

Queensland Crime Data Analysis

Roy Portas - s4356084

April 12, 2017

1 Introduction

Criminology is the study of causes, nature and prevention of criminal behaviour[8]. This study is important as it allows sociologists and psychologists to not only study the when and how crime occurs, but offers a way to predict crime trends in the future.

Crime is an evident part of our society and is an active topic during elections[7], even companies like IBM[2] and Palantir[6] are creating products to model, predict and prevent crimes before they happen.

As technology becomes more powerful and less expensive, the cost of implementing systems capable of predicting crimes becomes more feasible. However building these systems are difficult due to the sheer amount of data that needs processing. Nevertheless companies have started building such systems.

Crime prediction systems generally consist of the following components:

- Identify areas with frequent crime[2]
- Match trends in national and regional crimes with local crimes
- Identify circumstances for the cause of crimes[2]

2 Aims

This report aims to analyse and explore crime data for Queensland. A variety of techniques and datasets will be used to explore this problem, including:

- Analyse monthly offences to determine correlations during certain months of the year.
- Analyse reported offences by suburb, month and offence type to find relationships in crime over time.
- Analysis of geolocation data for crime locations and police districts to find divisions with high crime.

Data Source	URL
QPS Police Stations	https://data.qld.gov.au/dataset/qps-police-stations
Offence numbers by police division	https://data.qld.gov.au/dataset/offence-numbers-police-divisions-monthly-from-july-2001

3 Methods

3.1 Data Sourcing

The data was downloaded from the Queensland Government data website, data.qld.gov.au. The following table shows the data sources used for this report.

The files were provided as CSV and Shapefiles, which are both common storage formats for this kind of data. The files are maintained and updated by Queensland Police, who update it every month with new crime data. As this data comes directly from the Queensland Police service, it should be accurate and reliable. The data is accurate to individual months and police divisions, which gives more than enough detail to model crimes in Queensland.

Both CSV and Shapefiles are easy to import into Matlab, thus are suitable datasets to use.

3.2 Data Processing

The first dataset used is the offence rates per police division[1], which is a CSV file with 65000 rows and 90 columns. This is a lot of data to process, thus the data was preprocessed and saved in a Matlab format (.mat). This allowed the data to be easily loaded without having to process it again. Since the dataset has 90 dimensions, it is impossible to visualize it, so the total number of offences was found. This simple method of dimensionality reduction allowed easy plotting of the data to find crime trends in districts over time.

The second dataset is stored in a Shapefile format, which is a common format for geospatial vector data. This dataset was also preprocessed, as the shapefile stores the points of every boundary for every police region in Queensland. With the data processed, a map was created showing the boundaries of every police district in Queensland, see figure 6 for a visualization of the boundaries.

Both datasets use police divisions, thus allows the datasets to be joined. However the formatting of the police divisions are inconsistent, with one dataset using **BRISBANE CITY** with the other using **Brisbane City**. This was fixed by converting all the names to lowercase.

