

# spatial statistics with geoda

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Tuesday 28<sup>th</sup> November, 2023 14:46

## outline

spatial statistics intuition [wordy/lengthy: no time for this;  
do quick version posted on syllabus instead]

spatial weights

using spatial weights

K-means, medians etc DEFINITELY DO NEXT TIME



## reference

◇ <https://geodacenter.github.io>

◇ there are tutorials and data for practice:

- <https://geodacenter.github.io/documentation.html>
- <https://geodacenter.github.io/data-and-lab/>

## outline

spatial statistics intuition [wordy/lengthy: no time for this; do quick version posted on syllabus instead]

spatial weights

using spatial weights

K-means, medians etc DEFINITELY DO NEXT TIME

## correlation

◇ everyone heard of correlation, right? what is it?  
examples?

◇ many things correlate positively; people in space, too

◇ fat people like fat people; smokers like smokers, etc

◇ in short people you spend time with, are like you...

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3343772/pdf/nihms-216230.pdf> p11 cool vis

- <http://kelsocartography.com/blog/wp-content/uploads/2008/05/gr2008052600099.gif>

- <https://www.google.com/search?q=christakis+fowler+obesity>

- see ted talks by n christakis or j fowler

## same about anything in space

- ◇ <http://www.thebigsort.com/maps.php>
- ◇ hi-crime neighborhoods next to hi-crime neighborhoods
- ◇ poor blocks next to poor blocks
- ◇ even poor states are next to poor states (Mississippi, Alabama, etc)
- ◇ poor countries cluster together, too: Africa, Latin America, etc
- ◇ in short, things/areas that are close to each other in space are alike

# the first law of geography (Waldo Tobler)

- ◇ “Everything is related to everything else,
- ◇ but near things are more related than distant things”
- ◇ keep this in mind! almost always true!
  - do you see this in your research?



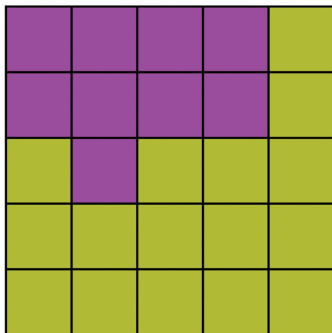
## positive v negative spatial autocorrelation

- ◇ note: autocorrelation
- ◇ correlate values of a var with values of the same var
- ◇ how?
- ◇ we spatially lag a variable (details in next section)
  - and we correlate value of that variable with
  - average value of the same variable in nearby polygons
- ◇ positive if similar values next to each other
- ◇ negative if dissimilar values next to each other
  - details in next section, but can already see it in plain thematic maps

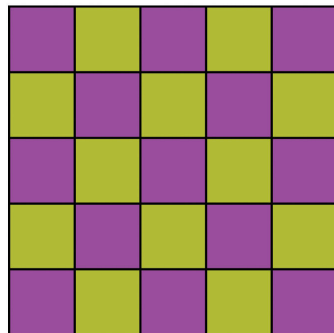
# pos and neg



POSITIVE : Pattern of Similarity



NEGATIVE : Pattern of Dissimilarity



## negative correlation is even more interesting

- ◇ less common than positive correlation: more interesting
- ◇ usually anything less common is more interesting
- ◇ if a dog bites a man its no news, if a man bites dog, that's news
- ◇ eg sometimes you will see rich area in the middle of poverty
- ◇ etc

## application: my paper about happiness in Europe

- ◇ <https://theaok.github.io/docs/gesis3.pdf>
  - see histogram and maps
- ◇ positive spatial autocorrelation
- ◇ clusters of happy and unhappy provinces
  - and they span across country boundaries
  - it is interesting to identify them and formally test it

## just a thematic map

- ◇ you'll already see or at least sense
  - spatial correlation from regular thematic maps
- ◇ just have a close look, and think about it
  - discuss in ps
- ◇ and now we'll use geoda to formally test if there is correlation

