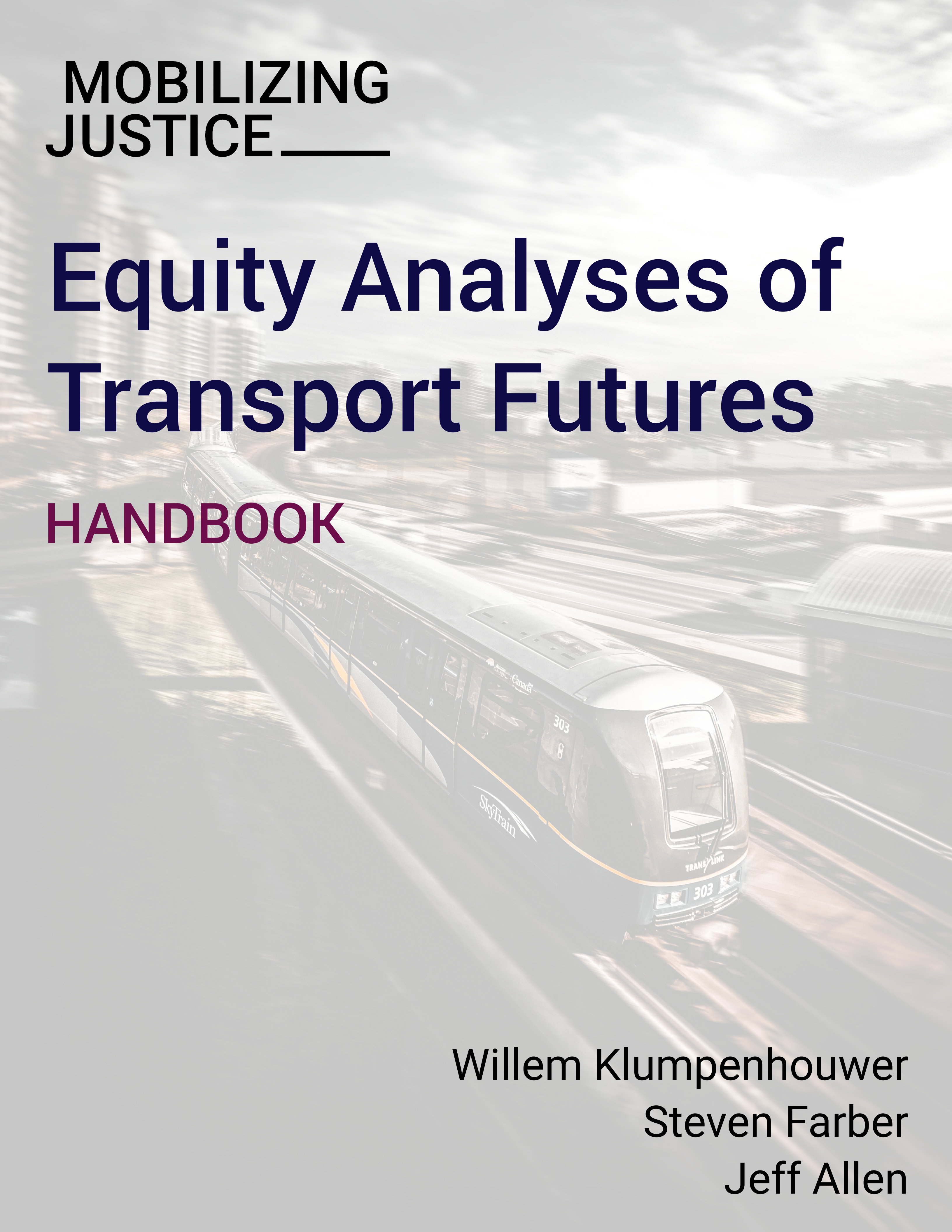
MOBILIZING JUSTICE

Equity Analyses of Transport Futures

HANDBOOK



Willem Klumpenhouwer
Steven Farber
Jeff Allen

Acknowledgements

This project was co-developed by:

Dr. Willem Klumpenhouwer
Postdoctoral Fellow
Transit Analytics Lab | University of Toronto;
Research Consultant
Klumpentown Consulting
`https://www.klumpentown.com/`

Dr. Steven Farber
Associate Professor
Department of Human Geography | University of Toronto Scarborough
`https://mobilizingjustice.ca`

Dr. Jeff Allen
Lead, Data Visualizaiton
School of Cities | University of Toronto
`https://jamaps.github.io/about.html`

The development of these resources was funded by:

MOBILIZING JUSTICE

Mobilizing Justice is a multi-sector research partnership committed to solving transportation inequities in Canadian cities. We are an unprecedented coalition that includes 33 academics from 6 Canadian provinces and 3 countries, 14 government agencies, 7 transportation companies, and 7 non-profit organizations. All of our partners are steadfastly committed to delivering equitable transportation systems. Mobilizing Justice is supported in part by funding from the Social Sciences and Humanities Research Council's partnership grant: Mobilizing justice: towards evidence-based transportation equity policy.

Cover image by Mado El Khouly

Contents

1 Introduction

This document is a reference guide and a companion handbook for a series of quantitative analyses that can be used to understand the distributive equity impacts of public transit or other transportation infrastructure scenarios or projects. The foundational work contained within this notebook and the accompanying analysis code was developed by Drs. Jeff Allen and Steven Farber at the University of Toronto in 2021,[1] but has been updated for new datasets and for use with Python notebooks.

This handbook outlines four ways to measure how changes in travel times and demand across various potential infrastructure scenarios are distributed across different demographic and socioeconomic groups. This includes station or project area access, changes in travel times, changes in access to opportunities, and changes in the number of trips made by a certain mode (normally transit) as a measure of induced utility. Finally, we use a simplified simulation model to project demographic shifts into the future based on broader demographic trends. Figure 1 diagrams the various inputs, analyses, and outputs that are discussed in this handbook, with the output steps matching the major headings in this document.

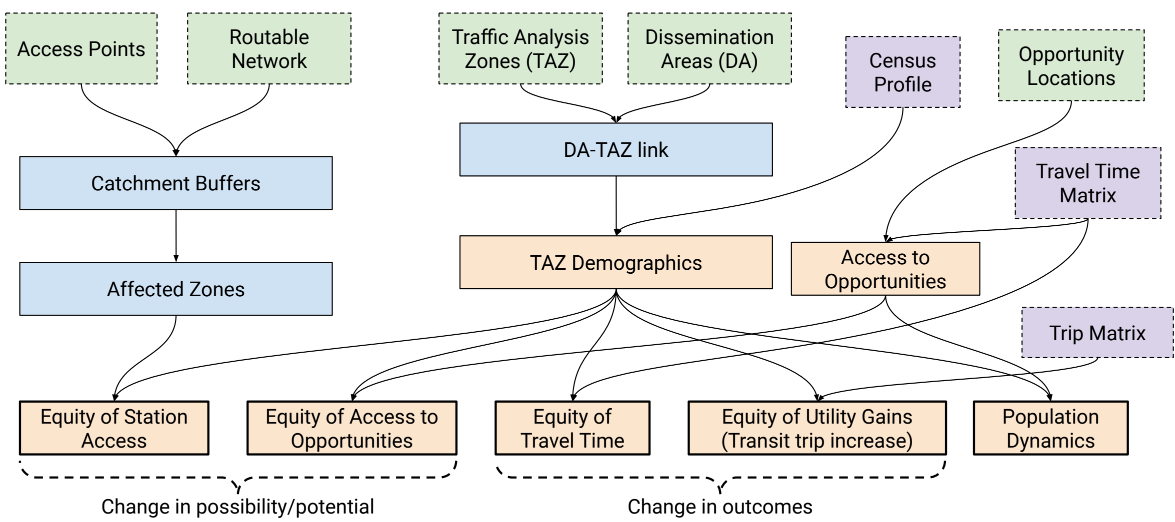


Figure 1: Data and work flow for calculating various measures of impacts of infrastructure projects on nearby populations. Blue tasks are spatial/GIS tasks, while orange-yellow tasks use tabular data analysis methods.

