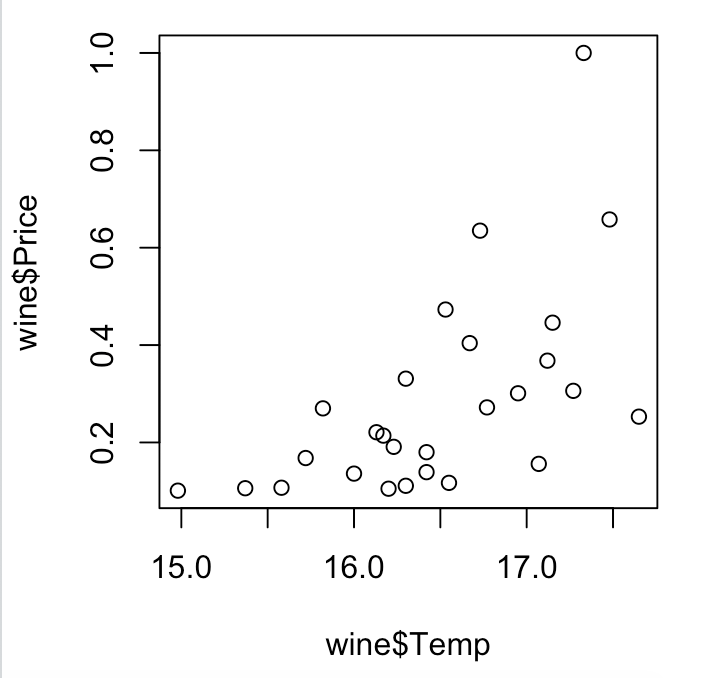
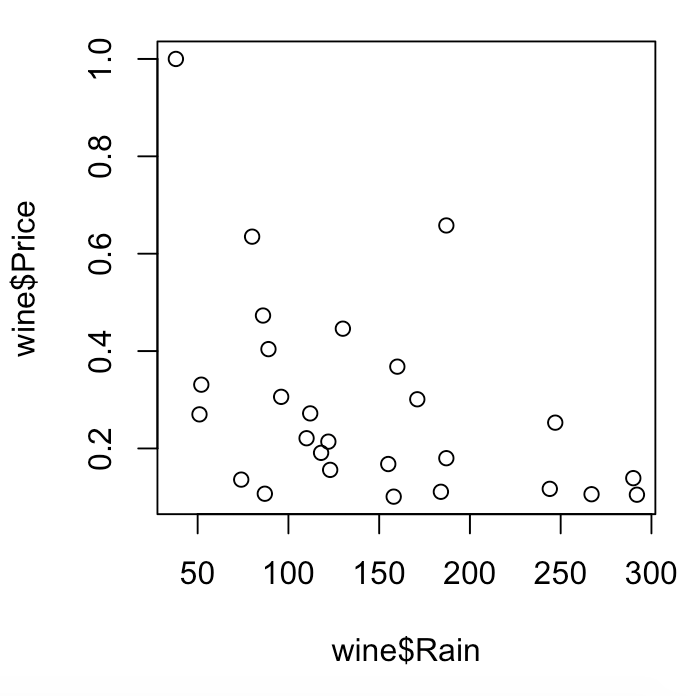
BSAN 450 Assignment 5

