gtfssupplyindex

James Reynolds

2024-05-28

Introduction

Use the gtfssupplyindex package to calculate Transit Supply Index (SI) scores from General Transit Feed Specification (GTFS) datasets. The Transit Supply Index (SI) was developed in Currie, G and Senbergs, Z (2007). A generalised equation[1] for the SI is as follows:

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

In which:

- $SI_{area,time}$ 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.

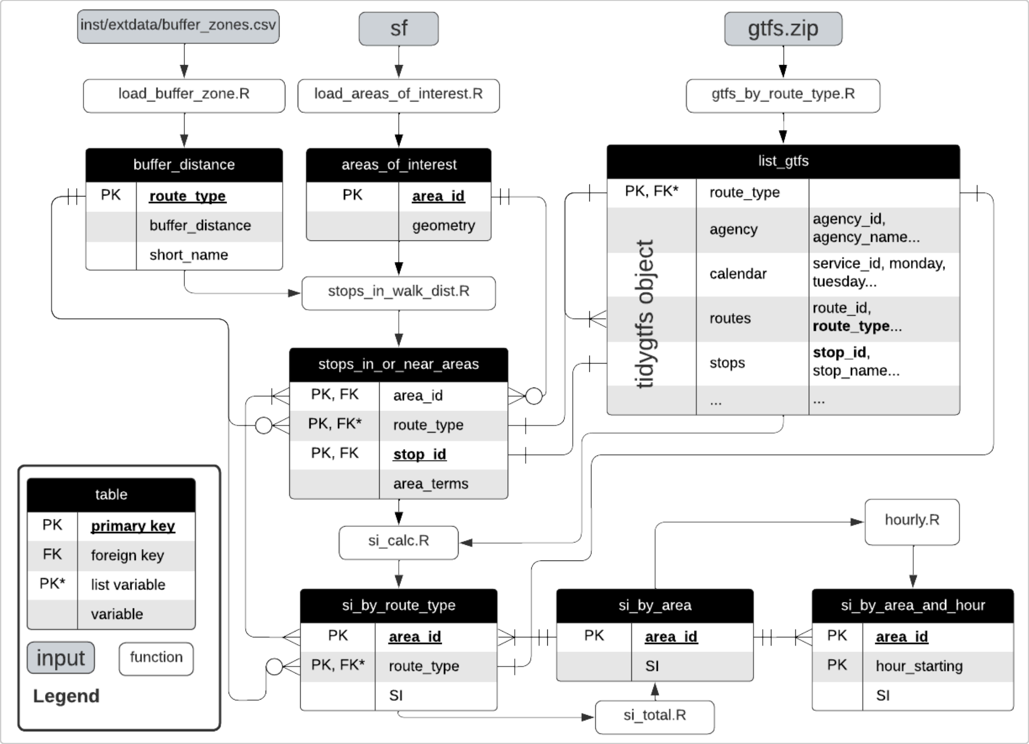
Installation & Dependencies

This package requires the sf, gtfstools and tidytransit packages (amongst others).

```
# Install devtools to allow package to be downloaded from github repository
install.packages("devtools")
devtools::install_github("James-Reynolds/gtfssupplyindex")
```

Package structure and functions

The structure and functions used to generate each table are shown in the below Entity Relationship Diagram (ERD).



Entity Relationship Diagram (ERD) showing the data structure and functions

The package takes input from three files:

- a gtfs feed (gtfs.zip);
- a sf object describing the geometry of the areas for which the SI is to be calculated; and
- a csv file defining the buffer zone distances (in metres) for each route_type - a version of this file is included in the package.

Outputs can be generated in various formats including:

- an si_by_route_type list of tables (bottom-left), showing SI scores by area for each route type (mode) on a specific day;
- an si_by_area table (bottom-centre), showing SI scores for each area for all route types (modes) on a specific day; and
- an si_by_area_and_hour table (bottom-right), showing a SI score for each area and each hour of a specific day.

