Tableau Visualisation Software

Tableau Desktop is a data visualisation tool that allows the analysis and visualisation of data for the purposes of easily seeing and understanding information and patterns (Tableau, 2016).

This author determined to use Tableau as the visualisation tool for this paper because it provides a variety of visualisation methods for analysis, it can be used to combine data from various data sources, and provides tools to enable the manipulation of data to analyse and investigate interesting properties.

Visualisation and Analysis

Multidimensional Data

Multidimensional (or hypervariate) data analysis involves the study of data that is grouped into two categories, data dimensions and measurements (Wikipedia, Multivariate Analysis, 2016). The data provided for this paper is multidimensional in that it contains information for categories of crimes, in different local government areas, over a period of several years.

The dimensionality of the data can cause problems in analysis, so it is important to plan the implementation of the visualisation. This author intends to use the data file **Crime_Yearly_FINAL.csv** and supplemental data obtained from online sources.

There are 4 steps that need to be taken when producing graphical representations from multidimensional data:

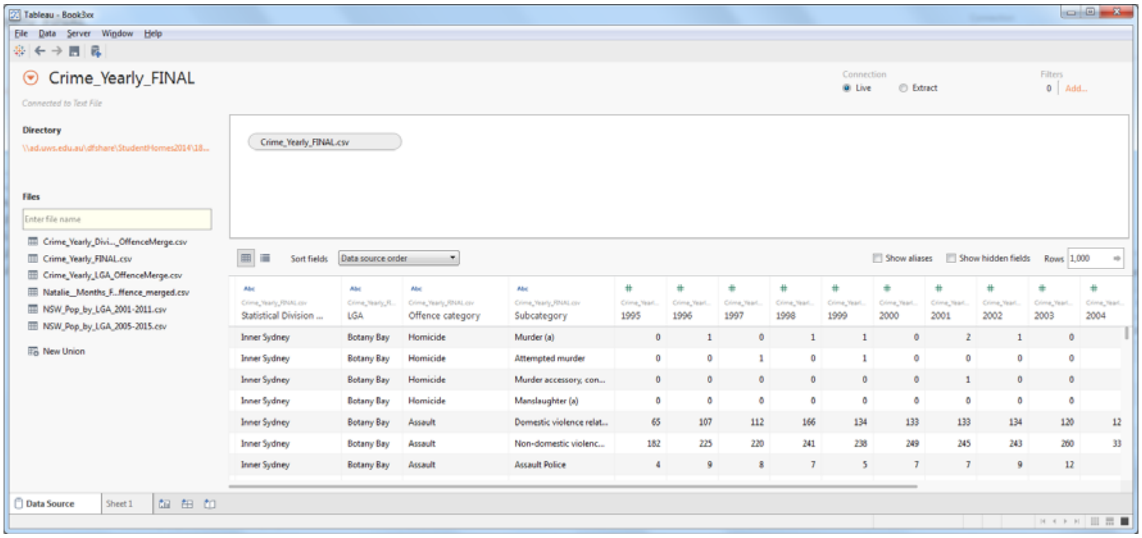
1. Data gathering, formatting and storing.
2. Data mining and enrichment
3. Data mapping and layout
4. Rendering

1. Data Gathering

Loading the Data

A number of steps were undertaken when importing the data into Tableau.

The data was imported as a csv file, with the first row configured as containing the column names.



Crime_Yearly_FINAL

Connected to Text File

Directory

\\ad.uws.edu.au\dfshare\StudentHomes2014\18...

Files

Enter file name

- Crime_Yearly_Divi..._OffenceMerge.csv
- Crime_Yearly_FINAL.csv
- Crime_Yearly_LGA_OffenceMerge.csv
- Natalie__Months_F...ffence_merged.csv
- NSW_Pop_by_LGA_2001-2011.csv
- NSW_Pop_by_LGA_2005-2015.csv
- New Union

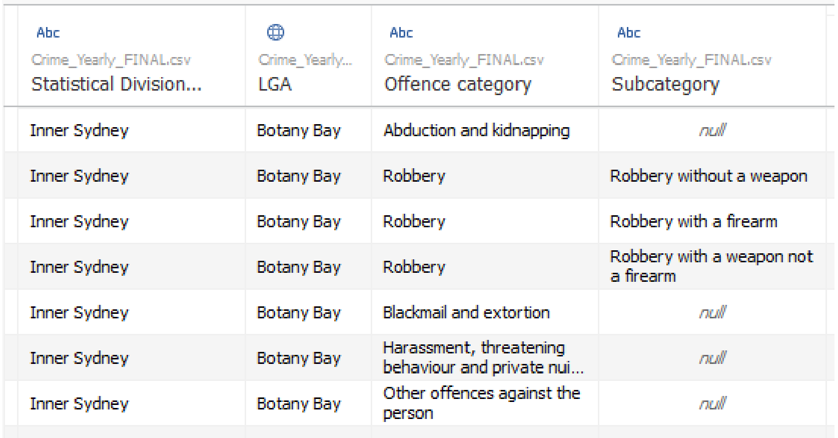
Crime_Yearly_FINAL.csv

Sort fields: Data source order

□ Show aliases □ Show hidden fields Rows 1,000

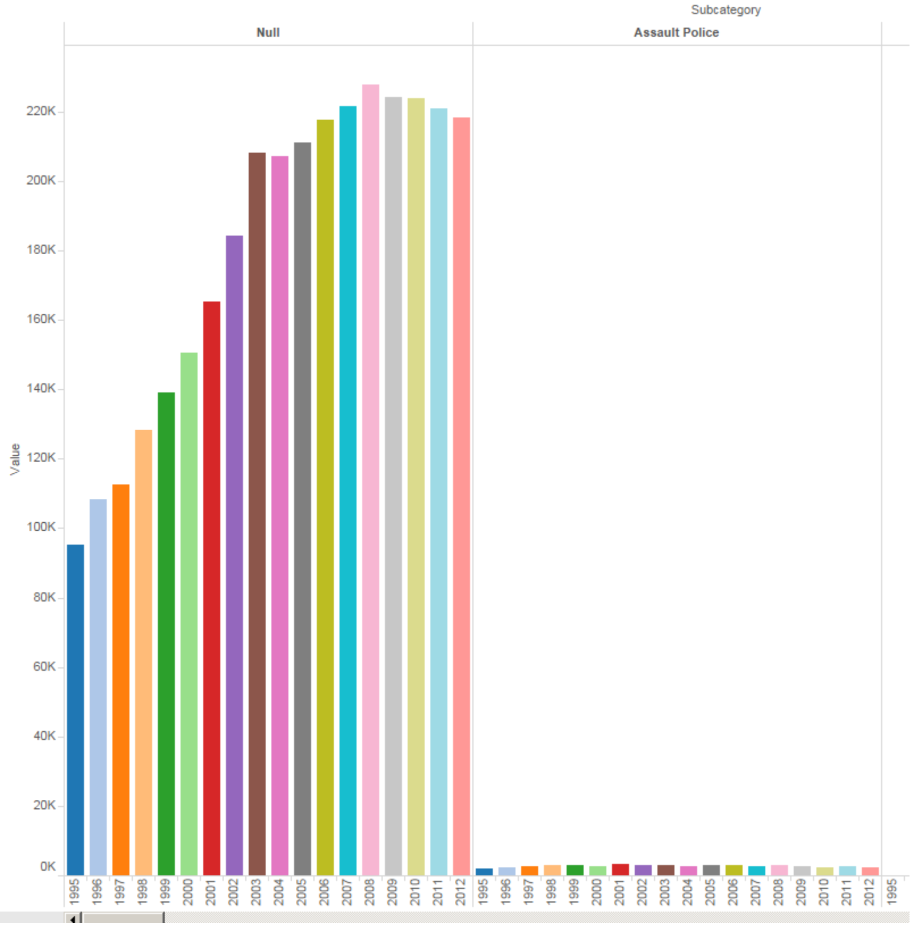
Statistical Division ...	LGA	Offence category	Subcategory	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Inner Sydney	Botany Bay	Homicide	Murder (a)	0	1	0	1	1	0	2	1	0	
Inner Sydney	Botany Bay	Homicide	Attempted murder	0	0	1	0	1	0	0	0	0	
Inner Sydney	Botany Bay	Homicide	Murder accessory, con...	0	0	0	0	0	0	1	0	0	
Inner Sydney	Botany Bay	Homicide	Manslaughter (a)	0	0	0	0	0	0	0	0	0	
Inner Sydney	Botany Bay	Assault	Domestic violence relat...	65	107	112	166	134	133	133	134	120	12
Inner Sydney	Botany Bay	Assault	Non-domestic violenc...	182	225	220	241	238	249	245	243	260	33
Inner Sydney	Botany Bay	Assault	Assault Police	4	9	8	7	5	7	7	9	12	

On initial inspection of the data, a side by side bar chart was used with offense subcategories as a column field, and the sum of yearly data in the row field. However, it was found that there were certain gaps in the data. Some crime categories did not contain corresponding subcategories, with the subcategory field containing a ***NULL*** value.

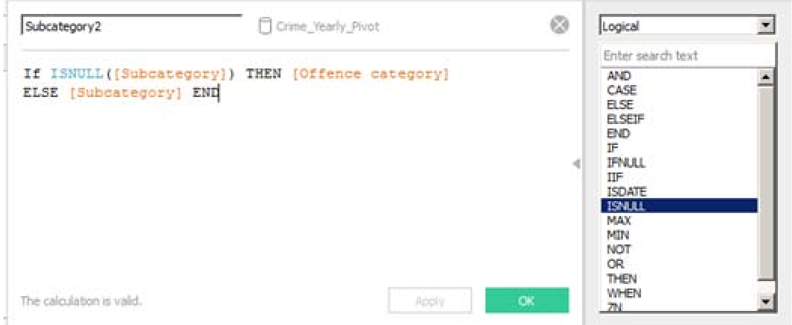


Crime_Yearly_FINAL.csv Statistical Division...	Crime_Yearly... LGA	Crime_Yearly_FINAL.csv Offence category	Crime_Yearly_FINAL.csv Subcategory
Inner Sydney	Botany Bay	Abduction and kidnapping	*null*
Inner Sydney	Botany Bay	Robbery	Robbery without a weapon
Inner Sydney	Botany Bay	Robbery	Robbery with a firearm
Inner Sydney	Botany Bay	Robbery	Robbery with a weapon not a firearm
Inner Sydney	Botany Bay	Blackmail and extortion	*null*
Inner Sydney	Botany Bay	Harassment, threatening behaviour and private nui...	*null*
Inner Sydney	Botany Bay	Other offences against the person	*null*

This meant that when analysing data by subcategory, anomalies were seen as a *Null* column, amalgamating offence types that were not normally grouped.



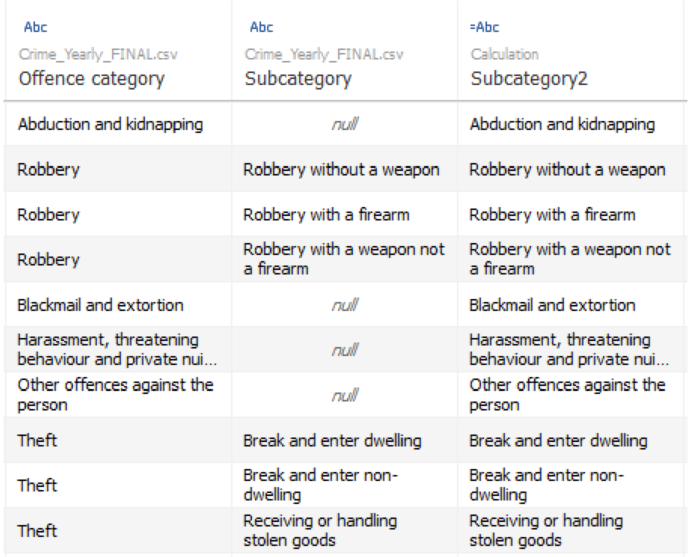
A calculated field was created, called ***Subcategory2***, which filled the null values with the offence category values.



```
If ISNULL([Subcategory]) THEN [Offence category]
ELSE [Subcategory] END
```

The calculation is valid.

The resulting column can then be used to classify data more accurately.



