

Package ‘olympicAthletes’

July 31, 2026

Type Package

Title Olympic Athlete Event Data, Athens 1896 to Milano-Cortina 2026

Version 0.5.10

Description A tidy, long-format dataset of every athlete-event participation in the modern Olympic Games, spanning Athens 1896 through Milano-Cortina 2026 (about 315,000 rows). Extends the rgriff23 'Olympic_history' dataset (1896-2016) with five additional editions scraped from Olympedia (<<https://www.olympedia.org>>): PyeongChang 2018, Tokyo 2020, Beijing 2022, Paris 2024, and Milano-Cortina 2026. Companion datasets cover edition-level metadata and verified medal tables for every edition from 1896 to 2026.

License CC BY 4.0

Language en-US

Encoding UTF-8

LazyData true

LazyDataCompression xz

Depends R (>= 4.1)

Suggests dplyr, knitr, pkgdown, rmarkdown, testthat (>= 3.0.0), tidyr

VignetteBuilder knitr

Config/testthat/edition 3

RoxygenNote 7.3.3

URL <https://github.com/moderndive/olympicAthletes>,
<https://moderndive.github.io/olympicAthletes/>

BugReports <https://github.com/moderndive/olympicAthletes/issues>

NeedsCompilation no

Author Chester Ismay [aut, cre]

Maintainer Chester Ismay <chester.ismay@gmail.com>

Repository CRAN

Date/Publication 2026-07-31 14:00:02 UTC

Contents

art_competitions_athletes	2
athletics_athletes	3
basketball_athletes	3
curling_athletes	4
editions	4
gymnastics_athletes	5
medal_table	6
olympic_athletes	7
olympic_athletes_2024	8
paris_2024_top_medals	9
season_counts	10
team_sport_athletes	10
usa_summer_medals	11
volleyball_athletes	12
Index	13

art_competitions_athletes
<i>Olympic Art Competitions entries, 1912-1948</i>

Description

The subset of [olympic_athletes](#) with sport == "Art Competitions": the painters, sculptors, writers, musicians, and architects who won Olympic medals when the arts were medal events. A convenience dataset for exploring how these entrants differ from athletic competitors (notably in age). Columns are identical to [olympic_athletes](#).

Usage

`art_competitions_athletes`

Format

A data frame with 3,578 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(art_competitions_athletes)
head(art_competitions_athletes)
```

athletics_athletes	<i>Olympic athletics (track and field) athlete-event participations, 1896-2026</i>
--------------------	--

Description

The subset of [olympic_athletes](#) with `sport == "Athletics"`. A convenience dataset so examples can focus on a single sport without first filtering. Columns are identical to `olympic_athletes`.

Usage

```
athletics_athletes
```

Format

A data frame with 43,062 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(athletics_athletes)
head(athletics_athletes)
```

basketball_athletes	<i>Olympic basketball athlete-event participations, 1896-2026</i>
---------------------	---

Description

The subset of [olympic_athletes](#) with `sport == "Basketball"`. A convenience dataset so examples can focus on a single sport without first filtering. Columns are identical to `olympic_athletes`.

Usage

```
basketball_athletes
```

Format

A data frame with 5,159 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(basketball_athletes)
head(basketball_athletes)
```

curling_athletes	<i>Olympic curling athlete-event participations, 1896-2026</i>
------------------	--

Description

The subset of [olympic_athletes](#) with sport == "Curling". A convenience dataset so examples can focus on a single sport without first filtering. Columns are identical to [olympic_athletes](#).

Usage

```
curling_athletes
```

Format

A data frame with 961 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(curling_athletes)
head(curling_athletes)
```

editions	<i>Edition-level metadata for every Olympic Games, 1896-2026</i>
----------	--

Description

One row per Olympic edition (including cancelled Games) with verified totals for participants, NOCs, medal events, and disciplines, plus opening/closing ceremony dates.

Usage

```
editions
```

Format

A data frame with 62 rows and 15 variables:

edition_id Integer. Olympedia's internal edition number.

games Character. e.g. "2024 Summer".

year Integer.

season Character. "Summer" or "Winter".

city_local_latin Character. Host city in the local-language form rendered in the Latin alphabet, i.e. native Latin-script names retain their diacritics ("Athína", "München", "Montréal", "Ciudad de México"), and non-Latin-script names use a standard romanization rather than the city's native script ("Moskva", "Tokyo", "Beijing", "PyeongChang"). The column is therefore *always* ASCII-or-Latin-1 and never contains the city's native non-Latin alphabet.

city_english Character. Common English name of the host city (e.g. "Athens", "Munich", "Moscow", "Mexico City").

country Character.

opening_ceremony, closing_ceremony Date or NA. ISO-8601. NA for cancelled editions and a few early Games where Olympedia does not record exact dates.

participants Integer or NA. Olympedia's headline participant count. NA for cancelled editions.

nocs Integer or NA. Number of NOCs represented.

medal_events Integer or NA. Number of medal events.

disciplines Integer or NA. Number of distinct disciplines.

notes Character. Free-text edition notes (e.g. cancellation reasons, intercalated-Games status).

source Character. Olympedia URL.

