Bayesian Network Regularized Regression for Crime Modeling

Luis Carvalho

Joint work with Liz Upton

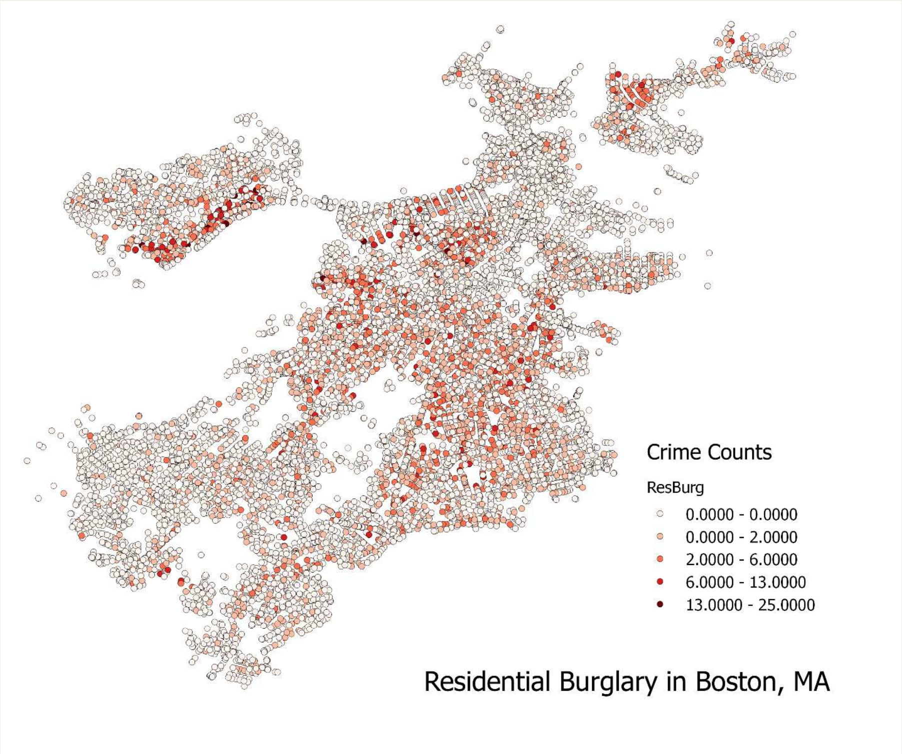
Dept. of Mathematics and Statistics

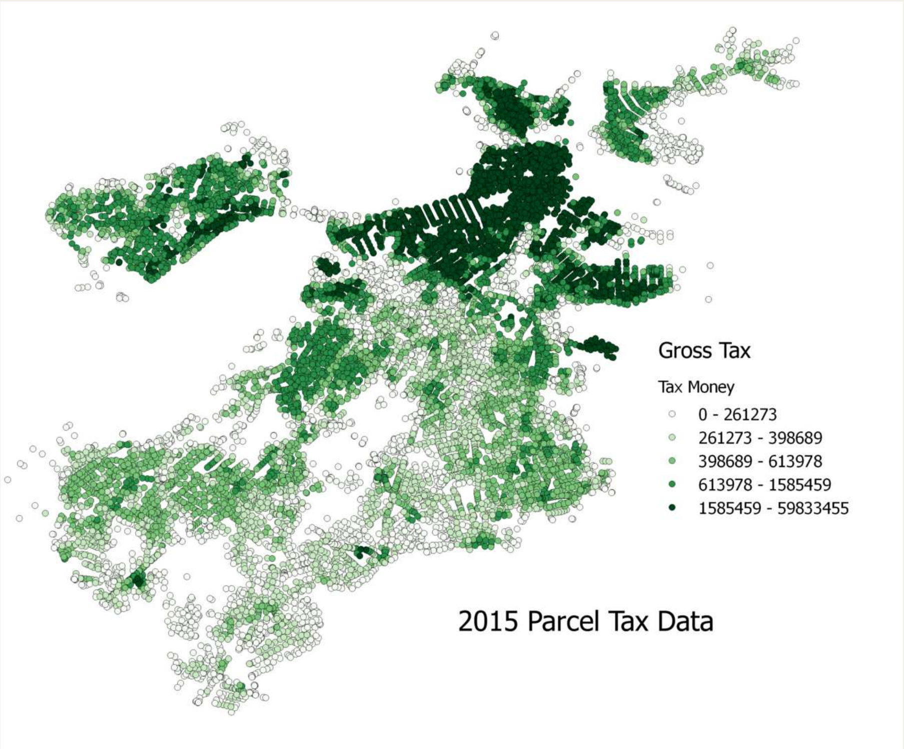
Boston University

lecarval@math.bu.edu

BU-Keio Workshop, August 2016

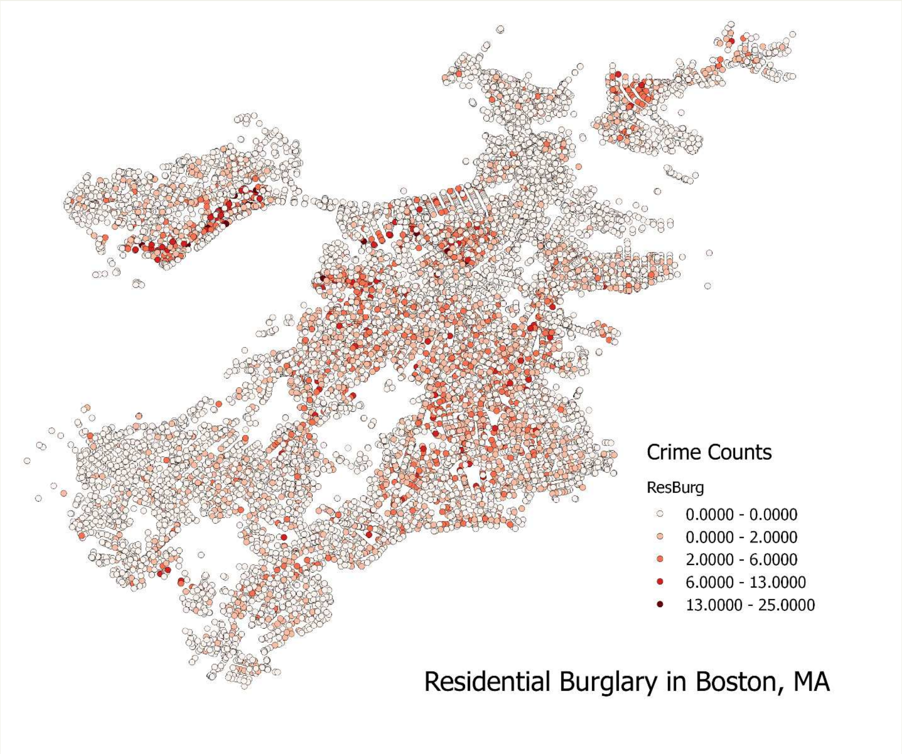
A motivating example: residential burglary in Boston

