21st Annual
GIS/CAMA
Technologies Conference
Chattanooga Convention Center

Continuing Education (CE) Credit

Recertification Credit forms for CE credit can be collected from the Registration Desk on <u>Thursday</u>.

Housekeeping

The conference proceedings will be available approximately 8 weeks after the conference.

Sub-neighborhood and baseline value study based on geographical grouping analysis for New York City

Min Zhu, Property Modeling and Research Unit, NYC DOF
Michael Cui, Property Modeling and Research Unit, NYC DOF

Study Goal:

- Seek to use statistical methods based on data(condo sales data used for this talk)
  – Guide the subdivision and grouping of existing neighborhoods
  – Produce baseline values based on location
- This is a pilot study conducted as part of a neighborhood delineation project.
- It is still an ongoing study…

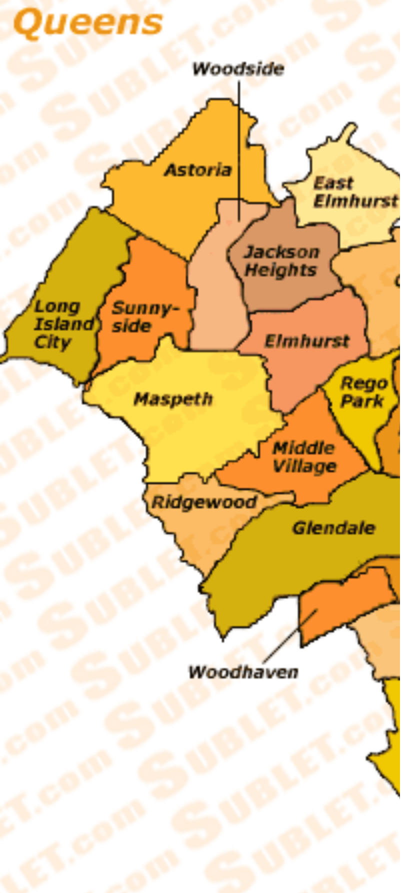
About Neighborhoods

- Traditional neighborhoods?
- Existing neighborhood boundaries?
  - Market change: new emerging areas
  - Neighborhood boundary and purpose
  - New York City Released Neighborhoods
- New York City: more heterogeneity

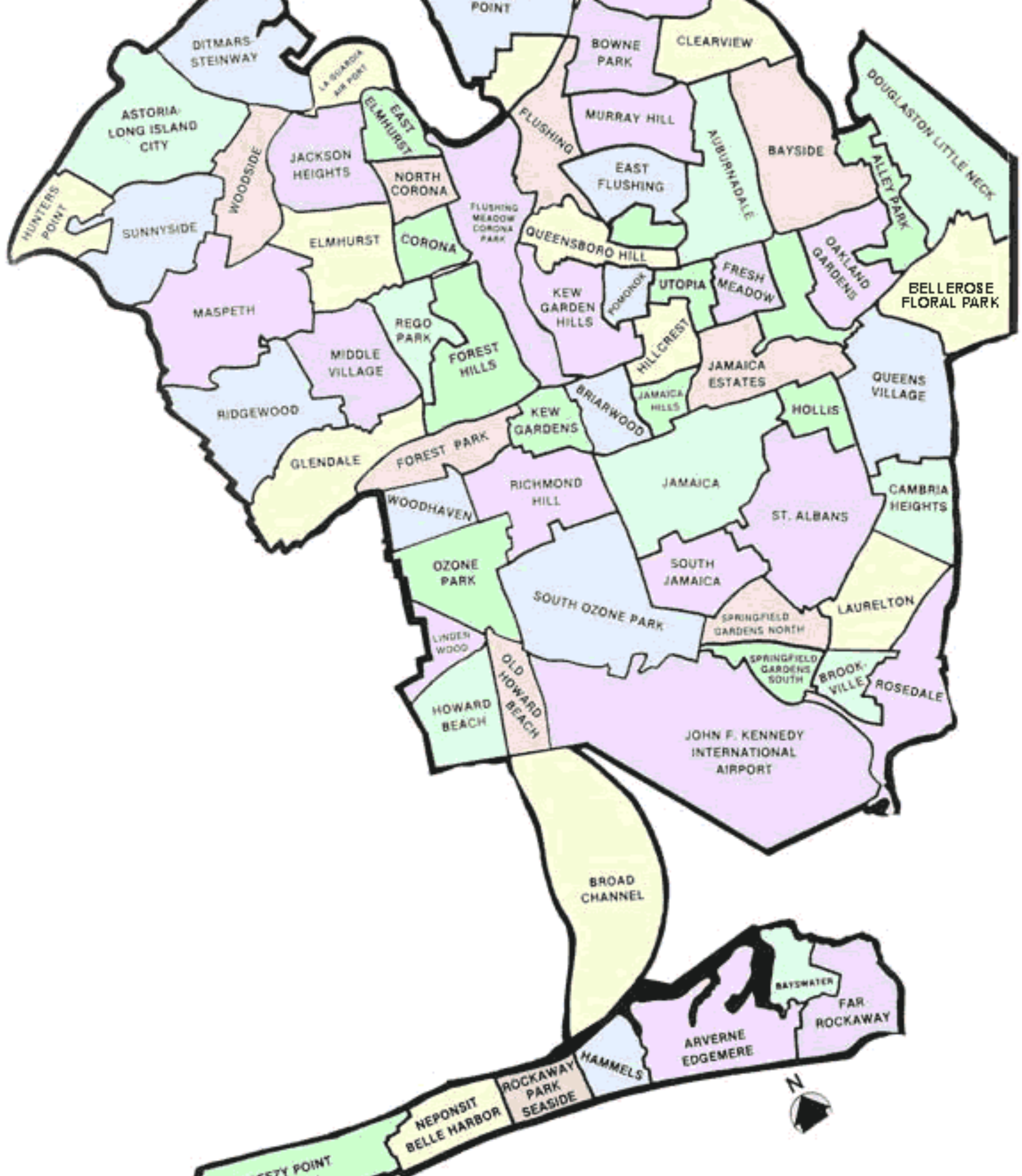
We Asked Yo
Nigel Chiway
map-heres-w

About Neighborhoods

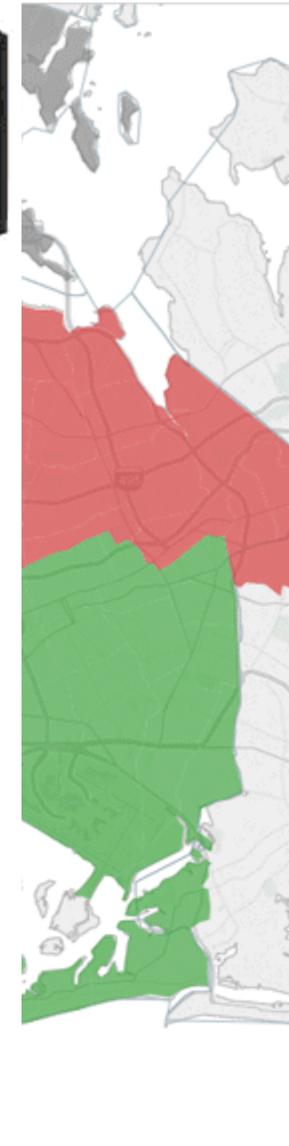
- Traditional neighborhoods?
- Existing neighborhood boundaries?
  - Market change: new emerging areas
  - Neighborhood boundary and purpose
  - New York City Released Neighborhoods
- New York City: more heterogeneity



Rentals in Queens! . Retrieve
2017, from https://queens.su



HNN - Cutting up NYC: The story of a market realignment
http://www.hotelnewsnow.com/Articles/91935/Cutting-up-NYC-The-story-of-a-market-realignment

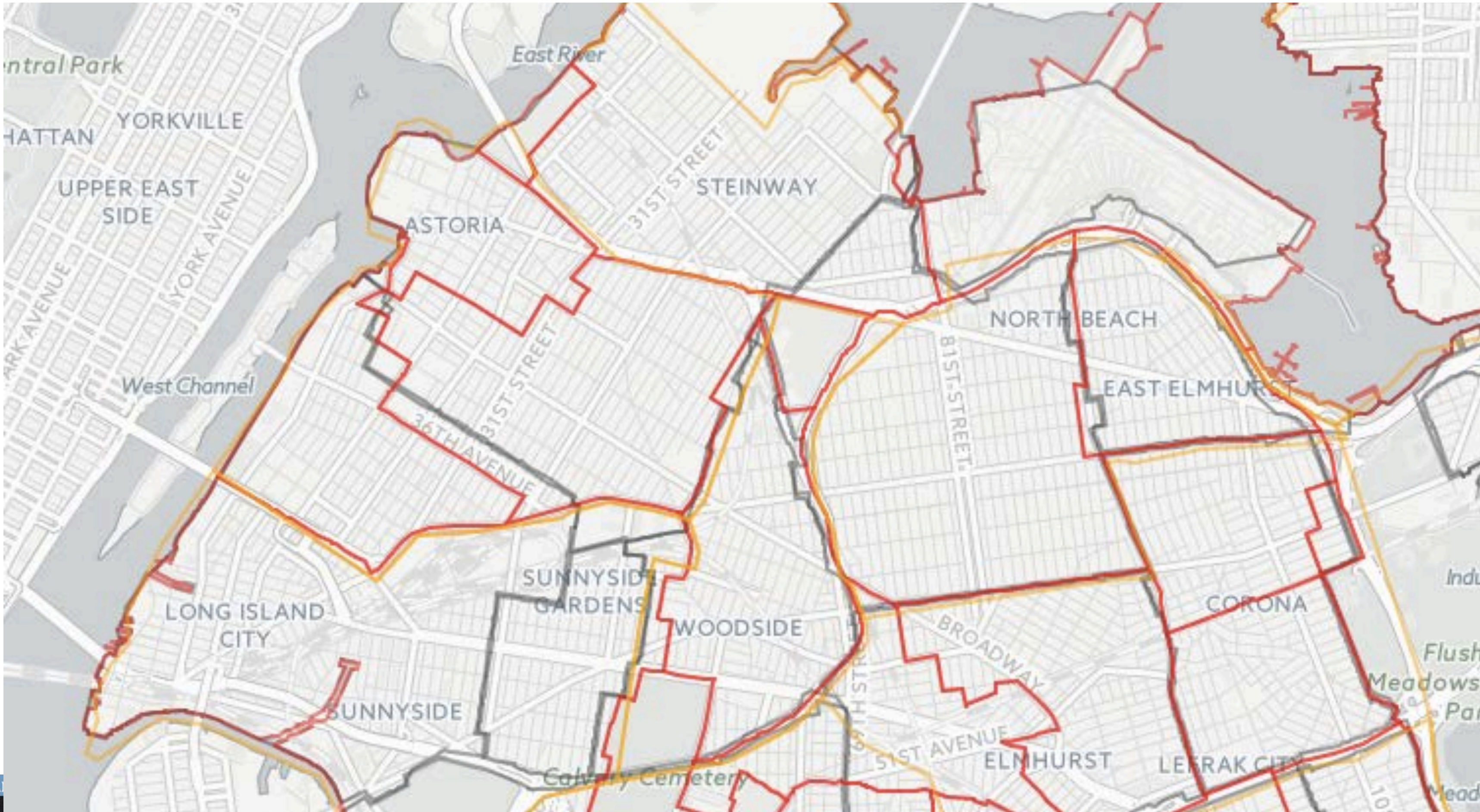


What is the borderline queens neighborhood to Long island ? From http://www.City-Data.com

