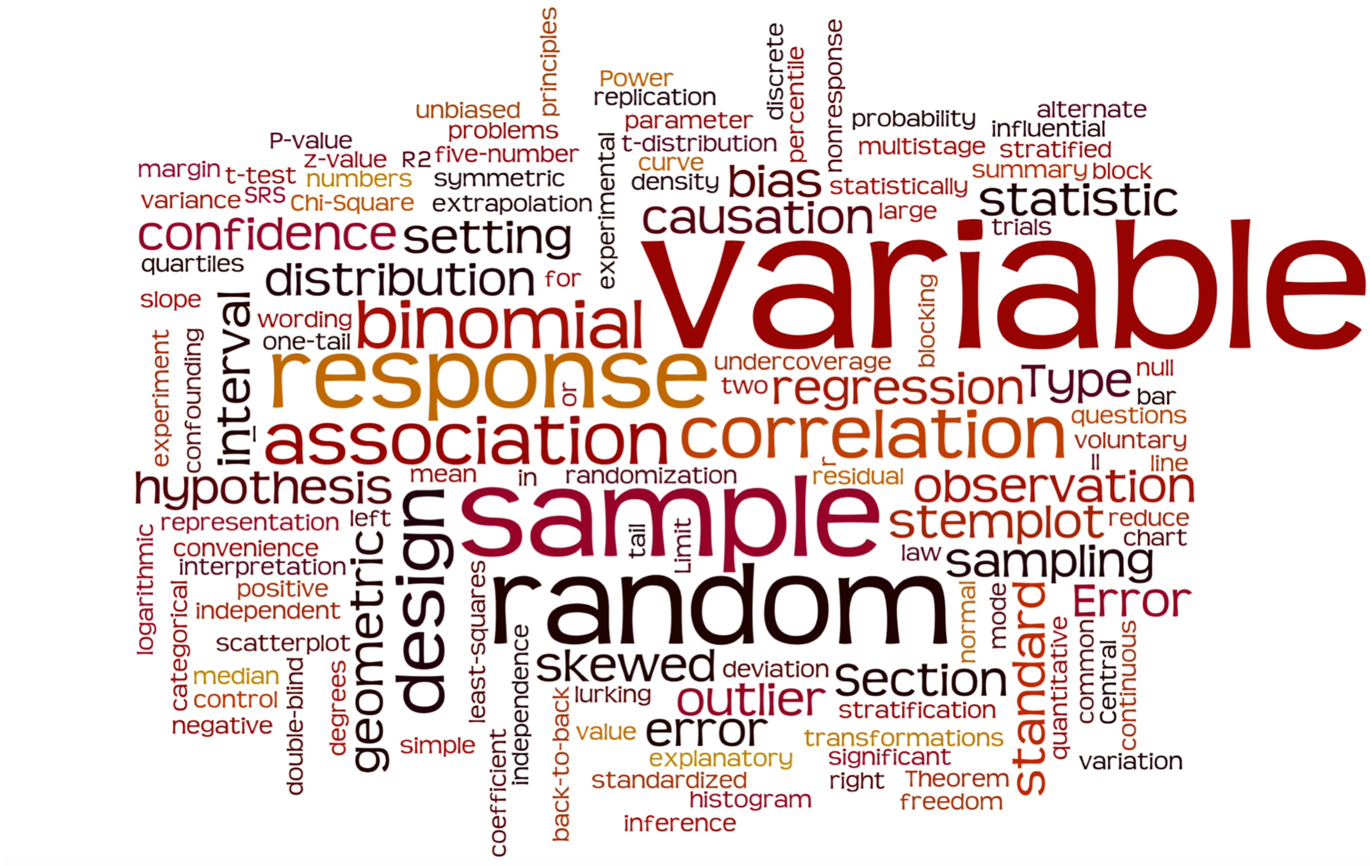
Introduction to Quantitative Methods



Kristin Eccles

Outline

- Properties of spatial data
- Autocorrelation
  - Global
  - Local
- Spatial aggregations
- Ecological fallacy
- MAUP
- Guest Lecture: Philippe Thomas, Environment and Climate Change Canada

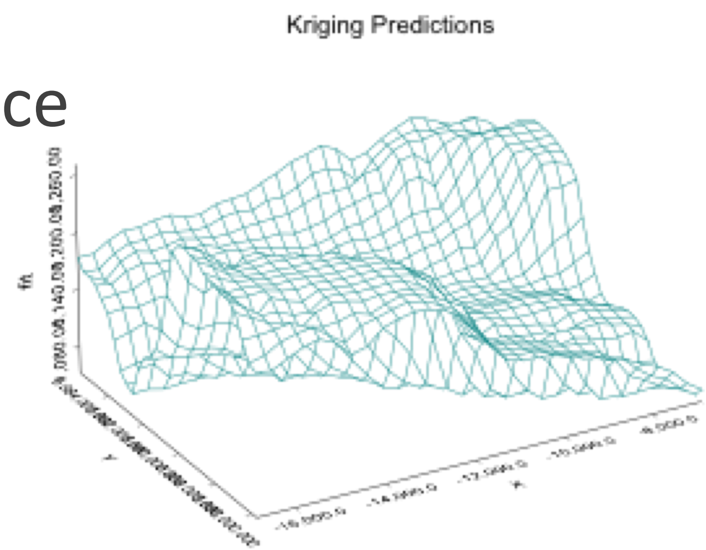
Unique Properties of Spatial Data

Spatial Dependence

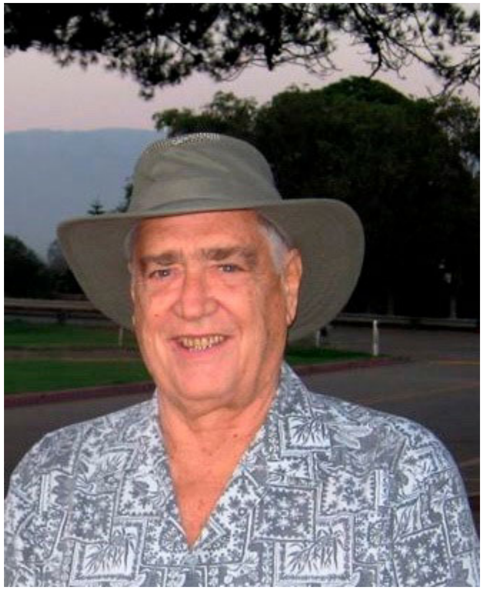
- Items that are closer are more similar than items that are further apart
  - Tobler's First Law of Geography
  - Spatial autocorrelation

Spatial Non-Stationarity

- Patterns change unevenly over space
  - Heterogeneous
  - Anisotropic



First Law of Geography

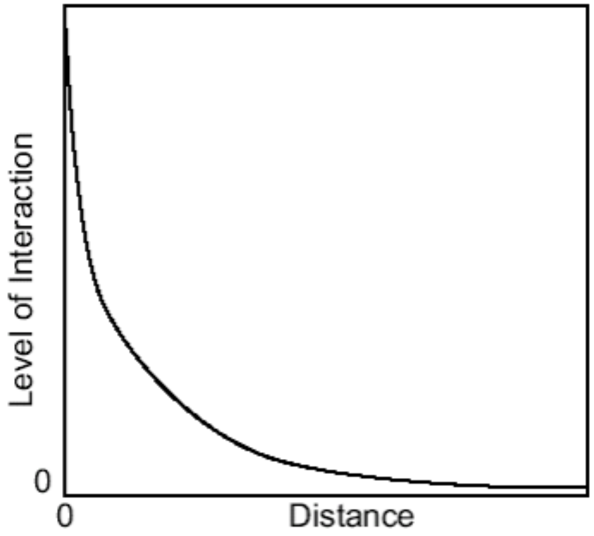


- Waldo Tobler (b. 1930; prof. emeritus of Geography and Statistics, U California Santa Barbara)
- Trained as a mathematical cartographer, and is best known for his contributions to that field.
- Coined the term analytical cartography to reflect the combination of mapping and analysis.

