Leveraging GTFS data to assess transit supply, and unmet social needs

James Reynolds and Graham Currie

2023-08-08

Introduction

Transit service level indicators include those in the Transit Capacity and Quality of Service Manual (TCQSM) [Kittleson & Associates et al., 2013], the Transit Score metric [Walk Score, 2023] and many more. Practitioners, researchers and advocates seeking to use such metrics may face two inter-related challenges: firstly, there is the problem of calculating the metrics themselves for a specific location and service pattern; secondly, is the challenge of explaining the metrics, their meaning and importance those who are not specialists in transit, such as politicians, other decision-makers or the general public.

The TCQSM specifies Levels of Service (LOS) between A and F across a range of factors[1]. This scoring scheme appears likely to help towards the second aforementioned challenge; it is relatively simple to explain that A is good and F is bad[2], and the detail within Kittleson & Associates et al. [2013] provides a resource for anyone wanting to better understand what the scores mean. However, calculation of many of TCQSM metrics may need specialised software and datasets[3] and it might be challenging to explain these measures[4] or how to improve them to non-technical decision-makers, stakeholders or others involved in transit management or advocacy. The introduction of the General Transit Feed Specification (GTFS) [MobilityData, undated] [5] and widespread release of schedule data in this format, however, has helped towards making transit metrics more broadly availabile. An example is provided in previous research by Wong [2013] that developed code to calculate of some of the TCQSM metrics[6] directly from GTFS feed data. GTFS data also underlies many online journey planning systems[7], and Transit Score, which scores locations out of 100 for transit service levels [Walk Score, 2023][8]. However, the TransitScore algorithm is patented and effectively a black box, meaning that it is not possible to calculate scores independently or understand how the metric might change with alteration to the transit system or services, or the surrounding environment. TransitScore, therefore fails the first of the aforementioned challenges; in that practitioners, researchers and advocates can only use those scores provided by Walk Score [2023], which reflect only current service patterns.

[1] Including service span, frequency, speed, the proportion of the population serviced, competitiveness of travel times to car-based travel, and many more.

[2] It also matches to the A to F LOS scoring used in many traffic capacity analysis calculation systems and software.

[3] For example, the Service Coverage Area metric in the TCQSM (pp. 5-8 to 5-21) may require GIS or other analysis, on-top of accurate data about population densities, stop locations and service schedules.

[4] Beyond a simplistic A is good and F is bad

[5] GTFS is an open, text-based format that was developed originally to allow transit information to be included in the Google Maps navigation platform [MobilityData, undated].

[6] Wong [2013] calculated daily average headways, route length and stop numbers for 50 transit operators.

[7] Notably Google Maps, from back when it was called the Google Transit Feed Specification.

[8] 100 representing the highest levels of transit service provision, and being roughly equivalent to what might be experienced in the centre of New York; but which was actually calibrated using scores for the centre of San Francisco, Chicago, Boston, Portland, and Washington, D.C. [Walk Score, 2023]

Previous research by Currie and Senbergs [2007] developed a Transit Supply Index that is relatively easy to calculate, open (rather than a black box), and relatively simple for a non-technical audience to understand, engage with and use. This Index is based on calculating the number of transit arrivals at stops within an area of interest, with an adjustment made for the amount of the area of interest that is within a typical walk-access distance of each stop. However, it does not appear to have been widely used, perhaps in part because it still required an analyst to obtain timetable and geographic data and undertake the calculation themselves. Since the publication of Currie and Senbergs [2007], however, such data has become much easier to obtain with more than 10,000 agencies now providing GTFS data [MobilityData, undated]. A gap, however, is that there is not yet a method for calculating the Currie and Senbergs [2007] Supply Index directly from GTFS data.

This paper reports the development of R code to calculate the Supply Index of Currie and Senbergs [2007] directly from GTFS data. The code is developed using data from a single case: the GTFS for Victoria in Australia, which includes Greater Melbourne. Cross-case comparisons to Toronto, Canada, and Washington DC, USA, are also undertaken to test the results and gain understanding of how the Supply Index might be useful for practitioners, researchers and advocates. The motivation for this research is to better understand how GTFS data might be used to produce benchmarking metrics that can be calculated using open-source code, that can be used to access proposed network changes, and that may be relatively easy for non-technical specialists to understand and use when making decisions about or advocating for changes to existing services.

Research context

Even a brief search shows that there is a very large number of metrics available for benchmarking transit services, for example: the Transit Cooperative Research Program (TCRP) Report 88 provides an extensive guidebook on developing a performance-measurement system [Ryus et al., 2003]; online databases are provided by the Florida Transit Information System (FTIS) [Florida Transit Information System, 2018] and the International Association of Public Transport (UITP) [International Association of Public Transport (UITP), 2015]; while the Transport Strategy Centre of Imperial College London runs extensive annual benchmarking programmes across over 100 transit provides around the world [Imperial College London, undated].

The Fielding Triangle [Fielding, 1987] 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 [2003] and Litman [2016] 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. [2017] 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. [2017], 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 introduced by Creutzig et al. [2020].

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. Where pre-calculated metrics are immediately available it may not be possible for practitioners, researchers or advocates to independently generate metrics for proposed system changes or to even know exactly how scores for the existing services levels 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[9]. With respect to the second challenge, the meaning of the Transit Score appears easy to explain[10], but as the score is calculated by a patented algorithm 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[11]. While Wong [2013] provides open-source code for calculating some TCQSM metrics [12] this is now 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, with the objective being to develop code to calculate the Suppy Index metric from Currie and Senbergs [2007].

[9] The Walk Score [2023] website shows scores for locations with a published GTFS feed, eliminating the need for any calculations.

[10] The closer to 100, the better.

[11] In that Kittleson & Associates et al. [2013] provide a manual describing all the metrics and how to calculate them.

[12] `https://github.com/jcwong86/GTFS_Explore_Tool`

The Suppy Index

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

Equation 1[13] shows the Supply Index[14]. An advantage of the Supply Index is that it is a relatively simple number to calculate, understand and explain. It describes the number of transit arrivals at stops within an area of interest and time frame, multiplied by a factor accounting for the proportion of the area of interest that is within typical walking distance of each stop. Hence, more services, more stops and higher frequencies would all result in an increase in Supply Index score. The Supply Index does not incorporate further aspects, such as service span, off-peak share of service or service speed. However, including such metrics may increase the complexity of calculating and describing the index to non-transit specialists. Such simplicity is also helped by the way that the Index is additive, in that $\mathrm{SI}_{[\text{area, time}]}$ scores can be aggregated to calculate an overall score across multiple time periods or for a region encompassing multiple areas of interest.

Currie and Senbergs [2007] calculated the $\mathrm{SI}_{[\text{area, time}]}$ for various Census Collection Districts (CCDs)[15] in Melbourne using a timetable database provided by the Victorian Public Transport Authority (PTA). This predated the widespread availability of GTFS data. A question, therefore, is how to calculate the SI using GTFS data so that $\mathrm{SI}_{\text{areas}}$ can be calculated and compared for any area of interest where transit service information in that format.

[13] In Equation 1 $SI_{area,time}$ is the Supply Index for the area of interest and a given period of time. $Area_{Bn}$ is the buffer area for each stop (n) within the area of interest. In Currie and Senbergs [2007] this was based on a radius of 400 metres for bus and tram stops, and 800 metres for railway stations. $Area_area$ is the area of the area of interest, and $SL_{n,time}$ is the number of transit arrivals for each stop for a given time period.

[14] Minor adjustments have been made to generalise the equation, as Currie and Senbergs [2007] focused on the context of Melbourne's Census Collection Districts (CCD) and calculations based on a week of transit service.

[15] CCDs predate the introduction of Statistical Areas 1, 2, 3, and 4 (SA1, SA2, SA3, SA4), and other geographical divisions currently used by the Australian Bureau of Statistics (ABS), which may be more familiar to readers

Transport Needs Index(es)

Currie and Senbergs [2007] also adopted the ABS Index for Relative Socio-Economic Advantage/Disadvantage (IRSAD), and developed a Transport Needs index. This Transport Needs index was based on combining standardised indicators[16] of: adults without cars[17]; accessibility[18]; persons aged over 60 years[19]; the number of people on a disability pension[20]; the number of low income households[21]; the number of adults not in the labour force[22]; the number of students[23]; and the number of people aged 5 to 9 years[24].

Four separate indicators provided input to the Currie and Senbergs [2007] assessment, reflecting different approaches to understanding the scale of the need for transit services. These are: 1) the Total Need IRSAD Index, based on the ABS ISRAD, weighted by the size of the population in each area of interest; 2) the Relative Need IRSAD Index, based on weighting the indicators by the proportion of the area of interest population in the social need component groups; 3) the Total Transport Need Index; and 4) the Relative Transport Need Index. These four indicators were combined in Currie and Senbergs [2007], with equal weighting[25] to provide a single, unified needs index (standardised to provide a combined relative score out of 100).