This handbook also contains a series of tasks which can be completed using the data and workbooks provided at `https://github.com/wklumpen/equity-transport-futures`. With these workbooks and the steps provided, you can produce the same data outputs and visualizations shown throughout this handbook. You can learn more about setting up your computer to follow along in Section 2.

It is important to approach the analyses in this handbook with an appropriate mindset. The goal is not to assign a quantitative "equity score" to a project or scenario; this handbook is designed to enable the kind of quantitative analysis that can be added to the strategic case portions of an initial business case, or wherever the discussion warrants. The specific groups used in this handbook are commonly used as representative equity groups in academic literature, however they are intended here as an example, not a prescription. Depending on the goals of your specific analysis, the context of the project, or the community in which you are working, different population groups may be more appropriate.

1.1 TERMINOLOGY AND NOTATION

Quite often, spatial data and census categories have interchangeable and somewhat ambiguous terms. To minimize confusion, in this handbook we use the following terminology and notation:

- **Areas, zones, origins, and destinations** refer generally to places that individuals travel *from* and *to*. In many cases, we represent a larger area of origin or destination with a single point, such as the centroid of a traffic analysis zone (TAZ) or other area. When referring to a general origin we will use the symbol and subscript i, and for a destination we use the symbol j.

- **Travel times** are used in a number of ways in the analyses in this handbook. Depending on the context, travel times can be derived from measured travel times, or modelled real or perceived travel times. We use the symbol $t_{i,j}$ to refer to travel times from origin i to destination j.

- **Trip counts or demand** measures the total number of trips flowing between two areas. We will use the notation $\tau_{i,j}$ for the count of the number of trips flowing from origin i to destination j. Many models use the approach of modelling an average number of trips on a "typical Tuesday in the fall", or some other day where demand is most stable. For this reason, demand value may not always represent whole numbers of trips.

- **Scenarios** refer to various possible situations or model outputs that we wish to compare. We will use superscript notation to compare two scenarios, for example $\tau^0_{i,j}$ refers to the trip counts flowing between origin i and destination j under scenario 0

(typically a "business as usual" scenario), while $\tau^1_{i,j}$ refers to the same type of data but for scenario 1.

- **Population and household subgroups**[2] are subsets of a given population or household count that fall under a certain categorization. For example, the number of visible minorities in a dissemination area is a population subgroup of the total number of people in the dissemination area. Similarly, the number of zero-car households in Toronto is a subset of the total number of households in Toronto. We will use p_i to represent a population subgroup in an area i while P_i represents the total population count in the same area.

1.2 SCENARIO ANALYSIS CONTEXT

For a comparative analysis of travel times, trip flows, and access to opportunities, we will need to have two or more scenarios which we can use to compute changes. These two scenarios can result from any sort of different modelling choice such as changes in economic forecasting, different infrastructure planning, different operational schemes for a transit service, or any other possible model change that would result in differences between the travel time or demand for travel within the study area.

In this handbook we are using data from a series of models investigating the effects of adding additional rail infrastructure and service in the Greater Toronto Area in Ontario, Canada. This project was dubbed "SmartTrack" and was studied by the City of Toronto with modelling assistance from the University of Toronto. The scenarios described below are examples from this study, however as of this publication date the SmarTrack program as modelled here has not been implemented. All model runs are forecasts of demand and travel times in the year 2041, and include the following scenarios:

- Baseline or "Business as Usual" Scenario (BAU)
- Scenario A: SmartTrack with 5 Minute Headways, using a base transit fare
- Scenario B: SmartTrack with 15 Minute Headways, using a base transit fare
- Scenario C: SmartTrack 15 Minute Headways, using a premium fare based on GO Transit fare structures

It is important to note that while these scenarios reflect real data used for planning and policy purposes, the results that we produce here are intended for teaching and illustrative purposes. This handbook is not intending to make a statement about the efficacy or quality of the SmartTrack program, but rather demonstrate how changes in scenarios can have a rather extensive effect on access and its distribution among population groups. Finally, while the model includes the entire Greater Toronto and Hamilton Area surrounding Toronto, the demand matrices provided are filtered to only include the City of Toronto itself. This makes the matrices a more manageable size and the visualization and interpretation of the results somewhat simpler for this handbook.

2 Getting Set Up

In order to perform the analyses described in this handbook, we will need a computational workspace environment that is capable of doing spatial calculations, the ability to generate semi-realistic walk sheds or walk buffers, and the ability to manipulate, aggregate, and perform other statistical calculations on large data sets.

There are a myriad of ways to set up your workspace to accomplish the analyses outlined in this handbook. If you are following along directly with the provided code handbooks and data sets, we will be using an open-source geographic information system software package called QGIS to perform spatial calculations and walk buffer analyses, and the programming language Python to accomplish the data manipulation, calculation, and visualizations.

2.1 SETTING UP YOUR WORKSPACE

Task 1 provides details on obtaining and setting up the appropriate software and data sources, Task 2 provides instructions on setting up a Python environment for the analyses in this handbook, and Task 3 provides an additional configuration for the QGIS software to allow you to calculate walking isocrhones for station area analysis.

The data used for computing walking distances from stations uses OpenStreetMap, a community-edited map of the world. This resource provides a generally more comprehensive walking network compared to other open data sets provided by cities. Nevertheless, some limitations exist in the data, such as missing paths through subway stations that can lead to walking sheds that may not be entirely accurate.

Task 1: Set up QGIS, data sources, and coding notebooks

1. Download and install QGIS from `https://www.qgis.org/en/site/`. The latest stable or long-term release version should be sufficient for the analysis we are conducting.

2. Download or clone the code and data sources from `https://github.com/wklumpen/equity-transport-futures`. You can download a zipfile of the code repository by selecting "Code" followed by "Download ZIP".

3. In the `counts` folder inside the `data` folder, create a folder named `output`. This is where the workbooks are set to produce output data.

Task 2: Set up the Python environment

Depending on your level of experience with Python and whether you have a version of Python set up on your system, choose one of the approaches below to complete your setup.

New to Python: The easiest way to set things up from scratch is using a Python distribution known as *Miniconda*. You can download and install Miniconda from `https://docs.conda.io/en/latest/miniconda.html`.

New to Python/Conda User: If you have set up your environment using Miniconda as described above or with a different Conda distribution, you can set up a new Conda environment by opening an Acaconda console and running

```
conda create -n equity-futures python=3.10
```

choosing `y` when prompted to proceed. Activate the environment with

```
conda activate equity-futures
```

and then install the required libraries

```
conda install -c conda-forge pandas scipy altair jupyterlab
```

again choosing `y` when prompted.

Existing Python User: You can install the required libraries from the standard Python package installer. It's recommended to set up a new environment first by following the instructions at `https://packaging.python.org/en/latest/guides/installing-using-pip-and-virtual-environments`.
You can then install the required packages with

```
pip install pandas scipy altair jupyterlab
```

Task 3: Set up Open Routing Service with QGIS

There are a number of tools available to calculate walking buffers and isochrones from points. Here, we will use a free QGIS plugin and web API tool called Open Routing Service (ORS).

1. Sign up for an API key from Open Routing Service. You can do that at `https://openrouteservice.org/dev/#/signup`.

2. Using QGIS:

   2.1. Go to Plugins → Mananage and Install Plugins

   2.2. Search for "ORS Tools" and install the plugin

   2.3. Once installed, go to Web → ORS Tools → Provider Settings and paste your API key into the first box.

2.2 DATA COLLECTION AND WRANGLING

In order to perform all of the analyses outlined in this handbook, we need to assemble a number of data sources:

- **Location of access points to the study system.** This typically consists of a point file containing station locations.

- **Travel time matrices between origin and destination areas.** These travel times can be calculated for existing transit services using tools such as OpenTripPlanner or R5(py),[3] however we assume here that travel time matrices are readily available to the analyst, for example from model outputs from the Greater Toronto Area Model.

