# MODULE 16: Spatial Statistics in Epidemiology and Public Health
# Lecture 7: Slippery Slopes: Spatially Varying Associations

Jon Wakefield and **Lance Waller**

# What are we doing?

- Alcohol
- Illegal drugs
- Violent crimes
- Regression
- Breaking the rules

# 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.

# What do we want to do?

- Quantify associations between outcomes and covariates as observed in data.
- What if strength of association varies across space?

# Motivating example

- Outcome: Number of violent crime reports by census tract
- Covariates: Alcohol sales, illegal drug arrests (also by census tract).
- Interaction of people and place (social disorganization, routine activities, crime potential)
- Q: What if the association depends on location?

# What analytic tools do we have?

Linear regression:

$$Y_i = \mathbf{X}_i'\boldsymbol{\beta} + \epsilon_i$$

$$\epsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

(alternately, $Y_i \overset{ind}{\sim} N(\mathbf{X}_i'\boldsymbol{\beta}, \sigma^2)$)

- Assumptions: independence, Gaussian errors, constant variance, linear association, constant $\boldsymbol{\beta}$.
- We will break all but one of these.

# How do you measure association?

- $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ from OLS (= MLE if Gaussian)
- What if an assumption doesn't hold?
- Two common strategies:
  - Fix it (adjustments to LS)
  - Model it (adjustments to ML)

# "Fix it" approach

- Not Gaussian? Transformation (log, Box-Cox, ...).
- $\sigma^2$ not constant? WLS: $\widehat{\boldsymbol{\beta}}_{WLS} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$
- $\epsilon_i$ ($Y_i$) not independent? GLS: $\widehat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Y}$
- Fixing multiple problems?

# Fixing multiple problems one at a time

# "Model it" approach

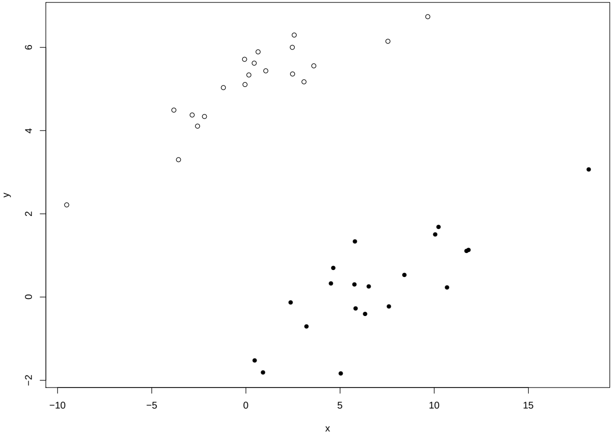
- Random effects (mixed models).
- Modeling non-independence hierarchically:

$$Y_{ij}|b_j \overset{ind}{\sim} N(\mathbf{X}'\beta + b_j, \sigma^2),$$

$$b_j \sim N(0, \tau^2),$$

where $b_j$ = random intercept for elements of $j$th subgroup.

# Example of subgroup random intercepts



# Impact of random effects

- Impact: Adds extra noise, adds (positive) correlation within subgroups.
- Borrow information across subgroups.
- Allows modeler to specify variance-covariance structure.
- Fit with SAS `PROC MIXED` or `R`.
- What about a random intercept for each census tract with spatial correlation?

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

# Inference for hierarchical models

$$[\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \mathbf{b}][\mathbf{b}|\tau^2][\boldsymbol{\beta}][\tau^2] = A \times B \times C \times D.$$

- Bayesian: $A =$ (conditionally independent) likelihood, $B, C =$ priors, $D =$ hyperprior, inference based on $[\boldsymbol{\beta}, \sigma^2, \mathbf{b}|\mathbf{Y}] \propto A \times B \times C \times D$.
- Classical: $A \times B =$ (correlated) likelihood, $C, D$ N/A, inference based on likelihood.
- Both computationally intensive.

# 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)

# Geographically weighted regression

- OLS: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
- WLS: $\hat{\beta}_{WLS} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$.
- GWR: $\hat{\beta}_{GWR}(\mathbf{s}) = (\mathbf{X}'\mathbf{W}(\mathbf{s})\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\mathbf{s})\mathbf{Y}$.
- $\mathbf{s}$ represents any spatial location in the study area.
- $\mathbf{W}(\mathbf{s})$ = diagonal matrix of weights for each observation ($Y_i$) with higher weights given to observations closer to $\mathbf{s}$.

