

# Final report

## Introduction

Domestic violence is a social issue which need to be noticed, with profound impacts on individuals and communities about health, psychology and economics. In New South Wales (NSW), variations in domestic violence rates across Local Government Areas (LGAs) highlight a complex problem shaped by socioeconomic conditions, cultural norms, and policy measures (Mitchell, 2011; Grech & Burgess, 2011). This research seeks to answer what are the key socioeconomic factors that influence domestic violence rates and how do these factors vary across LGAs in NSW. The aim is to search for the relationships between domestic violence rates and factors in order to provide effective suggestions to the governments according to the findings.

The motivation of the research is to reduce the harm of the domestic violence to the human society. Because the domestic violence not only brings severe physical and mental harm to victims but also leads to significant economic loss, such as the increase of medical expenses, the reduce of productivity and the pressure of social services. For example, the children who suffer the domestic violence are more probable to face psychological issues.

The research supposes that the rates of domestic violence are positively related to higher unemployment rates and rented rates, while the higher diplomas of education and incomes may reduce the rates. To be more specific, unemployment rates are expected to have a strong positive correlation with domestic violence rates because the instability of the work may tend to increase the probability of the conflicts of families. Education levels are expected to have a medium negative correlation because higher education levels could have more skills to solve conflicts of families and more chances to get higher salaries. Rented rates are presumed to have a moderate positive correlation because families without stable places could create more conflicts and stress. Finally, incomes are expected to have a medium negative correlation because more funds in families could give them more chances to deal with some sudden situations (Payne, Morgan, & Piquero, 2020; Stavrou, Poynton, & Weatherburn, 2016).

## Literature and theory

### Unemployment

The academic community generally believes that unemployment is an important predictor of domestic violence. The economic uncertainty caused by unemployment will increase the pressure within the family, which in turn intensifies conflict (Mitchell, 2011; Dowling et al., 2014). The study found that families with long-term unemployment were more likely to face a higher risk of violence than those with stable employment. Moreover, these trends are even more pronounced during a recession, as economic distress heightens family tensions.

### Housing situation

Housing situation could be another factor which influence the domestic violence. In specifically, when considering the rents, the families without their own house may face more stress, which could be one of the triggers (Grech & Burgess, 2011). It is especially in the places where the rents are higher and vacancy rates are low. High domestic violence could be more obvious in this phenomenon.

## Level of education

Education could be considered as one main point to prevent the domestic violence. To be more specific, the people in the regions with higher level of education could have more opportunities to accept the knowledge to solve the conflicts and avoid the violence. Moreover, they could get access to the wealth (Attah & Owoyemi, 2024). Additionally, the regions with higher level of education will provide more chances for people to understand violence is not a proper way to deal with the dilemmas. They create an environment where the victims are more willing to seek for help, and abusers are less likely to conduct violence.

## Income level

Low income has close connection with rates of domestic violence. The families with low income have to face economic pressures and will suppress their inner feelings, which will eventually explode at a certain point. In this situation, the tragedy could be easier to happen (Grech & Burgess, 2011). In the contrary, the families with high income are more confident to deal with dilemmas. However, the gap of income in the communities may cause the resentment, which could indirectly lead to domestic violence.

## Confounding variables

Several confounding variables are mentioned in the literature, such as population density, cultural norms, and access to support services. For example, population density affects the visibility and reporting rates of domestic violence incidents. Cultural factors may influence attitudes towards domestic violence, allowing it to be ignored or discouraging victims from seeking help (Grech & Burgess, 2011). In addition, regional differences in support services, such as shelters and counseling services, further complicate the relationship between socioeconomic factors and rates of domestic violence.

## Data and methodology

### Data Sources

**Primary Data:** This study uses data from the New South Wales Bureau of Crime Statistics and Research (BOCSAR, 2024). The dataset contains detailed information on LGAs for various criminal offences. The research will focus on domestic violence-related assaults categorized under the “Subcategory” variable, which may be set as a dependent variable.

**Additional Data:** The factors such as unemployment rates, income levels, education levels and rented rates are sourced from the Australian Bureau of Statistics (ABS). Additionally, the digital boundary files are also downloaded in ABS.

### Data Bias and Mitigation

One of the data bias is that domestic violence is often under-reported, especially in rural and marginalized areas. Inconsistent reporting across LGAs also creates gaps. Several steps are taken to address these issues. The research could utilize a variety of analytical methods to verify the accuracy of the results. For example, the comparison of the results between multi-linear regression and spatial linear regression could be used to mitigate the bias.

