

A close-up photograph of a white baseball with red stitching, resting on a light-colored wooden surface. The background is blurred, showing a green field.

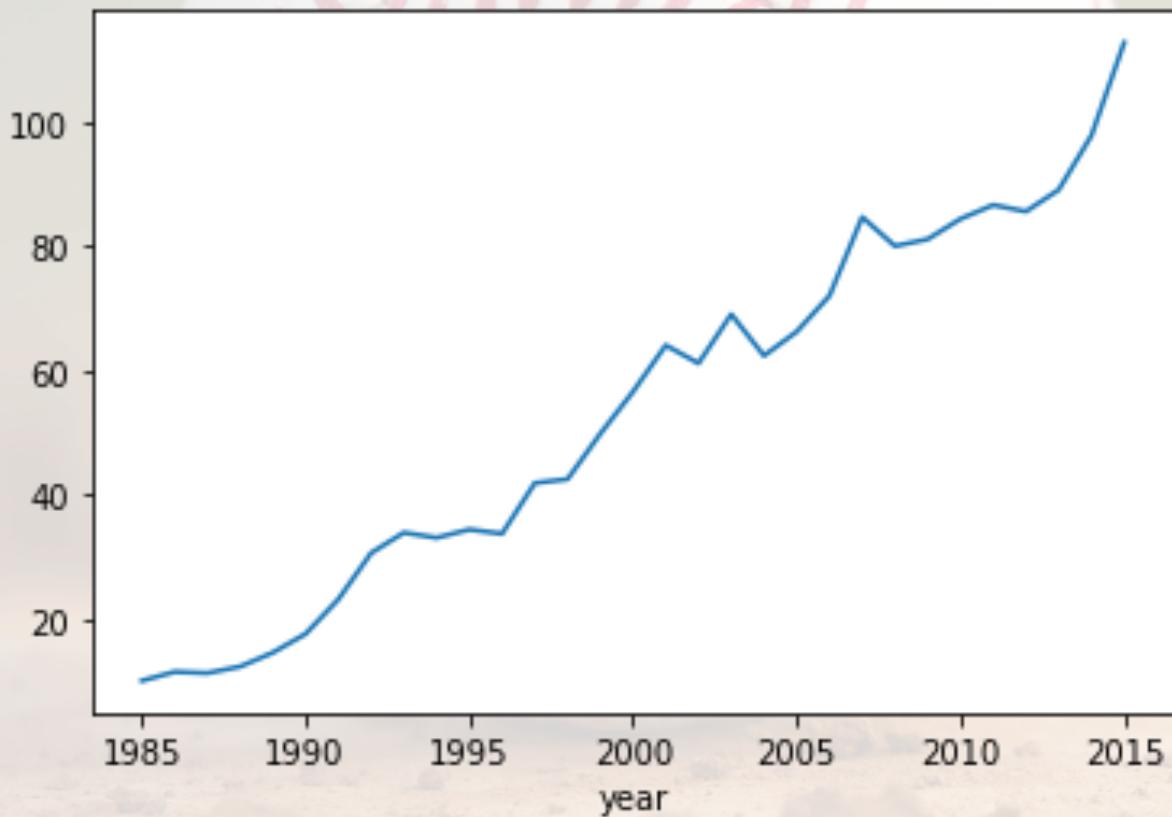
