

Package ‘metrosp’

September 3, 2026

Type Package

Title São Paulo Metro Passenger Demand Data

Version 1.2.1

Description Provides passenger demand data for the São Paulo metro system, covering 2012 to 2026. Datasets include monthly passenger entries and transported counts by line, average weekday station entries, daily station entries, and spatial geometries for metro and commuter train lines and stations. The bundled datasets are a fixed snapshot, so analyses stay reproducible and examples run offline; more recent data is published separately as the upstream sources are updated.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

LazyData true

Depends R (>= 3.5.0)

Imports cli, jsonlite, tools, utils

Suggests bizdays, spelling, dataverse, digest, dplyr (>= 1.2.0),
ggplot2, knitr, piggyback, quarto, rmarkdown, sf, tarchetypes,
targets, testthat (>= 3.0.0), waldo, withr

Config/testthat/edition 3

RoxygenNote 7.3.3

URL <https://github.com/viniciusoike/metrosp>,
<https://viniciusoike.github.io/metrosp/>

BugReports <https://github.com/viniciusoike/metrosp/issues>

VignetteBuilder quarto

NeedsCompilation no

Author Vinicius Oike [aut, cre, cph] (ORCID:
<<https://orcid.org/0009-0005-8015-9189>>)

Maintainer Vinicius Oike <viniciusoike@gmail.com>

Repository CRAN

Date/Publication 2026-09-03 08:00:02 UTC

Contents

calendar_spo	2
lines	3
metro_colors	4
passengers_entrance	4
passengers_transported	6
stations	8
station_averages	9
station_daily	10
station_inauguration	12

Index	14
--------------	-----------

calendar_spo	<i>São Paulo Holiday and Business-Day Calendar</i>
--------------	--

Description

A daily calendar for São Paulo (city) covering 2012–2030, classifying each date as a holiday or business day. Includes national, state, and municipal holidays in São Paulo, with flags for optional work days (`is_ponto_facultativo`) and extended holiday weekends (`is_feriadao`).

Usage

calendar_spo

Format

A data frame with one row per day and the following columns:

date Calendar date (Date).

year Calendar year (integer).

weekday Day of week from `lubridate::wday()`: 1 = Sunday, 2 = Monday, ..., 7 = Saturday (integer).

is_weekend TRUE for Saturdays and Sundays (logical).

is_holiday TRUE when the date is a gazetted holiday at any scope (logical).

is_business_day TRUE when the date is neither a weekend nor a holiday (logical).

holiday_name Name of the holiday in Portuguese (character). NA on non-holiday dates.

holiday_scope Scope of the holiday (character). One of "national", "state", or "municipal"; NA on non-holiday dates.

is_ponto_facultativo TRUE for holidays that are technically optional at the federal level (Carnaval, Corpus Christi) but observed as holidays in São Paulo (logical).

is_feriadao TRUE when a holiday falls on Monday, Tuesday, Thursday, or Friday, creating a potential extended weekend with the adjacent Saturday/Sunday (logical).

Details

The calendar covers the full date range of the [station_daily](#) dataset (Lines 4/5 from January 2012) and extends through 2030 for forecasting use.

See Also

[station_daily](#) for daily passenger data that can be joined on date.

lines	<i>Metro and Train Line Routes</i>
-------	------------------------------------

Description

Spatial line geometries for São Paulo metro (METRO SP) and commuter train (CPTM) lines, including both currently operating lines and planned future expansions.

Usage

```
lines
```

Format

An sf data frame with LINESTRING geometry (CRS: WGS84 / EPSG:4326) and the following columns:

line_number Official line number (integer).

line_name_pt Portuguese color name of the line (character).

line_name English color name of the line (character).

company_name Operating company name (character).

type Either "metro" (METRO SP) or "train" (CPTM) (character).

status Either "current" (operating) or "future" (planned expansion) (character).

geometry Line route geometry (sfc_LINESTRING).

Details

Requires the **sf** package to work with spatial features. The distinction between types follows GeoSampa's classification. Broadly, "metro" lines run underground as a subway and "train" lines run above ground as commuter rail, though exceptions exist.

