Package ‘rHealthDataGov’

February 20, 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

Retrieves 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 you 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
# 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 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.
resources

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 hACP
    haip haiS hn hophc hophp hopn hopqdrpq hopsp hosp hp hv 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 ahrq HRQI Healthcare Research and Quality Indicators, National data
# 2 ahrqph Healthcare Research and Quality Indicators, Providers
# 3 ahrqss Healthcare Research and Quality Indicators, State data
# 4 cacn CACN Childrens Asthma Care National (CACN)
# 5 cacop CACOP Childrens Asthma Care Only Providers (CACOP)
# 6 cacp 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