Crime Dynamics: How Socioeconomic Factors Influence Mischief and Break-and-Enter Crimes

Dalal Hrishikesh, Vashisht Raghav and Sanghera Srijan

Introduction

Crime has a lasting impact on our communities, creating a sense of insecurity and fear. Victims frequently suffer from chronic trauma, and even those who are not directly targeted may suffer from increased anxiety and a lower quality of life. This widespread uneasiness can undermine social cohesiveness and trust. Crime rates in Vancouver have continued to decline year over year (Griffiths, 2024). It is essential to understand the root causes of crime and what factors might affect it. The analysis of these factors can help us develop strategic plans that focus on suggesting policies or ideas for creating safer urban areas. Furthermore, the effects of important and basic demographics can help us identify early warning signs and risk factors, allowing for proactive intervention before crimes occur.

As crime rates in Vancouver fall and meet their pre-pandemic levels, the question remains: How do demographic and socioeconomic factors, such as average age, income inequality, and education, affect Mischief and Break-and-Enter Residential/Other (BER) crimes?

Our paper particularly focuses on Mischief and Break & Enter Residential/Other (BER) crimes due to the high degree of impact on neighbourhoods. The economic toll of these types of crime is substantial. Direct costs, such as property damage, theft, and the expenses associated with law enforcement and the justice system, place a huge burden on society. Additionally, since the VPD shows crime statistics by when they're reported and not when they happen, this adds an additional caveat of unreported crimes. We aim to minimise this effect by using Mischief and BER since they are among the most reported crimes in the community.

Background

The connection between crime rates and income inequality has been the subject of a substantial amount of research. Higher income inequality, as indicated by the Gini coefficient, is positively correlated with crime rates, especially property crimes (Fajnzylber et al., 2002).

Important insights can also be gained from studies on crime rates and educational attainment, where an extensive inverse relationship has been noted between years of education and crimes like vandalism, theft and assault (Groot & van den Brink, 2010). According to Machin & Meghir (2004), neighbourhoods with lower average levels of education also tend to have higher crime rates. In the context of crime, age has also been extensively researched. According to Hansen (2003), the fact that there is an increase in criminal activity from teens to the ages of middle 20 has been widely accepted.

Data Description

The data used in this paper is primarily from 2 sources – the 2021 Canadian Census and VPD Crime Statistics of 2023. The Vancouver Police Department's Crime Data provides detailed records of crimes occurring within the city of Vancouver, with geolocated information down to the neighbourhood level. The data includes various types of crimes, categorised into 11 distinct classifications: Break and Enter Residential/Other, Break and Enter Commercial, Homicide, Mischief, Offence Against a Person, Theft from Vehicle, Theft of Bicycle, Theft of Vehicle, Other Theft, Vehicle Collision or Pedestrian Struck with Fatality, and Vehicle Collision or Pedestrian Struck with Injury. Each crime entry includes the type of crime, its occurrence date, and the neighbourhood's longitudinal and latitudinal coordinates.

The 2021 Canadian Census Data offers socioeconomic information at the neighbourhood level, specifically at the dissemination area level. The data includes key socioeconomic variables such as average age, income inequality measured by the Gini coefficient, and educational attainment levels. These variables provide insights into the demographic and economic characteristics of each neighbourhood.

This paper examines the relationship between neighbourhood-level socioeconomic characteristics and crime rates. The Gini coefficient is used to determine whether a correlation exists between increased income inequality and crime rates, specifically for crimes like Break-and-Enter Residential/Other and mischief. The relationship between crime frequency and lower educational attainment is investigated to determine if neighbourhoods with lower average levels of education have higher crime rates. The impact of average age on crime rates is studied to understand if specific age profiles are associated with higher or lower crime incidence. Furthermore, Population Density is used as a control variable since higher populations may lead to more incidences of crime, but it might not necessarily explain why crimes are committed in the first place.

