MEASURES OF PUBLIC TRANSIT ACCESSIBILITY TO CULTURAL AND ART FACILITIES IN VANCOUVER

Statistics Canada

ABSTRACT

Proposal to develop a standardized metric for measuring public transit accessibility to urban amenities using travel time. The final deliverable is to provide a user friendly interactive dashboard presenting the public transit accessibility metrics on a 3D geomap in plotly. General Transit Feed Specification (GTFS) data, with open source geographic software such as open trip planner, will be used for constructing city transit network graphs. An Origin-Destination transit time matrix will be used in network analysis to obtain the shortest travel time from the centroid of a dissemination block to the cultural point of interest (museum, art gallery, etc.).

Graham Kerford
Rain Shen
Yuxuan Cui
Luka Vukovic

Partners:
Joseph Kuchar
Bjenk Ellefsen

Instructors:
Scott Fazackerley
Irene vrbik

1 INTRODUCTION

Currently, there is no standardized metric for measuring and visualizing public transit accessibility across urban centres in Canada. The availability of such measures provides a better understanding of an urban city system and can additionally be used as a reference for urban developers and policy makers in the context of city planning and optimization. The development of a standardized method for measuring and visualizing public transit accessibility to urban amenities is of high value particularly in rapidly growing Canadian cities.

Dissemination blocks (DBs) are the smallest unit for quantifying areas in Canada, provided by a Statistics Canada 2020 intercensus file with the polygonal shapes, coordinates, and unique ID's for each DB. DBs act as the points from which we can measure distances across urban landscapes. For example, in a previous study by Statistics Canada, walking/driving distance was mainly used for measuring DB proximity or closeness to services and amenities [1], such as educational institutions, health care, and cultural facilities. However, this study does not include a standardized metric for measuring accessibility using public transit which is a crucial part of the modern urban transportation system. In Canada, 12.4% of the population use public transit as the primary mode of commuting [2]. The percentage of commuters using public transit as their main mode of transportation is especially higher in large cities such as Toronto, Vancouver, and Montreal with a population average of 22.3% [2]. Therefore, this project will take into account transportation via public transit to provide a reliable and standardized measure of "transit efficiency" across Canadian cities.

We must note that computing transit times between all amenities and DBs across all cities in Canada requires enormous computational power and time, which may be impossible to do within the timeframe of this project. As such, this project is to serve as a proof of concept in one city, which we selected to be Vancouver, and with one type of amenity, cultural points of interest (museums, art galleries, etc.), from which later projects can then scale our method across Canada. We will start by applying graphic and network analysis in our transportation model [3]. More specifically, we will be using an open source geographic software *open trip planner* to conduct transportation models [4]. An origin-destination cost matrix will be used to calculate the average travel time from the centroid of each DB to each cultural point of interest in Vancouver [5]. Eventually, our model will be able to scale up nationwide with different amenities of interest. Last but not the least, final deliverable will include an interactive dashboard with an interactive Vancouver city heat map, where each DB's corresponding colour is determined by the transit accessibility score.

2 AIMS AND OBJECTIVES

There are two main objectives for this project:

- To map and visualize Vancouver's measure of accessibility to cultural points of interest such as museums and art galleries as determined by public transit travel time.
- To establish methodologies of data collection, data wrangling, and statistical algorithms that are easy to implement and scale to other municipalities and points of interests.

Stemming from the following primary research question:

- How accessible are Vancouver's cultural points of interest via the current transit system?

Time permitting, we may also be curious in investigating the following:

- Can our transit accessibility technique be used in the context of urban equity analysis? Sustainability analysis (transit vs driving)? Spatial planning for urban development?

3 METHODOLOGY

Data Collection

In order to calculate the measures of accessibility of various facilities via public transit, we will need General Transit Feed Specification (GTFS) data, which defines a common format for public transportation schedules and the associated geographic information. For this project, we will mainly use GTFS static data, which contains different text files of particular aspects of public transit information: trips, routes, stops etc. These files can be parsed into a spatial database which makes schedule information accessible.