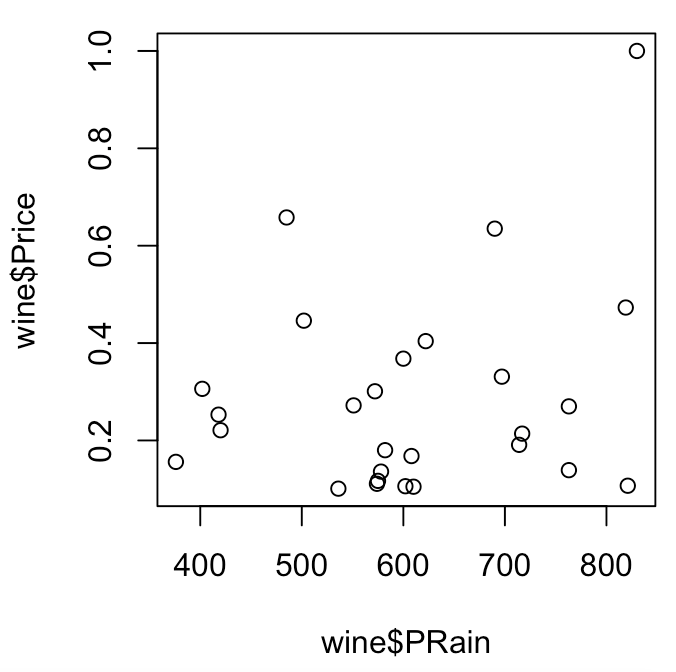
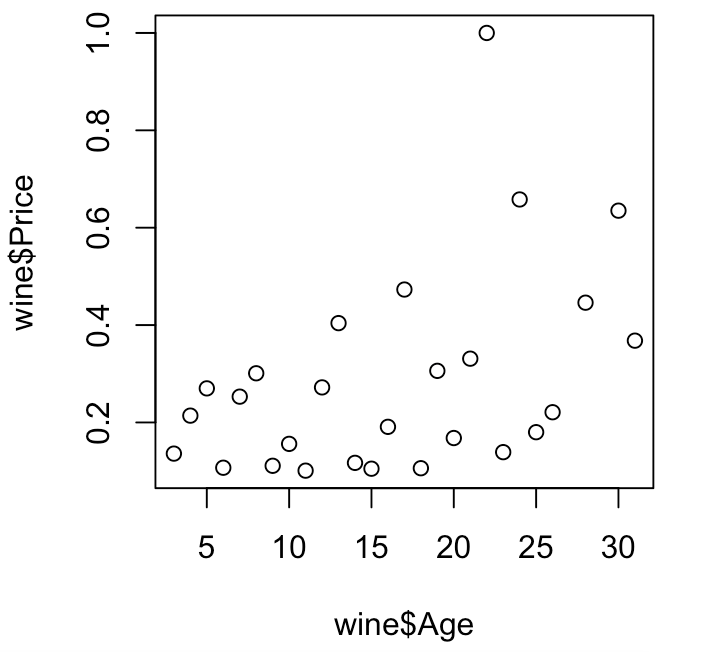
1) The purpose of this example is to develop a regression model to predict the price of Bordeaux wines based on the following variables: Temp = the average temperature between April and September in the year the grapes were grown, Rain = the total rainfall in the months of August and September in the year the grapes were grown, PRain = the total rainfall in the months October to March in the year the grapes were grown, and Age = the age of the wine in years in the year 1983. The variable we wish to predict is Price = the fraction of the price of the 1961 vintage. The prices were obtained from auctions in 1990 and 1991 in the London market. The data are in a file named “wine.csv”.

a) Read in the data and plot scatter plots of the variable Price versus all the independent variables.

Comment on these plots. For each plot, is there a relationship between the independent variable and Price? What is the nature of this relationship?

