Package: ncaavolleyballr (via r-universe)

January 24, 2025

Type Package

Title Extract Data from NCAA Women's and Men's Volleyball Website

Version 0.4.1

Maintainer Jeffrey R. Stevens <jeffrey.r.stevens@protonmail.com>

Description Extracts team records/schedules and player statistics for the 2020-2024 National Collegiate Athletic Association (NCAA) women's and men's divisions I, II, and III volleyball teams from <https://stats.ncaa.org>. Functions can aggregate statistics for teams, conferences, divisions, or custom groups of teams.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 4.1)

RoxygenNote 7.3.2

Imports cli, curl, dplyr, httr2, lifecycle, purrr, rlang, rvest, stringr, tibble, tidyr, xml2

Suggests knitr, quarto, testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/JeffreyRStevens/ncaavolleyballr, https://jeffreyrstevens.github.io/ncaavolleyballr/

BugReports https://github.com/JeffreyRStevens/ncaavolleyballr/issues

VignetteBuilder quarto

NeedsCompilation no

Author Jeffrey R. Stevens [aut, cre, cph] (<https://orcid.org/0000-0003-2375-1360>)

Repository CRAN

Date/Publication 2025-01-23 17:30:09 UTC

Config/pak/sysreqs libicu-dev libxml2-dev libssl-dev

Contents

---

conference_stats	*Aggregate player statistics for a NCAA conference and seasons*

Description

This is a wrapper around [group_stats()](group_stats) that extracts season, match, or pbp data from players in all teams in the chosen conference. For season stats, it aggregates all player data and team data into separate data frames and combines them into a list. For match and pbp stats, it aggregates into a data frame. Conferences names can be found in ncaa_conferences.

Usage

```
conference_stats(
  year = NULL,
  conf = NULL,
  level = NULL,
  sport = "WVB",
  save = FALSE,
  path = "."
)
```

Arguments

year	Numeric vector of years for fall of desired seasons.
conf	NCAA conference name.
level	Character string defining whether to aggregate "season", "match", or play-by-play ("pbp") data.
sport	Three letter abbreviation for NCAA sport (must be upper case; for example "WVB" for women's volleyball and "MVB" for men's volleyball).
save	Logical for whether to save the statistics locally as CSVs (default FALSE).
path	Character string of path to save statistics files.

Value

For season level, returns list with data frames of player statistics and team statistics. For match and pbp levels, returns data frame of player statistics and play-by-play information respectively.

See Also

Other functions that aggregate statistics: division_stats(), group_stats()

Examples

```
conference_stats(year = 2024, conf = "Peach Belt", level = "season")
```

---

division_stats	*Aggregate player statistics for a NCAA division and seasons*

Description

This is a wrapper around group_stats() that extracts season, match, or pbp data from players in all teams in the chosen division. For season stats, it aggregates all player data and team data into separate data frames and combines them into a list. For match and pbp stats, it aggregates into a data frame.

Usage

```
division_stats(
  year = NULL,
  division = 1,
  level = NULL,
  sport = "WVB",
  save = FALSE,
  path = "."
)
```

Arguments

year	Numeric vector of years for fall of desired seasons.
division	NCAA division (must be 1, 2, or 3).
level	Character string defining whether to aggregate "season", "match", or play-by-play ("pbp") data.
sport	Three letter abbreviation for NCAA sport (must be upper case; for example "WVB" for women's volleyball and "MVB" for men's volleyball).
save	Logical for whether to save the statistics locally as CSVs (default FALSE).
path	Character string of path to save statistics files.

Value

For season level, returns list with data frames of player statistics and team statistics. For match and pbp levels, returns data frame of player statistics and play-by-play information respectively.

See Also

Other functions that aggregate statistics: `conference_stats()`, `group_stats()`

---

find_team_contests   *Extract date, opponent, and contest ID for team and season*

Description

NCAA datasets use a unique ID for each sport, team, season, and match. This function returns a data frame of dates, opponent team names, and contest IDs for each NCAA contest (volleyball match) for each team and season.

