Assessing Equality in Canada

Vy Tran

August 2025

1. Introduction

Canada has historically been portrayed as a "Land of Opportunity", offering life-changing opportunities to its large and diverse immigration population, regardless of their origin and ethnicity (InvestImmi – Canadian Immigration Services, 2015). While the various individual success stories seem to support this notion, there remains a lack of concrete quantitative evidence of the existence and accessibility of these opportunities. This report aims to address this problem by analyzing and then presenting empirical evidence supporting or challenging this belief in equality.

Quantifying "opportunities" is a significant challenge, as is measuring "equality" across Canada. Various indicators such as education levels, employment rate, and social mobility can be used to represent "opportunities", and many more methods exist to measure the equality of these factors. However, this report chooses to concentrate on income, specifically the average income of selected populations, due to the data's relative accessibility and clarity as a quantitative variable of opportunity. Thus, to evaluate income equality, the analysis will employ the Gini coefficient to address the overall income distribution of those populations.

2. Data's Source, Overview, and Sampling

The dataset, released on October 4th, 2023, was sourced from Statistics Canada's "Employment income statistics by visible minority, immigrant status, highest level of education and location of study: Canada, provinces and territories" and retrieved on June 12th, 2025. It is derived from the Canada 2021 Census of Population and contains 296,352,000 unique datapoints (Statistics Canada, 2023b).

The dataset includes the following attributes and their respective number of attribute values in parentheses:

1. Age (15)
2. Employment Income Statistic (7)
3. Gender (3)
4. Geography (14)
5. Highest Certificate, Diploma or degree (16)
6. Immigration Status (4)
7. Location of Study (5)
8. Major field of study - Classification of Instructional Programs (CIP) 2021 (14)
9. Visible Minority (15)

The attribute "Employment Income Statistic" further contains 7 key observations:

1. Total - Employment Income
2. With Employment Income
3. Median Employment Income ($)
4. Average Employment Income ($)
5. With Wages, Salaries and Commissions
6. Median Wages, Salaries and Commissions ($)
7. Average Wages, Salaries and Commissions ($)

For this report, the analysis will focus on the following attributes: Geography, Highest Certificate, Diploma or degree, Immigration Status, and Visible Minority. Additionally, only 2 income-related metrics will be kept from the attribute "Employment income Statistic":

1. With Employment Income
2. Average Employment Income ($)

In addition, for the attribute “Highest Certificate, Diploma or degree”, only the following values will be used to target subsets of the population with only a diploma or with a bachelor’s degree or higher, respectively:

1. High (secondary) school diploma or equivalency certificate.
2. Bachelor’s degree or higher.

Similarly, for the attribute “Immigrant Status”, only the following values will be used to target the subsets of the population who are immigrants or non-immigrants:

1. Immigrant
2. Non-immigrants

Finally, for the attribute ‘Visible Minority’, the following value would be removed during evaluation for the Gini coefficient, as it would double-count some of the population’s subsets. Specifically, the below 2 values double-count the entire population of the visible minority and a certain subset of the visible minority, respectively:

1. “Total – Visible Minority”
2. “Total – Visible Minority Population”

However, the attribute ‘Visible Minority ’does contain potentially confusing values, displayed below with clarification:

1. Visible Minority: ” This category includes persons who provided responses that are classified as a visible minority, but that cannot be classified with a specific visible minority group. Such responses include, for example, "Guyanese," "Pacific Islander," "Polynesian," "Tibetan," and "West Indian.” (Statistics Canada, 2023b).
2. Multiple Visible Minorities: Assumed to be individuals from more than 1 background.
3. Not a Visible Minority: “In 2021 Census analytical and communications products, this category is referred to as ‘the rest of the population."(Statistics Canada, 2023b).

The remaining attributes will be analyzed using their full domain. For example, the attribute “Location of Study” will be “Total - Location of Study”.

For formulation on QGIS, which functions require the key attributes to have a consistent data format, which in the 5 above cases will be in the integer format:

1. Undefined value caused by the division of zero in the key attribute “Average Employment Income ($)” where their respective population is empty will be converted to zero.
2. Suppressed datapoints in the key attributes, which are denoted with X due to the confidentiality required by the Statistic Act, will be converted to zero (Statistic Canada, 2023b).
    a. These conversions ensure consistency in data format and prevent “Null” or disrupting data or spatial operation in QGIS.

3. Methods and Data Sampling

The full database from "Employment income statistics by visible minority, immigrant status, highest level of education and location of study: Canada, provinces and territories" after sampling by the restrictions mentioned above in 2-Data's Source, Overview and Sampling is edited into a different format to be used in QCIS (Statistics Canada, 2023b):

1. After downloading, the data package is unzipped, opened, and edited using Excel.
2. The spreadsheet typically contains non-data elements such as introductory texts, annotations, and metadata around the actual data. All non-data elements in the rows above the data are removed; other data is deleted. The data elements are then aligned with the first row. Column headers are double-checked and filled in from Statistics Canada.
3. Undefined or suppressed datapoints are replaced with zero.
4. The following attributes: Geography, Highest Certificate, Diploma or degree, Immigration Status, and Visible Minority contain some blank entries, where repeated entries are only shown once. In addition, each of their respective attribute values is annotated numerically. The cells containing the above attributes and their attribute values' annotation are removed, and the columns are forward-filled.
5. The spreadsheet should be saved as a CSV file, duplicated, and opened as a GeoPackage in QGIS, which allows edits to be made to the data. This GeoPackage will be referred to as the data GeoPackage onwards. In addition, add Canada's provinces and territories to QGIS as a GeoPackage will be referred to as the shapefiles GeoPackage onwards. The shapefile used for this report is provided by Statistics Canada, 2023a.
6. In QGIS, using the "Field Calculator" and the "to_int" function to duplicate and replace the above-mentioned income metrics in "Employment Income Statistic" with integer values. Make sure the new columns have the same data format, being 32-bit integers.
7. Since QGIS's Join function can't do a one-to-many data joining, which, for the interest of this project, is each of Canada's provinces and territories' shapefiles to the data points in the data GeoPackage, virtual fields will be used as a workaround. For example, for Ontario, after turning on "Edit" for the data GeoPackage, create 2 virtual fields with "Field Calculator" in 32-bit integer format, "Ontario - With Employment Income" and "Ontario - Average Employment Income ($)". When creating these fields, use the "Expression" tab and the expression:

   ```
   CASE
   WHEN "Geography" = X
   THEN  Y
   ELSE NULL
   END
   ```

8. Where X is replaced by 'Ontario' and Y is replaced by "Average Employment Income ($) " or "With Employment Income", depending on the virtual field.
9. Repeat step 7 for each province or territory of Canada. Afterwards, each of the produced virtual fields will only have values for their respective province and income-related metric, and the rest of their cells will be Null.

10. Create a new GeoPackage, referred onwards as the bridge GeoPackage. This GeoPackage will be used to process and join the data with the shapefiles GeoPackage.
11. In “Edit” mode, and the bridge GeoPackage's attribute table opened, create a field “Geography” in string format with each cell containing Canada’s provinces and territories, identical to the shapefiles GeoPackage’s field containing the provinces and territories’ names. Afterwards, using the “Field Calculator” to create a field with string format, named “Average Employment Income ($)”, with the expression as follows:

    CASE

    WHEN "Geography" = A

    THEN array_to_string(array_filter(aggregate(B, 'array_agg', C) ,@element IS NOT NULL))

        ...

    END

12. Where A is the name of the province or territories in single quotes, B being the string for the data Geo Package's name with single quotes in the system and C refers to the virtual field created by step 7 and 8, which should be a double quote string with A connected with “Average Employment Income ($)” with a dash. The triple dot above refers to the repeat of the middle section of the code for every province or territory, enclosed within “CASE” and “END”.
13. Repeat step 10, but for another field in string format, named “With Employment Income” instead of “Average Employment Income ($)”. Steps 10 and 11 will create 2 fields for the bridge GeoPackage and contain the “Average Employment Income ($)” and “With Employment Income” from the data GeoPackage, split up into their respective provinces. The multiple datapoints associated with each province or territory from the data GeoPackage are stored in a single cell for each province as an array of integers, converted to a string. This workaround allows any datapoints’ “Average Employment Income ($)” and “With Employment Income” to be linked after the conversion into a string, where the values share the same position in the converted array.
14. Using the “Field Calculator”, open the “Function Editor” and add the following function:

```
def gini(people_str, income_str):
        people = list(map(int, people_str.split(',')))
        income = list(map(float, income_str.split(',')))
        paired = list(zip(people,income))
        paired.sort(key=lambda x: x[1])
        people, income  = zip(*paired)
        people = list(people)
```

```
    income = list(income)
    n = sum(people)
    total_income = 0
    i = 1
    summation = 0
    for subset in range(len(people)):
        j=people[subset]
        x = income [subset]
        total_income+= j*x
        summation += x * (j * (n - i +1) - (j * (j - 1)) // 2)
        i+= j
return 1+ (1/n) - (2*summation/(total_income*n))
```

The above function takes 2 strings of arrays of integers, where people_str should be the value of "With Employment Income" and income_str should be "Average Employment Income ($)" from the GeoPackage, and produces the Gini coefficient for each province or territory. The function Gini is the optimized code form of the following derived formula to find the Gini coefficient (Kanel, 1993):

$$G = 1 + \frac{1}{\mathrm{n}} - \frac{2}{n^2\mu}(y_n + 2y_{n-1} + \cdots + ny_1)$$

15. The Gini coefficient, the derived formula, and the original formula will be explained further in 4 – Data Analysis.
16. Create a virtual field in 32-bit integer format in the bridge GeoPackage, named "Gini coefficient", and call the function Gini in the expression tab, with the input being the fields "With Employment" and "Average Employment Income" consecutively. The field should produce the Gini coefficient for each province or territory.
17. Using the Join function in QGIS to link the shapefiles GeoPackage and the bridge GeoPackage by the field "Geography", which allows the shapefiles GeoPackage to display labels according to the Gini coefficient.
18. To change the population of evaluation required below, use the "Query Builder" in the "Source" tab of the data GeoPackage and build the required filter for the evaluations below.

To evaluate Canada as a land of opportunity and equality, it is important to analyze different population groups. The following groups will be evaluated and presented in 4-Data Presentation:

1. Different ethnicities with both education levels and both immigrant status. This analysis will help to identify the availability and extent of opportunities between ethnicities. This will create 14 figures from the 13 retained values of the attributes "Visible Minority", where 13 figures for each ethnicity and 1 figure that includes all ethnicities.
2. Different education levels, all 13 ethnicities, and both immigrant statuses. This analysis will help to identify the availability and extent of opportunities between

different educational levels. There will be 2 figures for the 2 included values of the attribute “Highest Certificate, Diploma or degree”.

3. Different immigrant status with all 13 ethnicities and both education levels. This analysis will help to identify the availability and extent of opportunities between immigrants and non-immigrants. There will be 2 figures for the 2 included values of the attribute “Immigrant Status”.

During evaluation for equality, the Gini coefficient will be applied to each population group. For a population with a non-negative income, the Gini coefficient will be between 0 and 1. A value of 0 represents perfect equality, where each person makes the same income. On the other hand, a value of 1 represents perfect inequality, where 1 person makes the entire population’s income. The Gini coefficient can be calculated with the following formula, original and optimized (as $n^2\mu = n^2\frac{\sum_{i=1}^{n} y_i}{n} = n\sum_{i=1}^{n} y_i$) (Kanel, 1993):

$$G = 1 + \frac{1}{\mathrm{n}} - \frac{2}{n^2\mu}(y_n + 2y_{n-1} + \cdots + ny_1) = 1 + \frac{1}{n} - \frac{2\sum_{i=1}^{n}(n-1+i)y_i}{n\sum_{i=1}^{n} y_i}$$

Where G is the Gini coefficient of the population, n is the number of individuals, y_i is the income of the i-th individual, μ is the mean income of the population, and such that $0 \leq y_1 \leq y_2 \leq \cdots \leq y_{n-1} \leq y_n$.

