

- Predicting Residential Real Estate Prices in King County, WA Using MLR

# 1

## METHODOLOGY

- Scrub  $\Leftrightarrow$  Explore  $\Leftrightarrow$  Modify
- Prep  $\Leftrightarrow$  Select
- Model ↻

## Scrubbing, Exploring, and Modifying Data



### Scrubbing

Remove duplicates

Categorize data

Clean data



### Exploring

Identify patterns

Develop mental model

Establish system relationships



### Modifying

Transform data

Create data

Cast data

### King County Housing Data

id	Price	Bedrooms	Bathrooms
<del>1000102</del>	280000	6	3
<del>1000102</del>	300000	6	3
<del>7200179</del>	150000	2	1
<del>7200179</del>	175000	2	1
<del>109200390</del>	245000	3	1.75
<del>109200390</del>	250000	3	1.75
<del>123039336</del>	148000	1	1
<del>123039336</del>	244900	1	1
<del>251300110</del>	225000	3	2.25
<del>251300110</del>	358000	3	2.25

## Scrubbing, Exploring, and Modifying Data



### Scrubbing

- Remove duplicates
- Correctly categorize data
- Clean nonsense data



### Exploring

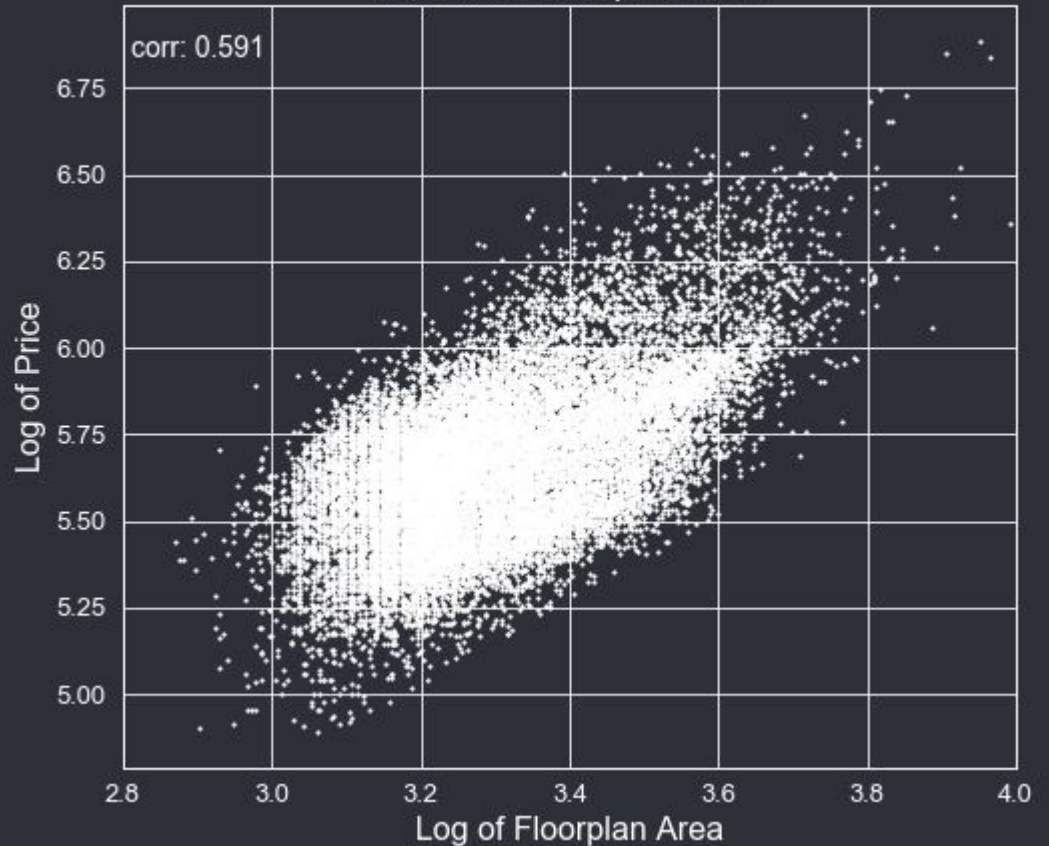
- Identify patterns
- Develop mental model
- Establish system relationships



