Package: golfastr (via r-universe)

October 30, 2026

Type Package

Title Efficiently Access Pro Golf Data

Version 0.3.0

Description Designed for quickly grabbing PGA Tour data. Includes a large data update with shot-level PGA Tour 'ShotLink' data, tee times and course setup, strokes gained, win probability after every hole, a what's in the bag database (WITBDB) of player equipment, an ID crosswalk between 'ESPN' and the PGA Tour, and logo image URLs. Leaderboards and hole-by-hole scoring come from 'ESPN' <https://www.espn.com/golf/> for the PGA Tour, LIV Golf, LPGA, DP World Tour and Champions Tour. Completed seasons are served from hosted files for fast repeated access, and data can be saved as RDS, 'Apache Arrow' Parquet or a 'SQLite' database.

License MIT + file LICENSE

Depends R (>= 4.0)

Encoding UTF-8

RoxygenNote 7.3.3

Imports httr2, dplyr, tibble, lifecycle, cli

Suggests testthat (>= 3.0.0), arrow, DBI, ggplot2, RSQLite

Config/testthat/edition 3

Config/Needs/website pkgdown, knitr, rmarkdown, ggplot2, dplyr

URL https://golfastr.com/, https://github.com/array-carpenter/golfastr

BugReports https://github.com/array-carpenter/golfastr/issues

LazyData true

Config/pak/sysreqs libssl-dev

Repository https://array-carpenter.r-universe.dev

Date/Publication 2026-10-30 00:17:09 UTC

RemoteUrl https://github.com/array-carpenter/golfastr

RemoteRef HEAD

RemoteSha 1c39f10d1da2749332d86d21e247ca0a81aeaa82

Contents

add_win_probability . . . 3
build_season . . . 4
cache_info . . . 5
calculate_strokes_gained . . . 6
check_season . . . 7
clear_cache . . . 8
compare_players . . . 8
field_strength . . . 9
get_majors . . . 10
get_player . . . 10
get_winners . . . 11
leaderboard . . . 12
list_tournaments . . . 12
load_course_setup . . . 13
load_data . . . 14
load_event_ids . . . 14
load_from_parquet . . . 15
load_from_rds . . . 16
load_holes . . . 16
load_leaderboard . . . 18
load_logos . . . 19
load_pga_hbh . . . 20
load_pga_leaderboards . . . 21
load_pga_schedule . . . 21
load_player_ids . . . 22
load_players . . . 23
load_schedule . . . 23
load_shots . . . 24
load_strokes_gained . . . 25
load_tee_times . . . 26
load_tournament . . . 27
load_tournament_detail . . . 28
load_witb . . . 28
load_witb_switches . . . 29
made_cuts_leaders . . . 30
pga_field_descriptions . . . 31
pga_majors . . . 31
pga_score_types . . . 32
player_summary . . . 32
plot_head_to_head . . . 33
plot_leaderboard . . . 34
plot_player . . . 34
plot_scoring . . . 35
plot_wins . . . 36
save_to_parquet . . . 37
save_to_rds . . . 37

scoring_avg_leaders 38
strokes_gained 39
summarise_strokes_gained 39
top10_leaders 40
tournament_history 41
update_db 42
win_leaders 43

Index **44**

`add_win_probability` *Add Win Probability to Hole-by-Hole Data*

Description

Estimates each player's chance of winning after every hole, and the win probability added (WPA) by each hole, for PGA Tour stroke play events.

Usage

```
add_win_probability(
  holes,
  sims = 10000,
  seed = 1,
  ratings = NULL,
  cut = 65,
  rounds = 4,
  start = NULL
)
```

Arguments

`holes`	Hole-by-hole data with `tournament_id`, `player_id`, `round`, `hole`, `par` and `score`, such as the output of [load_holes()]. Several events can be passed at once.
`sims`	Number of simulations used while a cut is still to come. After the last cut the calculation is exact.
`seed`	Random seed, so results are repeatable.
`ratings`	Optional skill ratings: a named numeric vector (names are ESPN player IDs) or a data frame with `player_id` and `rating`. By default the published pre-event ratings are used, or each player's latest rating for events not yet in them.
`cut`	Cut rule ("top `cut` and ties" after round 2) assumed for an event whose cut has not happened yet. Use `NA` for events with no cut. Cuts already made are read from the data.
`rounds`	Number of regulation rounds in the event: 3 for an event shortened to 54 holes, 5 for the old 90-hole Bob Hope Classic.
`start`	Optional starting strokes as a named numeric vector (names are ESPN player IDs), as at the 2019-2024 Tour Championship.

Details

Hole files from [load_holes()] already carry these columns for completed events from 2004 on. Use this function for an event in progress or for your own hole data.

For an event in progress, players who are behind in their round (or have not teed off) stay in the field at their latest score, and each player's chance right now is returned as `attr(x, "current")`.

The model plays out the rest of the event from each hole. Each remaining hole is drawn from how the field played that hole that day, shifted for the player's skill rating (strokes per round better than the field, from every earlier round since 2004) and for how the player has done so far this week and this round. Cuts are applied and ties for first are settled as a playoff (split evenly). Once no cut is left the answer is exact; before that it is simulated.

Fit on 2004-2019 and checked on 2020-2026, the probabilities are calibrated (for example, players given 10-20 those given 70-90 is 3.31 in rounds 1-2 and 1.62 in rounds 3-4, against 3.79 and 1.79 for a logistic model on strokes behind the leader and holes left.

Holes are assumed to be played 1 to 18. Players who start on the 10th tee are therefore placed in the wrong order within a round; their totals at the end of each round are unaffected. Playoff holes (rounds after `rounds`) split the chances evenly among the players still in until the playoff is decided. Modified Stableford and team events get 'NA'.

