ISM6137 Statistical Data Mining
Graduate MS-BAIS, Spring 2022

Welfare Spending Impact on Crime Rate

Group 1 Shyam Parmar | Dalton Anderson | Pankaj Tiwari | Uma Sankar Koramutla

Executive Summary

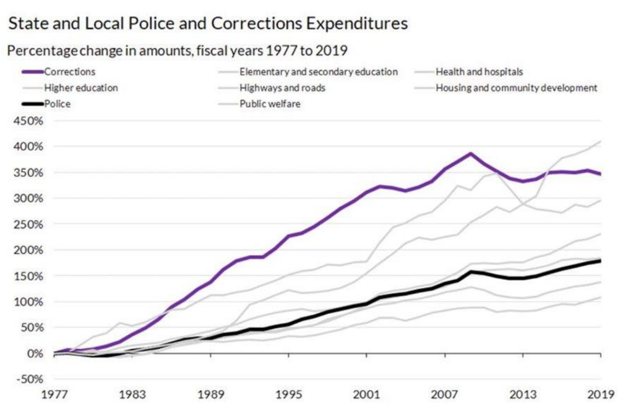
Over the last 60 years, welfare spending has increased at rates that taxpayers may be concerned about how much of the money spent is advantageous for the community both at a federal and local level. This report uses FBI crime data and city spending data over the last nine years. The project goal was to model the relationship between non-violent crimes and welfare spending in the cities with populations over 100,000 while comparing their crime rates in those cities from FBI crime data. Our model results show that welfare spending does not have a statistically significant effect on non-violent crime.

Table of Contents

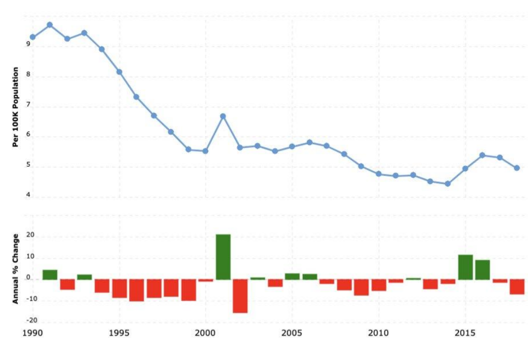
Problem Definition & Significance

Our target audience is municipalities and their city council members. Cities have a limited budget every year, and to take on new initiatives, they either have to optimize their current spending or increase revenue in the form of taxation. Cities like to keep in favor with their current residents; no matter what party one belongs to, either Republican or Democratic, no one is a massive fan of taxation. Thus it is essential to optimize spending to free up capital for new initiatives. Every city is interested in learning how spending over time affects its goals. To be more specific, every city wants to know the marginal effect of certain areas of spending on specific key city initiatives. There are three types of cities tier-1, tier-2, and tier 3 cities, and they all compete against each other within their respective tier. For a city to increase its attractiveness to the desired residents, perform well on three metrics - crime rate, cost of living, and high-paying jobs. An ideal situation for a future resident is high-paying jobs with low crime and below-average living expenses. High-paying jobs and increased cost of living are correlated. The only metric a city can have a marginal effect on is the crime rate regarding increasing city attractiveness to prospects.

State Spending (Trends)



U.S Crime Rate ("Justice Expenditures")



The above two graphs show the state - Local Police & Corrections Expenditures and Crime rate. From those two graphs, we can see a slight negative relationship between state spending and the crime rate, i.e., as the expenditures increase over the years, the crime rate has a downtrend.

Prior Literature

Title	Predictors	Findings
Striking at the roots of crime: The impact of social welfare spending on crime. (*Fishback et al.*)	Property crimes, Balanced panel, Larcenies, Burglaries, Robberies, Auto thefts, Murders	1. Analysis suggests that: a 10 percent increase in relief spending during the 1930s reduced property crime by roughly 1.5 percent. 2. By limiting the amount of relief recipients' free time, work relief may have been more effective than direct relief in reducing crime. 3. Our results indicate that social insurance, which tends to be understudied in economic analyses of crime, should be more explicitly and more carefully incorporated into analyzing temporal and spatial variations in criminal activity.
Reconsidering the Relationship Between Welfare Spending and Serious Crime: A Panel Data Analysis with Implications for Social Support Theory. (*Worrall*)	Aid to Families with Dependent Children, Per capita social service spending, General relief per recipient, Per capita general relief, Per capita family assistance	The results of the analysis by the authors are that there is little to no relationship between welfare spending and serious crime. **Note:** This result could be because they only looked at one state (California) and looked at the data at the county level. The author only considered male genders and African Americans as Race.
Analysis on Welfare Payments and Crime (*Foley*)	Burglary, Larceny-theft, Motor vehicle theft, Robbery, Arson, Assault, Homicide, Rape	Temporal patterns in crime were observed in jurisdictions in which disbursements are focused on the beginning of monthly welfare payment cycles and not in jurisdictions in which disbursements are relatively more staggered. Results indicate that crime rates increase with the amount of time since welfare payments occurred. (Foley)
Inequality and Crime (*Kelly*)	Population, Density, Income Gini, Female head, Nonwhite, Unemployed, Poverty, Movers, Police, College	For violent crime, inequality is significant, even after controlling for the effects of poverty, race, and family composition. Although the most disadvantaged members of society commit most crimes, these individuals face more significant pressure and incentives to commit crimes in areas of high inequality.
Dynamic linkages between poverty, inequality, crime, and social expenditures in a panel of 16 countries (*Anser and Nassani*) (*Anser et al.*)	Crime rate, GINI index, National estimates of unemployment, Education expenditures (% GDP), Per capita health expenditure, Per capita income, Trade openness as % GDP	No/flat relationship between per capita income and crime rate; U-shaped relationship between poverty headcount and per capita income Inverted U-shaped relationship between income inequality & economic growth in a selected country panel.
Estimating the impact of state government spending and the economy on crime rates (*Akpom and Doss*)	State population, Poverty rate, Median income, Unemployment rate, Gross State Product, Govt spending	The determinants of crime varied in both the category of crime and the period of study. In addition, government spending on welfare and education was not significant in 1990, but became more significant for the 2000 and 2010 samples.

