DATA5207: Major Research Project

Alvis Palmucci | 460270999

17 June 2018

A research study on the predictors of domestic violence in NSW

Aim: The aim of this study was to explore the relationship between a number of economic, financial stress and family status predictors, including male unemployment, weekly income, sole parent households and its impact on domestic violence incidents in NSW.

Method:

The study used data from the Australian Census provided by the Australian Bureau of Statistics for the years 2006, 2011 and 2016. A generalised linear Poisson model was used to study the relationship male unemployment, weekly income, sole parent households and its impact on domestic violence incidents in NSW Local Government Areas (LGA's).

Results:

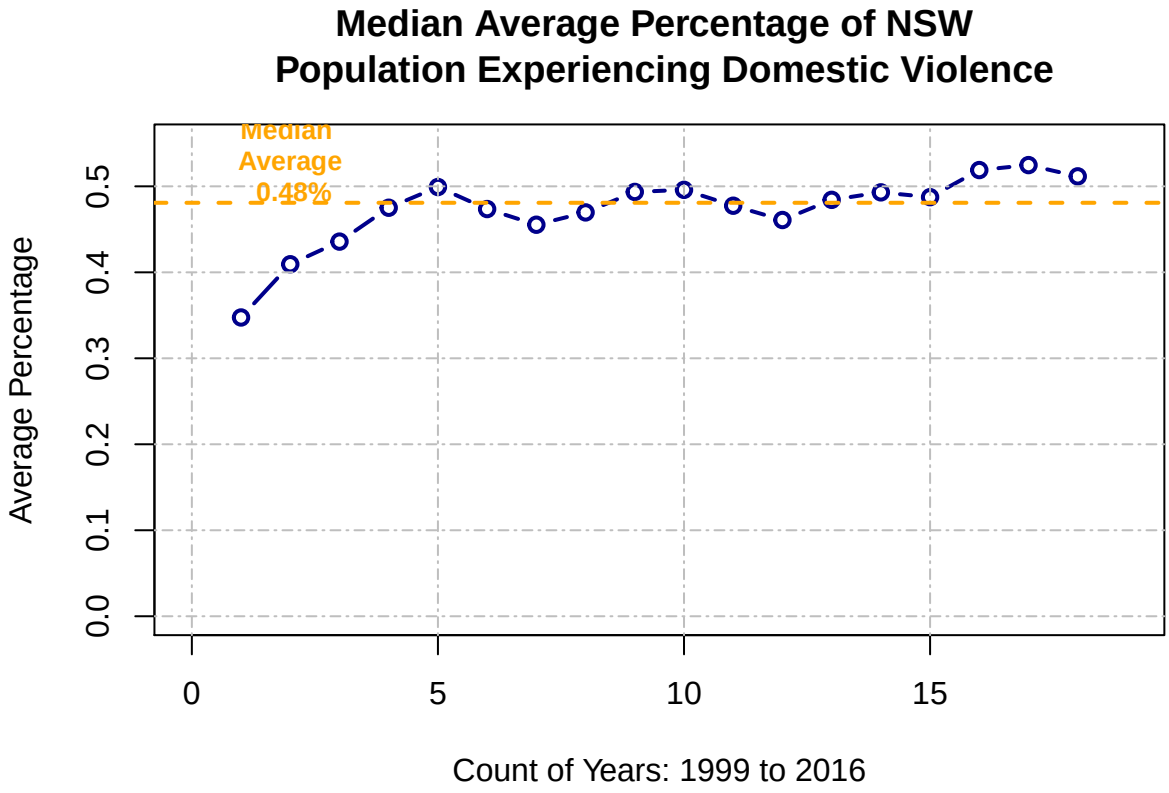
The risk of domestic violence was significantly higher for those areas where the male was unemployed, the victim is a sole parent and weekly personal income is between $800 and $1249. The results reveal a 69.7% ($\exp\{0.52888\} = 1.697$) increased risk of domestic violence incidents per 1000 people in NSW compared to those people not exposed to these factors.

Conclusion:

Economic situation, financial stress and family status predictors are strongly correlated variables for the incidents of domestic violence in NSW Local Government Areas. Further research however would be necessary to better understand if these factors are causal or there are other confounding factors at play as a result of domestic violence events.

Introduction

Domestic violence is a major concern for Australia's local and state authorities and under the Australian Government's National Plan to Reduce Violence against Women and their Children 2010-2022, it defines domestic violence as acts of violence that occur between people who have or have had, an intimate relationship (Dunkley and Phillips, 2015, pp.1). This is further defined by The Health Costs of Violence report (2003, pp.5) as any form of intimate partner violence that can occur on a continuum of economic, psychological and emotional abuse, through physical and sexual violence. As demonstrated in the chart below levels of domestic violence in NSW have increased in frqeuncy from 2006 to 2016:



Cox (2015) cites a number of risk factors with women aged 25-34 years experiencing the highest rate of male intimate domestic violence of all age groups. Furthermore, women from non-English speaking countries experienced domestic violence at double the rates compared to those from English speaking countries. Stavrou et al. (2016) cite sole parents, financial stress and those with a disability or long-term health condition as also being at risk. The latest Domestic Violence Statistics for NSW report (2013-2017) indicates alcohol was involved in almost one-third of all domestic violence incidents, while the male unemployment rate and the level of residential instability are also cited as key risk factors (People, 2005).

Data and Methodology

The study sourced data on male unemployment, weekly income and sole parents from the Australian Census provided by the Australian Bureau of Statistics for the years 2006, 2011 and 2016. Data from Census 2001 was not available from the ABS Census TableBuilder and has therefore been excluded from this study[1]. The assumption made to compile data for the years in-between Census was populated as a percentage of the moving average growth rate from one Census date to the next. Due to the fact 2001 Census data was not available, domestic violence incidents were analysed from 2006 through to 2016. The additional year of domestic violence data was sourced from the NSW Bureau of Crime Statistics and Research (BOSCAR). Additional data from BOSCAR, in particular alcohol related domestic violence incidents, was unavailable for the required time frame and was therefore not included in the study. Estimated resident population count data was obtained from the Australian Bureau of Statistics for each year of analysis.

One issue which arose when analysing Census data from 2006 to 2016 was the change in council regions in 2016 where a number of different councils were amalgamated. To counter this, all historical data prior to 2016 was matched to the new LGA regions so the results are relevant and reproducible in the future. Refer to Appendix.

This research study and other research studies which rely on Census data are however limited by the ability to effectively understand economic and financial stress indicators, such as employment status and weekly income, making the general assumption those on lower incomes and those who are unemployed experience more financial stress than those on higher incomes or with jobs (Weatherburn, 2011). Weatherburn (2011) claims "this may be true as a rough generalisation but the level of financial stress experienced by an individual depends not just on their income but also on their financial commitments and liabilities... and is best measured in terms reflective of the gap between income and expenditure."