- Tobler's First Law of Geography:
  - "Everything is related to everything else, but near things are more related than distant things" (1970)
- Does this make sense?
- Climate, Topography, Pollution, Geomorphology, Natural disasters

Tobler’s Law: Limitations

- Tobler’s law implies a distance decay function
- However, only true to a **certain location** OR at a **certain scale**
  - Not everything in geography varies smoothly and continuously
  - Continuous and discrete representations
- Some geographic phenomena vary smoothly across space
- Others exhibit extreme irregularity
  - Violation of Tobler’s Law?
  - I.e. abruptness
- Can we quantify the Law’s limits?
  - Determining relatedness



Quantifying Spatial Autocorrelation:

- Spatial autocorrelation measures attempt to deal simultaneously with the location of spatial objects and their attributes

  - **Positive** spatial autocorrelation: features that are *closer* in location are *more similar* in attributes

  - **Negative** spatial autocorrelation: features that are *close* in location are *more dissimilar* in attributes

  - **Zero** spatial autocorrelation: attributes are *independent* of proximity of location

Measuring Spatial Autocorrelation: Moran's I

$$I = \frac{N \sum_i \sum_j w_{ij}(X_i - \overline{X})(X_j - \overline{X})}{\sum_i \sum_j w_{ij} \sum_i (X_i - \overline{X})^2}$$

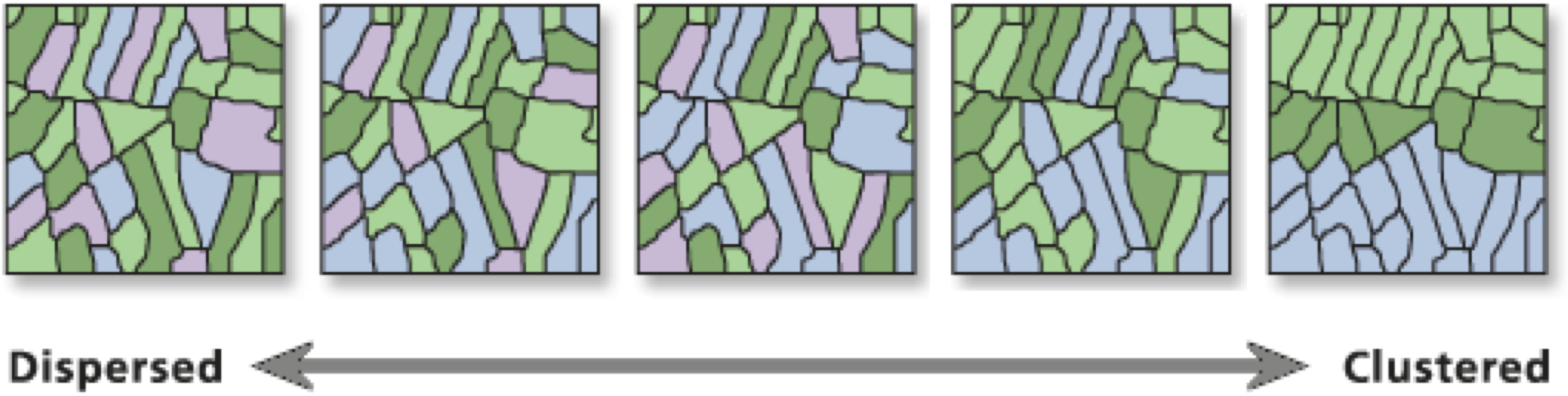
- Where N=number of spatial objects, indexed by *i* and *j*,
- *X* is the attribute of interest,
- $\bar{X}$ is the mean of *X*,
- w_{ij} is a matrix of spatial weights

Moran's I

Moran's I is a widely-used measure of global spatial autocorrelation, which ranges from -1 to +1

- Negative values indicate negative spatial autocorrelation
- Positive values indicate positive spatial autocorrelation
- Zero value indicates spatial independence (i.e. spatially random)

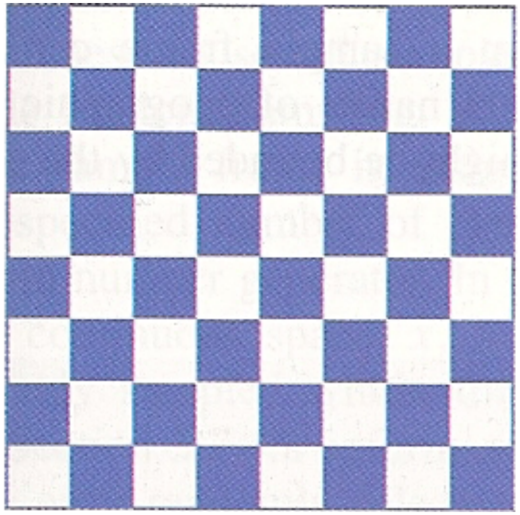
Source: ESRI



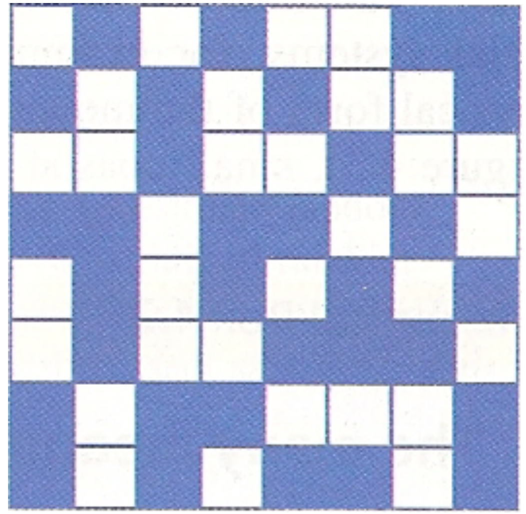
Perfect clustering = 1

Perfect dispersion =-1

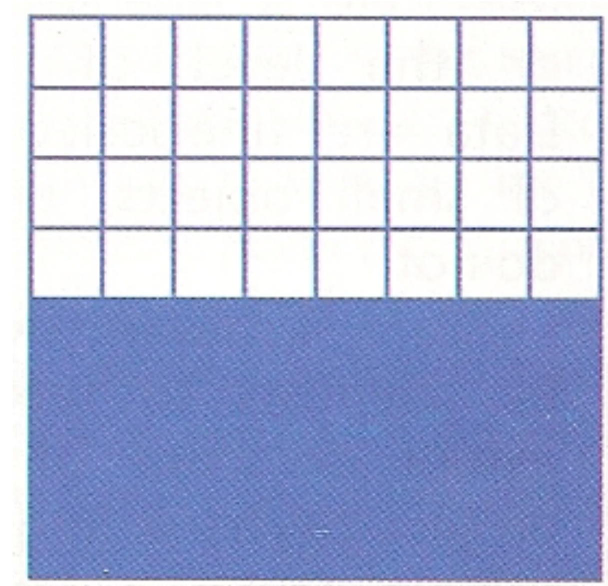
Nominal Patterns and Moran’s I



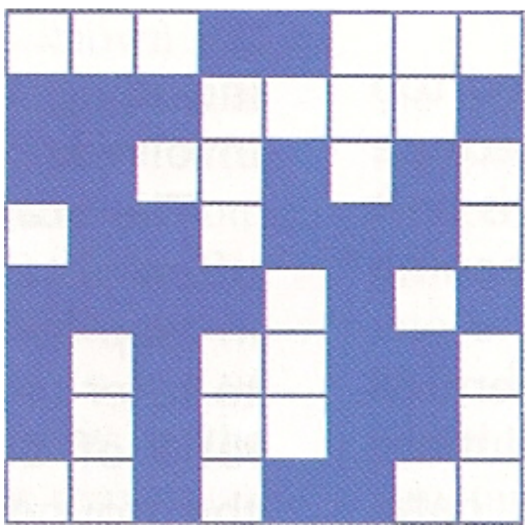
I = -1.00



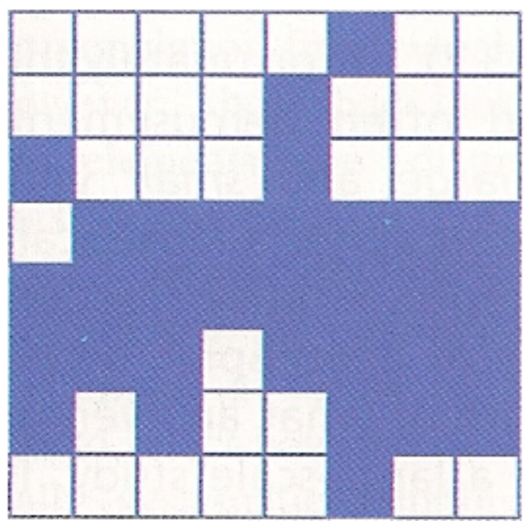
I = -0.393



I = 0.857



I = 0.000



I = 0.393

Correlogram

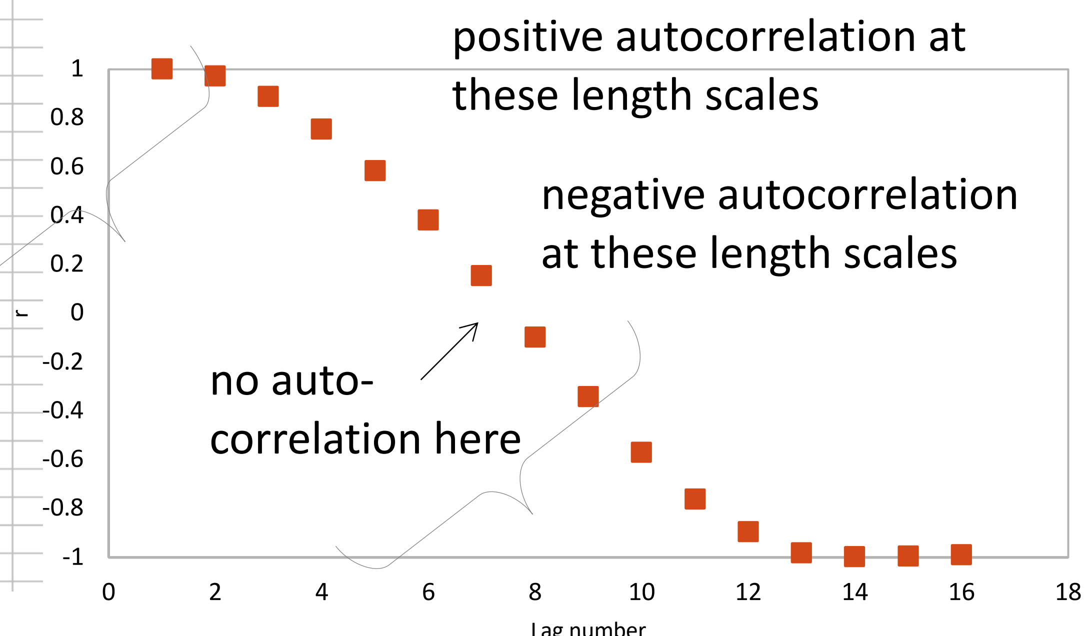
Correlate the observations against themselves (r = 1)

