
TTS2016R: A data set to study population and employment patterns from the 2016 Transportation Tomorrow Survey (TTS) in the Greater Toronto and Hamilton Area in Canada

Journal Title
XX(X):1–15
©The Author(s) 0000
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

Anastasia Soukhov, Antonio Paez

Abstract

This paper describes and visualises the data contained within the {TTS2016R} data-package created in R, the statistical computing and graphics language. In addition to a synthetic example, {TTS2016R} contains home-to-work commute information for the Greater Golden Horseshoe area in Canada retrieved from the 2016 Transportation Tomorrow Survey (TTS). Included are all Traffic Analysis Zones (TAZ), the number of people who are employed full-time per TAZ, the number of jobs per TAZ, and the origin and destination trips. {TTS2016R} also includes the calculated car travel time from TAZ origin-destination centroid pairs. This information is useful to understand patterns of commuting in the region. We illustrate the estimation of a distance-decay curve using these data. {TTS2016R} can be freely downloaded and explored at: <https://github.com/soukhova/TTS2016R> where the documentation and code involved in data creation, manipulation, and the final data products are detailed.

Keywords

Jobs; population; travel time; impedance; Greater Toronto and Hamilton Area; Ontario, Canada; R

Introduction

This manuscript presents the open data product `{TTS2016R}`. Open data products are the result of turning source data (open or otherwise) into accessible information that adds value to the original inputs (see Arribas-Bel et al., 2021). The product presented in this paper is an R data package which currently consists of three objects which are sourced from the 2016 Transportation Tomorrow Survey (TTS) or are curated to facilitate the use and analysis of TTS data. This package includes person-to-jobs origin-destinations, traffic analysis zone (TAZ) boundaries, and planning/municipality boundaries for the Greater Golden Horse area (GGH) located in southern Ontario, Canada (Data Management Group, 2018). In addition, the package includes TAZ centroid-to-centroid travel times by car computed using package `{r5r}` (Pereira et al., 2021). The aim of this paper is to walk readers through the empirical home-based work commute data set, illustrate the calculation of an impedance function, showcase how it can be used to calculate accessibility to employment, and invite its use in other applications.

Data from the TTS are in principle available to the public but are not fully open, since permission to access the data retrieval system is required. In addition, the raw data can be technically demanding, cumbersome to work with, and requires multiple software to process. By pre-processing the data in a R environment, `{TTS2016R}` offers a slice of the TTS data useful to understand patterns of commuting to work in the region. It also provides open infrastructure for additional TTS or complimentary data sets to be amended by the authors or wider open-source community in the future.

Home-to-work commute data

`{TTS2016R}` includes counts of fully-employed population by place of residence and counts of full-time jobs by place of work aggregated by TAZ (n=3,764 within the survey boundaries). Unique identifiers use the GTA06 Zoning System of TTS. The number of jobs (3,081,900) and workers (3,446,957) are organized in the form of an origin-destination table, indicative of patterns of home-to-work commute (there are 3,446,957 potential interactions). These data were retrieved from the Transportation Tomorrow Survey Data Retrieval System on October 28, 2021 and reflect the potential interaction of full-time employed people and jobs within the GGH survey boundaries shown in Figure 1 as defined by the 2016 TTS methodology (Data Management Group, 2018).

Also included in `{TTS2016R}` are travel times and cost of travel from origin to destination by car; travel times are calculated using the R package `{r5r}`. These travel times are useful to estimate the cost of travel and to calculate impedance functions, among other possible uses. It is important to note that for simplicity, all interactions within `{TTS2016R}` are assumed to be taken by car, and the travel time is calculated from an origin TAZ centroid to a destination TAZ centroid. The centroid is snapped to the nearest street line by `r5r` and the travel time is calculated for all trips assuming a car travel mode. Additionally, only travel times less than or equal to 180 mins (3 hrs) are

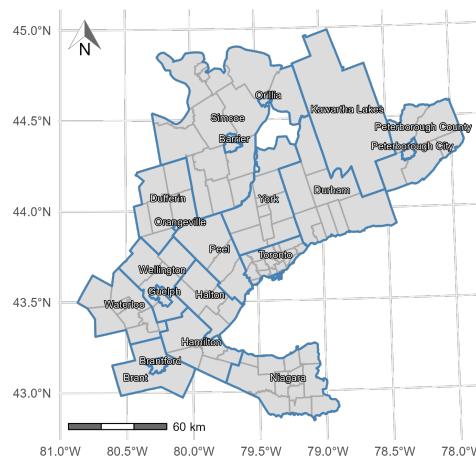


Figure 1. The TTS 2016 study area within the Greater Golden Horseshoe in Ontario, Canada.