Building a Better Baseball Team

The Problem

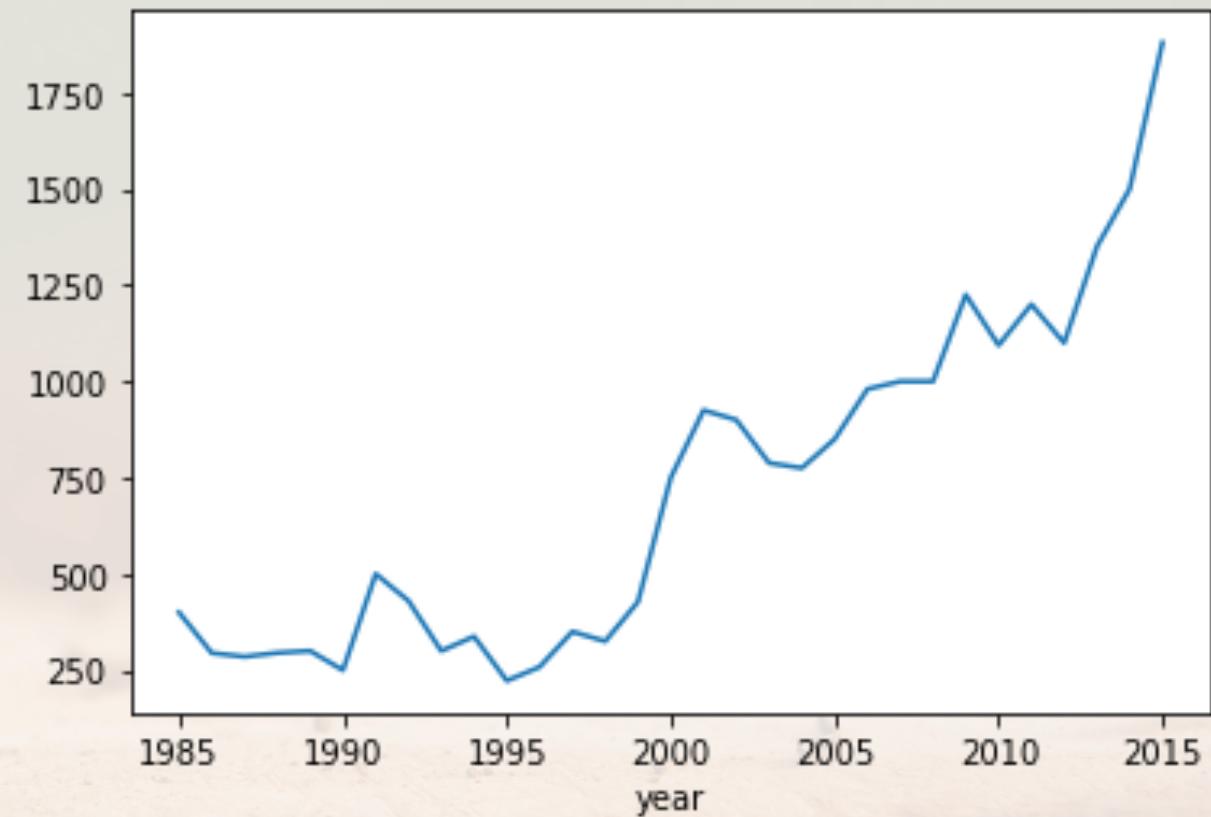
How does an MLB team owner minimize the amount of money spent on player salaries while still winning at least 81 games?

