**A blue text on a white background

Description automatically generated**

HUMBER INSTITUTE OF TECHNOLOGY

AND ADVANCED LEARNING

(HUMBER POLYTECHNIC)

**Big Data 2 - BIA-5303-0LB**

**Group Project: Data visualization for decision making**

**Submitted by:**

Harshad Valiya – N01650172

Abhishek Kumar Yadav – N01649428

Palash Sharma – N01649355

Shobhit Batra – N01648961

Sayam Kumar – N01606417

**Submitted to:** Professor Ammar Al-Qaraghuli

**INTRODUCTION**

This project conducts a comprehensive analysis of Canadian well-being indicators, with a primary focus on the Health and Environment domains. Leveraging the power of the R programming language, we employed a range of data manipulation and visualization techniques to uncover meaningful insights from publicly available datasets, primarily sourced from Statistics Canada and Environment and Climate Change Canada.

Our analytical process began with data extraction and cleaning, ensuring consistency across time series and geographic levels. We then performed targeted data transformation, including grouping, filtering, and aggregation, to align indicators for comparative analysis. Using R packages such as tidy verse, ggplot2, and plotly, we visualized trends across multiple variables, enabling clearer interpretation of changes over time and between provinces.

In the Health domain, we analyzed indicators such as life expectancy, perceived mental health, and physical activity compliance, highlighting regional disparities and temporal patterns. For the Environment domain, we examined greenhouse gas (GHG) emissions, air quality relative to CAAQS standards, and their interplay over time.

The following sections detail the methodology used for data handling, the rationale behind our visualization approaches, and the key insights derived from these indicators. Together, these findings provide a nuanced perspective on how Canadians’ well-being is evolving in response to health challenges and environmental changes, with implications for policy, public awareness, and future research.

**HEALTH DOMAIN ANALYSIS**

**1. Life Expectancy and Mental Health Indicators**

**Objective:** To analyze life expectancy across Canadian provinces and compare it with national mental health statistics.

**Data Processing and Visualization:**

* Data was loaded and cleaned using tidyverse functions.
* Life expectancy data was filtered for relevant age groups and regions.
* Mental health data was structured to represent different categories of mental well-being.
* A combined dataset was created to visualize provincial life expectancy along with national mental health statistics.

**R Code Overview:**

**Prepared Combined Dataset & unified visualization:**

|  |
| --- |
| **A screenshot of a computer code  AI-generated content may be incorrect.** |

|  |
| --- |
| A screenshot of a computer program  AI-generated content may be incorrect. |

|  |
| --- |
| A screenshot of a computer code  AI-generated content may be incorrect. |

*Graph 1. Unified Health Metrics Visualization: Provincial Life Expectancy vs National Mental Health Data*

A graph with different colored bars

AI-generated content may be incorrect.

**Key Insights:**

* Life expectancy at birth shows significant variation across Canadian provinces. For instance, **Newfoundland and Labrador reports the lowest healthy life expectancy at 66.3 years**, while **Quebec leads with 71.6 years**. These differences reflect disparities in healthcare access, socioeconomic factors, and regional supports (Statistics Canada, 2024a).
* Mental health outcomes are similarly uneven: **47.3% of Canadians rate their mental health as "Excellent"**, while **20.7% report it as "Poor"**, highlighting an urgent need for accessible mental health services and prevention strategies (Statistics Canada, 2024a).
* When combined, these indicators offer a broader understanding of health outcomes, suggesting that **longevity must be assessed alongside mental wellness** to develop holistic health policies (Statistics Canada, 2024a).

**2. Physical Activity and Well-being Indicators**

**Objective:** To analyze the percentage of Canadians meeting 24-hour movement guidelines over different years.

**Data Processing and Visualization:**

* Data was filtered for relevant categories.
* Average percentages of population meeting the guidelines were calculated.
* A bar chart was created to visualize trends over time.

**R Code Overview:**

A screenshot of a computer program

AI-generated content may be incorrect.

