VIS 2023

VDS@VIS 2023

NeighViz: Towards Better Understanding of Neighborhood Effects on Social Groups with Spatial Data

Yue Yu, Hong Kong University of Science and Technology
Yifang Wang, Northwestern University
Qisen Yang, Zhejiang University
Di Weng, Zhejiang University
Yongjun Zhang, Stony Brook University
Xiaogang Wu, New York University Shanghai
Yingcai Wu, Zhejiang University
Huamin Qu, Hong Kong University of Science and Technology

Introduction

- How local environments shape individual behavior (i.e., neighborhood effects) is a critical research question in social science.
- Identifying determinants and understanding such effects on social groups can promote social well-being and reduce spatial disparities.

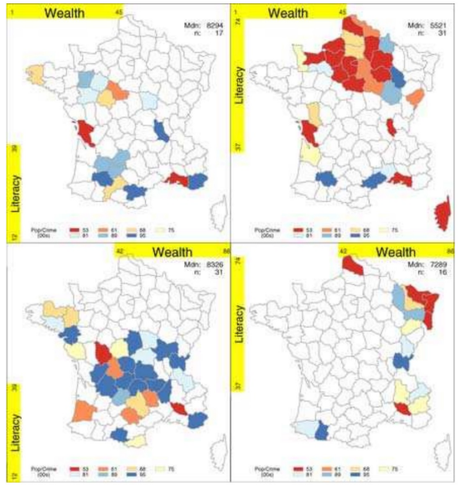
White voters' civic engagement due to neighborhood racial threat

R. D. Enos, 2016

Diverse neighbor exposure improve financial outcomes

R. Chetty et al., 2020

Neighborhood Effect Analysis

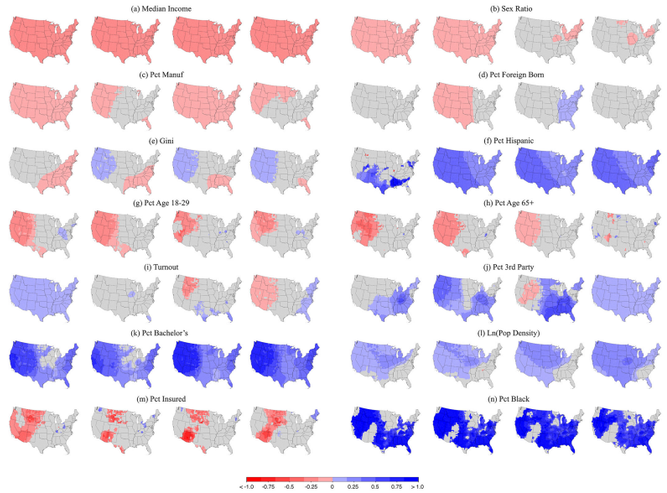


Conditioned Choropleth Maps

M. Friendly, 2007

spatial co-variance of variables

Fail to capture spatial dependence



Local coefficients of model

Z. Li and A. S. Fotheringham, 2022

spatial local model

Laborious to observe patterns in coefficients

	1			2			3			4	
	200 Meters (n=2727)			**250 Meters (n=2727)**			**350 Meters (n=2727)**			**450 Meters (n=2727)**	
VARIABLES	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min
Exposure (ref: nonexposed)											
Exposed	1.30	1.15	1.48	1.38	1.24	1.46	1.27	1.19	1.49	1.24	1.07
Age	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.98
Years of schooling	0.98	0.96	0.98	0.98	0.96	0.98	0.98	0.96	0.98	0.98	0.96
Economic status (ref: employed)											
Unemployed/inactive	1.48	1.30	1.51	1.49	1.30	1.52	1.49	1.29	1.52	1.48	1.28
Logged household income	0.88	0.86	0.89	0.88	0.86	0.89	0.88	0.86	0.89	0.88	0.86
Previous symptoms (ref: no)											
Yes	3.94	3.86	4.41	3.95	3.88	4.43	3.97	3.90	4.45	3.97	3.90
Constant	0.90	0.75	1.71	0.87	0.72	1.69	0.86	0.71	1.66	0.85	0.68
AIC	2512.41			2510.02			2511.29			2511.85	

