Diving Deeper into Deprivation within New Zealand

Project Report

Authors:
Josh Atwal
Cadence (Yifei) Liu

Prepared for:
Dr. Daniel Exeter

University of Auckland Department of Statistics

October, 2021

Abstract

Inequality and deprivation are major issues in New Zealand. The 2018 Index of Multiple Deprivation (IMD18) is a tool developed to map the relative deprivation between geographical regions in New Zealand. IMD18 is distributed online as a web app but lacks the ability to explain the underlying drivers behind deprivation. Detailed reports on deprivation are available but are dense, difficult to interpret for non-technical audiences, and only come at a broad territorial region.

With the purpose of better understanding and explaining deprivation, we propose the New Zealand Deprivation Explorer (NZDE). Built in Tableau, NZDE incorporates multiple sources of data to recreate and extend the functionality of IMD18 with easily interpretable visualisations and information. NZDE is designed to be easily interpretable and is currently in the process of being hosted online for public access. NZDE lays a blueprint for effective communication of data to the public by presenting interpretable statistics from various sources in a user-friendly tool. We believe NZDE will help to foster new discussions around deprivation while empowering policy and decision-makers to create the change necessary to address it.

Contents

Introduction

New Zealand is a developed oceanic nation with a reputation for beautiful landscapes, welcoming people, and a laid-back atmosphere. Tourism and global perception are an important component of New Zealand's economy, but the reality for many living in New Zealand is in stark contrast to the portrayed image. New Zealand struggles with inequality and deprivation on a wide range of fronts. In terms of both the "top 20% vs bottom 20% household income" and "top 10% wealth holdings" measures, New Zealand is worse than the OECD average as well as comparable countries like Australia, the UK, and Canada [1]. On top of this, New Zealand is a multi-ethnic nation and issues of inequality and deprivation are often worse for ethnic groups like Māori and Pasifika. For example, **Figure 1.1** shows the proportion of Māori living in deprived areas (according to the 2013 NZ deprivation framework). For a more recent example, see the discrepancy in Covid-19 vaccination rates between Māori/ Pasifika and the population averages.

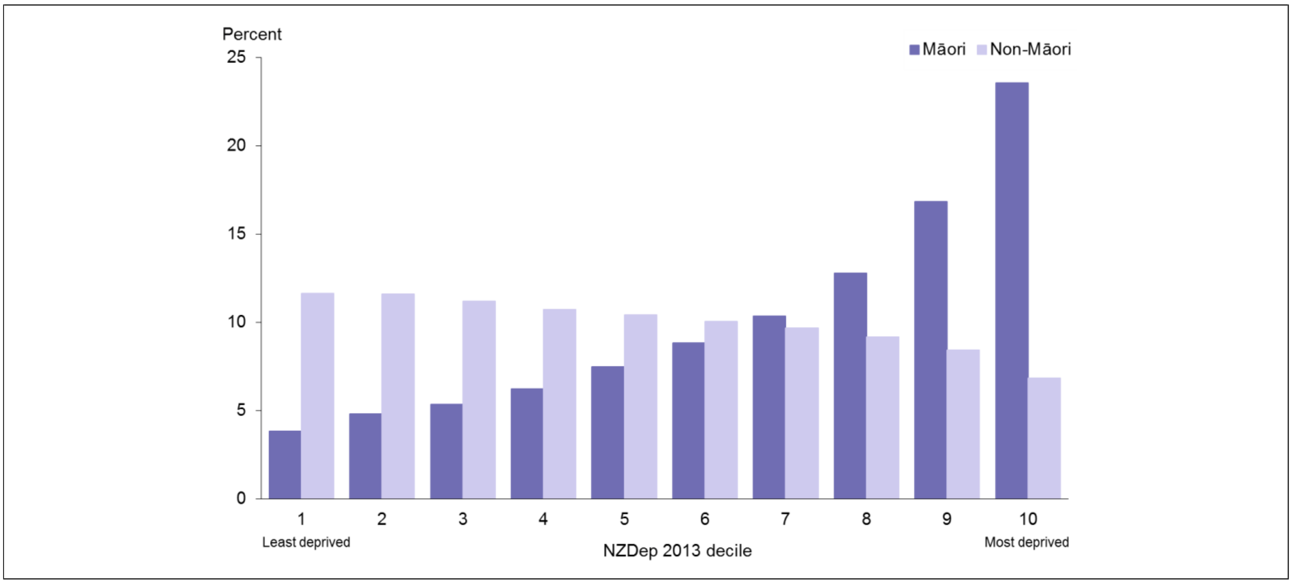


Figure 1.1: Neighbourhood deprivation distribution (NZDep 2013), Māori and non-Māori. Reproduced from the Ministry of Health [2].

The 2018 Index of Multiple Deprivation (IMD18) [3] considers deprivation in terms of seven dimensions of social conditions: Access, Crime, Education, Employment, Health, Housing, and Income. Various measures and social determinants are used to calculate an index of deprivation for each dimension and an additional summarising dimension that considers all of the above. IMD18 partitions New Zealand into *datazones*: small geographical regions containing ∼700 residents in each. Depending on the area, this may be roughly equivalent to a suburban street. Datazones fit into the New Zealand statistical standard for geographic areas [4] which defines other geographical regions: Statistical Areas 1 and 2 (SA1, SA2) and meshblocks. In terms of relative size, SA2 > datazone > SA1 > meshblock. Publicly available data is often reported at the SA2 level, and data specified at the smaller regions may require special access.

IMD18 ranks each datazone between 1 and 6,181 and partitions the datazones into deciles for each dimension of deprivation. The authors of IMD18 claim that mapping relative deprivation is a valuable way of demonstrating inequalities between geographical regions and can provide insight into many social and health outcomes [5].

IMD18 is distributed via an interactive web app [6] where users can select, query, and compare areas or regions of interest. The app is available publicly for use by anyone, with applications in decision making, research, and policy. While the app is useful for visualising deprivation, it is unable to explain the underlying drivers of deprivation. Detailed statistics are currently only available within static PDF reports and only for the major regions of New Zealand. Furthermore, these reports are dense and challenging to interpret without technical expertise. There are times when a user may have more specific questions such as "what sort of crime is being committed in this area" or "what is the income breakdown of residents in this area" which are not currently being addressed. Dr Daniel Exeter [7] is the principal author behind IMD18 and a frequent collaborator with the University of Auckland's Department of Statistics. He served as the client for this project, having instigated it when contacted by a District Health Board (DHB) decision-maker with such questions about IMD18.

The IMD18 app is an off-the-shelf solution that lacks the customisability needed for deeply exploring deprivation. *Tableau* [8] is data visualisation software originating from the Department of Computer Science at Stanford University. Acquired in 2019 for $15.7 billion [9], Tableau is a commercial product known for producing powerful dashboards and visual data storytelling. It is already well established in the business and government sectors. For example, the New Zealand police publish crime victimisation data on their website using a Tableau dashboard [10].

1.1 Objectives

The problem this project aimed to solve can be broken down as follows:

- Visualising and understanding the geographical variation of deprivation is essential for providing support and allocating resources to those most in need.
- The existing IMD18 app measures and visualises deprivation but does little to explain the underlying factors behind deprivation. Additionally, reporting made available by IMD18 is difficult to interpret and only available at a territorial region level.
- The software behind the IMD18 app lacks the customisability to provide avenues for exploring deprivation.

Therefore, the overall objective of this project was to improve on the existing IMD18 web app in terms of what insights were available while maintaining a user-friendly and broadly accessible interface. Tableau was used to develop the *New Zealand Deprivation Explorer*: (NZDE), a tool to supplement and potentially replace the existing IMD18 app. New data sources were accessed, and easily-interpretable visualisations were created to answer the more specific questions a user might have about the driving factors behind deprivation. NZDE was even extended to include Covid-19 vaccination data, adding an additional and socially relevant dimension of deprivation to be explored. NZDE lays a blueprint for effective communication of data to the public by presenting interpretable statistics from various sources in one user-friendly tool.

This project was completed over 10 weeks at the University of Auckland for Statistics 780.

Data

This chapter covers the methodology and technical details involved in developing NZDE. It begins with dataset acquisition and data pre-processing before covering how NZDE was built in Tableau.

2.1 Data Sources

A thorough survey of data quality and availability was conducted prior to any form of data acquisition—data sources were evaluated with considerations regarding usefulness, reliability, and granularity. Six sources of data made the final list of data sources for NZDE, with five of them being publicly available and one private data source only available to academics and researchers.

IMD18

The 2018 Index of Multiple Deprivation breaks New Zealand down into 6,181 datazones. Each datazone is given a deprivation score in each of the following eight deprivation domains [5]:

- *Access*: Distance to amenities including healthcare facilities, supermarkets, service stations, schools.
- *Crime*: Crime victimisation rates.
- *Education*: School leavers, working-age people without qualifications, youth not in education, employment, or training.
- *Employment*: Number of people receiving payments for job seeker support.
- *Health*: Mortality ratios, hospitalisations, rates of selected cancers.
- *Housing*: Renting, overcrowding, dampness.
- *Income*: Income tested benefits, Working for Families payments.
- *Overall*: This combines the ranking and score of all the above into an overall deprivation ranking and score.

Based on the deprivation scores, IMD18 computes rankings from 1–6181 and deciles from 1–10 for each datazone. The rankings, deciles, and geographic datazone boundaries are available for public use and were used within NZDE to recreate the functionality of the existing IMD18 app: summarising deprivation in the various domains and visualising relative geographic deprivation. The underlying data used in developing IMD18 was sensitive and only available within the New Zealand Integrated Data Infrastructure (IDI).

