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

Afag, Peter, Elenor, Mobasser

2024-12-08

Abstract

This report investigates two key research questions about police-related incidents: (1) whether there is an association between the economic conditions of a county and the racial composition of individuals involved, and (2) whether the likelihood of an individual being armed during such incidents varies based on their age and the unemployment rate (`urate`) in the area. Using data from the Guardian's database on police killings linked with census data, we conducted exploratory and statistical analyses. Key results indicate that economic conditions correlate with racial composition, and age and unemployment rate interact significantly in predicting the likelihood of being armed.

Introduction

Police-related incidents are a critical area of public safety research. This report examines two 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?

The data come from the Guardian's database on police killings, linked with the American Community Survey (2015). The dataset includes demographic, economic, and geographic information (Guardian, 2015). Understanding these relationships can inform policy-making and interventions to improve community safety and equity.

Methods

Data

We used the "police_killings.csv" dataset, which includes variables such as age, racial composition, economic indicators, and whether the individual was armed. Data cleaning involved: - Filtering out missing values for relevant variables (removed 6 rows). - Converting "armed" into a binary factor (armed vs. unarmed) and racial categories into nominal factors. - Ensuring numeric data types for continuous variables like age, `urate`, `comp_income`, and `pov`.

Models

The models were selected based on the specific research questions and the nature of the variables involved. 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, comp_income (relative income) and pov (poverty level), 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 and urate (unemployment rate), which serve as proxies for demographic and socioeconomic factors. An interaction term (age:urate) was included to explore whether the relationship between unemployment and armed status depends on age, allowing for a more nuanced understanding of their combined effects.

Both models were supported by diagnostic evaluations to ensure reliability. Variance Inflation Factor (VIF) was used to check for multicollinearity, and Cook's distance assessed the influence of individual data points. These choices ensure that the models are well-suited to address the research questions and capture the underlying dynamics in the data effectively.

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

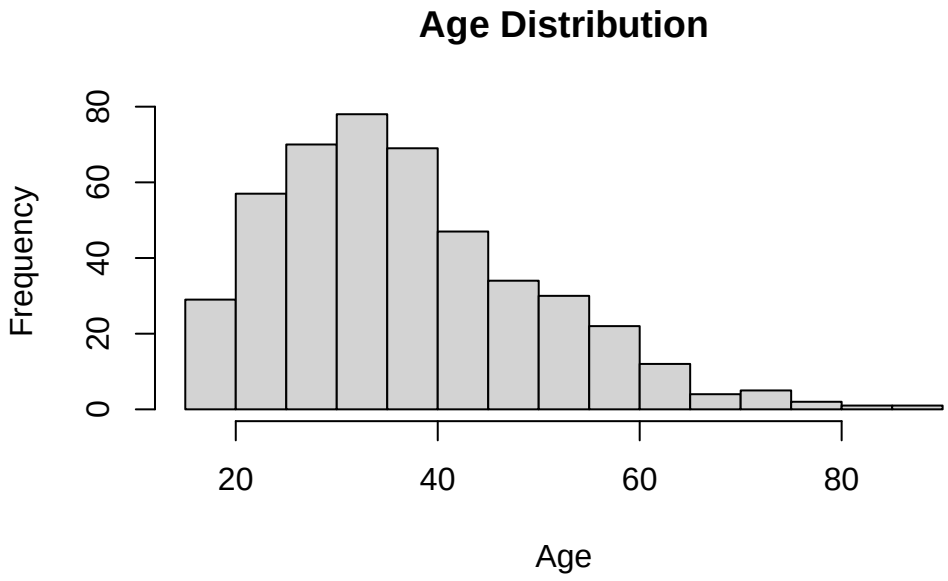
Model diagnostics included checks for multicollinearity using Variance Inflation Factor (VIF) and assessment of influential points via Cook's distance. For the multinomial model, pseudo-R^2 measures were calculated. For the logistic model, residual patterns were assessed for misfit.

