The Relationship Between Unemployment and Rates of Specific Crimes

Rowan McEvoy

Domain

Employment and Communities

Question

How does the unemployment rate affect the rate of different crimes?

Criminologists and police departments would benefit from this information. Depending on the rise and fall of unemployment, these people could be put on special alert for certain crimes that would be more likely than normal to happen. Catching and prosecuting criminals would be easier if increases in certain types of crime could be anticipated in advance. Additionally, public officials could use this information to launch programs during periods of high unemployment designed to prevent certain crimes from happening. For example, if burglary increases with unemployment, public officials could put more money into neighborhood watch programs when unemployment starts to increase.

Datasets

The following data sets are used:

- *2015 Local Government Area Profiles* Excel
  Provided by the Department of Human and Health Services, this dataset gives extensive profiles of Local Government Areas (LGAs) including statistics on employment, housing, transportation, education, health, community engagement, and more. From this dataset, only unemployment rate and LGA name are used. It is important to note that this dataset DOES NOT include highly detailed information about crime.
  https://www.data.vic.gov.au/data/dataset/2015-local-government-area-profiles

- *Crime Statistics Agency Data Tables – Crime by location* Excel
  This dataset comes from the Crime Statistics Agency which publishes statistics on Victorian crime. Crimes are listed by LGA, offence division and group, location division and group, and investigation status. Statistics on number of incidents recorded and rate per 100,000 population are both given for each aggregation. From this dataset, only the LGA and offence aggregations are used. A screenshot of this dataset is shown below (**Figure 1**).
  https://www.data.vic.gov.au/data/dataset/crime-by-location-data-table

Year ending September	Local Government Area	Offence Division	Offence Subdivision	Offence Subgroup	Incidents Recorded	Rate per 100,000 population
2017	ALPINE	A Crimes against the person	A10 Homicide and related offences	A10 Homicide and related offences	2	16.9
2017	ALPINE	A Crimes against the person	A20 Assault and related offences	A211 FV Serious assault	9	76.2
2017	ALPINE	A Crimes against the person	A20 Assault and related offences	A212 Non-FV Serious assault	13	110.0
2017	ALPINE	A Crimes against the person	A20 Assault and related offences	A22 Assault police, emergency services or other authorised officer	2	16.9
2017	ALPINE	A Crimes against the person	A20 Assault and related offences	A231 FV Common assault	11	93.1
2017	ALPINE	A Crimes against the person	A20 Assault and related offences	A232 Non-FV Common assault	6	50.8
2017	ALPINE	A Crimes against the person	A30 Sexual offences	A30 Sexual offences	16	135.4
2017	ALPINE	A Crimes against the person	A70 Stalking, harassment and threatening behaviour	A711 FV Stalking	1	8.5
2017	ALPINE	A Crimes against the person	A70 Stalking, harassment and threatening behaviour	A731 FV Threatening behaviour	2	16.9
2017	ALPINE	A Crimes against the person	A70 Stalking, harassment and threatening behaviour	A732 Non-FV Threatening behaviour	4	33.8

Figure 1: Crime Statistics Agency Data Tables – Crime by location

Preprocessing and Integration

Preprocessing of both datasets was accomplished in a Jupyter Notebook using Python. Neither dataset had easily identifiable missing data (more on that later) but it was important to check for potential outliers, convert all data to a useable format, and eliminate unnecessary data. The following actions were taken during preprocessing and integration:

- All unemployment data taken from the *2015 Local Government Area Profiles* dataset came in string form and had to be converted to numeric data. Additionally, '%' characters were removed from each string before the conversion function could operate. While the unemployment data ranges in value from 2 to 12 as stored, we remember that it always represents a rate.

- Unemployment data was checked for outliers using a boxplot from **matplotlib**. Greater Dandenong, the one outlier, had an unemployment rate of 12% which was higher than that of any other LGA by more than 2%.

