

1 **Leveraging GTFS data to calculate an open-source Transit Supply Index**

5 **James Reynolds**

6 *Corresponding Author*

7 Research Fellow

8 Public Transport Research Group, Institute of Transport Studies, Department of Civil
9 Engineering, Monash University, Victoria, Australia

10 james.reynolds@monash.edu

12 **Graham Currie**

13 Professor

14 Public Transport Research Group, Institute of Transport Studies, Department of Civil
15 Engineering, Monash University, Victoria, Australia

16 graham.currie@monash.edu

21 Word Count: 684 words + 0 table(s) \times 250 = 684 words

23 Submission Date: July 2, 2023

1 **ABSTRACT**

2 TBC.

3

4 *Keywords:* Transit, Public transport, GTFS, Benchmarking,

1 INTRODUCTION

2 Transit service level indicators include those in the Transit Capacity and Quality of Service Manual
3 (TCQSM) (1), the Transit Score metric and many more. Practitioners, researchers and advocates
4 seeking to use such metrics may face two inter-related challenges: firstly, there is the problem of
5 calculating the metrics themselves for a specific location; secondly, is the challenge of explain-
6 ing the metrics, their meaning and importance those who are not specialists in transit, such as
7 politicians, other decision-makers or the general public.

8 The TCQSM specifies Levels of Service (LOS) between A and F across a range of fac-
9 tors including service span, frequency, speed, and the proportion of population serviced. Previous
10 research by Wong (2) overcame some challenges of using the TCQSM, by using Python, Post-
11 greSQL and R software and GTFS feeds as input to automate the calculation of daily average
12 headways, route length and stop numbers. This indicators, however, are route based and so do not
13 include any consideration of geographic or population coverage. Further metrics addressing these
14 topics and much detail about their calculation and meaning are included in the TCQSM, such as
15 the Service Coverage Area (pp. 5-8 to 5-21). However, these appear highly detailed, may required
16 bespoke GIS or other analysis, and it might be challenging to explain these measures (beyond the
17 fact at A is good and F is bad) to non-technical decision-makers, stakeholders or others who might
18 be involved. Transit Score provides a similarly easily understood rating scale, scoring locations
19 out of 100 (3). However, the algorithm is patented and effectively a black box, meaning that it
20 is not possible to calculate scores independently to understand how the metric might change with
21 cahnges to the transit system or surrounding environment.

22 The Supply Index developed by Currie and Senbergs (4) may provide a metric that is rel-
23 atively easy to calculate, open (rather than a black box), and relatively simple for a non-technical
24 audience to understand, engage with and use. This Index is based on calculating the number of
25 transit arrivals at stops within an area of interest, with an adjustment made for the amount of the
26 area of interest that is within a typical walk access distance of each stop. However, it does not ap-
27 pear to have been widely used, perhaps in part because it still required an analyst to obtain sources
28 of timetable and geographic data. Since the publication of Currie and Senbergs (4) such data has
29 become much easier to obtain with more than 10,000 agencies now providing timetable and net-
30 work data using the General Transit Feed Specification (GTFS) format (5). A gap, however, is that
31 there is not yet a method for calculating the Currie and Senbergs (4) Supply Index directly from
32 GTFS data.

33 This paper reports the development of R code to calculate the Supply Index of Currie and
34 Senbergs (4) directly from GTFS data. The code is developed using data from a single case: the
35 GTFS for Victoria in Australia, which includes Greater Melbourne. Cross-case comparison to
36 Toronto, Canada, and Washington DC, USA, is also undertaken to test the results and gain un-
37 derstanding of how the Supply Index might be useful for practitioners, researchers and advocates.
38 The motivation for this research is to better understand how GTFS data might be used to produce
39 benchmarking metrics that can be calculated using open-source code, that can be used to access
40 proposed network changes and which may be relatively easy for non-technical specialists to un-
41 derstand.

