



baseballr :: CHEAT SHEET

v2.0.0

The SportsDataverse R package for baseball data: one tidy interface over the MLB Stats API, Baseball Savant / Statcast, FanGraphs, Baseball Reference, the Chadwick register, Retrosheet, Spotrac, Fox Sports, ESPN (MLB + college) and stats.ncaa.org college baseball. Sister package [softballr](#); Python mirror [sportsdataverse.mlb](#) / [sportsdataverse.baseball](#) in [sportsdataverse.py](#). Page 1 of 2: install, MLB Stats API, Statcast, FanGraphs, Baseball Reference, Chadwick, Retrosheet.

R ≥ 4.1.0 · tibles out
baseballr.sportsdataverse.org

GET STARTED

```
install.packages("baseballr") # CRAN
pak::pak("BillPetti/baseballr") # dev
version
library(baseballr)
most_recent_mlb_season() #
default_season
```

Public sources, no API key. Functions return tibles; may accept ... passed to httr. Legacy names (get_*_mlb, scrape_statcast_savant*) still export as aliases.

MLB STATS API · SCHEDULE & GAMES

```
mlb_schedule(season = 2024, level_ids = "1")
mlb_game_pks("2024-04-29", level_ids = c(1))
mlb_pbp(game_pk = 632970) # pitch-level
feed/live
mlb_pbp(game_pk, add_base_state = TRUE) #
pre_on_1b/2b/3b
mlb_pbp_diff(game_pk, start_timecode,
end_timecode)
mlb_batting_orders(game_pk, type = "starting")
# | "all"
mlb_probables(game_pk)
mlb_game_info(game_pk)
mlb_game_linescore(game_pk)
mlb_game_wp(game_pk)
mlb_game_content(game_pk)
mlb_game_timecodes(game_pk)
mlb_schedule_postseason(season = 2023,
sport_id = 1)
mlb_schedule_postseason_series(...)
mlb_venues(venue_ids = 4781)
mlb_venues(sport_ids = 1)
mlb_attendance(team_id = 109, season = 2024)
```

level_ids / sport_id: 1=MLB; 11 AAA · 12 AA · 13 High-A · 14 Low-A · 16 Rookie · 22 College (see mlb_sports()). Also mlb_game_pace(), mlb_game_changes(), mlb_game_context_metrics(), mlb_schedule_games_tied().

BASEBALL REFERENCE

```
bref_daily_batter(t1 = "2015-05-10", t2 =
"2015-06-20")
bref_daily_pitcher(t1 = "2015-04-05", t2 =
"2015-04-30")
bref_team_results(Tm = "NYM", year = 2015)
bref_standings_on_date(date = "2015-08-04",
division = "AL East")
# era-aware: pre-1969 league-only · 1969-93
E/W · 1994+ E/C/W
```

PYTHON MIRROR

sportsdataverse.mlb mirrors this surface (mlb_*, mlb_statcast_*, espn_mlb_pbp, load_mlb_* models); college lives in sportsdataverse.baseball.

MLB STATS API · TEAMS, PLAYERS, STANDINGS

```
mlb_teams(season = 2024, sport_ids = 1)
mlb_team_info(team_id = 137)
mlb_rosters(team_id = 109, season = 2018,
roster_type = "active") #
mlb_roster_types()
mlb_team_stats(team_id = 137, stat_type =
"season",
stat_group = "hitting", season
= 2021)
mlb_team_leaders(team_id = 137,
leader_categories =
"homeRuns")
mlb_standings(season = 2021,
league_id = 103) # 103 AL ·
104 NL
mlb_people(person_ids = c(502671, 605151))
mlb_player_game_stats(person_id, game_pk)
mlb_stats(stat_type = "season", stat_group =
"hitting",
season = 2021) # league-
wide table
mlb_stats_leaders(leader_categories =
"homeRuns",
sport_id = 1, season = 2021)
mlb_teams_stats(stat_type = "season",
stat_group = "pitching",
season = 2021)
```

Options lookups: mlb_stat_types(), mlb_stat_groups(), mlb_league_leader_types(), mlb_standings_types(), mlb_positions(). Situational splits: stat_type = "statSplits", sit_codes = c("h", "a", "risp") on mlb_stats() / mlb_teams_stats().

