

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 referred 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. The dataset also contains demographic proportions, such as **share_white**, **share_black**, and **share_hispanic**, representing the racial composition within the county.

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. Additionally, the outcome

variable was structured to focus on three primary racial groups: “White,” “Black,” and “Hispanic/Latino,” with “White” designated as the reference category. This decision was made due to the disproportionately low representation of individuals categorized as “Asian/Pacific Islander” and “Native American,” which would have hindered the reliability of the multinomial logistic regression analysis.

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

[## Need to remove these as per feedback from Andrea, I couldn’t figure out how to include these below in above content]

Research Question 1: Economic Conditions and Racial Composition

- **Outcome Variable:** raceethnicity (Nominal) – categorizes the racial/ethnic group of the deceased.
- **Independent Variables:**

- **comp_income**: A measure of relative income (household income divided by county income), indicating economic status.
- **pov**: Poverty level in the area, providing additional context on socio-economic conditions.

A multinomial logistic regression model was fitted to examine the association:

- `raceethnicity ~ comp_income + pov`

Research Question 2: Likelihood of Being Armed

- **Outcome Variable**: **armed** (Binary) – indicates whether the deceased was armed at the time of the incident.
- **Independent Variables**:
 - **age**: Age of the deceased (continuous variable).
 - **urate**: Unemployment rate in the area, representing socio-economic stress.
- **Interaction Term**: An interaction between **age** and **urate** to assess their combined influence on the likelihood of individuals being armed.

A logistic regression model was fitted:

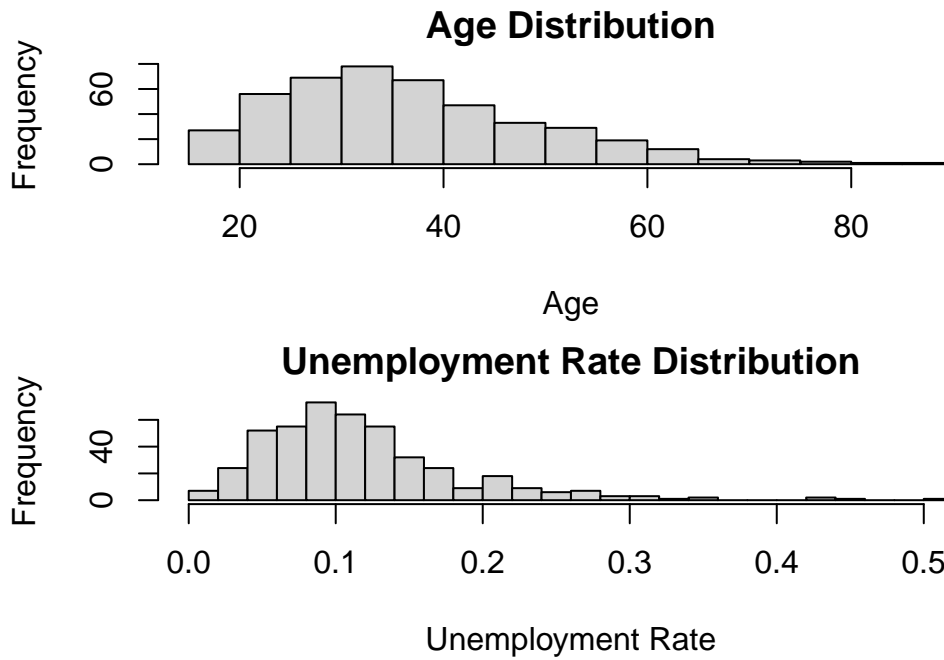
- `armed ~ age + urate + age:urate`

Model Assessment

Model diagnostics included checks for multicollinearity using Variance Inflation Factor (VIF) and assessment of influential points via Cook's distance. For the multinomial model, pseudo- R^2 and AIC measures were calculated. For the logistic model, residual patterns were assessed for misfit.