[16] Set as relative scores between 0 and 100 for all areas in the analysis.

[17] Based on census data for the number of cars per household and number of people aged 18 or over, weighted at 19% of the Transport Needs Index.

[18] Being the straight line distance to the CBD, weighted at 15%.

[19] Weighted at 14%.

[20] Wrighted at 12%.

[21] Weighted at 10%.

[22] Based on the number of people over 15 not in the labour force, weigthed at 9%.

[23] Weighted at 9%.

[24] Weighted at 12%.

[25] i.e. 25% each.

Methodology

This study adopts a case research approach by developing code to calculate Supply Indexes for Melbourne (Australia), Toronto (Canada) and Washington D.C. (USA). The case research approach serves two purposes: firstly the selected cases where used to develop, test and verify the functionality of the code produced in this study. Secondly, aspects of the the selected cases were explored; such as the impact of forthcoming upgrades and issues of level-boarding access (i.e. non-step suitable for wheelchair, pram, mobility-aid etc.) to the Supply Index scores[26].

Code development

Various analysis tools are available that make us of GTFS data, including the tidytransit package [Poletti et al., 2023] for the R statistical programming language [R Core Team, 2022]. Poletti [undated] provides code to calculate a departure timetable from a GTFS feed, and this was adapted to calculate arrivals at a stop and the $\mathrm{SL_{Bn}}$ term.

The gtfstools R package [Herszenhut et al., 2022] was used to split input GTFS feeds by mode to facilitate the buffer zone calculation. Buffer zones of 400 metres for bus and Light Rail Transit (LRT) services and 800 metres for heavy rail, as per Currie and Senbergs [2007][27].

Where transit stops are located close to boundaries their catchment areas may fall into multiple areas of interest. The sp package [Pebesma, 2023] provides tools for manipulating geographic data and shape files in R. This was used to calculate the proportion of each stop's catchment area that falls into each geographical area of interest[28].

The $\mathrm{SI_{area}}$ was calculated on a mode-by-mode and stop-by-stop basis, by first determining the amount of the catchment area ($\mathrm{Area_{Bn}}$) that falls into each geographical area of interest for the stop in question. This is combined with the area for each geographical area of interest ($\mathrm{Area_{area}}$) and the number of stop arrivals within the ($\mathrm{SL_{Bn}}$) to

[26] A key issue in case-based research is the duality criterion, being the need to simultaneously: be engaged with depth of the individual cases included in the study; yet also be seeking findings that are generalisable to more than just the case that are studied. This study responds to this criterion by first testing the developed code across three different transit systems, which is expected to be sufficient to validate that it is generalisable to most GTFS feeds, not specific to the quirks of the Victorian GTFS feed. Additionally, this study engages with a range of issues (level-boarding access, new routes and services, and cancelled projects) across these three cities, again seeking to demonstrate how the Supply Index code might be adapted to a range of other issues and places.

[27] There is an extended mode definition that includes modes beyond the 10 in the GTFS standard [Herszenhut et al., undated], but these are not dealt with by the gtfstools package. Further research may seek to extend this such that other modes can be included, but for the purposes of this study the buffer zone was set at 400 metres for cable trams, aerial lifts such a gondolas and trolleybuses, and at 800 metres for ferries, funiculars and monorails.

[28] GTFS files define stop locations based on latitude and longitude [MobilityData, undated], whereas the $\mathrm{Area_{Bn}}$ calculation needs to be provided in the same units as the $\mathrm{Area_{area}}$ variable, necessitating the use of a geographic transform as part of the code.

calculate the contribution to the index scores made by just that single stop for every area of interest; these are then added to a cumulative total field for each area of interest; and the calculations are repeated until all stops and modes in the GTFS file have been included.

Case research approach

To test the developed code results were generated for a single case: Victoria, Australia. This case was selected primarily for convenience, as the authors are familiar with the Victorian transit network and so well placed to assess whether output results are correct. Results were processed using the ggmaps [Kahle et al., 2023], ggplot [Wickham et al., 2023], ggstatsplot [Patil, 2023] and kable [Zhu, 2021, Xie, 2023] packages, with data processing leveraging the tidyverse approach [Wickham, 2023].

Victoria, Australia

Victoria is the southern-most state on the Australian mainland. The state capital is in Melbourne, which has a similar metropolitan area to of Paris or London. However, with only around 5 million people Greater Melbourne has about one-third of the population density. It has an inner Central Business District (CBD) with apartments, commercial skyscrapers and extensive sporting facilities nearby; surrounded by low-density, predominately single-family-housing-dominated, inner, middle and outer suburbs.

There are train and tram networks radiating from the CBD, but for most of the suburban areas the reality is that transit is provided by circuitious bus routes that are mostly used by those who cannot otherwise drive. An extensive freeway (and tollway) network provides connections across the Greater Melbourne area, further around Port Phillip Bay to Geelong (south-west) and the Mornington Penninsula (south-east) as well as to regional centres elsewhere in Victoria. There is a state-wide regional train and bus network, which also provides connections into South Australia, New South Wales and the Australian Capital Territory (Canberra) and local bus services in many regional towns and cities. However, accessibility to most of the city and state tends to be car-dominated.

Victoria's GTFS feed is published by Public Transport Victoria (PTV)[29]. The Australian census is undertaken in early August every 5 years [With the last two being in 2016 and 2021.]. GTFS feeds were therefore selected for the first week of August of each year for the purposes of this test analysis.

[29] There are over 400 historical releases of the available on the transitfeeds.com website, with the first dating from March 2015 [Transit Mobility Data,, 2023].

Minor corrections were made to the GTFS files to remove duplicate stop_ids[30]. The Australian Bureau of Statistics (ABS) provides a range of shape files and other resources, and this study made use of the absmapsdata R package [Mackey, 2023] to access the 2021 Local Government Area (LGA) boundaries for the Greater Melbourne area. The EPSG:28355 transform [EPSG, 1995] was used to shift longitude and latitude into metres, as per the Geocentric Datum of Australia 1994 (GDA95 / MGA zone 55) coordinates.

[30] These involved minor discrepancies in either the stop name, latitude or longitude.

The ABS provides IRSAD datasets for LGAs in excel format. Data for 2016 and 2021 was included in this study, with the scores shown in Figures ?? and ??