Our research found that income inequality exerted the highest influence on the occurrence of both crimes. Education level was a much bigger factor in Mischief crimes and had no impact in BER crimes. Looking into the effect of Average age, we found a non-linear relationship for both crimes, which was noteworthy. To analyse this non-linear relationship of average age we conducted a robustness analysis and found different peak ages for both crime types. This indicates that different factors affect both Mischief and BER crimes due to the difference in their inherent nature.

Summary Statistics

A data.frame: 6 × 5

variable	mean	sd	max	min
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Average Age	43.0246292	5.559827e+00	86.000	21.000
Gini Index	0.3308630	7.996311e-02	0.752	0.194
Education Level	626.3098907	4.484760e+02	7855.000	40.000
Break and Enter Crime Residential/Other	0.3091922	1.026915e+00	15.000	0.000
Mischief Crime	1.8484680	1.088369e+01	352.000	0.000
Population Density	6457.5899578	7.609169e+03	76474.359	0.000

Figure 1. Table of Summary Statistics for variables used in models

Model Overview

The primary regression model examines socioeconomic characteristics like Average Age, Gini Index, and Education Level to explain Vancouver neighborhood Mischief and Break and Enter crimes. This model explains how demographics and economics affect urban crime. The error term captures unobservable crime rate components.

The analysis assumes no multicollinearity, as VIFs below 2 show. Additionally, neighboring observations are assumed to be independent. Although the Breusch-Pagan test shows heteroskedasticity (p = 1.205e-05), robust standard errors are used to reduce its impact on standard error estimations.

Mischief Crimes:

$$(1)\ (Mischief)i = \beta_0 + \beta_1(Gini\ Index)_i + \beta_2(Education\ Level)_i + \beta_3(Average\ Age)_i + \epsilon_i$$

The baseline model (1) examines how Average Age, Gini Index, and Education Level affect Mischief Crime prevalence. Average Age has a statistically significant negative coefficient, showing that older communities have less mischief crimes. The Gini Index is positive and highly significant, showing that income disparity increases mischief offenses. Education Level is positively associated and significant, although with a lesser impact size. These findings lay the groundwork for understanding demographic and economic crime causes.

$$(2)\ (Mischief)i = \beta_0 + \beta_1(Gini\ Index)_i + \beta_2(Average\ Age)_i + \epsilon_i$$

A second specification (2) omits Education Level to test Average Age and Gini Index robustness. Education Level elimination does not affect remaining predictor coefficients or model fit. This implies that Education Level is substantial but not a large confounder in crime, Average Age, and income disparity. The major results' robustness emphasizes the Gini Index and Average Age's central significance in mischief crime variation.

$$(3)\ (Mischief)i = \beta_0 + \beta_1(Gini\ Index)_i + \beta_2(Education\ Level)_i + \beta_3(Average\ Age)_i + \beta_4(PD)_i + \epsilon_i$$

Urbanization impacts may increase crime rates through interactions and conflict, hence a third specification (3) adds Population Density as a control variable. Population density is extremely significant and strongly associated with mischief crimes, showing that denser communities have higher crime rates. The Gini Index and Education Level coefficients are marginally smaller with Population Density, showing partial mediation by urbanization variables.

$$(4)\ (Mischief)i = \beta_0 + \beta_1(Gini\ Index)_i + \beta_2(Education\ Level)_i + \beta_3(Average\ Age)^2_i + \epsilon_i$$

The final specification (4) includes a quadratic Average Age element to investigate a non-linear age-crime link. The squared term for Average Age is marginally significant, suggesting that crime rates initially drop with age but stable or slightly increase at later ages. The model's explanatory power matches the baseline specification, showing that the quadratic term captures modest age-crime association fluctuations.

A similar analysis of Break and Enter Residential/Other Crimes uses Average Age, Gini Index, and Education Level as predictors. The baseline model (5) shows that the Gini Index is again the most important factor, strongly related with Break and Enter Crimes. Average Age is negative but not significant, while Education Level is unrelated.