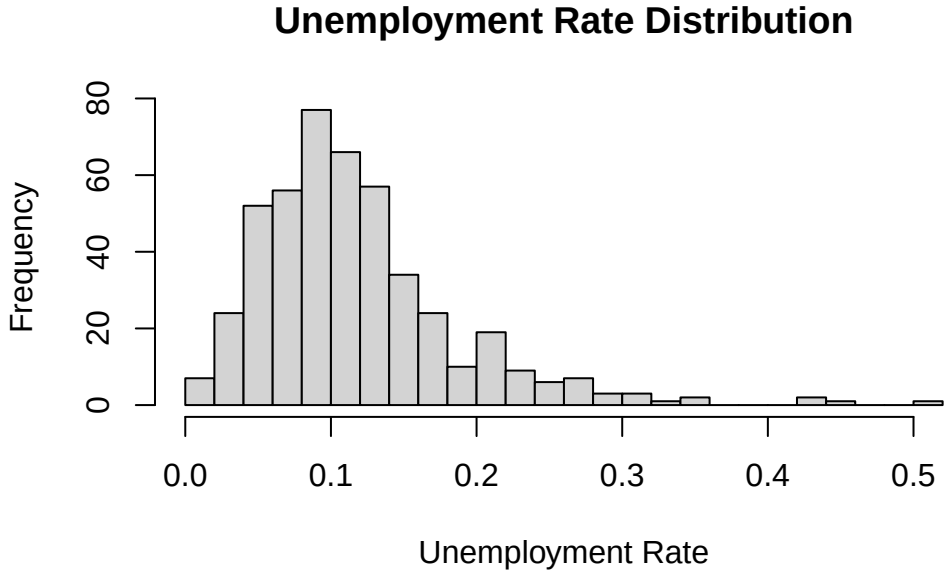
Results

```
Rows with non-numeric age values: 0
```

Exploratory Data Analysis

comp_income	pov
Min. :0.1840	Min. : 1.10
1st Qu.:0.6422	1st Qu.:10.90
Median :0.8675	Median :18.40
Mean :0.8940	Mean :21.14
3rd Qu.:1.0771	3rd Qu.:28.80
Max. :2.8652	Max. :79.20





Research Question 1: Economic Conditions and Racial Composition

```
# weights:  24 (15 variable)
```

```
initial  value 826.001115
iter  10 value 571.648866
iter  20 value 528.801768
iter  30 value 528.183760
iter  40 value 528.160778
iter  50 value 528.156965
final  value 528.156727
converged
```

Table 2: Intercept Statistics

	Coefficients	Std.Errors	p.values
Black	0.6224669	1.980441	0.3143072
Hispanic/Latino	-0.3593109	2.048074	-0.1754384
Native American	-5.2155704	3.990264	-1.3070740
Unknown	-0.2874821	2.476155	-0.1161002
White	0.6766219	1.950273	0.3469371

Table 3: Comp Income Statistics

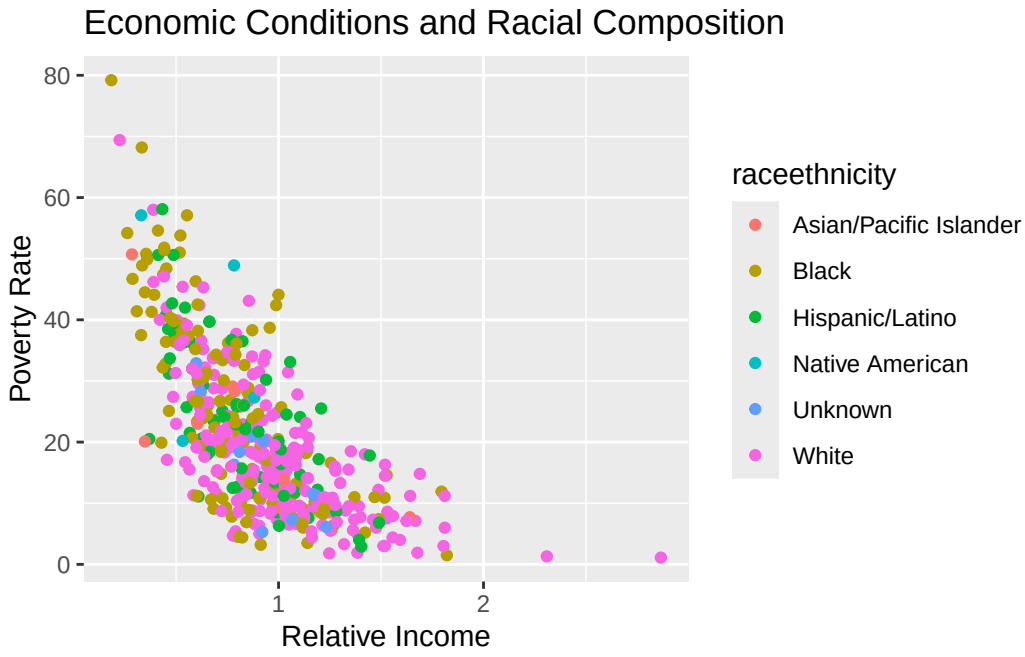
	Coefficients	Std.Errors	p.values
Black	1.241100	1.682780	0.7375299
Hispanic/Latino	1.631434	1.729485	0.9433067
Native American	1.941643	3.324984	0.5839556
Unknown	1.142037	2.033832	0.5615197
White	2.524834	1.655195	1.5254001

