

10 años
CEDEUS



The book Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny

Paula Moraga, 2019
Published by Chapman & Hall/CRC Biostatistics Series

Alejandra Vega
6 noviembre 2023

Capítulo 5 Areal data

Capítulo 6 Spatial modeling of areal data. Lip cancer in Scotland

Geospatial health data and
INLA

1 **Geospatial health**

2 Spatial data and R packages
for mapping

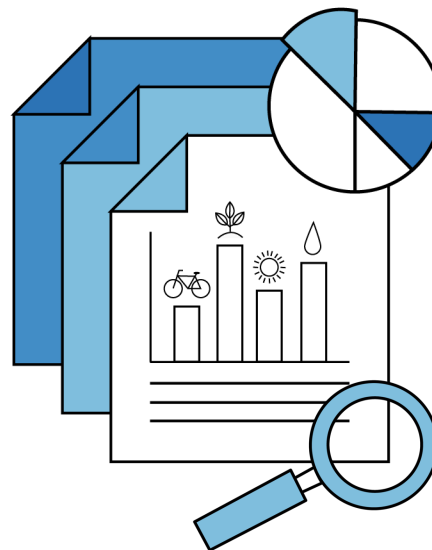
3 Bayesian inference and INLA

4 The R-INLA package

Modeling and visualization

5 Areal data

6 Spatial modeling of areal
data. Lip cancer in Scotland



$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}.$$

Here, Z denotes the attribute we observe, for example, the number of sudden infant deaths or the level of rainfall, and \mathbf{s} refers to the location of the observation. Cressie (1993) distinguishes three basic types of spatial data through characteristics of the domain D , namely, areal data, geostatistical data, and point patterns.

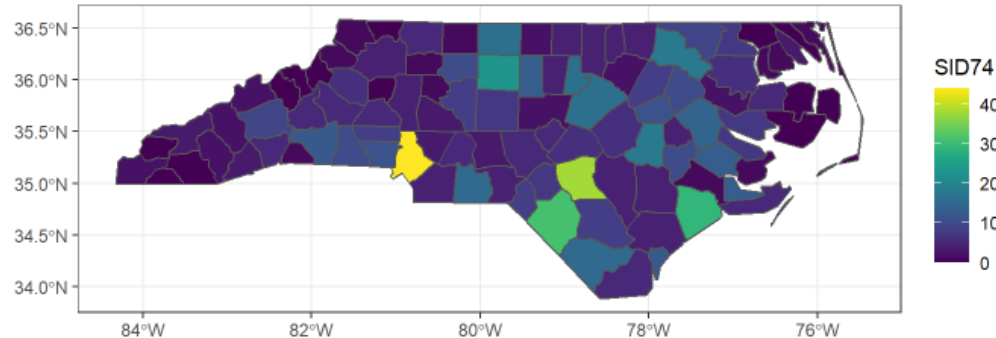
Areal data

Geostatistical data

Point patterns

Datos areales o reticulares

- Dominio dividido en subregiones que agregan los resultados
- Ejemplos: número de casos de cáncer en condados, el número de accidentes de tráfico en provincias y la proporción de personas que viven en la pobreza en secciones censales
- Los modelos de riesgo de enfermedad pretenden obtener estimaciones del riesgo de enfermedad dentro de las mismas áreas



Sudden infant deaths in North Carolina in 1974

- SIR standardized incidence ratio (la relación entre los recuentos observados y los esperados)
- Es preferible estimar el riesgo de enfermedad utilizando modelos jerárquicos bayesianos que permiten tomar prestada información de zonas vecinas e incorporar información sobre covariables, lo que da lugar a la suavización o reducción de los valores extremos

Aprenderemos a calcular matrices de vecindad, recuentos esperados y SIR

ajustar modelos de riesgo de enfermedad espaciales y espaciotemporales utilizando el paquete R-INLA

Ann. Inst. Statist. Math.
Vol. 43, No. 1, 1-59 (1991)

