



Environmental Health
Intelligence New Zealand
Rapu Mātauranga
Hauora mo te Taiao – Aotearoa

Analytical Standards

CITATION

Environmental Health Intelligence. 2021. Analytical Standards. Wellington: Environmental Health Intelligence NZ, Massey University.

VERSION CONTROL

The current version was published in September 2021.

AMENDMENTS LOG

Date of Amendment	Nature of Amendment	Amended by

Introduction

This Analytical Standards document is used by the Environmental Health Intelligence NZ (EHINZ) team when conducting and interpreting analyses, and presenting factsheets, reports, and other outputs. It draws on the standards used by the Ministry of Health (MoH) and Statistics New Zealand (StatsNZ) and aims to use Plain English principles.

The document is divided into three sections covering:

- 1** **Data acquisition** *page 3*
 - [1.1 Datasets and sources](#)
 - [1.2 Privacy and security](#)

- 2** **Analysis** *page 4*
 - [2.1 Data cleaning and processing \(including National Minimum Dataset\)](#)
 - [2.2 Demographic groupings](#)
 - [2.3 Other subgroups](#)
 - [2.4 Analytical toolkit](#)

- 3** **Presentation** *page 10*
 - [3.1 Data visualisation](#)
 - [3.2 Writing guide \(including font styles, Plain English, and disclaimer\)](#)

1. Data acquisition

The EHINZ team acquires data on *environmental exposures*, *health outcomes*, *population denominators*, and *geographical information* from various source agencies, and it is stored securely on the Massey server.

1.1 Datasets and sources

The full range and detail of these data sets and their data sources are stored in a spreadsheet that includes the following fields:

Data type	<ul style="list-style-type: none"> • Environmental exposure data • Health outcomes data • Population denominator data • Geographic Information System (GIS) data
Acquisition types	<ul style="list-style-type: none"> • request from the source (with a signed agreement), or • download from online portals or websites, or • information collated from published reports
Format (file type)	<ul style="list-style-type: none"> • SAS dataset, or • Excel spreadsheet, or • CSV file, or • text file, or • shapefile, geodatabase, raster, or web layers and services
Format (content)	<ul style="list-style-type: none"> • unit record level (ie, per person/event with a unique identifier), or • summarised output (ie, counts or rates)
Frequency	Mostly annual , but this can vary
Storage	File path to where the raw data is stored

1.2 Privacy and security

It is critical that all members of the EHINZ team ensure the privacy and security of the data we acquire and output.

For data with unit record level information, we will usually have a signed agreement with the source agency about the data's privacy and security. The associated electronic folder(s) storing the data may be password protected with access restricted to specific members of the EHINZ team.

Please note, we do not share unit record level data outside the EHINZ team, as per our signed agreement with the source agency when requested.

For more information about the EHINZ security procedures please refer to *A Quick Tour of the Health Information Privacy Code rules* document from

<https://www.privacy.org.nz/privacy-act-2020/codes-of-practice/hipc2020/>

2. Analysis

The EHINZ team conducts basic checks of the data we acquire and produce summary results by various demographic groupings and other subgroups.

2.1 Data cleaning and processing

The EHINZ team does not take responsibility for the quality of data we acquire from other agencies. However, all team members are required to conduct basic checks for quality, accuracy, and completeness of all data we receive, analyse, interpret, and disseminate.

Examples of this include checking:

- if the variables received are what was requested
- if the variables have sensible values and investigate any missing values
- against previous extracts, if available, to see if values for previous years are the same/similar, with justification for any differences
- that the sub-categories add up to the totals

Some data will have specific filters and/or exclusion criteria that need to be adhered to during analysis, so check any documentation (eg, data dictionary, metadata, or methods section) for details.

2.1.2 National Minimum Dataset (NMDS)

The National Minimum Dataset (NMDS) is a complex dataset requiring filters to be considered to reduce overcounting of hospitalisation events. The set of filters we have available to use include:

- transfers (within or between hospitals)
- Emergency Department short stays
- day cases
- deaths
- overseas patients
- readmissions for the same condition
- waiting list or elective cases.

APPENDIX 1 is a guide to applying these filters. This appendix also contributed to a guide for analysing hospital discharge data in the StatsNZ Integrated Data Infrastructure (IDI) within the Virtual Health Information Network (VHIN):

<https://vhin.co.nz/guides/hospital-discharge-data-in-the-idi/>

2.2 Demographic groupings

Where possible, we report by standard demographic groupings (ie, Total NZ, age, sex, ethnicity, NZDep and geographic area groupings), and set up variables in the dataset accordingly.

