Factors affecting the student attendance rates in Queensland schools

Group 3:
Matt Wang, Mohini Yadav, Natalie Manahan, Alan Seed

Introduction

Student attendance at school is an important part of a child's development into adulthood, and a predictor for achievement and also increased opportunities for economic advancement later in life. There are many reasons why a student may not attend school, ranging from individual preference for non-scholastic activities, an adverse situation at school (i.e., bulling), an adverse situation at home (i.e., relationship breakdowns, or learning at home not being encouraged), illness or mental health, disability or developmental delays, or cultural differences between home and school.

This project will use data published by the Queensland State Government on attendance rates for students enrolled in both primary and secondary school for the years 2015 to 2019, and offence rates for Local Government Areas together with Australian Bureau of Statistics data on relative Socio-Economic Index for Areas to explore some aspects of how crime rates and socio-economic factors affect student attendance rates.

The relative performance of Schools and the factors that affect school attendance rates has implications for individuals and society, and therefore is a subject of research by professionals in this field. This project is a very simple analysis that aims to show-case a range of data analysis techniques rather than to provide insights into a significant issue.

Data

1. The Queensland Department of Education maintains a record of attendance rates at the 1237 State Schools and publishes the data on the Queensland Government Open Data Portal at
   https://www.data.qld.gov.au/dataset/state-school-attendance-rate
   The published data are for the period 2015 – 2019 and include the school ID, post code and other meta-data together with the attendance rates for the 5-year period.
2. The location of each school is published in the Queensland School Directory, which is found at
   https://www.data.qld.gov.au/dataset/state-and-non-state-school-details
   The data for each school includes the latitude and longitude of the School.
3. The Queensland Police Service provide information of the rates of reported offences in Local Government Areas at monthly intervals. The rates are reported as per 100,000 people and (an impressive) range of crimes. The data are updated annually and are available at
   https://www.data.qld.gov.au/dataset/lga_reported_offences_rates

4. The Australian Bureau of Statistics published an index of Socio-Economic Indexes for Areas (SEIFA) after each Census, the data for 2016 can be found by Postal Area Code at https://stat.data.abs.gov.au/Index.aspx?DataSetCode=ABS_SEIFA_LGA
   We selected the Index of Relative Socio-economic Advantage and Disadvantage since we wanted to include the areas of relative advantage in the analysis. The data are reported as a relative measure, calculated for the residents in each post code and reported as percentiles of all post code areas in Queensland.

Summary Statistics

The first step was to download the Excel spread sheets from the Queensland Government Open Data Portal and to export the relevant workbooks as CSV files with the meta-data at the head of each file removed.

Student Attendance Rates

The school locations from the school directory were added to the attendance rate data table and the attendance rate was converted from a string to a floating-point number.

Figure 1 and Figure 2 show probability distributions and Box plots for student attendance rates from 2015 – 2019 respectively and Table 1 shows the summary statistics for each year.

It is immediately evident from the number of points below the Inter Quartile Range in Figure 2 that there are a significant number of schools with very low attendance rates relative to the bulk of the population. There is no reason to believe that these data are mis-reported and therefore should be removed from the analysis, but it does indicate that there is a set of schools that have a very different probability distribution of attendance rates from the rest of the population of schools in Queensland.