	- Education, Govt spending - Protection, Population urban areas (%), Govt spending - Welfare	
Testing the Relationship Between Welfare Spending and Property Crimes (*Burek*)	Welfare, Unemployment, Population Size, Racial Composition, Age Structure of Population, Female-headed Household	Found a positive relationship between property crime and welfare spending. The authors noted a study conducted by Messner (1986), which found a positive relationship between larceny and welfare spending. They noted that the minuscule welfare amount leaves a person with unmet needs that can be supplemented with crime.
How does the welfare state reduce crime? The effect of program characteristics and decommodification across 18 OECD-countries. (*Maximilian and Starke*)	AFDC payments per recipient, per capita social service spending, general relief per recipient, per capita general relief, per capita family assistance	The welfare state suppresses crime mainly through social support via generous unemployment benefits. Overall decommodification, the critical measure to test IAT, however, does not affect homicide.

Data Source/Preparation

Data Source

We sourced our project data from:

1. FBI to get the crime data based on city size https://ucr.fbi.gov/crime-in-the-u.s
2. Lincoln Institute of Land Policy for city municipalities' spending data http://www.lincolninst.edu/research-data/data-toolkits/fiscally-standardized-cities/search-database

FBI data was from 2010 to 2019; each year was a separate download. FBI data contained a total of counts by crime type (violent, and non-violent) then split out by crime type. The column format is the following Violent crime: Murder and nonnegligent manslaughter, Forcible rape, Robbery, Aggravated assault. Property crime (changed non-violent in R code): Burglary, Larceny-theft, Motor vehicle theft. The crime data was not ready for statistical analysis, especially not formatted for joins to another dataset.

Preparation

A	B	C	D	E	F	G	H	I	J	K	L
Table 6											
Crime in the United States											
by Metropolitan Statistical Area, 2010											
Metropolitan Statistical Area	Counties/principal cities	Population	**Violent crime**	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Aggravated assault	**Property crime**	Burglary	Larceny-theft	Motor vehicle theft
Abilene, TX M.S.A.		**159,566**									
	Includes Callahan, Jones, and Taylor Counties										
	City of Abilene	116,938	578	4	68	112	394	4,897	1,340	3,375	182
	Total area actually reporting	100.0%	675	5	78	116	476	5,772	1,610	3,925	237
	Rate per 100,000 inhabitants		423.0	3.1	48.9	72.7	298.3	3,617.3	1,009.0	2,459.8	148.5
Akron, OH M.S.A.		**698,613**									
	Includes Portage and Summit Counties										
	City of Akron	205,760	1,665	22	159	602	882	10,851	4,247	5,855	749
	Total area actually reporting	94.2%	2,063	26	276	709	1,052	21,127	6,389	13,637	1,101
	Estimated total	100.0%	2,129	26	286	734	1,083	22,255	6,621	14,493	1,141
	Rate per 100,000 inhabitants		304.7	3.7	40.9	105.1	155.0	3,185.6	947.7	2,074.5	163.3

The issues were columns A and B in the exhibit shown above. The merged cells in column A made it challenging to correct the format for column B. Our group's solution was to text-split column A and then concatenated city and state names into separate columns. The biggest issue was if the state had multiple cities within one merged cell in column A.

Allentown-Bethlehem-Easton, PA-NJ M.S.A.		**828,068**									
	Includes Warren County, NJ and Carbon, Lehigh, and Northampton Counties, PA										
	City of Allentown, PA	108,473	732	9	67	460	196	5,089	1,327	3,345	417
	City of Bethlehem, PA	73,634	234	0	18	110	106	2,069	445	1,522	102
	Total area actually reporting	99.6%	1,883	29	168	773	913	18,967	3,569	14,508	890
	Estimated total	100.0%	1,890	29	168	775	918	19,029	3,579	14,558	892
	Rate per 100,000 inhabitants		228.2	3.5	20.3	93.6	110.9	2,298.0	432.2	1,758.1	107.7

Allowing our group a way to split cities into two cells. The wrangling was a lengthy process because of the different lengths of city names, with the addition to a cell having many cities within a cell. Every time this problem occurred, the cities were separated with a hyphen. The city name column was created once the text split and concatenation were complete.

Baton	Rouge					Baton Rouge
Battle	Creek					Battle Creek
Bay	City					Bay City
Beaumont						Beaumont
Port	Arthur					Port Arthur
Beckley						Beckley

A	B	C	D	E	F	G	H	I	J	K	L
State	City	Year	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft
TX	Abilene	2019	123665	458	6	87	68	297	3112	576	2330
OH	Akron	2019	197882	1782	27	181	328	1246	6568	1686	4305
GA	Albany	2019	74989	790	12	32	165	581	3452	729	2489

The above image shows the result of the wrangling process. This process was repeated for FBI crime data from 2010 to 2019.

The city data was spent per resident for cities over 250k population. The city dataset had spending data from all areas of spending, from education to sewer expenditures. The data included many exciting areas of spending, just not relevant to crime. Our group chose to remove most city expenditure categories and only keep those spending variables that affect the welfare of a city resident, either health or education. The city spend dataset did not need any wrangling for analysis.

