Identifying and Quantifying Strong Communities

Domain: Community

How do we identify strong communities and quantify them for better analysis of benefits and problem areas? This report seeks to identify through correlation, the factors that can be used to identify and quantify strong communities, allowing for focused analysis on its benefits in all areas including mental and physical health and crime. This is a useful question for local governments, who can use this information to target issues in their communities. Through promoting highly correlated activities, communities will benefit through greater community engagement, and reap the benefits.

By identifying features of and ranking communities, allowing for the quantifying of strong communities, further analysis into the effects of community on families and members of the community is available. Only through identification of and quantifying of strong communities can local government make changes to suit residents.

The following two data sets were used:

• LGA11 Community Strength - Torren University of Australia (csv)
Contains modelled estimates of community strengths for the year 2011. Data is
divided by Local Government areas, containing information about volunteering in
the last 12 months, receiving or giving support and percentage of people that feel
safe walking at night.

Link: https://data.aurin.org.au/dataset/tua-phidu-lga11-communitystrength-lga2011

• Local government Area (LGA) profiles data 2011 – Government of Victoria – Department of Health (csv)

Contains annually updated Local Government Profiles containing a broad range of data. Used a subset of the data set, analysing relevant family, behavioural and community data.

Link: https://data.aurin.org.au/dataset/vic-govt-doh-health-lga-profilesdata2011-lga

Through the processing of data, I was able to create a more legible code, changing the names of unclear columns, keeping consistency in column titles between datasets and sorting in terms of LGA codes. I made the datasets more python readable to allow for easier analysis, removing whitespace from column headings, changing numerical inputs to their required types. Unreliable data was removed from the data sets, avoiding all data with high uncertainty, and eliminated major outliers, allowing for more accurate calculation of correlation.

With just the raw data, it is difficult to understand or visualise which communities are strong. Through the figures below and creating a scale to measure community strength, correlations between communities and benefits such as better health become visible. Through processing an in particular visualisation, comparing and contrasting different LGAs is made much easier.

Initial transformations:

• Cleaned up the information so that it would be easier to process and visualise, removing whitespace from column headings, changing numerical inputs from string to double or integer types.

- Removed unreliable data as defined by the data sets by RRMSE, the relative square root mean square error, **removing** any unreliable data, (those with a value of 2).
- Restricted the datasets to only relevant data, prioritising data that was measured per 100 people and removing columns containing data not concerning the proposed question
- Increased clarity of column headers, renaming used columns in LGA11 Community Strength to human understandable names.
- Used LGA codes as a key, sorting sets in ascending order in terms of LGA codes for comparison of datasets.
- Removed rows with inconsistent LGAs, in particular unincorporated Vic, which only contained insufficient data and had an extremely small sample size.
- Keeping consistency in the column titles and allowing for some operations to be applied to both sets, renaming 'area_code' in community strength to 'lga_code'.

Identifying strong communities:

A way to identify strong communities is through looking at correlating factors. I looked at the Pearson values between 5 areas support, volunteering, membership groups, safety and parental involvement below (figure 1)

Support -> Received support in times of crisis
Volunteering -> Volunteered in the last 12 months
Membership -> % of people in membership groups
Safety -> % of people that feel safe walking alone in the local area at night
Parents -> % of parents that are involved in their local schools

Pearson coefficients - r values: (figure 1)

	Support	Volunteering	Membership	Safety	Parents
Support		0.839	0.688	0.809	0.260
Volunteering	0.839		0.69	0.925	0.225
Membership	0.688	0.69		0.580	0.369
Safety	0.809	0.925	0.580		0.209
Parent inv	0.260	0.225	0.369	0.209	

As seen, the Pearson values between support, volunteering and safety all have very large values of r > 0.8 (Jacob Cohen), which indicates an extremely high correlation between these subsets. There is also a high correlation between these sets and membership, having a large coefficient of r > 0.5. On the other hand, parent involvement seems to have only small correlation between 0.2 and 0.37. Although some correlation is present, it can be said that it is not a good indicators of a strong community.

Quantifying strong communities

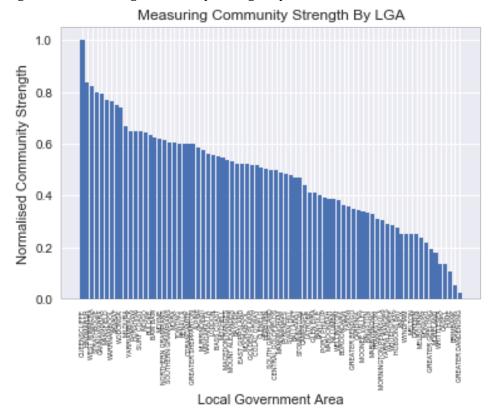
Assuming that support, volunteering and safety are good indicators of a strong society, we can rank different LGAs. To find a basis to rank communities, I normalised the values of support, volunteering, and membership for each LGA, then taking the average for each (figure 2). I chose to use normalisation as major outliers have been removed, thus the normalised data is not present only in a small range, which can occur.

figure(2) - strong communities based off of normalised data

$$\left(\sum_{i=0}^{3} \left(\frac{xi - min(x)}{max(x) - min(x)}\right)\right) \div 3$$

Visualising rank

figure 3 - Measuring community strength by LGA



Using a bar chart one is able to ranking the community strength of each LGA using the normalised values mentioned previously. This is useful for identifying the stronger communities by name, allowing for in depth analysis on problem areas in order to find problems, or on stronger communities.

How this will help:

Only when I had the normalised data was I able to find correlations between the strength of community and other factors, as seen in the next section.

Using the quantified strength to find the benefits of strong communities

By using scatter plot and line of best fit, one can see trends in other areas. This requires a scale to be compared against. This can be used to find correlations in areas such as physical and mental health, negative behaviours and crime. Below are some correlations including community strength and health, and

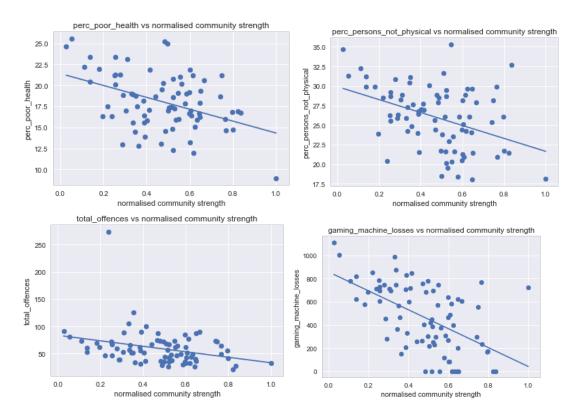


Figure 4, 5 – Gathering data of interest by comparing community strength against health problems. R values of -0.421 and -0.403 respectively, showing that community may have a substantial negative correlation to health.

Figure 6 – The medium negative correlation between number of offences (r=0.30). However, slightly skewed by the high crime rate in Melbourne.

Figure 7 – The mid-large negative correlation r=(0.48) in gambling losses, showing that there could be a link between community support and

The remainder of the project is feasible and likely to yield interesting results. As seen in figures 4-7, there are clear correlations between community strength and health, and community and negative behaviours such as gambling. As comparisons were only able to be made through quantifying community strength, the identification and quantifying of strong communities is important for further investigation into other areas. There may be many other correlations with community to reducing negative behaviour, or helping quality of life for residents through encouragement in factors.