

The wheels on the bus: Where do we go to school and why?

Part A: Data analysis

Introduction:

“Which school did you go to?” As a North Islander, I find it strange that this question is so popular in Christchurch. Why are people judged by their alma mater? What is so meaningful about a school’s location or demographics? In this project I aim to investigate the relationships between various characteristics of schools and their neighbourhoods in the North Island, in order to see whether a similar bias (conscious or unconscious) exists on my home island.

In any debate about which school is best, it is inevitable that deciles will be mentioned, despite the Ministry of Education (MoE) cautioning that they do not “indicate the overall socio-economic mix of the school or reflect the quality of education the school provides” (n.d.A). Rather, a school’s decile is calculated by comparing the proportion of its students coming from meshblocks with low household income, low-skilled workers, crowded households, lack of qualifications, and income support (MoE, n.d.A).

It is worth restating that deciles are calculated according to the location of a school’s students, not the location of the school. This is important because these locations can be very different. For example, University of Canterbury researcher Andrew Devonport found that Christchurch secondary school students would save travelling a collective 71,000km per school day if they attended their closest school (Gates, 2017). Interestingly, the average distance a student travels to a state school was 6.3km, compared to 9.3km and 9.8km for state-integrated and private schools respectively (Gates, 2017). This suggests that parents (and students) are willing to commute longer for what they believe is a better education.

It is clear that which school you (or your child) goes to is considered a very important decision, one worth making substantial sacrifices for. I have already discussed some of the variables that influence this decision: decile and authority (whether a school is state, state-integrated or private). In my project I hope to gather data on these and other relevant variables such as deprivation, ethnicity, and co-educational status. I believe my findings on how these variables interact will be useful for parents, schools and policy-makers, as they will reveal biases parents may hold when choosing a school, whether those biases are well-founded, and how they could be addressed.

Data and methods:

Equipment and materials:

I used the R programming language through the integrated development environment RStudio to manipulate and display my data. For Part A of the project, I used the tidyverse, sf and GGally packages.

My data is from the 2018 Deprivation Index (University of Otago, n.d.), 2018 Census (Stats NZ, 2020), and MoE School Directory (MoE, 2021). I chose the Deprivation Index as it is a key part of this topic, and I wanted to see just how strongly it correlated with decile. As for the census data, I chose to use the individual (part 2) dataset displayed by Statistical Area (SA)1, as it included education-related variables. I also used the SA2 clipped dataset (Stats NZ, 2018), as I wanted to aggregate my data by SA2s. This is because SA2s were made to “reflect communities that interact together socially and economically”, with “shared community facilities” and “socio-economic similarity” (Stats NZ, 2017). I felt that SA1s were too small, while my third areal option, enrolment zones, was not available in a very high quality nor would it have been as easy to aggregate the deprivation and census data by this areal unit. Finally, the School Directory provided various variables and the coordinates of all schools in New Zealand. I would have liked to include a variable that measured academic performance, such as National Certificate of Educational Achievement (NCEA) pass rates per school, however this data is not publicly available; therefore I used the highest qualification variable from the census dataset, which was less than ideal as it could only be attributed to SA1/SA2s instead of individual schools.

SA2 data (sa22018):

I began by filtering the SA2 data to the North Island. This was straightforward as SA2s are ordered by island, with all those with a code less than 300000 being in the North Island (Stats NZ, 2017). I then transformed its coordinate reference system (CRS) to Web Mercator, since I planned to create interactive maps in Part B whose basemap would most likely use this projection.

School Directory data (schools):

Moving on to the School Directory data, I removed columns that I did not need, leaving some that I thought might be interesting to analyse, including the isolation index. This value determines how much targeted funding each school receives, with more isolated schools receiving more (MoE, n.d.B). It is calculated according to the school's distance from the nearest settlements with populations of 5000 or more, 20,000 or more, and 100,000 or more (MoE, n.d.B). I removed entries without geometry, as I would not be able to plot them. Several columns included entries along the lines of “Not applicable” and “Not calculated”, so I changed these to “0”. One such column was the isolation index column, so I then had to convert it from character to numeric so that I could analyse it better later on. In the decile column, some schools were listed as decile “99”, which would have skewed my results. Decile 99 schools are those which do not have a decile, whether that is because they are new, correspondence schools, or “private schools which opt not to be a part of the decile system” (Dyslexia Foundation, n.d.). Because there were several of these schools in my data, I wanted to recategorize them instead of just removing them. Since most of these schools are private ones, and most private schools are high-decile, I decided to change “99” to “10”. One school had its own category in the co-educational status variable, as it was boys-only for junior students and co-educational for senior students. I chose to recategorize it as co-educational, as I had decided to focus on secondary schools and NCEA results, therefore the senior students were most relevant. For this reason I also filtered out schools which ended at Year 10 (as well as proposed schools which do not yet exist) and narrowed down the entries to NCEA-age schools using the organisation type column. I calculated percentage columns for European and Māori students to make comparison between schools easier and fairer – I could have done this for the other ethnicity columns, but I chose to focus on these two as I would already be working with plenty of variables. Finally, I converted schools to a spatial features (sf) object (schools_sf), converted the CRS to Web Mercator, and used sa22018 to filter schools to the North Island.

