

PREDICTING HOUSING PRICES

Informative Pricing Website

PROBLEM STATEMENT

- Given a housing dataset, create a model that predicts the sale price of a house
- Location: King County, Washington

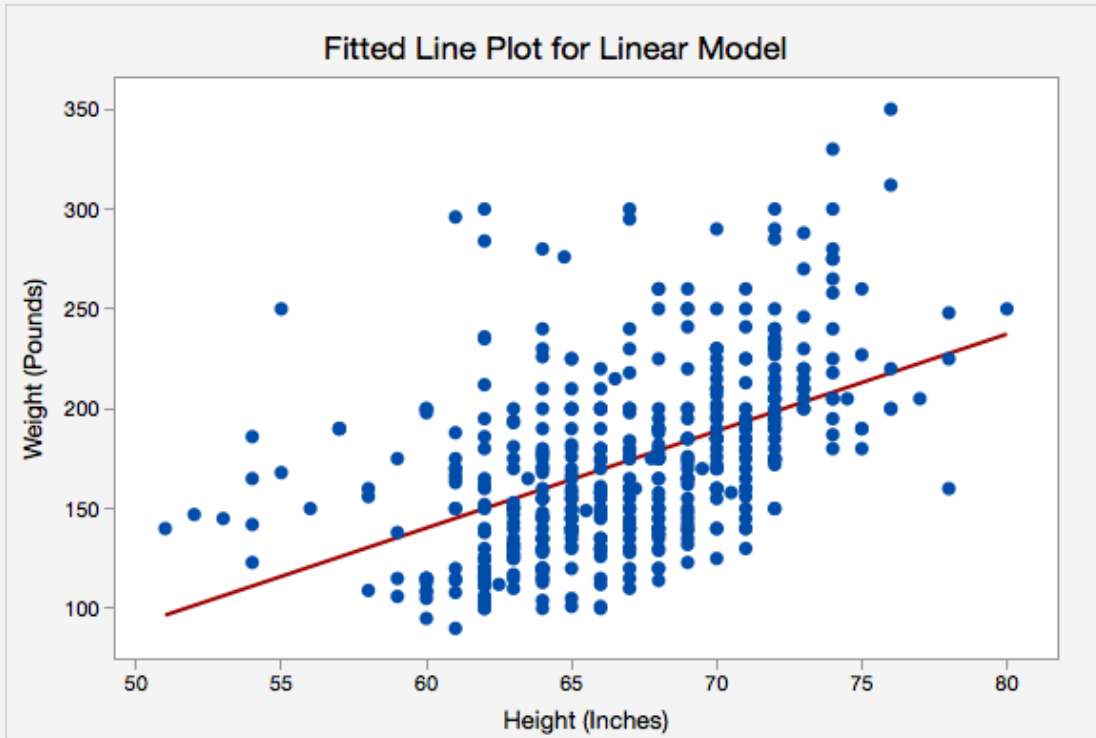


BUSINESS VALUE

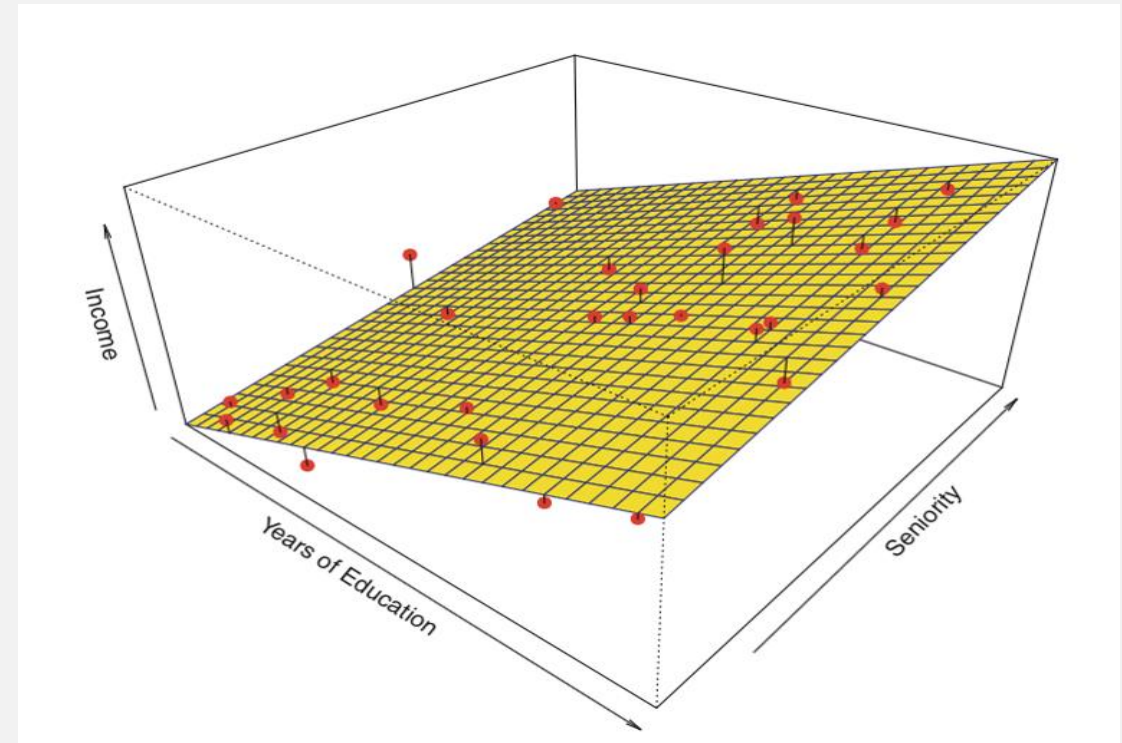
- Create a website that predicts the selling price of a home
- Users input information about their home / a theoretical home
 - Info is run through a predictive model
 - Suggested price = output
 - Useful for
 - Homeowner looking to sell
 - People searching for home with particular features
 - Home builder
 - Anyone with a simple interest in what a certain type of home would cost



