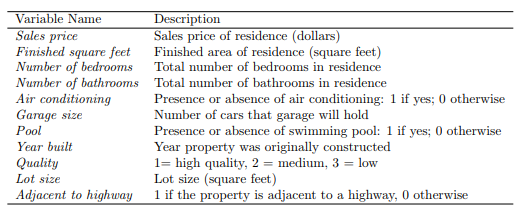
Assignment 3

Due: Wednesday, October 9, 2019

This assignment relates to the ‘Real-estate’ data set. People buying or selling houses would like to know how much they can expect to get, or pay, for a property. This is also a concern for those who are making mortgage loans, or for those taxing real estate (and who are more likely to commission statistical studies than individual home-owners). The price of a house depends on its physical characteristics, including size, features, quality of construction, age, etc. It also depends on location, and current market characteristics. You are approached by a research group which has a data on a sample of residential sales in a midwestern city; the variables are described in Table below.



1. Read the data into R. Call the loaded data “real.estate”.
2. Answer the following sub-questions
3. Use the “summary()” function to identify the types of variables. Which variables are categorical? Which variables are quantitative? Are there any concerns in the summary table? Explain.
4. Use the “pairs()” or “gpairs()” function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first then columns of a matrix A using A [,1:10]. Is there any interesting patterns? Which variables seem associated with the sales price? Explain.
5. Use the “as.factor” function to regenerate categorical variables.
6. Fit the models and address the following when building that model:
7. Fit the null model and the full model
8. Find ***the best sets*** of predictors using the stepwise procedures.
9. Find ***the best sets*** of predictors using the best subset approach.
10. Considering the models in part (ii) and part(iii), choose the best model.
11. Interpret the coefficients of the best model ***in this context***.
12. Evaluate the best model.
13. Check the assumptions using “plot()” function.
14. Continue exploring the data, and provide a brief summary of what you discover.