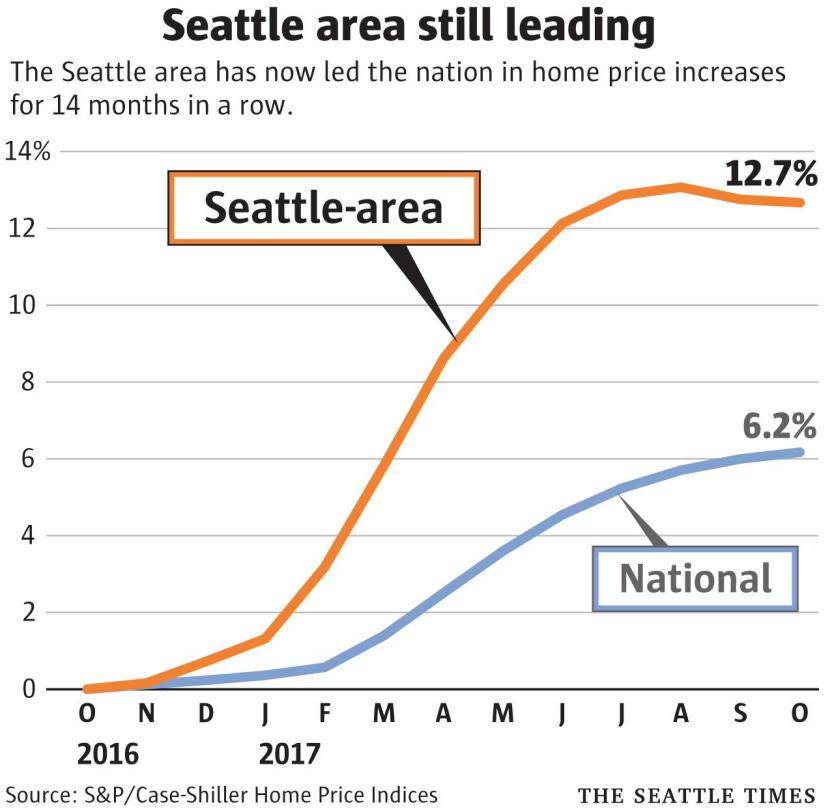
Beating the Zestimate: Predicting Home Price Values in Seattle

Brenner Heintz

Why Study Home Price Data?

- Hottest housing market around
- Companies: Buy undervalued assets
- Buyers vs. Sellers

