

Introduction to Stochastics

Exercise 3: Group Work

1 Reading Plots

The following plots visualize certain characteristics of a synthetic dataset about student performance factors. Answer the questions about the plots.

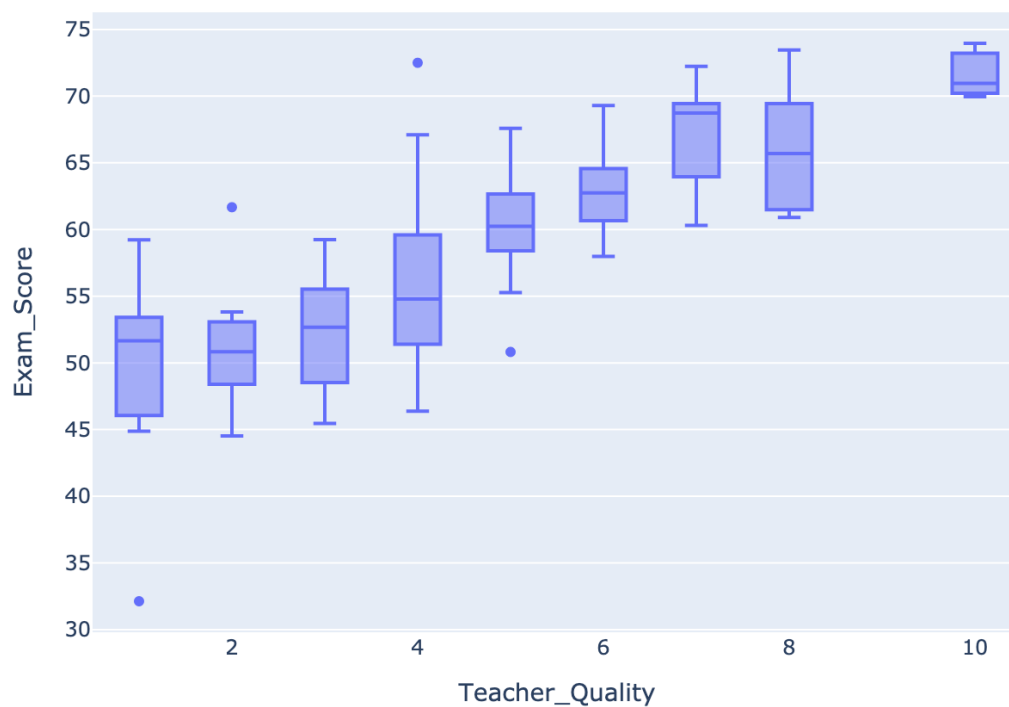


Figure 1: Boxplot

- What is the median exam score (approximately) for a teacher quality level of 4?
- Which teacher quality level contains outliers in its exam score distribution?

- Between teacher quality levels 2 and 8, which has a higher exam score?
- Estimate the five-number summary for teacher quality level 8 and 4.

