Beyond simple maps - Integrating space and time with Bayesian models

Corey S. Sparks, Ph.D.

Univerity of Texas at San Antonio - Department of Demography

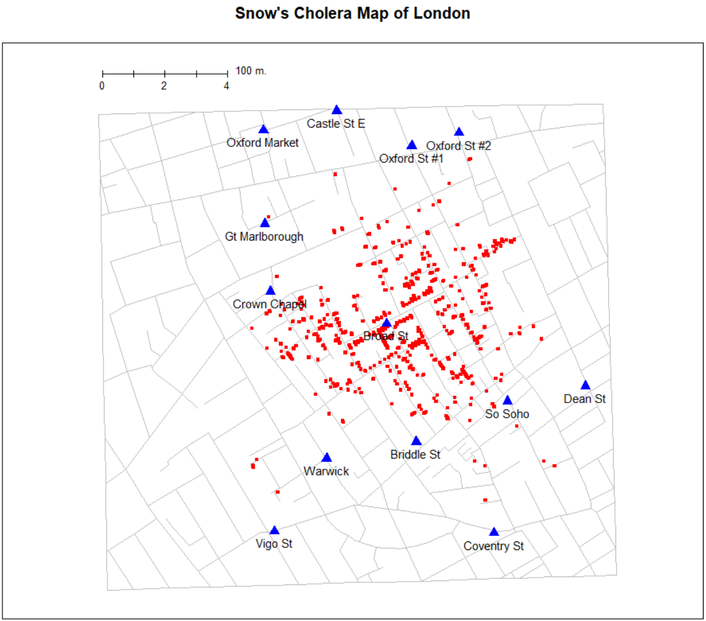
https://hcap.utsa.edu/demography

July 11, 2022

Presentation Structure

- Spatial and temporal demography
- Data sources
- Modeling strategies
- Empirical analysis of Florida mortality rates
- Results & visualizations
- Wrap up

Beyond maps. . .


Snow's Cholera Map of London
100 m.
0
2
4
Castle St E
Oxford Market
Oxford St #2
Oxford St #1
Gt Marlborough
Crown Chapel
Broad St
Dean St
So Soho
Briddle St
Warwick
Vigo St
Coventry St

Spatial Demography

- "Putting people into place" (Entwisle, 2007)
- Need to think about:
- Context
- Dynamics
- Processes
- Macro - demography (Voss, 2007)
  - Places as observations
  - Pre - 1960's
  - Ecological inference
- Micro - demography
  - People as observations
  - Social theory
  - Individual choices
- Multilevel - demography

Space & Time

- Future directions in spatial demography report
  - Most participants listed time or temporal data as integral to the future of the field
- Time allows for dynamics of humans and environment
  - Snap shots/cross sections tell us nothing of this

Space & Time data models

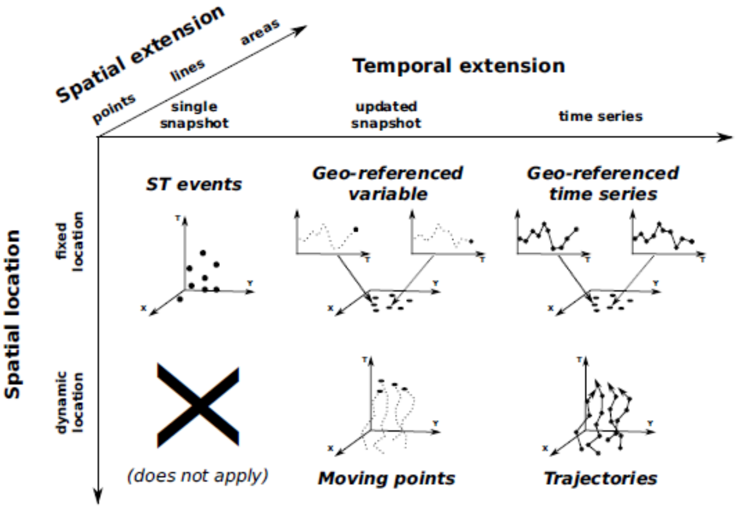


Fig. 44.1. Context for ST Clustering

Complexities

.pull-left[- Humans, I mean c'mon]

.pull-right[]

Complexities

- Data sources ?
  - Surveys
  - **Administrative data**
- Data management
  - Combining and merging data
- Analysis/methods
  - Problems with space
  - Problems with time
- Advantages
  - Rich, dynamic contexts
  - Policy relevance of timely, prospective analysis

Data sources

- NCHS/CDC
- Census/ACS
- DHS
- IPUMS
- International agencies
- Various administrative orgs.
  - State government
  - Private companies/Nonprofits

How to combine these things?