2.2.1 Total NZ

We report environmental exposures and health outcomes for total New Zealand.

2.2.2 Age

We analyse data and output our information by age groups consistent with the Ministry of Health's output as much as possible. Generally, it will be 5-year bands for children (0–14 years), and 10-year bands for adults (15+ years). Thus, the age groups (in years) are:

- 0–4, 5–9, 10–14, 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–84, 85+

There are variations to this grouping for some work, for example, infants (0–12 months).

Which grouping to use can depend on the data we acquire and the purpose of the analysis. We state the age groups used and ensure that the same approach is used for both the numerator and denominator data.

2.2.3 Sex

We use the term 'sex' (and not 'gender') and currently analyse the data by males and females only. Sex refers to a person's biological sexual characteristic, while gender is a person's internal sense of identity.

For more information, see the following link from StatsNZ:

www.stats.govt.nz/reports/sex-gender-and-sexual-orientation

2.2.4 Ethnicity

Ethnicity is self-perceived, and a person can affiliate with more than one ethnic group.

By default, we use a modified Level 1 classification from StatsNZ in grouping ethnicities, which is the Level 0 classification from the MoH (ie, ethnic group categories are: Māori, Pacific Peoples, Asian, and European/Other).

For more information, see the following link from MoH:

www.health.govt.nz/system/files/documents/publications/hiso_10001-2017_ethnicity_data_protocols_may-21.pdf

Two main approaches are used to analyse ethnicity data. We use 'Prioritised' ethnic groups and also 'Total response' ethnic groups. The approach we use can depend on the data that is available.

Regardless of the approach chosen, it is important to clearly state the approach used and to apply this to both the numerator and denominator data.

	Prioritised	Total response
Description	<ul style="list-style-type: none"> Every person in the dataset is assigned to ONE ethnic group. Priority of ethnic group categorisation: Māori, Pacific Peoples, Asian, European/Other. <p>For example, if a person selects Māori and European, then we categorise them as Māori.</p>	<ul style="list-style-type: none"> Every person in the dataset is counted in EACH ethnic group they report. Used when comparing Māori vs non-Māori, or Pacific Peoples vs non-Pacific Peoples.
Pros	<ul style="list-style-type: none"> Can compare between ethnic groups. The sum of the ethnic group populations will equal the total population of New Zealand. 	<ul style="list-style-type: none"> Allows better representation for those who identify with multiple ethnic groups.
Cons	<ul style="list-style-type: none"> Everyone is assigned to only one ethnic group, even if they identify with more than one ethnic group. 	<ul style="list-style-type: none"> Cannot compare between ethnic groups. The sum of the ethnic group populations will exceed the total population of New Zealand

Please note, when using ethnicity data from Census 2018, careful consideration should be taken when cross-tabulating with other demographic groups. Stats NZ has noted that the 2018 Census had a lower-than-expected response rate resulting in the introduction of new methods to produce the dataset, including using data from alternative sources. Stats NZ and the 2018 Census External Data Quality Panel (EDQP) have produced a rating system to help the users understand the quality-related issues and impacts of the 2018 Census dataset. Further information about the Stats NZ and EDQP documentation can be found in <https://www.stats.govt.nz/2018-census/data-quality-for-2018-census>

2.2.5 New Zealand Index of Deprivation (NZDep)

The University of Otago, Wellington (UOW) calculates NZDep using a set of variables from the Census to estimate the level of area deprivation for people in each small area geography in New Zealand. The calculations for the current NZDep (ie, NZDep2018) are based on Statistical Area 1 (SA1), whereas previous iterations of the NZDep were based on meshblocks. We use the NZDep ordinal scale, which divides New Zealand into tenths (deciles) based on the distribution of a continuous score.

For details about how the NZDep is derived and its recommended usage, see the following link from the University of Otago:

<https://www.otago.ac.nz/wellington/otago730394.pdf>

Decile 1 represents the areas with the least deprived NZDep scores, and Decile 10 represents the areas with the most deprived NZDep scores. Deciles are grouped in pairs, with Quintile 1 consisting of Deciles 1 and 2, and Quintile 5 consisting of Deciles 9 and 10. According to the UOW report, we should refer to 'areas that have the most deprived NZDep scores' rather than 'the most deprived areas'.