The Open Database of Cultural and Art Facilities (ODCAF) dataset comprises a list of cultural and art facilities across Canada. The key attributes that are provided along with its names include the type of institution, its address, city, and province. The latitude and longitude of the institution are also provided. Very little wrangling is required for this dataset with it only having to be filtered for the desired municipalities, cultural institutions, and key attributes. Since we require the location data all institutions without a latitude or longitude were removed from the dataset.

The Census DB dataset represents individual blocks used for the Canadian census. Each block consists of the latitude and longitude of the centroid which is used to indicate the starting point for each trip in our model. The original dataset consists of 40 attributes however the only required attributes are the ids of the datablocks (DBUID), the Census Subdivision Name (CSDNAME), and the latitude and longitude. This would provide a dataset of all of the datablocks in the Greater Vancouver Area (GVA) and the locations to compare the access to other datablocks via public transit.

Vancouver OpenStreetMap Data was provided via *Mapzen* [6] which contains information such as street names, street direction, speed limits, and road laws for the entire road network in the GVA. By overlaying the datasets with OpenTripPlanner, it is possible to visualize the extent of the transic network in the GVA with the route and stop information (**Figure 1**).

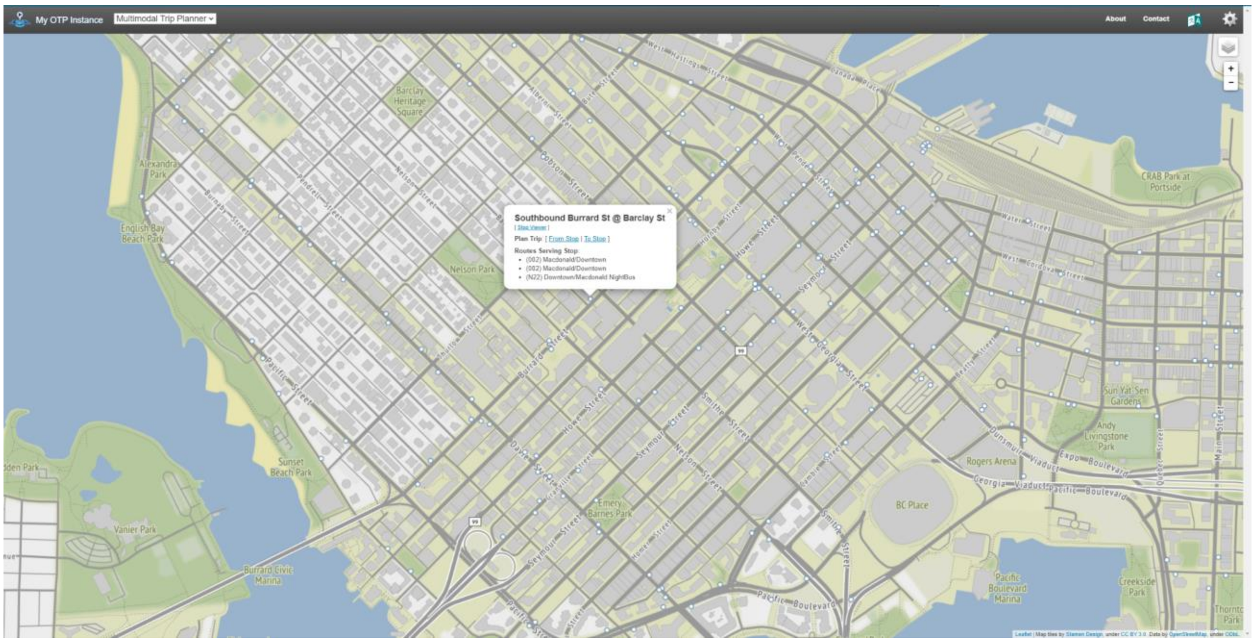

Figure 1. Illustration of the transit network on the Vancouver Peninsula using the OSM and GTFS datasets.

Accessibility Computation

The primary unit of the measures will be travel time via public transit. As we will obtain the public transportation network and proximity measures between DBs, the travel time can be computed by utilizing some open source software listed below. An advantage of using these is that they automatically account for the complexity and variable nature of transit time based on departure times, transfers, waiting times, walking times and so forth.