Value

`holes` with these columns added, and the current chances in `attr(, "current")`:

- `win_prob`: Chance the player wins, after this hole
- `win_prob_before`: Chance before this hole
- `wpa`: `win_prob` - `win_prob_before`
- `skill_rating`: Pre-event rating, strokes per round better than an average tour player

Examples

```
holes <- load_holes(2025, "Sony")
holes <- add_win_probability(holes, sims = 2000)
```

`build_season` *Build Season Data File*

Description

Incrementally load all tournaments for a season into a local file. Skips tournaments already saved. Safe to interrupt and resume.

Usage

```
build_season(year, file_path, tour = "pga")
```

Arguments

year	Season year
file_path	Path to data file (.rds or .parquet). Must be specified by user.
tour	Tour name (default: "pga")

Value

Invisibly returns number of tournaments added (integer).

Examples

```
## Not run:
# Build 2025 season (run multiple times if needed)
build_season(2025, file_path = tempfile(fileext = ".rds"))

## End(Not run)
```

---

cache_info *Display Cache Information*

Description

Shows information about the current cache including location, number of files, and total size.

Usage

```
cache_info()
```

Value

Invisible NULL. Called for side effects (prints to console).

Examples

```
cache_info()
```

`calculate_strokes_gained` *Calculate Strokes Gained for Every Shot*

Description

Scores each shot from [load_shots()] against a baseline of how many strokes a PGA Tour player needs to hole out from a given lie and distance. A shot's strokes gained is the expected strokes before it, minus the expected strokes after it, minus the strokes it cost (one, plus any penalty strokes before the next swing).

Usage

```
calculate_strokes_gained(shots, baseline = NULL)
```

Arguments

`shots`	A tibble from [load_shots()]. Holes without a holed final stroke are dropped.
`baseline`	Optional data frame with columns `lie`, `distance` (yards) and `expected` (strokes to hole out) to use in place of the bundled PGA Tour baseline, for example one fit to a different tour.

Details

The baseline is fit from ShotLink strokes and ships with the package. Over the 2025 and 2026 seasons, round totals built from it and adjusted with [summarise_strokes_gained()] track the PGA Tour's official strokes gained total closely (correlation 0.97 to 0.98, mean difference about 0.1 strokes per round). The split between off-the-tee and approach is less exact (correlation 0.82 to 0.92). Native area and trees count as rough, and "around the green" is defined as within 30 yards of the hole rather than of the green's edge.

Value

A tibble with one row per stroke hit (penalty and drop rows are folded into the shot that caused them), with the columns of 'shots' plus:

- `start_lie`: tee, fairway, rough, sand, recovery (water or unknown) or green
- `start_distance`: Yards to the hole before the shot
- `strokes_used`: Strokes the shot cost, including penalties
- `expected_before`, `expected_after`: Baseline strokes to hole out before and after the shot
- `sg`: Strokes gained by the shot
- `sg_category`: `"off_tee"`, `"approach"`, `"around_green"` or `"putt"`

See Also

[summarise_strokes_gained()] for round, event and season totals.

Examples

```
shots <- tibble::tibble(
  tournament_id = "R1", player_id = "1", round = 1L, course_id = "C1",
  hole = 1L, par = 4L, yardage = 420L, hole_score = 4L, stroke = 1:4,
  stroke_type = "STROKE", from_code = c("OTB", "OFW", "OGR", "OGR"),
  distance_remaining = c(150, 10, 1, 0),
  final_stroke = c(FALSE, FALSE, FALSE, TRUE)
)
calculate_strokes_gained(shots)[, c("stroke", "start_lie", "sg", "sg_category")]

sg <- calculate_strokes_gained(load_shots(2026, "Sony"))
```

---

check_season *Check Season Progress*

Description

See which tournaments are loaded vs missing for a season.

Usage

```
check_season(year, file_path, tour = "pga")
```

Arguments

year	Season year
file_path	Path to data file (.rds or .parquet). Must be specified by user.
tour	Tour name (default: "pga")

Value

A tibble showing status of each tournament with columns: event_id, tournament_name, start_date, end_date, and status (either "loaded" or "missing").

Examples

```
## Not run:
progress <- check_season(2025, file_path = "my_golf_data.rds")

## End(Not run)
```

clear_cache *Clear golfastr Cache*

Description

Removes all cached data files from the local cache directory.

Usage

```
clear_cache(confirm = TRUE)
```

Arguments

confirm	Logical. If TRUE (default), prompts for confirmation before clearing. Set to FALSE to skip confirmation.

Value

Invisible NULL. Called for side effects.

Examples

```
clear_cache()
clear_cache(confirm = FALSE)
```

compare_players *Compare Players*

Description

Side-by-side comparison of multiple players.

Usage

```
compare_players(players, year = NULL, file_path = NULL, data = NULL)
```

Arguments

players	Vector of player names
year	Optional year filter
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble comparing player statistics

Examples

```
## Not run:
season <- load_leaderboard(2025)
compare_players(c("Scheffler", "McIlroy"), data = season)

## End(Not run)
```

---

field_strength *Field Strength Analysis*

Description

Analyze the strength of a tournament field.

Usage

```
field_strength(tournament, year, file_path = NULL, data = NULL)
```

Arguments

tournament	Tournament name
year	Season year
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

Tibble with field statistics

Examples

```
## Not run:
season <- load_leaderboard(2025)
field_strength("Masters", 2025, data = season)

## End(Not run)
```

get_majors *Get Major Championships*

Description

Get results from the four major championships.

Usage

```
get_majors(year, file_path = NULL, data = NULL)
```

Arguments

year	Season year
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with major championship results

Examples

```
## Not run:
season <- load_leaderboard(2025)
get_majors(2025, data = season)

## End(Not run)
```

get_player *Get Player Results*

Description

Look up a player's results across tournaments. Pass a data frame from [load_leaderboard()] or a local file created with [save_to_rds()].