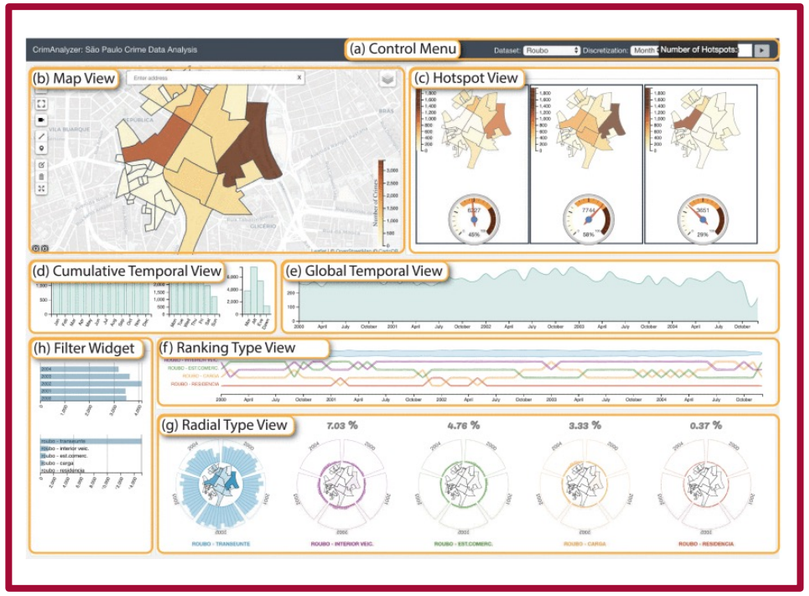
Summary of models

D. Zeng and X. Wu, 2021

hypothesis-driven & predefined social group

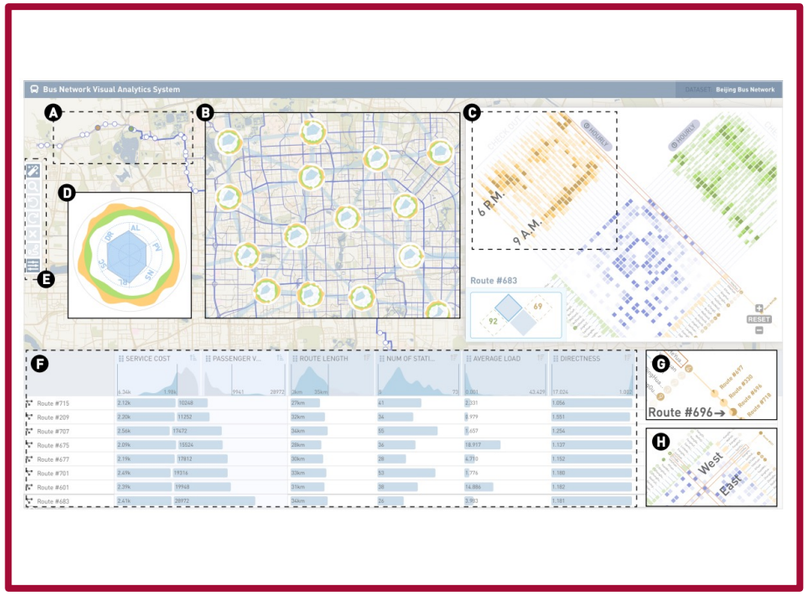
limited insights from large individual datasets

Spatial Data Visualization



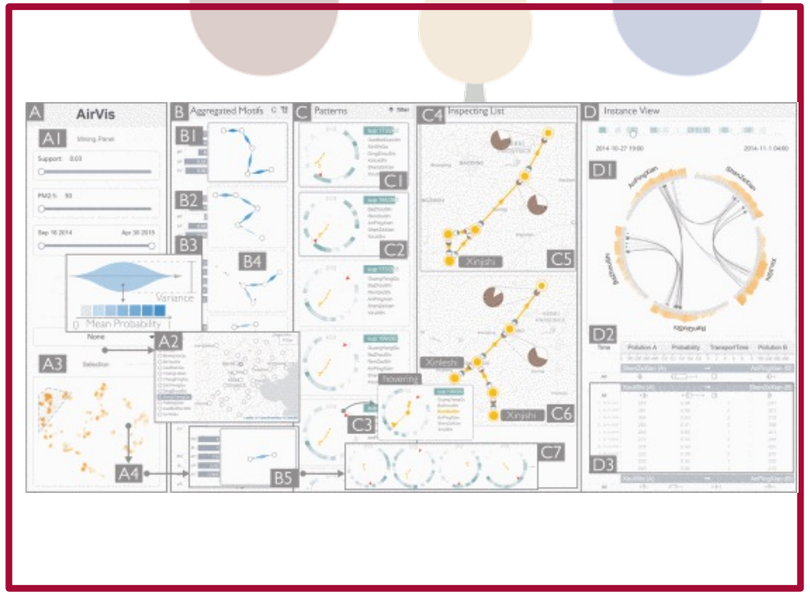
G. García et al., 2019

Pattern of crime hotspots



D. Weng et al., 2020

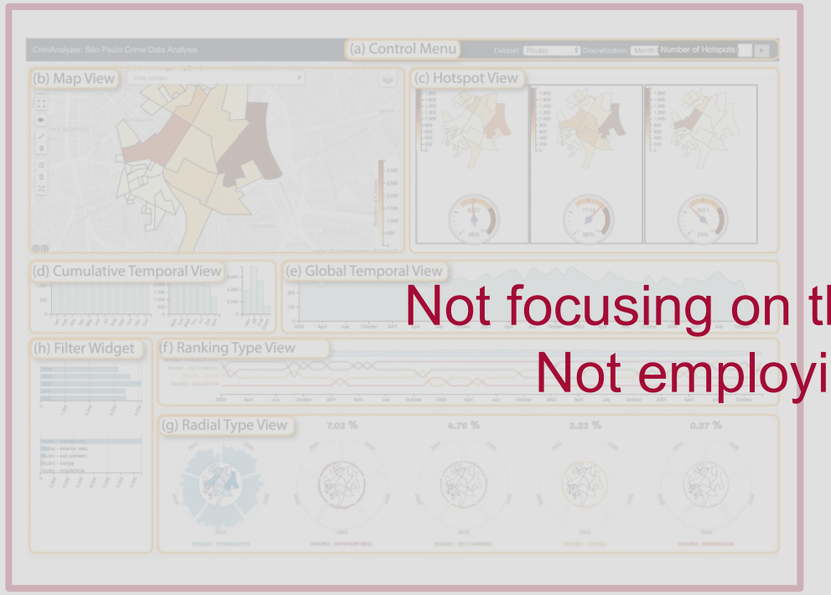
Planning of bus transit routes



Z. Deng et al., 2021

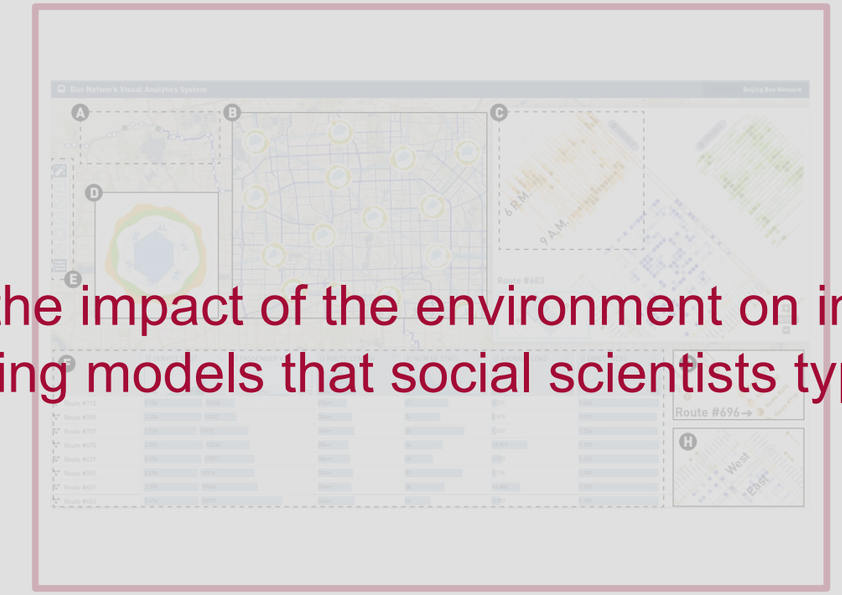
Propagation of air pollution

Spatial Data Visualization



G. García et al., 2019

Pattern of crime hotspots



D. Weng et al., 2020

Planning of bus transit routes

D. Zeng and X. Wu, 2021

Propagation of air pollution

Not focusing on the impact of the environment on individual behavior
Not employing models that social scientists typically adopt

Challenges

- Lack of a coherent **workflow** for multivariate spatial analysis.
- Difficulty in **visually presenting** complex spatial and social relationships.
- Lack of an interactive visualization **system** for social scientists.

