Measuring Housing Quality Using Revealed Preference: A Geographic PageRank Approach

NBER Summer Institute
CRIW Pre-Conference
July 16, 2025

Alex Bell (Georgia State University)

Sophie Calder-Wang (University of Pennsylvania)

Shusheng Zhong (Northwestern University)

Introduction

- **Research Question:**
  - **How can one measure the value of a place for households?**
    - The value of a place is reflected through
      - the cost of housing (observed) + the value of amenities (not directly observed)
- **Approach:**
  - **Geographic PageRank (GPR)**
    - A novel ranking of places revealed by household migration choices
    - A measure that captures a networked-based measure of centrality
      - Places attracting people are ranked as good
      - Places attracting people from other attractive places are ranked as even better
    - Interpretation through a revealed preference framework as households generally try to move to better places
  - **Today's focus:**
    1. Apply the algorithm to rich data to obtain rankings across geography, time periods, and sub-populations
    2. Use the Geographic PageRank as a measure of unobserved housing quality, which then can be used in pricing amenities (e.g., air quality)

Outline

I. Motivation
II. Theory of the PageRank Algorithm
III. Empirics and Measurement
IV. Application for pricing amenities

The PageRank Algorithm

(12) **United States Patent**
Page

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Google

(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

- PageRank, at its core, is an ***iterative*** eigenvector computation to produce a measure of network centrality
- Additional features added to resist manipulation (e.g., damping factor)

The PageRank Algorithm: A Brief History

(12) **United States Patent**
Page

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

Google

- The Eigenvalue problem was discovered and rediscovered:
  - In 1895, Edmund Landau (mathematician) suggested for ranking chess players
  - In 1976, Gabriel Pinski and Francis Narin used it to rank scientific journals
  - …
  - In 1996, Robin Li patented RankDex and founded Baidu
  - In 2001, Larry Page patented PageRank
  - Nowadays, it is used in a wide variety of settings
    - Bibliometrics, network analysis, link prediction and recommendation
    - Biology, chemistry, neuroscience, and physics
    - Sorkin (2018): Ranking Firms Using Revealed Preference

Related Literature

- **PageRank:**
  - Web: Brin and Page (1998), Page et. al. (1999), Page (2001)
  - Numerous applications across many fields(survey by Gleich 2015)
    - Including a ranking of roads in transportation studies (Jiang 2009)
  - Labor economics:
    - Firm rankings: Sorkin (2018), Lachowska, Mas, Saagio, and Woodbury (2023); Morchio and Moser (2024)

- **Housing:**
  - Modeling housing markets through a single index of quality:
    - Epple, Quintero, and Sieg (2020), Landvoigt, Piazzesi, and Schneider (2015), Ekeland, Heckman, Nesheim (2004)
  - Role of amenities in household location choices:
    - Diamond (2016), Almagro, Dominguez-Iino (2024)

Define Geographic PageRank (GPR)

- Geographic PageRank:
  - ***Adjacency matrix M:***
    - M_{ij} represents the fraction of households migrating from location *j* to *i*
  - ***Eigenvector formulation:***

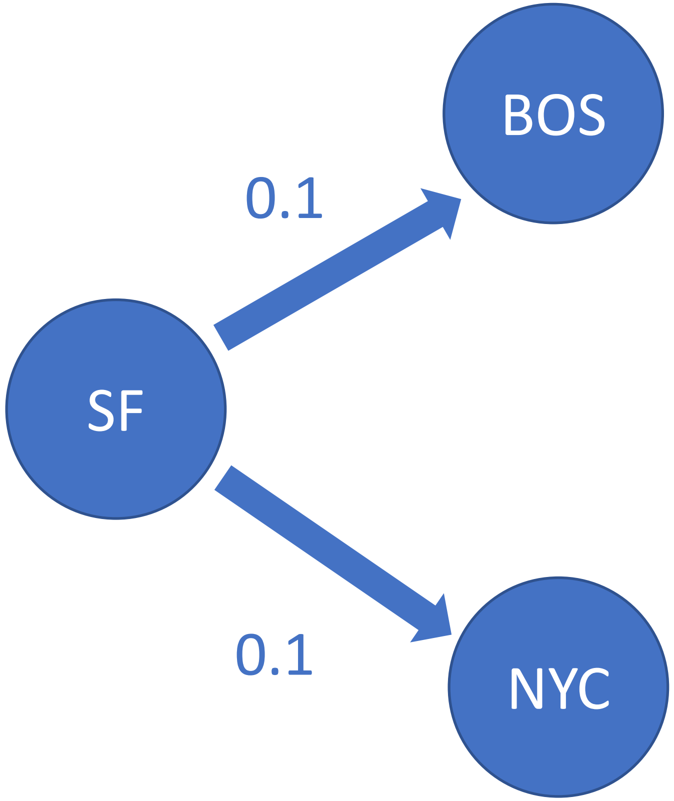
    $$Mv = v$$

    - v is the stationary distribution of the Markov process M
      - If people keep migrating with the migration matrix M forever, v represents the eventual population distribution that it converges to
      - Places with higher probabilities are ranked higher
  - ***Geographic PageRank formulation***

    $$(dM + (1-d)U)v = v$$

    - The damping factor d (=0.85) and U represent a uniform random teleport
    - Better computational properties (connected, no cycling) and robust to manipulation (link farms)

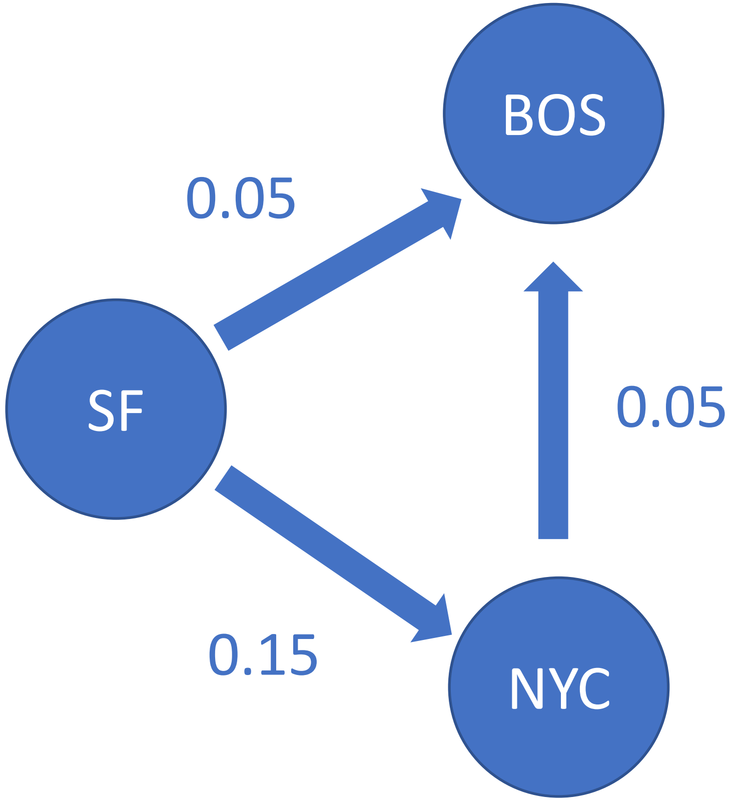
PageRank Example



Starting Input:
- Equal starting size
- 10% leaves SF for Boston
- 10% leaves SF for NYC

Output:
- Net migration rates
  - SF: -20%; NYC: +10%, BOS: +10%
- Ranking:
  - Boston = NYC > SF

PageRank Example



Starting Input:
- Equal starting size
- 5% leaves SF for Boston
- 15% leaves SF for NYC
- 5% leaves NYC for Boston

Output:
- Same net migration rates as before
  - SF: -20%; NYC: +10%, BOS: +10%
- Ranking
  - Boston > NYC > SF

Outline

Overview of Empirics

- **Data:**
  1. IRS County-to-county migration counts
     - Available from 1991 to present

