

Package ‘sportsR’

August 26, 2026

Type Package

Title A Comprehensive Collection of Sports and Athletics Datasets

Version 0.1.0

Maintainer Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

Description Offers a rich and diverse collection of datasets focused on sports, athletics, physical performance, and related disciplines. The package includes professional and amateur sports data covering team sports such as soccer, basketball, baseball, American football, volleyball, rugby, cricket, hockey, and handball, as well as individual sports including tennis, badminton, table tennis, golf, swimming, cycling, athletics, gymnastics, wrestling, boxing, martial arts, weightlifting, triathlon, rowing, canoeing, climbing, surfing, skiing, snowboarding, and motorsports. Datasets cover player and team performance, match statistics, tournament results, championship standings, Olympic and international competitions, rankings, player demographics, coaching and training, biomechanics, sports medicine, injuries, exercise physiology, fitness assessment, sports nutrition, wearable sensor measurements, talent identification, and sports analytics. Additional datasets include historical competitions, referee decisions, fan engagement, economic indicators, and sports management data obtained from public repositories, official organizations, research publications, and educational resources. Designed for sports scientists, coaches, analysts, researchers, educators, students, and data scientists, this package facilitates exploratory data analysis, statistical modeling, machine learning, visualization, and sports analytics research.

License GPL-3

Language en

URL <https://github.com/lightbluetitan/sportsr>,
<https://lightbluetitan.github.io/sportsr/>

BugReports <https://github.com/lightbluetitan/sportsr/issues>

Encoding UTF-8

LazyData true

Suggests ggplot2, testthat (>= 3.0.0), dplyr, knitr, rmarkdown
Depends R (>= 4.1.0)
Imports utils
Config/roxygen2/version 8.0.0
Config/testthat/edition 3
VignetteBuilder knitr
NeedsCompilation no
Author Renzo Caceres Rossi [aut, cre] (ORCID:
 <<https://orcid.org/0009-0005-0744-854X>>)
Repository CRAN
Date/Publication 2026-08-26 20:20:02 UTC

Contents

atp_matches_2019	2
english_football	4
italian_football	5
mlb_teams_2019	6
mlb_teams_2024	7
pga_results	8
sportsR	9
view_datasets_sportsR	10
warriors_2016	10
warriors_2019	12
wta_matches_2019	13
Index	15

atp_matches_2019	<i>ATP Matches in 2019</i>
------------------	----------------------------

Description

This dataset, atp_matches_2019, is a data frame containing match-level data for men’s professional tennis matches played on the ATP Tour during 2019. It includes information on tournament details, court and surface conditions, player rankings and points, set-by-set scores, and betting odds from multiple bookmakers for each match.

Usage

data(atp_matches_2019)

Format

A data frame with 2610 observations and 36 variables:

ATP Integer vector indicating the ATP tournament identification number

Location Character vector indicating the city where the tournament was played

Tournament Character vector indicating the name of the tournament

Date Character vector indicating the date the match was played

Series Character vector indicating the ATP series or category of the tournament

Court Character vector indicating whether the match was played indoors or outdoors

Surface Character vector indicating the court surface (e.g., Hard, Clay, Grass)

Round Character vector indicating the round of the tournament

Best.of Integer vector indicating the maximum number of sets played (3 or 5)

Winner Character vector indicating the name of the match winner

Loser Character vector indicating the name of the match loser

WRank Character vector indicating the ATP ranking of the winner

LRank Character vector indicating the ATP ranking of the loser

WPts Character vector indicating the ATP ranking points of the winner

LPts Character vector indicating the ATP ranking points of the loser

W1 Integer vector indicating the games won by the winner in set 1

L1 Integer vector indicating the games won by the loser in set 1

W2 Integer vector indicating the games won by the winner in set 2

L2 Integer vector indicating the games won by the loser in set 2

W3 Integer vector indicating the games won by the winner in set 3

L3 Integer vector indicating the games won by the loser in set 3

W4 Integer vector indicating the games won by the winner in set 4

L4 Integer vector indicating the games won by the loser in set 4

W5 Integer vector indicating the games won by the winner in set 5

L5 Integer vector indicating the games won by the loser in set 5

Wsets Integer vector indicating the total number of sets won by the winner

Lsets Integer vector indicating the total number of sets won by the loser

Comment Character vector indicating the match outcome status (e.g., Completed, Retired, Walkover)