About Neighborhoods

- Traditional neighborhoods?
- Existing neighborhood boundaries?
  - Market change: new emerging areas
  - Neighborhood boundary and purpose
  - New York City Released Neighborhoods
- New York City: more heterogeneity

Overlay of different neighborhood boundaries

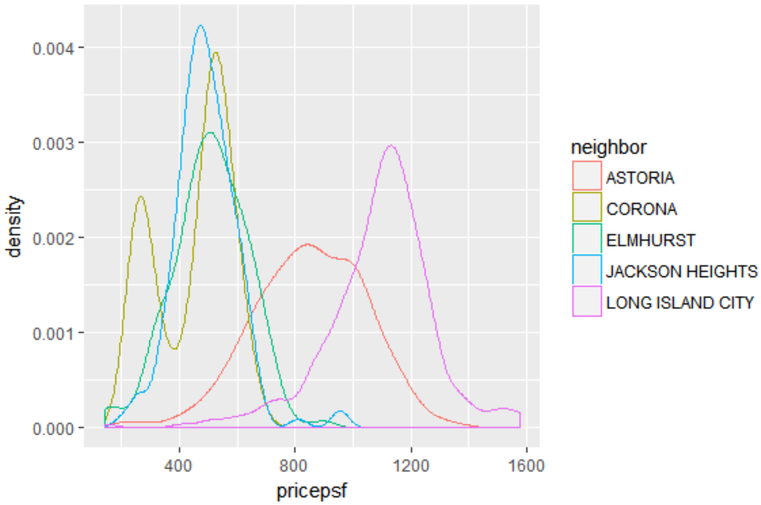


About Neighborhoods

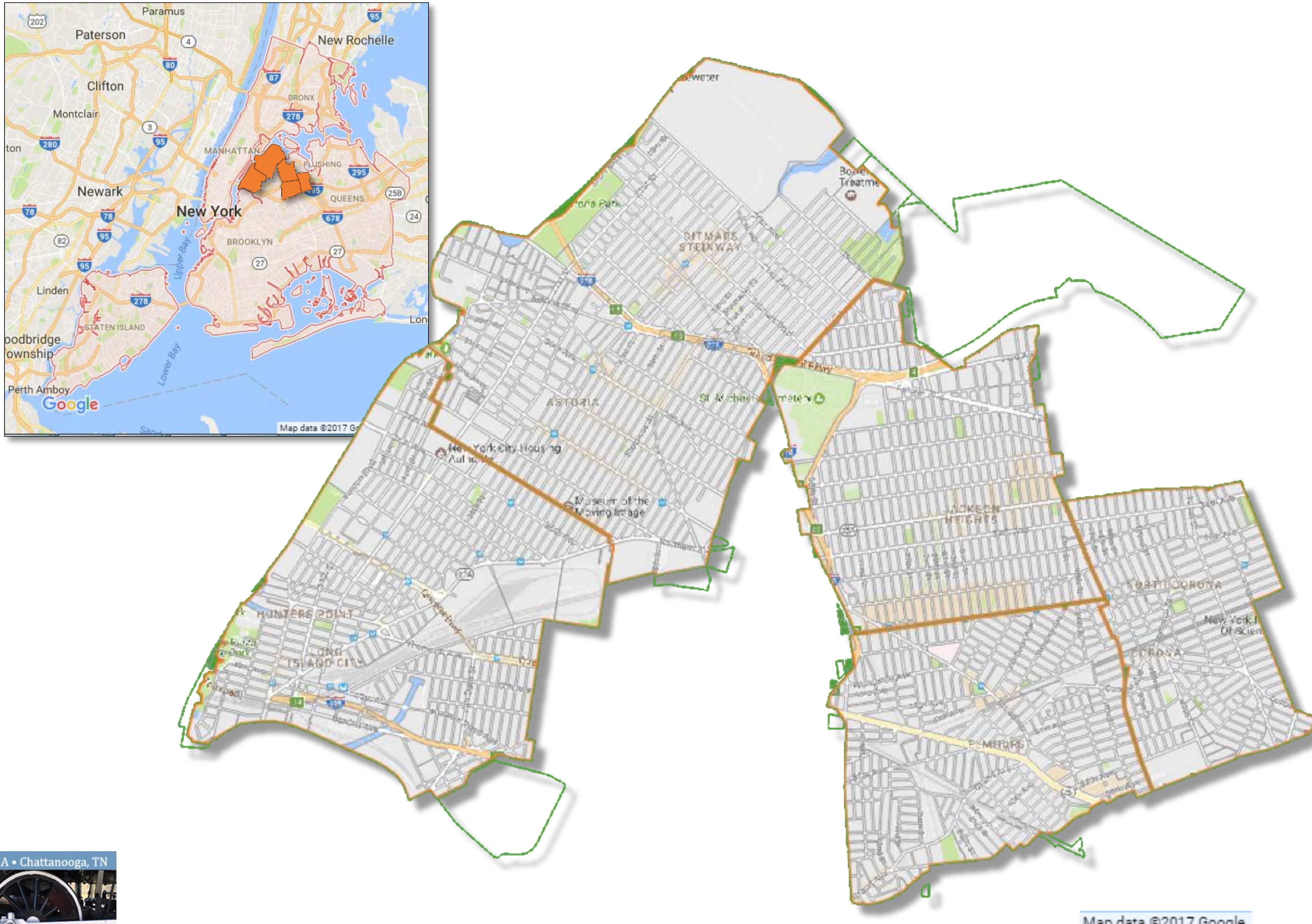
- Traditional neighborhoods?
- Existing neighborhood boundaries?
  - Market change: new emerging areas
  - Neighborhood boundary and purpose
  - New York City Released Neighborhoods
- New York City: more heterogeneity

Study Area

- Study area:
  - 5 neighborhoods in Queens:
    - ASTORIA, CORONA, ELMHURST, JACKSON HEIGHTS, LONG ISLAND CITY
    - Condo Sales Price per sf (2015 -2016)

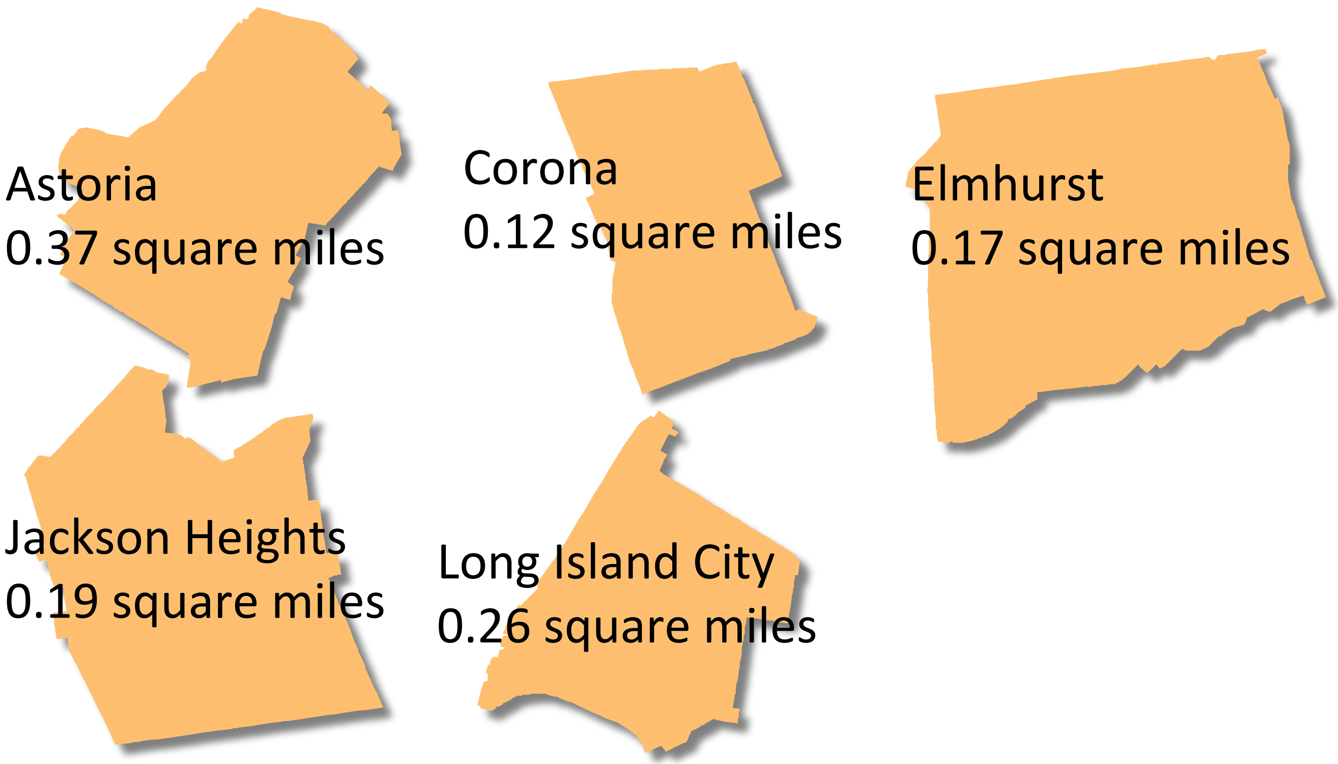


neighbor	n	Mean	Median	Q1	Q2	Q3
ASTORIA	200	$ 850.78	$ 853.08	$ 729.91	$ 853.08	$ 993.44
CORONA	134	$ 447.49	$ 502.99	$ 300.09	$ 502.99	$ 542.94
ELMHURST	132	$ 507.97	$ 511.41	$ 435.29	$ 511.41	$ 600.93
JACKSON HEIGHTS	143	$ 494.04	$ 485.75	$ 428.24	$ 485.75	$ 551.14
LONG ISLAND CITY	245	$ 1,089.66	$ 1,108.70	$ 1,001.16	$ 1,108.70	$ 1,188.17

