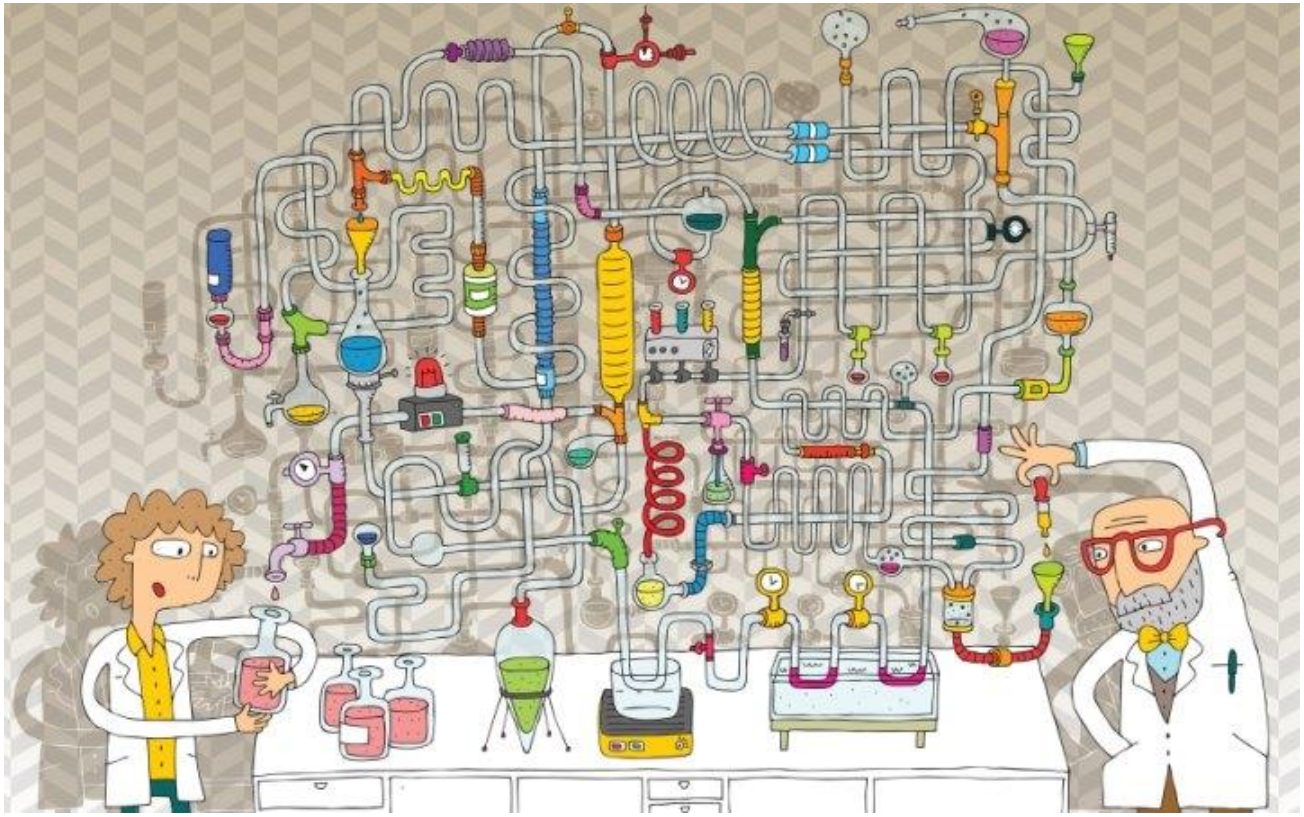
Reducing the number of predictors – selecting subsets of variables

(using the Toyota example in chap 6)

The kitchen sink temptation…



Simpler is often better

Selection Algorithms

Backward

1. Start with all predictors
2. Drop least significant predictor
3. Repeat until all remaining predictors are statistically significant

Foreward

1. Start with 0 predictors
2. Add predictor with the biggest boost in R^2
3. Repeat until contribution to R^2 is not statistically significant

Subset Selection

Like forward except at each stage, variables are eligible for being dropped, as well as added

Exhaustive Search (Best Subset)

Evaluate all possible subsets (very computationally intensive)

Possible Criteria for Evaluating Predictors Included in Subsets

- Statistical significance of predictors
- Contribution to R^2
- Goodness-of-fit metrics with a penalty based on # of predictors
  - Akaike Information Criterion (AIC)
  - Bayesian information Criterion (BIC)
- Mallows Cp

Note that all these metrics are based on the training data, whereas we are interested in predictive performance with new data. You can think of these criteria as useful, somewhat arbitrary ways to produce simpler models, and needn't spend too much time on the distinctions among them.