Lab_1

ESDA of Deprivation in New Zealand

Name: Wenjuan Wang

Student ID:59612505

Date: 22/03/2025

Content

1. Introduction

This report employs Exploratory Spatial Data Analysis (ESDA) techniques to examine deprivation patterns across New Zealand at three geographic scales: Meshblock, Area Unit (AU), and Territorial Authority (TA), with a focus on two major cities: Wellington and Christchurch. Based on the analysis results and conclusions, it advocates for interventions to improve the situation of deprivation.

1.1 Dataset

- New Zealand Deprivation Index 2006 (spatial shapefile)
- Census information of New Zealand by Meshblock (Excel file)

1.2 Software

- Microsoft Excel (data preparation)
- ArcGIS Pro (spatial operations)
- GeoDa (statistical/spatial analysis)

2. Data Preparation

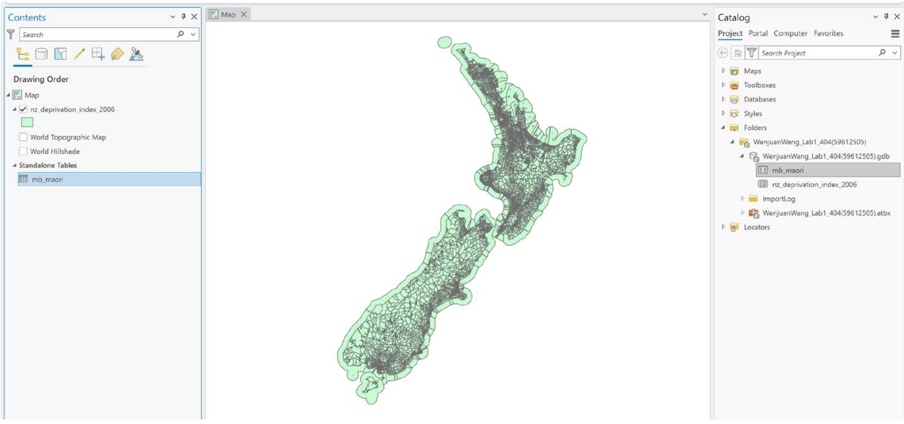
2.1 Excel Processing

- Add new sheet *'mb_maori'*.
- Add and fill two columns, *'mbcode' and 'maori'*.
- Replace null entries '..C' with -999 to prepare for filtering

  The 'maori' column contains many '..C' values, indicating no entries. However, all values should be numeric so that we can perform attribute filtering in ArcGIS Pro. Therefore, we replace all instances of 'C' with a negative number, '-999', to distinguish these 'null values' from actual values greater than or equal to zero. A total of 7,295 values are replaced.

2.2 ArcGIS Pro Operations

- Import shapefile and Excel data.



Picture_1 Import the shapefile and Excel file into ArcGIS

- Join layers using the common attribute (mbcode/MB06)

nz_deprivation_index_2006

Field: Add Calculate Selection: Select By Attributes Zoom To Switch Clear Delete Copy

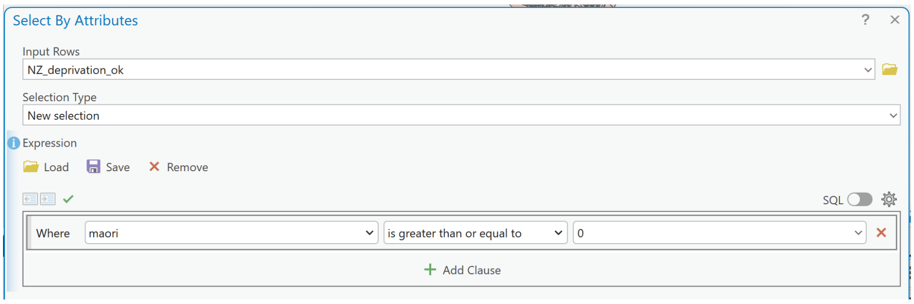
	OBJECTID *	Shape *	MB06 *	mbcode *	maori	TA06	WARD06	CB06	AU06	UA06	GED02	MED02	TA06V2
1	1	Polygon	1608300	1608300	3	033	03303	03304	552702	502	050	4	033
2	2	Polygon	1333001	1333001	-999	025	02504	02503	542550	502	048	7	025
3	3	Polygon	1020700	1020700	9	020	02001	02099	531731	502	050	2	020
4	4	Polygon	1416000	1416000	36	030	03001	03001	545911	502	029	1	030
5	5	Polygon	2224906	2224906	-999	048	04801	04899	578402	502	056	1	048
6	6	Polygon	2287502	2287502	-999	053	05303	05399	580444	502	021	6	053
7	7	Polygon	2417701	2417701	3	057	05703	05799	585321	502	059	6	057
8	8	Polygon	0059400	0059400	-999	002	00203	00299	501814	502	035	5	002
9	9	Polygon	1552701	1552701	9	033	03303	03301	551112	502	032	4	033
10	10	Polygon	1048500	1048500	-999	011	01101	01101	619301	507	009	2	011
11	11	Polygon	1110000	1110000	0	015	01502	01502	534604	502	041	2	015
12	12	Polygon	2755700	2755700	-999	063	06301	06399	597720	502	043	6	063

Picture_2 The attribute table after joining

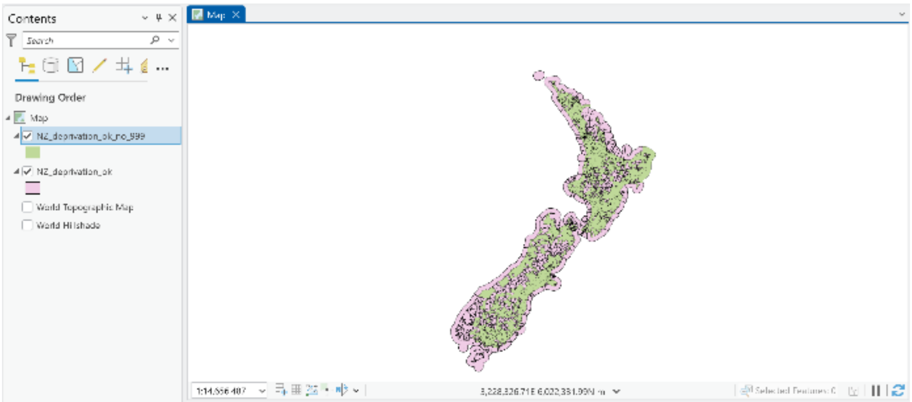
2.3 Data Cleaning

- Use the 'Select by Attributes' tool in ArcGIS Pro to exclude -999 (null) values, reducing the number of records from 42,946 to 34,702.

  From the previous steps, we know that the '-999' values in the 'NZ_deprivation_ok' layer represent 'null'. Therefore, in this step, we aim to eliminate the '-999' values using the 'Select by Attributes' function in ArcGIS, retaining only the meaningful values.



Picture_3 'Select by Attributes' tool in ArcGIS Pro



Picture_4 The Meshblock shapefile after eliminating -999 values

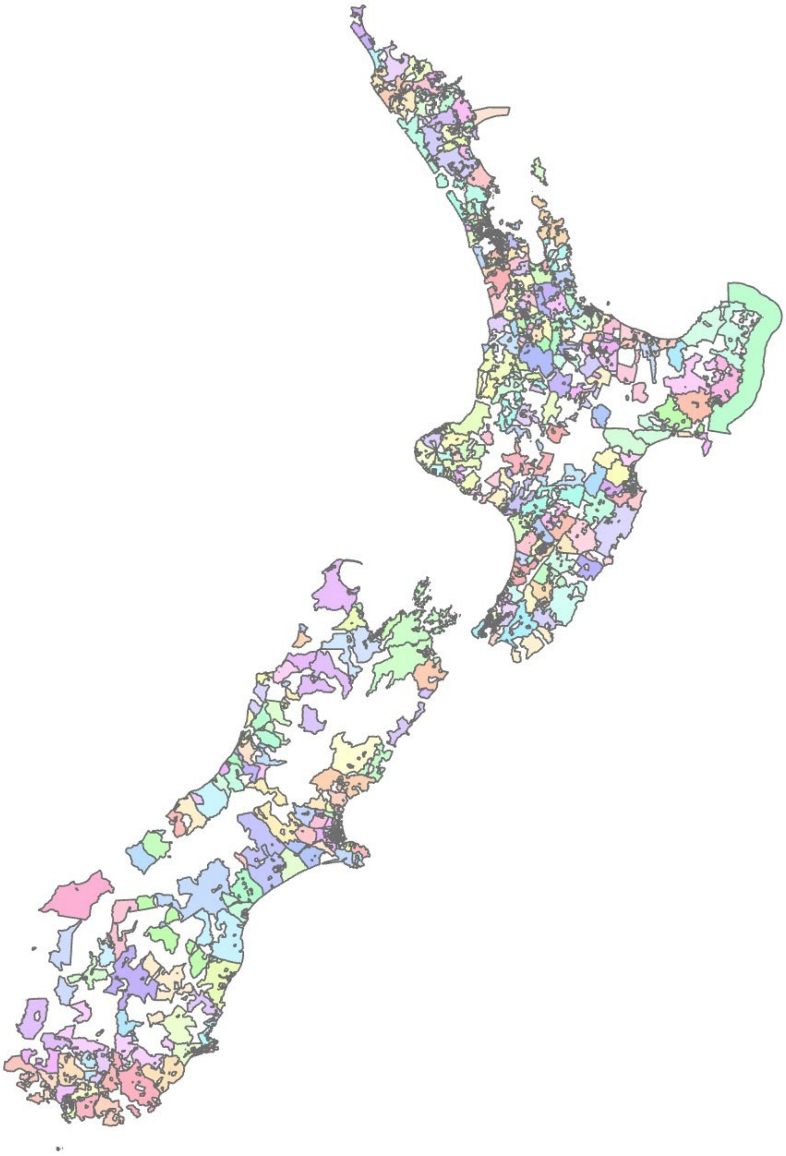
- **Question: Why are there null entries?**
  One reason is that in the New Zealand census if the number of people is very small (usually fewer than 3-5), the actual number is marked as 'Null' to protect privacy. In some records, the total number of people is not 0, but the populations of all the groups are marked as 'Null'; this is usually the case (in the picture below). Another reason is that the data is not accessible, or there could be a loss of information during data processing.

