

# Geo-epidemiológia

Solymosi Norbert

*Kvantitatív állatorvosi epidemiológia*

Járványtani és Mikrobiológiai Tanszék

Állatorvostudományi Egyetem

2020. november 4.

Az angol szakirodalomban *spatial epidemiology*-nak nevezik az epidemiológiának azt az ágát, amely az egészséggel összefüggő adatok földrajzi feldolgozásával, elemzésével foglalkozik.

A *geo-epidemiológia* megnevezést azért használjuk itt, mert az egyéb lehetőségek még inkább félrevezetőek lehetnek.

A *geo* szó helyett használhatnánk a *térbeli* előtagot is, a térinformatika kifejezés magyarázatához hasonlóan. De mint, ahogy annál is tér helyett egyértelműbb a földrajzi információs rendszer a GIS (Geographic Information System) megjelölésére, úgy a *térbeli* megnevezés a legtöbbekben nem pusztán kétdimenziós megközelítést jelent.

A geomatematika és geostatisztika elnevezése nyomán tehát addig, amíg nem találunk jobb nevet, a *geo-epidemiológia* használatát javasoljuk.

Főbb területei:

- betegség-térképezés
- földrajzi mintázatelemzés
- összefüggés-elemzés

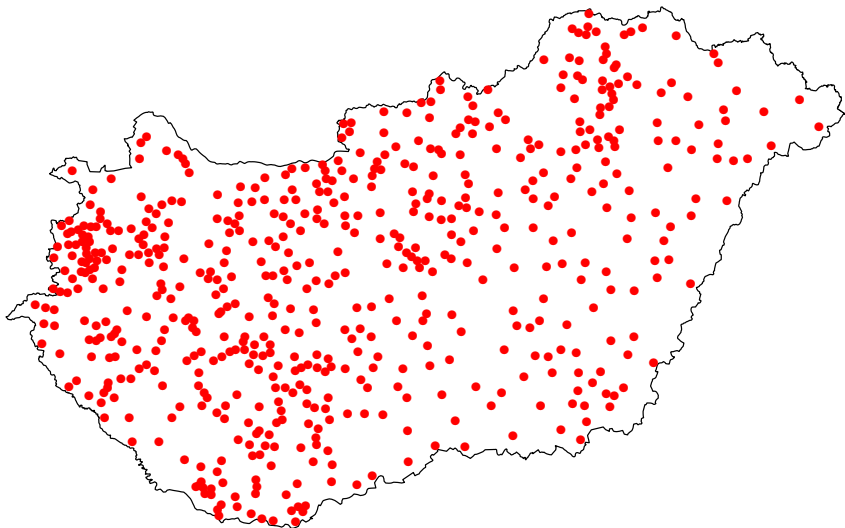
Az egészséggel kapcsolatos információk földrajzi ábrázolásának célja a kockázat térképezése.

A kockázat térképi megjelenítése jelentős többletinformációval szolgál az egyszerű táblázatos megjelenítéshez képest.

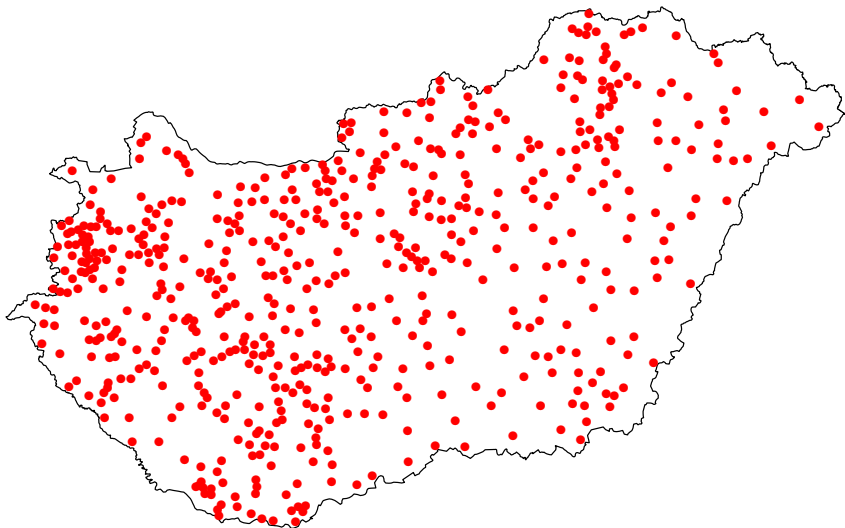
A betegség-térképezéshez számos módszert alkalmaznak az epidemiológiában.

A legegyszerűbb térképezési módszer, amikor a betegség előfordulásának helyét ponttal jelöljük a térképen.

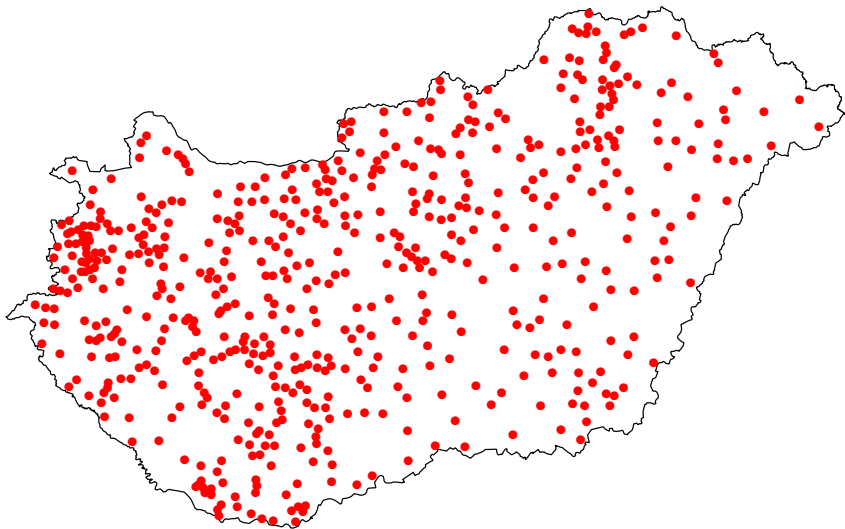
Az egészséggel kapcsolatos információk leggyakrabban pontszerű földrajzi referenciával rendelkeznek nyers formájukban (pl. GPS-koordináta, lakcím, tartási hely).



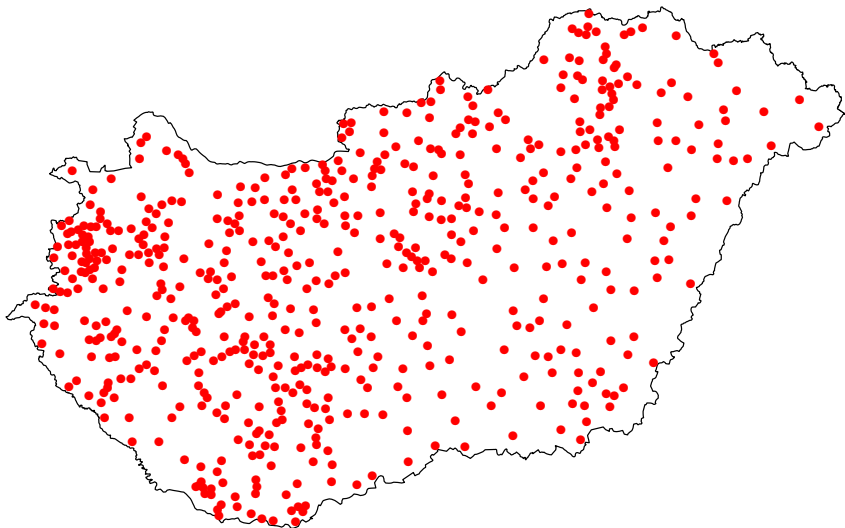
A piros pontok azokat a településeket jelölik, ahonnan származó rókákban az 1990. év során veszettséget állapítottak meg.



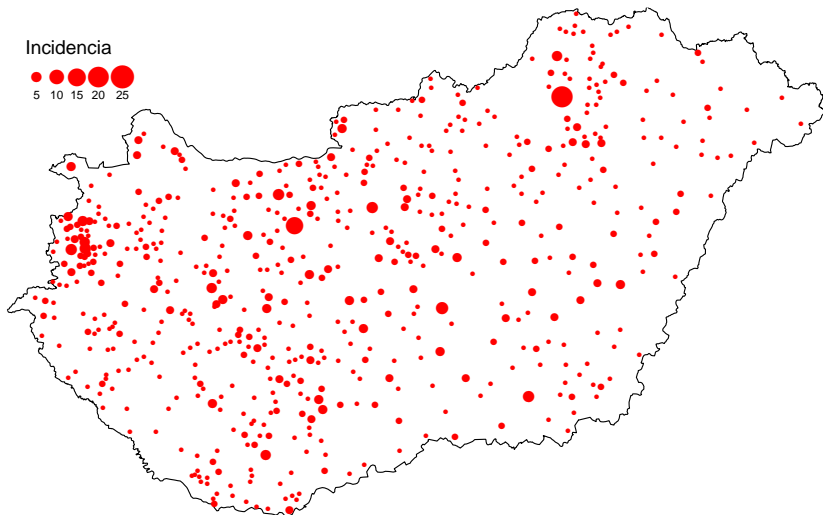
Azonban a pont alapú térképezésből legtöbbször csak kevés információ fogalmazható meg a kockázat földrajzi mintázatára vonatkozóan.



A *pont-térképek* alapján kialakított benyomásokat jelentősen befolyásolhatja a térképezésnél használt felbontás, a pontok mérete, alakja és színe.

