Part II

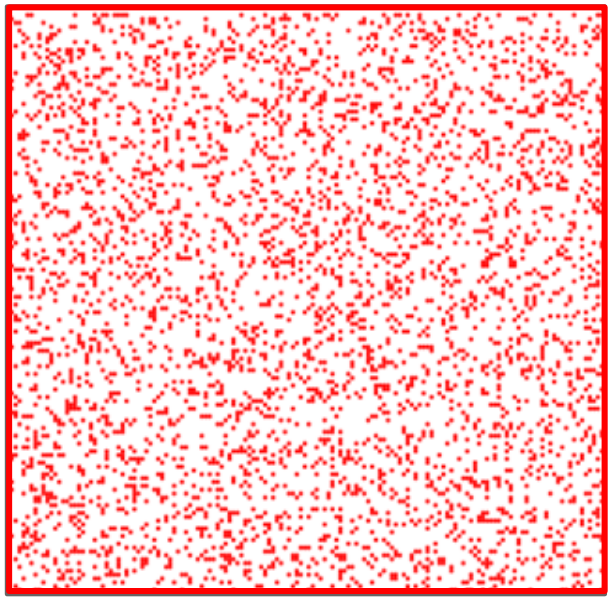
Local Modeling Perspectives on Scale Issues

Modifiable Areal Unit Problem

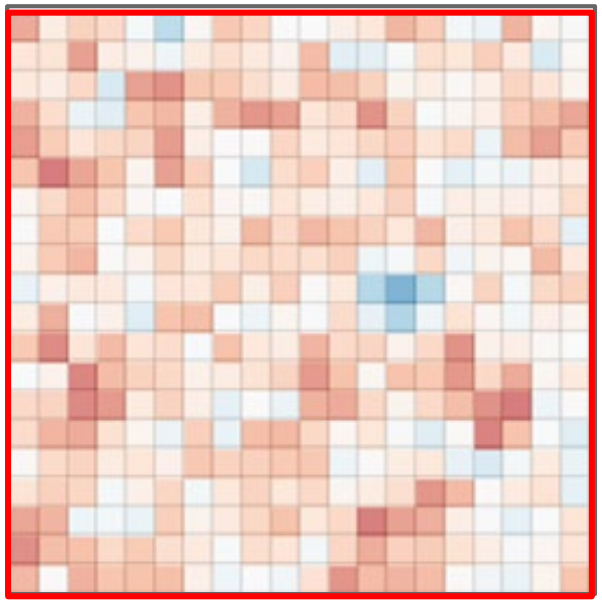
Publication : Fotheringham, A. S., & Sachdeva, M. (2022). Scale and local modeling: New perspectives on the modifiable areal unit problem and Simpson's paradox. ***Journal of Geographical Systems***. https://doi.org/10.1007/s10109-021-00371-5

Background and Problem Statement

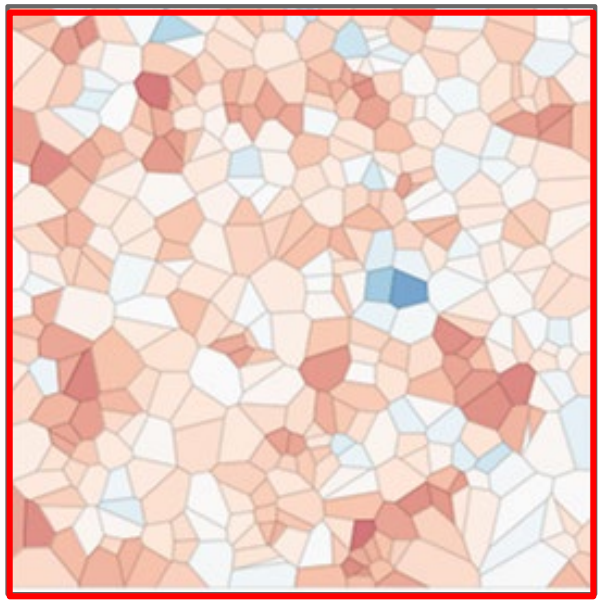
Classic approach to explore MAUP - *Data view*



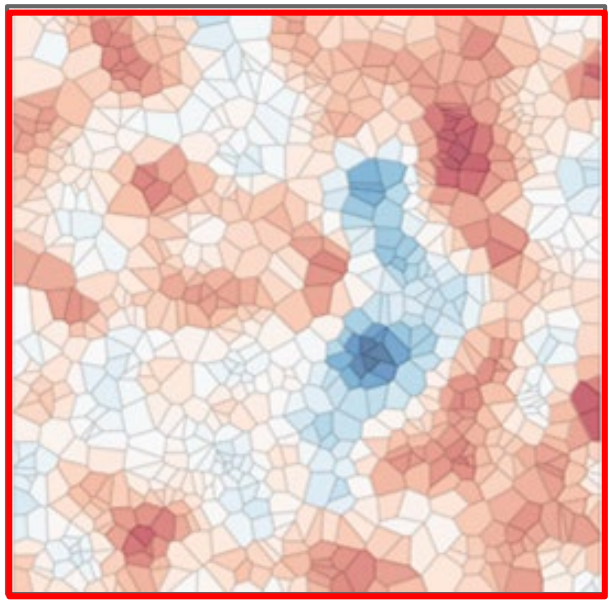
Data at the most disaggregated level



Data aggregated - Type I



Data aggregated - Type II

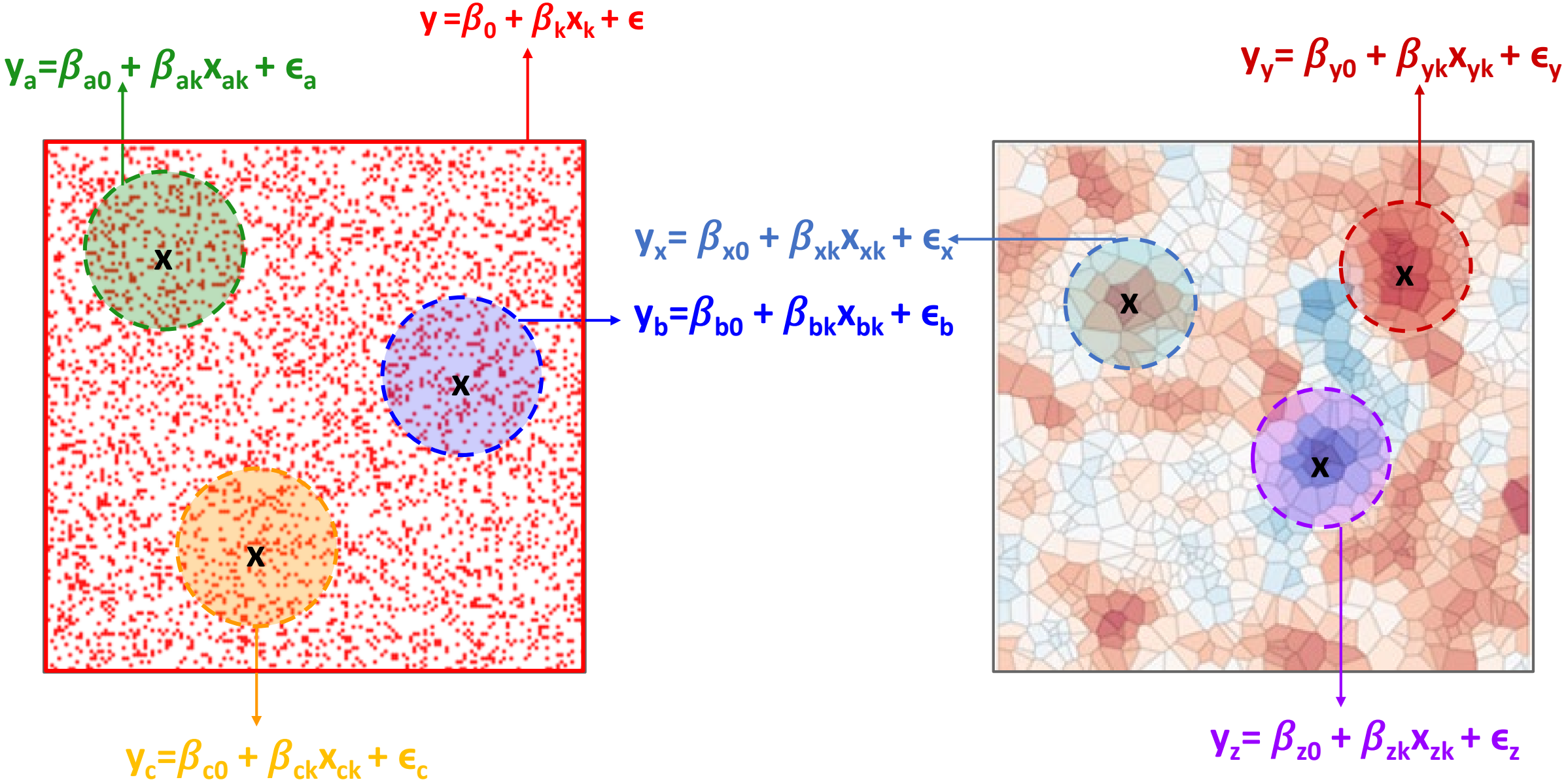


Data aggregated - Type III

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

Background and Problem Statement

A new approach to explore MAUP - *Process view*



Research Objective

What is the focus of the research?

- Can the **focus on processes help understand the classic MAUP?**

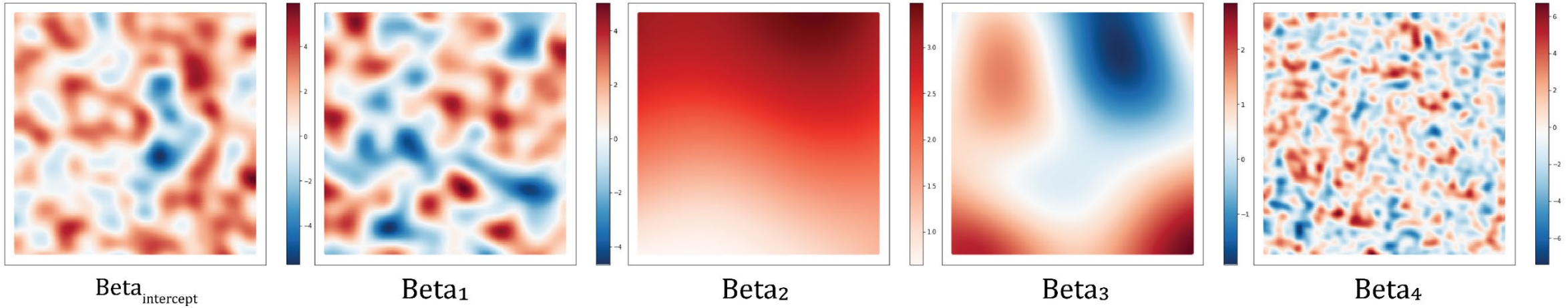
- Does **process heterogeneity affect sensitivity to the MAUP?**

  - Processes which are relatively stable across space, should exhibit less sensitivity to the MAUP than processes which vary greatly.

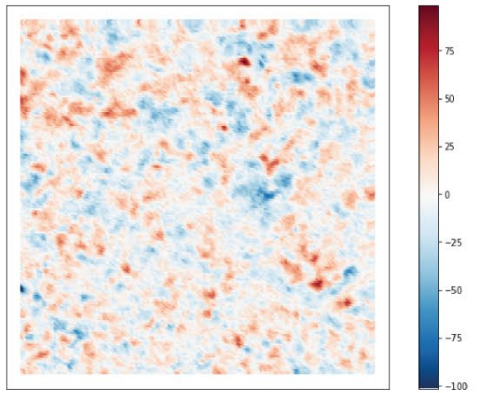
Research Design

Experiment 1: Mix of local and global processes

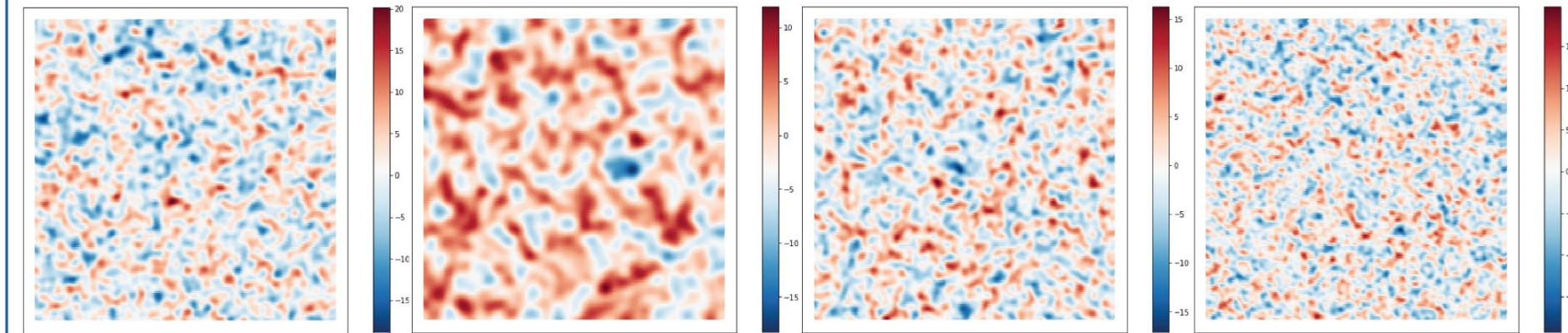
Simulated Processes



Response variable



Covariates

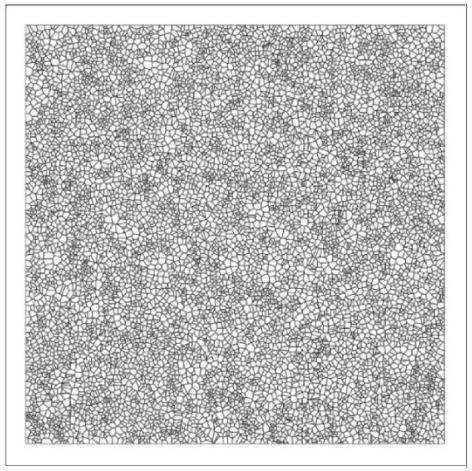


Research Design

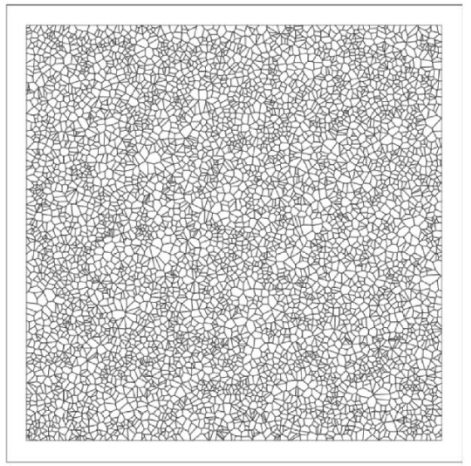
Experiment 1 with irregular grid aggregations