Then shift one set of observations down (a lag of 1)

Repeat with lag of 2 and so on

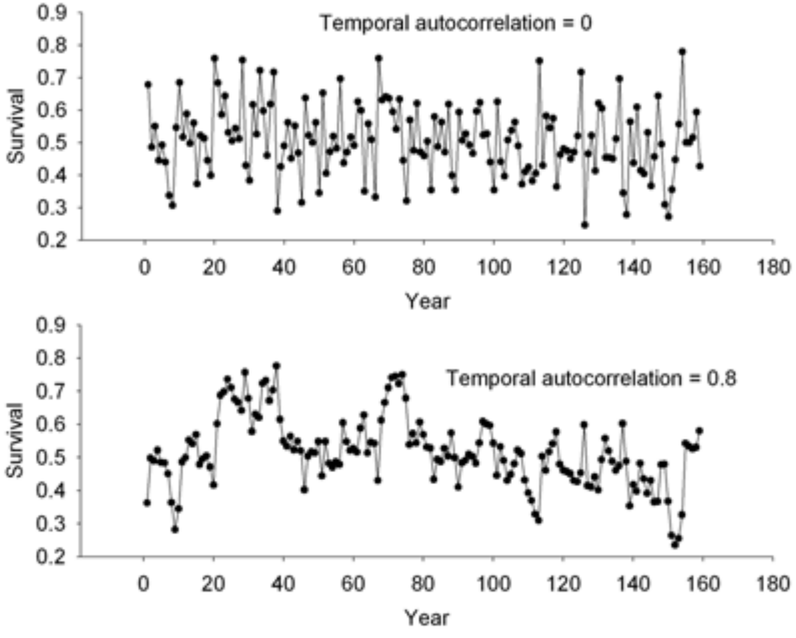
CORREL =CORREL(B$2:B$24,C3:C25)

	A	B	C	D	E	F	G
1	Distance (m)	Noise level (dB)	Noise level (dB)		Lag	r	
2	0	3	3		1	1	
3	100	4	4		2	=CORREL(B$2:B$24,C3:C25)	
4	200	5	5		3	0.886405989	
5	300	8	8		4	0.753542932	
6	400	10	10		5	0.581716947	
7	500	12	12		6	0.381718218	
8	600	15	15		7	0.151340664	
9	700	18	18		8	-0.099337417	
10	800	21	21		9	-0.34401352	
11	900	24	24		10	-0.570558277	
12	1000	29	29		11	-0.763387552	
13	1100	33	33		12	-0.896343485	
14	1200	40	40		13	-0.981937135	
15	1300	44	44		14	-0.998665577	
16	1400	42	42		15	-0.9954039	
17	1500	41	41		16	-0.992184796	
18	1600	38	38				
19	1700	36	36				
20	1800	33	33				
21	1900	29	29				
22	2000	25	25				
23	2100	22	22				
24	2200	19	19				
25	2300						
26							



Temporal autocorrelation

- Time series are prone to the same thing
- Is the temperature now statistically independent from the temperature 5 minutes ago?
- Diel cycles, seasonal cycles, etc. may add structure to data that violates the assumption of independence



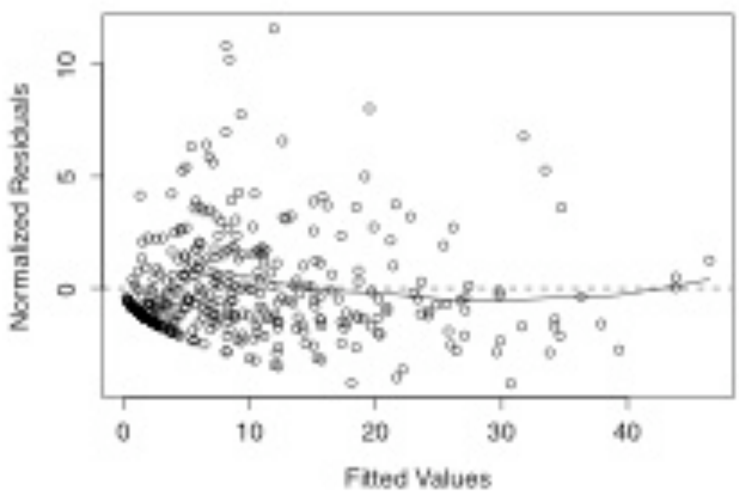
Test assumptions - revisited

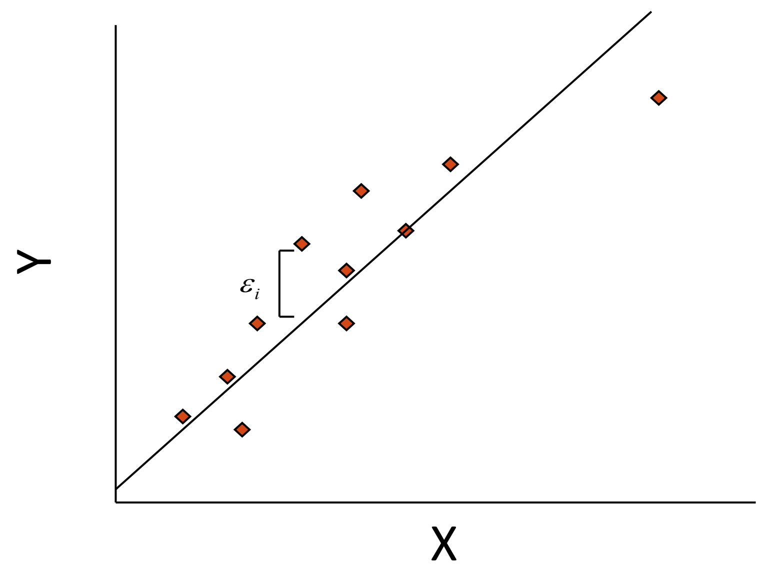
- Many [most] tests assume that samples are statistically independent
  - Hence the reason for probabilistic sampling schemes
  - Avoid bias, get a representative sample
- Spatial autocorrelation (positive or negative) will violate this assumption!
  - Residuals will show a pattern related to distance (not random)
- Need to use advanced spatial statistics to account for this
- The spatial structure could be interesting!

Global Indicators

- Global spatial autocorrelation and global statistics (i.e. regression) yields only one statistic to summarize the whole study area
- BUT a global analysis assumes homogeneity (spatial stationarity) over the study area
  - This is also an assumption of regression- homogeneity of residuals

$$Y_i = b_0 + b_1 X_{i1} + \varepsilon$$

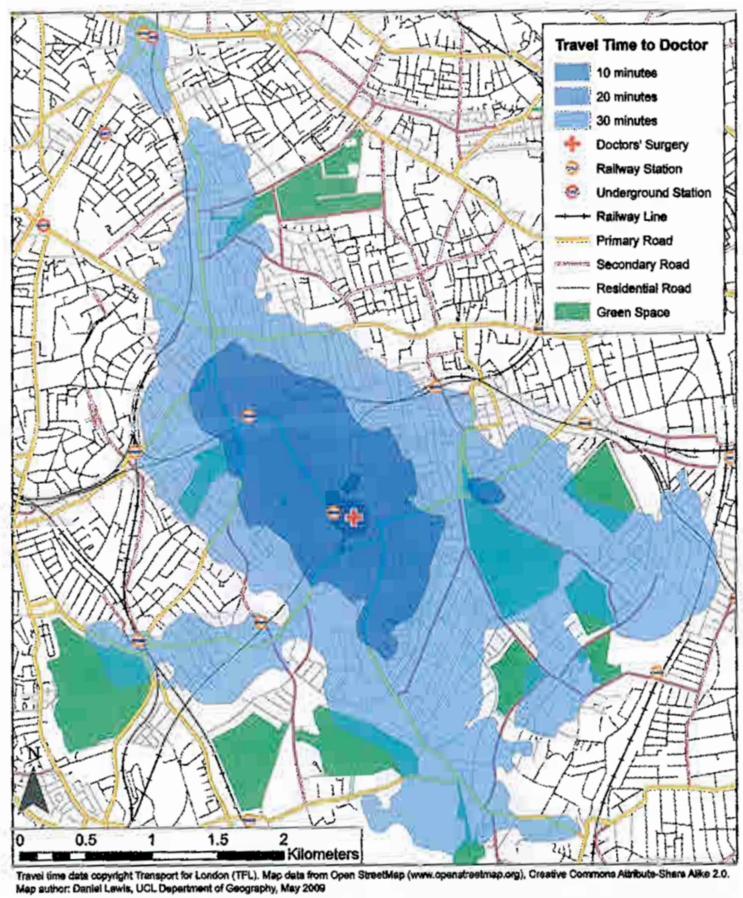




For more information: http://blog.liyunchen.com/variance-regression-clustering-residual-and-variance/