*Graph 2. Physical Activity Compliance Trends: 2016-2021*

A graph of blue rectangular bars

AI-generated content may be incorrect.

**Key Insights:**

* The proportion of Canadians meeting recommended physical activity levels fluctuated between **2016 and 2021**, peaking in 2019 at **approximately 43%**, before declining during the COVID-19 pandemic (Statistics Canada, 2024a).
* This trend demonstrates how **public health emergencies can disrupt positive lifestyle behaviors**, especially among youth and NEET (Not in Education, Employment, or Training) populations (Statistics Canada, 2024c).
* Recognizing these fluctuations is essential for designing **policies and programs that encourage physical well-being**, particularly in the face of societal disruptions (Statistics Canada, 2024a; Statistics Canada, 2024c).

**ENVIRONMENT DOMAIN ANALYSIS**

**1. Air Quality and GHG Emissions Indicators**

**Objective:**

The objective of this analysis is to examine health-related statistics, specifically focusing on air quality and greenhouse gas (GHG) emissions in Canada. The goal is to assess how air pollution levels have changed over the years and their correlation with emissions.

**Data Processing and Visualization:**

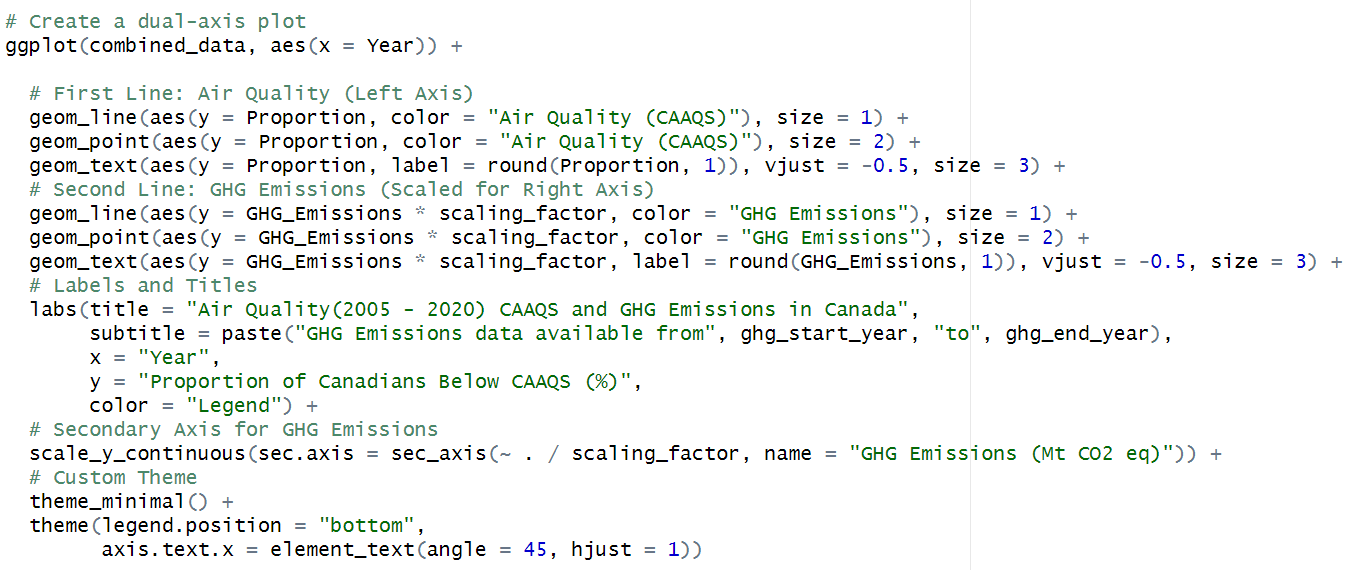
* Air quality and GHG emissions data were merged and scaled for comparative analysis.
* A dual-axis line plot was used to visualize trends over time.

**R Code Overview:**

A screenshot of a computer program

AI-generated content may be incorrect.