Usage

```
find_team_contests(team_id = NULL)
```

Arguments

team_id	Team ID determined by NCAA for season. To find ID, use `find_team_id()`.

Value

Returns a data frame that includes date, team, opponent, and contest ID for each season's contest.

Examples

```
find_team_contests(team_id = "585290")
```

---

find_team_id *Find team ID for season*

Description

NCAA datasets use a unique ID for each team and season. To access a team's data, we must know the volleyball team ID. This function looks up the team ID from wvb_teams or mvb_teams using the team name. Team names can be found in ncaa_teams or searched with find_team_name().

Usage

```
find_team_id(team = NULL, year = NULL, sport = "WVB")
```

Arguments

team   Name of school. Must match name used by NCAA. Find exact team name with
              find_team_name().

year   Numeric vector of years for fall of desired seasons.

sport   Three letter abbreviation for NCAA sport (must be upper case; for example
              "WVB" for women's volleyball and "MVB" for men's volleyball).

Value

Returns a character string of team ID.

See Also

Other search functions: find_team_name()

Examples

```
find_team_id(team = "Nebraska", year = 2024)
find_team_id(team = "UCLA", year = 2023, sport = "MVB")
```

---

find_team_name *Match pattern to find team names*

Description

This is a convenience function to find NCAA team names in ncaa_teams. Once the proper team name is found, it can be passed to find_team_id() or group_stats().

Usage

```
find_team_name(pattern = NULL)
```

Arguments

pattern   Character string of pattern you want to find in the vector of team names.

Value

Returns a character vector of team names that include the submitted pattern.

See Also

Other search functions: find_team_id()

Examples

```
find_team_name(pattern = "Neb")
```

---

get_teams   *Extract data frame of team names, IDs, conference, division, and season*

Description

NCAA datasets use a unique ID for each sport, team, and season. This function extracts team names, IDs, and conferences for each NCAA team in a division. However, you should not need to use this function for volleyball data from 2020-2024, as it has been used to generate wvb_teams and mvb_teams. However, it is available to use for other sports, using the appropriate three letter sport code drawn from ncaa_sports (e.g., men's baseball is "MBA").

Usage

```
get_teams(year = NULL, division = 1, sport = "WVB")
```

Arguments

year   Single numeric year for fall of desired season.
division   NCAA division (must be 1, 2, or 3).
sport   Three letter abbreviation for NCAA sport (must be upper case; for example "WVB" for women's volleyball and "MVB" for men's volleyball).

Value

Returns a data frame of all teams, their team ID, division, conference, and season.

Note

This function is a modification of the ncaa_teams() function from the {baseballr} package.

---

group_stats   *Aggregate player statistics and play-by-play information*

Description

This function aggregates player statistics and play-by-play information within a season by applying
[player_season_stats()](), [player_match_stats()](), or [match_pbp()]() across groups of teams (for
[player_season_stats()]()) or across contests within a season (for [player_match_stats()]() and
[match_pbp()]()). For season stats, it aggregates all player data and team data into separate data
frames and combines them into a list. For instance, if you want to extract the data from the teams
in the women's 2024 Final Four, pass a vector of c(″Louisville″, ″Nebraska″, ″Penn State″,
″Pittsburgh″) to the function. For match or play-by-play data for a team, pass a single team name
and year. Team names can be found in [ncaa_teams]() or by using [find_team_name()]().

Usage

```
group_stats(
  teams = NULL,
  year = NULL,
  level = ″season″,
  unique = TRUE,
  sport = ″WVB″
)
```

Arguments

teams	Character vector of team names to aggregate.
year	Numeric vector of years for fall of desired seasons.
level	Character string defining whether to aggregate "season", "match", or play-by-play ("pbp") data.
unique	Logical indicating whether to only process unique contests (TRUE) or whether to process duplicated contests (FALSE). Default is TRUE.
sport	Three letter abbreviation for NCAA sport (must be upper case; for example "WVB" for women's volleyball and "MVB" for men's volleyball).

