Environment and Planning B:
Urban Analytics and City Science

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

Journal:	*Environment and Planning B: Urban Analytics and City Science*
Manuscript ID	EPB-2022-0164.R1
Manuscript Type:	Urban Data/Code
Keywords:	spatial data, spatial analysis, open data products, Transportation Tomorrow Survey, travel-to-work areas

Journal Title
XX(X):2–9
©The Author(s) 0000
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

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

Anon1, Anon2

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 (GGH) 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, the count of origin destination (OD) pairs and trips by mode per origin TAZ, calculated car travel time from TAZ OD centroid pairs, and associated spatial boundaries to link TAZ to the Canadian Census. 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} is a growing open data product built on R infrastructure that allows for the immediate access of home-to-work commuting data alongside complimentary objects from different sources. The package will continue expanding with additions by the authors and the community at-large by requests in the future. {TTS2016R} can be freely explored and downloaded in the associated Github repository where the documentation and code involved in data creation, manipulation, and all open data products are detailed.

Keywords

Jobs; population; work; commute; travel time; impedance; Greater Toronto and Hamilton Area; Greater Golden Horshoe 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 a R data package which currently consists of a fusion of objects from a variety of sources: home-to-work flows sourced from the 2016 Transportation Tomorrow Survey (TTS) (Data Management Group, 2018b), estimated travel times (calculated using {r5r} (Pereira et al., 2021)), and boundary files from the TTS (Data Management Group, 2018a) and from the Canadian Census (Statistics Canada, 2017).

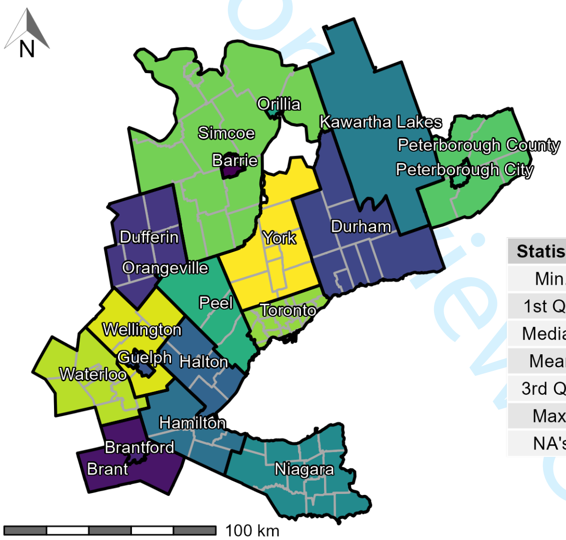
What is a R data package? A R data package contains code, data, and documentation in a standardised collection format that can be installed by R users through a centralized software repository such as CRAN (the Comprehensive R Archive Network) and GitHub. {TTS2016R} is freely available on GitHub for all to install and freely use in the spirit of open and reproducible research. Currently and in more detail, {TTS2016R} includes full-time home-based work-to-job origin destinations (OD) counts and mode-specific trip numbers retrieved from the 2016 TTS, traffic analysis zone (TAZ) boundaries, and municipality, planning, and census metropolitan area boundaries for the Greater Golden Horse area (GGH) located in southern Ontario, Canada. In addition, the package includes TAZ centroid-to-centroid travel times by car, transit, cycling, and walking mode computed using package {r5r} (Pereira et al., 2021).

The aim of this paper is to walk readers through the data sets, illustrate a use case (i.e., the calculation of an impedance function that can be used to calculate accessibility to employment), and invite others to experiment in its uses and applications. Though data from the TTS is freely available to the public through the TTS Data Retrieval System, the raw data can be technically demanding, cumbersome to work with, and requires multiple software to process. By pre-processing the data, packaging it with complimentary data, and providing explicit documentation in a R environment, {TTS2016R} offers a slice of the TTS data that can be immediately used by R users to analysis patterns of commuting to work in the region. Anticipate this package to grow in the future: it currently provides an open infrastructure for additional TTS or complimentary data sets to be amended by the authors and the open-source community in the future by request.

Home-to-work commute data

Currently, {TTS2016R} includes counts of full-time employed population by place of residence (origin), counts of full-time usual place of work (destination), number of trips to work by mode, and the calculated potential travel time of the trips in the GGH. The GGH (and hence the TTS survey area) is displayed in Figure 1.

This data is aggregated and available at the level of TAZ: TAZ are a spatial unit of analysis typically used to estimate the number of trips produced and attracted to each zone (Meyer and Miller, 2001). They are thus defined by transportation planners for a region based on intra-similarity and inter-dissimilarity between land-use and population demographics. Within the GGH boundaries, 3,764 TAZ are specified and each TAZ is uniquely identified using the GTA06 Zoning System: the survey boundary is discussed in the 2016 TTS methodology and defined by the TTS (Data Management Group, 2018b). The TAZ range between ≥ 0.019 km^2 in spatial area to a maximum of 879 km^2 (median: 1.3 km^2 and 3rd quantile: 2.8 km^2).



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

Figure 1. TTS 2016 study area within the GGH in Ontario, Canada along with associated descriptive statistic of workers and jobs per TAZ, OD links (count of workers potentially interacting with their place of employment) by origin TAZ, and calculated OD car travel time (TT) per origin TAZ. 3,507 trips were not assigned TT as they are longer than 180 mins. Spatial boundary files are retrieved from the TTS which define the survey area (Data Management Group, 2018a): the 20 regions in the GGH are represented by black lines and labelled, the dark gray lines are planning boundaries.

