Leveraging GTFS data to assess transit supply, and needs-gaps

James Reynolds and Graham Currie

2023-08-10

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[1].

The TCQSM specifies Levels of Service (LOS) between A and F across a range of factors[2]. This scoring scheme appears relatively simple to explain[3], 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[4] and it might be challenging to explain the detail of these measures 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)[5] [MobilityData, undated] and widespread release of schedule data in this format, however, has helped towards making transit metrics more broadly availabile. Wong [2013] provides an example, having 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], rather than being able to calculate scores independently.

Previous research by Currie and Senbergs [2007] developed a transit

[1] For example, politicians, other decision-makers or the general public.

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

[3] A is good and F is bad. Also this scoring system matches to the A to F LOS scoring used in many traffic capacity analysis software and manuals.

[4] 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.

[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]

Supply Index (SI) that is relatively easy to calculate and understand, and which is open[9]. This Index is based on calculating the number of transit arrivals at stops within an area of interest, adjusted to account for the typical walk-access catchment for each stop. Unfortunately, the index 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] SI directly from GTFS data.

[9] Rather than a black box, closed-source or otherwise secret algorithm.

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.

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. Such metrics might then be able to be used to access proposed network changes without the need for specialist software or analysis. A related objective is to increase the availability of metrics that are relatively easy to understand and use when making decisions about or advocating for changes to existing services, including for those who may not be technical specialists in transit planning and scheduling.

This rest of this document is structured as follows: the next section discusses the research context of transit metrics and the the Supply Index. The Transport Needs Index used in Currie and Senbergs [2007][10] is also briefly discussed. In the third section the methodology for the code development is outlined, including discussion of the case study GTFS for Victoria, Australia, that was used to test and verify the code output. This includes presentation and mapping of the Australian Bureau of Statistics (ABS) Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) scores from the 2016 and 2021 censuses, which are used later in the paper to explore the needs-gaps for Victorian Local Government Areas (LGAs). In the fourth section results are presented, starting with verification of the code output through hand-calculation of SI scores for Statistical Area 1 (SA1) areas in the Victorian Alps and Talbot. SI scores for LGAs across Greater Melbourne and Victoria are also presented, comparing transit service levels on the day of the census in 2016 with those on the day of the census in 2021. Mode-by-mode SI scores are also explored, followed by an examination of needs-gaps across the IRSAD scores, ranks and population levels. The document then closes with a brief

[10] Being the other part of a needs-gap assessment.

discussion and conclusion section.

Research context

Even a brief search shows that there is a very large number of metrics available for benchmarking transit services[11].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].

[11] 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].

However, many of these metrics appear 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 (the Walk Score [2023] website shows scores 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 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. [2013] provide a manual describing all the metrics and how to calculate them).]. 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].

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

The Suppy Index

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, which are a feature of the TQCSM. 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 $SI_{area,time}$ scores can be aggregated to calculate an overall score across multiple time periods or for a region encompassing multiple areas of interest.

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

[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.

Currie and Senbergs [2007] calculated the $SI_{area,time}$ 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 $SI_{area,time}$ scores can be calculated and compared for any area of interest where transit service information is available in that format.

[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

[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 Central Business District (CBD), weighted at 15%.

[19] Weighted at 14%.

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). Later in this document a similar needs-gap analysis is presented for Victorian LGAs. However, in the needs-gap analysis here only the IRSAD scores and population data are reported. Calculating the combined Transport Needs index as per Currie and Senbergs [2007] may be a direction for future research.

[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 Greater Melbourne and Victoria, Australia

Code development

This document has been prepared using Rmarkdown, which allows the intermingling of written text, code segments and code outputs. Code segements developed in this research are shown in the following, together with the relevant descriptive text[26]

Various analysis tools are available that make use 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 SL_{Bn} term in the Currie and Senbergs [2007] SI equation.

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 were adopted, 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 SI_{area} term in the SI equation was calculated on a mode-by-mode and stop-by-stop basis, by first determining the amount of the catchment area ($Area_{Bn}$) that falls into each geographical area of interest for the stop in question. This is then combined with the area for each geographical area of interest ($Area_{area}$) and the number of stop arrivals (SL_{Bn}) to calculate the contribution to the SI 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

[26] The Rmarkdown file is available at `https://github.com/James-Reynolds/Transit_Supply_Index_GTFS` and this can be read in a plain-text editor to view the code snippets themselves. If you are reading this in a PDF document you are seeing just the descriptive text, and outputs from the code where it has been run to produce maps, charts etc.

[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 coded 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 $Area_{Bn}$ calculation needs to be provided in the same units as the $Area_{area}$ variable, necessitating the use of a geographic transform as part of the code.

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[29] However, with only around 5 million people 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.

[29] Greater Melbourne is the term used to describe the larger metropolitan area, encompassing 30 LGAs. The City of Melbourne LGA covers only a small portion of the inner city.

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 (VLine), 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. The Overland train service to Adelaide and the XPT to Sydney are provided seperately to VLine services. Victoria's GTFS feed is published by Public Transport Victoria (PTV)[30].

[30] 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].

The developed code was tested for both SA1- and LGA-level areas of interest. Output was obtained for SA1-level areas using a GTFS file from late April, 2023, running for a full 7-day period. For the LGA-level output the results were compared between 2016 and 2021, to match the Australian census dates. The Australian census is undertaken in early August every 5 years. GTFS feeds were therefore selected for the first week of August of each year, with code output produced for only the day of the census itself[31]. Minor corrections were made to the GTFS files to remove duplicate stop_ids[32].

[31] It takes about a day of processing time to run the code for all of the stops in Vicotira for a single ‘day’ of service. Hence, only the census days (rather than weeks) were analysed to speed development.

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

The Australian Bureau of Statistics (ABS) provides a range of shape files and other resources. 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.

The ABS provides IRSAD datasets for LGAs in excel format. Data for 2016 and 2021 was included in this study. Scores are shown in Figures 1 and 2. Figure 3 shows changes in IRSAD scores between 2016 and 2021.

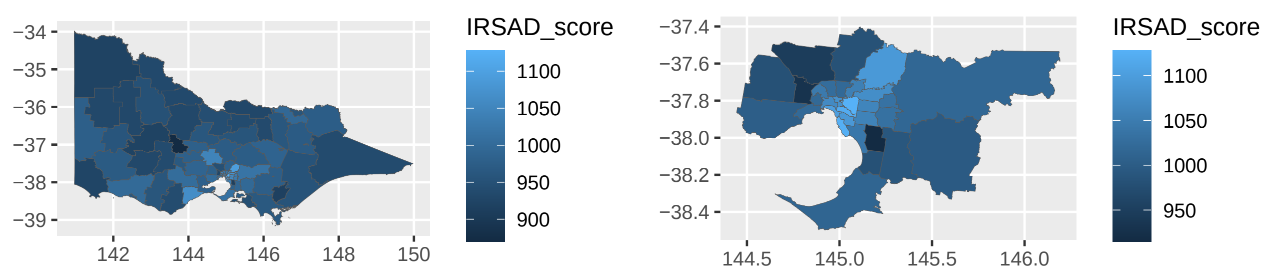


Figure 1: 2016 IRSAD for Victorian and Greater Melbourne LGAs

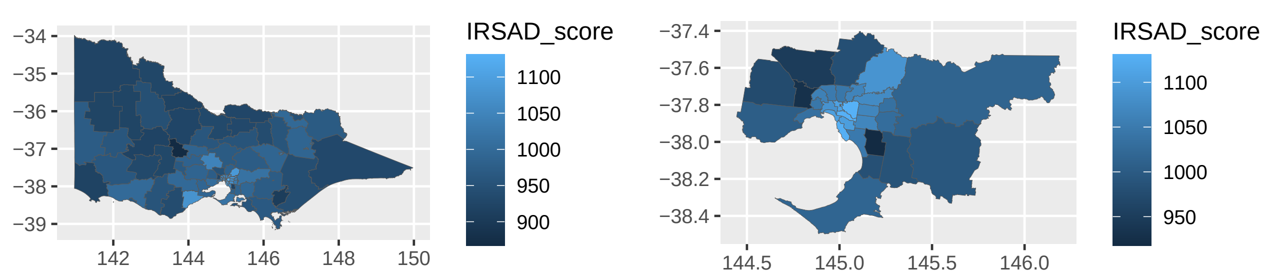


Figure 2: 2021 IRSAD for Victorian and Greater Melbourne LGAs

Figure 3 indicates that has not been a significant change in IRSAD scores for Victorian LGAs between 2016 and 2021. Figure 4 shows