Variable choice

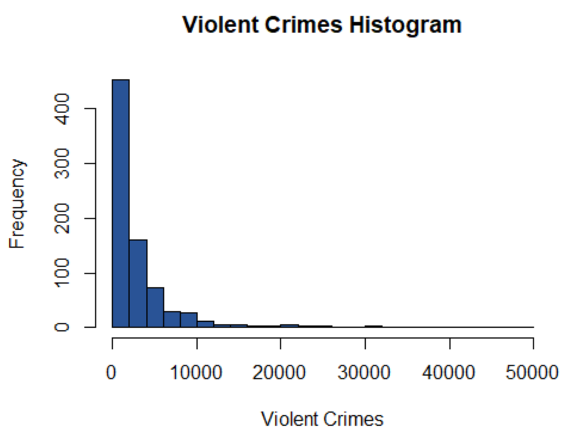
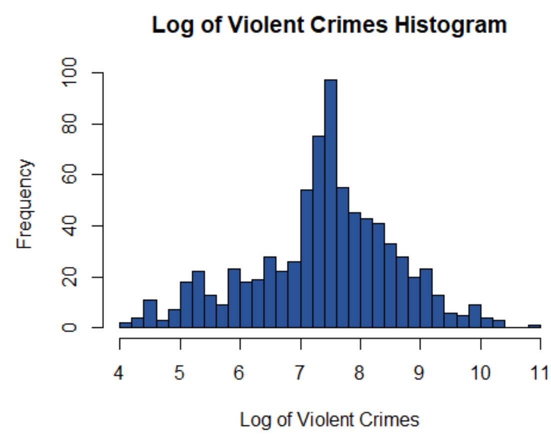
Over the last 60 years, welfare spending has increased at rates that taxpayers may be concerned about how much of the money spent is advantageous for the community both at a federal and local level. This report uses FBI crime data and city spending data over the last ten years. The project goal was to model the relationship between non-violent crimes and welfare spending in the cities with populations over 100,000 while comparing their crime rates in those cities from FBI crime data. Our model results show that welfare spending does not have a statistically significant effect on non-violent crime.

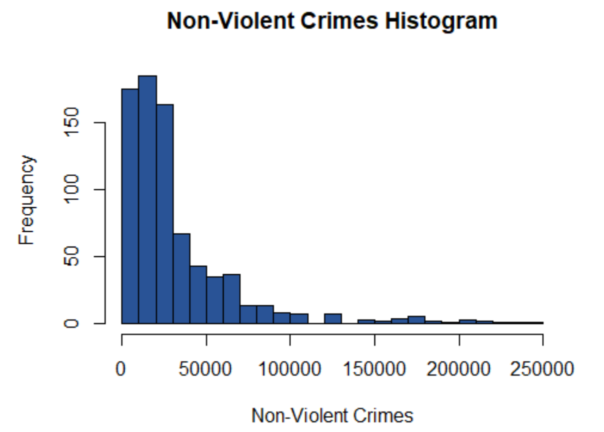
Predictor Table

Predictor	Effect	Rationale
education_services_exp	(-)	A significant percentage of criminals are school dropouts, so there is a strong relationship between crime in the city and school spending.
public_welfare_exp	(-)	More generous benefits such as insurance and social assistance are associated with a lower crime related to needs.
hospital_exp	(-)	Improved treatment facilities provide better health care to people reducing need-based crime.
health_exp	(-)	Expenditure on health, especially mental health, will help reduce crimes committed due to high stress.
police_protection_exp	(-)	Police protection is mandatory while dealing with criminals..
fire_protection_exp	(-)	Fire protection is required for dealing with arson.
corrections_exp	(-)	Corrections expenditures are for the operation, maintenance, and construction of prisons and jails and the activities of probation officers and parole boards. A higher budget could encourage more criminals to be in jail.
inspection_and_regulation	(-)	Regulatory inspection is essential for the workplace's safety, performance, and quality. So, a budget for this will help reduce crimes at the workplace.
housing_&_community_development_exp	(-)	Schemes of housing for the homeless and unemployed may help reduce crimes as homeless and unemployed people may be focused

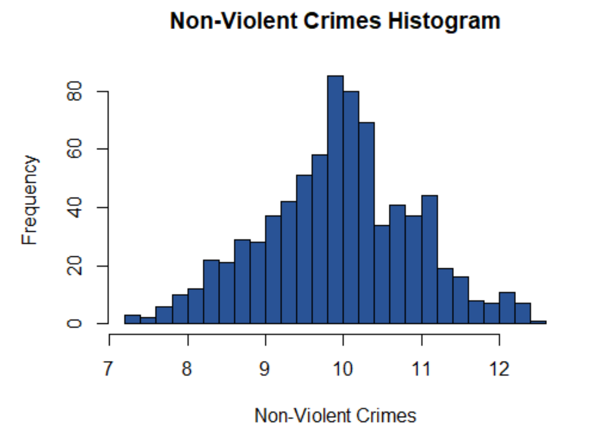
		on proper means to earn money to meet basic needs rather than focusing on paying rent which gives high stress and may lead to crime.
financial_administration_exp	(-)	Financial administration expenditure provides financial assistance to people with low income, which might help reduce crimes in need of money.
judicial_and_legal_exp	(-)	Expenditure on judicial and legal will help the judiciary work on cases fast and lessen the extended time, which helps maintain law and order.
other_governmental_admin_exp	(-)	Expenditure on governmental administration will support maintaining law and order.

