Journal Title
XX(X):1–8

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

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

Anastasia Soukhov, Antonio Páez

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, origin-destination trips, and calculated car travel time from TAZ origin-destination centroid pairs. To illustrate how this information can be analysed to understand patterns in commuting, we estimate a distance-decay curve (i.e., impedance function) for the region. {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 that can be used to calculate accessibility to employment, and invite its use in other applications.

Home-to-work commute data

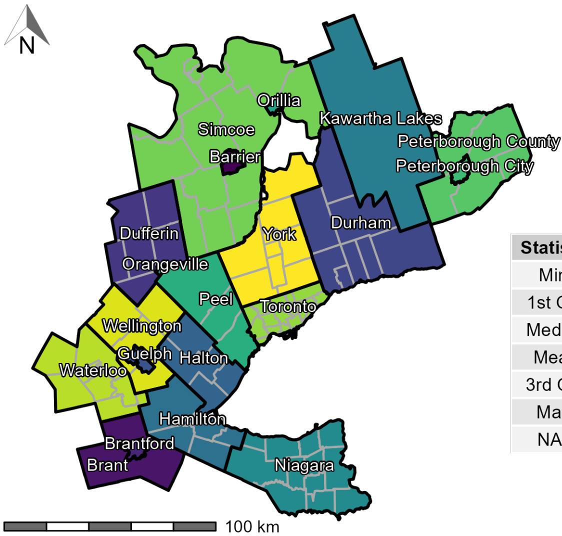
{TTS2016R} includes counts of fully-employed population by place of residence (origin) and counts of full-time jobs by place of work (destination) aggregated by TAZ (n=3,764 within the survey boundaries). TAZ typically are defined based on land-use and population demographics in order to estimate the number of trips produced and attracted to each zone (Meyer and Miller, 2001). As such, each TAZ is uniquely identified using the GTA06 Zoning System which can be used to join to the origin-destination table (i.e., trips taken).

The number of jobs (3,081,885) and workers (3,446,957) in this package are organized in the form of an origin-destination table which is indicative of home-to-work commute patterns (there are 3,446,957 potential interactions). These data were retrieved from the TTS 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 applications. 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 calculated; this threshold represents 99% of trip's travel times which are summarized in Figure 1.

Employed individuals and jobs

The origin-destination table (i.e., trips) consists of a cross-tabulation of people who are employed full-time by place of GGH residence (origin) and places of GGH employment



Statistic	Trips	TT	Workers	Jobs
Min.	1	0	0	0
1st Qu.	14	13	25	64
Median	22	20	464	244
Mean	33	23	916	819
3rd Qu.	38	30	1378	700
Max.	1129	179	8491	41821
NA's	NA	3507	NA	NA

Figure 1. TTS 2016 study area within the GGH in Ontario, Canada along with the descriptive statistics of the trips, calculated origin-destination car travel time (TT), workers per TAZ, and jobs per TAZ. Contains 20 regions (black boundaries) and sub-regions (dark gray boundaries).

(destination) using the GTA06 TAZ zoning system. It is important to note that the total number of workers and jobs is the TTS 2016 region are not equal but the number of trips taken are equal to the number of workers. Since the outer boundaries of the TTS are permeable, 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 3,081,885 jobs). As such, the origin-destination table contained in {TTS2016R} offers a perspective on all workers in the GGH and their home-based trips to places of GGH employment.

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 a jobs-housing imbalance that can impact accessibility in a region (Levine, 1998). It can also be seen that 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

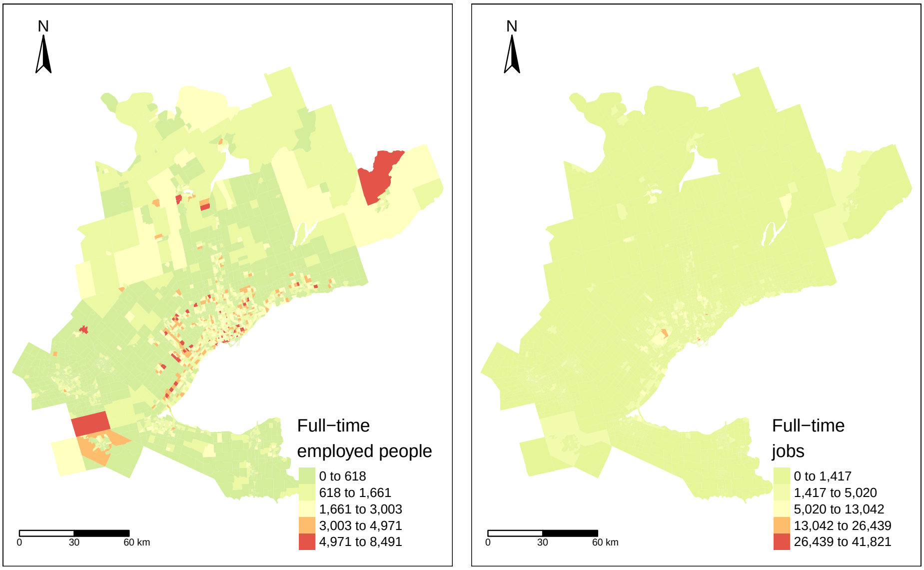


Figure 2. Number of workers (left) and jobs (right) in each TAZ in the 2016 TTS.

of jobs at the highest and lowest percentiles is significantly higher than the number of workers.

Calculated travel time

As mentioned, {TTS2016R} also includes travel time data for each home-to-work trip as displayed in Figure 3. This travel time corresponds to a car commute calculated using the R package {r5r} (see descriptive statistics in Figure 1). It is important to note that travel times within this data set are calculated assuming only 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 so we carry on with the assume that all trips are taken by one-time departure car trip.