Contributions

An analytical framework for multivariate spatial analysis.

Visualizations to reveal neighborhood effects and inter-group differences.

NeighViz, an interactive visual analysis system for social scientists.

Task Analysis

- **Model Generation**
  - **T1** Selecting variables for the spatial model.
  - **T2** Identifying social groups for further exploration.

- **Geographical Exploration**
  - **T3** Exploring geographical distribution.
  - **T4** Detecting areas with potential neighborhood effect.

- **Comparisons and Verifications**
  - **T5** Comparing across social groups.
  - **T6** Verifying and explaining neighborhood effects.

Data Description (1/3): Data Source

https://l2-data.com/

L2 Voter and Consumer Data

Microdata that contains information of 180 million registered voters in the United States

- Partisan
- Voting
- Consumption

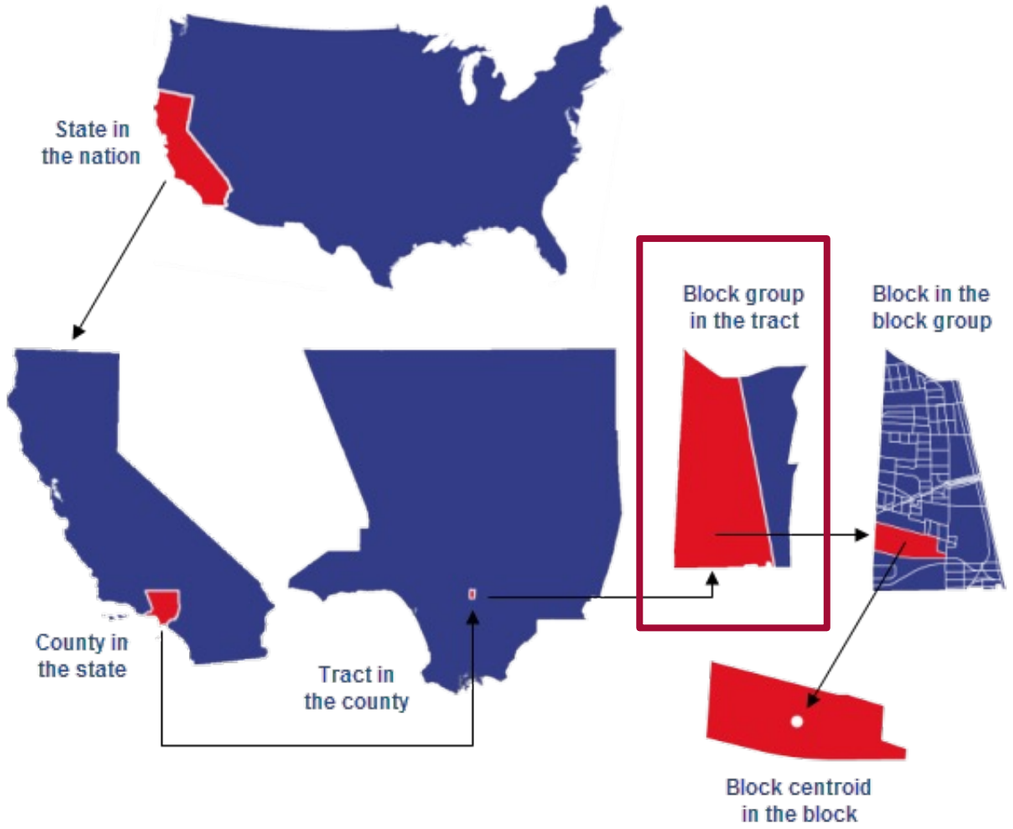
https://www.census.gov/programs-surveys/acs

American Community Survey 5-year Census Data

Aggregated dataset that contains survey data at different geographic levels in the United States

- Edu level
- Income level
- Race

Data Description (2/3): Spatial Unit



https://learn.arcgis.com/en/related-concepts/united-states-census-geography.htm

Basic Spatial Unit: Census Block Group (CBG)

250 - 550 housing units

600 - 3,000 people

Data Description (3/3): Data Preparation & Query

Processed offline

L2 BETTER DATA BETTER DECISIONS

Aggregate to Subgroup Dataset for privacy

CBG	race	voted in 2016	…	population
000001	European	Yes	…	1500
000001	European	No	…	500
000002	Asian	Yes	…	100

Each row is a minimal group of individuals who share similar characteristics in the same CBG

Queried on-the-fly

Race = "Asian" & Gender = "Male"