## Data description

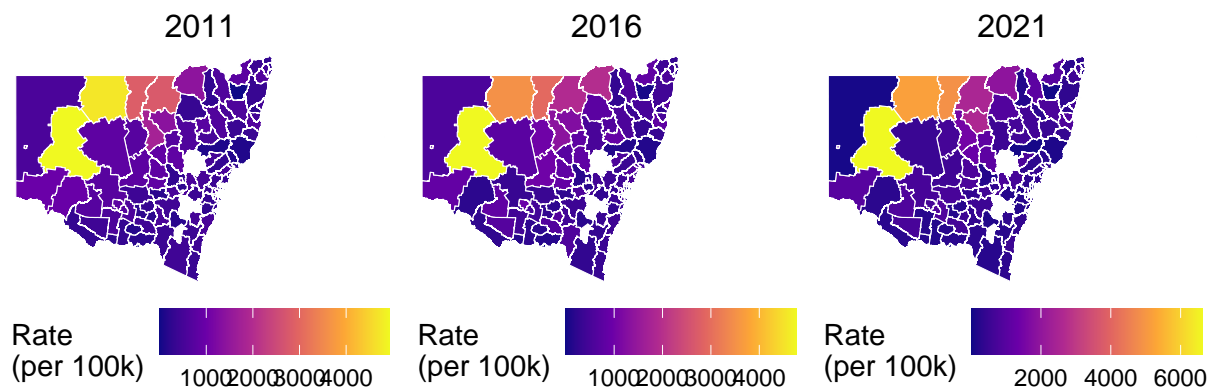
The primary data are classified by 131 LGAs. The range of data is from 1995 to Sep 2024 by month, and it contains several offence types. The additional data is DataPacks sorted in the Australian Bureau of Statistics. The metadata show the contents of census data statistically counted in 2011, 2016 and 2021 in 35 time series profile tables in the package. Moreover, the digital boundary files are provided by ABS as the same. They are available by ESRI shapefile format according to Australian Statistical Geography Standard (ASGS).

## Variables

**Dependent Variable:** Domestic violence rates are calculated as the number of domestic violence-related assaults per 100,000 population for each LGA. This variable is derived from the BOCSAR dataset. To ensure comparability across LGAs, domestic violence incidents were normalized using population data from the ABS, to be more specific, the features about the total persons in 2011, 2016 and 2021 census persons in 'T01-Selected Person Characteristics by sex' are used.

The distribution of the domestic violence rates are showed as maps as follows:

### NSW Domestic Violence Rate by LGA (2011, 2016, 2021)



It is evident from the visualization that there is a clear spatial pattern in domestic violence rates over the three years. In north-western New South Wales, the rate has remained high for three years, reflecting ongoing socio-economic challenges in these regions. At the same time, domestic violence rates have gradually declined in some local municipalities in the south-east, particularly around urban centers such as Sydney, possibly due to better access to support services and intervention programmes.

**Independent Variables:** The independent variables used in this study include unemployment rates, rented housing rates, education levels, and income levels. Unemployment rates are calculated as the percentage of unemployed individuals within the total labor force for each LGA, based on ABS census files in 'T33-Labour Force Status by Age by Sex' from 2011, 2016, and 2021. Rented housing rates represent the percentage of households living in rental properties, also derived from ABS housing datasets in 'T18-Tenure and Landlord Type by Dwelling Structure'. Education levels are expressed through an education index that incorporates weighted averages of postgraduate degrees, bachelor's degrees, and vocational certifications, capturing the overall educational attainment in each LGA in 'T31-Highest Non-School Qualification: Level of Education by Age by Sex'. Income levels are computed using weighted midpoints of income brackets provided in ABS census data in 'T22-Total Family Income (weekly) by Number of Children for Couple Families', offering an estimate of average household income for each region.

## Methodology

### Simple Linear Regression

Simple linear regression is a common method to explore the relationship between the domestic violence and the other independent variables. It is often used to reach the probable conclusion and help the researchers gain a simple understanding of the features in the dataset. What's more, this method has advantages in visualization due to its simplicity. So this method is considered to be the basic method of research.

However, the disadvantage of this method is also the simplicity because it does not take into account the influence of other variables. This limitation may reduce the robustness of this model. Moreover, in the real world, data may not be completely linear, which result in the reduce of model accuracy.

### Multivariate Logistic Regression

This method can integrate different characteristics to study the domestic violence rate at the same time, through this method, it is easier to find the common influence of multiple factors on the domestic violence rate.

It helps to improve the performance of the model, but the multicollinearity between variables has a small impact on the model. If the VIF value of a variable is too high, it will greatly affect the interpretability of the model.

### Generalized Additive Mixed Models (GAMM)

GAMM is a perfect method to explore the relationships when the dependent and independent variables are more likely to be nonlinear, which is to say, GAMM is ideal for capturing complex nonlinear interactions that linear models may miss.

It is flexible in modeling nonlinear relationships and robust to changes in data distribution. But it may cost higher computation due to the complexity of this model.