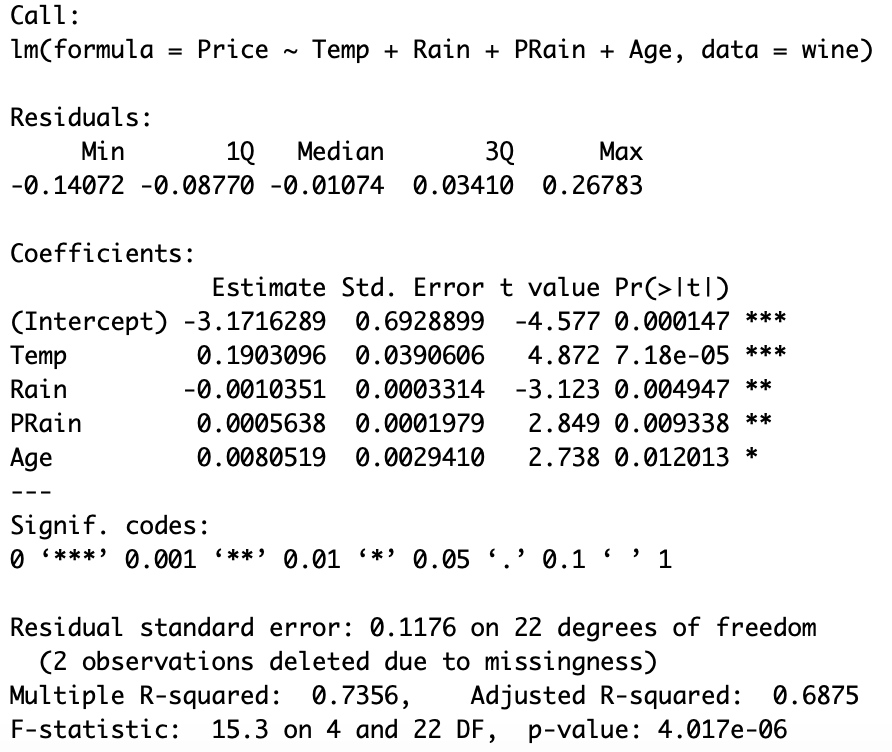
 

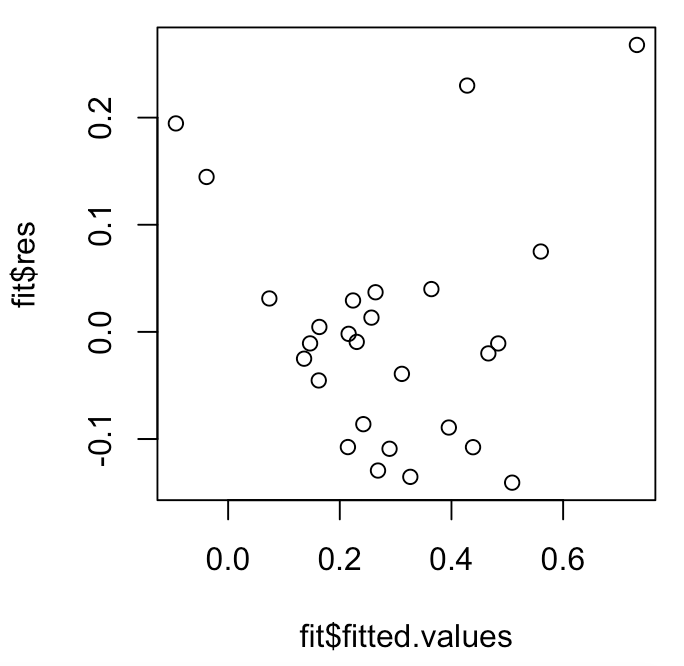
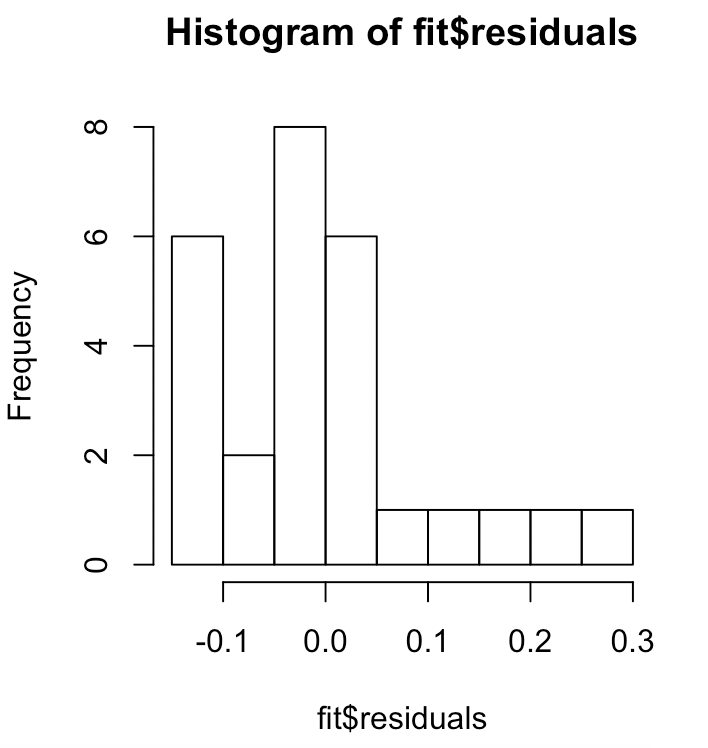
Temp appears to have a positive linear relationship with Price. Rain appears to have a negative linear realtionship with Price.

