2018 IEEE Fourth International Conference on Big Data Computing Service and Applications

A Relationship Between Fines and Violent Crimes

Samuel Smith
New York University
Courant Institute of Mathematical Sciences
New York, NY USA
samuelsmith@nyu.edu

Simranjyot Singh Gill
New York University
Courant Institute of Mathematical Sciences
New York, NY USA
simranjyot@nyu.edu

Kedar Gangopadhyay
New York University
Courant Institute of Mathematical Sciences
New York, NY USA
kedarg@nyu.edu

Suzanne McIntosh
New York University
Courant Institute of Mathematical Sciences
Center for Data Science
New York, NY USA
mcintosh@cs.nyu.edu

Abstract—**Inspired by the Ferguson, Missouri (MO) riots of 2014, we were motivated to study the influence of city and county revenue generated from fines and forfeitures (primarily traffic violations) on the incidence of violent crimes. For our analysis, we leveraged five orthogonal data sets aggregated at the state level, all concerning calendar year 2013. Resulting from our analysis of rate of searches, stops, and contraband found in seven states, we found a positive linear relationship between fines and violent crimes at the state level. We also found that subsets of the population in certain states are pulled over and searched disproportionately more despite finding relatively less contraband.**

Keywords-Analytical models, Big data applications, Regression analysis, Data engineering, Distributed computing, Knowledge discovery, Stanford Open Policing Project, Sunlight Foundation, Violent crime rates

I. Introduction

This paper was inspired by the Ferguson, MO riots of 2014. It analyzes and discusses the dependence of city and county revenue generated from fines (primarily traffic violations) and their potential effects on the incidence of violent crimes on an aggregated state level. Following the riots, several press articles pointed to Ferguson's elevated levels of municipal court fines (again, usually for traffic violations) and how they reduced the local population's faith in the police and overall city government. We tested whether the practice of collecting significant municipal revenue from low-level offenses had an impact on violent crimes not only in Missouri, but in other states as well.

Of the five data sets we employed, the three core data sets required extensive ETL processing (**E**xtract, **T**ransform, and **L**oad) in order to integrate them for mining and knowledge extraction:

1. U.S. Census data on tax and fine revenue for over 3,500 reporting jurisdictions (counties and cities) from the Sunlight Foundation [4].
2. FBI data on violent crimes reported by city [9].
3. Stanford Open Policing project data identifying all traffic stops by state annotated with reason(s) the driver was pulled over by law enforcement [8].

Two other public use data sets were aggregated and combined with the three core data sets, but required *de minimis* processing: U.S. Census 2010 population data [10], and Bureau of Labor Statistics average annual unemployment rates by state for 2013 [7].

We relied on data from the Sunlight Foundation (the "Foundation") for figures on local fines for 2013; 2013 is the only calendar year for which data was available. With respect to our analytics, we utilized linear regression to examine the relationship between fines and violent crime rates. An initial look at the plots revealed that there were no clear patterns at the city levels. Aggregating the local information to the state level made the relationship more apparent.

II. Motivation

After the shooting of Michael Brown in Ferguson (August 2014), politicians discussed how levels of policing activity made it harder for law enforcement to gain the trust of the public. Part of our motivation was to understand why this might true. As part of our "Search Rate" analysis, we looked at the rate of searches and pullovers for blacks *vs*. the general population in seven states, including Missouri. We found that in at least five other states in 2013, police arrests for petty violations were not generally distributed; the black population was being unfairly targeted. A closer look at the design and results of our Search Rate analysis is explained in detail in Section V.

0-7695-6396-1/18/$31.00 ©2018 IEEE
DOI 10.1109/BigDataService.2018.00025

The point of our Search Rate analysis was not to insinuate racially biased motivations, but rather the possibility that police and courts tend to mete out unduly harsh treatment, particularly with black people who tend to have less financial resources in certain areas. The literature in this area explains how courts may be imposing excessive fines on lower income residents without regard to their ability to pay. Although these related consequences were outside the scope of this paper, our findings point to the need for new practices in certain locales. We continue to be motivated by the idea that policing and court enforcement should be driven by public safety to better balance making the community whole without putting an undue burden on the offender.

III. RELATED WORK

Three related works had a similar premise to ours— a relationship between policing and law-related behaviors [1] [4] [5]. In fact, the Foundation's paper was also the source for one of our data sets— a 2013 extract of U.S. Census data on city and county-level fines and taxes [4].

