

# OpenWAR:

## An Open Source System for Overall Player Performance in MLB

Ben Baumer<sup>1</sup>    Shane Jensen<sup>2</sup>    Gregory J. Matthews<sup>3</sup>

<sup>1</sup> Smith College

<sup>2</sup> The Wharton School  
University of Pennsylvania

<sup>3</sup> Loyola University Chicago

Loyola Academy  
April 11, 2018



**LOYOLA**  
UNIVERSITY CHICAGO



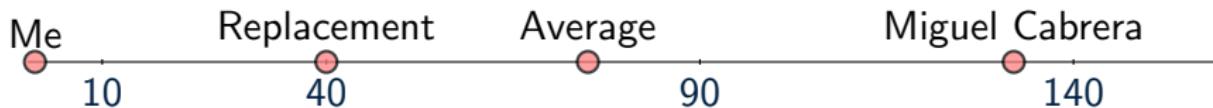
# WAR - What is it good for?

- Wins Above Replacement
- Question: How large is the contribution that each player makes towards winning?
- Four Components:
  - ① Batting
  - ② Baserunning
  - ③ Fielding
  - ④ Pitching
- Replacement Player: Hypothetical 4A journeyman
  - ▶ Much worse than an average player

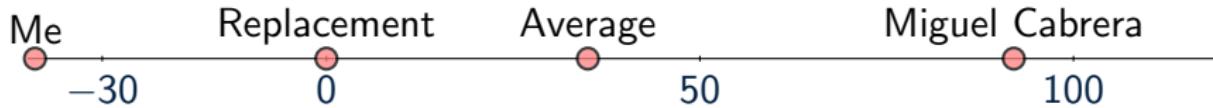


# Units and Scaling

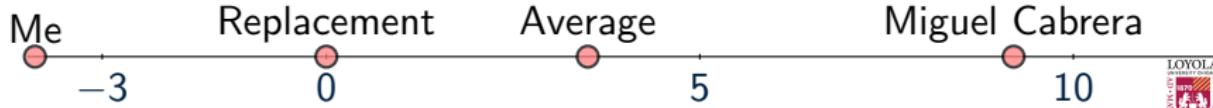
- In terms of **absolute runs**:



- In terms of **Runs Above Replacement (RAR)**:



- In terms of **Wins Above Replacement (WAR)**:



## Example: 2017 WAR leaders

FanGraphs	fWAR	BB-Ref	rWAR
Aaron Judge	8.2	Jose Altuve	8.3
Joe Altuve	7.6	Corey Kluber	8.2
Giancarlo Stanton	7.1	Aaron Judge	8.1
Mike Trout	6.9	Giancarlo Stanton	7.6
Anthony Rendon	6.8	Joey Votto	7.5
Jose Ramirez	6.7	Max Scherzer	7.4
Kris Bryant	6.6	Nolan Arenado	7.2
Joey Votto	6.6	Andrelton Simmons	7.1

Table: 2017 WAR Leaders

- Baseball Prospectus also publishes *WARP*
- There is no ONE formula for WAR!



# How much is a win worth?



**SITW** @StatsInTheWild

31m

@harrypav How much did you say one MLB win was worth during your talk?

[Details](#)



**Harry Pavlidis**

@harrypav

@StatsInTheWild "it depends"  
5, 10, 20 million. A lot.



# WAR is the Answer

**MICHAEL JORDAN AT 50** by WRIGHT THOMPSON  
"Can I find peace away from basketball?" PAGE 36

THE ANALYTICS ISSUE

WE DID THE MATH  
ANDREW LUCK HAS BETTER LEGS THAN RG3,  
STEPHEN CURRY IS THE NBA'S GREATEST SHOOTER EVER AND THE ORIOLES WILL FINISH LAST. PLUS THE GAMBLING MACHINE THAT'S BEATING VEGAS.

MARCH 4, 2013 \$5.99

THIS IS WAR  
HOW ONE GEEKED-OUT STAT IGNITED BASEBALL'S BIGGEST DEBATE

American League MVP MIKE TROUT andathletic sensation MIKE TROUT photographed in New York City, Jan. 19

# WAR in action

*Over the past four years, Mr. Zobrist has led baseball in WAR, ahead of stars like Albert Pujols, Ryan Braun and Robinson Cano. – The New York Times, 4/1/2013*

*According to baseball-reference.com, Trout's WAR was 10.7, far above Cabrera's 6.9. – Sports Illustrated, 11/15/2012*

*No player in history has been a plus-10 runner and a plus-20 fielder and a plus-30 hitter. – ESPN The Magazine, 3/4/2013*



# What's Wrong with WAR?

- Not Reproducible
  - ▶ WAR is an unknown hypothetical quantity – not a **statistic**
  - ▶ No reference implementation of WAR
  - ▶ No open data set
  - ▶ No open source code
- No unified methodology
  - ▶ Each component of WAR is viewed as a separate problem – not a piece of the same problem
  - ▶ Ad hoc definitions: what is replacement level?
- No error estimates
  - ▶ Only reported as **point estimates**
  - ▶ Only hand-wavy estimates of variability or margin or error
- Bug or Feature?: Competing black-box implementations



# Our Contribution: *openWAR*

- *openWAR*: a reproducible reference implementation of WAR
  - ▶ Principled **estimate** of WAR
  - ▶ Fully open-source R package (free as in freedom)
  - ▶ Partially open data (free as in beer)
- Unified Methodology:
  - ▶ Conservation of Runs
  - ▶ Each component is estimated as a piece of the larger problem
- Error estimates:
  - ▶ Use resampling methods to report WAR **interval** estimates
- Version 0.1: Emphasis at this stage on **reproducibility**



# openWAR

- R package to be submitted to CRAN
- Currently available for download on GitHub

<https://github.com/beanumber/openWAR>

(<https://github.com/beanumber/openWARDATA>)

- Scrapes XML files from MLBAM GameDay server
- Processes using XSLT and compiles detailed play-by-play info into a data frame
- Computes *openWAR*
- Diagnostic and visualization tools