42 RESEARCH CONTEXT

43 Even a brief search shows that there is a very large number of metrics available for benchmark-
44 ing transit services, for example: the Transit Cooperative Research Program (TCRP) Report 88

provides an extensive guidebook on developing a performance-measurement system (6); online databases are provided by the Florida Transit Information System (FTIS) (7) and the International Association of Public Transport (UITP) (8) have online databases, while the Transport Strategy Centre of Imperial College London runs extensive annual benchmarking programmes across over 100 transit providers around the world (9). The Fielding Triangle (10) provides a framework for understanding how such metrics combine service inputs, service outputs and service consumption to describe cost efficiency, cost effectiveness or service effectiveness measures. At a larger scale, Litman (11) and Litman (12) discuss some of the traffic, mobility, accessibility, social equity, strategic planning and other rational decision-making frames that might underlie such transit metrics, while Reynolds et al. (13) extends this into models of how institutionalism, incrementalism and other public policy models might apply to decision-making processes. Further examples are provided by Guzman et al. (14), who develop a measure of accessibility in the context of policy development and social equity for Latin American Bus Rapid Transit (BRT) based networks, and the street space allocation metrics based around 10 ethical principles from Creutzig et al. (15).

However, many of these metrics are difficult to calculate, complex to explain or understand, and likely not well suited to communication with those who are not transit planners or engineers, or otherwise technical specialists. However, where pre-calculated metrics are immediately available it may not be possible to generate metrics for proposed system changes or know exactly how scores are calculated. The TCQSM and Transit Score may provide contrasting examples: with respect to the first challenge, TCQSM metrics may require large amounts of network, service, population and other data to be assembled before the indicators can be calculated; whereas Transit Scores are readily available on the Walk Score (3) website for locations with a published GTFS feed (eliminating the need for any calculations). With respect to the second challenge, the meaning of the Transit Score appears easy to explain (the closer to 100, the better), but as the score is calculated by a patented algorithm (effectively a black-box) it may not be easy to understand or explain the connection between real-world conditions and the score, or what might need to be done to improve the score and service levels. Nor does it appear to be possible for Transit Scores to be generated for proposed changes to networks. The TCQSM, in contrast is open-source, in that Kittleson & Associates et al. (1) provide a manual describing all the metrics and how to calculate them. However, the calculations themselves appear to be complex, which may be a barrier to use by practitioners, researchers, advocates or others who are not transit scheduling specialists. While Wong (2) provides open-source code (https://github.com/jcwong86/GTFS_Explore_Tool) this is 11 years old and does not appear to be currently maintained. Future research may involve reviewing this code and using it to analyse modern GTFS feeds. However, in this paper the aim is more modest, in that the objective is to develop code to calculate the simpler Supply Index metric from Currie and Senbergs (4).

The Supply Index

The Supply Index is shown in Equation 1, where SI_{Area} is the Supply Index for the area of interest, $Area_{Bn}$ is the buffer area for each stop within the area of interest (in Currie and Senbergs (4) taken as 400 metres for bus and tram stops, and 800 metres for railway stations), $Area_{area}$ is the area of the area of interest, and SL_{Bn} is the number of transit arrivals for each stop for a given time period (in Currie and Senbergs (4) taken as one week).

$$SI_{area} = \sum \frac{Area_{Bn}}{Area_{area}} * SL_{Bn} \quad (1)$$

The SI_{area} metric describes the number of transit arrivals at stops within an area of interest, multiplied by a factor accounting for the proportion of the area of interest that is within typical walking distance of each stop. An advantage of the Supply Index is that it is a relatively simple number to calculate, understand and explain. It is also additive, in that SI_{area} scores can be aggregated to calculate an overall score for a region encompassing multiple areas of interest.

