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

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

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

Abstract

This is the abstract.

It consists of two paragraphs.

Keywords: keyword1, keyword2

1. Introduction

A key reason for providing public transport services is to providing a basic level of mobility for those who cannot otherwise drive themselves. In many places, especially those with low-density landuse patterns and/or limited transit servcies, it may be difficult or even impossible to get around without a private vehicle. Disability, youth or old age, socio-economic status, lack of a drivers license, and many other factors might also mean someone is reliant on transit services for some or all of their travel beyond their local neighbourhood. Adopting social equity-focused approaches to transport planning might therefore suggest targeting services towards spatial areas where there are higher social and transport needs for transit.

Approaches for assessing spatial gaps between social needs and transit supply have been reported in previous research including Currie et al. (2003), Currie (2004), Currie and Senbergs (2007) and Currie (2010). These presented a transit Supply Index (SI), which was used to identify geographic areas in a few Australian cities where there were very high social needs for transport, but very little transit service or none at all. This approach suggested a new way for planners to identify and prioritise service provision towards locations and populations more likely to be lacking sufficient mobility to meet their basic needs or fully participate in society.

However, in almost two decades since the SI was introduced there does not appear to have been much further use or development of this approach. As well, it is unclear whether gaps between social needs and transit supply have increased or reduced in Australian cities, and if the spatial patterns identified in Hobart (Currie et al., 2003; Currie, 2004) and Melbourne (Currie and Senbergs, 2007; Currie, 2010) are representative. This may in part be because applying such methodologies elsewhere would, at the time, likely have required bespoke data collection, cleaning and analysis. Currie et al. (2003), Currie (2004), Currie and Senbergs (2007) and Currie (2010) had data provided
direct from transit operators and authorities, but at the time transit schedule data was typically only publically available in paper or PDF format.
As well, the volume of spatial and service data and bespoke processing needed to produce a social needs-gap analysis may have put such analyses beyond reach in most places.

*Corresponding author

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

[1]Research Fellow

[2]PhD Student

[3]Professor

Nowadays, however, the General Transit Feed Specification (GTFS) and the internet allows timetable data to be published in a standardized and accessible format, and
more than 10,000 transit agencies publicly release data this way. Various tools for manipulating such data are also now available, including software for validating, analysing and visualizating GTFS, as well as a separate standard (GTFS-Realtime) for sharing live vehicle locations (MobilityData, undated). However, software tools for examining spatial patterns and gaps in transit supply with respect to social needs for transport appears limited.

This gap and the lack of direct follow up to Currie et al. (2003), Currie (2004), Currie and Senbergs (2007) and Currie (2010) provide the motivation for the research reported in this paper.
The overall aim of this research is to apply the developments in spatial data science in the intervening years to social and transport needs-gaps analysis. This encompasses three main objectives: (1) to develop tools for needs-gap analysis using GTFS datasets; (2) to explore whether such gaps have changed in the intervening years; and (3) to better understand whether spatial patterns in Melbourne and Hobart are representative of those in other places.

This research included the development of a new R package (gtfssupplyindex) for calculating SI scores. Also presented in this paper are results for Australian cities in 2016 and 2021, matching the most recent censuses, allowing comparison to the 2006 analysis reported in Currie (2010) and between locations.

The remainder of this paper is structured as follows: the next section outlines the background to this research, including the formulation of the Transit Supply Index (SI). Section 3 describes the study methodology, followed by presentation of results in Section 4 and discussion in Section 5. Limitations of this study, directions for future research and a brief conclusion are provided in Section 6.

2. Background

2.1. Transit metrics

There are many metrics available for benchmarking transit services, including: those in the Transit Cooperative Research Program (TCRP) Report 88 guidebook on developing performance-measurement systems (Ryus et al., 2003) and in the Transit Capacity and Quality of Service Manual (TCQSM)(Kittleson & Associates et al., 2013); those used across benchmarking databases and programs (e.g. Florida Transit Information System (2018), International Association of Public Transport (UITP) (2015) Imperial College London (undated); and the results available on the .Transit Score website (Walk Score, 2023). The Fielding Triangle (Fielding, 1987) provides a framework for combining indicators of service inputs, outputs and consumption to describe cost efficiency, cost effectiveness and service effectiveness. More broadly: Litman (2003) and Litman (2016) discuss some of the traffic, mobility, accessibility, social equity, strategic planning and other rational decision-making-based perspectives underling transport indicators; Reynolds et al. (2017) extends these into models of how institutionalism, incrementalism and other public policy analysis concepts might apply to decision-making processes relating to transit prioritization; Guzman et al. (2017) developed a measure of accessibility in the context of policy development and social equity for Latin American Bus Rapid Transit (BRT) networks; and Creutzig et al. (2020) introduced street space allocation metrics based around ten ethical principles.

However, many of these metrics may be difficult to calculate, explain or understand, especially for those who are not planners, engineers or other technical specialists. Where pre-calculated transit metrics are immediately available, such as on a website or other online platform, it may not be possible to independently generate scores, for instance to assess proposed system changes. Contrasting examples are provided by the metrics in the TCQSM and the Transit Score metric (Walk Score, 2023):

- Transit Scores are readily available on the Transit Score website for locations with a published GTFS feed. The meaning of these Transit Scores also appears easy to explain, with the highest possible score of 100 representing the sort of transit accessibility that might be xperienced in the center of New York. However, the Transit Score algorithm is a black box, and scores cannot be calculated independently or generated for proposed changes to networks.

