Measuring Umpire Consistency
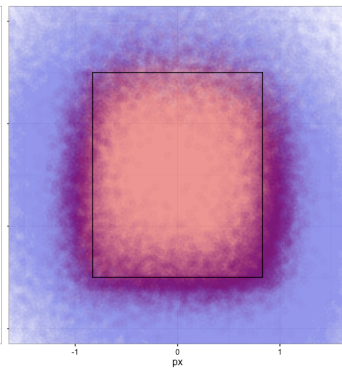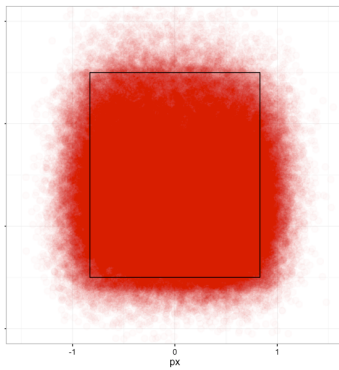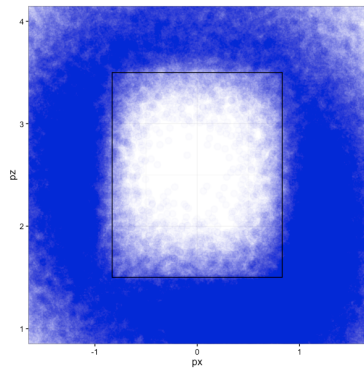
David J. Hunter

January 11, 2018

Assumptions/Motivations

- Consistency is more important than conformity.
- A consistent zone need not be rectangular, but should be convex.
- Consistency *within* a game is important.
- Different zones for LH and RH batters are OK.
- One egregiously bad call is worse/as bad as several marginally bad calls.

Assumptions/Motivations

- Consistency is more important than conformity.
- A consistent zone need not be rectangular, but should be convex.
- Consistency *within* a game is important.
- Different zones for LH and RH batters are OK.
- One egregiously bad call is worse/as bad as several marginally bad calls.

Assumptions/Motivations

- Consistency is more important than conformity.
- A consistent zone need not be rectangular, but should be convex.
- Consistency *within* a game is important.
- Different zones for LH and RH batters are OK.
- One egregiously bad call is worse/as bad as several marginally bad calls.

Assumptions/Motivations

- Consistency is more important than conformity.
- A consistent zone need not be rectangular, but should be convex.
- Consistency *within* a game is important.
- Different zones for LH and RH batters are OK.
- One egregiously bad call is worse/as bad as several marginally bad calls.

Assumptions/Motivations

- Consistency is more important than conformity.
- A consistent zone need not be rectangular, but should be convex.
- Consistency *within* a game is important.
- Different zones for LH and RH batters are OK.
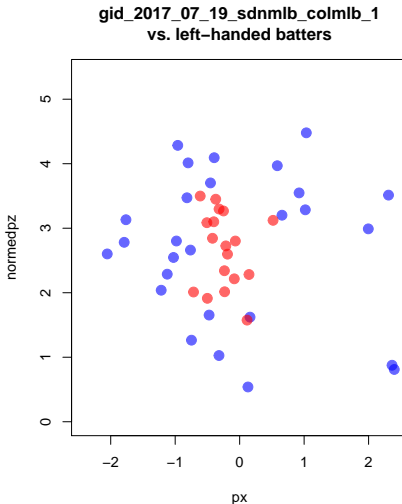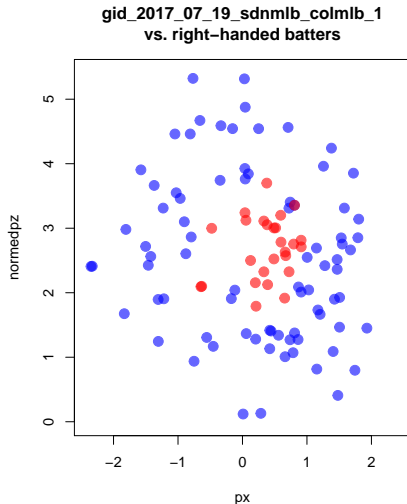- One egregiously bad call is worse/as bad as several marginally bad calls.

Define established ball and strike zones

Established Strike Zone: Convex Hull

Let H_l denote the open half-plane bounded by the line l. The convex hull of a set of points P is the set

$$S = \bigcap_{\{H_l \mid H_l \cap P = \emptyset\}} H_l^c$$

When P contains the locations of all called strikes, S is the *established strike zone*.