Integrated Data Infrastructure

New Zealand's IDI [11] is an extensive research database containing microdata about New Zealanders and the households they live in. It can be considered a central source of truth for many governmental and NGO statistics regarding New Zealanders—mainly education, income, benefits, migration, justice, and health. Everything is centrally linked and aggregable at varying levels of granularity, making it a first choice for research regarding complex issues regarding our society and economy.

The main advantage of using IDI data was that all the data were available at the datazone level, making it a good fit for the IMD18 data. Additionally, IDI data follows a strict standard, ensuring a high level of accuracy and reliability. From the IDI, the following data was retrieved at the datazone level:

- Age distribution
- Ethnicity distribution
- Never smoked vs Ex-smoker vs Regular smoker distribution
- Household access to vehicles (0, 1, or 1+ cars)

Due to the sensitive information it contains, extensive measures have been taken to ensure anonymity and security of the data in the IDI. All data stored in the IDI is removed of identifying information, and summaries of small geographic regions can be censored to avoid identifying individuals. IDI data is only able to be accessed by vetted researchers at approved physical locations. The lack of internet access to the IDI meant that, because of a Covid-19 lockdown, alternative and publicly available data sources had to be considered.

NZ Census

Every five years, a nationwide Census survey is run by Stats NZ to capture a comprehensive snapshot of New Zealanders. The Census is the only survey in NZ to cover the entire population. The data collected is used to reach higher levels of transparency within NZ and assist with the planning and development of governmental services. As such, all data collected in the Census are made available to the public online via Stats NZ [12].

The Census has historically been sent out via postal mail to physical household addresses of all persons presently living in NZ as a questionnaire about themselves and their households. An exception to the physical model was made in 2018, whereby Stats NZ trialled a digital-first approach, preferring to reach New Zealanders via an E-Mailed questionnaire. Compared to the 92.2% response rate from 2013, the 2018 census had a comparatively lower response rate at 88.3%. From Stats NZ the following data were extracted at the SA2 level:

- Housing Mould (Mould A4 size: Always, Sometimes, Never)
- Housing Dampness (Dampness: Always, Sometimes, Never)
- Family composition (Couple with children, Couple without children, One parent with children)
- Income distribution
- Sources of income

The data being at the SA2 meant that downsampling was required to bring the data to the datazone level for consistency with the other data sources. Further details on this is available in section 2.2 on Data Pre-processing.

Unfortunately, it is often deprived areas that are over-represented in non-response, but this is another limitation of the data available.

Ministry of Health

Covid-19 is a present-day example of how deprivation can vary geographically and with factors like ethnicity. The Ministry of Health provides continuously updated Covid-19 vaccination counts at the SA2 level [13]. These are available for the total population and separately for the Māori and Pasifika populations. Vaccination data were used to provide an alternate map overlay to the IMD deprivation data, allowing users to explore deprivation alongside vaccination rates. Because the vaccination data was published at the SA2 level, the same pre-processing approach as with the Census data was needed to downsample to the datazone level.

NZ Police

From the official NZ Police website, we were able to download crime victimisation data [10] broken down into six categories:

- Abduction and Harassment
- Sexual Assault
- Robbery and Extortion
- Acts intended to cause injury
- Burglary
- Theft

The crime data came as counts of each type of crime at the meshblock level, meaning aggregation was necessary to get this data to the datazone level. Further pre-processing was done to convert the counts into rates per 1,000 population, allowing for more equitable comparisons between datazones of varying sizes.

OpenStreetMap

When deciding between services offering location-based data for exploring the relationship between deprivation and proximity to amenities, Google Maps Platform (GMP) and Openstreetmap were shortlisted during our survey of available data in this domain.

During our comparison of the two services, GMP offered a suite of geospatial data solutions with a heavy-handed commercial approach targeting transport and logistics application developers especially. For example, an extensive list of features such as traffic optimisation and dynamic street view were identified, most of which would be underutilised in this project's scope. Though GMP did contain the core functionality of location search and coordinate extraction, it was concluded that paying for a suite of features outside the scope of our research was unnecessary.

OpenStreetMap was, therefore, the preferred alternative, providing all the core functionality required at no cost. It was also noted that many research projects involving geospatial visualisation preferred to use OpenStreetMap due to it being free and highly descriptive, with excellent support for data export. OpenStreetMap (OSM) was inspired by the success of Wikipedia and set out with the goal of becoming the open-source geographical database of the world [14]. OSM users crowdsource data using GPS devices, aerial photography, or their local knowledge of their area. OSM data is re-licensed under the Open Database License by the non-profit OpenStreetMap Foundation before being published to the main OSM service.

Published OSM data points are all categorised and tagged according to strictly defined guidelines [15]. By referencing the tags of each data point, a JSON query was constructed to retrieve data from the OSM API. This was exposed via *Overpass Turbo*, an open-source data mining tool explicitly created for OSM [16]. The exact JSON query used can be found in Appendix 6.3.

The following categories of locations of interest were obtained from OSM:

- Fast food outlets
- Liquor stores/Pubs/Bars
- Grocery and Convenience stores
- Healthcare facilities
- Parks and Playgrounds
- Schools
- Geysers
- Bus stops and Train stations

Even more detailed tags were available, but these broad categories were sufficient for purpose. Many more categories of locations were also available, but these served as a demonstration of what was possible. In total, ∼23,000 point locations of interest were plotted in NZDE as an additional overlay atop the IMD18 map. Geysers may seem an odd choice of amenity. However, our client was specifically interested in visualising how poor air conditions near geysers in Rotorua may be driving deprivation in terms of health outcomes. Locations of bus stops and train stations were initially included, but were later dropped due to the large number of points adversely affecting performance.

2.2 Data Pre-processing

Table 2.1 summarises each data source and the granularity it was available at. IMD18 uses data at the datazone level, and this was the preferred granularity. Pre-processing of the statistical (i.e. the non-geospatial) data was necessary to aggregate it when at the meshblock level, and downsample it when at the SA2 level. **Figure 2.1** provides visual reference to the relative sizes of SA2, datazone, and meshblock regions. Fortunately, the New Zealand statistical standard means that all of these regions nest perfectly within each other. E.g. each meshblock belongs to only one datazone and so forth.

Data pre-processing was performed in R with heavy usage of the *tidyverse* [17] package. Appendix 6.1 contains the R code in full.

Source	Granularity	Description
OpenStreetMap	Coordinates	Fast food and Alcohol, Geysers, Groceries and Convenience, Healthcare, Parks and Playgrounds, Schools.
NZ Police	Meshblock	Crime victimisations by crime type.
IMD18	Datazone	Ranking and decile of deprivation in each domain.
IDI	Datazone	Age, Ethnicity, Smoking, Vehicle access.
Census	SA2	Household mould/dampness, Income, Family composition, Financial assistance.
Ministry of Health	SA2	Covid-19 Vaccination rates.

Table 2.1: Summary of data sources and granularity.
IDI*: Note the Integrated Data Infrastructure is not a publicly available data source.

Figure 2.1: Comparison of different levels of geographic granularity.

Aggregation and Downsampling

The NZ police dataset came at the meshblock level. A mapping table provided by IMD18 specified, for each meshblock, the datazone it fits within. Multiple meshblocks can fit within a datazone, but each meshblock resides within only one datazone at a time. Aggregating the meshblocks to the datazone level was then a matter of: for each datazone, summing the statistics over all meshblocks within that datazone.

The NZ Census and Ministry of Health data came at the SA2 level. For these data, a similar mapping table was used to downsample from the SA2 to the datazone level. The statistics reported at the SA2 level were distributed among the datazones within the SA2 based on the population of each datazone. For example, consider an SA2 S containing two datazones d_{10} and d_{90} with populations 10, 90 respectively. If S had 50 single parents, 5 of them would be attributed to d_{10} and 45 to d_{90}. This involves the major assumption that the population of S is uniform across datazones but is the best that can be done given the data available.

Additional Pre-processing

The NZ Census data on housing and financial assistance required additional pre-processing after downsampling. The housing data contained two separate variables measuring reported mould and dampness. These were combined into a single combined measure of mould and dampness. First, the variables were recoded into numeric data. E.g. for “Mould over A4 size”, “Always” was coded as 3, “Sometimes” as 2, and “Never” as 1. The same idea was applied to the dampness variable, with the same “Always”, “Sometimes”, “Never” scale being recoded as

3, 2, 1, respectively. Next, the two variables were summed together, with the combined Mould and Dampness index being 1 if the sum was greater than 3 and 0 otherwise. This simplified things considerably and allowed for easier identification of areas with poor housing standards.

The NZ Census data on financial assistance also needed simplification. The Census specified 13 different income sources, which were grouped into three simple categories: "Independent", "Assistance", and "No Income". The full definition of how these categories were defined is available in Appendix 6.2.

Geospatial Data

Unfortunately, Overpass turbo only supported GEOJSON file format exports while our IMD18 map was provided in a shapefile format. QGIS [18] is free and open-source software allowing cross-format editing and analytics over all common formats of geospatial files, with extensive integration for the combination and export of graphical maps. QGIS offered conversion between file formats and other useful tools for processing geospatial data. Large locations such as parks and schools came with an area/border definition—consisting of multiple joined X-Y coordinates instead of a single X-Y coordinate (like liquor stores). During further processing, this caused major consistency issues as each location needed to be associated with a single datazone. The "Calculate Centroids" function from QGIS was used to ensure each location was represented as a single point location. See Appendix 6.3 for a step-by-step guide on how QGIS was used.

Once the coordinate data were consistent, the GEOJSON layer was exported as shapefile coordinates aligned with the IMD18 map using the 'Merge vector layers' tool. This also enabled a count of amenities within each datazone via the "Count points in polygon" feature. The shapefile coordinates could then be overlaid on top of the IMD18 map to visualise the relationship between deprivation and proximity to amenities.

