The Value of a View

Predicting home prices based on natural amenities

My Questions

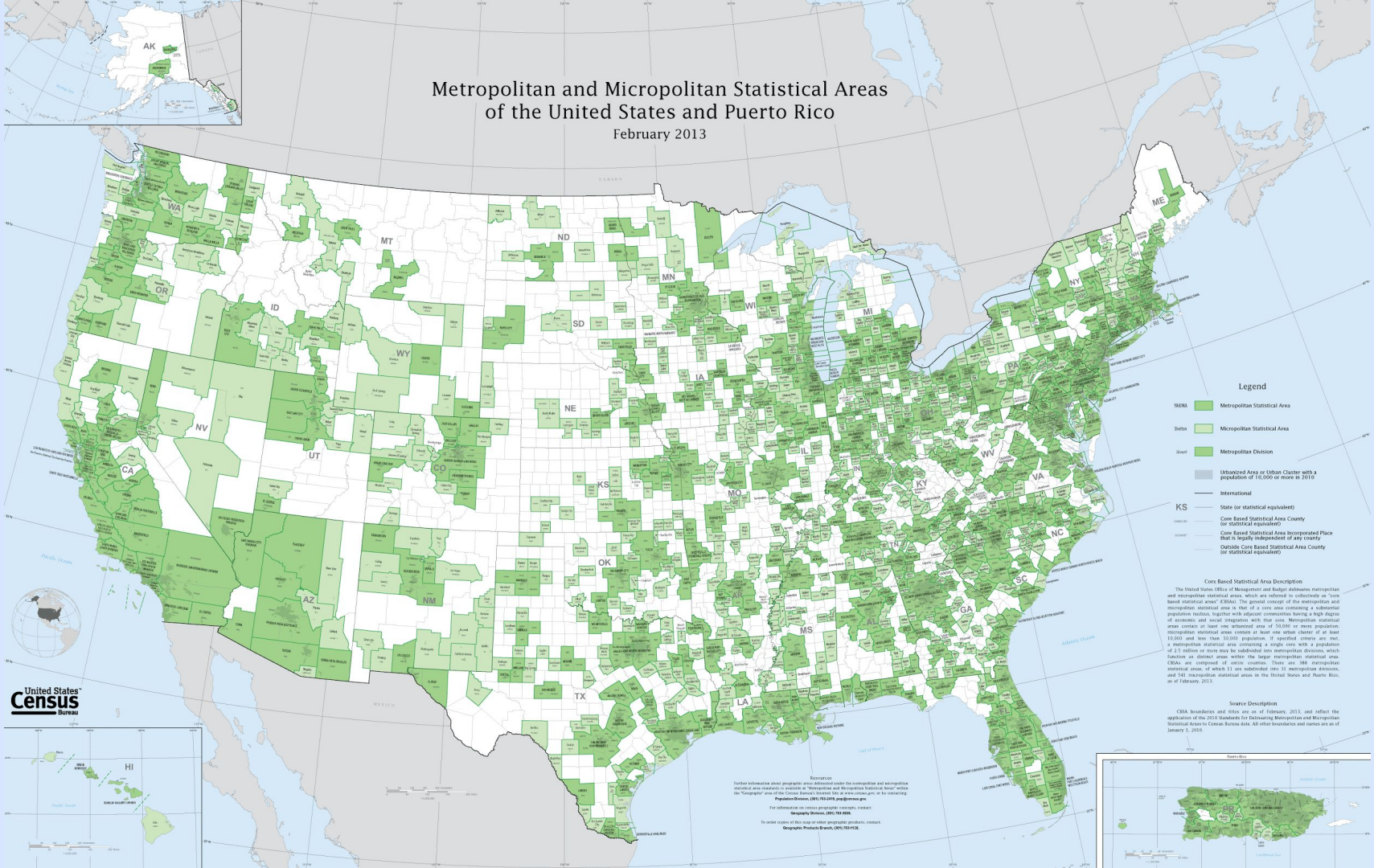
- Does natural beauty have predictive power?
- Are natural features more or less predictive than human-created features?