Source

GeoSampa, Prefeitura de São Paulo. <https://geosampa.prefeitura.sp.gov.br/>

See Also

[stations](#) for station point locations.

metro_colors	<i>Metro SP Official Line Colors</i>
--------------	--------------------------------------

Description

A named character vector of official hex color codes for the six metro lines operated by METRO SP (Lines 1–3 and 15) and ViaMobilidade (Lines 4 and 5).

Usage

metro_colors

Format

A named character vector of length 6. Names are English color names; values are hex color codes:

- Blue** Line 1 — "#171796"
- Green** Line 2 — "#007A5E"
- Red** Line 3 — "#ED2E38"
- Yellow** Line 4 — "#FFD525"
- Lilac** Line 5 — "#874ABF"
- Silver** Line 15 — "#8F8F8C"

Details

Colors follow official METRO SP and ViaMobilidade branding. Only the six currently operating metro lines are included; CPTM train lines and planned future lines (e.g., Line 6 Orange, Line 17 Gold) are not covered.

See Also

[lines](#) for the full line reference (numbers, names, and route geometries).

passengers_entrance	<i>Passengers Entering Metro SP Stations by Line</i>
---------------------	--

Description

Monthly count of passengers entering São Paulo metro stations, aggregated by metro line. Data covers January 2016 through 2026 for Lines 1, 2, 3, and 15; Line 4 from January 2012; Line 5 from January 2016. July 2017 is the one missing month, absent for every line the METRO portal covers. Sourced from the METRO SP transparency portal and the Insper Dataverse.

Usage

passengers_entrance

Format

A data frame with the following columns:

date First day of the month (Date).

line_number Metro line number: 1, 2, 3, 4, 5, 15, or 99 for the network total (integer).

metric_abb Abbreviated metric code (character). One of: "total", "mdu", "msa", "mdo", "max".

value Passenger count, in individual passengers (numeric).

metric Measurement type in English (character). One of: "Total", "Average on Business Days", "Average on Saturdays", "Average on Sundays", "Daily Peak".

metric_pt Measurement type in Portuguese (character). One of: "Total", "Média dos Dias Úteis", "Média dos Sábados", "Média dos Domingos", "Máxima Diária".

line_name English name of the metro line (character).

line_name_pt Portuguese name of the metro line (character).

year Calendar year (integer).

Details

Data by source and line:

- Lines 1, 2, 3, and 15: METRO SP transparency portal, January 2016–2026, except July 2017.
- Line 4 (Amarela/ViaQuatro): Insper Dataverse, January 2012–2026.
- Line 5 (Lilás/ViaMobilidade): METRO SP transparency portal, January 2016–July 2018, except July 2017; Insper Dataverse, August 2018–2026.
- Network total (line_number = 99): METRO SP transparency portal only; may not be available for all years.

METRO published January–September 2017 only as PDFs, with no machine-readable equivalent. Those months were transcribed from the reports and reconciled against the printed line and network totals. July 2017 has no entrance table at all, because the file METRO published under that name repeats the transported figures. Lines 1, 2, 3, 5, and 15 and the network total therefore carry no value for that month; Line 4 comes from the Dataverse and is unaffected.

Metrics:

- total: Total passengers in the month.
- mdu: Average daily entries on business days (Média dos Dias Úteis).
- msa: Average daily entries on Saturdays (Média dos Sábados).
- mdo: Average daily entries on Sundays (Média dos Domingos).
- max: Daily maximum (Máxima Diária).

Months beyond the last published data point for each line are trimmed during assembly; interior NAs (e.g. operational outages) are preserved.

Data vintage

This dataset is a fixed snapshot, current through July 2026. It ships with the package so examples, vignettes, and offline analysis always have data to hand. The snapshot moves only when the column schema changes or a release deliberately carries new data, not when new months are published upstream.

METRO SP publishes on an irregular schedule and revises already-published years, so the numbers here will drift from the source over time. Freshly rebuilt data is published on every pipeline run at <https://github.com/viniciusoike/metrosp/releases>.