**Code for Dual-axis line plot**

****

*Graph 3. Air Quality and GHG Emissions trends.*

A graph with lines and numbers

AI-generated content may be incorrect.

* Canada’s **GHG emissions declined by 8.5% between 2005 and 2023**, dropping from 759 to **694 Mt CO₂ eq**, largely due to reductions in the electricity and heavy industry sectors (Government of Canada, 2024a).
* However, overall emissions are still **14.4% higher than in 1990**, driven by increases in emissions from oil and gas, transport, and agriculture (Government of Canada, 2024a).
* In terms of air quality, **only 74% of Canadians lived in areas meeting the Canadian Ambient Air Quality Standards (CAAQS) from 2020–2022**, down from 85% in the previous reporting period, primarily due to **ground-level ozone exceedances in Ontario and wildfire impacts in BC and Alberta** (Government of Canada, 2024b).
* These findings reflect an **environmental trade-off**: while GHG reductions signal progress in climate policy, worsening air quality in some regions points to the **complexity of balancing industrial, environmental, and public health goals** (Government of Canada, 2024a; Government of Canada, 2024b).

**2. Physical flow account for water usage (trends) indicator**

**Objective:**

This indicator focuses on water usage trends across industries and households in Canada. The main aim of this kind of indicator  is to analyze how water consumption has evolved over the years.

**Data Processing and Visualization**

* Data Loading: The dataset containing water usage statistics was imported into R.
* Data Filtering: The dataset was filtered to include only total water use across industries and households.
* Visualization: A line graph with point markers was generated to illustrate the changes in water usage over time.

**#R Code Overview**

**A screenshot of a computer program

AI-generated content may be incorrect.**

***Graph 4. Total Water Use Trend (2009 - 2021)***

**A graph with a line and numbers

AI-generated content may be incorrect.**

**Key Insights**

* Need for Sustainable Management: Given the increasing pressure on natural resources, implementing efficient water conservation strategies is essential.
* Policy Implications: Understanding water use trends can aid policymakers in designing regulations for better resource management.

**SOCIAL WELL-BEING ANALYSIS**

**1. Sense of belonging to the local community & having someone to count on indicators**

This analysis explores the Sense of Belonging to Local Community and Having Someone to Count On indicators in Canada. The data is analyzed by gender and province over the period from Q2 2022 to Q4 2024. The focus of this analysis is to assess the degree of social connectedness across Canadian populations.

**Objective:** The main objective of this analysis is to explore the trends of:

1. **Sense of Belonging to Local Community** – Understanding the connection people feel to their local communities by gender and province.
2. **Having Someone to Count On** – Exploring the social support networks available to individuals by gender and province.

**Data Processing and Visualization**

* **Data Loading & Cleaning:** The two datasets were loaded into R and filtered for the **Sense of Belonging** and **Having Someone to Count On** indicators. The **gender-specific data** (Men, Women, and Total) was selected for further analysis.
* **Data Transformation:** The Year was extracted from the REF\_DATE field. The VALUE field was converted to numeric for analysis. We also filtered out any missing values (NA).
* **Data Grouping:** The data was grouped by **Year**, **Province/Territory (GEO)**, and **Gender**, and then averaged for each group.
* **Combining Datasets:** Both datasets were processed separately and then combined into one unified dataset for easy comparison.

**#R Code Overview**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**A screenshot of a computer program

AI-generated content may be incorrect.**

**Graph: Dot Plot of Social Connectedness**

A dot plot was generated to display the yearly averages of both indicators by province, gender, and year (2022 vs. 2024). Each data point represents the average percentage of people reporting a sense of belonging or social support, differentiated by gender and year. The plot allows for a clear visual comparison between the two indicators.

A computer screen shot of a code

AI-generated content may be incorrect.