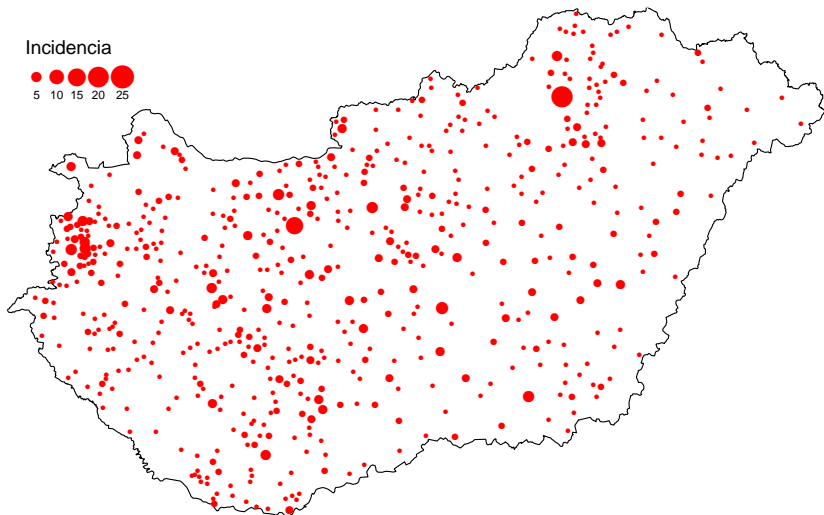


Gyengesége a pont-térképeknek, hogy az esetszámot nem tudjuk megjeleníteni velük, csupán a pozitív/negatív pozíciókat.

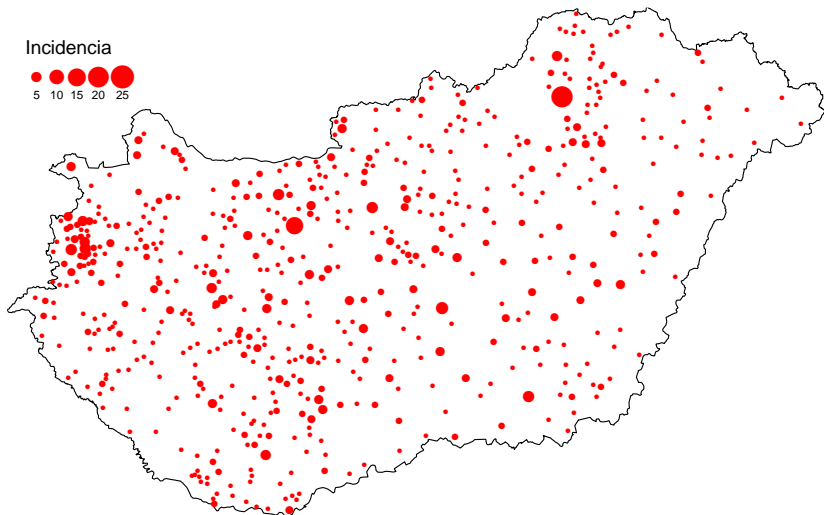


Az ún. *buborék-térkép* segítségével már megjeleníthető, hogy adott helyen, településen hány esetet találtak az adott időszakban.

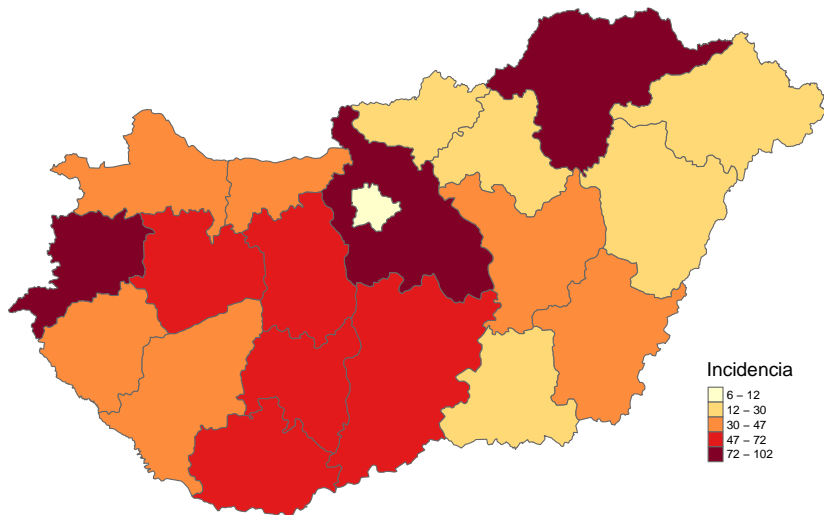




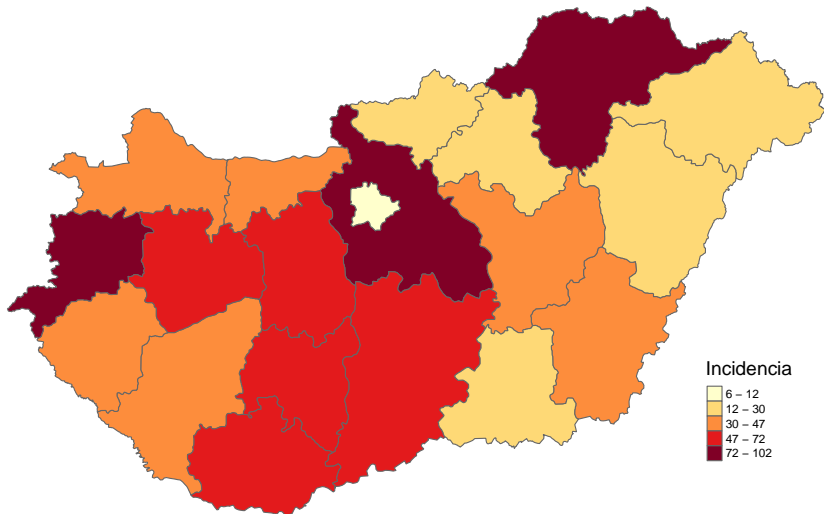
A térképen alkalmazott jel (esetünkben korong) mérete mutatja a veszett róókák számát.



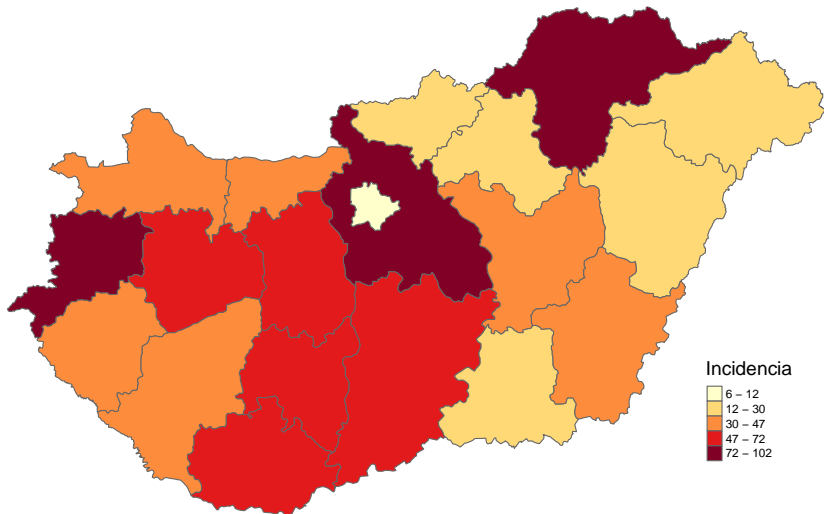
Itt a pont-térképeknél említett technikai tényezők mellett további torzító mozzanat, hogy ha van olyan település, melyben sok eset volt, akkor a szomszédságában lévő pozitív helyet eltakarhatja az esetszámot megjelenítő jel.



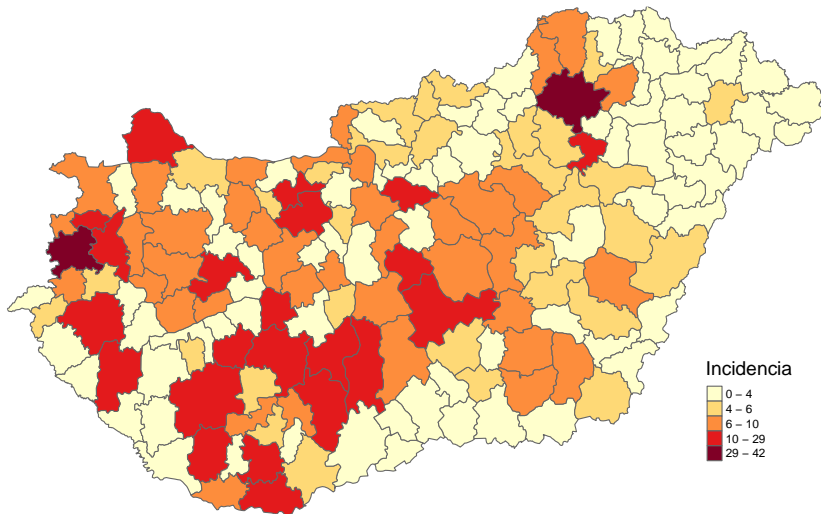
A kockázat térképezésében további lehetőségeket jelentenek az ún. *choropleth* térképek. Az elnevezés görög eredetű, a *khora* a település külterületét, a *plethos* pedig a sokaságot jelenti.