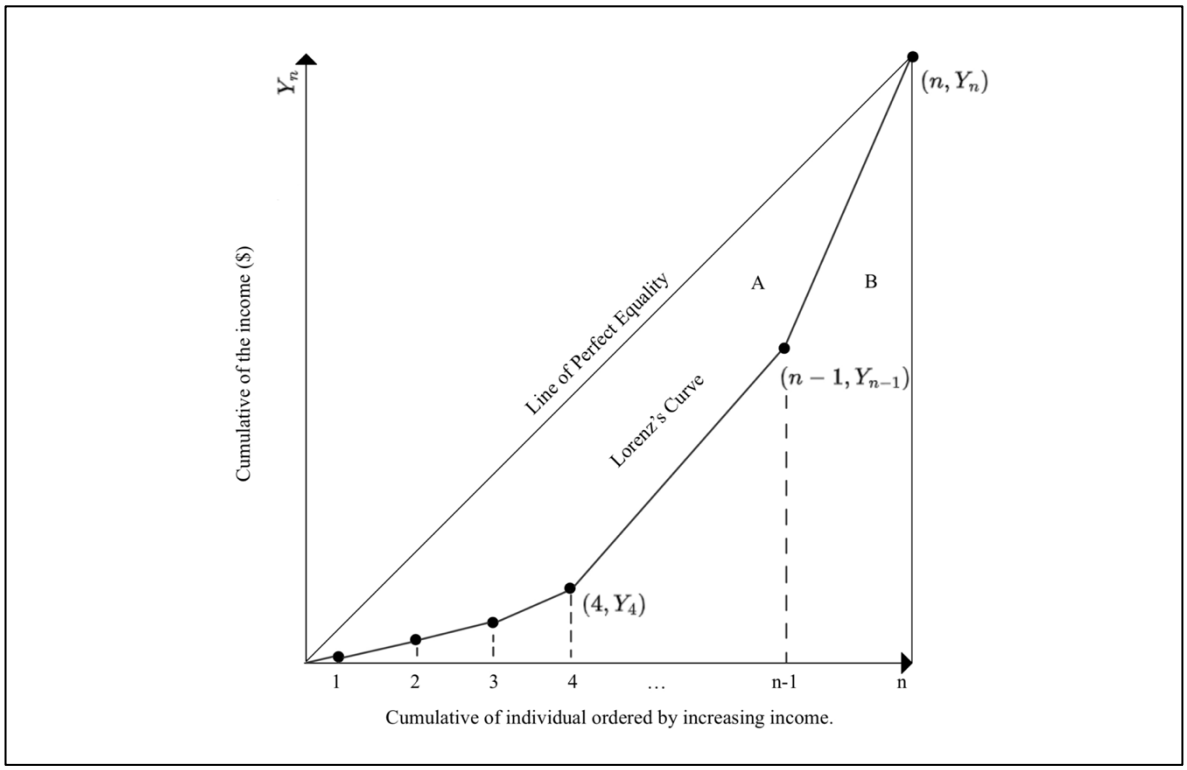

Figure 1. The graphical representation of the line of perfect equality, the Lorenz curve of an example population of n individuals, where $Y_j = \sum_{i=1}^{j} y_i$, the cumulative income of the population up to the i-th individual.

With the population being ordered by increasing income, the x-axis being the cumulative number of individuals, and the y-axis being the cumulative income of those individuals, the Gini coefficient of a population is how far that population's Lorenz curve deviates from the line of perfect equality. The Lorenz curve is formed by the data points of the i-th individual and the cumulative income of the population up to the i-th individual (Christian, n.d). It can be understood that the line of perfect equality is the Lorenz's curve of a second hypothetical population with the same total individual and income, with everyone making the same amount instead.

In theory, the Gini coefficient is defined as $G = \frac{A}{A+B}$ on Figure 1, where A is the area between the Lorenz curve and the line of perfect equality, while B is the area below the Lorenz curve (Kanel, 1993).

The Gini can also be expressed as:

$$G = \frac{A}{A+B} = \frac{A+B}{A+B} - \frac{B}{A+B} = 1 - \frac{B}{A+B}$$

From Figure 1, the area of $A + B$ can be calculated as a right-angle triangle with the base of n and height of Y_n, (Kanel, 1993):

$$A + B = \frac{nY_n}{2} = \frac{n\sum_{i=1}^{n} y_n}{2}$$

Furthermore, the area of B can be calculated as the total area of $n-1$ trapezoids, each with a base of 1 unit, the parallel sides being Y_j and Y_{j+1}, and a right-angle triangle with a base of 1unit and height of Y_1 which simplified to (Kanel, 1993):

$$\begin{aligned} B &= \frac{Y_1}{2} + \frac{Y_1 + Y_2}{2} + \frac{Y_2 + Y_3}{2} + \cdots + \frac{Y_{n-1} + Y_n}{2} \\ &= Y_1 + Y_2 + \cdots + Y_{n-1} + \frac{Y_n}{2} \\ &= \sum_{i=1}^{n-1} Y_i + \frac{Y_n}{2} = \sum_{i=1}^{n} Y_i - \frac{Y_n}{2} \\ &= \frac{-Y_n}{2} + y_1 + \\ &\quad y_1 + y_2 + \\ &\quad \cdots \\ &\quad y_1 + y_2 + \cdots + y_{n-1} + y_n \\ &= \sum_{i=1}^{n} (n - i + 1) y_i - \frac{Y_n}{2} \end{aligned}$$

When put together, the formula for the Gini coefficient provided by Kanel (1993) is produced:

$$G = 1 - \frac{B}{A+B} = 1 - \frac{2\left(\sum_{i=1}^{n}(n-i+1)y_i - \frac{Y_n}{2}\right)}{nY_n}$$

$$G = 1 - \frac{2\sum_{i=1}^{n}(n-i+1)y_i}{nY_n} - \frac{Y_n}{nY_n} = 1 - \frac{1}{n} - \frac{2\sum_{i=1}^{n}(n-i+1)y_i}{n\sum_{i=1}^{n} y_i}$$

$$G = 1 + \frac{1}{\mathrm{n}} - \frac{2}{n^2\mu}(y_n + 2y_{n-1} + \cdots + ny_1)$$

However, in the data provided by Statistics Canada, individual income has been aggregated into average income for each population subset (Statistics Canada, 2023b). For this report, it's assumed that every individual of a subset has the subset's average income. However, due to each individual in a subset sharing the same income, the summation process in the numerator can be further simplified as below:

$$S = \sum_{i=a}^{a+p_k} x_k(n-i+1) = \sum_{Z=0}^{p_k-1} x_k(n-\ a+1-z)$$

$$= x_k\left(n-a+1-\sum_{Z=0}^{p_{k-1}} z\right) = x_k\left(n-a+1-\frac{(p_k-1)p_k}{2}\right)$$

Where S refers to the summation for the subset k in the above-mentioned formula, when calculating the Gini coefficient, x_k is the average income of the subset k, p_k is the population of the subset k, a is the cumulative of individuals in the subset with less than or equal average income of the subset. Hence, the optimized Python function for the Gini coefficient is as below:

```
def gini(people_str, income_str):
    people = list(map(int, people_str.split(',')))
    income = list(map(float, income_str.split(',')))
    paired = list(zip(people,income))
    paired.sort(key=lambda x: x[1])
    people, income  = zip(*paired)
    people = list(people)
    income = list(income)
    n = sum(people)
    total_income = 0
    i = 1
    summation = 0
    for subset in range(len(people)):
        j=people[subset]
        x = income [subset]
        total_income+= j*x
         summation += x * (j * (n - i +1) - (j * (j - 1)) // 2)
        i+= j
    return 1+ (1/n) - (2*summation/(total_income*n))
```

The function Gini takes in 2 strings, each are integer separated by commas, being the number of individuals in each subset and the income of that subset, and returns the Gini coefficient as a float using the optimized formula for the Gini coefficient.

After the evaluation mentioned above, the Gini coefficient of each province or territory for each different population group is displayed in 4-Data Presentation, the next step will be compiling these Gini values into a consolidated summary table. For each province or territory and each population group, all the produced Gini coefficients would be aggregated into their respective mean and standard deviations. The formula for the mean of the Gini coefficient is as follows:

$$\bar{G} = \frac{\sum_{i=1}^{f(g)} g_i}{f(g)}$$

Where $f(n)$ is the number of values in the set n, g_i is the Gini coefficient of the i-th population in the set g , and $\bar{G}$ is the mean Gini coefficient of the set of Gini coefficients of $f(g)$ populations g.

The formula for the standard deviation for the set g, σ_g is as follows:

$$\sigma_g = \sqrt{\frac{\sum_{i}^{f(g)} (g_i - \bar{G})^2}{f(g) - 1}}$$

4. **Data Presentation**

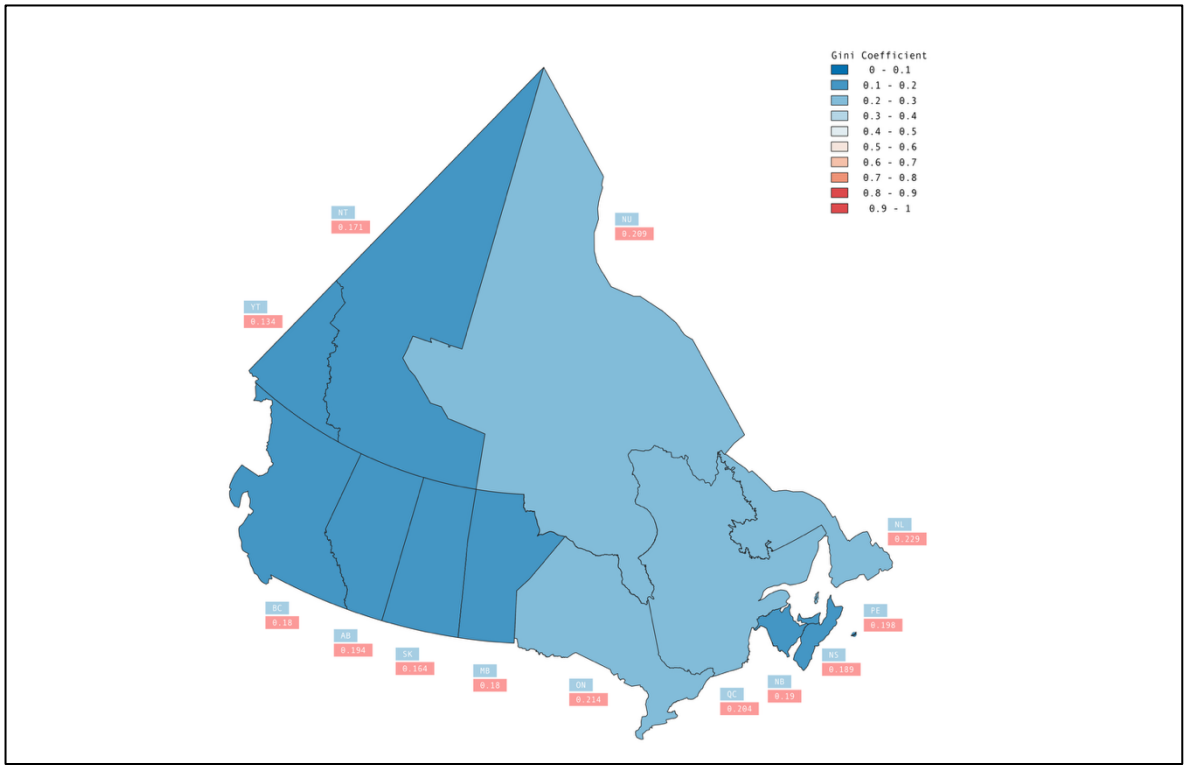

Figure 2. Gini Coefficient across Canada's Provinces and Territories for all ethnicities, with both immigration statuses and education levels.

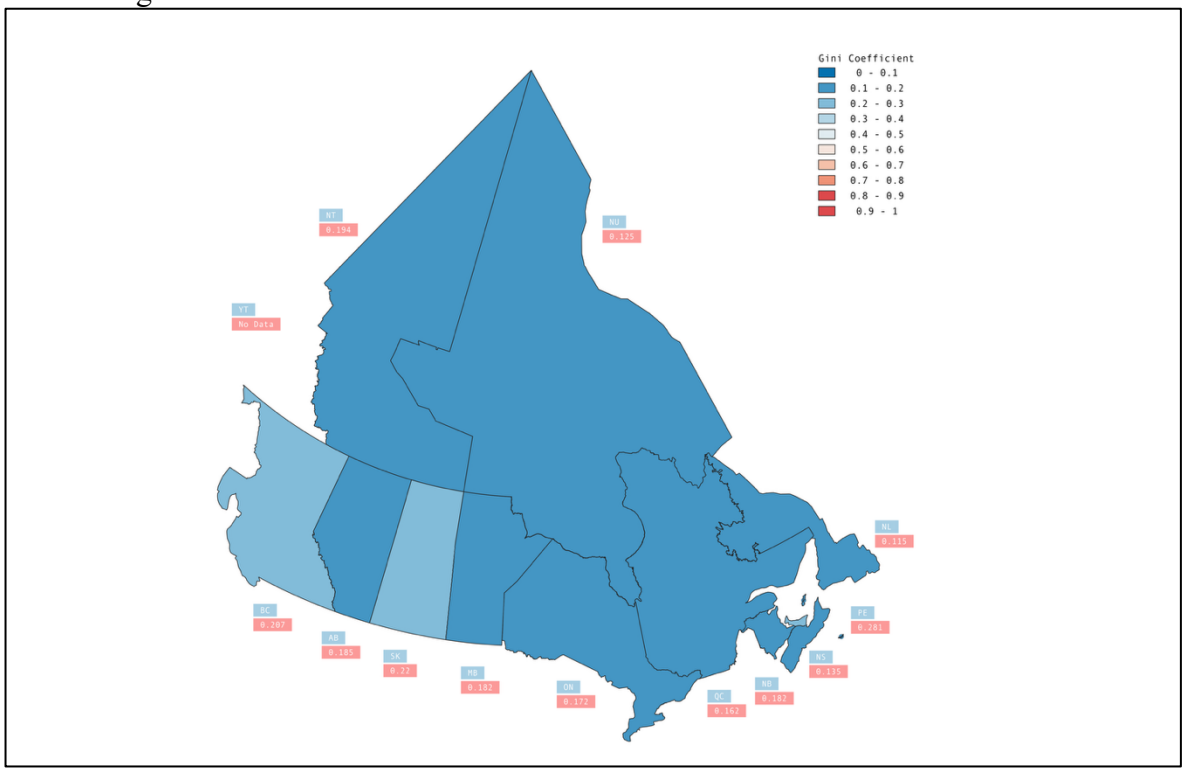

Figure 3. Gini coefficient across Canada's Provinces and Territories for the Arab population, with both immigration statuses and education levels.