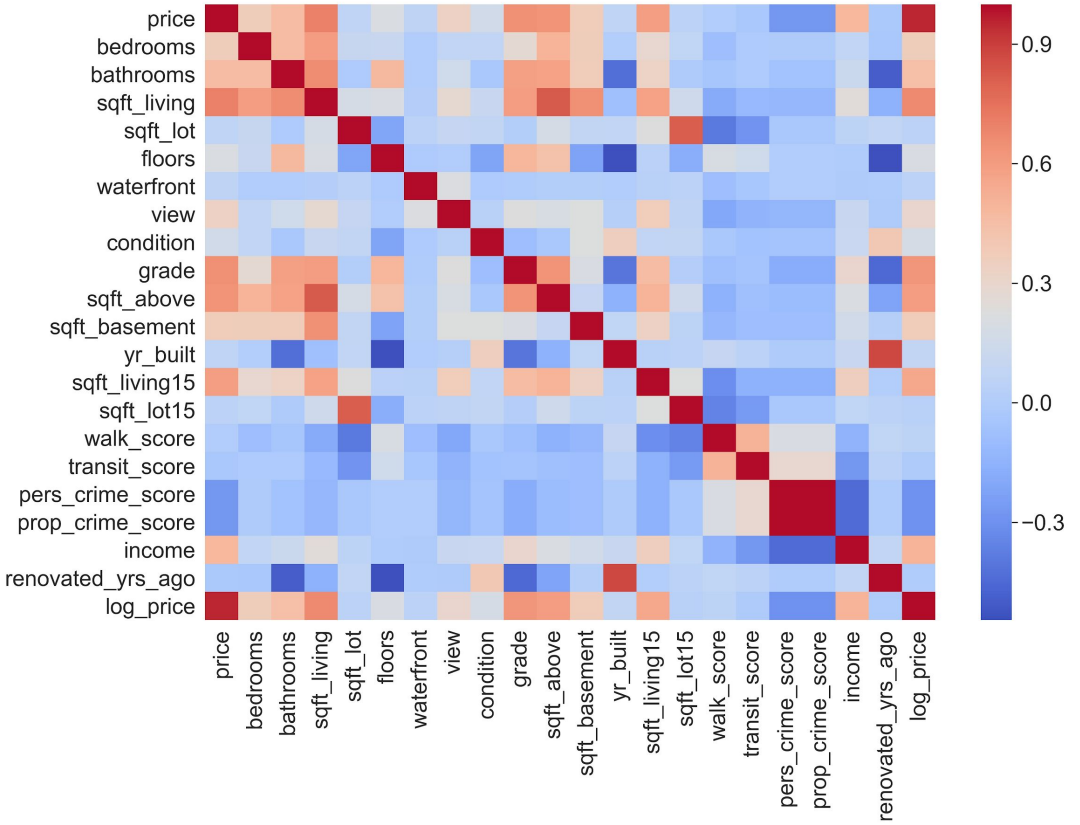
Zillow® Zestimates



Data Sources - Public Data, Web Scraping, APIs

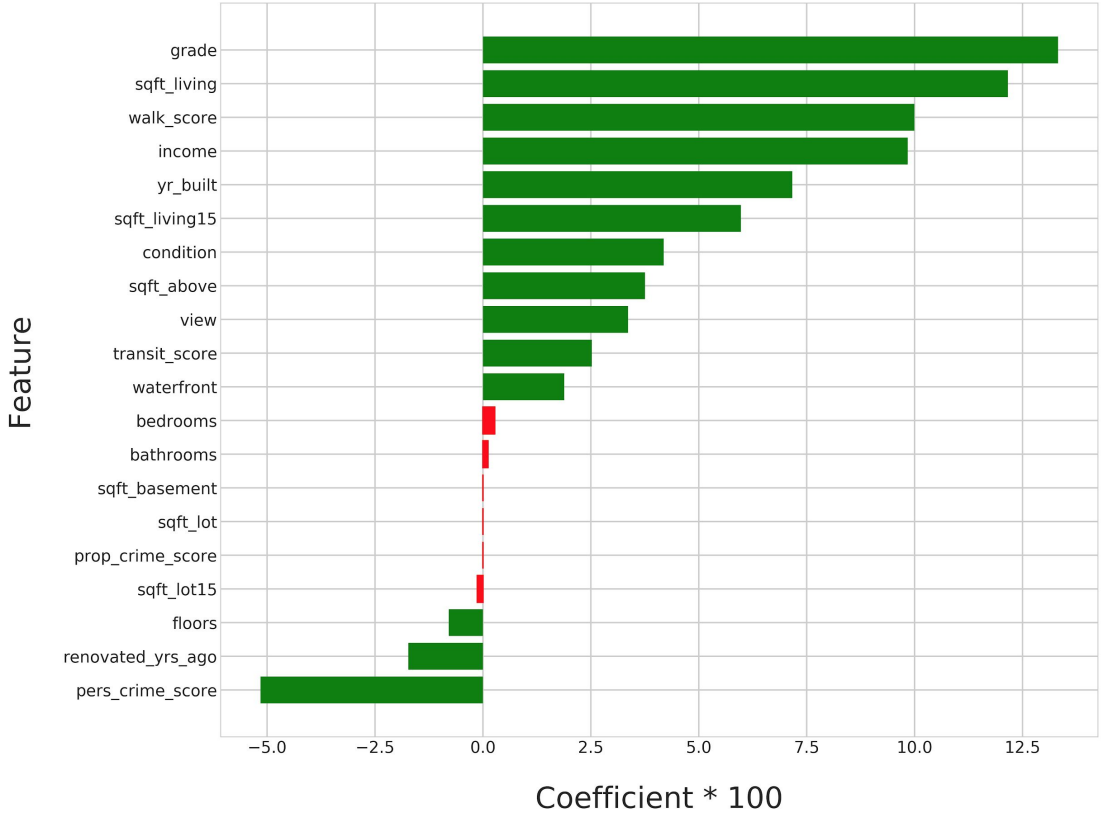
Methodology

- Focus on Seattle
- Dependent Variable: Log of home prices
- Bring in additional features
- Find the signal...