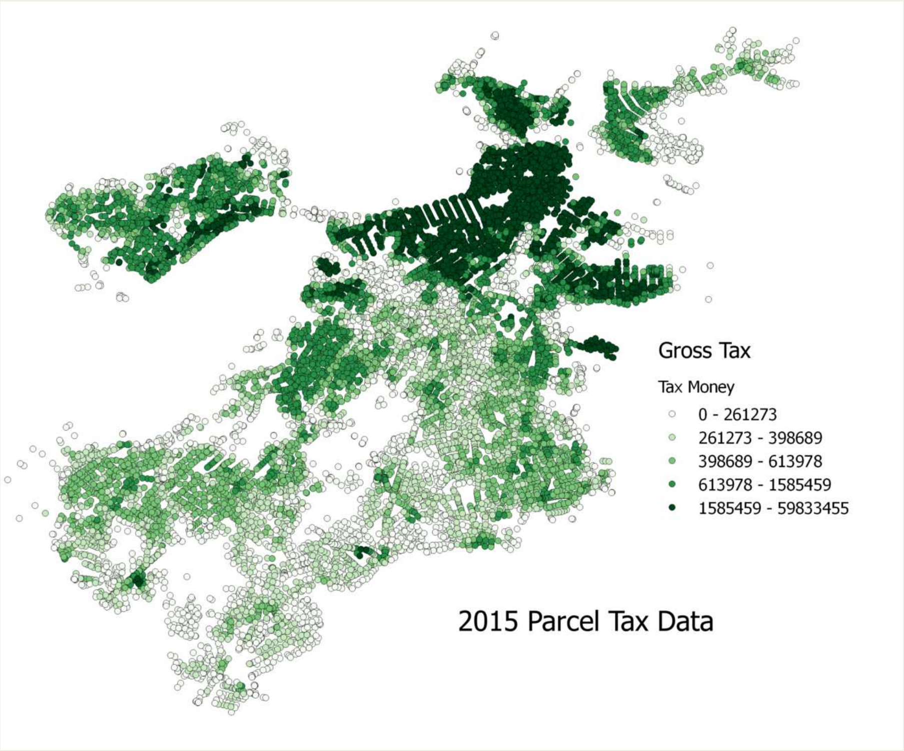




Introduction

A motivating example: residential burglary in Boston

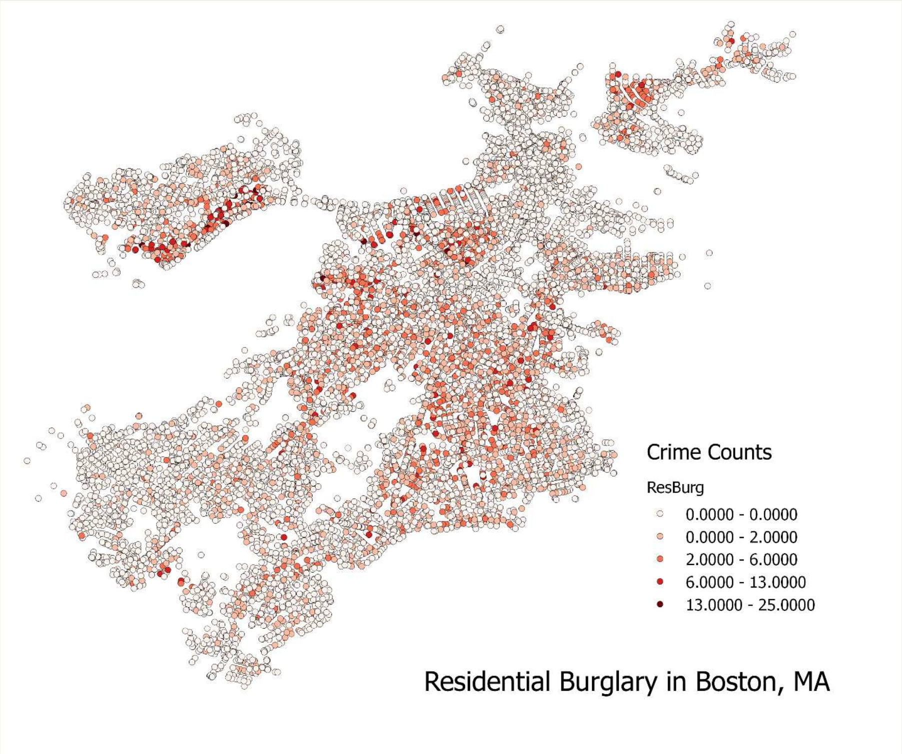


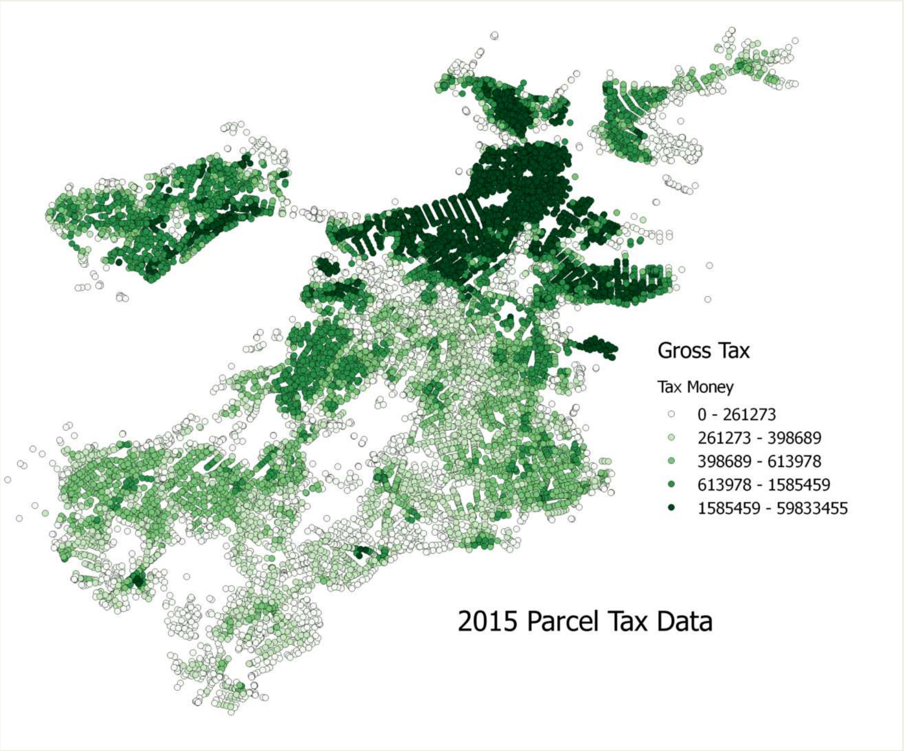


Potential goals:

- Understanding crime rates: covariates? predictions?