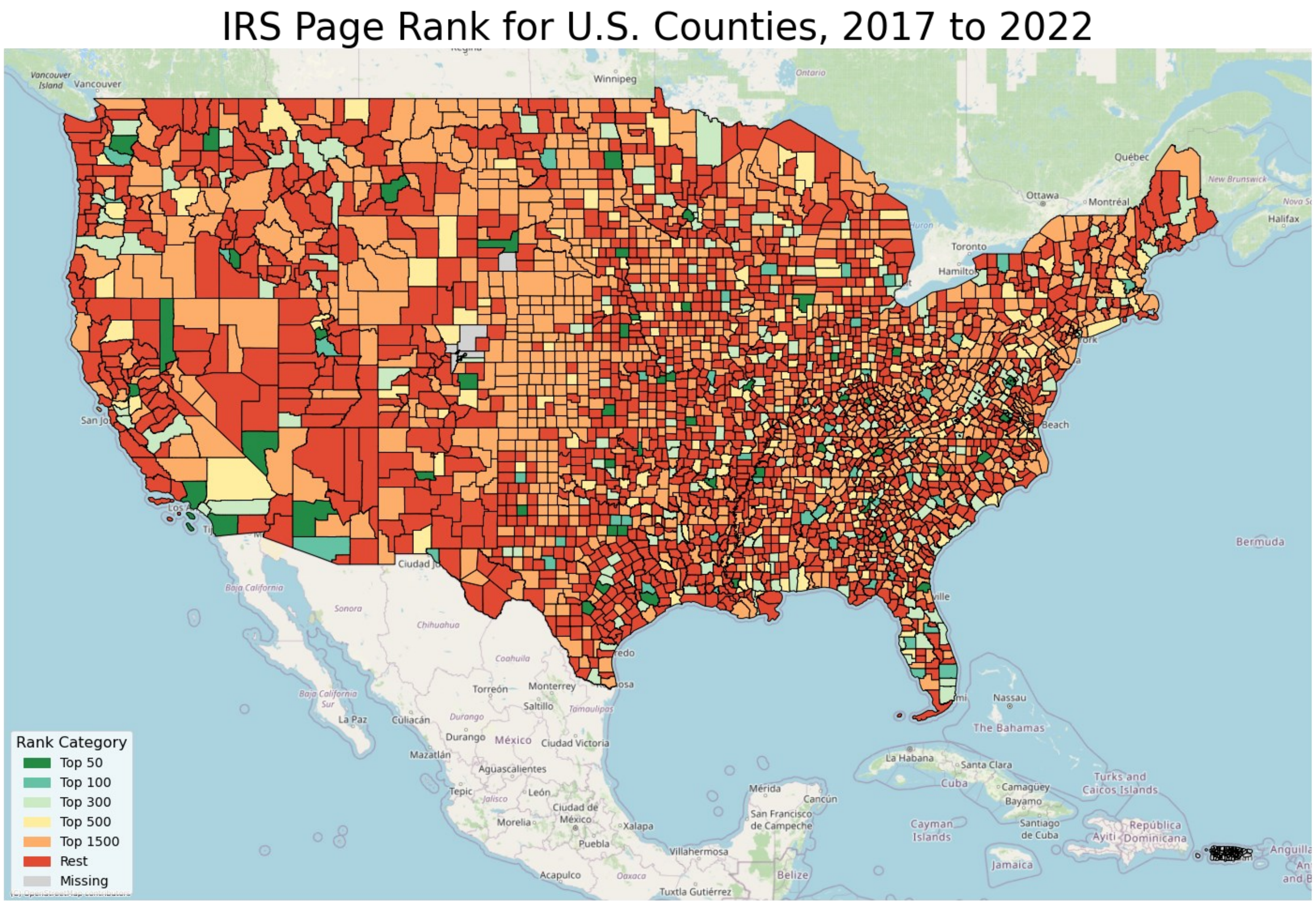
  2. DataAxle data
     - Available from 2009 to present
     - Address level moves; allowing us to rank not only counties, but also finer geographic areas
       - Focus is on finer neighborhood-level rankings

  3. ACS micro data
     - Available from 2005 to present
     - Rich household and individual demographics
     - Migration PUMA is created for data privacy, not consistent over time
       - Focus is on metropolitan level rankings to compare ranks for different demographic groups

Outline

I. Motivation
II. Theory of the PageRank Algorithm
III. Empirics and Measurement
   a) Cross-sectional rankings
   b) Time-series rankings
   c) Rankings by demographic subgroups
IV. Application for pricing amenities

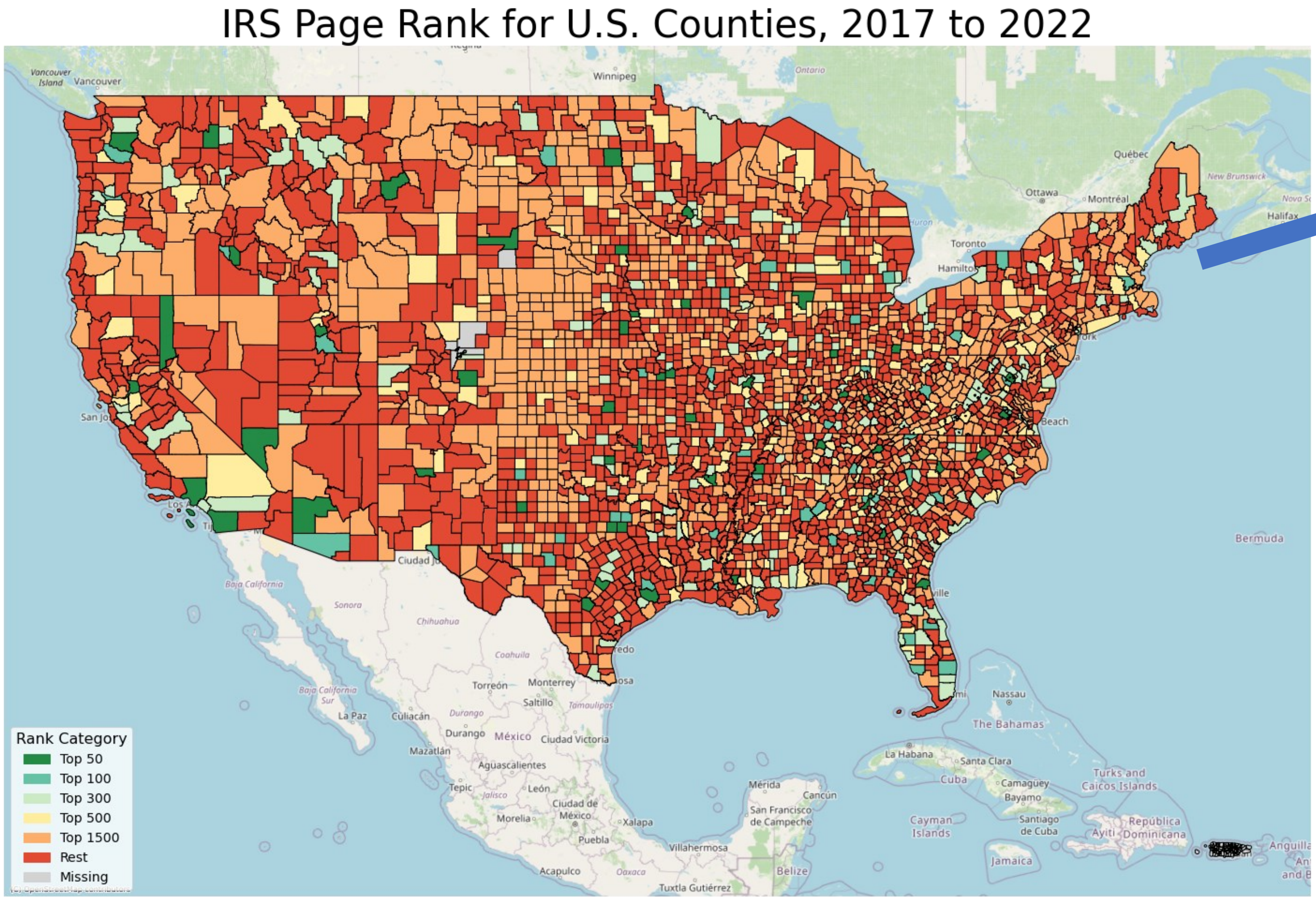
Geographic PageRank for US Counties: 2017-2022

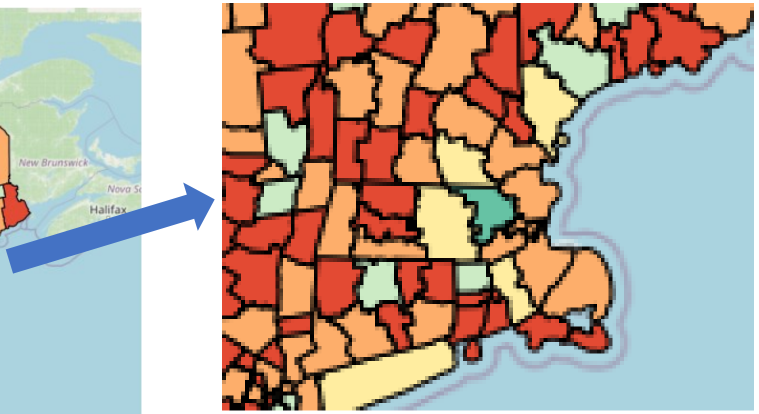


Top 10 Ranked Counties: 2017 - 2022

Rank	State	County	City
1	Texas	Harris	Houston
2	Arizona	Maricopa	Phoenix
3	Texas	Bexar	San Antonio
4	Texas	Tarrant	Fort Worth
5	Illinois	Cook	Chicago
6	Texas	Dallas	Dallas
7	California	Los Angeles	Los Angeles
8	Iowa	Polk	Des Moines
9	Minnesota	Hennepin	Minneapolis
10	Oklahoma	Oklahoma	Oklahoma City

