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. The findings of this research indicate the possibility that

We find ...

*Acknowledgements here, if any.

†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. ChatGPT-4.0 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" $(_{c}ite).This is a popular perspective: .The FBI report from which Oklahoma Watch draw$
$of the 20 highest violent crime rates for cities in the United States, the average violent crime rate (per 100,000 residents)$

[1] Although the hallucination rate for ChatGPT-4.0 has been dramatically reduced from previous versions, all cited 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.

At first glance, this would suggest a correlation between crime and population. However, a careful examination of the data and claims reveals that the data-supported statement "violent crime corresponds to high population" was inappropriately extended to the unsubstantiated claim "crime (here, including property crimes) corresponds to a high 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.

3 Literature Review

Extant literature focusing on criminality tends to focus on ecological factors such as socioeconomic status or social institutions. Indeed, an individual-level predictor makes little sense without an adjoining theory as to why criminally predisposed individuals are attracted to cities. In particular, significant attention has been given to Strain Theory.

Strain Theory is an economic perspective emphasizing the competition for resources.

However, this opens the door for a seemingly contradictory perspective: as there are more jobs in cities $(_{c}ite)$ $and as these jobs are generally better paid(_{c}ite), it would seem that economic competition for resources$

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 (**?**). This discrepancy leads to strain, which can manifest in deviant behavior when the means to achieve goals are limited by the social struc-

ture.

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 (**?**).
- 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 (**?**).
- Agnew (1992) significantly broadened the scope of Strain Theory with his General Strain Theory (GST), which considers a wider range of stressors beyond economic constraints, including emotional triggers and personal loss, as causes of strain (**?**).

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.

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:** 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.

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 suitable for analysis. Special characters and non-numeric strings inadvertently extracted during the scraping process were 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.

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.

4 Research Findings

OklahomaWatched claimed that "From year to year, it's not generally the largest cities in Oklahoma that have the highest crime rates." However, we find little $({}_{d}owe?)support\,for\,this\,broad\,claim$.