BAYESIAN IMAGE RESTORATION,
WITH TWO APPLICATIONS IN SPATIAL STATISTICS* **

JULIAN BESAG^{1***}, JEREMY YORK¹ AND ANNIE MOLLIE²

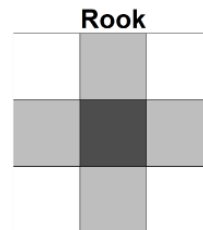
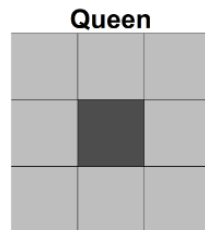
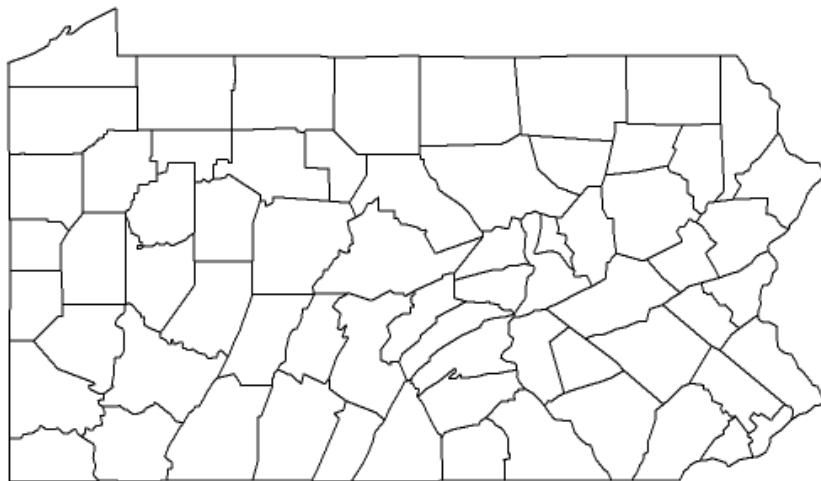
¹Department of Statistics GN-22, University of Washington,
Seattle, WA 98195, U.S.A.

²Institut Gustave Roussy, INSERM U287, 94805 Villejuif Cedex, France

(Received November 7, 1989; revised June 18, 1990)

Vecindad espacial o matriz de proximidad

```
library(SpatialEpi)  
map <- pennLC$spatial.polygon  
plot(map)
```



Vecindad espacial o matriz de proximidad

```
class(map)
```



```
library(spdep)  
nb <- poly2nb(map)  
head(nb)
```



Continuar en R....

A	B	
C	D	E

	A	B	C	D	E	Sum
A	0	1	1	1	0	3
B	1	0	1	1	1	4
C	1	1	0	1	0	3
D	1	1	1	0	1	4
E	0	1	0	1	0	2

- Estimación de riesgo
- la relación entre los recuentos observados (Y_i) y los recuentos esperados (E_i)

$$SIR_i = Y_i / E_i.$$

- Los recuentos previstos E_i representan el número total de casos que cabría esperar si la población del área i se comportara como la población estándar (o regional). E_i puede calcularse utilizando la estandarización indirecta como