- Run a local and global model each for different grid aggregations
- Highlight sensitivity of results from global models to MAUP
- Investigate whether process heterogeneity affects sensitivity to the MAUP

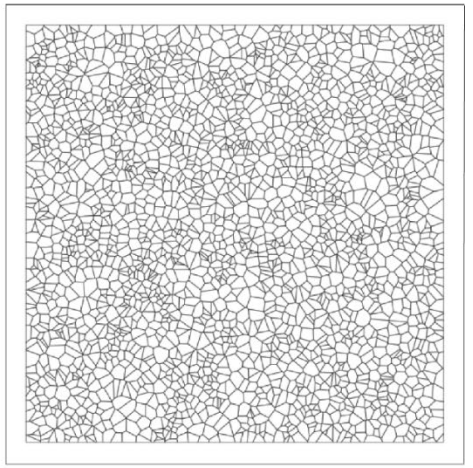
Irregular grids with different scales



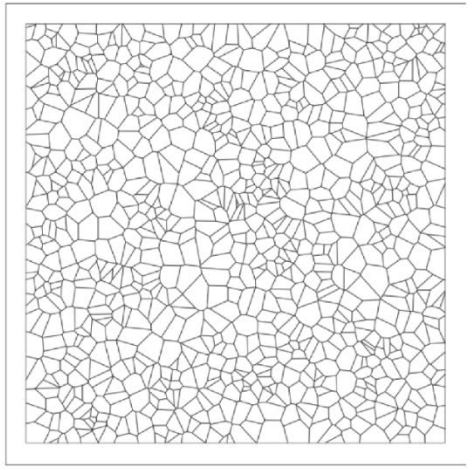
10000 random voronoi



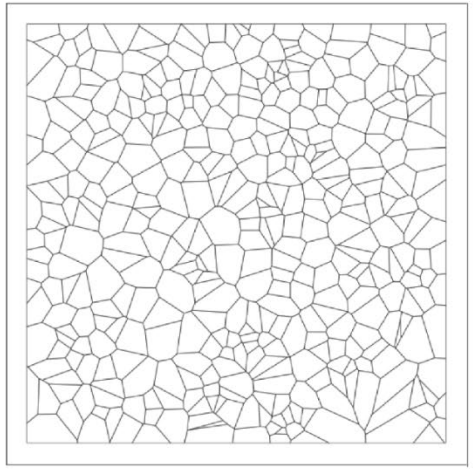
5625 random voronoi



2500 random voronoi

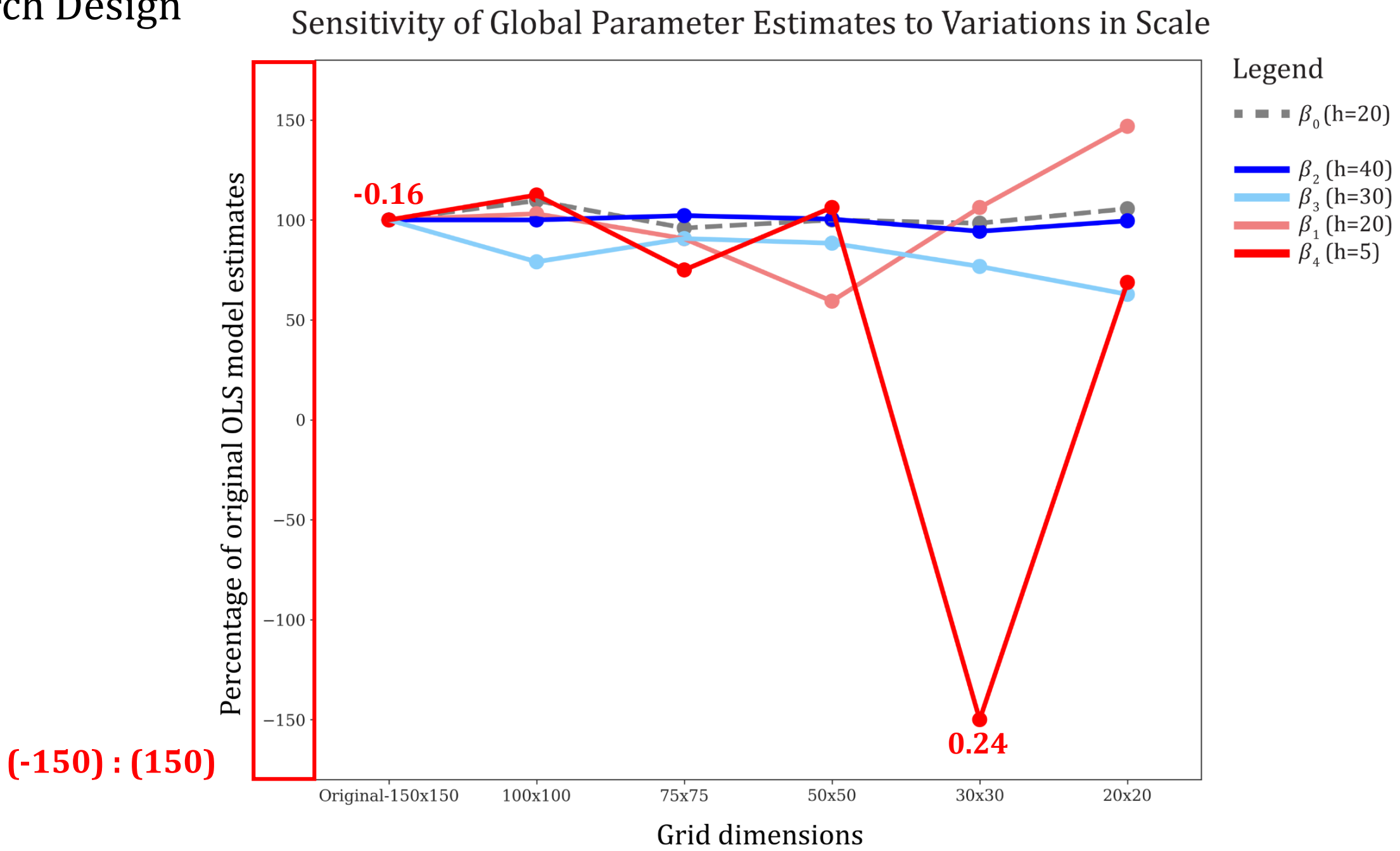


900 random voronoi



400 random voronoi

Research Design

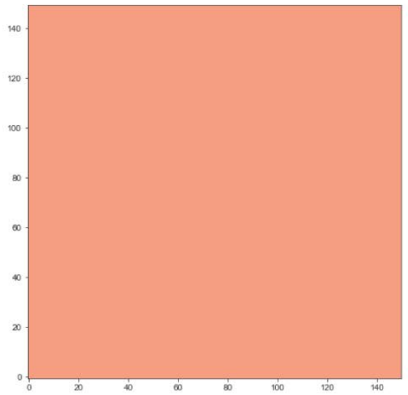


(-150) : (150)

