**Data Science II: Homework #1 Strategy**

**Goal:** Predict sale price of a house given a set of numeric and categorical predictors (and interaction terms?)

**Methods:**

* Linear model
* Lasso model
* Elastic net model
* Partial least squares model

**Linear modeling process:**

* Data preprocessing (cleaning, tidying)
  + Seems unnecessary; maybe omit any rows with NA?
* Exploratory data analysis / summary statistics
  + Probably just using training data? Look at summary statistics, perhaps correlation matrix
* Data partition into train and test sets
  + Seems done for us already
* Reduce predictors using best subsets on training data with resampling / cross-validation; also make sure predictors are regularized/scaled
* Train model on training data, with resampling / cross-validation
* Evaluate model with predictions on test data
  + Obtain RMSE / MSE
* Model diagnostics?
  + Probably out of scope for this assignment