Results

Exploratory Data Analysis

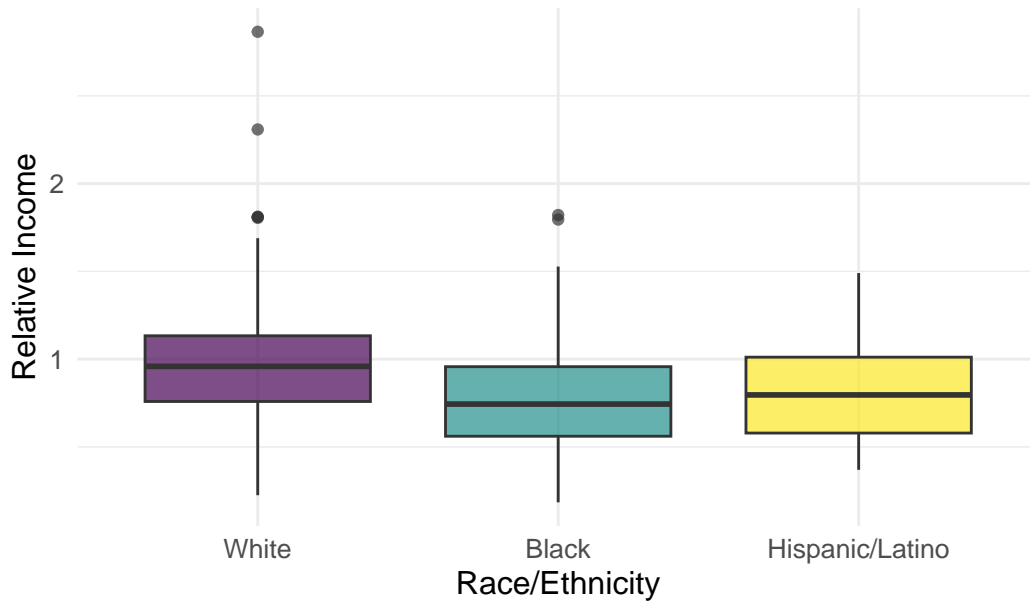


The demographic analysis shows that the deceased had a mean age in their 30s, with the age range extending to over 80. The unemployment rate in their residential tracts averaged 10%, reaching a maximum of 51%. The gender distribution was heavily skewed, with 428 males and 20 females.

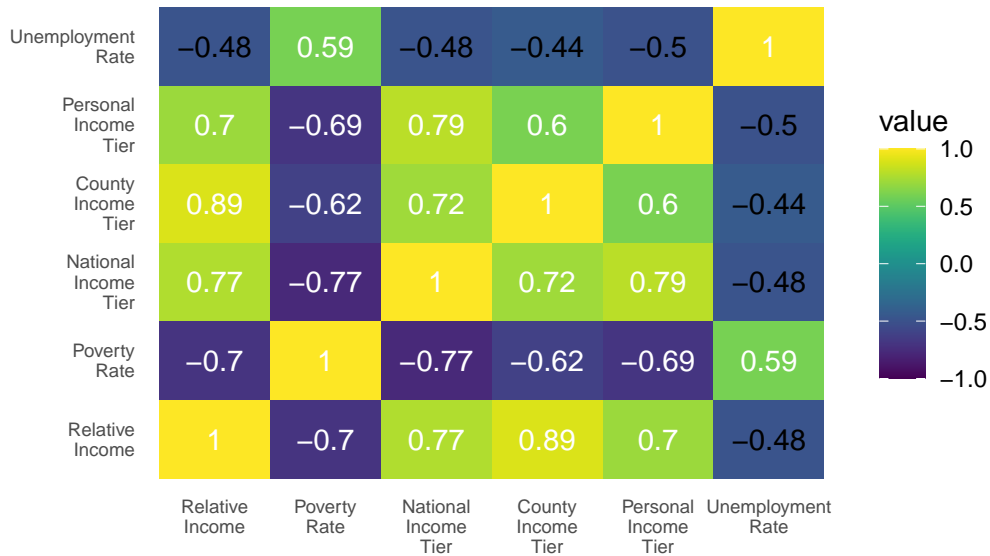
Table 1: Race Ethnicity Summary Statistics

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

Economic Conditions and Racial Composition



Correlation Plot of Economic Related Indicators



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 final multinomial logistic regression model are summarized in the table below. The predictors included in the model were chosen based on their relevance to the research question.

The findings reveal important relationships between economic conditions and racial classifications in police-related incidents. Counties with lower **relative income** (a measure of household income as a share of the county median) were more likely to see individuals involved in these incidents categorized as “Black” rather than “White.” Similarly, counties with higher **unemployment rates** were associated with an increased likelihood of individuals being classified as “Black” compared to “White.” These findings highlight the disproportionate impact of economic inequality and joblessness on Black individuals in the context of police-related incidents.