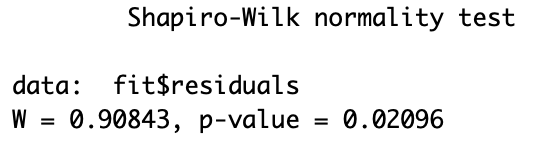
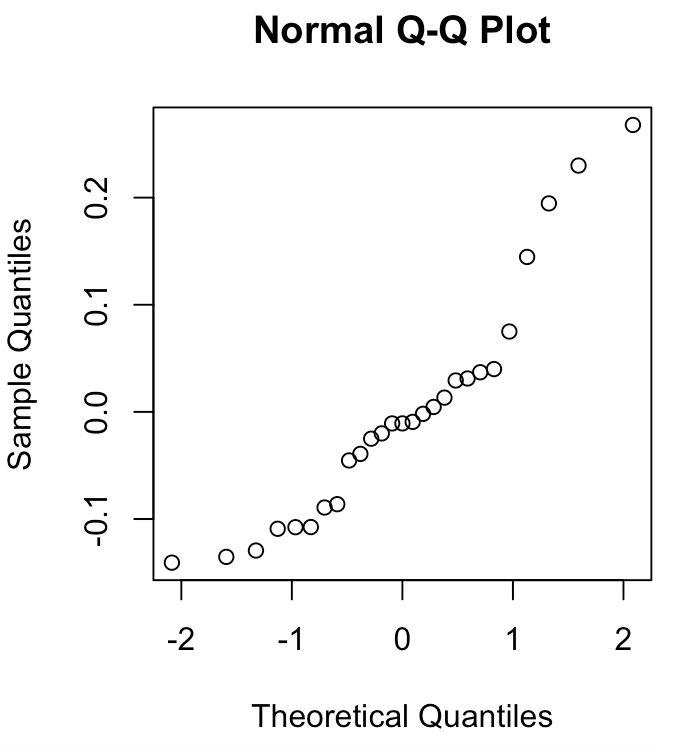
 

PRain appears to have a slight positive linear relationship with Price. Age appears to have a positive linear relationship with Price.

b) Based on the plots in part a) propose a regression model for this data and fit this model. Perform the diagnostic checks for this model. Are there any problems with your model? Does your model need to be modified?



