OpenWAR:

An Open Source System for Overall Player Performance in MLB

Ben Baumer[1] Shane Jensen[2] Gregory Matthews[3]

[1] Smith College

[2] The Wharton School
University of Pennsylvania

[3] University of Massachusetts

Amherst College
April 11, 2014

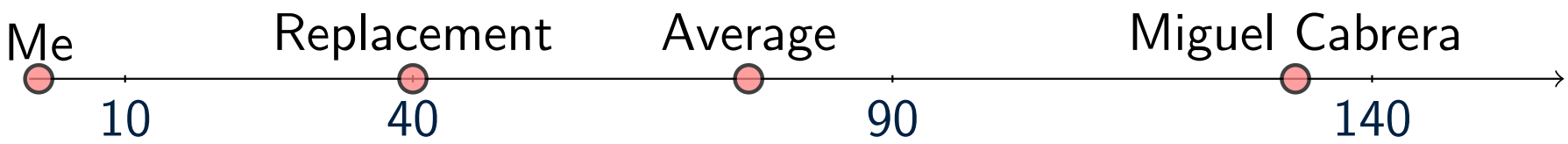
SMITH COLLEGE

WAR - What is it good for?

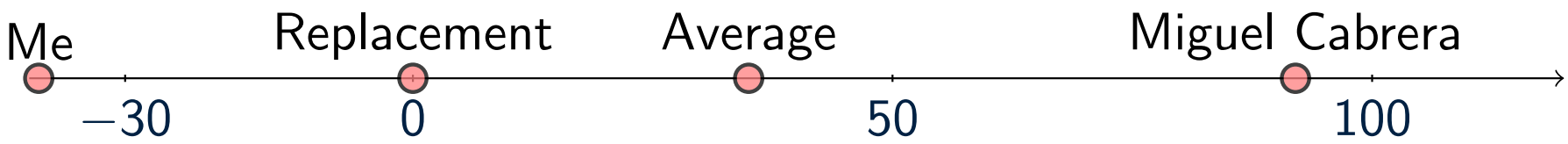
- Wins Above Replacement
- Question: How large is the contribution that each player makes towards winning?
- Four Components:
  1. Batting
  2. Baserunning
  3. Fielding
  4. Pitching
- Replacement Player: Hypothetical 4A journeyman
  - Much worse than an average player

Units and Scaling

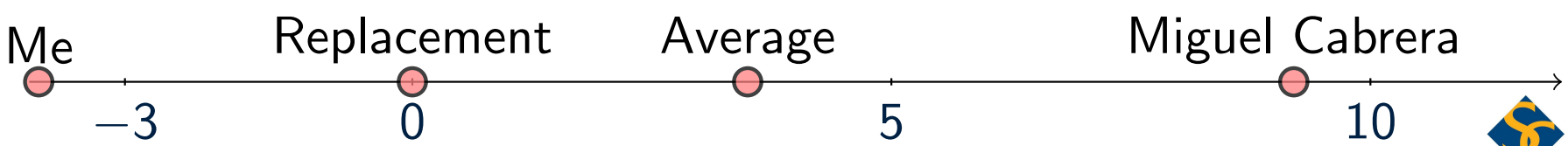
- In terms of absolute runs:



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



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



Example: 2012 WAR leaders

FanGraphs	*fWAR*	BB-Ref	*rWAR*
Mike Trout	10.0	Mike Trout	10.9
Robinson Cano	7.8	Robinson Cano	8.5
Buster Posey	7.7	Buster Posey	7.4
Ryan Braun	7.6	Miguel Cabrera	7.3
David Wright	7.4	Andrew McCutchen	7.2
Chase Headley	7.2	Adrian Beltre	7.0
Miguel Cabrera	6.8	Ryan Braun	7.0
Andrew McCutchen	6.8	Yadier Molina	6.9

Table : 2012 WAR Leaders

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

WAR is the Answer

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`
- 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

```
# Dependency not on CRAN
install.packages("Sxslt", repos = "http://www.omegahat.org/R", type = "sour
devtools::install_github("openWAR", "beanumber")
```