## Results from Analysis

### Simple Linear Regression

The tables show the correlation between the domestic violence rates and factors as follows:

Table 1: Correlation between Domestic Violence Rate and Unemployment Rate by Year

Year	Correlation between Domestic Violence Rate and Unemployment Rate
2011	0.3109150
2016	0.3695193
2021	0.0403399

Table 2: Correlation between Average Income and Domestic Violence Rate by Year

Year	Correlation between Average Income and Domestic Violence Rate
------	---

2011	-0.3468827
2016	-0.3966862
2021	-0.0580153

Table 3: Correlation between Education Index and Domestic Violence Rate by Year

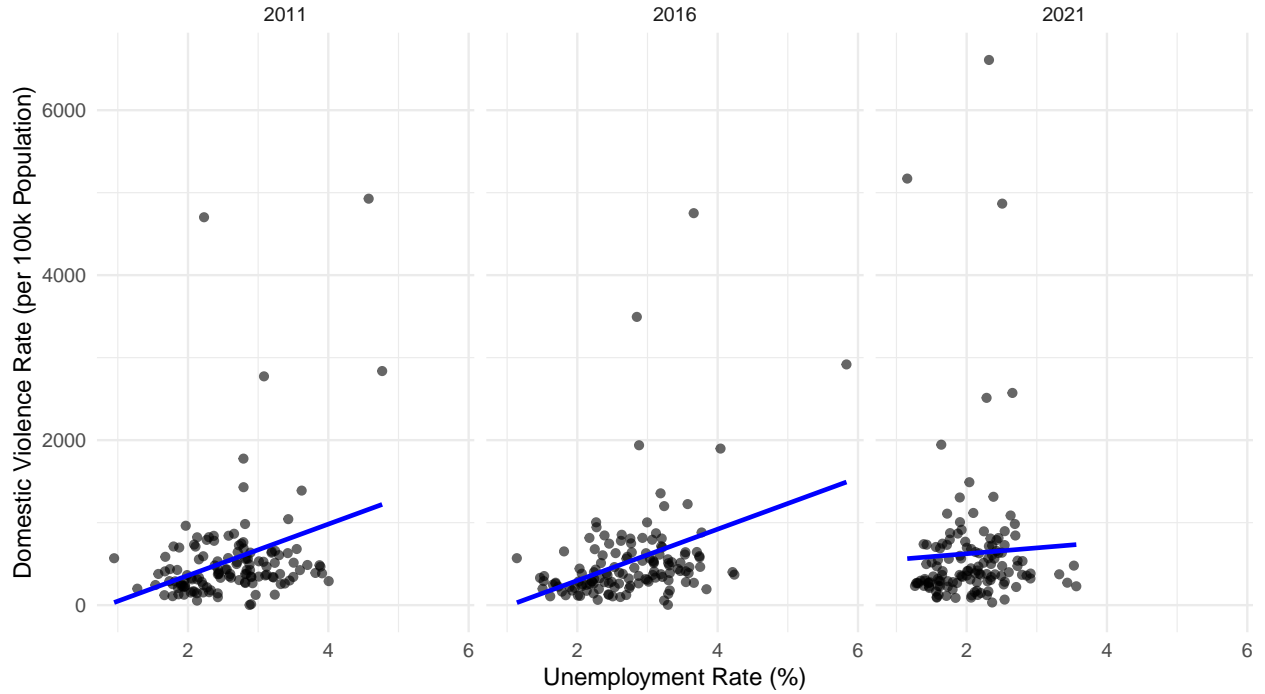
Year	Correlation between Education Index and Domestic Violence Rate
2011	-0.2108860
2016	-0.3223156
2021	-0.2117392

Table 4: Correlation between Rented Rate and Domestic Violence Rate by Year

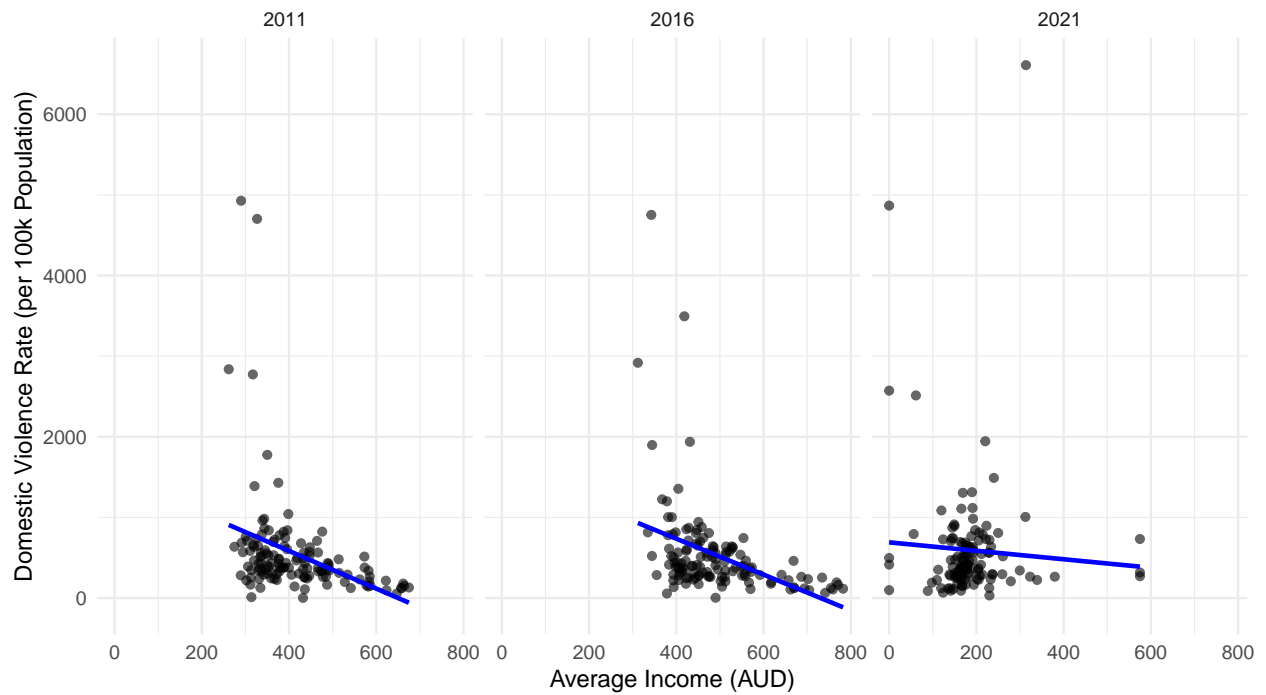
Year	Correlation between Rented Rate and Domestic Violence Rate
2011	0.1443804
2016	0.1662685
2021	0.1614455

