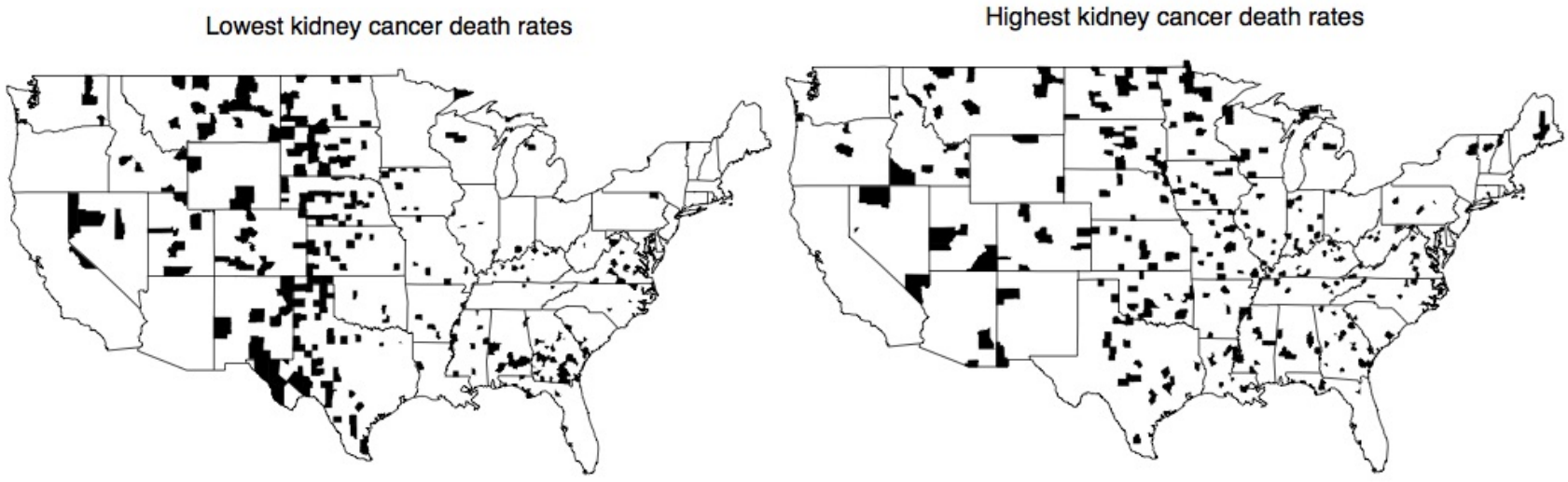
Spatial Models in Stan

Intrinsic Auto-Regressive Models for Areal Data

Mitzi Morris
Stan Developer, Columbia University

Areal data

Areal data consists of a single aggregated measure per areal unit, which may be a binary, count, or continuous value.



- Event counts for low-population counties display greater variance - noisy data acts like scratches, dirt on an image
- Spatial smoothing removes noise by borrowing information from *neighboring* regions; allows big picture trends to be seen more clearly

Areal data, continued

Areal units:

- partition a multi-dimensional volume D into a finite number of sub-volumes with well-defined boundaries
- the set of areal units is fixed

Geospatial data in R:

- Shapefile format: contains geometric locations (points, lines, and polygons) and associated attributes.
- R package `spdep` provides functions to compute and extract neighbor relationships from shapefile data.

Data structures for encoding neighbor relationship

$N \times N$ Adjacency matrix:

- entries {i,j} and {j,i} are 1 when regions n_i and n_j are neighbors, zero otherwise

Undirected graph:

- regions are vertices
- pairs of neighbors are edges

For sparse matrices, graph representation is more efficient (memory, I/O, processing)

Neighbor releationship is an abstraction.

Common metric: shared boundary - line (rook) or point (queen).

For large maps, this is (generally) very sparse.

Example: NYC census tract data used for Stan Case Study.

- 1921 census tracts: 3,690,241 cells in a 1921 X 1921 matrix, of which 3,679,319 are zero.
- representation as an edgeset: 10,922 cells (2 parallel arrays, 5461 edges), i.e., just the non-zero entries.

Graph of adjoining census tracts for Brooklyn

Algorithmic complexity of Operations on Matrices

Big-O notation classifies algorithms according to how processing time and/or memory usage grows as the size of the input grows.