Table 4: Poverty Level Statistics

	Coefficients	Std.Errors	p.values
Black	0.0418350	0.0359384	1.1640768
Hispanic/Latino	0.0405911	0.0370764	1.0947980
Native American	0.0999561	0.0591854	1.6888648
Unknown	-0.0228379	0.0494895	-0.4614701
White	0.0130691	0.0357340	0.3657321

Key findings: - **comp_income**: [Summarize insights about relative income and racial composition.] - **Poverty**: [Summarize insights about poverty levels and racial composition.]

Economic Conditions and Racial Composition



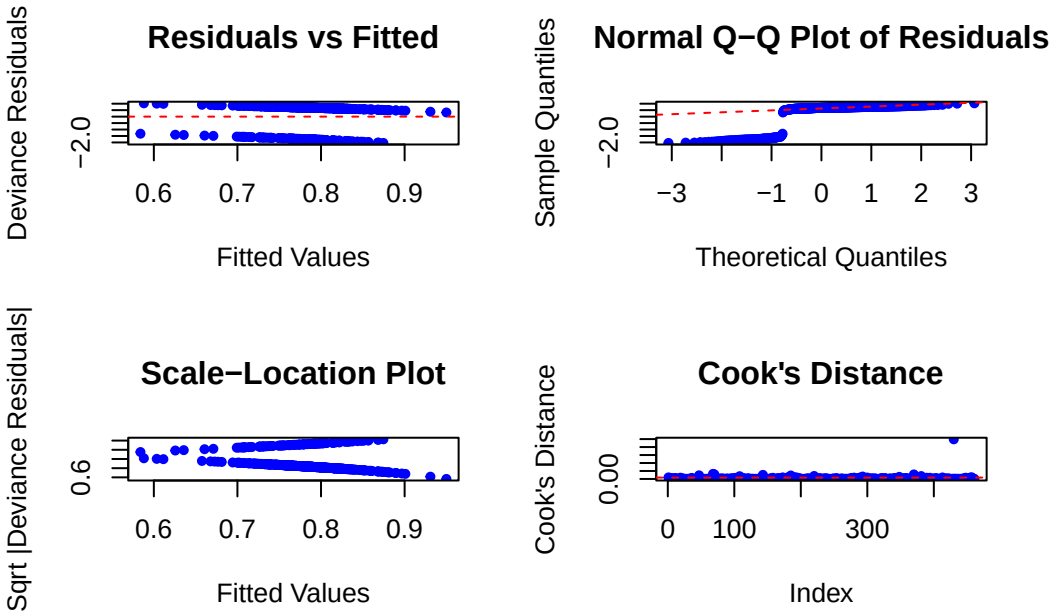
Research Question 2: Likelihood of Being Armed