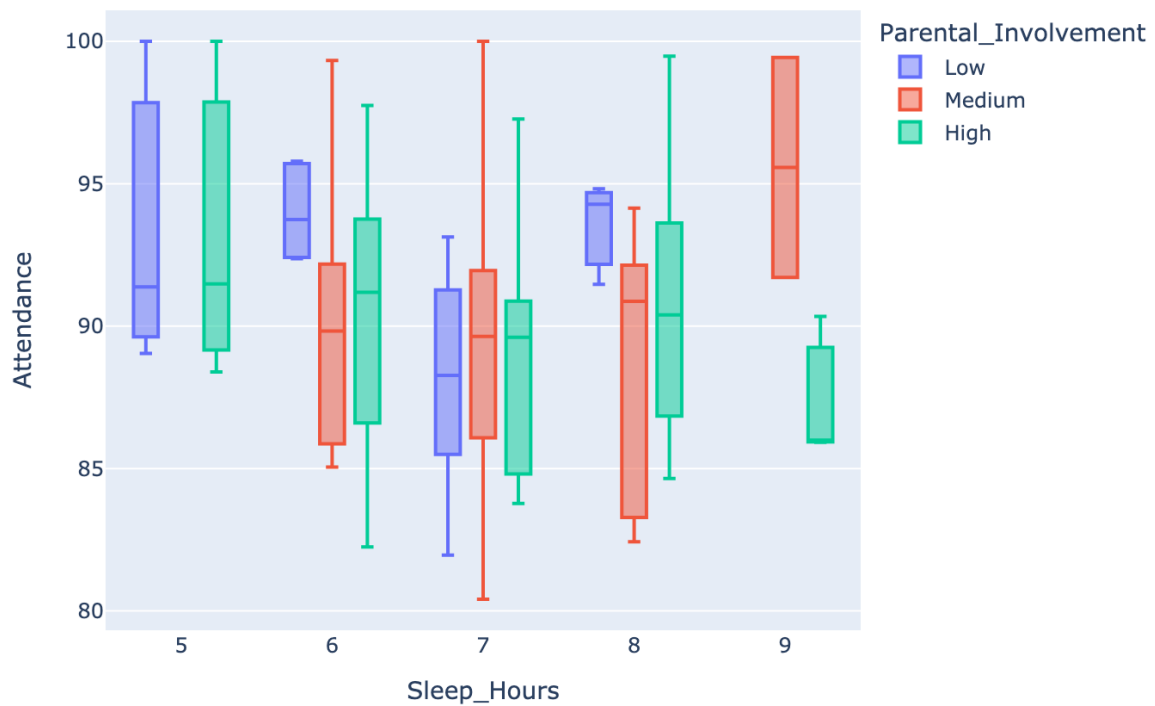


Figure 2: Boxplot 2

- Which parental involvement class includes students who sleep 5 hours?
- Which amount of sleep corresponds to the lowest median attendance rate when parental involvement is high?

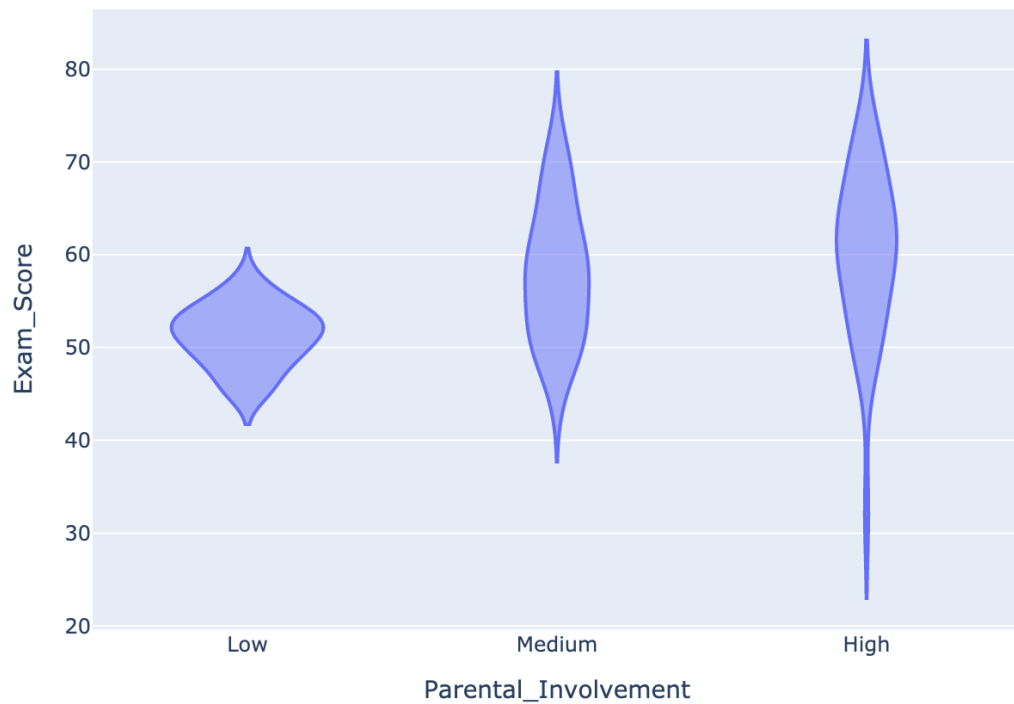


Figure 3: Violin plot

- Which parental involvement level shows the most spread in exam scores, and how can you tell from the shape of the violin plot?
- In which parental involvement category are exam scores most concentrated around a single value, and how does the shape (width) of the violin plot help you see this?
- Do you observe a possible relationship between parental involvement and exam score? Explain briefly.

