Please do this quiz using an R script and submit it via Moodle. All questions should be answered using R. Make sure your code is well-organized by each question, commenting the question numbers. Avoid hardcoding.

1. For this problem we will be working with the cornnit data set, which is available in the package faraway. Please inspect this dataset and note that the response is yield. Do the following:
   1. Store and display n. Set a seed to 27, then create a train vector by randomly selecting 2/3rd of the indices for the observations.
   2. Set a seed to 27, then find the training MSE and the 10-fold CV MSE for linear models for each of yield versus 1 through 5 polynomial powers of nitrogen on the training data set. Note that the cv.glm function cannot handle the subset variable, so you when creating your model, you cannot use this argument and you must subset the dataset instead. **NOTE**: You should end up with 5 training MSEs and 5 10-fold MSEs, one for each of five powers. ***Only set a seed at the start of this section before creating your loop!***
   3. Print plots for the training MSE and the 10-fold CV MSE versus the power of nitrogen in the same window.
   4. Find the “best” power model using the 10-fold CV MSE.
   5. Create a linear model on the training set for the best power you found in d) as well as for power 2, using the lm function and the subset argument. Do NOT use poly. Then choose between the two powers by calculating the testing MSE for each. **Hint:** Remember the hierarchical principle.
   6. State the winning model (properly).
   7. Plot yield versus nitrogen (full dataset) and plot the winning regression line in a different color (make sure it is ordered correctly). Looking at the plot, does it look like we did well? **NOTE**: The regression line should plot predictions for the full dataset versus nitrogen, using the model created on the training set.