IDS 702 Final Report: Analysis of Police-Related Incidents: Economic and Demographic Influences

Afag, Peter, Eleanor, Mobasser

2024-12-15

Abstract

This report investigates the relationship between economic conditions, racial composition, and the likelihood of being armed in police-related incidents. Using data from the Guardian's database on police killings linked to census data, we analyzed the role of unemployment rates, income, and poverty in shaping these outcomes. Multinomial regression revealed significant associations between economic conditions and racial composition, particularly for Black individuals in counties with lower income and higher unemployment. Logistic regression found no significant predictors of armed status, even after addressing model assumption violations. These findings highlight the influence of structural inequality on racial disparities while underscoring the need for further research on armed status predictors.

Introduction

Police-related incidents have become a focal point of public safety research and policy discussions in recent years. The complex interplay between socioeconomic factors, racial dynamics, and law enforcement practices has sparked intense debate and calls for reform across the United States. Understanding the underlying patterns and potential contributing factors to these incidents is crucial for developing effective strategies to enhance community safety and promote equitable policing practices.

Research has shown that economic conditions and racial demographics can significantly influence crime rates and police-community interactions. Studies have found that neighborhoods with higher concentrations of low-income residents and residents of color tend to experience higher rates of police-related deaths. Additionally, unemployment and economic stress have been associated with increased rates of certain types of crime, potentially affecting police encounters. However, the relationship between economic factors and crime is complex, with some research suggesting that economic downturns may actually decrease criminal opportunities in certain contexts.

This report aims to contribute to this body of knowledge by examining two key research questions:

1. Is there an association between the economic conditions of a county and the racial composition of individuals involved in police-related incidents?
2. Does the likelihood of an individual being armed during such incidents vary based on their age and the unemployment rate in their area?

By analyzing data from the Guardian's database on police killings, linked with the American Community Survey (2015), this study seeks to provide insights into these critical questions (Guardian, 2015). The dataset, which includes demographic, economic, and geographic information, offers a unique opportunity to explore the potential relationships between socioeconomic factors and the characteristics of police-related incidents. Understanding these dynamics can inform evidence-based policy-making and interventions aimed at improving community safety and promoting equity in law enforcement practices.

Methods

Data

The dataset consists of **467 observations** and **34 columns**, providing detailed information on police-related incidents across different counties in the United States. The dataset includes variables that capture demographic information (e.g., race, gender, age), geographic details (e.g., city, state, latitude, longitude), economic indicators (e.g., household income, poverty rate, unemployment rate), and other socioeconomic factors. The primary outcome variable, `raceethnicity`, categorizes individuals involved in police-related incidents into racial groups, including "White," "Black," "Hispanic/Latino," "Asian/Pacific Islander," and "Native American." For this analysis, categories with very low representation, such as "Asian/Pacific Islander" and "Native American," were excluded to ensure model reliability.

Key economic variables include: Tract-level median household income as a share of county-level median household income (will be refered as a relative income), poverty rate at the county level, unemployment rate at the county level, the proportion of the population with a Bachelor's degree or higher.

To prepare the data for analysis, entries with missing values were excluded. Since the number of missing values was minimal (approximately 2 or 3 for some variables), these rows were dropped without significantly affecting the overall dataset size or quality.

Model fitting and evaluation

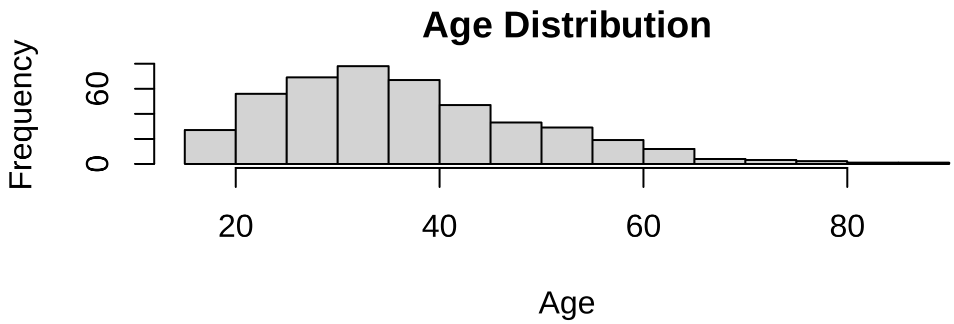
For Research Question 1, which examines the relationship between economic conditions and racial composition, a multinomial logistic regression model was appropriate given the categorical nature of the outcome variable, race/ethnicity, with multiple nominal categories. The independent variables, relative income, poverty level, college level and unemployment rate, were chosen to capture key socioeconomic dimensions influencing racial or ethnic group differences. This model enables the examination of how economic factors are associated with the likelihood of belonging to different racial or ethnic categories.

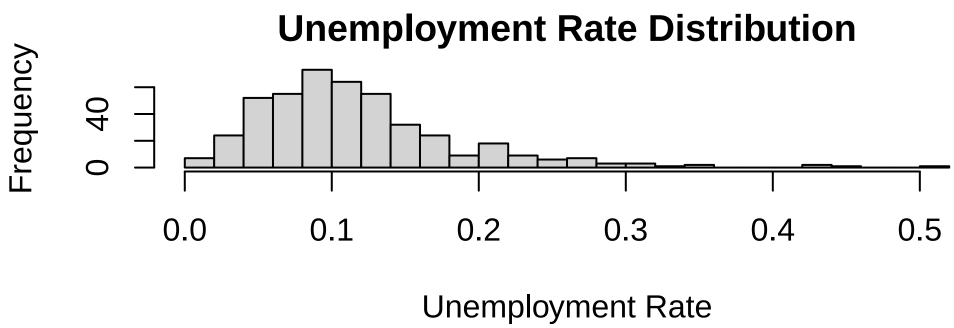
For Research Question 2, a logistic regression model was employed to analyze the binary outcome variable, armed (armed status of the individual). The independent variables include age, urate (unemployment rate), gender, race/ethnicity, and college level which serve as proxies for demographic and socioeconomic factors. All interaction terms between these variables are also included to adjust for possible multicollinearity between the variable.

