Final Project Rough Draft

Alex Schmidt*

Apr. 23, 2024

Abstract

We analyze data on population of various townships and jurisdictions, alongside the population of the town, to investigate the claimed correlation between urbanization and criminal enterprises. To this end, data was scraped from publicly available collections, cleaned, and analyzed in RStudio. Results were analyzed in a Hierarchical Linear Model (HLM) and hypothesized in line with Strain Theory. The findings of this research contradict the hypotheses; while economic motivations due to low income were predicted to be the primary driver of crimes, there was no significant effect between the two. Population was also tested, to investigate the possibility of lawlessness based on opportunism. While population had no effect on the rate of violent crime, we find that property crimes correlate significantly with the population of a city, where bigger cities exhibit more property crime.

*Department of Marketing and Supply Chain Management, University of Oklahoma. E-mail address: student.name@ou.edu

1 Artificial Intelligence Disclosure

This paper utilized artificial intelligence as a tool for the completion of the research conducted, namely ChatGPT-3.5 and ChaGPT-4.0. ChatGPT was used in the following roles:

- Creating, troubleshooting, annotating, interpreting, and amending code, including for RStudio (which was used to harvest, clean, and interpret data, as well as to generate displays) and LaTeX (which was used to produce this document format)
- Searching the internet for existing academic research[1]
- Compiling perspectives on methodologies and arguments
- Low-level mathematics and data cleaning.

The research team feels that the utilization of artificial intelligence in this project amount to the utilization of any other research tool or methodology and stands by the academic integrity of the paper. All other aspects of the paper, including the review of the literature, the selection of theories, and the interpretation of data, were fundamentally human in conduct.

2 Introduction

OklahomaWatch claims ”From year to year, it’s not generally the largest cities in Oklahoma that have the highest crime rates” (Oklahoma Watch, 2017). The FBI report from which OklahomaWatch draws data seems to oppose this at the national level: of the 20 highest violent crime

[1] Although the hallucination rate for ChatGPT-4.0 has been dramatically reduced from previous versions, all referenced research was verified in accuracy by the research team, including the references to the articles and the relevance of the claims made to the content and context of the article.

rates for cities in the United States, the average violent crime rate (per 100,000 residents) was 1464.65 (Federal Bureau of Investigation, 2019), and the average population at the nearest census was 580,175 (United States Census Bureau, 2021), far higher than the average violent crime rate for the United States as a whole (368.9 per 100,000).

At first glance, this would suggest a correlation between crime and population. However, a careful examination of the data and claims reveals that there are multiple types of crimes, which may be affected alternatively by population. We extend the analysis of this information by breaking down the types of crimes and examining how the population of the reporting district affects criminality. In so doing, we seek to expand perspectives of criminality and shed light on how different factors and assumptions might drive different forms of crime.

Moreover, breaking down the types of crimes— such as property crimes and violent crimes— may provide a nuanced view of criminality. For example, larger cities might report higher rates of property crimes and theft, possibly driven by greater anonymity and a larger number of potential targets, while violent crimes could be concentrated in areas with specific socio-economic issues regardless of city size.

By examining the population of the reporting district in conjunction with these variables, we seek to expand perspectives of criminality and shed light on how different factors and assumptions might drive different forms of crime.

3 Literature Review

Extant literature focusing on criminality tends to concentrate on ecological factors such as socioeconomic status or the roles of social institutions. Indeed, considering an individual-level predictor

in isolation makes little sense without an accompanying theoretical framework to explain why criminally predisposed individuals might be drawn to urban environments. In this vein, significant scholarly attention has been directed towards Strain Theory, which is a critical viewpoint in the field of criminology.

Strain Theory is primarily an economic perspective that underscores the intense competition for scarce resources. According to this theory, criminal activity is often seen as a viable method for achieving socially-desirable goals that are in alignment with the normative expectations of the actor's environment. This perspective suggests that when legitimate means to achieve these goals are blocked, individuals may turn to crime as an alternative strategy to fulfill their needs and ambitions.

However, this theory opens the door for a seemingly contradictory perspective. It is generally acknowledged that urban areas, or cities, tend to offer more employment opportunities (Glaeser and Maré, 2001), and these job opportunities are typically better paid *ibid*. From this standpoint, one might logically assume that the economic competition for resources should be mitigated in these densely populated areas, which could theoretically lead to a reduction in crime rates. This presents a complex scenario where the availability of more and better economic opportunities in urban settings might paradoxically both decrease the need to engage in criminal behavior due to improved access to resources, and increase criminal opportunities due to larger populations and greater anonymity.