The plots show the visualization between the domestic violence rates and factors as follows:

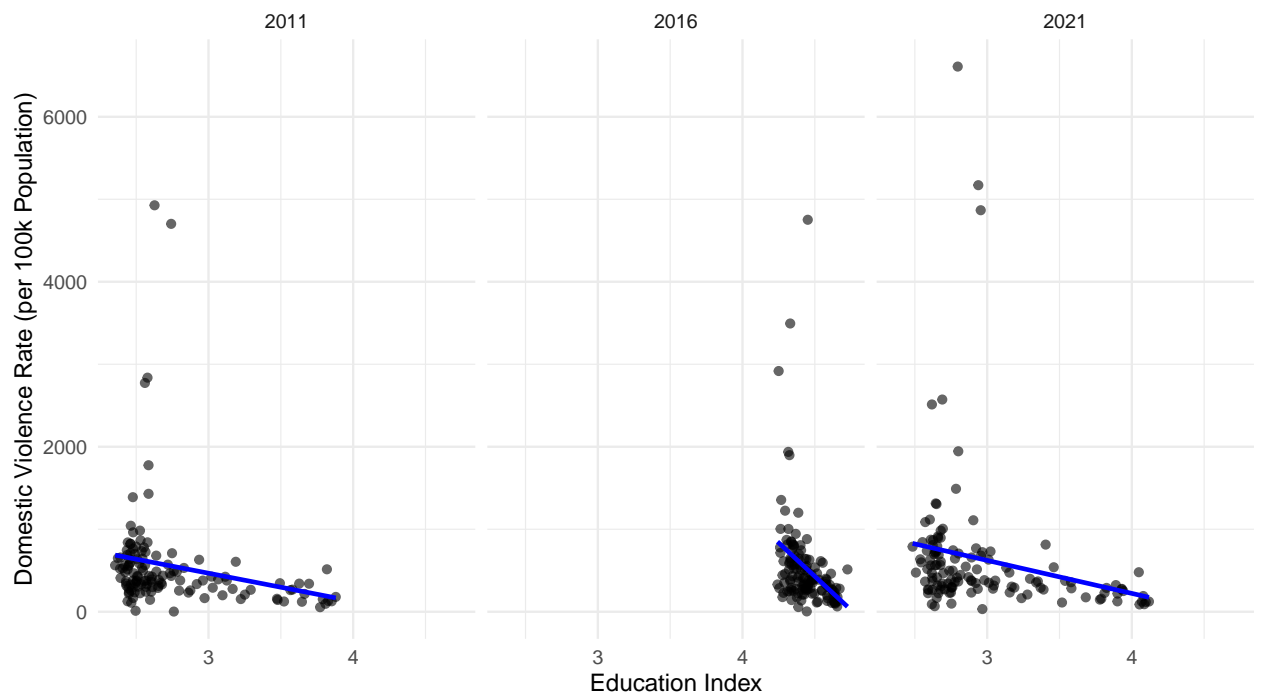
#### Relationship Between Unemployment Rate and Domestic Violence Rate