Getting Data

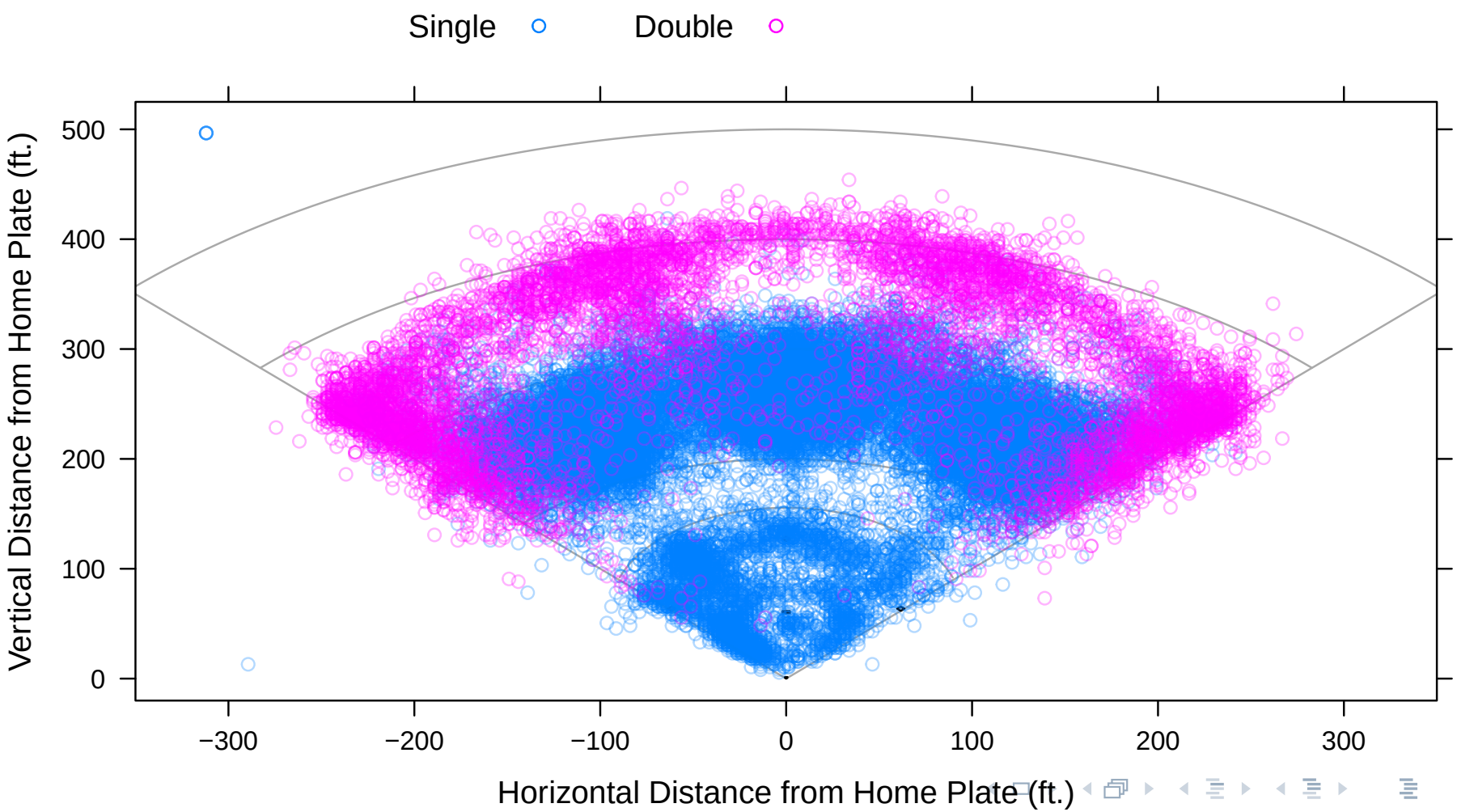
```
require(openWAR)
ds = getData(start = "2013-06-24")

##
## Retrieving data from 2013-06-24 ...
## ...found 4 games

dim(ds)

## [1] 298  62

head(ds$description)

## [1] Michael Bourn called out on strikes.
## [2] Mike Aviles lines out to right fielder Nick Markakis.
## [3] Jason Kipnis hit by pitch.
## [4] Nick Swisher singles on a line drive to left fielder Nate McLouth.   Jason K
## [5] Carlos Santana flies out to center fielder Adam Jones.
## [6] Nate McLouth called out on strikes.
## 345 Levels: Adam Jones flies out to right fielder Drew Stubbs.   ...
```