- Geocodes are essential
  - Limitation for many surveys
- **Caveats**
- Levels of geography
  - The evil tracts
- MAUP
- Changing boundaries
- Analytically
  - Lots of ways, but are they all ideal?
  - These data can often be *very* large in size

Hierarchical Models

- Allow for nesting of individuals by many different levels
- People within places, within time periods
- Different types of outcomes
  - Continuous/discrete observations/outcomes
- Can include correlation between higher level units
  - Autocorrelation between places/time periods
- Dynamic modeling
  - Place - specific time trends for example

Empirical example

- US County Mortality Rates
- NCHS Compressed Mortality File
  - County - level counts of deaths by year, age, sex, race/ethnicity and cause of death
  - 1980 to 2010
  - Age, sex and race *(white & black)* specific rates for all US counties
  - In total: 35,748,276 deaths in the data
  - Standardized to 2000 Standard US population age structure
  - Rates stratified by race and sex for each county by year
  - n = 2 sexes $*$ 2 races $*$ 3106 counties $*$ 31 years = 385,144 observations
  - *Analytic* n = 315,808 nonzero rates

—

\-

You

can

ba-

si-

cally

Data example

County	Year	Race-Sex	Rate
12073	1980	White Female	7.238632
12073	1980	Black Female	8.958174
12073	1980	White Male	11.840842
12073	1980	Black Male	15.907688
12073	1981	White Female	7.383039
12073	1981	Black Female	9.379846
12073	1981	White Male	10.518428
12073	1981	Black Male	16.626825
12073	1982	White Female	7.370335
12073	1982	Black Female	8.695655
12073	1982	White Male	11.902308
12073	1982	Black Male	12.149819

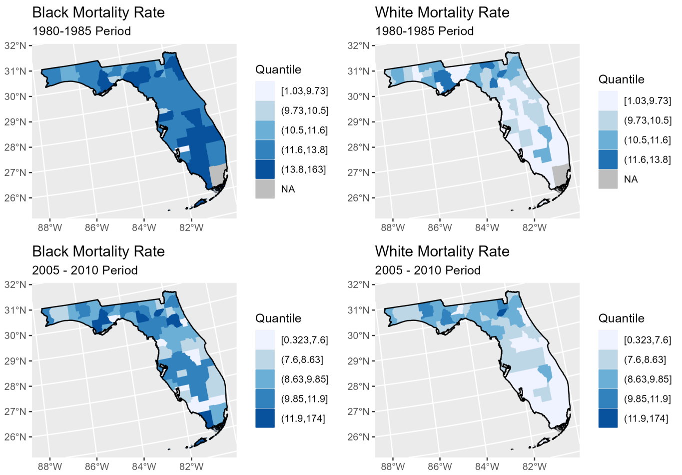
County specific temporal trends 1980 - 2010

Union County, FL, 1980 − 2010

Leon County, FL, 1980 − 2010

Florida Example

- $n = 67$ counties $*$ 31 years $*$ 2 Races $*$ 2 Sexes $= 8{,}308$



Methods - Bayesian Hierarchical models

- Example case of Florida counties
- Examine county-specific time trends in Black/White mortality rates
- I specify a Bayesian Hierarchical model for the age-standardized mortality rate
- Controls for sex and county SES
- Spatial correlation in overall rate u_j
- Time varying Black/white disparity parameter ν_{t2}
- Spatially varying Black/White disparity parameter γ_j

$$
\begin{aligned}
\mathrm{y}_{ij} &\sim N(\mu, \tau_y) \\
\mu_{ij} &= \beta_0 + x'\beta + \gamma_j * Black + u_j + \nu_{t1} + \nu_{t2} * Black \\
\gamma_j &\sim \mathrm{CAR}(\bar{\gamma}_j, \tau_\gamma / n_j) \\
u_j &\sim \mathrm{CAR}(\bar{u}_j, \tau_u / n_j) \\
\nu_{t2} &\sim RW1(time) \\
\nu_{t1} &\sim N(0, \tau_t)
\end{aligned}
$$