Data Sources

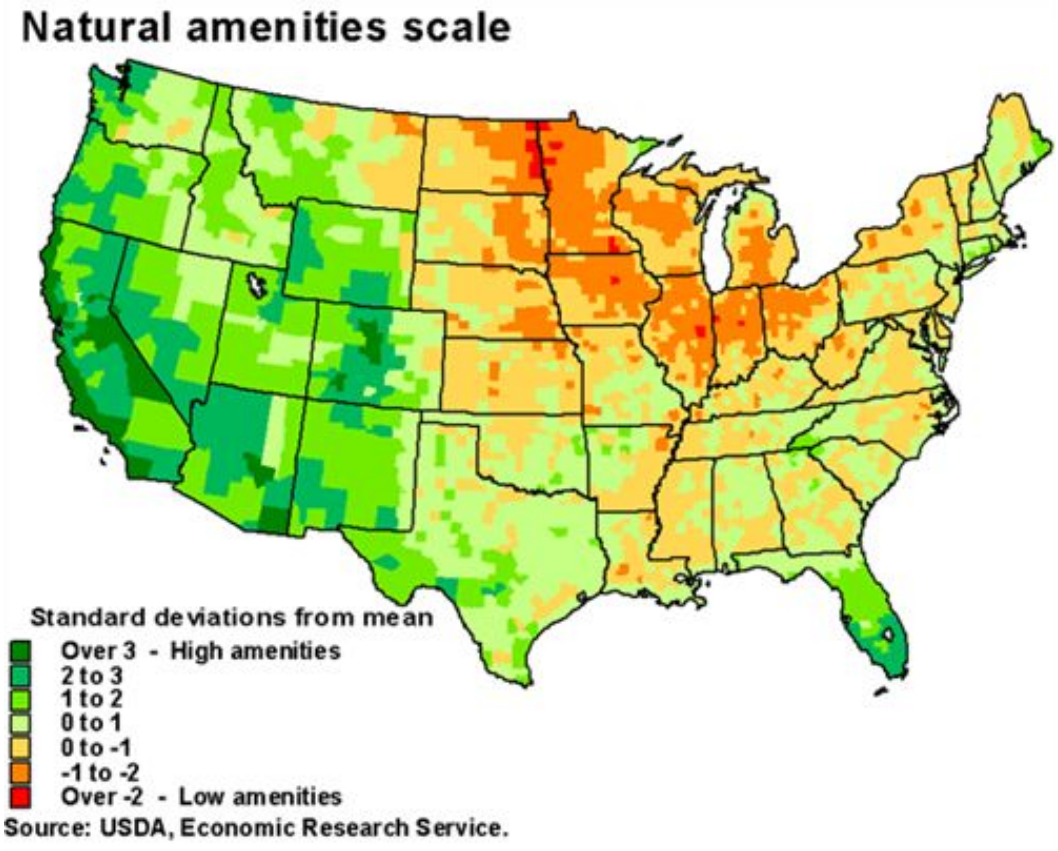
- USDA's Natural Amenity Scale (county-level)
- USDA's Rural-Urban-Continuum Codes (county-level)
- Home prices from National Association of Realtors (MSA-level)
- Population density from Wikipedia (county-level)
- MSA-to-county transition info from National Bureau of Economic Research

Metropolitan and Micropolitan Statistical Areas of the United States and Puerto Rico

