

Data and the State

PUBPOL 2130 / INFO 3130



Redlining

Lecture 19, Thursday Apr 10

Projects!

Dataset Ideas/Brainstorm:

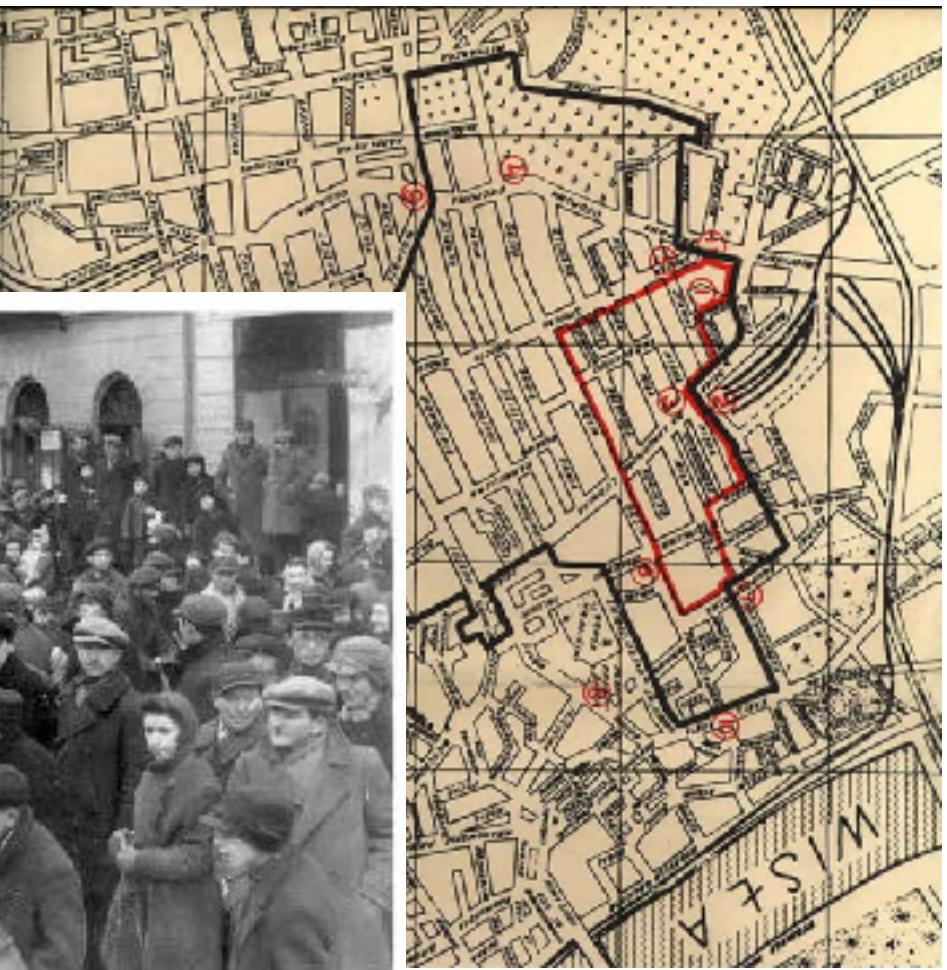
- [LODES](#) – Longitudinal Employer-Household Dynamics
- [HMDA Data](#) – Home Mortgage Disclosure Act
- [CRA Data](#) – Community Reinvestment Act, aimed at ending redlining
- Bike Share Data
 - [NYC \(CitiBike\)](#)
- [EPA Data](#) – Environmental Protection Agency
- Open Data Portals
 - [NYC](#), [New York State](#), [Minneapolis](#), etc etc
- [VEQ](#) – Longitudinal Employer-Household Dynamics / Veteran Employment Outcomes
- [FRED](#) – Federal Reserve Economic Data
- [Health Insurance Datasets from the Census](#)
- [Cyber Events Database](#)
- [CDC WONDER](#) – mortality data (Wide-ranging ONline Data for Epidemiologic Research)
- [PSID](#) – Panel Study of Income Dynamics

The trope of the “ghetto”

- The term “**ghetto**” comes from 16th century Venice, where Jews were confined by law to a specific district
- Across medieval and early modern Europe, Jewish ghettos became common, often legally prescribed, walled, and locked at night. Notorious starved and imprisoned ghettos in the Third Reich (e.g., Warsaw ghetto incarcerated 400,000)
- In the 19th to early 20th century, with mass immigration to the U.S., the term became applied to ethnic enclaves for immigrant groups
- Only in the midst of the Great Migration did the term begin to apply to largely Black “inner city” neighborhoods — sometimes anchored by public housing, often walled off by highway construction, calcified by formal policy (loans), unequal service provision, and informal practice (real estate)



