Modeling of Crime Data to Detect Social and Spatial Proximity

Claire Kelling*

Advisors: Murali Haran*, Corina Graif**

Penn State University
Department of Statistics*, Department of Sociology and Criminology**

STAT 544: Categorical Data Analysis

Topics

1. Introduction
2. Data Sources
   - Crime Data
   - Demographic Data
   - Social Proximity Data
3. Literature Review Plan
4. Analysis Plan

Motivation

Geographic proximity has been studied over many years

- mostly focusing on identifying hotspots in certain communities
- led to many controversial policing strategies, such as predictive policing, which was referenced repeatedly in "Weapons of Math Destruction" (O'Neill, 2016).

Our Approach:

- Demographics
- Geographic Proximity
- Social Proximity

Data Source

POLICE DATA INITIATIVE

POLICE DATA INITIATIVE

HOME DATA & AGENCIES JOIN THE INITIATIVE RESOURCES FAQ ABOUT CONTACT

Name	City	State	Link	Title
Baltimore Police Department	Baltimore	MD	https://goo.gl/KvCHMn	Calls for Service
Bloomington Police Department	Bloomington	IN	https://goo.gl/ewqJg2	Calls for Service
Burlington Police Department	Burlington	VT	https://goo.gl/02x19G	Calls for Service
Chandler Police Department	Chandler	AZ	https://goo.gl/xbLPce	Calls for Service
Charleston Police Department	Charleston	SC	https://goo.gl/Mk9idR	Calls for Service

Police Data Initiative

Call for Service (CFS) data:

- 911 Call
- Officer-Initiated OR Call-Initiated
- collected in computer-aided dispatch systems

Variation in Data:

- contain sensitive call types?
- contain all emergency and non-emergency calls?

Dataset Size

- Usually a small number of variables (longitude, latitude, time, brief description)
- Usually **many** calls (4 million+ in some cases)

Detroit Police Data Initiative Dataset

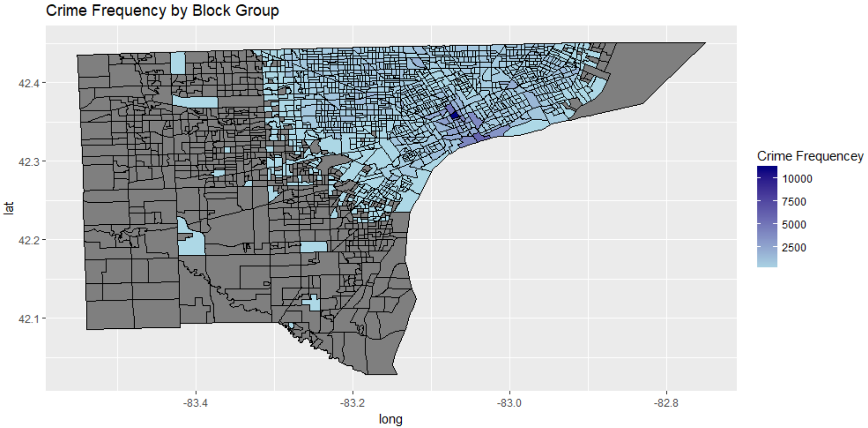
POLICE DATA INITIATIVE

5.5+ million observations with complete information (09/16-present)

Detroit Variables:

- Measures of Response Time:
  - Intake, Dispatch, Travel, Time On Scene, Response Time
- Priority (1-5)
- Call Code, Call Description, Category
- Call Time/Date
- Officer Initiated (Yes/No)
- Neighborhood
- Longitude/Latitude