Clusters

- Tobler's Law implies that distance would impose a smooth, continuous decay effect on attribute values
- Spatial structure is not always smooth and continuous like Tobler suggests
  - Physical factors, infrastructure, administrative boundaries
  - Particularly a problem in human-built environments: we like hard borders!
- Quantifying clusters: Local Indicators of Spatial Association (LISA)
  - Getis and Ord's Gi*
    - A group of areal units with high Gi∗ indicates a 'hotspot' where as low Gi∗ means a 'coldspot'



Local Clustering Example

QUANTITATIVE RESEARCH

Applications of geographic information systems in public health: A geospatial approach to analyzing MMR immunization uptake in Alberta

Kristin M. Eccles, MSc, Stefania Bertazzon, PhD

ABSTRACT

OBJECTIVE: This study evaluates the temporal, spatial, and spatio-temporal variation of immunization rates for measles, mumps and rubella (MMR) immunization in the province of Alberta. The study uses yearly immunization rate data for Health Zones and Local Geographic Areas (2004–2012), which were obtained from Alberta Health's Interactive Health Data Application (IHDA).

METHODS: Spatial analyses include a global spatial analysis, Moran's I, and local indicators of spatial association (LISA) analysis – Getis and Ord's G* – to identify clusters of high or low immunization rates. Spatial methods are then applied to a time series analysis to examine how the immunization rates change over time in conjunction with space.

RESULTS: Mapped results indicate decreasing immunization rates over time for the majority of the province where most local geographic areas (LGAs) fall short of the 95% herd immunity threshold. Clusters of high immunization rates in the metropolitan centres, and clusters of low immunization rates in the southern and northern region of the province exist spatially and spatio-temporally. Over time, the high rate clusters are decreasing in size and the low rate clusters are increasing.

CONCLUSION: This research provides a localized geographic approach to assessing MMR immunization rates in Alberta. Findings from this research can be used to target public health interventions to specific areas that exhibit the lowest immunization rates. These results can also be used for hypothesis generation in future research on barriers to immunization uptake.

KEY WORDS: Immunizations; MMR; public health; GIS; spatial analysis

La traduction du résumé se trouve à la fin de l'article.

Can J Public Health 2015;106(6):e1–e19
doi: 10.17269/CJPH.106.4981

Immunization uptake for preventable illnesses has been a struggle for Canada in the recent past. In 2013, Alberta had 29 measles cases occur in individuals younger than 15 years, and 14 cases occur in individuals aged 15–24. These cases were primarily recorded in Lethbridge, a city with a population just over 90,000, located in the South Health Zone.[1] The source case was exposed to the measles virus from travel to the Netherlands, where a measles outbreak had begun in May 2013.[2] While the outbreak in the South Health Zone ended at the end of 2013, measles outbreaks were declared in the Edmonton, Calgary, and Central Health Zones in April to June 2014. Prior to this, Alberta's most recent outbreaks of an immunization-preventable disease were pertussis outbreaks in 2009 and 2012.[2]

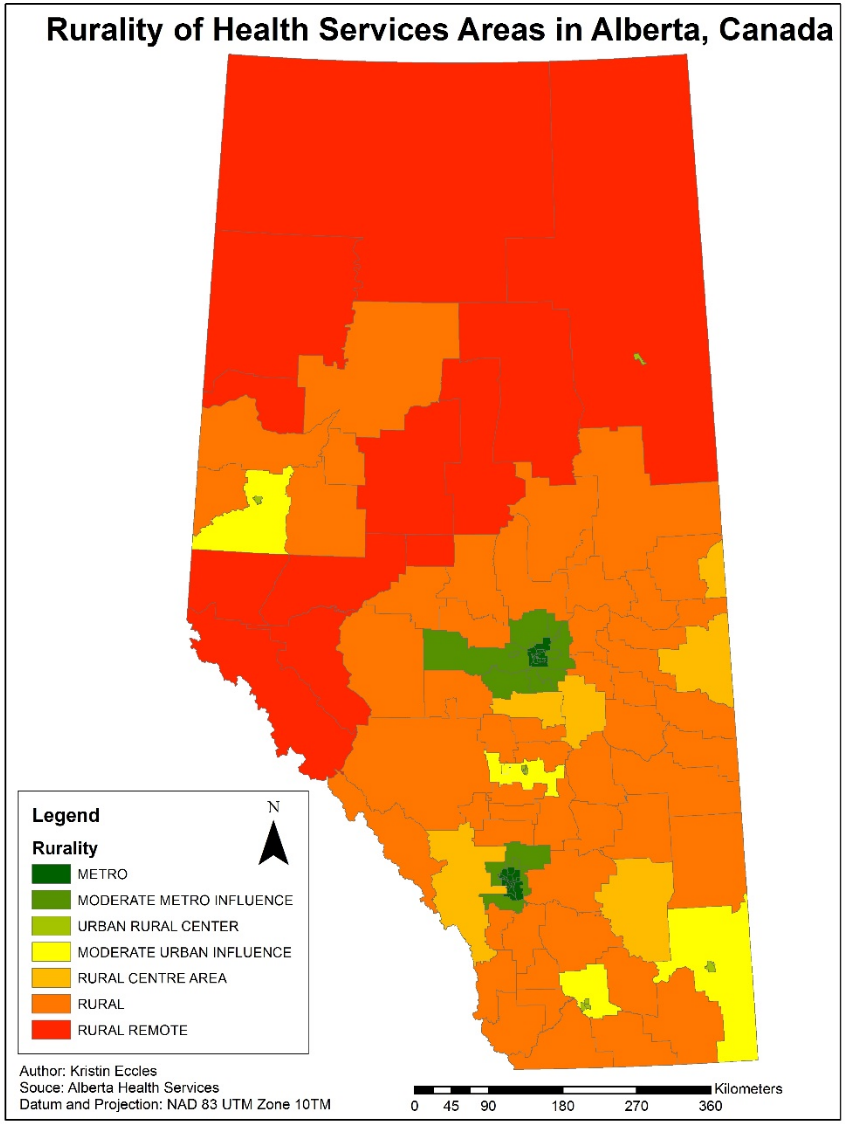
A study by the American Academy of Pediatrics reported that the measles, mumps and rubella (MMR) vaccine is the most frequently refused immunization.[3] Parents who have refused to have their children immunized have cited doing so due to religious, philosophical or personal beliefs. Additionally, some children are not able to receive the immunizations for medical reasons.[4] Recognizing that the entirety of a population may not be immunized or effectively protected by an immunization, herd immunity is a way to extend protection to the unimmunized and unprotected individuals based on the acquired immunity of a large proportion of the population.[5] However, threshold required to achieve herd immunity through MMR immunization is 95%.[6]