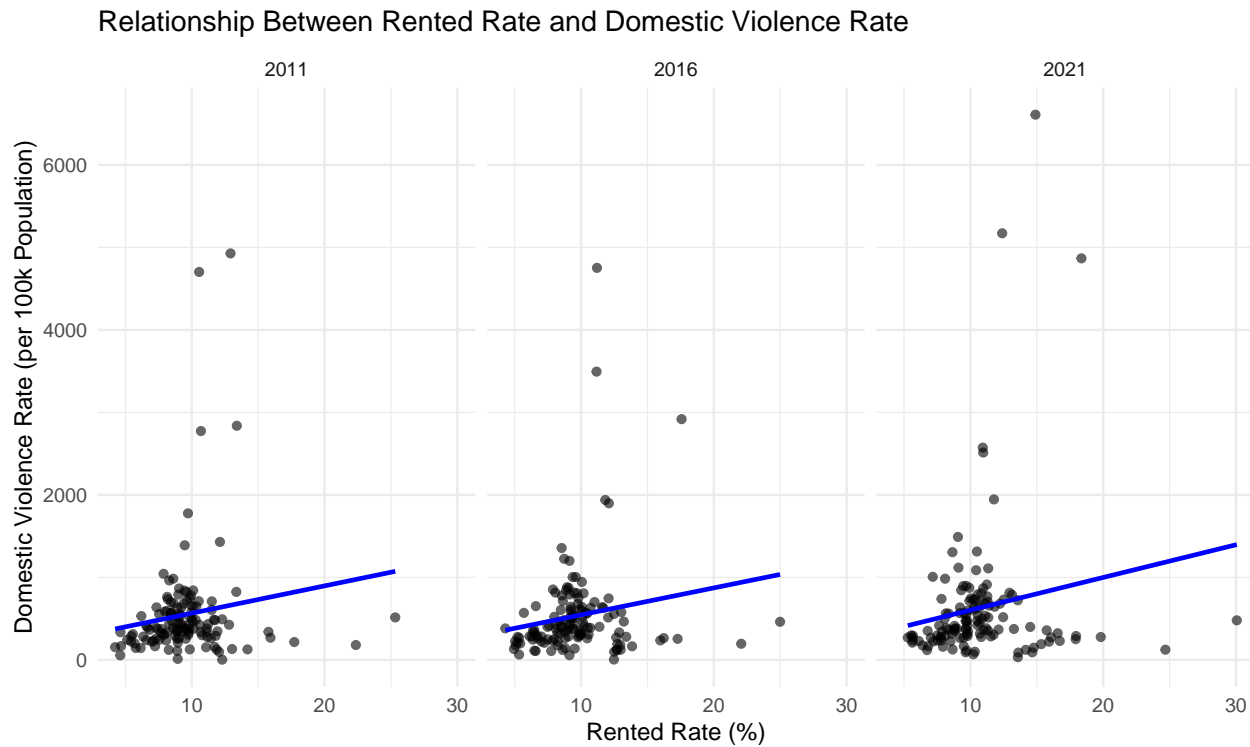


Relationship Between Average Income and Domestic Violence Rate



Relationship Between Education Index and Domestic Violence Rate





The results of simple linear regression provide clear insights into the relationship between domestic violence rates and individual socioeconomic variables. The detailed correlation analysis of the respective variables is as follows:

**Unemployment rate:** The correlation between unemployment rate and domestic violence rate is consistently positive across years (2011, 2016, and 2021). For example, the correlation is stronger in 2016 (0.369) compared to 2011 and 2021. The scatter plot shows a positive trend, with local governments with higher unemployment rates tending to show higher domestic violence rates, suggesting that economic instability contributes to family conflicts to a large extent.

**Average income:** As expected, the relationship between income level and domestic violence rate is negative. The correlation coefficient strengthens from -0.346 in 2011 to -0.396 in 2016 and slightly weakens in 2021. The scatter plot indicates that local governments with higher average incomes generally report lower domestic violence rates, confirming the protective effect of income on violence.

**Education index:** Education level, as represented by the education index, is moderately negatively correlated with domestic violence rates. The correlation coefficients remain relatively stable across years, with the strongest negative correlation in 2016 (-0.322). The scatter plot suggests that LGAs with higher levels of education experience fewer incidents of domestic violence, consistent with previous literature highlighting education as a mitigating factor.