A motivating example: residential burglary in Boston



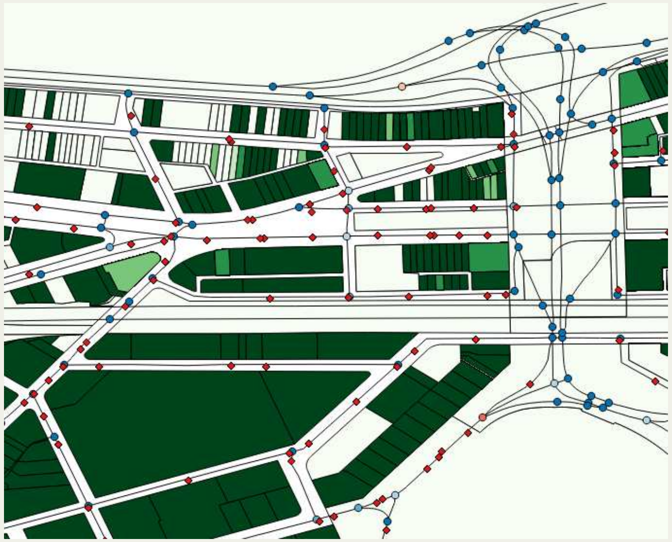


Potential goals:

- Understanding crime rates: covariates? predictions?
- Identifying "hot zones" for intervention

Residential Burglary

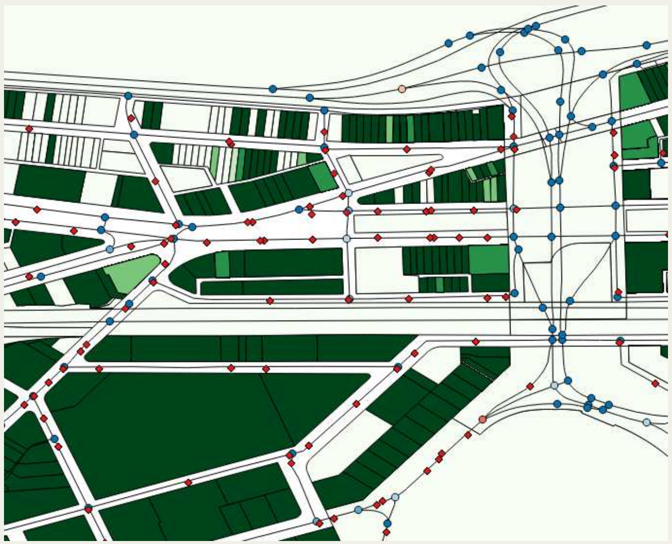
- Data description: $\sim$ 7K crimes occurring from July 2012 to October 2015 in Boston, provided by `data.cityofboston.gov`



- Reported occurrences are pooled in time and by intersection
- Covariates: averaged tax income, district type, distance to nearest police station
- Network with $\sim$ 13K nodes, provided by `boston.opendata.arcgis.com`

Residential Burglary

- Data description: $\sim$ 7K crimes occurring from July 2012 to October 2015 in Boston, provided by `data.cityofboston.gov`



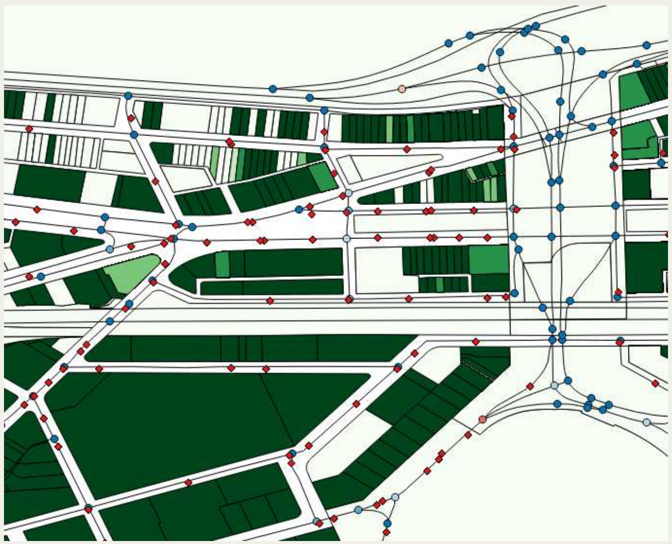
- Reported occurrences are pooled in time and by intersection
- Covariates: averaged tax income, district type, distance to nearest police station
- Network with $\sim$ 13K nodes, provided by `boston.opendata.arcgis.com`

- First take: for v in a network (undirected simple graph) G,

$$Y_v \overset{\text{iid}}{\sim} \text{Po}\left[\exp\left(\mathbf{x}_v^\top \beta\right)\right]$$

Residential Burglary

- Data description: $\sim$ 7K crimes occurring from July 2012 to October 2015 in Boston, provided by `data.cityofboston.gov`



- Reported occurrences are pooled in time and by intersection
- Covariates: averaged tax income, district type, distance to nearest police station
- Network with $\sim$ 13K nodes, provided by `boston.opendata.arcgis.com`

- First take: for v in a network (undirected simple graph) G,

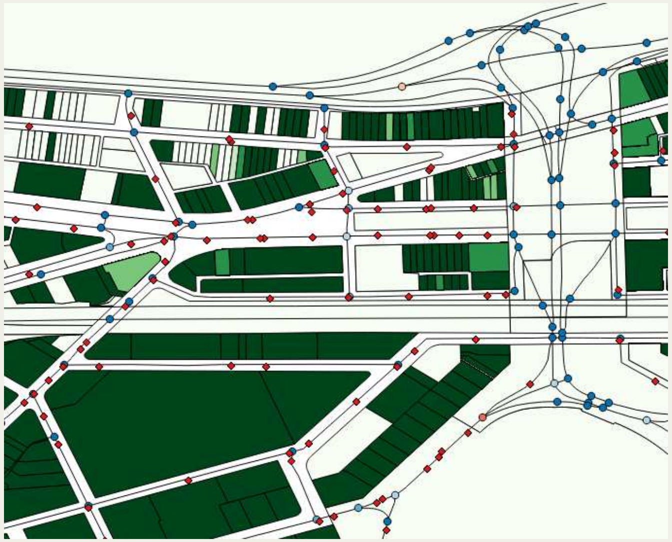
$$Y_v \overset{\text{iid}}{\sim} \text{Po}\left[\exp\left(\mathbf{x}_v^\top \beta\right)\right]$$

but:

  - Crime rates are not spatially homogeneous

Residential Burglary

- Data description: $\sim$ 7K crimes occurring from July 2012 to October 2015 in Boston, provided by `data.cityofboston.gov`



- Reported occurrences are pooled in time and by intersection
- Covariates: averaged tax income, district type, distance to nearest police station
- Network with $\sim$ 13K nodes, provided by `boston.opendata.arcgis.com`

- First take: for v in a network (undirected simple graph) G,

$$Y_v \overset{\text{iid}}{\sim} \text{Po}\left[\exp\left(\mathbf{x}_v^\top \beta\right)\right]$$

but:

  - Crime rates are not spatially homogeneous
  - Crime rates can vary sharply

Network Regularized Regression

- Addressing the first issue,

$$Y_v \overset{\text{ind}}{\sim} \text{Po}\left[\exp\left(\mathbf{x}_v^\top \beta(v)\right)\right]$$

where β is now network indexed

Network Regularized Regression

- Addressing the first issue,

$$Y_v \overset{\text{ind}}{\sim} \text{Po}\left[\exp\left(\mathbf{x}_v^\top \beta(v)\right)\right]$$

  where β is now network indexed

- To avoid overfitting, we impose smoothness on β, *e.g.*, under a single intercept model,

$$\begin{aligned}\widehat{\beta} &:= \arg\min_{\beta} \|\mathbf{Y} - \beta\|_2^2 + \lambda\|M\beta\|_2^2 \\ &= \arg\min_{\beta} D(\mathbf{Y};\beta) + \lambda\beta^\top M^\top M\beta\end{aligned}$$

  where M is a differential operator and λ is a roughness penalty

Network Regularized Regression

- Addressing the first issue,

$$Y_v \overset{\text{ind}}{\sim} \text{Po}\left[\exp\left(\mathbf{x}_v^\top \beta(v)\right)\right]$$

  where β is now network indexed

- To avoid overfitting, we impose smoothness on β, *e.g.*, under a single intercept model,

$$\begin{aligned}\widehat{\beta} &:= \arg\min_{\beta} \|\mathbf{Y} - \beta\|_2^2 + \lambda \|M\beta\|_2^2 \\ &= \arg\min_{\beta} D(\mathbf{Y};\beta) + \lambda \beta^\top M^\top M \beta\end{aligned}$$

  where M is a differential operator and λ is a roughness penalty

- Similar works: network kernel-based regression (Smola and Kondor, 2003; Kolaczyk, 2009), and, more generally, functional data analysis (Ramsay and Silverman, 1996)

Network Regularized Regression

- For network indexed coefficients, with M the oriented weighted incidence matrix:

$$\beta^\top M^\top M \beta := \beta^\top L_w \beta = \sum_{(u,v)\in E(G)} w_{uv} (\beta(u) - \beta(v))^2$$

i.e., L_w is weighted Laplacian

Network Regularized Regression

- For network indexed coefficients, with M the oriented weighted incidence matrix:

$$\beta^\top M^\top M\beta := \beta^\top L_w\beta = \sum_{(u,v)\in E(G)} w_{uv}(\beta(u)-\beta(v))^2$$

i.e., L_w is weighted Laplacian

- With $L_w := \Phi\Xi\Phi^\top$, $\Xi :=$ Diag$_{i=1,\dots,|V(G)|}(\xi_i)$, we adopt a basis expansion for β, $\beta = \Phi_{1:k}\theta$, $k \le |V(G)|$, so the penalty becomes:

$$\beta^\top L_w\beta = \theta^\top\Phi_{1:k}^\top\Phi\Xi\Phi^\top\Phi_{1:k}\theta = \theta^\top \text{Diag}_{i=1,\dots,k}(\xi_i)\theta$$

Network Regularized Regression

- For network indexed coefficients, with M the oriented weighted incidence matrix:

$$\beta^\top M^\top M\beta := \beta^\top L_w\beta = \sum_{(u,v)\in E(G)} w_{uv}(\beta(u)-\beta(v))^2$$

  i.e., L_w is weighted Laplacian

- With $L_w := \Phi\Xi\Phi^\top$, $\Xi :=$ Diag$_{i=1,\ldots,|V(G)|}(\xi_i)$, we adopt a basis expansion for β, $\beta = \Phi_{1:k}\theta$, $k \le |V(G)|$, so the penalty becomes:

$$\beta^\top L_w\beta = \theta^\top\Phi_{1:k}^\top\Phi\Xi\Phi^\top\Phi_{1:k}\theta = \theta^\top \text{Diag}_{i=1,\ldots,k}(\xi_i)\theta$$

- Under a Bayesian formulation, $\widehat{\beta}$ is the posterior mode when

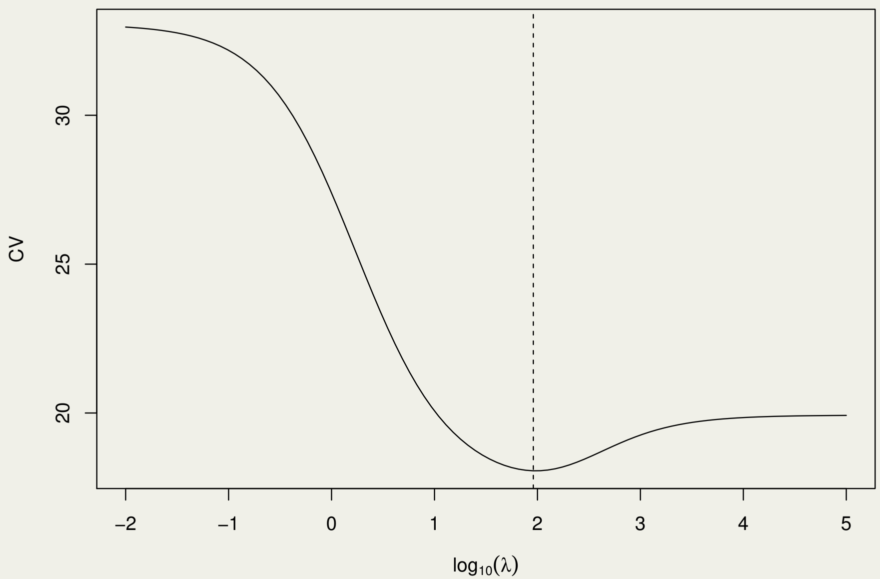
$$Y_v \,|\, \theta \overset{\text{ind}}{\sim} \text{Po}\Big[\exp\left({\phi_{kv}}^\top\theta\right)\Big]$$

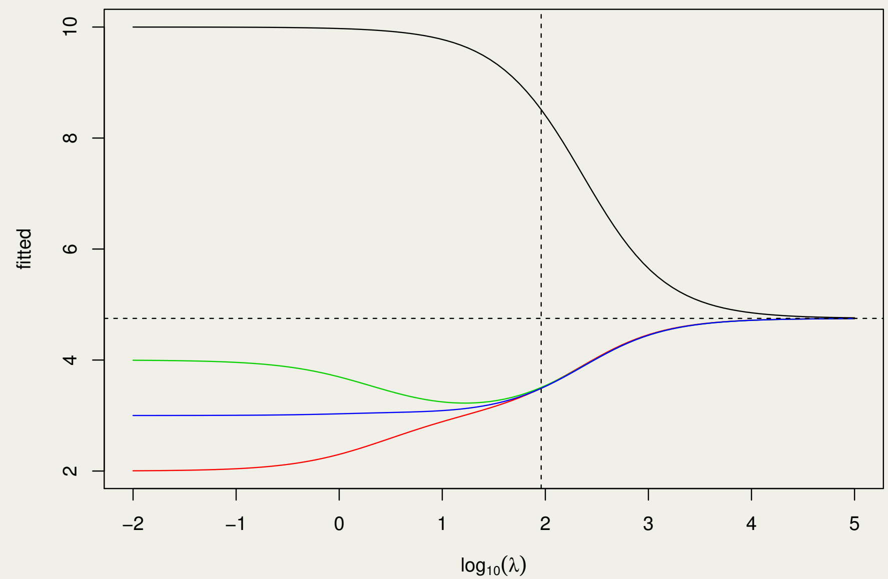
$$\theta \sim N\Big(0, \text{Diag}_{i=1,\ldots,k}\{(\lambda\xi_i)^{-1}\}\Big)$$

Network Regularized Regression

Toy example: $\mathbf{Y} = (10, 2, 3, 4)$, vertex 1 connected to triangle with vertices 2, 3, and 4, $w(u,v) \propto \exp\{-d(u,v)/2\} I[d(u,v) > 0]$, and

$$D = \begin{bmatrix} 0 & 10 & 0 & 0 \\ 10 & 0 & 5 & 3 \\ 0 & 5 & 0 & 2 \\ 0 & 3 & 2 & 0 \end{bmatrix}, \quad L = \begin{bmatrix} \mathbf{0.02} & -0.02 & 0 & 0 \\ -0.02 & \mathbf{0.85} & -0.22 & -0.61 \\ 0 & -0.22 & \mathbf{1.22} & -1 \\ 0 & -0.61 & -1 & \mathbf{1.61} \end{bmatrix}$$





