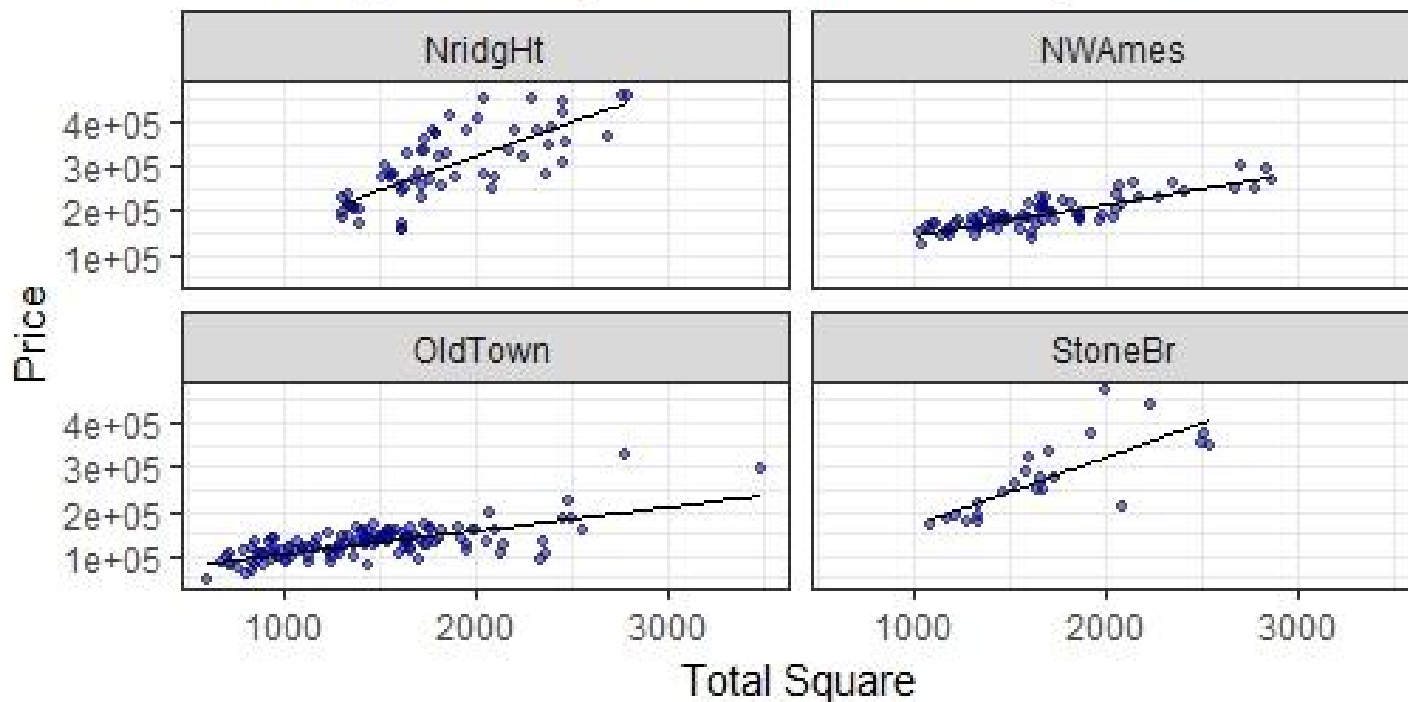# STA 521 Final Project
Prediction of House Prices