$$E_i = \sum_{j=1}^m r_j^{(s)} n_j^{(i)}, \quad E_i = r^{(s)} n^{(i)},$$

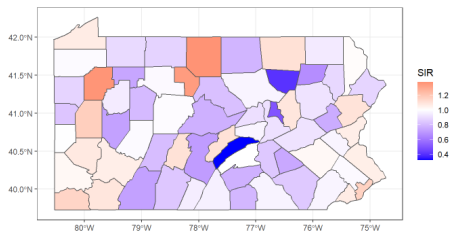
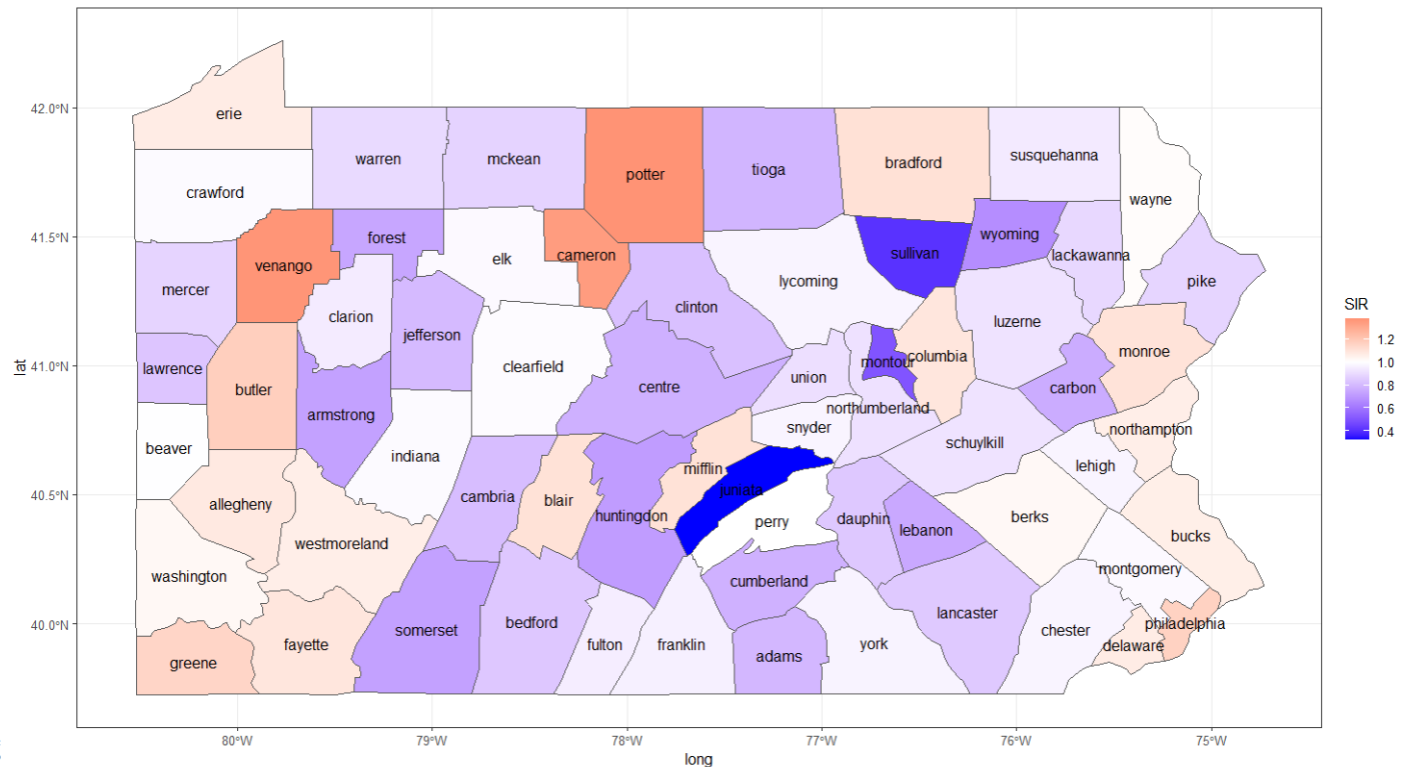
whether area i has higher ($SIR_i > 1$), equal ($SIR_i = 1$) or lower ($SIR_i < 1$) risk.

Standardized incidence ratio

- SIRs of lung cancer in Pennsylvania in 2002 using the data frame `pennLC$data` from the `SpatialEpi` package
- number of lung cancer cases and the population of Pennsylvania at county level, stratified on race (white and non-white), gender (female and male) and age (under 40, 40-59, 60-69 and 70+). We obtain the number of cases for all the strata together in each county, Y , by aggregating the rows of `pennLC$data` by county and adding up the number of cases. We can do this using the functions `group_by()` and `summarize()` of the `dplyr` package

Continuar en R....

Standardized incidence ratio



- Aunque los SIR pueden ser útiles en algunos contextos, en regiones con poblaciones pequeñas o enfermedades raras los recuentos esperados pueden ser muy bajos y los SIR pueden ser engañosos e insuficientemente fiables para la notificación. Por lo tanto, es preferible estimar el riesgo de enfermedad **utilizando modelos que permitan tomar prestada información de zonas vecinas e incorporar información sobre covariables que dé lugar a la suavización o reducción de los valores extremos sobre la base de muestras de pequeño tamaño.**

- Normalmente, los recuentos observados Y_i en el área i se modelizan mediante una distribución de Poisson con media $E_i\theta_i$ donde E_i son los recuentos esperados y θ_i es el riesgo relativo en el área i .

