Using Machine Learning to Generate New, Valuable Zoning Data

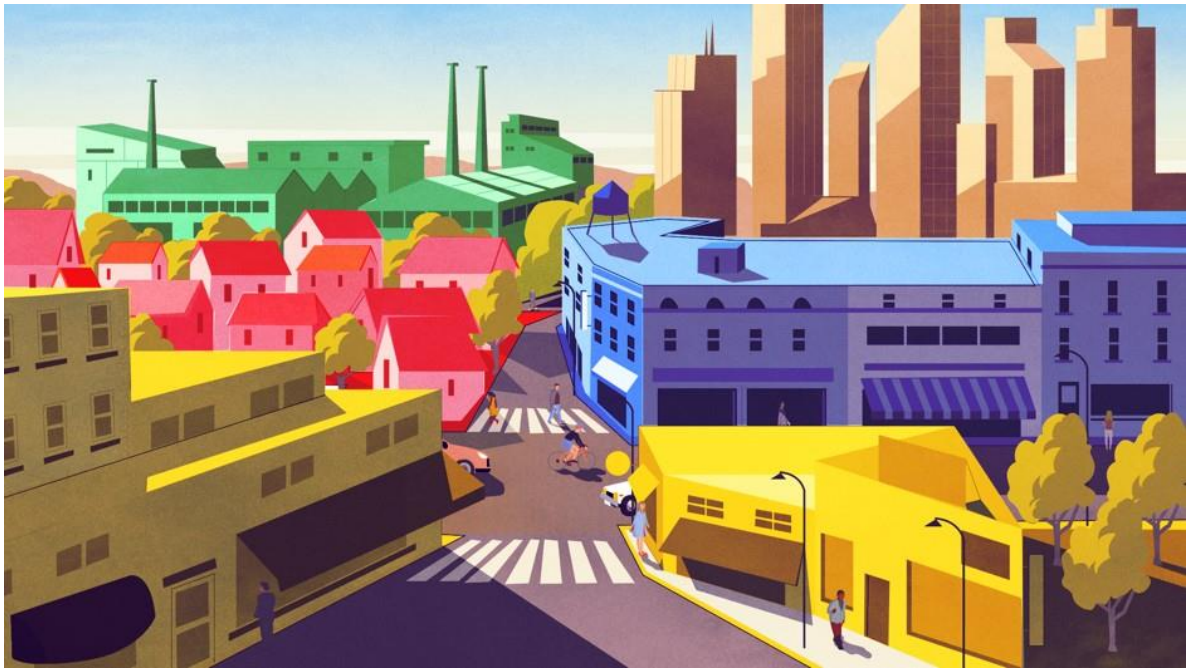
Graham MacDonald, Chief Data Scientist

URBAN INSTITUTE

What is zoning?

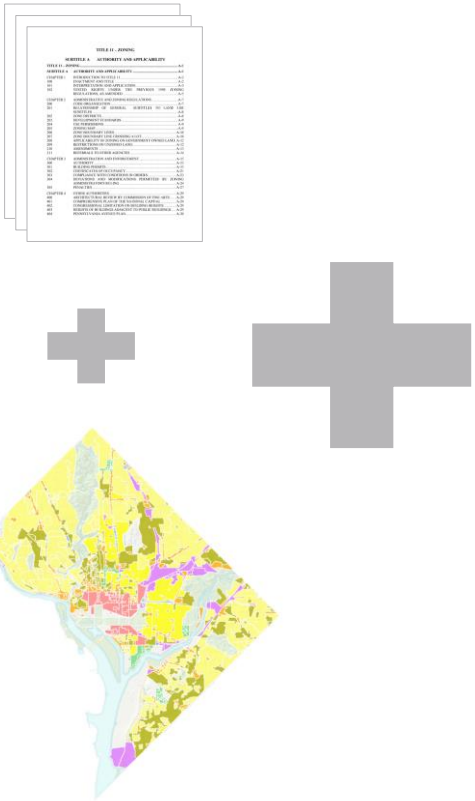
What is zoning?

"From the start, zoning has separated more than just land uses. It also separates people."



There are no (comprehensive, comparable, and current) zoning data?

There are no (comprehensive, comparable, and current) zoning data?



306 REAR YARD

306.1 A minimum rear yard of twenty-five feet (25 ft.) shall be provided in the R-1-A and R-1-B zones.

306.2 A minimum rear yard of twenty feet (20 ft.) shall be provided in the R-2 and R-3 zones.

306.3 Notwithstanding Subtitle D §§ 306.1 and 306.2, a rear wall of an attached or semi-detached building shall not be constructed to extend farther than ten feet (10 ft.) beyond the farthest rear wall of any adjoining principal residential building on an adjoining property.

306.4 A rear wall of an attached or semi-detached building may be constructed to extend farther than ten feet (10 ft.) beyond the farthest rear wall of any adjoining principal residential building on an adjoining property if approved as a special exception pursuant to Subtitle X, Chapter 9 and as evaluated against the criteria of Subtitle D §§ 5201.3(a) through 5201.3(d) and §§ 5201.4 through 5201.6.

SOURCE: Final Rulemaking published at 63 DCR 2447 (March 4, 2016 – Part 2); Final Rulemaking & Order No. 14-11B published at 64 DCR 4055 (April 28, 2017).

How can data science help?

How can data science help?

1. Use natural language processing to extract zoning rules directly from local zoning codes

2. Use machine learning to predict zoning rules based on property assessment data

Property assessment data?

Use machine learning to predict zoning rules based on property assessment data

Property assessment data?

Use machine learning to predict zoning rules based on property assessment data

- Lot size
- Year built/remodeled
- Land use description
- Geographic information

Zoning rules?

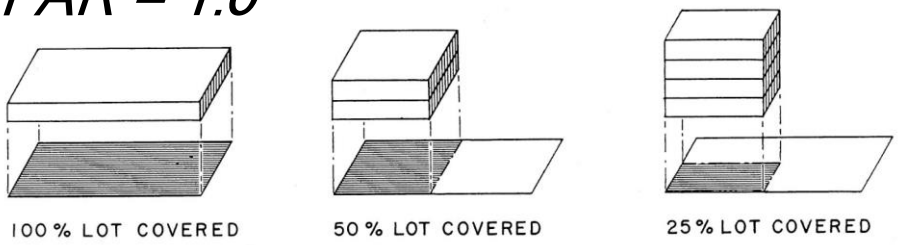
Use machine learning to predict zoning rules based on property assessment data

Zoning rules?

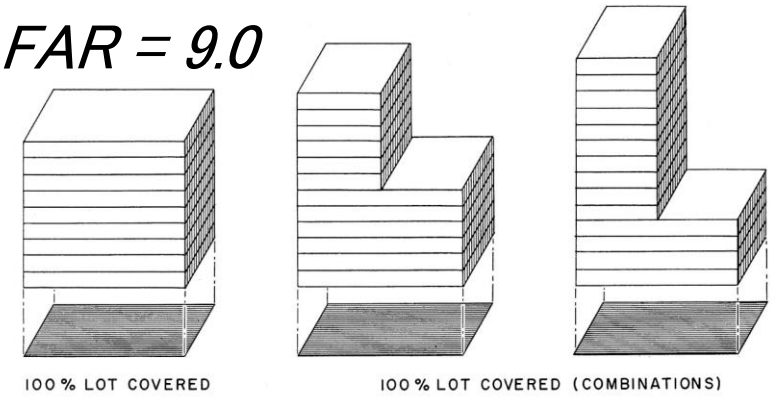
Use machine learning to predict zoning rules based on property assessment data

Maximum allowed by-right floor area ratio (FAR)

FAR = 1.0



FAR = 9.0



Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Step 1: Transform property assessment data into meaningful features

Step 2: Build a predictive model

Step 3: Evaluate our model

Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Step 1: Transform property assessment data into meaningful features

Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Step 1: Transform property assessment data into meaningful features

Property-Level		Zone-Level
Lot size	→	Average lot size per home
Land use description	→	Share of low-density homes

Machine learning?