- **OpenTripPlanner** (OTP), which provides transportation network analysis. By using the datasets mentioned above, a travel time matrix can be estimated by OTP and Python.
- **UrbanAccess**, which can create multi-modal graph networks for use in multi-scale transit accessibility analysis with a Python library, **Pandana**, which is a network analysis tool.

- **R5**, which is Conveyal's engine for multi-modal networks, with a particular focus on public transit. It works by planning many trips at different departure times in a time window and handles both scheduled public transit and headway-based lines.

The main language for writing the scripts to compute the travel time matrix will be Python. If distributed computing becomes a necessity for our project, **GraphX**, a component in Spark for graph-parallel computation via distributed clusters can compute the output efficiently. As another option, a geographic information system such as **ArcGIS** may be used to compute the network distance and estimate travel times.

Visualization

After the network of accessibility to cultural and art facilities via public transit has been obtained, a dashboard can be created to provide visualization of the accessibility as well as interactions with users. Visualization tools, such as **Plotly**, can be used as a Python library to create the final product.

Additional enhancements can be performed by using **Kepler.gl**, which can render millions of points representing thousands of trips and perform spatial aggregations.

Workflow

The project can generally be broken down into three components: graph building, many-to-many transit points computation, and finally dashboard visualization. The first part involves overlaying Vancouver's GTFS data on street data provided by OpenStreetMaps to create a functional transit graph. One useful tool that can be used for this task is OpenStreetPlanner, a widely available open source software. The second part of the project will be computing transit accessibility from each point of origin (dissemination blocks) to each cultural point of interest (museums, libraries, theatres, etc.). Transit travel times will be averaged by block and converted to an index. This leads us to part three of the project where the computed indices will be mapped to GEOJson data and visualized in a final dashboard as an interactive heatmap. The dashboard visualization will be built using either Kepler.gl or Mapbox, the tool used by the client in the predecessor project. The overall workflow for this project can be summarized as shown in **Figure 2**.