Full-time employed people and associated places of employment

In the GGH, there are 3,446,957 workers, 3,081,900 jobs, and 3,282,611 work-related trips (for the 2016 TTS survey day). The values are organized within the origin

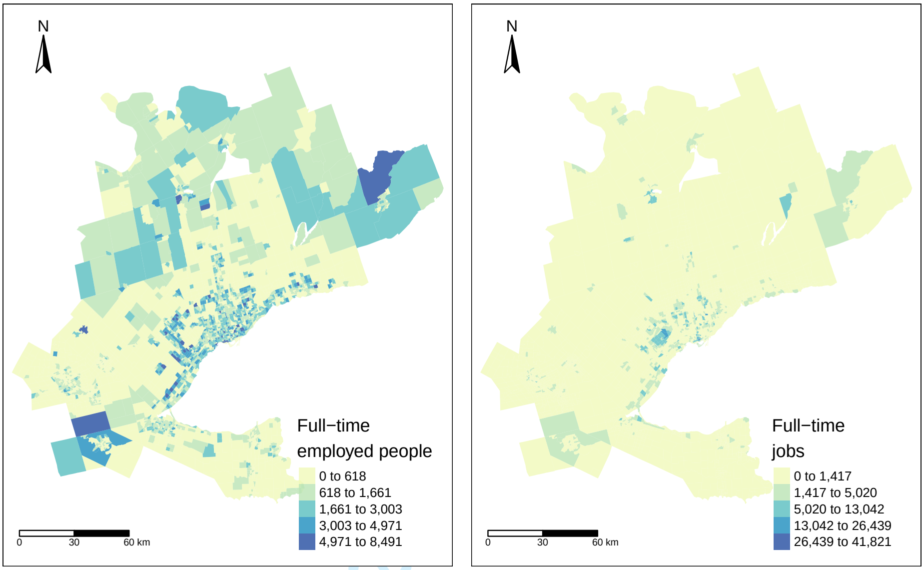


Figure 2. Number of workers (left) and jobs (right) in each TAZ retrieved from the 2016 TTS (Data Management Group, 2018b). Spatial boundary files are retrieved from the TAZ defined by the TTS (Data Management Group, 2018a).

destination (OD) table in the {TTS2016R} package and are derived from the cross-tabulation by person and by trip for the full-time employed population and associated places of employment.

Additionally, the total number of full-time workers and jobs in the TTS 2016 region are not equal. Since the outer boundaries of the TTS are permeable, workers who reside within the boundaries but have workplaces that are 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. The TTS is a proportionally representative survey, hence the values included in {TTS2016R} are adjusted to reflect the GGH working population and their home-based trips to places of GGH employment.

The count of links and trips made by the full-time working population and associated full-time place of employment per unique OD pair are quite variable. TAZ contain between 0 to 8,491 workers (median: 464, 3rd quantile: 1,378), 0 to 41,821 jobs (median: 244, 3rd quantile: 700), and generate between 0 to 241 trips (median: 15, 3rd quantile: 42).

Figure 2 presents the number of employed people and associated 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). Also, there is a higher number of TAZ with no workers than zones with no jobs (i.e., 791 TAZ with no workers and 396 TAZ with no jobs) and the mean of workers per TAZ is higher than the mean of jobs. The number TAZ with an extreme number of jobs at the highest and lowest percentiles is significantly higher than the number of workers.

Calculated travel time

Also included in {TTS2016R} are the estimated travel times between OD as summarized in the descriptive statistics table in Figure 3; travel times are calculated using the package {r5r}. {r5r} interfaces with the java-based R5 routing engine developed separately by Conveyal (Conveyal, 2022). The inputs to {r5r} for this data package were: the desired mode, a maximum travel time threshold of 180 minutes, the geo-coded origin destination pairs based on the centroids of the TAZ, and the static OpenStreetMap road network of Ontario (retrieved using Geofabrick (Geofabrik, 2022)). A travel time threshold of 180 minutes was selected since it captures almost all potential OD interactions.

Additionally, car travel is included in this data package because it is a critically important commute mode in the GGH. 2,598,379 of the trips are made using a car out of the total 3,282,611 work-related trips according to the TTS 2016 data (i.e., 79% of trips are taken by car).

These travel times are a useful addition to {TTS2016R} since they are not included in the TTS Data Retrieval System but they are vitally important to estimate the cost of travel and associated impedance functions, among other possible applications. If the readership is interested in additional information regarding the travel time computation, please see the calculation notebook in the documentation of {TTS2016R} and details about {r5r} at the r5r package website.

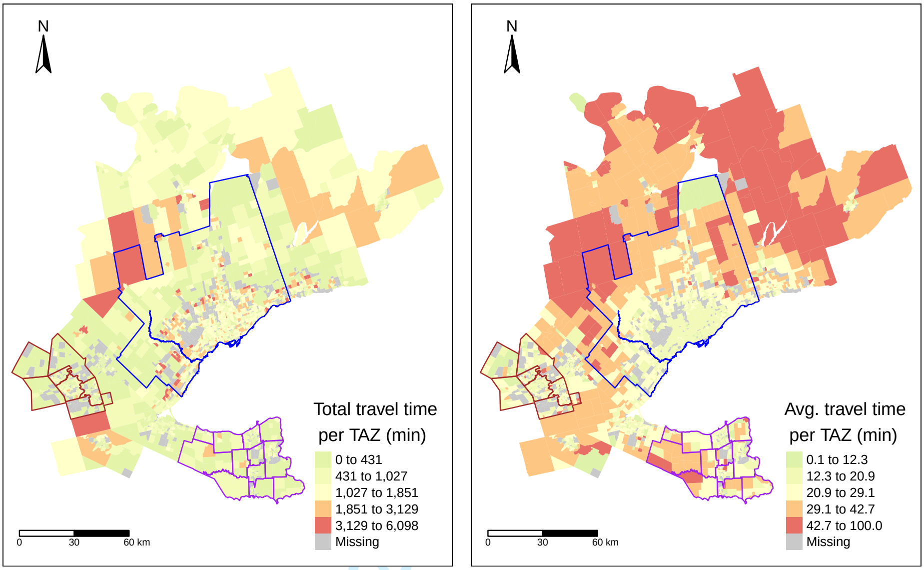


Figure 3. Calculated total worker travel time by car (left) and average worker travel time by car (right) for each TAZ in the 2016 TTS. Car travel times calculated using r5r (Pereira et al., 2021). Planning boundaries of Niagara and Waterloo (Data Management Group, 2018a), and the Toronto census metropolitan area (Statistics Canada, 2017) are drawn with purple, brown and blue borders, respectively.

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. For instance, we can see that in areas around the south-eastern border such as Niagara and Waterloo (purple and brown borders), the average travel times are moderately low. Additionally, travel times (by car) within the core of the Toronto census metropolitan area (CMA) (blue) is also moderately low since traffic congestion is not reflected in the travel time calculations. Further from these areas, travel times are higher.

