

# Analysis on MLB teams

Haorui Zhang | May 2019



# I. Obtaining the Data

- Use **rvest** package in R.
- For each of the 30 **Active Franchises**, download the “**Franchise History**” table on each team’s page and combine all of the tables into one.
- The resulted dataset has **2684 rows** and **22 columns**. Variables include wins, losses, ties, runs scored, runs allowed, batter’s average age, pitcher’s average age, etc.

MLB Team History

[Back to top](#) ▲

## MLB Teams and Baseball Encyclopedia

### Active Franchises

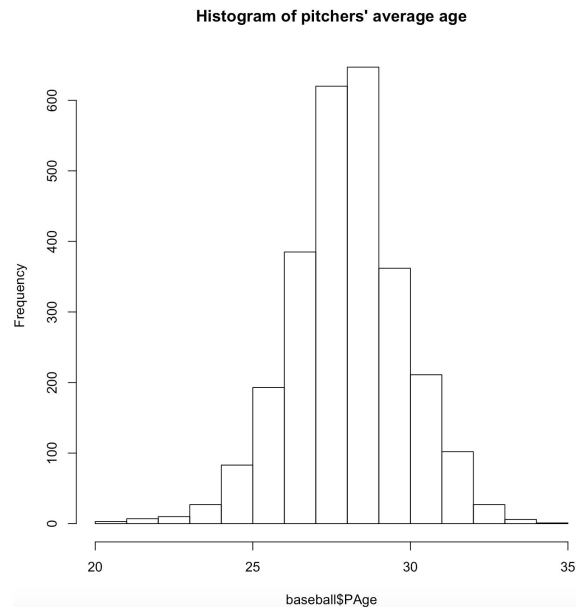
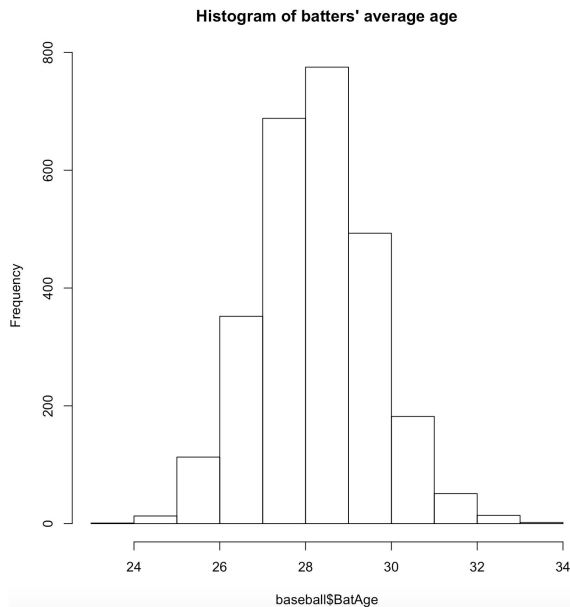
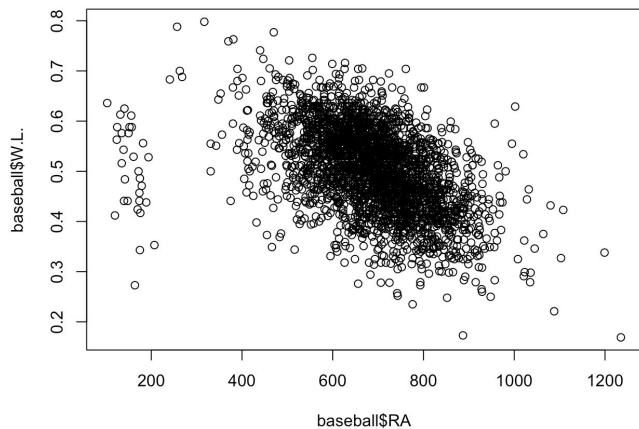
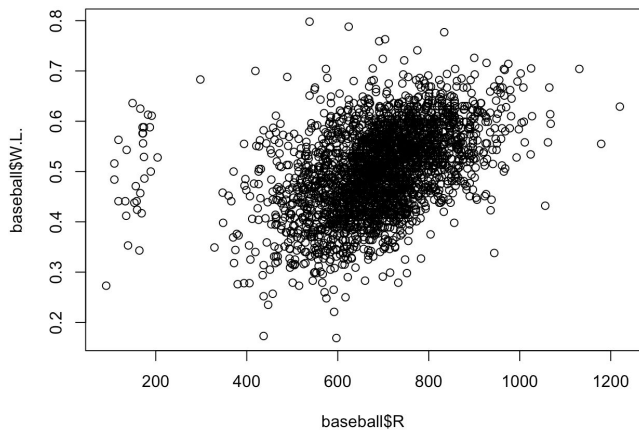
Share & more ▼