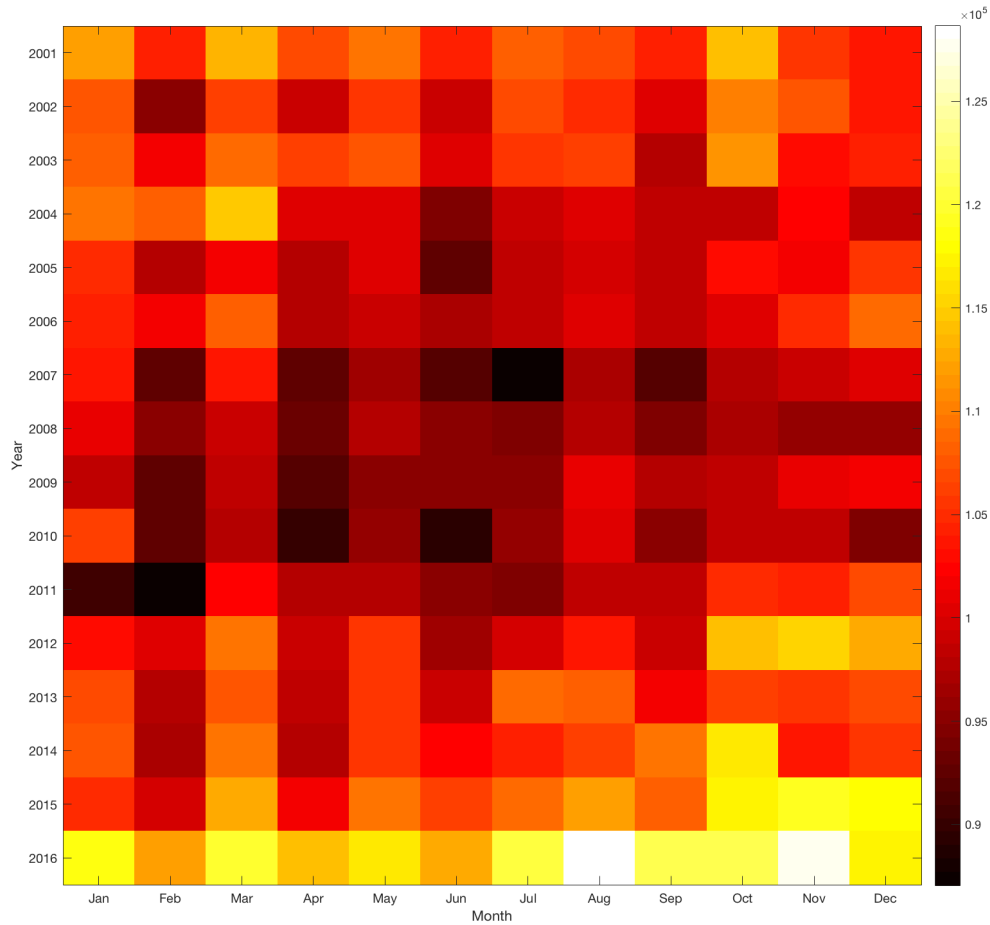
3.3 Data Analysis

4 Results

4.1 Analysing historic crime data for Queensland

The offences per district were aggregated into totals for every month. This shows the total number of offences in Queensland each month between January 2001 and December 2016.

Figure 1: Heatmap of crime over time

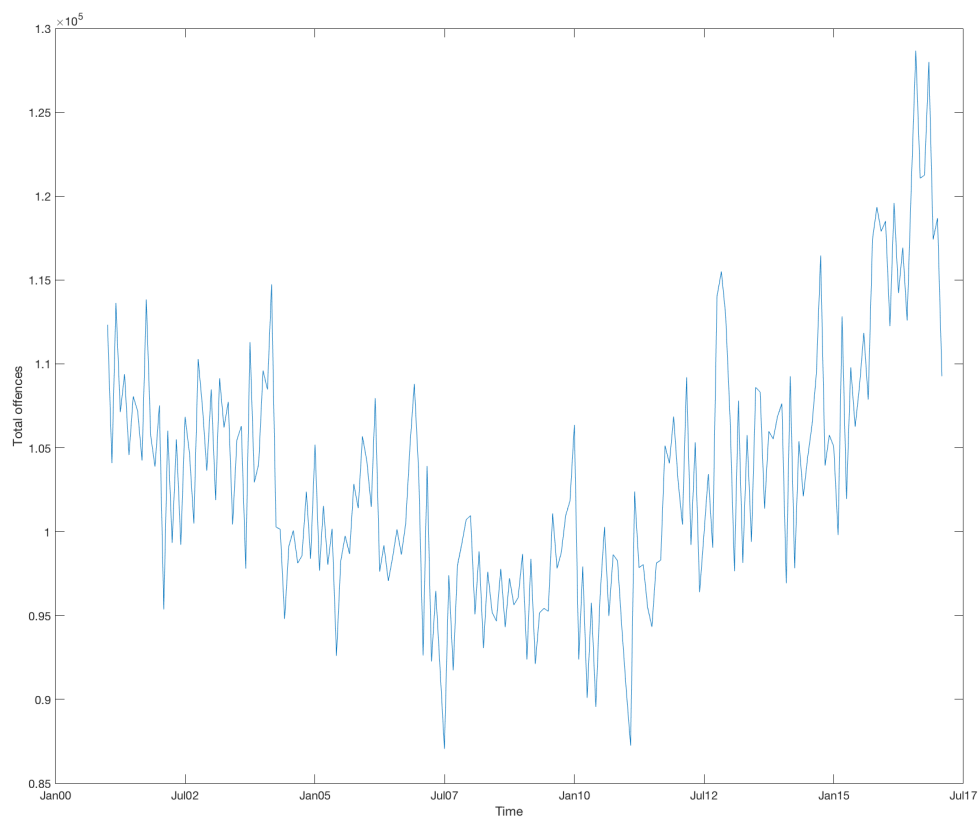


The heatmap shows that more crime occurs during October, November, December and January than the other months. Additionally overall offences has been steadily increasing since 2014, with 2016 having by far the most offences for any year on record. August and November of 2016 has spikes in the number of offences.

The data also shows a significant drop in crime from December 2010 to February 2011. This correlates with the 2010-2011 floods in Queensland, in which three quarters of the council areas in Queensland became disaster zones[4]. The heatmap indicates that crime dramatically decreased during that time but picked up quickly in March.

A spike in crime was noticed during March 2004, this can also be seen in the other charts below, historical records show that March was when the Brisbane City Council elections happened and Campbell Newman became the Lord Mayor of Brisbane[3]

Figure 2: Total offences over time



The data was also plotted in a line chart to display the total offences over time. The chart shows that overall crime has been decreasing from 2001 through to 2010, but starting 2011 crime rates started increasing with the peak in 2016.

Additionally the chart shows that there is a lot of fluctuation from month to month.