Table 1. Descriptive statistics of the trips, workers, and jobs for the traffic analysis zones (TAZ) from the TTS 2016 dataset along with estimated car origin-destination travel times.

	Trips	Car Travel Time	Area	Workers	Jobs
	(#)	(min)	(km ²)	(#)	(#)
Min.	1	0	0	0	0
1st Qu.	14	13	1	25	64
Median	22	20	1	464	244
Mean	33	23	7	916	819
3rd Qu.	38	30	3	1378	700
Max.	1129	179	879	8491	41821
NA's	NA	3507	NA	NA	NA

calculated; this threshold represents 99% of trip's travel times which are summarized in the descriptive statistics in Table 1.

Employed individuals and jobs

The origin-destination information consists of a cross-tabulation of people who are employed full-time by place of residence (origin) and places of employment in the GGH (destination) using the GTA06 zoning system. The number of workers and jobs is not equal; the boundaries of the survey are permeable, so workers who reside within the boundaries but travel outside of the boundaries are counted as workers within an origin TAZ, while jobs in TAZ that are filled by workers who reside outside the GGH boundaries are *unknown* since they were not surveyed. This mismatch results in the total number

of workers being 1.12 times larger than the number of jobs (i.e., 3,446,957 workers to 3,081,900 jobs). While the 2016 TTS survey boundaries are drawn to minimize the difference between opportunities at the destination and supply at the origins they are still not equal . That said, these data offer a perspective of the potential of home-based trips to places of employment and as such, the number of trips taken are equal to the number of workers in the GGH.

Figure 2 presents the number of workers and jobs per TAZ. It can be observed that the spatial distribution of jobs and workers is unequal, which is indicative of jobs-housing imbalance, which can impact accessibility in a region (Levine, 1998). Workers are concentrated in TAZ within the center of the GGH and along the south-east and northern border of the GGH. The center of the GGH corresponds to the Greater Toronto Area (GTA) which is the most densely populated area in southern Ontario (Statistics Canada, 2022). The south east border of the GGH neighbours Lake Ontario and is delineated by the urban built boundary of the Ontario Growth Plan being home to the highest density of working population in the GGH (Ontario, 2019; Auditor General of Ontario, 2021). The northern GGH border corresponds to the Simcoe, Dufferin, Kawartha Lake, and Peterborough regions which are home to lower density of worker population density population (Auditor General of Ontario, 2021). Conversely, the spread of jobs in the GGH is lower than the number of workers indicating population is more spatially dispersed than jobs.

It can also be seen that from the bottom plot in Figure 2 that high to medium-low concentrations of jobs are often present in the same areas as workers but only when the scale is transformed. In other words, though there is a higher number of TAZ with no workers than zones with no jobs (i.e., 791 TAZ with no workers : 396 TAZ with no jobs) and the mean of workers per TAZ is higher than the mean of jobs (i.e., 916 workers : 819 jobs) the number TAZ with an extreme number of jobs at the highest and lowest percentiles is significantly higher than the number of workers; see the following cumulative probability distribution in Figure 3 in which the 94th to 100th percentile and the 0th to 34th percentile of jobs in TAZ is higher than the number of workers in TAZ. This means that between these ranges, TAZ have a higher number of workers than they do jobs, echoing the more even spatial distribution of workers observed in Figure 2.



Figure 2. Number of workers (top) and jobs (bottom) in each TAZ in the GGH area as specified in the 2016 TTS.