Use machine learning to predict zoning rules based on property assessment data

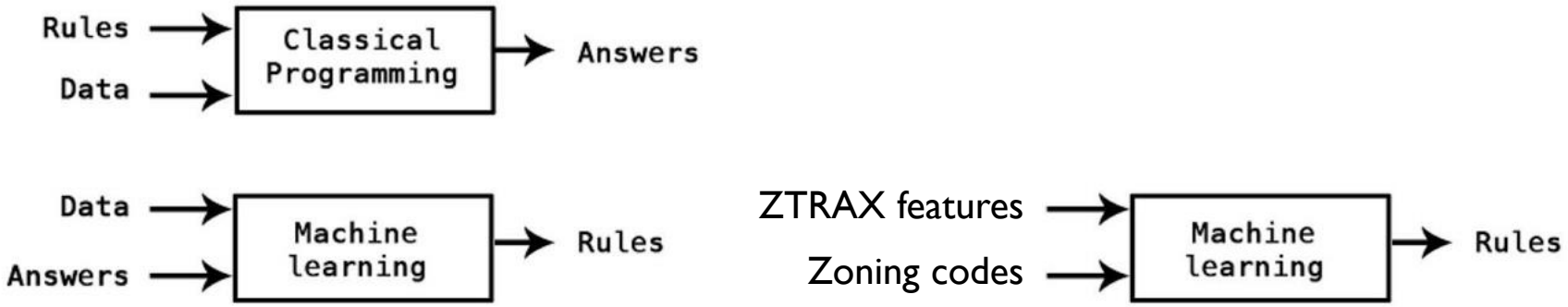
Step 1: Transform property assessment data into meaningful features

Step 2: Build a predictive model

Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Step 2: Build a predictive model



Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Step 1: Transform property assessment data into meaningful features

Step 2: Build a predictive model

Step 3: Evaluate our model

Machine learning?

Use machine learning to predict zoning rules based on property assessment data

Step 3: Evaluate our model

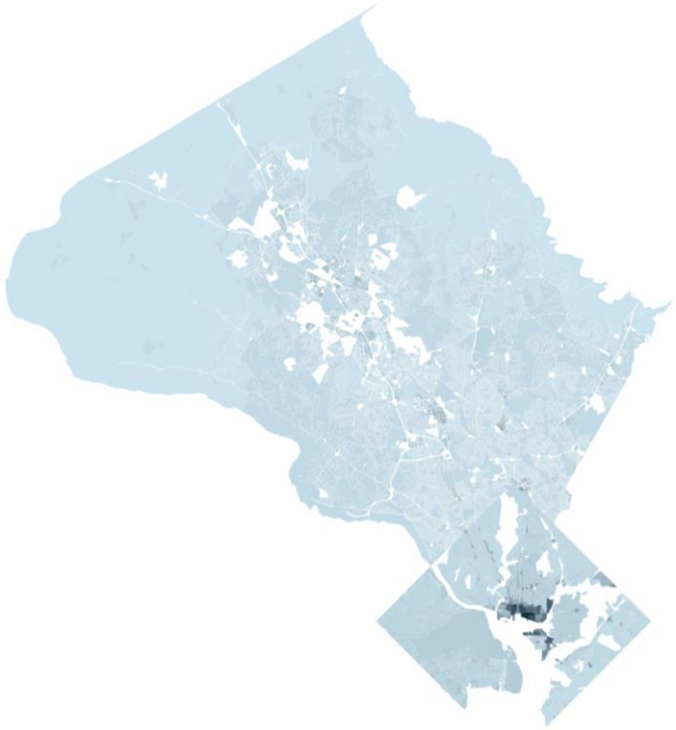
How close are our predicted FARs to the true FARs?

- In-sample? Out-of-sample?
- Weighted RMSE? Weighted Relative MAE?

How do we do?

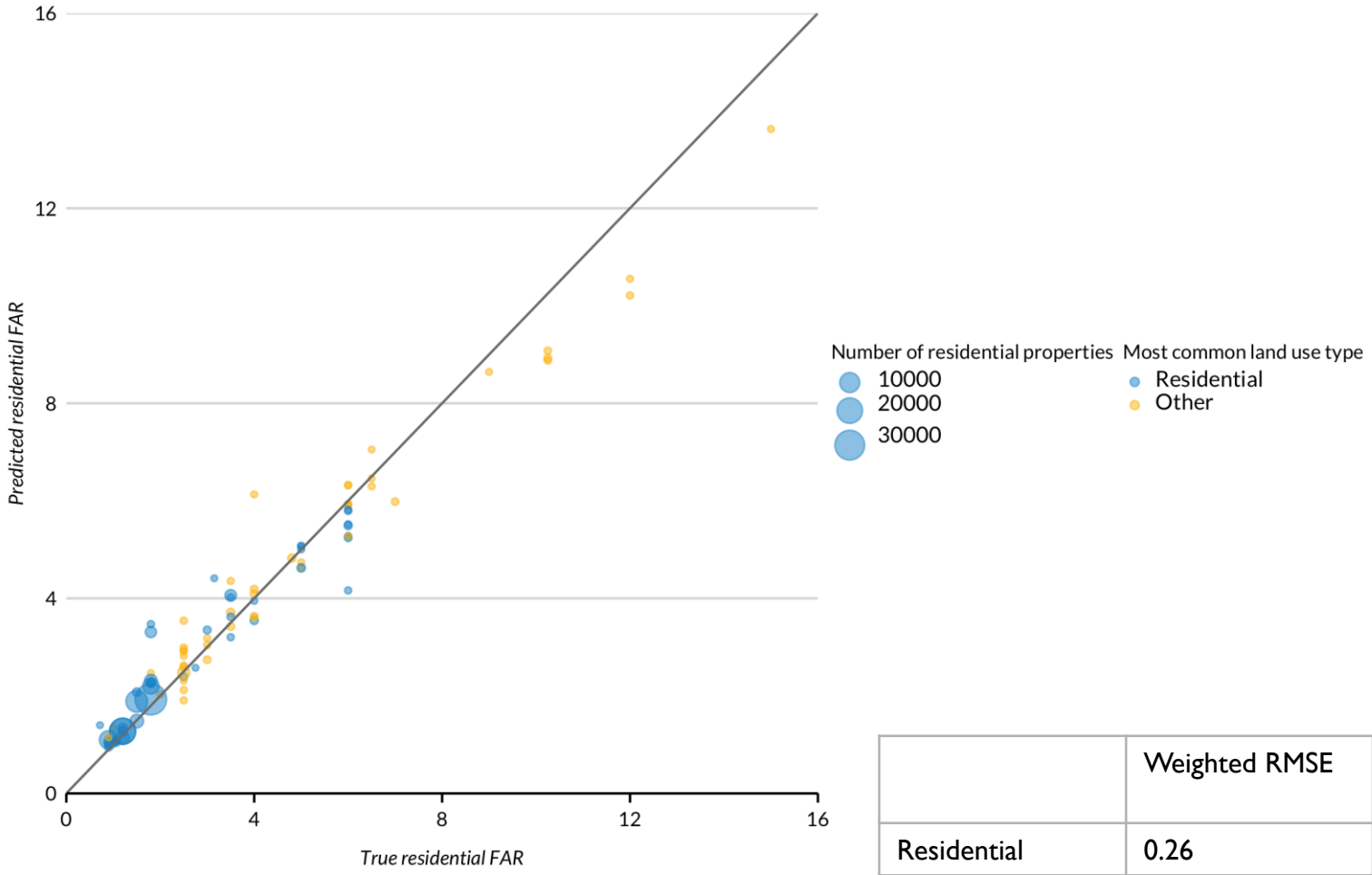
How do we do?

1. Predicting Washington, DC in-sample?
2. Predicting DC, Montgomery, and Arlington County in-sample?
3. Predicting Montgomery County out-of-sample?
4. Predicting Arlington County out-of-sample?



How do we do?

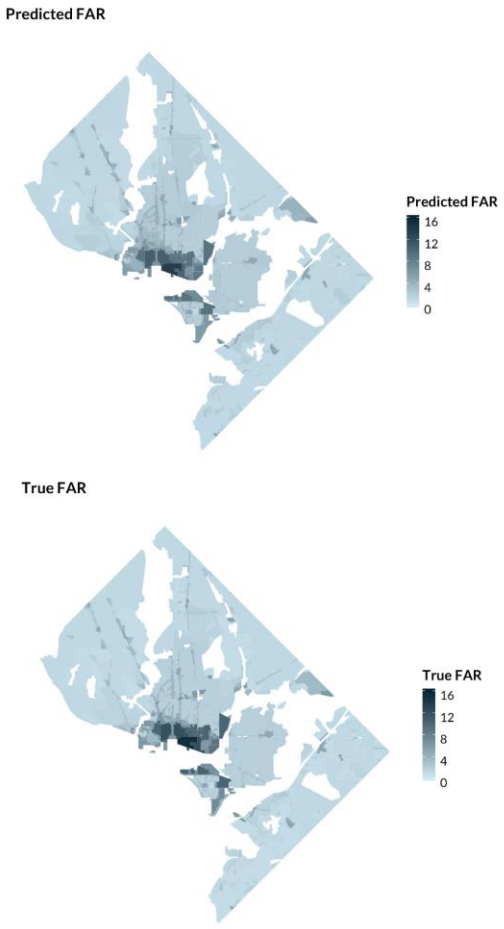
Predicting Washington, DC in-sample?

