Leveraging GTFS to explore spatial patterns in transit supply with respect to social needs

James Reynolds[a,1,*], Yanda Qu[a,2], Graham Currie[a,3]

[a]*Public Transport Research Group (PTRG), Institute of Transport Studies, Department of Civil Engineering Engineering, Monash University, Clayton Campus, Melbourne, 3800, Victoria, Australia*

Abstract

This is the abstract.

It consists of two paragraphs.

Keywords: keyword1, keyword2

1. Introduction

Currie (2010) introduced a transit Supply Index (SI) and demonstrated how it could be used to assess spatial gaps between the social need for public transport and what is supplied. The SI is based on calculating a score, for each area of interest, based on the rate of transit arrivals at nearby stops, adjusted for the amount of each area that is within a typical walking distance. Currie (2010) calculated scores for areas within Greater Melbourne in Victoria, Australia, combined these with census data to map locations with very high social needs for transit but very low or zero supply.

Unfortunately, this approach does not appear to have been widely used, perhaps in part because at the time it was first published the transit Supply Index (SI) was not easy to calculate. The Currie (2010) analysis was based on combining multiple operator databases and service frequency data that had been manually extracted from transit agency websites.

Nowadays, however, the General Transit Feed Specification (GTFS) allows timetable data to be published in a standardized format, and more than 10,000 agencies provide feeds(MobilityData, undated). Tools for analysing GTFS data are widely available, but there is a gap as there is not yet a tool to calculate SI scores directly from GTFS datasets. It is also unclear whether the gaps between social needs and transit supply identified in Currie (2010) in Greater Melbourne have gotten better in the almost two decades since the original analysis. Nor is it clear whether the spatial patterns of transit need, supply and gap in Melbourne are representative of other cities.

This provides the motivation for the research reported in this paper, in which a new R package (gtfssupplyindex) specifically developed to calculate SI scores is presented. The paper also reports results for Greater Melbourne in 2016 and 2021, matching the most recent censuses and allowing comparison to the 2006 result reported in Currie (2010). Comparisons are also made to other parts of Australia, so as to explore whether findings about Greater Melbourne can be confidently generalized to other places.

The remainder of this paper is structured as follows: the next section outlines the background to this research, including the original formulation of the Transit Supply Index, and an explanation of the GTFS.

*Corresponding author

Email addresses: `james.reynolds@monash.edu` (James Reynolds), `yanda.qu@monash.edu` (Yanda Qu), `graham.currie@monash.edu` (Graham Currie)

[1]Research Fellow

[2]PhD Strudent

[3]Professor

Section 3 then describes the study methodology, followed by presentation of results in Section 4. Section 5 discusses the results and the limitations of this stud, and outlines directions for future research. A brief conclusion is provided in Section 6.

2. Background

2.1. Transit metrics

Even a brief search reveals many metrics available for benchmarking transit services. Examples include: (1) those in the Transit Cooperative Research Program (TCRP) Report 88, which is an extensive guidebook on developing a performance-measurement system (Ryus et al., 2003); (2) online databases provided by the Florida Transit Information System (FTIS) (Florida Transit Information System, 2018) and International Association of Public Transport (UITP) (2015); (3) those used in the extensive annual benchmarking program undertaken yearly by the Transport Strategy Centre in the United Kingdom, including over 100 transit providers around the world (Imperial College London, undated); and (4) a recently developed methodology to calculate 'blank spots', beyond typical walking access distances to/from transit stops(Alamri et al., 2023).

The Fielding Triangle (Fielding, 1987) provides a framework for combining indicators of service inputs, outputs and consumption to describe cost efficiency, cost effectiveness and service effectiveness. More broadly:

- Litman (2003) and Litman (2016) discuss some of the traffic, mobility, accessibility, social equity, strategic planning and other rational decision-making-based perspectives underling transport indicators;
- Reynolds et al. (2017) extends these into models of how institutionalism, incrementalism and other public policy analysis concepts might apply to decision-making processes relating to transit prioritization;
- Guzman et al. (2017), developed a measure of accessibility in the context of policy development and social equity for Latin American Bus Rapid Transit (BRT) networks; and
- Creutzig et al. (2020) introduced street space allocation metrics based around 10 ethical principles

However, many transit metrics may be difficult to calculate, complex to explain or understand, and likely not well suited to communication with those who are not planners, engineers or other technical specialists. Where pre-calculated metrics are immediately available it may not be possible for practitioners, researchers or advocates to independently generate scores so as to test proposed system changes. Sometimes it is not even possible to know precisely how scores for the existing services levels are calculated. The Transit Score metric, which is readily available on the Walk Score (2023) website, provides an example. It is available for locations with a published GTFS feed, eliminating the need for any calculations. The meaning of these Transit Scores also appears easy to explain, with the highest possible score of 100 representing what might be experienced in the center of New York. However, the Transit Score algorithm is unpublished, and effectively a black box. It does not appear that Transit Scores can be calculated independently or generated for proposed changes to networks.

A contrasting example is the Transit Capacity and Quality of Service Manual (TCQSM), which provides a wealth of metrics for measuring different aspects of a transit system. The TCQSM scores themselves are easy to understand or explain, ranging from A to F, although the number of metrics is very large, which might limit the practicality of using the TCQSM in practice for communicating wiht non-technical audiences. All of these can be calculated independently, given sufficient data, and Wong (2013) provides an example reporting various TCQSM metrics across 50 transit operators. The Wong (2013) analysis is made possible by the availability of General Transit Feed Specification (GTFS) datasets for each of the transit systems

The GTFS is an open, text-based format, developed originally to allow transit to be included in the Google Maps navigation platform (MobilityData, undated). Figure @ref(fig:GTFS_ERD) shows an Entity Relationship Diagram (ERD) of the GTFS data structure. This indicates how GTFS data is stored as a series of tables (agency, routes, trips etc.) with primary and foreign keys (agency_id, route_id, trip_id etc.) providing links.

