The Impact of Economic and Education Factors on Crime Rates: A Neighborhood-Level Analysis

Wanling Ma

26 January, 2024

This study delves into the multifaceted factors influencing neighborhood crime rates, employing a robust analytical approach using R (R Core Team 2023). By meticulously analyzing diverse data sources, we uncover significant correlations between socio-economic indicators and crime frequencies. Our findings reveal that specific socio-economic conditions, notably income levels and educational attainment, exhibit a strong predictive relationship with neighborhood crime rates. This research not only enhances our understanding of the dynamics of urban crime but also offers valuable insights for policymakers and community leaders in formulating targeted strategies to mitigate crime in vulnerable areas.

Table of contents

1. Acknowledgments

The source code and data for this report are available on GitHub at link.[1]

2. Introduction

In the ever-evolving landscape of urban development and public safety, the intricate relationship between socio-economic factors and crime rates stands as a critical area of study. Recent studies, such as the one conducted by Cai (2021), have highlighted the profound impact of socio-economic factors, including median household income, income inequality, and unemployment rates, on hate crime rates in the United States. This paper delves into the depths of this complex interaction, focusing specifically on the influence of income levels and educational attainment on neighborhood crime rates in Toronto. Amidst growing concerns about urban safety, this research addresses a significant gap in our understanding of how varied socio-economic backgrounds can shape crime dynamics within city environments.

Our investigation utilizes comprehensive data sets from the City of Toronto, including the "Neighborhood Crime Rates" and "Neighbourhood Profiles", to explore the correlations between economic circumstances, educational backgrounds, and crime occurrences. This analysis, grounded in 2015 and 2016 data, reveals insightful patterns and correlations, particularly highlighting the nuanced role of education (Section 3.3.2 Analysis 1: Crime Rate vs. Education Attainment and Section 3.3.3 Analysis 2: Crime Rate vs. Major Field of Study) and household income level (Section 3.3.4 Analysis 3: Crime Rate vs. Household After-Tax Income) in the context of neighborhood crime rates. We observe, for instance, a notable negative correlation between higher income brackets and shooting crime rates, and intriguing associations between educational levels and specific crime types.

The findings of this study are pivotal in extending our knowledge of urban crime dynamics. They challenge preconceived notions about socio-economic influences on crime and offer essential insights for policymakers and community leaders. This research underscores the necessity of considering economic and educational factors in crafting effective strategies to reduce crime and enhance public safety in diverse urban settings.

In detailing our journey, the paper is structured as follows: We begin with an in-depth methodology section, outlining our data collection and analysis approach. This is followed by the presentation of our findings, discussing the implications of these results. Finally, we conclude with practical recommendations for policymakers and suggestions for future research in this vital field.