Both models were supported by diagnostic evaluations to ensure reliability. Correlation matrix and Variance Inflation Factor (VIF) was used to check for multicollinearity and correlation, and Cook's distance assessed the influence of individual data points. Classification performance was further evaluated using a confusion matrix, accuracy, precision, recall, and F1 scores, providing a detailed breakdown of the model's predictive capabilities for each category. The area under the Receiver Operating Characteristic (ROC) curve (AUC) was used as an additional metric to assess the model's ability to distinguish between categories. Residual diagnostics, including residual vs. fitted plots, were used to evaluate linearity and homoscedasticity. These diagnostic evaluations and performance metrics ensure that the models are well-suited to address the research questions, effectively capturing the underlying dynamics in the data. All statistical analyses were conducted using R programming language

Results

Exploratory Data Analysis



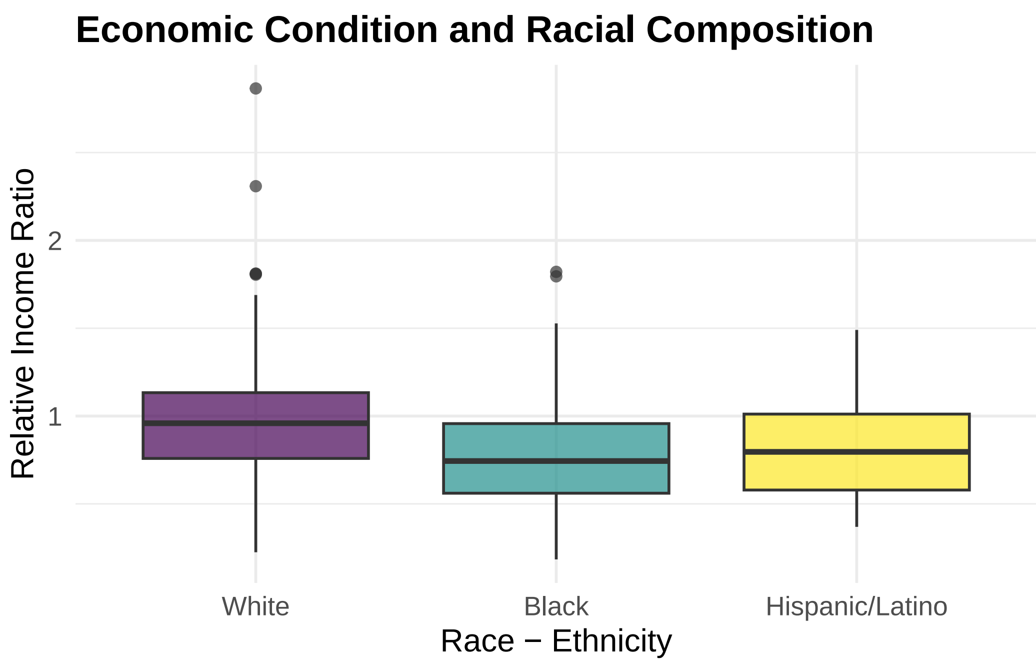


The demographic analysis shows that the deceased had a mean age in their 30s, with the age range extending to over 80. The unemployment rate in their residential tracts averaged 10%, reaching a maximum of 51%. The gender distribution was heavily skewed, with 428 males and 20 females.

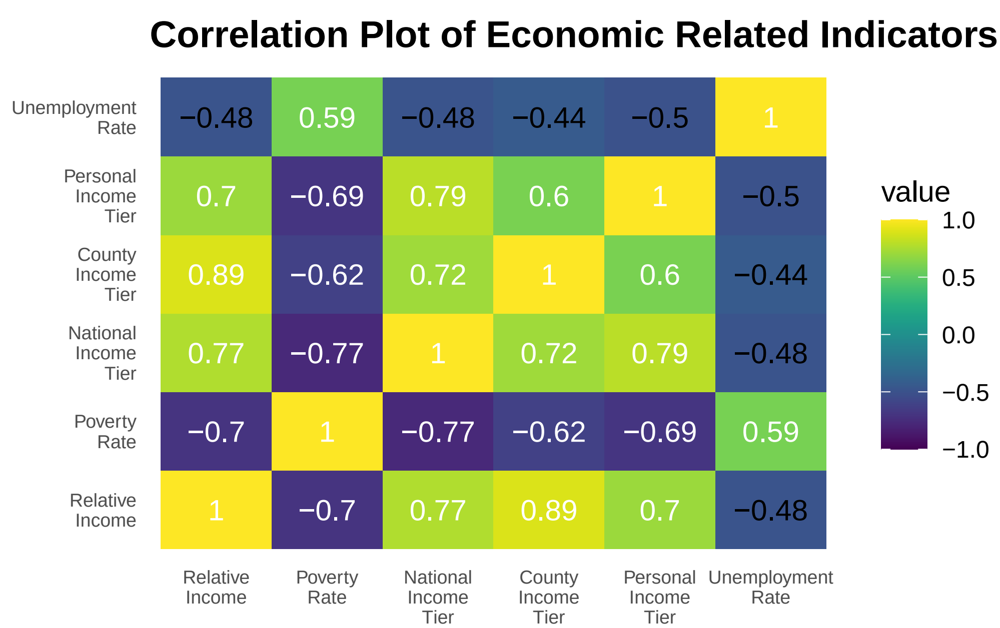
Table 1: Race Ethinicity Summary Statistics

raceethnicity	n	Percentage
White	235	52.46
Asian/Pacific Islander	10	2.23
Black	133	29.69
Hispanic/Latino	66	14.73
Native American	4	0.89

Due to sample size limitations, the racial composition analysis was restricted to three main groups: White (52%), Black, and Hispanic/Latino, with Asian and Native American groups (each <5%) with less than 30 people being excluded from the analysis.



