Package ‘squashinformr’

October 26, 2021

Title Politely Web Scrape Data from SquashInfo

Version 0.2.5

Author Hayden MacDonald [aut, cre]

Maintainer Hayden MacDonald <hayden.macdonald.8778@gmail.com>

Description Scrape SquashInfo <http://www.squashinfo.com/> for data on the Professional Squash Association World Tour and other squash events. ‘squashinformr’ functions scrape, parse, and clean data associated with players, tournaments, and rankings.

Depends R (>= 4.0)

Imports dplyr (>= 1.0.2), plyr (>= 1.8.6), tibble (>= 3.0.4), lubridate (>= 1.7.4), stringr (>= 1.4.0), tidyr (>= 1.1.2), rvest (>= 0.3.6), polite (>= 0.1.1), naniar (>= 0.5.0), janitor (>= 2.1.0), Hmisc (>= 4.4.2), rlang (>= 0.4.10)

Suggests testthat, covr, knitr, rmarkdown

License GPL-3

Encoding UTF-8

URL https://github.com/HaydenMacDonald/squashinformr

BugReports https://github.com/HaydenMacDonald/squashinformr/issues

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2021-10-26 18:10:02 UTC

R topics documented:

aggregate_profile_tables .................................................. 2
get_matchup ................................................................. 3
get_matchup_profiles ...................................................... 4
get_match_game_data ....................................................... 5
get_players ................................................................. 5
get_player_profile_table .................................................. 6
aggregate_profile_tables

Aggregate Profile Tables from Player Profiles

Description

Given two data frames with profile urls, fetch and aggregate summary tables from player profiles.

Usage

aggregate_profile_tables(x = NULL, y = NULL, identifier = NULL)

Arguments

x, y  data frames containing

identifier  unique string that used to identify the desired table.

Value

Data frame containing aggregated summary tables.
get_matchup

Get a player matchup data from SquashInfo

Description

Given the full names or ranks of players, and the competition category, get_matchup() returns recent matchup data for PSA ranked players.

Usage

get_matchup(
    player_1 = NULL,
    player_2 = NULL,
    ranks = NULL,
    category = NULL,
    tidy = FALSE,
    match_spread = FALSE
)

Arguments

player_1 character string of the first player’s name.
player_2 character string of the second player’s name.
ranks integers indicating the rank of the PSA players to return.
category character string indicating the competition category. Must be one of "mens" or "womens".
tidy logical indicating whether to organize results according to tidy principles.
match_spread logical indicating whether to only return match spread statistics.

Value

Tibble containing each player’s rank, name, total matches played, number of matches won, the match results spread (relative to player 1), the average match time, the number of games played, the number of games won, average point advantage in a won game, the average point difference in final scores, the number of tie-break wins, and the percentage of games that go to a tie-breaker.

Note

This function only returns data from players ranked in the most recent PSA rankings table for Men’s and Women’s singles competitions.

References

http://www.squashinfo.com/rankings/men
http://www.squashinfo.com/rankings/women
Examples

```r
## Get tidy matchup data for Mohamed Elshorbagy vs Ali Farag
get_matchup(player_1 = "Mohamed Elshorbagy",
             player_2 = "Ali Farag",
             category = "mens",
             tidy = TRUE)

## Get non-tidy matchup data for Nouran Gohar vs Nour El Sherbini
get_matchup("Nouran Gohar", "Nour El Sherbini", category = "womens", tidy = FALSE)

## Get tidy match spread data for Paul Coll and Ali Farag
get_matchup("Paul Coll", "Ali Farag", category = "mens", tidy = TRUE, match_spread = TRUE)
```

get_matchup_profiles  Get profile urls for players in matchup

Description

Get profile urls for players in matchup

Usage

```r
get_matchup_profiles(
  player_1 = NULL,
  player_2 = NULL,
  players = NULL,
  ranks = NULL,
  category = NULL
)
```

Arguments

- `player_1` character string of the first player’s name.
- `player_2` character string of the second player’s name.
- `players` character string vector of player names.
- `ranks` integers indicating the rank of the PSA players to return.
- `category` character string indicating the competition category. Must be one of "mens" or "womens".