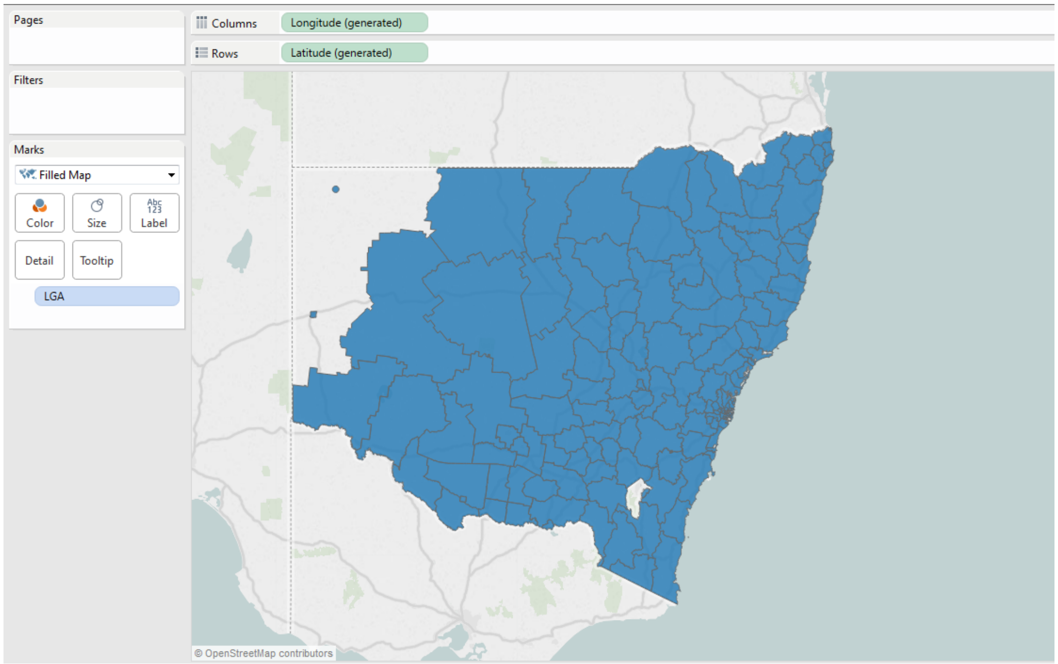
Abc	Abc	=Abc
Crime_Yearly_FINAL.csv	Crime_Yearly_FINAL.csv	Calculation
Offence category	Subcategory	Subcategory2
Abduction and kidnapping	*null*	Abduction and kidnapping
Robbery	Robbery without a weapon	Robbery without a weapon
Robbery	Robbery with a firearm	Robbery with a firearm
Robbery	Robbery with a weapon not a firearm	Robbery with a weapon not a firearm
Blackmail and extortion	*null*	Blackmail and extortion
Harassment, threatening behaviour and private nui...	*null*	Harassment, threatening behaviour and private nui...
Other offences against the person	*null*	Other offences against the person
Theft	Break and enter dwelling	Break and enter dwelling
Theft	Break and enter non-dwelling	Break and enter non-dwelling
Theft	Receiving or handling stolen goods	Receiving or handling stolen goods

Supplemental Data

Location

It is important to contextualise the crime data in terms of points of reference or commonalities. To compare crime across NSW, a geographic property needed to be assigned to either the **Statistical Division or Subdivision** field, or the **LGA** (Local Government Area).

BOSCAR groups data by NSW LGA, and it was decided to give this field a "*Geographic Role*" in Tableau. The ***County*** geographic role allows Tableau to generate a latitude and longitude based on LGA.



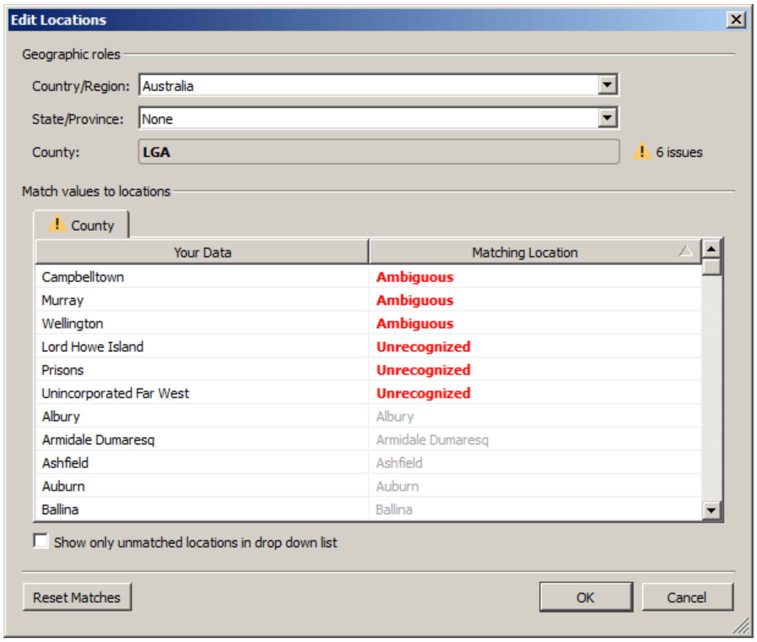
Columns Longitude (generated)

Rows Latitude (generated)

Filled Map

Detail LGA

Tableau was not perfect in matching locations though, due to either not knowing the location specified, or the location being referred to by a different name.



Geographic roles

Country/Region: Australia

State/Province: None

County: LGA ! 6 issues

Match values to locations

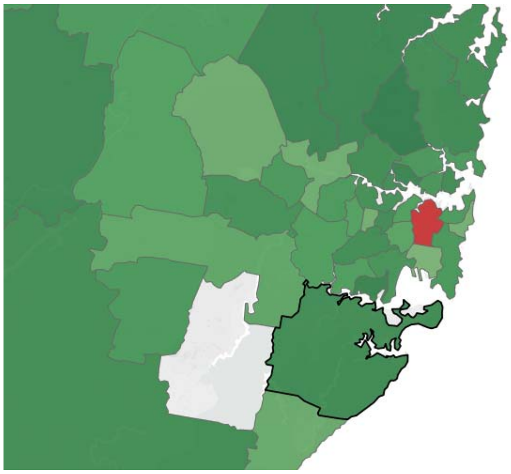
! County

Your Data	Matching Location
Campbelltown	Ambiguous
Murray	Ambiguous
Wellington	Ambiguous
Lord Howe Island	Unrecognized
Prisons	Unrecognized
Unincorporated Far West	Unrecognized
Albury	Albury
Armidale Dumaresq	Armidale Dumaresq
Ashfield	Ashfield
Auburn	Auburn
Ballina	Ballina

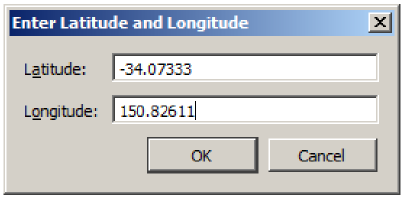
□ Show only unmatched locations in drop down list

The **Edit Locations** window allows for manual matching of data. Latitude and longitude can be entered manually from data sourced from Google Maps and Wikipedia.

For example, a missing area apparent on the map is shown below:



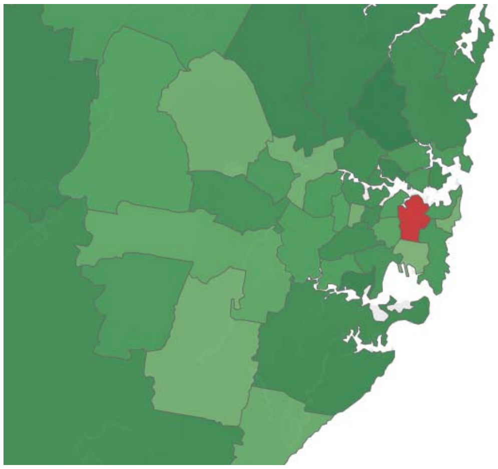
This area is in southern Sydney, and could be the ambiguous location of Campbelltown. The latitude and longitude are entered manually from the **Edit** Location window.



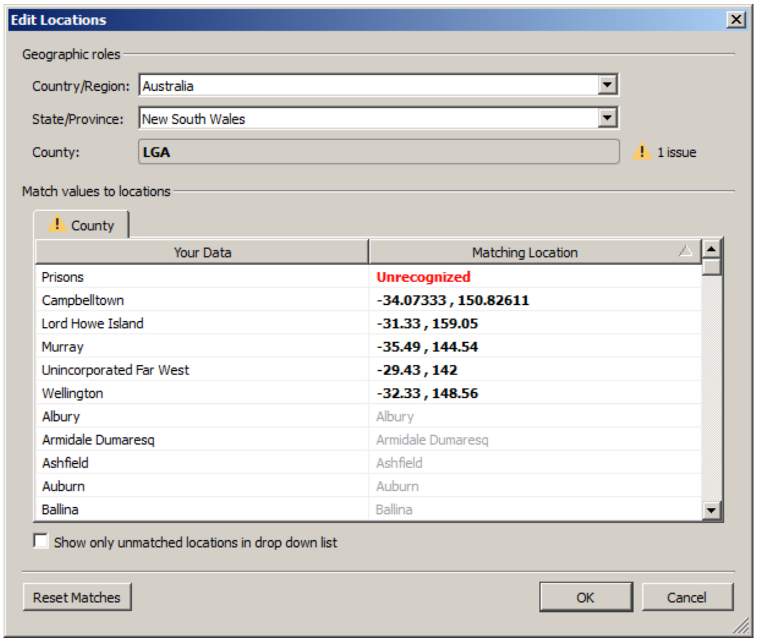
Latitude: -34.07333

Longitude: 150.82611

It can be seen after this that the LGA is now filled in by Tableau with relevant data.



The Lord Howe Island LGA, although correctly added to the geographic location settings, will be omitted from most map representations, it causes problems with map scaling as it is an outlier past the east coast of NSW, and will not be in the scope of this paper.



Edit Locations

Geographic roles

Country/Region: Australia

State/Province: New South Wales

County: LGA ⚠ 1 issue

Match values to locations

⚠ County

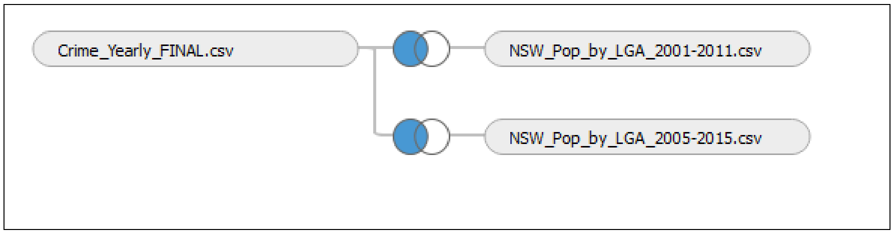
Your Data	Matching Location
Prisons	Unrecognized
Campbelltown	-34.07333 , 150.82611
Lord Howe Island	-31.33 , 159.05
Murray	-35.49 , 144.54
Unincorporated Far West	-29.43 , 142
Wellington	-32.33 , 148.56
Albury	Albury
Armidale Dumaresq	Armidale Dumaresq
Ashfield	Ashfield
Auburn	Auburn
Ballina	Ballina

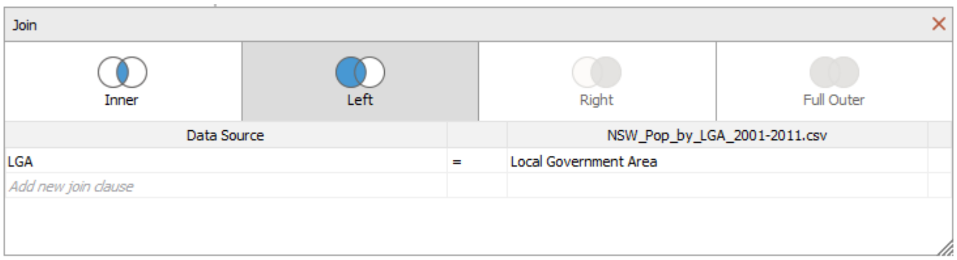
☐ Show only unmatched locations in drop down list

Population Density

To make more sense of the data, it was decided to compare offense counts per LGA to the LGA population. This is a common analysis for spatial data, and can show more effectively crime hotspots and trends.

Population estimates per LGA were obtained from the Australian Bureau of Statistics web site (ABS, 2016), and merged into the Tableau data source using a left join based on LGA.