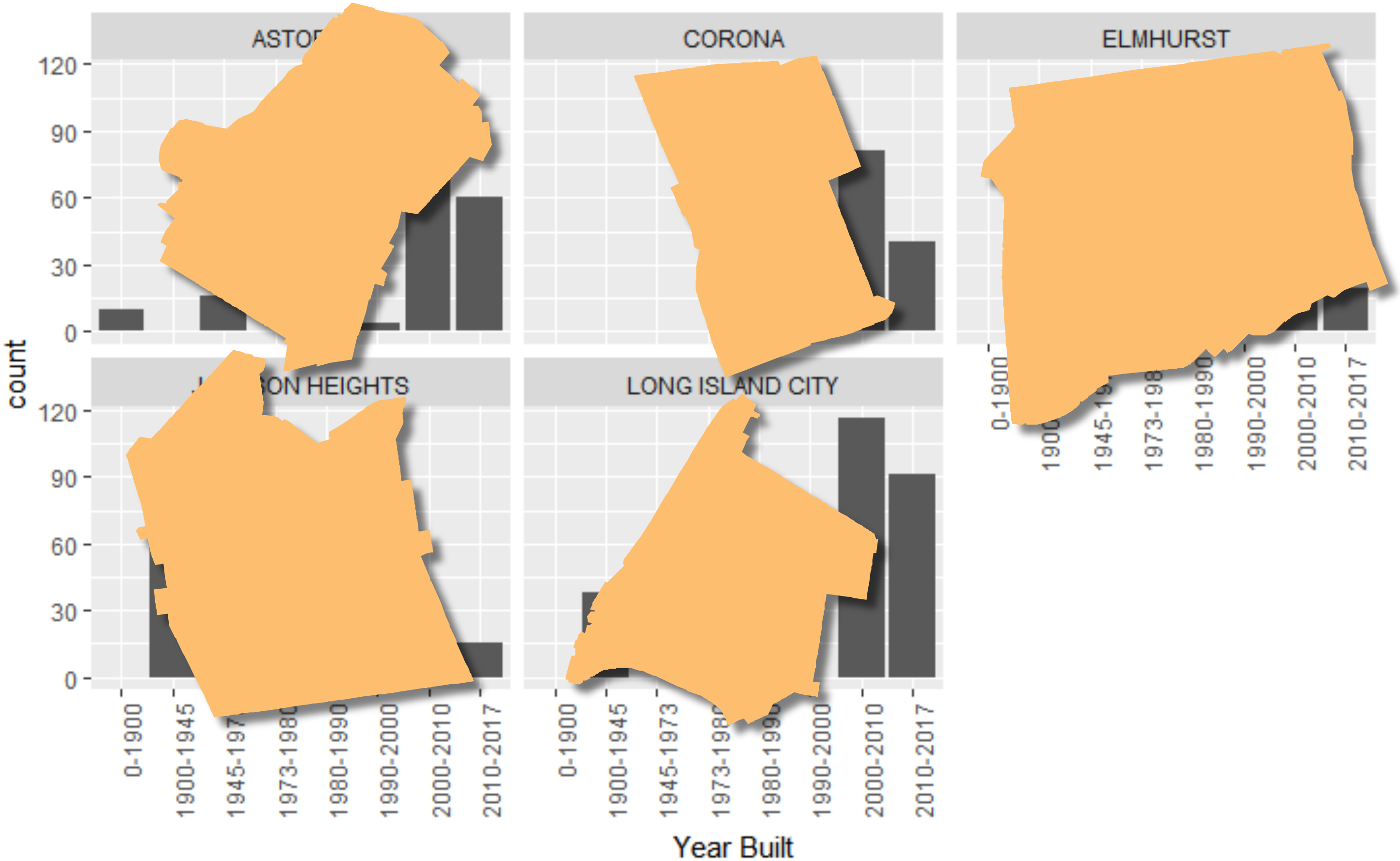


Paramus
Paterson
New Rochelle
Clifton
Montclair
BRONX
MANHATTAN
FLUSHING
Newark
New York
QUEENS
BROOKLYN
Upper Bay
Linden
STATEN ISLAND
Lower Bay
Perth Amboy
Google
DITMARS STEINWAY
ASTORIA
New York City Housing Authority
Museum of the Moving Image
JACKSON HEIGHTS
NORTH CORONA
CORONA
HUNTERS POINT
LONG ISLAND CITY
Map data ©2017 Google

Descriptive Data on Neighborhoods

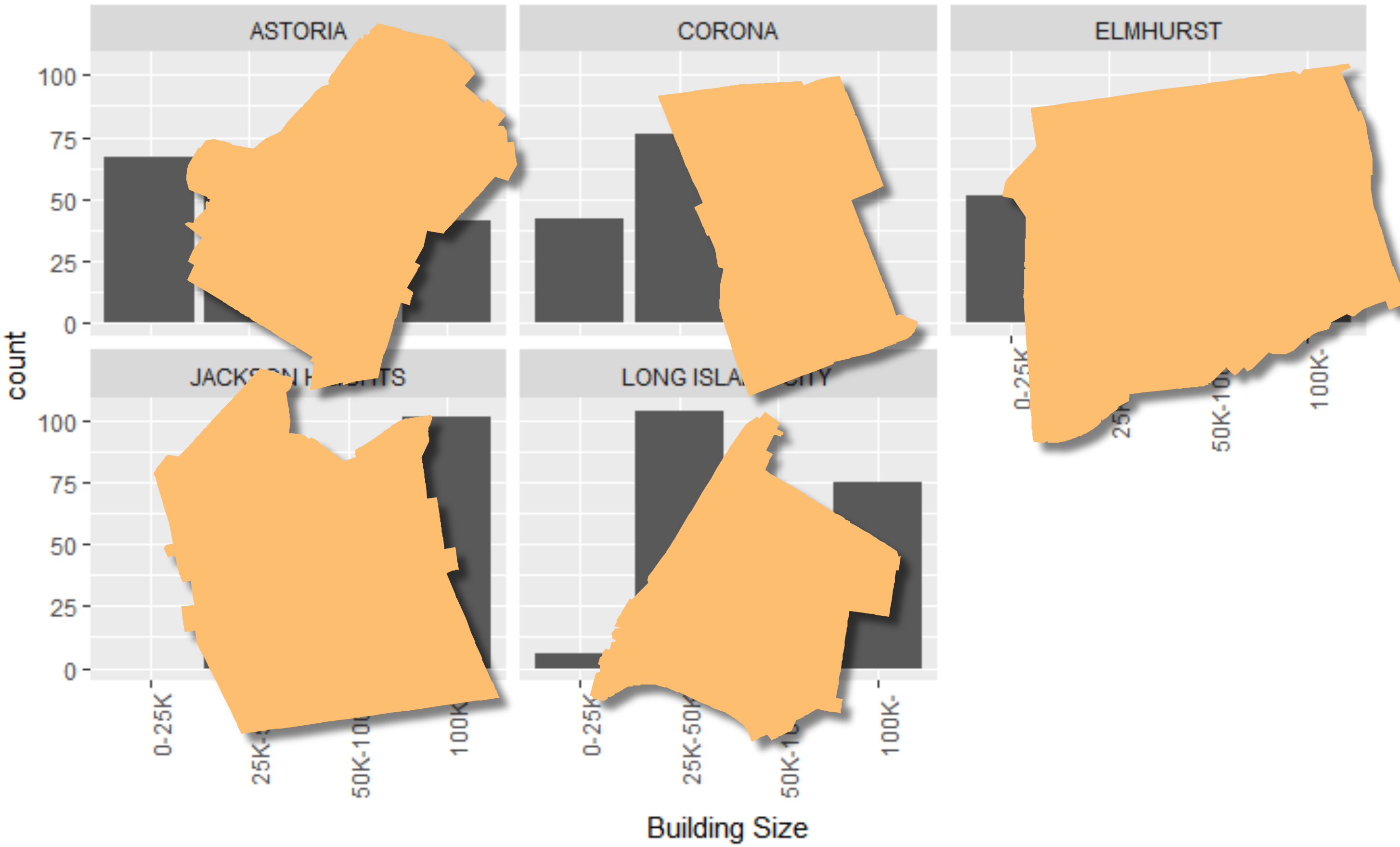


Descriptive Data on Neighborhoods



Descriptive Data on Neighborhoods

Descriptive Data on Neighborhoods



Geographical Methods:

- General discussion
  - Thin plate spline and classification
  - K nearest neighbor and classification
  - GWR and cluster analysis

Thin plate spline

- Thin plate splines (TPS) is a spline based technique for data interpolation and smoothing
- The minimization problem is:

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda * J_m(f)$$

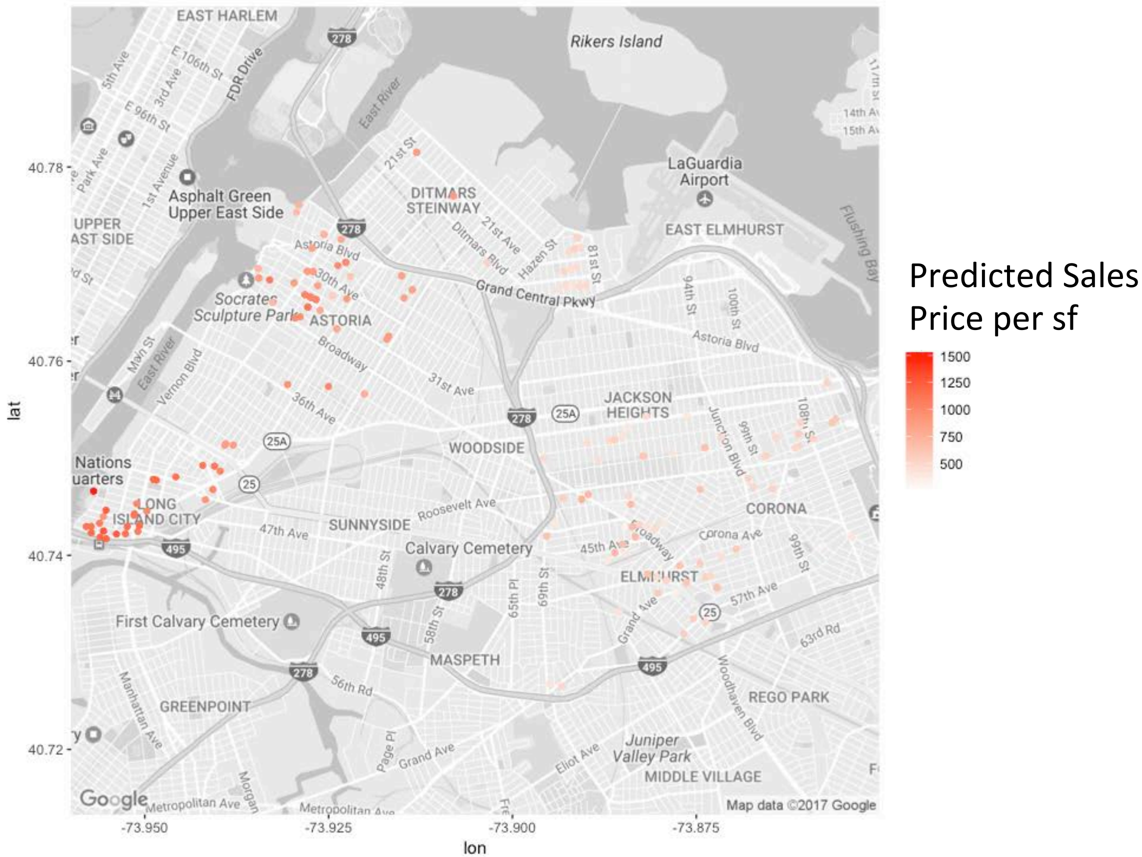
- Where Jm(f) is the m-th order thin plate spline penalty functional.
- The smoothing variant, uses a tuning parameter λ to control how non-rigid is allowed for the deformation