- **Origin population characteristics**, including demographic and socioeconomic status. This information is available at various levels of spatial granularity from the Census.[4] In this handbook we are going to assume that the analyst is able to retrieve DA-level census profile information for the extent of the analysis area.

- **Destination and opportunity data.** Access to opportunities or destinations is required to calculate access measures. For each destination of interest, data of where these opportunities are located is required.

2.3 LINKING TRAFFIC ANALYSIS ZONES AND DISSEMINATION AREAS

Since census demographic information and travel times generated by models use different geographical boundaries (e.g. dissemination areas vs traffic analysis zones), we need to link these two boundaries together. In this example we will use the traffic analysis zone as the "base unit of measurement", since travel times are associated with these zones. The goal is to calculate how much of each dissemination area falls within a given traffic analysis zone. We will do this with a process called *areal apportionment*.

With areal apportionment, we assume that the number of people of any population group in a given zone is evenly distributed within that zone, and so the fraction of that population zone's area that overlaps with another zone is equal to the fraction of population that falls within that zone. Figure 2 shows an example of how a population zone with 100 people gets apportioned. In this example, 35% of the blue zone (the green area) falls within the analysis zone shown in orange, and so 35 people from that particular blue zone are allocated to that particular orange zone. Task 4 outlines the steps required to complete this apportionment.

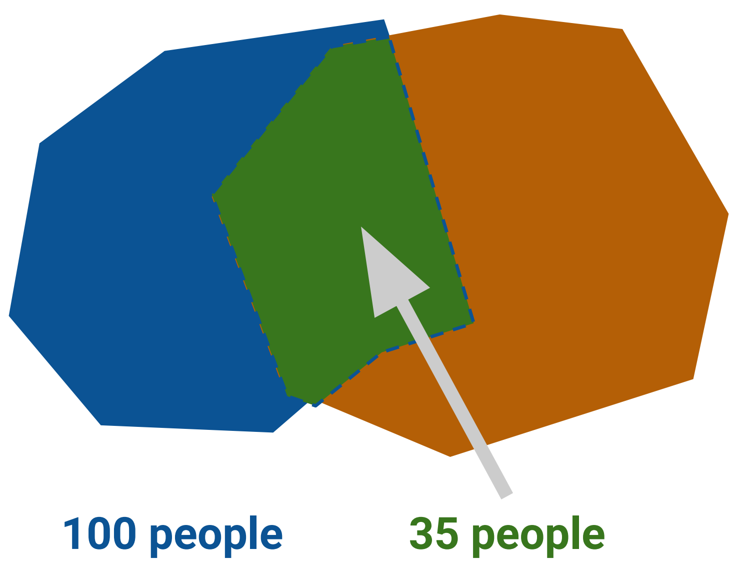


Figure 2: An example of apportioning 35 of the 100 people in the blue zone to the orange zone.

This type of direct areal apportionment is not sensitive to the distribution of the population *within* these zones. Ideally, population distribution-based approaches to apportionment would be more accurate. In this case, we will consider this method detailed enough for our analysis purposes.

Task 4: Perform an areal apportionment in QGIS

1. Load the layer to be apportioned (`dissemination_areas.geojson`) and ensure it's selected. You will also want to ensure the layer is in a projected format, which allows you to perform distance and area-based calculations.

2. Create a new field `da_area` for the layer to be apportioned type "decimal" and determine the area in square km with

$$\text{da\_area} = \text{\$area}/1000000 \tag{1}$$

3. Load the underlying geographical layer (traffic analysis zones) called `traffic_analysis_zones.geojson`

4. Compute the *intersection* of the underlying layer with the layer to be apportioned. The intersection tool can be found in the QGIS Processing Toolbox (View > Panels > Processing Toolbox).

5. Using the newly produced intersection layer, create a new field of type "decimal" and calculate the fractional area of the intersected piece with:

$$\text{frac\_da\_in\_taz} = (\text{\$area}/1000000)/(\text{zone\_area}) \tag{2}$$

6. Export the intersected layer as a CSV. Keep the two linked zone ID columns and the fractional area calculated.

3 Station Areas

The goal of this analysis is to determine the demographic composition of the area near stations or access points of interest. This particular analysis focuses solely on the populations that live nearby stations, and does not account for the level of use of the system or any other operational aspect. Equity of station access is a good first look at the composition of the communities and neighbourhoods that make up the station areas, and can afford you a better understand who might be affected by their proximity to new infrastructure.

To do this, we will need to calculate walking isocrhones from stations, often called “walking buffers” or “walk-shed analysis”. The simplest way to do this is by drawing a circle of a given radius around the study stations, however circular buffers often greatly overestimate the distance reachable on foot from a given station (see Figure 3). A better alternative is to use the underlying walking network to calculate walking buffers which can be done for projects with a relatively small number of points using the service set up in Task 3 and the steps described in Task 5.

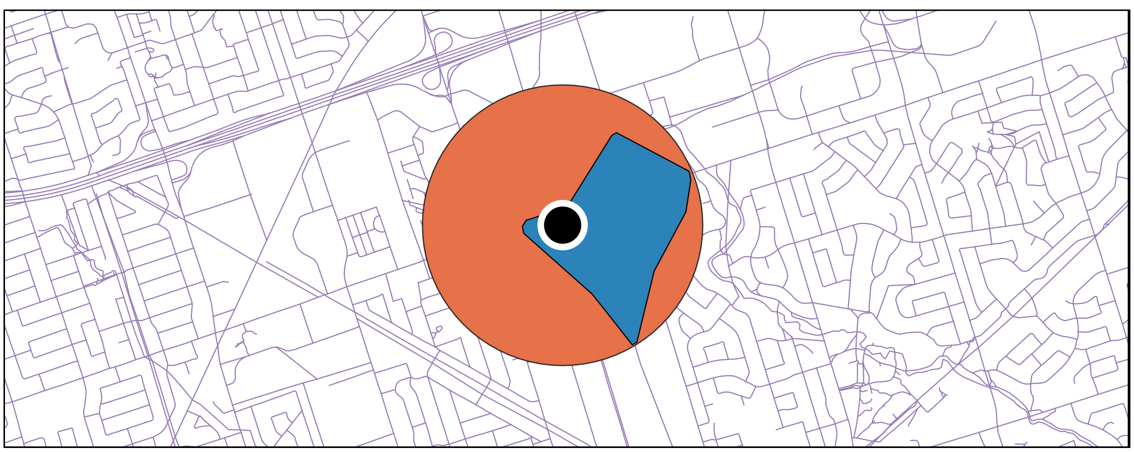


Figure 3: An example of the contrast in area between a circular buffer (orange) and a walking network based buffer (blue) at Ellesmere GO station. Note that the walking buffer is subject to the location of the station point and the underlying walking network supplied by OpenStreetMap.

Task 5: Perform a walking buffer calculation in QGIS

Using QGIS:

1. Load the point layer `smart_track_stations.geojson` containing the station names and locations,
2. Ensure your project/data is in a projected coordinate system (set the project projection appropriately)
3. Go to Web → ORS Tools → ORS Tools
4. Choose the “Batch Jobs” tab → Isochrones from Layer
5. Choose the loaded layer as the input layer, `station` as the Input layer ID field, `foot-walking` as the travel mode, `distance` as the Dimension, and is typical), the dimension and `800` in the “Comma-separated ranges” option.
6. Click on “Run” and wait for the process to finish. During the calculation you may see a `OverQueryLimit` warning - this means that to avoid being overloaded the service is temporarily limited, but will resume shortly. A new layer will be generated with the appropriate isochrones.
7. Export the buffer layer as `station_buffers.geojson`, using the standard EPSG:4326 projection, and keeping only the `station` field. Ensure that "Add saved file to map" is checked for future tasks.

Make sure to check the resulting buffers for any odd issues or errors. Walking networks aren’t perfect and you may end up with some oddly-shaped buffers if there is a critical link missing. You can edit the buffer shape after to be more realistic.

