

Package ‘mtaOpenData’

May 9, 2026

Type Package

Title Convenient Access to MTA Open Data API Endpoints

Version 0.1.0

Description Provides helper functions to access datasets from the Metropolitan Transportation Authority (MTA) portion of the New York State Open Data platform <<https://data.ny.gov/>>. Returns results as tidy tibbles with support for optional filtering, sorting, and row limits through the Socrata API.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.3

Imports dplyr, tibble, jsonlite, httr, janitor, rlang

Suggests curl, covr, knitr, testthat (>= 3.0.0), vcr, withr, webmockr, ggplot2

URL <https://martinezc1.github.io/mtaOpenData/>,
<https://github.com/martinezc1/mtaOpenData>

BugReports <https://github.com/martinezc1/mtaOpenData/issues>

VignetteBuilder knitr

Config/testthat/edition 3

Depends R (>= 4.1.0)

NeedsCompilation no

Author Christian Martinez [aut, cre] (GitHub: martinezc1, ORCID:
<<https://orcid.org/0009-0005-6026-6454>>)

Maintainer Christian Martinez <c.martinez0@outlook.com>

Repository CRAN

Date/Publication 2026-04-01 08:30:02 UTC

Contents

| | |
|-------------------|---|
| mta_any_dataset | 2 |
| mta_list_datasets | 3 |
| mta_pull_dataset | 3 |

| | |
|--------------|----------|
| Index | 5 |
|--------------|----------|

| | |
|-----------------|---------------------------------------|
| mta_any_dataset | <i>Load Any MTA Open Data Dataset</i> |
|-----------------|---------------------------------------|

Description

Downloads any MTA Open Data dataset given its Socrata JSON endpoint.

Usage

```
mta_any_dataset(
  json_link,
  limit = 10000,
  timeout_sec = 30,
  clean_names = TRUE,
  coerce_types = TRUE
)
```

Arguments

| | |
|--------------|--------------------------------------------------------------------------------------------|
| json_link | A Socrata dataset JSON endpoint URL (e.g., "https://data.ny.gov/resource/2ucp-7wg5.json"). |
| limit | Number of rows to retrieve (default = 10,000). |
| timeout_sec | Request timeout in seconds (default = 30). |
| clean_names | Logical; if TRUE, convert column names to snake_case (default = TRUE). |
| coerce_types | Logical; if TRUE, attempt light type coercion (default = TRUE). |

Value

A tibble containing the requested dataset.

Examples

```
# Examples that hit the live MTA Open Data API are guarded so CRAN checks
# do not fail when the network is unavailable or slow.
if (interactive() && curl::has_internet()) {
  endpoint <- "https://data.ny.gov/resource/2ucp-7wg5.json"
  out <- try(mta_any_dataset(endpoint, limit = 3), silent = TRUE)
  if (!inherits(out, "try-error")) {
    head(out)
  }
}
```

| | |
|-------------------|-----------------------------------------------|
| mta_list_datasets | <i>List datasets available in mtaOpenData</i> |
|-------------------|-----------------------------------------------|

Description

Retrieves the current MTA Open Data catalog and returns datasets available for use with ‘mta_pull_dataset()’.

Usage

```
mta_list_datasets()
```

Details

Keys are generated from dataset names using ‘janitor::make_clean_names()’.

Value

A tibble of available datasets, including generated ‘key’, dataset ‘uid’, and dataset ‘dataset_title’.

Examples

```
if (interactive() && curl::has_internet()) {  
  mta_list_datasets()  
}
```

| | |
|------------------|--------------------------------------------------------------------|
| mta_pull_dataset | <i>Pull a MTA Open Data dataset from the MTA Open Data catalog</i> |
|------------------|--------------------------------------------------------------------|

Description

Uses a dataset ‘key’ or ‘open_dataset_id’ from ‘mta_list_datasets()’ to pull data from MTA Open Data.

Usage

```
mta_pull_dataset(  
  dataset,  
  limit = 10000,  
  filters = list(),  
  date = NULL,  
  from = NULL,  
  to = NULL,  
  date_field = NULL,  
  where = NULL,  
  order = NULL,  
  timeout_sec = 30,
```

```

    clean_names = TRUE,
    coerce_types = TRUE
  )

```

Arguments

| | |
|--------------|----------------------------------------------------------------------------------------------------------------------------|
| dataset | A dataset key or open_dataset_id from 'mta_list_datasets()'. |
| limit | Number of rows to retrieve (default = 10,000). |
| filters | Optional named list of filters. Supports vectors (translated to IN()). |
| date | Optional single date (matches all times that day) using 'date_field'. |
| from | Optional start date (inclusive) using 'date_field'. |
| to | Optional end date (exclusive) using 'date_field'. |
| date_field | Optional date/datetime column to use with 'date', 'from', or 'to'. Must be supplied when 'date', 'from', or 'to' are used. |
| where | Optional raw SoQL WHERE clause. If 'date', 'from', or 'to' are provided, their conditions are AND-ed with this. |
| order | Optional SoQL ORDER BY clause. |
| timeout_sec | Request timeout in seconds (default = 30). |
| clean_names | Logical; if TRUE, convert column names to snake_case (default = TRUE). |
| coerce_types | Logical; if TRUE, attempt light type coercion (default = TRUE). |

Details

Dataset keys are generated from dataset_title using 'janitor::make_clean_names()'. Because keys are derived from live catalog metadata, dataset open_dataset_ids are the more stable option.

Value

A tibble.

Examples

```

if (interactive() && curl::has_internet()) {
  # Pull by key
  mta_pull_dataset("mta_bus_stops", limit = 3)

  # Pull by open_dataset_id
  mta_pull_dataset("2ucp-7wg5", limit = 3)

  # Filters
  mta_pull_dataset("2ucp-7wg5", limit = 3, filters = list(route_id = "QM3"))
}

```

Index

`mta_any_dataset`, 2
`mta_list_datasets`, 3
`mta_pull_dataset`, 3