Descriptive Analysis & Data Visualizations

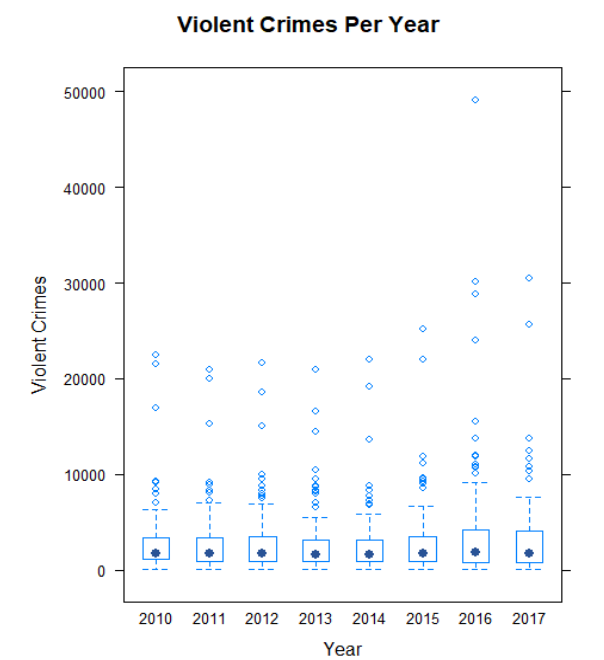
	
The distribution for violent crimes is very skewed to the right, which may cause errors when used as a predictor.	Using the log of violent crimes, the distribution becomes somewhat normal and usable.



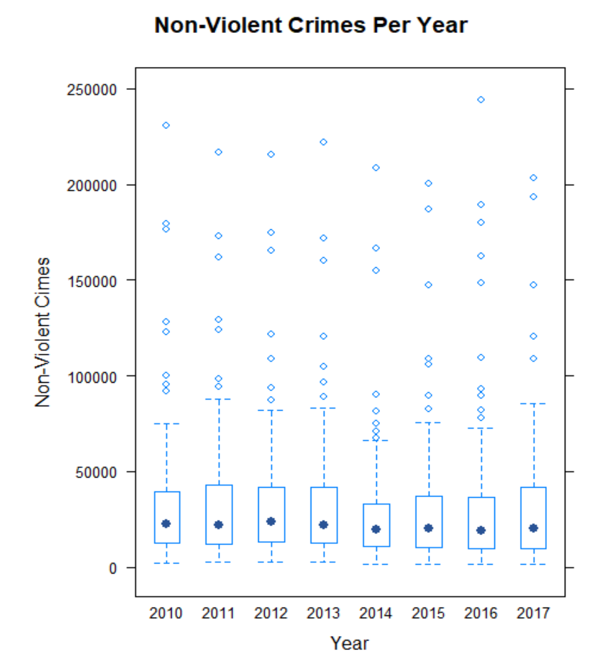
Looking at the distribution of non-violent crimes is skewed to the right.



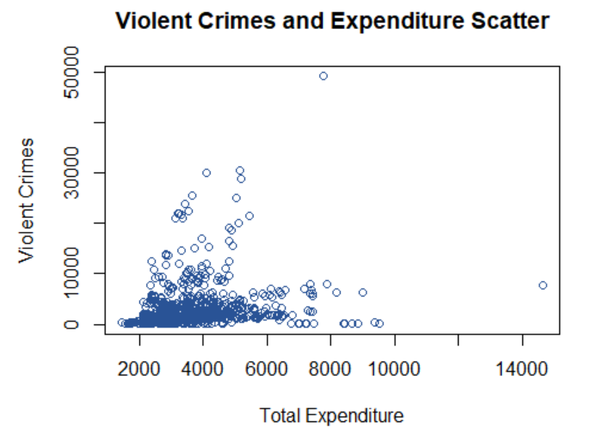
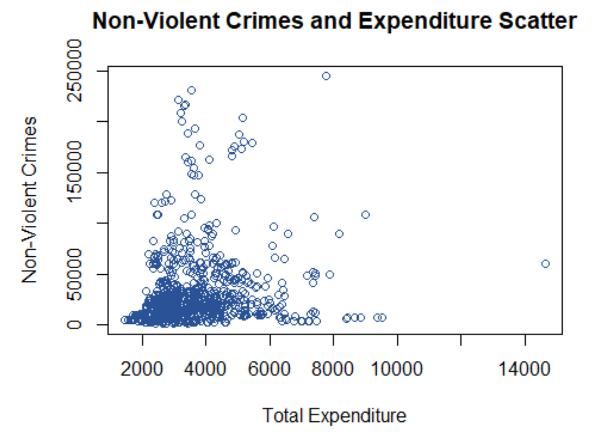
Looking at the distribution of non-violent crimes is skewed to the right.

