Zurich University of Applied Sciences

Crime and Society:
Analyzing the role of education, income inequality, unemployment, and GDP in shaping crime rates

Elena Pietroforte
Stanisław Zapała

Introduction

Crime rates are a critical indicator of a country's quality of life, as they directly affect public safety and social well-being. This study aims to analyze how educational attainment, income inequality, GDP per capita and unemployment collectively influence crime rates. By understanding the factors driving crime, policymakers can develop targeted strategies to enhance public safety and societal stability over the long term. The significance of this study lies in its ability to inform policy, offer valuable insights to design evidence-based interventions to reduce crime and promote social equity.

Literature Review

Previous research has highlighted the significant influence of income inequality, educational disparities, and income level and distribution on crime rates. A study by the International College Subang[1], found out that higher levels of income inequality are associated with increased crime rates, though strong institutional quality can mitigate this effect. Instead, a global analysis[2] showed mixed results, with some suggesting inequality promotes crime, while others point to increased security investments in unequal areas reducing crime.

Regarding educational level, an area-based analysis[3] found a significant relationship between educational inequality and juvenile conviction rates for violent crimes. Higher educational inequality was associated with more convictions for violent and racially motivated crimes, although no significant link was found with property-related crimes

Similarly, research in Indonesia[4] found that wider income gaps, particularly in discretionary spending, correlate with higher crime rates, suggesting perceptions of relative deprivation play a critical role.

Methodology Approach and Data

This study utilizes a dataset in the period from 1990 to 2022. The data is merged from multiple sources to create a comprehensive cross-sectional dataset, which is subsequently converted into a panel format. Interpolation techniques are employed to address missing values, ensuring data completeness. The analysis includes a series of diagnostic tests to ensure robust model specification:

- **Poolability Test**: To determine whether regional and temporal differences are significant or if the data can be pooled into a single model.
- **Autocorrelation Test (Breusch-Godfrey)**: To check for serial correlation in the panel data.**Heteroscedasticity**
- **Test (Breusch-Pagan)**: To detect variability in the error terms across observations.
- **Cross-Sectional Dependence Test**: To account for interdependencies between cross-sectional units.
- **Collinearity Analysis:** To ensure no collinearity in the model.

Empirical analyses involve error component regression models, including various panel estimators such as Panel OLS, Fixed Effects, and Multi Equation Panel Models. The robustness of findings is ensured through model comparison and statistical validation.

[1] “Income Inequality and Crime: Evidence from a Dynamic Panel Data Approach”
[2] "Income Inequality and Violent Crime: Evidence from Mexico's Drug War"
[3] “Educational Inequality and Juvenile Crime: An Area Based Analysis
[4] "How Economic Indicators Drive Crime? Empirical Study in Developing Country, Indonesia”

Data Section

Sources and Sample

For this analysis, six countries were selected: Germany, the United Kingdom, Canada, the United States, Costa Rica, and Brazil. These countries were chosen to represent a diverse range of sizes, backgrounds, and geographical regions (Europe, North and South America). While initially aimed to include larger nations such as Russia and China, unavailability of consistent and reliable data limited data selection. Some data remained unavailable despite consulting multiple sources due to incompleteness and different time spans. This issue was finally addressed by applying polynomial interpolation using Python.

The datasets utilized in this analysis were obtained from trusted sources, including Our World in Data, the Human Development Report, World Bank database and Macrotrends, which provided a solid foundation for investigating the factors influencing crime. The sample covers yearly data from 1990 to 2022, with a total of 32 observations per variable. The key variables under examination are:

- **Crime rates**: Intentional homicides per 100'000 citizens, with "*the core element of intentional homicide being the complete liability of the direct perpetrator, which thus excludes killings directly related to war or conflicts, self-inflicted death (suicide), killings due to legal interventions or justifiable killings (such as self-defence), and those deaths caused when the perpetrator was reckless or negligent but did not intend to take a human life (non-intentional homicide).*" by the United Nations Department on Drugs and Crime.
- **Educational attainment**: measured by expected years of schooling (number of years of education new students are anticipated to complete) and average years of schooling (the average number of years of education completed by individuals).
- **Income inequality**: represented by the Gini index (a value between 0 and 1, where values closer to 1 indicate greater inequality).
- **Unemployment**: expressed as a percentage of the total labor force.
- **GDP per capita**: expressed in USD with const=2015.

Overall summary statistics

After transforming data, descriptive statistics were conducted on crime rates, expected and mean years of school, unemployment, Gini coefficient and GDP per capita. Relevant indicators are presented in Table 1. "NA's" row represents the count of NA's in the dataset for each variable, they are not included in further models and calculations. Accounting for the possibility of differences between countries, additional descriptive statistics were conducted for each country; they are available as Attachment 1 in appendix.

Table 1: *Descriptive statistics for numeric variables in dataset*

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

Histograms of occurrences in Appendix 2 show that variables most probably do not have a normal distribution, closest to normal being percentage of unemployment. That was expected as the dataset has a relatively small number of observations for each country and each country possesses a different distribution (with different mean and standard deviation) of their own crime rates, education variables, unemployment and Gini coefficient. After taking into account also scatter plots of each explanatory variable with crime rates (Appendix 3), first intuition is that these differences between countries will have a significant and substantial impact on regression models and that the data cannot be pooled together but has to be analysed taking into account these divergences.

Moreover, scatter plots show an evident relationship between explanatory and explained variable in certain countries (for example Brazil and Costa Rica education vs crime rates), and less evident in other countries. Surprisingly, relationships for different countries seem to even be opposing. Further exploration of correlations is visible in Appendix 4 (overall correlation) and Appendix 5 (individual correlations for each country). Mean Years of Education and Expected Years of Education have high correlation (overall and individually in different countries), therefore multicollinearity is to be expected.