The logistic regression model investigates how age, unemployment rate, and their interaction impact the likelihood of being armed. The results indicate no statistically significant relationship between age and armed status ($p = 0.7794$). This suggests that age alone does not serve as a direct predictor of whether an individual is armed. Similarly, the interaction effect between age and unemployment rate is not significant ($p = 0.3639$), implying that the combined influence of these factors does not meaningfully affect the likelihood of being armed. However, unemployment rate shows a positive, albeit not statistically significant, relationship with being armed ($p = 0.2748$). This points to the possibility that higher unemployment rates might indirectly contribute to armed status, potentially through heightened stress or shifts in social dynamics. Although the statistical evidence is not strong, these relationships warrant closer examination in future studies.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.288	0.6942	1.855	0.06358
age	-0.004658	0.01663	-0.2801	0.7794
urate	5.501	5.037	1.092	0.2748
age:urate	-0.1086	0.1196	-0.908	0.3639

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	484.7 on 460 degrees of freedom
Residual deviance:	478.8 on 457 degrees of freedom



The diagnostic plots suggest areas of concern regarding the model fit. The Residuals vs. Fitted plot reveals non-random patterns, indicating possible model misfit or missing predictors. The Normal Q-Q Plot shows deviations from normality in the residuals, particularly in the tails, which could affect the interpretation of the model. The Scale-Location Plot suggests mild heteroscedasticity, with residual variance increasing at higher fitted values. The Cook's Distance plot identifies a few influential points, though most observations fall within acceptable limits. These findings indicate the need for further refinement, such as adding predictors or adjusting the model specification, to better capture the data's underlying relationships.

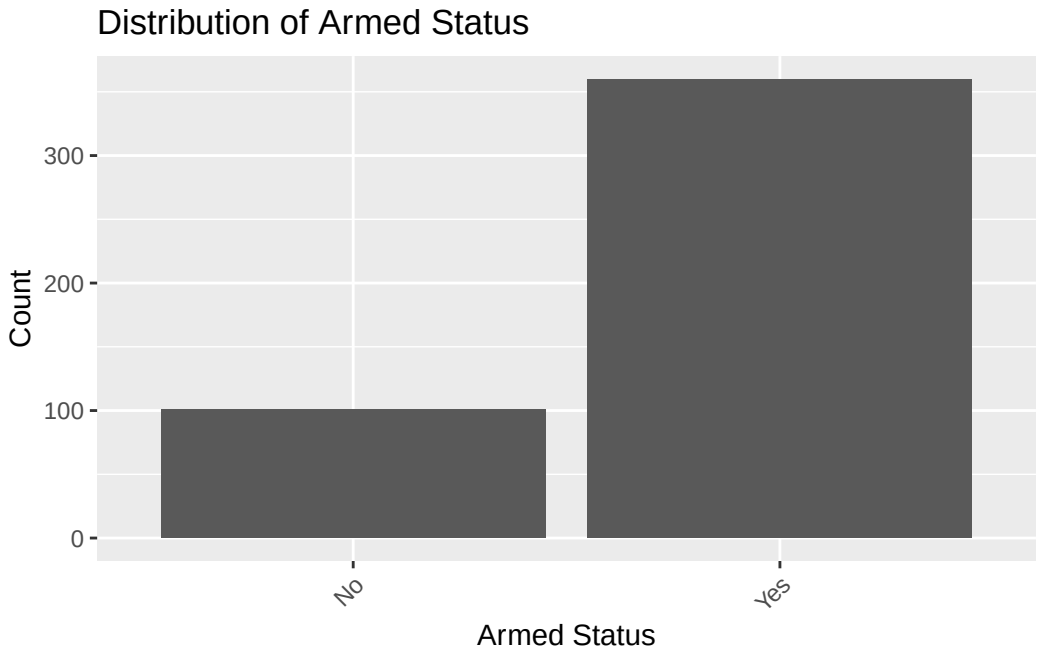
Distribution of Armed Status

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.