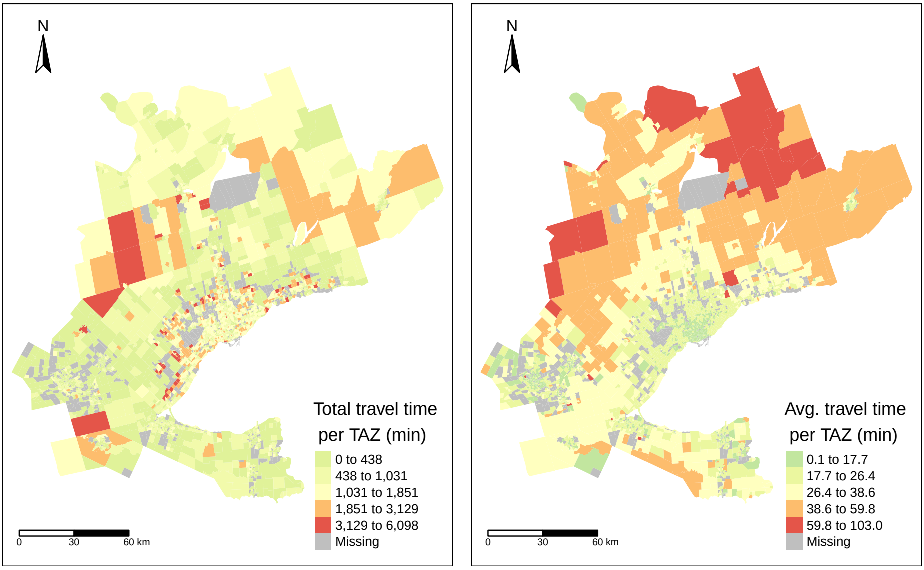


Figure 3. Calculated total worker travel time (left) and average worker travel time (right) for each TAZ in the 2016 TTS.

As can be observed in Figure 3, the total travel time resembles the spatial trend distribution in the number of employed people in the previous plot (Figure 2) and the spatial distribution of the average travel time is distinct from 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), 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 behaviour and are used to estimate gravity models of spatial interaction (Wilson, 1971; Haynes and Fotheringham, 1985) and applied in accessibility analysis (Hansen, 1959; Talen and Anselin, 1998; Páez et al., 2013; Barboza et al., 2021). An impedance function $f(\cdot)$ depends on the cost of travel c_{ij} between locations i and j (all which is supplied in the travel time and origin-destination table within $\{$TTS2016R$\}$).

A useful technique to calibrate an impedance function is to use the trip length distribution (TLD) as measured from origin-destination data (Horbachov and Svichynskyi, 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

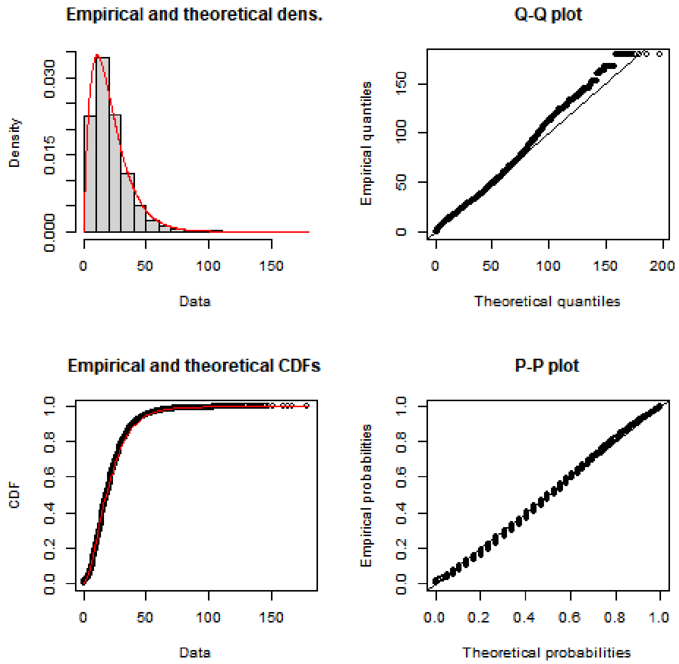


Figure 4. Empirical TTS 2016 home-based car TLD (black) and calibrated gamma distribution impedance function (red) with associated Q-Q and P-P plots

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.

Using the data contained in {TTS2016R}, we fit the empirical TLD to a density distribution using maximum likelihood techniques and the Nelder-Mead method for direct optimization available within the R package {fitdistrplus} (Delignette-Muller and Dutang, 2015). Based on goodness-of-fit criteria and diagnostics seen in Figure 4, the gamma distribution is selected. The 'shape' parameter is $\alpha = 2.019$, the estimated 'rate' is $\beta = 0.094$, and $\Gamma(\alpha)$ is defined in Equation (1).

$$\begin{aligned} f(x, \alpha, \beta) &= \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \quad \text{for } 0 \leq x \leq \infty \\ \Gamma(\alpha) &= \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx \end{aligned} \tag{1}$$

Concluding remarks

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

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

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.

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.

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.

Levine J (1998) Rethinking accessibility and jobs-housing balance. *Journal of the American Planning Association* 64(2): 133–149. URL ISI:000073499600007. JSPR.

Meyer MD and Miller EJ (2001) *Urban transportation planning: a decision-oriented approach*. McGraw-Hill series in transportation, 2nd ed edition. Boston: McGraw-Hill. ISBN 978-0-07-242332-7.

Pereira RHM, Saraiva M, Herszenhut D, Braga CKV and Conway MW (2021) r5r: Rapid realistic routing on multimodal transport networks with r^5 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.

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.