Complexity of the following operations on an $N \times N$ matrix:

- multiplication
- inversion
- determinant

$\mathbf{O}(N^3)$

Space required for a full $n \times n$ matrix: $\mathbf{O}(n^2)$

Space required for a sparse $n \times n$ matrix as an edgeset: $\mathbf{O}(k\,n)$ where k is the average fan-out (neighbers per node).

Conditional Auto-Regressive (CAR) models, Besag 1974

- data consists of per-region observations for area with n fixed regions
- random effects for each region is conditional only on neighboring regions

Use an $n \times n$ adjacency matrix W to specify neighborhood structure:

- entries $w_{i,j}$ and $w_{j,i}$ are positive when regions n_i and n_j are neighbors, zero otherwise
- *neighbor* relationship $i \sim j$ - the neighbors of a region i are the regions which have non-zero entries in row or column i of W

Conditional specification of the CAR model

Model spatial interactions as an n-length vector $\phi = (\phi_1, \ldots, \phi_n)^T$, where each ϕ_i is a normal random variate with a mean which is conditional on the values of its neighbors and unknown variance:

$$\phi_i \mid \phi_j, j \neq i \sim Normal(\sum_{j=1}^{n} w_{ij}\phi_j, \sigma)$$

ϕ is a Gaussian Markov Random Field (GMRF)

Joint specification of the CAR model

$$\phi \sim Normal\left(0, [\tau D(I - \alpha W)]^{-1}\right)$$

ϕ is a multi-variate normal, with mean 0 and covariance matrix $[D(I - \alpha W)]^{-1}$

- τ is the precision (inverse variance)
- D is an $n \times n$ diagonal matrix where $D_{i,i}$ is the number of neighbors for region w_i and all other entries are 0.
- I is an $n \times n$ identity matrix.
- α is a parameter between 0 and 1 which controls the amount of spatial correlation, 0 corresponds to spatial independence and 1 corresponds to complete spatial correlation.
- W is the $n \times n$ adjacency matrix

$$\phi \sim \mathit{Normal}\left(0, [\tau D(I - \alpha W)]^{-1}\right)$$

When $0 < \alpha < 1$, the covariance matrix is *positive definite*.

BUT computing the log probability density requires taking the determinant of the covariance matrix.

When $\alpha = 1$, the covariance matrix is singular, improper distribution - can only be used as a prior

Additional constraint needed to provide centering: e.g. $\sum_i \phi_i = 0$.

Intrinsic Conditional Auto-Regressive (ICAR) Models, Besag, 1974

Spatial interactions for fixed set of N areal units is N-length vector

$$\phi = (\phi_1, \ldots, \phi_n)^T$$

Neighborhood structure specified by $N \times N$ adjacency matrix W:

- W entries $w_{i,j}$ and $w_{j,i}$ are 1 when regions n_i and n_j are neighbors, zero otherwise

Number of neighbors for each region specified using $N \times N$ diagonal matrix D:

- diagonal elements $d_{i,i}$ contain number of neighbors for region n_i, all other elements zero

Conditional specification:

$$p\left(\phi_i \mid \phi_j \, j \neq i\right) = Normal\left(\frac{\sum_{i \sim j} \phi_i}{d_{i,i}}, \frac{1}{d_{i,i}\tau_i}\right)$$

where τ_i is the precision (inverse variance)

Joint specification:

$$\phi \sim Normal(0, [\tau\,(D - W)]^{-1}).$$

Unit mulitivariate Gaussian: $\tau = 1$, joint distribution rewrites to *pairwise difference* formulation:

$$p(\phi) \propto \exp\left\{-\frac{1}{2}\sum_{i \sim j}\left(\phi_i - \phi_j\right)^2\right\}$$