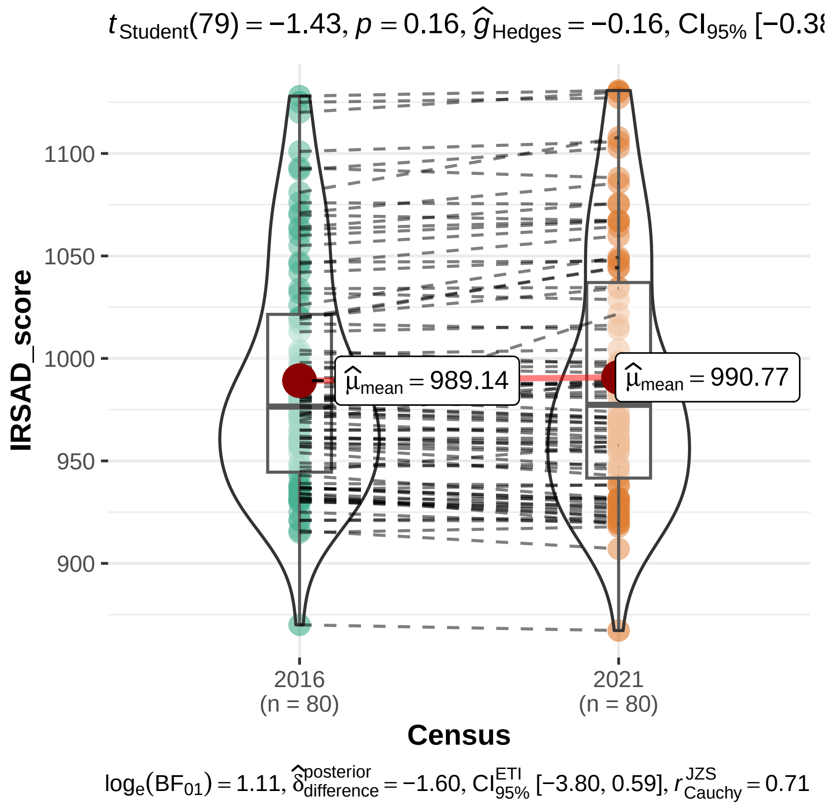


Figure 3: Comparing 2016 and 2021 IRSAD scores for Victorian LGAs. Analysis using ggstatsplot package.

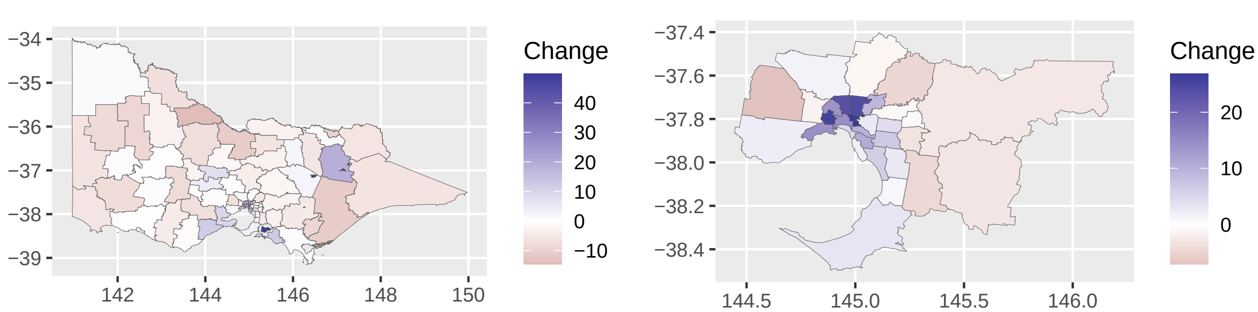


Figure 4: Change in IRSAD scores

the percentage change in IRSAD scores for individual LGAs across Victoria and Greater Melbourne. Alpine Shire and the unincorporated Victorian areas (Mount Buller, French Island in Westernport etc.) appear to have had the largest increases in IRSAD in Victoria, while in Greater Melbourne it appears to have been mostly in inner areas where there have been increases.

There appears to have been some decreases in IRSAD scores in outer regions of Victoria. Outer-eastern parts of Melbourne, and Melton (outer-west) also also appear to have had lower IRSAD scores in 2021 than 2016.

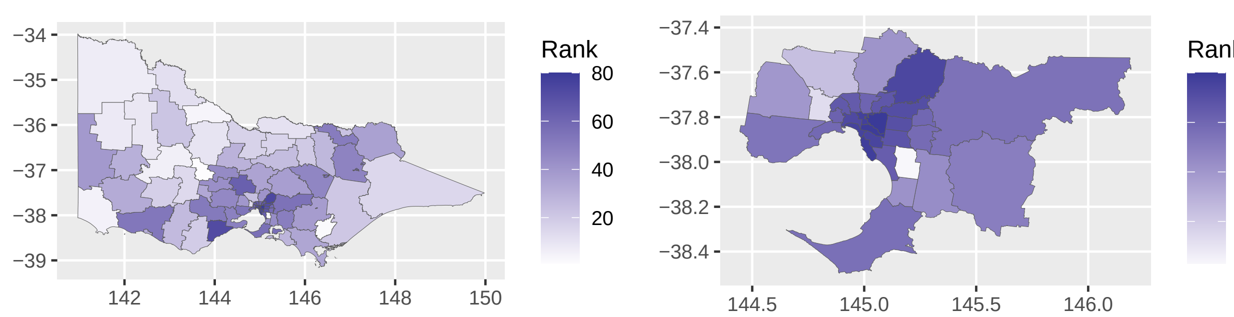


Figure 5: IRSAD rank 2021

Figure 5 shows the 2021 IRSAD ranking for Victoria and Greater Melbourne. A lower rank indicates a LGA with a lower IRSAD score. Inner Melbourne LGAs appear to have the highest ranks, whereas Greater Dandenong has the 3rd lowest IRSAD in the state. For regional LGAs the ranking appears to be generally lower in the north-west, with Central Goldfields having the lowest IRSAD score for the entire state. However, the Latrobe LGA in the east has the second lowest rank, while Glenelg in the south-west is the fifth lowest.

Figure 6 shows the change in ranking between the 2016 and 2021 for each LGA in Victoria and Greater Melbourne. A positive number indicates that the LGA in question had a higher IRSAD score than more of the other Victorian LGAs in 2021 than in it did in 2016. The state ranking appears to have fallen for most of the outer suburbs in Greater Melbourne, especially for Melton, Whittlesea and Casey. Rankings have improved particularly in Brimbank, but also in some of the other nearby LGAs. In regional Victoria Gannawarra's ranking has fallen the most, while Alpine Shire and Ararat's ranking have improved the most.

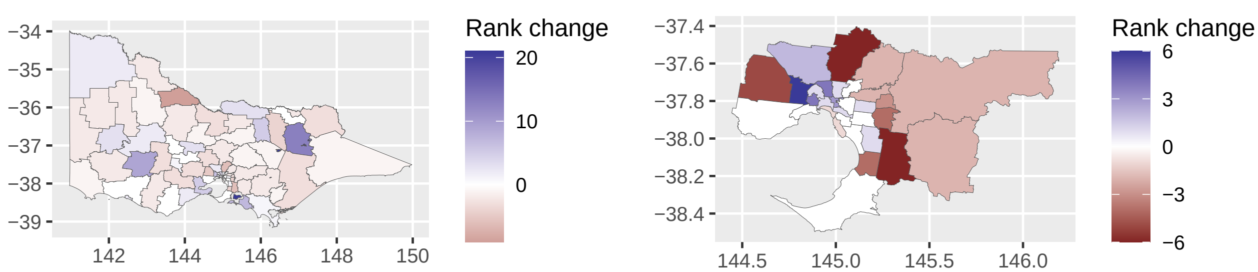


Figure 6: Change in IRSAD rank between 2021 and 2016 (lower rank = worse)

Results

The following subsections discuss the results of cases studies used to develop and test the SI calculating code[33].

[33] 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[34]. 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. Relevant geographic statistics are shown in the following.