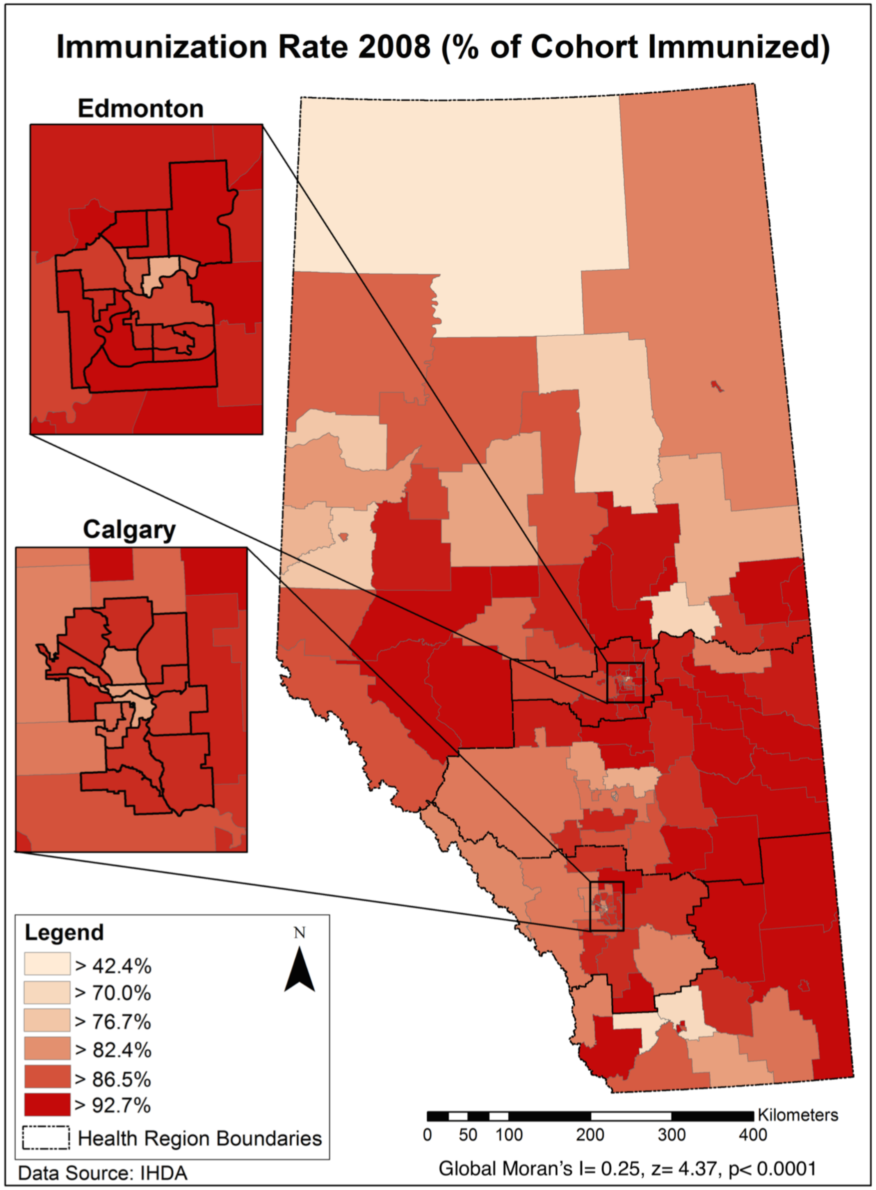
There is a strong geographical component to immunization rates, rates of refusal, and outbreaks. Studies have shown that individuals who choose not to have their children immunized tend to cluster together. These clusters of under-immunized populations also tend to overlap with outbreaks of immunization-preventable illnesses.[4] The relationship between unimmunized clusters and outbreaks was found statistically significant during a pertussis outbreak in Michigan between 1993 and 2004.[7]

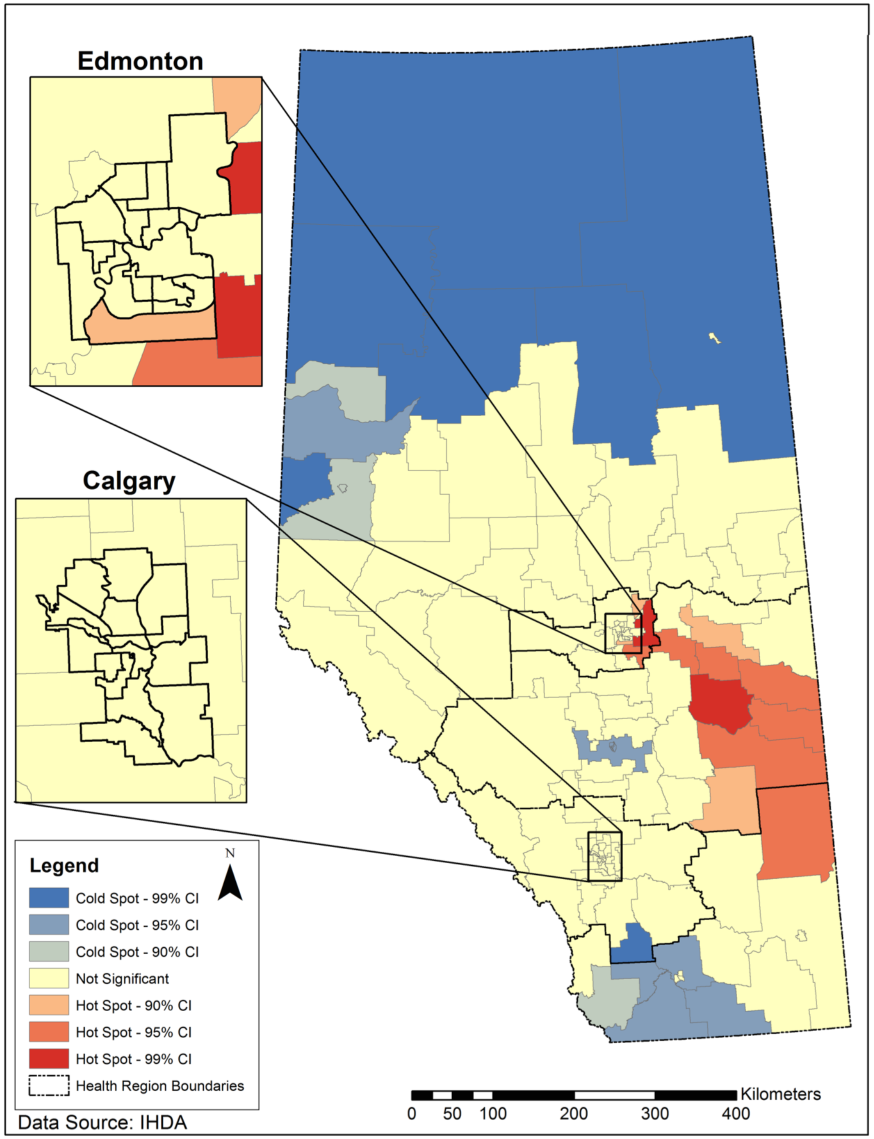
Within the literature, many health geography studies use spatial methods, temporal methods, and less commonly spatio-temporal methods. Spatial analyses of the relationship between space and any given phenomena – for example, place and health – as well as temporal analyses assessing how various health conditions change over time are commonplace in the literature.[8] Conversely, spatio-temporal analyses, and particularly analyses of