B365W Numeric vector indicating the Bet365 odds for the winner

B365L Numeric vector indicating the Bet365 odds for the loser

PSW Numeric vector indicating the Pinnacle Sports odds for the winner

PSL Numeric vector indicating the Pinnacle Sports odds for the loser

MaxW Numeric vector indicating the maximum odds offered by any bookmaker for the winner

MaxL Numeric vector indicating the maximum odds offered by any bookmaker for the loser

AvgW Numeric vector indicating the average odds offered across bookmakers for the winner

AvgL Numeric vector indicating the average odds offered across bookmakers for the loser

Details

The dataset name has been kept as 'atp_matches_2019' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the welo package version 0.1.4

english_football	<i>English Football League Results 1888-2022</i>
------------------	--

Description

This dataset, english_football, is a data frame containing results for English soccer games in the top 4 tiers from the 1888/89 season to the 2021/22 season. It includes information on match dates, seasons, home and visiting teams, full-time scores, goals scored, division, tier, and match outcomes.

Usage

```
data(english_football)
```

Format

A data frame with 203956 observations and 12 variables:

Date Character vector indicating the date of the match

Season Numeric vector indicating the season

home Character vector indicating the home team

visitor Character vector indicating the visiting team

FT Character vector indicating the full-time score

hgoal Integer vector indicating the number of goals scored by the home team

vgoal Integer vector indicating the number of goals scored by the visiting team

division Character vector indicating the division

tier Numeric vector indicating the tier

totgoal Integer vector indicating the total number of goals scored in the match

goaldif Integer vector indicating the goal difference

result Character vector indicating the match result

Details

The dataset name has been kept as 'english_football' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the footBayes package version 2.0.0

italian_football	<i>Italian Football League Results 1934-2022</i>
------------------	--

Description

This dataset, `italian_football`, is a data frame containing results for Italian soccer games in the top tier from the 1934/35 season to the 2021/22 season. It includes information on match dates, seasons, home and visiting teams, full-time scores, and goals scored.

Usage

```
data(italian_football)
```

Format

A data frame with 27684 observations and 8 variables:

Date Date vector indicating the date of the match

Season Numeric vector indicating the season

home Character vector indicating the home team

visitor Character vector indicating the visiting team

FT Character vector indicating the full-time score

hgoal Integer vector indicating the number of goals scored by the home team

vgoal Integer vector indicating the number of goals scored by the visiting team

tier Numeric vector indicating the tier

Details

The dataset name has been kept as 'italian_football' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the footBayes package version 2.0.0

mlb_teams_2019

*Baseball Team Statistics (2019)***Description**

This dataset, `mlb_teams_2019`, is a data frame containing season-level team statistics for Major League Baseball teams during the 2019 season. It includes information on league affiliation, wins, and offensive statistics such as runs, hits, home runs, RBI, stolen bases, walks, strikeouts, and batting average.

Usage

```
data(mlb_teams_2019)
```

Format

A data frame with 30 observations and 14 variables:

Team Factor w/ 30 levels indicating the name of the MLB team

League Factor w/ 2 levels indicating the league the team belongs to (American or National)

Wins Integer vector indicating the number of games won by the team

Runs Integer vector indicating the total number of runs scored by the team

Hits Integer vector indicating the total number of hits by the team

Doubles Integer vector indicating the total number of doubles hit by the team

Triples Integer vector indicating the total number of triples hit by the team

HomeRuns Integer vector indicating the total number of home runs hit by the team

RBI Integer vector indicating the total number of runs batted in by the team

StolenBases Integer vector indicating the total number of stolen bases by the team

CaughtStealing Integer vector indicating the total number of times the team was caught stealing

Walks Integer vector indicating the total number of walks drawn by the team

Strikeouts Integer vector indicating the total number of strikeouts by the team

BattingAvg Numeric vector indicating the team's overall batting average

Details

The dataset name has been kept as 'mlb_teams_2019' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the `sportsR` package and assists users in identifying its specific characteristics.

Source

Data taken from the `Lock5Data` package version 4.0.1

mlb_teams_2024	<i>Baseball Team Statistics (2024)</i>
----------------	--

Description

This dataset, `mlb_teams_2024`, is a data frame containing season-level team statistics for Major League Baseball teams during the 2024 season. It includes information on league affiliation, wins, and offensive statistics such as runs, hits, home runs, RBI, stolen bases, walks, strikeouts, and batting average.

