Massey University

158741

Location Data: Mapping, Analysis and Visualisation

At-Home Activity 1: Choropleth Mapping

Luis Vieira

23012096

March, 2024

Luis Vieira [23012096]

At-Home Activity 1: Choropleth Mapping

Tasks:

The Topic 2 tutorial involves producing a choropleth map of census population data, and population density data for meshblocks[1] in New Zealand.

Select another data set of a different kind (see Topic 1 on Stream for a list of geospatial data resources, or use some other data set), and select an attribute to map.

Generate a non-normalised version of a choropleth map, and then generate two other choropleth maps for the same attribute, using the field calculator (try to use a different type of calculation this time), then address the following questions in up to 500 words:

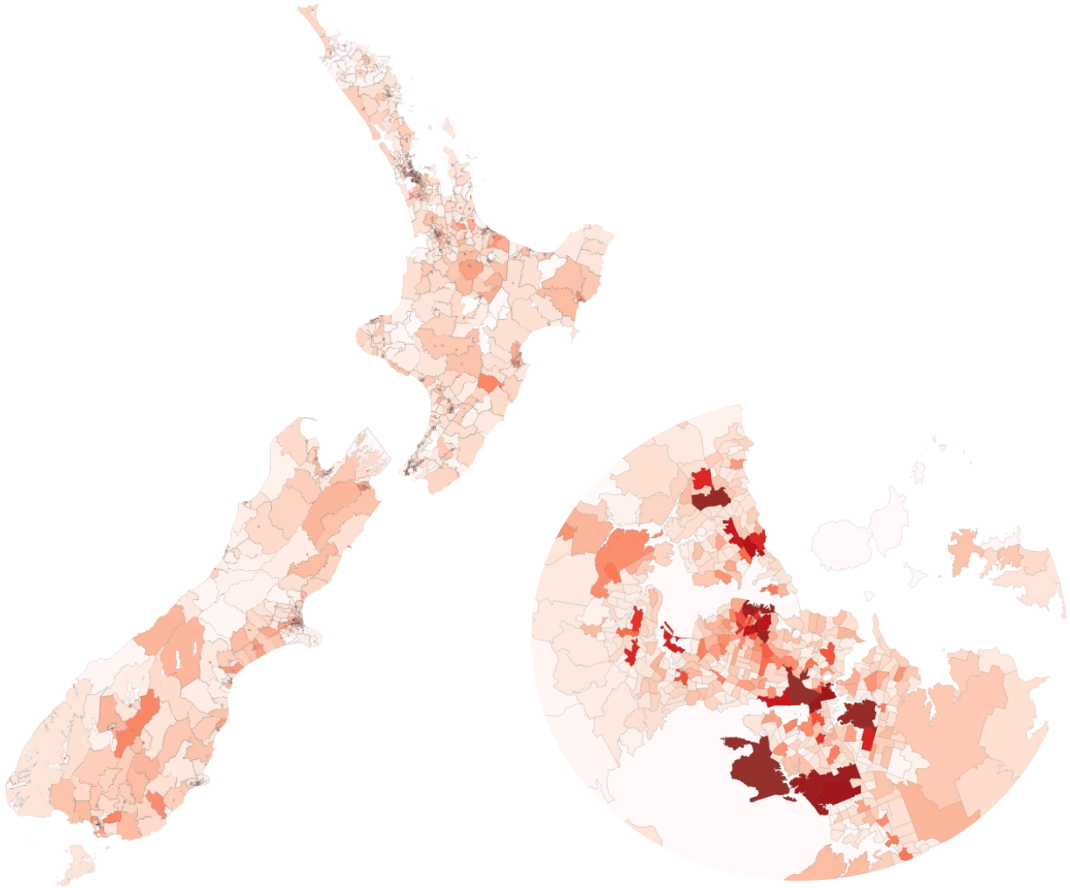
a. Which attribute did you choose to map and why?

The attribute selected was New Zealand Total Industry - Total employment count ("Total_empl") per meshblock, for year 2018 (Statistics New Zealand, 2021). The reason for the choice was that this attribute can provide a strong indication of socio-economic activity and may reveal disparities among regions. It can potentially aid policymakers, as well as businesses targeting areas for investment or development.

