Factors that Influence Rate of Serious Crimes per County

Bárbara Bettencourt* Chisom Onyishi† Jessica-Fuchs Shafer‡

Columbia University, Department of Statistics

April 2021

Summary

This paper examines factors that influence the rate of serious crimes per county in the United States through empirical analysis. In particular, we find that poverty is the biggest predictor of crime rate and surprisingly, that the counties in the South are the most likely to have serious crimes as opposed to the counties in the Northeast, which have the lowest crime rates in the country. We suggest policies whose implementation have the potential to lower the rate of serious crimes in these counties.

Contents

*bbf2118@columbia.edu
†co2501@columbia.edu
‡jf2677@columbia.edu

1 Introduction

Importance of Analysis

Given the recent attention to systemic issues in the police department and with police misconduct, the need to harness the growing power of data to address these issues is more important than ever. A vital concern in overall social well-being is, of course, decreasing the crime rate overall, and the corresponding likelihood that police will be called. Thus, in this paper, we will address possible predictive factors for crime rates, so the factors themselves can be addressed. The identification of predictive factors has the potential to decrease the crime rate overall, and the corresponding need for police intention. We focus on identifying features of United States counties between 1990 and 1992 that are associated with high levels of crime. Through identification of these factors, we recommend policy changes that have the potential to address some of the major underlying causes of crime in order to address the root of the issue.

1.1 Data

To determine factors that effect county-level crime, we analyze the *County Demographic Information* (CDI) data set, which contains characteristics of 440 counties in the United States collected from 1990 to 1992. In particular, we focus on a subset of this data set with 250 observations. The CDI data includes 17 variables. Our outcome variable of interest is rate of serious crimes per 1000 people (henceforth referred to as "crime rate").

Based on the correlations among the transformed variables (See *Appendix* (4)), we find that the following variables are highly correlated, and thus we leave some of them out (see first column of the table below) in our analysis as they generally convey similar information. This is done to avoid spurious results. We also take note of some observations that differ significantly from the rest, as these observations can have a big impact in our model and analysis.

Problematic Correlations Among the Predictor Variables

Will not use	Will keep	Correlation
Physicians per 1000 people	Beds per 1000 people	72%
Per capita income	Percent below poverty	-60%
Per capita income	Percent uni grad	65%
Percent highschool grad	Percent uni grad	72%
Percent of unemployment	Percent uni grad	-58%
Over 65 years	18 to 34 years	-60%

Note that these are pairwise correlations.

See the appendix for the full correlation heat map

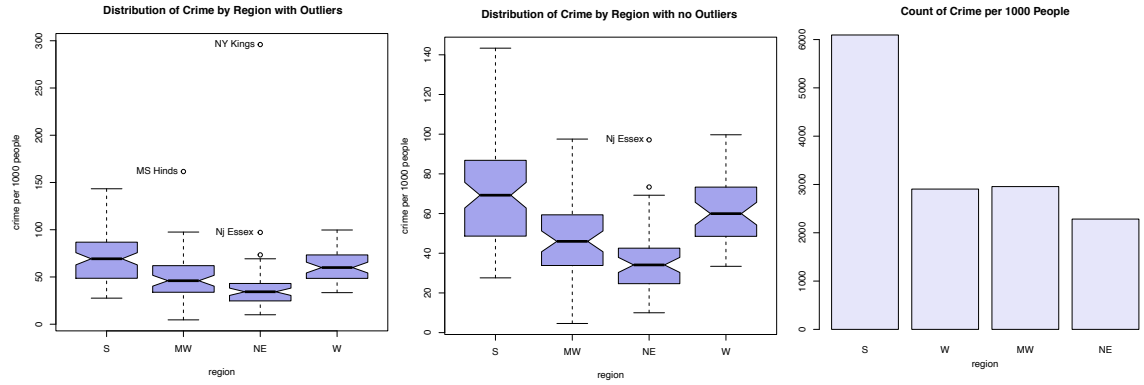
Here, we compare the outlier of Kings county with another county with the closest population density, Rockland county, in order to investigate how much the values differ. We can see that Kings' population density is almost 21 times bigger and the crime rate is 11 times bigger than those of Rockland county. This comparison illustrates the nature of the high crime rate in Kings county in this data.