Once an isocrhone or walking buffer has been calculated, we can intersect these buffers with our population data layer to determine the approximate demographic makeup of the various station areas. Task 6 provides the steps to complete this apportionment.

3.1 AREAL APPORTIONMENT OF BUFFERS

Once we have created our buffer(s), we can perform an areal apportionment of census or traffic analysis zones onto the buffers.[5] This is done via the following steps:

Task 6: Apportion the walking buffers over the demographic layer

Using QGIS:

1. If not already loaded, load the `station_buffers.geojson` file. We'll call this the "buffers layer".

2. Load the `dissemination_areas.geojson` layer file to be apportioned. We'll call this the "demographic layer".

3. If you have not already calculated the areas for the demographic layer in Task 4, select the demographic layer and create a new field named `zone_area` of type "decimal" (field length 10, precision 7) and determine the area in square km using the expression
$$\$area/1000000 \tag{3}$$
If you created this field, save it by toggling edit mode off.

4. Calculate the *intersection* of the zonal layer with Vector → Geoprocessing Tools → Intersection. Set the demographic layer as the input layer and the buffers layer as the overlay. A new layer will be created called Intersection.

5. Select the newly created Intersection layer and create new field named `frac_area` of type "decimal" (field length 10, precision 9) and use the following expression to calculate the fractional area of the intersection:
$$(\$area/1000000)/(zone\_area) \tag{4}$$

6. Export the intersected layer as a comma separated value (CSV) file with the dissemination area ids (`DAUID`), the station names (`station`) and the fractional area (`frac_area`) columns.

Once we have produced a file containing the fractional overlap between our demographic layer and our station areas, we can determine the composition of nearby communities. For example, if we were interested in determining the percentage of people nearby a given station who are identified as visible minorities, we would use the following steps:

1. Multiply the number of visible minorities in a given dissemination area by the fraction of that zone's area that overlaps the station area.

2. Multiply the "total" or summary category for visible minorities in a dissemination area

by the fraction of that zone's area that overlaps the station area.

3. Sum all of the visible minority and total population counts for each given station.
4. Divide the two summed values calculated for each station and multiply by 100 to get the fractional composition near the station.

Task 7: Compute station-level equity

Work through the accompanying Jupyter notebook entitled `Station Areas.ipynb`. You should be able to reproduce Figure 4.

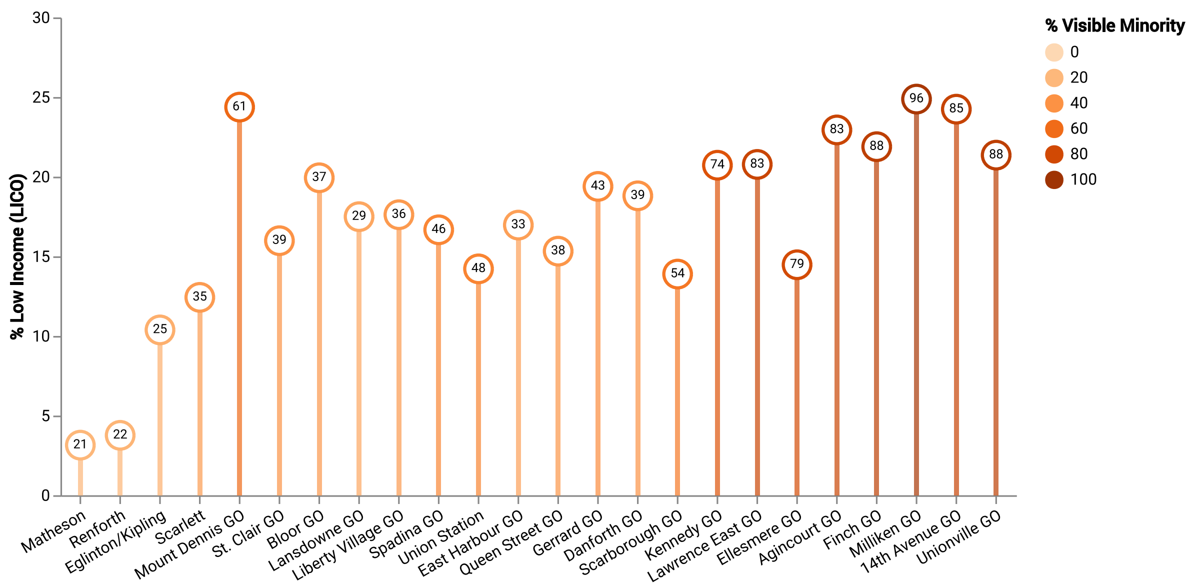


Figure 4: Low-income and visible minority compositions within a 10-minute walk of SmartTrack stations.

4 Access to Opportunities

Spatial access to opportunities involves a calculation of the total number of a given opportunity that can be reached with a certain level of effort from a given origin. Mathematically, there are many ways to formulate this type of measure, and there is extensive ongoing research and debate about the best ways to formulate the measure and what opportunities to access. This handbook outlines one approach, which can be adapted and adjusted as deemed appropriate for any type of application or based on updated research. A good access measures tries to strike a balance between realism and interpretability.

Broadly speaking, there are two families of measures to consider when calculating access, as diagrammed in Figure 5:

- **Cumulative measures:** the sum of items of interest reachable from a location, weighted by their proximity. Mathematically, we can define cumulative accessibility by[6]

$$A_i = \sum_{j \in J} X_j I(\cdot) \tag{5}$$

  Where A_i is the access measure at an origin i, j is a destination in a set of destinations J, X_i is the number of opportunities at the destination j, and $I(\cdot)$ is an *impedance* function which weights the total number of opportunities by some function.[7]

- **Minimum cost measures**: the minimum travel cost to reach X opportunities (e.g. what is travel time to the nearest grocery store, or the minimum travel time to the nearest 3 libraries). Mathematically:

$$A_i' = \max\{\min_n\{C_{i,j} X_j\}\} \tag{6}$$

  where $C_{i,j}$ is the cost of travel between origin i and destination j, X_j is the number of opportunities at the destination j, and we are choosing the n smallest.

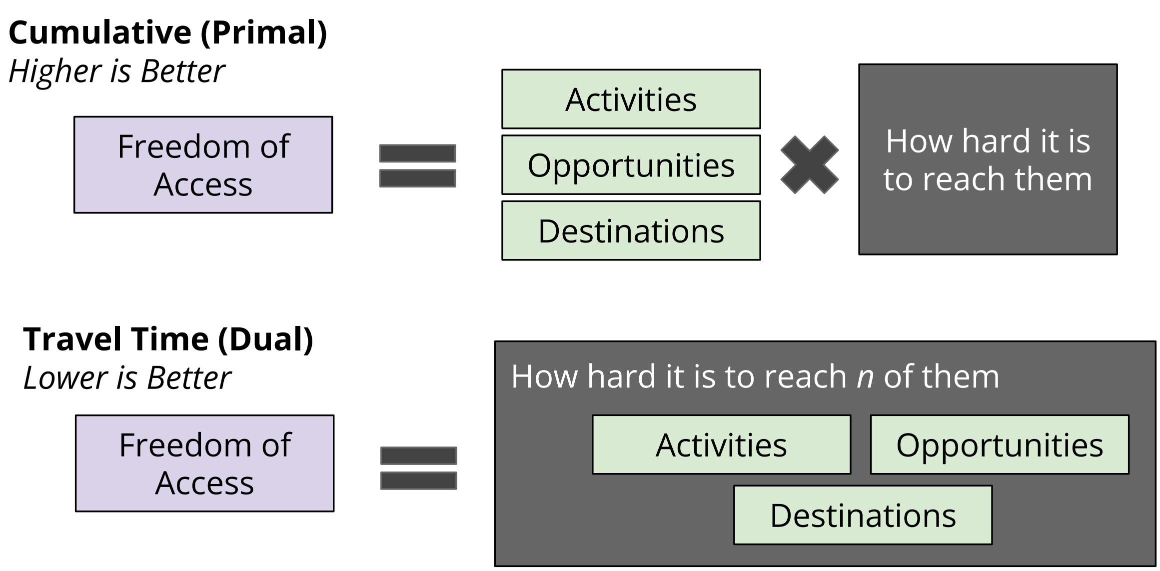


Figure 5: Diagrammatic explanation of cumulative and travel time measures of access.