MORE MLB STATS FAMILIES

Gro up	Functions
Draft	mlb_draft(year) mlb_draft_latest(year) mlb_draft_prospects(year)
Awards	mlb_awards() mlb_awards_recipient() mlb_all_star_*(*) mlb_homerun_derby(game_pk)
League	mlb_league() mlb_divisions() mlb_conferences() mlb_sports() mlb_seasons(sport_id = 1) mlb_seasons_all()
Team org	mlb_team_history() mlb_team_affiliates() mlb_team_coaches() mlb_team_personnel() mlb_team_alumni()
Jobs	mlb_jobs(job_type = "UMPR") · mlb_jobs_*(*)
People	mlb_people_free_agents() mlb_player_game_stats_current() mlb_high_low_stats()
Codes	mlb_stat_types() · mlb_*_codes() · mlb_hit_trajectories() · mlb_metrics()

BASEBALL SAVANT · STATCAST SEARCH

```
sc <- statcast_search(start_date = "2024-04-06",
end_date = "2024-04-15",
player_type = "batter")
# | "pitcher"
statcast_search_batters("2016-04-06", "2016-04-15",
batterid = 621043)
statcast_search_pitchers(start_date, end_date,
pitcherid = NULL)
process_statcast_payload(payload) # raw ->
typed
statcast_search_minors(...)
statcast_search_wbc(...)
statcast_pitch_colors() # Savant pitch-type
palette
```

One pitch per row; Savant truncates big result sets, so loop short date windows. playerid is the MLBAM id (Chadwick).

SAVANT LEADERBOARDS

```
statcast_leaderboards(leaderboard =
"expected_statistics",
year = 2018)
statcast_leaderboards(leaderboard =
"arm_strength",
year = 2020, min_throws
= 100)
statcast_leaderboards("exit_velocity_barrels",
year = 2023,
abs = 50, player_type =
"batter")
```

leaderboard =	key args
exit_velocity_barrels	abs, player_type
expected_statistics	min_pa, player_type
pitch_arsenal	min_pitches, arsenal_type, hand
outs_above_average	min_field, fielding_type, oaa_position, oaa_roles
pop_time	min2b, min3b
sprint_speed	min_run, position
arm_strength	min_throws
+ directional_oaa_running_splits_90_ft	catch_probability

SPRAY CHARTS

```
ggspraychart(sc, x_value = "hc_x", y_value =
"-hc_y",
fill_value = "events")
ggspraychart(sc, density = TRUE, bin_size =
15)
ggpitchzone(sc) # pitch locations + strike
zone
```

Statcast metric helpers (code_barrel, edge_code, run

FANGRAPHS

```
fg_batter_leaders(startseason = 2023,
endseason = 2023,
qual = "0", ind = "0") #
ind=1 by season
fg_pitcher_leaders(startseason = 2023,
endseason = 2023)
fg_fielder_leaders(startseason = 2023,
endseason = 2023)
fg_team_batter(startseason = 2023, endseason =
2023)
fg_team_pitcher(...) fg_team_fielder(...)
fg_batter_game_logs(playerid = 19755, year =
2023)
fg_pitcher_game_logs(playerid, year)
fg_mlb_batter_game_logs(playerid, year) # +
pitcher
fg_projections(type = "steamer") #
ZIPS · ATC · THE BAT · RoS
fg_guts() fg_park(yr) fg_park_hand(yr)
```

Leader args mirror the FanGraphs URL (pos, lg, qual, month, hand, team, type, sortstat).

CHADWICK IDS & RETROSHEET

```
lu <- chadwick_player_lu() # full public
register
playerid_lookup(last_name = "Garcia",
first_name = "Karim")
playername_lookup(id) # mlbam | fangraphs |
bbref | retro
chadwick_is_installed(); chadwick_path() #
cwevent
retrosheet_data(path_to_directory = NULL,
years_to_acquire = 2023) #
Chadwick CLI
```

OTHER MLB SOURCES

```
spotrac_league_payrolls(year) #
Spotrac
spotrac_team_active_payroll("BAL", year)
fox_mlb_team_roster(team_id)
fox_mlb_standings(team_id)
fox_mlb_odds(game_id)
fox_mlb_league_leaders("batting")
espn_mlb_scoreboard(season)
espn_mlb_pbp(game_id)
espn_mlb_team_box(game_id)
espn_mlb_player_box(game_id)
espn_mlb_standings(year)
espn_mlb_wp(game_id)
```

Plus 107 espn_mlb_* wrappers in all — college twins on page 2.