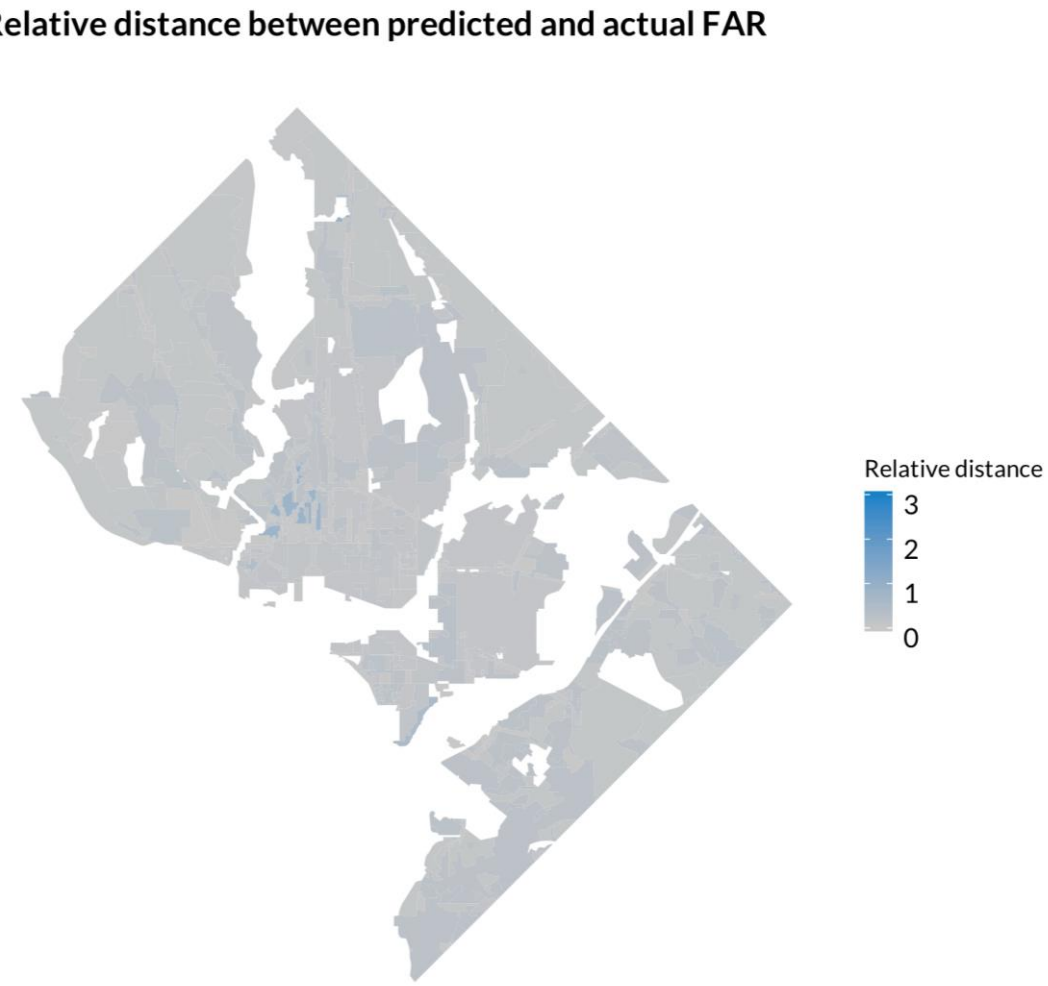


	Weighted RMSE	Weighted Relative MAE
Residential	0.26	0.12
Non-Residential	0.39	0.04

How do we do?

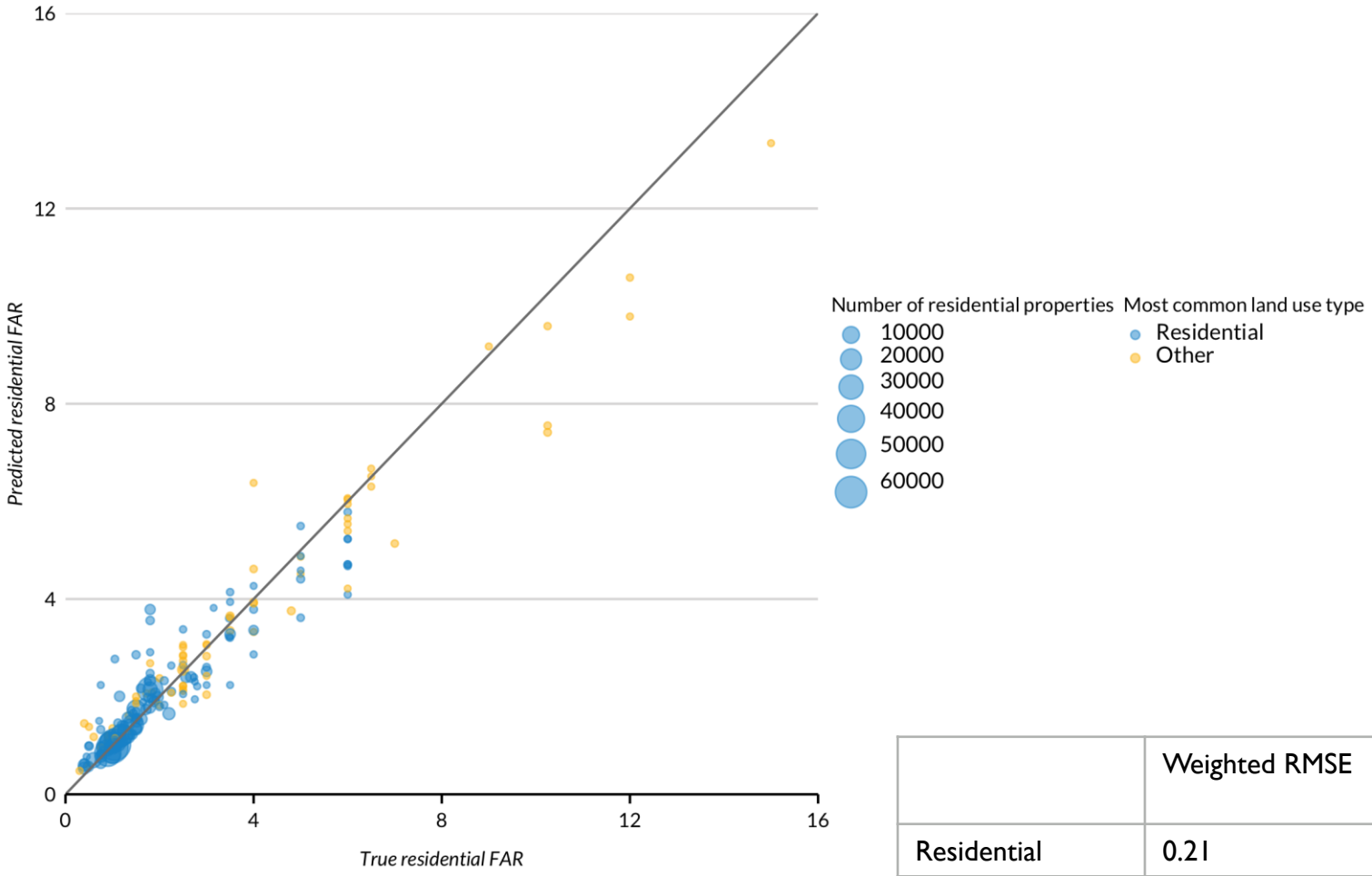
Predicting Washington, DC in-sample?





How do we do?

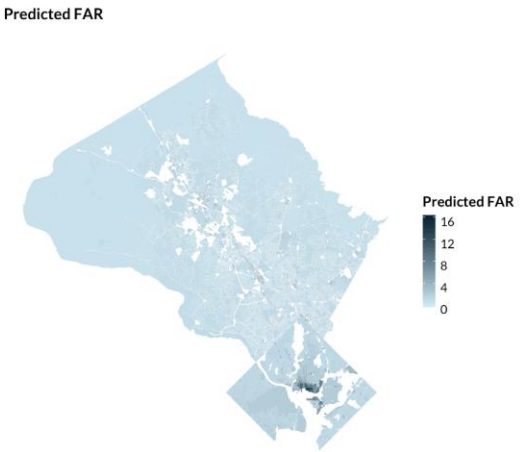
Predicting DC, Montgomery, and Arlington County in-sample?