Calibrating an impedance function

An application of the {TTS2016R} package is the calculation of 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}).

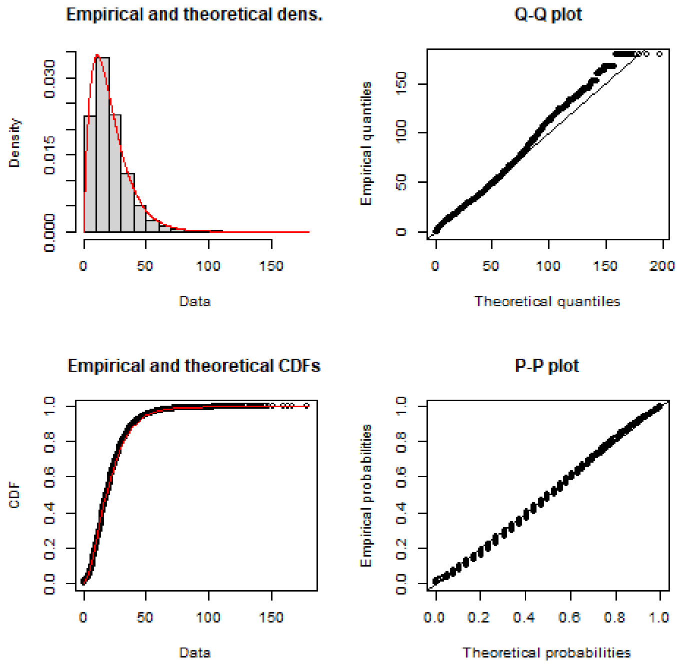


Figure 4. Empirical TTS 2016 home-based car TLD (black) and calibrated gamma distribution impedance function (red) with associated Q-Q and P-P plots

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 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

This paper introduces {TTS2016R}, an open data product in the form of a R data package. This package is a fusion of data from multiple sources and we demonstrate the spatial and numeric extent of the data contained within. It includes an OD cross-tabulation by person and by trip mode table for home-to-work commute data from the 2016 TTS alongside complimentary boundaries and estimated travel times. The value of this data package is in its transparency, easy of access, and its open infrastructure for the addition of complimentary data sets in the future. R users can immediately and easily explore GGH commute flow trends as well as suggestion further amendments to the package by request. 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.

Conveyal (2022) Conveyal R5 Routing Engine. URL https://github.com/conveyal/r5.

Data Management Group (2018a) Survey Boundary Files. URL http://dmg.utoronto.ca/survey-boundary-files.

Data Management Group (2018b) 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.

Geofabrik (2022) Ontario OpenStreetMap - Geofabrik Download Server. URL https://download.geofabrik.de/north-america/canada/ontario.html.

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.

Statistics Canada (2017) Boundary Files, 2016 Census. URL https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm.

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.

Reviewer Response Letter

The authors thank the reviewers for their time and thoughtful commentary in response to our paper. We have extensively revised the paper to clarify the comments in the paper. In this respect, we present the comments in blue text, our responses and comments in black text, and the relevant revised text is presented in red.

1. Response to Referee #1:

1.1. Firstly, I should say that I've never used R, so I can't comment on how unique or useful these techniques are in terms of the libraries being used. What I would like to have seen, though, is how computationally efficient this method is and how it would scale to larger datasets (e.g. memory requirements or time to compute). There is no mention of how long the cost function took to compute using R5R, or what network was used for the road travel times. I assume OpenstreetMap, but this needs to be stated.

I think that the shapefile in your GitHub repo has about 100 geographic areas, which would make for 100,000 interactions from every possible pair? I would like to see more on how big the inputs are as 100x100 is very different from 1000x1000. I assume it scales with O(n^2), and that you're using R5R to compute 100,000 shortest paths? Is this Dijkstra, an optimised contracted hierarchies algorithm, CPU or a GPU algorithm? This would limit the applicability of the technique to larger sized metropolitan areas if the compute scales as n^2 and the calculation of the cost function is computationally intensive.

- Thank you for these suggestions to clarify what is a R data-package and how aspects of the R data package were calculated for those who are unfamiliar to R.
- To more clearly define a R package for new-to-R users, the Introduction Section is revised as follows:

> *What is a R data package? A R data package contains code, data, and documentation in a standardised collection format that can be installed by R users through a centralized software repository such as CRAN (the Comprehensive R Archive Network) and GitHub. {TTS2016R} is freely available on GitHub for all to install and freely use in the spirit of open and reproducible research. Currently and in more detail, {TTS2016R} includes full-time home-based work-to-job origin destinations (OD) counts and mode-specific trip numbers retrieved from the 2016 TTS, traffic analysis zone (TAZ) boundaries, and municipality, planning, and census metropolitan area boundaries for the Greater Golden Horse area (GGH) located in southern Ontario, Canada. In addition, the package includes TAZ centroid-to-centroid travel times by car, transit, cycling, and walking mode computed using package {r5r} (Pereira et al., 2021).*

- There are 103,076 origin TAZ to destination TAZ pairs in the 'od' table. Inspecting this table, each origin TAZ is matched with between 1 to 241 destination TAZ. This means practically across the GGH, the total number of workers at an average origin TAZ have a total of approximately 35 work place TAZ destinations on average. There are 3,764 TAZ in the GGH, and not all TAZ have workers and/or places of employment, thus the TTS 2016 yields 103,076 origin destination pairs, so for those pairs we associate a travel time.

