1. Read the dataset in [Boston.csv](https://maryville.instructure.com/courses/71038/files/16328256/download?wrap=1)[Download Boston.csv](https://maryville.instructure.com/courses/71038/files/16328256/download?download_frd=1)into R. Call the loaded data Boston. Make sure that you have the directory set to the correct location for the data.2.

Text

Description automatically generated

1. The response is **nox** and the predictor is **dis**. Use the poly() function to fit a cubic polynomial regression to predict **nox** using **dis**. Report the regression output.

Table

Description automatically generated

1. Your assistant data scientist, Tom Johnson, is considering predicting **nox** using **dis** as a predictor. He proposes models from degree 5, degree 4, and degree 3, and degree 2 polynomial regression. Please perform cross-validation using caret package to select the optimal degree for the polynomial and justify your answer.

Text

Description automatically generated

Text

Description automatically generated

Text, letter

Description automatically generated

Of the models using the degrees of freedom 2, 3, 4, and 5, the best model is the one using 5 degrees of freedom. It has the highest RMSE score of all the models we ran.

1. Tom just took the DSCI 512. You recommend that he perform the following GAM analysis.

* Predict **nox** using a smoothing spline of degree 2 in dis and a smoothing spline of degree 1 in **medv**.
* Perform **anova** analysis. Recommend the best model and justify your answer.

Text, letter

Description automatically generated

The anova model will signal to you the best p-value with 3 stars. In this case the second model, gamB, would be the best of the two models.