[34] Stop:ID 45125, Running Creek Rd/Kiewa Valley Hwy (Running Creek) and Stop ID: 45124, Kiewa Valley Hwy (Mongans Bridge).

```
##   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 that services these stops, but stop times are provided on the PTV website[35]. This indicates a total of 26 arrivals to each of the two bus stops in a seven day week [36]

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

Figure 7: SA1 20403106915, with approximate location of Stop:ID 45125 highlighted, sources ABS and Google Maps

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.

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.

Talbot, Avoca line

Talbot, north-west of Ballarat, provides a location where the accuracy of the developed code can be tested for both: 1) multiple modes; and 2) station/stop catchment areas spanning across multiple areas of interest. As shown in the below figure, the Talbot township is covered by SA1 20103101708, which lies within the SA1 20103101707.

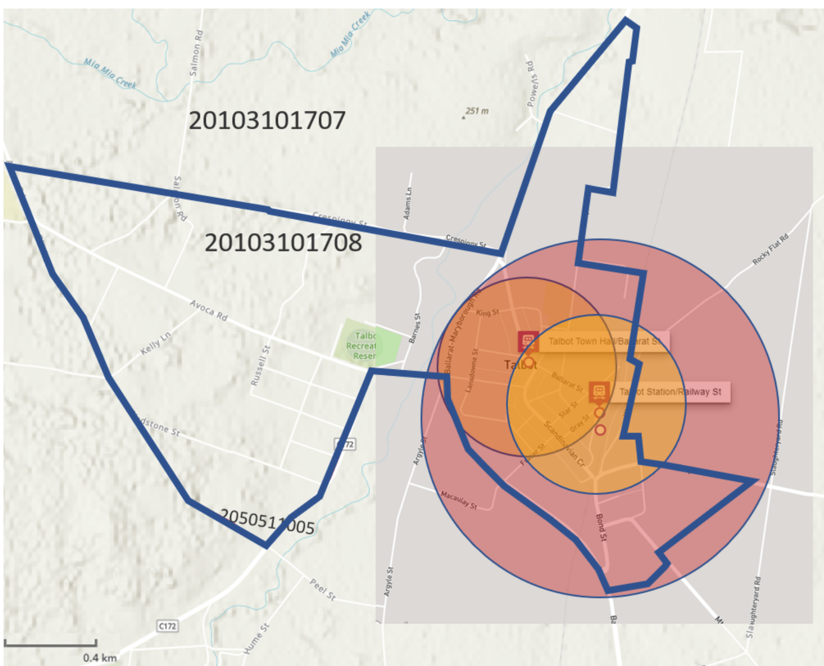


Figure 8: SA1 Location of 20103101707 and 20103101708 boundaries in relation to Talbot railway station and 800m catchment (red), and bus stops and 400m catchments (yellow), source ABS and PTV

The Talbot Railway station's 800m catchment area is roughly three-fifths in SA1 20103101708. A bus stop located close to the railway station has roughly three-quarters of its catchment in 20103101708. The bus stop at the Talbot Town Hall has almost its entire catchment area within 20103101708.

The area of SA1 20103101707 is 237km^2[37].

[37] via LGA data from abs_map package.

Hence the $Area_{Bn}/Area_{SA120103101708}$ term for the Talbot Town Hall bus stop is approximately equal to $(\pi.4^2)/3.05 = 0.165$.

For the bus stop at the Talbot Station the $Area_{Bn}/Area_{SA120103101708}$ term is approximately equal to $3/4(\pi.4^2)/3.05 = 0.124$. For the Railway Station itself the $Area_{Bn}/Area_{SA120103101708}$ term is approximately equal to $(3/5)(\pi.8^2)/3.05 = 0.396$.

Review of the PTV website stop departures indicates that the typical service pattern appears to be:

- Talbot Town Hall bus stop - 42 services per week[38]
- Talbot Station bus stop - 32 services per week[39]
- Talbot Railway Station - 28 services per week^{Two services towards Maryborough every day of the week; and two services towards Melbourne every day of the week]

[38] One Vline bus service towards Maryborough on Thursdays, Fridays, Saturdays, Mondays and Tuesdays(5 per week); 3 Vline services towards Melbourne and/or Ballarat on Thursdays and Fridays, 2 on Saturdays, 1 on Sundays, 3 on Mondays and Tuesdays, and 1 on Wednesdays (16 per week); two Vline bus serices towards Mildura on Thursdays and Fridays, 1 on Satudays and Sundays, 2 on Mondays and Tuesdays and 1 on Wednesdays (11 per week); local bus services to Ballarat on Thursdays (1), Fridays (1), Monday (1), Tuesday (1) and Wednesday(1) (5 per week); and local bus services towards Bendigo on Thursdays (1), Fridays (1), Mondays (1), Tuesdays (1) and Wednesdays (1) (5 per week).

[39] One Vline bus service towards Maryborough on Thursdays, Fridays, Saturdays, Mondays and Tuesdays(5 per week); 3 Vline services towards Melbourne and/or Ballarat on Thursdays and Fridays, 2 on Saturdays, 1 on Sundays, 3 on Mondays and Tuesdays, and 1 on Wednesdays (16 per week); and two Vline bus serices towards Mildura on Thursdays and Fridays, 1 on Satudays and Sundays, 2 on Mondays and Tuesdays and 1 on Wednesdays (11 per week)

Therefore, the $SI_{20103101708}$ is approximately equal to: (Talbot Town Hall bus stop) 42 services per week x 0.165 = 6.92; plus (Talbot Station bus stop) 32 services per week x $3/4(\pi.4^2)/3.05 = 0.00124 = 3.96$; plus (Talbot Railway Station) 28 services per week x 0.00396 = 11.1 = 22