Data Source		NSW_Pop_by_LGA_2001-2011.csv
LGA	=	Local Government Area
Add new join clause		

Ambiguous LGA names were corrected, such as *Unincorporated Far West* being referred to as *Unincorporated NSW* in the ABS data, and the *Prisons* LGA was ignored, as the author will not be analysing prison crime data in this paper. Note that this author was able to obtain population estimates for the years 2001 to 2015 only. Earlier data was only available in PDF format, and would have taken lengthy steps to import into Tableau.

2. Data Mining and Enrichment

Several calculations and data analysis techniques were used to enrich the crime statistics data.

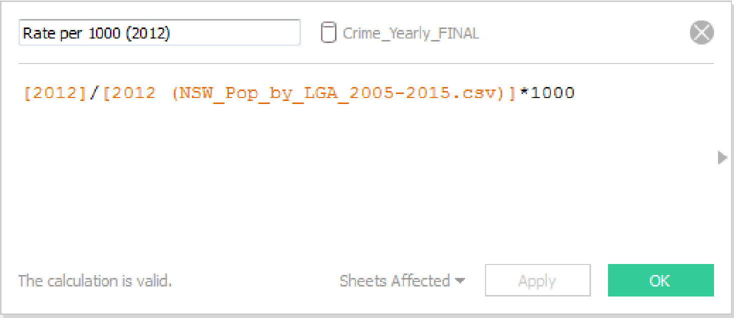
Rate of Crime

The crime rate is calculated by dividing the number of reported crimes by the total population. In statistical analysis, it is common to multiply the result by 100,000, mostly to portray orders of magnitude or significance, and in this instance BOSCAR does use “rate per 100,000 population” in its online tools. However, this author has decided to use “rate per 1000” to show more granularity for crime incidents by subcategory.

The rate is calculated as follows:

$$\frac{\text{Number of criminal incidents}}{\text{Resident population of LGA}} \times 1000$$

In Tableau, calculated fields are created in the **Data Source** pane.



Rate per 1000 (2012)

Crime_Yearly_FINAL

```
[2012]/[2012 (NSW_Pop_by_LGA_2005-2015.csv)]*1000
```

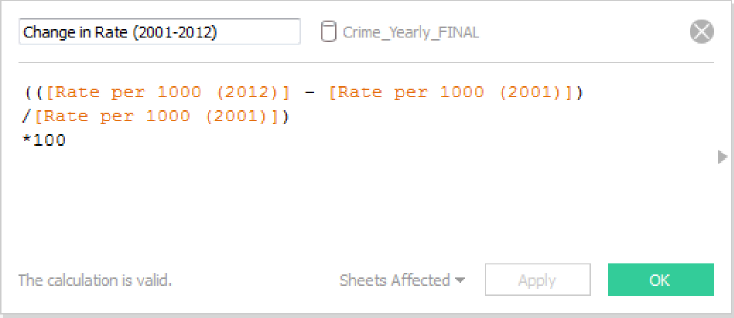
The calculation is valid.

Rates were calculated for the years 2001, 2004, 2008 and 2012.

The rate of change in crime incidents was also calculated from 2001 to 2012, as follows:

$$\frac{\text{Rate per 1000 (2012)} - \text{Rate per 1000 (2001)}}{\text{Rate per 1000 (2001)}} \quad \text{X } 100\%$$

Again, from the Tableau **Data Source** pane, as follows:



Change in Rate (2001-2012)

Crime_Yearly_FINAL

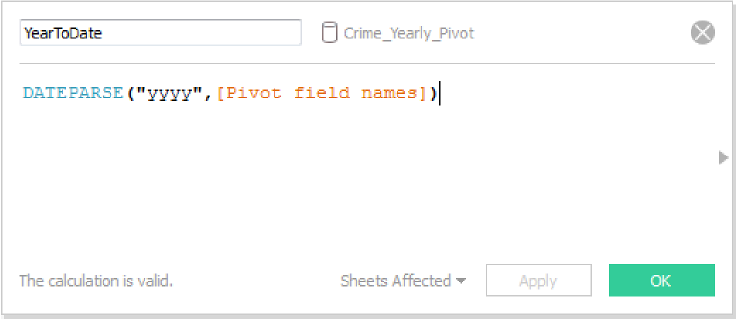
```
(([Rate per 1000 (2012)] - [Rate per 1000 (2001)])
/[Rate per 1000 (2001)])
*100
```

The calculation is valid.

Pivot Table

While investigating the data, it was also found that Tableau did not allow the visualisation of data by date, for example, using line charts or scatter plots. This is because the year data was imported as a column, and was recognised as ***a continuous measure*** value by Tableau. To plot data by date, the year needs to be a ***date value***, and also have a ***discrete*** data role, which means that there are a finite amount of values the data field can take, eg. years from 1995 to 2012.

In the **Data Source** pane a calculation was created to convert the text year value to a date, as follows:



```
DATEPARSE("yyyy",[Pivot field names])
```

The calculation is valid.

To show all offence values by year, the ***pivot*** option can be used to change the format of the existing fields so that all year values are contained in one field and all offence values are contained in another (Tableau-Help, 2016).

The columns containing data by year were then selected and the pivot option selected by right clicking on the columns and selecting **Pivot**.

The original fields in the data source are then replaced with new fields called "*Pivot field names*" in the **Dimensions** pane, and "*Pivot field values*" in the **Measures** pane.

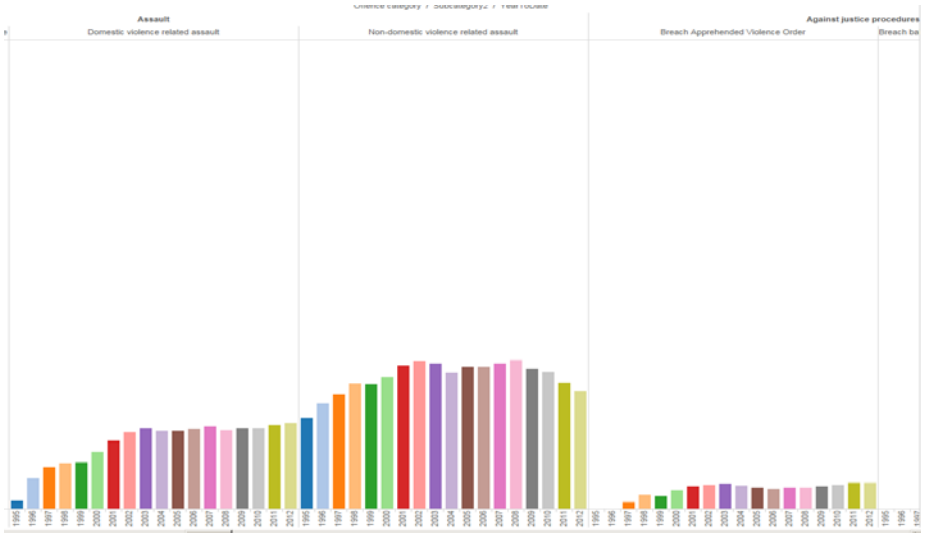
Unfortunately, this author found it difficult to analyse data created for crime rates and change in crimes rates with the pivoted data. This is because adding the columns for population estimates to the pivot just added the values to the pivot field values, causing confusion to the data. Further research would need to be undertaken to manipulate the data appropriately.

3. Data Mapping and Visualisation

Bar Charts

Initial investigations of the **Crime_Yearly_FINAL.csv** data produced limited results. As previously discussed, only bar charts and circle plots could be produced with the data loaded straight from the file. However, bar charts are useful for comparing things between different groups.

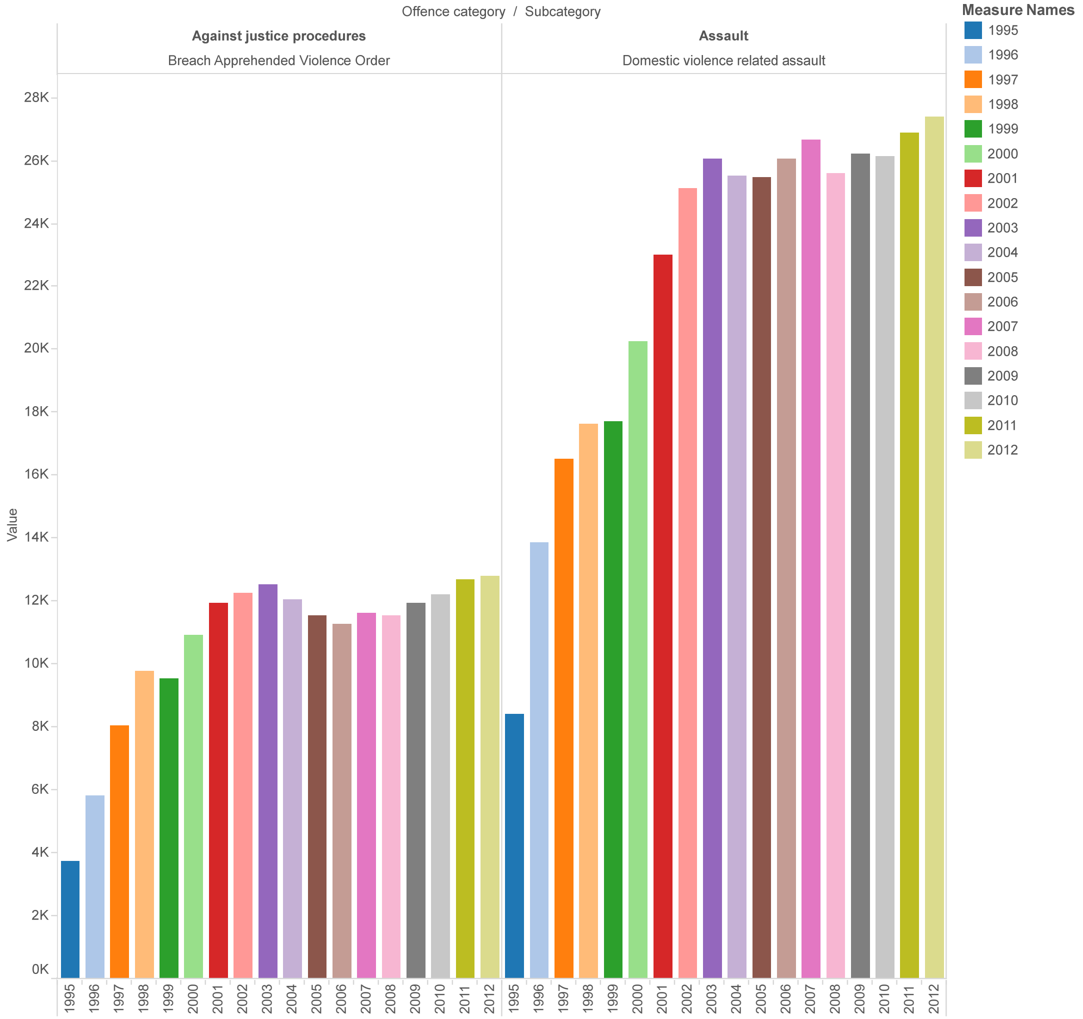
Below is a chart showing the comparison of offence category and subcategory by year. Interesting features could be seen however, by examining the bar plots and scrolling horizontally through the bar graph data.



One interesting feature observed was a visual correlation between subcategories Breach Apprehended Violence Order (Offence category Against Justice Procedures), and Domestic Violence Related Assault (Offence category Assault). This prompted the author to investigate these properties further.

Below is a bar chart of the subcategories Domestic Violence Related Assault (DVA) and Breach Apprehended Violence Order (BAVO).