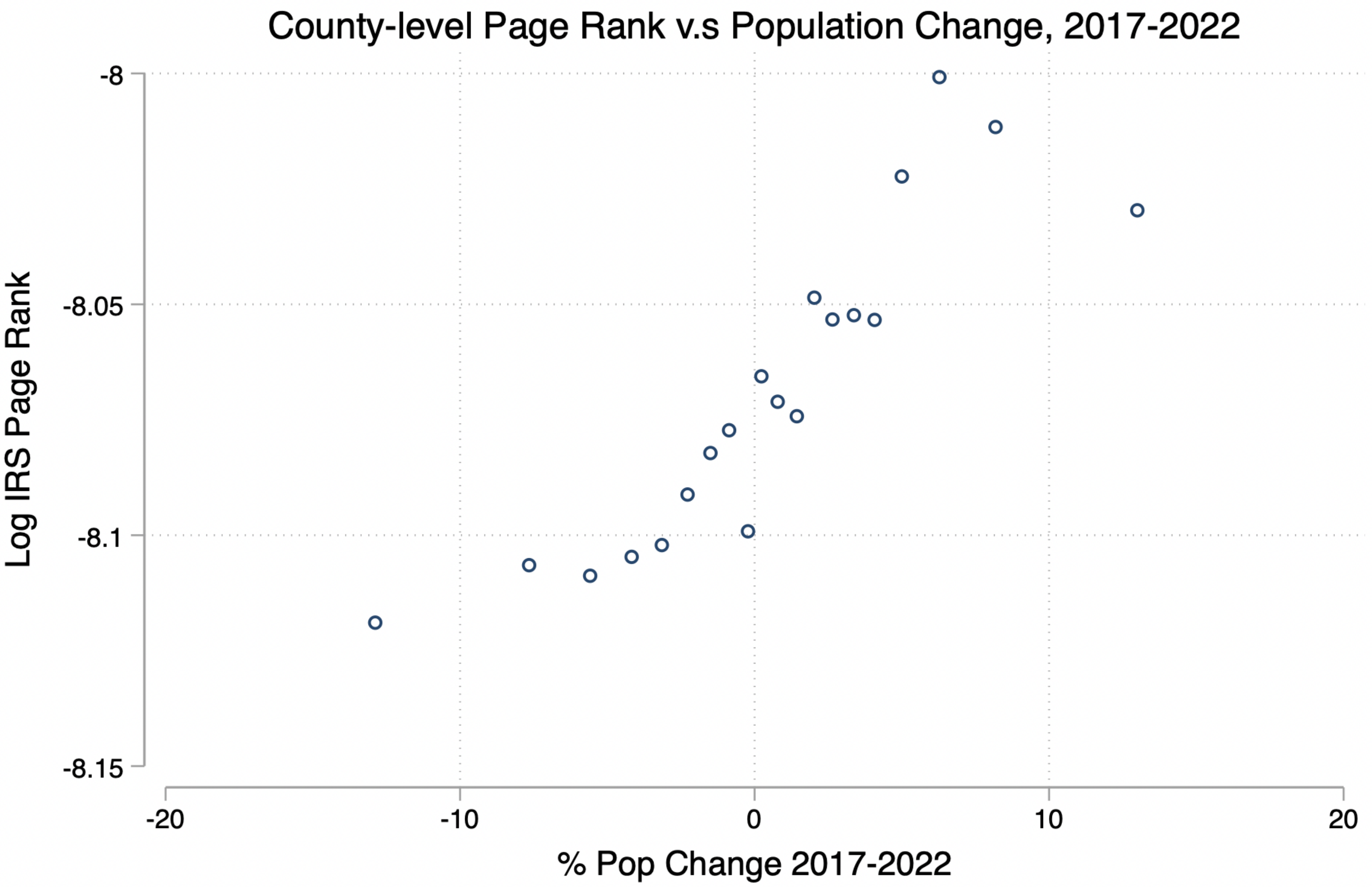
Geographic PageRank for US Counties: 2017-2022





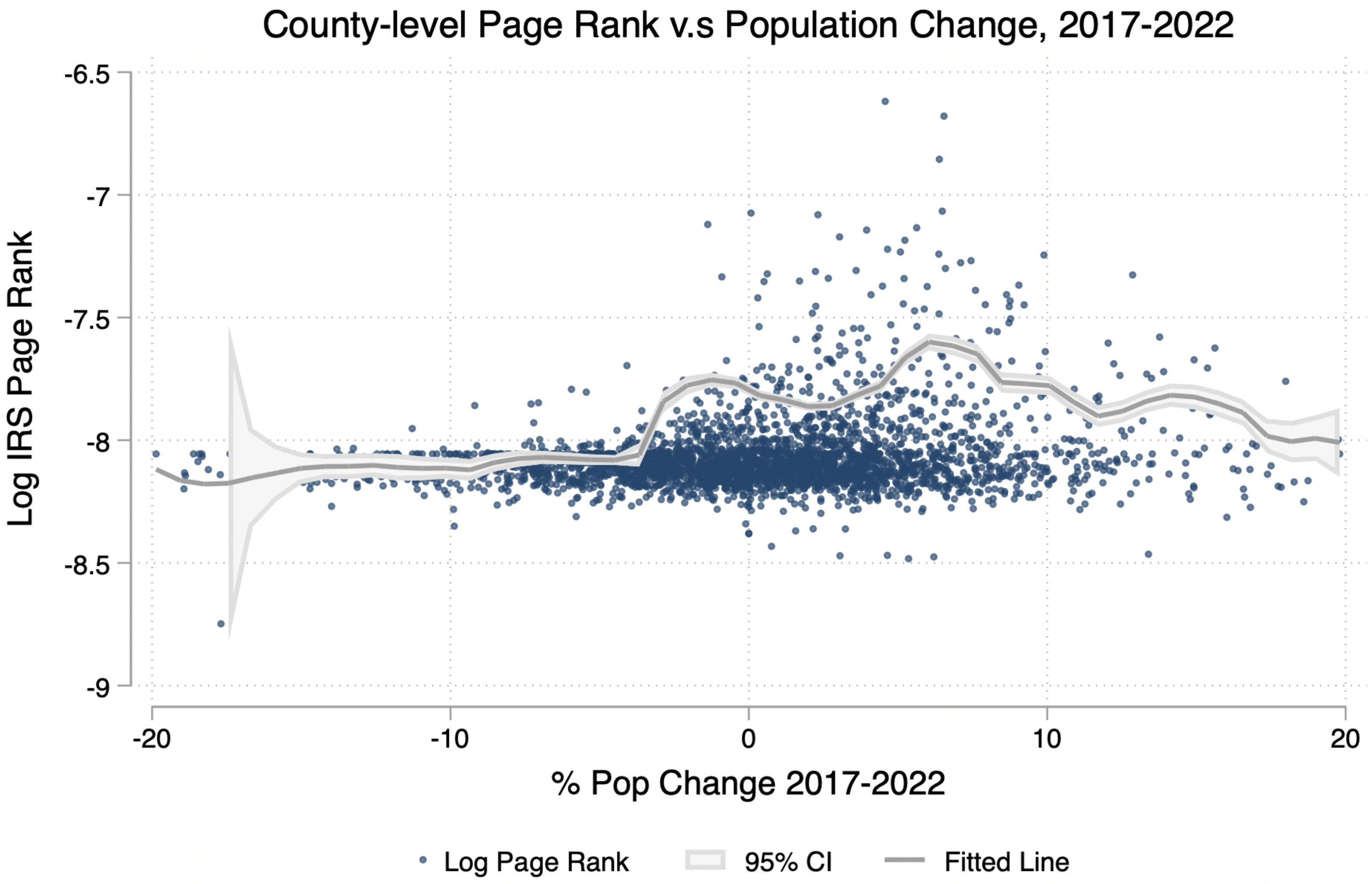
- Rankings of counties in the Boston-metro area:
  - Middlesex (85) – Cambridge
  - Worcester (361) - Worcester
  - Essex (629) - Salem
  - Norfolk (717) - Norfolk
  - Suffolk (770) – Boston
  - Plymouth (1138) – Plymouth