Considering the type of count data available and its inherent distribution characteristics a generalised linear Poisson model with a link log was the first model considered. The Poisson model was considered due to the relationship between the mean and the variance and the nature of count data where the variance typically increases with the mean (Christensen, 1997). The disadvantage of a glm Poisson model is that while the linear predictor can represent any real world value, the Poisson mean, which represents a count, has to be non-negative. Given this a simple solution is to log transform the mean and assume the transformed mean follows a linear pattern. As cited by Lillis (2015) one further drawback to this model, which was experienced in this study, is that it can suffer from over-dispersion which occurs when the variance is larger than the mean in our dependent variable, or in other words, the residual variance is larger than the conditional mean. Lillis recommends the application of a Quasi-Poisson model fitting an extra dispersion parameter to account for the extra variance and confirming the over-dispersion.

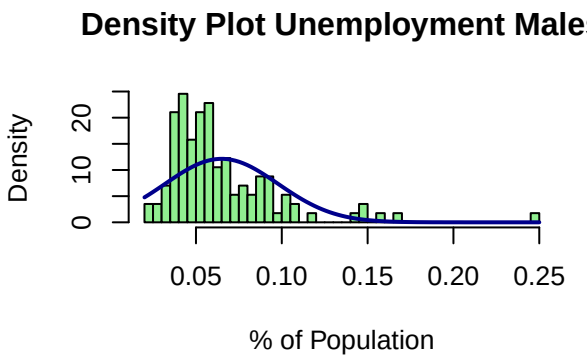
In an attempt to address this over-dispersion a Negative Binomial Regression model has been applied, which unlike the Poisson model, assumes the mean and the variance are not equal. As explained by Kleiber et al.(2016) "as the dispersion parameter gets larger and larger, the variance converges to the same value as the mean, and the negative binomial turns into a Poisson distribution".

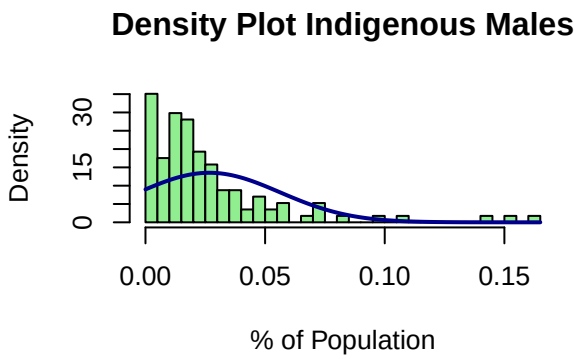
Finally, to improve the validity of the data to be integrated into the model any NSW LGA regions which had domestic violence counts outside of the interquartile range were excluded from the analysis and considered outliers. A total of 45 LGA regions and their data were included in the analysis.

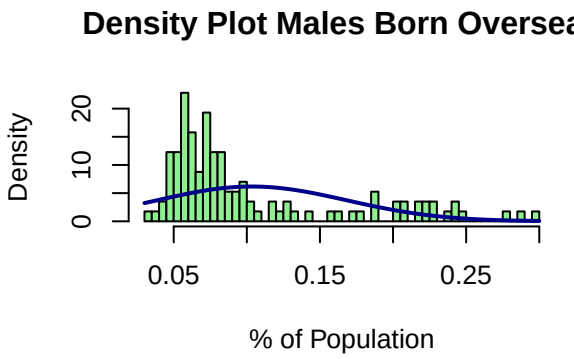
Results and Implications

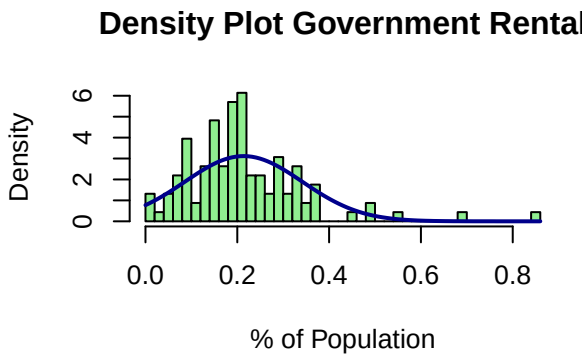
Understanding the distribution of the independent | predictor variables

After consulting numerous literacy studies as mentioned earlier in the report, a number of independent or predictor variables were considered as having a potential relationship to the dependent or outcome variable of domestic violence incidents in NSW local government areas. The first step to understanding the data was to plot the distribution of the data to understand its typical spread. Some examples of this are provided below where we can see a typical poisson type of distribution:





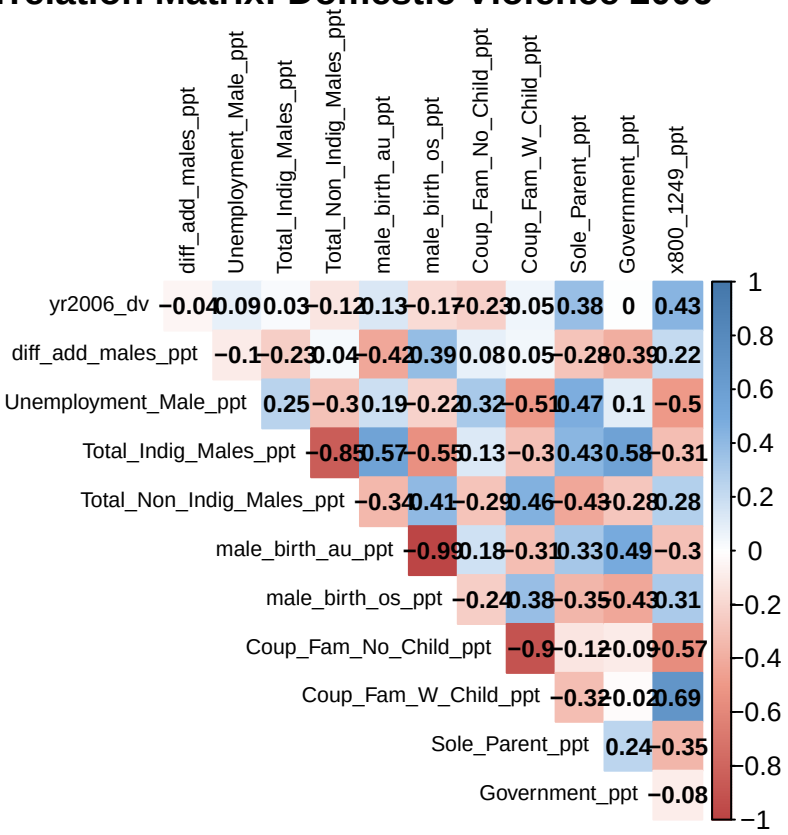


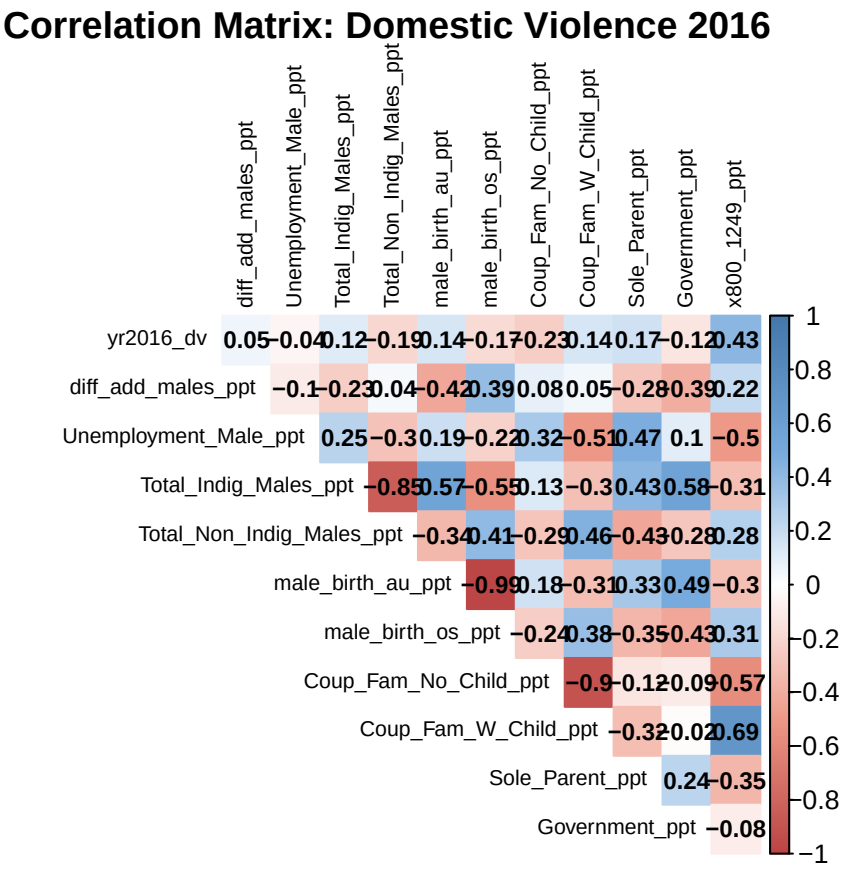


Understanding correlation domestic violence and independent variables

With our independent | predictor variables combined into the one dataframe the next step was to identify any possible correlation to domestic violence as well as any potential multicollinearity which may exist. Due to the possiblity of these factors changing over the time period of the domestic violence data (from 2006 to 2016) this was taken at a point in time in both years to see if there was evidence of any significant change. The two correlation matricies are shown below:

Correlation Matrix: Domestic Violence 2006



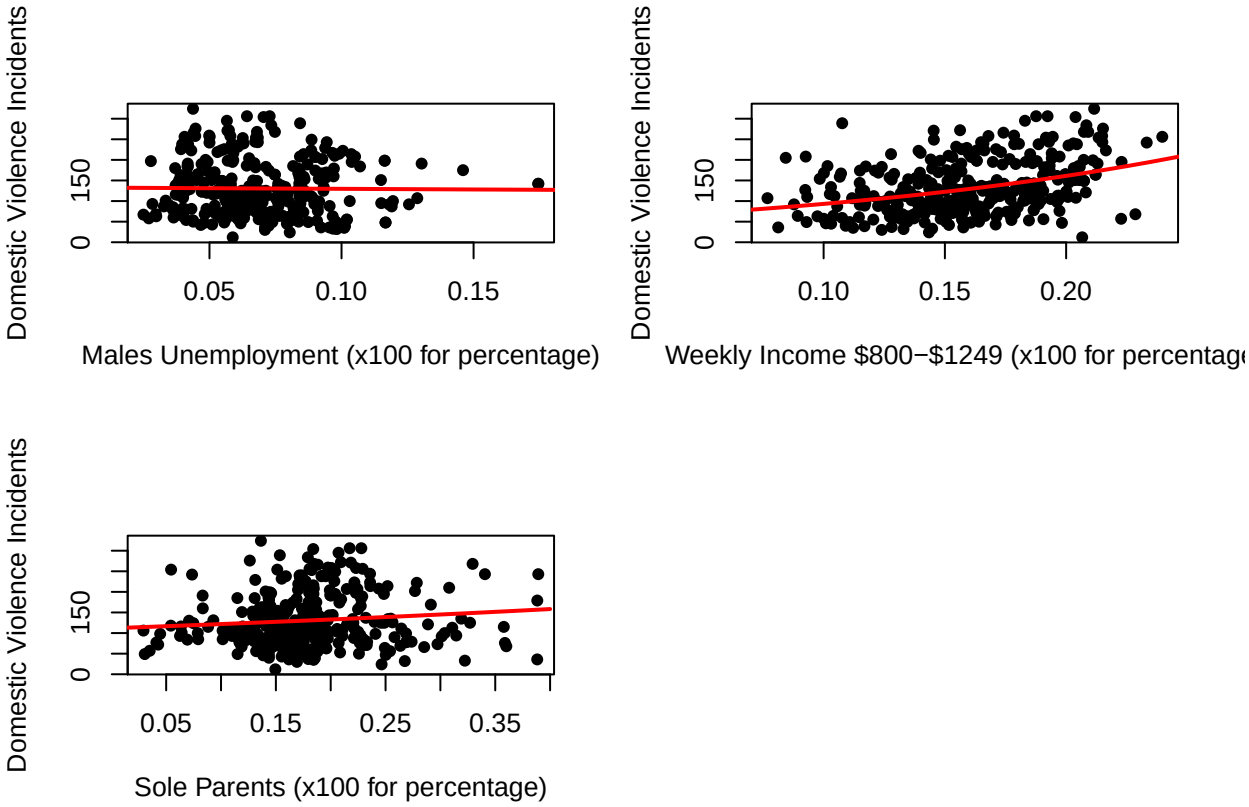


Those people earning a weekly income of between $800 and $1249 is the only factor which continued to show the highest level of correlation in both Census years (0.43). Those on Nil-to-low and higher incomes were found to show no positive correlation. Sole parents had the second highest level of correlation although this did drop from 0.38 in 2006 to 0.17 in 2016. Couples with children and Indigenous males were the only other two variables to show some movement, although minimal. Thus the key predictors brought into the model are Male Unemployment, Weekly Income between $800-$1249 and Sole Parents. Indigenous Males were not brought into the model due to the higher level of correlation to Male Unemployment (0.25).