- the TCQSM provides a wide range of metrics for measuring different aspects of a transit system. The TCQSM scores themselves appear easy to understand or explain, ranging from A (good) to F (bad), although the large number of metrics might be somewhat overwhelming for some users. The scores, however, can be calculated independently, given sufficient data.

Wong (2013) provides an example of what can be done with GTFS data, open metrics and coding by reporting the distribution of various TCQSM metrics across 50 USA transit operators. Code used in the Wong (2013) analysis is available for those who might wish to produce a similar study for other locations and systems. Producing code that allows similar calculations from GTFS data, except for the SI metric, is included in the objectives of this study.

2.2. The Transit Suppy Index

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

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

where:

- $SI_{area,time}$ 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 (2010) 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.

Currie (2010) reported SI scores for Census Collection Districts (CCDs) across Greater Melbourne in 2006, as shown in Figure 1, and identified a general pattern of: more transit supply in the middle and inner suburbs and along passenger railway lines; and outer areas tending to have very low SI scores or no transit supply at all.

2.3. Social need and needs gap

As well as measuring transit supply, Currie (2010) also assessed the social need for transit across Greater Melbourne using: the Australian Bureaus of Statistics' Index of Related Socio-Economic Advantage/Disadvantage (IRSAD) and a transport needs index derived from eight weighted indicators. The spatial distribution of this composite social needs index in 2006, reproduced in Figure \ref{fig:Currie_maps_needs), indicates areas of above average, high and very high social needs in 2006 were located in: some outer areas, particularly in the east and south-east; and in some middle areas in the south-east, north and west. Currie (2010) also identified areas with very high transport needs, but very low or no transit supply,
as reproduced in Figure 3. This indicated areas where service gaps might be of particular concern, Most of these were located in outer parts of Melbourne in the north-east, south-east and south, although there were also some pockets in the middle suburbs in the west, north and south east.

Overall Currie (2010) found that "8.2% of Melbourne residents have 'very high' needs but 'zero', 'low' or 'very low' public transport supply." Using this methodology in transit planning appraoch was also stated as being "substantially more useful than the presentation of anecdotal evidence, which is the most common means of identifying transport needs in local transport studies throughout the world."(Currie, 2010)

However, it does not appear that this approach has been widely adopted in practice or further developed by researchers. Our suspicion is that while the SI has a relatively simple formula and requires only geographic and timetable data, the lack of a software tool to perform these calculations may be part of the reason that it has not been more widely adopted. It is also unclear whether the patterns in Melbourne identified in Currie (2010), have changed since the 2006 analysis. Developing a software tool to calculate SI tools from GTFS data, and then using it to comparing current conditions and other locations to the findings of Currie (2010), therefore, provides motivation for this research.

