

What explains domestic violence rates in NSW?

Masud Jubaier

Introduction

Domestic violence (DV) is dark stain in modern society. DV can be defined as any form of mistreatment including physical, psychological, financial, sexual take place within domestic circumstances perpetrated by member of intimate relationship. Any member of household can be victim of DV, while women and children are most vulnerable to this crime. The World Health Organization (WHO) has described the violence against women is a global problem of public health which need immediate attention[1]. Australia is not Immune to this crime, especially NSW suffer with this crime heavily. Hence it is very important to understand the nature of this crime and identify the factors having most influence. In this project we shall try find these factors and try to prove following hypothesis by implementing a linear model.

Null Hypothesis (H_0): There is no relationship between income, education and employment with DV.

Alternative Hypothesis (H_A): There is strong correlation between income, education and employment with DV.

Literature and theory

In 2007 Drabsch T. stated that the crime of DV exists in all level of society. However, the true magnitude of the level of crime is impossible to measure as most incidents are not reported and in many instances the victim has a tendency to deny the crime ever happened[2].

In 2009 Morgan A. done an extensive research on DV. He has defined DV as acts of violence take place within intimate relationship in domestic environment. He has try to quantify the cost of the crime and found in FY2008-09 totalling \$13.6b and if no action taken by FY2021-22 the cost is estimated to be \$15.6B[3].

In 2014 Phillips J. has presented men are more likely to experience violence at the hand of strangers, while women are more like to experience at the hand of someone who are known to them[1]. In 2017 Stubbs J. identified that the DV was considered as private issue until 1970, when government and different activist group campaigned this is a social problem effecting mainly women[4].

In 2018 Aziz N. has developed a conceptional model on connecting the association of committer and victim factors that consequence in violence against women.

In 2000 Astbury J. has presented an very important finding, the idea of “Home” point toward physical and emotional safety and security, person affected by violence experiences a concealed homelessness[6].

In 1998 Gwenneth L. has done a study on sample of 335 women (mean age 45.5 years) from the Royal Brisbane Hospital Emergency Department to establish a solid scientific basis for recognition and treatment of post-traumatic psychiatric injury related with DV. The finding is resounding, about one-third of the psychiatric diagnoses can be attributed to DV[7].

In 2011 Richards K. has characterize, children witness DV is a form of child abuse. Children witness DV are likely to suffer Psychological and behavioural impacts in future[8]. In the same year Ghafournia N. has explored the status of abused migrant women in Australian migration policy. Although the policies

has been adopted in order to defend non-resident women from being battered by their partners due to their uncertain residence status, still some significant concerns related to social, economic, and cultural factors are overlooked[9].

In 2003 Mulrone J. has presented a survey using a representative sample of 6300 Australian women provided by ABS. It was found that 23% of women married or in a de-facto relationship had experienced violence by a partner[10].

Data and methodology

We have obtained following 3 CSV files

1. DV_NSW_by_LGA.csv
2. NSW_LGA.csv
3. labels.csv

The 1st CSV file contains the count of DV crime in NSW LGA from Jan-99 to Dec-15. The 2nd CSV file contain 2011 census data of NSW LGA. In the 2nd CSV file there are around 8K features on 2011 NSW LGA census data. The features are in encoded form, where the description of the features is found in the 3rd CSV file. At first we have performed some exploratory analysis to understand the dataset. We have generated some visualization to show the trend of the crime. From the visualization it is clear the trend is upward. To identify the features contributes most towards the crime we have implemented Forward stepwise Feature selection.

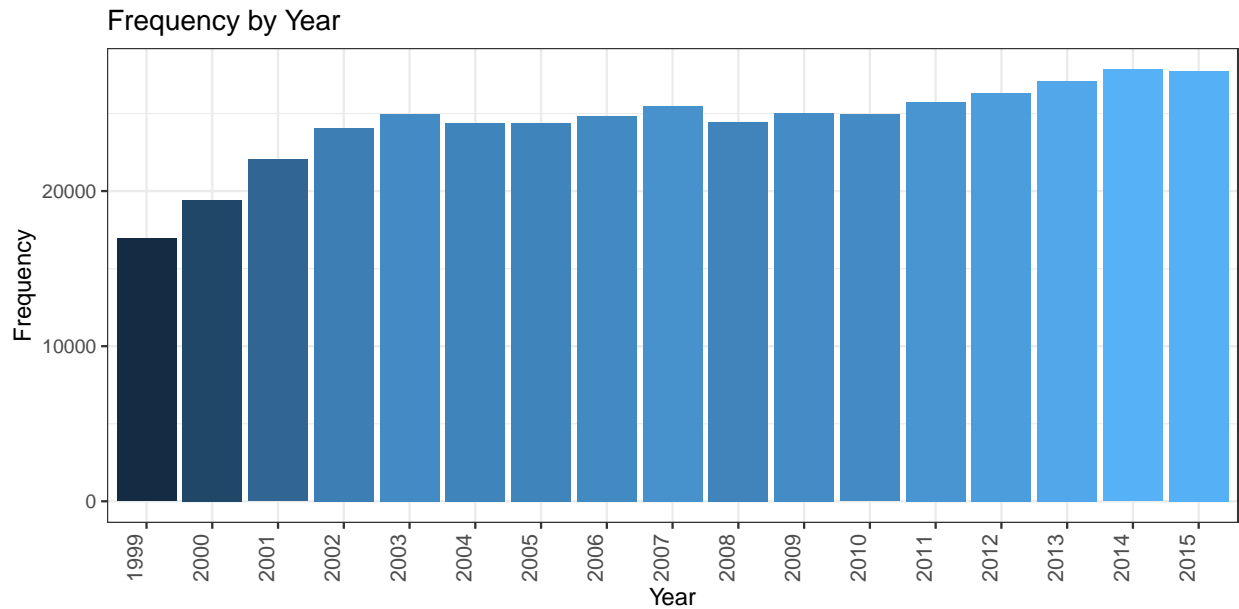
Forward stepwise Feature selection is a computationally efficient algorithm to select the best features that produces best outcome. The algorithm starts with a null model and add predictors one by one until all predictors are tried. At each step, the predictor that give the largest improvement on the model is added to the final predictor list. The steps of the algorithm are as follows.