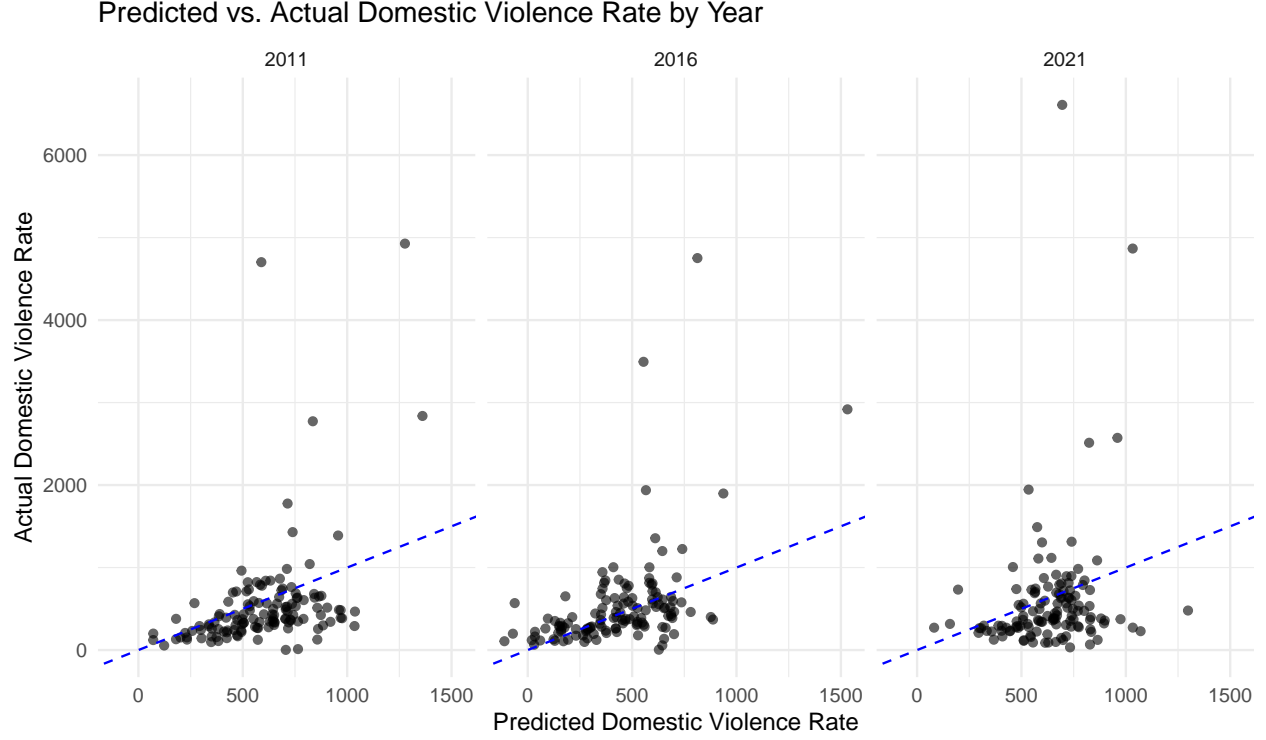
**Rental Housing Rate:** The correlation between the rental housing rate and domestic violence rates is positive, but weaker than that of unemployment or income. The scatter plot shows a consistent upward trend, with correlations ranging from approximately 0.144 to 0.166 over the years, suggesting that housing instability modestly contributes to higher rates of domestic violence.

## Multivariate Logistic Regression

Multivariate logistic regression integrates the independent variables to measure the comprehensive influence to the domestic violence rate.

Table 5: Model Performance Metrics by Year

Year	R <sup>2</sup>	Adjusted R <sup>2</sup>	p-value
2011	0.207	0.182	0
2016	0.287	0.264	0
2021	0.186	0.160	0



To ensure comparability and avoid scaling bias among predictors, all variables were standardized before fitting the model.

The R square values after model adjustments for 2011, 2016, and 2021 are 0.182, 0.264, and 0.160 respectively. These values indicate that the model accounts for about 18.2%, 26.4%, and 16.0% of the variation in domestic violence rates over the years. This performance peaked in 2016, which shows that the included predictors had a relatively large impact on explaining changes in domestic violence rates that year.

For all years, the model's p-values were statistically significant ( $< 0.001$ ), which shows that the combination of unemployment, income, education level, and rent significantly predicted the incidence of domestic violence.

The predicted and actual domestic violence rates for the year are displayed on a scatter plot. The dashed line represents an ideal scenario where the predicted values exactly match the actual values. Although the model showed reasonable prediction accuracy, it suggests that there is a significant bias, especially when the incidence of domestic violence is high, and that other unmeasured factors may influence these results.