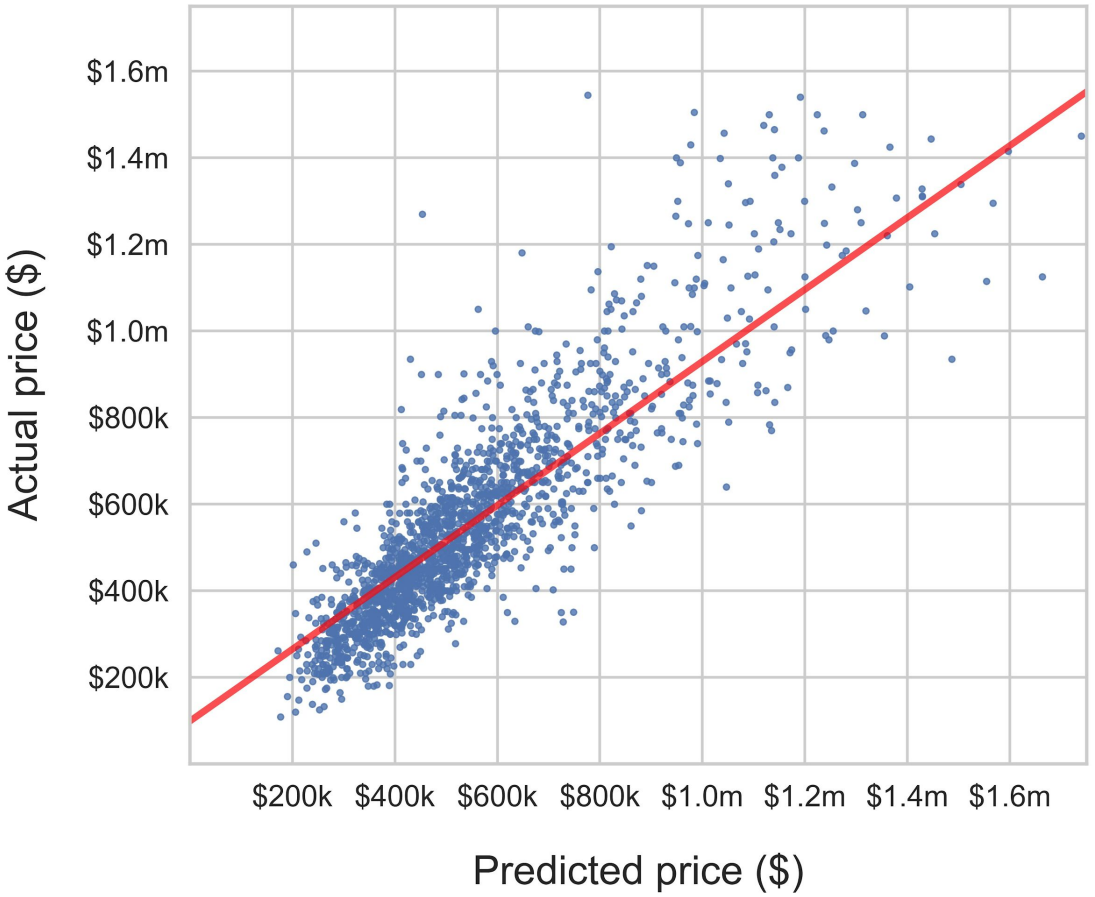
Incremental Progress

- Data Normalization
- Lasso Regression
- Walk Score, Income = Important Features
- Personal vs. Property Crime Scores
- R-squared: **0.741**
- MSE: **.047**



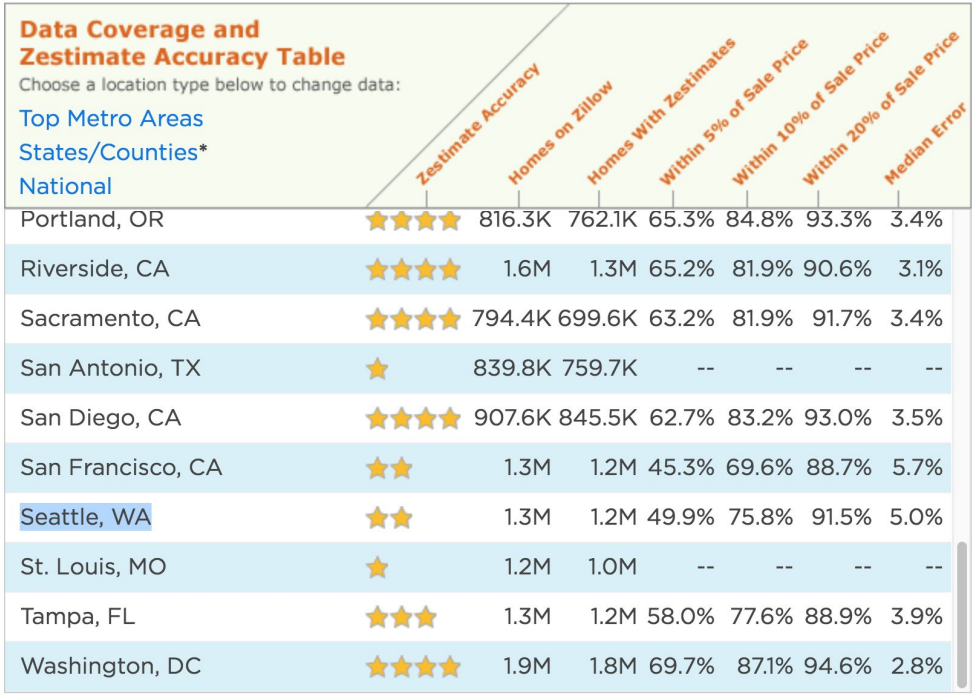
Results

- Removed redundant variables, normalized
- Added polynomial features
- Ridge regression
- R-squared: **0.781 (train), 0.742 (test)**
- MSE: **.040 (train), .046(test)**



Beating the Zillow “Zestimate”

- Zestimate is not that accurate
- 5% median error
- We could beat it, with more features and time



Data Coverage and Zestimate Accuracy Table

Choose a location type below to change data:

Top Metro Areas
States/Counties*
National

	Zestimate Accuracy	Homes on Zillow	Homes With Zestimates	Within 5% of Sale Price	Within 10% of Sale Price	Within 20% of Sale Price	Median Error
Portland, OR	★★★★	816.3K	762.1K	65.3%	84.8%	93.3%	3.4%
Riverside, CA	★★★★	1.6M	1.3M	65.2%	81.9%	90.6%	3.1%
Sacramento, CA	★★★★	794.4K	699.6K	63.2%	81.9%	91.7%	3.4%
San Antonio, TX	★	839.8K	759.7K	--	--	--	--
San Diego, CA	★★★★	907.6K	845.5K	62.7%	83.2%	93.0%	3.5%
San Francisco, CA	★★	1.3M	1.2M	45.3%	69.6%	88.7%	5.7%
Seattle, WA	★★	1.3M	1.2M	49.9%	75.8%	91.5%	5.0%
St. Louis, MO	★	1.2M	1.0M	--	--	--	--
Tampa, FL	★★★	1.3M	1.2M	58.0%	77.6%	88.9%	3.9%
Washington, DC	★★★★	1.9M	1.8M	69.7%	87.1%	94.6%	2.8%

Future Work

- Expand scope to King County
- High end / low end
- School scores

Q & A

Appendix

Importance of Walk Scores

Crime Grade

51 University Street compared to all of Seattle.

C

Personal Crime Grade

Average crime area.

