

Package ‘hockeyR’

August 22, 2022

Title Collect and Clean Hockey Stats

Version 1.0.0

Description Various functions to scrape hockey play-by-play data from the [<https://www.nhl.com/>](https://www.nhl.com/). It also contains functions to scrape data from [<https://www.hockey-reference.com/>](https://www.hockey-reference.com/), including standings, player stats, and jersey number history.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

URL <https://github.com/danmorse314/hockeyR>

BugReports <https://github.com/danmorse314/hockeyR/issues>

Depends R (>= 3.5.0)

Imports dplyr, glue, httr, janitor, jsonlite, magrittr, lubridate, polite, purrr, rvest, stringr, tidyr, tidyselect, utils, zoo

Suggests ggimage, ggplot2, ggrepel, knitr, rmarkdown, scales, sportyR

VignetteBuilder knitr

NeedsCompilation no

Author Daniel Morse [aut, cre]

Maintainer Daniel Morse <danmorse8642@gmail.com>

Repository CRAN

Date/Publication 2022-08-22 08:10:06 UTC

R topics documented:

fix_player_columns	2
get_current_rosters	3
get_draft_class	3
get_game_ids	4

get_game_info	5
get_game_rosters	6
get_game_shifts	6
get_goalie_stats_hr	7
get_jersey_players	7
get_player_stats_hr	8
get_rosters	9
get_skater_stats_hr	10
get_team_records	10
get_team_rosters	11
load_pbp	12
scrape_day	12
scrape_game	13
scrape_season	17
team_logos_colors	17

Index	19
--------------	-----------

fix_player_columns	<i>Fix event player columns</i>
--------------------	---------------------------------

Description

Helper function to parse player columns from JSON pbp data & clean up the column names

Usage

```
fix_player_columns(players, col)
```

Arguments

players	A tibble of event players from JSON pbp data
col	Integer between 1 & 4 for which event player is being evaluated

Value

A tibble of an individual event player with cleaned column names

get_current_rosters *Get current NHL rosters*

Description

Get current NHL rosters

Usage

```
get_current_rosters()
```

Value

A tibble containing the current rosters for every team per the NHL

player_id NHL player ID, as an integer

player Player name as a string

jersey_number Player jersey number, as an integer

position Abbreviated official player position, as a string

position_type Abbreviated position group: F, D, or G

team_id NHL integer team ID

full_team_name Full team name as a string

team_abbr Team abbreviation, as a string

Examples

```
## Not run:  
current_rosters <- get_current_rosters()  
  
## End(Not run)
```

get_draft_class *Get draft classes*

Description

Get all selections for any single NHL draft class back to 1963

Usage

```
get_draft_class(  
  draft_year = as.numeric(format(Sys.Date() - 181, "%Y")),  
  player_details = FALSE  
)
```

Arguments

draft_year Draft year to scrape
player_details If true, returns more detailed data on each prospect

Value

A tibble containing all selections for the supplied draft year.

Basic draft class variables:

- Draft year
- Round
- Overall Pick #
- Round Pick #
- Drafting team
- Player ID
- Player name
- Player link

Detailed draft class variables:

- Player birthdate
- Player birthplace & nationality
- Player height & weight
- Player shoots/catches
- Player position
- Player amateur team & league

Examples

```
## Not run:  
draft_2022 <- get_draft_class(draft_year = 2022, player_details = TRUE)  
  
## End(Not run)
```

get_game_ids

Fetch game IDs for a single day or a full season

Description

Fetch game IDs for a single day or a full season

Usage

```
get_game_ids(season = NULL, day = as.Date(Sys.Date(), "%Y-%m-%d"))
```

Arguments

season An integer value denoting the end year of the season to scrape
day A day in the format of 'YYYY-MM-DD'

Value

A tibble containing game IDs for specified time frame

Examples

```
## Not run:  
get_game_ids(season = 2020)  
get_game_ids(day = "2015-03-12")  
  
## End(Not run)
```

get_game_info *Gather basic game information*

Description

Scrapes basic game info like date, venue, & information about the home and away teams for a given game

Usage

```
get_game_info(game_id)
```

Arguments

game_id Game ID to scrape (Can be found using get_game_ids function)

Value

A 1xN tibble containing N pieces of information about the specified game

get_game_rovers	<i>Scrape game day rosters for individual game</i>
-----------------	--

Description

Scrapes the game-day rosters for both teams in the given game ID

Usage

```
get_game_rovers(game_id)
```

Arguments

game_id Game ID to scrape (Can be found using get_game_ids function)

Value

A tibble containing player names, ids, and positions for both team rosters in a given game.

get_game_shifts	<i>Fetch game shift data</i>
-----------------	------------------------------

Description

A function to gather shift data from a given game. Shifts are turned into events to match the style of events in standard game pbp.