GTFS software tools are available for performing various analyses, creating visulizations or other purposes. However, software to calculate the Transit Suppy Index (SI) is not yet available.

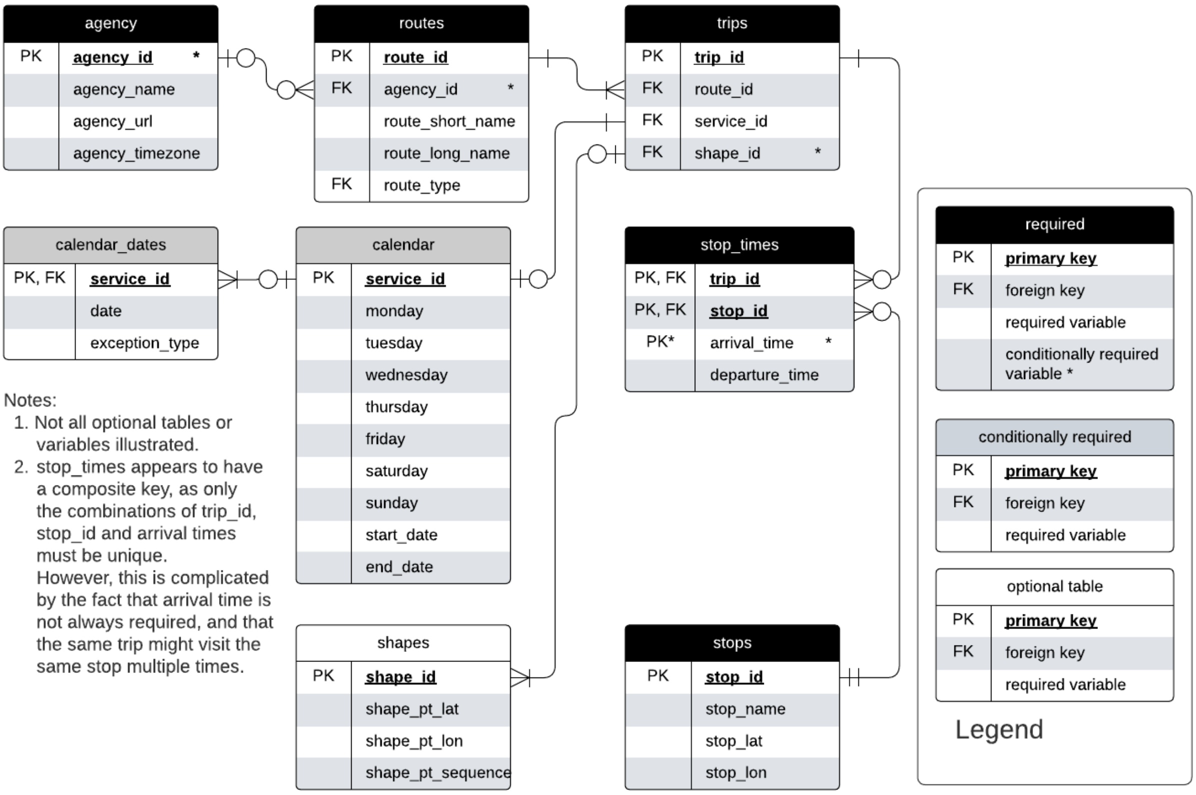

Figure 1: GTFS entity relationship diagram. Source: adapted by author from Alamri et al (2023) and the GTFS Schedule Reference (16/11/2023 revision).