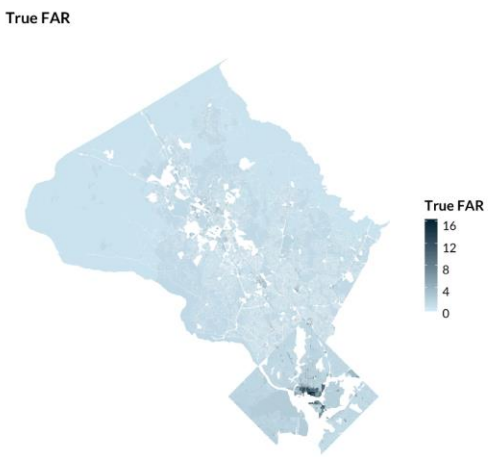


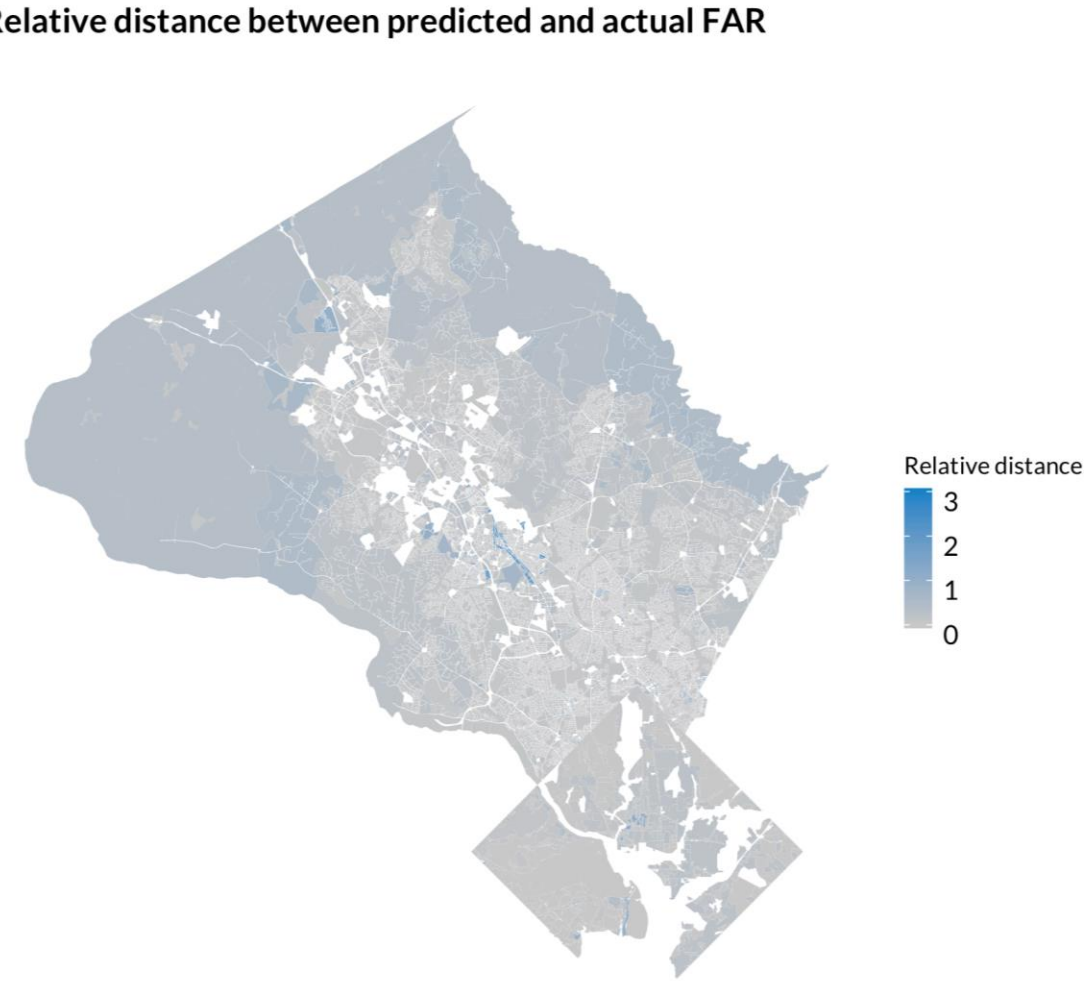
	Weighted RMSE	Weighted Relative MAE
Residential	0.21	0.09
Non-Residential	0.67	0.13

How do we do?

Predicting DC, Montgomery, and Arlington County in-sample?

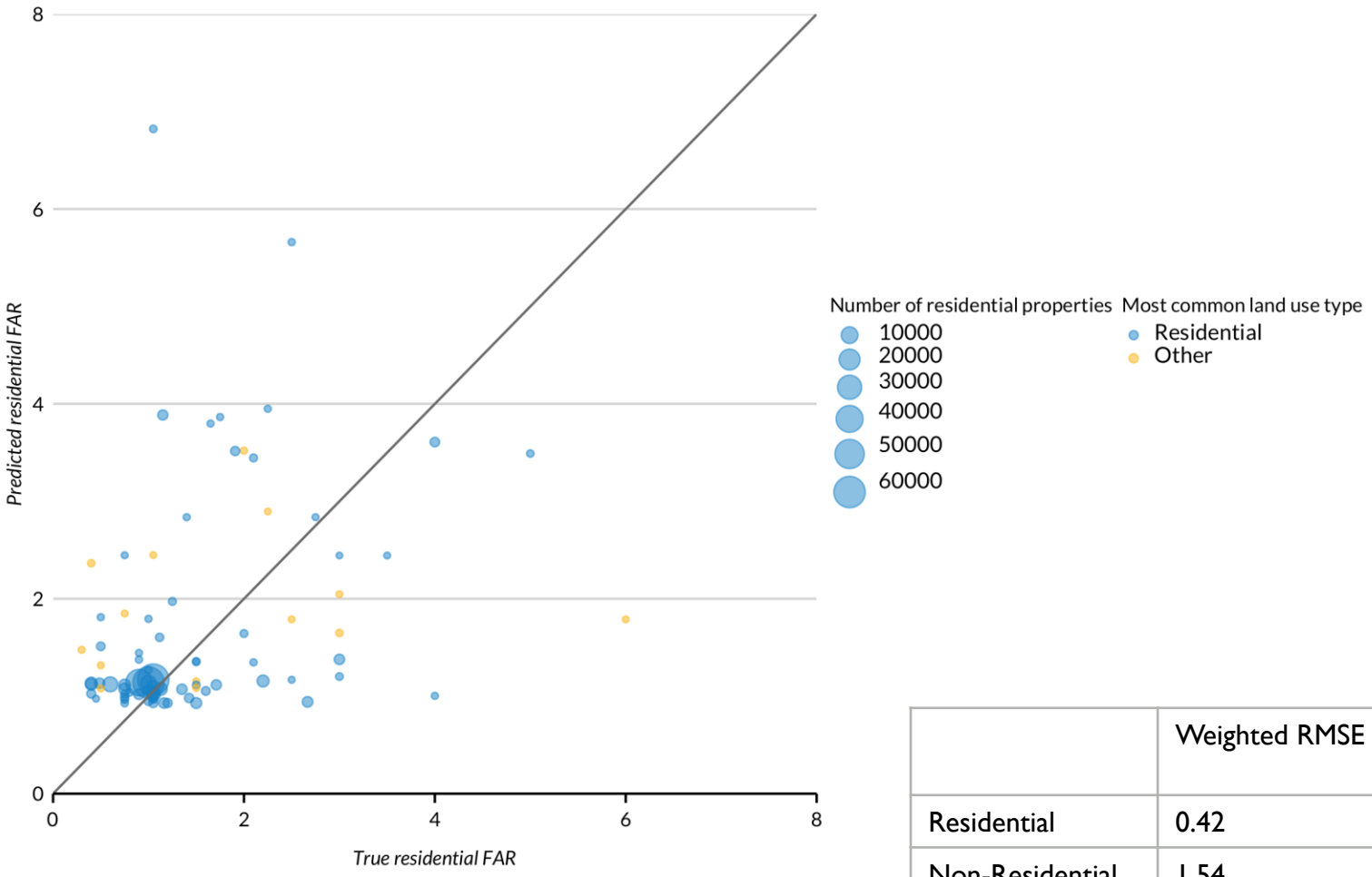






How do we do?