Geographic PageRank for US Counties: 2017-2022



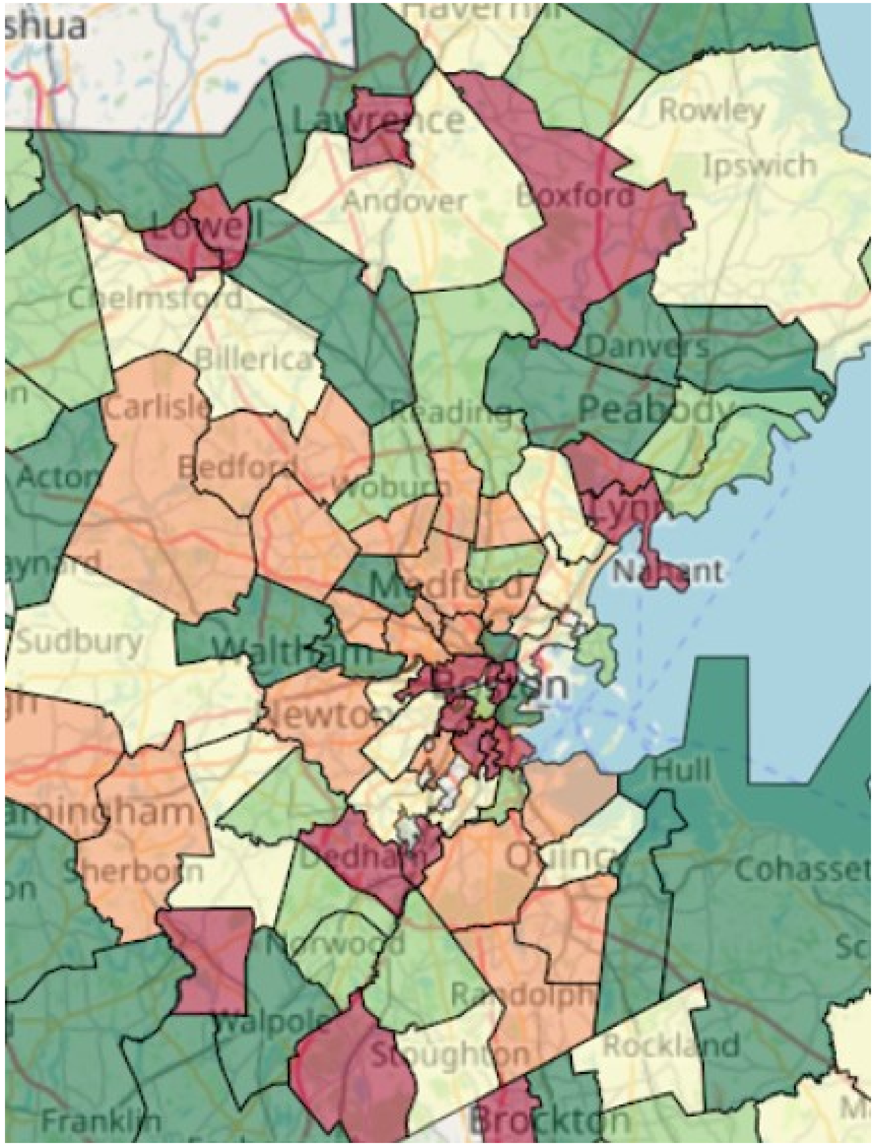
- Places with more inflows are ranked higher
- Places with more inflows from other higher-ranked places are ranked even higher

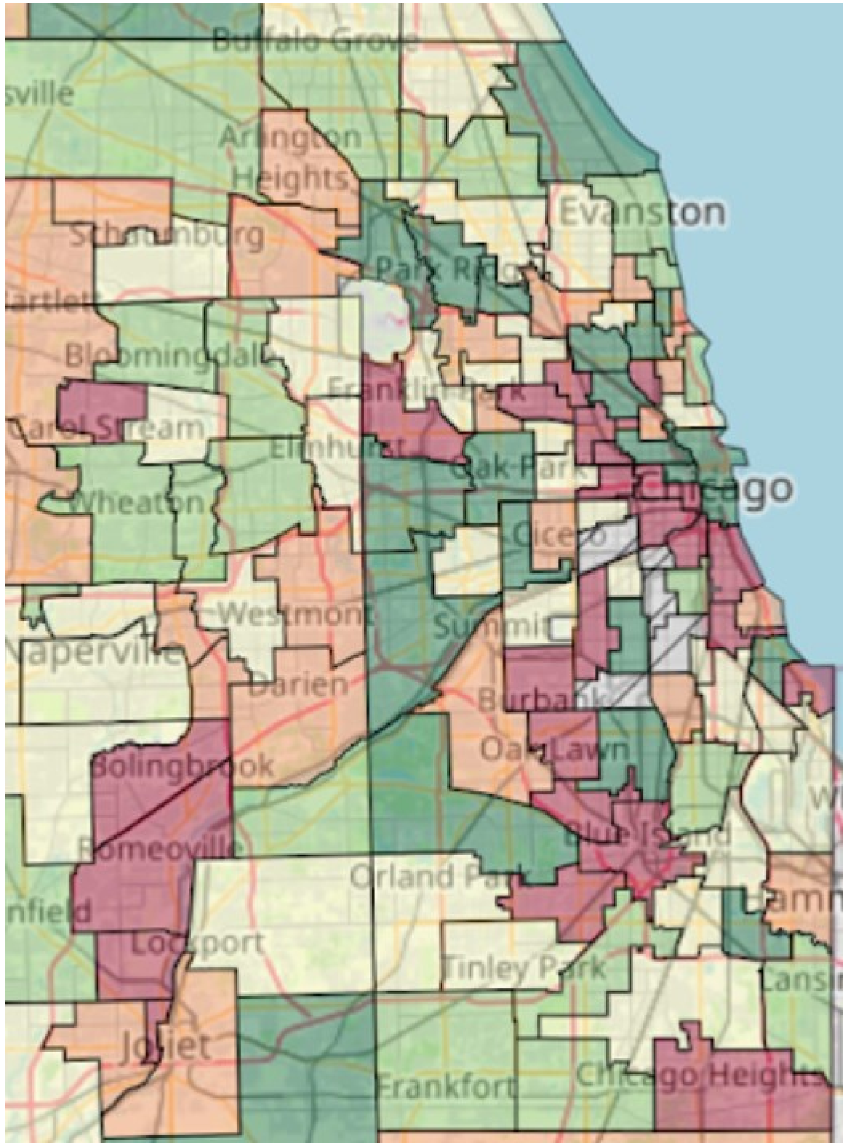
Geographic PageRank for US Counties: 2017-2022



- Places with more inflows are ranked higher
- Places with more inflows from other higher-ranked places are ranked even higher
- Still significant dispersion compared to pure migration measures

Geographic PageRank for Neighborhoods





- To getting ranking of finer smaller geographic units, we use Data Axle
  - Based on address changes (e.g., USPS, utility etc.)
  - Aggregate to neighborhood as defined by Mast (2025)
  - Fairly intuitive estimates within a metropolitan area