Usage

```
get_player(name, file_path = NULL, data = NULL)
```

Arguments

name	Player name (partial match, case-insensitive)
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with player's tournament results

Examples

```
## Not run:
season <- load_leaderboard(2025)
get_player("McIlroy", data = season)
get_player("Scheffler", file_path = "golf_data.rds")

## End(Not run)
```

---

get_winners *Get Tournament Winners*

Description

Get all tournament winners from leaderboard data.

Usage

```
get_winners(year = NULL, file_path = NULL, data = NULL)
```

Arguments

year	Optional year filter
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with tournament winners

Examples

```
## Not run:
season <- load_leaderboard(2025)
get_winners(data = season)
get_winners(2025, file_path = "golf_data.rds")

## End(Not run)
```

leaderboard *Leaderboard Snapshot*

Description

Get a compact leaderboard for display.

Usage

```
leaderboard(tournament, year, top_n = 10, file_path = NULL, data = NULL)
```

Arguments

tournament	Tournament name
year	Season year
top_n	Number of players to show (default: 10)
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

Formatted tibble for display

Examples

```
## Not run:
season <- load_leaderboard(2025)
leaderboard("Masters", 2025, data = season)

## End(Not run)
```

list_tournaments *List Available Tournaments*

Description

Get a list of all available tournaments for a given year.

Usage

```
list_tournaments(year, tour = "pga")
```

Arguments

year	Season year (e.g., 2025)
tour	Tour name: "pga" (default), "liv", "lpga", "euro", or "champions".

Value

Tibble with event_id, tournament_name, start_date, end_date

Examples

```
# See all 2025 PGA tournaments
list_tournaments(2025)

# See LIV Golf schedule
list_tournaments(2026, tour = "liv")
```

---

load_course_setup *Load Course Setup by Round*

Description

Returns par, yardage and the pin position for every hole in every round of a PGA Tour season.

Usage

```
load_course_setup(year, tournament = NULL)
```

Arguments

year	Season year (e.g., 2026).
tournament	Optional PGA Tour tournament ID (e.g., '"R2025011"') or partial tournament name match.

Value

A tibble with one row per hole per round: tournament_id, tournament_name, round, course_id, hole, par, yardage, and pin_x, pin_y (pin position on the green diagram, scaled 0 to 1; NA in older seasons).

Examples

```
setup <- load_course_setup(2026, "Sony")
```

load_data *Load Tournament Data*

Description

Load tournament data from a file. Auto-detects format based on extension.

Usage

```
load_data(file_path, tournament = NULL)
```

Arguments

file_path	Path to data file (.rds or .parquet)
tournament	Optional tournament name filter (partial match)

Value

A tibble with tournament leaderboard data

Examples

```
## Not run:
# Load all data
data <- load_data("golf_data.rds")

# Load and filter to Masters
masters <- load_data("golf_data.rds", tournament = "Masters")

## End(Not run)
```

load_event_ids *Load Event IDs*

Description

Links ESPN tournament IDs to PGA Tour tournament IDs for events in both sources.

Usage

```
load_event_ids(year = NULL)
```

Arguments

year	Optional season filter.

Value

A tibble with espn_tournament_id, pga_tournament_id, year, tournament_name, pga_tournament_name, start_date and field_overlap (share of the smaller field found in both sources).

Examples

```
load_event_ids(2025)
```

---

load_from_parquet *Load Tournament Data from Parquet*

Description

Load tournament data from a Parquet file. Requires the arrow package.

Usage

```
load_from_parquet(file_path)
```

Arguments

file_path	Path to Parquet file. Must be specified by user.

Value

A tibble with tournament leaderboard data containing columns such as position, player name, scores, and tournament metadata.

Examples

```
## Not run:
data <- load_from_parquet(file_path = "my_golf_data.parquet")

## End(Not run)
```

load_from_rds *Load Tournament Data from RDS*

Description

Load Tournament Data from RDS

Usage

```
load_from_rds(file_path)
```

Arguments

file_path	Path to RDS file. Must be specified by user.

Value

A tibble with tournament leaderboard data containing columns such as position, player name, scores, and tournament metadata.

Examples

```
## Not run:
data <- load_from_rds(file_path = "my_golf_data.rds")

## End(Not run)
```

load_holes *Load Hole-by-Hole Scoring*

Description

Retrieves hole-by-hole scoring data for tournaments.

Usage

```
load_holes(
  year = as.integer(format(Sys.Date(), "%Y")),
  tournament = NULL,
  top_n = NULL,
  tour = "pga",
  live = FALSE
)

get_player_scorecards(event_id, athlete_id)
```

Arguments

year	Season year (e.g., 2026). Defaults to current year.
tournament	Tournament identifier - either event_id or partial name match.
top_n	Number of top finishers to include. Default NULL returns all.
tour	Tour identifier: ‘"pga"‘ (default), ‘"liv"‘, ‘"lpga"‘, ‘"euro"‘, or ‘"champions"‘.
live	If TRUE, skip the hosted data and fetch directly from the ESPN API. Defaults to FALSE.
event_id	ESPN event identifier (for legacy function)
athlete_id	ESPN athlete identifier (for legacy function)

Details

For PGA Tour data, completed tournaments from 2004 on are served from pre-built full-field files hosted on GitHub releases (updated daily), avoiding the one-request-per-player cost of the live ESPN API. The live API is used as a fallback for in-progress tournaments, other tours, or when the hosted data is unavailable. Hosted files also carry win probability for every hole; see [add_win_probability()] for the method.

