

Package ‘HockeyScrapR’

January 16, 2017

Type Package

Title Tools to scrape hockey data

Version 0.1

Date 2017-01-06

Author Philip Bulsink

Maintainer Philip Bulsink <bulsinkp@gmail.com>

Description Tools to scrape data from Hockey-Reference.com, and DailyFaceoff.com.

License MIT + file LICENSE

Imports RCurl, plyr, stringr, XML, utils

Suggests testthat

RoxygenNote 5.0.1

NeedsCompilation no

R topics documented:

flattenTables	1
getCurrentRosters	2
getPlayerList	3
getPlayerStats	3
processPlayerData	4
scrapePlayers	4
scrapeRosters	5
scrapeScores	6
updatePlayers	6
updateScores	7
Index	8

flattenTables	<i>Scrape & compile a player's info Read each table on hockey-reference.com and flatten to single useful tables. Avoids tables such as last5, awards, etc. Melts all NHL (regular season and playoff) and non-NHL, recording league and playoff status</i>
---------------	--

Description

Scrape & compile a player's info Read each table on hockey-reference.com and flatten to single useful tables. Avoids tables such as last5, awards, etc. Melts all NHL (regular season and playoff) and non-NHL, recording league and playoff status

Usage

```
flattenTables(tables)
```

Arguments

tables	Tables output from XML : : readHTMLTables of the player's page
--------	--

Value

single combined stats Table as data.frame

getCurrentRosters	<i>Get Current estimated rosters</i>
-------------------	--------------------------------------

Description

Get Current estimated rosters

Usage

```
getCurrentRosters(sleep = 30, teamUrllist = teamURLList)
```

Arguments

sleep	The amount of time to sleep between requests
teamUrllist	A list of team roster page urls on dailyfaceoff.com. Defaults to correct team URLs.

Value

a long data frame with three columns:

Team	The team
Player	The name of the Forward, Defence or Goalie playing
updateDate	The date the Team's page was updated

getPlayerList	<i>Get a list of Players Download the full list of active and past players from Hockey-Reference.com</i>
---------------	--

Description

Get a list of Players Download the full list of active and past players from Hockey-Reference.com

Usage

```
getPlayerList(sleep = 30, letters = letters)
```

Arguments

sleep	The sleep time between scrape requests
letters	The letters to scrape player names from. Default: letters

Value

a data.frame containing:

Complete	The complete line that data was scraped from
BlnNHL	Whether the player played any time in the NHL
URL	Player page URL
Name	Player Name
Active	Whether the player is currently active

getPlayerStats	<i>Get stats for a list of players A function to get all player stats from a list of players (in the form provided by getPlayerList)</i>
----------------	--

Description

Get stats for a list of players A function to get all player stats from a list of players (in the form provided by [getPlayerList](#))

Usage

```
getPlayerStats(player_list, sleep = 30)
```

Arguments

player_list	A player list (data.frame) from getPlayerList
sleep	Time to sleep between player scrapings

Value

a list of three data.frames, containing

PlayerStats	Combined player statistics
GoalieStats	Combined goalie statistics
PlayerMeta	Meta statistics for all (goalies and players)

processPlayerData	<i>Clean Player Data This function will process player data, returning clean data frames as a list</i>
-------------------	--

Description

Clean Player Data This function will process player data, returning clean data frames as a list

Usage

```
processPlayerData(player_data, drop_awards = TRUE, ...)
```

Arguments

player_data	The player_data to clean up
drop_awards	Whether to drop awards column.
...	Additional parameters to pass

Value

a list of three cleaned data.frames, containing

PlayerStats	Combined player statistics
GoalieStats	Combined goalie statistics
PlayerMeta	Meta statistics for all (goalies and players)

scrapePlayers	<i>Scrape and clean all player data</i>
---------------	---

Description

This is a one-command function to scrape and clean all player data available from Hockey-Reference.com. This takes many hours. Also saves results to a dated .RDS file.

Usage

```
scrapePlayers(data_dir = "../data/players/", ...)
```

Arguments

data_dir	Directory to store data
...	Additional parameters to pass

Value

a list of three cleaned data.frames, containing

PlayerStats	Combined player statistics
GoalieStats	Combined goalie statistics
PlayerMeta	Meta statistics for all (goalies and players)

Examples

```
## Not run: scrapePlayers()  
## Not run: scrapePlayers(sleep=15, long_sleep=180)
```

scrapeRosters	<i>Scrape all roster data</i>
---------------	-------------------------------

Description

This is a one-command function to scrape all roster data available from DailyFaceoff.com. Also saves results to a dated .RDS file.

Usage

```
scrapeRosters(data_dir = "../data/rosters/", ...)
```

Arguments

data_dir	Directory to store data in
...	Additional parameters to pass

Value

roster information data.frame

Examples

```
## Not run:  
scrapeRosters()  
  
## End(Not run)
```

scrapeScores	<i>Scrape and clean all score data</i>
--------------	--

Description

This is a one-command function to scrape and clean all score data available from Hockey-Reference.com. Also saves results to a dated .RDS file.

Usage

```
scrapeScores(data_dir = "./data/scores/", ...)
```

Arguments

data_dir	Directory to store data in.
...	Additional parameters to pass

Value

scores information data.frame

Examples

```
## Not run:
scrapeScores()

## End(Not run)
```

updatePlayers	<i>Update player information (don't rescrape old players).</i>
---------------	--

Description

Update player information (don't rescrape old players).

Usage

```
updatePlayers(player_data, data_dir = "./data/players/", years_back = 1,
  player_list = NULL, ...)
```

Arguments

player_data	The player_data data.frame to update
data_dir	The data dir of stored player information
years_back	The number of years to go back (will scrape all players currently active + past years_back years)
player_list	Optional player list to scrape from (default NULL → get updated player list)
...	Additional parameters to pass

Value

a list of three updated cleaned data.frames, containing

PlayerStats	Combined player statistics
GoalieStats	Combined goalie statistics
PlayerMeta	Meta statistics for all (goalies and players)

Examples

```
## Not run: updatePlayers()  
## Not run: updatePlayers(data_dir = './data/', years_back=2)
```

updateScores	<i>Update scores information (don't rescrape old scores).</i>
--------------	---

Description

Update scores information (don't rescrape old scores).

Usage

```
updateScores(score_data, data_dir = "./data/scores/", ...)
```

Arguments

score_data	The player_data data.frame to update
data_dir	The data dir of stored player information
...	Additional parameters to pass

Value

updated scores information data.frame

Examples

```
## Not run: updateScores()  
## Not run: updatePlayers(score_data, data_dir = './data/')
```

Index

flattenTables, [1](#)

getCurrentRosters, [2](#)
getPlayerList, [3](#), [3](#)
getPlayerStats, [3](#)

processPlayerData, [4](#)

scrapePlayers, [4](#)
scrapeRosters, [5](#)
scrapeScores, [6](#)

updatePlayers, [6](#)
updateScores, [7](#)