### Modifying

- Transform data
- Create data
- Cast data

Exploring for relationships between  
Price and Floorplan Area



## Scrubbing, Exploring, and Modifying Data



### Scrubbing

- Remove duplicates
- Correctly categorize data
- Clean nonsense data



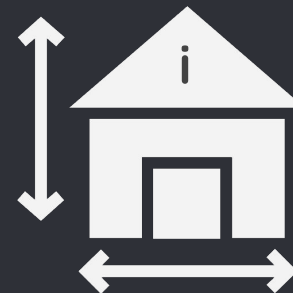
### Exploring

- Identify patterns
- Develop mental model
- Establish system relationships



### Modifying

- Transform data
- Create data
- Cast data



## Prep Data and Select Features



### Prep

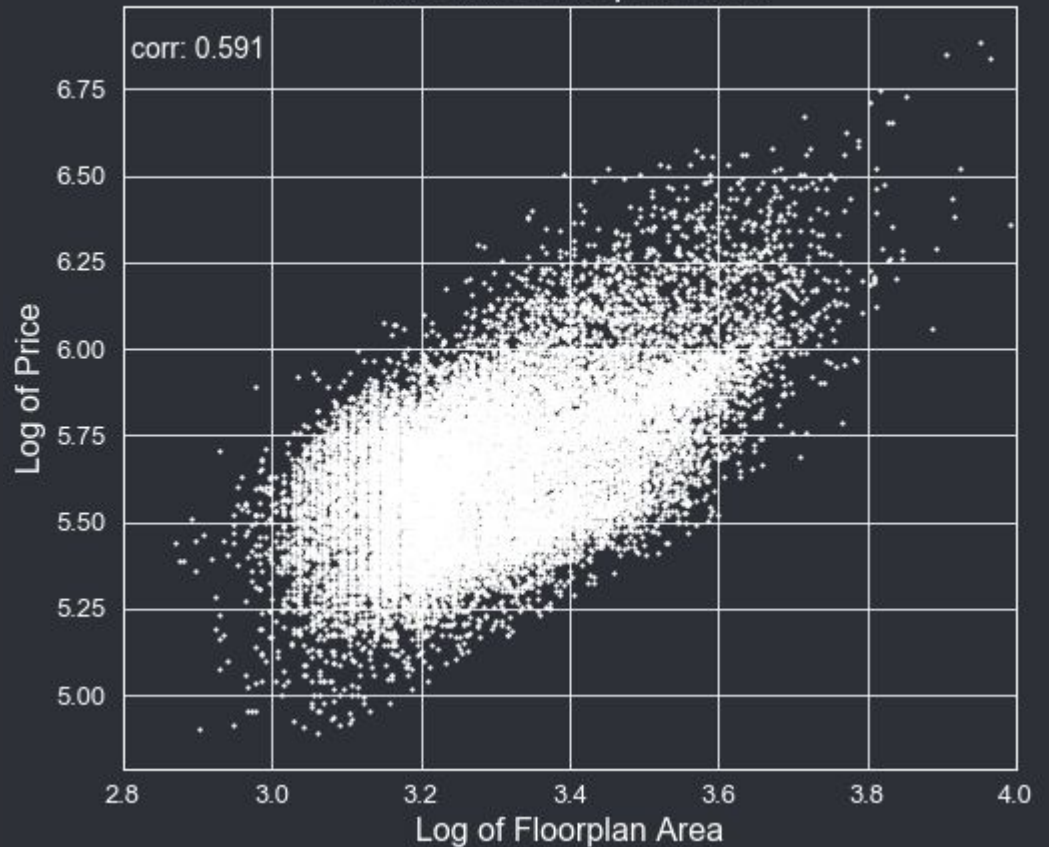
- Linearize data
- Normalize data
- Split data