Value

For season level, returns list with data frames of player statistics and team statistics. For match and
pbp levels, returns data frame of player statistics and play-by-play information respectively.

See Also

Other functions that aggregate statistics: [conference_stats](), [division_stats]()

Examples

```
group_stats(teams = c("Louisville", "Nebraska", "Penn St.", "Pittsburgh"),
year = 2024, level = "season")
```

---

match_pbp   *Extract play-by-play information for a particular match*

Description

The NCAA's page for a match/contest includes a tab called "Play By Play". This function extracts the tables of play-by-play information for each set.

Usage

```
match_pbp(contest = NULL)
```

Arguments

contest   Contest ID determined by NCAA for match. To find ID, use `find_team_contests()`
                     for a team and season.

Value

Returns a data frame of set number, teams, score, event, and player responsible for the event.

Examples

```
match_pbp(contest = "6080706")
```

---

mvb_teams   *NCAA Men's Volleyball Teams 2020-2024*

Description

This data frame includes all men's NCAA Division 1 and 3 teams from 2020-2024.

Usage

```
mvb_teams
```

Format

A data frame with 873 rows and 6 columns:

team_id Team ID for season/year

team_name Team name

conference_id Conference ID

conference Conference name

div NCAA division number (1 or 3)

yr Year for fall of season

Source

<https://stats.ncaa.org>

See Also

Other data sets: ncaa_conferences, ncaa_sports, ncaa_teams, wvb_teams

Examples

```
head(mvb_teams)
```

---

ncaa_conferences	*NCAA Conference Names*

Description

This vector includes names for all NCAA volleyball conferences.

Usage

```
ncaa_conferences
```

Format

A character vector with 111 conference names.

Source

<https://stats.ncaa.org>

See Also

Other data sets: mvb_teams, ncaa_sports, ncaa_teams, wvb_teams

Examples

```
head(ncaa_conferences)
```

---

ncaa_sports   *NCAA Sports and Sport Codes*

Description

This data frame includes all NCAA women's and men's sports and the codes used to refer to the sports.

Usage

```
ncaa_sports
```

Format

A data frame with 100 rows and 2 columns:

code Sport code

sport Sport name

Source

https://ncaaorg.s3.amazonaws.com/championships/resources/common/NCAA_SportCodes.pdf

See Also

Other data sets: [mvb_teams](), [ncaa_conferences](), [ncaa_teams](), [wvb_teams]()

Examples

```
head(ncaa_sports)
```

---

ncaa_teams   *NCAA Team Names*

Description

This vector includes names for all NCAA volleyball teams.

Usage

```
ncaa_teams
```

Format

A character vector with 1,089 team names.

Source

<https://stats.ncaa.org>

See Also

Other data sets: mvb_teams, ncaa_conferences, ncaa_sports, wvb_teams

Examples

```
head(ncaa_teams)
```

---

player_match_stats   *Extract player statistics for a particular match*

Description

The NCAA's page for a match/contest includes a tab called "Individual Statistics". This function extracts the tables of player match statistics for both home and away teams, as well as team statistics (though these can be omitted). If a particular team is specified, only that team's statistics will be returned.

Usage

```
player_match_stats(
  contest = NULL,
  team = NULL,
  team_stats = TRUE,
  sport = "WVB"
)
```

Arguments

contest	Contest ID determined by NCAA for match. To find ID, use find_team_contests() for a team and season.
team	Name of school. Must match name used by NCAA. Find exact team name with find_team_name().
team_stats	Logical indicating whether to include (TRUE) or exclude (FALSE) team statistics. Default includes team statistics with player statistics.
sport	Three letter abbreviation for NCAA sport (must be upper case; for example "WVB" for women's volleyball and "MVB" for men's volleyball).

Value

By default, returns data frame that includes both home and away team match statistics. If team is specified, only that team's data are returned.