Usage

```
data(mlb_teams_2024)
```

Format

A data frame with 30 observations and 14 variables:

- Team** Character vector indicating the name of the MLB team
- League** Character vector indicating the league the team belongs to (American or National)
- Wins** Integer vector indicating the number of games won by the team
- Runs** Integer vector indicating the total number of runs scored by the team
- Hits** Integer vector indicating the total number of hits by the team
- Doubles** Integer vector indicating the total number of doubles hit by the team
- Triples** Integer vector indicating the total number of triples hit by the team
- HomeRuns** Integer vector indicating the total number of home runs hit by the team
- RBI** Integer vector indicating the total number of runs batted in by the team
- StolenBases** Integer vector indicating the total number of stolen bases by the team
- CaughtStealing** Integer vector indicating the total number of times the team was caught stealing
- Walks** Integer vector indicating the total number of walks drawn by the team
- Strikeouts** Integer vector indicating the total number of strikeouts by the team
- BattingAvg** Numeric vector indicating the team’s overall batting average

Details

The dataset name has been kept as `'mlb_teams_2024'` to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the `sportsR` package and assists users in identifying its specific characteristics.

Source

Data taken from the `Lock5Data` package version 4.0.1

 pga_results

PGA Tournament Data

Description

This dataset, `pga_results`, is a data frame containing player-level results and performance statistics from PGA Tour tournaments. It includes information on tournament and player identifiers, scoring, fantasy points (DraftKings, FanDuel, and SuperDraft), cut status, finishing position, tournament details such as course, date, purse and season, and strokes gained statistics across different aspects of the game.

Usage

```
data(pga_results)
```

Format

A data frame with 3676 observations and 34 variables:

Player_initial_last Character vector indicating the player's name in initial-last format

tournament.id Integer vector indicating the tournament identification number

player.id Integer vector indicating the player identification number

hole_par Integer vector indicating the par for the hole

strokes Integer vector indicating the number of strokes taken

hole_DKP Numeric vector indicating the DraftKings points earned per hole

hole_FDP Numeric vector indicating the FanDuel points earned per hole

hole_SDP Integer vector indicating the SuperDraft points earned per hole

streak_DKP Integer vector indicating the DraftKings streak bonus points

streak_FDP Numeric vector indicating the FanDuel streak bonus points

streak_SDP Integer vector indicating the SuperDraft streak bonus points

n_rounds Integer vector indicating the number of rounds played

made_cut Integer vector indicating whether the player made the cut

pos Integer vector indicating the player's finishing position

finish_DKP Integer vector indicating the DraftKings points earned for finishing position

finish_FDP Integer vector indicating the FanDuel points earned for finishing position

finish_SDP Integer vector indicating the SuperDraft points earned for finishing position

total_DKP Numeric vector indicating the total DraftKings points earned

total_FDP Numeric vector indicating the total FanDuel points earned

total_SDP Integer vector indicating the total SuperDraft points earned

player Character vector indicating the full name of the player

tournament.name Character vector indicating the name of the tournament

course Character vector indicating the name of the golf course
date Character vector indicating the date of the tournament
purse Numeric vector indicating the total prize money offered at the tournament
season Integer vector indicating the season or year of the tournament
no_cut Integer vector indicating whether the tournament had no cut
Finish Character vector indicating the player's final finishing position
sg_putt Numeric vector indicating strokes gained putting
sg_arg Numeric vector indicating strokes gained around the green
sg_app Numeric vector indicating strokes gained approach
sg_ott Numeric vector indicating strokes gained off the tee
sg_t2g Numeric vector indicating strokes gained tee to green
sg_total Numeric vector indicating total strokes gained

Details

The dataset name has been kept as 'pga_results' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the ISAR package version 1.0.5

sportsR

sportsR: A Comprehensive Collection of Sports and Athletics Datasets

Description

This package provides a diverse collection of datasets focused on sports, athletics, physical performance, and related disciplines. The package includes professional and amateur sports data covering team sports such as soccer, basketball, baseball, American football, volleyball, rugby, cricket and more.

Details

sportsR: A Comprehensive Collection of Sports and Athletics Datasets

A Comprehensive Collection of Sports and Athletics Datasets.

Author(s)

Maintainer: Renzo Caceres Rossi <arenzocaceresrossi@gmail.com>

See Also

Useful links:

- <https://github.com/lightbluetitan/sportsr>

view_datasets_sportsR *View Available Datasets in sportsR*

Description

This function lists all datasets available in the 'sportsR' package. If the 'sportsR' package is not loaded, it stops and shows an error message. If no datasets are available, it returns a message and an empty vector.