# Installing the Package

```
install.packages("xslt")
devtools::install_github("beanumber/openWAR")
#devtools::install_github("znmeb/openWAR")
devtools::install_github("beanumber/openWARData")
```



# Getting Data

```
library(openWAR)
ds = getData(start = "2013-06-24")
dim(ds)
head(ds$description)
```



# Getting Data

```
library(openWARData)
dim(MLBAM2017)

## [1] 185704      62

head(MLBAM2017$description)

## [1] "Kyle Schwarber singles on a line drive to right fielder Stephen Piscotty."
## [2] "Kris Bryant strikes out swinging.  "
## [3] "Anthony Rizzo singles on a line drive to right fielder Stephen Piscotty."
## [4] "Ben Zobrist grounds into a double play, second baseman Jedd Gyorko to short"
## [5] "Dexter Fowler lines out to center fielder Jason Heyward.  "
## [6] "Aledmys Diaz doubles (1) on a sharp line drive to right fielder Ben Zobrist"
```



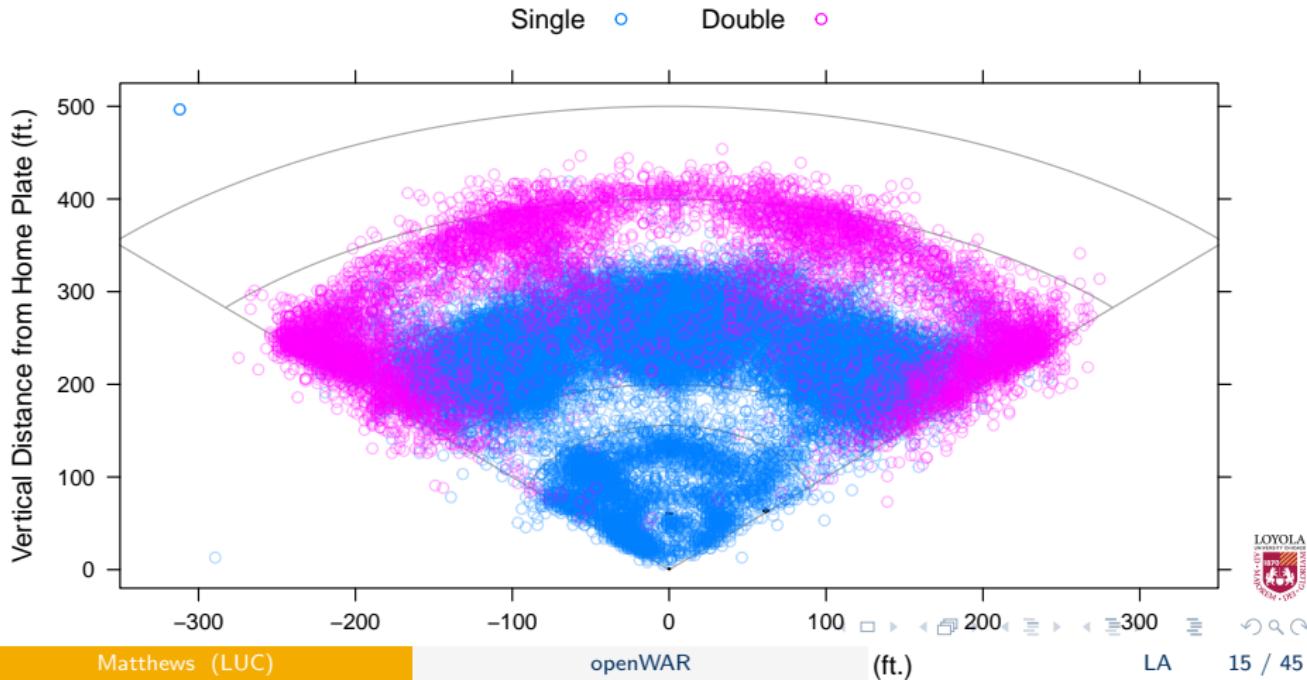
# Data Contents

- Each row represents a plate appearance
- Complete description of the game state (i.e. balls, strikes, ballpark, outs, etc.)
- Baserunner movement (before and during) is captured
- Hit location given as  $(x, y)$ -coordinate
- Fielder and baserunner identities
- PITCHf/x not currently captured, but could be added in a future release



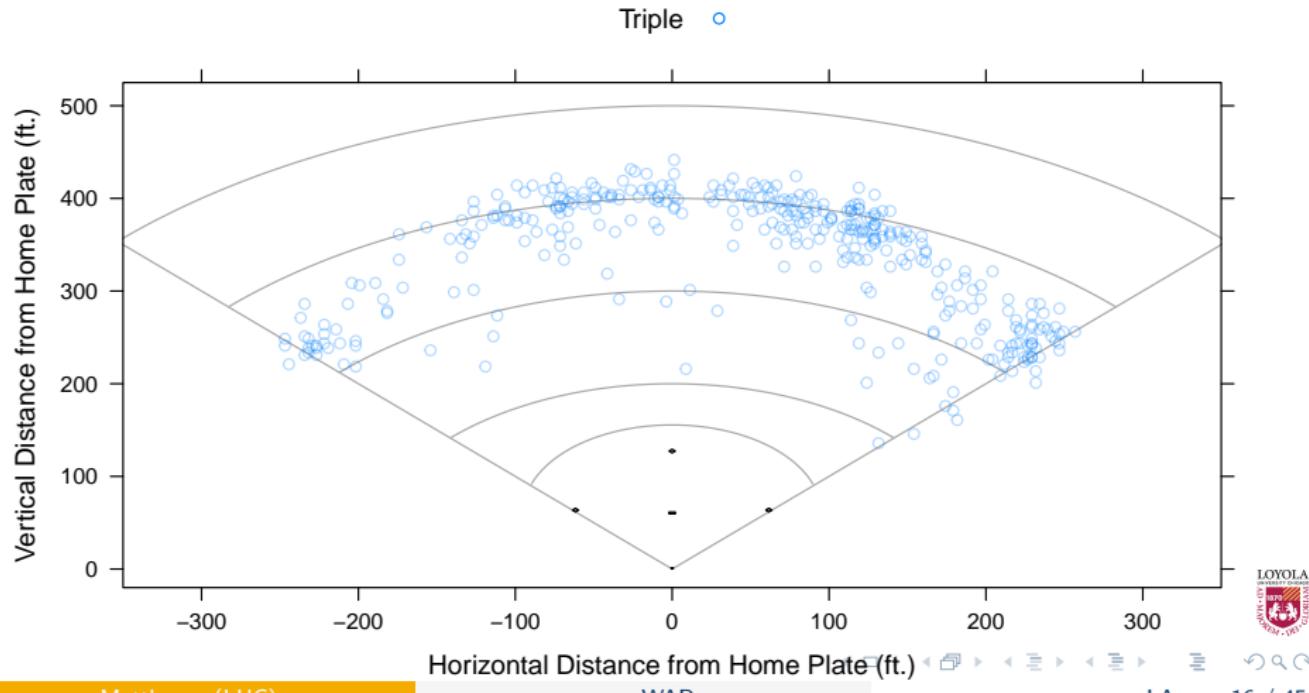
# Visualizing the Data

```
library(openWARData)
data(MLBAM2013)
plot(subset(MLBAM2013, event %in% c("Single", "Double")))
```