Sample calculation

The package includes a GTFS dataset from the Mornington Penninsula Tourist Railway as sample data. Minor alternations have been made to this GTFS dataset[2].

1. The GTFS data is first loaded, with the gtfs_by_route_type function splitting this into a list (by route_type) of tidygtfs objects. This is achieved using the filter_by_route_type() function from the gtfstools package.

```
#load the revised mornington GTFS data
list_gtfs = gtfssupplyindex:::gtfs_by_route_type(system.file(
  "extdata/mornington180109",
  "gtfs.zip",
  package = "gtfssupplyindex",
  mustWork = TRUE))
```

2. Geographical data about the areas of interest are loaded by the load_areas_of_interest() function into an sf object. The resultant areas_of_interest table contains each area_id and its associated geometry.

```
areas_of_interest <- load_areas_of_interest(areas_of_interest = sf::st_read(system.file(
  "extdata",
  "mornington_sa12021.geojson",
  package = "gtfssupplyindex",
  mustWork = TRUE)),
  area_id_field = "sa1_code_2021")
#> Reading layer `mornington_sa12021' from data source
#> 
`/private/var/folders/tw/zt27pfld0gn750zkn_y4zr440000gn/T/RtmpPuqDml/temp_libpath17e25febe89b/gtfssupplyindex/extdata/morningto
#>   using driver `GeoJSON'
#> Simple feature collection with 392 features and 1 field
#> Geometry type: MULTIPOLYGON
#> Dimension:     XY
#> Bounding box:  xmin: 144.6514 ymin: -38.49738 xmax: 145.2615 ymax: -38.16239
#> Geodetic CRS:  WGS 84
head(areas_of_interest) %>% knitr::kable(caption = "First 6 entries in areas_of_interest
table")
```

First 6 entries in areas_of_interest table

area_id	geometry
21402137701	MULTIPOLYGON (((144.9617 -3…
21402137702	MULTIPOLYGON (((144.9916 -3…
21402137703	MULTIPOLYGON (((144.9923 -3…
21402137704	MULTIPOLYGON (((144.9878 -3…
21402137705	MULTIPOLYGON (((145.0015 -3…
21402137706	MULTIPOLYGON (((145.0088 -3…

3. Data about buffer zones, specifically the walking distance threshold assigned to each route_type (mode) is then loaded with the load_buffer_zone() function. The package includes this information in a csv file, in which it is assumed that the buffer zone is defined in metres.

```
buffer_distance <- gtfssupplyindex:::load_buffer_zones()
head(buffer_distance) %>% knitr::kable(caption = "First six entries in buffer_distance
definitions")
```