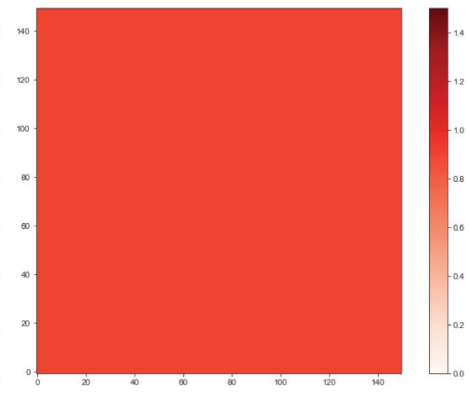
Research Design

Experiment 2: All global processes

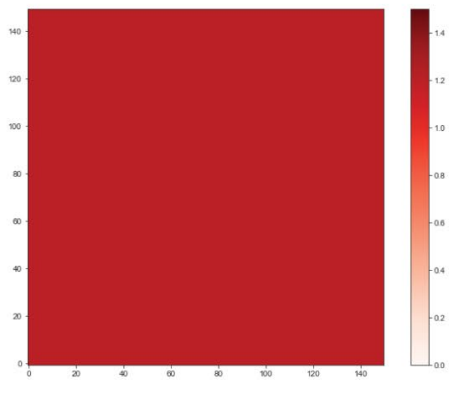
Simulated Processes



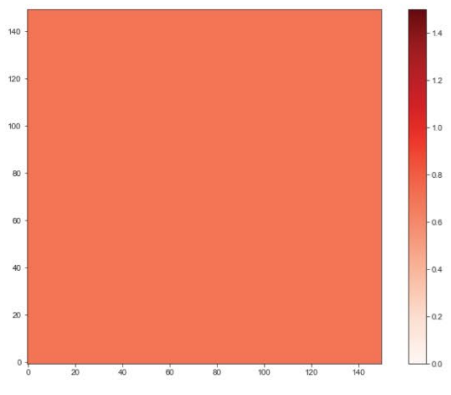
$Beta_{intercept}$



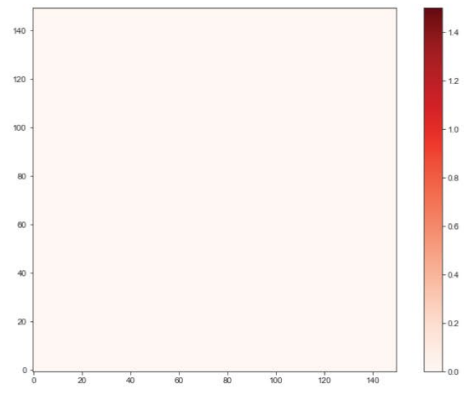
$Beta_1$



$Beta_2$

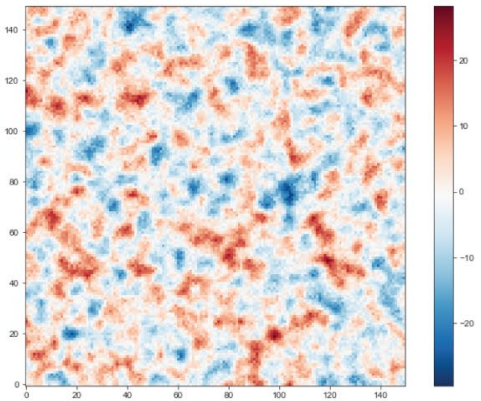


$Beta_3$



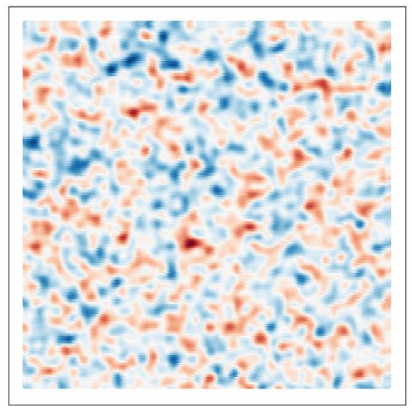
$Beta_4$

Response variable

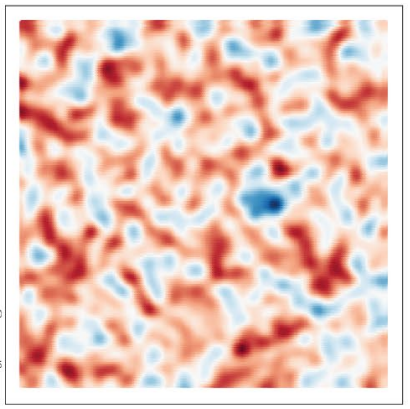


y

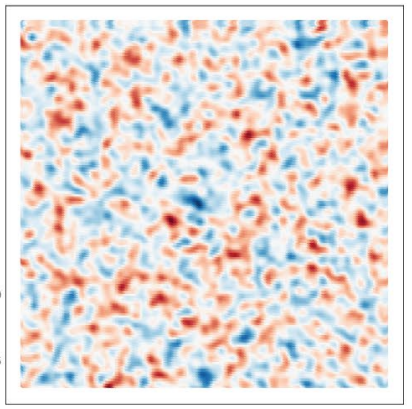
Covariates



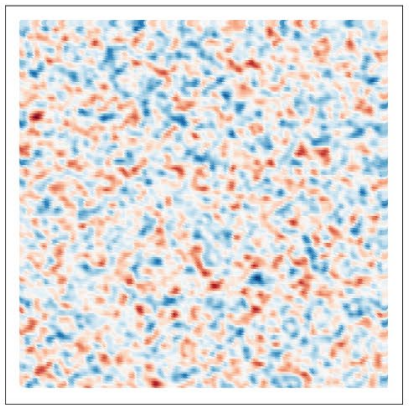
x_1



x_2



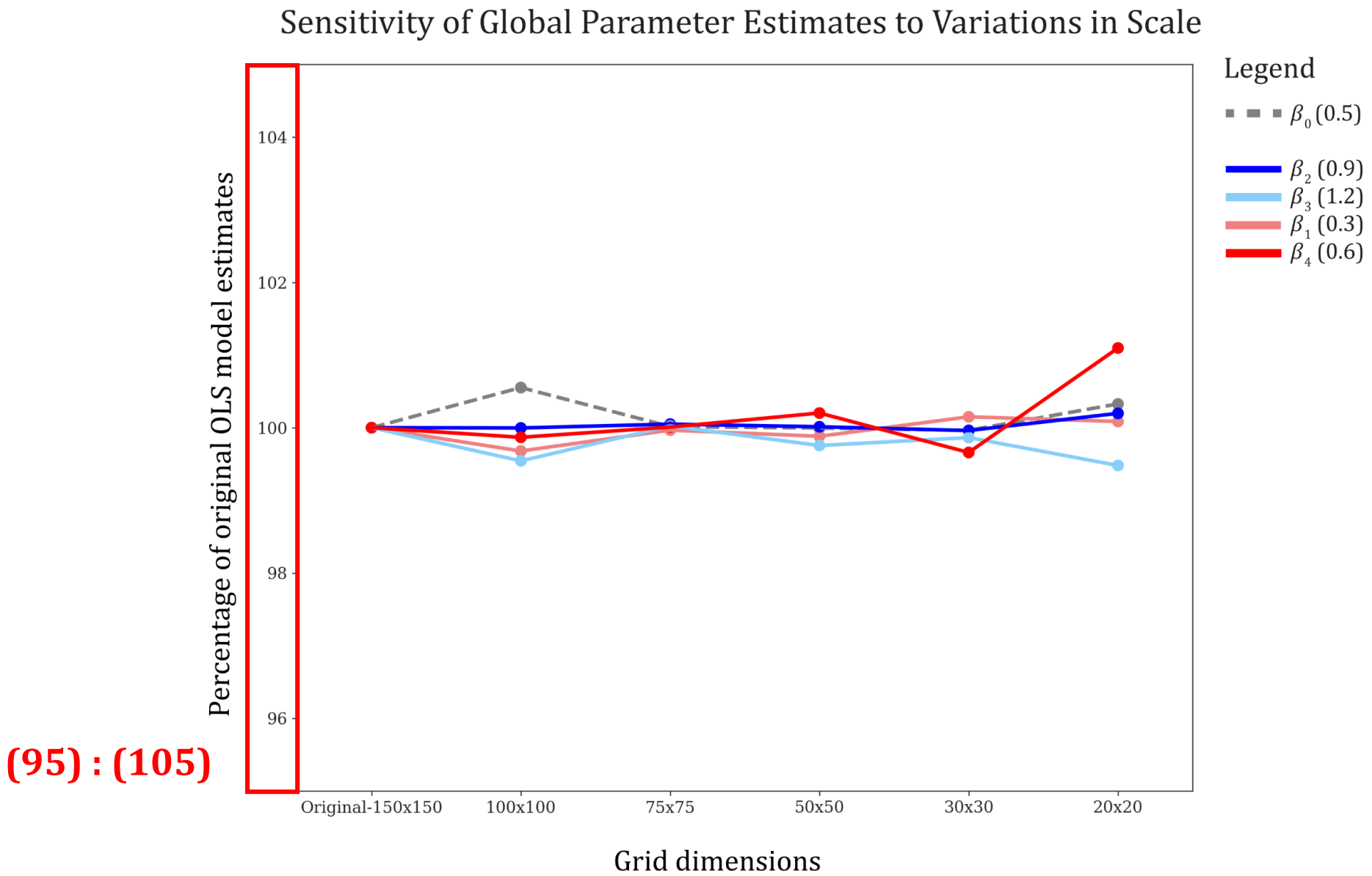
x_3



x_4

Research Design

Experiment 2 with regular grid aggregations



Intellectual merits and Broader impacts

- An examination of the sensitivity of global and local models to the MAUP suggests that **the effects of MAUP in global models are a function of the degree to which processes vary over space.**

- This generates a new insight into the MAUP: **it results from the properties of *processes* rather than the properties of *data*.**

- The observations, while important in **understanding the operational scale of processes within geographical sciences**, are also **expected to inform analysis in various applied fields using aggregated spatial data of any kind.**

Part II
Local Modeling Perspectives on Scale Issues

Simpson's Paradox

Submitted Manuscript : Sachdeva, M., & Fotheringham, A. S. (*under review*). A Geographical Perspective on Simpson's Paradox. ***Journal of Spatial Information Science***.

Background

What is Simpson's Paradox?

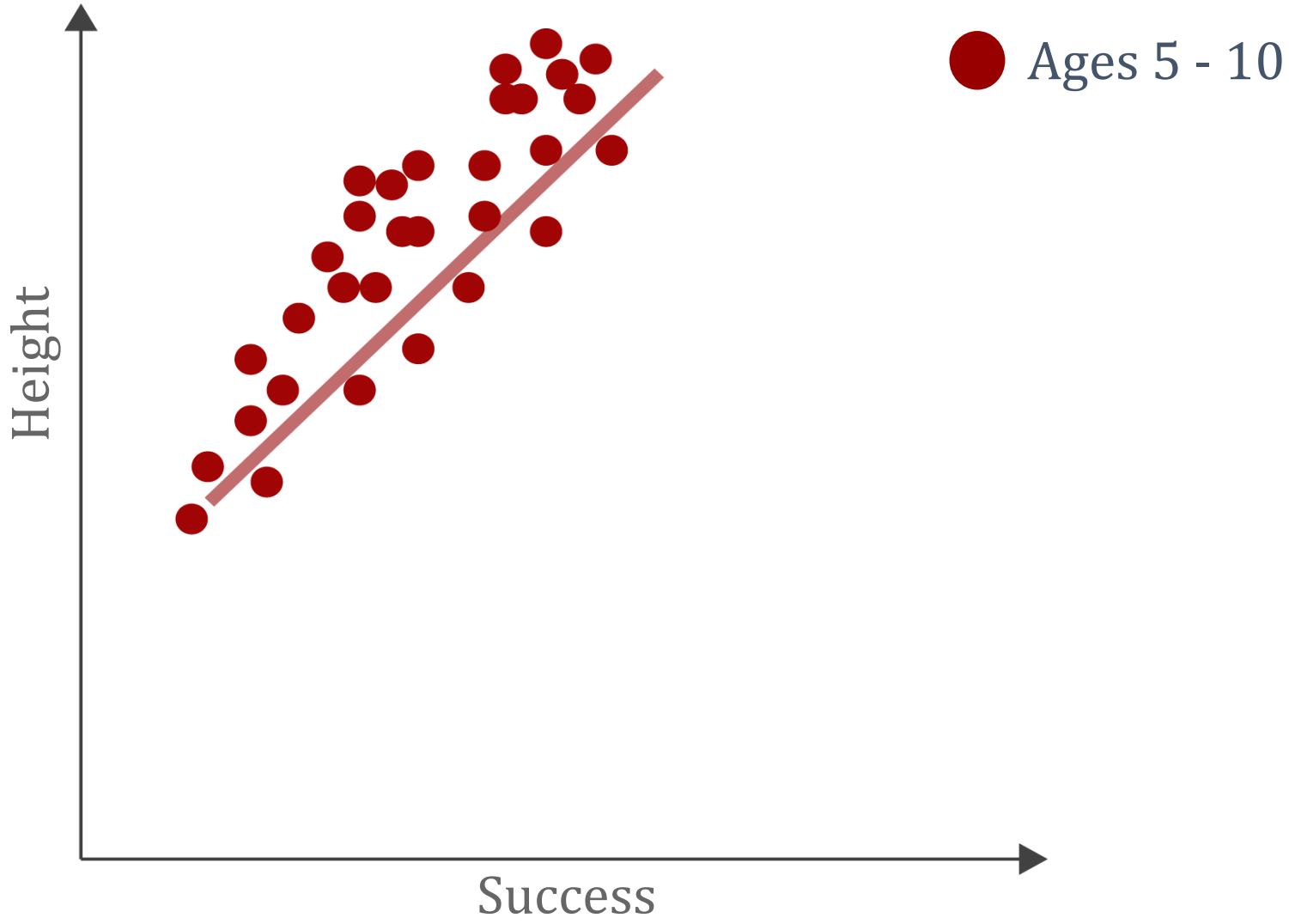
- A trend appears in aggregated data but disappears or reverses when the data are disaggregated

- Reported chiefly in social-science and medical-science applications

- Major implications for causal relations and their interpretations

Background

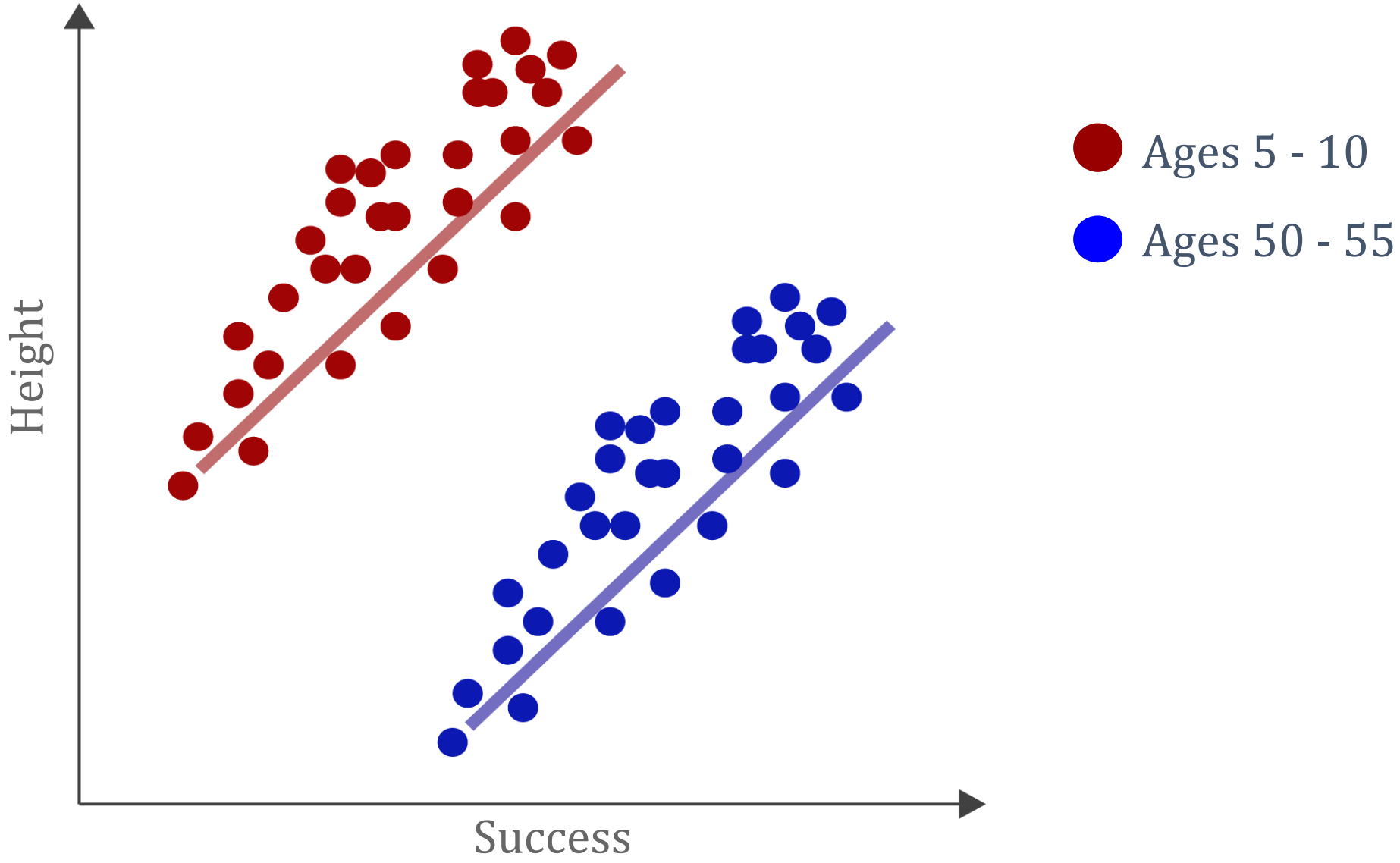
Example: Correlation between height and basketball success



Inference: Height **positively** affects basketball success

Background

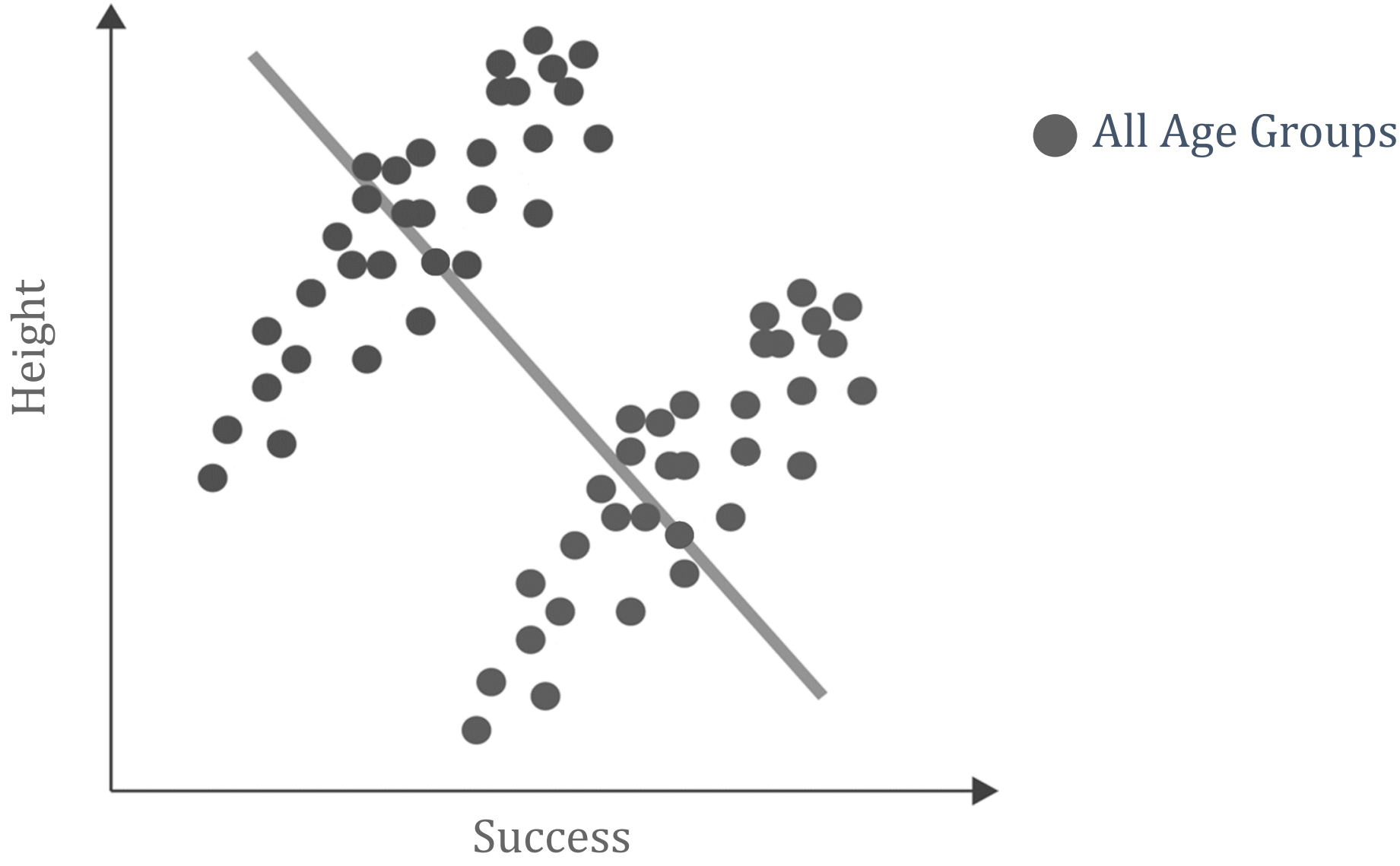
Example: Correlation between height and basketball success