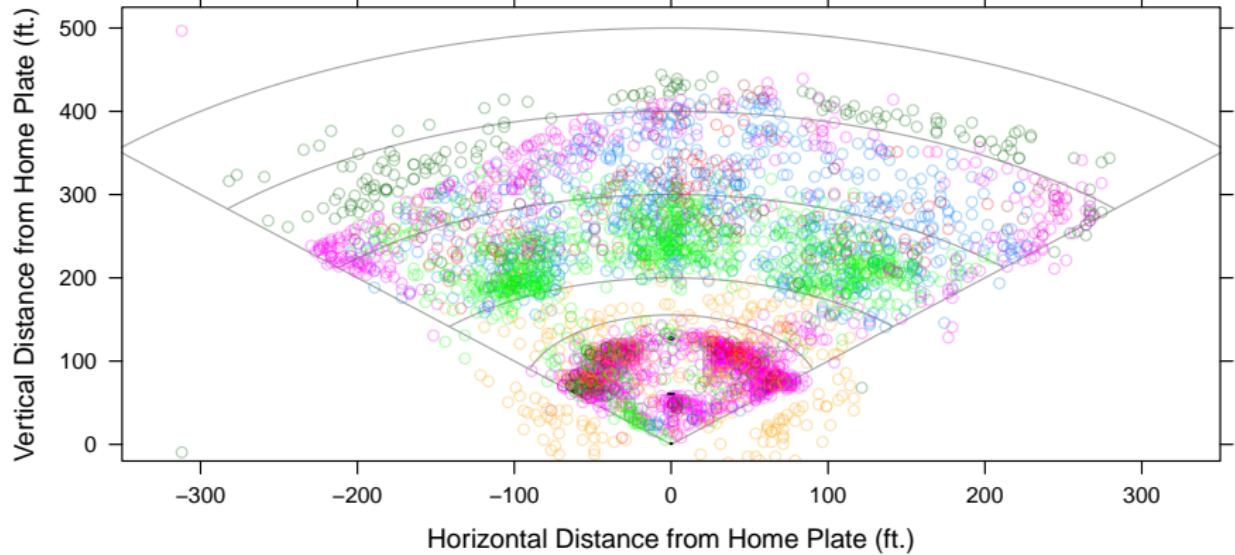
# Visualizing the Data

```
data(MLBAM2013)
plot(subset(MLBAM2013, event=="Triple" & stand=="R"))
```



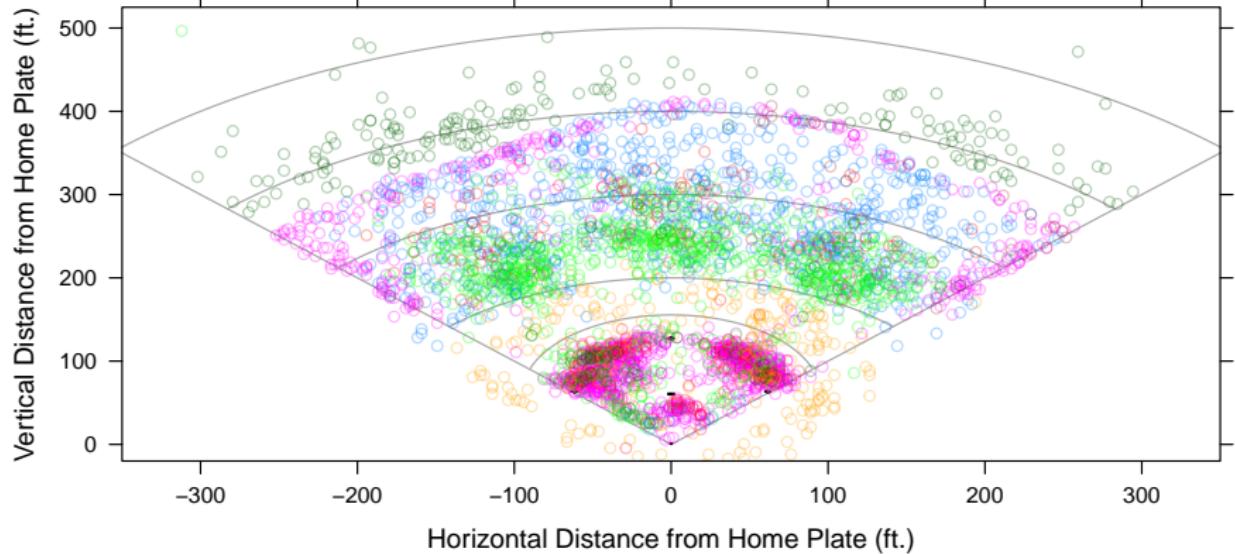
# Visualizing the Data

your roundout home Run     Lineout Pop Out Single     Triple Bunt Groundout Double     Field Error Bunt Pop Out Batter Interference     Catcher Interferer