Empirical Methods

Relationship between variables

This study employed a panel data analysis to examine the impact of education level, unemployment and Gini coefficient on crime rate. Fixed Effects model and multiple linear regression using pooled data were used to explore this relationship. The Fixed Effects model was chosen over the First Differences approach based on the analysis of time plots, which did not suggest a random walk and possibility of eliminating probable serial correlation during analysing models.

Estimation equation and approach

Multiple linear regression estimation equation for pooled dataset is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

y - dependent variable
x_k - explanatory variables
β_0 - intercept
β_k - corresponding coefficients
u - error term

The Fixed Effects estimation equation is as follows:

$$y_{t,i} = \beta_1 * x_{1,i,t} + \beta_2 * x_{2,i,t} + \beta_3 * x_{3,i,t} + \ldots + \beta_k * x_{k,i,t} + c_i + u_{i,t}$$

$y_{t,i}$ - dependent variable in country i at time t
$x_{k,i,t}$ - independent variables in country i at time t
β_k - corresponding coefficients
c_i - unobserved time-variant individual effect for country i
$u_{i,t}$ - error term

Two models were considered - both multiple linear regression and Fixed Effects model, due to uncertainty, whether the data could be pooled together. Next steps included aggregating the data, polynomial interpolation to handle missing data, performing descriptive statistics and fitting regression models to estimate impact of explanatory variables on crime rates. Further actions included testing for poolability, homoscedasticity, autocorrelation, normality of residuals and cross-sectional dependence.

Assumptions and tests

MLR base assumptions consist of linear relationship, independence of errors, homoscedasticity, no perfect collinearity and strict exogeneity. The Fixed Effects model assumes a linear relationship, independence of errors across entities, no perfect collinearity, cross-sectional independence of errors, homoscedasticity, and strict exogeneity. Additional poolability test was conducted to decide whether the MLR or FE model was to be applied. Linear relationship was ensured through analysis of scatter plots of residuals vs fitted values (Appendix 6). Homoscedasticity was analyzed using the **Breusch-Pagan test**, which detects whether the variance of residuals is constant across all levels of the independent variables. Serial correlation was examined using the **Breusch-Godfrey test**, designed to identify whether error terms in the model are correlated across observations, a condition that could compromise the efficiency of regression estimates. Collinearity was addressed through Variance Inflation Factor (**VIF**) analysis, which quantifies the extent to which predictor variables are linearly related to one another, helping to identify potential multicollinearity issues. Cross-sectional dependence was assessed using **Pesaran's**

test, which evaluates whether residuals from different cross-sectional units are correlated, a critical consideration in panel data analysis where cross-sectional interdependencies can affect the reliability of model estimates.

Limitations of approach

The short observation period limits the ability to study the relationship between variables in each country separately, which can cause potential bias in the model. Despite varied data from both socially and economically different countries, this model does not cover neither the whole world nor its majority, therefore real relationships between variables can differ in different countries and circumstances. Moreover, R-squared value suggests that crime rates are influenced by other factors than included in the model, so results of this study should be treated with caution.

Poolability

First step was to analyse whether data from separate countries could be pooled together. This step has a significant impact on further analysis and choice of proper model. The Chow test was employed for this purpose, a statistical test used to determine whether the relationships between variables differ significantly across groups—in this case, countries. The test evaluates whether pooling the data into a single model would result in a significant loss of explanatory power compared to using separate models for each group. Result was equal to 15.565 with p-value $< 2.2e - 16$, what indicated that the data could not be pooled together, confirming substantial heterogeneity between countries. Consequently, the Fixed Effects model was selected as the appropriate approach, as it accounts for unobserved, country-specific factors that remain constant over time.

Collinearity

Possibility of multicollinearity was analysed in a few steps. First step included analysis of both overall and separate correlation matrices. In both cases GDP per capita and Mean Years of Education have shown high correlation (higher than 0.9) and were suspected of collinearity. Due to the Fixed Effects model standard VIF method to access collinearity could not have been used, therefore the condition index was utilized, showing no signs of multicollinearity.

Autocorrelation

Breusch - Godfrey/Wooldridge test for serial correlation in panel models was utilized with result 44.248 (29 degrees of freedom and p-value = 0.03475), showing serial correlation. To address this issue two methods were tested: introducing dummy variable Year and introducing lags on both explained and explanatory variables. Second method gave better results measured by change in adjusted R-squared, therefore in the end lags on Crime Rates and GDP per capita remained in the model.

Homoscedasticity

In the initial model heteroscedasticity was detected after conducting the Breusch - Pagan test with result equal to 7.2762 (4 degrees of freedom and p-value = 0.122). This issue was addressed by changing the specification of the model and introducing heteroscedasticity and autocorrelation consistent (HAC) covariance matrix, ensuring proper measures for standard errors.

Cross - sectional dependence

Pesaran's test, which was conducted in order to assess cross - sectional dependence, showed result equal to -1.1935 and p-value = 0.2327, confirming no cross - sectional dependence.

Results

This study compared multiple linear regression models for changes in crime rates assessing impact of changes in numerous explanatory variables. Specification of the model was improved by removing statistically insignificant variables and introducing lags on both crime rates and GDP per capita. Final results are presented in the following table (Table 2) with the best model being model 4.