Source

Companhia do Metropolitano de São Paulo (METRO SP). <https://transparencia.metrosp.com.br/dataset/demanda>

See Also

[passengers_transported](#) for transported counts, [station_averages](#) for station-level weekday averages.

passengers_transported
<i>Passengers Transported by Metro SP Line</i>

Description

Monthly count of passengers transported by São Paulo metro, aggregated by metro line and reported in **thousands of passengers**. Data covers January 2016 through 2026 for Lines 1, 2, 3, and 15, and January 2016 through August 2018 for Line 5. Sourced from the METRO SP transparency portal.

Usage

passengers_transported

Format

- A data frame with the following columns:
- date** First day of the month (Date).
 - line_number** Metro line number: 1, 2, 3, 5, 15, or 99 for the network total (integer).
 - metric_abb** Abbreviated metric code (character). One of: "total", "mdu", "msa", "mdo", "max".
 - value** Passenger count, in thousands of passengers (numeric).
 - metric** Measurement type in English (character). One of: "Total", "Average on Business Days", "Average on Saturdays", "Average on Sundays", "Daily Peak".
 - metric_pt** Measurement type in Portuguese (character). One of: "Total", "Média dos Dias Úteis", "Média dos Sábados", "Média dos Domingos", "Máxima Diária".

line_name English name of the metro line (character).

line_name_pt Portuguese name of the metro line (character).

year Calendar year (integer).

Details

Values are in thousands of passengers, as published by METRO SP. The other demand datasets count individual passengers, so multiply by 1000 before comparing value with [passengers_entrance](#).

A transported passenger is one who crossed a turnstile plus one who transferred between lines at an interchange station, so transported counts run above entry counts for the same line and month.

All data comes from the METRO SP transparency portal. Line 4 (Amarela) is not available in this dataset — the Insper Dataverse source does not include transported counts for Lines 4 or 5. Line 5 (Lilás) is available from the METRO portal only for January 2016–August 2018: the line was handed over to ViaMobilidade in August 2018 and the portal stopped reporting its transported counts afterwards. The network total (`line_number = 99`) may not be available for all years.

METRO published January–September 2017 only as PDFs, with no machine-readable equivalent. Those months were transcribed from the reports and reconciled against the printed line and network totals. June 2017 is the one month whose network total is missing, because the report reprinted May’s network column; the per-line values for June are sound.

Metrics:

- **total**: Total passengers in the month.
- **mdu**: Average daily entries on business days (Média dos Dias Úteis).
- **msa**: Average daily entries on Saturdays (Média dos Sábados).
- **mdo**: Average daily entries on Sundays (Média dos Domingos).
- **max**: Daily maximum (Máxima Diária).

Months beyond the last published data point for each line are trimmed during assembly; interior NAs (e.g. operational outages) are preserved.

Data vintage

This dataset is a fixed snapshot, current through July 2026. It ships with the package so examples, vignettes, and offline analysis always have data to hand. The snapshot moves only when the column schema changes or a release deliberately carries new data, not when new months are published upstream.

METRO SP publishes on an irregular schedule and revises already-published years, so the numbers here will drift from the source over time. Freshly rebuilt data is published on every pipeline run at <https://github.com/viniciusoike/metrosp/releases>.

Source

Companhia do Metropolitano de São Paulo (METRO SP). <https://transparencia.metrosp.com.br/dataset/demanda>

See Also

[passengers_entrance](#) for entry counts, [station_averages](#) for station-level weekday averages.

stations

*Metro and Train Station Locations***Description**

Spatial point locations for São Paulo metro (METRO SP) and commuter train (CPTM) stations, including both currently operating stations and planned future stations.

Usage

stations

Format

An sf data frame with POINT geometry (CRS: WGS84 / EPSG:4326) and the following columns:

station_name Station name in title case (character).

line_number Line number the station belongs to (integer).

line_name_pt Portuguese color name of the line (character).

line_name English color name of the line (character).

company_name Operating company name (character).

type Either "metro" (METRO SP) or "train" (CPTM) (character).

status Either "current" (operating) or "future" (planned expansion) (character).

geometry Station location (sfc_POINT).

Details

Requires the **sf** package to work with spatial features. The distinction between types follows GeoSampa's classification. Broadly, "metro" lines run underground as a subway and "train" lines run above ground as commuter rail, though exceptions exist.