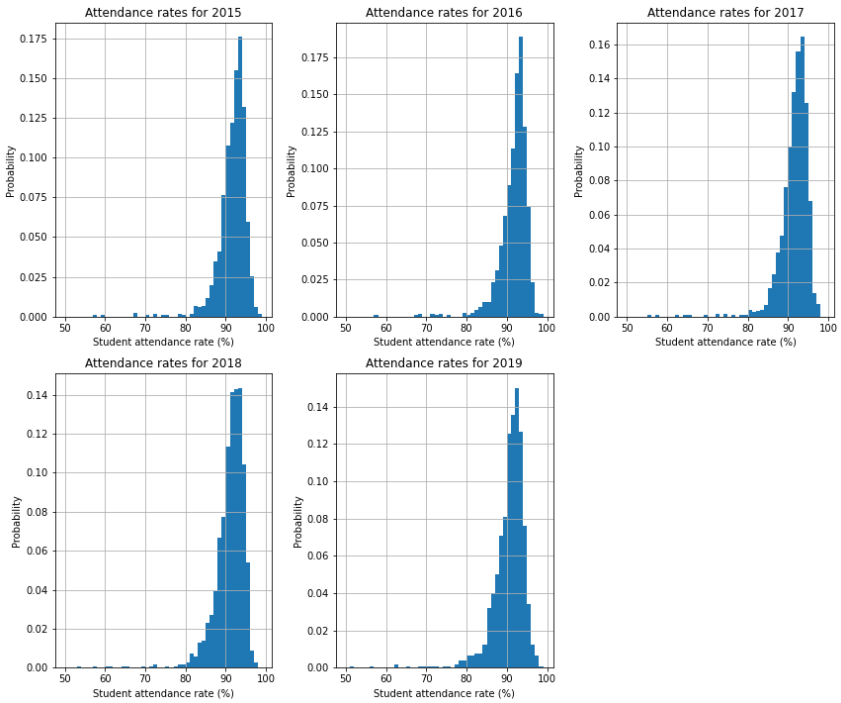


Figure 1 Probability distribution of student attendance rates for the years 2015 – 2019.

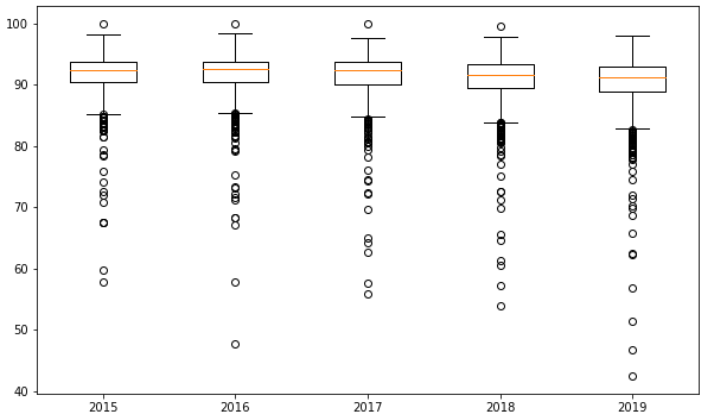


Figure 2 Box plots of student attendance rates for the years 2015 – 2019.

Table 1 Summary statistics for attendance rates for 2015-2019.

	min	25%	median	75%	mean	std	max
2015	57.90	90.35	92.30	93.80	91.67	3.53	100.00
2016	47.80	90.40	92.50	93.80	91.71	3.69	100.00
2017	55.90	90.10	92.30	93.70	91.52	3.69	100.00
2018	53.90	89.50	91.65	93.30	90.94	3.95	99.50
2019	42.50	88.90	91.30	93.00	90.44	4.48	98.00

Rates of reported offices

The LGA reported offences rates data provide a monthly rate for a wide range of reported offences for the past 20 years. The data are in rates per 100,000 of population, which is important since we are interested in comparing the crime rates between Local Government Areas that have widely different populations. The rates of the reported offences were summed each month to a total rate of reported crime per month and then averaged over the year. The histogram of the reported crime rates for 2019 is shown in Figure 3.

The maximum crime rate is for Woorabinda Shire Council with 37459 offences per 100,000 (!). This Local Government Area had the highest rate of reported crime consistently through the 20-year record and the datum is not considered to be a reporting error. The Shire had a population of 962 people in the 2016 census and has a total of around 200 students in the two schools and is considered to be amongst Queensland's most disadvantaged communities.

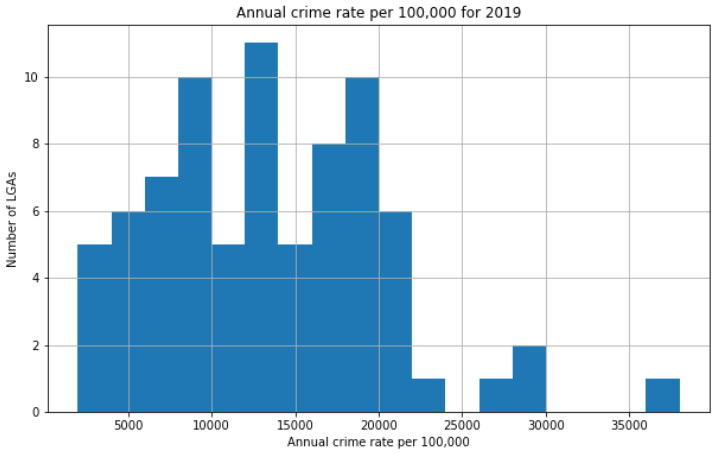


Figure 3 Histogram of frequency of total offences rate per Local Government Area for 2019.