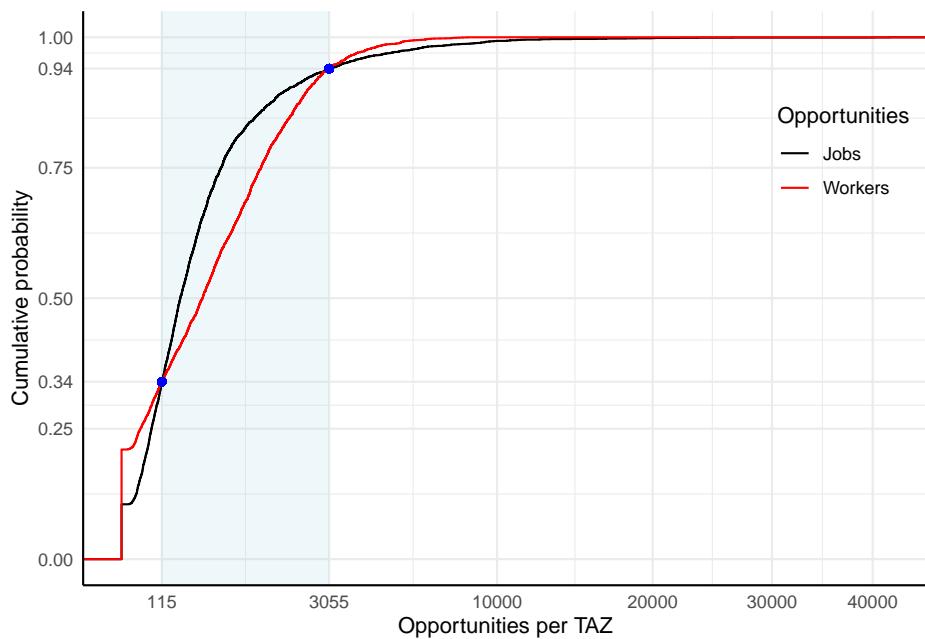


Figure 3. The cumulative distribution of the number of jobs and workers per TAZ from the 2016 TTS data set. Light blue shaded ranges correspond to all cumulative probabilities where the number of workers per TAZ are larger than jobs per TAZ.

Calculated travel time

{TTS2016R} also includes travel time data for each home-to-work trip as displayed in Figure 4. This travel time corresponds to a car commute calculated using the R package {r5r} and is interpreted as the travel time for a work commute for full-time employed people in the GGH. The travel times were calculated assuming the following input parameters: a maximum travel time less than or equal to 180 mins (3 hrs) and a street network retrieved from OpenStreetMaps for the GGH area. The 3 hr threshold was selected as it captures 99% of the trips taken (see the travel times summarized in Table 1).

It is important to note that all travel times within this data set are calculated assuming car travel and one departure time for all origins. These assumptions are not completely realistic since we know a small proportion of trips are taken by non-car modes and travel time departures varies, however, it is not possible from the data retrieval system to obtain higher order tabulations that would allow us to split the data further. Thus we do not know which trips are made with non-car modes nor exact departure times in these tables. Though modal split and travel times can be estimated through other methods (e.g., Allen and Farber, 2021; Higgins et al., 2021), for simplicity, we assume that all trips are taken by one-time departure car trip.