Baseline Model(5) -

(5) $$(BER)i = \alpha_0 + \alpha_1(Gini\ Index)_i + \alpha_2(Education\ Level)_i + \alpha_3(Average\ Age)_i + e_i$$

Model excluding Education variable(6) :

(6) $$(BER)i = \alpha_0 + \alpha_1(Gini\ Index)_i + \alpha_3(Average\ Age)_i + e_i$$

(7) Population Density (PD) is positively correlated with Break and Enter Crimes, implying that urbanization increases crime in densely populated places.

(7) $$(BER)i = \alpha_0 + \alpha_1(Gini\ Index)_i + \alpha_2(Education\ Level)_i + \alpha_3(Average\ Age)_i + \alpha_4(PD)_i + e_i$$

Using a quadratic specification (8) for Average Age shows that crime rates rise with age before peaking and falling for older populations. This pattern supports demographic ideas that mid-aged people are more vulnerable to crime.

(8) $$(BER)i = \alpha_0 + \alpha_1(Gini\ Index)_i + \alpha_2(Education\ Level)_i + \alpha_3(Average\ Age)^2_i + e_i$$

Across all models, the consistent significance of the Gini Index highlights the critical role of income inequality in shaping crime rates. Population density is another major factor, demonstrating how urbanization affects crime. Since Average Age and Education Level are not significant they may be influenced by other factors. These studies provide a thorough understanding of Mischief and Break and Enter Crimes, helping policymakers offering insights for policymakers to address crime in urban settings.

Table of Results

Model Comparison

	Dependent variable:			
	Mischief Crimes			
	Baseline	Without Education	Quadratic	With Density
	(1)	(2)	(3)	(4)
Average_Age	−0.145***	−0.160***	−0.622**	−0.130***
	(0.035)	(0.035)	(0.280)	(0.035)
Average_Age_Squared			0.005*	
			(0.003)	
Gini_Index	22.352***	22.359***	22.786***	20.988***
	(2.315)	(2.319)	(2.328)	(2.307)
Education_Level	0.001***		0.001***	0.001**
	(0.0004)		(0.0004)	(0.0004)
Population_Density				0.0002***
				(0.00002)
Constant	−0.202	1.317	10.784	−1.232
	(1.590)	(1.529)	(6.593)	(1.585)
Observations	3,519	3,519	3,519	3,519
R2	0.031	0.027	0.031	0.044
Adjusted R2	0.030	0.027	0.030	0.043
Residual Std. Error	10.684 (df = 3515)	10.700 (df = 3516)	10.681 (df = 3514)	10.609 (df = 3514)
F Statistic	36.968*** (df = 3; 3515)	49.457*** (df = 2; 3516)	28.479*** (df = 4; 3514)	40.819*** (df = 4; 3514)
Note:				*p<0.1; **p<0.05; ***p<0.01

FIGURE 2. TABLE OF COMPARISON OF MODELS FOR MISCHIEF CRIMES DEPENDENT VARIABLE

Figure 2 compares baseline, quadratic, omitted variable and control variable (population density) models. The inclusion of the quadratic term for Average Age assesses non-linear relationships, while the Population Density variable controls for contextual density effects.

Model Comparison

	Dependent variable:			
	Break and Enter Residential/Other Crimes			
	Baseline	Quadratic	Without Education	With Density
	(1)	(2)	(3)	(4)
Average_Age	−0.004	0.044*	−0.004	−0.002
	(0.003)	(0.026)	(0.003)	(0.003)
Average_Age_Squared		−0.001*		
		(0.0003)		
Gini_Index	2.773***	2.730***	2.774***	2.649***
	(0.218)	(0.219)	(0.218)	(0.217)
Education_Level	0.00004	0.0001		0.00000
	(0.00004)	(0.00004)		(0.00004)
Population_Density				0.00002***
				(0.00000)
Constant	−0.467***	−1.577**	−0.419***	−0.561***
	(0.150)	(0.620)	(0.144)	(0.149)
Observations	3,519	3,519	3,519	3,519
R2	0.045	0.046	0.045	0.058
Adjusted R2	0.044	0.045	0.044	0.057
Residual Std. Error	1.005 (df = 3515)	1.005 (df = 3514)	1.005 (df = 3516)	0.999 (df = 3514)
F Statistic	55.396*** (df = 3; 3515)	42.424*** (df = 4; 3514)	82.406*** (df = 2; 3516)	53.923*** (df = 4; 3514)
Note:				*p<0.1; **p<0.05; ***p<0.01

