Development of an open-source pipeline for the replication and extension of the Canadian Active Living Environments (Can-ALE) measure.

Daniel Fuller, University of Saskatchewan

Meisam Ghasedi, University of Saskatchewan

Gorham Achot, Queen's University

Nancy A. Ross, Queen's University

User Manual & Technical Document

Table of Contents

1. Introduction and Study Framework

1.1. Introduction

This study builds upon the existing Can-ALE measure, which was previously available for 2006 and 2016. We have updated the Active Living Environment (ALE) index to include 2011 and 2021 data, in addition to the 2016 data. We calculated the index using five key factors, including Weighted population density, Weighted dwelling density, Number of Transit Stops, Number of Intersections with equal or more than Legs density and Weighted Points of interest. Although the Can-ALE is now available for four census years, it is not considered reliable for longitudinal analysis. This limitation stems from inconsistencies in the underlying data, such as the progressive completeness of OpenStreetMap (OSM) POIs over time and the availability of transit stop data for only 2016 and 2021. However, the measures can be used effectively to develop and analyze the ALE index for each census year independently. All code, original data, and Can-ALE data are provided on our website (https://walkabillylab.github.io/Can-ALE/).

1.2.Software

All code was written in RStudio (Version 2025.05.1, Build 513), and has been made available as open-source for use in future census years or for other analyses. Additionally, ArcGIS Pro 3.0.1 was used as validation software for all measures to confirm that the results from the R scripts were consistent with those produced by ArcGIS.

1.3.Data Collection and Definition

The updated Active Living Environment (ALE) index for 2011, 2016, and 2021 is based on five core measures: weighted population density, weighted dwelling density, transit stop counts, intersection density (≥3 legs), and a weighted points of interest (POI). For detailed definitions and data sources for each of these measures, please refer to Table 1.

Table 1. Definition of measures and their sources

Measure	Definition	Data Source
Weighted population density	The number of people per square kilometer within a 1-km circle centered on a Dissemination Area's (DA) population-weighted centroid.	Census (Statistics Canada)
Weighted dwelling density	The number of dwelling per square kilometer within a 1-km circle centered on a Dissemination Area's (DA) population-weighted centroid.	Census (Statistics Canada)
Transit Stops	The number of available transit stops within 1 kilometer of population weighted centroid of the DA	General Transit Feed Specification (GTFS)
Intersections with ≥3 Legs density	The number of ≥ three-way intersections on roads within 1 kilometer of population weighted centroid of the DA, excluding roads classified as motorways (highways, freeways) or slip roads (e.g., highway entrance and exit ramps)	Road Network File (Statistics Canada)
Weighted Points of interest	The number of points of interest (e.g., libraries, schools, hospitals) within 1 kilometer of population weighted centroid of the DA and weighted according to their importance and distance from weighted centroid of the DA.	OpenStreetMap

1.4.Geographic Unit of Analysis

To ensure the index accurately reflects the environments that people actually live in, all measures were computed for the area within a 1-kilometer circular (Euclidean) buffer centered on the population-weighted centroid of each Dissemination Area, rather than a simple geometric centroid. This weighted centroid represents the population's "center of mass" and was computed by taking the weighted average of the centroids of its constituent Dissemination Blocks (DBs), using the population of each block as its weight. In cases where a Dissemination Area had zero population, the geometric centroid was used instead to ensure complete coverage and prevent any area from being omitted. Finally, the Can-ALE measures were estimated within each buffer using a precise areal interpolation method at the Dissemination Area (DA) level.

2. Methodology and Code Implementation

This section details the methodology for calculating the measures, provides the corresponding code, and presents the results. Each code link includes a readme file with a step-by-step guide to facilitate future implementation and analysis.

2.1. Population and Dwelling Densities

This section outlines the method used for calculating population and private dwelling densities. The script calculates these densities within a 1-kilometer radius of a Dissemination Area's (DA) population-weighted centroid.

The core of this improved method is the use of census data at the Dissemination Block (DB) level, the smallest geographic unit for which census data is available. This approach provides a more granular and accurate spatial distribution of population and dwellings compared to methodologies that rely on DA-level data.

The key assumption of this method is the uniform distribution of population and dwellings within each DB. Because DBs represent a very small geographic area (often just a few city blocks), this assumption is far more reliable and leads to a more accurate density estimation than assuming uniform distribution across a much larger DA.

Data Sources:

The analysis relies on four key datasets:

Census Data (DB-Level): Population and private dwelling counts at the DB level are retrieved from the Canadian Census using the cancensus package. A valid API key, which can be obtained by registering on the CensusMapper website, is required for this process.