2006 Census, Ethnic Groups (Grouped Total Responses), for the Census Usually Resident Population Count						
European Ethnic Groups	Māori Ethnic Group	Pacific Peoples' Ethnic Groups	Asian Ethnic Groups	MELAA Ethnic Groups	Other Ethnic Groups	Total People
..C	..C	..C	..C	..C	..C	3
..C	..C	..C	..C	..C	..C	12
21	15	0	0	0	0	33

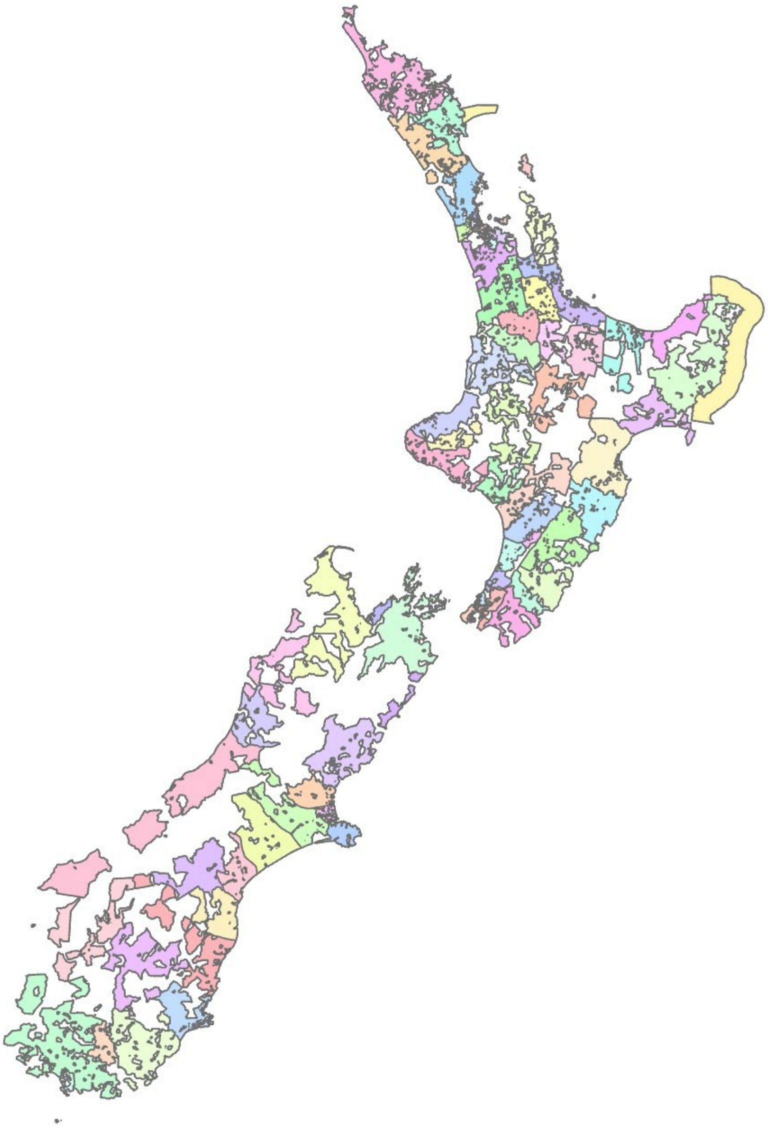
Picture_5 The total population is not null, but groups are all null

2.4 Dissolving for AU and TA Levels

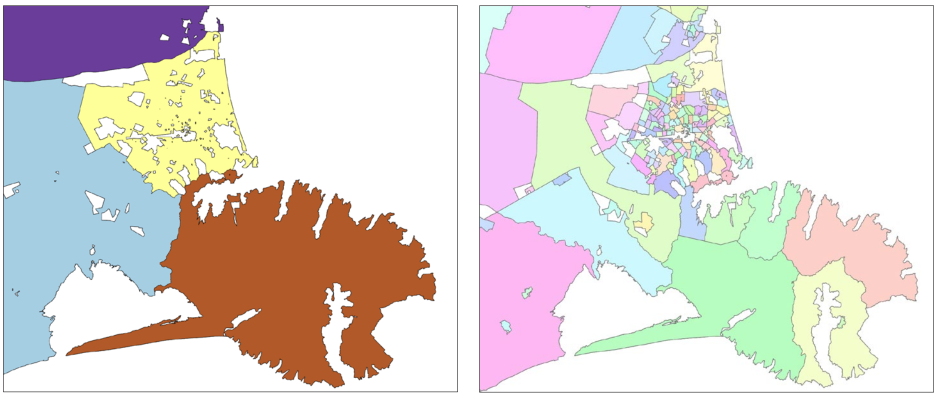
- Created Area Unit and Territorial Authority layers using 'AU06' and 'TA06' and Dissolve function.



Picture_6 The results dissolved based on Area Unit



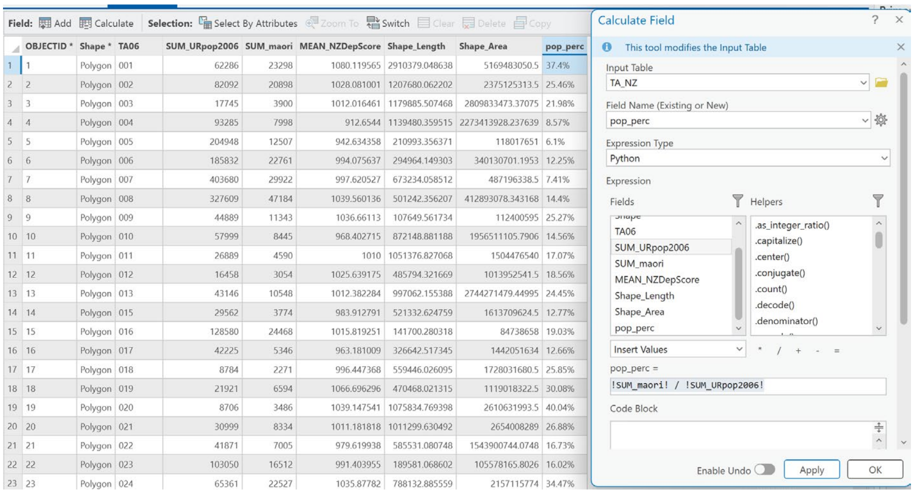
Picture_7 The results dissolved based on Territory Authority



Picture_8 Comparison of the same place between AU and TA

2.5 Calculate Maori Percentage

- Created the Māori%' field for Meshblock, AU, and TA by dividing the Māori population by the total population.