Source

GeoSampa, Prefeitura de São Paulo. <https://geosampa.prefeitura.sp.gov.br/>

See Also

[lines](#) for line route geometries, [station_averages](#) for passenger data by station.

station_averages	<i>Average Weekday Passenger Entries by Station</i>
------------------	---

Description

Monthly average of weekday (business day) passenger entries for each station in the São Paulo metro system. Data covers January 2016 through 2026 for Lines 1, 2, 3, and 15; Line 4 from January 2012; Line 5 from January 2016. Sourced from the METRO SP transparency portal and the Insper Dataverse.

Usage

```
station_averages
```

Format

A data frame with the following columns:

date First day of the month (Date).
line_number Metro line number (integer).
station_name Name of the metro station (character).
avg_passenger Average weekday passenger entries (numeric).
line_name English name of the metro line (character).
line_name_pt Portuguese name of the metro line (character).
year Calendar year (integer).

Details

Only the weekday average (mdu) metric is available at the station level. For line-level data with all five metrics, see [passengers_entrance](#). Months beyond the last published data point for each line are trimmed during assembly; interior NAs (e.g. operational outages) are preserved.

Station coverage by line and source:

- Line 1 (Azul/Blue): 23 stations, January 2016–2026 (METRO SP portal).
- Line 2 (Verde/Green): 14 stations, January 2016–2026 (METRO SP portal).
- Line 3 (Vermelha/Red): 18 stations, January 2016–2026 (METRO SP portal).
- Line 4 (Amarela/Yellow): January 2012–2026 (Insper Dataverse).
- Line 5 (Lilás/Lilac): January 2016–July 2018 (METRO SP portal) and August 2018–2026 (Insper Dataverse).
- Line 15 (Prata/Silver): 2 stations in 2016–2017 (assisted operation: Vila Prudente and Oratório), 10 stations in 2020, 11 from January 2021 onward (Jardim Colonial added), January 2016–2026 (METRO SP portal).

METRO published January–September 2017 only as PDFs, with no machine-readable equivalent. Those months were transcribed from the reports and reconciled against the published line totals.

February–June 2016 carries a defect in the Line 1 values. Across those five months the station figures fall well short of what the surrounding months and the line total in `passengers_entrance` imply, and they are misallocated across stations, with Santa Cruz and Sé too high and São Bento and Portuguesa-Tietê too low. The defect comes from METRO’s retroactive publication of 2016 and is not corrected here, so exclude those five months from station-level baselines.

Data vintage

This dataset is a fixed snapshot, current through July 2026. It ships with the package so examples, vignettes, and offline analysis always have data to hand. The snapshot moves only when the column schema changes or a release deliberately carries new data, not when new months are published upstream.

METRO SP publishes on an irregular schedule and revises already-published years, so the numbers here will drift from the source over time. Freshly rebuilt data is published on every pipeline run at <https://github.com/viniciusoike/metrosp/releases>.

Source

Companhia do Metropolitano de São Paulo (METRO SP). <https://transparencia.metrosp.com.br/dataset/demanda>

See Also

`station_daily` for daily station entries, `passengers_entrance` for monthly line-level totals.

station_daily	<i>Daily Passenger Entries by Metro SP Station</i>
---------------	--

Description

Daily passenger entries at each station in the São Paulo metro system. Data covers January 2012 through 2026 for Lines 4 and 5 (Insper Dataverse), and 2020 through 2026 for Lines 1, 2, 3, and 15 (METRO SP transparency portal).

Usage

`station_daily`

Format

- A data frame with the following columns:
- date** Date of observation (Date).
 - line_number** Metro line number: 1, 2, 3, 4, 5, or 15 (integer).
 - station_name** Full station name (character).

passengers Daily passenger entries (numeric).
line_name English name of the metro line (character).
line_name_pt Portuguese name of the metro line (character).
station_code Three-letter station abbreviation used internally by METRO SP (character). NA for Lines 4 and 5 (Dataverse source).
year Calendar year (integer).