[1] Github link: https://github.com/WanlingMa/EconomicImpactOnCrime.git

3. Data

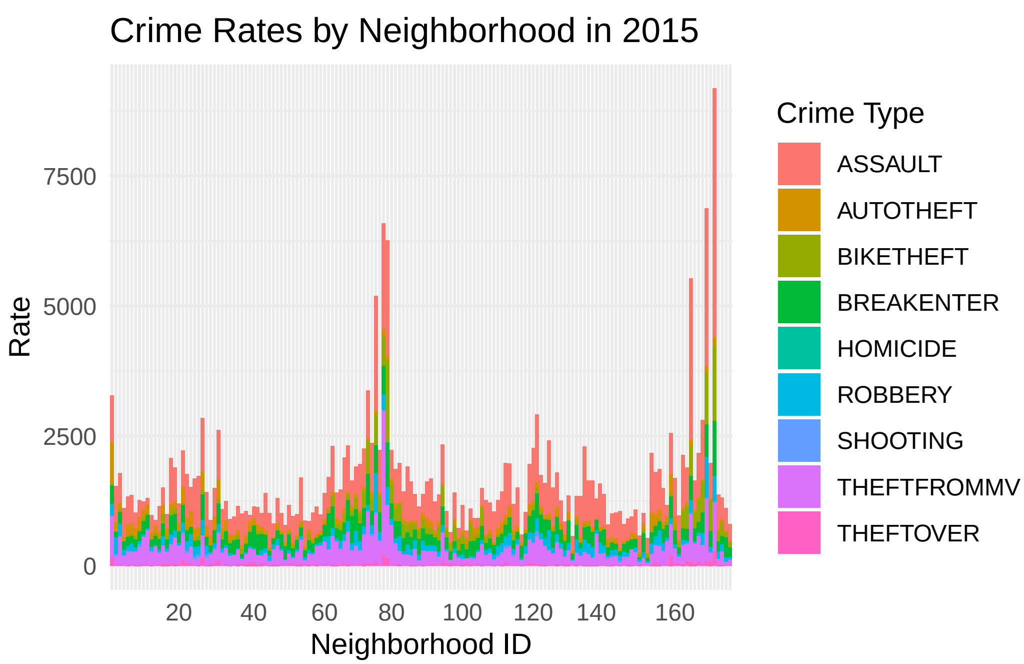
3.1 Data Sources

This analysis utilizes two primary data sets, both published by authoritative bodies within the City of Toronto. The first data set, Neighborhood Crime Rates, is published annually by the Toronto Police Services and includes crime counts and rates for various offences across the 158 City of Toronto Neighborhoods. The crime rates are calculated per 100,000 population, using estimates from Environs Analytics, which allows for consistent comparison across geographic areas and over time (Toronto Open Data 2024).

The second data set, Neighborhood Profiles, is derived from the Census of Population conducted every 5 years across Canada. Published by the Social Development, Finance & Administration division of the City of Toronto, this data set provides a comprehensive portrait of the demographic, social, and economic characteristics of individuals and households within the same 158 neighborhoods (Toronto Open Data 2023).

3.2 Data Overview

Here we will visualize the processed Neighborhood Crime Rates data set to present a overview of what the data set looks like:



The bar graphs illustrate substantial disparities in crime rates across different neighborhoods, prompting a critical inquiry into the underlying factors contributing to this phenomenon and seeking preventive measures to mitigate future occurrences of criminal activities.

3.3 Data Analysis

3.3.1 Data Analysis Methodology

3.3.1.1 Correlation Matrix

A correlation matrix is an organized representation of the correlation coefficients between a set of variables. In essence, it is a table where each element quantifies the degree to which two variables share a linear relationship.

To illustrate, consider the matrix $\mathbf{R}$. The elements of $\mathbf{R}$, denoted as r_{ij}, measure the strength and direction of the association between the i^{th} and j^{th} variables. These coefficients range from -1 to 1. A value of 1 indicates a perfect positive linear relationship, where an increase in one variable exactly predicts an increase in the other. Conversely, a value of -1 signifies a perfect negative linear relationship, implying that an increase in one variable predicts a decrease in the other. When r_{ij} is 0, it indicates no linear relationship between the variables.

In another words, a correlation matrix is a table that tells us how related different things are to each other. For each pair of items we're comparing, we calculate a number known as a 'correlation coefficient.' This number can be anywhere between -1 and 1. If the number is close to 1, it means that there's a strong positive relationship – as one item increases, the other one tends to as well. If the number is close to -1, it means there's a strong negative relationship – as one item goes up, the other goes down. A number close to 0 means there isn't much of a relationship at all.

Correlation matrices are particularly useful in identifying patterns within large sets of data, where manual assessment of relationships would be impractical. For example, in a dataset containing various socio-economic factors and crime rates, a correlation matrix can rapidly illuminate which factors share the strongest relationships with specific types of crime.

However, while a correlation matrix can be revealing, it is imperative to remember that correlation does not imply causation. The presence of a correlation does not guarantee that one variable influences the other; instead, it merely suggests a relationship that may be due to a variety of underlying causes or coincidences. Thus, any observed correlations warrant further research and analysis to determine the nature of the relationship and the potential influence of other variables.

In summary, a correlation matrix is a powerful statistical tool that provides initial insights into potential relationships between variables, necessitating cautious interpretation and subsequent in-depth investigation to uncover the underlying causal mechanisms.

3.3.2 Analysis 1: Crime Rate vs. Education Attainment

In this analysis, we investigate the hypothesis that educational attainment and field of study are key factors influencing neighborhood crime rates in 2015. Using data from Toronto Open Data Portal, we compare various educational levels and majors against crime rates, employing a correlation matrix and heatmap visualization to identify potential correlations. This approach seeks to uncover whether education plays a significant role in the dynamics of neighborhood crime.