Portions of this code are modified versions of code from the NHL scraper by Evolving-Hockey, which in turn were modified from the NHL scraper by Manny Perry.

Usage

```
get_game_shifts(game_id)
```

Arguments

game_id Game ID to scrape

Value

A tibble containing each player change as an event

Examples

```
## Not run:  
get_game_shifts(2020020561)  
  
## End(Not run)
```

get_goalie_stats_hr *Scrape goalie stats*

Description

A function to scrape all goalie stats from a single season via hockey-reference.com

Usage

```
get_goalie_stats_hr(season = as.numeric(format(Sys.Date() + 81, "%Y")))
```

Arguments

season Integer value denoting the end year of the season to scrape

Value

A tibble containing all goalie stats found on hockey-reference.com for the given season

Examples

```
## Not run:  
get_goalie_stats_hr(2022)  
  
## End(Not run)
```

get_jersey_players *Get Players by Jersey Number*

Description

Get the name of every player to wear a specific jersey number in the NHL and the season in which they wore it.

Usage

```
get_jersey_players(jersey)
```

Arguments

jersey An integer or a vector of integers between 0 & 99

Value

A tibble containing each player-season where a player wore the specified number

Examples

```
## Not run:  
#' get_jersey_players(69)  
get_jersey_players(c(99, 66))  
  
## End(Not run)
```

get_player_stats_hr *Get player counting stats from hockey-reference.com*

Description

Get player counting stats from hockey-reference.com

Usage

```
get_player_stats_hr(player_name, season = "career", league = "NHL")
```

Arguments

player_name	A player name or vector of player names
season	An integer value denoting the end year of the season(s) to scrape
league	The league stats to scrape, either 'NHL' or 'WHA'

Value

A tibble containing goals, assists, and various other stats for the specified player(s) from hockey-reference.com

Examples

```
## Not run:  
#' get_player_stats("Wayne Gretzky")  
get_player_stats_hr(c("Wayne Gretzky", "Mario Lemieux"))  
  
## End(Not run)
```

`get_rosters`*Get team rosters from hockey-reference.com*

Description

Get the latest roster for any team from hockey-reference.com. You may enter either the team abbreviation or the full team name. Seasons must be 4-digit integers denoting the end-year of the regular season desired (ie 2021-22 season should be 2022)

Please note that this uses the hockey-reference.com team abbreviations, the oddest of which is Vegas being 'VEG' instead of 'VGK'. If you are unsure of the team abbreviation, enter the full team name instead, or check the full team abbreviations data and filter to your desired season:

```
filter(team_abbr_yearly, season_short == {season})
```

Usage

```
get_rosters(  
  team = "all",  
  season = as.numeric(format(Sys.Date() + 184, "%Y")),  
  include_stats = FALSE  
)
```

Arguments

<code>team</code>	A character vector of team names or abbreviations
<code>season</code>	An integer value denoting the end year of the season to scrape
<code>include_stats</code>	Set to TRUE to return player counting stats for the season

Value

A tibble containing the latest roster for the specified team(s) in the specified season

Examples

```
## Not run:  
get_rosters("SEA", 2022)  
  
## End(Not run)
```

```
get_skater_stats_hr      Scrape skater stats
```

Description

A function to scrape all skater stats from a single season via hockey-reference.com

Usage

```
get_skater_stats_hr(season = as.numeric(format(Sys.Date() + 81, "%Y")))
```

Arguments

season Integer value denoting the end year of the season to scrape

Value

A tibble containing all skater stats found on hockey-reference.com for the given season

Examples

```
## Not run:
get_skater_stats_hr(2022)

## End(Not run)
```

```
get_team_records      Get team records from hockey-reference.com
```

Description

Get team records from hockey-reference.com

Usage

```
get_team_records(
  season = as.numeric(format(Sys.Date() + 184, "%Y")),
  league = "NHL",
  include_records = TRUE
)
```

Arguments

season An integer value denoting the end year of the season(s) to scrape
 league The league stats to scrape, either 'NHL' or 'WHA' or get both with c('NHL','WHA')
 include_records Option to exclude records from the function, used to gather full team names & abbreviations for every season since 1918

Value

A tibble containing full team names & win-loss records for teams in all desired seasons

Examples

```
## Not run:  
get_team_records(2021)  
  
## End(Not run)
```

get_team_rovers	<i>Get current team roster</i>
-----------------	--------------------------------

Description

Get current team roster

Usage

```
get_team_rovers(team)
```

Arguments

team A team name or abbreviation, as a string - or the NHL integer team ID

Value

A tibble containing the current official team roster per NHL.com

Examples

```
## Not run:  
get_team_rovers("SEA")  
  
## End(Not run)
```

load_pbp	<i>Load season play-by-play</i>
----------	---------------------------------

