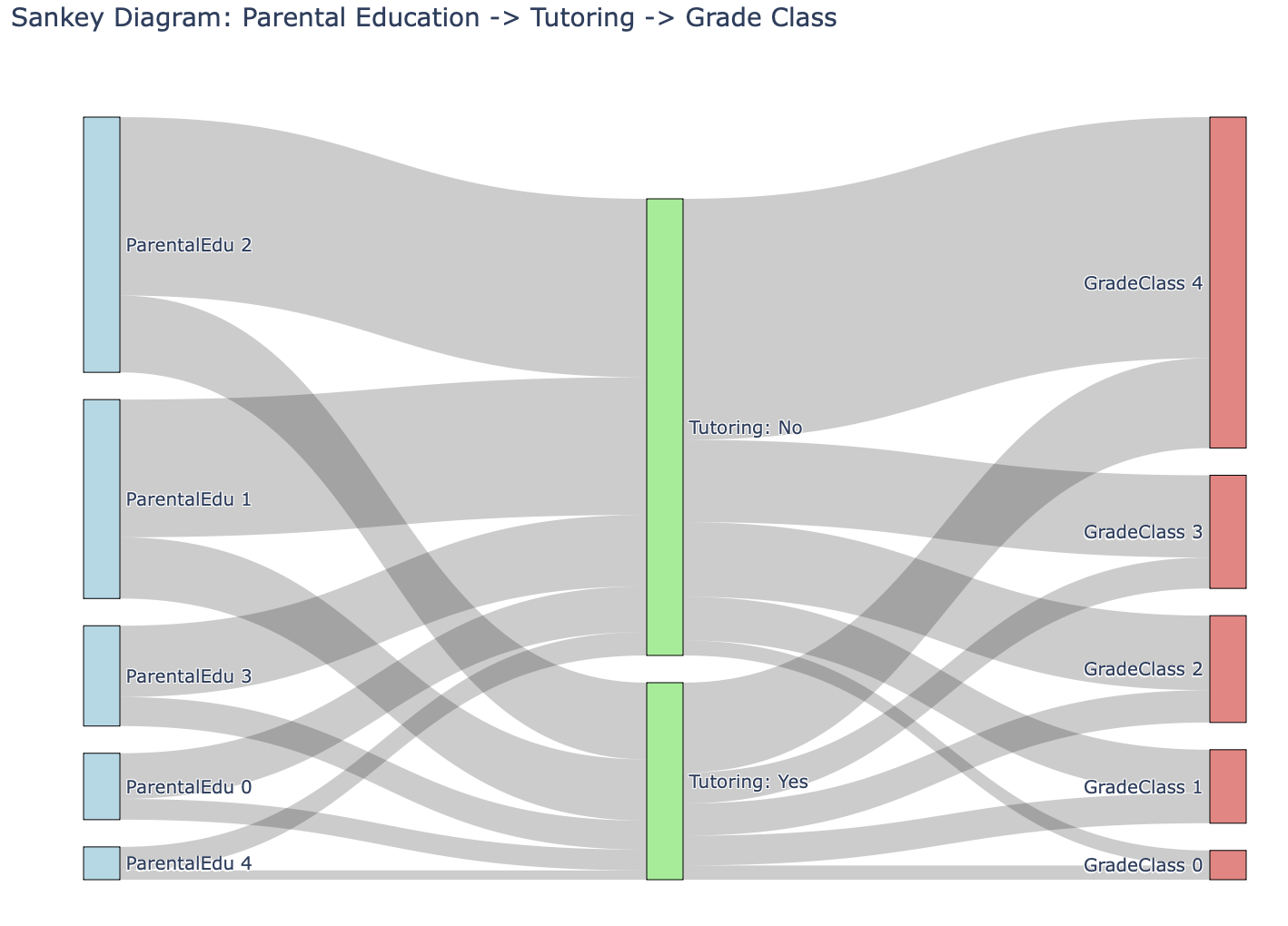
**Visualization 3:**



**Takeaway:** We next wanted to explore the average GPA of students based on their parents’ level of education, ranging from no formal education to higher education. Surprisingly, the chart reveals minimal variation in GPA across all education levels. Students whose parents completed college or higher education do not appear to have significantly higher GPAs than those whose parents did not complete high school. In fact, students with parents who only completed high school or some college show slightly higher GPAs on average. This challenges common assumptions about the direct influence of parental education and suggests that other variables may play a more important role in academic 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.