A related paper studied the link between police street stops and perceptions of policing legitimacy in local New York City (NYC) communities. The authors segmented NYC into 146 *neighborhood clusters* and subsequently surveyed 1,261 individuals from clusters having a minimum of 25 individuals. The authors found strong links between whether an individual had been stopped and their perception of police, particularly when the stop was viewed as being intrusive. Individuals who believed they had been stopped fairly and treated appropriately in their neighborhood clusters considered police activity to be more legitimate [5].

This paper was related because the perception was not primarily a consequence of the number of stops. The impact on police evaluations had to do with whether the police were exercising their authority fairly. Similarly, we posit that the incidence of violent crimes is influenced by what is considered an "unfair" way to generate local revenue [5]. The general ideas in this paper motivated our Search Rate analysis as described in Sections IV and V.

Another paper examined how incidents of excessive police violence influenced citizen crime reporting. Using a time series design, the study analyzed patterns of police-related 911 calls before and after the Frank Jude beating was made public in the Milwaukee press. (Frank Jude was a Wisconsin man who was beaten by off-duty Milwaukee police officers in 2004.) This paper was relevant because the methodology did not rely simply on accounts of past behavior, but on the complete universe of every reported crime in Milwaukee for a period of seven years [1]. Similarly, we aggregated a large sample of reported police data— every police stop in 31 states which includes information on reasons for any given pullover.

Finally, the paper by the Sunlight Foundation looked at 2013 Census Data. The Foundation looked at how a local government's collection of fines compared to its collection of taxes. Our explanatory variable, fines per capita for those cities and counties reporting, is based on the approach in [4].

The Foundation looked at the ratio of fines to taxes to see which local governments rely heavily on the former to pay for basic services. It arrived at a median ratio of fines to tax revenue for city and county governments of 0.02. In other words, the median city/county collected two cents in fines for every dollar it collected. Relative to the median, the five states with the highest fine ratios were Louisiana, Arkansas, Georgia, Illinois, and Mississippi [4].

The Foundation showed that certain states generated a disproportionate amount of revenue through policing activities relative to taxation. The Foundation argued that although the absolute dollar total of fines may be large, an economically healthy state should derive a small proportion of local revenue from fines relative to taxes [4].

IV. METHODOLOGY

Figure 1. depicts our data flow. The left side shows Stage I (ETL) where we utilized Hadoop MapReduce (MR) to clean the three core data sets: Sunlight, FBI, and Stanford. For Sunlight and FBI, MR was also used to sort the respective data sets by city/county and city, respectively. For Stanford, MR was necessary early in the process to eliminate states with incomplete data. The boxes in Figure 1. characterize the sets after MR was applied. The three cleaned data sets were saved to HDFS on NYU's High Performance Computing cluster ("Dumbo") for Stage II (Analytics).

The right side of the diagram depicts Stage II where we utilized Apache Hive and Cloudera Impala to aggregate and combine the three core data sets with two secondary data sets— 2010 U.S. Census and BLS. Hive was used to aggregate the local data to the state level.

The Stanford data was already at the state level but needed to be combined with population data based on the 2010 U.S. Census. This was achieved with Impala. Since the Stanford data set was significantly larger, we chose Impala because of its ability to execute queries natively. The newly aggregated and combined data sets were again saved to HDFS on Dumbo for analysis.

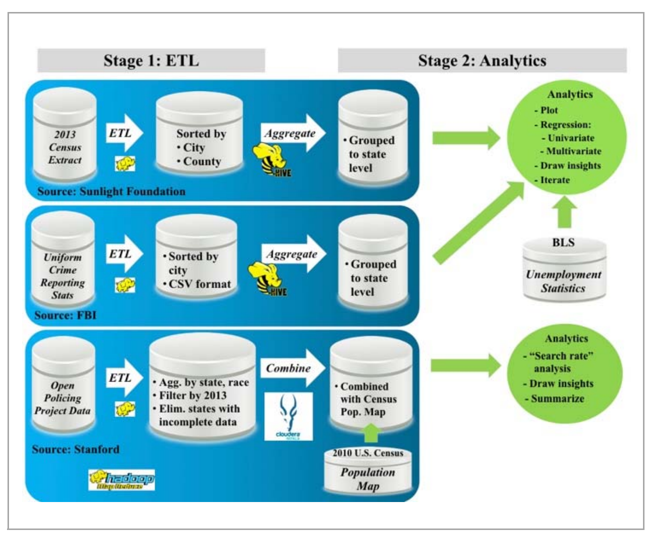

Figure 1. Data Flow

First, for an initial view we plotted fines per capita for 1,926 cities against violent crimes per capita for 2013 (Figure 2.). This immediately alerted us to a couple aspects of the data. There was a high degree of scatter, suggesting no relationship at the city level. There appeared to be no slope as well.