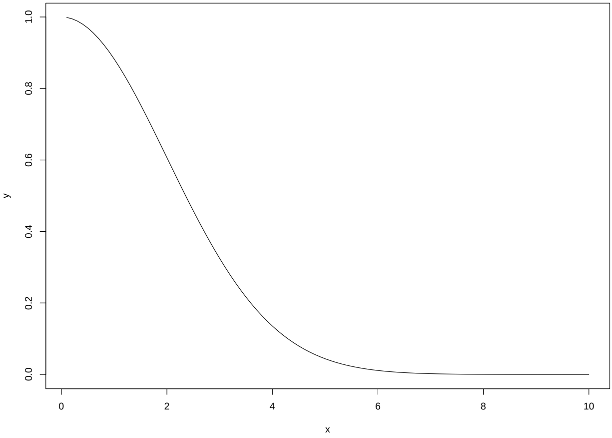
# GWR

- GWR: $\widehat{\beta}_{GWR}(\mathbf{s}) = (\mathbf{X}'\mathbf{W}(\mathbf{s})\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\mathbf{s})\mathbf{Y}$.
- Similar to local regression smoothing (e.g., loess), but here "local" means in geographic space, not covariate space.
- $\mathbf{W}(\mathbf{s})$ typically defined by a distance-decay kernel function such as the Gaussian kernel

$$\mathbf{W}(\mathbf{s})_i = \exp\left(-\frac{1}{2}\left(\frac{\mathbf{s}-\mathbf{s}_i}{bw}\right)^2\right),$$

  where $\mathbf{s}_i$ = location of observation $Y_i$ and $bw$ = "bandwidth" (smoothing parameter).
- Larger $bw$, smoother $\widehat{\beta}_{GWR}(\mathbf{s})$ surface.

# Example of Gaussian kernel



# Notes

- Fotheringham et al. (2002) (and their software) suggest using *bw* which minimizes the AIC.
- GWR a descriptive technique. Inference is complicated.
- Observation $Y_i$ receives different weight for each $\mathbf{s}$.
- Provides a smooth estimate of the $\widehat{\beta}_{GWR}(\mathbf{s})$ surface, but not a probability model of $\mathbf{Y}$ as a function of $\beta(\mathbf{s})$.
- Testing (e.g., $\beta(\mathbf{s}) = 0$ at particular locations) difficult due to correlation induced by kernel function.
- Most testing approaches based on ad hoc adjustments.

# Spatially varying coefficient models

- How could we do this with random effects?
- Consider a single covariate model

$$Y_i | b_{0,i} \overset{ind}{\sim} N(\beta_0 + x_{1,i}\beta_1 + b_{0,i}, \sigma^2)$$

$$\mathbf{b} \sim MVN(0, \mathbf{\Sigma})$$

  and $\mathbf{\Sigma}$ defines a spatial correlation matrix.
- $\beta_0 + b_{0,i}$ sort of like $\beta_0(\mathbf{s})$ in GWR, but...
  - GWR: $\beta_0(\mathbf{s})$ is a smooth surface in GWR with smoothness defined by $bw$.
  - SVC: Estimate $\beta_0$ and elements of $\mathbf{\Sigma}$ from data. These parameters induce spatial correlation between intercept associated with $Y_i$ and that of its neighbors.

# SVC

- Now, let's add spatial variation to $\beta_1$.

$$Y_i | b_{0,i}, b_{1,i} \overset{ind}{\sim} N(\beta_0 + x_{1,i}\beta_1 + b_{0,i} + x_{1,i}b_{1,i}, \sigma^2)$$

$$\mathbf{b}_0 \sim MVN(\mathbf{0}, \mathbf{\Sigma}_{b_0})$$
$$\mathbf{b}_1 \sim MVN(\mathbf{0}, \mathbf{\Sigma}_{b_1})$$

# Notes

- Good news: $[\mathbf{b}_1|\mathbf{Y}]$ gives inference we want (but at a computational cost).
- Not-so-good news:
  - Identifiability
  - Often complicated to fit via MCMC (hence the flurry of papers in 2003).

# So...

- What if the data are non-Gaussian?
- Model it: GLMs
  - $Y$ =0/1 or proportions, logistic regression.
  - $Y$ = counts or rates, Poisson regression.
  - Some additional baggage (mean and variance related)
  - Estimate $\beta$ via iteratively reweighted least squares.
- GWR for Poisson regression (Nakaya et al. 2005).
- GLMM for adding random effects (Agresti et al. 2000).

# My background: Spatial epidemiology

- Poisson regression with spatially correlated random effects.
- Outcomes: Counts of (incident or prevalent) cases of disease in small areas.
- Covariates: Environmental exposures.
- Potential confounders: Demographics.
- Data often from different sources (health department, EPA, census), linked by location via geographic information systems (GISs).
- Spatial components: Adjusting estimation for residual spatial correlation.