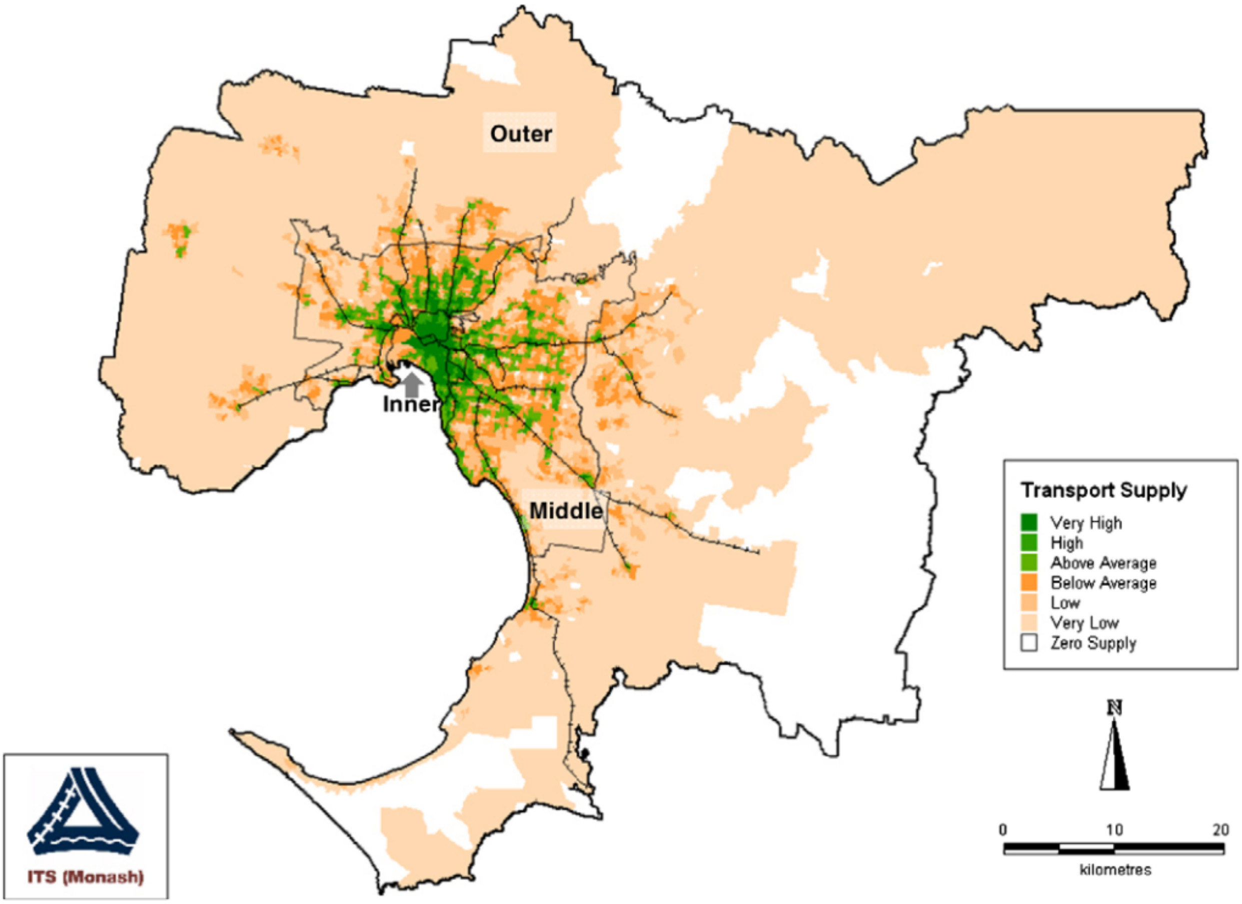

Figure 2: Distribution of supply measure scores – Metropolitan Melbourne (2006), Source: Currie (2010)

2.2. The Transit Suppy Index

A generalized form of the SI equation , adapted from Currie (2010), is:

$$SI_{area,time} = \sum \frac{Area_{Bn}}{Area_{area}} * SL_{n,time}$$

where: (1) $SI_{area,time}$ is the Supply Index for the area of interest and a given period of time; (2) $Area_{Bn}$ is the buffer area for each stop (n) within the area of interest. In Currie (2010) this was based on a radius of 400 metres for bus and tram stops, and 800 metres for railway stations; (3) $Area_{area}$ is the area of the area of interest; and (4) $SL_{n,time}$ is the number of transit arrivals for each stop for a given time period.

The SI is additive, in that $SI_{area,time}$ scores can be aggregated to calculate an overall score across multiple time periods or for a region encompassing multiple areas of interest.

Currie (2010) presented maps of SI scores and areas with very high social needs but zero or very low transit supply, and a chart contrasting needs and supply (see Figures 2, 3 and 4. However, it doesn't appear that there have been similar analyses undertaken since then, so it is unclear whether gaps between transit needs and supply have improved, remained similar or become worse. Nor does it appear that this methodology has been applied to other places, except for similar studies of Hobart and Adelaide discussed in Currie (2010). Hence, it is unclear whether the patterns in Melbourne, where areas with very high transport needs but zero or very low transit supply tend to be in middle and outer areas of the city serviced by buses, is consistent with patterns in other cities. Comparing current conditions and other locations to the findings of Currie (2010), therefore, is an aim of the research discussed in this paper.

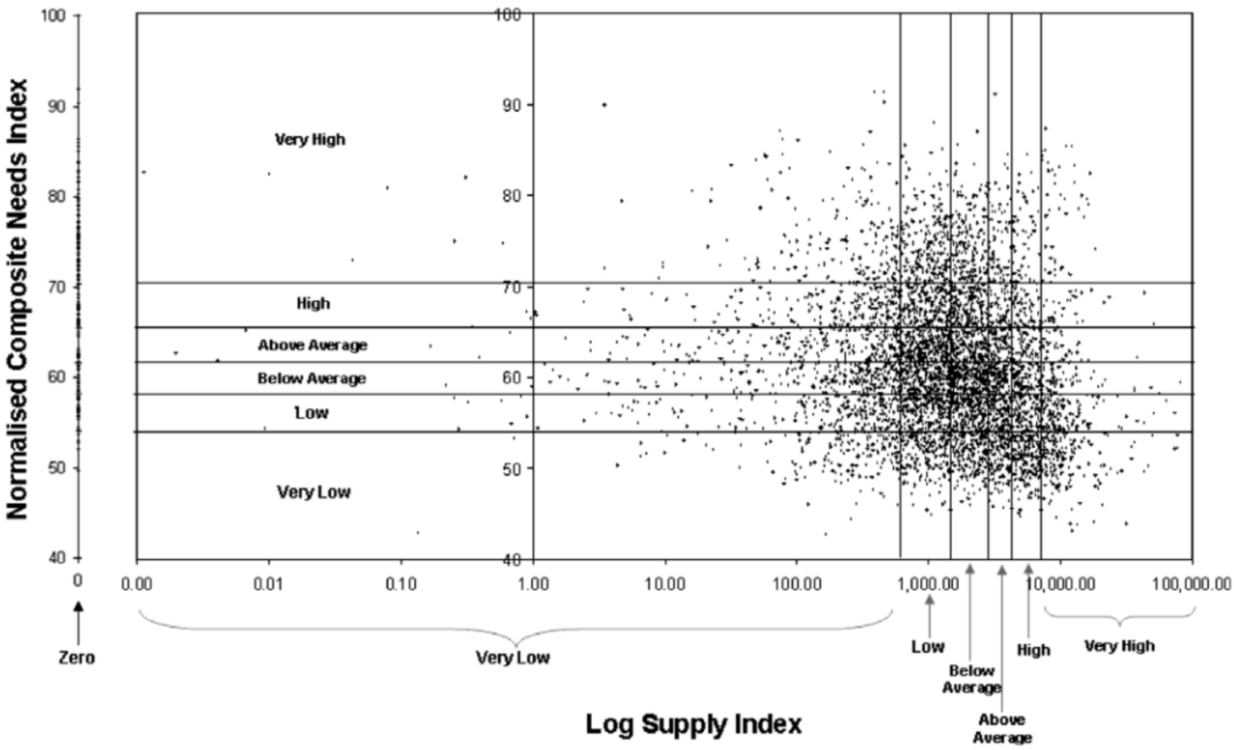

Figure 3: Log supply score and need index values – Melbourne needs-gap study, Source: Currie (2010)

3. Methodology

This study developed a package with tools for calculating the SI from GTFS data. The R programming language (R Core Team, 2023) was adopted for code development. Package development setup and workflow as described by Wickham and Bryan (2023) was adopted. Various existing packages were relied upon including: the sf package (Pebesma, 2023) for geospatial analysis; the tidyverse (Wickham et al., 2019); gtfstools (Herszenhut et al., 2022); and tidytransit (Poletti et al., 2023). Some code was adapted from examples, vignettes and other documentation in the tidytransit, gtfstools and other packages.