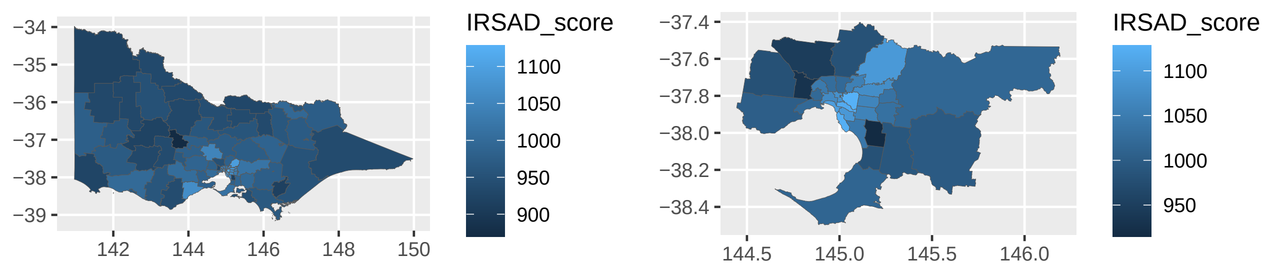


Figure 1: 2016 IRSAD for Victorian and Greater Melbourne LGAs

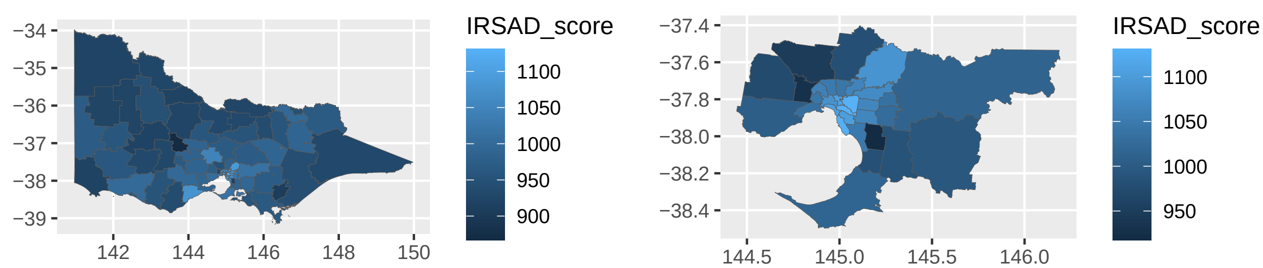


Figure 2: 2021 IRSAD for Victorian and Greater Melbourne LGAs

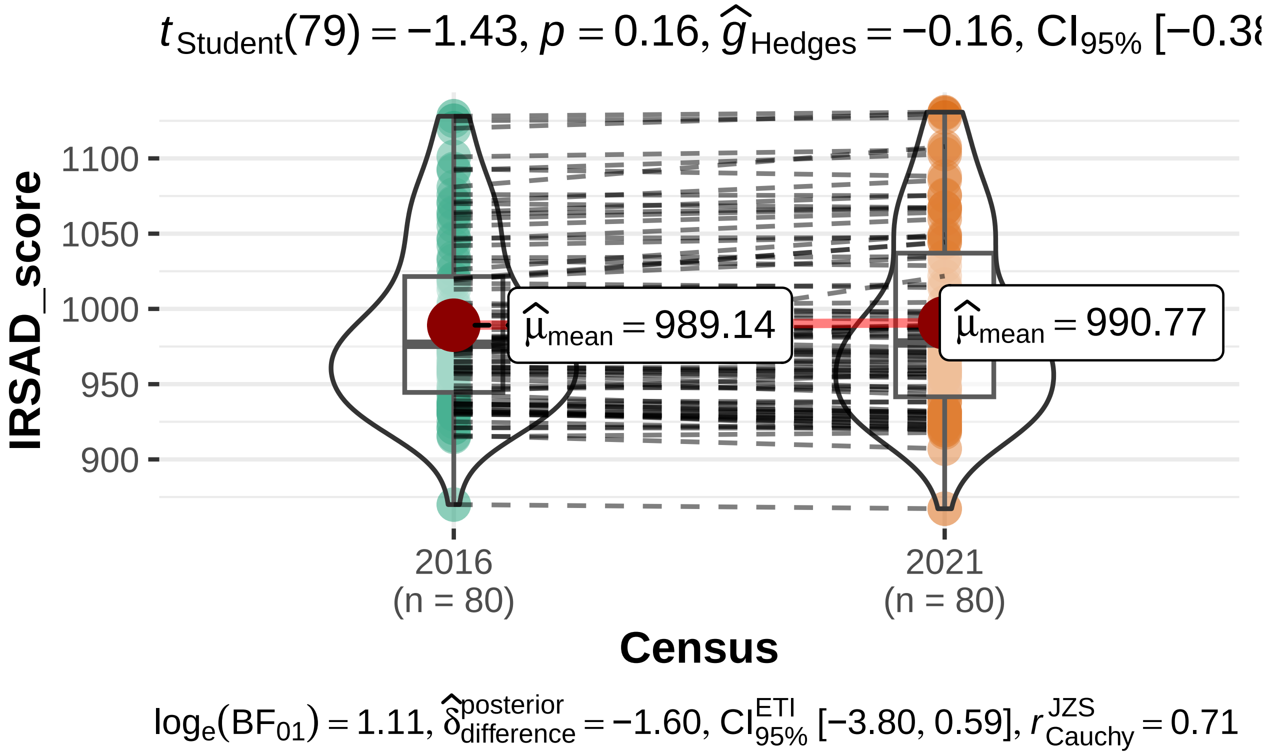


Figure 3: Comparing 2016 and 2021 IRSAD scores for Victorian LGAs

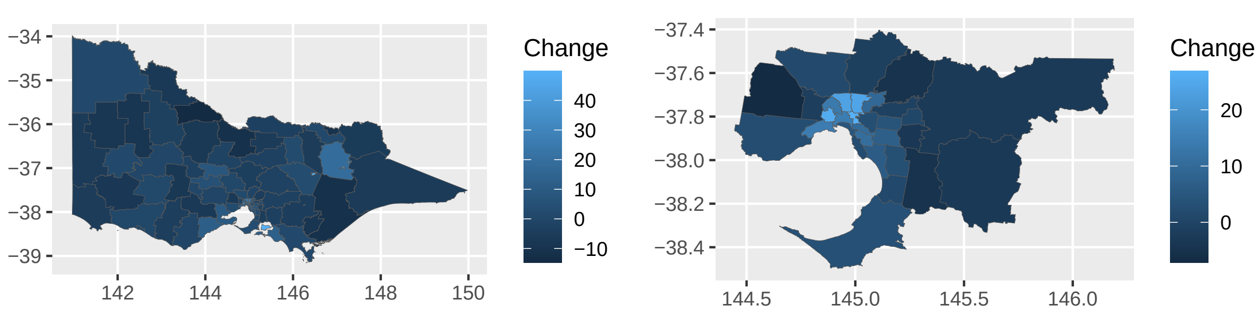


Figure 4: Change in IRSAD scores (2021 minus 2016)

Results

The following subsections discuss the results of cases studies used to develop and test the code[31].

[31] The code is available at `https://github.com/James-Reynolds/Transit_Supply_Index_GTFS` as a Rmarkdown file (used to typeset this paper).

Verifing the code output

Code output results were verified by comparison to by-hand calculations for selected combinations of date and location in Victoria.

Running Creek and Morgans Bridge, Kiewa Valley Hwy

The SA1 area 20403106915 covers Running Creek and Morgans Bridge, two localities in the Victorian Alps. Within this SA1 area there are only two V/Line bus stops^{Stop:ID 45125, Running Creek Rd/Kiewa Valley Hwy (Running Creek) and Stop ID: 45124, Kiewa Valley Hwy (Mongans Bridge).}. This SA1 was selected for the purposes of verifying the code output, as it is relatively easy to calculate the relevant SI values as a cross-check, because there is only one bus service and two stops to include. The location of the SA1 20403106915 is shown in the following figure, together with relevant geographic statistics.

```
##   sa1_code_2021 sa2_code_2021         sa2_name_2021 sa3_code_2021
## 1   20403106915     204031069 Bright - Mount Beauty         20403
##      sa3_name_2021 sa4_code_2021 sa4_name_2021 gcc_code_2021 gcc_name_2021
## 1 Wodonga - Alpine           204          Hume         2RVIC  Rest of Vic.
##   state_code_2021 state_name_2021 areasqkm_2021  cent_lat cent_long
## 1               2        Victoria       284.598 -36.57879    147.05
```

The area of SA1 20403106915 is 284.598km^2. By inspection, the entire 400m radius catchment area of both of the bus stops lie entirely within the SA1 20403106915 boundaries.

Hence the $Area_{Bn}/ Area_{SA1_Area} $ term for each of the bus stops is equal to $(\pi 400^2)/284598000 = 1.77\text{e-}03$.
No printed timetable has been located for the Albury - Mt Beauty via Baranduda and Tawonga South route, but stop times are provided on

Figure 5: SA1 20403106915, with approximate location of Stop:ID 45125 highlighted, source ABS

the PTV website[32], which indicates services on April 27 to Albury at 7:25am, 9:25am and 9:30am. For April 28 there are services at 7:25am and 9:30am, but then there are no services over the weekend. May 1 has two services (7:25am and 9:30am), May 2nd and 3rd both have three (7:25, 9:25 and 9:30am), bringing the total to 13. In the opposite direction the PTV website indicates a similar service pattern, only in the opposite direction and at 2:50pm, 4:20pm and 4:40pm[33]. This brings the total to 26 arrivals to each of the two bus stops in a seven day week [34] Therefore the total $SI_{20403106915,27/4-3/5/23}$ score is equal to $(2*(26*pi*400*400/284598000))$ which is equal to 0.091842.

[32] See `https://tinyurl.com/5n83ryhy`

[33] In general, this appears to be a school bus service pattern. It is unclear, however, why some days have three services but others have only two.

[34] The $\mathrm{SL}_{\mathrm{Bn}}$ term.

The $SI_{20403106915,27/4-3/5/23}$ score calculated by the developed code is shown in the following table.

Table 1: Developed code output for SA1 20403106915, seven days starting April 27, 2023

sa1_code_2021	sa2_name_2021	SI_CCD
20403106915	Bright - Mount Beauty	0.0918

The hand-calculated $SI_{20403106915,27/4-3/5/23}$ matches that produced by the developed code, suggesting that the developed code is providing the expected output.

Further verification tests

Further verification tests are yet to be completed. They may include selecting: a SA1 area that has multiple modes; selecting a SA1 area that has transit stops close to its boundary (such that only part of the catchment area is within the SA1 boundaries); and selecting an SA1 area where a transit stop is located outside the boundary, but the catchment area overlaps.

Some of these verification calculations appear likely to require the use of GIS software to determine how much of a catchment area lies within a SA1 boundary[35].

[35] TO BE COMPLETED.

