Measuring spatial polarization and inequality in cities using housing transaction big data

Jinhyung Lee[1,3], Harvey J. Miller[1,3], Nicholas B. Irwin[2] & Elena G. Irwin[2,3]

[1)] Department of Geography

[2)] Department of Agricultural, Environmental, and Development Economics

[3)] Center for Urban and Regional Analysis (CURA)

The Ohio State University

Great Recession in the U.S.

- Economic earthquake
  - The U.S. labor market lost 8.4 million jobs

Recovery from recession

- Positive signs
  - Increase in GDP percent change
  - Urban revitalization
- However, **unequal distribution of benefits**

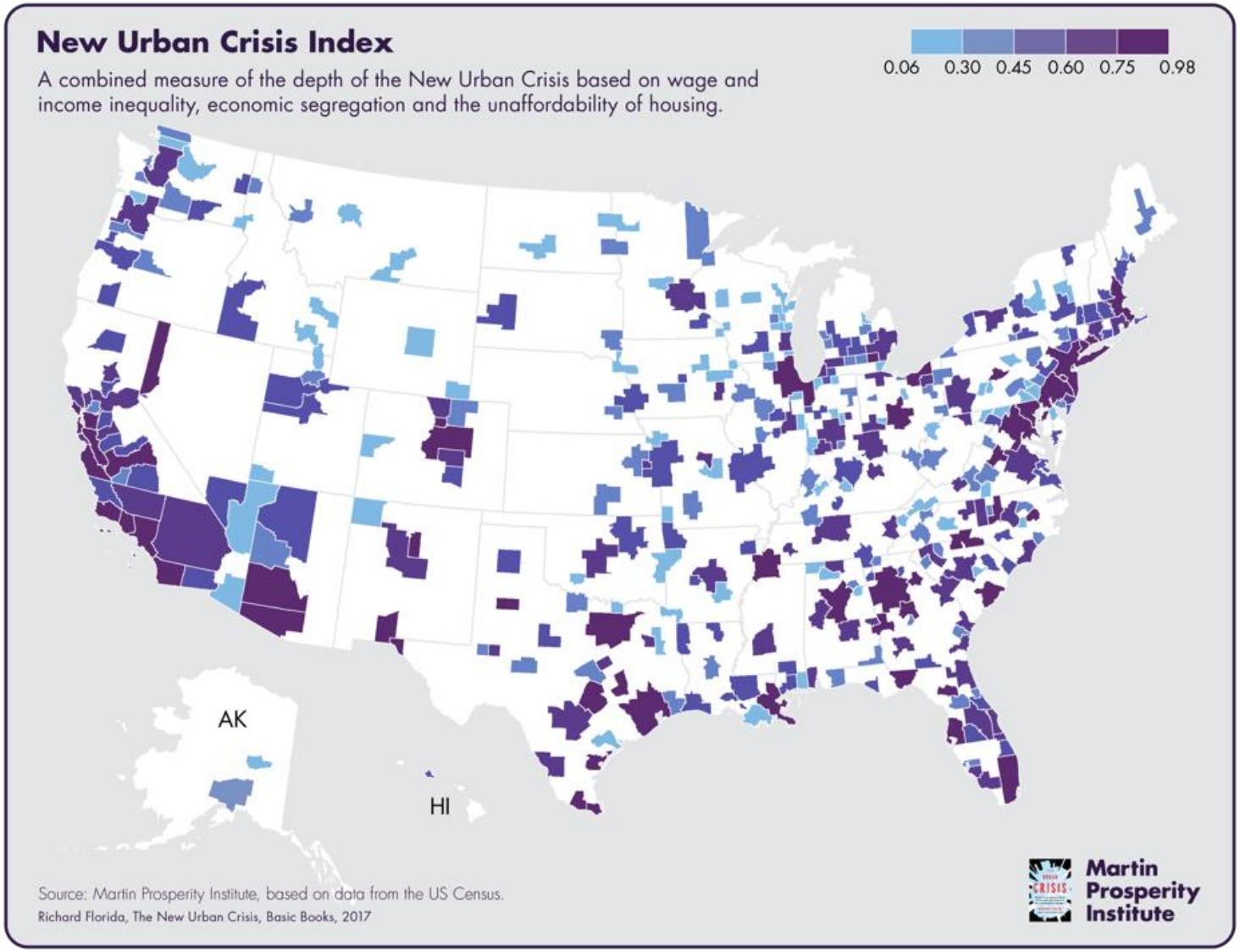
  : concentration of wealth and economic growth in small parts of city

