



جامعة مصر للمعلوماتية
EGYPT UNIVERSITY
OF INFORMATICS

Egypt University of Informatics
Computer and Information Systems
Data Analysis Course

The Analysis of The Relationship Between Unemployment Rates and Crime in the U.S.

Submitted by: Kirolos Rafaat 23-101223 || Omar Mohamed Abdelsalam 23-101207 || Ali Tamer 23-101210 || Youssef Walid Salah 23-101117

24/05/2025

Introduction

Comprehending the relationship between economic conditions and crime is essential for formulating effective public policy. Unemployment has historically been regarded as a variable that may affect crime rates due to its influence on financial stability, social stress, and opportunity structures. This study examines the statistical correlation between unemployment rates and crime levels in the United States, employing data from all 50 states spanning multiple decades (1976–2015). Historical events, such as the 2008 financial crisis, the COVID-19 pandemic, and other substantial economic transformations, have highlighted the importance of this topic, rendering it pertinent and critical for policymakers, criminologists, and economists.

What occurs when individuals become unemployed? Does escalating unemployment result in a surge in criminal activity, or is the correlation more intricate than it seems? This paper examines the impact of the unemployment rate on crime rates, specifically violent and property crimes, utilizing empirical data and sophisticated analytical methods. Through the analysis of temporal and regional trends, we seek to identify patterns that could guide future policy decisions.

Alongside the analysis of raw data, we also take into account external factors such as poverty levels and population size, which can profoundly affect both unemployment and crime rates. This comprehensive approach enables us to transcend mere correlations and attain a more profound comprehension of the socioeconomic dynamics involved.

Research Question

Is there a statistically significant relationship between unemployment rates and crime rates across U.S. states?

Population of Interest:

The population of interest includes all 50 U.S. states, with data ranging from the 1970s up to 2015. The dataset comprises annual crime and unemployment statistics, yielding a total of 1,989 observations.

Sampling Method:

This study utilizes a complete enumeration of publicly available state-level data, incorporating all U.S. states across multiple years. Although the data is not randomly sampled, its comprehensive coverage enhances the validity and generalizability of the findings. By including the entire population of interest rather than a sample, the study eliminates sampling bias and avoids sampling error. This approach enables more accurate and detailed analysis of trends and patterns over time and across regions, thereby strengthening the reliability of the results.

Bias Identification:

In designing this study, we have taken several steps to identify and minimize potential sources of bias. One key source of bias could be selection bias, as our dataset includes all 50 U.S. states over multiple years, but not all years are represented uniformly. To mitigate this, we included only complete data for each year and state, ensuring that missing values were handled consistently through imputation or exclusion.

Another potential bias is confounding variables —factors like poverty and population that may influence both unemployment and crime rates. To address this, we incorporated additional datasets (e.g., poverty rates) and used regression analysis to control for these variables in our models.

The dataset provided is publicly available and standardized, which further reduces the risk of researcher bias.

Survey Questions/Collected Data/Dataset:

The main dataset used is a cleaned version of [Crime rate and Unemployment rate by state-by-state](#), containing the following variables:

Variables	Description
State	U.S. state abbreviation (e.g., AL, CA, NY).
Year	Calendar year the data corresponds to.
Population	Used for visual scaling in bubble plots.
Unemployment	Percentage of the labor force that is unemployed and actively seeking work.
Violent crime	Incidents such as assault, robbery, and homicide.
Property Crime	Incidents like burglary, theft, and motor vehicle theft.
Total Crime	The sum of violent and property crime incidents.

Plus, we used another dataset, [hstpov9.csv](#), for the poverty rate (confounding variable):

Variables	Description
Year	The year for which poverty data is reported. Data spans from 1959 to 2023.
Total Population	Total U.S. population in thousands, as of March of the following year.
Population Below Poverty	Number of people living below the poverty level, expressed in thousands.
Poverty Rate (%)	Percentage of the total U.S. population that lives below the poverty level.