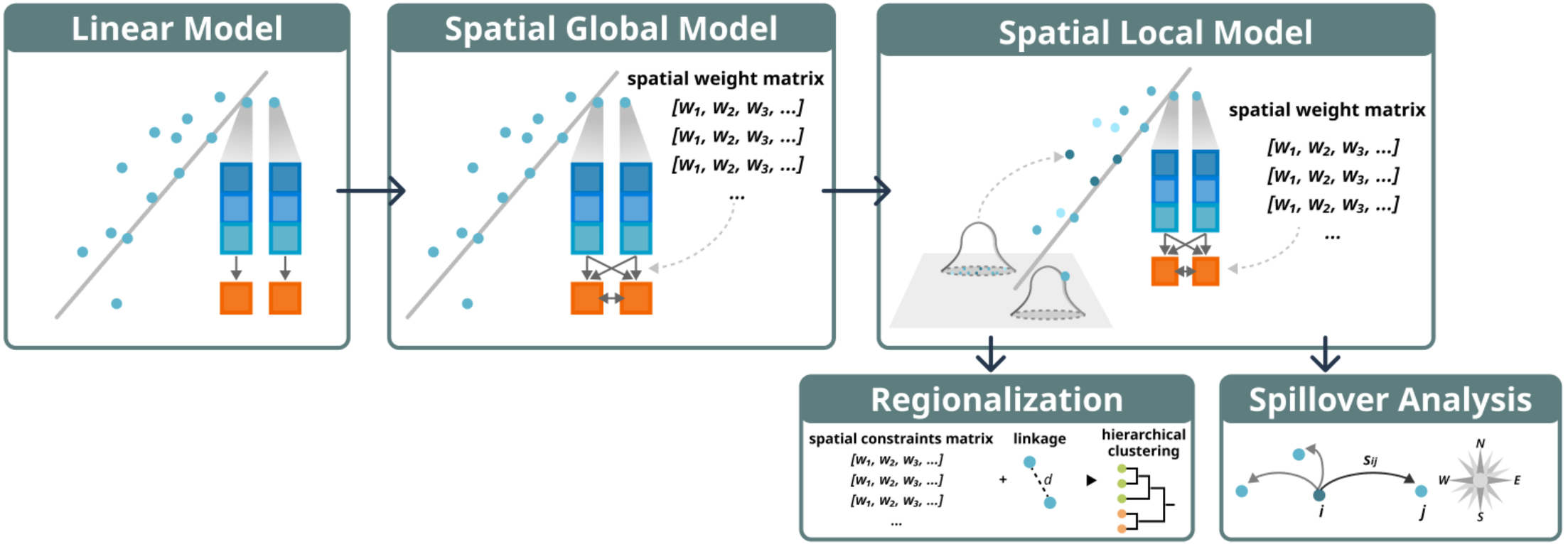
Query subgroup indicated by user

Join census statistics

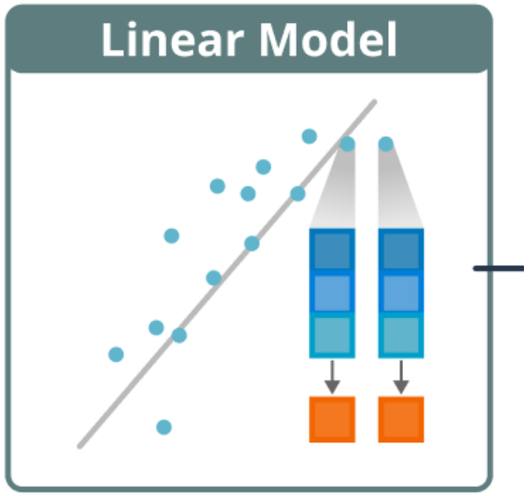
CBG	Voting rate 2016	College ratio	…	Census Asian ratio	Census European ratio	…
000001	0.7	0.5	…	0.3	0.6	…
000002	0.6	0.4	…	0.4	0.3	…
000003	0.3	0.2	…	0.5	0.2	…

Each row is a CBG's statistics of the queried subgroup and census data

Data Model: Overview

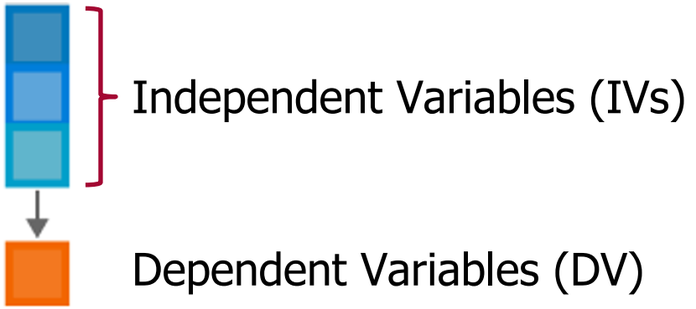


Data Model (1/4)

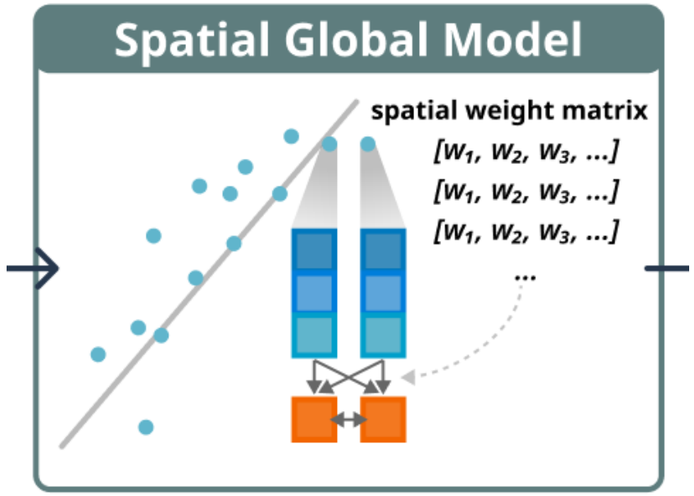


Ordinary Least Square (OLS)

- Baseline model to explore variable relationships
- Each data point is a geographic unit
- Treats spatial units as independent observations, neglecting spatial dependencies

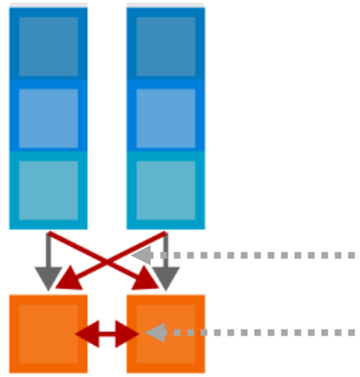


Data Model (2/4)



Spatial Durbin Model (SDM)

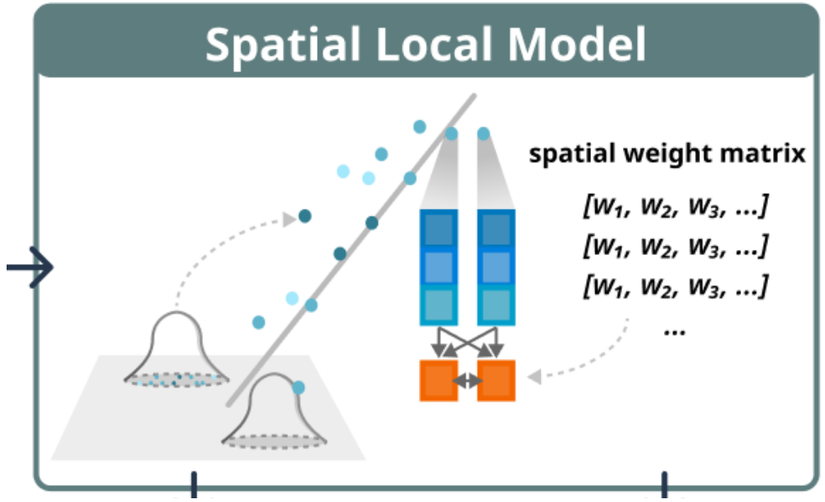
- Capture spatial dependencies (influence from neighboring spatial units)
- Cannot effectively address spatial heterogeneity due to location-specific factors (e.g., culture and history)