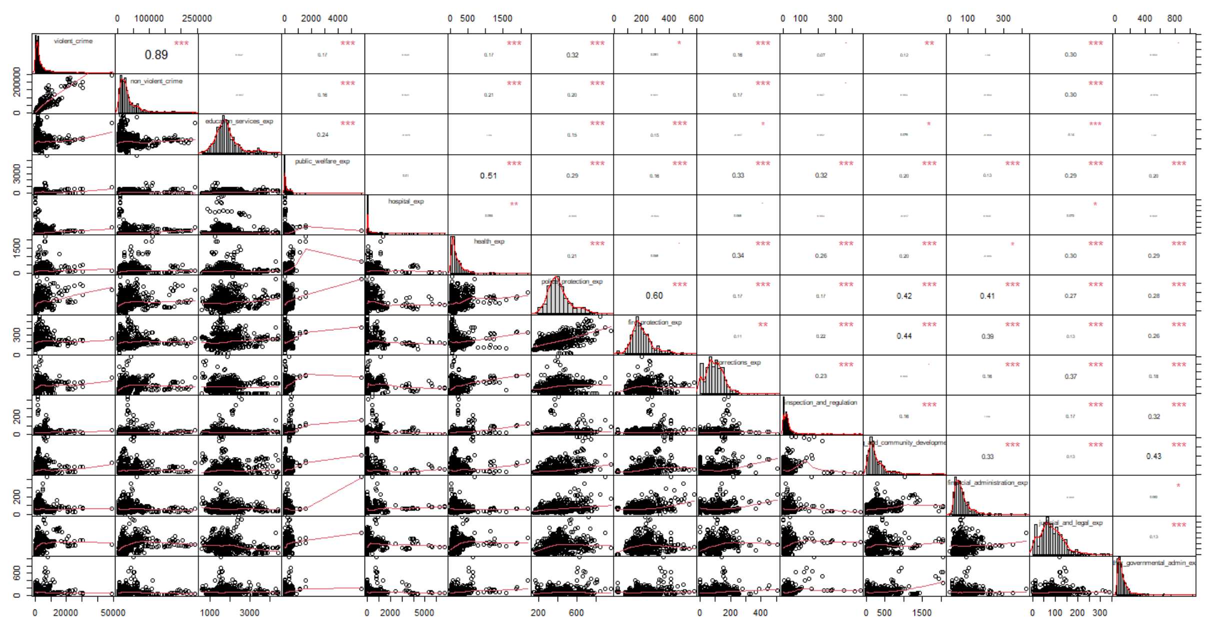


Looking at violent crimes over the years creates somewhat of a U-shape where 2013 and 2014 were the lowest years, but 2016 was extremely high. It even has an outlier whereby 2016 had the highest violent crime rate.



Looking at the non-violent crimes over the years, it is somewhat steady. We can see a similarity with violent crime; however, 2014 had the lowest crime rate, and 2016 had the highest crime rate.

	
Looking at the scatter plot to see if there is any relation between expenditure and crimes. It does not seem to have any homoscedasticity; it has a fanning shape.	We were looking at the non-violent crimes with the expenditure to see if there is any relation. There does not seem to be any relationship. However, there is a fanning shape.



The correlation matrix above shows how strongly/poorly each variable is correlated with one another. The only correlation that is high is between non-violent and violent variables. This is fine because these two variables are our Y variables and will not be used in models as an X variable.

Models

The models we used were mixed-level on random effects, Poisson, and Negative Binomial for both y variables (violent crimes, and non-violent crimes). We applied mixed-level random effects because our data is multi-leveled with city and state levels. We decided to use the Poisson Model as well because our Y variable is a "count" variable, whereby we are counting the number of crimes that have occurred in a city. We model both violent crimes and non-violent crimes on each of the selected models for comparability.

Violent Crimes

Random effects (state level):

lmer(log(violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp + year + (1 | state), data=df_m, REML=FALSE)

Random effects (city level):

lmer(log(violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp + year + (1 | city), data=df_m, REML=FALSE)

Poisson:

glm(violent_crime ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp, data=df_m, family = poisson (link = log))

Negative Binomial:

glm.nb(log(violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp +

inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp + year, data=df_m)

Non-Violent Crimes

Random effects (state level):

lmer(log(non_violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp + year + (1 | state), data=df_m, REML=FALSE)

Random effects (city level):

lmer(log(non_violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp + year + (1 | city), data=df_m, REML=FALSE)

Poisson:

glm(non_violent_crime ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp, data=df_m, family = poisson (link = log))

Negative Binomial:

glm.nb(log(non_violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp + other_governmental_admin_exp + year, data=df_m)

Quality Checks

Violent Crimes

Model	Test	Interpretation
Random Effects (State Level)	AIC = 1764.358 BIC = 1866.862	The BIC seems a little high for the State level. This value needs to be as low as possible.
Random Effects (City Level)	AIC = 167.5 BIC = 270.01	The BIC is much lower at the City level than at the State level. This value needs to be as low as possible.
Negative Binomial	AIC = 13877.24 BIC = 13975.08 Residual Dev = 885.17 on 760 D.F	The dispersion is better when compared to Poisson based on the Residual Deviance. However, the BIC value is high compared to the Random Effects Models.

Non-Violent Crimes

Model	Test	Interpretation
Random Effects (State Level)	AIC = 1467.18 BIC = 1569.69	The BIC seems a little high for the State level. This value needs to be as low as possible.
Random Effects (City Level)	AIC = -33.21 BIC = 69.29	The BIC is much lower at the City level than at the State level. This value needs to be as low as possible.
Negative Binomial	AIC = 17594.2 BIC = 17692.05 Residual Dev = 862.19 on 760 D.F	The dispersion is better when compared to Poisson based on the Residual Deviance. However, the BIC value is high compared to the Random Effects Models.

Stargazer Output for Random Effect Model

Crime Rate Analysis

	Dependent variable:	
	Violent Crime	Non-Violent Crime
Education Services	0.00001 (0.00004)	0.00004 (0.00004)
Public Welfare	-0.0004*** (0.0001)	-0.0001 (0.0001)
Hospital	0.0001* (0.0001)	0.00002 (0.0001)
Health	0.0003*** (0.0001)	0.0004*** (0.0001)
Police Protection	0.00002 (0.0002)	0.0002 (0.0002)
Fire Protection	0.0002 (0.0003)	-0.0004* (0.0002)
Correction	0.0002 (0.0002)	0.001*** (0.0002)
Inspection and Regulation	0.001* (0.001)	0.001*** (0.0005)
Housing and Community Dev	0.0001 (0.0001)	0.0001 (0.0001)
Financial Administration	0.001** (0.0003)	0.001* (0.0003)
Judicial and Legal	0.003*** (0.001)	0.002*** (0.0005)
Other Governmental Admin	0.0005*** (0.0002)	0.0002 (0.0001)
Year 2011	-0.020 (0.024)	0.012 (0.022)
Year 2012	0.001 (0.026)	0.033 (0.023)
Year 2013	-0.027 (0.026)	0.009 (0.023)
Year 2014	-0.017 (0.027)	-0.045* (0.024)
Year 2015	0.025 (0.026)	-0.062*** (0.023)
Year 2016	0.086*** (0.025)	-0.103*** (0.023)
Year 2017	0.068*** (0.026)	-0.100*** (0.024)
Constant	6.890*** (0.157)	9.522*** (0.131)