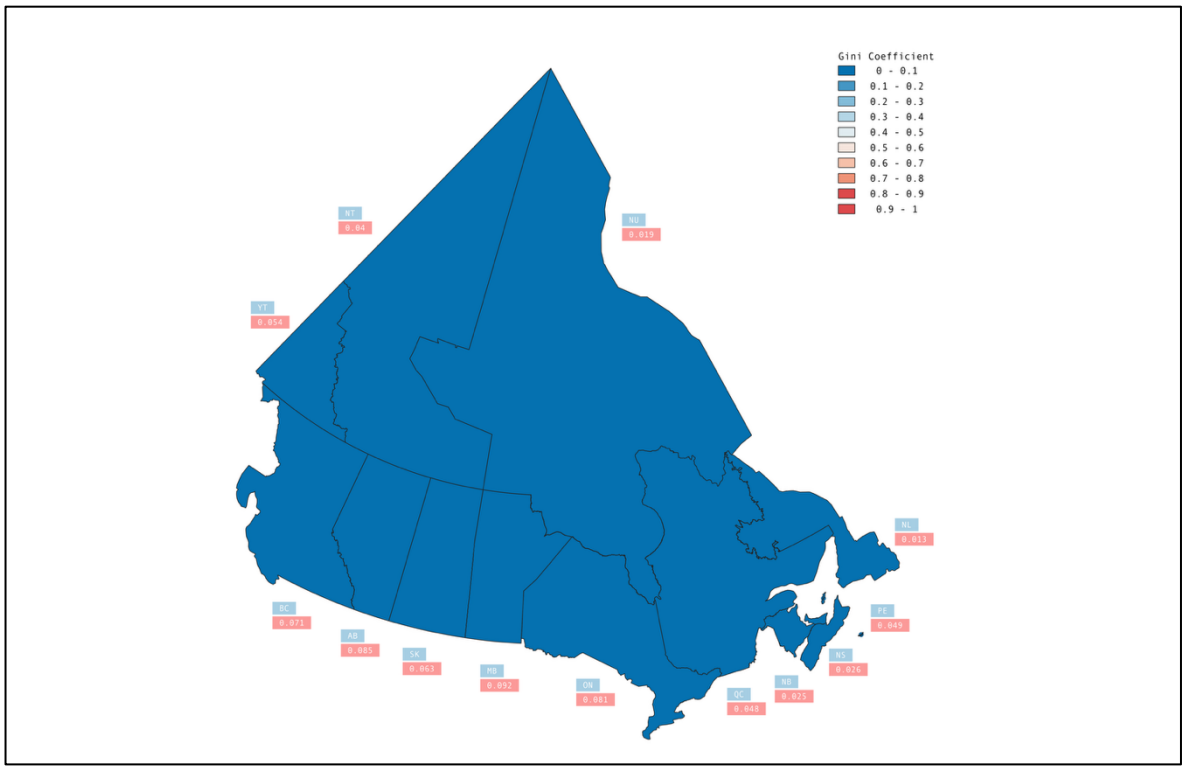

Figure 4. Gini coefficient across Canada's Provinces and Territories for the population of all ethnicities, with both immigration statuses, and bachelor's degree or higher.

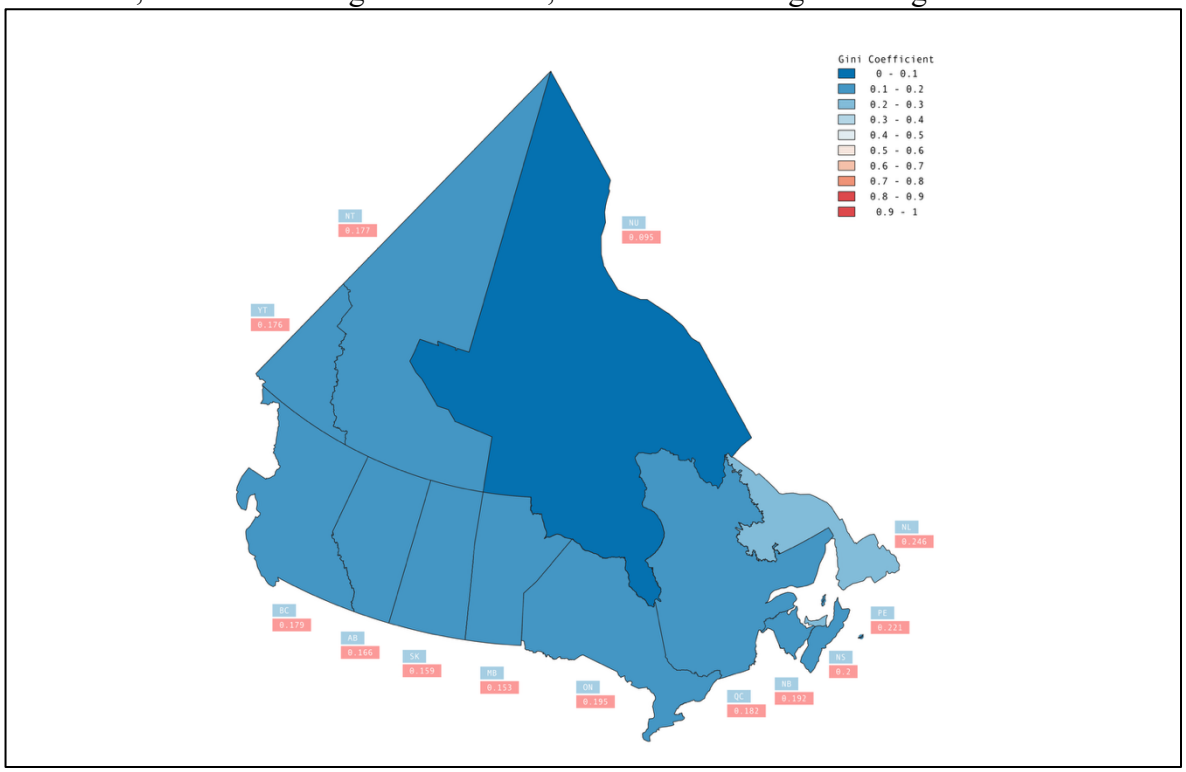

Figure 5. Gini coefficient across Canada's Provinces and Territories for the Black population, with both immigration statuses and education levels.

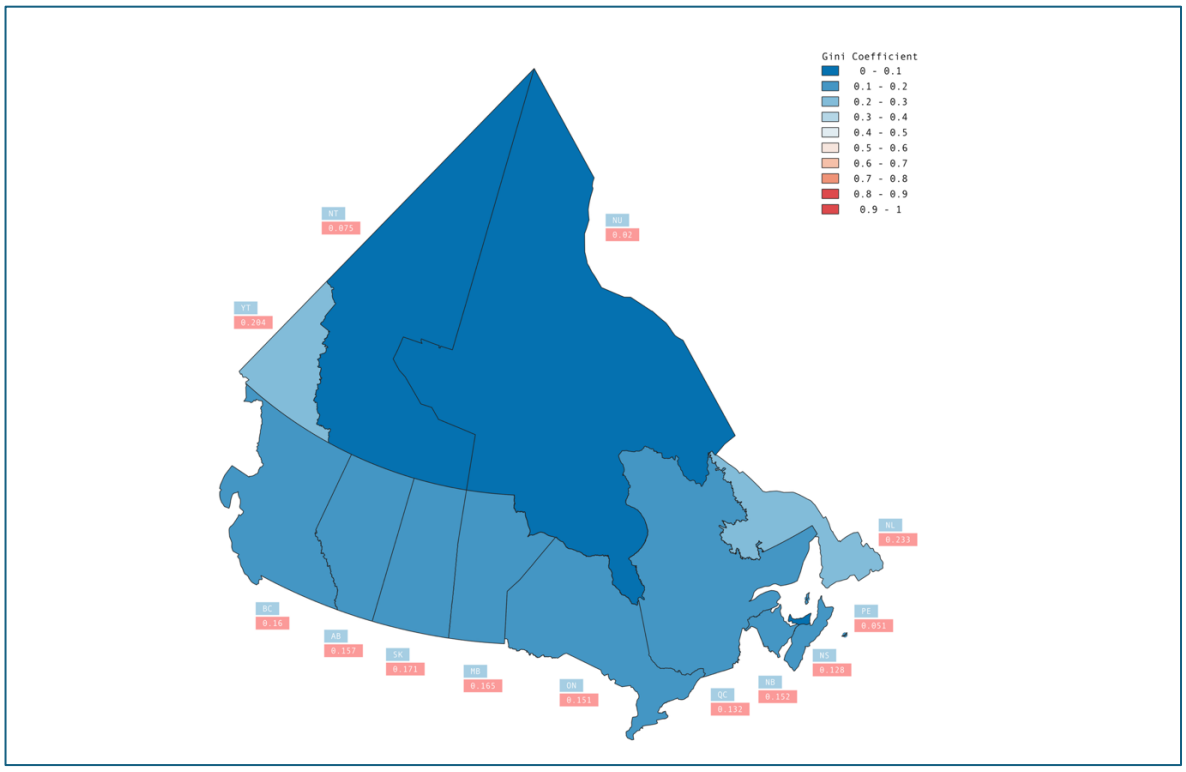

Figure 6. Gini coefficient across Canada's Provinces and Territories for the Chinese population, with both immigration statuses and education levels.

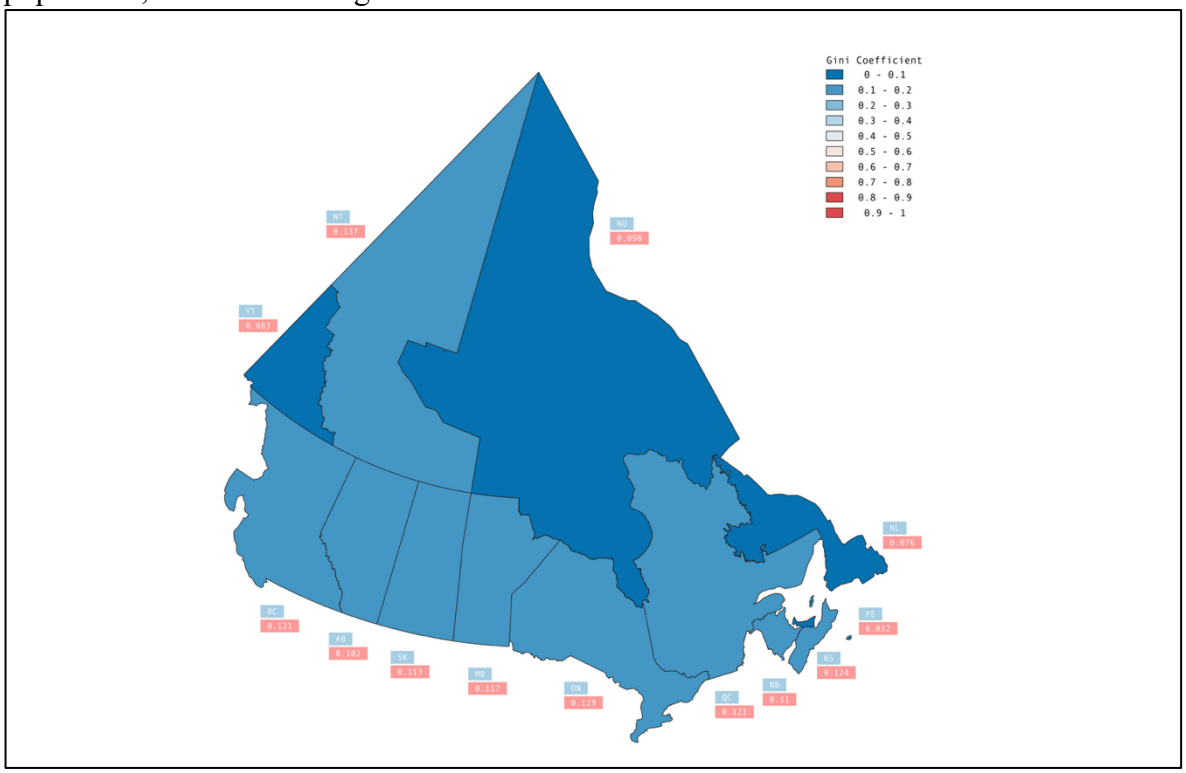

Figure 7. Gini coefficient across Canada's Provinces and Territories for the Filipino population, with both immigration statuses and both education levels.