Two cases where used during the code development and testing, such that results might be generated for real GTFS data: the Mornington Peninsula Tourist Railway GTFS feed and the Public Transport Victoria (PTV) GTFS feed, both in Victoria, Australia. Both were selected primarily for convenience, given that the authors are familiar with the typical service patterns and geography. The Mornington Peninsula Tourist Railway network, consisting of only three stations, also facilitated hand calculation of the SI as a cross-check of the results produced by the developed package.

Figure @ref(Melbourne_map)) shows the areas of interest relevant to the code development and testing, and selected railway stations. Statistical Area (SA) zones from the Australian Bureau of Statistics (Australian Bureau of Statistics, undated) Areas of interest included Greater Melbourne (main) and SA1 zones within 800 metres of the Mornington Penninsula railway (right). SA1 zones are the smallest geographical areas for which results are reported in the Australian census, while the main image of Figure @ref(Melbourne_map)) shows the boundary of the Greater Melbourne Greater Capital City (GCC) zone and SA3 zone boundaries, which are generally similar to Local Government Area (LGA) boundaries (albeit with some LGAs split into two zones).

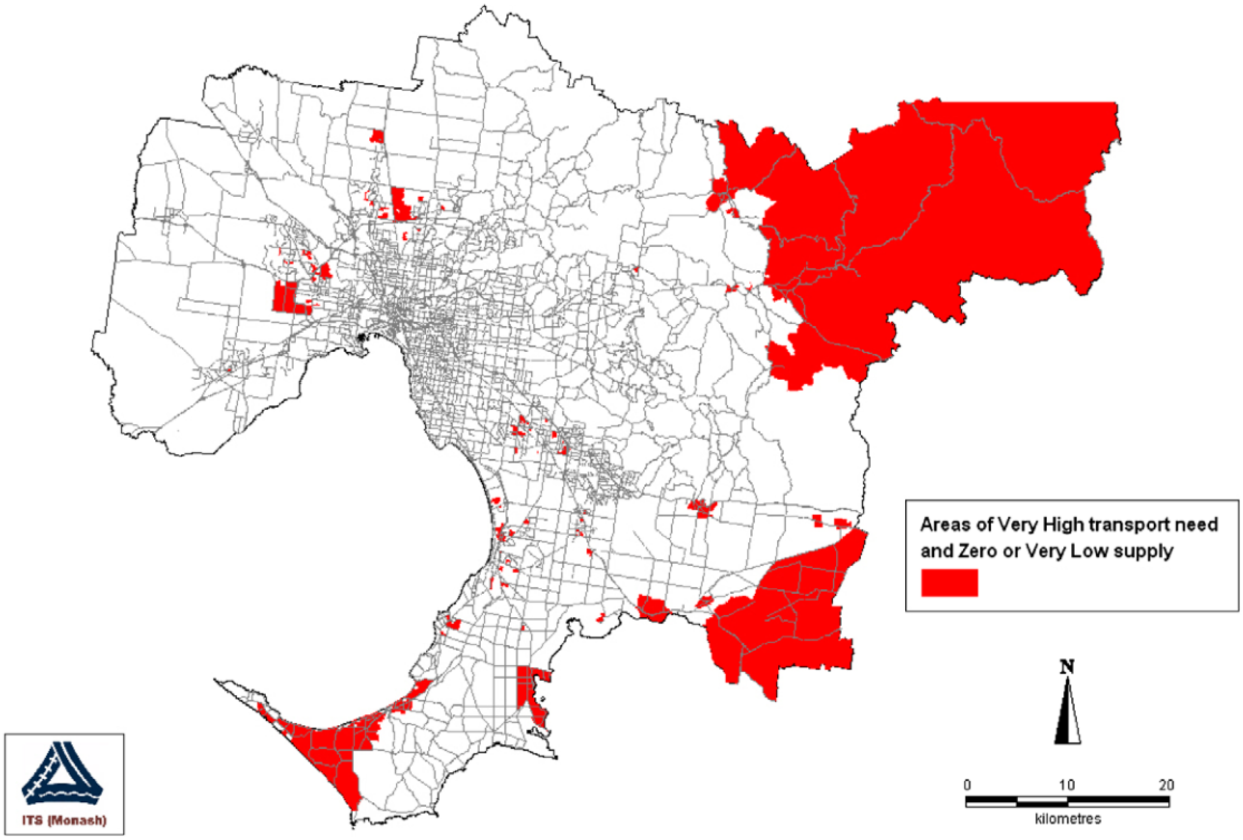

Figure 4: Melbourne needs-gap – very high transport need areas with zero or very low public transport supply, Source: Currie (2010)

