* MVU Module/Slide Notes
* Intro to Statistical Learning with R (ISLA)
  + Chapters 4, pages 127-137
    - If you have 2 categorical variables, linear regression can still work out. >0.5 = 1 everything else is 0
    - Similar to linear in that we still get intercepts and coefficients
    - Can also make predictions based on “multiple” attributes
  + Chapters 5, pages 178-184
    - Re-sample, take smaller samples of overall data, take them multiple different times and draw conclusions
    - Model assessment = judge performance
    - Model selection = choosing based on flexibility
      * Validating on a 50/50 split can produce highly variable test errors
      * Accuracy gets worse as training data size shrinks, thus a larger validation set makes our training worse and could overestimate true test error
      * “Leave One Out Cross Validation” LOOCV:
        + Validation data set is only 1 single observation
        + But then we do this n times for the n observation and average all of the test errors
        + Expensive and time consuming to run the model n times
  + Chapters 5, pages 205-213