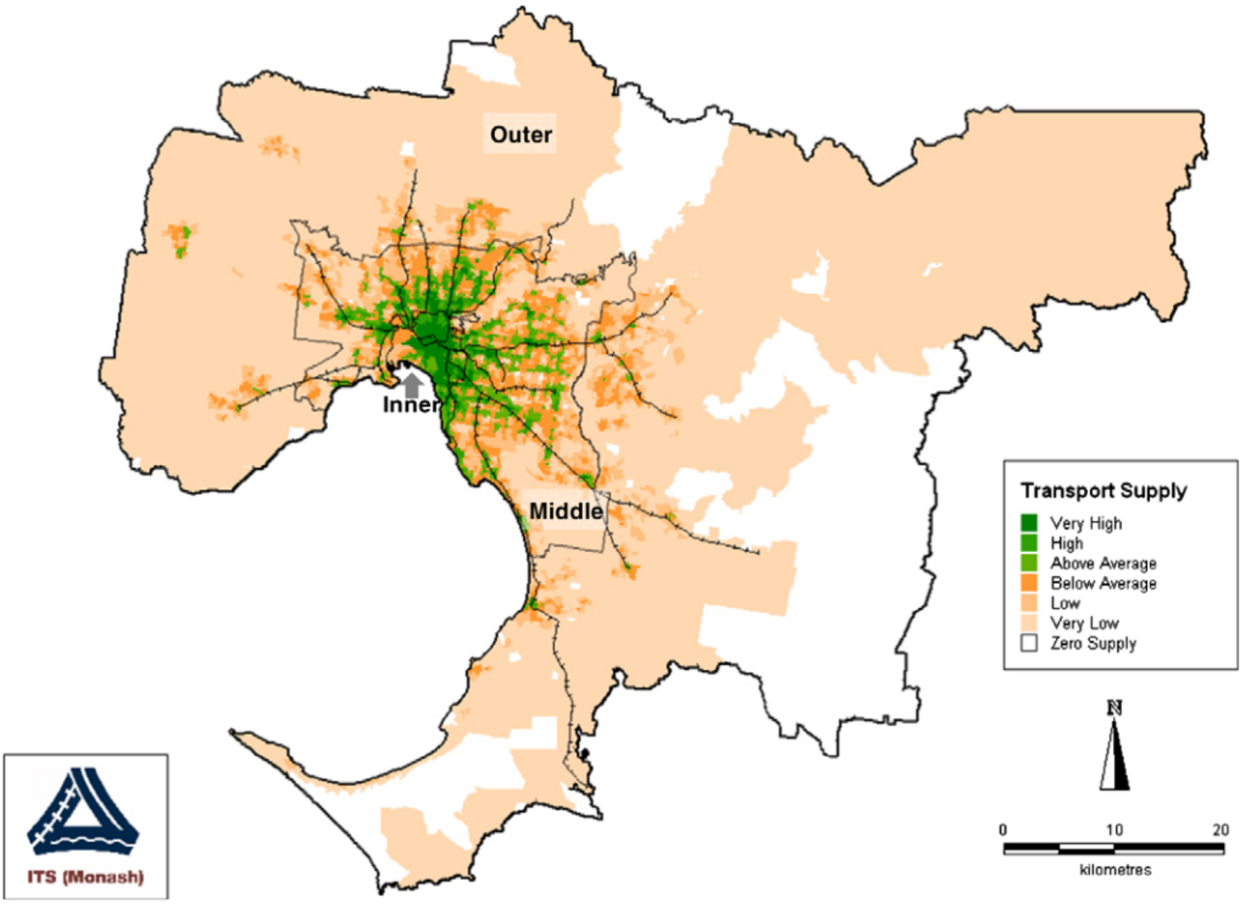

Figure 1: Distribution of supply measure scores – Metropolitan Melbourne (2006), Source: Currie (2010)

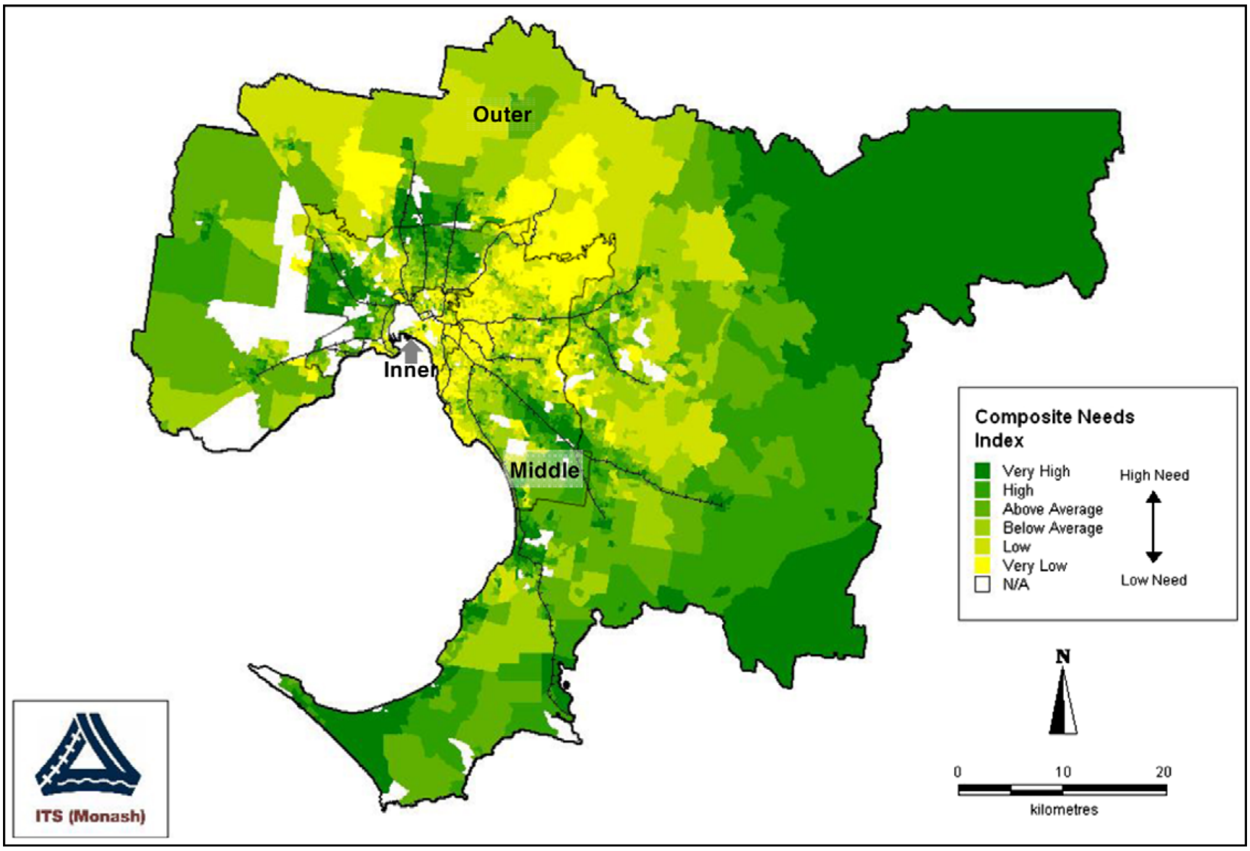

Figure 2: Distribution of categories of composite social need index scores in 2006, Source: Currie (2010)