1. Let, M_0 be a null model, which has no predictor.
2. For $k = 0, \dots, p - 1$:
 - (a) Consider all $p - k$ models that supplement the predictors in M_k with one added predictor.
 - (b) Chose the *best* between these $p - k$ models, and name it M_{k+1} . Here *best* is considered as the model having minimum RSS or maximum R^2 .
3. Select the best model from M_0, \dots, M_p using any one of these metric cross-validated prediction error or C_p (*AIC*) or (*BIC*) or adjusted R^2 [11].

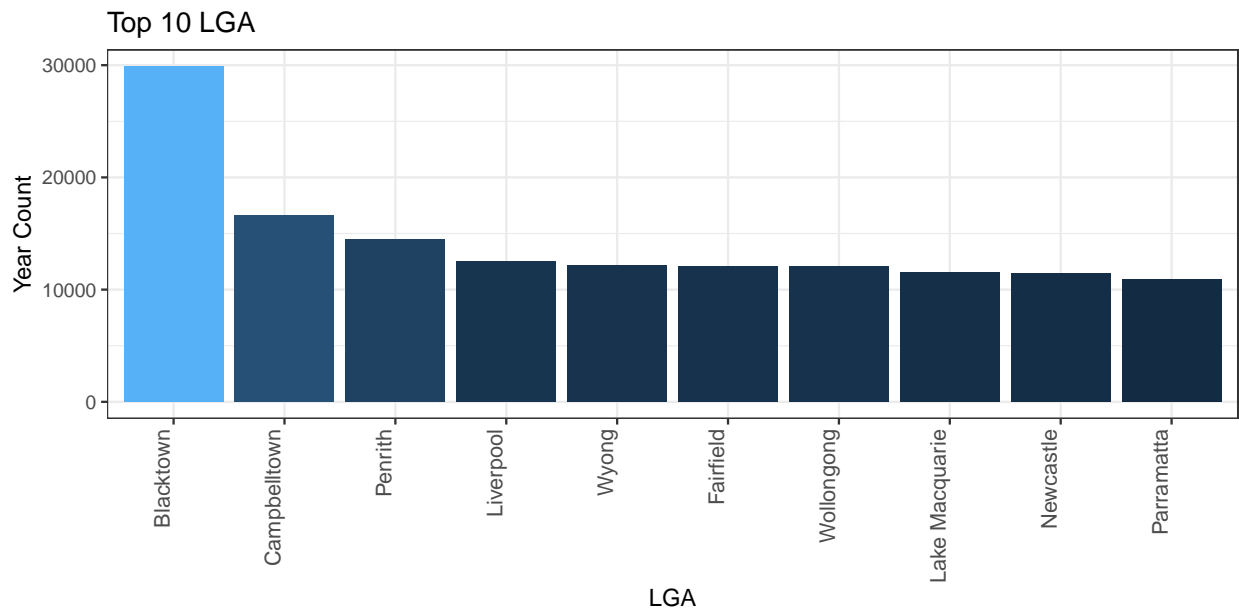
Using the features identified we shall build a linear model to predict DV.

Results and implications

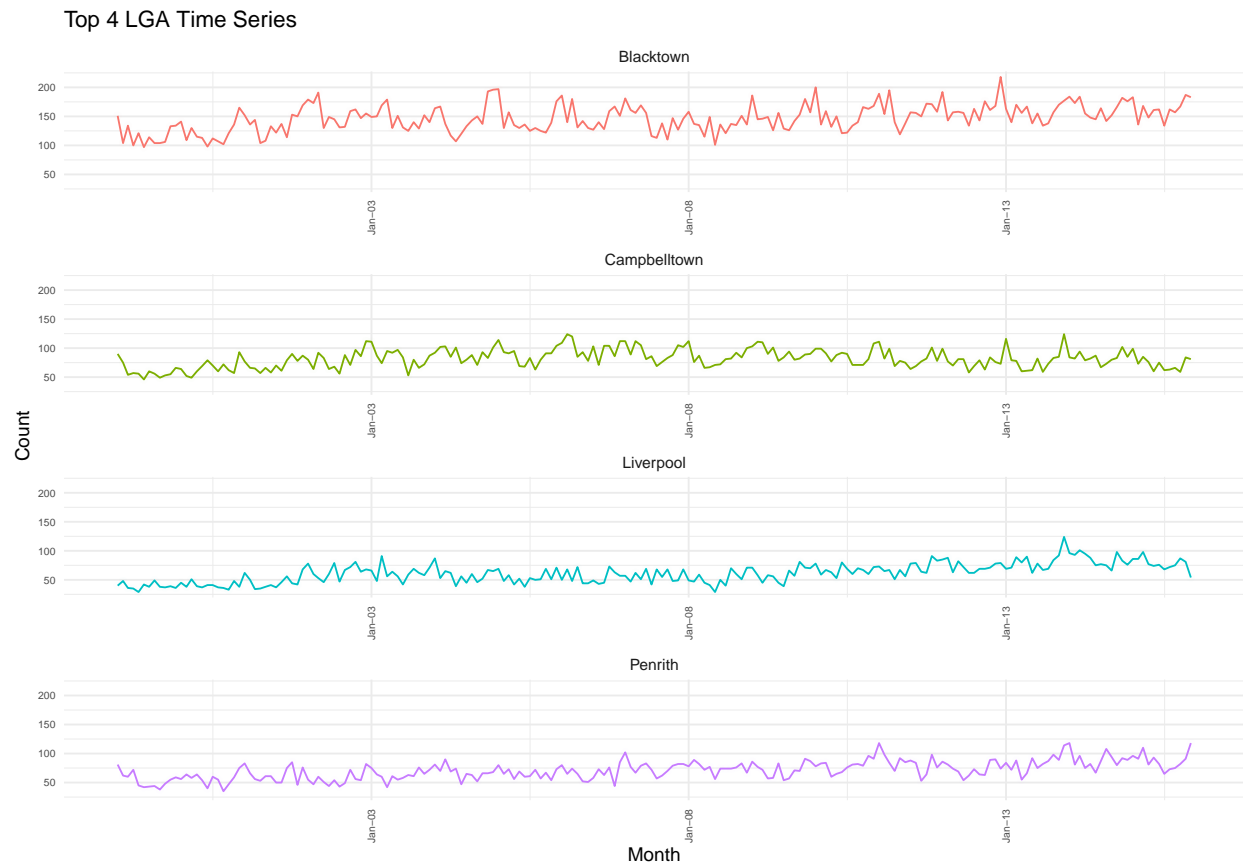
The plot below visualizes yearly frequencies of all across NSW. From the plot it is clear the crime rate has an upward trend.



