

TABLE: game_events



VARIABLES	game_str	play_id	at_bat	play_per_game	timestamp	player_position	event_code
OTHER TABLES							
	ball_pos	ball_pos			ball_pos		
	player_pos	player_pos			player_pos	player_pos	
	game_info		game_info	game_info		(game_info)	
DESCRIPTION	Each game string identifies the year and the sequential day within the season.	A play is defined as a situation where the ball is live. Play ids begin at 1 and are listed consecutively for each game.	An at-bat corresponds to a batter at the plate. They are listed consecutively for each game.	Plays-per-game are comparable to play ids.	Times are shown in milliseconds.	Numbers corresponding to positions are given in the Glossary.	Numbers corresponding to ball events are given in the Glossary.
	Fields are structured as Year_Day_AwayTeam_HomeTeam.	Play ids generally include at least three events: the pitch, what happens to the ball (often a catch of the pitch by the catcher), and the end of the play.	An at-bat includes at least one play id and one play-per-game.	Some plays (e.g. pickoff throws) occur during an at-bat, but are not associated with the at-bat.	Timestamps start at the beginning of each game, with the time of the initial play beginning at t < 60 seconds.	Each player position corresponds to a column given in Game Info, with the column names connected through the Glossary. That information corresponds to player ids at each position.	

TABLE:
ball_pos



VARIABLES	game_str	play_id	timestamp	ball_position_x	ball_position_y	ball_position_z	
OTHER TABLES	game_events	game_events	game_events				
	player_pos	player_pos	player_pos				
	game_info						
DESCRIPTION	Each game string identifies the year and the sequential day within the season.	A play is defined as a situation where the ball is live. Play ids begin at 1 and are listed consecutively for each game.	Times are shown in milliseconds.	Dimensions are given in feet.	Dimensions are given in feet.	Dimensions are given in feet.	
	Fields are structured as Year_Day_AwayTeam_HomeTeam.	Play ids generally include at least three events: the pitch, what happens to the ball (often a catch of the pitch by the catcher), and the end of the play.	Timestamps start at the beginning of each game, with the time of the initial play beginning at t < 60 seconds.	x = 0 is defined at the line between home plate and second base, with x > 0 towards first base.	y = 0 is at the back of home plate, with y > 0 towards second base.	z is the height with respect to z=0 in real-world coordinates. z=0 does not necessarily correspond to the ground, which is neither flat nor level.	

TABLE:
player_pos



VARIABLES	game_str	play_id	timestamp	player_position	field_x	field_y	
OTHER TABLES	game_events	game_events	game_events	game_events			
	ball_pos	ball_pos	ball_pos				
	game_info			(game_info)			
DESCRIPTION	Each game string identifies the year and the sequential day within the season.	A play is defined as a situation where the ball is live. Play ids begin at 1 and are listed consecutively for each game.	Times are shown in milliseconds.	Numbers corresponding to positions are given in the Glossary.	Dimensions are given in feet.	Dimensions are given in feet.	
	Fields are structured as Year_Day_AwayTeam_HomeTeam.	Play ids generally include at least three events: the pitch, what happens to the ball (often a catch of the pitch by the catcher), and the end of the play.	Timestamps start at the beginning of each game, with the time of the initial play beginning at t < 60 seconds.	Each player position corresponds to a column given in Game Info, with the column names connected through the Glossary. That information corresponds to player ids at each position.	x = 0 is defined at the line between home plate and second base, with x > 0 towards first base.	y = 0 is at the back of home plate, with y > 0 towards second base.	

TABLE:
game_info (p. 1)



VARIABLES	game_str	home_team	away_team	at_bat	play_per_game	inning	top_bottom_inning
OTHER TABLES	game_events			game_events	game_events		
	ball_pos						
	player_pos						
		team_info	team_info				
DESCRIPTION	Each game string identifies the year and the sequential day within the season.			An at-bat corresponds to a batter at the plate. They are listed consecutively for each game.	Plays-per-game begin at 1 and are listed consecutively for each game.	Innings begin at 1 and correspond to both the home and away team getting up to bat.	The away team bats in the top of the inning. The home team bats in the bottom.
	Fields are structured as Year_Day_AwayTeam_HomeTeam.			An at-bat includes at least one play id and one play-per-game.	Some plays (e.g. pickoff throws) occur during an at-bat, but are not associated with the at-bat.	Typical games have 8.5-9 innings (depending on whether the home team is ahead), though they can go from 5 (shortened due to weather) to >9 (extra innings).	