Student attendance vs school type

The Queensland Department of Education classifies schools into 6 types: Community, Special Purpose, Special, Specific Purpose, State High, and State Schools. Community Schools are those that are run by the local communities in remote and very remote areas.
Figure 4 shows the attendance rate per school type and year.
Following are the observations:

1. The attendance rate for Community schools has been consistently lower in comparison to other School types.
2. The attendance rate for the special school type had remained almost constant throughout the years.
3. The attendance rate for the State High School and Sate school was also the same for all the years.
4. There is no data available for Specific Purpose School for the years 2018-19
5. Data for special purpose school is only available for 2019

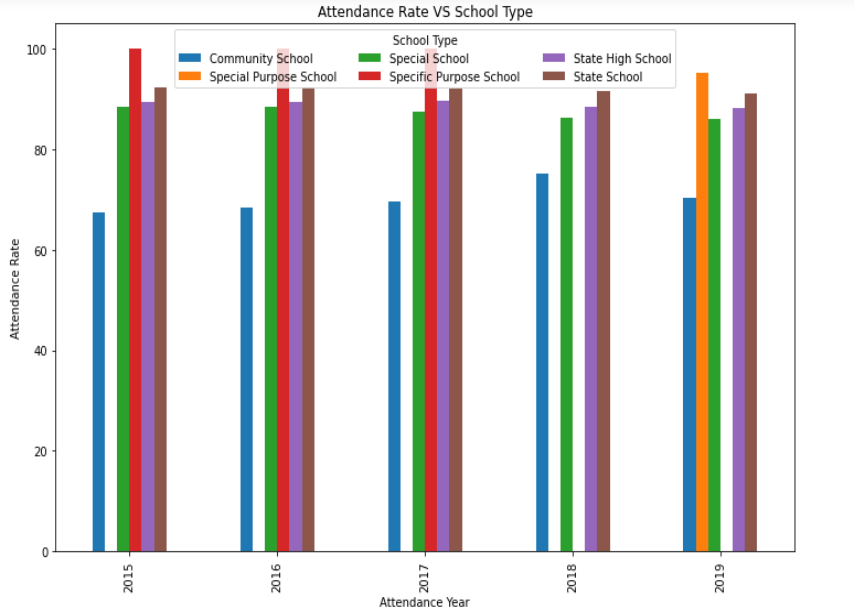


Figure 4 Attendance rates per school type and year.

Student attendance vs location

To understand the school location, we need to see an overview distribution of the schools within Queensland. With the aid from Google maps API, the heat map of student attendance rate per school is displayed as below for a typical year in 2019.

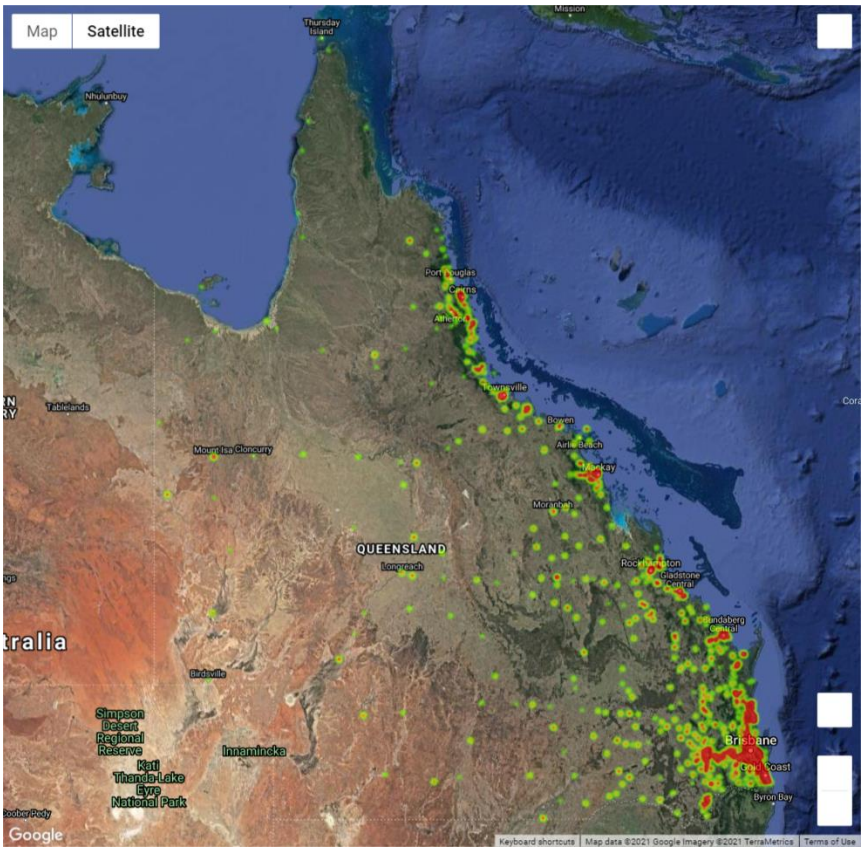


Figure 5 Heat map of student attendance rate for all schools in Queensland.