[Glossary](#)

[Hide Partial Rows](#)

Rk	Franchise	From	To	G	W	L	W-L%	G>.500	Divs	Pnnts	WS	Playoffs	Players	HOF#	R	AB	H	HR	BA	RA	ERA
Anaheim Angels, see <a href="#">Los Angeles Angels</a>																					
1	<a href="#">Arizona Diamondbacks</a>	1998	2019	3,436	1,698	1,738	.494	-40	5	1	1	6	480	2	15,597	117,565	30,229	3,625	.257	15,727	4.21
2	<a href="#">Atlanta Braves</a>	1876	2019	21,382	10,618	10,610	.500	8	18	17	3	24	2004	52	95,066	732,261	190,330	13,696	.260	94,354	3.66
Also played as Milwaukee Braves, Boston Braves, Boston Bees, Boston Rustlers, Boston Doves, Boston Beaneaters and Boston Red Stockings																					
3	<a href="#">Baltimore Orioles</a>	1901	2019	18,478	8,726	9,642	.475	-916	9	7	3	14	1844	34	79,631	627,737	162,362	13,341	.259	84,711	4.03
Also played as St. Louis Browns and Milwaukee Brewers																					
Boston Braves, see <a href="#">Atlanta Braves</a>																					
4	<a href="#">Boston Red Sox</a>	1901	2019	18,466	9,535	8,848	.519	687	10	14	9	24	1801	37	86,036	631,729	168,781	13,509	.267	82,495	3.90
Also played as Boston Americans																					
Brooklyn Dodgers, see <a href="#">Los Angeles Dodgers</a>																					
California Angels, see <a href="#">Los Angeles Angels</a>																					
5	<a href="#">Chicago Cubs</a>	1876	2019	21,416	10,917	10,338	.514	579	7	17	3	20	2078	45	98,617	733,581	192,671	14,168	.263	95,118	3.68
Also played as Chicago Orphans, Chicago Colts and Chicago White Stockings																					
6	<a href="#">Chicago White Sox</a>	1901	2019	18,472	9,225	9,144	.502	81	5	6	3	9	1774	32	80,431	625,580	162,840	11,369	.260	80,122	3.78
7	<a href="#">Cincinnati Reds</a>	1882	2019	21,003	10,538	10,326	.505	212	10	10	5	15	2003	36	94,274	716,331	187,208	13,256	.261	93,339	3.74
Also played as Cincinnati Redlegs and Cincinnati Red Stockings																					
8	<a href="#">Cleveland Indians</a>	1901	2019	18,475	9,402	8,982	.511	420	10	6	2	14	1887	34	83,647	630,220	167,402	12,976	.266	81,672	3.83

# II. Exploratory Data Analysis



# III. Proposal

- Fit a **multiple linear regression** model to **predict the win rate** of these MLB teams. Potential predictors include: runs scored, runs allowed, batters' average age, pitchers' average age, team's finish, and some other newly created variables.
- **“Product”**: Provide baseball teams with a reference to see what factors would contribute to their success in a certain game so that they would know how to prepare better for future games.