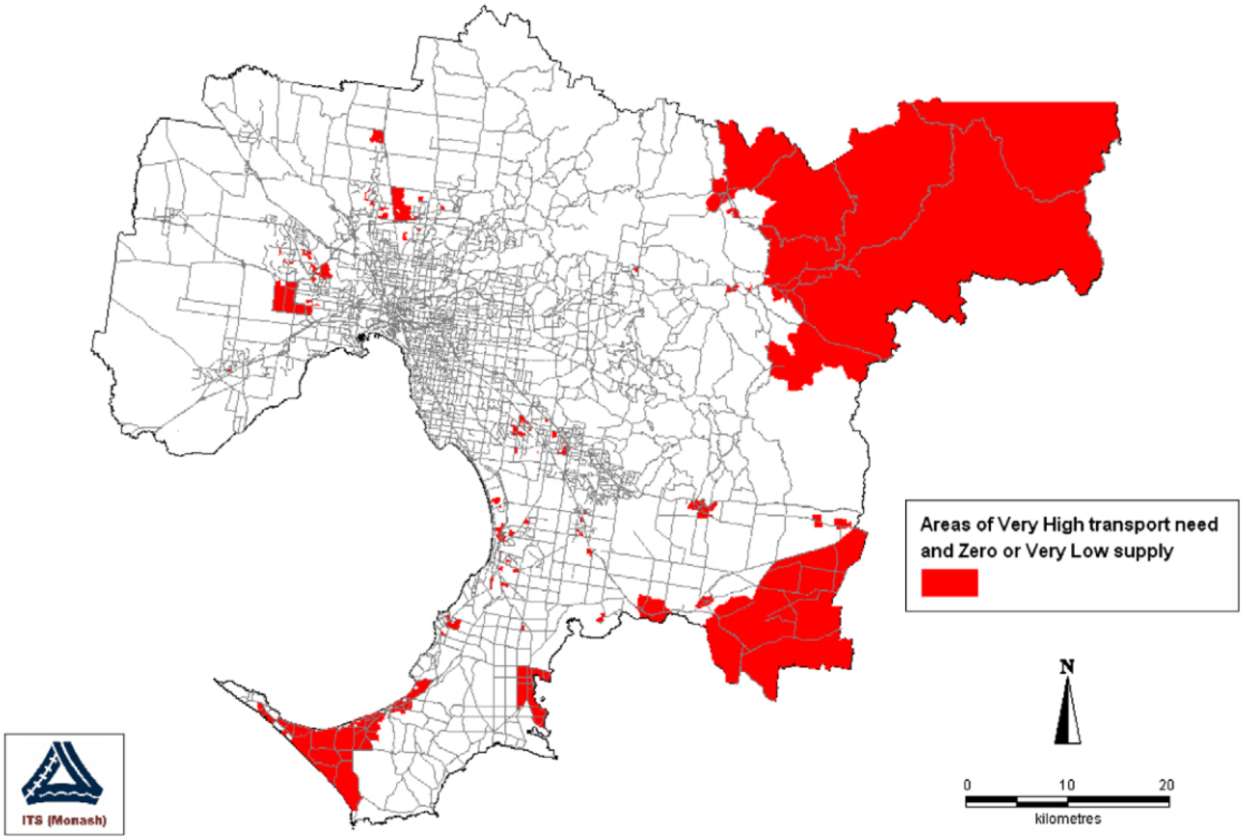

Figure 3: Melbourne needs-gap in 2006 – very high transport need areas with zero or very low public transport supply, Source: Currie (2010)

3. Methodology

3.1. Code development

This study developed a package of tools for calculating the SI from GTFS data using the R programming language (R Core Team, 2023). The recommendations of Wickham and Bryan (2023) informed the package setup and development approach. Various existing packages and code examples were relied upon including: the sf package (Pebesma, 2024) for geospatial analysis; the tidyverse (Wickham et al., 2019); gtfstools (Herszenhut et al., 2022); and tidytransit (Poletti et al., 2023). Australian Bureau of Statistics (ABS) data was also used, sourced via the strayr and absmapsdata packages (Mackey et al., 2023).

Code was developed and tested on the Mornington Peninsula Tourist Railway GTFS feed. This was selected primarily for convenience, given that the authors are familiar with the surrounding geography and that the feed covers a small number of trips across just three stations.

Statistical Area 1 (SA1) zones
were adopted from the Australian Bureau of Statistics (Australian Bureau of Statistics, undated) as the areas of interest. SA1 zones are the smallest geographical areas for which results are reported in the Australian census. SA3 zones, which are generally similar to Local Government Area (LGA) boundaries, where used when aggregating.

3.2. Changes since 2006: Greater Melbourne

Much has changed since 2006, including the spatial geography used by the Australian Bureau of Statistics (ABS) to collect census data. To allow direct comparison between 2006 and now, therefore, this study first calculated SI scores for the week starting the day of the 2021 census using the same Census Collection Districts (CCDs) used by Currie (2010). The Victorian GTFS feed, published by Public Transport Victoria (PTV), was used, with historical feeds sourced via Transit Mobility Data, (2023).

Unfortunately, it is not possible to obtain 2016 or 2021 social disadvantage data for CCDs, as the ABS now releases data for SA zones only. SA1 zones, therefore, are adopted for the comparison of social needs-gaps in 2016 and 2021, as discussed in the following.

3.3. Variation in spatial patterns across location.

SI scores were also calculated
for the weeks starting the day of the 2016 and 2021 censuses in Greater Melbourne, Greater Brisbane, Greater Perth, Greater Adelaide and Greater Hobart. Historical GTFS data was again sourced via the Transit Feeds website, Unfortunately it was not possible to locate historical data for Greater Sydney, so the latest data set was sourced directly from Transport for NSW. SA1 zones were adopted as the areas of interest so as to allow direct comparison to ABS-reported values for social needs, transport needs, population and other census data.