February 2013

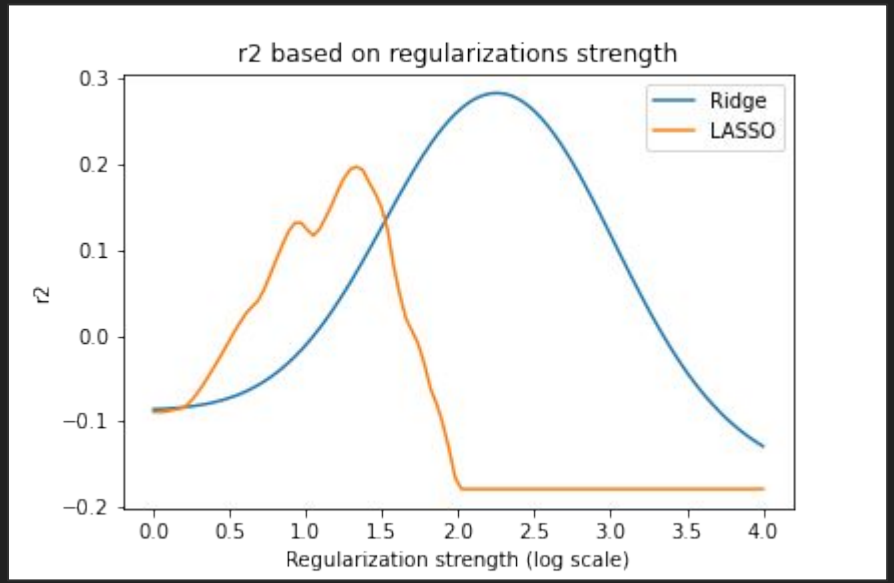


Natural Amenities Scale Breakdown



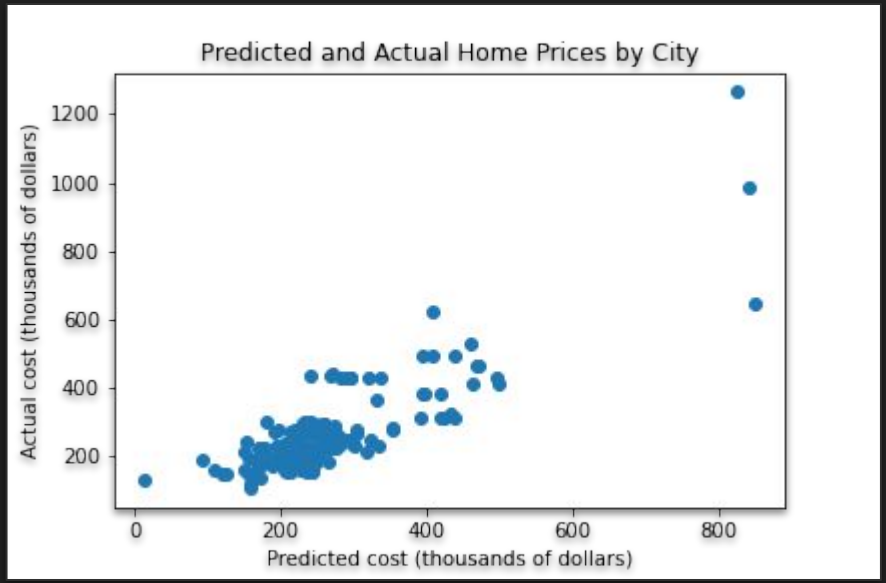
Applying LASSO and Ridge Regression

- Cross validated
- Various regularization strengths
- Pretty bad results
  - LASSO explained 20% of variation
  - Ridge explained 28%
  - Both models off by over $70,000 on average



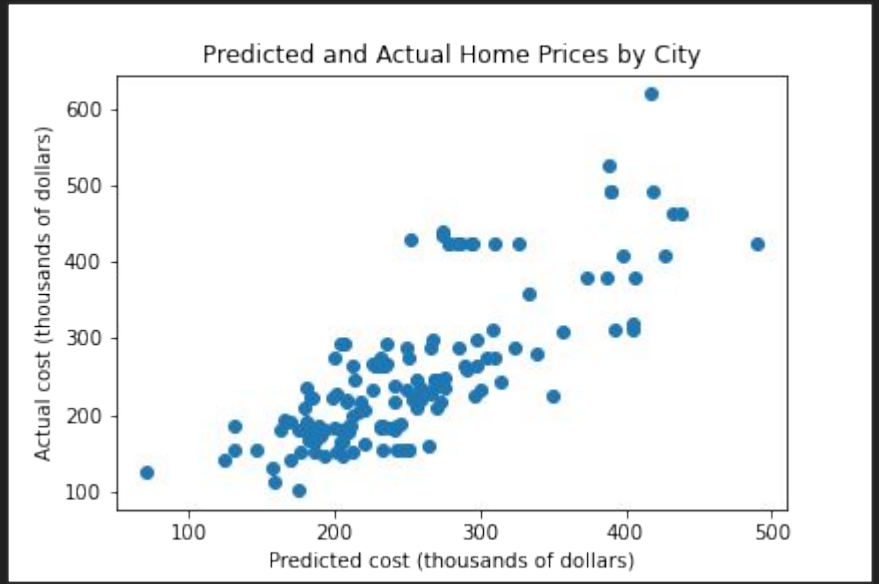
Model Diagnostics

- Three clear outliers



Model Optimization

- Removed CA outliers
- Added squared natural amenity feature
- Plotted predicted Ridge regression values against the actual values in the validation set



