# Multiple Regression Lab

Research Question: What influences housing prices in my area?

Goals for Lab:

* Download and prepare data for Lab
* Preliminary data analysis
* Run Multiple Regression
* Post-regression analysis

### Deliverable:

### Write-Up

Write a 2-3-page report of your results. The file should include:

* Plots
* Regression output
* Post-regression tests/analyses with output
* Explanation of results

Any claims made must be supported by the information you provide in the write-up.

### Steps for Lab:

Download Data: For this lab, we will be using housing data from Redfin

* Go to https://www.redfin.com/
* Search by Zip Code
* Click on “More Filters”
* Set Filters to eliminate non-single-family housing, to include sold properties, and to have at least 200 observations
* Go to Table View
* Scroll down and click on “Download All”

Save Master copy

Variables needed in this lab:

Price (in thousands)

Beds

Bath

Square Feet (in thousands)

Lot Size (in categories)

Year Built

Load Data into Rstudio

Necessary Packages: [XLConnect](https://cran.rstudio.com/web/packages/XLConnect), [tidyverse](https://cran.r-project.org/web/packages/tidyverse/index.html),

Create Information for Report, based on the questions that follow.

Prepare and submit report

1. Create multiple scatterplots (5) of each explanatory variables versus price (y).
2. Identify which variable do you think best explains sales price.
3. Create histograms (5) of all variables.
4. Identify any variables that might benefit from data transformation, and transform as necessary.
5. Perform an OLS regression including the variables above as explanatory variables, and price as the dependent variable
6. Discuss the significance of each of your variables.
7. Is the model significant?
8. How much variation in price is explained by this model?
9. Create a residual plot for each of your explanatory variables (5).
10. Describe and interpret the residuals for each explanatory variable.