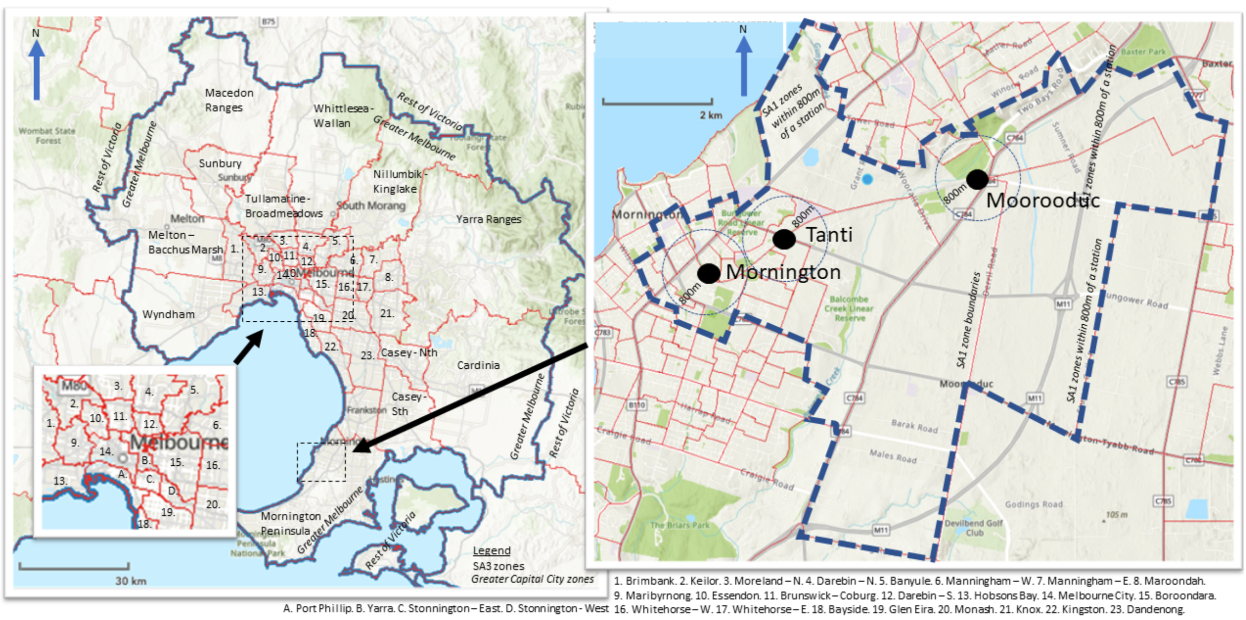

Figure 5: Areas of interest

3.1. *Mornington Penninsula Tourist Railway*

The Morning Peninsula Tourist Railway is in the outer south-east of Melbourne, running on Sundays and Wednesdays between Mornington and Moorooduc, with an intermediate stop at Tanti Park (see https://transitfeeds.com/p/mornington-railway/806/latest/stops). A GTFS feed from 2018 was selected for the purposes of tests and demonstrating the code and output. Australian Bureau of Statistics (ABS) data was also used, sources via the strayr and absmapsdata packages (Mackey et al., 2023). The Mornington Peninsular Statistical Area 3 (SA3) zone and the Statistical Area 1 (SA1) zones contained within it were adopted as the areas of interest.

3.2. *Public Transport Victoria (PTV)*

The Victorian GTFS feed, published by Public Transport Victoria (PTV) and with historical feeds sourced via Transit Mobility Data, (2023), was used for analysis of Victoria. SI scores were obtained for the weeks starting on the day of the census in 2016 and 2021, which were on Tuesday 9th and 10th of August respectively.

4. Results

4.1. *Code structure and functionality*

Developed code is available and documented on github (Reynolds, 2024). The structure of the package, functions developed, and data tables are shown in Figure @ref(fig:SI_ERD). This shows how the package takes input from three files: a gtfs feed (gtfs.zip); a sf object describing the geometry of the areas for which the SI is to be calculated; and a csv file (included in the package) defining the buffer zone distances for each route type. The ultimate output is a si_by_area_and_hour table (bottom-right), which reports the SI score for each hour of the day across dates specified by the user.

Various functions and their output are explained in the following, using the Mornington Peninsula GTFS for December 30th, 2018, and SA1 zone boundaries as a worked example. Individual steps are:

(1) loading the gtfs.zip file: the gtfs_by_route_type function loads the gtfs data and splits it into a list (by route_type) of tidygtfs objects, using the filter_by_route_type function from the gtfstools package (Herszenhut et al., undated).
(2) loading geometry information about the areas of interest: geographical data about the areas of interest are loaded by the load_areas_of_interest.R function into an sf object, using the sf package (Pebesma, 2023). The resultant areas_of_interest table contains each area_id and its associated geometry. Data about buffer zones, specifically the walking distance threshold assigned to each route_type (mode) is then loaded, again through a function (load_buffer_zone.R).
(3) calculating which stops are within the catchment walking distance of which areas: using the stops_in_walk_dist function. Figure @ref(fig:calculate_stop_in_or_near_areas_verbose)) shows how this function identified SA1 areas within the 800 metre catchment of the three Mornington stations.

(4) Calculating SI scores for a given time period: The si_calc.R function calculates the number of arrivals in a given time period, using code adapted from an article included in the tidytransit package (Poletti, undated), and combines this with the calculated area components. The si_total.R and hourly.R functions provided aggregation, giving the results mapped in Figure @ref(fig:SI_mornington_20181230_output).