- 'r5r' is an interface for the java-based r5 routing engine (https://github.com/conveyal/r5). The outputs from r5r was a 3756*3756 travel times, but only the travel times from existing OD pairs was retained. The aim of including the travel time within this data package is not to demonstrate the benefits of 'r5r', but to associate an estimated 'cost of travel' variable to the TTS 2016 data set in an R environment.
- Your questions on computational efficiency, and details about the 'r5r' method are important. We've added additional details about 'r5r' and direct readers to additional a source to see details about commutation efficiency associated with the 'r5r' package as follows:

> *Also included in {TTS2016R} are the estimated travel times between OD as summarized in descriptive statistics table in Figure 1; travel times are calculated using the package {r5r}. {r5r} interfaces with the java-based {r5} routing engine developed separately by Conveyal (Conveyal, 2022). The inputs to {r5r} for this data package were: the desired mode, a maximum travel time threshold of 180 minutes, the geo-coded origin destination pairs based on the centroids of the TAZ, and the static OpenStreetMap road network of Ontario (retrieved using Geofabrick (Geofabrik, 2022)). A travel time threshold of 180 minutes was selected since it captures almost all potential OD interactions.*
>
> *Additionally, the car mode was included since it is a critically important commute mode in the GGH. 2,598,379 of the trips are made using a car mode out of the total 3,282,611 work-related trips according to the TTS 2016 data (i.e. 79% of trips are taken by car).*
>
> *These travel times are useful addition to {TTS2016R} since they are not included in the TTS Data Retrieval System but they are vitally important to estimate the cost of travel and associated impedance functions, among other possible applications. If the readership is interested in additional information regarding the travel time computation, please see the calculation notebook in the documentation of {TTS2016R} and details about {r5r} at the [r5r package website](https://ipeagit.github.io/r5r/index.html).*

- We also add the following to the text discussing Figure 3 (travel times) to emphasis the uncongested nature of the road network:

> *Additionally, travel times (by car) within the core of the Toronto census metropolitan area (CMA) (blue) is also moderately low since traffic congestion is not reflected in the travel time calculations. Further from these areas, travel times are higher.*

I've got a few minor points as follows:

1.2. P2, line 26: What is "n=3764 within the survey boundaries"? I couldn't understand what data you were using to make the travel-to-work flows.

- Thank you for your comment asking to clarify what data is used to make the travel-to-work flows, specifically "P2, line 26: What is "n=3764 within the survey boundaries"?". We removed the 'n' terminology all together and added additional explanation regarding the significance of TAZ:

> *This data is aggregated and available at the level of TAZ: TAZ are a spatial unit of analysis typically used to estimate the number of trips produced and attracted to each zone (Meyer and Miller, 2001). They are thus defined by transportation planners for a region based on intra-similarity and inter-dissimilarity between land-use and population demographics. Within the GGH boundaries, 3 ,764 TAZ are specified and each TAZ is uniquely identified using the GTA06 Zoning System: the survey boundary is discussed in the 2016 TTS methodology and defined by the TTS (Data Management Group, 2018b). The TAZ range between ≥ 0.019 km² in spatial area to a maximum of 879 km² (median:1.3 km² and 3rd quantile: 2.8 km²).*

1.3. P2, line 38: What is the network used for travel time calculations? (also see P.4 line 35)

- Thank you for you question asking to clarify travel time calculations (P2, line 38 and P.4 line 35). Good point. In the response to your first comment, we clarify the use of the network. We repeat the text that is added to the revised manuscript as follows:

> *Also included in {TTS2016R} are the estimated travel times between OD as summarized in descriptive statistics table in Figure 1; travel times are calculated using the package {r5r}. {r5r} interfaces with the java-based {r5} routing engine developed separately by Conveyal (Conveyal, 2022). The inputs to {r5r} for this data package were: the desired mode, a maximum travel time threshold of 180 minutes, the geo-coded origin destination pairs based on the centroids of the TAZ, and the static OpenStreetMap road network of Ontario (retrieved using Geofabrick (Geofabrik, 2022)). A travel time threshold of 180 minutes was selected since it captures almost all potential OD interactions.*

1.4. P3, line 34: "and jobs is the TTS 2016 region" should be "jobs in the..."?

- Thank you for pointing out the grammatical error in P3 line 34, it has now been corrected.

1.5. P4, line 35: What is the departure time used? Is the travel time the same across the day? Does R5R take into account rush hour traffic congestion or this a simple "20 miles driven on a 40mph limit road takes 30 minutes" calculation? I'd like to see a bit more information. Also, one mode (road) for all commuter transport seems like an over-simplification. How much rail travel is happening in the area? Walk/bike etc?

- Thank you for your questions regarding the need for clarity in the travel time calculations (P4, line 35).
- Details about the departure time were removed – since this a static OpenStreetMap road network was used, traffic conditions are neglected so time of departure is not used in the car travel time calculations. We hope we sufficiently clarified the type of network used for the travel time calculations in addressing your comment above.
- We chose to only include the travel times for car commuting because they are so critically important to the GGH. Reflecting on your comment, we added the number of work-related trips for each OD by primary mode, to the OD table in the package. With this data, the modal split was calculated and overall, 79% of trips taken using a car. We added in this context into the manuscript as follows:

> *Additionally, car travel is included in this data package because it is a critically important commute mode in the GGH. 2,598,379 of the trips are made using a car*

out of the total 3,282,611 work-related trips according to the TTS 2016 data (i.e., 79% of trips are taken by car).

1.6. P5, line 29: Where you reference Toronto and Hamilton on the choropleth maps, it took me a while to realise that I had to refer back to Fig 1 which has the names. The "south-eastern border" and "eastern areas" descriptions had me looking in the wrong areas on Fig 3, until I looked back to Fig 1. It needs something on the Fig 3 map to highlight the areas for those of us who don't know Canada. Or at the very least "refer to Fig 1 for names".

- Thank you for identifying this difficult in interpreting Figure 3 (P5, line 29). We updated Figure 3's palette and added your suggestions in the text describing Figure 3 as follows:

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. For instance, we can see that in areas around the south-eastern border such as Niagara and Waterloo (purple and brown borders), the average travel times are moderately low. Additionally, travel times (by car) within the core of the Toronto census metropolitan area (CMA) (blue) is also moderately low since traffic congestion is not reflected in the travel time estimations. Further from these areas, travel times are higher.

We hope that these modifications satisfy all the comments you communicated. Thank you greatly for your time in reviewing our paper, we believe its quality has been enhanced.

2. Response to Referee #2:

Here are a few comments that might be helpful:

2.1. The link to access the original dataset through the 'http://dmg.utoronto.ca/survey-boundary-files' page provided in the GitHub documentation was not working at the time of this review.

- The link (http://dmg.utoronto.ca/survey-boundary-files) seems to be working now! Thanks for flagging this, we will monitor it to ensure this link remains live.

2.2. p. 2, l. 24-30: TAZ are central to the dataset and, as such, should be described more precisely. Could you provide an indication of their sizes and regularity? -- "Based on land-use and population demographics", such as? and how?