For our **research question 1**, it is clear that a multinomial logistic regression is proficient in this case. The box plot reveals distinct income disparities across racial groups, suggesting relative income as a potentially significant predictor.



Analysis of economic indicators showed high correlations among most variables, with tract-level unemployment rate being the exception. Based on this correlation analysis, we selected unemployment rate and one additional economic indicator for the final model.

Research Question 1: Economic Conditions and Racial Composition

The results of the multinomial logistic regression model indicate that specific economic conditions are significantly associated with the racial composition of individuals involved in police-related incidents. Counties with **lower relative income** were significantly more likely to have individuals categorized as *Black* rather than *White*, suggesting that economic disadvantage disproportionately impacts Black individuals. Additionally, counties with **higher unemployment rates** showed a strong association with the likelihood of individuals being classified as *Black*, emphasizing the role of joblessness in shaping these disparities.

For *Hispanic/Latino* individuals, economic predictors such as relative income and poverty rate did not show statistically significant relationships. However, these findings highlight the complexity of factors influencing racial composition in police-related incidents and suggest that the effects of economic conditions may vary across racial groups.

The confidence intervals provide a range of values within which the true effects of these factors are likely to fall, and they support the reliability of these findings. Additionally, odds ratios greater than 1 suggest an increase in likelihood, while those less than 1 suggest a decrease. For instance, the odds of being classified as Black increase significantly with higher unemployment, whereas the odds decrease with higher relative income. These conclusions are further supported by the low p-values observed for several predictors in the model, indicating that these relationships are statistically significant. One special note: the unusually large confidence interval for unemployment rate suggests a high degree of uncertainty in the estimate. This may be due to sparse data for counties with very high unemployment, the influence of outliers, or multicollinearity with other predictors. While the effect remains significant, the wide confidence interval warrants caution in interpreting its precise impact. Future research should address this by examining extreme values or transforming the variable to stabilize the estimates.

Overall, these results demonstrate how structural factors like economic inequality and unemployment disproportionately impact certain racial groups in police-related incidents. By identifying these relationships, policymakers and researchers can better understand and address the systemic factors contributing to these disparities.

Fitting model on comp_income, poverty rate, unemployment and college, share_black

Table 2: Combined Table: Coefficients, Std.Errors, Odds Ratios, Confidence Intervals, and p-values

	Variable	Coef.	SE	OR	CI Low	CI Up	p-val	Sig.
Black	Intercept	-1.28	0.74	0.28	0.07	1.19	0.08	
Hispanic/Latino	Intercept	-0.76	0.96	0.47	0.07	3.08	0.43	
Black1	Relative Income	-1.57	0.55	0.21	0.07	0.61	0.00	***
Hispanic/Latino1	Relative Income	-0.82	0.70	0.44	0.11	1.72	0.24	
Black2	Poverty Rate	0.02	0.01	1.02	0.99	1.04	0.22	
Hispanic/Latino2	Poverty Rate	0.03	0.02	1.03	0.99	1.06	0.10	
Black3	Unemployment Rate	8.39	2.38	4409.65	41.74	465808.93	0.00	***
Hispanic/Latino3	Unemployment Rate	-1.02	2.96	0.36	0.00	118.10	0.73	
Black4	Higher Education Share	3.26	0.90	26.03	4.44	152.67	0.00	***
Hispanic/Latino4	Higher Education Share	-1.14	1.29	0.32	0.03	4.00	0.38	

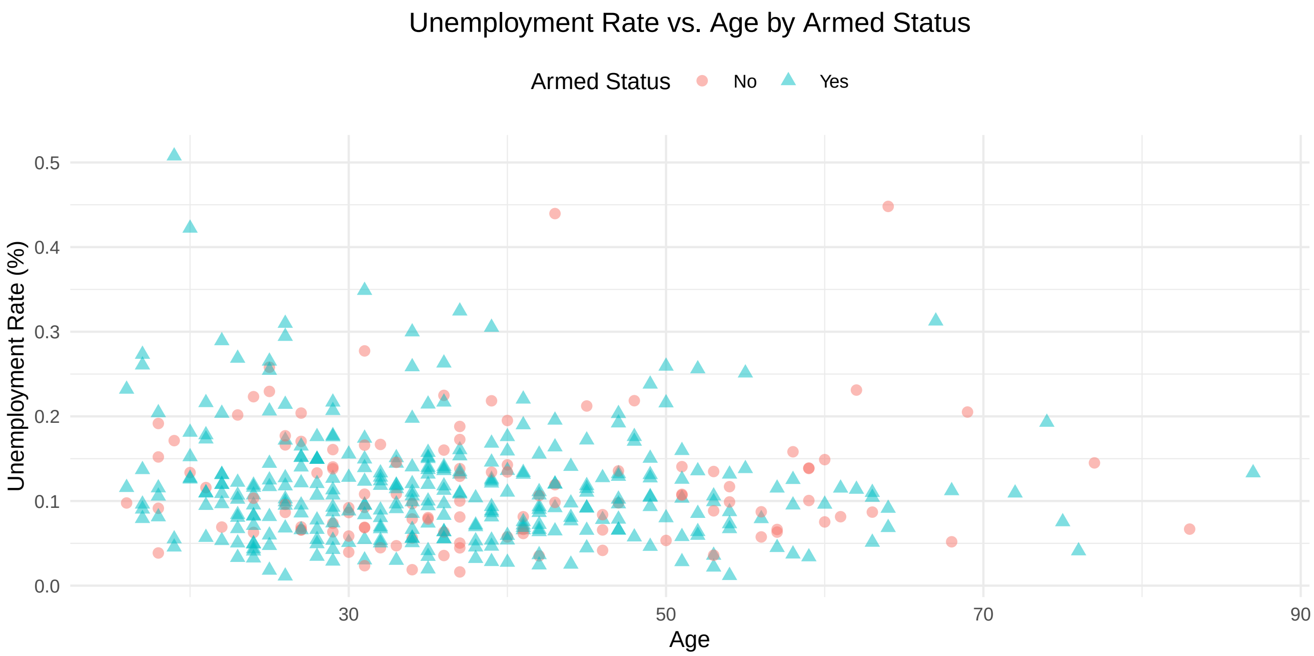
Research Question 2: Likelihood of Being Armed