Bayesian Network Regularized Regression

- Addressing the issue of abrupt rate changes,

$$Y_v \,|\, \zeta, \beta, Z_v \overset{\text{ind}}{\sim} \text{Po}\Big[\exp\Big(Z_v \zeta + (1 - Z_v)\mathbf{x}_v^\top \beta(v)\Big)\Big]$$
$$Z_v \,|\, \gamma \overset{\text{ind}}{\sim} \text{Bern}\Big[\text{logit}^{-1}\Big(\mathbf{u}_v^\top \gamma(v)\Big)\Big]$$

  where:

  - ζ is the “background” crime rate
  - Z_v codes for v being in a “hot zone”, also varying smoothly
  - Both β and γ are network indexed and assume a basis expansion as before

Bayesian Network Regularized Regression

- Addressing the issue of abrupt rate changes,

$$Y_v \,|\, \zeta, \beta, Z_v \overset{\text{ind}}{\sim} \text{Po}\left[\exp\left(Z_v \zeta + (1 - Z_v)\mathbf{x}_v^\top \beta(v)\right)\right]$$
$$Z_v \,|\, \gamma \overset{\text{ind}}{\sim} \text{Bern}\left[\text{logit}^{-1}\left(\mathbf{u}_v^\top \gamma(v)\right)\right]$$

  where:

  - ζ is the "background" crime rate
  - Z_v codes for v being in a "hot zone", also varying smoothly
  - Both β and γ are network indexed and assume a basis expansion as before

- Using basis coefficients, $\mathbf{x}_v^\top \beta(v) \to D_X(v)^\top \theta$ and $\mathbf{u}_v^\top \gamma(v) \to D_U(v)^\top \omega$,

$$Y_v \,|\, \zeta, \theta, Z_v \overset{\text{ind}}{\sim} \text{Po}\left[\exp\left(Z_v \zeta + (1 - Z_v) D_X(v)^\top \theta\right)\right]$$
$$Z_v \,|\, \omega \overset{\text{ind}}{\sim} \text{Bern}\left[\text{logit}^{-1}\left(D_U(v)^\top \omega\right)\right]$$

Bayesian Network Regularized Regression

- Quick methodological recap:
  - Network regularized regression as a building block,

$$Y_v \mid \theta \overset{\text{ind}}{\sim} \mathbf{F}\left[g^{-1}\left(D_X(v)^\top \theta\right)\right], \quad \theta \sim N\left(0, \lambda_\theta^{-1} \Omega(X, L_w(G))^-\right)$$

  - Change regions using latent network-indexed indicators Z and conditional responses

$$Z_v \mid \omega \overset{\text{ind}}{\sim} \text{Bern}\left[\operatorname{logit}^{-1}\left(D_U(v)^\top \omega\right)\right], \quad \omega \sim N\left(0, \lambda_\omega^{-1} \Omega(U, L_w(G))^-\right)$$

Bayesian Network Regularized Regression

- Quick methodological recap:
  - Network regularized regression as a building block,

$$Y_v \,|\, \theta \overset{\text{ind}}{\sim} \mathbf{F}\Big[g^{-1}\big(D_X(v)^\top\theta\big)\Big], \quad \theta \sim N\big(0, \lambda_\theta^{-1}\Omega(X, L_w(G))^-\big)$$

  - Change regions using latent network-indexed indicators Z and conditional responses

$$Z_v \,|\, \omega \overset{\text{ind}}{\sim} \text{Bern}\Big[\text{logit}^{-1}\big(D_U(v)^\top\omega\big)\Big], \quad \omega \sim N\big(0, \lambda_\omega^{-1}\Omega(U, L_w(G))^-\big)$$

- There are now two main practical problems:
  - How to define the hyper-parameters controlling the *smoothness* of β and γ?

Bayesian Network Regularized Regression

- Quick methodological recap:
  - Network regularized regression as a building block,

$$Y_v \,|\, \theta \overset{\text{ind}}{\sim} \mathbf{F}\Big[g^{-1}\big(D_X(v)^\top \theta\big)\Big], \quad \theta \sim N\Big(0, \lambda_\theta^{-1}\Omega(X, L_w(G))^-\Big)$$

  - Change regions using latent network-indexed indicators Z and conditional responses

$$Z_v \,|\, \omega \overset{\text{ind}}{\sim} \text{Bern}\Big[\text{logit}^{-1}\big(D_U(v)^\top \omega\big)\Big], \quad \omega \sim N\Big(0, \lambda_\omega^{-1}\Omega(U, L_w(G))^-\Big)$$

- There are now two main practical problems:
  - How to define the hyper-parameters controlling the *smoothness* of β and γ?
  - How to fit this model *efficiently* for large scale datasets?

Prior Elicitation: Guidelines

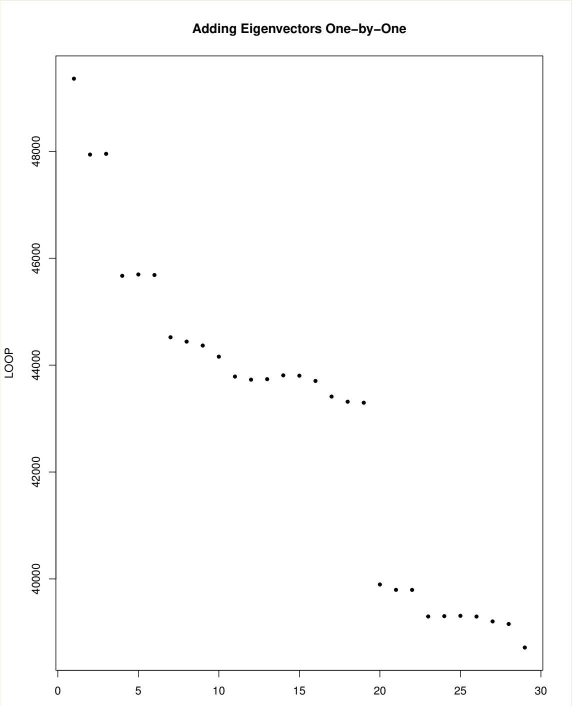
- Three main sets of hyper-parameters: $\theta \sim N\big(0, \lambda^{-1}\Omega(X, L_w(G))^{-}\big)$, where $\Omega(X, L_w(G)) := D_X^\top L_w D_X$ and D_X depends on $\Phi_{1:k}$