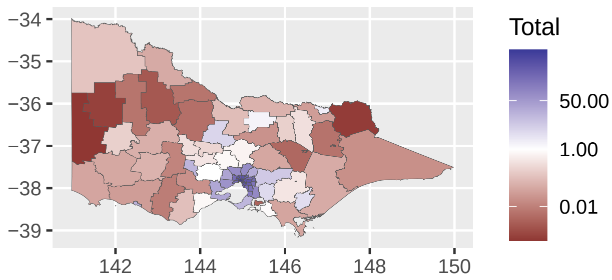
Victoria, Australia: Supply Index results for LGAs

To better understand whether the SI calculation from GTFS data can be practically applied, this section presents results when the SI is calculated for Local Government Areas(LGAs) within Victoria. For the purposes of this section, the SI_{LGA} is calculated for only a single day of service, being the date of the 2021 census[36]. Analysis of a single day was selected so as to reduce the processing time required to output the SI scores, as the code current takes about one day to run for all the stops in Victoria for every day that is analysed[37].

[36] Tuesday August 10, 2021

[37] Code performance could be improved, but appears good enough for the purposes of this research. The rmarkdown used to typeset this paper loads pre-calculated SI scores.

Figure 6 show the $SI_{LGA2021,10/8/21}$ values for Victoria (left) and Greater Melbourne (right).



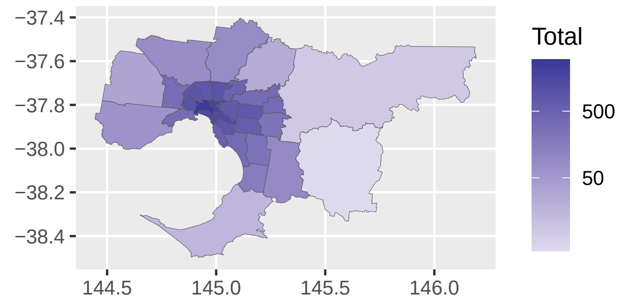


Figure 6: $SI_{LGA2021,10/8/21}$

In general, Figure 6 indicates that the amount of transit service is higher for LGAs that are closer to the centre of Melbourne.

Figure 7 compares the SI for 2021 and 2016 for each LGA. It indicates that there are significant differences in the SI scores in 2016 and 2021. However, it appears that the change has been relatively small, with the average SI increasing from 208.6 to 218.1.

Figure 8 shows that there was a significant change in SI for LGAs within Greater Melbourne, rising from 548.4 to 573.2.

There was no significant change amongst LGAs in the rest of Victoria (Figure 9).

Figure 10 maps percentage changes in SI between 2016 and 2021. These are also shown in Figure 11.

There was no significant difference in changes in SI between LGAs in Greater Melbourne or the Rest of Victoria. The SI increased by an average of 12% for LGAs in Victoria. However, there were some LGAs that saw decreases in SI between 2016 and 2021 (Table 2).

Table 2: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for those LGAs where SI has fallen, by mode

		Change in SI			
LGA	Location	LRT	Rail	Bus	Total
Hindmarsh	Elsewhere	0%	-100%	8%	-52%
East Gippsland	Elsewhere	0%	0%	-30%	-30%
West Wimmera	Elsewhere	0%	0%	-25%	-25%
Gannawarra	Elsewhere	0%	0%	-18%	-13%
Swan Hill	Elsewhere	0%	0%	-6%	-6%
Ararat	Elsewhere	0%	0%	-6%	-6%
Indigo	Elsewhere	0%	0%	-6%	-5%
Wangaratta	Elsewhere	0%	-25%	-3%	-4%
Loddon	Elsewhere	0%	0%	-5%	-3%
Warrnambool	Elsewhere	0%	33%	-1%	-1%
Melbourne	Greater Melbourne	-2%	2%	-3%	-1%
Central Goldfields	Elsewhere	0%	100%	-2%	-1%

The largest relative changes in SI were in Hindmarsh (-52%), East

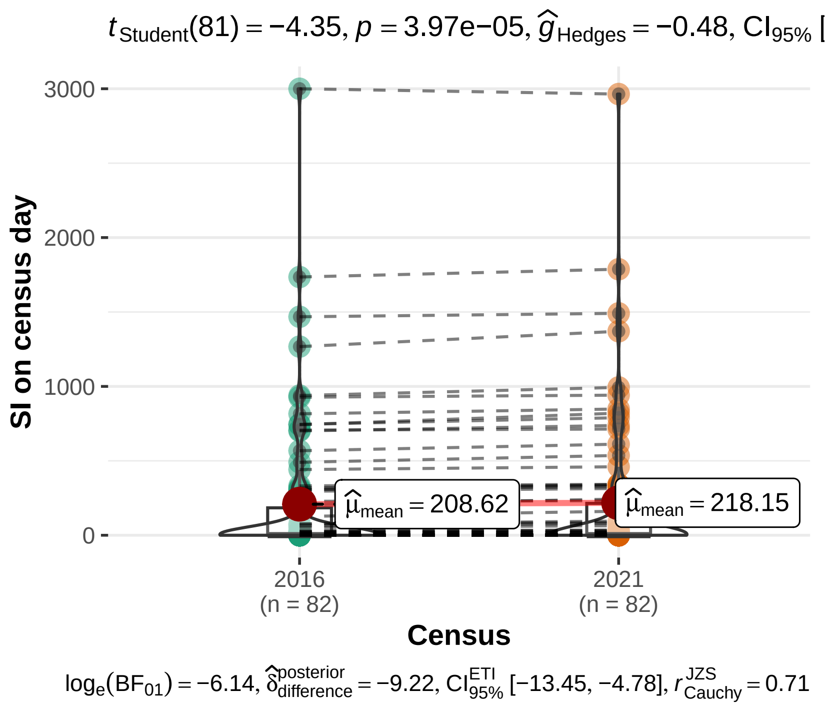


Figure 7: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ violin and box plot generated using ggstatsplot package

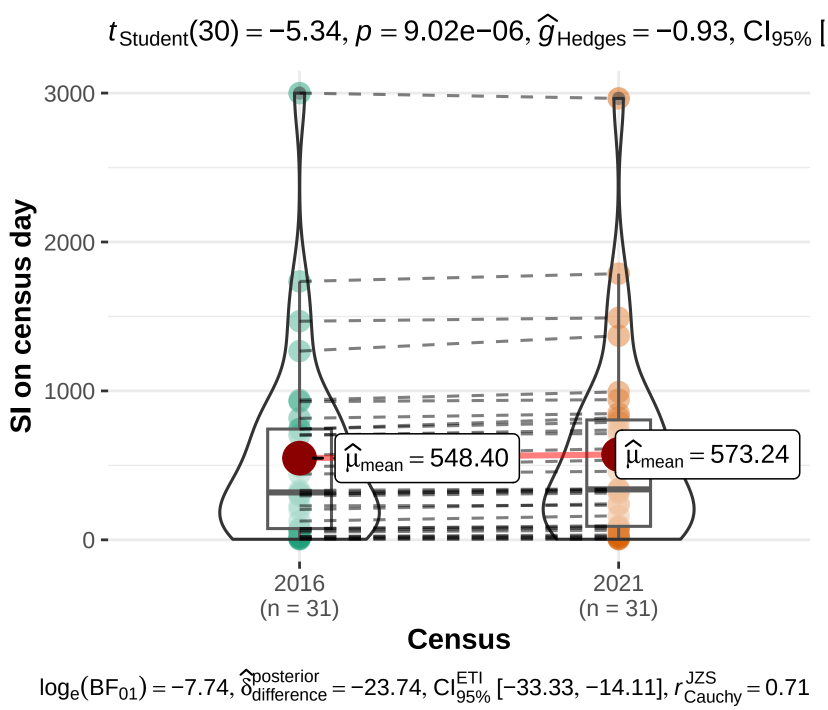


Figure 8: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ Greater Melbourne, violin and box plot generated using ggstatsplot package