DA Shapefile: Contains the geographic boundaries for all Dissemination Areas.

- **2011**: lda_000b11a_e.shp
- **2016**: lda_000b16a_e.shp
- **2021**: lda_000b21a_e.shp

DB Shapefile: Contains the geographic boundaries for all Dissemination Blocks.

- **2011**: ldb_000b11a_e.shp
- **2016**: ldb_000b16a_e.shp
- **2021**: ldb_000b21a_e.shp

Population-Weighted Centroids: A pre-calculated file with a single representative point for each DA, based on where the population is concentrated. The file corresponding to the analysis year should be used.

- **2011:** DA_2011_hybrid_centroids_Canada.gpkg
- **2016:** DA_2016_hybrid_centroids_Canada.gpkg
- **2021:** DA_2021_hybrid_centroids_Canada.gpkg

Methodology

The calculation is performed for each province through a series of spatial operations as described below.

Step 1: Buffer Creation

First, the analysis area for each DA is defined. The script takes the pre-calculated population-weighted DA centroids and generates a 1-kilometer radius buffer around each point. This buffer specifies the specific area for which the density will be calculated.

Step 2: Interpolation and Proportional Allocation

This is the central step of the analysis, where population and dwelling counts are allocated to each buffer. For each 1-km DA buffer, the script identifies all of the smaller DB polygons that it intersects. A DB may only partially fall within a buffer. The script calculates the exact proportion of each intersecting DB's area that lies inside the buffer. It then attributes the population and dwelling counts from that DB based on this area proportion. For example, if 30% of a DB's area is inside the buffer, 30% of its population and 30% of its dwellings are assigned to that buffer.

Step 3: Density Calculation

For each DA buffer, the proportionally-allocated population and dwelling counts from all the intersecting DBs are summed up. This gives a highly accurate estimate of the total population and total dwellings within the 1-km radius. Finally, these total counts are divided by the buffer's area (which is $\pi \times (1\text{ km})2 \approx 3.14\text{ km}2$) to produce the final population density and dwelling density measures.

2.2. Intersection Density

This section outlines the method used for calculating intersection density of dissemination areas (DAs) across Canadian provinces. In this study, intersection density, a key measure of walkability, is defined as the number of 3-way or more road intersections within a 1-kilometer buffer around the population-weighted centroid of each DA.

Data Sources

The analysis relies on three primary datasets for a given census year.

DA Shapefile: Contains the geographic boundaries for all Dissemination Areas.

- **2011**: lda_000b11a_e.shp
- **2016**: lda_000b16a_e.shp
- **2021**: lda_000b21a_e.shp

Population-Weighted Centroids: A pre-calculated file with a single representative point for each DA, based on where the population is concentrated. The file corresponding to the analysis year should be used.

- **2011**: DA_2011_hybrid_centroids_Canada.gpkg
- **2016**: DA_2016_hybrid_centroids_Canada.gpkg
- **2021**: DA_2021_hybrid_centroids_Canada.gpkg

Road Network File (RNF): Contains the road segments for all of Canada.

- **2011**: lrnf000r11a_e.shp
- **2016**: lrnf000r16a_e.shp
- **2021**: lrnf000r21a_e.shp

Methodology

The script processes data on one province at a time, following these steps:

Step 1: Preparing Buffers and Roads

1. **Buffer Creation**: For each province, the script loads the pre-calculated population-weighted centroids for its DAs and creates a 1-kilometer circular buffer around each one. This buffer defines the area of analysis for each DA.

2. **Filtering the Road Network**: The provincial road network is filtered to exclude limited-access highways and ramps (class codes 10, 11, 12, and 13), focusing the analysis on roads relevant to local connectivity and walkability.

Step 2: Tiling and Intersection Analysis

To handle large provincial road networks efficiently, the script divides the entire province into a grid of **10 km x 10 km tiles**. The intersection analysis is then performed independently within each tile.

For each tile, the script performs a sophisticated process to identify true intersections:

1. **Node the Network**: It first finds all points where road lines cross or touch.
2. **Split the Roads**: The road network is then split at these intersection points, breaking down long roads into smaller segments that connect nodes. This process is known as "noding."
3. **Count Connections**: The script then counts how many of these new, smaller road segments connect at each node (endpoint).
4. **Identify 3+ Way Intersections:** Only nodes where three or more road segments meet are classified as valid intersections.