Assessing model fit and residual deviance

A 70% training set was established and randomly shuffled to ensure randomness in the selection of the data. The training data set will also allow us to train our final model against a a final testing set of data following a 10 k-fold cross-validation to ensure our model does not over-fit the data. Using the training set data a Poisson model was created for each predictor variable in order to gain a better understanding of the effect of each predictor on the outcome variable. The output below demonstrates the effect of each of these predictor variables on domestic violence incidents with no relationship to Male Unemployment and a positive relationship to Weekly Income and Sole Parents.

```
## Using LGA, region_id as id variables
## Using LGA, region_id as id variables
## Using LGA, region_id as id variables
## Using LGA, region_id as id variables
## Using LGA, region_id as id variables
```



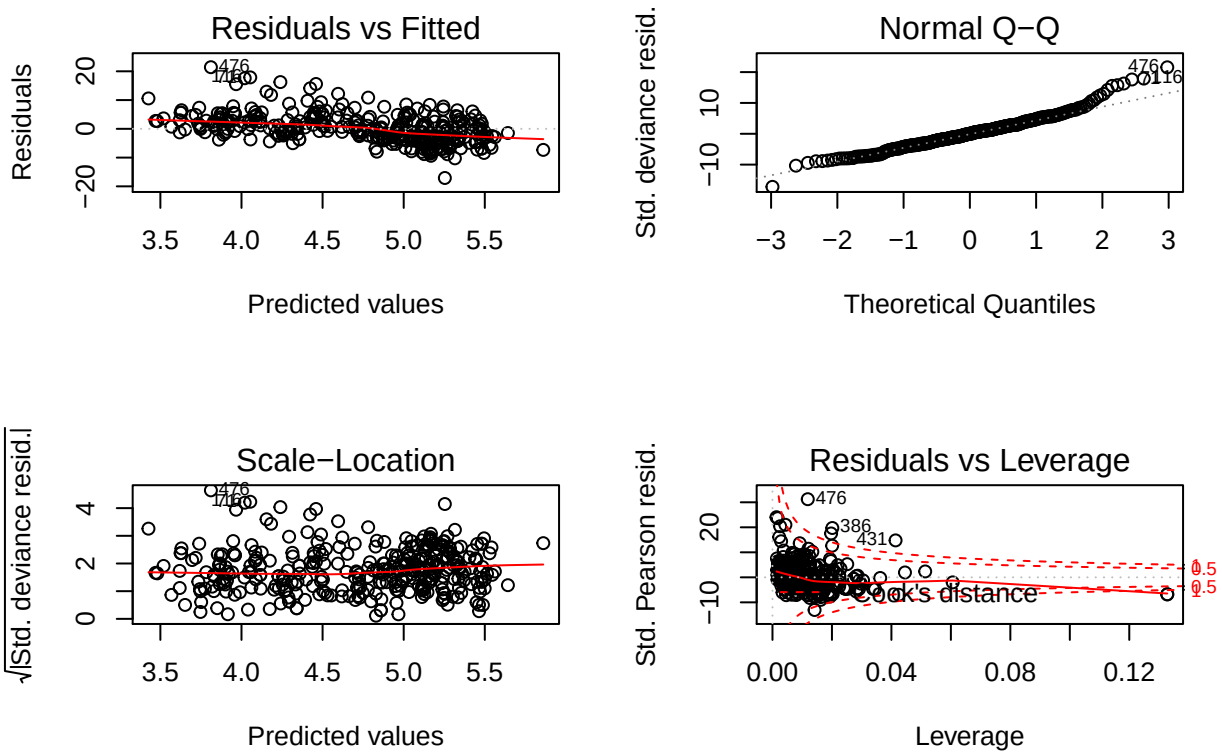
After analysing each of the predictor variables and the impact on domestic violence, all of the predictor variables are combined into the Poisson model. While each predictor variable is represented as a percentage of the population for each LGA a population count has been implemented as the offset on a per 1,000 person basis.

```
##
## Call:
## glm(formula = Domestic_Violence_Count ~ +M_Unemployment_ppt +
##     Income_x800_1249_ppt + Sole_parent_ppt + offset(log(Population_Count/1000)),
##     family = poisson(link = log), data = dataTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -17.090   -3.118    0.007    2.865   21.428
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.52888    0.03629  14.573  < 2e-16 ***
## M_Unemployment_ppt     5.99781    0.22443  26.725  < 2e-16 ***
## Income_x800_1249_ppt   0.64537    0.16742   3.855 0.000116 ***
## Sole_parent_ppt        2.31194    0.09361  24.698  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 10830.7  on 345  degrees of freedom
```

```
## Residual deviance:  8620.7  on 342  degrees of freedom
## AIC: 10907
##
## Number of Fisher Scoring iterations: 4
```

```
##           llh       llhNull            G2      McFadden          r2ML
## -5449.4121590 -6480.5385900  2062.2528620     0.1591112     0.9974208
##          r2CU
##     0.9974208
```

