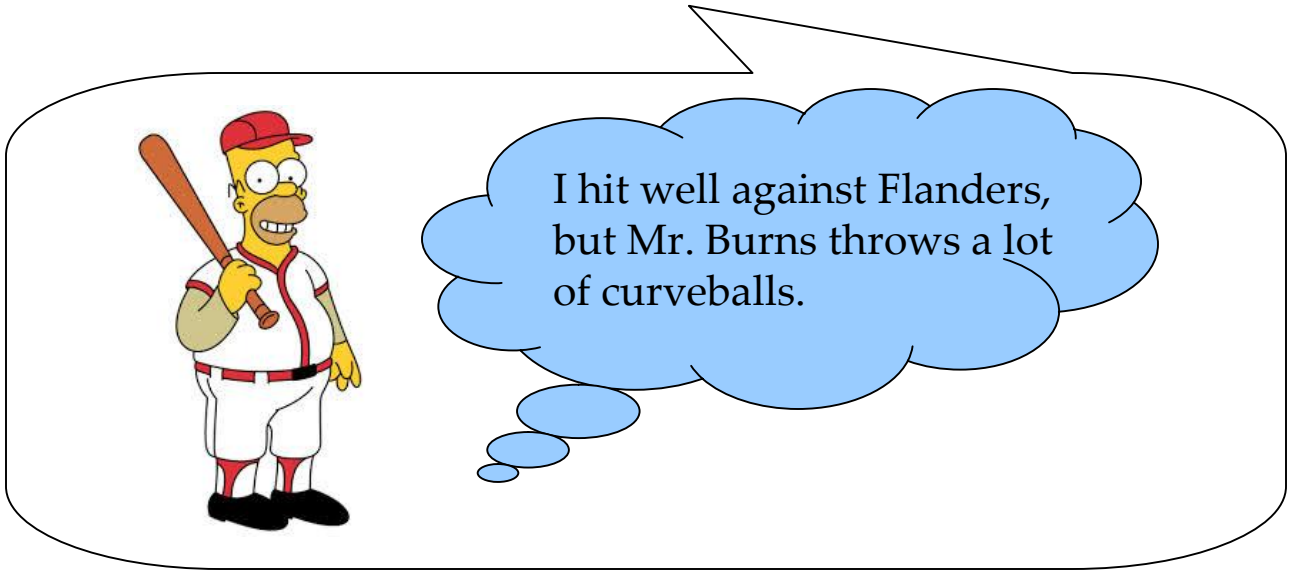
Who's On First

Shanshan Ding
Insight Data Science

Fantasy Baseball

March	April to September	October
Get together with friends to each draft a team.	Everyday, decide who to play and who to bench.	Overall winner gets prize or bragging rights.