The plot below visualizes the frequency of top 10 LGA in descending order. Blacktown LGA outcast all other LGA by very clear margin. The concern authority might need to pay extra attention on Blacktown.



The plot below visualizes the frequency timeseries of top 4 LGA. The timeseries plot shows normal distribution of the frequency, no sudden abrupt spike or drop of the frequency.



After applying the Forward stepwise Feature selection algorithm, the features selected by the algorithm are listed in the table below. From the feature list we can see, no features on income education and employment as our hypothesis are selected by the algorithm. Therefore, we are manually adding feature B115, B126, B5503 and B2847 in the predictor list covering these aspects. With this combined predictor list, we shall build a linear model.

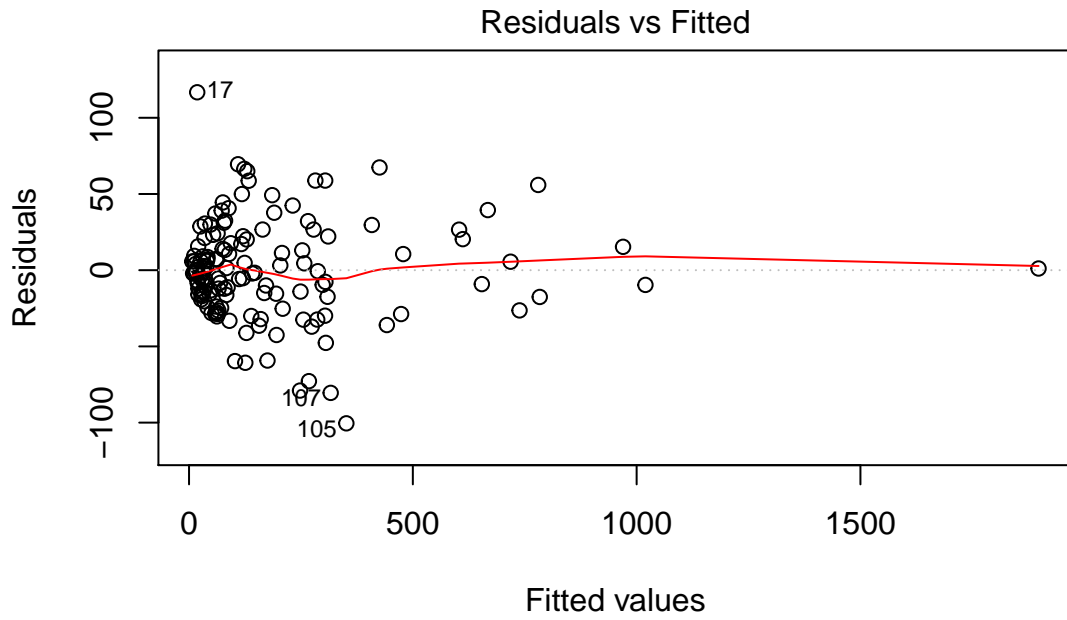
We are building a linear model with the selected features. The summary of the linear model as presented below.

```
##
## Call:
## lm(formula = DV_Count_2011 ~ B115 + B126 + B5503 + B2847 + B4849 +
##      B4702 + B2252 + B4258 + B2894 + B1758 + B3851 + B7452 + B2409 +
##      B1390, data = DV_NSW_2011_all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -100.473  -17.105   -1.651   18.293  116.720
##
## Coefficients:
```

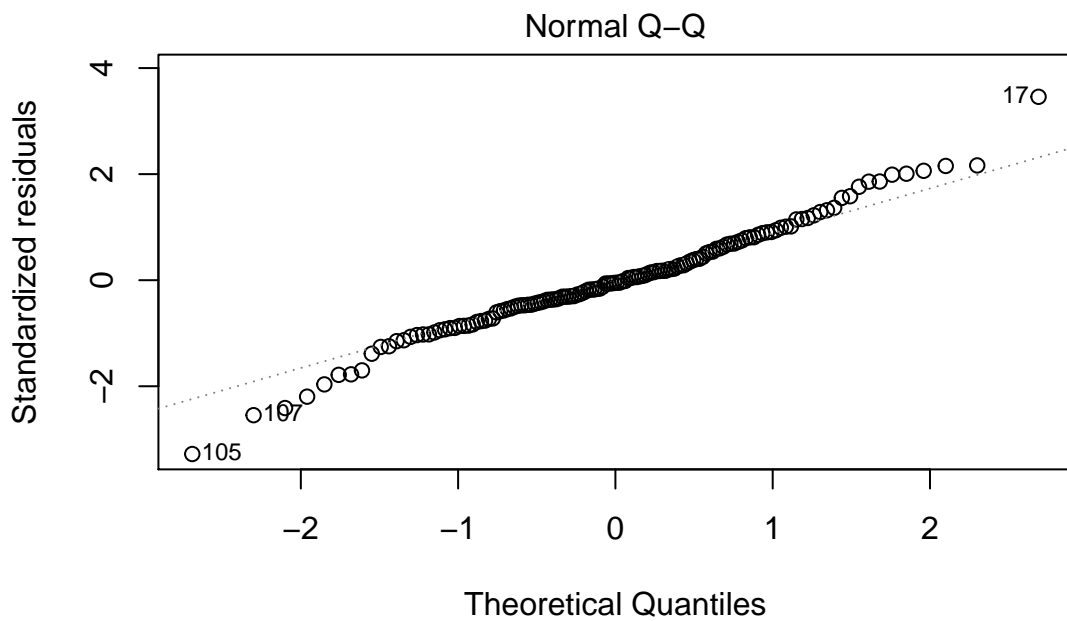
Sequential	Long
B1390	Cambodia Year of arrival 2011
B1758	Poland Year of arrival 2011
B2252	Dependent children aged 5 9 years female parent Language and proficiency in English not stated male parent Speaks English only
B2409	Dependent children aged 15 17 years female parent Total male parent Speaks other language and speaks English Proficiency in English not stated
B2894	Males Year 11 or equivalent Age 15 19 years
B3851	Females 15 19 years Did unpaid domestic work 30 hours or more
B4258	Persons 20 24 years Cared for Total
B4702	Age group of parent 25 29 years Number of children ever born Six or more children
B4849	One parent family with children under 15 and no dependent students and non dependent children Persons
B7452	Professional scientific and technical services Occupation Labourers