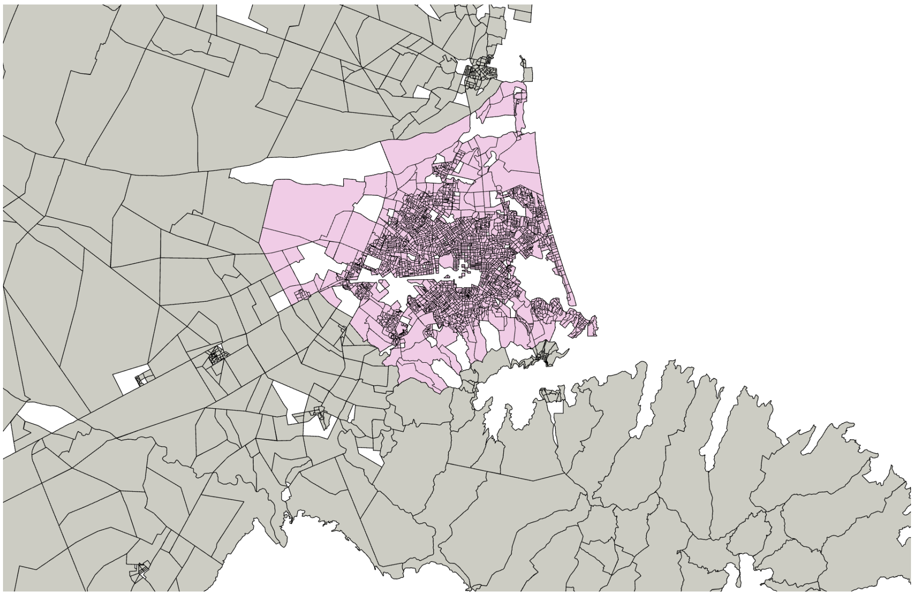


	OBJECTID *	Shape *	TA06	SUM_URpop2006	SUM_maori	MEAN_NZDepScore	Shape_Length	Shape_Area	pop_perc
1	1	Polygon	001	62286	23298	1080.119565	2910379.048638	5169483050.5	37.4%
2	2	Polygon	002	82092	20898	1028.081001	1207680.062202	2375125313.5	25.46%
3	3	Polygon	003	17745	3900	1012.016461	1179885.507468	2809833473.37075	21.98%
4	4	Polygon	004	93285	7998	912.6544	1139480.359515	2273413928.237639	8.57%
5	5	Polygon	005	204948	12507	942.634358	210993.356371	118017651	6.1%
6	6	Polygon	006	185832	22761	994.075637	294964.149303	340130701.1953	12.25%
7	7	Polygon	007	403680	29922	997.620527	673234.058512	487196338.5	7.41%
8	8	Polygon	008	327609	47184	1039.560136	501242.356207	412893078.343168	14.4%
9	9	Polygon	009	44889	11343	1036.66113	107649.561734	112400595	25.27%
10	10	Polygon	010	57999	8445	968.402715	872148.881188	1956511105.7906	14.56%
11	11	Polygon	011	26889	4590	1010	1051376.827068	1504476540	17.07%
12	12	Polygon	012	16458	3054	1025.639175	485794.321669	1013952541.5	18.56%
13	13	Polygon	013	43146	10548	1012.382284	997062.155388	2744271479.44995	24.45%
14	14	Polygon	015	29562	3774	983.912791	521332.624759	1613709624.5	12.77%
15	15	Polygon	016	128580	24468	1015.819251	141700.280318	84738658	19.03%
16	16	Polygon	017	42225	5346	963.181009	326642.517345	1442051634	12.66%
17	17	Polygon	018	8784	2271	996.447368	559446.026095	1728031680.5	25.85%
18	18	Polygon	019	21921	6594	1066.696296	470468.021315	1119018322.5	30.08%
19	19	Polygon	020	8706	3486	1039.147541	1075834.769398	2610631993.5	40.04%
20	20	Polygon	021	30999	8334	1011.181818	1011299.630492	2654008289	26.88%
21	21	Polygon	022	41871	7005	979.619938	585531.080748	1543900744.0748	16.73%
22	22	Polygon	023	103050	16512	991.403955	189581.068602	105578165.8026	16.02%
23	23	Polygon	024	65361	22527	1035.87782	788132.885559	2157115774	34.47%

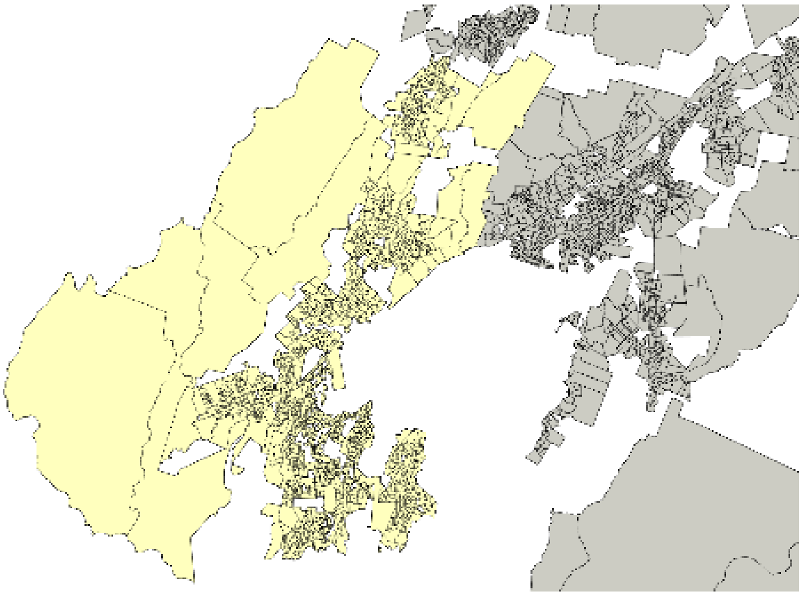
Picture_9 Calculating Maori percentage in the attribute table

2.6 City Subsets

- Extract Wellington and Christchurch Meshblock using attribute queries



Picture_10 Extract Christchurch Meshblock from Meshblock_NZ

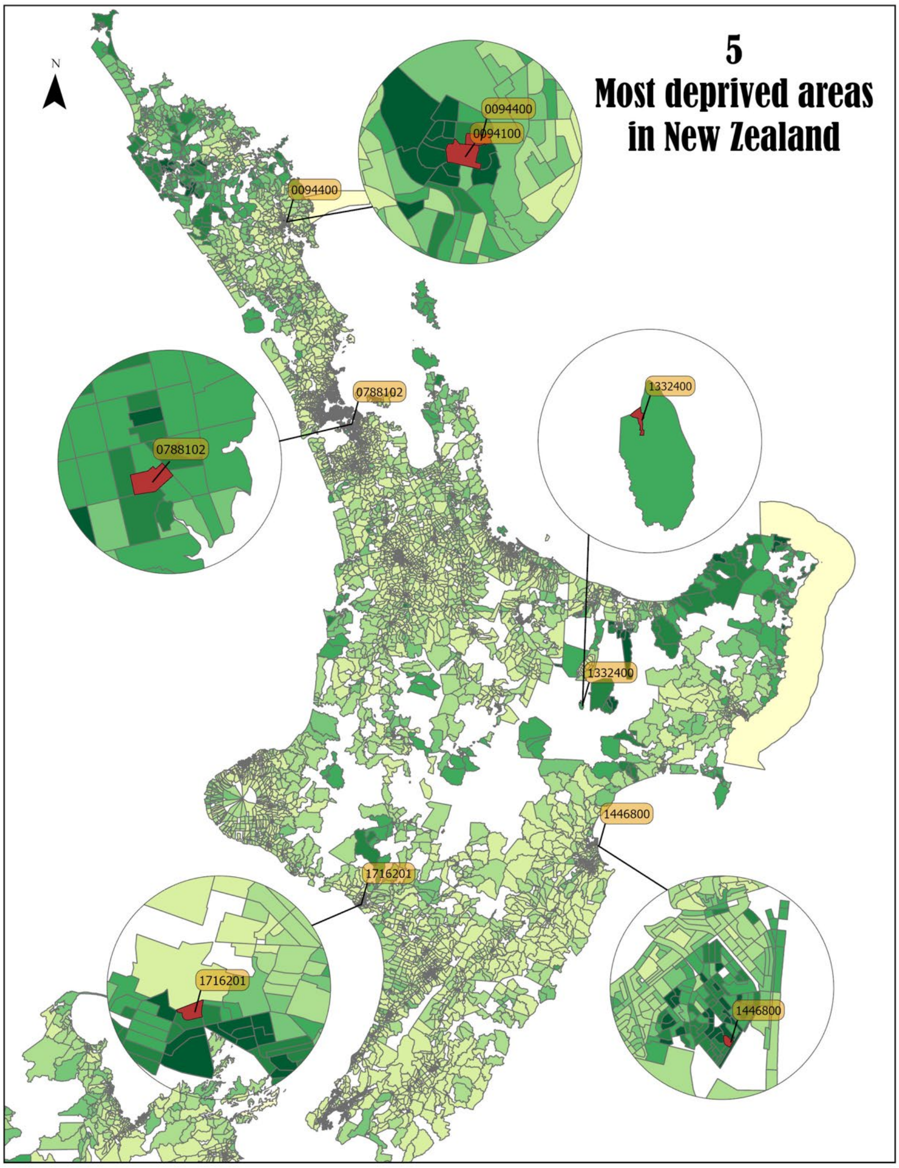


Picture_11 Extract Wellington Meshblock from Meshblock_NZ

3. Deprivation Analysis

3.1 Top 5 Most Deprived Areas

- <u>Identify and visualize based on the highest NZDep scores</u>
  1) Sort the NZDep scores from high to low.
  2) Then, select the first 6 areas (since the fifth value equals the sixth value) with the highest deprivation in New Zealand.
  3) Symbolize them with a red color to highlight, along with a label to indicate the Mbcode of a specific area.
  4) From the NZDep scores histogram, we can see that the data is left-skewed. Therefore, we use a natural break classification method in ArcGIS to concentrate the 7 class breaks more on the lower values.
  5) Utilize a green sequential color scheme to symbolize the polygons according to the NZDep Score, with lower values represented by lighter colors.
  6) Finally, create a separate map frame and zoom in to each mesh block unit listed as top 5, providing a detailed location display. (The 5 areas are too small to see on the North Island map, so I added additional 5 map frames to showcase them in detail.)

