MODULE 5: Spatial Statistics in Epidemiology and Public Health
Lecture 5: Spatial regression

Jon Wakefield and **Lance Waller**

References

- Waller and Gotway (2004, Chapter 9) *Applied Spatial Statistics for Public Health Data*. New York: Wiley.
- Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data, 2nd Ed.* Boca Raton, FL: CRC/Chapman & Hall.
- Blangiardo, M. and Cameletti, M. (2015) *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Chichester: Wiley.

What do we have so far?

- Point process ideas (intensities, K-functions).
  - Data: (x, y) event locations.
  - *Where* are the clusters? Use intensities.
  - *How* are events clusters (in average)? Use K-functions.
- Inhomogeneous Poisson process $\rightarrow$ regional counts are Poisson distributed.
- Non-overlapping areas *should* be independent.
- Point process results provide a basis for the small area estimation methods from yesterday.
- Tension between *statistical precision* (want large local sample sizes $\rightarrow$ big regions), and *geographic precision* (want small regions for more detail in map).

What's left?

- So we know how to describe and evaluate spatial patterns in health outcome data.
- What about linking patterns in health outcomes to patterns in exposures?
- With *independent* observations we know how to use *linear* and *generalized linear* models such as linear, Poisson, logistic regression.
- What happens with *dependent* observations?

Caveat

"...all models are wrong. The practical question is how wrong do they have to be to not be useful."
Box and Draper (1987, p. 74)

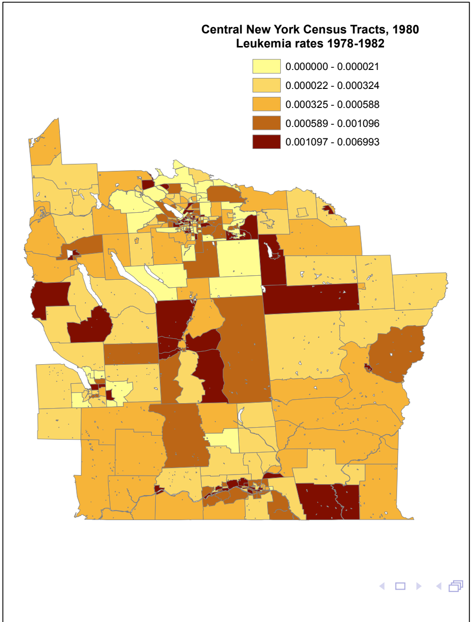
What changes with dependence?

- In statistical modeling, we are often trying to describe the mean of the outcome as a function of covariates, assuming error terms are mutually independent.
- Where do correlated errors come from?
- Perhaps outcomes truly correlated (infectious disease).
- Perhaps we omitted an important variable that has spatial structure itself.

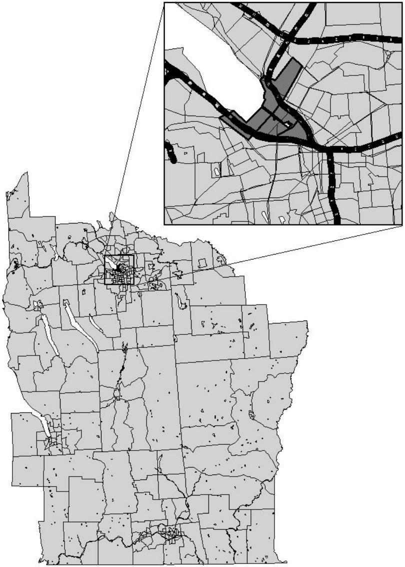
NY leukemia data

- NY leukemia data and covariates (Waller and Gotway, 2004).
- 281 census tracts (1980 Census).
- 8 counties in central New York.
- 592 cases for 1978-1982.
- 1,057,673 people at risk.

Crude Rates (per 100,000)



Outliers, where are the top 3 rates?



Building the model: Poisson regression

- Let Y_i = count for region i.
- Let E_i = *expected* count for region i.
- Let $(x_{i,TCE}, x_{i,65}, x_{i,home})$ be the associated covariate values.
- Poisson regression:

$$Y_i \sim Poisson(E_i \zeta_i)$$

where

$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home}.$$

Details

- Poisson distribution for counts.
- *Link function*: Natural log of mean of Y_i is a linear function of covariates.
- βs represent multiplicative increases in expected counts, e^{β} a measure of relative risk associated with one unit increase in covariate.
- E_i an *offset*, what we expect if the covariates have no impact.
- Age, race, sex adjustments in either E_i (standardization) or covariates.