Source

<https://www.olympedia.org/>

Examples

```
data(editions)
editions[, c("games", "city_local_latin", "city_english",
            "participants", "medal_events")]
```

gymnastics_athletes	<i>Olympic gymnastics athlete-event participations, 1896-2026</i>
---------------------	---

Description

The subset of `olympic_athletes` with `sport == "Gymnastics"`. A convenience dataset so examples can focus on a single sport without first filtering. Columns are identical to `olympic_athletes`.

Usage

gymnastics_athletes

Format

A data frame with 29,413 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(gymnastics_athletes)
head(gymnastics_athletes)
```

medal_table	<i>Verified medal tables for every Olympic edition, 1896-2026</i>
-------------	---

Description

One row per (Games, NOC) with verified Gold/Silver/Bronze counts cross- checked against olympedia.org and Wikipedia/IOC. Use this rather than aggregating `olympic_athletes` when you want the per-team-event medal convention (one medal per gold-winning ice hockey team, not 25).

Usage

medal_table

Format

A data frame with 1,929 rows and 11 variables:

- edition_id** Integer. Olympedia’s internal edition number.
- games** Character. e.g. "2026 Winter".
- year** Integer.
- season** Character. "Summer" or "Winter".
- noc** Character. Three-letter IOC code.
- country** Character. Full country/team name.
- gold, silver, bronze** Integer. Medal counts.
- total** Integer. Sum of the three.
- notes** Character. Free-text notes (e.g. shared medals, post-hoc reallocations).

Source

<https://www.olympedia.org/> (manually verified per edition).

Examples

```
data(medal_table)
subset(medal_table, year == 2024)[1:10, ]
```

olympic_athletes	<i>Olympic athlete-event participations, 1896-2026</i>
------------------	--

Description

One row per (athlete, Games, event) participation across every modern Summer and Winter Olympic Games from Athens 1896 through Milano-Cortina 2026. Rows for 1896-2016 come from the rgriff23 *Olympic_history* dataset (originally scraped from sports-reference.com); rows for 2018-2026 were scraped from [olympedia.org](https://www.olympedia.org/).

Usage

```
olympic_athletes
```

Format

A data frame with 315,094 rows and 16 variables:

id Integer. Unique athlete ID. Original IDs are 1..135571; athletes new to the 2018-2026 editions get IDs starting at 1,000,000 to avoid collision.

name Character. ASCII-folded athlete name.

sex Character. "M" or "F".

age Integer or NA. Approximate age at the start of the Games, computed from date of birth (Feb 1 reference for Winter, Jul 15 for Summer).

height Numeric or NA. Height in centimetres.

weight Numeric or NA. Weight in kilograms.

team Character. Country/team name as displayed on the source.

noc Character. Three-letter IOC code (e.g. "USA", "NOR", "ROC").

games Character. "<year> <Summer/Winter>", e.g. "2026 Winter".

year Integer. Calendar year of the Games. Tokyo 2020 is coded as 2020 even though it was held in 2021 (IOC convention).

season Character. "Summer" or "Winter".

city_local_latn Character. Host city in the local-language form rendered in the Latin alphabet, i.e. native Latin-script names retain their diacritics ("Athína", "München", "Montréal", "Ciudad de México"), and non-Latin-script names use a standard romanization rather than the city's native script ("Moskva", "Tokyo", "Beijing", "PyeongChang"). The column is therefore *always* ASCII-or-Latin-1 and never contains the city's native non-Latin alphabet.

city_english Character. Common English name of the host city (e.g. "Athens", "Munich", "Moscow", "Mexico City").

sport Character. Sport name. The 2018-2026 rows are normalised to the original sport names where possible (e.g. "Cycling Track" and "Cycling Road" both collapse to "Cycling").

event Character. Event name including the sport prefix.

medal Character or NA. "Gold", "Silver", "Bronze", or NA. **Per-player:** every member of a medal-winning team roster gets a row with the team's medal.

Source

- 1896-2016: Griffin, R. (2018). Olympic_history. https://github.com/rgriff23/Olympic_history
- 2018-2026: <https://www.olympedia.org/> (scraped April 2026).

Examples

```
data(olympic_athletes)
head(olympic_athletes)
table(olympic_athletes$year, olympic_athletes$season)

# Top medal-winning NOCs at Paris 2024 (one row per athlete-medal)
library(dplyr)
olympic_athletes |>
  filter(year == 2024, !is.na(medal)) |>
  count(noc, medal) |>
  tidyr::pivot_wider(names_from = medal, values_from = n, values_fill = 0) |>
  arrange(desc(Gold + Silver + Bronze))
```

olympic_athletes_2024 *Olympic athlete-event participations at the Paris 2024 Summer Games*

Description

The subset of `olympic_athletes` from the Paris 2024 Summer Games (`year == 2024 & season == "Summer"`). A single-Games convenience dataset: small enough that a height vs weight scatterplot shows over-plotting without the multi-hundred-thousand-row scale of the full data, and no `filter()` is required. Note that bio data (height/weight) is missing for many recent athletes, so roughly a third of rows have complete height and weight. Columns are identical to `olympic_athletes`.

Usage