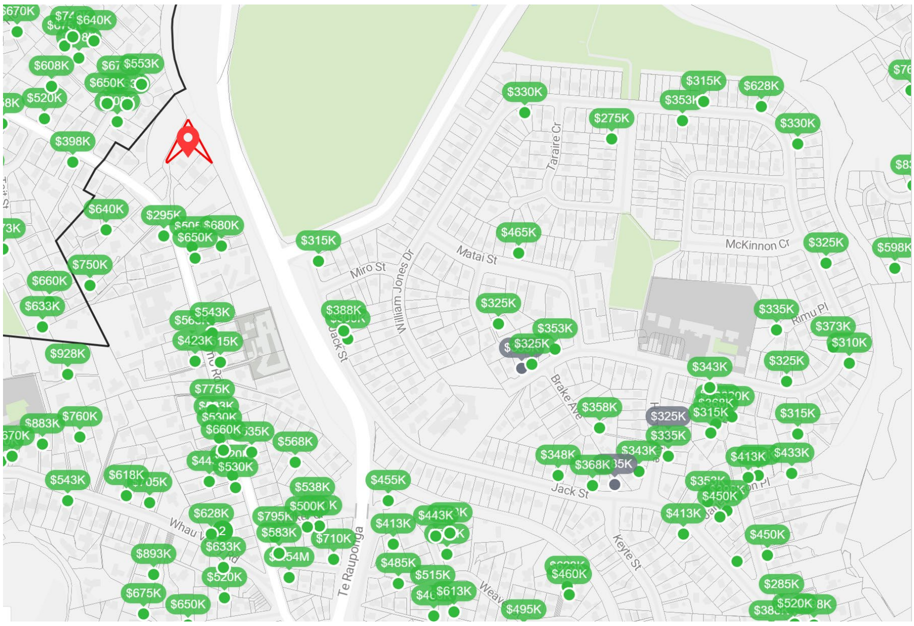


Picture_12 Top five most deprived areas in New Zealand

- Locations of the TOP5 deprived areas:

  Overall, the five most deprived areas are all located in the North Island.

  1) **Area 1** (Mbcode = 0094400 & 0094100): **Tikipunga**, a large suburb of Whangārei located in the very north of New Zealand, with a median income lower than the national median income. I also searched for the rent in Tikipunga, and it turned out to be lower than in surrounding areas. In picture_13, Area 1 is located on the right-hand side with an average rent of $ 350, which is lower than the areas on the left, which are mostly above $500.



Picture_13 The rent prices in Tikipunga and the surroundings

2) **Area 2** (Mbcode = 0788102): **East Tamaki**. East Tamaki is a suburb in the southeast of Auckland, primarily used for industrial purposes.

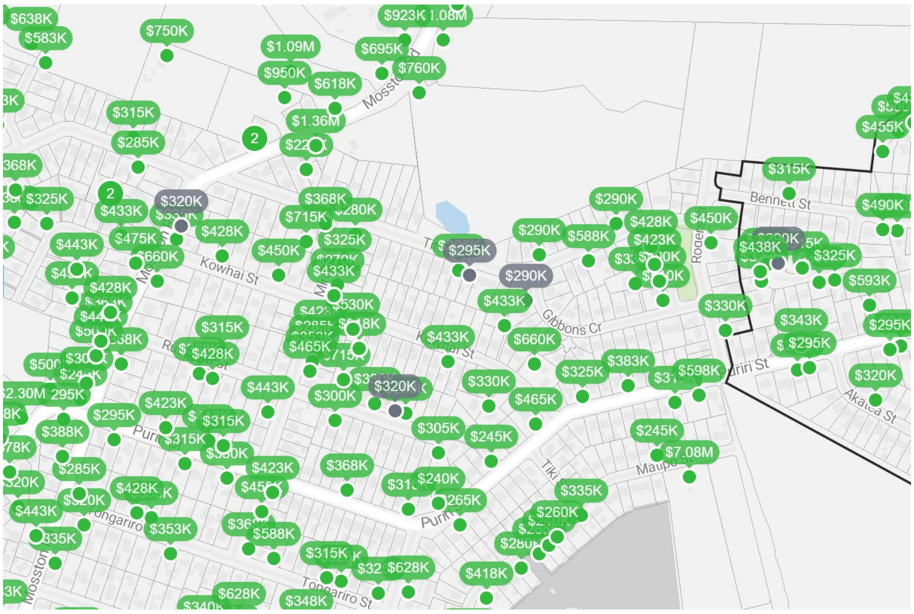
3) **Area 3** (Mbcode = 1332400): **Minginui**. Minginui is a small town in the Whakatāne District of the Bay of Plenty region on New Zealand's North Island, situated next to a Conservation Park. As shown in Picture 14, the area is surrounded by isolated mountains, and its population is in decline.

Picture_14 The hill shade map of the Minginui area

4) **Area 4** (Mbcode = 1446800): **Maraenui**. Maraenui is a suburb of Napier, located

in the Hawke's Bay region of New Zealand's North Island. The area is a lower socio-economic community, with residents including both homeowners and those living in state housing provided by Housing New Zealand. Additionally, Maraenui has a high rate of synthetic drug use. The suburb faces high levels of poverty and unemployment, as well as relatively low educational attainment. These factors have had a significant impact on the area's socio-economic conditions.

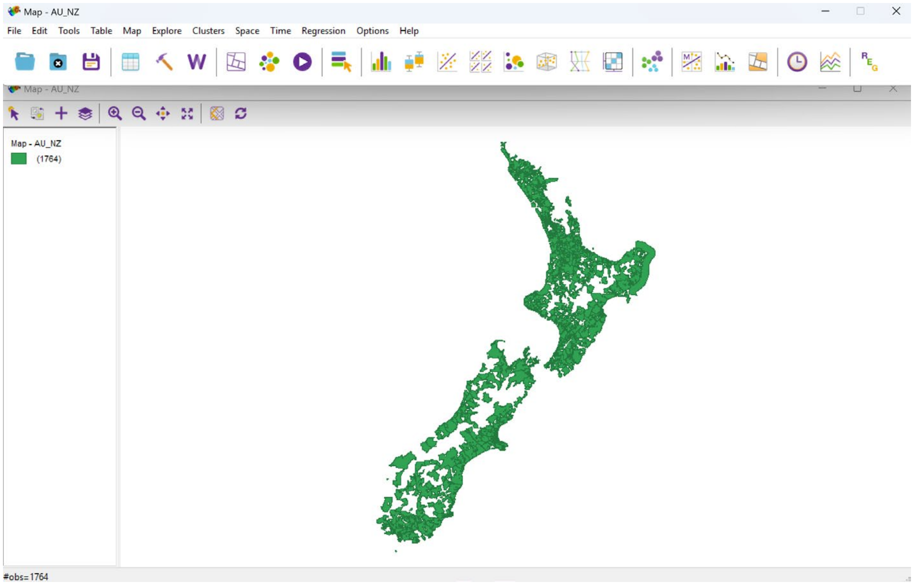
5) **Area 5** (Mbcode = 1716201): **Castlecliff**. Castlecliff is a suburb of Whanganui, located in the Manawatū-Whanganui region of New Zealand's North Island. Castlecliff is a multicultural community with rich educational resources and a growing population. However, the median income of residents is lower than the national average, and the unemployment rate is relatively high, posing certain socio-economic challenges. From the rent prices, we can see that the average price is around \$ 300-\$ 350, which is relatively lower.



Picture_15 The rent prices in Castlecliff and the surroundings

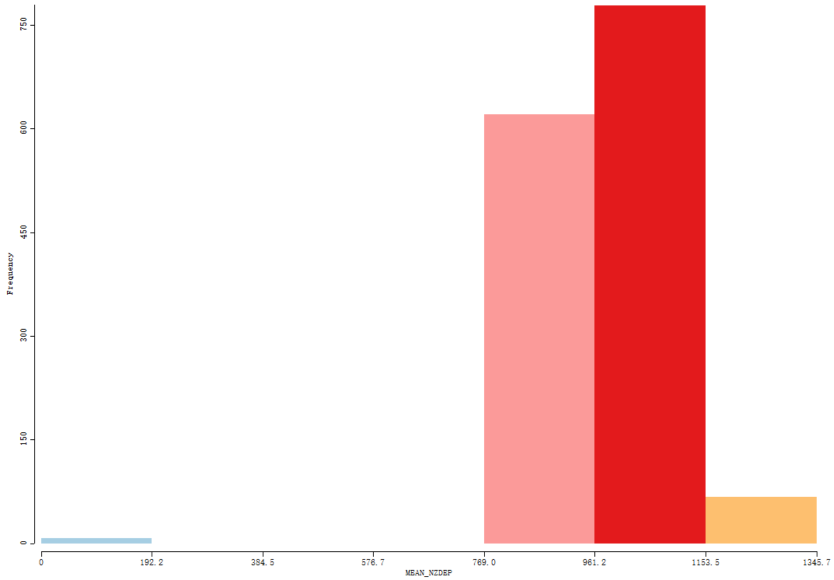
3.2 Histogram and Boxplot

- <u>Used GeoDa to analyze NZDep distribution at AU level</u>
  1) Load Shapefile AU_NZ in GeoDa. In this process, the shapefile can't be dragged directly to GeoDa. Otherwise, there would be a mistake. The reason for this is that the shapefile format contains multiple related files, so the loading process should be initiated through the 'ESRI shapefile' option.