Distribution of Armed Status

To better understand the prevalence of being armed, Appendix 1 summarizes the distribution of armed status. The analysis reveals that the majority of individuals in the dataset—73.96%—are armed, with only 26.04% classified as unarmed. This substantial difference underscores the predominance of armed individuals within the sample. This imbalance raises questions about the underlying reasons for such a distribution, whether cultural, legal, or related to other societal factors.

Relationship Between Armed Status, Age and Unemployment Rate

To take an initial look at the correlation between these three variables, we first plot a simple scatter plot between unemployment rate and age, colored by armed status.



The scatter plot reveals that most individuals are aged 20–50, with unemployment rates clustering below 20%. The "Yes" (armed) group is more evenly distributed across ages and leans toward lower unemployment rates, while the "No" (unarmed) group has slightly more cases with unemployment rates above 20%. Outliers above 40% unemployment are mostly in the armed group. Overall, no clear trend links age and unemployment, with patterns between armed and unarmed groups being similar.

The box plots, see Appendix 1, further compares age distributions for armed and unarmed individuals, showing similar ranges and medians. This suggests age is not a significant factor in determining armed status, aligning with regression results. The limited variability indicates other contextual factors may play a larger role.

Similarly, we explore the relationship between armed status and unemployment rate using a box plot with jittered points. The plot shows that the median and interquartile ranges of unemployment rates are similar for both groups. However, outliers, especially in the armed group, highlight individuals with unusually high unemployment rates. Despite this, the overall spread suggests unemployment rate, like age, shows little variability across armed status groups, indicating other factors likely play a greater role in determining armed status.

Logistic Regression

The earlier exploratory data analysis (EDA) found no strong relationships between age, unemployment rate (urate), and armed status, so a logistic regression model was used to formally test these relationships. To control for potential confounders, college education status and poverty levels (pov) were included, as they may influence both unemployment rates and armed status (see Appendix 2 for details).

The college variable captures higher education attainment, which is linked to lower unemployment rates and may mitigate social or economic stress factors tied to armed status. Poverty (pov) reflects regional poverty levels, potentially affecting armed status through economic insecurity or limited resources. Including these variables helps account for confounding effects and ensures observed relationships are not due to omitted variable bias.

The initial regression results are summarized below:

Table 3: Logistic Regression Results with Proper Variable Names (continued below)

	Estimate	Std. Error	Z value	P value
(Intercept)	0.512	1.413	0.3622	0.7172
Age	-0.004622	0.02838	-0.1629	0.8706
Unemployment Rate (%)	13.33	8.64	1.543	0.1227
College Completion (%)	1.166	3.198	0.3645	0.7155
Poverty Index	0.01906	0.04206	0.4531	0.6505
Age X Unemployment	-0.1457	0.1549	-0.9409	0.3468
Age X College	0.01473	0.06769	0.2177	0.8277
Age X Poverty	0.0001798	0.0008812	0.2041	0.8383
Unemployment X College	-30.59	17.32	-1.766	0.0774
Unemployment X Poverty	-0.1719	0.1114	-1.543	0.1228
College X Poverty	0.008119	0.07837	0.1036	0.9175

	OddsRatio	2.5%	97.5%
(Intercept)	1.669	0.1005	26.34
Age	0.9954	0.9419	1.054
Unemployment Rate (%)	618215	0.07395	3.906e+13
College Completion (%)	3.208	0.006702	2021
Poverty Index	1.019	0.9395	1.109
Age X Unemployment	0.8644	0.6281	1.163
Age X College	1.015	0.8878	1.16
Age X Poverty	1	0.9984	1.002
Unemployment X College	5.163e-14	1.165e-29	6.762
Unemployment X Poverty	0.8421	0.6662	1.042
College X Poverty	1.008	0.8727	1.186

The initial results indicate that age is not a significant predictor of armed status (OR = 0.995, 95% CI: 0.942–1.054, p = 0.8706), consistent with similar age distributions across groups. Unemployment rate shows a large but uncertain effect (OR = 618,215, 95% CI: 0.074–3.91e+13, p = 0.1227), suggesting a potential positive association, though not statistically significant. College completion (OR = 3.208, 95% CI: 0.007–2021, p = 0.7155) and poverty (OR = 1.019, 95% CI: 0.940–1.109, p = 0.6505) also lack significant individual effects, though they may control for confounding. Notably, the interaction between unemployment and college completion approaches significance (OR = 5.16e-14, 95% CI: 1.17e-29–6.76, p = 0.0774), suggesting that unemployment may more strongly impact those with lower education levels. Other interactions, such as age with unemployment or poverty, show weak and non-significant effects, reflecting a complex but largely inconclusive interplay between variables.

