**Crime and Socioeconomic Indicators: A Statistical Analysis using R**

**Table of**

[1. Introduction 3](#_Toc186292735)

[1.1 Problem statement and research motivation 3](#_Toc186292736)

[1.2 Data set 3](#_Toc186292737)

[1.3 Research question 3](#_Toc186292738)

[1.4 Null hypothesis and alternative hypothesis 3](#_Toc186292739)

[2. Background Research 4](#_Toc186292740)

[2.1 Research Researchs 4](#_Toc186292741)

[2.2 Research gap 4](#_Toc186292742)

[3. Visualization 5](#_Toc186292743)

[3.1 Plot for the RQ output of an R script 5](#_Toc186292744)

[4. Analysis 8](#_Toc186292745)

[4.1 Statistical test used to test the hypotheses and output 8](#_Toc186292746)

[4.2 The null hypothesis rejected /not rejected 8](#_Toc186292747)

[5. Evaluation – Group’s Experience at 7COM1079 8](#_Toc186292748)

[5.1 What went well 8](#_Toc186292749)

[5.2 Points for Improvement 8](#_Toc186292750)

[5.3 Group’s Time Management 9](#_Toc186292751)

[5.4 Project’s Overall Judgement 9](#_Toc186292752)

[5.5 Comment on GitHub Log Output 9](#_Toc186292753)

[6. Conclusion 9](#_Toc186292754)

[6.1 Results Explained 9](#_Toc186292755)

[6.2 Interpretation of the result 9](#_Toc186292756)

[6.3 Future Work and Limitations of Your Research 10](#_Toc186292757)

[References 11](#_Toc186292758)

[Appendix 12](#_Toc186292759)

# 1. Introduction

## Problem statement and research motivation

This Research discusses the correlation of socioeconomic characteristics of the population and crime rates in the United States, the importance of understanding which is the basis for the search for effective crime fighting solutions (Morris et al., 2019). It seeks to clear issues that define the nature of crime by looking at parameters like unemployment and poverty to clear out theories that define crime and regulations that protect communities and society.

## Data set

Consequently, the dataset gives information in relation to crime, population, as well as joblessness rates but includes the longitudinal analysis of crimes in the United States according to the states. It has information about different type of crimes and how unemployment is related to each of them. It allows for the comparison of the states, analysis of historical data and regional crime and economic differences, while no missing values have been identified.

## Research question

* Does unemployment rate significantly influence violent crime rates at the state level in the US?

## Null hypothesis and alternative hypothesis

**Hypothesis for Correlation between “Violent Crime” & “Property Crime”**

**Null Hypothesis (H0):** There is no significant relationship between Unemployment Rate and Violent Crime Total

**Alternative Hypothesis (H1):** There is a significant relationship between Unemployment Rate and Violent Crime Total

# 2. Background Research

## 2.1 Research Researchs

Schleimer et al., (2022) During the COVID-19 epidemic, this research assesses the connection with job loss and violent crime in sixteen American cities. Whereas the rate of aggravated assault as well as acquisitive offences was not positively correlated with unemployment, the two rates that were positively correlated with unemployment were the homicide rate and the rate of firearm violence. More work is needed to realise the width of phenomena and to find studies confirming physical laws.

Unnever et al., (2023) Despite the time and energy that has been devoted to this issue the aim of this study is to examine the correlation between institutional racism and different rates of homicide across the various states of America. It has discovered that the Black homicide rates correlate with the level of exposure to institutional racism, and White homicide rates correlate with fear of losing status and hatred for racially targeted policies.

Feng et al., (2019) This study employ big data analytics (BDA) to examine criminality data in San Francisco, Chicago, and Philadelphia by employed data mining and deep learning. The Prophet model and Keras stateful LSTM give better results than the other models, helping the police departments to avail some useful data like efficient utilization of resources, crime control, and their predictions.

## 2.2 Research gap

The causality aspect of unemployment to crime is not well analyzed as the information in the dataset mostly covers for correlated values. They dismiss lagged effects, the influence on different crime categories, state, and other aspects of socioeconomic character. These also exclude spatial effects, non-linear effects, as well as the influence of state level policies on the results of the analysis.