Currie and Senbergs (4) calculated the SI for various CCDs in Melbourne using a timetable database provided by the Victorian Public Transport Authority (PTA). This predated the widespread availability of GTFS data, which provides a standardised format for timetable data that is produced by many transit systems. GTFS is an open, text-based format that was developed originally to allow transit information to be included in the Google Maps navigation platform (5). A question, therefore, is how to calculate the SI using GTFS data so that SI_{areas} can be calculated and compared for any area of interest where transit service information is available in the GTFS format.

METHODOLOGY

Various analysis tools are available that make use of GTFS data, including the tidytransit package (?) for the R statistical programming language (16). Poletti (17) provides code to calculate a departure timetable from a GTFS feed, and this was adapted to calculate arrivals at a stop and the SL_{Bn} term. The code is available at https://github.com/James-Reynolds/Transit_Supply_Index_GTFS (and included in the Rmarkdown file used to typeset this paper).

The gtfstools R package (18) was used to split input GTFS feeds by mode to facilitate the buffer zone calculation. The extended GTFS mode definitions based on Herszenhut et al. (19) were used with default buffer zones of 400 metres (bus and Light Rail Transit (LRT)) and 800 metres (rail, monorail, air etc.)

RESULTS

R code

Case studies

Melbourne

Overall

Wheelchair accessibility?

The metro tunnel - adding services

Toronto

Washington DC

Cross-case comparison

DISCUSSION

CONCLUSIONS

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: A. Anonymous, D. Zoolander; data collection: B. Security; analysis and interpretation of results: A. Anonymous, B. Security; draft manuscript preparation: A. Anonymous. All authors reviewed the results and approved the final version of the manuscript.

1 **ACKNOWLEDGEMENTS**

2 This document was prepared using the `rticles` template, created by Gregory Macfarlane, which
3 is based on the \LaTeX originally posted by David Pritchard in 2009 and updated it in 2011, soon
4 after TRB began allowing PDF submissions. Gregory Macfarlane and Ross Wang made adjust-
5 ments to the template, and Ross Wang now maintains the \LaTeX template at [https://github.com/](https://github.com/chiehrosswang/TRB-LaTeX_tex)
6 [chiehrosswang/TRB-LaTeX_tex](https://github.com/chiehrosswang/TRB-LaTeX_tex). Gregory Macfarlane created the `rticles` template in 2021.

References

1. Kittleson & Associates, Parsons Brinckerhoff, KFH Group, Texas A&M Transportation Institute, and ARUP. *Transit Capacity and Quality of Service Manual, Third Edition*. Transportation Research Board, Washington DC, third edition, tcrp report 165 edition, 2013. URL <http://www.trb.org/Main/Blurbs/169437.aspx>.
2. 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.
3. Walk Score. Transit score methodology. 2023. URL <https://www.walkscore.com/transit-score-methodology.shtml>.
4. Graham Currie and Zed Senbergs. Identifying spatial gaps in public transport provision for socially disadvantaged australians: the melbourne needs-gap study. In *Australasian Transport Research Forum*. Australasian Transport Research Forum, 2007.
5. MobilityData. *General Transit Feed Specification (GTFS)*, undated. URL <https://gtfs.org/>.
6. 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.
7. Florida Transit Information System. Urban integrated national transit database, 2018. URL http://www.ftis.org/urban_intd.aspx.
8. International Association of Public Transport (UITP). Mobility in cities database 2015, 2015. URL uitp.org/publications/mobility-in-cities-database/.
9. Imperial College London. Transport strategy centre (tsc); applied research, undated. URL <https://www.imperial.ac.uk/transport-engineering/transport-strategy-centre/applied-research/>.
10. 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.
11. Todd Litman. Measuring transportation: traffic, mobility and accessibility. Technical Report 10, Institute of Transportation Engineers, Washington, D.C., 2003.
12. 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>.
13. 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.
14. 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.
15. Felix Creutzig, Aneeqe 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>.
16. 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/>.
17. 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.
18. 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://CRAN.R-project.org/package=gtfstools>. R package version 1.2.0.
19. 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.