Transit Supply Index scores on the days of the 2016 and 2021 censuses: using Statistical Area Level 1 (SA1) 2016 boundaries

James Reynolds and Graham Currie

2023-08-23

Introduction

Previous research by Currie and Senbergs [2007] developed a transit Supply Index (SI), 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. The Public Transport Research Group (PTRG) has been developing R code to calculate this Supply Index directly from GTFS data.

This document describes results from using the code to output SI scores for the day of the 2016 census and the day of the 2021 census for each of the Australian Bureau of Statistics (ABS) 2016 Statistical Area Level 1 (SA1) zones in Victoria[1]. It also presents verification checks to determine the accuracy of the output scores, and shows some statistical analysis of the scores as a way of exploring the output.

[1] These scores have been requested by Maryam Jafari as an input to her PhD project.

This rest of this document is structured as follows: the next section discusses the research context of the Supply Index. In the third section the methodology for the code development is outlined, including discussion of the overall context of the Victoria, Australia, case for which the scores are being calculatted. It also discusses the individual (sub-)cases within Victoria^[An SA1 in the Alpine are, an SA1 in Talbot (near Ballarat), SA1s within the Clayton SA2 area and SA1s witihin the Greater Melbouren SA3 area) used to verify and explore the scores outputed by the code. In the fourth section results are presented, including the verification results, exploration of the scores across Clayton and Melbourne City, and a review of the SI scores for SA1s across all of Greater Melbourne and all of Victoria. Mode-by-mode SI scores are also explored, followed by an examination of needs-gaps across the ABS Index for Relative Socio-Economic Advantage/Disadvantage (IRSAD) scores, ranks and population levels. The document then closes with a brief discussion and conclusion section.

Research context

The Suppy Index

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

Equation 1[2] shows the Supply Index[3]. 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 Transit Capaicity and Quality of Service Manual (TC-QSM) [Kittleson & Associates et al., 2013] and other transit supply metrics. However, including such factors may increase the complexity of calculating and describing the index.

Simplicity is also helped by the way that the SI is additive. Hence, $SI_{area,time}$ scores can be aggregated to calculate an overall score across multiple time periods or for a region encompassing multiple areas of interest.

Currie and Senbergs [2007] calculated the $SI_{area,time}$ for various Census Collection Districts (CCDs)[4] 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.

[2] 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.

[3] 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.

[4] 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 developed a Transport Needs Index based around population, transport and employment data. This was combined with the ABS' Index for Relative Socio-Economic Advantage/Disadvantage (IRSAD) to produce a combined index addressing social and transport needs. This was then used to assess the gap between transit needs and supply.

Later in this document a similar needs-gap analysis is presented for Victorian SA1s. 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.

Methodology

This document has been prepared using Rmarkdown, which allows the intermingling of written text, code segments and code outputs. Code segments developed in this research are shown in the following, together with the relevant descriptive text[5]. This specific document is part of a branch of the developed code, specifically created for reporting the calculation of the SI scores for the SA12016 zones.

[5] The Rmarkdown file is available at `https://github.com/James-Reynolds/Transit_Supply_Index_GTFS/tree/SA12016_analysis` 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.

Various analysis tools are available that make use of GTFS data, including the tidytransit package [**?**] for the R statistical programming language [R Core Team, 2023]. Poletti [undated] provides code to calculate a departure timetable from a GTFS feed, and this was adapted to calculate arrivals at a stop and the $\mathrm{SL_{Bn}}$ term in the Currie and Senbergs [2007] SI equation.

```
##
## Attaching package: 'tidytransit'

## The following objects are masked from 'package:gtfstools':
##
##     get_trip_geometry, read_gtfs, validate_gtfs, write_gtfs
```

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][6].

[6] 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.

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[7].

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

The $\mathrm{SI_{area}}$ 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 ($\mathrm{Area_{Bn}}$) 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 calculations are repeated until all stops and modes in the GTFS file have been included.

Case research approach

This document reports results for a single case, the state of Victoria. The state capital is in Melbourne, which has a similar metropolitan area to of Paris or London[8]. 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.

[8] 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)[9].

[9] 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].

For the results reported here output was obtained for SA1-level areas using GTFS files from August 2016 and 2021, running for just the day of the census in each year. 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[10]. Minor corrections were made to the GTFS files to remove duplicate stop_ids[11].

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

[11] 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 SA1 boundaries for Victoria used during the 2016 census (SA1_2016)[12]. The EPSG:28355 transform [EPSG, 1995] was used to shift longitude and latitude into

[12] Note, there is also a set of SA1 boundaries that are relevant to the 2021 census (SA1_2021).

metres, as per the Geocentric Datum of Australia 1994 (GDA95 / MGA zone 55) coordinates.