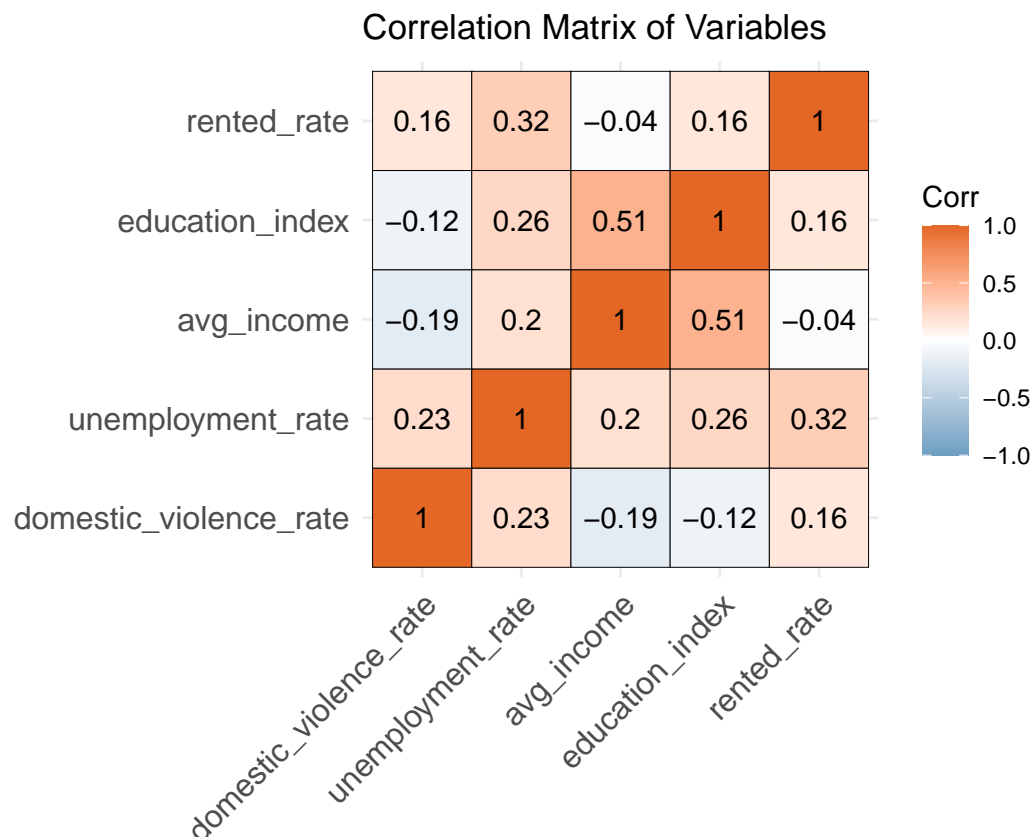


Table 6: Table: Variance Inflation Factors for Independent Variables

	Variable	Variance Inflation Factor (VIF)
unemployment_rate	unemployment_rate	1.192727
avg_income	avg_income	1.406979
education_index	education_index	1.439199
rented_rate	rented_rate	1.157056

Correlation Matrix and Variance Inflation Factor (VIF) above are used to validate the model of multivariate logistic regression. The correlation matrix and VIF analysis results show that the overall validity of the model is high. The correlations between the dependent variable and the arguments are theoretically expected, and the direction and strength are consistent with the literature, which indicates that the unemployment rate and rental housing rate are positively correlated with domestic violence, and income and education level are negatively correlated. The low correlation between the arguments and the VIF value of less than 2 indicate that the model does not have serious multicollinearity problems.

## Predicted vs. Actual Domestic Violence Rate by Year (GAMM)

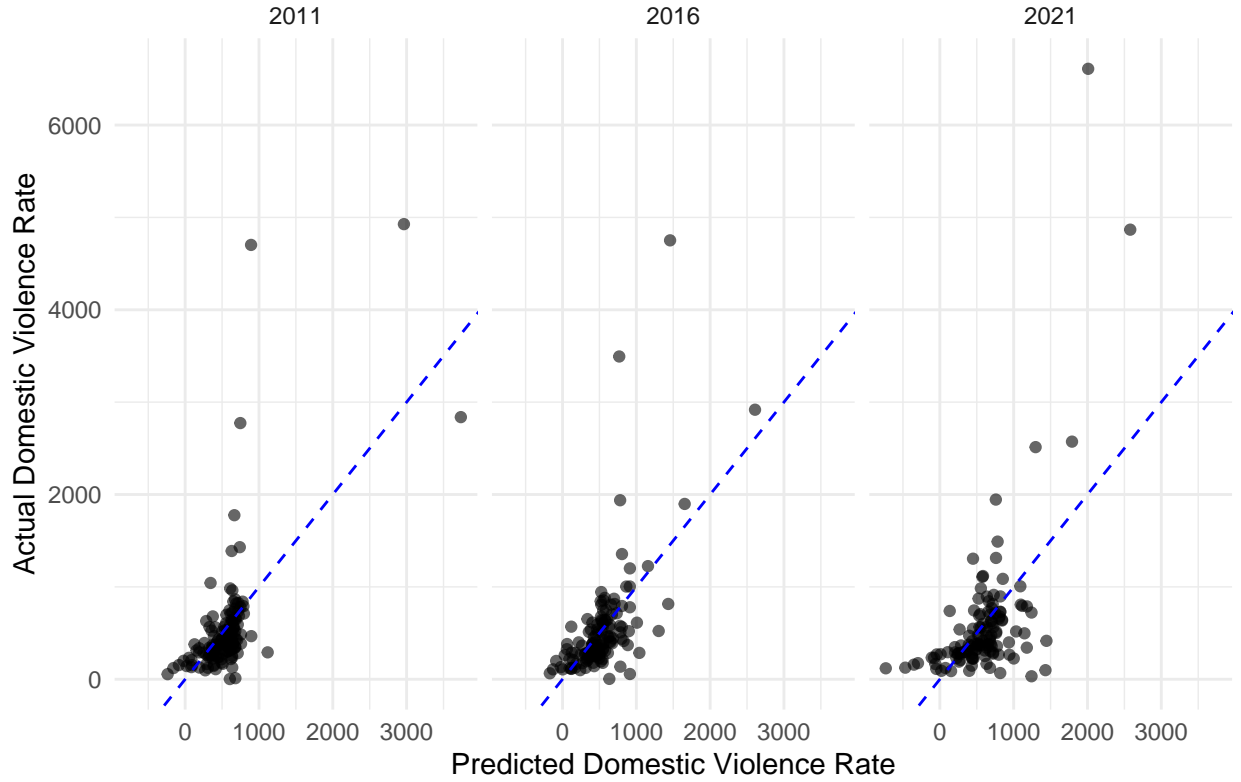


Table 7: Table: Parametric Coefficients of the GAM Model

Predictor	Estimate	Standard Error	t-Value	P-Value
(Intercept)	558.6772	28.76708	19.42071	0