These competing hypotheses offer paradoxical predictions that challenge simplistic interpretations of urban crime dynamics. They suggest a nuanced landscape where the effects of urbanization on crime are mediated by a complex interplay of socioeconomic factors, availability of resources, and individual motivations. As such, the relationship between urban density and crime rates re-

mains a fertile area for further research, necessitating more sophisticated analytical frameworks to unravel the intricate patterns of influence exerted by various ecological and economic factors. This rich area of study invites researchers to delve deeper into the socioeconomic underpinnings of criminal behavior, particularly in urban settings, to better understand the mechanisms through which environmental factors influence crime.

Foundational Framework

The inception of Strain Theory was marked by Merton's 1938 article, 'Social Structure and Anomie,' where he proposed that society's structure exerts pressure on individuals to achieve culturally defined goals, often without providing adequate means (Merton, 1938). This discrepancy leads to strain, which can manifest in deviant behavior when the means to achieve goals are limited by the social structure.

Extensions and Refinements

Strain Theory was further refined and extended by researchers interested in the nuances of individual reactions to societal pressures:

- Cohen (1955) introduced the concept of status frustration to explain delinquency among lower-class boys who lack the means to achieve success in middle-class terms. Under this perspective, rejection of the dominant culture leads to adoption of the counterculture, which (by virtue of opposing the law-making dominant culture, is often opposed to the law itself).
- Cloward and Ohlin (1960) expanded on this by distinguishing between different types of delinquent subcultures that arise from variations in access to illegitimate means, proposing

that opportunity structures shape the forms of delinquency.

Contemporary Applications

Today, Strain Theory continues to influence criminological research and policy development. It is particularly relevant in studies exploring the impact of economic disparities and the effects of rapid social change on crime rates. Strain Theory principles are cited widely in the academic literature as well as in debating public policy and informating judicial sentencing.

Data

This study utilized data extracted from the Oklahoma Watch website, which provides comprehensive crime statistics across various towns within the state. The data were collected using the `rvest` package in R, facilitating the scraping of information directly from the website's crime statistics page.

The variables extracted include town name, total offenses, offense rate, violent crimes, violent crime rate, property crime, and property crime rate.

The data collection process involved identifying the appropriate CSS selectors to effectively extract each piece of information. The population and crime rates, originally in string format and often containing commas as thousand separators, were converted into numeric data types more suitable for analysis. Special characters and non-numeric strings which had been inadvertently extracted during the scraping process were also cleaned and removed.

Post-extraction, the processed data were stored in a structured data frame within R. Entries with missing or incomplete data were removed to maintain the integrity of the dataset. The final

dataset was then saved for further analysis, containing 325 complete entries.

The cleaned data enabled the construction of visualizations to explore relationships between population size and various crime rates, employing the `ggplot2` package. Plots were generated to display the distribution of offense, violent crime, and property crime rates across different population sizes, with additional adjustments made to exclude statistical outliers and focus on core trends.

After cleaning the data, another section of the script was created to scrape population density from Wikipedia, where it was automatically generated from a combination of the nearest census (United States Census Bureau, 2021) and the United States Geological Survey [2]. The script then created a dataframe of the towns and density. Towns which were previously removed from the list due to incomplete data were also excluded here.

These towns were then assigned to counties by scraping the Wikipedia page for the associated town (also informed by the U.S. Geological Survey)(*ibid*). Those which could not be identified positively (ex. when a town name coincides with the county name, and cannot be discerned with certainty) were removed. This left 243 entries representing 75 of Oklahoma's 77 counties. This is possibly due to tribal nations electing not to report crimes to the FBI (Oklahoma Watch, 2017).

The towns and associated counties were then assigned a county-wide income based on the United States Bureau of Economic Analysis ((Bureau of Economic Analysis, 2023)). Unfortunately, this was the most granular data available. Because this data could not be directly associated with the effects of population, a Hierarchical Linear Model (HLM) was constructed. This approach was selected because it can appropriately handle the nested nature of the data (i.e. towns within

[2] It would, of course, be possible to create the density ourselves from the population we've already scraped and the size of the town, but the U.S. Geological Survey website is much harder to scrape.

counties) and help to attribute the outcomes seen to the individual reporting agency or the broader area it was a part of. Through the HLM research design, it was possible to isolate the effects of population in the precinct separately from the income of the county.

$$\text{Offense Rate} = \beta_0 + \beta_1 \times \text{Per Capita Income} + \beta_2 \times \text{Population} + b_{\text{County}}$$

$$\text{Violent Crime Rate} = \beta_0 + \beta_1 \times \text{Per Capita Income} + \beta_2 \times \text{Population} + b_{\text{County}}$$

$$\text{Property Crime Rate} = \beta_0 + \beta_1 \times \text{Per Capita Income} + \beta_2 \times \text{Population} + b_{\text{County}}$$

Methodlogy

The equations describe the hierarchical linear models used to analyze crime rates based on socio-economic factors. Each equation represents a different crime rate (offense, violent crime, and property crime) as a function of Per Capita Income and Population, with a random effect for County.