3.4. Measuring social disadvantage

This study adopts a similar approach to measuring social disadvantage as used in Currie (2010), using: the ABS' Index of Relative Socio-Economic Advantage/Disadvantage (IRSAD); and a transport needs index[4]. A composite needs indicator was derived based on the IRSAD and the transport needs index, again as per the Currie (2010) approach[5].

[4]The same need indicators and weightings used in Currie (2010) were adopted, although $799 or lower per week was used as the threshold for low income households rather than $499 to account for inflation (as per Reserve Bank of Australia's online inflation calculator).

[5]

However, changes to the ABS reporting systems mean that the composite needs indicator had to based on weighting both the IRSAD index and the transport need index by the total population of each SA1 zone, which were then added, standardised and split into six groups.

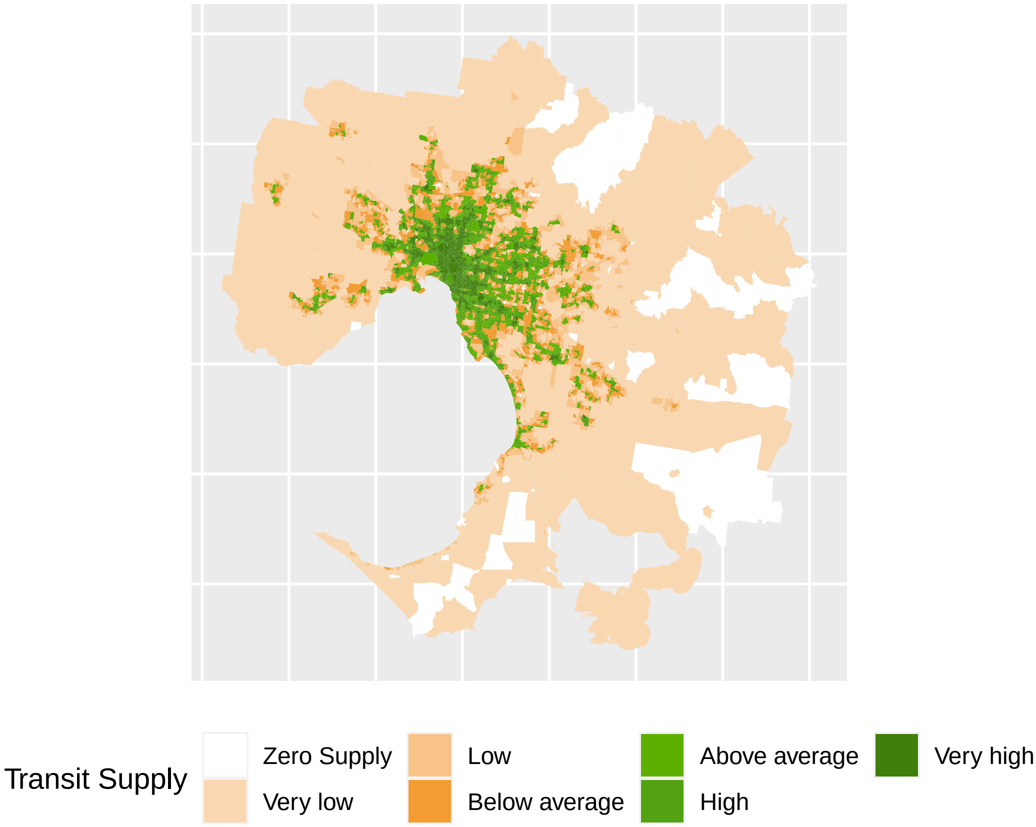

Figure 4: Transport Supply by CCD, weeks starting the dates of the 2016 and 2021 censuses

4. Results

4.1. The gtfssupplyindex Package

Code developed to calculate SI scores is available as an R package on github as Reynolds (2024). Included in the package is a vignette that outlines the structure of the calculations, the developed functions (LINK HERE), The vignette also includes step-by-step calculations for the Mornington Peninsula Railway that provide a worked example and comparison to SI scores calculated manually.

4.2. Changes since 2006: Greater Melbourne

```
## Reading layer ‘census_collection_district_2006’ from data source
##   ‘/Users/jreynold/Documents/0001_project/gtfssupplyindex_main_paper/data/asgc2006.gpkg’
##   using driver ‘GPKG’
## Simple feature collection with 38612 features and 41 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 96.81677 ymin: -43.74051 xmax: 159.1092 ymax: -9.142176
## Geodetic CRS:  GDA94
```

Figure **??** shows SI scores
for Melbourne in the week of the 2021 census, using the same (2006) CCD boundaries as in Figure 1. In general, the 2021 spatial patterns appear similar to those in 2006, with higher levels of transit supply in