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 41.0157167 32.3057487   1.270 0.206582
## B115         0.0015837  0.0141017   0.112 0.910762
## B126        -0.0002338  0.0005186  -0.451 0.652966
## B5503        -0.6446707  0.7519112  -0.857 0.392878
## B2847         0.0306113  0.0207253   1.477 0.142191
## B4849         0.2614197  0.0394617   6.625 9.30e-10 ***
## B4702         7.7695686  1.6862621   4.608 9.90e-06 ***
## B2252         9.1174831  2.3752840   3.838 0.000196 ***
## B4258         0.1956317  0.0460197   4.251 4.13e-05 ***
## B2894        -0.8215502  0.1847910  -4.446 1.91e-05 ***
## B1758        11.3768055  3.1433709   3.619 0.000427 ***
## B3851         2.3630352  0.5996256   3.941 0.000134 ***
## B7452         0.8505166  0.3691961   2.304 0.022891 *
## B2409        -3.4327696  1.7094242  -2.008 0.046784 *
## B1390        -2.5787578  0.9509693  -2.712 0.007636 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.06 on 125 degrees of freedom
## Multiple R-squared:  0.9831, Adjusted R-squared:  0.9812
## F-statistic: 519.3 on 14 and 125 DF,  p-value: < 2.2e-16
```