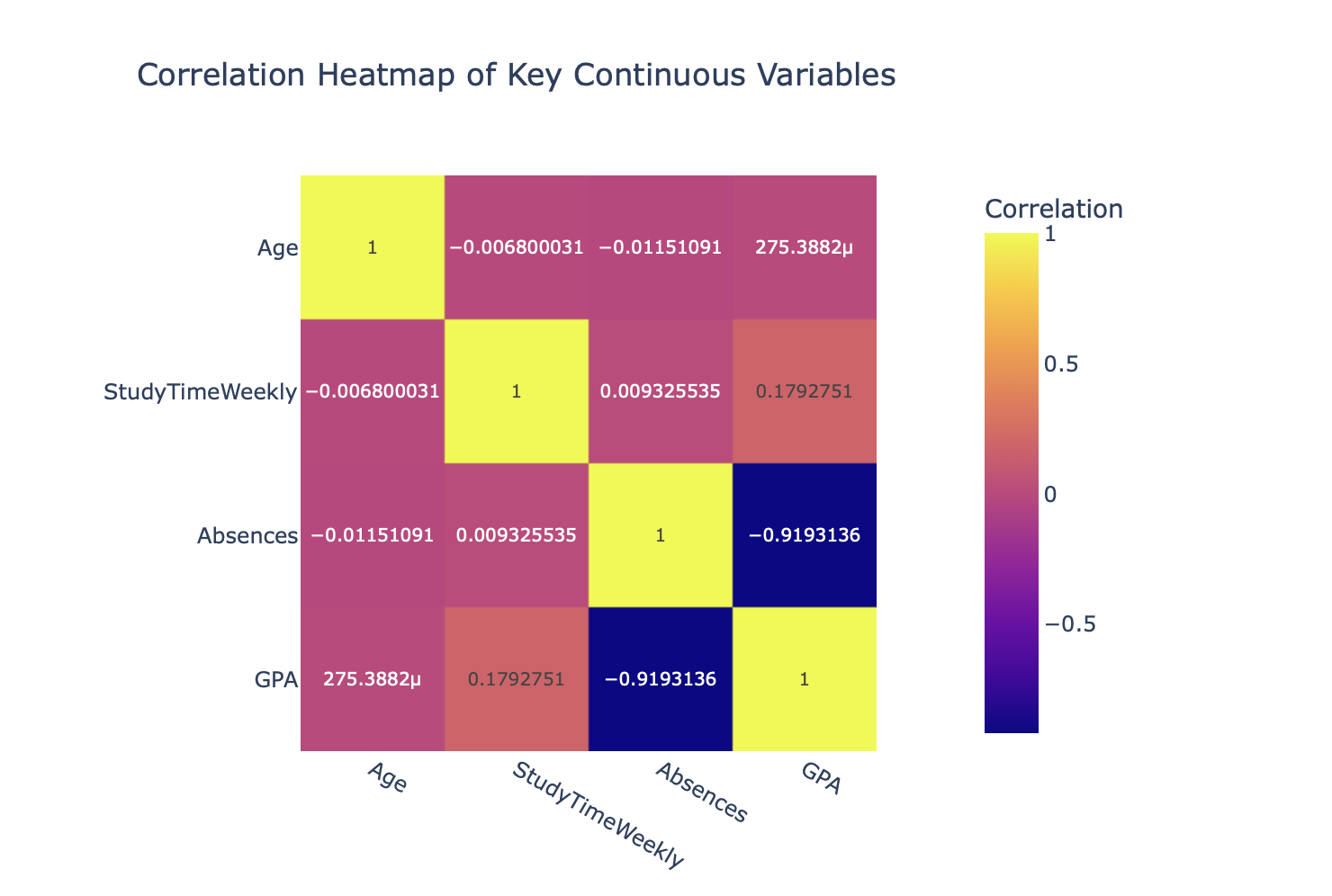
**Visualization 4:**

****

**Takeaway:** Continuing our exploration of parental education level effect on GPA, this Sankey diagram traces the pathway from parental education levels through tutoring status to students’ grade classifications, offering a layered look at how background and support factors might influence academic outcomes. One notable observation is that a large number of students from lower to mid-level parental education groups do not receive tutoring, and a substantial portion of those students end up in the lower grade categories. In contrast, students who do receive tutoring are more evenly distributed across all grade levels, with some representation in the higher performance bands. While the diagram doesn’t show a direct cause-effect relationship, it highlights how tutoring is a strong support mechanism for students, potentially improving outcomes regardless of parental education level.