Inference: Height **positively** affects basketball success

Background

Example: Correlation between height and basketball success



Inference: Height **negatively** affects basketball success

Background

A Geographical Perspective on Simpson's Paradox

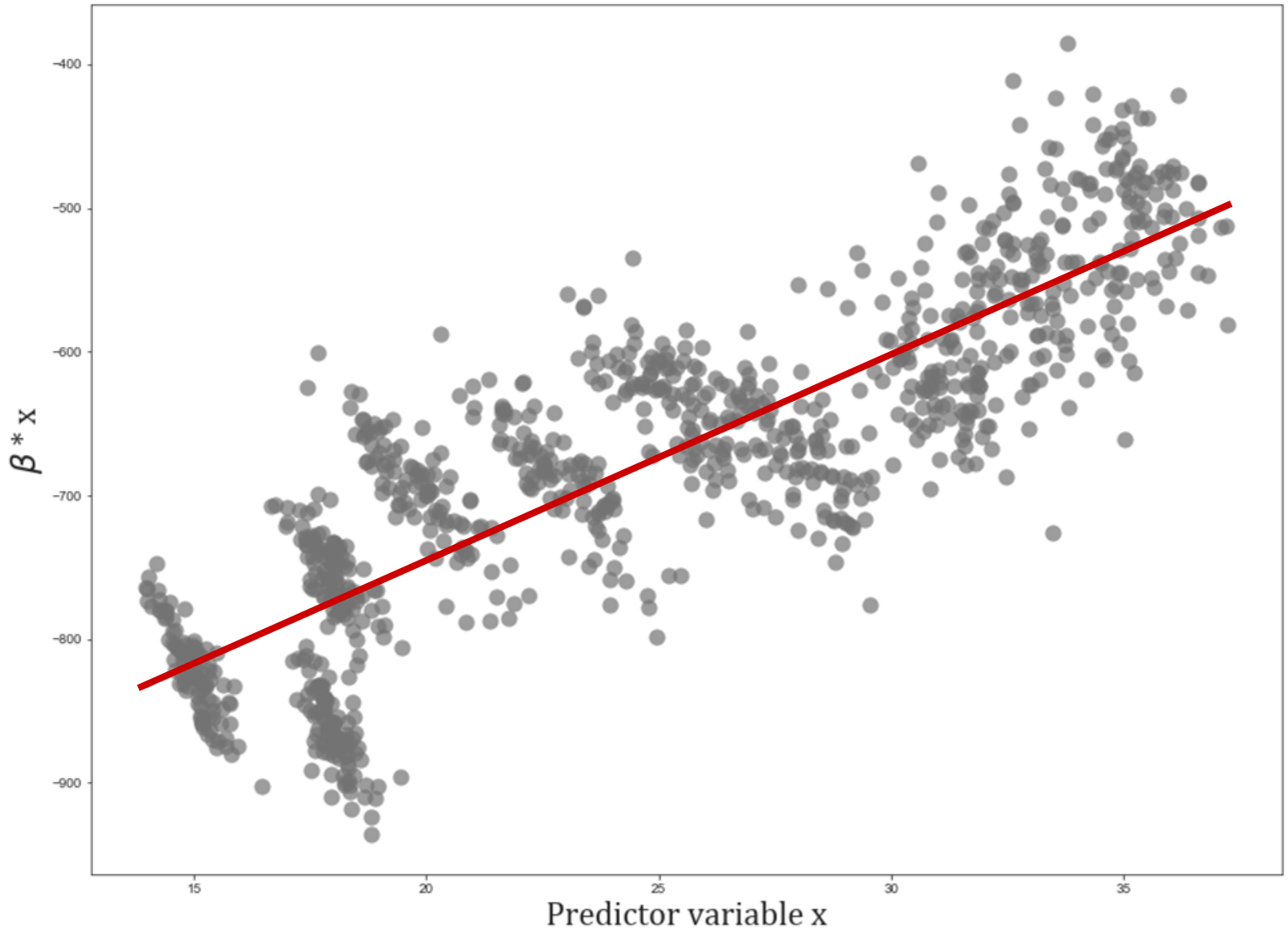
- While such aspatial examples are well-known, the presence of Simpson's Paradox in spatial analysis has largely been ignored

- Mathematically, Simpson's paradox is observed when the partitioning criterion used to stratify the population into groups is correlated with both the predictor and the response variables

- It can be imagined that *space* is correlated with both the response and predictor variables – Hence, omitting *space* could lead to a case of Simpson's paradox in spatial analysis.

Background

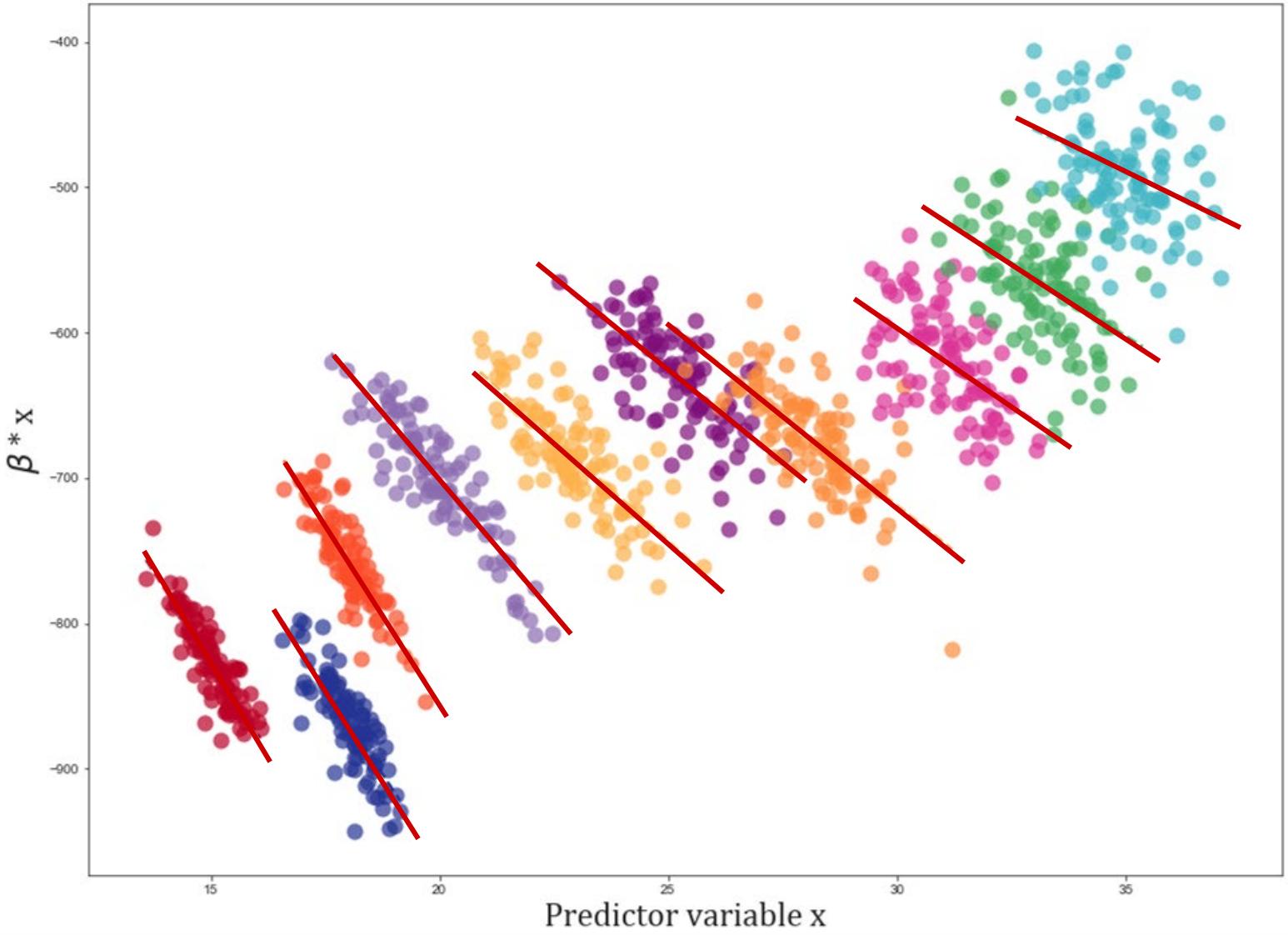
A Geographical Perspective on Simpson's Paradox



Inference: *x* **positively** affects *y*

Background

A Geographical Perspective on Simpson's Paradox



Inference: *x* **negatively** affects *y*

Problem Statement

Why is Simpson's Paradox a problem?

- Global models are calibrated using aggregated data

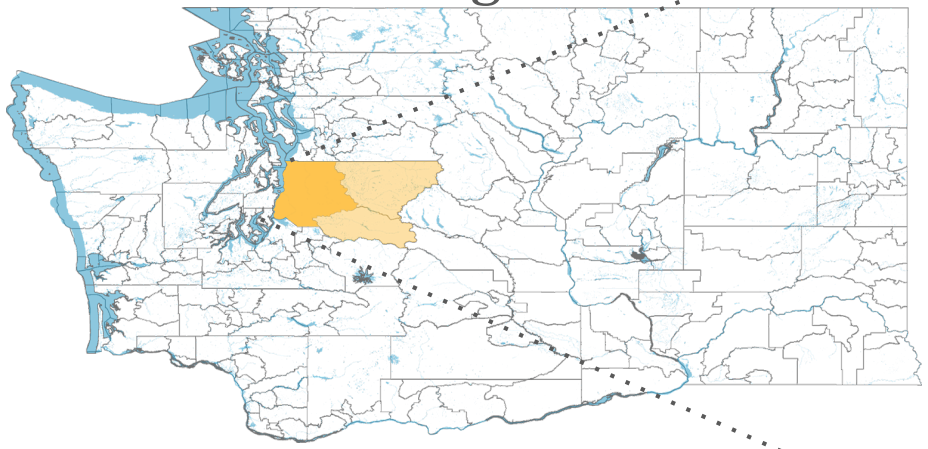
- Local models are calibrated using disaggregated data

- With the popularity of local models, this abstract problem is now a real concern since a comparison between local and global models can lead to different inferences