These models aim to understand how variations in income and population size at the county level relate to different types of crime rates. The coefficients ($\beta 0$, $\beta 1$, $\beta 2$) represent the impact of income and population on crime rates, while the random effect for County captures unobserved heterogeneity across counties. By accounting for these factors, the models provide insights into the complex relationship between socio-economic conditions and crime rates, which is crucial for informing effective policy interventions and crime prevention strategies.

An interactive effect between population and income was considered, but it was found to exhibit excessive co-linearity (.993), indicating a high likelihood that small towns were more or less inherently predisposed to having low incomes. Researchers speculate that low incomes resulting from a lack of available work contribute directly to the lack of population in a town.

4 Research Findings

OklahomaWatch claimed that "From year to year, it's not generally the largest cities in Oklahoma that have the highest crime rates." This is more or less true: there is no statistically significant effect between the population of a city and the overall crime rate. However, this is not the full story.

Digging deeper reveals that the different rates of crime *are* affected differently based on population. Breaking down the offense rate to violent crimes and property crimes allows them to be analyzed together. The findings indicate that while violent crime is not significantly associated with larger cities, property crimes have a very significant (¡ 0.01) and positive correlation with the population of a reporting precinct. However, the size of this effect is negligible. Additional testing was done in an attempt to normalize the crimes and see if the effect could be assessed more readily by including the variance within the model. However, this did not alter the effect size, which remained negligible. It is possible that the significance of the effect is the result of the large data, although future studies should work to test this possibility.

This result also contradicts the expected result that property crimes (according to Strain Theory, the most likely to be financially motivated) are to be blamed on resource scarcity. Instead, we find no evidence of any significant relationship between income and criminality of any of the three

forms tested. The findings imply a different narrative: that property crimes may instead be an issue of opportunity, which may present itself more often in communities in which residents do not often know each other.

Future research should be conducted with the aim of exploring population density, as we only measure population in the total area. For some locations, the total population might be small principally because the jurisdiction is small, which does not qualify it to be categorized as ”small” in the same way that small towns are. In addition, spatial analysis should examine how districts might aid each other by sharing policing resources; if a county's neighbors are struggling with lawlessness, they may receive less aid than counties whose neighbors have an excess of resources. Finally, dummy variables should represent which state or states a county might border. For instance, McCurtain County borders both Arkansas and Texas, which may lead to their resources being dedicated to supporting federal coordination efforts on interstate issues.

Resolving these issues promises to improve the ability to successfully allocate police resources to supporting communities struggling with crime by directing equitable policing to underserved districts, which might receive insufficient attention due to population or income. In addition, this may help to inform sentencing guidelines for allegedly financially motivated crimes by opposing the common notion that poverty is the main driver of crime, and may alleviate stigmatization of poverty-stricken regions and individuals by dispelling the myth that they are predisposed to criminal behavior.

5 References

Bureau of Economic Analysis. (2023, November 16). *Personal Income by County and Metropolitan Area, 2022*. Retrieved from https://www.bea.gov/sites/default/files/2023-11/lapi1123.pdf

Cloward, R. A., & Ohlin, L. E. (1960). *Delinquency and Opportunity: A Theory of Delinquent Gangs*. New York, NY: Free Press.

Cohen, A. K. (1955). *Delinquent Boys: The Culture of the Gang*. Glencoe, IL: Free Press.

Federal Bureau of Investigation. (2019). *Crime in the United States, 2018*. Retrieved from https://www.fbi.gov/stats-services/crimestats

Glaeser, E. L. & Maré, D. C. (2001). Cities and skills. *Journal of Labor Economics*, *19*(2), 316-342.

Merton, R. K. (1938). Science and the social order. *Philosophy of Science*, *5*(3), 321-337.

Oklahoma Watch. (2017). *Crime in Oklahoma*. Retrieved from https://oklahomawatch.opalstacked.com/crime/ (Originally published 2017; last updated September 24, 2018)

United States Census Bureau. (2021). *2020 Census of Population and Housing*. Retrieved from https://www.census.gov/programs-surveys/decennial-census/2020-census.html

@bookcloward1960, author = Richard A. Cloward and Lloyd E. Ohlin, title = Delinquency and Opportunity: A Theory of Delinquent Gangs, publisher = Free Press, year = 1960, address = New York, NY

@bookcohen1955, author = Albert K. Cohen, title = Delinquent Boys: The Culture of the Gang, publisher = Free Press, year = 1955, address = Glencoe, IL

@miscfbi2019, author = Federal Bureau of Investigation, title = Crime in the United States, 2018, year = 2019, howpublished = https://www.fbi.gov/stats-services/crimestats

@articleglaeser2001, author = Edward L. Glaeser and David C. Maré, title = Cities and Skills, journal = Journal of Labor Economics, year = 2001, volume = 19, number = 2, pages = 316–342