FIGURE 3. TABLE OF COMPARISON OF MODELS FOR BER CRIMES DEPENDENT VARIABLE

Figure 3 compares baseline, quadratic, missing variable and control variable (population density) models. The inclusion of the quadratic term for Average Age assesses non-linear relationships, while the Population Density variable controls for contextual density effects.

Specification Checks

Our regression models for Mischief Crimes and Break and Enter Crimes were tested for robustness and reliability using the Variance Inflation Factor (VIF) test for multicollinearity and the Breusch-Pagan test for heteroskedasticity. The models must pass both tests to accurately and meaningfully interpret the dependent-independent variable relationships.

As shown in Figure 4,the VIF test was used to detect multicollinearity, a condition where independent variables are highly correlated, potentially inflating standard errors and distorting coefficient estimates. In both the Mischief Crimes and Break and Enter Crimes models, the VIF values for all variables—Average Age, Gini Index, and Education Level—were well below the commonly accepted threshold of 5. This indicates that multicollinearity is not a significant concern in these models.

Average_Age: 1.07310020580567 **Gini_Index:** 1.05636203489737 **Education_Level:** 1.01668922138194

Figure 4: Results of VIF test

They were tested for heteroskedasticity, which violates the Ordinary Least Squares (OLS) assumption of constant residual variance, using the Breusch-Pagan test (Figure 5). The test result rejected the null hypothesis of homoskedasticity at 1% for the Mischief Crimes and BER models, suggesting heteroskedasticity. Heteroskedasticity implies residual variance not constant. This circumstance can limit regression analysis by causing wasteful standard error estimates and inconsistent hypothesis testing.

```
        studentized Breusch-Pagan test

data:  model_main
BP = 32.687, df = 6, p-value = 1.205e-05

        studentized Breusch-Pagan test

data:  model_main_BE
BP = 47.392, df = 6, p-value = 1.563e-08
```

Figure 5: Results of White test

To fix this, all models used robust standard errors. By compensating for heteroskedasticity, robust standard errors maintain coefficient estimates and provide reliable standard error estimates. Using robust standard errors did not change the analysis's main conclusions. The Gini Index drives crime dynamics, as both models showed the Gini Index to be a strong predictor. All specifications maintained a relationship with Average Age and crime .

VIF and Breusch-Pagan tests ensure model robustness. VIF test showed no multicollinearity, ensuring regression coefficient stability and interpretability. The Breusch-Pagan test revealed heteroskedasticity, which robust standard errors addressed. This dual strategy strengthens findings' credibility.

Discussion

This study aims to investigate the impact of socioeconomic factors like age distribution, income inequality, education level, and population density on the number of Break and Enter Residential/Other and Mischief offenses in Vancouver communities. The results suggest that economic inequality, which is measured by the Gini Index, is the primary predictor of crime even though population density significantly influences crime rates, especially in urban areas.

The Gini Index showed a strong and consistent relationship with crime. A higher Gini Index was strongly associated with higher crime rates, for both Break and Enter Residential/Other and Mischief offenses. This result is consistent with earlier studies that found that neighborhoods with more economic inequality are more likely to have social conflicts. The disparity in wealth can cause societal tensions that increase criminal incidents. Therefore, economic disparity is an important factor to consider when examining crime trends.