Picture_16 Display AU_NZ in GeoDa

2) Create the histogram based on ‘NZDep’ in AU_NZ



Picture_17 The histogram of field ‘NZDep’ in AU_NZ

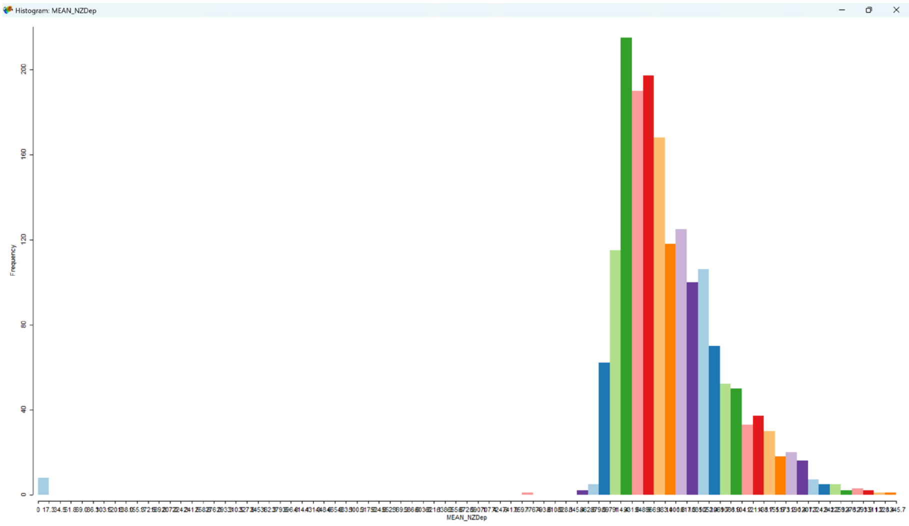
- <u>Applied Freedman-Diaconis Rule for refined histogram binning (78 bins)</u>
  1) There are seven bins in the default histogram. However, the classification is not detailed in the main range between 769 and 1153. Therefore, more bins are required to distinguish the interior differences within this range.
  2) Therefore, I calculate the new bin width according to the ‘Freedman-Diaconis Rule,’ which is suitable for large datasets containing more than 1,000 records

(we have 1,764 records in AU_NZ). The formula is illustrated in Picture 18, and the results are as follows: the new bin width is 17, and the number of bins is 78. After increasing the number of intervals, the histogram can reveal more details about the distribution within the main range of 769~1153.

$$h = 2 \times \frac{\text{IQR}}{n^{1/3}}$$

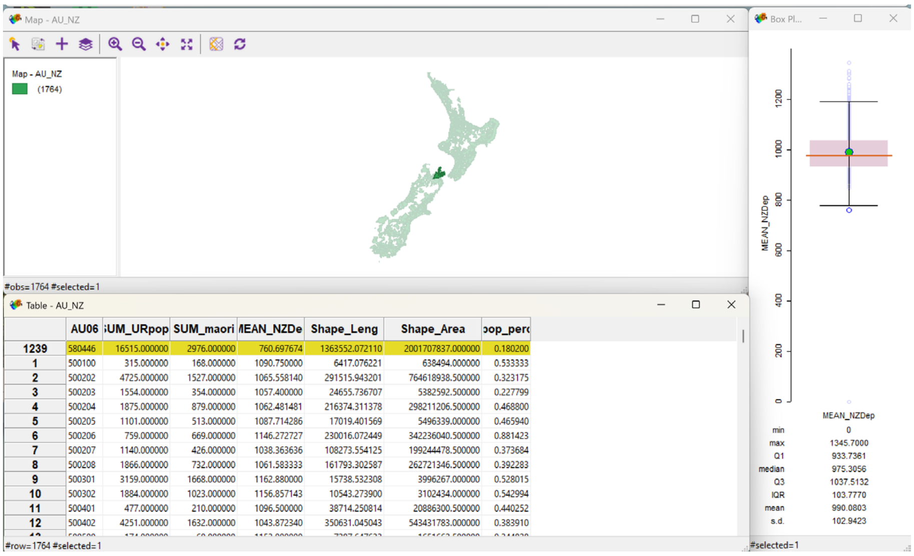
Picture_18 The formula of 'Freedman-Diaconis Rule'

3) Finally, we got the fixed detailed histogram with 78 bins.



Picture_19 New histogram of NZDep in AU_NZ with bin width = 17

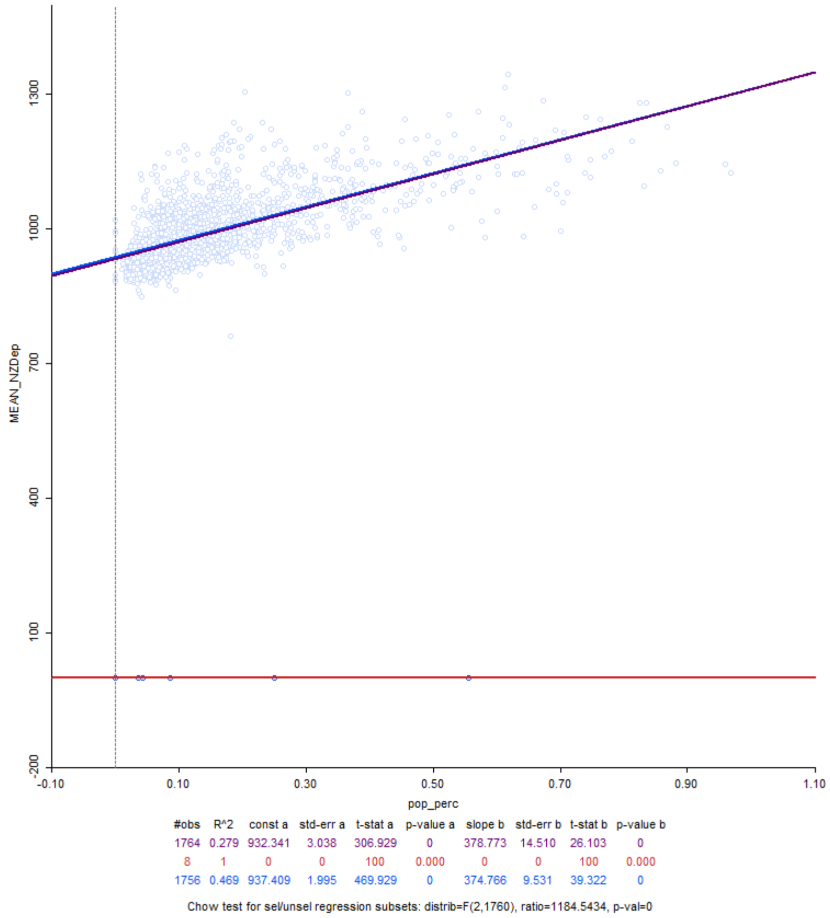
- <u>Identified 'MarlboroughSounds Terrestrial' as an outlier via boxplot</u>
  1) Create a boxplot in GeoDa using the AU_NZ dataset.
  2) Select the outlier just below the bottom hinge.
  3) From the table, we can see that the **AU code is 580446**, and the location is **Marlborough Sounds Terrestrial**. It is an extensive network of sea-drowned valleys **at the northern end of the South Island** of New Zealand.



Picture_20 Select outliers in the boxplot and check their AUcode and location

3.3 Scatterplot: NZDep vs. Maori %

- <u>Created regression scatterplot.</u>
  Using the Māori percentage as the independent variable (x) and NZDep as the dependent variable (y), create a scatterplot using a simple linear regression equation.



Picture_21 The regression scatterplot with Māori percentage (as X) and NZDep (as Y)

- Found moderate positive relationship (R^2 = 0.469, p-value = 0)
  1) Overall, the scatterplot illustrates a clear positive linear relationship between the two variables, ‘Maori percentage’ and ‘Mean Deprivation Score.’
  2) The p-value equals 0, indicating that the ‘Maori percentage’ has a statistically significant effect on the variance of the mean Deprivation Score.
  3) However, the R^2 in the statistics is only 0.469, indicating that the formula can explain 46% of the Mean Deprivation Score, which is moderate but not very effective.
  4) The points concentrate mainly on low values (where Māori percentages are between 0 and 0.5). There are only a few outliers at the bottom, suggesting that some errors in the investigation or other factors may be at play.