We first compare various crime rates versus the highest degrees that individuals acquired:

Here we use kable to present some samples of processed data (Xie 2023):

	HOOD_ID	ASSAULT_RATE_2015	AUTOTHEFT_RATE_2015
24	139	897.3905	88.55827
25	138	775.0925	91.44350
26	136	1429.3468	65.13480
27	135	783.6725	95.84484
28	134	392.9727	84.75882
29	133	277.8387	51.18081

	No certificate, diploma or degree	Secondary (high) school diploma or equivalency certificate	Trades certificate or diploma other than Certificate of Apprenticeship or Certificate of Qualification
24	2955	3785	355
25	3925	5645	605
26	4990	6425	765
27	2465	4345	450
28	1625	3200	345
29	1295	2820	235

3.3.2.1 Data Visualization

We used cor() function to create a correlation matrix to quickly determine whether or not some specific crimes rate have correlation to the education levels that each individual in the neighbourhood acquired (R Core Team 2023). And then we use ggplot2 library to generate a heatmap to provide an elegant and efficient data visualization (Wickham et al. 2023). The heat map can be found as Figure 1.

3.3.2.2 Implication

From the Figure 1, we observe some interesting data points from the correlation matrix.

1. Earned Doctorate and Bike theft Rate:

The correlation coefficient between Bike theft Rate and Earned Doctorate is relatively high, which is 0.42. This would suggest that the number of residents that earned doctorate degree has a relatively strong correlation with the amount of bike theft cases. A higher incidence of bike theft in areas with more residents holding doctorates might be due to the prevalence of biking in university-centric urban areas, where doctorate holders often work and live. Increased opportunity for theft, alongside a potentially higher reporting rate, could account for this correlation.

2. Decrease in Crime Rates with Higher Education Levels:

As education levels rise, crime rates tend to fall, likely due to improved employment prospects and income levels, which reduce economic incentives for crime. Additionally, neighborhoods with more educated residents may experience enhanced community engagement and social cohesion, contributing to a safer environment.

Our observation of the negative correlation between higher education levels and crime rates is echoed in the study by Dr Muhammad Ramzan Sheikh et al., which identifies education level as a key factor influencing crime rates in Multan District Prison, Pakistan. This finding strengthens our argument about the critical role of education in shaping crime dynamics (Sheikh et al. 2022).

3. Bachelor's Degrees or Above and Shooting Cases:

People who gets the Bachelor's degrees or above would significantly contribute to the decrease of shooting case in a neighbourhood, which could be summarized from the correlation matrix. The negative correlation between higher education and shootings suggests that areas with a larger proportion of residents holding at least a Bachelor's degree may enjoy more robust community resources and higher levels of civic involvement, which can lead to better crime prevention and intervention efforts.

3.3.3 Analysis 2: Crime Rate vs. Major Field of Study

Next, we compare various crime rates versus some popular field of study:

Sample of processed data:

	HOOD_ID	ASSAULT_RATE_2016	AUTOTHEFT_RATE_2016
23	140	444.8488	40.44080
24	139	1127.8418	53.14437
25	138	894.7277	96.01955
26	136	1520.6111	54.05016
28	134	553.3043	62.34414
29	133	350.1605	36.47505

	52. Business, management, marketing and related support services	11. Computer and information sciences and support services	27. Mathematics and statistics
23	965	175	40
24	1355	285	100
25	1820	485	65
26	2125	415	85
28	1425	325	90
29	1610	300	90

3.3.3.1 Data Visualization

A correlation matrix is calculated to quickly determine whether or not some specific crimes rate have correlation to the major of study that each individual in the neighbourhood focus using cor() function (R Core Team 2023). Then, we use ggplot2 to visualize the correlation matrix (Wickham et al. 2023). The data visualization can be found as Figure 2.

3.3.3.2 Implication

From the Figure 2, we observe some interesting data points from the correlation matrix.

1. Mechanic and Precision Production and Auto Theft Rate:

The correlation between auto theft and the field of Mechanic and Precision Production could be partially due to the practical skills gained in this major, which might inadvertently provide the technical knowledge necessary for committing auto thefts. It's important to clarify that this does not imply that individuals in these fields are committing crimes, but rather that the technical expertise prevalent in these areas could raise the potential for such activities. The negative correlation with bike theft might suggest that the focus in these areas is more on motorized vehicles than on bikes, which could influence the types of theft that are more common.

2. Break Enter Crime and Communication and Journalism Programs:

The positive correlation with break-and-enter crimes and Communication and Journalism majors may suggest that these areas, often associated with active, bustling urban campuses, may also be more attractive targets for such crimes due to higher concentrations of valuable electronics and equipment used in communication and media studies. Students and facilities may possess more items that are easily resold, such as cameras, audio equipment, and computers, potentially making them more likely targets for thefts.