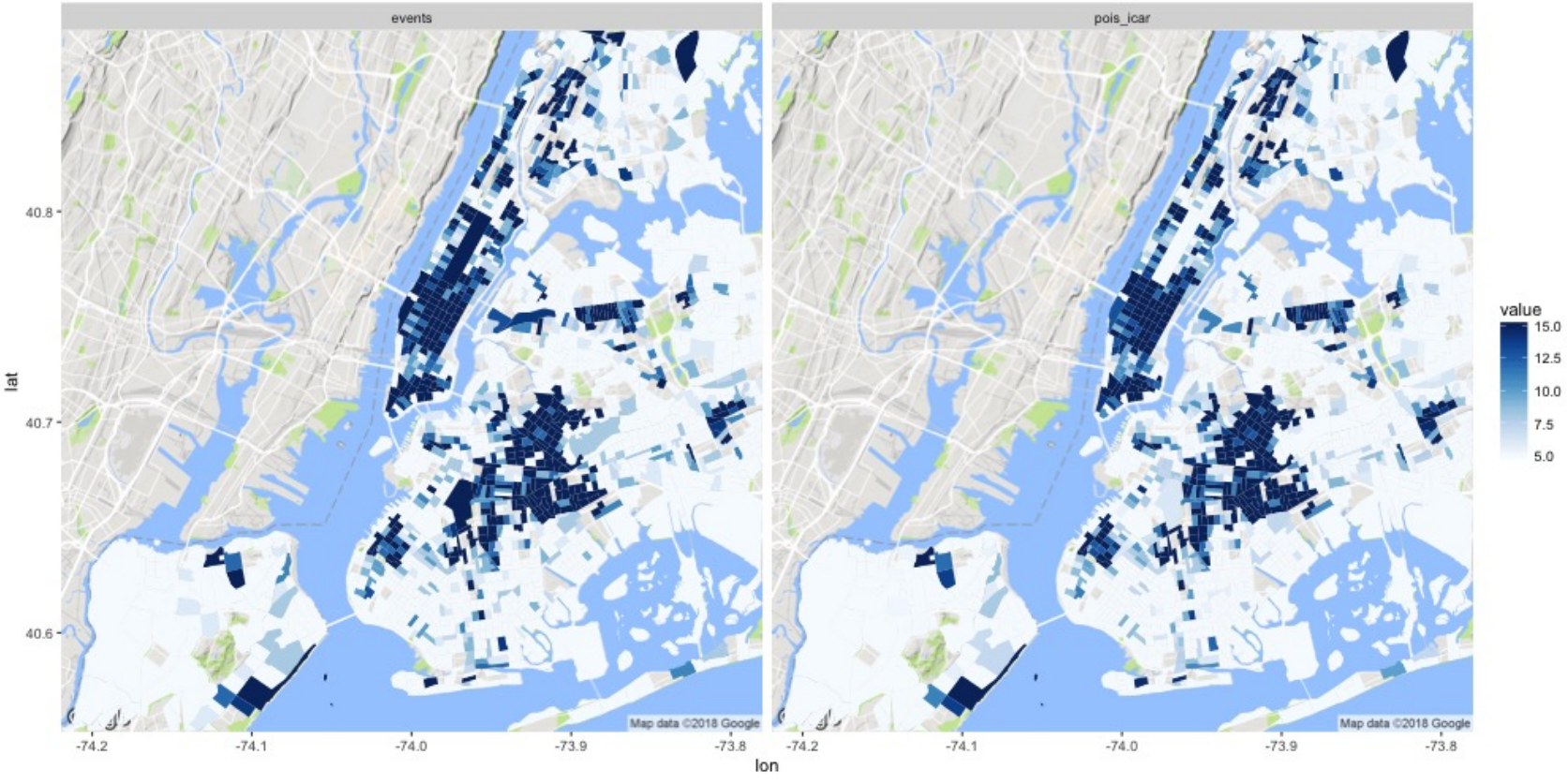
NOTE: ICAR model is non-identifiable, must add the constraint $\sum_i \phi_i = 0$.