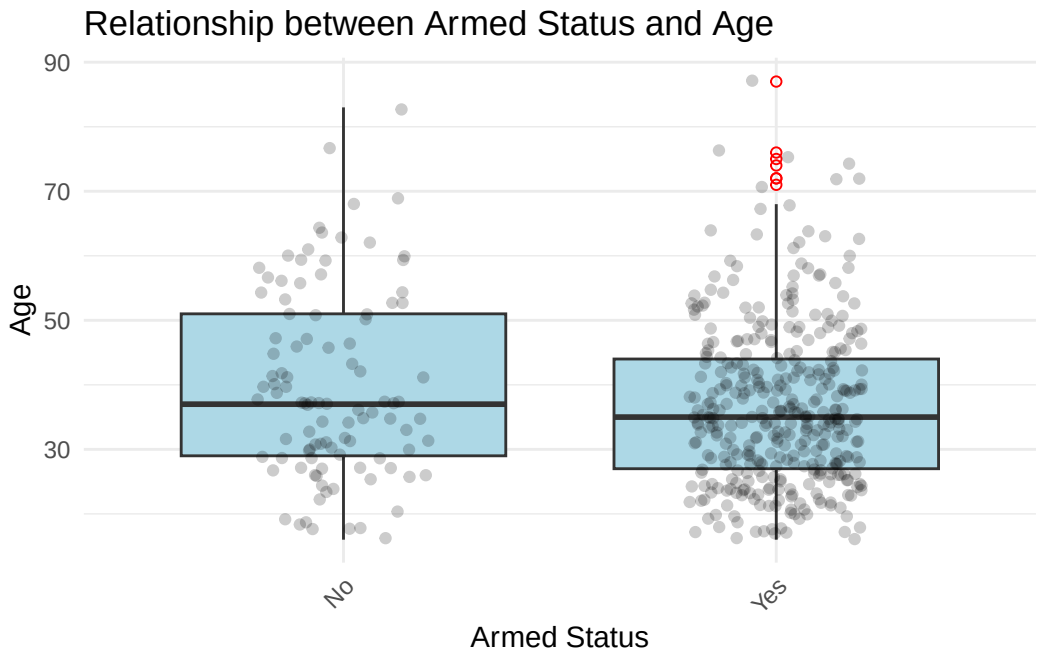
Table 7: Armed Status Distribution

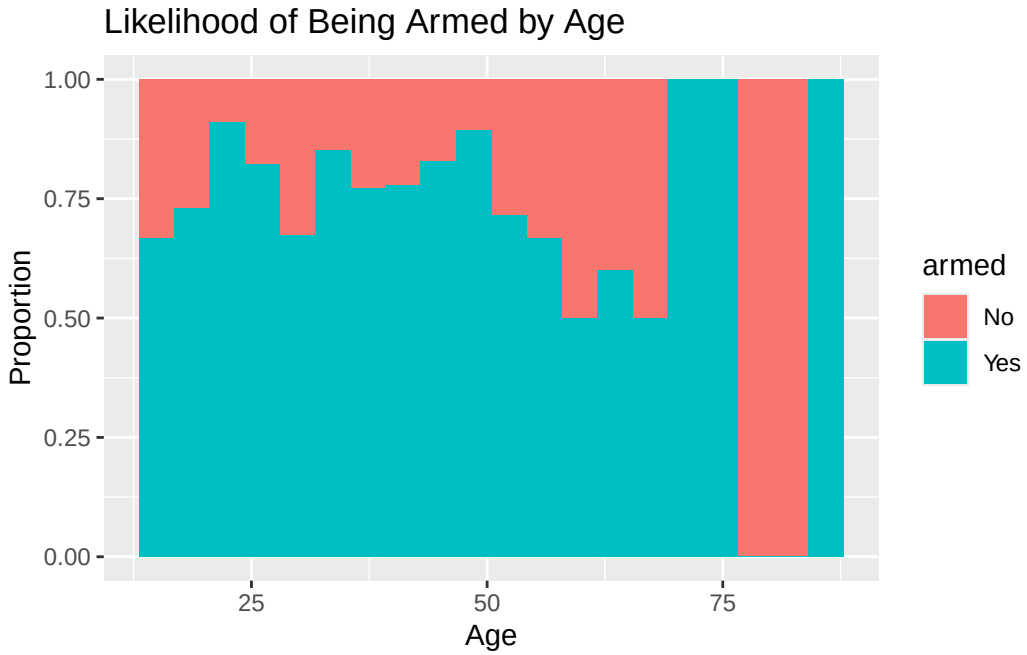
Armed_Status	Count	Percentage
No	101	21.91
Yes	360	78.09