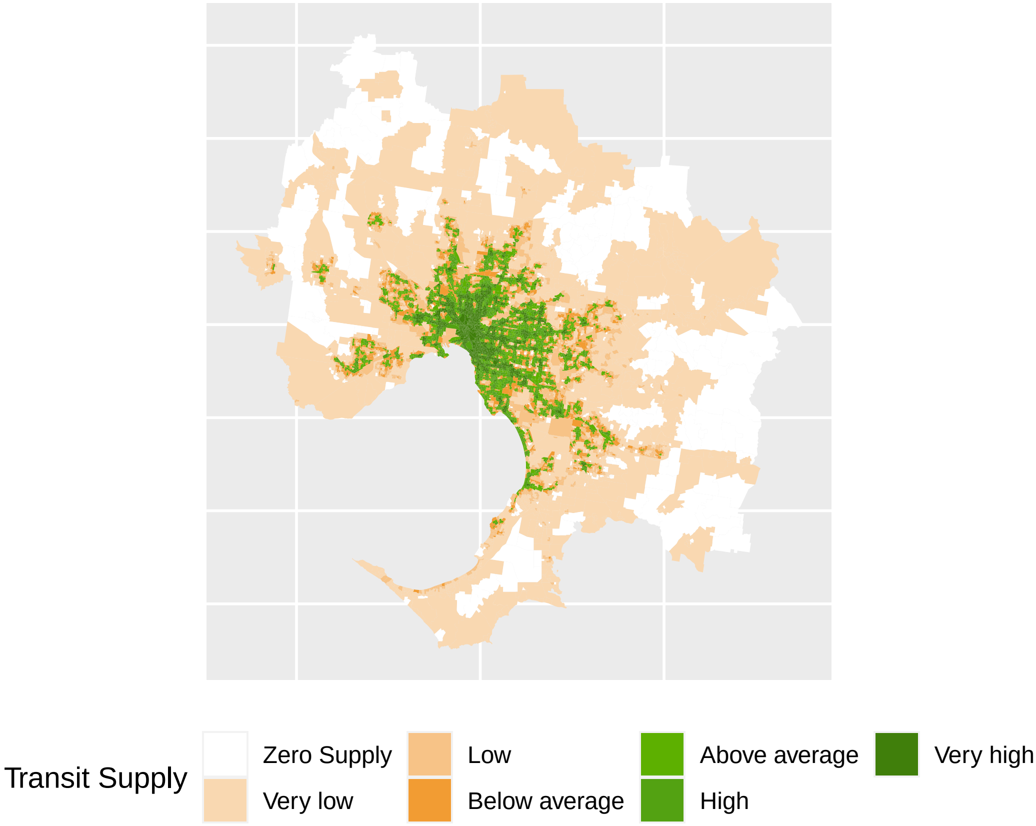

Figure 5: Transit Supply by SA1, weeks starting the date of 2021 census

inner areas and close to railway lines. There are still some areas with zero transit supply in outer areas, although fewer than in 2006.
Transit supply also appears to be more spread out in 2021 than in 2006, with higher levels of transit supply provided in many middle and outer locations.

4.3. Variation in spatial patterns across Australian cities

<NEED TO ADD OTHER CITIES. MIGHT BE BEST TO CHANGE ALL OF THE NEXT CHUCK INTO A FUNCTION TO THEN APPLY TO EACH CITY IN TURN>

Figure **??** shows SI values for the week starting on the day of the 2021 census for all Australian Capital Cities except Greater Sydney, for which the SI values are calculated for the week starting .

4.3.1. Social needs

Figure **??** shows the distribution of categories of social need index scores across Greater Melbourne for 2021. This figure is analogous to the 2006 value from Currie (2010) shown in Figure 2 although, as discussed in the methodology section above, it was not possible to exactly replicate the Currie (2010) approach as the total number of people within one or more of social need component groups (necessary to calculate the two relative need indicators) are not reported in the 2021 census.

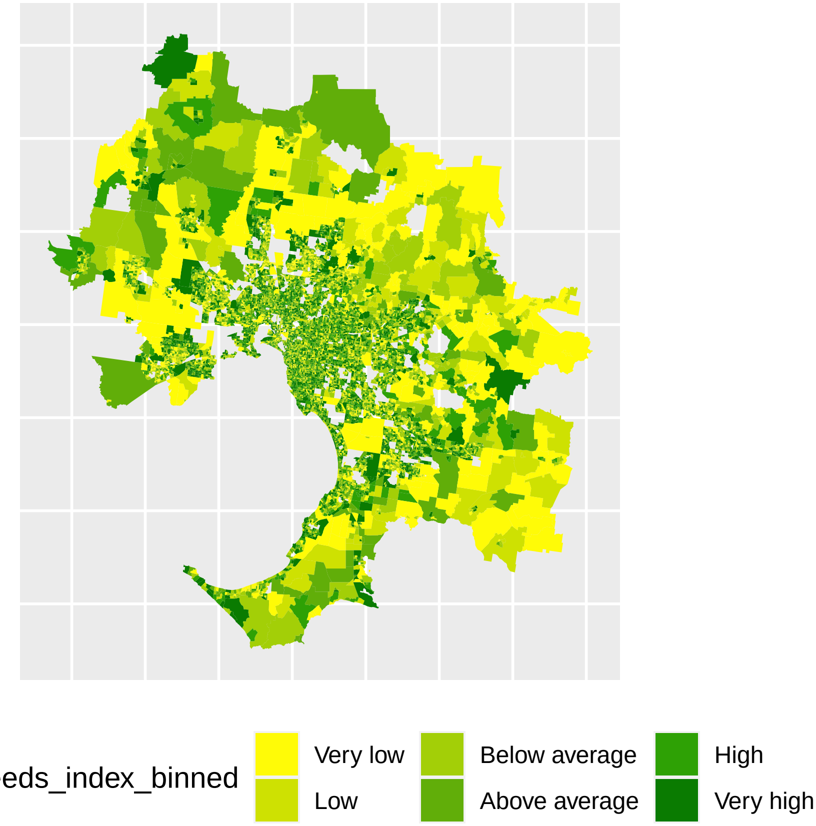

Figure **??** appears to indicate that there is no clear spatial pattern to the distribution of the categories of the composite need index scores. This appears to contrast to trend, albeit with some exceptions, towards very high social needs scores in outer areas identified by Currie (2010) (Figure 2). This may, however, be an artifact of either the differences in the composite needs scores used in this analysis (due to the lack of data to assess relative needs) compared to the Currie (2010) analysis. As well, the 2021 census SA1 zones generally appear to smalller than the 2006 Census Collection Districts (CCDs) used in Currie (2010), especially in outer areas. This will be associated with the growth of Greater Melbourne's population and spatial dispersment, with many of the large outer 'Very High' CCDs shown in the 2006 map now split into many more SA1 zones.

