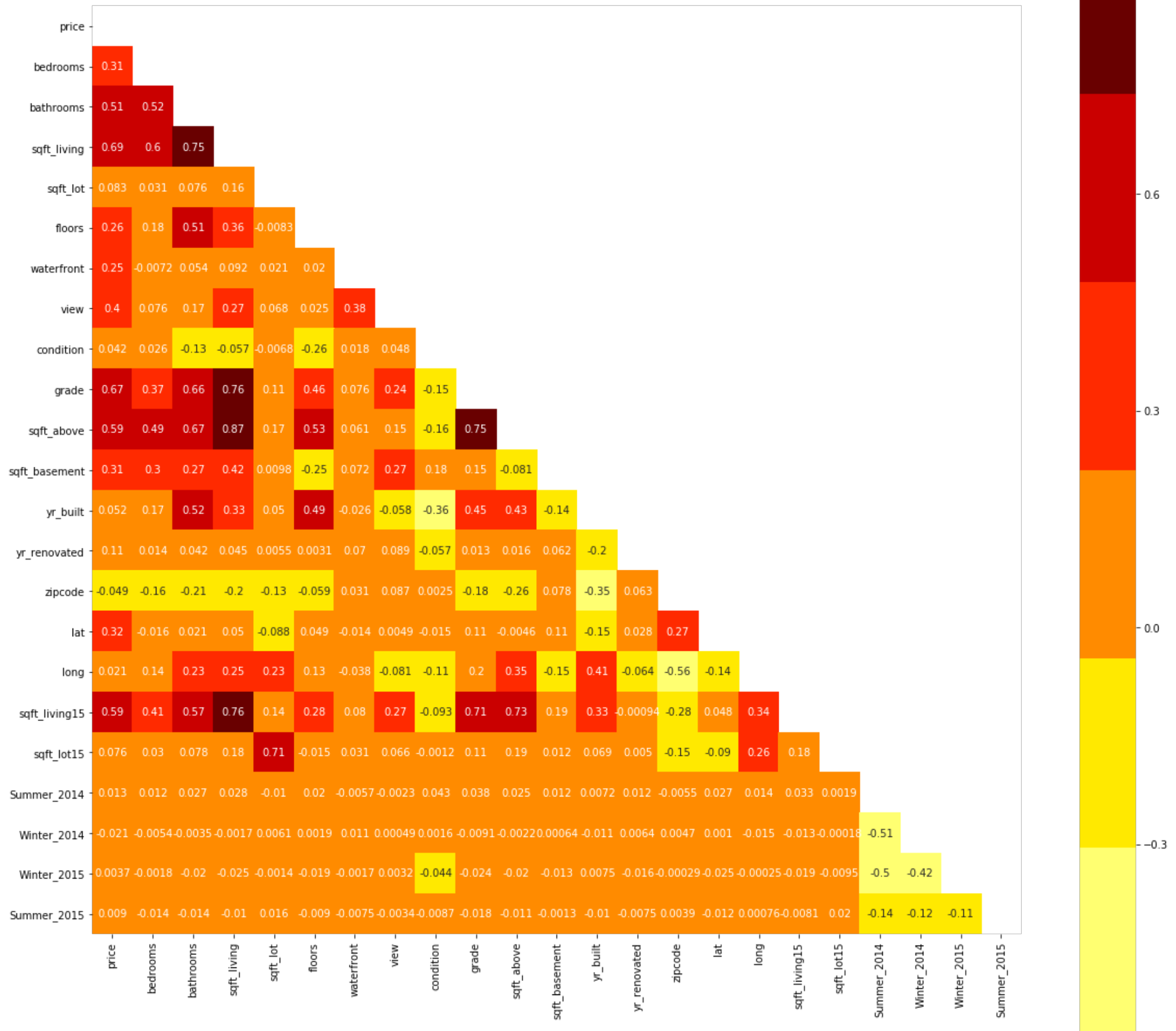


KING COUNTY HOUSING DATA 2014/2015

A Multilinear Regression Odyssey

Highly Correlated Features

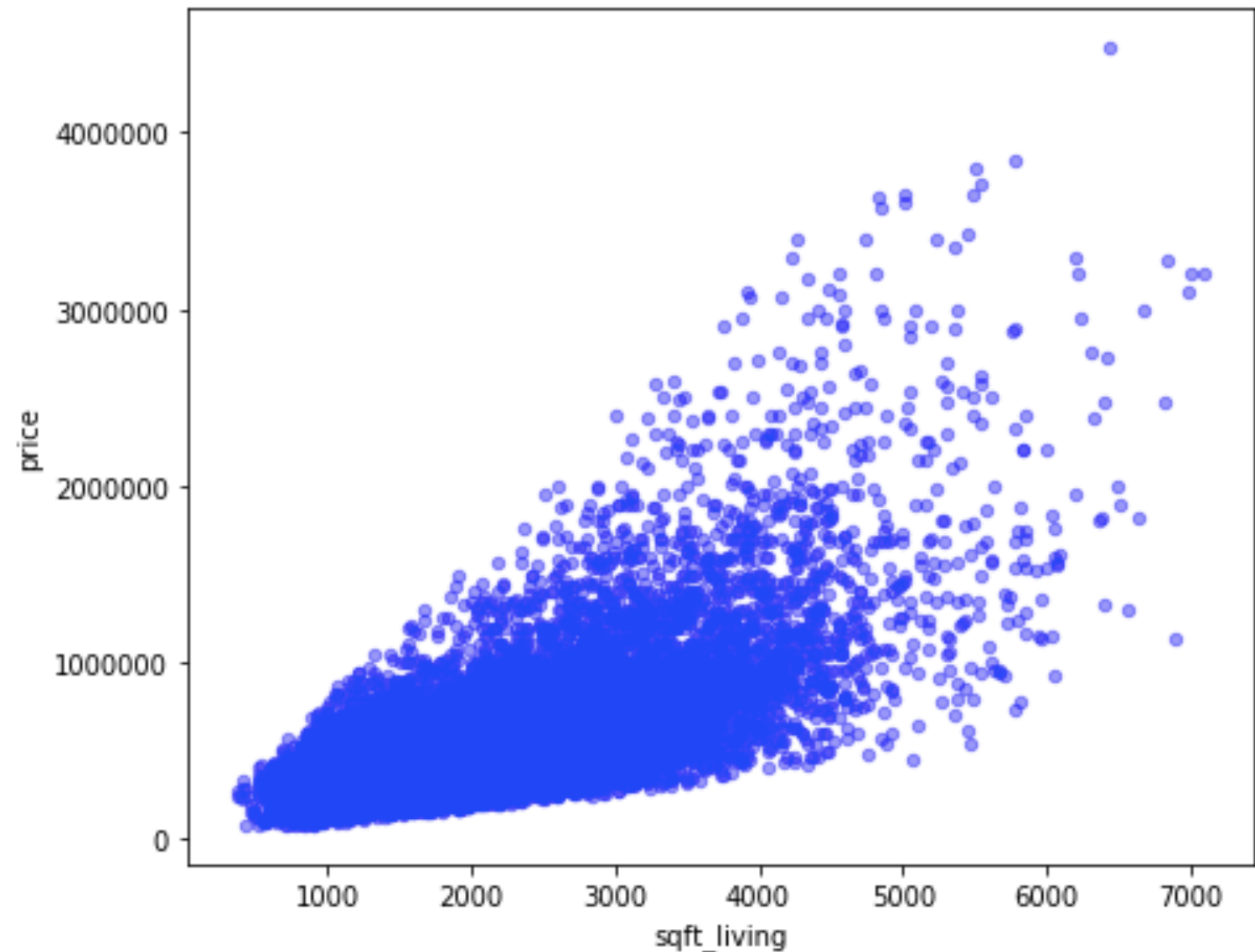


MOST HIGHLY CORRELATED INDIVIDUAL FEATURES

- ▶ Square footage of the living area (incl. square footage of the basement, if applicable).
- ▶ Square footage of the living area minus basement.
- ▶ Square footage of the living area of the closest 15 neighbors (see a pattern?).
- ▶ Grade, or construction quality of the house.

Square Footage of the Living Area

- Fairly linear (one to one) relationship between price and square footage.
- Not perfect though. Some 5000sqft houses sold for less than one million dollars, while some sold for almost four million.