For “Hispanic/Latino” individuals, the relationships with economic variables were less consistent. For example, relative income showed no significant relationship with the likelihood of individuals being classified as “Hispanic/Latino” versus “White.” However, counties with slightly higher **poverty rates** were marginally more likely to have individuals categorized as “Hispanic/Latino” rather than “White,” though the effect was small. Counties with a higher proportion of **college-educated residents** were associated with a reduced likelihood of individuals being classified as “Hispanic/Latino” compared to “White,” suggesting the potential role of education in shaping these patterns.

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.

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, Table 7 and the accompanying bar chart summarize the distribution of armed status. The analysis reveals that the majority of individuals in the dataset—78.09%—are armed, with only 21.91% classified as unarmed. This substantial difference underscores the predominance of armed individuals within the sample. The bar chart visually reinforces this finding, with the “Yes” category towering over the “No” category. This imbalance raises questions about the underlying reasons for such a distribution, whether cultural, legal, or related to other societal factors.

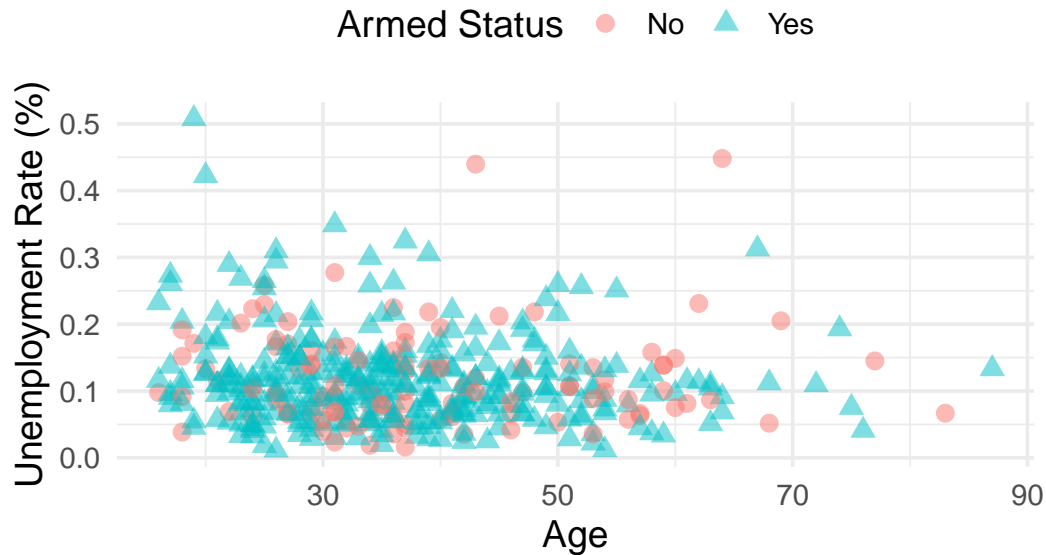
Table 3: Armed Status Distribution

Armed_Status	Count	Percentage
No	113	26.04
Yes	321	73.96

Relationship Between Armed Status, Age and Unemployment Rate