Thin plate spline

- Original Sales data: Tax Class 2

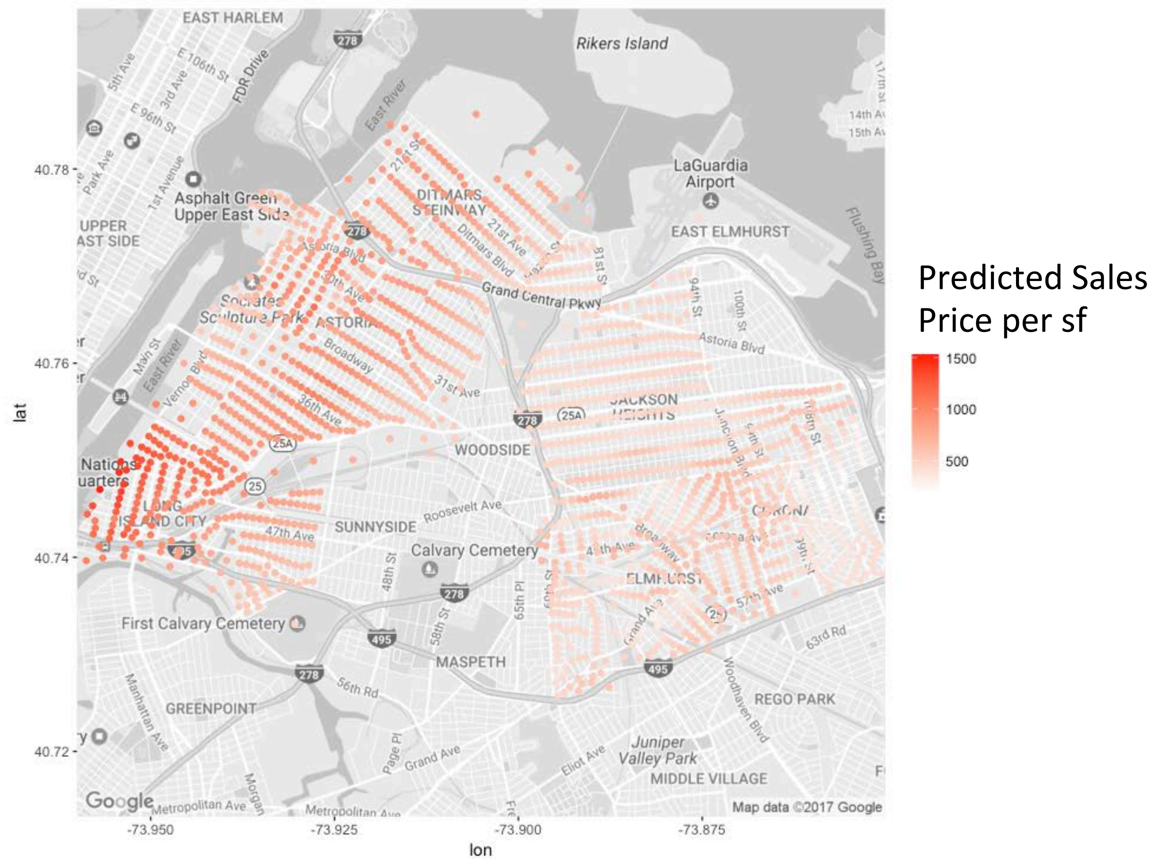
Thin plate spline

- Predictions: use optimized λ



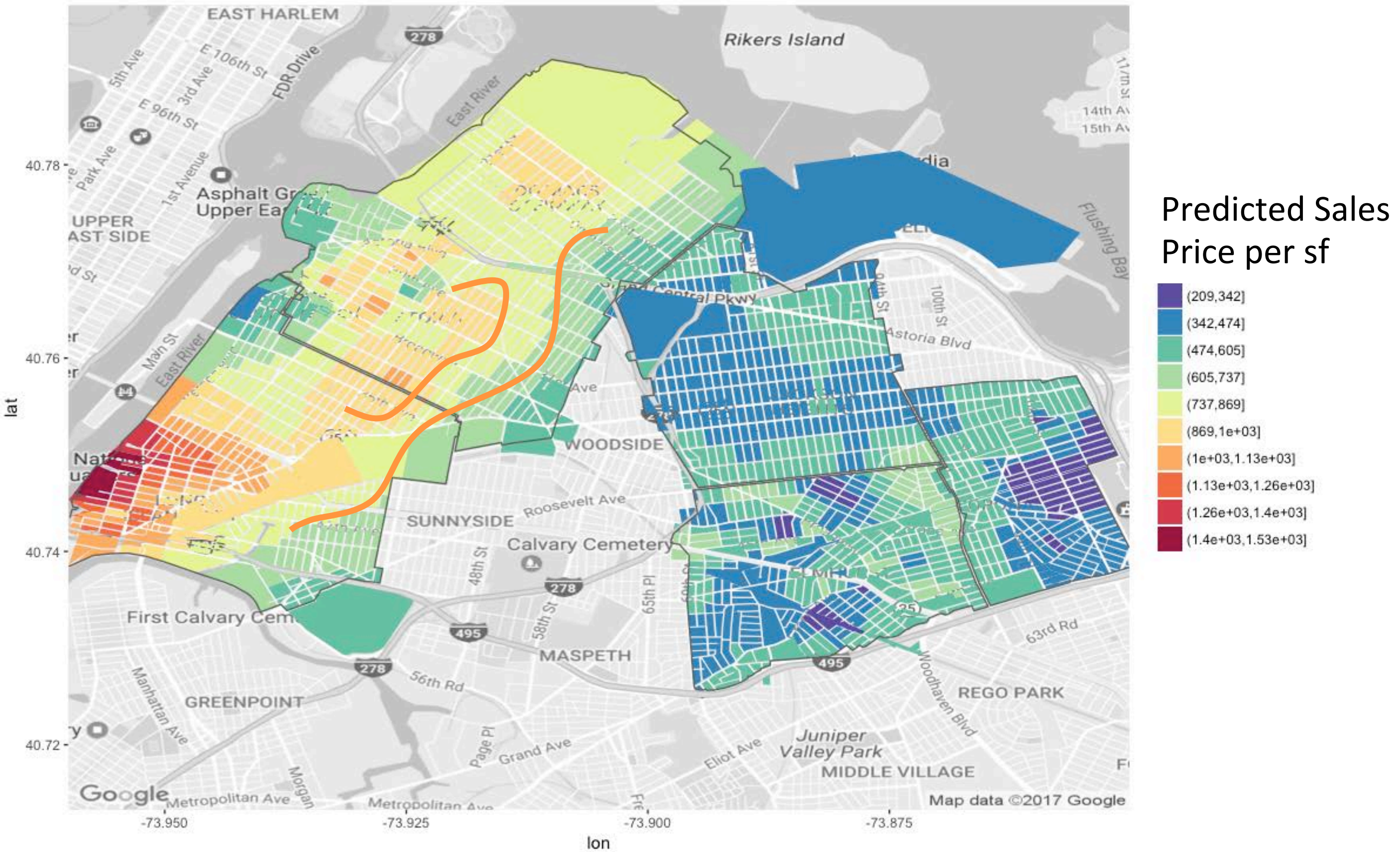
Thin plate spline

- Predictions for all blocks: use optimized λ



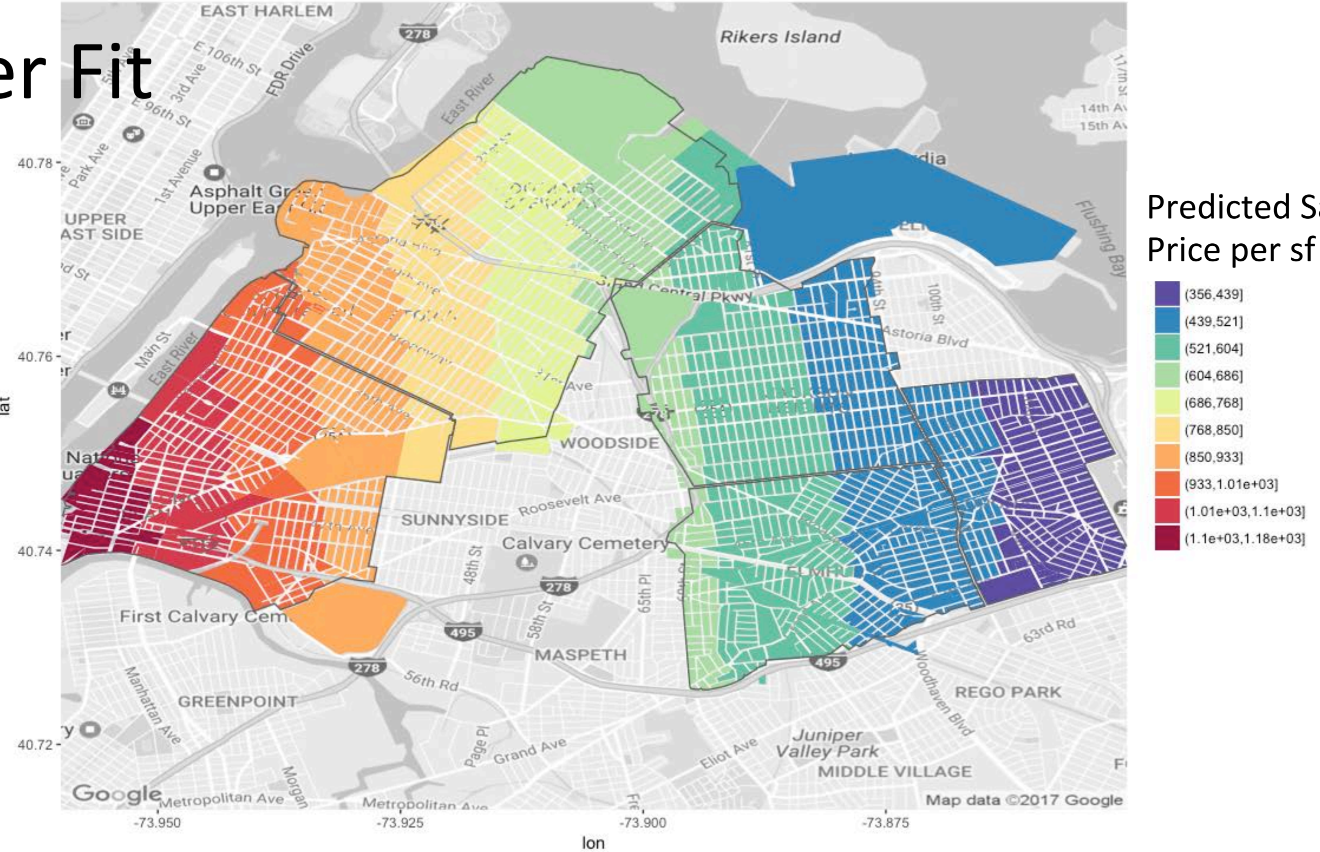
Thin plate spline

- Predictions for all blocks: use optimized λ



Thin plate spline

- Predictions for all blocks: use a larger λ
- Under Fit

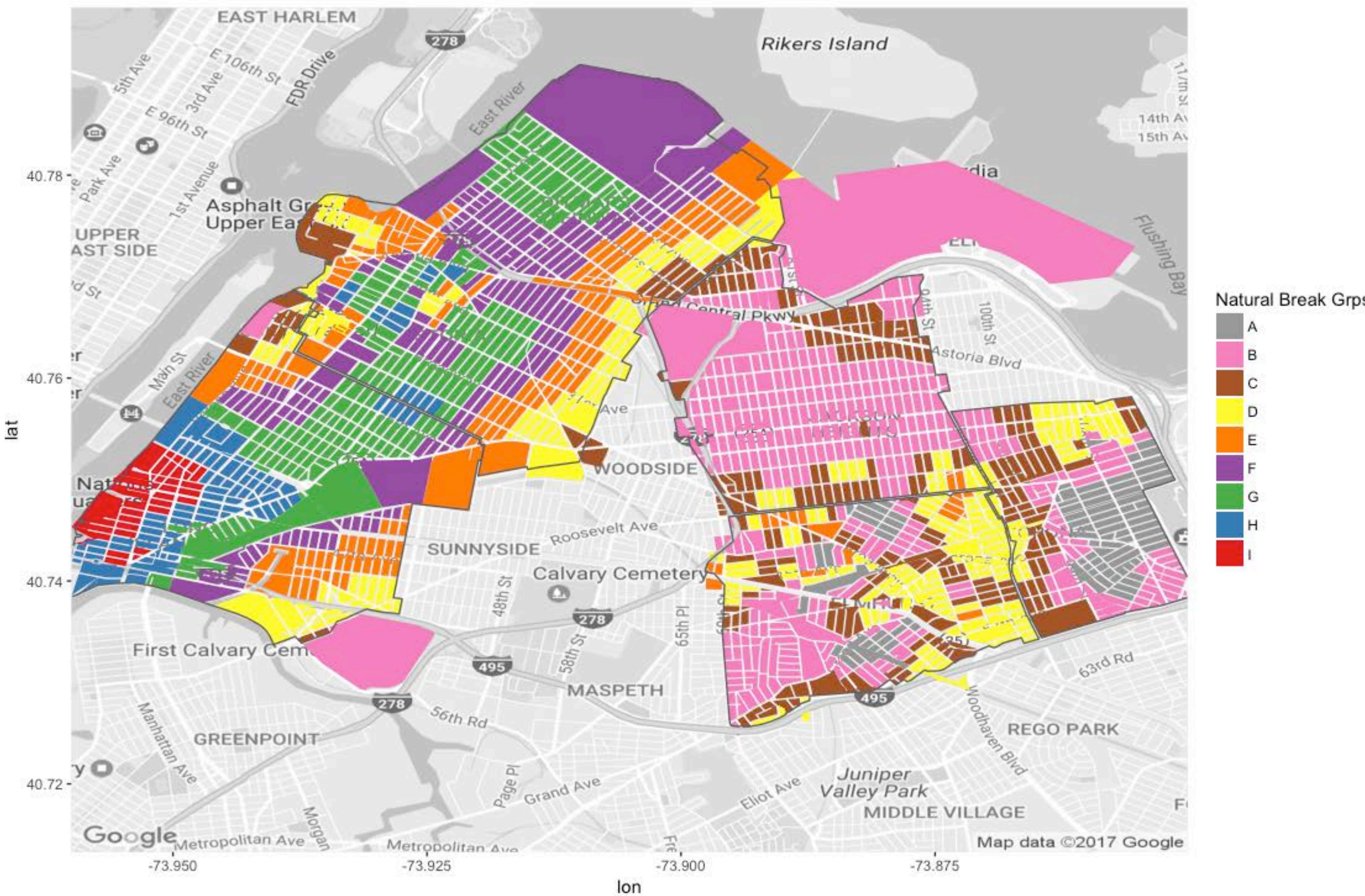


Classification: Jenks Natural Breaks

- After filling up the entire area with predicted values
- Data clustering method: Jenks Natural Breaks Classification
  - Determines the best arrangement of values into different classes
  - Divide blocks into sub-neighborhood areas
  - By reducing the variance within classes and maximizing the variance between classes

Thin plate spline

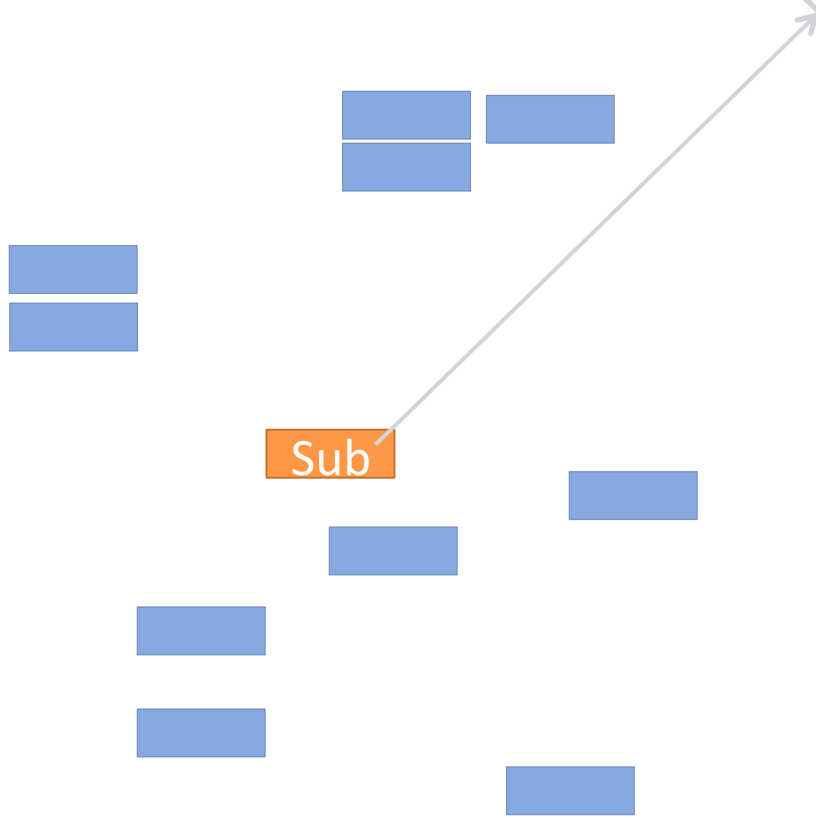
- Grouped to sub-neighborhoods: Jenks Natural Breaks



Thin plate spline

- Discussion:
  - Pros:
    - quick and straight forward,
    - use coordinates and specify lambda
  - Cons:
    - choice of lambda, can be tricky; Can over fit.
    - When data is sparse, confident level for predictions near boundary are low. May end up with interpolation values from closest point looking like contours.
    - When data variation is larger in small area, extreme peaks can form.