@articlemerton1938, author = Robert K. Merton, title = Science and the Social Order, journal = Philosophy of Science, year = 1938, volume = 5, number = 3, pages = 321–337

@miscokwatch2017, author = Oklahoma Watch, title = Crime in Oklahoma, year = 2017, howpublished = https://oklahomawatch.opalstacked.com/crime/, note = Originally published 2017; last updated September 24, 2018

@misccensus2021, author = United States Census Bureau, title = 2020 Census of Population and Housing, year = 2021, howpublished = https://www.census.gov/programs-surveys/decennial-census/2020-census.html

6 Appendix

Name	Description
Town Name	The name of the town for which crime data were reported.
Population	The total population of the town.
Total Offenses	The total number of criminal offenses reported.
Offense Rate	The rate of offenses per 100,000 people.
Violent Crimes	The count of violent crimes reported.
Violent Crime Rate	The rate of violent crimes per 100,000 people.
Property Crimes	The count of property crimes reported.
Property Crime Rate	The rate of property crimes per 100,000 people.

Table 1: Description of Variables Used in the Study

Table 2: Summary of total crime rate model

Coefficient	Estimate	SE	df	T_value	P_value
(Intercept)	52.81	10.89	365.10	4.85	0.00
Per_Capita_Income	-0.00	0.00	379.65	-0.57	0.57
Population	0.00	0.00	349.68	1.43	0.15

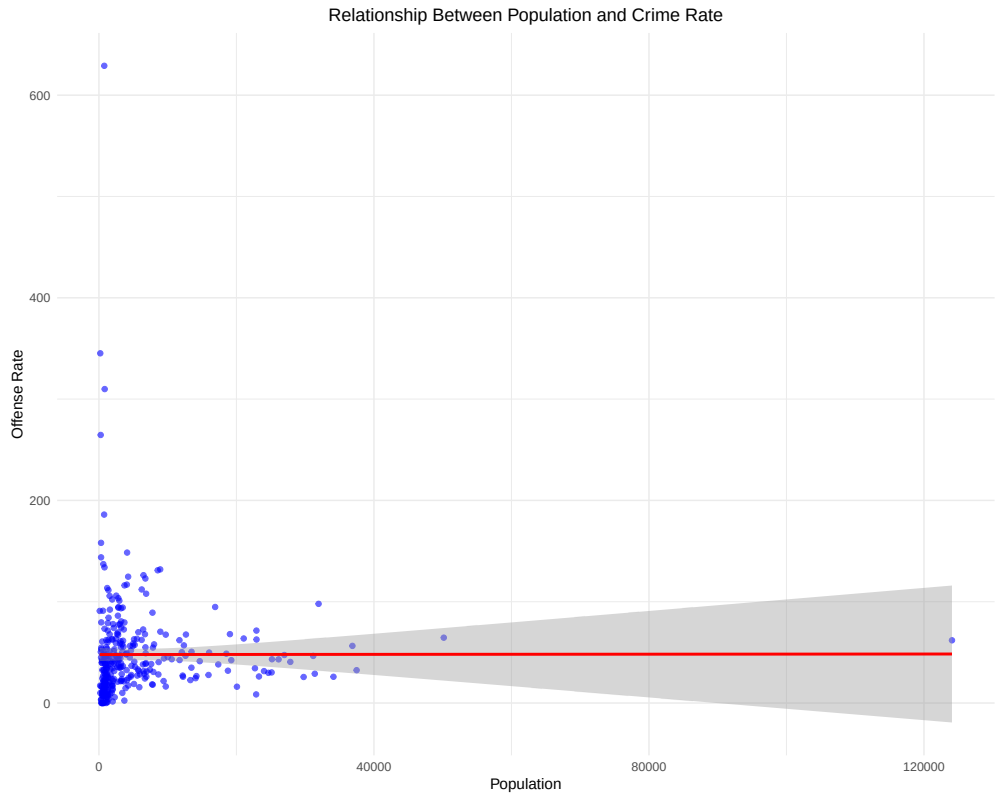

Figure 1: The relationship between population size and crime rate. A linear model fit highlights the trend.

Table 3: Summary of violent crime rate model

Coefficient	Estimate	SE	df	T_value	P_value
(Intercept)	2.15	0.45	323.22	4.81	0.00
Per_Capita_Income	-0.00	0.00	334.45	-1.21	0.23
Population	0.00	0.00	381.38	0.50	0.62

Table 4: Summary of property crime rate model

Coefficient	Estimate	SE	df	T_value	P_value
(Intercept)	17.81	3.68	356.12	4.85	0.00
Per_Capita_Income	-0.00	0.00	379.98	-0.51	0.61
Population	0.00	0.00	352.69	4.00	0.00