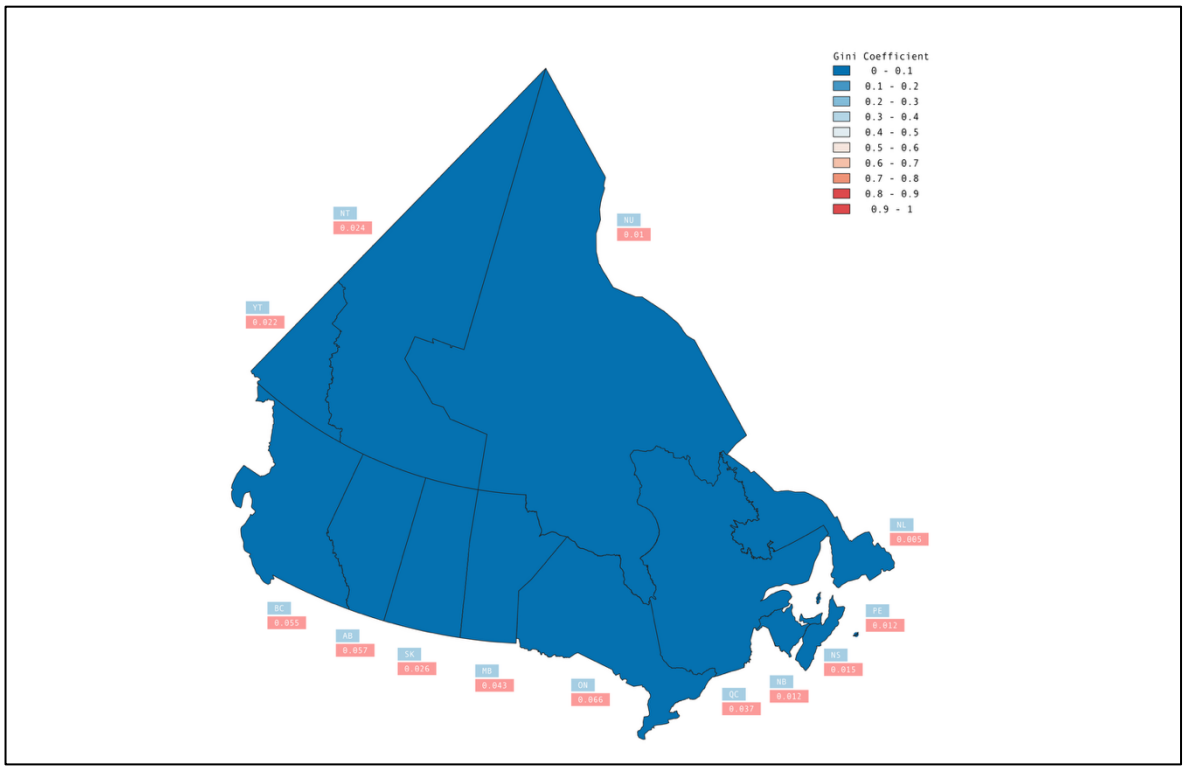

Figure 8. Gini coefficient across Canada's Provinces and Territories for the population of all ethnicities, with both immigration statuses and a high school diploma.

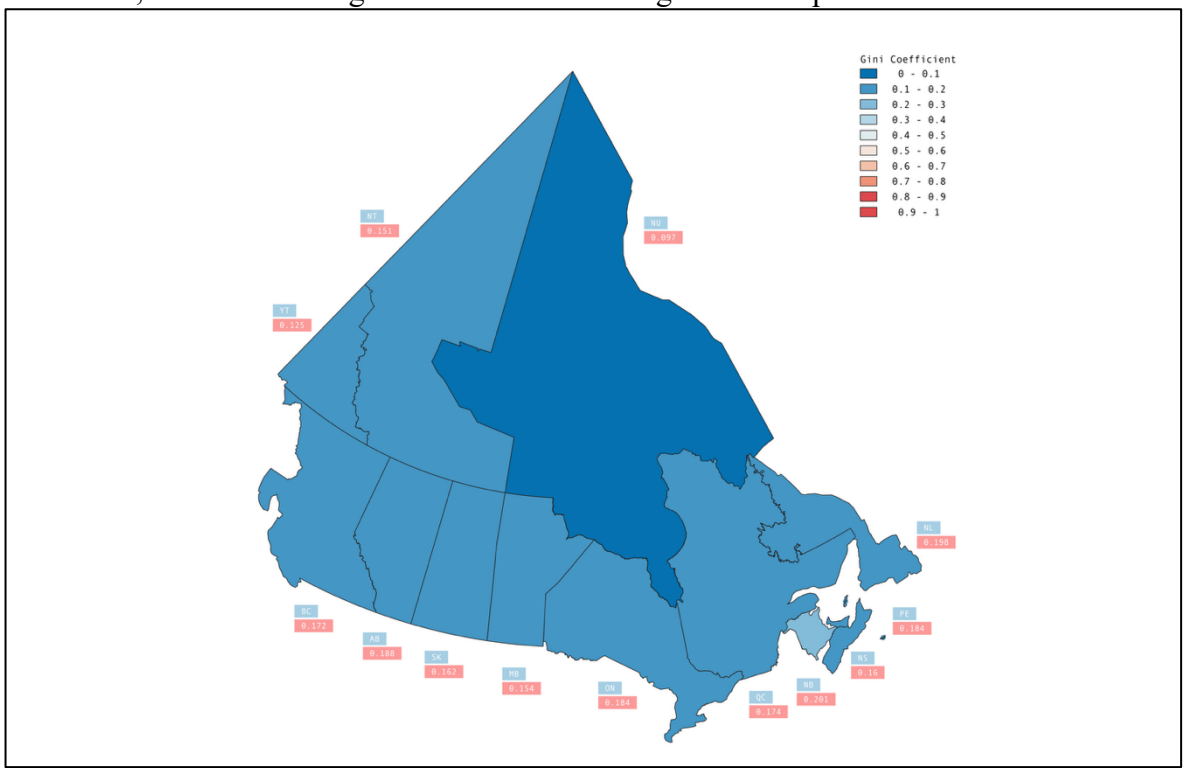

Figure 9. Gini coefficient across Canada's Provinces and Territories for the population of all ethnicities, with immigrant status and both education levels.

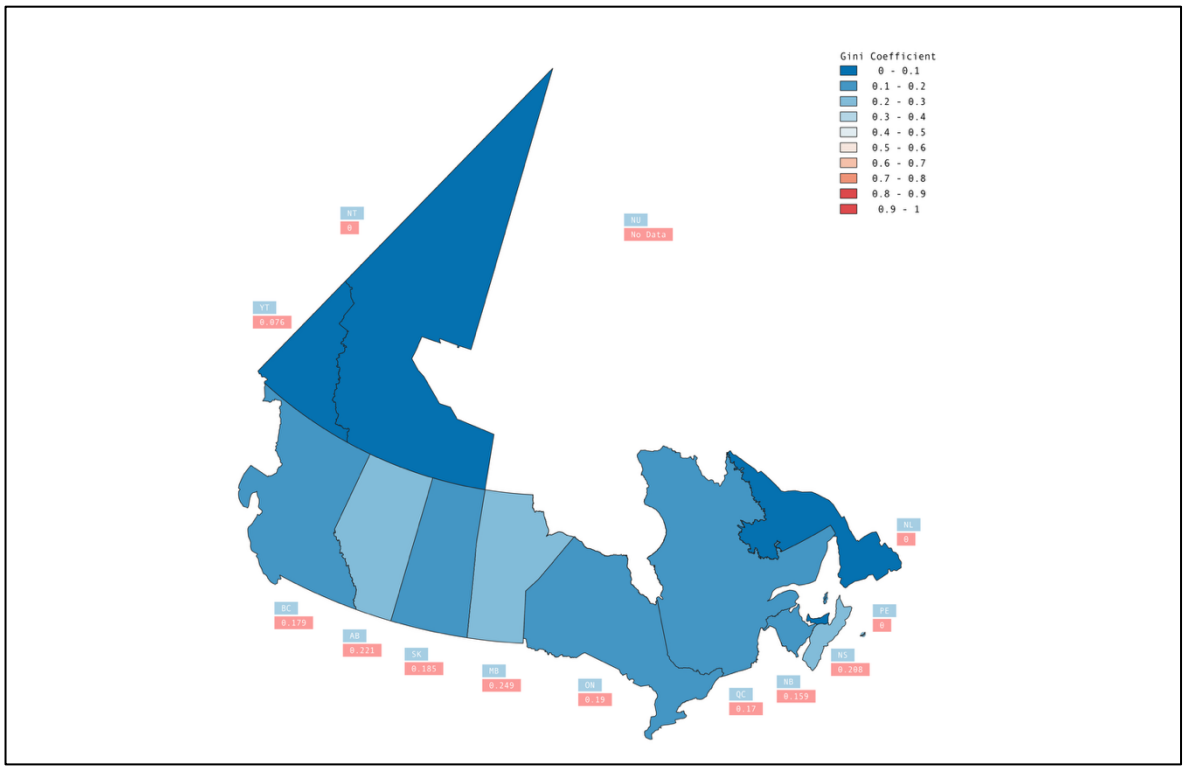

Figure 10. Gini coefficient across Canada's Provinces and Territories for the Japanese population, with both immigration statuses and both education levels.

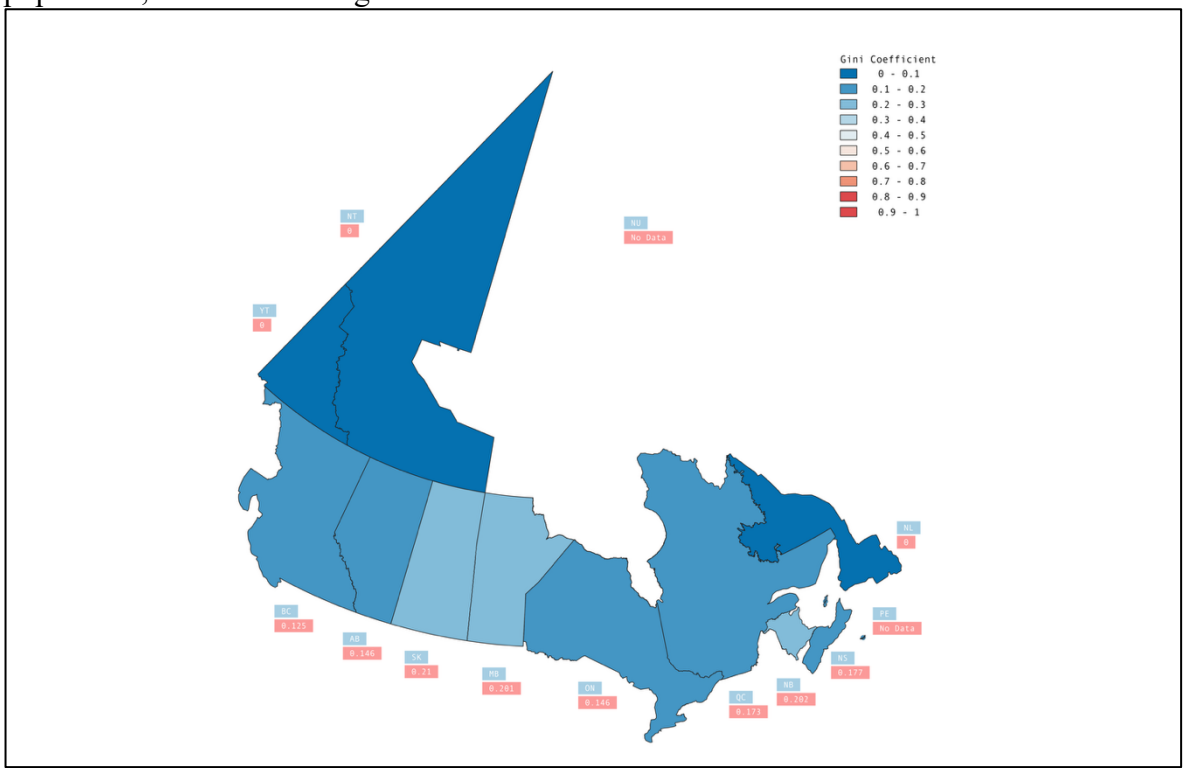

Figure 11. Gini coefficient across Canada's Provinces and Territories for the Korean population, with both immigration statuses and both education levels.