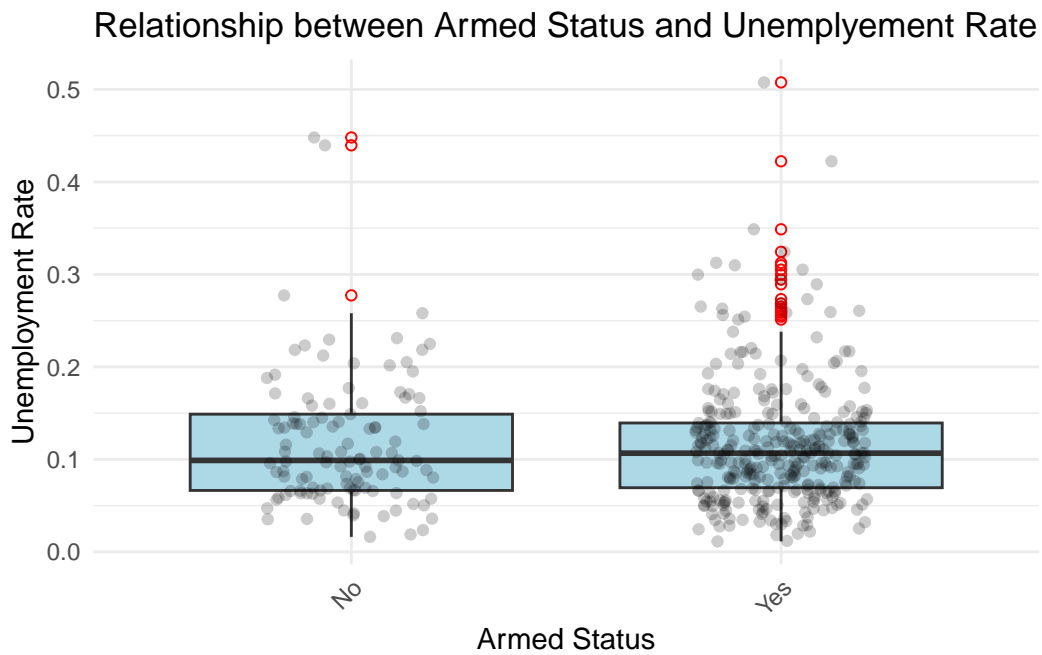
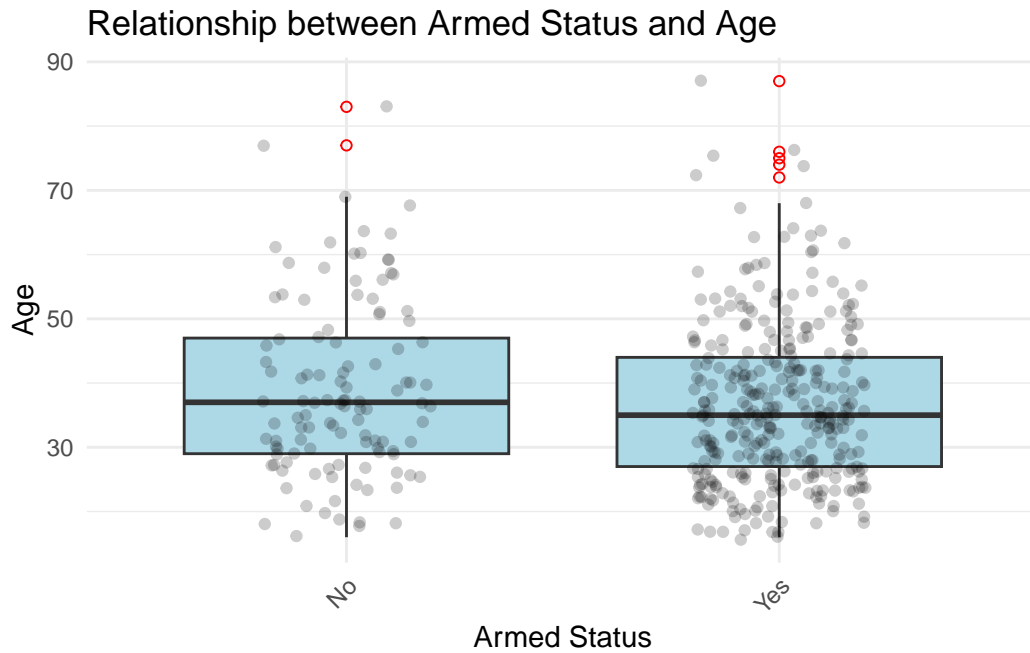
To take an initial look at the correlation between these three variables, we first plot a simple scatter plot between unemployment rate and age, colored by armed status.

catter Plot of Unemployment Rate vs. Age by Armed



The scatter plot shows, firstly most individuals are between 20 and 50 years old, with unemployment rates clustering below 20%. The “Yes” (armed) group is more evenly distributed across ages and slightly concentrated at lower unemployment rates, while the “No” (unarmed) group shows a few more cases of elevated unemployment rates above 20%. Outliers with unemployment rates exceeding 40% appear primarily in the armed group. Overall, no clear trend emerges between age and unemployment rate, and the patterns between armed and unarmed groups are broadly similar.

We further explore the relationship between age and armed status through a box plot, which compares the age distribution for armed and unarmed individuals. The plot illustrates that the age ranges and medians for both groups are remarkably similar, suggesting no significant difference in age distribution between those who are armed and those who are not. This visual evidence supports the regression findings, reinforcing the conclusion that age alone is not a strong determinant of armed status. Nonetheless, the lack of significant variability in the data suggests that other contextual factors beyond age may play a more crucial role in influencing armed status.