Step 3: Calculating Final Density

After all tiles in a province are processed, the unique intersection points are combined.

1. **Count Intersections in Buffers**: The script counts how many of these valid 3+ way intersections fall within each DA's 1-km buffer.
2. **Calculate Density**: The final intersection density is calculated by dividing the total intersection count by the area of the buffer (πr^2, where $r=1$km, so the area is π km^2).

The results are aggregated for each DA and saved as a provincial Excel file.

2.3. Point of Interest (POI)

This section outlines the method used for calculating weighted Point of Interest (POI) index for each dissemination area (DA). The script generates a comprehensive index reflecting the availability and accessibility of destinations relevant to active living environments. The final output is an Excel file listing each DAUID alongside both a raw POI count and the calculated weighted POI index.

Data Sources

The analysis uses three main types of data:

- **DA Shapefile**: A file containing the geographic boundaries for all dissemination areas in Canada.
    - **2011:** lda_000b11a_e.shp
    - **2016:** lda_000b16a_e.shp
    - **2021:** lda_000b21a_e.shp
- **POI Shapefiles**: Two separate OpenStreetMap (OSM) shapefiles containing POIs:
    - One file contains POIs as points (e.g., gis_osm_pois_free_1.shp).
    - The other contains POIs as polygons (e.g., gis_osm_pois_a_free_1.shp).
- **Population-Weighted Centroids**:
    - 2011: DA_2011_hybrid_centroids_Canada.gpkg
    - 2016: DA_2016_hybrid_centroids_Canada.gpkg
    - 2021: DA_2021_hybrid_centroids_Canada.gpkg

Methodology

Step 1: Generating Buffers and Preparing POIs

- **Creating Buffers**: Using the filtered population-weighted centroids for the province, the script generates a circular buffer with a 1-kilometer radius around each centroid. These buffers define the catchment area for counting and weighting POIs for each DA.
- **POI cleaning:** Some POI records are filtered out because their "code" values are not considered relevant to active living. POIs with the following codes are excluded:
"code" NOT IN (2423, 2725, 2424, 2951, 2961, 2734, 2422)

Step 2: Processing POI Data

- **Unifying POI Files**: For each province, the two POI files (points and polygons) are read and transformed into the same projection (EPSG:3347) to align with the DA data.
- **Geometry Conversion and Cleaning**: For the POI file containing polygons, the script converts each polygon feature into its centroid. This step ensures all POI features are uniformly represented as points.
- **Merging and Filtering**: The two processed POI datasets (original points and polygon-centroids) are merged into a single dataset. POIs with the irrelevant OSM codes listed in Section 1 are then removed.

Step 3: POI Weighting

Instead of a simple count, each POI is assigned a final weight based on two factors: its category and its distance from the DA's population weighted centroid.

Category-Based Weights

POIs are assigned a base weight to reflect their relative importance to creating an active living environment. POIs are categorized into four tiers based on their OSM code, with a higher weight indicating greater importance.

- **Weight of 1**: All other relevant POIs (default).
- **Weight of 2**: Other useful destinations (e.g., Cinema, Doctors, Hotel).
- **Weight of 3**: Important secondary destinations (e.g., University, Library, Cafe).
- **Weight of 4**: Key destinations (e.g., Supermarkets, Park, Playground).

Distance-Decay Weights

A second weight is applied to account for proximity. POIs closer to a DA's population-weighted centroid receive a higher weight than those farther away within the 1 km buffer. This is calculated using an exponential distance-decay function:

$$W\_decay = 1.0126 \times e^{(-0.0013 \times d)}$$

Where d is the distance in meters from the POI to the DA's population-weighted centroid.

Step 4: Spatial Join and Index Calculation

- **Spatial Join**: All processed POIs are spatially joined to the 1-km buffers. This step identifies which DAs each POI is associated with.
- **Weight Calculation:** For each POI within a DA's buffer, a final weight is calculated by multiplying its two component weights:

Final Weight = Category Weight × Distance-Decay Weight

- **Final Aggregation and Summarization**: The script groups all POIs by their associated DAUID. It then calculates two summary values for each DA:
    1. **raw_poi_count**: The simple count of all POIs within the buffer.
    2. **weighted_poi_index**: The sum of the Final Weight of all POIs within the buffer. This is the primary output metric.
- **Final Output**: The summary results are joined back to the complete list of DAs for the province. This ensures that DAs with no POIs in their buffer are included in the final report with their values set to 0. The final table is then exported as an Excel file for the province.

