M 358K Regression Project

You have been hired as a consultant by a university to give recommendations to improve their Graduation Rate. To do this, you will construct a multiple linear regression model (with at most three predictors) to predict College Graduation Rate.

Use the College data set posted on Canvas. A file explaining each variable is also posted on Canvas. You may use any of the quantitative variables in the data. You may also derive new variables from the data!

Please submit a file to Canvas with all code, all plots, and answers to the questions below.

Do/answer the following:

1. Before you perform the regression, create scatter plots showing the association between the response variable and each of your chosen predictor variables. Make sure the axes are labeled and the plot is titled. Describe the associations (shape, strength, sign). Check the "straight enough" condition.
2. Construct the multiple regression model. Write the equation of the model, and carefully interpret each of the coefficients.
3. What percent of the variability of the response variable is explained by your model? What does the adjusted R-squared tell you?
4. Plot the residuals $(y - \hat{y})$ against the predicted (fitted) values $(\hat{y})$. Create a normal probability plot or a histogram of the residuals. Check the remaining assumptions required for multiple regression.
5. Test the hypothesis that the overall model is significant. This involves an F statistic.
6. Test the hypotheses that each predictor is significant for the model. This involves multiple t statistics.
7. "All models are wrong, but some are useful." Is your model useful? That is, would you use your explanatory variables together to predict graduation rate? Finally, please explain your recommendations to the university.