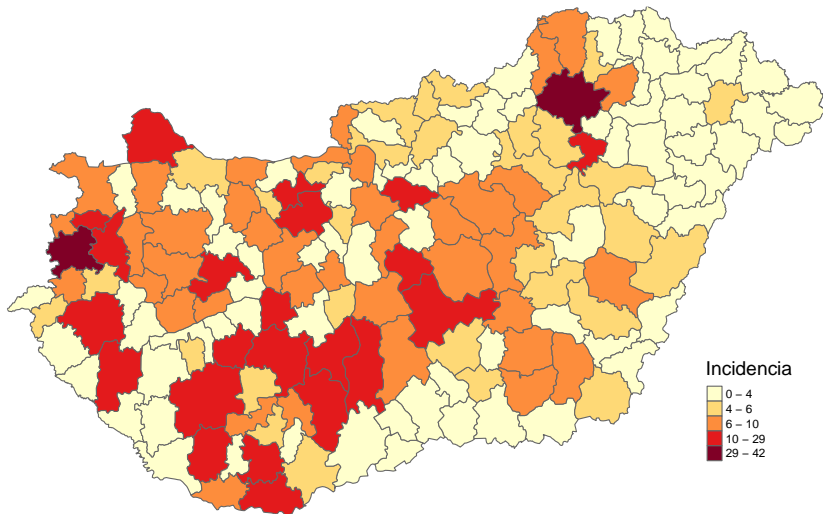
A betegség-térképezésben ennek megfelelően olyan térképeket jelöl, amelyek valamilyen területi egységen aggregálva mutatják be az előfordulási kockázatot.



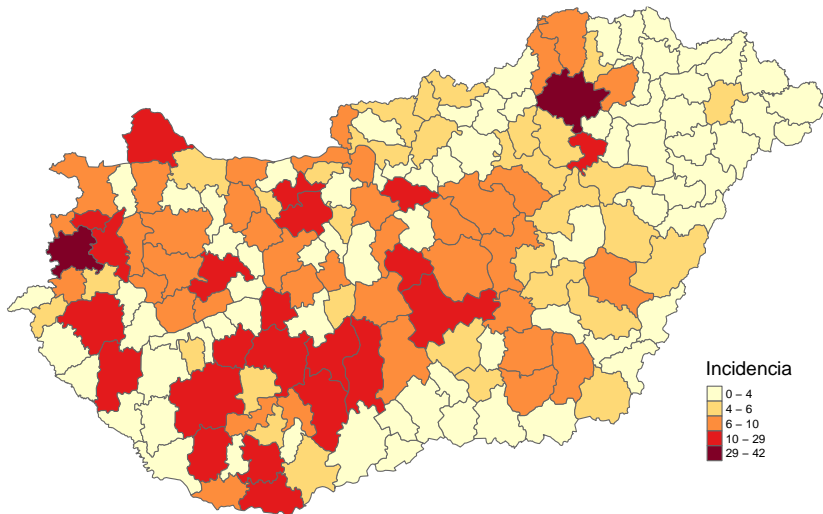
Az aggregálás alapja tetszőleges lehet, habár jelentősen befolyásolhatja a térképből levont következtetéseinket. Megyei szintű aggregálás.



Kistérségi szintű aggregálás.

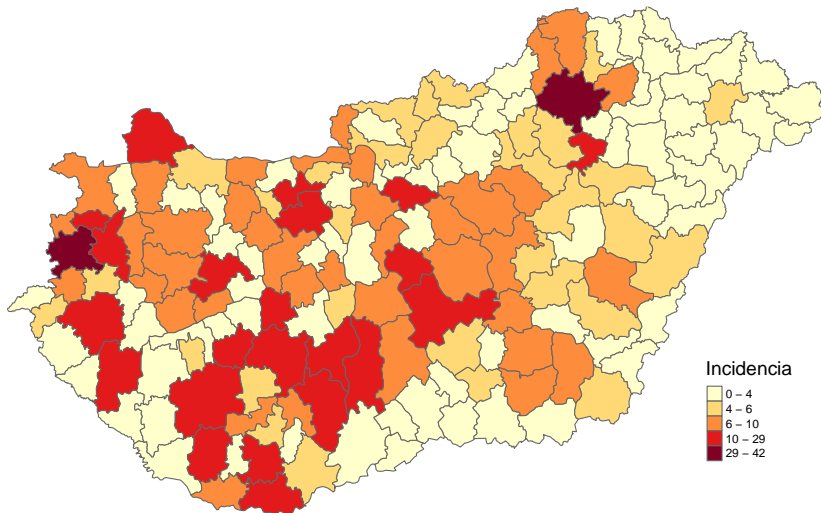


Térinformatikai szempontból a choropleth térképeken az aggregálás és ábrázolás alapegységei poligonok. Ezek a poligonok szomszédaikkal szorosan érintkezve kitöltik a vizsgált területet.

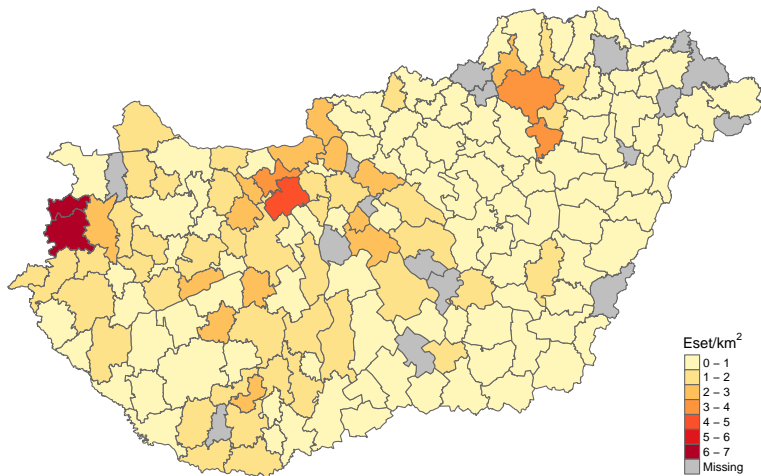


Így elkerülhetővé válik a buborék-térképnél említett szomszédos adatok kitakarása. Továbbá így egy határokkal ugyan megszakított, de folytonos felületen ábrázolhatjuk az adatainkat, ami megkönnyíti a kockázatra vonatkozó következtetések kialakítását.

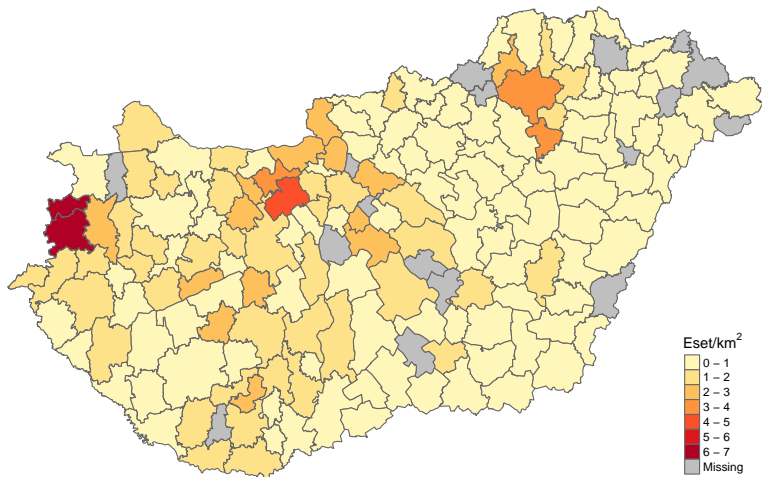




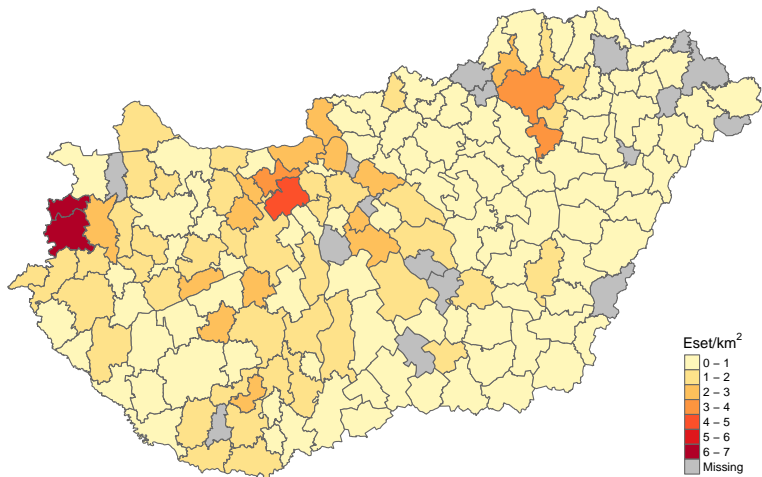
Úgy tűnhet, hogy a nagyobb területű kistérségek nagyobb kockázattal bírnak. Mielőtt ezt elfogadnánk, gondoljunk arra, hogy ha a rókák veszzettségi kockázata homogén lenne az országban, akkor nagyobb területű településeken több esetünk lenne.