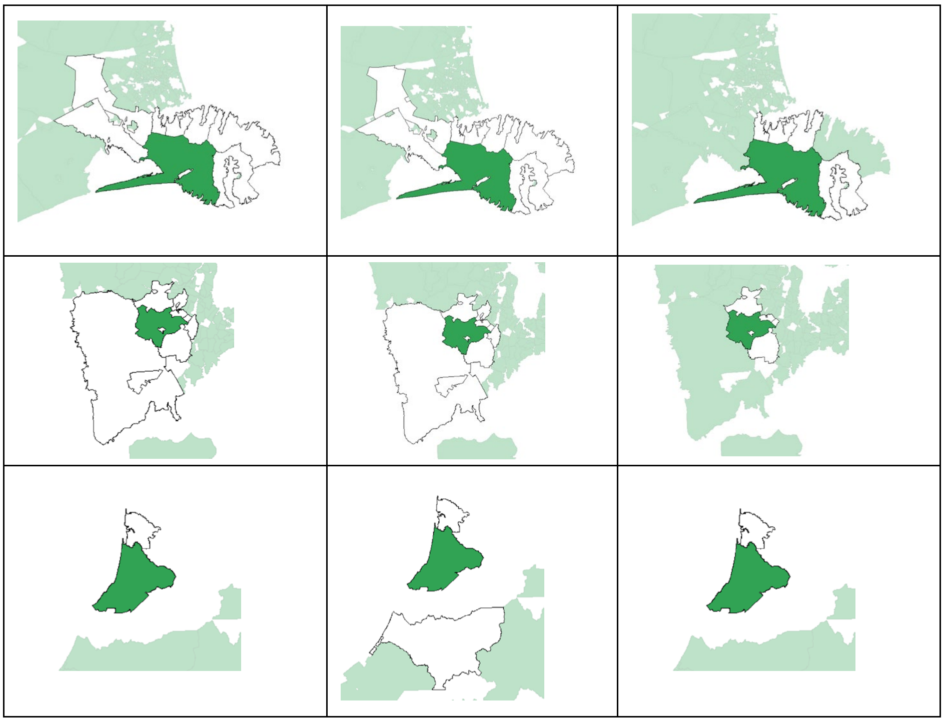
4. Spatial Autocorrelation Analysis

4.1 Moran's I at AU Level

- Created Queen, Rook, and KNN spatial weights

1) Create three weight files based on the ID field in AU_NZ; the order is always set to 1 because we consider only direct neighbors. All the other parameters are default. From the connectivity maps, it's indicated that one polygon always has four neighbors in K-Nearest (k=4) but has a quite different number of neighbors ranging from 0 to 14 in Queen and Rook.

2) Compare the neighbors of three different blocks in different weights files.



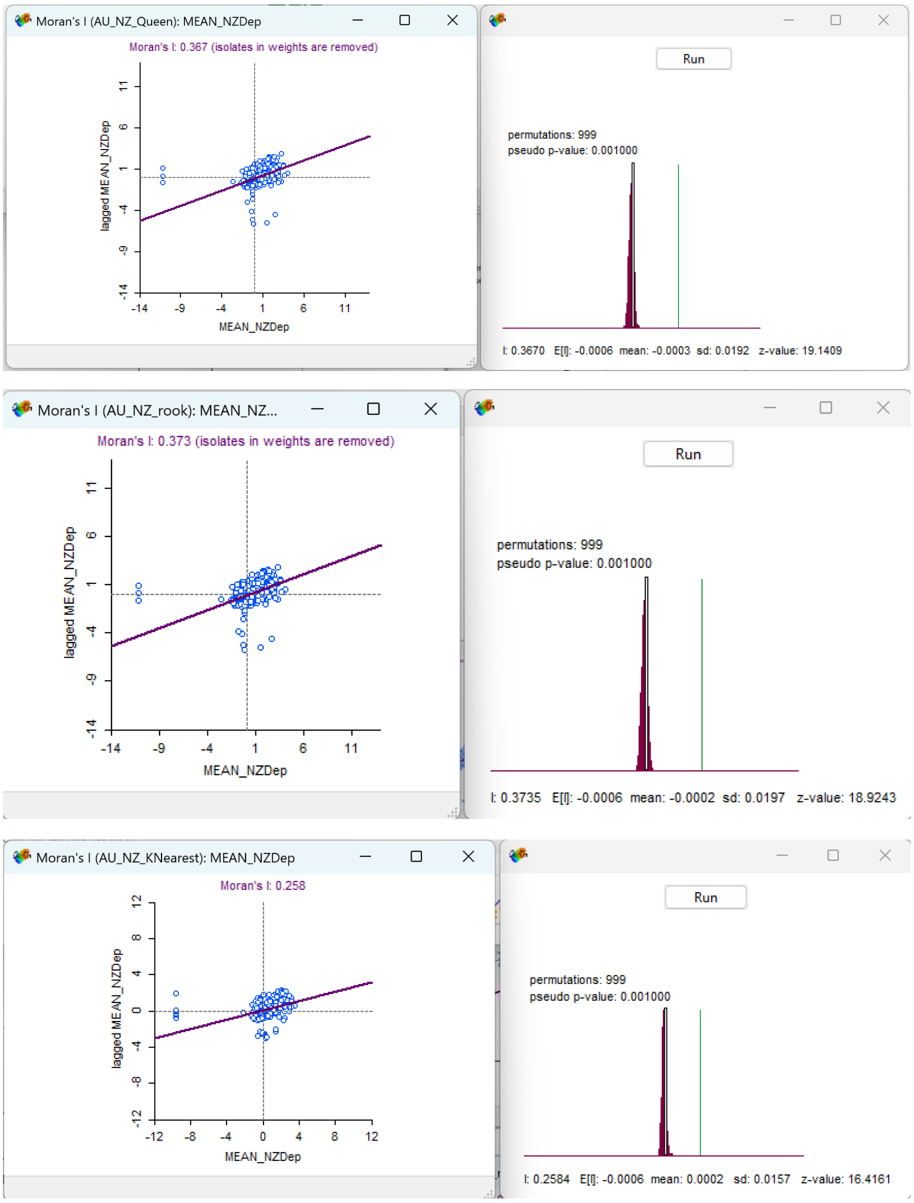
Picture_22 The neighbors of blocks in Queen, Rook, and KNN weights file from left to right

- <u>Moran's I: Queen = 0.367, Rook = 0.373, KNN = 0.258 (all p = 0.001)</u>

1) Using GeoDa, I computed three Univariate Moran's I based on three weight configurations:

Spatial weights	**Moran's I**	**z- value**	**p-value**
Queen contiguity	0.367	19.1409	0.0010
Rook contiguity	0.373	18.9243	0.0010
K-Nearest neighbors (k = 4)	0.258	16.4161	0.0010

2) The plots and statistics of the three Univariate Moran's I calculations are as follows.



Picture_23 The scatterplots of Moran's I based on Queen, Rook, and KNN using AU_NZ

- <u>Positive spatial autocorrelation found in all</u>

  1) **The autocorrelation is significant.** All the p-values based on three weight configurations are less than 0.05, indicating the autocorrelation is all statistically significant. Meanwhile, all three responding Univariate Moran's I are greater than 0, suggesting a positive autocorrelation in the deprivation across New Zealand, which means areas surrounding a high-deprivation area tend to be high-deprivation as well, and vice versa. Similar areas have a cluster trend.
  2) **Compare different weight configurations.** The Queen and Rook methods

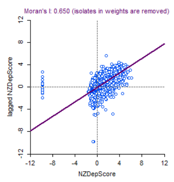
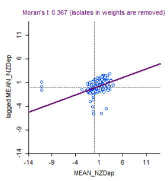
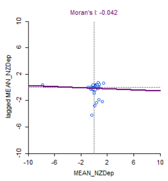
gave a similar Univariate Moran’s I result (0.367 and 0.373), while the K-Nearest Neighbor method produced a lower result (0.258). Therefore, the former two methods indicate a moderately strong positive spatial autocorrelation, and the third method shows a relatively weaker positive one.

3) **Despite the nuances between the three results, the overall conclusion stays consistent across three weight configurations. However, the choice of weight configuration can affect the degree of autocorrelation.**

4.2 Comparison Across Scales (Queen Weights)

- The Moran’s I analysis results:

  There are three aggregation scales data ordered from small to large: Meshblock, AU, and TA. For unification and comparability, the calculations of Univariate Moran’s I proceeded all based on the ‘Queen contiguity weights configuration’. Choosing this configuration because it’s more suitable for polygon data and can be applied to all three levels of aggregation data. The results are as follows:

Scale	Moran’s I	p-value	scatterplot
Meshblock	0.650	0.001	
AU_NZ	0.367	0.001	
TA_NZ	-0.042	0.286	

Picture_24 The Moran’s I analysis results on three levels

- Interpretations:

  Meshblock shows statistically significant strong positive autocorrelation. **AU** level shows a statistically significant but only a moderate positive spatial autocorrelation of deprivation; **TA** level shows no significant pattern.

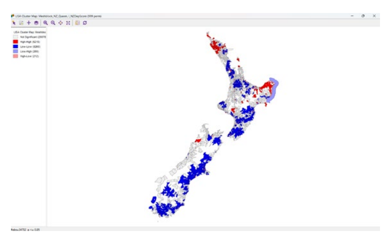
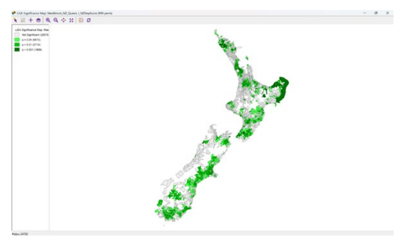
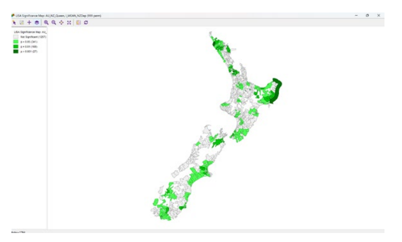
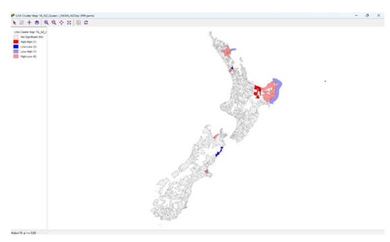
- Conclusion:

  Among the three levels of aggregation data, Meshblock with the lowest aggregation and smallest unit represents the strongest spatial autocorrelation. As the aggregation goes, the autocorrelation becomes weaker and weaker and finally disappears on the TA (the highest aggregation) level. This is because, during the aggregation, small units with high value and low value are mixed and averaged to a 'medium value.' Then, the original pattern existing at a low aggregation level is lost. The original neighbors no longer exist. Instead, there are lots of new neighbors with a 'far' relationship. That's why autocorrelation becomes weaker at high aggregation levels.