Current years of data: NZDep2001, NZDep2006, NZDep2013, NZDep2018.

We use the year closest to the year of the outcome or exposure being analysed.

Please note, health outcomes or environmental exposures by NZDep can be reported over time. However, it is not advisable to compare actual NZDep scores, deciles or quintiles for an area over time.

For further information about NZDep refer to the University of Otago website:

<http://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html>

Please note, when using NZDep data based on Census 2018, careful consideration should be taken when cross-tabulating with other demographic groups. Stats NZ has noted that the 2018 Census had a lower-than-expected response rate resulting in the introduction of new methods to produce the dataset, including using data from alternative sources. Stats NZ and the 2018 Census External Data Quality Panel (EDQP) have produced a rating system to help the users understand the quality-related issues and impacts of the 2018 Census dataset. Further information about the Stats NZ and EDQP documentation can be found in <https://www.stats.govt.nz/2018-census/data-quality-for-2018-census>

2.2.6 Geographic area groupings

Results can be produced at varying geographical levels. Geographic area groupings are based on the document 'Statistical standard for geographic areas 2018' published by Stats NZ: <http://archive.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/geographic-areas.aspx>

Please note, a summary document providing further background information, maps outlining different area boundaries and how they align with each other can be found in <https://www.stats.govt.nz/methods/statistical-standard-for-geographic-areas-2018>.

2.2.6.1 Administrative, statistical and electoral area groupings

The Statistical Standard for Geographic Areas 2018 divides geographic areas into two main groups.

Statistical geographies are defined and maintained by Stats NZ and include four levels:

- Meshblock (MB)
- Statistical Area 1 (SA1)
- Statistical Area 2 (SA2), which replaced Area Units (AU) in 2018
- Urban Rural (UR) and Urban Rural Indicator (IUR), which replaced the geography of Urban Area (UA) in 2018.

Each of the four levels is nested within the level below, and boundaries align mostly with territorial authority and regional council geographies.

Administrative and electoral boundaries were established in legislation under the Local Government Act 2002 and the Electoral Act 1993. The main areas in use by the EHINZ team are:

- Regional Council (REGC)
- Territorial Authority (TA)
- Community Board/Local Board (CB/LB)
- District Health Board (DHB)

Please note, due to changes in geographical boundaries, some data will need to be adjusted to align with these changes over time.

2.3 Other subgroups

Apart from demographic groupings (ie age, sex, ethnicity, NZDep, geographic area groupings, and Urban Rural), some indicators require data to be disaggregated by other subgroups. Some examples include:

- total energy consumed, by fuel type and sector: data is broken down by the type of fuel used (eg, oil) and by the sector consuming the fuel (eg, domestic transport sector)
- number and density of livestock in New Zealand, by type: data is broken down by the type of livestock (eg, dairy cattle)
- number of motor vehicles in the fleet, by vehicle type and fuel type: data is broken down by the type of vehicle (eg, truck) and fuel type (eg, diesel)
- monitoring sites exceeding the national environmental standard (one-hour average) for nitrogen dioxide (and similar indicators monitoring other air pollutants): data is broken by monitoring stations for the respective air pollutant.

2.4 Analytical toolkit

Key components to monitoring the environmental health of New Zealand are to analyse demographic and other subgroups of New Zealand, compare against New Zealand, and observe trends over time.

To do this, the EHINZ team uses a selection of analytical techniques to calculate common statistical estimates including **counts**, **proportions**, **means**, **rates**, and **ratios**. Our analytical methodologies draw heavily from the United Kingdom Association of Public Health Observatories (APHO) documentation and spreadsheet. Their work provides a toolkit to calculate proportions, means, crude rates and age-standardised rates based on the Poisson and Binomial distribution, along with their corresponding confidence intervals.

The APHO documentation (particularly, the APHO Technical Briefing 3 – Commonly used public health statistics and their confidence intervals) and analytical tool (specifically, the Tool for calculating common public health statistics and their confidence intervals) can be found on <https://fingertips.phe.org.uk/profile/guidance>

Age-standardisation is a key analytical technique used in our analyses, and a detailed explanation about this is found on https://www.health.govt.nz/system/files/documents/publications/standardising-rates-disease_0.pdf

APPENDIX 2 is a summary table of the common output estimates and statistical terms we use, including confidence intervals and statistical significance. A description of the key statistical tests we use are also summarised here.