Number of Samples: $n = 1989$, which represents the number of rows, where each row represents a unique observation (typically a specific U.S. state in a specific year).

Analysis:

The relationship between unemployment and crime was explored using

Descriptive Statistics:

We calculated basic descriptive statistics such as mean, median, and standard deviation for key variables including unemployment, violent crime, and property crime. These statistics helped us understand the central tendencies and variability in the data.

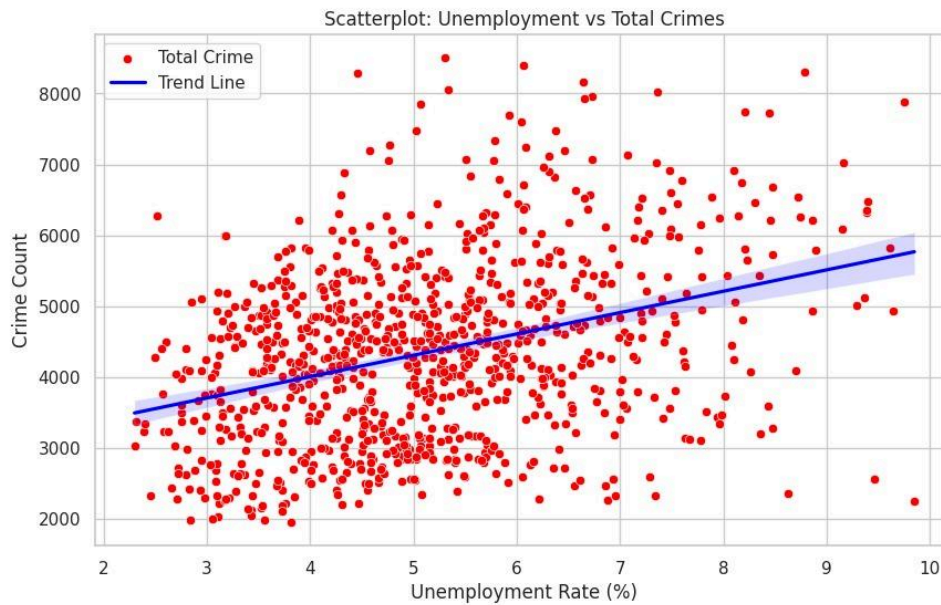
For example:

- The average unemployment rate was around 5.14%.
- The average violent crime rate was approximately 416 crimes per 100,000 people.
- The average property crime rate was approximately 3,933 crimes per 100,000 people.

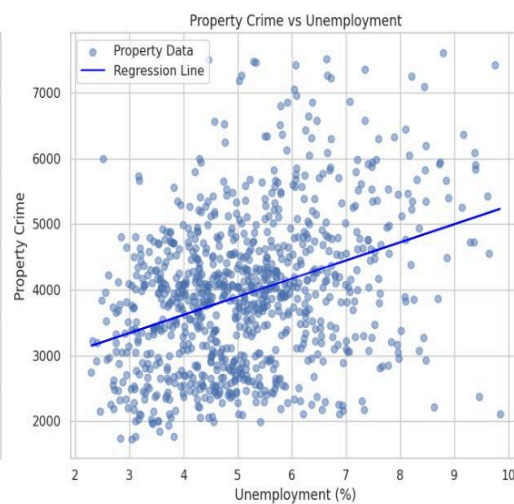
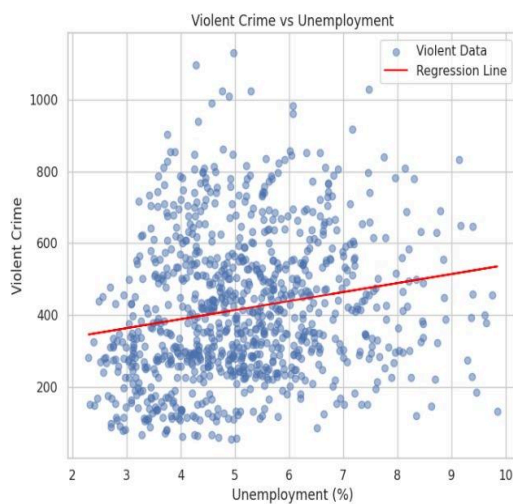
These statistics provide a baseline understanding of the data before conducting more advanced analysis.