In a similar manner, we can take a deeper dive into the relationship between armed status and unemployment rate by plotting a box plot with jittered points for both groups. The visualization demonstrates that the median and interquartile range of unemployment rates are relatively similar for armed and unarmed individuals. However, the presence of outliers, partic-

ularly for the armed group, suggests that there may be certain individuals with unusually high unemployment rates. Despite this, the overall spread of the data indicates that unemployment rate, much like age, does not exhibit strong variability across armed status groups. This lack of pronounced difference implies that other contextual factors beyond unemployment rate likely play a more significant role in influencing armed status.

Logistic Regression

Motivation and Model Design

While the earlier exploratory data analysis (EDA) did not reveal pronounced relationships between **age**, **unemployment rate (urate)**, and **armed status**, a logistic regression model was employed to formally quantify these relationships and test for statistical significance. To account for potential confounding effects, **college education status** and **poverty levels (pov)** were included in the model. These variables are theoretically relevant, as they may influence socioeconomic conditions that impact both unemployment rates and armed status.

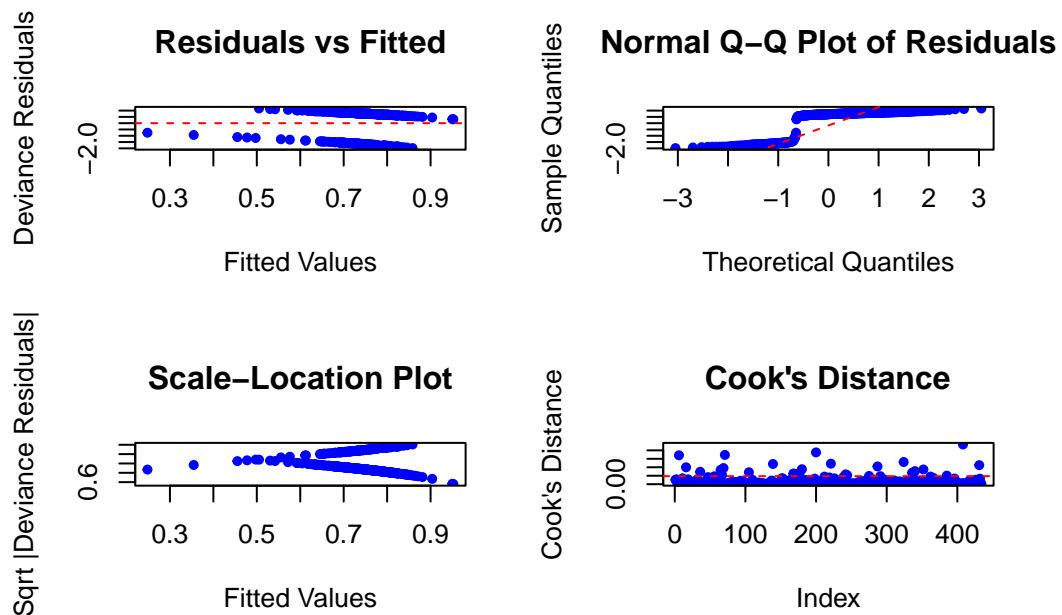
The **college** variable represents whether individuals have attained higher education. Lower education levels are often associated with higher unemployment rates, which can indirectly contribute to social or economic stress factors that correlate with armed status. Similarly, **poverty (pov)** measures the proportion of individuals in poverty within a region. Poverty may directly or indirectly influence armed status through economic insecurity, limited access to resources, or broader community-level factors. By including these variables, the model adjusts for their potential confounding effects and ensures that any observed relationships between age, unemployment rate, and armed status are not falsely attributed due to omitted variable bias.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.512	1.413	0.3622	0.7172
age	-0.004622	0.02838	-0.1629	0.8706
urate	13.33	8.64	1.543	0.1227
college	1.166	3.198	0.3645	0.7155
pov	0.01906	0.04206	0.4531	0.6505
age:urate	-0.1457	0.1549	-0.9409	0.3468
age:college	0.01473	0.06769	0.2177	0.8277
age:pov	0.0001798	0.0008812	0.2041	0.8383
urate:college	-30.59	17.32	-1.766	0.0774
urate:pov	-0.1719	0.1114	-1.543	0.1228
college:pov	0.008119	0.07837	0.1036	0.9175

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	497.7 on 433 degrees of freedom
Residual deviance:	486.7 on 423 degrees of freedom

age	urate	college	pov	age:urate
11.294619	27.067226	20.022708	24.452561	19.196168
age:college	age:pov	urate:college	urate:pov	college:pov
15.095022	19.352511	4.495927	13.967057	4.097839



