Local House Price Modeling:

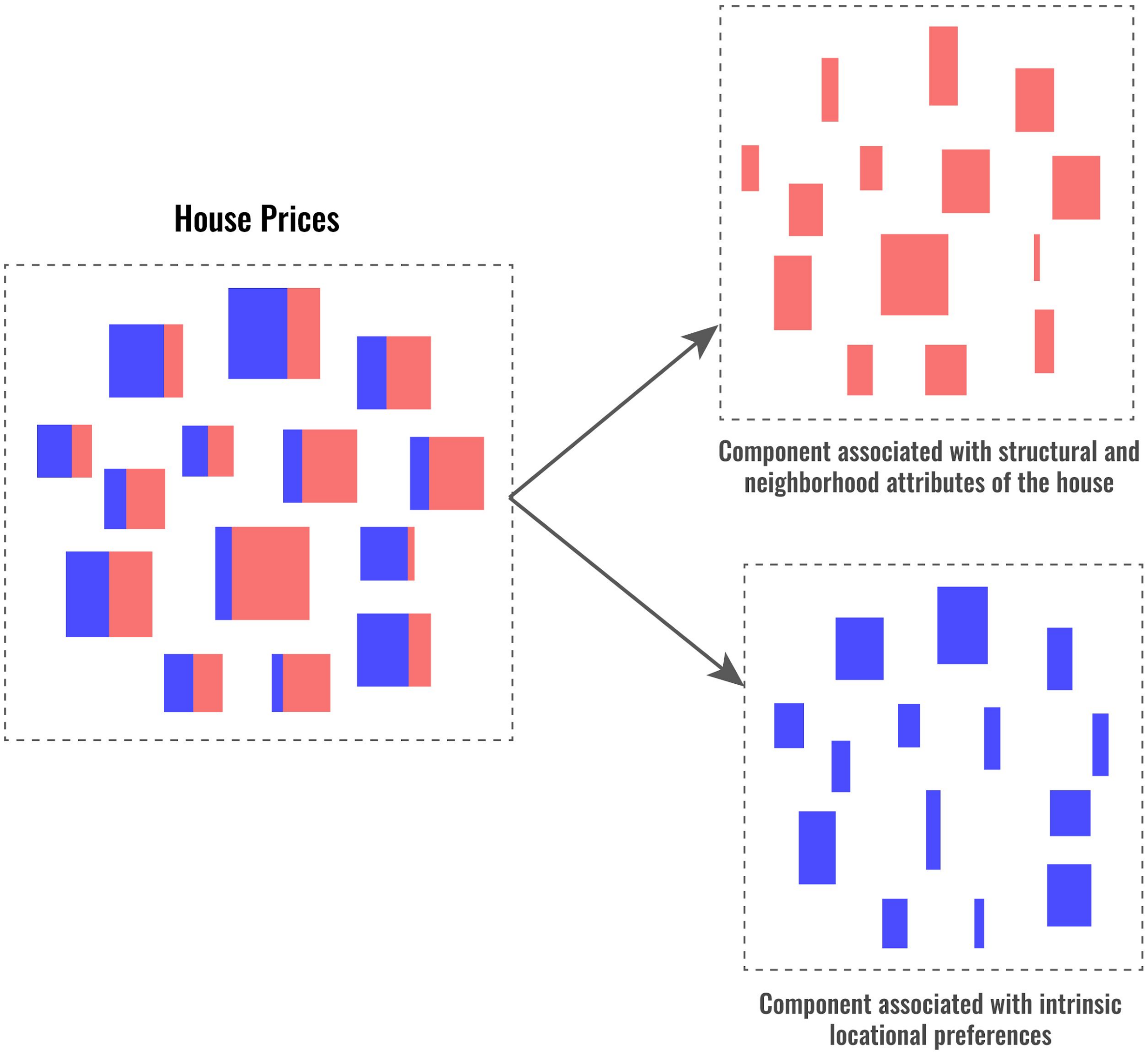
A Multiscale Approach Incorporating Spatial Non-stationarity

Mehak Sachdeva

Spatial Analysis Research Center
School of Urban Planning and Geographical Sciences
Arizona State University

Paper: Sachdeva, M., Fotheringham, A.S. and Li, Z. (2022). “Do places have value? Quantifying the intrinsic value of housing neighborhoods using MGWR” Journal of Housing Research.

Premise of the study



Traditional Hedonic Models

$$p = f(S, \quad N, \quad L)$$

Property price

Structural features

x1 Square feet living area
x2 Age of residence
x3 Basement present or not (categorical)

Neighborhood features

x5 Distance to the nearest waterfront
x6 Unemployment rate
x7 Percentage of technology sector jobs

Locational/Contextual features

$\boxed{?}$

MGWR as a model

$$p = f(S, N, L)$$

Property price

Structural features

x1 Square feet living area
x2 Age of residence
x3 Basement present or not (categorical)

Neighborhood features

x5 Distance to the nearest waterfront
x6 Unemployment rate
x7 Percentage of technology sector jobs

Locational/Contextual features

Response Outcome

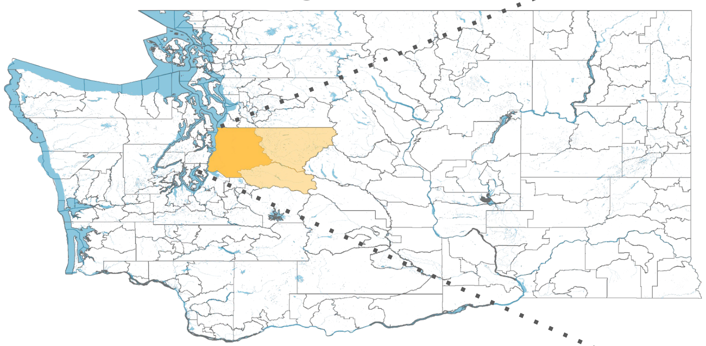
Measurable Independent Variables

Unmeasurable Contextual Effects

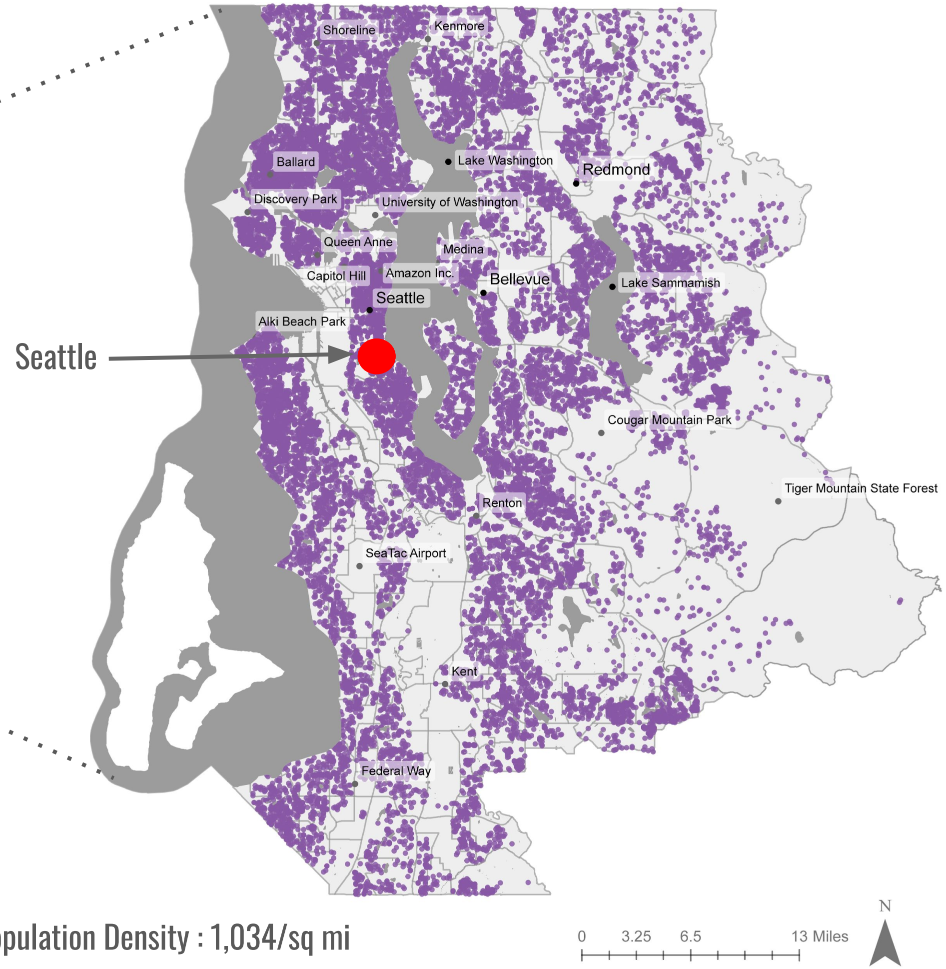
Study Area

Location

Washington State



King County, WA



Area : 2,116 sq mi (5,480 km^2) | Population Density : 1,034/sq mi

Data source: https://www.kaggle.com/harlfoxem/housesalesprediction

date - Date house was sold

price - Price is prediction target

bedrooms - Number of Bedrooms/House

bathrooms - Number of bathrooms/bedrooms

sqft_living - square footage of the home

sqft_lot - square footage of the lot

floors - Total floors (levels) in house

waterfront - House which has a view to waterfront

grade - grade of housing unit

sqft_above - sq.ft. of house apart from basement

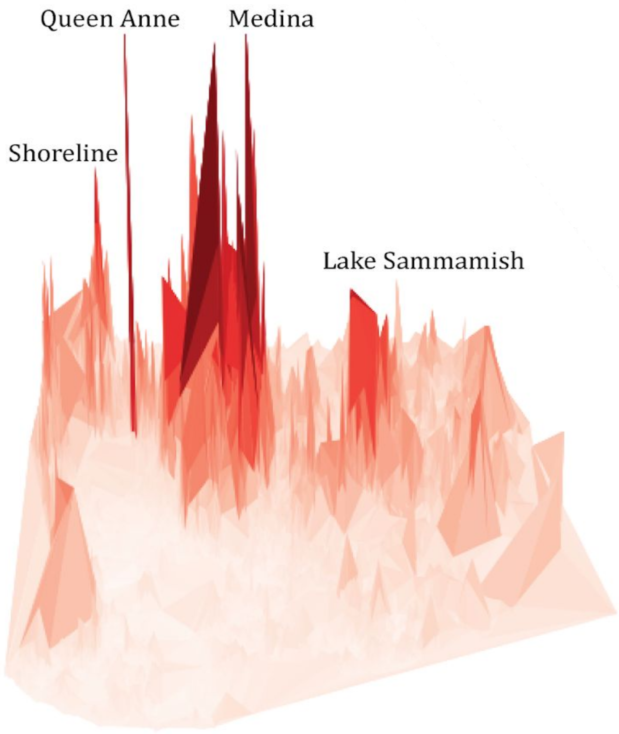
sqft_basement - square footage of the basement

yr_built - Built Year

yr_renovated - Year when house was renovated

sqft_living15 - Living room area in 2015

sqft_lot15 - Lot size area in 2015



21,613 points → 19,832 points

Maximum price - $7.7 Million , Minimum price - $75,000

Mean price - $0.54 Million

Variable selection

Literature review

Hedonic Models

In traditional linear regression form and calibrated using the ordinary least squares (OLS) technique

Issue: Ignores the spatial effects commonly existing in housing prices

$$P_i = \sum_j \boldsymbol{\beta}_j X_{ij} + \varepsilon_i$$

Global Spatial Hedonic Models

These address spatial dependence or spatial autocorrelation in spatial processes assuming spatial autocorrelation to be either in the response variables or in the error term

Issue: Housing price processes are assumed to be constant or stationary over space

$$\mathrm{W}y = \mathrm{W}_{\mathrm{NT}}\, y = (\mathrm{I}_{\mathrm{T}} \otimes \mathrm{W}_{\mathrm{N}})\, y,$$

$$\mathrm{W}X = \mathrm{W}_{\mathrm{NT}}\, X = (\mathrm{I}_{\mathrm{T}} \otimes \mathrm{W}_{\mathrm{N}})\, X,$$

$$\mathrm{W}\varepsilon = \mathrm{W}_{\mathrm{NT}}\, \varepsilon = (\mathrm{I}_{\mathrm{T}} \otimes \mathrm{W}_{\mathrm{N}})\, \varepsilon$$

Local Spatial Hedonic Models

Linear models where parameters are allowed to vary over space to better represent processes generating housing prices