2.3 Data Summary

With the geospatial and statistical steps outlined above, we have created not just a tangible product, but a blueprint for further expansion. Future researchers looking to add additional amenities can edit the JSON query to download the required data points and follow the steps outlined in Appendix 6.3 to prepare the data for Tableau import. The statistical data follows an even simpler process: the R code was written with generalisability in mind. All operations were done by referring to columns by name rather than index. Once data is loaded into Tableau, the next chapter will demonstrate how easily additional visualisations can be created and incorporated within Tableau.

Visualisation

This chapter introduces the New Zealand Deprivation Explorer (NZDE), explaining how it was built before describing the interface and functionality.

Figure 3.1 shows the full display of NZDE. The main content area is a map of New Zealand broken down by datazone for visualisation of deprivation. Two options are available for the map overlay: 1) IMD deprivation domains and 2) Covid-19 vaccination rates. In each case, the map overlay can be basic (Low-Medium-High) or advanced (decile from 1 to 10). The colour scheme of the existing IMD18 app has been retained for consistency and familiarity. The interface is reactive even when zooming and panning. Hovering over a datazone displays a tooltip with the datazone ID, DHB, and the national rank of the datazone (based on the currently selected IMD domain overlay). The right-hand side pane contains supplementary visualisations and an amenity location overlay for exploring deprivation. Options are available for DHB/national level comparisons against the datazone statistics.

3.1 Creating NZDE in Tableau

Once the data were ready, they were read into Tableau as geospatial shapefiles and CSVs. Common to every data file was a datazone column holding the unique ID of the datazone created during data preprocessing. This enabled efficient joining of the data sources to the original IMD18 data. Visualisations were composed together into the dashboard that is NZDE.

The visualisations used in NZDE fall into two categories: the main map visualisation of New Zealand and all the supplementary visualisations used for exploring deprivation. In each case, generating visualisations in Tableau followed two main steps. This is best illustrated with the example: see the age distribution in the bottom right-hand corner of **Figure 3.1**.

The first step was to assign variables to the rows and columns of the visualisation. Variables can be *dimensions* (e.g. Age group, Crime type) or *measures* (e.g. counts, proportions). For the age distribution example, datazone and age group are the columns and are both dimensions. The variable assigned to the rows is the count of people per age group per datazone. This means that one plot is created for every datazone, and within each plot a point is plotted for every age group.

The second step is to specify the *marks* of the plot. The marks specify the visual representation of the plot: e.g. a line, bar, pie chart. For the age distribution, an area chart was chosen. Within the marks, additional variables can be assigned to different visual roles (e.g. the size, colour, or tooltip of points). For example, the datazone population could be assigned to the size component of the map to create a map visualisation scaled by population.

All visualisations in NZDE followed the above basic process but involved some additional complexities and features of Tableau. *Dual-axes* allowed for a second data series to share the same set of axes and were also used in displaying DHB and national level averages on the supplementary visualisations like the orange line on the age distribution. Rather than compute these averages in R and append them to the data, these were computed within Tableau

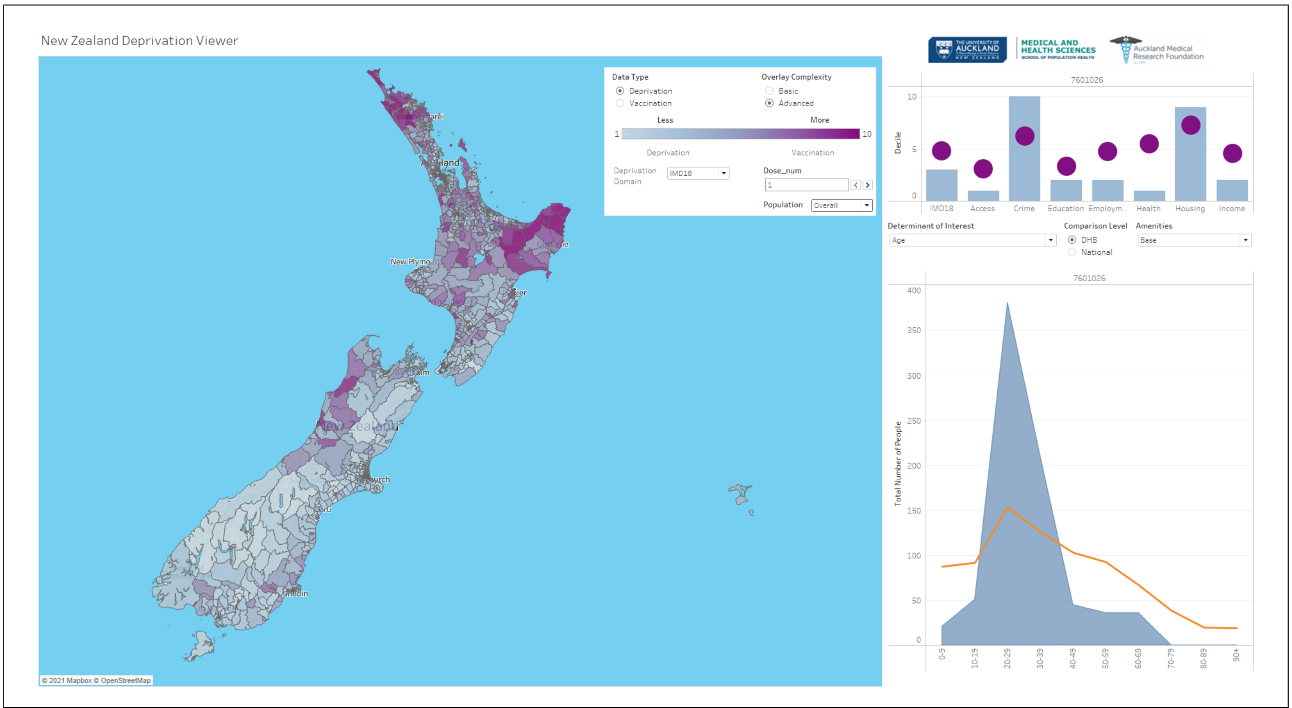


Figure 3.1: New Zealand Deprivation Explorer

using *Parameters* and *Calculated Fields*. Parameters were global variables that were modifiable through user interaction with the visualisation. Calculated Fields were variables defined within Tableau using variables in the data, parameters, and programming constructs like functions. For example, displaying the DHB/national average on top of an existing plot was done by:

1. Creating a parameter for a user to toggle between displaying a DHB/national average.
2. Creating a calculated field that computes the DHB/national average. This involved an `IF` statement for control flow and a `FIXED` expression for computing aggregate statistics. If the value was set to "DHB", compute the average of all age observations `FIXED` on the DHB and age group.
3. Using a dual-axis to display the calculated field alongside the original age distribution. The colour and type of mark could be changed independently of the age distribution.

Each dropdown and toggle in NZDE was associated with its own parameter that drove a calculated field determining what was displayed on-screen. Calculated fields were also used to convert count data into rates. For example, the NZ Police crime data contained counts of victimisations which were transformed into counts per 1,000 by dividing by the population of each datazone. A similar procedure was used to convert the counts in the family composition and financial assistance visualisations into proportions. Examples of calculated fields are available in Appendix 6.4.

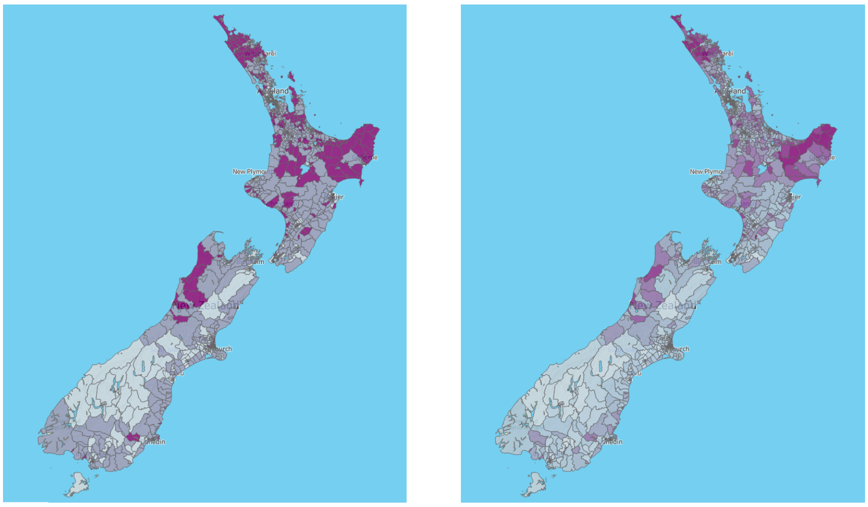
The main map visualisation further serves as an example of Tableau's functionality. The point locations of amenities were visualised on top of the base map using a *dual-axis*. The colour of the map overlay was a complex calculated field involving parameters such as the one controlling whether deprivation or vaccination data was displayed. The control panel in the top right corner of the map shows interactive parameters that control different aspects of the map. The colour scale limits are automatically scaled based on the range of data visualised on the map overlay.

Selecting datazones on the map drove an *action* linked to a *filter*. Based on the contents of the filter, only a subset of the data was exposed to the supplementary visualisations. This meant that a visualisation (such as the age distribution) only produced a plot for the currently selected datazones, with the filter being updated by selecting datazones from the map. Point locations of amenities were also implemented using a filter to allow for multiple amenities to be displayed simultaneously.

Finally, creating the dashboard was a matter of dragging and dropping visualisations and parameter controls into either a resizable grid or floating layout. Images were also included attributing funding and support. The following sections go into detail regarding the interface of NZDE.

3.2 Exploring Deprivation

Exploring deprivation is the main focus of NZDE. The main map overlay visualises the relative geographic deprivation between datazones in New Zealand. The “Deprivation Domain” dropdown specifies which of the eight IMD domains (including the summary domain) are currently visualised on the map with the colour scale. **Figure 3.2** compares the basic and advanced options for the colour scale. The basic view allows for quick decision making and identification of deprived areas. This keeps high-level decision-makers in mind, as often getting bogged down in the details may hinder decision making. The advanced view still provides nuance and depth to comparisons.