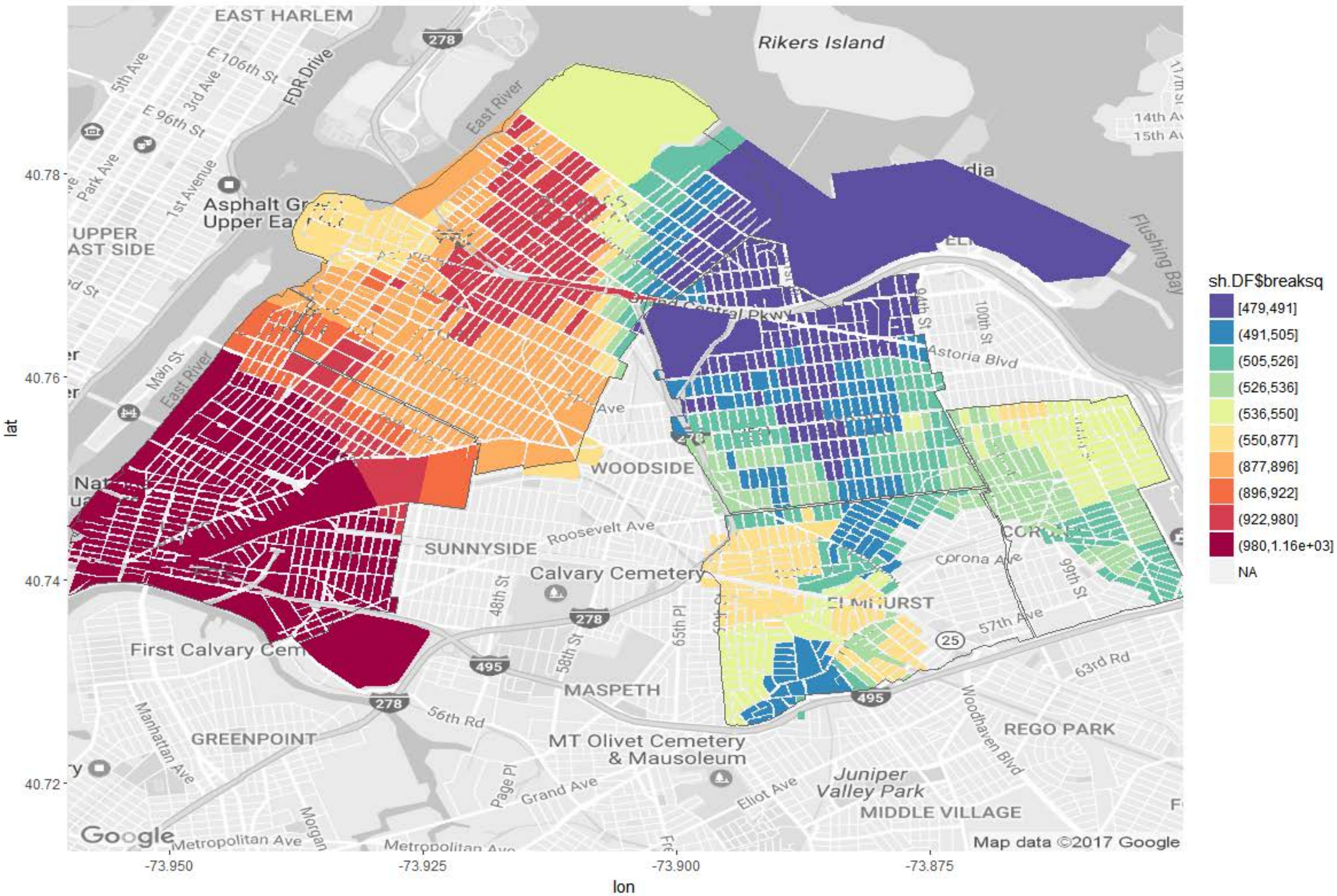
K nearest neighbor



- K-nearest neighbor:
  - data interpolation and smoothing technique
  - Value of a given location determined by the k nearest neighbors around
  - Flexibility on distance
- For our purposes, we look for a subject block's value based on the median of its k nearest block values
- Classification: Jenks Natural Breaks