Value

Tibble containing player names, rank, and profile url slugs.
**get_match_game_data**

*Given recent match data, return a cleaned list of recent match and recent game data*

**Description**

Given recent match data, return a cleaned list of recent match and recent game data.

**Usage**

```
get_match_game_data(data = NULL, players = NULL, ranks = NULL)
```

**Arguments**

- `data`: recent match data frame.
- `players`: a vector containing player names as strings.
- `ranks`: a vector of integers representing the player ranks.

**Value**

list containing recent match and recent games data frames.

---

**get_players**

*Get PSA Squash Player Data from SquashInfo*

**Description**

Given the rank(s) and competition category, get_players() returns profile data of ranked players in PSA World Tour competitions.

**Usage**

```
get_players(top = NULL, rank = NULL, category = NULL)
```

**Arguments**

- `top`: integer indicating the number of top PSA players by rank to return.
- `rank`: integer indicating the rank of the PSA player(s) to return.
- `category`: character string indicating the competition category. Must be one of "mens", "womens", or "both".

**Value**

Tibble containing first name, last name, age, gender, birthplace, nationality, residence, height in cm, weight in kg, plays (handedness), racket brand, year of joining PSA, university, and club.
Note

This function only returns players ranked in the most recent PSA rankings table for Men’s and Women’s singles competitions.

References

http://www.squashinfo.com/rankings/men
http://www.squashinfo.com/rankings/women

Examples

---

## Return the top 5 ranked players from the Women’s PSA rankings
get_players(top = 5, category = “womens”)

## Return the 5th ranked player from both Men's and Women's PSA rankings
get_players(rank = 5, category = “both”)
---

get_player_profile_table

Get Player Summary Tables from SquashInfo

Description

Given a data frame with player ranks, names, and profile slugs, `get_player_profile_urls()` returns the Event Summary Table of ranked players in PSA World Tour competitions.

Usage

```
get_player_profile_table(data = NULL, identifier = NULL)
```

Arguments

- `data` data frame with columns for player rank (int), names (chr), and profile slugs (chr).
- `identifier` unique string that used to identify the desired table.

Value

Data frame containing Event Summary Table(s).
get_player_profile_urls

Get Player URLs from SquashInfo

Description

Given the player name, rank(s), and competition category, get_player_profile_urls() returns profile slugs of ranked players in PSA World Tour competitions.

Usage

```r
get_player_profile_urls(
  player = NULL,
  rank = NULL,
  category = c("mens", "womens")
)
```

Arguments

- `player` character string of player name.
- `rank` single integer or vector of integers indicating the rank of the PSA player(s) to return.
- `category` character string indicating the competition category. Must be one of "both", "mens", or "womens".

Value

Data frame containing player ranks, names, and profile slugs.

get_player_recent_

Get a player’s recent data from SquashInfo

Description

Given level of detail (e.g. 'results', 'matches', 'games'), the full name or rank of a player, and the competition category, get_player_recent_() returns the recent event data for PSA ranked players.

Usage

```r
get_player_recent_(
  level = c("results", "matches", "games"),
  player = NULL,
  rank = NULL,
  category = NULL
)
```
get_player_recent_games

Arguments

level character string indicating the level of detail of the data to be returned. Must be one of "results", "matches", or "games".
player character string of player name.
rank single integer or vector of integers indicating the rank of the PSA player(s) to return.
category character string indicating the competition category. Must be one of "both", "mens", or "womens".

Value

Tibble containing the player rank, player name, player’s seeding, round reached, event name, event date, event location, and event tour.

Description

Get a player’s recent games from SquashInfo

Given the full name or rank of a player and the competition category, get_player_recent_games() returns recent game data for PSA ranked players.