(a) Basic (b) Advanced

Figure 3.2: Deprivation in New Zealand using the summary domain as map overlay.

The right-hand side pane of NZDE contains visualisations and information for exploring deprivation. The bar plot in the top portion of the pane shows the deprivation profile of the currently selected datazone(s). This provides a quick summary and quickly identifies the domains in which datazones are particularly deprived. The bar plot is similar to what is seen in the existing IMD18 app.

The remaining portion of the pane below the bar-plot is the main area for explaining the factors driving deprivation. The “Determinant of Interest” dropdown selects the displayed visualisation; see **Figure 3.3** for all visualisations available. The “Comparison Level” selects whether comparisons are made to the DHB or national level. This affects both the IMD summary bar plot and the determinant of interest visualisations.

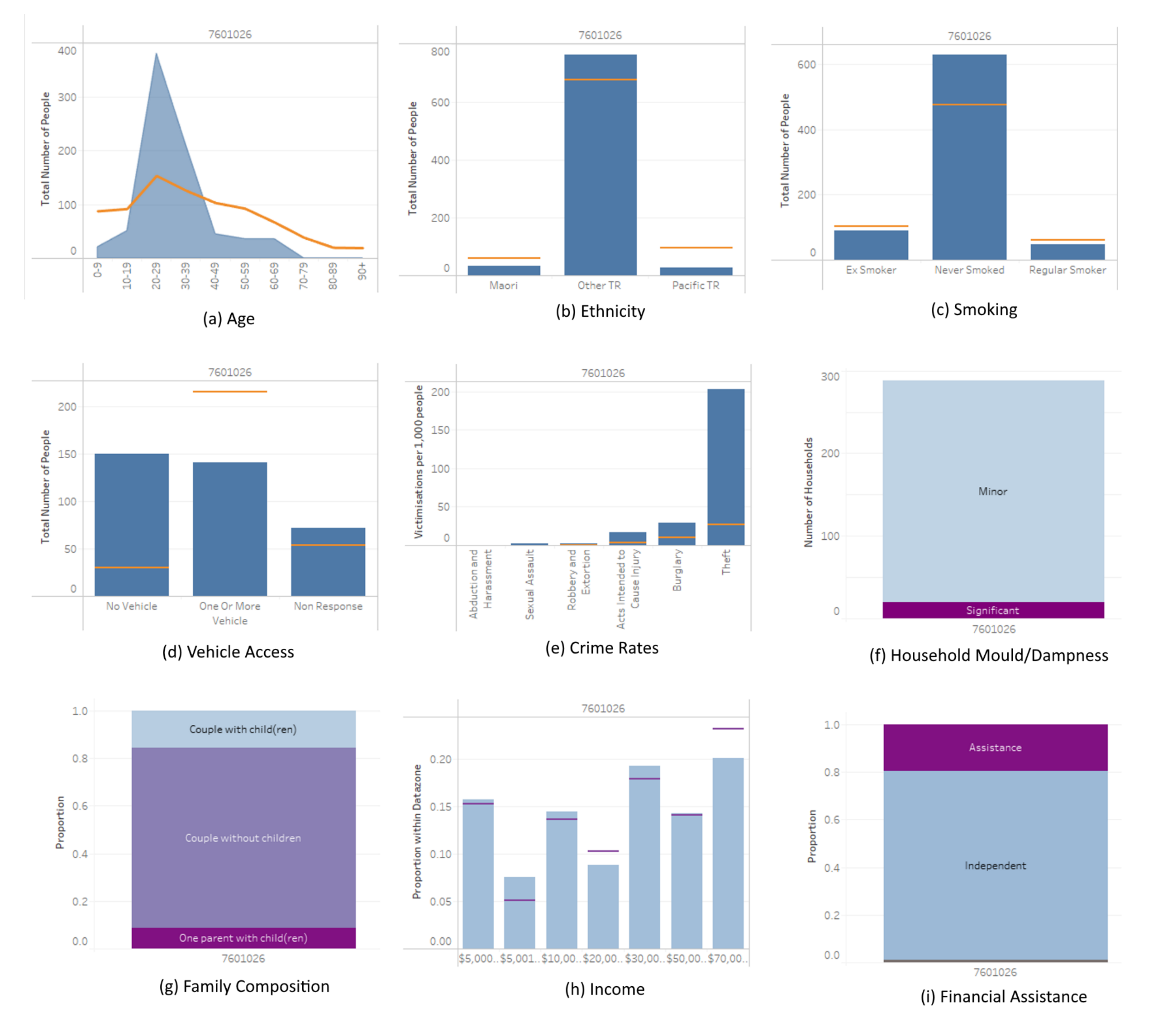


Figure 3.3: Visualisations available for deprivation determinants. Datazone *7601026* (for which these visualisations are displayed) is in Auckland and encompasses the main city campus of the University of Auckland. Comparisons are to Auckland DHB and visualisations (f) onwards use downsampled data.

The "Amenities" dropdown allows the user to visualise the locations of various amenities on top of the map display. This allows for the association of deprivation with proximity to things like fast food and alcohol outlets. **Figure 3.4** shows the North Island of New Zealand with each of the amenities overlaid. OpenStreetMap contains many more categories which could have been included—the current set of amenities are some of the most relevant to deprivation.

(a) Fast food and Alcohol

(b) Geysers

(c) Groceries and Convenience

(d) Healthcare

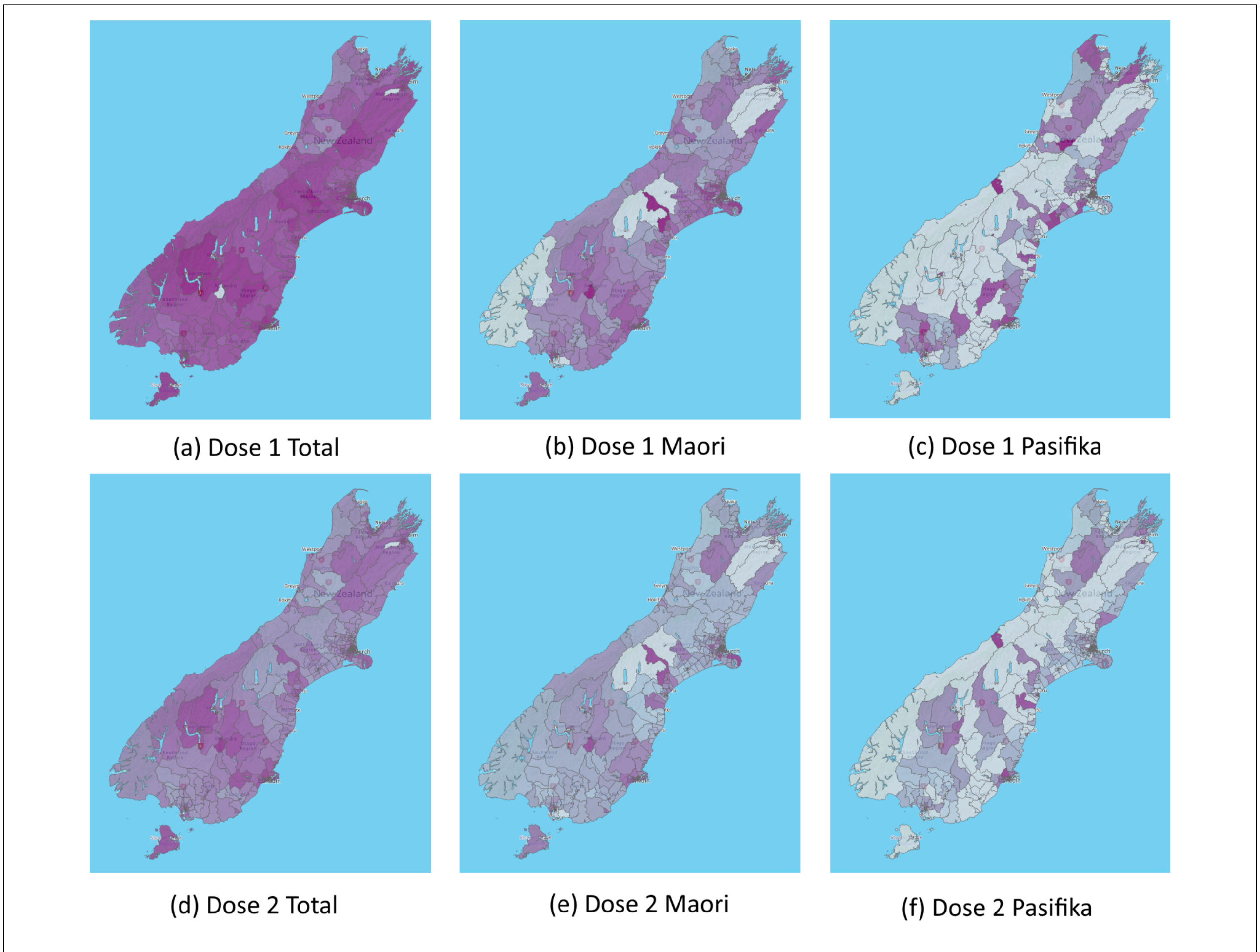
(e) Parks and Playgrounds

(f) Schools

Figure 3.4: Visualisation of amenities in the North Island.

3.3 Vaccination Rates

Covid-19 is a present-day example of how deprivation can vary geographically and with factors like ethnicity. NZDE provides the ability to visualise Ministry of Health vaccination data as a map overlay along with all the other tools for exploring deprivation. **Figure 3.5** visualises vaccination rates in the South Island. This provides a powerful visualisation of the discrepancy in vaccination rates between Māori/Pasifika and the rest of the population. Visualisations like this have the potential to be far more impactful than raw statistics and drive home the need for initiatives to boost Māori and Pasifika vaccination rates. This is one example, but the principle applies more broadly to all areas of deprivation.