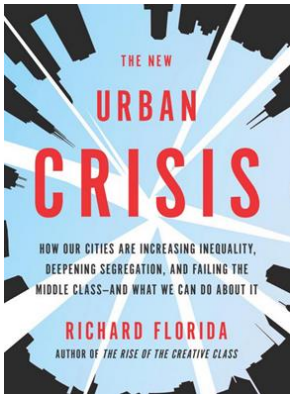
Possible outcomes

1. Deepening **spatial polarization** (clusters of the rich vs. poor)
2. Increasing **inequality** (e.g., income)

➡ New Urban Crisis! (Florida, 2017)

Geography of "New Urban Crisis"





How have polarization and inequality in cities **evolved**?

Research objectives

1. Investigate the evolution of **spatial polarization** in the U.S. cities
2. Investigate the evolution of **inequality** in the U.S cities

- Study area
  - State of Ohio: as a pilot study
  - Cities with different economic growth trajectories:
    - Columbus (high), Cincinnati (medium), Cleveland (low)
- Methods
  - Big data analysis using high-resolution housing transaction data
  - Geovisualization of polarization over time
  - Measuring spatial polarization (Moran's I) and inequality (Gini coefficient)

Data

Housing transaction data

- High spatial resolution: parcel-level
- Time range: 2000-2015
- Big data: # of observations in Columbus: 482,478
- City boundary: Urbanized Areas (UAs)

Data pre-processing

1. Removing housing transactions missing key characteristics
2. Removing outliers
   - Housing transactions less than $10,000 and more than $2.5 million are dropped

Housing price hedonic model

- Model specification
  - Estimate hedonic price models to measure location premium
    - Disaggregate price into attributes of house

$$lnP_{it} = \beta_0 + \beta_X \mathrm{X}_i + \beta_Q Quarter_q + \epsilon_{it}$$

- Natural log of price for house i in period t
- Function of a vector of house characteristics (X_i) and a quarter-by-year fixed effect ($Quarter_q$)
- Remaining unexplained variation → Location premium

Variables	Cincinnati	Cleveland	Columbus
House size (sqft)	0.000564	0.000660	0.000573
House size squared	-4.36e-08	-4.33e-08	-3.28e-08
Parcel size (acres)	0.173	0.0948	0.170
Parcel size squared	-0.0189	-0.0126	-0.0178
Age	-0.000318	0.00276	-0.0109
Age squared	-5.04e-05	-0.000100	5.94e-05
Bedrooms	-0.0563	-0.0511	-0.152
Full bathrooms	0.135	0.0347	0.145
Stories	-0.0149	-0.106	-
Pool	0.0420	-0.162	0.0904
Finished basement	-0.0517	0.0763	-0.150
Fireplace	0.192	0.239	0.116
Central air conditioning	0.130	0.194	0.581
Garage	0.138	0.0700	-0.0417
Constant	10.38	10.70	10.60
R-squared	0.475	0.526	0.447
Observations	414,665	512,844	482,477
Quarter-year fixed effects	Yes	Yes	Yes

Note: LogPrice is the dependent variable. All variables are significant at 0.01 levels. Stories is not reported for Columbus due to a lack of observations.

