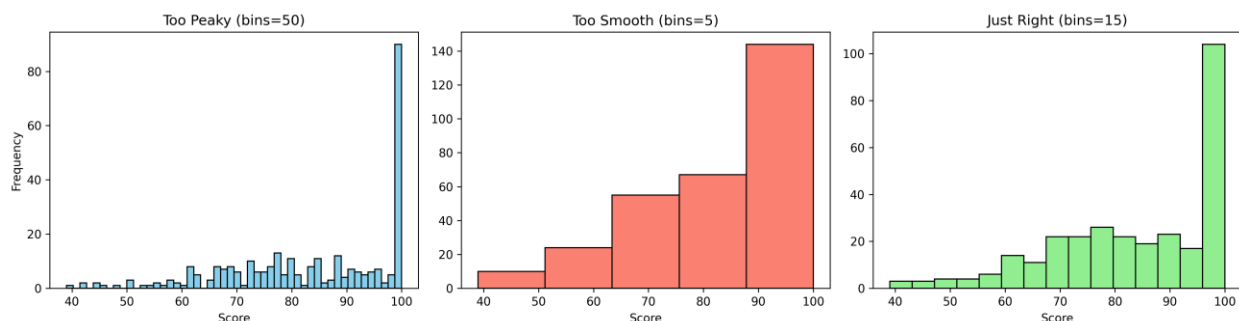


1 Exercises (100 pts)

1. (25 pts) Create three histograms for their scores: one that's too peaky, one that's too smooth, and one that is just right. Justify your choice of bin width.

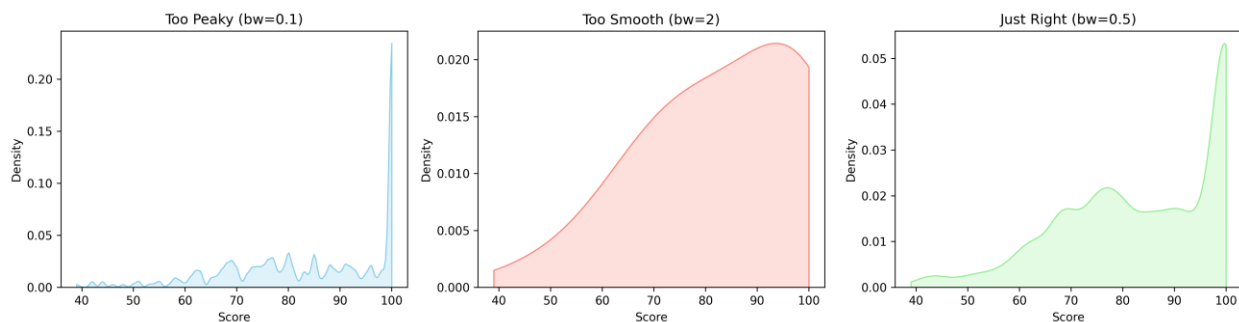


The histogram on the left is too peaky. There are too many bars that only cover a very small range of scores. This histogram looks spiky and uneven, implying that there's a lot of ups and downs in the data when that's really not the case.

The histogram in the middle is too smooth. There are too many different scores that fit in the same bin. This hides important patterns that we could locate with a smaller bin width.

The histogram on the right is just right. It is clear enough to show the general trend without oversimplifying the data. It shows the variance in the students' performance without unnecessarily making too many jagged edges like the too peaky. This is the best chart to use for interpretation.

2. (25 pts) Create three density plots for their scores: one that's too peaky, one that's too smooth, and one that is just right. For each plot trim the tails that extend beyond the range of values. Justify your choice of bandwidth.



The density plot on the left is too peaky. It makes the whole graph overwhelming to look at. It also makes the graph harder when trying to identify patterns.

The density plot in the middle is too smooth. It hides patterns that we could find easier in a smaller bandwidth.

The density plot on the right is just right. The plot shows patterns without overwhelming the reader. This plot also does a good job at highlighting certain patterns and does not round them out therefore misleading the reader.

3. (50 pts) Assume that the course instructor plans to develop an intervention strategy to help students

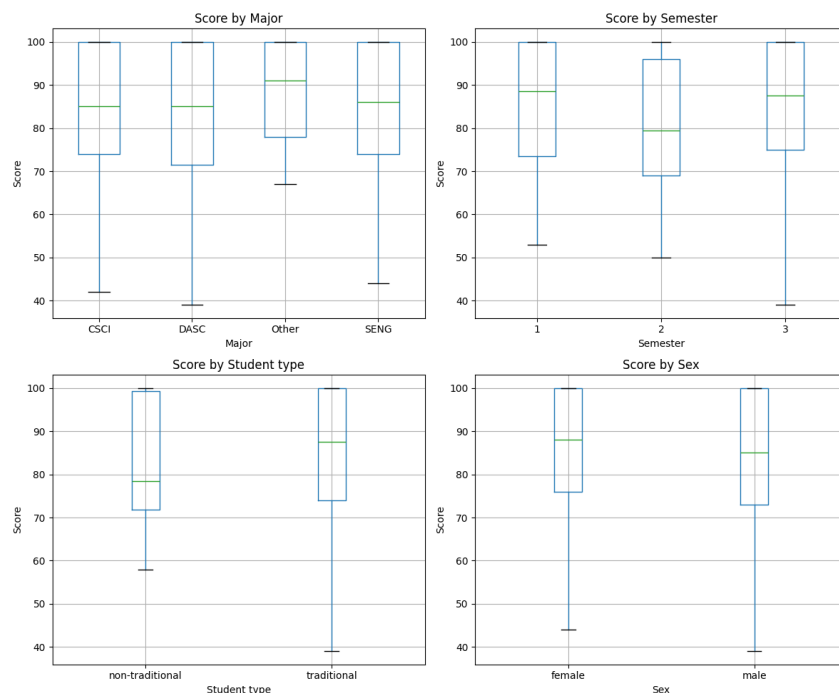
that are likely to fail (based on the past performance of similar students in this course).

- Create one or more visualizations that show the correlation between the student characteristics

and the score in the course, to help the instructor identify at-risk students.

- Describe the pattern seen in the data, thanks to these visualizations.

- Justify your choice of visualization(s).



I created boxplots that score by major, semester, student type, and by sex. The patterns I observed are when comparing majors to scores, everything is similar. When I compare semester to scores, you can kind of see that students in semester 2 are slightly more at risk. When I compared Student type to score, it shows that students who take the non-traditional route struggle a little more than the traditional students. Finally, when comparing sex vs score you can't really see a big difference in the scores.