To better isolate the individual effects from the interaction terms, we've also calculated estimates from the implied model, see Appendix 2. At an unemployment rate (urate) of 0.1167, with age = 37.08, college = 0.2216, and pov = 21.11, the model predicts a log-odds of about 1.012 for "armed = 1." This corresponds to a roughly 73% chance, with a 68% to 78% confidence range. These estimates, derived from the implied model, provide insights into the combined effects of predictors and their interactions, offering a clearer understanding of the outcome at specific predictor levels. Yet, findings are largely inconclusive, and highlight the need for further investigation to clarify these relationships.

Refitting the Model

The initial diagnostic plots in Appendix 2 reveal issues with the model fit. The Residuals vs. Fitted plot shows non-random patterns, indicating violations of independence. The Q-Q plot highlights deviations from normality, particularly in the tails, which, though less critical for logistic regression, can still impact performance. The Scale-Location plot suggests heteroscedasticity, and Cook's Distance identifies influential points that may skew results, thought heteroscedasticity is not an assumption for logistic regression.

After removing observations with high Cook's Distance and refitting the model, see Appendix 3, no predictors were significant: age (p = 0.9417), unemployment rate (p = 0.3164), college education (p = 0.4203), and poverty (p = 0.2609). Interaction terms also failed to produce meaningful effects. While residual deviance slightly improved (416.8 on 396 degrees of freedom), overall performance remained unchanged, and infact led to the model producing "yes" predictions unanimously.

Conclusion

The key takeaway from this analysis is that **age**, **unemployment rate (urate)**, **college education status**, and **poverty levels (pov)** do not significantly predict the likelihood of being armed in this dataset. Despite exploring interaction effects and accounting for potential confounding factors, the logistic regression model yielded no statistically significant results. While the **unemployment rate** showed a positive but inconclusive effect and the **Urate × College** interaction approached significance, these relationships remain weak and require further investigation. Additionally, for the factors associated with racial disparities in police-related incidents in the United States. Counties with **lower relative income** were more likely to have individuals categorized as *Black* rather than *White*. Additionally, counties with a **higher share of college-educated residents** were associated with a lower likelihood of individuals being classified as *Hispanic/Latino*.

This study is not without limitations. The dataset may lack key explanatory variables that capture the broader socioeconomic or contextual factors influencing armed status. Correlation between several economic variables, which may have influenced the observed results. Addressing economic inequalities and exploring additional contextual factors can provide further insights into reducing these disparities. Additionally, the relatively small sample size and presence of influential observations may have limited the model's ability to detect subtle relationships.

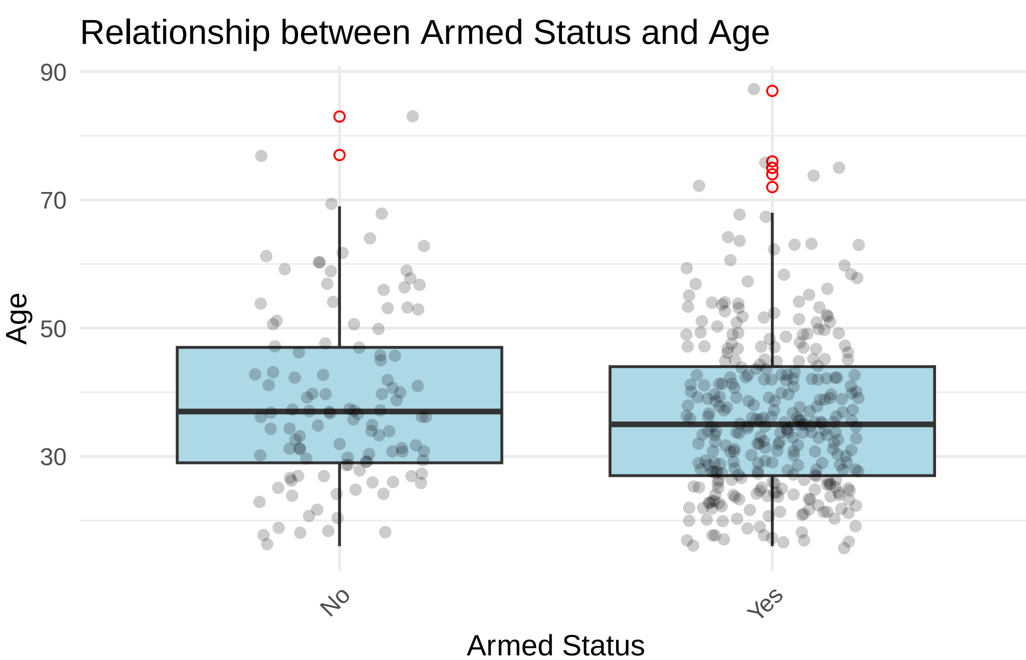
Future work in this area can focus on incorporating additional predictors, such as crime rates, regional policies, or mental health indicators, which may provide a more comprehensive understanding of the factors influencing armed status. Expanding the dataset and exploring advanced modeling techniques, such as mixed-effects models or machine learning approaches, could further enhance predictive power and uncover nuanced relationships. Additionally, future research could evaluate the impact of specific policy interventions, such as economic revitalization programs, community policing initiatives, or education-based reforms, to assess their effectiveness in addressing these disparities. Examining temporal trends using longitudinal data would help determine whether racial disparities in police-related incidents are improving, worsening, or remaining stable over time in response to broader social and economic changes. By expanding the scope of this analysis and addressing these limitations, future research can provide a more comprehensive understanding of the systemic factors that contribute to racial disparities in police-related incidents and drivers behind armed status.