Census data (census), deprivation data (nzdep2018), and combination of both (dep_census):

As for the census data, I removed unneeded columns, changed -999-suppressed data to "0", and removed entries with empty geometries. All I had to do for the deprivation data was remove NAs. I then joined these two by first changing column names and types to match, then left-joining `nzdep2018` to `census` by SA1 to create `dep_census`.

Combination of `dep_census` and `sa22018` (`sa2_dep_census`):

The next step was to join `dep_census` to `sa22018`. I did this by selecting relevant columns in `dep_census` and grouping by SA2, then summarising for the median deprivation score and sums of usually-resident population and each highest qualification column. I then calculated percentage columns for these qualification columns so that they would be easier and fairer to compare. Finally, I right-joined this to `sa22018` by SA2, removed unneeded columns, grouped by SA2, removed duplicates, and converted to sf (`sa2_dep_census_sf`).

Combination of `schools_sf` and `sa2_dep_census_sf` (`schools_data`):

To create `schools_data`, I spatially joined `schools_sf` to `sa2_dep_census_sf` so that each school was linked to the data for its SA2.

ESDA base data (`allcorr`):

However, for the exploratory spatial data analysis (ESDA) I wished to carry out, some more tweaks needed to be made. I removed geometry from `schools_data`, narrowed down columns to those I deemed most interesting or useful (as I still had too many variables to analyse at once), and finally removed NAs to create `allcorr`.

Statistical techniques:

To analyse the seven numeric variables I chose to focus on, I created a correlation matrix including correlation values, density plots, and scatter plots with regression lines and transparent points.

As for the two categorical variables, I created a set of box plots for each showing their relationship with the seven numeric variables.

Results and discussion:

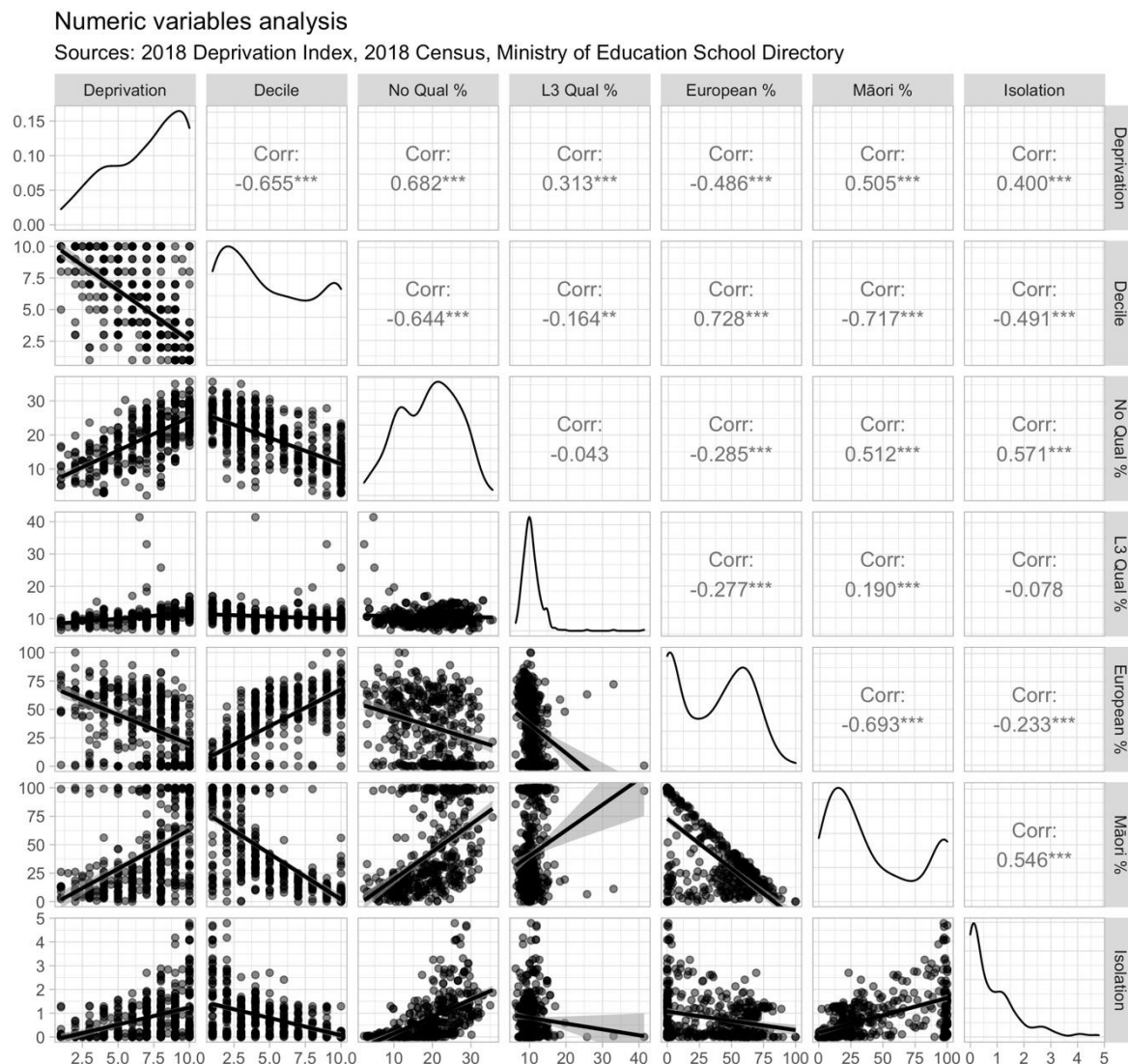


Figure 1: Numerical variables analysis

As shown by the correlation value of -0.655, there is a strong negative relationship between decile and deprivation, as expected. The high density of schools in highly-deprived areas (scores between 7.5 and 10) is worth noting. This suggests that if some parents are trying to avoid sending their children to schools in deprived areas, they are greatly narrowing down their options.

The correlation value of 0.682 shows there is a strong positive relationship between deprivation and the percentage of the population without a qualification. This makes sense as lack of qualifications is one of the variables factored into the Deprivation Index (Atkinson et al., 2019). The correlation value of -0.644 between decile and lack of qualifications is similarly justified due to the percentage of unqualified people per meshblock being part of decile calculations (MoE, n.d.A). However, the correlation value of 0.313 between deprivation and Level 3 qualifications contradicts this, as does the correlation value of -0.164 for Level 3 qualifications versus decile. This is potentially important, as it would debunk the mindset of parents who send their children to higher-decile schools in less-deprived areas because they think they will achieve better there. Despite this, I must keep in mind the outliers visible on the scatterplot which affect the correlation value, as well as what the variables actually mean. Since the Level 3 qualification variable refers to the entire population of an area aged 15 or older (Stats NZ, n.d.), it is only

useful as a reference of how qualified that area is, not how effective any schools within that area are at qualifying students.

With a correlation value of -0.486 between the percentage of European students and deprivation compared to 0.505 for Māori students, there is clearly some segregation occurring here. This is compounded by the correlation values of 0.728 and -0.717 for European and Māori students respectively in relation to decile. These figures suggest European students are less likely to attend schools in deprived areas and/or low-decile schools, which is a phenomenon dubbed “white flight” (Boyack, 2019). It is worth noting the density plots for the European and Māori student percentages, as they have two peaks – one at the expected point which is equivalent to that ethnicity’s percentage of the overall population, and another which is near 0% for European students and near 100% for Māori. This can be attributed to the existence of kura Kaupapa Māori, which are schools that revolve around the Māori language and culture (Donnelly, 1998). To drive the point home, the correlation value between percentage of European and Māori students is -0.693, which brings into question the idea of equality and hints at the existence of racism in school selection, be it conscious or unconscious.

There is a correlation value of -0.285 between lack of qualifications and European students compared to 0.512 for Māori students, meaning Māori students are more likely to attend schools in areas with more unqualified people. This meets the stereotypes, however it is contradicted by the correlation values for Level 3 qualifications -0.277 for European students and 0.190 for Māori students. Once again, the meaning of the qualification variables and the effect of outliers must be acknowledged.

