ANALYSIS OF SPATIAL DATA IN EPIDEMIOLOGY

Prof. Dr. Maria A Barceló and Prof. Dr. Marc Saez

September 8, 10, 14 and 16, 2021

Research Group on Statistics, Econometrics and Health (GRECS), University of Girona
CIBER of Epidemiology and Public Health (CIBERESP)

GRECS
GRUP DE RECERCA EN ESTADÍSTICA ECONOMETRIA I SALUT

UdG

ciberesp
Centro de Investigación Biomédica en Red
Epidemiología y Salud Pública

COURSE INTRODUCTION

1. Course introduction
2. Introduction to epidemiology and spatial statistics
3. Overview of mixed models
4. Overview of mixed models - Practicals
5. Introduction to INLA and R INLA
6. R INLA - Practicals

Wednesday 8

Friday 10

Marc Saez Maria Antònia Barceló

COURSE INTRODUCTION

7. Disease mapping. Standardisation of incidence and mortality rates
8. Disease mapping. Smoothing standardised incidence and mortality rates
9. Disease mapping – Practicals
10. Geographical association studies. Spatial ecological regression
11. Spatial ecological regression - Practicals

Tuesday 14

Marc Saez Maria Antònia Barceló

COURSE INTRODUCTION

12. **Clustering**
13. Extensions: BYM2, point processes, leaflet, pc priors
14. Extensions – Practicals

Thursday 16

Marc Saez Maria Antònia Barceló

CLUSTERING

Methods for aggregation analysis

- Evaluation (detection) of the presence of clusters
- Localization of clusters
- Clusters around a source

Aggregation analysis works with areal data and with point processes of the health event.

CLUSTERING

What is a **cluster**?

- According to Knox (1989), *'A cluster is a geographically and/or temporally bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance'*.
- That is, a cluster occurs when there is a general tendency for a more non-random or 'clustered' distribution of the disease than would be expected to result from variations in population structure and probability fluctuations.

CLUSTERING. RISK HETEROGENEITY ANALYSIS

- There are differences among the risks in each of the spatial units analysed.
- Whether there is a spatial structure in these differences is not evaluated
- **Causes of heterogeneity:**
  - Presence of a pollution source that would cause increased risk in its environment.
  - Spatial variation of a risk factor (more risks related to higher exposure to this factor).
  - Other sources

CLUSTERING VS. CLUSTER DETECTION

Disease Clustering and Cluster detection: conclusion

- **Clustering**
  - We have defined a number of statistics (Moran, Tango) to determine the level of clustering in a set of data
  - In the context of count data in spatial epidemiology these methods have some drawbacks due to non-constant variance of the response

    when the response is the SMR, variance of the SMR depends on the expected numbers

  - These methods are useful in an *exploratory* step in an analysis
  - Hierarchical modelling provides greater information...
- **Cluster detection**
  - The most popular method for cluster detection is the scan statistic
  - Crucial choice: maximum size of the cluster (population)

CLUSTERING VS. CLUSTER DETECTION

Spatial aggregation analysis (clustering) in general, consists of assessing (detecting) and locating areas where risks tend to be higher than expected.

Cluster detection consists of the identification of "unusual" groupings of cases, allowing the identification of hotspots.

CLUSTER DETECTION

Cluster detection

- Methods discussed so far evaluate the tendency for global disease clustering, but no information on the location of the clusters

- Different methods available for cluster detection, depending on the nature of the data
- Methods using 'windows' to investigate spatial patterns:
  - Superimpose a number of circular windows on the region of interest
  - Determine whether the number of cases in each window is larger than expected
- Different methods define the circles in terms of:
  - distance (Openshaw)
  - number of cases (Besag and Newell)
  - population size (Scan statistic)
- These methods may be used as particular areas may be highlighted and subsequently investigated

CLUSTER DETECTION

- Scan Statistics: family of methods
- Kulldorff
  - Establishment of a circular window of variable size. It can be based on fraction of areas or populations
  - Compare the overall risk of the regions inside and outside the window

CLUSTER DETECTION

- Kulldorff
  - Null hypothesis: no aggregation = the two risks are equal
  - Alternative hypothesis: the risk is larger within the window.
  - Likelihood ratio test
  - The most likely cluster is the one with the highest likelihood ratio with the best significance.