- Crime data proved more difficult to check for outliers given the number of features contained in the data. To provide an initial check, crime data was aggregated by LGA and normalized by summing over the crime rate per 100,000 population to produce a set of total crime rates per 100,000 population. Using a boxplot from **matplotlib**, Latrobe and Melbourne both registered as outliers, having unusually high crime rates. Melbourne specifically had almost twice the crime rate of the highest non-outlier and four times the median crime rate.

- Although the data for Greater Dandenong, Latrobe, and Melbourne is believed to be accurate, it was removed from both datasets before further analysis. While this means that the findings of this report cannot be applied to any of these LGAs, it ensures that the generated predictive models are not skewed by these extreme pieces of data.

- The crime dataset initially had two different metrics for crime: 'Incidents recorded' and 'Rate per 100,000 population'. 'Rate per 100,000 population' was chosen over 'Incidents recorded' to be used in this report because it normalized the data based on the population size of the respective LGA. The 'Incidents recorded' dimension was then discarded. For the rest of this report, any crime rate refers to a rate per 100,000 population.

- The crime dataset was filtered for data from 2015 and the 'Year' dimension was then discarded.

- The crime dataset was split by 'Offence division' and then 'Offence subdivision' for analyzing specific types of crime. These two dimensions along with 'Offence subgroup' were then discarded.

Results

At the 'Offence division' level the crime data splits into six categories:

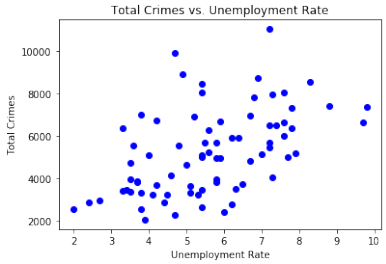
- 'Crimes against the person'
- 'Property and deception offences'
- 'Drug offences'
- 'Public order and security offences'
- 'Justice procedures offences'
- 'Other offences'

After the data was split along these lines, data in the 'Other offences' category was discarded. This category generally had only 1-10 total crimes per LGA and thus could not have provided any useful policy suggestion for larger crime reduction. The five remaining categories were separated by LGA, summed over offence

subdivision, and plotted against unemployment (**Figures 3-7)** using **matplotlib**. Total crimes by LGA was plotted against unemployment (**Figure 2**) using the same technique.

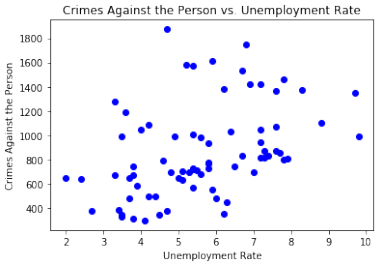
The corresponding Pearson correlation was calculated to accompany each graph using **numpy**. This identified which offence divisions were correlated with unemployment and which offence subdivisions should be further investigated for correlation. Out of the five offence divisions studied, 'Crimes against the person' and 'Property and deception offences' had the highest correlation with unemployment. Therefore, these kinds of crimes deserved further investigation by means of drilling down into offence subdivision.