4.3.2. Needs-gap analysis

```
##  [1] Very high     High          Above average Below average Low
##  [6] Very low      Very high     High          Above average Below average
## [11] Low           Very low      Very high     High          Above average
## [16] Below average Low           Very low      Very high     High
## [21] Above average Below average Low           Very low      Very high
## [26] High          Above average Below average Low           Very low
## [31] Very high     High          Above average Below average Low
## [36] Very low      Very high     High          Above average Below average
## [41] Low           Very low      Very high     High          Above average
## [46] Below average Low           Very low      Very high     High
## [51] Above average Below average Low           Very low      Very high
## [56] High          Above average Below average Low           Very low
## [61] Very high     High          Above average Below average Low
```

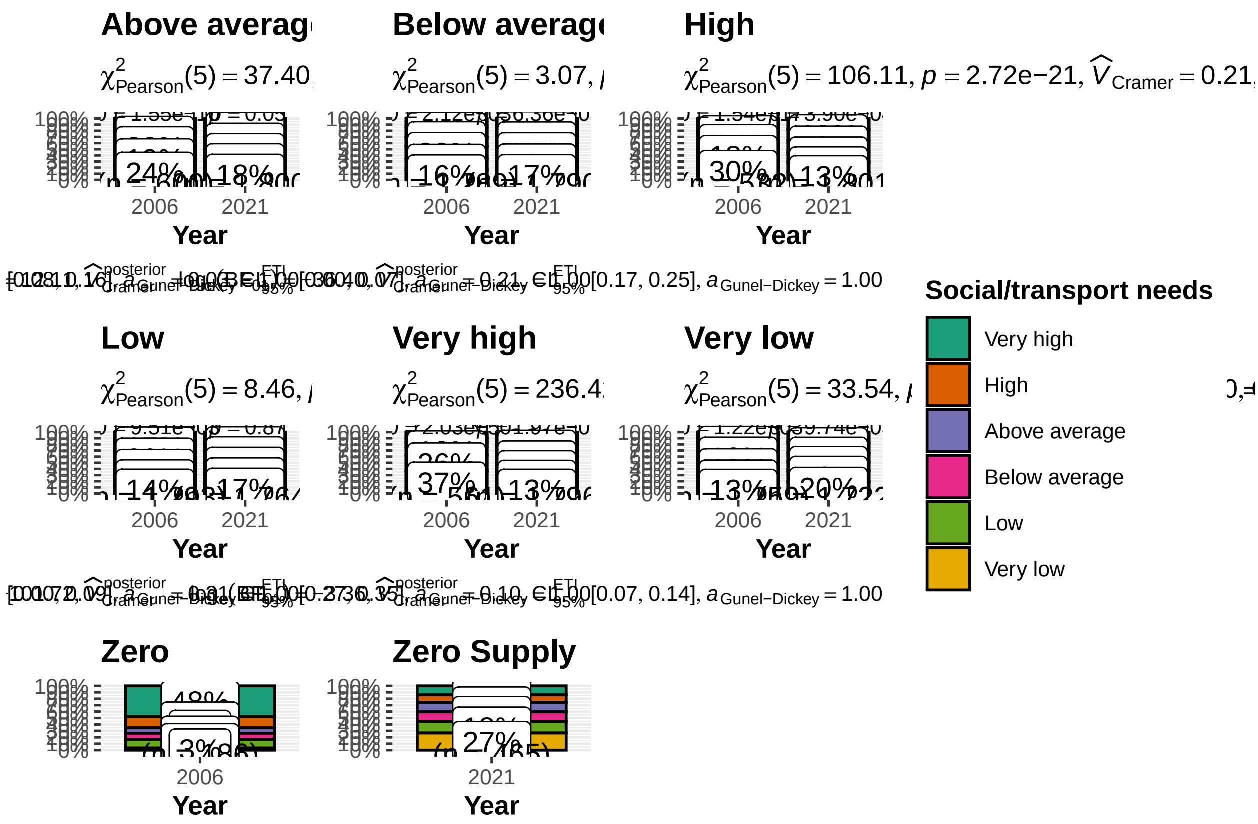

Figure 6: Number of areas by SI and social/transport need category, comparison between 2006 and 2021

```
## [66] Very low       Very high      High           Above average Below average
## [71] Low            Very low       Very high      High          Above average
## [76] Below average Low             Very low       Very high     High
## [81] Above average Below average Low              Very low
## Levels: Very low Low Below average Above average High Very high
```

Figure **??** compares the number of zones[6] by SI and social/transport need category between 2006 and 2021. Population is compared in Figure **??**

These results indicate that, in 2021, 65 SA1 zones, (representing 0.6%of the total 11,138SA1 zones in Greater Melbourne) had no public transport, but very 'very high' social/transport needs. These zones have a combined population of

4.4232×10^4 SA1 zones, representing 44,232(0.9% of the total 4,885,773 population) live in areas with no public transport but have 'very high' social/transport needs. This compares to the 89 CCDs (1.6% of the 5,720 total), representing 37,699 Melbourne residents (1.1% of the population) living in areas with no public transport, but very high social/transport needs reported for 2006 in Currie (2010).