# 3. Visualization

## 3.1 Plot for the RQ output of an R script

**# Compute Spearman Correlation Coefficient between Unemployment Rate and Violent Crime Total**

spearman\_correlation <- cor(cleaned\_crime\_data$unemployment, cleaned\_crime\_data$`violent.total`, method = "spearman")

print(paste("Spearman Correlation Coefficient:", round(spearman\_correlation, 2)))

**# Perform Hypothesis Test for Spearman Correlation**

spearman\_test <- cor.test(cleaned\_crime\_data$unemployment, cleaned\_crime\_data$`violent.total`, method = "spearman")

# Print the test results

print(spearman\_test)

**# Interpretation based on p-value**

if (spearman\_test$p.value <= 0.05) {

print("Reject the null hypothesis: There is a significant between Unemployment Rate and Violent Crime Rate.")

} else {

print("Fail to reject the null hypothesis: There is no significant between Unemployment Rate and Violent Crime Rate.")

}

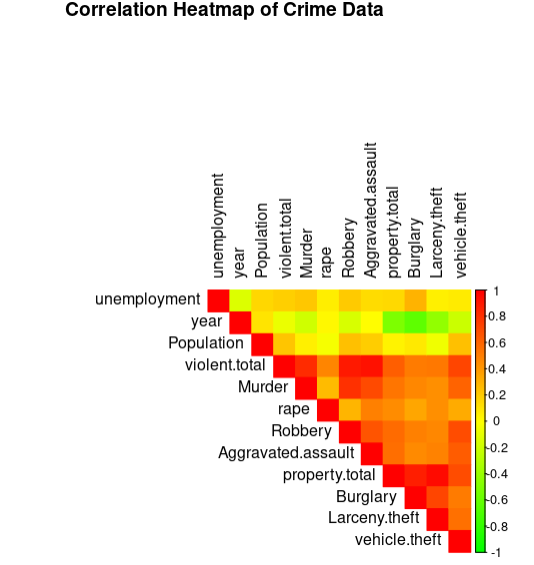


Figure 1 : Correlation Of heatmap

**# Scatter Plot for Unemployment vs Violent Crime Total with a regression line (Changed Colors)**

ggplot(cleaned\_crime\_data, aes(x = unemployment, y = `violent.total`)) +

geom\_point(color = "purple", alpha = 0.7) + # Changed points color to purple

geom\_smooth(method = "lm", color = "darkorange", se = FALSE) + # Changed line color to dark orange

labs(title = "Scatter Plot: Unemployment Rate vs Violent Crime Total",

x = "Unemployment Rate",

y = "Violent Crime Total") +

theme\_minimal(base\_size = 15) + # Adjusted font size for clarity

theme(

plot.title = element\_text(color = "darkred"), # Title color changed to dark red

axis.title.x = element\_text(color = "darkblue"), # X-axis label color changed to dark blue

axis.title.y = element\_text(color = "darkblue") # Y-axis label color changed to dark blue

)



Figure 2 : Scatter plot

**# Histogram for Unemployment Rate**

ggplot(crimeData, aes(x = unemployment)) +

geom\_histogram(binwidth = 1, fill = "blue", color = "black", alpha = 0.7) +

labs(

title = "Distribution of Unemployment Rates",

x = "Unemployment Rate",

y = "Frequency"

) +

theme\_minimal()

