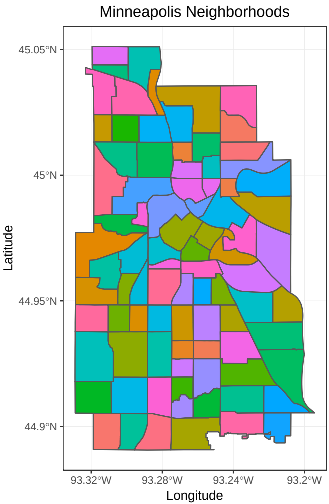
Modeling Crime: A Bayesian Hierarchical Approach to Modeling Minneapolis Crime Rates

Andrew DiLernia

Division of Biostatistics
University of Minnesota

May 1, 2019

Minneapolis Neighborhoods



Data

Crime Data

- Crime data for Minneapolis in 2011 and neighborhood shape files were obtained from the web site *Open Minneapolis* (2019).
- Consists of types of crimes committed, the dates and times of crimes, and also the location of crimes in longitude and latitude.
- Overall crime frequency was aggregated by day within each neighborhood for analysis.

Data

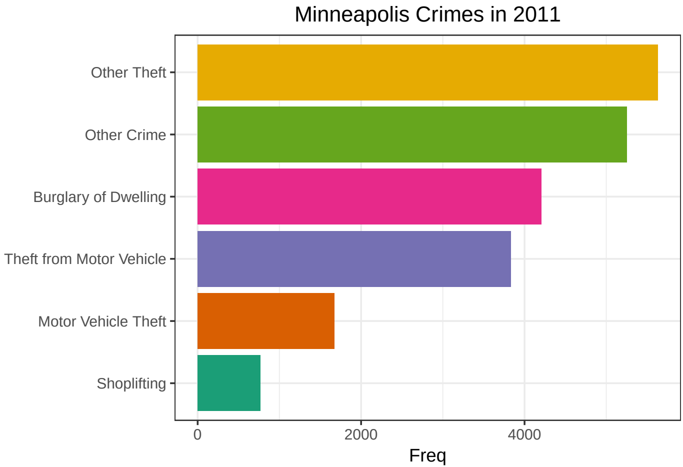
Neighborhood Data

- Data on Minneapolis neighborhoods from the 2010 US Census was obtained from *Minneapolis 2010 Census Results* (2019).
- Population size and percent of vacant housing units were used for analysis.

Weather Data

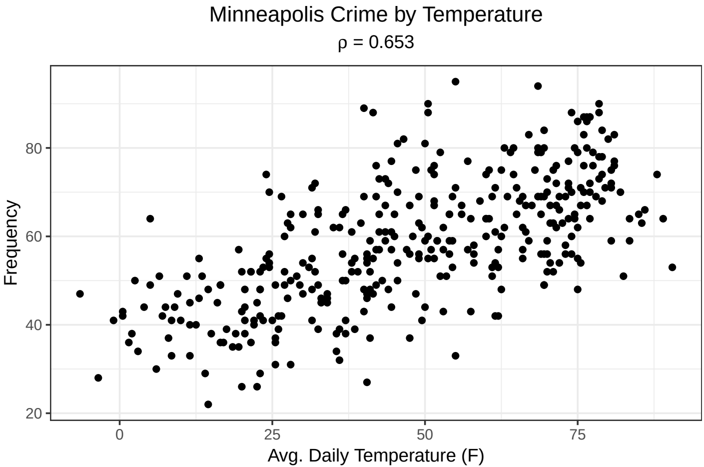
- Data on Minneapolis weather was obtained from *Minnesota DNR* (2019).
- Consisted of average daily temperature (daily max - daily min), amount of rain (inches), snow (inches), and snow accumulation (inches).

Crime Types

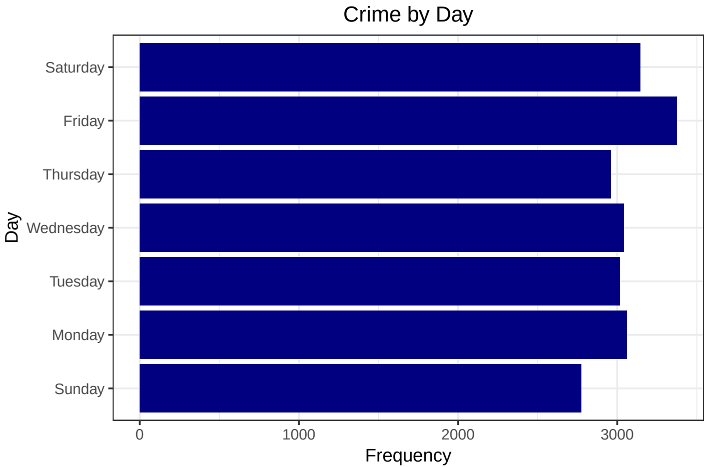


Minneapolis Crimes in 2011
Other Theft
Other Crime
Burglary of Dwelling
Theft from Motor Vehicle
Motor Vehicle Theft
Shoplifting
0
2000
4000
Freq