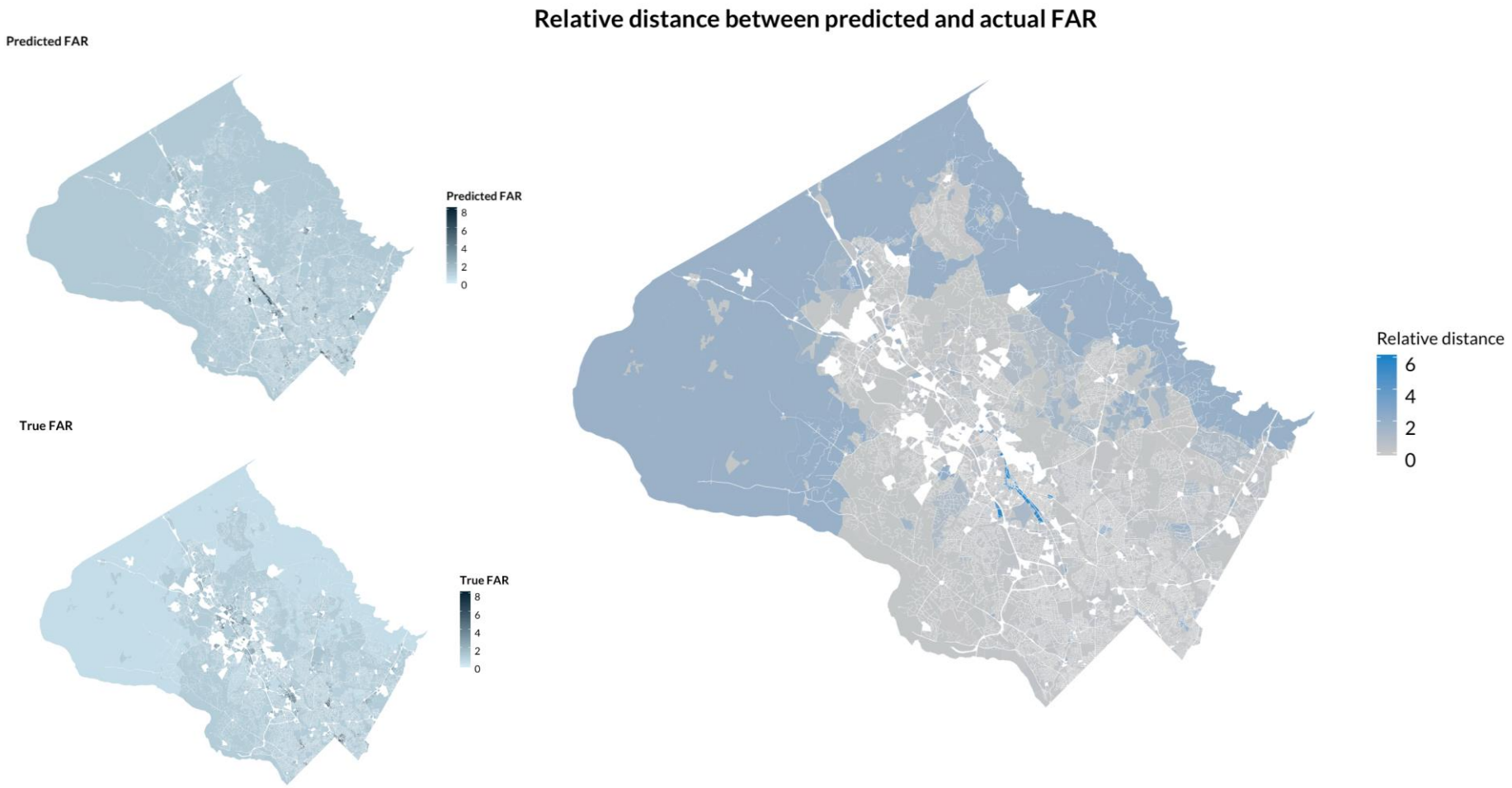
Predicting Montgomery County out-of-sample?
(Training using DC)



	Weighted RMSE	Weighted Relative MAE
Residential	0.42	0.26
Non-Residential	1.54	2.23

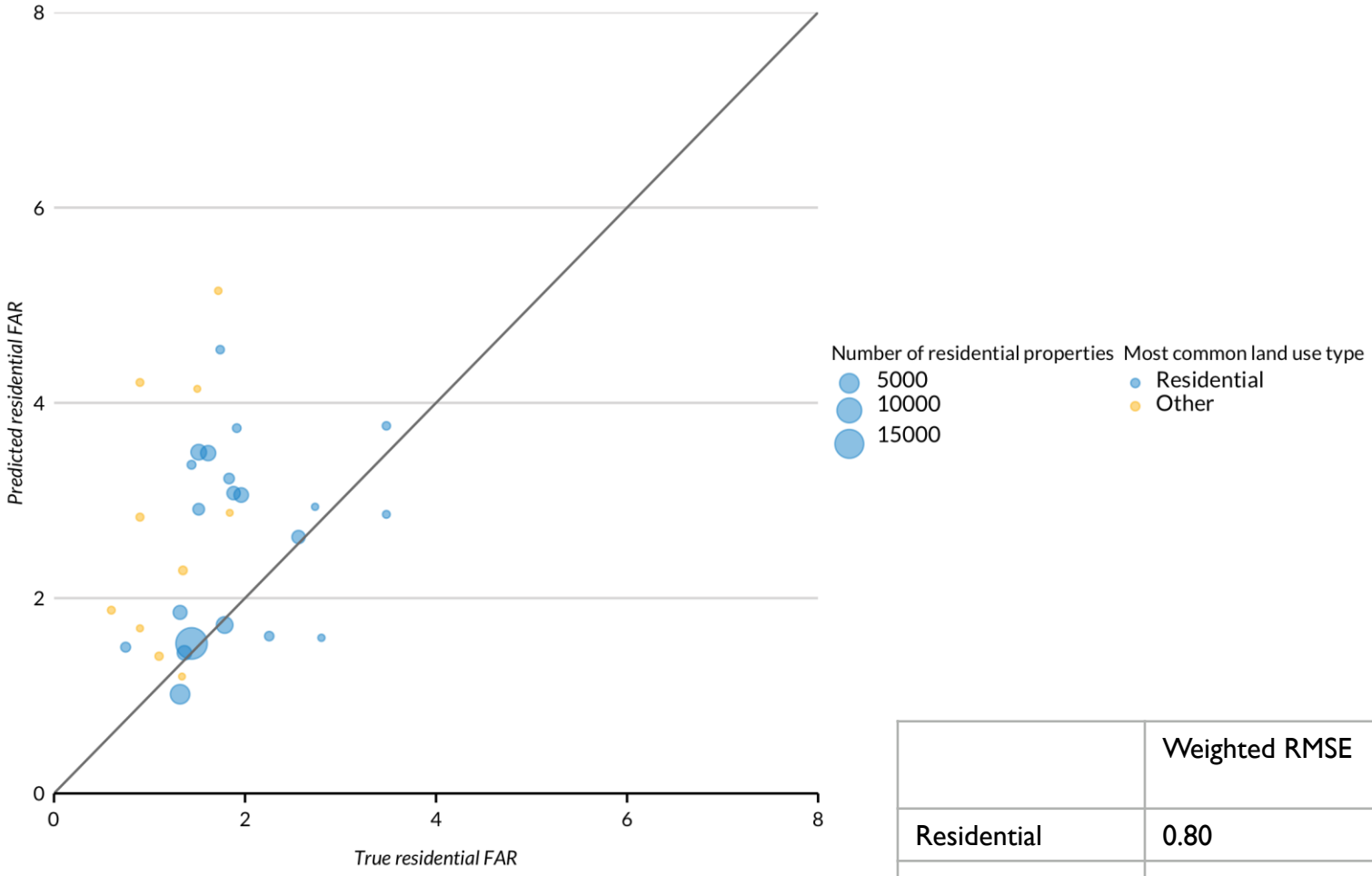
How do we do?

Predicting Montgomery County out-of-sample?
(Training using DC)



How do we do?

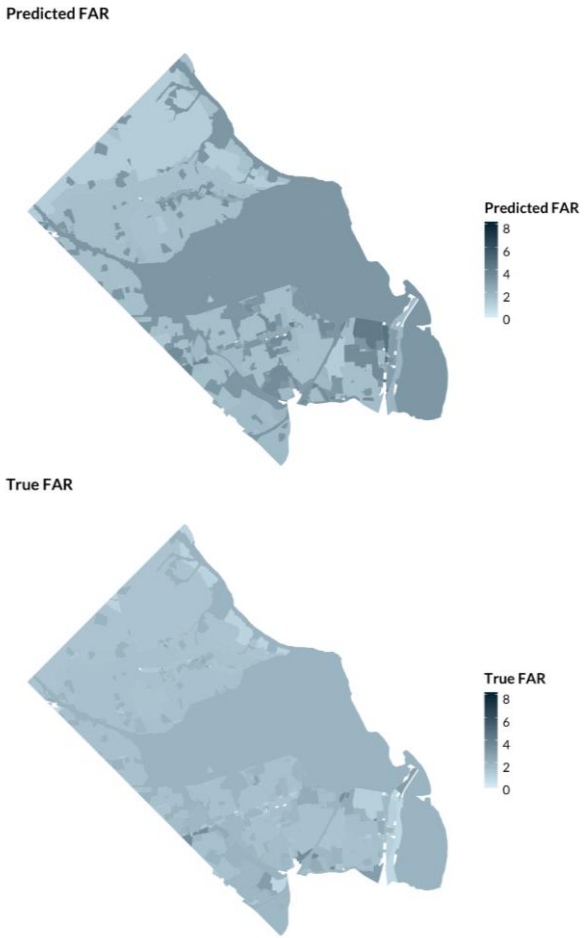
Predicting Arlington County out-of-sample?
(Training using DC & Montgomery County)