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

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



Run Expectancy

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

```
round(rem, 3)
```

```
##       0     1     2
## 0 0.456 0.240 0.091
## 1 0.812 0.491 0.211
## 2 1.096 0.617 0.301
## 3 1.382 0.838 0.402
## 4 1.261 0.925 0.344
## 5 1.828 1.108 0.480
## 6 2.080 1.390 0.558
## 7 2.179 1.568 0.714
```

Conservation of Runs

- Conservation of Runs:
  - Every run gained by the offense is a run lost by the defense
- δ_i: Change in expected runs occuring on the i^{th} play:

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

- 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`

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

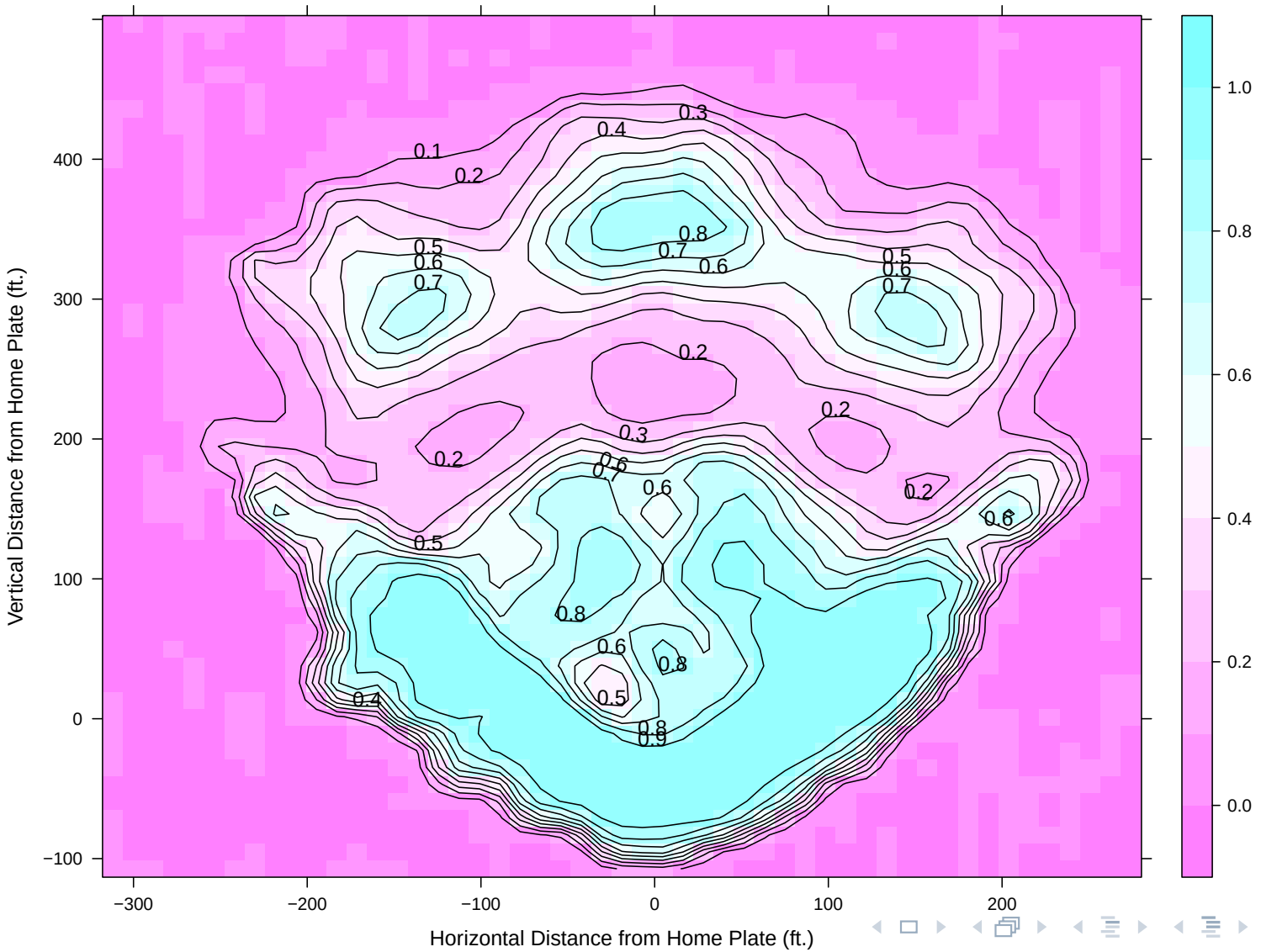
openWAR accounting

- $\delta = 1.90$ 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

1. Build model for p, the probability of any fielder making a play, in terms of location
2. $\delta_{field} = \hat{p} \cdot \delta$
   - For each position, estimate the probabilty 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
3. $\delta_{pitch} = (1 - p) \cdot \delta$
   - Construct a linear model for $\delta_{pitch} \sim ballpark + hasPlatoonAdv$
   - RAA_{pitch} = residuals from this model