## outline

spatial statistics intuition [wordy/lengthy: no time for this; do quick version posted on syllabus instead]

spatial weights

using spatial weights

K-means, medians etc DEFINITELY DO NEXT TIME

## the first step

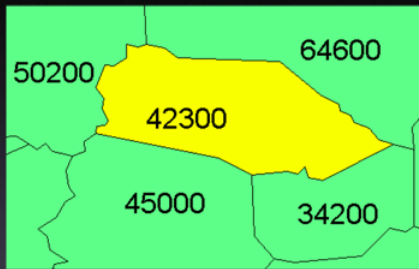
- ◇ the first step before producing spatial corr
- ◇ is to produce spatial weights
- ◇ or spatially lag a variable

## we will spatially lag a variable

- ◇ it's like time lagging a variable **draw a var and its lag**
  - time lagging is useful in exploring temporal precedence
  - eg you may want to know what is the corr/effect of unemployment last year on this year's poverty
- ◇ spatially lagged var: want to know the relationship of
- ◇ a place to its neighbors
- ◇ spatially lagged variable is just
  - an average of values for its neighbors
- ◇ for elaboration see geoda documentation



# Spatial Lag Example



Average Neighbor Land Values

$$1/4 \times 50200 + 1/4 \times 45000 + 1/4 \times 34200 + 1/4 \times 64600$$

Copyright © 2004–2007 by Luc Anselin, All Rights Reserved

## 2 ways

- ◇ contiguity based (we'll just do these):
  - neighbor of place A touches on place A
- ◇ distance based: neighbor of place A is within some distance of place A

## 2 types of contiguity weights

- ◇ usually just pick queen contiguity—neighbor is any place that neighbors our place
  - at least must share a vertex, say North, North-East, etc
- ◇ can do rook: must share a border, not just vertex
  - so *\*not\** North-East

## rook v queen



- ◇ Rook: only 2,4,6, 8; Queen: all (i.e. 1-8)

## order of contiguity

- ◇ can choose higher orders
- ◇ i.e. neighbors of my neighbors are my neighbors...
- ◇ we'll just stick with 1st order

## outline

spatial statistics intuition [wordy/lengthy: no time for this; do quick version posted on syllabus instead]

spatial weights

using spatial weights

K-means, medians etc DEFINITELY DO NEXT TIME

## reference

◇ again, see geoda workbook's appropriate chapter

`https://geodacenter.github.io/documentation.html`

# Moran's I

- ◇ it's a basic spatial statistic
- ◇ just like regular correlation (from -1 to 1)



# application

- ◇ <https://theaok.github.io/docs/gesis3.pdf>
- see Moran's I scatterplot

## so what?

- ◇ Moran's I and LISA help make sense of thematic maps
- ◇ they identify patterns, clusters, outliers
- ◇ very useful!
- ◇ eg is poverty concentrated?
- ◇ I would be really happy if I see them in final project
- ◇ likewise, histograms are very nice for paper
- ◇ and histogram for your key variable is necessary
  - (don't forget about interpretation!)
  - (don't ever show anything that you don't interpret)

## so what?

- ◇ and it does matter where in the cluster one is located
- ◇ eg being poor in the middle of poverty may be better
- ◇ than being poor next to rich
- ◇ suicide among females in rural china:
  - not absolute but relative deprivation

## we're doing space, but think about time, too

- ◇ not only focus on location of greatest poverty, crime etc
- ◇ over-time changes matter, too
- ◇ greatest or smallest increase
- ◇ largest change from well-established trend
- ◇ trend
- ◇ etc
- ◇ show 2 maps, say 1950 map next to 2000 map
- ◇ or calculate new var  $(2000-1950)/1950$

## outline

spatial statistics intuition [wordy/lengthy: no time for this;  
do quick version posted on syllabus instead]

spatial weights

using spatial weights

K-means, medians etc DEFINITELY DO NEXT TIME