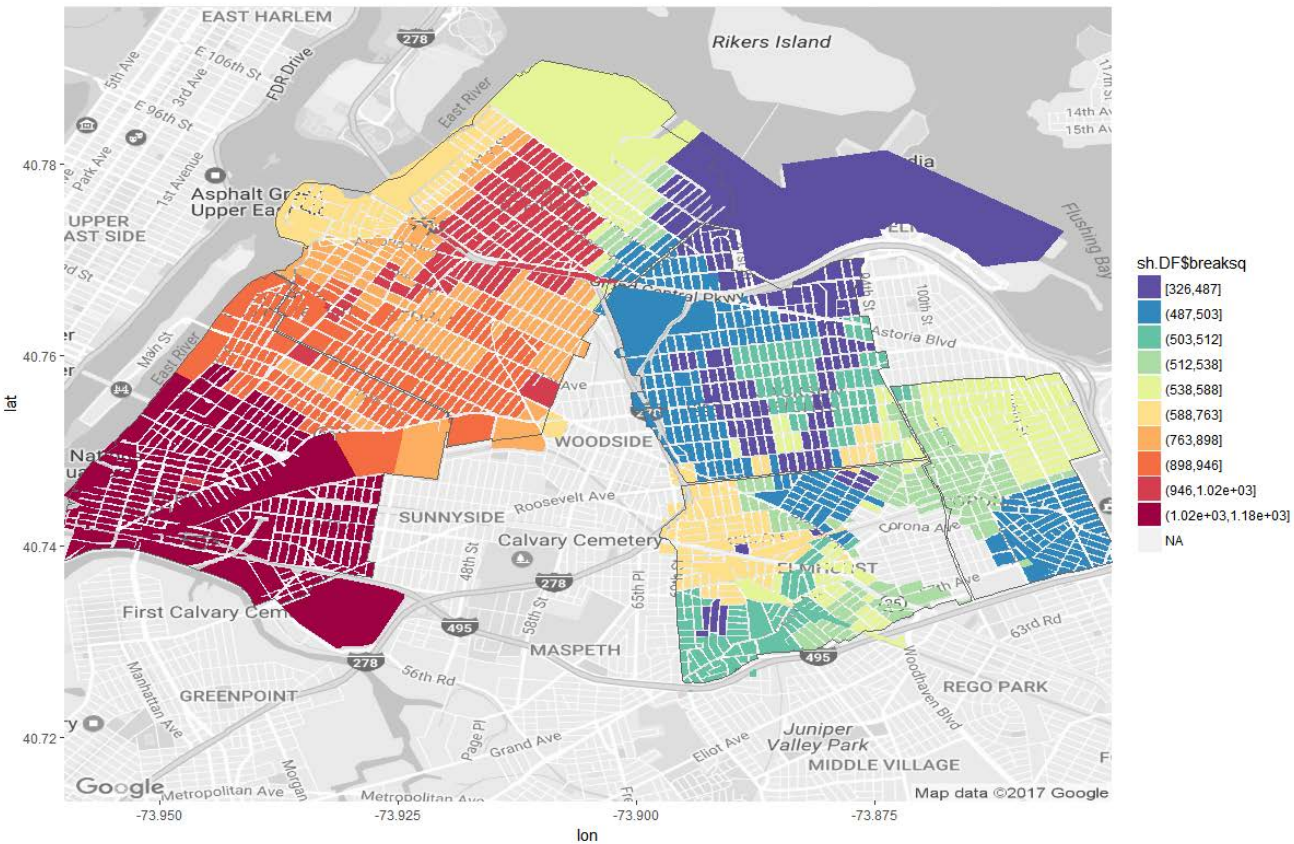
K nearest neighbor

Tc2, 10 nearest neighbor block values
Based on tc2 sales



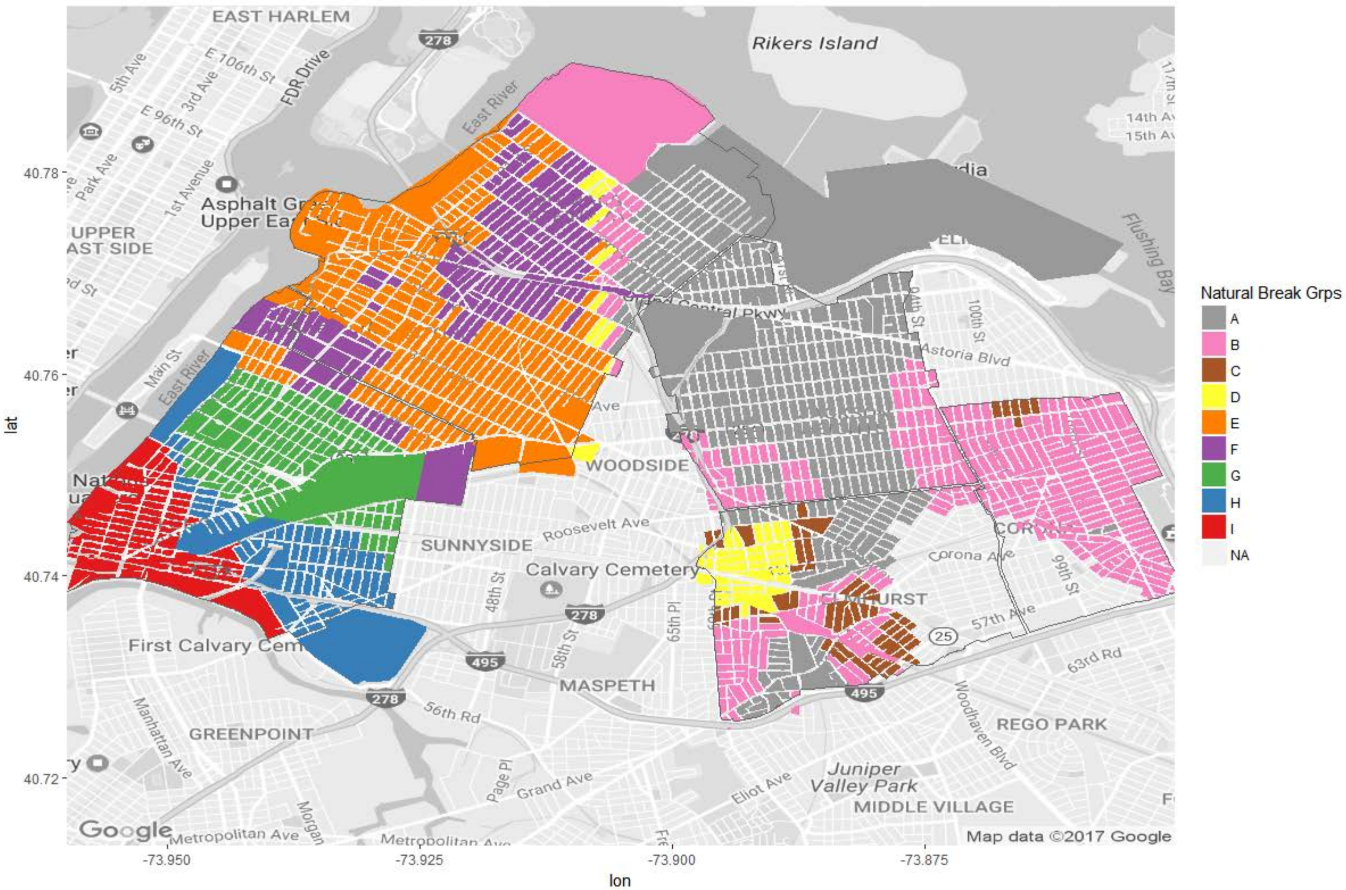
K nearest neighbor

Tc2, **5** nearest neighbor block values
Based on tc2 sales



K nearest neighbor group: Jenks Natural Break

Tc2, 10 nearest neighbor groups
Based on tc2 sales



K nearest neighbor

- Discussion
  - Pros:
    - Easy to interpret
    - Flexible searching distance
  - Cons:
    - Based on one variable: price psf
    - Small pieces of land when grouping

Geographically Weighted Regression (GWR)

- The basic form of the GW regression model is:

Dependent variable

Regression coefficients

Residual variable

$$y_i = \beta_{i0} + \sum_{k=1}^{M} \beta_{ik} x_{ik} + \varepsilon_i$$

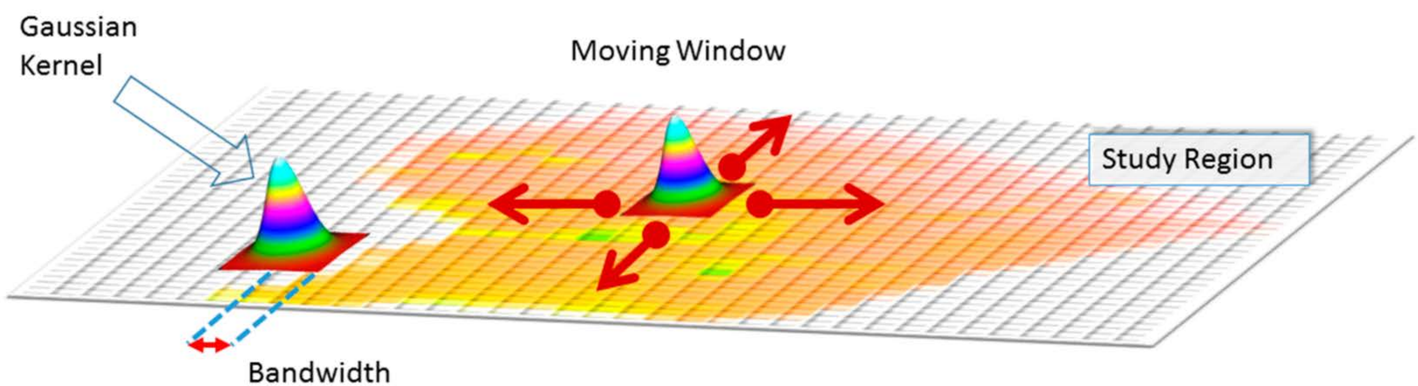
The k^th variable at location i

i=1-N number of observations
J=0-M number of ind. variables

  - As data are geographically weighted, nearer observations have more influence in estimating the local set of regression coefficients than observations farther away.
  - Difference in local coefficients reflect the difference in sensitivity to how price respond

Geographically Weighted Regression (GWR)

- Geographically Weighted Regression scheme. Geographical weighting is achieved by a fixed Gaussian kernel function with a given bandwidth moving across the spatial domain. Bandwidth determines the rate at which the weights decay around each cell, and reflects the degree of spatial variation: if the bandwidth gets larger, the model will tend to a global regression model.



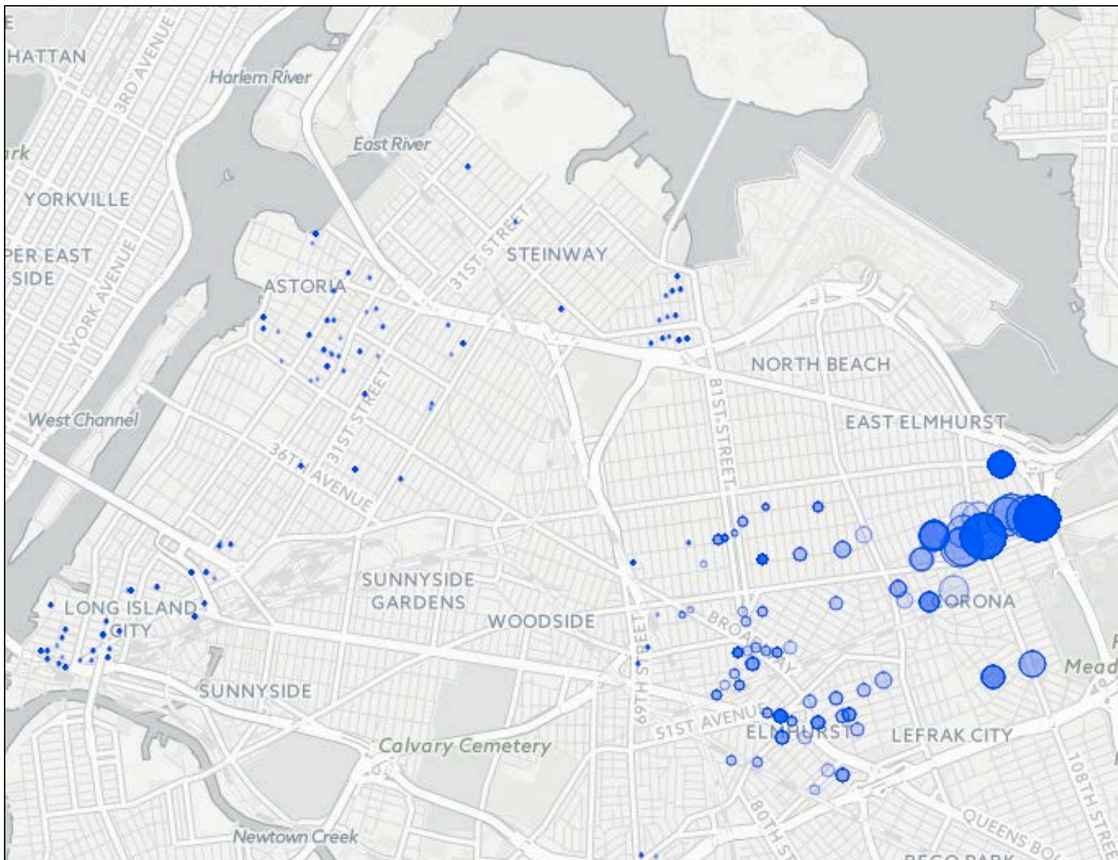
* Ceccherini, Guido; Ameztoy, Iban; Hernández, Claudia P.R.; Moreno, Cesar C. 2015. "High-Resolution Precipitation Datasets in South America and West Africa based on Satellite-Derived Rainfall, Enhanced Vegetation Index and Digital Elevation Model." *Remote Sens.* 7, no. 5: 6454-6488.

Geographically Weighted Regression (GWR)

- GWR variables
  - Predicting variables: pricepsf
  - Explaining variables:
    - Aptsize
    - Total resid unit
    - Building story
    - Year built
    - Elev/Walkup
    - Building size
    - Resratio, retratio …
    - yearqtr
  - * Variables used in model are after transformation and scaled

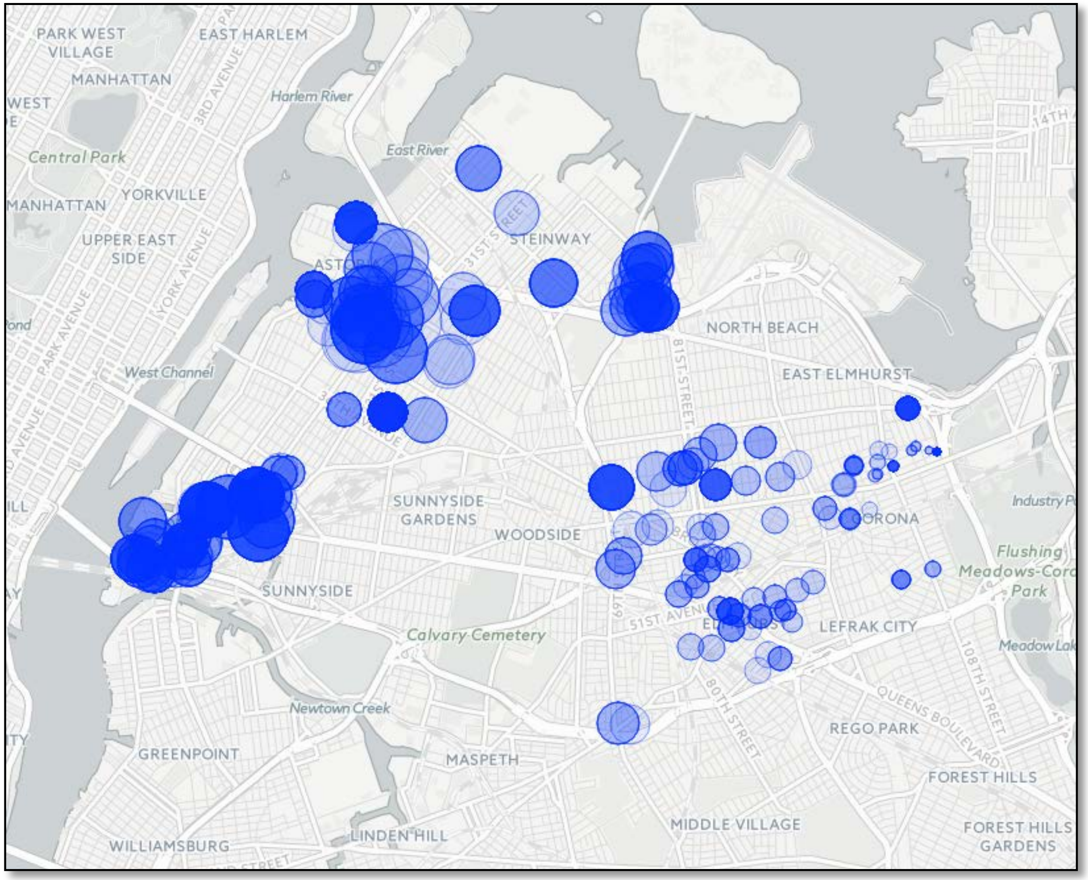
Geographically Weighted Regression (GWR)