Description

Load season play-by-play

Usage

```
load_pbp(  
  season = as.numeric(substr(Sys.Date() + 184, 1, 4)),  
  shift_events = FALSE  
)
```

Arguments

season	An integer value or vector of values denoting the end year of the season(s) to scrape. <code>load_pbp</code> also accepts character strings with more explicit definitions of the season to scrape: '2020-2021', '2020-21', '2020_21' are also acceptable. The default value is the current season, switching to the next year on July 1st when the new league year begins.
shift_events	Logical value; when set to FALSE this function returns a smaller dataset that excludes specifically shift change events

Value

A tibble containing all play-by-play data for a given season(s) in the same format as the output of [scrape_game](#)

Examples

```
## Not run:  
pbp <- load_pbp(2021)  
  
## End(Not run)
```

scrape_day	<i>Scrape play-by-play for single day's games</i>
------------	---

Description

Scrape play-by-play for single day's games

Usage

```
scrape_day(day = as.Date(Sys.Date(), "%Y-%m-%d"))
```

Arguments

`day` A day in the format of 'YYYY-MM-DD'; defaults to system date.

Value

A tibble containing all play-by-play data for a given day in the same format as the output of [scrape_game](#)

Examples

```
## Not run:
bbp_day <- scrape_day(day = "2015-01-06")

## End(Not run)
```

scrape_game	<i>Scrape game play-by-play</i>
-------------	---------------------------------

Description

Scrapes play-by-play data for a specified game ID.

Usage

```
scrape_game(game_id)
```

Arguments

`game_id` Game ID to scrape

Value

A tibble containing event-based play-by-play data for an individual NHL game. The resulting data will have columns for:

event String defining the event

event_type String with alternate event definition; in all caps

secondary_type String defining secondary event type

event_team String defining the primary team involved in the event

event_team_type String indicator of event team type: home or away

description String detailed description of event

period Integer value of the game period

period_seconds Numeric value of the seconds into the period of the event

period_seconds_remaining Numeric value of the seconds remaining in the period

game_seconds Numeric value of the seconds into the game of the event

game_seconds_remaining Numeric value of the seconds remaining in the game; negative for over-time periods

home_score Integer value of the home team score after the event

away_score Integer value of the away team score after the event

event_player_1_name String name of the primary event player

event_player_1_type String indicator for the role of event_player_1

event_player_2_name String name of the secondary event player

event_player_2_type String indicator for the role of event_player_2

event_player_3_name String name of the tertiary event player

event_player_3_type String indicator for the role of event_player_3

event_goalie_name String name of the goalie involved in the event

strength_code String indicator for game strength: EV, SH, or PP

strength String name for game strength: Even, Shorthanded, or Power Play

strength_state String name for detailed game strength in the form of '(event team skaters)v(opponent skaters)'

penalty_minutes Integer value of the penalty minutes on penalty events

penalty_severity String name for penalty severity: Minor or Major

num_on Integer value of the number of skaters substituted on during a shift change event

players_on String of player names substituted on during a shift change event

num_off Integer value of the number of skaters substituted off during a shift change event

players_off String of player names substituted off during a shift change event

extra_attacker Logical indicator of whether or not the event team had their goalie pulled

x Numeric x-coordinate of event in feet, with origin at center ice

y Numeric y-coordinate of event in feet, with origin at center ice

x_fixed Numeric transformed x-coordinate of event in feet, where the home team always shoots to the right, away team to the left

y_fixed Numeric transformed y-coordinate of event in feet, where the home team always shoots to the right, away team to the left

shot_distance Numeric distance (in feet) to center of net for unblocked shot events

shot_angle Numeric angle (in degrees) to center of net for unblocked shot events

home_skaters Numeric value for number of skaters on the ice for the home team, excluding the goalie

away_skaters Numeric value for number of skaters on the ice for the away team, excluding the goalie