That leaves the isolation variable, which follows similar trends. The correlation value between isolation and deprivation is 0.400 while between isolation and decile it is -0.491, meaning isolated areas are more likely to be deprived and more likely to be home to lower-decile schools. Isolated populations are more likely to be unqualified, with a correlation value of 0.571 compared to -0.078 for Level 3 qualifications. Finally, European students are less likely to attend isolated schools (correlation value of -0.233) compared to Māori students (0.546). The sharp peak between 0 and 0.5 on the density plot shows that the majority of schools are in accessible urban areas.

Categorical variables analysis: Authority

Sources: 2018 Deprivation Index, 2018 Census, Ministry of Education School Directory

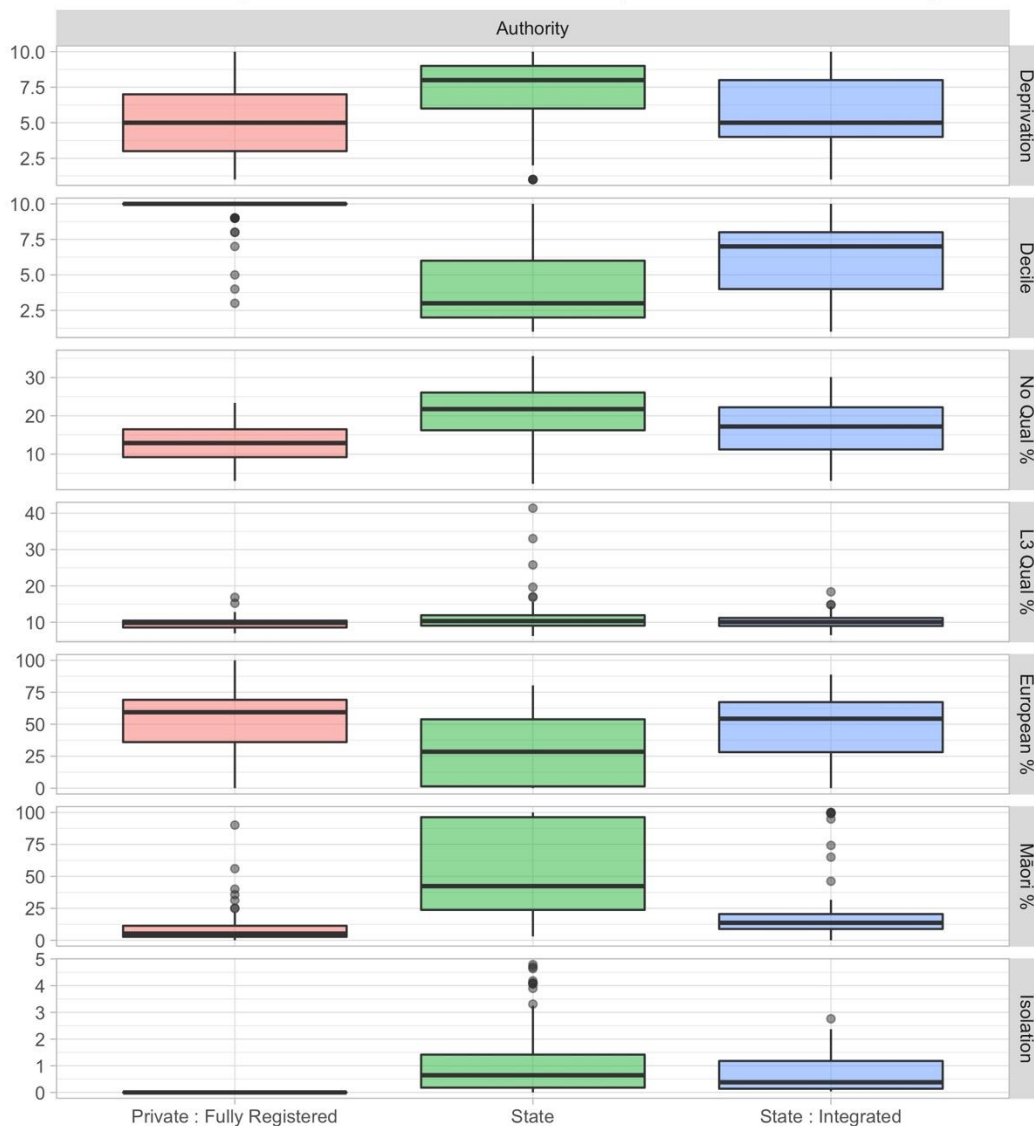


Figure 2: Categorical variables analysis: Authority

The median deprivation score for state schools is above the upper quartiles of both private and state-integrated schools, meaning state schools are more likely to be in deprived areas, as expected.