First six entries in buffer_distance definitions

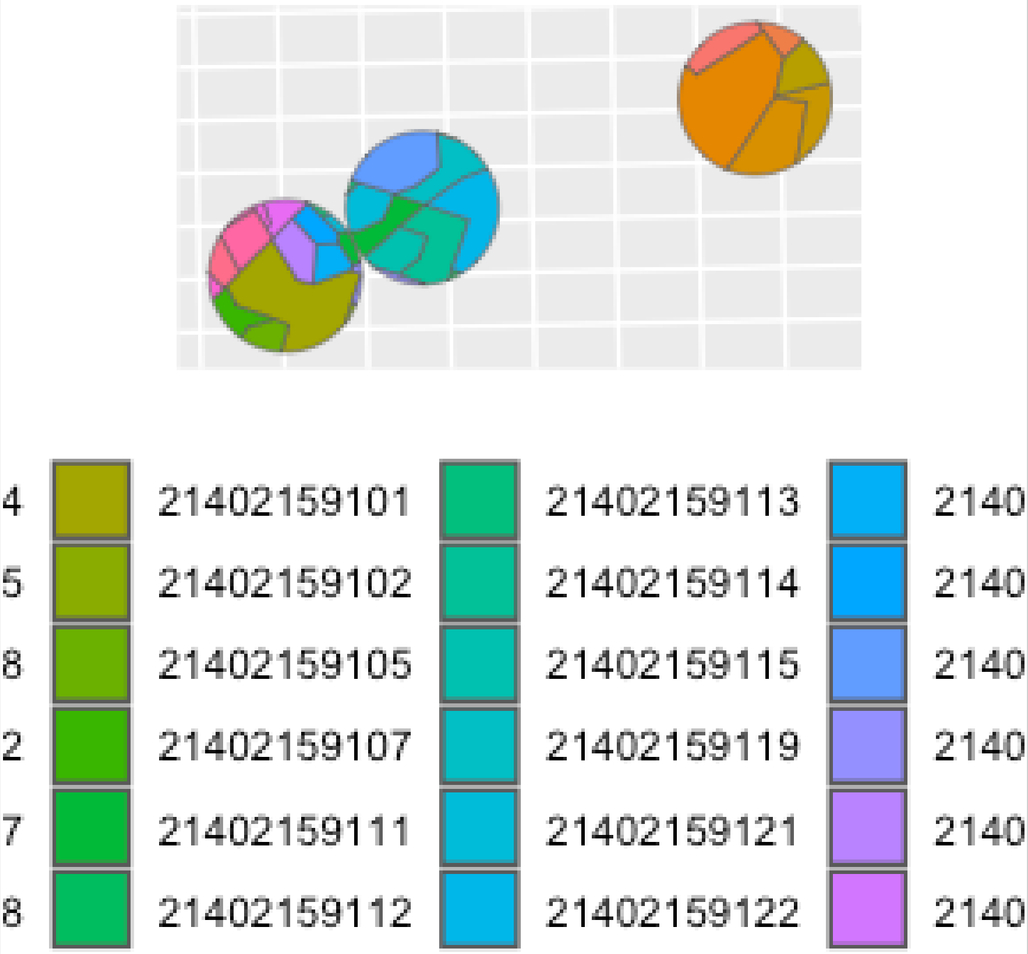
route_type	buffer_distance	short_name
0	400	lrt
1	800	subway
2	800	rail
3	400	bus

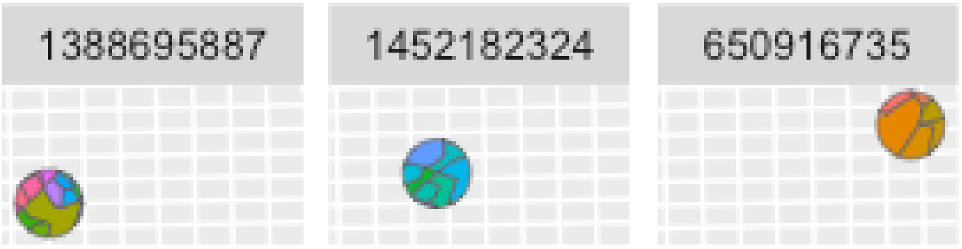
route_type	buffer_distance	short_name
4	800	ferry
5	400	cable_tram

4. The stops_in_walk_dist() function is then used to generate a list (by route_type) in which each element is a datatable describing which stops are within walking distance of which areas of interest. Each table entry includes a stop_id, an area_id that is within the buffer_distance threshold of that stop, and the corresponding area related terms in the SI calculation[3].

With verbose = TRUE the stops_in_walk_dist() function will plot the areas_of_interest in each stop's catchment (for the first 6 stops only), as a single plot and then facetted by stop_id, as shown below. Stop area_terms are also reported.

```r
stops_in_or_near_areas <- gtfssupplyindex:::stops_in_walk_dist(
  list_gtfs = list_gtfs,
  areas_of_interest = areas_of_interest,
  EPSG_for_transform = 28355,
  verbose = TRUE
)
```


4
5
8
2
7
8
21402159101
21402159102
21402159105
21402159107
21402159111
21402159112
21402159113
21402159114
21402159115
21402159119
21402159121
21402159122
2140
2140
2140
2140
2140
2140