Comparing Kings to Next Most Dense County in New York

County	State	Pop density	Crime rate
Kings	NY	32404	29.60%
Rockland	NY	1525	2.70%

We also note that the South has a higher crime rate than the other regions in the United States This is somewhat counter-intuitive, as the population density is higher in the Northeast than in the South. And consistent with above, we find that the most outlying observation is Kings county, New York. This is

due to its high population density and crime rate. All other characteristics are within a reasonable range.



Finally, based on the scatter plot matrix (see *Appendix* (4)) the variables `Region, Percent of population 18 to 34 years old, Percent of population with bachelor's degree, Percent of population below poverty, Hospital beds per 1000 people,` and `Population density` are chosen as the potential explanatory variables for crime rate.

1.2 Model

After careful data analysis, we employ the following regression model:

$$\begin{aligned}\texttt{crimerate}_{1000} = \mu + \tau_1\,\texttt{Northeast} + \tau_2\,\texttt{Midwest} + \tau_3\,\texttt{South} \\ + b_1\,\texttt{log(\%poor)} + b_2\,\texttt{log(popdensity)} + b_3\,\texttt{log(\%unigrad)}\end{aligned} \tag{1}$$

Overall, this model shows that the effect of population density, percent of university graduates, and percent of people living under the poverty line is essentially the same for all the regions. The effects of the regions, however, are different, and this difference is captured by the various τ's. At first glance, it might seem that the `West` region has no effect. However, the effect is simply $\tau_4 = -\tau_1 - \tau_2 - \tau_3$. We report the results in the following section.

2 Results

The table below contains the coefficients from our regression analysis.

Regression Summary for Model (1)

Variable		Estimate
Grand mean	(μ.)	-115.6***
log(pct poor)	(b_1)	29.7***
log(pct uni grad)	(b_3)	13.6***
log(pop density)	(b_2)	11.8***
Northeast	(τ_1)	-12.6***
Midwest	(τ_2)	-4.4*
South	(τ_3)	8.1***

Note that effect of `West` *would be 8.9(*τ_4*)*

*Significance: * is 0.1 level *** is at 0.005 level*

We note that each predictive variable in the regression has a statistically significant relationship with the outcome variable of crime rate. It is worth mentioning that the grand mean is negative. This is the

crime rate we expect when all the independent variables have zero effect. The value of -115.6 means that the crime rate would be an impossible value without the influence of the predictor variables. This is, in part, an artifact of the model type, and is a hypothetical reference value, as a population density of 1, for example, is far outside the scope of the data.

Region variables: Controlling for the non-regional predictors, on average, (1) counties in the Northeast have approximately 12.6 less crimes per 1000 people (2) counties in the Midwest have 4.4 less crimes per 1000 people (3) counties in the South have 8.1 more crimes per 1000 people, and (4) counties in the West have 8.1 more crimes per 1000 people.

Other predictors: In order to look at the predictive effects of the non-regional explanatory variables, we examine the table below. For example, the first two entries of the "Percent poor" row states that a 10% reduction in the percent of people living under the poverty line would reduce crime rate by 2.7, while a 10% increase would increase the crime rate by 2.5. The other values are interpreted similarly. The largest effect we observe on change in crime rate comes from percent of individuals living below the poverty line. We will discuss the implications of these findings in the following sections.

Effect of Percent Change in Predictor Variables on Crime Rate per 1000

	Percent Increase		Percent Decrease	
	10%	20%	-10%	-20%
Percent poor	2.5	**2.9**	-2.7	**-5.8**
Percent uni grad	1.3	2.5	-1.4	-3
Population density	1.1	2.2	-1.2	-2.6

We note that the biggest reduction or increase in crime rate is associated with poverty

2.1 Interpretation of Results

Our regression analysis demonstrates the predictive power of poverty in a community on crime rates. The significant coefficient of 29.7 on percent below the poverty line means, on average with all other variables held constant, decreases to percent of people living below poverty has a significant impact on reducing crime rate (as seen on the table above). Although this is a very strong relationship, this finding is unsurprising. Individuals with fewer resources are more at risk for joining gangs, becoming involved in drugs, and finding themselves at risk of not being able to pay bills and meet basic needs. The interconnected nature of crime and poverty is undeniable, and a seemingly insurmountable and unsolvable problem. We will discuss possible interventions in the next subsection.