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

Table 2: Developed code output for SA1 20103101707, seven days starting April 27, 2023

sa1_code_2021	sa2_name_2021	SI_CCD
20103101708	Maryborough Surrounds	25.1404

The hand-calculated $SI_{20103101708,27/4-3/5/23}$ is approximately 13% lower than the value calculated by the SI code. However, this appears to be sufficient to accept the result given that:

- the overlaps of the station and stop catchment areas have been estimated based on PTV and ABS maps, rather then precisely calculated using GIS software;
- the hand-calculated SI is based on current (August 2023) service schedules, whereas the code-calculated value is based on a GTFS file from April 27, 2023[40]; and
- Overall the SI scores vary greatly across LGAs[41], meaning that a difference of just 3.19 between the hand-calculated and code-calculated score is likely an issue of precision, rather than accuracy.

[40] The online PTV timetables only go back a month or two, so any changes in the timetables between April and now could not be checked without looking directly at the GTFS files, which sort of defeats the purpose of doing the verification calculations.

[41] Ranging between 0 and 143,006.

Further verification work might involve:

- updating the code calculated values from $SI_{20103101708,27/4-3/5/23}$ to $SI_{20103101708,now+7days}$ such that the current PTV timetables could be directly compared to the code output using the current GTFS files[42];

- obtaining more precise $Area_{Bn}/Area_{areaofinterest}$ terms using GIS or similar;
- stepping through the coding calculations one at a time to verify the intermediate calculations, and hence confirm that this disrepancy is simply a matter of precision;
- undertaking the verification calculations for the SA1 20103101707 area[43] or for some other area where hand calculations are relatively easy to perform.

[42] This would remove the chance that the service patterns have actually changed between April and now, although it is expected that even if there has been a change it is unlikely to have been large.

[43] Unfortunately this is a large SA1 and there are other bus stops (outside of Talbot) that will contribute to the SI score.

Anyhow, this concludes the hand-calculation verification efforts to date. The next sections, therefore, look at the code outputs alone, with a focus on LGAs in Greater Melbourne and Victoria.

Victoria, Australia: Supply Index results for LGAs

To better understand how 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 2016 and 20121, but only a single day of service, being the date of the census in each year[44].

Figure 9 show the $SI_{LGA2021,10/8/21}$ values for Victoria (left) and Greater Melbourne (right).

[44] Tuesday August 9, 2016, and Tuesday August 10, 2021. Analysis of a single day was selected so as to reduce the processing time required to output the SI scores, as the code currenly takes about one day to run for all the stops in Victoria for

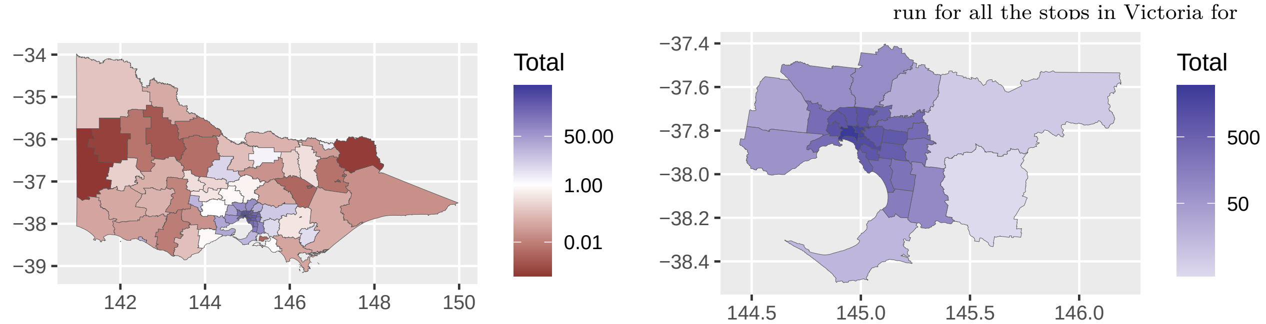


Figure 9: $SI_{LGA2021,10/8/21}$

In general, Figure 9 indicates that the amount of transit service is higher for LGAs that are closer to the centre of Melbourne.

Figure 10 compares the SI for 2016 and 2021 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.

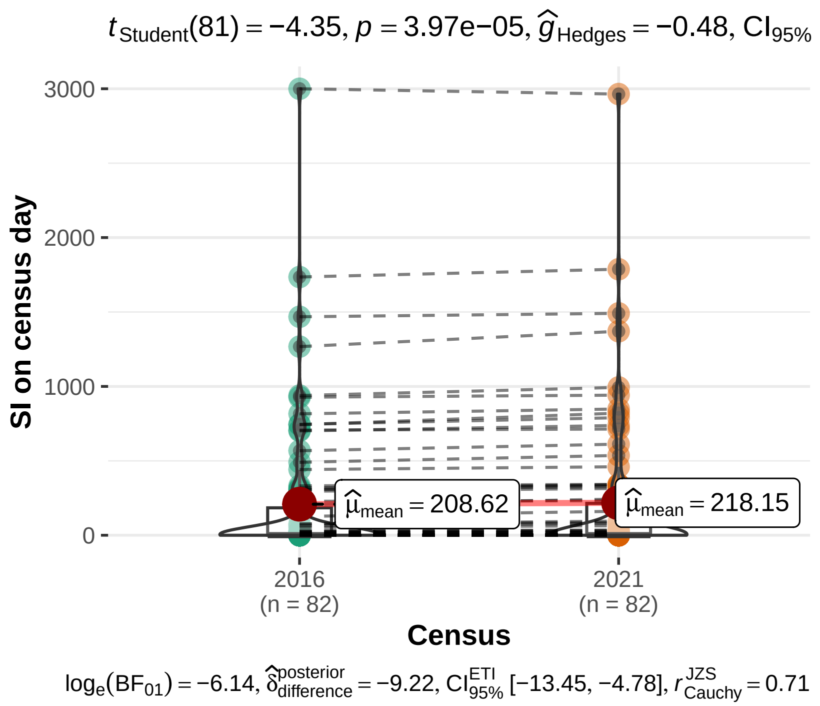


Figure 10: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ violin and box plot generated using ggstatsplot package

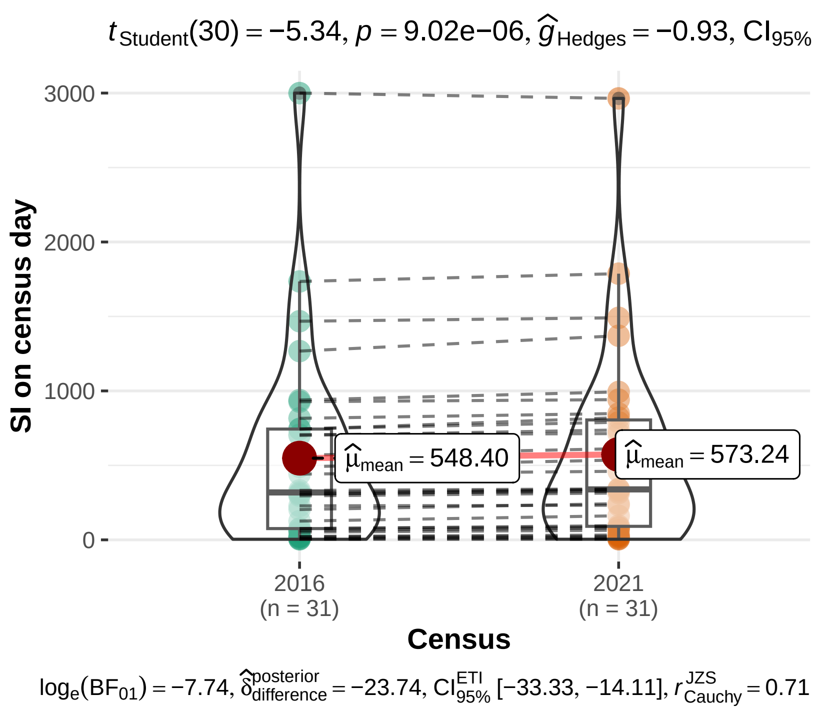


Figure 11: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ Greater Melbourne, violin and box plot generated using ggstatsplot package