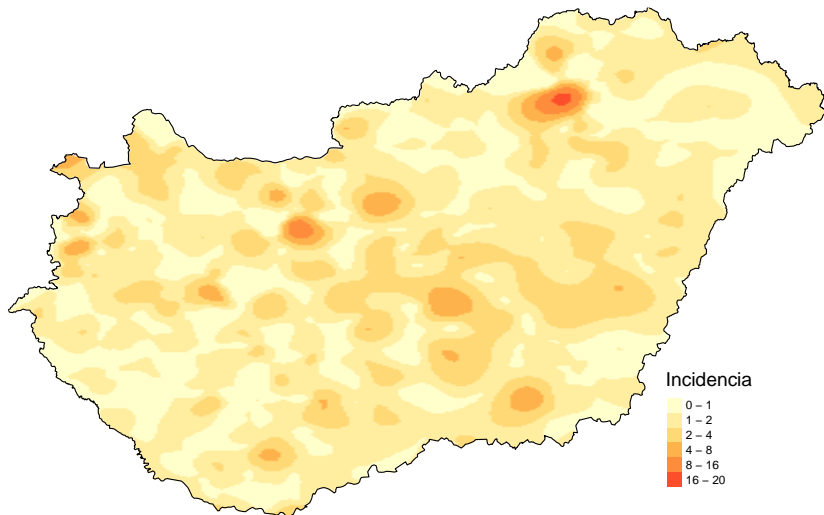
Mivel rókákra vonatkozó megbízható populációs adatok nem állnak rendelkezésünkre, a populáció helyett használjuk a kistérség területének méretét az esetszám normálására.



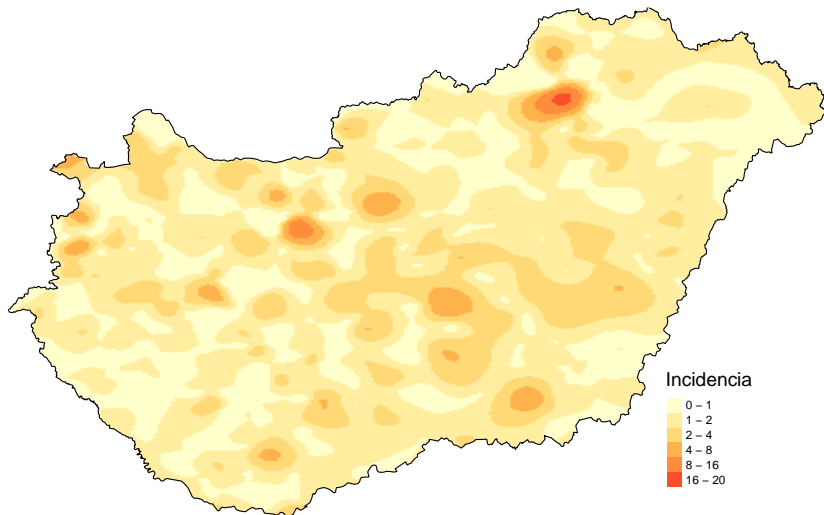
Ez alapján úgy tűnik, hogy egy magas esetszámú góc volt 1990-ben a szombathelyi kistérségben és környezetében, illetve néhány kisebb kockázatos góc a Dunántúlon és egy Miskolc környezetében.



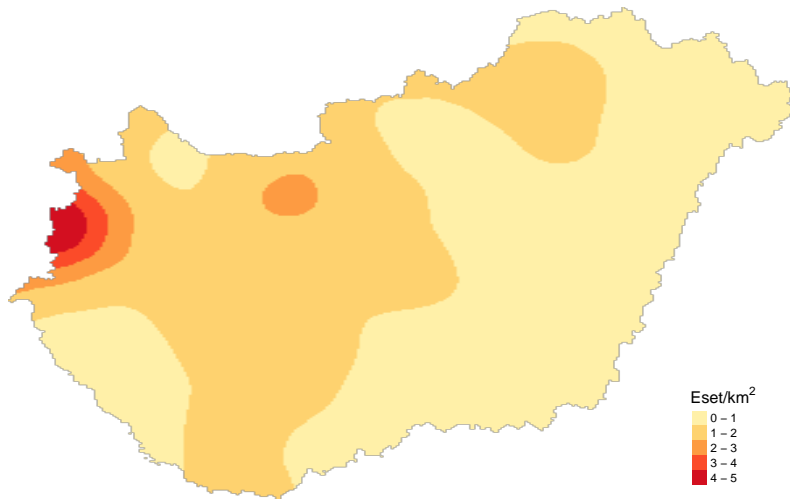
A choropleth térképek értékelhetőségét jelentősen befolyásolhatja a területi egységek színezésére használt színskála és az értéktartományok megválasztása.



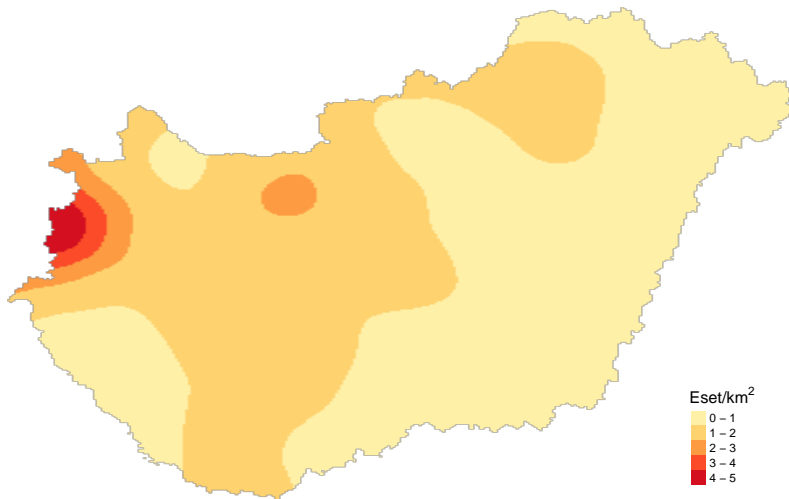
A kockázat földrajzi mintázatának bemutatására további lehetőséget nyújt az ún. *isopleth* térképezés. Az isopleth térképek létrehozásánál a pontszerű forrásadatokat valamilyen simítási módszerrel interpoláljuk a kétdimenziós térben.



Az 1990. év során diagnosztizált esetek települési szintű adataiból létrehozott isopleth térkép.

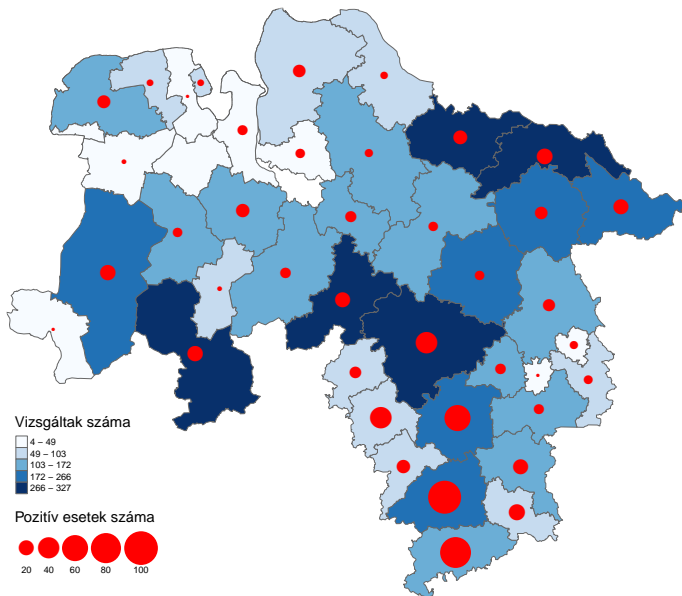


Az 1990. év során diagnosztizált esetek kistérségi területtel normált adataiból létrehozott isopleth térkép.

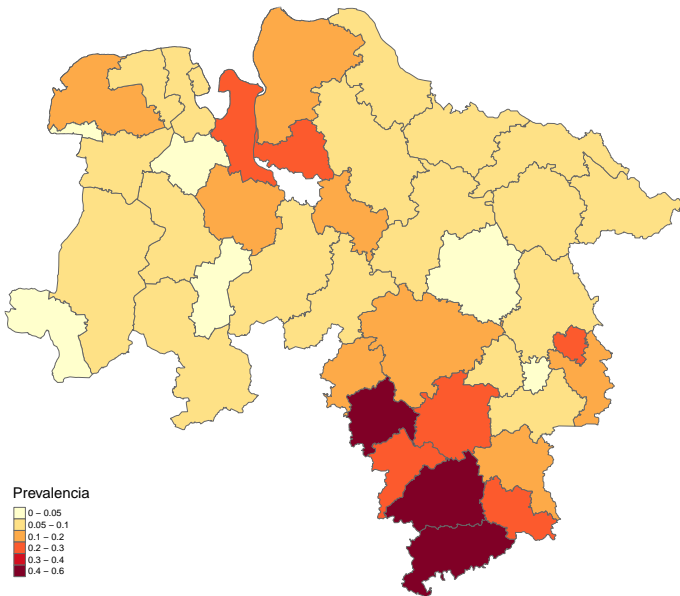


Az isopleth térképek létrehozására használt simítási eljárás, illetve annak paraméterezése nagy mértékben befolyásolhatja a kockázati térképet, így az abból levonható epidemiológiai következtetéseinket.

