An Empirical Analysis of Stop-and-Frisk in New York City

Md.Afzal Hossain
NYC College of Technology
afzal.cuny@gmail.com

Khanna Pugach
Baruch College
kpugach@gmail.com

Derek Sanz
Brooklyn College
lkb_su13@hotmail.com

Siobhan Wilmot-Dunbar
Pace University
s.wilmot.dunbar@gmail.com

1. INTRODUCTION

Stop-and-frisk is a practice used by the New York City Police Department in which officers stop and question pedestrians, and then potentially frisk them for weapons and other contraband. Between 2006 and 2012, approximately 530,000 people were stopped annually. However, only six percent of these stops led to an arrest, which suggests that hundreds of thousands of innocent individuals were stopped each year. Moreover, the vast majority (87%) of stopped individuals were black or Hispanic. While there are undoubtedly social benefits of stop-and-frisk—including crime reduction and creating an atmosphere of safety—we focus here on assessing the social cost of the policy. Namely, we quantify the burden that is placed on individuals of various demographic groups by estimating yearly per capita stop rates. To mitigate the social costs of stop-and-frisk, we then develop two statistical methods to aid officers in making optimal stop decisions.

Data. We use two primary datasets in our analysis. First, we use the stop-and-frisk data publicly released by the NYC police department, which lists information about each of the 3.8 million documented stops from 2006–2012. Second, we use population statistics released by the Census Bureau. The Census data are broken into two parts: the Public Use Microdata Sample (PUMS), and block-level data. The PUMS data give detailed individual-level statistics for a random sample of the population; and the block-level data give aggregate demographic information by location.

2. RESULTS

Measuring Social Cost. In order to understand and analyze the social cost of stop-and-frisk, it must first be quantified. Here we define the social cost to be the average number of stops of an "innocent" individual (i.e., someone who is stopped but not arrested), normalized by the total number of people in the population, and further broken down by demographic group. Specifically, we define yearly per capita stops (YPCS) as follows:

$$\text{YPCS} = \frac{\text{number of stops of innocent people per year}}{\text{number of people in the population}}$$

The number of stops of innocent people is obtained from the NYPD stop-and-frisk data, while the population count is obtained from the Census PUMS data. Figure 1 shows the yearly per capita stops for each major race group recorded in the stop-and-frisk data, across all the recorded ages of those stopped, and further broken down by sex. It can be seen that females have drastically lower yearly per capita

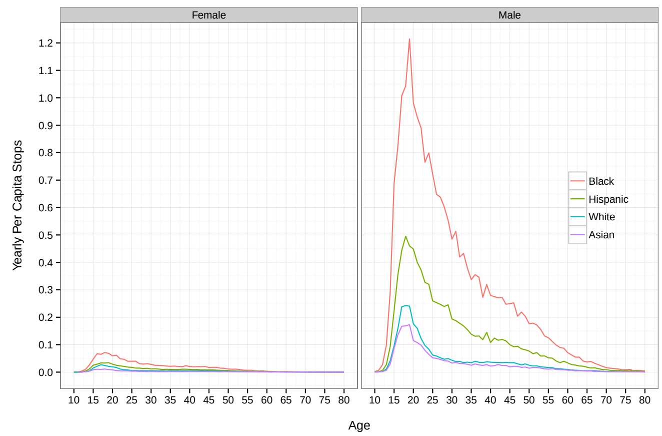
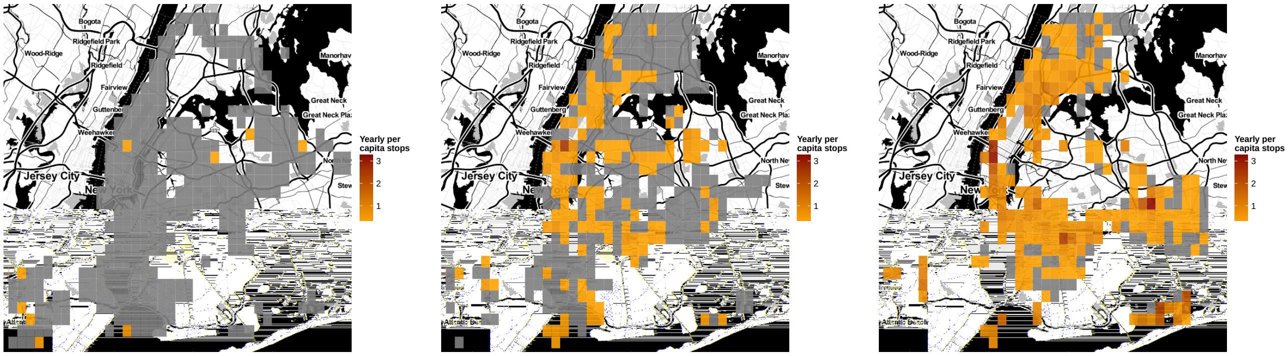


Figure 1: Yearly per capita stop rate by demographic group.

stops than males. Nonetheless, for both men and women, black individuals have the highest yearly per capita stops. The highest ranking group is 19 year-old black males, with a stop rate of 1.2. This means such individuals are on average stopped more than once each year. In contrast, a 19 year-old white male is only expected to be stopped about 0.2 times. Finally, a 19 year-old black female is only expected to be stopped about 0.06 times each year.