NCAA · STATS.NCAA.ORG

TEAMS & IDS

```
library(baseballr)
ncaa_teams(year = 2024, division = 1)      #
1 | 2 | 3
ncaa_school_id_lu(team_name = "Vanderbilt") #
-> team_id
id <-
ncaa_school_id_lu("Vanderbilt")$team_id[1]
load_ncaa_baseball_teams()                # bundled
team lookup
load_ncaa_baseball_season_ids()            # season_id
crosswalk
most_recent_ncaa_baseball_season()
```

team_id is the stats.ncaa.org school id; year is the season end year. ncaa_school_id_lu() resolves against load_ncaa_baseball_teams(), no live scrape.

SCHEDULES, ROSTERS & STATS

```
sched <- ncaa_schedule_info(team_id = id, year = 2024,
                             pbp_links = TRUE)
# + game_info_url / game_pbp_url per game
ncaa_roster(team_id = id, year = 2024)
ncaa_team_player_stats(team_id = id, year = 2024,
                        type = "batting") #
|pitching|fielding
ncaa_game_logs(player_id, year = 2024,
                type = "batting", span =
"game") # | "career"
ncaa_park_factor(team_id = id, years = 2019:2024,
                  type = "conference") # |
"division"
```

stats.ncaa.org sits behind Akamai bot protection; live calls can fail or rate-limit. Prefer the loaders (col 2) for bulk history and scrape only the gap.

LEGACY ALIASES

get_ncaa_schedule_info get_ncaa_baseball_roster
get_ncaa_baseball_pbp get_ncaa_lineups
get_ncaa_game_logs get_ncaa_park_factor
ncaa_scrape school_id_lu ncaa_baseball_pbp
ncaa_baseball_roster — all point at the ncaa_* names above.

PLAY-BY-PLAY & LINEUPS

```
pbp <- ncaa_pbp(game_info_url =
sched$game_info_url[1])
pbp <- ncaa_pbp(game_pbp_url =
sched$game_pbp_url[1],
               raw_html_to_disk = TRUE,
               raw_html_path = "raw/") #
cache html
ncaa_pbp(read_from_file = TRUE, file =
"raw/<id>.html")
ncaa_lineups(game_info_url =
sched$game_info_url[1])
```

One row per play: inning, inning_top_bot, running score, batting / fielding team, description, game_pbp_id. Pass game_info_url or game_pbp_url (or a saved html file). ncaa_lineups() returns one row per batter per team.

DATA LOADERS

```
load_ncaa_baseball_pbp(seasons = 2021:2024)
load_ncaa_baseball_schedule(seasons = 2024)
load_ncaa_baseball_teams()
load_ncaa_baseball_season_ids()
load_game_info_sup()      # supplementary MLB
game info
load_umpire_ids()         # MLB umpire ids
# 11 MLB model loaders (sportsdataverse-data):
load_mlb_{expected_stats,expected_hr,oaa,stuff_plus,
command_plus,xera,batter_projection,catcher_framing,
re24_matrix,we_table,wpa}()
# write straight into a DB instead of
returning:
load_ncaa_baseball_pbp(2024, dbConnection =
con,
                      tablename = "ncaa_pbp")
```

Loader	Reads from
load_ncaa_baseball_pbp	sportsdataverse-data release ncaa_baseball_pbp (rds per season)
load_ncaa_baseball_schedule	sportsdataverse-data release ncaa_baseball_schedules
load_ncaa_baseball_teams / _season_ids	baseballr-data repo (ncaa/teams_info, ncaa/seasons_info)
load_game_info_sup / load_umpire_ids	baseballr-data repo

BUNDLED DATA

teams_lu_table	MLB team lookup (used by bref_team_results, team_consistency)
statcast_impute	imputed-value table for label_statcast_imputed_data()
column_structur e_draft_mlb	column skeleton for mlb_draft()
stats_api_live_ empty_df	empty live-feed schema