```
olympic_athletes_2024
```

Format

A data frame with 13,660 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

See Also

For the same data a generation earlier (Rome 1960), filter the full data: `olympic_athletes |> filter(year == 1960, season == "Summer")`.

Examples

```
data(olympic_athletes_2024)
head(olympic_athletes_2024)
```

paris_2024_top_medals *Top-10 countries by medals at the Paris 2024 Summer Olympics*

Description

A small, tidy summary of the ten countries that won the most total medals at the Paris 2024 Summer Games, in long format with one row per country-and-medal-type. Built for barplot examples: ten ordered categories (rather than the ~90 medal-winning nations in [medal_table](#)) plus a medal fill variable, so stacked, dodged, and proportional barplots all read cleanly without any filtering.

Usage

```
paris_2024_top_medals
```

Format

A data frame with 30 rows (10 countries x 3 medal types) and 4 variables:

country Factor. Country/NOC name, levels ordered from the most to the fewest total medals (United States first).

noc Character. Three-letter IOC code (e.g. "USA", "CHN").

medal Factor with levels "Gold", "Silver", "Bronze".

count Integer. Number of that medal type won by that country.

Source

Derived from [medal_table](#) (games "2024 Summer").

Examples

```
data(paris_2024_top_medals)
head(paris_2024_top_medals)
```

season_counts	<i>Athlete-event rows per Olympic season, pre-counted</i>
---------------	---

Description

A tiny pre-counted table: one row per Olympic season (Summer/Winter) with the number of athlete-event rows in [olympic_athletes](#) for that season. Added so plotting exercises can contrast `geom_bar()` (counts raw rows for you) with `geom_col()` (plots pre-counted values) without introducing base R counting code.

Usage

```
season_counts
```

Format

- A data frame with 2 rows and 2 variables:
- season** Factor. "Summer" or "Winter".
- n** Integer. Number of athlete-event rows in [olympic_athletes](#) for that season.

Source

Derived from [olympic_athletes](#) (one count per season).

Examples

```
data(season_counts)
season_counts
```

team_sport_athletes	<i>Olympic team-sport athlete-event participations (Basketball, Volleyball, Curling)</i>
---------------------	--

Description

The subset of [olympic_athletes](#) with sport in Basketball, Volleyball, or Curling: three team sports with contrasting physiques, handy for comparing distributions across groups (boxplots, violins, facets) without first filtering. Columns are identical to `olympic_athletes`.

Usage

```
team_sport_athletes
```

Format

A data frame with 10,912 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(team_sport_athletes)
head(team_sport_athletes)
```

usa_summer_medals

USA Summer Olympics medal-table rows, 1896-2024

Description

The subset of [medal_table](#) with `noc == "USA"` and `season == "Summer"`: one row per Summer Games the USA attended. A convenience dataset for time-series plots of one nation's medal haul without first filtering. Columns are identical to `medal_table`. Note the USA-boycotted 1980 Moscow Games have no row at all.

Usage

```
usa_summer_medals
```

Format

A data frame with 30 rows and 11 variables; see [medal_table](#) for the full column descriptions.

Source

See [medal_table](#).

Examples

```
data(usa_summer_medals)
head(usa_summer_medals)
```

volleyball_athletes	<i>Olympic volleyball athlete-event participations, 1896-2026</i>
---------------------	---

Description

The subset of [olympic_athletes](#) with `sport == "Volleyball"`. A convenience dataset so examples can focus on a single sport without first filtering. Columns are identical to `olympic_athletes`.

Usage

```
volleyball_athletes
```

Format

A data frame with 4,792 rows and 16 variables; see [olympic_athletes](#) for the full column descriptions.

Source

See [olympic_athletes](#).

Examples

```
data(volleyball_athletes)
head(volleyball_athletes)
```

Index

* datasets

- art_competitions_athletes, [2](#)
- athletics_athletes, [3](#)
- basketball_athletes, [3](#)
- curling_athletes, [4](#)
- editions, [4](#)
- gymnastics_athletes, [5](#)
- medal_table, [6](#)
- olympic_athletes, [7](#)
- olympic_athletes_2024, [8](#)
- paris_2024_top_medals, [9](#)
- season_counts, [10](#)
- team_sport_athletes, [10](#)
- usa_summer_medals, [11](#)
- volleyball_athletes, [12](#)

- art_competitions_athletes, [2](#)
- athletics_athletes, [3](#)

- basketball_athletes, [3](#)

- curling_athletes, [4](#)

- editions, [4](#)

- gymnastics_athletes, [5](#)

- medal_table, [6](#), [9](#), [11](#)

- olympic_athletes, [2–6](#), [7](#), [8–12](#)
- olympic_athletes_2024, [8](#)

- paris_2024_top_medals, [9](#)

- season_counts, [10](#)

- team_sport_athletes, [10](#)

- usa_summer_medals, [11](#)

- volleyball_athletes, [12](#)