

# Investigating the Socioeconomic Dynamics of Crime: Exploring the Relationship Between Income and Crime Rates in Canada

Yinuo Zhao

<sup>1</sup> Department of Computer Science & Statistics,  
Faculty of Arts and Science, University of Toronto

April 2024

**Abstract:** This project explores the correlation between crime rates and income in Canada, utilizing integrated datasets from Statistics Canada. It finds a significant negative correlation between average income and total crime rates across provinces, with regional nuances. The analysis uncovers intriguing associations, including an inverse relationship between robbery rates and employment income, and varied impacts of income on prostitution rates. Statistical modeling confirms these findings, emphasizing the complex interplay between income, geography, and crime rates. These insights are vital for informing policy interventions and social welfare programs to address crime and socioeconomic disparities in Canadian society. **Keywords:** *Crime rates, Income levels, Socioeconomics*

## 1 Introduction

### 1.1 Topic Background

The intersection of crime rates and income levels is a complex issue with far-reaching implications. Numerous factors intertwine to shape this relationship, revealing compelling insights into societal dynamics[1]. In many cases, areas with lower income levels exhibit higher crime rates, a phenomenon rooted in various socioeconomic realities. Limited economic opportunities prevalent in these regions often propel individuals towards illicit activities as a means of survival[2]. Moreover, disparities in education and access to essential resources further compound this correlation. Wealthier neighborhoods typically enjoy better-funded educational institutions and greater access to vital services, acting as mitigating factors against criminality.

However, understanding this relationship transcends simplistic associations. It demands a nuanced approach, one that acknowledges the multifaceted nature of the issue. Unraveling the intricate connections between income levels and crime rates necessitates comprehensive statistical modeling and reasoned analysis.

Exploring the correlation between crime rates and income in Canada holds paramount importance for policymakers, law enforcement agencies, and social welfare programs [3]. Delving into this relationship illuminates the socioeconomic drivers of criminal behavior, offering invaluable insights for crafting targeted interventions aimed at poverty alleviation and crime reduction. Furthermore, elucidating this correlation contributes to broader societal discussions on social inequality, justice, and community well-being within Canadian society.

### 1.2 Data Sources

Addressing this inquiry hinges upon access to pertinent data on crime and socioeconomic indicators. Due to the absence of a single comprehensive dataset, a meticulous approach involving the merging of multiple datasets becomes imperative. Statistics Canada serves as the primary source for such data, offering a diverse array of information encapsulating various facets of the nation's status. Further details regarding the acquisition and compilation of these datasets are expounded upon in Section 2.1: Obtaining Data.

## 2 Methods

### 2.1 Data Acquisition

Two primary datasets were sourced from Statistics Canada's official website, both freely accessible:

1. *Income of individuals by age group, sex, and income source, Canada, provinces, and selected census metropolitan areas* (Table 11-10-0239-01)[4], last updated on May 2, 2023. This dataset is regularly updated and collected through the Survey of Labor and Income Dynamics, Survey of Consumer Finances, and Canadian Income Survey.
2. *Incident-based crime statistics, by detailed violations, Canada, provinces, territories, Census Metropolitan Areas, and Canadian Forces Military Police* (Table 35-10-0177-01, formerly CAN-SIM 252-0051)[5], last updated on July 27, 2023. Similarly, this dataset is annually updated and obtained through the Uniform Crime Reporting Survey.

### 2.2 Data Preparation

As both datasets originate from the same source, they share a consistent structure with variables denoting geographical regions and the year of data collection. Consequently, merging was facilitated using these shared variables. However, due to the extensive nature of the datasets containing extraneous information, preprocessing steps including cleaning and wrangling were necessary to ensure ease of analysis and computational efficiency. Notably, the datasets were transformed from long to wide formats for convenience. Moreover, meticulous examination of dataset dimensions, variable types, and values was undertaken, with any suspicious observations cross-validated against external sources to ensure data reliability.

