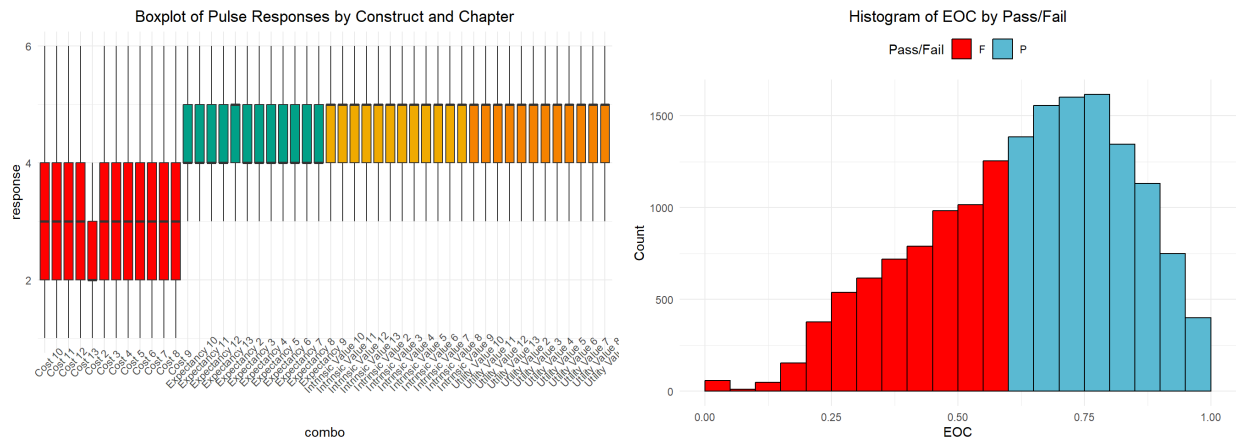


CourseKata
.gif or .gif
Truman State University

Our initial area of interest upon reception of the data was the subjective student responses in the checkpoints_pulse table, as the developers of CourseKata, Jim Stigler and Ji Son, were primarily interested in the students' opinions of the textbook. Unfortunately, this data proved unfruitful, as there was no variation in response for any factor we could find. Since subjective measures had no analytical value, we pivoted to looking at objective measures, starting with the EOC variable in the checkpoints_eoc table, which is the final percentage of questions each student answered correctly on the end of chapter (EOC) quiz. This data was much more diverse and had some interesting potential factors that might influence it. We ended up modifying the EOC data into a binary pass/fail variable with a division at 0.6 (for a 60% pass rate) that focused on book College(ABC).



Our group used a gradient boosted classification tree to chunk down the variables due to the high cardinality so we could use it as an exploratory model. The model started with 20 gradient classification models and picked the best one utilizing racing anova. The final model had 1649 trees. The importance of variables was calculated by how often they were utilized in the final fitted model to make a decision. The important variables were sum of engagement, average attempt, institution, and chapter. The model produced an AUC of 0.847 with an accuracy of 78.3%.

Of the variables our model found to be important, two were student-determined variables and two were environment-determined variables. The total engagement time as well as the average attempts per question, the later of which we engineered ourselves based on n_possible and n_attempted, were the two most influential variables regarding student pass/fail rate, and were the two student-determined variables. Students who obtained over a 60% on the EOC quizzes spent more time utilizing the textbook than students who obtained less than a 60%. Additionally, students with an average attempts per question over 3 were more likely to have an EOC score below 60%, with less students being over that threshold the higher the average attempts were.

The book version for College (ABC) is important to the pass/fail rate of the students with more students improving with newer book versions. There is a large variance in pass/fail rates for institutions with some passing at 75% and others failing at 75%.

Our next steps, if there was more time, would be to remove or rework the subjective "pulse" questions, looking into different book versions, and investigating discrepancies amongst institutions.