### Select

- Computer selection
- Statistical elimination
- Common sense elimination

Exploring for relationships between  
Price and Floorplan Area



## Prep Data and Select Features



### Prep

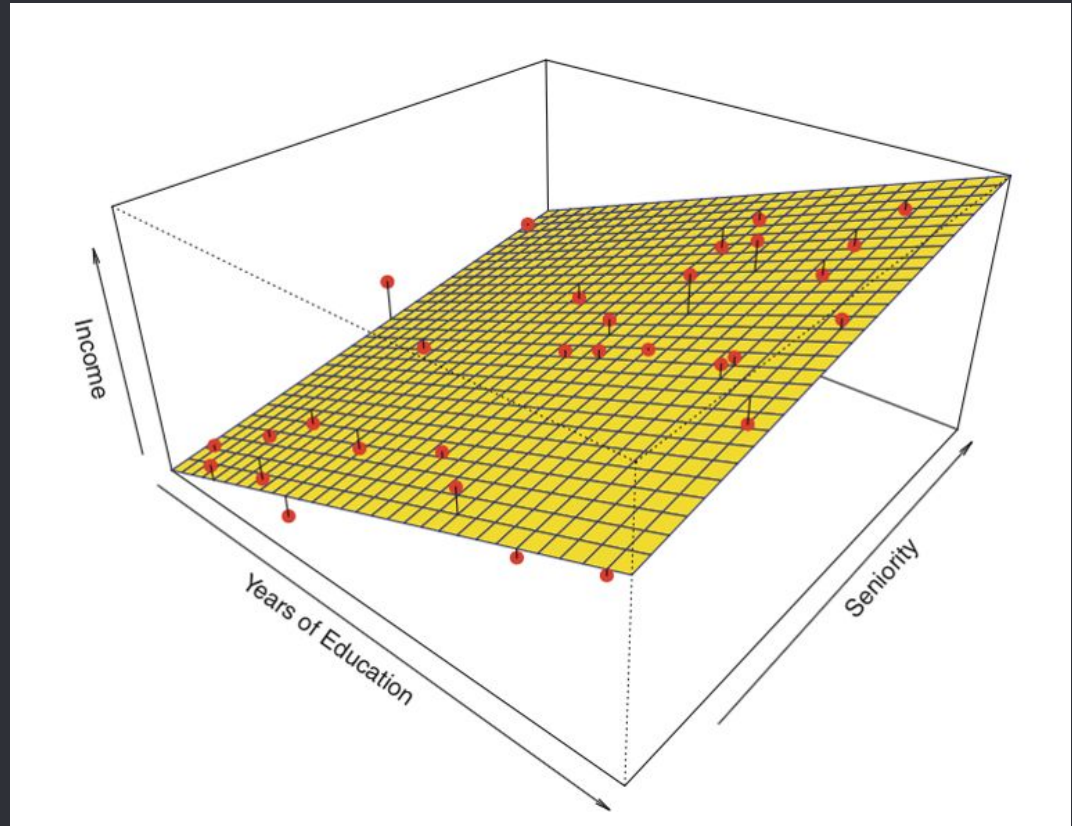
Linearize data  
Normalize data  
Split data



### Select

Computer selection  
Statistical elimination  
Common sense elimination

	coef	std err	t	P> t
<b>Zipcode Average</b>	0.3573	0.002	164.692	0
<b>Floor Plan Area</b>	0.2702	0.006	42.658	0
<b>Grade</b>	0.2082	0.005	41.052	0
<b>Area Compared to Neighbors</b>	0.1502	0.005	27.45	0
<b>Basement Area</b>	0.0514	0.006	9.069	0
<b>Bedrooms</b>	0.0132	0.004	3.302	0.001
<b>Bathrooms</b>	-0.0143	0.005	-3.07	0.002
<b>Waterfront (y/n)</b>	0.1076	0.005	22.632	0
<b>View (y/n)</b>	0.0425	0.001	31.11	0
<b>Renovated (y/n)</b>	0.0191	0.002	9.445	0
<b>Basement (y/n)</b>	0.016	0.002	7.18	0
<b>April (y/n)</b>	0.0174	0.001	14.077	0
<b>March (y/n)</b>	0.0099	0.001	7.479	0



# Multiple Linear Regression

Calculates the “plane” that is closest to all of the data.