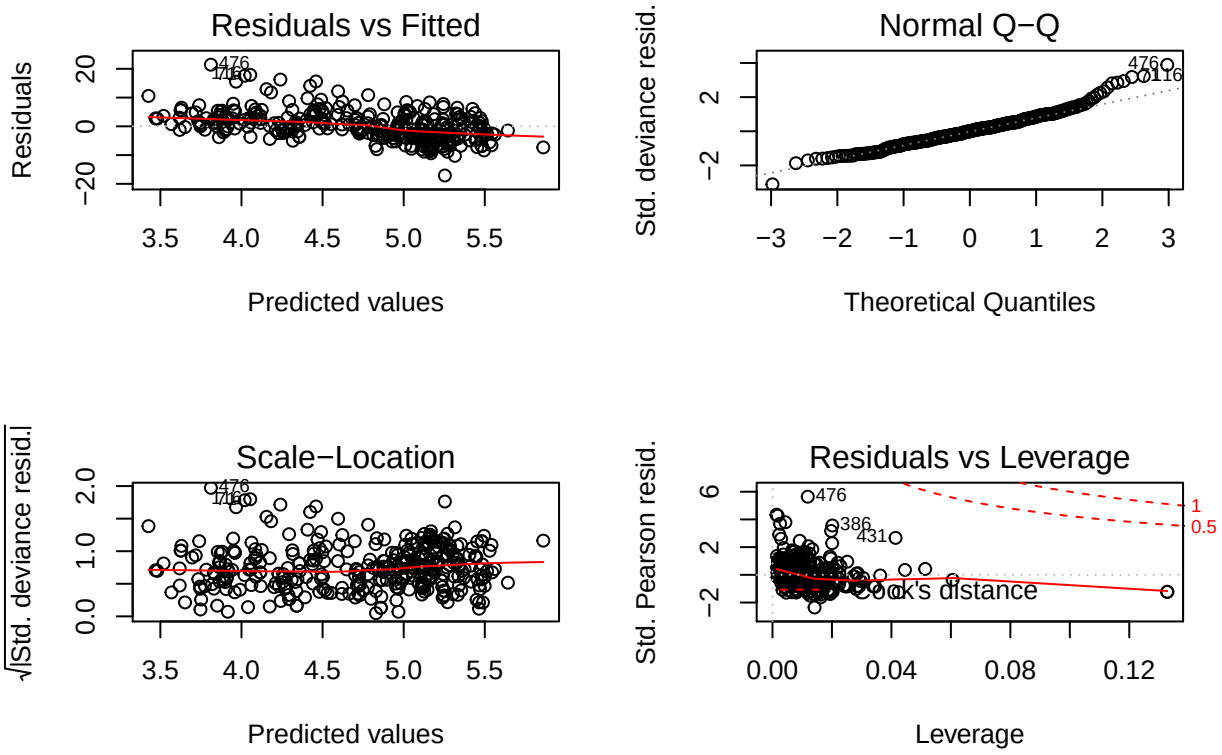
The output demonstrates for each additional male who is unemployed in NSW per 1000 people there is a 5.99 log count increase in domestic violence incidents, a 0.64 log count increase in domestic violence incidents for people with a weekly income between $800 and $1249 and a 2.31 log count increase in domestic violence incidents for sole parents. While not an exact equivalent to the R2 of linear regression exists, the McFadden R2 index can be used to assess the model fit and at 0.16 it doesn't appear to be a good fit. There also appears to be possible over-dipsersion based on the deviance and degrees of freedom while the addional plots below indicate there are problems with normality and potentially heterogeneity in the Poisson model.



To better understand the potential over-dispersion a Quasi-Poisson model was analysed below:

```
##
## Call:
## glm(formula = Domestic_Violence_Count ~ +M_Unemployment_ppt +
##     Income_x800_1249_ppt + Sole_parent_ppt + offset(log(Population_Count/1000)),
##     family = quasipoisson(link = log), data = dataTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -17.090   -3.118    0.007    2.865   21.428
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              0.5289     0.2010   2.631  0.00889 **
## M_Unemployment_ppt       5.9978     1.2430   4.825 2.11e-06 ***
## Income_x800_1249_ppt     0.6454     0.9273   0.696  0.48693
## Sole_parent_ppt          2.3119     0.5185   4.459 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 30.67796)
##
##     Null deviance: 10830.7  on 345  degrees of freedom
## Residual deviance:  8620.7  on 342  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

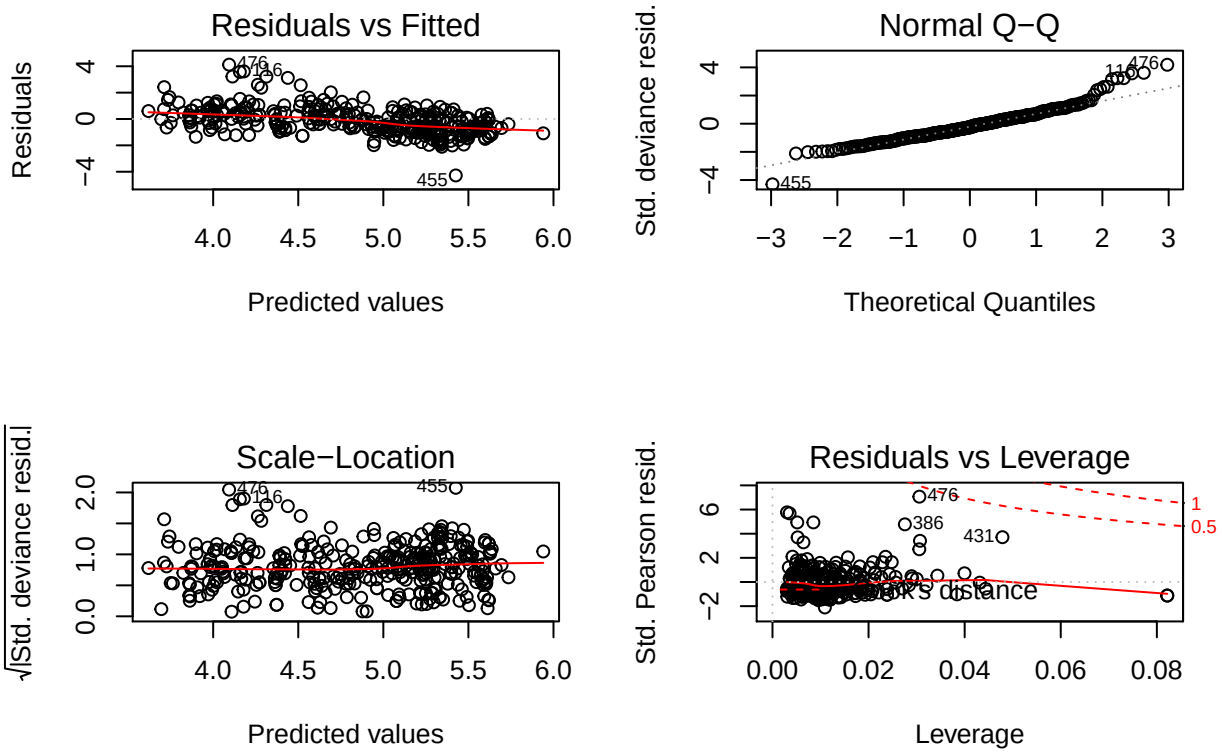


The dispersion parameter (deviance divided by the residual degrees of freedom) was 30.7, which is much greater than 1, indicating over-dispersion. In an attempt to deal with this phenomena a negative binomial distribution model was selected, but unlike the Poisson distribution, the variance and the mean are not equivalent indicating it may better represent modeling counts with variability that is different from the mean.