Interestingly, the Northeast has a low poverty rate relative to its population density. One possible explanation for this phenomenon is that the Northeast contains a few very densely populated areas and some very sparsely populated areas that skew the average. The Northeast and Midwest showed, on average, a statistically significant lower crime rate than we observe in other parts of the country. While these places have densely populated cities, which was among the strongest predictors of rate of serious crimes, these areas also include very sparsely populated areas. This is one possible explanation for the lower crime rates in these regions. The South, however, showed a statistically significant positive relationship with crime rates. We must, of course, also consider the lasting affects of the South's history of slavery when discussing its social structures. While difficult to tease apart from other potential factors leading to social inequality, it is worth bearing in mind when considering these problems.

Perhaps most surprising of the predictive factors was the positive relationship between the percent of the populations that are college-educated and the crime rates in these same areas. As we can see on the heatmap, this correlation is 0.1. But, in the regression when we control for other variables, this effect is even stronger. To fully investigate this connection, more data across a longer period of time is warranted. It is possible that this relationship is a function of income inequality in more densely populated areas, and places with more educated people also have greater numbers of individuals living below the poverty

line. It is also possible that the volume of people may be large enough to generate more crime. Further analysis will better elucidate the nature of this counterintuitive relationship.

In conjunction with our prior understanding of crime rates, higher average income levels are negatively predictive, while higher unemployment rates are positively predictive. The two are of course interconnected. Data about the income distribution might better represent the issue of wealth in communities. For example, communities with a couple of very rich individuals cannot truly be said to have higher income. A better measure of income would the median as it is more resistant to outliers. This information would improve future exploration of the vitally important issue of predictive factors for crime rates.

3 Implications and Policy Recommendations

In this paper, we sought to answer the following question: what are the predictive factors for crime rate increase? While it is not possible to claim direct causal relationships, we propose the following recommendations to potentially mitigate the crime rate based on the findings of our empirical analysis.

3.1 Suggested Poverty Reduction Policies

According to our model, the most significant predictive factor for an increase in crime rate, across all the regions, is the percentage of the population living below the poverty line. In this section, we will discuss policies that could aid in raising communities above the poverty line. We propose that engaging the community in ways that could naturally lower the crime rate by supporting the individuals and community as a whole will have an overall positive affect. Thus, we posit that lowering the crime rate is not about catching "bad apples", but about treating the problems endemic to the community that may lead to higher crime rates. These methods include increasing employment opportunities, community engagement, healthcare interventions, and improving education and youth support. From the model, we can infer that areas in most need of this support through policy changes are communities in more densely populated areas.

A. Increasing Employment Opportunities

In the heatmap (*Appendix* (4)), we observe, unsurprisingly, a 0.42 correlation between unemployment and poverty. This data was gathered from the years 1990 to 1992, at the time the United States was going through the first recession since the stagflation, simultaneous economic stagnation and inflation, of the 1970s. This recession increased both unemployment and underemployment.

Unemployment is an important issue for policy makers across state and party lines. The link between many of these policies, aimed at solving the problem, and marked increases in employment rates are specious. One demonstrated way to both increase employment opportunities, however, is through federal work programs. This job creation not only stimulates the economy and generates purpose, but it also improves communities. One such example is President Franklin D. Roosevelt's Work Progress Administration, 1935 to 1943, prior to American involvement in World War II. This program created full-time jobs and developed infrastructure such as building highways. Future research endeavoring to find plausibly causal relationships to unemployment would be invaluable to considerations of crime rates and subsequent policy reform.