SPRAY 25°

LAUNCH ANGLE 0.59°

EXIT VELOCITY 104.9 MPH

2023
SMT
DATA CHALLENGE

HOSTED BY Carnegie Mellon University Statistics & Data Science

[illegible]

TABLE:
game_info (p. 3)



VARIABLES	left_field	center_field	right_field	batter	first_baserunner	second_baserunner	third_baserunner
OTHER TABLES	(game_events)	(game_events)	(game_events)	(game_events)	(game_events)	(game_events)	(game_events)
	(player_pos)	(player_pos)	(player_pos)	(player_pos)	(player_pos)	(player_pos)	(player_pos)
	team_info	team_info	team_info	team_info	team_info	team_info	team_info
DESCRIPTION	The left fielder corresponds to player_position = 7.	The centerfielder corresponds to player_position = 8.	The right fielder corresponds to player_position = 9.	The batter corresponds to player_position = 10.	The runner on first base corresponds to player_position = 11.	The runner on second base corresponds to player_position = 12.	The runner on third base corresponds to player_position = 13.
	The number corresponds to a player id given in team info.	The number corresponds to a player id given in team info.	The number corresponds to a player id given in team info.	The number corresponds to a player id given in team info.	The number corresponds to a player id given in team info.	The number corresponds to a player id given in team info.	The number corresponds to a player id given in team info.
	This player is on the home team in the top of inning and the away team in the bottom.	This player is on the home team in the top of inning and the away team in the bottom.	This player is on the home team in the top of inning and the away team in the bottom.	This player is on the away team in the top of inning and the home team in the bottom.	This player is on the away team in the top of inning and the home team in the bottom.	This player is on the away team in the top of inning and the home team in the bottom.	This player is on the away team in the top of inning and the home team in the bottom.
					This field is not necessarily populated for each play per game.	This field is not necessarily populated for each play per game.	This field is not necessarily populated for each play per game.

TABLE:
team_info



VARIABLES	team_id	player_id	team_year	
OTHER TABLES				
	game_info	game_info		
DESCRIPTION	TeamA# and TeamB are always home teams. They all appear over multiple years.	Player ids for home teams (TeamA#, TeamB) can repeat within the data set, carrying over team years or between teams.	Games occur during years 1900-1903.	
	Teams with two-letter designations are always away teams. They are unique to each year.	Player ids for away teams are unique to each year. If a player has been traded between away teams in a given year (highly unlikely), the same player id may appear on two teams.	Data for a given home team covers multiple years. TeamA1: [1902, 1903] TeamA2: [1902, 1903] TeamA3: [1901, 1902, 1903] TeamB: [1900, 1901, 1902, 1903]	
	Teams A are part of the same organization, such that a given player_id may appear on more than one team if the player is promoted or sent down. The levels correspond to: TeamA1: Low TeamA2: Middle TeamA3: High	The data set includes 1342 player ids (583 home, 759 away).	Every home team plays three 2-3 game series each year. (The data set includes 97 games.)	



GLOSSARY

VARIABLE	TABLE	CODE	DEFINITION
player_position	game_events player_pos (game_info)	1	pitcher
		2	catcher
		3	first baseman
		4	second baseman
		5	third baseman
		6	shortstop
		7	left field
		8	center field
		9	right field
		10	batter
		11	runner on first base
		12	runner on second base
		13	runner on third base
		255	ball event with no player (e.g., ball bounce)
event_code	game_events	1	pitch
		2	ball acquired
		3	throw (ball-in-play)
		4	ball hit into play
		5	end of play
		6	pickoff throw
		7	ball acquired - unknown field position
		8	throw (ball-in-play) - unknown field position
		9	ball deflection
		10	ball deflection off of wall
		11	home run
		16	ball bounce