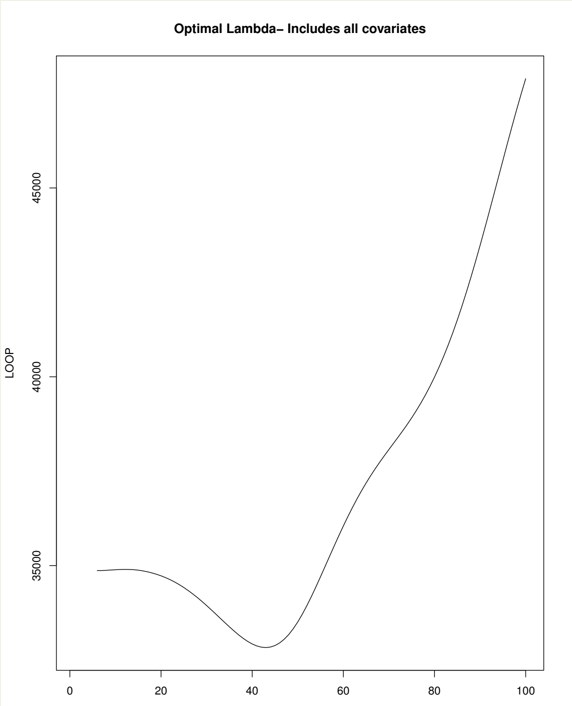
Prior Elicitation: Guidelines

- Three main sets of hyper-parameters: $\theta \sim N\Big(0, \lambda^{-1}\Omega(X, L_w(G))^-\Big)$, where $\Omega(X, L_w(G)) := D_X^\top L_w D_X$ and D_X depends on $\Phi_{1:k}$
- To define L_w we need a measure of similarity as weights in G: in our application, we use $w(u, v) \propto \exp\{-d(u, v)/\psi\}$ and set the "network range" ψ such that median similarity is 0.8

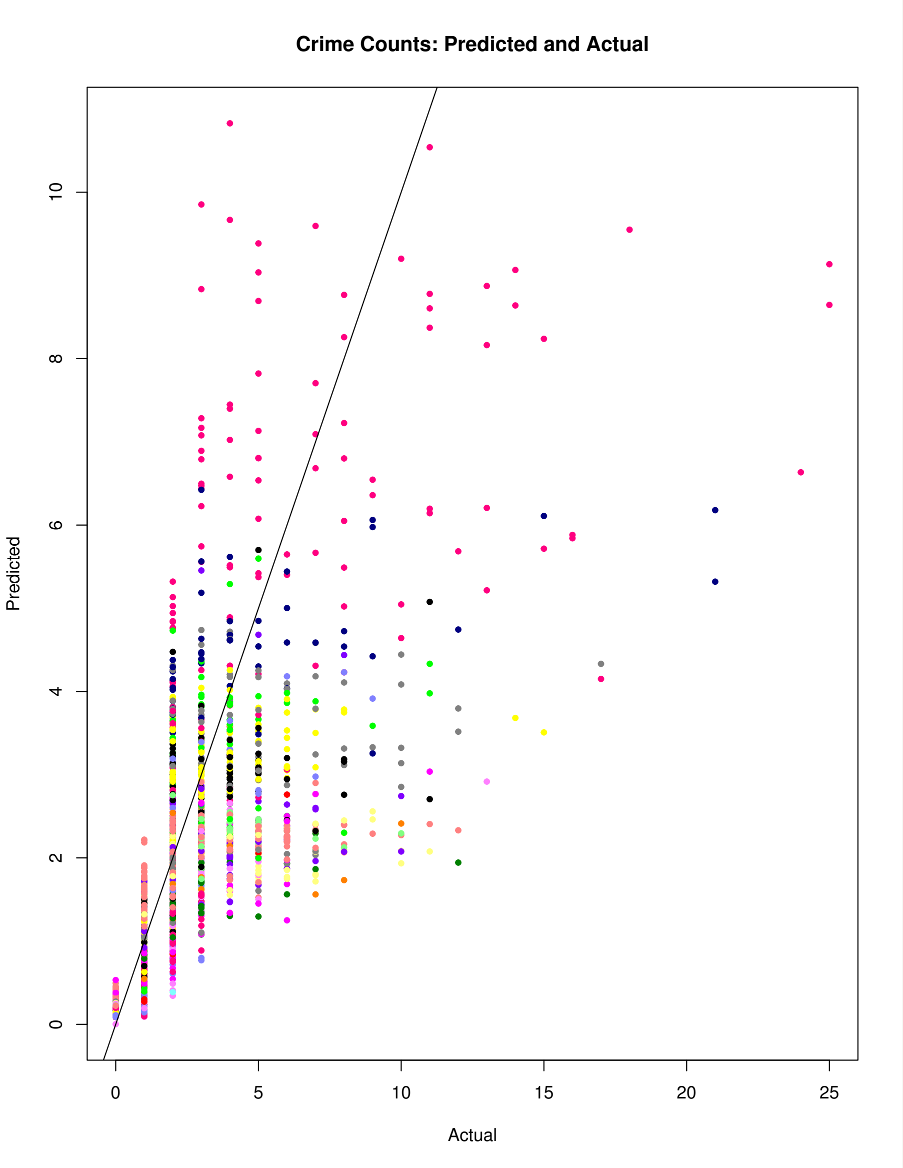
Prior Elicitation: Guidelines

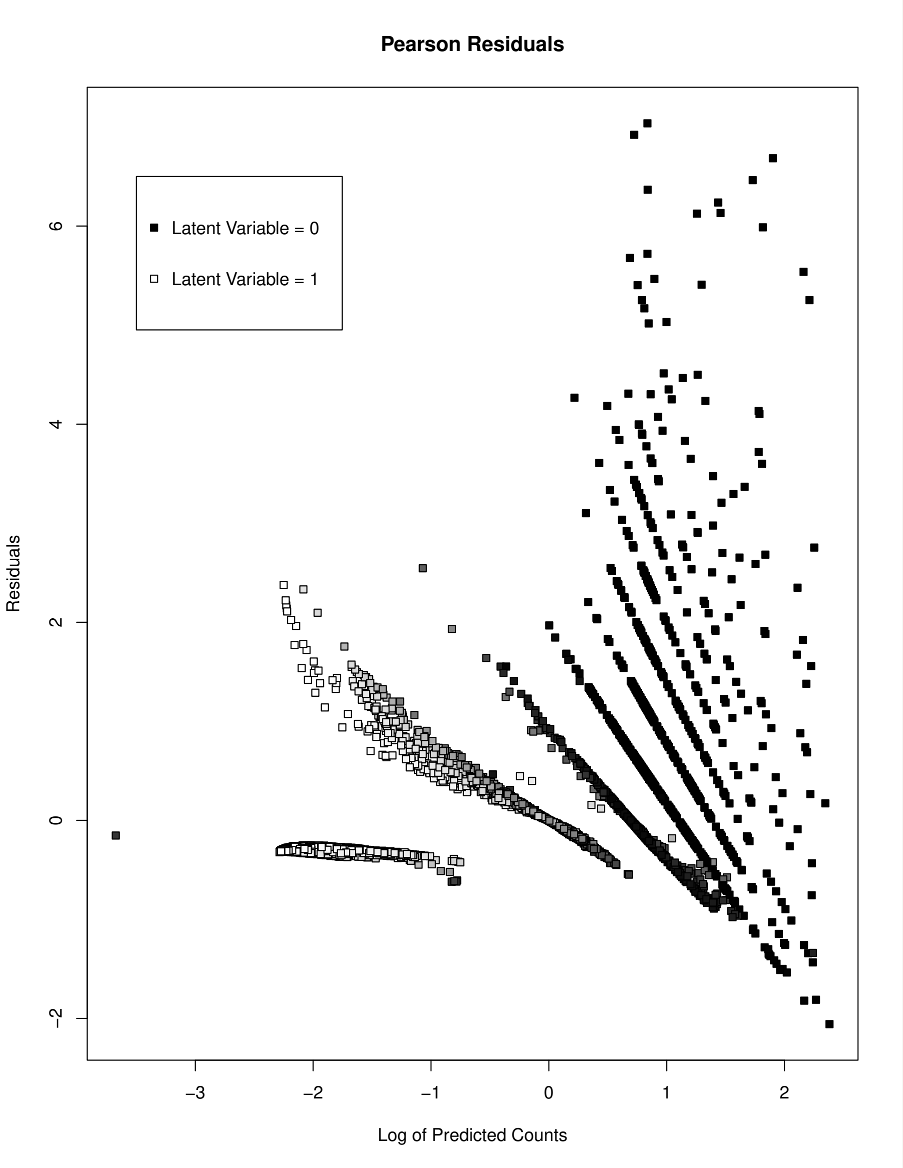
- Three main sets of hyper-parameters: $\theta \sim N\left(0, \lambda^{-1}\Omega(X, L_w(G))^{-}\right)$, where $\Omega(X, L_w(G)) := D_X^\top L_w D_X$ and D_X depends on $\Phi_{1:k}$
- To define L_w we need a measure of similarity as weights in G: in our application, we use $w(u,v) \propto \exp\{-d(u,v)/\psi\}$ and set the "network range" ψ such that median similarity is 0.8
- Penalty λ and basis rank k can be defined jointly using leave-one-out cross-validation via PRESS working residuals ("LOOP")