Figure 3: Total offences over time with 95% confidence intervals

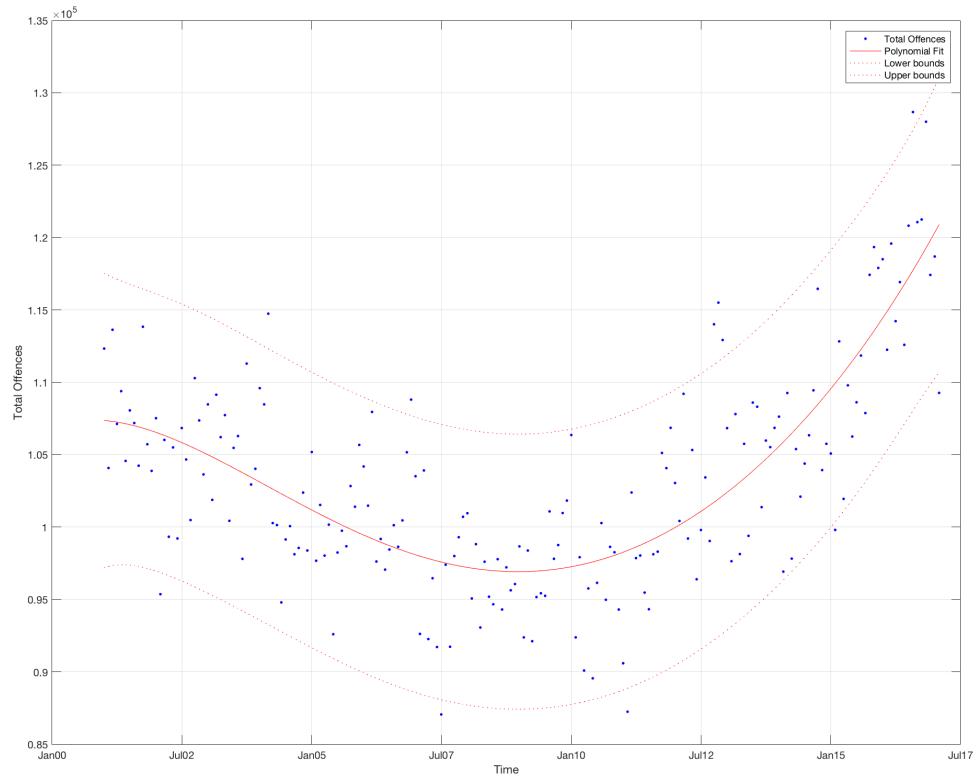


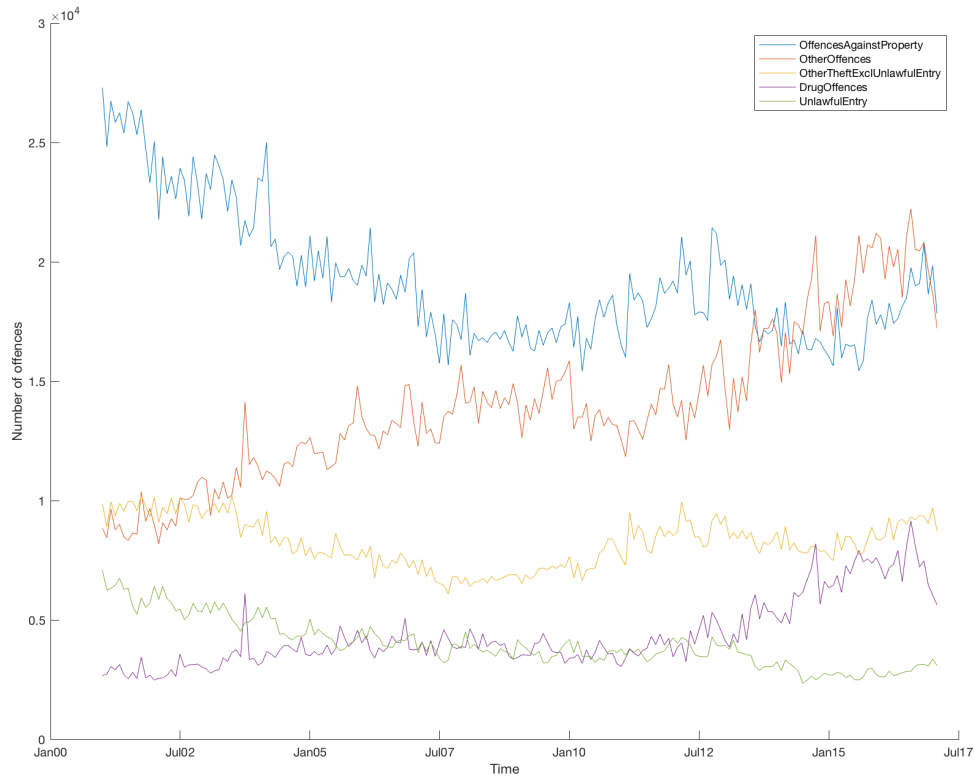
Figure 3 displays the 95% confidence intervals when the data is plotted to a 5th degree polynomial. The points above and below the confidence intervals are outliers, which could suggest that events occurred during that time to cause the crime rates to increase or decrease.