Table 2: *Final regression models and coefficients' estimations*

	Dependent variable:			
	crime_rate			
	(1)	(2)	(3)	(4)
lag(crime_rate, 1)			0.411***	0.779***
			(0.046)	(0.046)
exp_years_of_school	0.946***	0.936***	0.559***	0.171**
	(0.095)	(0.081)	(0.083)	(0.085)
mean_years_of_school	0.288***	0.337***		
	(0.104)	(0.100)		
unemployment_	0.004			
	(0.038)			
gini_coef	0.052	0.062	0.011	0.010
	(0.039)	(0.038)	(0.031)	(0.028)
GDP	−0.0002***	−0.0002***	−0.0001***	−0.0004***
	(0.00002)	(0.00002)	(0.00002)	(0.0001)
lag(GDP, 1)				0.0004***
				(0.0001)
Observations	184	189	185	185
R^2	0.579	0.595	0.692	0.840
Adjusted R^2	0.555	0.575	0.676	0.831
F Statistic	47.586*** (df = 5; 173)	65.811*** (df = 4; 179)	98.373*** (df = 4; 175)	686.009*** (df = 5; 174)
Note:				*p<0.1; **p<0.05; ***p<0.01

Interpretation of estimated parameters

Variable Expected Years of Education (Model 4) shows a statistically significant positive relationship with crime rates equal to 0.171 and significance level of less than 0.05 (**). That means that a change in Expected Years of Education by 1 year would increase Crime Rate by 0.171 percentage point. Howether that seems to be irrational, this study does not assess quality of education, education inequality and other factors that can influence both crime rates and education.

Gross Domestic Product per capita (Model 4) shows significant negative relationship with crime rates equal to -0.0004 at significance level of less than 0.01 (***). Although this value seems small, it is to be remembered that GDP per capita occurs in large numbers. This value can be interpreted as: when GDP per capita rises by 1 USD (USD as 2015=const) crime rate decreases by 0.0004 percentage points.

Gini Coefficient (Model 4) shows a statistically insignificant slightly positive relationship with crime rates. That might suggest that crime rates are less affected by changes in Gini coefficient. Despite insignificance Gini coefficient proved to have a positive impact on quality of the model and is supported by theoretical research given in Introduction.

Lagged Crime Rates and lagged GDP per capita both state statistically significant positive relationships with crime rate at significance level less than 0.01 (***). Lagged Crime Rate equals 0.779, meaning that with change of 1 percentage point in crime rate for period t-1 comes change of 0.779 percentage point in crime rates for period t. Lagged GDP per capita means that change of 1 USD in GDP per capita for period t-1 results in change of 0.004 percentage points in crime rates for period t.

The interpretation of the estimated parameters suggests a complex relationship between crime rates and the included explanatory variables. The positive relationship between *Expected Years of Education* and crime rates is counterintuitive, highlighting the need for further research into factors such as educational quality and inequality. On the other hand, *GDP per capita* demonstrates a significant negative relationship with crime rates, emphasizing the role of economic prosperity in reducing crime, while the *Gini Coefficient* shows minimal impact, indicating that income inequality alone might not be a major driver of crime in this context.

Conclusion

This study examined the factors influencing crime rates across several countries, focusing on the roles of educational attainment, income inequality, unemployment, and GDP per capita. By utilizing panel data and employing various regression models, it aimed to identify significant relationships between these factors and crime rates, offering insights for policymakers and contributing to the broader understanding of crime dynamics.

The analysis revealed that educational attainment, income inequality, and unemployment have nuanced and varied effects on crime rates, with important differences across countries. One of the most notable findings is the positive relationship between *Expected Years of Education* and crime rates, which, despite its seeming irrationality, highlights the complex interactions between educational inequality, quality, and broader societal factors that may drive crime. It suggests that the mere quantity of education does not necessarily result in lower crime rates without accounting for other influencing factors like educational quality or inequality.

Additionally, the *Gini coefficient* (income inequality) exhibited a slightly positive but statistically insignificant relationship with crime rates in the final model, aligning with some theoretical perspectives but diverging from others. This indicates that while inequality might be one factor contributing to crime, it is not the sole determinant, and its effects may be overshadowed by other variables in the model. These results are in line with the broader body of research that suggests a complex and sometimes indirect relationship between income inequality and crime, often moderated by factors such as social policies, law enforcement, and local socioeconomic conditions.

Unemployment, despite being a well-documented driver of crime in the literature, did not show the expected strong correlation with crime rates in this study. The role of unemployment is likely indirect, with its effects possibly mediated by education and income inequality, making it harder to isolate its direct impact on crime.

The negative relationship between *GDP per capita* and crime rates in the models was another crucial finding. As expected, higher economic output, represented by GDP per capita, is generally associated with lower crime rates. This relationship reinforces the idea that economic prosperity and opportunities can reduce the motivations for crime, likely by improving living standards and reducing the frustrations associated with poverty.

Importantly, the analysis revealed the necessity of accounting for cross-country differences when analyzing such data. The *Poolability Test* confirmed that data could not be pooled together due to the substantial heterogeneity between countries. As a result, the Fixed Effects model was more appropriate, allowing for the examination of within-country variations over time.

The study's limitations, such as the relatively short time frame of 1990 to 2022 and the incomplete nature of available data, should be taken into account when interpreting the findings. Additionally, the model did not account for all potential variables influencing crime rates, such as law enforcement practices, social safety nets, or political stability.

Overall, the study contributes to a deeper understanding of how various economic and educational factors influence crime rates. The findings suggest that while education, economic inequality, and unemployment play important roles in shaping crime, their effects are not as straightforward or uniform across different contexts as might be expected. Policymakers must consider these complexities when designing interventions aimed at reducing crime, particularly through education reform, addressing inequality, and promoting economic growth.