### 2.3 Exploratory Data Analysis and Modeling

EDA was conducted individually on each dataset before merging, followed by exploration of the combined dataset. This involved analyzing total crime rates and specific crime rates across various provinces over time, alongside assessments of average income per province per year. Correlation analyses between crime rates and income levels were performed, with notable features warranting further investigation. Zooming into specific geographical regions, analyses were refined to

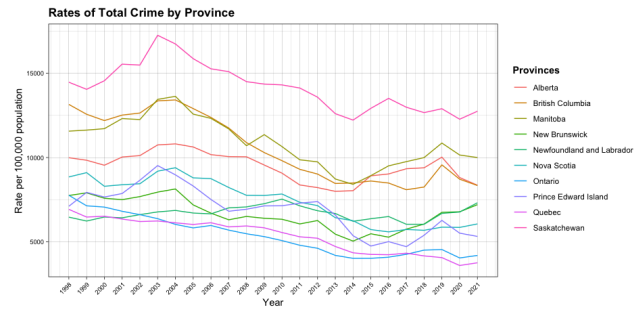


Figure 1: Total Crime Rate by Province and Year

examine the relationship between crime rates and specific income categories. Additionally, observations were grouped into four income levels for concise interpretation. To quantify the relationship, four distinct models—linear regression, random forest, generalized additive model, and XGBoost—were employed. Model outcomes were compared and contrasted in Section 3: Results.

## 3 Results

### 3.1 Separate Investigation

Across all provinces, there is a discernible decreasing trend in crime rates over the years, with many provinces experiencing peak crime rates in 2003-2004. Figure 1 depicts the total crime rate, encompassing all types of crimes, from 1998 to 2021, with each province represented by a distinct color. The crime rate is measured per 100,000 population. Notably, Saskatchewan consistently exhibits a significantly higher crime rate compared to other provinces throughout the period of 1998 to 2021. Conversely, Quebec and Ontario consistently demonstrate the lowest crime rates.

In addition to analyzing total crime, individual crime categories often associated with poverty is investigated, namely break and enter, robbery, and prostitution. Overall, there is a decreasing trend in the rates of all three (Figure 2), with occasional exceptions such as robbery rates in Manitoba. British Columbia stands out with a significantly high rate of prostitution in 2004, doubling the number reported in Saskatchewan, which held the second-highest rate that year.

The average total income for all provinces exhibits a steady upward trend as in Figure 3. Since 2003, Alberta has surpassed Ontario to become the province with the highest average total income. Additionally, there is slight decrease in income across all provinces around 2019, likely attributed to the impact of the