3. Presentation

As part of the EHINZ branding, we ensure consistency in style, formatting and referencing of our data visualisation and writing.

3.1 Data visualisation

The EHINZ team currently uses several data visualisation tools, which include:

- Excel
- ArcGIS
- InDesign
- InstantAtlas
- Datawrapper
- Tableau

In general, tables, graphs, and maps should be able to stand alone. This means that the table/graph/map and its key components (ie, titles, headings, axes labels, legends, and footnotes) should contain all the key information needed.

3.1.1 Tables

Tables are useful when readers need precise values, or when we want to show different types of information. Tables are generally not the best way of showing patterns or trends over time – a graph may be better for this.

Example:

Year	Unintentional		Intentional		Unknown	
	Male	Female	Male	Female	Male	Female
2006	21.7	7.6	3.0	4.5	0.3	0.6
2007	18.9	6.1	3.2	3.5	0.7	0.4
2008	20.0	6.3	3.3	4.2	1.6	0.8
2009	21.2	8.5	3.3	4.6	1.3	1.2
2010	17.2	5.8	4.0	3.9	0.9	0.9
2011	17.8	7.7	3.6	3.9	0.7	0.6
2012	17.2	6.1	3.2	4.7	1.1	0.6
2013	14.7	5.9	3.2	4.2	0.8	1.1

3.1.2 Graphs

Graphs are helpful in ‘telling a story’, highlighting trends over time, and making comparisons. A well-presented graph should make the key information ‘hit you in the face’ – you should be able to grasp the key message of the graph within a split second of looking at it.

Consider using a graph when you want to illustrate:

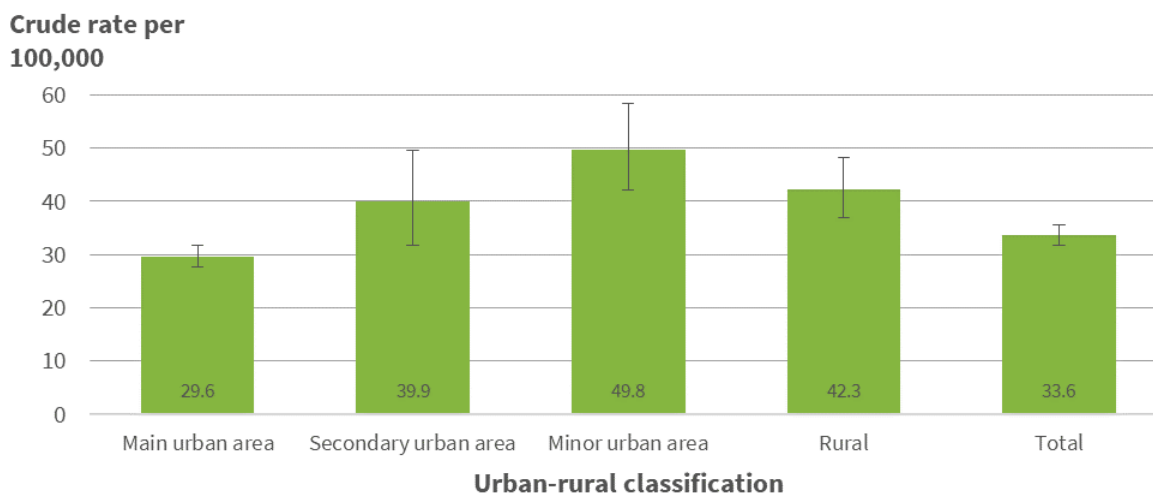
- patterns
- trends.

Types of graphs

Bar charts

- to **compare** different values (eg, age-standardised results)
- to **emphasise** individual values
- when the variable on the horizontal axis is **categorical**