Current crime literature explains the significant challenges inherent in comparing crime rates across U.S. cities and counties. Some of these challenges include differences in local criminal justice systems, rates at which crimes are reported by victim and recorded by police, crime counting methods, definitions, etc. Even within an individual city, the nature of reporting in one neighborhood versus another produces different distributions of crime rates. Rather than adjusting the FBI's standard definition of violent crime, we followed the Foundation's approach by looking at the relationship between fines and crimes on an aggregated state level. The analyses of our aggregated data are discussed in the next section.

V. RESULTS

Our objective in this research is to provide insights that build on those discovered by the Sunlight Foundation to better understand and quantify the relationship between high rates of fines and higher incidence of violent crimes.

The first part of our analysis involved two regressions shown in Figure 3. and Figure 4. We tested whether the commonly held belief that high unemployment rates lead to higher crimes held true. Although many academic studies have debunked this myth, we ran a regression for annual average unemployment rates of 48 states against violent crime/capita to confirm the weak relationship. Two states were excluded, Hawaii and Wyoming, because they were missing in the Foundation data.

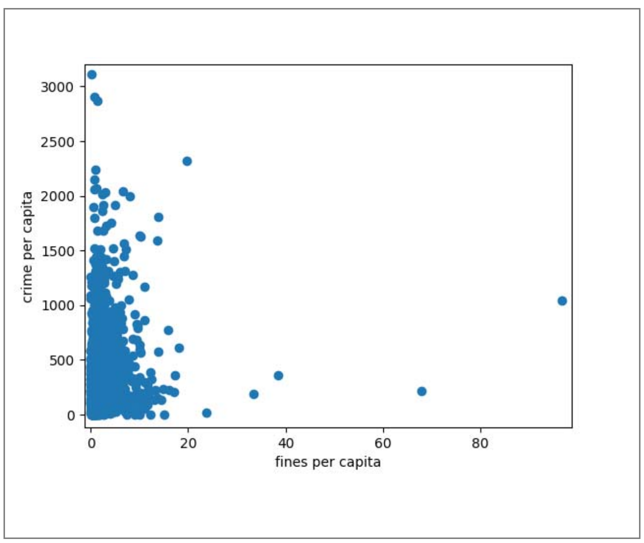

Figure 2. Initial Plot of City Data Prior to Aggregation

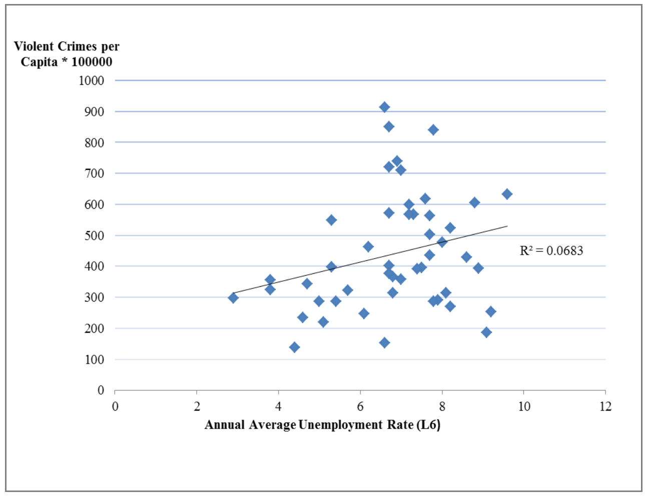

Figure 3. Regression of Aggregated State Results - 2013 Avg. Unemployment Rate[1] vs. 2013 Violent Crimes per Capita

The regression only yielded an Adjusted R2 of 0.06 and was over the level of statistical significance ($p > 0.05$). This finding was in line with several studies which cite state poverty rates, among other factors, as better explanatory variables than high unemployment. We note that unemployment tends to have a stronger relationship with property crime whereas we were concerned with violent crime.

The second regression for the 48 states ran fines per capita against violent crimes per capita. Results are shown in Figure 4. Unlike the plot in Figure 2. we were able to fit a line with an Adjusted R2 of 0.21 at a statistically significant level ($p = 0.05$). Although we had hoped for a higher R2, there are a few considerations to note.