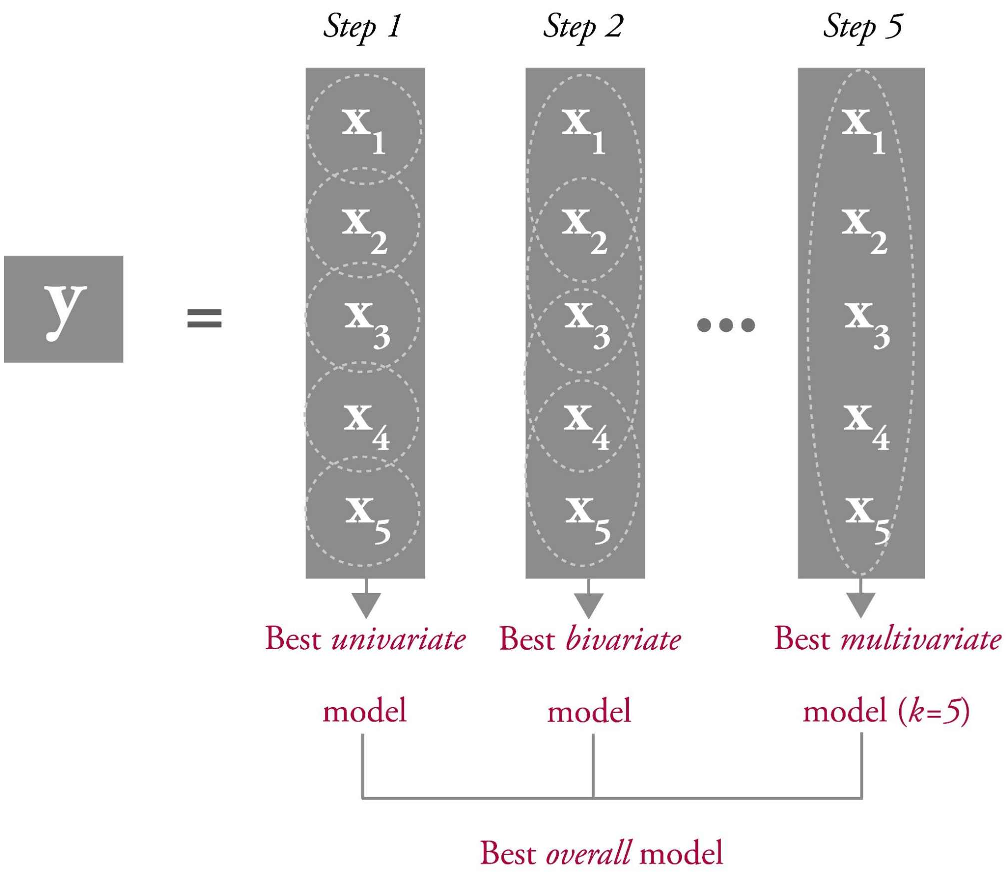
Issue: Does not account for temporal effects on housing processes

$$P_i = \sum_j \boldsymbol{\beta}_{ij} (u_i, v_i)\, X_{ij} + \varepsilon_i$$

Variable selection

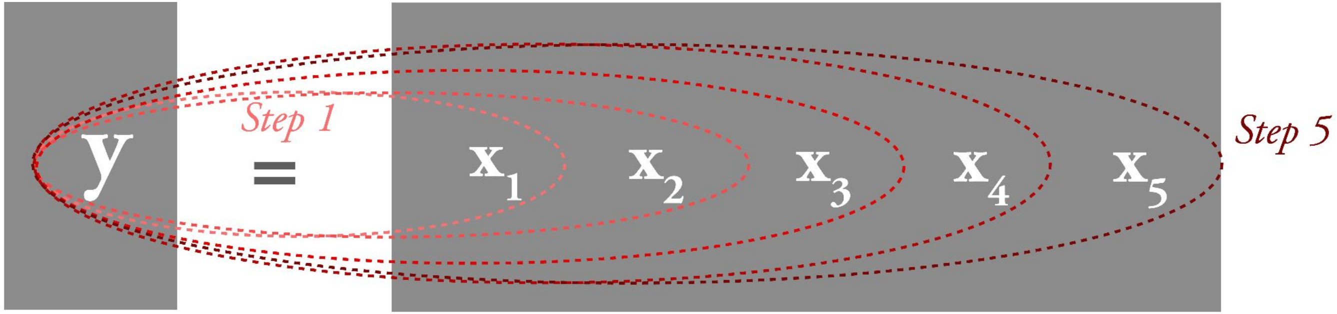
Total = 2^5 = *32 possible combinations*

Best subset



Variable selection

Forward selection



The variable with the **greatest additional improvement** *to the fit is added to the model*

Variable selection

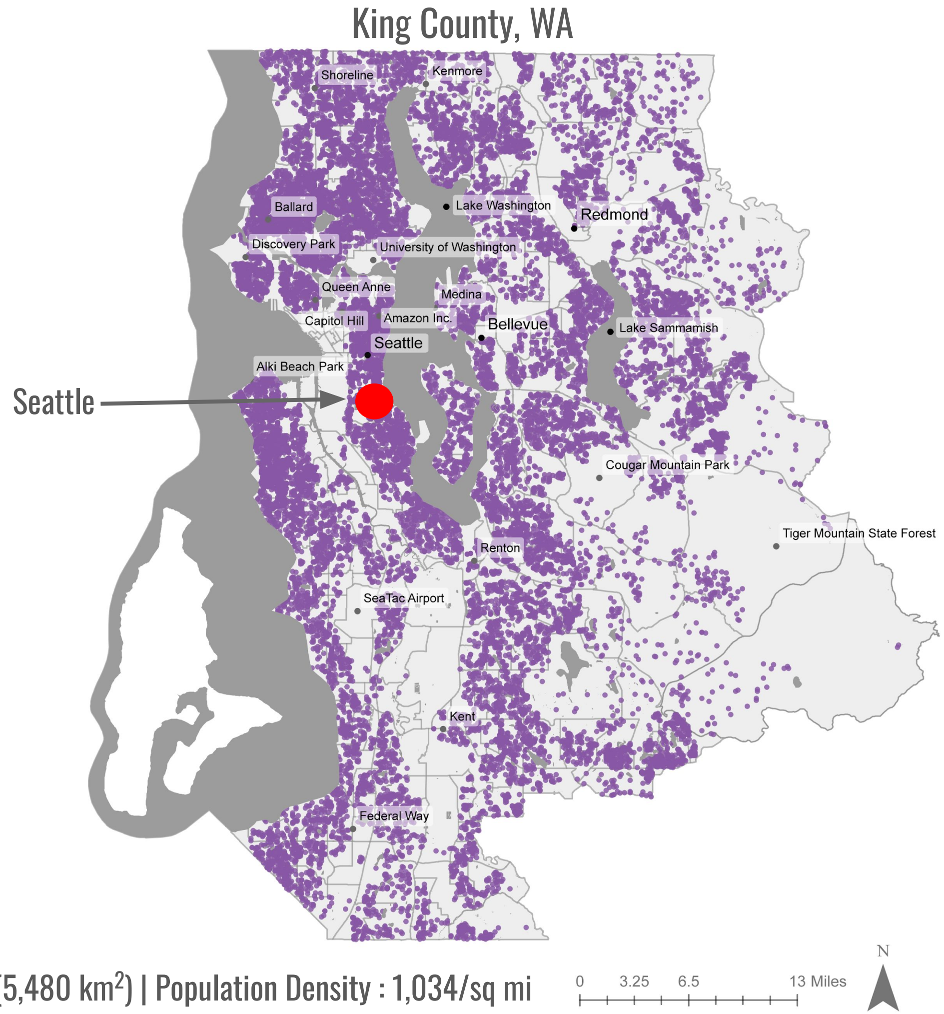
Contextual

Constructed an "index" variable:

Houses close to the waterfront and at high elevation = 1

Houses away from the waterfront and at a lower elevation = 0

Approximates: "Waterfront view"



Area : 2,116 sq mi (5,480 km^2) | Population Density : 1,034/sq mi

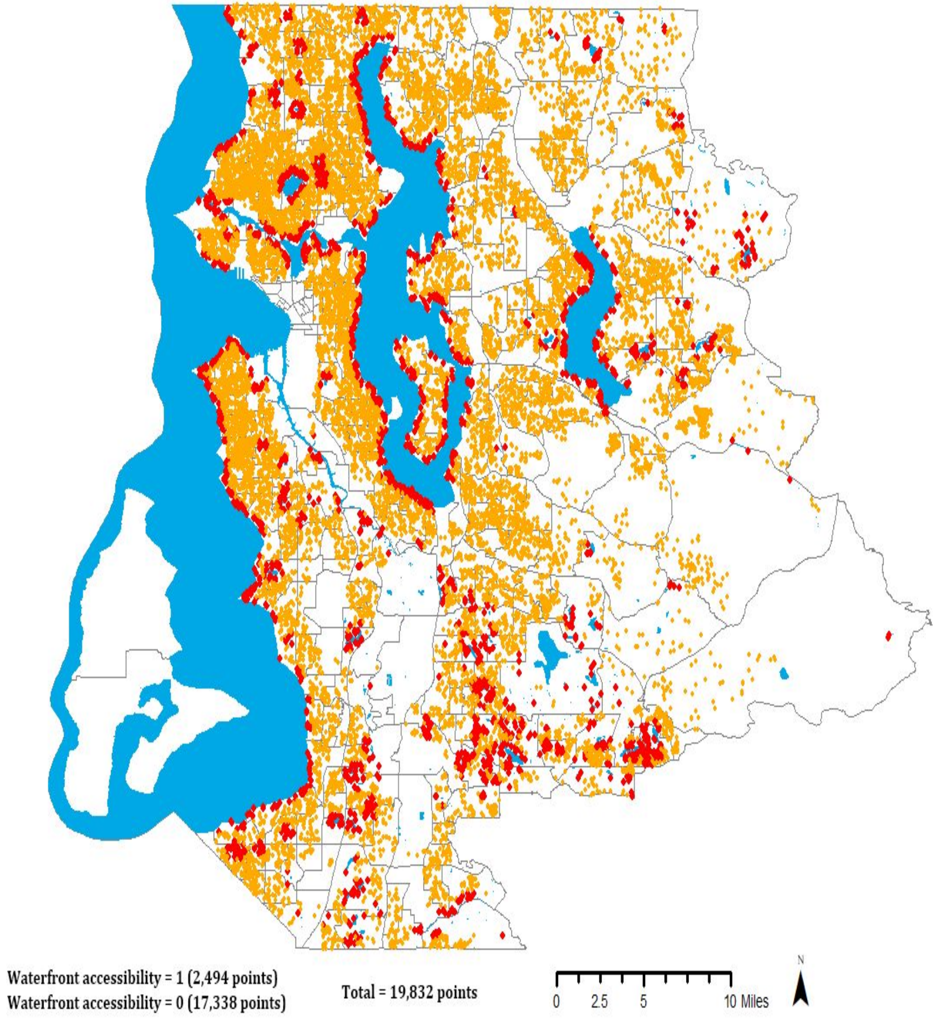
Variable selection

Categorical variables

Converted waterfront accessibility from (0,1)

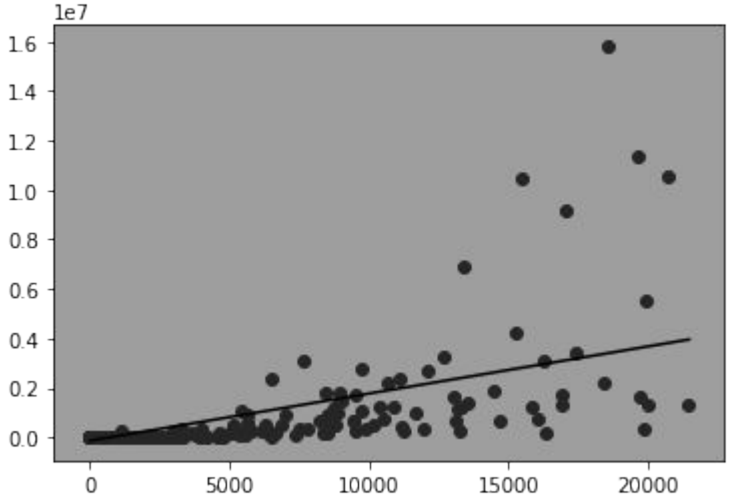
to

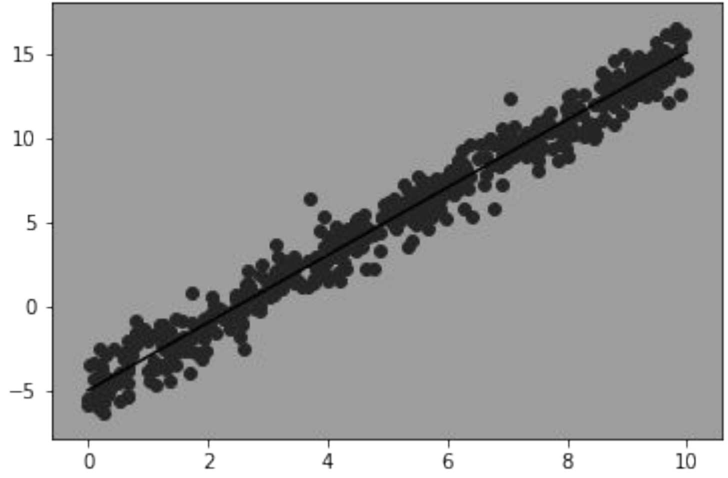
Distance to nearest waterfront (continuous)



Log Transformation

$$\ln y_i = \sum_j \beta_{ij} (u_i , v_i) \ln X_{ij} + \varepsilon_i$$





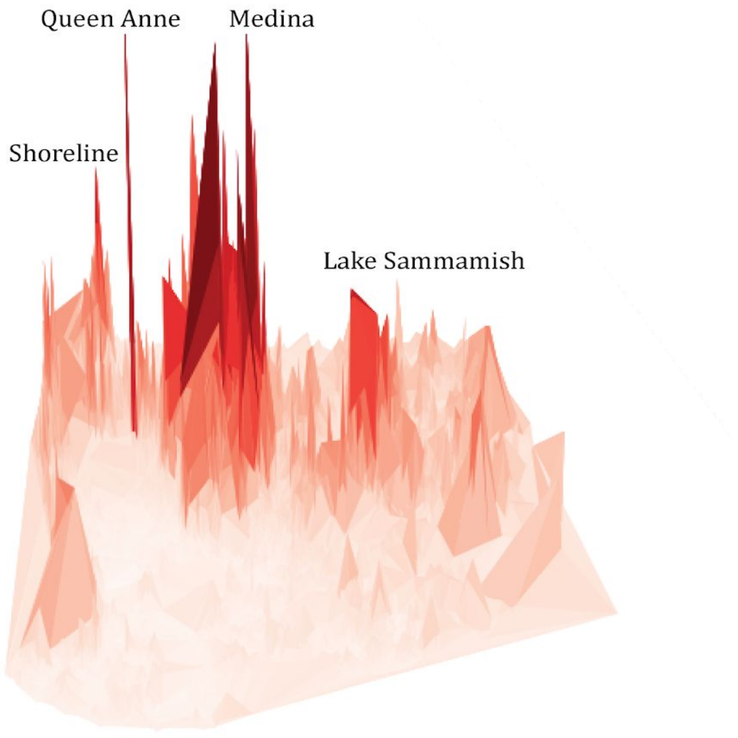
Data description

Dependent Variable

House Sales Price (May, 2015 to May, 2016) - in dollars

Independent Variables

1. Square Footage of living area
2. Age of the structure
3. Presence of basement in a residence
4. Distance to the nearest waterfront (constructed using Near Distance Tool - ESRI ArcMap Software)
5. Unemployment Rate (2014 ACS - 5 year estimate interpolated from census tracts)
6. Percentage of technology sector jobs (2014 ACS - 5 year estimates, interpolated from census tracts)
7. Index - composite measure of waterfront access and elevation (capturing view from the house to waterfront)