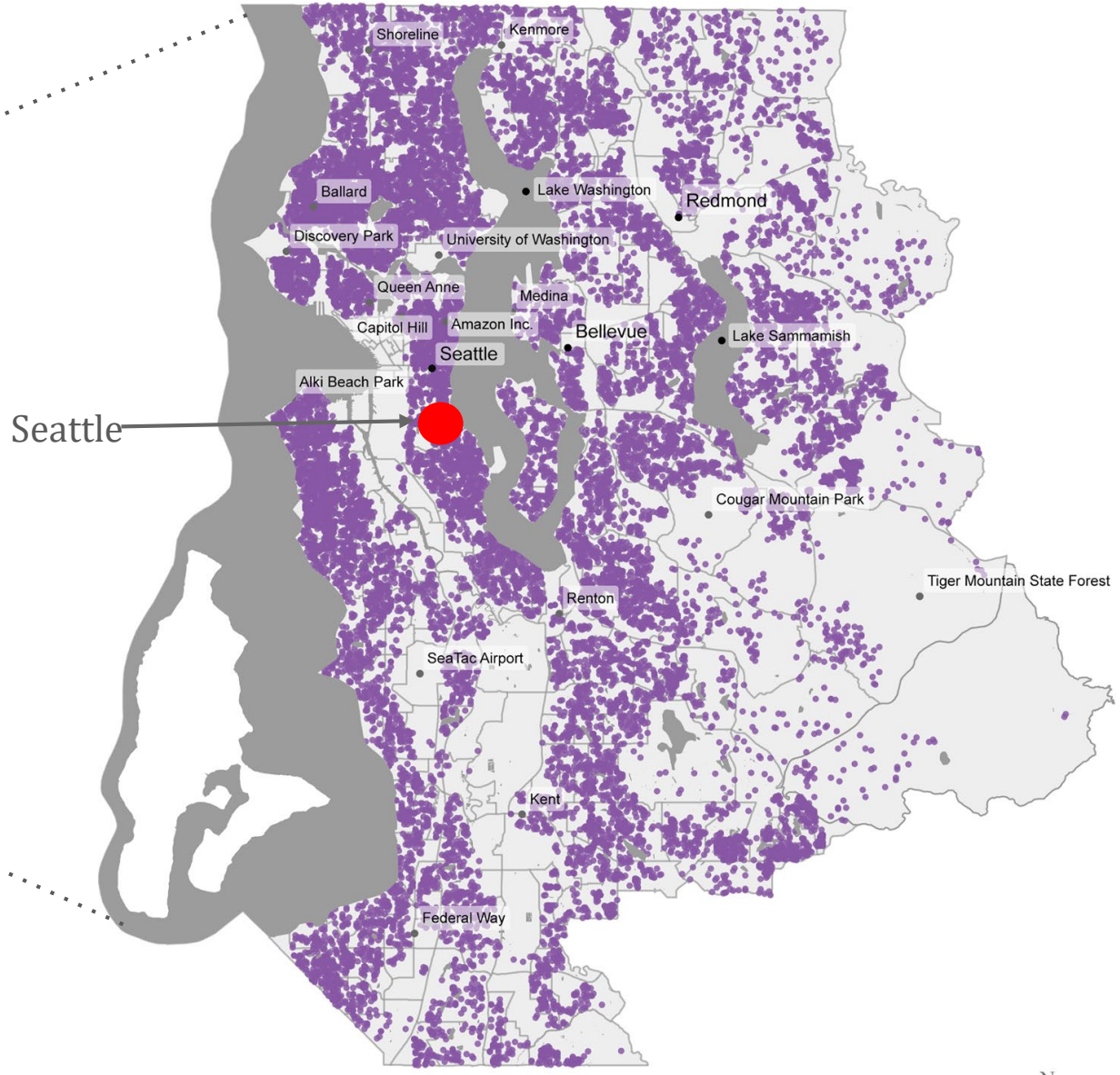
Research Design

Empirical example: Location

Washington State



King County, WA

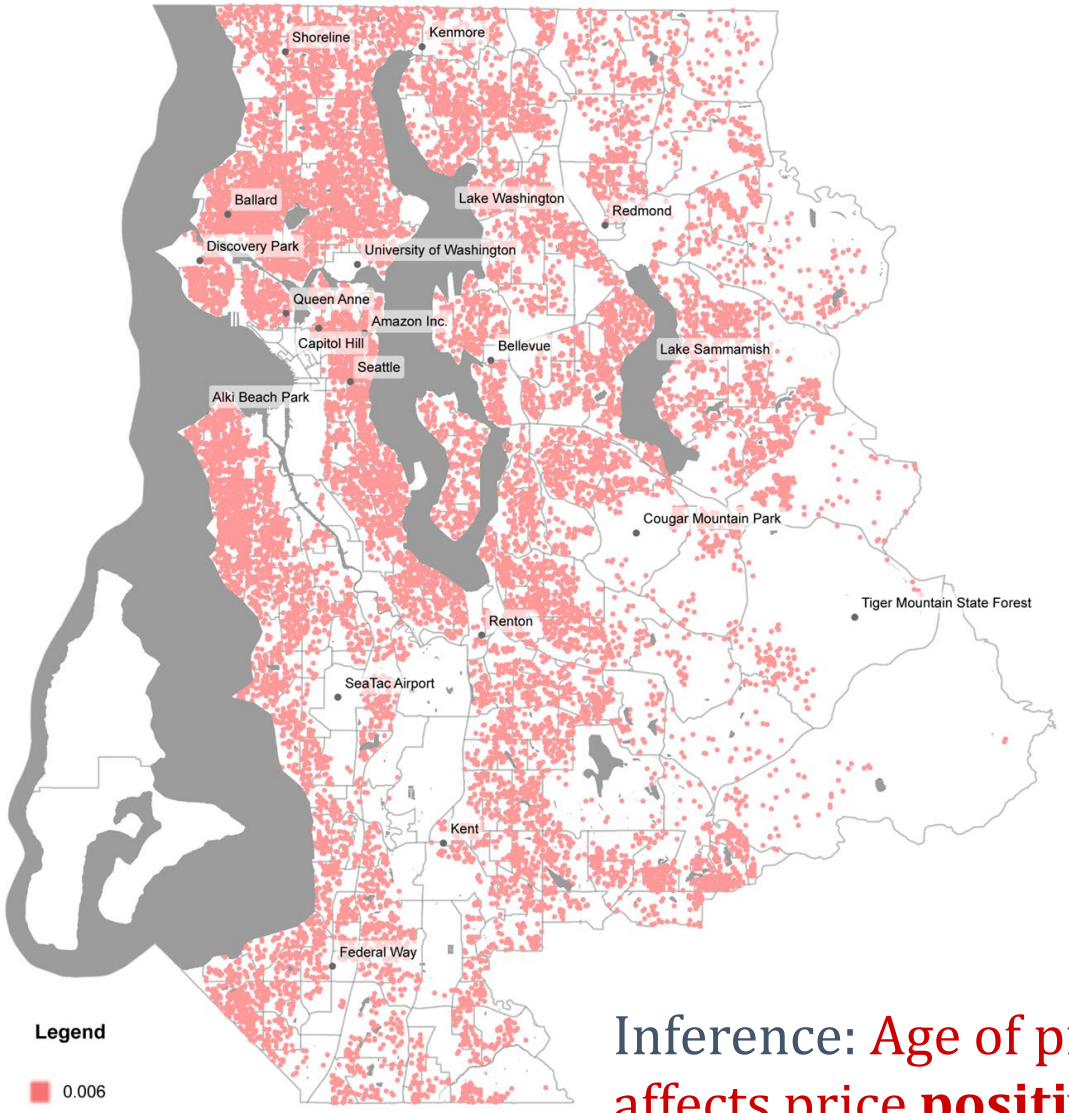


Area : 2,116 sq mi (5,480 km^2) | Population Density : 1,034/sq mi

Mehak Sachdeva, Stewart Fotheringham & Ziqi Li (2022) Do Places Have Value?: Quantifying the Intrinsic Value of Housing Neighborhoods Using MGWR, Journal of Housing Research, DOI: 10.1080/10527001.2021.2003505

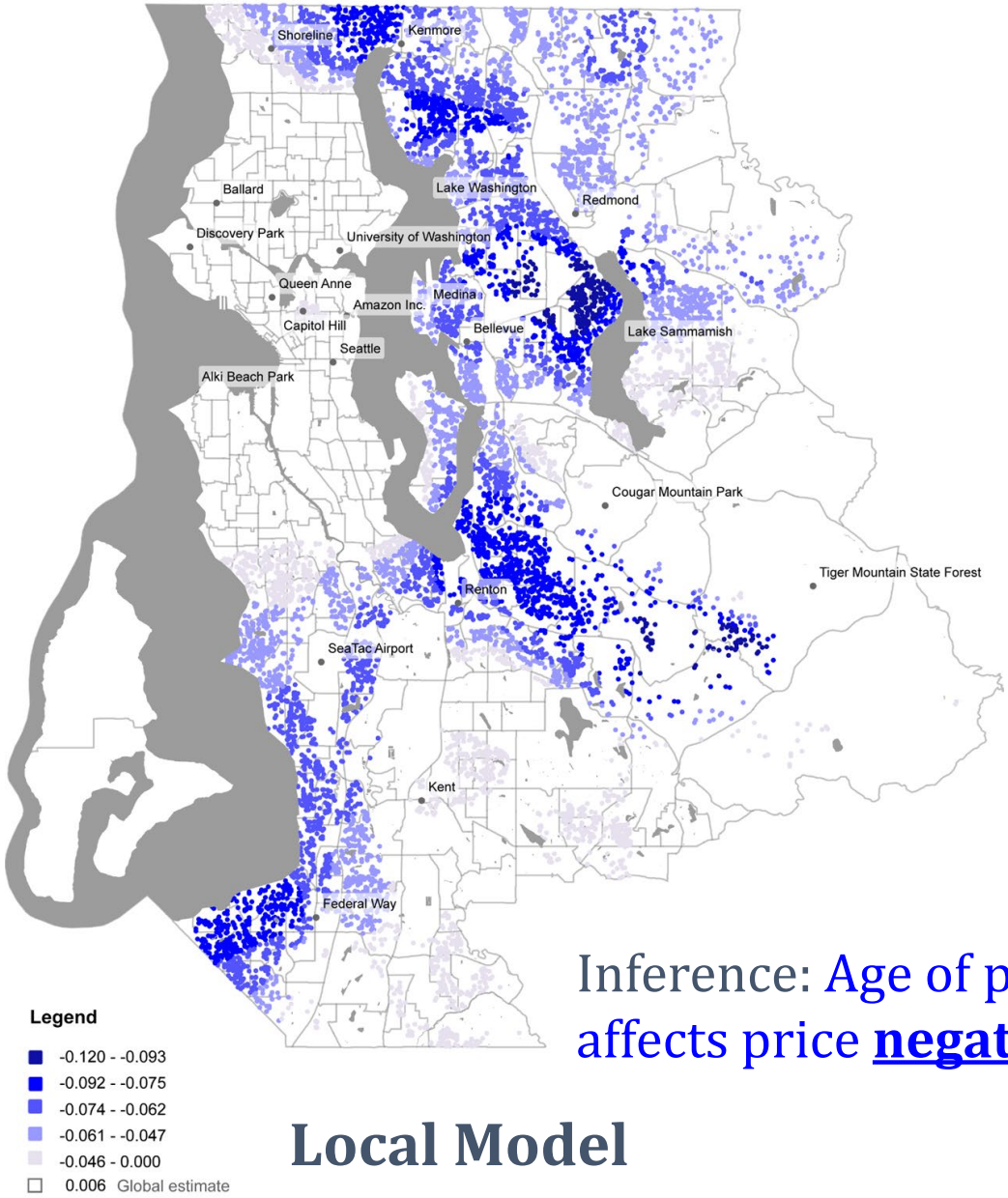
Research Design

MGWR Parameter Estimates: Age of the property



Inference: Age of property affects price **positively**

Global Model

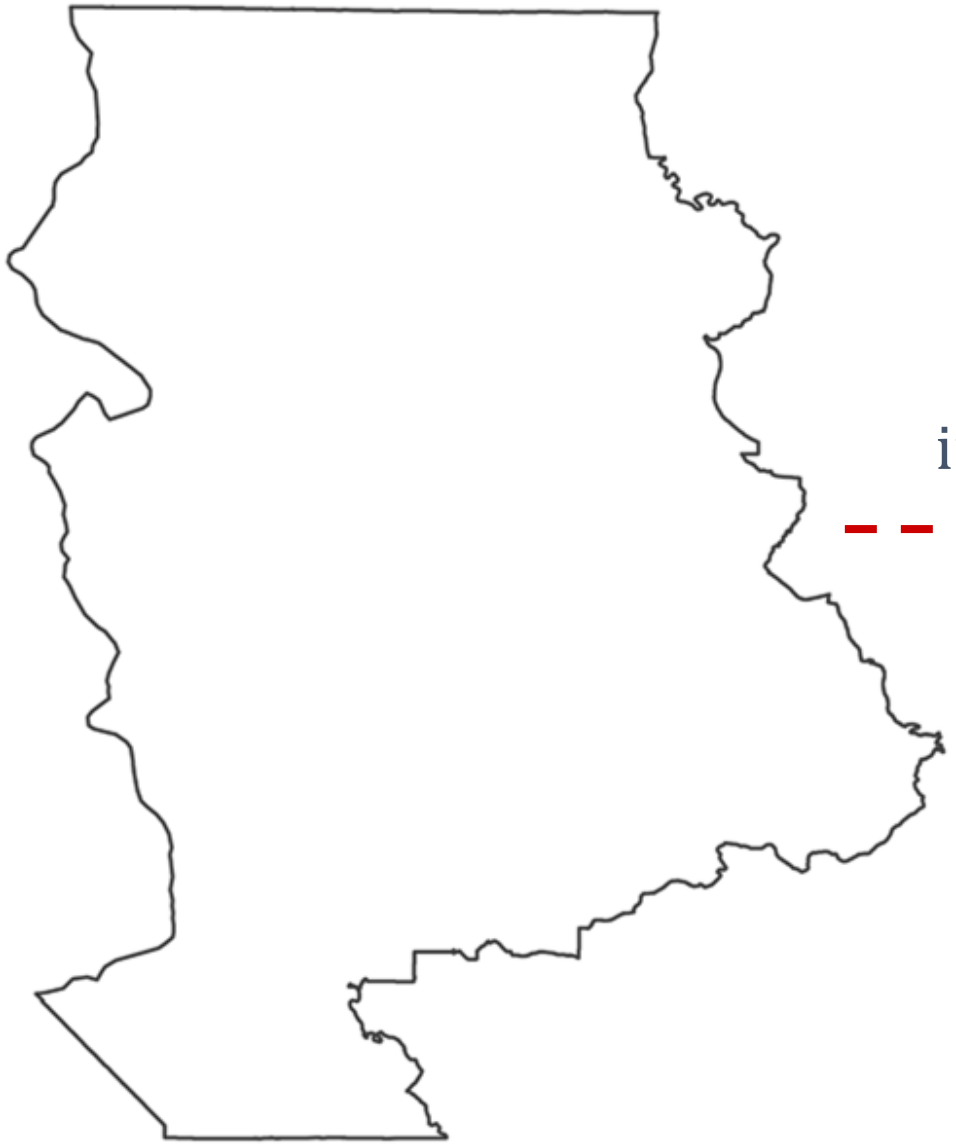


Inference: Age of property affects price **negatively**

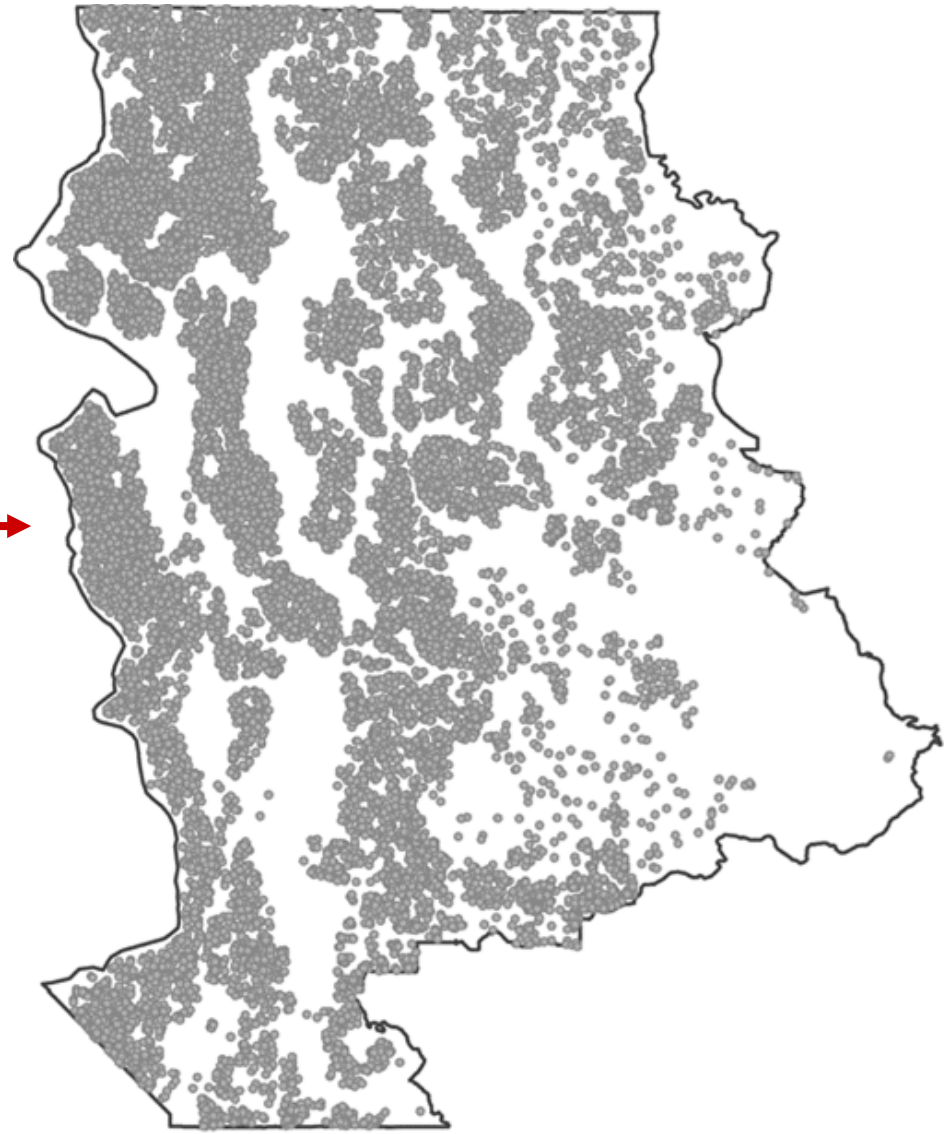
Local Model

Research Design

Further Exploration



At **what scale** does the inference change?