Figure 11 shows that there was a significant change in SI for LGAs within Greater Melbourne, rising from 548.4 to 573.2.

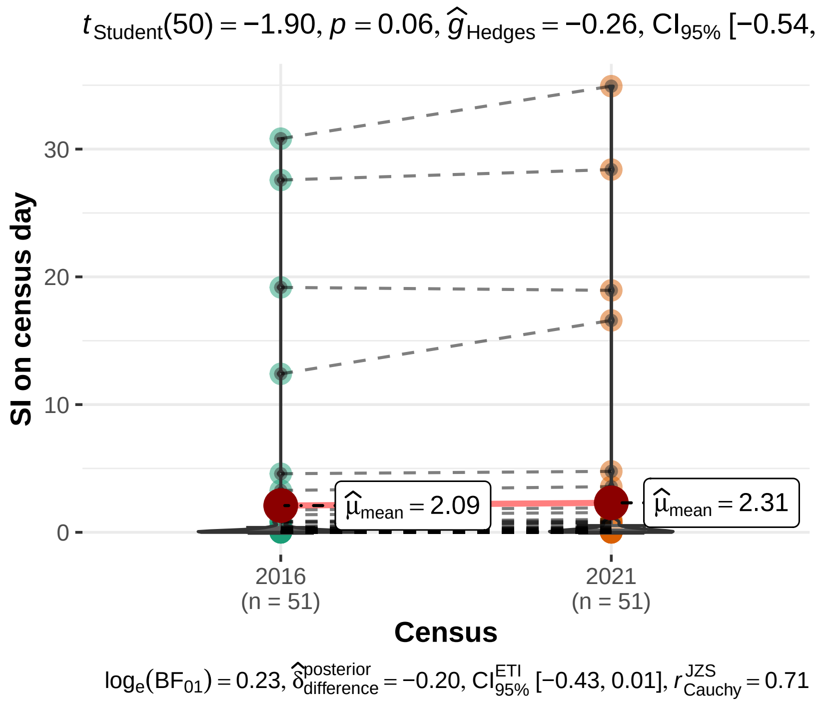


Figure 12: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ non-Greater Melbourne parts of Victoria, violin and box plot generated using ggstatsplot package

There was no significant change amongst LGAs in the rest of Victoria (Figure 12).

Figure 13 maps percentage changes in SI between 2016 and 2021. These are also shown in Figure 14. There was no significant difference in changes in SI between LGAs in Greater Melbourne or the Rest of Victoria (Figure 14). 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 3).

The largest relative changes in SI were in Hindmarsh (-52%), East Gippsland (-30%) and West Wimmera (-25%). For Hindmarsh this appears to have involved the entire removal of the rail service[45], 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 are shown in Table 4. These 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% and

[45] This might be related to the Overland rail service to Adelaide. Further work might involve looking at whether this service was operational on the day of the census in 2021, and whether this service is included in the PTV GTFS feed.

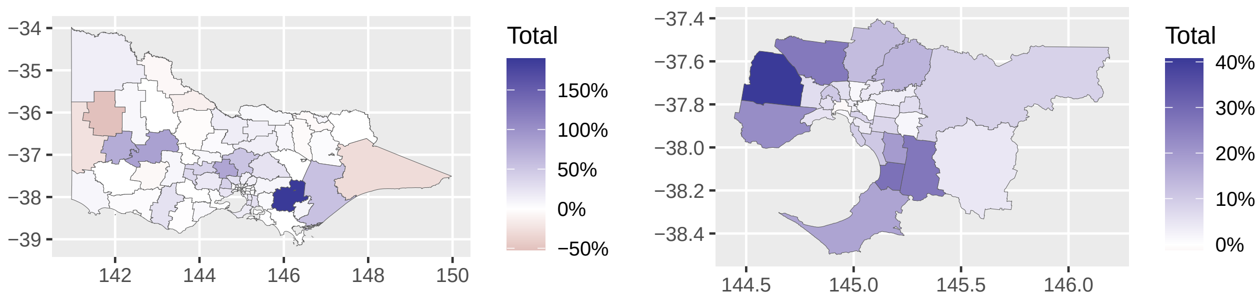


Figure 13: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$

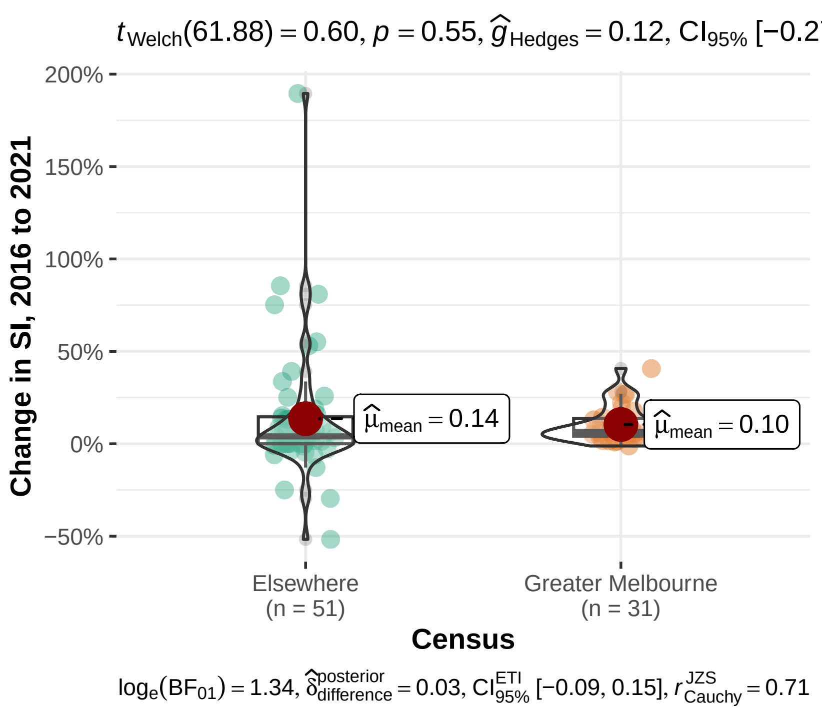


Figure 14: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ violin and box plot generated using ggstatsplot package

Table 3: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for those LGAs where SI has fallen, by mode

LGA	Location	Change in SI			
		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%

Table 4: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for those LGAs where SI has risen 25 percent or more

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%

+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[46], but an increase of 77% in SI_{bus}.

[46] Again, this might be related to the Overland rail service to Adelaide.

Supply Index by mode

Figure 15 shows changes in the proportion of the SI score[47] delivered by train, tram and bus across Victoria.

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

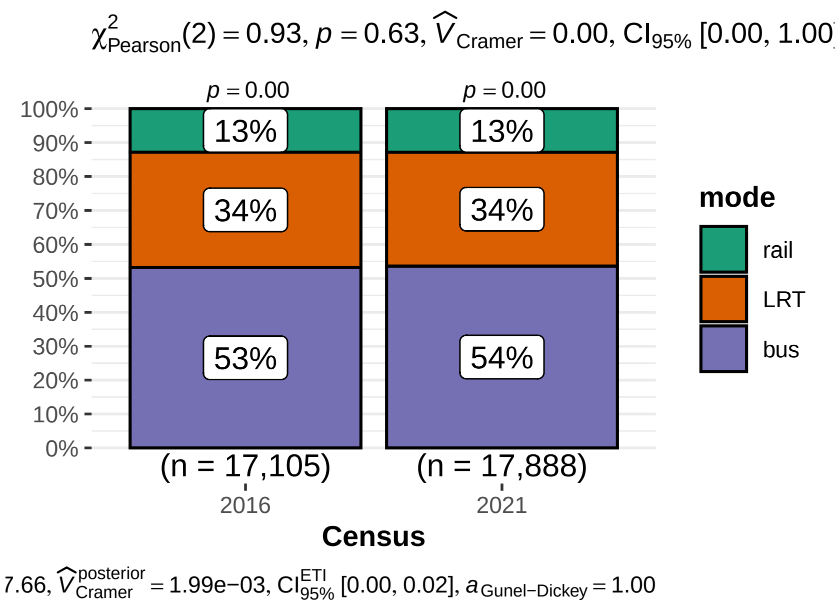


Figure 15: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ proportion by mode, bar chart generated by ggstatsplot package

Figure 15 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[48].

[48] Not shown.

There were small but significant changes in SI_{rail} and SI_{bus} scores for LGAs in Victoria[49] between 2016 and 2021, as shown in Figures 16 and 17. However, Figure 18 indicates that there have not been any significant change in SI tram scores for those LGAs that have tram services.

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

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.

Rail Supply Index

Figure 19 maps percentage changes in SI_{rail} between 2016 and 2021[50]. These are also shown in Figure 20.

[50] The City of Manningham is excluded from these results as it has a very high percentage change, which may be misleading. The reality is that 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.

The maps indicate reductions in SI_{rail} in western Victoria and also

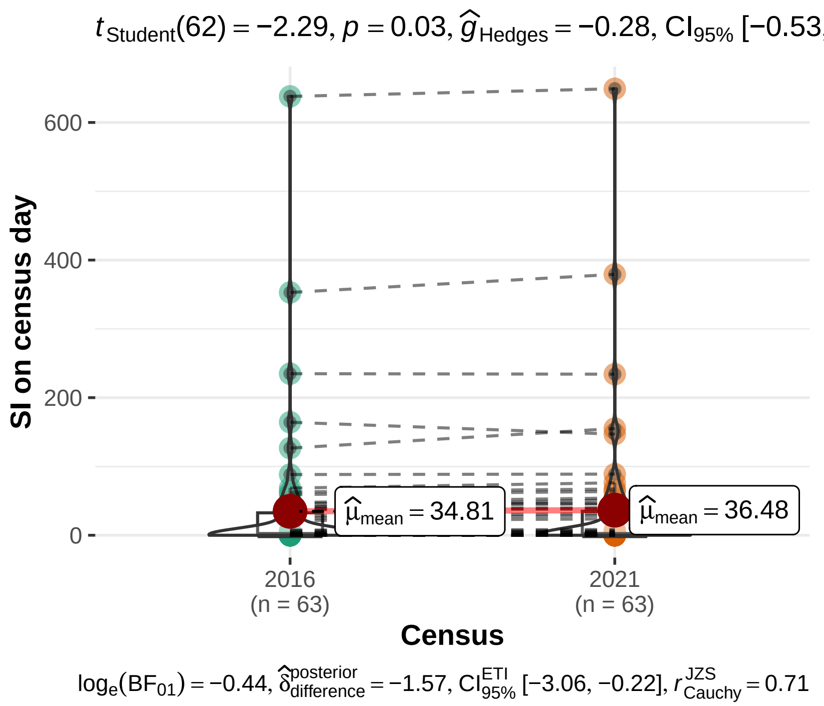


Figure 16: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ rail mode only, violin and box plot generated using ggstatsplot package

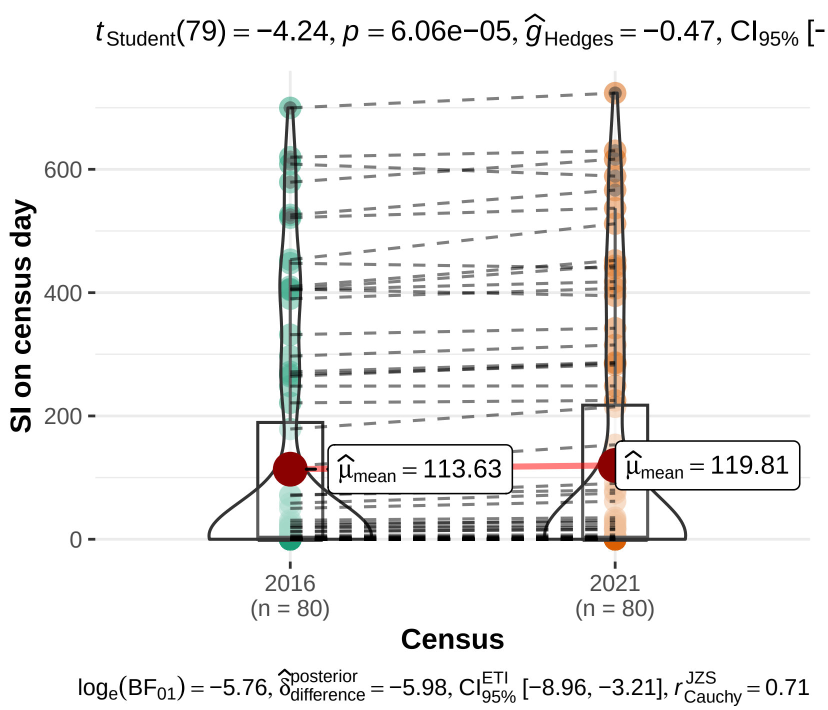


Figure 17: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ bus mode only, violin and box plot generated using ggstatsplot package

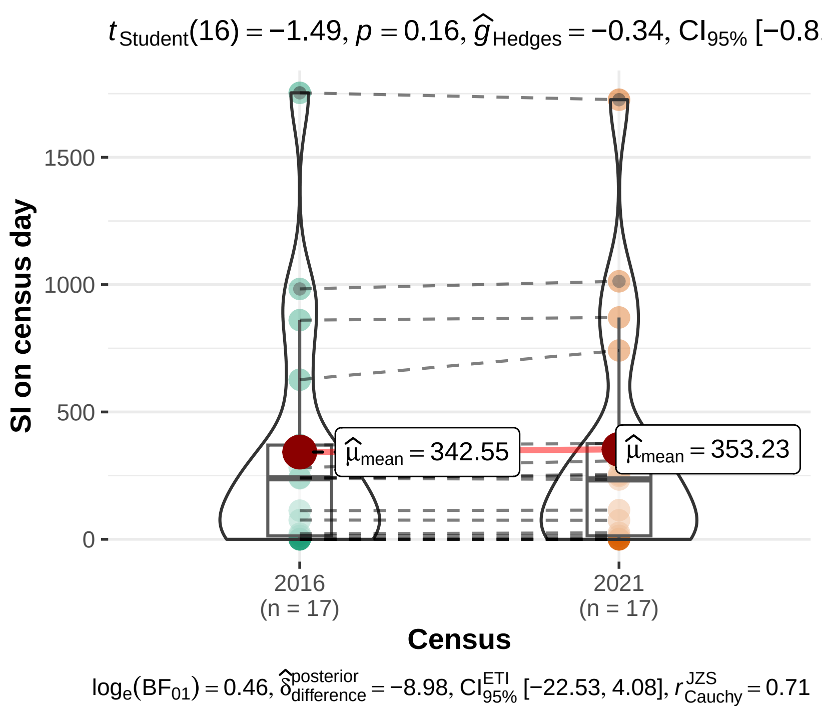