Outline

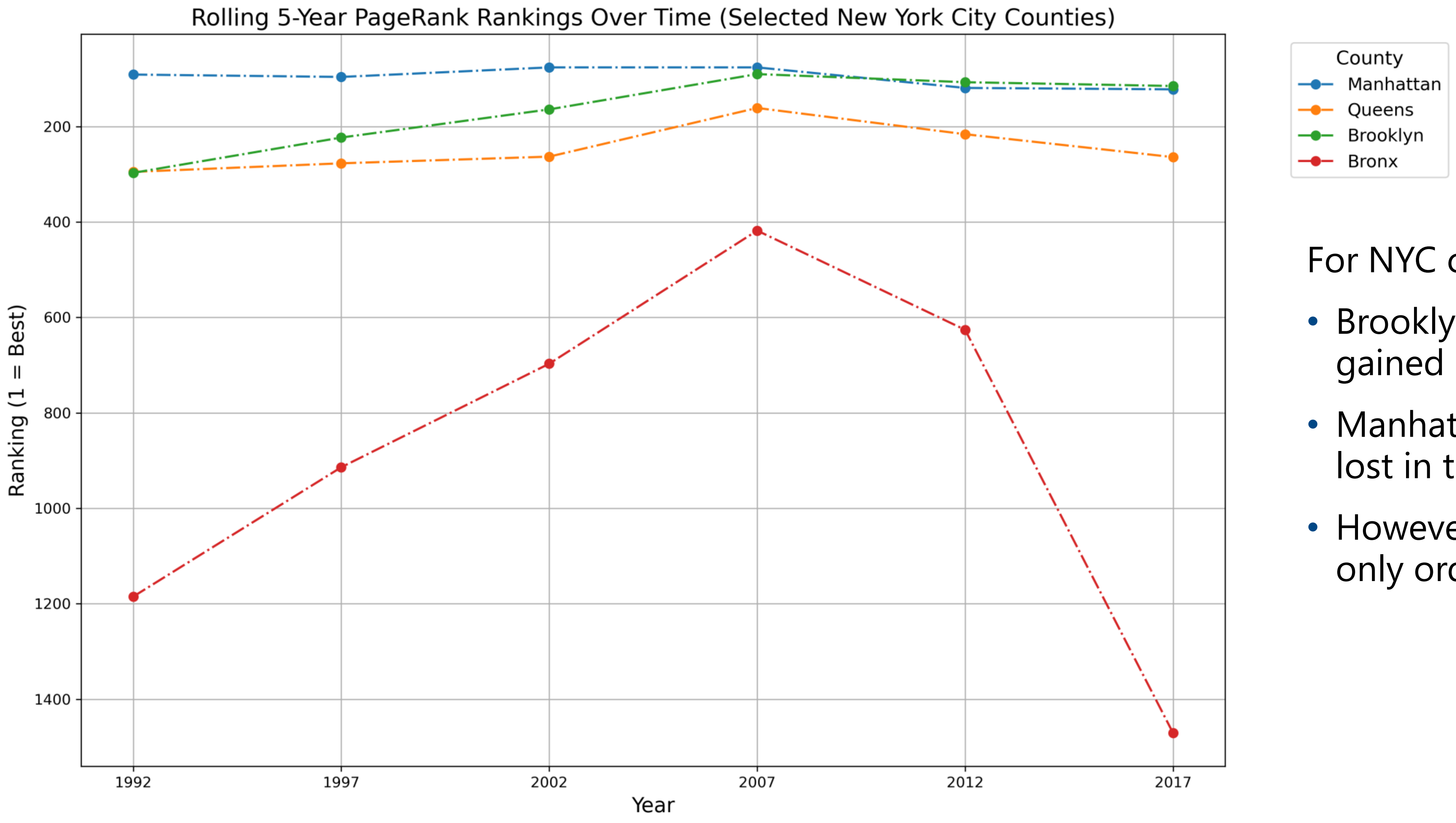
Geographic PageRank for US Counties: 1997-2022

Top 20 Ranked Counties: 1997 - 2002			
Rank	State	County	City
1	Texas	Harris	Houston
2	Arizona	Maricopa	Phoenix
3	Texas	Dallas	Dallas
4	Minnesota	Hennepin	Minneapolis
5	Texas	Tarrant	Fort Worth
6	Texas	Bexar	San Antonio
7	California	Los Angeles	Los Angeles
8	Illinois	Cook	Chicago
9	Washington	King	Seattle
10	Nebraska	Lancaster	Lincoln
11	South Dakota	Minnehaha	Sioux Falls
12	Nevada	Clark	Las Vegas
13	California	San Diego	San Diego
14	Iowa	Polk	Des Moines
15	Ohio	Franklin	Columbus
16	Oklahoma	Oklahoma	Oklahoma City
17	Texas	Travis	Austin
18	Indiana	Marion	Indianapolis
19	Tennessee	Shelby	Memphis
20	Utah	Salt Lake	Salt Lake City

Top 20 Ranked Counties: 2017 - 2022			
Rank	State	County	City
1	Texas	Harris	Houston
2	Arizona	Maricopa	Phoenix
3	Texas	Bexar	San Antonio
4	Texas	Tarrant	Fort Worth
5	Illinois	Cook	Chicago
6	Texas	Dallas	Dallas
7	California	Los Angeles	Los Angeles
8	Iowa	Polk	Des Moines
9	Minnesota	Hennepin	Minneapolis
10	Oklahoma	Oklahoma	Oklahoma City
11	Ohio	Franklin	Columbus
12	Texas	Lubbock	Lubbock
13	Oklahoma	Tulsa	Tulsa
14	Washington	King	Seattle
15	North Carolina	Wake	Raleigh
16	South Dakota	Minnehaha	Sioux Falls
17	Nevada	Clark	Las Vegas
18	Utah	Salt Lake	Salt Lake City
19	Indiana	Marion	Indianapolis
20	Kentucky	Jefferson	Louisville

- Ranking of the top places are similar
- Some differences remain:
  - San Diego (13 ->28)
  - Austin (17 ->29)
  - Raleigh (36->15)
  - Lubbock (28->12)

Geographic PageRank for US Counties: 1997-2022



For NYC counties:

- Brooklyn and Queens gained in the rankings
- Manhattan and Bronx lost in the rankings
- However, rank itself is only ordinal, not cardinal

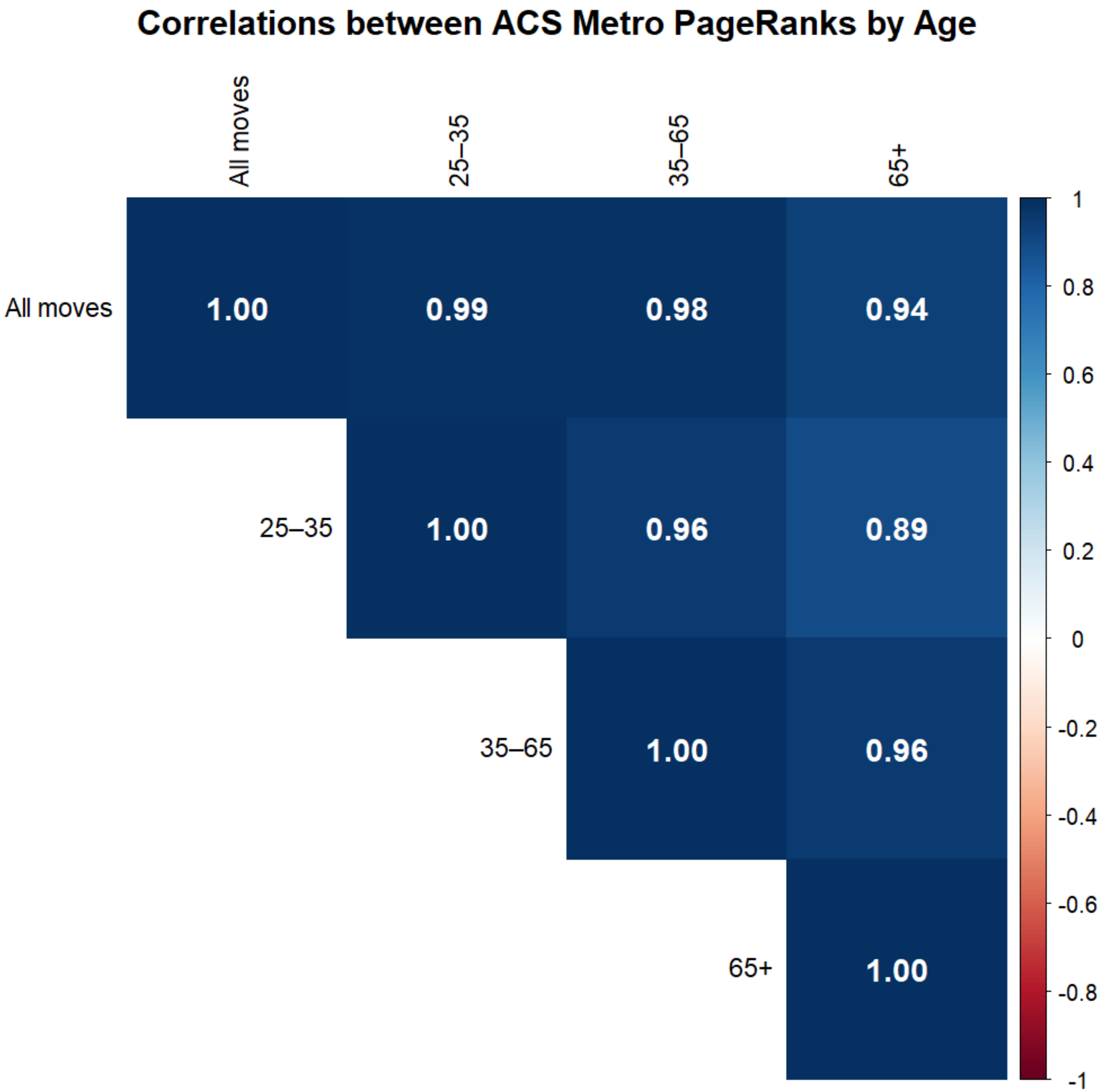
Outline

Geographic PageRank By Subgroups

- **Data: ACS micro data**
  - Includes the migration PUMA from last year (if moved)
    - Aggregate to metro measures to be consistent across years
    - Aggregate for multiple years (5yr) given that ACS is only a sample
  - A wide array of household and individual characteristics:
    - Age, Education, Race, and Industry
  - The value of a place may vary for these sub-groups