CLUSTER DETECTION - KULLDORF

Scan statistics (Kulldorff)

- spatial scan statistic implemented in the software Satscan (http://www.satscan.org/)
- drawing all possible circles/ellipses in the study region with areas' centroids as the centers
- size of the circles/ellipses based on population or number of areas in the cluster
- determining whether there were more cases inside the circle than expected
- the most likely cluster is found by a likelihood ratio test

CLUSTER DETECTION - KULLDORF

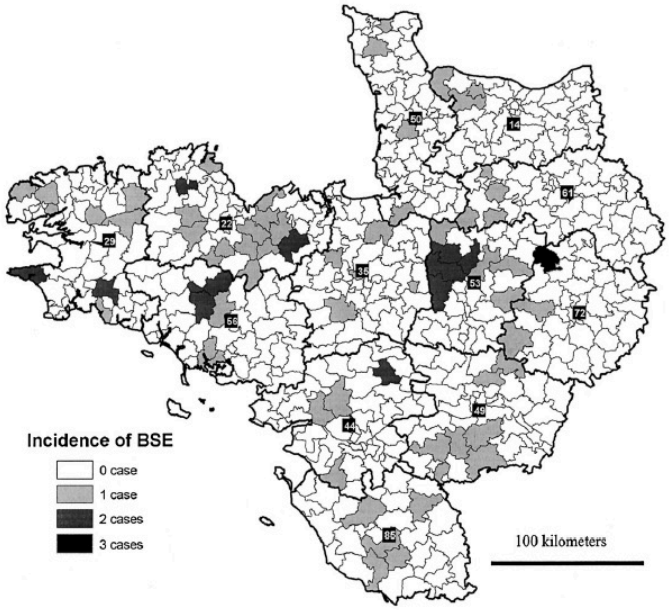


Figure 1. Incidence of bovine spongiform encephalopathy in western France ("canton" level) between August and December 2000.

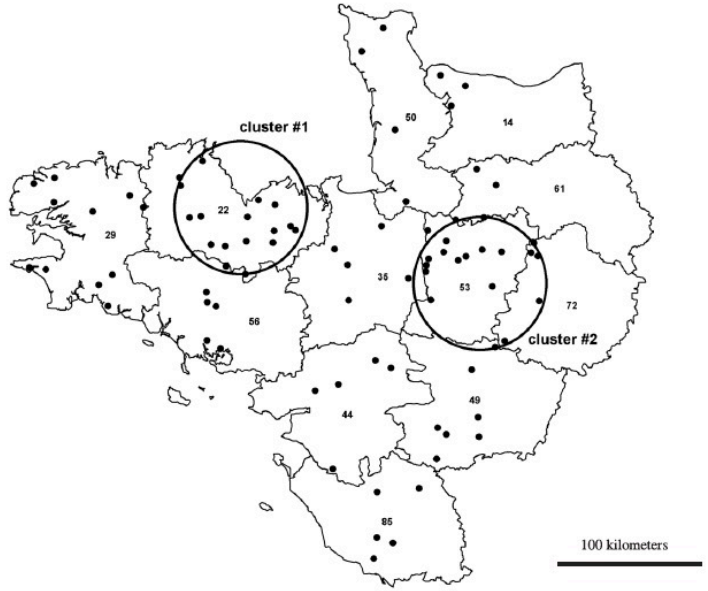


Figure 3. Spatial distribution of the 84 bovine spongiform encephalopathy cases (black dots) in western France between August and December 2000. Two clusters appeared with the method of Kulldorff: Cluster #1: 18 observed cases and 6 expected; 45 km radius. Cluster #2: 18 observed cases and 7 expected; 44 km radius.

Abrial D, Calavas D, Lauvergne N, Morignat E, Ducrot C. Descriptive spatial analysis of BSE in western France. *Vet Res.* 2003; 34(6):749-760. doi: 10.1051/vetres:2003032