Figure 2: Total Crimes



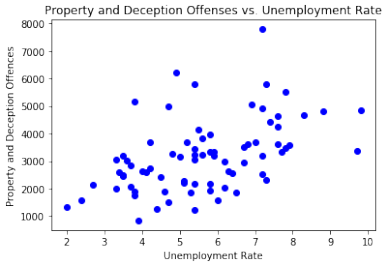
r = 0.51

Figure 3: Against the Person



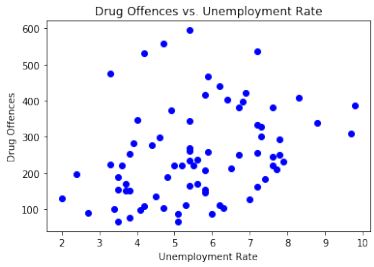
r = 0.41

Figure 4: Property/Deception



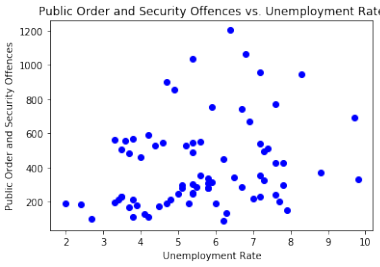
r = 0.48

Figure 5: Drug Offences



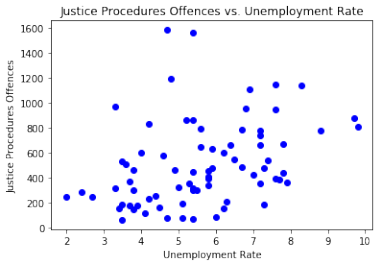
r = 0.31

Figure 6: Public Order/Security



r = 0.26

Figure 7: Justice Procedures



r = 0.35

At the offence subdivision levels of the 'Crimes against the person' and 'Property and deception offences' offence divisions, the crime data split into 14 total categories:

- 'Homicide'
- 'Robbery'
- 'Arson'
- 'Deception'
- 'Assault'
- 'Blackmail and extortion'
- 'Property damage'
- 'Bribery
- 'Sexual offences'
- 'Stalking'
- 'Burglary'
- 'Abduction'
- 'Negligent acts'
- 'Theft'

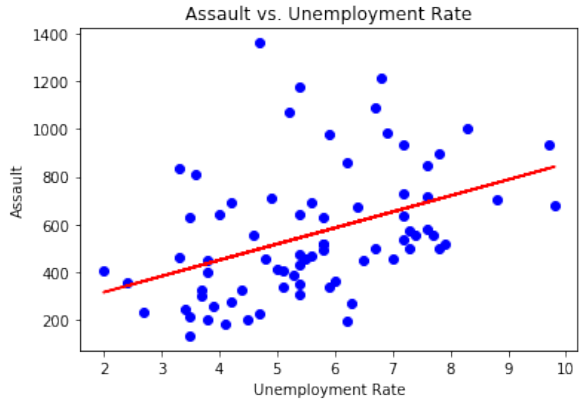
After the data was split along these lines, the categories were analyzed for widespread prevalence among LGAs. 'Homicide', 'Abduction', 'Robbery', 'Blackmail and extortion', and 'Bribery' averaged a rate of less than 15 per 100,000 population. Given the aim of this report to provide policy suggestions for large crime reduction, these categories were discarded.

A few of the remaining categories had missing values for some LGAs. Since the crime data used included all reported crimes in Victoria, this means that no crimes of some types were reported in some LGAs. Because of this, values of 0 were inserted where data was missing for an LGA within a specific crime category.

The Pearson correlation was calculating for the remaining nine offence subdivisions using **numpy**. The four that correlated the highest with unemployment are plotted below (**Figures 8-11**). Additionally, least-squares linear models were calculated using **numpy** and appear in each graph as well as below in equation form.

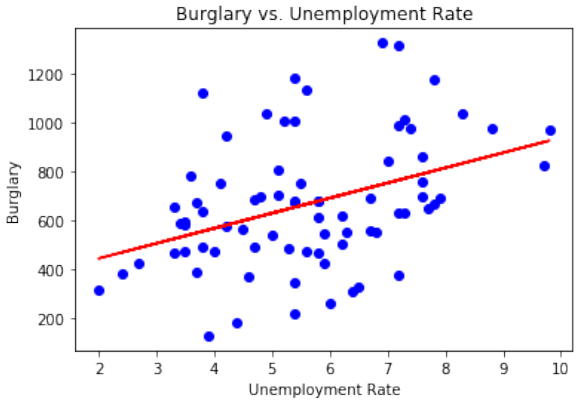
The linear models show that LGAs with an unemployment rate one point higher than others tend to have higher assault rates by 65.66 per 100,000 population and higher burglary rates by 61.70 per 100,000 population. They also have higher theft rates by 218.42 per 100,000 population and higher deception rates by 30.70 per 100,000 population. As a note, the intercepts of these linear models cannot be analyzed for meaning because doing would be extrapolation since there is no data for LGAs with unemployment rates below 2%.