# Visualizing the Data

you roundout home Run Lineout Pop Out Single Triple Bunt Groundout Double Field Error Bunt Pop Out Batter Interference Catcher Interferer Bunt Lineout



# Run Expectancy

- $\rho(\text{baseCode}, \text{outs})$ : expected number of runs scored in remainder of inning, from the state  $(\text{baseCode}, \text{outs})$
- Incomplete innings (i.e. walk-offs, rainouts) are removed
- Empirically estimate  $\hat{\rho}(\text{row}, \text{column})$ :

```
round(rem, 3)

##      0      1      2
## 0 0.516 0.272 0.108
## 1 0.916 0.525 0.225
## 2 1.157 0.712 0.337
## 3 1.509 0.947 0.450
## 4 1.436 0.954 0.381
## 5 1.840 1.215 0.496
## 6 2.048 1.516 0.602
## 7 2.242 1.635 0.804
```



# Conservation of Runs

- Conservation of Runs:
  - ▶ Every run gained by the offense is a run lost by the defense
- $\delta_i$ : Change in expected runs occurring on the  $i^{th}$  play:

$$\delta_i = \rho(b_{i+1}, o_{i+1}) - \rho(b_i, o_i) + \text{runsOnPlay}_i$$

- Since  $\rho(0, 3) = 0$  and

$$\hat{\rho}(0, 0) \approx \frac{\text{total observed runs scored}}{\text{number of observed complete innings}}$$

it follows that  $\sum_i \hat{\delta}_i \approx 0$



# Sample Play

- 5/08/2013: 2 outs, Nick Markakis on 2B, Adam Jones on 1B
- Matt Wieters doubles to right center and both runs score

$$\hat{\delta}_i = \hat{\rho}(2, 2) - \hat{\rho}(3, 2) + 2 = 0.31 - 0.41 + 2 = 1.90$$

[https://cvmdo.bamnetworks.com/mlbam/2013/05/08/347228/  
coaching\\_video/cv\\_26934817\\_4500K.mp4](https://cvmdo.bamnetworks.com/mlbam/2013/05/08/347228/coaching_video/cv_26934817_4500K.mp4)

- How to allocate responsibility among the offensive and defensive players?



# openWAR accounting

- $\delta = 1.90$  runs
- $\delta_{br} = 0.32$  runs, after controlling for ballpark and platoon advantage
  - ▶ The runner on first (Jones) gets 91% of the baserunning credit
  - ▶ The runner on second (Markakis) gets 9% of the baserunning credit
- $\delta_{bat} = 1.58$  runs goes to the batter (Wieters)
  - ▶ Remains 1.58 runs after controlling for the fact that Wieters is a catcher
- $\delta_{field} = -0.70$  runs (37% of the blame) go to the fielders
  - ▶ 68% of that blame ( $-0.47$  runs) goes to the CF
  - ▶ 32% of that blame ( $-0.22$  runs) goes to the RF
  - ▶ Negligible amounts go to the other fielders
- $\delta_{pitch} = -1.20$  runs (63% of the blame) goes to the pitcher