It is obvious that more schools and higher attendance rates are recorded along the East to South-east coastline of Queensland in the major urban areas. The reduction in attendance rate as the remoteness of the locations increases is noticeable.

Under this part of report, it is mainly focused on 3 types of location to compare and study at if there is any corresponded relationship to the student attendance rate.

Student Attendance Rate vs Remoteness Area

In Figure 6 shown below there are 5 different Remoteness areas in the data namely 'Inner Regional Australia', '`Major Cities of Australia`', 'Outer Regional Australia', 'Remote Australia' and 'Very Remote Australia'.

According to the plotted data, attendance rate in Very Remote Australia has been the lowest when compared to other remoteness areas and also has been constantly decreasing when compared to previous years.

Attendance Rate in other Remoteness areas other than Very Remote Australia has been pretty much consistent over the years.

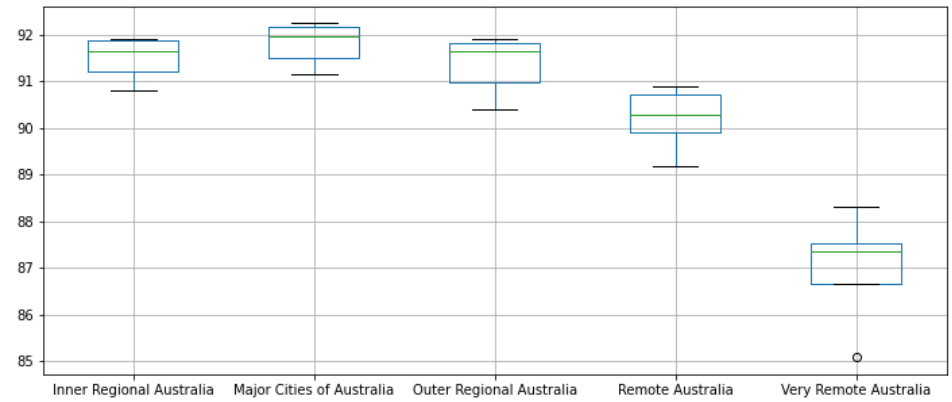
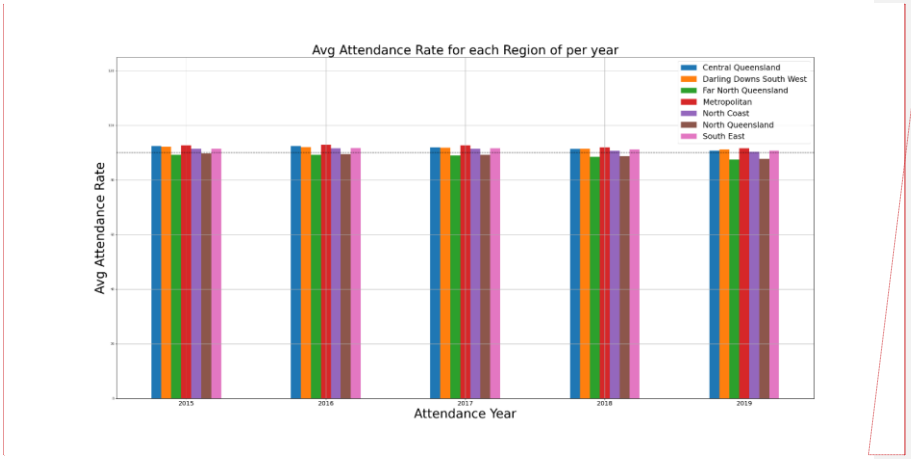


Figure 6 Box plots of the attendance rates conditioned on the Region.

Student Attendance Rate vs Region

There are 7 different Region areas in the data viz. 'Central Queensland', 'Darling Downs South West', 'Far North Queensland', 'Metropolitan', 'North Coast', 'North Queensland' and 'South East'. For all the years from 2015 to 2019, 'Metropolitan' has the greatest number of schools (it has by far the largest population) and 'Far North Queensland' has the least. The variation in the attendance rates for each year is shown in Figure 7.



Commented [AS1]: Matt, could you make the legend a bit larger?

Figure 7 Bar charts of student attendance rate for each region area each year.

It is clear to observe that the percentage does not change much for a typical region over the years. It shows that there is no direct relation between the attendance rate against different region over the 5 years or alternatively, the region area does not affect the attendance change rate much.

Figure 8 shows a set of pie charts with the ratio of number of attendance rate counts (I.e., the number of schools) providing another view demonstrate that there is not difference in the number of schools within a particular region area.