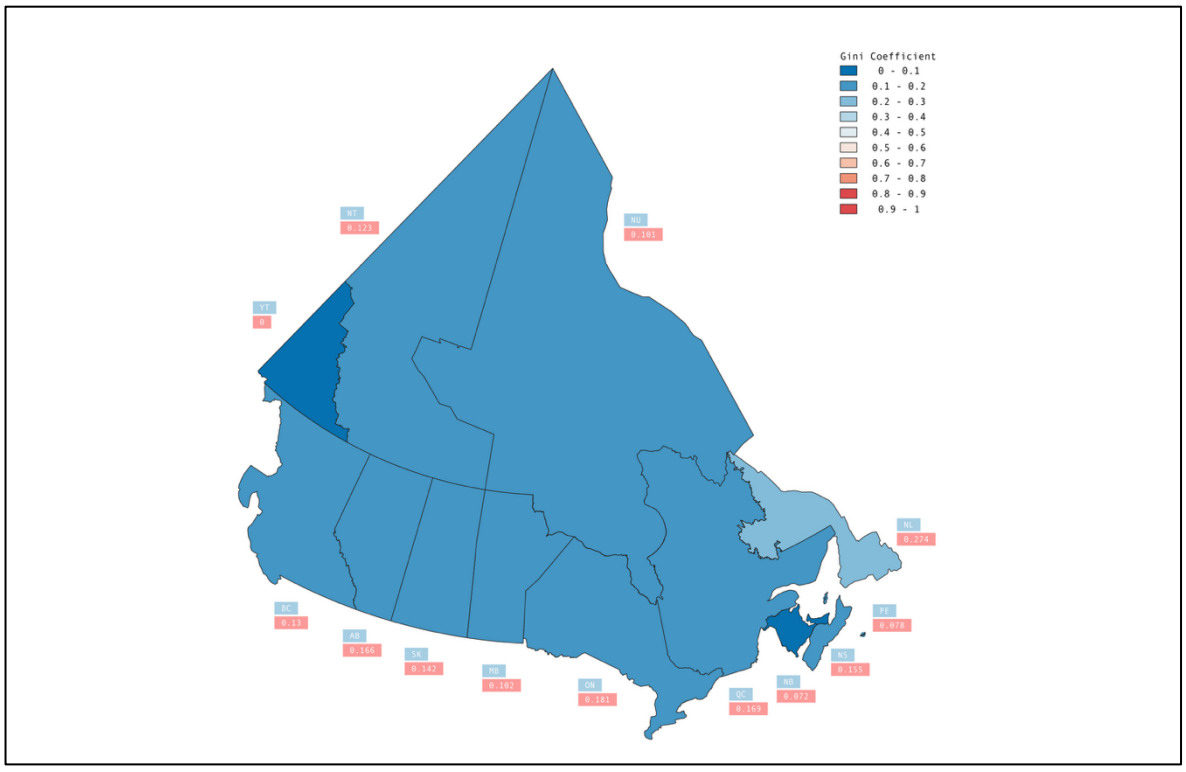

Figure 12. Gini coefficient across Canada's Provinces and Territories for the Latin American population, with both immigration statuses and both education levels.

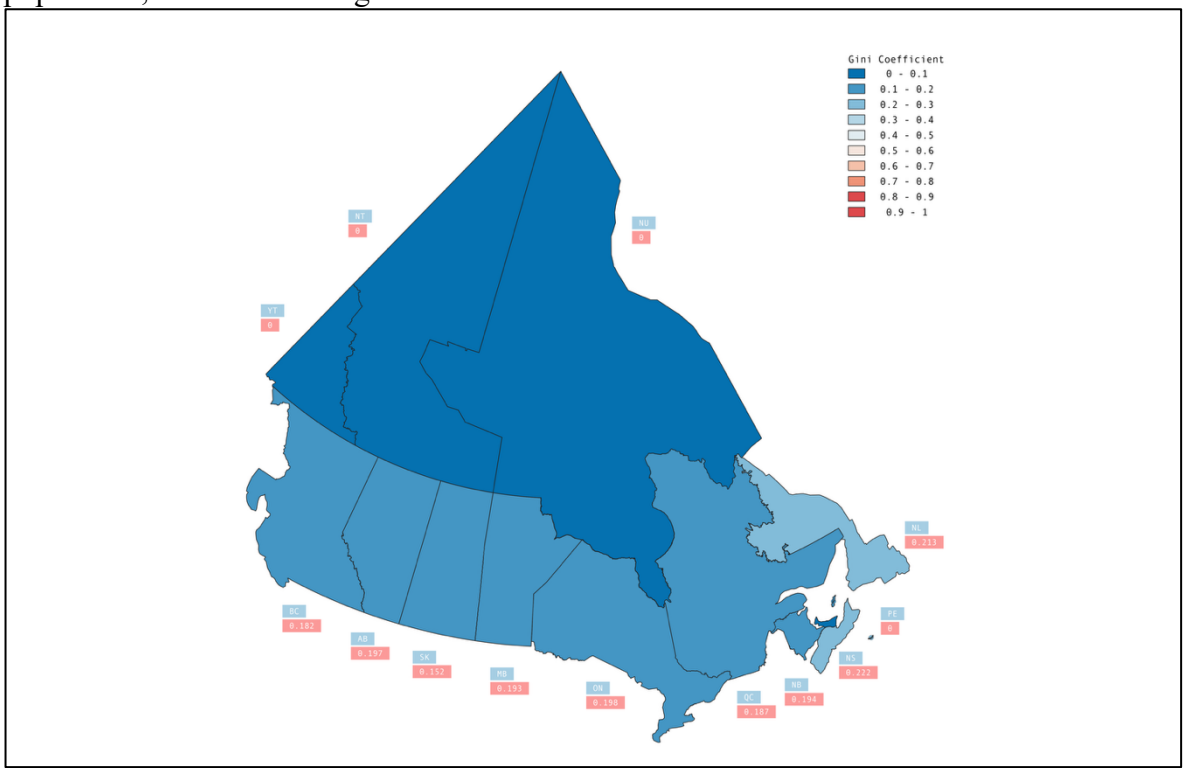

Figure 13. Gini coefficient across Canada's Provinces and Territories for the population with multiple visible minorities, with both immigration statuses and both education levels.

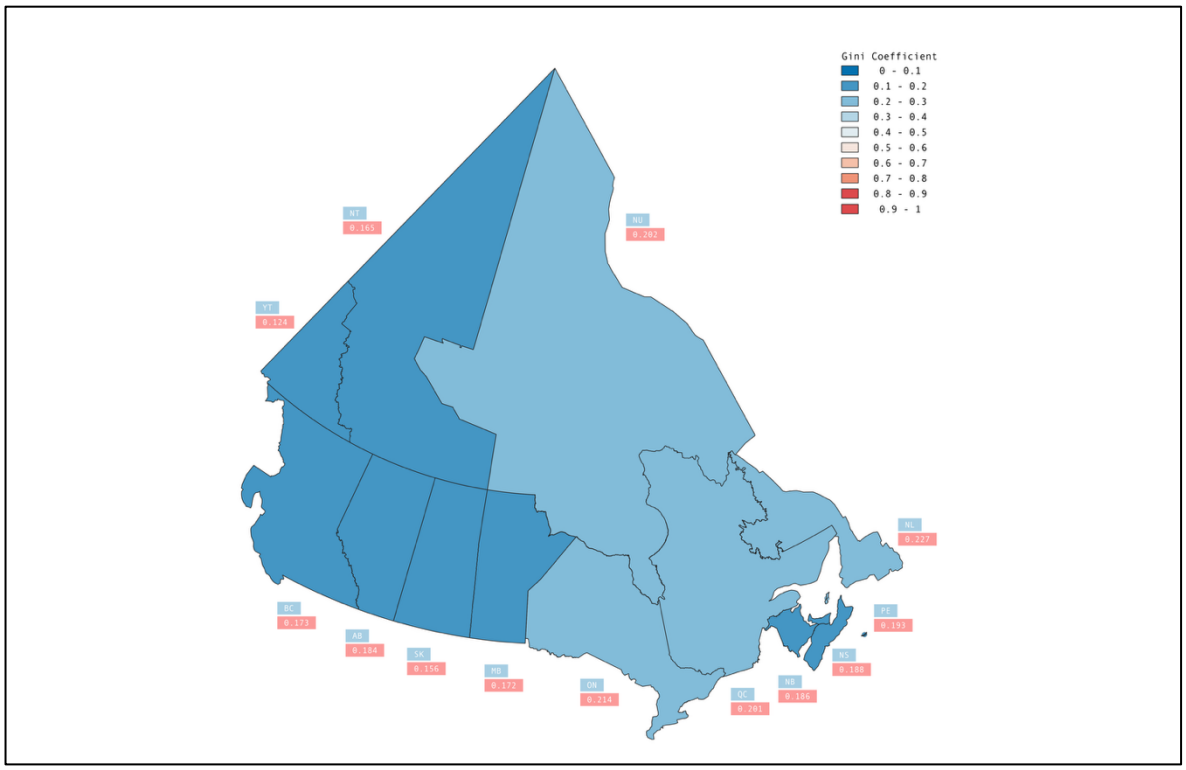

Figure 14. Gini coefficient across Canada's Provinces and Territories for the population of all ethnicities, with no immigrant status, and both education levels.

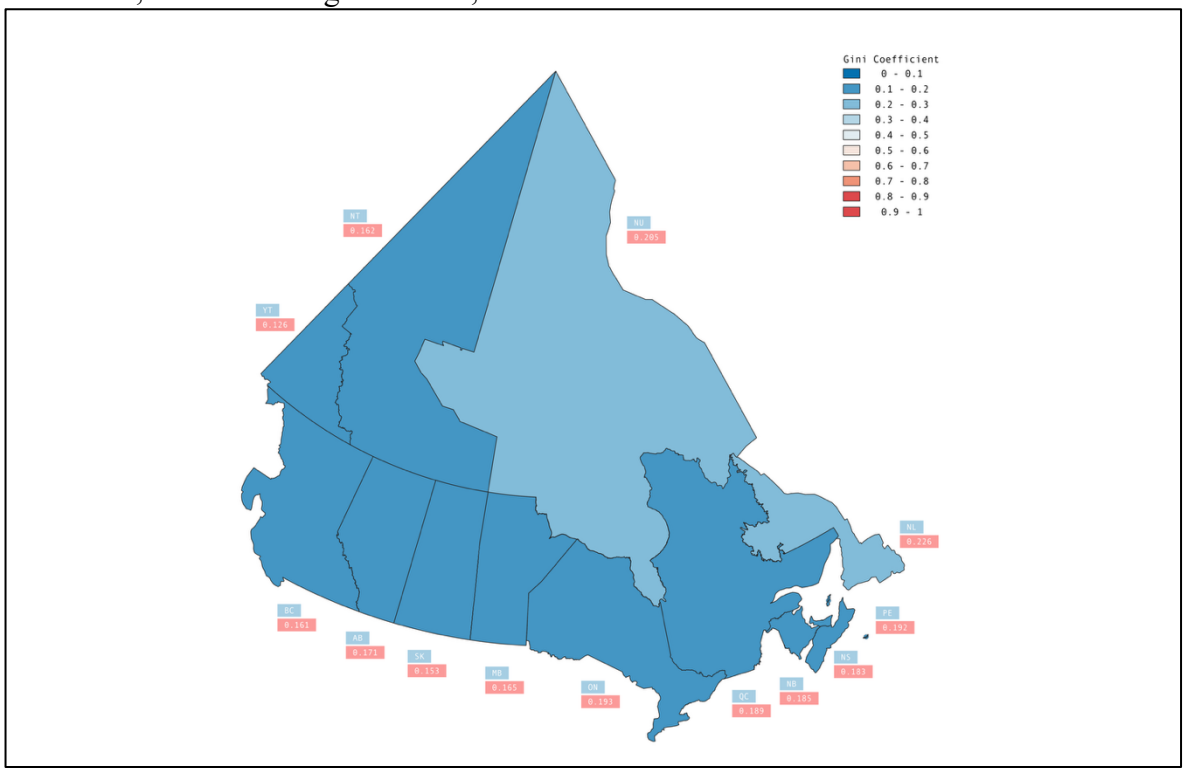

Figure 15. Gini coefficient across Canada's Provinces and Territories for the " not a visible minority" population, with both immigration statuses and both education levels.

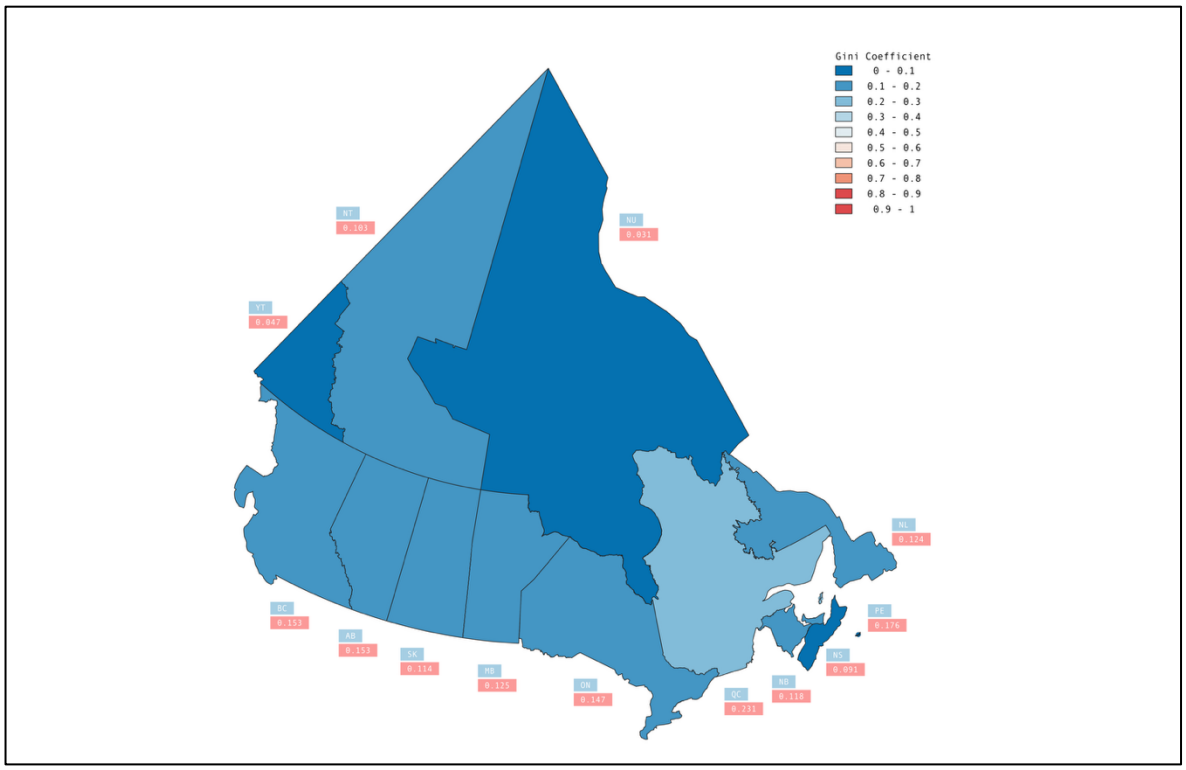

Figure 16. Gini coefficient across Canada's Provinces and Territories for the South Asian population, with both immigration statuses and both education levels.