(a) Dose 1 Total (b) Dose 1 Maori (c) Dose 1 Pasifika

(d) Dose 2 Total (e) Dose 2 Maori (f) Dose 2 Pasifika

Figure 3.5: Visualisation of vaccination rates in the South Island.

Discussion

Overall, the project was a major success. The client was pleased with NZDE, and work has begun on online hosting for public access in late 2021. NZDE is intuitive and user-friendly, maintaining a broad range of use cases and not limited to technical audiences. Some potential examples of how NZDE could be used include:

- A young family are looking at moving to an unfamiliar city. They use NZDE to identify suburbs low in crime and with schools nearby.
- A community advocate uses NZDE to visualise the income deprivation and the number of fast food and alcohol outlets in their community. They use this as evidence to oppose new liquor licenses.
- A DHB member uses NZDE to visualise areas deprived in health and lacking access to healthcare facilities. They use this information to determine the distribution of funding.

Even if a user does not have a specific objective in mind, NZDE confronts the user with the reality of deprivation. For example, there is a great deal of difference between being told an area is deprived and seeing what that means regarding crime rates, income distributions, and family compositions. This naturally raises questions in the mind of the user. For example, "why are these areas experiencing this kind of deprivation?" "Is this fair?" "What can be done about it?"

The methodology behind NZDE has laid a blueprint for further work on including more data and improving the tool's value. NZDE serves as an example of effective data communication and transparency towards the public and what is possible by incorporating a range of data sources.

We believe that having a tool like NZDE available for the general public to use and explore deprivation will lead to questions being raised, conversations broached, perspectives broadened, and steps towards change being made. Academics, researchers, agencies, policy and decision-makers are all empowered by NZDE in having the tools necessary at hand to understand deprivation and bring change into reality.

4.1 Limitations and Future Work

This project entails exciting opportunities for further development towards better explaining deprivation.

NZDE allows multiple datazones can be selected, but only for direct comparison. It may be desirable to understand the deprivation profiles of broader regions. Similarly, datazones or regions may need to be compared to other regions. For example, a more advanced comparison may compare a suburb of datazones to a similar suburb in another city. Currently, datazones can only be compared to DHB or National averages. Future work could therefore look at allowing for selections of custom regions for inference or comparison.

One of the areas of improvement for IMD18 was in its reporting. Reports were dense, technical, and only available at a territorial region level. NZDE has gone a long way towards incorporating a breadth of information in an approachable alternative to the report format, but a report may still be preferred in some scenarios. Tableau has built-in "Explain data" functionality, and future work could investigate automating report generation through Tableau. Tableau also offers a data pre-processing suite *Tableau Prep.* This integrates directly into the visualisation workflow and could be used instead of R to pre-process the data, allowing for automatic updates of the underlying data used in the visualisations.

A limitation of the data available for NZDE was that some of it came at the SA2 level rather than the more granular datazone level. This necessitated downsampling and meant that the figures reported for the data from these sources were less reliable. The assumption is that all datazones in an SA2 have similar characteristics. Analysis could be done to test the validity of this assumption and features could be built to warn a user about interpreting downsampled data. However, in the future, these data sources could just be replaced with data from the IDI natively at the datazone level, avoiding downsampling and its limitations entirely. Additional data sources could also be investigated, such as house prices (with homeownership being a major inequality issue in New Zealand). Data on court cases (for example relating to domestic violence or drug abuse) were also of interest. However, upon investigation, it was not available at a granular enough level.

4.2 Tableau Evaluation

While Tableau may be widely used in certain sectors like business and government, it is something the client and the University of Auckland Department of Statistics was less familiar with. This section justifies the choice of Tableau and evaluates it as a visualisation tool.

After an initial investigation into solutions that would be able to achieve our objectives, *Shiny* [19] and Tableau was shortlisted as the two most viable candidates. Shiny is a package for the R programming language designed for building interactive web apps. Given that R was developed by the University of Auckland's Department of Statistics, Shiny apps are ubiquitous and have a proven track record. While the client was open to both options, they were also interested in trying something new and seeing what Tableau offered.

The main distinguishing factor of Tableau is its user interface; Tableau is targeted at a less technical audience and assumes no prior programming skills. Although it comes with a learning curve, compared to creating visualisations programmatically, the mouse-based user interface is far more intuitive. Creating plots can be done by dragging and dropping, and editing plot elements is done by directly interacting with the visualisation. Individual visualisations can be composed into dashboards, and dashboards into "stories"—allowing complex narratives to be easily constructed. Supplementary elements like images, text, plot controls, and colour scales can be dragged and dropped into a grid layout or left floating. Tableau offered options for composing layouts for a range of displays, including mobile phones and smaller screens. Tableau's user-friendliness may be particularly relevant for this project as continuing development on the app can be done without technical expertise. Future data refreshes mean that the visualisations will eventually need to be updated.

Tableau does a lot of the hard work in creating visualisations for the user. The default options were modern and attractive, types of plots are suggested given the input data, and interactivity like tooltips and linked brushing are included by default. Geospatial data can be difficult to work with, but Tableau natively handles a variety of geospatial formats. Creating geographic visualisations follows the same procedure as producing any other kind of visualisation. Tweaking parameters like map appearance and layers are also natively handled by Tableau.

Being an established commercial product with a solid reputation and proven track record also means that users can be at ease that any bugs will be fixed, and ongoing development means more features in the future. Tableau is well documented online with plenty of support available and many online courses offering to teach it. The broader Tableau product ecosystem includes free hosting of (non-confidential) data and visualisations. For sensitive data, a hosting and collaboration cloud platform is available. Tableau has basic support for common data pre-processing tasks like joins, variable refactors, and type conversions. For more complex tasks, Tableau Prep could have been used, but this was outside of the scope of this project.

Tableau also came with its share of downsides. Doing everything with a mouse means the reproducibility of code is lost. Rather than being able to reuse code, procedures must be followed through manually and must be documented precisely for accurate reproduction with a mouse. Being a commercial product also comes with licensing fees. Each fully-privileged user license currently costs $US 70 per month [20]. However, the University of Auckland does already offer Tableau licenses.

4.3 Reflection on Consulting

Both our client and their DHB client were pleased with the final product, evident from their plans for NZDE to be published alongside the IMD18 as an additional visualisation tool. Daniel himself took the time out of his busy schedule to attend our final project presentation and offer his kind comments praising our work. It is rewarding to know our efforts will make a meaningful difference and help those in need, something rarely offered by other courses at University. These core principles of consulting advice were referenced from STATS 780 lectures and applied throughout the project:

- *Adapting to challenges.* Although progress on this project was generally smooth, there were significant challenges that needed to be overcome. Because of a lack of access to the IDI, a significant amount of time had to be invested in finding alternative data sources. These data sources then needed processing like aggregation or downsampling to get them to the correct granularity level. Similarly, the geospatial data needed pre-processing and conversion to the correct file format. Neither of us had any prior experience working with geospatial data or with Tableau at all.

  Despite these challenges, we were able to rely on our skills and abilities as consultants to adapt and learn what was necessary to overcome them.

- *Time and Workload Management.* Weekly recurring meetings were scheduled with our client from the first consultation session, and the originally agreed schedule was kept up for all of 10 weeks. This is even despite Covid alert level changes. The regular meetings ensured constant progress on the project and the satisfaction of the client.

  Between the two team members, agreed blocks of time were allocated before/after each STATS 780 lecture to be dedicated exclusively towards progress on the project. The workload was split into manageable self-contained portions that could be worked on simultaneously. A shared Dropbox directory was created, and each team member's role was defined early on in the project. Cadence would be mainly responsible for data acquisition and pre-processing, while Josh would take on the chief responsibility of designing and developing the Tableau dashboard. Though our tasks were capable of being completed individually, we still spent much time bouncing ideas and helping each other out as we were both quite interested in each other's section of work. This had the added benefit of learning more from the project and providing some much needed social interaction during Auckland's level 4 lockdown.

- *Client Management*. A proposal document draft was presented to the client at the consultation session of the second week, which was then iteratively discussed and completed with input from the client by the third week. Extra care was taken to make sure we understood every nuance of the problem and that all assumptions were clearly stated and confirmed before the scope was firmly locked in. We were fortunate to be paired up with a highly supportive client that encouraged us to bring new ideas to the table. Thus it was crucial to ensure both parties understood what would be included and why it was included in the scope of the final deliverable. The creation of the proposal document enabled us to set a clear goal for the final product that we could deliver on.

  Throughout the project and the weekly meetings, our relationship with the client developed. Establishing a rapport and getting to know the client only led to better levels of understanding around our other commitments and a better outcome overall. We appreciated Daniel taking time out of his busy schedule to meet with us and attend our final project presentation. His appreciation for our work galvanised us to deliver a successful project. For example, incorporating the Covid-19 vaccination data was done even after the final presentation of the project. It was too good of an opportunity to add relevance and appeal to our product, elevating the profile of the discussion around deprivation.

Summary and Conclusion

Inequality and deprivation are major issues in New Zealand. New Zealand ranks worse in measures like income equality than the OECD average and other comparable nations like Australia, the UK, and Canada. IMD18 is a tool developed to map the relative deprivation between geographical regions in New Zealand and aid in decision making and policy around addressing deprivation. IMD18 is distributed online as a web app but lacks the ability to explain the underlying drivers behind deprivation. Detailed reports on deprivation are available but are dense, difficult to interpret for non-technical audiences, and only come at a broad territorial region. Our team of two postgraduate statistical consultants addressed these concerns over 10 weeks at the University of Auckland.