Geovisualization of polarization

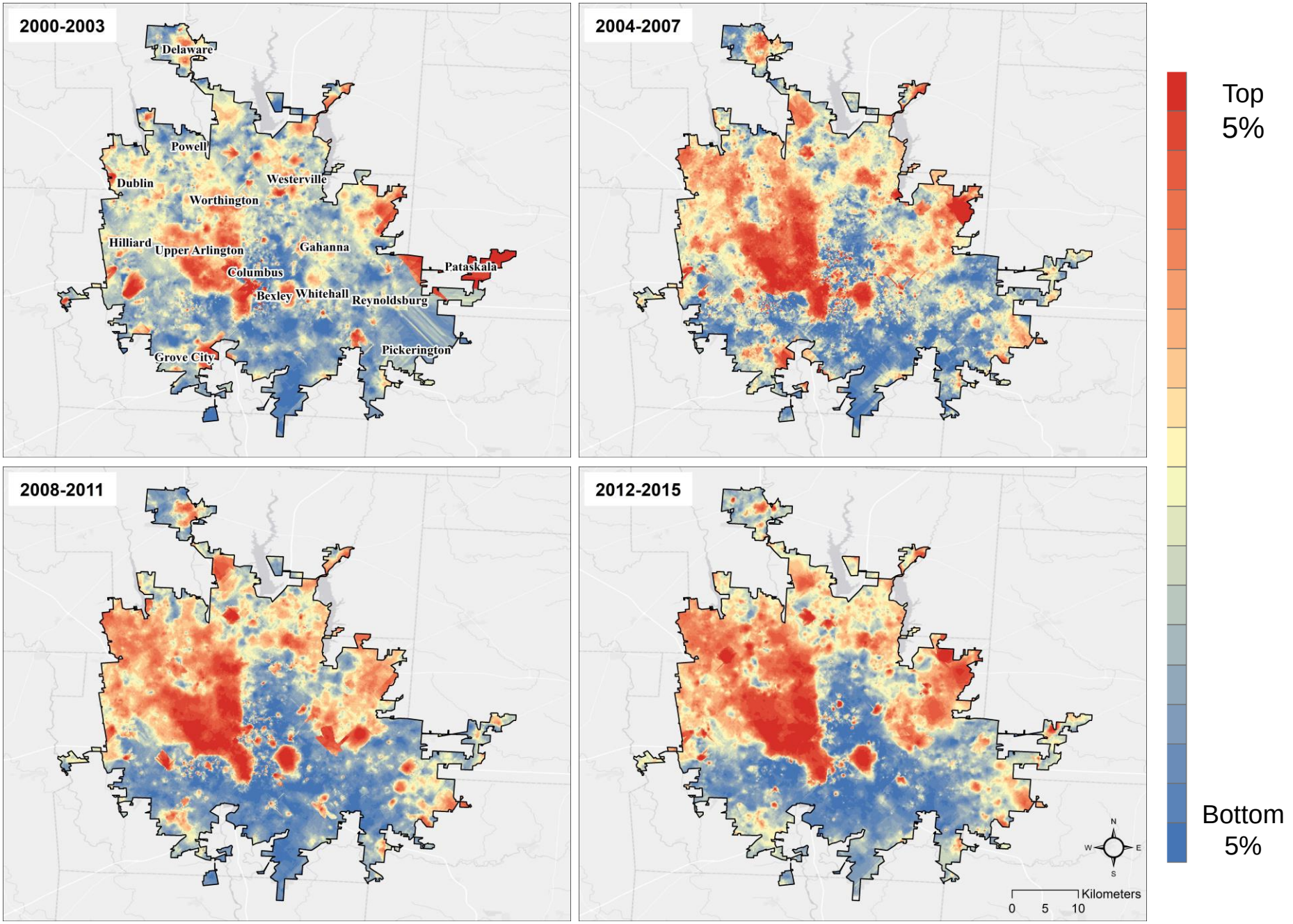
- Spatial interpolation: Kriging
  - Creating surfaces of location premium over time
  - Ordinary Kriging / Exponential semivariogram model (lowest RMSE)

Measuring polarization and inequality

- Moran's I values
  - Measure spatial polarization of location premium
  - Unit of analysis: Census block group
- Gini coefficients
  - Measure inequality of location premium