Details

Station coverage and date range by line:

- Line 1 (Azul/Blue): 23 stations, 2020–2026 (METRO SP portal).
- Line 2 (Verde/Green): 14 stations, 2020–2026 (METRO SP portal).
- Line 3 (Vermelha/Red): 18 stations, 2020–2026 (METRO SP portal).
- Line 4 (Amarela/Yellow): January 2012–2026 (Insper Dataverse); station_code is NA.
- Line 5 (Lilás/Lilac): August 2018–2026 (Insper Dataverse); station_code is NA.
- Line 15 (Prata/Silver): 10 stations in 2020, 11 from 2021 onward (Jardim Colonial added), 2020–2026 (METRO SP portal).

Some stations appear on multiple lines (e.g., Ana Rosa on Lines 1 and 2, Paraíso on Lines 1 and 2, Sé on Lines 1 and 3). These are recorded separately for each line.

Days beyond the last published data point for each line are trimmed during assembly; interior NAs (e.g. operational outages) are preserved.

Data vintage

This dataset is a fixed snapshot, current through July 2026. It ships with the package so examples, vignettes, and offline analysis always have data to hand. The snapshot moves only when the column schema changes or a release deliberately carries new data, not when new months are published upstream.

METRO SP publishes on an irregular schedule and revises already-published years, so the numbers here will drift from the source over time. Freshly rebuilt data is published on every pipeline run at <https://github.com/viniciusoike/metrosp/releases>.

Source

Companhia do Metropolitano de São Paulo (METRO SP). <https://transparencia.metrosp.com.br/dataset/demanda>

See Also

[station_averages](#) for monthly weekday averages, [passengers_entrance](#) for monthly line-level totals.

station_inauguration	<i>Station Commercial Opening Dates</i>
----------------------	---

Description

Inauguration (commercial opening) dates for São Paulo metro stations, covering stations whose opening falls within or near the [station_daily](#) / [station_averages](#) window. Used to flag ramp-up periods in which monthly ridership is still climbing toward steady-state and should generally be excluded from year-on-year or baseline comparisons.

Usage

```
station_inauguration
```

Format

A data frame with one row per (line, station):

line_number Metro line number (integer).

station_name Full station name (character).

inauguration_date Date of commercial opening (Date). NA for stations whose opening predates the dataset window (i.e., they were already operating when the data record begins).

phase Short label identifying the expansion phase, e.g. "L15 Fase 4" (character).

verified Whether the inauguration date has been cross-checked against the operator's announcement or an equivalently reliable source (logical). Stations with `verified = FALSE` carry best-effort dates and should not be relied on for legal or publication purposes without re-checking.

notes Free-text annotations about the source or any caveats (character, possibly NA).

pre_data_window TRUE when `inauguration_date` is NA because the station opened before the data starts (logical).

ramp_up_end `inauguration_date + 180` days — a heuristic end of the initial ramp-up period (Date). NA when `pre_data_window` is TRUE.

Details

The table is compiled by hand from `data-raw/station_inauguration.csv` in the package repository. Contributions that extend the table or verify uncertain dates are welcome.

When computing pre/post comparisons (e.g. 12m-vs-prior-12m or recovery-vs-2019), exclude stations where either window overlaps `ramp_up_end` to avoid mistaking ramp-up growth for organic demand change.

Source

Compiled from operator announcements (Companhia do Metropolitano de São Paulo, ViaQuatro, ViaMobilidade).

See Also

[stations](#) for spatial point locations, [station_averages](#) for monthly weekday averages.

Index

* datasets

- calendar_spo, [2](#)
- lines, [3](#)
- metro_colors, [4](#)
- passengers_entrance, [4](#)
- passengers_transported, [6](#)
- station_averages, [9](#)
- station_daily, [10](#)
- station_inauguration, [12](#)
- stations, [8](#)

calendar_spo, [2](#)

lines, [3](#), [4](#), [8](#)

metro_colors, [4](#)

passengers_entrance, [4](#), [7](#), [9–11](#)

passengers_transported, [6](#), [6](#)

station_averages, [6–8](#), [9](#), [11–13](#)

station_daily, [3](#), [10](#), [10](#), [12](#)

station_inauguration, [12](#)

stations, [3](#), [8](#), [13](#)