Methods - Bayesian Hierearchical models

- This type of model is commonly used in epidemiology and public health
- Various types of data likelihoods may be used
- Need to get at:

*

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

- Traditionally, we would get $p(\theta|y)$ by:
  - either figuring out what the full conditionals for all our model parameters are (hard)
  - Use some form of MCMC to arrive at the posterior marginal distributions for our parameters (time consuming)

Methods - INLA approach

- Integrated Nested Laplace Approximation - Rue, Martino & Chopin (2009)
- One of several techniques that approximate the marginal and conditional posterior densities
  - Laplace, PQL, E-M, Variational Bayes
- Assumes all random effects in the model are latent, zero-mean Gaussian random field, x with some precision matrix
  - The precision matrix depends on a small set of hyperparameters
- Attempts to construct a joint Gaussian approximation for $p(x|\theta, y)$
  - where θ is a small subset of hyper-parameters

Methods - INLA approach

- Apply these approximations to arrive at:
- $\tilde{\pi}(x_i|y) = \int \tilde{\pi}(x_i|\theta, y)\tilde{\pi}(\theta|y)d\theta$
- $\tilde{\pi}(\theta_j|y) = \int \tilde{\pi}(\theta|y)d\theta_{-j}$
- where each $\tilde{\pi}(.|.)$ is an approximated conditional density of its parameters
- Approximations to $\pi(x_i|y)$ are computed by approximating both $\pi(\theta|y)$ and $\pi(x_i|\theta, y)$ using numerical integration to integrate out the nuisance parameters.
    - This is possible if the dimension of θ is small.
- Approximations to $\tilde{\pi}(\theta|y)$ are based on the Laplace appoximation of the marginal posterior density for $\pi(x, \theta|y)$
- Their approach relies on numerical integration of the posterior of the latent field, as opposed to a pure Gaussian approximation of it

INLA in R

`library(INLA)`

`std_rate~male+black+scale(lths)+`

`f(year2, model = "rw1",constr = T, scale.model = T)+`
nonparametric time trend

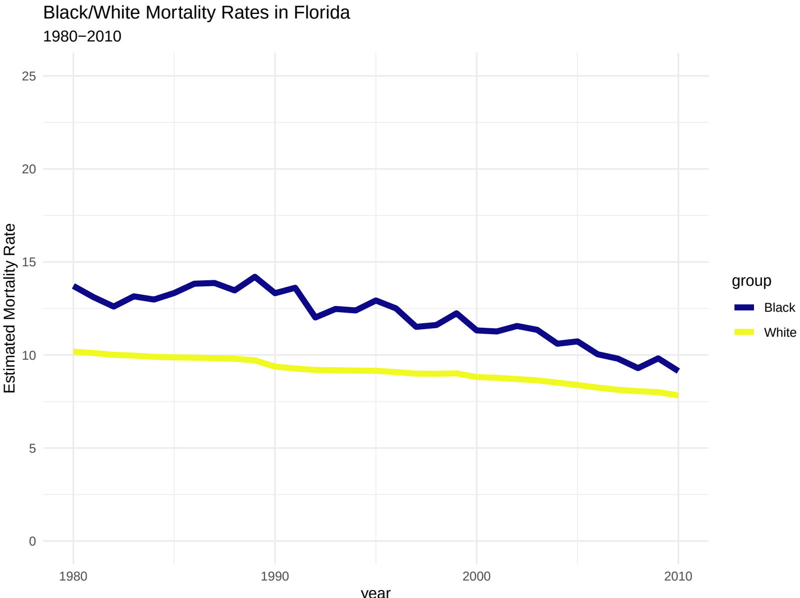
`f(struct, model="besag", graph="cl_graph", constr = T, scale.model = T)+` **spatial correlation**