Value

A tibble with hole-by-hole data including:

- player_id: ESPN athlete ID
- player_name: Player display name
- position: Final leaderboard standing
- round: Round number (1-4; playoff holes come as an extra round)
- hole: Hole number (1-18)
- par: Par for the hole
- score: Player’s score on the hole
- score_type: BIRDIE, PAR, BOGEY, EAGLE, etc.
- tournament_id: Event ID
- field_avg, vs_field, to_par, cume_to_par: Field average on the hole that round, strokes against it, score to par and running score to par (hosted data)
- win_prob, win_prob_before, wpa: Chance of winning after and before the hole, and the difference (hosted PGA Tour stroke play events)
- skill_rating: Pre-event rating, strokes per round better than an average tour player
- hole_order_source: "shotlink" when the player’s starting tee is known, else "assumed" (holes 1 to 18)

A tibble with hole-by-hole scoring data

Functions

- get_player_scorecards(): Deprecated, use ‘load_holes()‘ instead

Examples

```
# Load hole-by-hole for top 10 at Sony Open
holes <- load_holes(2026, "Sony", top_n = 10)
```

`load_leaderboard` *Load Tournament Leaderboard*

Description

Retrieves leaderboard data for tournaments. Can load a single tournament or all tournaments for a year.

Usage

```
load_leaderboard(
  year = as.integer(format(Sys.Date(), "%Y")),
  tournament = NULL,
  tour = "pga",
  live = FALSE
)

get_tournament_leaderboard(event_id)
```

Arguments

`year`	Season year (e.g., 2026). Defaults to current year.
`tournament`	Tournament identifier - either event_id or partial name match. If NULL, returns all tournaments for the year.
`tour`	Tour identifier: ‘"pga"‘ (default), ‘"liv"‘, ‘"lpga"‘, ‘"euro"‘, or ‘"champions"‘.
`live`	If TRUE, skip the hosted data and fetch directly from the ESPN API. Defaults to FALSE.
`event_id`	ESPN event identifier (for legacy function)

Details

For PGA Tour data, completed tournaments are served from pre-built files hosted on GitHub releases (updated daily), which is much faster than querying ESPN tournament by tournament. The live ESPN API is used as a fallback for in-progress tournaments, other tours, or when the hosted data is unavailable.

Value

A tibble with leaderboard data including:

- `position`: Final standing
- `player_id`: ESPN athlete ID
- `player_name`: Player display name
- `total_score`: Total strokes
- `score_to_par`: Score relative to par (e.g., "-11")
- `tournament_id`: Event ID
- `tournament_name`: Tournament name

A tibble with leaderboard data

Functions

- `get_tournament_leaderboard()`: Deprecated, use ‘load_leaderboard()‘ instead

Examples

```
# Load specific tournament by name
sony <- load_leaderboard(2026, "Sony")
```

---

load_logos *Load Logos*

Description

Logo image URLs for tours, tournaments, ShotLink courses and famous clubs (every host of a men’s or women’s major, the Ryder Cup, Presidents Cup or Solheim Cup since 1960). Only URLs are provided, not images; logos remain the property of their owners.

Usage

```
load_logos(type = NULL, name = NULL)
```

Arguments

`type`	Optional filter: one or more of `"tour"`, `"tournament"`, `"course"` and `"club"`.
`name`	Optional name filter (partial match, case-insensitive).

Details

Tournament logos come from the PGA Tour schedule (2012 on, joined to [load_event_ids()]) and from Wikimedia. A course takes its club’s logo, from Wikimedia or else the club’s own site. Every URL returned an image on `checked_at`.

Value

A tibble with `logo_type`, `name`, `logo_url`, `logo_source`, `source_page`, `tour`, `pga_tournament_id`, `espn_tournament_id`, `course_id`, `club_id`, `primary_color`, `secondary_color` and `checked_at`.

Examples

```
load_logos("club", "Augusta")
```

`load_pga_hbh` *Load PGA Hole-by-Hole Data*

Description

Deprecated. Use [load_holes()] instead, which serves full-field scorecards for completed tournaments from pre-built hosted files.

Usage

```
load_pga_hbh(years, tournaments = NULL, top_n = 10, tour = "pga", dir = NULL)
```

Arguments

`years`	Numeric vector. Year(s) to load.
`tournaments`	Character vector. Optional tournament event IDs or names. If NULL, loads all tournaments.
`top_n`	Numeric. Number of top finishers to include scorecards for. Default is 10. Set to NULL for all players.
`tour`	Character. Tour type, default "pga".
`dir`	Character. Optional directory to save CSV files.

Value

A tibble with hole-by-hole scoring data.

Examples

```
load_pga_hbh(2025, "Masters Tournament", top_n = 3)
```

load_pga_leaderboards *Load PGA Leaderboards*

Description

Deprecated. Use [load_leaderboard()] instead, which serves completed tournaments from pre-built hosted files rather than one live request per event.

Usage

```
load_pga_leaderboards(years, tournaments = NULL, tour = "pga", dir = NULL)
```

Arguments

years	Numeric vector. Year(s) to load (e.g., 2025 or 2023:2025).
tournaments	Character vector. Optional tournament event IDs or names to filter. If NULL, loads all tournaments.
tour	Character. Tour type, default "pga".
dir	Character. Optional directory to save CSV files.

Value

A tibble with leaderboard data.

Examples

```
load_pga_leaderboards(2025, "Masters Tournament")
```

load_pga_schedule *Load PGA Tour Schedule*

Description

Deprecated. Use [load_schedule()] instead, which supports all tours and shares its output format with the rest of the package.

Usage

```
load_pga_schedule(year, tour = "pga")
```

Arguments

year	Numeric. The year to load (e.g., 2025).
tour	Character. Tour type, default "pga".

Value

A tibble with tournament schedule data.

Examples

```
load_pga_schedule(2025)
```

`load_player_ids` *Load Player IDs*

Description

Links ESPN player IDs (used by [load_holes()], [load_leaderboard()] and [add_win_probability()]) to PGA Tour player IDs (used by [load_shots()], [load_tee_times()] and [load_strokes_gained()]).

Usage

```
load_player_ids(player = NULL)
```

Arguments

`player`	Optional player name filter (partial match, case-insensitive).