Comparing Natural and Human-Created Features

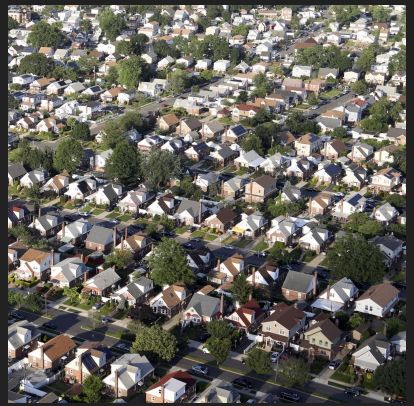
- Which factors are most important for predicting home prices?

Natural Features vs Human Features

- Neither the “natural-only” or “human-only” regression was very predictive of home prices.
- Between the two, the “natural-only” regression was slightly better, explaining 49% of the variation.

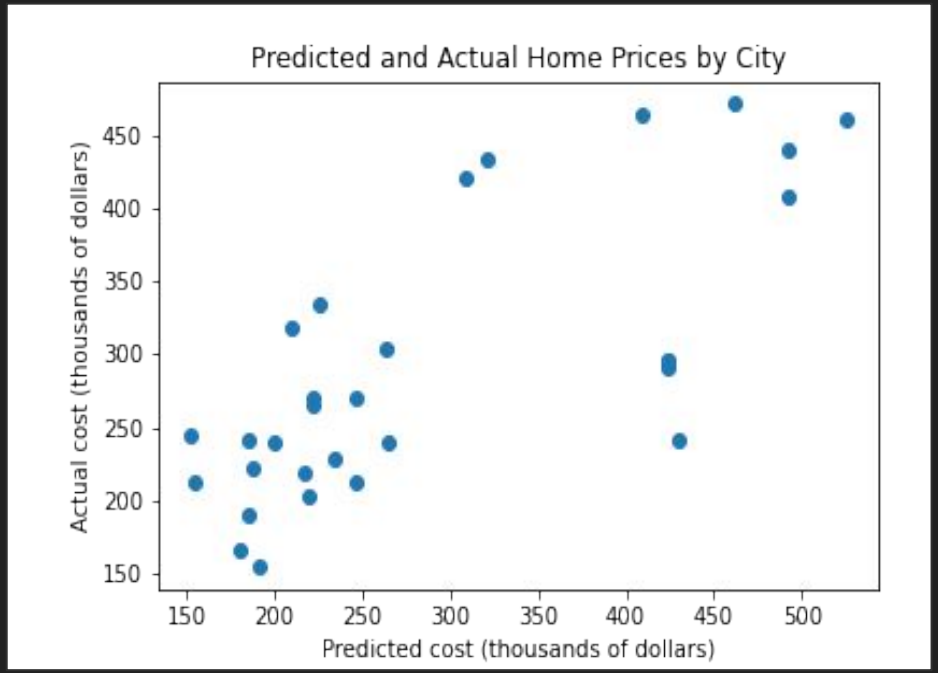
Combined Features

- This was the best model overall
- Ridge had best validation score
  - The model explained 60% of variation
  - Model was off by an average of $47,350



Testing the Model

- Final model
  - Ridge
  - Combination of natural and human features
  - Outliers excluded
  - Exponential term
- Explained 52% of variation
- Model was off by $72,660 on average



Applications for this Model

- Housing developers
  - Natural features + conventional metrics can predict price point

Possible Improvements and Further Research

- Increase size of dataset (only 182 cities)
- Application of transformations to reduce variance of accuracy
- What’s up with California?
- Incorporate spatial features within cities
  - Distance to green space
  - Distance to water

Conclusion

- Model has issues generalizing to new data.
- Natural features *are* valuable in predicting median home prices per city.
- The most predictive model is a combination of natural and human-created features.
- In a more developed form, this sort of model could augment traditional methods of modeling housing prices.

Sources

https://www.brookings.edu/research/whos-to-blame-for-high-housing-costs-its-more-complicated-than-you-think/

https://www.countryliving.com/life/entertainment/g18924890/nature-quotes/

https://nymag.com/intelligencer/2020/07/trumps-racist-defense-of-nimbyism-may-aid-housing-justice.html

https://i.pinimg.com/originals/c8/21/5b/c8215bc0865fa387251c3eeab574b8db.jpg