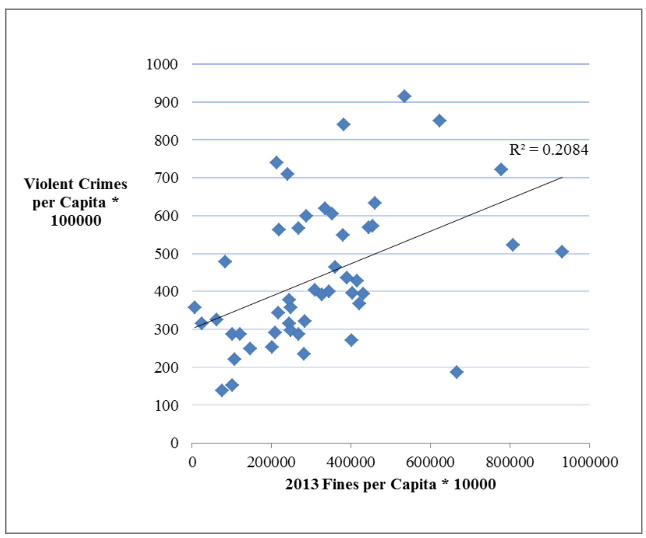

Figure 4. Regression of Aggregated State Results - 2013 Fines per Capita vs. 2013 Violent Crimes per Capita

[1] Defined broadly by L6: total unemployed, plus all marginally attached workers, plus total employed part time for economic reasons as percent of civilian labor force plus all marginally attached workers [7].

First, the low value shows that even noisy, high-variability data can have a positive trend. We argue that this trend indicates that the predictor variable, fines per capita, still provides information on violent crimes even though data points fall further from the regression line. In the context of a low *Adjusted* R^2 but statistically significant level, one might choose to increase the explanatory power of the model by adding variables, but this direction was not the point of the analytic. Instead, we analyzed the results of our aggregated data by looking at the top eleven states in our dependent variables (unemployment and fines). The results are presented in Table I.

Table I shows that five of the top eleven states with the highest fines per capita also had the highest violent crimes per capita. These states are shaded in green and include Louisiana, Delaware, Maryland, Nevada, and Missouri. We listed the top eleven states rather than the top ten because we wanted to include Missouri in the comparison, which placed eleventh on the list of the most violent states.

When looking at unemployment, only two states— Nevada and Michigan— also fell in the top eleven violent states; they are shaded in yellow. This analysis should be considered in conjunction with the positive trend in the regression because it allows a policy maker to hone in on certain problem areas. Let us consider Maryland, for example. One would expect its violent crime ranking to be lower given its low unemployment rate (6.6% in 2013), high percentage of bachelor's degrees, etc. However, the cities around Baltimore contribute to the sixth highest murder rate in the country— 249 per 100,000. Moreover, compared to the median fines per capita of $1.20 for all cities and counties, Baltimore is 293% higher at $4.73 [4].

Finally, we used our aggregated state level data to consider the consequences of excessive policing of petty crimes. We consider the possibility that citizens are simply committing more crimes, resulting in more fines. We conducted a Search Rate analysis to see whether the practice of excessive policing helps to reduce crime. By combining our statewide data of pullovers with U.S Census population data, we aggregated information on pullovers for the overall population *vs*. black population to analyze the rate at which blacks were being searched, stopped, and whether contraband was found. Due to significant differences in how the two sources defined race and ethnicities, it was not possible to look at other minority populations.

The results for these seven states are summarized in Table II: Florida, Texas, Missouri, California, Arizona, Massachusetts, and South Carolina. Of these seven states, Missouri and California (shaded green in Table II) also happened to be "high fine" states, as shown in Table I. Note that California cannot be used for comparison purposes because few consent searches are conducted in California relative to other states.

TABLE I. 2013 AGGREGATED RESULTS BY STATE[a]

Rank	Violent Crimes per Capita	Fines per Capita	Unemployment Rate
1	Maryland	New York	Nevada
2	Delaware	Georgia	Rhode Island
3	Tennessee	Louisiana	Illinois
4	Alaska	Illinois	California
5	Louisiana	Delaware	Michigan
6	New Mexico	Maryland	Mississippi
7	Nevada	Nevada	Georgia
8	S. Carolina	Missouri	New Jersey
9	Michigan	Arkansas	Kentucky
10	Alabama	California	N. Carolina
11	Missouri	Colorado	Oregon

a. The results of aggregation comprised 48 states because Hawaii and Wyoming do not report fines in the Sunlight Foundation data.