NZDE, our main contribution, was successfully delivered to better understand and explain deprivation in New Zealand. NZDE extends IMD18 using Tableau, incorporating data from multiple sources such as the New Zealand Census, Police, Ministry of Health, and OpenStreetMap. These data sources were transformed to a common datazone granularity level and processed using geospatial tools. The resulting dashboard offers a series of easily interpretable visualisations for exploring the underlying drivers behind deprivation, alongside vaccination rates and the locations of various amenities. Tableau was an effective tool for rapidly developing visualisations but lacked the reproducibility that programming offers.

NZDE was designed with an emphasis on usability and interpretability as reflected by the needs of the end-users. At the same time, the methodology behind NZDE serves as an example of effective data communication to the public and of how multiple sources of data can be incorporated to produce a truly useful and relevant tool. NZDE is currently in the process of being hosted online for public access in late 2021 as a supplementary tool to the existing IMD18 app. NZDE's broad accessibility means both technical and non-technical audiences can use it. We believe NZDE will help to foster new discussions around deprivation while empowering policy and decision-makers to create the change necessary to address it.

Throughout this project, we developed as statistical consultants—learning how to work with and manage a client. We were lucky to have this opportunity to develop something with real-world relevance that can help to make large-scale social change in New Zealand a reality.

Appendices

6.1 Data Preprocessing R Code

```
---
title: "Data Preprocessing"
author: "Josh Atwal and Cadence Liu"
date: "`r format(Sys.time(), '%d %B, %Y, %H:%M')`"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE)
library(pacman)
p_load(tidyverse, readxl)

# Custom-defined function for reading excel files
read_excel_data <- function(filename,
                            data.path='../data/raw/',
                            col.trunc = NULL,
                            ...) { # Additional args passed to read_xlxs
  df <- paste0(data.path, filename) %>%
          read_xlsx(..., na=c("S", "")) %>% # Suppressed values recoded as NA
          mutate_at(1, as.numeric) %>%
          drop_na(1)

  # Columns to ignore at the end of the dataframe
  if (!is.null(col.trunc)) {
    df <- df %>%
              select(all_of(1:col.trunc)) %>%
              # Regex matches dot dot digit(s) at end of column name
              rename_with(~ gsub('\\.{3}\\d{1,2}$', '', .))
  }

  return(df)

}

# Custom function for writing to file
write_to_csv <- function(df, fname){
  write.csv(df, file = paste0('../data/', fname), row.names=FALSE)
}

```


Datazone level data (from IDI)
Main data source
```{r}

imd18 <- read_excel_data('IMD2018.xlsx',
                         sheet=2) %>%
            rename(datazone2018='DZ2018')

```
```

````
imd18_long <- imd18 %>% select(datazone2018, Dec_IMD18, contains('Decile_')) %>%
                          pivot_longer(!datazone2018,names_to='IMD_measure',
                             values_to = 'Decile')

dim(imd18_long)
write_to_csv(imd18_long, 'IMD_cols.csv')

```


```{r}
# Age groups
age.grps <- read_excel_data('census_age_grps_OUTPUT.xlsx',
                            skip = 5)
dim(age.grps)
write_to_csv(age.grps, 'ages.csv')
```

```{r}
# Ethnicity
eth <- read_excel_data('ethnicity_cen2018_OUTPUT_Checked.xlsx',
                       skip = 7,
                       col.trunc = 7)

dim(eth)
write_to_csv(eth, 'ethnicity.csv')

```

```{r}
# Smoker data
smoke <- read_excel_data('smokers_cen2018_checked.xlsx',
                         skip = 5,
                         col.trunc = 5)

dim(smoke)
write_to_csv(smoke, 'smoke.csv')
```


```{r}
# Vehicle information
vehicles <- read_excel_data('vehicles_cen2018_Checked.xlsx',
                            skip = 7,
                            col.trunc = 5)

dim(vehicles)
write_to_csv(vehicles, 'vehicles.csv')
```

# Meshblock Data

## Meshblock to datazone mapping
```{r}
mb_to_dz <- read_excel_data('mb18_dz13_dz18.xlsx',
                            sheet = 2)

```
````
````

```
## Crime data
```{r}
crime <- read.csv('../data/raw/2018_crime_meshblock.csv') %>%
  group_by(Meshblock, ANZSOC.Division) %>%
  # Take sum of victimisatinos per meshblock and crime type over entire time
     range of data
  summarise(sum(Victimisations)) %>%
  ungroup() %>%
  spread(key = 2, value = 3) %>%
  # Join meshblock to datazone conversion table
  merge(mb_to_dz, by.x = 'Meshblock', by.y = 'MB18' ) %>%
  # Drop unnecessary columns
  select(-Meshblock, -DataZone2013, -Census18_Pop) %>%
  # Replace missing observations with 0
  replace(is.na(.), 0) %>%
  group_by(DataZone2018) %>%
  # Sum over datazones
  summarise_all(sum) %>%
  # Rename colummn
  rename(datazone2018 = DataZone2018)

write_to_csv(crime, 'crime.csv')
```


Census SA2 Data
SA1 to datazone mapping

```{r}
sa1 <- read_xlsx(paste0('../data/Meshblock to IMD18 conversion/', '2018sa1_
    dz2013_dz2018.xlsx'),
                    na=c("S", ""),
                    sheet=2) %>% select(-DataZone2013)

```

datazone to SA2 mapping
```{r}
# Create dataframe with SA2 <-> datazone mapping

mapping <- read.csv('../data/Meshblock to IMD18 conversion/2018FullMapping.csv')

mapping %>% distinct(SA12018_V1_00, .keep_all = TRUE) %>%
                select(SA12018_V1_00, SA22018_V1_00_NAME) %>%
                right_join(sa1, by =c('SA12018_V1_00' = 'SA1_2018')) %>%
                select(-SA12018_V1_00) %>%
                right_join(imd18, by=c('DataZone2018' = 'datazone2018')) %>%
                group_by(SA22018_V1_00_NAME) %>%
                # Use datazone populations to calculate SA2 population
                summarise(DataZone2018, Census18Pop, SA2Population=sum(Census18Pop
                    , na.rm=TRUE)) %>%
                distinct() %>%
                ungroup() %>%
                rename('Area' = 'SA22018_V1_00_NAME' ) -> dz2sa2

```

```

```


Mould
```{r}
mould <- read.csv('../data/raw/mould_dampness_income.csv')

mould.processed <- full_join(mould, dz2sa2) %>% # Join with SA2 mapping dataframe
  select(-c(Household.income, Year, Flags, Area)) %>%
  filter(Dwelling.dampness.indicator %in% c('Always damp', 'Sometimes damp', 'Not damp')) %>%
  filter(Dwelling.mould.indicator %in% c("Mould over A4 size - always", "Mould over A4 size - sometimes", "No mould/mould smaller than A4 size")) %>%
  # Recode mould indicators as numeric values
  mutate(Dwelling.mould.indicator = recode(Dwelling.mould.indicator,
                                           "Mould over A4 size - always" = 3,
                                           "Mould over A4 size - sometimes" = 2,
                                           "No mould/mould smaller than A4 size" = 1),
         # Recode dampness indicators as numeric values
         Dwelling.dampness.indicator = recode(Dwelling.dampness.indicator,
                                              'Always damp'=3,
                                              'Sometimes damp'=2,
                                              'Not damp'=1),
         # Create binary output variable
         damp.mould = (Dwelling.dampness.indicator + Dwelling.mould.indicator) > 3 ) %>%
         select(-c(Dwelling.dampness.indicator, Dwelling.mould.indicator)) %>%
         # Downsample by population
         mutate(Value = Value * Census18Pop / SA2Population) %>%
         select(-c(Census18Pop, SA2Population)) %>%
         group_by(damp.mould, DataZone2018) %>%
         summarise(sum(Value))


write.csv(mould.processed, file = '../data/mould.csv', row.names = FALSE)
```

Income
```{r}

census_income <- read.csv('../data/raw/income_dz2.csv')

income.processed <- inner_join(census_income, dz2sa2) %>% # Join with SA2 mapping dataframe
  # Remove unneeded columns
  select(-c(Year, Flags, Area, Individual.home.ownership, Age.group, Sex)) %>%
  # Remove summary rows
  filter(!(Grouped.total.personal.income %in% c('Total people - grouped total personal income', 'Total people stated', 'Not stated'))) %>%
  group_by(Grouped.total.personal.income, DataZone2018) %>%
  summarise(Value=mean(Value)) %>%
  ungroup()

# Sum over each datazone
total.counts <- income.processed %>% group_by(DataZone2018) %>%
                        summarise(total=sum(Value))

```
```

```
# Compute as proportion
income.processed <-  income.processed %>% left_join(total.counts) %>%
                              mutate(Value = Value/total) %>%
                              select(-total)


write.csv(income.processed, file = '../data/income.csv', row.names = FALSE)
```
```


## Family
```{r}
family <- read.csv('../data/raw/family.csv')

family.processed <- right_join(family, dz2sa2) %>%
  select(-c(Year, Flags, Area)) %>%
  # Downsample by population
  mutate(Value = Value * Census18Pop / SA2Population) %>%
  select(-c(Census18Pop, SA2Population))


write.csv(family.processed, file = '../data/family.csv', row.names = FALSE)
```