Sorting the crimes by the average number of offences determined that the most common crimes are:

1. Offences against property
2. Other offences
3. Other theft exculding unlawful entry
4. Drug offences
5. Unlawful entry

These offences can be plotted against the months to see how they change over time.

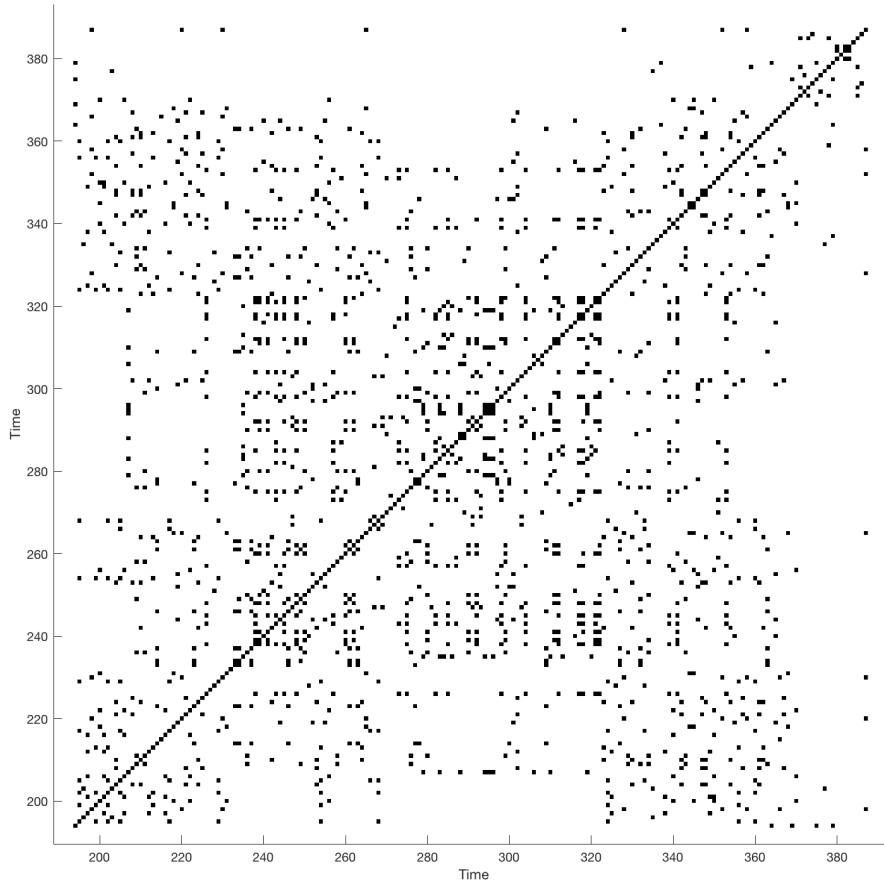
Figure 4: Top offences over time



Displaying the types of offences over time shows a few trends. Offences against properties has been on the decline, whereas other types of offences has been increasing. Here ‘other’ means the offences that do not fit any existing categories according to the Queensland Police. Additionally drug related offences has been increasing, but started to fall again at the start of 2017.

There is a spike in crime related offences between 2003 and 2004, where drug offences, other offences and offences against properties all spiked for a short period. This suggests that there could have been an event in 2003 that caused these to increase.