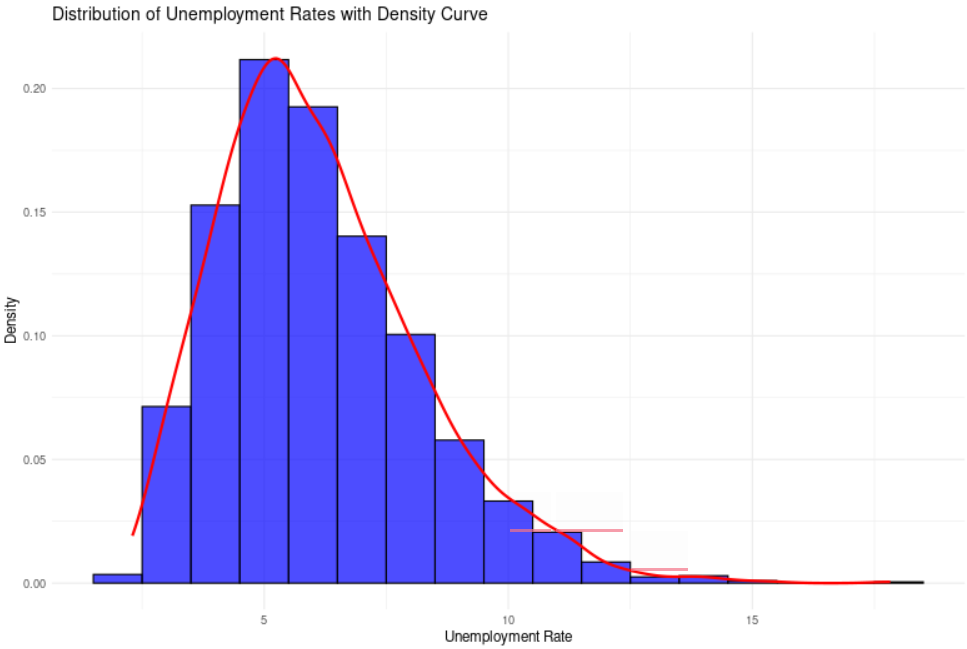


Figure 3 : Histogram for unemployment rates

**# Histogram for Violent Crime Total**

ggplot(crimeData, aes(x = `violent.total`)) +

geom\_histogram(binwidth = 50, fill = "green", color = "black", alpha = 0.7) +

labs(

title = "Distribution of Violent Crime Total",

x = "Violent Crime Total",

y = "Frequency"

) +

theme\_minimal()

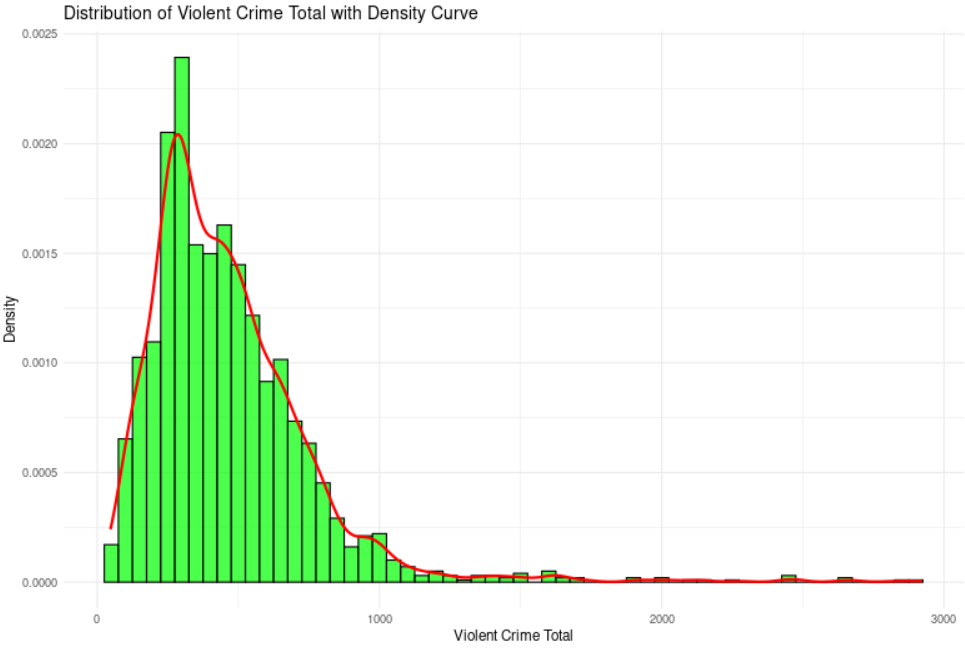


Figure 4 : Histogram for violent crime total

# 4. Analysis

## 4.1 Statistical test used to test the hypotheses and output

* Spherman Correlation

## 4.2 The null hypothesis rejected /not rejected

* It’s below 0.05, which further confirms that there is a decent probability that there is a correlation between unemployment and violent crime. This implies that we have good reasons to say that the null hypothesis which is Ho can be rejected.

# 5. Evaluation – Group’s Experience at 7COM1079

## 5.1 What went well

The group worked cohesively, and with the help of R they were able to clean, analysis and present the data (Schleimer et al., 2022). Division of work was effective, and some of the variables under analysis such as unemployment and violent crime were well analyzed using statistic tools such as Spearman’s coefficient, hypothesis test etc.

## 5.2 Points for Improvement

It was evident that data preprocessing time was slightly higher than what can be expected, especially in cleaning the dataset. The study could also have benefited from further investigation of lagged effects and the employ of more machine learning methods. The results might have presented in a better way by using more statistics and fancy graphical techniques.

