Class 3

- First look at the p-value to see if any values have any impact on the output.
- Normalization (esdcaling) has not effect on the quality of model. It just gives different numbers.
- R-Sqaure doesn't account for ovrerfitting. That's why we use adjust RSqaure.
- the process is, include the squared version of vars in the model, and see if this improves Adjusted R and p-value
- Median residual better be zero. Compare medain with the range
- Plot fitted versus residuls, if they are outside of zero range you can see where you make errors.
- That shows if we need to squares, root

Model Selection and Feature Selection

Two reason for not using all vars

- p>n
- Variance-bias trade-off which is the idea overfitting *** Two types of errors: **** Bias: How well we work with training data **** Variane: how well do we do with other training or test data? ** Focus on the big picture and weed out predictors that are irrelavant
- Best subset selection - (2^P so we can't test all subsets) select the best
- Cp estimate of MSE how good your model would do if you run the model on data
- Cp, AIC and BIC smaller is better, Higher adjust sqaure is better
- Cp and AIC are porportional so you can use just one. BIC penalizes the num of predicators for large n - log(n) _ BIC grows faster than Cp
- Adjusted RSquare first goes up and then goes down. RSquare always goes up.