**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:**

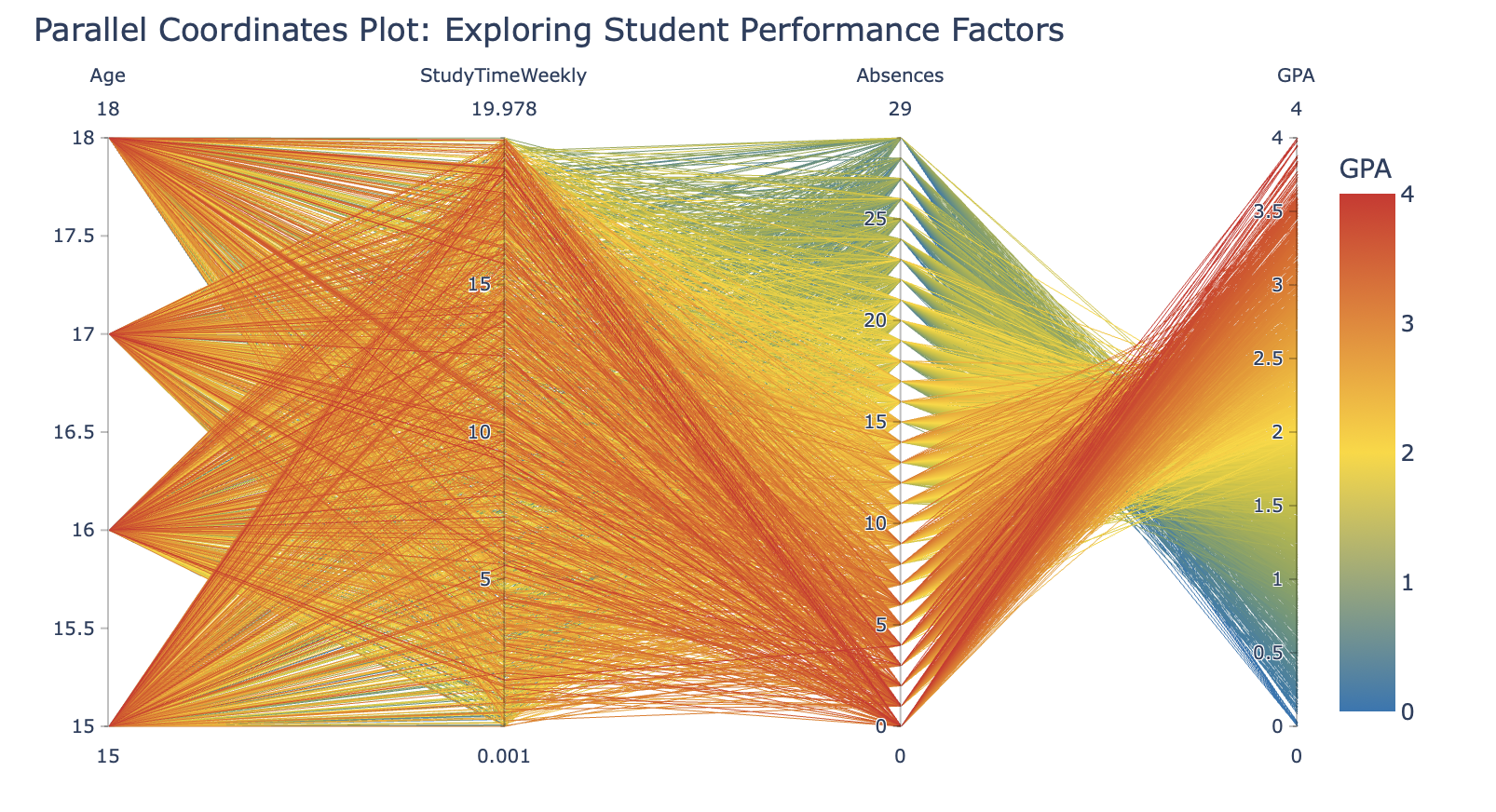
****

**Takeaway:** This visualization, a correlation heatmap reveals the strength and direction of linear relationships between key continuous variables: Age, Study Time, Absences, and GPA. An important observation is the strong negative correlation between Absences and GPA (-0.91), indicating that students with more absences tend to have significantly lower GPAs. Not surprisingly, there is a weak positive correlation between Study Time and GPA (0.18), suggesting that studying more may have a small positive impact on academic performance, something we observed in the earlier scatter plot. Age appears to have no meaningful relationship with any of the other variables. Overall, the heatmap emphasizes that school attendance is a more critical factor in GPA than either age or weekly study time alone.

**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:**

****

**Takeaway**: The parallel coordinates plot visualizes how multiple continuous variables, Age, Study Time, Absences, and GPA, interact on an individual student level. A key pattern that emerges is the relationship between Absences and GPA: students with low absences (near 0) tend to have higher GPAs, while those with more absences cluster around lower GPA values. Study time shows some positive variation, with a number of students who study more hours per week trending toward higher GPA values. Age, however, appears to have little variation in GPA, consistent with earlier findings. This plot reinforces the idea that attendance is a major contributor to academic performance.

**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.