Usage

```
view_datasets_sportsR()
```

Value

A character vector with the names of the available datasets. If no datasets are found, it returns an empty character vector.

Examples

```
if (requireNamespace("sportsR", quietly = TRUE)) {
  library(sportsR)
  view_datasets_sportsR()
}
```

warriors_2016 *Golden State Warriors Basketball - 2016*

Description

This dataset, warriors_2016, is a data frame containing game-by-game team statistics for the Golden State Warriors during the 2016 NBA season. It includes information on game location, opponent, win/loss outcome, points scored, and detailed shooting, rebounding, and other box score statistics for both the Warriors and their opponents.

Usage

```
data(warriors_2016)
```

Format

A data frame with 82 observations and 33 variables:

Game Integer vector indicating the game number in the season

Date Factor w/ 82 levels indicating the date the game was played

Location Factor w/ 2 levels indicating whether the game was played at home or away

Opp Factor w/ 29 levels indicating the opposing team
Win Factor w/ 2 levels indicating whether the Warriors won or lost the game
Points Integer vector indicating the points scored by the Warriors
OppPoints Integer vector indicating the points scored by the opponent
FG Integer vector indicating the number of field goals made by the Warriors
FGA Integer vector indicating the number of field goals attempted by the Warriors
FG3 Integer vector indicating the number of three-point field goals made by the Warriors
FG3A Integer vector indicating the number of three-point field goals attempted by the Warriors
FT Integer vector indicating the number of free throws made by the Warriors
FTA Integer vector indicating the number of free throws attempted by the Warriors
Rebounds Integer vector indicating the total rebounds by the Warriors
OffReb Integer vector indicating the offensive rebounds by the Warriors
Assists Integer vector indicating the assists by the Warriors
Steals Integer vector indicating the steals by the Warriors
Blocks Integer vector indicating the blocks by the Warriors
Turnovers Integer vector indicating the turnovers committed by the Warriors
Fouls Integer vector indicating the personal fouls committed by the Warriors
OppFG Integer vector indicating the number of field goals made by the opponent
OppFGA Integer vector indicating the number of field goals attempted by the opponent
OppFG3 Integer vector indicating the number of three-point field goals made by the opponent
OppFG3A Integer vector indicating the number of three-point field goals attempted by the opponent
OppFT Integer vector indicating the number of free throws made by the opponent
OppFTA Integer vector indicating the number of free throws attempted by the opponent
OppRebounds Integer vector indicating the total rebounds by the opponent
OppOffReb Integer vector indicating the offensive rebounds by the opponent
OppAssists Integer vector indicating the assists by the opponent
OppSteals Integer vector indicating the steals by the opponent
OppBlocks Integer vector indicating the blocks by the opponent
OppTurnovers Integer vector indicating the turnovers committed by the opponent
OppFouls Integer vector indicating the personal fouls committed by the opponent

Details

The dataset name has been kept as 'warriors_2016' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the Lock5Data package version 4.0.1

warriors_2019

Golden State Warriors Basketball - 2019

Description

This dataset, `warriors_2019`, is a data frame containing game-by-game team statistics for the Golden State Warriors during the 2019 NBA season. It includes information on game location, opponent, win/loss outcome, points scored, and detailed shooting, rebounding, and other box score statistics for both the Warriors and their opponents.

Usage

```
data(warriors_2019)
```

Format

A data frame with 82 observations and 33 variables:

Game Integer vector indicating the game number in the season

Date Factor w/ 82 levels indicating the date the game was played

Location Factor w/ 2 levels indicating whether the game was played at home or away

Opp Factor w/ 29 levels indicating the opposing team

Win Factor w/ 2 levels indicating whether the Warriors won or lost the game

Points Integer vector indicating the points scored by the Warriors

FG Integer vector indicating the number of field goals made by the Warriors

FGA Integer vector indicating the number of field goals attempted by the Warriors

FG3 Integer vector indicating the number of three-point field goals made by the Warriors

FG3A Integer vector indicating the number of three-point field goals attempted by the Warriors

FT Integer vector indicating the number of free throws made by the Warriors

FTA Integer vector indicating the number of free throws attempted by the Warriors

