Assignment 8

From the three methods (best subset, forward stepwise, and backward stepwise):

- Which of the three models with k predictors has the smallest training RSS?
    - Best subset has the smallest training RSS
- Which of the three models with k predictors has the smallest test RSS?
    - Also Best subset would be smallest test RSS
- Application exercise: Generate simulated data, and then use this data to perform best subset selection.
    - 1.Use the rnorm () function to generate a predictor X of length n = 100, as well as a noise vector ε of length n = 100. Hint: set.seed(1)X= rnorm(100)eps = rnorm(100)
    - 2.Generate a response vector y of length n = 100 according to the model:$y=\beta 0+\beta 1x+\beta 2x2+\beta 3x3+\varepsilon$,where $\beta 0,\beta 1,\beta 2$ and$\beta 3$are 4, 9, 2, 1 respectively.Plot x and y.
    - 3.Usethe leapspackage: require(leaps)
    - 4.Use the regsubsets() functionfrom the leaps packageto perform best subset selection in order to choose the best model containing the predictors. $x,x2...x10$.Hint:regsubsets(Y~poly(X,10,raw=T), data=data.frame(Y,X), nvmax=10)
        - What is the best model obtained according to Cp, BIC, and adjusted $R2$?
            - I was unable to compart them due to R studio being down
        - Show some plots to provide evidence for your answer, and report the coefficients of the best model obtained. Note you will need to use the data.frame() function to create a single data set containing both $x$and y.
            - I cannot
    - 5.Repeat 3, using forward stepwise selection and using backwards stepwise selection.
        - How does your answer compare to the results in 3?
            - I was unable to compart them due to R studio being down