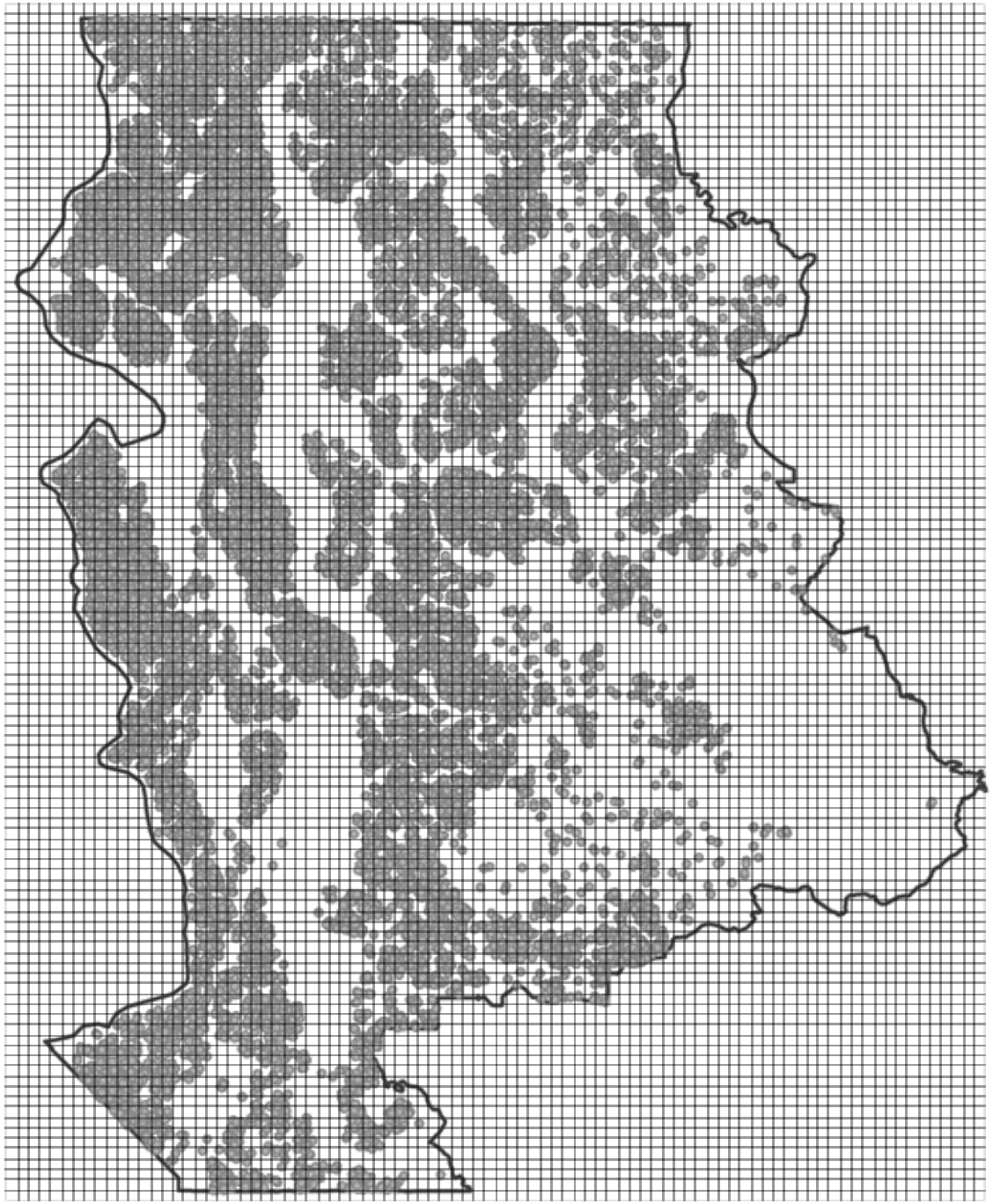
Global Scale

Local Scale

Research Design

Experiment

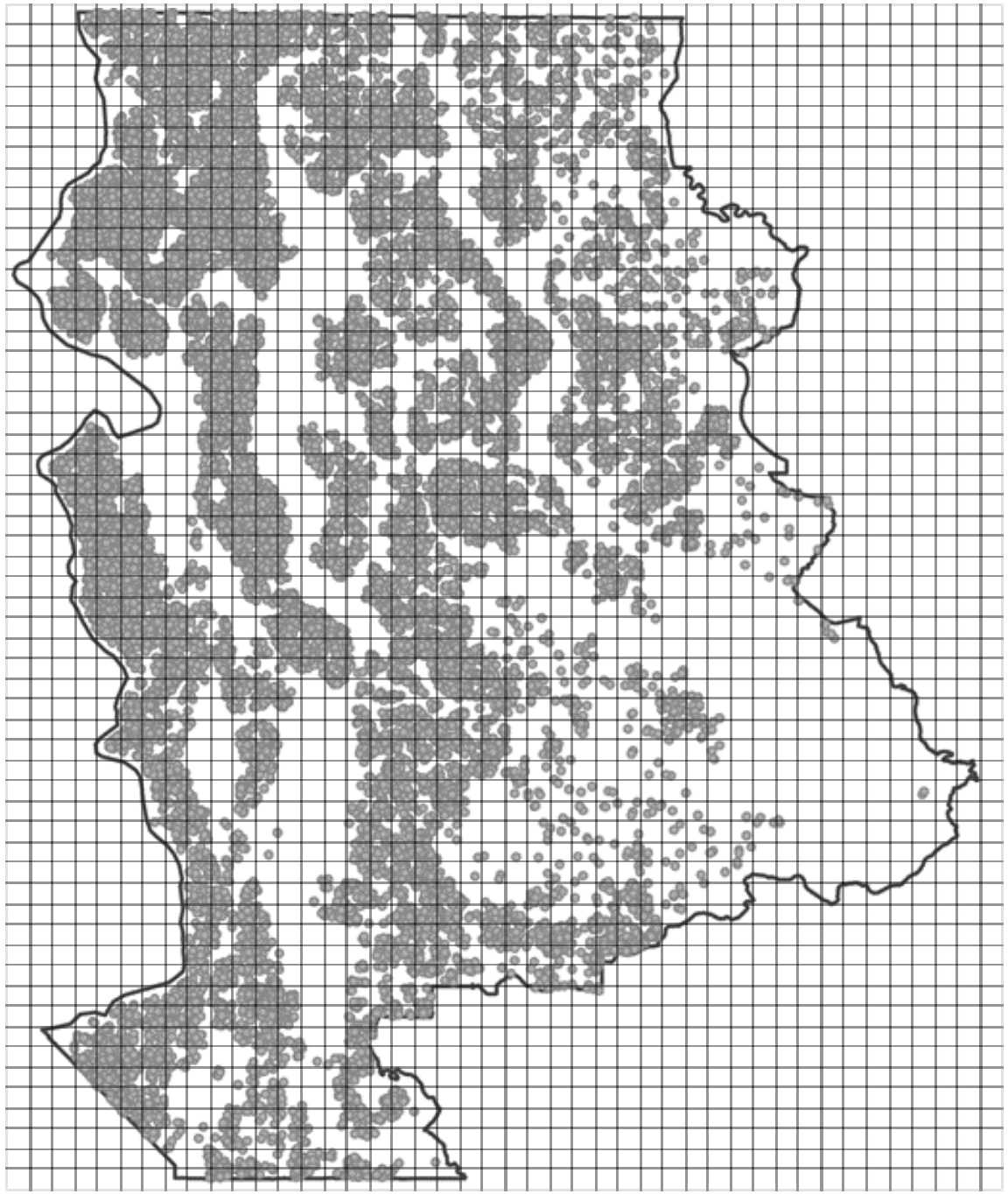
Aggregate the data at different grid levels and calibrate a local model at each grid scale



Research Design

Experiment

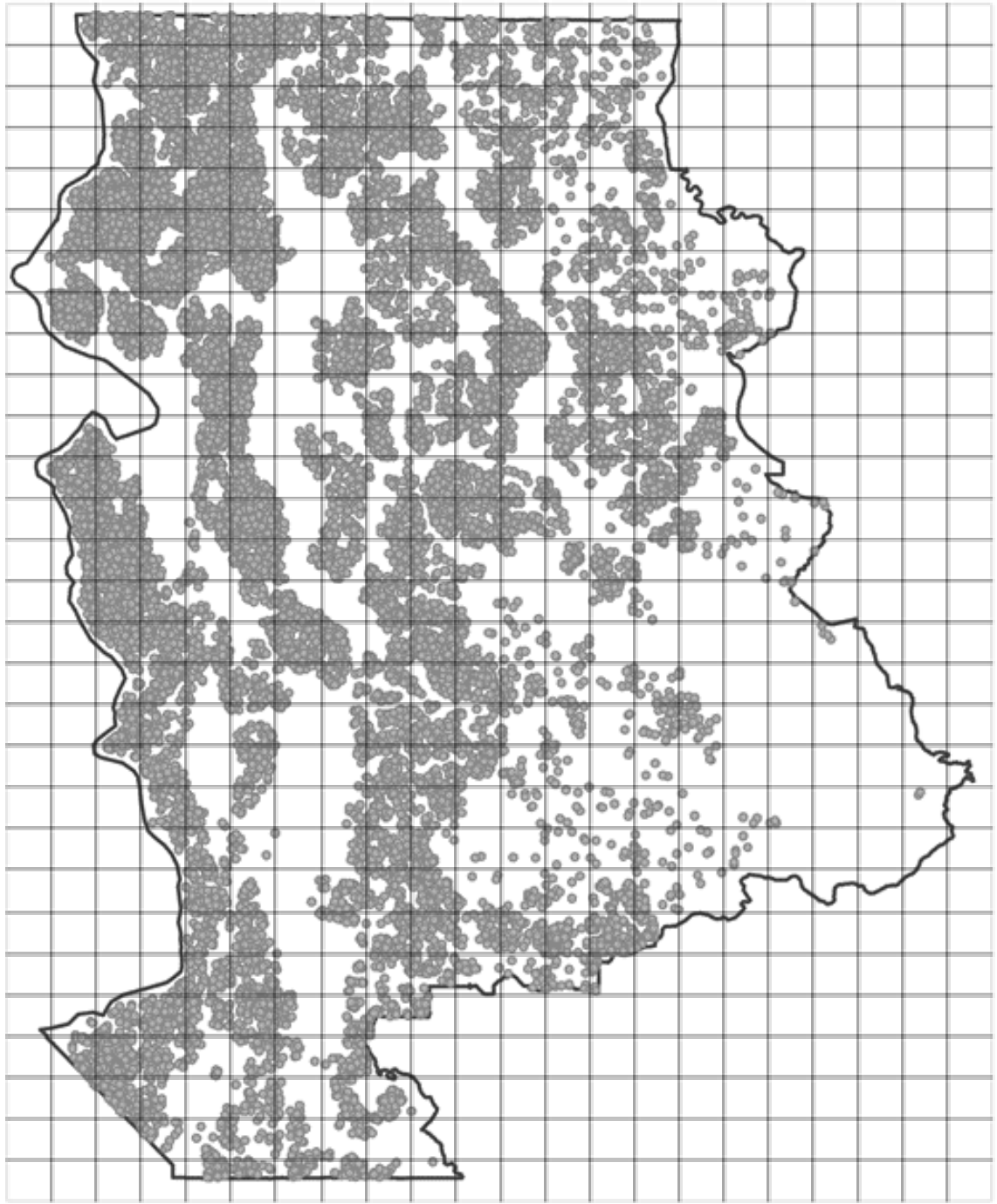
Aggregate the data at different grid levels and calibrate a local model at each grid scale



Research Design

Experiment

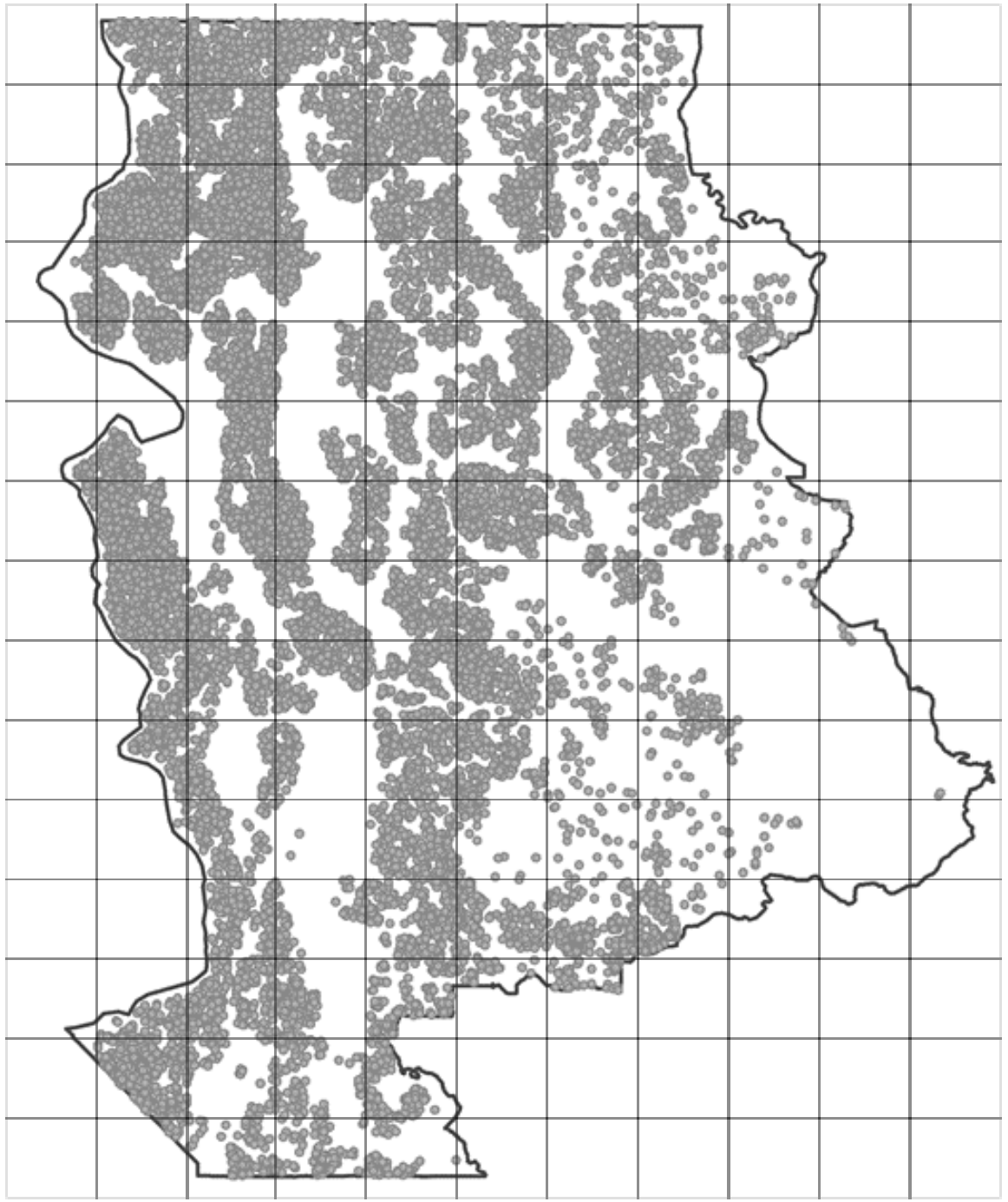
Aggregate the data at different grid levels and calibrate a local model at each grid scale