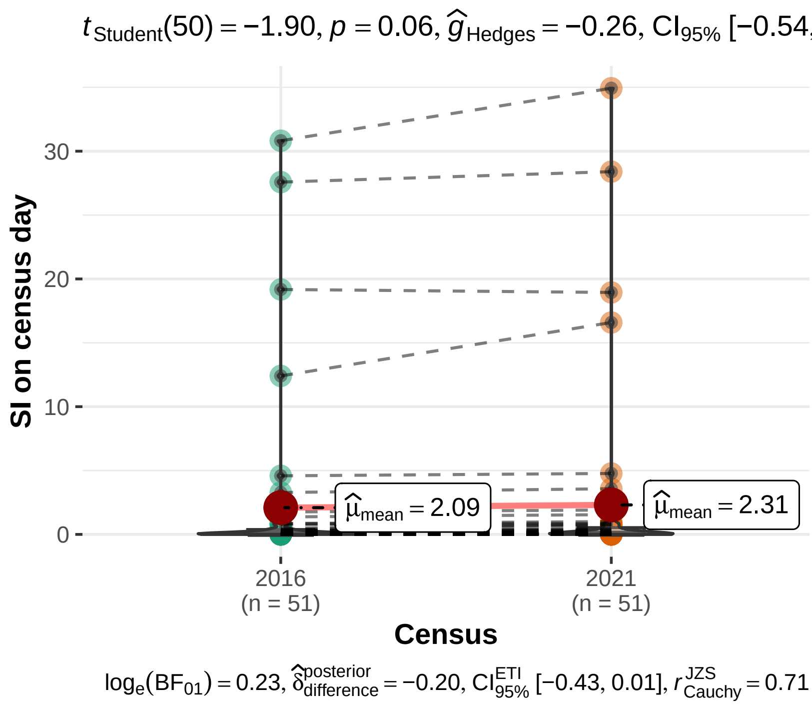


Figure 9: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ non-Greater Melbourne parts of Victoria, violin and box plot generated using ggstatsplot package

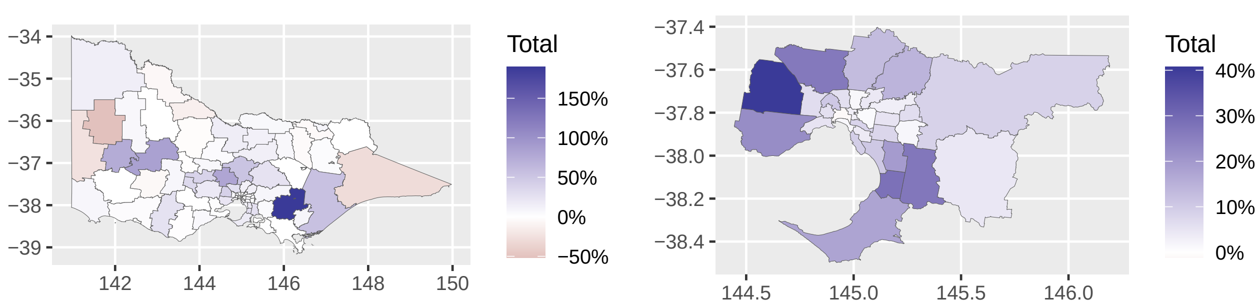


Figure 10: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$

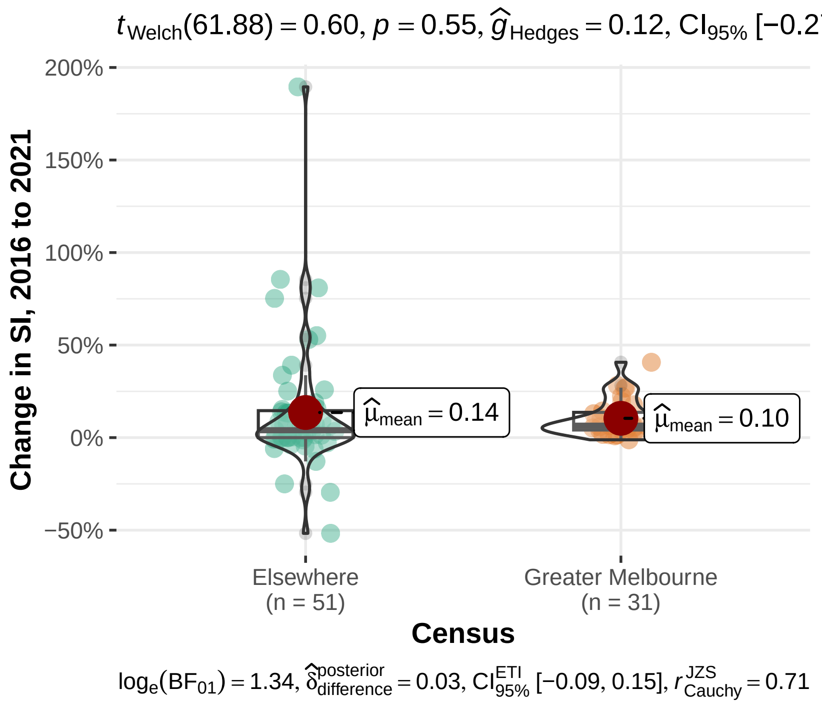


Figure 11: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ violin and box plot generated using ggstatsplot package

Gippsland (-30%) and West Wimmera (-25%). For Hindmarsh this appears to have involved the entire removal of the rail service, but offset by a small increase in bus service levels (+8%). Central Goldfields, in contrast had a doubling of train services, but an overall change of -1%. However, some LGAs had relatively large increases in SI between 2016 and 2021.

The largest relative increases in SI were in Baw Baw (190%), Northern Grampians (85%) and Macedon Ranges (81%). For Baw Baw and Macedon Ranges this appears to have involved large increases in bus service levels (+401%, 172%). Northern Grampians, lost all of its train services, but saw SI_{bus} increase by 90%. Horsham, likewise, saw the loss of all of its rail services, but an increase of 77% in SI_{bus}.

Supply Index by mode

Figure 12 shows changes in the proportion of the SI score[38] delivered by train, tram and bus across Victoria.

[38] Note, SI scores have been rounded to the nearest integer.

Figure 12 indicates there has not been a significant change. Changes were also not statistically significant when including only LGAs in Greater Melbourne or only LGAs elsewhere in Victoria[39].

[39] Not shown

There were small but significant change in SI_{rail} and SI_{bus} scores

LGA	Location	Change in SI			
		LRT	Rail	Bus	Total
Baw Baw	Elsewhere	0%	15%	401%	190%
Northern Grampians	Elsewhere	0%	-100%	90%	85%
Macedon Ranges	Elsewhere	0%	25%	172%	81%
Horsham	Elsewhere	0%	-100%	77%	75%
Wellington	Elsewhere	0%	0%	61%	55%
Mitchell	Elsewhere	0%	-2%	85%	53%
Melton	Greater Melbourne	0%	147%	38%	41%
Hepburn	Elsewhere	0%	94%	38%	39%
Ballarat	Elsewhere	0%	55%	33%	34%
Frankston	Greater Melbourne	0%	11%	29%	28%
Casey	Greater Melbourne	0%	7%	28%	27%
Hume	Greater Melbourne	0%	12%	28%	27%
Corangamite	Elsewhere	0%	33%	0%	26%
Murrindindi	Elsewhere	0%	0%	25%	25%

Table 3: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for those LGAs where SI has risen 25 percent or more

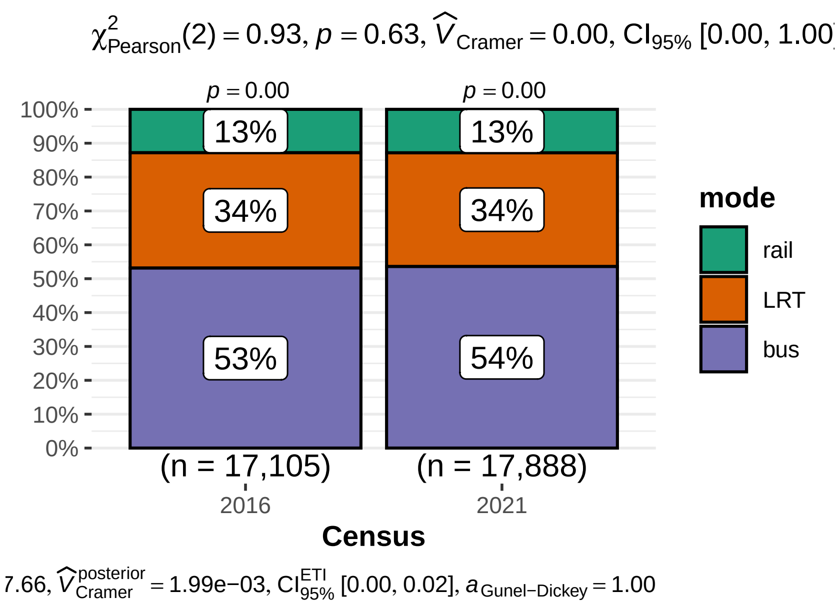


Figure 12: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ proportion by mode, bar chart generated by ggstatsplot package