Future research should focus on exploring the deeper mechanisms at play, including the quality of education and broader social policies, to provide a more holistic understanding of crime dynamics. Moreover, extending the observation period and including more countries could help refine these results and offer more generalized conclusions.

Appendix

Appendix 1: *Additional descriptive statistics for individual countries*

Descriptive statistics for Germany

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

Descriptive statistics for United Kingdom

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

Descriptive statistics for United States

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

Descriptive statistics for Brazil

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

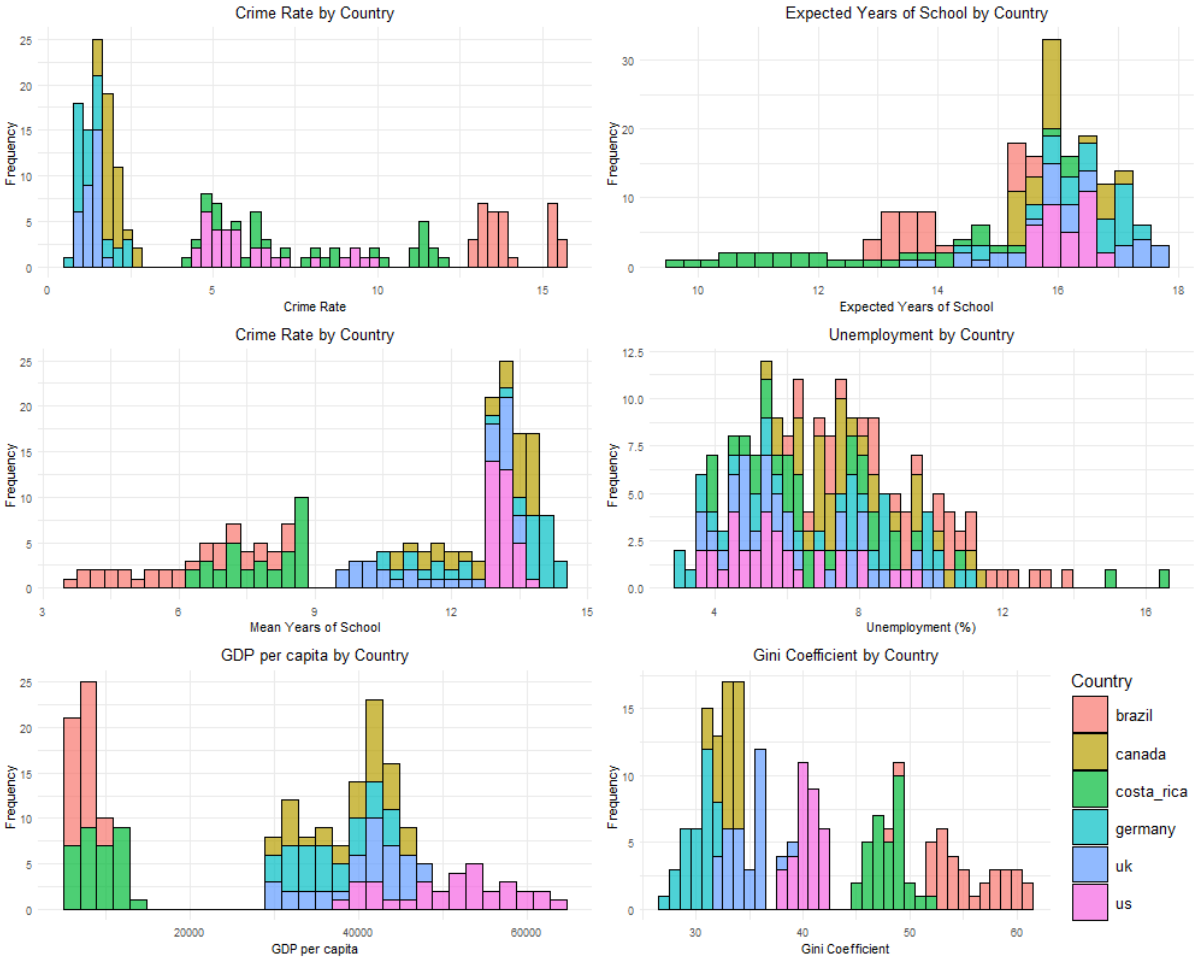
Descriptive statistics for Canada

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

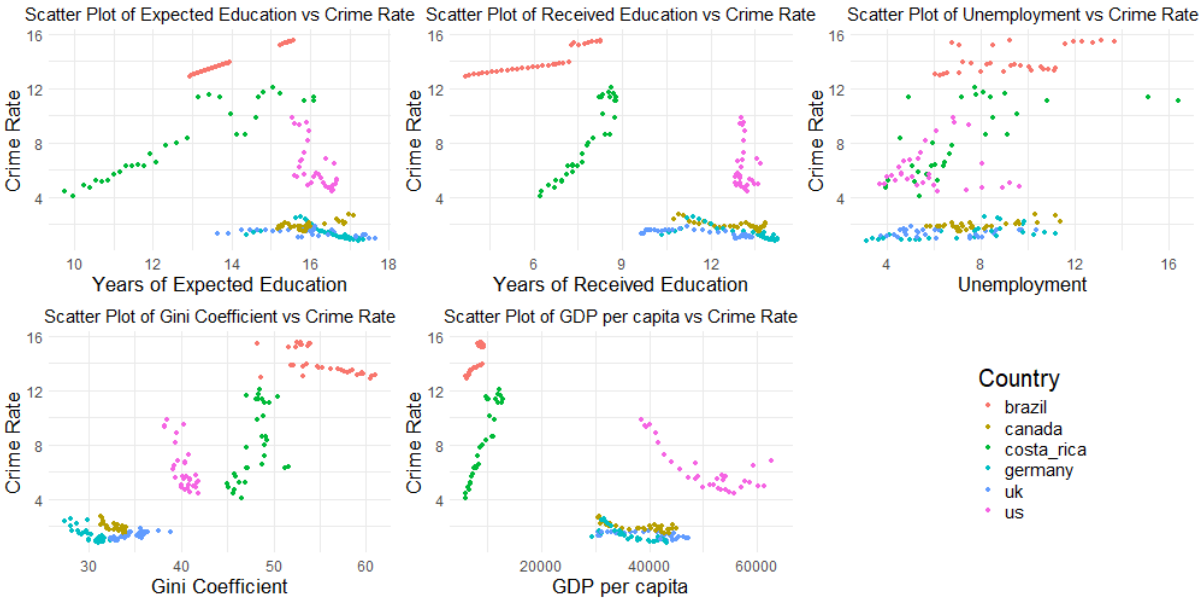
Descriptive statistics for Costa Rica

	crime_rate	exp_years_of_school	mean_years_of_school	unemployment_	gini_coef	GDP
Min.	0.749	9.740	3.685	3.140	27.378	5911.687
1st Qu..25%	1.548	14.349	8.278	5.340	32.975	9324.064
Median	4.237	15.784	12.077	7.070	38.675	35769.078
Mean	5.547	15.246	10.811	7.248	40.406	30707.514
3rd Qu..75%	9.345	16.413	13.250	8.710	48.286	42921.562
Max.	15.579	17.668	14.256	16.430	61.032	63720.764
Std Dev	4.852	1.721	2.988	2.422	9.086	17053.325
NA's	6.000	0.000	0.000	7.000	4.000	0.000

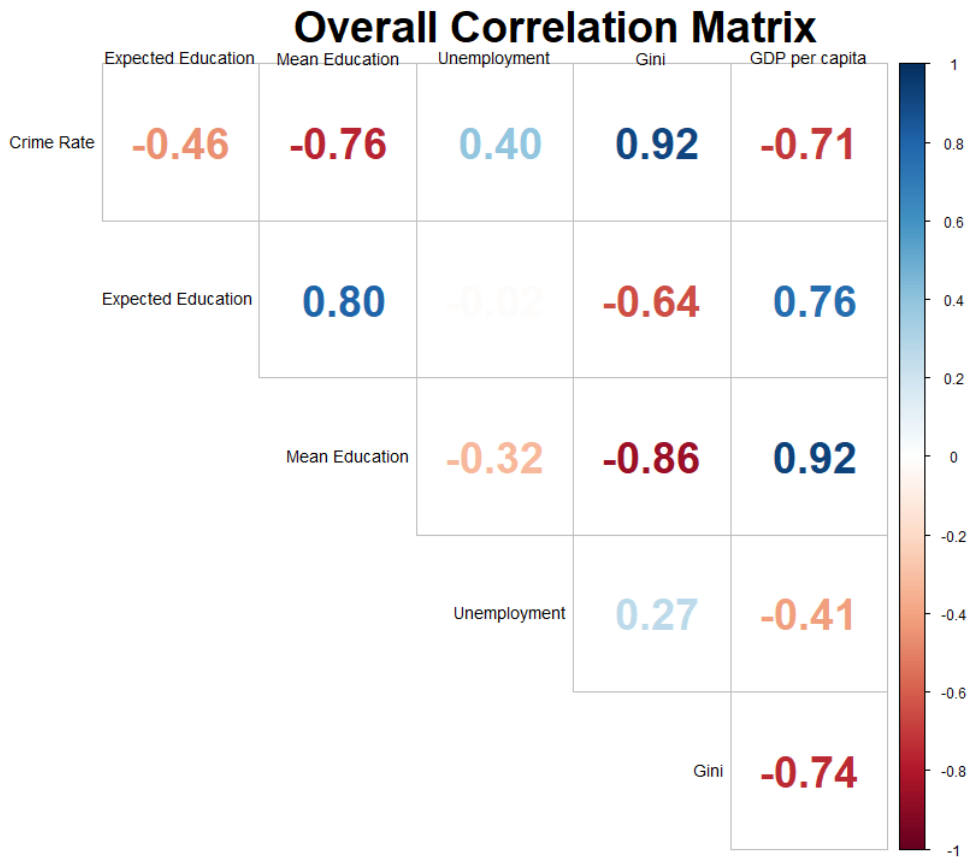
Appendix 2: *Frequencies of values by countries*