2.4. Transit Stop Density

This section outlines the method used for calculating transit stop density, from initial data acquisition to the final analysis. The process involves three main stages: downloading General Transit Feed Specification (GTFS) data, organizing the raw files, and then calculating the number of unique transit stops within a 1-kilometer buffer of each Dissemination Area's (DA) population-weighted centroid.

Data Sources

We require three datasets:

- **Population-Weighted Centroids**:
    - 2016: DA_2016_hybrid_centroids_Canada.gpkg
    - 2021: DA_2021_hybrid_centroids_Canada.gpkg
- **DA Shapefile**: A file containing the geographic boundaries for all dissemination areas in Canada.
    - **2016:** lda_000b16a_e.shp
    - **2021:** lda_000b21a_e.shp
- **GTFS dataset:**
    - GTFS 2016, downloaded from TransitFeeds website for each province separately.
    - GTFS 2021, downloaded from TransitFeeds website for each province separately.

Step 1: GTFS Data Acquisition

The first step is to automate the process of downloading GTFS data from the main sources (TransitFeeds website). Scripts are customized for each province, using URLs specific to the various transit agencies.

To ensure the data reflects typical ridership, a representative day is carefully selected based on the following criteria:

- **Exclusion of Atypical Days**: Statutory holidays and weekends are excluded to avoid unusual transit usage patterns. Certain non-statutory holidays are also omitted, as service levels can vary.
- **Priority for High-Usage Days**: The selection process prioritizes typical weekdays to capture standard commuting patterns.

The script iterates through a date range, checks each day against these criteria, and downloads the GTFS data for the first valid day found. The output of this step is a collection of raw GTFS files containing stops.txt files, which provide the geographic locations of transit stops.

Step 2: Data Processing — Organizing GTFS Files

Once downloaded, the raw GTFS files are processed and organized to prepare them for analysis.

- **Unzipping and Organization**: The downloaded .zip files are unzipped and stored in province-specific folders. Within each provincial folder, subdirectories are created for each transit agency, ensuring the data is clearly structured and easy to access.
- **Categorization by Area**: The script also identifies transit agencies that operate within a Census Metropolitan Area (CMA). An agency is classified as a "CMA agency" if it serves any part of a CMA. This allows for future analyses that can differentiate between urban and non-urban transit systems.

The primary output of this stage is a set of cleanly organized stops.txt files, ready for the spatial analysis phase.

Methodology

1-kilometer circular buffer is created around each population-weighted centroid. This buffer serves as the catchment area for counting transit stops. The script then aggregates all stops.txt files for the province into a single spatial layer. A spatial join is then performed to identify every transit stop that falls within each DA's 1-km buffer. Finally, it groups the results by each DA and counts the number of distinct transit stop IDs, which prevents any stop from being counted more than once.

The results are compiled into a final table. DAs that have no transit stops within their buffer are assigned a count of 0. This ensures a complete dataset for the province. The final table, containing each DAUID and its corresponding transit_stop_count, is exported as an Excel file for further use.

3. Appendix

Table A1: Field Descriptions

Field name	Full name	Definition	Source
DAUID	Dissemination area - unique ID	Unique identifier of the dissemination area (eight digits)	Statistics Canada (Boundary Files)
INT_D	Intersection density	The number of ≥3-way intersections per square kilometre in the buffer around a dissemination area centroid	OpenStreetMap road network
DWL_D	Dwelling density	The number of dwellings per square kilometre in the buffer around a dissemination area centroid	Statistics Canada
POI	Points of Interest	The number of points of interest in the buffer around a dissemination area centroid	OpenStreetMap
Z_INT_D	Intersection density (z-score)	Z-score of the intersection density measure	OpenStreetMap road network
Z_DWL_D	Dwelling density (z-score)	Z-score of the dwelling density measure	Statistics Canada
Z_POI	Points of Interest (z-score)	Z-score of the points of interest measure	OpenStreetMap
ALE_INDEX	Active Living Environment Index	Sum of the z-score of the intersection density, dwelling density, and points of interest measure	-
ALE_CLASS	Active Living Environment class	Categorical value characterizing the favourability of the ALE on a scale from 1 (very low) to 5 (very high)	-
TRANSIT	Transit stops	The number of public transit stops or stations in the buffer around a dissemination area population weighted centroid	TransitFeeds
Z_TRANSIT	Transit stops (z-score)	Z-score of the transit measure	TransitFeeds
ALE_TRANSIT	Transit stops (z-score)	Sum of the z-score of the intersection density, dwelling density, points of interest, and transit measures	-
ALE_TRANSIT_CLASS	Active Living Environment + Transit class	Categorical value characterizing the favourability of the ALE in CMAs on a scale from 1 (very low) to 5 (very high)	-