- **positive:** Yes
- **table:**

	No	Yes
No	5	0
Yes	108	321

- **overall:**

Table 7: Table continues below

Accuracy	Kappa	AccuracyLower	AccuracyUpper	AccuracyNull
0.7512	0.06409	0.7077	0.7911	0.7396

AccuracyPValue	McnemarPValue
0.3137	7.339e-25

- **byClass:**

Table 9: Table continues below

Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision
1	0.04425	0.7483	1	0.7483

Table 10: Table continues below

Recall	F1	Prevalence	Detection Rate	Detection Prevalence
1	0.856	0.7396	0.7396	0.9885

Balanced Accuracy
0.5221

- **mode:** everything
- **dots:**

Key Results

The regression results are summarized below:

Table 12: Regression Results

Variable	Estimate	Std. Error	z-value	p-value
Intercept	0.512	1.413	0.3622	0.7172

Variable	Estimate	Std. Error	z-value	p-value
Age of Individuals	-0.0046	0.0284	-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 Rate	0.0191	0.0421	0.4531	0.6505
Age \times Unem- ployment Rate	-0.1457	0.1549	-0.9409	0.3468
Unemployment \times College	-30.59	17.32	-1.766	0.0774

Interpretation of Results

The results show that **age** does not significantly predict armed status ($p = 0.8706$), consistent with earlier findings where age distributions were similar across groups. While the **unemployment rate (urate)** has a positive relationship with armed status (Estimate = 13.33), it is not statistically significant ($p = 0.1227$), suggesting a possible but inconclusive link.

Both **college education** and **poverty levels (pov)** have weak individual effects ($p > 0.65$), but their inclusion helps account for confounding factors. Interestingly, the interaction term **Urate \times College** is close to significance ($p = 0.0774$), hinting that unemployment may have a stronger impact on individuals with lower education levels.

Overall, interactions like **Age \times Urate** and **Urate \times College** suggest that age, education, and poverty interact in complex ways, shaping how unemployment relates to armed status.