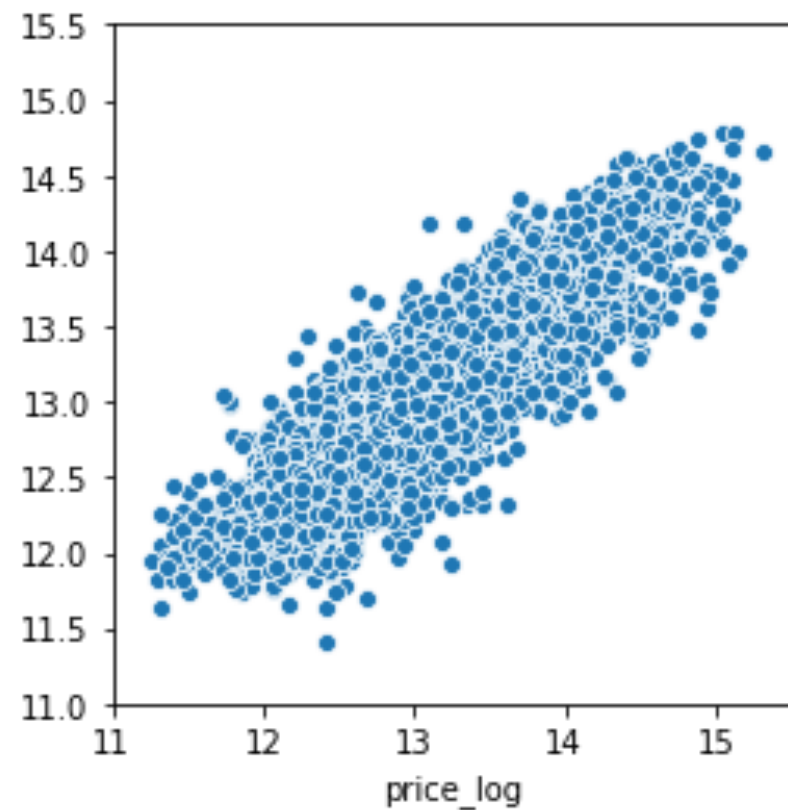


SO WHAT FEATURES DO WE USE?

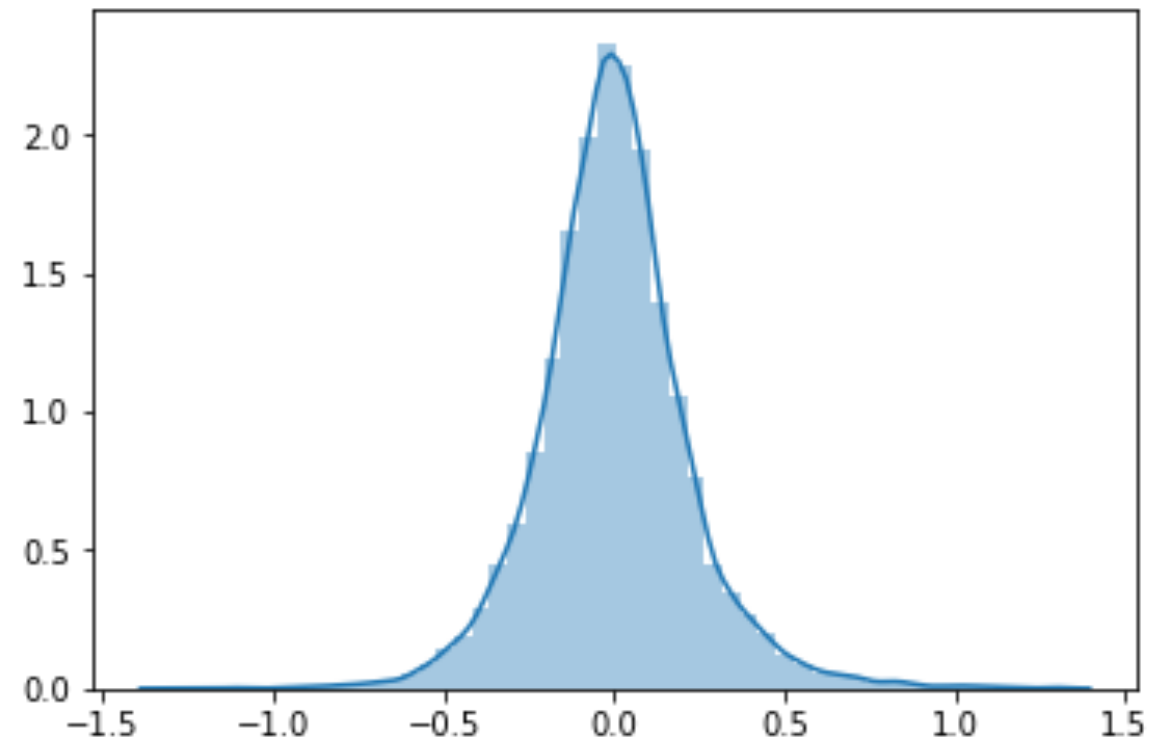
- ▶ Square footage of the living area.
- ▶ Zipcode

And here's what we get:

Prediction vs. actual



Residual errors



QQ Plot