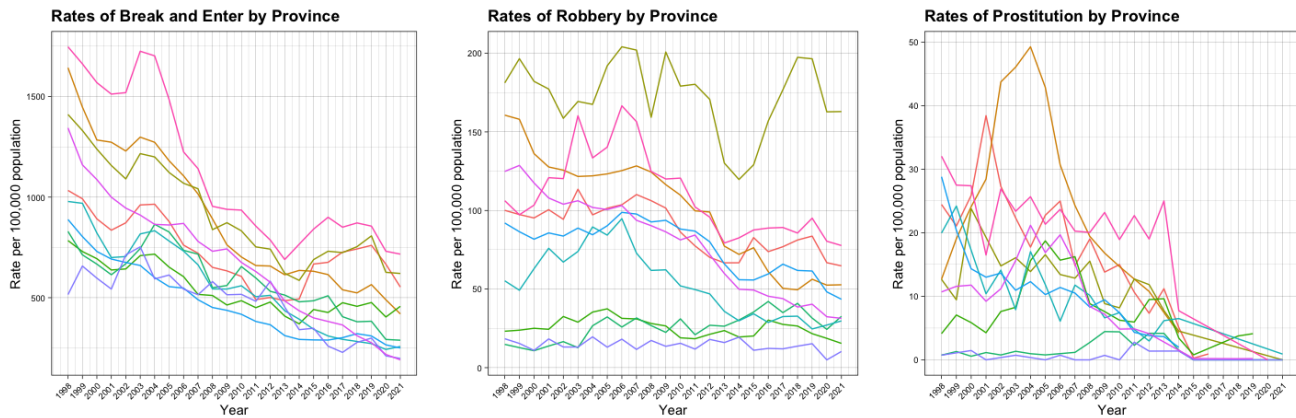


Figure 2: Break and Enter, Robbery, Prostitution

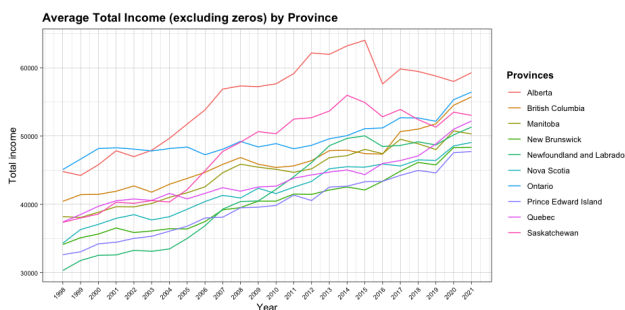


Figure 3: Average Total Income by Province

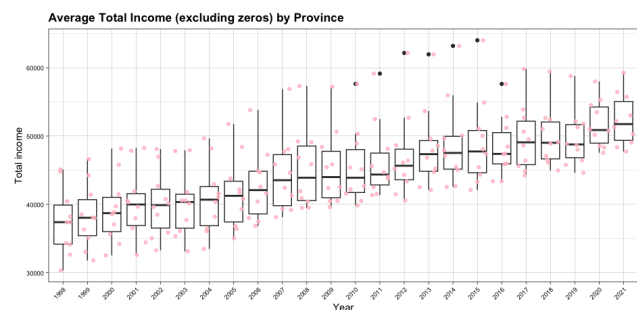


Figure 4: Total Income by Year

COVID-19 pandemic. Figure 4 shows the average total income by year, combining data of all provinces. The pink dots represent the actual values for each year in each province. It's clear from the boxes that the average income is increasing year by year. Notably, between 2010 and 2016, there are some outliers with very high incomes, which corresponds to Alberta when compared with the previous plot.

Specific income sources: employment income, investment income, and market income, which are major income categories are taken closer look as in Figure 5. All three types of income are increasing. However, employment and market income show a more steady growth pattern, while investment income fluctuates dramatically from year to year.

### 3.2 Combining Crime and Income

Average total income and total crime rate are negatively correlated across all provinces (Figure 6), with possibly the exception of Newfoundland and Labrador where the relationship appears to be less pronounced, indicated by a line that is nearly horizontal. This graphical observation is verified by the negative numerical

Quebec	-0.9447648
Nova Scotia	0.9236636
British Columbia	-0.8154907
Prince Edward Island	-0.8082244
Saskatchewan	-0.7695228
Ontario	-0.7583329
Manitoba	-0.7392303
Alberta	-0.7023852
New Brunswick	-0.5966958
Newfoundland and Labrador	-0.0025743

Table 1: Correlation: Income and Crime Rate

correlations in Table 1. Quebec exhibits the strongest correlation between average total income and total crime rate as the correlation is close to -1.

The relationship between robbery crime rate and employment income is negative for all provinces except Newfoundland and Labrador. Manitoba demonstrates a weak relationship, as evidenced by the considerable dispersion of points around the line. Property crime rate shows strong and negative relationship with market income. Employment income and prostitution crime rate has varying degrees of association across