Figure 18: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ tram mode only, violin and box plot generated using ggstatsplot package

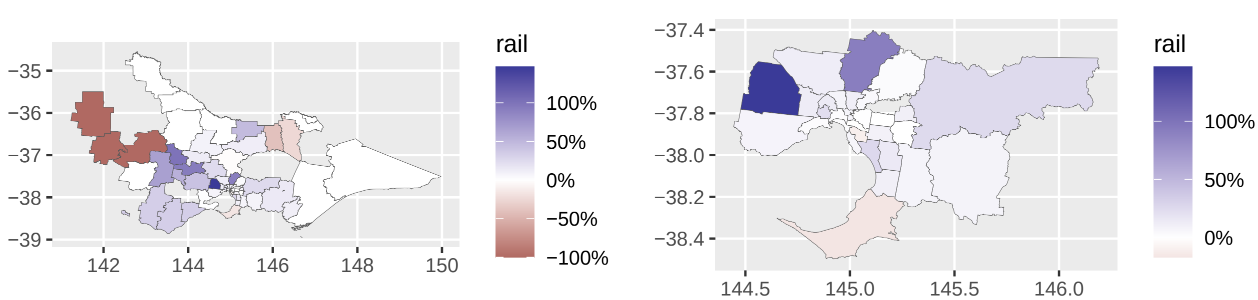


Figure 19: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ rail mode only

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 the most for Melton (west)and Whittlesea (north).

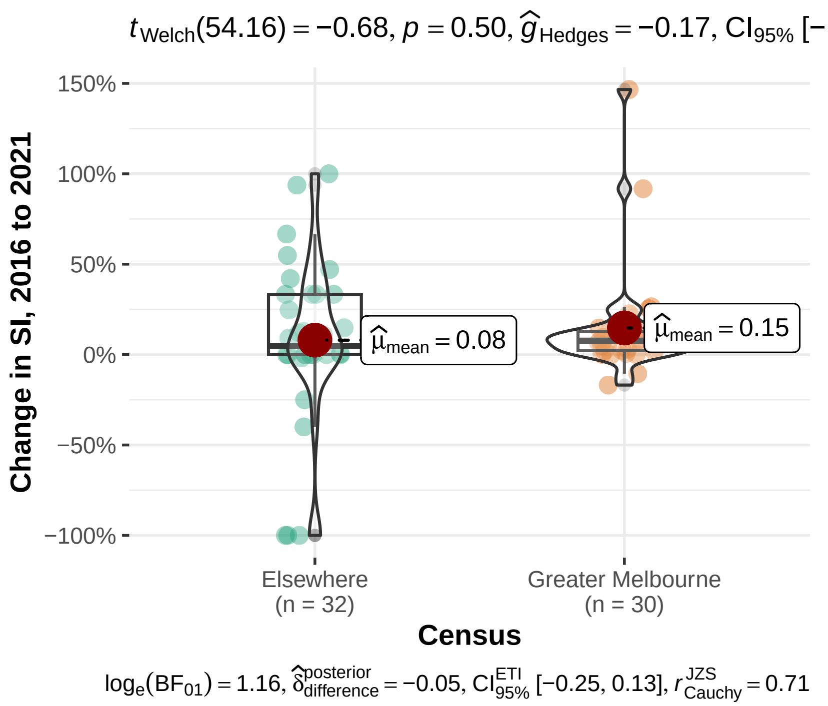


Figure 20: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ rail mode only, violin and box plot generated using ggstatsplot package

Figure ?? shows no significant difference in changes in SI between those LGAs in Greater Melbourne versus those in the Rest of Victoria.

Table 5: $SI_{LGA2021,10/8/21}$ 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%

The largest relative changes in SI were the loss of rail services in Northern Grampians, Horsham and Hindmarsh[51]. 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%).

The largest relative increases in SI_{rail} were in Melton (+147%),

[51] 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 2021, rather than an actually cessation of services. TO DO: look more closely at the GTFS file.

LGA	Location	Change in SI			
		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%

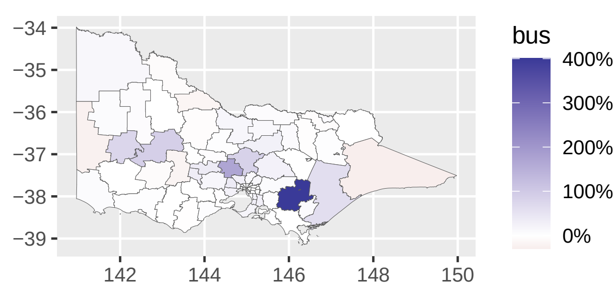
Table 6: $SI_{LGA2021,10/8/21}$ percent change $SI_{LGA2021,9/8/16}$ for those LGAs where SI_{rail} has risen 25 percent or more

Central Goldfields (+100%) and Hepburn (+94%)[52]. However, Melton's SI_{rail} in 2021 is only 1.5, the Central Goldfields is 0.0079 and Hepburn's is 0.012, suggesting the the large relative change in SI_{rail} is more to do with having had low values in 2016.

[52] 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 21 maps percentage changes in $\mathrm{SI}_{\mathrm{bus}}$ between 2016 and 2021. These are also shown in Figure 22.



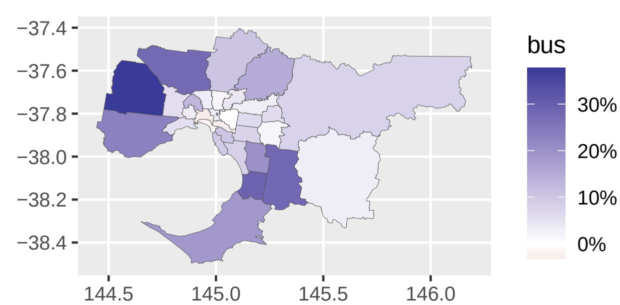


Figure 21: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ bus mode only, violin and box plot generated using ggstatsplot package

The maps indicate that the larger reductions in SI_{bus} have occurred in far western and eastern Victoria. Within Greater Melbourne there appears to have been a small decrease in some inner LGAs, but large increases for many outer LGAs.

Overall SI_{bus} increased by an average of 17% for LGAs in Victoria. There was no significant difference in changes in SI_{bus} between LGAs in Greater Melbourne or the Rest of Victoria (Figure 22). However,

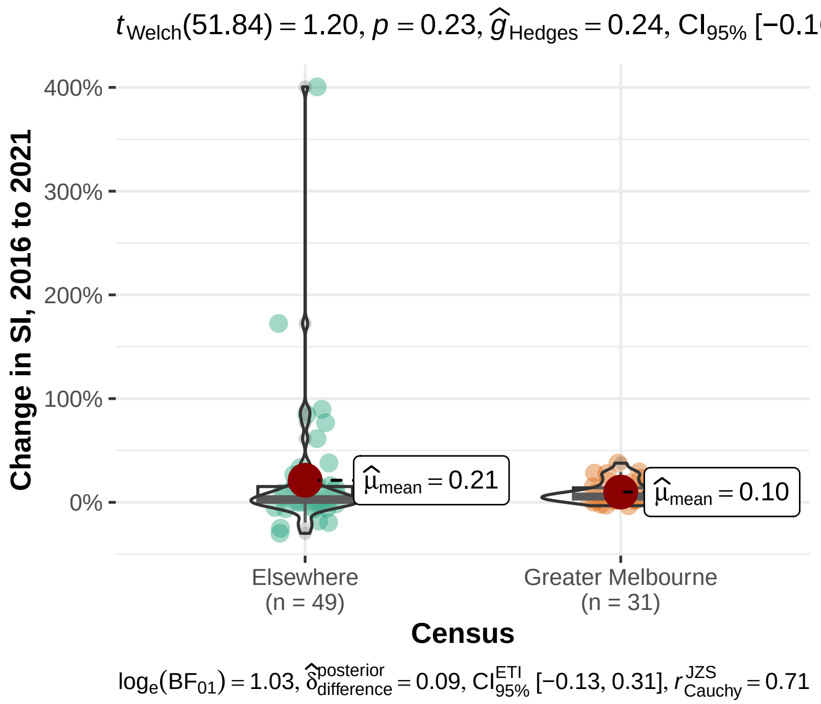


Figure 22: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ bus mode only

there were some LGAs that saw decreases in SI between 2016 and 2021 (Table 7).

The largest relative changes in SI were in East Gippsland (-30%), West Wimmera (-25%), Pyrenees (-19%) and Gannawarra (-18%). For Pyrenees this appears to have been offset by a large increase in rail service levels (+67%).

In contrast, some LGAs had relatively large increases in SI_{bus} between 2016 and 2021, as shown in Figure 8. The largest relative increases in SI_{bus} were in Baw Baw (+401%), Macedon Ranges (+172%) and Northern Grampians (+90%). Northern Grampians, lost all of its train services[53], but saw SI_{bus} increase by 90%. Horsham, likewise, saw the loss of all of its rail services[54], but an increase of 77% in SI_{bus}.

[53] Again, this may be related to the Overland.

[54] Again, this appears to be Overland related

LRT Supply Index

Figure 23 maps percentage changes in SI_{LRT} between 2016 and 2021. These are also shown in Figure 24.

Table 7: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ for those LGAs where SI_{bus} has fallen

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