Observations	780	780
Log Likelihood	-61.750	38.607
Akaike Inf. Crit.	167.500	-33.213
Bayesian Inf. Crit.	270.005	69.291
Note:		*p<0.1; **p<0.05; ***p<0.01

Recommendations

Our recommendation is to re-evaluate the relationship between city crime rates for nonviolent crimes and welfare spending programs. Our model results show a nonsignificant relationship between nonviolent crimes and welfare spending programs. This model result could be because the welfare amount is not significant enough to reduce crime, as crime is still needed to supplement household income. Cities could increase allocated amounts of welfare to decrease the reliance on crime for additional household income.

References

Works Cited

Akpom, Uchenna N., and Adrian D. Doss. "Estimating the impact of state government spending and the economy on crime rates." *Law and Conflict Resolution*, vol. 10, no. 2, 2018, p. 9. *Estimating the impact of state government spending and the economy on crime rates*, https://academicjournals.org/journal/JLCR/article-abstract/40D750956167. Accessed 2022.

Anser, Muhammad Khalid, et al. "Dynamic linkages between poverty, inequality, crime, and social expenditures in a panel of 16 countries: two-step GMM estimates." *Journal of Economic Structures*, vol. 43, no. 9, 2022, p. 25. *Dynamic linkages between poverty,* *inequality, crime, and social expenditures in a panel of 16 countries: two-step GMM estimates*, https://journalofeconomicstructures.springeropen.com/track/pdf/10.1186/s40008-020-00220-6.pdf. Accessed 2022.

Burek, Melissa W. “Now Serving Part Two Crimes: Testing the Relationship Between Welfare Spending and Property Crimes.” *Criminal Justice Policy Review*, vol. 16, no. 3, 2005, p. 25, https://journals.sagepub.com/doi/abs/10.1177/0887403405274782?casa_token=zBZCiMTfh3gAAAAA:_CcTJKO2KrD84fmWtXv9dWMsIL1qRQ8xL92xBIehTYDrdHY6uUNe3hgPpXGbqc-gUtK_9p_Asqq4sQ. Accessed 2022.

Fishback, Price V., et al. “Striking at the Roots of Crime: The Impact of Welfare Spending on Crime during the Great Depression | The Journal of Law and Economics: Vol 53, No 4.” *The Journal of Law and Economics*, vol. 53, no. 4, 2010, p. 41. *The University of Chicago Press: Journals*, https://www.journals.uchicago.edu/doi/abs/10.1086/655778?journalCode=jle. Accessed 23 April 2022.

Foley, C. Fritz. *The Review of Economics and Statistics*, vol. 93, no. 1, 2011, p. 39. *Welfare Payments and Crime*, https://direct.mit.edu/rest/article-abstract/93/1/97/57922/Welfare-Payments-and-Crime?redirectedFrom=fulltext. Accessed 23 April 2022.

Kelly, Morgan. “Inequality and crime.” *Rev Econ Stat*, vol. 82(4), 2000, pp. 530–539. https://watermark.silverchair.com/003465300559028.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAA0IwggM-BgkqhkiG9w0BBwagggMvMII

DKwIBADCCAyQGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMSWOegAzr3zUE-9L
PAgEQgIIC9XaaclRB3gNymPTgyKlIRG98VwPvwxtwhh8NeJVp.

Marco Trends. “U.S. -2022.” *U.S. Crime*, Marco Trends, 2022, https://www.macrotrends.net/countries/USA/united-states/crime-rate-statistics'%3EU.S.%20Crime%20Rate%20&%20Statistics%201990-2022%20%E2%80%93%20World%20Bank. Accessed 23 April 2022.

Rudolph, Maximilian, and Peter Starke. “How does the welfare state reduce crime? The effect of program characteristics and decommodification across 18 OECD-countries.” *Journal of Criminal Justice*, vol. 68, no. 0, 2020, p. 12. *How does the welfare state reduce crime? The effect of program characteristics and decommodification across 18 OECD-countries*, https://www.sciencedirect.com/science/article/abs/pii/S0047235220300623?via%3Dihub. Accessed 2022.

Urban Institute. “Justice Expenditures.” *Criminal Justice Expenditures: Police, Corrections, and Courts*, Urban Institute, https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders/criminal-justice-police-corrections-courts-expenditures#Question2P. Accessed 23 April 2022.

Worrall, John L. “Reconsidering the Relationship Between Welfare Spending and Serious Crime: A Panel Data Analysis with Implications for Social Support Theory.” *Justice Quarterly*, vol. 3, no. 3, 2018, p. 29. *Reconsidering the Relationship Between Welfare Spending and Serious Crime*, https://www.tandfonline.com/doi/full/10.1080/07418820500219094?casa_token=Enf

Fyr5jUP8AAAAA%3AoZVxHh-yydS-r6qAf2T50qhsTO8dK-AaCAB2b3fNhF3VhrMqaU72

LWcKlvtgNyLH_PsQtp1ey0N4XA. Accessed 23 April 2022.