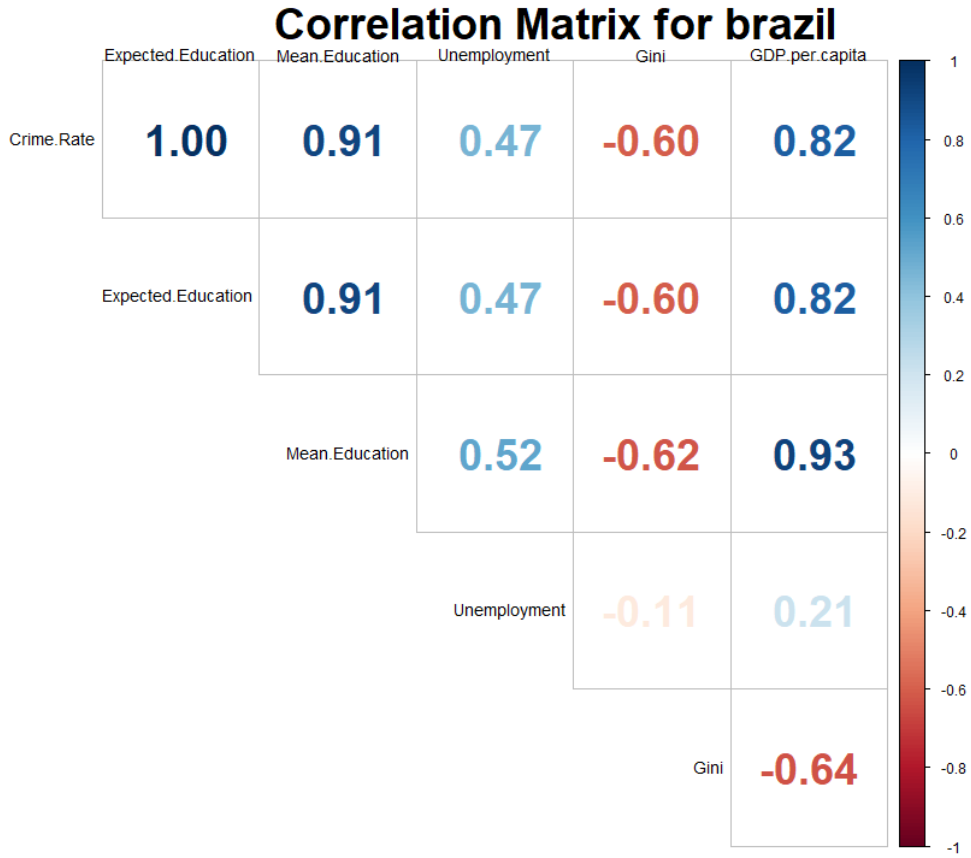
Appendix 3: *Scatter plots of explanatory variables against Crime Rates*