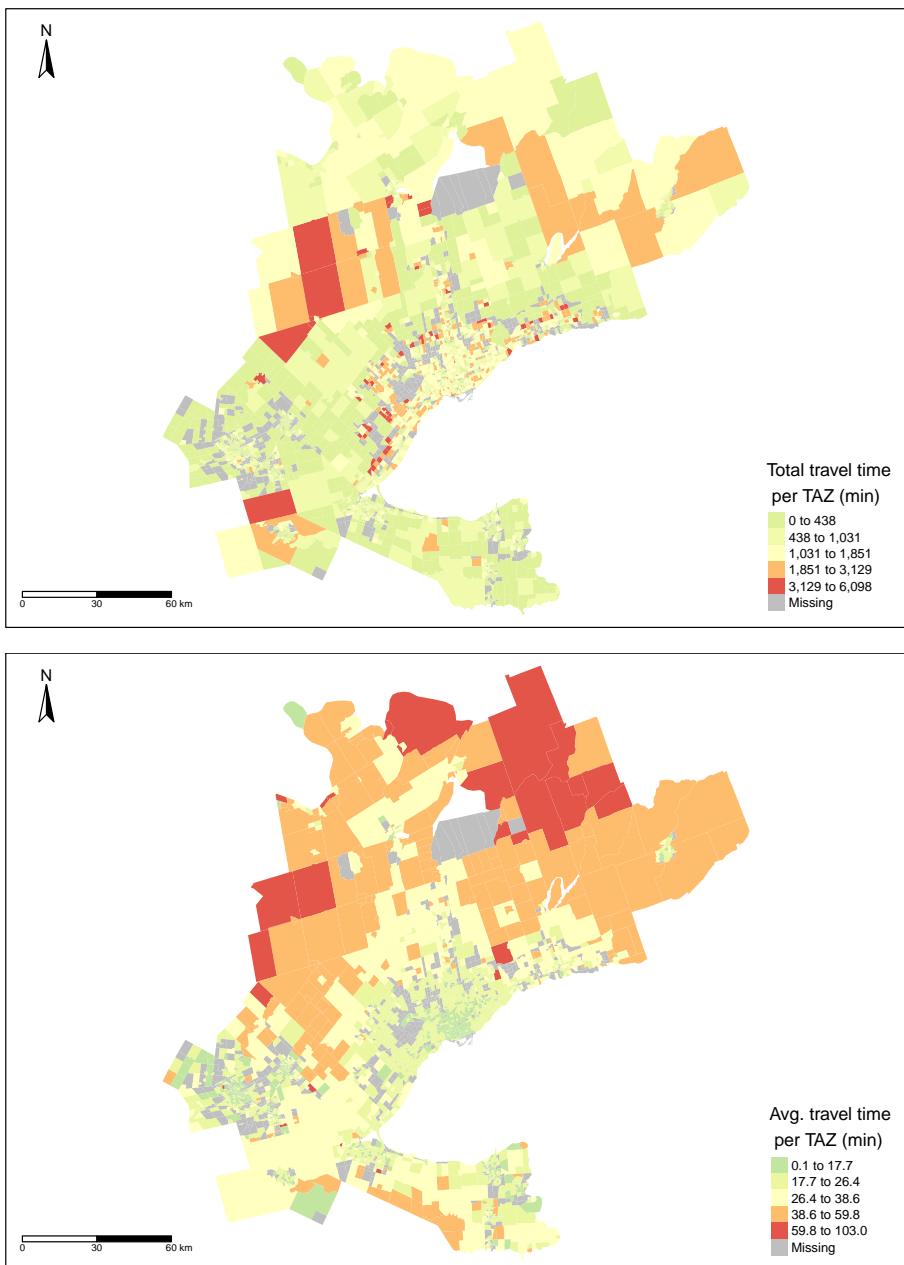


Figure 4. Estimated total worker travel time (top) and average worker travel time (bottom) for each TAZ in the GGH.

As can be observed in Figure 4, the total travel time (min) resembles the spatial trend distribution in the number of employed people in the previous plot (Figure 2). However, when the average travel time per trip in each TAZ is presented, the spatial distribution is distinct from all other plots presented so far. We can see that in areas around the south-eastern border that make up the Greater Toronto and Hamilton Area (GTHA) (e.g., Hamilton, Halton, Peel, Toronto, York, Durham) and Niagara and Waterloo, the average travel times are moderately low. Further from these areas, travel times are higher. Interestingly, even in eastern areas (e.g., Peterborough) with high employment and high job concentration, average travel time is higher than within the GTHA.

Calibrating an impedance function

Impedance functions are useful to understand mobility behavior and are part, for instance, of gravity models of spatial interaction (Wilson, 1971; Haynes and Fotheringham, 1985) and accessibility analysis in many applications (Hansen, 1959; Levinson, 1998; Talen and Anselin, 1998; Reggiani et al., 2011; Páez et al., 2013; Kuai and Zhao, 2017; Barboza et al., 2021). An origin destination matrix and a cost matrix (i.e., with travel times) can be used in combination to estimate impedance functions. An impedance function $f(\cdot)$ depends on the cost of travel between locations i and j c_{ij} ; it usually is a monotonically decreasing function, although sometimes the function can increase to reflect patterns of separation between activities; for instance, separation of land uses means that very short commuting trips are relatively rare. There can be sometimes also fluctuations due to hierarchical patterns, where travelers bypass opportunities in favor of more distant destinations that offer economies of agglomeration.

A useful technique to calibrate an impedance function is to use the trip length distribution (TLD) as measured from origin-destination data (Horbachov and Sivchynskyi, 2018; Batista et al., 2019). The TLD is the representation of the likelihood that a proportion of trips are taken at a specific travel cost. In our data set, where we assume cost is travel time, the impedance function maps low travel times to higher proportions of trips, and high travel times are mapped to low proportion of trips.