Figure 5: Recurrence Plot

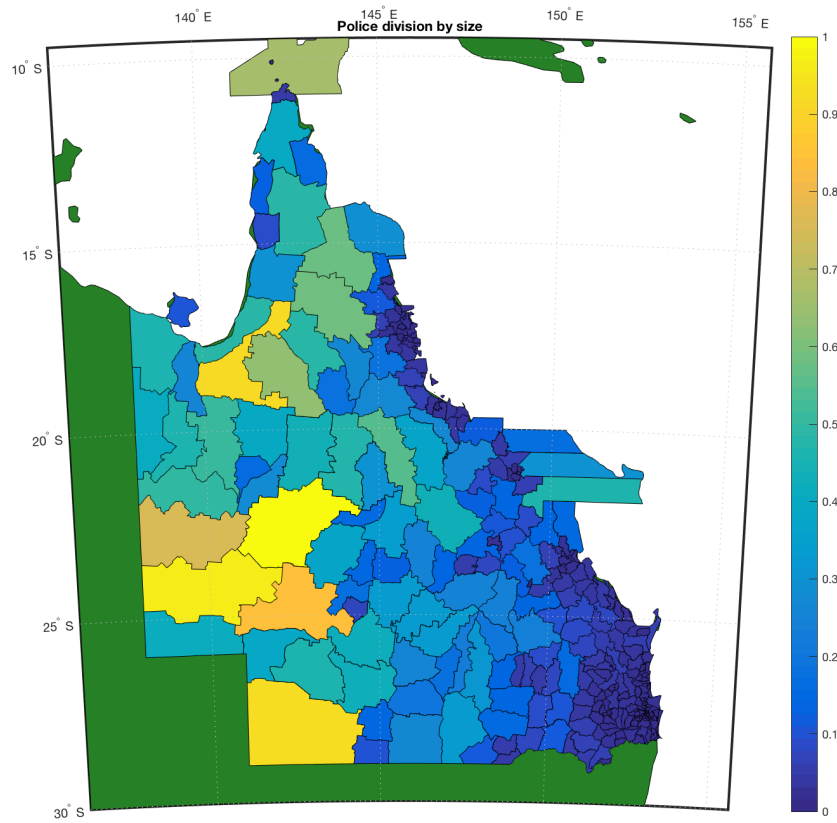


The above graph is a recurrence plot of the crime data over time. This shows how the data recurs over time.

4.2 Correlating crime data with police divisions

The crime data can be correlated against police divisions on a map, which gives a geospatial representation of the data. For this analysis only the crimes occurring in February 2017 is used.

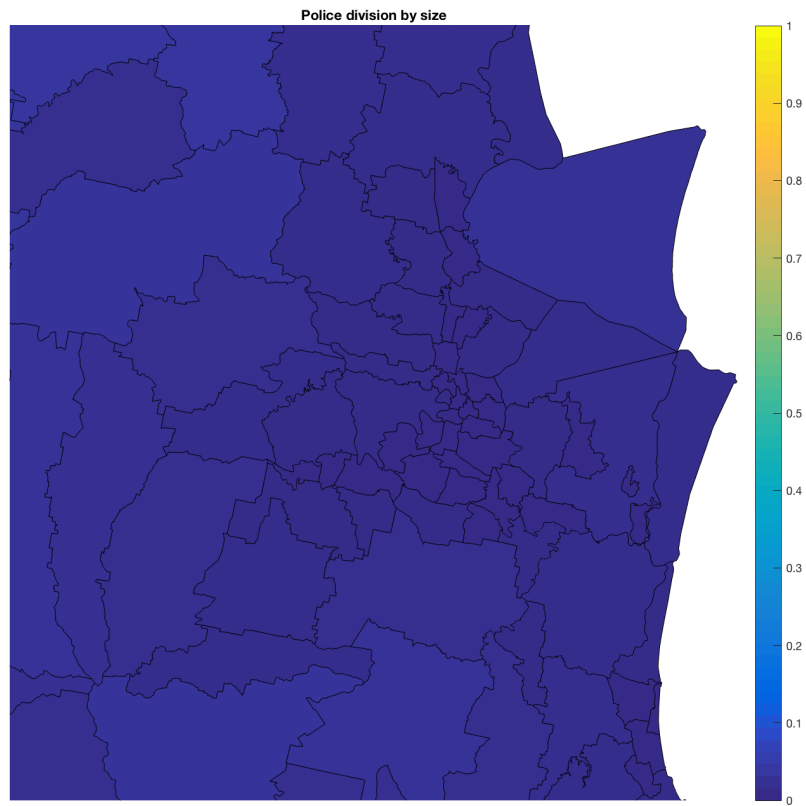
Figure 6: Police Division by Size



The police divisions in Queensland can be plotted onto a map. The divisions with more land area are displayed in a lighter colour. The solid green is the base map, thus an area that the Queensland Police force does not patrol.

Larger police divisions indicate more rural areas with smaller populations, whereas smaller police divisions are within cities with a larger population density. This makes sense as the crime rate increases as the population of a given area increases[5]. So logically the Queensland Police will have more divisions where the crime rate is higher.

Figure 7: Police Division by Size (Brisbane Area)



The police divisions around the Brisbane area are much smaller than anywhere else in Queensland, with the smallest ones being around Brisbane City and Fortitude Valley.