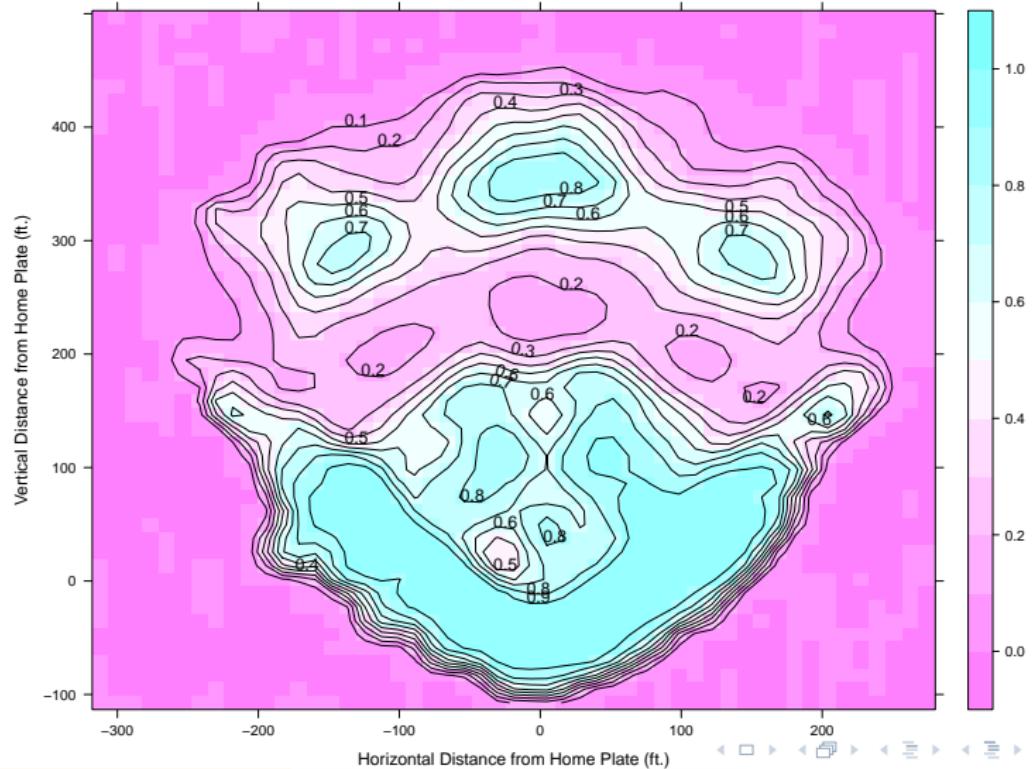


# Modeling Defense

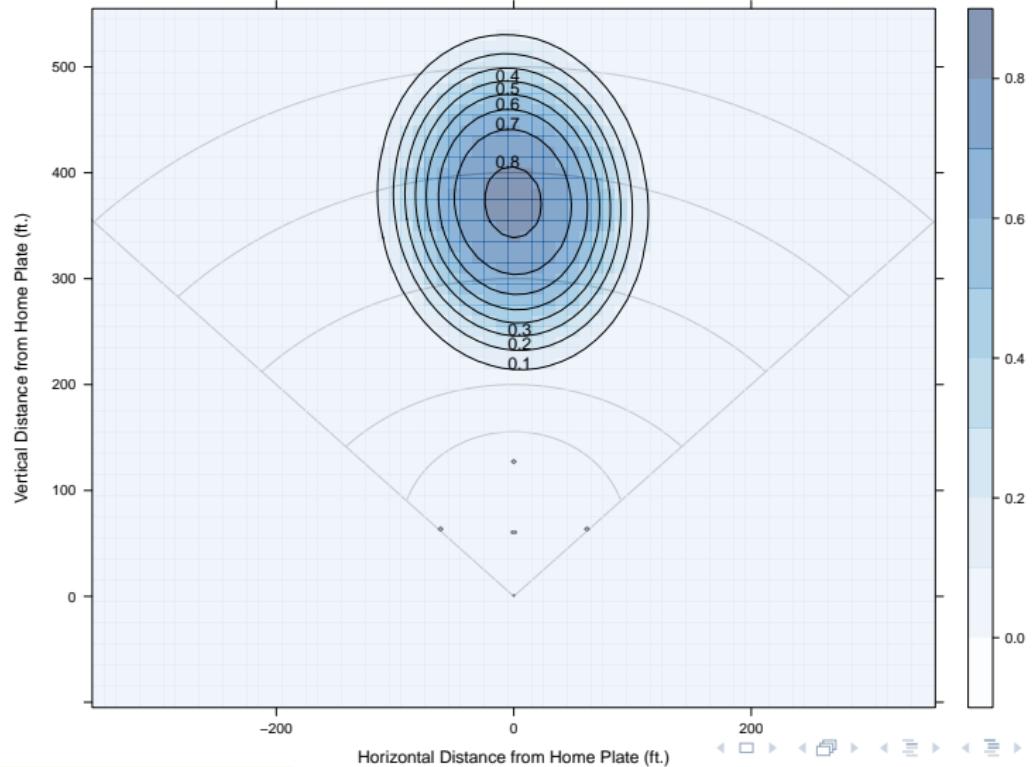
- ➊ Build model for  $p$ , the probability of any fielder making a play, in terms of location  $(x, y)$ .
- ➋  $p = f(x, y)$  where  $f$  is estimated using a 2-dimensional kernel density smoother.
- ➌  $\delta_{field} = \hat{p} \cdot \delta$ 
  - ▶ For each position, estimate the probability  $p_k$  of that position making a play
  - ▶  $\delta_{position_k} = \frac{p_k}{\sum_k p_k} \cdot \delta_{field}$
  - ▶ Construct a linear model for  $\delta_{position_k} \sim x^2 + y^2 + ballpark$
  - ▶  $RAA_{position_k}$  = residuals from these models
- ➍  $\delta_{pitch} = (1 - p) \cdot \delta$ 
  - ▶ Construct a linear model for  $\delta_{pitch} \sim ballpark + hasPlatoonAdv$
  - ▶  $RAA_{pitch}$  = residuals from this model



# Cumulative Fielding Model



# Cumulative Fielding Model

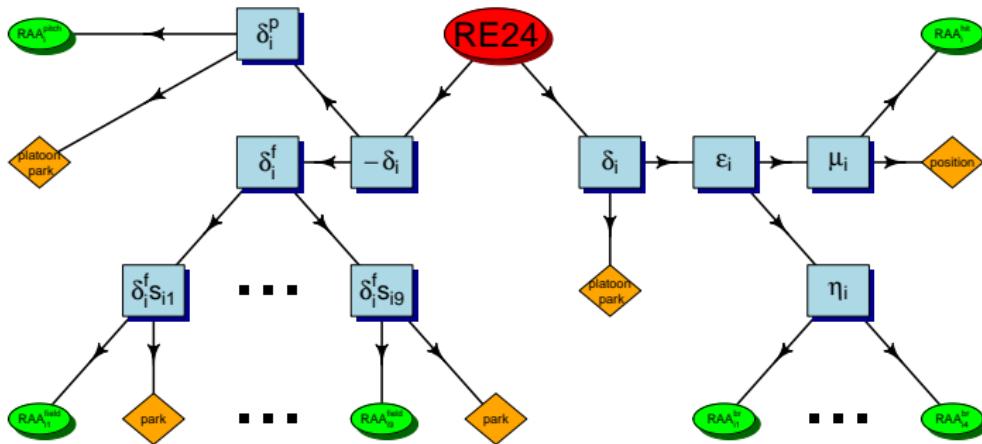


# Modeling Offense

- ➊ Build model for  $\delta$  in terms of *ballpark* and *hasPlatoonAdv*
  - ▶ Residuals from this model go to the offense ( $\delta_{off}$ )
- ➋ Build model for  $\delta_{off}$  in terms of *event*, *startCode*, *startOuts*
  - ▶ Residuals from this model go to the baserunners ( $\delta_{br}$ )
  - ▶ Compute empirical probability of each baserunner doing **as well**
  - ▶ Normalize these probabilities to compute a **share**
  - ▶ Credit to baserunners apportioned based on shares
- ➌ Remainder goes to batter ( $\delta_{bat} = \delta - \delta_{br}$ )
- ➍ Build model for  $\delta_{bat}$  in terms of *batterPos*
  - ▶  $RAA_{bat}$  = residuals from this model