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].

Sub-cases

The developed code has been separately tested using SA1- and LGA-level areas of interest, including hand verification of some example SA1 areas[13]. Here the results are examined in detail for:

- SA1 zone 20403106915, which covers Running Creek and Morgans Bridge, two localities in the Victorian Alps[14]. Within this SA1 area there are only two V/Line bus stops[15]. 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 Figure 1.

[13] This testing is reported in the main branch of the project on GitHub. This document is instead specific to the calculation of the SI scores for the 2016 SA1 boundaries.

[14] The aforementioned hand verification of example SA1 areas reported in the main branch of the project on GitHub also examined SA1 zone 20403106915. This is repeated here for consistency.

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

Figure 1: SA1 20403106915 with approximate location of Stop:ID 45125 highlighted, sources ABS and Google Maps

- Output results for SA1 zones within the Clayton Statistical Area Level 2 (SA2) were also examined. This SA2 includes the Monash University Clayton campus, and is selected because it is familiar to authors and likely readers.

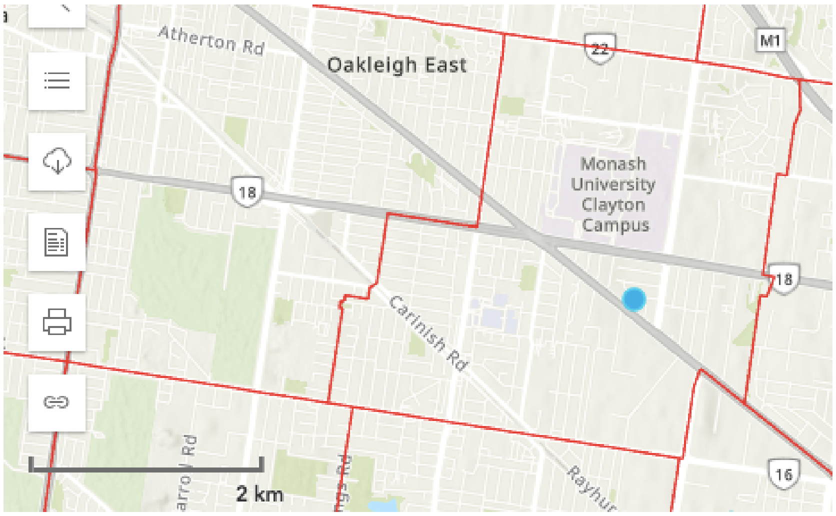

Figure 2: Clayton SA2 zone. Source ABS.

- Output results for SA1 zones within the Melbourne City Statistical Area Level 3 (SA3) were also examined. This SA3 includes Melbourne CBD, north Melbourne, Royal Park, Carlton, East Melbourne. parts of South Yarra and Prahan, and Southbank. Again, it has been selected for familiarity so as to help assess the accuracy of the reported SI results.

Transit needs: ISRAD scores

The ABS provides IRSAD datasets for SA1 in excel format. Data for 2016 and 2021 was included in this study[16].

ISRAD Scores for 2021 are shown in Figure 4 for the Melbourne City and Clayton case study areas.

The ISRAD scores appear to meet expectations. In Clayton there is no score reported for the Monash University campus, Likely due to its low resident population. Higher scores are reported for SA1s to the north-east of Clayton, in the Notting Hill area.

For the Melbourne City SA3 zone, there are no ISRAD scores reported for the various parks[17]. Areas with lower IRSAD scores are located in parts of North Melbourne and in the north-east [18], which meets expectations.

[16] The IRSAD scores for 2021 appear to only be available for the SA12021 boundaries. A correspondance file was used to match the SA12021 to SA12016 boundaries, but it was not possible to factor the scores.

[17] Royal Park to the north, Fitzroy Gardens and the sporting district to the east and south east, and the SA1 with the Docklands stadium in it.

[18] Lygon Street and Nicholson Street areas.