Adding spatial correlation: New York data

- Assume E_i known, perhaps age-standardized, or based on global (external or internal) rates.
- Our model is

$$Y_i|\boldsymbol{\beta}, \psi_i \overset{ind}{\sim} \text{Poisson}(E_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \psi_i)),$$

$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \psi_i.$$

- The ψ_i represent the *random intercepts*.
- Add *overdispersion* via $\boldsymbol{\psi}_i \overset{ind}{\sim} N(0, v_\psi)$.
- Add spatial correlation via

$$\boldsymbol{\psi} \sim MVN(\mathbf{0}, \Sigma).$$

Priors and "shrinkage"

- Overdispersion model (i.i.d. ψ_i) results in each estimate being a compromise between the *local* SMR and the *global average* SMR.
- "Borrows information (strength)" from other observations to improve precision of local estimate.
- "Shrinks" estimate toward global mean. (Note: "shrink" does not mean "reduce", rather means "moves toward").

Local shrinkage

- Spatial model (correlated ψ_i) results in each estimate begin a compromise between the *local* SMR and the *local average* SMR.
- Shrinks each ψ_i toward the average of its *neighbors*.
- Can also include *both* global and local shrinkage (Besag, York, and Mollié 1991).
- How do we fit these models?

Bayesian inference

Bayesian inference regarding model parameters based on *posterior distribution*

$$Pr[\boldsymbol{\beta}, \psi | \boldsymbol{Y}]$$

proportional to the product of the likelihood times the prior

$$Pr[\boldsymbol{Y} | \boldsymbol{\beta}, \psi] Pr[\psi] Pr[\boldsymbol{\beta}].$$

Defers spatial correlation to the prior rather than the likelihood.

Spatial priors

- Could model *joint* distribution

$$\boldsymbol{\psi} \sim MVN(\mathbf{0}, \Sigma).$$

- Could also model *conditional* distribution

$$\psi_i|\psi_{j\neq i} \sim N\left(\frac{\sum_{j\neq i} c_{ij}\psi_j}{\sum_{j\neq i} c_{ij}}, \frac{1}{v_{CAR}\sum_{j\neq i} c_{ij}}\right), i = 1, \ldots, N.$$

  where c_{ij} are *weights* defining the neighbors of region i.
- Adjacency weights: $c_{ij} = 1$ if j is a neighbor of i.

CAR priors

- The conditional specification defines the *conditional autoregressive* (CAR) prior (Besag 1974, Besag et al. 1991).
- Under certain conditions on the c_{ij}, the CAR prior defines a valid multivariate joint Gaussian distribution.
- Variance covariance matrix a function of the *inverse* of the matrix of neighbor weights.

Fitting Bayesian models

- Posterior often difficult to calculate mathematically.
- Markov chain Monte Carlo: Iterative simulation approach to model fitting.
- Given *full conditional* distributions, simulate a new value for each parameter, holding the other parameter values fixed.
- The set of simulated values converges to a sample from the posterior distribution.
- Alternative: *integrated nested Laplace analysis* using the `inla` package (example code).

Complete model specification

$$Y_i|\boldsymbol{\beta}, \psi_i \overset{ind}{\sim} \text{Poisson}(E_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \psi_i)),$$

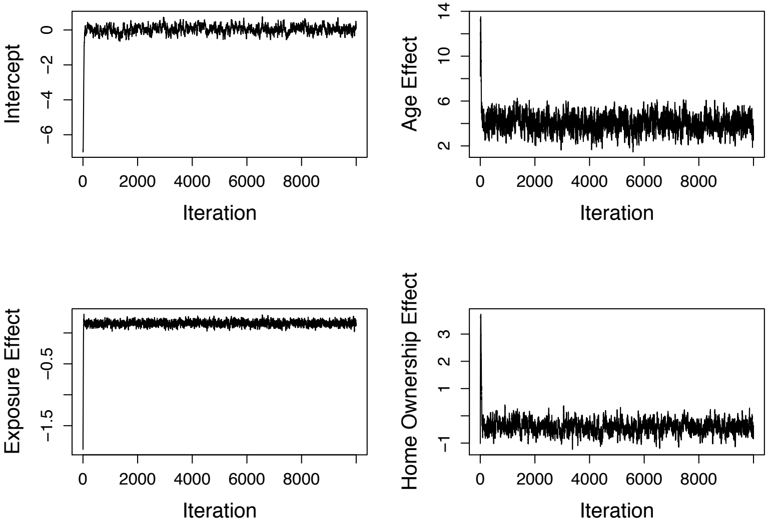
$$\log(\zeta_i) = \beta_0 + x_{i,TCE}\beta_{TCE} + x_{i,65}\beta_{65} + x_{i,home}\beta_{home} + \psi_i.$$