ESPN COLLEGE BASEBALL

GAMES, TEAMS & SEASON

```
espn_college_baseball_scoreboard(season = 2024)
espn_college_baseball_pbp(game_id)
espn_college_baseball_team_box(game_id)
espn_college_baseball_player_box(game_id)
espn_college_baseball_teams()
espn_college_baseball_conferences()
espn_college_baseball_standings(year = 2024)
espn_college_baseball_team_roster(team_id, season = 2024)
espn_college_baseball_team_schedule(team_id, season = 2024,
season_type = 2)
espn_college_baseball_season_rankings(season = 2024)
espn_college_baseball_leaders(season = 2024)
most_recent_college_baseball_season()
```

70 espn_college_baseball_* exports — thin twins of the espn_mlb_* family over the same league-parameterised helpers; extra query args pass through ...

Fa mil y	Functions
Game	_game_all _game_play _game_rosters _game_officials _game_situation _game_team_statistics _game_team_linescores _game_broadcasts
Team	_team _team_record _team_leaders _team_news _team_season_roster _team_season_statistics _team_season_profile
Player	_player_info _player_gamelog _player_splits _player_career_stats _player_awards _athletes_index
Season	_seasons _season_info _season_types _season_weeks _season_groups _season_group_teams _week_rankings
Misc	_coaches _calendar _news _tournaments _venues

METRICS

WOBA, FIP, CONSISTENCY

```
bat <- bref_daily_batter("2015-08-01", "2015-10-03")
woba_plus(bat) # wOBA + wOBA_CON via fg_guts()
pit <- bref_daily_pitcher("2015-04-05", "2015-04-30")
fip_plus(pit) # FIP + FIP-team_consistency(year = 2015) # run scoring / prevention
```

Both take any frame with the bref-style columns (season, uBB, HBP, X1B ...); the season is used to pick the right linear weights.

RUN EXPECTANCY & LINEAR WEIGHTS

```
sc <- statcast_search(start_date = "2016-04-06",
                      end_date = "2016-04-15",
                      player_type = "batter")
sc <- run_expectancy_code(sc,
                          level = "plate appearance") # |
"pitch"
linear_weights_savant(sc, level = "plate appearance")
```

run_expectancy_code() appends final_pitch_*, bat_score_*, runs_to_end_inning and the run-expectancy columns per plate appearance (or pitch); linear_weights_savant() turns that into event run values for the window.

STATCAST CLASSIFIERS

```
sc <- code_barrel(sc) # + barrel
(launch_speed/angle)
sc <- edge_code(sc, height_var_name =
"b_height")
edge_frequency(sc, group = "pitcher") # Edge% by group
statline_from_statcast(sc, base = "pa") #
swings|contact|pa
sc <- label_statcast_imputed_data(sc) # +
imputed flag
```

Helper	Needs columns
code_barrel	launch_angle, launch_speed
edge_code	b_height, stand, pz, px
edge_frequency	output of edge_code()
statline_from_statcast	events, description, game_date, type
label_statcast_imputed_data	raw Savant frame + statcast_impute

PYTHON MIRROR

sportsdataverse.baseball.college_baseball:
stats.ncaa.org pbp + box parsers and run-expectancy helpers; load_ncaa_baseball_pbp() / load_ncaa_baseball_schedule() read the same releases. sportsdataverse.mlb:
mlb_run_expectancy_matrix, mlb_expected_stats, mlb_fielding_oaa, mlb_catcher_framing, load_mlb_re24_matrix, load_mlb_wpa.

UTILITIES

most_recent_mlb_season() ·
most_recent_ncaa_baseball_season() ·
most_recent_college_baseball_season() ·
chadwick_is_installed() · chadwick_find_lib() ·
chadwick_set_ld_library_path().



baseballr :: CHEAT SHEET

v2.0.0

The SportsDataverse R package for baseball data: one tidy interface over the MLB Stats API, Baseball Savant / Statcast, FanGraphs, Baseball Reference, the Chadwick register, Retrosheet, Spotrac, Fox Sports, ESPN (MLB + college) and stats.ncaa.org college baseball. Sister package **softballR**; Python mirror `sportsdataverse.mlb / sportsdataverse.baseball` in `sportsdataverse.py`. Page 1 of 2: install, MLB Stats API, Statcast, FanGraphs, Baseball Reference, Chadwick, Retrosheet.