In the GGH data presented, the empirical TLD (i.e., proportion of trips taken vs. travel time in minutes) is fitted to a density distribution using maximum likelihood techniques and the Nelder-Mead method for direct optimization available within the `fitdistrplus` package (Delignette-Muller and Dutang, 2015). Based on goodness-of-fit criteria and diagnostics seen in Figure 5 and Figure 6, the gamma distribution is selected for the presented data.

The resulting calibrated impedance function is given in the following general form where the estimated ‘shape’ is $\alpha = 2.019$, the estimated ‘rate’ is $\beta = 0.094$, and $\Gamma(\alpha)$ is defined in Equation (1).

$$f(x, \alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} \quad \text{for } 0 \leq x \leq \infty \quad (1)$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

summary statistics

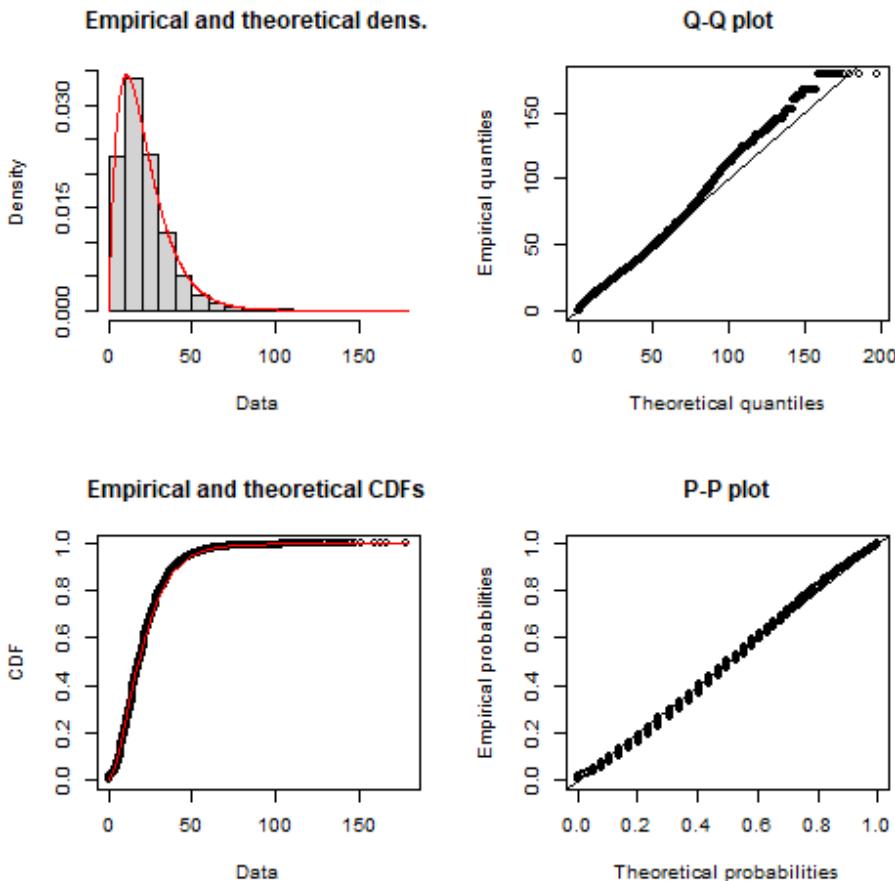


Figure 5. Empirical TTS 2016 home-based car trip length distribution (black) and calibrated gamma distribution impedance function (red) with associated Q-Q and P-P plots

```
-----
min: 0.1    max: 179
median: 18
mean: 21.4344
estimated sd: 14.61254
estimated skewness: 1.703326
estimated kurtosis: 8.353363
```