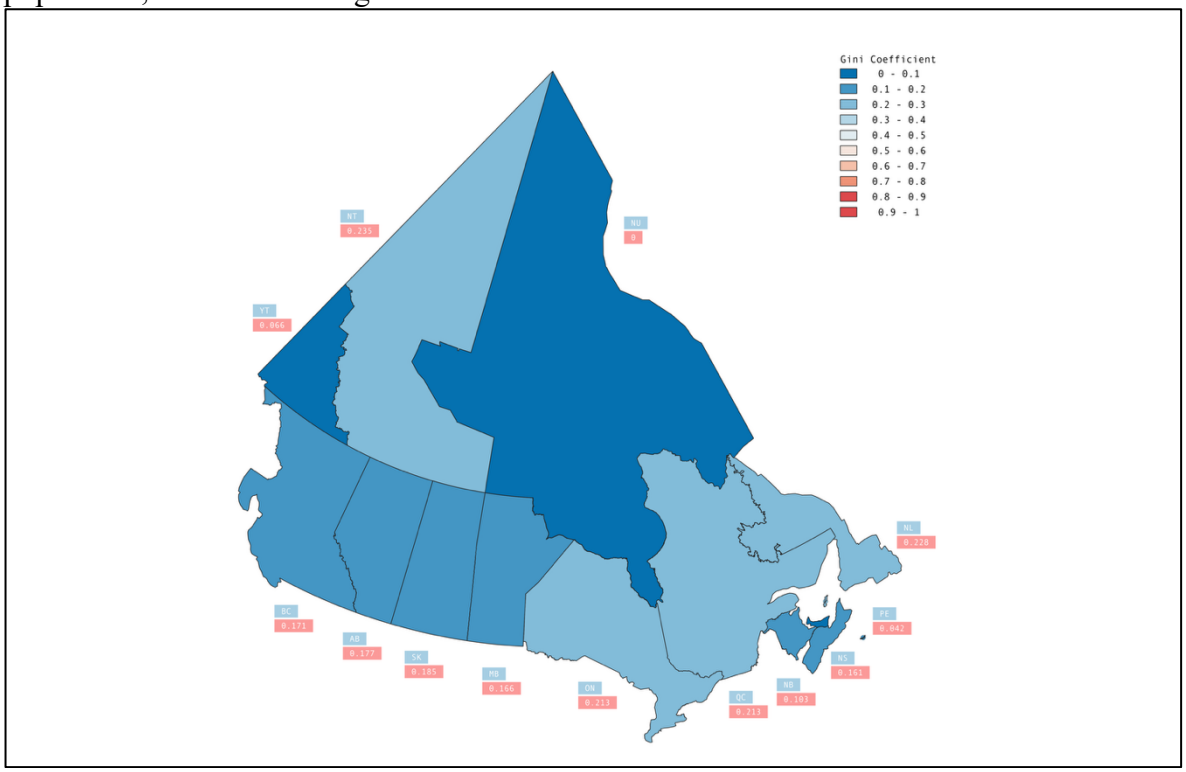

Figure 17. Gini coefficient across Canada's Provinces and Territories for the Southeast Asian population, with both immigration statuses and both education levels.

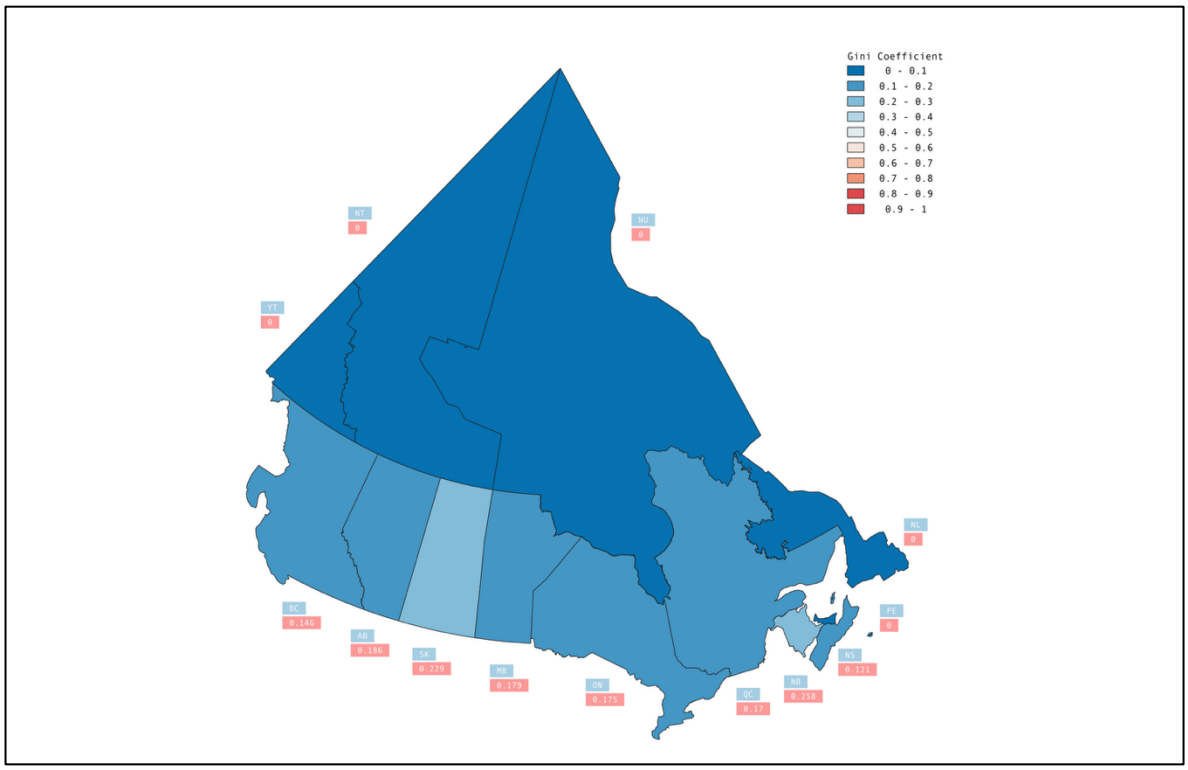

Figure 18. Gini coefficient across Canada's Provinces and Territories for the population with visible minorities, with both immigration statuses and education levels.

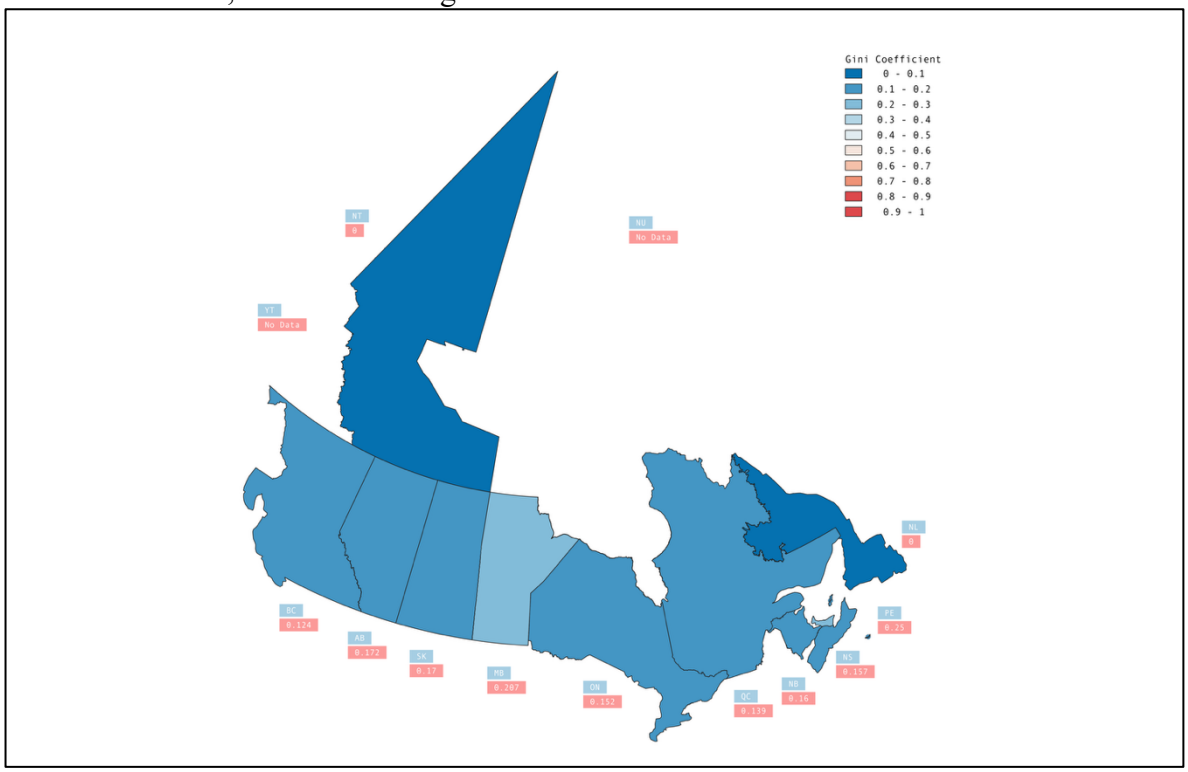

Figure 19. Gini coefficient across Canada's Provinces and Territories for the West Asian population, with both immigration statuses and both education levels.

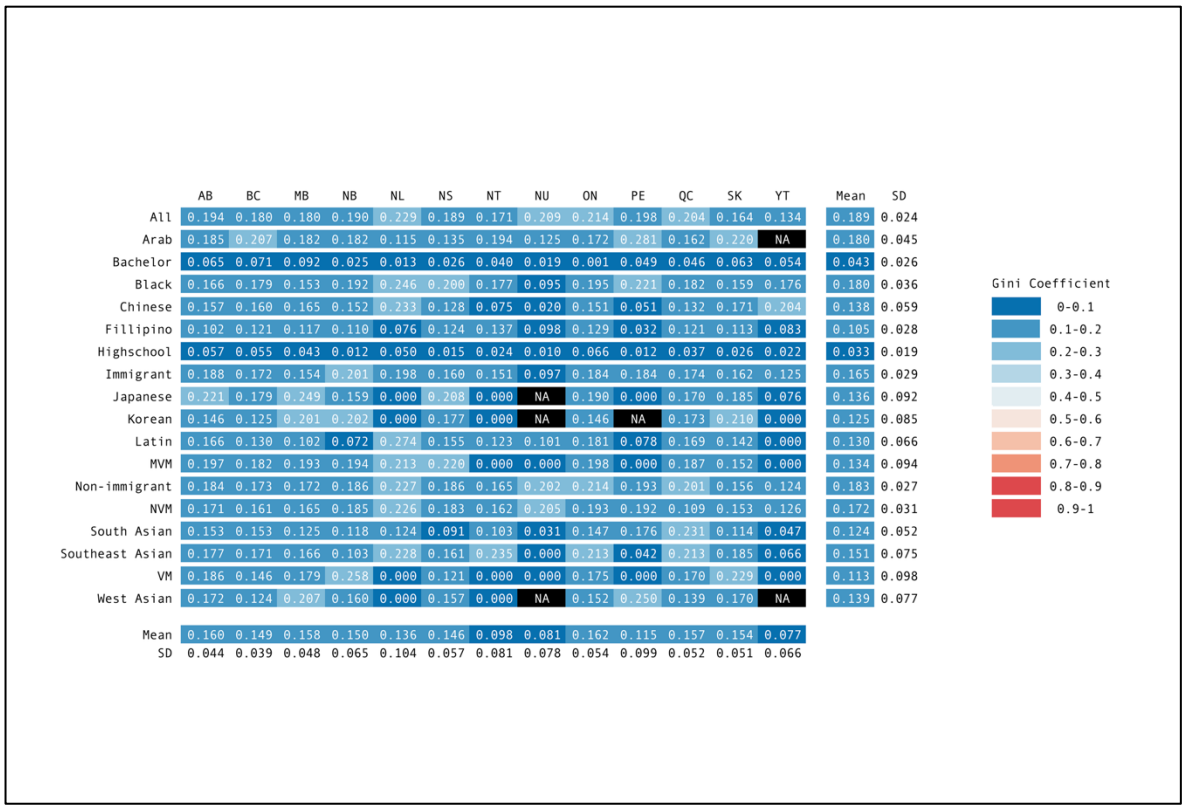

Figure 20. The Gini coefficient across Canada's provinces and territories by population ethnicity, immigration status, and education level, including mean and standard deviation of the Gini coefficient by population and province/territory.