21,613 points - - - - - - - - - ► 19,832 points

Maximum price - $7.7 Million , Minimum price - $75,000

Mean price - $0.54 Million

Global Model Results

Covariates

x1 Square feet living area

x2 % of technology sector jobs

x3 Unemployment rate

x4 Basement present or not (categorical)

x5 Distance of the nearest waterfront from the property

x6 Age of the structure

x7 Composite index

	Coefficients
Constant	6.89 e-17
β_1	0.57***
β_2	0.422***
β_3	-0.08***
β_4	- 0.032***
β_5	- 0.257***
β_6	0.011***
β_7	- 0.031***
R^2	0.764

Covariate Effect

x1	Square feet living area
x2	% of technology sector jobs
x6	Age of the structure
x7	Composite index
x4	Basement present or not (categorical)
x3	Unemployment rate
x5	Distance of the nearest waterfront from the property

	Coefficients
Constant	6.89 e-17
β_1	0.57***
β_2	0.422***
β_6	0.011***
β_7	- 0.031***
β_4	- 0.032***
β_3	-0.08***
β_5	- 0.257***
R^2	0.764

Interpreting Log-Log Model Estimates

Square feet living area - 1% increase in sq.ft. living area increases price by 0.57 %

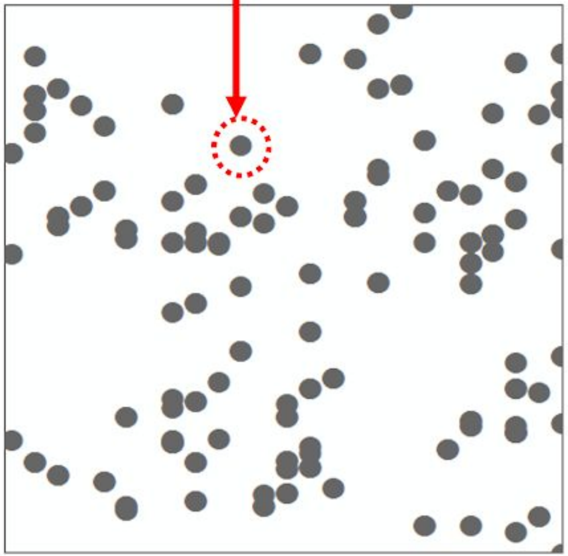
⇉ 1 sq.ft. increase leads to $130 increase in price

	Coefficients
Constant	6.89 e-17
β_1	0.57***
β_2	0.422***
β_6	0.011***
β_7	- 0.031***
β_4	- 0.032***
β_3	-0.08***
β_5	- 0.257***
R^2	0.764

MGWR Results

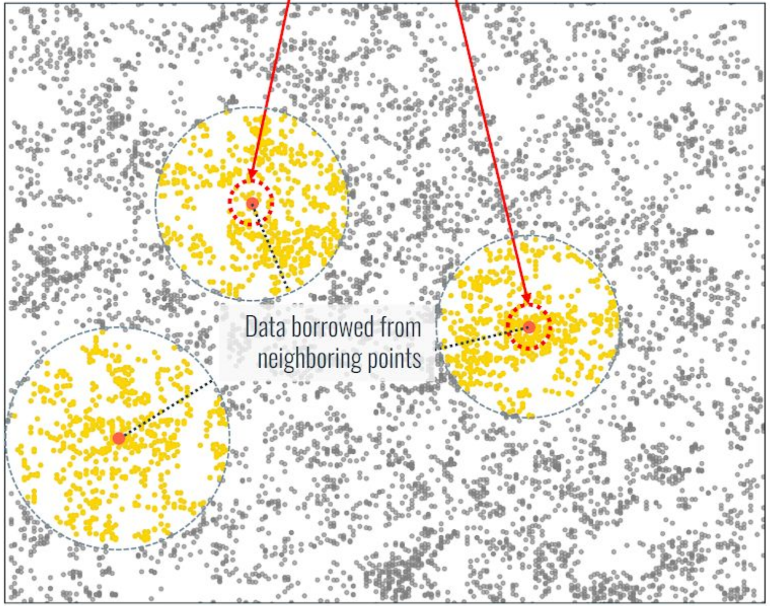
MGWR model

$$y_i = \sum_j \boldsymbol{\beta}_{ij}(u_i, v_i) X_{ij} + \boldsymbol{\varepsilon}_i$$



MGWR model

$$y_i = \sum_j \beta_{ij}\underbrace{(u_i, v_i)} X_{ij} + \varepsilon_i$$

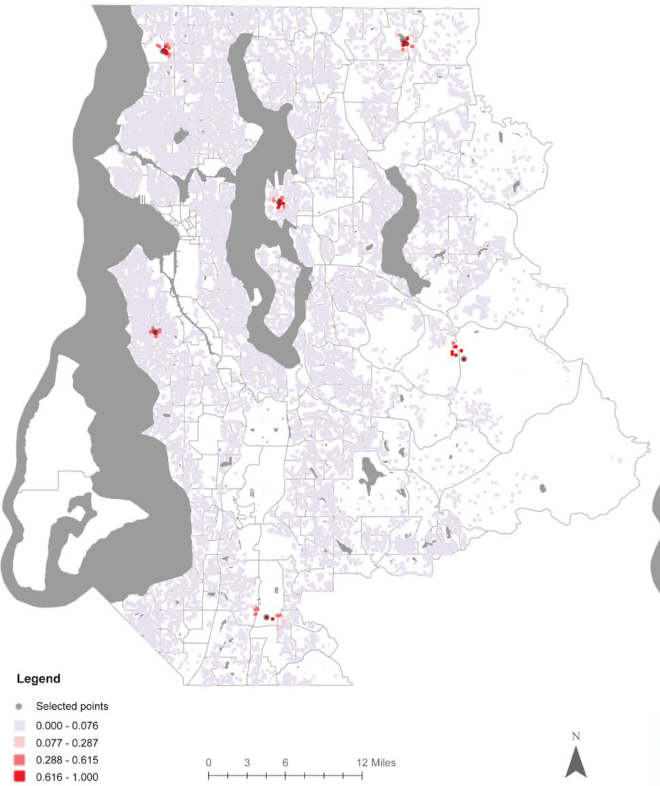


MGWR model - R-squared = 0.91

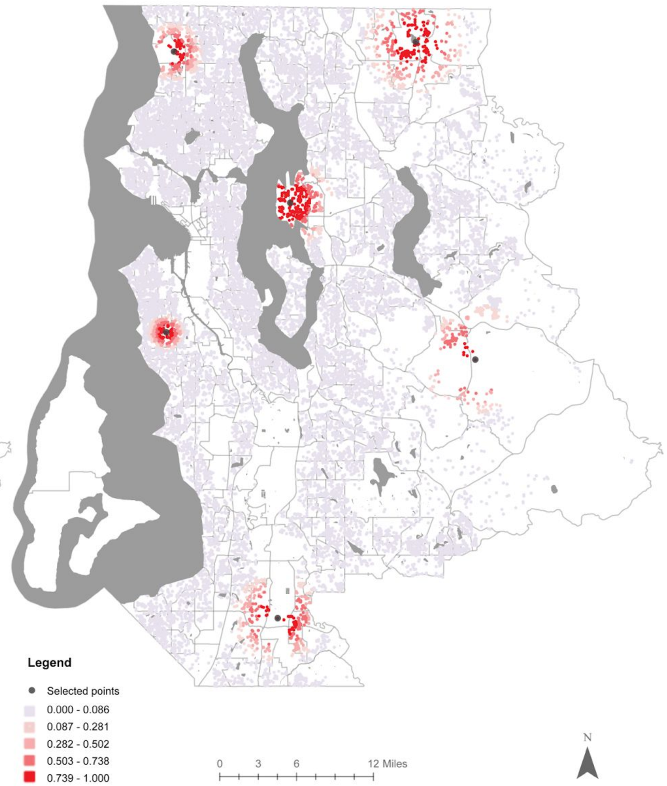
Variable	Bandwidths	
Unemployment rate	2990	Global
Basement present or not (categorical)	2072	Global
% of technology sector jobs	1552	Global
Age of the structure	845	Regional
Constant	463	Regional
Distance to nearest waterfront from the property	179	Local
Square feet living area	62	Local
Composite Index	45	Local

Bandwidth Visualization

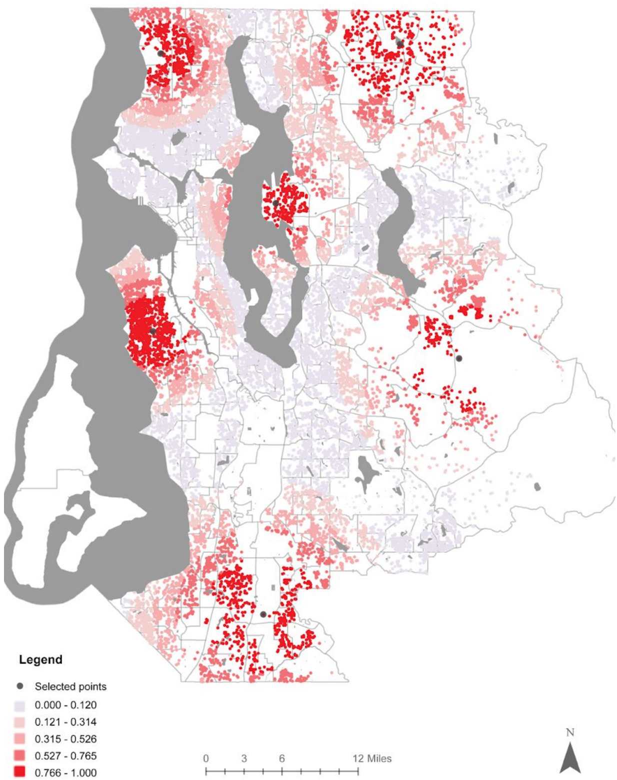
Bandwidth = 45 nearest neighbors



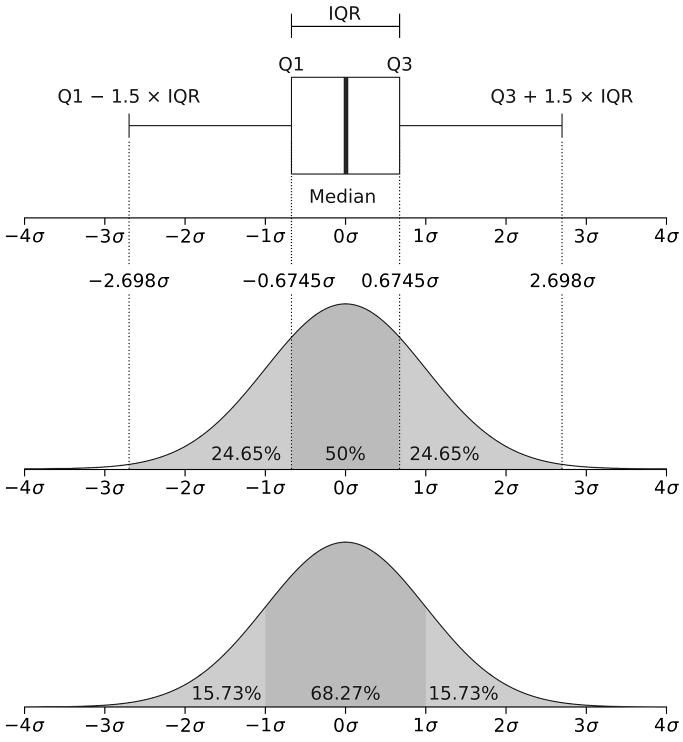
Bandwidth = 463 nearest neighbors



Bandwidth = 2,990 nearest neighbors



IQR - Variability tests of local parameter estimates



IQR of local estimates and Standard Errors of Global estimates

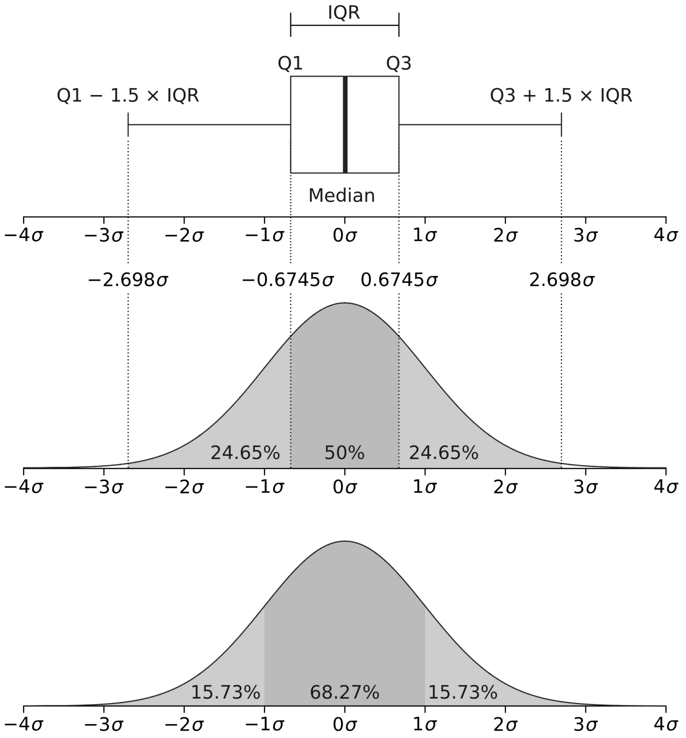
Empirically, 2*SE is considered the expected variation in the values (contains about 60% of all the values)

Indicates a possible nonstationary process if IQR (which includes 50% values) is larger than 2*SE

Test for Spatial Non Stationarity

Indicates a possible nonstationary process if IQR (which includes 50% values) is larger than 2*SE

