canaccessR: An open data product for analyzing transportation accessibility to employment and grocery stores in Canada's largest metropolitan areas.

Journal Title
XX(X):2–8

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

João Pedro Figueira Amorim Parga, Anastasia Soukhov, Robert Nutifafa Arku, Christopher Higgins, Antonio Páez

Prepared using sagej.cls

Abstract

In this paper, we describe the {canaccesR} package, an open data product (ODP) created in R that contains public transit travel time estimates to employment locations and grocery stores across Canada's 12 largest metropolitan areas. We calculate travel time matrices (TTM) from and to each Dissemination Area (DA) within these regions for the years 2019 and 2023. We add value to the urban analytics community by processing and integrating raw data, and disseminating user-ready data in the domain of transportation accessibility in Canada. To do so, we use the {r5r} R package, General Transit Feed Specification (GTFS), OpenStreetMap (OSM), DMTI's Enhanced Points of Interest, and Statistics Canada Census data. This data package can be used by researchers, practitioners, and transit agencies to estimate accessibility levels to these two essential destinations within these urban areas. Moreover, these estimations can be used as inputs in equity and inequalities assessments, through the comparisons of within and across accessibility levels found throughout Canada's largest metropolitan areas. Consequently, we expect to contribute to informed and data-based decision making in transportation by disseminating these data. We hope that these datasets can substantiate future improvements in policy-making that may lead to greater justice in the country's urban transportation systems. The package is still in its initial phase and may undergo expansions in the future by adding TTM's for other destinations (e.g., schools, healthcare facilities). Finally, as an ODPs, the {canaccess} package allows for open exploration, use, and contribution by users through its GitHub repository.

Keywords

Public transit accessibility; open data products (ODPs); R data package; travel time matrices.

Introduction

The objective of this paper is to describe the {canaccesR} open data package. Its main contents are a set of public transit travel time matrices (TTM) estimates to employment and groceries stores from the 12 largest Canadian metropolitan areas in 2019 and 2023. The results are stored at the Dissemination Areas (DA) * level, yielding origin-destination pairs containing information on travel time by public transit, population, and total employment within each metropolitan area. These estimates were created by leveraging expertise in data science, computer programming, and transportation accessibility using the {r5r} package (Pereira et al., 2021b). We used public transit schedule, transport network, population, and business location data from different sources to estimate these TTM's datasets.

*Dissemination Areas are the smallest publicly available spatial unit provided by Statistics Canada (Government of Canada, 2021).

Estimating accessibility - the potential offered by the transportation system to reach destinations (Páez et al., 2012) - requires specialized datasets and technical expertise. Recent efforts following an open-source and transparent philosophy have been made to disseminate useful data and information on transportation in the Canadian context (Soukhov and Páez, 2023). However, despite these initiatives, pre-processed and available data that allow for the ease estimation of accessibility indicators are still scarce. Within this context, we expect to help filling this gap by the processing raw into user-ready data and making them publicly available to advance knowledge on the field. Our main contribution is to provide analysis-ready data for Canada's largest cities on the topic of transportation accessibility, thus making urban analytics in the country more accessible and contributing to future research and data-based decision making.

The package's main audiences are Canadian researchers in urban planning and transportation and transportation system agencies. We anticipate three primary uses for the open data product (ODP) described in this paper. First, the datasets allow for static assessment of the level of public transit accessibility across the country's largest cities before and after the COVID-19 pandemic. In other words, {canaccesR} makes it easier for those interested in comparing cities regarding their level of public transit accessibility to essential destinations (such as employment centers and groceries stores) to do so. Second, the temporal and spatial characters of the datasets made available here allow researchers to evaluate accessibility changes through time and across space within the largest Canadian urban areas. Third, as is now common practice in transportation accessibility research, used as inputs, these estimates can substantiate broader investigations on transportation justice and equity (Higgins et al., 2021; Humberto, 2023; Pereira et al., 2021a). For example, the TTM estimates allow for evaluating the evolution of public transit's accessibility by income or spatial distribution across all Dissemination Areas (DA's) of each of the 12 cities in the sample (Parga et al., 2024). In other words, the package's contents can be used from straightforward assessments of accessibility in Canadian urban areas to more theoretically and morally complex evaluations of justice in the country's urban transportation system.