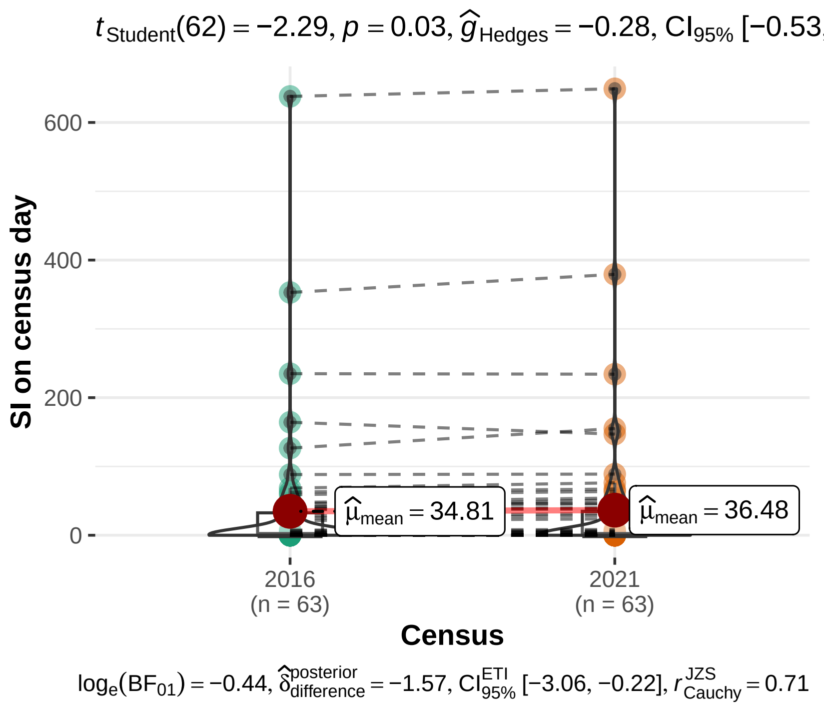


Figure 13: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ rail mode only, violin and box plot generated using ggstatsplot package

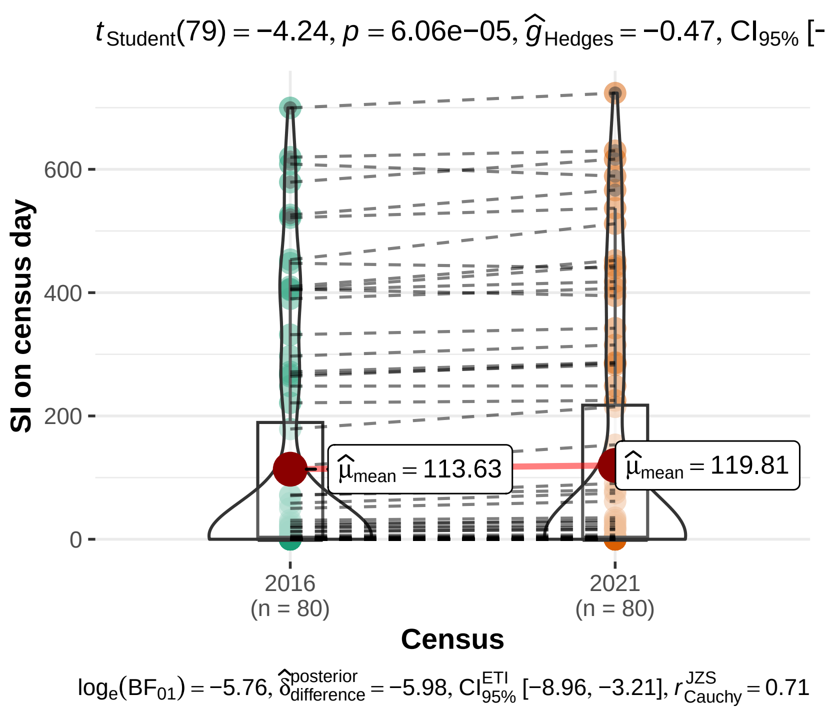


Figure 14: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ bus mode only, violin and box plot generated using ggstatsplot package

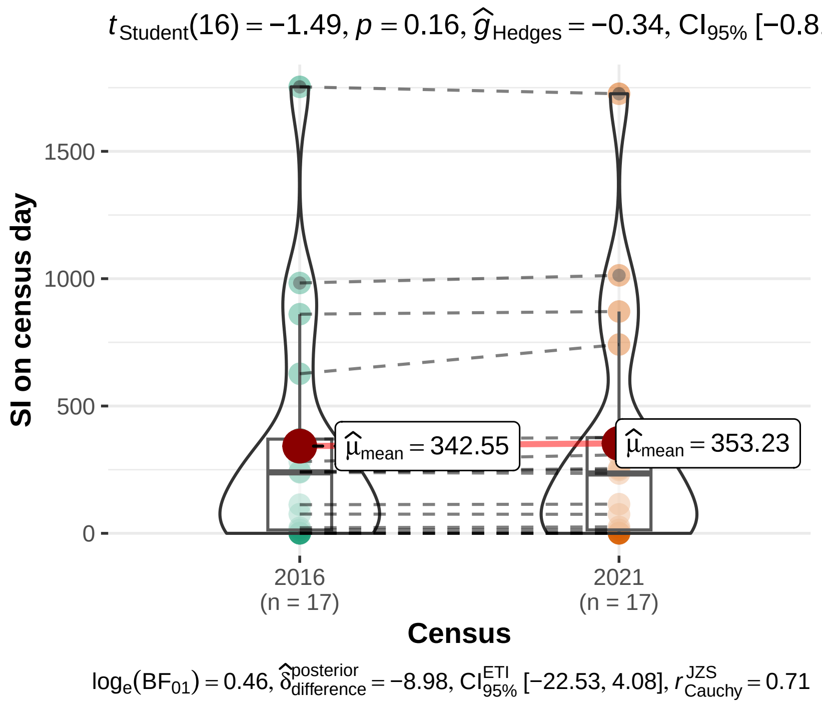


Figure 15: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ tram mode only, violin and box plot generated using ggstatsplot package

for LGAs in Victoria[40] between 2016 and 2021, as shown in Figures 13 and 14. However, Figure 15 indicates that there have not been any significant change in SI tram scores those LGAs that have tram services.

Entirely new modes of transit were introduced in 0 LGAs between 2016 and 2021. The following sections further examine changes in SI scores for each mode.

[40] Considering only those LGAs that had rail or bus services in 2016 or 2021

Rail Supply Index

Figure 16 maps percentage changes in SI_{rail} between 2016 and 2021[41]. These are also shown in Figure 17.

The maps indicate reductions in SI_{rail} in western Victoria and also in the north east. Within Greater Melbourne there appears to have been a small decrease on the Mornington Penninsula, while the SI_{rail} has increased by a large amount in one of the inner east LGAs.

There was no significant difference in changes in SI between LGAs in Greater Melbourne or the Rest of Victoria.

The largest relative changes in SI were the loss of rail services in Northern Grampians, Horsham and Hindmarsh. These LGAs are serviced by the Overland, rather than regular V/line services, so it may be that the Overland service was not included in the GTFS file in

[41] The City of Manningham is excluded from these results as there are no railway stations within its boundaries. Only a tiny fraction of the LGA is within the 800m buffer zone of any railway stations, meaning that the SI_{rail} scores are very low (i.e. <0.1). However, as a result the percentage change between 2021 and 2016 is proportionally very high.

