

# Package ‘nysOpenData’

April 1, 2026

**Type** Package

**Title** Convenient Access to NYS Open Data API Endpoints

**Version** 0.1.1

**Description** Provides helper functions to access datasets from the NYS Open Data platform <<https://data.ny.gov/>>. Functions return results as tidy tibbles and support optional filtering, sorting, and row limits via the Socrata API.

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.3.3

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

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

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

**BugReports** <https://github.com/martinezc1/nysOpenData/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 08:00:14 UTC

## Contents

nys_any_dataset . . . . .	2
nys_list_datasets . . . . .	3
nys_pull_dataset . . . . .	3

---

nys_any_dataset	<i>Load Any NYS Open Data Dataset</i>
-----------------	---------------------------------------

---

## Description

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

## Usage

```
nys_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/28gk-bu58.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 nys 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/28gk-bu58.json"  
  out <- try(nys_any_dataset(endpoint, limit = 3), silent = TRUE)  
  if (!inherits(out, "try-error")) {  
    head(out)  
  }  
}
```

---

nys\_list\_datasets      *List datasets available in nysOpenData*

---

**Description**

Retrieves the current Open NY catalog and returns datasets available for use with 'nys\_pull\_dataset()'.

**Usage**

```
nys_list_datasets()
```

**Details**

Keys are generated from dataset titles using 'janitor::make\_clean\_names()'.

**Value**

A tibble of available datasets, including generated 'key', dataset 'uid', and dataset 'title'.

**Examples**

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

---

nys\_pull\_dataset      *Pull a NYS Open Data dataset from the NYS Open Data catalog*

---

**Description**

Uses a dataset 'key' or 'uid' from 'nys\_list\_datasets()' to pull data from NYS Open Data.

**Usage**

```
nys_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 UID from <code>'nys_list_datasets()'</code> .
limit	Number of rows to retrieve (default = 10,000).
filters	Optional named list of filters. Supports vectors (translated to <code>IN()</code> ).
date	Optional single date (matches all times that day) using <code>'date_field'</code> .
from	Optional start date (inclusive) using <code>'date_field'</code> .
to	Optional end date (exclusive) using <code>'date_field'</code> .
date_field	Optional date/datetime column to use with <code>'date'</code> , <code>'from'</code> , or <code>'to'</code> . Must be supplied when <code>'date'</code> , <code>'from'</code> , or <code>'to'</code> are used.
where	Optional raw SoQL WHERE clause. If <code>'date'</code> , <code>'from'</code> , or <code>'to'</code> 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 titles using `'janitor::make_clean_names()'`. Because keys are derived from live catalog metadata, dataset UIDs are the more stable option.

**Value**

A tibble.

**Examples**

```
if (interactive() && curl::has_internet()) {
  # Pull by key
  nys_pull_dataset("311_service_requests", limit = 3)

  # Pull by UID
  nys_pull_dataset("28gk-bu58", limit = 3)

  # Filters
  nys_pull_dataset("28gk-bu58", limit = 3, filters = list(award_name = "MBA"))
}
```

# Index

nys\_any\_dataset, 2  
nys\_list\_datasets, 3  
nys\_pull\_dataset, 3