In this handbook, we will use a **cumulative** measure calculation, as it is more commonly used, better suited for the opportunity we are using (employment), and mathematically more intuitive.

4.1 IMPEDANCE

For a cumulative measure, we need to decide on how to quantify the concept of “how hard it is to reach a destination”. It is generally accepted that destinations that require more travel time to reach them, are more difficult to reach, so most measures are a function of travel time. While other costs and difficulties can provide more nuanced insight, for clarity and simplicity we will use a measure that is a function of travel time alone.

Since we are multiplying counts of opportunities by a weighting function, an impedance function is used that varies between 1 (low impedance, no diminished access) and 0 (low impedance, no access counted at all). There are many philosophies behind choosing the shape of this function,[8] mostly centered around the trade-off between precision and clarity in explanation. Many planners and anlaysts use a simple threshold-based method, where any opportunity beyond a cutoff point (30 minutes, for example) is not counted in the measure, and all opportunities within that range are given full weighting in the measure.

A travel time threshold impedance function for a cutoff time of t_0 can be mathematically defined as

$$I(t_{i,j}) = \begin{cases} 1 & t_{i,j} \leq t_0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The main issue is that this creates quite an abrupt cutoff point. For a $t_0 = 30$ minute cutoff, this means that a destination 31 minutes away is worth nothing in the measure, while one 29 minutes away is worth full credit.

4.2 IMPEDANCE ESTIMATION FROM TRIP LENGTHS

In this handbook, we will “smooth out” this abrupt cutoff, but keep the general shape of that function by using a cumulative normal distribution curve. Specifically, we define our impedance function as

$$I(t_{i,j}) = 1 - \text{CDF}(t_{i,j}) \tag{8}$$

To estimate this function, we will use the following steps:

1. Calculate the weighted mean of trip lengths in a given set of origins and destinations as

$$\mu = \frac{\sum_{i,j} \tau_{i,j} t_{i,j}}{\sum_{i,j} \tau_{i,j}} \tag{9}$$

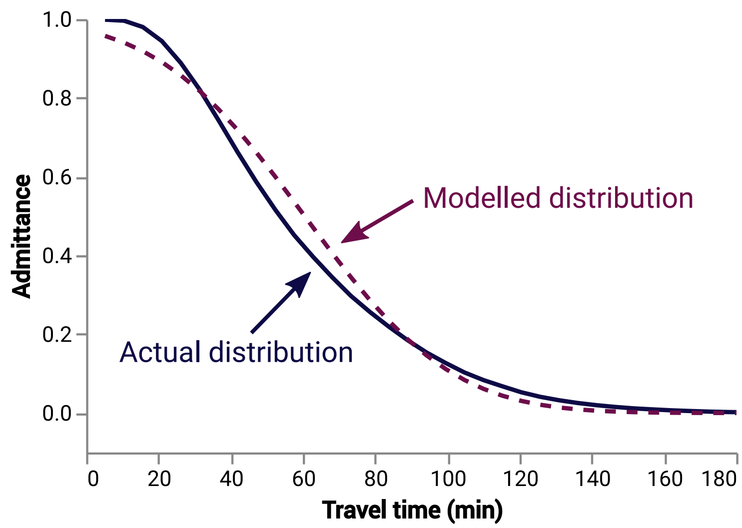


Figure 6: An impedance function estimated from trip lengths in the business as usual scenario of the study model.

2. Calculate the weighted standard deviation of trip lengths as

$$\sigma = \sqrt{\frac{\sum_{i,j} \tau_{i,j}(t_{i,j} - \mu)^2}{\sum_{i,j} \tau_{i,j}}} \tag{10}$$

   Where $\tau_{i,j}$ is the flow of trips between zones i and j.

3. Determine the impedance function as:

$$I(t_{i,j}) = 1 - \text{CDF}(t_{i,j}) \tag{11}$$

   for a Normal distribution, this is

$$I(t_{i,j}) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{t_{i,j} - \mu}{\sigma\sqrt{2}}\right)\right] \tag{12}$$

   Where erf$(\cdot)$ is the Gauss error function. We are making an assumption here that the distribution is a Normal distribution. If you are unsure, it can be helpful to plot the data as a histogram to see for yourself. Figure 6 shows an example of a modelled and actual distribution.

4.3 CALCULATING ACCESS

Calculating access requires the following steps for each origin zone i.

1. Determine the impedance function value from origin i to a given destination j.

2. Multiply this impedance value by the number of opportunities at the destination j.

3. Sum the resulting value for each j.

The resulting "score" will capture the level of access to the given opportunity from that origin.

If we repeat these calculations for two scenarios, we can compute a percent difference in access by each origin i. For example:

$$\Delta_{A_i} = 100 \left(\frac{A_i^1 - A_i^0}{A_i^0} \right) \tag{13}$$

Where Scenario 0 is a "do nothing" scenario, and Scenario 1 includes the addition of an intervention such as a new transit service.

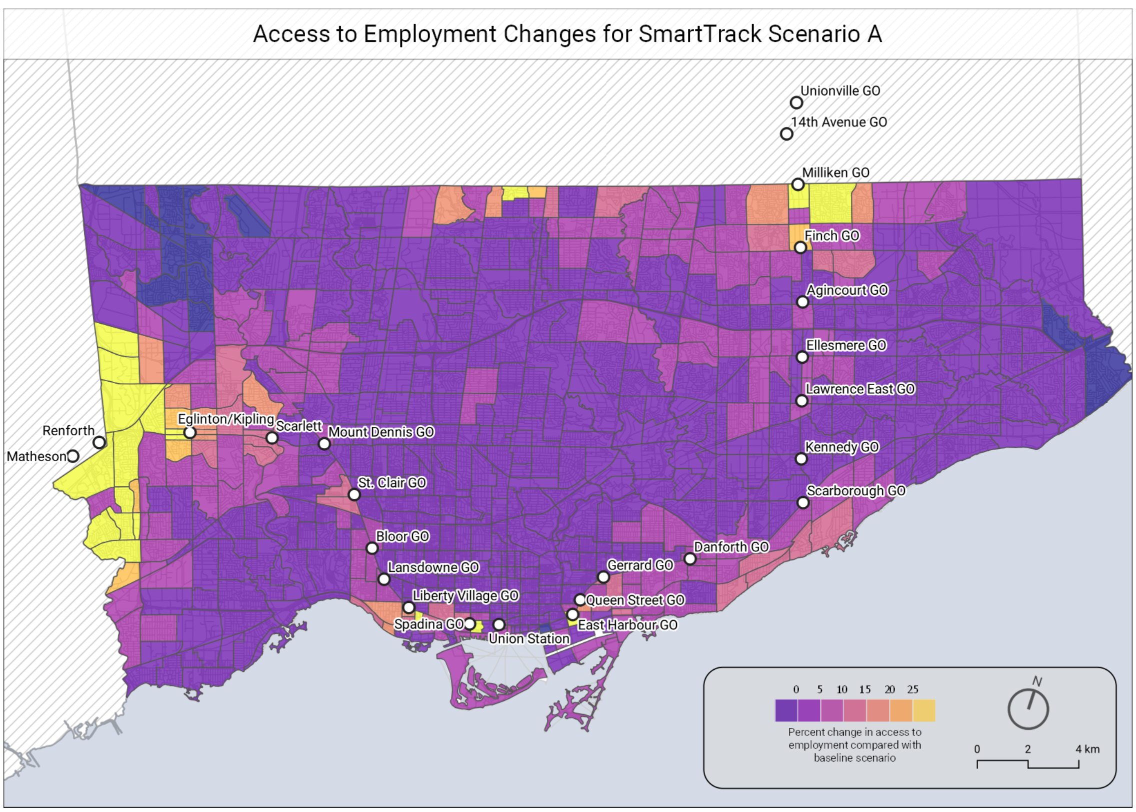


Figure 7: Comparative access to employment induced by Scenario A.