Average age had a less consistent impact on crime. Mischief showed a negative correlation which suggests that older communities have less incidents of mischief. In the Break and Enter Residential/Other model, average Age lacks significant predictive ability. The complex nature of crime dynamics may explain this disparity. In areas with aging populations, there may be less opportunities to commit or witness property-related crimes.

Education Level showed different effects across both models for both crime types. In the Mischief model, Education Level was statistically significant, showing a small positive relationship with crime. However, it did not play a considerable role in the Break and Enter Residential/Other model. This could mean that educational qualifications may be a more important predictor for crimes like mischief, linking to the behavior of youth in certain neighborhoods. However, its lack of significance in the BER model suggests that other structural elements may play a stronger role in property crimes.

The effect of Population Density as a control was significant in both models. In areas with high population density, both Mischief and BER crimes were more common, supporting the idea that congested urban areas increase the likelihood of crimes. High Population Density leads to increased social interactions, which may boost offenses such as Mischief, while also providing more targets for crimes like Break and Enter Residential/Other.

Robustness Analysis

The Breusch-Pagan test showed the presence of heteroskedasticity. We used robust standard errors for this. This adjusted for heteroskedasticity and provided more reliable coefficient estimates. The use of robust standard errors did not alter the significance of the Gini Index or Population Density which confirms the robustness of these findings. The analysis of the relationship between Average Age and crime rates reveals a non-linear pattern that is quite insightful.

For **Mischief Crimes**, Figure 6 indicates that neighborhoods with average ages in the 35-45 range experience higher crime rates with crime occurrences decreasing as the population ages. Considering that older populations are generally less likely to commit disruptive or vandalism-related acts, this pattern illustrates how aging has a negative influence on crime. The marginal effect(Figure 7) becomes positive after the middle-aged range (about 60 years) which indicates that mischief crimes may resume in elderly neighborhoods. This might be explained by certain traits of older populations such as reduced social interactions and increased vulnerability.