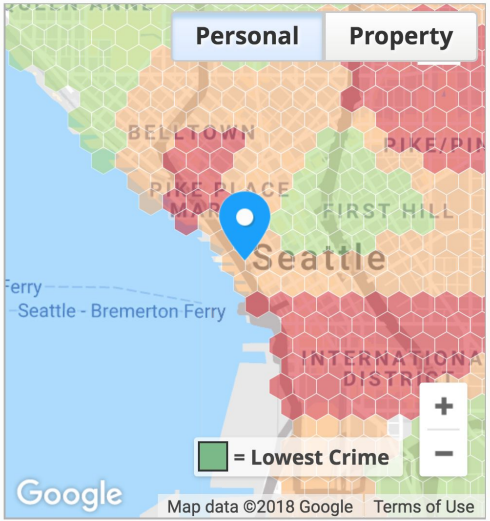
Less Average More

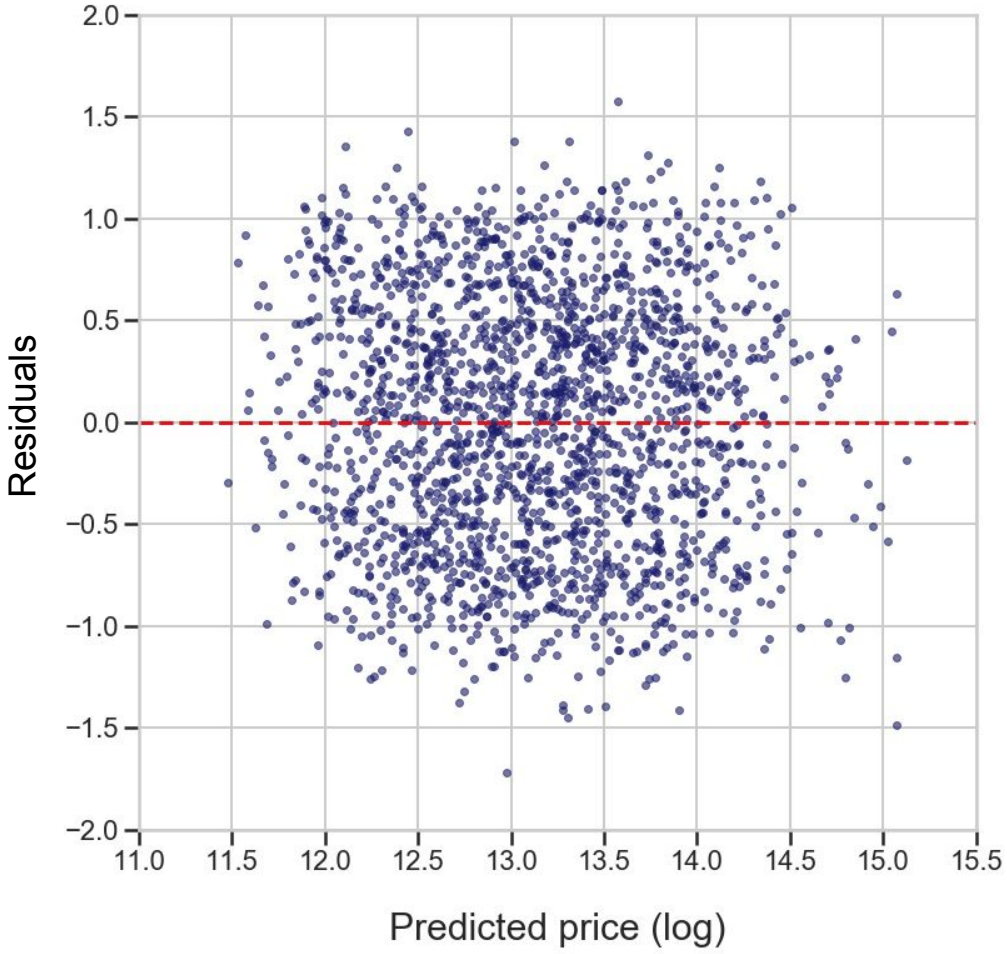
A