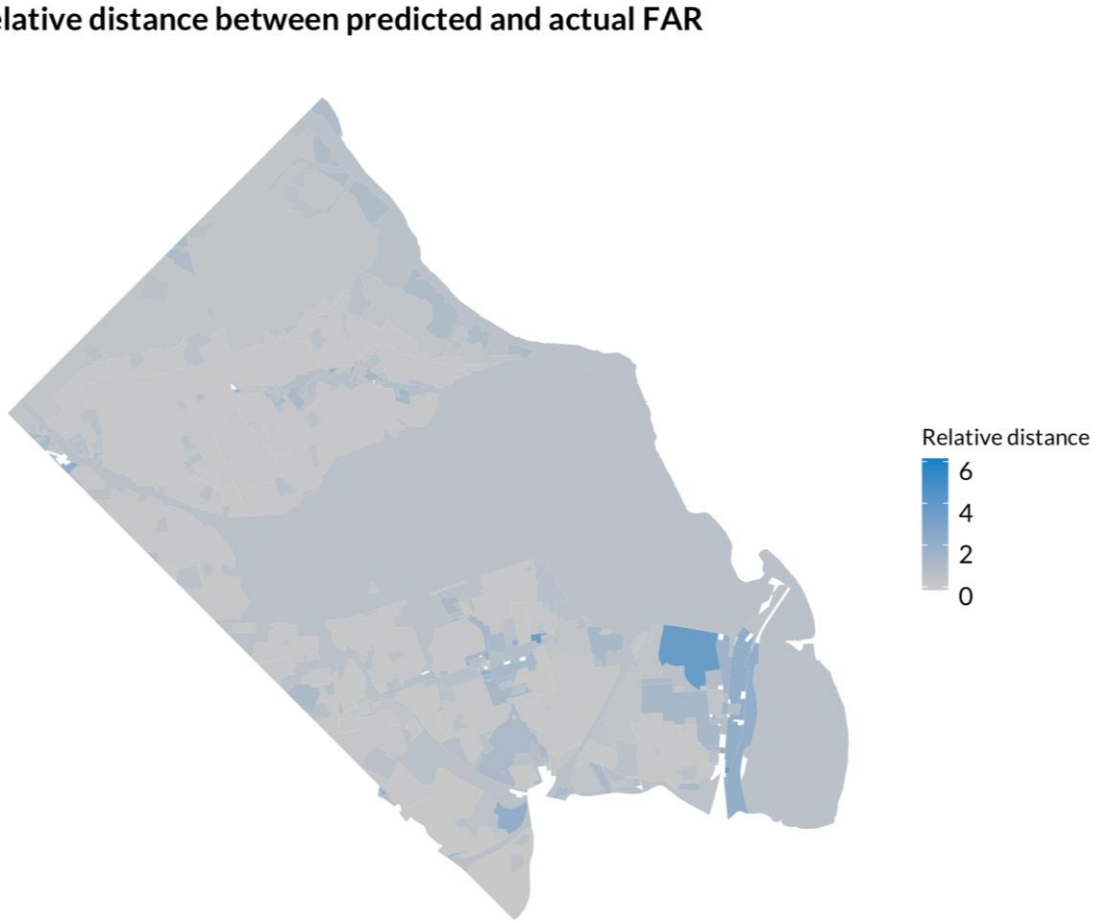


	Weighted RMSE	Weighted Relative MAE
Residential	0.80	0.31
Non-Residential	1.65	1.35

How do we do?

Predicting Arlington County out-of-sample?
(Training using DC & Montgomery County)





What comes next?

What comes next?

1. How can we improve the model?

   More features and more jurisdictions!

What comes next?

1. How can we improve the model?

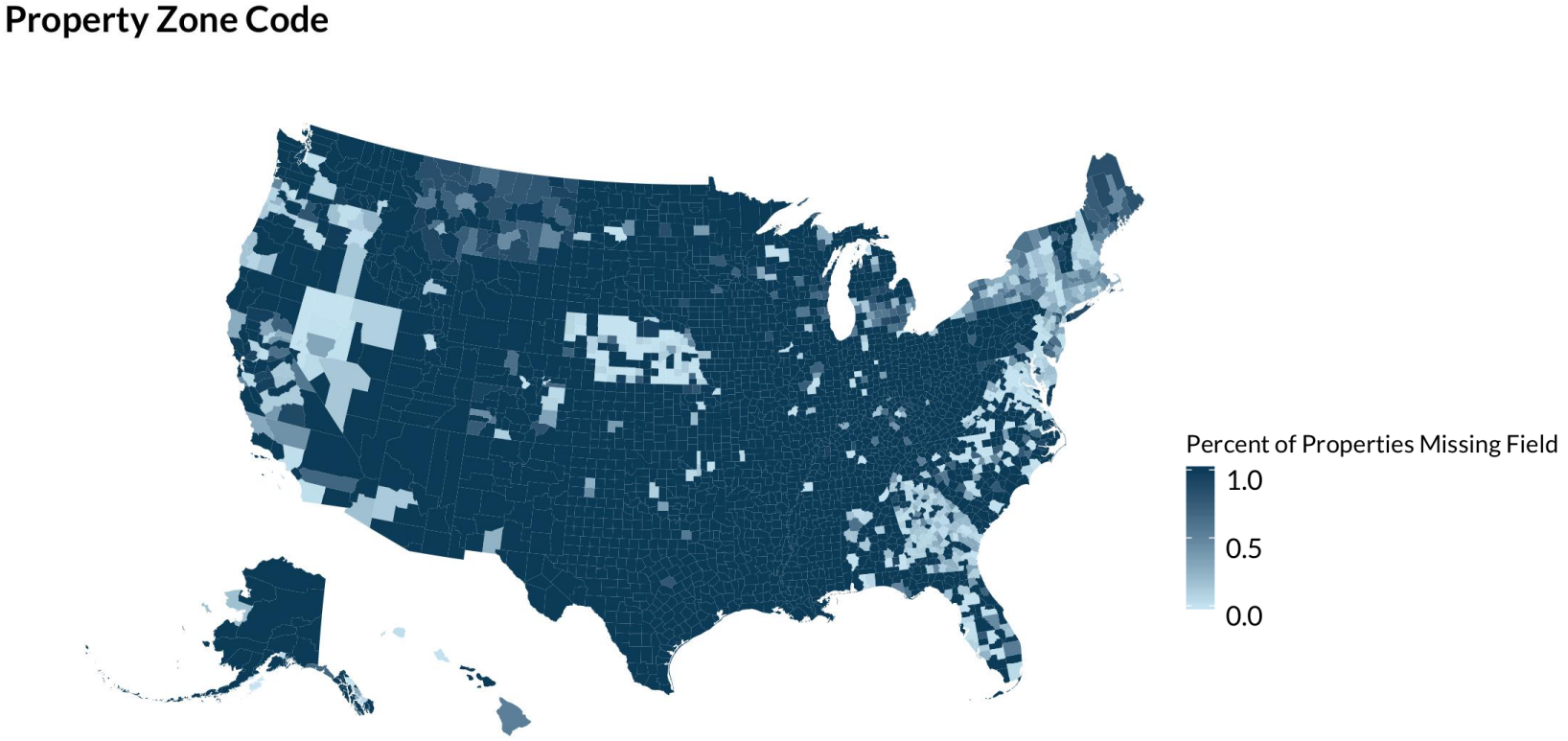
   More features and more jurisdictions!

2. Can this generalize?

   To other regions?
   To other built environments?
   To other zoning characteristics?

Can this generalize to other regions?

In some cases – yes!
In other cases – not with ZTRAX alone



Can this generalize?

Not with ZTRAX alone – we need open data!

US CITY OPEN DATA CENSUS | About | Changes | FAQ | Support | SUNLIGHT FOUNDATION Logo1

Datasets / Zoning (GIS)

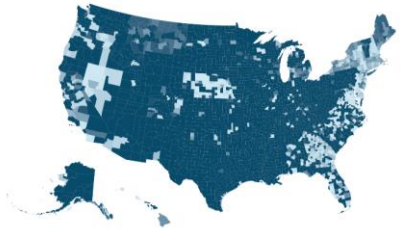
On this page you can see the state of open data for Zoning (GIS) in all the places for which we have information.