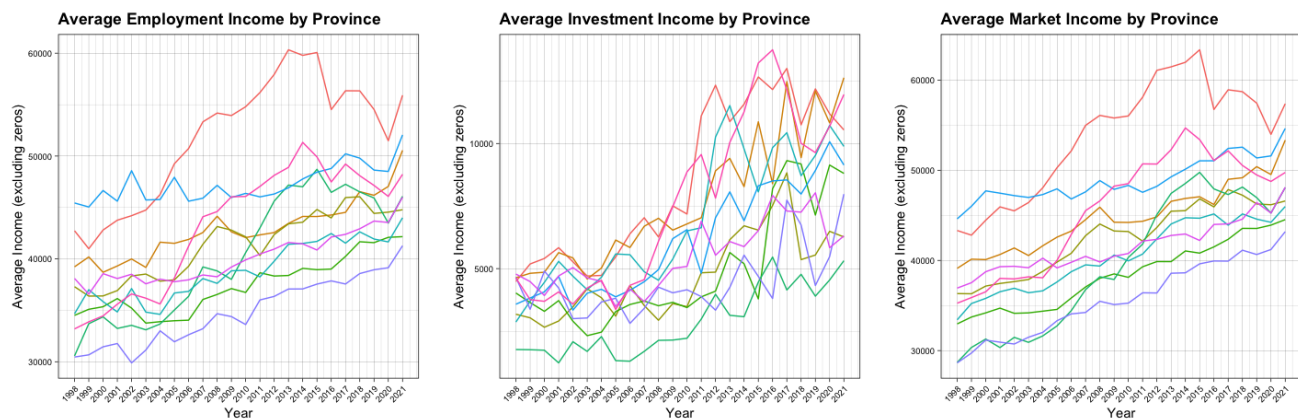


Figure 5: Employment Income, Investment Income, Market Income

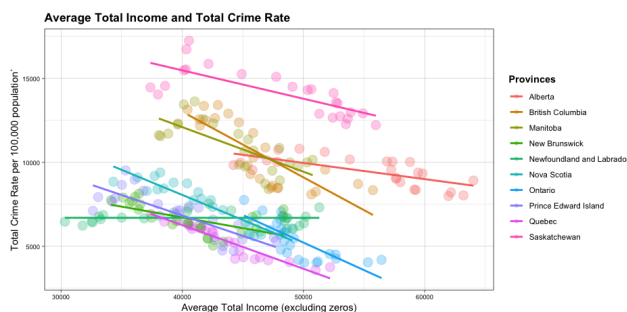


Figure 6: Average Total Income and Total Crime Rate

provinces, with many showing a weak relationship. Notably, Ontario exhibits a positive relationship between employment income and prostitution crime rate. Figure 7 provides graphical representation of the findings.

After grouping the observation by income to 4 levels by quartile: Low, Med-Low, Med-High, High, it appears that the total crime rate does not exhibit a clear trend of decreasing with higher levels of total income (Figure 8). This finding contradicts previous observations when examining the relationship between total crime rate and average total income by province. Therefore, it is conceivable that while a relationship exists, it may be influenced by other factors related to the demographics of each province. Consequently, when considering all observations collectively, the relationship becomes less apparent.

### 3.3 Case Study: Ontario

The major categories in Ontario are property crime and weapon violations. The heat map (Figure 9) summarizes the correlation between all types of violations and income sources in Ontario. COVID-19 benefits are ex-

cluded from the analysis due to their availability only during the pandemic years.

The strongest correlation is observed between self-employment income and production under the Cannabis Act. However, this relationship may not be entirely reliable due to the limited data available (3 observations), as in Figure 10. The second strongest correlation, with a coefficient of 0.98, is observed between total income and incidents of possession of other Controlled Drugs and Substances Act drugs. As income levels increase, the number of incidents of possession of these drugs also tends to rise. Despite the negative correlation between total income and total crime rate in Ontario, as found in previous analyses, a positive relationship exists between total income and incidents of possession of other Controlled Drugs and Substances Act drugs (Figure 11).

### 3.4 Statistical Modelling

All models use average income and interaction with regions as predictors, a summary of model outcome is in Table 2.