Usage

get_player_recent_games(player = NULL, rank = NULL, category = NULL)

Arguments

player character string of player name.
rank single integer or vector of integers indicating the rank of the PSA player(s) to return.
category character string indicating the competition category. Must be one of "both", "mens", or "womens".

Value

Tibble containing the player rank, name, opponent, points won, points lost, game result, tournament round, event, PSA designation, and event location.

Note

This function only returns data from players ranked in the most recent PSA rankings table for Men’s and Women’s singles competitions.
get_player_recent_matches

References

http://www.squashinfo.com/rankings/men
http://www.squashinfo.com/rankings/women

Examples

```r
## Get Mohamed Elshorbagy's most recent game data
get_player_recent_games(player = "Mohamed Elshorbagy", category = "mens")

## Get Nour El Tayeb's recent game data
get_player_recent_games("El Tayeb", category = "womens")

## Get recent game data from the top two players in both Men's and Women's competitions
get_player_recent_games(rank = 1:2, category = "both")
```

get_player_recent_matches

*Get a player's recent matches from SquashInfo*

Description

Given the full name or rank of a player and the competition category, `get_player_recent_matches()` returns recent match data for PSA ranked players.

Usage

```r
get_player_recent_matches(player = NULL, rank = NULL, category = NULL)
```

Arguments

- **player**: character string of player name.
- **rank**: single integer or vector of integers indicating the rank of the PSA player(s) to return.
- **category**: character string indicating the competition category. Must be one of "both", "mens", or "womens".

Value

Tibble containing the player rank, name, opponent, match result, games won, games lost, match time, date, tournament round, event, PSA designation, and event location.

Note

This function only returns data from players ranked in the most recent PSA rankings table for Men’s and Women’s singles competitions.
get_player_recent_results

References

http://www.squashinfo.com/rankings/men
http://www.squashinfo.com/rankings/women

Examples

```r
## Get Mohamed Elshorbagy's most recent match data
get_player_recent_matches(player = "Mohamed Elshorbagy", category = "mens")

## Get Nour El Tayeb's recent match data
get_player_recent_matches("El Tayeb", category = "womens")

## Get recent match data from the top two players in both Men's and Women's competitions
get_player_recent_matches(rank = 1:2, category = "both")
```

get_player_recent_results

Get a player’s recent event results from SquashInfo

Description

Given the full name or rank of a player and the competition category, `get_player_recent_results()` returns the recent event results table for PSA ranked players.

Usage

```r
get_player_recent_results(player = NULL, rank = NULL, category = NULL)
```

Arguments

- **player** character string of player name.
- **rank** single integer or vector of integers indicating the rank of the PSA player(s) to return.
- **category** character string indicating the competition category. Must be one of "both", "mens", or "womens".

Value

Tibble containing the player rank, player name, player’s seeding, round reached, event name, event date, event location, and event tour.

Note

This function only returns results data from players ranked in the most recent PSA rankings table for Men’s and Women’s singles competitions. Recent results are limited to events that occurred within the current and previous calendar years.
get_rankings

References

http://www.squashinfo.com/

Examples

## Get Mohamed Elshorbagy's most recent results data
get_player_recent_results(player = "Mohamed Elshorbagy", category = "mens")

## Get Nour El Tayeb's recent results data
get_player_recent_results("El Tayeb", category = "womens")

## Get recent results data from the top two players in both Men's and Women's competitions
get_player_recent_results(rank = 1:2, category = "both")

get_rankings

Get the most recent PSA rankings from SquashInfo

Description

Given a competition category, get_rankings() returns the most recent PSA rankings table.

Usage

get_rankings(top = NULL, category = NULL)

Arguments

top integer indicating the number of top PSA players by rank to return.
category character string indicating the competition category. Must be one of "both", "mens", or "womens".