[1] Stats NZ defines meshblock as the smallest geographic unit, used for statistical data reporting, and covers all New Zealand land and marine areas (Statistics New Zealand, 2016).

Figure 1

Non-normalised choropleth maps of New Zealand and Auckland region (partial) of total employee count 2018



Note. Maps generated using QGIS 3.36. Map shows non-normalised "Total Industry - Total employee count" per meshblock. Darker colour indicate higher employee counts. NZ Stats dataset (Statistics New Zealand, 2021).

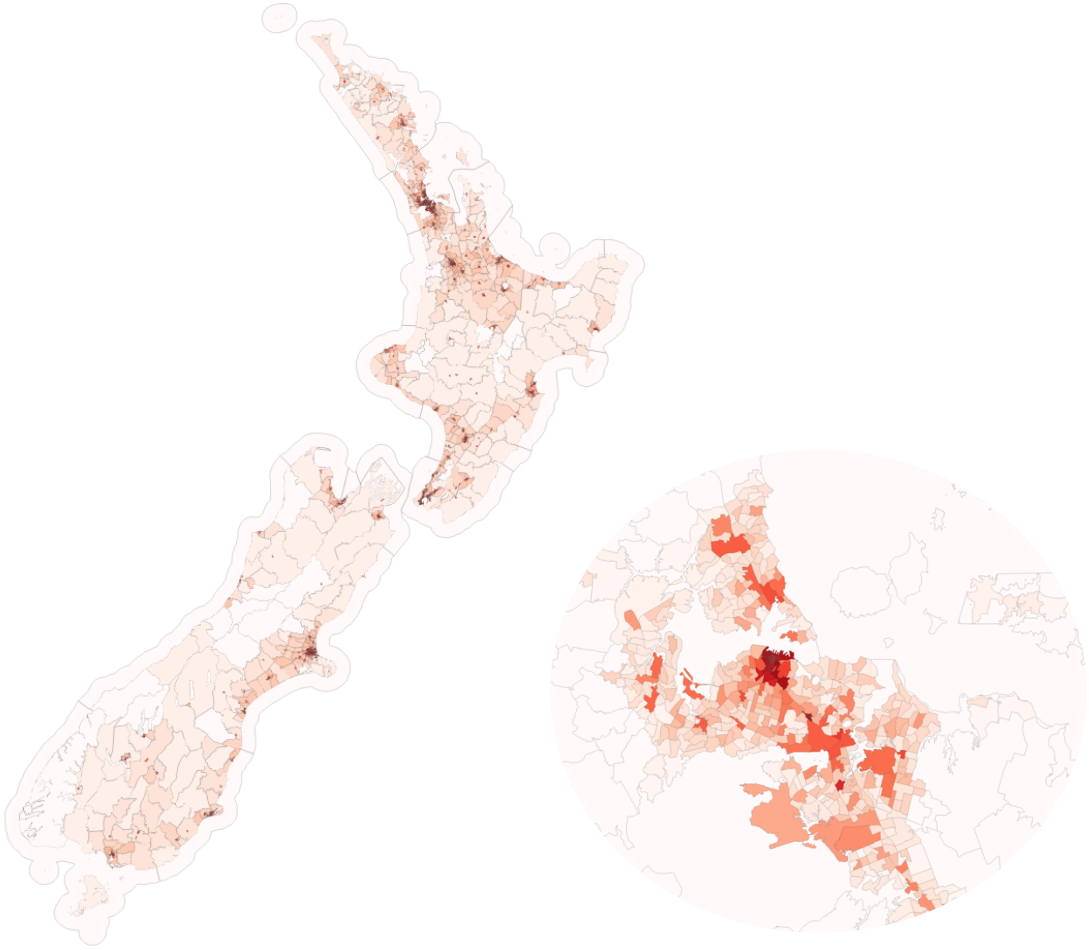
b. How did you calculate the values for the normalised versions of the map?