4.4 REGIONAL POPULATION DISTRIBUTION SUMMARIES

Once an access score A_i for each zone or origin i has been calculated, we can use the origin zone's characteristics to summarize the sociodemographic distribution of the access throughout a given region. For a given population subset (e.g. unemployed individuals), we can calculate the regional population weighted summary as follows:[9]

1. Determine the fractional population of each origin i by dividing the count of the population subset in the zone N_i by the total sum of that population throughout the region:

$$n_i = \frac{p_i}{\sum_i p_i} \tag{14}$$

2. Multiply that fractional population n_i by the access score A_i and sum over all zones:

$$A = \sum_i A_i n_i \tag{15}$$

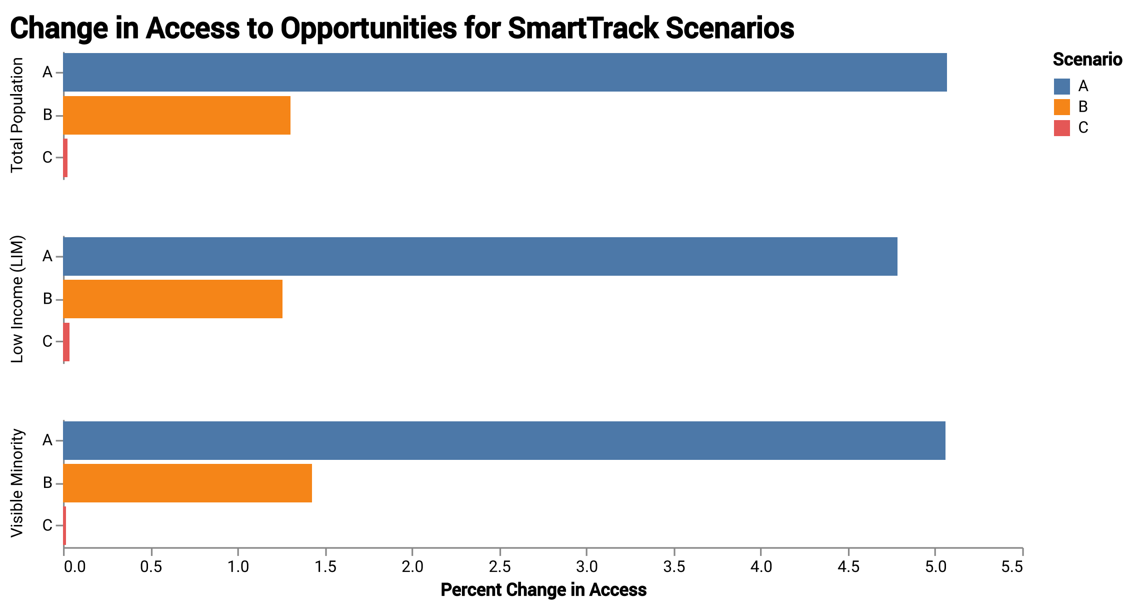


Figure 8: Modelled percent change in access to employment induced by various study scenarios across different demographic groups.

Task 8: Compute equity of access to opportunities

Work through the Jupyter notebook entitled `Access to Opportunities.ipynb`. You should be able to reproduce Figures 6 and 8, and the data required to produce the map in Figure 7.

5 Travel Time

Travel time savings allow us to learn how individuals might benefit generally from travel anywhere in the city. If those travel time savings are demand-weighted, we can have a more complete picture of this benefit that is sensitive to *where* people travel within the city. Unlike access to opportunities, however, we do not need to have data on the locations of opportunities nor make assumptions about *why* people are travelling to a specific destination (though most transportation models have those assumptions coded into the model behaviour).

Our goal here is to find the difference in demand-weighted travel time between our three scenarios and the baseline business as usual case. A difference in two travel time and demand matrices can be calculated with the following formula:

$$\Delta_i = \frac{\sum_{i,j} \tau^1_{i,j} t^1_{i,j}}{\sum_{i,j} \tau^1_{i,j}} - \frac{\sum_{i,j} \tau^0_{i,j} t^0_{i,j}}{\sum_{i,j} \tau^0_{i,j}} \tag{16}$$

Where $\tau_{i,j}$ represents the demand between an origin i and destination j, and $t_{i,j}$ represents the travel time between the origin-destination pair. Scenario '0' is our baseline scenario, and Scenario '1' includes the addition of an intervention or infrastructure project. The steps to compute this are as follows:

1. Multiply the travel times $t_{i,j}$ by the demand $\tau_{i,j}$ for each scenario
2. Sum this product as well as the demand value $t_{i,j}$ for each destination from a given origin i.
3. Divide the sum of the product by the sum of the demand to calculate weighted travel times
4. Repeat for a second scenario
5. Subtract the weighted travel times for an intervention scenario from the "do nothing" scenario to calculate Δ_i

Once this metric is generated, we can use the same process as Section 4.4.

Task 9: Compute equity of travel time savings

Work through the Jupyter notebook entitled `Travel Time.ipynb`. You should be able to reproduce Figure 9 and the data required for the maps in Figures 10 and 11.

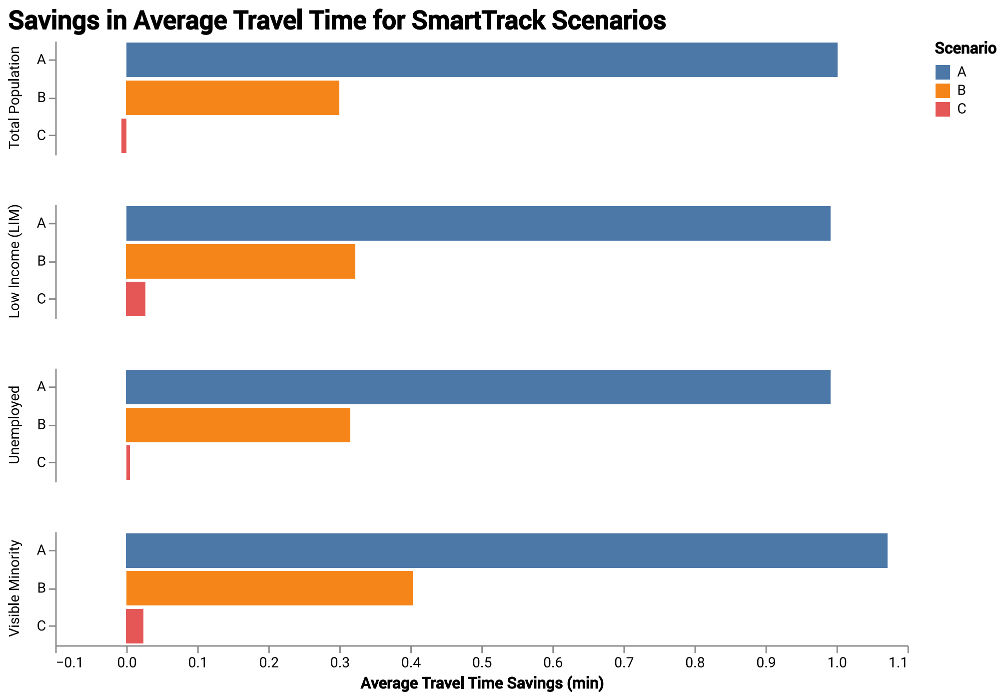


Figure 9: Demand-weighted average travel time savings on transit within Toronto as a result of various SmartTrack scenarios across various socioeconomic groups. groups.