Benefits
```{r}
benefit <- read.csv('../data/raw/benefit.csv') %>%
                group_by(Area, Sources.of.family.income) %>%
                summarise(Value=sum(Value, na.rm = TRUE)) %>%
                rename(Income.Source = 'Sources.of.family.income') %>%
                # Recode data based on defined categories
                mutate(Income.Source = recode(Income.Source,
                                              "Interest, dividends, rent, other
                                                investments"='Independent',
                                              "Self-employment or business I own
                                                and work in"='Independent',
                                              "Wages, salary, commissions, bonuses
                                                 etc paid by my employer"='
                                                Independent',
                                              "Jobseeker Support" = 'Assistance',
                                              "New Zealand Superannuation or
                                                Veteran's Pension"= 'Assistance',
                                              "Other government benefits,
                                                government income support
                                                payments, war pensions or paid
                                                parental leave"= 'Assistance',
                                              "Other sources of income, including
                                                support payments from people who
                                                do not live in my household"= '
                                                Assistance',
                                              "Other superannuation, pensions, or
                                                annuities (other than NZ
                                                Superannuation, Veteran's Pension
                                                 or war pensions)"= 'Assistance',
                                              "Regular payments from ACC or a
                                                private work accident insurer"= '
                                                Assistance',
                                              "Sole Parent Support"= 'Assistance',
```
```

```
                                                    "Student Allowance"= 'Assistance',
                                                    "Supported Living Payment"= 'Assistance',
                                                    "No source of income during time"='No Income'
                                                    ))



benefit.processed <- right_join(benefit, dz2sa2) %>%
  select(-Area) %>%
  # Downsample by population
  mutate(Value = Value * Census18Pop / SA2Population) %>%
  select(-c(Census18Pop, SA2Population))


write.csv(benefit.processed, file = '../data/benefit.csv', row.names = FALSE)


```

Covid-19 data
```{r}
filepaths <- c('https://raw.githubusercontent.com/minhealthnz/nz-covid-data/main/vaccine-data/sa2-data/uptake_sa2_dhb_pacific_latest.csv',
               'https://raw.githubusercontent.com/minhealthnz/nz-covid-data/main/vaccine-data/sa2-data/uptake_sa2_dhb_maori_latest.csv',
               'https://raw.githubusercontent.com/minhealthnz/nz-covid-data/main/vaccine-data/sa2-data/uptake_sa2_dhb_latest.csv')

read.covid <- function(fp, eth) {
  read.csv(fp) %>%
    right_join(dz2sa2, by=c('sa2_name'='Area')) %>%
    select(-c(dhb,SA2Population), -contains('sa2_code')) %>%
    mutate(across(!sa2_name, as.numeric)) %>%
    mutate(dose1_uptake=dose1_uptake/10,
          dose2_uptake=dose2_uptake/10,
          ethnicity=eth)


}


covid <-
  rbind(read.covid(filepaths[1], 'pacific'),
        read.covid(filepaths[2], 'maori'),
        read.covid(filepaths[3], 'other'))

covid <-
  covid %>%
  group_by(DataZone2018) %>%
  summarise(dose1_uptake = 100*sum(dose1_cnt, na.rm = T)/sum(pop_cnt, na.rm = T),
            dose2_uptake = 100*sum(dose2_cnt, na.rm = T)/sum(pop_cnt, na.rm = T)) %>%
  ungroup() %>%
  mutate(ethnicity='Overall') %>%
  full_join(covid) %>%
```

```
  filter(ethnicity!='other') %>%
  select(-c(dose1_cnt, dose2_cnt, Census18Pop, sa2_name, pop_cnt))


covid <-
  covid %>%
  distinct() %>%
  pivot_wider(DataZone2018,
              names_from = ethnicity,
              values_from = c(dose1_uptake, dose2_uptake),
              values_fn = mean)

write_to_csv(covid, 'covid_all.csv')

```
```

data_preprocessing.Rmd
```

6.2 Financial Assistance

Table 6.1 defines the grouping of income types based on the sources of income specified in the NZ Census data. “Independent”, “Assistance”, and “No Income” are the levels seen within the NZDE visualisation for financial assistance.

Grouping	Sources of Income
Independent	Interest, dividends, rent, other investments
	Self-employment or business I own and work in
	Wages, salary, commissions, bonuses etc paid by my employer
Assistance	Jobseeker Support
	New Zealand Superannuation or Veteran’s Pension
	Other government benefits, government income support payments, war pensions or paid parental leave
	Other sources of income, including support payments from people who do not live in my household
	Other superannuation, pensions, or annuities (other than NZ Superannuation, Veteran’s Pension or war pensions)
	Regular payments from ACC or a private work accident insurer
	Sole Parent Support
	Student Allowance
	Supported Living Payment
No Income	No source of income during that time

Table 6.1: Definition of financial assistance.

6.3 Geospatial Data Processing

JSON Query

```

[out:json][timeout:999];
(
  // query part for:   "fast food"
  node["amenity"="fast_food"]({{bbox}});
  way["amenity"="fast_food"]({{bbox}});
  relation["amenity"="fast_food"]({{bbox}});

  // query part for:   "liquor store"
  node["shop"="alcohol"]({{bbox}});
  way["shop"="alcohol"]({{bbox}});
  relation["shop"="alcohol"]({{bbox}});

  // query part for:   pub
  node["amenity"="pub"]({{bbox}});
  way["amenity"="pub"]({{bbox}});
  relation["amenity"="pub"]({{bbox}});
  // query part for:   bar
  node["amenity"="bar"]({{bbox}});
  way["amenity"="bar"]({{bbox}});
  relation["amenity"="bar"]({{bbox}});

  // query part for:   supermarket
  node["shop"="supermarket"]({{bbox}});
  way["shop"="supermarket"]({{bbox}});
  relation["shop"="supermarket"]({{bbox}});
  // query part for:   "convenience store"
  node["shop"="convenience"]({{bbox}});
  way["shop"="convenience"]({{bbox}});
  relation["shop"="convenience"]({{bbox}});

  // query part for:   healthcare   =
```

```
  node["healthcare"]({{bbox}});
  way["healthcare"]({{bbox}});
  relation["healthcare"]({{bbox}});

  // query part for: park
  node["leisure"="park"]({{bbox}});
  way["leisure"="park"]({{bbox}});
  relation["leisure"="park"]({{bbox}});
  // query part for: playground
  node["leisure"="playground"]({{bbox}});
  way["leisure"="playground"]({{bbox}});
  relation["leisure"="playground"]({{bbox}});

  // query part for: school
  node["amenity"="school"]({{bbox}});
  way["amenity"="school"]({{bbox}});
  relation["amenity"="school"]({{bbox}});