Figure 8: Police Division by Crime Density

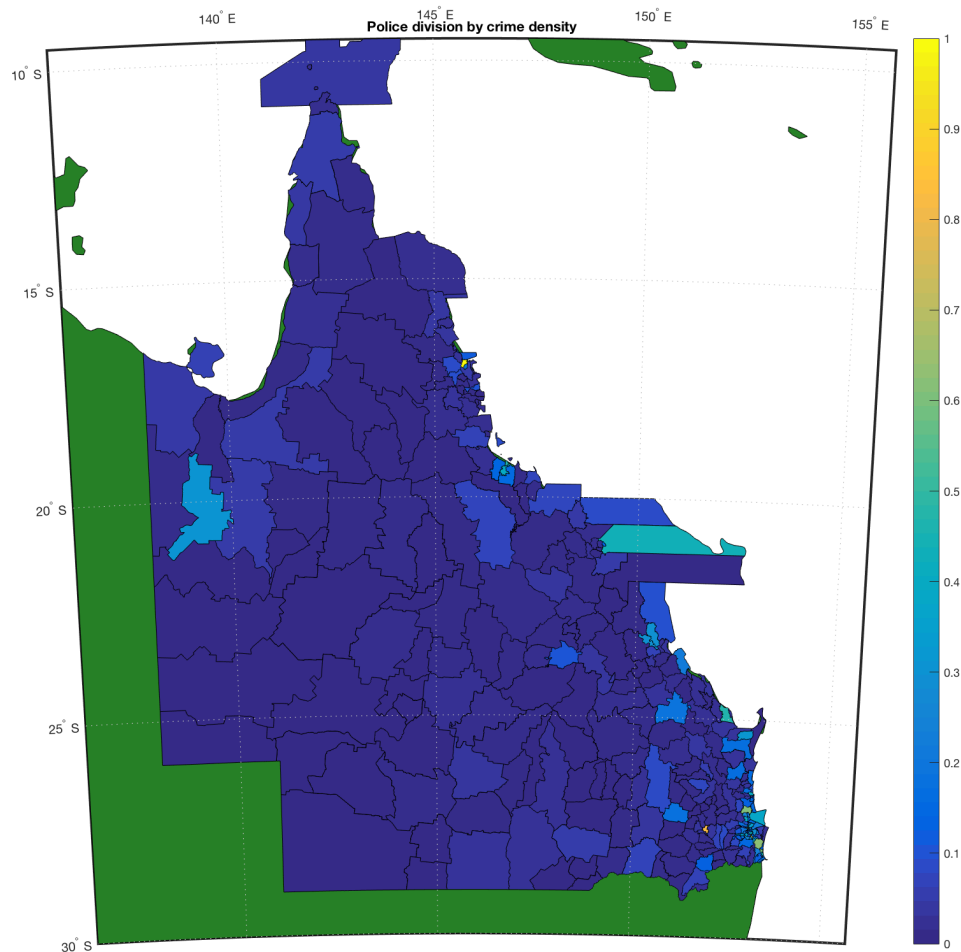


Figure 8 shows the density of crime in police divisions, the lighter the colour, the more crime happens in that division. Immediately it can be seen that there are a few brightly coloured points that stand out.

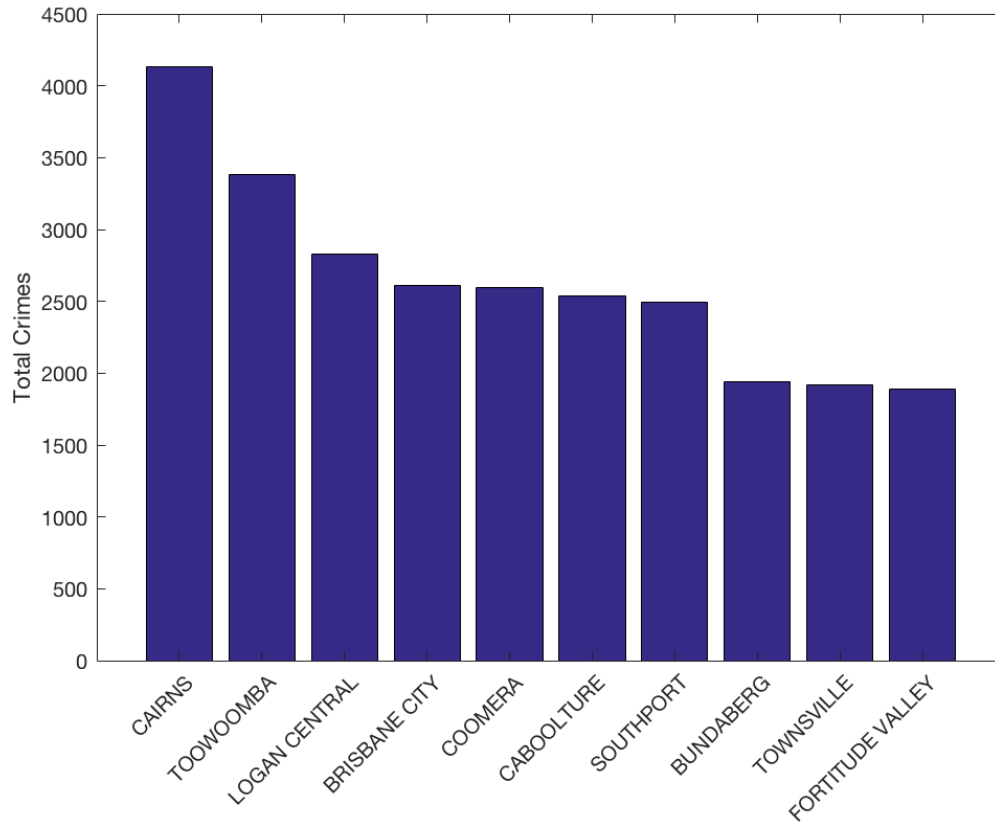
Furthermore by comparing figure 6 and figure 8, it can be seen that most of the police divisions with the largest land areas also has the lowest reported offences. There are two large rural police divisions that do have higher crime rates. One is Mount Isa, in the west, and the other is Mackay, in the east. This could be an indication that those two rural divisions could be understaffed.