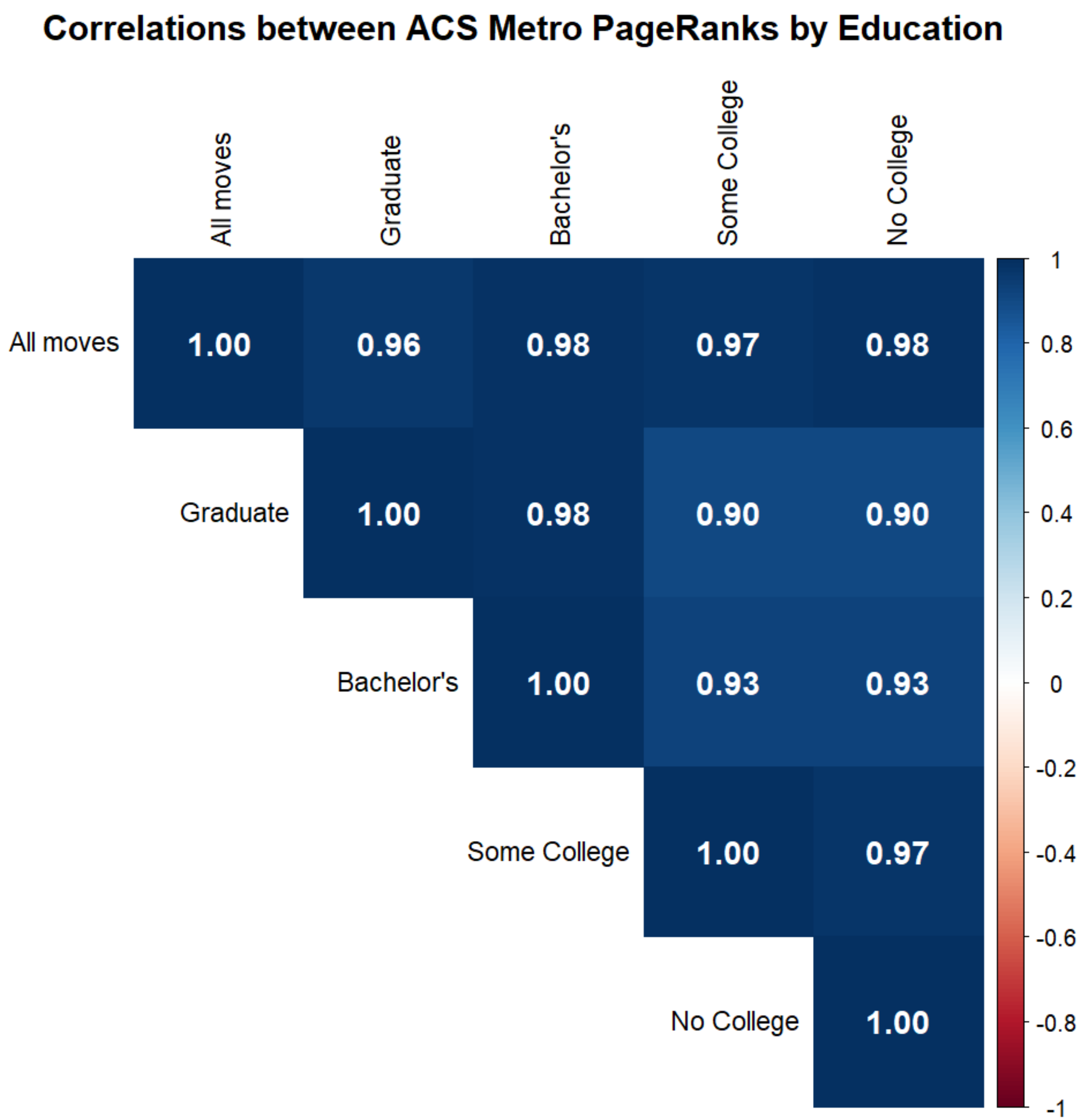
- **Main findings:**
  - For most demographic characteristics, the ranking of metros are highly correlated
    - However, important differences remain

Geographic PageRank for US Metros: By Age Group



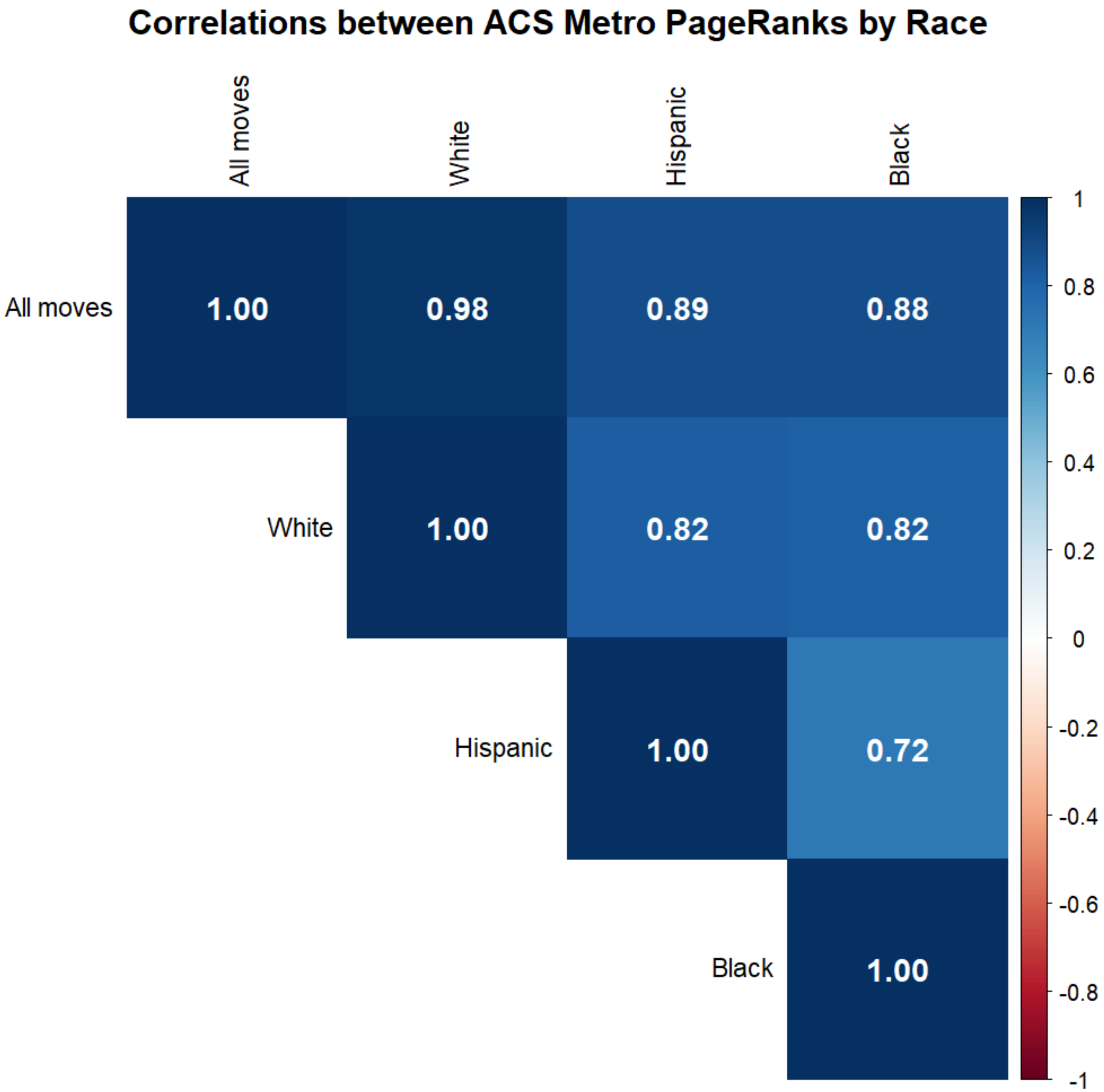
- Highly correlated for all age groups
- The rankings for the oldest movers and the youngest movers are least correlated

Geographic PageRank for US Metros: By Education



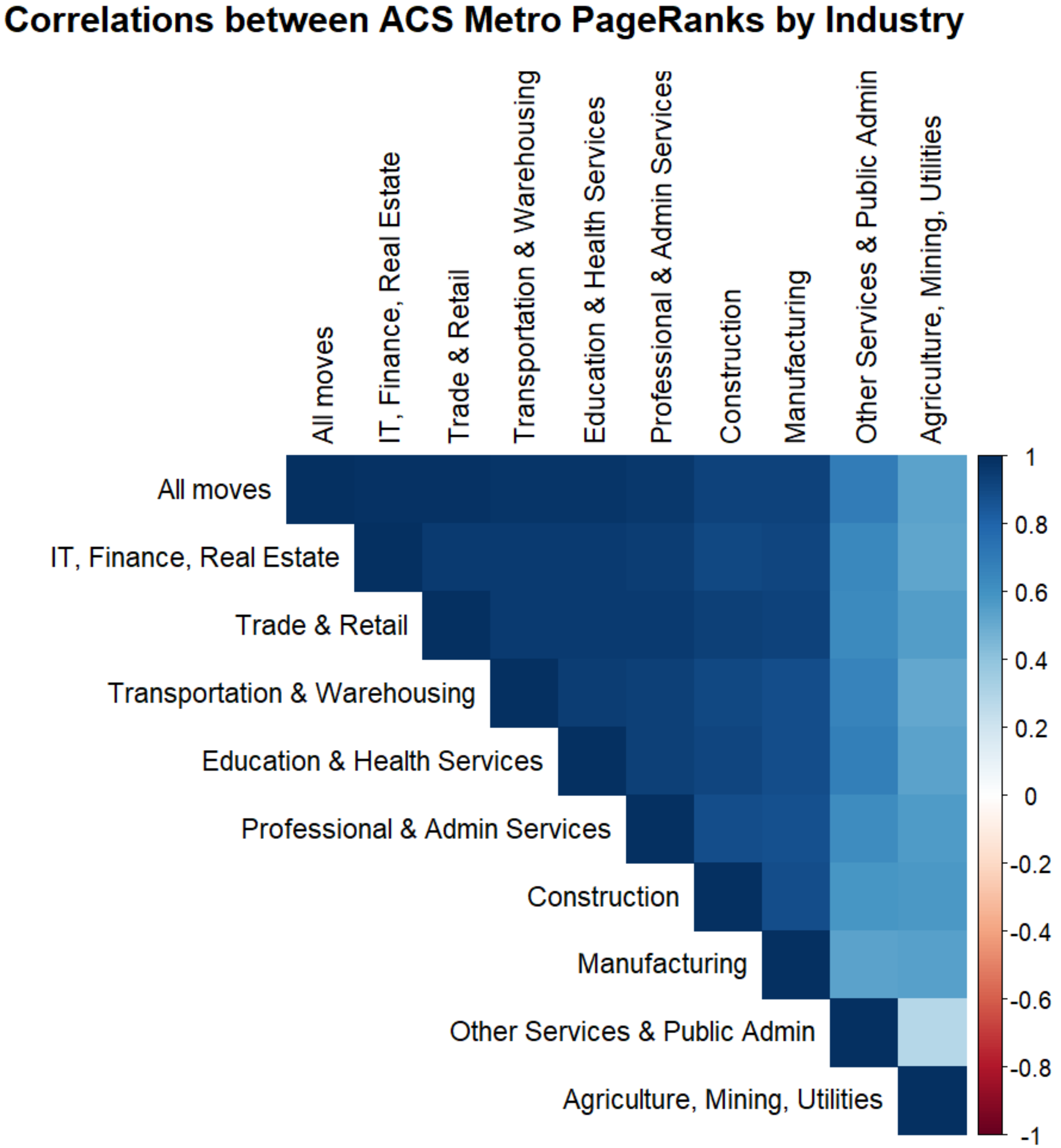
- Highly correlated for levels of education
- The rankings for those without college are least correlated with those with graduate degrees

Geographic PageRank for US Metros: By Race and Ethnicity



- Mostly correlated across racial and ethnicity groups
  - The rankings for Blacks and Hispanics are also different

Geographic PageRank for US Metros: By Industry

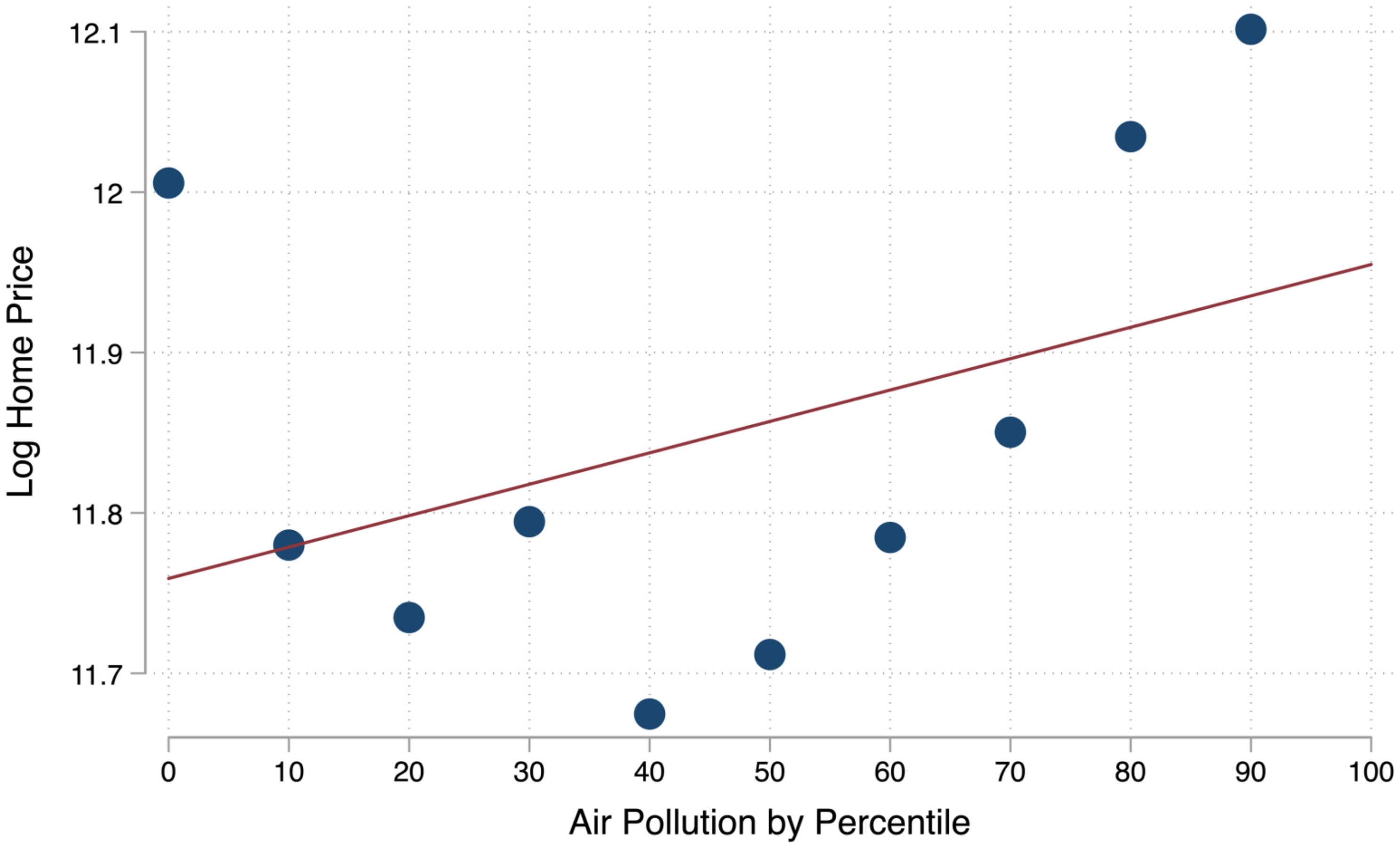


- Mostly correlated across industries
  - Industry measured at the destination due to ACS
  - Rankings become noticeably different for those in agriculture, mining, and utilities
- Framework can be extended to the ranking of Place-Industry pairs
  - Requires more granular data at both the source and the destination

Outline

Application: Pricing Neighborhood Amenities

- **Research question:**
  - What is the implicit price of air quality?
    - Empirical challenge: amenities such as clean air can often be correlated with the unobserved quality of the place -> “The wrong-signed problem”