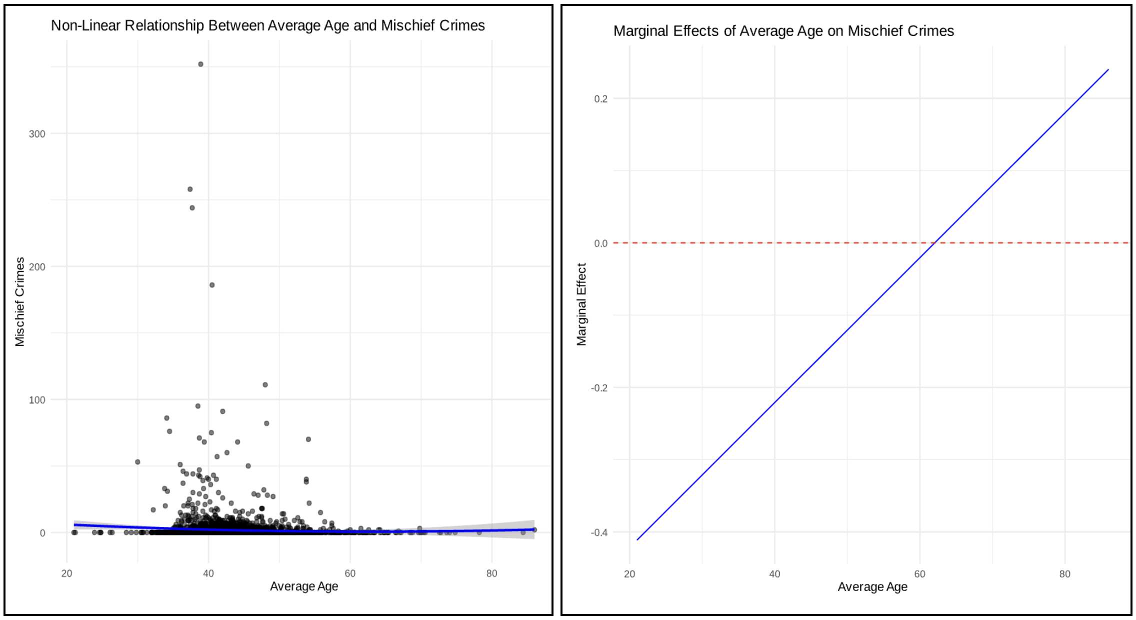

FIGURE 6 & 7: NON-LINEAR RELATIONSHIP (ON LEFT) & MARGINAL EFFECT(ON RIGHT) BETWEEN AVERAGE AGE AND MISCHIEF CRIMES

For **Break and Enter Crimes**, Figure 8 depicts a nonlinear relationship where crime rates initially increase with age, peaking around mid-aged populations (40–50 years). This peak can be seen in figure 9 where marginal effect crosses 0. This point represents increased sensitivity during mid-life when a growth in wealth or property is paired with a lack of stability in the society. In older communities, crime rates gradually drop after this peak potentially due to better security, less mobility, or more deep connections with the community.

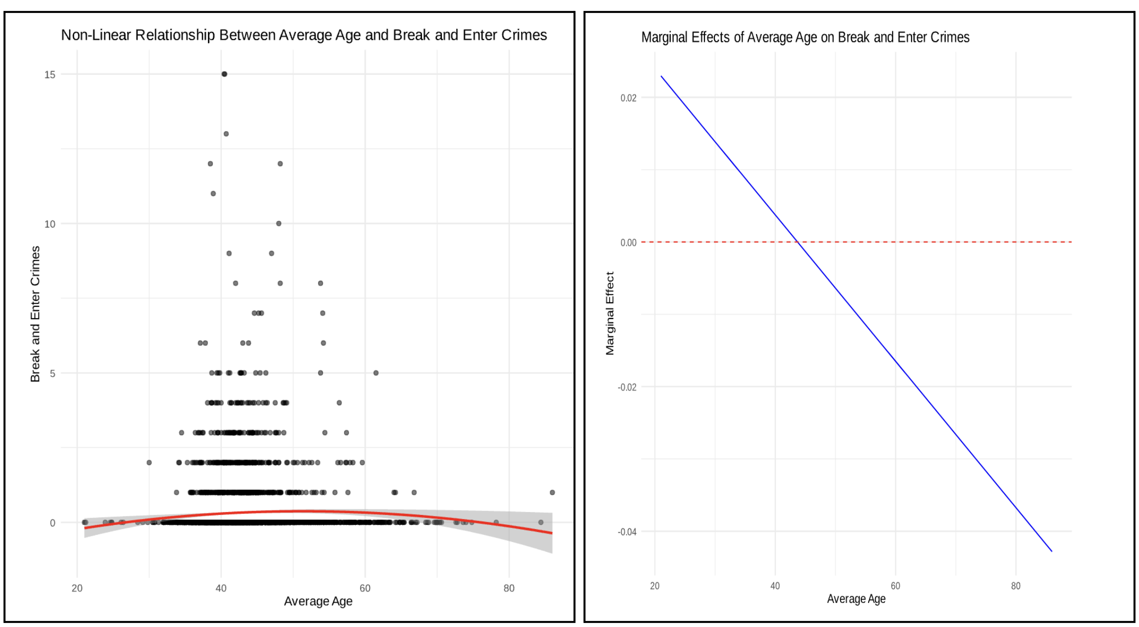

FIGURE 8 & 9 NON-LINEAR (ON LEFT) & MARGINAL EFFECT (ON RIGHT) BETWEEN AVERAGE AGE AND MISCHIEF CRIMES

Although the scatterplots for the average age and crime types as well as the graphs of marginal effects provide robust evidence of non-linear relationships, it is essential to recognize the potential drawbacks. Unmeasured factors like housing instability can confound these relationships.