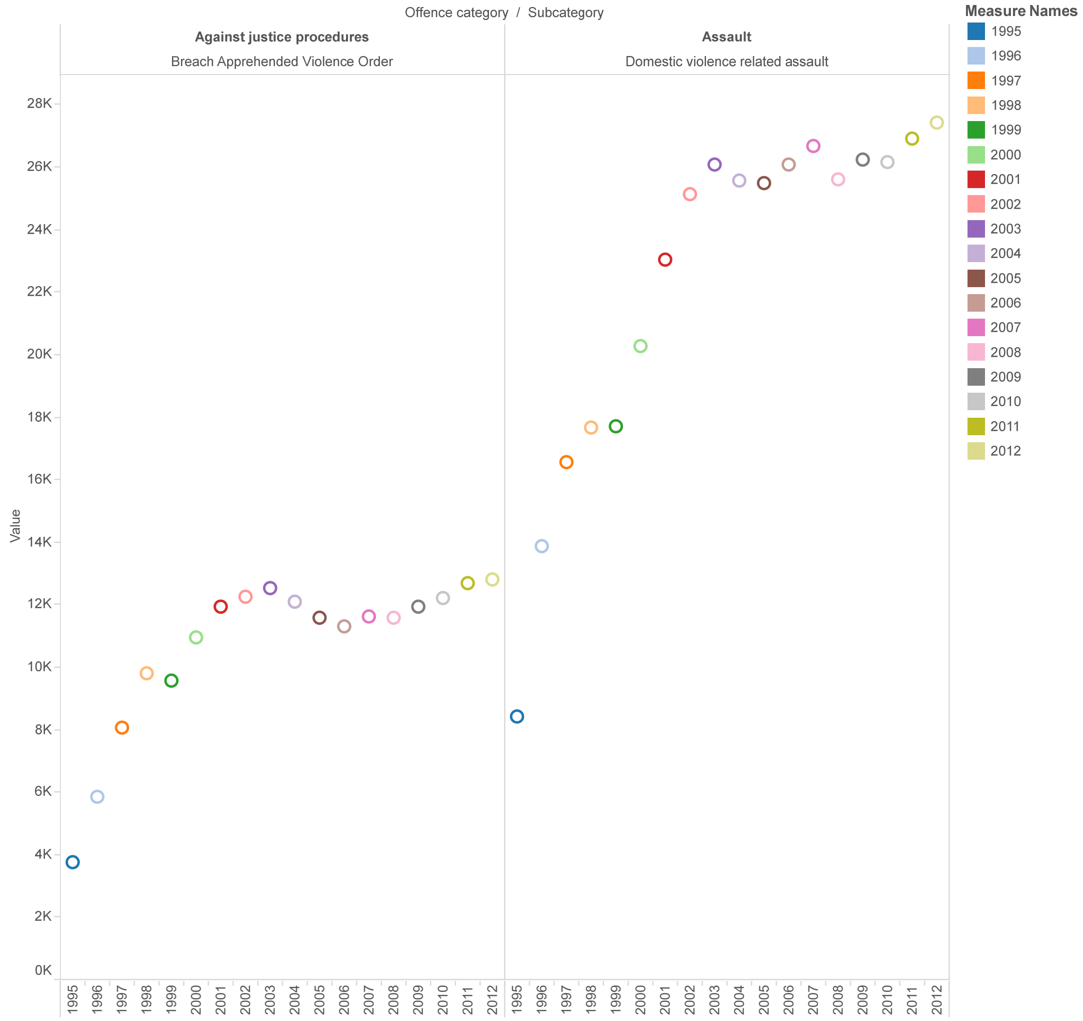
Sheet 2



1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011 and 2012 for each Subcategory broken down by Offence category. Color shows details about 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011 and 2012. The view is filtered on Subcategory, which keeps Breach Apprehended Violence Order and Domestic violence related assault.

A scatter plot of the data was attempted, see below, and this was when the author decided to perform the data mining and enrichment procedures above (section 2. Data Mining and Enrichment).

Sheet 2



1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011 and 2012 for each Subcategory broken down by Offence category. Color shows details about 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011 and 2012. The view is filtered on Subcategory, which keeps Breach Apprehended Violence Order and Domestic violence related assault.

Line Graphs and Crosstab Analysis

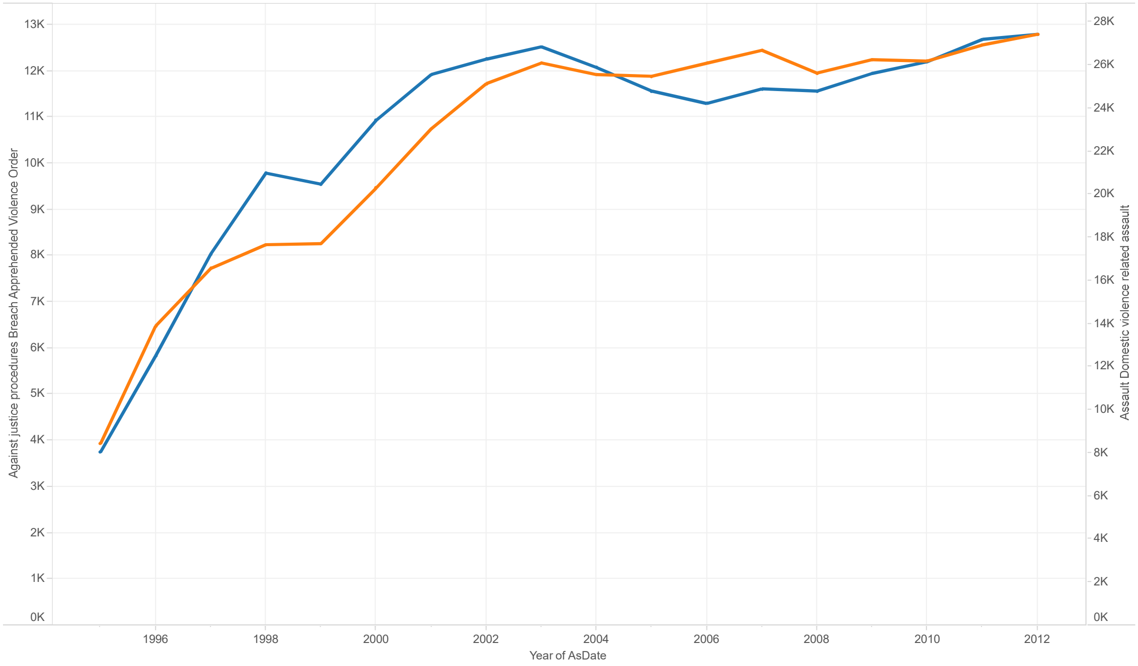
Line charts track changes over short and long periods of time, but do not work that well for categorical data, and are limited for large scale analysis. Therefore data was filtered to include only the DVA and BAVO information.

A **crosstab**, or text version of the selected DVA and BAVO data was exported to Excel and imported again as a data source.

	Offence category / Subcategory	
	Against justice procedures	**Assault**
	Breach Apprehended Violence Order	Domestic violence related assault
1995	3,736	8,403
1996	5,818	13,861
1997	8,028	16,533
1998	9,777	17,637
1999	9,538	17,687
2000	10,917	20,259
2001	11,912	23,018
2002	12,243	25,111
2003	12,511	26,067
2004	12,063	25,529
2005	11,553	25,453
2006	11,283	26,058
2007	11,601	26,656
2008	11,552	25,599
2009	11,934	26,223
2010	12,188	26,155
2011	12,671	26,904
2012	12,781	27,399

The Year field was transformed into a calculated date field so the results were able to be plotted as a line chart by year. The plot below shows the DVA and BAVO data compared on the same graph.

AVO + DVA



The trends of Against justice procedures Breach Apprehended Violence Order and Assault Domestic violence related assault for AsDate Year. Color shows details about Against justice procedures Breach Apprehended Violence Order and Assault Domestic violence related assault.

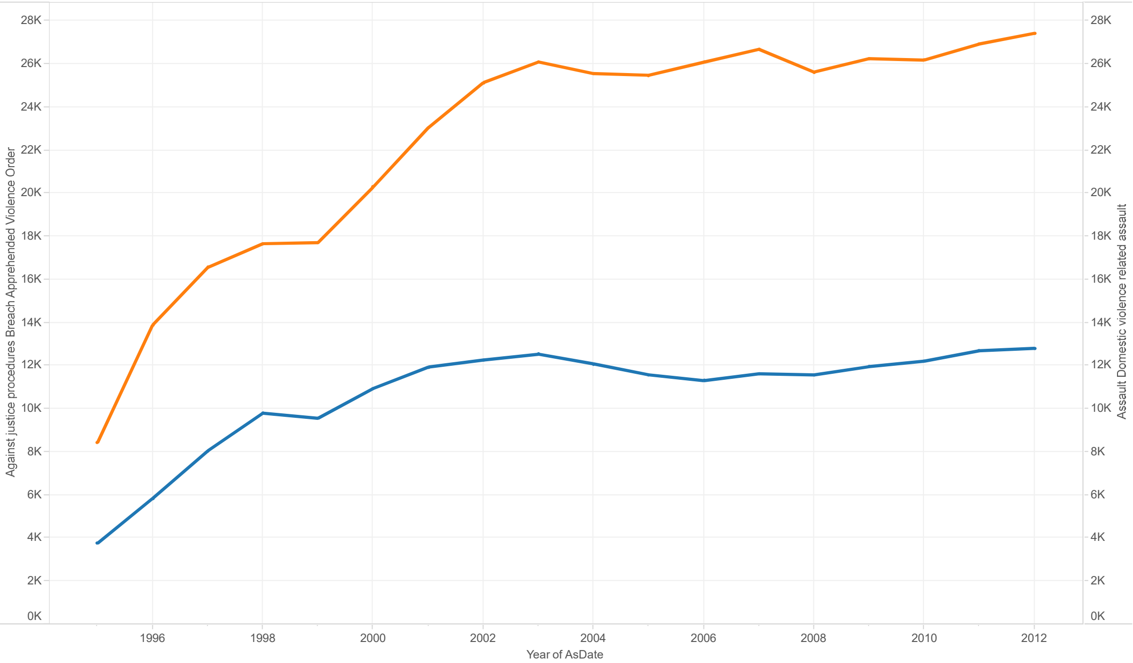
Measure Names

🟦 Against justice procedures Breach Apprehended Violence Order

🟧 Assault Domestic violence related assault

However, the initial results are misleading. Although a correlation can be seen in the data, the y-axis values differ, and a distortion of the actual link is shown. The graph below shows the data more clearly, with the axis for each subcategory equalised.

AVO + DVA



The trends of Against justice procedures Breach Apprehended Violence Order and Assault Domestic violence related assault for AsDate Year. Color shows details about Against justice procedures Breach Apprehended Violence Order and Assault Domestic violence related assault.

Measure Names

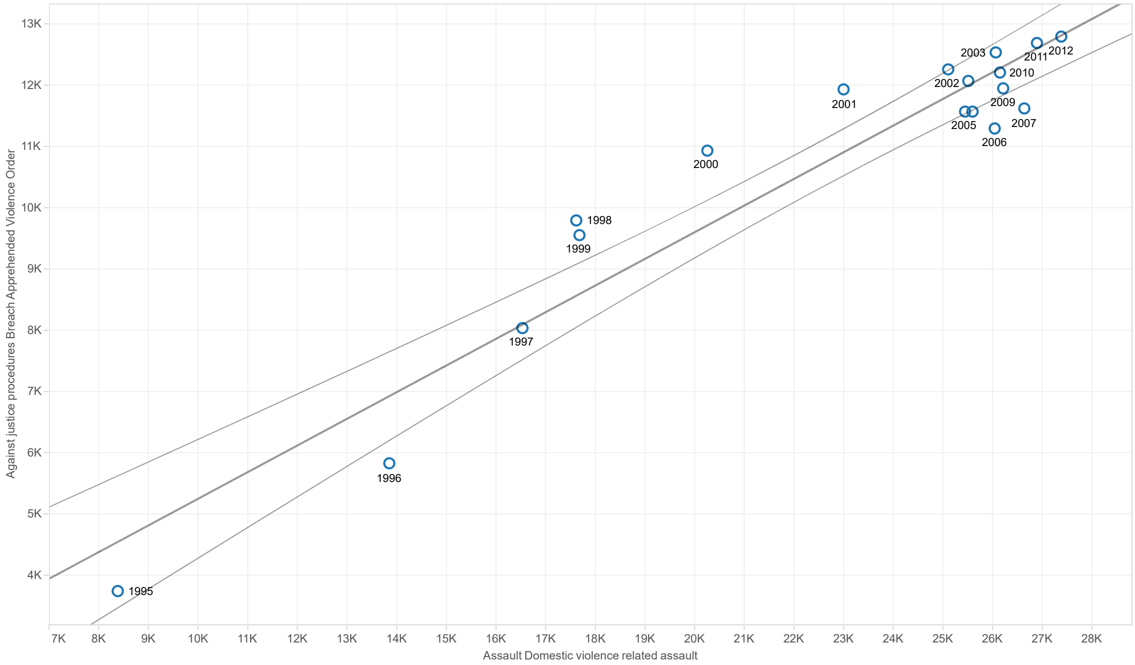
🟦 Against justice procedures Breach Apprehended Violence Order

🟧 Assault Domestic violence related assault

The interesting correlation pattern can still be seen between the two subcategories though, and Tableau allows further analysis of this property with its built in tools.