Dataset Description

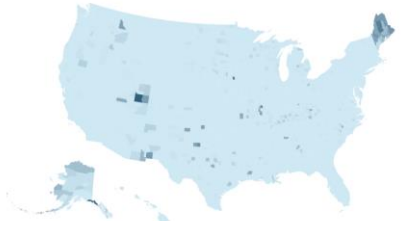
The mapped zone (GIS) shapefiles of designated permitted land use in your city. (More info)

Place	Score	Breakdown	Year	Location (URL)	Information	
Anchorage, AK	100%		2017	http://munimaps.muni....		Propose Revisions
Ann Arbor, MI	100%		2014	http://data.a2gov.org/f...		Propose Revisions
Asheville, NC	100%		2014	http://opendatacatalog...		Propose Revisions
Austin, TX	100%		2017	https://data.austintexa...		Propose Revisions
Baton Rouge, LA	100%		2016	http://gis.brla.gov/		Propose Revisions
Boulder, CO	100%		2014	https://bouldercolorad...		Propose Revisions

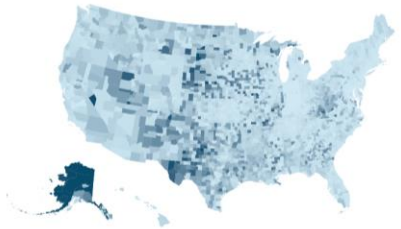
Property Zone Code



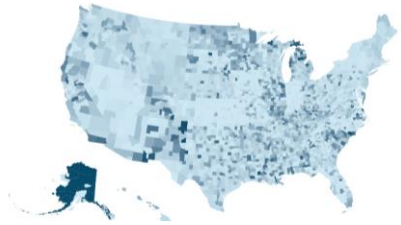
Property Land Use Code



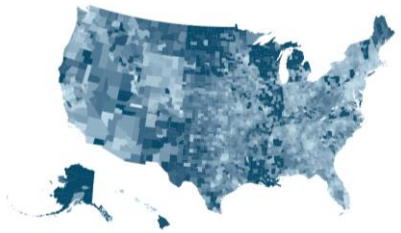
Property Latitude/Longitude



Property Lot Size



Property Year Built



"What are we zoned for and what have we built – what is the delta? It's nearly impossible to know."

– Ruby Bolaria, Chan Zuckerberg Initiative

We can use data science to unlock zoning data.

Links

Blog: https://greaterdc.urban.org/blog/we-need-better-zoning-data-data-science-can-help

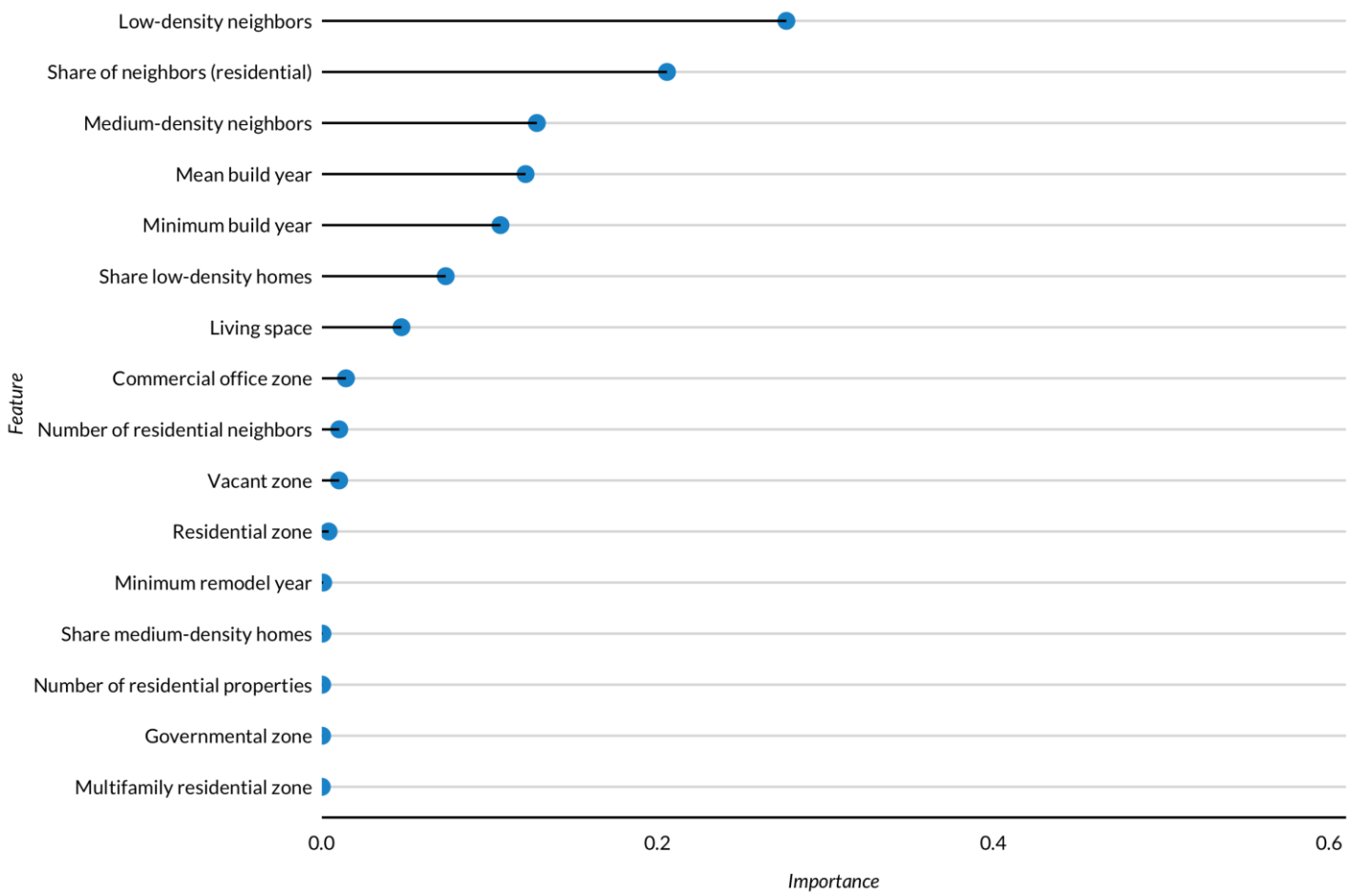
Technical Appendix:
https://www.urban.org/sites/default/files/2019/10/15/predicting_zoned_density_using_property_records_next_steps.pdf