Private schools have a much higher median decile and much lower interquartile range than the other two categories, as expected. State-integrated schools also have a higher median decile than state schools.

State schools have the highest median percentage of unqualified people around them while private schools have the lowest, which meets the stereotypes. However the median percentage of Level 3 qualifications is surprisingly even across categories; in fact, state schools' median is higher than the other two, but not enough to be statistically significant.

Private schools have the highest median percentage of European students and state schools have the lowest, as expected. Likewise, private schools have the lowest median percentage of Māori students and state schools have the highest. Private and state-integrated schools both have much lower interquartile ranges, meaning there is little variation in these schools' percentage of Māori students.

State schools are more isolated as expected, while private schools are the least isolated and also have a low interquartile range.

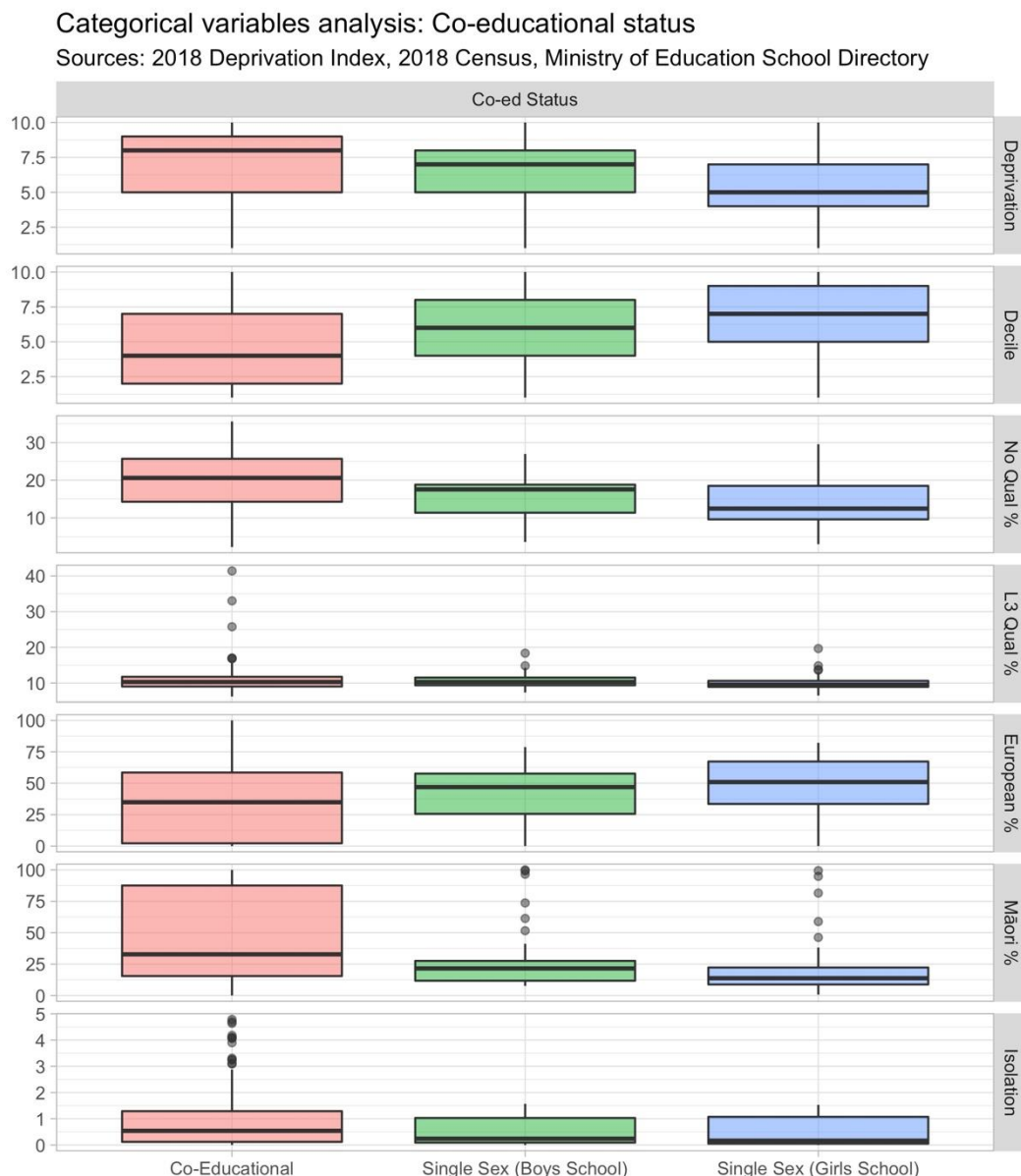


Figure 3: Categorical variables analysis: Co-educational status

The median deprivation score for co-educational schools is above the upper quartiles of both boys' and girls' schools, meaning co-educational schools are more likely to be in deprived areas, as expected. Interestingly girls' schools have the lowest median deprivation score, perhaps because society sees it as more important that girls are sent to schools with a clean, safe image.