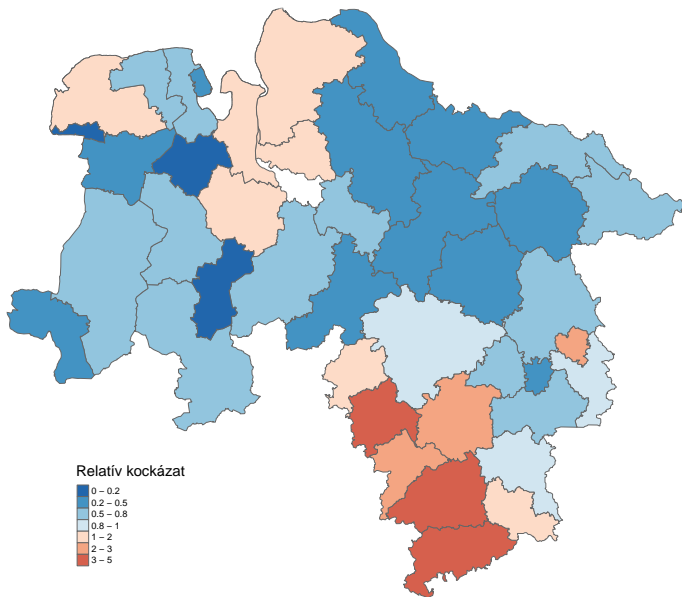




A betegség-térképezés eddigi eljárásai kombinálódhatnak is, illetve egyéb információkkal egészülhetnek ki. Alsó-szászországi járásokban megvizsgált és *E. multilocularis*-pozitív vörös rókok száma (Berke, 2001).



Az *E. multilocularis*-prevalencia járásonként, a prevalenciát egyszerűen a pozitív esetek részarányaként becsülve.



Rókák *E. multilocularis* fertőzöttségének relatív kockázat térképe. Az RR a teljes terület átlagos kockázat arányában fejezi ki az egyes járások kockázatát.

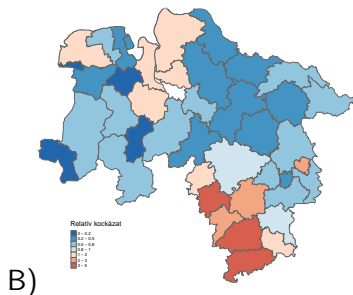
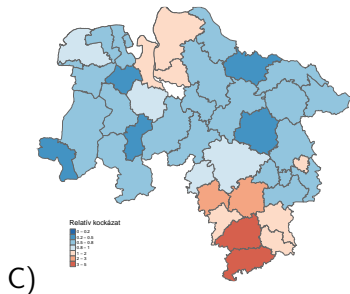
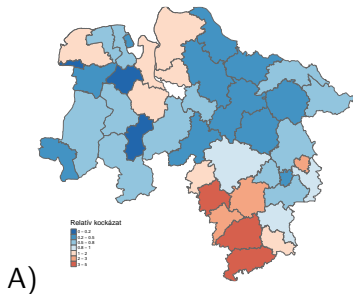
A kockázat-térképezésünk eddigi megközelítései a kétdimenziós simítástól eltekintve nem valódi térbeli modellek, mivel az események egymástól való földrajzi távolságát/közelségét nem kezelik. Eddig a térképeket csak vizualizációra használtuk, de a becslésekben azok szerkezetének nem volt szerepe.

Azonban könnyen belátható, hogy legalább a közvetlen szomszédságban lévő járásokban megfigyelt *E. multilocularis*-fertőzöttségi kockázatok nem lehetnek teljesen függetlenek egymástól. Gondoljunk csak arra, hogy a közigazgatási határok a legtöbb esetben nem jelentenek fizikai korlátot, illetve nem esnek egybe olyan természeti határokkal, amelyek a rókák mozgását akadályoznák.

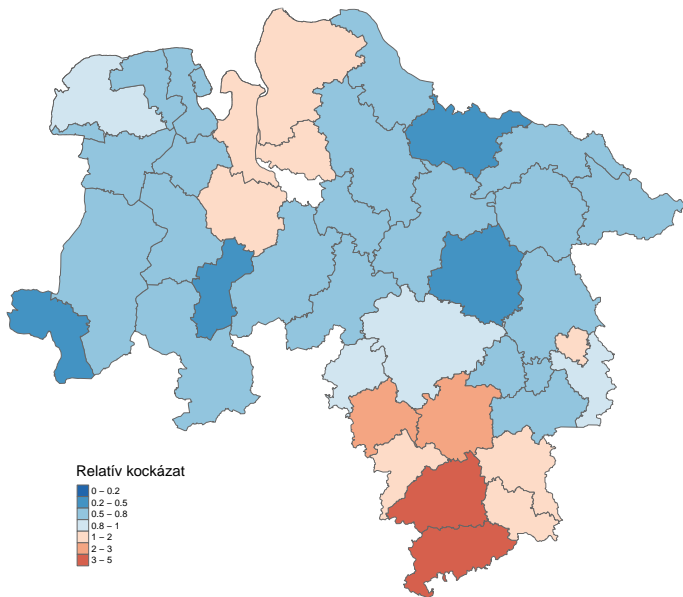
Így könnyen lehet, hogy az egyik járásban kilőtt róka territóriumra más járásra is kiterjed, vagy valami kivételes okból ment át egyik járásból a másikba. Illetve lehetséges, hogy a fertőződése nem abban a járásban történt, amelyikben regisztrálták az esetet. Ennek következtében a kockázatbecsléseink torzítottak lehetnek (ecological bias).



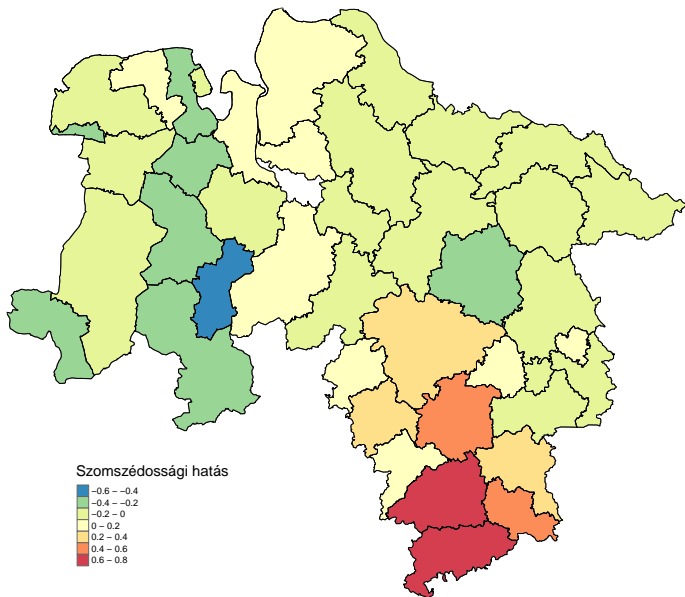
Szomszédossági térkép. A poligonok középpontjai közti kék egyenesek jelzik azt, hogy a szomszédsági struktúrában a poligonok szomszédosak.



- A) Poisson-gamma modell
- B) log-normális modell
- C) log-normális modell, conditional autoregressive (CAR) priorral

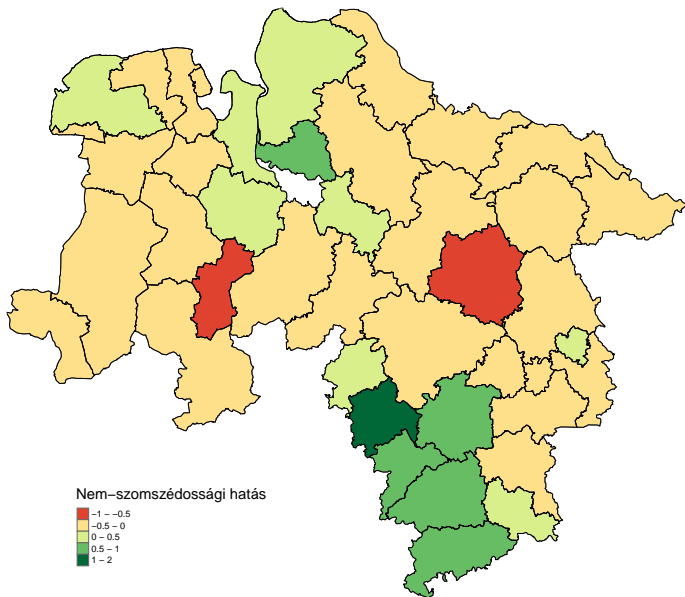


Komplexebb modellek segítségével a **járásonkénti kockázat** felbontható a szomszédsági és a nem-szomszédsági hatásnak tulajdonítható részekre. Ilyen pl. a Besag-York-Mollié modell (Besag et al., 1991)



Komplexebb modellek segítségével a járásonkénti kockázat felbontható a **szomszédsági** és a nem-szomszédsági hatásnak tulajdonítható részekre.





Komplexebb modellek segítségével a járásonkénti kockázat felbontható a szomszédsági és a **nem-szomszédsági** hatásnak tulajdonítható részekre.

A betegség-térképezési példákból láthattuk, hogy a térképi megjelenítést több tényező is erősen befolyásolhatja (pl. pontok mérete, színskála, simítási eljárás, illetve annak paraméterezése).

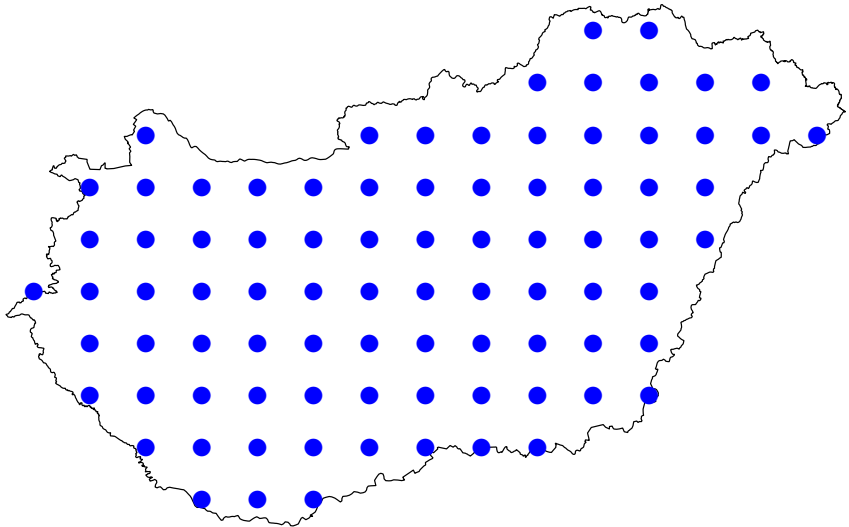
Ezért a kockázat földrajzi mintázatának pusztán vizuális értékelése szubjektív következtetésekhez vezethet.