CLUSTER DETECTION - KULLDORF

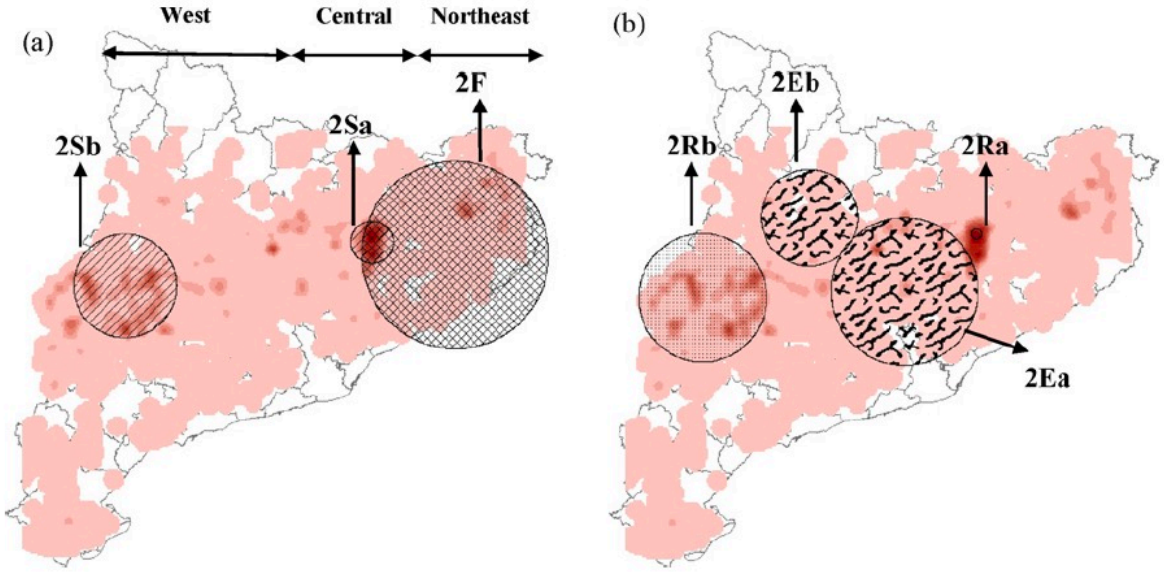
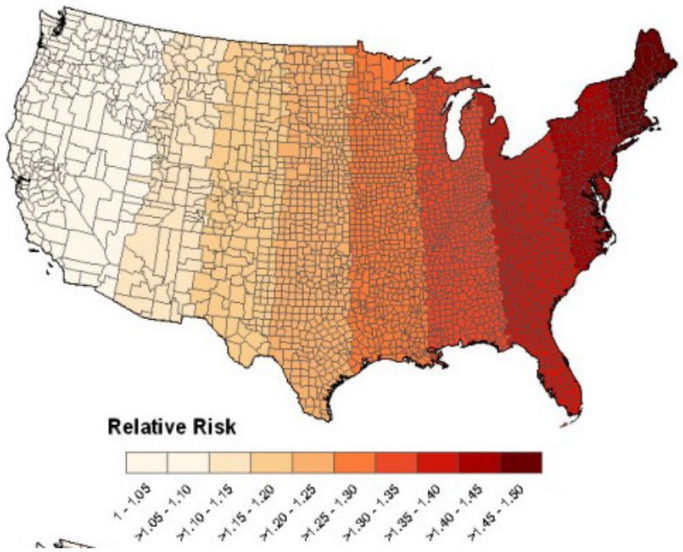


Fig. 2. Clusters identified in the second period (June 2004–May 2005) with the spatial scan statistic (Bernoulli model). (a) Clusters of positive sow farms (S) and fattening farms (F), (b) clusters of elimination (E) and reinfection (R). Clusters are represented over a Kernel density surface of pig farms.

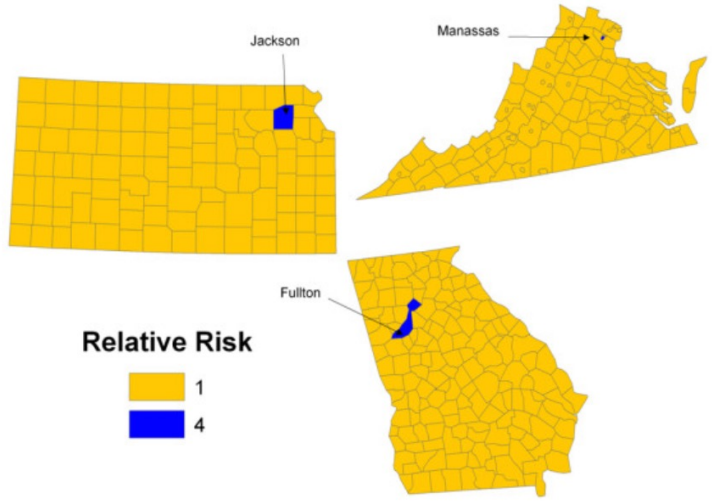
Allepuz A, Saez M, Alba A, Napp S, Casal J. Exploratory spatial analysis of Aujeszky's disease during four phases of the eradication programme in Catalonia, Spain (2003-2007). *Prev Vet Med.* 2008; 86(1-2):164-75. doi: 10.1016/j.prevetmed.2008.04.005