The Small Sample Size Challenge

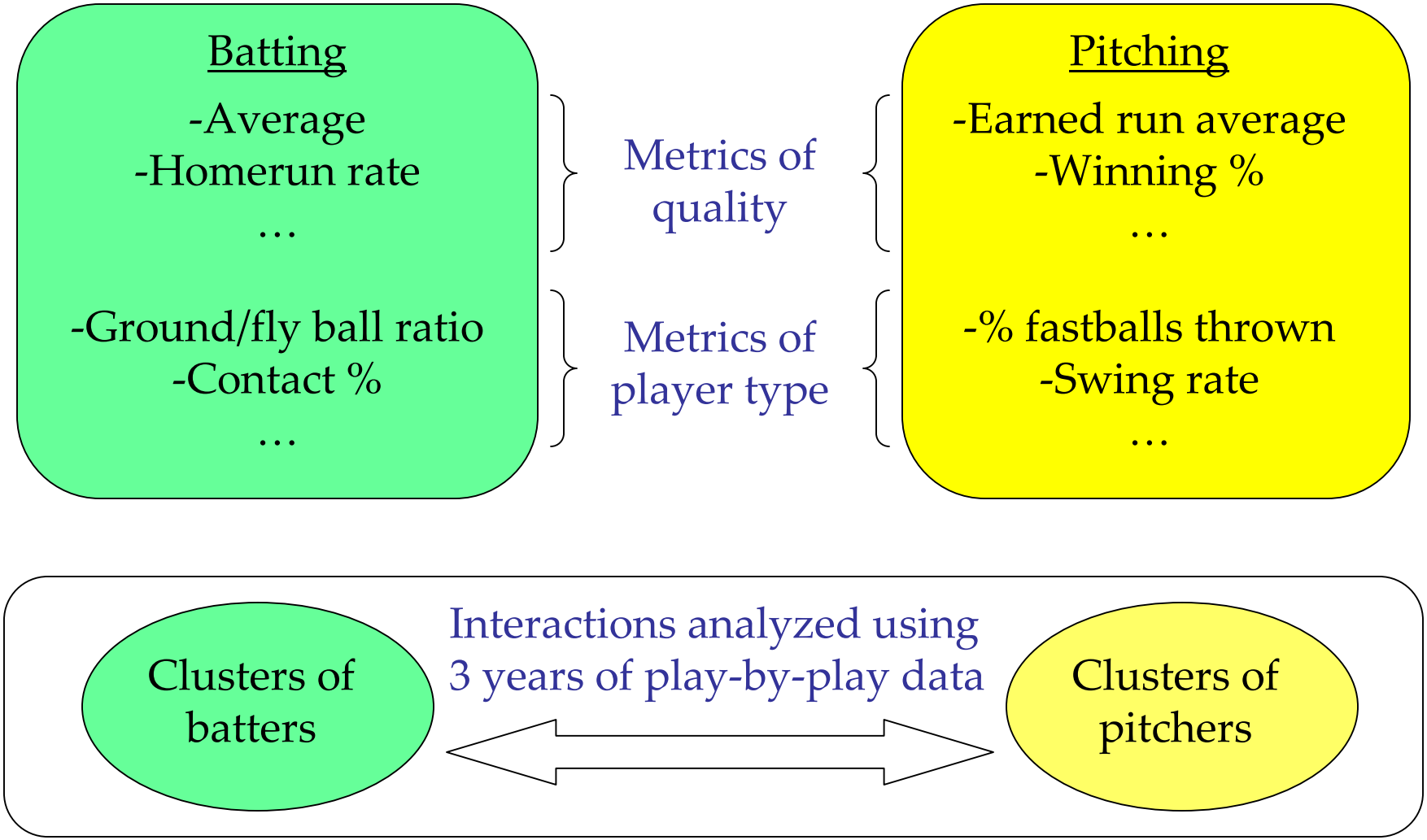
ROTOWIRE
YOUR PREMIUM SOURCE FOR FANTASY SPORTS

Traditional predictions: simply look up the batter-pitcher matchup history.