# Our data for today

- Outcome: Rates (number of cases per person per year) of violent crimes (police/sheriff reports).
- Covariates: Alcohol distribution (licenses and sales), illegal drug arrests (police/sheriff reports).
- Potential confounders: Sociodemographics (census).
- Linked to common spatial framework (census tracts) via GIS.

# Translation complications

- When are crime data like disease data?
  - Counts from small areas.
  - Per person “rate” of interest.
- When are crime data not like disease data?
  - Outcome not as “rare”.
  - Police vs. medical records.
  - Residents not only ones at risk.

# Background: Alcohol, drug arrests, and violent crime

- Large body of research showing association between alcohol distribution and the incidence of violence.
- Usually focuses on characteristics of:
  - People (social normative, social disorganization theories)
  - Places (routine activities theory)
  - Interactions of people and places (crime potential, ecology of crime)
- Alcohol distribution of interest since it is regulated and we have data on what and how much is sold where.

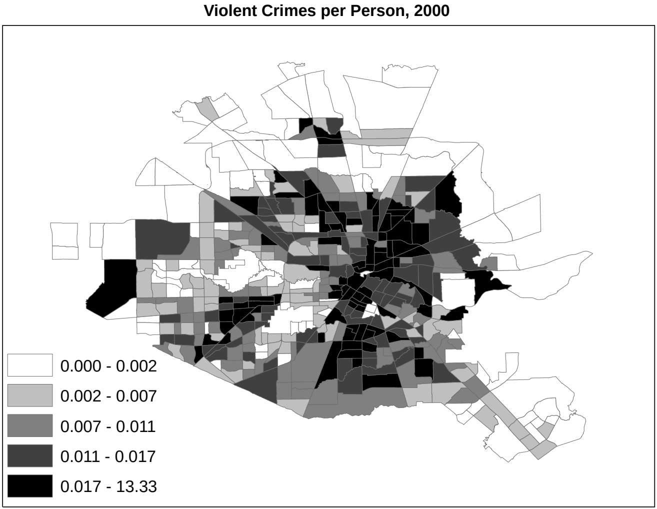
# Three questions and who asks them

- What population characteristics are associated with increased incidence of violence? (Criminologists, Sampson and Lauritsen 1994).
- In what places is violence more likely? (Routine activity theorists, Felson et al. 1997).
- What spatial interactions link people and place characteristics to violence? (Environmental criminologists, Brantingham and Brantingham 1993, 1999).

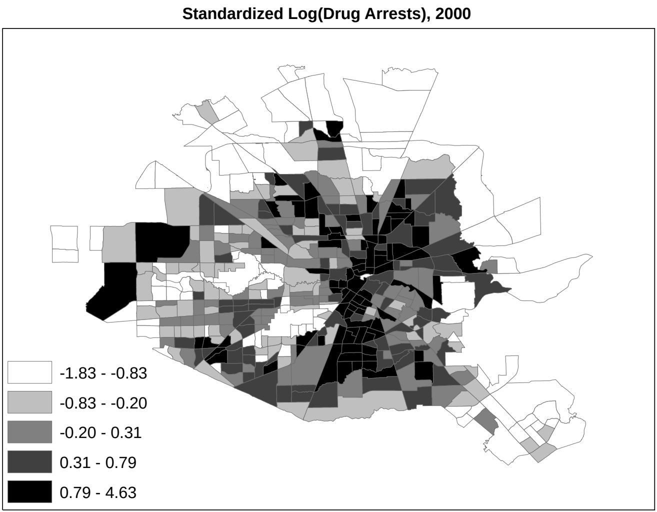
# Data description

- 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.

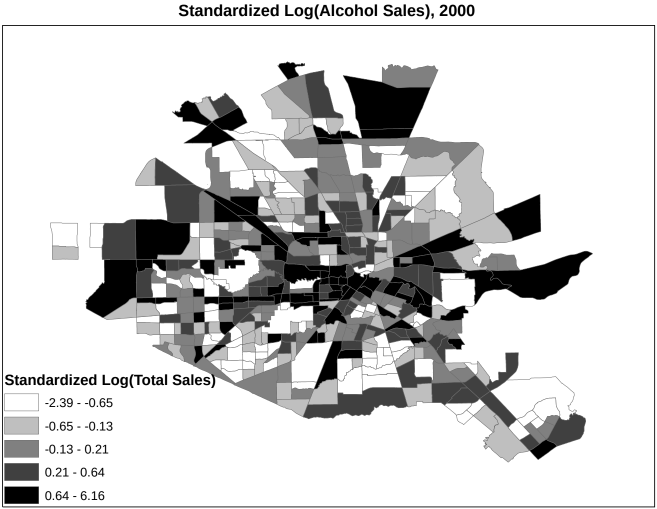
# Violent Crime reporting rates, Houston, 2000