Cumulative Fielding Model



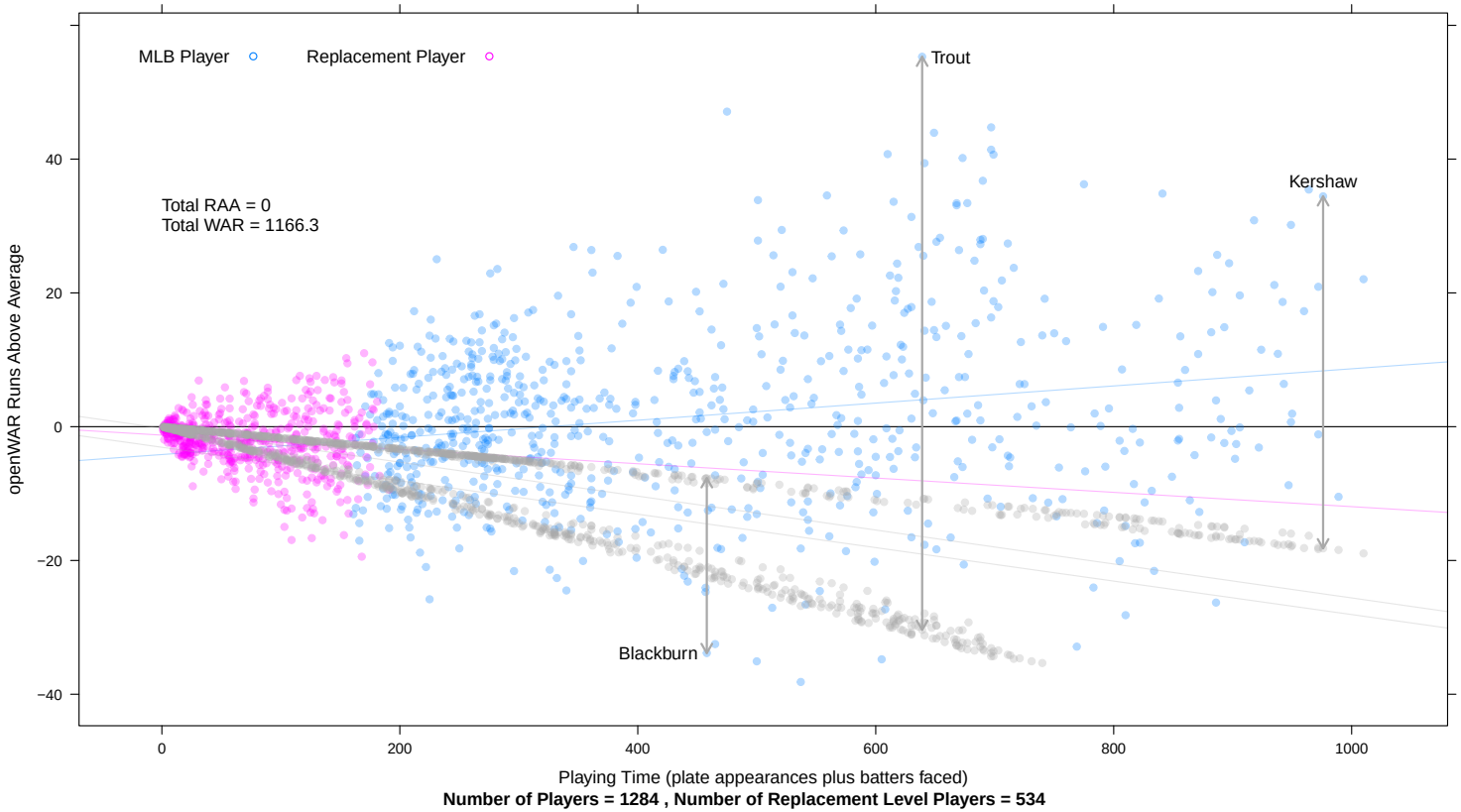
Modeling Offense

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