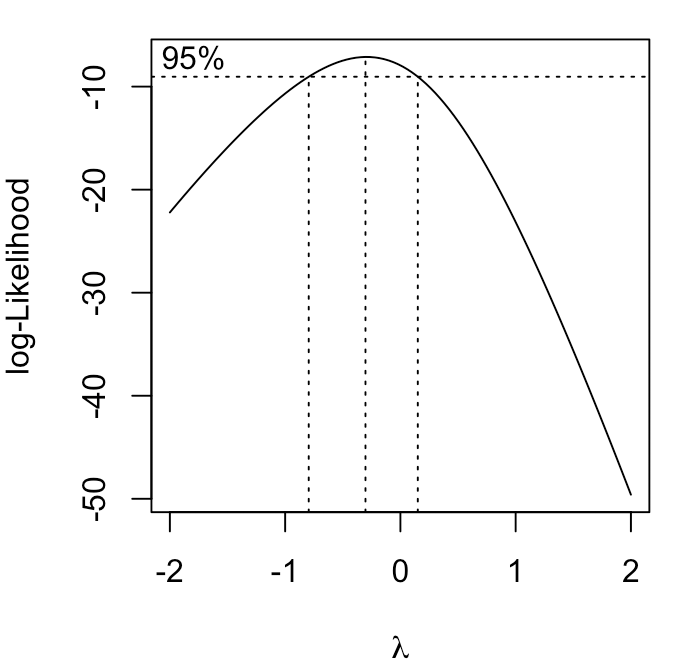
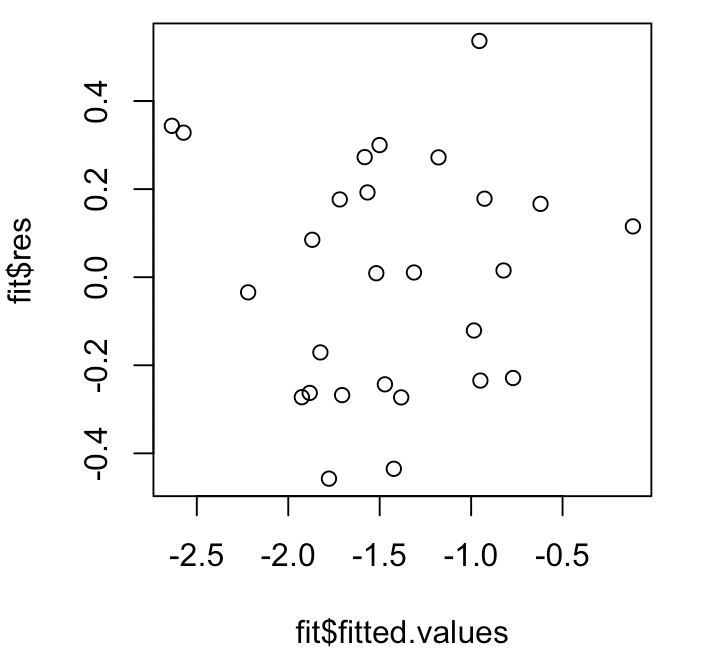
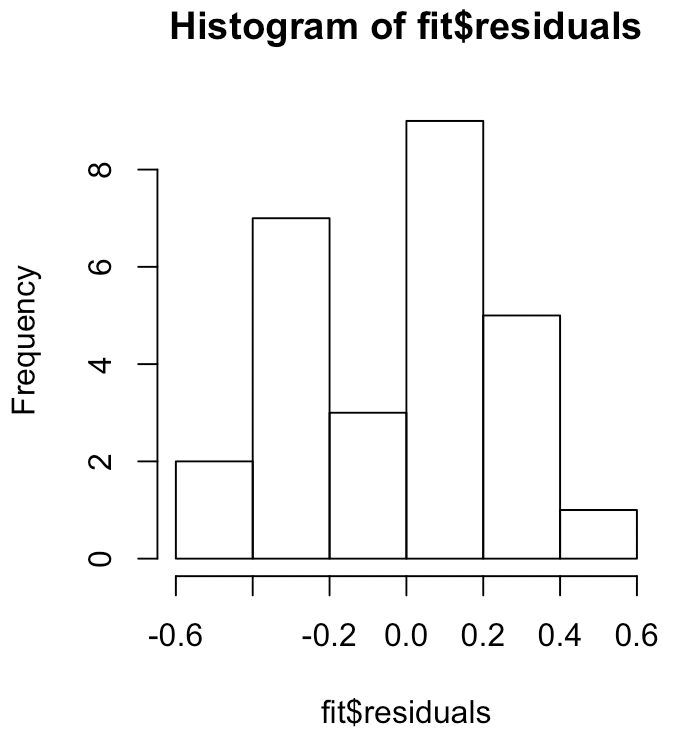
