The Data Incubator Project Pitch

Predicting Housing Price Growth Using County-Level Data

or: where should I buy a house?

Sam Dillavou

Many, Many Factors in Housing Prices

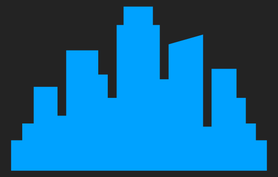
Individual

- Size/Style
- Bed/Bathrooms
- Location
- Year Constructed
- Upkeep
- Quirks

Properties Mostly Static
Heavily Studied
High Variance Locally

Which house do I buy?

+



Neighborhood/County

- Schools (DoEd.)
- Unemployment (BLS)
- Crime Rates (DoJ)
- Demographics
- Local Businesses (USDA)
- Location

Slowly Varying
Fewer Forecasts
Data Available

Where should I buy a house?

+

Global/Market

- Supply/Demand
- Mortgage Rates
- Rent Prices
- Market Fluctuations
- Pandemics

Volatile in Time
Heavily Studied
Immensely Complex

When do I buy?

First Step: School and Housing Data

Dept of Education

Public school test assessments in math and reading by grade, demographic, and district since 2009.
~200 MB

Currently: dividing data by district and year

Ridge Regression with SciPy

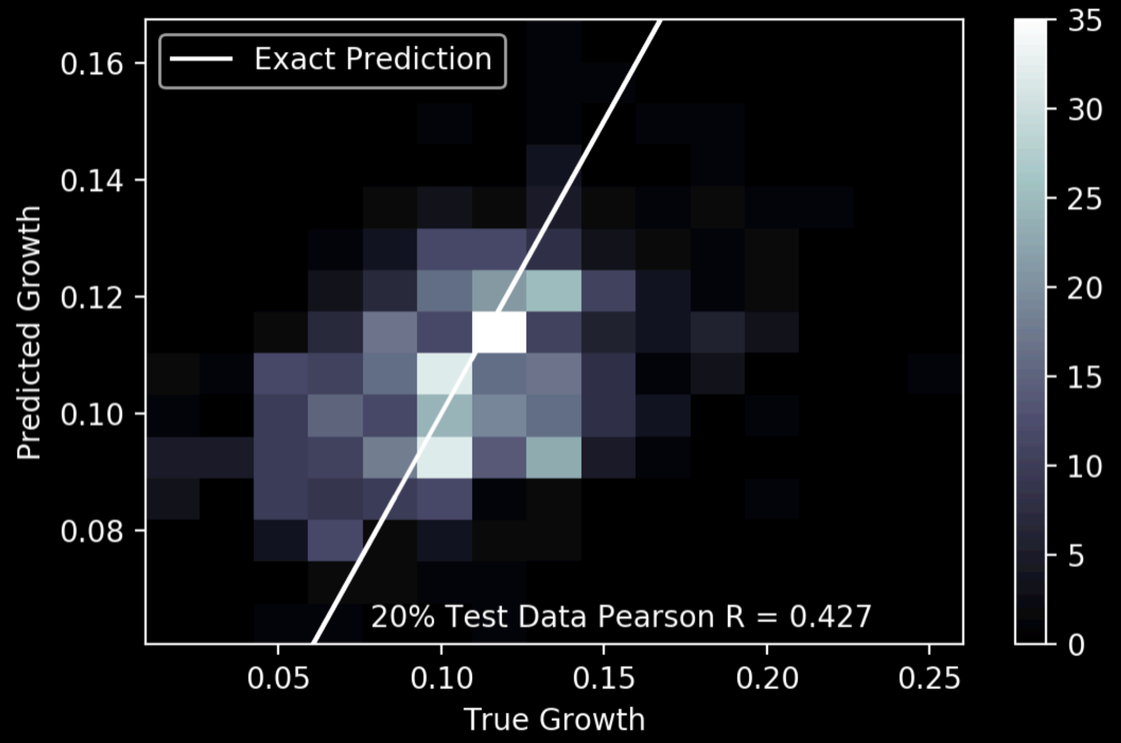
Zillow

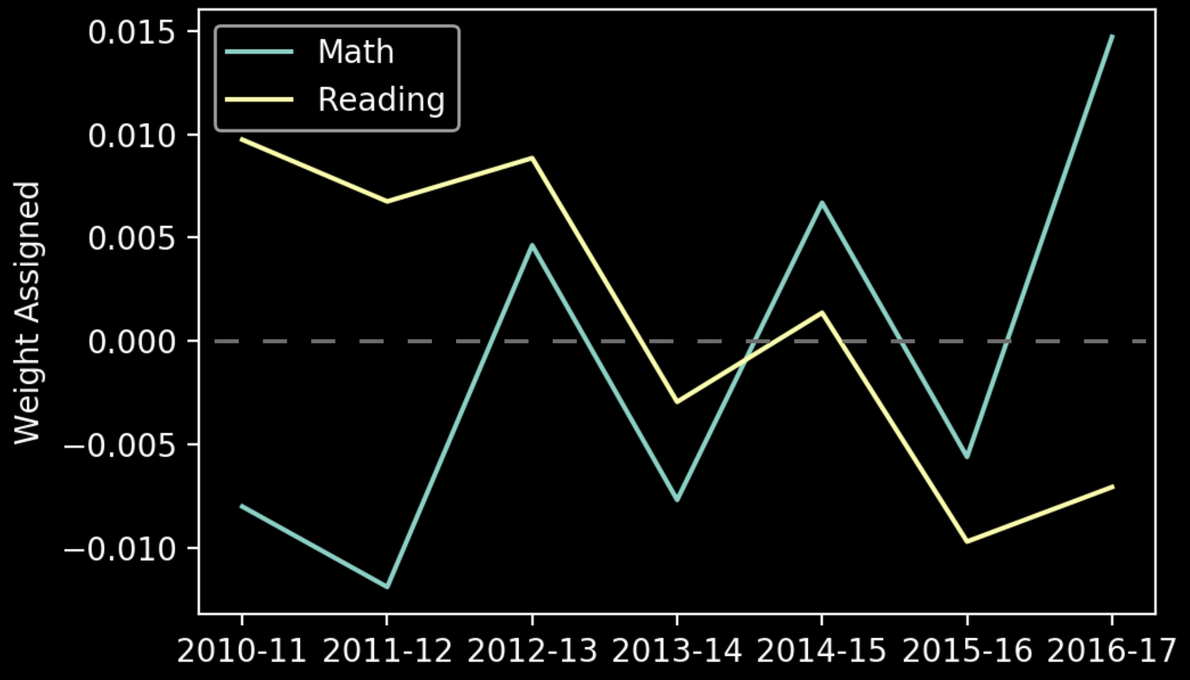
Seasonally adjusted house price index by county and month since 1996.
~6MB

Currently: dividing data by county and year

Math Up, Reading Down → Housing Price Rise!?

Predicting 2017-19 Housing Price Changes Using Ridge Regression on Public School Test Data 2010-16





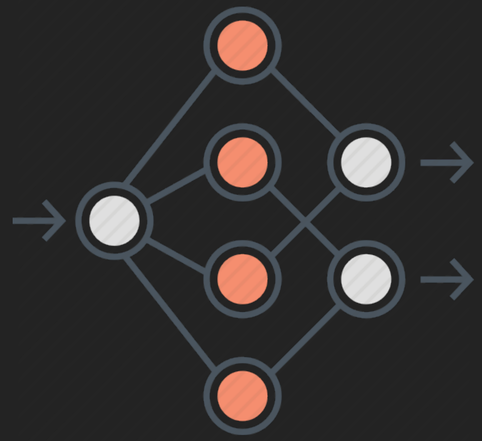
Past school test data correlates with future housing price changes: current algorithm adds 1.9% (2 year) growth over median of 5 options.

Surprisingly, the model consistently correlates *falling* reading performance with rising prices.

Next Steps

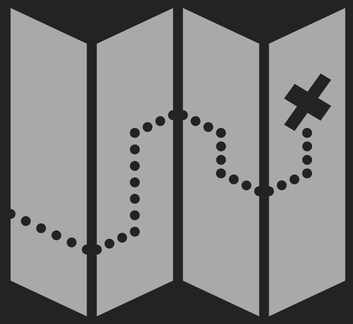
Add unemployment (BLS) and crime (DoJ) data. Both publicly available by county for decades.

https://www.bls.gov/lau/
https://www.ucrdatatool.gov/Search/Crime/Crime.cfm



Increase model capability by using neural network.

Separate training and test data in time.



Visualization tool for county selection and data source overlay.