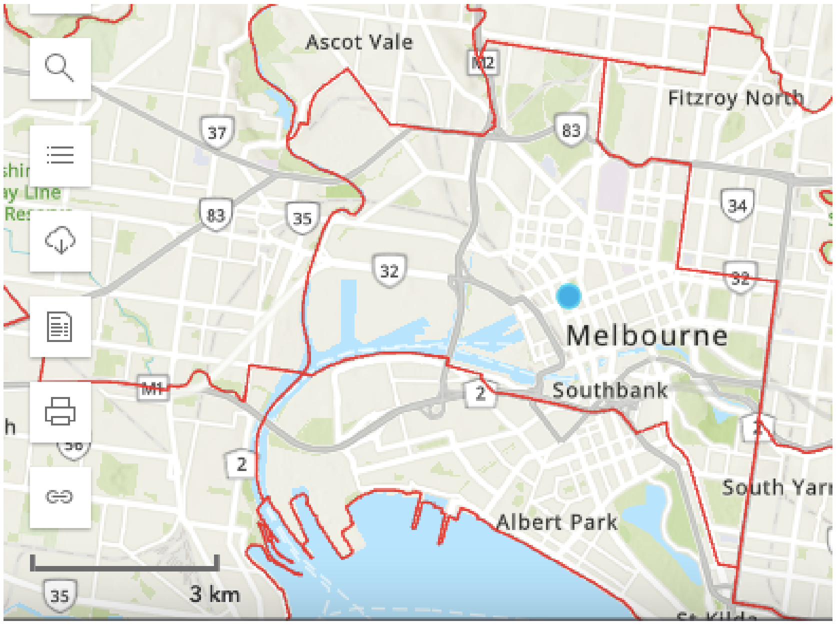

Figure 3: Melbourne City SA3 zone. Source: ABS.

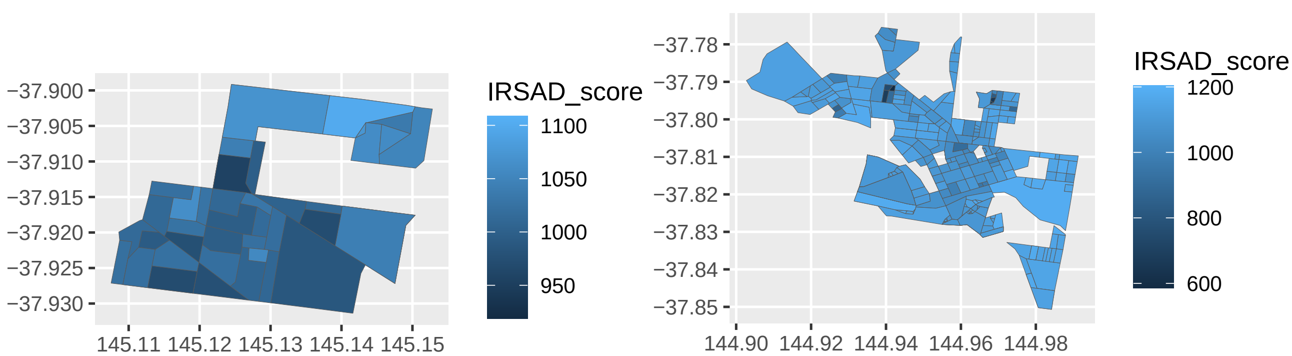

Figure 4: 2021 IRSAD score for SA1s: within the Clayton SA2 zone (left) and within the Melbourne City SA3 zone (right)

Results

The following subsections discuss the results of running the code for all of Victoria for 2016 and 2021, and the looks in detail at the two selected cases (Clayton and Melbourne City) as a validation of the results.

Supply Index results for all of Victoria

Four files are output by this document[19]. These include the total 2016 and 2021 SI scores for each SA12016 zone[20], and the SI scores by mode[21].

[19] Located in the 'results' subdirectory of the SA12016 branch in the github repository.

[20] Victoria_2016_SA_SA2016_160809.csv and Victoria_2021_SI_SA12016_210810.csv

[21] Victoria_2016_SI_df_by_mode...etc.

Verifing the code output: Running Creek and Mongans Bridge, Kiewa Valley Hwy

Code output results were verified by comparison to by-hand calculations for the SA1 area 20403106915. Within this SA1 area there are only two V/Line bus stops[22]. 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. Relevant geographic statistics are shown in the following.

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

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 $(\pi 400^2)/284598000 = 1.77$e-03.

The number of services stopping at the two stops were extracted from the GTFS files, using instructions provided by Poletti et al. [undated] on the tidytransit r package manual pages.