Spatially lagged independent variables of neighboring units

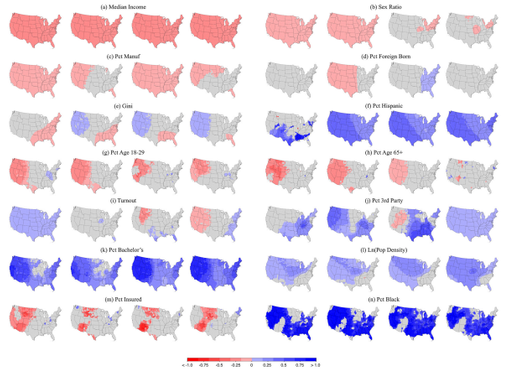
Spatially lagged dependent variable of neighboring units

Data Model (3/4)



Geographically Weighted Regression (GWR)

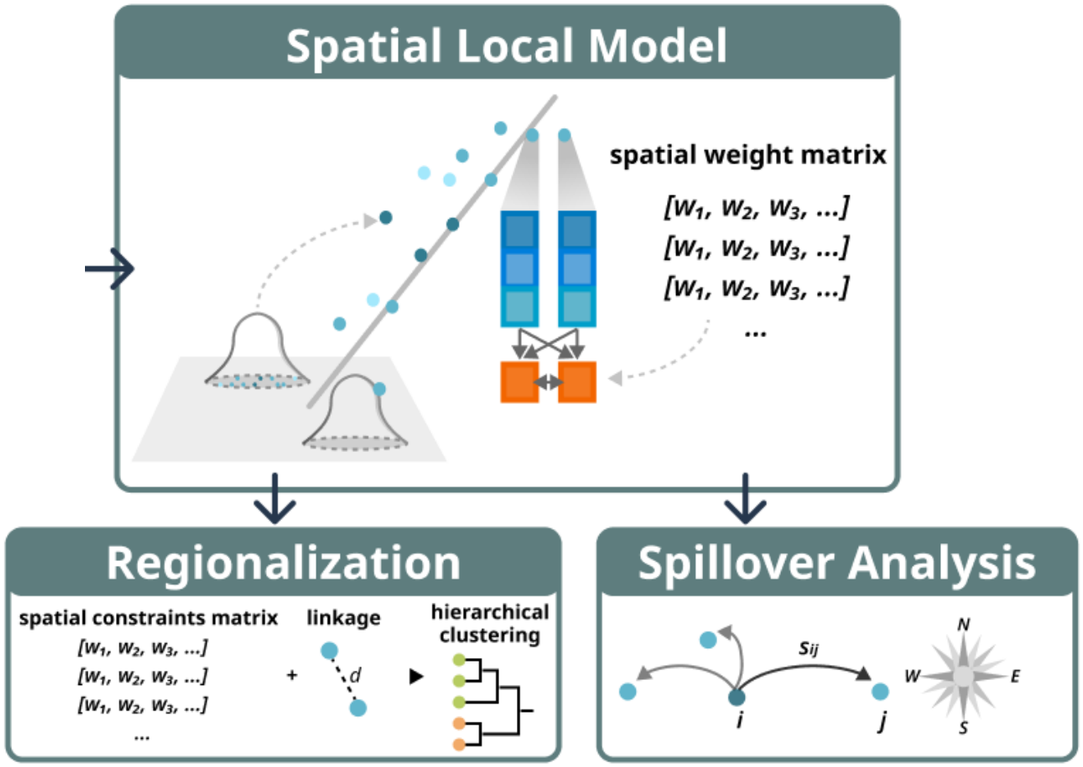
- Explore spatially varying neighborhood effects over places
- A set of regression coefficients for each spatial unit (hard to explore by manual inspection)



Z. Li and A. S. Fotheringham, 2022

Data Model (4/4)