3.3.4 Analysis 3: Crime Rate vs. Household After-Tax Income

Lastly, we want to see if the household after-tax income would have positive correlation to the Crime rate. Therefore, we compare various crime rates versus various income levels:

Sample of processed data:

	HOOD_ID	ASSAULT_RATE_2015	AUTOTHEFT_RATE_2015
24	139	897.3905	88.55827
25	138	775.0925	91.44350
26	136	1429.3468	65.13480
27	135	783.6725	95.84484
28	134	392.9727	84.75882
29	133	277.8387	51.18081

	Under $5,000	$5,000 to $9,999	$10,000 to $14,999
24	110	175	310
25	100	125	265
26	150	390	550
27	145	130	225
28	65	30	45
29	20	20	30

3.3.4.1 Data Visualization

We used a correlation matrix to quickly determine whether or not some specific crimes rate have correlation to various income levels. We used ggplot2 to visualize the correlation matrix as a heat map, which can be found as Figure 3.

3.3.4.2 Implication

From the Figure 3, we made some interesting observations as follow:

1. Lower Crime Correlation with Higher Household Income:

As household income increases, the resources available for crime prevention and community investment also typically rise. Higher-income areas might afford better home security, neighborhood watch programs, and more effective law enforcement services. There is also the sociological perspective that higher income correlates with lower stress levels regarding basic needs, which could reduce the incentive for crimes of necessity.

The study by Nur Farah Zafirah Zulkiflee et al. in Malaysia also supports our findings, suggesting that higher education levels and economic growth have a significant long-term positive impact on reducing crime rates. This parallel finding across different geographical contexts underscores the universal relevance of socio-economic factors in crime prevention (Zulkiflee, Borhan, and Hadrawi 2022).

2. Strong Correlation of Various Crimes with Low-Income Households:

Households with an annual income of less than $15,000 might experience various social challenges, including limited access to education and employment opportunities, which can be linked to higher crime rates. Low-income areas might have less community infrastructure to deter crime, such as fewer police patrols or underfunded social services, leading to a stronger correlation with various crimes. This demographic might also be less likely to have home security systems, increasing vulnerability to property crimes.