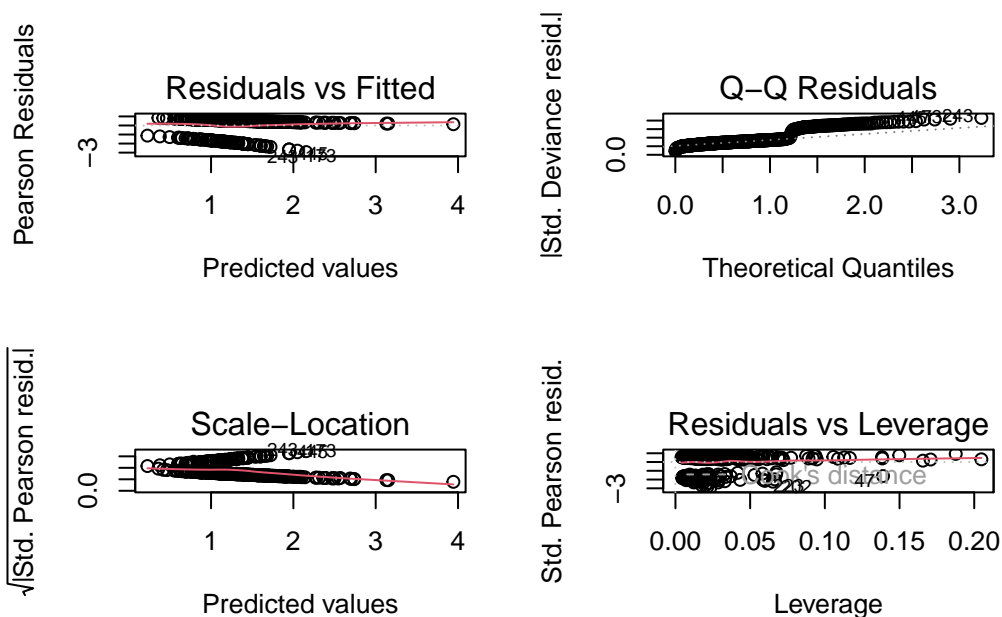
Model Diagnostics

The diagnostic plots reveal several issues with the current model fit. In the **Residuals vs. Fitted plot**, the residuals display a non-random pattern, suggesting violations of the assumptions of independence and homoscedasticity, indicating that the model may not fully capture the data's underlying structure. The **Normal Q-Q plot** shows deviations from the diagonal line, particularly at the tails, pointing to non-normally distributed residuals. While normality is less critical for logistic regression, extreme deviations can still affect the model's performance. Additionally, the **Scale-Location plot** highlights an uneven spread of residuals, reinforcing the presence of heteroscedasticity. Finally, **Cook's Distance** highlights influential points that may skew the model's results, making it necessary to address them for improved reliability.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.5439	1.816	-0.2995	0.7646
age	-0.002722	0.03719	-0.07319	0.9417
urate	11.46	11.44	1.002	0.3164
college	3.487	4.327	0.8058	0.4203
pov	0.06767	0.0602	1.124	0.2609
age:urate	-0.1411	0.2333	-0.6048	0.5453
age:college	0.03719	0.09468	0.3929	0.6944
age:pov	9.033e-05	0.001267	0.07129	0.9432
urate:college	-24.6	26.24	-0.9377	0.3484
urate:pov	-0.1875	0.1962	-0.9559	0.3391
college:pov	-0.07108	0.1107	-0.6422	0.5207

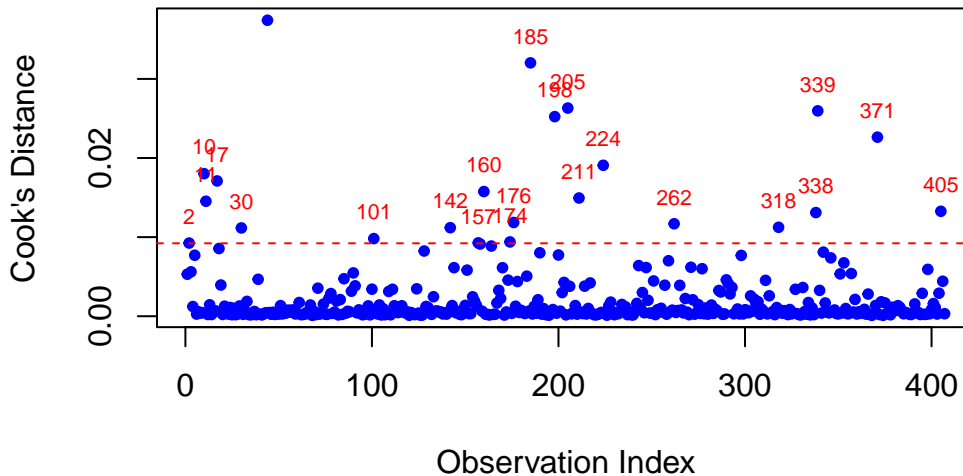
(Dispersion parameter for binomial family taken to be 1)

Null deviance:	430.1 on 406 degrees of freedom
Residual deviance:	416.8 on 396 degrees of freedom



[1] "Results remain insignificant; no significant relationships observed."

Cook's Distance After Removing Outliers



The initial diagnostic plots showed some issues with our model, like non-random residual patterns and influential points flagged by Cook's Distance. To fix this, we removed observations with Cook's Distance values above the threshold $((4/n))$ and refitted the logistic regression model using the cleaned data.

After refitting, the results still didn't show any significant predictors. **Age** ($p = 0.9417$), **unemployment rate (urate)** ($p = 0.3164$), **college education** ($p = 0.4203$), and **poverty (pov)** ($p = 0.2609$) all remained insignificant. Even the interaction terms, like **age:urate** and **urate:college**, didn't produce meaningful effects. Although the residual deviance improved slightly (416.8 on 396 degrees of freedom), the overall model performance didn't change much.

In summary, removing outliers didn't uncover any strong relationships between these predictors and armed status. It seems these variables, on their own or combined, don't explain much in this dataset. Exploring other variables or alternative modeling approaches might give us better insights.

+++++

[TODO: Comment on the model fit etc., talk about the model assumptions, and try refitting the model after excluding outliers based on cook's distance Logic: 1. We see that the variables are not significant -> comment on that 2. However, we also see that several model assumptions might have been violated based on the diagnostic plot 3. Refit the model 4. -> if results are significant, write an interpretation saying that we found significant correlation -> else: write how the results are still not significant, and there's no significant correlation between the variables etc.]

+++++

Conclusion (only includes research Q2 conclusion, content may need to be edited to reduce word limit after adding Research Q1 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.