GET STARTED

```
install.packages("baseballr") # CRAN
pak::pak("BillPetti/baseballr") # dev
version
library(baseballr)
most_recent_mlb_season() #
default_season
```

Public sources, no API key. Functions return tibbles; many accept ... passed to httr. Legacy names (`get_*_mlb` , `scrape_statcast_savant*`) still export as aliases.

MLB STATS API · SCHEDULE & GAMES

```
mlb_schedule(season = 2024, level_ids = "1")
mlb_game_pks("2024-04-29", level_ids = c(1))
mlb_pbp(game_pk = 632970) # pitch-level feed/live
mlb_pbp(game_pk, add_base_state = TRUE) # pre_on_1b/2b/3b
mlb_pbp_diff(game_pk, start_timecode, end_timecode)
mlb_batting_orders(game_pk, type = "starting")
# | "all"
mlb_probables(game_pk)
mlb_game_info(game_pk)
mlb_game_linescore(game_pk)
mlb_game_wp(game_pk)
mlb_game_content(game_pk)
mlb_game_timecodes(game_pk)
mlb_schedule_postseason(season = 2023, sport_id = 1)
mlb_schedule_postseason_series(...)
mlb_venues(venue_ids = 4781)
mlb_venues(sport_ids = 1)
mlb_attendance(team_id = 109, season = 2024)
```

`level_ids / sport_id`: 1 = MLB; 11 AAA · 12 AA · 13 High-A · 14 Low-A · 16 Rookie · 22 College (see `mlb_sports()`). Also `mlb_game_pace()` , `mlb_game_changes()` , `mlb_game_context_metrics()` , `mlb_schedule_games_tied()` .

BASEBALL REFERENCE

```
bref_daily_batter(t1 = "2015-05-10", t2 = "2015-06-20")
bref_daily_pitcher(t1 = "2015-04-05", t2 = "2015-04-30")
bref_team_results(Tm = "NYM", year = 2015)
bref_standings_on_date(date = "2015-08-04", division = "AL East")
# era-aware: pre-1969 league-only · 1969-93 E/W · 1994+ E/C/W
```

PYTHON MIRROR

`sportsdataverse.mlb` mirrors this surface (`mlb_*` , `mlb_statcast_*` , `espn_mlb_pbp` , `load_mlb_*` models; college lives in `sportsdataverse.baseball` .

MLB STATS API · TEAMS, PLAYERS, STANDINGS

```
mlb_teams(season = 2024, sport_ids = 1)
mlb_team_info(team_id = 137)
mlb_rosters(team_id = 109, season = 2018, roster_type = "active") #
mlb_roster_types()
mlb_teams_stats(team_id = 137, stat_type = "season", stat_group = "hitting", season = 2021)
mlb_team_leaders(team_id = 137, leader_categories = "homeRuns")
mlb_standings(season = 2021, league_id = 103) # 103 AL · 104 NL
mlb_people(person_ids = c(502671, 605151))
mlb_player_game_stats(person_id, game_pk)
mlb_stats(stat_type = "season", stat_group = "hitting", season = 2021) # league-wide table
mlb_stats_leaders(leader_categories = "homeRuns", sport_id = 1, season = 2021)
mlb_teams_stats(stat_type = "season", stat_group = "pitching", season = 2021)
```

Options lookups: `mlb_stat_types()` , `mlb_stat_groups()` , `mlb_league_leader_types()` , `mlb_standings_types()` , `mlb_positions()` . Situational splits: `stat_type = "statSplits"` , `sit_codes = c("h", "a", "risp")` on `mlb_stats()` / `mlb_teams_stats()` .