Besides this introduction, we organize this paper as follows. The next section contains a description of the data sources we used to construct the data package. Then, we recount the data processing necessary to create the package. Next, we go through the main contents of the data package, i.e., the travel time matrices estimated through our analysis. We present some basic descriptive statistics of these datasets, and elucidate how one can use them in accessibility analysis. Finally, we conclude by explaining how we expect {canaccesR} to contribute to the urban analytics and science community.

Data and methods

Raw data sources

The locations included in the data package comprise the 12 largest (population-wise) Census metropolitan areas (CMA's) based on the 2016 Canadian Census (Government of

Canada, 2016) [†]. These locations are Toronto, Montreal, Vancouver, Ottawa-Gatineau, Calgary, Edmonton, Quebec City, Winnipeg, Hamilton, Kitchener-Cambridge-Waterloo, London, and Halifax. We used four main data sources to construct the {canaccesR} data package: General Transit Feed Specification (GTFS), OpenStreetMap (OSM), DMTI's Enhanced Points of Interest, and Statistics Canada Census data.

We manually collected and processed the GTFS files from all transit agencies within the selected CMA's to use their information on the public transit schedule in 2019 and 2023. The OpenStreetMap data for the selected areas were collected through the {osmextract} package (Gilardi and Lovelace, 2025). We used OSM data from 2019 and 2023, which provided information on the areas' transit network in two points in time. We collected data from the 2016 Canadian Census using the {cancensus} package (von Bergmann et al., 2022) and used its information on the spatial distribution of the population and the number of workplace locations (employment) across the CMA's (Government of Canada, 2016). Finally, we gathered and cleaned the 2023 DMTI's Enhanced Points of Interest dataset to obtain the location of the groceries stores within every urban area selected (Inc., 2015). We filtered the locations within the DMTI dataset using the grocery stores code from the North American Industry Classification System (NAICS) and the Standard Industrial Classification (SIC). {**THE CODE FOR THIS ESTIMATIONS AND THE TRAVEL TIME MATRICES IS ON THE transit_death_spiral github repo. 1) SHOULD WE CITE IT? 2) IS THAT A PROBLEM?**}

Methods: travel time matrices processing

Using the {r5r} package, we estimated public transit travel times for two destination types, grocery stores and jobs. For each amenity type, we chose a likely travel time and day of the week. We set a 15 minutes time window and the maximum trip duration to 120 minutes. The estimated times are the median of the 15 minute time window. For groceries stores, we set the departure date to a weekend afternoon and the departure time to between 12:00 PM to 12:15 PM on April 20, 2019 and April 22, 2023. For employment, we ran the analysis on a typical weekday morning rush-hour commute, more specifically 8:00 to 8:15 AM departure on Tuesday, April 16, 2019 and Tuesday, April 18, 2023 [‡]. In both cases, we assumed that walking was the mode of travel from origin to transit stop and from transit stop to destination. We aggregated all the resulting travel time matrices at the Dissemination Area (DA) level, which comprise the fundamental unit of analysis in data package.

[†]We included Oshawa, Ontario, as part of the Greater Toronto Area (GTA) because of the former's proximity to the latter. We also included Abbotsford-Mission, British Columbia, as part of the Vancouver metropolitan area because of the former's proximity to a transit station on the region's West Coast Express commuter rail line.

[‡]The one exception is Quebec City, where the routing for 2019 occurs on a Saturday and Tuesday in June (instead of April) due to the GTFS data unavailability.

{canaccessR}'s contents

The package contains the following contents: travel time matrices, socio-economic and demographic data disaggregated at the DA level, the CMA areas' boundaries and backgrounds for plotting the data spatially, and aggregated statistics.

The main contents of the {canaccessR} package are the travel time matrices. These matrices (datasets) comprise the estimated travel time by transit from and to each origin and destination Dissemination Area pairs. The datasets also feature the total population in the origin DA and the total employment in the destination DA. Besides these information, the matrices contain the unique origin and destination DA codes, region and name identifiers, the year in which the travel times refers to, and the date and time of departure for a trip that originates at the Dissemination Area of origin. There are two matrices for each metropolitan area, one containing the travel times to jobs and the other to groceries stores [§]. These information can be turned into spatial data using the socio-economic and demographic datasets.

The socio-economic and demographic data contain other information disaggregated at the DA level for each CMA. These refer to total population by age groups, number of dwellings, number of individuals below the Low Income Measure, etc. The aggregated statistics refer to population aggregates (for the selected CMA's) and transit revenue and ridership data aggregated by regional and national scale.