1. **Linear Model:** Negative relationship exists between average total income and total crime rate. Various provinces exhibit differing baseline rates, with British Columbia displaying significantly higher rate compared to the reference province. Interaction terms between income and provinces indicate varying effects across regions.
2. **GAM:** Predicts crime rate with a smooth term of Year. Saskatchewan exhibits the strongest effect on crime rates among the provinces. There is a negative relationship between average income and crime rates.



Figure 7: Specific Income and Crime Types

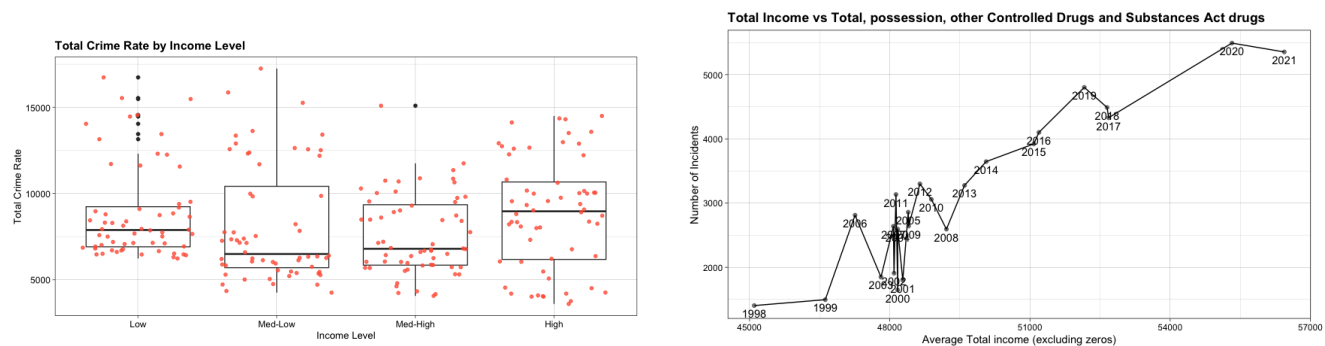


Figure 8: Total Crime Rate By Income Level

Figure 11: Total Income and Controlled Drug Act

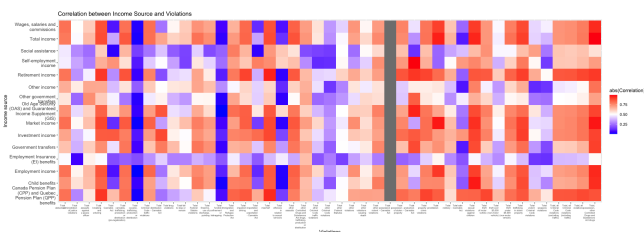


Figure 9: Correlation Heat Map

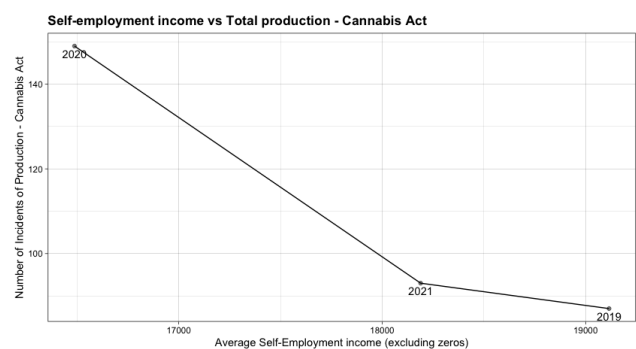


Figure 10: Self Employment Income and Cannabis Act

Model	Linear	GAM	RF	XGB
$R^2$	0.94	0.95	0.90	0.98

Table 2: Model Comparison

- RF:** Explains approximately 90.24% of the variance in crime rate, which suggests that the predictor variables collectively have a strong relationship with the crime rate. The mean of squared residuals is relatively high (859479.8), indicating that there is still some unexplained variability in the data that the model is not capturing well. This suggests that there may be other factors influencing the crime rate that are not included in the model.
- XGB:** The model has shown a RMSE of approximately 420.15, indicating a reduction in prediction error compared to the previous model. The coefficient of determination (R-squared) has also increased to approximately 0.9799, suggesting that the model explains almost 98% of the variance in the data.