$$\beta_k \sim \text{Uniform}.$$

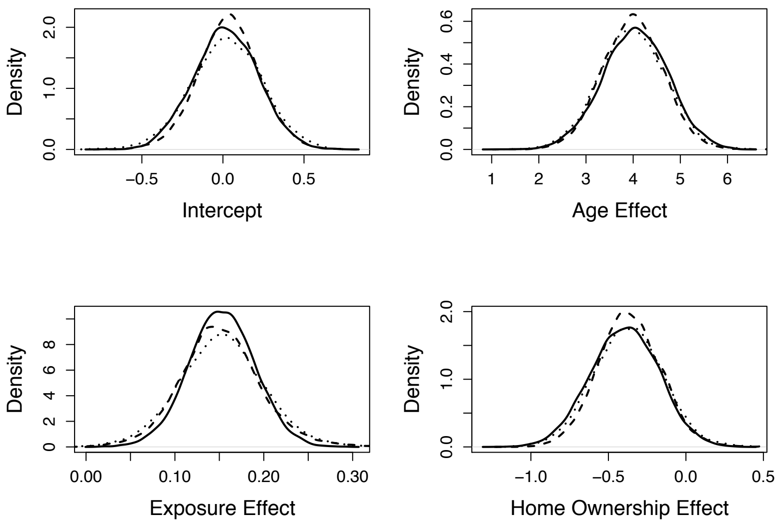
$$\psi_i|\psi_{j\neq i} \sim N\left(\frac{\sum_{j\neq i} c_{ij}\psi_j}{\sum_{j\neq i} c_{ij}}, \frac{1}{v_{CAR}\sum_{j\neq i} c_{ij}}\right), i = 1, \ldots, N.$$

$$\frac{1}{v_{CAR}} \sim \text{Gamma}(0.5, 0.0005).$$

MCMC trace plots



Posterior densities



MCMC posterior estimates

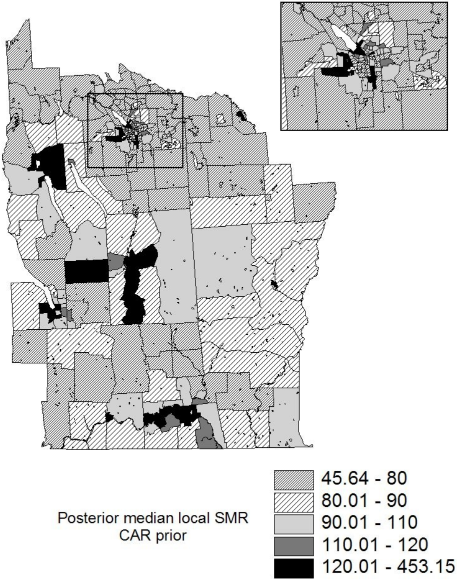
Covariate	Posterior Median	95% Credible Set
β_0	0.048	(-0.355, 0.408)
β_{65}	3.984	(2.736, 5.330)
β_{TCE}	0.152	(0.066, 0.226)
β_{home}	-0.367	(-0.758, 0.049)

But there's more!