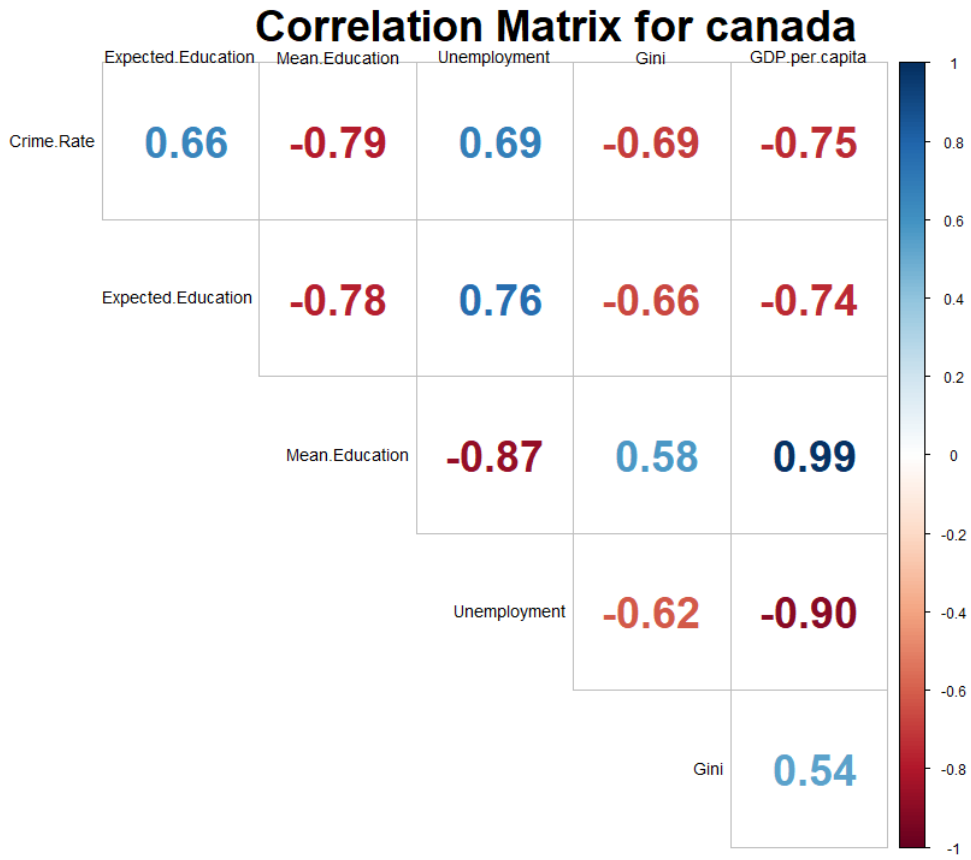


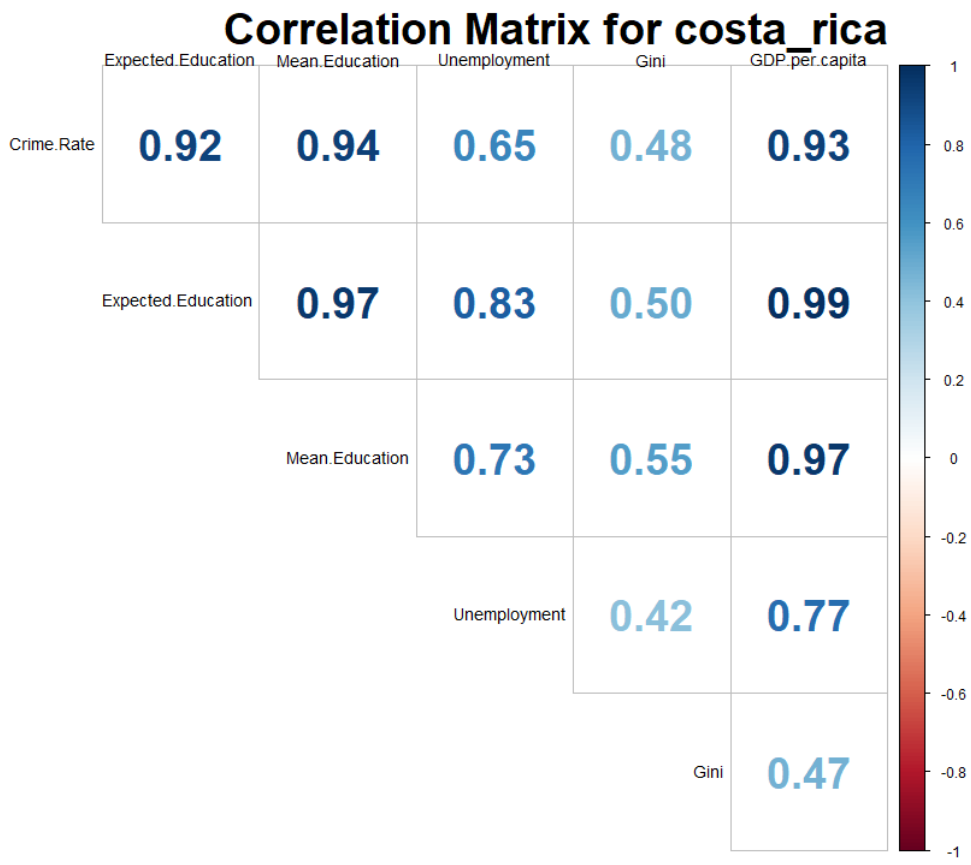
Appendix 4: *Overall Correlation Matrix*