Details

Pairs come from events that appear in both sources (2012 on): a player with the same name in the same event votes for a pair, and a pair needs at least two such events. IDs that pair with more than one partner (for example two players with the same name) are left out rather than guessed.

Value

A tibble with `player_name`, `espn_player_id`, `pga_player_id`, `first_season`, `last_season` and `events_matched`.

Examples

```
load_player_ids("Scheffler")
```

load_players *Load Player Directory*

Description

Retrieves a directory of players for a tour. Since ESPN doesn't have a dedicated player directory endpoint, this aggregates players from recent tournament leaderboards.

Usage

```
load_players(year = as.integer(format(Sys.Date(), "%Y")), tour = "pga")
```

Arguments

year	Season year to pull players from. Defaults to current year.
tour	Tour identifier: ‘"pga"‘ (default), ‘"liv"‘, ‘"lpga"‘, ‘"euro"‘, or ‘"champions"‘.

Value

A tibble with player data including:

- player_id: ESPN athlete ID
- player_name: Player display name

Examples

```
# Get player directory
players <- load_players()
```

load_schedule *Load Golf Schedule*

Description

Retrieves the tournament schedule for a given year and tour.

Usage

```
load_schedule(year = as.integer(format(Sys.Date(), "%Y")), tour = "pga")

get_pga_schedule(year = format(Sys.Date(), "%Y"))
```

Arguments

year	Season year (e.g., 2026). Defaults to current year.
tour	Tour identifier: ‘"pga"‘ (default), ‘"liv"‘, ‘"lpga"‘, ‘"euro"‘, or ‘"champions"‘.

Value

A tibble with columns:

- `event_id`: ESPN event identifier
- `tournament_name`: Tournament name
- `start_date`: Tournament start date
- `end_date`: Tournament end date

A tibble with tournament schedule data

Functions

- `get_pga_schedule()`: Deprecated, use ‘load_schedule()‘ instead

Examples

```
# Get current year schedule
schedule <- load_schedule()

# Get specific year
schedule_2025 <- load_schedule(2025)
```

---

load_shots *Load Shot-Level Data*

Description

Returns every stroke from PGA Tour ShotLink for a season: where each shot started and finished, its distance, and launch monitor (radar) readings when the tour captured them.

Usage

```
load_shots(year, tournament = NULL, player = NULL)
```

Arguments

`year`	Season year (e.g., 2026).
`tournament`	Optional PGA Tour tournament ID (e.g., ‘"R2025011"‘) or partial tournament name match.
`player`	Optional player name filter (partial match, case-insensitive).

Details

Data comes from pre-built files on the package’s GitHub releases, built from the PGA Tour’s shot tracking feed. Older seasons have strokes, lies and distances but no coordinates or radar data. Events without ShotLink coverage are not included.

Value

A tibble with one row per stroke, including:

- `tournament_id`, `tournament_name`: PGA Tour event
- `player_id`, `player_name`: PGA Tour player
- `round`, `hole`, `par`, `hole_score`
- `stroke`: Stroke number on the hole
- `stroke_type`: `"STROKE"`, `"PENALTY"`, `"DROP"` and similar
- `distance`: Shot distance in yards
- `distance_remaining`: Distance to the hole after the shot, in yards
- `from_location`, `to_location`: Lie before and after (e.g., `"Tee Box"`, `"Fairway"`, `"Green"`)
- `from_code`, `to_code`: ShotLink lie codes
- `play_by_play`: Text description of the shot
- `from_x`, `from_y`, `to_x`, `to_y`: Shot position on the hole diagram, scaled 0 to 1
- `club_speed`, `ball_speed` (mph), `smash_factor`, `spin_rate` (rpm), `spin_axis`, `launch_angle`, `launch_direction` (degrees), `apex_height`, `apex_range`, `apex_side` (feet), `flight_time`: radar readings, `NA` when not captured

Examples

```
shots <- load_shots(2026, "Sony")
```

---

load_strokes_gained *Load Strokes Gained Statistics*

Description

Returns strokes gained data from the PGA Tour, with all six SG categories per player: putting, around the green, approach, off the tee, tee to green, and total.

Usage

```
load_strokes_gained(player = NULL, use_bundled = FALSE, year = NULL)
```

Arguments

`player`	Optional player name filter (partial match, case-insensitive).
`use_bundled`	If `TRUE`, skip the download and use the dataset bundled with the package. Defaults to `FALSE`.
`year`	Optional season year filter (e.g., 2026).

Details

By default, downloads the latest data from GitHub releases (updated weekly). Falls back to the bundled dataset shipped with the package if the download fails (e.g., no internet connection).

Value

A tibble with one row per player and columns:

- `player_id`: PGA Tour player ID
- `player_name`: Player display name
- `country`: Three-letter country code
- `sg_putt`: Strokes Gained: Putting (per round avg)
- `sg_arg`: Strokes Gained: Around the Green (per round avg)
- `sg_app`: Strokes Gained: Approach the Green (per round avg)
- `sg_ott`: Strokes Gained: Off the Tee (per round avg)
- `sg_t2g`: Strokes Gained: Tee to Green (per round avg)
- `sg_total`: Strokes Gained: Total (per round avg)
- `rounds`: Number of measured (ShotLink) rounds
- `season`: PGA Tour season year

Examples

```
sg <- load_strokes_gained(use_bundled = TRUE)
load_strokes_gained("Scheffler", use_bundled = TRUE, year = 2026)

load_strokes_gained()
```

`load_tee_times` *Load Tee Times and Pairings*

Description

Returns the tee time, group and starting hole for every player in every round of a PGA Tour season.

Usage

```
load_tee_times(year, tournament = NULL, player = NULL)
```

Arguments

`year`	Season year (e.g., 2026).
`tournament`	Optional PGA Tour tournament ID (e.g., ‘"R2025011"‘) or partial tournament name match.
`player`	Optional player name filter (partial match, case-insensitive).