1388695887
1452182324
650916735

```
head(stops_in_or_near_areas[[1]]) %>% knitr::kable(caption = "'Rail' element of the
stops_in_or_near_areas list for the Mornington Pennisula datasets, first six entries")
```

‘Rail’ element of the stops_in_or_near_areas list for the Mornington Pennisula datasets, first six entries

stop_id	area_id	area_terms
1388695887	21402159101	0.7999912
1388695887	21402159102	0.0168220
1388695887	21402159105	0.6779951
1388695887	21402159107	0.6453927
1388695887	21402159111	0.2011127
1388695887	21402159113	0.1081424

In the plot above the catchments of three stops on the Mornington Peninsula Tourist Railway are shown. The catchments cover parts of around 30 different SA1 zones. If instead, the Monrington Peninsula SA3 zone is selected as the area of interest, as below:

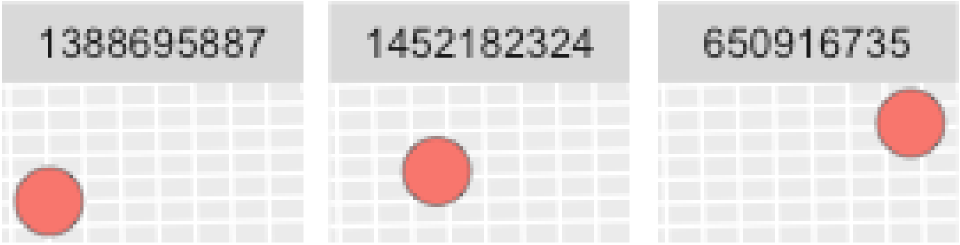
```
areas_of_interest <- load_areas_of_interest(areas_of_interest = sf::st_read(system.file(
  "extdata",
  "mornington_sa32021.geojson",
  package = "gtfssupplyindex",
  mustWork = TRUE)),
  area_id_field = "sa3_code_2021")
#> Reading layer `mornington_sa32021' from data source
#>
`/private/var/folders/tw/zt27pfld0gn750zkn_y4zr440000gn/T/RtmpPuqDml/temp_libpath17e25febe89b/gtfssupplyindex/extdata/morningto
#>   using driver `GeoJSON'
#> Simple feature collection with 1 feature and 1 field
#> Geometry type: MULTIPOLYGON
#> Dimension:     XY
#> Bounding box:  xmin: 144.6514 ymin: -38.49884 xmax: 145.2615 ymax: -38.16239
#> Geodetic CRS:  WGS 84
head(areas_of_interest) %>% knitr::kable(caption = "First 6 entries in areas_of_interest
table")
```

First 6 entries in areas_of_interest table

area_id	geometry
21402	MULTIPOLYGON (((145.2129 -3…

```
stops_in_or_near_areas <- gtfssupplyindex:::stops_in_walk_dist(
  list_gtfs = list_gtfs,
  areas_of_interest = areas_of_interest,
  EPSG_for_transform = 28355,
  verbose = TRUE
)
```

area_id
21402


1388695887
1452182324
650916735

```
head(stops_in_or_near_areas[[1]]) %>% knitr::kable(caption = "'Rail' element of the
stops_in_or_near_areas list for the Mornington Pennisula datasets, first six entries")
```

'Rail' element of the stops_in_or_near_areas list for the Mornington Pennisula datasets, first six entries

stop_id	area_id	area_terms
1388695887	21402	0.0027748
1452182324	21402	0.0027748
650916735	21402	0.0027748

The above shows how the catchments for all three stops are completely within the Mornington Peninsula SA3 zone, but make up only a small fraction of its area. The area_terms variable represents the ${} $ component of the SI calculation. In this case $Area_{Bn} = \pi 800^2$ and $Area_{area}$ = 723832572[4], giving $\frac{Area_{Bn}}{Area_{area}}$ = 0.00278

Continuing the SA3 calculation:

5. The si_calc function is then used to output the si_by_route_type table. In this instance we are looking at the services that ran on December 30th, 2018, which involved 4 trips to Mornington (trip_id 47-50) and 4 trips to Moorooduc (trip_id 54 to 57).

```
si_by_route_type <- gtfssupplyindex::si_calc(
    list_gtfs = list_gtfs,
    stops_in_or_near_areas = stops_in_or_near_areas,
    date_ymd = lubridate::ymd("2018-12-30"),
    start_hms = lubridate::hms("0:00:00"),
    end_hms = lubridate::hms("23:59:59"),
    verbose = TRUE)
#> [1] "Now calculating rail"
```

```
si_by_route_type %>% knitr::kable(caption = "SI values for Mornington Penninsula Railway
services on 30/12/2018 (full day)")
```

SI values for Mornington Penninsula Railway services on 30/12/2018 (full day)

area_id	SI
21402	0.0443966

This result is relatively easy to confirm by hand, as follows:

$$
\begin{aligned}
SI_{area,time} &= \sum \frac{Area_{Bn}}{Area_{area}} * SL_{n,time} \\
SI_{area=21402,date=30-12-18} &= \frac{Area_{stop1388695887}}{Area_{area=21402}} * SL_{stop1388695887,date=30-12-18} \\
&+ \frac{Area_{stop1452182324}}{Area_{area=21402}} * SL_{stop1452182324,date=30-12-18} \\
&+ \frac{Area_{stop650916735}}{Area_{area=21402}} * SL_{stop650916735,date=30-12-18} \\
SI_{area=21402,date=30-12-18} &= 0.0027748 * 4 \\
&+ 0.0027748 * 8 \\
&+ 0.0027748 * 4 \\
SI_{area=21402,date=30-12-18} &= 0.00444
\end{aligned}
$$

6. The si_calc() function used above returns a list, each element of which is a datatable containing the area_id and SI scores for an individual mode. The Mornington Peninsular GTFS dataset contains only one route type ("rail"), so the output of the si_calc() function in this case is a list with only one element. In the event, however, that there is more than one route type (mode) within the GTFS dataset the si_total() function can be used to aggregate the SI scores across modes into a single table

```
si_by_area <- si_by_route_type %>% si_total()
```

7. Finally, if SI values are desired on an hourly basis, the hourly() function can be used instead of the si_calc() function. hourly() simply calls the si_calc() and si_total functions across each hour of service (e.g. 10:00 to 11:00):

```
        si_by_area_and_hour <- gtfssupplyindex::hourly(list_gtfs, stops_in_or_near_areas, "2018-12-
30")

        si_by_area_and_hour %>% head() %>% knitr::kable(caption = "Mornington Penninsula Tourist
Railway hourly SI values for December 30, 2018, for SA3 zones")
```

Mornington Penninsula Tourist Railway hourly SI values for December 30, 2018, for SA3 zones

area_id	SI	hour_starting
21402	0.0055496	10:00
21402	0.0055496	11:00
21402	0.0110991	12:00
21402	0.0027748	13:00
21402	0.0083244	14:00
21402	0.0110991	15:00

---

1. The focus of Currie and Senbergs (2007) was on the context of Melbourne’s Census Collection Districts (CCD) and calculations based on a week of transit service. The equation shown has been adjusted so that it is generalised to all times and areas of interest.↩
2. The route_colour field has been updated as tidytransit throws an error if this is not present. As well, the route type has been changed to 2 (“rail”). The original feed uses a route_type of 107, which correlates with “tourist railway” in the extended GTFS route type definitions, which is not yet supported by all R packages related to GTFS analysis.↩
3. This is the $Area_{Bn}/Area_{Area}$ bit↩
4. Note, the ABS data gives a slightly different value for the area of the Mornington Peninsula SA3 zone than that output by the st_area function, but this is likely due to there being a slightly different methodology for calculating the area. The error is approximately 0.015%.↩