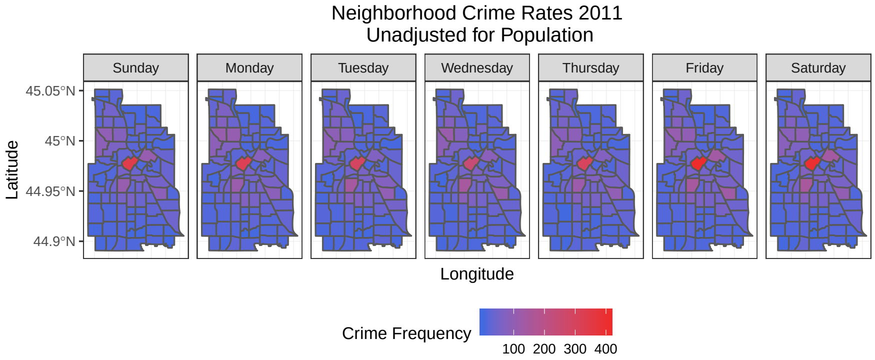
Temperature and Crime



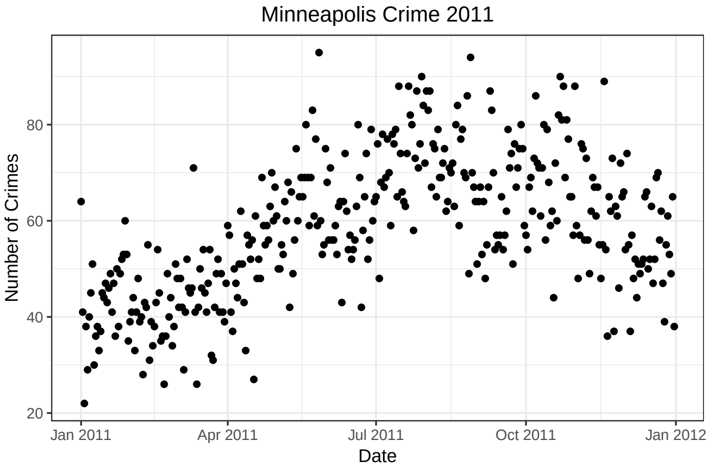
Crime by Day of the Week



Spatial Dependence



Temporal Dependence



Method

- Given the clear temporal and also possibly spatial dependence, I fit a spatio-temporal generalized linear mixed model using the *CARBayesST* package in R (Lee et al. 2018).
- Model assumes a spatio-temporal first-order auto-regressive process for areal unit data with a Poisson response as proposed by Rushworth et al. (2014).

$$Y_{kt} \sim \text{Poisson}(\lambda_{kt}) \text{ and } \log(\lambda_{kt}) = \mathbf{x}_{kt}^T \boldsymbol{\beta} + \phi_{kt}$$

where:

1. $k = 1, \ldots, K$ indexes the areal units (neighborhoods)
2. $t = 1, \ldots, T$ indexes the discrete time points (dates i.e. October 10th, 2011)
3. ϕ_{kt} are random effects which induce temporal correlation through its mean, and spatial correlation through its covariance.

University of Minnesota

Method

Specifically, the method assumes the following:

1. $\phi_t | \phi_{t-1} \sim \mathcal{N}_K \left(\rho_T \phi_{t-1}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_s)^{-1} \right)$ for $t = 2, \ldots, T$
2. $\phi_1 \sim \mathcal{N}_K \left(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_s)^{-1} \right)$
3. $\tau^2 \sim \text{Inverse-Gamma}(a, b)$
4. $\rho_s, \rho_T \sim \text{Uniform}(0, 1)$

- ρ_s is a spatial correlation parameter, ρ_T is a temporal correlation parameter, and $\mathbf{W}$ is a $K \times K$ adjacency matrix for the areal units (Minneapolis neighborhoods).
- The precision matrix $\mathbf{Q}$ is of the form proposed by Leroux et al. (2000), and is motivated by the CAR model.

Non-Spatial Model

Table 1: Coefficient estimates, exponentiated estimates, and standard errors for non-spatial Poisson GLM.

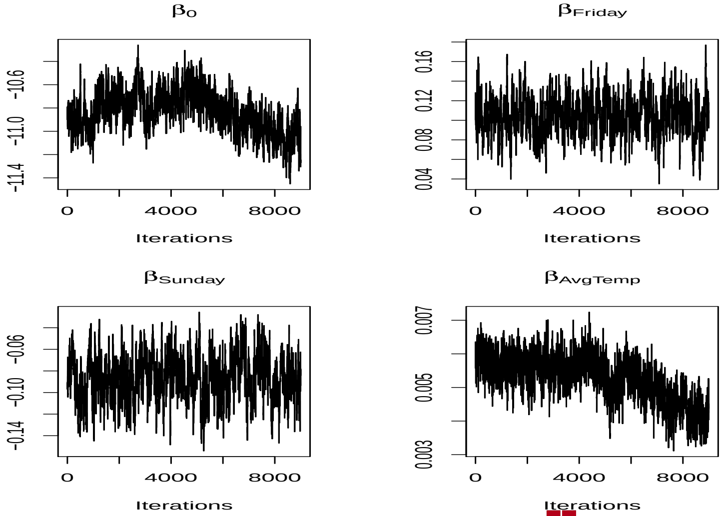
	Estimate	exp(Estimate)	SE
(Intercept)	-11.037	< 0.001	0.111
Friday	0.104	1.110	0.019
Sunday	-0.086	0.918	0.021
Temp	0.005	1.005	< 0.001
Snow	-0.037	0.964	0.011
Snow Accum.	-0.016	0.984	0.003
% Vacant	0.090	1.094	0.001
log(Pop)	1.142	3.134	0.012

