Regression analysis of used Mercedes cars: Data taken from the advertising pages of the Sunday Times a few years ago, presenting cars for sale in the UK (mainly in and around London). The asking prices (in pounds sterling) are classified according to type/model of car, age of car (in six-month units based on date of registration), recorded mileage, and vendor. The data in usedcars.xls are:

1. Case number 1:54;
2. price: Asking price in pounds;
3. type: Type/Model: 0=model 500, 1=450, 2=380, 3=280, 4=200;
4. age: Age of car in six-month units, based on registration;
5. mileage: Recorded mileage (in thousands);
6. vendor: Vendor (0,1,2,3 are dealerships, 4=“sale by owner”).

a) Run the simple linear regression of price on mileage.
b) Run the simple linear regression of price on age.
c) Which of the linear regressions fits price better in terms of $R^2$?
d) Run the multiple linear regression of price on mileage and age.
e) Based on $R^2$, is the multiple linear regression in d) better than the ones in a) and b)?
f) Do the residuals of the previous three linear regressions look i.i.d. normal?

The variables type and vendor are both categorical and need special consideration. Remember, from class, that in order to run regressions with these variables we need first to create dummy variables (0/1 variables) to account for the different categories. For example, type has 5 categories, so 4 dummy variables are necessary.

g) Run the regression of price on type.
h) Run the regression of price on vendor.
i) Run the regression of price on mileage, age, type and vendor.
j) Compare all 6 models based on $R^2$.

From a) to j) provide as much comments as possible. For instance, plot x against y, discuss the (lack of) importance of the various explanatory variables, check residuals. The more (useful) information the better!