- Thanks for this comment about the need to clarify TAZ (p. 2, l. 24-30). We agree TAZ are critical to the dataset. We have added an additional sentence to describe the TAZ used. The text is revised as follows:

> *This data is aggregated and available at the level of TAZ: TAZ are a spatial unit of analysis typically used to estimate the number of trips produced and attracted to each zone (Meyer and Miller, 2001). They are thus defined by transportation planners for a region based on intra-similarity and inter-dissimilarity between land-use and population demographics. Within the GGH boundaries, 3 ,764 TAZ are specified and each TAZ is uniquely identified using the GTA06 Zoning System: the survey boundary is discussed in the 2016 TTS methodology and defined by the TTS (Data Management Group, 2018b). The TAZ range between ≥ 0.019 km2 in spatial area to a maximum of 879 km2 (median:1.3 km2 and 3rd quantile: 2.8 km2).*

2.3. p. 3, l. 35-42: It would be interesting to indicate how the workers that are travelling outside of the GGH boundaries are identified in the data (i.e., what destination does the origin-destination table record for them).

- Thank you for this comment regarding workers traveling outside of the GGH boundaries (p.3, 1. 35-42) – we wondered about this too!
- To our knowledge, the TTS does not provide data on where workers inside the TTS travel for employment, only that they travel 'externally'. The majority of the workers who travel externally are in closer proximity to the edge of the GGH, which makes sense.
- Additionally, the GGH accounts for approximately 55% of Ontario's entire population and captures many of the provinces census agglomerations.
- Considering the significant of the GGH population and the limitations of the TTS, we can't provide any additional information about the externally-travelling GGH workers.

2.4. p. 4, l. 36: "a small proportion of trips are taken by non-car modes": is there a way to know how much from another source?

- Great suggestion regarding the proportion of trips taken by non-car modes (p. 4, l. 36).
- We added the number of trips taken from O to D using a primary mode (car, transit, active, etc.) to the OD table so we were able to identify how prominent car mode is. As such, we replaced the text describing the car-commute in the GGH as follows:

> *Additionally, car travel is included in this data package because it is a critically important commute mode in the GGH. 2,598,379 of the trips are made using a car*

out of the total 3,282,611 work-related trips according to the TTS 2016 data (i.e., 79% of trips are taken by car).

2.5. p. 4, Figure 2. The equivalent plots in the GitHub documentation are far more readable. Would it be possible to change the palette to something else?

- Great suggestion regarding p.4 Figure 2 – we updated the figure.
- In the GitHub readme, the figures shown just use a continuous scale – this makes the map look nicer. In the Figures included in the manuscript, we used a different plotting pack and set the scale to jenks distribution to emphasis the differences in the origin manuscript submission.
- However – we changed the palette to Blue-Yellow and muted the colours so the figures in this submission are more visually appealing from our perspective but still have a scale set to a jenks distribution. We like it more now - hope you like it too!

2.6. p. 5, l. 32: please indicate where Peterborough is located on the map.

- Thank you for pointing out that it is difficult to identify where Peterborough is located on the map (p.5, 1. 32). We've added a suggestion from referee #1 that clarifies this issue as follows:

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. For instance, we can see that in areas around the south-eastern border such as Niagara and Waterloo (purple and brown borders), the average travel times are moderately low. Additionally, travel times (by car) within the core of the Toronto census metropolitan area (CMA) (blue) is also moderately low since traffic congestion is not reflected in the travel time estimations. Further from these areas, travel times are higher.

2.7. p. 7, l. 7: "are included as a compliment"

- Nice catch (p. 7, l.7), we rephrased the concluding remarks sections. The parts pertaining to 'complimenting' data now read:

This paper introduces {TTS2016R}, an open data product in the form of a R data package. This package is a fusion of data from multiple sources and we demonstrate the spatial and numeric extent of the data contained within. It includes an OD cross-tabulation by person and by trip mode table for home-to-work commute data from the 2016 TTS alongside complimentary boundaries and estimated travel times.

We hope that the modifications made as suggested by your comments are to your satisfaction. Thank you for your time and insightful feedback.

3. Response to Editor

3.1. I have received two reports on your submission, and I have also inspected it myself in detail. My sense is that there is value in the manuscript and the data product, but it needs to be further "unearthed" to make it as useful as possible for the EP-B readership. Both referees provide useful comments I'd like you to address (in the case of R.1, I don't expect you to include detailed computational comparisons and this is not the venue for it, but I do think it'd be useful to try to use their comments as guidance to make your paper clearer as to what it tries and doesn't try to do). Beyond the two reports, there are two further aspects I think would be useful to incorporate in the paper:

- Thank you for these comments. We hope we addressed the comments from the two reviewers up to your satisfaction.

3.2. Contribution of the data product: what is new? From my reading, it is not entirely clear if what you have packaged is entirely provided by TTS or you have done an additional exercise in fusing the data with additional sources (e.g., census). Could you please walk the reader through what is the data product adding above and beyond what is already available for direct download from the official TTS page? If there is only data that is already available from the official source, I'd consider combining it with other data that you think would be useful in this context.

- Agreed. Your suggestion along with the other reviews inspired us to incorporate more data into the data package. We added the number of trips, by mode, to the origin destination table. This will facilitate the analysis of modal trends in commute.
- Furthermore, we added IDs to the TAZ (which are defined by the 2016 TTS) that correspond to the Canadian Census. This will facilitate the simple integration of any census variables which users of the package can easily extract. These variables can be widely used for a variety of analysis.
- We edited the introductory section to better capture the contribution of the data product as follows:

> *This manuscript presents the open data product [{TTS2016R}](https://github.com/soukhova/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 @Arribas2021open]. The product presented in this paper is a `R` data package which currently consists of a fusion of objects from a variety of sources: home-to-work flows and trips by mode are sourced from the 2016 Transportation Tomorrow Survey (TTS) [@data_management_group_tts_2018], estimated travel times are calculated, and boundary files from the TTS and from the Canadian Census.*
>
> *What is a `R` data package? A `R` data package contains code, data, and documentation in a standardised collection format that can be installed by `R` users through a centralized software repository such as CRAN (the Comprehensive R Archive Network) and GitHub. {TTS2016R} is freely available on GitHub for all to install and freely use in the spirit of open and reproducible research. Currently, {TTS2016R} includes full-time home-based work-to-job origin destinations counts and mode-specific trip numbers retrieved from the 2016 TTS, traffic analysis zone (TAZ) boundaries, and municipality, planning, and census metropolitan area*