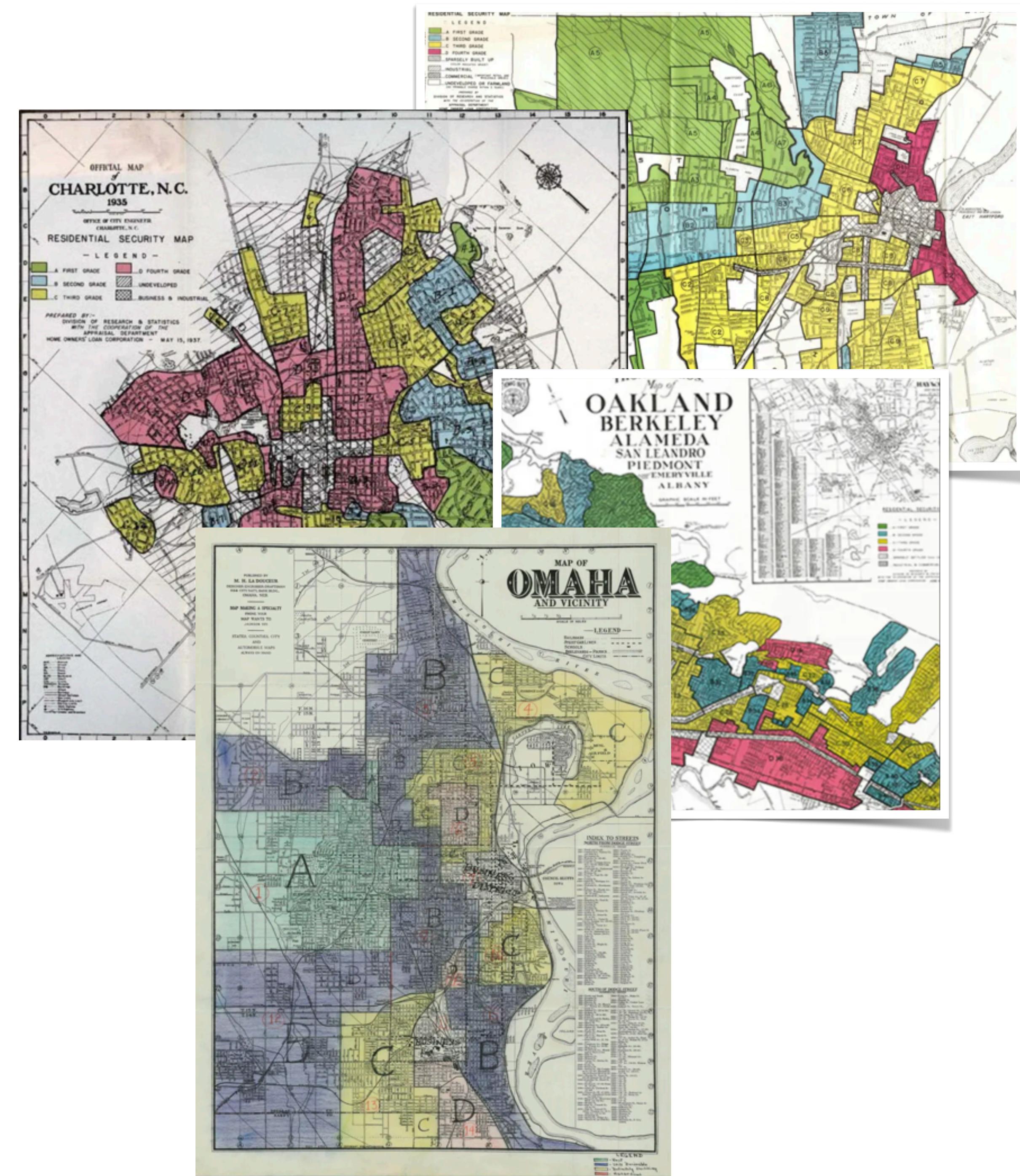
Warsaw 1942

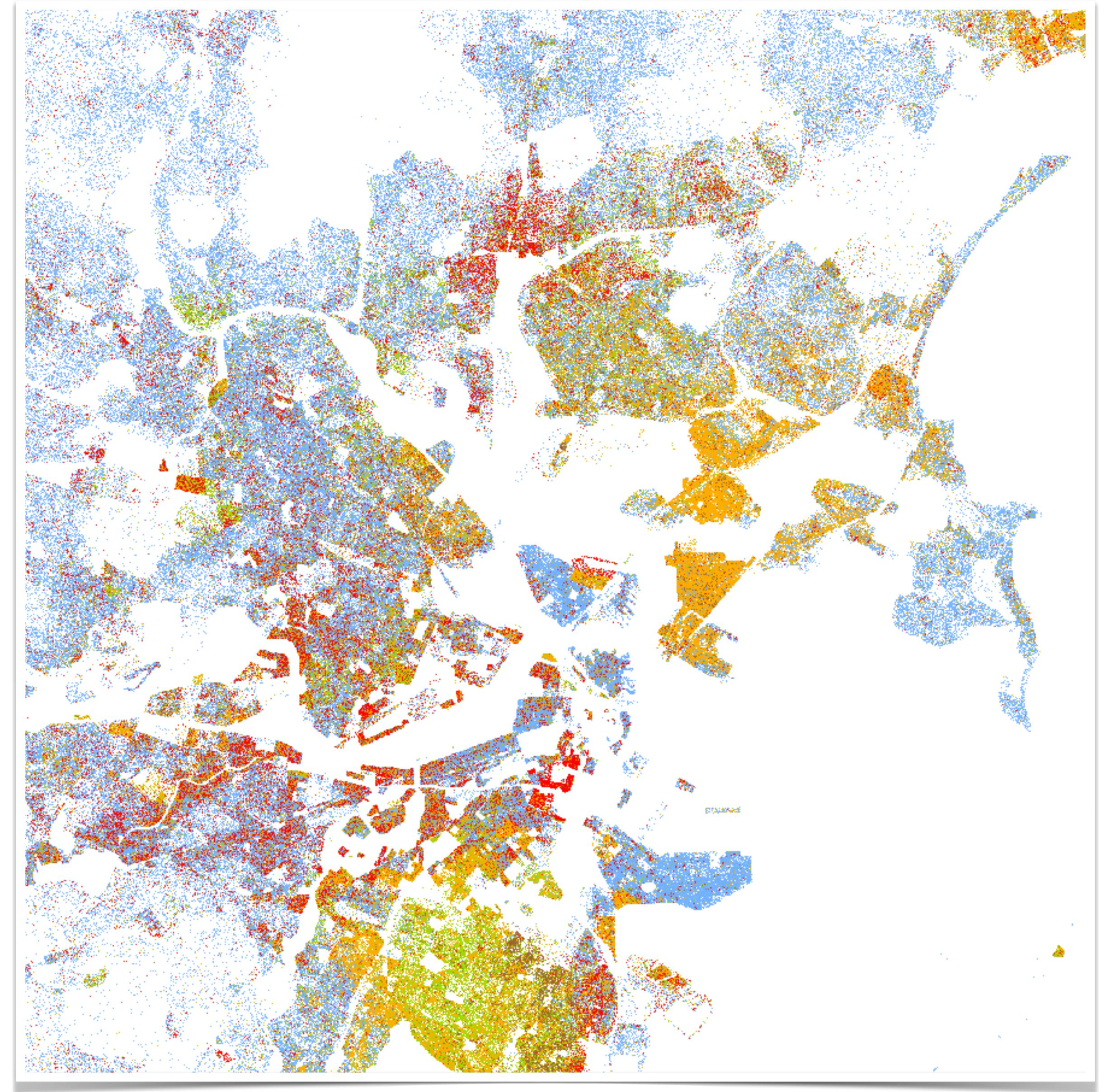


Camden 2015

Redlining

- Home lenders produced literally color-coded maps starting in 1930s with FDR's HOLC (Home Owners Loan Corporation)
- Neighborhoods were graded for investment risk, with red or "hazardous" areas being marked for denial of mortgage loans and refinancing
- Policies were overtly racial, targeting Black neighborhoods
- HOLC refused to **make** loans, FHA refused to **insure** them
- Officially outlawed 1968, further addressed in Community Reinvestment Act of 1977



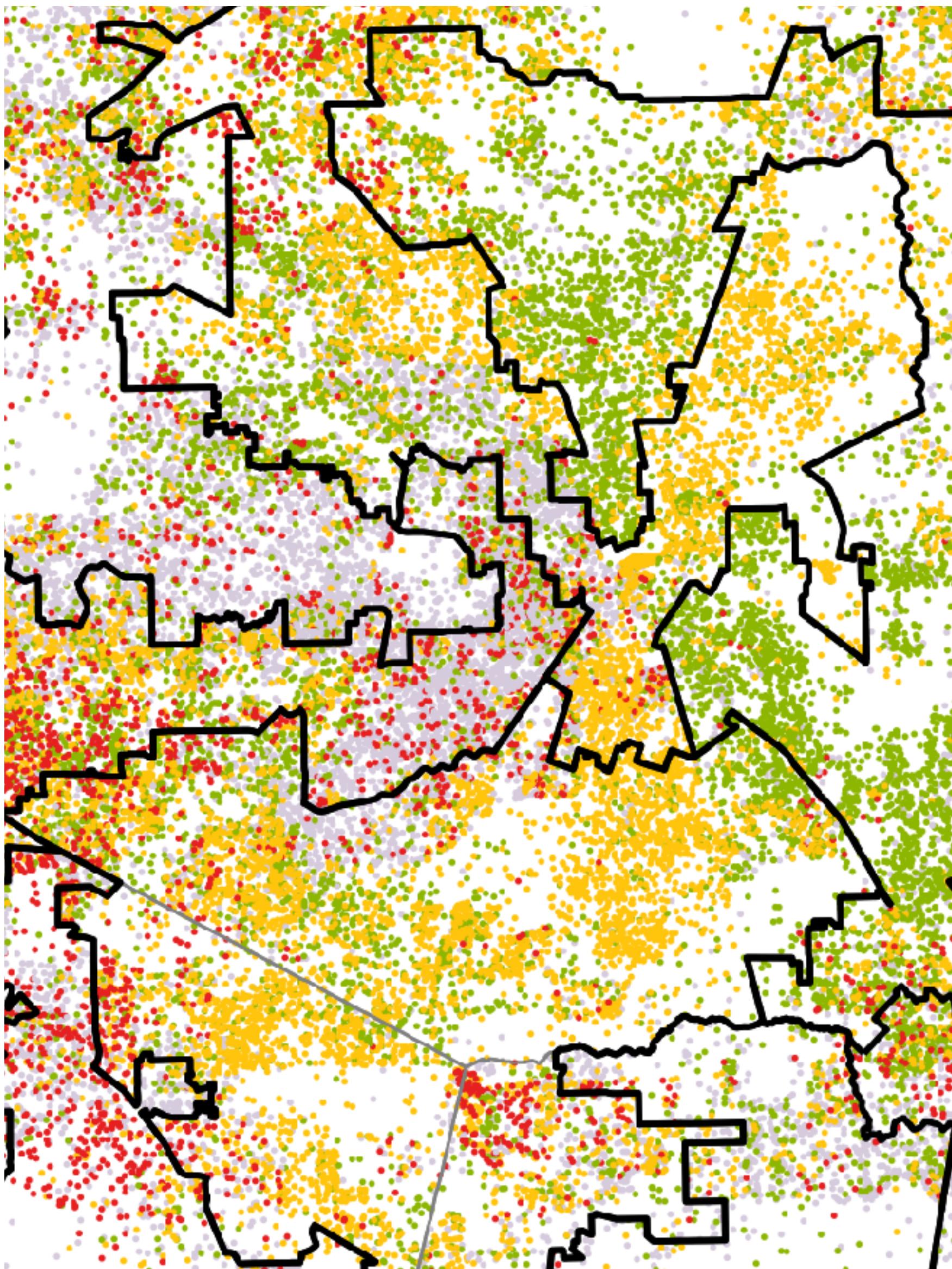


2010 Census Block Data

1 Dot = 1 Person

- White
- Black
- Asian
- Hispanic
- Other Race / Native American / Multi-racial

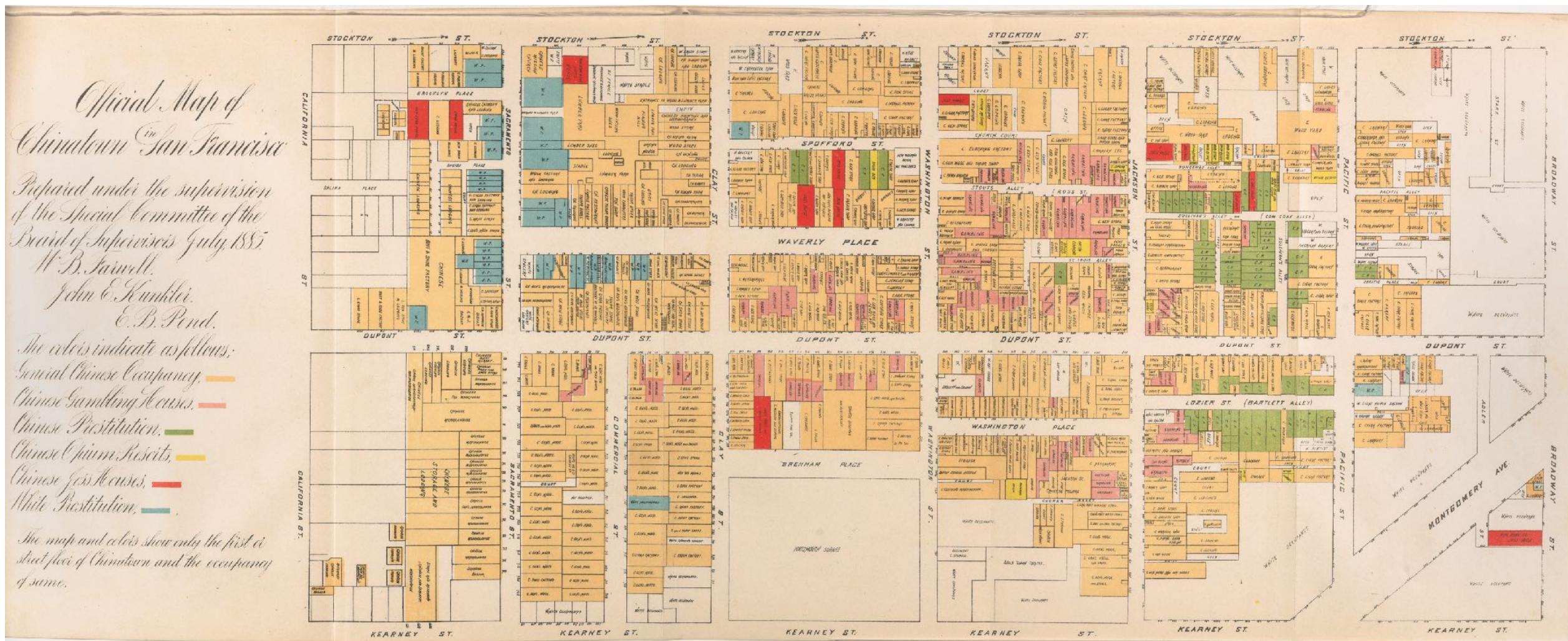
Boston area



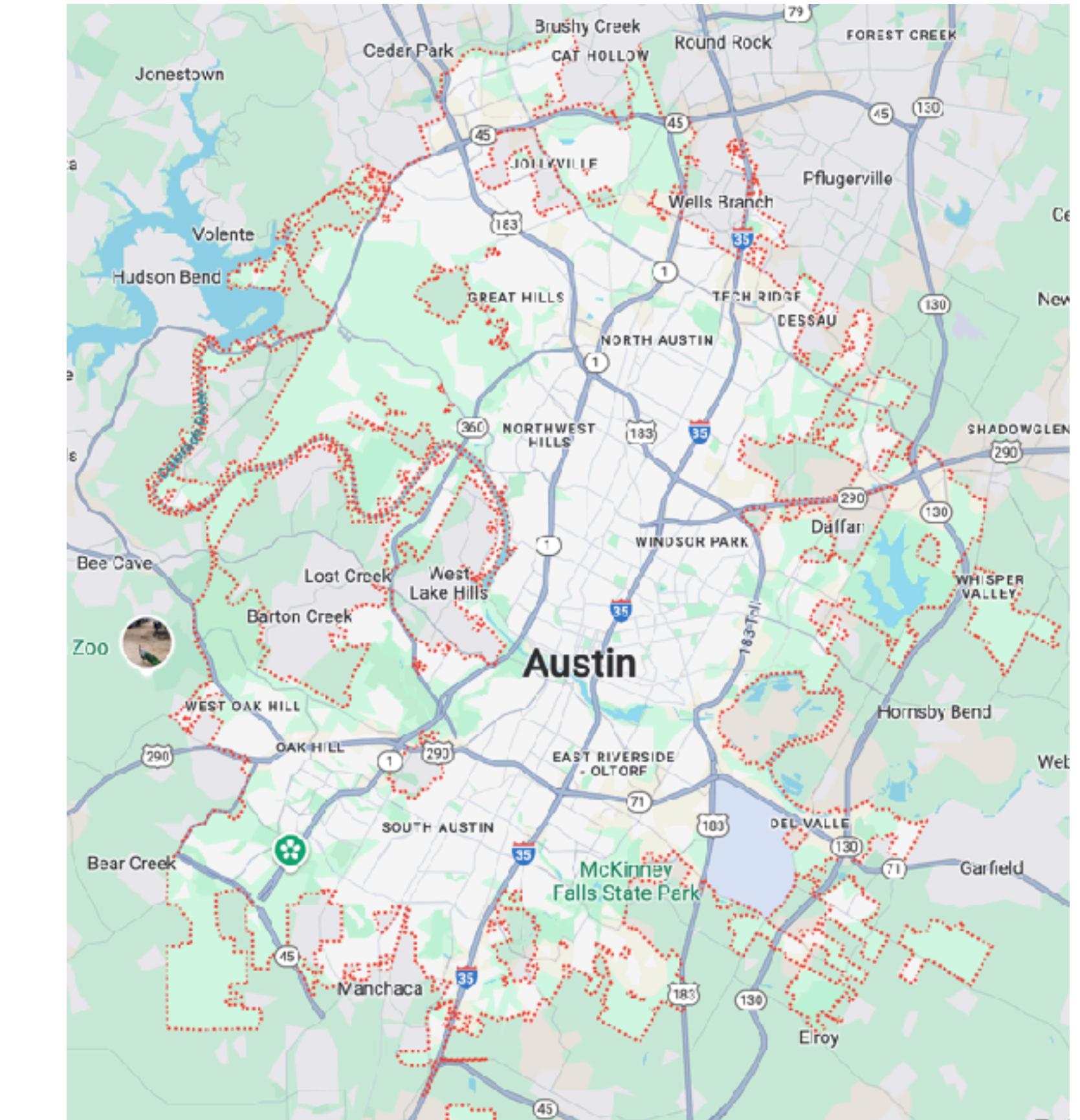
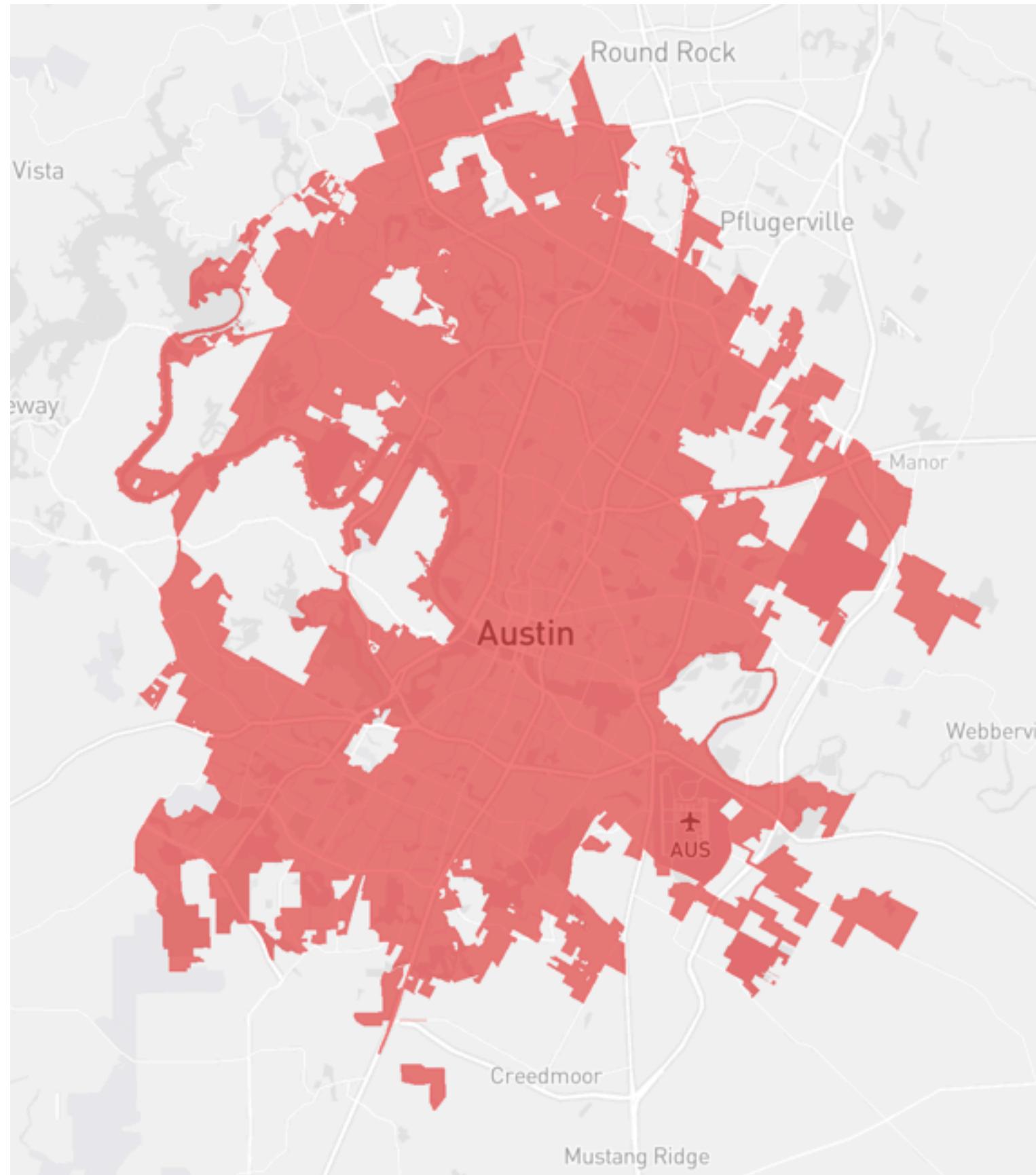
Houston area

Example: Asian segregation

- Alien Land Laws: Beginning with California (1913) and eventually adopted by many western states, these laws prohibited "aliens ineligible for citizenship" (which included all people of Chinese ancestry until 1943) from owning land or property.
- San Francisco Cubic Air Ordinance (1870): Required 500 cubic feet of air per person in living quarters, selectively enforced against Chinese residences to criminalize their housing arrangements. — coupled with Laundry Ordinance, Queue Ordinance, etc
- Residential pressure and restrictions
 - Restrictive covenants in property deeds prohibiting sale to "Mongolians"
 - Neighborhood improvement associations with explicit anti-Chinese policies
 - Municipal zoning that concentrated laundries and other Chinese businesses
 - Police non-protection



Example: Austin, TX



I-35 (late 50s/early 60s)

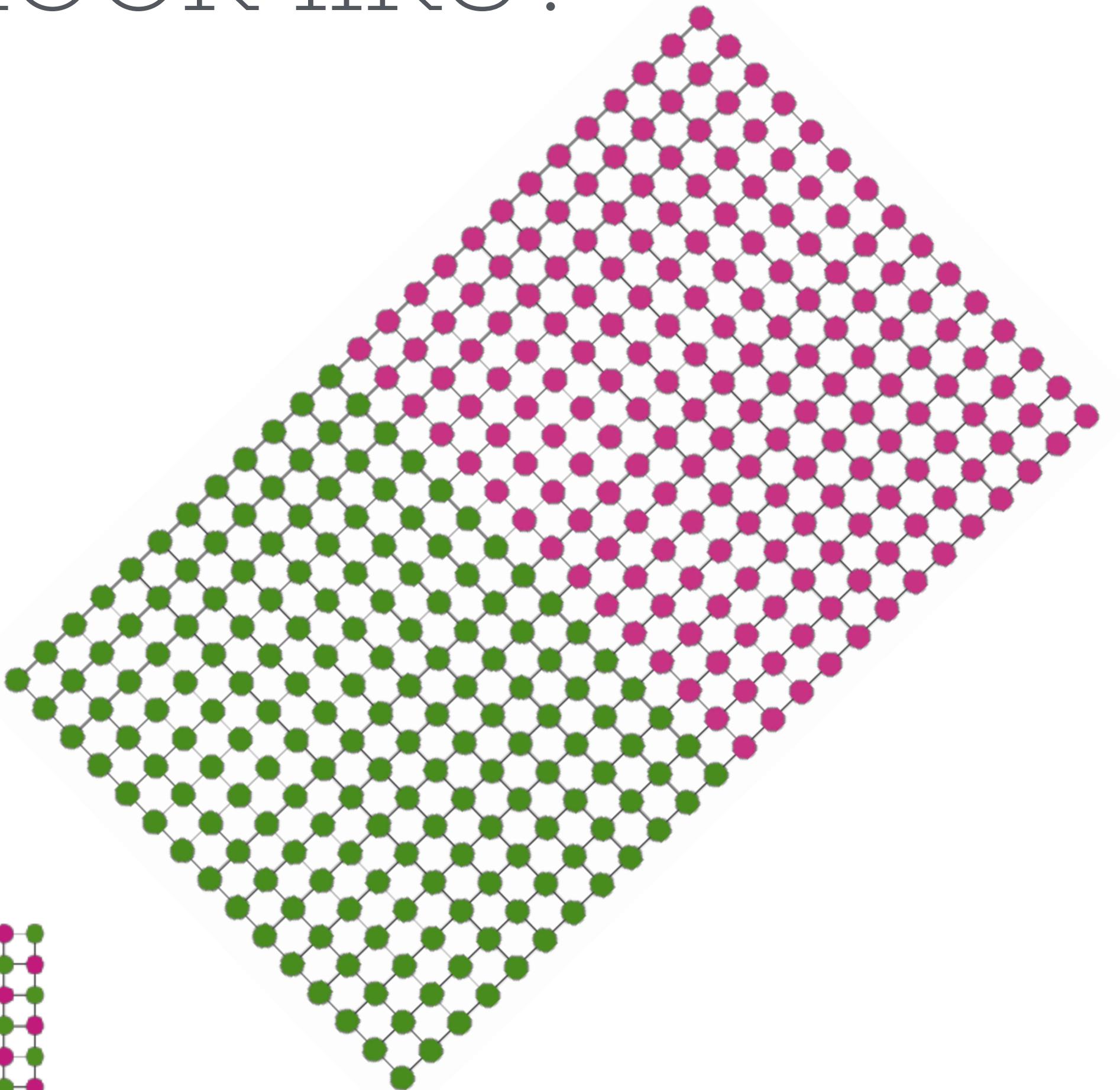
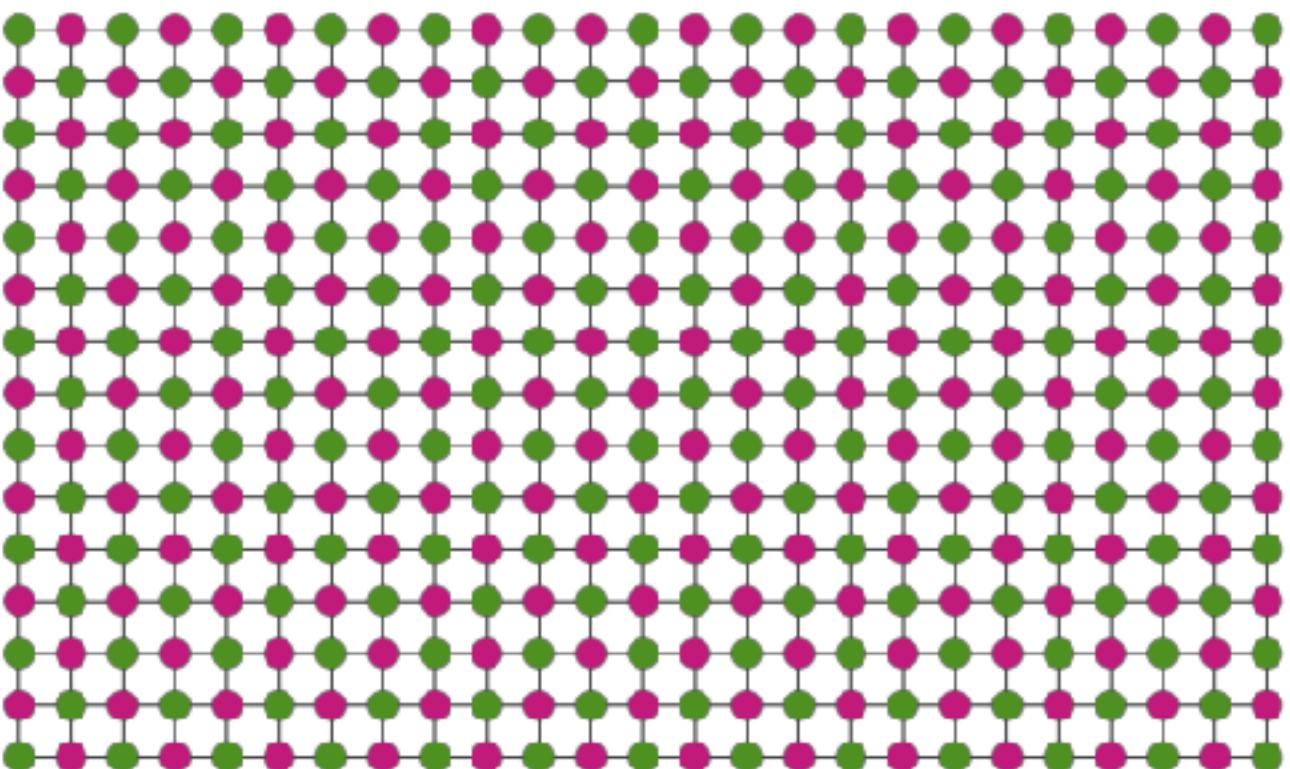
What does segregation look like?

Suppose you have a group **X** and its complement **Y**

Do they live interspersed? Clustered?

Units matter, scale matters!

recall MAUP - modifiable areal unit problem



Metrics of segregation

Suppose you have a group X and its complement Y
Have counts or shares x_i at every node of a network

Let's let \bar{x} be the average of the shares

- **Dissimilarity**: differences x_i vs \bar{x}
- **Gini**: all differences x_i vs x_j
- **Moran's I**: $I(\mathbf{x}) = \frac{\mathbf{v}^T W \mathbf{v}}{\mathbf{v}^T \mathbf{v}}$
- **Assortativity** ("capy"): share of XX edges
- **Assortativity'** : share of XX half-edges

This view is verified by historical studies that report quantitative indices of racial segregation. The standard measure of segregation is the index of **dissimilarity**, which captures the degree to which blacks and whites are evenly spread among neighborhoods in a city.⁶ Evenness is defined with respect to the racial composition of the city as a whole. If a city is 10% black, then an even residential pattern requires that every neighborhood be 10% black and 90% white. Thus, if a neighborhood is 20% black, the excess 10% of blacks must move to a neighborhood where the black percentage is under 10% to shift the residential configuration toward evenness. The index of dissimilarity gives the percentage of blacks who would have to move to achieve an “even” residential pattern—one where every neighborhood replicates the racial composition of the city.

Dissimilarity

- **Dissimilarity**: differences x_i vs \bar{x}

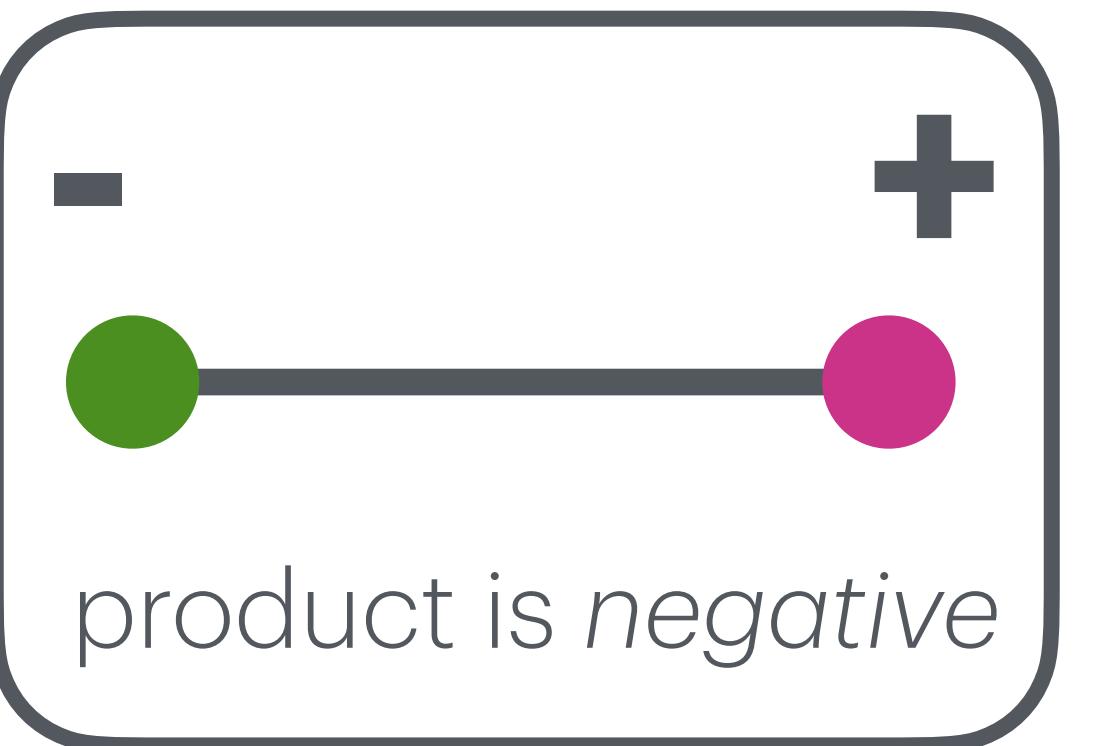
from Massey, American Apartheid

Table 2.1 Indices of black-white segregation (dissimilarity) in selected northern and southern cities: circa 1860–1870, 1910, and 1940

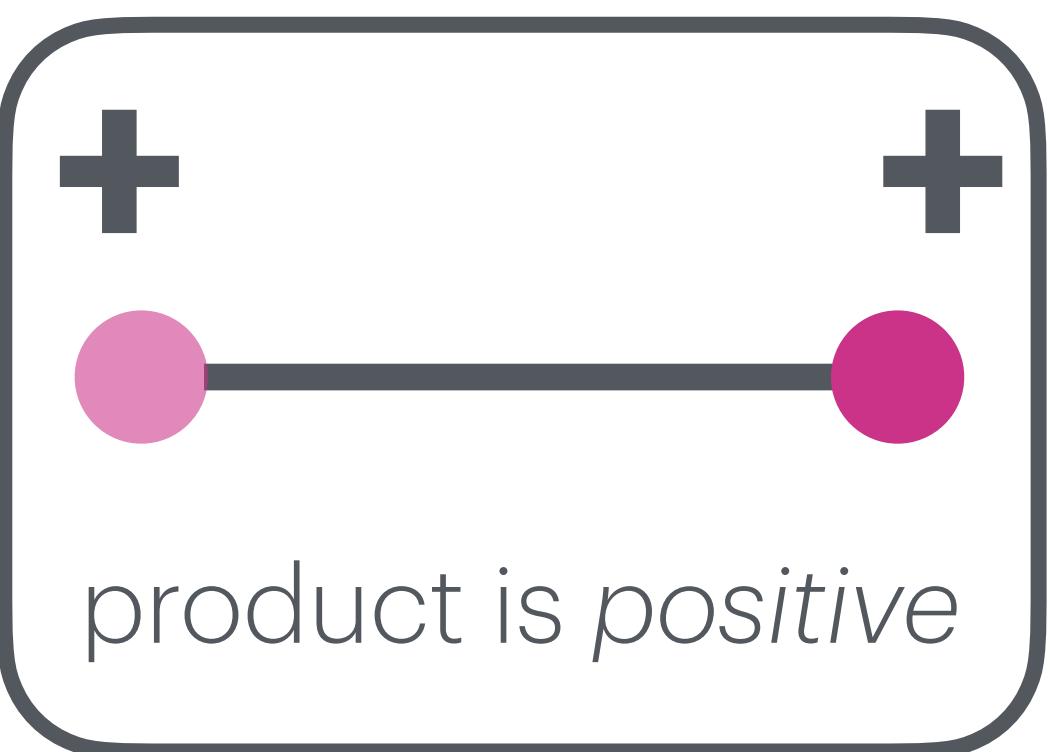
	Free blacks vs. whites, circa 1860	Blacks vs. native whites, 1910	Nonwhites vs. whites, 1940
<i>Northern cities</i>			
Boston	61.3	64.1	86.3
Chicago	50.0	66.8	95.0
Cincinnati	47.9	47.3	90.6
Cleveland	49.0	69.0	92.0
Indianapolis	57.2	—	90.4
Milwaukee	59.6	66.7	92.9
New York	40.6	—	86.8
Philadelphia	47.1	46.0	88.8
St. Louis	39.1	54.3	92.6
San Francisco	34.6	—	82.9
Wilmington	26.1	—	83.0
Average	45.7	59.2	89.2
<i>Southern cities</i>			
Augusta	—	58.8	86.9
Baltimore	22.1	—	90.1
Charleston	23.2	16.8	60.1
Jacksonville	—	39.4	94.3
Louisville	20.2	—	81.7
Mobile	29.8	—	86.6
Nashville	43.1	—	86.5
New Orleans	35.7	—	81.0
Average	29.0	38.3	81.0

Moran's I

- It's almost the definition of "spatial autocorrelation" in the whole field of geography and spatial statistics
- It comes standard in ArcGIS
- **Positive** means segregated
- **Negative** means checkerboarded
- **Zeroish** means random

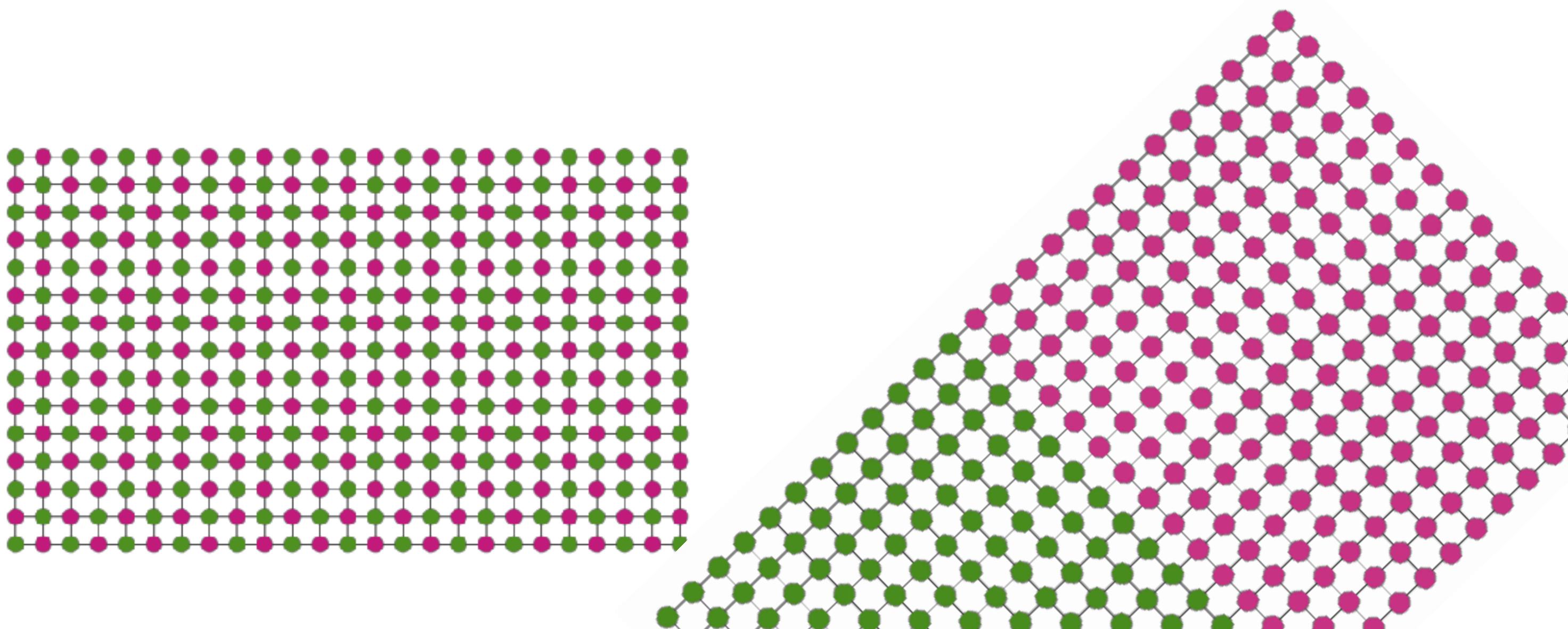


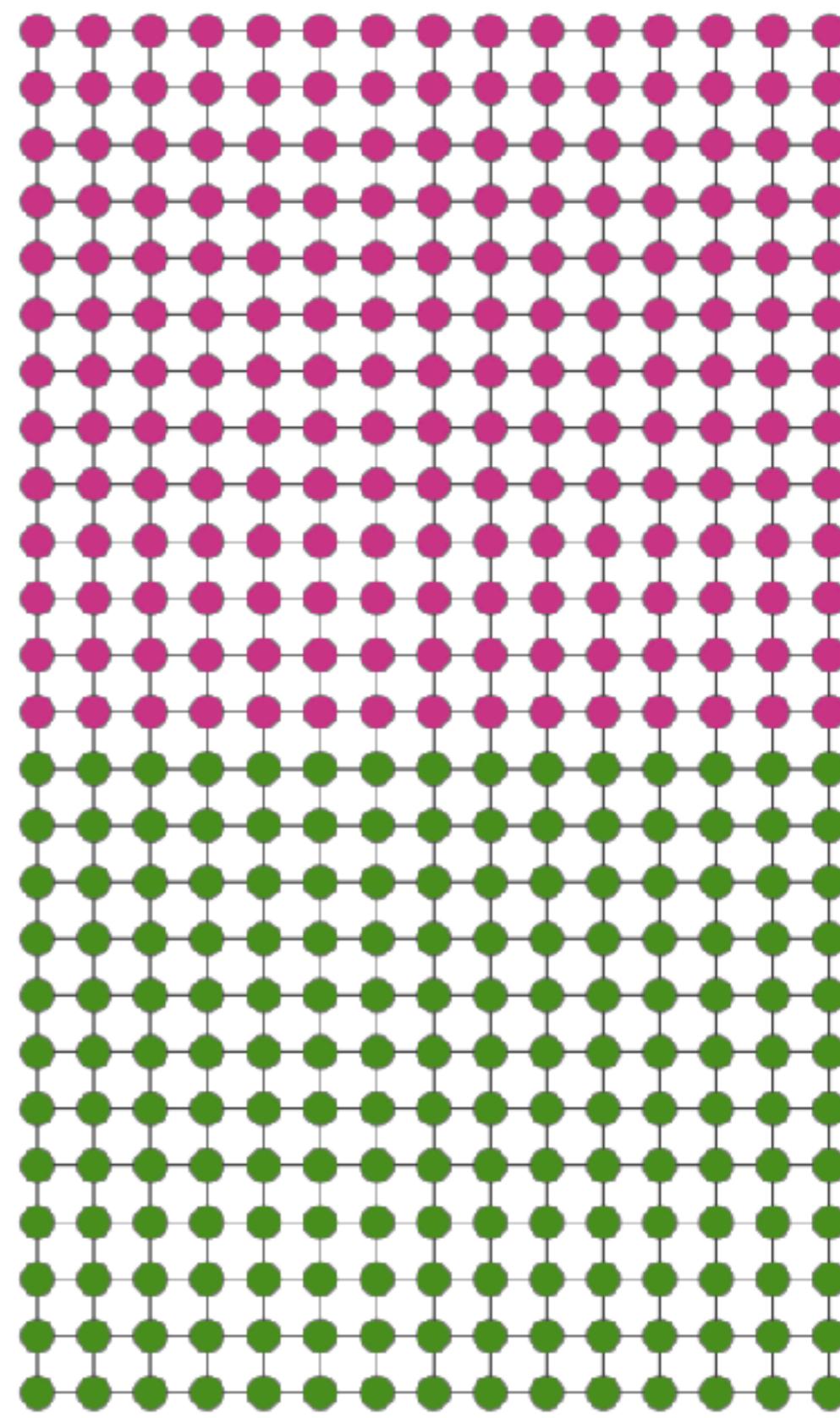
product is negative



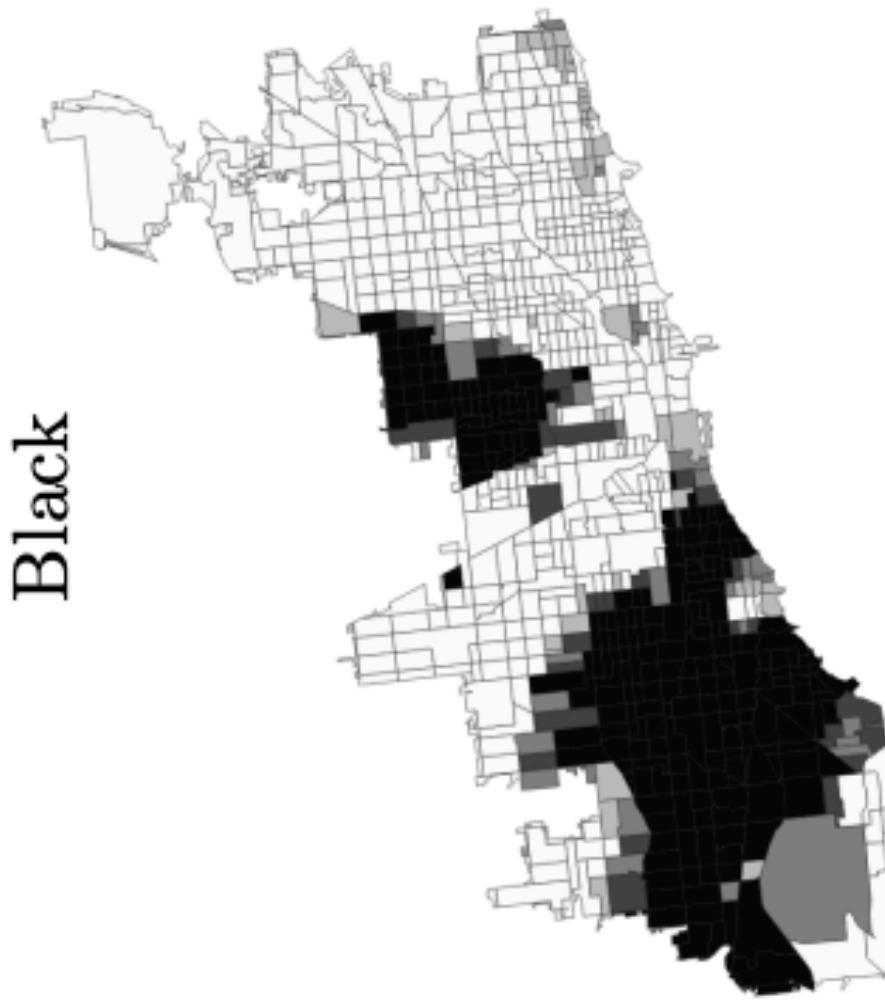
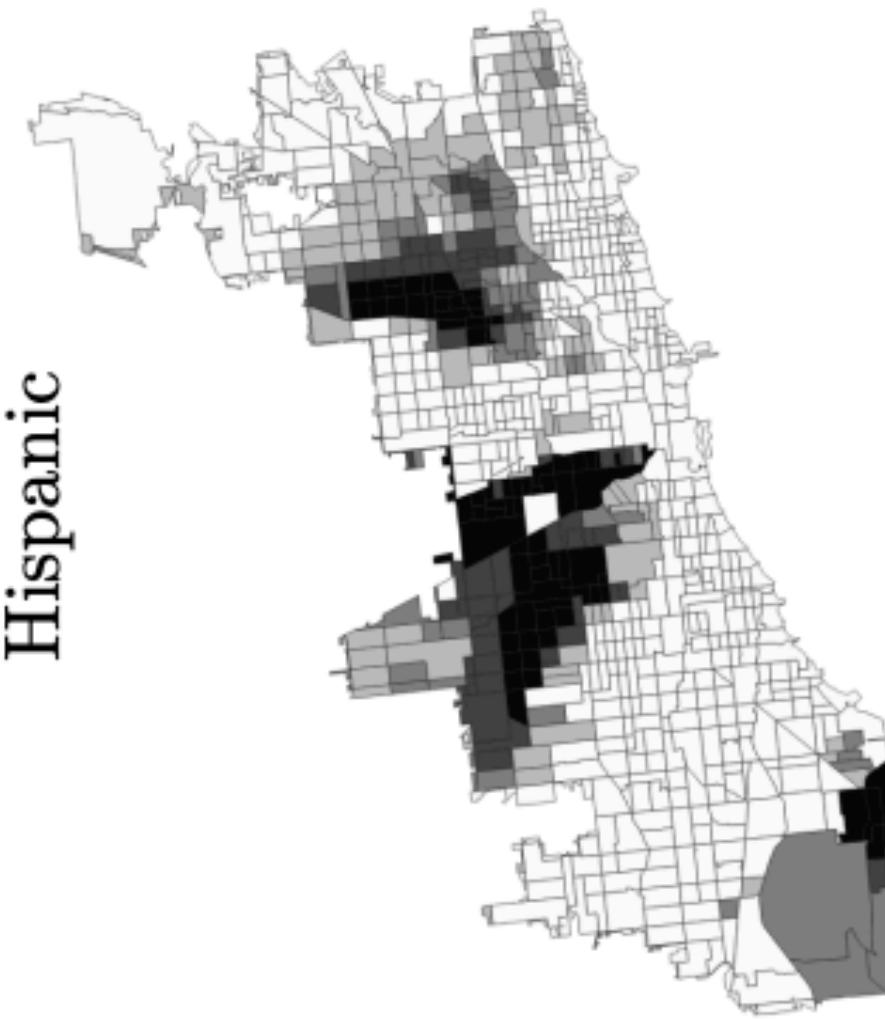
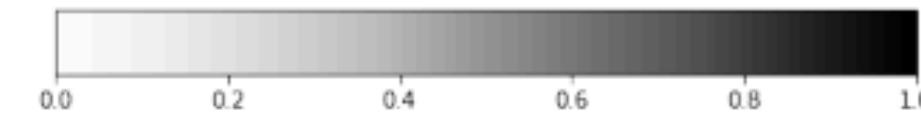
product is positive

- **Moran's I:** $I(\mathbf{x}) = \frac{\mathbf{v}^T W \mathbf{v}}{\mathbf{v}^T \mathbf{v}}$

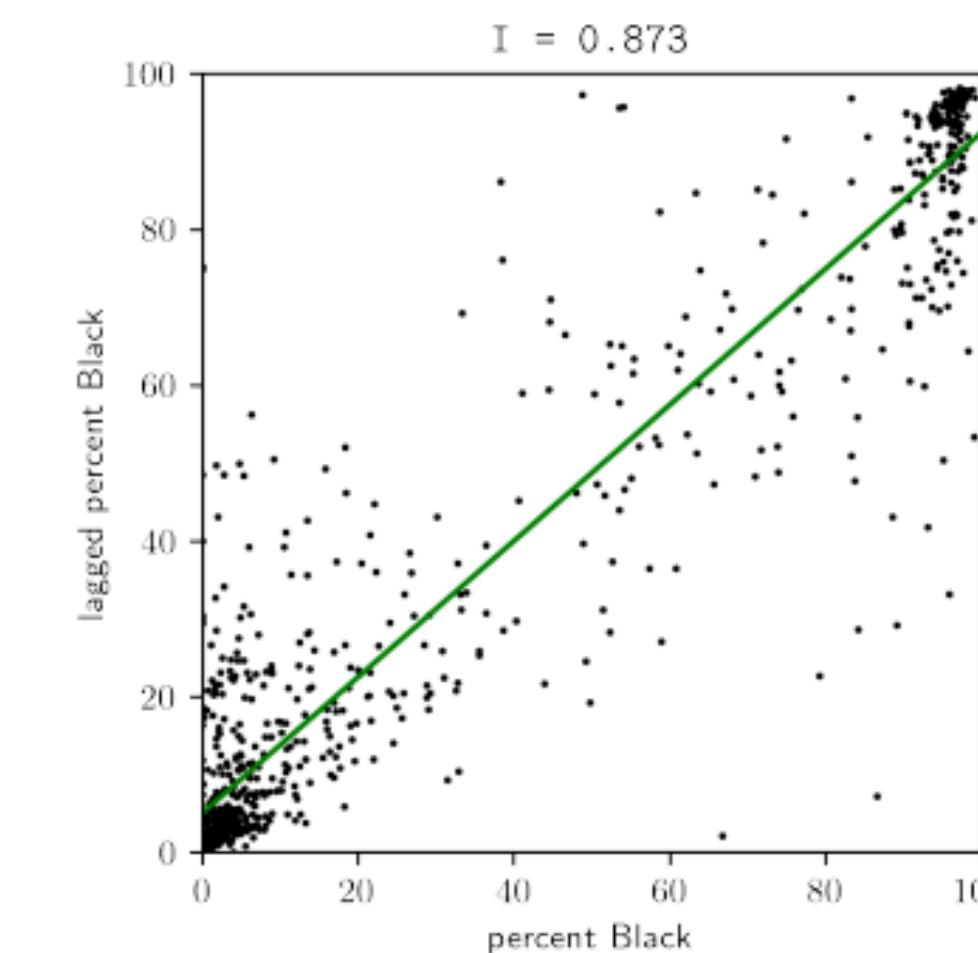
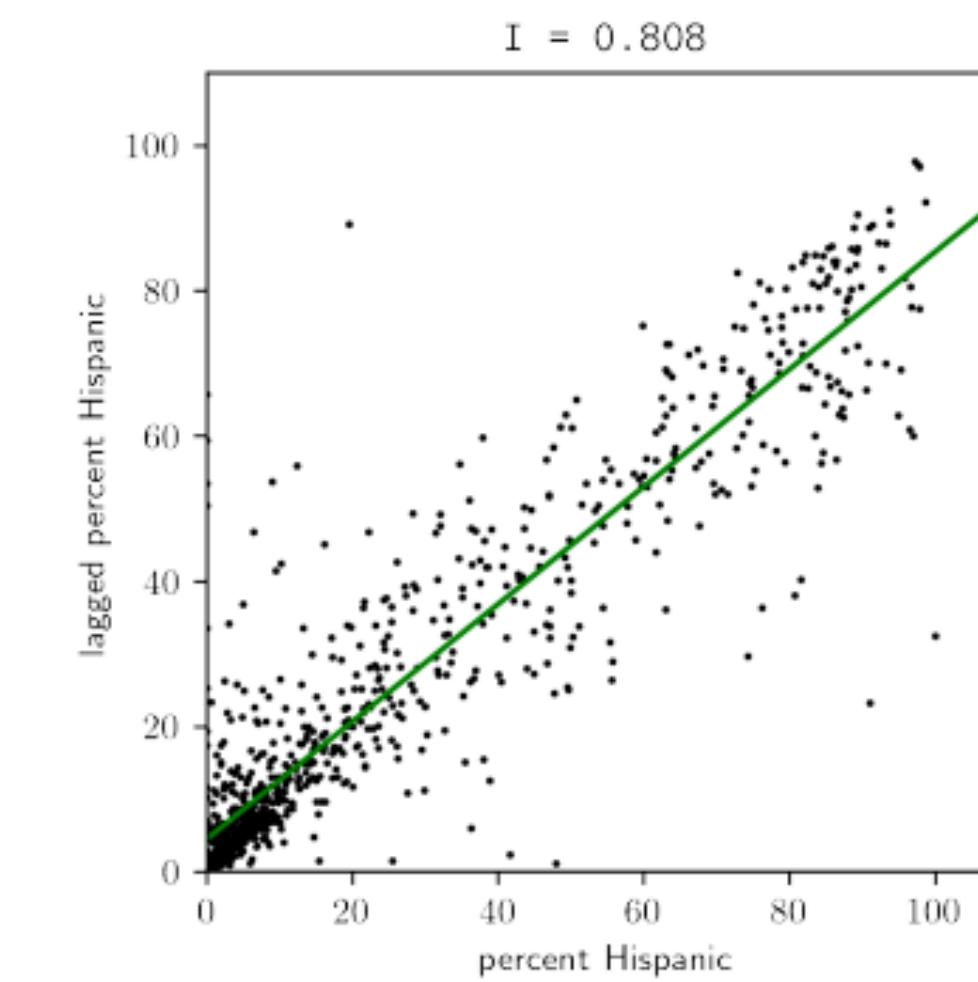




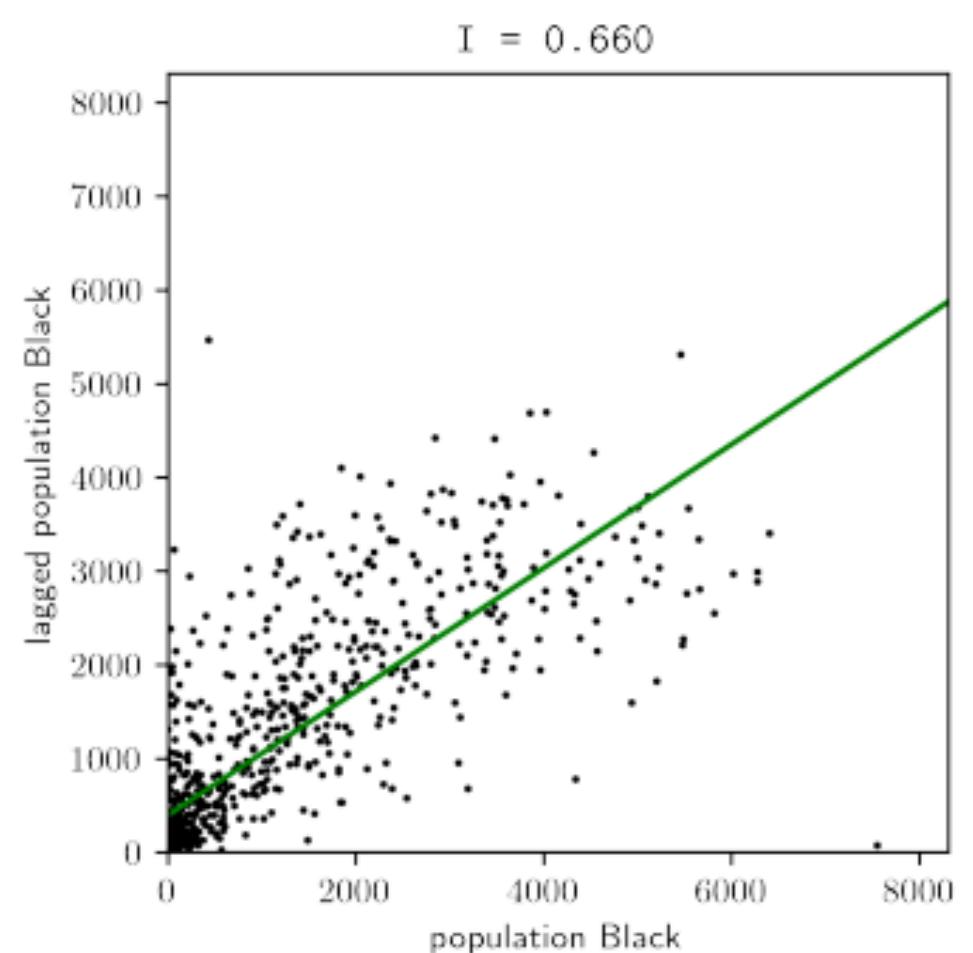
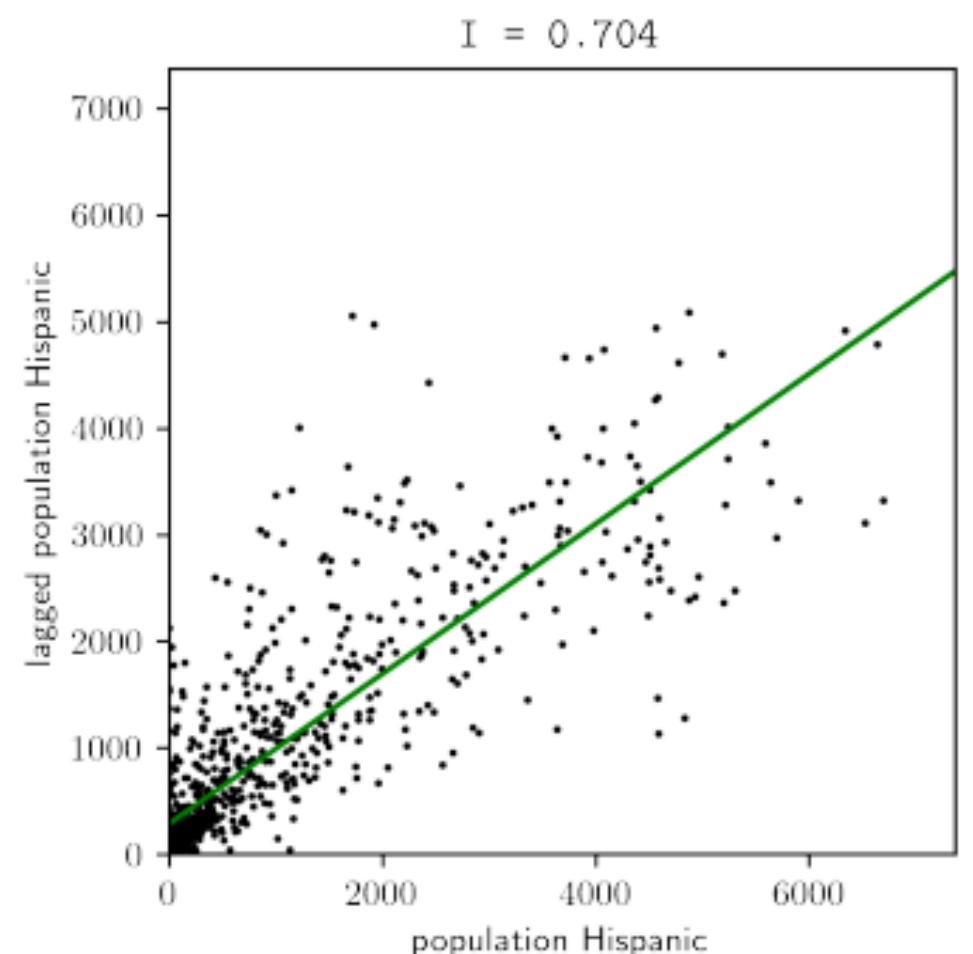
cool fact: I is the **slope** of the line that compares shares in a unit to shares in neighboring units

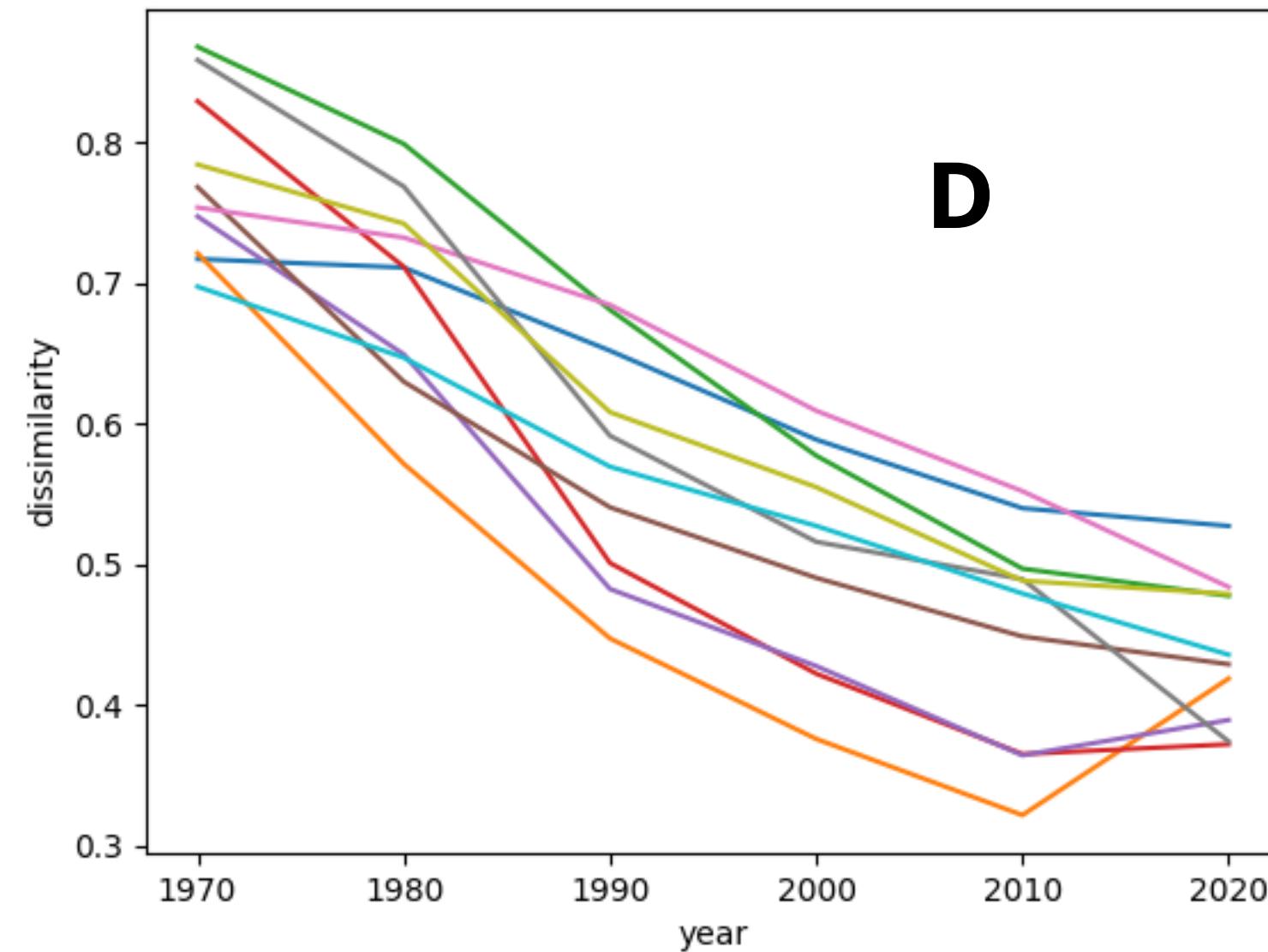


By share

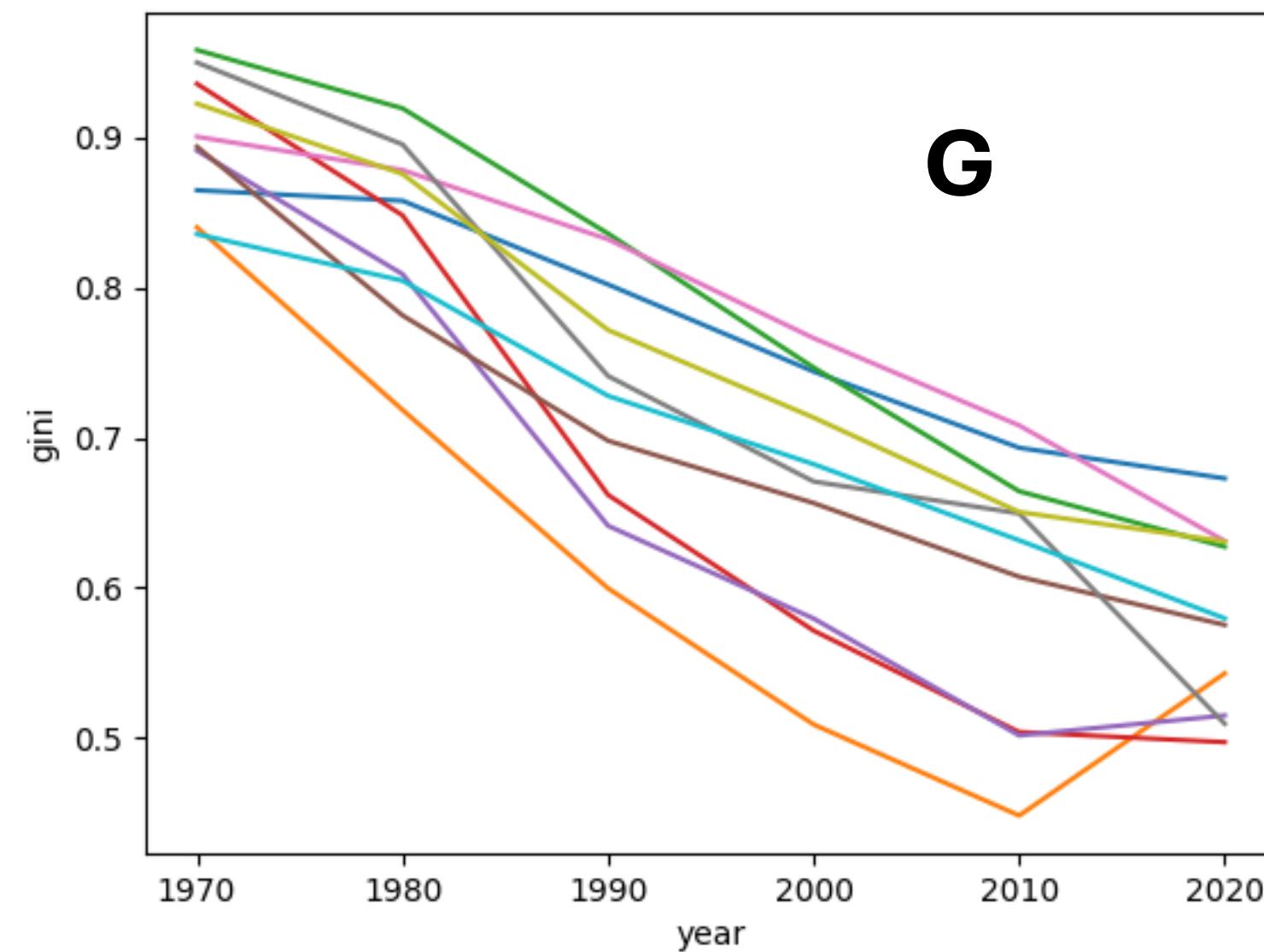


By count

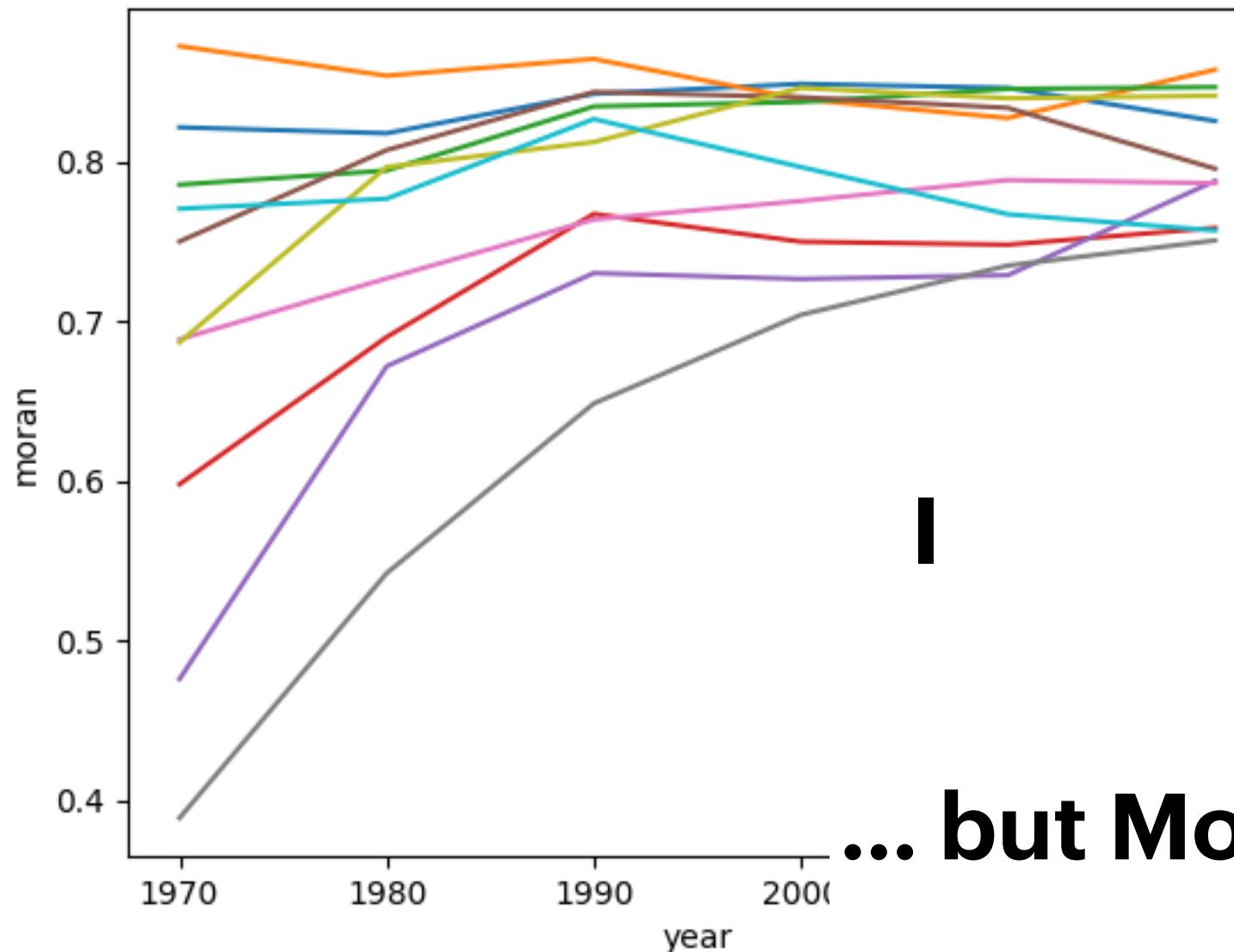
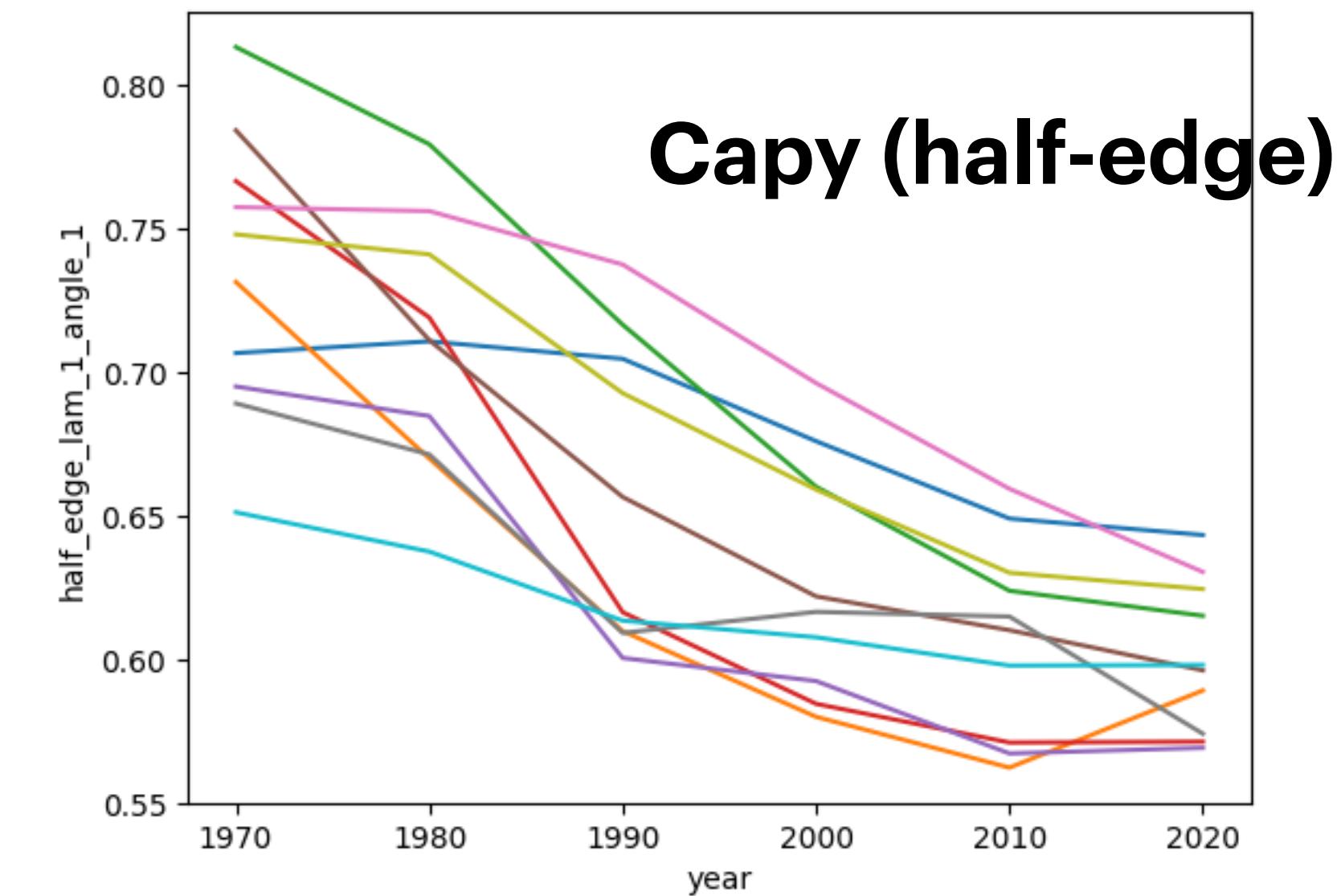
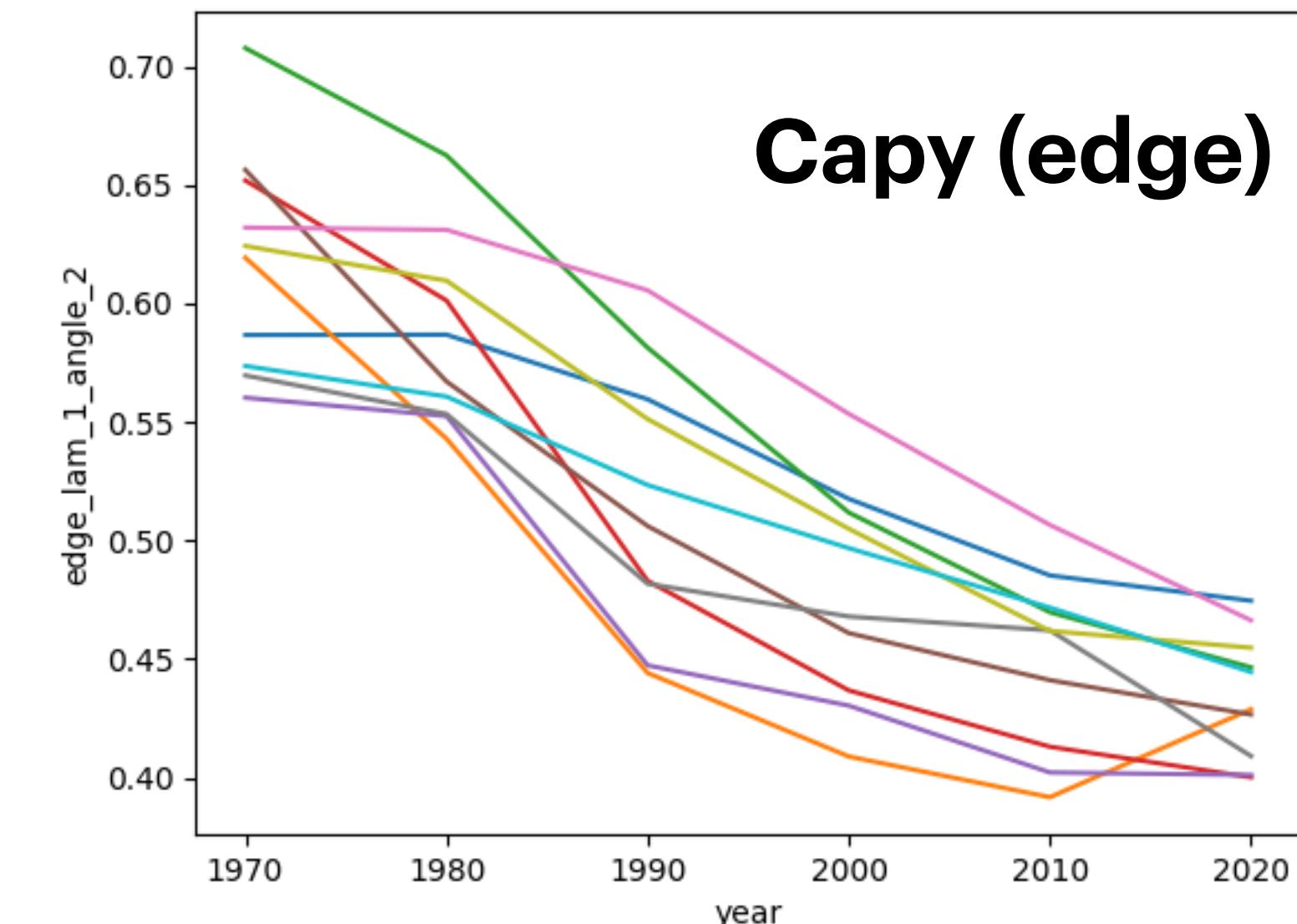




There are lots of open questions on the clustering scores.



One example: almost all of the scores agree that U.S. metros are getting less segregated over time.



... but Moran's I says the opposite! 🤔