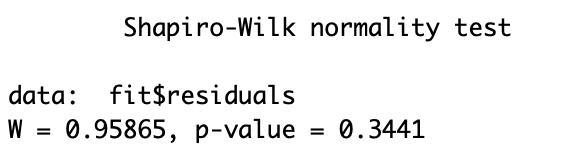
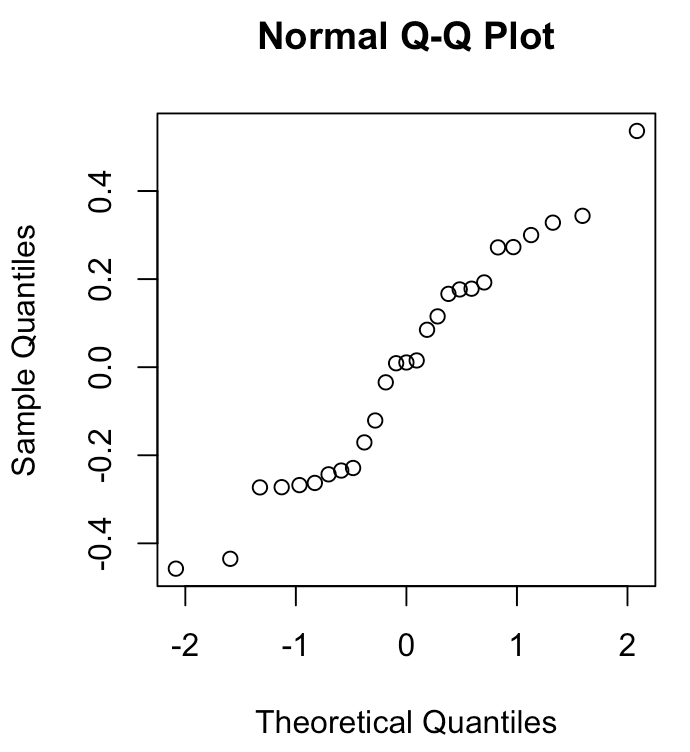
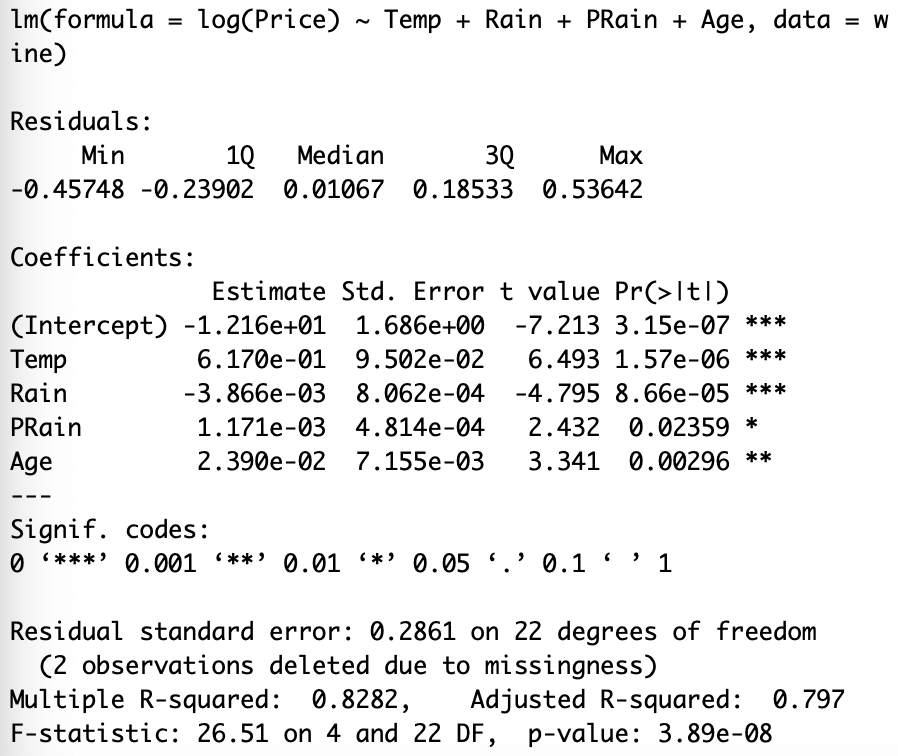
 