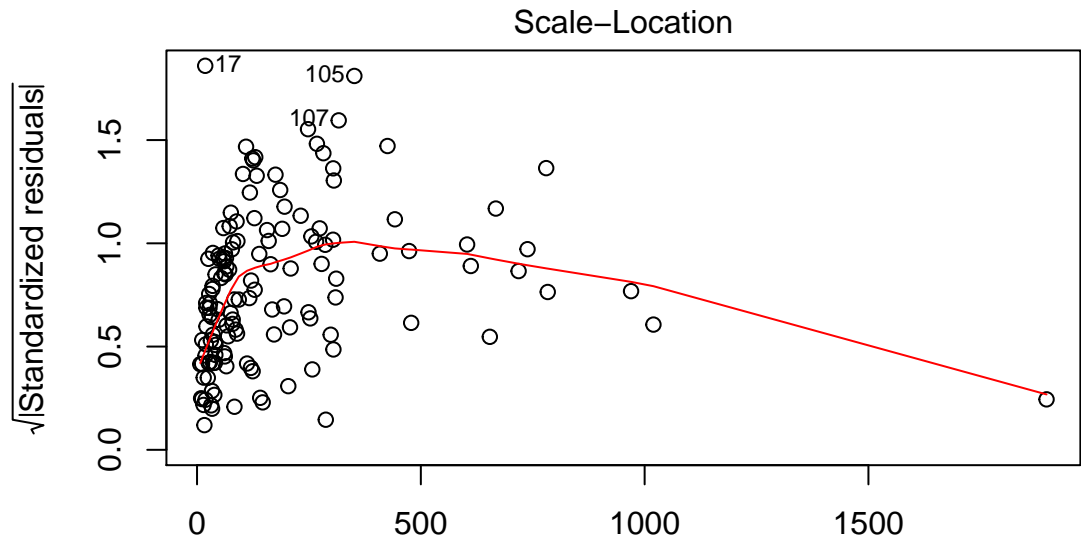
The model summary shows, the features added manually B115, B126, B5503 and B2847 has P-Value greater than 0.05, hence these are not a good predictor in the model. Therefore, we cannot reject H_0 , i.e. There is no relationship between income, education and employment with DV. The features selected by Forward stepwise Feature selection algorithm are the best predictor to predict DV. The regression summary we can see these features have very low P-Value indicating the features having very strong relationship with the response variable. This fact is marked by *** in the summary. Also, the Multiple R-squared of the model returned 0.9831, indicating the predictors can explain 98.12% of variance from the model.



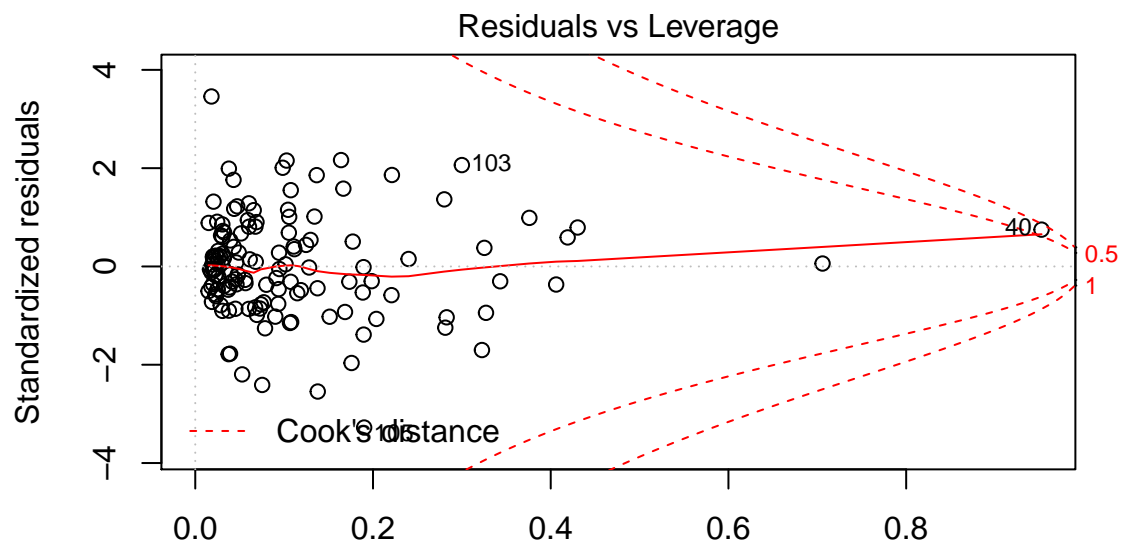
$n(\text{DV_Count_2011} \sim \text{B115} + \text{B126} + \text{B5503} + \text{B2847} + \text{B4849} + \text{B4702} + \text{B2252} +$