Abbreviation	2 * Standard Error	IQR	Monte Carlo Test
sqft_living	0.024	0.41	✓
age	0.017	0.08	✓
basement_p	0.015	0.025	✓
waterfront_dist	0.014	0.86	✓
tech_jobs	0.012	0.046	✓
unemp_rate	0.015	0.024	✓
index	0.013	0.37	✓



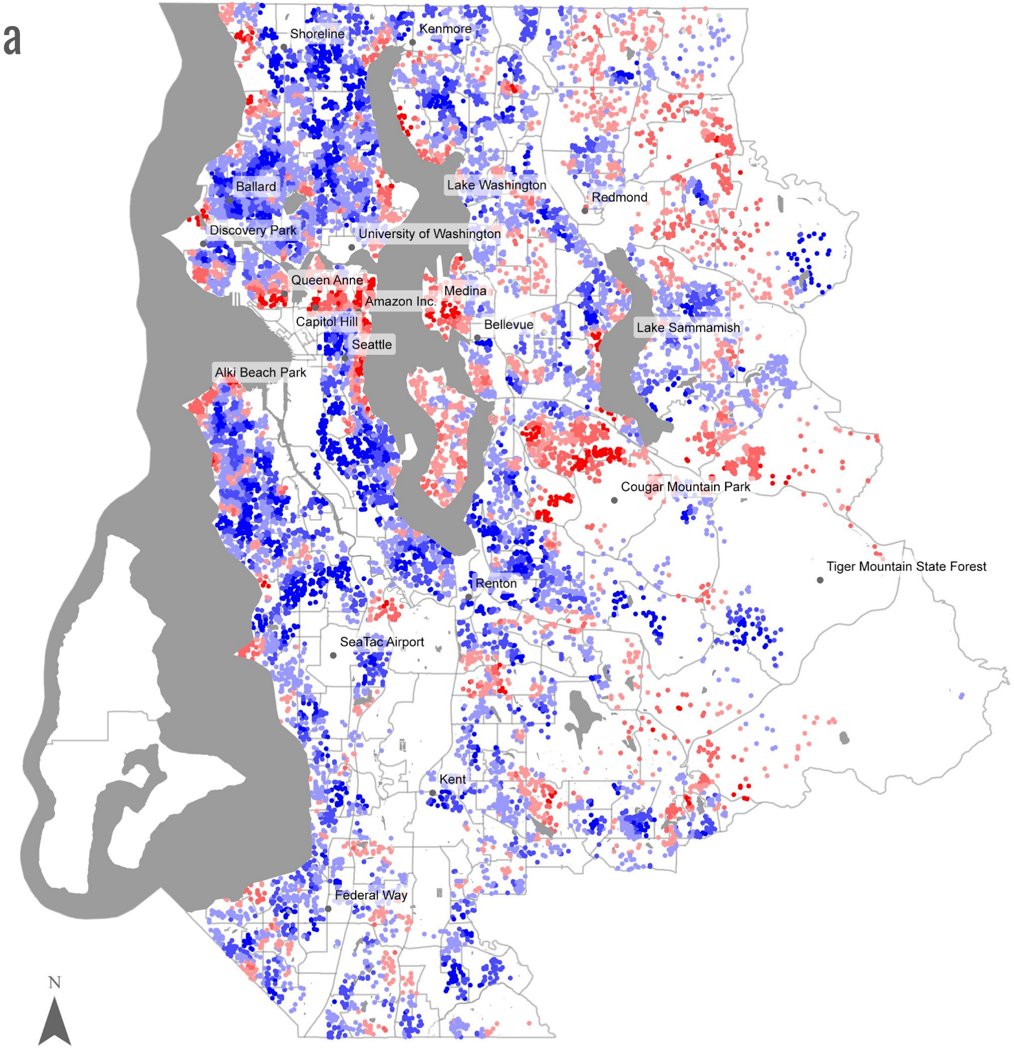
Parameter Estimates: Square footage of living area

OLS -> β = 0.57**

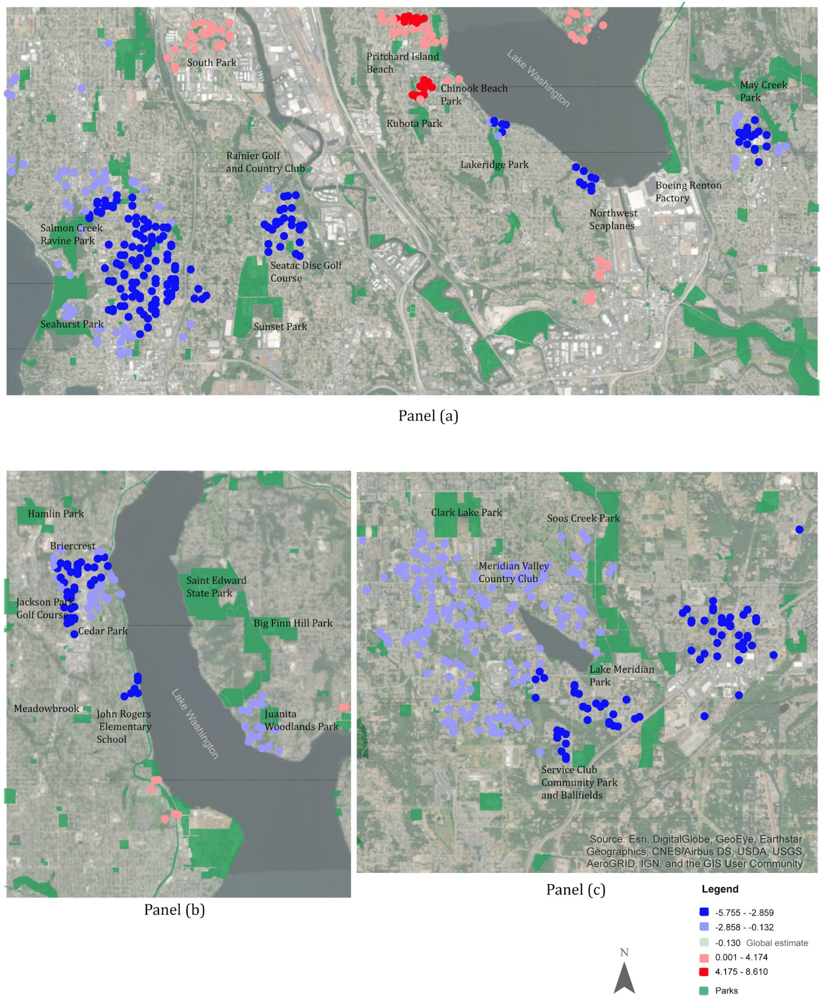
MGWR (BW = 62)

Legend

- 0.288 - 0.468
- 0.469 - 0.555
- 0.556 - 0.712
- 0.713 Global estimate
- 0.714 - 0.838
- 0.839 - 0.983
- 0.984 - 1.320



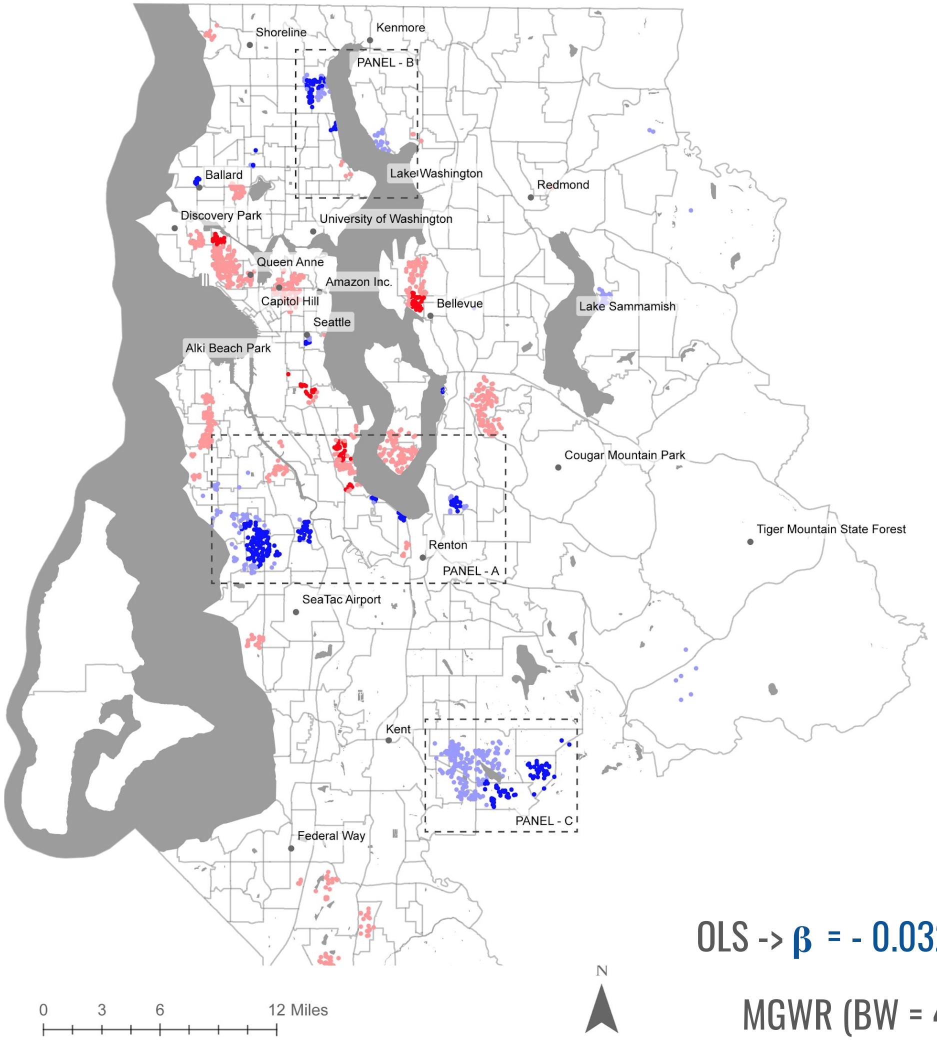
Parameter Estimates: Composite index



Panel (a)

Panel (b)

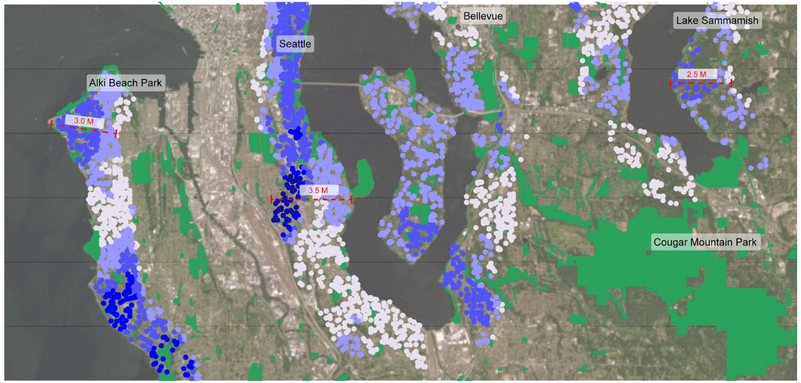
Panel (c)



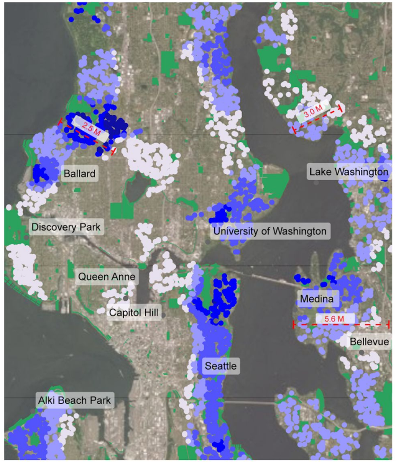
OLS -> β = - 0.032**

MGWR (BW = 45)

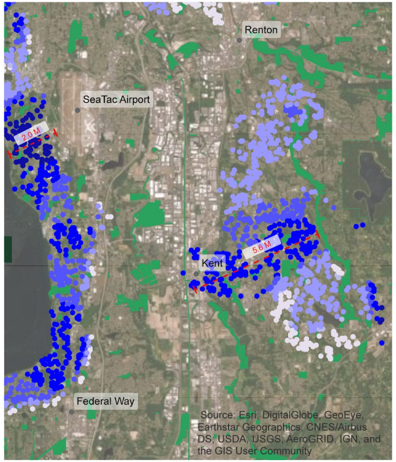
Parameter Estimates: Distance to nearest waterfront



Panel (a)



Panel (b)

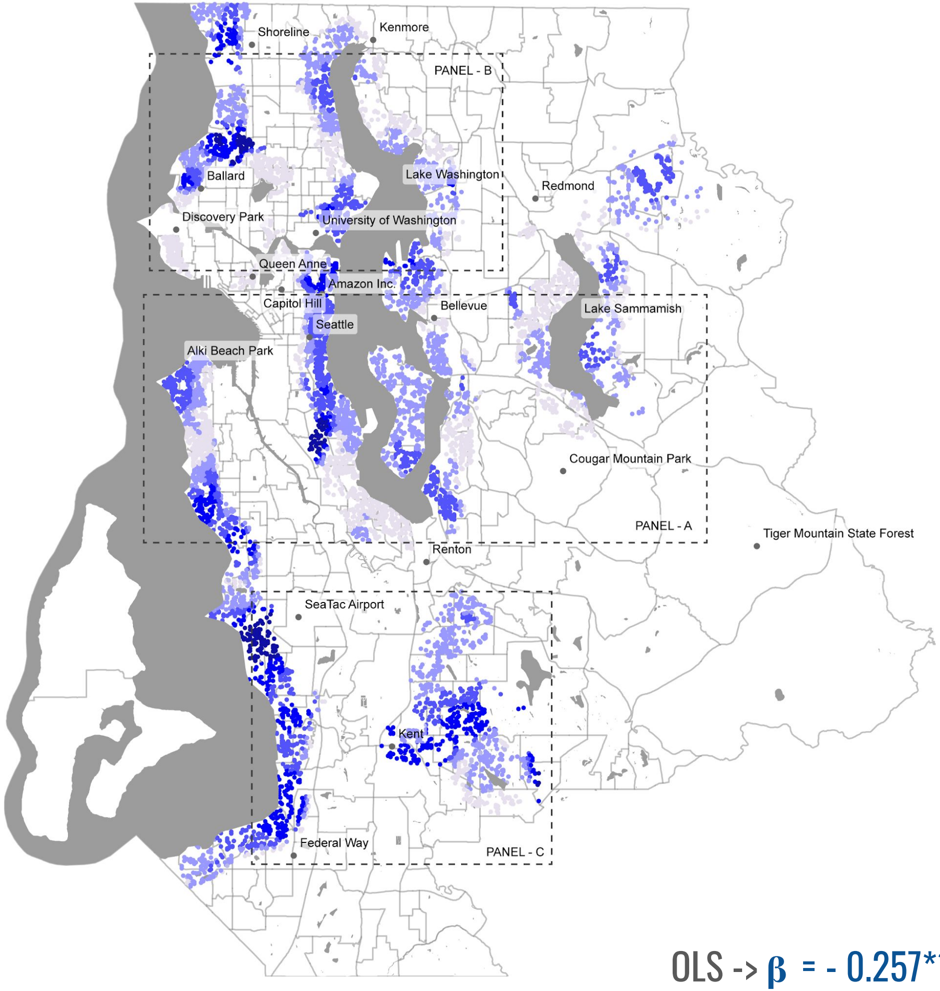


Panel (c)