Value

Tibble containing the player rank, previous month’s rank, name, highest ranking achieved, date of highest ranking, nationality, and competition category.

Note

This function only returns the most recent PSA rankings table for Men's and Women's singles competitions.

References

http://www.squashinfo.com/rankings/men
http://www.squashinfo.com/rankings/women
Examples

```r
## Get the top 10 ranked men's singles players
get_rankings(top = 10, category = "mens")

## Get the top 10 ranked women's singles players
get_rankings(top = 20, category = "womens")

## Get the top 20 players in both men's and women's singles competitions
get_rankings(top = 20, category = "both")
```

---

**get_rankings_table**  
Get the PSA rankings tables from SquashInfo

**Description**

Given a competition category, `get_rankings_table()` returns the most recent PSA rankings table.

**Usage**

```r
get_rankings_table(top = NULL, category = NULL)
```

**Arguments**

- `top`  
  integer indicating the number of top PSA players by rank to return.

- `category`  
  character string indicating the competition category. Must be one of "mens", or "womens".

**Value**

Tibble containing the player rank, previous month’s rank, name, highest ranking achieved, date of highest ranking, nationality, and competition category.

---

**get_recent_matches**  
Get a player’s recent matches

**Description**

Get a player’s recent matches

**Usage**

```r
get_recent_matches(profile_urls)
```
get_tournaments

Arguments

profile_urls a data frame containing a list of player names, ranks, and profile slugs.

Value

Tibble containing the player’s recent matches.

get_tournaments

Get tournaments from SquashInfo

Description

Given a year, get_tournaments() returns data for PSA World Tour tournaments and other events.

Usage

get_tournaments(year = 2021, world_tour = TRUE)

Arguments

year integer indicating the tournament year. Must be one of 2020 or 2021.
world_tour logical indicating whether to only return PSA World Tour tournaments.

Value

Tibble containing the league, competition category, name, date, city, and country.

Note

This function only returns tournaments from 2020 and 2021, as any other data are not available to non-premium members on SquashInfo.

References

http://www.squashinfo.com/results

Examples

## Get data on 2021 PSA World Tour tournaments
get_tournaments()

## Get data on 2020 non-PSA World Tour tournaments
get_tournaments(2021, world_tour = FALSE)
**get_tournament_games**

Scrape tournament metadata

**Description**

Scrape tournament metadata

**Usage**

```r
get_tournament_(tournament = NULL, year = NULL, world_tour = NULL)
```

**Arguments**

- `tournament` the name of the tournament.
- `year` the year the tournament was held.
- `world_tour` logical indicating whether the tournament was a part of the PSA World Tour.

**Value**

a data frame containing a specified tournament and its slug.

---

**get_tournament_games**

Get a tournament’s games from SquashInfo

**Description**

Given a tournament name or a year, `get_tournament_games()` returns match data for PSA World Tour tournaments and other events.

**Usage**

```r
get_tournament_games(tournament = NULL, year = NULL, world_tour = NULL)
```

**Arguments**

- `tournament` character string of name of the tournament. Partial tournament names are matched via regular expressions.
- `year` integer indicating year of competition. Must be 2020, 2021, or NULL if querying results for both years.
- `world_tour` logical indicating whether to only return PSA World Tour tournaments.

**Value**

Tibble containing the tournament name, competition category, tournament date, round, match number, game number, player 1, player 2, the game winner, player 1’s score, player 2’s score, player 1’s seed, player 2’s seed, player 1’s nationality, player 2’s nationality.
get_tournament_matches

Note

This function only returns game data from tournaments in 2020 and 2021, as any other data are not available to non-premium members on SquashInfo. Additionally, events that do not use a single elimination format are not included in the results (e.g. Karakal Premier League).