# openWAR Flowchart

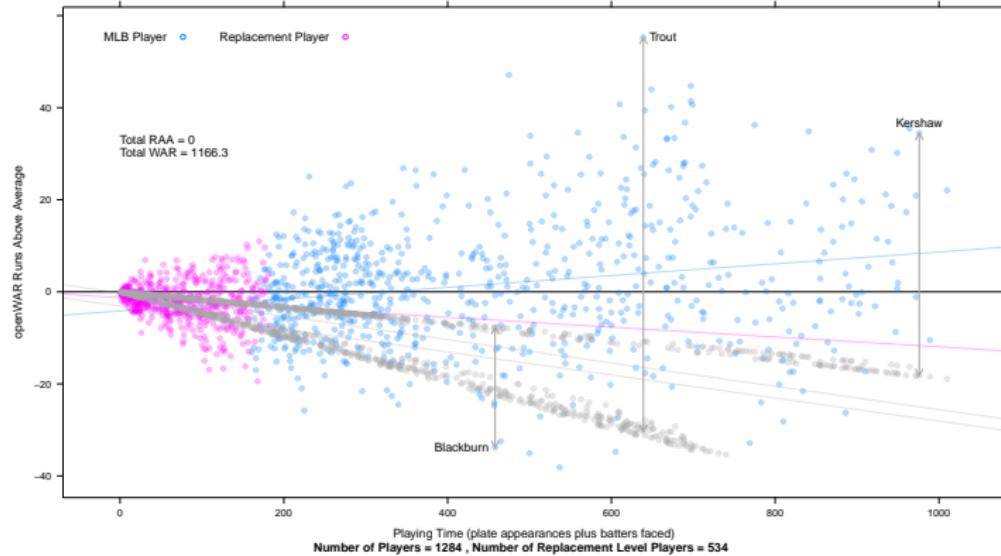


# Defining Replacement Level

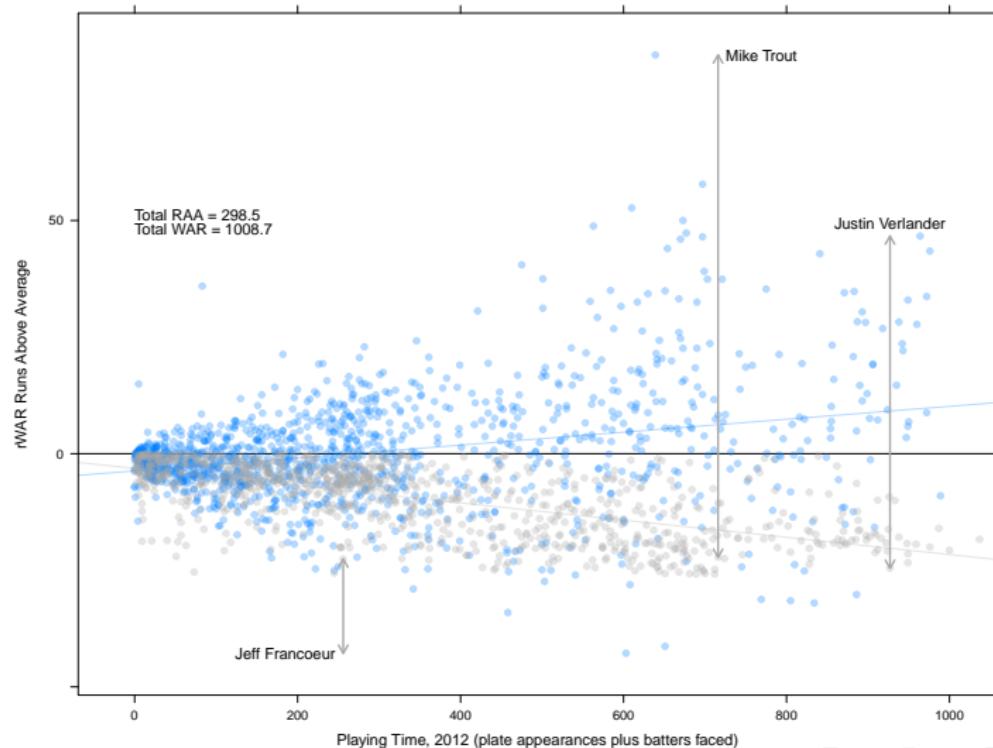
- Scarcity: Only  $30 \cdot 25 = 750$  roster spots
  - ▶ Take the 750 players who played the most
  - ▶ All other players are by definition “replacements”
- Replacement players have an average RAA per plate appearance
- Each player is assigned a *replacement-level shadow* based on their playing time (shown in gray in next slide)



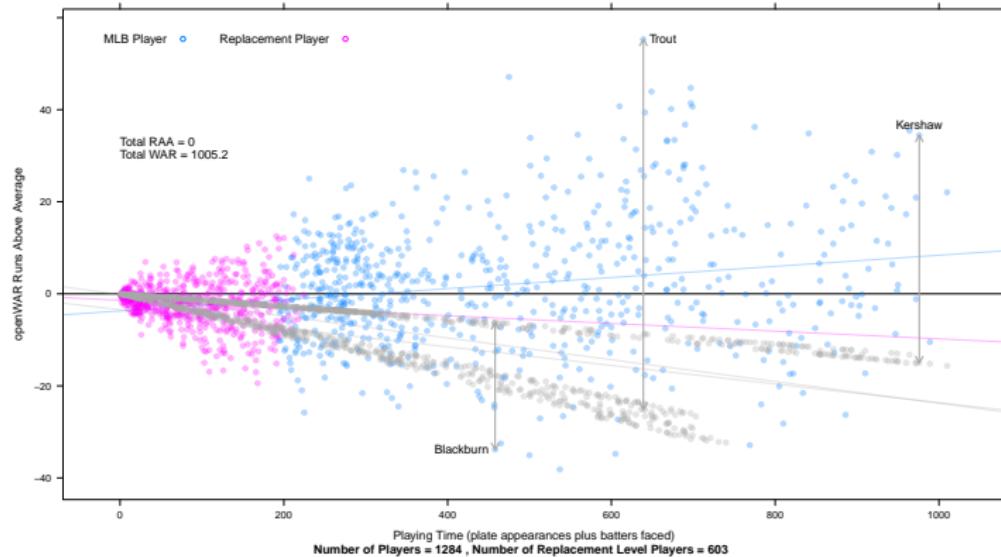
# Defining Replacement Level - openWAR 2012