**Continued ….**

**A screenshot of a computer program

AI-generated content may be incorrect.**

*Graph 5: Dot Plot of Social Connectedness*

**A graph with numbers and dots

AI-generated content may be incorrect.**

**Key Insights**

* **Sense of Belonging**: Significant regional differences exist, with varying perceptions of community connection between men and women across provinces.
* **Social Support**: The **Having Someone to Count On** indicator reveals important variations in social support networks, highlighting the strength of ties across gender and province.
* **Gender & Regional Trends**: The data highlights distinct gender and provincial differences, shedding light on diverse experiences of community and support.

**2. Sense of Belonging to Canada Indicator**

**Objective**: This analysis explores key indicators of social well-being in Canada, focusing on **Sense of Belonging to Canada**. The dataset covers gender-specific data for **Men** and **Women**, segmented by yearly trends.

The main objectives of this analysis are to examine:

* **Sense of Belonging to Canada**: Understanding how individuals across Canada feel about their connection to the country, segmented by gender.

**Data Processing and Visualization**

* **Data Loading & Preprocessing:** The dataset was loaded, filtering the data to focus on the Sense of Belonging and Having Someone to Count On indicators, with gender-specific data for Men and Women.
* **Data Transformatio**n: The Year was extracted from the REF\_DATE column, and the VALUE field was converted to numeric for analysis.
* **Grouping:** The data was grouped by Year and Gender, calculating the average percentage for each group.
* **Visualization:** Bar plots were generated to show trends in sense of belonging and support networks by year and gender.

**R Code Overview:**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**A white background with black text

AI-generated content may be incorrect.**

***Graph 6:*** *The bar plot below illustrates the average sense of belonging to Canada, segmented by gender and year. It clearly shows how perceptions shift over time, providing insights into social connectedness across regions.*

A graph with green and orange bars

AI-generated content may be incorrect.

**Key Insights**

* **Sense of Belonging to Canada**: The data shows how perceptions of belonging evolve over time across genders, revealing regional variations.
* **Gender and Regional Differences**: The analysis highlights the differences in how **Men** and **Women** feel about their connection to Canada, with variations across provinces.
* **Policy Implications**: The findings can inform policymakers aiming to strengthen social cohesion, suggesting areas where gender-specific or region-targeted programs may be needed.

**GOOD GOVERNANCE DOMAIN ANALYSIS**

**1. Discrimination and unfair treatment by gender and province**

**Objective:**

To analyze trends in self-reported discrimination or unfair treatment across Canadian provinces, comparing gender disparities and quarterly changes throughout 2024.

**Data Processing and Visualization:**

* Data was loaded and cleaned using dplyr function.
* Converted VALUE to numeric percentages and structured categorical variables:
  + Quarter (Q1–Q4) derived from REF\_DATE.
  + Gender standardized to "Men" and "Women".
* Visualized trends using ggplot2, with separate line charts for each province (GEO) and gender. Key features:
  + Faceted plot by province (facet\_wrap) to compare regional trends.
  + Quarterly trends highlighted with gender-specific colors (blue for men, orange for women).
  + Dynamic labels, titles, and themes to enhance readability.

**R Code Overview:**

A screenshot of a computer code

AI-generated content may be incorrect.

*Graph 7: Quarterly Discrimination Trends in Canada*

A graph with blue and orange lines and dots

AI-generated content may be incorrect.

**Key Insights:**

* **Provincial Variations in Discrimination Rates**
  + Provinces like Alberta, Ontario, and Quebec show higher percentages of reported discrimination compared to smaller provinces (e.g., Prince Edward Island or New Brunswick).
  + British Columbia and Manitoba exhibit moderate rates, with fluctuations across quarters.
* **Gender Disparities**
  + Women consistently report higher discrimination rates than men across most provinces (e.g., Ontario, Quebec, Alberta), highlighted by the orange line (women) trending above blue (men).
  + The gender gap widens in province like Nova Scotia during Q3 and Q4.
* **Regional Patterns**
  + Western provinces (Alberta, British Columbia) and central provinces (Ontario) exhibit more volatility in trends, while Atlantic provinces (Nova Scotia, New Brunswick) show steadier patterns.
* **Data Limitations**
  + Smaller provinces (e.g., PEI) have narrower confidence intervals due to lower sample sizes, potentially affecting trend reliability.