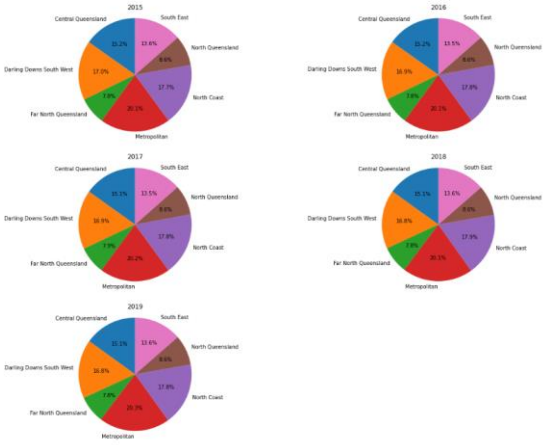


Figure 8 Pie charts of student attendance rate for each region area each year.

Student Attendance Rate vs Local Government Area

Local Government Area, LGA, is the next tier below the Regional Areas and covers smaller areas. There are fewer schools per LGA and show a more detailed geographical pattern of student attendance rate than the regional data. Within Queensland, there are 4 types of different LGAs, which are listed as 'Queensland: Cities (C), Shires (S), Towns (T) and Regional Councils (R)'. Within the dataset, 'Brisbane (C)' LGA has the highest school attendance rates recorded and the 'Hope Vale (S)' LGA has the lowest. There is a total of 74 LGAs in the data, but as 'Hope Vale (S)' only has valid data for one year, it is removed for below analysis.

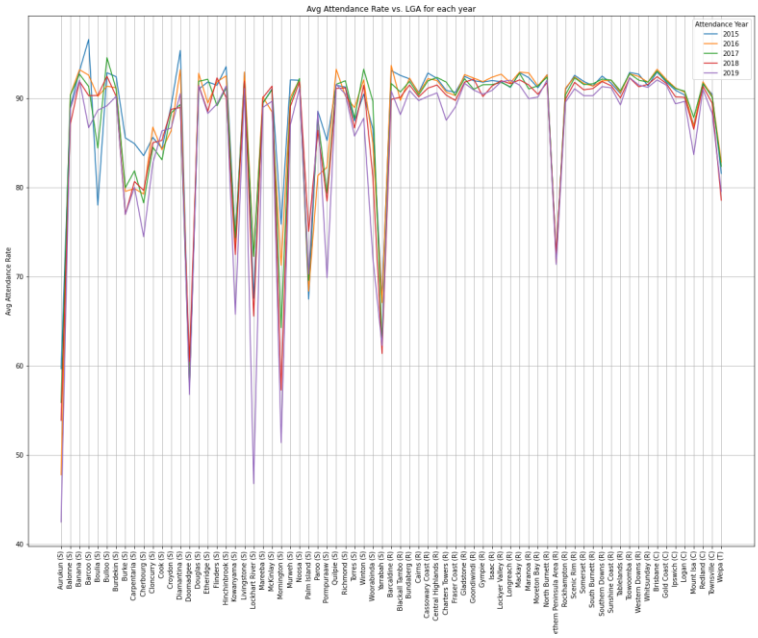


Figure 9 Line chart of student attendance rate for each LGA area each year.

This is the changing pattern for the attendance rates against a particular LGA over the years. It shows that most of the LGA do not have big changes of attendance rate over the 5 years excepts LGA such as 'Aurukun (S)', 'Barcoo (S)', 'Cherbourg (S)', 'Kowanyama (S)', 'Lockhart River (S)', 'Momington (S)' and 'Pormpuraaw (S)' have a drop more than 5%. There is only one LGA, the 'Boulia (S)' has a large increase (more than 5%) in attendance rate.

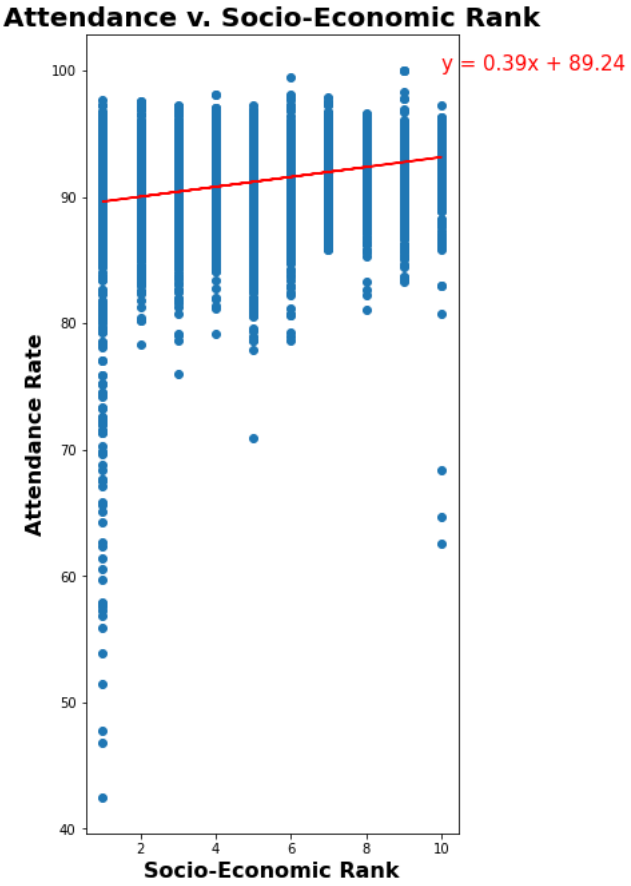
It is quite interesting to see that only few of the LGA have different pattern than the others and it is worth to study at these LGA individually. Example studies such as attendance rate change per school within these LGA and population change, the change of number of schools within each LGA etc.

