

## BIP\_dataset user notes

The .xlsx file contains season data for “balls in play,” derived from Retrosheet, and various other performance elements for every AL/NL pitcher from 1912-2014. Variables include:

Variable	Description	Variable	Description
playerID	Lahman player id	nameFirst	first name
yearID	Lahman season id	nameLast	last name
teamID	Lahman team id	ab	total batters faced
lgID	Lahman league id	SH	sacrifice bunts
bbrefID	BBR player id	SF	sacrifice flies
W	wins	babip	babip
L	losses	kb	knuckelball at least sometimes
G	games appeared	kb_plus	primarily knuckleball sometime career
GS	games started	hr_pip	hr/IP
CG	complete games	k_pip	K/IP
SHO	shutouts	bb_pip	bb/IP
SV	saves	hbp_pip	hbp/IP
H	hits allowed	rapg	runs allowed per game (r/IP*9)
ER	earned runs allowed	fipr	FIP regression
HR	home runs allowed	period	BIP study season bin
BB	walks allowed	teamid	Lahman team id
SO	strike outs	field_composite	team fielding composite (see paper)
WP	wild pitches	ifgb	proportion BIP infield GB
HBP	hit batters	ofgb	proportion BIP outfield GB
R	runs allowed	iffb	proportion BIP infield FB
IPouts	total inning outs	offb	proportion BIP outfield FB
abf	total batters faced	ifld	proportion BIP infield LD
IP	innings pitched (IPouts/3)	ofld	proportion BIP outfield LD
retroID	retrosheet iid	ifpop	proportion BIP infield popup
total_bip	total number balls in play	ofpop	proportion BIP infield popup
		bip_ip	BIP/IP

These data were used in the models featured in [“Empirically Testing the Impact of Pitchers on ‘Balls Hit in Play.’”](#) They also form the basis for the “runs saved” lists distributed along with these data.