

**Initial Problem:** Evaluate the impact of kitchen quality on a home's selling price in Ames, Iowa

**Secondary Problem:** Describe the results of the analysis to a home-seller and to a realtor

**Data:** in R: `library(AmesHousing)`

**More Info:** <http://jse.amstat.org/v19n3/decock.pdf>

---

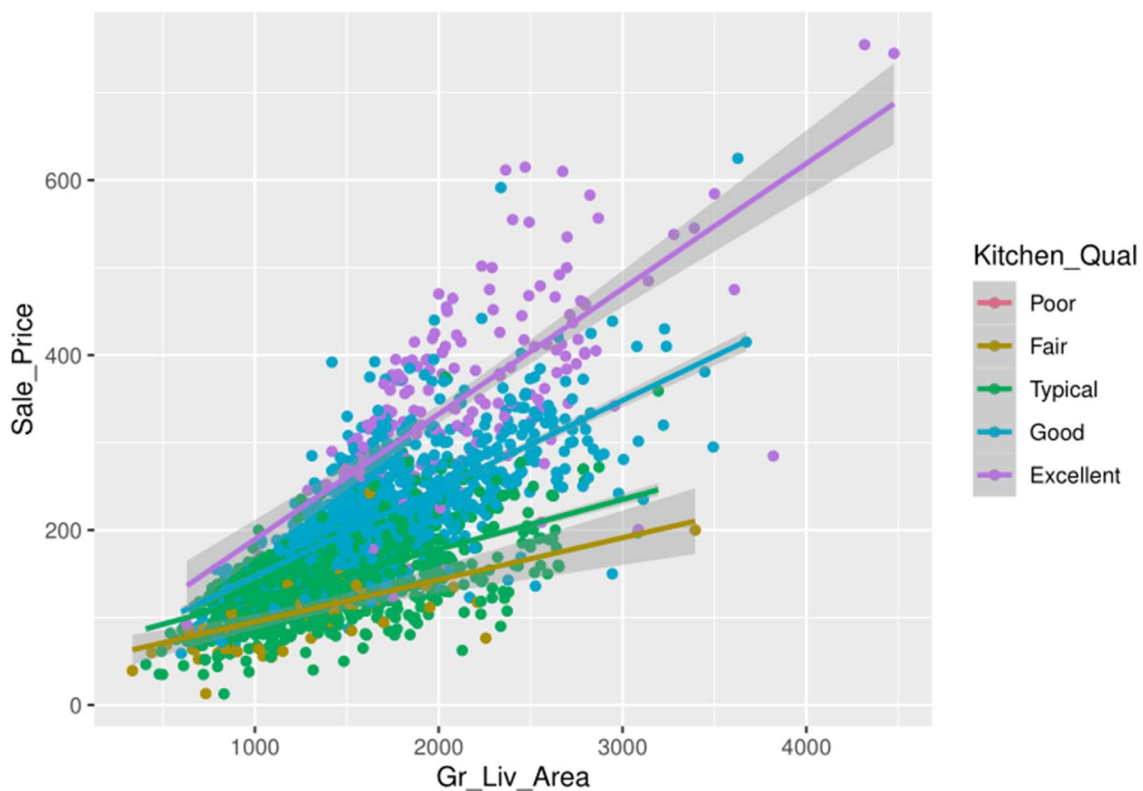
*First Step:*

The easiest way to determine a relationship is running a regression model

Regression analysis is fundamentally a quantitative measure

Kitchen quality is a categorical variable. The first step is change kitchen quality into a numeric variable. This is done in R by: `factor()`.

*Result:*



There is an obvious correlation between kitchen quality and selling price.

### Second Step:

Now we need to see the actual impact of kitchen quality. Remember that the numbers are in the thousands.

### Result:

```
##
## Call:
## lm(formula = Sale_Price ~ Gr_Liv_Area + Kitchen_Qual, data = ames_quick)
##
## Coefficients:
##           (Intercept)           Gr_Liv_Area           Kitchen_QualFair
##              3.00760              0.08537              -1.57710
## Kitchen_QualTypical      Kitchen_QualGood  Kitchen_QualExcellent
##              23.32711              68.25500              161.77383
```

The coefficients of “above ground living area in square feet” increases the price by \$85/sqft. Likewise, a “fair” quality kitchen decreases house price by roughly \$1600. In comparison, an “excellent” quality kitchen increases house price by roughly \$161 thousand – everything else being equal, a house with an excellent quality kitchen will cost \$163 thousand more than a house with a fair quality kitchen.

### Third Step:

Consider everything we have not considered:

Great Recession, neighborhood, day of month, days on market, business activity

Can you think of anything else that might impact the sale price other than above ground living area and kitchen quality? (interest rates, weather, quality of other rooms)

There are 80 variables in the dataset. We looked at two. There are dozens, if not hundreds, of other variables one could also consider when creating a predictive linear regression model.

This brings up another thing we have not considered – model type. Is linear regression the best model?

### Conclusion:

Realize data limitations (4 year time period, missing variables, etc.) and determine if conclusion is still valid. Then present your findings in a way **understandable to the recipient**