# Spatio-Temporal Point Processes for Crime

## STOPPER

Nate Garton

April 9, 2018

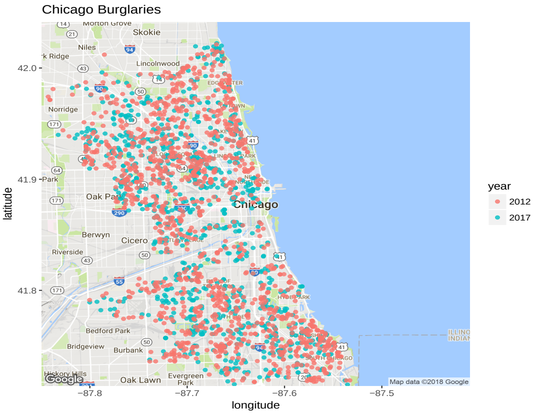
# Current trends in quantitative criminology

1. Predictive policing
   a. PredPol (2012)
2. Modeling and inference
   a. Crime from multiple data sources/geographic regions/cities/areas within cities/multiple offense types
   b. Push for more modeling of crime trends in the 2018 *Annual Review of Criminology*

# Objectives

1. Develop a Bayesian modeling and estimation framework for the analysis and prediction of urban crime.
2. With this, we hope to provide a way to:
    a. analyze the factors impacting urban crime and quantify their effects
    b. analyze shifts in the spatial distribution of crime over time
    c. forecast crime hotspots

# Chicago


Chicago Burglaries
latitude
longitude
42.0
41.9
41.8
-87.8
-87.7
-87.6
-87.5
year
2012
2017
Map data ©2018 Google

# What is a point process?

1. A probability distribution describing a random number of randomly located points over a domain $D$.
2. For a spatial point process, $D \subset \mathbb{R}^2$. For a spatio-temporal point process, $D \subset \mathbb{R}^2 \times \mathbb{R}^+$.
3. For any measurable subset of that domain, define the distribution of the number of events happening in that subset.
4. If we take that distribution to be Poisson with mean proportional to the area/volume of the subset, we have a homogeneous Poisson process.

# Homogeneous Poisson Process

1. Formally, for any $A \in \mathcal{B}(D)$, define the number of events in $A$, say $Y(A)$ to have pmf $p(Y(A) = y) = \frac{1}{y!}(\int_A \lambda m(dx))^y e^{-\int_A \lambda m(dx)}$
2. More interpretably, $p(Y(A) = y) = \frac{1}{y!}(volume(A)\lambda)^y e^{-volume(A)\lambda}$
3. $\lambda$ is called the intensity function.

# Nonhomogeneous Poisson Process

1. A more flexible model allows the intensity to vary over the domain.
2. $\lambda(s,t)$ is now a function of the spatial location and the time.
3. Introduce covariates into the model.
   a. $log(\lambda(s,t)) = X(s)^\top\psi + X(t)^\top\tau + X(s,t)^\top\gamma$

# What kinds of covariates might we consider?

1. Spatial variables
    a. Distance to certain locations/landmarks (maybe tourist attractions, bars, public transit stops)
2. Temporal variables
    a. Weather
    b. State of economy
3. Spatio-temporal variables
    a. Locations and times of 911 calls
    b. Population density

# Log Gaussian Cox Process

1. Usually people want more flexibility
2. Now, intensity function becomes
   $log(\lambda(s,t)) = X(s)^\top \psi + X(t)^\top \tau + X(s,t)^\top \gamma + Z(s,t)$
3. $Z(s,t) \sim \mathcal{GP}(0, C((s,t),(s',t')))$

# Offset for police presence

1. Instead of modeling the log intensity, maybe we want to model the log of the ratio of the intensity of crime over intensity of "police presence", $p(s, t)$.
2. $log(\frac{\lambda(s,t)}{p(s,t)}) = X(s)^\top \psi + X(t)^\top \tau + X(s, t)^\top \gamma + Z(s, t)$

# Dependencies between crime types

1. Do burglaries attract or repel robberies?
2. Multivariate point process

# Shiny application

1. Can we make a Shiny app that takes as input...
   a. a dataset of crime (locations, time, offense type)
   b. user specified locations (or locations and times) as spatial (or spatial and temporal) variables
   c. ...
2. and runs an analysis using the above LGCP model structure
3. provides the ability to make forecasts

# Challenges

1. computation
2. access to police presence data
3. learning how to automate pulling and combining data from multiple data sources
4. account for anonymized data/measurement error in space and time