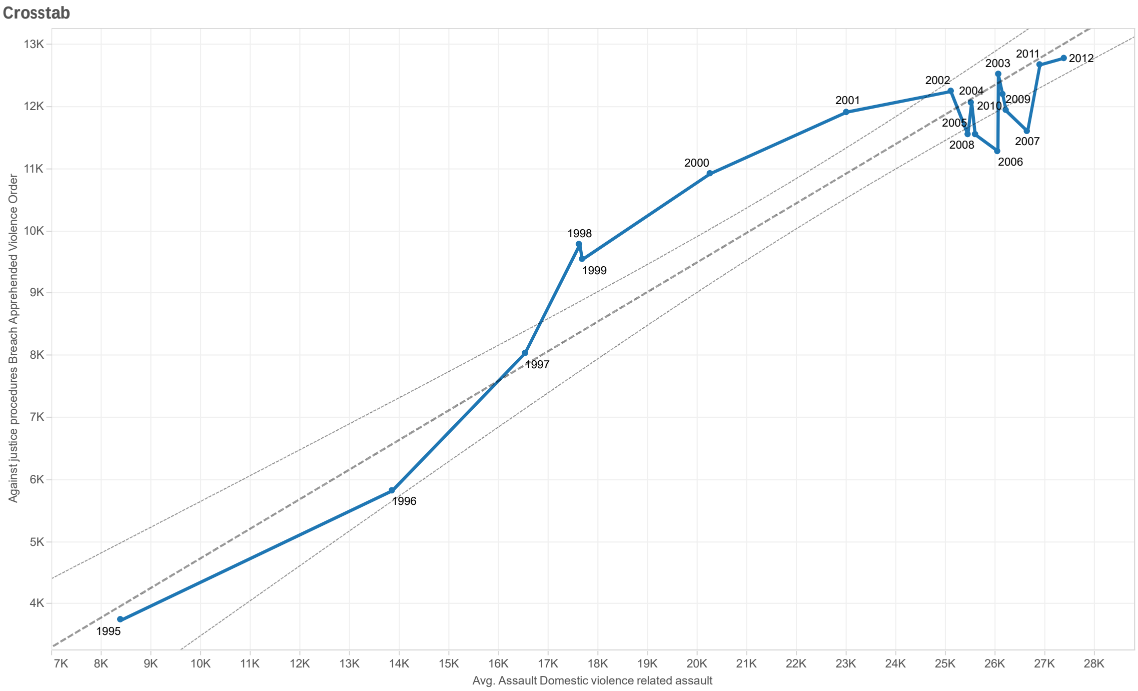
A scatter plot of the subcategories plotted against each other shows a correlation of the increase in rate over years.

Crosstab



Assault Domestic violence related assault vs. Against justice procedures Breach Apprehended Violence Order. The marks are labeled by AsDate Year.

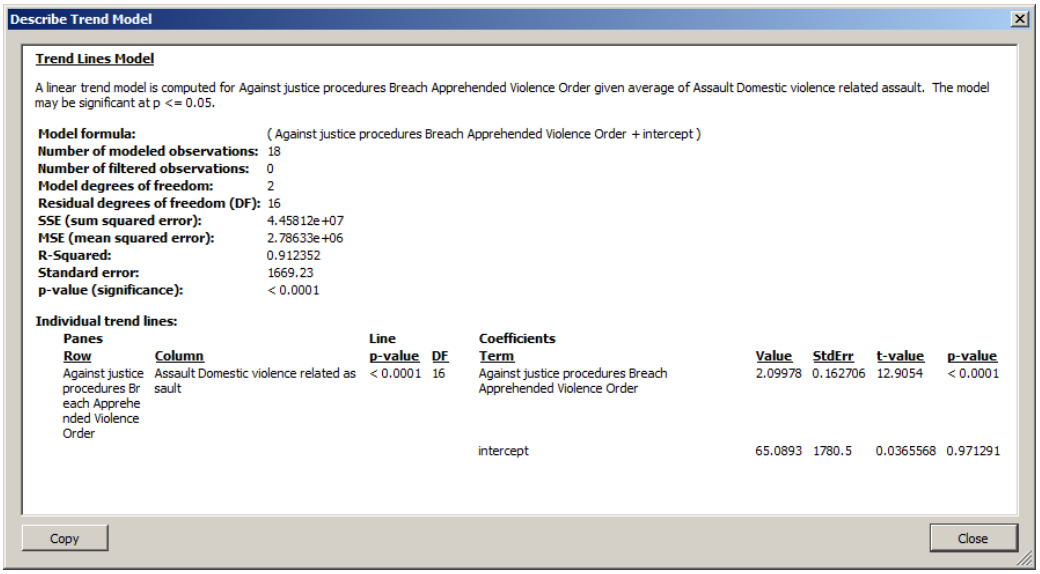
For some reason, Tableau cannot link the points into a line graph unless one of the data points is aggregated, this means using an aggregation function such as **SUM**, **AVG**, **MAX** or **MIN**. Here, **AVG** was used, although it does not change the data values as each set of data is discrete.



The trend of average of Assault Domestic violence related assault for Against justice procedures Breach Apprehended Violence Order. The marks are labeled by AsDate Year.

Linear Regression

A trend line was added to the graph, showing a linear model with 95% upper and lower confidence bands (Tableau-Help, 2016). The trend model details are shown below:



Trend Lines Model

A linear trend model is computed for Against justice procedures Breach Apprehended Violence Order given average of Assault Domestic violence related assault. The model may be significant at p <= 0.05.

Model formula:	(Against justice procedures Breach Apprehended Violence Order + intercept)
Number of modeled observations:	18
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	16
SSE (sum squared error):	4.45812e+07
MSE (mean squared error):	2.78633e+06
R-Squared:	0.912352
Standard error:	1669.23
p-value (significance):	< 0.0001

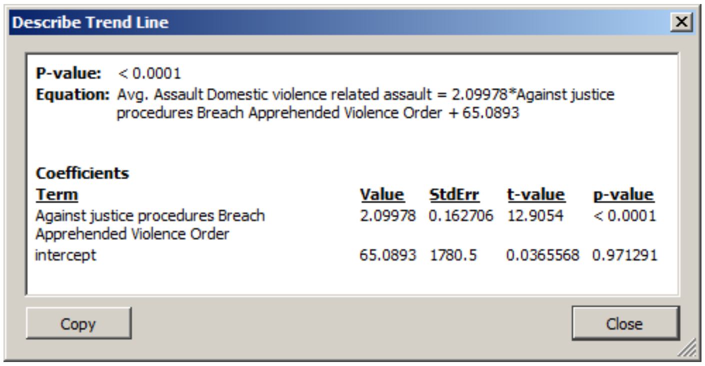
Individual trend lines:

Panes		Line		Coefficients				
Row	Column	p-value	DF	Term	Value	StdErr	t-value	p-value
Against justice procedures Breach Apprehended Violence Order	Assault Domestic violence related assault	< 0.0001	16	Against justice procedures Breach Apprehended Violence Order	2.09978	0.162706	12.9054	< 0.0001
				intercept	65.0893	1780.5	0.0365568	0.971291

The linear model can be described by the following equation:

$$Y = b0 + b1 * X + e$$

where X represents the explanatory variable, Y to response variable, (e) (epsilon) represents random error and b0 and b1 are the coefficients.



P-value: < 0.0001
Equation: Avg. Assault Domestic violence related assault = 2.09978*Against justice procedures Breach Apprehended Violence Order + 65.0893

Coefficients

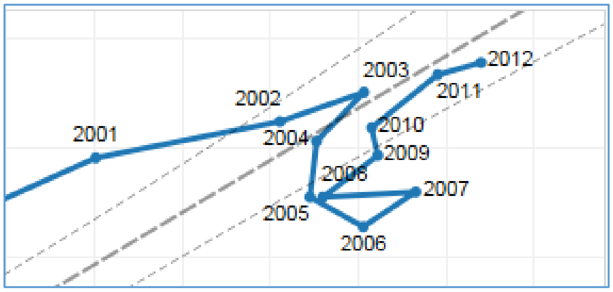
Term	Value	StdErr	t-value	p-value
Against justice procedures Breach Apprehended Violence Order	2.09978	0.162706	12.9054	< 0.0001
intercept	65.0893	1780.5	0.0365568	0.971291

As a translation of the linear equation, the data can be represented as follows:

$$DVA = 2.09978 * BAVO + 65.0893$$

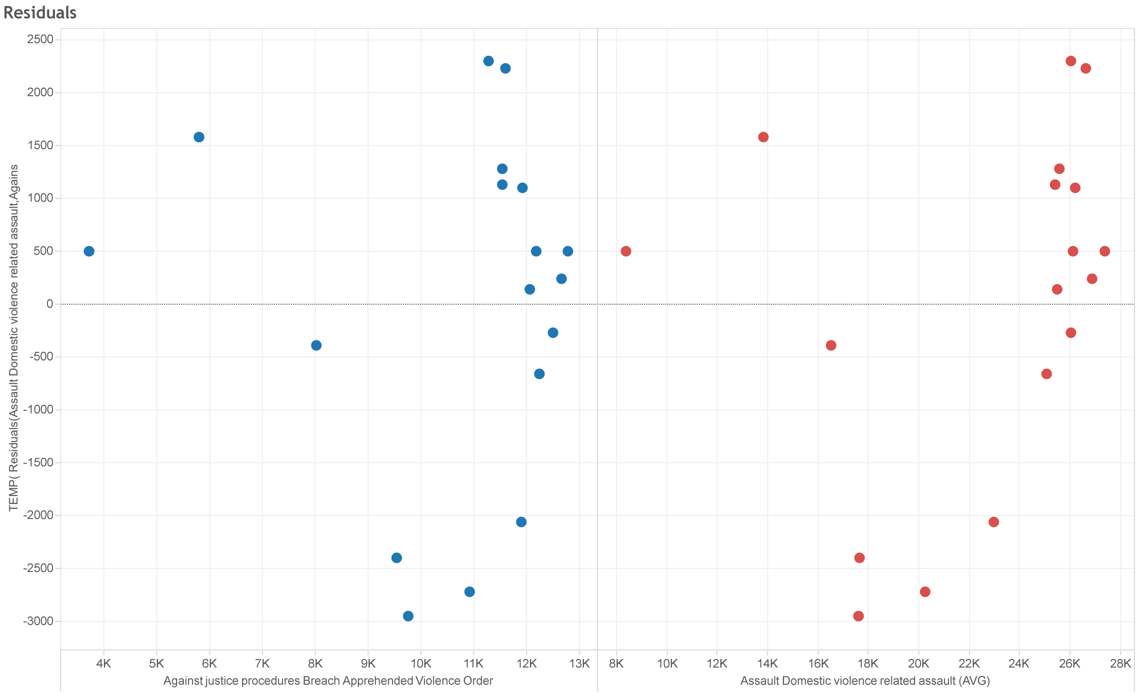
The p-value for the regression is less than 0.001, which shows that the model does have some significance.

However, on observation, there are strange behaviours of the data towards the end of the yearly data. A plot of the line in date order shows the following:



This means that a linear model may not be appropriate. Tableau allows the data from the linear regression model to be exported and re-plotted to show the relationship between the residuals in the model. This technique is used to test whether the data is Normally distributed, and if the linear model is appropriate.

A plot of the residuals versus data can be seen below.