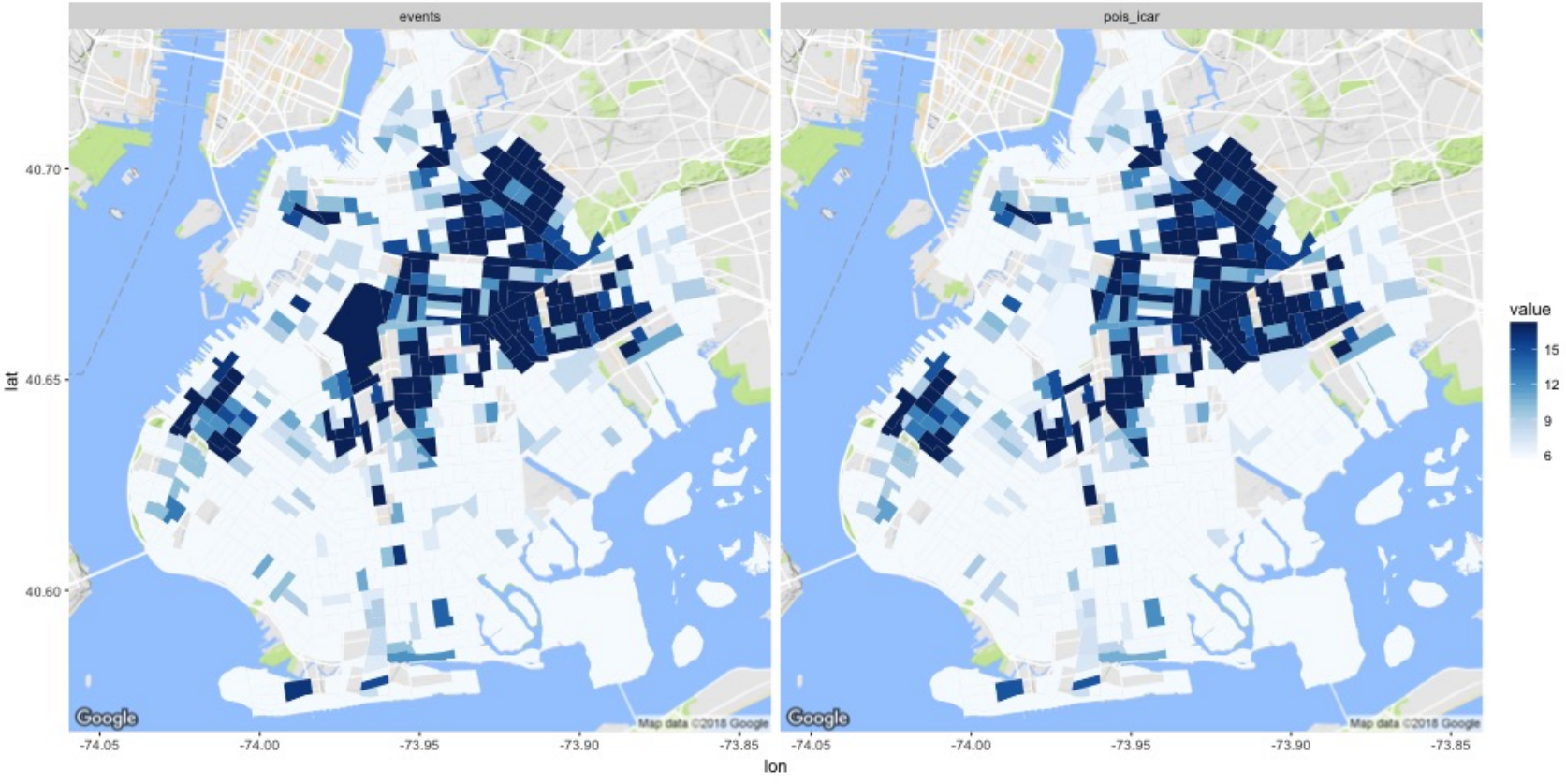
Stan program: ICAR prior, soft sum-to-zero constraint

The sum-to-zero constraint is implemented by putting a prior on phi as follows:

```
data {
  int<lower=0> N;
  int<lower=0> N_edges;
  int<lower=1, upper=N> node1[N_edges];  // node1[i] adj to node2[i]
  int<lower=1, upper=N> node2[N_edges];  // and node1[i] < node2[i]
}
parameters {
  vector[N] phi;
}
model {
  target += -0.5 * dot_self(phi[node1] - phi[node2]);

  // soft sum-to-zero constraint on phi,
  // equivalent to mean(phi) ~ normal(0,0.01)
  sum(phi) ~ normal(0, 0.01 * N);
}
```

```
data {
  int<lower=0> N;   int<lower=0> N_edges;
  int<lower=1, upper=N> node1[N_edges];
  int<lower=1, upper=N> node2[N_edges];
  int<lower=0> y[N];     // count outcomes
  vector<lower=0>[N] E;  // exposure
}
transformed data {
  vector[N] log_E = log(E);
}
parameters {
  real beta0;            // intercept
  vector[N] phi;         // spatial effects
  real<lower=0> sigma;   // variance (non-centered)
}
model {
  y ~ poisson_log(log_E + beta0 + phi * sigma);
  beta0 ~ normal(0.0, 2.5);
  sigma ~ normal(0.0, 5.0);
  target += -0.5 * dot_self(phi[node1] - phi[node2]);
  sum(phi) ~ normal(0, 0.001 * N);  // soft sum-to-zero constraint
}
generated quantities {
  vector[N] mu = exp(log_E + beta0 + phi * sigma);
}
```