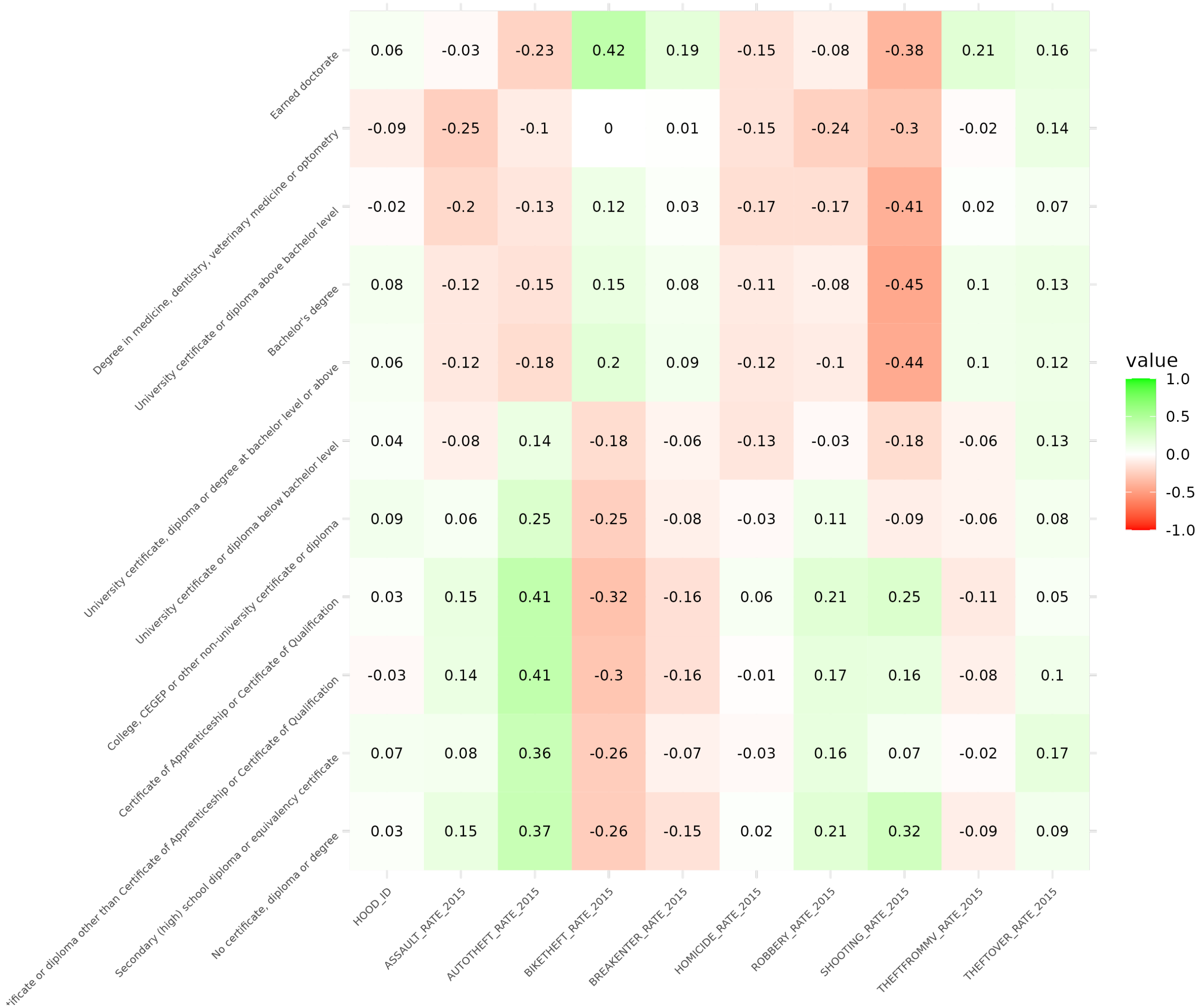


Figure 1: Correlation Matrix of Education Levels vs. Crime Rate

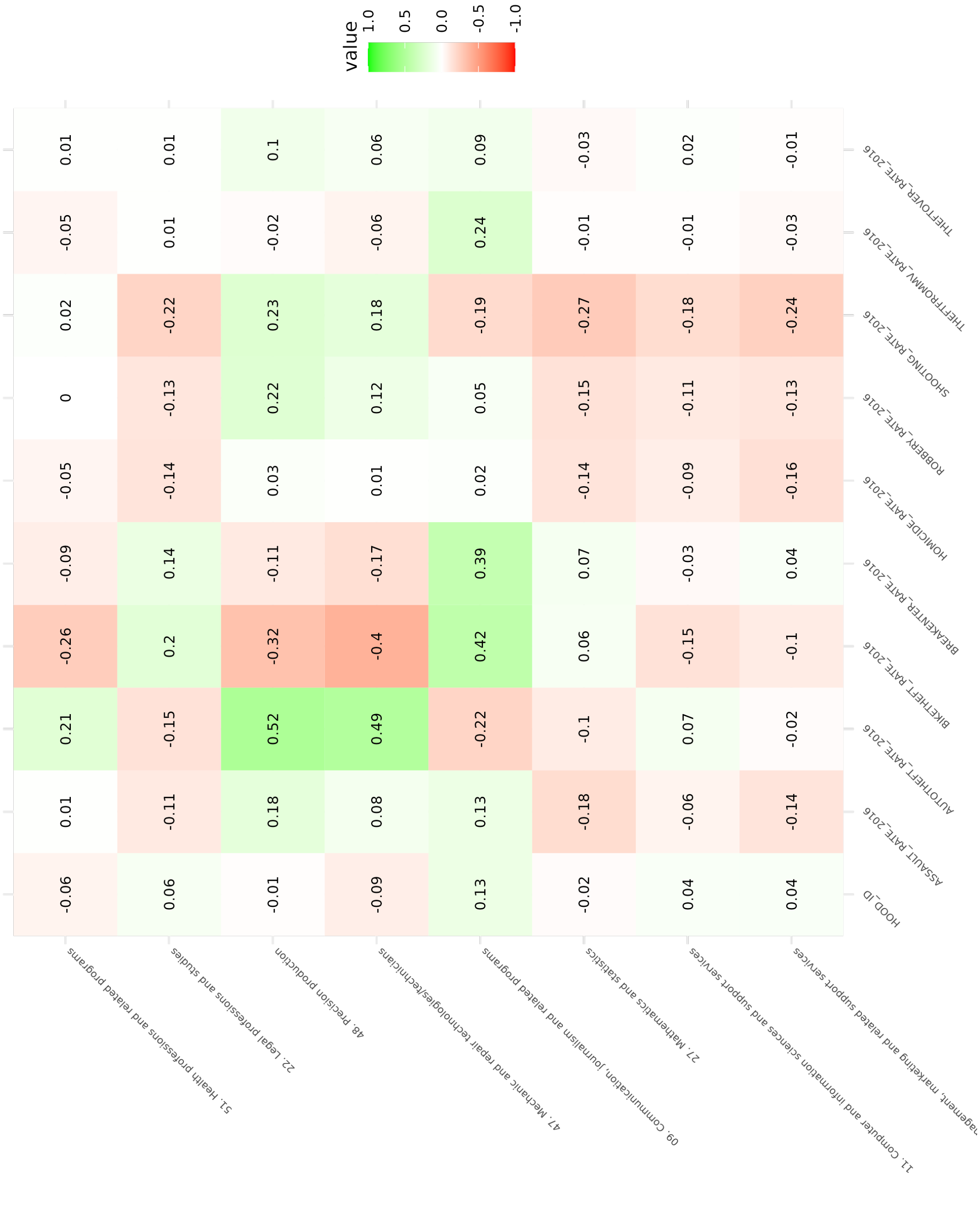


Figure 2: Correlation Matrix of Major of Study vs. Crime Rate

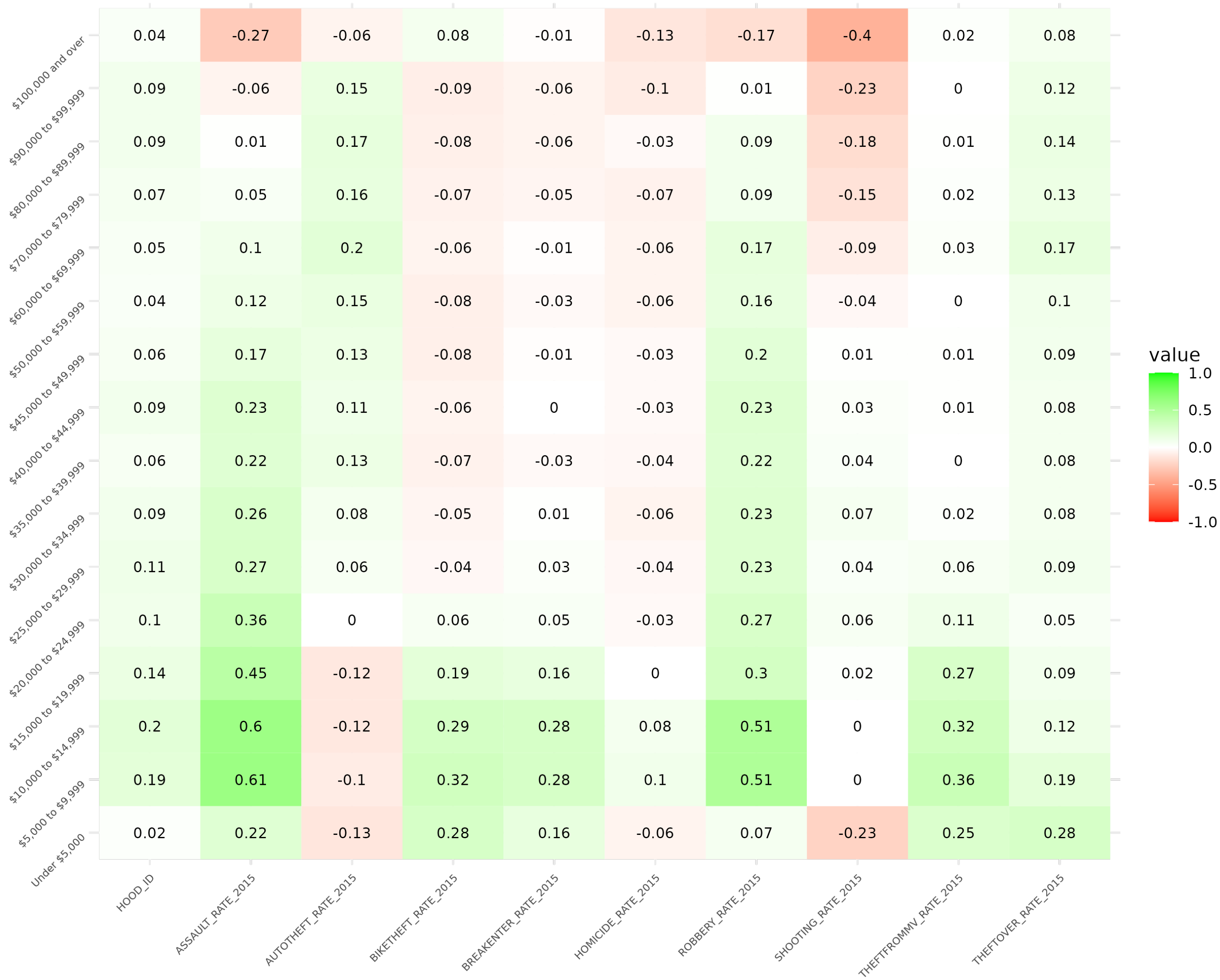


Figure 3: Correlation Matrix of Household After-Tax Income Levels vs. Crime Rate

4. Limitations

While this study offers significant insights into the interplay between socio-economic factors and crime rates at the neighborhood level, it is important to acknowledge certain limitations that may influence the interpretation and generalization of the findings.

