ANALYSIS

1)Backward elimination: find which variables are most significant

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const 83.2315 12.282 6.777 0.000 59.049 107.414

x1 0.5457 0.129 4.243 0.000 0.292 0.799

x2 0.0391 0.021 1.885 0.061 -0.002 0.080

x3 0.0578 0.049 1.171 0.243 -0.039 0.155

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

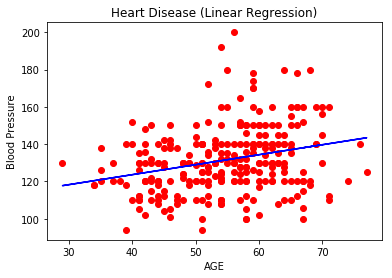
const 94.7481 7.359 12.875 0.000 80.259 109.237

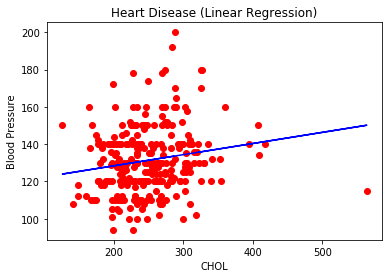
x1 0.4842 0.117 4.122 0.000 0.253 0.716

x2 0.0410 0.021 1.981 0.049 0.000 0.082

Here I determined that AGE and cholesterol were most significant (low p values)

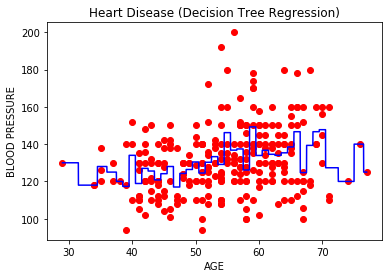
Next are my two simple linear regression models, one for each variable in relation to blood pressure.

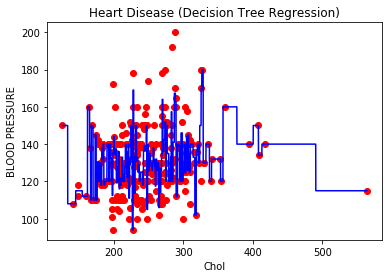




We can see a pretty good trendline in both of these, as predicted by the previous calculations.

Next, I will try to see how a decision tree model explains our data.





We can see the first one is a more stable pattern while the second, not so much. The first one tells us that as age increases, blood pressure increases relative up to around age 70. This is important information that we did not have from the SLR models. The second one appears too jagged to be of much help, so this one is better represented by the SLR model.