Claire Kelling

STAT 597: Final Project Proposal

In the Summer of 2017, I worked as a Graduate Research Fellow in the Social and Decision Analytics Lab in Arlington, Virginia. I worked closely with Dr. Josh Goldstein, a recent graduate from Penn State, and Dr. Gizem Korkmaz, a research fellow in the lab. The three of us, with other students and faculty, partnered with the local Arlington Police Department to model crime data to see if there was any spatial correlation in crime, and to see if certain events affected crime rates, both in time and space. This opportunity served as my introduction to this work, and I will continue collaborating with the lab throughout the next year.

Arlington’s police data is governed by several data usage agreements. Therefore, the main source of data for this project will be the Police Data Initiative, which houses publicly accessible data for many large cities across the United States. We will first focus on Detroit, but will plan to complete our analysis for more than one major city.

**Motivation:**

The motivation of our work is mainly being drawn from Dr. Corina Graif’s current work in criminology. She is interested in modeling the diffusion of crime between neighborhoods. In the current literature, neighborhoods are often treated as closed systems, even though we know this is not the case. We are interested in both the geographic and social proximity between neighborhoods. Geographic proximity has been studied over many years, mostly focusing on identifying hotspots in certain communities. This has led to many controversial policing strategies, such as predictive policing, which was referenced thoroughly in “Weapons of Math Destruction” (O’Neill, 2016). We seek to take this research in a new direction by not just focusing on hotspots, but what other ties might exist between neighborhoods, in order to create more effective interventions. Social proximity is a relatively new topic in the criminology literature, and Dr. Graif has experience with commuting data that could establish social ties between communities. We will also consider demographics in our modeling efforts. If we are able to, we would like to subset our data to focus on violent crime. We hypothesize that violent crime will be more affected by social and spatial dynamics than other crime types, such as property-related crimes.

So, to summarize, we are considering three main data sources to model crime:

1) Demographics (American Communities Survey, from Census)

2) Geographic proximity

3) Social proximity (LODES Data, from Census)

**Project Goals:**

Over the next couple weeks, there will be two main parts to my project:

1. We will create areal models based off of the geographic and social proximity (separately) and compare our results. We will make conclusions on the strength of the models and how they give different results. From initial exploration, we already see that without spatial information, we get a very poor estimate of crime.
2. Time permitting, we will also create a point process model where the dependent variable is the response time to the crime, or the time from when the call was received until the time when the officer was on the scene.