URBAN
INSTITUTE · ELEVATE · THE · DEBATE

Appendix Slides

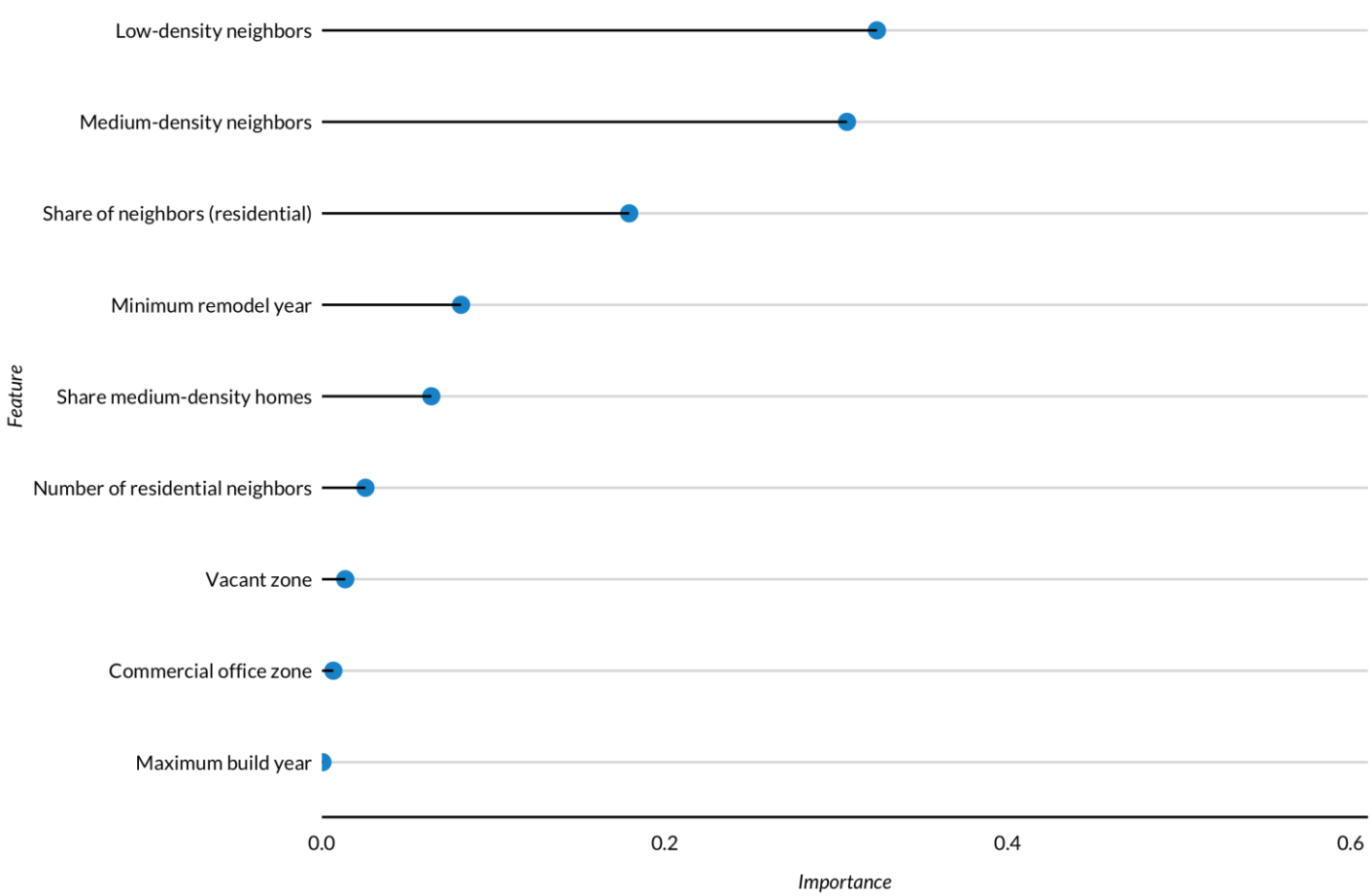
Feature importance

Washington, DC in-sample



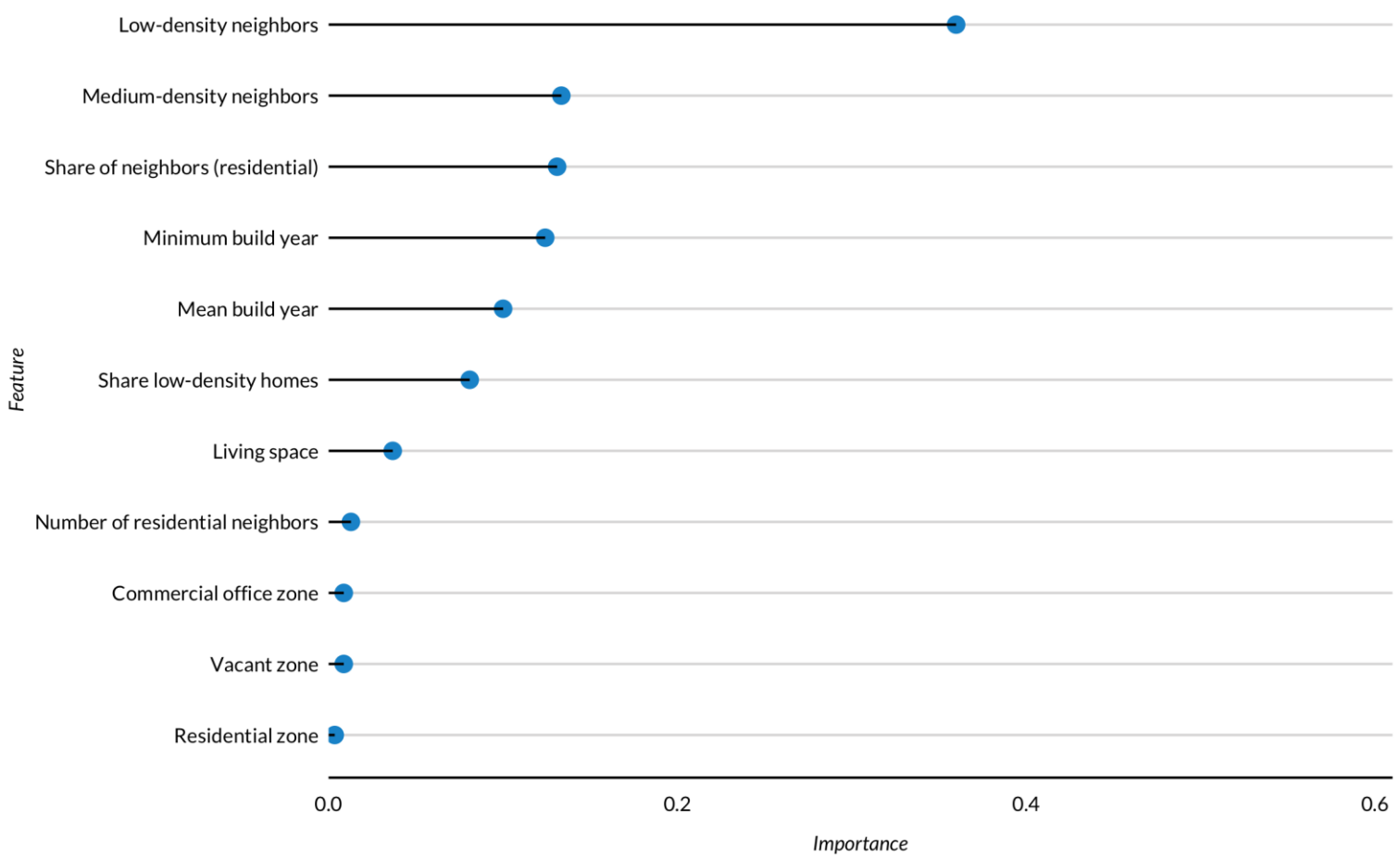
Feature importance

DC, Montgomery County, Arlington County in-sample



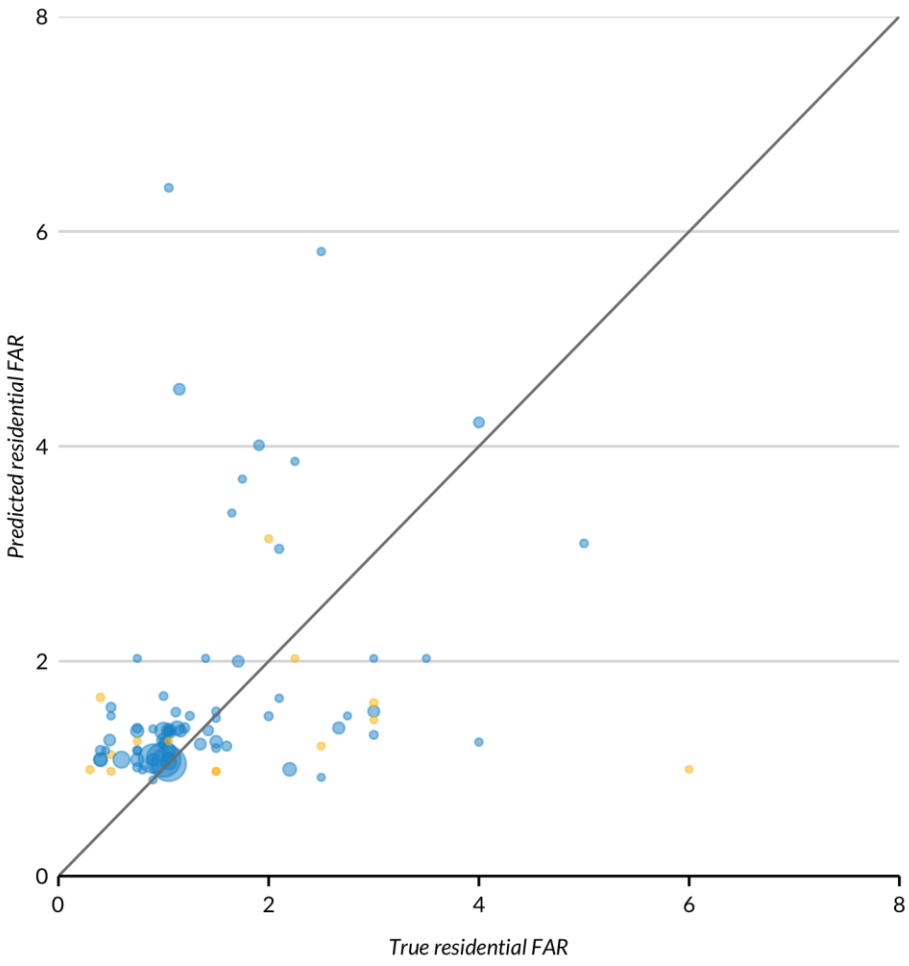
Feature importance

Montgomery County out-of-sample



Predicting Montgomery County

Training on DC & Arlington County



	Weighted RMSE	Weighted Relative MAE
Residential	0.43	0.22
Non-Residential	1.26	1.52