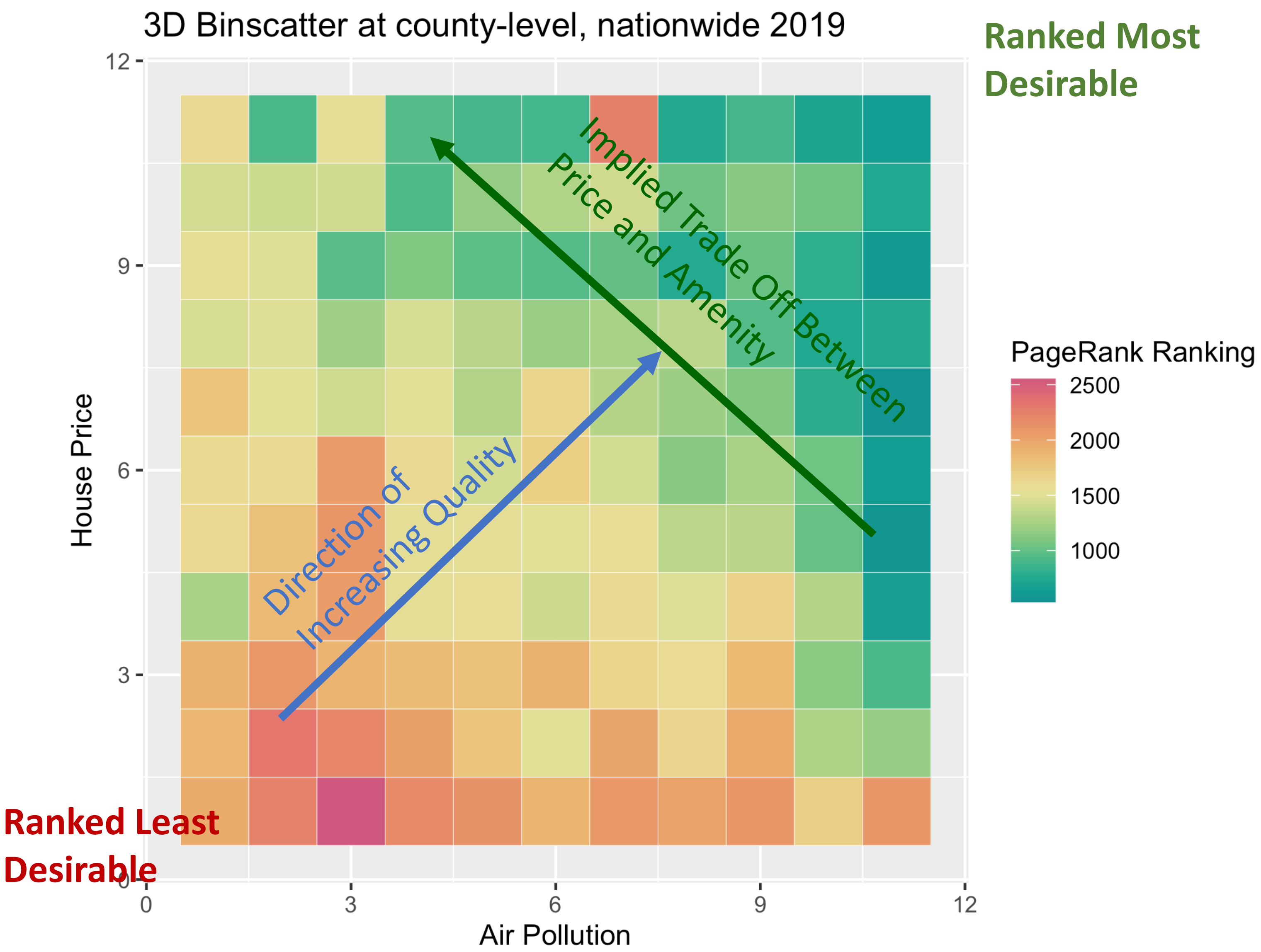


Application: Pricing Neighborhood Amenities

- **Research question:**
  - What is the implicit price of air quality?
    - Empirical challenge: amenities such as clean air can often be correlated with the unobserved quality of the place -> "The wrong-signed problem"

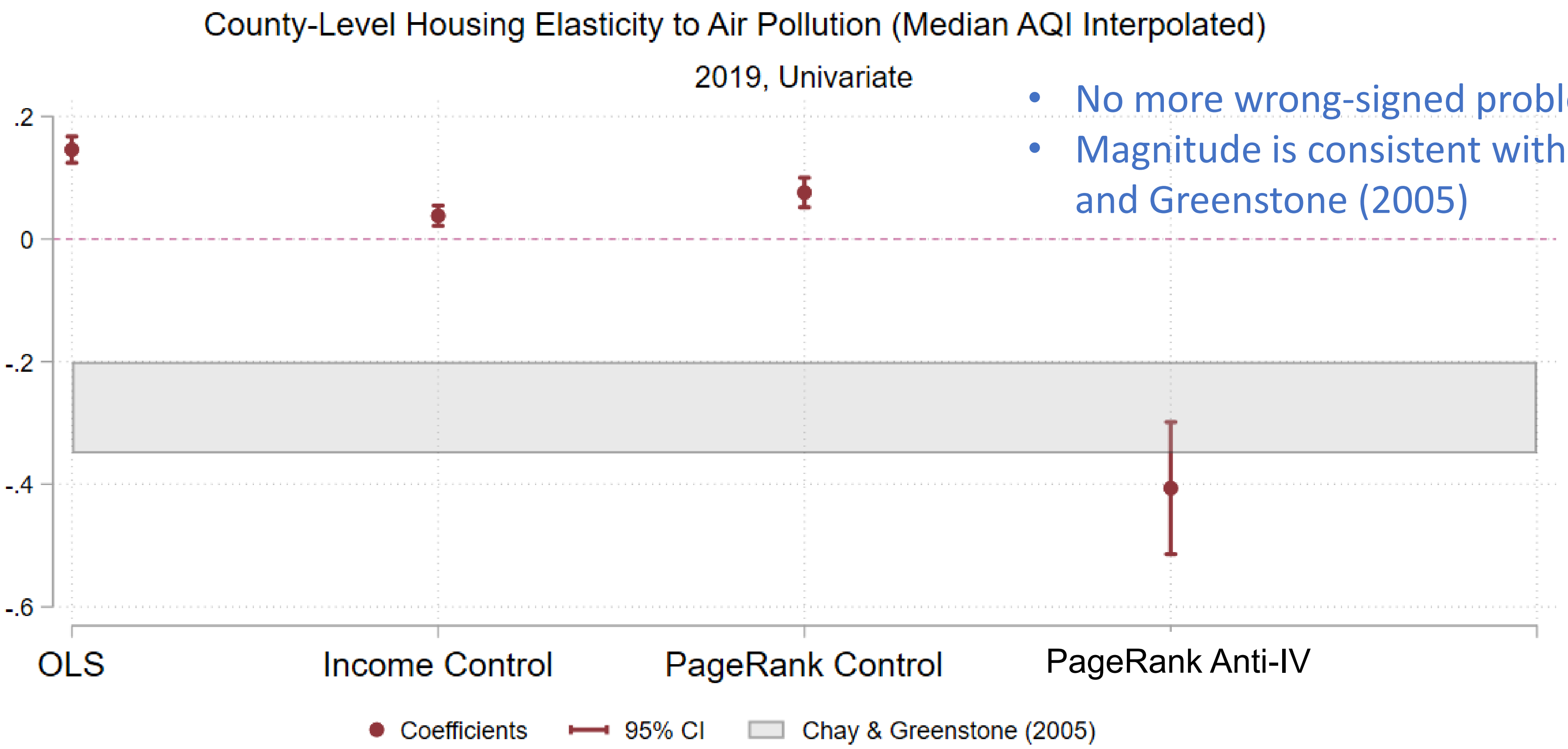
- **Method:**
  - Conventional approach:
    - Find more controls -> But can never exhaust all controls
    - Find quasi-experiments -> Limited in scope
  - The Anti-IV Approach (Bell, Billings, Calder-Wang, and Zhong, 2024)
    - County-level Geographic PageRank can be used as an anti-instrument
      - GPR co-moves with unobserved housing quality
      - The extent to which GPR mismeasures true housing quality is uncorrelated with air pollution

Application: Pricing Neighborhood Amenities



- X-axis: Deciles of air pollution
- Y-axis: Deciles of home prices
- Z-axis (color): Deciles of predicted Geographic PageRank:
  - E[H | P, Z]
  - Calculated as the average GPR for the bin

Application: Pricing Neighborhood Amenities



- No more wrong-signed problems!
- Magnitude is consistent with Chay and Greenstone (2005)

Conclusion

- **Theory:**
  - Use the well-known PageRank algorithm to rank places
    - A recursive measure based on network centrality
    - Interpreted as the value of place through revealed preference

- **Measurement:**
  1. Cross-sectional measures: County-level and neighborhood-level GPR
  2. Time-series measures: County-level GPR over time
  3. Geographic PageRank for different demographic groups

- **Application:**
  - Use GPR as an anti-IV for unmeasured quality of housing to price neighborhood amenities
    - Works well for obtaining the implicit price of air quality

Data and Visualization Available at: https://sophieqzwang.github.io/geopagerank/

Thanks!

Questions or comments?