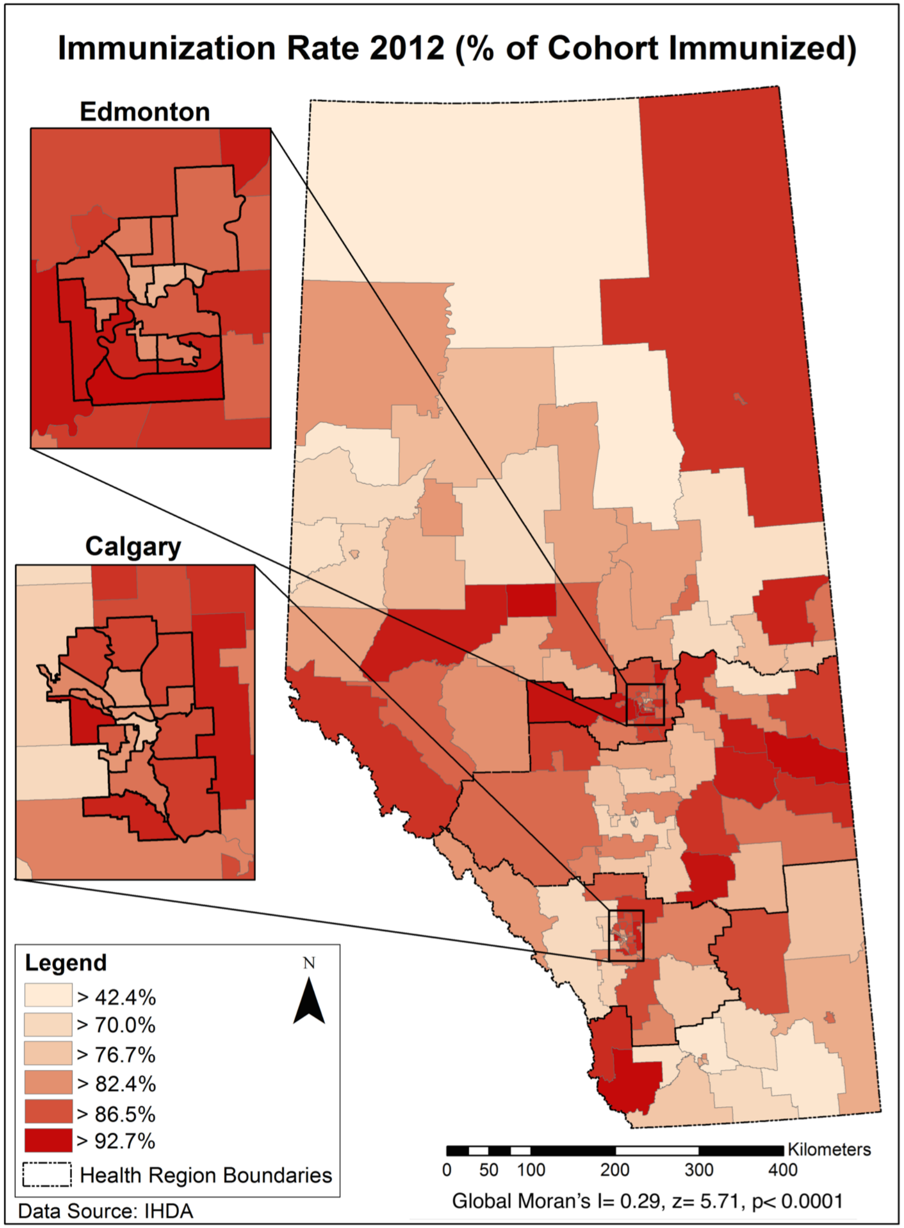
Authors' Affiliation

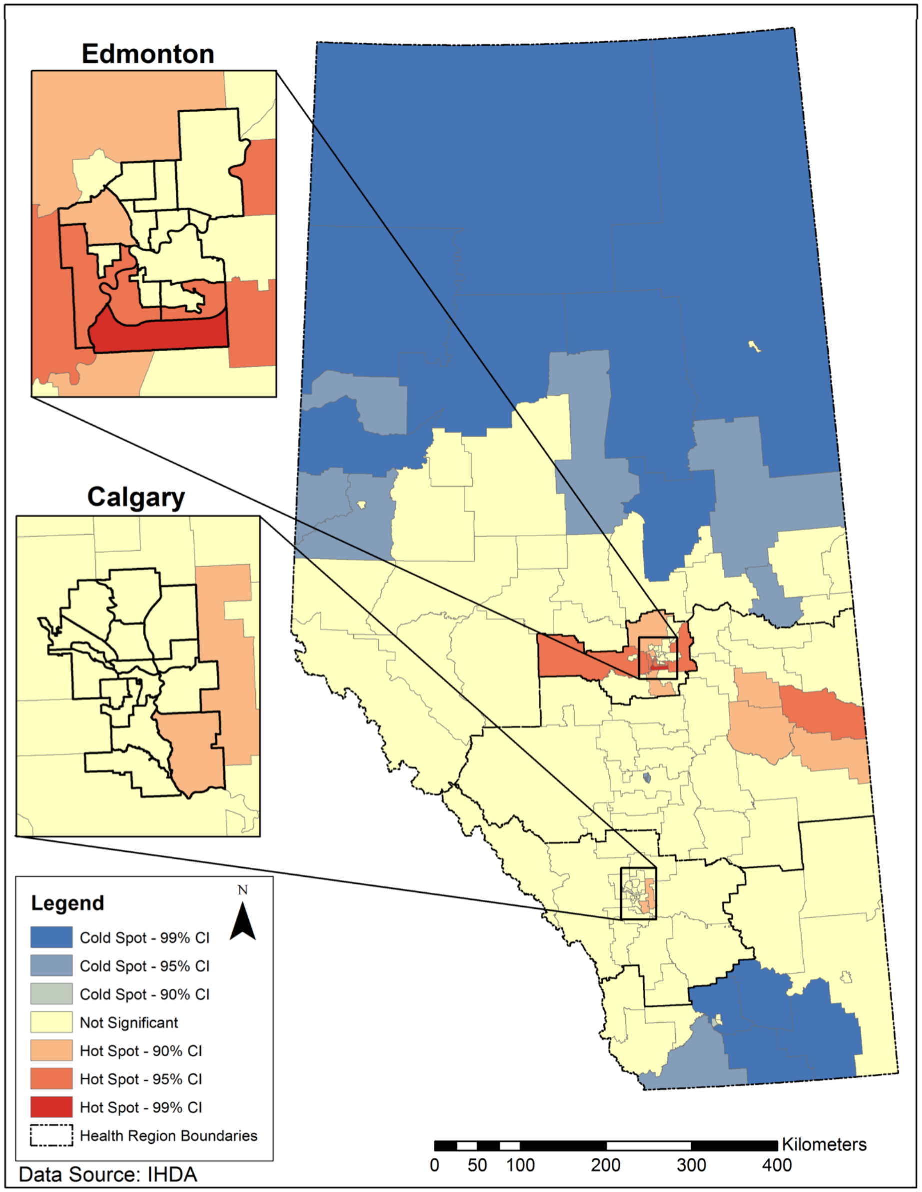
Geography Department, University of Calgary, Calgary, AB
Correspondence: Kristin Eccles, Department of Biology, University of Ottawa, Gendron Hall Room 160, 30 Marie Curie Private, Ottawa, ON K1N 6N5, Tel: ☎613-562-5800, ext. 7500, E-mail: kristin.eccles@uottawa.ca
Conflict of Interest: None to declare.