MORE MLB STATS FAMILIES

Gro up	Functions
Draft	<code>mlb_draft(year)</code> <code>mlb_draft_latest(year)</code> <code>mlb_draft_prospects(year)</code>
Awards	<code>mlb_awards()</code> <code>mlb_awards_recipient()</code> <code>mlb_all_star_*(())</code> <code>mlb_homerun_derby(game_pk)</code>
League	<code>mlb_league()</code> <code>mlb_divisions()</code> <code>mlb_conferences()</code> <code>mlb_sports()</code> <code>mlb_seasons(sport_id = 1)</code> <code>mlb_seasons_all()</code>
Team	<code>mlb_team_history()</code> <code>mlb_team_affiliates()</code> <code>mlb_team_coaches()</code> <code>mlb_team_personnel()</code>
Org	<code>mlb_team_alumni()</code>
Jobs	<code>mlb_jobs(job_type = "UMPR") · mlb_jobs_*(())</code>
People	<code>mlb_people_free_agents()</code> <code>mlb_player_game_stats_current()</code> <code>mlb_high_low_stats()</code>

`mlb_*_types()` · `mlb_*_codes()` · `mlb_hit_trajectories()` · `mlb_metrics()` · `mlb_baseball_stats()`

BASEBALL SAVANT · STATCAST SEARCH

```
sc <- statcast_search(start_date = "2024-04-06", end_date = "2024-04-15", player_type = "batter")
# | "pitcher"
statcast_search_batters("2016-04-06", "2016-04-15", batterid = 621043)
statcast_search_pitchers(start_date, end_date, pitcherid = NULL)
process_statcast_payload(payload) # raw -> typed
statcast_search_minors(...)
statcast_search_wbc(...)
statcast_pitch_colors() # Savant pitch-type palette
```

One pitch per row; Savant truncates big result sets, so loop short date windows. `playerid` is the MLBAM id (Chadwick).

SAVANT LEADERBOARDS

```
statcast_leaderboards(leaderboard = "expected_statistics", year = 2018)
statcast_leaderboards(leaderboard = "arm_strength", year = 2020, min_throws = 100)
statcast_leaderboards("exit_velocity_barrels", year = 2023, abs = 50, player_type = "batter")
```

leaderboard =	key args
exit_velocity_barrels	<code>abs</code> , <code>player_type</code>
expected_statistics	<code>min_pa</code> , <code>player_type</code>
pitch_arsenal	<code>min_pitches</code> , <code>arsenal_type</code> , <code>hand</code>
outs_above_averagerage	<code>min_field</code> , <code>fielding_type</code> , <code>oaa_position</code> , <code>oaa_roles</code>
pop_time	<code>min2b</code> , <code>min3b</code>
sprint_speed	<code>min_run</code> , <code>position</code>
arm_strength	<code>min_throws</code>
+ directional_oaa · catch_probability · running_splits_90_ft	

SPRAY CHARTS

```
ggspraychart(sc, x_value = "hc_x", y_value = "hc_y", fill_value = "events")
ggspraychart(sc, density = TRUE, bin_size = 15)
ggpitchzone(sc) # pitch locations + strike zone
```

Statcast metric helpers (`code_barrel` , `edge_code` , `run_expectancy`) — page 2.

FANGRAPHS

```
fg_batter_leaders(startseason = 2023, endseason = 2023, qual = "0", ind = "0") #
ind=1 by season
fg_pitcher_leaders(startseason = 2023, endseason = 2023)
fg_fielder_leaders(startseason = 2023, endseason = 2023)
fg_team_batter(startseason = 2023, endseason = 2023)
fg_team_pitcher(...) fg_team_fielder(...)
fg_batter_game_logs(playerid = 19755, year = 2023)
fg_pitcher_game_logs(playerid, year)
fg_mlb_batter_game_logs(playerid, year) # + pitcher
fg_projections(type = "steamer") # ZiPS·ATC·THE BAT·RoS
fg_guts() fg_park(yr) fg_park_hand(yr)
```

Leader args mirror the FanGraphs URL (`pos` , `lg` , `qual` , `month` , `hand` , `team` , `type` , `sortstat`).