home_on_1 String name of home team player on ice

home_on_2 String name of home team player on ice

home_on_3 String name of home team player on ice

home_on_4 String name of home team player on ice

home_on_5 String name of home team player on ice

home_on_6 String name of home team player on ice
home_on_7 String name of home team player on ice
away_on_1 String name of away team player on ice
away_on_2 String name of away team player on ice
away_on_3 String name of away team player on ice
away_on_4 String name of away team player on ice
away_on_5 String name of away team player on ice
away_on_6 String name of away team player on ice
away_on_7 String name of away team player on ice
home_goalie String name of home goalie on ice
away_goalie String name of away goalie on ice
game_id Integer value of assigned game ID
event_idx Numeric index for event
event_player_1_id Integer value of the player ID for the primary event player
event_player_1_link String value of the NHL.com player link for the primary event player
event_player_1_season_total Integer value for the total events for the primary event player this season
event_player_2_id Integer value of the player ID for the secondary event player
event_player_2_link String value of the NHL.com player link for the secondary event player
event_player_2_season_total Integer value for the total events for the secondary event player this season
event_player_3_id Integer value of the player ID for the tertiary event player
event_player_3_link String value of the NHL.com player link for the tertiary event player
event_player_3_season_total Integer value for the total events for the tertiary event player this season
event_goalie_id Integer value of the player ID for the event goalie
event_goalie_link String value of the NHL.com player link for the event goalie
event_goalie_type String indicator for the role of the event_goalie
game_winning_goal Logical indicator of whether or not the goal scored was the game-winning goal
empty_net Logical indicator of whether or not the goal scored was on an empty net
period_type String name of period type: REGULAR, OVERTIME, or SHOOTOUT
ordinal_num String name of the ordinal period: 1st, 2nd, 3rd, 4th...
period_time String value of the time into the period of the event
period_time_remaining String value of the time remaining in the period
date_time String value of the real-world timestamp of the event
event_team_id Integer value of the NHL ID of event_team
event_team_link String value of the NHL.com team link for the event_team

event_team_abbr String value of the 3-letter NHL abbreviation for the event_team

home_final Integer value of the final score for the home team

away_final Integer value of the final score for the away team

season String value of the official NHL season

season_type String indicator of season type: R, or P

game_date Date of game

game_start Date time of start of game in US/Eastern time zone

game_end Date time of end of game in US/Eastern time zone

game_length Period value of length of game, in hours:minutes

game_state String indicator of state of game

detailed_state String indicator of detailed game state

venue_id Integer value of the NHL ID for the venue

venue_name String name of the game venue

venue_link String value of the NHL.com link for the venue

home_name String name of the home team

home_abbreviation String value of the 3-letter NHL abbreviation of the home team

home_division_name String value of the name of the NHL division of the home team

home_conference_name String value of the name of the NHL conference of the home team

home_id Integer value of the NHL ID of the home team

away_name String name of the away team

away_abbreviation String value of the 3-letter NHL abbreviation of the away team

away_division_name String value of the name of the NHL division of the away team

away_conference_name String value of the name of the NHL conference of the away team

away_id Integer value of the NHL ID of the away team

Examples

```
## Not run:  
pbp <- scrape_game(2020020420)  
  
## End(Not run)
```

scrape_season	<i>Scrape full season play-by-play</i>
---------------	--

Description

Scrape full season play-by-play

Usage

```
scrape_season(season, type = "all")
```

Arguments

season	An integer value denoting the end year of the season to scrape
type	A character vector of the game types to include: REG, POST, or "all"

Value

A tibble containing all play-by-play data for a given season in the same format as the output of [scrape_game](#)

Examples

```
## Not run:  
# scrape all regular season & postseason games for the 2016-2017 season  
pbp_2016_2017 <- scrape_season(2017, type = "REG")  
  
## End(Not run)
```

team_logos_colors	<i>Team logos & colors</i>
-------------------	--------------------------------

Description

A dataset containing the full team names, abbreviations, colors & logos for all 32 NHL teams (plus some defunct teams)

Usage

```
team_logos_colors
```

Format

A data frame with 35 rows and 12 variables:

full_team_name full team name
team_abbr NHL.com team abbreviation
team_nick lowercase, no spaces team nickname
division current NHL division
conference current NHL conference
team_logo_espn primary team logo from ESPN.com
team_color1 current primary team color
team_color2 current secondary team color
team_logo_alterate alternate or throwback logo
team_color_alt1 alternate logo primary color
team_color_alt2 alternate logo secondary color
status active or inactive

Source

<https://www.espn.com/nhl/teams>

https://www.sportslogos.net/teams/list_by_league/1/National_Hockey_League/NHL/logos/

<https://teamcolorcodes.com/nhl-team-color-codes/>

Index

* datasets

team_logos_colors, [17](#)

fix_player_columns, [2](#)

get_current_rosters, [3](#)

get_draft_class, [3](#)

get_game_ids, [4](#)

get_game_info, [5](#)

get_game_rosters, [6](#)

get_game_shifts, [6](#)

get_goalie_stats_hr, [7](#)

get_jersey_players, [7](#)

get_player_stats_hr, [8](#)

get_rosters, [9](#)

get_skater_stats_hr, [10](#)

get_team_records, [10](#)

get_team_rosters, [11](#)

load_pbp, [12](#)

scrape_day, [12](#)

scrape_game, [12](#), [13](#), [13](#), [17](#)

scrape_season, [17](#)

team_logos_colors, [17](#)