> *boundaries for the Greater Golden Horse area (GGH) located in southern Ontario, Canada. In addition, the package includes TAZ centroid-to-centroid travel times by car, transit, cycling, and walking mode computed using package {r5r} [@Pereira2021r5r].*
>
> *The aim of this paper is to walk readers through the data sets, illustrate a use case (i.e., the calculation of an impedance function that can be used to calculate accessibility to employment), and invite others to experiment in its uses and applications. Though data from the TTS is freely available to the public through the [TTS Data Retrieval System](https://dmg.utoronto.ca/idrs/index), 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 and packaging it alongside complimentary data, {TTS2016R} offers a slice of the TTS data which can be immediately used by `R` users to analysis patterns of commuting to work in the region. Anticipate this package to grow in the future: it currently provides an open infrastructure for additional TTS or complimentary data sets to be amended by the authors and the open-source community in the future by request.*

3.3. Purpose and value: there is a useful example of what the package can be used for, but it is very specific. I'd like to see a more general reflection on how you expect this package to be used by the wider community (e.g., Urban Data/Code readership).

- Excellent suggestion. In addition to editing the introduction, we added a few addition remarks to the conclusion to crystallize the purpose and value of this data package to the Urban Data/Code readership.

> *{TTS2016R}, the open data package introduced in this paper fuses multiple sources of data. It includes an OD cross-tabulation by person and by trip mode table for home-to-work commute data from the 2016 TTS alongside complimentary boundaries and estimated travel times. The value of this data package is in its transparency, easy of access, and its open infrastructure for the addition of complimentary data sets in the future. Using `R` users can immediately and easily explore GGH commute flow trends as well as suggestion further amendments to the package by request. 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.*

We hope that these modifications satisfy your recommendations and the comments communicated by the reviewers. Thank you for taking the time to provide this insightful and constructive feedback!

Journal Title
XX(X):2–11

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

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

~~Anastasia Soukhov~~Anon1, ~~Antonio Páez~~Anon2

Abstract

This paper describes and visualises the data contained within the {TTS2016R} ~~data-package~~ 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 (GGH) 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~~ the count of origin destination (OD) pairs and trips by mode per origin TAZ, calculated car travel time from TAZ ~~origin-destination centroid pairs~~OD centroid pairs, and associated spatial boundaries to link TAZ to the Canadian Census. 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} is a growing open data product built on R infrastructure that allows for the immediate access of home-to-work commuting data alongside complimentary objects from different sources. The package will continue expanding with additions by the authors and the community at-large by requests in the future. {TTS2016R} can be freely ~~downloaded and explored at:~~ explored and downloaded in the associated Github repository where the documentation and code involved in data creation, manipulation, and ~~the final~~ all open data products are detailed.

Keywords

Jobs; population; work; commute; travel time; impedance; Greater Toronto and Hamilton Area; Greater Golden Horshoe 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~~ a R data package which currently consists of ~~three objects which are~~ a fusion of objects from a variety of sources: home-to-work flows sourced from the 2016 Transportation Tomorrow Survey (TTS) ~~or are curated to facilitate the use and analysis of TTS data. This packageincludes person-to-jobs origin-destinations~~(Data Management Group, 2018b), estimated travel times (calculated using {r5r} (Pereira et al., 2021)), and boundary files from the TTS (Data Management Group, 2018a) and from the Canadian Census (Statistics Canada, 2017).

What is a R data package? A R data package contains code, data, and documentation in a standardised collection format that can be installed by R users through a centralized software repository such as CRAN (the Comprehensive R Archive Network) and GitHub. {TTS2016R} is freely available on GitHub for all to install and freely use in the spirit of open and reproducible research. Currently and in more detail, {TTS2016R} includes full-time home-based work-to-job origin destinations (OD) counts and mode-specific trip numbers retrieved from the 2016 TTS, traffic analysis zone (TAZ) boundaries, and ~~planning/municipality~~municipality, planning, and census metropolitan area boundaries for the Greater Golden Horse area (GGH) located in southern Ontario, Canada~~(Data Management Group, 2018b)~~. In addition, the package includes TAZ centroid-to-centroid travel times by car, transit, cycling, and walking mode 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~~ data sets, illustrate a use case (i.e., the calculation of an impedance function that can be used to calculate accessibility to employment), and invite ~~its use in other applications.~~ others to experiment in its uses and applications. Though data from the TTS is freely available to the public through the TTS Data Retrieval System, the raw data can be technically demanding, cumbersome to work with, and requires multiple software to process. By pre-processing the data, packaging it with complimentary data, and providing explicit documentation in a `R` environment, {TTS2016R} offers a slice of the TTS data that can be immediately used by `R` users to analysis patterns of commuting to work in the region. Anticipate this package to grow in the future: it currently provides an open infrastructure for additional TTS or complimentary data sets to be amended by the authors and the open-source community in the future by request.

Home-to-work commute data

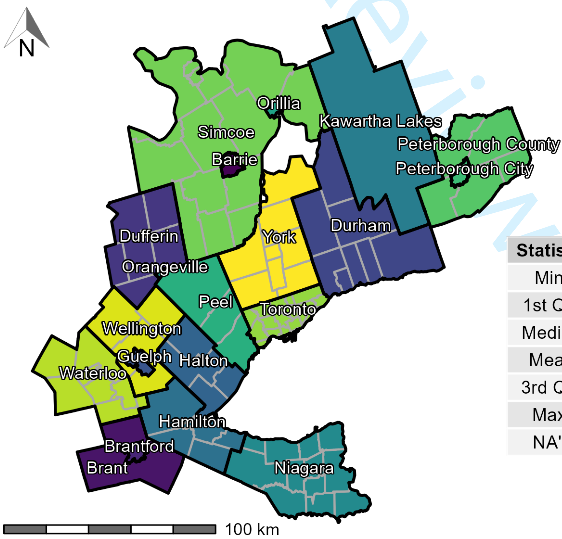
Currently, {TTS2016R} includes counts of ~~fully-employed~~ full-time employed population by place of residence (origin)~~and~~, counts of full-time ~~jobs by~~ usual place of work (destination)~~aggregated by TAZ (n=3, 764 within the survey boundaries) . TAZ typically are defined based on land-use and population demographics in order~~, number of trips to work by mode, and the calculated potential travel time of the trips in the GGH. The GGH (and hence the TTS survey area) is displayed in Figure 1.