`f(year3, bl2,  model="iid")+` **time - disparity**

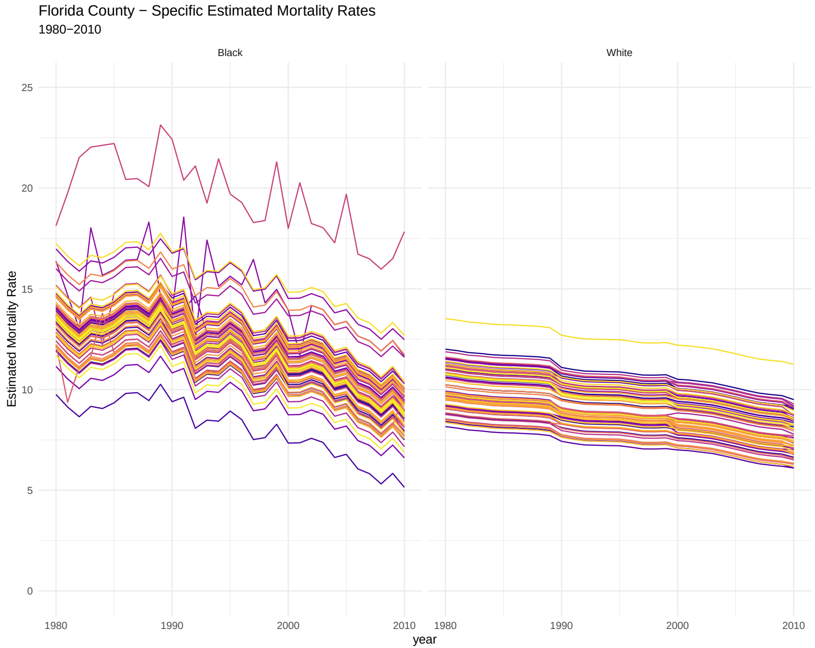
`f(struct2, bl2, model="besag", graph="cl_graph", constr = T, scale.model = T)` **spatial disparity**