4.4. Variation in spatial patterns across time: 2016 and 2021

INCLUDE GINI COEFFICIENT PLOTS HERE

[6]SA1s in 2021, versus CCD in 2006.

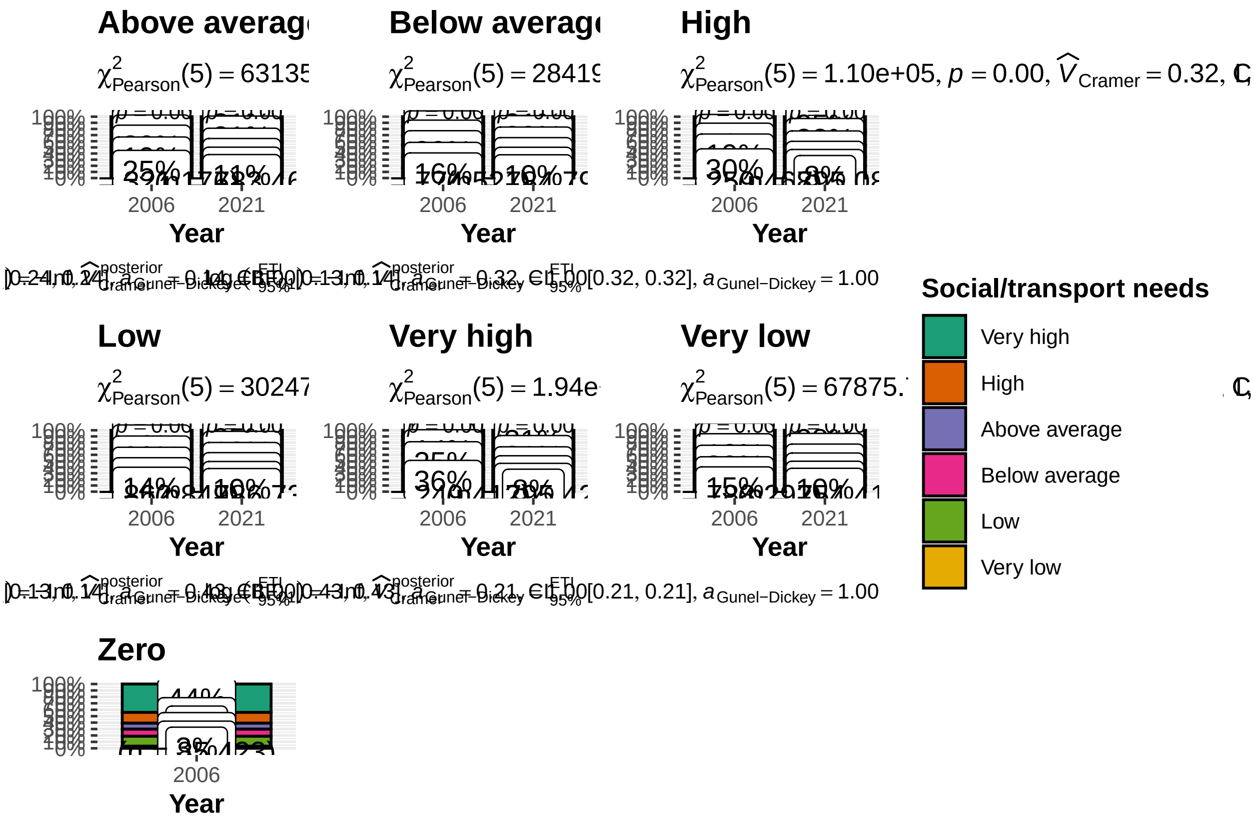

Figure 7: Population by SI and social/transport need category, comparison between 2006 and 2021

5. Discussion

5.1. Limitations

5.2. Directions for furture research

6. Conclusions

References

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

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

Graham Currie. Gap analysis of public transport needs:measuring spatial distribution of public transport needs and identifying gaps in the quality of public transport provision. *Transportation Research Record*, 1895:137 – 146, 2004. doi: 10.3141/1895-18. URL https://www.semanticscholar.org/paper/428fdef8b83dffb26b1ec4ff7e709e23e4dd8dc8.

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

Graham Currie and Zed Senbergs. Identifying spatial gaps in public transport provision for socially disadvantaged australians: the melbourne 'needs gap' study. 2007. URL https://www.semanticscholar.org/paper/9759acfacc8f91558157a7da8b85ad1847f35173.

Graham Currie, David Enright, Craig Hoey, and D. Paterson. Quantitative approaches to needs based assessment of public transport services: The hobart transport needs gap study. 2003. URL https://www.semanticscholar.org/paper/2c049091cafb56c66efc532ad2bdd774d8efc0eb.

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

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

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

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

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/.

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

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

Edzer Pebesma. *sf: Simple Features for R*, 2024. URL https://r-spatial.github.io/sf/. R package version 1.0-16.

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

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

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

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

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

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

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

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

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

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