



Impact of Household Income On Crime Rate

Dominic Lim
Emerald Swei
Shalini Chawla



Introduction

The DOJ has tasked us with finding correlations between community characteristics and crime rates.

- Lack of familial resources and economic opportunities → increase in crime
 - Focusing on socioeconomic characteristics
 - Income, level of education, employment/unemployment, and immigration
-

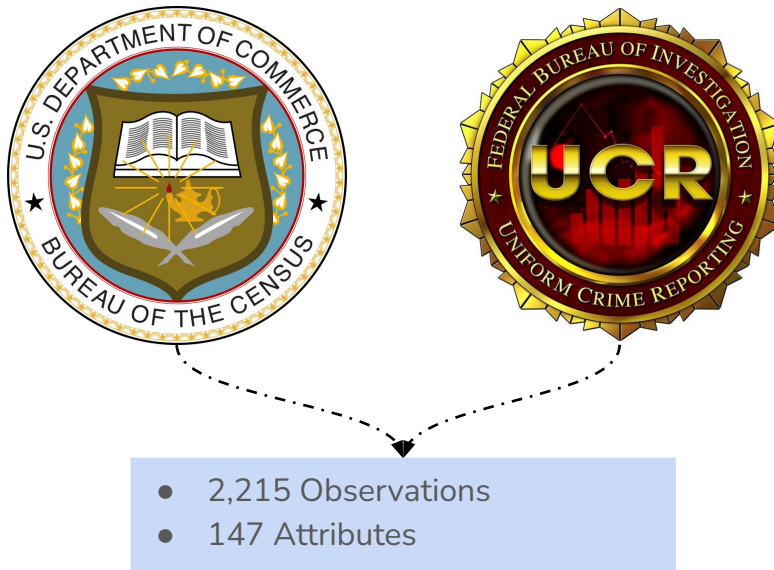
Our team has been tasked by the Department of Justice to find correlations between community characteristics and crime rates. By identifying communities that are likely to have higher crime rates as a product of these community characteristics, the DOJ may be able to prioritize high crime rate communities in allocating its \$29 billion budget.

For the purpose of this study, we hold a working hypothesis that a lack of familial resources and economic opportunities compel individuals to commit crimes. To operationalize our study, we are choosing to focus on predictor variables that relate to socioeconomic factors. Principally, we will be focusing on median household income as the primary predictive variable for crime rate. The covariates we have chosen are based on research that indicates these other factors may also have an impact on crime rate, such as level of education, levels of unemployment, and recent immigration into the community.

Do neighborhoods with lower median household incomes have higher crime rates?

The primary research question that we aim to address is:
Do neighborhoods with lower median household incomes have higher crime rates?

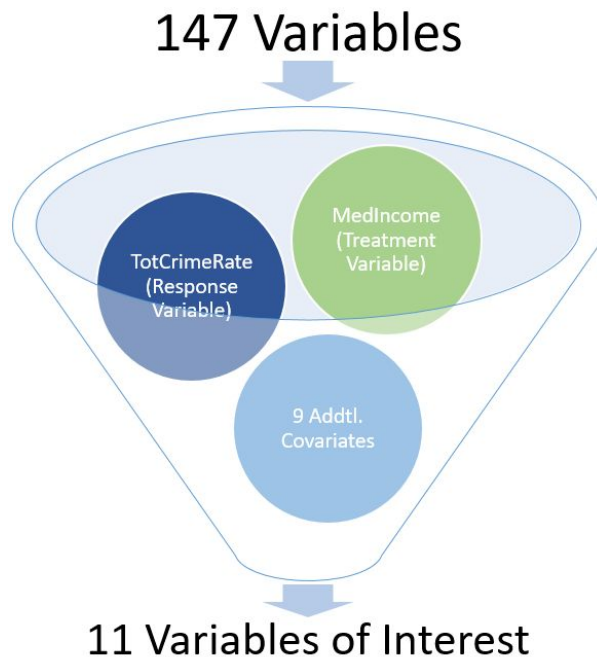
Dataset



For this analysis, we have chosen to use an existing dataset available through the University of Irvine data archive. This dataset combines socio-economic data from the 1990 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR. Communities not found in both census and crime datasets are not included in the dataset.