Ugyanazt a kockázati térképet eltérő módon értelmezhetik különböző személyek.

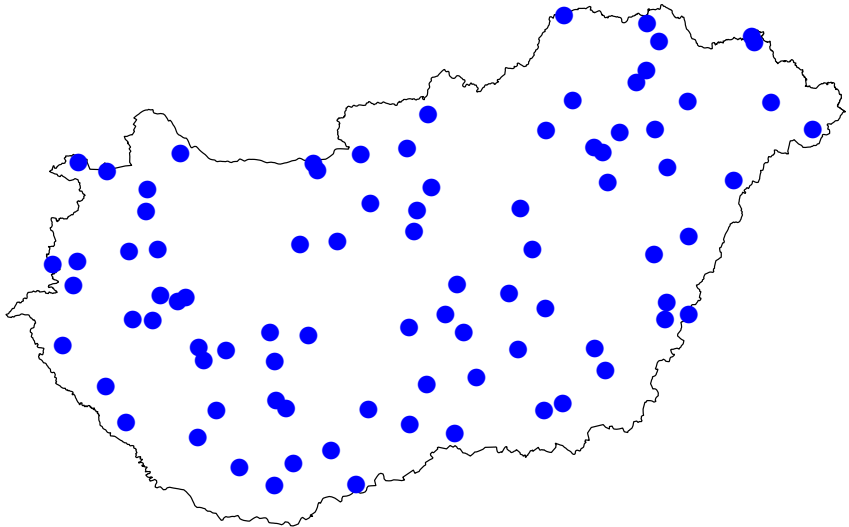
Ezért számos olyan módszert hoztak létre, amelyek a térképen megjelenített kockázatok értékelését objektívebbé teszi.

A geo-epidemiológiában a kockázat földrajzi mintázatának statisztikai módszerek segítségével történő elemzését nevezzük *földrajzi mintázatelemzésnek*.

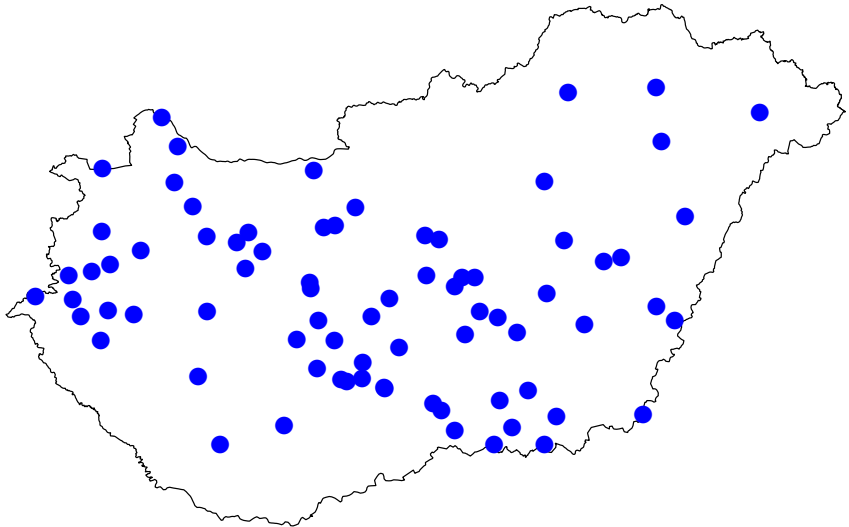
Az epidemiológiában leggyakrabban használt földrajzi *klaszterelemzések* mellett a *határkeresést* mutatjuk be a következőkben.



A geostatisztikában a pontok kétdimenziós térben való elhelyezkedésére vonatkozóan három típust szoktak megkülönböztetni: **szabályos**, véletlenszerű és aggregált.



A geostatisztikában a pontok kétdimenziós térben való elhelyezkedésére vonatkozóan három típust szoktak megkülönböztetni: szabályos, **véletlenszerű** és aggregált.



A geostatisztikában a pontok kétdimenziós térben való elhelyezkedésére vonatkozóan három típust szoktak megkülönböztetni: szabályos, véletlenszerű és **aggregált**.

A klaszterelemzés célja, hogy az esetek, kockázatok földrajzi **halmozódását**, klasztereződését vizsgálja. A CDC (1990) definíciója alapján *klaszternek* nevezzük az egészséggel kapcsolatos események látszólagos vagy valódi csoportosulását térben és/vagy időben.

Az állatorvosi epidemiológiai szakirodalomban számos földrajzi klaszterelemzési módszert találhatunk (Ward and Carpenter, 2000a,b).

A módszereket csoportosítani szokták aszerint, hogy *pontszerű* vagy *területi* adatokra alapozott eljárások. Megkülönböztetnek továbbá *globális*, *lokális* és *fókuszált* klaszterelemzési módszereket.

A **globális** klaszterelemzési módszerekkel az esetek földrajzi halmozódására vonatkozóan egy tesztstatisztika-értékhez juthatunk. Ez az érték arról tájékoztat pusztán, hogy a vizsgált területen az esetek halmozottan fordulnak elő vagy sem.

A **lokális** klaszterezési módszerek lehetővé teszik, hogy ha kimutathatók klaszterek a vizsgálati területen, akkor azok hol vannak.

A **fókuszált** klaszterezési módszerek alkalmazása esetén egy előre meghatározott földrajzi helyre vonatkozóan kapunk halmozódási információt.

A **globális** klaszterelemzés két gyakran használt módszere a **Moran  $I$**  és a **Geary  $c$  autokorrelációra** alapozott eljárások.

Az autokorrelációs módszerek a megfigyelési egységek közötti **szomszédsági** viszonyt, illetve a **vizsgált változó**beli hasonlóságot elemzik.

Területi elrendeződés	Geary $c$	Moran $I$
Aggregált	$0 \leq c < 1$	$I > 0$
Szabályos	$1 < c < 3$	$I < 0$
Véletlenszerű	$c = 1$	$I = 0$

*E. multilocularis* prevalencia térbeli halmozódása?

Területi elrendeződés	Geary $c$	Moran $I$
Aggregált	$0 \leq c < 1$	$I > 0$
Szabályos	$1 < c < 3$	$I < 0$
Véletlenszerű	$c = 1$	$I = 0$

Moran I statistic standard deviate = 5.5582, p-value = 1.363e-08  
 alternative hypothesis: greater  
 sample estimates:

Moran I statistic	Expectation	Variance
0.54047145	-0.02439024	0.01032791

Geary C statistic standard deviate = 4.5358, p-value = 2.87e-06  
 alternative hypothesis: Expectation greater than statistic  
 sample estimates:

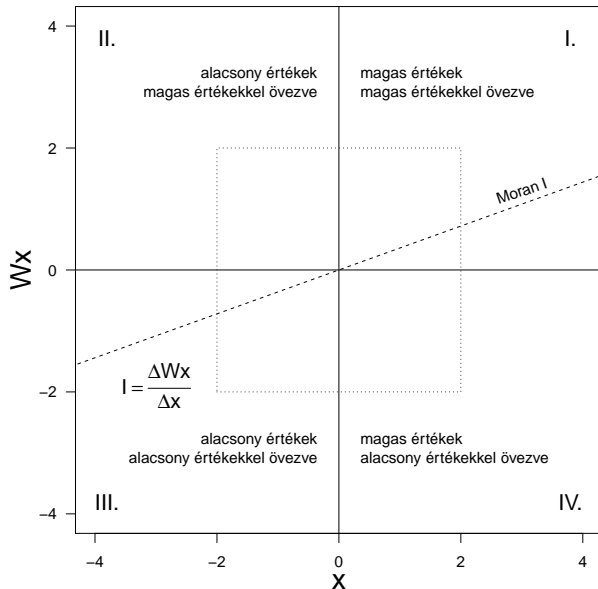
Geary C statistic	Expectation	Variance
0.43313436	1.00000000	0.01561929



Míg a *globális* Moran  $I$  az összes vizsgálati egységre egyetlen teststatisztika értéket és egy  $p$ -értéket ad, a **lokális** módszer (Anselin, 1992) **minden területi egységre** meghatározza ezeket, külön-külön.

	$I_i$	$E.I_i$	$Var.I_i$	$Z.I_i$	$Pr(z > 0)$
23041	0.254596615	-0.02439024	0.2057601	0.61503973	2.692642e-01
23042	0.086809574	-0.02439024	0.1094258	0.33615855	3.683757e-01
23043	0.003763955	-0.02439024	0.2057601	0.06206726	4.752546e-01
23044	6.562042766	-0.02439024	0.4305401	10.03791459	5.192317e-24
23045	0.021121822	-0.02439024	0.2057601	0.10033350	4.600398e-01
23046	4.111202504	-0.02439024	0.1608041	10.31309944	3.074288e-25
.					
.					
.					

