spatial statistics with geoda

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Wednesday 11th October, 2023 17:52

outline

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

why spatial statistics?

- sounds scary...there is word 'statistics'
  - but we'll only do maps and graphs
  - no formulas, no calculations–relax!
- all we will do is just correlation in space
- so called spatial autocorrelation
- and formally calculated with Moran's I
  - or Local Moran's I (LISA)

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...
  - `http://nicholaschristakis.net/wp-content/uploads/2015/03/Spread-of-Alcohol-Consumption-Behavior-in-a-Large-Soci pdf`
  - (last page)
  - `http://kelsocartography.com/blog/wp-content/uploads/2008/05/gr2008052600099.gif`
  - `https://www.google.com/search?q=christakis+fowler+obesity&tbm=isch`

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

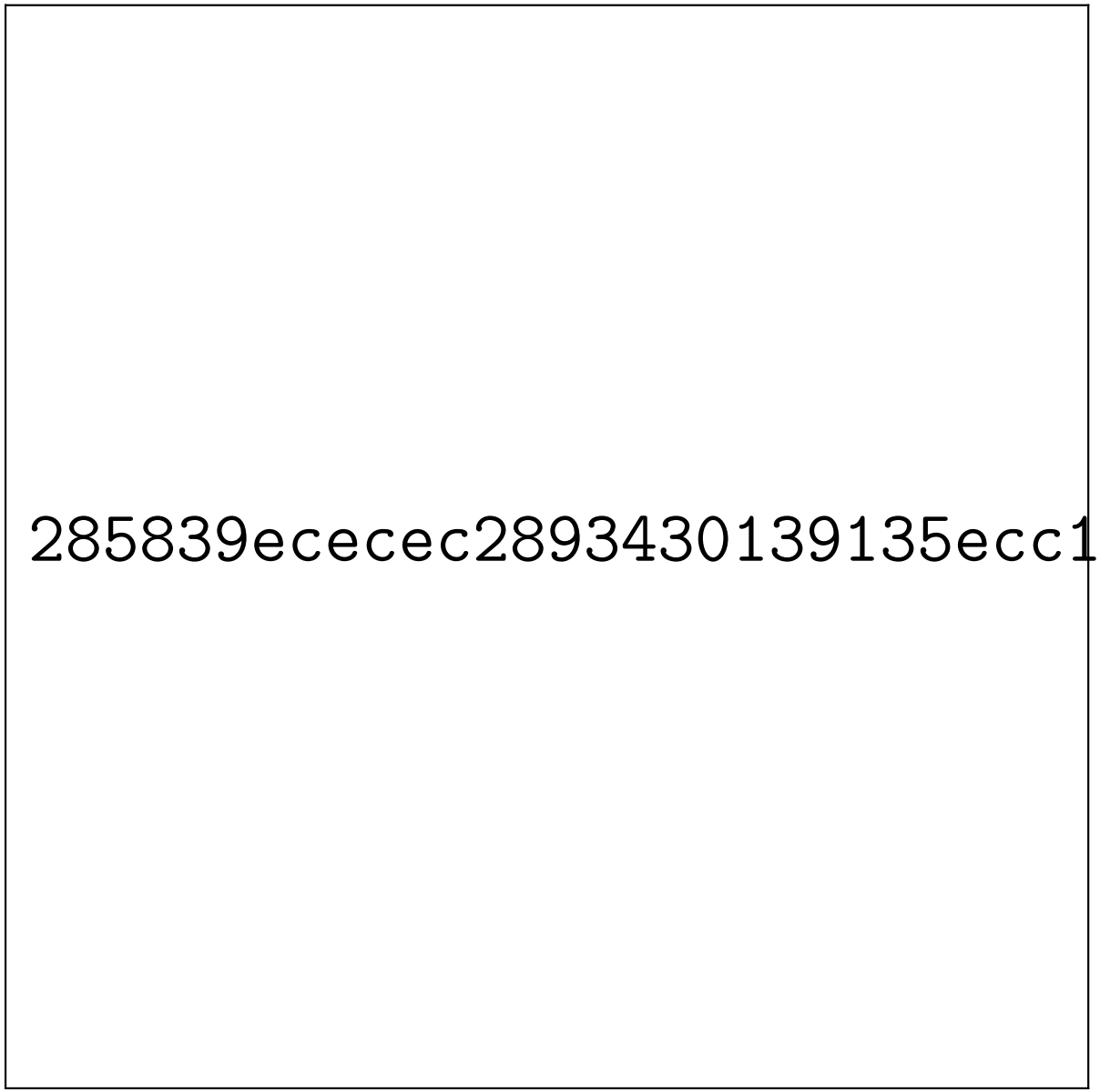
the first law of geography

- The first law of geography according to Waldo Tobler is:
- "Everything is related to everything else,
- but near things are more related than distant things"
- keep this in mind! it's 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



negative correlation is even more interesting

- less common than positive correlation: it's more interesting
- (usually anything less common is more interesting)
- eg sometimes you will see rich area in the middle of poverty
- etc

application: my paper about happiness in Europe

- `https://sites.google.com/site/adamokuliczkozaryn/pubs/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 ps6, paper
- and now we'll use geoda to formally test if there is correlation

outline

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 ex17 'spatially lagged vars' p124 of geoda workbook

`https://geodacenter.asu.edu/system/files/geodaworkbook.pdf`


KeynoteExport11823685777488888007001.pdf

slide_64.pdf

let's do it! create weights

- Weights File ID Variable: POLYID
  - usually fips or some unique ID/KEY var identifier of a place
  - (i think it must be numeric)
- and now the key part: defining neighbors
- who is a neighbor?

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

1476-072X-6-54-1.pdf

- 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

exploring weights

- 0 has 2 neighbors: 1,2
- 1 has 3 neighbors: 0,2,3
- and so on
- do not trust anybody
- let's look them up
- and confirm in map that indeed this is the case !

outline

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)
- Space-Univariate Moran's I: CRIME
- and it's .5 meaning that
  - there is a moderate positive spatial autocorr in CRIME
- we've expected that from thematic map
- note that y-axis is lagged crime
- select some obs and discuss: its and its nei crime
  - see in a map; select some other obs that is diff

Moran's I

- i can also rectangle select points in scatterplot
- let' select those in top-right (hi-hi): central city
- now bottom-left (lo-lo): outer areas
- now outlier in top-left (lo-hi: low crime but hi crime around)
- let's look back at thematic map–indeed that place is low crime
  - but its neighbors are high crime
- there isn't a clear outlier with hi-lo at bottom right

Moran's I: HOVAL

- how about housing value (HOVAL)
- make thematic map and Moran's scatter plot
- much less clear clustering, and few hi-lo, lo-hi
- highlight them in scatter and compare in thematic map

LISA

- LISA is a Local Moran's I
- Space-Univariate Local Moran's I: CRIME
  - and select all three maps
- it nicely identifies clusters
- again, compare with thematic map

application

- https://sites.google.com/site/adamokuliczkozaryn/pubs/gesis3.pdf
  - see Moran's I scatterplot
  - http://people.hmdc.harvard.edu/~akozaryn/myweb/papers/gesis/
  - see output from Geoda online

so what?

- Moran's I and LISA help make sense of thematic maps
- they identify patterns, clusters, outliers
- very useful !
- e.g. is poverty concentrated ? etc etc
- 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