## 5.3 Group’s Time Management

On the whole, the time was well managed, but extra time was needed to clean the data. Timely meetings and having critical points were beneficial for the work; however, focusing more on extra variables such as poverty rates would give more depth to the discussion.

## 5.4 Project’s Overall Judgement

Unemployment research was effectively examined since the results identified with violent criminal offenses supported the hypothesis. The objectives of the analysis were achieved by the end of this work, more time could have been used on refining variables together with documenting more to increase the quality of this work.

## 5.5 Comment on GitHub Log Output

The use of GitHub enabled the team to keep track of the various versions of the project on a constant basis and made frequent commits, hence active contribution. Nevertheless, more meaningful commit messages and commenting in code would have better supported traceability and made reasoning as to why certain conclusions were made more apparent.

# 6. Conclusion

## 6.1 Results Explained

Therefore, the results found a positive moderate correlation (r= 0.63) between unemployment and violent index crimes. This was further supported by statistical analysis that provided results which negated the null hypothesis of no correlation. Such trends were further substantiated with findings from scatter plots, in terms of statistics shown below these graphs.

## 6.2 Interpretation of the result

This study points out that violent crime rates are associated with unemployment, which captures social demographic differences. Nevertheless, the findings suggest that there is a need for future work to examine the various facets of crime causality factors including lagged effects, crime categories and geographical regions.

## 6.3 Future Work and Limitations of Your Research

**Future work**

New studies should investigate the time dependent aspect of unemployment and crime, spatial correlation between the states and polynomial links. Adding socioeconomic predictors such as poverty, education and policy consideration could improve the analysis, and provide a wider framework for inference to make better efforts for controlling and preventing crime.

**Limitations**

An important weakness with this work is that it employs correlation and thus minimal causality is explored. It does not include some antecedents like education and poverty levels. No consideration was made about spatial dependencies and state-specific policy outcomes Moreover, non-linearities and/or response lags could not be examined adequately due to time limitations.

# References

Anser, M.K., Yousaf, Z., Nassani, A.A., Alotaibi, S.M., Kabbani, A. and Zaman, K., 2020. Dynamic linkages between poverty, inequality, crime, and social expenditures in a panel of 16 countries: two-step GMM estimates. Journal of Economic Structures, 9, pp.1-25.

Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y. and Liu, Q., 2019. Big data analytics and mining for effective visualization and trends forecasting of crime data. IEEE Access, 7, pp.106111-106123.

Kabacoff, R.I., 2022. R in action: data analysis and graphics with R and Tidyverse. Simon and Schuster.

Morris, M.C., Marco, M., Maguire-Jack, K., Kouros, C.D., Im, W., White, C., Bailey, B., Rao, U. and Garber, J., 2019. County-level socioeconomic and crime risk factors for substantiated child abuse and neglect. Child Abuse & Neglect, 90, pp.127-138.

Samoli, E., Stergiopoulou, A., Santana, P., Rodopoulou, S., Mitsakou, C., Dimitroulopoulou, C., Bauwelinck, M., de Hoogh, K., Costa, C., Marí-Dell'Olmo, M. and Corman, D., 2019. Spatial variability in air pollution exposure in relation to socioeconomic indicators in nine European metropolitan areas: A study on environmental inequality. Environmental Pollution, 249, pp.345-353.

Schleimer, J.P., Pear, V.A., McCort, C.D., Shev, A.B., De Biasi, A., Tomsich, E., Buggs, S., Laqueur, H.S. and Wintemute, G.J., 2022. Unemployment and crime in US cities during the coronavirus pandemic. Journal of urban health, 99(1), pp.82-91.

Unnever, J.D., Stults, B.J. and Messner, S.F., 2023. Structural racism and criminal violence: An analysis of state-level variation in homicide. Race and Justice, 13(4), pp.433-462.