Firstly, the study's reliance on secondary data sources, particularly the "Neighborhood Crime Rates" and "Neighbourhood Profiles" from the City of Toronto, constrains the analysis to the parameters and accuracy of these datasets. The datasets' design and data collection methods, which are not under the control of this study, could inherently limit the depth and scope of the analysis. For instance, the accuracy of crime reporting and the comprehensiveness of socio-economic data could vary, potentially affecting the reliability of the correlations observed.

Secondly, the cross-sectional nature of the study, focusing on data primarily from the year 2015 and 2016, provides a snapshot view, limiting the ability to infer long-term trends or causal relationships. Socio-economic dynamics and crime patterns are subject to change over time, and the findings from a single year may not fully encapsulate these evolving trends.

Thirdly, while the study identifies correlations between socio-economic factors and crime rates, it is crucial to recognize that correlation does not imply causation. The observed relationships could be influenced by other unmeasured variables or underlying factors not captured in the available data. For instance, factors such as cultural dynamics, community engagement, law enforcement practices, and urban development policies could also play significant roles in influencing crime rates but were not directly accounted for in this analysis.

Finally, the study's focus on a single city, Toronto, while providing a detailed local context, may limit the generalizability of the findings to other urban settings with different socio-economic landscapes and crime dynamics. Urban environments vary widely in terms of demographics, economic conditions, cultural factors, and law enforcement policies, all of which can impact the relationship between socio-economic factors and crime.

In conclusion, while the study offers valuable insights into the relationship between education, income, and crime within the context of Toronto neighborhoods, these limitations highlight the need for cautious interpretation of the findings and suggest avenues for further research to deepen and broaden our understanding of these complex socio-economic interactions.

5. Conclusion

The study undertaken to analyze the relationship between socio-economic factors and crime rates in Toronto neighborhoods has yielded several profound insights, particularly concerning educational attainment and household income levels.

Firstly, the investigation into the correlation between educational levels and crime types has unveiled significant findings. We found out that higher educational achievement within a neighborhood generally correlates with lower crime rates. This trend underscores the pivotal role of education in fostering societal stability and reducing the propensity for criminal behavior. Moreover, specific correlations were observed between the levels of education, such as doctorates, and particular crime types like bike thefts. These nuances in the data suggest that the educational composition of a neighborhood can influence the nature and frequency of crime occurrences.

Secondly, the study delved into how the major field of study impacts neighborhood crime rates. Intriguing correlations were noted, such as the association between fields like Mechanic and Precision Production and certain types of crime, including auto theft. These correlations point towards the complex socio-economic fabrics that characterize urban neighborhoods, where the prevalent skills and knowledge bases can subtly influence crime patterns.

Thirdly, the research highlighted a notable negative correlation between higher income brackets and shooting crime rates. This finding is particularly significant, as it suggests that economic well-being within a neighborhood can act as a deterrent to certain violent crimes. The inverse relationship between income levels and shooting incidents underscores the importance of economic stability and prosperity as a cornerstone in the foundation of safe and secure communities.

Lastly, the analysis brought to light a compelling connection between lower household incomes and higher crime rates. This correlation is particularly stark, suggesting that economic hardships and limited financial resources are key drivers of criminal activity. This insight is crucial for understanding the broader socio-economic underpinnings of crime in urban settings and highlights the need for targeted economic interventions to address the root causes of crime in lower-income neighborhoods.

In conclusion, this study offers a comprehensive and detailed examination of the role of socio-economic factors, particularly education and income, in shaping neighborhood-level crime dynamics, aligning with international research like P. Boxer et al.'s work on the connection between neighborhood crime and academic performance (Boxer, Drawve, and Caplan 2020). These findings not only deepen our understanding of urban crime but also provide valuable guidance for policymakers in crafting informed and effective strategies for crime prevention and community development, underscoring the importance of socio-economic stability in fostering safer communities.

6. Furthur steps

Here we proposed two further steps following this paper:

6.1 Longitudinal Study to Trace Trends Over Time

A pivotal next step in this area of research involves conducting a longitudinal study that tracks socio-economic factors and crime rates over an extended period. Such an approach would allow for the observation of trends, shifts, and causal relationships that are not discernible in cross-sectional analyses. By examining data across multiple years, researchers can better understand how changes in economic conditions, educational attainment, and other socio-economic variables impact crime rates over time. This longitudinal perspective is essential to capture the dynamism of urban environments and to provide a more robust and nuanced understanding of the evolving relationship between socio-economic factors and crime.