Student attendance vs socio-economic score

The SEFIA (IRSAD) provides a general socio-economic index that summarises the economic and social attributes of people and households. A low score (1) could indicate many households with low income, educational attainment and low skill occupations, and a high score (10) could indicate households with higher income, educational attainment and high skill occupations. This SEIFA data set was used to explore any correlation between the socio-economic rank of an area and the attendance rate for schools of the same area. The results of the scatter plot and regression analysis (Figure 10) shows that schools in areas ranked 2 to 10 all had similar attendance rates ranging between 80 and 100 percent. There were however a few schools in areas ranked 10 with an attendance rate below 70 percent. For

schools located in areas ranked 1 on the SEFIA scale, the attendance ranged from just above 40 and 100 percent.

The low p – value (4.8386 E-113) indicates that there is no relationship between the SEFIA socio-economic rank of an area and the attendance rate of students at the schools in that same area, suggesting there are other factors are influencing attendance rate.



The p-value	**4.838562723842561 E-113**

Figure 10 Attendance rate vs. deciles of the socio-economic indicator of advantage and disadvantage index.

Further exploration of the data for schools in the area with a SEFIA rank of 1, show that most of these schools were located in north coast and far north coast region of Queensland (Figure 11), and that they were inner regional and out regional areas (Figure 12). As is explained in the next section, these areas are very remote and are often impacted by extreme weather events and conditions which do not allow for easy access to school.

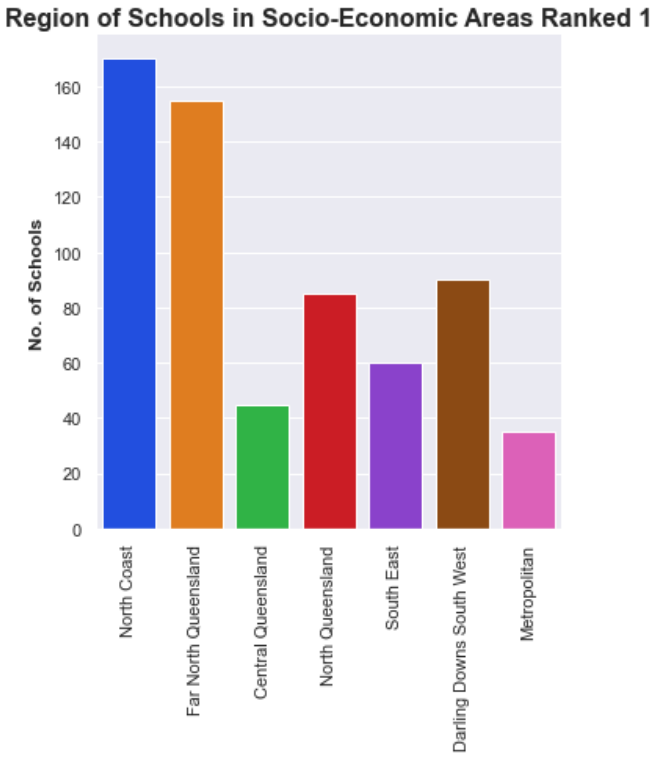


Figure 11 Number of schools in the first decile of the socio-economic index of advantage and disadvantage per region.