The top eleven results are shown to include Missouri and highlight the high level of fines and violent crimes.

Shaded green rows represent overlapping states from the first and second columns.

Shaded yellow rows represent overlapping states from the first and third columns.

The notes in Table II explain our methodology of comparing rates of being stopped, being searched, and finding contraband in the black driver population *vs*. general driver population. To interpret the first and second columns, for example, a ratio of 1.00 suggests that blacks are stopped and searched, respectively, at the same rates as the general population. The third and fourth columns are percentages of contraband found in the general population and black population, respectively.

Looking at Missouri, black drivers are 71% more likely to be searched after being stopped compared to the general population. That is despite the fact that searches of black drivers resulted in contraband 23% of the time while general drivers searched by police were found to have contraband 25% of the time (i.e., roughly equal). The effect is even more pronounced for Massachusetts. This tied into our review of related work Section III where we saw that it is not necessarily the number of police which helps reduce crime, but rather whether they are respected by the community and how law enforcement is managed [5].

TABLE II. SEARCH RATE ANALYSIS OF BLACK DRIVERS *VS.* GENERAL DRIVERS

	black_ stop [c]	**black_ search** [d]	**general_ contra** [e]	**black_ contra** [f]
Florida	1.09	1.33	51.1%	50.4%
Texas	0.79	1.07	22.4%	38.3%
Missouri	0.75	1.71	25.1%	22.9%
California[b]	1.34	*nmf*	*nmf*	*nmf*
Arizona	1.33	1.63	22.3%	28.1%
Massachusetts	1.59	1.45	56.6%	49.4%
S. Carolina	1.23	1.21	23.4%	26.0%

b. We use the label *nmf* to indicate that these values are not meaningful for purposes of comparison.

c. black_stop. $= \dfrac{\text{(\# blacks stopped / \# general stopped)}}{\text{(black pop. / general pop.)}}$

d. black_search $= \dfrac{\text{(\# blacks searched / \# blacks stopped)}}{\text{(total pop. searched / total pop. stopped)}}$

e. general_contra $= \dfrac{\text{\# contraband found in general pop.}}{\text{\# general pop. searched}}$

f. black_contra $= \dfrac{\text{\# contraband found in black pop.}}{\text{\# black pop. Searched}}$

Finally, this analysis allows the policy makers to focus their attention on problem areas. In Ferguson, there has been discussion on the large number of municipal warrants for missed court appearances, license suspensions, late fees, etc. Incidentally, a CNN article appeared last year after the completion of our analytic which discusses a travel advisory by the NAACP warning people of color that their civil rights could be violated in Missouri.

The travel advisory was the first of its kind in the 108-year history of the NAACP. According to the article, the NAACP cited a state attorney general report that found black drivers 75% more likely to be stopped and searched than white drivers. The NAACP's finding was similar to ours in Table II: black drivers are 71% more likely to be stopped and searched compared to the general population [3].

For some of the other states listed in the table, there are fees and penalties for offenders who fall behind on payments of accrued fines and fees. In Florida, for example, offenders are required to pay for the costs of prosecution irrespective of their ability to pay [2]. In California, ex-offenders are charged an additional $300 if they are unable to pay their fines [2]. The results of our analysis suggest law enforcement needs to consider how to change the restitution system in certain locales. New ideas may include taking into account the person's ability to pay, allowing for conversion of certain fines into community service, and imposing statutes of limitations on fines, a type of state law that restricts the time within which legal proceedings may be brought.

VI. FUTURE WORK

This paper looked at violent crime rates in states marked by high fines and forfeitures, and to a lesser extent, by high unemployment. It is important to underscore that discovery of a statistical relationship between a given factor and violent crime never provides a picture on causation.

As evidenced in Table I, our research cannot tell us whether individuals being levied with fines are actually committing violent crimes. In other words, "arresting people for minor violations is exactly the point ... they don't want to actually incarcerate people because it costs money, so they fine them."[2] Since violent crime frequently leads to arrest and imprisonment and this reduces an individual's earnings to pay (repay) fines and obligations, it is possible to argue that crimes lead to fines and not vice versa. To determine the causal direction of the relationship, future research might be a longitudinal study to see whether crime follows or precedes the burden of fines.

Even without changing the structure of this study, however, we propose leveraging other data in the Stanford Open Policing Project. For example, one frequently overlooked influence on long-term crime rates is age. Crime is overwhelmingly committed by individuals in the 15-24 age range. Since the Stanford data contains pullover by age and reason for stop, it would be interesting to test whether violent crime rates are higher in "younger" states. In a similar vein, we could also use the gender data supplied in the Stanford data.