$n(\text{DV_Count_2011} \sim \text{B115} + \text{B126} + \text{B5503} + \text{B2847} + \text{B4849} + \text{B4702} + \text{B2252} +$



Fitted values
 $n(\text{DV_Count_2011} \sim \text{B115} + \text{B126} + \text{B5503} + \text{B2847} + \text{B4849} + \text{B4702} + \text{B2252} +$

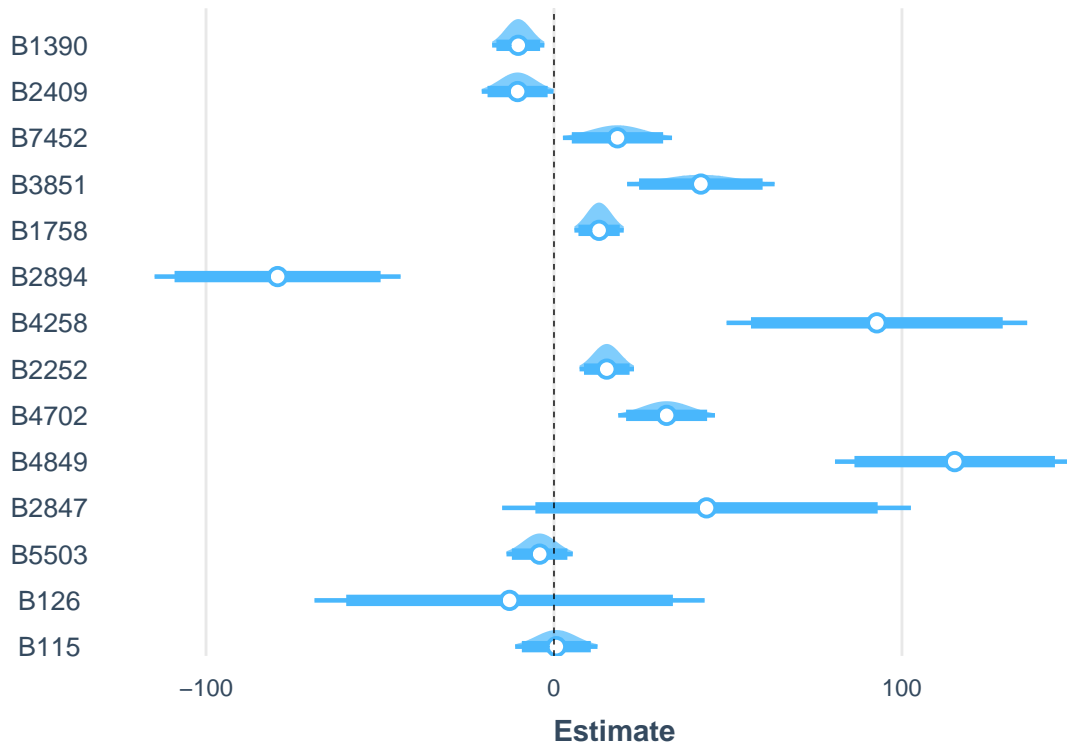


Leverage
 $n(\text{DV_Count_2011} \sim \text{B115} + \text{B126} + \text{B5503} + \text{B2847} + \text{B4849} + \text{B4702} + \text{B2252} +$

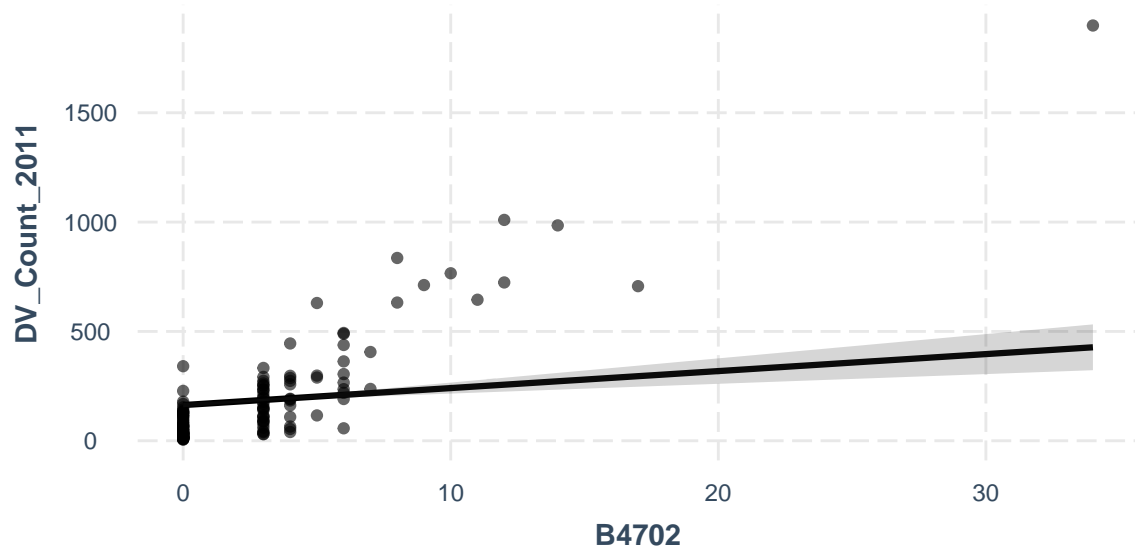
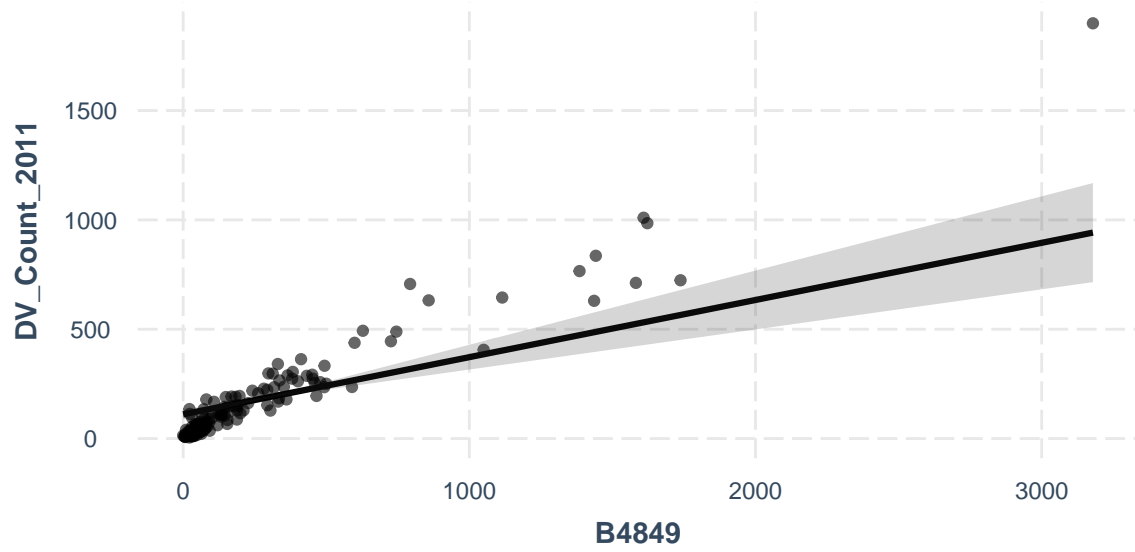
One of the best ways to present finding from a model is to generate `plot_summs` from `jtools` library. Using the plot, we can show the best predictor's coefficient estimations are normally distributed as presented in plot below.

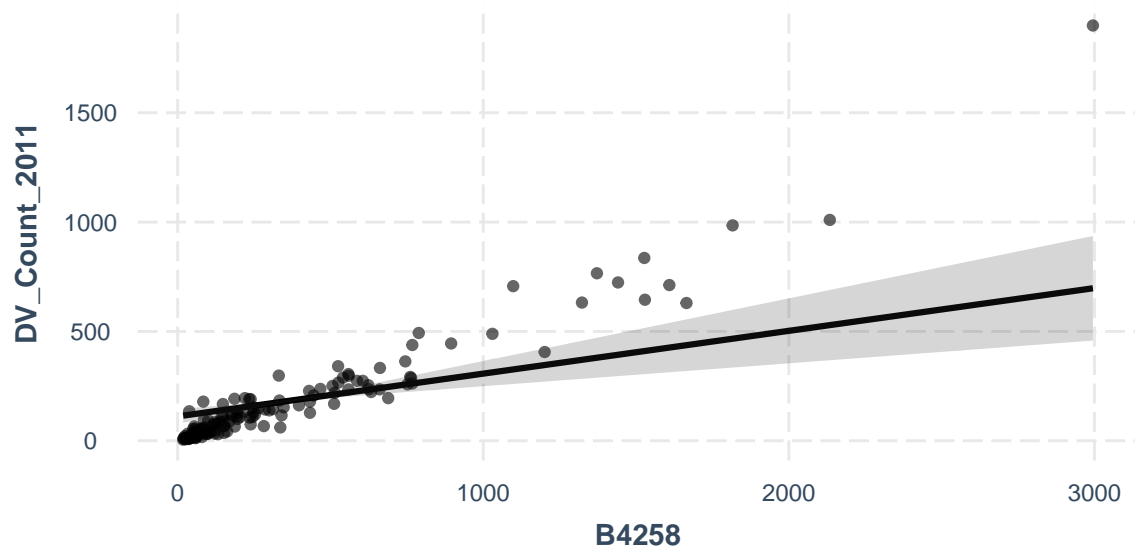
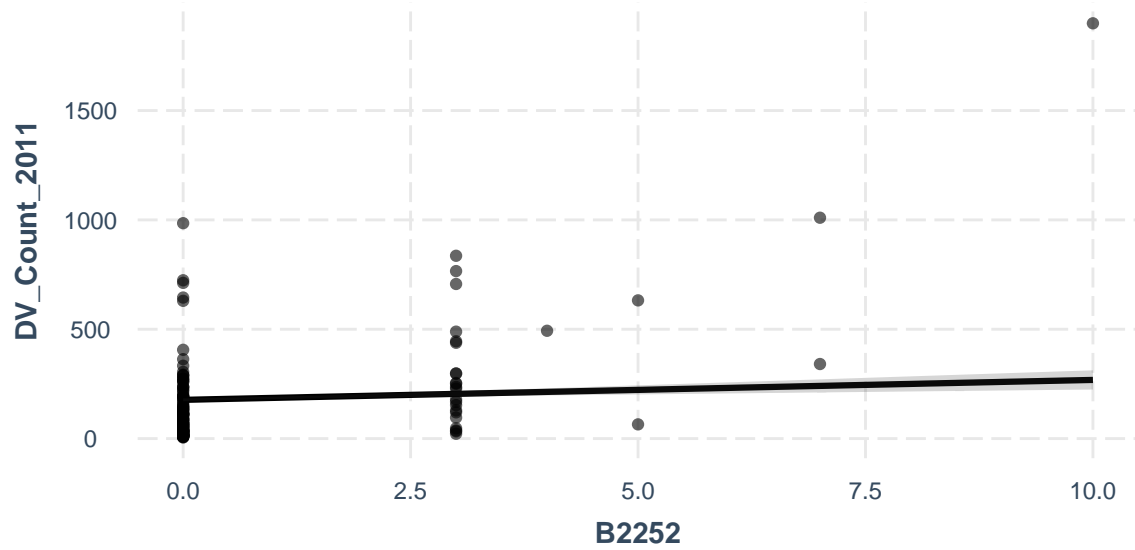
```
## Registered S3 methods overwritten by 'broom':
##   method      from
##   tidy.glht    jtools
##   tidy.summary.glht jtools