4.3 Local Moran's I (LISA Maps)

- The Local Moran's I analysis results:

  There are three aggregation scales data ordered from small to large: Meshblock, AU, and TA. For unification and comparability, the calculations of Local Moran's I proceeded all based on the 'Queen contiguity weights configuration.' The results are as follows:

Scale	Lisa cluster map	Lisa significance map
Meshblock		
AU_NZ		
TA_NZ		

Picture_25 The LISA maps on three levels

- **Meshblock level:**

1) The LISA cluster map indicates that a clear spatial pattern exists.
2) Overall, the North Island has many more high-high clusters than the South Island. Significant **high-high clusters** are concentrated in the northeastern and the very northern parts of the North Island. There are also some small clusters scattered inside the North Island and along the eastern and southern coastline of the South Island.
3) In contrast, **low-low clusters** are concentrated in central and southern parts of the North Island and western and southern South Island.
4) There are some **outliers** labeled as **high-low or low-high**, being possible transitional zones or other uncertain reasons. Luckily, the outliers have a small number.
5) The LISA significance map supports cluster map patterns.
6) Most of the high-high or low-low areas have p-values less than 0.05. Some big cities and surrounding areas (such as Christchurch, Auckland, Queenstown, Wellington, etc.) even have a p-value less than 0.01, while the northeastern and the very northern part of North Island have a p-value less than 0.001.
7) These results indicate that the cluster pattern is not random but a reflection of a meaningful trend in the real geographic world.
8) Many areas are not significantly autocorrelated, as the Meshblock is a fine statistical unit, and it can reflect detailed distribution patterns yet is not able to avoid ignoring some less significant clusters due to its too tiny and broken bits of units.

- **AU level: More generalized patterns; clearer regional clusters**
  1) At the Area Unit level, a more generalized and definite geographical pattern is indicated.
  2) The **high-high regions** are concentrated in the northeastern mountainous areas of the North Island, part of the central area of Auckland, Northland, etc., meaning being deprived.
  3) In contrast, the **low-low regions,** which means being affluent, are concentrated in main cities or towns, such as Auckland, Hamilton, Naper, Wellington, Nelson, Christchurch, Queenstown, Oamaru, Dunedin, etc.
  The significance map indicates that many of the high-high and low-low clusters in the cluster map are significant. Most of the p-values fall in the range of $p < 0.05$, while there are a few $p < 0.01$ and $p < 0.001$, being particularly concentrated in the northeastern and northern parts of the North

Island.

4) **Interpretation:** On the Area Unit level, the unit becomes more generalized than the Meshblock level. The pattern represented by AU_NZ data is more generalized and easier to interpret as the clusters become more concentrated and less localized due to the coarser resolution of the unit level.

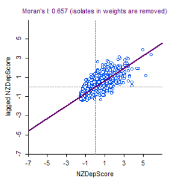
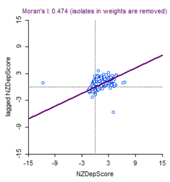
- <u>**TA: Most clusters disappear due to aggregation**</u>

1) In this level, the high-high and low-low clusters are limited to the northeastern part of the North Island and the South Island, respectively. Meanwhile, the high-low and low-high outliers are limited as well, suggesting the majority are ‘not significant.’ This phenomenon indicates that the clustering pattern seen in the lower aggregation levels (Meshblock and AU) averaged out due to further aggregation.
2) The significance map confirms the trend: the areas with a p-value less than 0.05 are quite limited, indicating only a few units at the TA level show significant autocorrelation. Strong significance can rarely be detected due to the large geographical unit area and small number of units.

5. City-Level Comparison: Wellington vs Christchurch

5.1 Univariate Moran’s I

- <u>The Moran’s I in two cities are as follows:</u>

City Name	Moran’s I	p-value	scatterplot
Christchurch	0.657	0.001	
Wellington	0.474	0.001	

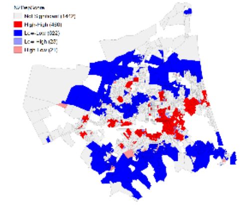
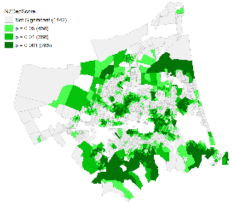
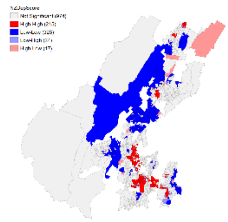
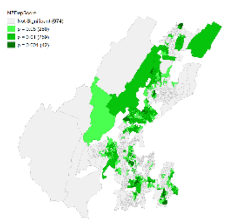
Picture_26 The Moran’s I analysis results in two cities

- Both Christchurch and Wellington show statistically significant positive spatial autocorrelation of NZDepScore with p-value = 0.001 < 0.05.
- **Christchurch** has a Univariate Moran’s I of 0.657, representing a strong positive spatial autocorrelation.

- **Wellington** also has a Univariate Moran's I of 0.474, but this is a moderate positive spatial autocorrelation slightly lower than Christchurch.
- **Comparing the two cities,** Christchurch exhibits a stronger clustering of deprivation, while Wellington implies a more dispersed pattern of deprivation.

5.2 LISA Interpretation

- The LISA (using Queen weight)analysis results are:

City Name	Lisa cluster map	Lisa significance map
Christchurch		
Wellington		

Picture_27 The LISA maps in two cities

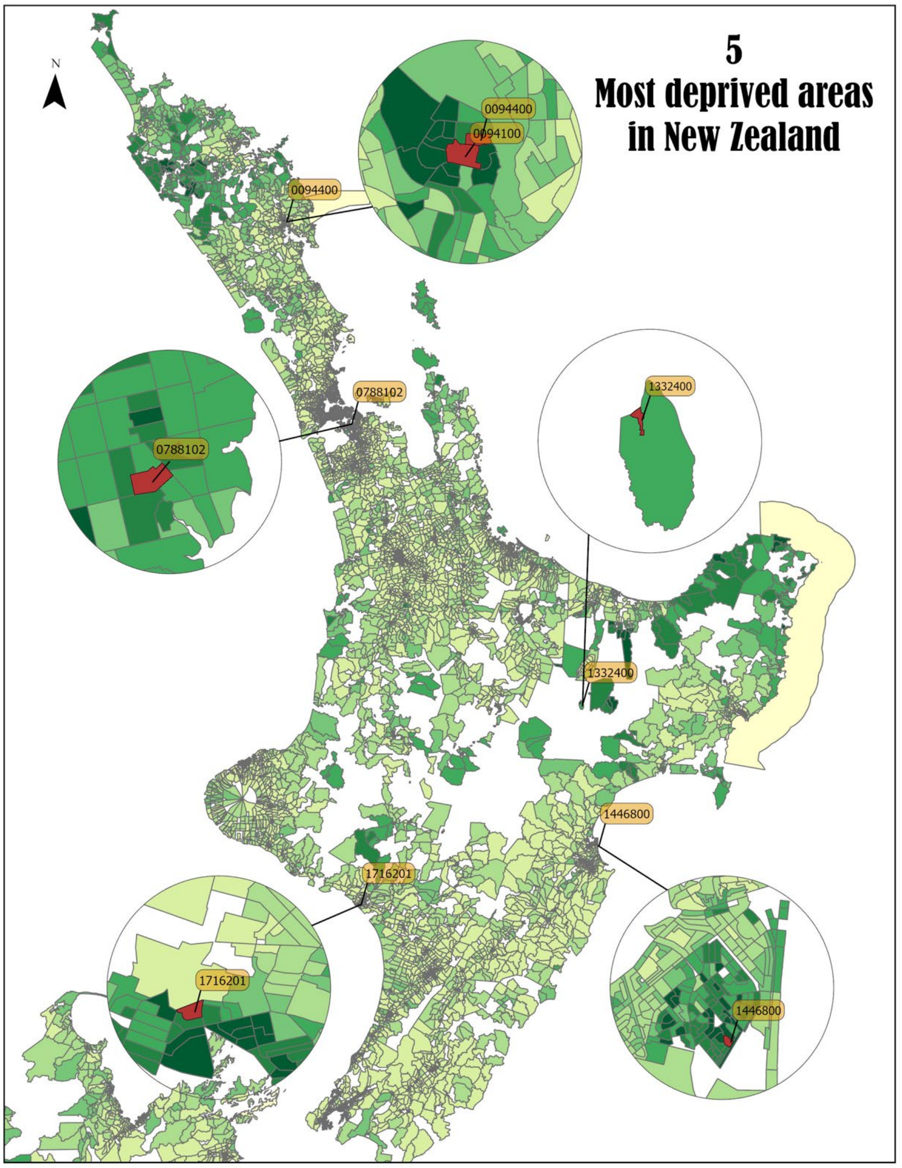
- Christchurch: High-high clusters in central-south; ring of low-low in outskirts
  1) **In Christchurch,** the **high-high clusters** are concentrated in the southern part of the central city, indicating a high deprivation clustering pattern, while the **low-low clusters are mainly on the edge of the central city,** being like a ring shape.
  2) This pattern is not random but a reveal of the geographic distribution because there is a widespread p-value that is less than 0.05.
  3) Additionally, there are many significant areas with p-values less than 0.01 or even less than 0.001, indicating that the clustering pattern is unlikely random.
- Wellington: Dominant low-low clusters in suburbs; few high-high zones
  1) **In Wellington,** the low-low clustering pattern takes a dominant place, widely clustered in the western and northern suburbs. Only a few high-high clusters are located in the southeastern corner.
  2) The significance map of Wellington can still confirm the cluster pattern; however, it is not as strong as Christchurch because there are few places with a p-value less than 0.001.
- These differences are consistent with the Univariate Moran's I analysis, indicating

that Christchurch has a more statistically significant and stronger positive spatial clustering of deprivation than Wellington.