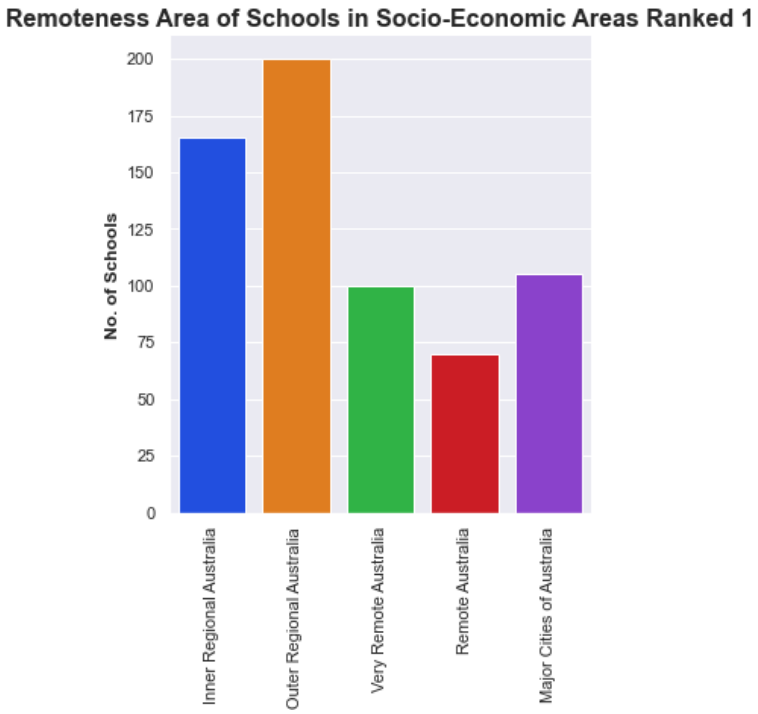


Figure 12 Number of schools in the first decile of the socio-economic index of advantage and disadvantage per remoteness area.

Student attendance vs reported offences rate

The relationship between attendance and rate of reporting offences was first evaluated using a scatter plot, as shown in Figure 13. While the Shire with the highest reported offence rate had an attendance rate of 72% in 2019, it has twice the crime rate as the school with the lowest attendance rate of 42.5% in 2019 in Aurukun Shire. The Aurukun Shire is on the west coast of the Cape York Peninsula. The location is extremely remote, and the roads become impassable during the monsoon wet season, which would make attending school during the wet season more difficult.

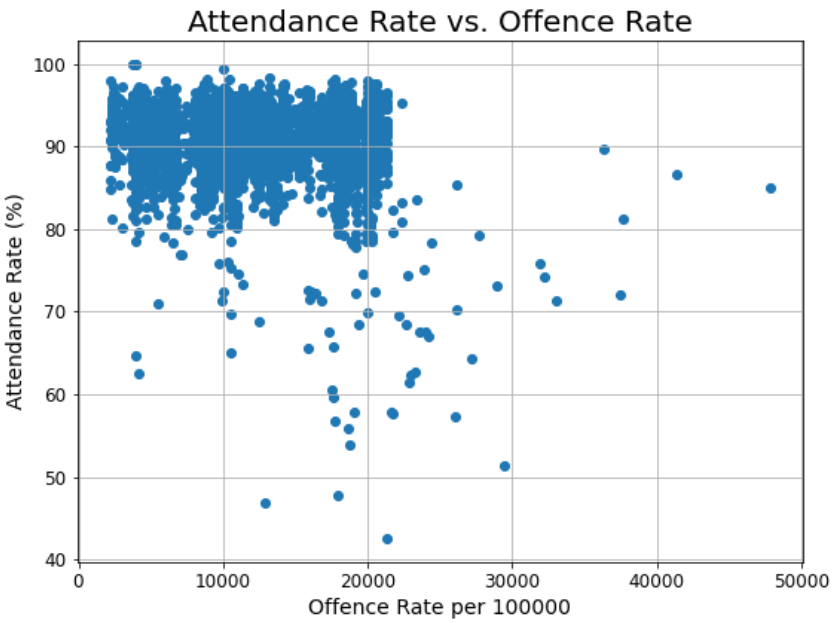


Figure 13 Scatter plot of Offence Rate vs. Attendance rate for Queensland schools between 2015 and 2019.

The probability distribution of attendance rate conditioned on the quartiles of the offences rates is shown in Figure 14. There is some evidence that the probability distribution for the first quartile (the least reported crime) also has the highest frequency of schools with a very high attendance rate, and that the mode of the distributions for the third and fourth quartiles is lower than that for the first and second quartiles. The left tail of the distribution for the fourth quartile is noticeably heavier that the others. Table 2 shows the summary statistics for the four conditional probabilities.

Table 2 Summary statistics for the probability distribution of attendance rate conditioned on quartiles of the reported offences.