Relationship Between Armed Status and Age

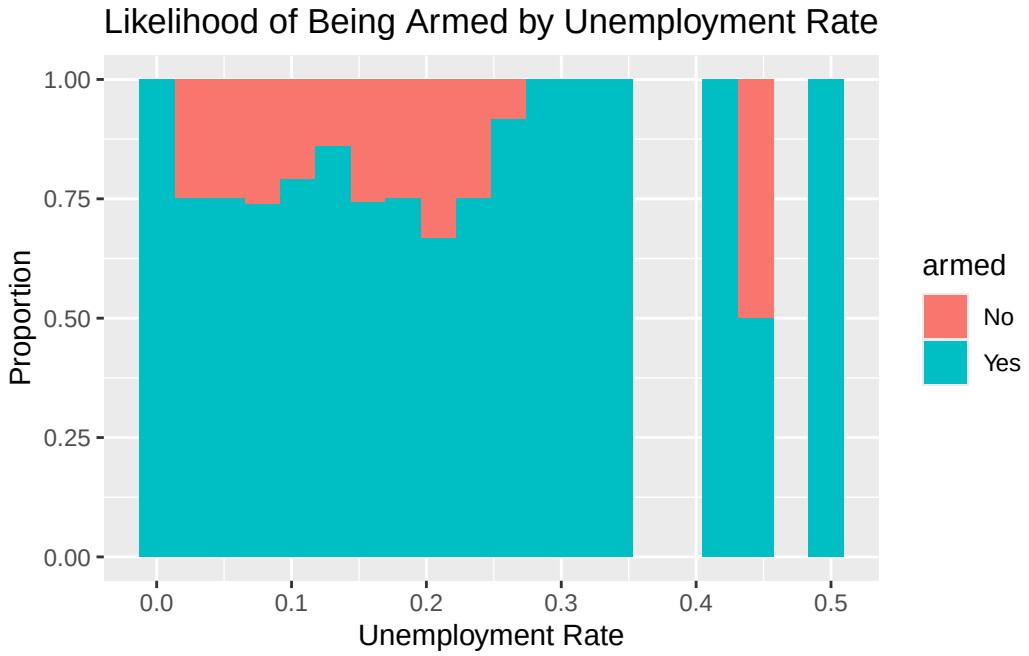
The relationship between age and armed status is further explored through a box plot, which compares the age distribution for armed and unarmed individuals. The plot illustrates that the age ranges and medians for both groups are remarkably similar, suggesting no significant difference in age distribution 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.


Relationship between Armed Status and Age
90
70
50
30
Age
No
Yes
Armed Status


Likelihood of Being Armed by Age
1.00
0.75
0.50
0.25
0.00
Proportion
25
50
75
Age
armed
No
Yes

Unemployment Rate and Likelihood of Being Armed

The role of unemployment rate is captured in the histogram, which displays the proportion of armed individuals across varying levels of unemployment. While the logistic regression did not find this relationship to be statistically significant, the visualizations hint at a potential pattern. Specifically, the proportion of armed individuals appears to increase slightly at higher unemployment rates. This observation suggests that unemployment could contribute indirectly to the likelihood of being armed, potentially reflecting economic pressures or broader social dynamics at play. The trend, while subtle, calls for more granular investigations to uncover the precise mechanisms underlying this association.



Conclusion

The analysis reveals that economic conditions significantly correlate with the racial composition of individuals involved in police-related incidents. Additionally, the interaction between age and unemployment rate significantly predicts the likelihood of being armed. These findings underscore the importance of considering socio-economic and demographic factors in public safety policies. Limitations of this study include [limitations]. Future work could explore [suggestions for future research].