Legend

- -0.484 - -0.279
- -0.278 - -0.210
- -0.209 - -0.172
- -0.171 - -0.130
- -0.129 Global estimate
- -0.127 - 0.000
- Park

N



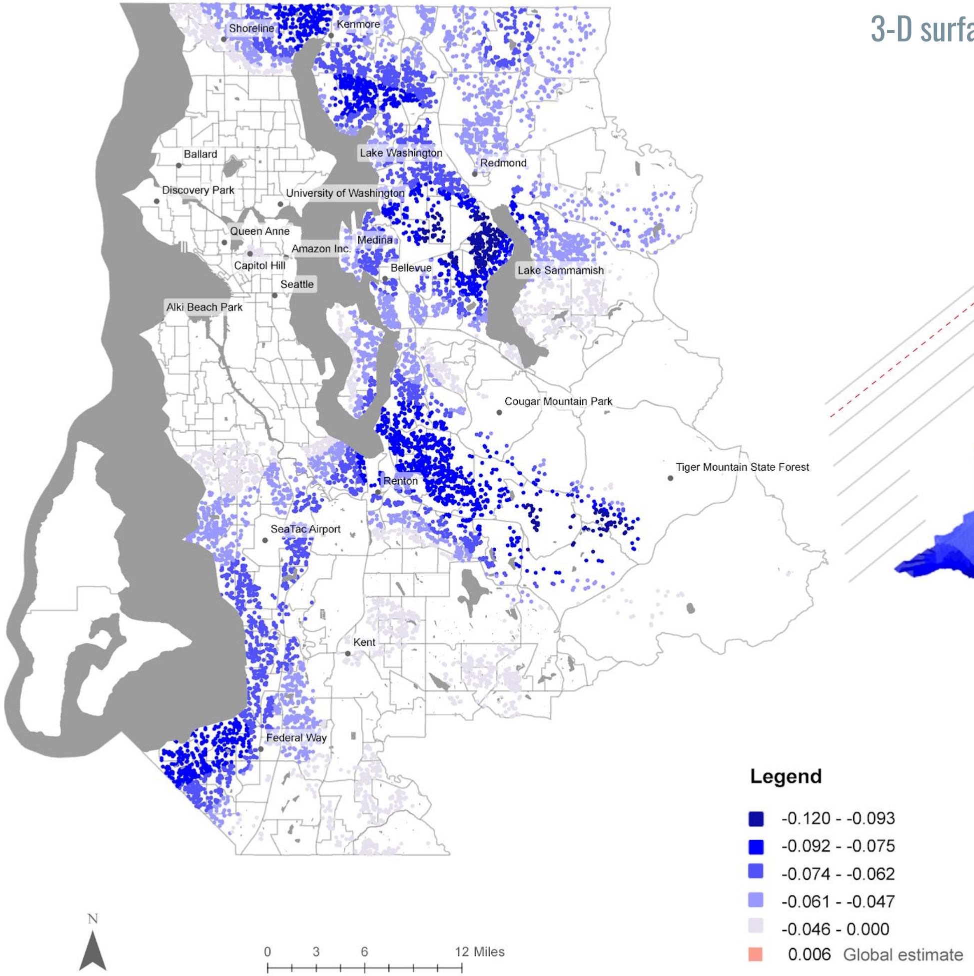
0 3 6 12 Miles

N

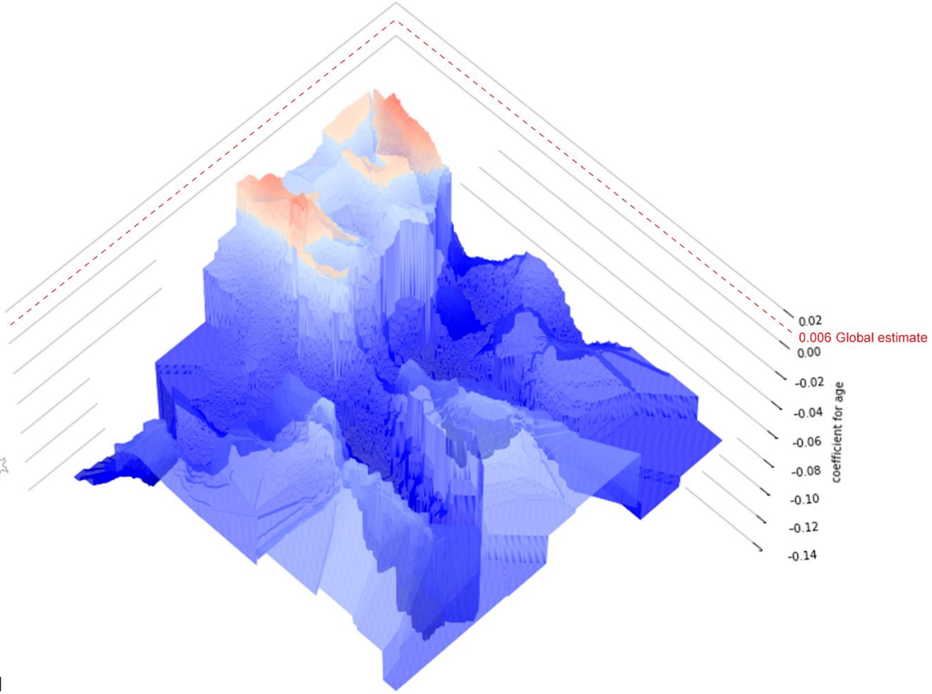
OLS -> β = - 0.257**

MGWR (BW = 179)

Parameter Estimates: Age of the structure



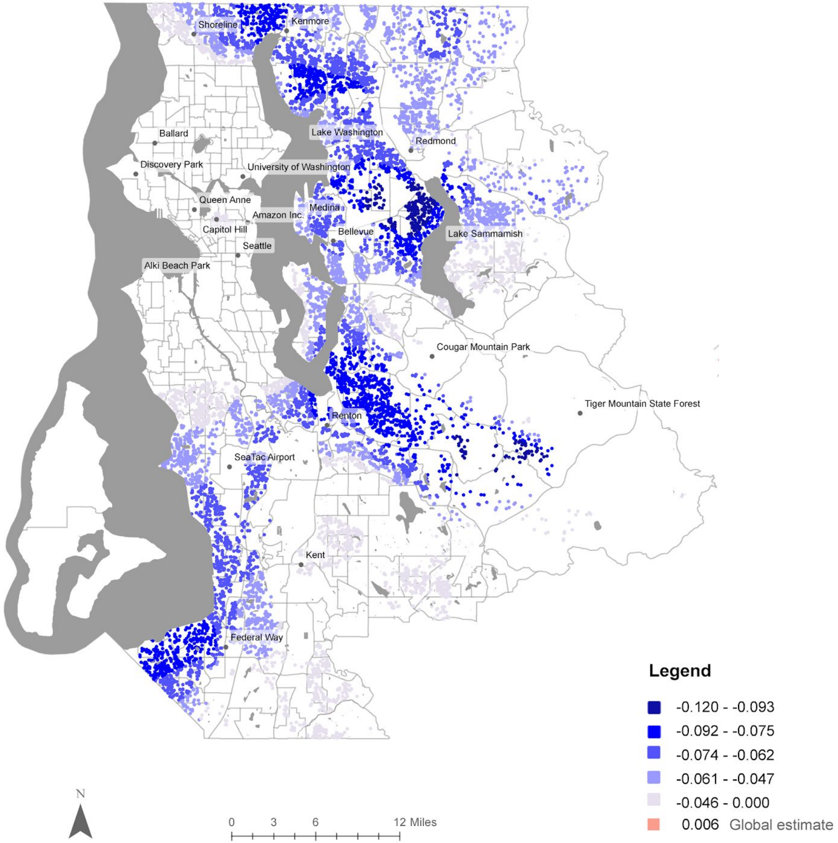
3-D surface of all parameter estimates from MGWR - Simpson's paradox



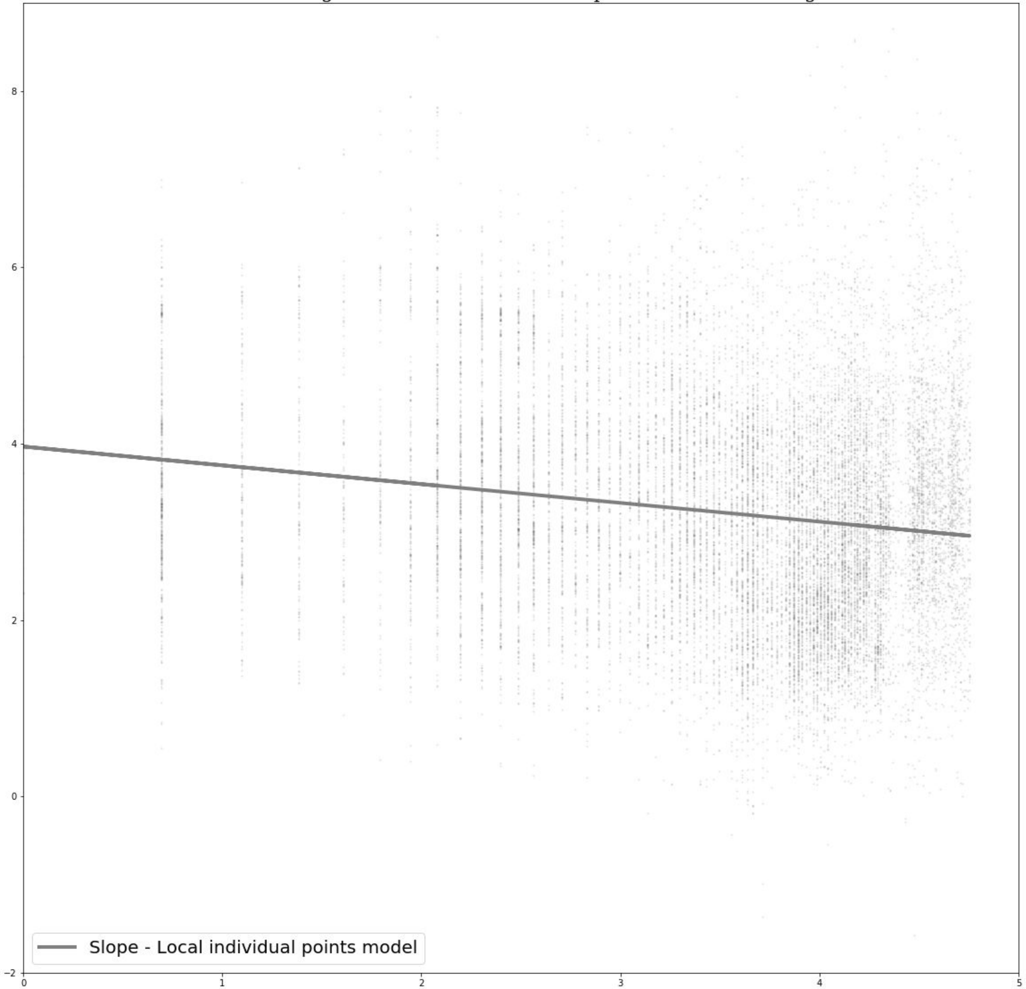
OLS -> β = 0.011***

MGWR (BW = 845)

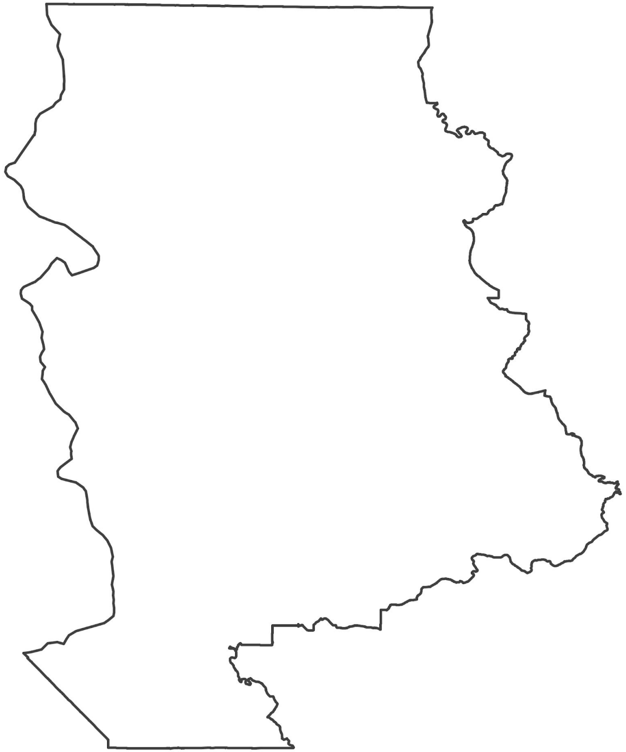
MAUP


Shoreline
Kenmore
Lake Washington
Redmond
Ballard
Discovery Park
University of Washington
Queen Anne
Amazon Inc
Medina
Capitol Hill
Bellevue
Lake Sammamish
Seattle
Alki Beach Park
Cougar Mountain Park
Tiger Mountain State Forest
Renton
SeaTac Airport
Kent
Federal Way
Legend
-0.120 - -0.093
-0.092 - -0.075
-0.074 - -0.062
-0.061 - -0.047
-0.046 - 0.000
0.006 Global estimate
N
0 3 6 12 Miles