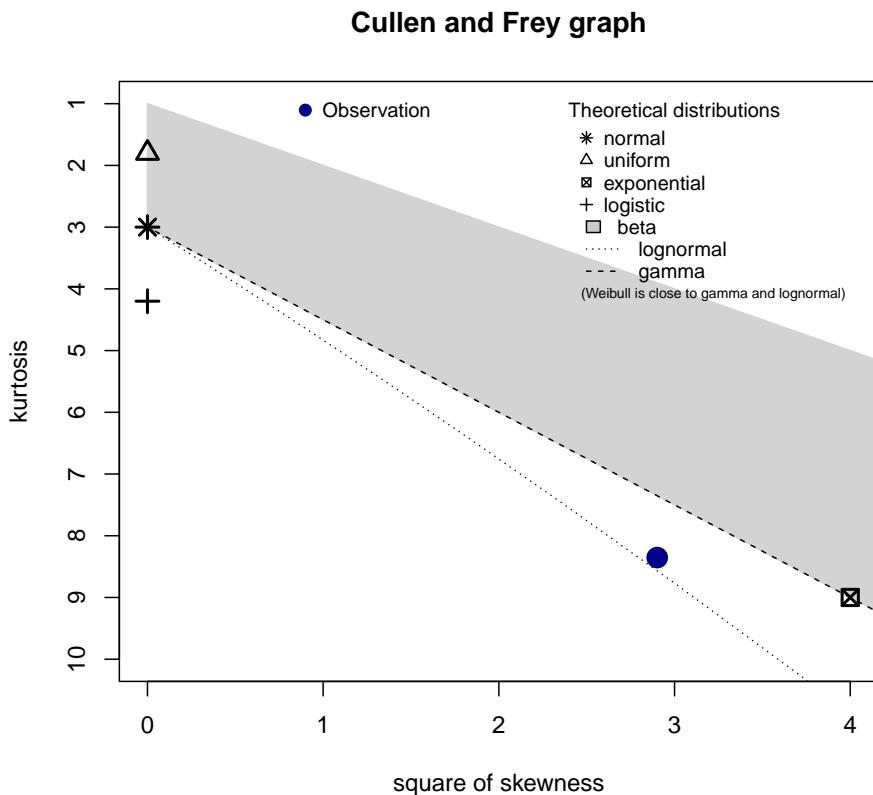


Figure 6. Cullen and frey graph for the 2016 TTS calculated travel times.

Accessibility to employment

As noted above, impedance functions are an essential component of accessibility analysis. Equation (2) shows that accessibility A_i is the weighted sum of opportunities that can be reached from location i , given the cost of travel c_{ij} . Summing the opportunities in the neighborhood of i as defined by the impedance function $f(\cdot)$, provides estimates of the number of opportunities that can be reached from i at a certain cost.

$$A_i = \sum_{j=1}^J O_j f(c_{ij}) \quad (2)$$

where:

- A is accessibility.
- i is a set of origin locations.
- j is a set of destination locations.
- O_j is the number of opportunities at location j . These are opportunities for activity and add some sort of *supply* to the area;
- c_{ij} is a measure of the cost of moving between i and j
- $f(\cdot)$ is an impedance function of c_{ij} .