## 4 Conclusion and Summary

### 4.1 Summary of Findings

The negative correlation between income and crime rates has the intricate interplay of socioeconomic factors and regional dynamics. Despite fluctuations, there was a general downward trend in total crime rates. Concurrently, average total income across provinces showed an upward trajectory. A negative association emerged between crime and income across all provinces, indicating that higher average income tended to coincide with lower crime rates proven by multiple statistical models. However, the relationship between specific crime types and income varied.

Further analysis at the provincial level revealed nuanced correlations between income and specific crime types as they differ by province. Despite these findings, group analysis based on income levels revealed complexities, suggesting that while a relationship between income and crime exists, it may be influenced by other demographic factors, challenging the notion of a straightforward causal relationship.

### 4.2 Limitations and Future Works

A primary limitation lies in the lack of inflation adjustment for income values, potentially confounding the observed relationship between income and crime rates. Without this adjustment, the findings may incorporate the effects of nominal income growth, which could exaggerate or diminish the true relationship. Future studies should prioritize incorporating inflation adjustments to provide a more accurate understanding of the relationship between income and crime rates over time.

Moreover, while this study focused on income as a primary socioeconomic factor, it's essential to recognize the influence of other economic variables as potential con-founders. Economic recessions, for instance, can significantly impact both income levels and crime rates, complicating the interpretation of their relationship. Future research endeavors should explore the multifaceted interplay between various economic factors and crime rates to elucidate the nuanced dynamics at play.

Additionally, extending the analysis to consider regional differences in economic policies, cultural norms, and demographic compositions could provide deeper insights into the observed relationships. By accounting for these contextual factors, researchers can refine their understanding of the socioeconomic determinants

of crime and develop more targeted interventions to address underlying drivers effectively.

## 5 Acknowledgements

This project is created by Yinuo Zhao as part of the course JSC370H1 offered by the University of Toronto in Winter 2024, instructed under Jun Ni (Jenny) Du and professor Meredith Franklin.

## References

- [1] M. T. Henderson, J. Wolfers, and E. Zitzewitz, "Predicting crime," *Ariz. L. Rev.*, vol. 52, p. 15, 2010.
- [2] M. Hooghe, B. Vanhoutte, W. Hardyns, and T. Bircan, "Unemployment, Inequality, Poverty and Crime: Spatial Distribution Patterns of Criminal Acts in Belgium, 2001–06," *The British Journal of Criminology*, vol. 51, no. 1, pp. 1–20, 12 2010. [Online]. Available: <https://doi.org/10.1093/bjc/azq067>
- [3] W. L. Perry, B. McInnis, C. C. Price, S. Smith, and J. S. Hollywood, *Predictive Policing: Forecasting Crime for Law Enforcement*. Santa Monica, CA: RAND Corporation, 2013.
- [4] Statistics Canada, "Income of individuals by age group, sex and income source, canada, provinces and selected census metropolitan areas," *Statistics Canada*, Year. [Online]. Available: <https://doi.org/10.25318/1110023901-eng>
- [5] —, "Incident-based crime statistics, by detailed violations, canada, provinces, territories, census metropolitan areas and canadian forces military police," *Statistics Canada*, Year. [Online]. Available: <https://doi.org/10.25318/3510017701-eng>

## A Artifact Appendix

The code base for running the analysis in this report is available at the GitHub repository [https://github.com/inorrr/income\\_and\\_crime](https://github.com/inorrr/income_and_crime).

Source code of analysis is available in `work.rmd`. Please also find the project website with interactive visualizations at [https://inorrr.github.io/income\\_and\\_crime](https://inorrr.github.io/income_and_crime).