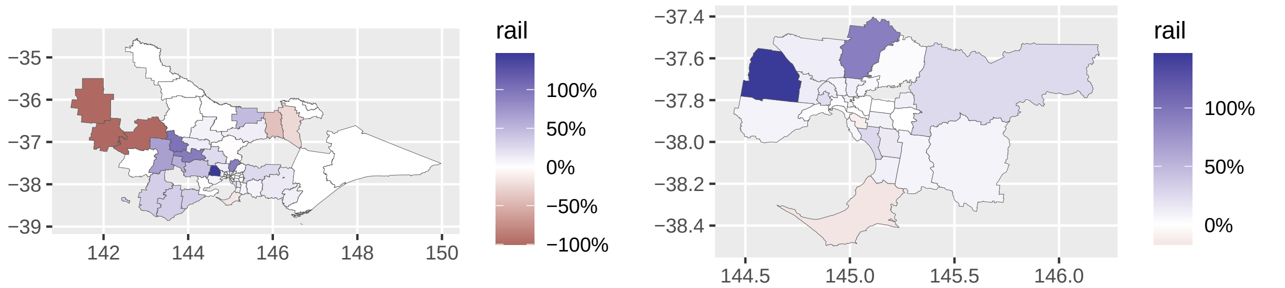


Figure 16: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ rail mode only

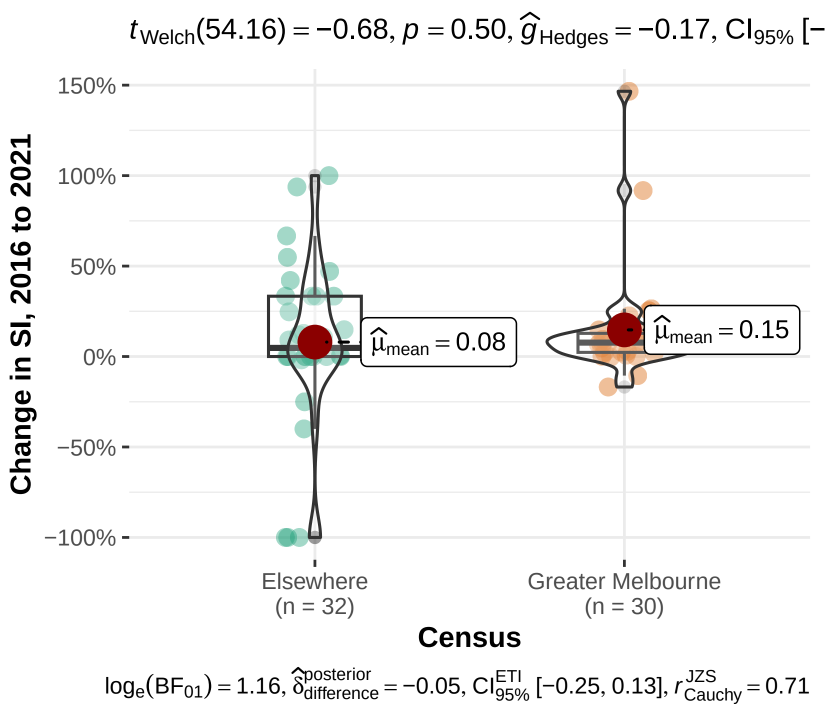


Figure 17: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ rail mode only, violin and box plot generated using ggstatsplot package

Table 4: $SI_{LGA2021,10/8/21}$ percent change $SI_{LGA2021,9/8/16}$ for those LGAs where SI_{rail} has fallen

		Change in SI			
LGA	Location	LRT	Rail	Bus	Total
Hindmarsh	Elsewhere	0%	-100%	8%	-52%
Horsham	Elsewhere	0%	-100%	77%	75%
Northern Grampians	Elsewhere	0%	-100%	90%	85%
Benalla	Elsewhere	0%	-40%	9%	7%
Wangaratta	Elsewhere	0%	-25%	-3%	-4%
Mornington Peninsula	Greater Melbourne	0%	-17%	19%	18%
Glen Eira	Greater Melbourne	3%	-11%	10%	4%
Mitchell	Elsewhere	0%	-2%	85%	53%
Stonnington	Greater Melbourne	18%	0%	-3%	8%

2021, rather than an actually cessation of services[42]. Benalla (-40%) and Wangaratta (-25%) in the north east of Victoria had the next largest reduction in SI_{rail} scores for 2021 compared to 2016, followed by the Mornington Peninsula (-17%).

[42] TO DO: look more closely at the GTFS file.

Table 5: $SI_{LGA2021,10/8/21}$ percent change $SI_{LGA2021,9/8/16}$ for those LGAs where SI_{rail} has risen 25 percent or more

		Change in SI			
LGA	Location	LRT	Rail	Bus	Total
Melton	Greater Melbourne	0%	147%	38%	41%
Central Goldfields	Elsewhere	0%	100%	-2%	-1%
Hepburn	Elsewhere	0%	94%	38%	39%
Whittlesea	Greater Melbourne	2%	92%	11%	13%
Pyrenees	Elsewhere	0%	67%	-19%	8%
Ballarat	Elsewhere	0%	55%	33%	34%
Greater Shepparton	Elsewhere	0%	47%	13%	13%
Moorabool	Elsewhere	0%	42%	16%	19%
Warrnambool	Elsewhere	0%	33%	-1%	-1%
Corangamite	Elsewhere	0%	33%	0%	26%
Colac Otway	Elsewhere	0%	33%	-1%	1%
Surf Coast	Elsewhere	0%	33%	5%	6%
Kingston (Vic.)	Greater Melbourne	0%	26%	8%	11%
Yarra Ranges	Greater Melbourne	0%	25%	8%	9%

The largest relative increases in SI_{rail} were in Melton (147%), Central Goldfields (100%) and Hepburn (94%)[43]. However, Melton's SI_{rail} in 2021 is only 1.5, the Central Goldfields is 0.0079 and Hepburn's is 0.01207615, suggesting the the large relative change in SI_{rail} is more to do with having low values in 2016.

[43] Again, Manningham has been excluded as it does not have any railway stations within its boundaries, and its large percentage increase is due to having a very low SI_{rail} score in 2016 (<0.1)

Bus Supply Index

Figure 18 maps percentage changes in $\mathrm{SI}_{\mathrm{bus}}$ between 2016 and 2021. These are also shown in Figure 19.

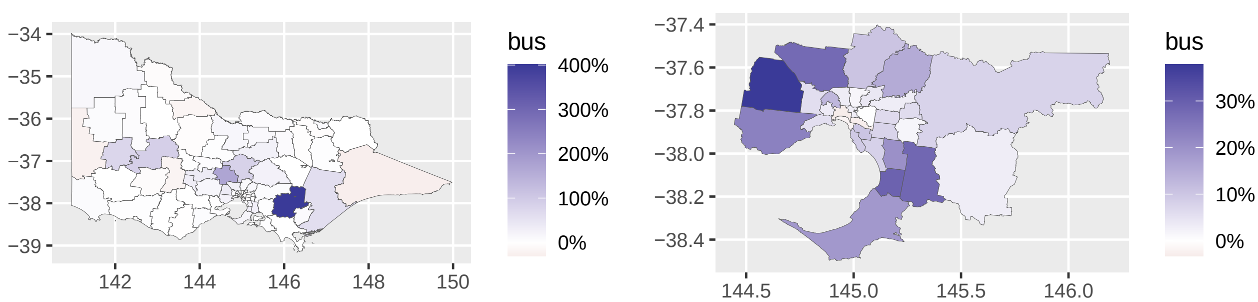


Figure 18: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ bus mode only, violin and box plot generated using ggstatsplot package

The maps indicate that the larger reductions in SI_{bus} have occurred in far western and eastern Victoria, and also in the north west. Within

Greater Melbourne there appears to have been a small decrease on the Mornington Penninsula, while the SI_{rail} has increased by a large amount in one of the inner east LGAs. The larger increases have occurred east of Melbourne, and in some of the outer parts of Greater Melbourne.

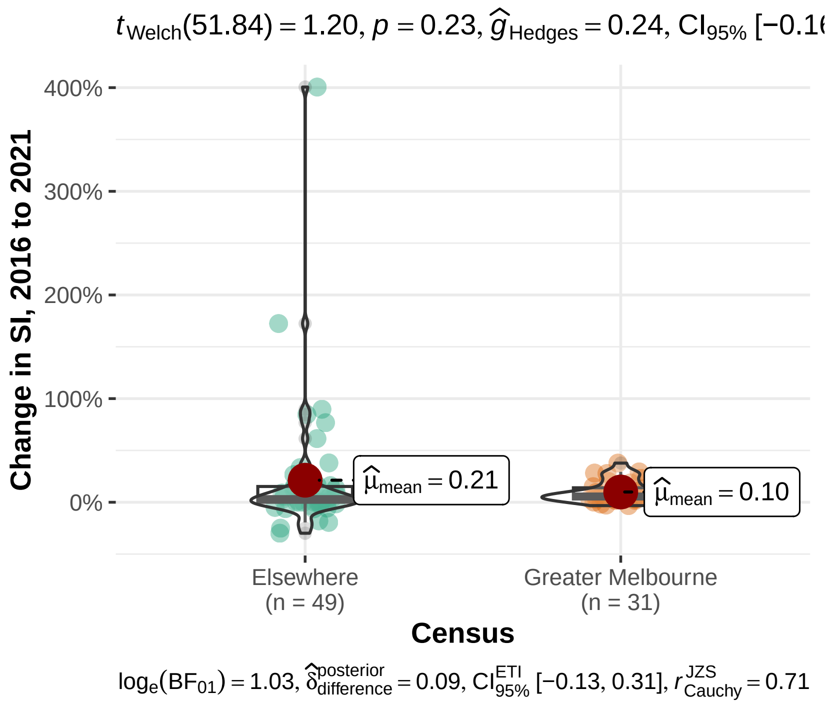


Figure 19: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ bus mode only

There was no significant difference in changes in SI_{bus} between LGAs in Greater Melbourne or the Rest of Victoria (Figure 19). SI_{bus} increased by an average of 17% for LGAs in Victoria. However, there were some LGAs that saw decreases in SI between 2016 and 2021 (Table 6).

The largest relative changes in SI were in East Gippsland (-30%), West Wimmera (-25%), Pyrenees () and Gannawarra (-18%). For Pyrenees this appears to have been offset by a large increase in rail service levels (+67%).