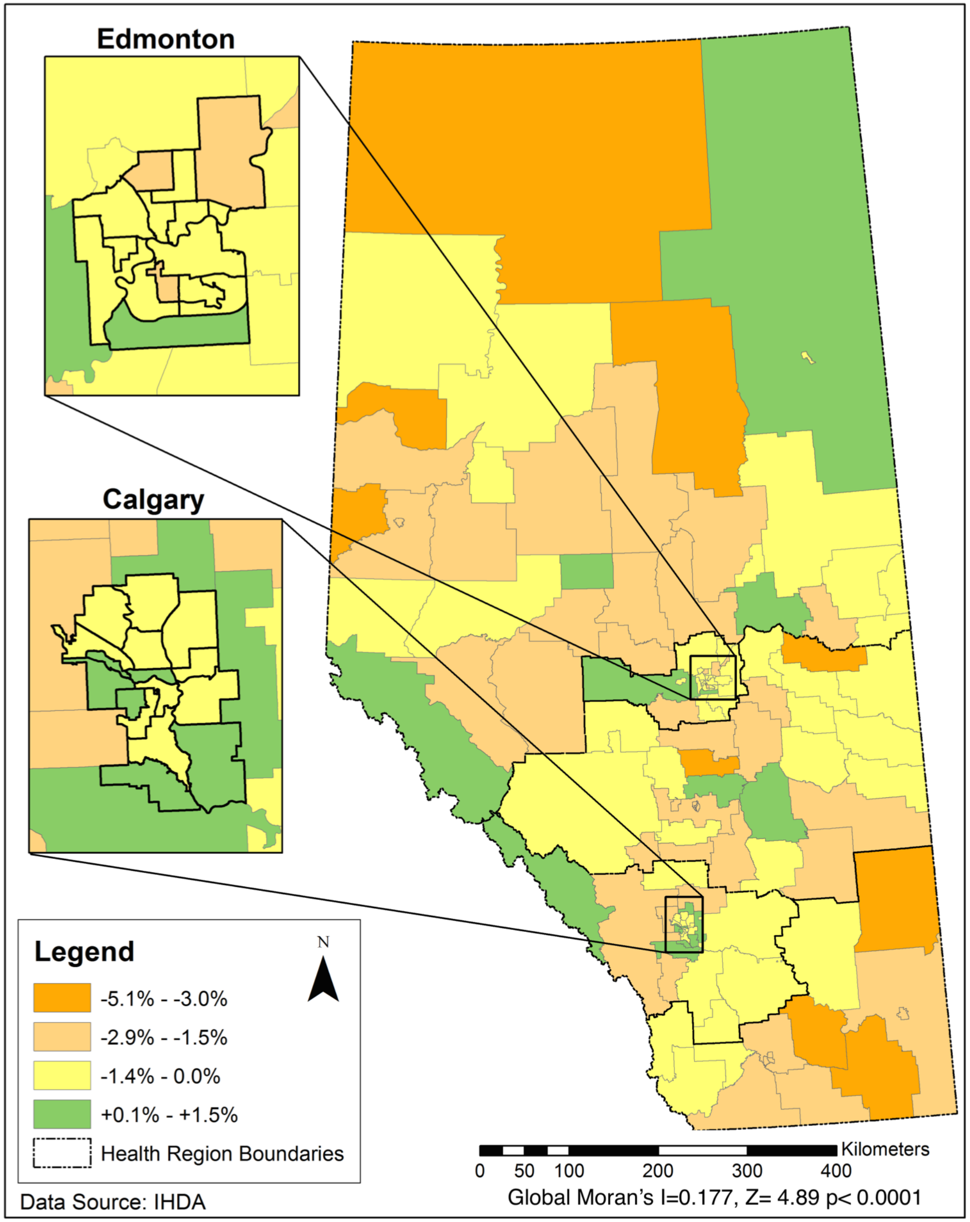


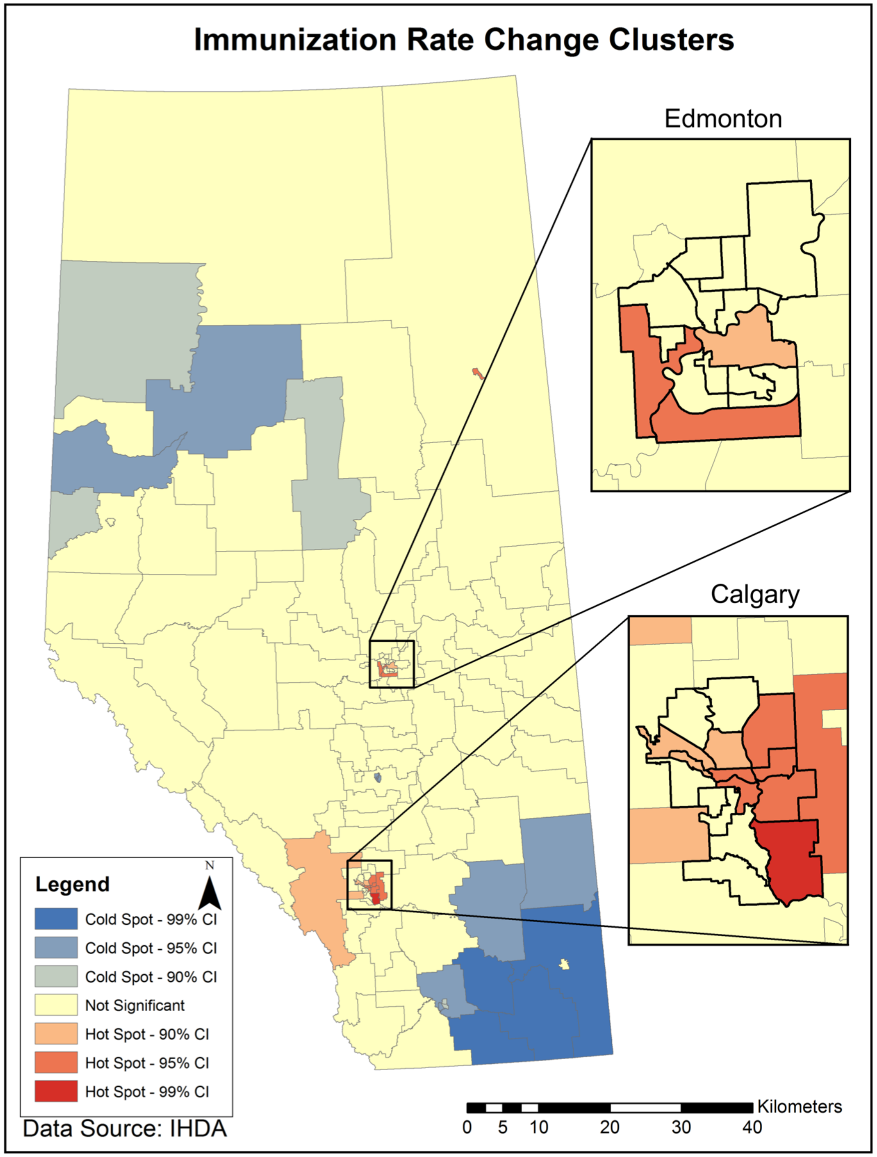

Immunization Rate 2008 (% of Cohort Immunized)
Edmonton
Calgary
Legend
> 42.4%
> 70.0%
> 76.7%
> 82.4%
> 86.5%
> 92.7%
Health Region Boundaries
N
0 50 100 200 300 400 Kilometers
Data Source: IHDA
Global Moran's I= 0.25, z= 4.37, p< 0.0001


Edmonton
Calgary
Legend
Cold Spot - 99% CI
Cold Spot - 95% CI
Cold Spot - 90% CI
Not Significant
Hot Spot - 90% CI
Hot Spot - 95% CI
Hot Spot - 99% CI
Health Region Boundaries
N
0 50 100 200 300 400 Kilometers
Data Source: IHDA


Immunization Rate 2012 (% of Cohort Immunized)
Edmonton
Calgary
Legend
> 42.4%
> 70.0%
> 76.7%
> 82.4%
> 86.5%
> 92.7%
Health Region Boundaries
N
Kilometers
0 50 100 200 300 400
Global Moran's I= 0.29, z= 5.71, p< 0.0001
Data Source: IHDA


Edmonton
Calgary
Legend
N
Cold Spot - 99% CI
Cold Spot - 95% CI
Cold Spot - 90% CI
Not Significant
Hot Spot - 90% CI
Hot Spot - 95% CI
Hot Spot - 99% CI
Health Region Boundaries
Kilometers
0 50 100 200 300 400
Data Source: IHDA


Edmonton
Calgary
Legend
-5.1% - -3.0%
-2.9% - -1.5%
-1.4% - 0.0%
+0.1% - +1.5%
Health Region Boundaries
N
Kilometers
0 50 100 200 300 400
Data Source: IHDA
Global Moran's I=0.177, Z= 4.89 p< 0.0001

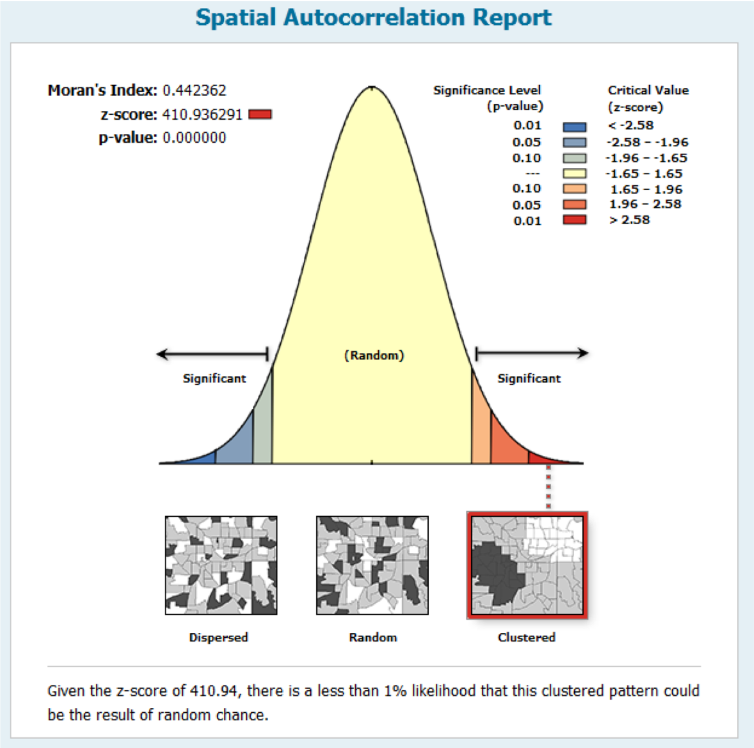

Immunization Rate Change Clusters
Edmonton
Calgary
Legend
N
Cold Spot - 99% CI
Cold Spot - 95% CI
Cold Spot - 90% CI
Not Significant
Hot Spot - 90% CI
Hot Spot - 95% CI
Hot Spot - 99% CI
Data Source: IHDA
Kilometers
0 5 10 20 30 40

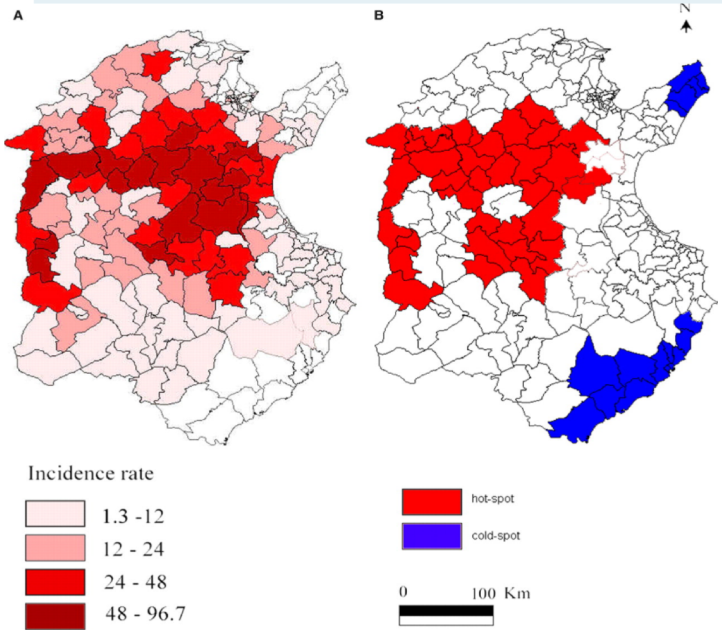
Summary: Detecting Patterns

- **Global spatial methods**
  - One result is generated for an entire area
  - One relationship assumed to represent the study region
  - Eg. Global Moran's I

- **Local Spatial methods**
  - Local methods look for differences across space
  - No assumption of homogeneity
  - Eg. Getis and Ord's G*