Growth in Salaries over Time

Median Log Team Salaries (in millions)
1985-2015

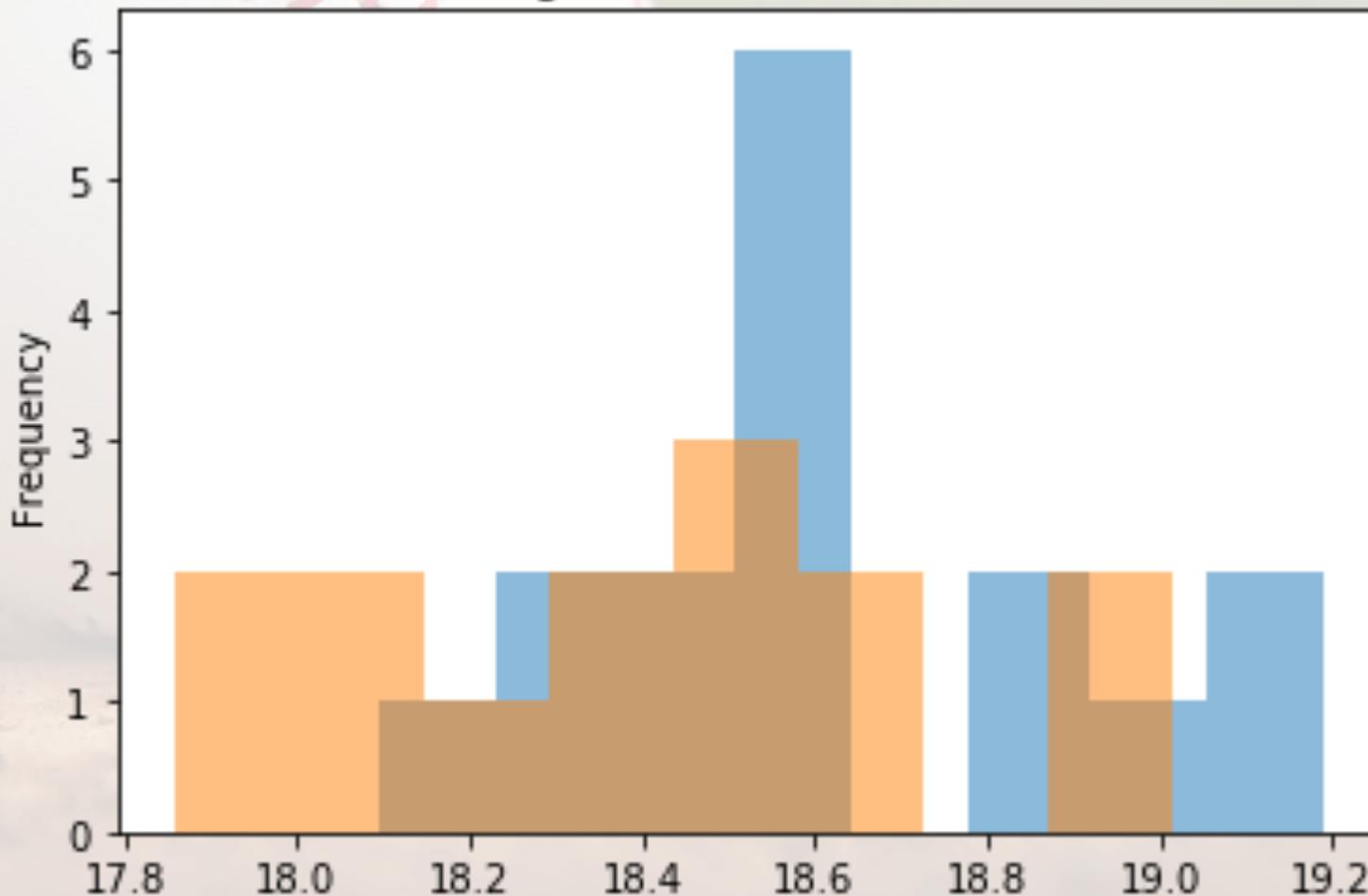


Median Individual Salaries (in thousands)
1985-2015



Distribution of Total Team Salaries in 2015

Log team salaries 2015



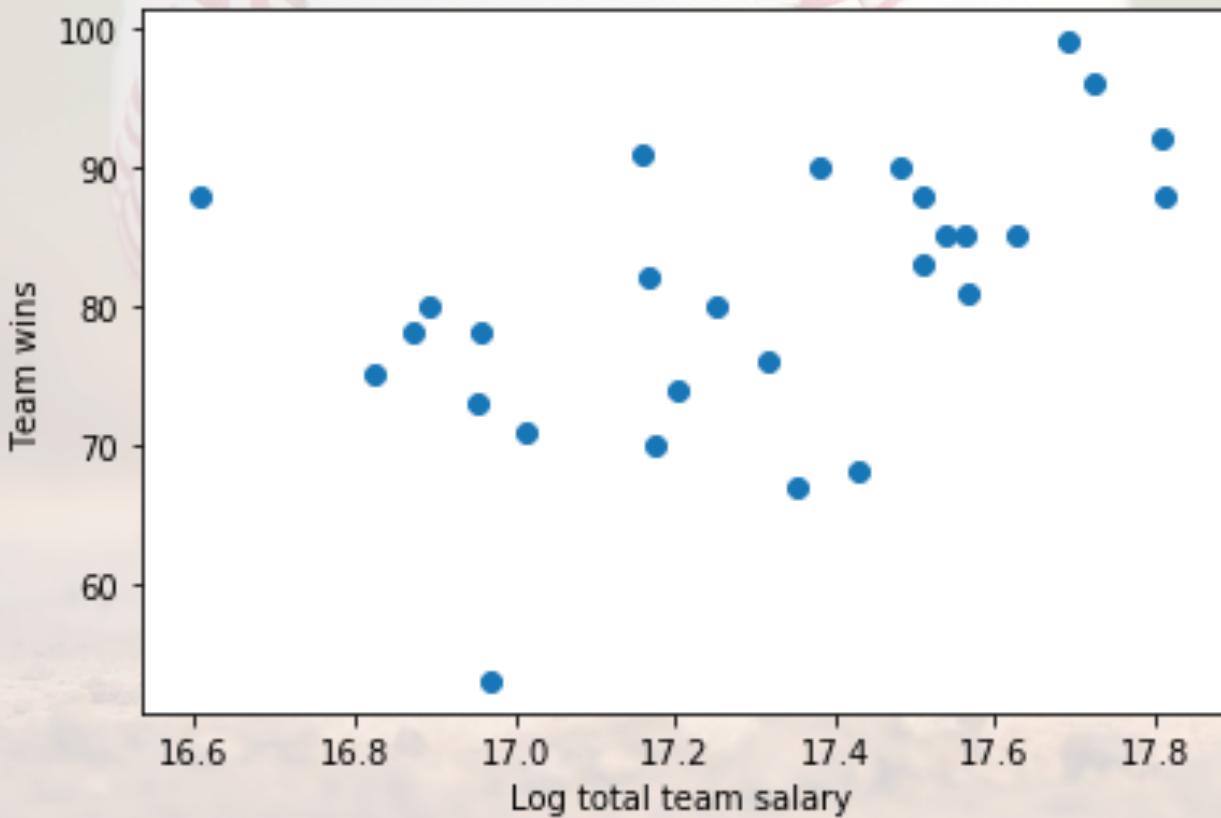
Impact of Salary on Number of Wins

How does relative salary impact wins?
1985-2015

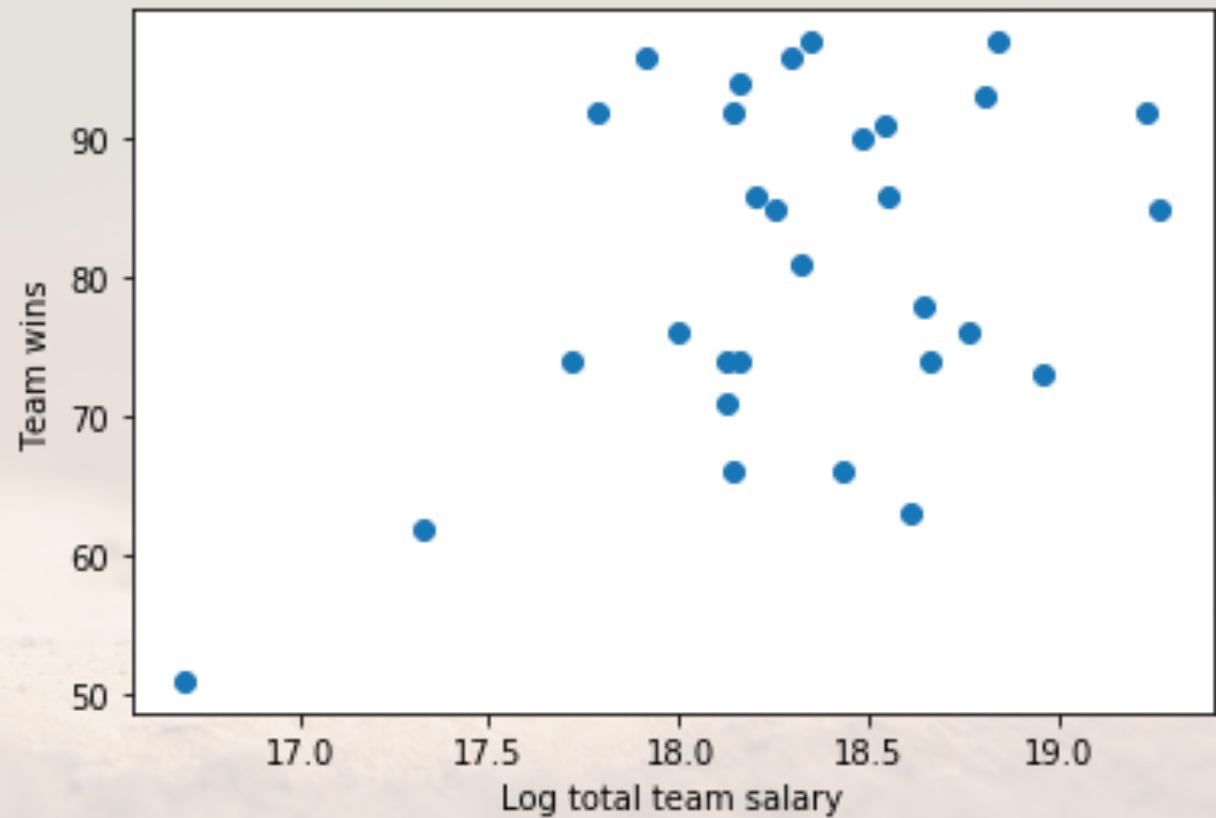


Impact of Salary on Number of Wins

How does salary impact wins?
1996 season



How does salary impact wins?
2013 season



What stats are important for winning?

Results of our decision tree model:

```
('r_avg', 0.39529150528029333)
('avg', 0.022069976424698174)
('bb_avg', 0.0066931475486664456)
('so_avg', 0.10790911123162597)
('era_adj', 0.45704845784772291)
('e_avg', 0.010987801666993107)
```

What stats are important for winning?

Results of our decision tree model:

```
('r_avg', 0.39529150528029333)
('avg', 0.022069976424698174)
('bb_avg', 0.0066931475486664456)
('so_avg', 0.10790911123162597)
