

Project 1: King County Housing

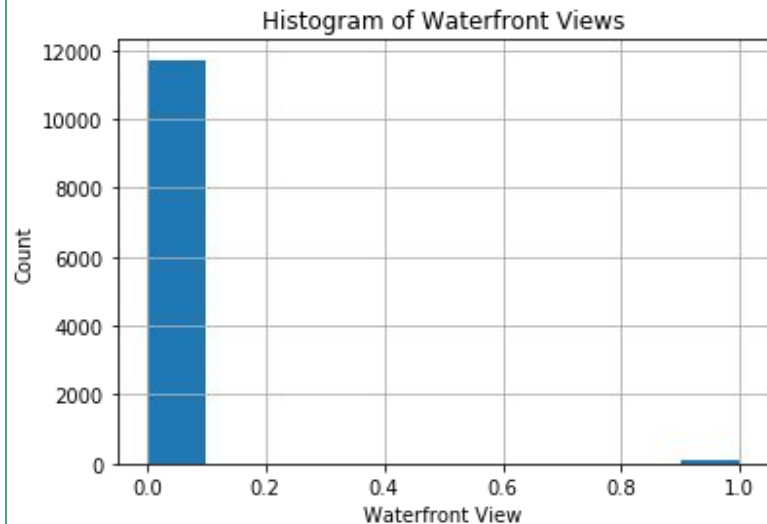
Liz Alvarez

Methodology

Focused on:

- houses with complete data records
- characteristics that differed between houses

id	date	price	bedrooms	bathrooms	...
7129300520	10/13/2014	221900	3	1.00	...
6414100192	12/9/2014	538000	3	2.25	...
5631500400	2/25/2015	180000	2	1.00	...
2487200875	12/9/2014	604000	4	3.00	...
1954400510	2/18/2015	510000	3	2.00	...



Methodology

Goals for Model:

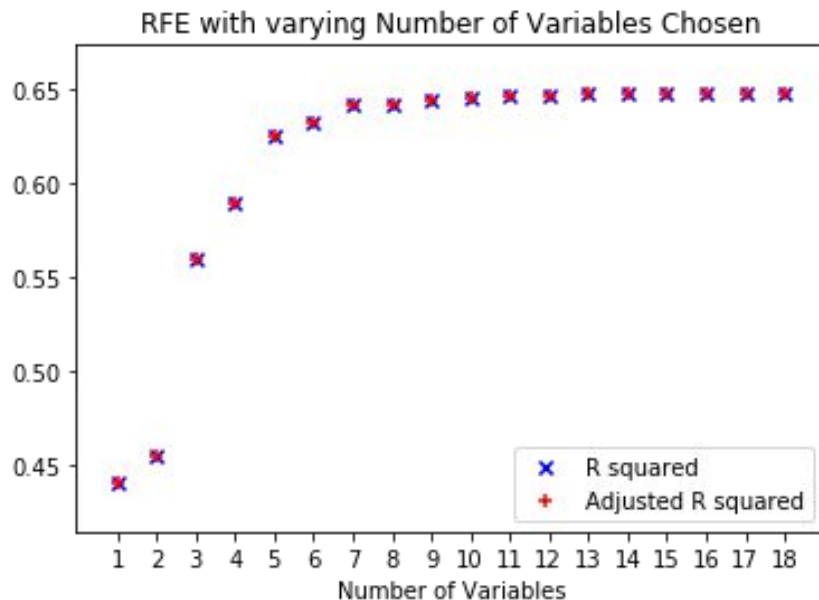
- Accurate (R^2)
 - Cleaned and processed data
 - Created multiple models
 - Refined the best model



Methodology

Goals for Model:

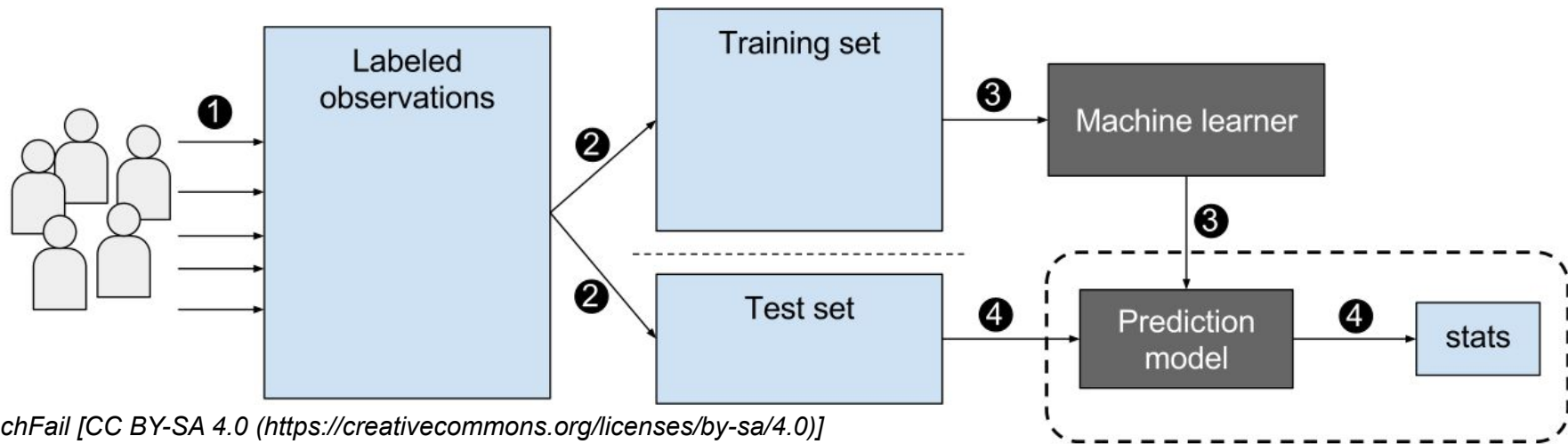
- Accurate (R^2)
- Streamlined (number of predictor variables)
 - Unique variables
 - Most influential variables
 - # of variables vs. accuracy



Methodology

Goals for Model:

- Accurate (R^2)
- Streamlined (number of predictor variables)
- Robust (Predictions with unknown data)



The Model

Base value: \$291,900

* log transformation
~ per standard deviation

Positive Influences	Negative Influences
Grade (+\$177,600)~ Bathrooms (+\$89,600) Views (+\$78,200) Sq. Ft. Above (+\$77,930)*~ Sq. Ft. Basement (+\$50,950)~ Renovated? (+\$32,350) Condition (+\$11,090)	Year Built (-\$111,300)~ Bedrooms (-\$26,450)~

Statistics Summary

- Accurate
 - R^2 : 0.606
 - r : 0.778
- Streamlined
 - 9 variables
- Robust
 - Mean Square Error Ratio (Test/Train): 0.987

Ideal value: 1