Table 8: Change in SI, 2016 to 2021, by LGA, for those LGAs where SI has risen 25 percent or more

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

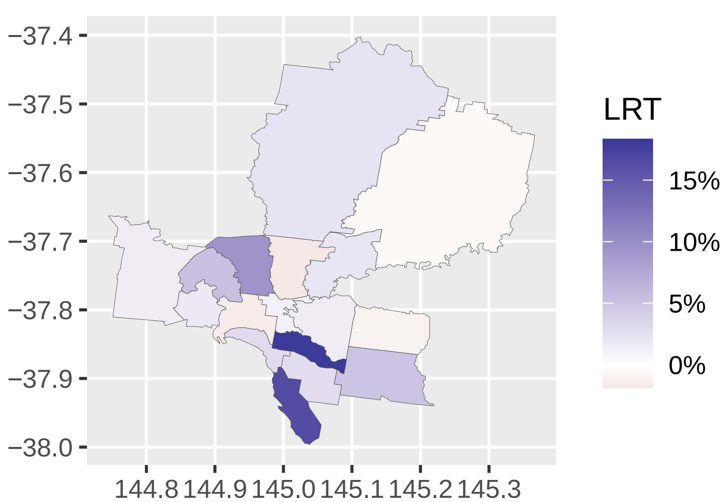


Figure 23: $SI_{LGA2021,10/8/21}$ and $SI_{LGA2021,9/8/16}$ tram mode only, violin and box plot generated using ggstatsplot package

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.

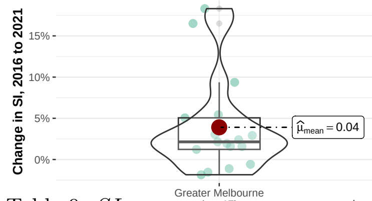


Figure 24: $SI_{LGA2021,10/8/21}$ percent change from $SI_{LGA2021,9/8/16}$ LRT mode only

Table 9: $SI_{LGA2021,10/8/21}$ 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 9 shows the 4 LGAs where SI_{LRT} fell. In general, the changes were relatively small.

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

		Change in SI			
LGA	Location	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 10 shows LGAs where SI_{LRT} increased. The largest relative increases in SI were in Stonnington (+18%) and Bayside (+17%).

Victoria, Australia: transit needs

In the following sections assessments of transport needs versus transport supply are reported.

IRSAD needs score and SI supply score

The IRSAD scores for Victorian LGAs where presented above in Figures 2 and 4. This section explores how these scores compare to the SI_{LGA} scores. Figure 25 compares the 2021 IRSAD and SI scores for LGAs across Victoria.

Figure 25 indicates a significant association between IRSAD and SI scores for LGAs in Victoria in 2021, with SI increasing with increase IRSAD. .

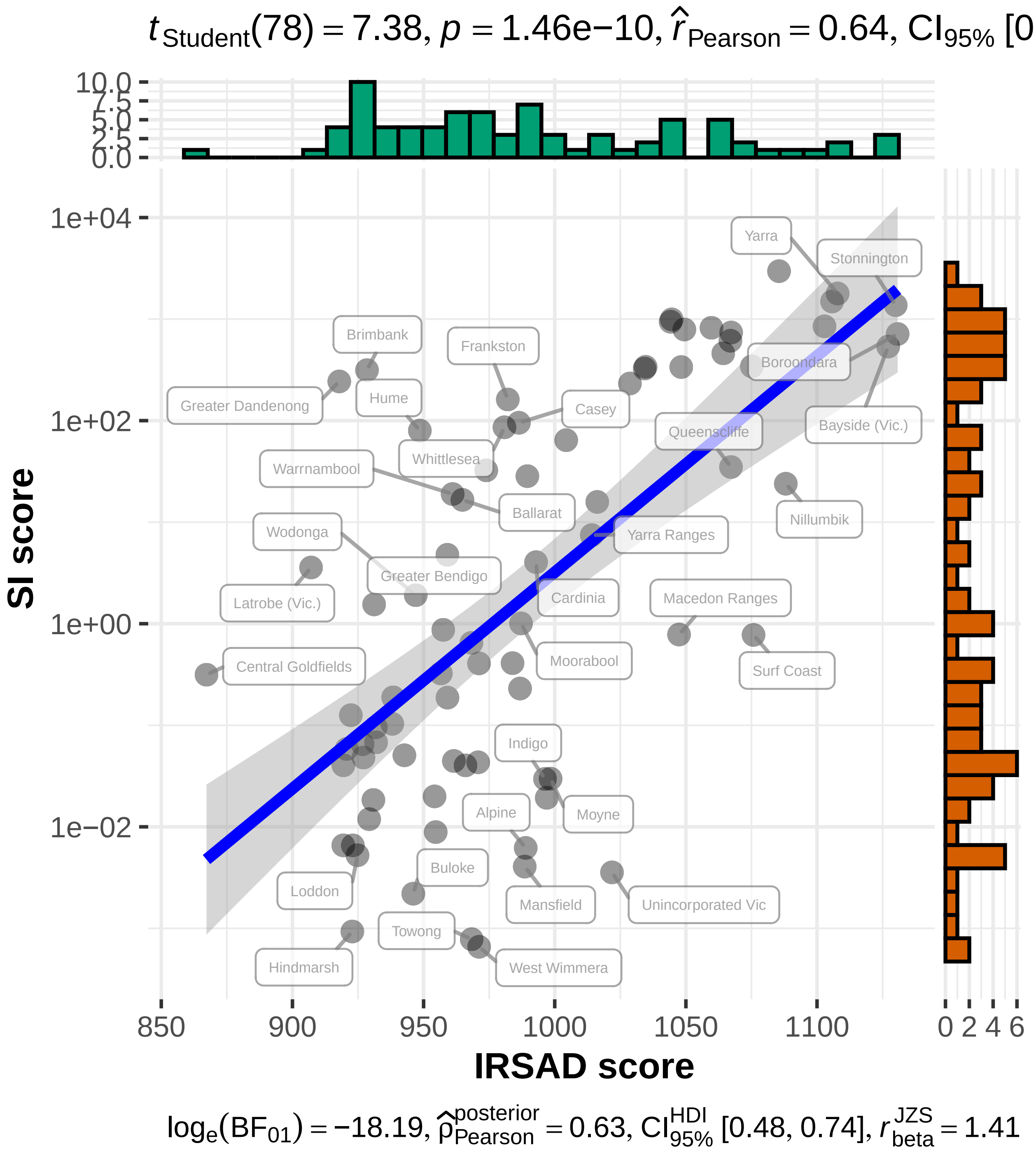


Figure 25: Relationship between IRSAD score and SI, by LGA, 2021

Population and SI supply score

Figure 26 compares population and SI score. It indicates a significant relationship, with SI score increasing with increasing population. The City of Casey and Greater Geelong appear to be notable examples of LGAs with higher populations but SI scores that appear lower than the trend.

IRSAD needs rank and SI supply rank

Figure 27 compares the rank (out of all Victorian LGAs) for each LGA's IRSAD score and SI score. It indicates a significant association between IRSAD rank and SI rank, with those LGAs that have higher SI scores (better transit) tending to have higher IRSAD ranks (implying better socio-economic conditions). Various LGAs that have SI ranks 10 or more places lower than their IRSAD rank are labeled.

Further analysis of population, IRSAD, SI rank and other needs-gaps metrics might be a direction for future research. In particular, mapping of those LGAs with low SI scores by higher needs, along the lines of the anlysis presented in Currie and Senbergs [2007], would appear to be an important next step.

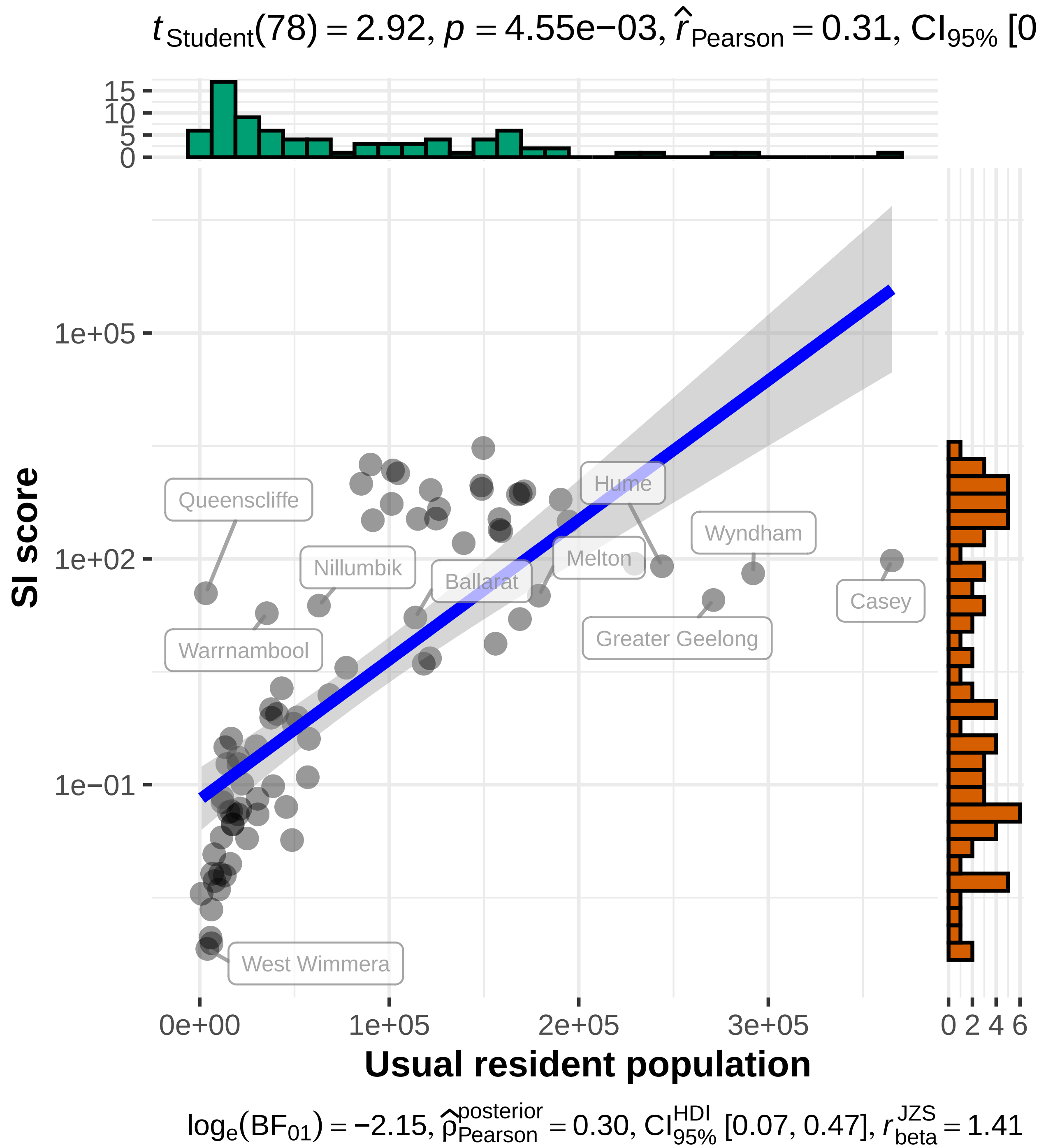


Figure 26: Relationship between population and SI, by LGA, 2021

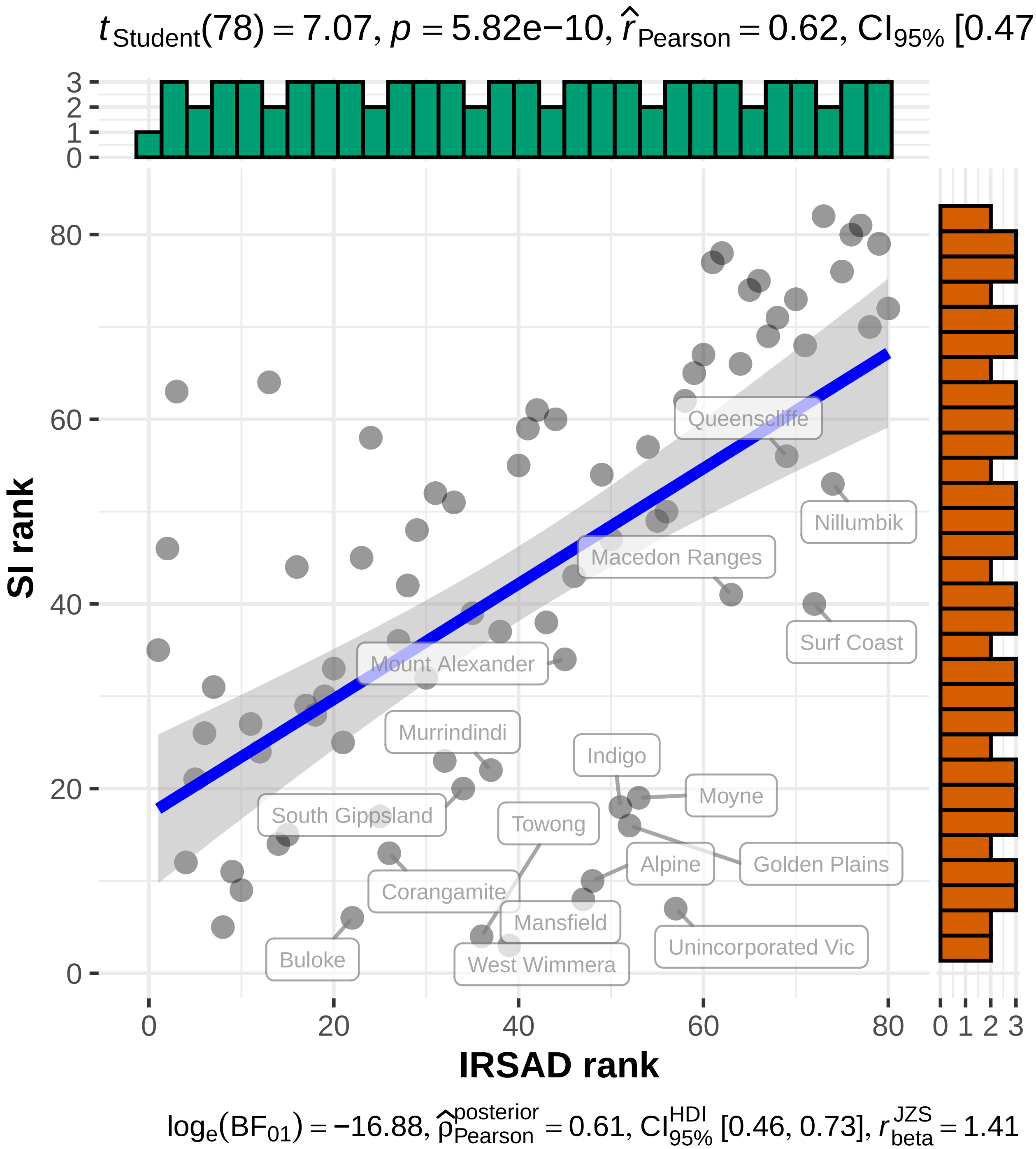


Figure 27: Relationship between IRSAD rank and SI rank, by LGA, 2021

Discussion and Conclusions

This document presents preliminary results from the development of code to calculate the Currie and Senbergs [2007] transit Supply Index (SI) directly from GTFS data. The developed code is available within the Rmarkdown file used to typesset this document, and available on github.

The code out has been verified by performing hand calculations for SA1 SI scores in the Victorian Alps (Running Creek) and Talbot. The Running Creek calculation, included two bus stops. This verification check resulted in an exact match between the hand- and code-calculated values.