('era_adj', 0.45704845784772291)
('e_avg', 0.010987801666993107)
```

Scoring runs and preventing runs are
the most important.

What stats predict scoring runs?

Results of our decision tree models:

```
('avg', 0.0)
('obp', 0.71129493684638967)
('slug', 0.28870506315361028)
```

```
('1b_avg', 0.12174123628549635)
('2b_avg', 0.073495933626006288)
('3b_avg', 0.16955756378379142)
('hr_avg', 0.42955547857826454)
('bb_avg', 0.1933259634922658)
('hbp_avg', 0.012323824234175764)
```

What stats predict scoring runs?

Results of our decision tree models:

```
('avg', 0.0)
('obp', 0.71129493684638967)
('slug', 0.28870506315361028)
```

```
('1b_avg', 0.12174123628549635)
('2b_avg', 0.073495933626006288)
('3b_avg', 0.16955756378379142)
('hr_avg', 0.42955547857826454)
('bb_avg', 0.1933259634922658)
('hbp_avg', 0.012323824234175764)
```

Homeruns, triples, and walks are most predictive of scoring runs.

Player Stats and Salary

OLS Regression Results

Dep. Variable:	salary	R-squared:	0.490			
Model:	OLS	Adj. R-squared:	0.481			
<hr/>						
	coef	std err	t	P> t	[0.025	0.975]
const	6.3813	0.591	10.801	0.000	5.219	7.543
1b_avg	8.2406	1.631	5.052	0.000	5.032	11.449
2b_avg	7.8214	3.750	2.086	0.038	0.444	15.199
3b_avg	-25.8545	9.226	-2.802	0.005	-44.003	-7.706
hr_avg	19.2107	3.523	5.453	0.000	12.281	26.141
bb_avg	2.0800	1.347	1.544	0.124	-0.571	4.731
age	0.2031	0.014	14.500	0.000	0.176	0.231

Triples and walks are the best indicators of undervalued players.

Conclusion

- Focus on hiring players who know how to get on base, reflected in the ability to hit singles and triples, and draw walks.
- Look for a few inexpensive players who are adept at hitting triples.
- Resist the urge to hire big homerun hitters because, while homeruns are a sure way to score runs, homeruns are expensive. Homerun hitters are flashier and may make a team more exciting to watch, but they also demand higher salaries.
- Finally, hire younger but promising players, because older players tend to be paid more money.