Spatio-Temporal Model

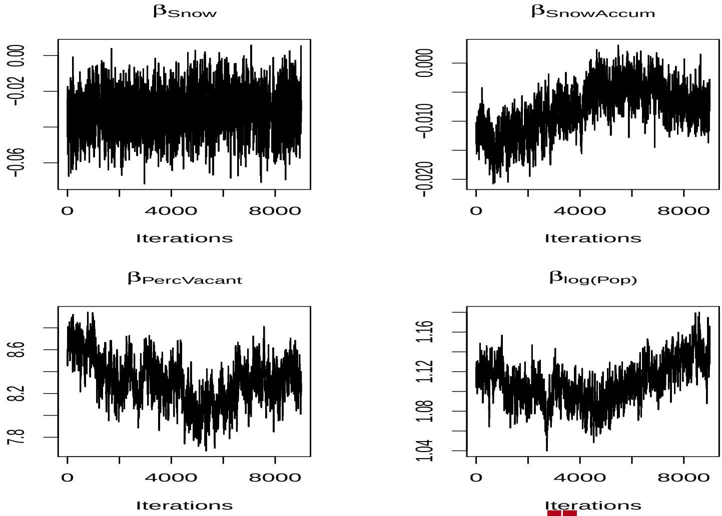
Table 2: Summary of posterior coefficient estimates from Metropolis sampling. The acceptance rate for the 9,000 β draws was about 45% (10,000 total with a 1,000 burn-in).

	Median	exp(Median)	2.5%	97.5%
(Intercept)	-10.834	< 0.001	-11.170	-10.493
Friday	0.105	1.110	0.064	0.144
Sunday	-0.087	0.917	-0.128	-0.046
Temp	0.005	1.005	0.004	0.006
Snow	-0.032	0.969	-0.055	-0.009
Snow Accum.	-0.007	0.993	-0.016	-0.001
% Vacant	0.083	1.087	0.079	0.087
log(Pop)	1.108	3.029	1.070	1.150

Trace Plots


β_0
-11.4
-11.0
-10.6
0
4000
8000
Iterations
β_{Friday}
0.04
0.08
0.12
0.16
0
4000
8000
Iterations
β_{Sunday}
-0.14
-0.10
-0.06
0
4000
8000
Iterations
$\beta_{AvgTemp}$
0.003
0.005
0.007
0
4000
8000
Iterations

Trace Plots

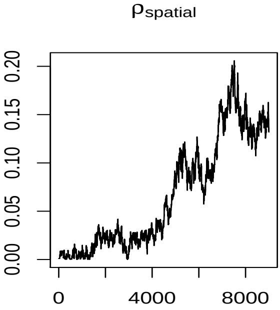


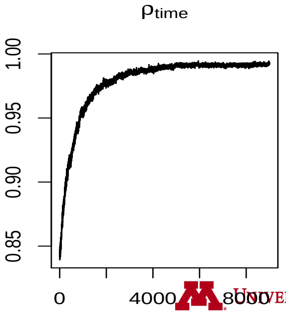
University of Minnesota

Correlation Estimates

Table 3: Summary of posterior correlation estimates.

	Median	2.5%	97.5%
ρ_{spatial}	0.040	0.001	0.173
ρ_{time}	0.989	0.892	0.993





Discussion

- Overall the non-spatial and spatio-temporal models yielded very similar estimates.
- The 95% credible intervals for the correlation parameters indicate a strong temporal dependence in the data, and a weak but still present spatial dependence.

Discussion

- Fridays have increased rates of crime, and Sundays have deflated crime rates.
- More vacant housing units in neighborhoods is associated with higher rates of crime.
- More snow is negatively associated with crime rates for a given day.

References I

Lee, D., Rushworth, A. & Napier, G. (2018), 'Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package', *Journal of Statistical Software* **84**(9), 1–39.

Leroux, B. G., Lei, X. & Breslow, N. (2000), *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, Springer-Verlag, New York, chapter Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence, pp. 179–191.
URL: *http://dx.doi.org/10.1007/978-1-4612-1284-3*$_4$

Minneapolis 2010 Census Results (2019).
URL: *http://www.minneapolismn.gov/census/2010/index.htm*

Minnesota DNR (2019).
URL: *https://www.dnr.state.mn.us/climate/twincities/listings.html*

Open Minneapolis *(2019).*
URL: *http://opendata.minneapolismn.gov/*

Rushworth, A., Lee, D. & Mitchell, R. (2014), 'A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London', Spatial and Spatio-temporal Epidemiology **10***, 29–38.*