- Employee density:
  Expression used in QGIS: coalesce("Total_empl", 0)/coalesce("AREA_SQ_KM", 1)

In order to calculate the normalised values of employee count density, the "Total Industry – Total employee count" was divided by the same meshblock area (in km^2), taking into account potential NULL and zero values.

Figure 2

Choropleth maps of New Zealand and Auckland region (partial) of total employee count 2018 by total area (km^2)



Note. Maps generated using QGIS 3.36. Map shows normalised "Total Industry - Total employee count" by "Square kilometres of total area" per meshblock. Darker colour indicate higher density of employee for the area. NZ Stats dataset (Statistics New Zealand, 2021).

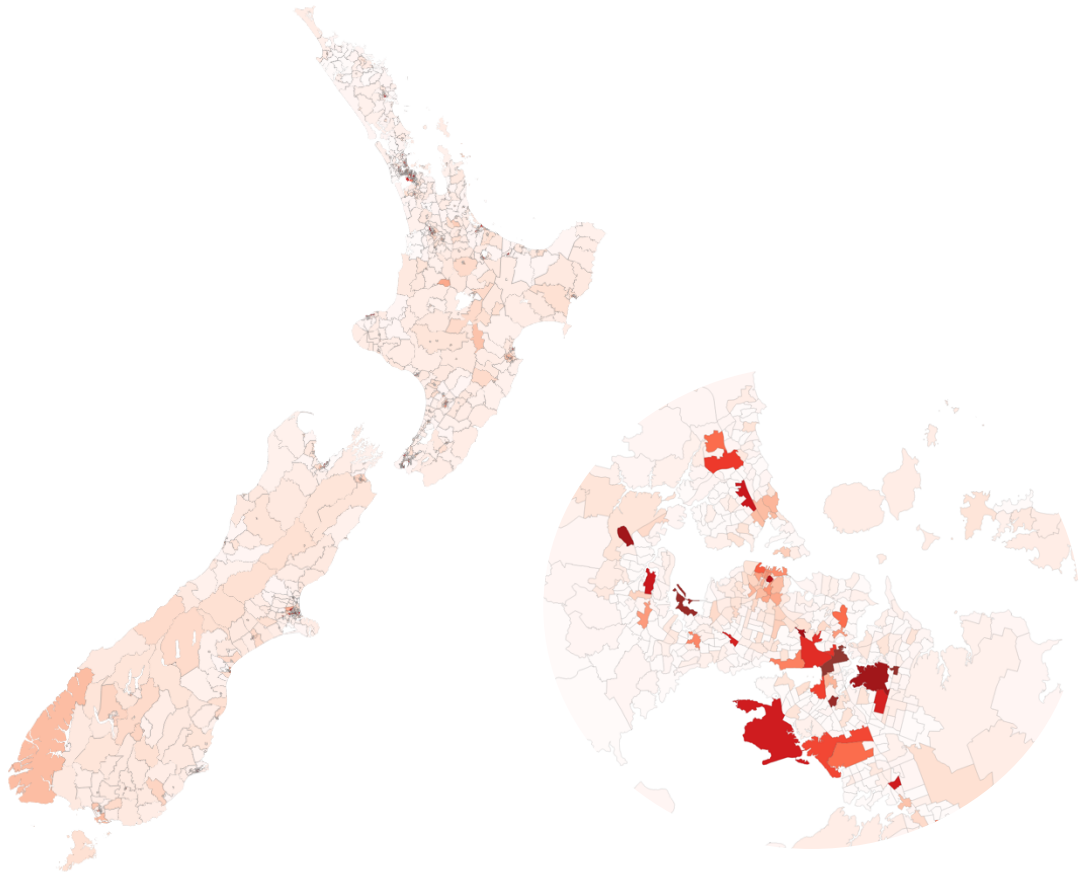
- Employee per capita:
  Expression used in QGIS: coalesce("Total_empl", 0)/nullif(coalesce("erp18", 0), 0)

  To obtain the employee count by resident population, it was necessary aggregate two datasets/layers. The total employee count was divided by the same meshblock estimated resident population. The latter was retrieved from "*Estimated Resident Population at 30 June 2018 by Statistical Area 2*" (Statistics New Zealand, 2020).