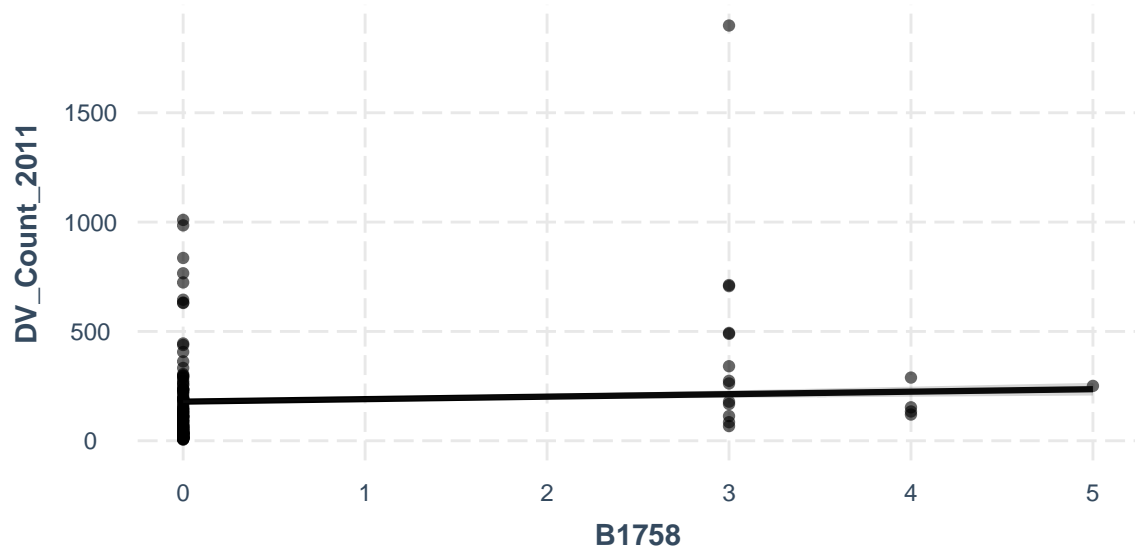
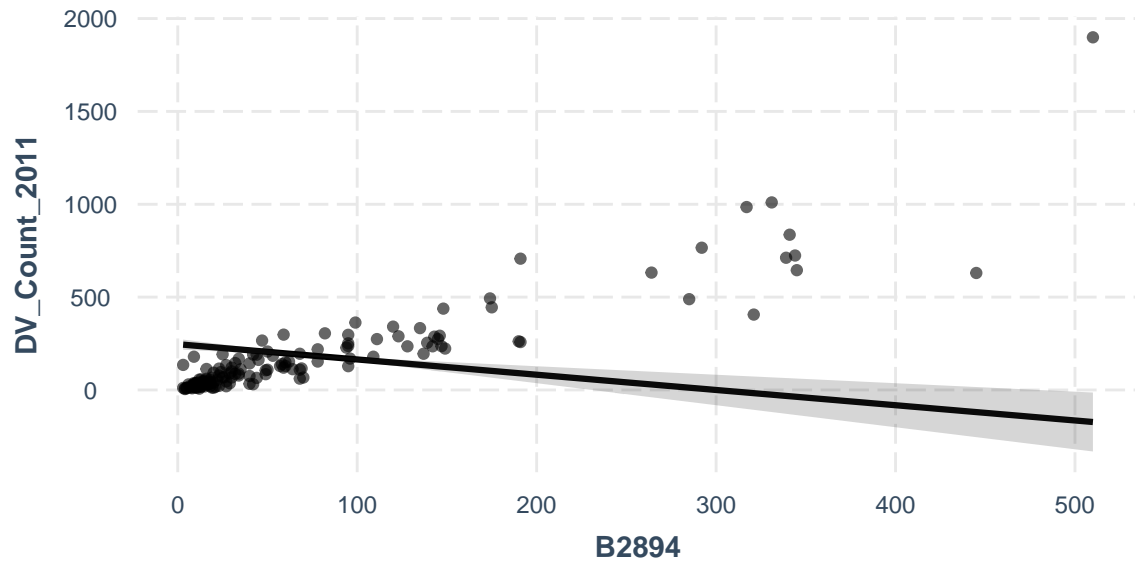
## Loading required namespace: broom.mixed
## Loading required namespace: broom.mixed
```

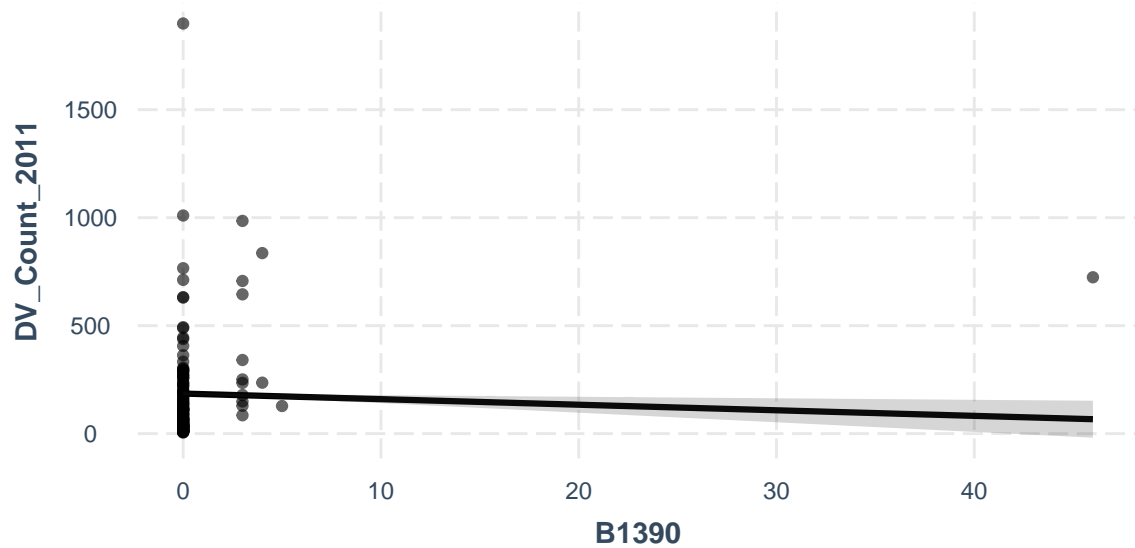
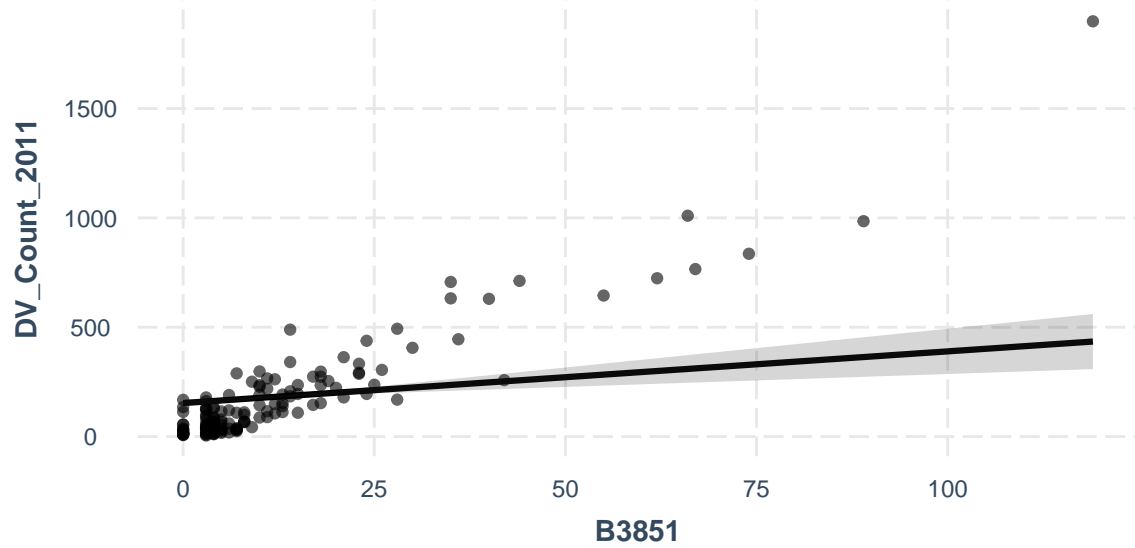


Another useful visualization is `effect_plot` to show the relationship between predictor and response variable. The predictors with low P-Value are visualized below. All these plot display very strong connection with the response variable.









Conclusion

DV is a social problem in NSW. The data we have used in this project, clearly shows the level of crime does not have a downward trend. Policy maker and law enforcing agencies need to have more clear insight on the crime, so that they can come up with a policy to reduce this crime. This project can add a very good value in this regard. We have shown the trend of the crime and the area where the crime is happening the most. We have also identified some key predictors and build a regression model that can predict the crime very with very high level of confidence. The data we have worked on is not enough to build a complete model. We had only census data of 2011. In future we shall add more census and DV data and try to build a complete model that can predict DV with higher accuracy.

Bibliography