Value

A tibble with one row per player per round: `tournament_id`, `tournament_name`, `round`, `group`, `tee_time` (UTC), `timezone` (local zone of the event), `start_tee` (1 or 10), `course_id`, `course_name`, `player_id`, `player_name`.

Examples

```
tee_times <- load_tee_times(2026, "Sony")
```

---

load_tournament *Load Tournament Data*

Description

Fetch leaderboard data for a tournament. You can specify either the event_id or search by tournament name.

Usage

```
load_tournament(year, tournament, tour = "pga")
```

Arguments

`year`	Season year (e.g., 2025)
`tournament`	Tournament name (partial match) or event_id
`tour`	Tour name: "pga" (default), "liv", "lpga", "euro", or "champions".

Value

Tibble with tournament leaderboard

Examples

```
# Load by name (partial match works)
masters <- load_tournament(2025, "Masters")
pga_champ <- load_tournament(2025, "PGA Championship")

# Load LIV Golf tournament
adelaide <- load_tournament(2026, "Adelaide", tour = "liv")
```

load_tournament_detail *Load Tournament with Hole-by-Hole Scores*

Description

Fetch full tournament data including hole-by-hole scorecards.

Usage

```
load_tournament_detail(year, tournament, top_n = 10, tour = "pga")
```

Arguments

year	Season year (e.g., 2025)
tournament	Tournament name (partial match) or event_id
top_n	Only fetch scorecards for top N finishers (default: 10)
tour	Tour name: "pga" (default), "liv", "lpga", "euro", or "champions".

Value

List with 'leaderboard' and 'scorecards' tibbles

Examples

```
# Get Masters with top 10 scorecards
masters_detail <- load_tournament_detail(2025, "Masters", top_n = 10)
```

load_witb *Load What's in the Bag*

Description

Player equipment from the what's in the bag database (WITBDB): one row per club, parsed from GolfWRX and PGA Tour Club Tracker bag posts. Each bag keeps its `source` and `source_url`.

Usage

```
load_witb(player = NULL, year = NULL, slot = NULL)
```

Arguments

player	Optional player name filter (partial match, case-insensitive).
year	Optional year filter on `as_of_date`.
slot	Optional filter: one or more of `"driver"`, `"fairway"`, `"hybrid"`, `"utility"`, `"iron"`, `"wedge"` and `"putter"`.

Details

player_id is the PGA Tour player ID used by [load_shots()]; it is NA for players without one (for example LPGA players and amateurs). Fields a post does not list are NA.

Value

A tibble with one row per club: bag columns (bag_id, player_id, player_name, as_of_date, event_name, tournament_id, source, source_url), the club (slot, club_label, club_order), head, shaft, grip and putter specs, and qc_flags. See [pga_field_descriptions()] with "witb".

Examples

```
load_witb("Scheffler", year = 2026, slot = "putter")
```

---

load_witb_switches *Load Equipment Switches*

Description

Equipment changes found by comparing each player's bags in date order (see [load_witb()]): one row per change of head, shaft or grip in a slot. Only players with a PGA Tour ID are included.

Usage

```
load_witb_switches(player = NULL, slot = NULL)
```

Arguments

player	Optional player name filter (partial match, case-insensitive).
slot	Optional slot filter, as in [load_witb()].

Value

A tibble with player_id, player_name, slot, component, old_value, new_value, last_seen_old, first_seen_new, first_seen_event, tournament_id, old_bag_id, new_bag_id, gap_days, uncertain, reason and brand_change. See [pga_field_descriptions()] with "witb_switches".

Examples

```
load_witb_switches("McIlroy", slot = "driver")
```

made_cuts_leaders *Made Cuts Percentage*

Description

Get players by percentage of cuts made.

Usage

```
made_cuts_leaders(
  year = NULL,
  min_events = 5,
  top_n = 10,
  file_path = NULL,
  data = NULL
)
```

Arguments

year	Optional year filter
min_events	Minimum events played
top_n	Number of players to show
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with made cuts leaders

Examples

```
## Not run:
season <- load_leaderboard(2025)
made_cuts_leaders(data = season)

## End(Not run)
```

pga_field_descriptions *Get Field Descriptions for PGA Data*

Description

Returns a data frame with field names and descriptions for leaderboard, hole-by-hole, logo or what's in the bag data.

Usage

```
pga_field_descriptions(
  data_type = c("leaderboard", "holes", "logos", "witb", "witb_switches")
)
```

Arguments

data_type	Character. One of "leaderboard", "holes", "logos", "witb" ([load_witb()]) or "witb_switches" ([load_witb_switches()]).

Value

A tibble with field and description columns.

Examples

```
pga_field_descriptions("leaderboard")
pga_field_descriptions("holes")
pga_field_descriptions("logos")
pga_field_descriptions("witb")
```

pga_majors *Get PGA Major Championships*

Description

Returns a data frame with the four major championships and their typical schedule.

Usage

```
pga_majors()
```

Value

A tibble with tournament, month, and course columns.

Examples

```
pga_majors()
```

`pga_score_types` *Get PGA Score Types*

Description

Returns a data frame with score type classifications used in hole-by-hole data.

Usage

```
pga_score_types()
```

Value

A tibble with score_type, strokes_vs_par, and description columns.

Examples

```
pga_score_types()
```

`player_summary` *Player Season Summary*

Description

Get aggregate statistics for a player's season.

Usage

```
player_summary(name, year = NULL, file_path = NULL, data = NULL)
```

Arguments

`name`	Player name (partial match)
`year`	Optional year filter
`file_path`	Path to data file (.rds or .parquet)
`data`	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

Tibble with season statistics

Examples

```
## Not run:
season <- load_leaderboard(2025)
player_summary("Scheffler", data = season)
player_summary("McIlroy", year = 2025, file_path = "golf_data.rds")

## End(Not run)
```

`plot_head_to_head` *Plot Head-to-Head Comparison*

Description

Compare two or more players' finishes across tournaments.