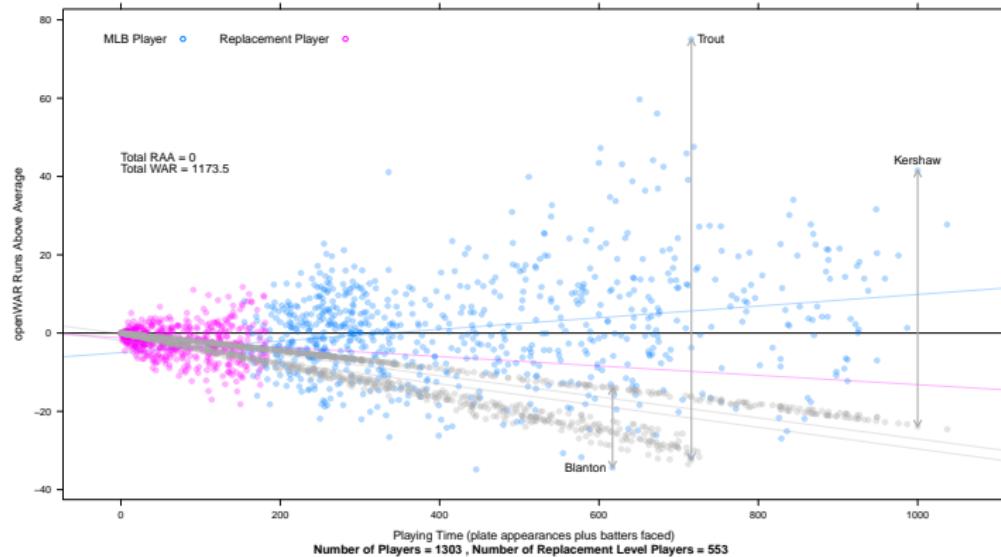
# Defining Replacement Level - rWAR 2012



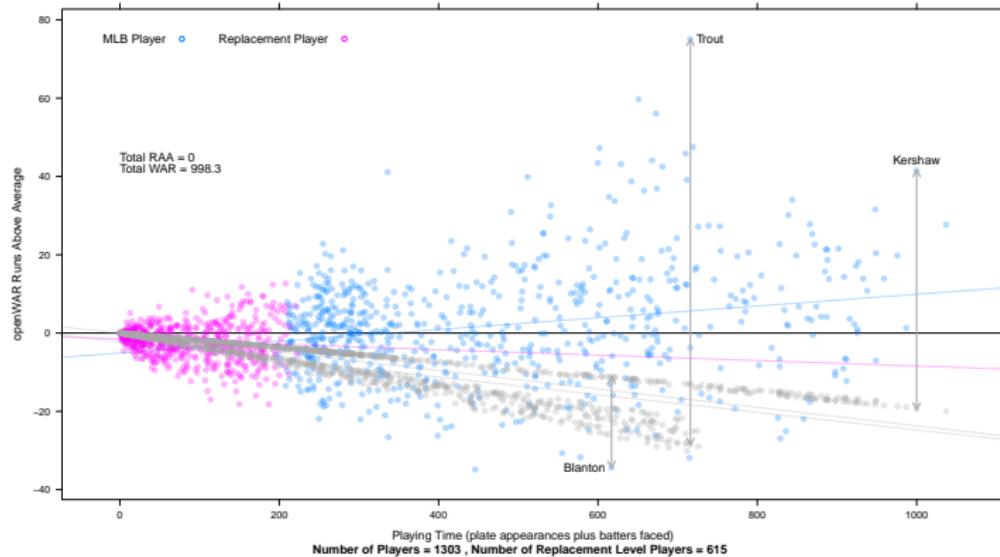
# Defining Replacement Level - openWAR 2012 normalized



# Defining Replacement Level - openWAR 2013



# Defining Replacement Level - openWAR 2013 normalized



# 2017 openWAR leaders

```
load("/Users/gregorymatthews/Dropbox/war2017.RData")
out <- war2017[order(-war2017$WAR),] [1:10,]
out[,c("Name", "TPA", "WAR", "RAA", "repl")]
```

	Name	TPA	WAR	RAA	repl
## 173	Blackmon	724	8.689068	55.20032	-31.69036
## 709	Arenado	680	7.701402	49.33395	-27.68006
## 218	Votto	707	6.851037	41.19690	-27.31348
## 658	Trout	507	6.649125	44.21665	-22.27460
## 859	Judge	678	6.641795	46.20091	-20.21704
## 84	Cruz	645	6.626615	42.08219	-24.18396
## 422	Altuve	662	6.624671	37.08417	-29.16254
## 167	Scherzer	851	6.481137	46.37177	-18.43960
## 249	Andrus	689	6.293150	34.88380	-28.04770
## 833	Bryant	665	6.158989	34.91433	-26.67555

# 2018 openWAR leaders

```
load("/Users/gregorymatthews/Dropbox/war2018.RData")
out <- war2018[order(-war2018$WAR),][1:10,]
out[,c("Name", "TPA", "WAR", "RAA", "repl")]
```

	Name	TPA	WAR	RAA	repl
## 403	Polanco	39	1.1100098	9.960885	-1.139213
## 369	Gregorius	42	1.0147995	8.480418	-1.667577
## 451	Pollock	40	1.0106474	8.382959	-1.723514
## 785	Ohtani	64	0.9444003	7.811270	-1.632733
## 172	Smoak	44	0.8973726	7.318298	-1.655428
## 681	Swanson	35	0.8627940	7.198783	-1.429157
## 262	Freeman, F	43	0.8577686	6.953067	-1.624619
## 535	Bogaerts	40	0.7585258	5.962711	-1.622548
## 386	Harper	43	0.7444915	6.159513	-1.285402
## 523	Taillon	55	0.7356530	6.157353	-1.199177

# Modeling Uncertainty

- Both *rWAR* and *fWAR* are published as **point** estimates, not interval estimates
- *openWAR* models two types of uncertainty:
  - ➊ Estimation Error: uncertainty associated with the models
    - ★ Resample the actual data, then build the models on the resampled data
    - ★ Use the resampled models to evaluate the actual data
    - ★ Should be very small, since sample size is large ( $\geq 100k$  plays per season)
    - ★ **Work in progress**
  - ➋ Sampling Error: uncertainty associated with the sample of data
    - ★ Keep the original models, but resample the plays