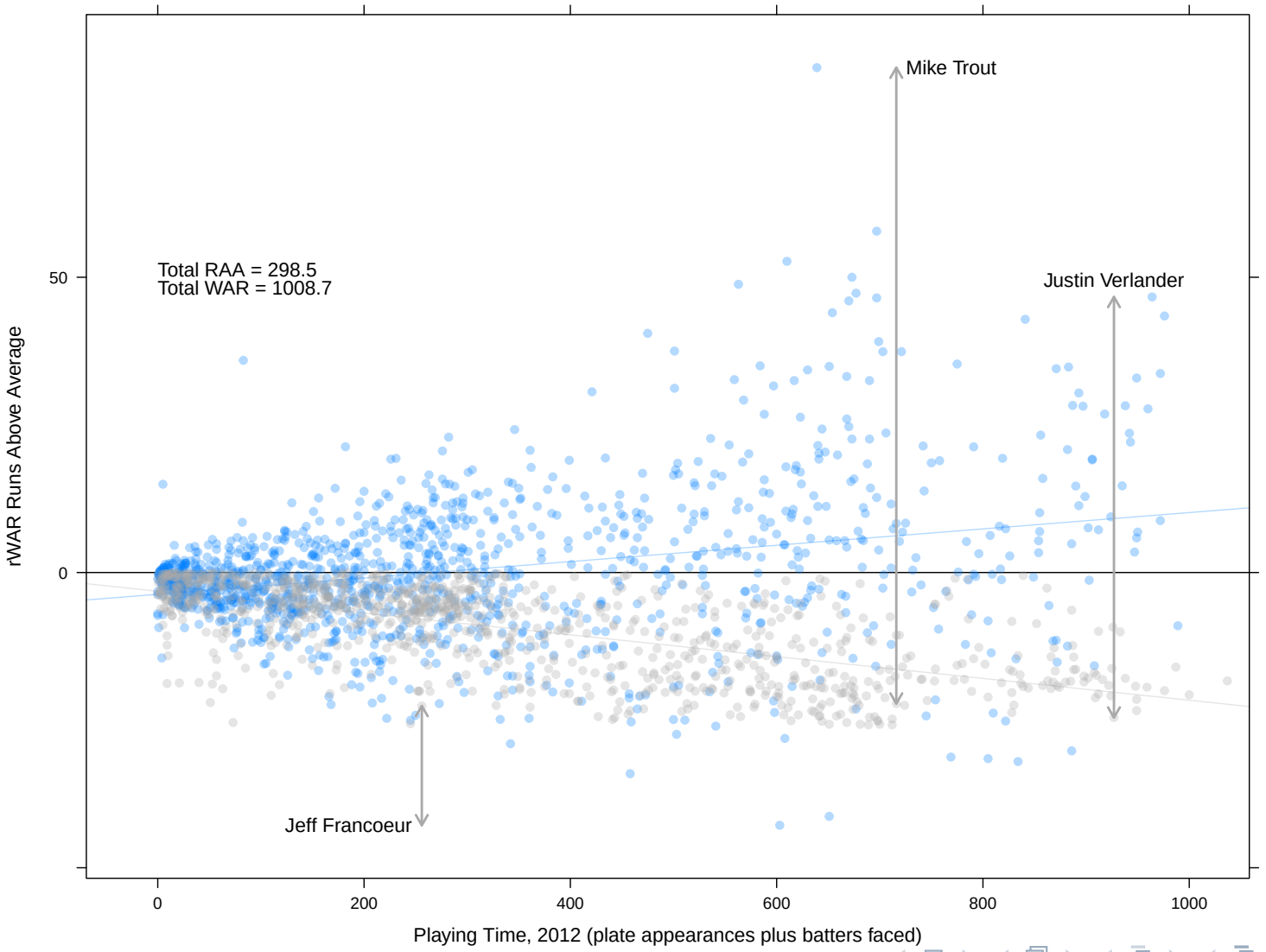
Defining Replacement Level

- 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)