Results

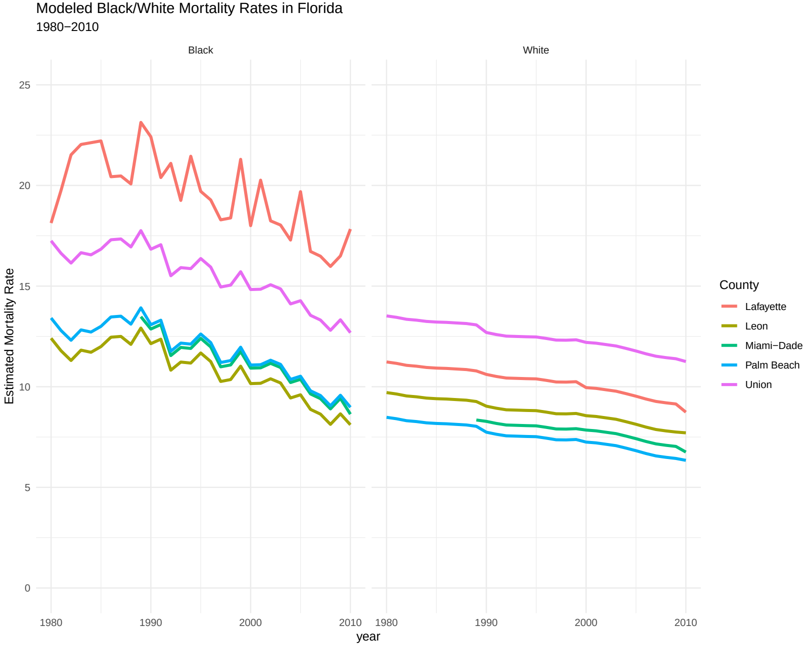
- Time trend in Black/white Mortality



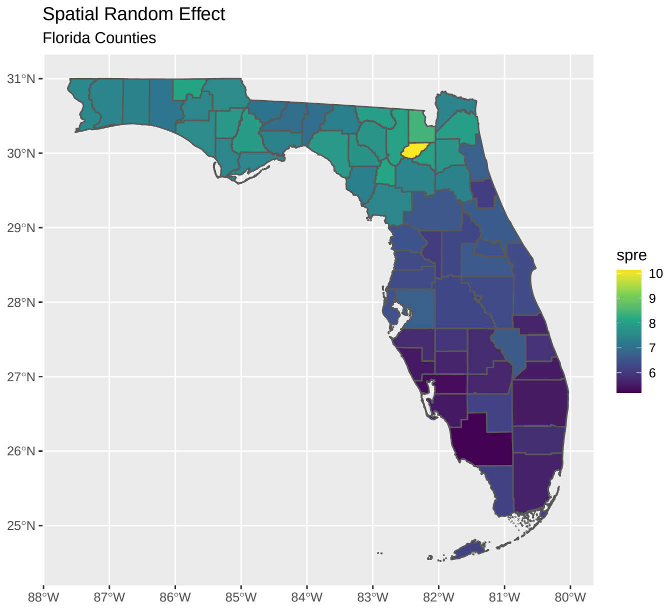
County time trends



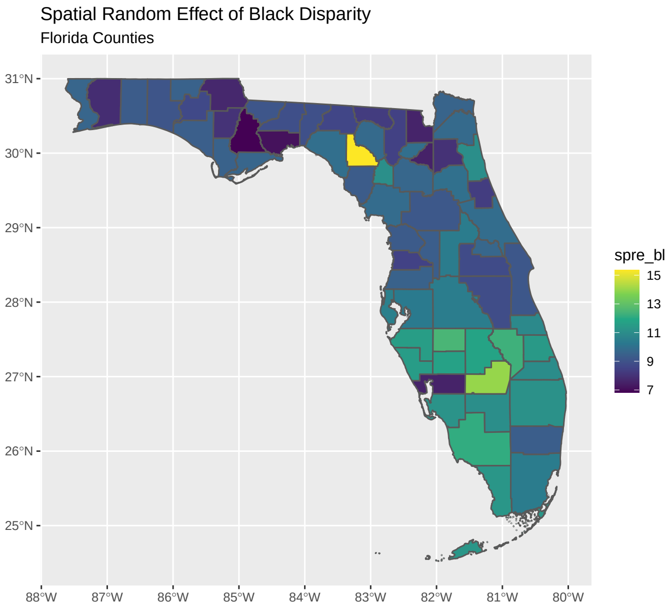
Highlighed trends



Spatial trend



Spatial disparity



Discussion

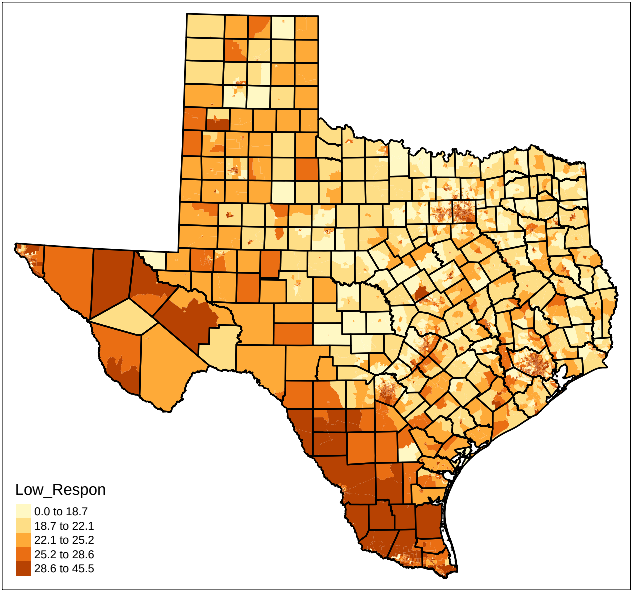
- We see that, while there is a persistence of the gap in black-white mortality:
  - The mortality gap appears to be fairly consistent over time
  - In some areas, the mortality difference are decreasing
  - Results point to higher disparities in several notable Florida rural areas
- Spatio-temporal modeling allows for the incorporation of dynamics that cross-sectional models cannot

Low Response Score Outcome

- INLA model for Low Response Score metric
- Considered both an unstructured and spatially structured random effect model
- Modeled LRS as Gaussian, considering how it is constructed
- Besag, York and Mollie specification for tract level heterogeneity

$$
\begin{aligned}
\mathrm{y}_i &\sim \mathrm{Normal}\,(\mu_i, \tau_y) \\
\mu_i &= \beta_0 + u_i + v_i \\
u_i &\sim \mathrm{CAR}(\bar{u}_i, \tau_u/n_j) \\
u_i &\sim \mathrm{Normal}(\bar{0}, \tau_v/n_j)
\end{aligned}
$$

Low Response Score in Texas



Spatial correlation random effect

More discussion

- INLA allows for rapid deployment of Bayesian statistical models with latent Gaussian random effects
  - Faster and *generally* as accurate as MCMC
  - Potentially an attractive solution for problems where large data/complex models may make MCMC less desirable

Thank you!

corey.sparks@utsa.edu

@Coreysparks1

UTSA Demography

Slides created via the R package **xaringan**

All talk materials available at my Github page

R-INLA examples available at my Rpubs page