CLUSTER DETECTION - KULLDORF



Global clustering

From Jackson et al, 2009



Localised clusters

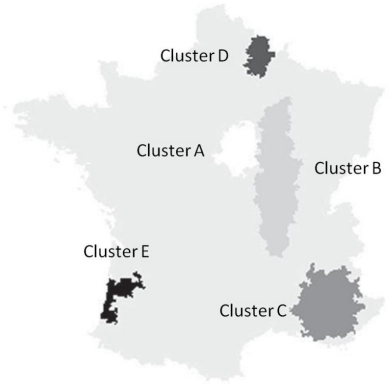
Jackson MC, Huang L, Luo J, Hackey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. *Int J Health Geogr.* 2009; 8:55. doi: 10.1186/1476-072X-8-55

CLUSTER DETECTION - KULLDORF

Example: Childhood acute leukaemia in France, 1990-2006

Demoury et al, 2012

- This study did not find evidence of any global spatial heterogeneity of AL incidence rates in France (Potthoff-Whittinghill test)
- Although no significant spatial cluster was detected over the whole period, the study identified a few significant spatial clusters in specific periods
- Even though the significance levels of those clusters do not strongly support the existence of local risk factors, the clusters may still reflect a slight impact of shared risk factors



- cluster A: AL for 90–94, P=0.08
- cluster B: AML for 95–00, P=0.08
- cluster C: ALL for 95–00, P=0.06
- cluster D: ALL 1–4 years for 95–00, P=0.09
- cluster E: ALL 1–4 years for 01–06 P=0.08

Demoury C, Goujoon-Bellec S, Guyot-Goubin A, Hémon D, Clavel J. Spatial variations of childhood acute leukaemia in France, 1990-2006: global spatial heterogeneity and cluster detection at 'living-zone' level. *Eur J Cancer Prev*. 2012; 21(4):367-374. doi: 10.1097/CEJ.0b013e32834e31d8