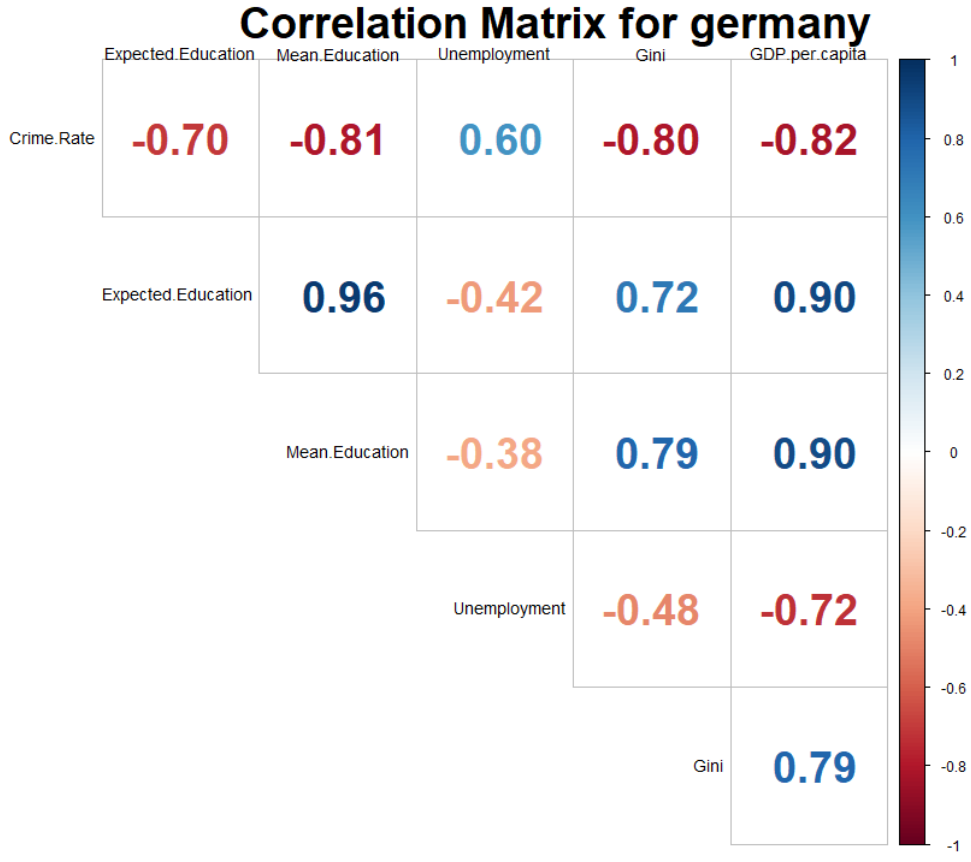


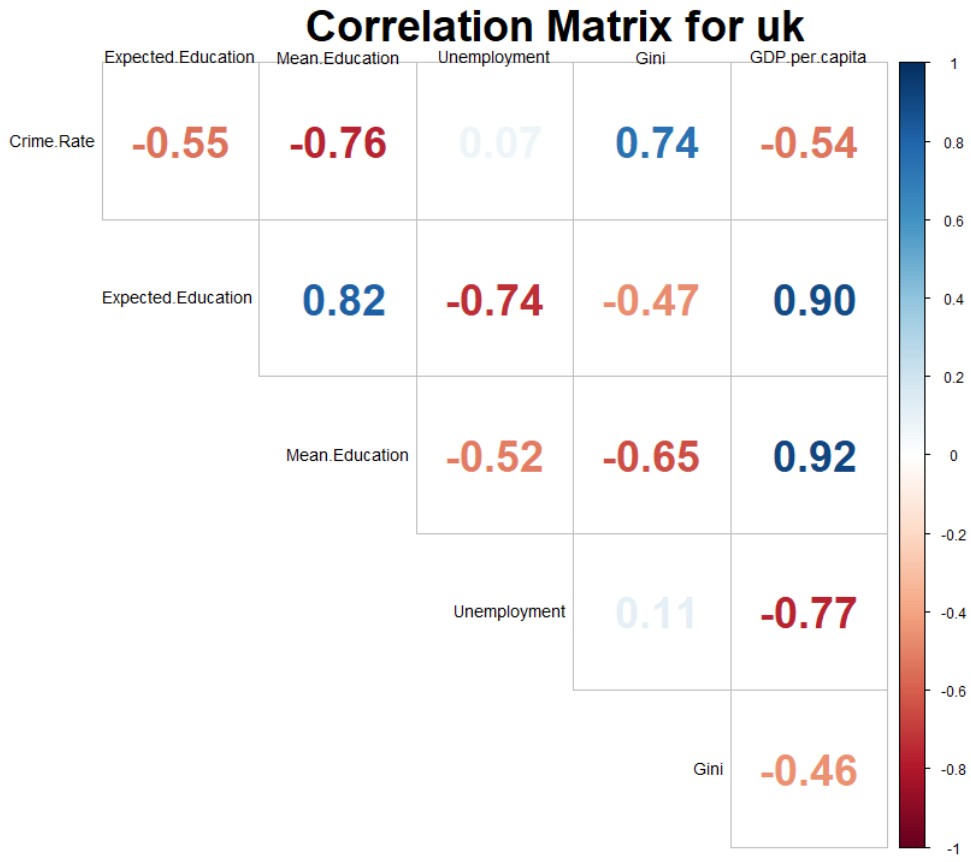
Appendix 5: *Additional correlation matrices for individual countries*