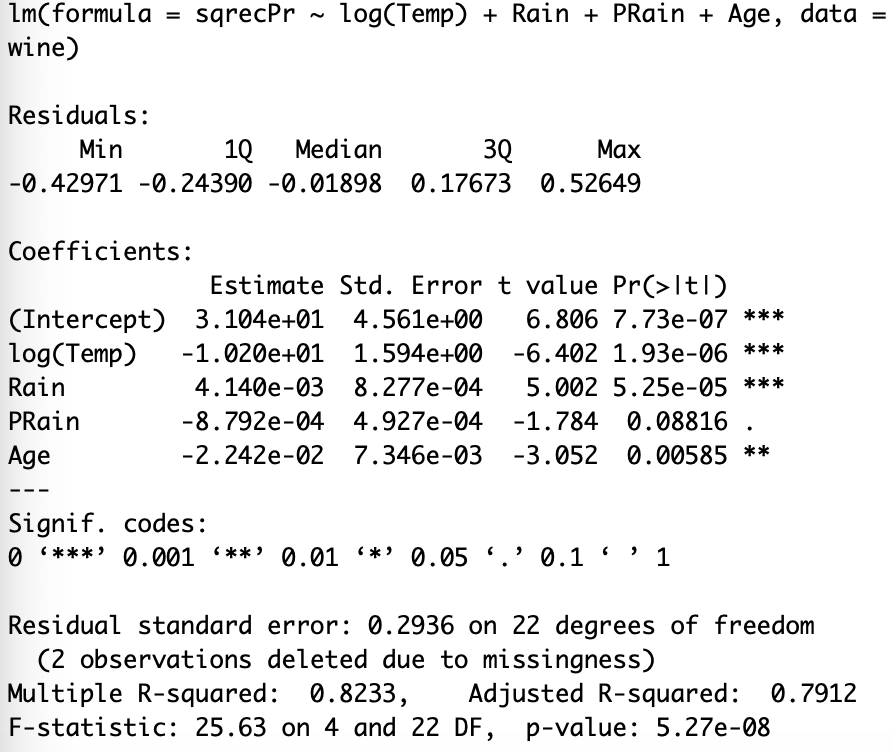
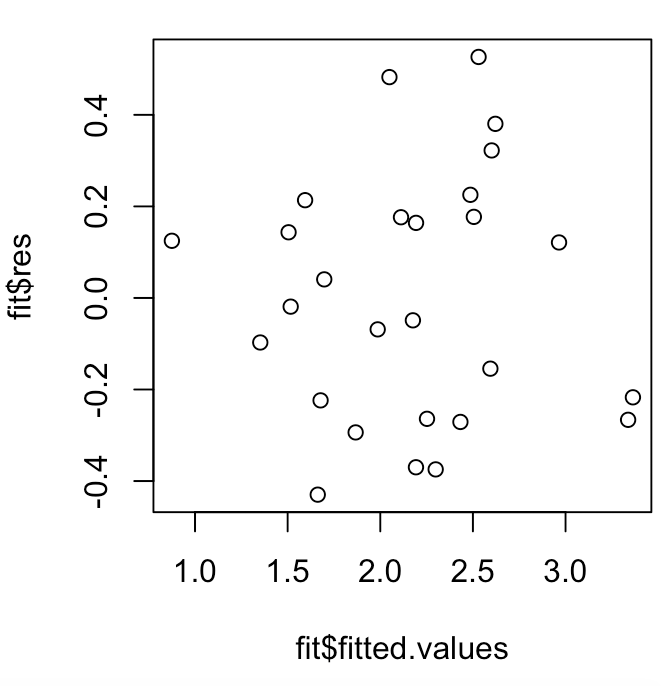
The diagnostics suggest a problem. The histogram is not a normal distribution. The residual looks almost parabolic.