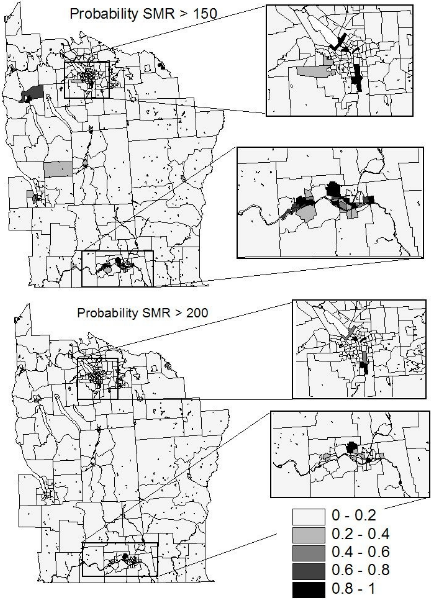
- A nifty thing about MCMC estimates:

  We get posterior samples from any function of model parameters by taking that function of the sampled posterior parameter values.
- Gives us posterior inference for $SMR_i = Y_{i,fit}/E_i$.
- Also can get $Pr[SMR_i > 200|\boldsymbol{Y}]$ and map these *exceedence probabilities*.

Posterior median SMRs



Posterior exceedence probabilities



What if associations vary across space?

- Usually, we assume β is the same at every location.
- What happens if this association varies across space?
- In the NY leukemia model we had a *random intercept* that was spatially correlated.
- What if we have a *random slope*? Can we use CAR priors for that?

What about spatially varying associations?

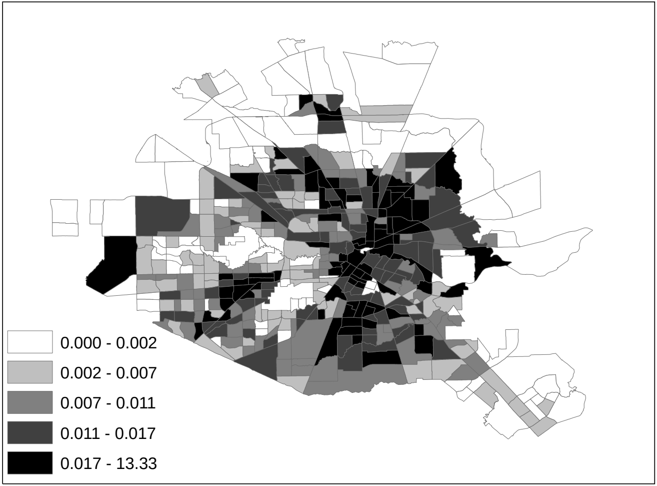
- Fix it: Geographically weighted regression (GWR)
  - Fotheringham et al. (2002)
- Model it: Spatially varying coefficient (SVC) models
  - Leyland et al. (2000), Assuncao et al. (2003), Gelfand et al. (2003), Gamerman et al. (2003), Congdon (2003, 2006)

Motivating example

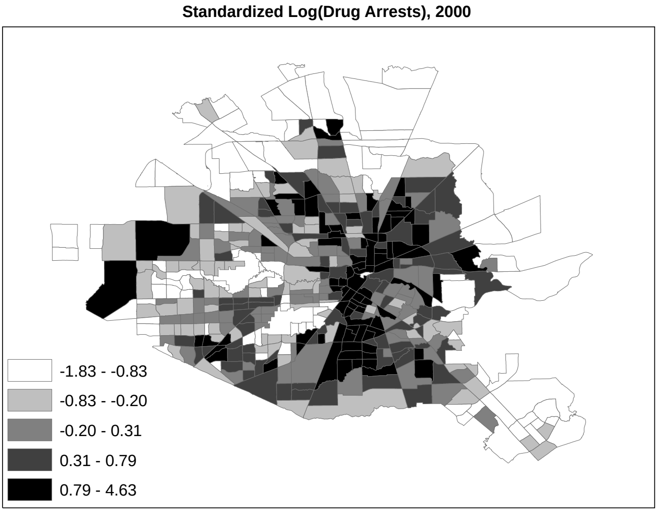
- Spatial support: 439 census tracts (2000 Census).
- Violent crime (murder, robbery, rape, aggravated assault) "first reports" for year 2000 from City of Houston Police Department website.
- Gorman et al. (2005, *Drug Alcohol Rev*) report less than 5% discrepancy with 2000 Uniform Crime Reports.
- 98% of reports geocoded to the census tract level.
- Alcohol data (locations of active distribution sites in 2000) from Texas Alcoholic Beverage Commission (6,609 outlets), 99.5% geocoded to the tract level.
- Drug law violations (also from City of Houston police data). 98% geocoded to the tract level.