# Standarized log(drug arrests), Houston, 2000



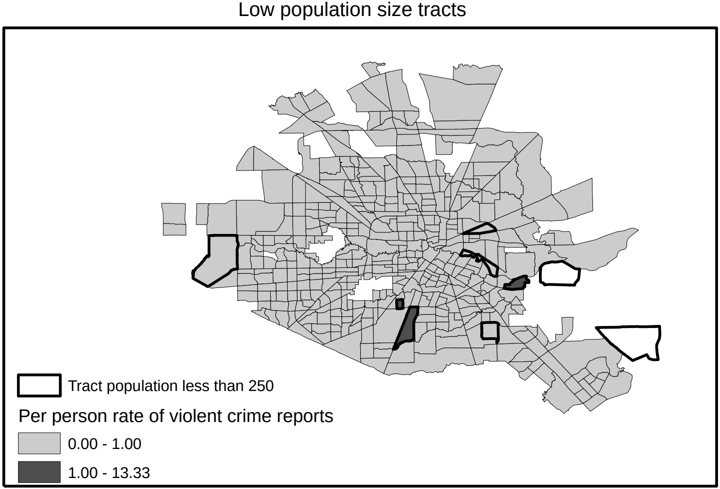
# Standarized log(alcohol sales), Houston, 2000



# 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.

# Why do we have $E_i$?

- $E_i = n_i R$ represents an "offset" in the model and lets us use Poisson regression to model *rates* as well as *counts*.

$$
\begin{aligned}
E[Y_i] &= E_i \exp(\beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}) \\
&= \exp(\ln(E_i) + \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}) \\
&= \exp(\ln(n_i) + \ln(R) + \beta_0 + \beta_1 x_{alc,i} + \beta_2 x_{drug,i}) \\
\log(E[Y_i]) &= \ln(n_i) + \ln(R) + \beta_0 + \beta 1 x_{alc,i} + \beta_2 x_{drug,i}
\end{aligned}
$$

- GWR offset: $\ln(n_i)$, SVC offset: $\ln(n_i) + \ln(R)$.

# GWPR (Nakaya et al., 2005)

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

# Fitting in *R*

- Waller et al. (2007) use `GWR 3.0` software.
- Now can use *R*.
- `maptools` will read in `ArcGIS`-formatted shapefile (files) into *R*.
- `spgwr` fits linear GWR and GLM-type GWR.
- Let's try it out!

# 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 $\lambda$ controls spatial similarity.
- $\theta_i$ allows overdispersion (smoothing to global mean).
- $\phi_i$ follows conditionally autoregressive distribution (smoothing to local mean), generates MVN but more convenient for MCMC.

# Defining the SVCs

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

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

.

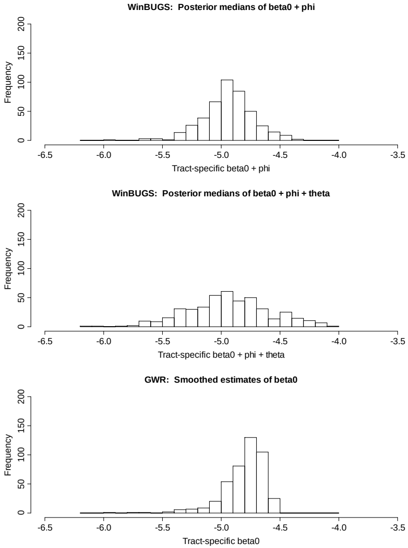
# Fitting in *WinBUGS*

- Define the model in WinBUGS.
- MCMC fit.
- Note: Runs sloooooooowly.

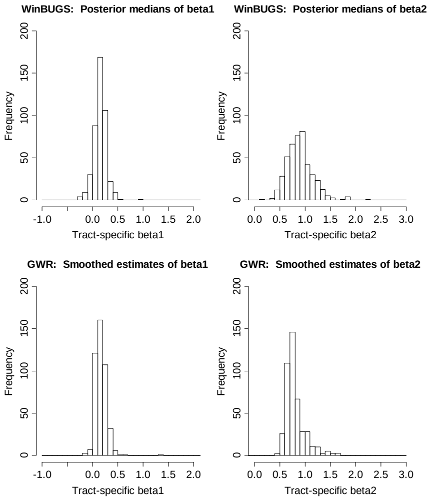
# Implementation

- Waller et al. (2007): `GWR3.0` used to fit the GWR Poisson model.
  - Converged to estimate in $\sim$ 100 iterations.
  - Minutes.
  - Example code using `spgwr` library in `R`.
- `WinBUGS 1.4.1` used to fit SVC model.
  - Converged to distribution in $\sim$ 2,000 iterations.
  - 8,000 additional iterations used for inference.
  - Hours.
- Fit several versions of SVC model and compared fit via deviance information criterion (Spiegelhalter et al., 2003).
- Best fit included spatial varying coefficients, random intercept, and correlation between alcohol and drug effects.