Notes in reference to the figures 2 to 20:

1. The figures were produced in QGIS.
2. AB, BC, MB, NB, NL, NS, NT, NU, ON, PE, QC, SK, YT refers to these respective Canadian provinces and territories: Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, Northwest Territories, Nunavut, Ontario, Prince Edward Island, Quebec, Saskatchewan, Yukon.
3. "No Data" or "NA" indicates that there is no applicable individual for the subset of the evaluated population. For example, in Figure 16, "Gini coefficient across Canada's Provinces and Territories for the West Asian population, with both immigration statuses and both education levels.", the province "Nunavut" displays "No Data", indicating there is no applicable West Asian individual at either immigrant status or education level located in Nunavut for evaluation.
4. MVM, NVM, and VM refer to the following ethnicities, respectively: Multiple Visible Minorities, Not a Visible Minority, and Visible Minority.
5. In Figure 17," *The Gini coefficient across Canada's provinces and territories by population ethnicity, immigration status, and education level, including mean and standard deviation of the Gini coefficient by population and province/territory*", SD refers to the standard deviation of a set of Gini coefficients.
6. The Gini coefficient is rounded to the thousandths for viewing purposes.

5. Data Analysis

The following analysis will examine income inequality across Canada's provinces and territories, with different population groups' respective Gini coefficients, mean values, and standard deviation.

In regard of interpreting the Gini coefficient, a population with a Gini coefficient of 0 represents perfect equality, and 1 represents perfect inequality; the intermediate values can be interpreted as below (Lee, 2025):

1. Low inequality: $0 < G \leq 0.2$
2. Moderate inequality: $0.2 < G \leq 0.4$
3. High inequality: $0.4 < G \leq 0.6$
4. Very high inequality: $0.6 < G < 1$

In regard of interpreting the standard deviation of each population group's Gini coefficient, values below 0.05 are considered to indicate a tight distribution, and values greater than or equal to 0.05 are considered to indicate a wider distribution.

It must be noted that the color differences displayed in the above-mentioned choropleth graphs in 4-Data Presentation are reflected from the Gini coefficient, where each color change corresponds to a 0.1 equal interval from 0 to 1. Hence, the distribution analysis from the choropleth differs from the distribution of the standard deviation, where the former focuses on the distribution of the Gini coefficient on Canadian geography.

For the population groups that include all ethnicities, immigration statuses, and education levels are examined in Figure 2. "*Gini Coefficient across Canada's Provinces and Territories for all ethnicities, with both immigration statuses and education levels.*", the figure highlights Ontario, Quebec, Nunavut, and Newfoundland and Labrador in a lighter blue, contrasting with the Western and Atlantic regions. This highlights the concentrated regional differences in income inequality across Canada. Furthermore, as shown in Figure 17, provincial and territorial Gini coefficients in this group range from 0.134 to 0.229, with a national mean of 0.189 and a standard deviation of 0.024. The low mean and tighter distribution indicate generally low income inequality, with most provinces/territories falling within this category. Overall, this group displays concentrated regions of low and moderate income inequality, where the highest income inequality is clustered around Northern and Central Canada.

For the 2 population groups that include all ethnicities and both immigration status, one consisting of individuals with a bachelor's degree or higher and the other with a high school diploma, both respective choropleth maps (Figure 4. "*Gini coefficient across Canada's Provinces and Territories for population of all ethnicities, with both immigration statuses, and bachelor's degree or higher."* and Figure 7. "*Gini coefficient across Canada's Provinces and Territories for population of all ethnicities, with both immigration statuses and a high school diploma.")* display a consistent colour across provinces, suggesting uniform regions. Furthermore, in Figure 17, the high school diploma group has a national mean of 0.033 and a standard deviation of 0.026, while the bachelor's degree or higher group has a national mean of 0.043 and a standard deviation of 0.019. In both cases, the low mean value suggests low-income inequality, and the low standard deviation suggests a generally low-income inequality, with provinces/territories generally falling within this category. Overall, both high school diploma and bachelor's degree or higher groups display similar, uniform, low-income inequality across Canada.

For the 2 population groups that include all ethnicities and both education levels, one consisting of individuals with immigrant status and the other with non-immigrant status, their respective choropleth maps (Figure 8. “Gini coefficient across Canada’s Provinces and Territories for population of all ethnicities, with immigrant status and both education levels.” and Figure 13. “*Gini coefficient across Canada’s Provinces and Territories for the population of all ethnicities, with no immigrant status and both education levels*.”) displays a distinct distribution of the Gini coefficient. In Figure 8, the choropleth map for the population group with immigrant status displays a consistent blue across Canada, with only Nunavut and New Brunswick showing darker and lighter blue, respectively. The map features largely uniform regions with localized deviations. Figure 13, on the other hand, the choropleth map for the population group with non-immigrant status displays distinct regions of colour differences, with Central Canada, Nunavut, Newfoundland and Labrador coloured lighter blue, with Atlantic Canada, Western provinces and territories being darker blue, highlighting the concentrated, clustered regional differences. Furthermore, from Figure 19, the population group with immigrant status has a Gini national mean of 0.165 and the standard deviation of 0.029, and the population group with non-immigrant status has the Gini national mean of 0.183 and the standard deviation of 0.027. Both groups display similarity, tight spreads, and low means, suggesting generally low-income inequality, with provinces/territories generally falling within this category. Overall, both population groups display similar low income inequality with distinct patterns across provinces, where the population group with immigration status exhibits a uniform pattern with localized deviations in Nunavut and New Brunswick, and the population group with non-immigrant status displays concentrated, clustered regional differences, with Western and Atlantic Canada having better income equality.

For the population groups with both immigration status and education levels, specifically those identified by the following ethnicities: Arab, Black, Chinese, Filipino, Japanese, Korean, Latin, Multiple Visible Minorities, Not a Visible Minority, South Asian, Southeast Asian, Visible Minorities, and West Asians, their respective choropleth maps are shown in the figures. These include Figure 3: "Gini coefficient across Canada’s Provinces and Territories for the Arab population, with both immigration statuses and education levels," Figure 5: "Gini coefficient across Canada’s Provinces and Territories for the Black population, with both immigration statuses and education levels," Figure 6: "Gini coefficient across Canada’s Provinces and Territories for the Chinese population, with both immigration statuses and education levels," Figure 7: "Gini coefficient across Canada’s Provinces and Territories for the Filipino population, with both immigration statuses and both education levels," Figure 10: "Gini coefficient across Canada’s Provinces and Territories for the Japanese population, with both immigration statuses and both education levels," Figure 11: "Gini coefficient across Canada’s Provinces and Territories for the Korean population, with both immigration statuses and both education levels," Figure 12: "Gini coefficient across Canada’s Provinces and Territories for the Latin American population, with both immigration statuses and both education levels," Figure 13: "Gini coefficient across Canada’s Provinces and Territories for the population with multiple visible minorities, with both immigration statuses and both education levels," Figure 15: "Gini coefficient across Canada’s Provinces and Territories for the 'not a visible minority” population, with both immigration statuses and both education levels," and Figure 16: "Gini coefficient across Canada’s Provinces and Territories for the South Asian population.” with both immigration statuses and both

education levels.", Figure 17. "Gini coefficient across Canada's Provinces and Territories for the Southeast Asian population, with both immigration statuses and both education levels.", Figure 18. "Gini coefficient across Canada's Provinces and Territories for the population with visible minorities, with both immigration statuses and education levels.", Figure 19. "Gini coefficient across Canada's Provinces and Territories for the West Asian population, with both immigration statuses and both education levels.") exhibits properties which fall within the below general 3 categories:

1. The population groups from Japanese and Korean backgrounds' choropleth maps (Figures 10 and 11) generally show different shades of colour across Canada, exhibiting random, scattered regional differences.
2. The population groups from Arab, Black, Latin, Not a Visible Minority and South Asians backgrounds' choropleth maps (Figure 3, 5, 12, 15 and 16) generally shows a consistent colour, whether dark blue or light blue across the map with 1 or 2 provinces/territories shaded different, highlighting large, uniformed regions with localized deviations.
3. The population groups from Chinese, Filipino, Multiple Visible Minorities, Southeast Asian, Visible Minority, and West Asian backgrounds' choropleth map (Figures 6, 7, 13, 17, 18, and 19) generally shows multiple regions shaded similarly, usually being Northern against Southern Canada, highlighting concentrated, clustered regional differences.

Furthermore, from Figure 20, the mean Gini coefficient of the mentioned population groups all falls below 0.2, being 0.105, 0.113, 0.124, 0.125, 0.130, 0.134, 0.136, 0.138, 0.139, 0.151, 0.172, 0.180, 0.180 for the Filipino, Visible Minority, South Asian, Korean, Latin, Multiple Visible Minorities, Japanese, China, West Asian, Southeast Asian, Not a Visible Minority, Arab, Black backgrounds respectively, with their standard deviations falls within 2 groups:

1. The population groups with Filipino, Not a Visible Minority, Arab, and Black backgrounds have a standard deviation below 0.05, being 0.028, 0.031, 0.045, and 0.036, respectively. These groups have a similar low mean and a tighter distribution, indicating a generally low income inequality with provinces/territories generally falling within this category.
2. The population groups with Visible Minority, South Asian, Korean, Latin, Multiple Visible Minorites, Japanese, Chinese, West Asian and Southeast Asian backgrounds have the standard deviation above 0.05, being 0.098, 0.052, 0.085, 0.066, 0.094, 0.092, 0.059, 0.077, 0.075 respectively, These groups have a similar low mean and a wider distribution, indication a generally low income inequality with provinces/territories possibly fall outside this category.

Overall, these population groups display similar general low income inequality with unique patterns across Canada that depend on their background, ranging from unique scattered regional differences, concentrated clustered regional differences, or a uniform pattern with some local deviations.

In addition, from Figure 20, the value for the Gini index of each population group can be seen by province, one by one. On average, all provinces have a mean Gini index of below 0.2, the lowest being Yukon with 0.077 and the highest being Ontario with 0.162. However, the standard deviation of each provincial/territorial mean falls within the following 2 categories:

1. Alberta, British Columbia, and Manitoba have their standard deviation below the value of 0.05, being 0.044, 0.039, and 0.048, respectively. These provinces have a similar low mean and tighter distribution, indicating a generally low income inequality with most population groups in these provinces falling within this category.
2. New Brunswick, Newfoundland and Labrador, Nova Scotia, Nunavut, Ontario, Prince Edward Island, Quebec, Saskatchewan, Yukon have their standard deviation above the value of 0.05, being 0.065, 0.104, 0.057, 0.081, 0.078, 0.054, 0.099, 0.052, 0.051, 0.066, respectively. These provinces/ territories have a similar low mean and a wider distribution, indicating that a generally low income inequality with population groups in these provinces/territories possibly falls outside this category.

Overall, most provinces/territories have generally low income inequality; however, only the Western provinces have population groups that generally fall within this category.

6. Conclusion

The report attempts to examine whether Canada's global reputation as the "Land of Opportunity" is true. By connecting income equality as a measure for opportunity, the report hopes to provide empirical evidence on the accessibility of these opportunities across different population groups in each province, territory, and Canada as a whole.

From 5-Data Analysis, with most population groups and provinces/territories consistently falling below the Gini value of 0.2, it can be said that Canada, in general, experiences low income inequality and thus low inequality. While most provinces, notably Western Canada, fall in this category, Central Canada, on the other hand, frequently displays deviations. On the other hand, different population groups experience case-by-case regional patterns. For example, population groups with either education level experience uniform low inequality across Canada, while population groups with either immigrant status experience both uniform and clustered differences across Canada. Furthermore, population groups with different ethnic backgrounds experience uniform, clustered, and scattered regional differences, differ by each background individually. In addition to the geographical distribution of the Gini coefficient, the standard deviation results indicate that some provinces have population groups that deviate outside of low income inequality range. Overall, even though Canada in general exhibits low income inequality, this is not consistent, as some population groups in some provinces will fall outside this category.

Overall, this report has successfully produced empirical evidence for Canada being the "Land of Opportunity" or specifically, the presence of opportunity in Canada, that income inequality is generally low across the country. However, the report fails to declare absolute equal accessibility of these opportunities. Instead, the report identifies that opportunity isn't evenly experienced by all population groups. The reports highlight that opportunities accessibility is most often disproportionately clustered into regions, such as Northern against Southern Canada, or uniform across the country with local deviations, usually in Canadian territories. The report also fails to measure other aspects of opportunities, such as social mobility. Hence, for the above reasons and the further reasons listed in Weaknesses of 7-Strengths, Weaknesses and Extensions, the report's conclusion lacks concrete support and doesn't accurately represent equality or opportunity in Canada. As a result, the conclusion lacks confidence, and the report are not to be used as further reference material.