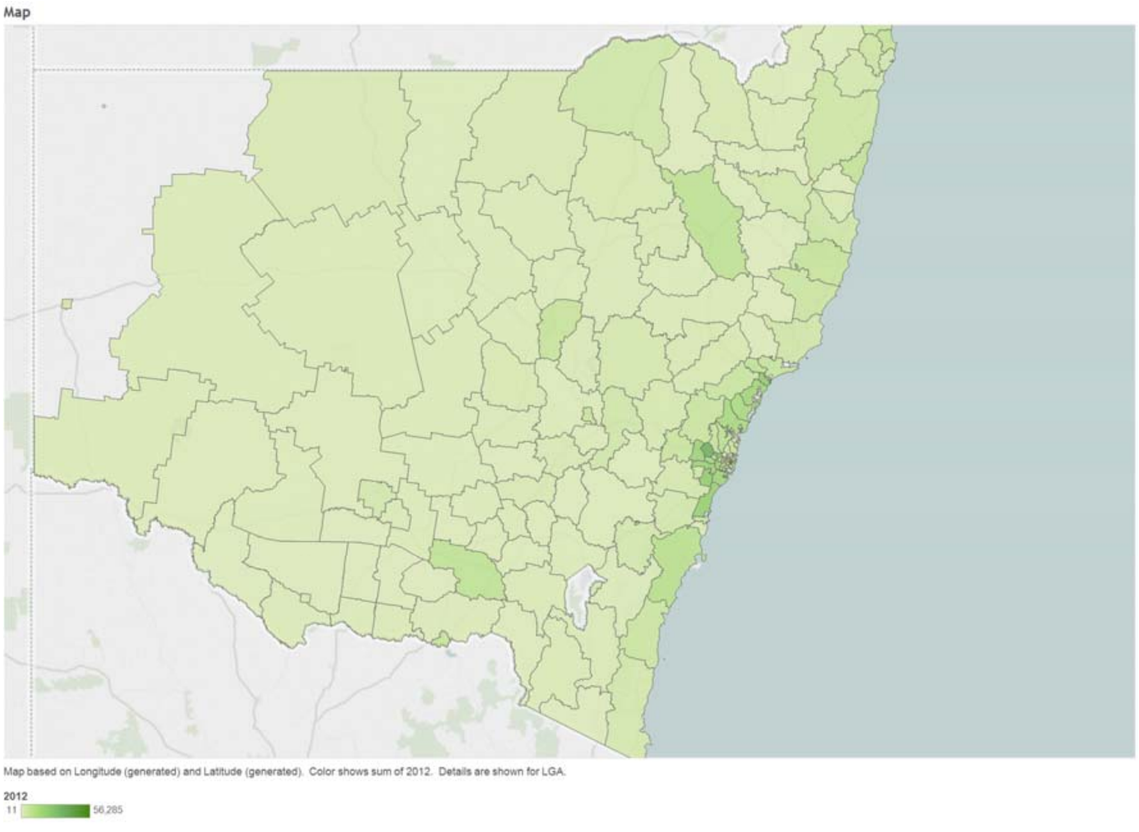
The plots of TEMP(Residuals(Assault Domestic violence related assault,Agains as an attribute for Against justice procedures Breach Apprehended Violence Order and Assault Domestic violence related assault (AVG).

This plot shows residuals distributed fairly randomly. However, it is difficult to know when the data set is small. Further analysis could be done, but is not in the scope of this paper.

Geospatial Graphs

The multidimensional data provided for this paper lends itself to the production of geospatial graphs. In section 1. Data Gathering, Local Government Area (LGA) was converted to a geographical property, enabling the mapping of crime statistics over a map of NSW with LGAs marked into polygons.

Crime data attributes can be associated with locations and compared visually. For example, below is a choropleth graph, called a ***filled map*** in Tableau, portraying number of offences per LGA for 2012 data.

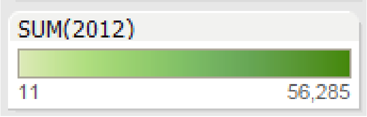


Map based on Longitude (generated) and Latitude (generated). Color shows sum of 2012. Details are shown for LGA.

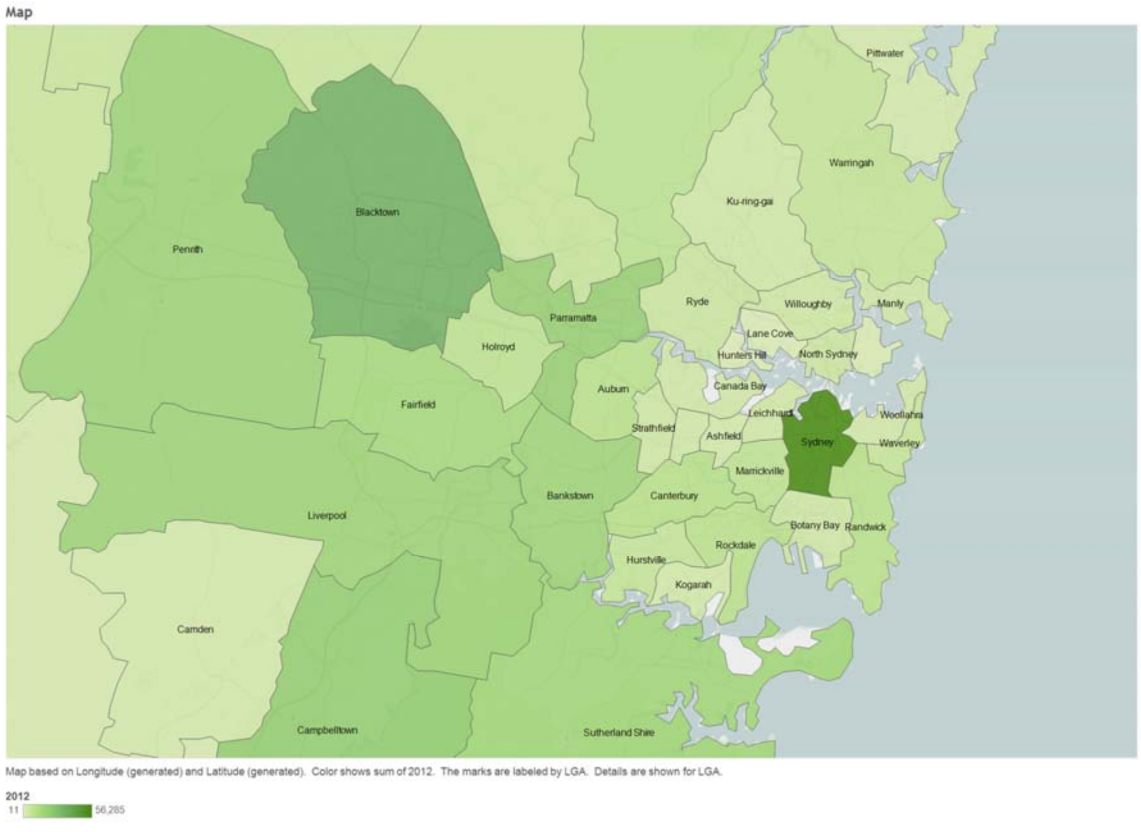
When mapping quantitative data, a specific colour progression should be used to depict the data properly (Wikipedia, Coropleth Maps, 2016). There are several different types of colour progressions used by cartographers. Single-hue progressions fade from a dark shade of the chosen colour to a very light or white shade of relatively the same hue. This is a common method used to map magnitude. The darkest hue represents the greatest number in the data set and the lightest shade representing the least number (Robinson, 1995).

Choropleth maps are sometimes incorrectly referred to as heat maps. A choropleth map features different shading or patterns within geographic boundaries to show the proportion of a variable of interest, whereas the coloration a heat map (in a map context) does not correspond to geographic boundaries (Wikipedia, Heat Maps, 2016).

Assigning measure values to colour ranges in Tableau enables easy visual analysis of data. For the following colour palette, lighter green corresponds to areas of low crime, and darker green to areas of high crime.

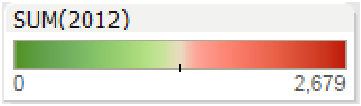


Zooming in on the city of Sydney shows how crime is distributed across LGAs.



Map based on Longitude (generated) and Latitude (generated). Color shows sum of 2012. The marks are labeled by LGA. Details are shown for LGA.

A more visually striking colour pattern is used here, a reversed red-green diverging pattern, where red indicates areas of high measure values, and green areas of low values.

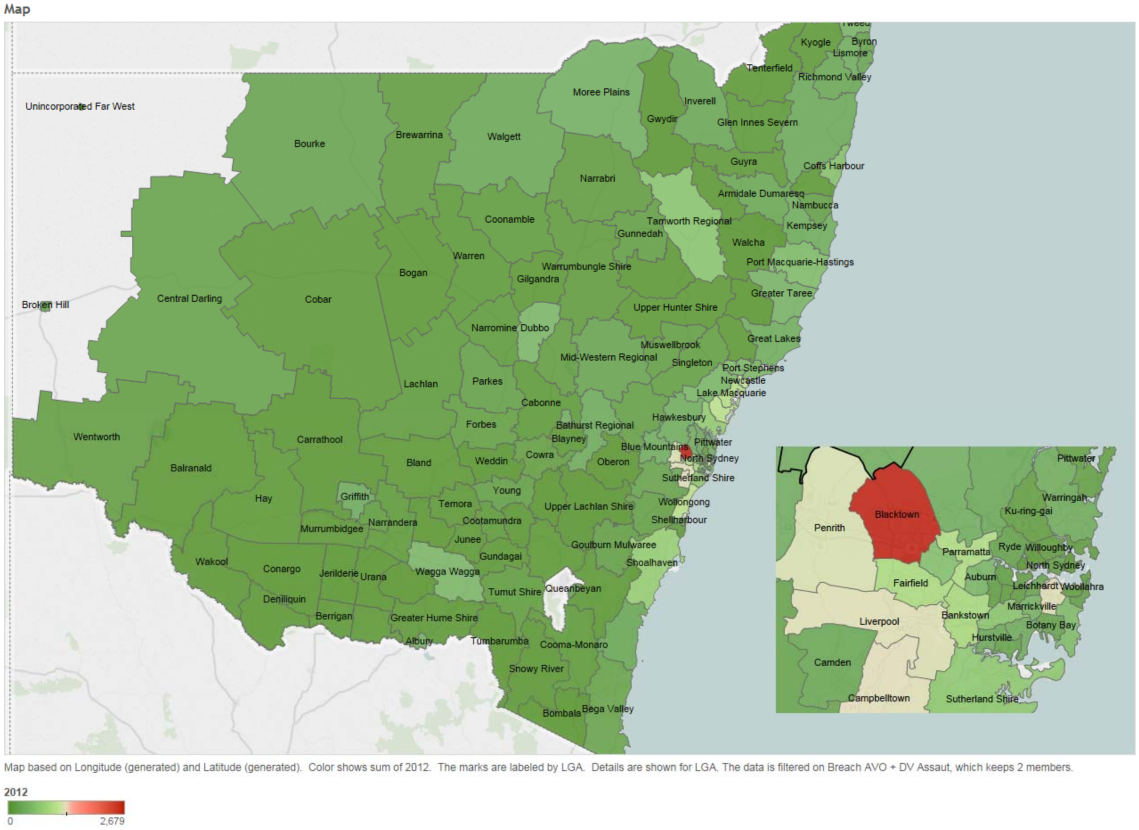


This bipolar colour progression is normally used with two opposite hues to show a change in value from negative to positive or on either side of some either central tendency, such as average values or temperature. When one extreme can be considered better than the other (as here with volume of crime) then it is common to denote the poor alternative with shades of red, and the good alternative with green (Wikipedia, Coropleth Maps, 2016).