Usage

```
plot_head_to_head(players, year = NULL, file_path = NULL, data = NULL)
```

Arguments

`players`	Vector of player names
`year`	Optional year filter
`file_path`	Path to data file (.rds or .parquet)
`data`	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

ggplot object

Examples

```
## Not run:
season <- load_leaderboard(2025)
plot_head_to_head(c("Scheffler", "McIlroy"), data = season)

## End(Not run)
```

plot_leaderboard *Plot Tournament Leaderboard*

Description

Bar chart of tournament leaderboard.

Usage

```
plot_leaderboard(tournament, year, top_n = 10, file_path = NULL, data = NULL)
```

Arguments

tournament	Tournament name
year	Season year
top_n	Number of players to show
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

ggplot object

Examples

```
## Not run:
season <- load_leaderboard(2025)
plot_leaderboard("Masters", 2025, data = season)

## End(Not run)
```

plot_player *Plot Player Results*

Description

Visualize a player's finishes across tournaments.

Usage

```
plot_player(name, year = NULL, file_path = NULL, data = NULL)
```

Arguments

name	Player name
year	Optional year filter
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

ggplot object

Examples

```
## Not run:
season <- load_leaderboard(2025)
plot_player("Scheffler", data = season)

## End(Not run)
```

---

plot_scoring *Plot Scoring Distribution*

Description

Histogram of tournament scores.

Usage

```
plot_scoring(tournament, year, file_path = NULL, data = NULL)
```

Arguments

tournament	Tournament name
year	Season year
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

ggplot object

Examples

```
## Not run:
season <- load_leaderboard(2025)
plot_scoring("Masters", 2025, data = season)

## End(Not run)
```

`plot_wins` *Plot Win Distribution*

Description

Bar chart of wins by player.

Usage

```
plot_wins(year = NULL, top_n = 10, file_path = NULL, data = NULL)
```

Arguments

`year`	Optional year filter
`top_n`	Number of players to show
`file_path`	Path to data file (.rds or .parquet)
`data`	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

ggplot object

Examples

```
## Not run:
season <- load_leaderboard(2025)
plot_wins(data = season)

## End(Not run)
```

save_to_parquet *Save Tournament Data to Parquet*

Description

Save tournament data to a Parquet file for cross-language compatibility. Requires the arrow package.

Usage

```
save_to_parquet(data, file_path, append = TRUE)
```

Arguments

data	Tibble of tournament data
file_path	Path to Parquet file. Must be specified by user.
append	If TRUE, append to existing data

Value

Invisible NULL. Called for side effects (writes to file).

Examples

```
## Not run:
masters <- load_tournament(2025, "Masters")
save_to_parquet(masters, file_path = tempfile(fileext = ".parquet"))

## End(Not run)
```

save_to_rds *Save Tournament Data to RDS*

Description

Save tournament data to an RDS file.

Usage

```
save_to_rds(data, file_path, append = TRUE)
```

Arguments

data	Tibble of tournament data
file_path	Path to RDS file. Must be specified by user.
append	If TRUE, append to existing data

Value

Invisible NULL. Called for side effects (writes to file).

Examples

```
## Not run:
masters <- load_tournament(2025, "Masters")
save_to_rds(masters, file_path = tempfile(fileext = ".rds"))

## End(Not run)
```

`scoring_avg_leaders` *Scoring Average Leaders*

Description

Get players with best scoring average.

Usage

```
scoring_avg_leaders(
  year = NULL,
  min_events = 5,
  top_n = 10,
  file_path = NULL,
  data = NULL
)
```

Arguments

`year`	Optional year filter
`min_events`	Minimum events played
`top_n`	Number of players to show
`file_path`	Path to data file (.rds or .parquet)
`data`	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with scoring average leaders

Examples

```
## Not run:
season <- load_leaderboard(2025)
scoring_avg_leaders(data = season)

## End(Not run)
```

strokes_gained *PGA Tour Strokes Gained Statistics*

Description

Pre-built dataset of strokes gained statistics from the PGA Tour. Contains per-round averages for all six strokes gained categories for players in the current season. Updated with each package release.

Usage

```
strokes_gained
```

Format

A tibble with one row per player and 11 columns:

player_id PGA Tour player ID

player_name Player display name

country Three-letter country code (e.g., "USA", "ENG")

sg_putt Strokes Gained: Putting (per round average)

sg_arg Strokes Gained: Around the Green (per round average)

sg_app Strokes Gained: Approach the Green (per round average)

sg_ott Strokes Gained: Off the Tee (per round average)

sg_t2g Strokes Gained: Tee to Green (per round average)

sg_total Strokes Gained: Total (per round average)

rounds Number of ShotLink-measured rounds

season PGA Tour season year

Source

PGA Tour (https://www.pgatour.com/stats)

summarise_strokes_gained *Summarise Strokes Gained by Round, Event or Season*

Description

Adds up shot-level strokes gained from [calculate_strokes_gained()] and adjusts each category to the field: in every round, each category is measured against the average of the players on the same course, so the four categories add up to the total and a round's total equals the strokes the player beat the field average by. Only complete 18-hole rounds are used, and team events (Zurich Classic, Presidents Cup) are left out.

Usage

```
summarise_strokes_gained(sg, by = c("round", "tournament", "season"))
```

Arguments

sg	A tibble from [calculate_strokes_gained()].
by	One of ‘"round"‘ (one row per player per round), ‘"tournament"‘ or ‘"season"‘ (per-round averages for each player).

Value

A tibble with `player_id`, `player_name` (when present), grouping columns, `rounds`, and `sg_ott`, `sg_app`, `sg_arg`, `sg_putt`, `sg_t2g` and `sg_total`.