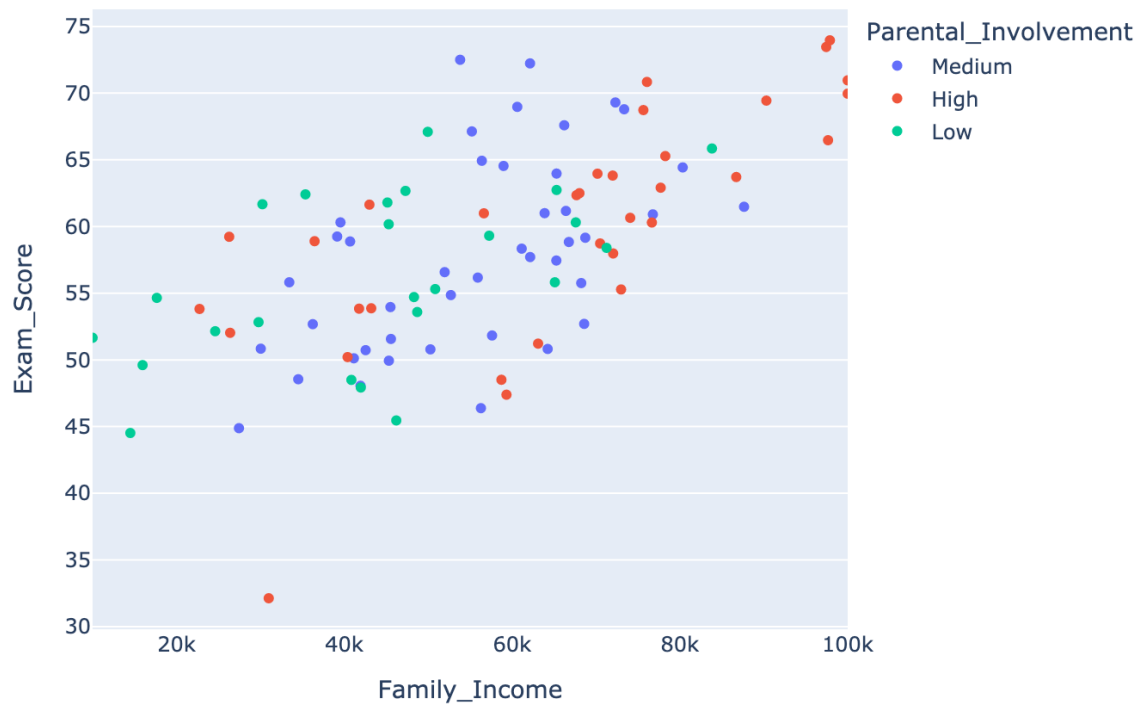


Figure 4: Scatter plot

- What is the highest exam score (approximately) achieved by a student whose parents earn between 20k and 40k?
- What is the highest exam score (approximately) achieved by a student whose parents earn between 80k and 100k, where parental involvement is low?
- Can you suggest a model that could explain the relationship between family income and exam score?

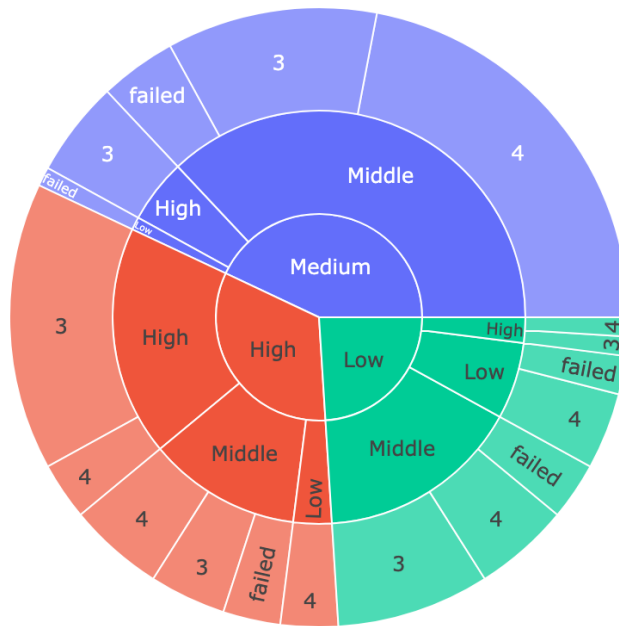


Figure 5: Sunburst chart: Parental_Involvement \rightarrow Income_Level \rightarrow Grade

- Which is the most common parental involvement level?
- Which is the most common grade per income level for medium parental involvement?
- List all combinations of parental involvement and income level where students failed the exam.