Violent Crime reporting rates, Houston, 2000

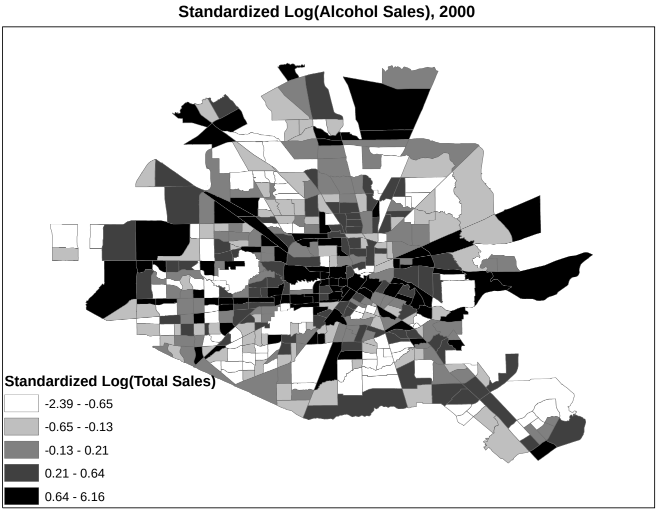
Violent Crimes per Person, 2000



Standarized log(drug arrests), Houston, 2000



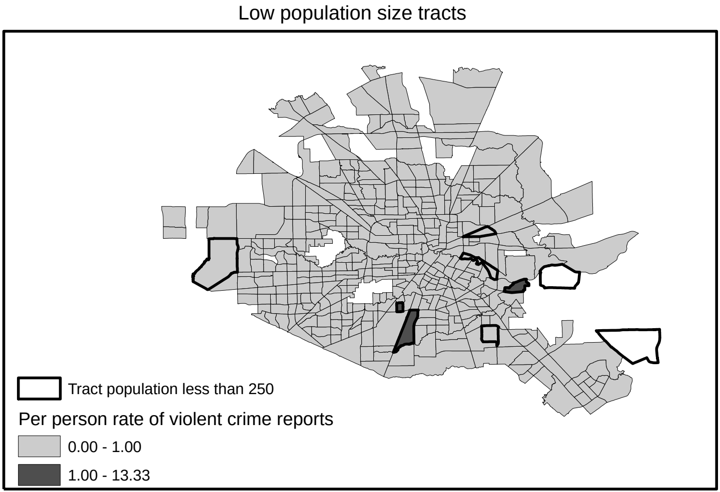
Standardized log(alcohol sales), Houston, 2000



Data “features”

- 7 of 439 tracts have extremely small population sizes: 1, 3, 4, 16, 34, 116, and 246.
- Tracts typically have 3,000-5,000 residents.
- Local rates for such tracts are extremely unstable (e.g., 40 reports, 3 residents).
- Actually a motivating a reason for including the spatially varying intercept: borrow information across regions.

Low population tracts and high rates



Basic Poisson regression

- Let Y_i = number of reports in tract $i, i = 1, \ldots, 439$.
- Suppose $Y_i \sim \text{Poisson}(E_i \exp(\mu_i))$, where E_i = the "expected" number of reports under some null model.
- Typically, $E_i = n_i R$ where all n_i individuals in region i are equally likely to report.
- $\exp(\mu_i)$ = "relative risk" of outcome in region i.
- We add covariates in linear format (within $\exp(\cdot)$): $\mu_i = \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}$.
- Same "skeleton" for both GWR and SVC.

Geographically Weighted Poisson Regression (Nakaya et al., 2005)

- $\widehat{\beta}_{GWPR} = (\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{s})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{s})\boldsymbol{A}(\boldsymbol{s})\boldsymbol{Z}(\boldsymbol{s})$.
- $\boldsymbol{A}(\boldsymbol{s}) =$ diagonal matrix of Fisher scores.
- $\boldsymbol{Z}(\boldsymbol{s}) =$ Taylor-series approximation to transformed outcomes.
- Update $\boldsymbol{A}(\boldsymbol{s})$, $\boldsymbol{Z}(\boldsymbol{s})$ and $\widehat{\beta}_{GWPR}$ until convergence.

Fitting in *R*

- Waller et al. (2007) use `GWR 3.0` software.
- In `R`: `maptools` will read in `ArcGIS`-formatted shapefile (files) into *R*.
- `spgwr` fits linear GWR and GLM-type GWR, including GWPR.