This data is aggregated and available at the level of TAZ: TAZ are a spatial unit of analysis typically used to estimate the number of trips produced and attracted to each zone (Meyer and Miller, 2001). ~~As such,~~ They are thus defined by transportation planners for a region based on intra-similarity and inter-dissimilarity between land-use and population demographics. Within the GGH boundaries, 3,764 TAZ are specified and 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~~ : the survey boundary is discussed in the 2016 TTS methodology and defined by the ~~2016 TTS methodology (Data Management Group, 2018b).~~

~~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~~TTS (Data Management Group, 2018b). The TAZ range between ≥ 0.019 km^2 in spatial area to a maximum of 879 km^2 (median: 1.3 km^2 and 3rd quantile: 2.8 km^2).



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

Figure 1. TTS 2016 study area within the GGH in Ontario, Canada along with ~~the~~ associated descriptive ~~statistics~~ statistic of ~~the trips~~workers and jobs per TAZ, OD links (count of workers potentially interacting with their place of employment) by origin TAZ, and calculated ~~origin-destination~~ OD car travel time (TT) ~~, workers~~ per origin TAZ. 3,~~and jobs per TAZ~~507 trips were not assigned TT as they are longer than 180 mins. ~~Contains~~ Spatial boundary files are retrieved from the TTS which define the survey area (Data Management Group, 2018a): the 20 regions ~~(~~in the GGH are represented by black ~~boundaries)~~ lines and ~~sub-regions (~~labelled, the dark gray lines are planning boundaries~~)~~.

Full-time employed people and associated places of employment

~~The origin-destination table (i.e., trips) consists of a~~ In the GGH, there are 3,446,957 workers, 3,081,900 jobs, and 3,282,611 work-related trips (for the 2016 TTS survey day). The values are organized within the origin destination (OD) table in the {TTS2016R} package and are derived from the cross-tabulation ~~of people who are employed~~ by person and by trip for the full-time ~~by place of GGH residence (origin) and places of GGH employment(destination) using the GTA06 TAZ zoning system. It is important to note that~~ employed population and associated places of employment.

Additionally, the total number of full-time workers and jobs ~~is~~ in 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~~ have workplaces that are 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~~ . The TTS is a proportionally representative survey, hence the values included in {TTS2016R} ~~offers a perspective on all workers in the GGH~~ are adjusted to reflect the GGH working population and their home-based trips to places of GGH employment.

The count of links and trips made by the full-time working population and associated full-time place of employment per unique OD pair are quite variable. TAZ contain between 0 to 8,491 workers (median: 464, 3rd quantile: 1,378), 0 to 41,821 jobs (median: 244, 3rd quantile: 700), and generate between 0 to 241 trips (median: 15, 3rd quantile: 42).

Figure 2 presents the number of ~~workers and~~ employed people and associated jobs per TAZ. It can be observed that the spatial distribution of jobs and workers is unequal, which is indicative of a ~~jobs-housing~~ jobs -housing imbalance that can impact accessibility in a region (Levine, 1998). ~~It can also be seen that~~ Also, there is a higher number of TAZ with no workers than zones with no jobs (i.e., 791 TAZ with no workers ~~:~~ and 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~~ . The number TAZ with an extreme number of jobs at the highest and lowest percentiles is significantly higher than the number of workers.

Calculated travel time

~~As mentioned,~~ Also included in {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~~ are the estimated travel times between OD as summarized in the descriptive statistics table in Figure 3; travel times are 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~~ . {r5r} interfaces with the java-based R5 routing engine developed separately by Conveyal (Conveyal, 2022). The inputs to

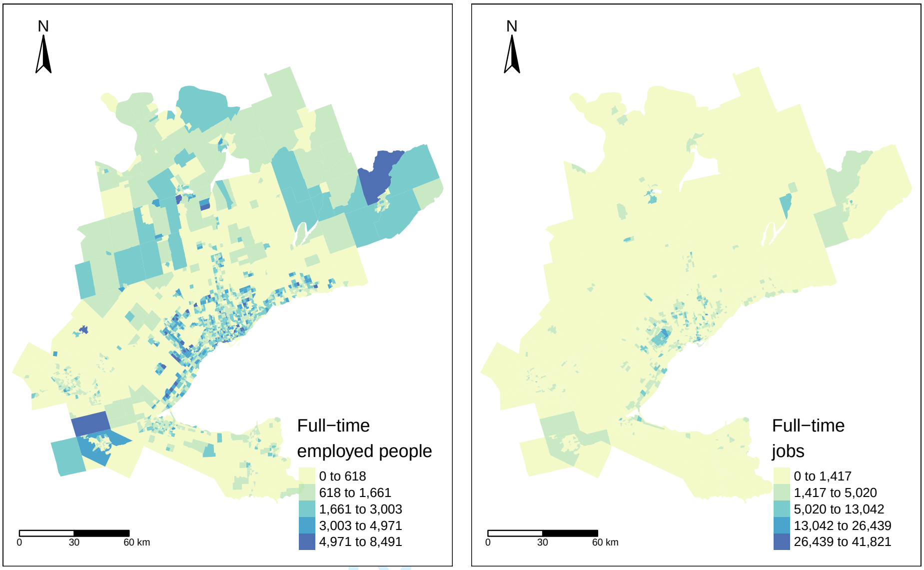


Figure 2. Number of workers (left) and jobs (right) in each TAZ ~~in~~ retrieved from the 2016 TTS (Data Management Group, 2018b). Spatial boundary files are retrieved from the TAZ defined by the TTS (Data Management Group, 2018a).

{r5r} for this data package were: the desired mode, a maximum 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~~ threshold of 180 minutes, the geo-coded origin destination pairs based on the centroids of the TAZ, and the static OpenStreetMap road network of Ontario (retrieved using Geofabrick (Geofabrik, 2022)). A travel time threshold of 180 minutes was selected since it captures almost all potential OD interactions.