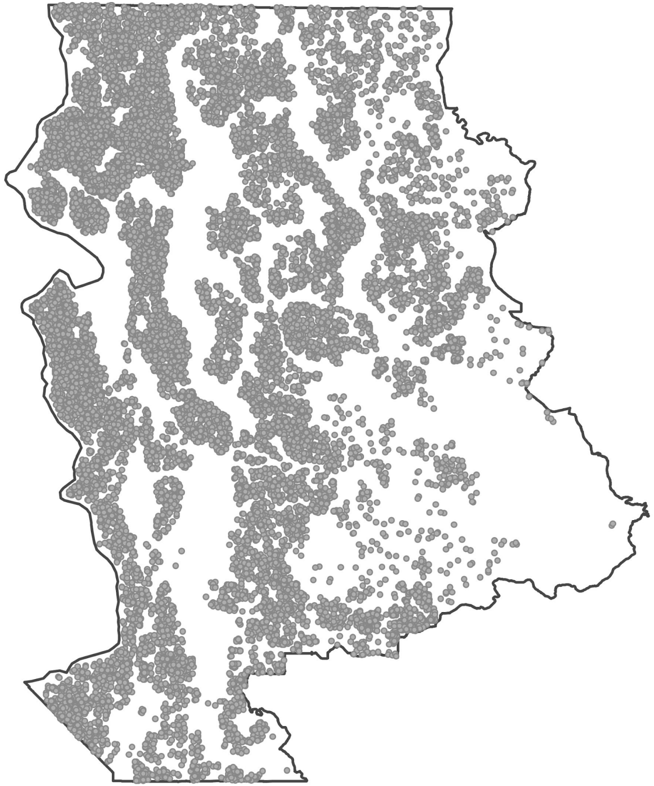
Conditional effect of age of a residence on house price value - local vs global models



Aggregation units

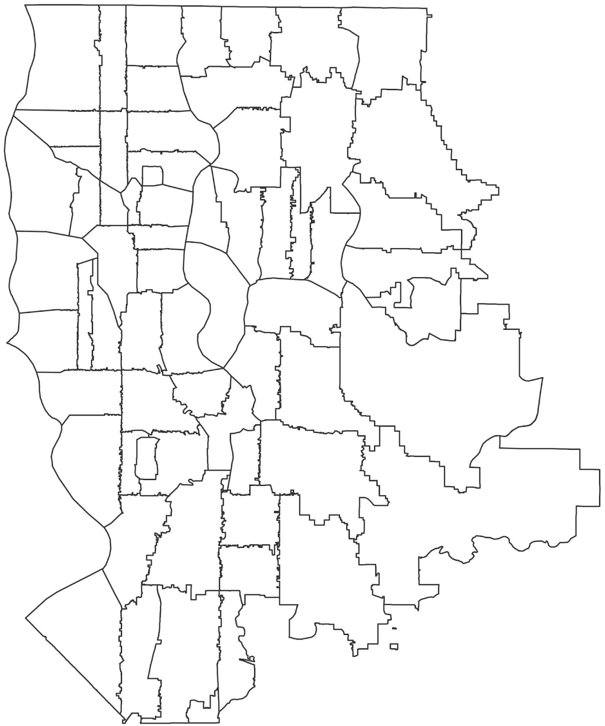


Global model
1

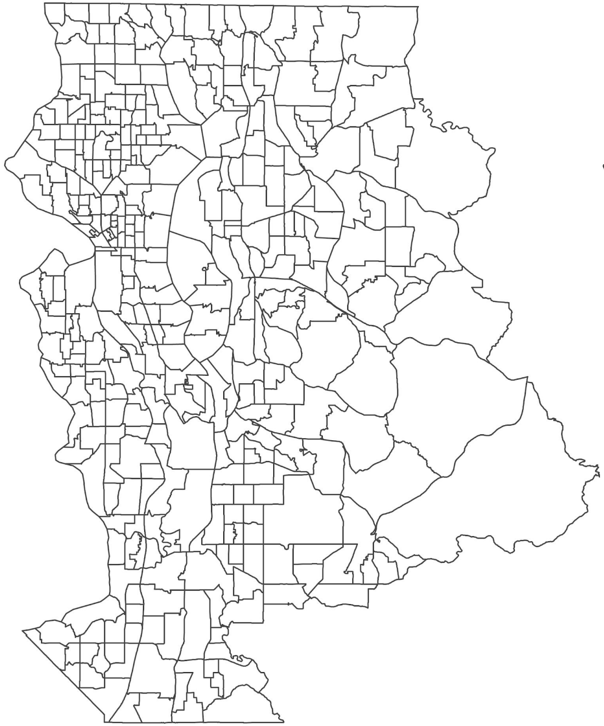


Individual points
19,832

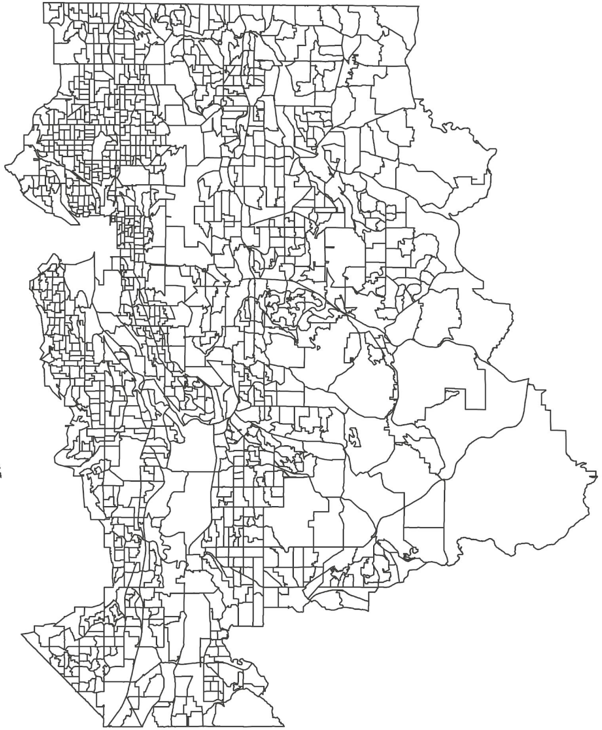
Aggregation units



Zipcodes
73



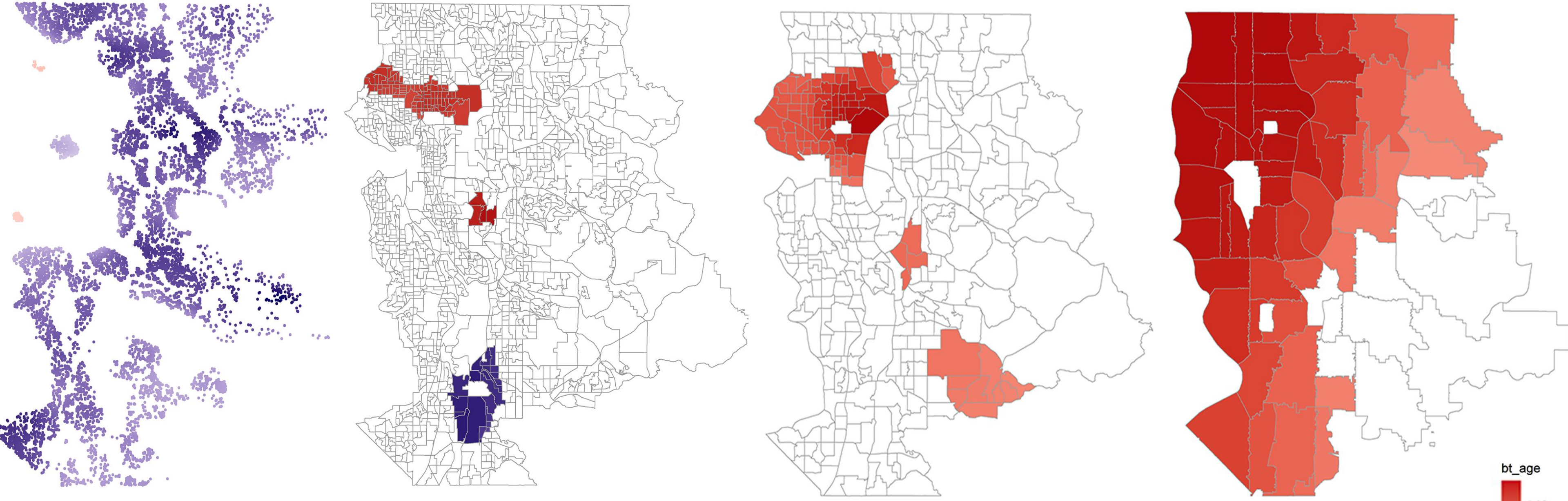
Census Tracts
373

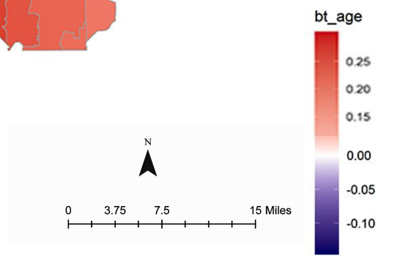


Block Groups
1,333

Simpson’s Paradox effect

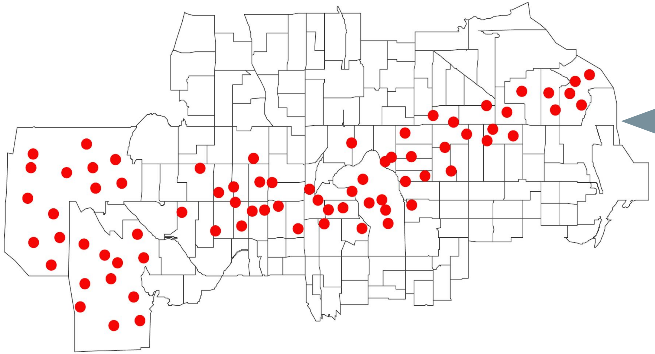
So which one of these maps is correct?



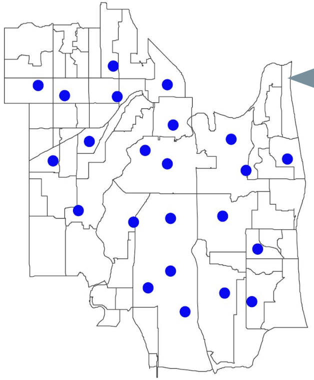


Depends on your question.

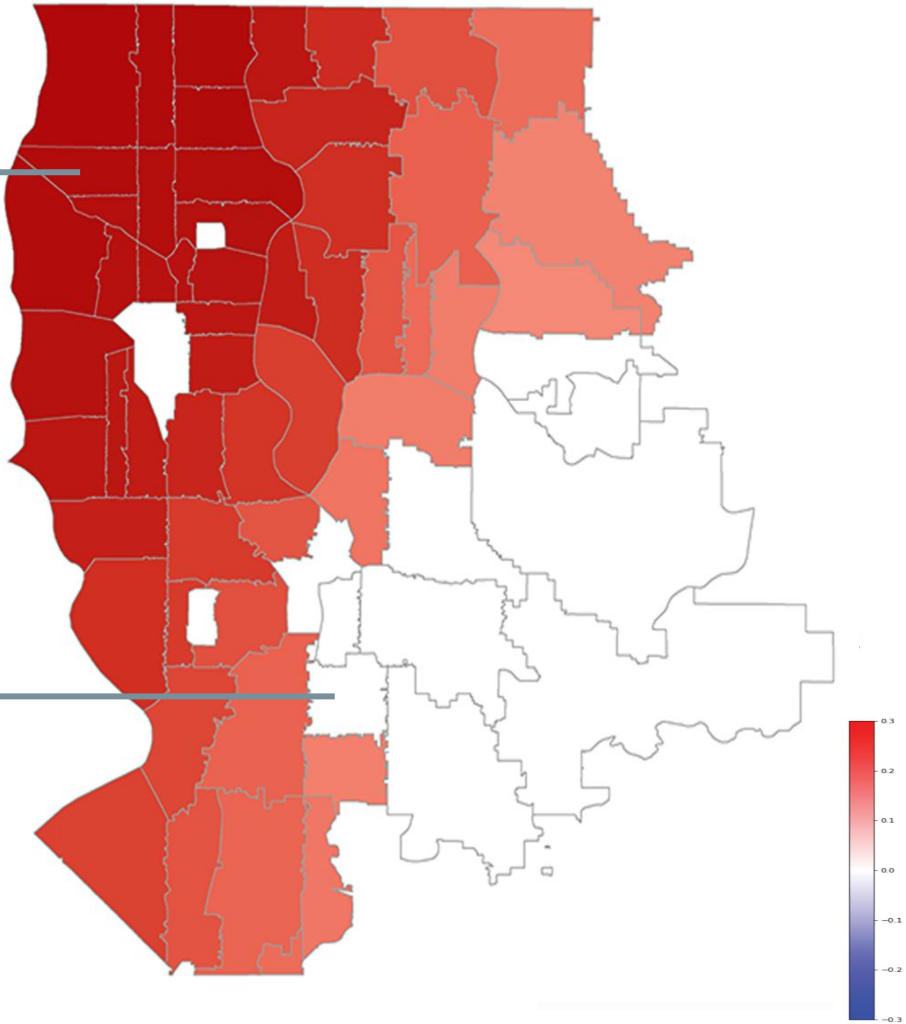
In global models



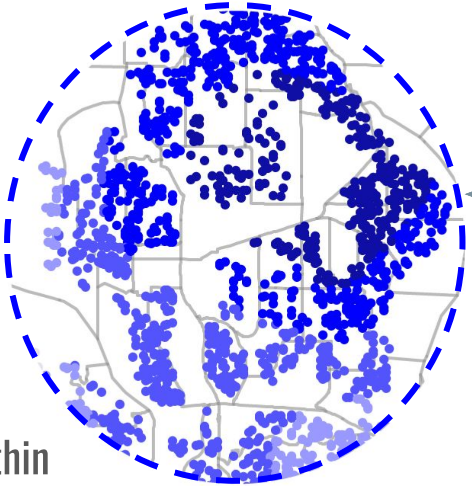
Areas with older and more expensive housing clusters are compared to areas with cheaper and newer housing clusters