Visualization:

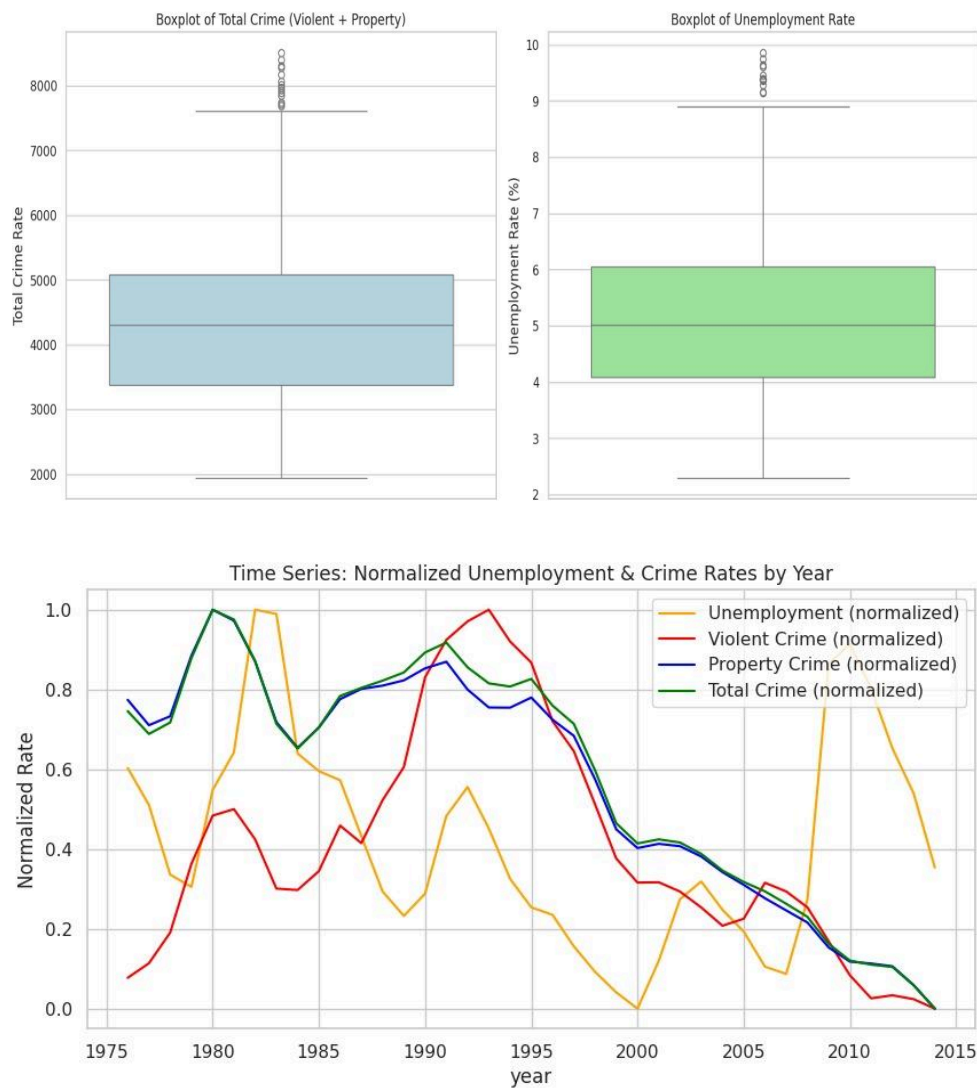
- **Scatterplots:** Showed positive trends between unemployment and both violent and property crimes.