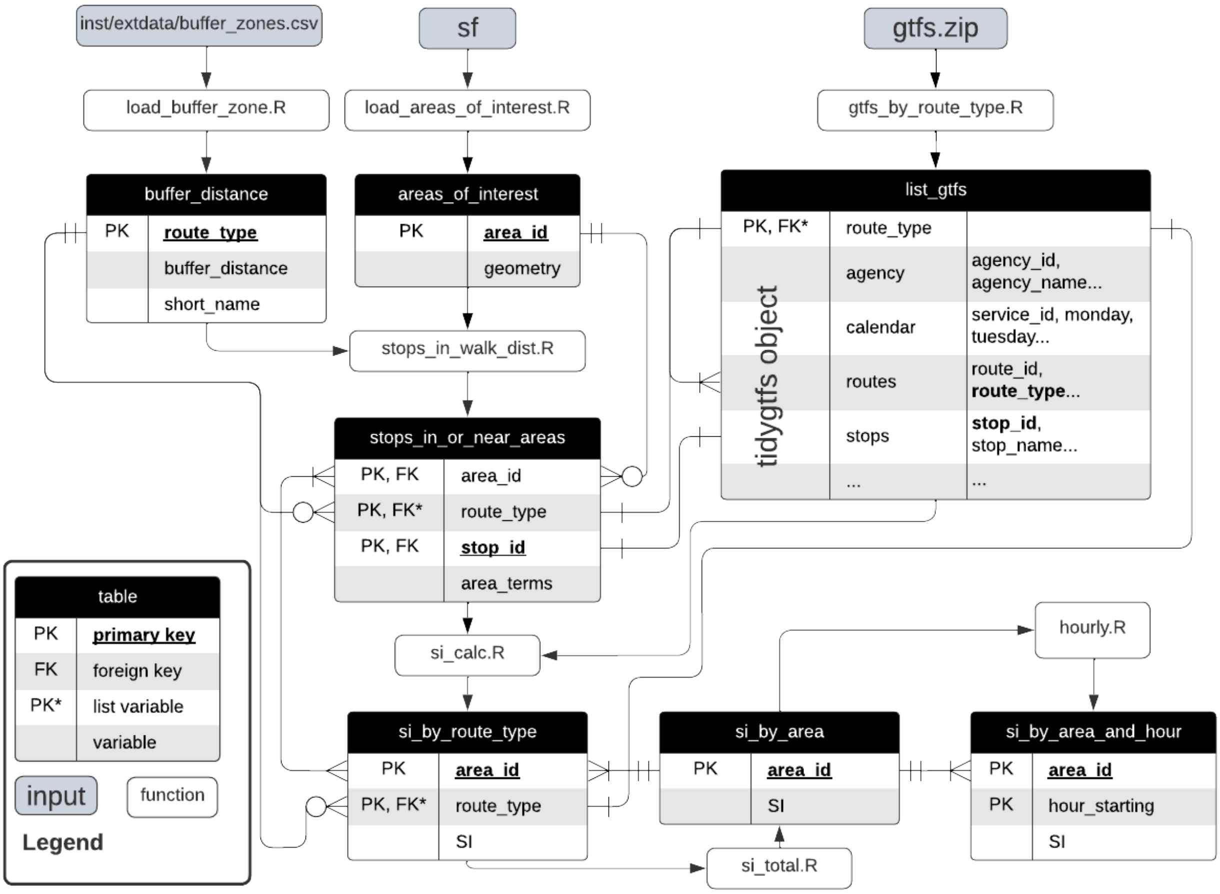

Figure 6: Entity Relationship Diagram (ERD) showing the data structure and functions related to the gtfssupplyindex package

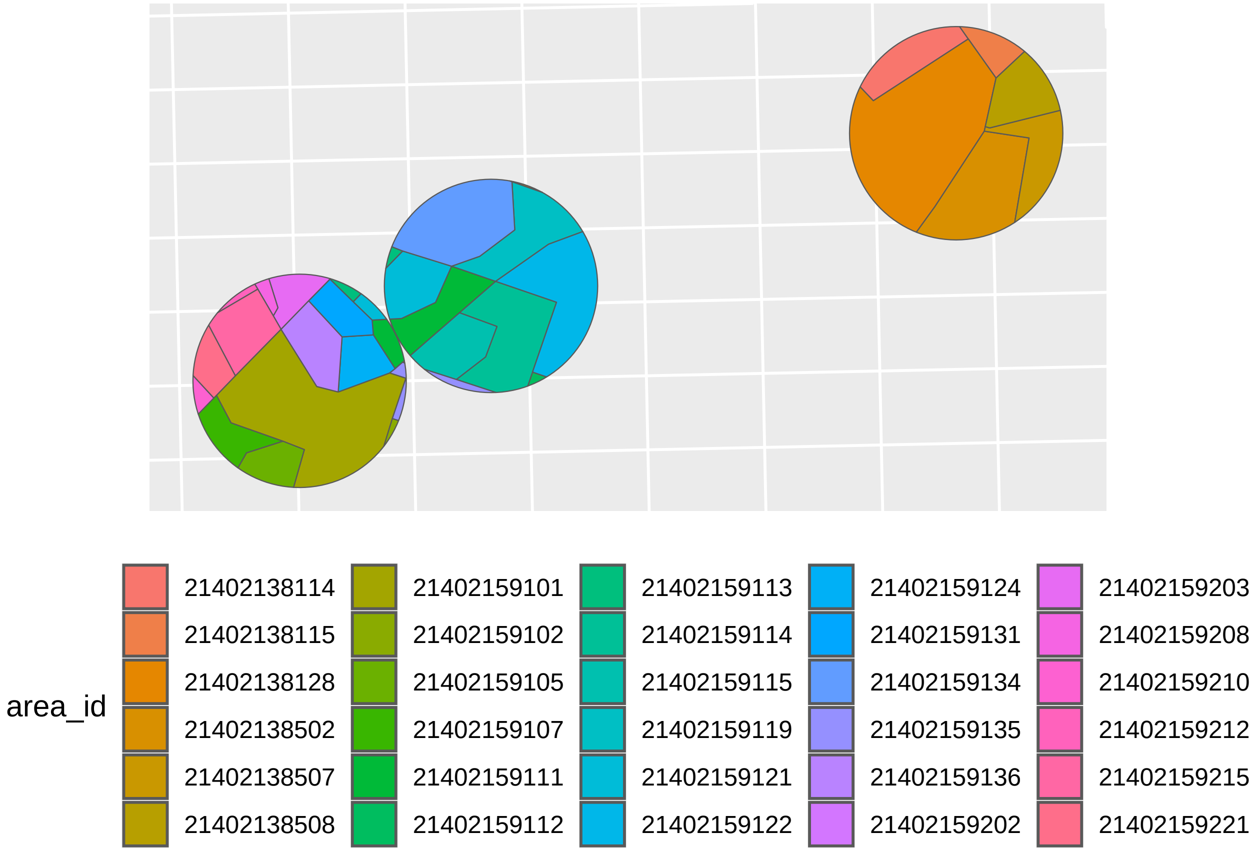

Figure 7: Step 3, stop catchments for the Mornington Penninsula Tourist Railway, showing intersections with SA1 zones