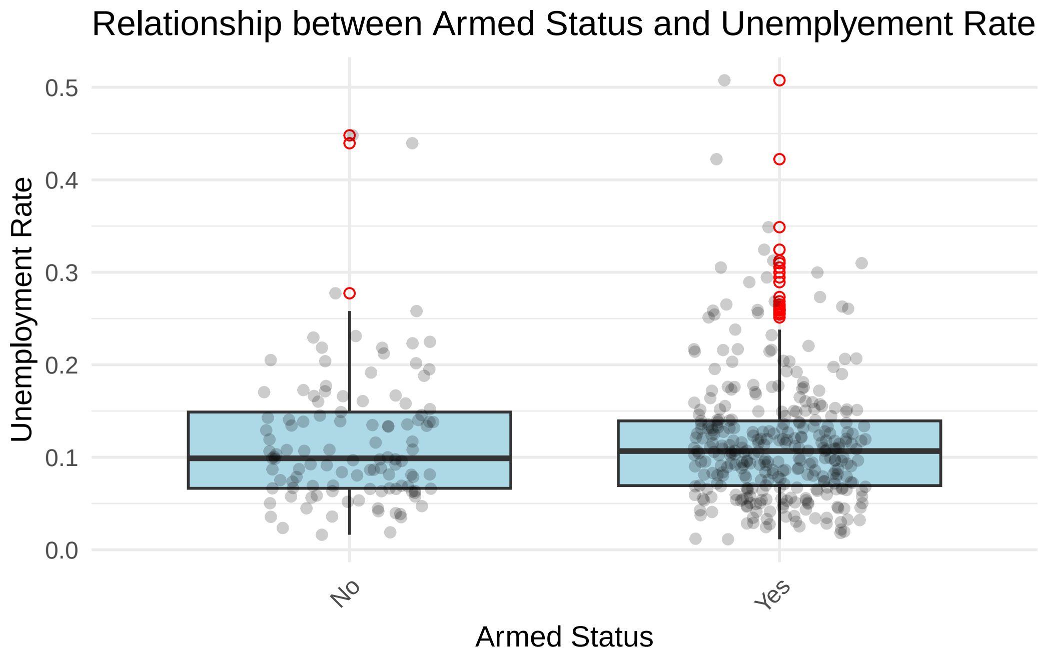
Appendix

1. Research Question 2 EDA

Table 5: Armed Status Distribution

Armed_Status	Count	Percentage
No	113	26.04
Yes	321	73.96





2. Research Question 2 Initial Model

Table 6: Logistic Regression Results with Proper Variable Names (continued below)

	Estimate	Std. Error	Z value	P value
(Intercept)	0.512	1.413	0.3622	0.7172
Age	-0.004622	0.02838	-0.1629	0.8706
Unemployment Rate (%)	13.33	8.64	1.543	0.1227
College Completion (%)	1.166	3.198	0.3645	0.7155
Poverty Index	0.01906	0.04206	0.4531	0.6505
Age X Unemployment	-0.1457	0.1549	-0.9409	0.3468
Age X College	0.01473	0.06769	0.2177	0.8277
Age X Poverty	0.0001798	0.0008812	0.2041	0.8383
Unemployment X College	-30.59	17.32	-1.766	0.0774
Unemployment X Poverty	-0.1719	0.1114	-1.543	0.1228
College X Poverty	0.008119	0.07837	0.1036	0.9175

	OddsRatio	2.5%	97.5%
(Intercept)	1.669	0.1005	26.34
Age	0.9954	0.9419	1.054
Unemployment Rate (%)	618215	0.07395	3.906e+13
College Completion (%)	3.208	0.006702	2021
Poverty Index	1.019	0.9395	1.109
Age X Unemployment	0.8644	0.6281	1.163
Age X College	1.015	0.8878	1.16
Age X Poverty	1	0.9984	1.002
Unemployment X College	5.163e-14	1.165e-29	6.762

	OddsRatio	2.5%	97.5%
Unemployment X Poverty	0.8421	0.6662	1.042
College X Poverty	1.008	0.8727	1.186

Table 8: Table continues below

age	urate	college	pov	age:urate	age:college	age:pov
11.29	27.07	20.02	24.45	19.2	15.1	19.35