We next focus on young (18–29 year-old) males, and compute stop rates by race and location. To measure the stop rate by location, we partition the city into one kilometer by one kilometer squares. The results are shown in Figure 2, where the three panels show stop rates for whites, Hispanics, and blacks respectively. The colors indicate the stop rate, with grey squares showing areas where the stop rate is less than 0.25, indicating that it is relatively unlikely for an individual to be stopped. As is readily apparent from the figure, throughout the city stop rates for whites are considerable lower than for Hispanics, which are in turn lower than for blacks. For example, The highest stop rate for young white males is 0.5, in Coney Island, whereas the highest stop rate for young black males is 3.2 in Jamaica, Queens.

Improving Stop Decisions. In order to reduce the burden placed on innocent individuals being stopped, we develop statistical models to help officers make "better" stops. In particular, our goal is to develop techniques for reducing the number of innocent people stopped while still stopping those guilty of a crime. Though imperfect, we use whether



(a) Stops of whites (b) Stops of Hispanics (c) Stops of blacks

Figure 2: Yearly per capita stop rate for young (18–29) men by location.

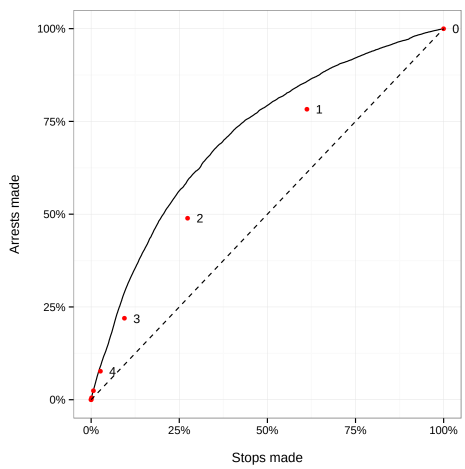


Figure 3: Performance of the full statistical model (solid line), the heuristic model (red dots), and random selection (dashed line).

or not a stopped individual was arrested as a proxy for guilt.

We start with a logistic regression model to predict the likelihood of a stop resulting in an arrest based on twenty predictor variables that are currently being used as *stop reasons* by NYPD officers. Examples of stop reasons are "carrying an object", "report from victim", "fits description", "drug transaction", and "ongoing investigation". The goal of the statistical model is to capture the importance (i.e., predictive power) of *stop reasons* being used by the officers, and weight them accordingly for future stop decisions. We envision the usage of this model as follows: the model is first trained using past stops performed by the NYPD; based on the results, the NYPD chooses a threshold p_t of probability of arrest; finally, an individual is stopped if and only if the model estimates a probability of arrest higher than p_t. A low threshold p_t will result in more stops—and hence greater social cost—but also more arrests. On the other hand, if a higher threshold is chosen, the result will be lower social cost and fewer arrests. This trade-off should be considered by the NYPD in choosing the threshold.

To evaluate our model, we train it on all 2011 stops as recorded in the stop-and-frisk data. We then rank the stops in 2012 from highest to lowest likelihood of resulting in an arrest based on our model. Note that the choice of p_t will dictate what fraction of the best stops are made. The results of this analysis are presented in Figure 3. In this graph, the x-axis represents the percentage of *best stops* that are made. The y-axis gives the percentage of arrests out of all arrests made when the corresponding percentage of *best stops* are made. We also mark the arrests made through a random sample of 2012 stops as a comparison point (dashed line). The results show that our model significantly improves on the current system. For instance, using this statistical model and setting a threshold of $p_t = 0.03$, the program would result in only 25% (x-axis) of all stops performed in 2012, but still lead to 56% (y-axis) of the arrests out of the total arrests.

As discussed above, there are twenty coefficients in the full statistical model that determine the likelihood of a suspect being arrested once a stop is made. However, this model may be hard for officers to implement in the field and can be difficult to interpret. Therefore, we propose the following *heuristic model*: we set each of the nine positive coefficients to 1, and each of the negative coefficients to 0. Determining an individual's score under this heuristic model then amounts to counting the number of criteria satisfied. We envision the NYPD selecting a stop score threshold s_t and stopping an individual if and only if the score of the individual passes that threshold.

The red points in Figure 3 represent the heuristic model, where the numbers indicate the stop threshold s_t. As can be seen from the graph, the heuristic model does surprisingly well, and is only slightly worse than the full statistical model. For instance, using a threshold of $s_t = 2$ results in 27% of stops and 49% of arrests. In comparison, the best 27% of stops made through the full statistical model results in 60% of arrests. Our results thus suggest that the heuristic model provides a useful trade-off between model complexity and accuracy.

Acknowledgments

We thank our mentors Ceren Budak (Microsoft Research), Sharad Goel (Microsoft Research), and Ravi Shroff (New York University) for their guidance on this project as part of the Microsoft Research Data Science Summer School.