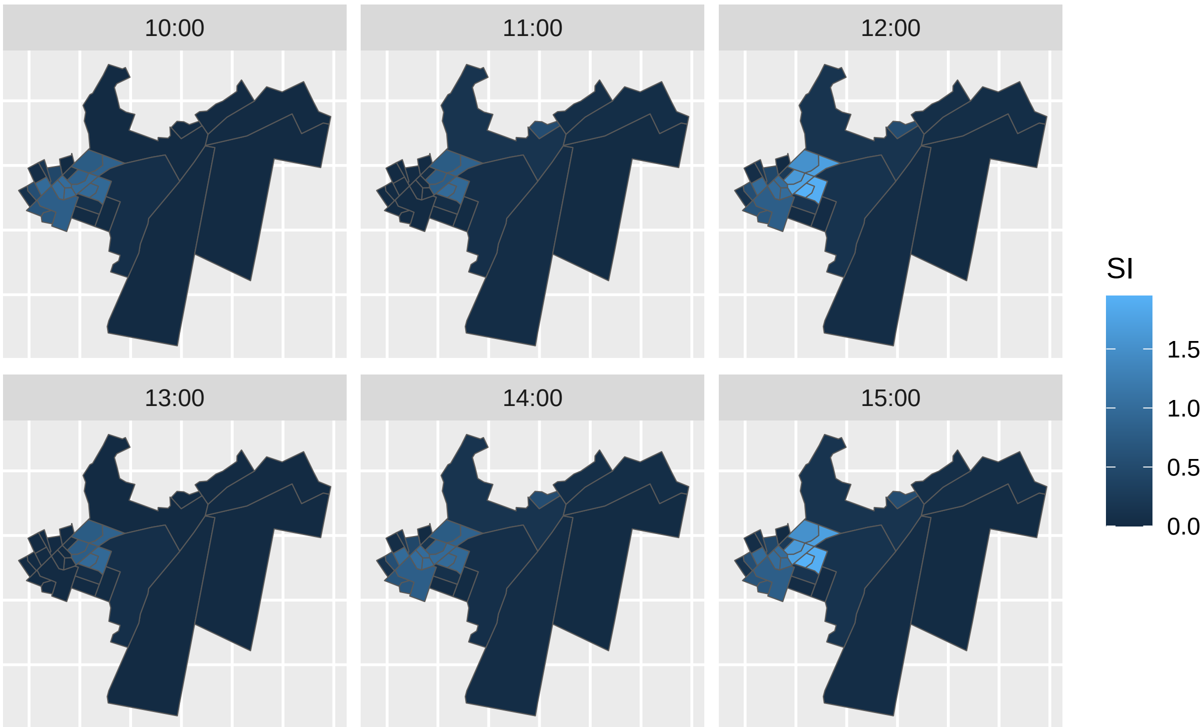

Figure 8: Mornington Penninsula Tourist Railway hourly SI values for December 30, 2018

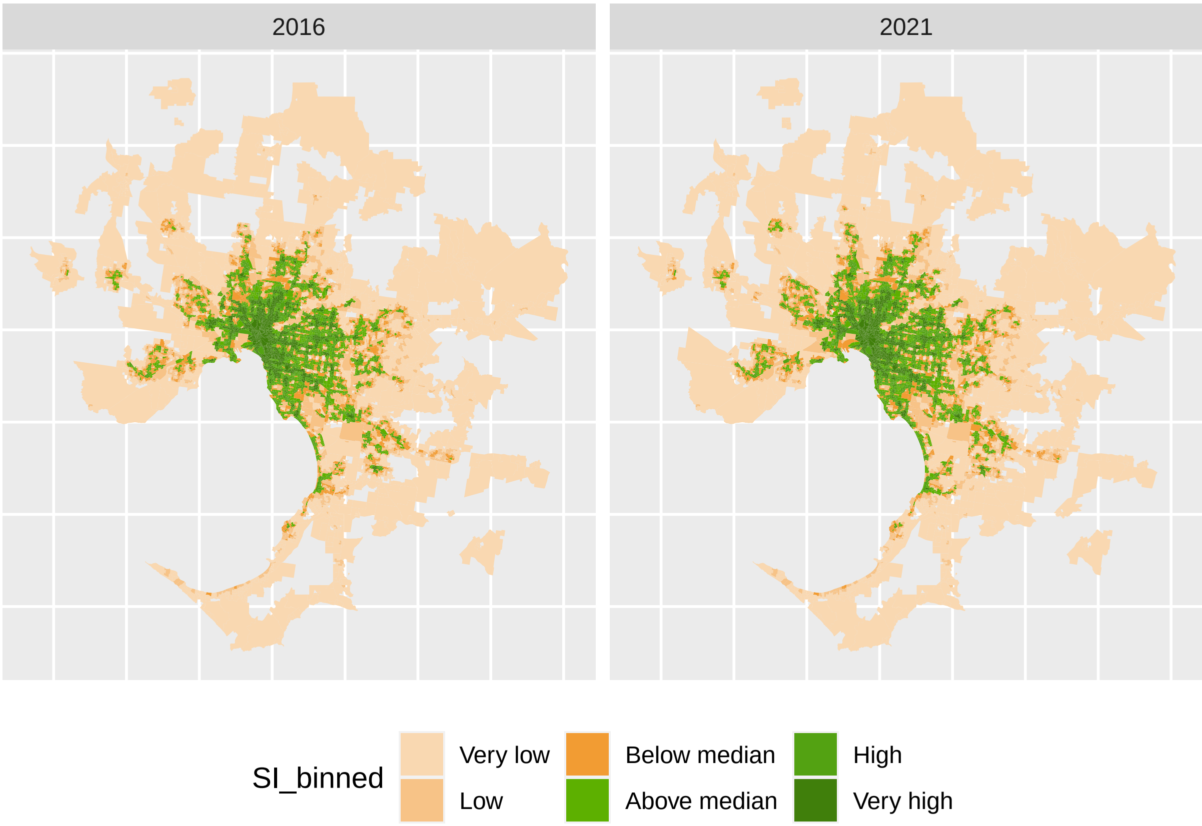

Figure 9: SI scores, census day 2016 and 2021

4.2. SI scores

4.2.1. IMRAD

4.3. Comparing cases

4.3.1. Population and equality

4.4. Purpose of transit in the city's transport policy

4.5. Indexes and comparing cities

5. Discussion

5.1. Limitations and directions for furture research

6. Conclusions

References

Sultan Alamri, Kiki Adhinugraha, Nasser Allheeib, and David Taniar. Gis analysis of adequate accessibility to public transportation in metropolitan areas. *ISPRS international journal of geo-information*, 12(5):180, 2023. ISSN 2220-9964.