We now present some descriptive statistics from the travel time matrices contained in the {canaccessR} package. The Table **?** below summarizes the travel time estimates. The travel time datasets contain, in total, 97,784,850 origin-destination pairs (observations) from population to employment locations for all the DA's in the sample and 18,519,897 pairs from population to groceries stores. Considering all areas combined, the mean travel time to jobs was 80 and 76 for groceries stores.

Study Region Name	Destination	Observations	Mean	Sd	P25	P50	P75
All regions	Employment	97,784,850	78	25	59	80	99
Toronto	Employment	41,038,062	82	25	64	85	103
Montréal	Employment	30,091,411	76	25	57	77	97
Vancouver	Employment	12,254,478	76	25	58	77	96
Calgary	Employment	3,022,244	75	23	59	75	91
Ottawa	Employment	2,936,912	74	24	56	74	93
Edmonton	Employment	2,057,486	70	23	53	70	87
Québec City	Employment	1,473,702	72	25	54	72	93
Winnipeg	Employment	1,574,205	62	22	46	61	76
Hamilton	Employment	2,089,172	82	28	61	87	106
Waterloo	Employment	552,826	71	27	49	69	93
London	Employment	415,174	61	21	46	60	74
Halifax	Employment	279,178	66	25	47	65	84
All regions	Groceries Stores	18,519,897	76	26	56	76	97
Toronto	Groceries Stores	8,512,874	80	25	61	82	101
Montréal	Groceries Stores	2,993,965	73	26	53	74	94
Vancouver	Groceries Stores	4,540,106	73	25	53	74	93

[§]The information from Toronto, Hamilton, and Waterloo are aggregated at the Greater Golden Horseshoe travel time matrices, making it 20 datasets in total.

Calgary	Groceries Stores	502,757	70	22	54	70	86
Ottawa	Groceries Stores	619,791	72	24	55	72	91
Edmonton	Groceries Stores	324,030	68	23	51	67	84
Québec City	Groceries Stores	234,600	70	26	51	70	89
Winnipeg	Groceries Stores	362,566	58	21	43	57	72
Hamilton	Groceries Stores	281,118	84	29	60	91	110
Waterloo	Groceries Stores	44,726	66	28	44	62	90
London	Groceries Stores	52,617	55	19	42	55	67
Halifax	Groceries Stores	50,747	63	26	42	63	83

How to use {canaccessR}

This section exemplifies how the package can be used through visual representation. In Figure **?**, we present the spatial representation of the travel time matrices for the metropolitan region of Montréal. In it, we see the median travel time from each DA to employment (left) and to groceries stores (right). The plot shows that moving away from the city core increases the necessary travel time by public transit to reach employment locations and groceries stores.

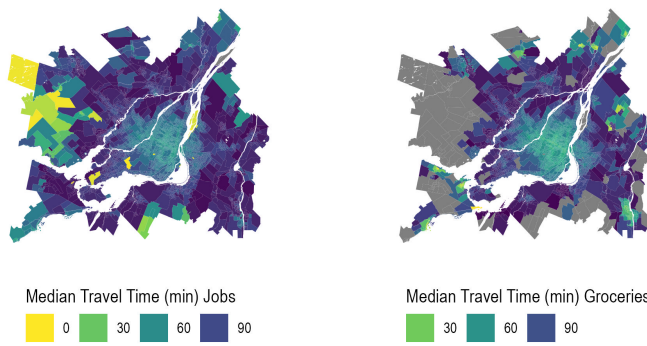

Figure 1. Estimated median travel time (minutes) per Dissemination Area to jobs (left) and groceries stores (right). Public transit travel times are calculated using r5r (Pereira et al., 2021). DA and planning boundaries of the Montréal metropolitan area (Statistics Canada, 2016).

A more thorough example of the package's use can be found on the School of Cities' recent report on Canada's Urban Infrastructure Deficit. In its 11th Chapter, we use the travel time matrices to estimate accessibility metrics to jobs and groceries stores before and after the pandemic (Parga et al., 2024). We then compare how changes affected groups differently according to their spatial distribution and income level, thus making explicit the connection of the package's information and matters of equity in

transportation. The report is freely available for download at the State of Cities Summit website.