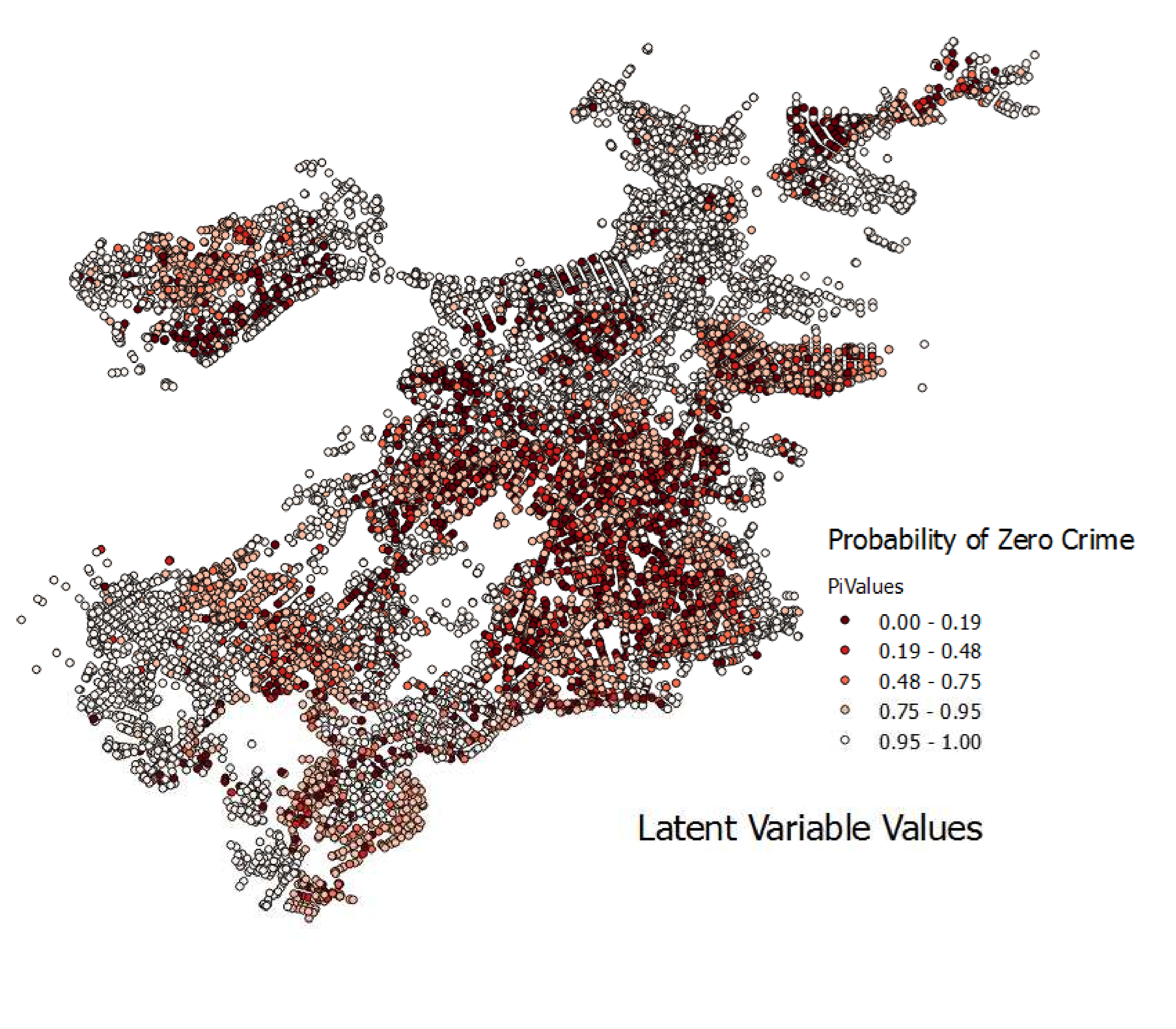


Model Fitting

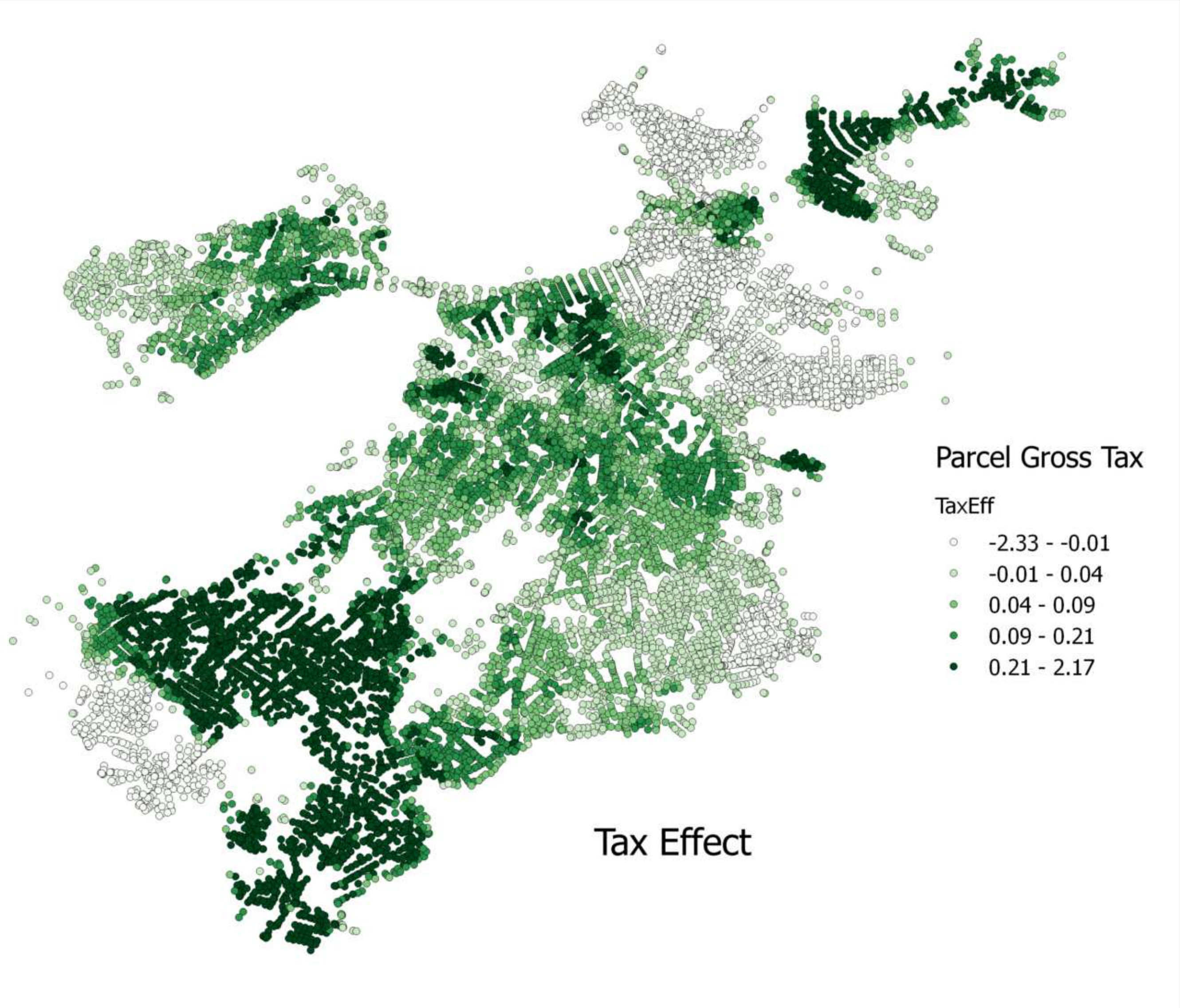




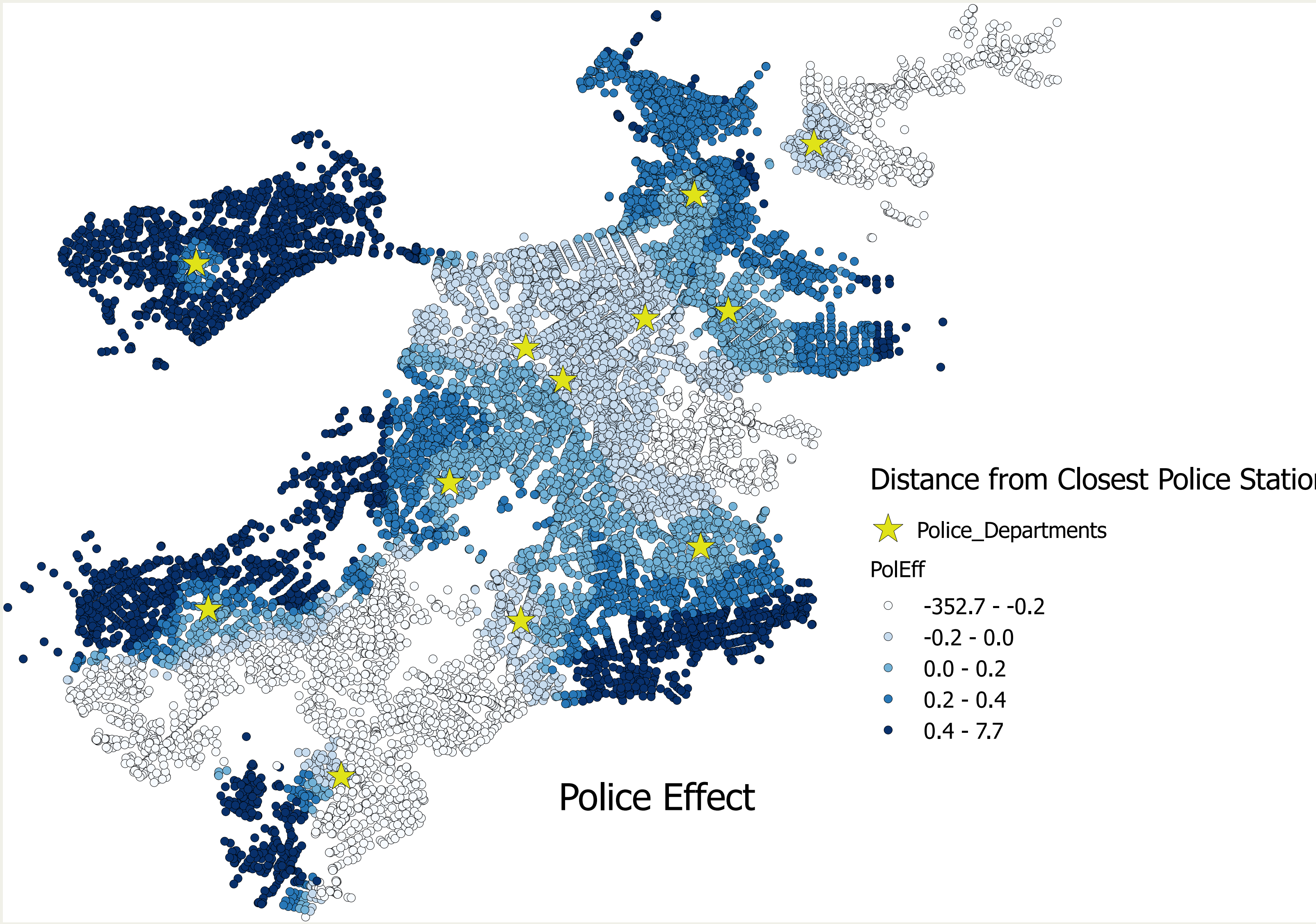
Model Fitting



Model Fitting



Model Fitting



Conclusions and future work

- Summary

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models
  - Main concern: representative models and computational efficiency

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models
  - Main concern: representative models and computational efficiency
- New challenges

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models
  - Main concern: representative models and computational efficiency
- New challenges
  - Refinements and extensions: dynamic model, basis selection, covariance structure

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models
  - Main concern: representative models and computational efficiency
- New challenges
  - Refinements and extensions: dynamic model, basis selection, covariance structure
  - Other applications in Biology, Epidemiology, and Engineering

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models
  - Main concern: representative models and computational efficiency
- New challenges
  - Refinements and extensions: dynamic model, basis selection, covariance structure
  - Other applications in Biology, Epidemiology, and Engineering
  - Bayesian *network* regression (for topology inference)

Conclusions and future work

- Summary
  - Network regularization is useful in a number of applications and can be more suitable than other types of regularization
  - Methodology can be used as a building block for more elaborated models
  - Main concern: representative models and computational efficiency
- New challenges
  - Refinements and extensions: dynamic model, basis selection, covariance structure
  - Other applications in Biology, Epidemiology, and Engineering
  - Bayesian *network* regression (for topology inference)

Thank you!