events
pois_icar
lat
lon
40.8
40.7
40.6
-74.2
-74.1
-74.0
-73.9
-73.8
Map data ©2018 Google
value
15.0
12.5
10.0
7.5
5.0


events
pois_icar
lat
40.70
40.65
40.60
-74.05
-74.00
-73.95
-73.90
-73.85
lon
Google
Map data ©2018 Google
value
15
12
9
6

Besag York Mollié: Poisson GLM + ICAR + normal RE

$$Y_i \mid \psi_i \sim Poisson(E_i\, e^{\psi_i}),$$

for $i \in 1 : N$, where

$$\psi = x\beta + \theta + \phi$$

and

- x is the matrix of explanatory spatial covariates such that x_i is the vector of covariates for areal unit i. The coefficients β are called "fixed effects."
- θ is an ordinary random-effects components for non-spatial heterogeneity.
- ϕ is an ICAR spatial component.
- strong hyperpriors on ϕ and θ equal for emphasis on both spatial and non-spatial variance (following Bernardinelli et al. 1995)

BYM2: Reibler et al 2016

Combined random effects component:

$$\theta + \phi = \sigma(\sqrt{1-\rho}\theta^* + \sqrt{\rho}\phi^*)$$

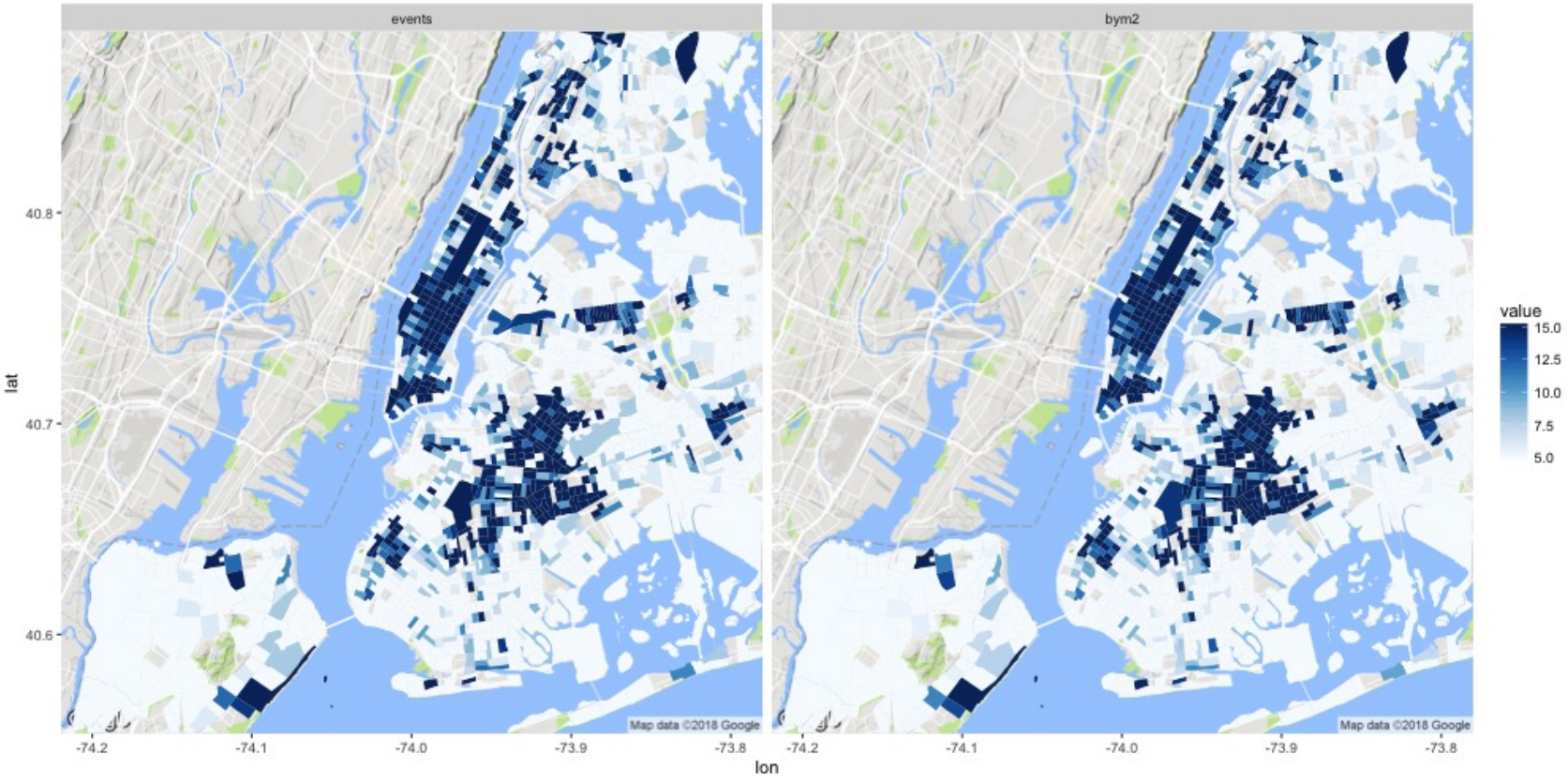
- $\sigma \geq 0$ *overall* standard deviation
- $\rho \in [0, 1]$ models how much of the variance comes from the spatially structured effect and how much comes from the spatially unstructured effect
- $\theta^* \sim N(0, I)$ is the unstructured random effect with fixed standard deviation 1
- ϕ^* is the ICAR model scaled so $\text{Var}(\phi_i) \approx 1$

```
data {
  int<lower=0> N;
  int<lower=0> N_edges;
  int<lower=1, upper=N> node1[N_edges];
  int<lower=1, upper=N> node2[N_edges];

  int<lower=0> y[N];              // count outcomes
  vector<lower=0>[N] E;           // exposure
  int<lower=1> K;                 // num covariates
  matrix[N, K] x;                 // design matrix

  // scales the variance of the spatial effects
  real<lower=0> scaling_factor;
}
transformed data {
  vector[N] log_E = log(E);
}
```