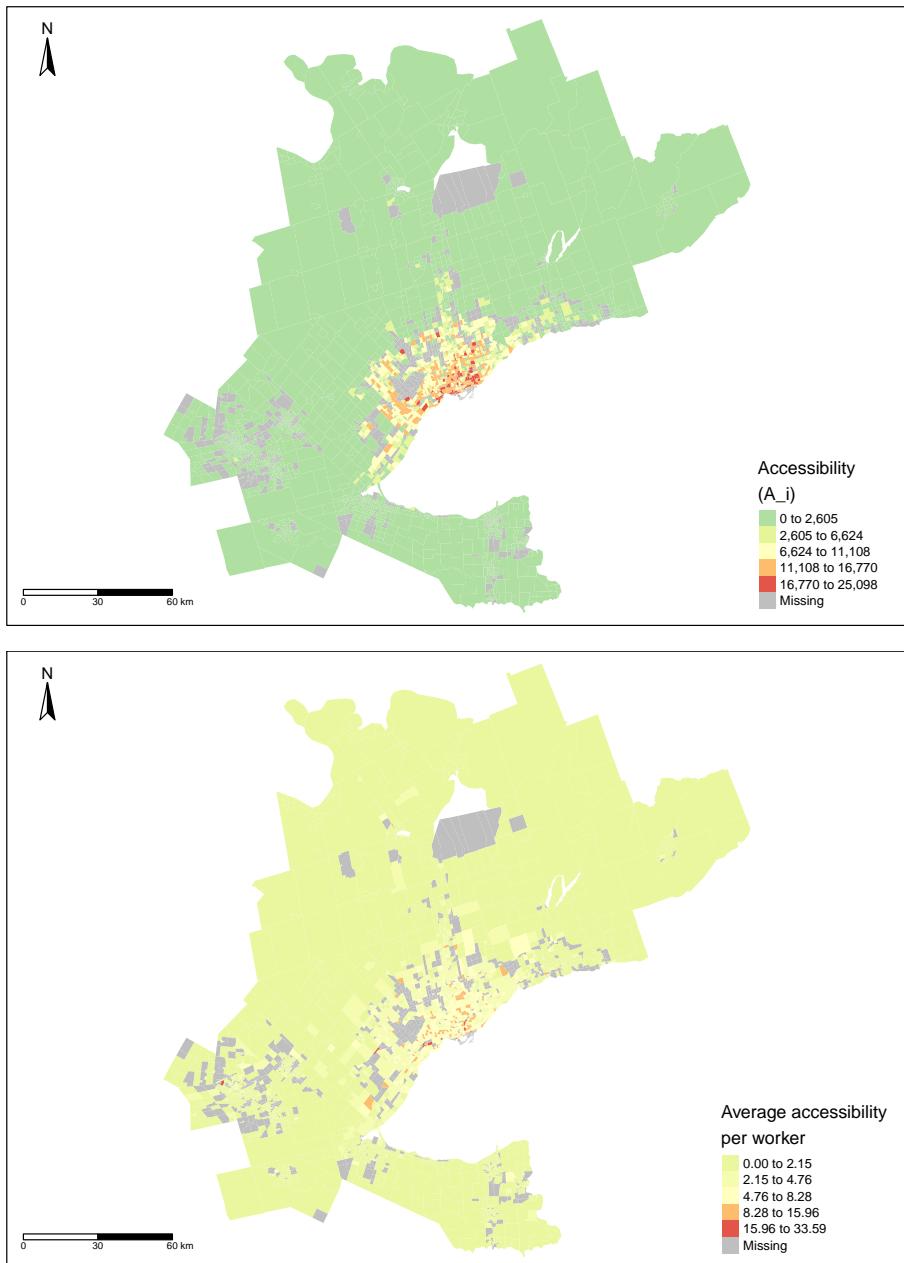


Figure 7. Accessibility (top) and normalized accessibility (bottom) to employment in the GGH area

Accessibility is a property of the origin (i.e., origin TAZ in our data), the landscape of opportunities (i.e., the TAZ destinations in our data), and transportation infrastructure (i.e., the travel time table). Figure 7 presents the accessibility estimates that result from the impedance function calibrated in the preceding section. The top plot, that of the raw accessibility values, presents a distinct radial trend where the majority of TAZ in and around Toronto have high accessibility values and values gradually decrease in TAZ which are further from Toronto's boundary. Alternatively, the bottom plot in Figure 7 present the worker-normalized accessibility values. In other words, the accessibility value for each TAZ is divided by the number of workers in each TAZ. The patterns in this plot are similar to the Greater Toronto Area radial trend observed in the top plot but are less extreme (i.e., smaller spread).

Concluding remarks

The open data product introduced in this paper shares tables for worker-to-employment data from the 2016 TTS aggregated by TAZ. In addition, inter-centroid travel time tables are calculated, and the planning/municipality boundaries are included to compliment the 2016 TTS data. This open data product, {TTS2016R}, is freely available to explore in an R environment. One possible use of these data, as showcased in this paper, is the calibration of impedance functions which in turn can be used for accessibility analysis.

New digital formats are increasingly complex and the explanation of the methods often do not concisely and intelligibly fit within the confines of a traditional article. With this motivation, we invite all who are interested to use {TTS2016R} to explore the worker-employment patterns contained in the {TTS2016R} package. In the spirit of novel and original research, we hope readers value the efforts made to detail the data in order to improve transparency in our work and encourage others to replicate and, hopefully, inspire research of their own. We see this product as providing open infrastructure for additional TTS or complimentary data sets to be amended by the authors or wider open-source community in the future.