Additionally, the relatively low explanatory power (R^2 values) of the models suggests that while Average Age, Gini Index, and Education Level are significant predictors, other critical factors influencing crime rates remain unknown. Running a time-series or panel data analysis might provide further insight into how these factors evolve over time.

Conclusion

Our research concentrated on investigating the probable factors influencing Mischief and Break and Enter offenses throughout Vancouver's neighborhoods. We chose to look at variables such as wealth Gini Coefficient, age, education, and population density to comprehend how these aspects influence individuals towards criminal activity.

Our investigation revealed that income inequality, as indicated by the Gini Index, is the most significant determinant of crime. Our findings indicate that communities exhibiting greater disparities between wealthy and low-income residents consistently experience elevated crime rates. When individuals experience economic marginalization, it potentially leads some to engage in criminal actions. Older areas appear to encounter less Mischief cases, indicating more stable community dynamics. However, in the context of Break and Enter offenses, age seems to have little impact. Education appeared to exert a more significant influence on minor charges such as Mischief, although showed approximately no correlation with Break and Enter crimes. Densely populated neighborhoods experienced higher crime rates, which is unsurprising. An increase in population results in potentially increased friction among community members.

We used Variance Inflation Factor (VIF) test to check for multicollinearity and found no multicollinearity in the model.The Breusch-Pagan test detected heteroskedasticity which we then minimized using robust standard errors to get more accurate coefficient estimations. Our statistical analyses validated that the relationships among income inequality, population density, and crime were stable across several specification models. To evaluate the strength of our findings, we investigated further and identified a non-linear correlation between Average Age and crime rates, with Mischief offenses declining in older demographics but reemerging in senior communities, whilst Break and Enter crimes reached their peak in middle-aged groups.

Although the results of our research provide valuable insights, it is merely a small component of a more complex picture of urban safety. Other critical factors influencing crime rates remain unaccounted for, requiring further research. For policymakers, addressing economic disparities could be the key to creating safer neighborhoods in Vancouver.

References

Fajnzylber, P., Lederman, D., & Loayza, N. (2002). Inequality and violent crime. The Journal of Law and Economics, 45(1), 1–39. https://doi.org/10.1086/338347

Griffiths, N. (2024, September 24). Violent crime drops in Vancouver but still higher than pre-pandemic | Vancouver Sun. https://vancouversun.com/news/local-news/violent-crime-drops-still-higher-pre-pandemic

Groot, W., & van den Brink, H. M. (2010). The effects of education on crime. Applied Economics, 42(3), 279–289. https://doi.org/10.1080/00036840701604412

Hansen, K. (2003). Education and the crime-age profile. *British Journal of Criminology*, *43*(1), 141–168. https://doi.org/10.1093/bjc/43.1.141

Machin, S., & Meghir, C. (2004). Crime and economic incentives. The Journal of Human Resources, 39(4), 958. https://doi.org/10.2307/3559034

Attribution

In the code file, Raghav wrote the code for loading the crime data and the summary statistics table for all the variables. Srijan wrote the code for filtering the data, selecting variables of interest and renaming them accordingly. Hrishikesh wrote the code for running the regression on the Mischief model and its different specification checks. Raghav performed the VIF test on the explanatory variables and used stargazer to display the table of results comparing the baseline Mischief model with all its different specifications. We collectively performed the regression and specification checks on The Break and Enter Residential/Other model in a similar manner. Srijan wrote the code for performing the White Test on both the models to check for heteroskedasticity. Hrishikesh wrote the code for testing and displaying the results with robust standard errors. We collectively discussed and wrote the code for the scatterplots of ‘Crime versus Average Age’ and marginal effects graphs for both times of crime. In the word document, the introduction and the data description sections were written by Srijan. Raghav wrote the summary statistics, the model and the table of results. Hrishikesh wrote the discussion and the conclusion section. Each member contributed to writing overall development of the paper by discussing and reviewing all sections, ensuring the consistency and quality of the final work.