Methods and Initial Results

Jie Heng

Research Question:

1. Is there a relation between the number of coffee shops and crime associated with gentrification in a city with high ethno-racial diversity (New York) from 2006 to 2017?
2. Whether the number of coffee shops of neighborhoods and other features of neighborhoods (race, income) are associated with violent and property crimes?

Data and Data Collection

The research uses data collected from nation website and database: New York City Open Data, New York City Police Department (NYPD), United States Census Bureau and New York Government. There are four parts of data: neighborhood, crime, coffee shops and census.

The paper focuses on the influence of gentrification on crime and the changes of the number of coffee shops. I would use neighborhood as units since gentrification is a process of a renovation of lower-income neighborhood (Hamnett, 1984: 284) and previous studies also study gentrification by analyzing it in neighborhoods (Ley, 1996; Lee, 2010; Barton and Gruner, 2014; Kreager, Lyons, & Hays, 2011). The data, ZIP Code Definitions of New York City Neighborhoods, was obtained from New York government website, which lists the zip codes of all neighborhoods in NYC.

As for crime data, the paper will use NYPD Complaint Data Historic and NYPD Complaint Data Current YTD obtained from NYPD website. The historic crime data includes all valid felony, misdemeanor, and violation crimes reported to the NYPD from 2006 to 2016, with 5,580,035 observations. The current crime data only includes crime records in 2007 with 468,761 observations.

To obtain the counts of coffee shops in neighborhoods, I use Sidewalk_Cafes(legally operating), Sidewalk Café Licenses and Applications and Sidewalk_Cafes collected from New York Open Data.

Theses three dataset features businesses/individuals holding a DCA license so that they may legally operate coffee shops in New York City. However, these datasets only include one record for chain stores, impossible to trace all branches of coffee shops from it. An alternative to NYC Open Data's data is crawled from Google Maps using Google Earth software. Google Earth saves all records of coffee shops in neighborhoods on Google Maps to KML files. This method allows me to get a complete data of coffee shops in NYC.

Census data, US 2000 Census and US 2010 Census, was collected from United States Census Bureau. These two censuses have detailed information about population, age, business and industry, education, governments, housing, income, origins and languages, poverty, race and Hispanic origin, veterans for every zip code in New York.

Analytic Strategy

There are two stages of this study. The first stage is conducting a descriptive analysis of the spatial changes of gentrification by ploting the changes of the number of coffee shops, violent and property crimes, some fields of census data on New York map. The goal is to display the trend of gentrification and observe the relations between variables. The second stage is predicting violent and property crimes in neighborhoods using coffee shop, census and crime data. Details of models and method are discussed in the following parts.

Model

Traditional research on gentrification uses Poisson model with neighborhood fixed effects (Kreager et. al, 2011; Papachristos et al, 2011). However, fixed effects models cannot control for variables that vary over time. Although we can introduce dummy variables for time or space units to the models, the more variables we introduce, the more noises we have-- the effects of group-level predictors are confounded with the effects of the group dummies. Recently, scholars have started to use multilevel model to analyze data that have a hierarchical or clustered structure. In a multilevel model, the effects of both types of

variable can be estimated, as groups in the sample are treated as a random sample from a population of groups. Multilevel models could recognize data hierarchies by allowing for residual components at each level in the hierarchy.

This study employs log linear multilevel modeling to examine the relationship between crime, gentrification and coffee shop. The modeling fits the data structure that coffee shop is nested in neighborhood. Employing multilevel modeling could minimize statistical issues associated with trying to handle multilevel scales with Ordinary Least Squares regression. In addition, log linear multilevel model is ideal for analyzing count data (Long and Freese 2006) and are also suited for analyzing crime data (Osgood 2000).

Level 1

$$n_{ij} \sim Poisson(\theta_{ij}) \tag{1}$$

$$log\theta_{i,j} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \tag{2}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \tag{3}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{4}$$

Coffee shops and neighborhoods are level 1 and level 2 in the multilevel models respectively. The above equations present the structure of the multilevel model. This model allows intercepts to vary across groups (neighborhood).

In level 1, $log\theta_{i,j}$ refers to the dependent variable, the counts of violent and property crime; X_{ij} refers to the predictors, the number of coffee shops; β_{0j} refers to the intercept of the dependent variable in

group j (level 2); β_{1j} refers to the slope for the relationship in group j (level 2) between the level 1 predictor and the dependent variable; ϵ_{ij} refers to random errors of prediction for the level 1 equation.

In level 2,u_{0j} refers to the overall intercept; W_j refers to the level 2 predictor (see details in the next part); γ_{01} refers to the overall regression coefficient , between the dependent variable and the level 2 predictor; γ_{00} refers to the random error component for the deviation of the intercept of a group from the overall intercept; γ_{10} refers to the overall regression coefficient between the dependent variable and the level 1 predictor; u_{1j} refers to the error component for the slope.

