

# Package ‘ucimlrepo’

July 22, 2025

**Title** Explore UCI ML Repository Datasets

**Version** 0.0.2

**Description** Find and import datasets from the University of California Irvine Machine Learning (UCI ML) Repository into R. Supports working with data from UCI ML repository inside of R scripts, notebooks, and 'Quarto'/RMarkdown' documents. Access the UCI ML repository directly at <https://archive.ics.uci.edu/>.

**URL** <https://r-pkg.thecoatlessprofessor.com/ucimlrepo/>,  
<https://github.com/coatless-rpkg/ucimlrepo>,  
<https://archive.ics.uci.edu/>

**BugReports** <https://github.com/coatless-rpkg/ucimlrepo/issues>

**License** MIT + file LICENSE

**Depends** R (>= 4.1)

**Imports** htr2, utils

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**Collate** 'constants.R' 'fetch-ucirepo.R' 'list-available-datasets.R'  
'ucimlrepo-package.R'

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**NeedsCompilation** no

**Author** James Joseph Balamuta [aut, cre, cph] (ORCID:  
<https://orcid.org/0000-0003-2826-8458>),  
Philip Truong [aut, cph]

**Maintainer** James Joseph Balamuta <james.balamuta@gmail.com>

**Repository** CRAN

**Date/Publication** 2024-08-31 07:20:02 UTC

## Contents

|                                   |          |
|-----------------------------------|----------|
| fetch_ucirepo . . . . .           | 2        |
| list_available_datasets . . . . . | 4        |
| <b>Index</b>                      | <b>5</b> |

---

|               |  |
|---------------|--|
| fetch_ucirepo | <i>Fetch UCI ML Repository Dataset</i> |
|---------------|--|

---

### Description

Loads a dataset from the UCI ML Repository, including the dataframes and metadata information.

### Usage

```
fetch_ucirepo(name, id)
```

### Arguments

|      |  |
|------|--|
| name | Character. Dataset name, or substring of name. |
| id   | Integer. Dataset ID for UCI ML Repository.     |

### Details

Only provide name or id, not both.

### Value

A list containing dataset metadata, dataframes, and variable info in its properties.

- **data**: Contains dataset matrices as pandas dataframes
  - **ids**: Dataframe of ID columns
  - **features**: Dataframe of feature columns
  - **targets**: Dataframe of target columns
  - **original**: Dataframe consisting of all IDs, features, and targets
  - **headers**: List of all variable names/headers
- **metadata**: Contains metadata information about the dataset.
  - **uci\_id**: Unique dataset identifier for UCI repository
  - **name**: Name of dataset on UCI repository
  - **repository\_url**: Link to dataset webpage on the UCI repository
  - **data\_url**: Link to raw data file
  - **abstract**: Short description of dataset
  - **area**: Subject area e.g. life science, business
  - **tasks**: Associated machine learning tasks e.g. classification, regression
  - **characteristics**: Dataset types e.g. multivariate, sequential

- **num\_instances**: Number of rows or samples
- **num\_features**: Number of feature columns
- **feature\_types**: Data types of features
- **target\_col**: Name of target column(s)
- **index\_col**: Name of index column(s)
- **has\_missing\_values**: Whether the dataset contains missing values
- **missing\_values\_symbol**: Indicates what symbol represents the missing entries (if the dataset has missing values)
- **year\_of\_dataset\_creation**: Year that data set was created
- **dataset\_doi**: DOI registered for dataset that links to UCI repo dataset page
- **creators**: List of dataset creator names
- **intro\_paper**: Information about dataset's published introductory paper
- **external\_url**: URL to external dataset page. This field will only exist for linked datasets i.e. not hosted by UCI
- **additional\_info**: Descriptive free text about dataset
  - \* **summary**: General summary
  - \* **purpose**: For what purpose was the dataset created?
  - \* **funded\_by**: Who funded the creation of the dataset?
  - \* **instances\_represent**: What do the instances in this dataset represent?
  - \* **recommended\_data\_splits**: Are there recommended data splits?
  - \* **sensitive\_data**: Does the dataset contain data that might be considered sensitive in any way?
  - \* **preprocessing\_description**: Was there any data preprocessing performed?
  - \* **variable\_info**: Additional free text description for variables
  - \* **citation**: Citation Requests/Acknowledgements
- **variables**: Contains variable details presented in a tabular/dataframe format
  - **name**: Variable name
  - **role**: Whether the variable is an ID, feature, or target
  - **type**: Data type e.g. categorical, integer, continuous
  - **demographic**: Indicates whether the variable represents demographic data
  - **description**: Short description of variable
  - **units**: Variable units for non-categorical data
  - **missing\_values**: Whether there are missing values in the variable's column

## Examples

```
# Access Data by Name
iris_dl <- fetch_ucirepo(name = "iris")

# Access original data
iris_uci <- iris_dl$data$original

# Access features and targets
iris_features <- iris_dl$data$features
iris_targets <- iris_dl$data$targets
```

```
# Access Data by ID
iris_dl <- fetch_ucirepo(id = 53)
```

---

```
list_available_datasets
```

*List Available Datasets from UCI ML Repository*

---

### Description

Prints a list of datasets that can be imported via the `fetch_ucirepo` function.

### Usage

```
list_available_datasets(filter, search, area)
```

### Arguments

|                     |   |
|---------------------|---|
| <code>filter</code> | Character. Optional query to filter available datasets based on a label.      |
| <code>search</code> | Character. Optional query to search for available datasets by name.           |
| <code>area</code>   | Character. Optional query to filter available datasets based on subject area. |

### Value

A data frame containing the list of available datasets with columns of:

- **id**: Integer ID for the data set.
- **name**: Name of Dataset
- **url**: Download location of the data set

In the event the search fails, the data frame returned will be empty.

### Examples

```
list_available_datasets(search = "iris")
list_available_datasets(area = "social science")
list_available_datasets(filter = "python") # Required for now...
```

# Index

`fetch_ucirepo`, [2](#)

`list_available_datasets`, [4](#)