Additionally, car travel is included in this data package because it is a critically important commute mode in the GGH. 2,598,379 of the trips are made using a car out of the total 3,282,611 work-related trips according to the TTS 2016 data (i.e., 79% of trips are taken by ~~one-time departure cartrip~~car).

These travel times are a useful addition to {TTS2016R} since they are not included in the TTS Data Retrieval System but they are vitally important to estimate the cost of travel and associated impedance functions, among other possible applications. If the readership is interested in additional information regarding the travel time computation, please see the calculation notebook in the documentation of {TTS2016R} and details about {r5r} at the r5r package website.

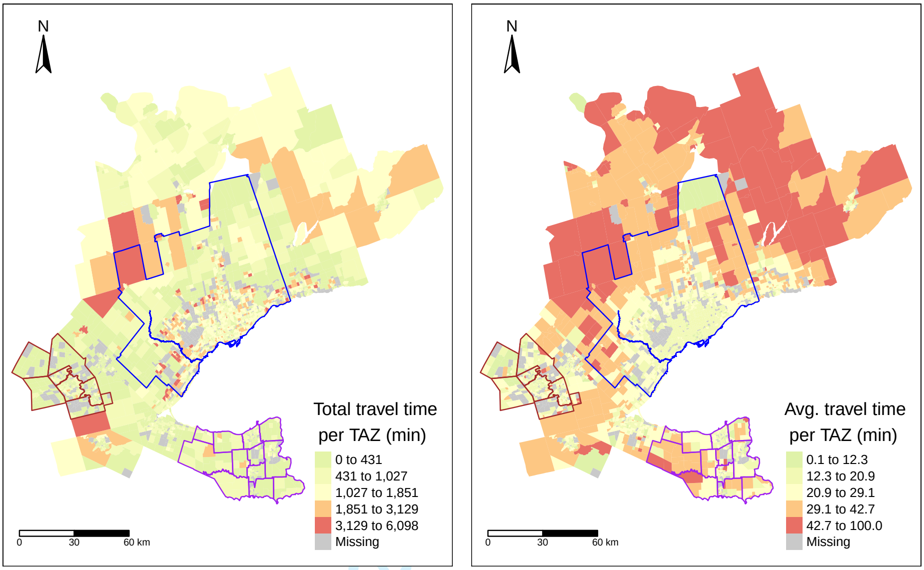


Figure 3. Calculated total worker travel time by car (left) and average worker travel time by car (right) for each TAZ in the 2016 TTS. Car travel times calculated using r5r (Pereira et al., 2021). Planning boundaries of Niagara and Waterloo (Data Management Group, 2018a), and the Toronto census metropolitan area (Statistics Canada, 2017) are drawn with purple, brown and blue borders, respectively.

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~~ For instance, we can see that in areas around the south-eastern border ~~that make up the Greater Toronto and Hamilton Area (GTHA), Niagara and Waterloo,~~ such as Niagara and Waterloo (purple and brown borders), the average travel times are moderately low. Additionally, travel times (by car) within the core of the Toronto census metropolitan area (CMA) (blue) is also moderately low since traffic congestion is not reflected in the travel time calculations. 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

An application of the {TTS2016R} package is the calculation of 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}

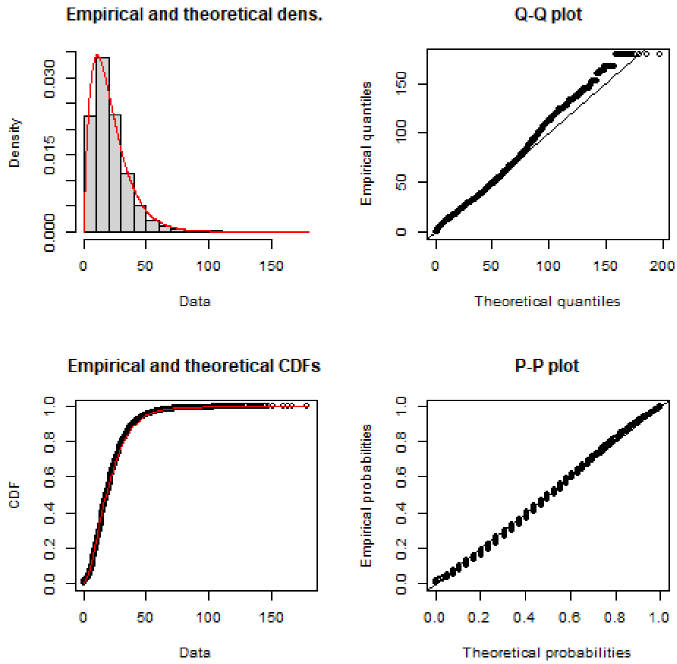


Figure 4. Empirical TTS 2016 home-based car TLD (black) and calibrated gamma distribution impedance function (red) with associated Q-Q and P-P plots

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 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).

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

$$f(x,\alpha,\beta) = \tfrac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \quad \text{for } 0 \leq x \leq \infty \qquad \Gamma(\alpha) = \textstyle\int_0^{\infty} x^{\alpha-1}e^{-x}\,dx \tag{1}$$

Concluding remarks

~~The~~ This paper introduces {TTS2016R}, an open data product ~~introduced in this paper shares tables~~ in the form of a R data package. This package is a fusion of data from multiple sources and we demonstrate the spatial and numeric extent of the data contained within. It includes an OD cross-tabulation by person and by trip mode table for home-to-work ~~related~~ commute 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~~ alongside complimentary boundaries and estimated travel times. The value of this data package is in its transparency, easy of access, and its open infrastructure for the addition of complimentary data sets in the future. R ~~environment~~users can immediately and easily explore GGH commute flow trends as well as suggestion further amendments to the package by request. 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.

Conveyal (2022) Conveyal R5 Routing Engine. URL https://github.com/conveyal/r5.

Data Management Group (2018a) Survey Boundary Files. URL http://dmg.utoronto.ca/survey-boundary-files.

Data Management Group (2018b) 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.

Geofabrik (2022) Ontario OpenStreetMap - Geofabrik Download Server. URL https://download.geofabrik.de/north-america/canada/ontario.html.

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.

Statistics Canada (2017) Boundary Files, 2016 Census. URL https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm.

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.