The ten police divisions with the most crime during the month of February 2017 was found to be:

Table 1: Divisions with most crime in February 2017

Police District	Offences
Cairns	4131
Toowoomba	3383
Logan Central	2829
Brisbane City	2610
Coomera	2598
Caboolture	2538
Southport	2495
Bundaburg	1943
Townsville	1922
Fortitude Valley	1886

Figure 9: Top crime in February 2017

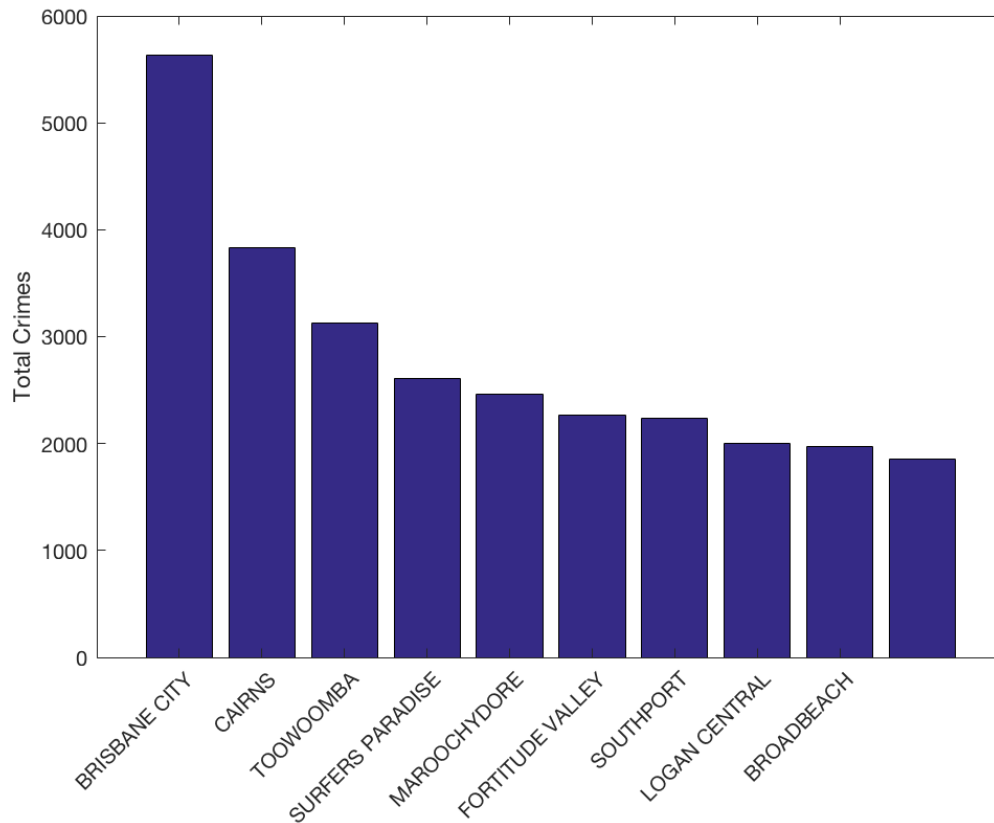


This can be compared to the divisions with the most crime in February 2001. By comparing with the same month of the year, it should avoid any error caused by fluctuations caused by crime happening in certain months more than others.

Table 2: Divisions with most crime in February 2001

Police District	Offences
Brisbane City	5636
Cairns	3833
Toowoomba	3127
Surfers Paradise	2613
Maroochydore	2466
Fortitude Valley	2266
Southport	2234
Logan Central	1999
Broadbeach	1973
Hendra	1851

Figure 10: Top crime in February 2001



The two tables show how crime rates change over time. One of the largest changes is the crime in Brisbane City, which decreased significantly over 16 years. Additionally Surfers Paradise and Maroochydore dropped off the list since 2001 and Coomera, Caboolture and Bundaburg had increases in crime over that period.