The verification calculation for Talbot involved regional train, regional bus and local bus services across two bus stops and one train station. The catchment areas for these stops and station were only partially within the SA1. The hand-calculated and code-calculated SI values were reasonably similar, with the different results likely due to differing levels of precision in the assessment of areas and the lack of available timetables for the service date in question. Further verification checks, including more accurate assessment of catchment area coverage, may be a direction for future research. Similarly, additional verification calculations or close, step-by-step review of the developed code may be a direction for future research.

The developed code was used to compare transit service levels for LGAs across Greater Melbourne and the rest of Victoria at the time of 2016 and 2021 censuses. Results appear to be largely as might be expected, with more transit service coverage generally being provided in more central LGAs. There were significant differences between the 2016 and 2021 service levels across all of Victoria and for LGAs in Greater Melbourne (average SI increased), but not for LGAs in Victoria but outside Greater Melbourne.

Changes in SI scores appeared to be unevenly distributed across the state, with outer regional areas having reduced service levels, while some other LGAs, especially those in outer Melbourne, had higher SI scores in 2021 than in 2016. There were significant changes in rail and bus service levels between 2016 and 2021, but not for trams.

Some of the results suggest that the Overland train services (to Adelaide) were not included in the GTFS file in 2021. It is unclear at the moment whether this was because to the service was not operational, or simply a data accuracy issue. However, this perhaps demonstrates the challenge of using GTFS data alone for analysis of this nature: the quality of the result is reliant on the quality of the underlying data.

Comparisons with IRSAD scores and ranks, and population suggest that SI score is sificiantly associated with all three of these metrics. In general, the SI increases with increases in population, IRSAD and IRSAD rank, although it is important to note that increased IRSAD implies the LGA has less social disadvantage. This would appear to be contrary to what might be desired, with greater service provided to those areas where there is greater disadvantage, but clearly there will be many confounding factors and unclear causality involved in these relationships. That said, Casey, Wyndham and Greater Geelong appear to have particularly low SI scores given their residential population.

There is clearly much more anlaysis that can be done to expand on this code. In particular, completing mapping of needs-gaps as per Currie and Senbergs [2007] would appear a promising direction for future efforts. However, in general the verification calculations appear to support the accuracy of the developed code, and the results from the analysis suggest that output about Victorian LGAs is largely as might be expected. Next steps appear to be to use GIS to improve the Talbot verification calculations, and perhaps select a SA1 in Greater Melbourne to do verification calculation on[55].

Additionally, the code will need to be forked so as to provide outputs for Deakin^{2023 and 2018 SI scores for 7 days, using SA1 areas from 2021.} and Monash[56]. At approximately one day of processing time per day of analysis, this would appear likely to take in the order of 4-5 weeks^{Although I'm going to try to get a few different computers going on this at the same time so that it takes less 'real' time.}.

[55] This may be challenging though, as the hand-calculations are already quite involved, even for rural SA1s with only a few stops.

[56] 2021 and 2016 SI scores for 1 and 7 days, using SA1 areas from 2016.

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.