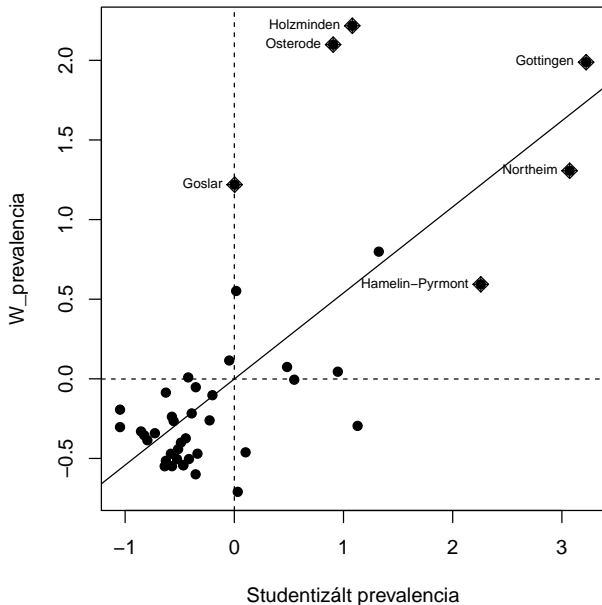
- $I_i$  a lokális Moran  $I$  teststatisztikája az egyes járásokban
- $Pr(z > 0)$  a  $p$ -érték



Moran-szórásdiagram.

Az  $x$ -tengelyen az egységeknél, az  $y$ -tengelyen pedig a szomszédaikban megfigyelt standardizált prevalenciát ábrázoljuk.

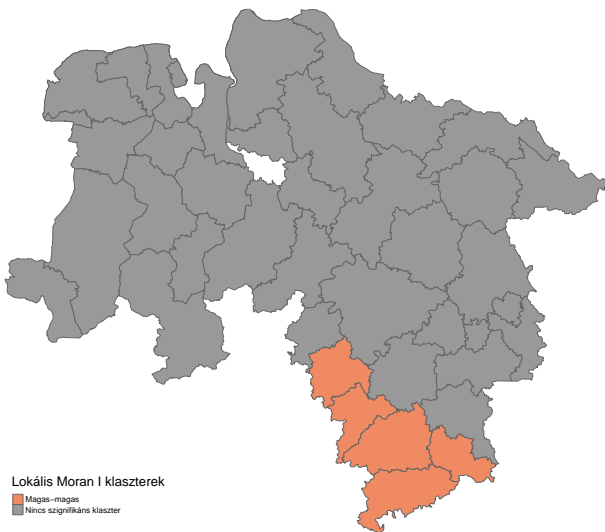
A  $Wx \sim x$  lineáris modell regressziós együtthatója a globális Moran  $I$ -vel egyezik meg.



Moran-szórásdiagram.

Az I. kvadrátban lévő négyzetek azokat a járásokat jelzik, amelyekben (és környezetükben is) magas prevalenciát figyeltünk meg.

Valamely tengelyen kettőnél magasabb értékkel rendelkező klaszter-elemek szignifikánsak ( $p < 0.05$ ).



Göttingen,  
Hamelin-Pyrmont,  
Holzminden,  
Northeim,  
Osterode járások olyan  
magas prevalenciájú  
járások, amelyek  
környezetében is magas  
prevalenciájú járások  
helyeződnek.

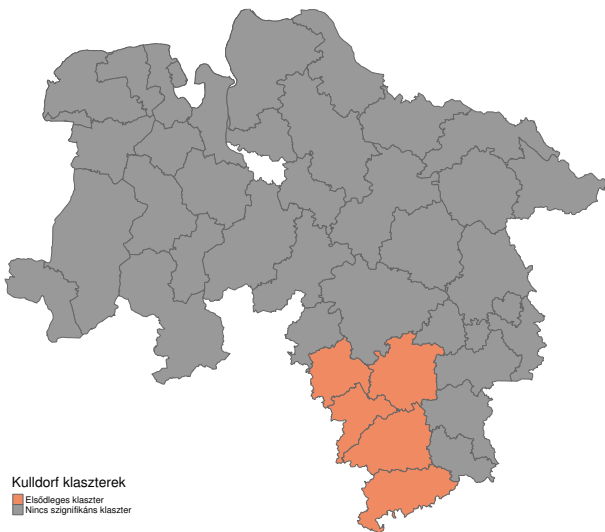
Számos ún. szken-statisztikát használnak földrajzi és/vagy időbeli klaszterek lokalizációjára. Az epidemiológiában talán leggyakrabban alkalmazott szken-statisztika a **Kulldorff-féle** megoldás (Kulldorff and Nagarwalla, 1995).

A vizsgált területre egy **rácsot** illesztünk, aminek rácspontjain végigmenve mindegyikből, mint középpontból, köröket képzünk.

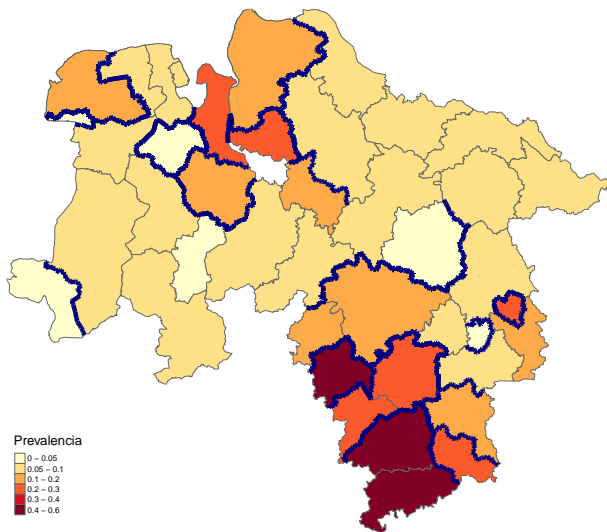
A **körök sugarát** folyamatosan növeljük, aminek felső korlátját általában úgy állapítják meg, hogy az azzal kialakított kör területe ne legyen nagyobb, mint az egész vizsgálati terület nagyságának a fele.

Minden létrehozott körhöz összeszámoljuk, hogy a **körön belül és kívül hány eset** fordul elő, illetve kiszámoljuk a körön belül és kívül várható esetek számát.

A megfigyelt esetszámokat **összevetjük** elméleti eloszlások alapján várható esetszámokkal, és azt a kört, amiben az esetek száma a körön kívüli esetszámokhoz viszonyítva a legkisebb valószínűségű, azonosítjuk, mint elsődleges klasztert.



A Kulldorff-eljárással a következő járások alkotnak magas prevalenciájú klasztert: Göttingen, Hamelin-Pyrmont, Holzminden, Lüneburg, Northeim.



Míg a klaszterelemzések arra a kérdésre keresik a választ, hogy a kockázat vonatkozásában földrajzilag meg lehet-e állapítani halmozódásokat, a határkeresési megközelítések a szomszédok közti különbségek alapján húznak meg határokat (Solymosi et al., 2005).

Monmonier (1973) legnagyobb különbségségek módszerével létrehozott határok (kék).

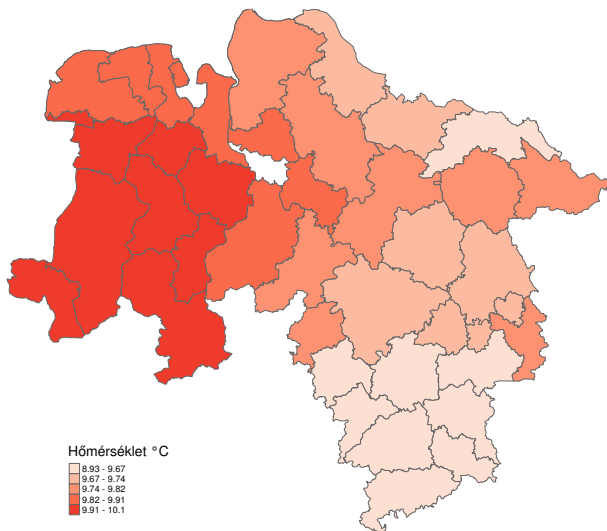
A betegség-térképezés és mintázatelemzés alapján egyértelműnek tűnik, hogy Alsó-Szászország délnyugati járásaiban jelentősen magasabb a fertőzöttség kockázata, mint a tartomány nagy részében. Állat-, és humán-egészségügyi szempontból is fontos lenne tudni, hogy ez a kockázattöbblet miből származhat.

A *E. multilocularis* fejlődésmenetében van olyan szakasz, amikor a pete közvetlenül ki van téve külső környezeti tényezőknek. Így számos, a fertőzöttségi kockázatot befolyásoló tényező között a külső környezeti feltételeknek is szerepe lehet.

Szakirodalmi adatok alapján a parazita fejlődését befolyásolhatja a környezeti hőmérséklet, a csapadék mennyisége, valamint a talaj nedvessége.

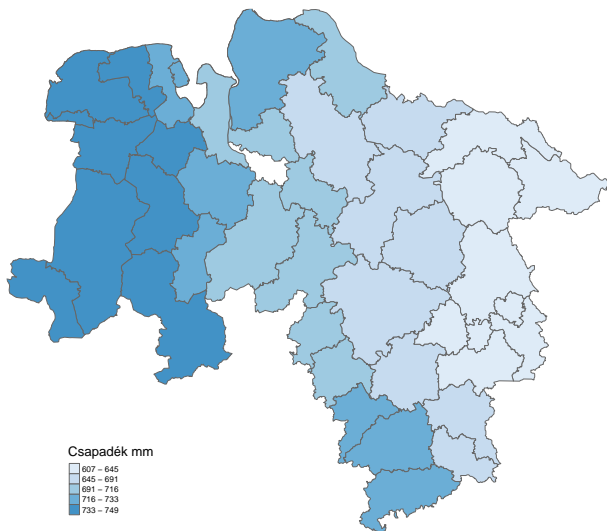
Az alsó-szászországi járásonkénti prevalenciának ezektől a környezeti változótól való függésének elemzéséhez ezeket be kell szerezni.