B. Community Building

In densely populated areas, there is greater potential for deindividuation and isolation. One way to mitigate these possible side effects of membership in a city is the creation of organizations, especially among the young. For youth in particular, isolation and lack of community activities are correlated with negative future outcomes. One way to tackle such issue is to implement policies that encourage a sense of ownership in the community. Programs that connect the community and help make the community stronger are possible ways to improve this sense of ownership over the community. Funding for projects such as community gardens, public arts, help lines, have the potential to decrease crime rates in communities with these programs. Community building projects, of course, cannot merely be implemented in

the same way across in-need communities. The engagement efforts must be tailored to specific communities. Thus, funds should be allocated to aid established leaders in the community in order to support them and expand their resources. Efforts must be taken to avoid imposing what outsiders deem the best way to speak to those in need of support. Instead, scaffolding for individual communities should be provided, on which each community can build their own best support structure.

C. Medical Resource Support

There is a 0.06 correlation between number of physicians per 1000 people and percent of people living in poverty (*Appendix* (4)), which implies there are few doctors in impoverished counties. This can lead to negligence, physician burn-out, and worse health outcomes for patients. Inadequate healthcare discourages individuals in need from seeking medical care. Furthermore, overwhelmed doctors are far less likely to conduct comprehensive evaluations.

A possible way to improve medical care is the inclusion of cultural liaisons. Because poverty can be racialized, and physician trust among historically marginalized communities is very low, these measures have the potential to make a tremendous difference in the healthcare system. This endeavor also has the added benefit of creating jobs for individuals who need only become certified. These programs are being implemented across the United States and would greatly benefit from more attention. Support staff, beyond doctors, are a vital component to medical care. Increasing the number of nurses aids and individuals who have the ability to help patients feel heard has the potential to support doctors, improve patient experience and likelihood of seeking treatment, and activate and build community.

Bringing health education to impoverished areas, too, can aid with family financial planning and improved sexual health. Increasing support in family planning programs, especially in impoverished areas, has the potential to allow individuals to plan for their ideal number of children, and have children when they are most ready. This type of program availability requires intervention at the policy level, and has the potential to have tremendous effect on both the individual and community levels.

D. Education

Although our model results (2) suggests that the percent of university graduates is a positive predictor of increase in crime rate, which is counterintutitive and inconsistent with existing research, we addressed possible explanations of this in the interpretation section of results.

We observe that communities with lower rates of high school graduation rates have higher rates of individuals falling below the poverty line (see *Appendix* (4)). Thus, policies in favor of bringing increased funding, allocated towards education, to communities with higher poverty rates has the potential to mitigate their correspondingly crime rates, especially in densely populated areas.

Areas stricken with poverty are also typically plagued with a lack of funding in their local public schools. Not only should programs not be cut, but they ought to be increased. While expensive, the social cost of students in these communities missing out on learning during this critical period is far greater.

After school programs, which both occupy students and help them to foster crucial skills, have tremendous potential to improve communities. Underprivileged groups, who are typically most effected by crime, could be significant beneficiaries. Additional funding dedicated to introducing more after school programs in high-crime rate communities could help students who are most at risk. Not only would participation in after school sports, art, or skill-building benefit the communities today, but this participation would also benefit the students in the long-run.

3.2 Conclusion

To summarize, we found that the percent of people living below poverty lines and population density are amongst the significant predictors of crime rates. In addition, we find that counties in the Northeast and Midwest are less likely to have crimes than counties in the South and West regions. This highlights areas that require special consideration and attention. Further analysis of predictive factors is warranted

in order to better understand the nature of crime rates, and what leads to higher crime rates, in order to implement policies that have the greatest potential to improve communities across the United States. Considerations of economic and financial factors that may influence employment rates, in addition to considerations of systemic social issues, can improve our understanding of the multifaceted issue of crime and its prevention. Endeavors toward policy reform should be bolstered by careful analysis of the data, and remain sensitive to future information as it becomes available. This type of systematic change, however, can be achieve only with community engagement, empirical analysis, and policy reform designed to support this all-important goal of decreasing crime rates across counties.

4 Appendix

Correlation Heatmap for Relevant Variables