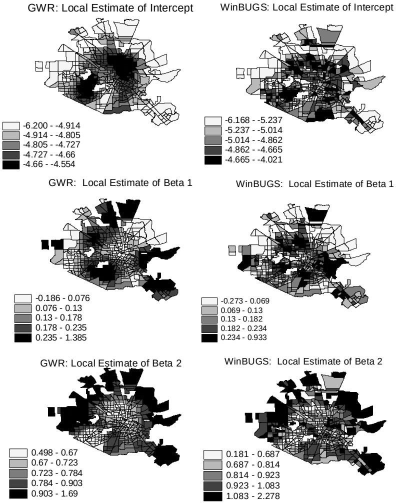
# Results: Intercept



# Results: Alcohol sales and drug arrests



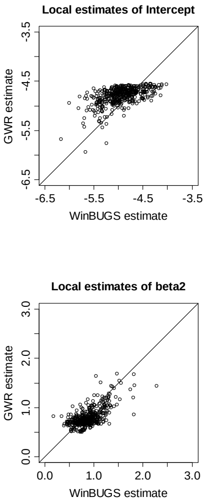
# Estimated effects



# 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...

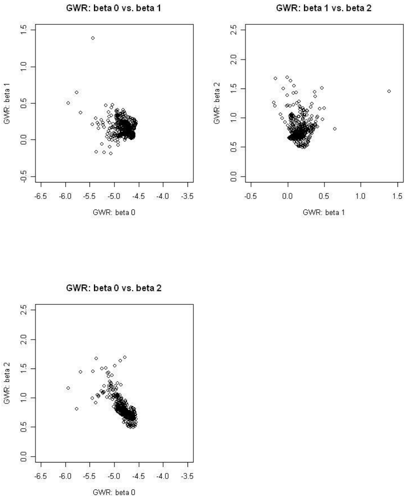
# Results: tract-by-tract



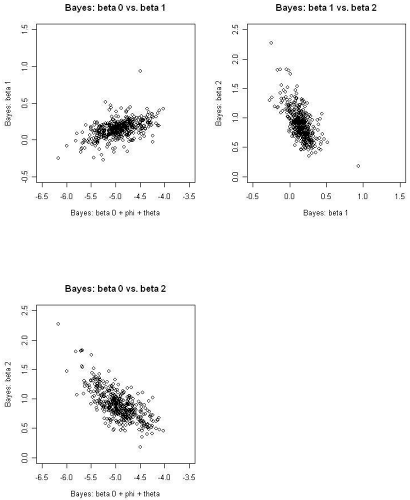
# Differences

- GWR much smoother based on global best fit for *bw*.
- SVC used adjacency-based smoothing and a different amount of smoothing for each covariate.
- GWR: collineary between surfaces (Wheeler and Tiefelsdorf, 2005).
- SVC: Model based approach removes (or at least reduces) collinearity.

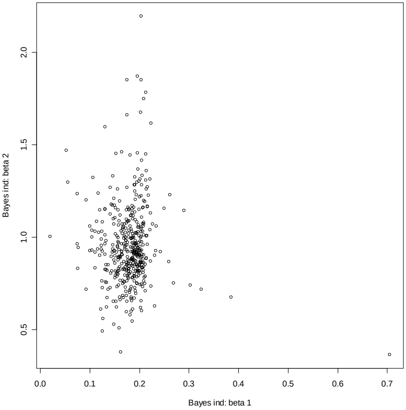
# GWR: Collinearity?



# SVC: Collinearity?



# SVC: No prior correlation



# Let's try it out!

- Houston data on violent crime, alcohol sales, and illegal drug arrests.
- `ArcGIS` shapefile.
- Required `R` libraries: `map tools` (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 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).
- Ongoing work:
  - Wheeler and Waller (2009): Attempt to set up SVC model to more closely mirror amount of smoothing in GWR.
  - Collinearity "ribbons".
  - Griffith (2002) eigenvector spatial filtering to adjust collinearity. Interpretability?