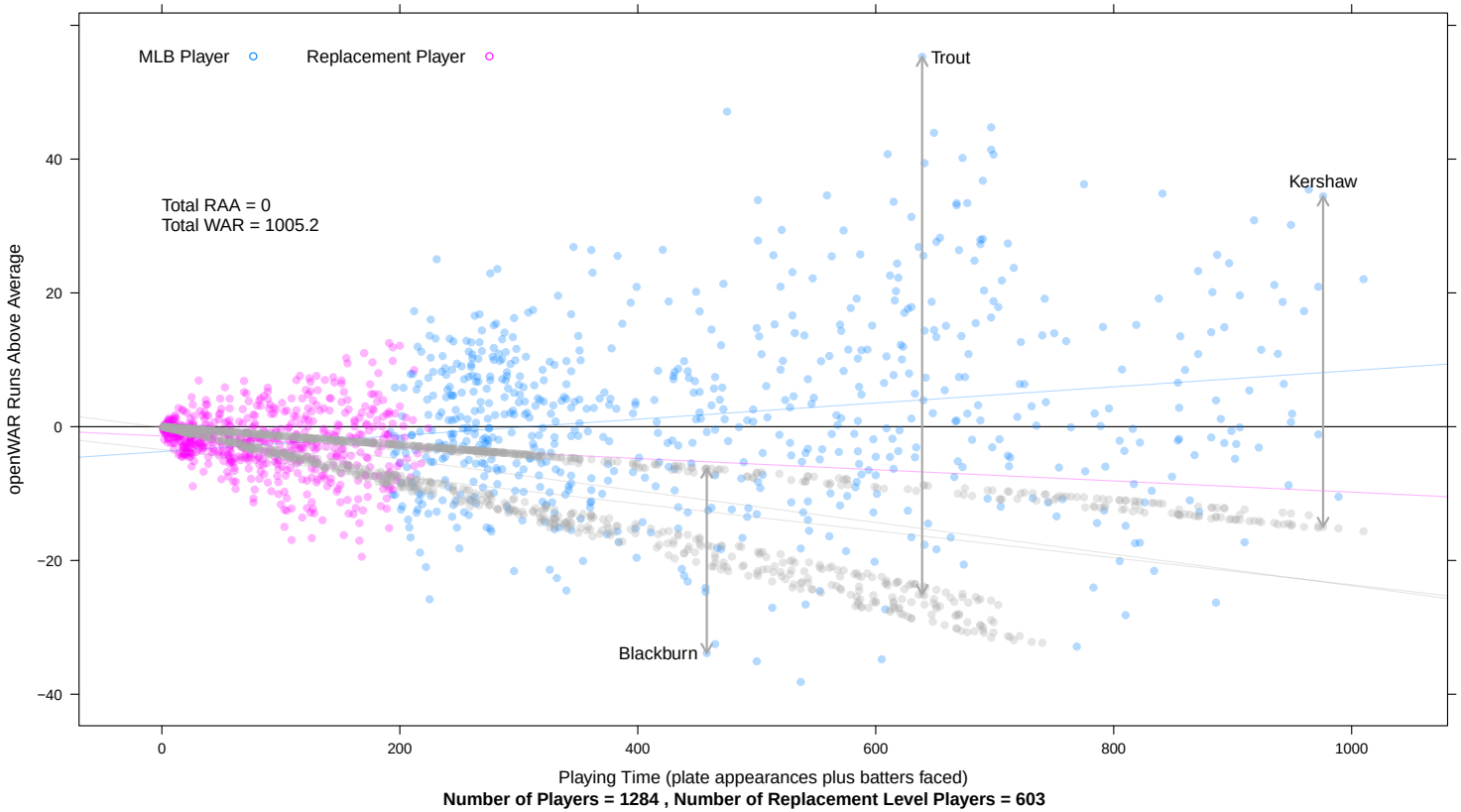
Defining Replacement Level - openWAR 2012