Exploratory Data Analysis- PDI

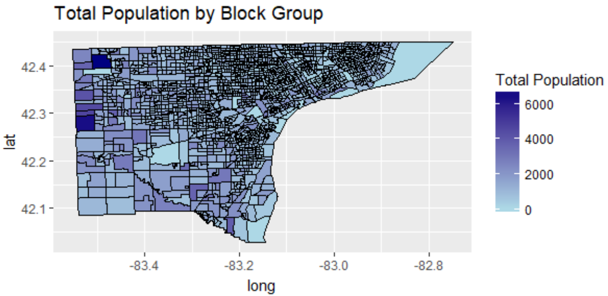


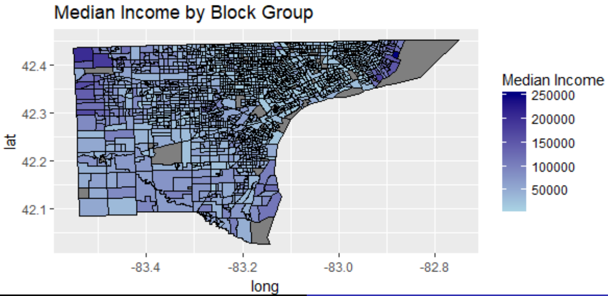
American Communities Survey

Managed through the US Census Bureau, available through 2015

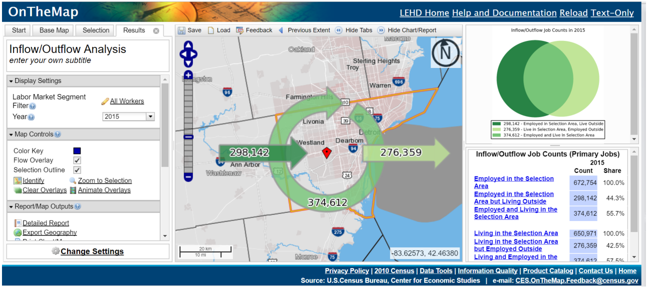
Variables for all 1,822 block groups:

- median income
- median age
- percentage female
- unemployment rate
- total population
- measure of racial diversity (?)


Total Population by Block Group
Total Population
6000
4000
2000
0
42.4
42.3
42.2
42.1
lat
-83.4
-83.2
-83.0
-82.8
long


Median Income by Block Group
Median Income
250000
200000
150000
100000
50000
42.4
42.3
42.2
42.1
lat
-83.4
-83.2
-83.0
-82.8
long

What is Social Proximity?



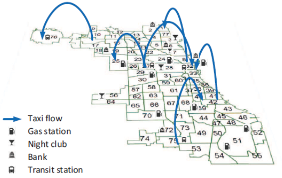


Figure 1: An illustration of various types of features we used in Chicago. The POI distribution across community areas reflects profiles of the region functionality. The taxi flow connects non-adjacent regions and act as "hyperlinks" on the space.

Social Proximity Data

More Census data

- used by OnTheMap (a Census App)
- LEHD Origin-Destination Employment Statistics (LODES)
- LEHD = Longitudinal Employer-Household Dynamics

Index of /data/lodes/LODES7/mi/od

Name	Last modified	Size	Description
Parent Directory		-	
mi_od_aux_JT00_2002.csv.gz	21-Sep-2017 20:57	347K	
mi_od_aux_JT00_2003.csv.gz	21-Sep-2017 20:56	341K	
mi_od_aux_JT00_2004.csv.gz	21-Sep-2017 20:57	320K	
mi_od_aux_JT00_2005.csv.gz	21-Sep-2017 20:57	323K	
mi_od_aux_JT00_2006.csv.gz	21-Sep-2017 20:57	319K	
mi_od_aux_JT00_2007.csv.gz	21-Sep-2017 20:56	337K	
mi_od_aux_JT00_2008.csv.gz	21-Sep-2017 20:56	500K	
mi_od_aux_JT00_2009.csv.gz	21-Sep-2017 20:57	419K	
mi_od_aux_JT00_2010.csv.gz	21-Sep-2017 20:56	447K	
mi_od_aux_JT00_2011.csv.gz	21-Sep-2017 20:57	444K	
mi_od_aux_JT00_2012.csv.gz	21-Sep-2017 20:56	487K	
mi_od_aux_JT00_2013.csv.gz	21-Sep-2017 20:57	512K	
mi_od_aux_JT00_2014.csv.gz	21-Sep-2017 20:57	553K	
mi_od_aux_JT00_2015.csv.gz	21-Sep-2017 20:56	537K	
mi_od_aux_JT01_2002.csv.gz	21-Sep-2017 20:56	330K	
mi_od_aux_JT01_2003.csv.gz	21-Sep-2017 20:56	323K	

Literature Review- Block Group Analysis

There are two main areas in the current literature:

- Analysis of Crime Data:
  - weighted spatial regression, Wang et al (2016) [1]
  - critique of spatial regression in econometric applications, Anselin (2002) [2]
  - Social ecology theory of crimes, Anselin et al (2000) [3]
- Statistical techniques for Areal Data:
  - CAR Model (conditional autoregressive)- CAR structure on random effects but GLMM framework
  - Chapters 12 and 13 of Handbook of Spatial Statistics [4]
  - Chapter 4 of Banerjee et al [5]

Analysis

There will be two main focuses of this project, with an additional possibility:

1. Model the aggregated crime rates (counts) for the block groups, as related to demographic variables
2. Model the response time to incidents at the point level, particularly in according to the categorical variable of priority
3. Consider neural models with negative binomial spiking, Pillow and Scott (2012) [6]

References

[1] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 635–644. ACM, 2016.

[2] Luc Anselin. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267, 2002.

[3] Luc Anselin, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita. Spatial analyses of crime. *Criminal justice*, 4(2):213–262, 2000.

[4] Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

[5] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. Crc Press, 2014.

[6] James Scott and Jonathan W Pillow. Fully bayesian inference for neural models with negative-binomial spiking. In *Advances in neural information processing systems*, pages 1898–1906, 2012.