CLUSTER DETECTION - KULLDORF

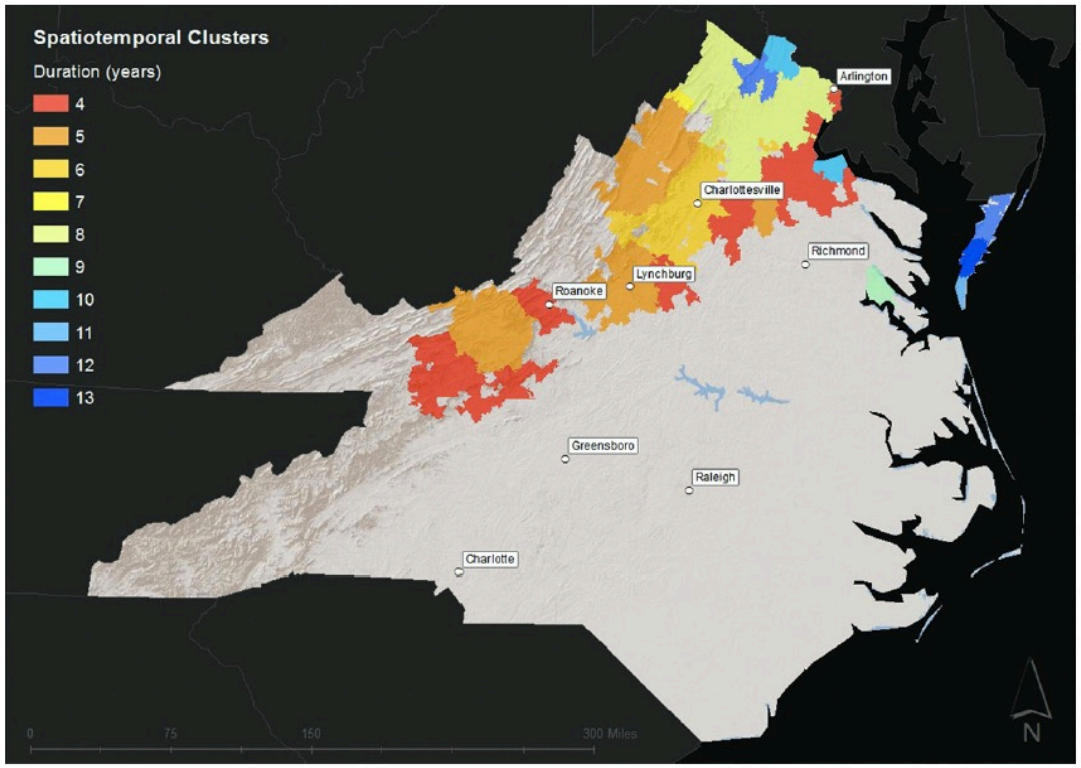


Figure 3. Spatiotemporal scanning statistic cluster analysis. This analysis was limited to spatial clusters of at least 3 years duration and a *P* value of ≤.0001, ensuring that only significant and durable trends were identified. Clusters of longest duration were in northern and eastern Virginia, and the most recently emergent clusters extended along the Appalachian Mountains towards southwest Virginia.

Lantos PM, Nigrovic LE, Auwaerter PG, Fowler VG, Ruffin F, Brinkerhoff RJ, Reber J, Williams C, Broyhill J, Pan WK, Gaines DN. Geographic Expansion of Lyme Disease in the Southeastern United States, 2000-2014. *Open Forum Infect Dis*. 2015; 2(4):ofv143. doi: 10.1093/ofid/ofv143