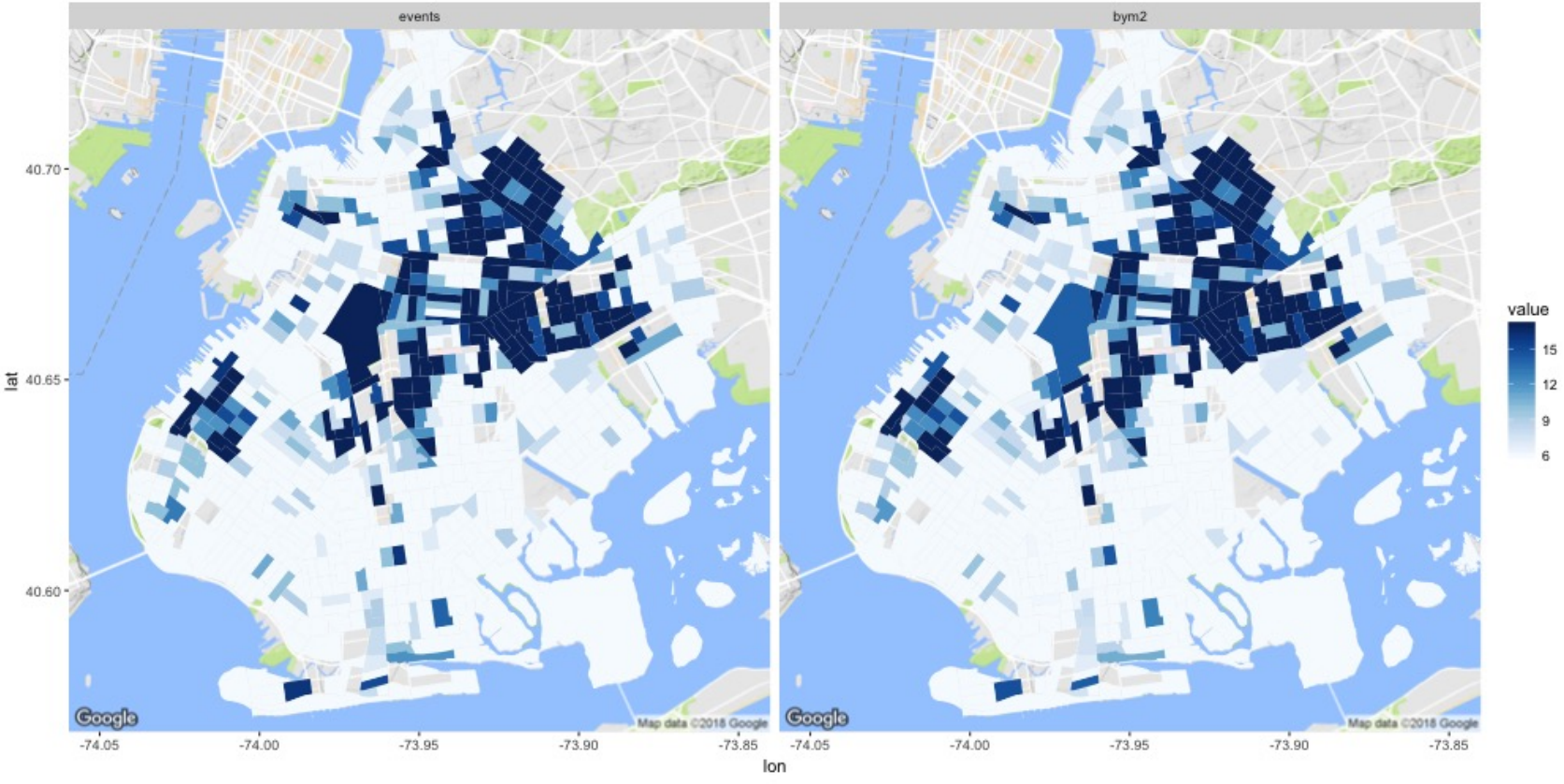
```
parameters {
  real beta0;              // intercept
  vector[K] betas;         // covariates
  real<lower=0> sigma;     // overall standard deviation
  real<lower=0, upper=1> rho;
  vector[N] theta;         // heterogeneous effects
  vector[N] phi;           // spatial effects
}
transformed parameters {
  vector[N] convolved_re;
  convolved_re =  sqrt(1 - rho) * theta
                + sqrt(rho / scaling_factor) * phi;
}
model {
  y ~ poisson_log(log_E + beta0 + x * betas + convolved_re * sigma);
  ...
```

```
model {
  y ~ poisson_log(log_E + beta0 + x * betas + convolved_re * sigma);

  beta0 ~ normal(0.0, 2.0);
  betas ~ normal(0.0, 2.0);
  theta ~ normal(0.0, 1.0);
  sigma ~ normal(0, 2.0);
  rho ~ beta(0.5, 0.5);

  target += -0.5 * dot_self(phi[node1] - phi[node2]);
  sum(phi) ~ normal(0, 0.001 * N);  // soft sum-to-zero constraint
}
generated quantities {
  vector[N] eta = log_E + beta0 + x * betas + convolved_re * sigma;
  vector[N] mu = exp(eta);
}
```

Raw events vs. BYM2 model fit, all NYC census tracts



Raw events vs. BYM2 model fit, Brooklyn



Case Study: "Spatial Models in Stan: Intrinsic Auto-Regressive Models for Areal Data"

http://mc-stan.org/users/documentation/case-studies/icar_stan.html

References

- Banerjee, Sudipto, Alan E. Gelfand, and Bradley P. Carlin. "Hierarchical modeling and analysis for spatial data." (2003).
- Besag, Julian. "Spatial interaction and the statistical analysis of lattice systems." Journal of the Royal Statistical Society. Series B (Methodological) (1974): 192-236.
- Riebler, Andrea, Sigrunn H. Sørbye, Daniel Simpson, and Håvard Rue. "An intuitive Bayesian spatial model for disease mapping that accounts for scaling." Statistical methods in medical research 25, no. 4 (2016): 1145-1165.