

Package ‘rHealthDataGov’

January 27, 2015

Type Package

Title Retrieve data sets from the HealthData.gov data API

Version 1.0.1

Date 2014-06-01

Author Erin LeDell

Maintainer Erin LeDell <ledell@berkeley.edu>

Description An R interface for the HealthData.gov data API. For each data resource, you can filter results (server-side) to select subsets of data.

License GPL-2

Depends R (>= 3.0.1), bit64, httr, jsonlite

LazyLoad yes

LazyData yes

NeedsCompilation no

Repository CRAN

Date/Publication 2014-06-02 01:14:25

R topics documented:

rHealthDataGov-package	2
fetch_healthdata	2
filters	4
resources	5
Index	7

rHealthDataGov-package

Retrieve data sets from the HealthData.gov data API

Description

An R interface for the HealthData.gov data API. For each data resource, you can filter results (server-side) to select subsets of data.

Details

Package: rHealthDataGov
Type: Package
Version: 1.0.1
Date: 2014-06-01
License: GPL-2

The main function of this package is `fetch_healthdata`, which is used to query (by filter) and retrieve data from the HealthData.gov data API. Currently, the only interesting data store available via the HealthData.gov API is the "Hospital Compare" data store. This is comprised of 33 data sets containing information about process of care, mortality, and readmission quality measures for U.S. hospitals. This package is part of the rOpenHealth project: <https://github.com/rOpenHealth>

Author(s)

Erin LeDell

Maintainer: Erin LeDell <ledell@berkeley.edu>

References

<http://www.healthdata.gov>
<http://www.healthdata.gov/data-api>
<http://hub.healthdata.gov/dataset/hospital-compare-api>

fetch_healthdata

Fetch HealthData.gov data sets

Description

Query and retrieve data from the HealthData.gov data API.

Usage

```
fetch_healthdata(resource = "hosp", filter = NULL)
```

Arguments

resource	A string that identifies the name of the desired data resource. See the resources object for names and descriptions of the available data resources. Any name from the resources\$resource column can be a value here.
filter	A list of named filters to apply to the API call. The named list elements must be a field in the given resource. To return all records from a particular data resource, set filter to NULL.

Details

The resources data frame and filters list will be lazy-loaded automatically when you load the package. You can also load them explicitly using the data(resources) and data(filters) commands. The filter(s) will be applied on the server-side.

Value

A data frame containing the results of the API query.

Note

The HealthData.gov API only returns 100 results per HTTP request, so if your query matches more than 100 rows, multiple HTTP requests will be made. After all the records that match your query are retrieved via the API, a data frame containing all the records will be returned. Field types will be converted automatically using the field type information returned by the API. Support for 64-bit integers is provided by the required **bit64** package. Some date fields are designated by the API as "text", and therefore will not be converted automatically. However, there are some fields that are designated as having type, "timestamp", and these columns will be converted from a UTC character string (e.g. "2011-01-01T00:00:00") to R base class "POSIXct".

Author(s)

