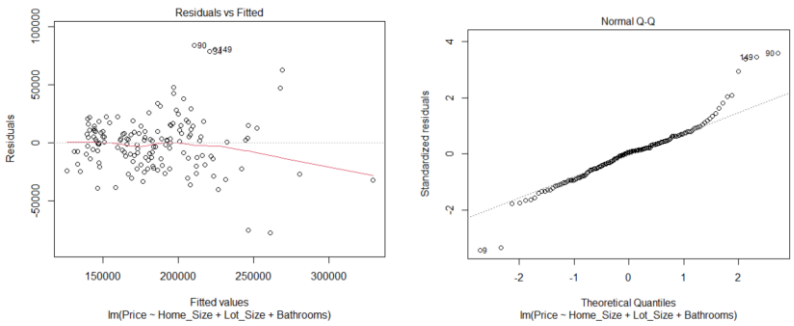
MTH 325, Quiz #7, Spring 2023 Name _____

Instructions: Answer each question as thoroughly as possible. Round answers to 4 decimal places as needed. Exact answers are best when possible. Be sure to answer all parts of each question.

1. Using the same data from Quiz #6 (**325quiz6data.xlsx**), perform three types of model selection procedures (you don't need to transform any variables for this):
   a. Best subset selection
   b. Backward selection
   c. LASSO (penalized) regression (see Lab #6 for code examples)

Best subset

```
(Intercept)    Home_Size     Lot_Size    Bathrooms
87979.41064     32.56303   7512.43041 14807.04471

Call:
lm(formula = Price ~ Home_Size + Lot_Size + Bathrooms, data = data6)

Residuals:
   Min     1Q Median     3Q    Max
-77647 -13457   1215  10732  84230

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 87979.411   6403.155  13.740  < 2e-16 ***
Home_Size      32.563      5.191   6.273 3.80e-09 ***
Lot_Size     7512.430    878.653   8.550 1.52e-14 ***
Bathrooms   14807.045   4314.867   3.432 0.000781 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 23700 on 146 degrees of freedom
Multiple R-squared:  0.6822,   Adjusted R-squared:  0.6756
F-statistic: 104.4 on 3 and 146 DF,  p-value: < 2.2e-16

AIC= 3453.585
BIS= 3468.638
```



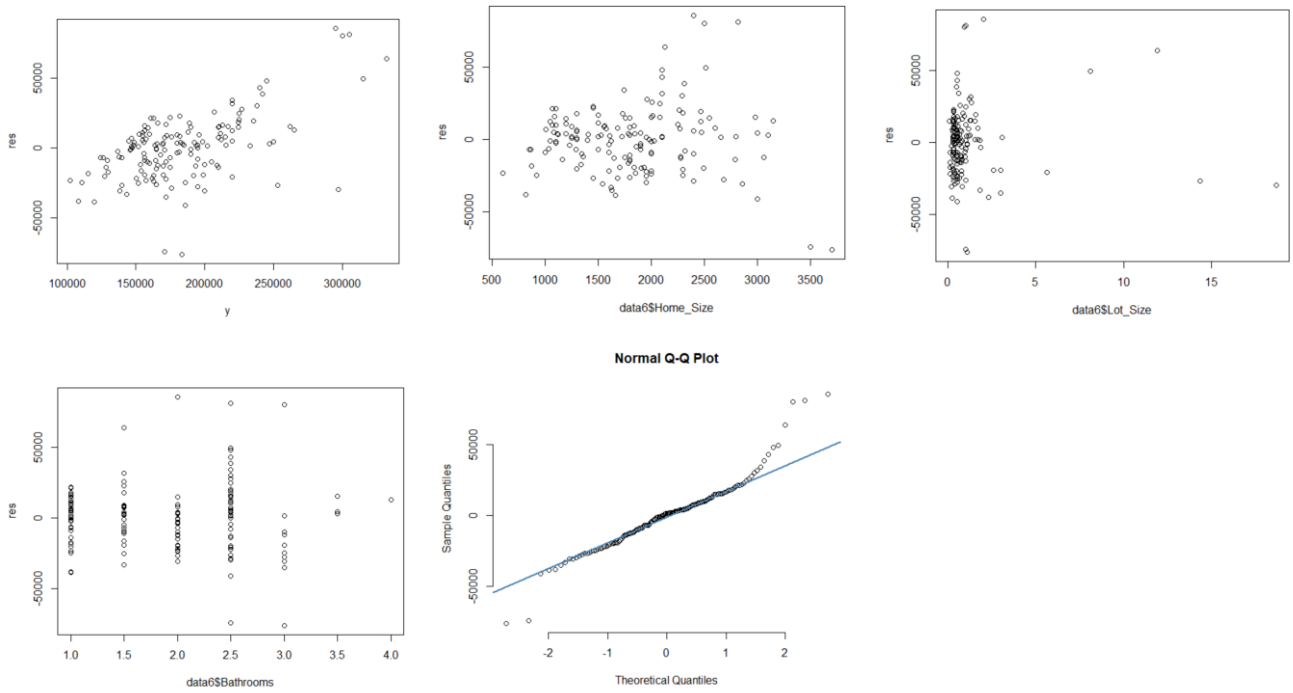
Backward selection

In backward selection, we start with all 4 variables. The rooms variable is not significant. If removed, we end up with the model above.

LASSO Regression

```
                    s0
(Intercept) 86404.27355
```

```
Home_Size       30.86998
Lot_Size      7432.83467
Rooms          718.27114
Bathrooms    14565.63278

    RMSE       s0
1 23373.6 0.6824218
```



Report the results of the coefficients and variables in the model in each case. Compare the results using the following criteria:

i. The R^2 value if available
ii. The values of the coefficients
iii. The residual standard error
iv. The AIC and BIC
v. Which model is the simplest (has the fewest variables)? Did any come out the same?
vi. Using plot(modelname) in R, create diagnostic plots for each model.

Based on this information, write a paragraph explaining how you would choose from among these models. You are free to bring in additional criteria as needed.

In this case, best subset regression and backward selection resulted in the same model with three variables, while the LASSO retained all variables (a little surprising). The three-variable model is the simplest. The R^2 values for the LASSO model and the other model are about the same, and the standard errors (RMSE) is also nearly the same, so the advantage goes to the simpler model unless there is some external reason to wish to retain Rooms in the model.