One of the key obstacles faced when researching for this paper was the lack of standardization across the Stanford pullover datasets for the different states. Not all states reported all columns. New Hampshire does not report the race of the driver who was pulled over, for example, and so New Hampshire was excluded from the study. Ultimately, this led to only a small set of states having provided enough data. Including the other variables would have allowed other states to have their data more thoroughly explored by the analysis.

Similarly, the study was limited to the year 2013 as this was the source year for the Sunlight Foundation's City and County fines analysis. Given adequate resources, revenue details for additional years could be sourced, allowing for the study to cover a wider timespan.

Research has uncovered variations in rates of assault and homicide by time of day, day of week, and month of year and the Stanford data also contains every pullover by time. Another application would therefore examine whether there is overlap between time of fines and time of violent crimes. The Stanford data gives us refined location data of the stops; it may be possible to more strongly relate the locations of frequent pullovers and searches to the incidence of violent crime in that more specific location.

In this paper, we also examined, to a limited degree, the relationship between the unemployment rate and violent crime. This section could be expanded by separating out short- and long-term unemployment metrics, as well as using the unemployment data broken down by the age, race, and gender demographics referenced above.

We know a good deal more about general characteristics of high fine and high crime states. Based on our study, we ought to know more about the *causes* of high fine-prone and violent crime-prone places.

VII. CONCLUSION

Much of the existing literature states that it is not common for local governments to generate excessive revenue through policing petty crimes. In our paper, we focus on exploring precisely why this is an unhealthy practice.

The results of our experiments confirm our alignment with previous studies which debunked the common myth that unemployment and violent crimes are strongly correlated. We then tested whether there was any positive linear relationship between fines and violent crimes. Due to the complex ways in which boundaries are drawn and crimes are defined, we saw no relationship at the local level. At the state level, however, the linear relationship became apparent and statistically significant. The results of our fit were confirmed by overlaps between the top fine states and top violent crime states. We also discussed equitable stop and search treatment with respect to subsets of the population.

The causes of violent crime are a highly nuanced topic. We showed that a relationship between areas marked by high fines and high rates of violent crime exists, and there are potential consequences of excess fining in certain areas. We have provided a preliminary overview in technical, yet simple terms, on what we know about the relationship between fines and crimes.

ACKNOWLEDGMENT

We are grateful to the Sunlight Foundation for collecting and providing us with data on fines and forfeitures.

REFERENCES

[1] M. Desmond, A. Papachristos, D. Kirk. Police Violence and Citizen Crime Reporting in the Black Community. Cambridge, MA: American Sociological Review Vol. 81(5) 857-876, 2016.

[2] G. Guilford. How Ferguson and Many Other U.S. Cities Wring Revenue from Black People and the Poor. Quartz: qz.com/257042/these-seven-charts-explain-how-ferguson-and-many-other-us-cities-wring-revenue-from-black-people-and-the-poor/ August 28 2014.

[3] R. Sanchez. Here's Why the NAACP says Missouri is unsafe for minorities. CNN: http://www.cnn.com/2017/08/06/us/missouri-naacp-travel-advisory/index.html August 6 2017.

[4] E. Shaw. Where Local Governments are Paying the Bills with Police Fines. Washington, DC: Sunlight Foundation, September 26 2016.

[5] T. Tyler, J. Fagan, A. Geller. Street Stops and Police Legitimacy: Teachable Moments in Young Urban Men's Legal Socialization. New Haven, CT: Journal of Empirical Legal Studies, October 20 2014.

[6] City of Ferguson Missouri. Annual Operating Budget: http://www.fergusoncity.com/DocumentCenter/View/1609 Fiscal Year 2013-2014.

[7] Local Area Unemployment Statistics. Unmployment Rates for States, 2013 Annual Averages (Bureau of Labor Statistics). Washington, DC: bls.gov/lau/lastrk13.htm, 2013.

[8] The Stanford Open Policing Project. Palo Alto, CA: openpolicing.stanford.edu/data, 2017.

[9] Uniform Crime Reporting Statistics. Department of Justice (Federal Bureau of Investigation). Clarksburg, West Virginia: ucrdatatool.gov, 2017.

[10] United States Census 2010. Interactive Population Map (U.S. Census Bureau). Washington, DC: census.gov/2010census/popmap/, 2010.