Methods

Based on 2000 census and 2010 census, we will create three-year period (four periods in total) using OLS regression to track changes in the neighborhood (Kreager et. al, 2011). We will regress the total number of neighborhood violent and property crime, respectively, on census factors, the number of coffee shops, and the lagged levels of neighborhood crime using multilevel model. Lagged levels of crime are added to (2) equation. As there are more than 15 variables in census data and we are not sure which of them relates to gentrification and crime, we would use five sub-groups of variables in census data as predictors in level 2 and choose one group that makes the model with higher prediction power.

Five groups of variables are:

(1) Education(bachelor degree), gentrifiers (residents move to a neighborhood); new housing built in 5 years, log of mean family income, percent of population that is Black, White, Hispanic, Asian, foreign born,

(2) Percent of population that is Black, White, Hispanic, Asian, foreign born;

(3) Log of mean family income, education(bachelor degree), gentrifiers (residents move to a neighborhood);

(4) Education (bachelor degree), log of mean family income;

(5) Gentrifiers (residents move to a neighborhood), percent of population have not moved for five years, new housing built in 5 years.

Summary Statistics for Key Variable

Table 1: The Number of Coffee Shop from 2006 to 2017 (four time periods)

Year	count
2006-2008	716
2009-2011	1339
2012-2014	2010
2015-2017	2075

Tentative Results

COFFEE AND CRIME IN NEW YROK, 2006 TO 2017: DESCRIPTIVE ANALYSIS

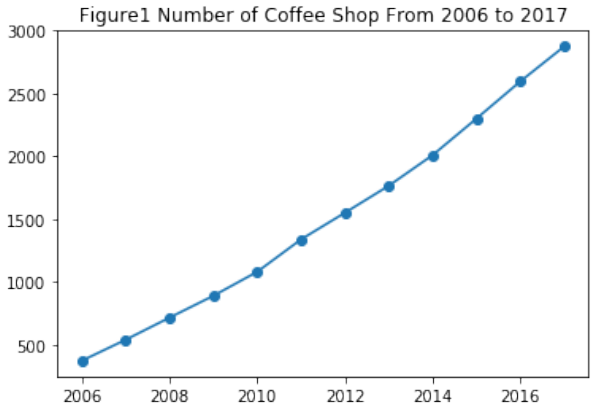


Figure2 Coffee Shop (2006)

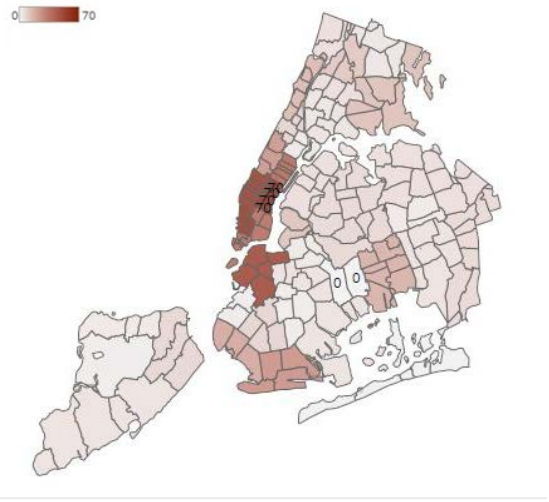
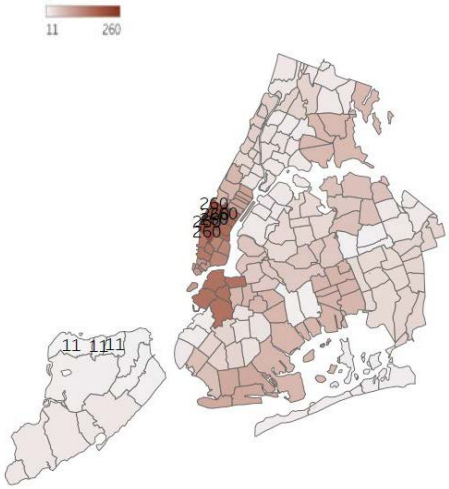


Figure3 Coffee Shop (2017)



The number of coffee shops in NYC keeps increasing from 2006 to 2017. Compared the distribution of Figure2 and Figure3, it is obvious that the color of neighborhoods in the south of NYC near the sea, Queen borough become darker and it might relate with gentrification.

Reference

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*, 7(2), 127-150.

Erhardt, V., & Czado, C. (2009). Generalized estimating equations for longitudinal generalized Poisson count data with regression effects on the mean and dispersion level. *Preprint.*

Kennedy, Maureen. and Leonard, Paul. "Dealing with Neighborhood Change: A Primer on Gentrification and Policy Changes." *Discussion Paper Prepared for the Brookings Institution Center on Urban and Metropolitan Policy*, 2001.