```
##
## Call:
## glm.nb(formula = Domestic_Violence_Count ~ +M_Unemployment_ppt +
```

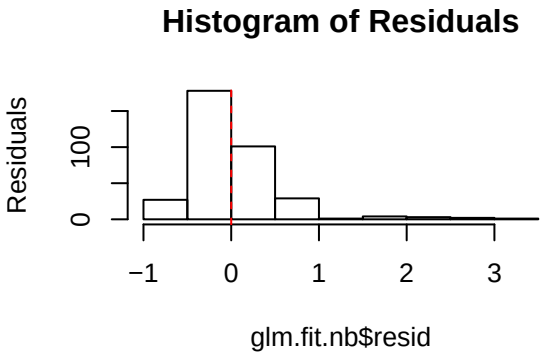
```
##     Income_x800_1249_ppt + Sole_parent_ppt + offset(log(Population_Count/1000)),
##     data = dataTrain, init.theta = 5.007049548, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2815  -0.8286  -0.2539   0.3985   4.1204
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)             0.7312     0.1831   3.994 6.49e-05 ***
## M_Unemployment_ppt      6.2625     1.2219   5.125 2.97e-07 ***
## Income_x800_1249_ppt    1.2128     0.8536   1.421  0.15535
## Sole_parent_ppt         1.2230     0.4436   2.757  0.00584 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(5.007) family taken to be 1)
##
##     Null deviance: 402.22  on 345  degrees of freedom
## Residual deviance: 357.72  on 342  degrees of freedom
## AIC: 3757.9
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  5.007
##           Std. Err.:  0.386
##
##  2 x log-likelihood:  -3747.945
```

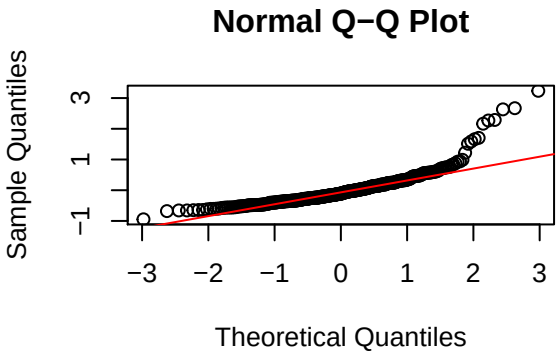
The output indicates Male Unemployment has a co-efficient of 6.2625, Weekly Income of $800-$1249 of 1.2128 and Sole Parents of 1.2230, which are all statistically significant. This means that for each one count increase in Male Unemployment, Weekly Income and Sole Parents per 1,000 people in NSW LGA's, the expected log count of the number of domestic violence incidents increases by 6.2625, 1.2128 and 1.2230, respectively. The Negative Binomial Distribution model has also addressed the over-dispersion issue with much closer deviance and degrees of freedom indicators.