CHADWICK IDS & RETROSHEET

```
lu <- chadwick_player_lu() # full public register
playerid_lookup(last_name = "Garcia", first_name = "Karim")
playername_lookup(id) # mlbam | fangraphs | bbref | retro
chadwick_is_installed(); chadwick_path() # cwevent
retrosheet_data(path_to_directory = NULL, years_to_acquire = 2023) #
Chadwick CLI
```

OTHER MLB SOURCES

```
sptrc_league_payrolls(year) #
Spotrac
sptrc_team_active_payroll("BAL", year)
fox_mlb_team_roster(team_id)
fox_mlb_standings(team_id)
fox_mlb_odds(game_id)
fox_mlb_league_leaders("batting")
espn_mlb_scoreboard(season)
espn_mlb_pbp(game_id)
espn_mlb_team_box(game_id)
espn_mlb_player_box(game_id)
espn_mlb_standings(year)
espn_mlb_wp(game_id)
```

Plus 107 `espn_mlb_*` wrappers in all — college twins on page 2.



NCAA · STATS.NCAA.ORG

TEAMS & IDS

```
library(baseballr)
ncaa_teams(year = 2024, division = 1) # 1
| 2 | 3
ncaa_school_id_lu(team_name = "Vanderbilt") #
-> team_id
id <-
ncaa_school_id_lu("Vanderbilt")$team_id[1]
load_ncaa_baseball_teams() # bundled
team lookup
load_ncaa_baseball_season_ids() # season_id
crosswalk
most_recent_ncaa_baseball_season()
```

team_id is the stats.ncaa.org school id; year is the season end year. ncaa_school_id_lu() resolves against load_ncaa_baseball_teams(), no live scrape.

SCHEDULES, ROSTERS & STATS

```
sched <- ncaa_schedule_info(team_id = id, year = 2024,
                           pbp_links = TRUE)
# + game_info_url / game_pbp_url per game
ncaa_roster(team_id = id, year = 2024)
ncaa_team_player_stats(team_id = id, year = 2024,
                       type = "batting") #
|pitching|fielding
ncaa_game_logs(player_id, year = 2024,
               type = "batting", span = "game")
# | "career"
ncaa_park_factor(team_id = id, years = 2019:2024,
                 type = "conference") # |
"division"
```

stats.ncaa.org sits behind Akamai bot protection; live calls can fail or rate-limit. Prefer the loaders (col 2) for bulk history and scrape only the gap.

LEGACY ALIASES

get_ncaa_schedule_info get_ncaa_baseball_roster
get_ncaa_baseball_pbp get_ncaa_lineups
get_ncaa_game_logs get_ncaa_park_factor
ncaa_scrape school_id_lu ncaa_baseball_pbp
ncaa_baseball_roster — all point at the ncaa_* names above.

PLAY-BY-PLAY & LINEUPS

```
pbp <- ncaa_pbp(game_info_url =
sched$game_info_url[1])
pbp <- ncaa_pbp(game_pbp_url =
sched$game_pbp_url[1],
               raw_html_to_disk = TRUE,
               raw_html_path = "raw/") #
cache html
ncaa_pbp(read_from_file = TRUE, file =
"raw/<id>.html")
ncaa_lineups(game_info_url =
sched$game_info_url[1])
```

One row per play: inning, inning_top_bot, running score, batting / fielding team, description, game_pbp_id. Pass game_info_url or game_pbp_url (or a saved html file). ncaa_lineups() returns one row per batter per team.

DATA LOADERS

```
load_ncaa_baseball_pbp(seasons = 2021:2024)
load_ncaa_baseball_schedule(seasons = 2024)
load_ncaa_baseball_teams()
load_ncaa_baseball_season_ids()
load_game_info_sup() # supplementary MLB
game info
load_umpire_ids() # MLB umpire ids
# 11 MLB model loaders (sportsdataverse-data):
load_mlb_{expected_stats,expected_hr,oa,stuff_plus,
command_plus,xera,batter_projection,catcher_framing,
re24_matrix,we_table,wpa}()
# write straight into a DB instead of
returning:
load_ncaa_baseball_pbp(2024, dbConnection =
con,
                      tablename = "ncaa_pbp")
```

Loader	Reads from
load_ncaa_baseball_pbp	sportsdataverse-data release ncaa_baseball_pbp (rds per season)
load_ncaa_baseball_schedule	sportsdataverse-data release ncaa_baseball_schedules
load_ncaa_baseball_teams / _season_ids	baseballr-data repo (ncaa/teams_info, ncaa/seasons_info)
load_game_info_sup / load_umpire_ids	baseballr-data repo