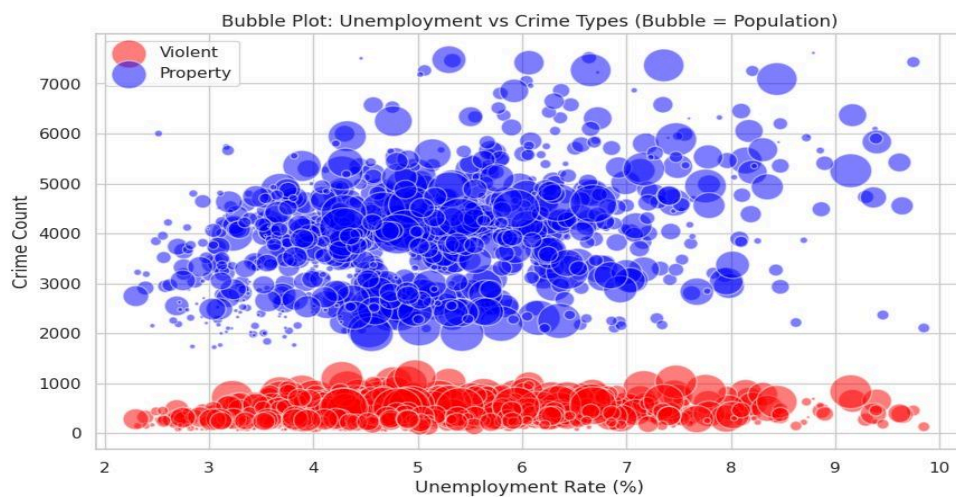
- **Linear Regression:** Demonstrated a statistically significant positive association.



- **Boxplots and Time Series Plots:** Helped assess variation across time and states.



- **Bubble Plots:** Illustrated regional differences.



- **Z-tests:**

Violent Crime Z-test: $z = 4.068$, $p\text{-value} = 0.000047$

Property Crime Z-test: $z = 9.1$, $p\text{-value} = 0.00000000000000000085$

Crime Rate Z-test: $z = 8.91$, $p\text{-value} = 0.00000000000000000048$

- **Pearson's correlation test:**

Violent Crime: $r = 0.18$, $p = 0.000000015$

Property Crime: $r = 0.35$, $p = 0.000000000000000000000000000012$

Crime Rate: $r = 0.35$, $p = 0.000000000000000000000000000042$

Hypothesis Testing Steps

1. Z-Test Analysis

Step 1. Define Null and Alternative Hypotheses

Null Hypothesis (H_0): No difference in crime rates (violent, property, total) between high and low unemployment groups.

$$H_0 : \mu_{\text{high}} - \mu_{\text{low}} = 0$$

Alternative Hypothesis (H_1): Significant difference in crime rates between high and low unemployment groups.

$$H1 : \mu_{\text{high}} - \mu_{\text{low}} \neq 0$$

Step 2. Choose the Appropriate Test

Two-sample Z-test, comparing crime rates between high (unemployment > median) and low (unemployment ≤ median) groups, assuming large sample size and normal distribution.

Step 3. Mention P-Values and Calculations

Violent Crime: $z = 4.0684$, $p = 0.0000473$

Property Crime: $z = 9.1072$, $p = 8.45e-20$

Total Crime: $z = 8.9179$, $p = 4.75e-19$

Step 4. Determine Statistical Significance

Significance level: $\alpha = 0.05$

All p-values (0.0000473, 8.45e-20, 4.75e-19) < 0.05, rejecting H_0 for all crime types.

Conclusion: High unemployment groups have significantly higher crime rates than low unemployment groups.

2. Pearson's Correlation Analysis

Step 1. Define Null and Alternative Hypotheses

Null Hypothesis (H_0): No correlation between unemployment and crime rates ($r = 0$).