2. Experience(s) of discrimination, reason(s) and context(s) of discrimination, 5 years before and since the beginning of COVID-19 pandemic, by groups designated as visible minorities and selected sociodemographic characteristics

**Objective:**

To analyze trends in self-reported experiences, reasons, and contexts of discrimination among groups designated as visible minorities versus non-visible minorities in Canada, comparing pre-COVID-19 (5 years prior) and since the onset of the COVID-19 pandemic, while incorporating selected sociodemographic characteristics.

**Data Processing and Visualization:**

1. **Data Loading and Cleaning:**
   * Data was imported from a CSV file and processed using dplyr.
   * Relevant columns (Visible.minority, Indicators, VALUE) were retained, and rows were filtered to include only entries related to *discrimination*.
2. **Variable Transformation:**
   * **Time Periods:** Created a Time\_Period variable categorizing data into *Pre-COVID* or *During COVID* based on keywords in the Indicators column.
   * **Discrimination Types:** Derived Indicator\_Type categories (e.g., *Ethnicity/Culture*, *Race/Color*, *Workplace*) using regex pattern matching on the Indicators column.
   * **Numeric Conversion:** Ensured VALUE was treated as numeric for percentage calculations.
3. **Visualization (ggplot2):**
   * **Chart Type:** Bar charts (geom\_col) comparing discrimination rates between *visible* and *non-visible minorities*.
   * **Faceting:** Split plots by Time\_Period (Pre-COVID vs. During COVID) using facet\_wrap.
   * **Color Coding:** Assigned distinct colors (blue for visible minorities, orange for non-visible minorities) via scale\_fill\_manual.
   * **Labels:** Added percentage labels on bars using geom\_text and improved readability with dynamic titles and axis labels.
   * **Theming:** Applied theme\_minimal and rotated x-axis labels to avoid overlap.

**R Code Overview:**

A screenshot of a computer program

AI-generated content may be incorrect.

*Graph 8: Discrimination Rates*

A graph of different colored bars

AI-generated content may be incorrect.

**Key Insights:**

* **Visible Minority vs. Non-Visible Minority Disparities:** Visible minorities consistently report higher discrimination rates than non-visible minorities. Pre-COVID rates were 50.4% (visible) vs. 40.0% (non-visible). During COVID, visible minorities remained disproportionately affected (e.g., 45.6% vs. 35.7%).
* **Subgroup Variations:** Subgroups within visible minorities show stark disparities, with some reporting very low rates (e.g., 1.7%, 1.9%), suggesting internal diversity in discrimination experiences.
* **Anomalies:** Certain percentages (e.g., 5%, 4.2%) deviate significantly from broader trends, potentially indicating unique subpopulations or data collection inconsistencies.

**PROSPERITY DOMAIN ANALYSIS**

**1. Age-specific labor force metrics, adjusted for seasonality (monthly)**

**Objective:** To compare and visualize labour force characteristics (employment, unemployment, and participation rates) across different age groups.

**Data Processing and Visualization:**

Data & Cleaning Loaded labour data from CSV. Split into:

* Counts (employment/unemployment in thousands)
* Rates (unemployment/participation/employment rates in %) using dplyr::filter.

**Visualization**

1. **Counts Plot:**
   * Bar charts by age group (geom\_bar), faceted by labour metric (4 categories, 2 columns).
   * Units: Thousands; stripped legend; rotated x-labels.
2. **Rates Plot:**
   * Similar structure, faceted into 3 columns for rate metrics.
   * Units: Percentages; minimalist theme.

**Key Features**

* Faceting for direct category/age comparisons.
* Clean design (theme\_minimal), dynamic labels.
* Avoided unit confusion by separating counts vs. rates.

**R Code Overview:**

A screenshot of a computer code

AI-generated content may be incorrect.