Examples

```
sg <- calculate_strokes_gained(load_shots(2026, "Sony"))
summarise_strokes_gained(sg, by = "tournament")
```

---

top10_leaders *Top 10 Leaders*

Description

Get players with most top 10 finishes.

Usage

```
top10_leaders(
  year = NULL,
  min_events = 5,
  top_n = 10,
  file_path = NULL,
  data = NULL
)
```

Arguments

year	Optional year filter
min_events	Minimum events played
top_n	Number of players to show
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with top 10 leaders

Examples

```
## Not run:
season <- load_leaderboard(2025)
top10_leaders(data = season)

## End(Not run)
```

`tournament_history` *Tournament History*

Description

Get historical results for a specific tournament.

Usage

```
tournament_history(tournament, file_path = NULL, data = NULL)
```

Arguments

`tournament`	Tournament name (partial match)
`file_path`	Path to data file (.rds or .parquet)
`data`	Optional leaderboard data frame. Use this instead of ‘file_path‘ when the data is already in memory.

Value

Tibble with tournament winners by year

Examples

```
## Not run:
season <- load_leaderboard(2025)
tournament_history("Masters", data = season)

## End(Not run)
```

update_db *Build or Update a Local golfastr Database*

Description

Writes golfastr's hosted data into a database on your computer, so it can be queried with SQL or dplyr. By default this is an SQLite file (`golfastr_db` in `dbdir`); pass `db_connection` to use any other DBI database, such as DuckDB or Postgres.

Usage

```
update_db(
  dbdir = getOption("golfastr.dbdirectory", default = "."),
  dbname = "golfastr_db",
 tables = c("events", "holes", "leaderboards", "strokes_gained", "skill_ratings",
    "player_ids", "event_ids"),
  force_rebuild = FALSE,
  db_connection = NULL
)
```

Arguments

`dbdir`	Folder for the SQLite file. Defaults to `getOption("golfastr.dbdirectory")`, or the working directory.
`dbname`	File name of the SQLite database.
`tables`	Tables to build. The default leaves out the ShotLink tables, which are large (`shots` is about a million rows per season); add `"shots"`, `"tee_times"` and `"course_setup"` to include them.
`force_rebuild`	If `TRUE`, drop the requested tables and build them again from scratch.
`db_connection`	Optional DBI connection to write to instead of the SQLite file. It is left open.

Details

The first call downloads everything requested. Later calls add seasons that are missing and refresh the latest season, so running it once a week keeps the database current. Lookup tables (ratings, IDs, strokes gained) are replaced in full each time.

Every season-based table has a `season` column. Tables built from ESPN data (`holes`, `events`, `leaderboards`, `skill_ratings`) use ESPN IDs in `tournament_id` and `player_id`. ShotLink tables (`shots`, `tee_times`, `course_setup`) name theirs `pga_tournament_id` and `pga_player_id`, and also carry `espn_tournament_id` and `espn_player_id` where the ID crosswalk has a match, so the two sources join directly. Dates are stored as `"YYYY-MM-DD"` text and times as ISO 8601 text in UTC.

Value

The DBI connection, invisibly. When golfastr opened the SQLite file itself it is closed and `NULL` is returned; reopen it with `DBI::dbConnect(RSQLite::SQLite(), file.path(dbdir, dbname))`.

Examples

```
if (requireNamespace("RSQLite", quietly = TRUE)) {
  dir <- tempdir()
  update_db(dbdir = dir, tables = c("events", "player_ids"))
  con <- DBI::dbConnect(RSQLite::SQLite(), file.path(dir, "golfastr_db"))
  DBI::dbGetQuery(con, "SELECT season, COUNT(*) AS events FROM events GROUP BY season")
  DBI::dbDisconnect(con)
}
```

---

win_leaders *Most Wins*

Description

Get players with most wins.

Usage

```
win_leaders(year = NULL, top_n = 10, file_path = NULL, data = NULL)
```

Arguments

year	Optional year filter
top_n	Number of players to show
file_path	Path to data file (.rds or .parquet)
data	Optional leaderboard data frame. Use this instead of 'file_path' when the data is already in memory.

Value

Tibble with win leaders

Examples

```
## Not run:
season <- load_leaderboard(2025)
win_leaders(data = season)

## End(Not run)
```

Index

∗ **datasets**
    strokes_gained, 39

add_win_probability, 3

build_season, 4

cache_info, 5
calculate_strokes_gained, 6
check_season, 7
clear_cache, 8
compare_players, 8

field_strength, 9

get_majors, 10
get_pga_schedule *(*load_schedule*)*, 23
get_player, 10
get_player_scorecards *(*load_holes*)*, 16
get_tournament_leaderboard
        *(*load_leaderboard*)*, 18
get_winners, 11

leaderboard, 12
list_tournaments, 12
load_course_setup, 13
load_data, 14
load_event_ids, 14
load_from_parquet, 15
load_from_rds, 16
load_holes, 16
load_leaderboard, 18
load_logos, 19
load_pga_hbh, 20
load_pga_leaderboards, 21
load_pga_schedule, 21
load_player_ids, 22
load_players, 23
load_schedule, 23
load_shots, 24
load_strokes_gained, 25
load_tee_times, 26
load_tournament, 27
load_tournament_detail, 28
load_witb, 28
load_witb_switches, 29

made_cuts_leaders, 30

pga_field_descriptions, 31
pga_majors, 31
pga_score_types, 32
player_summary, 32
plot_head_to_head, 33
plot_leaderboard, 34
plot_player, 34
plot_scoring, 35
plot_wins, 36

save_to_parquet, 37
save_to_rds, 37
scoring_avg_leaders, 38
strokes_gained, 39
summarise_strokes_gained, 39

top10_leaders, 40
tournament_history, 41

update_db, 42

win_leaders, 43