Crime Rates and Comparisons

However, the amount of crime per LGA does not necessarily provide the whole picture. For example, filtering the subcategory to show the map with our DVA and BAVO data shows the following map:

Map

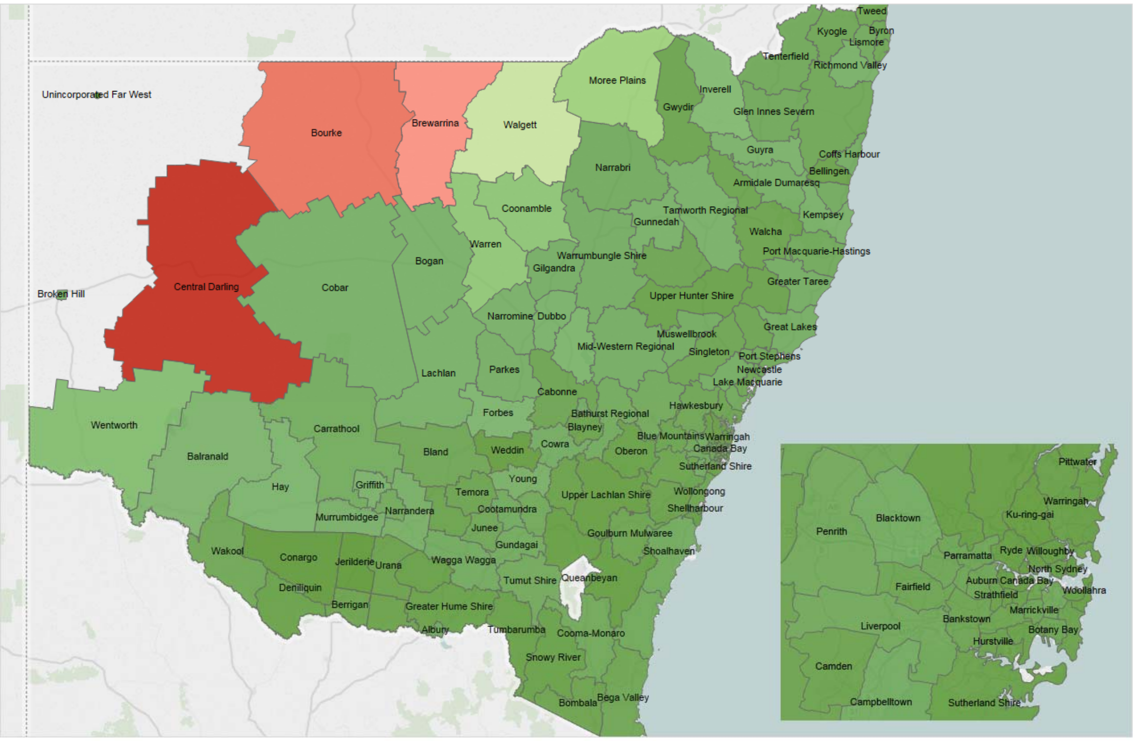


Map based on Longitude (generated) and Latitude (generated). Color shows sum of 2012. The marks are labeled by LGA. Details are shown for LGA. The data is filtered on Breach AVO + DV Assault, which keeps 2 members.

As expected, there are larger amounts of crime in the city of Sydney as compared to regional areas.

But when mapping the ***rate*** of crime by comparing it to the LGA population, a very different outcome emerges. Below is the same subcategories of Domestic Violence Assault and Breach Apprehended Violence Order for 2012 data:

Map



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Rate per 1000 (2012). The marks are labeled by LGA. Details are shown for LGA. The data is filtered on Breach AVO + DV Assaut, which keeps 2 members.

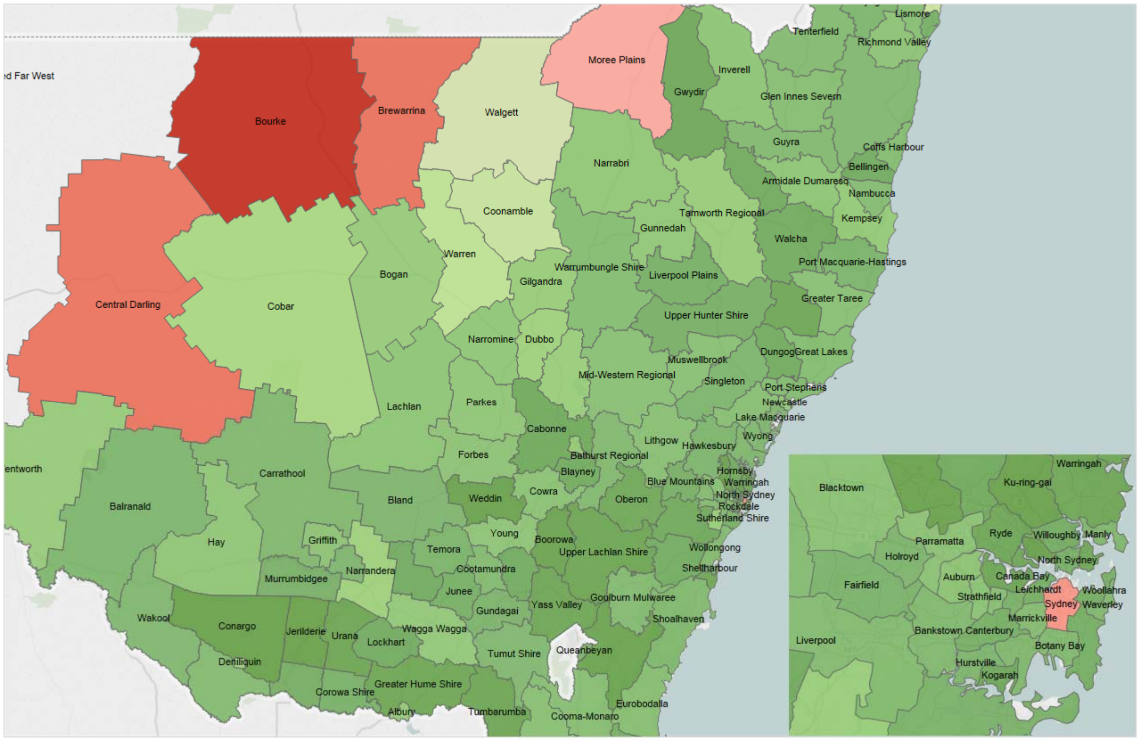
Rate per 1000 (2012)

0.00 86.41

This choropleth map shows that the rate per 1000 of population for the subcategories selected is actually higher in some regional areas, and lower in the city suburbs. Central Darling, Bourke and Brewarrina show the highest rates of offenses on a population factor as compared to Sydney city areas such as Blacktown which previously shown the largest amount of offences.

A similar result is shown for the rate of total offences per 1000 of population:

Map



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Rate per 1000 (2012). The marks are labeled by LGA. Details are shown for LGA.

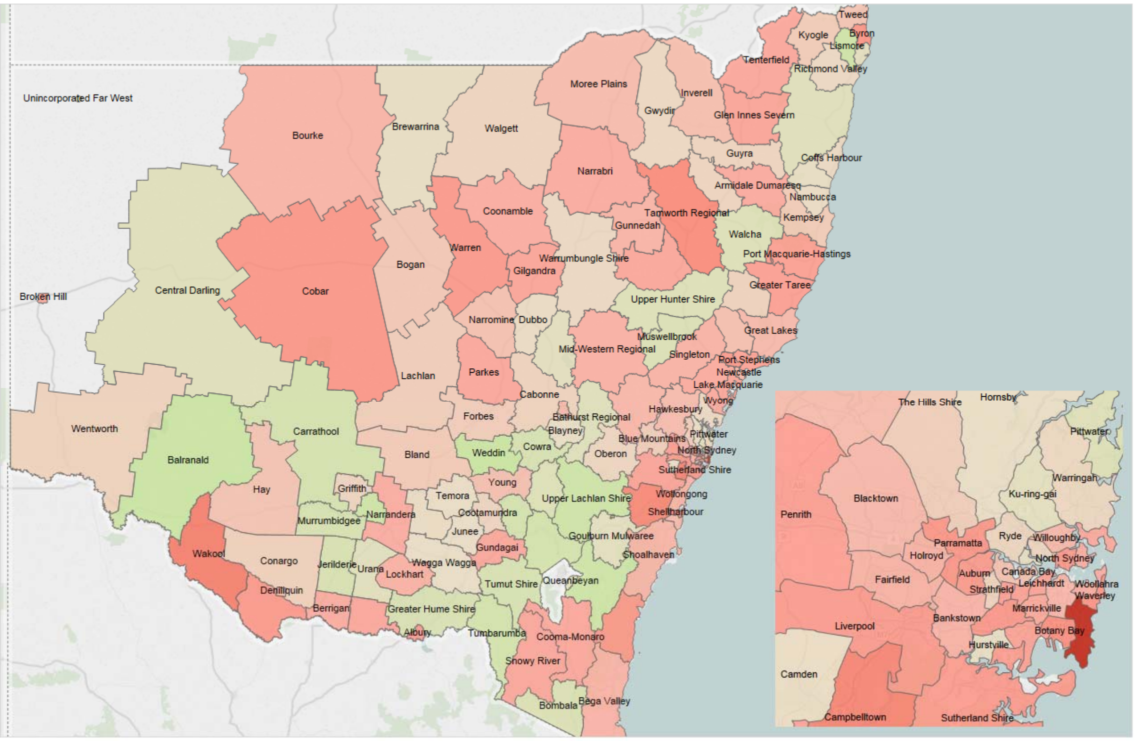
Rate per 1000 (2012)
9.7 476.4

Change in Rate of Crime

Another important measurement in the analysis of crime data is the change in rate of crime. This measurement was calculated during data mining and enrichment (section 2. Data Mining and Enrichment), and shows the change in rate of crime between 2001 and 2012.

As shown in the following choropleth map, this analysis shows varying results. Some LGAs show rising rates of crime, and other LGAs show reducing rates of crime.

Map



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Change in Rate (2001-2012). The marks are labeled by LGA. Details are shown for LGA.

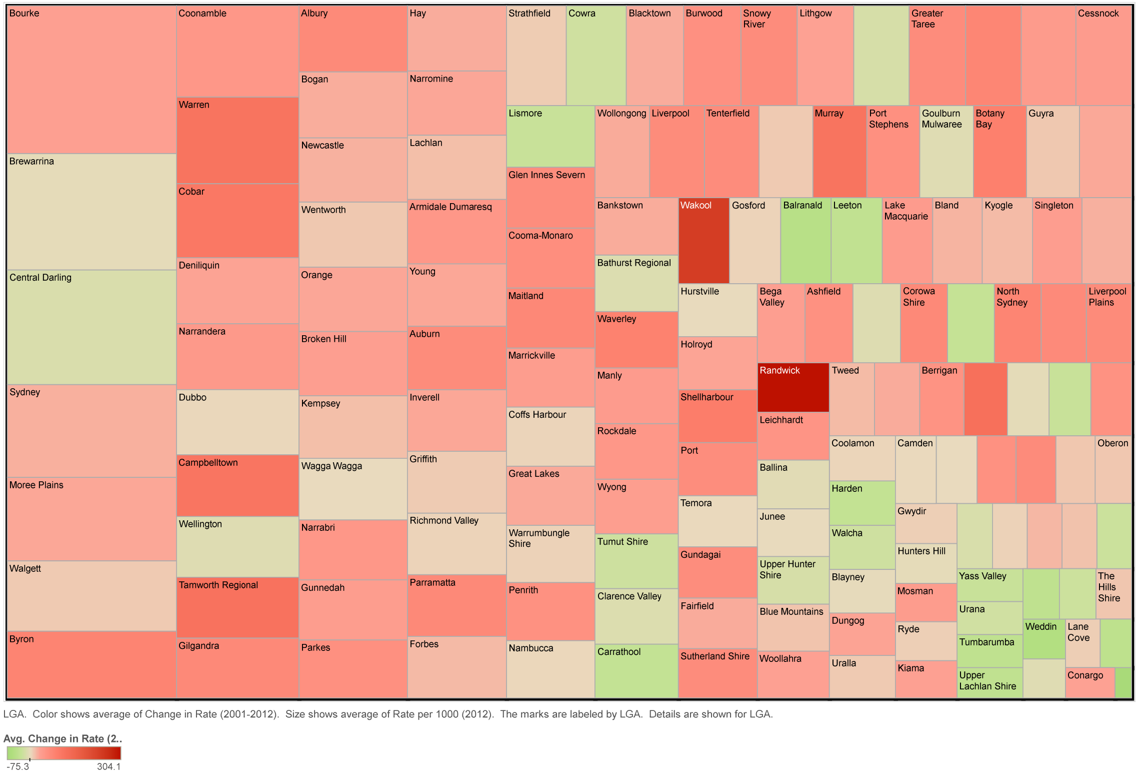
Change in Rate (2001-2..

-1,923 17,335

Tree Graphs

Although the geographic map above shown total rates of crime, it could be that some categories of crime are rising and some categories are falling. The following tree graph has been configured like a heat map, with the Change in Rate of Offences set as the colour mark, and the Rate per 1000 for 2012 data set as the size.

