Baseball Project

Team JBT

CS 122

Project Presentation

Goal

Create an interface that allows the user to query and manipulate baseball statistics in the following ways:

- Search for games with specified characteristics such as number of runs, teams playing, date/year, records
- Create fantasy teams assembled out of mixed and matched players from different eras and compare expected win totals based on wins above replacement (WAR) statistics
- Compare side by side two players or teams based on a number of statistics

Sources of Data

- **Baseball Reference**: The main source of baseball statistics. Contains stats for both players and games.
- **FanGraphs:** Contains baseball stats and also some calculations of interesting statistics, as well as visuals.
- **MLB:** The official website of the MLB with its own player and game statistics.
- **Baseball America and Baseball Savant**: Two more sites with calculations of interesting baseball statistics and graphs.

Baseball Reference

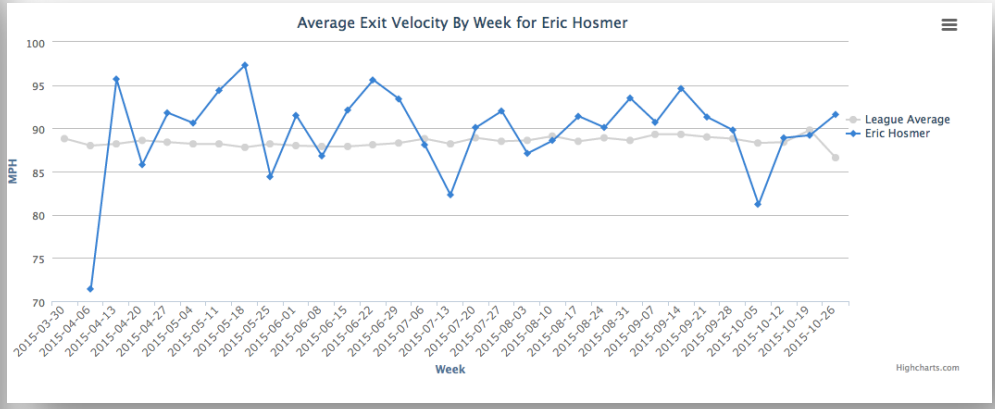
Standard Batting Glossary · Show Minors Stats · SHARE · Embed · CSV · Export · PRE · LINK · ?

Year	Age	Tm	Lg	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	Pos	Awards
1954	20	MLN	NL	122	509	468	58	131	27	6	13	69	2	2	28	39	.280	.322	.447	.769	104	209	13	3	6	4		*79	RoY-4
1955 ★	21	MLN	NL	153	665	602	105	189	**37**	9	27	106	3	1	49	61	.314	.366	.540	.906	141	325	20	3	7	4	5	*974	AS,MVP-9
1956 ★	22	MLN	NL	153	660	609	106	***200***	**34**	14	26	92	2	4	37	54	**.328**	.365	.558	.923	151	**340**	21	2	5	7	6	*9	AS,MVP-3
1957 ★	23	MLN	NL	151	675	615	**118**	198	27	6	***44***	***132***	1	1	57	58	.322	.378	.600	.978	166	***369***	13	0	0	3	15	*98	AS,**MVP-1**
1958 ★	24	MLN	NL	153	664	601	109	196	34	4	30	95	4	1	59	49	.326	.386	.546	.931	152	328	21	1	0	3	16	*98	AS,MVP-3,GG
1959 ★	25	MLN	NL	154	693	629	116	***223***	46	7	39	123	8	0	51	54	***.355***	.401	***.636***	***1.037***	***182***	***400***	19	4	0	9	17	*98/5	AS,MVP-3,GG
1960 ★	26	MLN	NL	153	664	590	102	172	20	11	40	***126***	16	7	60	63	.292	.352	.566	.919	156	***334***	8	2	0	***12***	13	*9/84	AS,MVP-11,GG
1961 ★	27	MLN	NL	**155**	671	603	115	197	**39**	10	34	120	21	9	56	64	.327	.381	.594	.974	163	**358**	16	2	1	9	20	*89/5	AS,MVP-8
1962 ★	28	MLN	NL	156	667	592	127	191	28	6	45	128	15	7	66	73	.323	.390	.618	1.008	170	366	14	3	0	6	14	*89/3	AS,MVP-6
1963 ★	29	MLN	NL	161	714	631	***121***	201	29	4	***44***	***130***	31	5	78	94	.319	.391	***.586***	***.977***	***179***	***370***	11	0	0	5	18	*9	AS,MVP-3

Fan Graphs

#	Name	Team	G	PA	HR	R	RBI	SB
1	Bryce Harper	Nationals	153	654	42	118	99	6
2	Mike Trout	Angels	159	682	41	104	90	11
3	Josh Donaldson	Blue Jays	158	711	41	122	123	6
4	Paul Goldschmidt	Diamondbacks	159	695	33	103	110	21
5	Joey Votto	Reds	158	695	29	95	80	11
6	Manny Machado	Orioles	162	713	35	102	86	20
7	Yoenis Cespedes	- - -	159	676	35	101	105	7
8	A.J. Pollock	Diamondbacks	157	673	20	111	76	39
9	Lorenzo Cain	Royals	140	604	16	101	72	28

Baseball Savant



Nearest Neighbor Search

- Finds closest and most similar points--closeness typically expressed in terms of a dissimilarity function: the less similar the objects, the larger the function values. This algorithm could be used to compare similar players.

Marcel Algorithm

- Developed by Tom Tango, Marcel examines a player's performance over the past three seasons, weighting more recent seasons more heavily, while regressing performance toward the league mean and applying an age adjustment to generate a season forecast, both of those at equal rates per player. Major League Equivalencies (MLEs) -- translations of a player's minor league stats into major league terms -- aren't included.

Runs-to-Wins

- Pythagorean Formula:
  - W-L% = (RS^x) / (RS^x + RA^x)
  - Where x = (runs/gm)^.285, RS = runs scored, RA = runs attempted

Timeline

- Write code to scrape data to create database
- Write SQL code to search through database and determine calculations necessary
- Write code to create fantasy teams based on WAR stats found in the database
- Determine how to compare two players and write code to extract these statistics
- Create website interface and design organization for each of the three tasks.
- Finalize project work, test through querying