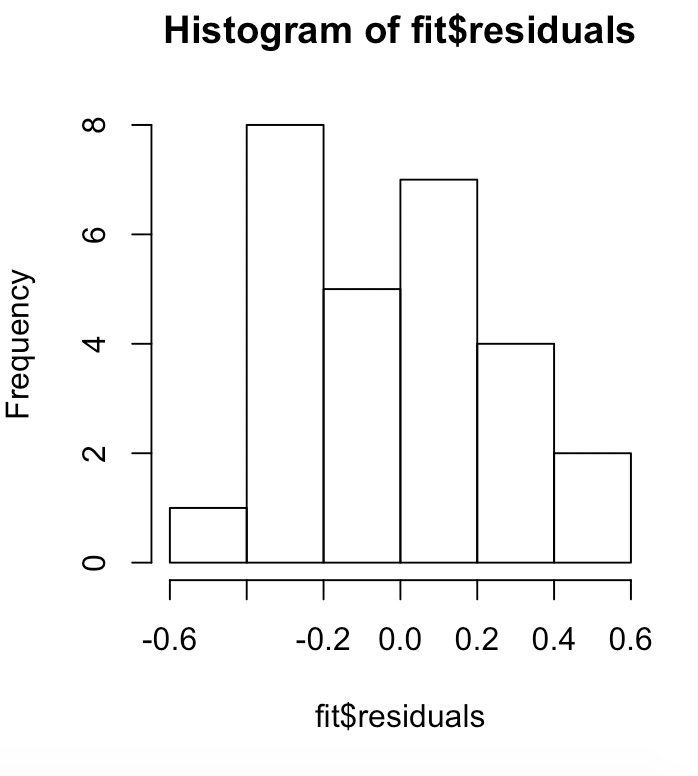
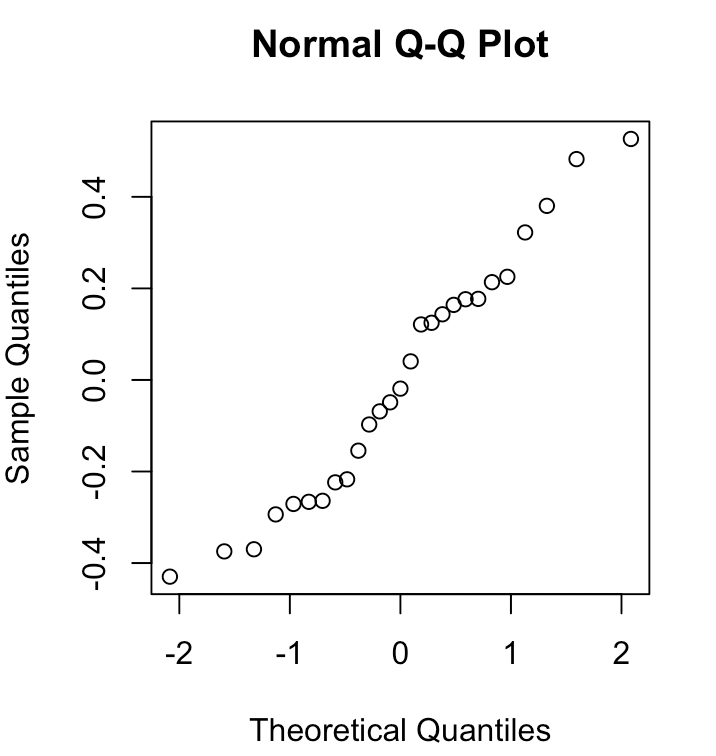
c) If the model you fit in part a) needs to be modified, suggest a new model and fit that model. Repeat the diagnostic checks. Is there a need for further modification? If so propose a new model, fit that new model and check that model. Continue until you have a model that you are satisfied with.

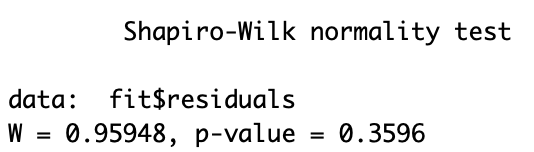
To try to fix the parabolic shape of the residual graph I did a boxcox to see if there was a transformation of Price that would look good. I tired doing the log.

The Q norm graph looks like it has a curve in it. I then tried a transformation on the Temp variable, because it looked like it might have a slight curve. I also tried a different transformation of the y variable.



I decided to leave PRain in even though its p-value is .08 which is greater than .05. The R-squared value was higher, and the histogram of the residuals looked better in this model.

In my last model I tried just doing the log of temp and rain and leaving the y-variable unchanged. I also removed PRain from the model.