   // query part for: geyser
  node["natural"="geyser"]({{bbox}});
);
// print results
out body;
>;
out skel qt;
```

Figures/jsonquery.txt

QGIS

File import

See **Figure 6.1**.

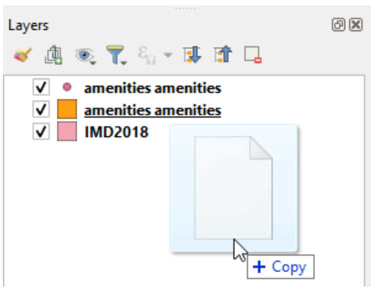


Figure 6.1: QGIS File Import—Drag and drop to "Layers".

Processing

See **Figures 6.2, 6.3, 6.4, 6.5, 6.6**

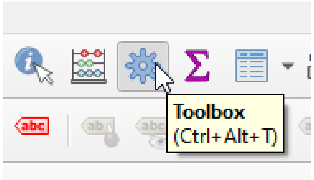


Figure 6.2: Open QGIS "Toolbox".

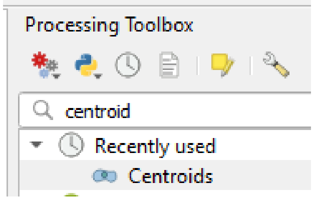


Figure 6.3: Search for function "Centroid"

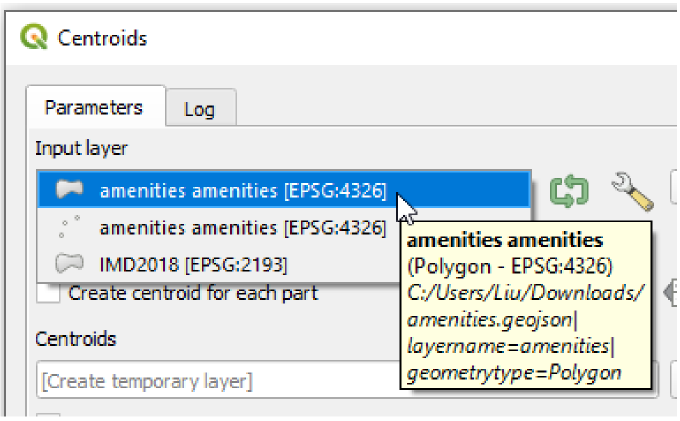


Figure 6.4: Select area-type layer and calculate centroids

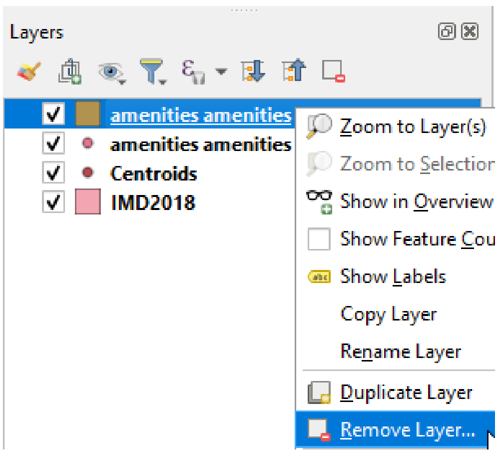


Figure 6.5: Remove redundant area-type layer from QGIS

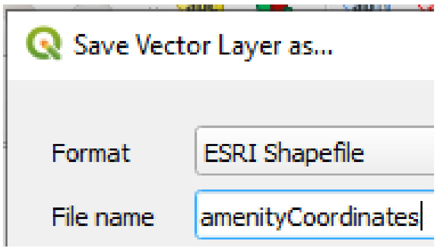


Figure 6.6: Select both coordinate-type layers and export as Shapefile

Merging Layers

See **Figures 6.7, 6.8**.

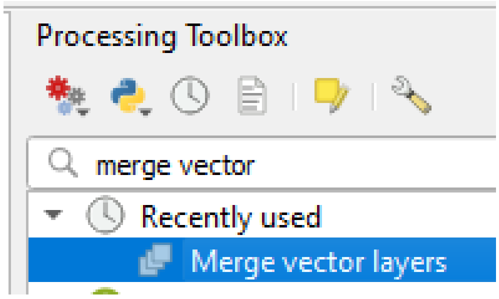


Figure 6.7: Search for function "Merge vector"

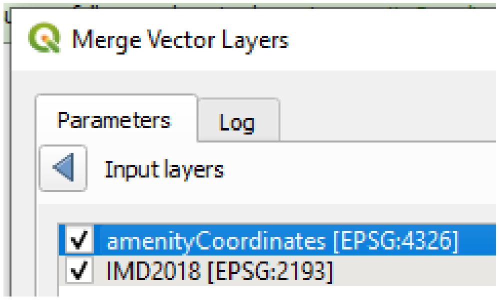


Figure 6.8: Select relevant layers to merge and proceed with "OK"

Counting Points

See **Figures 6.9, 6.10**.

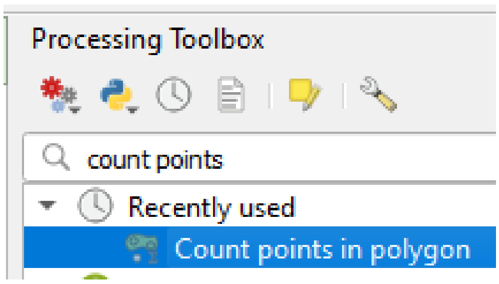


Figure 6.9: Search for function "Count points"

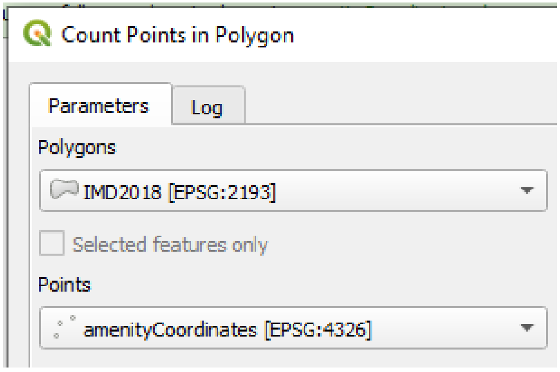


Figure 6.10: Select relevant layers to count and proceed with "OK"

Preview

See **Figure 6.11**.

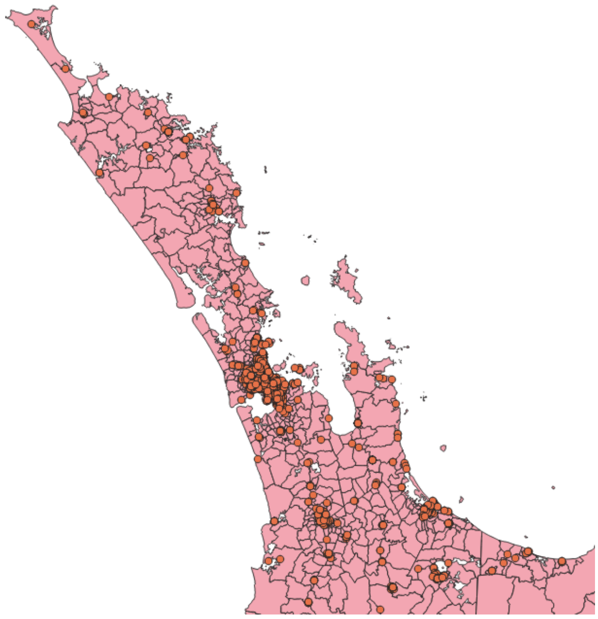


Figure 6.11: Preview of QGIS interface after following the above steps.

6.4 Calculated Fields

Aggregate Age

```
If [Aggregate Level] = 'None' THEN NULL
ELSEIF [Aggregate Level] = 'DHB' THEN {FIXED [Dhb2015 Na], [Age Group] : AVG([Age Total])}
ELSEIF  [Aggregate Level] = 'National' THEN {FIXED [Age Group] : AVG([Age Total])}
END
```

Figure 6.12: Calculated field computing the DHB or National average per age group

Family prop

```
SUM([Family.value]) / SUM({FIXED [Datazone2018] : SUM([Family.value])})
```

Figure 6.13: Calculated field converting the counts for each family type into proportions.

Map Overlay Raw

```
CASE [Dep_Vac]
when 'Deprivation' then
    CASE [View]
        when 'Dec' then [Dec IMD18]
        when 'Access' then [Decile Acc]
        when 'Crime' then [Decile Cri]
        when 'Education' then [Decile Edu]
        when 'Employment' then [Decile Emp]
        when 'Health' then [Decile Hea]
        when 'Housing' then [Decile Hou]
        when 'Income' then [Decile Inc]
    END
when 'Vaccination' then
    CASE [Dose_num]
        when '1' then
            CASE [Vac_Eth]
                when 'Overall' then [dose1 uptake Overall]
                when 'Maori' then [Dose1 Uptake Maori]
                when 'Pacific' then [Dose1 Uptake Pacific]
            END
        when '2' then
            CASE [Vac_Eth]
                when 'Overall' then [dose2 uptake Overall]
                when 'Maori' then [Dose2 Uptake Maori]
                when 'Pacific' then [Dose2 Uptake Pacific]
            END
    END
END
```

Figure 6.14: Calculated field computing the base of the map overlay based on whether deprivation or vaccination data is to be displayed.

Map Overlay

```
CASE [Overlay Complexity]
    when 'Basic' then
        CASE [Dep_Vac]
            when 'Deprivation' then 1+Round([Map Overlay Raw]/4.1)
            when 'Vaccination' then 1+Round([Map Overlay Raw]/41)
        END
    when 'Advanced' then [Map Overlay Raw]
END
```

Figure 6.15: Calculated field re-scaling the map overlay to a 1–3 scale.

References

[1] OECD. *Better Life initiative — New Zealand*. 2021. URL: https://www.oecd.org/statistics/Better-Life-Initiative-country-note-New-Zealand.pdf (visited on 10/27/2021).

[2] New Zealand Ministry of Health. *Neighbourhood deprivation*. 2021. URL: https://www.health.govt.nz/our-work/populations/maori-health/tatau-kahukura-maori-health-statistics/nga-awe-o-te-hauora-socioeconomic-determinants-health/neighbourhood-deprivation (visited on 10/27/2021).

[3] University of Auckland Faculty of Medical and Health Sciences. *Deprivation and Health Geography within NZ — 2018 New Zealand Index of Multiple Deprivation (IMD18)*. URL: https://www.fmhs.auckland.ac.nz/en/soph/about/our-departments/epidemiology-and-biostatistics/research/hgd/research-themes/imd18.html (visited on 08/12/2021).

[4] Stats NZ. *Statistical standard for geographic areas 2018*. 2017. URL: https://www.stats.govt.nz/methods/statistical-standard-for-geographic-areas-2018 (visited on 10/21/2021).

[5] Dr Daniel Exeter et al. *The 2018 New Zealand Index of Multiple Deprivation (IMD18)*. URL: https://www.fmhs.auckland.ac.nz/assets/fmhs/soph/epi/hgd/docs/Brief-report-IMD18.pdf (visited on 10/20/2021).

[6] The University of Auckland Faculty of Medical and Health Sciences. *2018 New Zealand Index of Multiple Deprivation*. URL: http://www.imd.ac.nz/NZIMD2018_Single_animation_w_logos/atlas.html (visited on 10/20/2021).

[7] The University of Auckland. *Dr Daniel John Exeter*. URL: https://unidirectory.auckland.ac.nz/profile/d-exeter (visited on 10/21/2021).

[8] Tableau. *Tableau*. 2021. URL: https://www.tableau.com/ (visited on 10/20/2021).

[9] Tableau. *Salesforce Completes Acquisition of Tableau*. 2021. URL: https://www.tableau.com/about/press-releases/2019/salesforce-completes-acquisition-tableau (visited on 10/26/2021).

[10] New Zealand Police. *Victimisation Time and Place*. 2021. URL: https://www.police.govt.nz/about-us/publications-statistics/data-and-statistics/policedatanz/victimisation-time-and-place (visited on 10/21/2021).

[11] Stats NZ. *Integrated Data Infrastructure*. 2021. URL: https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/ (visited on 10/27/2021).

[12] Stats NZ. *NZ.Stat table viewer*. 2021. URL: http://nzdotstat.stats.govt.nz/wbos/index.aspx# (visited on 10/28/2021).

[13] minhealthnz on GitHub. *nz-covid-data*. 2021. URL: https://github.com/minhealthnz/nz-covid-data/tree/main/vaccine-data/sa2-data (visited on 10/28/2021).

[14] OpenStreetMap. *About OpenStreetMap*. 2021. URL: https://wiki.openstreetmap.org/wiki/About_OpenStreetMap (visited on 10/27/2021).

[15] OpenStreetMap. *OpenStreetBrowser/Category list*. 2021. URL: https://wiki.openstreetmap.org/wiki/OpenStreetBrowser/Category_list (visited on 10/28/2021).

[16] OpenStreetMap. *Overpass turbo*. 2021. URL: https://wiki.openstreetmap.org/wiki/Overpass_turbo (visited on 10/28/2021).

[17] Hadley Wickham et al. "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.

[18] QGIS. *QGIS*. 2021. URL: https://www.qgis.org/en/site/ (visited on 10/28/2021).

[19] RStudio. *Shiny*. 2021. URL: https://shiny.rstudio.com/ (visited on 10/20/2021).

[20] Tableau. *Pricing for data people*. 2021. URL: https://www.tableau.com/pricing/ (visited on 10/23/2021).