Techniques to help interpret the output of GWR



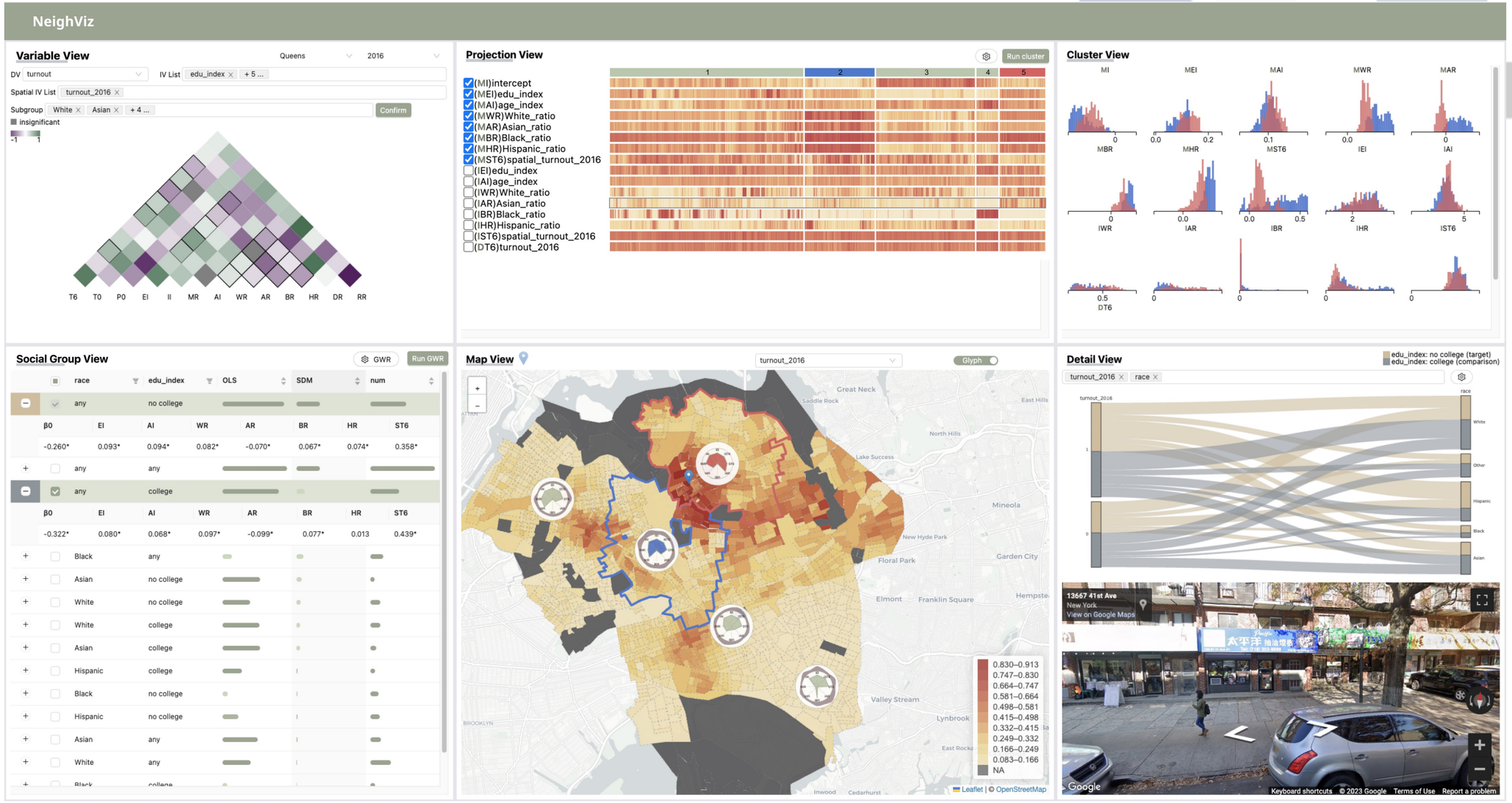
Regionalization

- Facilitate the discovery of latent neighborhoods with similar characteristics

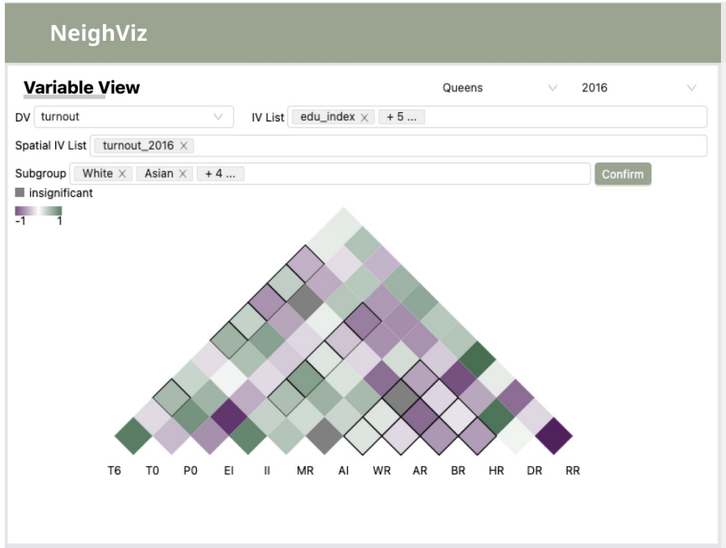
Spillover Analysis

- Explore dynamics of the spillover effect in different regions

System Design: Overview



System Design (1/4)



Variable View

Build a spatial model (T1)

Social Group View

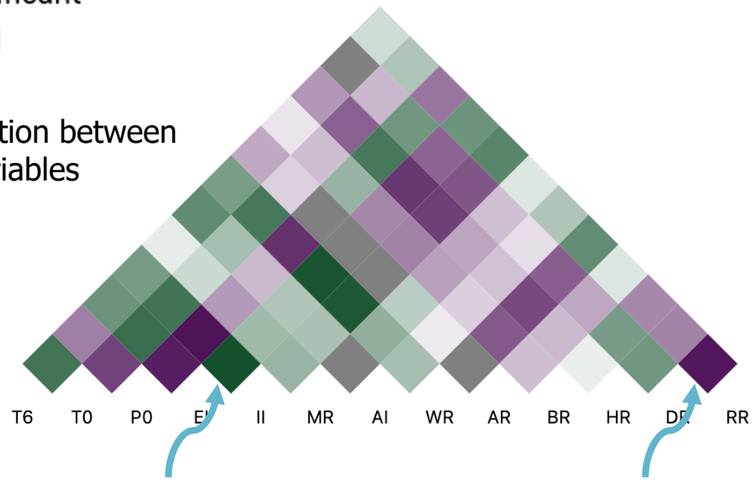
Identify social groups for in-depth analysis in a *LineUp Table* (T2)

Visual Encoding

insignificant

-1 1

Pearson correlation between each pair of variables



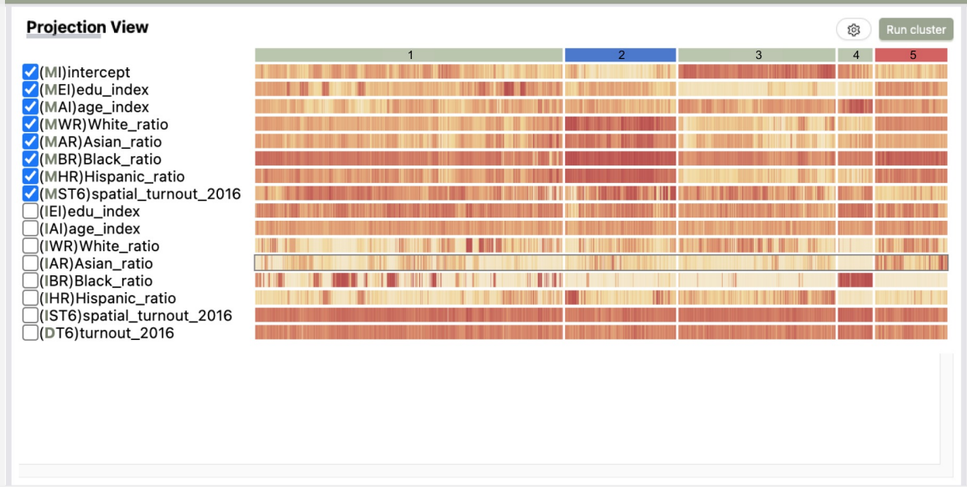
Strongly positive

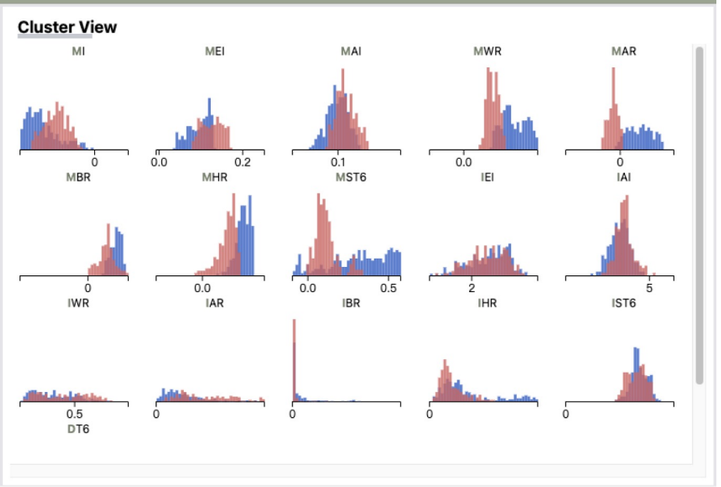
Strongly negative

System Design (2/4)