However, some LGAs had relatively large increases in SI_{bus} between 2016 and 2021.

The largest relative increases in $\mathrm{SI}_{\mathrm{bus}}$ were in Baw Baw (401%), Northern Grampians (90%) and Macedon Ranges (172%). For Baw Baw and Macedon Ranges this appears to have involved large increases in bus service levels (+401%, 172%). Northern Grampians, lost all of its train services, but saw SI_{bus} increase by 90%. Horsham,

LGA	Location	Change in SI			
		LRT	Rail	Bus	Total
Colac Otway	Elsewhere	0%	33%	-1%	1%
Warrnambool	Elsewhere	0%	33%	-1%	-1%
Port Phillip	Greater Melbourne	3%	2%	-2%	2%
Central Goldfields	Elsewhere	0%	100%	-2%	-1%
Stonnington	Greater Melbourne	18%	0%	-3%	8%
Melbourne	Greater Melbourne	-2%	2%	-3%	-1%
Wangaratta	Elsewhere	0%	-25%	-3%	-4%
Loddon	Elsewhere	0%	0%	-5%	-3%
Indigo	Elsewhere	0%	0%	-6%	-5%
Ararat	Elsewhere	0%	0%	-6%	-6%
Swan Hill	Elsewhere	0%	0%	-6%	-6%
Gannawarra	Elsewhere	0%	0%	-18%	-13%
Pyrenees	Elsewhere	0%	67%	-19%	8%
West Wimmera	Elsewhere	0%	0%	-25%	-25%
East Gippsland	Elsewhere	0%	0%	-30%	-30%

Table 6: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for those LGAs where SI_{bus} has fallen

LGA	Location	Change in SI			
		LRT	Rail	Bus	Total
Murrindindi	Elsewhere	0%	0%	25%	25%
Strathbogie	Elsewhere	0%	12%	27%	14%
Hume	Greater Melbourne	0%	12%	28%	27%
Casey	Greater Melbourne	0%	7%	28%	27%
Frankston	Greater Melbourne	0%	11%	29%	28%
Ballarat	Elsewhere	0%	55%	33%	34%
Melton	Greater Melbourne	0%	147%	38%	41%
Hepburn	Elsewhere	0%	94%	38%	39%
Wellington	Elsewhere	0%	0%	61%	55%
Horsham	Elsewhere	0%	-100%	77%	75%
Mitchell	Elsewhere	0%	-2%	85%	53%
Northern Grampians	Elsewhere	0%	-100%	90%	85%
Macedon Ranges	Elsewhere	0%	25%	172%	81%
Baw Baw	Elsewhere	0%	15%	401%	190%

Table 7: Change in SI, 2016 to 2021, by LGA, for those LGAs where SI has risen 25 percent or more

likewise, saw the loss of all of its rail services, but an increase of 77% in SI_{bus}.

LRT Supply Index

Figure 20 maps percentage changes in $\mathrm{SI_{LRT}}$ between 2016 and 2021. These are also shown in Figure 21.

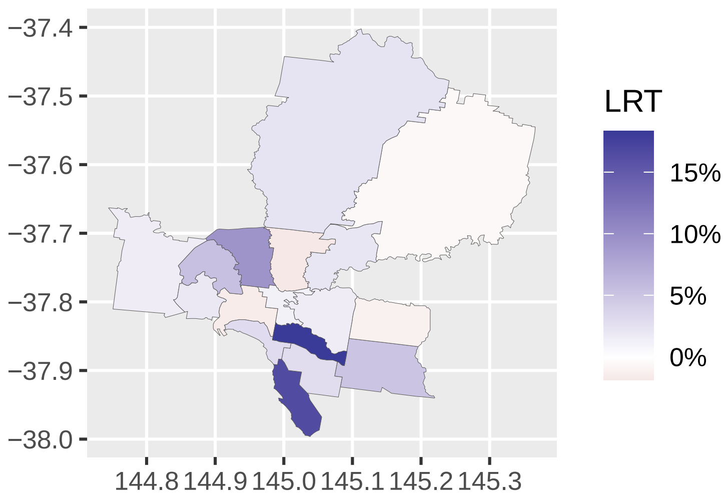


Figure 20: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ tram mode only, violin and box plot generated using ggstatsplot package

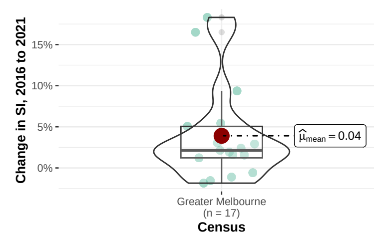


Figure 21: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ LRT mode only

There have been increases in SI_{LRT} in the south-east and some parts of the north of the Melbourne Tram network, but decreases in central Melbourne, some parts of the east and parts of the inner north.SI_{LRT} increased by an average of 1% for LGAs in Victoria.

Table 8: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for LGAs where SI_{LRT} fell

		Change in SI			
LGA	Location	LRT	Rail	Bus	Total
Nillumbik	Greater Melbourne	-1%	3%	15%	14%
Whitehorse	Greater Melbourne	-1%	0%	6%	5%
Melbourne	Greater Melbourne	-2%	2%	-3%	-1%
Darebin	Greater Melbourne	-2%	11%	2%	1%

Table 8 shows the 4 LGAs where SI_{LRT} fell. In general, the changes were relatively small.

Table 9 shows LGAs where SI_{LRT} increased. The largest relative increases in SI were in Stonnington (18%) and Bayside (17%).

Victoria, Australia: transit needs

LGA	Location	Change in SI			
		LRT	Rail	Bus	Total
Stonnington	Greater Melbourne	18%	0%	-3%	8%
Bayside (Vic.)	Greater Melbourne	17%	7%	9%	9%
Moreland	Greater Melbourne	9%	8%	3%	6%
Moonee Valley	Greater Melbourne	5%	14%	13%	11%
Monash	Greater Melbourne	5%	10%	8%	8%
Port Phillip	Greater Melbourne	3%	2%	-2%	2%
Glen Eira	Greater Melbourne	3%	-11%	10%	4%
Whittlesea	Greater Melbourne	2%	92%	11%	13%
Banyule	Greater Melbourne	2%	5%	4%	4%
Maribyrnong	Greater Melbourne	2%	23%	3%	6%
Boroondara	Greater Melbourne	2%	1%	0%	1%
Brimbank	Greater Melbourne	2%	13%	6%	6%
Yarra	Greater Melbourne	1%	7%	3%	3%

Table 9: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ LGAs where SI_{LRT} rose by at least 10 percent

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.

Acknowledgements

This document was prepared using the `rticles` template, created by Gregory Macfarlane, which is based on the LaTeXoriginally posted by David Pritchard in 2009 and updated it in 2011, soon after TRB began allowing PDF submissions. Gregory Macfarlane and Ross Wang made adjustments to the template, and Ross Wang now maintains the LaTeXtemplate at `https://github.com/chiehrosswang/TRB_LaTeX_tex`. Gregory Macfarlane created the `rticles` template in 2021.

Bibliography

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 and Zed Senbergs. Identifying spatial gaps in public transport provision for socially disadvantaged Australians: the Melbourne needs-gap study. 2007.

EPSG. *EPSG:28355; GDA94/MGA zone 55*, 1995. URL `https://epsg.io/28355`.

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://CRAN.R-project.org/package=gtfstools`. R package version 1.2.0.

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/.

David Kahle, Hadley Wickham, and Scott Jackson. *ggmap: Spatial Visualization with ggplot2*, 2023. URL https://github.com/dkahle/ggmap. R package version 3.0.2.

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.

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. *absmapsdata: A catalogue of ready-to-use ASGS (and other) sf objects*, 2023. R package version 1.3.3.

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

Indrajeet Patil. *ggstatsplot: ggplot2 Based Plots with Statistical Details*, 2023. URL https://CRAN.R-project.org/package=ggstatsplot. R package version 0.11.1.

Edzer Pebesma. *sf: Simple Features for R*, 2023. URL https://CRAN.R-project.org/package=sf. R package version 1.0-12.

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 1.5.0.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`.

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. *tidyverse: Easily Install and Load the Tidyverse*, 2023. URL `https://CRAN.R-project.org/package=tidyverse`. R package version 2.0.0.

Hadley Wickham, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*, 2023. URL `https://CRAN.R-project.org/package=ggplot2`. R package version 3.4.2.

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.

Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2023. URL `https://yihui.org/knitr/`. R package version 1.42.

Hao Zhu. *kableExtra: Construct Complex Table with kable and Pipe Syntax*, 2021. URL `https://CRAN.R-project.org/package=kableExtra`. R package version 1.3.4.