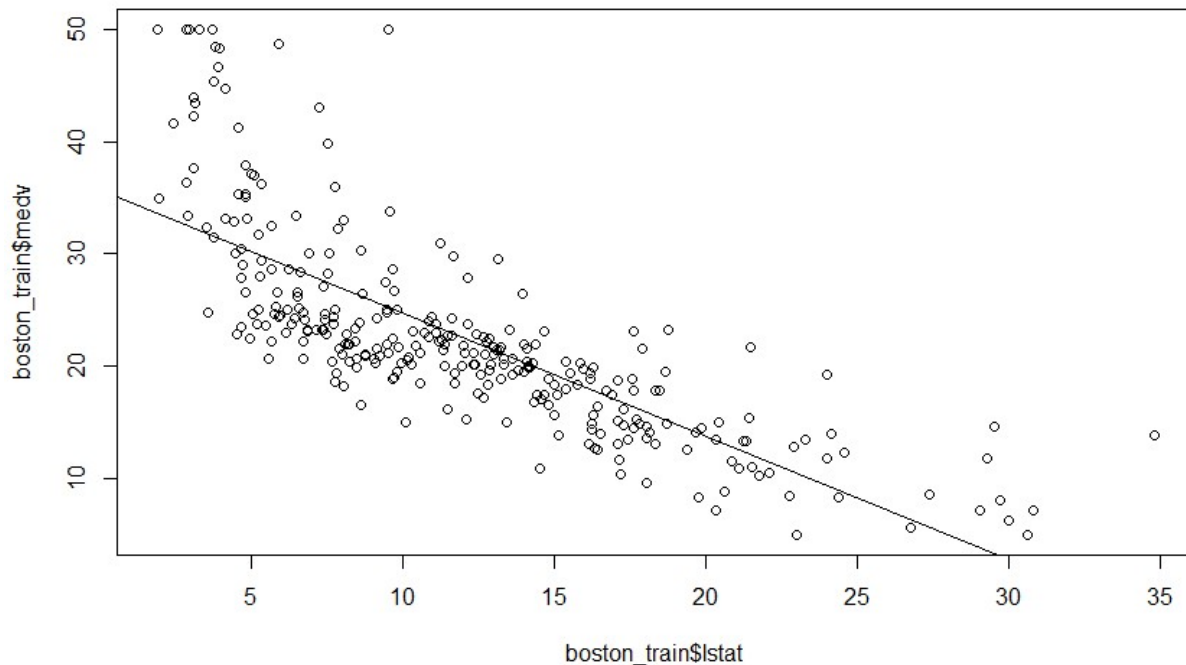


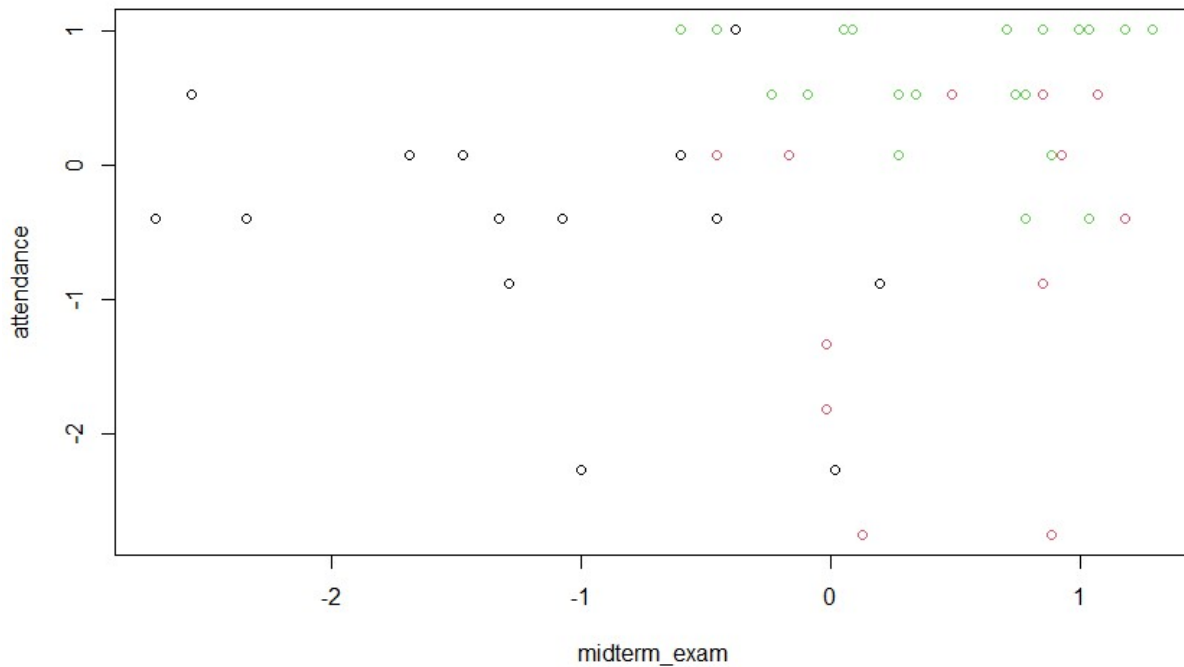
Exam 2

1. The prediction for MEDV using only the LSTAT column is 19.24. The 95% prediction intervals are 7.98 to 30.49 which tells us that 95% of the points are between those two values. This also shows us a huge range and predicting MEDV won't be accurate since it can be any value between those two values.
2. The regression line equation is $y = -1.10x + 35.69$. There is a slight pattern and slightly successful as the graph below shows where some of the points fit on the line and the rest are close, but the regression line could be better. It might be able to be improved by using other predictor variables such as age.



3. The RMSE calculated is 7.20. This shows that the range error is 7.20 for MEDV which is good since MEDV has an original range of 50. The lower we can get RMSE the better so even though 7.20 is better than the default it's still not the best.
4. Using all the other variables to predict MEDV results a RMSE of 6.35 which is a much better range than only using LSTAT to predict MEDV.
5. The best column to remove is CRIM which results in 6.27 RMSE which is much better than using all the columns to predict MEDV.
6. The first cluster has 14 students, the second has 12, and the third has 24.

7. The graph below shows the clusters and the corresponding attendance to midterm exam z-scores. The black is the first cluster which has low grades because their attendance was below average. The red is the second cluster which is the average with some students being below average attendance but had a close to average midterm exam. Finally, the green is the third cluster which has above average attendance and grades.



8. The results below show the centers of each cluster and their averages that correspond to the columns. This shows us that cluster one is below the average of every category which ends them with a below average exam. They also have the lowest midterm exam and paper grade which leads to the final exam having the lowest score of the other clusters. The second cluster shows the highest-class year average with the lowest attendance which results in their exams and paper having the highest average, The last cluster has the lowest class year average with above average exam grades and attendance which results in a above average exam score. The third cluster shows us an average cluster. In total the data shows us that the higher your class year is and if you do well on the paper and midterm you will have a high final exam grade.

```

class_year midterm_exam paper attendance final_exam
1 -0.3267698 -1.1912288 -1.1053071 -0.4682141 -1.2226242
2 1.3489726 0.4773768 0.7856708 -0.6754597 0.6260477
3 -0.4838706 0.4561951 0.2519271 0.6108547 0.4001736

```