I have used new data for this regression module since the previous one did not have much scope with numerical variables. My dataset is about automobiles (from the moodle-suggested UCI website) and insurance information related to that. My target variable is normalized losses (relative average loss payment per insured vehicle year), and there were 25 predictors involved. Since the list is too long, I have simply attached an excel document in case anyone is interested.

First I removed all the data with missing variables since regression can’t work with missing variables. Then, I made a regression model with all variables but there were only a few significant (I judged the significance of the predictors but the \*\*\* in the lm model printed by R) so my final model had the following predictors: make, num.of.doors, drive.wheels,height (the description of these variables is in the attached text document).

One problem that took me forever to debug: R kept complaining about my model with categorical variables and it turned out that one of the factor variables had all the data in one factor and zero in the other, and R doesn’t plot regression models with that so I had to remove that.

Final model:

