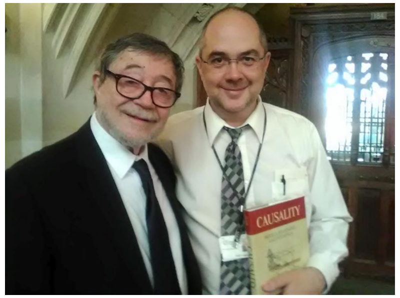
Investigating the Life Expectancies Differences in the US by Comparing Naïve and Spatial Analytic Methods across Census Tracts, Counties, and States

Modern Modeling Methods – 2024, Storrs CT,
June 25-26, 2024 **Session 1B: Modeling Spatial Data**

Slides at https://tinyurl.com/mmmlifeexp



Emil Coman Pstat, SEMNET 'moderator'; Jason Byers; Blair Johnson;

Sandro Steinbach; Peter (Xiang) Chen; Stewart Fotheringham

comanus@gmail.com

tinyurl.com/agecause

UCONN HEALTH
HEALTH DISPARITIES INSTITUTE

knowledge as a human right

Humanity, Earth, Milky Way, Universe

Research Questions (RQs) &General plan

RQ 1: Do residents from places with more racial/ethnic minorities live shorter/longer lifes? By how much?

RQ 2: Is “Socioeconomic Status” a stable construct across spatial levels?

More: How much does naïve analyses misdirect compared to proper spatial analyses?

1. Challenges of spatial data and analytics and solutions
2. Naïve/a-spatial vs. spatial modeling
3. Future extensions: 1-to-many relations; spatial factor analysis; dyadic modeling.

Common origin for:

i. Spatial ordering
ii. Path analysis
iii. Causal graphs (DAGs)
iv. Social network analysis

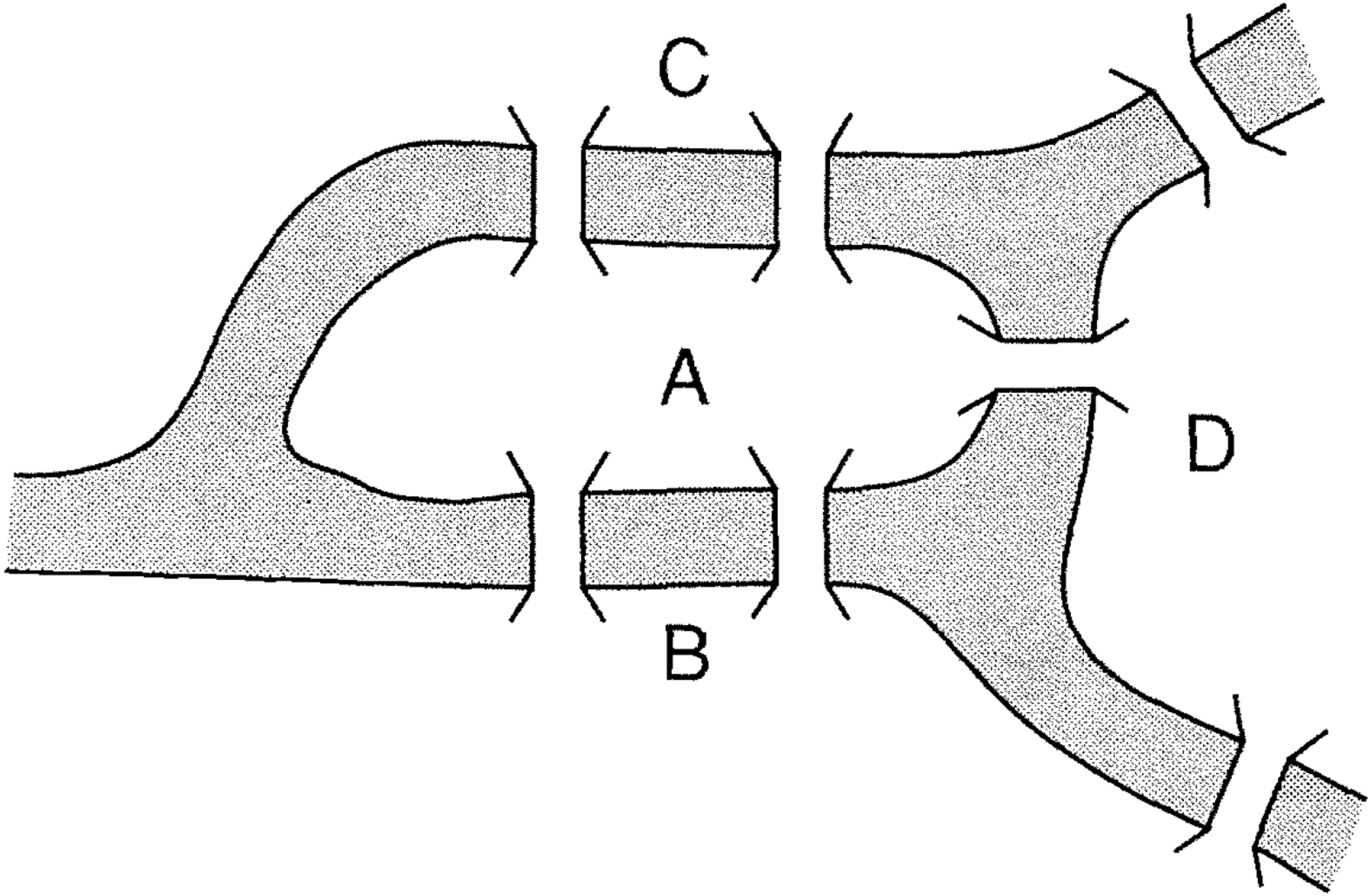
Wilson, R. J. (1996). Introduction to **Graph theory**: https://drive.google.com/file/d/1FaBttSZ_APfmD5b-CJ9TSldIxt8fSduJ/view?usp=sharing

p. 12: “The degree of a vertex v of *G* is the number of edges incident with v, and is written deg(v); in calculating the degree of v, we usually make the convention that a loop at v contributes 2 (rather than 1) to the degree of v. A vertex of degree 0 is an isolated vertex and a vertex of degree 1 is an end-vertex.”

“Note that in any graph the sum of all the vertex-degrees is an even number – in fact, twice the number of edges, since each edge contributes exactly 2 to the sum. This result, due essentially to Leonhard Euler in 1736, is called the **handshaking lemma**. It implies that if several people shake hands, then the total number of hands shaken must be even - precisely because just two hands are involved in each handshake.”

Graph theory and space

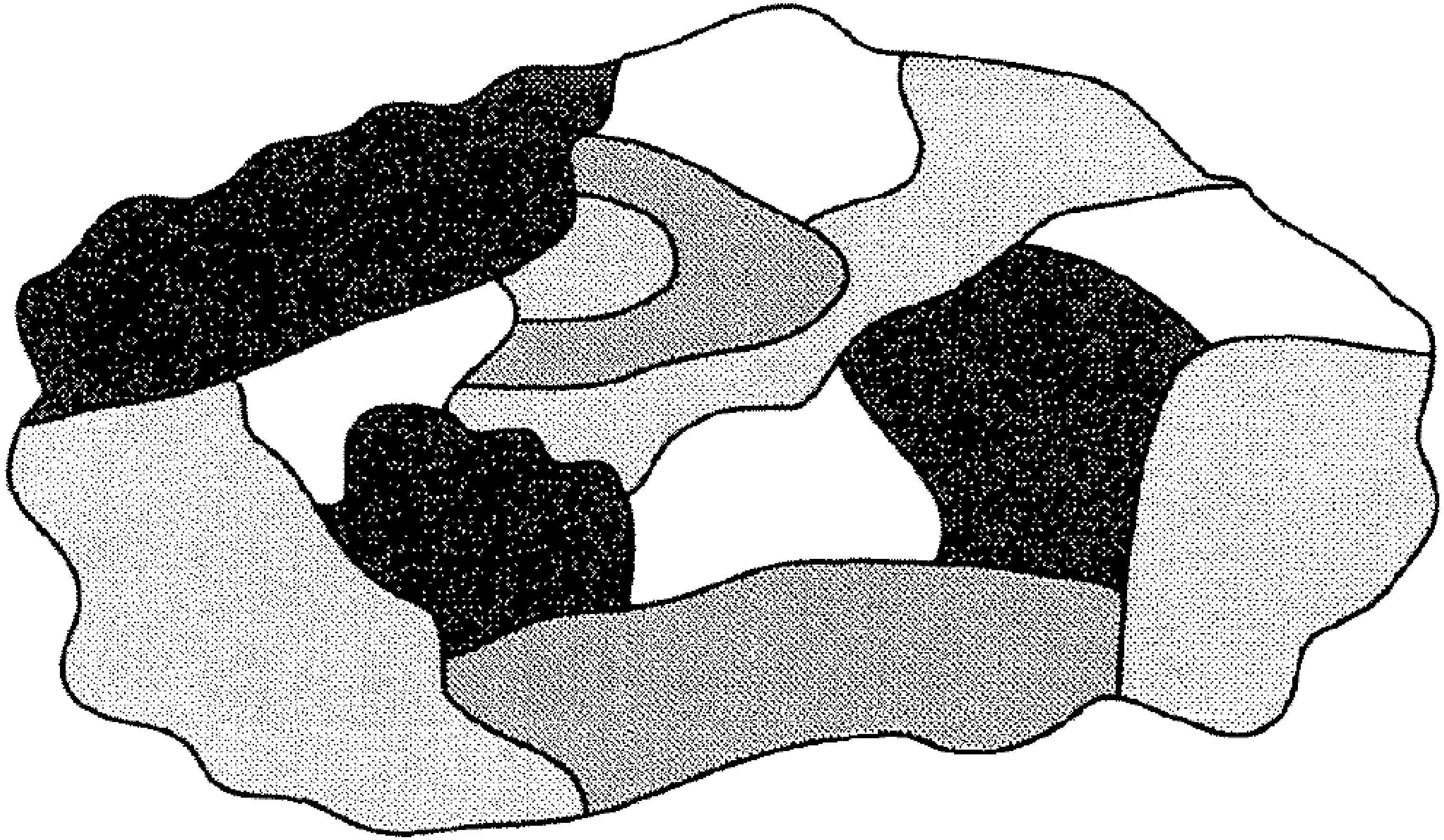
p. 31: ‘The name 'Eulerian’ arises from the fact that Euler was the first person to solve the famous Konigsberg bridges problem which asks whether you can cross each of the seven bridges in Fig. 6.4 exactly once and return to your starting point. This is equivalent to asking whether the graph in Fig. 6.5 has an Eulerian trail.”



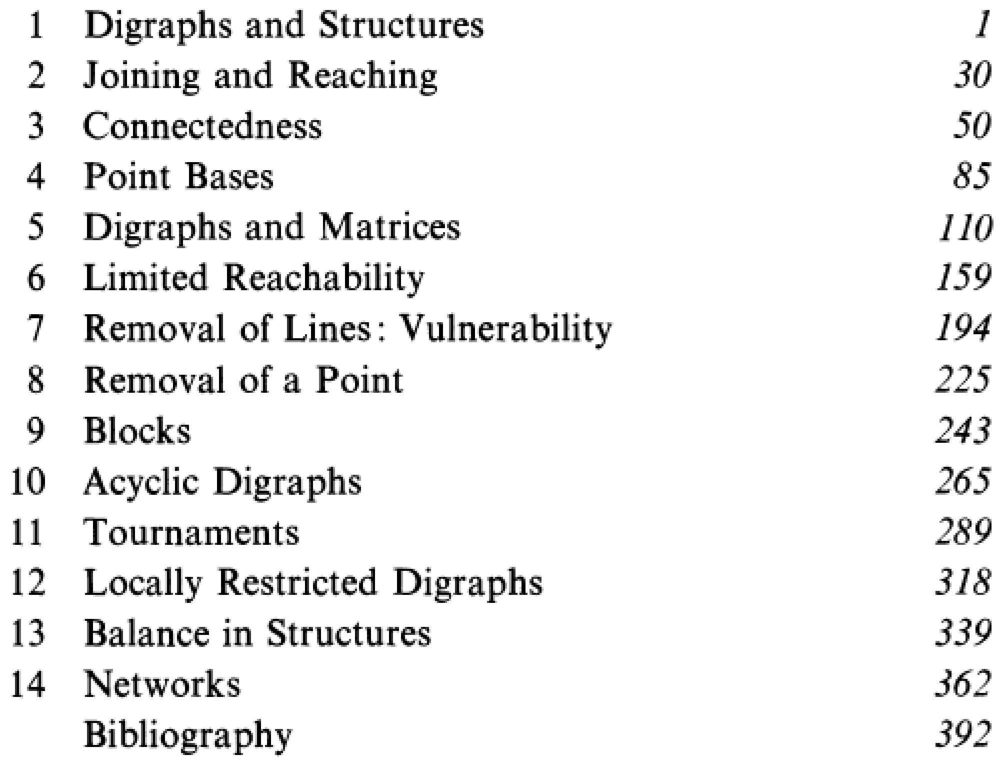
Wilson, R. J. (1996). Introduction to **Graph theory**: Longman. https://drive.google.com/file/d/1FaBttSZ_APfmD5b-CJ9TSldIxt8fSduJ/view?usp=sharing

Graph theory and space

p. 88: ‘The four-colour problem arose historically in connection with the colouring of maps. Given a map containing several countries, we may ask how many colours are needed to colour them so that no two countries with a boundary line in common share the same colour. Probably the most familiar form of the four-colour theorem is the statement that every map can be coloured with only four colours. For example, Fig. 19.1 shows a map that has been coloured with four colours.”



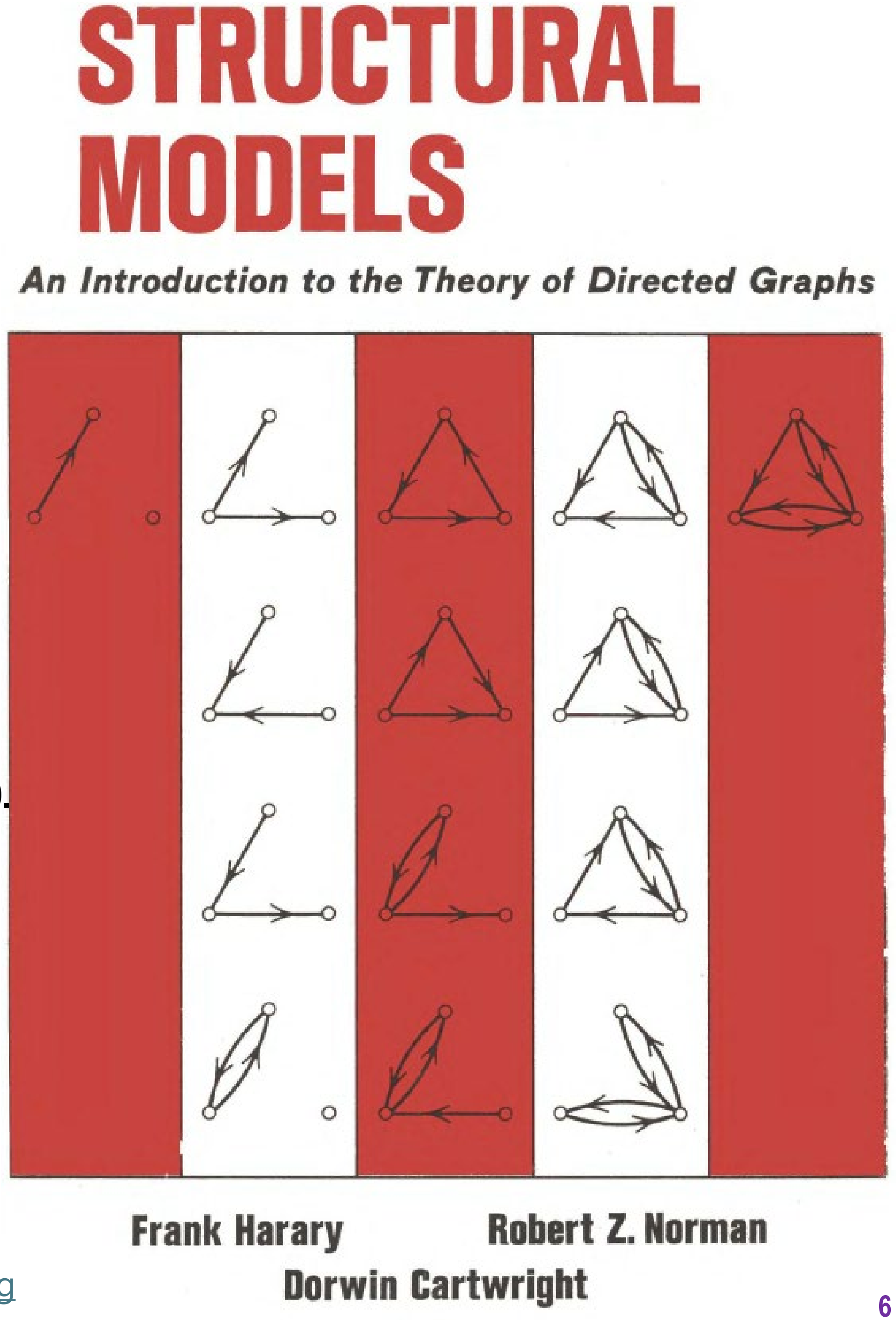
Wilson, R. J. (1996). Introduction to **Graph theory**: Longman. https://drive.google.com/file/d/1FaBttSZ_APfmD5b-CJ9TSldIxt8fSduJ/view?usp=sharing



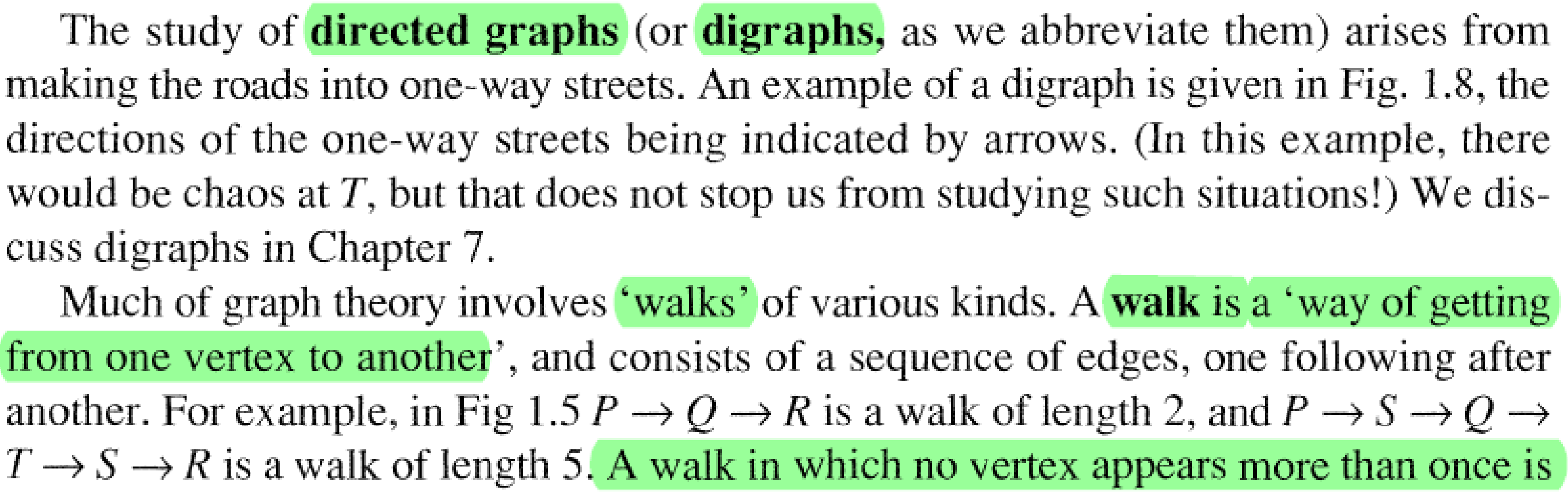
Theorem 2.1. If in a digraph, v is reachable from u and w is reachable from v, then d(u, w) ~ d(u, v) + d(v, w).

"Three matrices are of particular value : the ***reachability matrix*** *R(D), which indicates whether a point vi can reach a point vi; the* ***connectedness matrix*** *C(D), which shows the connectedness of every pair of points of D; and the* ***distance matrix*** *N(D), which gives the distance from any point to any other." p. 110*

Harary, F., Norman, R. Z., & Cartwright, D. **(1965).** Structural models: An introduction to the theory of directed graphs
https://drive.google.com/file/d/1_dsCnxuUXpF8DdMB2hA4NeA4hvYgRpGL/view?usp=sharing



Graph Grammar



The study of **directed graphs** (or **digraphs,** as we abbreviate them) arises from making the roads into one-way streets. An example of a digraph is given in Fig. 1.8, the directions of the one-way streets being indicated by arrows. (In this example, there would be chaos at T, but that does not stop us from studying such situations!) We discuss digraphs in Chapter 7.

Much of graph theory involves 'walks' of various kinds. A **walk** is a 'way of getting from one vertex to another', and consists of a sequence of edges, one following after another. For example, in Fig 1.5 $P \to Q \to R$ is a walk of length 2, and $P \to S \to Q \to T \to S \to R$ is a walk of length 5. A walk in which no vertex appears more than once is

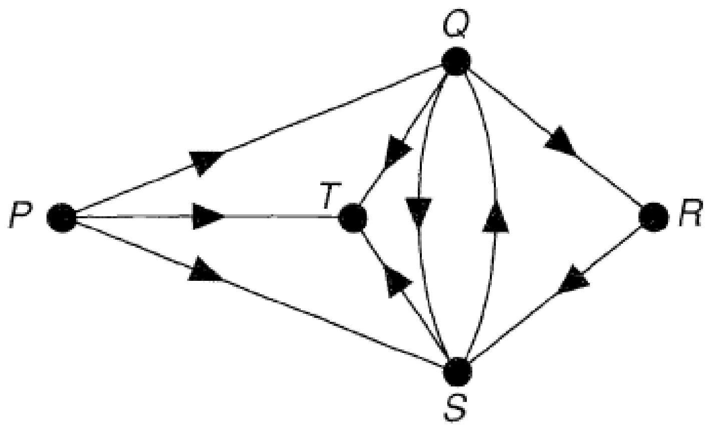


Fig. 1.8

called a **path**; for example, $P \to T \to S \to R$ is a path. A walk of the form $Q \to S \to T \to Q$ is called a **cycle**.

Wilson, R. J. (1996). Introduction to **Graph theory**: Longman. P. 3 https://drive.google.com/file/d/1FaBttSZ_APfmD5b-CJ9TSldIxt8fSduJ/view?usp=sharing

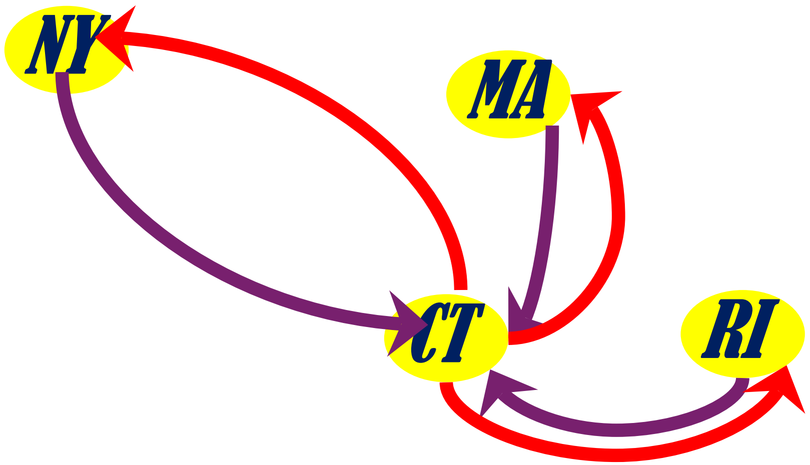
2 Networks with spatial data

Modeling/analyzing spatial data requires handling 2 OVERLAYED networks:

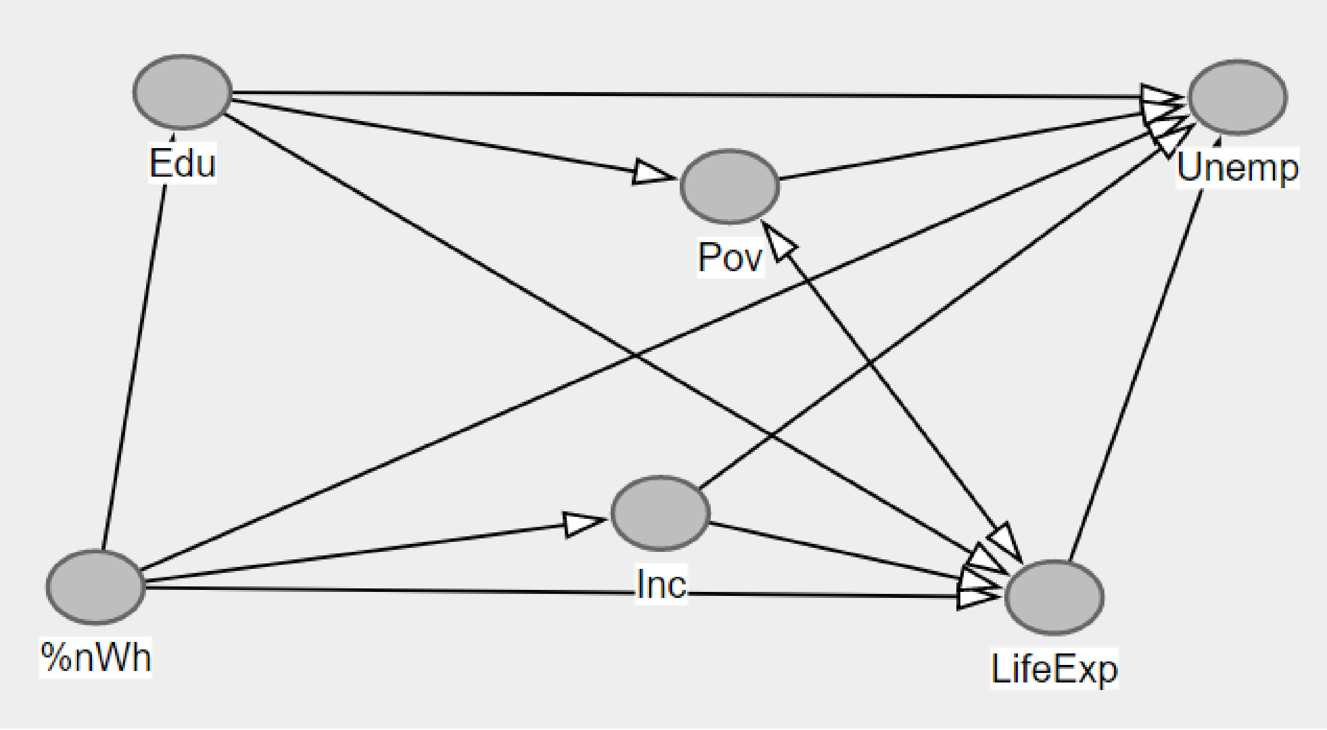
1. Among cases/regions in the data -> ‘contagion’ between ‘individuals’
   * Same happens with dyads, or groups, or time: different structures though
2. Among variables -> relations: causal or otherwise

2 Networks with spatial data

i. State level networking
CT only and neighbors: etc.



ii. Variable 'network'" e.g.
Life Expectancy data- informed model
http://dagitty.net/dags.html?id=4TETpl



Yes, there is cyclical/feedback influence at work: Stata's ***sp*** module estimates total effects too, that takes these back-and-forth's into account.

Interference and causal issues

“The principles of covariate control in the presence of interference are straightforward: like in the case of no interference, they follow from the fact that **all backdoor paths from treatment to outcome must be blocked by a measured set of covariates**.
However, without taking the time to draw the operative causal DAG with interference it is easy to make mistakes, like controlling only for individual-level covariates when block-level covariates are necessary to identify the causal effect of interest.” [1]:565

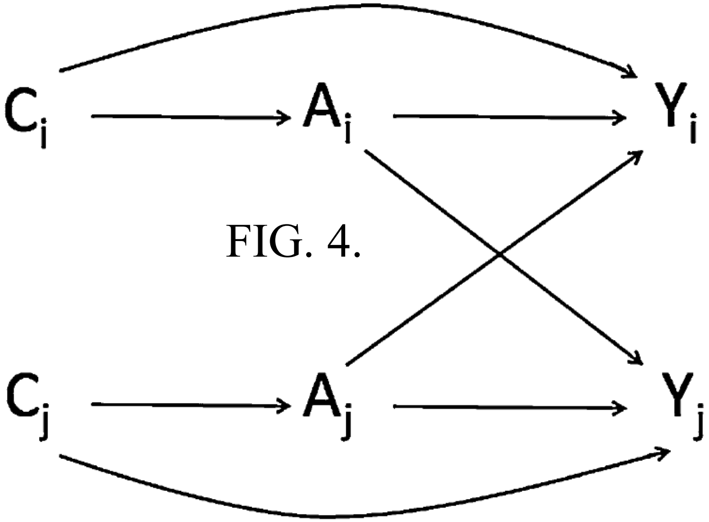
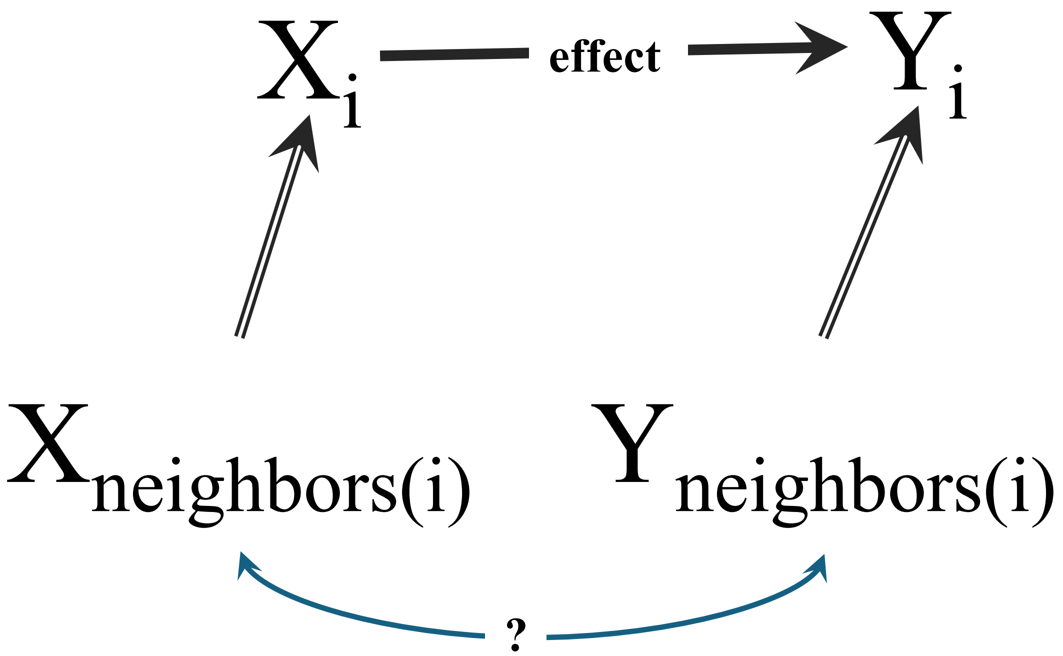


FIG. 4.

“If the individuals in the block share no common causes of A or Y , as in the DAG in Figure 4, then Ci suffices to block the backdoor paths from Ai to Yi and from Aj to Yi and, therefore, exchangeability for the effect of **A** on Yi holds conditional on Ci.
That is, Yi(ai, aj) ∐ **A** |Ci for all i.” [1]:565 *[subscripts upgraded for clarity]*

1. Ogburn, E. L., & VanderWeele, T. J. (2014). Causal Diagrams for Interference. Statistical Science, 29(4), 559-578.

Spatial interference: interference by contagion



X_i effect Y_i

$X_{neighbors(i)}$ $Y_{neighbors(i)}$

?

Direct interference
“if individual *i* receives treatment and individual *j* does not, individual *j* may be nevertheless be exposed to the treatment of individual *i*”

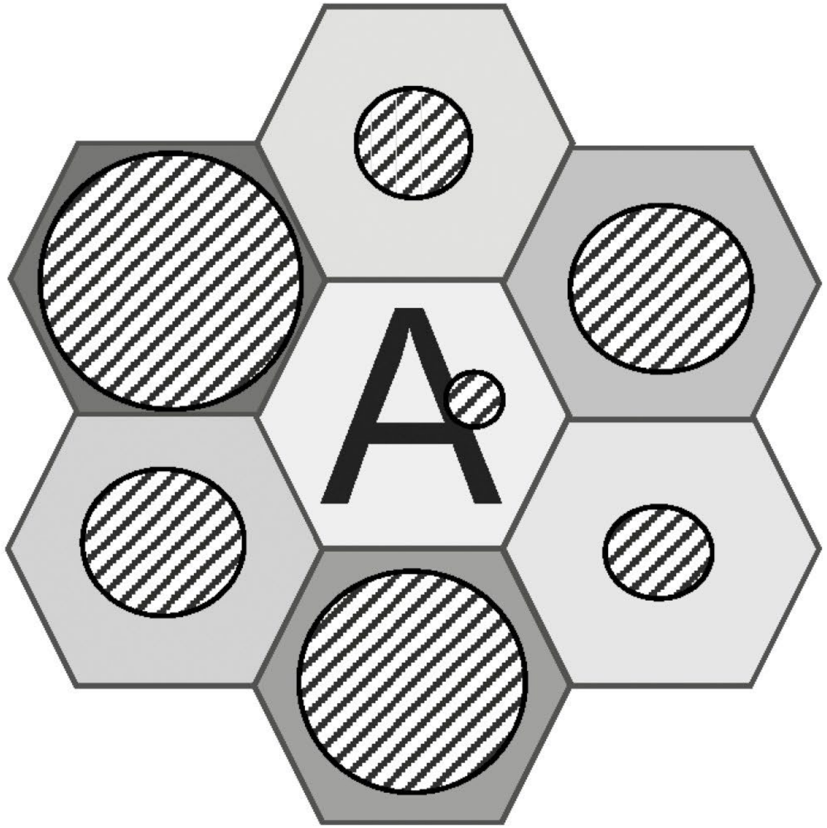
Interference by contagion
Via the first individual’s outcome - It does not represent a direct causal pathway from the exposed individual to another individual’s outcome, but rather a pathway mediated by the outcome of the exposed individual.

Allocational interference
Treatment in this setting allocates individuals to groups; through interactions within a group individuals’ characteristics may affect one another.
“An example that often arises in the social science literature is the allocation of children to schools or of children to classrooms within schools”
[1]:565

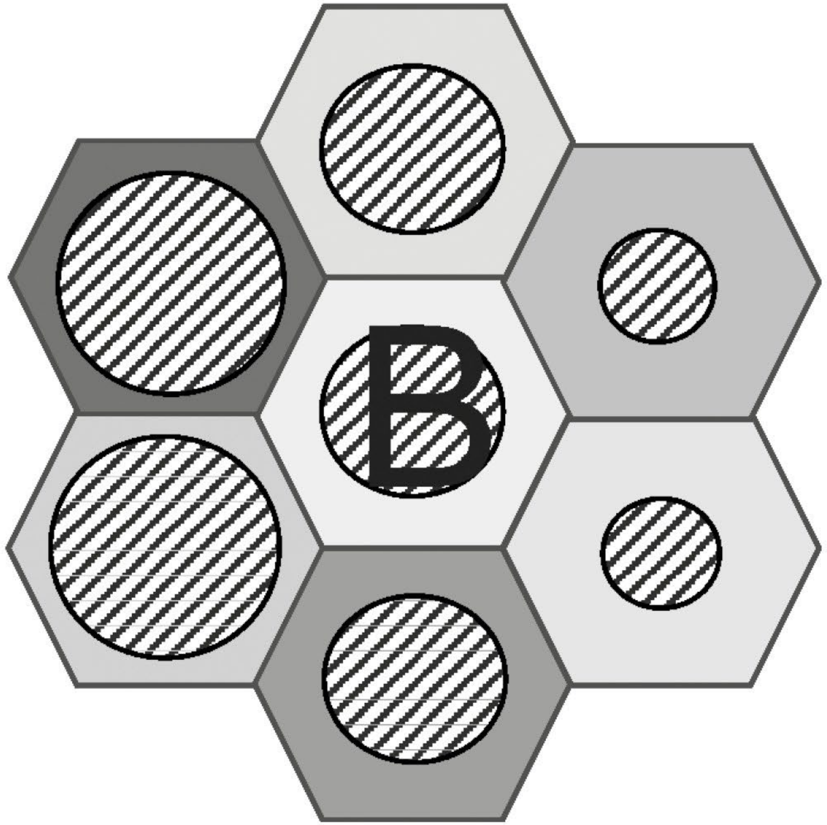
[1] Ogburn, E. L., & VanderWeele, T. J. (2014). Causal Diagrams for Interference. Statistical Science, 29(4), 559-578.

Why spatial analytics is needed

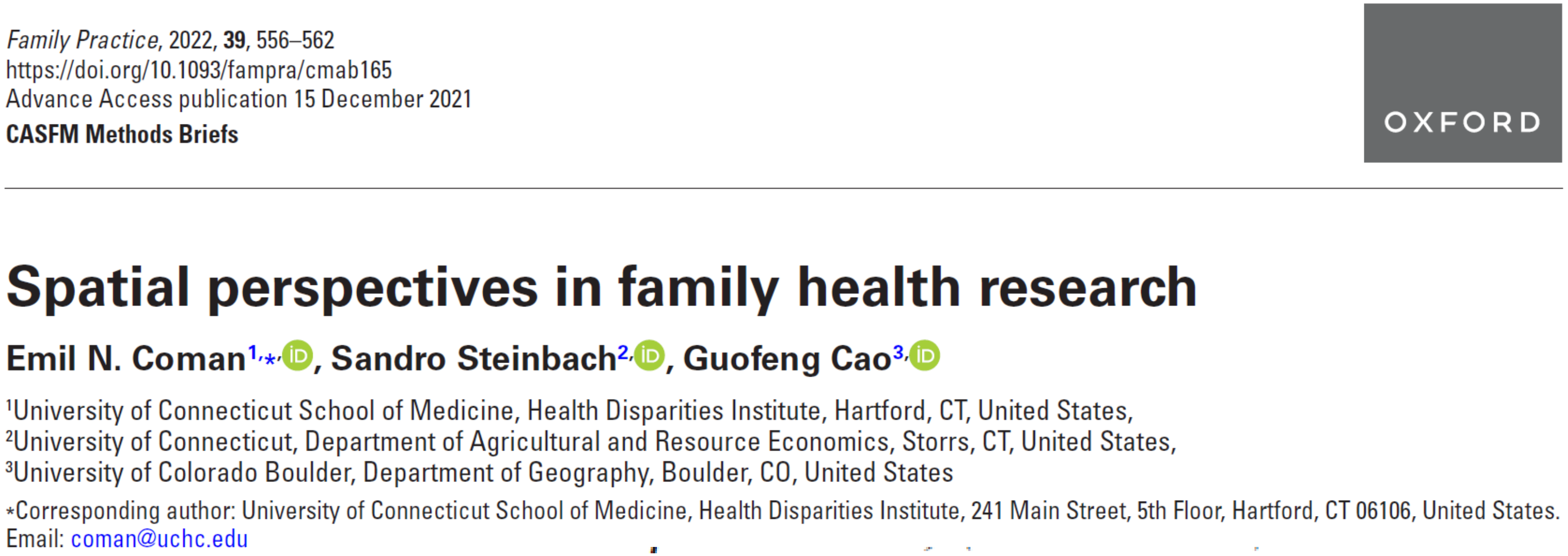
A: Spatial randomness (independence)
of both shades and circles, and
shade-circle association



B: Spatial similarity of circles
(decreasing from left to right), and
<u>no</u> shade-circle association



Spatial perspectives in family health research https://academic.oup.com/fampra/article/39/3/556/6463006



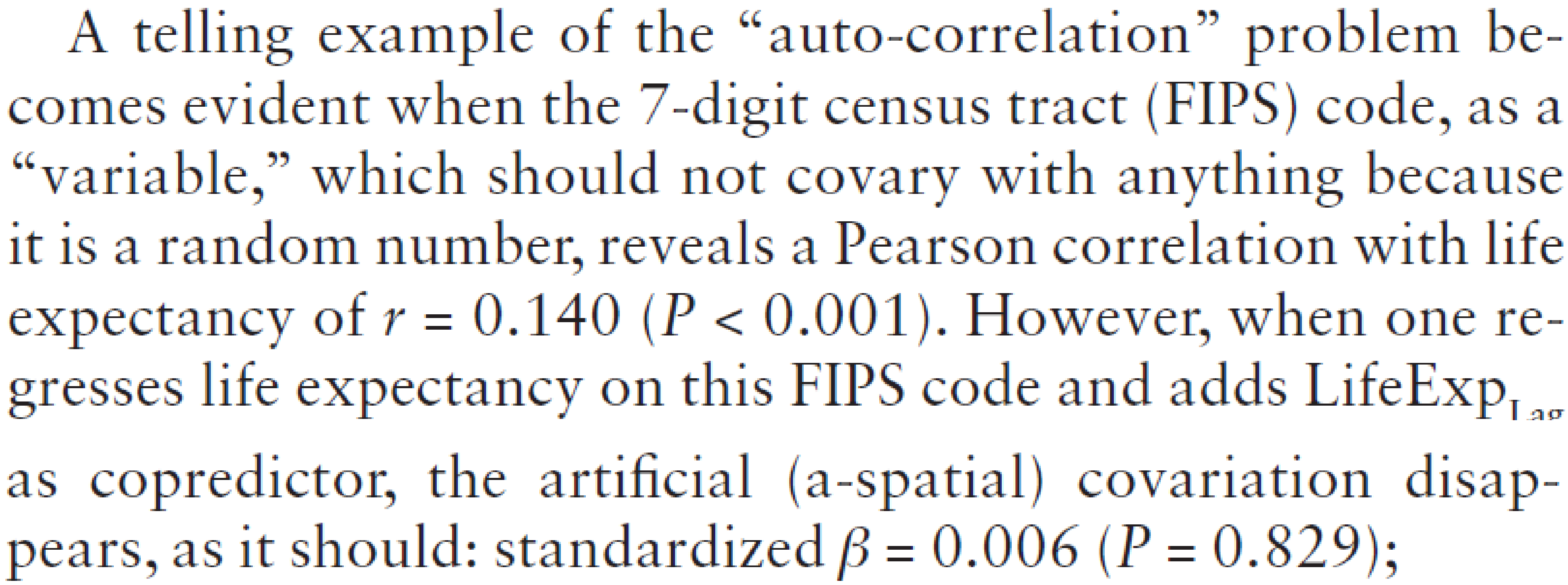
Family Practice, 2022, **39**, 556–562
https://doi.org/10.1093/fampra/cmab165
Advance Access publication 15 December 2021
CASFM Methods Briefs

Spatial perspectives in family health research

Emil N. Coman[1,*], **Sandro Steinbach**[2], **Guofeng Cao**[3]

[1]University of Connecticut School of Medicine, Health Disparities Institute, Hartford, CT, United States,
[2]University of Connecticut, Department of Agricultural and Resource Economics, Storrs, CT, United States,
[3]University of Colorado Boulder, Department of Geography, Boulder, CO, United States

*Corresponding author: University of Connecticut School of Medicine, Health Disparities Institute, 241 Main Street, 5th Floor, Hartford, CT 06106, United States.
Email: coman@uchc.edu

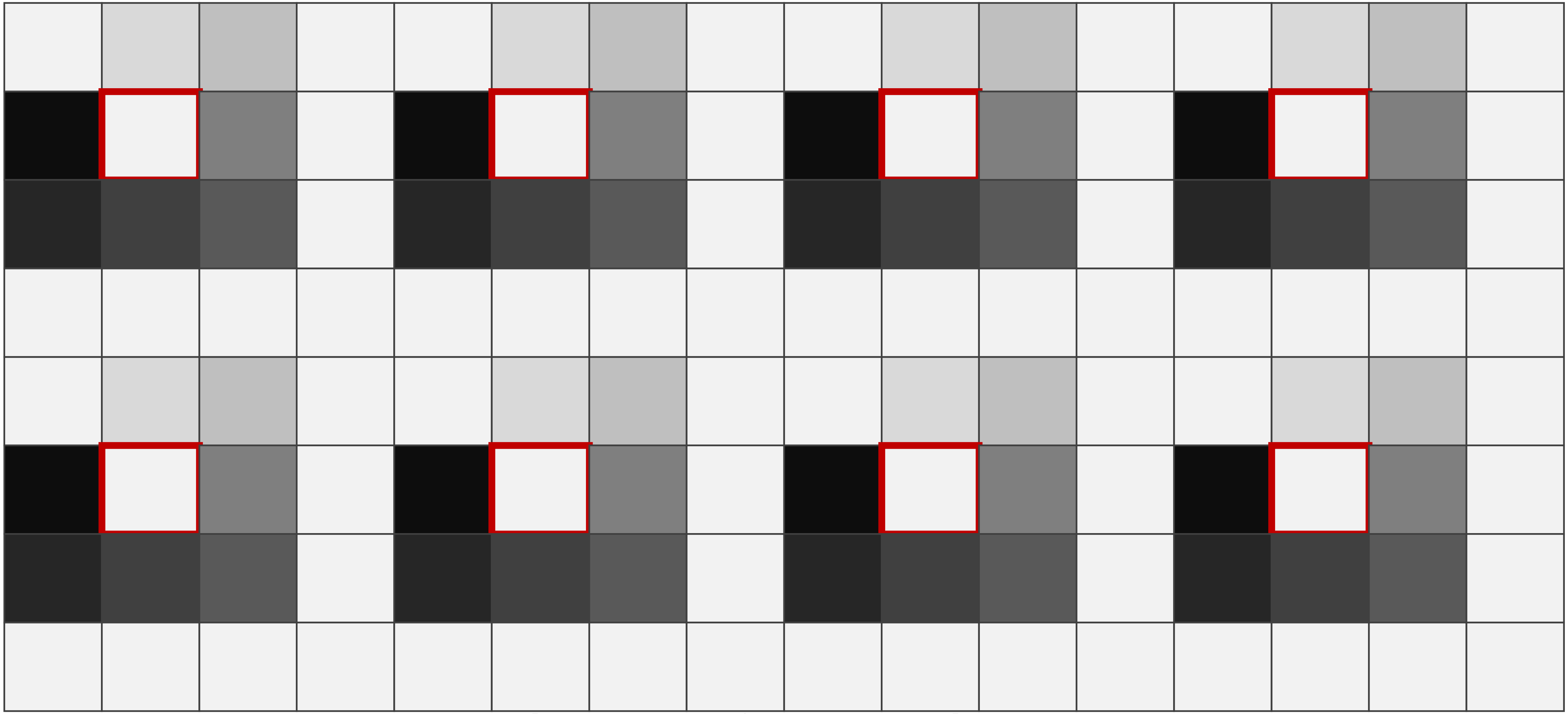


A telling example of the "auto-correlation" problem becomes evident when the 7-digit census tract (FIPS) code, as a "variable," which should not covary with anything because it is a random number, reveals a Pearson correlation with life expectancy of $r = 0.140$ ($P < 0.001$). However, when one regresses life expectancy on this FIPS code and adds LifeExp$_{Lag}$ as copredictor, the artificial (a-spatial) covariation disappears, as it should: standardized $\beta = 0.006$ ($P = 0.829$);

Spatial perspectives in family health research

Intuition for minimum Moran's I

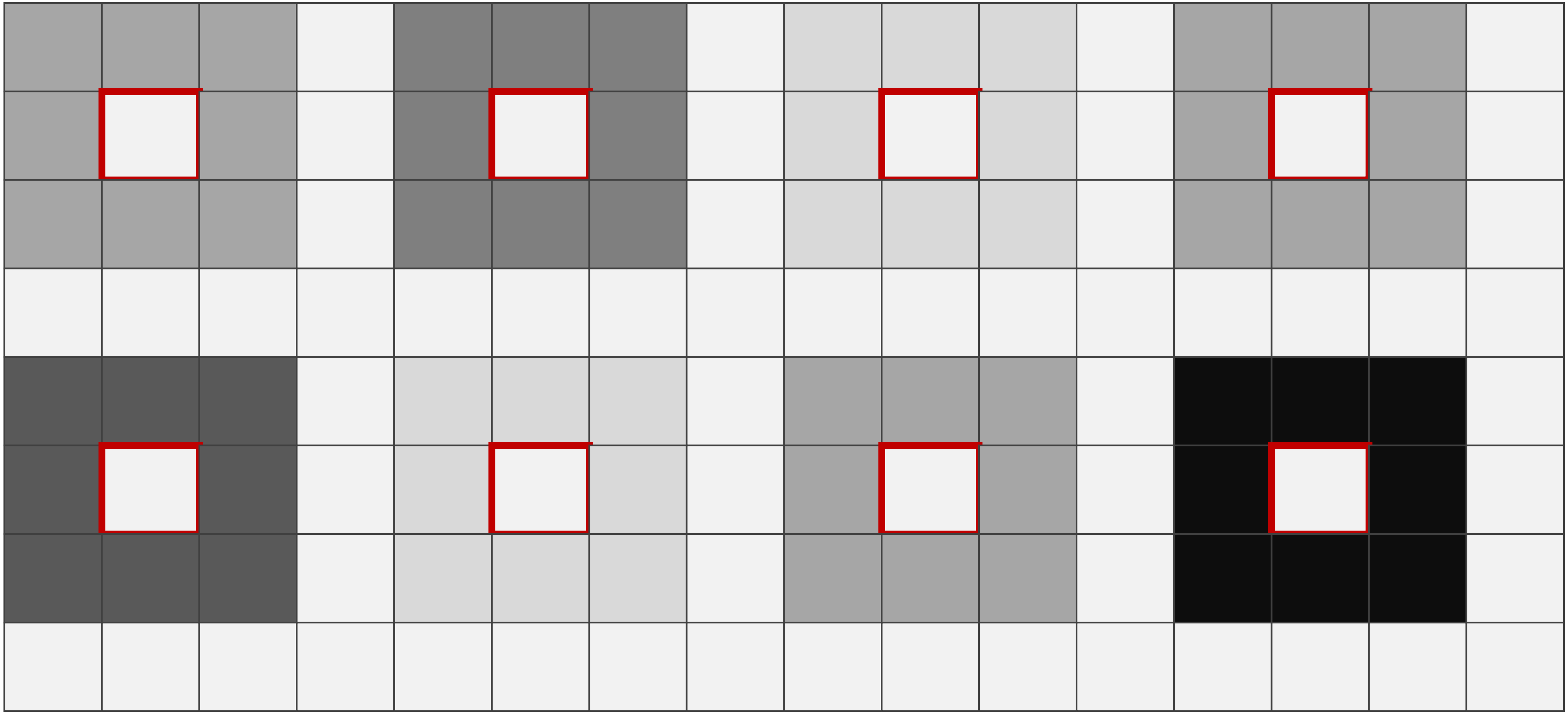
Haggard, E. A. (1958). Intraclass correlation and the analysis of variance



"all the variation is within classes [neighbors of red squares], with the result that there is no variation between class (i.e., each class sum equals [the same #])."

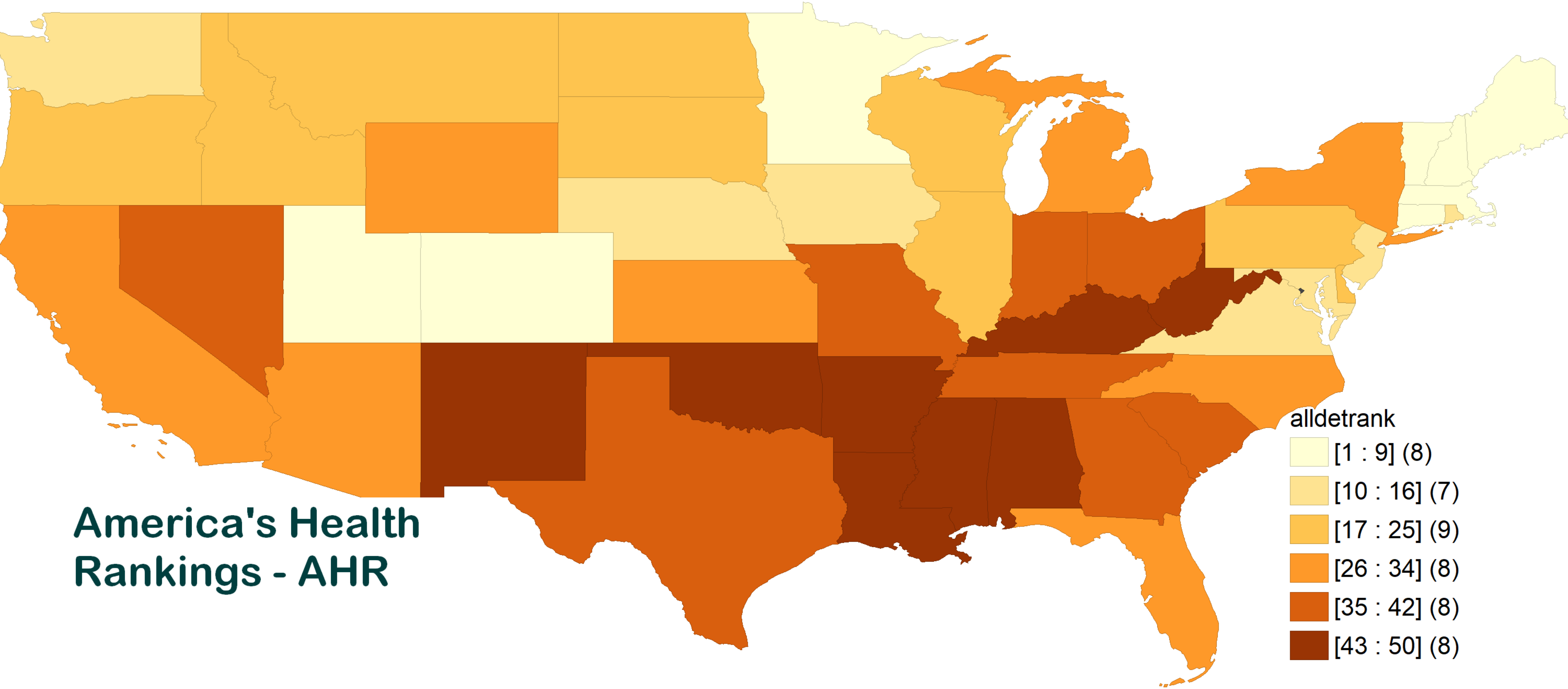
Intuition for Maximum Moran’s I

Haggard, E. A. (1958). Intraclass correlation and the analysis of variance



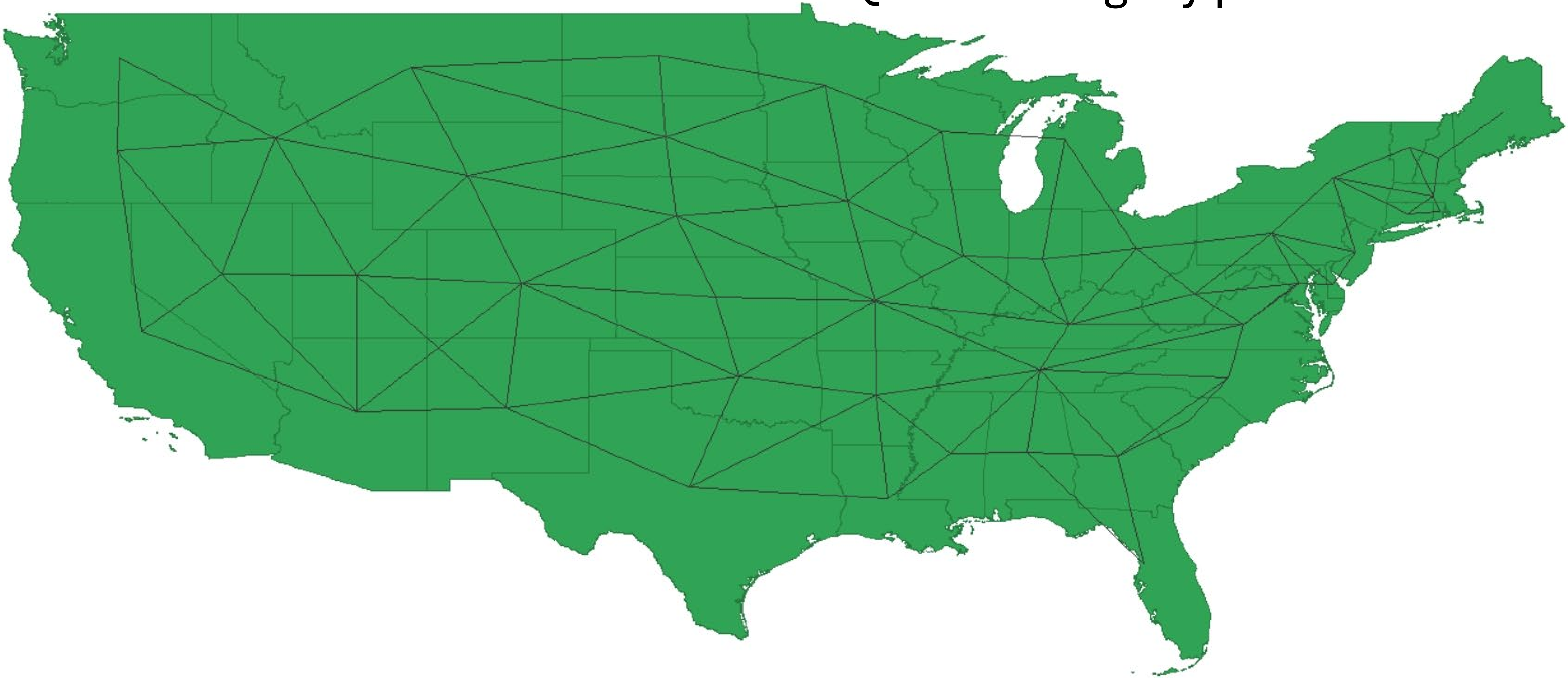
“there is no variation between the scores in any of the [classes [neighbors of red squares]; rather all the variation is between the [classes [the same #]).”

America's Health Rankings - AHR



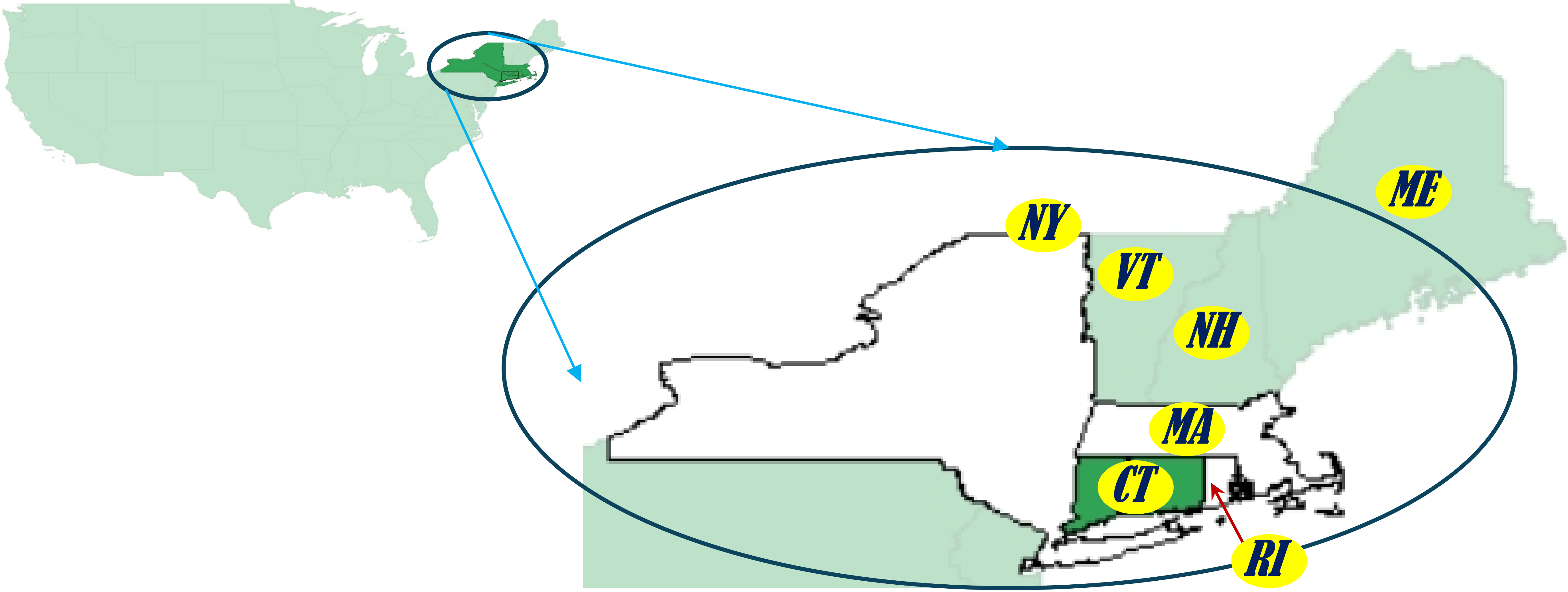
Networks spatial structure

States neighboring other states based on a Queen contiguity pattern



state2dig	AL	AR	AZ	CA	CO	CT	DC	DE	FL	GA	IA	ID	IL	IN	KS	KY	LA	MA	MD	ME	MI	MN	MO	MS	MT	NC	ND	NE	NH	NJ	NM	NV	NY	OH	OK	OR	PA	RI	SC	SD	TN	TX	UT	VA	VT	WA	WI	WV	WY
AL	0	0	0	0	0	0	0	0	0.25	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0
AR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0.17	0.17	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0.17	0.17	0	0	0	0	0	0	0
AZ	0	0	0	0.20	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0.20	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0
CA	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0
CO	0	0	0.14	0	0	0	0	0	0	0	0	0	0	0	0.14	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0	0	0.14	0	0	0	0.14	0	0	0	0	0	0	0	0.14	0	0	0	0	0	0.14
CT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0
DC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.50	0	0	0	0	0
DE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0
FL	0.50	0	0	0	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GA	0.20	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0.20	0	0	0	0	0	0	0	0
IA	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0.17	0.17	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0.17	0	0
ID	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0.17	0	0	0	0.17	0	0	0	0	0	0	0.17	0	0	0.17	0	0	0.17
IL	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0.20	0	0.20	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0
IN	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0.25	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KS	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0.25	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KY	0	0	0	0	0	0	0	0	0	0	0	0	0.14	0.14	0	0	0	0	0	0	0	0	0.14	0	0	0	0	0	0	0	0	0	0	0.14	0	0	0	0	0	0	0.14	0	0	0.14	0	0	0	0.14	0
LA	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0
MA	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0.20	0	0	0	0	0.20	0	0	0	0	0	0	0.20	0	0	0	0
MD	0	0	0	0	0	0	0.20	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0.20	0	0	0	0.20	0
ME	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MI	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0
MN	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0.25	0	0
MO	0	0.13	0	0	0	0	0	0	0	0	0.13	0	0.13	0	0.13	0.13	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0	0	0	0	0	0.13	0	0	0	0	0	0.13	0	0	0	0	0	0	0	0
MS	0.25	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0
MT	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0.25
NC	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0.25	0	0	0.25	0	0	0	0	0
ND	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0
NE	0	0	0	0	0.17	0	0	0	0	0	0.17	0	0	0	0.17	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0.17
NH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0
NJ	0	0	0	0	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0
NM	0	0	0.20	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0.20	0.20	0	0	0	0	0	0
NV	0	0	0.20	0.20	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0.20	0	0	0	0	0	0
NY	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0.20	0	0	0	0
OH	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0.20	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0	0	0	0	0	0	0	0	0	0	0.20	0
OK	0	0.17	0	0	0.17	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0
OR	0	0	0	0.25	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0
PA	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0.17	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0
RI	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SD	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0.17	0	0.17	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17
TN	0.13	0.13	0	0	0	0	0	0	0	0.13	0	0	0	0	0	0.13	0	0	0	0	0	0	0.13	0.13	0	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0	0	0	0
TX	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UT	0	0	0.17	0	0.17	0	0	0	0	0																									0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17
VA	0	0	0	0	0	0	0.17	0	0	0																									0	0	0	0	0	0	0.17	0	0	0	0	0	0	0.17	0
VT	0	0	0	0	0	0	0	0	0	0																									0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WA	0	0	0	0	0	0	0	0	0	0																									0	0.50	0	0	0	0	0	0	0	0	0	0	0	0	0
WI	0	0	0	0	0	0	0	0	0	0																									0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WV	0	0	0	0	0	0	0	0	0	0																									0	0	0.20	0	0	0	0	0	0	0.20	0	0	0	0	0
WY	0	0	0	0	0.17	0	0	0	0	0																									0	0	0	0	0	0.17	0	0	0.17	0	0	0	0	0	0

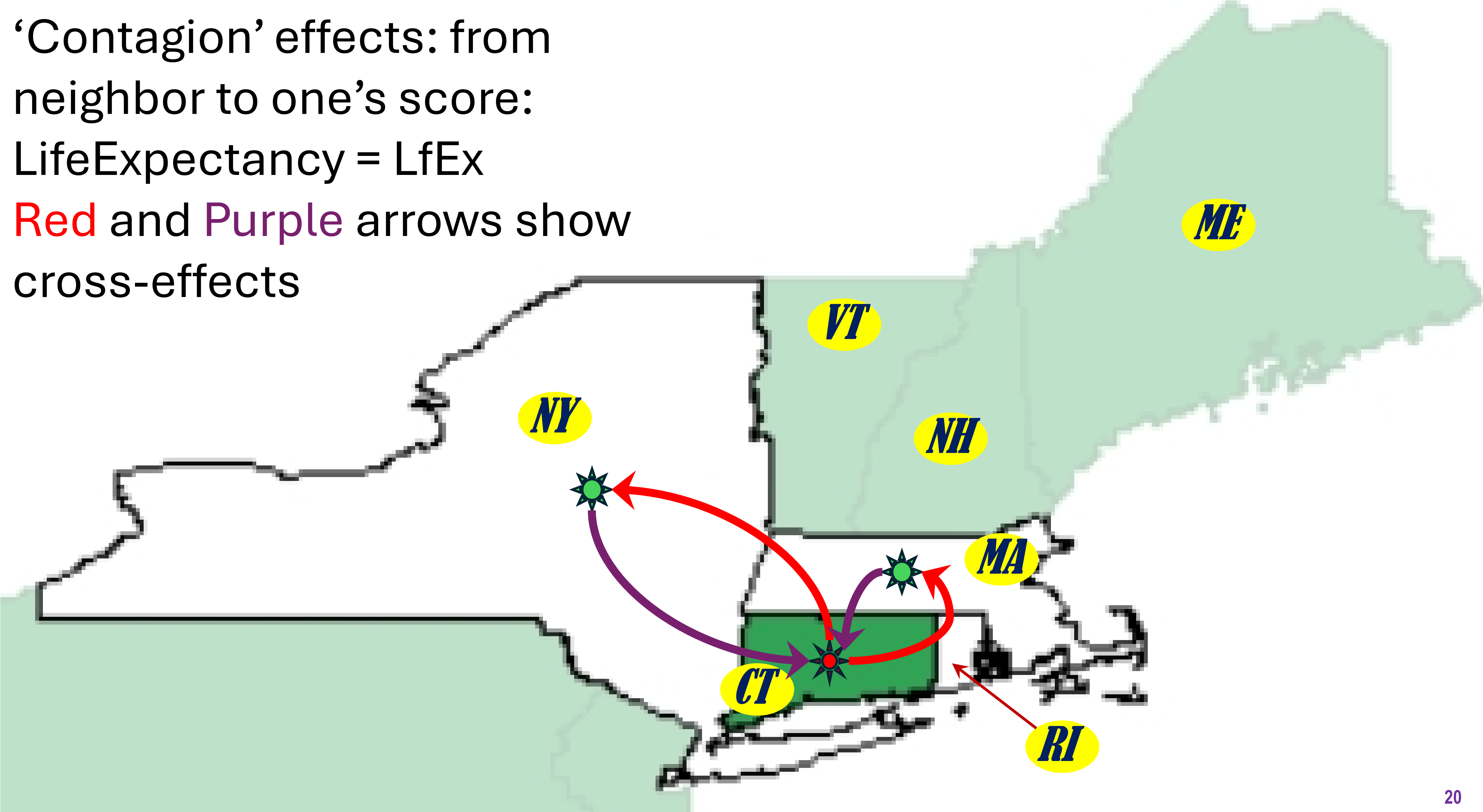
Queen contiguity type standardized Weight matrix



Lines from the actual *.gal weights file in GeoDa for CT:
CT 3
NY MA RI

‘Contagion’ effects: from neighbor to one’s score: LifeExpectancy = LfEx

Red and Purple arrows show cross-effects



From naïve/a-spatial to spatial regression

A classic regression $Y_i = \alpha. + \beta.\cdot X_i + \varepsilon_i$ would become for spatially connected/nonindependent data e.g.:

$Y_{CT} = \rho\cdot(1/3\cdot Y_{MA} + 1/3\cdot Y_{NY} + 1/3\cdot Y_{RI}) + \alpha. + \beta.\cdot X_{CT} + \varepsilon_{CT}$,
which says that MA, NY, and RI are neighbors of CT

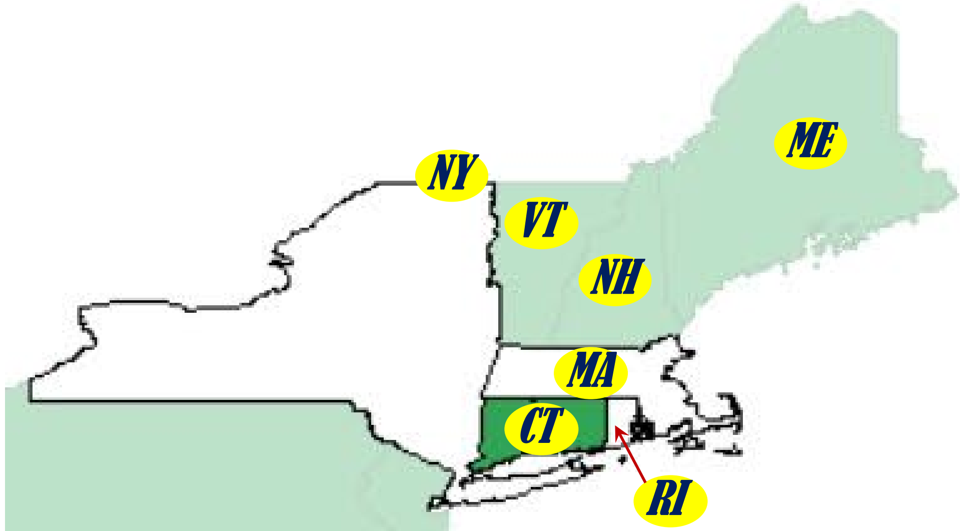
$Y_{ME} = \rho\cdot(1\cdot Y_{NH}) + \alpha. + \beta.\cdot X_{ME} + \varepsilon_{ME}$,
which says that only NH is a US state neighbor of ME

$Y_{MA} = \rho\cdot(1/5\cdot Y_{CT} + 1/5\cdot Y_{NY} + 1/5\cdot Y_{NH} + 1/5\cdot Y_{RI} + 1/5\cdot Y_{VT}) + \alpha. + \beta.\cdot X_{MA} + \varepsilon_{MA}$,
which says that CT, NY, NH, RI and VT and RI are neighbors of NY

$Y_{RI} = \rho\cdot(1/2\cdot Y_{CT} + 1/2\cdot Y_{MA}) + \alpha. + \beta.\cdot X_{RI} + \varepsilon_{RI}$,
which says that CT and MA are neighbors of RI

etc., 45 more times

Self & **Other**



Spelling out the ‘auto’-correlation meanings

“In essence, it is a cross-product statistic between a variable and its spatial lag, with the variable expressed in deviations from its mean.” GeoDa

$I_Y = \sum_i \sum_j [(W_{ij}\cdot(y_i - \bar{Y})\cdot(y_j - \bar{Y})]/\ S_0]/ [\sum_i (y_i - \bar{Y})^2/n]$

with w_{ij} as the elements of the spatial weights matrix, $S_0=\sum_i\sum_j w_{ij}$ as the sum of all the weights, and n as the number of observations. For the 49 contiguous US states, one then would get

$[(^{CT}[1/3\cdot(y_{CT} - \bar{Y})\cdot(y_{MA} - \bar{Y}) + 1/3\cdot(y_{CT} - \bar{Y})\cdot(y_{NY} - \bar{Y}) + 1/3\cdot(y_{CT} - \bar{Y})\cdot(y_{RI} - \bar{Y})] +$

$^{ME}[1/1\cdot(y_{ME} - \bar{Y})\cdot(y_{NH} - \bar{Y})] +$

$^{MA}[1/5\cdot(y_{MA} - \bar{Y})\cdot(y_{CT} - \bar{Y}) + 1/5\cdot(y_{MA} - \bar{Y})\cdot(y_{NY} - \bar{Y}) + 1/5\cdot(y_{MA} - \bar{Y})\cdot(y_{NH} - \bar{Y}) + 1/5\cdot(y_{MA} - \bar{Y})\cdot(y_{RI} - \bar{Y}) + 1/5\cdot(y_{MA} - \bar{Y})\cdot(y_{VT} - \bar{Y})] +$

$^{RI}[1/2\cdot(y_{RI} - \bar{Y})\cdot(y_{CT} - \bar{Y}) + 1/2\cdot(y_{RI} - \bar{Y})\cdot(y_{MA} - \bar{Y})] +$

$\ldots + {}^{CA}[(y_{CA} - \bar{Y})\cdot\ldots.\ ])/\ 49]\quad /$

$([(y_{AL} - \bar{Y})^2+(y_{AR} - \bar{Y})^2+\ldots + (y_{WV} - \bar{Y})^2+ (y_{WY} - \bar{Y})^2]/\ 49)$

(if we use the standardized weights, to sum up to 1 per case)

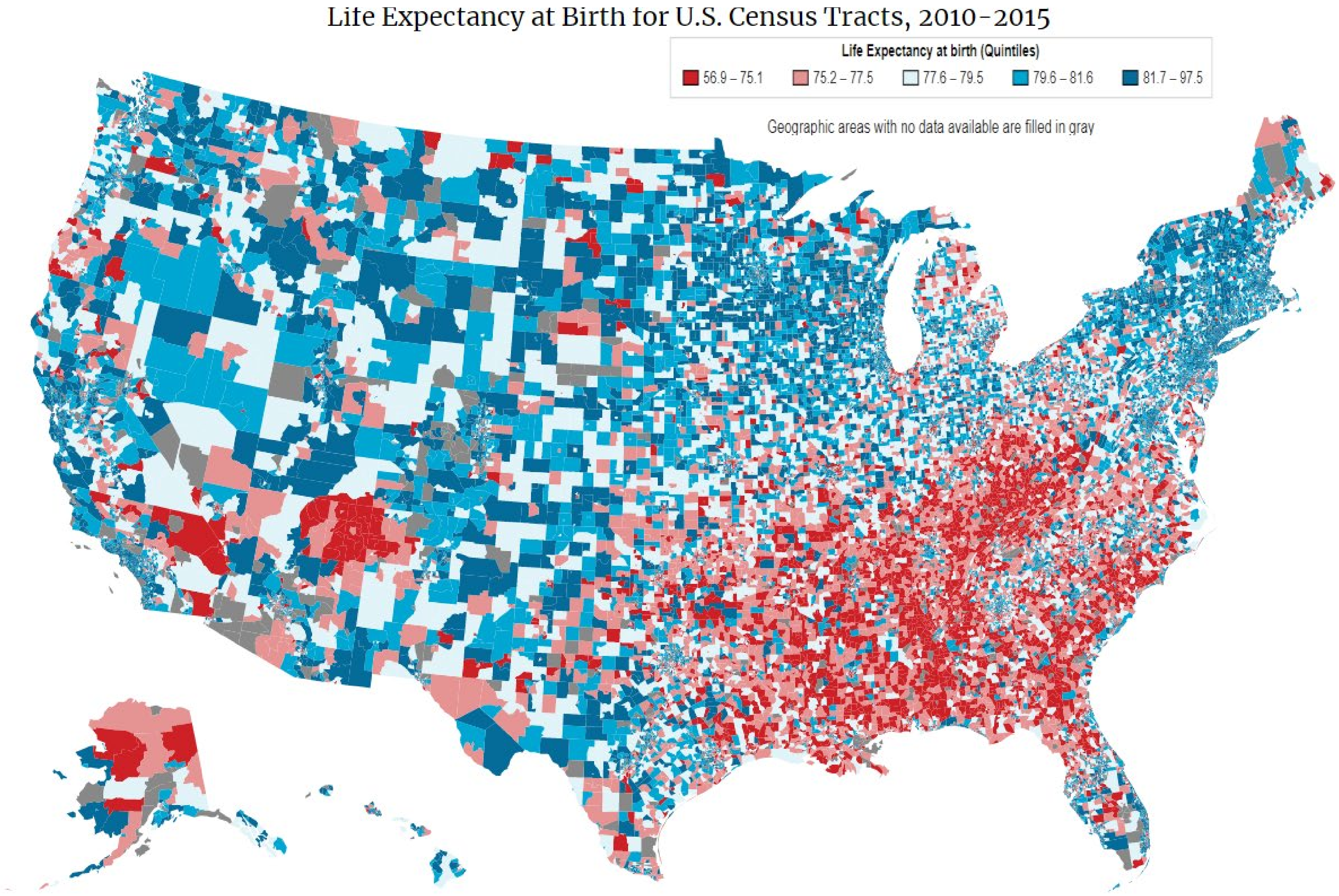
Note that Moran’s I applies to 1 variable + and some internal structure among cases (defined by a relationship matrix, who-with-whom, w_{ij}) whereas Pearson correlation applies to 2 variables, and is:

$\rho_{XY} = \sigma_{XY}/\ \sigma_X\cdot\ \sigma_Y = (E[(x_i - \bar{X})\cdot(y_i - \bar{Y})]\ /\ \mathrm{sqrt}[E(x_i - \bar{X})^2]\cdot\ \mathrm{sqrt}\ [E(y_i - \bar{Y})^2]\ =$

$[\sum_i(x_i - \bar{X})\cdot(y_i - \bar{Y})]/\ n\ ]/\ \mathrm{sqrt}([\sum_i(x_i - \bar{X})^2/n])\cdot\ \mathrm{sqrt}([\sum_i(y_i - \bar{Y})^2/n])$

Life Expectancy at Birth in the US by Census tracts (N = 67,148)

“US life expectancy dropped a full year in first half of 2020, according to CDC” Source: CNN


Life Expectancy at Birth for U.S. Census Tracts, 2010-2015
Life Expectancy at birth (Quintiles)
56.9 – 75.1
75.2 – 77.5
77.6 – 79.5
79.6 – 81.6
81.7 – 97.5
Geographic areas with no data available are filled in gray

Color code:
dark red = worst –
dark blue = best

Descriptives of the three main US regional variables, at census tract, county and state levels

	N	Mean	SD	Min	Max	Moran's I	I's p
%Minority$_{St}$	49	28.98	14.65	6.19	65.24	.43	.001
%Minority$_{Cnty}$	*3,087*	*22.42*	*19.34*	*0*	*99.28*	*.70*	*.001*
%Minority$_{CsTr}$	65,142	35.21	28.86	0	100	.70	.001
Life Expectancy$_{St}$	49	78.65	1.56	75.58	81.22	.54	.001
Life Expectancy$_{Cnty}$	*3,060*	*77.82*	*2.62*	*67.00*	*89.50*	*.56*	*.002*
Life Expectancy$_{CsTr}$	60,609	78.40	3.91	56.30	97.50	.41	.001
Income$_{St}$	49	32.24	5.48	23.55	53.32	.39	.002
Income$_{Cnty}$	*3,086*	*27.01*	*6.45*	*10.93*	*72.83*	*.56*	*.001*
Income$_{CsTr}$	64,683	32.85	16.13	0.04	221.60	.65	.001

Notes: The US counties and census tracts come from the 49 contiguous states; income 2014-2018 expressed in US $1,000s; Life Expectancy 2010-2015; % n-white minority 2014-2018;

Zero-order naïve/a-spatial Pearson correlations and spatial lag standardized regression coefficients (*direction* of effect is from row to column above diagonal) among state-level spatial variables

	%Minority$_{St}$	Income$_{St}$		Life Expectancy$_{St}$	
%Minority$_{St}$	*214.55*	.333 $_{\rightarrow}\uparrow$	*.534* $_{\leftarrow}\downarrow$	.101 $_{\rightarrow}\uparrow$	*.133* $_{\leftarrow}\downarrow$
p	---	.001	*.001* $_{\leftarrow}\downarrow$	.307	*.210*
Income$_{St}$	.198$^{\text{Naive}}$	*29.98*	*NA*	.333 $_{\rightarrow}\uparrow$	*.270* $_{\leftarrow}\downarrow$
p	.157 $^{\text{L}}$	---		.054	*.024*
Life Expectancy$_{St}$	-.210$^{\text{Naive.L}}$	.510$^{\text{Naive}}$		*2.45*	
p	.133	<.001 $^{\text{L}}$		---	

Notes: N = 49 (contiguous US states); spatial lag standardized regression coefficients above the diagonal are directional, the first value from row->column $_{\rightarrow}\uparrow$, the second value from column -> row $_{\leftarrow}\downarrow$; variances in diagonal (in *italics*); the standardized regression coefficients below diagonal are symmetric (hence 'same'); NA = non applicable: the variance is a same-variable (hence symmetric) parameter; $^{\text{L}}$: marks large discrepancies between the naïve/a-spatial and proper spatial estimates.

Unstandardized regression/path coefficients for the Life Expectancy regression on % non-White and income, at census tract, county and state levels, from naïve/a-spatial and proper/spatial models

	Census tracts			***Counties***			**States**		
LifeExp. predictor	Naïve	***Spatial***	%Δ	Naïve	***Spatial***	%Δ	Naïve	*Spatial*	%Δ
^{1}%nW	-0.312	***-0.224***	**0.088**	-0.224	***-0.182***	**0.042**	-0.224	0.108^{NS}	**0.332**
^{1}Inc	0.141	***0.116***	**0.026**	0.231	***0.206***	**0.024**	0.146	0.055	**0.091**
^{2}%nW	-0.064	***-0.051***	**0.013**	-0.078	***-0.068***	**0.010**	-0.346	0.020^{NS}	**0.366**
^{2}Inc	0.137	***0.113***	**0.025**	0.226	***0.203***	**0.023**	0.164	0.053^{NS}	**0.111**
$^{M.Tot}$%nW	-0.312	***-0.121***	**0.191**	-0.226	***-0.153***	**0.072**	-0.224	0.074^{NS}	**0.298**
$^{M.Dir}$%nW	-0.064	***-0.051***	**0.013**	-0.078	***-0.068***	**0.010**	-0.346	0.020^{NS}	**0.366**
$^{M.Indir}$%nW	-0.248	***-0.071***	**0.178**	-0.148	***-0.086***	**0.062**	0.122	0.054^{NS}	**0.067**
$^{M.b}$ Inc	0.137	***0.113***	**0.025**	0.226	***0.203***	**0.023**	0.164	0.053^{NS}	**0.111**
$^{M.a}$ *%nW→Inc*	-0.181	**-0.063**	**0.118**	-0.065	**-0.042**	**0.023**	0.074	0.104	**0.029**

Notes: ^{1} : Single predictor; ^{2}: Both predictors; M: %nW ->Inc->mediation models, Tot = total, Dir = direct, Indir = indirect effects, b is the Mediator → Outcome effect (when interaction is also included,); Unstandardized and standardized regression coefficients from CFAs and from naïve/a-spatial and proper/spatial models; unstandardized coefficients represent Life Expectancy years differences for 10% points difference in % non-White (p values for the unstandardized loadings not reported, as population data is analyzed; all coefficients were <.001, except for all states-level - ^{NS} statistically non-significant (except maybe A: p = .054); is absolute inflation of naïve estimates compared to proper spatial estimates; census tracts, counties, and states estimates of Life Expectancy outcome were inflated on average by 0.8, 0.4, and 2.6 months, respectively.

Research Questions answers

RQ 1: Do residents from places with more racial/ethnic minorities live shorter/longer lifes? By how much?

* **Yes**, at Census tract and County level:

i. Census tracts with 10% more non-White residents live **2.7 months** shorter lives.

ii. Counties with 10% more non-White residents live **2.2 months** shorter lives.

iii. There are no such differences seen across US states.

Standardized loadings from the four SVI (social vulnerability index) indicators of the Socioeconomic Status dimension (SVI-SES), from naïve/a-spatial and proper/spatial confirmatory factor analyses (CFA)

	Census tracts		Counties		US States	
SVI1 Indicator item$^{\text{Model}}$	λ's	% Expl.	λ's	% Expl.	λ's	% Expl.
% Poverty$^{\text{Naïve}}$	0.916	84%	0.850	72%	1.00	100%
% Poverty$^{\text{Spatial}}$	*0.523*	27%	*0.744*	55%	*0.84*	71%
% Unemployment$^{\text{Naïve}}$	0.678	46%	0.625	39%	0.32	10%
% Unemployment$^{\text{Spatial}}$	*0.393*	15%	*0.320*	10%	*0.35*	13%
Income $1,000s$^{\text{Naïve}}$	-0.780	61%	-0.718	52%	-0.75	56%
*Income $1,000s*$^{\text{Spatial}}$	*-0.585*	34%	*-0.311*	10%	*-0.62*	38%
% No high school$^{\text{Naïve}}$	0.709	50%	0.729	53%	0.84	70%
% No high school$^{\text{Spatial}}$	*0.435*	19%	*0.314*	10%	*0.44*	19%

Notes: Standardized loadings come from naïve CFAs and from spatial CFAs with added spatial lags behind each indicator of SVI-SES (all p values for the unstandardized loadings were <.001), and percent variance explained by the SVI-SES latent/common factor; the states level loadings come from CFA with n = 49, yet covering the entire population.

Research Questions answers

RQ 2: Is “Socioeconomic Status” a stable construct across spatial levels?
No: The indicators indicate SES stronger/weaker by levels.
i. Unemployment drops as an item at county and state levels.
ii. Income too drops as an item at county level, and only 33% and 38% of its variability is explained by the latent SES at census tract and state levels.

Conclusions

Effects depend on:

1. The level at which data and analysis are gotten/done

2. Analysis: naïve vs. spatial:
   Spatial models are many available.