Table A2. POI Weighting System

Weight = 2	**Weight = 3**	**Weight = 4**
Post Box	Library	School
Post Office	Community Centre	Park
Town Hall	University	Playground
Arts Centre	Kindergarten	Sports Centre
Public Building	College	Supermarket
Pharmacy	Dog Park	Bakery
Hospital	Pitch	Convenience
Doctors	Swimming Pool	Greengrocer
Dentist	Stadium	General Stores
Theatre	Ice Rink	Market Place
Cinema	Restaurant	-
Hotel	Fast Food	-
Motel	Cafe	-
Bookshop	Pub	-
Butcher	Bar	-
Optician	Food Court	-
Sports Shop	Biergarten	-
Bicycle Shop	Mall	-
Vending Machine	Department Store	-
Vending Parking	Newsagent	-
Bank	Bicycle Rental	-
Atm	Picnic Site	-
Attraction	Toilet	-
Museum	Bench	-
Theme Park	Bed and Breakfast	-
Drinking Water	-	-
Waste Basket	-	-
Clinic	-	-
Post Box	-	-

Table A3. Points of Interest (POIs) with a Weight of 1

Alpine Hut	Courthouse	Memorial	Tower
Archaeological	Do it yourself	Mobile Phone Shop	Toy Shop
Artwork	Embassy	Monument	Track
Battlefield	Florist	Nightclub	Travel Agent
Beauty Shop	Fountain	Nursing Home	Vending Any
Beverages	Fire Station	Observation Tower	Veterinary
Camera Surveillance	Fort	Outdoor Shop	Video Shop
Camp Site	Furniture Shop	Police	Viewpoint
Car Dealership	Garden Centre	Prison	Wastewater Plant
Car Rental	Golf Course	Recycling Glass	Water Works
Car Repair	Gift Shop	Recycling	Water Mill
Car Sharing	Graveyard	Recycling Metal	Wayside Cross
Car Wash	Guesthouse	Recycling Clothes	Water Tower
Caravan Site	Hairdresser	Recycling Paper	Wayside Shrine
Castle	Hostel	Ruins	Water Well
Chalet	Hunting Stand	Shelter	Windmill
Chemist	Jeweler	Shoe Shop	Zoo
Clothes	Kiosk	Stationery	-
Comms Tower	Laundry	Telephone	-
Computer Shop	Lighthouse	Tourist Info	-

Table A4. Descriptive and summary statistics of the five cluster groups

Year	Measure	Statistic	Overall	1 (very low)	2 (low)	3 (moderate)	4 (high)	5 (very high)
2011	ALE	Count	56204	15051	12528	15481	9936	3208
		Average	0	-2.46	-1.05	0.29	2.21	7.4
		Minimum	-2.61	-2.61	-2.002	-0.82	0.97	3.14
		Maximum	26.38	-1.23	0.97	1.26	5.49	26.38
2016	ALE	Count	56589	15015	11616	16041	10177	3740
		Average	0	-2.46	-1.1	0.32	1.86	6.87
		Minimum	-2.61	-2.61	-1.99	-0.73	0.24	2.54
		Maximum	28.66	-1.48	0.28	2.99	6.12	28.66
	ALE Transit	Count	30465	729	5059	12196	8752	3729
		Average	1.35	-3.66	-1.84	0.17	2.30	8.33
		Minimum	-4.33	-4.33	-3.64	-2.27	-1.34	2.19
		Maximum	32.8	-1.35	1.72	4.74	7.65	32.8
2021	ALE	Count	57932	15208	11764	16152	11334	3474
		Average	0	-2.48	-1.1	0.27	1.86	7.29
		Minimum	-2.64	-2.64	-2.02	-0.77	0.3	2.49
		Maximum	25.31	-1.47	0.25	2.68	5.9	25.31
	ALE Transit	Count	33697	791	5964	13124	10344	3474
		Average	1.25	-3.56	-1.81	0.08	2.35	8.76
		Minimum	-4.22	-4.22	-3.55	-2.19	-1.27	1.88
		Maximum	29.09	-1.86	1.52	4.88	9.73	29.09

Note. Averages for overall ALE scores are effectively zero because the ALE scores are z-scored with a mean of zero.

Figure A1: Reference Maps of Four Canadian Urban Areas by their 2021 ALE Class Value