Figure 11: Cairns crime density

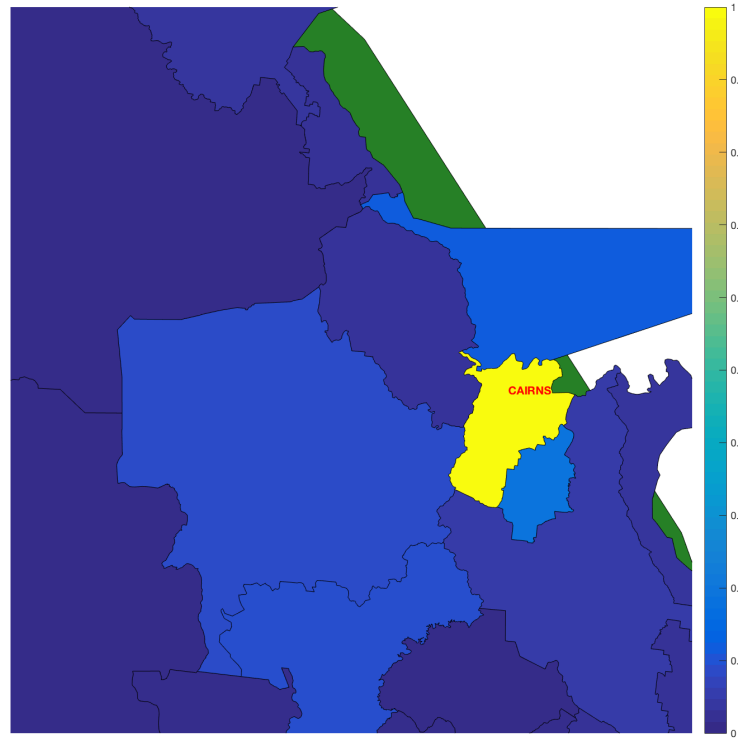
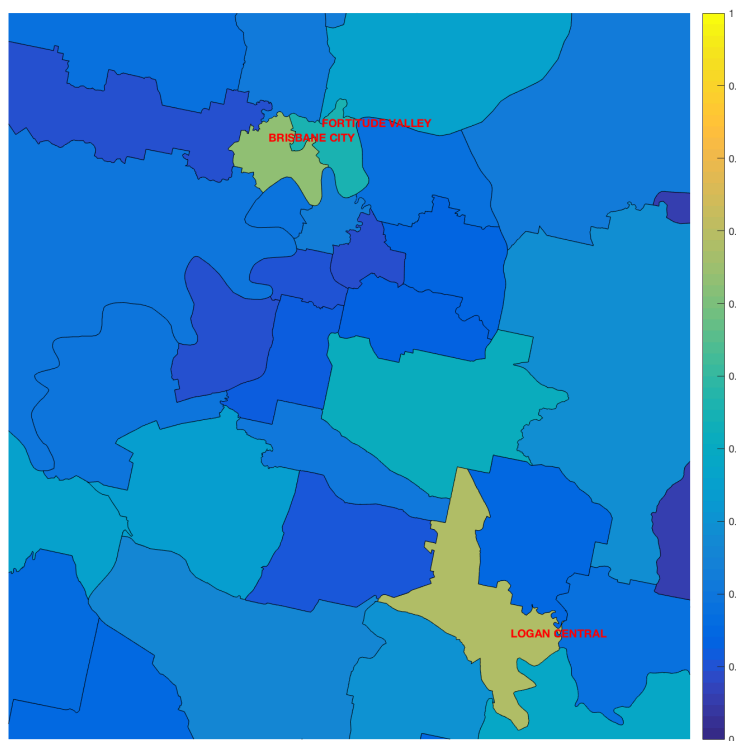


Figure 11 is interesting, as there is a large concentration of crime in one police division, yet the surrounding area has significantly lower crime rates. This differs from other areas, which has a much more gradual decrease in crime in surrounding areas, see figure 12. Also note the solid green indicates there is no police division for that area, thus the map either has errors or the area is a military base.

Figure 12: Brisbane crime density



The crime map for Brisbane shows that the areas with the most crime is Logan City, Brisbane City and Fortitude Valley. Additionally neighboring areas also have increased crime rates, which is expected.

5 Limitations

The accuracy of the crime data is limited to the police region. The New York government provides similar data, however they provide exact coordinates of where crime happens, which allows much finer analysis of where crimes happen.

Additionally finer grain time periods would allow better analysis of data, as it would allow for more accurate modelling of the data.

6 Conclusions

7 References

References

- [1] Offence numbers police divisions monthly from July 2001 - Data | Queensland Government. URL: <https://data.qld.gov.au/dataset/offence-numbers-police-divisions-monthly-from-july-2001>.
- [2] IBM SPSS Crime Prediction and Prevention, Mar. 2012. URL: <http://www.ibm.com/software/analytics/spss/11/na/cpp/>.
- [3] 2004 in Australia, Mar. 2017. Page Version ID: 768677697. URL: https://en.wikipedia.org/w/index.php?title=2004_in_Australia&oldid=768677697.
- [4] 201011 Queensland floods, Mar. 2017. Page Version ID: 772106382. URL: https://en.wikipedia.org/w/index.php?title=2010%E2%80%9311_Queensland_floods&oldid=772106382.
- [5] J. J. Nolan. Establishing the statistical relationship between population size and UCR crime rate: Its impact and implications. *Journal of Criminal Justice*, 32(6):547–555, Nov. 2004.
- [6] Palantir Technologies. Law Enforcement. URL: <https://palantir.com/solutions/law-enforcement/index.html>.
- [7] A. Remeikis. Queensland crime statistics not black and white. *Brisbane Times*, Aug. 2016. URL: <http://www.brisbanetimes.com.au/queensland/queensland-crime-statistics-not-black-and-white-20160807-gqn2g7.html>.
- [8] T. Roufa. The Early History of Crime and Criminology. URL: <https://www.thebalance.com/the-history-of-criminology-part-1-974579>.