Erin LeDell

References

<http://www.healthdata.gov/data-api>

Examples

```
## Not run:  
df <- fetch_healthdata(resource="hosp", filter=list(addr_city="SAN FRANCISCO"))  
head(df)
```

```
#   addr_city provider_id  tel_nbr seqn      addr_line_1  
# 1 SAN FRANCISCO      50076 4158332646  38      2425 GEARY BLVD  
# 2 SAN FRANCISCO      50228 4152068000  641      1001 POTRERO AVENUE
```

```

# 3 SAN FRANCISCO      50668 4157592300 660   375 LAGUNA HONDA BLVD
# 4 SAN FRANCISCO      50008 4156006000 1207   45 CASTRO STREET
# 5 SAN FRANCISCO      50152 4153536000 2353   900 HYDE ST
# 6 SAN FRANCISCO      50055 4156416562 2911 3555 CESAR CHAVEZ STREET
#
# ownership_type hsp_accreditation addr_postalcode
# 1 Government - Hospital District or Authority 94115
# 2 Government - Local 94110
# 3 Government - Local 94116
# 4 Voluntary non-profit - Other 94114
# 5 Voluntary non-profit - Private 94109
# 6 Voluntary non-profit - Church 94110
# emergency_serv_type addr_state _id hospital_type
# 1 Yes CA 38 Short-term
# 2 Yes CA 641 Short-term
# 3 Yes CA 660 Short-term
# 4 Yes CA 1207 Short-term
# 5 Yes CA 2353 Short-term
# 6 Yes CA 2911 Short-term
#
# hsp_name county_cd
# 1 KAISER FOUNDATION HOSPITAL - SAN FRANCISCO 480
# 2 SAN FRANCISCO GENERAL HOSPITAL 480
# 3 LAGUNA HONDA HOSPITAL & REHABILITATION CENTER 480
# 4 CALIFORNIA PACIFIC MEDICAL CTR-DAVIES CAMPUS HOSP 480
# 5 SAINT FRANCIS MEMORIAL HOSPITAL 480
# 6 CALIFORNIA PACIFIC MEDICAL CTR - ST. LUKES CAMPUS 480

```

```
## End(Not run)
```

filters

List of supported fields and field values for each of the data resources.

Description

This is a list containing elements for each resource. Each resource element is another list that contains named (filter names) vectors of values (unique filter values).

Usage

```
data(filters)
```

Format

The format is:
List of 33

Details

This list was created by by calling the `fetch_healthdata` function (with `filter = NULL`) on all of the resources and returning the unique values for each column in the resulting data frame.

Source

<http://hub.healthdata.gov/dataset/hospital-compare-api>

References

<http://www.healthdata.gov/data-api>

Examples

```
data(filters)

str(filters$hosp$addr_state)
# chr [1:55] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" ...

str(filters$hosp$addr_city)
# chr [1:2832] "ABBEVILLE" "ABERDEEN" "ABILENE" "ABINGDON" ...
```

resources

HealthData.gov resources metadata

Description

This is a data frame that contains the resource names which can be passed as a string to the resource argument of the `fetch_healthdata` function. It also contains resource descriptions and other metadata.

Usage

```
data(resources)
```

Format

A data frame with 33 observations on the following 6 variables.

`resource` a factor with levels `ahrqn ahrqp ahrqs cacn cacop cacp cacs cn cp cs hacn haccp haip hais hn hophc hophp hopnp hopqdrpq hopsp hosp hp hpv hs npv on op oq os ppv q sm spv`

`description` a character vector

`date` a Date

`nrow` a numeric vector

`ncol` a numeric vector

`resource_id` a factor (with levels identified by a hash assigned by HealthData.gov)

Details

There are 33 available resources which can be queried by their resource name. The names, descriptions, source dates, number of records and fields, and resource ids are available in this data frame.

Source

<http://hub.healthdata.gov/dataset/hospital-compare-api>

References

<http://www.healthdata.gov/data-api>

Examples

```
data(resources)
head(resources[,1:2])
#   resource                                     description
# 1   ahrqn Healthcare Research and Quality Indicators, Natioanal data
# 2   ahrqp Healthcare Research and Quality Indicators, Providers
# 3   ahrqs Healthcare Research and Quality Indicators, State data
# 4   cacn Childrens Asthma Care National (CACN)
# 5   cacop Childrens Asthma Care Only Providers (CACOP)
# 6   cacp Childrens Asthma Care Providers (CACP)
```

Index

*Topic **API**

fetch_healthdata, [2](#)

*Topic **datasets**

filters, [4](#)

resources, [5](#)

*Topic **package**

rHealthDataGov-package, [2](#)

fetch_healthdata, [2](#)

filters, [4](#)

resources, [5](#)

rHealthDataGov

(rHealthDataGov-package), [2](#)

rHealthDataGov-package, [2](#)