BUNDLED DATA

teams_lu_table	MLB team lookup (used by bref_team_results, team_consistency)
statcast_impute	imputed-value table for label_statcast_imputed_data()
column_struct_e_draft_mlb	column skeleton for mlb_draft()
stats_api_live_empty_df	empty live-feed schema

ESPN COLLEGE BASEBALL

GAMES, TEAMS & SEASON

```
espn_college_baseball_scoreboard(season = 2024)
espn_college_baseball_pbp(game_id)
espn_college_baseball_team_box(game_id)
espn_college_baseball_player_box(game_id)
espn_college_baseball_teams()
espn_college_baseball_conferences()
espn_college_baseball_standings(year = 2024)
espn_college_baseball_team_roster(team_id,
season = 2024)
espn_college_baseball_team_schedule(team_id,
season = 2024,
                                season_type
= 2)
espn_college_baseball_season_rankings(season = 2024)
espn_college_baseball_leaders(season = 2024)
most_recent_college_baseball_season()
```

70 espn_college_baseball_* exports — thin twins of the espn_mlb_* family over the same league-parameterised helpers; extra query args pass through

Fa mil y	Functions
Game	_game_all _game_play _game_rosters _game_officials _game_situation _game_team_statistics _game_team_linescores _game_broadcasts
Team	_team _team_record _team_leaders _team_news _team_season_roster _team_season_statistics _team_season_profile
Player	_player_info _player_game_log _player_splits _player_career_stats _player_awards _athletes_index
Season	_seasons _season_info _season_types _season_weeks _season_groups _season_group_teams _week_rankings
Misc	_coaches _calendar _news _tournaments _venues

METRICS

WOBA, FIP, CONSISTENCY

```
bat <- bref_daily_batter("2015-08-01", "2015-10-03")
woba_plus(bat) # wOBA + wOBA_CON via fg_guts()
pit <- bref_daily_pitcher("2015-04-05", "2015-04-30")
fip_plus(pit) # FIP + FIP-
team_consistency(year = 2015) # run scoring / prevention
```

Both take any frame with the bref-style columns (season, uBB, HBP, X1B ...); the season is used to pick the right linear weights.

RUN EXPECTANCY & LINEAR WEIGHTS

```
sc <- statcast_search(start_date = "2016-04-06",
                    end_date = "2016-04-15",
                    player_type = "batter")
sc <- run_expectancy_code(sc,
                        level = "plate appearance") # |
"pitch"
linear_weights_savant(sc, level = "plate appearance")
```

run_expectancy_code() appends final_pitch_*, bat_score_*, runs_to_end_inning and the run-expectancy columns per plate appearance (or pitch); linear_weights_savant() turns that into event run values for the window.

STATCAST CLASSIFIERS

```
sc <- code_barrel(sc) # + barrel
(launch_speed/angle)
sc <- edge_code(sc, height_var_name =
"b_height")
edge_frequency(sc, group = "pitcher") # Edge%
by group
statline_from_statcast(sc, base = "pa") #
swings|contact|pa
sc <- label_statcast_imputed_data(sc) # +
imputed flag
```

Helper	Needs columns
code_barrel	launch_angle, launch_speed
edge_code	b_height, stand, pz, px
edge_frequency	output of edge_code()
statline_from_statcast	events, description, game_date, type
label_statcast_imputed_data	raw Savant frame + statcast_impute

PYTHON MIRROR

sportsdataverse.baseball.college_baseball:
stats.ncaa.org pbp + box parsers and run-expectancy helpers;
load_ncaa_baseball_pbp() /
load_ncaa_baseball_schedule() read the same releases.
sportsdataverse.mlb: mlb_run_expectancy_matrix,
mlb_expected_stats, mlb_fielding_oaa,
mlb_catcher_framing, load_mlb_re24_matrix,
load_mlb_wpa.

UTILITIES

most_recent_mlb_season() ·
most_recent_ncaa_baseball_season() ·
most_recent_college_baseball_season() ·
chadwick_is_installed() · chadwick_find_lib() ·
chadwick_set_ld_library_path().