Appendix: R Code

```
################################
# Crime Rates Project

rm(list=ls())

################################
# Load Libraries
pacman::p_load(dplyr, tidyr, caret, ggplot2, caTools, MLmetrics, mlbench, mlTools, corrplot, expss,
PerformanceAnalytics, AER, MASS, stargazer, pscl, jtools, Hmisc, ggcorrplot, rpart, rpart.plot, readxl,
ROCR, lme4)

###############################
# Load crime data set
crime_df <- read.csv("C:/Users/Scott/Documents/Shyam_Personal/SDM
Assignment/Project/combined_crime_data.csv")

# Load expenditure data set
exp_df <- read.csv('C:/Users/Scott/Documents/Shyam_Personal/SDM
Assignment/Project/Expenditure_2010-2017_Updated_New.csv')

###############################
# Merge

# Ensure datatypes are matching for joining
str(crime_df)
str(exp_df)

# Convert the Year column to Integer for joining
crime_df$year <- as.integer(crime_df$year)

# Merge the two datasets
df <- merge(crime_df, exp_df, by=c("state", "city", "year"))

# Check for any NA values
colSums(is.na(df))

df <- df[complete.cases(df$forcible_rape), ]        # Drop incomplete rows

#############################
# Clean datasets
str(df)

# Convert values to factors
cols_factor <- c("state", "city", "year")
df[cols_factor] <- (lapply(df[cols_factor], factor))
```

```
levels(df$year)

# convert variables to integer
cols_num <- c("population", "violent_crime", "non_violent_crime",
"murder_and_nonnegligent_manslaughter", "forcible_rape",  "robbery", "aggravated_assault",
"property_crime", "burglary", "larceny_theft", "motor_vehicle_theft")
df[cols_num] <- (lapply(df[cols_num], as.integer))

# Create a calculated columns to see the total crimes in each city for that year
total_crimes <- df$violent_crime + df$non_violent_crime

# Create a new dataframe only with the columns that will be used
df_m <- subset(df, select = c(-population, -murder_and_nonnegligent_manslaughter, -larceny_theft,
-motor_vehicle_theft,   -forcible_rape, -robbery, -aggravated_assault, -property_crime, -burglary,
-parks_and_recreation_exp,  -general_public_buildings_exp))
str(df_m)

# Create a variable to get the total expenditure
total_expenditure <- df$education_services_exp + df$public_welfare_exp + df$hospital_exp +
df$health_exp + df$police_protection_exp + df$fire_protection_exp + df$corrections_exp +
df$inspection_and_regulation +   df$housing_and_community_development_exp +
df$financial_administration_exp + df$judicial_and_legal_exp +   df$other_governmental_admin_exp

##############################
# Visualization and Descriptive analysis

hist(df_m$violent_crime, main = "Violent Crimes Histogram", xlab = "Violent Crimes",  col = "#295396",
breaks = 30)
hist(log(df_m$violent_crime), main = "Log of Violent Crimes Histogram", xlab = "Log of Violent Crimes",
col = "#295396", breaks = 30)

hist(df$non_violent_crime, main = "Non-Violent Crimes Histogram", xlab = "Non-Violent Crimes",  col =
"#295396", breaks = 30)
hist(log(df_m$non_violent_crime), main = "Non-Violent Crimes Histogram", xlab = "Non-Violent Crimes",
col = "#295396", breaks = 30)

hist(total_crimes, main = "Total Crimes Histogram", xlab = "Total Crimes",  col = "#295396", breaks = 30)
hist(log(total_crimes), main = "Log of Total Crimes Histogram", xlab = "Log of Total Crimes",  col =
"#295396", breaks = 30)

options(scipen = 999)
bwplot(population ~ year, data=df, main = "Population Per Year", xlab = "Year", ylab = "Population", col =
"#295396")
bwplot(violent_crime ~ year, data = df, main = "Violent Crimes Per Year", xlab = "Year", ylab = "Violent
Crimes", col = "#295396")
bwplot(non_violent_crime ~ year, data = df, main = "Non-Violent Crimes Per Year", xlab = "Year", ylab =
"Non-Violent Cimes", col = "#295396")

plot(population ~ total_expenditure, data = df, main = "Population and Expenditure Scatter", xlab = "Total
Expenditure", ylab = "Population", col = "#295396")
plot(violent_crime ~ total_expenditure, data = df, main = "Violent Crimes and Expenditure Scatter", xlab =
"Total Expenditure", ylab = "Violent Crimes", col = "#295396")
plot(non_violent_crime ~ total_expenditure, data = df, main = "Non-Violent Crimes and Expenditure
Scatter", xlab = "Total Expenditure", ylab = "Non-Violent Crimes", col = "#295396")
```

```
plot(total_crimes~ total_expenditure, data = df, main = "Total Crimes Scatter", xlab = "Total Expenditure",
ylab = "Total Crimes", col = "#295396")
bwplot(total_crimes ~ year, data = df, main = "Total Crimes", xlab = "Year", ylab = "Total Crimes", col =
"#295396")

chart.Correlation(df_m[as.integer(which(sapply(df_m,class)=="integer"))]) # Plot for numeric variables

### Descriptive Analysis
# Sum for Totals
sum(total_expenditure)                      # Total Expenditure
sum(df_m$violent_crime)                     # Violent Crimes
sum(df$non_violent_crime)                   # Non-Violent Crimes
sum(total_crimes)                           # Total Crimes

# Summary
summary(total_expenditure)                  # Total Expenditure
summary(df_m$violent_crime)                 # Violent Crimes
summary(df$non_violent_crime)               # Non-Violent Crimes
summary(total_crimes)                       # Total Crimes

# Summary by Year
summary(total_expenditure ~ df_m$year)              # Total Expenditure by Year
summary(df_m$violent_crime ~ df_m$year)             # Violent Crimes by Year