urate:college	urate:pov	college:pov
4.496	13.97	4.098

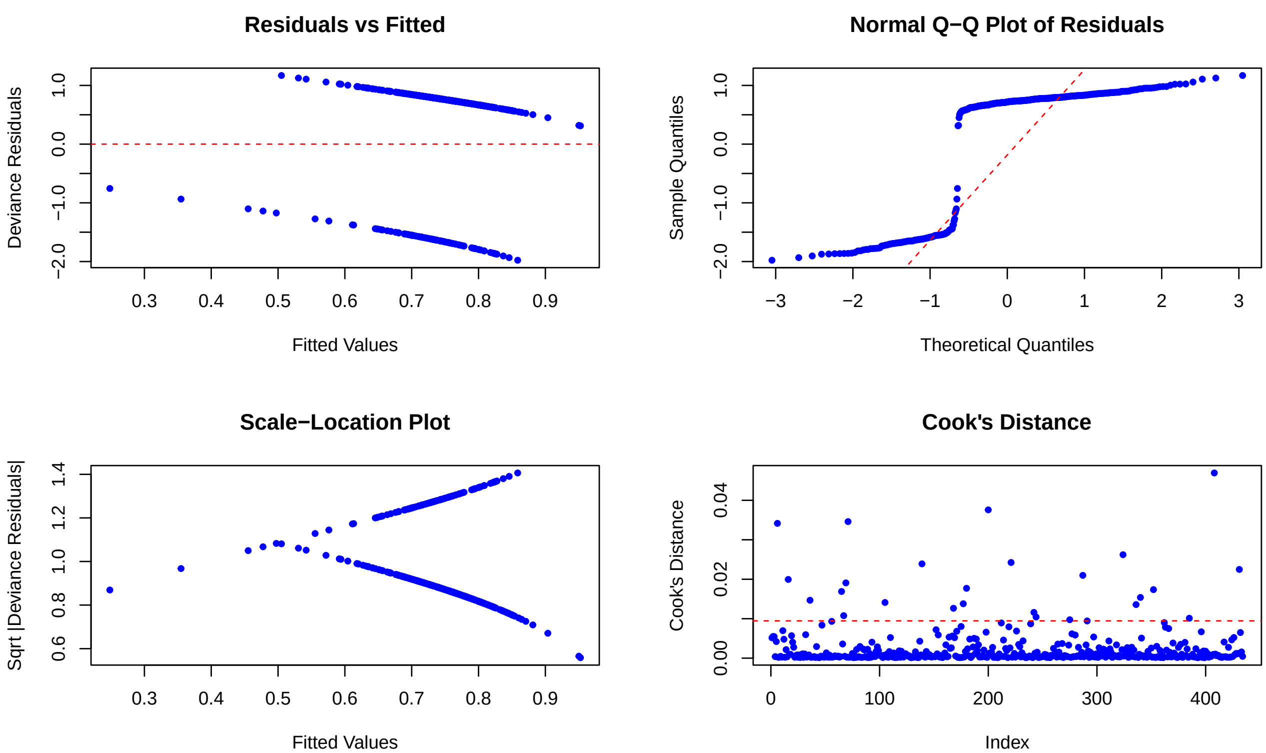


Table 10: Confusion Matrix (Predicted vs. Actual)

Actual	Predicted: No	Predicted: Yes
No	5	108
Yes	0	321

Table 11: Overall Statistics (continued below)

Accuracy	Kappa	AccuracyLower	AccuracyUpper	AccuracyNull
0.7512	0.06409	0.7077	0.7911	0.7396

AccuracyPValue	McnemarPValue
0.3137	7.339e-25

Table 13: Estimates from Implied Model (Age)

age	urate	college	pov	emmean	SE	df	asymp.LCL	asymp.UCL
37.08	0.1167	0.2216	21.11	1.012	0.1354	Inf	0.7464	1.277

Table 14: Estimates from Implied Model (Unemployment)

urate	age	college	pov	emmean	SE	df	asymp.LCL	asymp.UCL
0.1167	37.08	0.2216	21.11	1.012	0.1354	Inf	0.7464	1.277

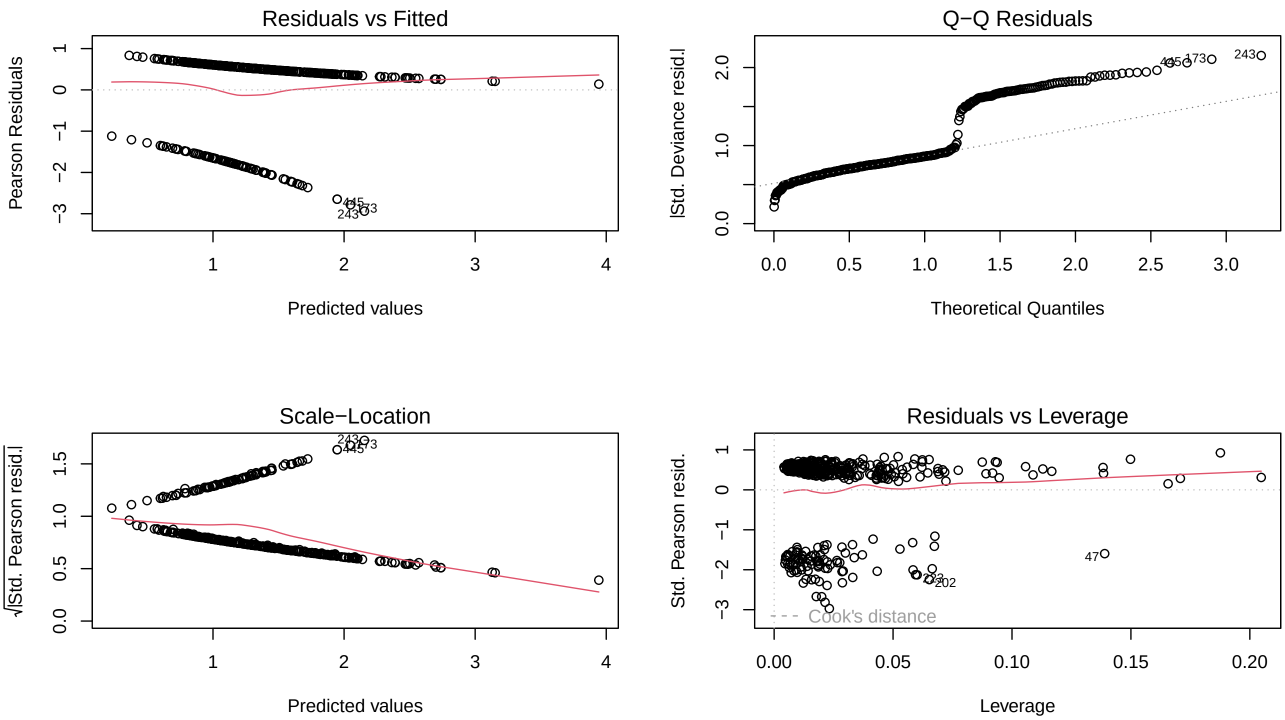
3. Research Question 2 Refitted Model

Table 15: Logistic Regression Results with Proper Variable Names (continued below)