6.2 Comparative Urban Analysis to Enhance Generalizability

Another crucial step forward would be to undertake a comparative analysis across different urban settings. By analyzing and contrasting cities with varying socio-economic landscapes, cultural backgrounds, and policy frameworks, researchers can gain a more comprehensive understanding of the interplay between socio-economic factors and crime. This comparative approach would not only enhance the generalizability of the findings but also illuminate the specific conditions under which certain socio-economic factors become more influential in shaping crime dynamics. Such a study would significantly contribute to the development of tailored and effective urban policies and crime prevention strategies, catering to the unique needs and challenges of diverse urban communities.

7. Appendix:

7.1 Data Cleaning:

In creating a comprehensive analysis of the relationship between socio-economic factors and neighborhood crime rates, we've undertaken a meticulous data cleaning and processing procedure. This process ensures that our data sets are compatible, relevant, and free of inaccuracies, allowing for a more accurate and insightful analysis. Here's an overview of the steps involved:

7.2 Transposing the 'Neighbourhood Profiles' Dataset:

The original structure of the 'Neighborhood Profiles' dataset presented each column as a neighborhood, with rows containing corresponding socio-economic data (such as education level, income, etc.). To facilitate analysis, we transposed this dataset. Now, each row represents an individual neighborhood, and each column corresponds to a specific piece of socio-economic information. This reorientation makes it easier to compare and analyze data at the neighborhood level.

7.3 Aligning Temporal Data Across Datasets:

The 'Neighborhood Crime Rates' data set, provided by the Toronto Police Services, includes crime rates for each neighborhood from 2015 to 2020. However, our socio-economic data from the 'Neighborhood Profiles' data set is specific to the year 2015 and 2016. To ensure consistency and comparability, we extracted only the 2015 and 2016 crime data from the 'Neighborhood Crime Rates' data set. This alignment of the time frame across both data sets is crucial for accurate comparative analysis.

7.4 Merging Datasets Using Neighborhood ID (HOOD_ID):

To integrate the two data sets, we performed a left-join operation using the Neighborhood ID (HOOD_ID) as the key. This technique ensures that each neighborhood's socio-economic data from the 'Neighborhood Profiles' is accurately matched with its corresponding crime data from the 'Neighborhood Crime Rates'. This merged dataset now offers a comprehensive view of each neighborhood, encompassing both its socio-economic characteristics and crime statistics.

7.5 Handling Missing Data:

In the process of merging and analyzing the datasets, we encountered instances of missing data (NA fields). To maintain the integrity and reliability of our analysis, we chose to omit these NA fields. This decision ensures that our findings and conclusions are based solely on complete and accurate data, thereby enhancing the validity of our study.

8. References

Boxer, P., G. Drawve, and Joel M. Caplan. 2020. "Neighborhood Violent Crime and Academic Performance: A Geospatial Analysis." *American Journal of Community Psychology*. https://doi.org/10.1002/ajcp.12417.

Cai, Lei. 2021. "Analysis of Hate Crime Rates in the United States: Statistical Modeling of Public Safety Issues Based on Socioeconomic Factors." *IEEE International Conference on e-Education, e-Business and e-Commerce*. https://doi.org/10.1109/ICECEM54757.2021.00082.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sheikh, Dr Muhammad Ramzan, Rashid Ahmad, F. Bashir, and S. Sultan. 2022. "Identifying Crime and Socio-Economic Factors Related to Crime in Multan District Prison, Pakistan." *Pakistan Journal of Humanities and Social Sciences*. https://doi.org/10.52131/pjhss.2022.1002.0226.

Toronto Open Data. 2023. *Neighbourhood Profiles*. Toronto, Ontario, Canada: Toronto Open Data. https://open.toronto.ca/dataset/neighbourhood-profiles/.

———. 2024. *Neighbourhood Crime Rates*. Toronto, Ontario, Canada: Toronto Open Data. https://open.toronto.ca/dataset/neighbourhood-crime-rates/.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2023. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://ggplot2.tidyverse.org.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zulkiflee, Nur Farah Zafirah, Nurbaizura Borhan, and Mohd Fikri Hadrawi. 2022. "Factors Affecting Crime Rate in Malaysia Using Autoregressive Distributed Lag Modeling Approach." *Pertanika Journal of Social Sciences & Humanities*. https://doi.org/10.47836/pjssh.30.3.13.