Source: OregonState.Edu

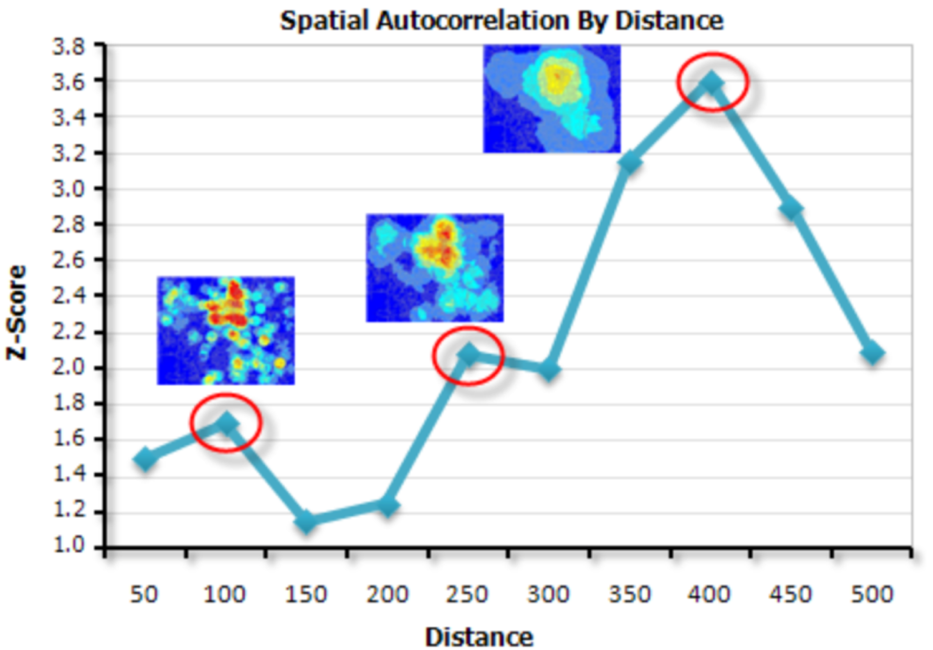




Source: AJTMH

Scale

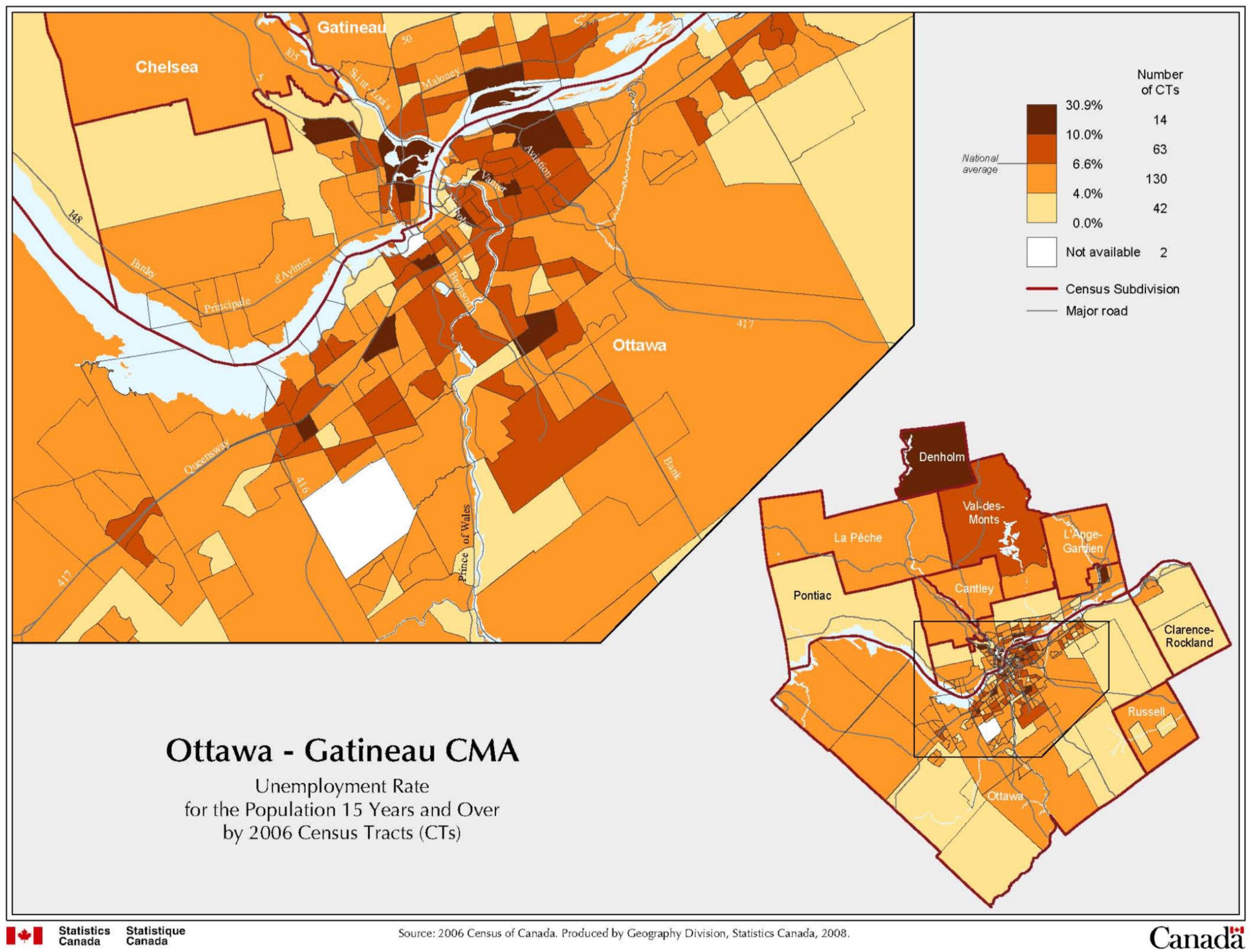
- Measures of spatial autocorrelation (and just about any other measure involving spatial data) are scale-dependent
  - Can cause problems (Uncertainities)
  - If our geographic representation does not match the scale that the phenomena that is a problem
    - False negative, false positive
- We would expect different measures of spatial autocorrelation at different scales



Z-score peaks reflect distances where the spatial processes promoting clustering are most pronounced.

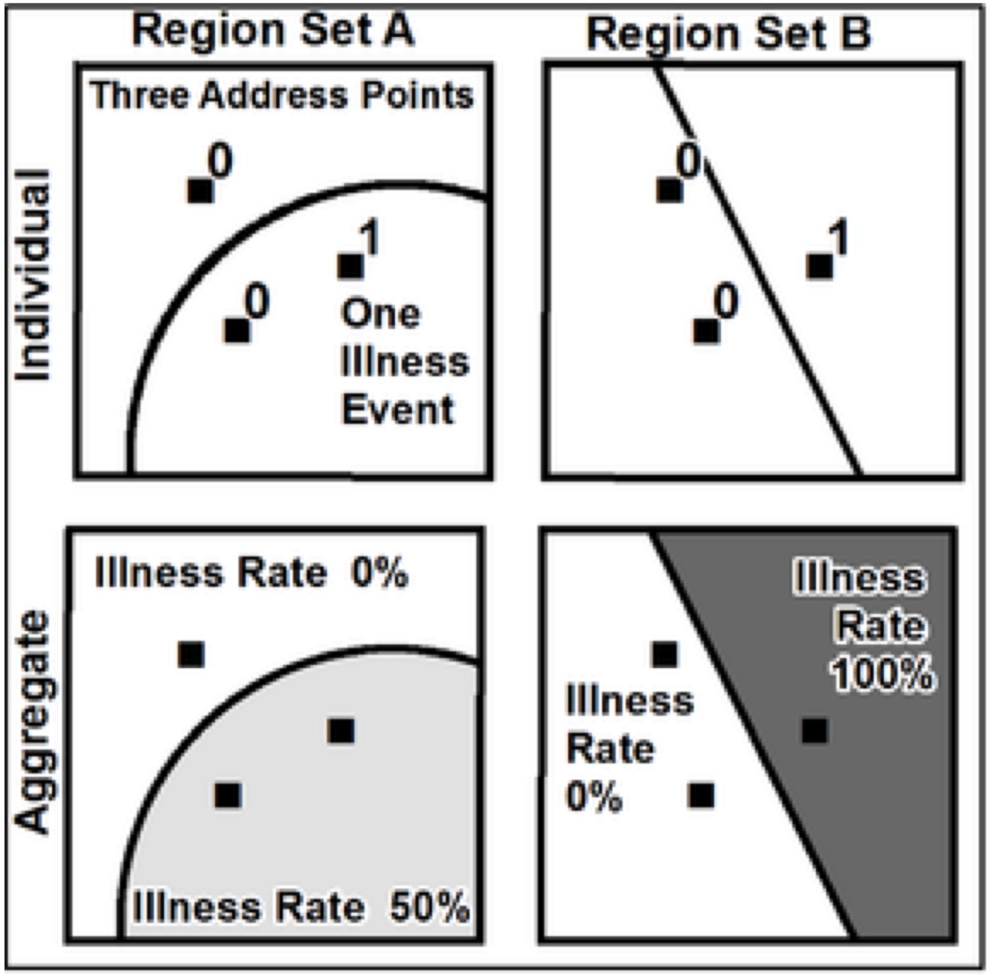
Aggregation of spatial data

- Boundaries are typically pre-selected and arbitrary:
  - E.g. Postal codes, census tract, neighbourhood, municipality, province



Why is this a problem?

- A change in boundaries can change results!



Ecological fallacy

- You cannot make inferences about an individual based on aggregated data
  - Remember that aggregates are averages!
  - Not just a *spatial* issue
  - Any aggregate data is prone to this

Summary

- Spatial properties of data can violate model assumptions:
  - Independence
  - Homogeneity
- Make sure you check for spatial autocorrelation if you use tests that demand statistically independent data
- Autocorrelation is not a bad thing, if you are aware
- Watch out for aggregate data
- MAUP
- Ecological fallacy
- Remember: Scale matters!