	Q1	Q2	Q3	Q4
number	1554.00	1501.00	1519.00	1525.00
min	62.60	65.10	46.80	42.50
max	100.00	99.50	98.30	98.10
mean	92.20	91.04	91.22	90.54
variance	9.89	10.85	10.28	28.91
skewness	-2.04	-1.74	-3.47	-3.65
kurtosis	10.59	6.48	32.29	19.53

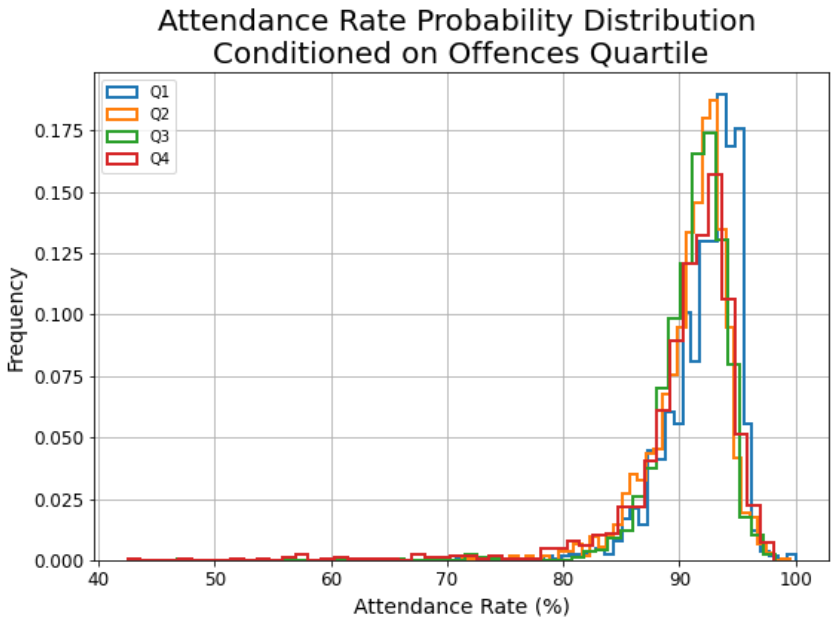


Figure 14 Probability distribution of attendance rates conditioned on the quartiles of the offence rates.

Conclusions

The first stage in the analysis was to plot the probability distributions of the student attendance rates over all schools in Queensland for each year in the attendance rate data and calculate the summary statistics for each year. The probability distributions have a heavy tail to the left with relatively more schools with a low attendance rate than would be expected from a Normal Distribution. This was the first indication that the school attendance rates are drawn from a mixture of at least two probability distributions, one of which has a significant number of schools with very low attendance rates.

The histogram of reported offences rates per Local Government Area (LGA) (Figure 3) shows a similar mixed distribution with a heavy tail to the right with some LGAs showing very high rates of reported offences. The maximum crime rate is for Woorabinda Shire Council with 37459 offences per 100,000 people. The Shire is in a remote area and had a population of 962 people at the 2016 census. It is also considered to be amongst Queensland's most disadvantaged communities.

The remote areas have schools that are run by the local community, and the same pattern of low attendance at community schools can be seen in Figure 4. The trend in a reduction in school attendance rates for remote areas can be seen in Figure 5 where the number of schools in the fourth quartile of attendance rate decreases away from the main city (Brisbane) and the developed areas along the coast.

The index of socio-economic advantage and disadvantage shows a similar pattern to the offences rates and school attendance rates. The highest number of schools in decile 1 (most disadvantaged) of the socio-economic index are in Outer-Regional areas, Figure 12.

It is evident that the probability distributions of attendance that are conditioned on the reported crime rate were not identical, Figure 14. The Kolmogorov-Smirnoff test was used to determine if the distribution of the first decile was identical to the other three and the Null Hypothesis was rejected at very low p-values. The probability of recording a very high attendance rate is increased for the schools in the areas with the most advantage (decile 1), but this does not show up in the mean and variance of the distributions, Table 2.

The relationship between the rate of reported offences and attendance is not straightforward. For example, in Figure 13 the Shire with the highest reported offence rate had an attendance rate of 72%. It had twice the crime rate of the school with the lowest attendance rate, 42.5% in the Aurukun Shire. The Aurukun Shire is on the west coast of the Cape York Peninsula. The location is extremely remote and the roads become impassable during the monsoon wet season, which would make attending school during the wet season more difficult.

The remote areas of Queensland have communities with relatively high reported crime rates and socio-economic disadvantage and are also the communities where the school attendance rate is generally lower. There are several reasons why these communities have low attendance rates that are outside this study. For example, the Monsoon wet season making roads impassable, the distance to school, and other cultural reasons for not attending school.

The signal in the data from the schools in the remote areas may have drowned out any more subtle relationships between offence rates, socio-economic conditions and school attendance in the city schools.