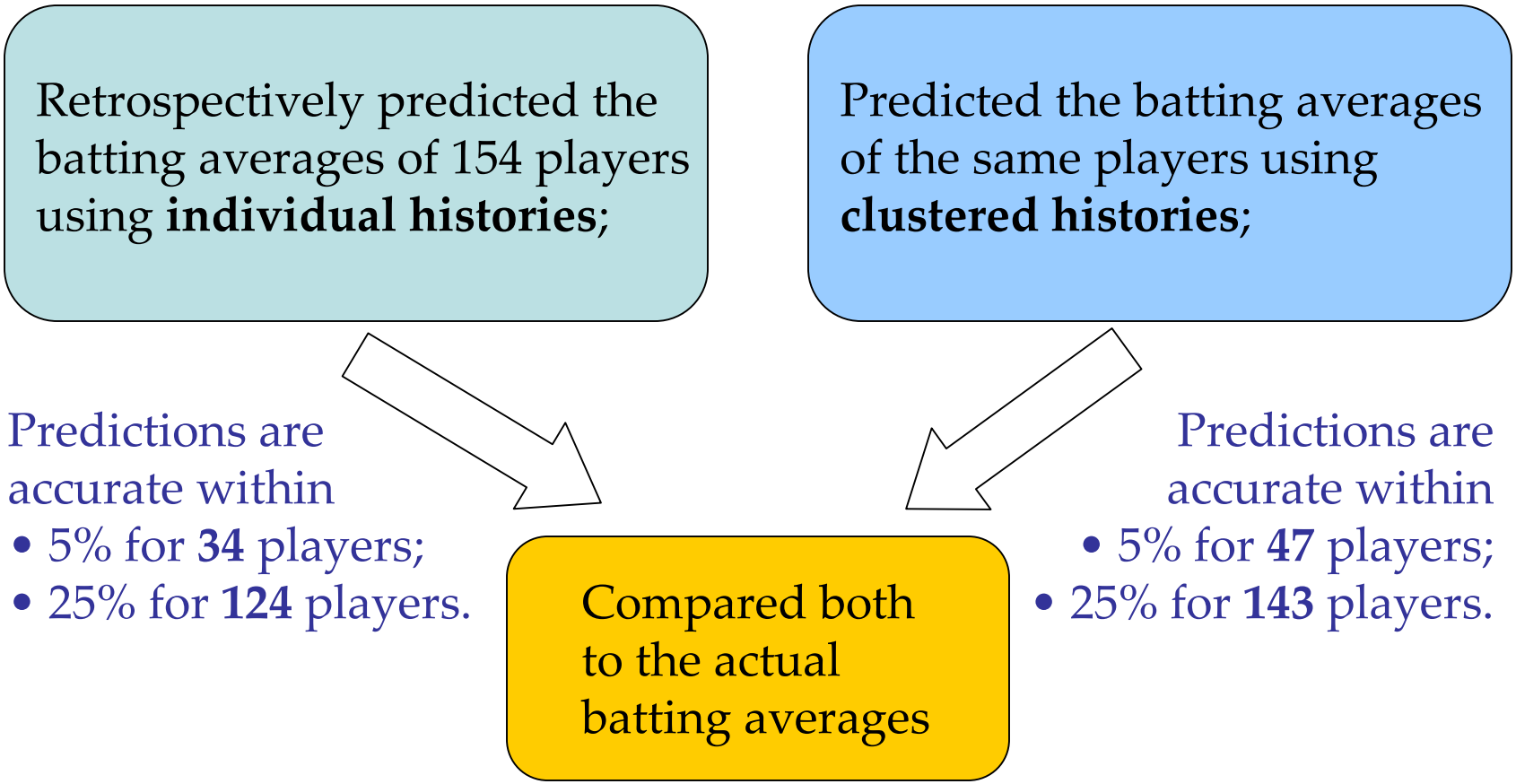
Hot Batter Matchups			At-Bats	AVG
Quintin Berry - OF	BOS vs NYY 9/14 1:05 PM	CC Sabathia	2-4	.500
Stephen Drew - SS	BOS vs NYY 9/14 1:05 PM	CC Sabathia	2-3	.667

- 37% of matchups involve players who have never faced each other before;
- Another 32% have histories of 5 or less at-bats.