- [1]Phillips J. and Vandenbroek P.(2014). Domestic, family and sexual violence in Australia: an overview of the issues. Department of Parliamentary Services, Parliament of Australia.
- [2]Drabsch T.(2007). Domestic Violence in NSW. NSW Parliamentary Library Research Service.
- [3]Morgan A. and Chadwick H.(2009). Key issues in domestic violence. Australian Institute of Criminology. Australian Government.
- [4]Stubbs J. UNSW, Wangmann J. UTS(2017). Australian Perspectives on Domestic Violence.
- [5]Aziz N. N. A., Ishak M., Wahid N. A., Yazid Z. N. A.(2018). Factors affecting domestic violence against woman: A conceptual model and research propositions. Universiti Teknologi MARA.
- [6] Astbury J., Atkinson J., Easta P., Turner J. (2000). The impact of domestic violence on individuals. The Medical journal of Australia. November 2000.
- [7] Gwenneth L. Roberts and Joan M. Lawrence. (1998). The impact of domestic violence on women's mental health. University of Queensland.
- [8] Richards K. (2011). Children's exposure to domestic violence in Australia. Australian Institute of Criminology.
- [9] Ghafournia N. (2011). Aggression and Violent Behavior. University of Sydney.
- [10] Mulroney J. (2003). Australian Statistics on Domestic Violence. Australian Domestic and Family Violence Clearinghouse. [11] James G., Witten D., Hastie T, Tibshirani R. (2013). An Introduction to Statistical Learning