SVC

- $\mu_i = \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i} + b_{1,i} x_{alc,i} + b_{2,i} x_{drug,i} + \phi_i + \theta_i$.
- $\beta_0, \beta_1, \beta_2 \sim$ Uniform.
- Random intercept has 2 components (Besag et al. 1991):

$$\theta_i \overset{ind}{\sim} N(0, \tau^2)$$

$$\phi_i | \phi_j \sim N\left(\frac{\sum_j w_{ij} \phi_j}{\sum_j w_{ij}}, \frac{1}{\lambda \sum_j w_{ij}} \right).$$

  where w_{ij} defines neighbors, and λ controls spatial similarity.
- θ_i allows overdispersion (smoothing to global mean).
- ϕ_i follows conditionally autoregressive distribution (smoothing to local mean), generates MVN but more convenient for MCMC.

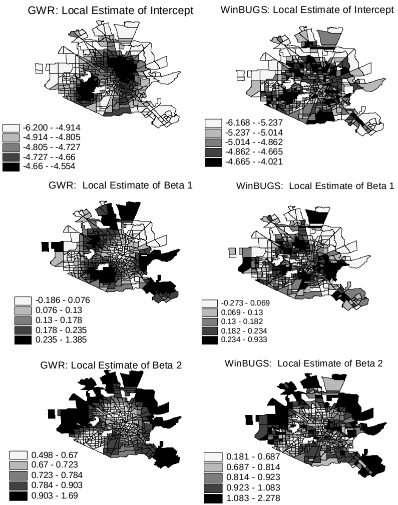
Defining the SVCs

- $\boldsymbol{b}_1, \boldsymbol{b}_2$ also given spatial priors and allowed to be correlated with one another.
- We use a formulation by Leyland et al. (2000) which defines

$$(b_{1,i}, b_{2,i})' \sim MVN((0,0)', \boldsymbol{\Sigma})$$

.

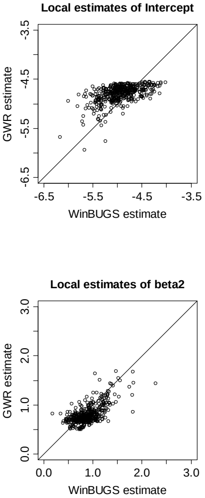
Estimated effects



Similarities

- Alcohol: Increased impact in western, south-central, and southeastern parts of city.
- Illegal drug: Increased impact on periphery, lower influence in central and southwestern parts of city.
- Intercept: Increased risk of violence in central area, above and beyond that predicted by alcohol sales and illegal drug arrests.
- But, associations not too close...

Results: tract-by-tract



Let's try it out!

- Houston data on violent crime, alcohol sales, and illegal drug arrests.
- `ArcGIS` shapefile.
- Required `R` libraries: `maptools` (to read in shape file), `RColorBrewer` (to set colors), `classInt` (to set intervals of values for mapping), and `spgwr` (for GWR).

Conclusions

- GWR and SVC very different approaches to the same problem.
- Qualitatively similar in results, but not directly transformable.
- GWR (GWPR) fixed problems within somewhat of a black box.
- SVC allows probability model-based inference with lots of flexibility but at a computational cost (both in set-up and implementation).

Acknowledgements and References

- Collaborators: Paul Gruenewald, Dennis Gorman, Li Zhu, Carol Gotway, and David Wheeler
- References:
  - Waller et al. (2008) Quantifying geographical associations between alcohol distribution and violence... *Stoch Environ Res Risk Assess* **21**: 573-588.
  - Wheeler and Caldor (2009) As assessment of coefficient accuracy...*J Geogr Systems* **9**: 573-588.
  - Wheeler and Waller (2009) Comparing spatially varying coefficient models... *J Geogr Systems* **11**: 1-22.
  - Finley (2011) Comparing spatially-varying coefficient models...*Methods in Ecology and Evolution* **2**: 143-154.