Table 8: Table: Smooth Terms of the GAM Model

Predictor	Effective Degrees of Freedom (edf)	Reference Degrees of Freedom (Ref.df)	F-Value	P-Value
s(unemployment_rate)	4.774761	5.888151	4.176864	0.0004808
s(avg_income)	6.750313	7.866730	5.362840	0.0000035
s(education_index)	5.856162	7.036610	3.667714	0.0007395
s(rented_rate)	2.690942	3.400603	7.195570	0.0000575

Table 9: Table: Performance Metrics of the GAM Model

Adjusted R-Squared	Deviance Explained (%)
0.30073	33.6812

The analysis showed that GAMM outperformed multiple logistic regression in explaining the differences in domestic violence rates, with an adjusted R square of 0.30, while the maximum R square of the logistic model in 2016 was 0.264. GAMM captured 33.68% of the difference, demonstrating its ability to simulate complex nonlinear relationships between domestic violence rates and predictors such as unemployment rates,

average income, education index, and rental rates. It is noteworthy that predictors such as average income and education index have strong non-linearity and high effective degrees of freedom (EDF), indicating that the impact of these factors on domestic violence rates is more complex than previously assumed.

## Conclusion

The study found a significant association between socio-economic factors in New South Wales and levels of domestic violence. Unemployment and rent levels were positively related with domestic violence, while income and education levels were negatively related. The results of this research support this assumption and are in accordance with existing literature. For example, economic and housing anxiety increase the risk of domestic violence, while income and education levels were found to be with less risk.

However, the limitation of this research is that the  $r$  square of the results obtained by the models used is not very high, which indicates that the models used so far do not fully explain the rate of domestic violence. This suggests that there may be other important factors not explored in this article. At the same time, this research should also consider the influence of more real environment and uncontrollable variables. For example, after the covid-19 epidemic in 2021, many variables in this environment may have deviated, that is, there will be an important variable with a particularly high weight during this period, which is to be discovered in the future.

(2490 words)

## Bibliography

1. Attah, L. L., & Owoyemi, J. O. (2024). Public Perception of Informal Strategies for Curbing Domestic Violence against Women in Ofu Local Government Area of Kogi State, Nigeria. *International Scholar Journal of Arts and Social Science Research*, 7(1).
2. Australian Bureau of Statistics (ABS). (2021). Australian Statistical Geography Standard (ASGS), Edition 3, Digital Boundary Files. Retrieved from <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files>.
3. Australian Bureau of Statistics (ABS). (2021). Census DataPacks: Guide to Census Data. Retrieved from <https://www.abs.gov.au/census/guide-census-data/about-census-tools/datapacks>.
4. Buzawa, E. S., & Buzawa, C. G. (Eds.). (2017). *Global responses to domestic violence*. Springer International Publishing.
5. Dowling, N. A., Jackson, A. C., Suomi, A., Lavis, T., Thomas, S. A., Patford, J., . . . & Bellringer, M. E. (2014). Problem gambling and family violence: Prevalence and patterns in treatment-seekers. *Addictive Behaviors*, 39(12), 1713-1717.
6. Grech, K., & Burgess, M. (2011). Trends and patterns in domestic violence assaults: 2001 to 2010. Issue paper.
7. Jiang, H., Riordan, B., Laslett, A. M., Livingston, M., Lee, K., James, D., . . . & Room, R. (2023). Final report Association of liquor outlet density with domestic and non-domestic assault in New South Wales.
8. Mitchell, L. (2011). Domestic violence in Australia: an overview of the issues.
9. New South Wales Bureau of Crime Statistics and Research (BOCSAR). (2024). Criminal Offences Data. Retrieved from <https://bocsar.nsw.gov.au/statistics-dashboards/open-datasets/criminal-offences-data.html>.
10. Payne, J., & Morgan, A. (2020). COVID-19 and Violent Crime: A comparison of recorded offence rates and dynamic forecasts (ARIMA) for March 2020 in Queensland, Australia.
11. Payne, J. L., Morgan, A., & Piquero, A. R. (2020). COVID-19 and social distancing measures in Queensland, Australia, are associated with short-term decreases in recorded violent crime. *Journal of experimental criminology*, 1-25.

12. Smyth, C., Cullen, P., Breckenridge, J., Cortis, N., & Valentine, K. (2021). COVID-19 lockdowns, intimate partner violence and coercive control. *Australian Journal of Social Issues*, 56(3), 359-373.
13. Stavrou, E., Poynton, S., & Weatherburn, D. (2016). Intimate partner violence against women in Australia: related factors and help-seeking behaviours. NSW Bureau of Crime Statistics and Research.
14. Woods, J. A., Ward, A. C., Greville, H. S., Moran, M. C., Nattabi, B., Martin, K. E., & Thompson, S. C. (2023). Measuring for primary prevention: An online survey of local community perspectives on family and domestic violence in regional Australia. *PLOS ONE*, 18(4), e0284302.