Research Design

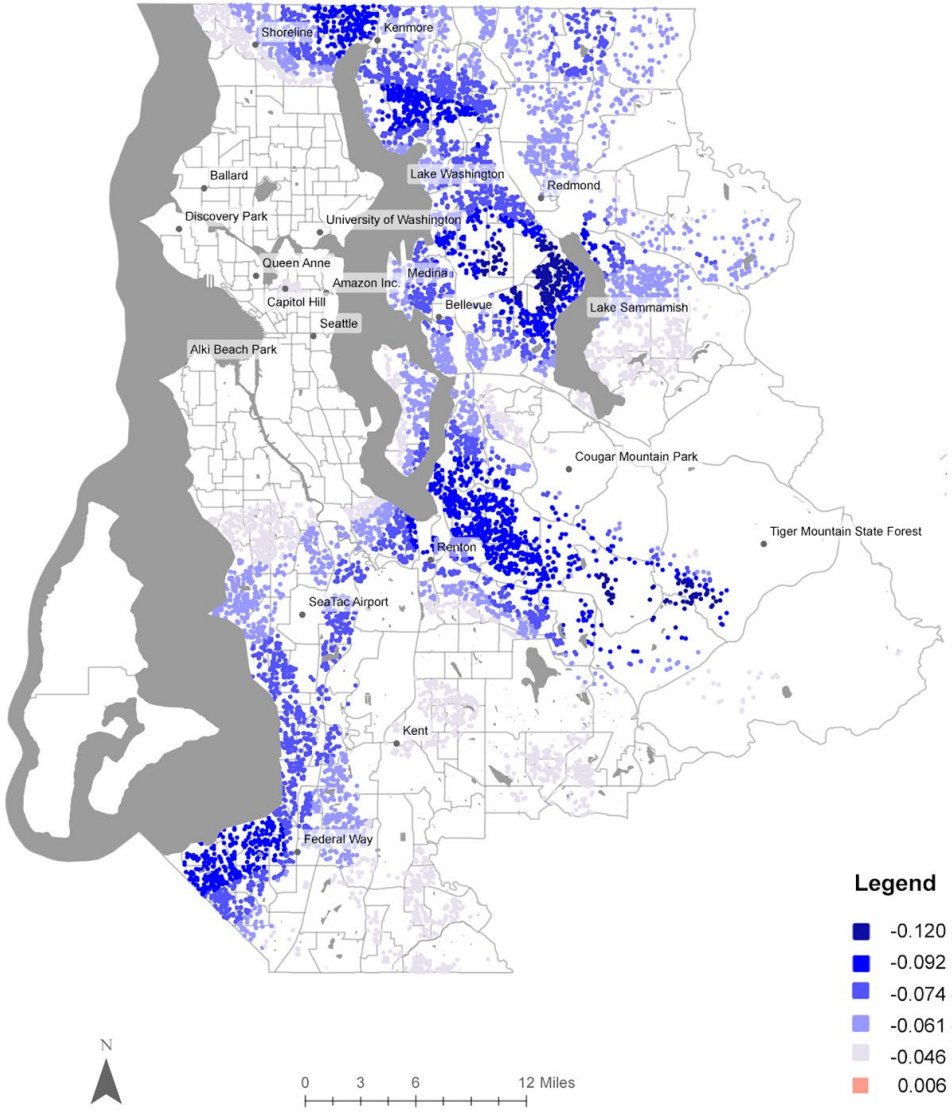
Experiment

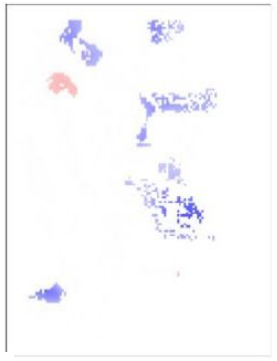
Aggregate the data at different grid levels and calibrate a local model at each grid scale



Research Design

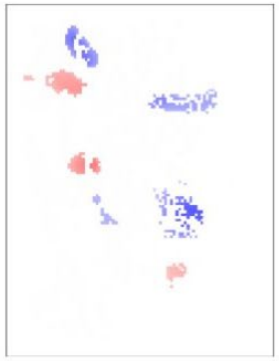
At what scale does the inference change?





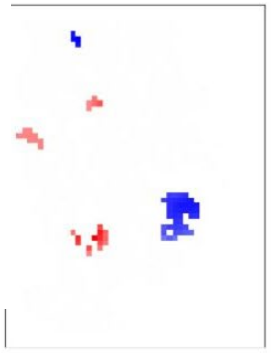
400m x 400m grid

450m x 450m grid

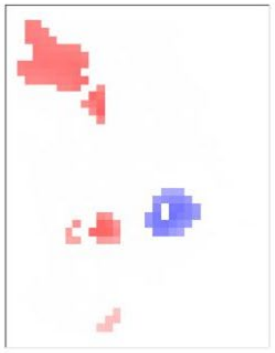


500m x 500m grid

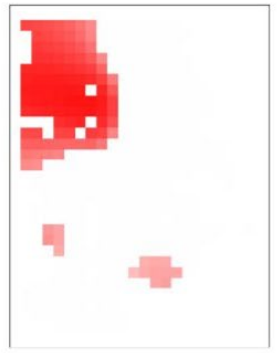
750m x 750m grid



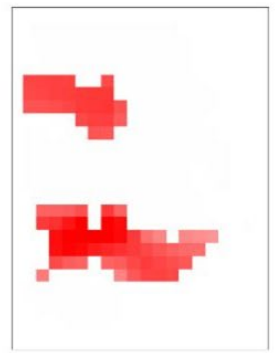
1000m x 1000m grid



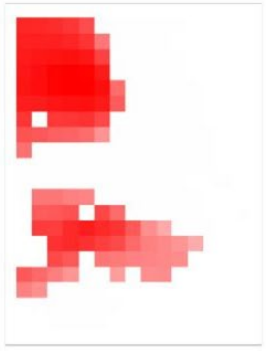
1500m x 1500m grid



2000m x 2000m grid

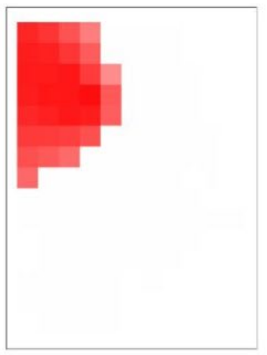


2500m x 2500m grid



3000m x 3000m grid

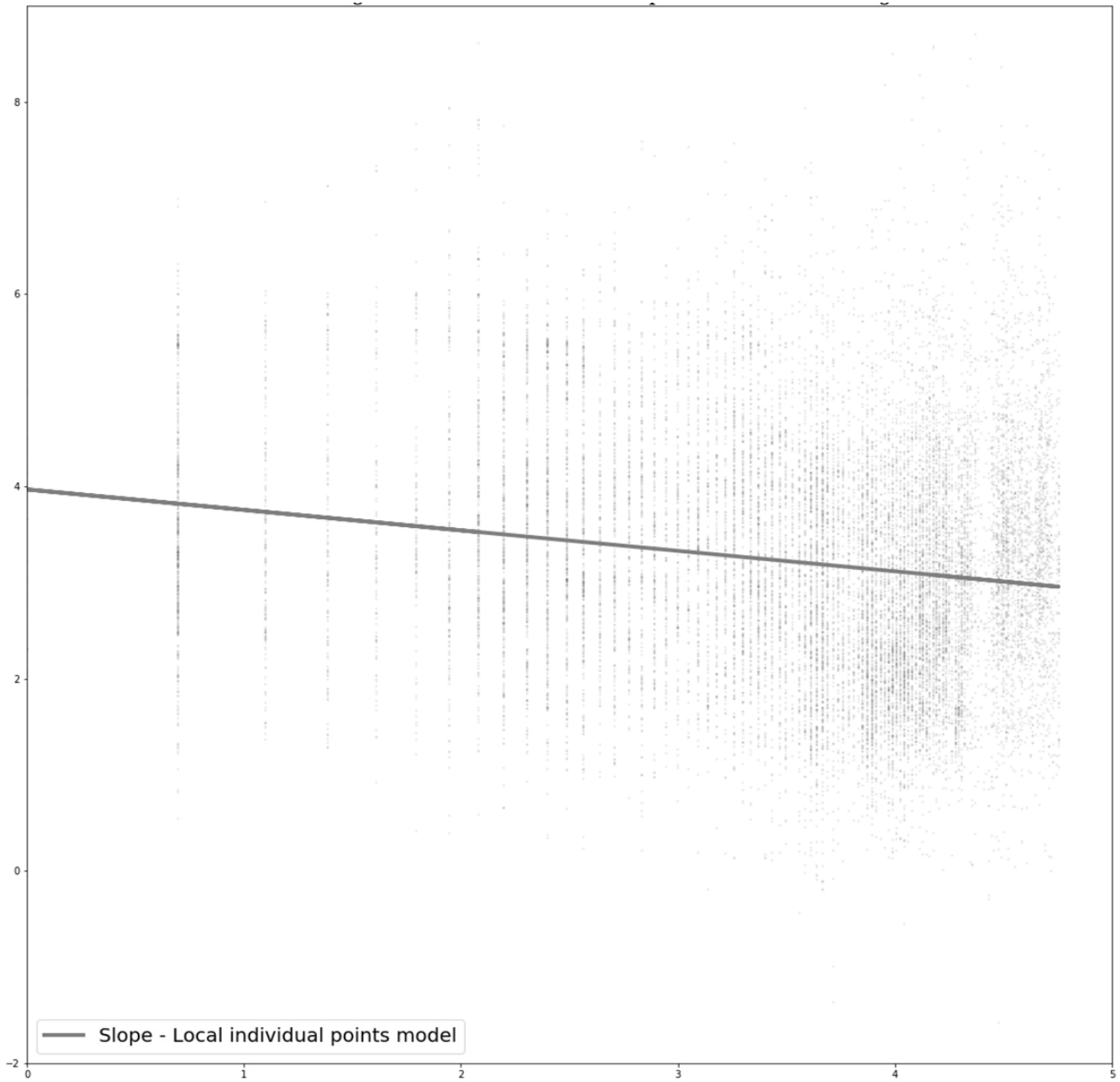
3500m x 3500m grid



4000m x 4000m grid

Research Design

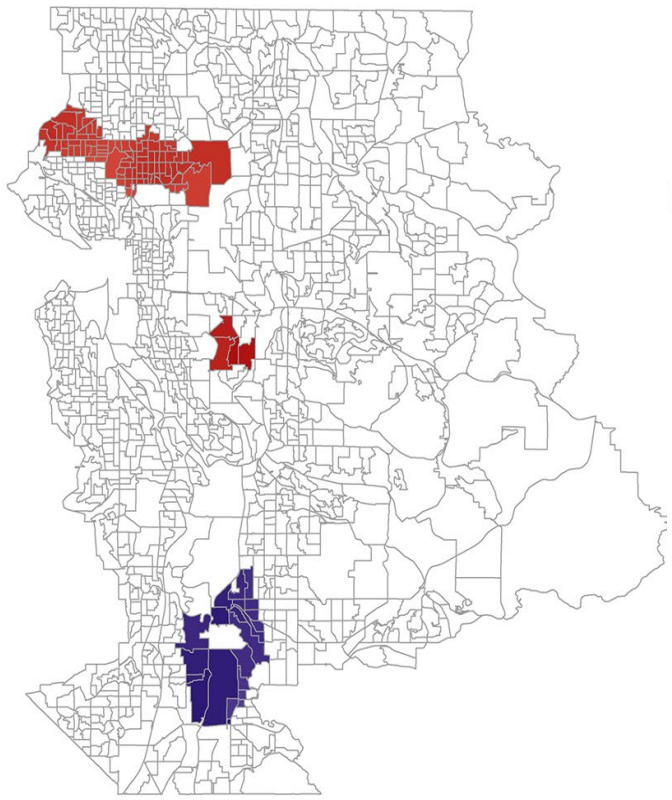
Conditioned relationship between age of property and house price