- GWR local coefficients rendering: BLDG Story



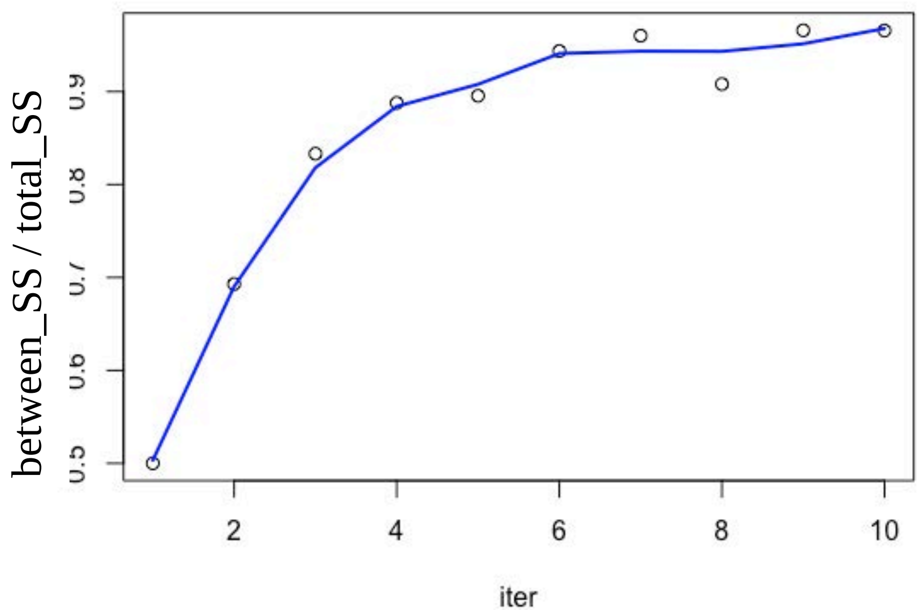
Geographically Weighted Regression (GWR)

- GWR local coefficients rendering: year built



Geographically Weighted Regression (GWR)

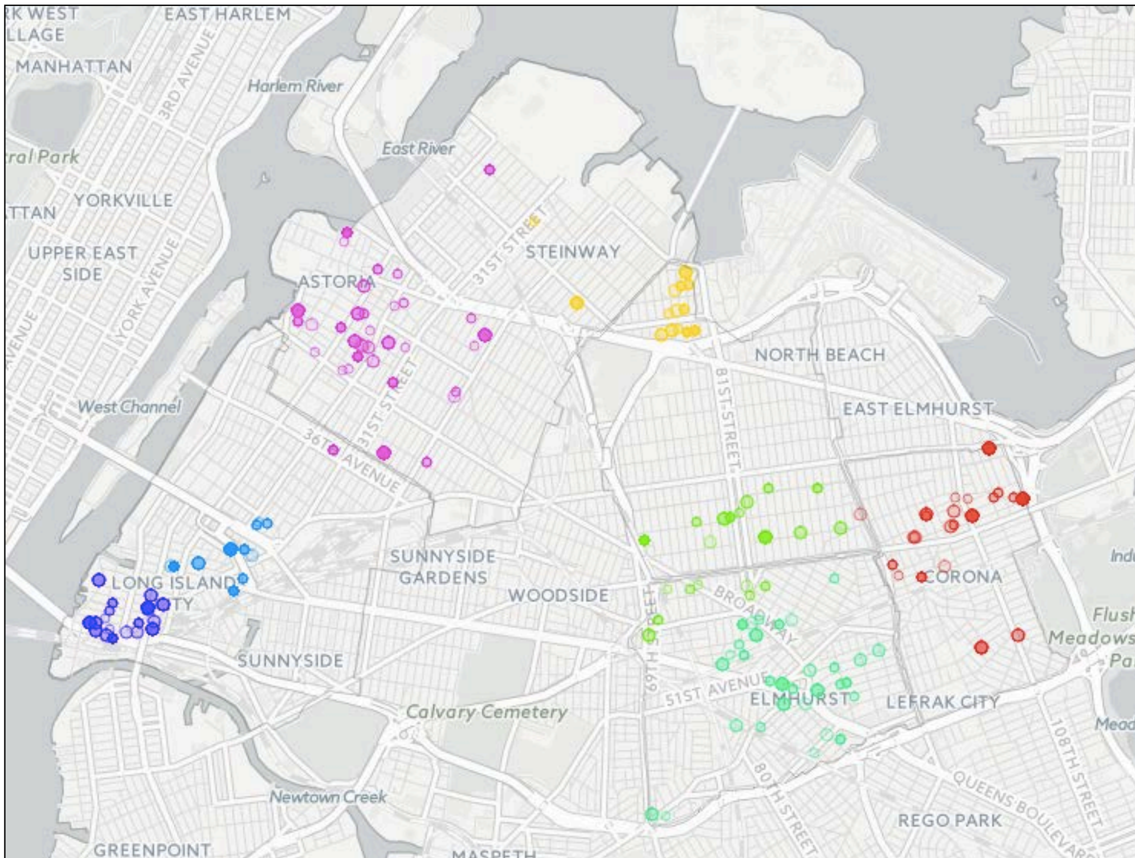
- K- means clustering
- Find optimized number of groups



Geographically Weighted Regression (GWR)

- Clustering result

Go to Interactive map



What's Next?

- Application on larger geographical area and expand sales data
- Look for systematic method to identify boundary considering the hierarchy of road system or other geo boundaries in the area
- Test in practice

References

- HNN - Cutting up NYC: The story of a market realignment http://www.hotelnewsnow.com/Articles/31935/Cutting-up-NYC-The-story-of-a-market-realignment
- We Asked You to Draw Your Own Neighborhood Map: Here Are the Results. Nigel Chiwaya - https://www.dnainfo.com/new-york/20150928/inwood/we-asked-you-draw-your-own-neighborhood-map-heres-what-you-did
- Gollini, Isabella, Lu, Binbin, Charlton, Martin, Brunsdon, Christopher, AND Harris, Paul. "GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models" *Journal of Statistical Software* [Online], Volume 63, Issue 17 (16 February 2015)
- Gaudart, Jean, Graffeo, Nathalie, Coulibaly, Drissa, Barbet, Guillaume, Rebaudet, Stanilas, Dessay, Nadine, Doumbo, Ogobara, AND Giorgi, Roch. " SPODT: An R Package to Perform Spatial Partitioning" *Journal of Statistical Software* [Online], Volume 63, Issue 16 (16 February 2015)
- Xianhong Xie (2013). rgcvpack: R Interface for GCVPACK Fortran Package. R package version 0.1-4. https://CRAN.R-project.org/package=rgcvpack
- Roger Bivand and Danlin Yu (2017). spgwr: Geographically Weighted Regression. R package version 0.6-29. https://CRAN.R-project.org/package=spgwr
- D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

References

- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.
- Hadley Wickham and Romain Francois (2016). dplyr: A Grammar of Data Manipulation. R package version 0.5.0. https://CRAN.R-project.org/package=dplyr
- Roger Bivand, Tim Keitt and Barry Rowlingson (2016). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.2-5. https://CRAN.R-project.org/package=rgdal
- Roger Bivand and Nicholas Lewin-Koh (2017). maptools: Tools for Reading and Handling Spatial Objects. R package version 0.8-41. https://CRAN.R-project.org/package=maptools
- David Robinson (2016). broom: Convert Statistical Analysis Objects into Tidy Data Frames. R package version 0.4.1. https://CRAN.R-project.org/package=broom
- Roger Bivand (2015). classInt: Choose Univariate Class Intervals. R package version 0.1-23. https://CRAN.R-project.org/package=classInt
- Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2016). Hmisc: Harrell Miscellaneous. R package version 4.0-2. https://CRAN.R-project.org/package=Hmisc
- Gregory R. Warnes, Ben Bolker and Thomas Lumley (2015). gtools: Various R Programming Tools. R package version 3.5.0. https://CRAN.R-project.org/package=gtools
- Ceccherini, Guido; Ameztoy, Iban; Hernández, Claudia P.R.; Moreno, Cesar C. 2015. "High-Resolution Precipitation Datasets in South America and West Africa based on Satellite-Derived Rainfall, Enhanced Vegetation Index and Digital Elevation Model." *Remote Sens.* 7, no. 5: 6454-6488. http://www.mdpi.com/2072-4292/7/5/6454
- Rabosky Daniel L, Michael C Grundler, Carlos J Anderson, Pascal O Title, Jeff J Shi, Joseph W Brown, Huateng Huang and Joanna G Larson. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution 5:701-707.


GIS-Pro 2017
October 23-26, 2017
Jacksonville, Florida


Assessment Leadership
Beyond All Limits
IAAO LAS VEGAS
ANNUAL CONFERENCE
& Exhibition
September 24-27
2017
MGM