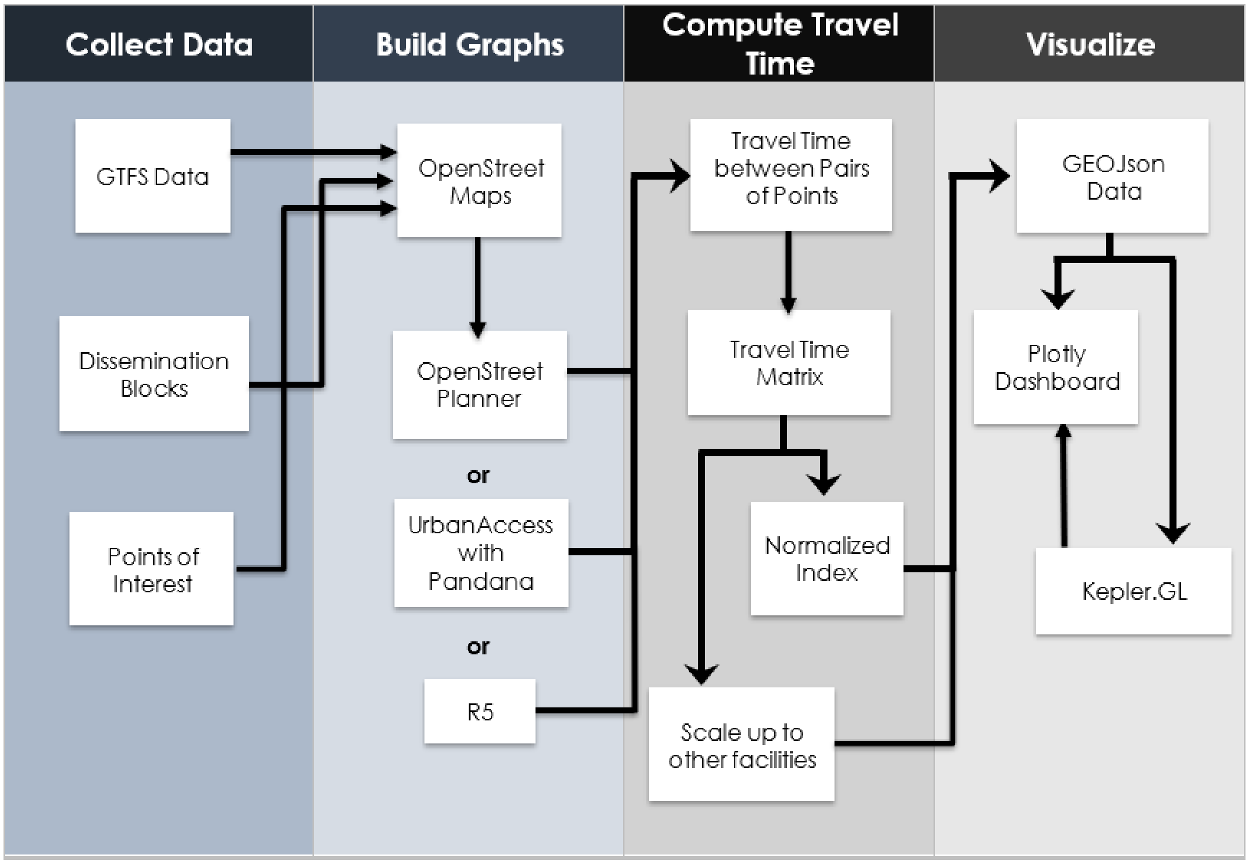

Figure 2. High Level Project Workflow

4 DELIVERABLE SCHEDULE AND TEAM DUTIES

The duration of the project will be from May 4th to June 25th with a mid-project presentation near the end of May. The weekly deliverables and their target dates are broken down in **Table 1**, with our final deliverables (Report and Dashboard) due June 22nd.

Since the nature of our project is very linear, our duties will be split on a weekly basis during our team meetings every Monday, after a quick consultation with Joseph and Bjenk via Microsoft Teams. This allows for the possibility of weekly rotations on general duties, such as tasks related to coding, software implementation, dashboard building, testing, deployment, and documentation. We can usually expect one person to work on documentation for big weekly updates, and all members to log smaller updates in a general updates file. Within this file, we will be able to see general work credit and project status as time goes. Having flexible roles also implies that our deliverable schedule can be a living document. For example, if all goes as planned, we may have someone begin working on a general dashboard format week in advance.

Target Date	Deliverable	Tasks
Week 0: Getting Organized		
May 4 - May 9	Proposal Development	Client Outreach and Meetings Background Research Tool Familiarization
Week 1: Basic Tool Familiarization and Graph Setup		
May 10 - May 11	Raw Data Collection and Organization	To Acquire: - Census DB data (DBUIDs and coordinates) - Cultural points data (IDs and coordinates) - Vancouver GTFS data - Vancouver OSM data - GeoJSON polygon data for DBs
May 11 - May 14	Graph Building and Local Visualization with OTP	Tool installation and familiarization. Data layering (performed by OTP). Graph building and saving so loading can be quick in subsequent local visualization (visualized by OTP). Local StreetGraph rendering.
May 11 - May 14	Small Scaled many-to-many point transit time computation using OTP.	Building a sample set of dissemination blocks and points of interest. Python script preparation for batch transit time computation. Small scale testing of computation times. Test set transit accessibility visualization via the OTP Analyst extension.
Week 2: Transit Time Computation		
May 17 - May 21	Fully Scaled Many-to-many transit time computation.	Optimization of last week's batch transit computation script. Compiling full pointset of all DBs and points of interest. Full scaled many-to-many transit point computation to output a CSV with a point to point time matrix. OTP Heatmap Visualization of transit accessibility to cultural points of interest. If transit times are to be weighted in next week's score building, then this week we will start a small web scraping side project to collect google map review counts for each corresponding point of interest as a metric for popularity.
Week 3: Accessibility Scoring		
May 24 - May 28	DB Accessibility Scores	Trip times will be grouped by DB, averaged, and scaled to create an accessibility score between 0 and 1. If realistic, trip times will be weighed by the popularity or size of a destination point (ie. more popular points of interest have higher weighted transit times).