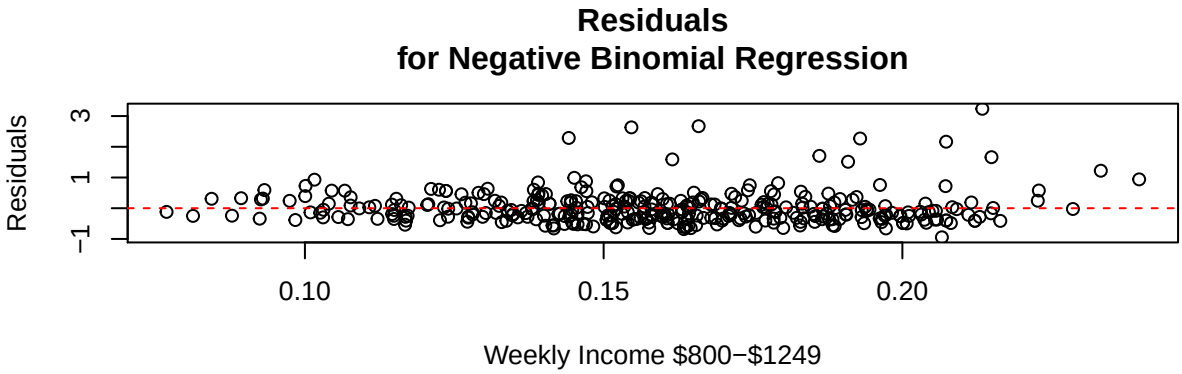


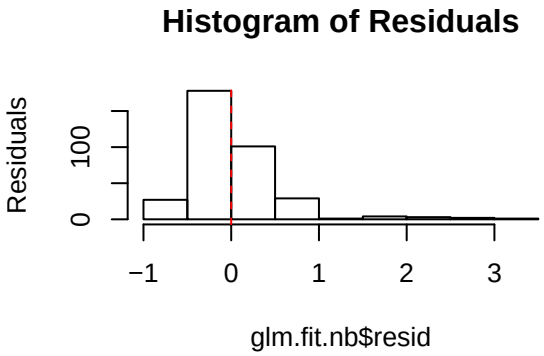
The below charts also demonstrate the Negative Binomial Distribution residuals for each of the predictor variables:

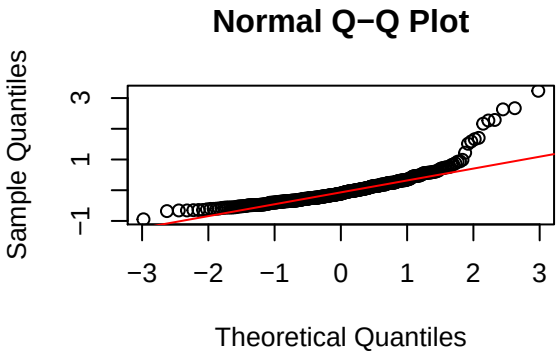

Residuals
for Negative Binomial Regression
Residuals
3
1
−1
0.05
0.10
0.15
Males Unemployment

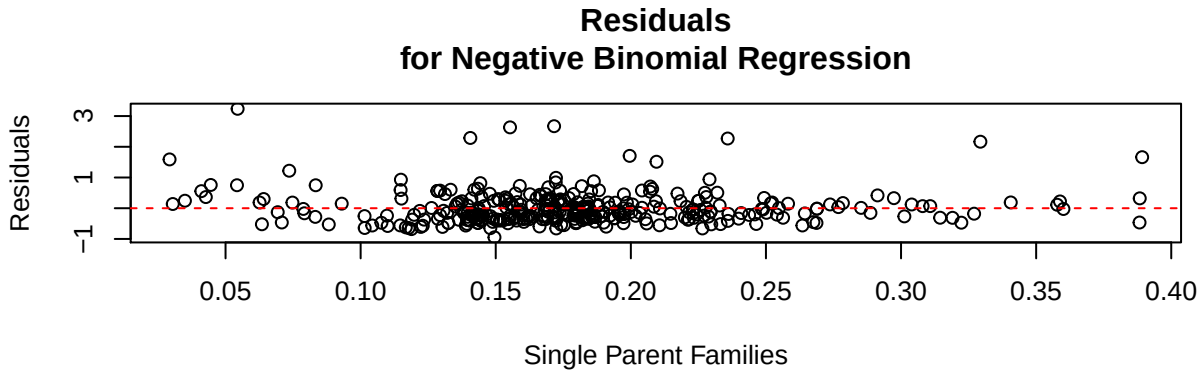

Histogram of Residuals
Residuals
100
0
−1
0
1
2
3
glm.fit.nb$resid

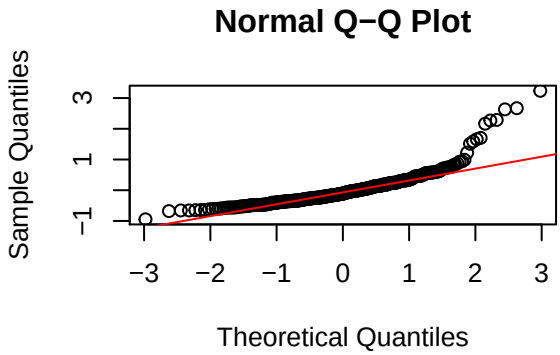

Normal Q−Q Plot
Sample Quantiles
3
1
−1
−3
−2
−1
0
1
2
3
Theoretical Quantiles


Residuals
for Negative Binomial Regression
Residuals
3
1
−1
0.10
0.15
0.20
Weekly Income $800−$1249


Histogram of Residuals
Residuals
100
0
−1
0
1
2
3
glm.fit.nb$resid


Normal Q−Q Plot
Sample Quantiles
3
1
−1
−3
−2
−1
0
1
2
3
Theoretical Quantiles





Assessing the predictive ability of the model

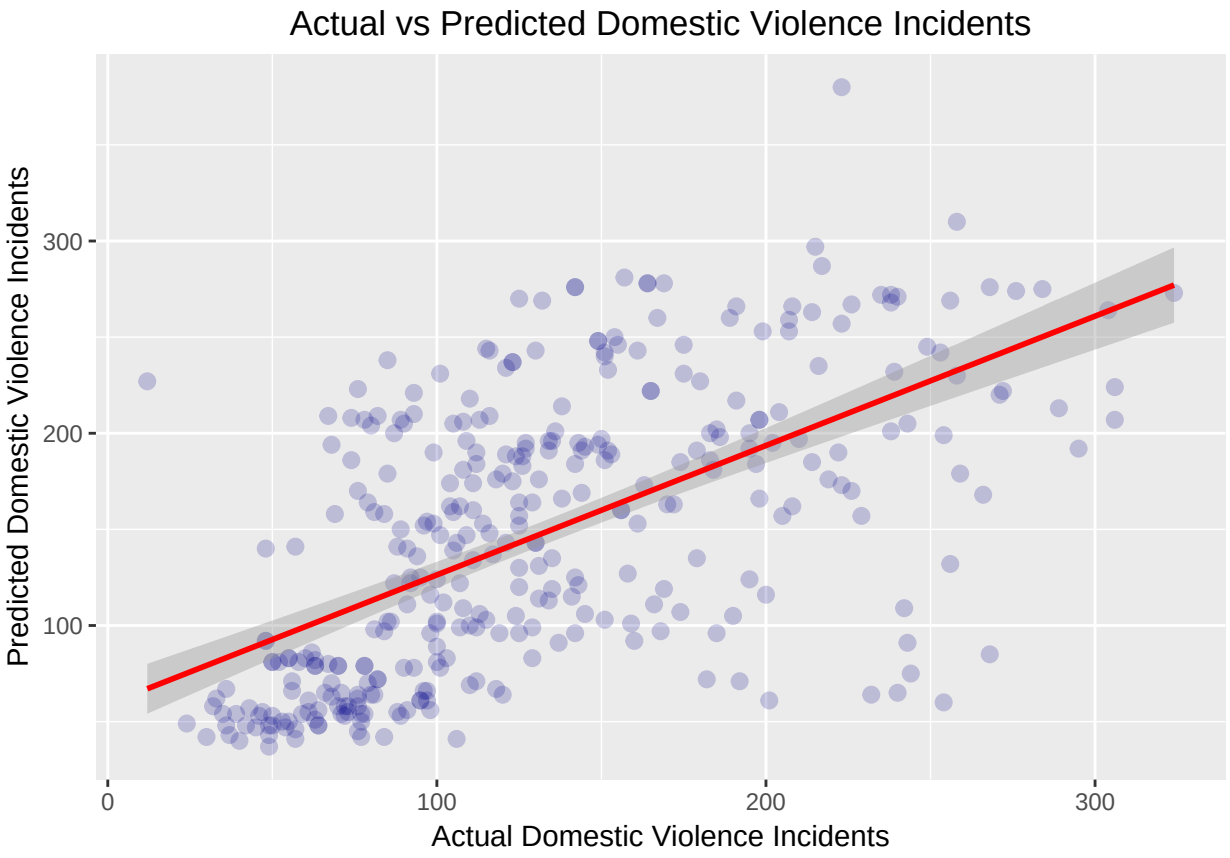
A 10k fold cross-validation was performed on the training set to avoid over-fitting the model on the final testing data set. A sample of the predicted domestic violence count rates versus the actual count rates are shown below. Over the testing data set the prediction accuracy of the model has a median difference of 12.9%.