See Also

Other functions that extract player statistics: player_season_stats()

Examples

```
player_match_stats(contest = "6080706")
```

---

player_season_stats   *Extract player statistics from a particular team and season*

Description

The NCAA's main page for a team includes a tab called "Team Statistics". This function extracts the table of player statistics for the season, as well as team and opponent statistics (though these can be omitted).

Usage

```
player_season_stats(team_id, team_stats = TRUE)
```

Arguments

team_id   Team ID determined by NCAA for season. To find ID, use find_team_id().

team_stats   Logical indicating whether to include (TRUE) or exclude (FALSE) team statistics. Default includes team statistics with player statistics.

Value

Returns a data frame of player statistics. Note that hometown and high school were added in 2024.

See Also

Other functions that extract player statistics: player_match_stats()

Examples

```
player_season_stats(team_id = "585290")
```

---

team_match_stats *Extract team summary statistics for all matches in a particular season*

Description

The NCAA's main page for a team includes a tab called "Game By Game" and a section called "Game by Game Stats". This function extracts the team's summary statistics for each match of the season.

Usage

```
team_match_stats(team_id = NULL, opponent = FALSE, sport = "WVB")
```

Arguments

team_id	Team ID determined by NCAA for season. To find ID, use `find_team_id()`.
opponent	Logical indicating whether to include team's stats (FALSE) or opponent's stats (TRUE). Default is set to FALSE, returning team stats.
sport	Three letter abbreviation for NCAA sport (must be upper case; for example "WVB" for women's volleyball and "MVB" for men's volleyball).

Value

Returns a data frame of summary team statistics for each match of the season.

See Also

Other functions that extract team statistics: `team_season_info()`, `team_season_stats()`

Examples

```
team_match_stats(team_id = "585290")
```

---

team_season_info *Extract arena, coach, record, and schedule information for a particular team and season*

Description

The NCAA's main page for a team includes a tab called "Schedule/Results". This function extracts information about the team's venue, coach, and records, as well as the table of the schedule and results. This returns a list, so you can subset specific components with $ (e.g., for coach information from an object called `output`, use `output$coach`).

Usage

```
team_season_info(team_id = NULL)
```

Arguments

team_id   Team ID determined by NCAA for season. To find ID, use find_team_id().

Value

Returns a list that includes arena, coach, schedule, and record information.

See Also

Other functions that extract team statistics: team_match_stats(), team_season_stats()

Examples

```
team_season_info(team_id = "585290")
```

---

team_season_stats   *Extract teams statistics for season statistics from 2001-2024*

Description

The NCAA's main page for a team includes a tab called "Game By Game" and a section called "Career Totals". Though the page only shows one season's worth of information, this function extracts season summary stats starting with 2001. We have included the conference starting with 2020 (conference data for previous seasons is not currently available).

Usage

```
team_season_stats(team = NULL, opponent = FALSE, sport = "WVB")
```

Arguments

team   Name of school. Must match name used by NCAA. Find exact team name with
                 find_team_name().

opponent   Logical indicating whether to include team's stats (FALSE) or opponent's stats
                 (TRUE). Default is set to FALSE, returning team stats.

sport   Three letter abbreviation for NCAA sport (must be upper case; for example
                 "WVB" for women's volleyball and "MVB" for men's volleyball).

Value

Returns a data frame of summary team statistics for each season.

See Also

Other functions that extract team statistics: [team_match_stats](), [team_season_info]()

Examples

```
team_season_stats(team = "Nebraska")
```

---

wvb_teams   *NCAA Women's Volleyball Teams 2020-2024*

Description

This data frame includes all women's NCAA Division 1, 2, and 3 teams from 2020-2024.

Usage

```
wvb_teams
```

Format

A data frame with 5,289 rows and 6 columns:

team_id Team ID for season/year

team_name Team name

conference_id Conference ID

conference Conference name

div NCAA division number (1, 2, or 3)

yr Year for fall of season

Source

<https://stats.ncaa.org>

See Also

Other data sets: [mvb_teams](), [ncaa_conferences](), [ncaa_sports](), [ncaa_teams]()

Examples

```
head(wvb_teams)
```

Index