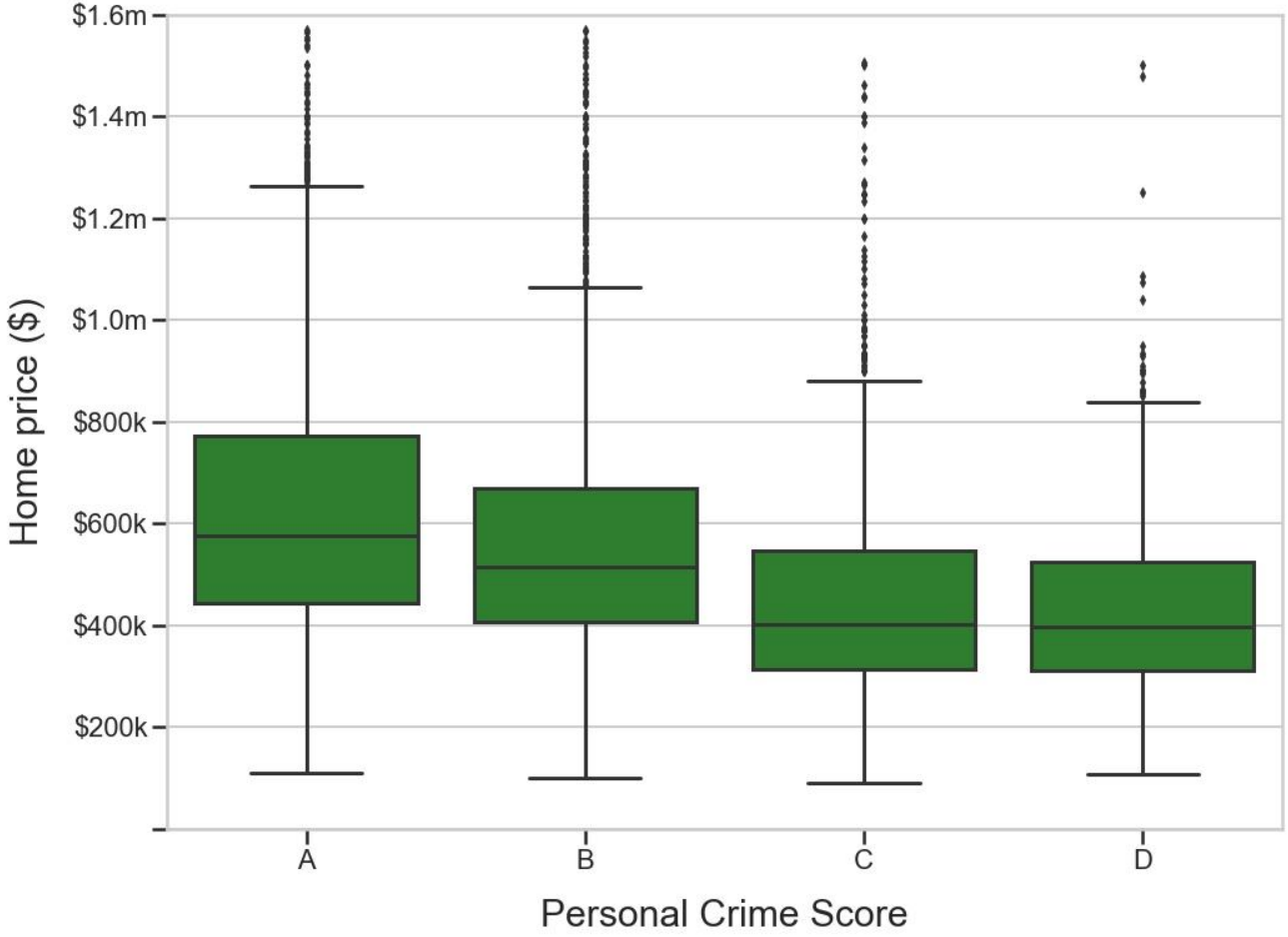
Property Crime Grade

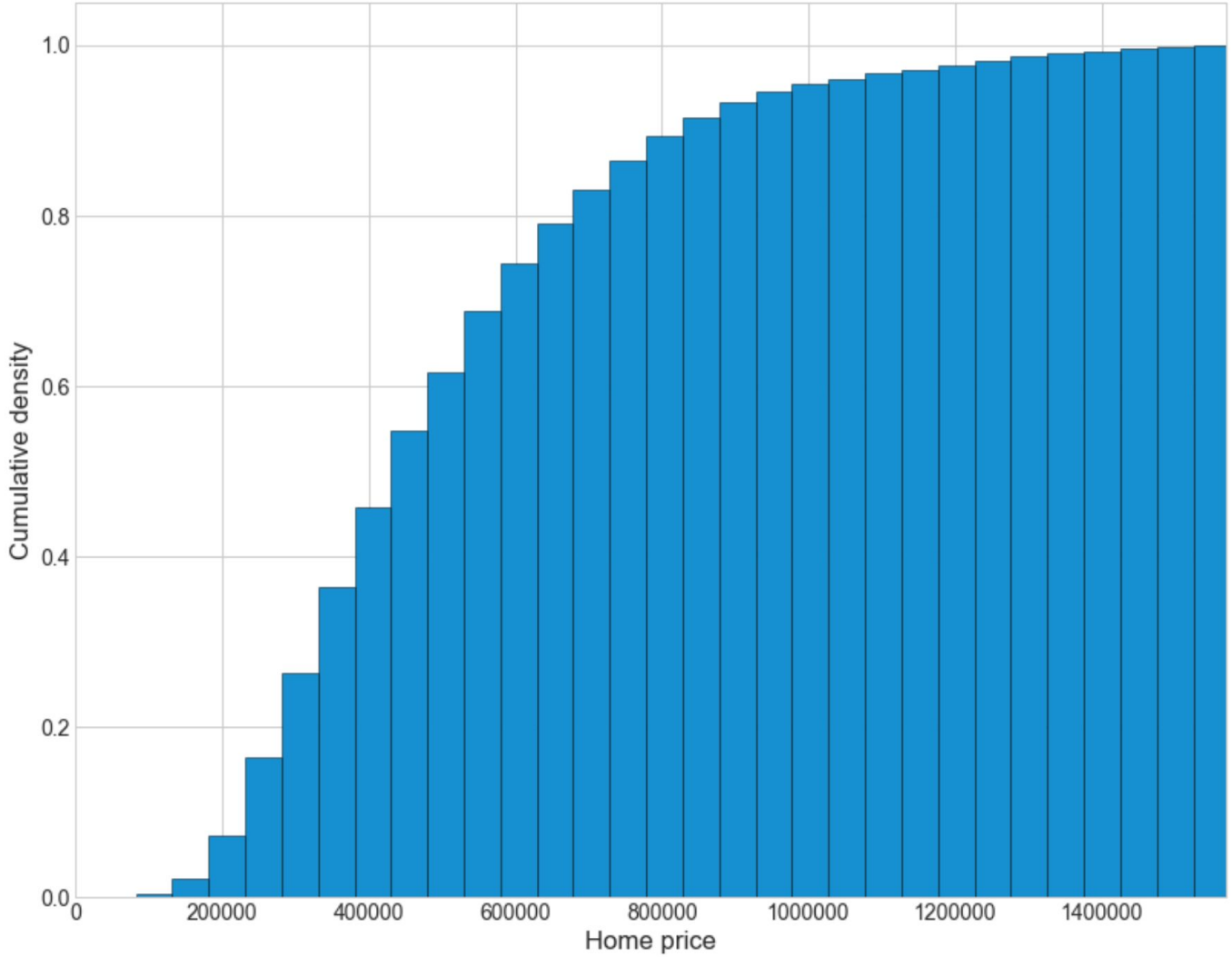
Lowest crime area.