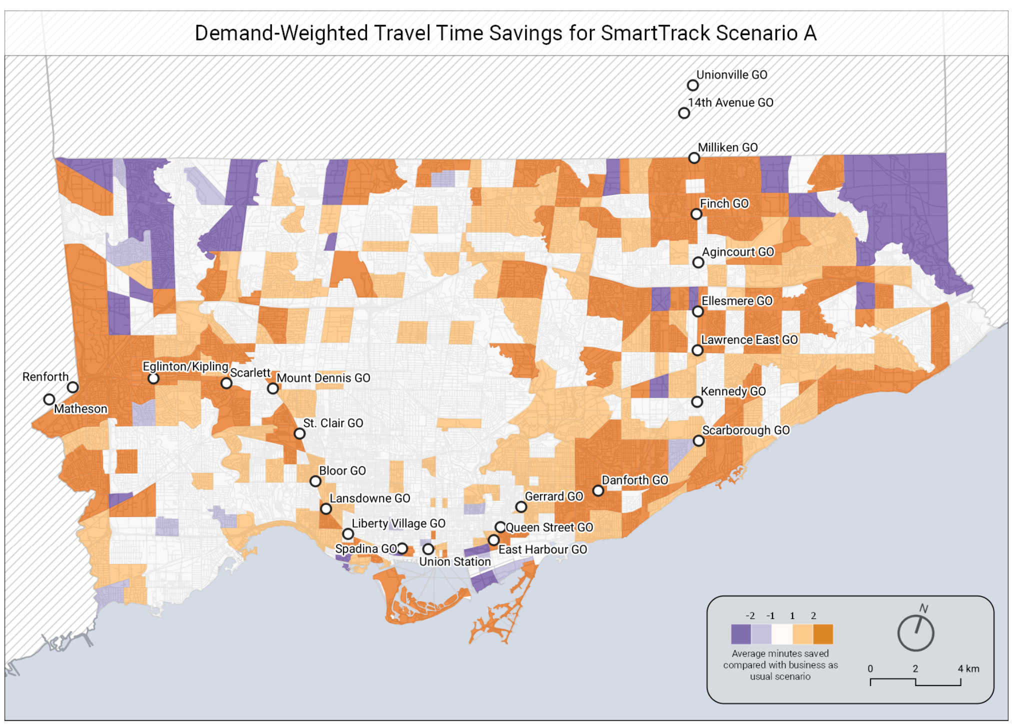


Figure 10: Flow-weighted travel time differences between Scenario A and "Business as Usual" scenarios for our study area model.

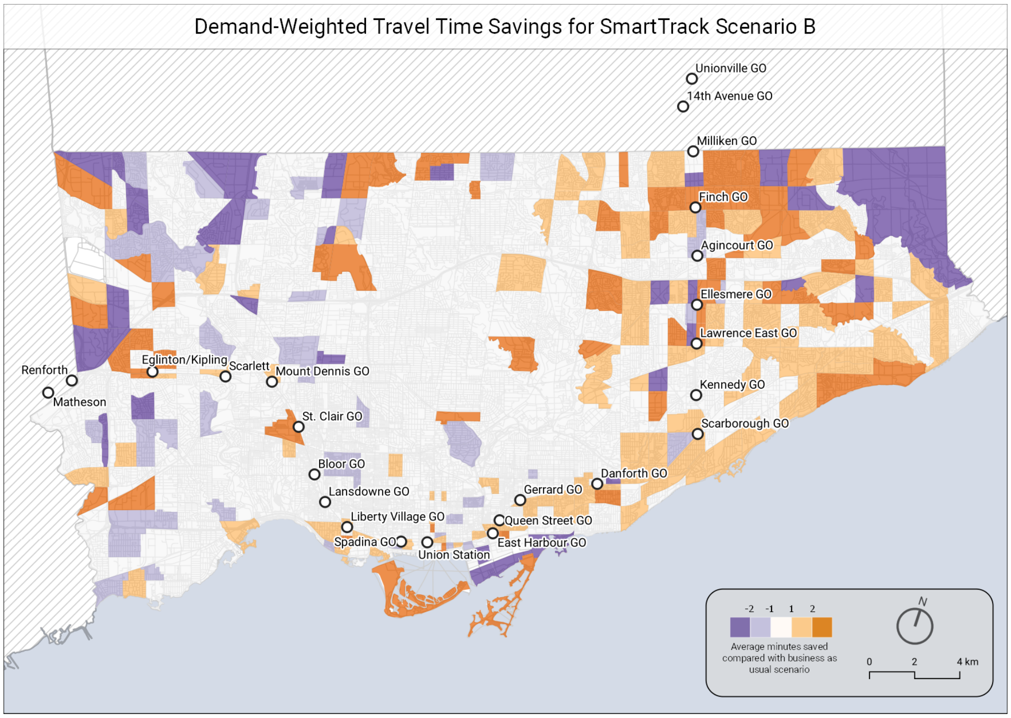


Figure 11: Flow-weighted travel time differences between Scenario B and "Business as Usual" scenarios for our study area model.

6 Utility Gains through Induced Trips

Most transportation models model mode switching using an econometric calculation of utility. In these cases, a modelled traveller may switch modes in a given scenario because that mode now offers additional utility compared to others. In this case, an increase in transit mode share implies an increased benefit to a traveller as this causes a shift in behaviour. In this analysis we will consider the extra transit trips induced by the study project as a proxy for implied utility gains compared with a business as usual case.

We make the calculation as follows:

1. For each origin zone i, determine difference in the number of transit trips originating from each zone.

2. Multiply the difference in trips by the number of individuals in a population subset in a zone i

3. Sum the resulting values over all zones i

4. Divide the total person-trips by the sum total of the population of interest in each zones.

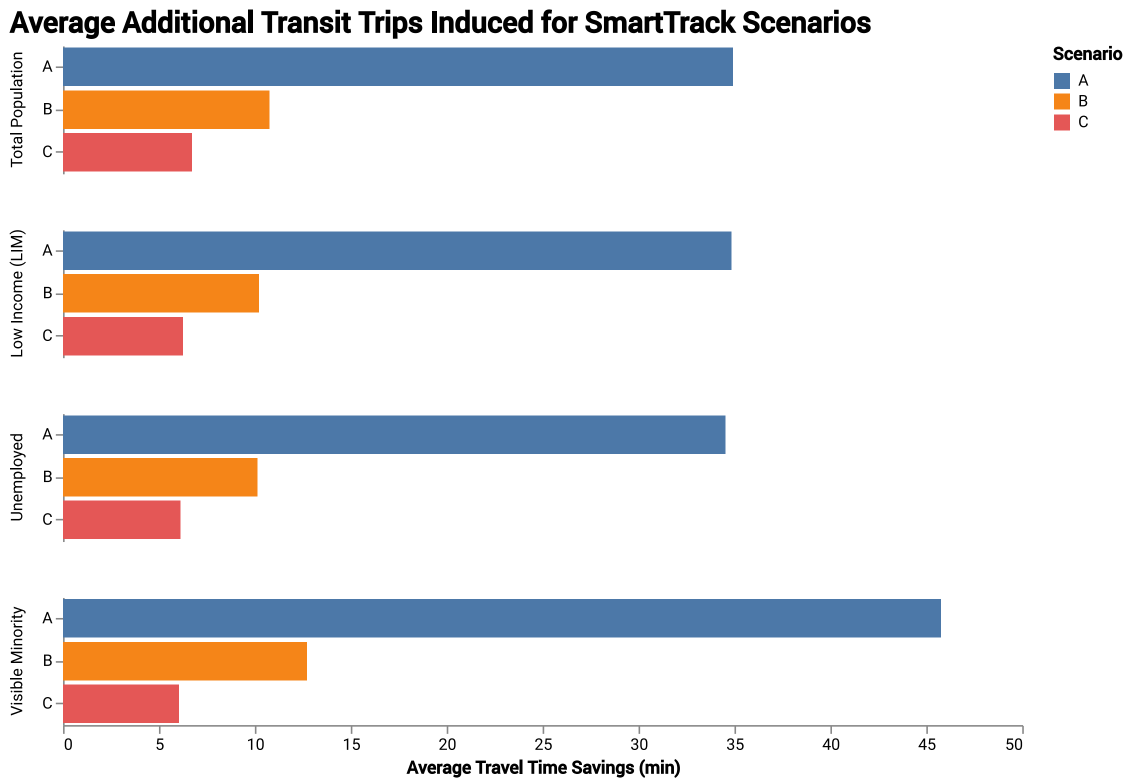


Figure 12: Distribution of implied utility gains as induced transit trips across various demographic groups for three scenarios.