# 2017 openWAR leaders

```
load("/Users/gregorymatthews/Dropbox/shakeWARplays2017.RData")
summary(test)

## # A tibble: 1,358 x 10
##   playerId Name      N    q0   q2.5   q25   q50   q75   q97.5   q100
##   <dbl> <chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 453568. Blackmon 200  4.34  5.57  7.63  8.77  9.74 11.9   13.4
## 2 571448. Arenado  200  3.95  4.73  6.72  7.93  9.12 11.5   12.9
## 3 458015. Votto    200  3.23  4.52  6.04  7.01  7.86  9.53 12.2
## 4 545361. Trout    200  3.36  4.39  5.99  6.89  7.84  9.29 10.1
## 5 592450. Judge    200  3.20  3.95  5.64  6.79  7.96  9.59 11.7
## 6 443558. Cruz     200  2.89  3.94  5.82  6.78  7.69  9.59 10.4
## 7 514888. Altuve   200  3.65  4.30  5.73  6.63  7.50  9.12 10.4
## 8 453286. Scherzer 200  4.02  4.62  5.74  6.52  7.16  8.67  9.28
## 9 462101. Andrus   200  2.71  3.80  5.30  6.44  7.37  9.64 10.3
## 10 502517. Murphy   200  2.05  3.57  5.46  6.40  7.34  9.21 10.6
## # ... with 1,348 more rows
```



# Reliability and Comparison to Other Implementations

- Reliability: autocorrelation of *openWAR* is comparable

	<i>rWAR</i>	<i>fWAR</i>	<i>openWAR</i>
Autocorrelation	0.522	0.596	0.571
	<i>rWAR</i>	<i>fWAR</i>	<i>openWAR</i>

- Correlation:

	<i>rWAR</i>	<i>fWAR</i>	<i>openWAR</i>
<i>rWAR</i>	1	0.918	0.881
<i>fWAR</i>	0.918	1	0.875



# The Future State of WAR

*Over the past four years, sabermetricians estimate that there is a 63% chance that Mr. Zobrist has led baseball in WAR, ahead of stars like Albert Pujols, Ryan Braun and Robinson Cano.* – The New York Times, 4/1/2013

*According to baseball-reference.com, Trout's openWAR was  $10.7 \pm 1.2$ , almost certainly exceeding Cabrera's  $6.9 \pm 0.9$ .* – Sports Illustrated, 11/15/2012

*There is only a 3% chance that any player in history has been a plus-10 runner and a plus-20 fielder and a plus-30 hitter.* – ESPN The Magazine, 3/4/2013



# Limitations

- Data integrity
- Stolen bases and wild pitches not properly accounted
- Cannot distribute data with the R package (sort of)
- Can't distinguish between batted ball trajectories or speeds
- Batted ball locations are approximate.
- Defense measures only **range** – not sure-handedness, throwing, etc.



# Summary

- *openWAR*: a reference implementation of WAR
- Open source R package (*openWAR*) on GitHub
- Holistic methodology – Conservation of Runs
- Variances estimates
- Diagnostic & Visualization tools



# Summary

Cheers!!



# 2012 openWAR leaders

Name	TPA	WAR	RAA	repl
Trout	639	8.6	55.3	-30.4
Cano	697	7.9	44.7	-34.3
Cabrera, Mi	697	7.5	41.4	-33.8
Headley	699	7.5	40.7	-34.3
Encarnacion	649	7.3	43.9	-28.9
McCutchen, A	673	7.2	40.2	-32.3
Votto	475	7.0	47.1	-22.6
Fielder	690	6.9	36.8	-32.5
Mauer	641	6.7	39.4	-27.9
Posey	610	6.7	40.7	-26.3
Hill, A	668	6.6	33.4	-32.8
Braun	677	6.6	33.4	-32.3
Zobrist	668	6.5	33.1	-31.8
Willingham	615	6.3	33.6	-29.0
Ramirez, Ar	630	6.2	31.4	-30.8

Table: 2012 openWAR Leaders



# 2013 openWAR leaders

Name	TPA	WAR	RAA	repl
Trout	716	10.7	75.1	-32.4
Cabrera, M	651	8.7	59.7	-26.9
Davis, C	673	8.5	56.1	-29.0
Carpenter, M	719	7.8	47.6	-30.0
Goldschmidt	710	7.7	45.9	-30.8
Choo	712	7.3	39.1	-33.5
Donaldson	668	7.2	44.1	-27.9
Holliday	602	7.1	47.2	-24.0
Cano	681	7.1	42.5	-28.4
Freeman, F	629	7.1	43.1	-27.4
McCutchen	674	6.9	36.9	-32.0
Ortiz, D	600	6.7	43.4	-23.2
Kershaw	1000	6.5	41.5	-23.9
Santana, C	642	6.4	38.8	-25.1
Kipnis	658	6.4	36.4	-27.4

Table: 2013 openWAR Leaders



# 2014 openWAR leaders

Name	TPA	WAR	RAA	repl
Trout	705	9.1	63.5	-27.7
Kershaw	823	7.0	50.0	-20.1
McCutchen	648	6.7	41.2	-25.9
Dozier	707	6.7	33.0	-33.6
Cabrera, M	685	6.4	38.9	-24.9
Brantley	676	6.4	39.4	-24.3
Kendrick, H	674	6.3	30.5	-32.4
Stanton	637	6.3	38.8	-23.7
Werth	629	6.2	37.6	-23.8
Bautista	673	6.1	35.2	-25.5
Gordon, A	643	6.1	38.4	-22.1
Abreu	624	6.0	36.4	-23.0
Kluber	957	6.0	37.1	-22.3
Martinez, V	641	5.9	34.7	-24.0
Puig	646	5.8	34.0	-24.7

Table: 2014 openWAR Leaders