Co-educational schools have the lowest median decile as expected, while girls' schools have the highest, perhaps for the same reason.

The median percentage of unqualified people is highest for co-educational schools as expected, meanwhile it is lowest for girls' schools. The median percentage of Level 3 qualifications is very similar across categories; in fact, it may be slightly higher in co-educational schools, but not enough to be statistically significant.

Co-educational schools have the lowest median percentage of European students and the highest median percentage of Māori students, as expected. The interquartile ranges of single-sex schools in terms of Māori student percentage are both much narrower, meaning there is less variation in these schools' percentage of Māori students.

Co-educational are more isolated as expected, however there is not as much variation across the three categories in comparison with other variables.

Conclusion:

In this report I have processed and analysed data relating to North Island secondary schools and their neighbourhoods. I have found some interesting relationships between variables including ethnicity, deprivation, authority and co-educational status. I will explore these further for Part B of this project, as they could highlight particular biases held by parents and students when choosing a school, and may help schools and policy-makers to combat these biases.

References:

- Atkinson, J., Salmond, C., & Crampton, P. (2019). *NZDep2018 Index of Deprivation User's Manual*. University of Otago. <https://www.otago.ac.nz/wellington/otago730391.pdf>
- Boyack, N. (2019, July 22). *Call for debate on 'white flight' from our low decile schools*. Stuff. <https://www.stuff.co.nz/national/education/114067757/call-for-debate-on-white-flight-from-low-decile-schools>
- Donnelly, B. (1998, June 04). *Kura Kaupapa Māori*. New Zealand Government press release. <https://www.beehive.govt.nz/release/kura-kaupapa-maori>
- Dyslexia Foundation of New Zealand. (n.d.). *2015 SAC DATA*. https://www.dyslexiafoundation.org.nz/dyslexia_advocacy/2015-sac-data.php
- Gates, Charlie. (2017, May 08). *Christchurch students travel the equivalent of a trip to the moon to avoid local school*. Stuff. <https://www.stuff.co.nz/the-press/news/92208905/christchurch-students-travel-the-equivalent-of-a-trip-to-the-moon-to-avoid-local-school>
- Ministry of Education. (n.d.A). *School deciles*. <https://www.education.govt.nz/school/funding-and-financials/resourcing/operational-funding/school-decile-ratings/>
- Ministry of Education. (n.d.B). *Operational funding components*. <https://www.education.govt.nz/school/funding-and-financials/resourcing/operational-funding/operational-funding-components/#Isolation>
- Ministry of Education. (2021, September 26). *New Zealand Schools*. <https://catalogue.data.govt.nz/dataset/directory-of-educational-institutions/resource/20b7c271-fd5a-4c9e-869b-481a0e2453cd>
- Stats NZ. (n.d.). *Qualifications: highest secondary school qualification (information about this variable and its quality)*. http://datainfoplus.stats.govt.nz/Item/nz.govt.stats/2628f2f2-be94-4132-96e9-dbea88dd7c07/?_ga=2.254203336.1376150028.1632529238-1524024078.1632438216
- Stats NZ. (2017). *Statistical standard for geographic areas 2018*. <https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Methods/Statistical-standard-for-geographic-areas-2018/statistical-standard-for-geographic-areas-2018.pdf>
- Stats NZ. (2018, May 17). *Statistical Area 2 2018 Clipped (generalised)*. <https://datafinder.stats.govt.nz/layer/92213-statistical-area-2-2018-clipped-generalised/>
- Stats NZ. (2020, May 21). *2018 Census Individual (part 2) total New Zealand by Statistical Area 1*. <https://datafinder.stats.govt.nz/layer/104616-2018-census-individual-part-2-total-new-zealand-by-statistical-area-1/>
- University of Otago. (n.d.). *Socioeconomic Deprivation Indexes: NZDep and NZiDep, Department of Public Health*. <https://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html#2018>