Task 10: Compute equity of travel time savings

Work through the Jupyter notebook entitled `Transit Trips.ipynb`. You should be able to reproduce Figure 12.

7 Population Dynamics

The analyses conducted thus far are based on a static snapshot in time, generally using future predictions of demand and travel time, and current or historical population and demographic data. In reality, the demographic topography of cities change over time, and this can affect many of the distributive equity measures calculated above.

This section introduces a simple simulation technique to model potential socioeconomic changes over time that may be induced by large transit infrastructure projects. The goal of these models and scenarios is not realism. Instead, they are intended as a sensitivity analysis to better understand how larger trends in demographic changes might affect the studied access to opportunity metrics above.

Each scenario in this section follows a similar process: In each iteration, a number of individuals are swapped between two zones according to the criteria of a specific scenario. The swapping process is as follows:

1. Divide all the zones in the study area into two sets A and B based on the criteria of the specific scenario outlined below.

2. For each of the 1000 swap actions in an iteration:

   2.1. Choose at random a zone a from A and a zone b from B

   2.2. Subtract one from the number of low-income individuals in a

   2.3. Add one to the number of low-income individuals in b

At the end of each iteration, two values are calculated. The access scores for low-income individuals[10] are calculated as A_L, as well as the access measures for everyone else A_L^C using the process outlined in Section 4.3. We can calculate the access ratio as:

$$R_A = \frac{A_L}{A_L^C} \tag{17}$$

This value can be plotted for each iteration to see how the ratio of access for low-income individuals compares with everyone else. Values of R_A less than one indicate poorer access for low-income individuals compared to everyone else, and values of R_A greater than one indicate better access for low-income individuals compared to everyone else.

While there are many possible scenarios to model, we are considering the following three:

- **Income polarization**: Based on the current distribution of low-income households, low-income areas continue to get poorer, wealthier areas experience lower poverty rates. In this case, group A zones are those with lower than average concentrations of low-income households, and B zones are those with higher than average concentrations of low-income households.

- **Gentrification around transit**: Areas with higher transit access become more wealthy, areas with low access see larger concentrations of low-income households. In this case, group A zones are those with higher than average access values, and group B zones are those with lower than average access values.

- **Affordability around transit**: Areas with higher transit access see higher concentrations of low-income households (opposite of gentrification scenario). In this case, group A zones are those with lower than average access values, and group B zones are those with higher than average access values.

Task 11: Compute population dynamics simulation

Work through the Jupyter notebook entitled `Population Dynamics.ipynb`. You should be able to reproduce Figure 13.

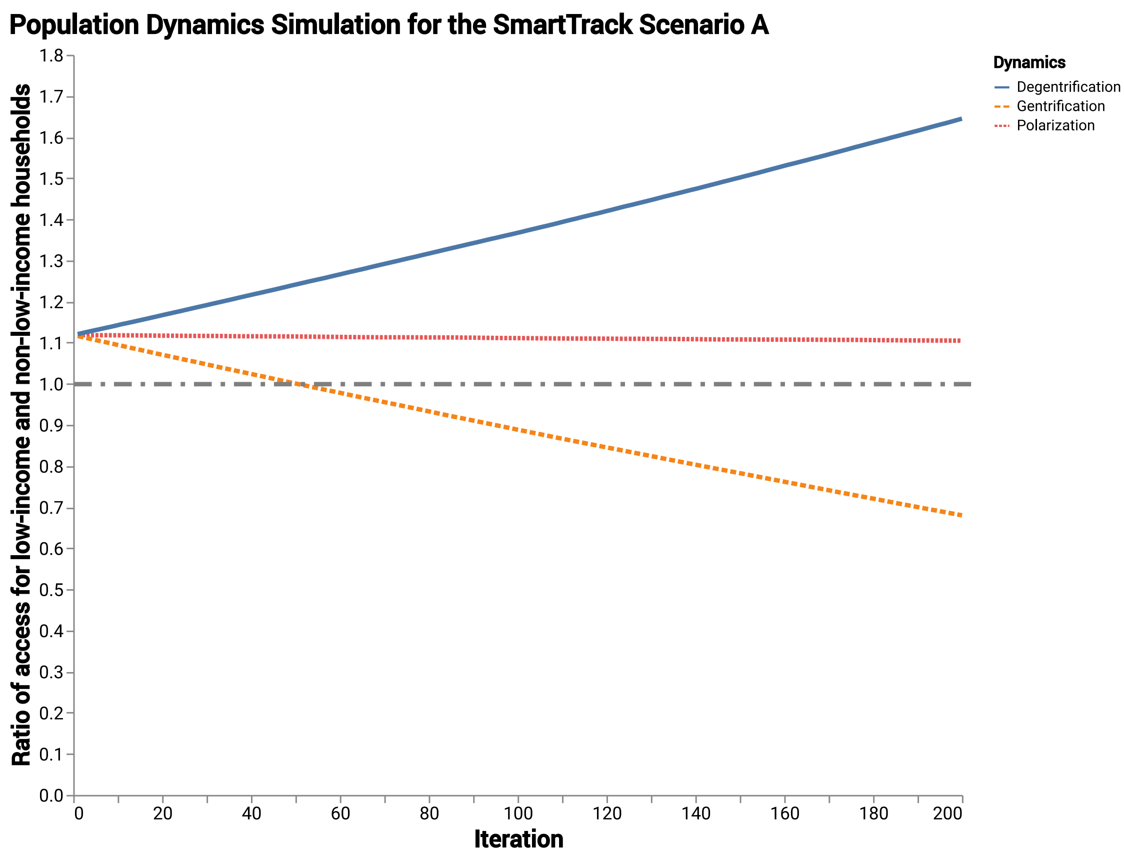


Figure 13: Population dynamics simulation for Scenario A of our study project. Three population dynamics scenarios outlined above for 200 iterations of 1,000 swaps each.

Notes

[1]Jeff Allen and Steven Farber (2021). "Equitable Transit Futures: The Distribution of Benefits of a Proposed Rapid Transit Line in Toronto". In: *100th Annual Meeting of the Transportation Research Board* January. URL: `https://www.researchgate.net/publication/352296205_Equitable_Transit_Futures_The_Distribution_of_Benefits_of_a_Proposed_Rapid_Transit_Line_in_Toronto`.

[2]The census provides subtotal counts for various categories. For example, there is a different count for the total number of individuals who responded to a "visible minority" question versus the total population count in a given area. It is important to use the counts from a particular question as the population total.

[3]Travel times can be real estimates, or "percieved" travel time which better capture the time as experienced by the user. The analysis can be conducted using any set of travel time matrices.

[4]Sourcing DA-level census data at scale can be somewhat tricky. One option is to use the StatCan Python library at `https://stats-can.readthedocs.io/en/latest/`. Some institutions have portals to query specific data categories.

[5]Depending on whether you are interested in individual station-level access or line-level access, you may need to perform a *dissolve* first to create a single shape out of overlapping buffers.

[6]Typically, $I(\cdot)$ is in the range $0 \leq I(\cdot) \leq 1$, but this is not strictly required.

[7]Cumulative measures can be origin or destination oriented, e.g. "what is the total number of jobs reachable from an origin in 45 minutes on transit?" or "how many people live within 30 minutes travel from this destination?".

[8]Mei-Po Kwan (2010). "Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework". en. In: *Geographical Analysis* 30.3, pp. 191–216. (Visited on 11/18/2021).

[9]You may need to first calculate the number of people in a population group within a a given traffic analysis zone based on the areal apportionment performed in Section 2.3.

[10]In the handbook, individuals falling below the low income cutoff are used. This can be adjusted for other low-income criteria