		If the above point is implemented, last week's mini web scraping project will need to be completed as well.
Week 4: Preliminary Deployment		
May 31 - June 4	Initial Dashboard Visualization	Initial dashboard Building with Plotly and Dash. Familiarization with Keplr.gl and/or MapBox for DB Score visualization. Preliminary deployment on Heroku.
Week 5: Deployment		
June 8 - June 11	Fully Functional Deployment of the Interactive Heatmap Dashboard	Dashboard aesthetics, optimization, and official deployment via Heroku.
Week 6: Documentation and Buffer Week		
June 14 - June 18	Additional Analysis and Final Touch Ups Drafting of the Final Report Project Packaging	Any additional analysis or fixes we may need to do. Writing of the final report. Documentation, code packaging, report drafting, and presentation outlining.
Project Presentation Week		
June 22	Final Report	-
June 25	Capstone Presentations	-

DB: Dissemination Block; **DBUID**: Dissemination block unique identification **GTFS**: General Transit Feed Specification **OSM**: OpenStreetMaps; **OTP**: OpenTripPlanner

Table 1: Target deliverable timeline and required tasks for completion.

REFERENCES

[1] Statistics Canada, Alasia, A., Newstead, N., Kuchar, J., & Radulescu, M. (2021, February). *Measuring proximity to services and amenities: An experimental set of indicators for neighbourhoods and localities*. https://www150.statcan.gc.ca/n1/pub/18-001-x/18-001-x2020001-eng.htm

[2]Statistics Canada. (2017, November). *Commuters using sustainable transportation in census metropolitan areas*. https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016029/98-200-x2016029-eng.cfm

[3]Mavoa, S., Witten, K., McCreanor, T., & O'sullivan, D. (2012). GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of transport geography*, *20*(1), 15-22

[4] Pereira, R. Grégoire, L., Karner, A. (2019). *otp-travel-time-matrix* (version 1.3.0) [Tutorial with reproducible example to estimate a travel time matrix using OpenTripPlanner and Python]. https://github.com/rafapereirabr/otp-travel-time-matrix

[5] Krishnakumari, P., van Lint, H., Djukic, T., & Cats, O. (2020). A data driven method for OD matrix estimation. *Transportation Research Part C: Emerging Technologies*, *113*, 38-56.

[6] Interline Technologies LLC. OSM Extracts by Interline https://www.interline.io/osm/extracts/

APPENDIX

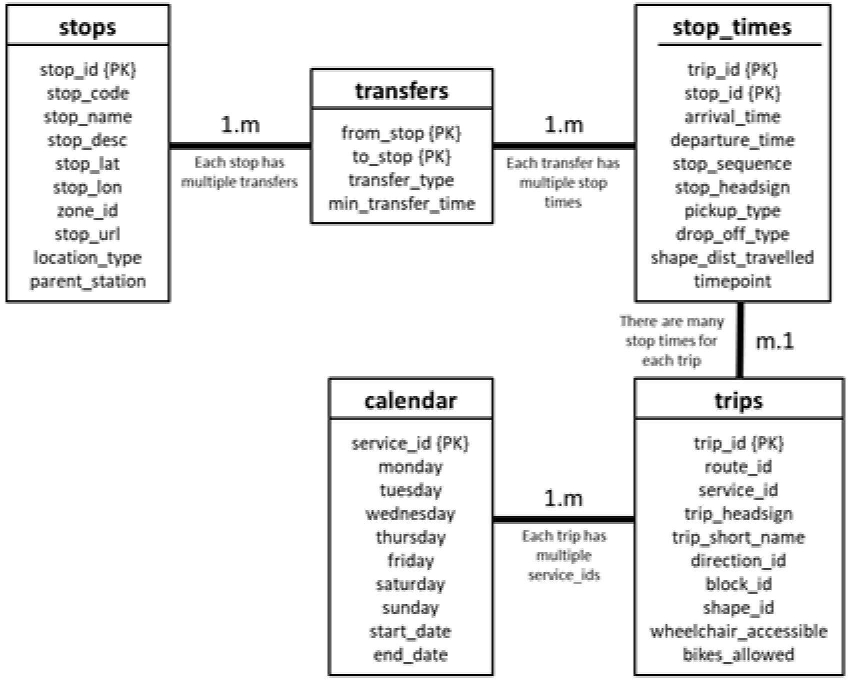

Figure 3. Relational database of the GTFS renabling to identify the time between stops based upon the stopd_ids, time and date.