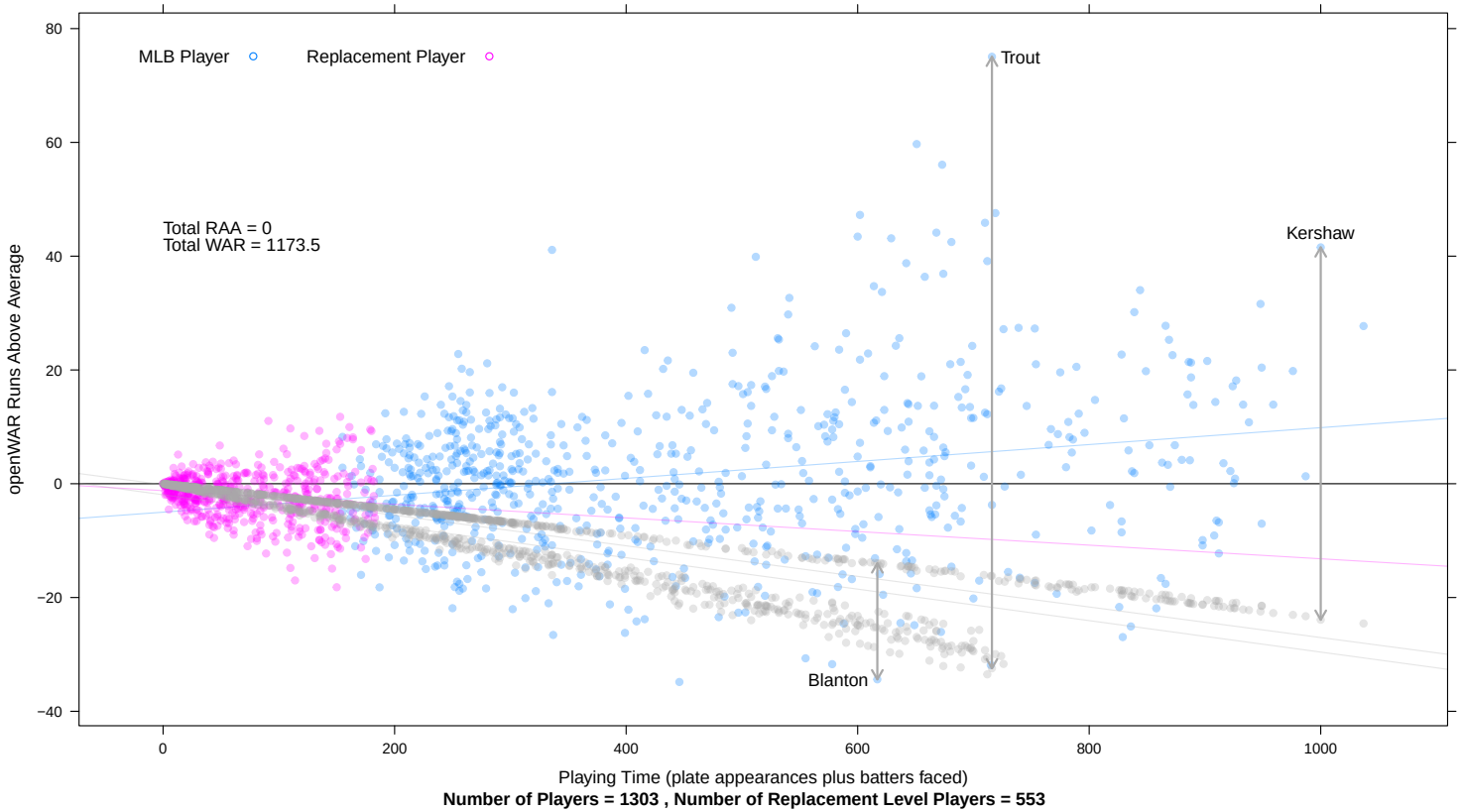
Defining Replacement Level - rWAR 2012



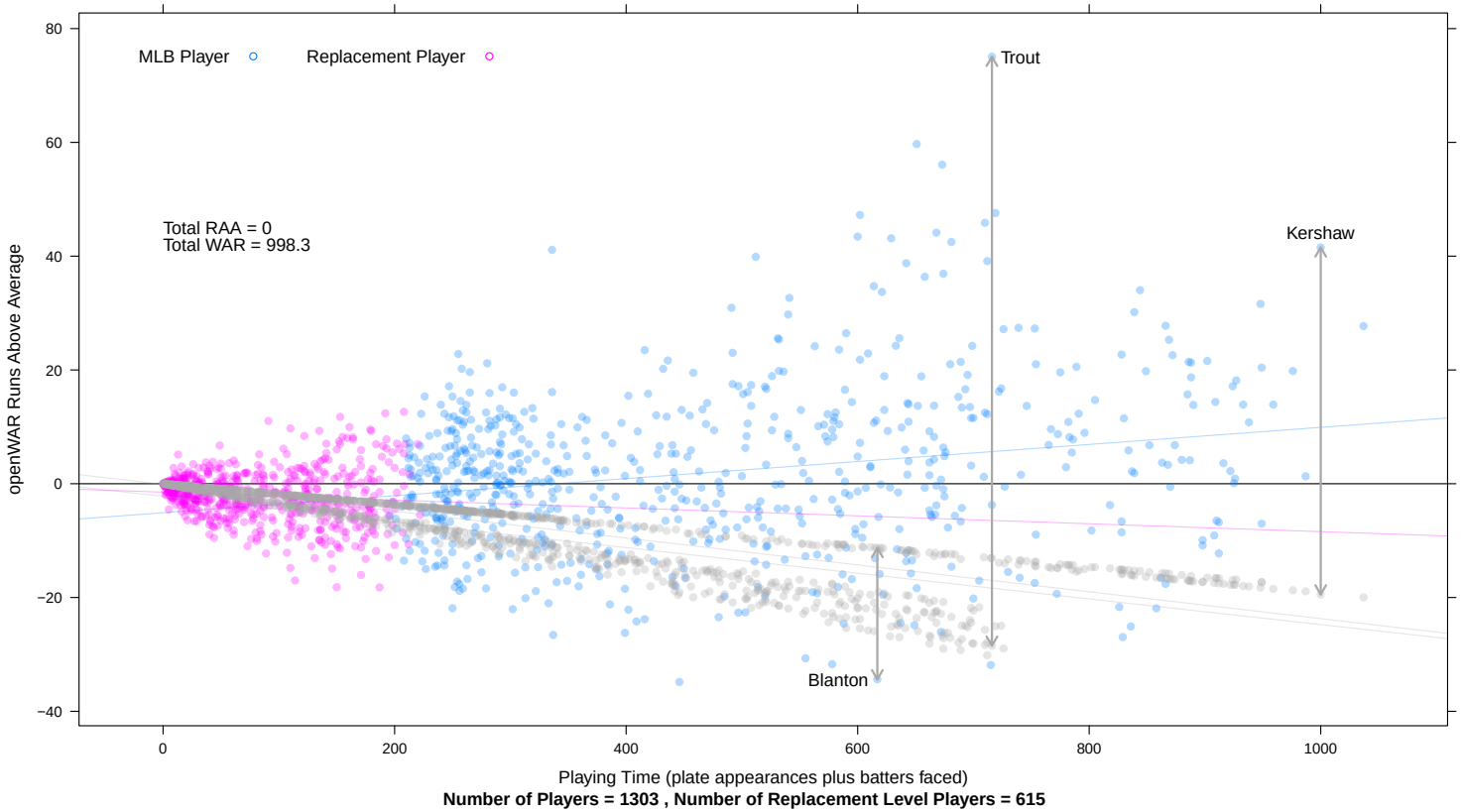
Defining Replacement Level - openWAR 2012 normalized



Defining Replacement Level - openWAR 2013



Defining Replacement Level - openWAR 2013 normalized



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

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

Modeling Uncertainty

- Both *rWAR* and *fWAR* are published as point estimates, not interval estimates
- *openWAR* models two types of uncertainty:
  1. 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
  2. Sampling Error: uncertainty associated with the sample of data
     - ★ Keep the original models, but resample the plays

Uncertainty (Sampling)

Name	q0	q2.5	q25	q50	q75	q97.5	q100
Trout	3.96	5.90	7.66	8.58	9.50	11.26	13.34
Cano	2.43	4.77	6.82	7.94	9.06	11.17	13.85
Cabrera, Mi	2.17	4.40	6.40	7.51	8.55	10.72	13.24
Headley	2.53	4.56	6.39	7.46	8.47	10.42	13.00
Encarnacion	2.45	4.47	6.33	7.32	8.27	10.09	13.08
McCutchen, A	2.33	4.41	6.25	7.27	8.26	10.27	12.21
Votto	2.68	4.80	6.23	7.01	7.81	9.28	11.40
Fielder	2.11	4.12	5.99	6.98	7.95	9.82	12.22
Posey	2.09	4.14	5.80	6.75	7.67	9.61	11.68
Mauer	2.63	4.30	5.88	6.74	7.60	9.27	10.91

Table : 2012 openWAR Leaders with Uncertainty Estimates: 3423 simulations

Reliability and Comparison to Other Implementations

- Reliability: autocorrelation of *openWAR* is comparable

	rWAR	*fWAR*	*openWAR*
Autocorrelation	0.522	0.596	0.571

- Correlation:

	rWAR	*fWAR*	*openWAR*
rWAR	1	0.918	0.881
fWAR	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 ± 1.2*, almost certainly exceeding Cabrera's* 6.9 ± 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
- Can't distinguish between batted ball trajectories or speeds
- Defense measures only range – not sure-handedness, throwing, etc.

Summary

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

Summary

THANK YOU!!