References

- Allen J and Farber S (2021) Suburbanization of Transport Poverty. *Annals of the American Association of Geographers* 111(6): 18.
- 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. URL <https://dx.doi.org/10.1007/s10109-021-00363-5>.
- Auditor General of Ontario (2021) Value for Money Audit: Land-Use Planning in the Greater Golden Horseshoe. Technical report. URL https://www.auditor.on.ca/en/content/annualreports/arreports/en21/AR_LandUse_en21.pdf.
- Barboza MHC, Carneiro MS, Falavigna C, Luz G and Orrico R (2021) Balancing time: Using a new accessibility measure in Rio de Janeiro. *Journal of Transport Geography* 90: 102924. DOI:10.1016/j.jtrangeo.2020.102924. URL <https://www.sciencedirect.com/science/article/pii/S0966692320310012>.
- Batista S, Leclercq L and Geroliminis N (2019) Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. *Transportation Research Part B: Methodological* 122: 192–217. DOI:10.1016/j.trb.2019.02.009. URL <https://linkinghub.elsevier.com/retrieve/pii/S0191261518311603>.
- Data Management Group (2018) TTS - Transportation Tomorrow Survey 2016. URL <http://dmg.utoronto.ca/transportation-tomorrow-survey/tts-introduction>.
- Delignette-Muller ML and Dutang C (2015) fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software* 64(4): 1–34. URL <https://www.jstatsoft.org/article/view/v064i04>.
- Hansen WG (1959) How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* 25(2): 73–76. DOI:10.1080/01944365908978307. URL <http://www.tandfonline.com/doi/abs/10.1080/01944365908978307>.
- Haynes KE and Fotheringham AS (1985) *Gravity and Spatial Interaction Models*. Reprint. WVU Research Repository. URL <https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=1010&context=rri-web-book>.
- Higgins CD, Páez A, Ki 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* : 114442 DOI:10.1016/j.socscimed.2021.114442. URL <https://dx.doi.org/10.1016/j.socscimed.2021.114442>.
- Horbachov P and Svichynskyi S (2018) Theoretical substantiation of trip length distribution for home-based work trips in urban transit systems 11(1): 593–632. URL <https://www.jstor.org/stable/26622420>. Publisher: Journal of Transport and Land Use.
- Kuai X and Zhao Q (2017) Examining healthy food accessibility and disparity in Baton Rouge, Louisiana. *Annals of GIS* 23(2): 103–116. DOI:10.1080/19475683.2017.1304448. URL <https://www.tandfonline.com/doi/full/10.1080/19475683.2017.1304448>.

- Levine J (1998) Rethinking accessibility and jobs-housing balance. *Journal of the American Planning Association* 64(2): 133–149. URL [ISI:000073499600007](https://doi.org/10.1080/000073499600007). JSPR.
- Levinson DM (1998) Accessibility and the journey to work. *Journal of Transport Geography* 6(1): 11–21. DOI:10.1016/S0966-6923(97)00036-7. URL <https://linkinghub.elsevier.com/retrieve/pii/S0966692397000367>.
- Ontario (2019) Built Boundary for the Growth Plan for the Greater Golden Horseshoe, 2006. URL <https://www.ontario.ca/document/built-boundary-growth-plan-greater-golden-horseshoe-2006-built-boundary-maps>.
- Pereira RHM, Saraiva M, Herszenhut D, Braga CKV and Conway MW (2021) r5r: Rapid realistic routing on multimodal transport networks with r⁵ in r. *Findings* DOI:10.32866/001c.21262.
- Páez A, Farber S, Mercado R, Roorda M and Morency C (2013) Jobs and the Single Parent: An Analysis of Accessibility to Employment in Toronto. *Urban Geography* 34(6): 815–842. DOI: 10.1080/02723638.2013.778600. URL <http://www.tandfonline.com/doi/abs/10.1080/02723638.2013.778600>.
- Reggiani A, Bucci P and Russo G (2011) Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network. *International Regional Science Review* 34(2): 230–252. DOI:10.1177/0160017610387296. URL <http://journals.sagepub.com/doi/10.1177/0160017610387296>.
- Statistics Canada (2022) The Daily — Canada's large urban centres continue to grow and spread. URL <https://www150.statcan.gc.ca/n1/daily-quotidien/220209/dq220209b-eng.htm>. Last Modified: 2022-02-09.
- Talen E and Anselin L (1998) Assessing Spatial Equity: An Evaluation of Measures of Accessibility to Public Playgrounds. *Environment and Planning A: Economy and Space* 30(4): 595–613. DOI:10.1068/a300595. URL <http://journals.sagepub.com/doi/10.1068/a300595>.
- Wilson A (1971) A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space* 3(1): 1–32. DOI:10.1068/a030001. URL <http://dx.doi.org/10.1068/a030001>.