A computer screen shot of a code

AI-generated content may be incorrect.

*Graph 9: Labour Force Rates*

A graph of different colored squares

AI-generated content may be incorrect.

**Key Insights:**

* **Labor Market Metrics:**
  + The three key rates (employment, participation, and unemployment) across different age groups in March 2025.
* **Age Group Comparisons:**
  + A comparison of labor trends for groups like:
    - Younger workers (15–24 years)
    - Prime-age workers (25–54 years)
    - Older workers (55+ years).
* **Youth (15–24 years):**
  + Likely lower participation (due to education) and higher unemployment (entry-level job challenges).
* **Prime Working Age (25–54 years):**
  + Highest employment and participation rates, as this group is typically the core workforce.
* **Older Workers (55+ years):**
  + Lower participation (retirement trends) but potentially stable employment for those still active.

2.  Proportion of 15- to 29-year-olds enrolled in or outside formal education, categorized by employment status, highest educational attainment, age subgroups, and sex.

**Objective:**

To compare key youth (ages 15–29) education and employment indicators for Canada against the OECD average over time.

**Data Processing and Visualization:**

**Data Import & Cleaning:**

* Loaded CSV and filter for Canada and OECD Average (renaming OECD string).
* Select indicators: *Total in education*, *Employed in education*, *Employed not in education*, and *NEET* for ages 15–29.
* Convert VALUE and REF\_DATE to numeric.

**Visualization:**

* Create a line plot with points using ggplot2 (x = year, y = value, colored by region).
* Facet by labour/education status (free y-scales, 2 columns) to compare trends side by side.
* Apply custom colors, minimal theme, rotated x-axis labels, and set x-axis breaks (2020–2024).

**Objective:**

* Compare key youth education and employment indicators between Canada and the OECD average (2020–2024).

**R Code Overview:**

A screenshot of a computer program

AI-generated content may be incorrect.

*Graph 10: Education and Employment*

A graph with red and blue lines and numbers

AI-generated content may be incorrect.

**Key Insights:**

* **Higher Dual Engagement:** A larger proportion of Canadian youth (15–29) are both employed and in education compared to the OECD average, indicating a trend of balancing work and studies.
* **Lower NEET Rate:** Fewer Canadian youth are categorized as NEET (not in employment, education, or training) relative to the OECD, suggesting better overall engagement in either work or education.
* **Increased Educational Enrollment:** Canada shows consistently higher rates of youth enrollment in education, meaning more young people are pursuing studies.
* **Comparable Employment Outside Education:** The rate of youth employed but not in education is generally in line with, or slightly below, the OECD average, highlighting a focus on continuing education alongside part-time employment rather than full-time work after leaving school.

**INFOGRAPHIC**

A screenshot of a graph

AI-generated content may be incorrect.

**REFERENCES**

Environment and Climate Change Canada. (2024, May 2). *2024 National Inventory Report 1990–2022: Greenhouse Gas Sources and Sinks in Canada*. Government of Canada. <https://publications.gc.ca/collections/collection_2024/eccc/En81-4-2022-1-eng.pdf>

Environment and Climate Change Canada. (2025, March 7). *Population exposure to outdoor air pollutants*. Government of Canada. <https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/population-exposure-outdoor-air-pollutants.html>

Statistics Canada. (n.d.). *Table 13-10-0893-01: Youth not in employment, education, or training for persons with disabilities aged 15 to 29 years by gender*. Retrieved April 6, 2025, from <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310089301>

Statistics Canada. (2024, July 11). *Canadian System of Environmental–Economic Accounts: Water use, 2021* (The Daily). <https://www150.statcan.gc.ca/n1/daily-quotidien/240711/dq240711a-eng.pdf>

Statistics Canada. (2025, March 5). *Key findings from the Health of Canadians report, 2024* (The Daily). <https://www150.statcan.gc.ca/n1/daily-quotidien/250305/dq250305a-eng.htm>