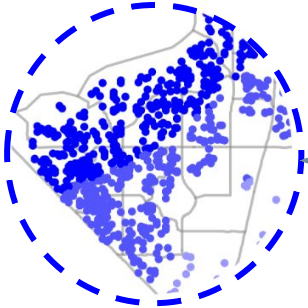
And hence, the results suggest a preference for older housing



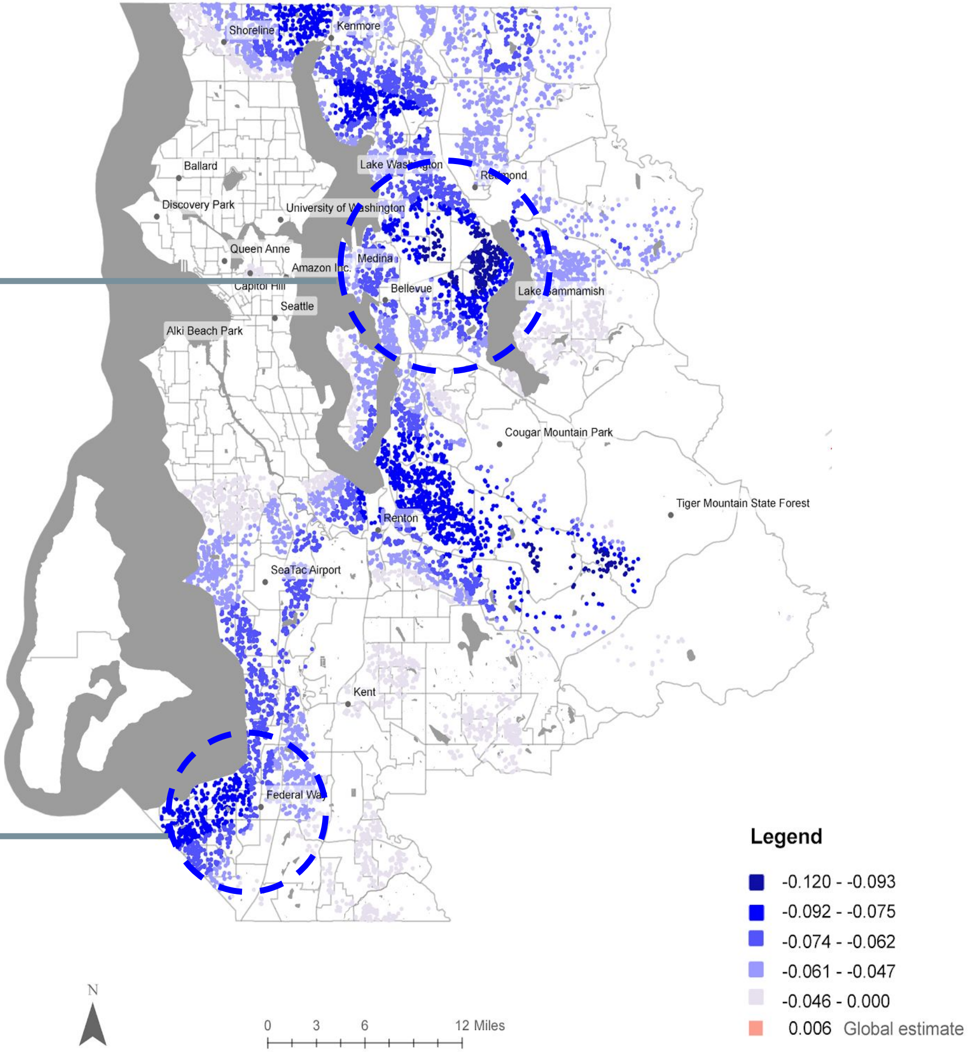
In local models with disaggregated data



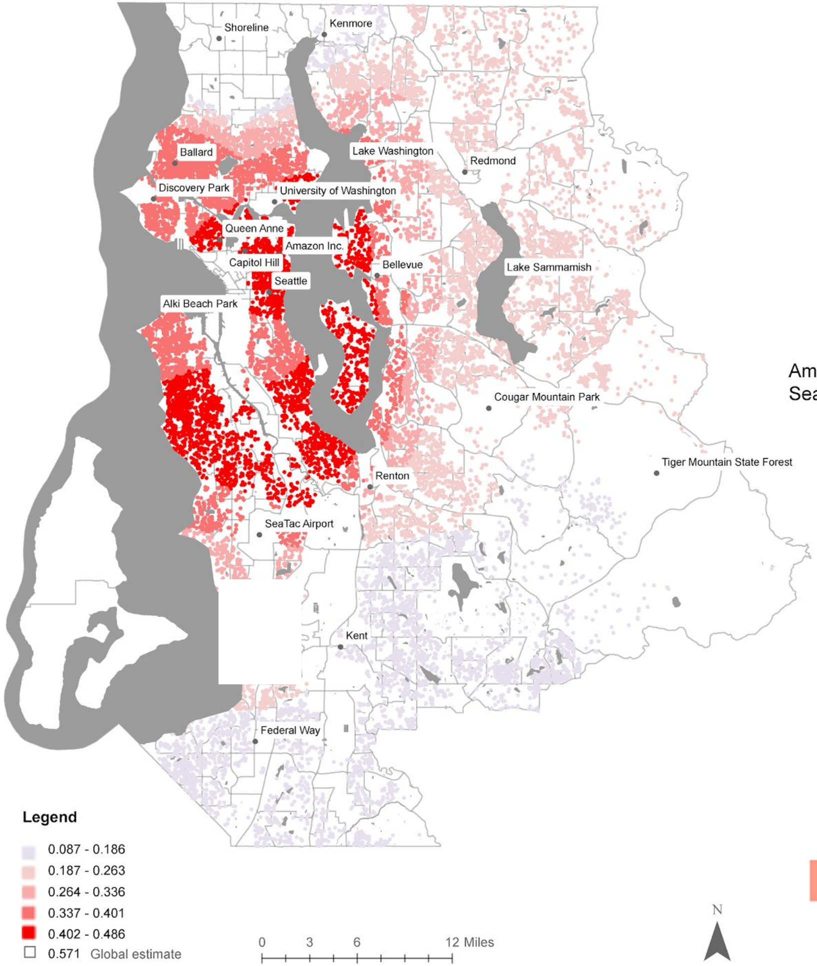
Similar houses within small subsets are compared with one another

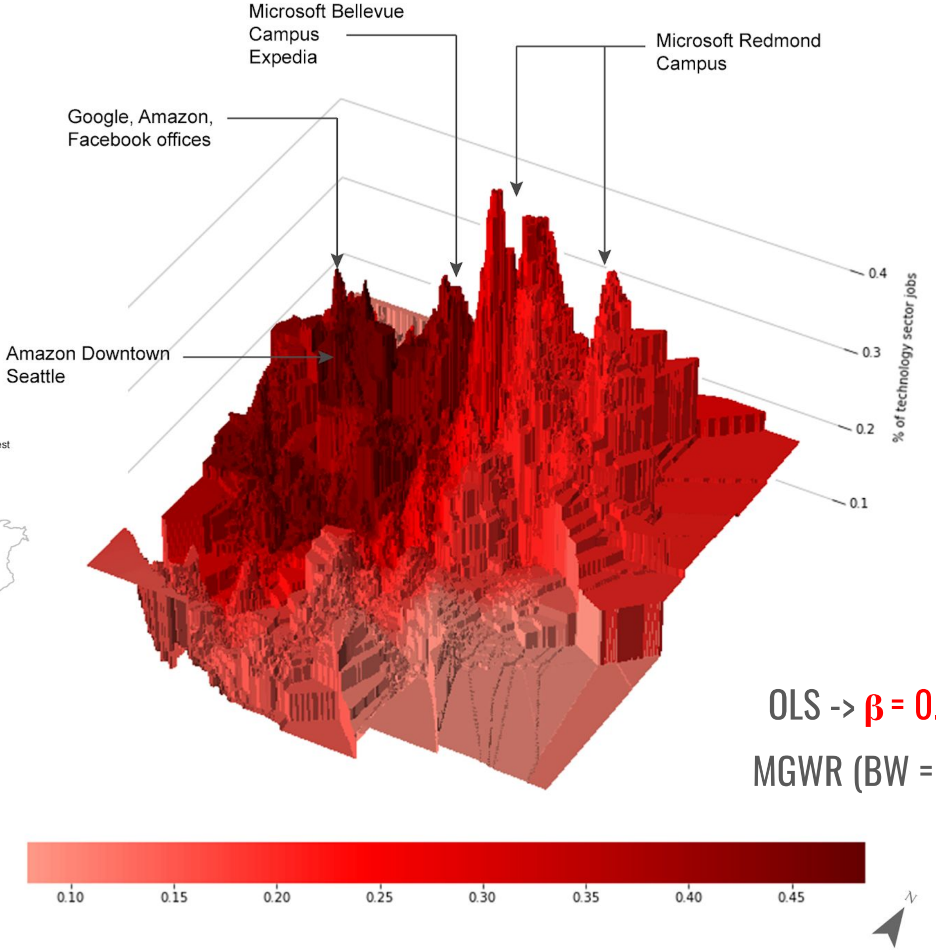


Hence, within neighborhoods of similar housing, newer houses are preferred over older ones.



Parameter Estimates: Technology sector jobs





OLS -> β = 0.42***

MGWR (BW = 1552)

Measuring intrinsic neighborhood value

$$yi = y\_mean + \boldsymbol{\alpha}_i \boldsymbol{\sigma}_y + \boldsymbol{\sigma}_y \left(\Sigma_{ij} \boldsymbol{\beta}_{ij} \left(x_{ij} - x_j\_mean\right)\right) / \boldsymbol{\sigma}_x$$