CLUSTER DETECTION - KULLDORF

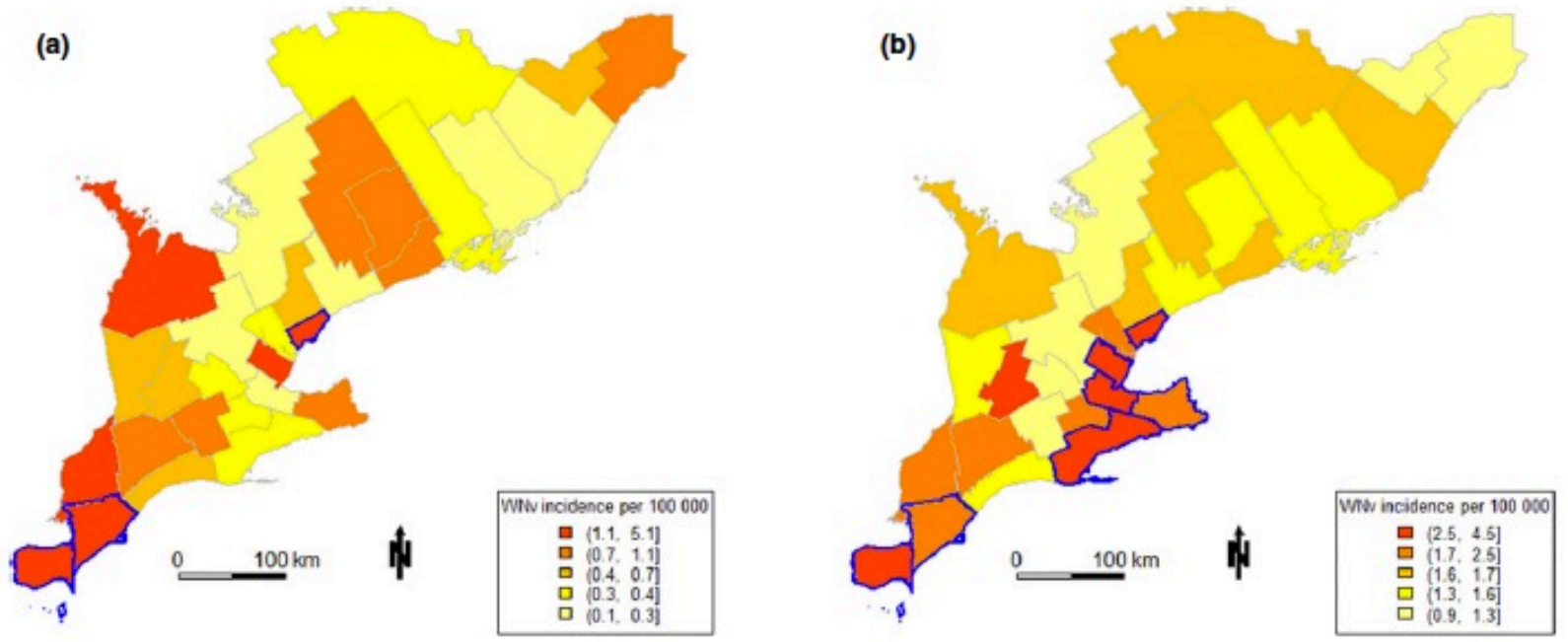
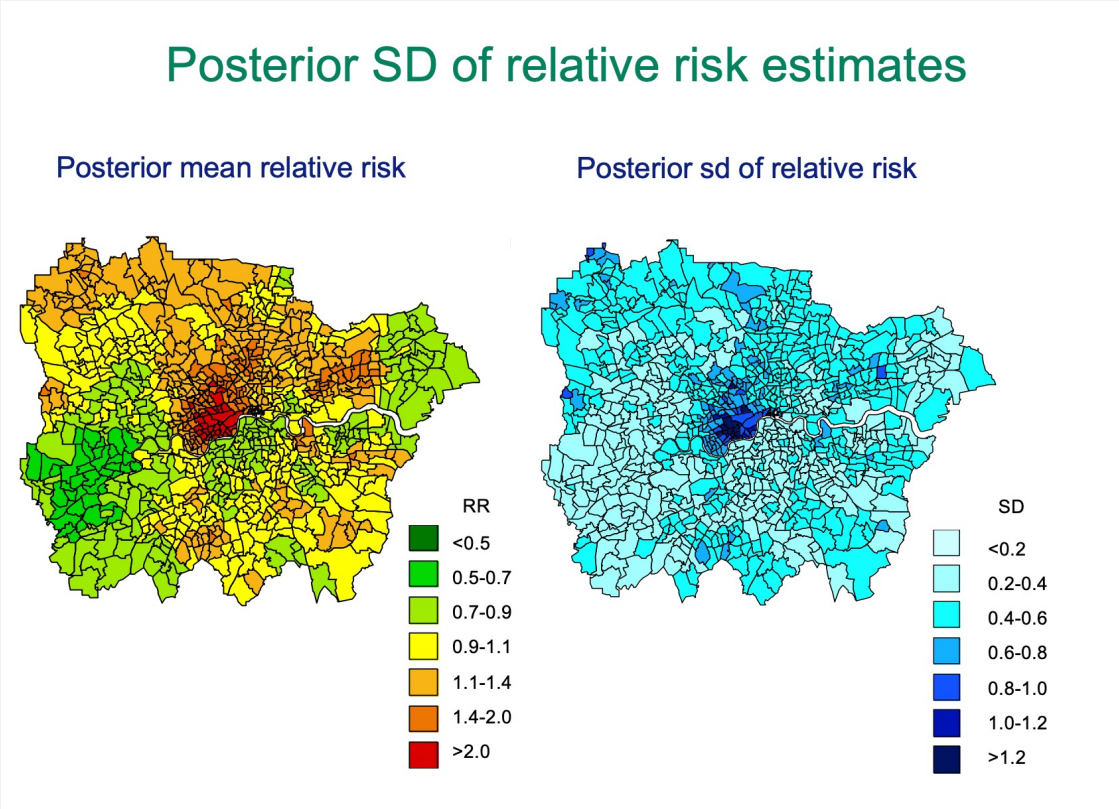


Fig. 1. Choropleth maps of empirical Bayesian smoothed annual human risk incidence estimates of WNV disease per 100 000 population for the 29 health units of southern Ontario for the years a) 2005, b) 2012 and c) 2012 adjusted. Outlined areas represent health units where clusters of disease were identified by the spatial scan statistic.