6. Insights and Recommendations

6.1 Layout_1 Top 5 Deprived Areas

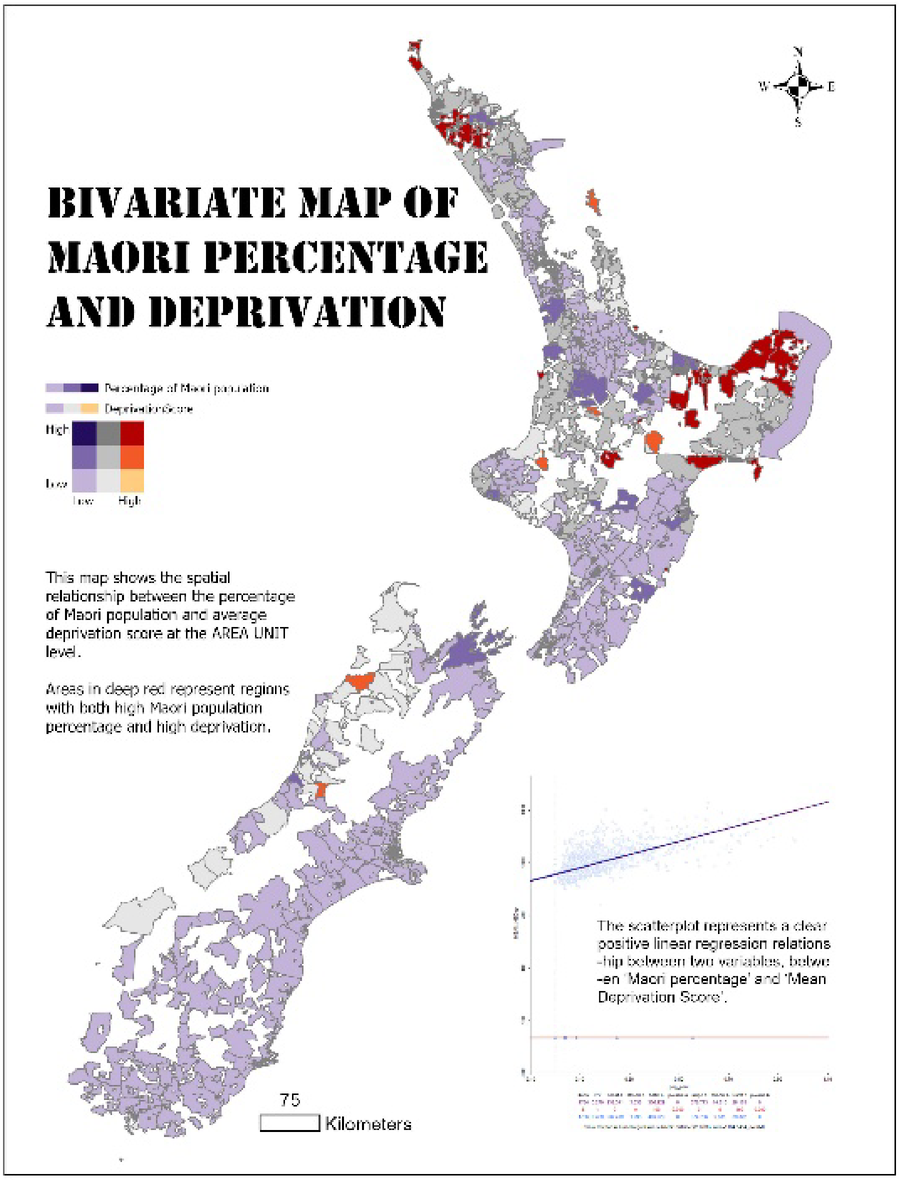
- Contributing factors: geographic isolation, industrial zoning, low income, drug use, unemployment.
- Interventions: Improve access, career training, healthcare, and drug prevention.



Picture_28 Top five most deprived areas in New Zealand

6.2 Layout_2 Māori Population Correlation

- Higher Māori population associated with higher deprivation
- Recommendations: fund education, bilingual support, job training

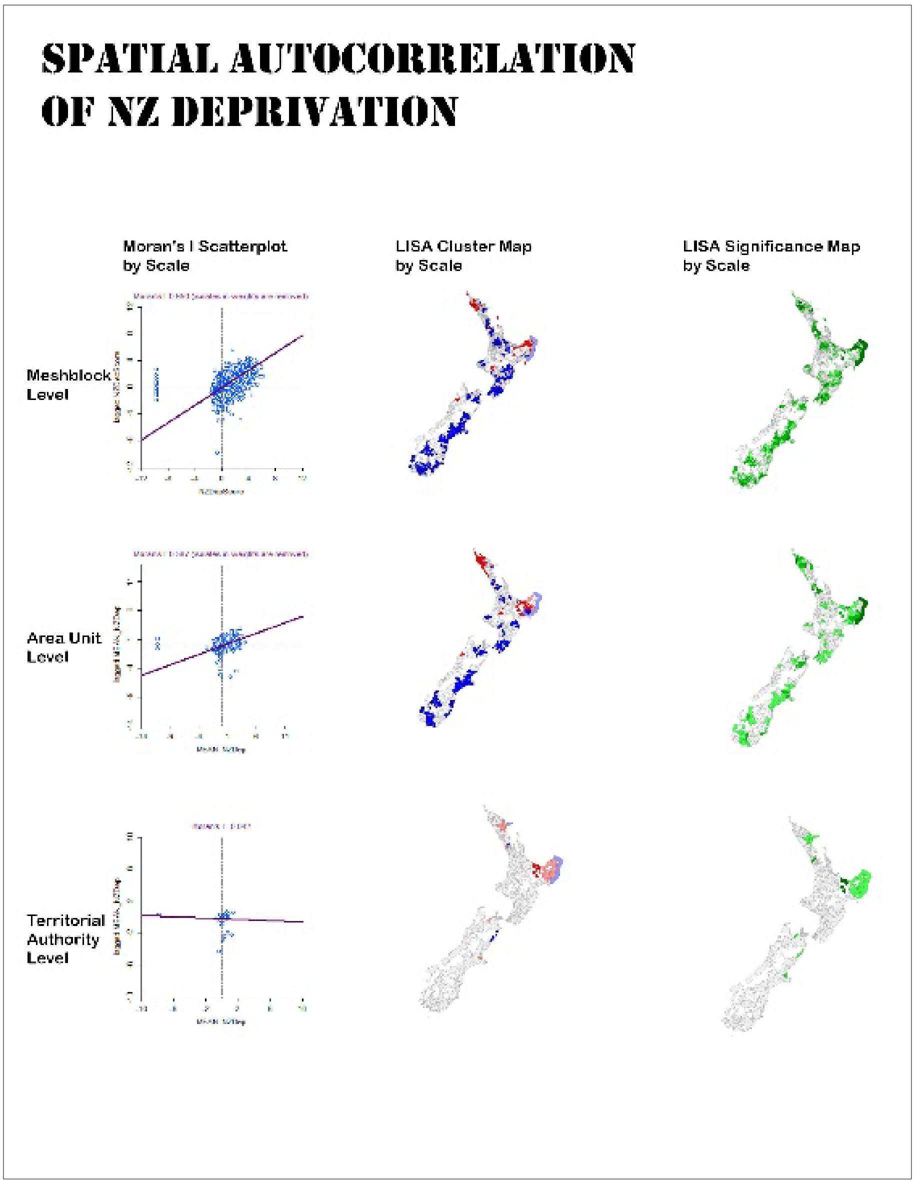


Picture_29 Bivariate map of Maori percentage and deprivation

6.3 Layout_3 Spatial Autocorrelation Patterns

- Strongest clustering at fine scale (Meshblock)
- AU level shows a more generalized pattern
- Aggregation dilutes spatial patterns
- Interventions: We should make plans considering responding levels of data,

otherwise we may achieve inaccurate results.



Picture_30 Comparison of spatial autocorrelation on three levels

7. Conclusion

This report demonstrates the significant impact of spatial resolution on observed deprivation patterns. A detailed spatial autocorrelation analysis reveals strong clustering at finer levels, underscoring the importance of granularity in policy planning. Christchurch exhibits more concentrated deprivation compared to Wellington, and the Māori population is a statistically significant factor in the distribution of deprivation.