Rebounds Integer vector indicating the total rebounds by the Warriors

OffReb Integer vector indicating the offensive rebounds by the Warriors

Assists Integer vector indicating the assists by the Warriors

Steals Integer vector indicating the steals by the Warriors

Blocks Integer vector indicating the blocks by the Warriors

Turnovers Integer vector indicating the turnovers committed by the Warriors

Fouls Integer vector indicating the personal fouls committed by the Warriors

OppPoints Integer vector indicating the points scored by the opponent

OppFG Integer vector indicating the number of field goals made by the opponent

OppFGA Integer vector indicating the number of field goals attempted by the opponent

OppFG3 Integer vector indicating the number of three-point field goals made by the opponent

OppFG3A Integer vector indicating the number of three-point field goals attempted by the opponent

OppFT Integer vector indicating the number of free throws made by the opponent

OppFTA Integer vector indicating the number of free throws attempted by the opponent

OppRebounds Integer vector indicating the total rebounds by the opponent

OppOffReb Integer vector indicating the offensive rebounds by the opponent

OppAssists Integer vector indicating the assists by the opponent

OppSteals Integer vector indicating the steals by the opponent

OppBlocks Integer vector indicating the blocks by the opponent

OppTurnovers Integer vector indicating the turnovers committed by the opponent

OppFouls Integer vector indicating the personal fouls committed by the opponent

Details

The dataset name has been kept as 'warriors_2019' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the Lock5Data package version 4.0.1

wta_matches_2019	<i>WTA Matches in 2019</i>
------------------	----------------------------

Description

This dataset, wta_matches_2019, is a data frame containing match-level data for women's professional tennis matches played on the WTA Tour during 2019. It includes information on tournament details, court and surface conditions, player rankings and points, set-by-set scores, and betting odds from multiple bookmakers for each match.

Usage

```
data(wta_matches_2019)
```

Format

A data frame with 2472 observations and 32 variables:

WTA Integer vector indicating the WTA tournament identification number

Location Character vector indicating the city where the tournament was played

Tournament Character vector indicating the name of the tournament

Date Character vector indicating the date the match was played

Tier Character vector indicating the WTA tier or category of the tournament

Court Character vector indicating whether the match was played indoors or outdoors

Surface Character vector indicating the court surface (e.g., Hard, Clay, Grass)

Round Character vector indicating the round of the tournament

Best.of Integer vector indicating the maximum number of sets played

Winner Character vector indicating the name of the match winner

Loser Character vector indicating the name of the match loser

WRank Character vector indicating the WTA ranking of the winner

LRank Character vector indicating the WTA ranking of the loser

WPts Character vector indicating the WTA ranking points of the winner

LPts Character vector indicating the WTA ranking points of the loser

W1 Integer vector indicating the games won by the winner in set 1

L1 Integer vector indicating the games won by the loser in set 1

W2 Integer vector indicating the games won by the winner in set 2

L2 Integer vector indicating the games won by the loser in set 2

W3 Integer vector indicating the games won by the winner in set 3

L3 Integer vector indicating the games won by the loser in set 3

Wsets Integer vector indicating the total number of sets won by the winner

Lsets Integer vector indicating the total number of sets won by the loser

Comment Character vector indicating the match outcome status (e.g., Completed, Retired, Walkover)

B365W Numeric vector indicating the Bet365 odds for the winner

B365L Numeric vector indicating the Bet365 odds for the loser

PSW Numeric vector indicating the Pinnacle Sports odds for the winner

PSL Numeric vector indicating the Pinnacle Sports odds for the loser

MaxW Numeric vector indicating the maximum odds offered by any bookmaker for the winner

MaxL Numeric vector indicating the maximum odds offered by any bookmaker for the loser

AvgW Numeric vector indicating the average odds offered across bookmakers for the winner

AvgL Numeric vector indicating the average odds offered across bookmakers for the loser

Details

The dataset name has been kept as 'wta_matches_2019' to avoid confusion with other datasets in the R ecosystem. This naming convention helps distinguish this dataset as part of the sportsR package and assists users in identifying its specific characteristics.

Source

Data taken from the welo package version 0.1.4

Index

atp_matches_2019, [2](#)

english_football, [4](#)

italian_football, [5](#)

mlb_teams_2019, [6](#)

mlb_teams_2024, [7](#)

pga_results, [8](#)

sportsR, [9](#)

sportsR-package (sportsR), [9](#)

view_datasets_sportsR, [10](#)

warriors_2016, [10](#)

warriors_2019, [12](#)

wta_matches_2019, [13](#)