METHODOLOGY: PREDICTIVE MODELING

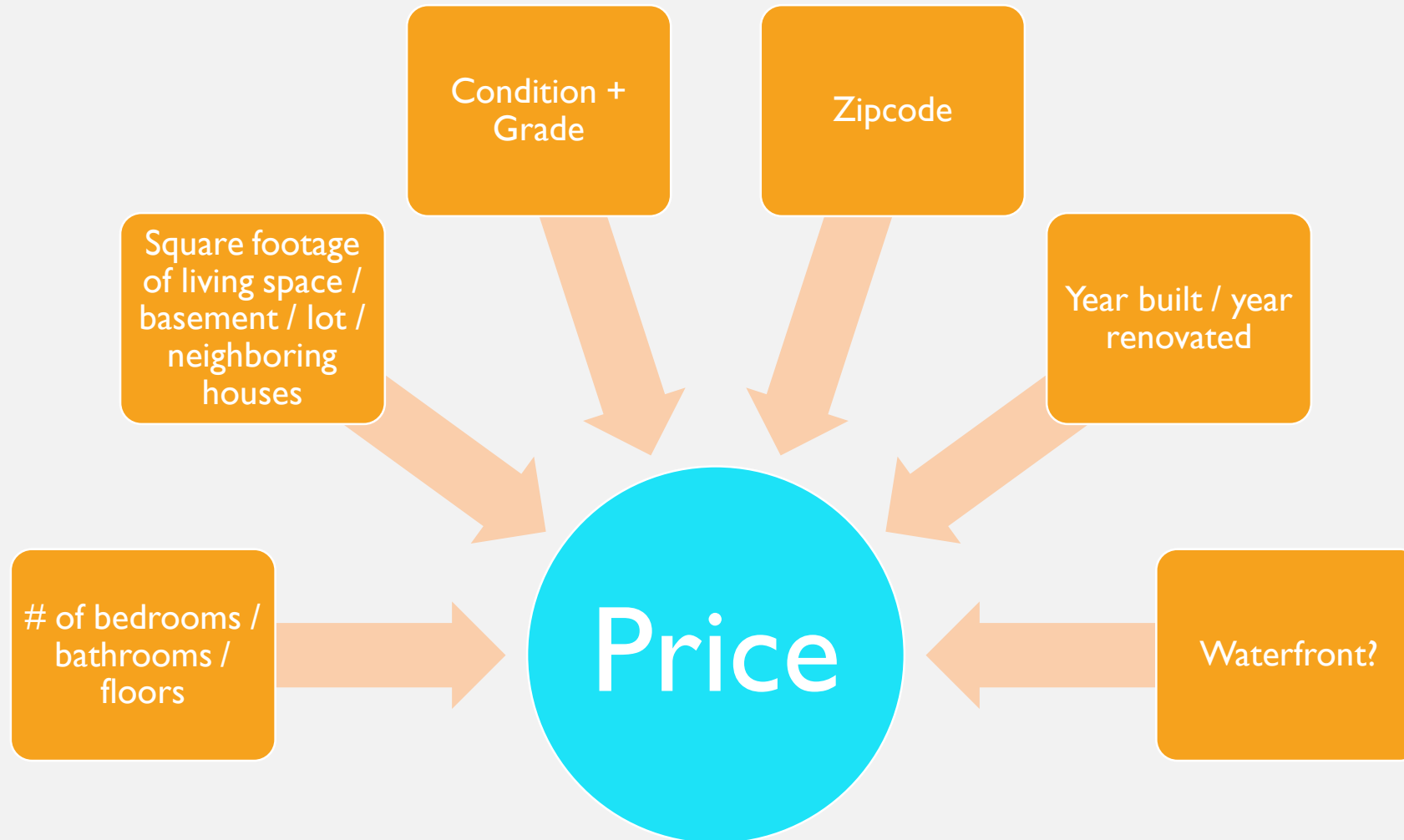


Simple Linear Regression

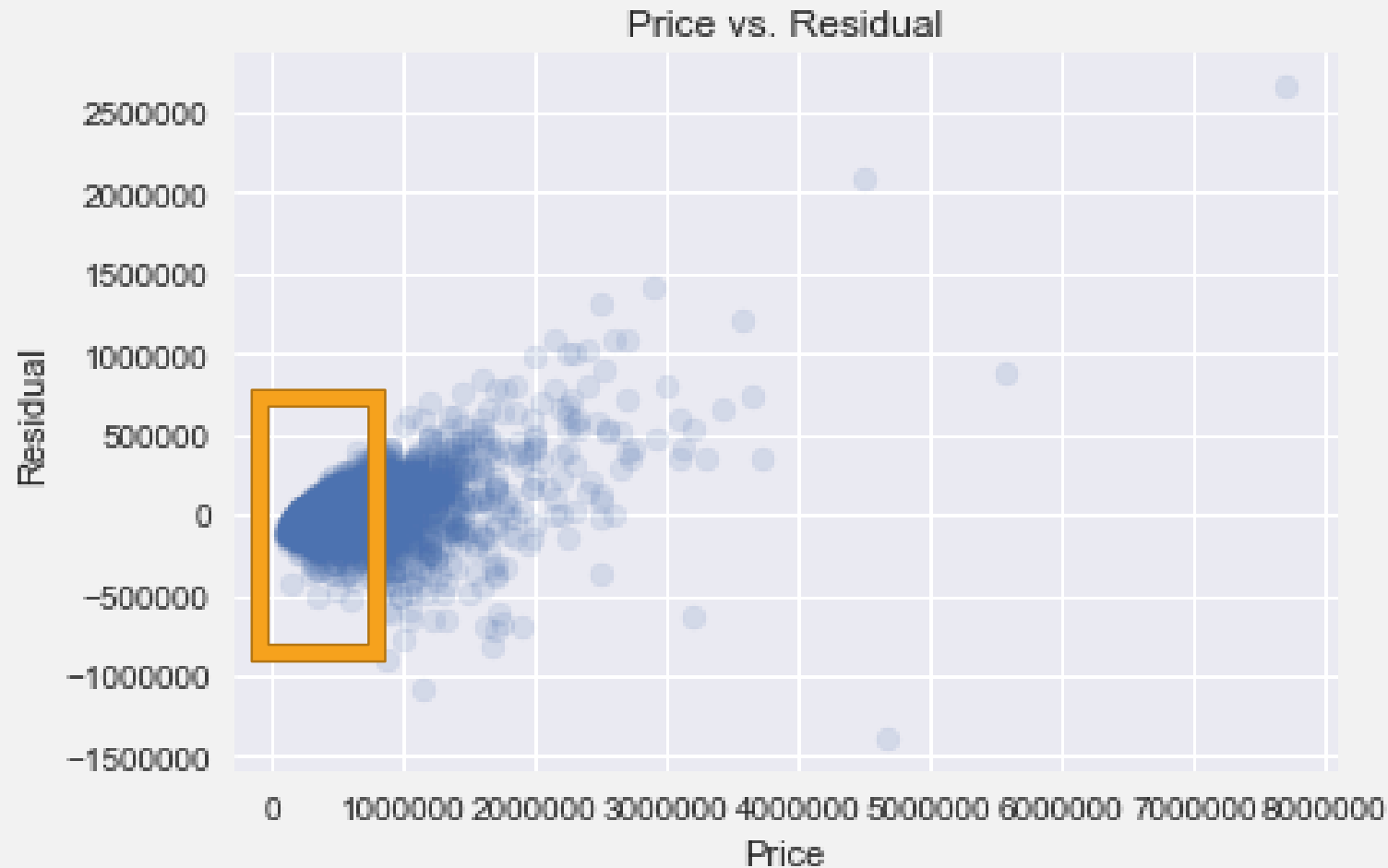


Multiple Linear Regression

VARIABLES: MODEL INPUTS



FINDINGS



$\text{Residual} = \text{Actual} - \text{Prediction}$



Comprises 70.5% of dataset | Average error \approx \$54,200



Comprises 58.1% of dataset | Average error \approx \$50,000



Comprises 40.9% of dataset | Average error \approx \$45,500



Comprises 20.5% of dataset | Average error \approx \$41,300

BUSINESS RECOMMENDATIONS

MODEL HOMES UNDER \$320,000

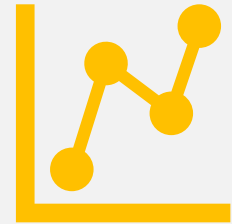