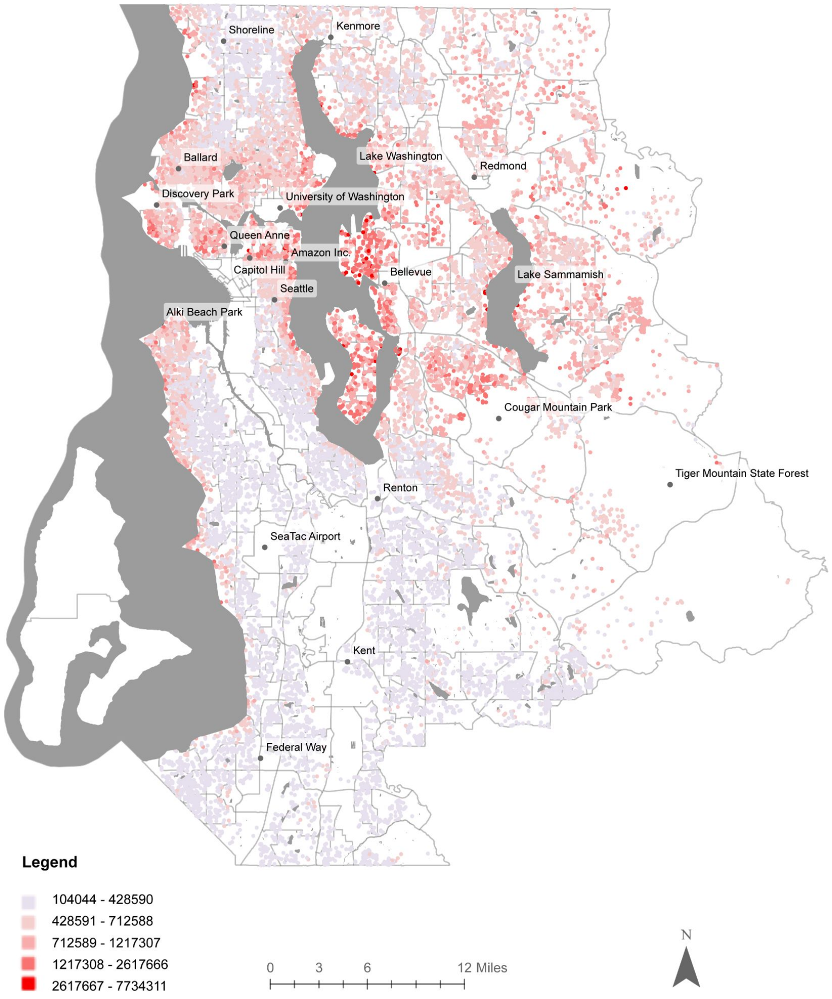
Predicted house prices

Base house prices

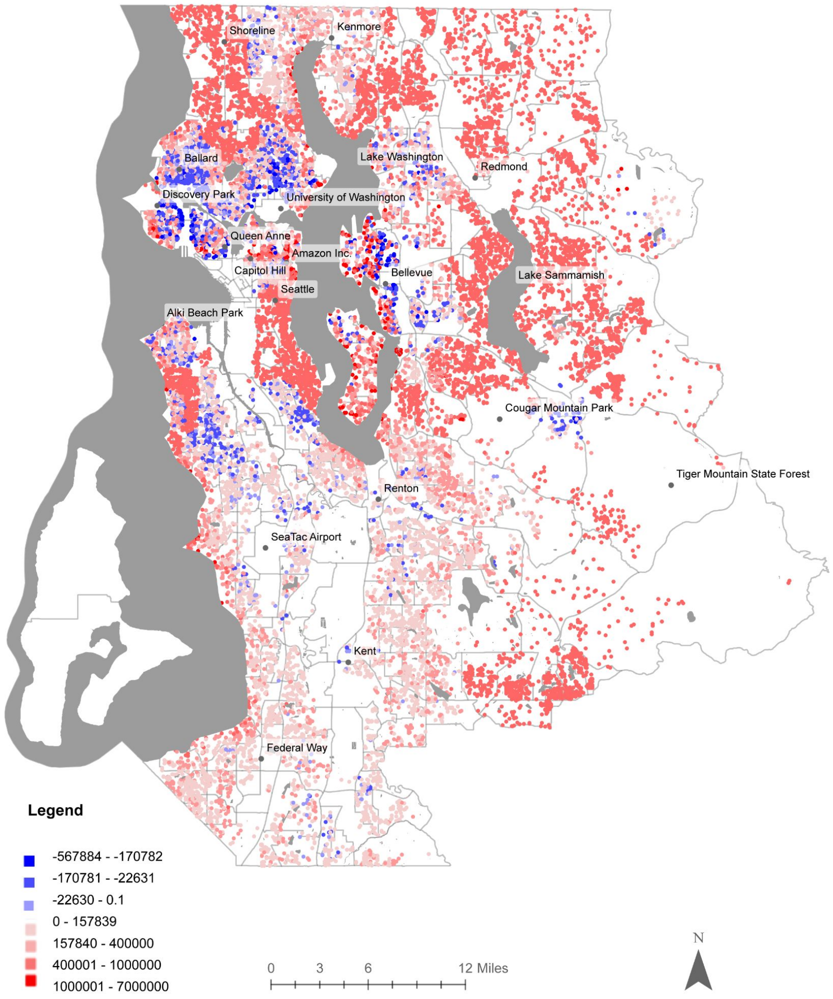
Intrinsic location effect

House prices explained through structural and neighborhood attributes

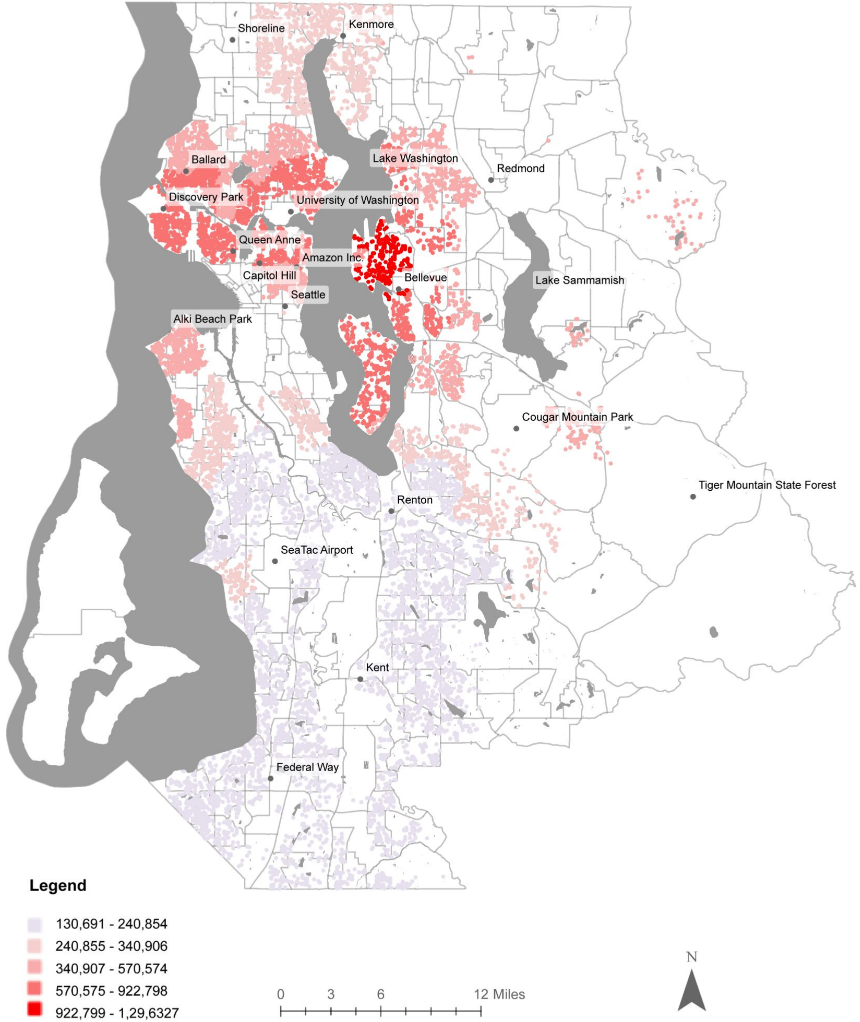
Predicted house prices



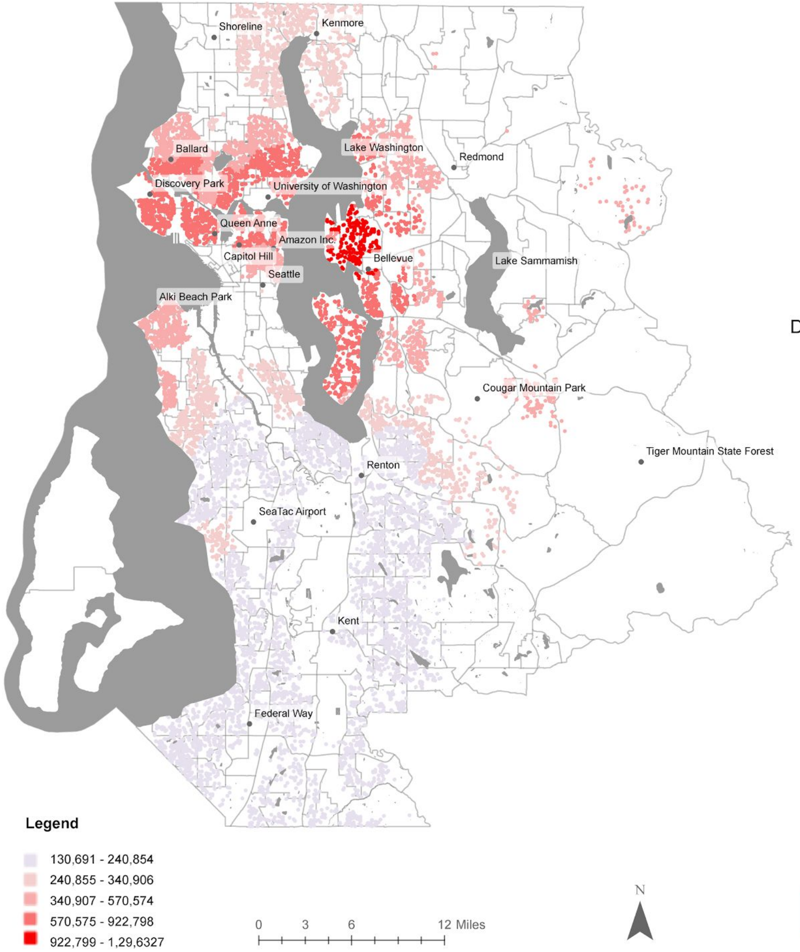
House prices explained through structural and neighborhood attributes

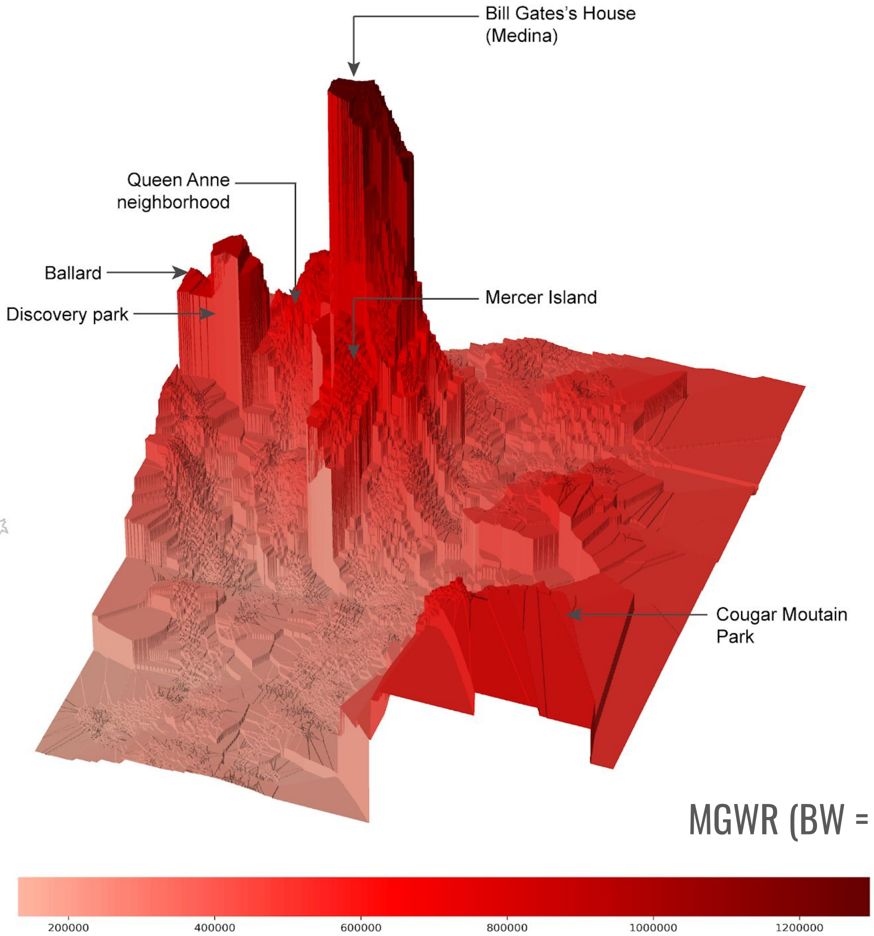


Intrinsic location effect



Parameter Estimates: Intrinsic location value

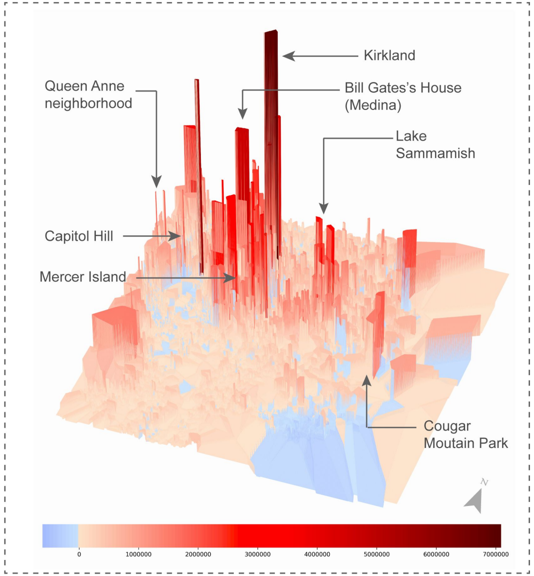


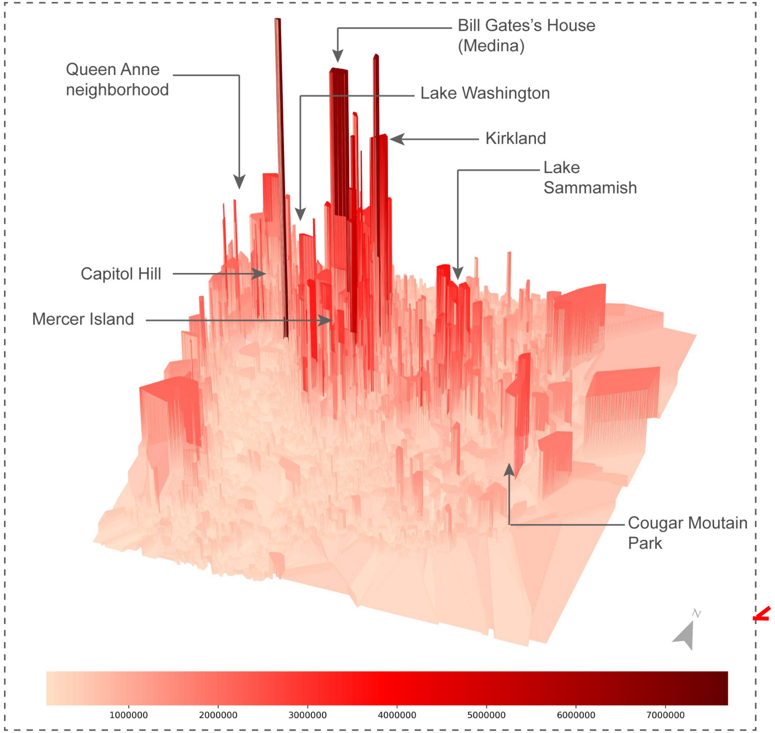


MGWR (BW = 463)

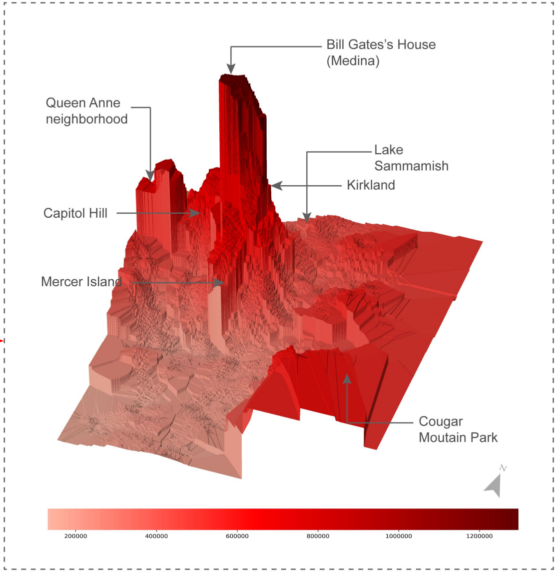
Conclusion: Measuring context using MGWR

Component associated with measurable structural and neighborhood attributes of a house





House Price



Component associated with intrinsic locational preferences

References for detailed description of MGWR:

1. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression : the analysis of spatially varying relationships*. Wiley.

2. Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

3. Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Stewart Fotheringham, A. (2019). MGWR: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, *8*(6). https://doi.org/10.3390/ijgi8060269

4. Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2019). Inference in Multiscale Geographically Weighted Regression. *Geographical Analysis*, gean.12189. https://doi.org/10.1111/gean.12189

Thank you!