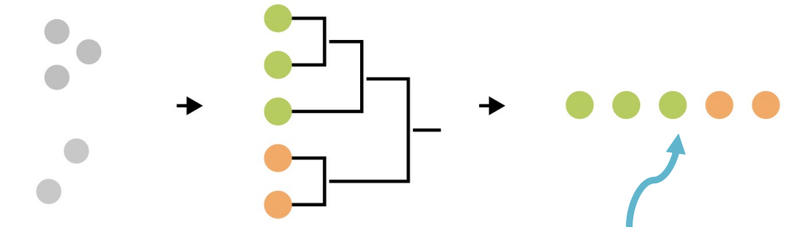
Projection View

Show the spatial distribution of attributes on *Projection Bars* (T3)



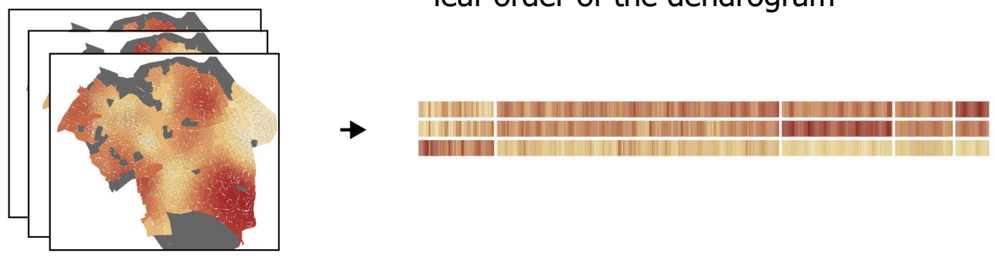


Visual Encoding



Spatially Constrained Hierarchical Clustering

leaf order of the dendrogram

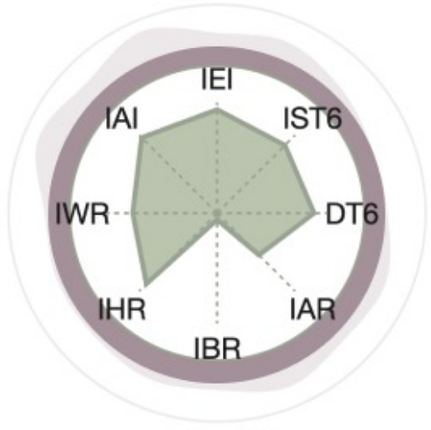


Cluster View

Compare the variable distributions across clusters with *Density Histograms* (T3)

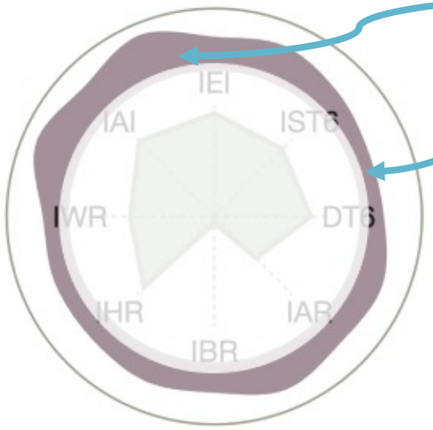
System Design (3/4)

Visual Encoding



Inner Layer

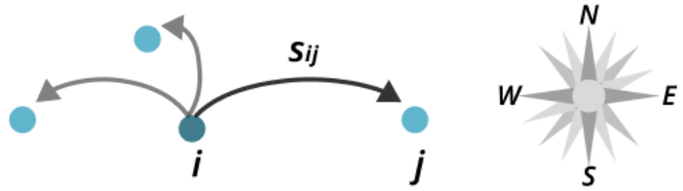
Cluster's mean values displayed on a *Radar Chart*



Outer Layer

Strong average spillover along **NW** direction

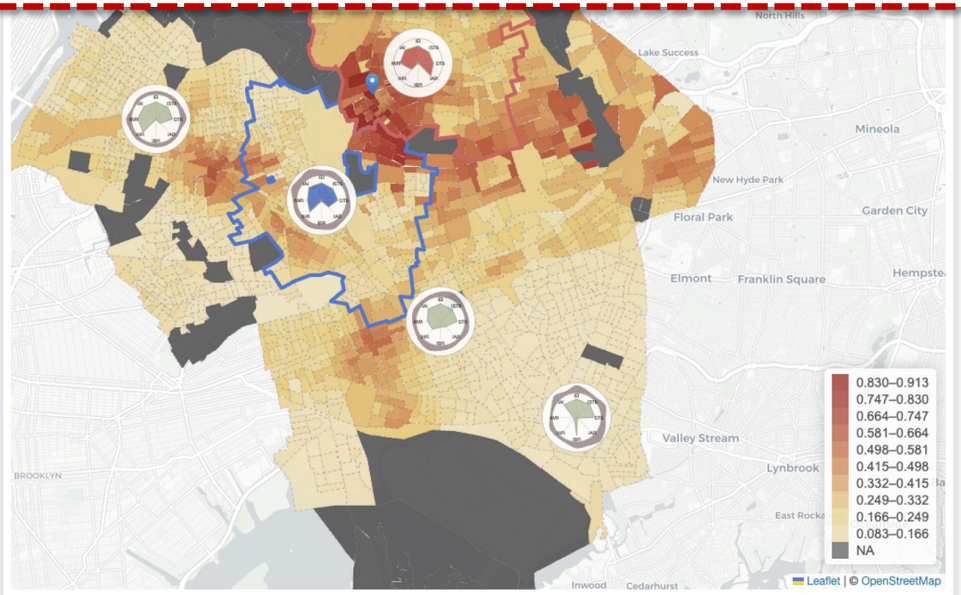
Weak average spillover along **E** direction



Map View

Provide spatial details on the *Map* (T3)

Show spillover effect in clusters with *Glyphs* (T4)



System Design (4/4)

Detail View

Compare social group data with the *Parallel Sets Chart* (T5)

Visually show the neighborhood context with *Google Street View* (T6)