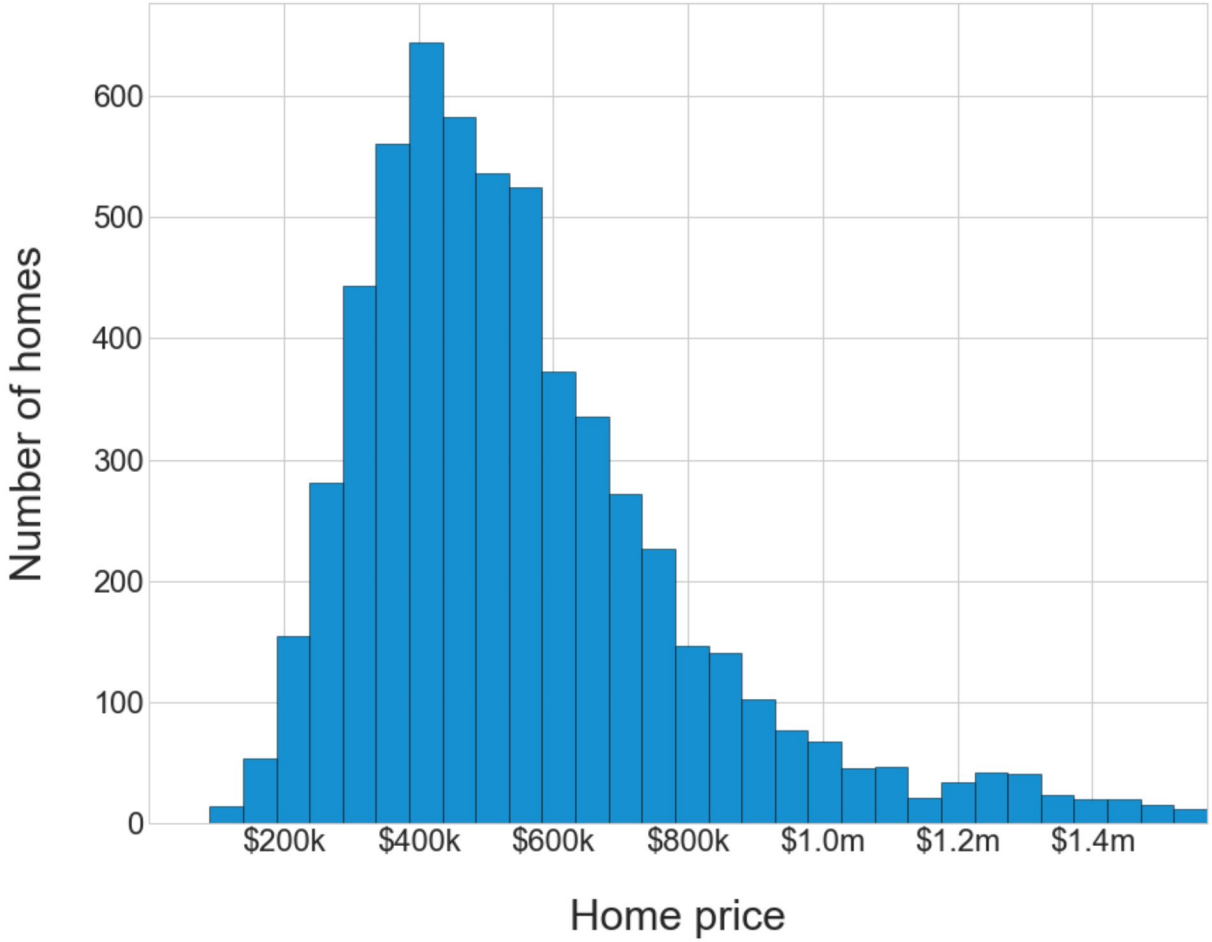
Less Average More

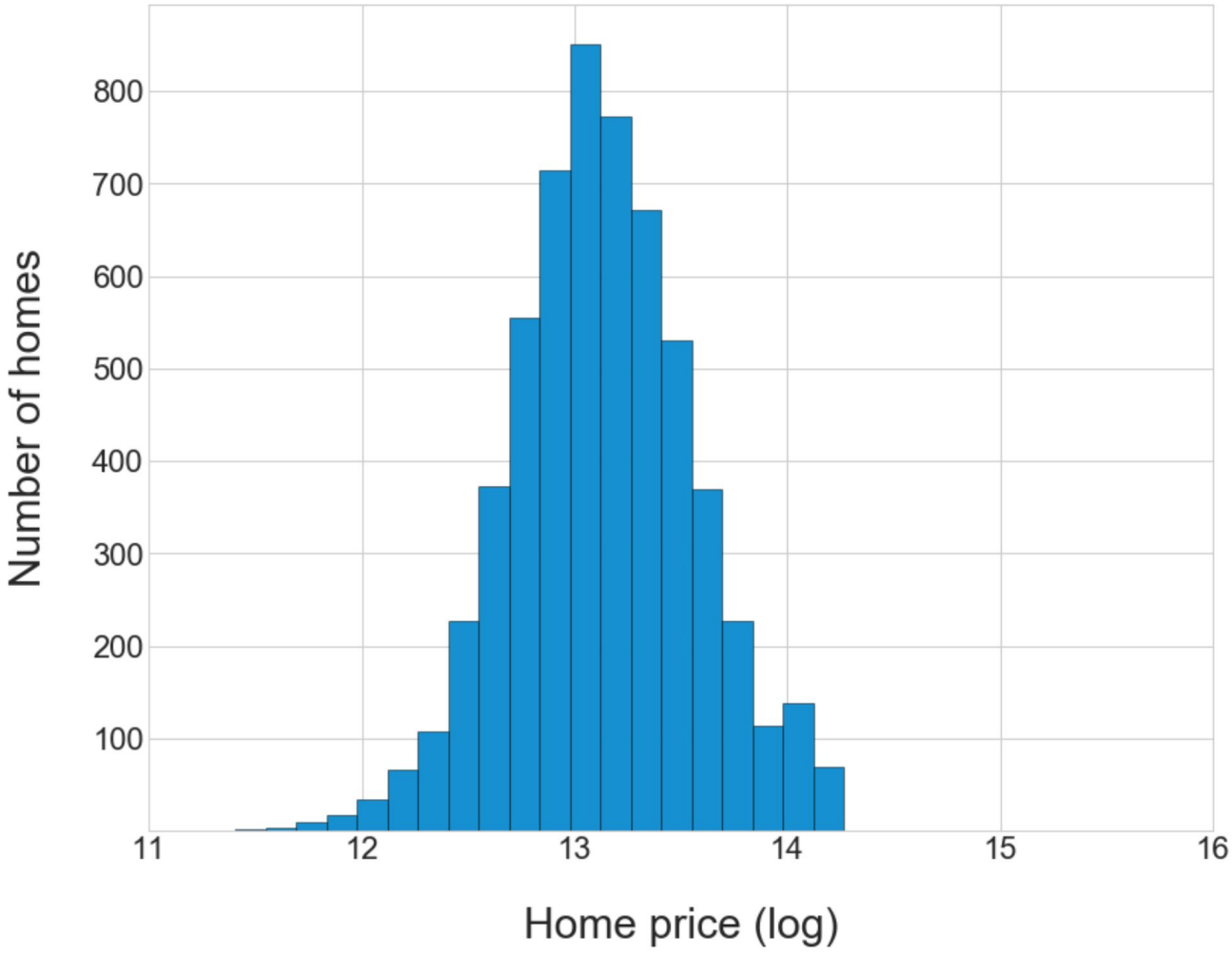


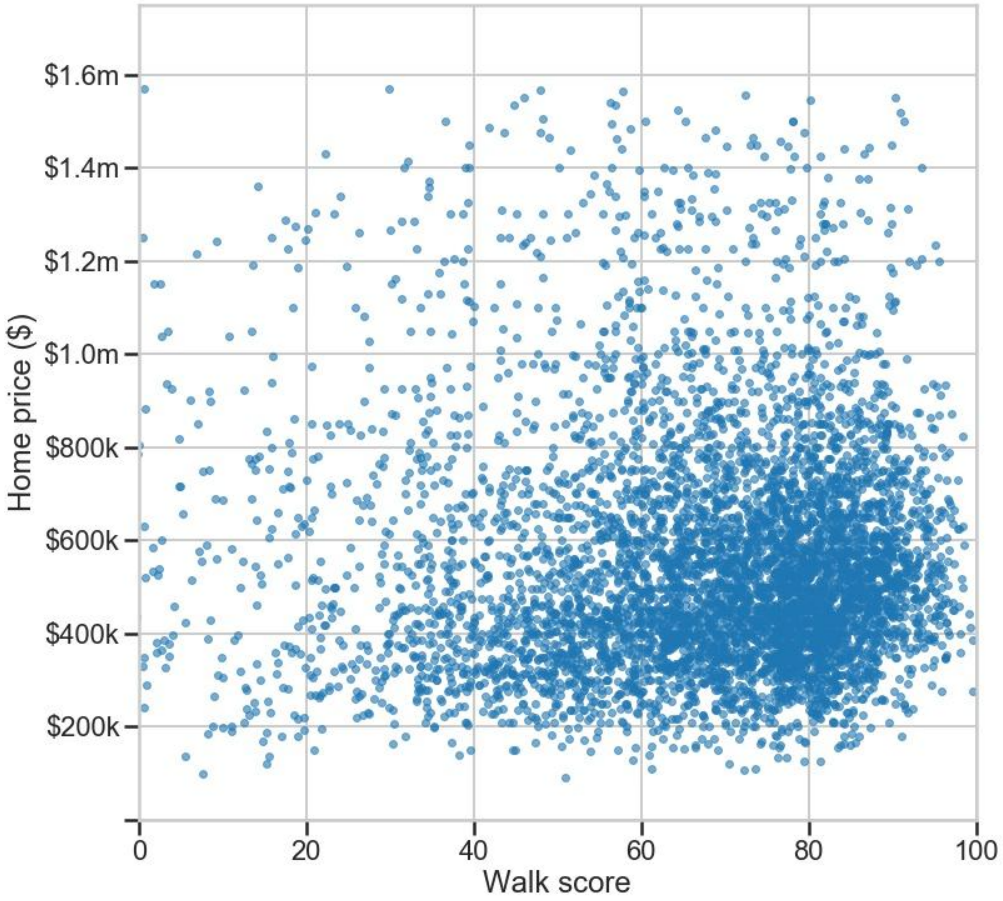

Residuals
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
Predicted price (log)


$1.6m
$1.4m
$1.2m
$1.0m
$800k
$600k
$400k
$200k
Home price ($)
A
B
C
D
Personal Crime Score


Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
0
200000
400000
600000
800000
1000000
1200000
1400000
Home price


Number of homes
600
500
400
300
200
100
0
$200k
$400k
$600k
$800k
$1.0m
$1.2m
$1.4m
Home price


Number of homes
800
700
600
500
400
300
200
100
11
12
13
14
15
16
Home price (log)


$1.6m
$1.4m
$1.2m
$1.0m
$800k
$600k
$400k
$200k
Home price ($)
0
20
40
60
80
100
Walk score