Az 1991–1997. időszakban az évi átlaghőmérsékletek átlaga az ERA-Interim reanalízis adatbázis alapján.

A legalacsonyabb átlaghőmérsékletű járások nagy része a legmagasabb *E. multilocularis* prevalenciájú járásokkal azonos.



Az 1991–1997. időszakban az éves összes csapadék átlaga az ERA-Interim reanalízis adatbázis alapján.

A térkép alapján a legmagasabb prevalenciájú járások nem a legtöbb csapadékot kapó járások. De ha a hőmérsékleti térképpel is összevetjük, akkor azt láthatjuk, hogy a legtöbb csapadékot kapó járásokban az átlaghőmérséklet is a legmagasabb tartományba esik.

A ESDAC-adatbázisból  
származó hidrogeológiai  
típusok mintázata  
Alsó-Szászországban:



- ❶ áteresztő altalajú talaj,  
a talajvíz mélyen van:  
ritkán nedves
- ❷ alföldi talaj, amire a  
talajvíznek hatása van,  
szezonálisan vagy  
folyamatosan nedves,  
vagy mesterségesen  
csatornázott
- ❸ olyan talaj, amiben 80  
cm-en belül vízzáró  
réteg van, szezonálisan  
vagy folyamatosan  
nedves
- ❹ felföldi vagy hegyvidéki  
talaj

Az eddigiekben **szubjektív, vizuális** alapon fogalmaztunk meg lehetséges összefüggéseket a fertőzöttség kockázata és néhány környezeti tényező között.

Azonban ez csak a bevezetése az **összefüggések matematikai statisztikai** leírásának. Ebben a megközelítésben a függőváltozónak (pl. prevalencia) a független változóktól (pl. hőmérséklet, csapadék, talajnedvesség) való függését valamilyen matematikai modell segítségével írjuk le.

Ezek a modellek egyrészt lehetővé teszik, hogy **számszerűsítsük** a független változók hatásának mértékét a függő változóra, másrészt ezeknek a hatásoknak a számszerű értékei lehetővé teszik, hogy predikciókat adhassunk.

Mit jelent a geo-epidemiológiában a **predikció**; Egyrészt jelenti azt, hogy a modellünk azokra a helyekre (pl. járások) vonatkozóan, amelyekből van információnk a függő változóra (pl. prevalencia) milyen függőváltozó értékeket becsül, prediktál.

Másrészt, mivel azt hisszük, hogy a fizikai világban vannak szabályok és ezek a szabályok matematikailag leírhatók, azt is hisszük, hogy ezek a szabályok nem csak egy adott helyen érvényesek, hanem más területekre is **kiterjeszthetők**.

A célunk leírni egy olyan modellt, amelyből származó **predikció a legkisebb eltérést** mutatja a megfigyelt értékektől.

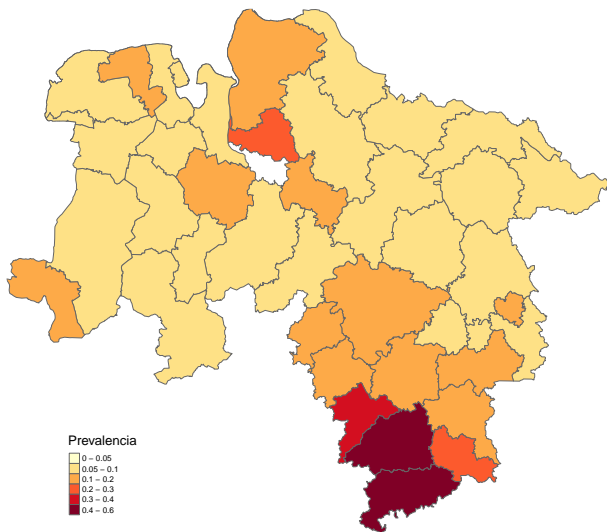
Ehhez a három független változónkból (hőmérséklet, csapadék, talajnedvesség) az összes lehetséges kombináció szerint létrehozunk modelleket és azokat **illesztjük**.

Minden illesztett modell alapján **prediktáljuk** az egyes járásokban várható prevalenciát.

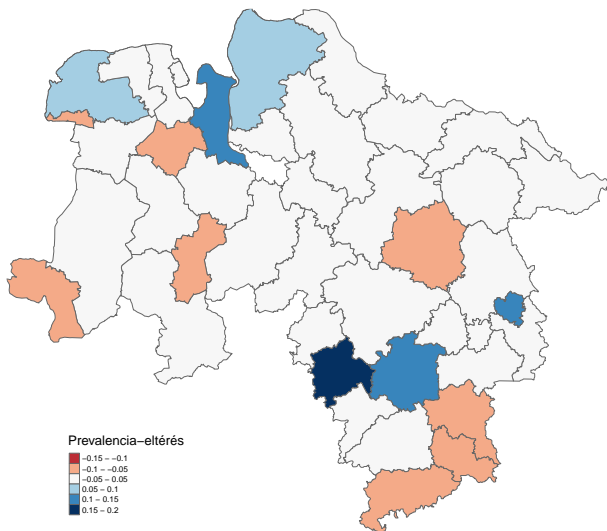
Ezen prediktált prevalenciák felhasználásával mindegyik modellre kiszámítjuk a **megfigyelt és prediktált értékek különbségét**. Erre a mean absolute error (MAE) használjuk:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i|$$

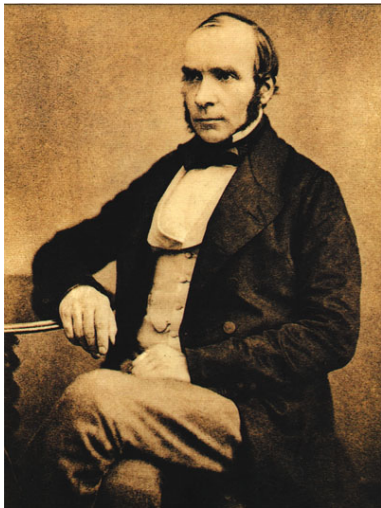
Azt a modellt tekintjük a **legjobb**nak, amelyik a **legkisebb MAE**-értéket adó prediktált prevalenciát eredményezte.



A legkisebb MAE-értékű  
modell alapján  
prediktált, járásonkénti  
prevalencia.

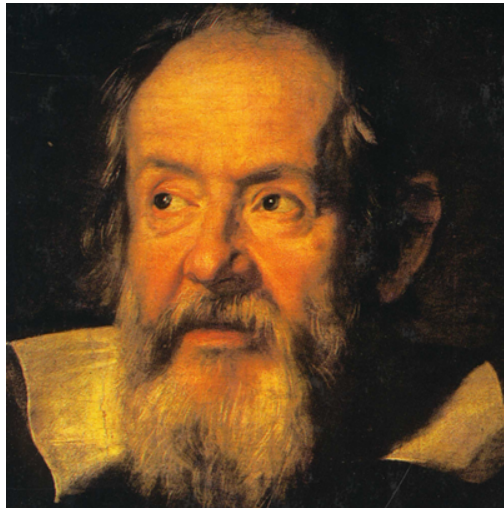


A megfigyelt és a prediktált prevalencia különbsége. A felülprediktált járások a piros skálával, az alulprediktáltak pedig a kék skálával vannak színezve.



John Snow (1813–1858)

A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system (Glasziou et al., 2008).



Galileo Galilei (1564–1642)



- Anselin, L. (1992). Local indicators of spatial association – LISA. *Geogr. Anal.* 27(2), 93–115.
- Berke, O. (2001). Choropleth mapping of regional count data of *Echinococcus multilocularis* among red foxes in Lower Saxony, Germany. *Preventive veterinary medicine* 52(3), 119–131.
- Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43(1), 1–20.
- CDC (1990). Guidelines for investigating clusters of health events. Technical Report RR-11.
- Glasziou, P., A. Burls, and R. Gilbert (2008). Evidence based medicine and the medical curriculum. *BMJ (Clinical Research Ed.)* 337(7672), 704–705.
- Kulldorff, M. and N. Nagarwalla (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine* 14, 799–810.
- Lawson, A. B. (2013a). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.
- Lawson, A. B. (2013b). *Statistical methods in spatial epidemiology*. John Wiley & Sons.
- Lawson, A. B., S. Banerjee, R. P. Haining, and M. D. Ugarte (2016). *Handbook of spatial epidemiology*. CRC Press.
- Monmonier, M. S. (1973). Maximum-difference barriers: An alternative numerical regionalization method. *Geogr. Anal.* 3, 245–261.
- Solymosi, N., J. Reiczigel, A. Harnos, J. Mészáros, L. D. Molnár, and F. Rubel (2005). Finding spatial barriers by Monmonier's algorithm. In *International Society for Clinical Biostatistics Conference*, Szeged, Hungary.
- Ward, M. P. and T. E. Carpenter (2000a). Analysis of time-space clustering in veterinary epidemiology. *Prev. Vet. Med.* 43(4), 225–237.
- Ward, M. P. and T. E. Carpenter (2000b). Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev. Vet. Med.* 45, 257–284.