Thompson M, Berke O. Evaluation of the control of West Nile virus in Ontario: Did risk patterns changre from 2005 to 2012?. *Zoonoses Public Health.* 2017; 64(2):100-105. doi: 10.1111/zph.12285

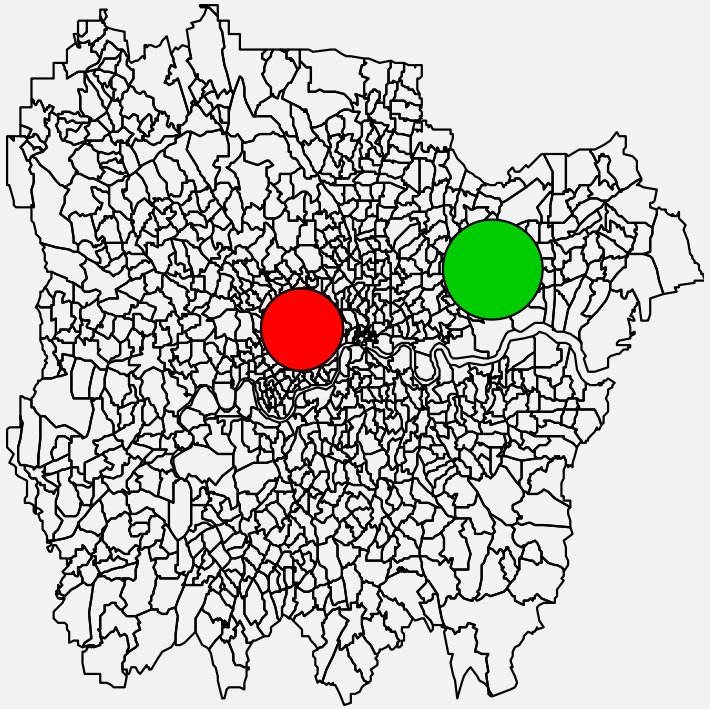
CLUSTER DETECTION - KULLDORF

Example: Childhood leukaemia incidence in London, 1986-1998

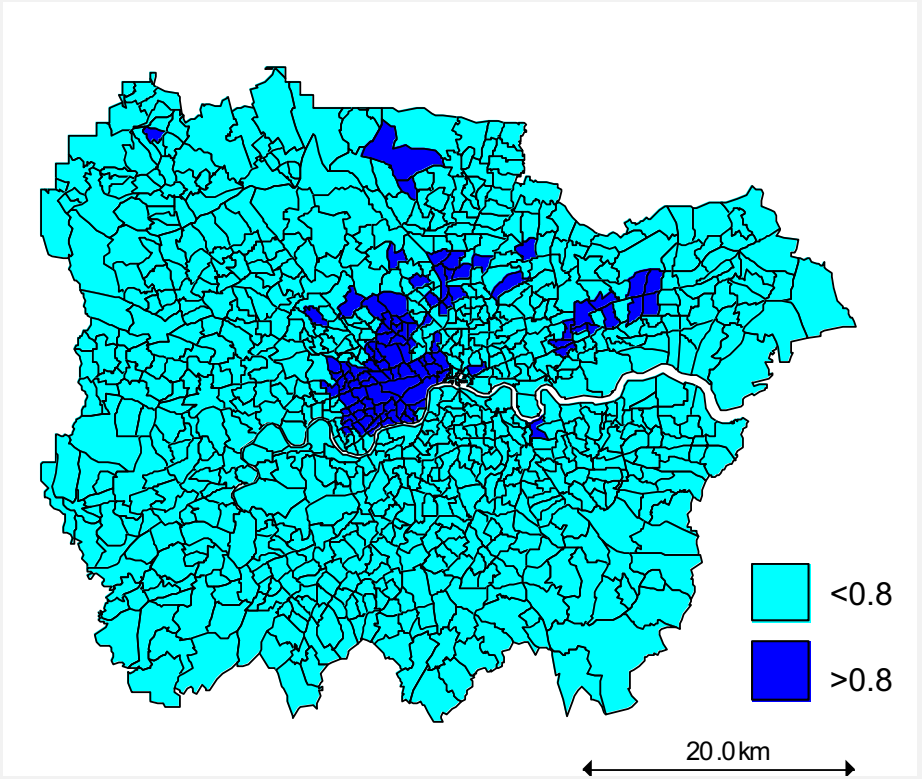


CLUSTER DETECTION - KULLDORF

Example: **Childhood leukaemia incidence in London, 1986-1998**



- Most likely cluster; p<0.001
- 2nd most likely cluster; p = 0.2



Teaching statistics

Doing Statistics