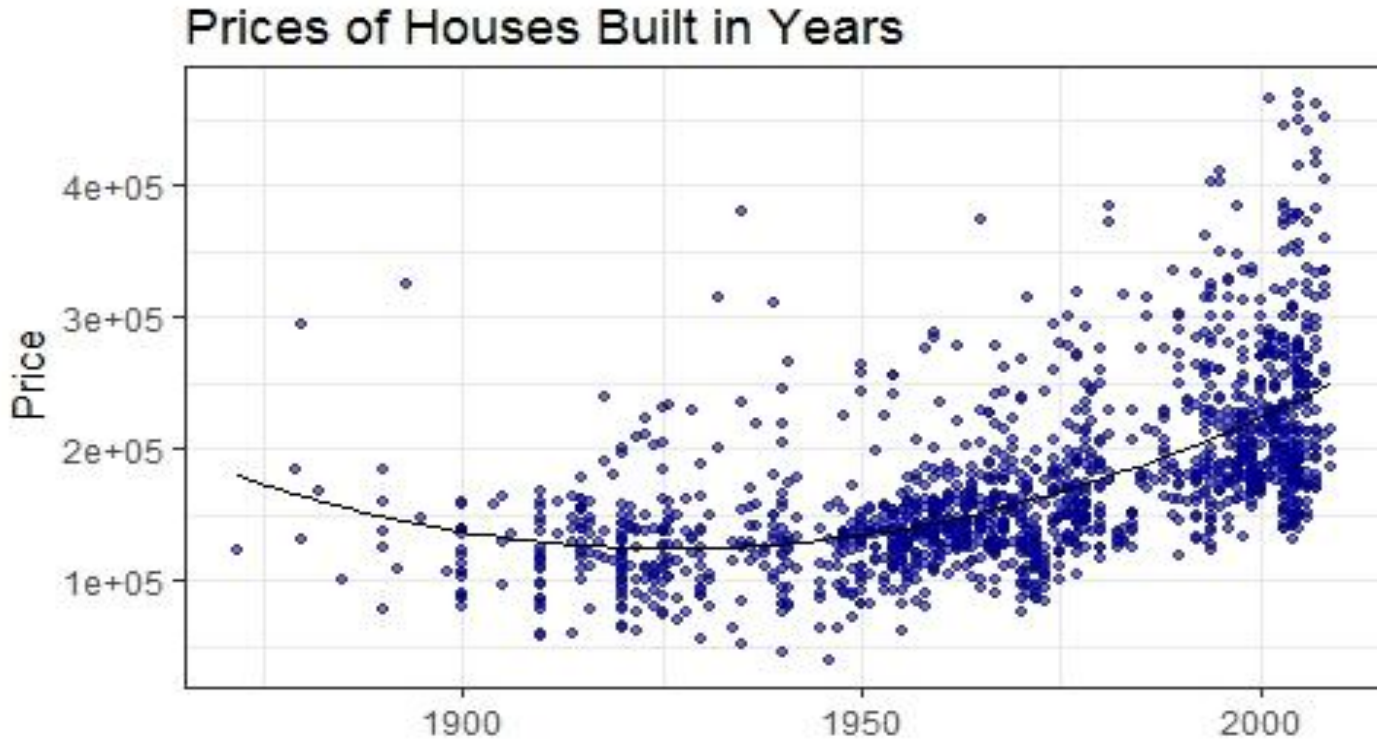BayeStar

Exploratory Data Analysis

Prices by Total Square in Different Neighborhoods
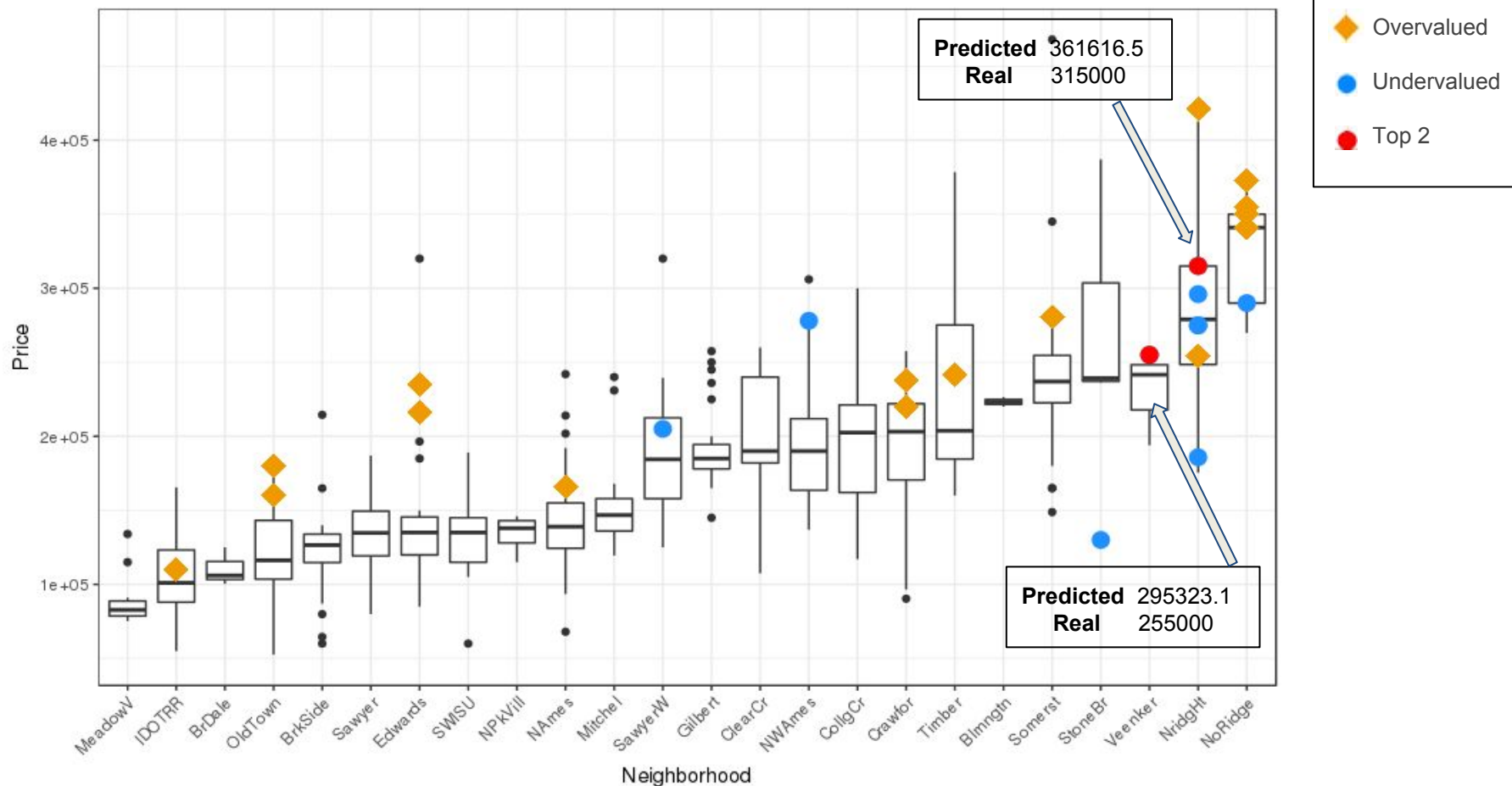
Exploratory Data Analysis

Prices of Houses Built in Years

Linear Regression

- Remove variables with few unique values
  - Contain no information (E.g. Utilities)
- Transformations
  - Log transformation (E.g. log(Lot.Area))
  - Nonlinear relationship (E.g. poly(Year.Built,3))
  - Numerical -> Categorical (E.g. factor(MS.SubClass))
- Interactions
  - Numerical-categorical pair
  - Describe the same features

    (E.g. Lot.Area & Lot.Shape)

Prices by Neighborhood

Predicted 361616.5
Real 315000

Predicted 295323.1
Real 255000

Overvalued
Undervalued
Top 2

Price

Neighborhood

MeadowV IDOTRR BrDale OldTown BrkSide Sawyer Edwards SWISU NPkVill NAmes Mitchel SawyerW Gilbert ClearCr NWAmes CollgCr Crawfor Timber Blmngtn Somerst StoneBr Veenker NridgHt NoRidge

Duke UNIVERSITY

Thank You!