**Motivating reasons for subset selection:**

* Model interpretability – If we can narrow down the subset of predictor variables to just those that explain the majority of the variance in predictions, we gain better information about our problem.
* p > n – If we have more predictor variables than observations, we require a smaller subset of predictors to even use OLS. Even if n is less than p but by a small margin, reducing the subset of predictors is still desirable to reduce variance in our model predictions.

**When to use:**

Best subset selection: Use best subset selection when p is reasonably small. The book gives a number of less than 40 predictors while the instructors in the video series suggest an even smaller number of predictors. They made it sound like there aren’t many situations in general anymore where you would use this approach. We are in an age with very large data sets with many predictors being the norm. That and the fact that we have more methods to choose from including the hybrid approach discussed below that give comparable results with much less computational cost, best subset seems to be continually losing ground in its application.

Forward stepwise selection: Use forward stepwise selection when p is too large to apply best subset selection and when p > n since backward stepwise selection will no longer work.

Backward stepwise selection: Use backward stepwise selection when p is too large to apply best subset selection and when the number of samples n is larger than p.

Hybrid approach: If n > p and p is very large, you would want to use forward or backward stepwise selection over best subset selection, but there isn’t a clear advantage of using forward stepwise selection over backward stepwise selection and vice versa. It would be useful to try out a hybrid approach that combines forward and backward stepwise selection that adds a variable at each step but also removes a variable that is no longer provide improvements on the model. This approach is both computationally similar to backward and forward selection and may give better results than either on its own since it closely mimics best subset selection.