The dataset consists of 2,215 observations with 147 attributes.

**Dataset
(continued)**



We are working with 11 variables in our data analysis

- 1 Response variable: Total crime rate per capita
- 1 Treatment variable: Median household income
- 9 covariates

Our source has data on 8 different crimes that are identified by FBI as Index Crimes. For our analysis we are adding up all 8 individual crime rates into a single response variable - Total Crime Rate per Capita, where capita is defined as per 100,000 of the population.

Our primary treatment variable Median Household Income comes from the census data and we have converted the values to be represented as 1000s for our analysis.

Variables of Interest

Response Variable:

Per Capita Crime Rate

Treatment Variable:

Median Household Income

Covariates:

Percent Immigrants in Last 5 Years*
Percent of Non-High School Graduates*
Percent Unemployed*
Population Density*

Percent 2 Parent Households**
Percent Under Poverty**
Percent Vacant Homes**
Percent Living in Shelters**
Median Year of Houses Built**

* included in model 2

** included in model 3

After carefully examining the list of fields as well as considering data availability, we landed on these 11 variables: the response variable and treatment variable are the crime rate and median household income, as previously mentioned. In our second model, we include variables that measure the percentage of immigrants in the last 5 years (i.e., the percentage of the community that are considered recent immigrants, having moved to the U.S. in the last 5 years), the percentage of non-high school graduates, the percentage of people in the community who are unemployed, and the population density, which is measured in persons per square mile. We chose these variables because we wanted to represent factors that we believe would have an impact on the crime rate, in addition to the median household income: immigration, education, unemployment, and population density. These variables serve as our controls in model 2.

In model 3, our kitchen-sink model, we included an additional 5 variables: percentage of families with 2 parents, percentage of people living under poverty, percentage of vacant homes, percentage of people living in shelters, and median year of houses built. We wanted to see the relationship between single parent vs. dual parent households and crime rate, and measure the impact of poverty and homelessness on crime. We also wanted to measure variables not directly related to residents and citizens, but that were characteristics of the community through real estate/property; in particular, we wanted to look at the age, desirability, inhabitation, and new investment into the neighborhood. With these variables, we wanted to evaluate any variables that just might have an impact on crime rate.

Model Selection

$$\log(\text{totCrimesPerPop}) = \beta_0 + \beta_1 \text{medIncomein000s} + \epsilon$$

$$\log(\text{totCrimesPerPop}) = \beta_0 + \beta_1 \text{medIncomein000s} + \beta_2 \text{PctRecImmig5} + \beta_3 \text{PctNotHSGrad} \\ + \beta_4 \text{PctUnemployed} + \beta_5 \text{PopDens} + \epsilon$$

$$\log(\text{totCrimesPerPop}) = \beta_0 + \beta_1 \text{medIncomein000s} + \beta_2 \text{PctRecImmig5} + \beta_3 \text{PctNotHSGrad} \\ + \beta_4 \text{PctUnemployed} + \beta_5 \text{PopDen} + \beta_6 \text{PctFam2Par} + \beta_7 \text{PctPopUnderPov} \\ + \beta_8 \text{PctVacMore6Mos} + \beta_9 \text{PctShelter} + \beta_{10} \text{MedYrHousBuilt} + \epsilon$$

Our first model consists of only median income, while our second model includes an additional four variables as mentioned previously: a variable to measure recent immigration, one to measure completion of high school, a measure of unemployment, and a measure of population density.

Our third model, which is the model we ended up choosing as our best model, includes the five variables in model 2 as well as the additional variables that measure the percentage of families with 2 parents, percentage of the population living below the poverty line, the percentage of properties that have been vacant for more than six months, the percentage of people living in shelters, and the median year of houses built in the community.

We found that as we added more variables to the models, the model performance increased and incrementally fit the data better.