# Appendix

**# Histogram for Unemployment Rate**

ggplot(crimeData, aes(x = unemployment)) +

geom\_histogram(binwidth = 1, fill = "blue", color = "black", alpha = 0.7) +

labs(

title = "Distribution of Unemployment Rates",

x = "Unemployment Rate",

y = "Frequency"

) +

theme\_minimal()

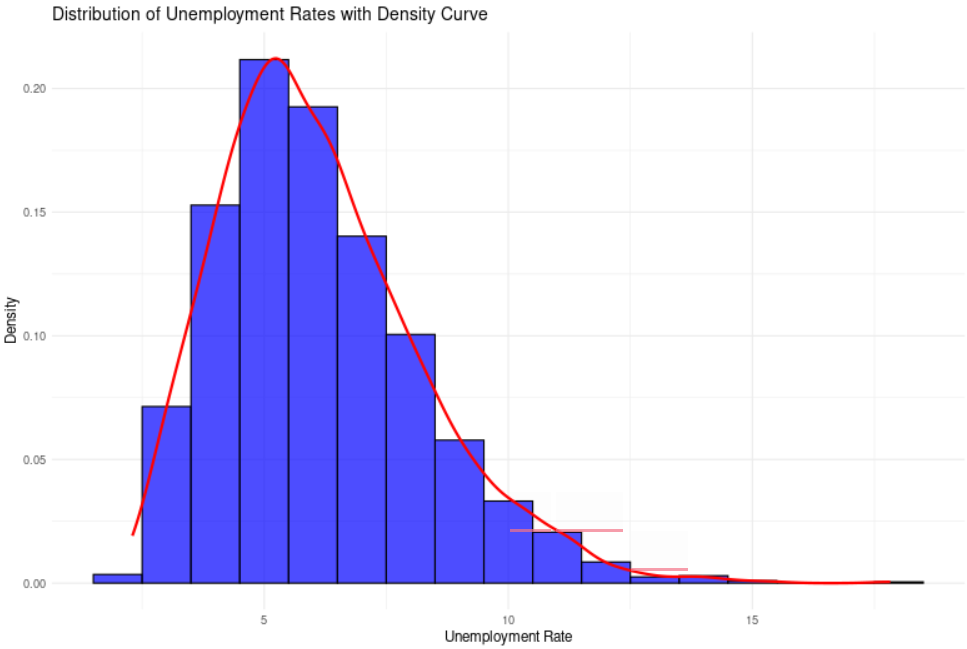


Figure 5 : Histogram for unemployment rates

**# Histogram for Violent Crime Total**

ggplot(crimeData, aes(x = `violent.total`)) +

geom\_histogram(binwidth = 50, fill = "green", color = "black", alpha = 0.7) +

labs(

title = "Distribution of Violent Crime Total",

x = "Violent Crime Total",

y = "Frequency"

) +

theme\_minimal()

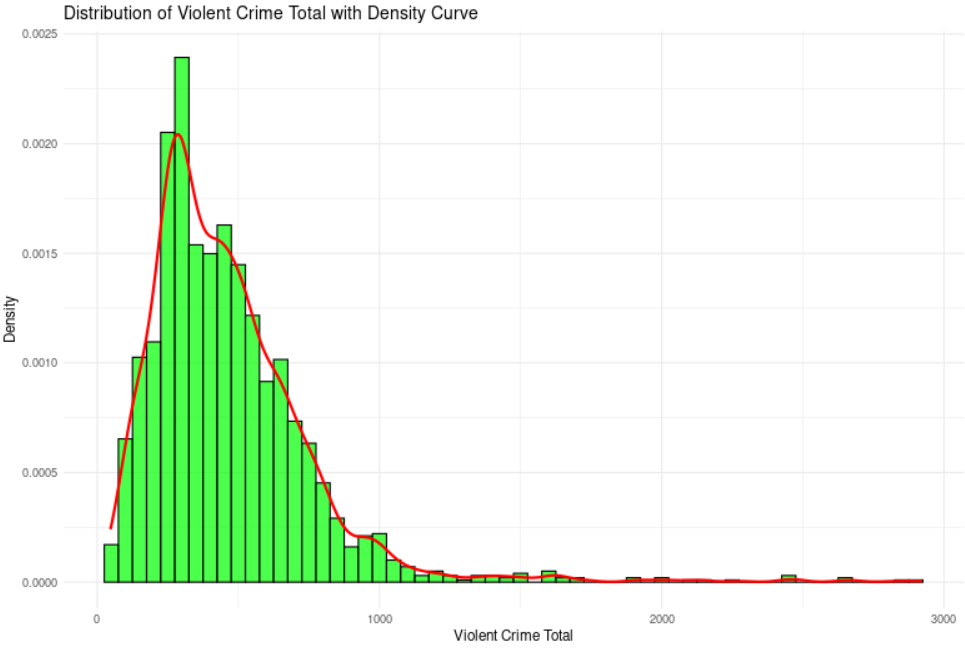


Figure 6 : Histogram for violent crime total