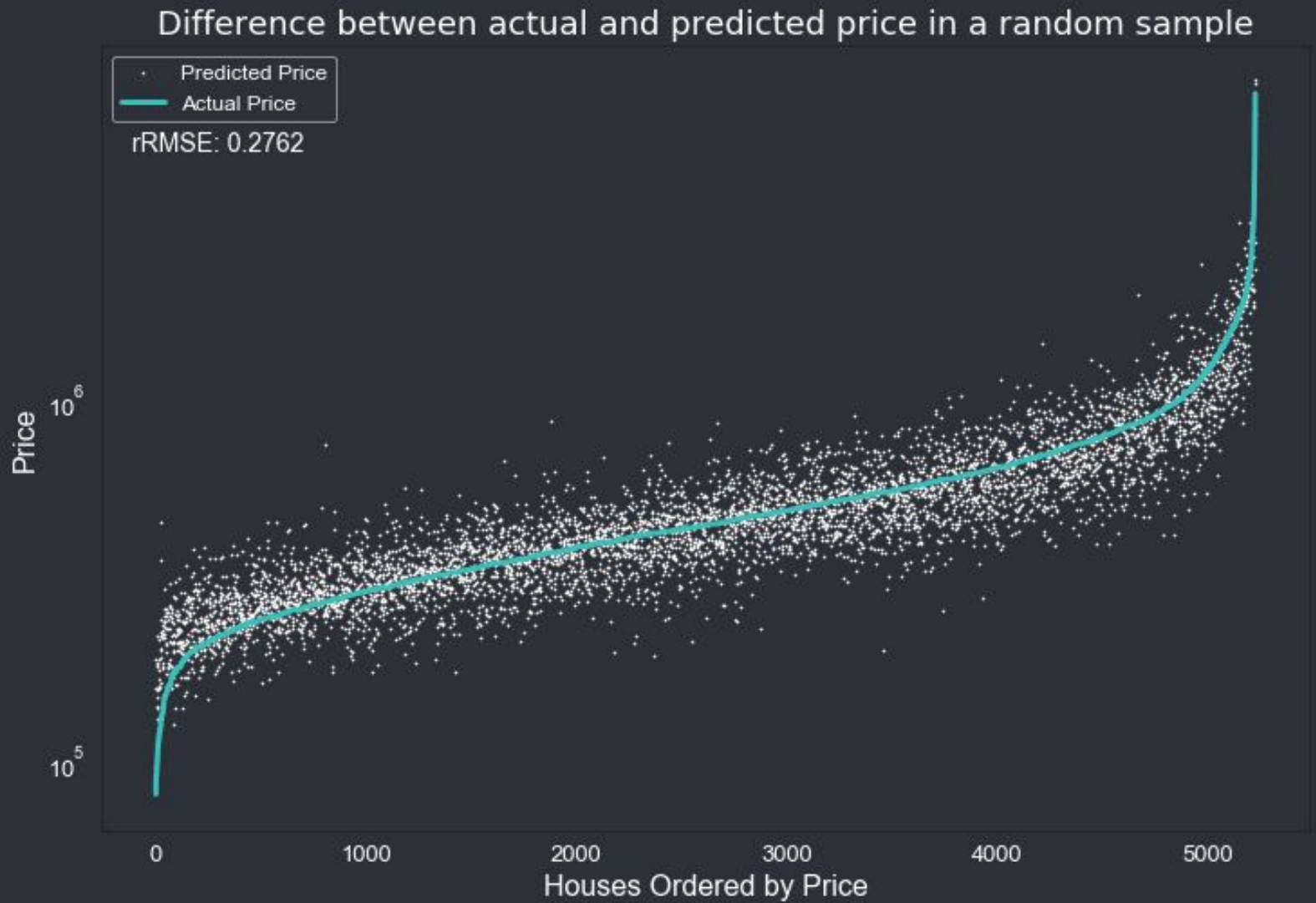


## 2

## FINDINGS

- Results
- Interpretation
- Recommendations

## Results



## Interpretation



### Location

Most impactful predictor

Accounts for the majority of the price



### House Size

Above-ground size is most important

Peers are important



### Conditions

Have a significant effect despite their binary nature

Other conditions modify price

Feature	Coefficient
Zipcode Average	0.3556
Area Above	0.266
Grade	0.1869
Area Compared to Neighbors	0.1308
Area Basement	0.0814
Year Built	0.0314
Waterfront (y/n)	0.1191
View (y/n)	0.0428
Renovated (y/n)	0.0205
April (y/n)	0.0155
March (y/n)	0.0086
Basement (y/n)	0.0076

## Interpretation



### Location

Most impactful predictor

Accounts for the majority of the price



### House Size

Above-ground size is most important

Peers are important



### Conditions

Have a significant effect despite their binary nature

Other conditions modify price

Feature	Coefficient
Zipcode Average	0.3556
Area Above	0.266
Grade	0.1869
Area Compared to Neighbors	0.1308
Area Basement	0.0814
Year Built	0.0314
Waterfront (y/n)	0.1191
View (y/n)	0.0428
Renovated (y/n)	0.0205
April (y/n)	0.0155
March (y/n)	0.0086
Basement (y/n)	0.0076

## Interpretation



### Location

Most impactful predictor

Accounts for the majority of the price



### House Size

Above-ground size is most important

Peers are important



### Conditions

Have a significant effect despite their binary nature

Other conditions modify price

Feature	Coefficient
Zipcode Average	0.3556
Area Above	0.266
Grade	0.1869
Area Compared to Neighbors	0.1308
Area Basement	0.0814
Year Built	0.0314
Waterfront (y/n)	0.1191
View (y/n)	0.0428
Renovated (y/n)	0.0205
April (y/n)	0.0155
March (y/n)	0.0086
Basement (y/n)	0.0076

**+** *Accurate predictions of average price for a particular set of features ( $\pm 1.4\%$ ) with 99% confidence*

**△** *Poor ability to predict the price for a particular house in an actionable way ( $\pm \$125,000$ ) for a consumer*

**=** *Use model to understand the market, but develop a hierarchical bayesian model to improve predictive ability*