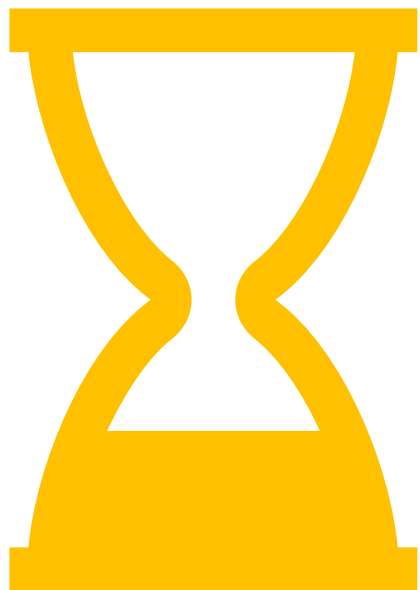
Comprises 24.4% of dataset | Average error \approx \$42,000

- Includes 24.4% of all homes in dataset
- Over 21,000 sales tracked
 - May 2nd, 2014 – May 27th, 2015
- Average error \approx \$42,000
- King County ranked 12th most populous county in US
 - > 2.2 million people (July 1, 2018 official estimate)
- As of July 1, 2018, 45 counties have population > 1 million people
- Potential for scores of thousands of website users

SUMMARY

- Modeling → Predict housing prices
- Creation of a website that uses predictive model
 - Users enter housing features
 - Output = pricing prediction
- Potential users
 - Homeowners looking to sell
 - People seeking a home
 - Home builders
 - Those curious about prices in certain county / zip code





FUTURE WORK

- Tune the current model to create more accurate estimates
- Test different modeling techniques
- Model other counties / zip codes
- Analyze residual vs price ratio trend

THANK YOU

SOURCES

- Single linear regression picture:
 - <https://online.stat.psu.edu/stat200/lesson/12/12>.
- County population information
 - https://en.wikipedia.org/wiki/List_of_the_most_populous_counties_in_the_United_States