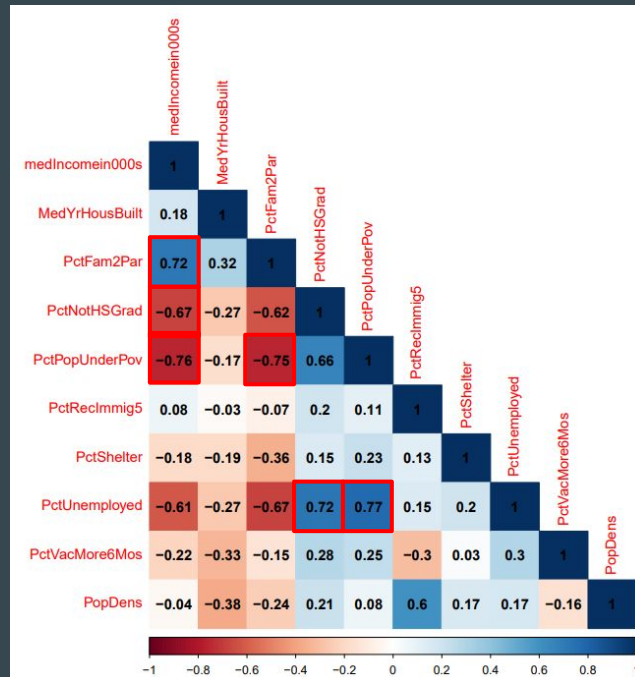
Results

	<i>Dependent variable:</i>		
	Crime Rate Per Capita		
	(1)	(2)	(3)
Median Income in 000s	-0.024*** (0.001)	-0.021*** (0.001)	-0.005*** (0.001)
Percent Immigrants		0.035*** (0.006)	0.027*** (0.005)
Percent Non HS Graduates		-0.001 (0.002)	0.001 (0.001)
Percent Unemployed		0.030*** (0.005)	-0.008 (0.005)
Population Density		0.00001 (0.00001)	-0.00001** (0.00000)
Percent 2 Parent Families			-0.038*** (0.002)
Percent Under Poverty			-0.00002 (0.002)
Percent Vacant Homes			-0.004*** (0.001)
Percent Living in Shelters			0.225*** (0.070)
Median Yr House Built			0.006*** (0.001)
Constant	9.270*** (0.034)	8.916*** (0.080)	-0.106 (1.841)
Observations	2,215	2,215	2,215
R ²	0.302	0.352	0.550
Adjusted R ²	0.301	0.350	0.547
Residual Std. Error	0.493 (df = 2213)	0.476 (df = 2209)	0.397 (df = 2204)

Note: *p<0.1; **p<0.05; ***p<0.01

Displayed here is a stargazer table that compares the coefficients and robust standard errors of the three models. At the bottom of the table highlighted in red you will see the adjusted R², which climbs from 0.301 in the first model to 0.547 in the third. Percentage immigrants, population density, percent two parent families, percent vacant homes, percent living in shelters, and median year house built are all significant covariates in the third model. Percent living in shelters has the highest absolute coefficient, contributing a 22.5% increase in total crime rate with each percentage increase. We also see that median income remains a significant variable for modeling the crime rate in all three models and thus we have evidence to answer our research question of whether median income has an impact on crime rate. The coefficient for median income is negative, indicating an inverse relationship with crime rate; as median income increases, the crime rate decreases and vice versa. However, it is important to note the decrease in strength of the coefficient for median income, at first a minor decrease of 0.3% from the first to the second, and then dramatically from the second to the third by 0.16%. We believe this may be due to some direct and indirect collinearity with some of the other variables in the third model, which Dominic will discuss more in detail in the next slide.

Results (continued)