Heat/Tree Map

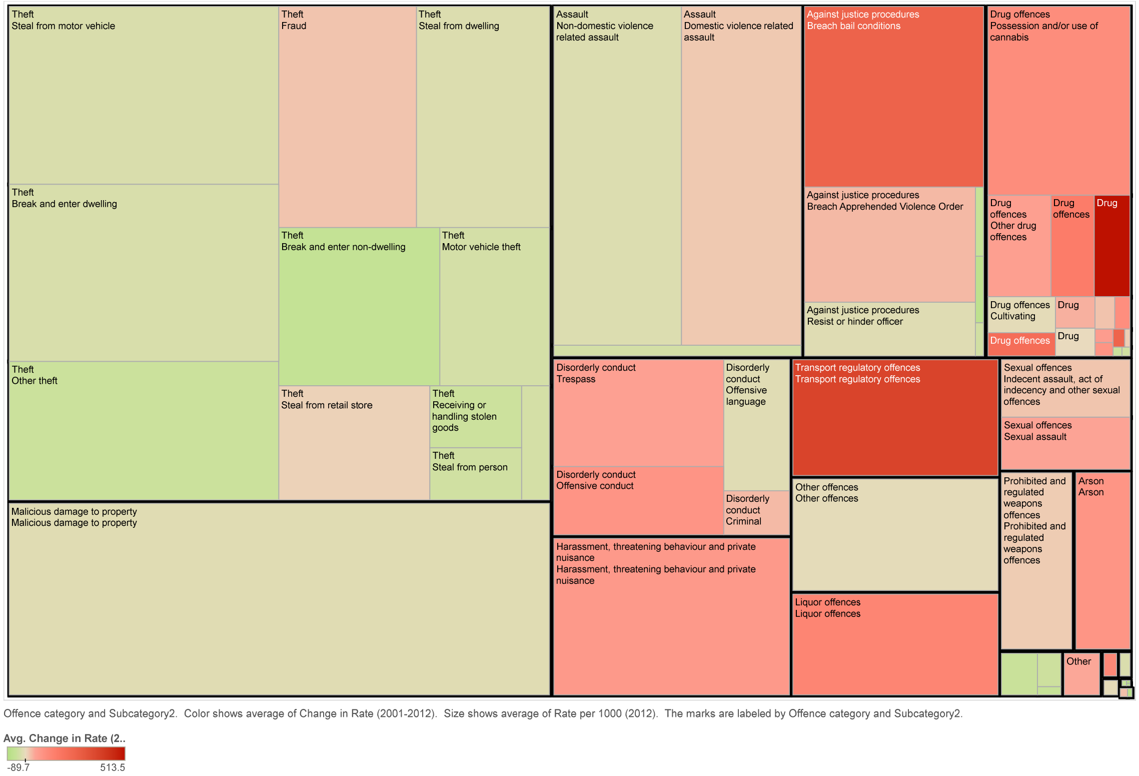


LGA. Color shows average of Change in Rate (2001-2012). Size shows average of Rate per 1000 (2012). The marks are labeled by LGA. Details are shown for LGA.

This analysis can show some interesting properties. For example LGAs such as Randwick and Rakool, with low rates of crime, have the highest increase in crime rates. Other LGAs, mentioned previously, such as Bourke, Brewarrina and Central Darling have varying changes in crime rate, even though they have the largest rates per 100 population. Bourke goes up, while Brewarrina and Central Darling go down.

Generating a tree graph using subcategories can also show whether specific crime offense rates are rising or falling. See below for a graph where Change in Rate (2001 to 2012) is allocated to the colour mark, and Rate per 1000 (2012) is assigned to size.

Heat/Tree Map



Offence category and Subcategory2. Color shows average of Change in Rate (2001-2012). Size shows average of Rate per 1000 (2012). The marks are labeled by Offence category and Subcategory2.

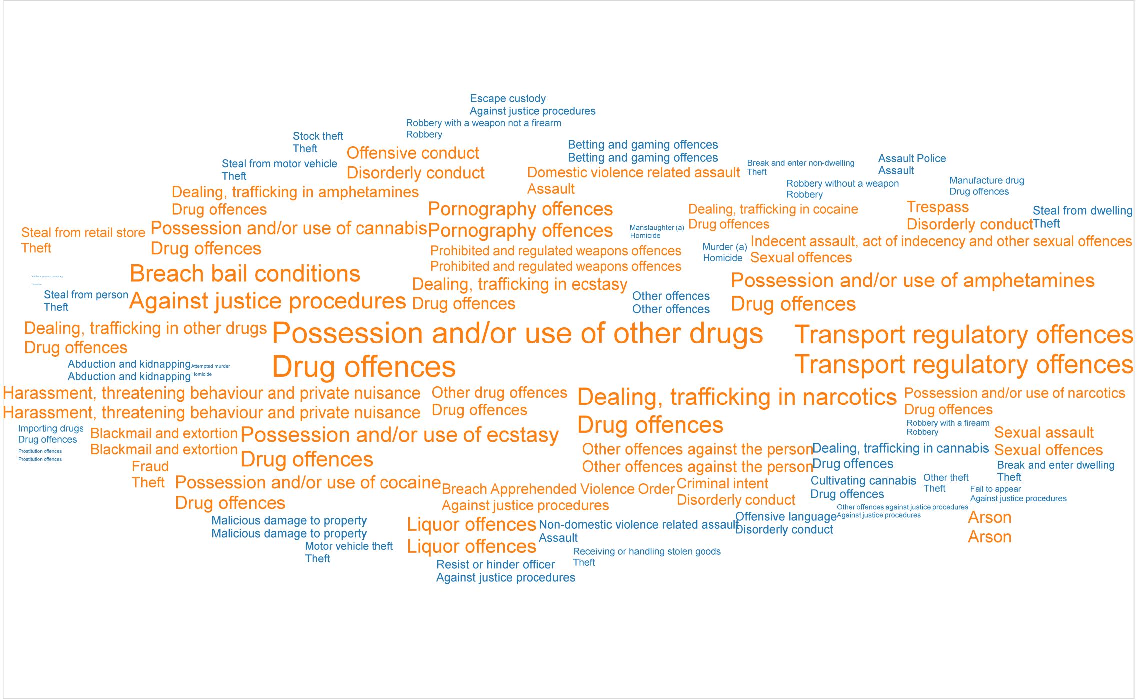
Interesting properties can be seen, such as an overall decrease in rates of most thefts, non-domestic violence assaults and malicious property damage, but increases in domestic violence assaults, drug offenses and transport regulatory offences (includes offences on the rail network such as travelling without a valid ticket, smoking, drinking or using offensive language on a train or railway land).

Subsets

Tableau allows the configuration of ***sets***, where data is grouped by a calculated field, in this case whether the crime rate has increased or decreased.

Interesting visualisations can be generated from this data, such as a word map, where size relates to the amount of change, and colour relates to the value rising or falling.

In/Out



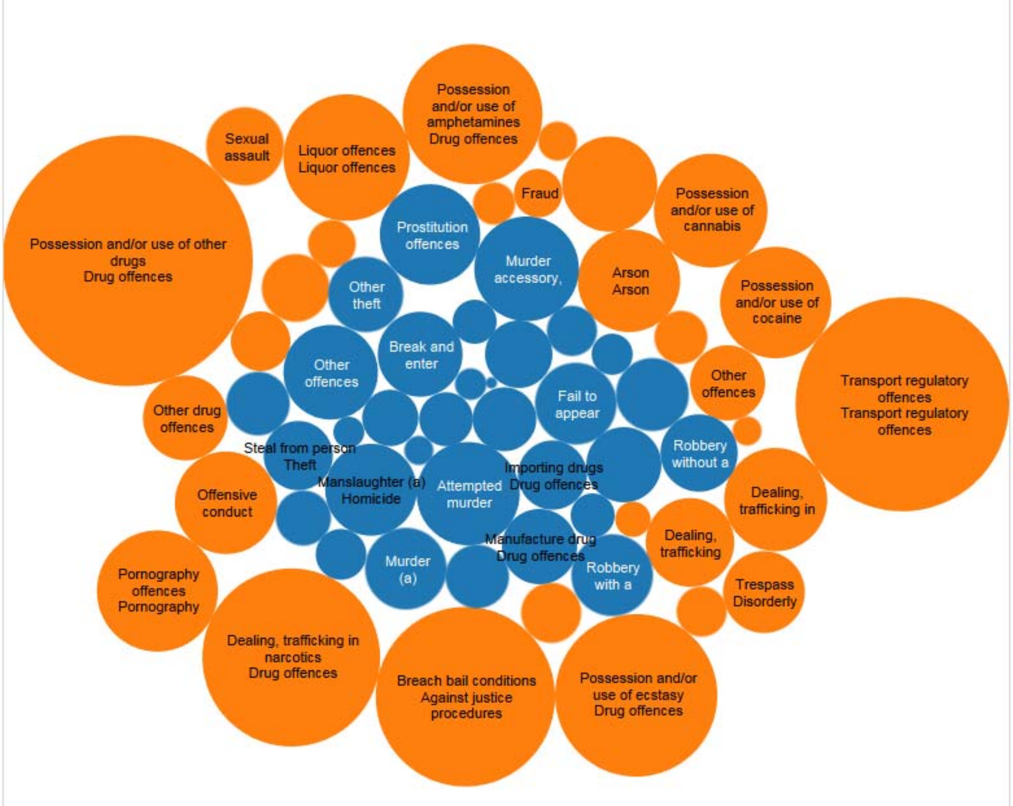
Subcategory2 and Offence category. Color shows details about In / Out of Rate Change < 0. Size shows average of Change in Rate (2001-2012). Details are shown for In / Out of Rate Change < 0.

In / Out of Rate Change < 0

🟦 In

🟧 Out

Or a packed bubble map, conveying the most relevant rate changes by size and colour.



Discussion

Visualisation Methods

Visualisation methods for multidimensional data include:

- Mapping onto 2D or 3D space before visualisation
- Parallel coordinate plots
- Star plots
- Scatter plot matrix
- Linked Scatter plots
- Linked histograms
- Mosaic plots
- Glyphs, Icons
- Pixel-based Visualisations and heat maps

This paper shows the limitations of line graphs and bar graphs for large scale analysis, and shows that interesting analysis can be performed using geospatial modelling, tree graphs and heat maps.

Other visualisation methods such as multiple scatter plots or multiple line charts are difficult to read because of the large number of dimensions in the data set.

Visualisation Tools

Tableau provides a broad tool set to analyse multidimensional (or hypervariate) data such as the BOCSAR crime data provided. However, visualisations of this data can encounter problems related to the size and number of dimensions of the data. Careful planning and analysis should take place in order to generate useful graphs.

Tableau was selected as the visualisation tool for this paper because it provides a variety of visualisation methods for analysis, it can be used to combine data from various data sources, and provides tools to enable the manipulation of data to analyse and investigate interesting properties of the visualisations.

Conclusions

Multidimensional data visualisation involves the observation and analysis of more than one measure variable at a time. In design and analysis, the method is used to perform analysis across multiple dimensions while taking into account the effects of all variables on the responses of interest (Wikipedia, Multivariate Analysis, 2016).

It is easy to become distracted by the large amount of dimensions in the data, and planning and preparation must be undertaken before any useful visualisations can be generated.

This paper developed several visualisation techniques to view the multidimensional data, and showed that interesting analysis can be performed using geospatial modelling, tree graphs and heat maps.

Other, more detailed analysis can be performed by filtering and sub setting the data, such as extracting specific offence categories or subcategories of interest, and finding patterns and relationships. However, this method reduces the dimensionality to a more manageable level in order to generate useful visualisations.

References

ABS. (2016, 06 01). *3218.0 - Regional Population Growth, Australia, 2014-15* . Retrieved from Australian Bureau of Statistics: http://www.abs.gov.au/AUSSTATS/abs@.nsf/MF/3218.0

BOCSAR. (2016, 06 01). *Bureau of Crime Statistics and Research*. Retrieved from Bureau of Crime Statistics and Research: http://www.bocsar.nsw.gov.au/

Jane Law, M. Q. (2014). Bayesian Spatio-Temporal Modeling for Analysing Local Patterns of Crime Over Time. *J Quant Criminol*, 3-:57-78.

Robinson, A. M. (1995). *Elements of Cartography* (6th Edition ed.). New York: Wiley.

Spence, R. (2001). *Information Visualization - An Introduction, 3rd Edition.* London, UK: Springer.

Tableau. (2016, 06 01). *Tableau Products*. Retrieved from Tableau Software: http://www.tableau.com/products/desktop

Tableau-Help. (2016, 06 01). *Online Help* . Retrieved from Tableau Desktop: http://onlinehelp.tableau.com/v9.3/pro/online/windows/en-us/help.html

Wikipedia. (2016, 06 01). *Coropleth Maps*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Choropleth_map

Wikipedia. (2016, 06 01). *Heat Maps*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Heat_map

Wikipedia. (2016, 06 01). *Multivariate Analysis*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Multivariate_analysis