  The merging process and minor data leakage is documented in the *Assumptions & Data Processing* section.

Figure 3

Choropleth maps of New Zealand and Auckland region (partial) of total employee count 2018 by total area population



Note. Maps generated using QGIS 3.36. Map shows normalised "Total Industry - Total employee count" by "Estimated Resident Population" per meshblock. Darker colour

indicate higher ratio of employees to residents. NZ Stats dataset (Statistics New Zealand, 2020; 2021).

c. Why did you select those normalised versions, and what differences can you see between the three maps?

The selection of these normalised versions focuses on highlighting nuances and patterns in the distribution and concentration of employment throughout New Zealand.

On one hand, having the employee count density allows to normalise the visualisation to account for an area dimension, so large areas are not overrepresented and vice versa.

On the other hand, normalising the employee count by resident population can reveal interesting insights into potential employment opportunities, areas where employment count is outpacing resident population.

When comparing the maps, there are clear distinctions in particular the rural areas between non-normalised and normalised maps. In the non-normalised it is skewed towards larger areas, which tend to be over-represented. In the employee density map it focuses on the spread of employment across an area, more noticeable in Figure 2 the clustering of employment around larger business cities.

The employee count by residential population gives a good distinction between residential areas in lighter colour and business centres in darker colour in Figure 3, whereas in between we see rural areas with a better balance between population and employment. This contrast between maps is more evident when looking at the Auckland map.

d. Which of the three maps is most informative, and why?

The choice of map depends on the specific goal of our analysis. Should the focus be on illustrating the socio-economic activity, then employee density is the most informative map as it is able to highlight socio-economic clusters independently of population figures.

In contrast, if the goal is to assess employment availability relative to population, employment per capita displays areas' differences in job access and labour market health.

Each map provides unique insights, aiding policy makers' strategic planning, companies deciding on a new location, or even job seekers identifying areas with more job abundance.

However, it is crucial to recognise any map has limitations, and no single map is able to provide complete information. In conclusion, the best map is dictated by the context and objective of the analysis.

References

Jenks, G. F. (1967). *The data model concept in statistical mapping*. International Yearbook of Cartography, 7, 186–190.

Statistics New Zealand. (2016). *Statistical standard for meshblock*. Available from https://www.stats.govt.nz/

Statistics New Zealand. (2020). *Estimated Resident Population at 30 June 2018 by Statistical Area 2* [Data set]. Stats NZ. Retrieved March 8, 2024, https://datafinder.stats.govt.nz/layer/105008-estimated-resident-population-at-30-june-2018-by-statistical-area-2/

Statistics New Zealand. (2021). *New Zealand business demography statistics at February 2020 (on statistical area 2 2020)* [Data set]. Stats NZ. Retrieved March 8, 2024, https://datafinder.stats.govt.nz/layer/105388-new-zealand-business-demography-statistics-at-february-2020-on-statistical-area-2-2020/

APPENDIX

Assumptions & Data Processing

> "*Values of -999 indicates data are not available. (…)*
>
> *Individual figures may not sum to stated totals due to rounding.*" (Statistics New Zealand, 2020; 2021).

Data set covers New Zealand statistical area 2 (SA2), and refers to data from 2018.

NULL values were not removed nor imputed, but also not considered.

Estimated resident population (ERP) at 30 June 2018 was aggregated into the business demography statistics.

The merging process using "Join attributes by field value" of the layers based on key column "SA22020_V1", led to data leakage. Among the unmerged data The Inlet Tauranga Harbour South, Inlets Far North District, Marlborough Sounds Coastal Marine, Inlet Port Lyttelton and Inlets other Southland District, had employee counts of 60, 45, 20, 12 and 3 respectively. However, since these meshblocks did not have an estimated resident population data.

Jenks natural breaks classification was the preferred method, due to the large variance in the data, with the exception of the NZ Employee count density to provide a clearer map contrast (Jenks, 1967).