Concluding remarks

In this paper, we describe the {canaccesR} data package, created using the {r5r} package and transit schedule, street network, employment, and population data. The package's main contents refers to the ready-to-use travel time matrices for public transit to reach employment and groceries stores in Canada's 12 largest urban areas. We expect the contents of the package to be used in transportation accessibility evaluations within and across those regions. Moreover, these datasets can be used in further equity assessments that evaluate the distribution of accessibility across space and between social groups. Furthermore, in the spirit of open data products (Arribas-Bel et al., 2021), the package can be expanded through collaboration with other researchers by, for example, including travel time matrices to other essential destinations within the DMTI's dataset (*e.g.*, schools, healthcare, etc.). In other words, we hope that by making these datasets publicly available, future analysis can contribute to making Canada's transportation system more just and fair, considering accessibility's as the main social good of transportation (Martens, 2016), and the inherent connection between public transit and the "right to the city" (Coggin and Pieterse, 2015).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Social Sciences and Humanities Research Council of Canada (*More description about the funding source after the review process*).

ORCID

- name: João Pedro Figueira Amorim Parga orcid: 0000-0002-4105-5927
- name: Anastasia Soukhov orcid: 0000-0003-4371-4831
- name: Robert Nutifafa Arku orcid: 0000-0002-2018-886X
- name: Christopher Higgins orcid: 0000-0002-3551-7750
- name: Antonio Páez orcid: 0000-0001-6912-9919

Data availability statement

The {canaccessR} data package can be found and installed on its Github respository.

References

Arribas-Bel D, Green M, Rowe F and Singleton A (2021) Open data products-A framework for creating valuable analysis ready data. *Journal of Geographical Systems* 23(4): 497–514. DOI: 10.1007/s10109-021-00363-5.

Coggin T and Pieterse M (2015) A Right to Transport? Moving Towards a Rights-Based Approach to Mobility in the City. *South African Journal on Human Rights* 31(2): 294–314. DOI: 10.1080/19962126.2015.11865248.

Gilardi A and Lovelace R (2025) *Osmextract: Download and Import Open Street Map Data Extracts*.

Government of Canada SC (2016) 2016 Census of Population.

Government of Canada SC (2021) Dictionary, Census of Population, 2021 – Dissemination area (DA). https://www12.statcan.gc.ca/census-recensement/2021/ref/dict/az/definition-eng.cfm?ID=geo021.

Higgins CD, Páez A, Kim G and Wang J (2021) Changes in accessibility to emergency and community food services during COVID-19 and implications for low income populations in Hamilton, Ontario. *Social Science & Medicine* 291: 114442. DOI:10.1016/j.socscimed.2021.114442.

Humberto M (2023) How to translate justice theory into urban transport metrics? Synchronic assessment of Latin American cities based on equality, priority and sufficiency. *Journal of Transport Geography* 110: 103630. DOI:10.1016/j.jtrangeo.2023.103630.

Inc DS (2015) Enhanced Points of Interest (EPOI).

Martens K (2016) *Transport Justice: Designing Fair Transportation Systems*. New York: Routledge. ISBN 978-1-315-74685-2. DOI:10.4324/9781315746852.

Páez A, Scott DM and Morency C (2012) Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Special Section on Accessibility and Socio-Economic Activities: Methodological and Empirical Aspects* 25: 141–153. DOI: 10.1016/j.jtrangeo.2012.03.016.

Parga JPFA, Soukhov A, Arku RN, Higgins CD and Páez A (2024) Democratic Access to our Cities: The impacts of recent changes to transit services in major Canadian metropolitan areas. Technical report, University of Toronto School of Cities, Toronto, ON.

Pereira RHM, Braga CKV, Servo LM, Serra B, Amaral P, Gouveia N and Paez A (2021a) Geographic access to COVID-19 healthcare in Brazil using a balanced float catchment area approach. *Social Science & Medicine* 273: 113773. DOI:10.1016/j.socscimed.2021.113773.

Pereira RHM, Saraiva M, Herszenhut D, Braga CKV and Conway MW (2021b) R5r: Rapid Realistic Routing on Multimodal Transport Networks with R^5 in R. *Findings* DOI:10.32866/001c.21262.

Soukhov A and Páez A (2023) TTS2016R: A data set to study population and employment patterns from the 2016 Transportation Tomorrow Survey in the Greater Golden Horseshoe area, Ontario, Canada. *Environment and Planning B: Urban Analytics and City Science* 50(2): 556–563. DOI:10.1177/23998083221146781.

von Bergmann J, Shkolnik D and Jacobs A (2022) *Cancensus: R Package to Access, Retrieve, and Work with Canadian Census Data and Geography*.