

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

Journal Title

XX(X):2-4

©The Author(s) 0000

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/ToBeAssigned

www.sagepub.com/

SAGE

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

Abstract

In this paper, we describe the {canaccessR} package, an open data product (OPS) created using the R statistical language. {canaccess} is a data package that provides public transit travel time estimates (travel time matrices - TTM) to employment locations and grocery stores across the 12 largest Canadian metropolitan areas. We calculate these estimates for each Dissemination Area (DA) within these regions for the years 2019 and 2023. 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 to essential services across these regions. These estimates can be used to compare different regions across Canada in terms of their accessibility and to conduct within-region equity assessments regarding access to services, which can inform improvements in transportation policies related to accessibility. 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 OPS, the {canaccess} package allows for open exploration, use, and contribution by users through its GitHub repository.

Keywords

Accessibility; public transit; open data products; OPS; travel time; employment; grocery stores.

Introduction

The objective of this paper is to describe the {canaccesR} Open Data Product (OPS), an open data package. The package's main contents are public transit travel time matrices (TTM) estimates to employment and groceries stores from the 12 largest Canadian metropolitan areas in the years 2019 and 2023. These estimates were created by leveraging expertise in data science, computer programming, and transportation accessibility, using the {r5r} R package [PEREIRA CITATION]. We used transport and street network, population, and employment data from different sources to arrive at our estimations.

The main contribution of this data package is to provide analysis-ready data for Canada's largest cities, thus making the field of accessibility research and urban analytics more accessible. Estimating accessibility, i.e., the potential offered by the transportation system to reach destinations [Páez, Scott, and Morency, "Measuring Accessibility."] requires specialized datasets and technical expertise. By integrating and processing raw data from diverse sources, estimating TTM's for two destinations types (e.g., jobs and groceries) across Canada's largest cities of the country, and making these findings publicly available, we hope this OPS can advance the field of urban analytics in Canada.

We anticipate that these datasets will enable researchers to analyze changes in public transit accessibility to essential destinations, such as employment centers and grocery stores, in the largest urban areas of the country. Consequently, the package addresses the need for ready-to-use public transit accessibility data, helping researchers, transit agencies, and the general public in evaluating equity in transportation accessibility within and across Canada.

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 Sources

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.

General Transit Feed Specification

OpenStreetMaps

Files used for routing.

Census Data

Employment etc.

DMTI's Enhanced Points of Interest

Data processing

Routing, cleaning, etc.

{canaccessR}'s contents

The main contents of the {canaccessR} package are the travel time matrices estimates for all the 12 largest Canadian cities.

Descriptive statistics

Below, we present some of the basic statistics of the

How to use {canaccessR}

This section presents some potential applications of the data package.

Concluding remarks

In this paper, we describe the {canaccessR} data package, created using the {r5r} R 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. 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 JUSTICE], and the inherent connection between public transit and the "right to the city" [COGGIN].

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

Author 1

Author 2

Author 3

Data availability statement

The {canaccessR} data package can be found and installed on its Github [respository](#).