Australian Bureau of Statistics. Abs maps, undated. URL `https://maps.abs.gov.au/`.

Felix Creutzig, Aneeque Javaid, Zakia Soomauroo, Steffen Lohrey, Nikola Milojevic-Dupont, Anjali Ramakrishnan, Mahendra Sethi, Lijing Liu, Leila Niamir, Christopher Bren d'Amour, Ulf Weddige, Dominic Lenzi, Martin Kowarsch, Luisa Arndt, Lulzim Baumann, Jody Betzien, Lesly Fonkwa, Bettina Huber, Ernesto Mendez, Alexandra Misiou, Cameron Pearce, Paula Radman, Paul Skaloud, and J. Marco Zausch. Fair street space allocation: ethical principles and empirical insights. *Transport Reviews*, 40(6):711–733, 2020. doi: 10.1080/01441647.2020.1762795. URL `https://doi.org/10.1080/01441647.2020.1762795`.

Graham Currie. Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, 18 (1):31–41, 2010. ISSN 0966-6923. doi: https://doi.org/10.1016/j.jtrangeo.2008.12.002. URL `https://www.sciencedirect.com/science/article/pii/S0966692308001518`.

Gordon J Fielding. *Managing public transit strategically: a comprehensive approach to strengthening service and monitoring performance.* Jossey-Bass public administration series. Jossey-Bass Publishers, San Francisco, 1st ed. edition, 1987. ISBN 1555420680.

Florida Transit Information System. Urban integrated national transit database, 2018. URL `http://www.ftis.org/urban_intd.aspx`.

Luis A. Guzman, Daniel Oviedo, and Carlos Rivera. Assessing equity in transport accessibility to work and study: The bogotá region. *Journal of transport geography*, 58:236–246, 2017. ISSN 0966-6923.

Daniel Herszenhut, Rafael H. M. Pereira, Pedro R. Andrade, and Joao Bazzo. *gtfstools: General Transit Feed Specification (GTFS) Editing and Analysing Tools*, 2022. URL `https://ipeagit.github.io/gtfstools/`. R package version 1.2.0, https://github.com/ipeaGIT/gtfstools.

Danile Herszenhut, Rafael H. M. Pereira, Pedro R. Andrade, and Joao Bazzo. *gtfstools; filter GTFS object by route type (transport mode)*, undated. URL `https://ipeagit.github.io/gtfstools/reference/filter_by_route_type.html`. R package version 1.2.0.9000, last accessed June 30, 2023.

Imperial College London. Transport strategy centre (tsc); applied research, undated. URL `https://www.imperial.ac.uk/transport-engineering/transport-strategy-centre/applied-research/`.

International Association of Public Transport (UITP). Mobility in cities database 2015, 2015. URL `uitp.org/publications/mobility-in-cities-database/`.

Todd Litman. Measuring transportation: traffic, mobility and accessibility. Technical Report 10, Institute of Transportation Engineers, Washington, D.C., 2003.

Todd Litman. When are bus lanes warranted? considering economic efficiency, social equity and strategic planning goals. Technical report, Victoria Transport Policy Institute, 2016. URL `http://www.vtpi.org/blw.pdf`.

Will Mackey, Matt Johnson, David Diviny, Matt Cowgill, Bryce Roney, William Lai, and Benjamin Wee. strayr, 2023. URL `https://runapp-aus.github.io/strayr/`.

MobilityData. *General Transit Feed Specification (GTFS)*, undated. URL `https://gtfs.org/`.

Edzer Pebesma. *sf: Simple Features for R*, 2023. URL `https://r-spatial.github.io/sf/`. R package version 1.0-15.

Flavio Poletti. *tidytransit: generate a departure timetable*, undated. URL `https://r-transit.github.io/tidytransit/articles/timetable.html`. R package version 1.5.0, last accessed June 22, 2023.

Flavio Poletti, Daniel Herszenhut, Mark Padgham, Tom Buckley, and Danton Noriega-Goodwin. *tidytransit: Read, Validate, Analyze, and Map GTFS Feeds*, 2023. URL `https://github.com/r-transit/tidytransit`. R package version 1.6.1.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

James Reynolds. gtfssupplyindex, 2024. URL `https://github.com/James-Reynolds/gtfssupplyindex`.

James Reynolds, Graham Currie, Geoff Rose, and Alistair Cumming. Moving beyond techno-rationalism: new models of transit priority implementation. In *Australasian Transport Research Forum 2017*, Auckland, New Zealand, 2017.

Paul Ryus, M Connor, S Corbett, A Rodenstein, L Wargelin, L Ferreira, Y Nakanishi, and K Blume. Tcrp report 88: a guidebook for developing a transit performance-measurement system. Technical report, 2003.

Transit Mobility Data,. Ptv gtfs - openmobilitydata, 2023. URL `https://transitfeeds.com/p/ptv/497`.

Walk Score. Transit score methodology. 2023. URL `https://www.walkscore.com/transit-score-methodology.shtml`.

Hadley Wickham and Jennifer Bryan. *R packages.* " O'Reilly Media, Inc.", 2023. URL `https://r-pkgs.org/`.

Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.

James Wong. Leveraging the general transit feed specification for efficient transit analysis. *Transportation Research Record*, 1 (2338):11–19, 2013. doi: 10.3141/2338-02.