$$Y_i \sim Po(E_i\theta_i), i = 1, \dots, n,$$

$$\log(\theta_i) = \alpha + u_i + v_i.$$

- El logaritmo del riesgo relativo θ_i se expresa como la suma de un intercepto que modela el nivel de riesgo global de la enfermedad, y efectos aleatorios para tener en cuenta la variabilidad extra-Poisson. El riesgo relativo θ_i cuantifica si el área i tiene mayor ($\theta_i > 1$) o inferior ($\theta_i < 1$) que el riesgo medio de la población estándar. Por ejemplo, si $\theta_i = 2$ significa que el riesgo de la zona i es el doble del riesgo medio de la población estándar.

$$\log(\theta_i) = d_i\beta + u_i + v_i,$$

Spatial small area disease risk estimation

- Un modelo espacial popular en las aplicaciones de cartografía de enfermedades es el modelo Besag-York-Mollié (BYM)
- En este modelo, el efecto aleatorio espacial u_i se le asigna una distribución Autorregresiva Condicional (CAR) que suaviza los datos de acuerdo con una cierta estructura de vecindad que especifica que dos áreas son vecinas si comparten un límite común. En concreto,

$$u_i | \mathbf{u}_{-i} \sim N \left(\bar{u}_{\delta_i}, \frac{\sigma_u^2}{n_{\delta_i}} \right),$$

- representan, respectivamente, el conjunto de vecinos y el número de vecinos del área i . El componente no estructurado se modela como variables normales independientes e idénticamente distribuidas con media cero y varianza σ_v^2 .

Continuar en R...In R-INLA, the formula of the BYM model is specified as follows. Spatial modeling of lung cancer in Pennsylvania

```
formula <- Y ~  
  f(idareau, model = "besag", graph = g, scale.model = TRUE) +  
  f(idareav, model = "iid")
```

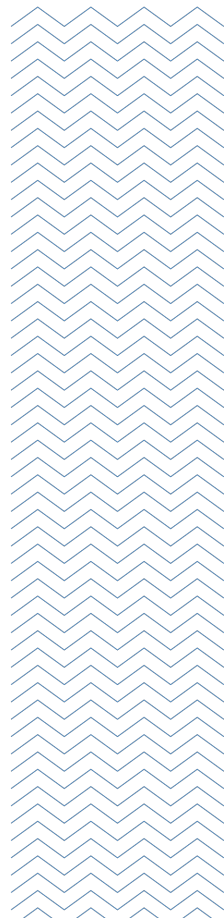
```
formula <- Y ~ f(idarea, model = "bym2", graph = g)
```

- Spatio-temporal small area disease risk estimation (Capitulo 7)
- Issues with areal data.



Spatial modeling of areal data. Lip cancer in Scotland

Continuar en R....



6.4.1 Model

We specify a model assuming that the observed counts, Y_i , are conditionally independently Poisson distributed:

$$Y_i \sim \text{Poisson}(E_i \theta_i), \quad i = 1, \dots, n,$$

where E_i is the expected count and θ_i is the relative risk in area i . The logarithm of θ_i is expressed as

$$\log(\theta_i) = \beta_0 + \beta_1 \times AFF_i + u_i + v_i,$$

where β_0 is the intercept that represents the overall risk, β_1 is the coefficient of the AFF covariate, u_i is a spatial structured component modeled with a CAR distribution, $u_i | \mathbf{u}_{-i} \sim N\left(\bar{u}_{\delta_i}, \frac{\sigma_u^2}{n_{\delta_i}}\right)$, and v_i is an unstructured spatial effect defined as $v_i \sim N(0, \sigma_v^2)$. The relative risk θ_i quantifies whether area i has higher ($\theta_i > 1$) or lower ($\theta_i < 1$) risk than the average risk in the standard population.

Continuar en R....

6.6 Exceedance probabilities

We can also calculate the probabilities of relative risk estimates being greater than a given threshold value. These probabilities are called exceedance probabilities and are useful to assess unusual elevation of disease risk. The probability that the relative risk of area i is higher than a value c can be written as $P(\theta_i > c)$. This probability can be calculated by subtracting $P(\theta_i \leq c)$ to 1 as follows:

$$P(\theta_i > c) = 1 - P(\theta_i \leq c).$$

In **R-INLA**, the probability $P(\theta_i \leq c)$ can be calculated using the `inla.pmarginal()` function with arguments equal to the marginal distribution of θ_i and the threshold value c . Then, the exceedance probability $P(\theta_i > c)$ can be calculated by subtracting this probability to 1:

```
1 - inla.pmarginal(q = c, marginal = marg)
```



Continuar en R....