Alternative Hypothesis (H_1): Correlation exists between unemployment and crime rates ($r \neq 0$).

Step 2. Choose the Appropriate Test

Pearson's correlation test, assessing the linear relationship between unemployment and crime rates (violent, property, total).

Step 3. Mention P-Values and Calculations

Violent Crime: $r = 0.1842$, $p = 1.57e-8$

Property Crime: $r = 0.3531$, $p = 1.16e-28$

Total Crime: $r = 0.3496$, $p = 4.20e-28$

Step 4. Determine Statistical Significance

Significance level: $\alpha = 0.05$

All p-values ($1.57e-8$, $1.16e-28$, $4.20e-28$) < 0.05 , rejecting H_0 .

Correlation strength: Very weak (violent, $r = 0.1842$), Weak (property, $r = 0.3531$; total, $r = 0.3496$).

Conclusion: Weak positive correlations exist between unemployment and crime rates, suggesting other factors also influence crime.

3. Linear Regression Analysis

Step 1. Define Null and Alternative Hypotheses

Null Hypothesis (H_0): Unemployment has no effect on crime rates (coefficient = 0).

Alternative Hypothesis (H_1): Unemployment affects crime rates (coefficient $\neq 0$).

Step 2. Choose the Appropriate Test

Linear regression (OLS), modeling crime rates (violent, property, total) as a function of unemployment.

Step 3. Mention P-Values and Calculations

Violent Crime: Coefficient = 25.10, $p < 0.001$, $R^2 = 0.034$

Property Crime: Coefficient = 276.23, $p < 0.001$, $R^2 = 0.125$

Total Crime: Coefficient = 301.34, $p < 0.001$, $R^2 = 0.122$

Step 4. Determine Statistical Significance

Significance level: $\alpha = 0.05$

All p-values (< 0.001) < 0.05 , rejecting H_0 for all crime types.

Interpretation:

Violent crime: Each 1% increase in unemployment increases violent crime by ~25.1 incidents per 100,000 people ($R^2 = 3.4\%$).

Property crime: Each 1% increase in unemployment increases property crime by ~276.2 incidents per 100,000 people ($R^2 = 12.5\%$).

Total crime: Each 1% increase in unemployment increases total crime by ~301.3 incidents per 100,000 people ($R^2 = 12.2\%$).

Conclusion: Unemployment significantly predicts higher crime rates, but its explanatory power is limited, especially for violent crime.

Overall Conclusion:

All tests confirm significant relationships between unemployment and crime rates. Z-tests show higher crime rates in high unemployment groups, Pearson's correlations indicate weak positive relationships, and linear regression quantifies unemployment's impact. However, low R^2 values (3.4–12.5%) suggest unemployment is not a dominant factor, and further analysis of other variables is recommended.

Conclusion

The data demonstrates a distinct positive correlation between unemployment and crime in the United States. Although causality cannot be established due to the observational nature of the data, the evidence indicates that increased unemployment correlates with heightened levels of both violent and property crimes. These findings highlight the necessity of considering economic policy and employment policies as essential components of crime reduction initiatives.

This research promotes a multidisciplinary approach to policy formation, fostering collaboration among economists, sociologists, and law enforcement officials. Initiatives that enhance employment accessibility, especially in areas with a history of elevated crime rates, may alleviate the social repercussions of unemployment. Furthermore, comprehending the intricacies of this link, including regional trends and the delayed impacts of economic recessions, could improve the accuracy of treatments.

As crime and unemployment persist as concerns for American communities, data-driven policies based on empirical research, such as this study, can offer a framework for fostering safer, more resilient societies.

Any potential issues

- **Confounding Variables:** Other factors not accounted for in the dataset, such as education and healthcare access, may influence crime rates.
- **Geographic Variability:** Crime and unemployment rates vary significantly across states due to differences in demographics, economy, and policy, which may affect the generalizability of the findings.

Contribution:

Contribution for every team member is on GitHub README.md