* (Download data (King County housing price dataset))
* Check data : price; sqft\_living; grade; basement; lat; long; condition
* **Model comparison**
* **Hiearchical regression**
  + quantify the amount of **information gained by adding (a) new predictor**(s)
  + this model comparison is “simple” because it compares **nested** models -> **models are similar**
* Build TWO regression models
  + price; sqft\_living; grade
  + price; sqft\_living; grade; lat; long
  + Compare change in the adju. R^2
* Build a hierarchical regression model
  + ASK FOR R^2 change
  + F test for R^2 change
* This only works for NESTED models!!!
* Ask for AIC -> add Statistics SELECTION in syntax
  + AIC is more established comparison tool
  + It can be used for comparing non nested models as well to some extent, but may not be as reliable
* Hierarchical regression with more than two blocks
* **Model selection**
* **First rule of model selection:** Always go with the model that is grounded in theory and prior research
* Predicting data in the original dataset is easy
* We can create a line which is super-flexible, -> prediction will be perfect
  + Chart - Element – interpolate line
  + On new data, it will produce bad model fit
* Useful to have training set vs. test set
* Advantage of linear regression: **Linear regression is very inflexible**, so it is less prone to overfitting
* **Comparing model performance on the training set and the test set**
* Generate random variables
* Separate dataset to 2 100 data segments
* Build hierarchical regression
  + With original predictors
  + with all predictors
* Compare models
  + RSS
  + Adj.R^2
  + AIC
  + hierarchical regression ANOVA
* Result-based models selection