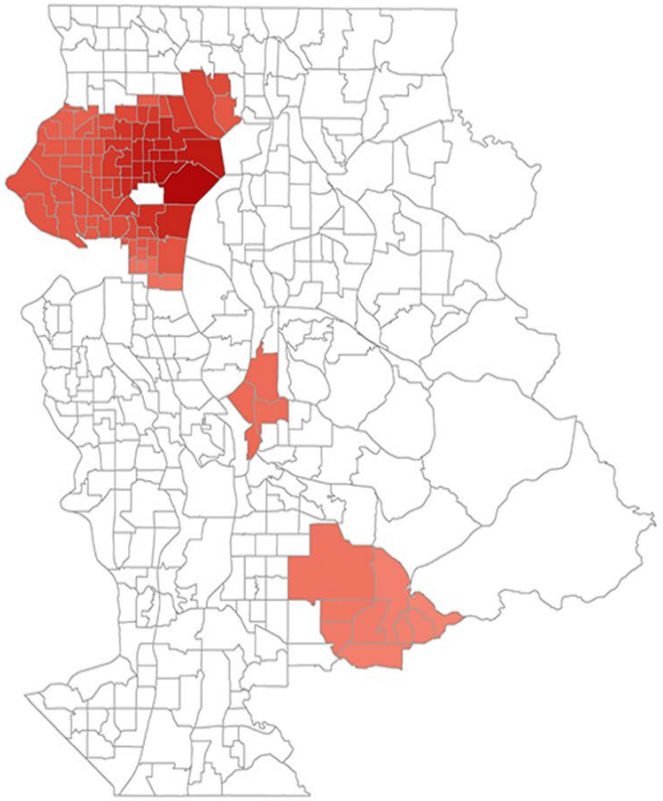


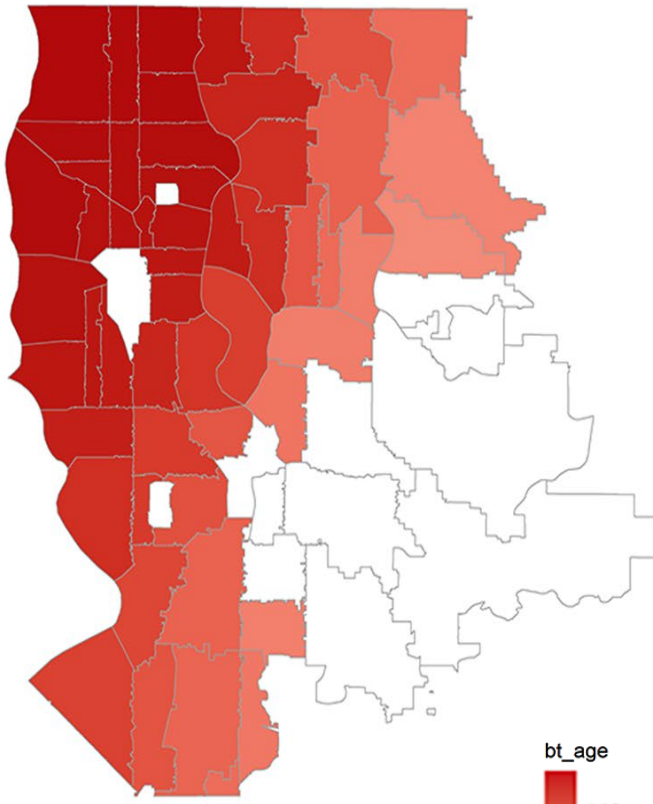
Research Design

What to do with contrasting inferences?

So which one of these maps is correct?





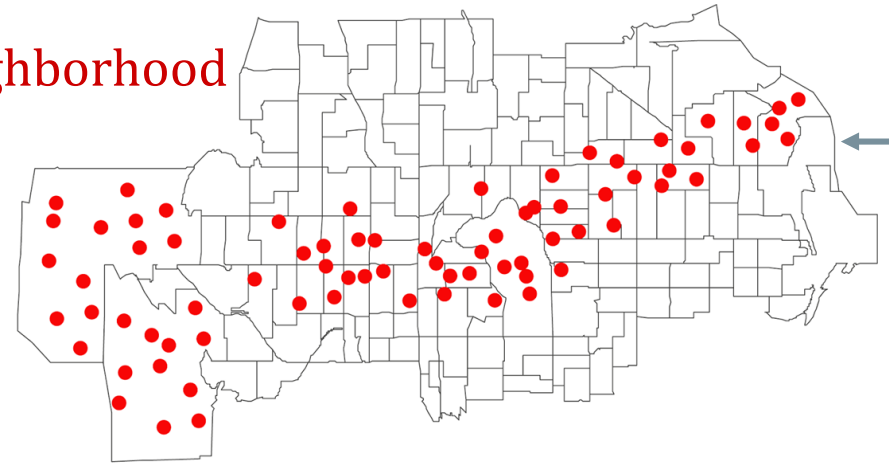


Depends on your question.

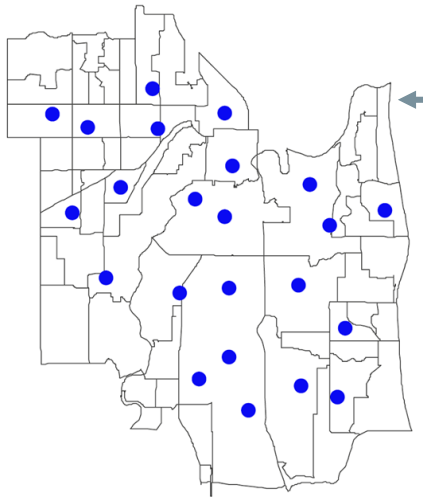
Research Design

In global models

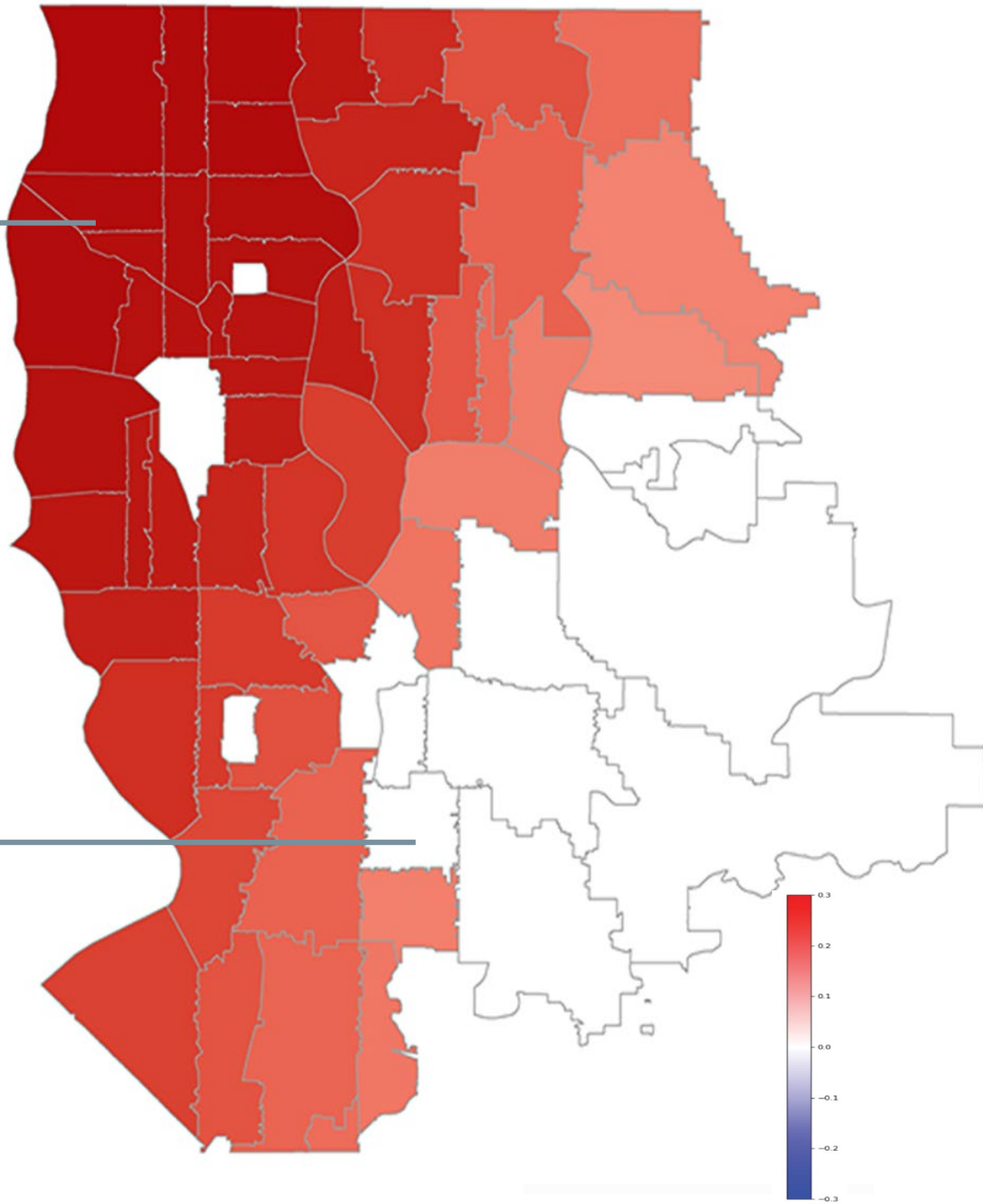
Between neighborhood comparison



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



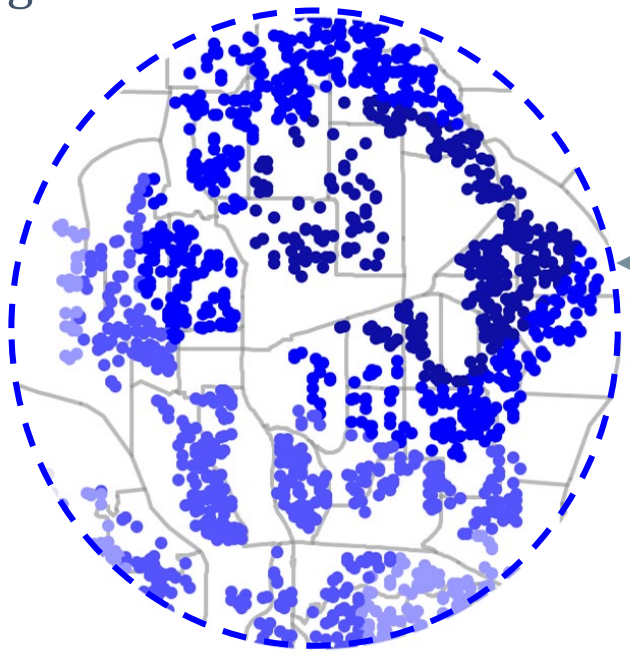
Research Design

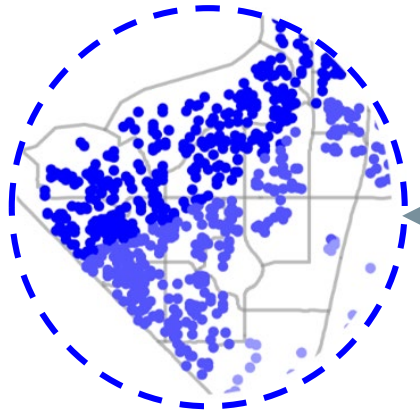
In local models with disaggregated data

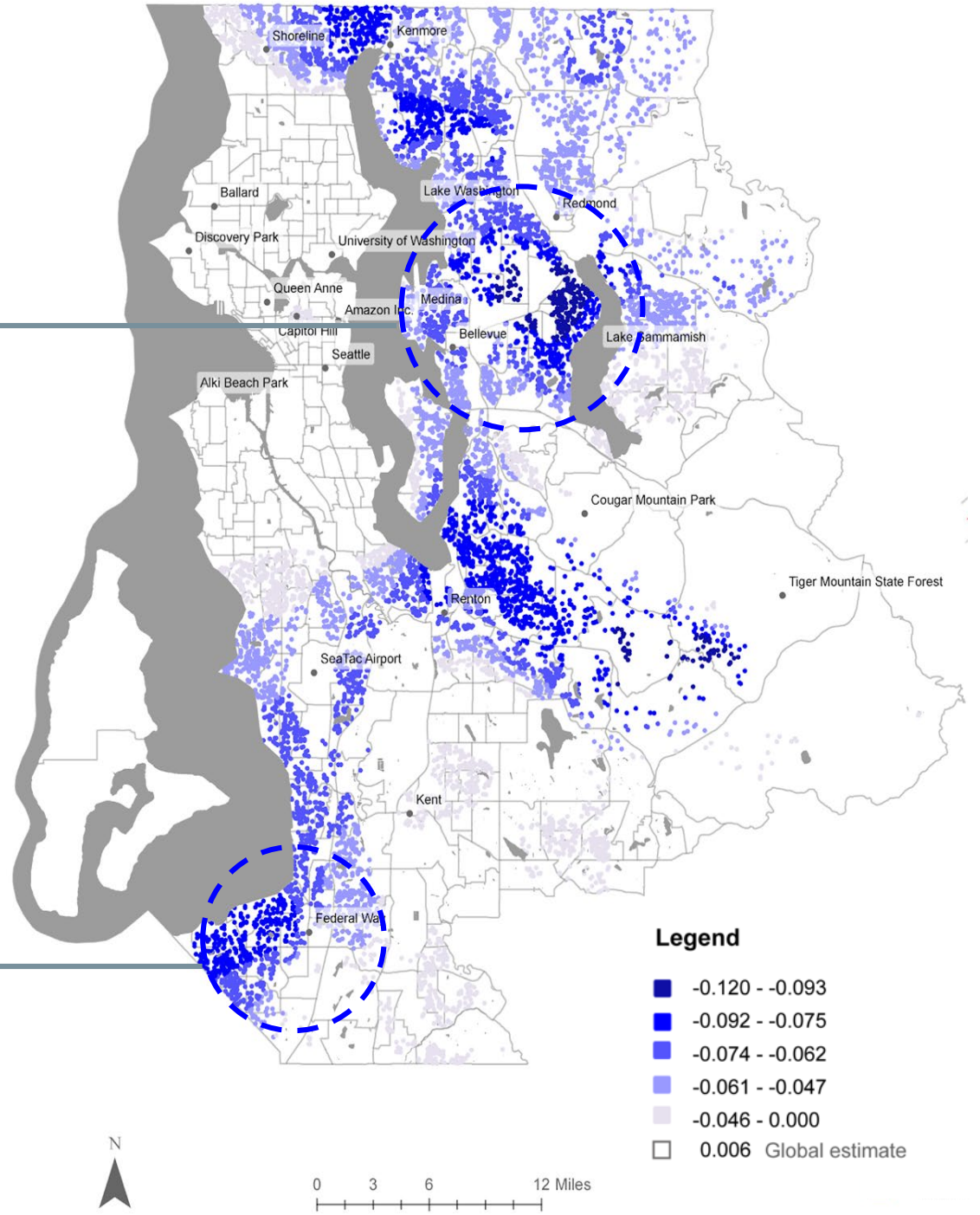
Within neighborhood comparison

Similar houses within small subsets are compared with one another

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







Intellectual merits and Broader impacts

- This research highlights that questions within spatial analysis research have an inherent scale which must be addressed in choice of modeling scale, results and interpretation

- Further, local and global models answer different questions related to spatial scale and the parameter estimates from global models are not necessarily ‘averages’ of their local equivalents

- With the growing popularity of local models, these issues related to scale are likely to be encountered more frequently and hence need to be better understood.

- Overall, this part of my dissertation advances local modeling as an effective way to explore and inform on commonly observed statistical biases such as Simpson’s paradox, MAUP and ecological fallacy

Thank you!

Any Questions?