References

http://www.squashinfo.com/results

Examples

## Return game data for 2021's CIB Egyptian Open.
get_tournament_games("CIB Egyptian Open", year = 2021, world_tour = TRUE)

## Return game data for all PSA World Tour tournaments in 2021
get_tournament_games(year = 2021, world_tour = TRUE)

get_tournament_matches

Get a tournament's matches from SquashInfo

Description

Given a tournament name or a year, get_tournament_matches() returns match data for PSA World Tour tournaments and other events.

Usage

get_tournament_matches(tournament = NULL, year = NULL, world_tour = NULL)

Arguments

- **tournament**: character string of name of the tournament. Partial tournament names are matched via regular expressions.
- **year**: integer indicating year of competition. Must be 2020, 2021, or NULL if querying results for both years.
- **world_tour**: logical indicating whether to only return PSA World Tour tournaments.

Value

Tibble containing the tournament name, competition category, tournament date, round, player 1, player 2, the match winner, games won (by player 1), games lost (by player 1), the match time, player 1’s seed, player 2’s seed, player 1’s nationality, player 2’s nationality.
### get_tournament_objects

Get players, matches, or games from a given tournament

#### Description

Get players, matches, or games from a given tournament.

#### Usage

```r
get_tournament_objects(tournaments = NULL, level = NULL)
```

#### Arguments

- `tournaments`: a data frame containing a set of tournaments.
- `level`: the level of object to return. Must be one of "players", "matches", or "games".

#### Value

A tibble containing the tournament objects.

---

### Note

This function only returns match data from tournaments in 2020 and 2021, as any other data are not available to non-premium members on SquashInfo. Additionally, events that do not use a single elimination format are not included in the results (e.g. Karakal Premier League).

### References

http://www.squashinfo.com/results

### Examples

```r
## Return match data for 2021's CIB Egyptian Open.
get_tournament_matches("CIB Egyptian Open", year = 2021, world_tour = TRUE)

## Return match data for all PSA World Tour tournaments in 2021
get_tournament_matches(year = 2021, world_tour = TRUE)
```
get_tournament_players

Get a tournament’s players from SquashInfo

Description

Given a tournament name or a year, get_tournament_players() returns player registrants for PSA World Tour tournaments and other events.

Usage

get_tournament_players(tournament = NULL, year = 2021, world_tour = TRUE)

Arguments

tournament character string of name of the tournament. Partial tournament names are matched via regular expressions.
year integer indicating year of competition. Must be 2020, 2021, or NULL if querying results for both years.
world_tour logical indicating whether to only return PSA World Tour tournaments.

Value

Tibble containing the tournament name, competition category, tournament date, player, seed, nationality, round_reached.

Note

This function only returns player registrant data from tournaments in 2020 and 2021, as any other data are not available to non-premium members on SquashInfo. Additionally, events that do not use a single elimination format are not included in the results (e.g. Karakal Premier League).

References

http://www.squashinfo.com/results

Examples

## Who played in the CIB Egyptian Open in 2021?
get_tournament_players(tournament = "CIB Egyptian Open", year = 2021)

## Return player registrant data for all PSA World Tour tournaments in 2021
get_tournament_players(year = 2021, world_tour = TRUE)
Description

The squashinformr package allows sports enthusiasts to easily access data related to the Professional Squash Association’s World Tour and other squash tournaments from SquashInfo. The functions within this package scrape, parse, and clean data associated with players, tournaments, and rankings data.

Details

See the README on GitHub
Index

aggregate_profile_tables, 2
get_match_game_data, 5
get_matchup, 3
get_matchup_profiles, 4
get_player_profile_table, 6
get_player_profile_urls, 7
get_player_recent_, 7
get_player_recent_games, 8
get_player_recent_matches, 9
get_player_recent_results, 10
get_players, 5
get_rankings, 11
get_rankings_table, 12
get_recent_matches, 12
get_tournament_, 14
get_tournament_games, 14
get_tournament_matches, 15
get_tournament_objects, 16
get_tournament_players, 17
get_tournaments, 13

squashinfromr, 18