HW 10

Due: May 14th at 3:00 pm

1. Use the real estate dataset from this link:
   http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Appendix%20C%20Data%20Sets/APPENC07.txt

   The variables are:

   - Identification number — 1-522
   - Sales price — Sales price of residence (dollars)
   - Finished square feet — Finished area of residence (square feet)
   - Number of bedrooms — Total number of bedrooms in residence
   - Number of bathrooms — Total number of bathrooms in residence
   - Air conditioning — Presence or absence of air conditioning: 1 if yes; 0 otherwise
   - Garage size — Number of cars that garage will hold
   - Pool — Presence or absence of swimming pool: 1 if yes; 0 otherwise
   - Year built — Year property was originally constructed
   - Quality — Index for quality of construction: 1 indicates high quality; 2 indicates medium quality; 3 indicates low quality
   - Style — Qualitative indicator of architectural style
   - Lot size — Lot size (square feet)
   - Adjacent to highway — Presence or absence of adjacency to highway: 1 if yes; 0 otherwise

   a) Select a random sample of 300 observations to use as a training dataset.

   b) Develop a neural network model for predicting sales price. Try your best to find a good number of hidden nodes and other tuning parameters.

   c) Assess your model's ability to predict and discuss its usefulness as a tool for predicting sales prices. (Here you need to use the test dataset.)

   d) Compare your neural network to a regression model with your choice of best subset selection method. Which model is easier to interpret?