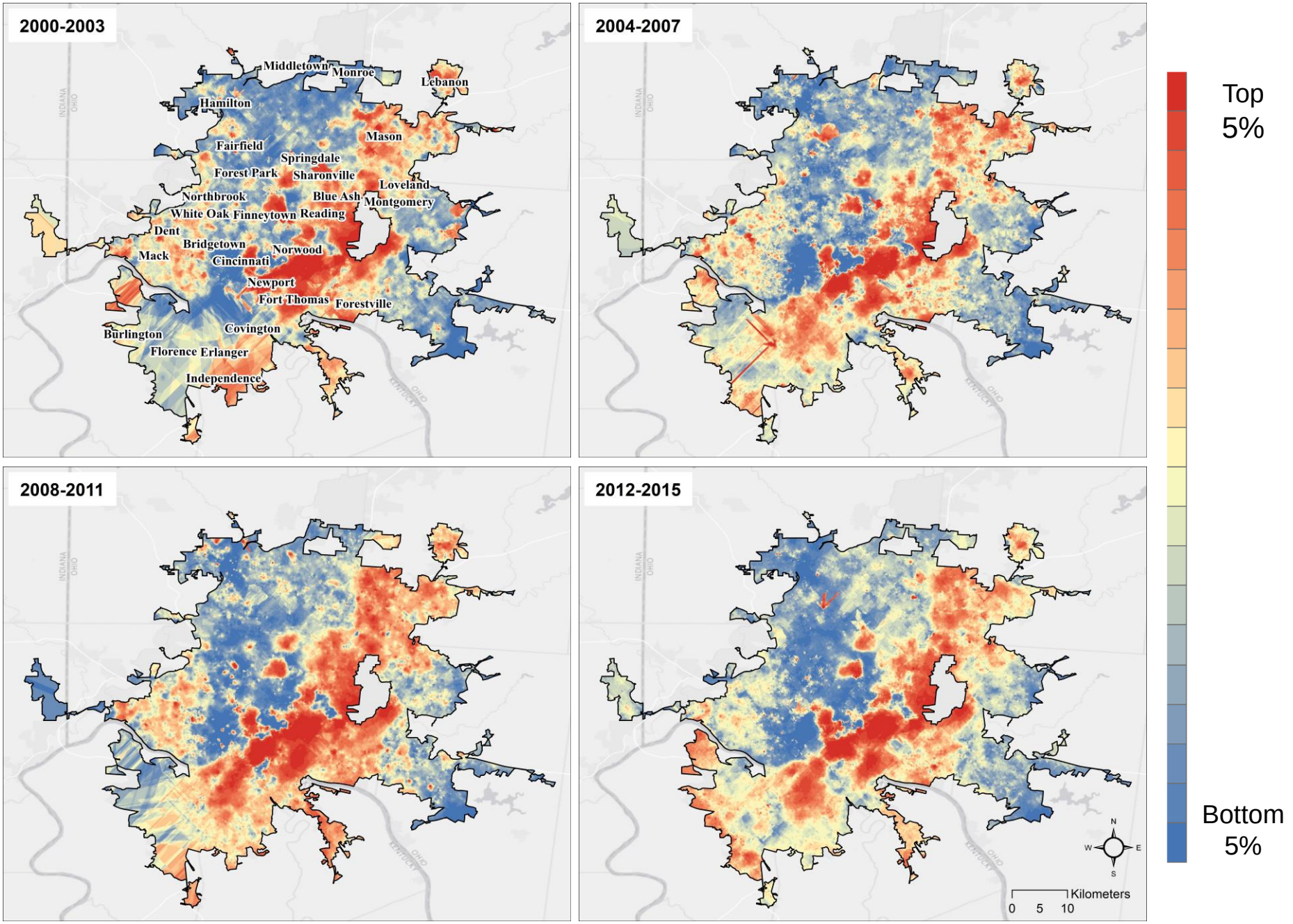
Location premium surfaces

Columbus



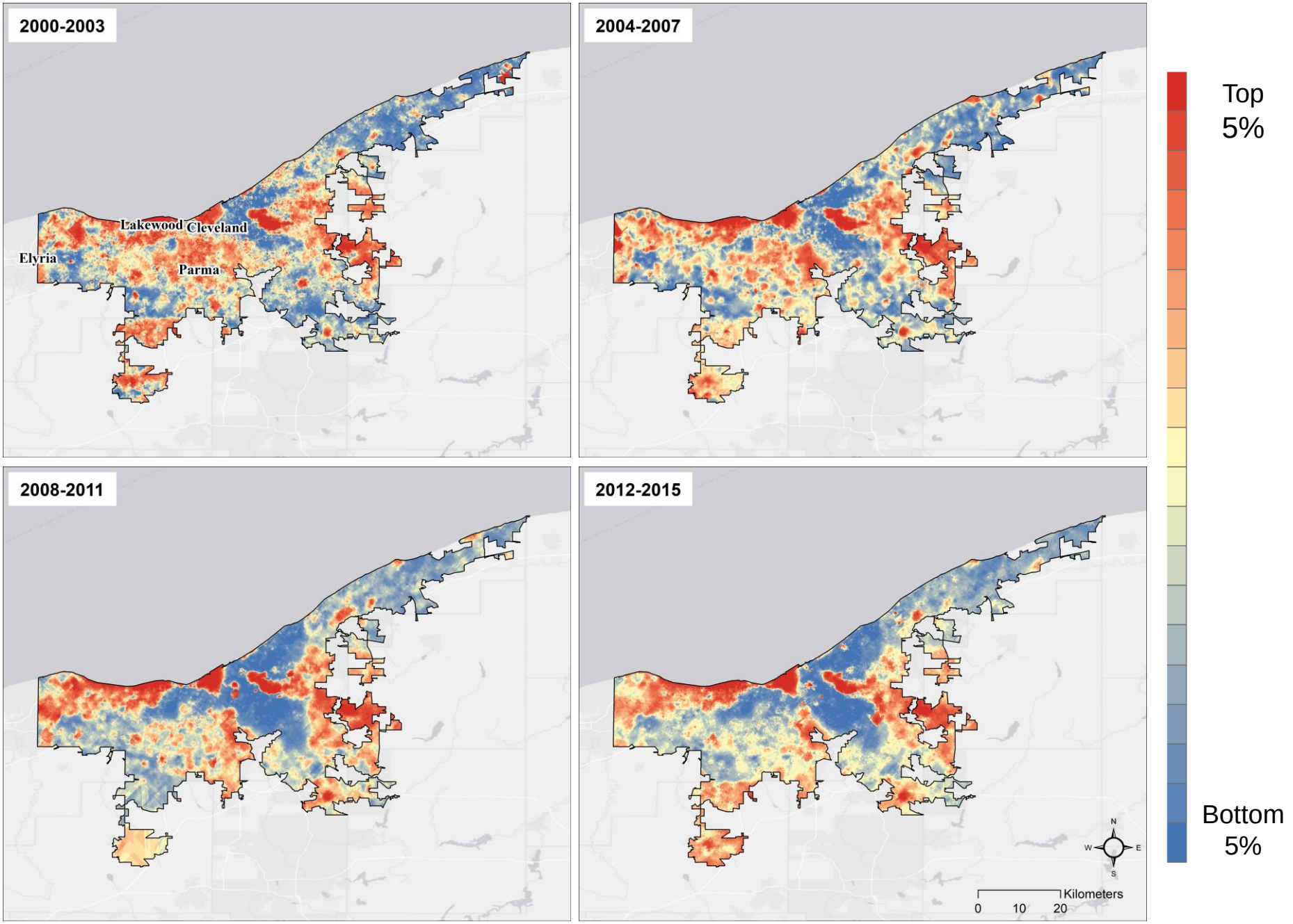
Location premium surfaces

Cincinnati



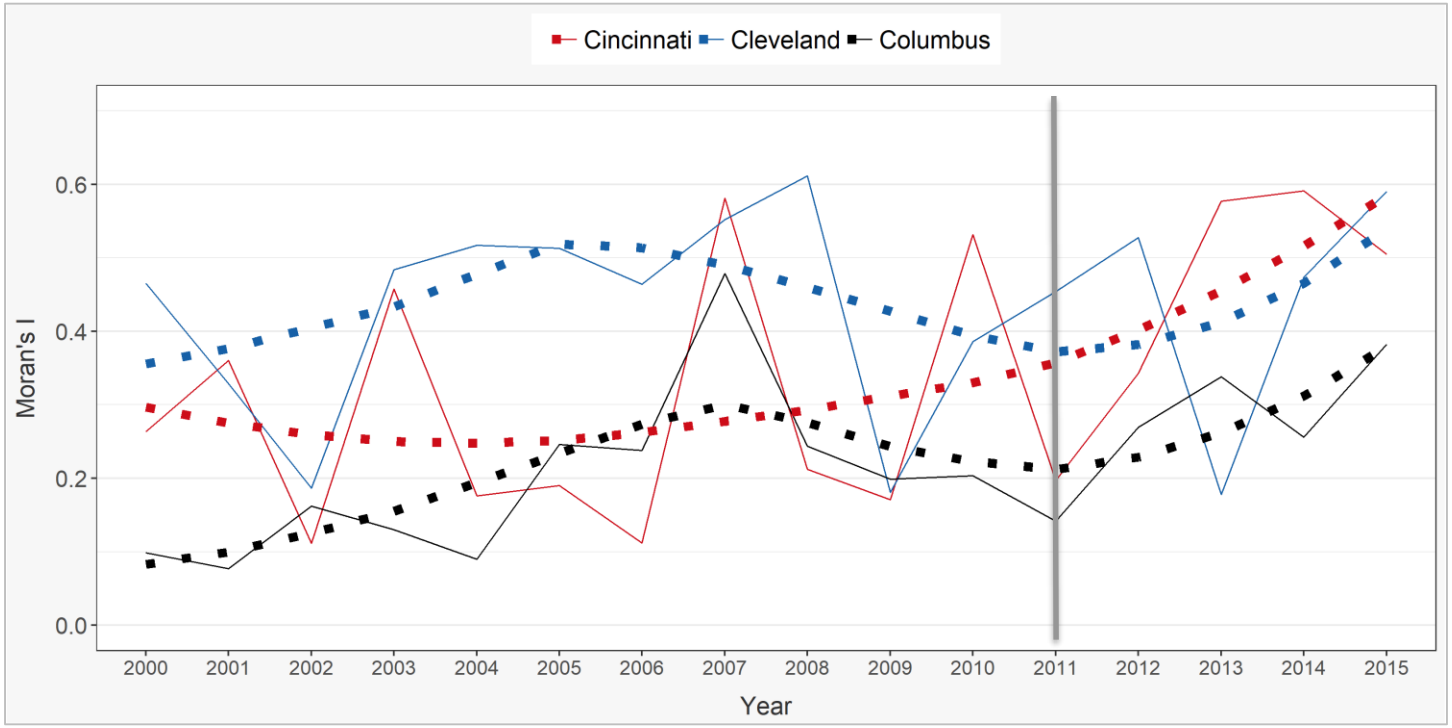
Location premium surfaces

Cleveland



Measuring spatial polarization: Moran's I

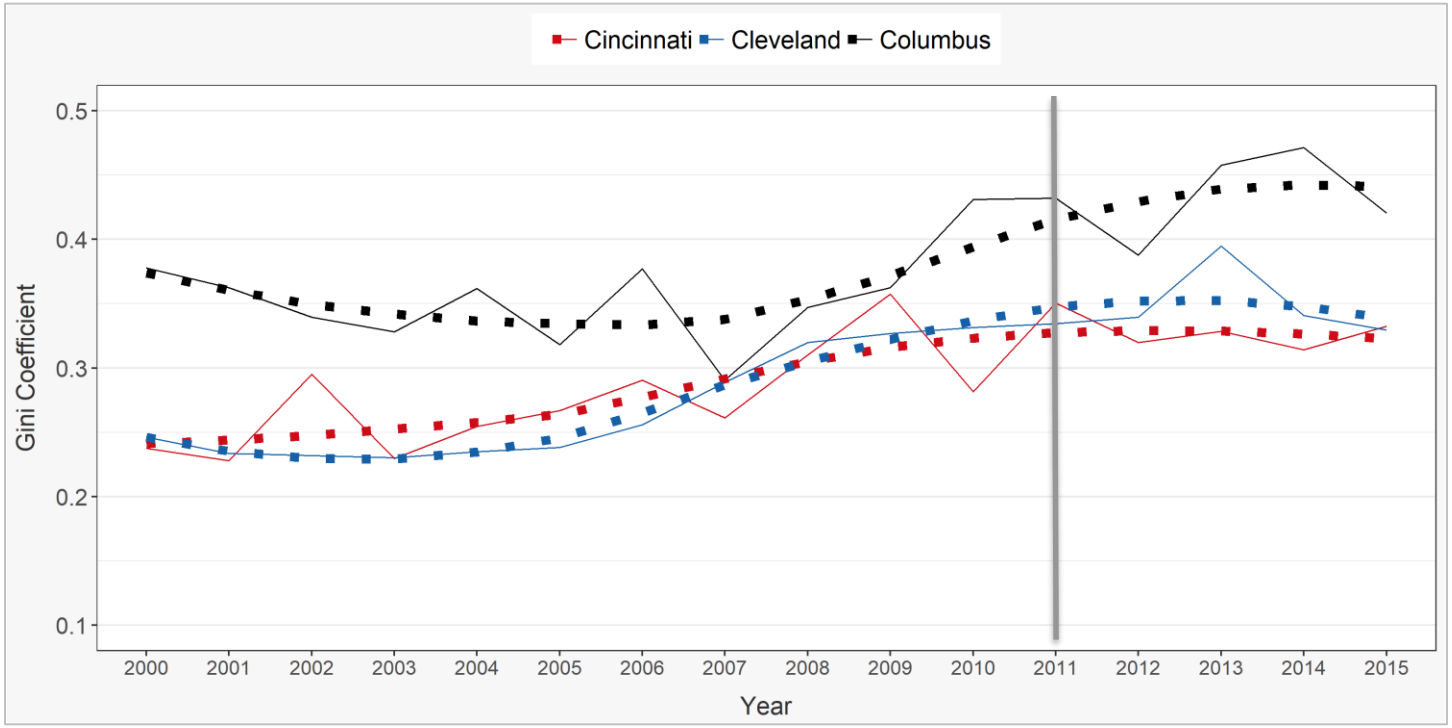
Results



- Increasing spatial polarization trends
- Peak right before GR (2007), sharp drop after GR
- Growing spatial polarization trends during recovery (after 2011)

Measuring inequality: Gini coefficient

Results



- Similar to the Moran's I values, general increasing trends of inequality

Conclusion

- Three biggest Ohio cities are becoming more polarized!
  - **Location premium surfaces:** clustering of high (or low) values → Divided cities
  - **Spatial polarization**: increasing trends in Moran's I values
  - **Inequality**: growing trends in Gini coefficients

Implications

1. **Methodological**: unprecedent high-resolution data
2. **Empirical**: evidence of growing polarization / inequality trends
3. **Policy / planning**: data-driven, map-based decision making

Next steps

- Scale up to the national level
- GWR against location premium → Why does an area have high LP values?
- Predicting future polarization using GWR-TS (time series)

Thank you!

Any Questions?

Contact

- **Email:** lee.7738@osu.edu (Jinhyung Lee)
  miller.81@osu.edu (Harvey J. Miller)
  irwin.114@osu.edu (Nicholas B. Irwin)
  irwin.78@osu.edu (Elena G. Irwin)