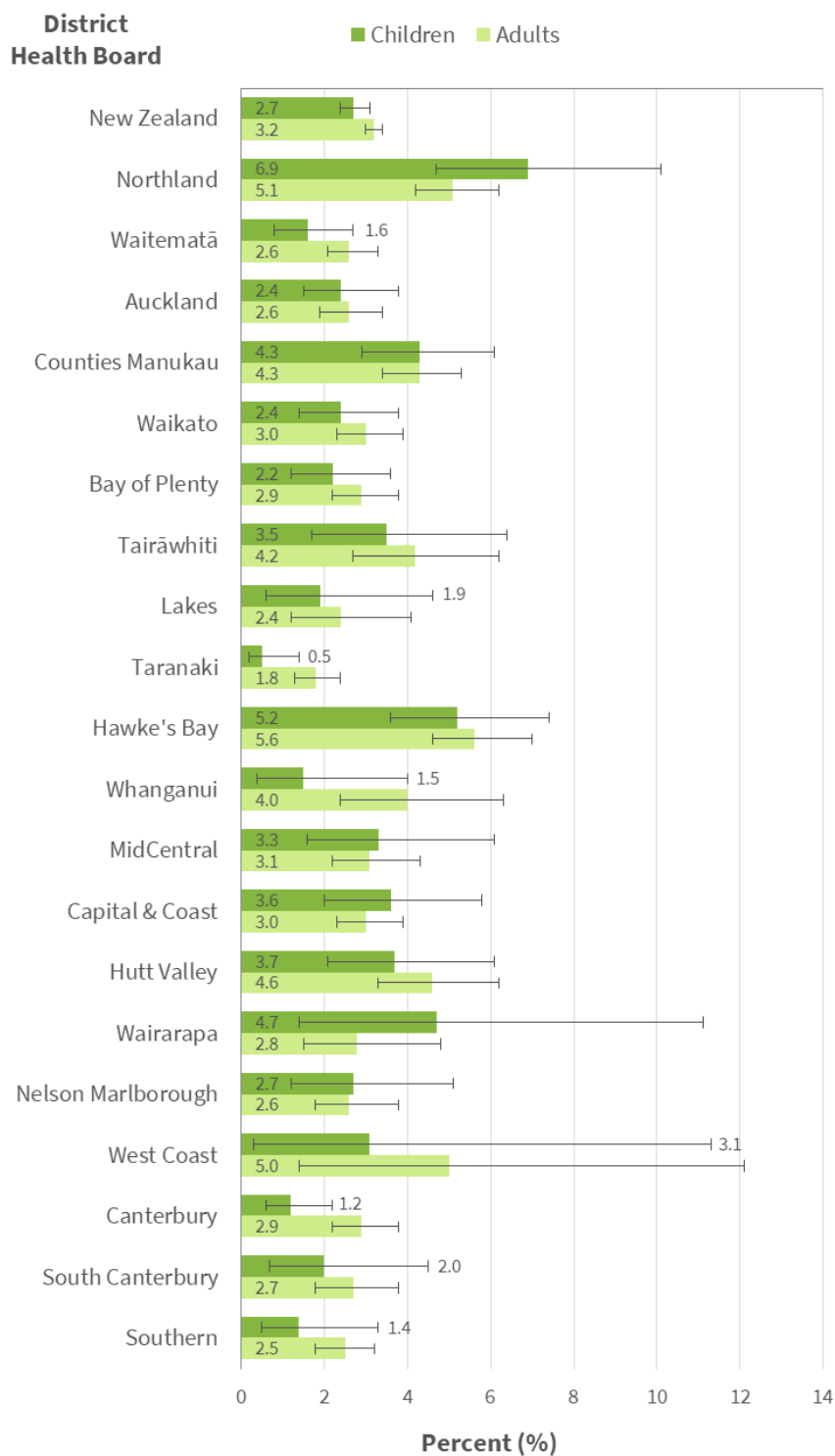
Example:



Horizontal bar charts

- when there are too many bars to fit comfortably from left to right
- when the labels for the bars are very long

Example:



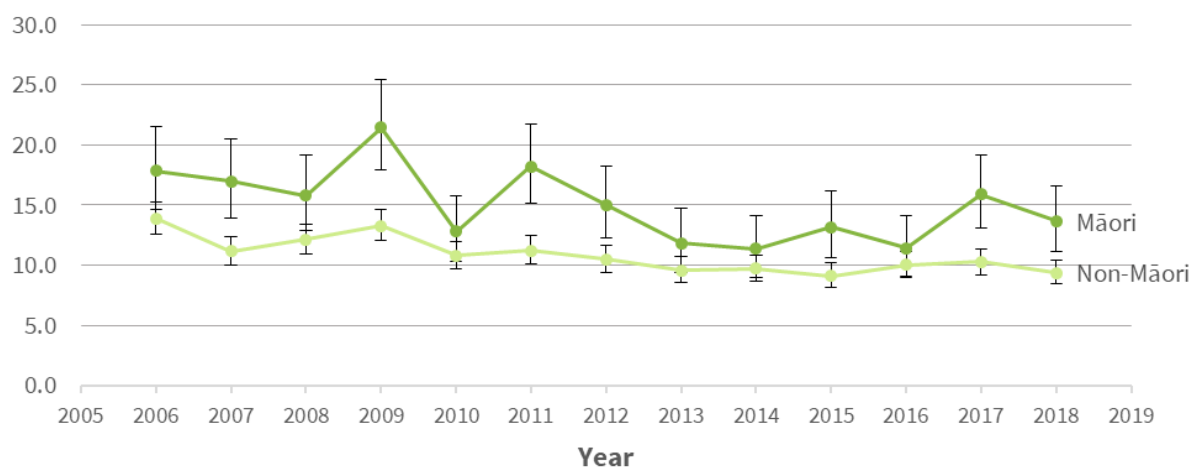
Line graphs

- to show the **trend** of the values
- to show the **shape of change** from one value to the next

Note: Use continuous or interval variables on the x-axis, like time (eg, year or month) and age (eg, age in years or months, age in equal sized groups), never discrete values (eg, region).

Example:

**Age-standardised
rate per 100,000**



EHINZ – Analytical Standards

Small multiples

- to show a series of graphs or maps, arranged in a grid – for comparison

Note: All the graphs need to have the same scale for x and y axis.

Example:



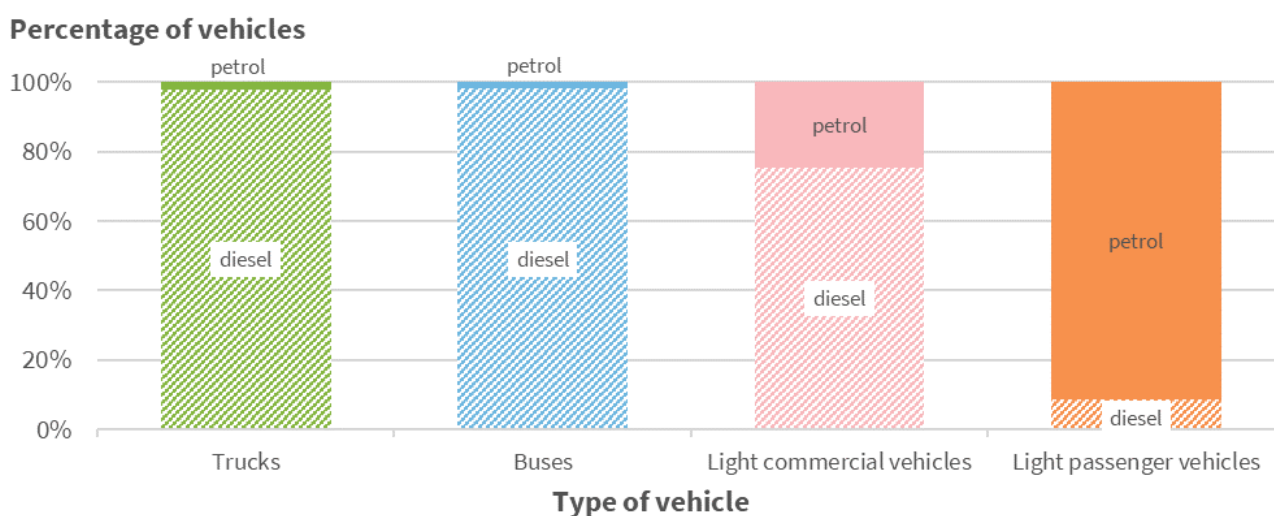
Stacked bar chart

- to show **proportions** of a whole
- useful when the focus is to **compare the totals** as well as provide additional information about the **composition of each bar**

Notes:

- present 'parts of a whole' type data in a bar chart first.
- comparisons between the different categories can be difficult in a stacked bar chart. Add data labels to help the reader interpret the size of the categories.

Example:



3.1.3 Maps

Maps are useful when showing spatial patterns in data that might not be obvious in tables or graphs. They allow readers to look up data in specific locations and enable comparison of data across geographical regions.