	Estimate	Std. Error	Z value	P value
(Intercept)	-0.5439	1.816	-0.2995	0.7646
Age	-0.002722	0.03719	-0.07319	0.9417
Unemployment Rate (%)	11.46	11.44	1.002	0.3164
College Completion (%)	3.487	4.327	0.8058	0.4203
Poverty Index	0.06767	0.0602	1.124	0.2609
Age X Unemployment	-0.1411	0.2333	-0.6048	0.5453
Age X College	0.03719	0.09468	0.3929	0.6944
Age X Poverty	9.033e-05	0.001267	0.07129	0.9432
Unemployment X College	-24.6	26.24	-0.9377	0.3484
Unemployment X Poverty	-0.1875	0.1962	-0.9559	0.3391
College X Poverty	-0.07108	0.1107	-0.6422	0.5207

	OddsRatio	2.5%	97.5%
(Intercept)	0.5805	0.01614	20.68
Age	0.9973	0.9272	1.074
Unemployment Rate (%)	94916	4.118e-05	1.676e+15
College Completion (%)	32.69	0.009132	239437
Poverty Index	1.07	0.952	1.205

	OddsRatio	2.5%	97.5%
Age X Unemployment	0.8684	0.5417	1.357
Age X College	1.038	0.8617	1.252
Age X Poverty	1	0.9976	1.003
Unemployment X College	2.068e-11	2.259e-35	2.781e+10
Unemployment X Poverty	0.829	0.5695	1.239
College X Poverty	0.9314	0.7613	1.172



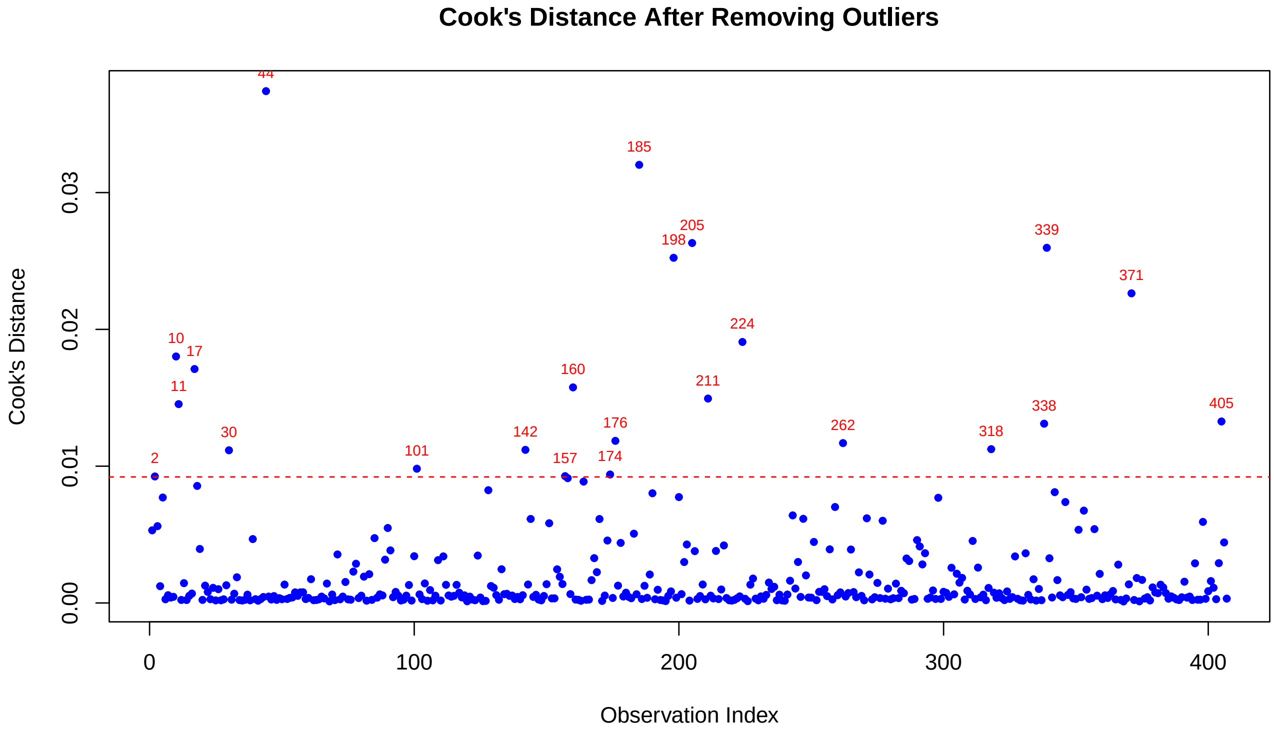


Table 17: Confusion Matrix (Predicted vs. Actual)

Actual	Predicted: No	Predicted: Yes
No	0	90
Yes	0	317

Table 18: Overall Statistics (continued below)

Accuracy	Kappa	AccuracyLower	AccuracyUpper	AccuracyNull
0.7789	0	0.7354	0.8183	0.7789

AccuracyPValue	McnemarPValue
0.5282	6.509e-21

Table 20: Estimates from Implied Model (Age)

age	urate	college	pov	emmean	SE	df	asymp.LCL	asymp.UCL
36.65	0.1139	0.2151	20.57	1.219	0.1554	Inf	0.9147	1.524

Table 21: Estimates from Implied Model (Unemployment)

urate	age	college	pov	emmean	SE	df	asymp.LCL	asymp.UCL
0.1139	36.65	0.2151	20.57	1.219	0.1554	Inf	0.9147	1.524