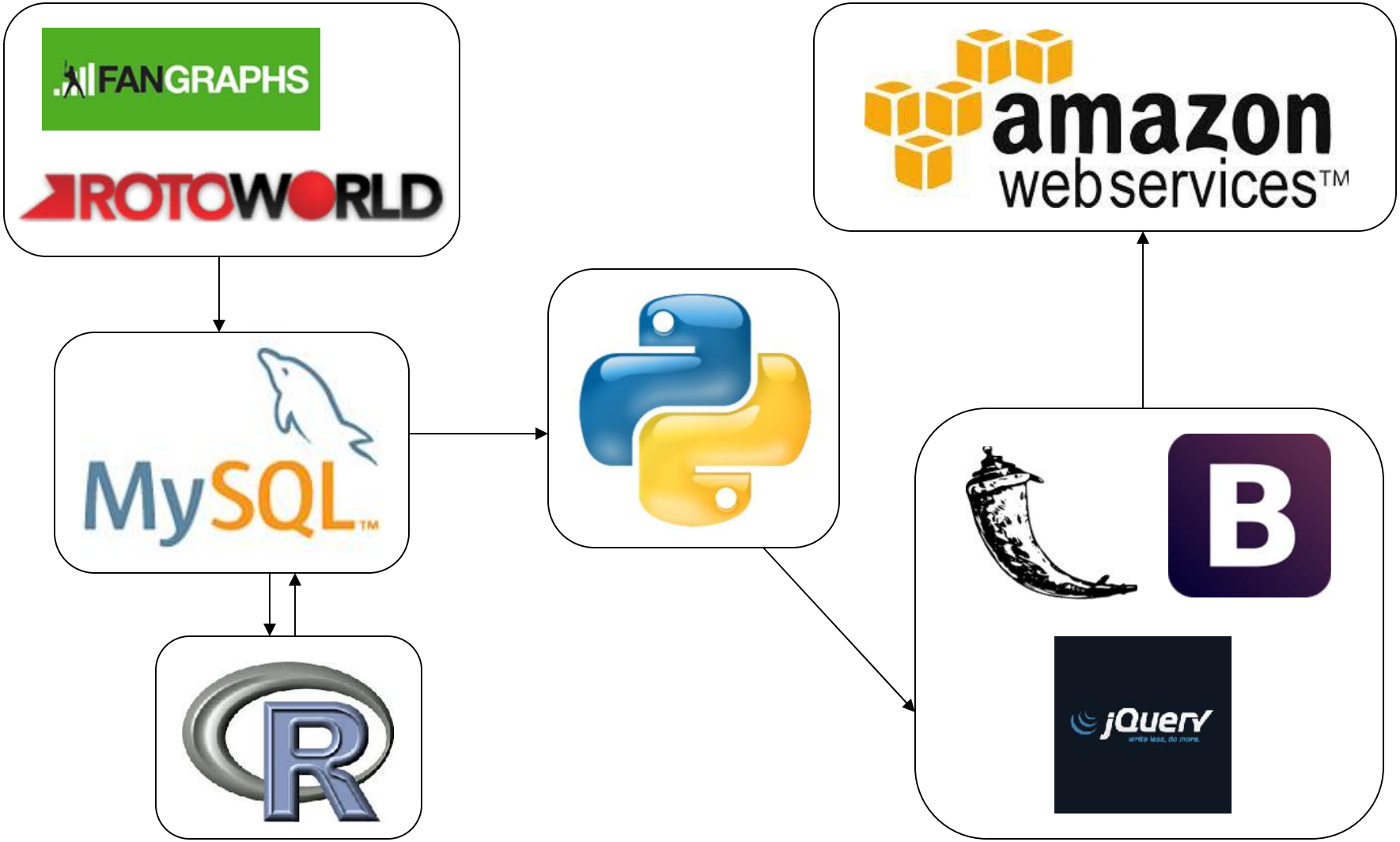
K-Means Clustering



Validation


Retrospectively predicted the batting averages of 154 players using **individual histories**;
Predicted the batting averages of the same players using **clustered histories**;
Predictions are accurate within
• 5% for **34** players;
• 25% for **124** players.
Compared both to the actual batting averages
Predictions are accurate within
• 5% for **47** players;
• 25% for **143** players.

Tools Used


FANGRAPHS
ROTOWORLD
amazon
web services™
MySQL™
B
jQuery
write less, do more.

Shanshan Ding

$$\mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_t \in A|X_s)$$