Case Study

- Invited a political scientist to freely explore the system
- Focused on Queens County, New York State
- Relationship between racial diversity and the neighborhood effect on the political engagement

NeighViz

Model Diagnositcs

Variable View

Weight | Queens | 2016

DV | IV List

Spatial IV List

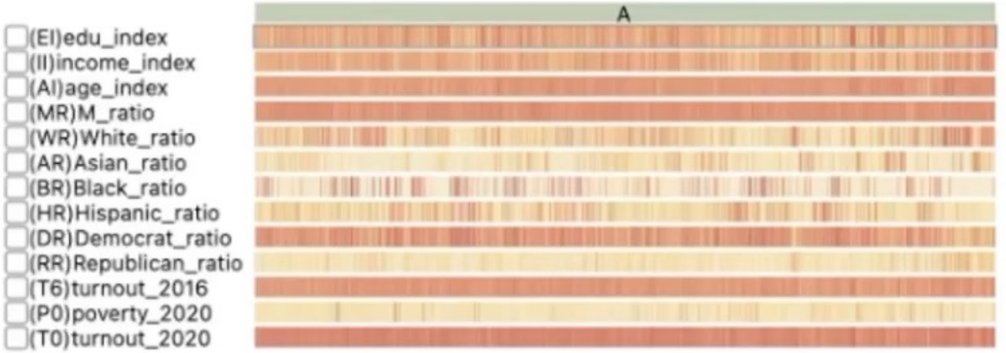
Subgroup | Confirm

insignificant

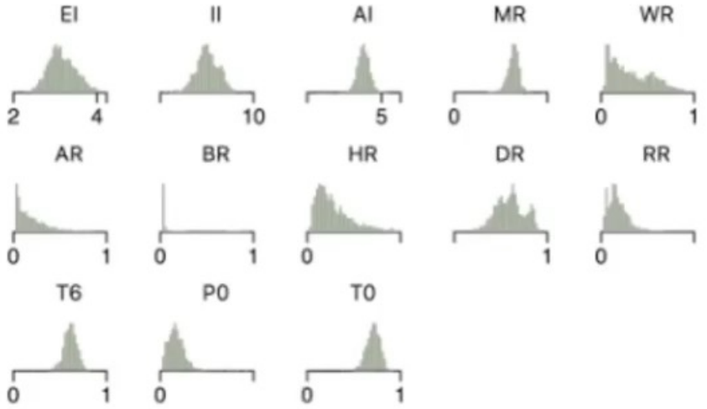
-1 1

Projection View

Clustering | Run cluster



Cluster View



Social Group View

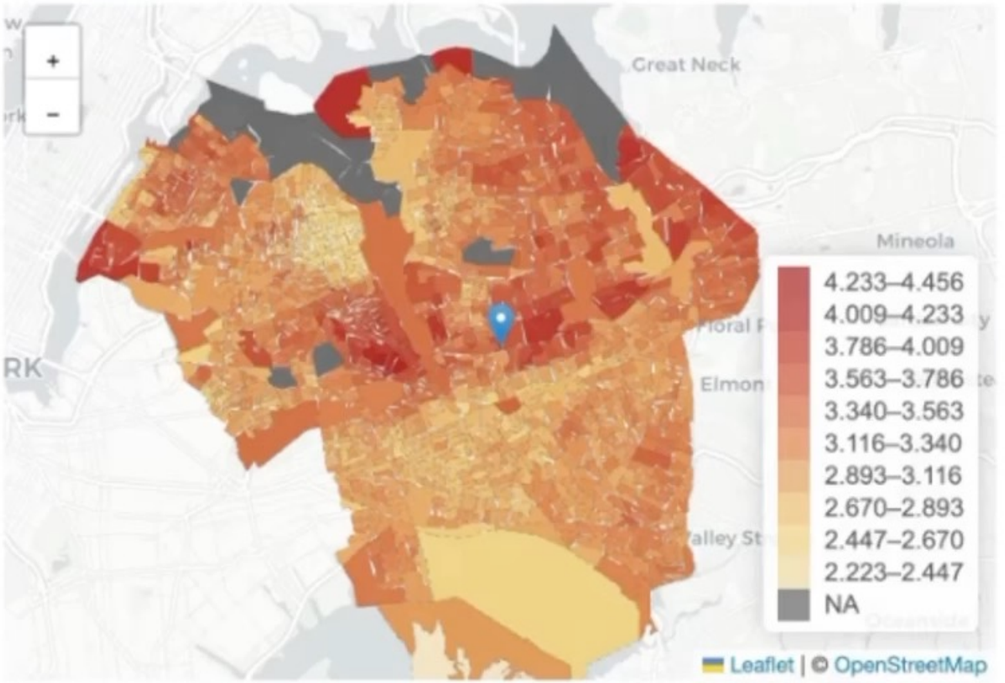
GWR | Run GWR

OLS	SDM	num

No data

Map View

No Glyph



Detail View

Conclusions

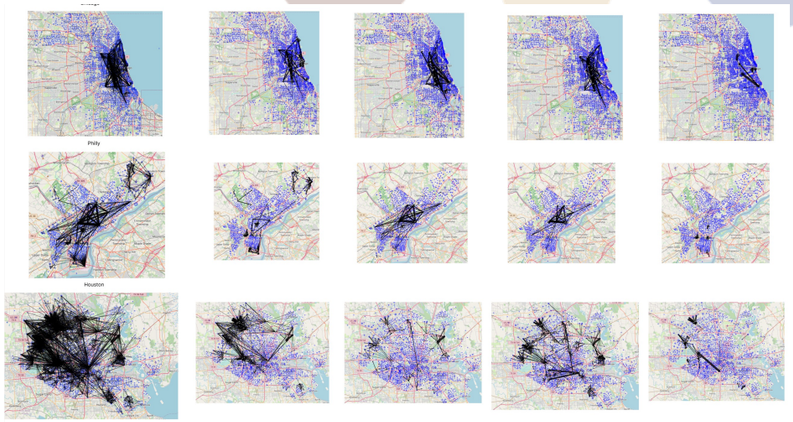
An analytical framework for multivariate spatial analysis.

Visualizations to reveal neighborhood effects and inter-group differences.

NeighViz, an interactive visual analysis system for social scientists.

Future Works

1. (WIP) Integration of mobility data. Consider the neighborhood effects in daily activity space
2. Enhance the capability to handle time-varying data
3. Incorporate datasets from various research domains



VIS 2023

Thank you!

Slides available on https://bruceyyu.github.io/