7. Strengths, Weaknesses, and Extension

1. Strengths:

The report was able to examine the population across the entirety of Canada and into each province and territory, respectively. This allows for a more comprehensive national and provincial/territorial comparison. In addition, with the assistance of QGIS, the scope of the report allows qualitative observation on the spread and the pattern of the Gini coefficient across Canada. This supports the analysis by identifying the general spread of these values and categorizing them into uniform, clustered, and scattered differences.

In addition, the scope of evaluation includes multiple different population groups from multiple ethnicities, study levels, and immigrant statuses. This allows the report to perform a comprehensive dissection into the Canadian population and identify meaningful differences in equality between groups. This supports the analysis by highlighting some demographic differences across Canada and evaluating the opportunity in each province.

2. Weakness:

The dataset before data sampling, sourced from Statistics Canada's "Employment income statistics by visible minority, immigrant status, highest level of education and location of study: Canada, provinces and territories," is a sample from the 2021 Census, with 25% of the sample data (Statistics Canada, 2023b). While the sample dataset provided sufficient data for the report to analyze and produce the required Gini coefficient, the sample dataset omits 75% of the population. This limited sample could potentially be skewed or underrepresent a population group in a province/territory, which potentially decreases or increases the group's Gini coefficient and thus the national mean and standard deviation. As a result, the limited data sample reduces the overall confidence in the report's conclusion. This weakness can potentially be addressed by utilizing a different data source.

In addition, in the original data set from Statistics Canada, the income for each population group is aggregated in the attribute value "Average Employment Income ($)". While it is possible to produce the Gini coefficient by assuming that each individual makes the same income, as mentioned in 3-Methods and Data Sampling, the process removes variation and straightens out portions of the Lorenz curve. For example, for the population group with a Chinese background in Prince Edward Island, there exist 4 subsets with their respective number of individuals and average income (Statistics Canada, 2023b).

1. The subset with non-immigrant status and a high school diploma: 30 individuals and $30000 average employment income.
2. The subset with non-immigrant status and bachelor's degree or higher: 50 individuals and $38000 average employment income.
3. The subset with immigrant status and a high school diploma: 125 individuals and $22800 average employment income.
4. The subset with immigrant status and a bachelor's degree or higher: 665 individuals and $33700 average employment income.

Due to approximately 76.3% of the population with a Chinese background in Prince Edward Island being assumed to make the same income of $33700, the Lorenz

curve of this population set is straightened out in the middle and produces the Gini coefficient of only 0.051. While this effect is less pronounced when there are more population subsets and the number of individuals of these subsets is more proportional, the cases such as above, this effect skews the Gini coefficient towards zero. As a result, the nature of this dataset being aggregated reduces confidence in this report's conclusion. This weakness can potentially be addressed by a different data source where individual income is displayed.

Furthermore, as briefly mentioned above, population groups from ethnic backgrounds, when evaluating for the provincial/territorial Gini coefficient, would only contain 4 subsets. On the contrary, population groups with either immigrant status or education level would have 26 subsets to evaluate for the Gini coefficient. While the Gini coefficients are analyzed under the same interpretation, the disproportional number of subsets in these population groups during evaluation causes inconsistency and reduces confidence in the analysis of this Gini coefficient. The reduced confidence in the analysis further propagates to the reduced confidence in this analysis report. This weakness can potentially be addressed by introducing new attributes and attribute values, and hence, more population groups for evaluation. While this approach would maintain the disproportional number of subsets, it would increase the subsets when evaluating the Gini index, and it would increase confidence in the analysis result and the report's conclusion.

Finally, when calculating the national Gini coefficient mean for a population group or a Gini mean of a province and its respective standard deviation in Figure 20, some set of Gini coefficients contains the value 0 or "No Data". Despite the established definition of the Gini coefficient of 0 indicating perfect equality, all cases of 0 in this report are due to some provinces/territories of a population group containing a single subset. Under the established assumption that all individuals from the same subset would make the subset's average income, this results in no variation on the Lorenz curve and artificially produces a value of 0. On the one hand, the result of "No Data" is due to the provinces for some population group containing no subset, hence no population to produce the Gini coefficient. Both these issues are caused by the limited scope of the sampled data set and lead to an inconsistent Gini coefficient mean and standard deviation. The inconsistency reduces the confidence in the data's analysis and hence the report's conclusion.

3. Extension:

Before producing the modified Python function mentioned above in 3 – Methods and Data Sampling to optimize the large number of population input when calculating the Gini coefficient, the following method was theorized.

For example, a population with subset A making $10000 on average with 100000 people, subset B making $50000 on average with 25000 people, and so on. By assuming that all individuals of a subset make their subset's average income, with a function in Python following the operations of the outlined function for the Gini coefficient, said function would have to go through every individual of each subset.

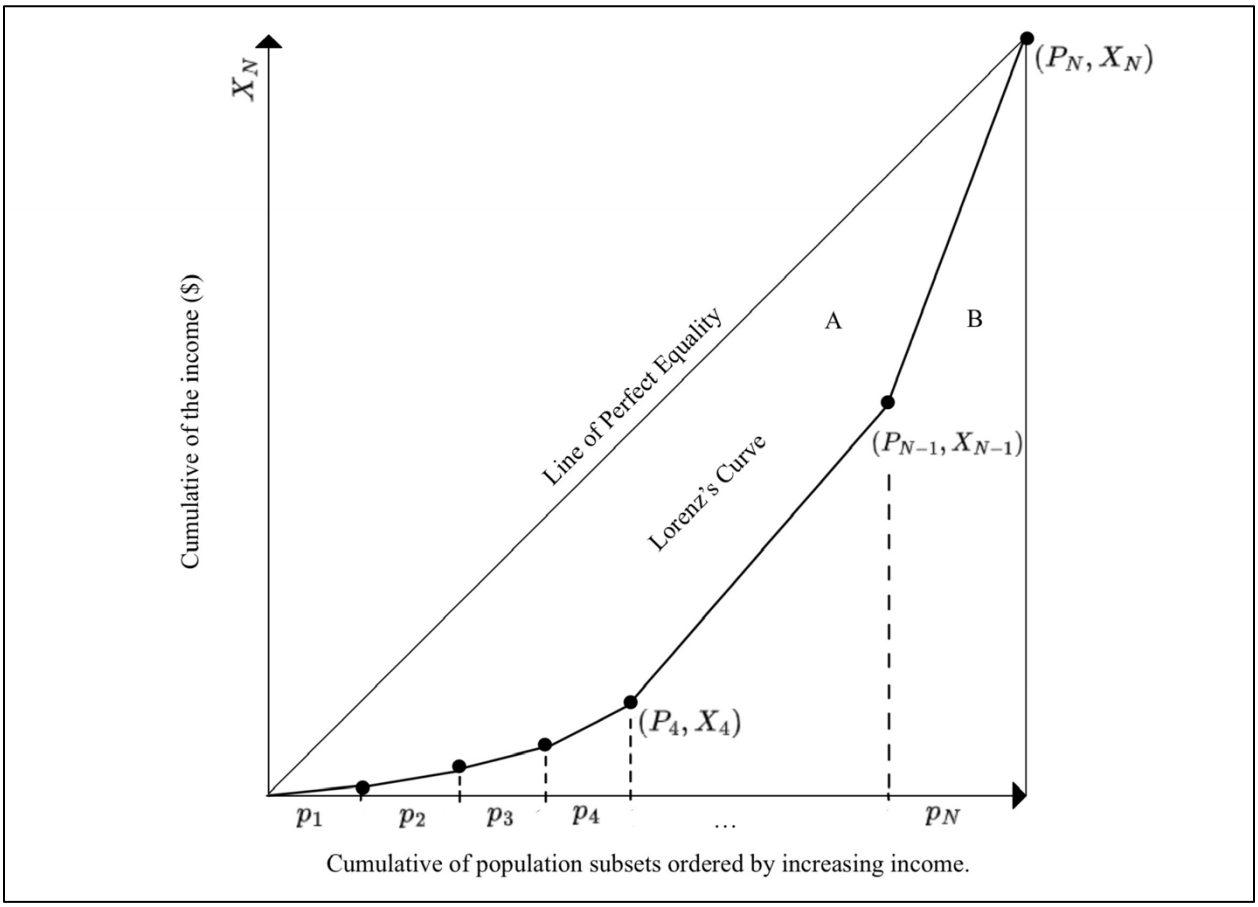

Figure 2. The graphical representation of the adjusted formula for the Gini coefficient, where $X_k = \sum_{j=1}^{k} x_j$ is the cumulative income of the population subsets up to the k-th subset, $P_k = \sum_{j=1}^{k} p_j$ is the cumulative population up to k-th subset, x_j is the income of the j-th subset, p_j is the population of the j-th subset, N is the total number of subsets, with the subset ordered with $0 \leq x_1 \leq x_2 \leq \cdots \leq x_{N-1} \leq x_N$.

Hence, by applying the same logic of approximating the Lorenz's curve as trapezoids and the area below the line of perfect equality as a right-angle triangle, the area of B can be calculated as a total area of $N-1$ trapezoid(s), each with a base of, the parallel sides being X_k and X_{k+1}, and a right-angle triangle with a base of p_1 and height of X_1, which simplified to (Kanel, 1993):

$$B = \frac{p_1 X_1}{2} + \frac{p_2(X_1 + X_2)}{2} + \cdots + \frac{p_{N-1}(X_{N-2} + X_{N-1})}{2} + \frac{p_N(X_{N-1} + X_N)}{2}$$

$$= \frac{p_1 X_1 + p_2 X_1 + p_2 X_2 + \cdots + p_{N-1} X_{N-2} + p_{N-1} X_{N-1} + p_N X_{N-1} + p_N X_N}{2}$$

$$= \frac{X_1(p_1 + p_2) + X_2(p_2 + p_3) + \cdots + X_{N-1}(p_{N-1} + p_N) + p_N X_N}{2}$$

$$= \frac{\sum_{j=1}^{N-1} X_j(p_j + p_{j+1}) + p_N X_N}{2}$$

For simplification, with $A + B = \frac{nY_n}{2} = \frac{P_N X_N}{2}$, the revised formula for the Gini coefficient would be:

$$G = 1 - \frac{B}{A+B} = 1 - \frac{\sum_{j=1}^{N-1} X_j(p_j + p_{j+1}) + p_N X_N}{P_N X_N}$$

However, when the formula is converted to a Python function and implanted for different groups, this formula did not produce reliable or accurate results. Unfortunately, further optimization is needed to fix any logic errors in this adaptation.

On the hand, the report could implement a different metric to measure income inequality. While the formula of the Gini coefficient could produce a value between 0 and 1 to indicate the income equality of a province, it's not a good indicator to use for comparison between the different provinces. For example, given 2 populations, the first with a Gini coefficient of 0.1 and the second with 0.2, which would place both populations as populations with low income inequality. While the statement that the latter population faces more income inequality would be true, the statement that the second population faces double the income inequality would not necessarily be true. In addition, the middle of the population typically has consistent income and often influences the produced Gini coefficient to show that the population has higher income equality (Cobham, 2013). Hence, in conjunction with the Gini coefficient, the Palma ratio would also be used.

The Palma ratio is another measure of income equality, dividing the income of the richest 10 percent by the income of the poorest 40 percent. While the Palma ratio can't produce a number in a set range, it does allow for numerical comparison and shows the income disparity between the top and the bottom of the population (Cobham, 2013).

The implementation of Palma ratio with or instead of the Gini coefficient during analysis can potentially produce different conclusions or provide confidence of this report's conclusion.

8. References

1. Christian, D. (n.d.). *Lorenz Curve*. MathWorld--A Wolfgram Resource, created by Eric W. Weisstein. https://mathworld.wolfram.com/LorenzCurve.html
2. Cobham, A. (2013, April 5). *Palma vs Gini: Measuring post-2015 inequality | center for global development*. Center for Global Development. https://www.cgdev.org/blog/palma-vs-gini-measuring-post-2015-inequality
3. InvestImmi - Canadian Immigration Legal Services. (2015). *Why Canada*. Life in Canada. https://www.canadaimmigration.services/Home/WhyCanada
4. Kanel, N. R. (1993). Lorenz curve and Gini coefficient: Conceptual considerations. *Economic Journal of Nepal*, *16*(4), 221–230. https://doi.org/10.3126/ejon.v16i4.71678
5. Lee, S. (2025, May 25). *Understanding gini coefficient in development*. Number Analytics // Super Easy Data analysis tool for Research. https://www.numberanalytics.com/blog/ultimate-guide-gini-coefficient-international-development-policy
6. Statistics Canada. (2023a, September 29). 2021 Census - Boundary files. https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21
7. Statistics Canada. (2023b, October 4). Table 98-10-0399-01: Employment income statistics by visible minority, immigrant status, highest level of education and location of study: Canada, provinces and territories. https://doi.org/10.25318/9810039901-eng