summary(df_m$non_violent_crime ~ df_m$year)         # Non-Violent Crimes by Year
summary(total_crimes ~ df_m$year)                   # Total Crimes by Year

#############################
# Models

################
# Violent crimes

# Baseline
vio_b <- lmer(log(violent_crime) ~ 1 + (1 | city) + (1 | state), data=df_m, REML=FALSE)
summary(vio_b)

# Model using mixed level
# state level
vio_m0 <- lmer(log(violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp +
health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation
+ housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp +
other_governmental_admin_exp + year + (1 | state), data=df_m, REML=FALSE)
summary(vio_m0)
confint(vio_m0)
AIC(vio_m0)
BIC(vio_m0)
fixef(vio_m0)                        # Magnitude of fixed effect
ranef(vio_m0)                        # Magnitude of random effect
coef(vio_m0)                         # Magnitude of total effect

# city level
vio_m1 <- lmer(log(violent_crime) ~ education_services_exp + public_welfare_exp + hospital_exp +
health_exp +police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation
```

```
+ housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp +
other_governmental_admin_exp + year + (1 | city), data=df_m, REML=FALSE)
summary(vio_m1)
confint(vio_m1)
AIC(vio_m1)
BIC(vio_m1)
fixef(vio_m1)                          # Magnitude of fixed effect
ranef(vio_m1)                          # Magnitude of random effect
coef(vio_m1)                           # Magnitude of total effect

# Model using poisson method
vio_m2 <- glm(violent_crime ~ education_services_exp + public_welfare_exp + hospital_exp +
health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation
+ housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp +
other_governmental_admin_exp, data=df_m, family = poisson (link = log))
summary(vio_m2)
dispersiontest(vio_m2)

# Negative Binomial
vio_m3 <- glm.nb(violent_crime ~ education_services_exp + public_welfare_exp + hospital_exp +
health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation
+ housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp +
other_governmental_admin_exp + year, data=df_m)
summary(vio_m3)
AIC(vio_m3)
BIC(vio_m3)

# Stargazer between models
stargazer(vio_m1, vio_m3, type="text", single.row=TRUE)

################
# Non-Violent crimes

# Baseline
non_vio_b <- lmer(log(non_violent_crime) ~ 1 + (1 | city), data=df_m, REML=FALSE)
summary(non_vio_b)

# Model using mixed level
# state level
non_vio_m0 <- lmer(log(non_violent_crime) ~ education_services_exp + public_welfare_exp +
hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp +
inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp
+ judicial_and_legal_exp + other_governmental_admin_exp + year + (1 | state), data=df_m,
REML=FALSE)
summary(non_vio_m0)
confint(non_vio_m0)
AIC(non_vio_m0)
BIC(non_vio_m0)
fixef(non_vio_m0)                      # Magnitude of fixed effect
ranef(non_vio_m0)                      # Magnitude of random effect
coef(non_vio_m0)                       # Magnitude of total effect

# city level
```

```
non_vio_m1 <- lmer(log(non_violent_crime) ~ education_services_exp + public_welfare_exp +
hospital_exp + health_exp + police_protection_exp + fire_protection_exp + corrections_exp +
inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp
+ judicial_and_legal_exp + other_governmental_admin_exp + year + (1 | city), data=df_m,
REML=FALSE)
summary(non_vio_m1)
confint(non_vio_m1)
AIC(non_vio_m1)
BIC(non_vio_m1)
fixef(non_vio_m1)                    # Magnitude of fixed effect
ranef(non_vio_m1)                    # Magnitude of random effect
coef(non_vio_m1)                     # Magnitude of total effect

# Model using poisson method
non_vio_m2 <- glm(non_violent_crime ~ education_services_exp + public_welfare_exp + hospital_exp +
health_exp + police_protection_exp + fire_protection_exp + corrections_exp + inspection_and_regulation
+ housing_and_community_development_exp + financial_administration_exp + judicial_and_legal_exp +
other_governmental_admin_exp + year, data=df_m, family = poisson (link = log))
summary(non_vio_m2)
dispersiontest(non_vio_m2)

# Negative Binomial
non_vio_m3 <- glm.nb(non_violent_crime ~ education_services_exp + public_welfare_exp + hospital_exp
+ health_exp + police_protection_exp + fire_protection_exp + corrections_exp +
inspection_and_regulation + housing_and_community_development_exp + financial_administration_exp
+ judicial_and_legal_exp + other_governmental_admin_exp + year, data=df_m)
summary(non_vio_m3)
AIC(non_vio_m3)
BIC(non_vio_m3)

# Stargazer between models
stargazer(non_vio_m1, non_vio_m3, type='text', single.row = TRUE)

###########################
# Stargazer between the two variable types
options(max.print=10000)

stargazer(vio_m1, non_vio_m1, type="text", title = "Crime Rate Analysis", dep.var.labels = c("Violent
Crime", "Non-Violent Crime"), covariate.labels = c("Education Services", "Public Welfare", "Hospital",
"Health", "Police Protection", "Fire Protection", "Correction", "Inspection and Regulation", "Housing and
Community Dev", "Financial Administration", "Judicial and Legal", "Other Governmental Admin", "Year
2011", "Year 2012", "Year 2013", "Year 2014", "Year 2015", "Year 2016", "Year 2017"), align=TRUE,
single.row=TRUE, no.space = TRUE, out = "crime_rate_analysis.html")
```