```
##                                LGA Domestic_Violence_Count Prediction
## 474                    Lithgow (C)                      85        102
## 488                Wollondilly (A)                     185        202
## 317                    Ballina (A)                     136        201
## 495 Queanbeyan-Palerang Regional (A)                    161        243
## 16                    Griffith (C)                      142         96
## 451                      Albury (C)                     284        275
## 104           Glen Innes Severn (A)                      40         40
## 52                     Burwood (A)                       69        158
## 425                    Kempsey (A)                     243        205
## 40           Armidale Regional (A)                     130        143
##     Difference
## 474         17
## 488         17
## 317         65
## 495         82
## 16         -46
## 451         -9
## 104          0
## 52          89
```

```
## 425        -38
## 40          13
```

```
## [1] 12.9
```

Finally, a plot of the actual and predicted domestic violence counts on the testing data set is shown on the scatter plot below indicating a positive relationship.



Conclusion

References

Census Table Builder, https://auth.censusdata.abs.gov.au/webapi/jsf/dataCatalogueExplorer.xhtml

Christensen,R. 1997, Log-Linear Models and Logistic Regression, Chapter 9, Springer-Verlag New York. Dunkley, A and Phillips, J. 2015, Domestic Violence in Australia: A quick guide to the issues, Parliament of Australia, Department of Parliamentary Services.

Kleiber, Christian & Zeileis, Achim (2016), Visualizing Count Data Regressions Using Rootograms, The American Statistician.

Lillis, D. (2015), Generalized Linear Models in R, Part 7: Checking for Overdispersion in Count Regression, available: https://www.theanalysisfactor.com/glm-r-overdispersion-count-regression.

People, Julie. 2015, Trends and Patterns in Domestic Violence Assaults, NSW Bureau of Crime Statistics and Research, Number 89.

Stavrou, E, Poynton, S & Weatherburn, S. November 2016, Intimate partner violence against women in Australia: related factors and help-seeking behaviours, NSW Bureau of Crime Statistics and Research.

Weatherburn, D (2011), Personal Stress, financial stress and violence against women, Contemporary Issues in Crime and Justice, NSW Bureau of Crime Statistics and Research, Number 151.

Appendix

Canterbury-Bankstown (A) The Canterbury-Bankstown Council is a local government area located in the south-western suburbs of Sydney, in New South Wales, Australia. The council was formed on 12 May 2016 from a merger of the City of Canterbury and the City of Bankstown,[2] after a review of local government in New South Wales by the state government. Source: Canterbury-Bankstown Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 14 May 2016.

Snowy Monaro Regional (A) The Snowy Monaro Regional Council is a local government area located in the Snowy Mountains and Monaro regions of New South Wales, Australia. The council was formed on 12 May 2016 through a merger of the Bombala, Cooma-Monaro and Snowy River shires.[1] Source: Snowy Monaro Regional Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016.

Hilltops (A) Hilltops Council is a local government area in the South West Slopes region of New South Wales, Australia. This area was formed on the 12 May 2016 from the merger of Boorowa Council, Harden Shire and Young Shire.[2] Source: Hilltops Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 14 May 2016.

Central Coast (C) (NSW) The Central Coast Council is a local government area serving the Central Coast region of New South Wales, Australia, established on 12 May 2016 following the amalgamation of Gosford City and Wyong Shire councils. Source: Central Coast Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 14 May 2016.

Edward River (A) The Edward River Council is a local government area in the Riverina region of New South Wales, Australia. This area was formed in 2016 from the merger of the Deniliquin Council with the surrounding Conargo Shire.[2] Source: Edward River Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016.

Federation (A) A 2015 review of local government boundaries by the NSW Government Independent Pricing and Regulatory Tribunal recommended that the Corowa Shire merge with the Lockhart and Urana shires to form a new council with an area of 8,581 square kilometres (3,313 sq mi) and support a population of approximately 16,000.[3] The Council was dissolved on 12 May 2016 and along with Urana Shire the area became part of the new Federation Council [4] Source: Federation Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016.

Western Plains Regional (A) The Dubbo Regional Council is a local government area located in the Central West and Orana regions of New South Wales, Australia. The council was formed on 12 May 2016 through a merger of the City of Dubbo and Wellington Council.[2] Source: Dubbo Regional Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 7 January 2016.

Cumberland (A) The Cumberland Council is a local government area located in the western suburbs of Sydney in the state of New South Wales, Australia. The Council was formed on 12 May 2016 from the merger of parts of the Auburn City, the former Parramatta City (Woodville Ward), and Holroyd City councils.[2][3] Source: Cumberland Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016. Local Government (City of Parramatta and Cumberland) Proclamation 2016 [NSW] - Schedule 2 - Provisions for Cumberland Council. NSW Government. 12 May 2016. Retrieved 29 September 2017.

Georges River (A) The Georges River Council is a local government area located in the St George region of Sydney located south of the CBD, in New South Wales, Australia. The Council was formed on 12 May 2016 from the merger of the Kogarah City Council and Hurstville City Council.[3]

Mid-Coast (A) Mid-Coast Council is a local government area located in the Mid North Coast region of New South Wales, Australia. The council was formed on 12 May 2016 through a merger of the Gloucester

Shire, Great Lakes and City of Greater Taree councils.[1] Source: Mid-Coast Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 24 May 2016.

Gloucester (A) A 2015 review of local government boundaries by the NSW Government Independent Pricing and Regulatory Tribunal recommended that the Gloucester Shire merge with adjoining councils. The government considered two proposals. The first proposed a merger of Gloucester Shire and Dungog Shire councils to form a new council with an area of 5,200 square kilometres (2,000 sq mi) and support a population of approximately 14,000.[6] Following the lodging of an alternate proposal by Gloucester Shire Council to amalgamate the Gloucester, Great Lakes and Greater Taree councils, the NSW Minister for Local Government proposed a merger between the Dungog Shire and City of Maitland.[7] Source: Mid-Coast Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 20 May 2016.

Armidale Regional (A) Guyra Shire is the name of a former local government area located in the New England region of New South Wales, Australia. The shire was abolished on 12 May 2016, where the council, together with the Armidale Dumaresq Shire, was subsumed into the Armidale Regional Council with immediate effect.[3] Source: Armidale Regional Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 12 May 2016.

Inner West (A) Inner West Council is a local government area located in the inner western region of Sydney in the state of New South Wales, Australia. The Council was formed on 12 May 2016 from the forced merger of the Ashfield, Leichhardt, and Marrickville councils.[2][3] Source: Inner West Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 15 May 2016.

Murrumbidgee (A) Jerilderie Shire was a local government area in the Riverina region of southern New South Wales, Australia. The Shire was located adjacent to the Newell Highway. The Shire was declared in 1918 after the amalgamation of the former Municipality of Jerilderie (1889 - 1918) and Wunnamurra Shire (1906 - 1918). It was dissolved in 2016 after its amalgamation with Murrumbidgee Shire to create Murrumbidgee Council. Source: Murrumbidgee Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016.

Northern Beaches (A) The Northern Beaches Council is a local government area located in the northern beaches region of Sydney, in the state of New South Wales, Australia. The Council was formed on 12 May 2016 replacing Manly, Pittwater and Warringah Councils.[2] Source: Northern Beaches Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 12 May 2016.

Murray River (A) The Murray River Council is a local government area in the Riverina region of New South Wales, Australia. This area was formed in 2016 from the merger of Murray Shire with Wakool Shire.[2] Source: Murray River Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016.

Snowy Valleys (A) The Snowy Valleys Council is a local government area located in the South West Slopes region of New South Wales, Australia. This area was formed in 2016 from the merger of Tumbarumba Shire and Tumut Shire councils.[2] Source: Snowy Valley Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 13 May 2016.

Queanbeyan-Palerang Regional (A) Queanbeyan-Palerang Regional Council is a local government area located in the Southern Tablelands region of New South Wales, Australia. The council was formed on 12 May 2016 through a merger of the City of Queanbeyan and Palerang Council.[2] Source: Queanbeyan-Palerang Regional Council. Stronger Councils. Government of New South Wales. 12 May 2016. Retrieved 21 May 2016.

Bayside Bayside Council is a local government area located around Botany Bay which is split between the eastern suburbs and St George areas of Sydney, located between 7 kilometres (4.3 mi) and 12 kilometres (7.5 mi) south of the CBD[3] in the state of New South Wales, Australia. The Council was formed on 9 September 2016 from the merger of the Botany Bay and the Rockdale councils.[4] Source: Bayside Council. Stronger Councils. Government of New South Wales. 9 September 2016. Retrieved 9 September 2016.