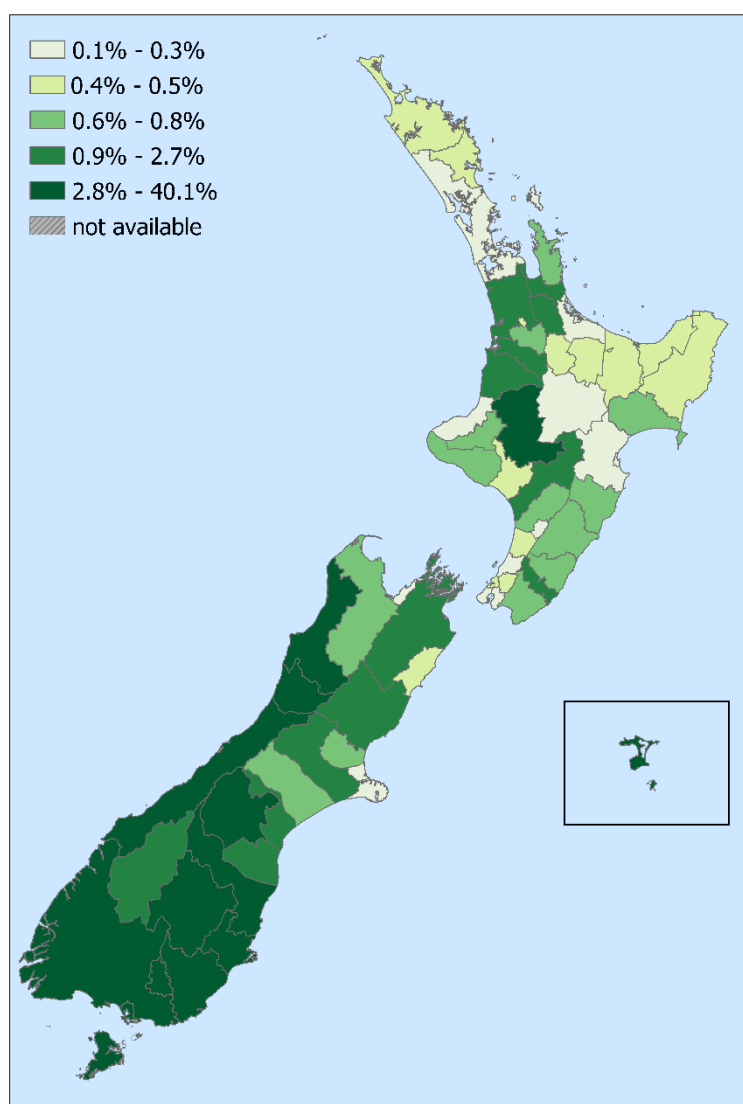
Consider using a map when you want to:

- show the **location and spatial distribution** of data
- **compare different areas** and show clear **spatial patterns**

Notes:

- the EHINZ team has developed its own map template.
- by default, quintiles are to be used for classification. However, some data needs to be normalised. For example, counts of livestock per TA are replaced by density (number of livestock per km²) to be comparable between areas of different size.
- zero or null values are to be appropriately labelled in order to avoid confusion between data that is not available, suppressed or an actual zero value.

Example:



For more information on selecting and designing tables and graphs, have a look at these resources:

- Show me the Numbers book by Stephen Few
- Perceptual Edge website www.perceptualedge.com

3.2 Writing guide

The *Communication Standards for the Ministry of Health* is a core reference document for the EHINZ team. It is the basis for our style of communication, certain grammatical preferences, and referencing style.

APPENDIX 3 is a copy of the Ministry of Health Communication Standards.

3.2.1 Font styles

We use **Arial** font for all documents and presentations, except factsheets. The font for factsheets, EHINZ and Healthspace websites is **Source Sans Pro**.

3.2.2 Plain English

The EHINZ team adhere to Plain English principles. Therefore, irrespective of the method of dissemination, authors need to:

1. identify their intended audience
2. keep words simple
3. keep sentences and paragraphs short and focussed
4. put the most crucial information at the beginning
5. have clear and succinct rationale
6. proofread their work.

Some helpful resources to help you understand and apply plain English principles are:

- *How to tell a story using statistics* (Statistics New Zealand)
- *Your guide to our Plain English Standard* (Statistics New Zealand).

APPENDIX 4A is a copy of How to tell a story using statistics.

APPENDIX 4B is a copy of Your guide to our Plain English Standard.

3.2.3 Disclaimer

All EHINZ publications, including the EHINZ and Healthspace websites, contain the following disclaimer:

Environmental Health Intelligence NZ – Rapu Mātauranga Hauora mo te Taiao - Aotearoa, makes no warranty, express or implied, nor assumes any legal liability or responsibility for the accuracy, correctness, completeness or use of any information that is available on or through this [website]/ [report]/ [factsheet].