Table 1: SA1 zone 20403106915 geographic data

	20136
sa1_code_2016	20403106915
sa1_7dig_2016	2106915
sa2_code_2016	204031069
sa2_5dig_2016	21069
sa2_name_2016	Bright - Mount Beauty
sa3_code_2016	20403
sa3_name_2016	Wodonga - Alpine
sa4_code_2016	204
sa4_name_2016	Hume
gcc_code_2016	2RVIC
gcc_name_2016	Rest of Vic.
state_code_2016	2
state_name_2016	Victoria
areasqkm_2016	284.598
cent_long	147.05
cent_lat	-36.57882

departure_time	stop_id	route_short_name	trip_headsign
07:25:00	45125	NA	Albury
09:25:00	45125	NA	Albury
09:30:00	45125	NA	Albury
14:50:00	45125	NA	Mt Beauty
16:20:00	45125	NA	Mt Beauty
16:40:00	45125	NA	Mt Beauty

The GTFS feed shows six departures for the stop, with three services running towards Albury and three towards Mt Beauty. This suggests a total of 6 arrivals to each of the two bus stops in a single day week [23] Therefore the total $SI_{SA1,2016,20403106915,10/8/21}$ score is equal to $(2*(6*pi*400*400/284598000))$ which is equal to 0.0211943.

[23] The SL_{Bn} term.

The $SI_{SA1,2016,20403106915,10/8/21}$ score calculated by the developed code is 2.1194406×10^{12}. The hand-calculated $SI_{SA1,2016,20403106915,10/8/21}$ matches that produced by the developed code, suggesting that the developed code is providing the expected output.

Clayton SA2: SI scores for SA1 zones

This section briefly reviews the SI scores for SA1 zones within the Clayton SA2 area. SI scores for the day of the 2016 and 2021 censuses are compared, together with review of the ratio of the ISRAD to SI scores

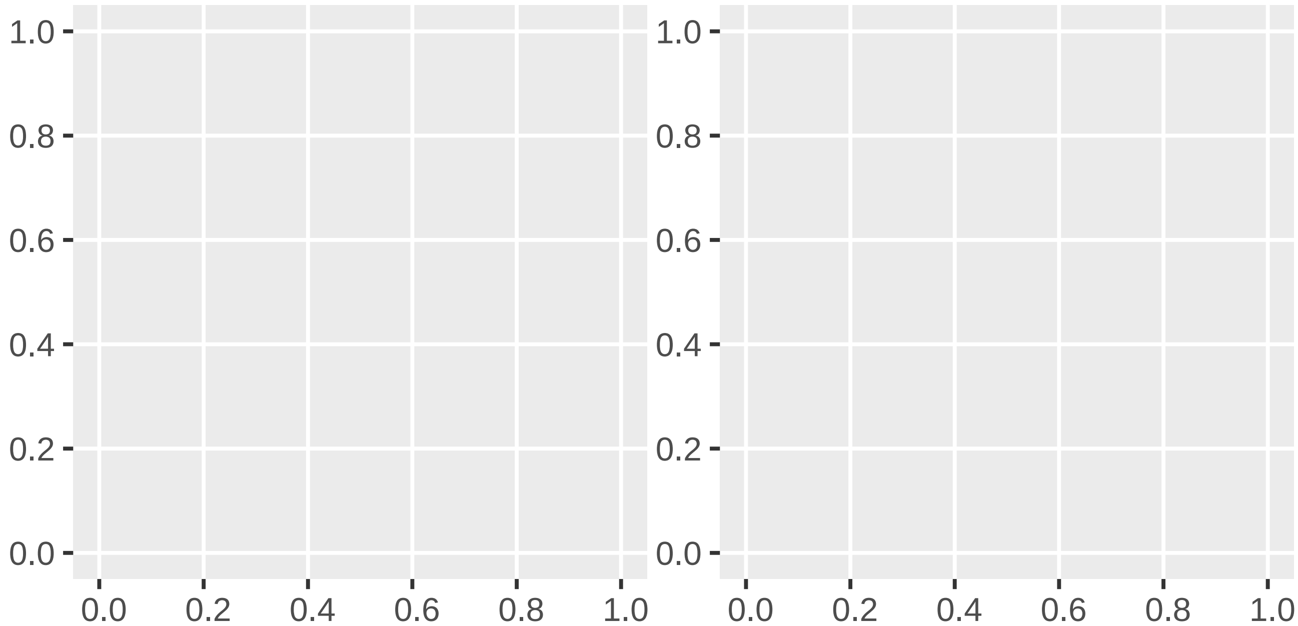

Figure 5: SI scores for SA1 zones within the Clayton SA2 area, 2016 census day (left) and 2021 census day (right)

Melbourne City SA3 zone

Examining SA1s across all of Greater Melbourne and all of Victoria

Bibliography

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`.

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.

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`.

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, generate a depature timetable*, undated. URL `https://r-transit.github.io/tidytransit/articles/timetable.html`.

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/`.

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

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.

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.