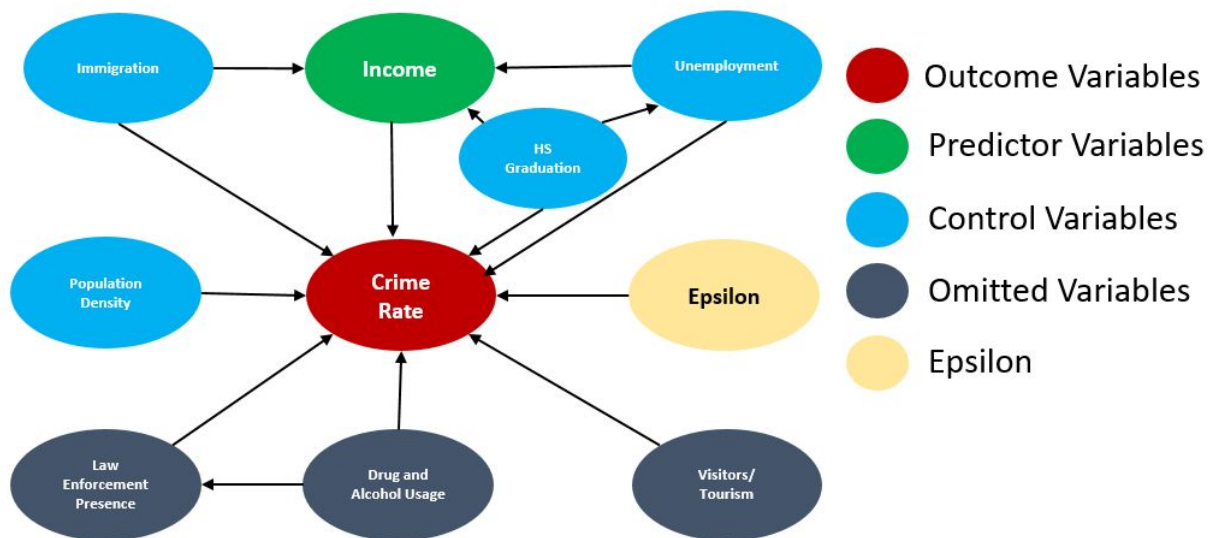
DL:

To further drive the point that that our third model include collinear variables, we present this correlation heatmap. We find that correlations between pairs of 'socioeconomic' variables tend to be higher. In particular, PctFam2Par, Pct with NO HS degree, PctPopUnderPov, have strong linear relationships with medIncome with correlations of 0.72 - 0.67 and -0.76 respectively. The relationship between median income and the percentage of families with two parents is not surprising because a household with two parents is likely to have a higher combined household income.

Practically, these variables likely have collinearity as factors such as income, employment, and higher education are generally related.

Despite this, when calculating the Variance Inflation Factor for our model 3 covariates, only % Population under poverty registered a VIF greater than 4.

Statistical Limitations - Omitted Variable Bias



We would also like to discuss Omitted Variable Bias in our models. Our select list of omitted variable biases includes 'Law Enforcement Presence', 'Drug and Alcohol Usage' and 'Visitors/Tourism'.

We believe that 'Law Enforcement Presence' has a negative relationship with crime rate, and higher income areas can generate more taxable income to fund additional officers, the overall bias pushes our coefficient away from zero (the coefficient for median income is negative). While our dataset included LE data such as "FT police officers", we intentionally omitted this variable because only the largest police departments (>100 officers) were included in dataset.

- + (- * +) (awayfrom 0)

2. Another omitted variable is drug and alcohol usage. With greater drug and alcohol usage in a community, we believe that crime rates would also increase. Drug and Alcohol Usage may impact a number of coefficients. For example, high drug and alcohol usage may affect employment and adversely impact income. Drug usage may also impact Law Enforcement Presence as addtl narcotics officers are hired to combat illicit substances.

- + (+ * -) (net impact is coeff away from 0)

Another omitted variable is the number of visitors. Communities with large numbers of visitors will have higher per capita crime than communities with fewer visitors. More

visitors may also mean additional business for the local economy bringing the household income up. The overall bias in this case is pushing our coefficient towards zero.

- $(+ * +)$ towards zero.

However we don't believe we have an exhaustive list of omitted variables as we believe crime a complex phenomena.

Conclusion

A \$1,000 increase in median household income is associated with a 0.5% reduction in total crime rate.

In conclusion, we have demonstrated that according to our model, lower median household income communities tend to have higher total crime rates. We estimate that an incremental \$1,000 in median household income is associated with a 0.5% reduction in total crime rate, all things equal.

This is practically important as the DOJ may be able to target communities with lower median household incomes as candidates for intervention programs. Such example programs might include job and skills training that allow members of that community to access better economic opportunities. We would also encourage additional investigation in which the impact of household income on crime rates is studied through direct intervention.

This may come in the form of experimentation with tax breaks and incentives for employers to provide jobs in lower income communities, or direct cash assistance akin to a universal basic income. With an estimated annual aggregate cost of 2.8 trillion dollars to the US from violent and non-violent crimes, we believe that such preventative measure may help to reduce crime and their impact on our communities.

<https://www.journals.uchicago.edu/doi/abs/10.1086/715713>