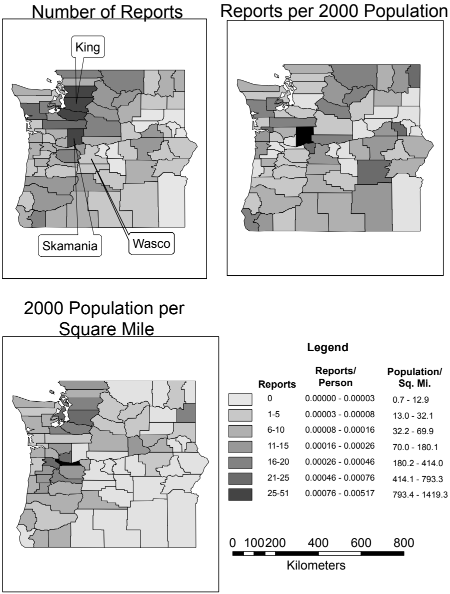
Bonus Example

- Cryptozoology Example: Waller and Carlin (2010) Disease Mapping. In *Handbook of Spatial Statistics*, Gelfand et al. (eds.). Boca Raton: CRC/Chapman and Hall.

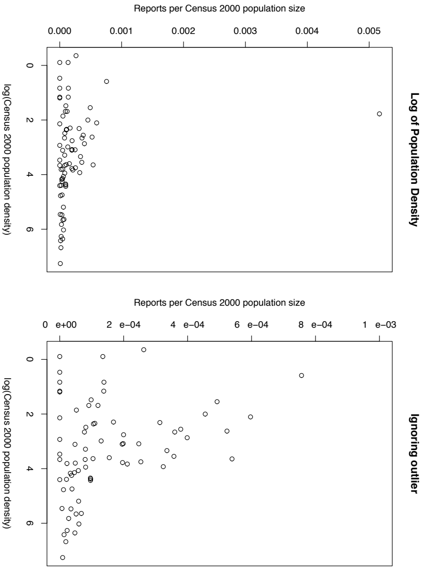
Cryptozoology example

- County-specific reports of encounters with *Sasquatch* (Bigfoot).
- Data downloaded from `www.bfro.net`
- Sightings from counties in Oregon and Washington (Pacific Northwest).
- Probability of report related to population density?
- (Hopefully) rare events in small areas.
- Perhaps spatial smoothing will stabilize local rate estimates.

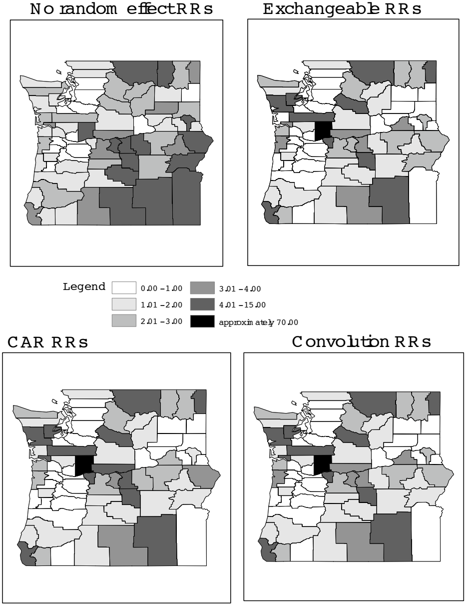
Sasquatch Data



Reports vs. Population Density



Mapped relative risks



Skamania Sasquatch Ordinances

- http://www.skamaniacounty.org/commissioners/homepage/ordinances-2/
- Big Foot Ordinance 69-1: "THEREFORE BE IT RESOLVED that any premeditated, willful and wonton slaying of any such creature shall be deemed a felony punishable by a fine not to exceed Ten Thousand Dollars ($10,000.00) and/or imprisonment in the county jail for a period not to exceed Five (5) years. ADOPTED this 1st day of April, 1969."
- Big Foot Ordinance 1984-2:
  - Repealed felony and jail sentence.
  - Established a Sasquatch Refuge (Skamania County).
  - Clarified penalty (gross misdemeanor vs. misdemeanor) and penalty (fine and jail time), disallowed insanity defense, and clarified distinction between coroner designation of victim as humanoid (murder) or anthropoid (this ordinance).

And...

www.amazon.com/Skamania-County-Washington-Bigfoot-Vintage/dp/B076PWN7ZM

Conclusions

- What method to use depends on what you want data you have and what question you want to answer.
- All methods try to balance trend (fixed effects) with correlation (here, with random effects).
- All models wrong, some models useful.
- Trying more than one approach often sensible.