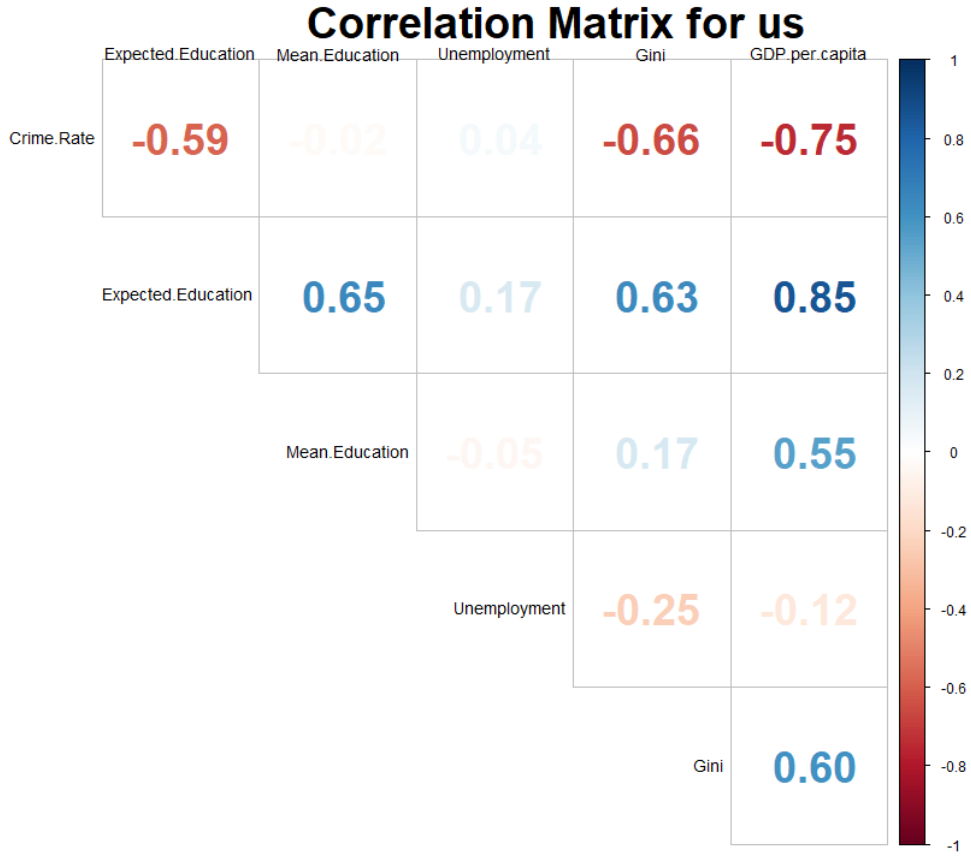





Correlation Matrix for costa_rica
Expected.Education
Mean.Education
Unemployment
Gini
GDP.per.capita
Crime.Rate
0.92
0.94
0.65
0.48
0.93
Expected.Education
0.97
0.83
0.50
0.99
Mean.Education
0.73
0.55
0.97
Unemployment
0.42
0.77
Gini
0.47
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Correlation Matrix for germany
Expected.Education
Mean.Education
Unemployment
Gini
GDP.per.capita
Crime.Rate
-0.70
-0.81
0.60
-0.80
-0.82
Expected.Education
0.96
-0.42
0.72
0.90
Mean.Education
-0.38
0.79
0.90
Unemployment
-0.48
-0.72
Gini
0.79
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Correlation Matrix for uk
Expected.Education
Mean.Education
Unemployment
Gini
GDP.per.capita
Crime.Rate
-0.55
-0.76
0.07
0.74
-0.54
Expected.Education
0.82
-0.74
-0.47
0.90
Mean.Education
-0.52
-0.65
0.92
Unemployment
0.11
-0.77
Gini
-0.46
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Correlation Matrix for us
Expected.Education
Mean.Education
Unemployment
Gini
GDP.per.capita
Crime.Rate
-0.59
-0.02
0.04
-0.66
-0.75
Expected.Education
0.65
0.17
0.63
0.85
Mean.Education
-0.05
0.17
0.55
Unemployment
-0.25
-0.12
Gini
0.60
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Appendix 6: *Residuals vs fitted values for pooled and Fixed Effects models*