Figure 8: Assault



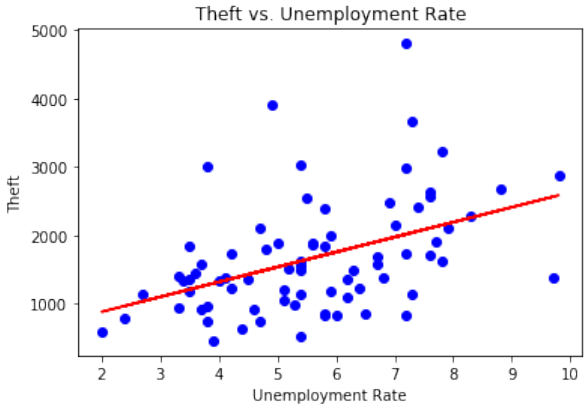
A = 65.66*U + 178.50 r = 0.42

Figure 9: Burglary



B = 61.70*U + 319.57 r = 0.39

Figure 10: Theft



T = 218.42*U + 445.25 r = 0.44

Figure 11: Deception

D = 30.70*U + 18.78 r = 0.41

Limitations

This method of analyzing the relationship between unemployment and the rates of specific crimes only uses data from one year. Further research could be done using data from multiple years but would require finding unemployment data to match the other available crime data from 2008-2014 and 2016-2017. This would provide more data to work with and ideally produce more accurate results.

Additionally, having data spanning a range of years would allow for times series analysis. Unemployment and the rate of specific crimes could be analyzed individually for each LGA. With enough data (20-30 years' worth), predictive models could be made specific to each LGA.

Finally, the current report completely excludes Greater Dandenong, Melbourne, and Latrobe from its scope based on outlier status of their total crime or unemployment data. While this has been done to improve the accuracy of the predictive models for the remaining LGAs, this report would ideally create models that could be used for all LGAs.

Value

While the unemployment data was usable for analysis after conversion from string to numeric form, the crime data set had too many dimensions to feasibly compare types of crimes rates against unemployment without processing and separating the data. Even after separating the data by offence division and then again by offence subdivision, missing values needed to be inserted before further analysis. On top of this, outliers needed to be removed along the way to prevent skewing the eventual linear models. Given these issues, preprocessing, integration, and analysis/visualization added significant value to the raw data.

Challenges and Reflections

The biggest challenge I faced involved figuring out which LGAs did not have any reported crimes for certain offence subdivisions. The complete dataset did not appear to have any missing values so I only figured out that some were missing after trying to create scatterplots, some successfully and some unsuccessfully, for different offence subdivisions against unemployment. After about an hour of trying to figure out my matplotlib error messages, I realized that some of my Series objects were not the correct size. I then had to manually compare each problematic offence subdivision Series against a known list of LGAs to figure out which LGAs were missing before inserting 0s.

Question Resolution

Assault, burglary, theft, and deception each exhibit a moderate correlation with unemployment. If the unemployment rate naturally changes, its correlation with assault, burglary, theft, and deception offers some insight into how these crime rates may change.

Criminologists and police departments would benefit from this information. Using the linear models created, these people could anticipate the change in assault, burglary, theft, and deception when unemployment fluctuates in each LGA.

Code

The nicely formatted version of my code that was submitted in a zip file contains around 200 lines of original code. Very small portions, for example the steps taken to run linear regressions, were loosely modelled off code from stack overflow but the rest was completed by referencing labs as well as through trial and error. In addition to this, another ~200 lines of original code were written for exploratory analysis with a bunch of other variables when deciding on a topic for this report. This code was not included in the zip file. All code was written in python using a Jupyter Notebook. Pandas, matplotlib, and numpy were the only libraries used and were all used extensively.