Established Strike Zone: Convex Hull

gid_2017_07_19_sdnmlb_colmlb_1
vs. right-handed batters

gid_2017_07_19_sdnmlb_colmlb_1
vs. left-handed batters
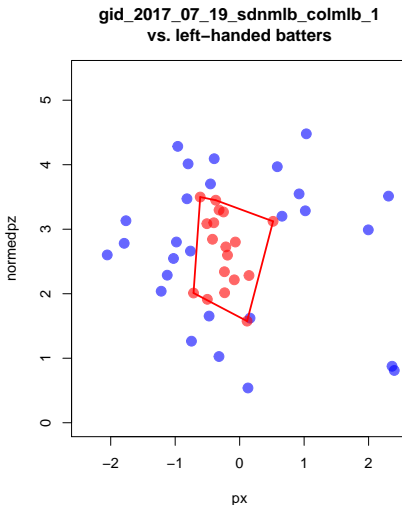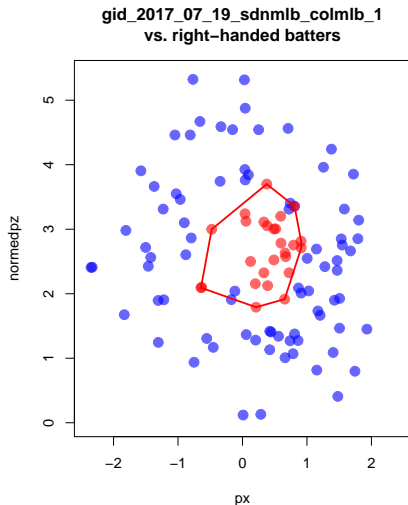
Established Ball Zone: α-**Shape?**

For $\alpha > 0$, let $B_{x,\alpha}$ denote the open ball in $\mathbb{R}^2$ of radius α centered at the point x. Given a set of points $P \subset \mathbb{R}^2$, two points $p_1, p_2 \in \mathbb{R}^2$ are α-*neighbors* if p_1 and p_2 lie on the boundary of some $B_{x,\alpha}$ such that $B_{x,\alpha} \cap P = \emptyset$.

The α-*shape* is the straight line graph formed by drawing line segments between α-neighbors.

- α-neighbors are close together but next to big empty regions.
- The α-shape is the outline of the points P.
- Does not need to be convex or even simply-connected.
- For large α, the α-shape is the boundary of the convex hull.

Established Ball Zone: α-Shape?

gid_2017_07_19_sdnmlb_colmlb_1
vs. right–handed batters

gid_2017_07_19_sdnmlb_colmlb_1
vs. left–handed batters

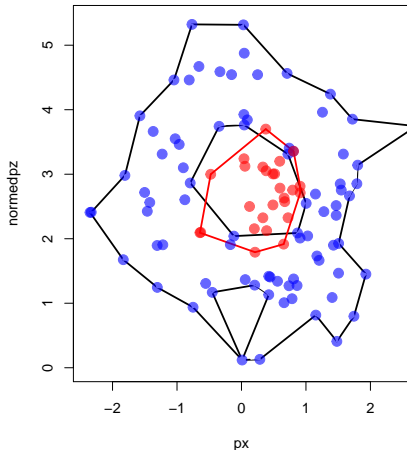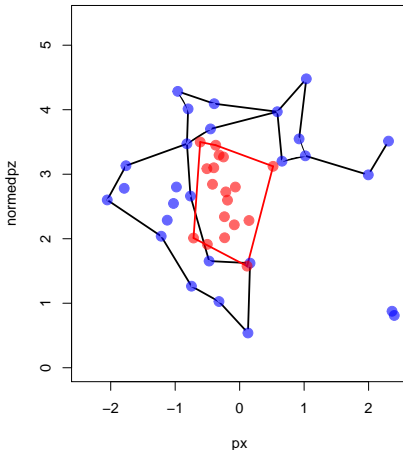# Established Ball Zone: α-Hull

For $\alpha > 0$, let $B_{x,\alpha}$ denote the open ball in $\mathbb{R}^2$ of radius α centered at the point x. Given a set of points $P \subset \mathbb{R}^2$, the α-hull of P is the set

$$X = \bigcap_{\{B_{x,\alpha} | B_{x,\alpha} \cap P = \emptyset\}} B_{x,\alpha}^c$$

When P contains the locations of all called balls, X is the *established ball zone*.

- Approaches convex hull $S = \bigcap_{\{H_l | H_l \cap P = \emptyset\}} H_l^c$ as $\alpha \to \infty$.
- Intuitively: hole punch.
- May not be simply connected for smallish α.

Established Ball Zone: α-Hull

gid_2017_07_19_sdnmlb_colmlb_1 vs. right-handed batters
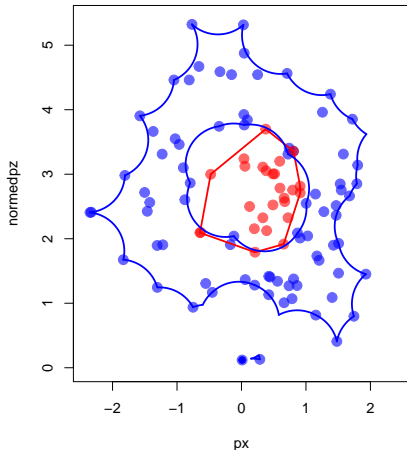
gid_2017_07_19_sdnmlb_colmlb_1 vs. left-handed batters
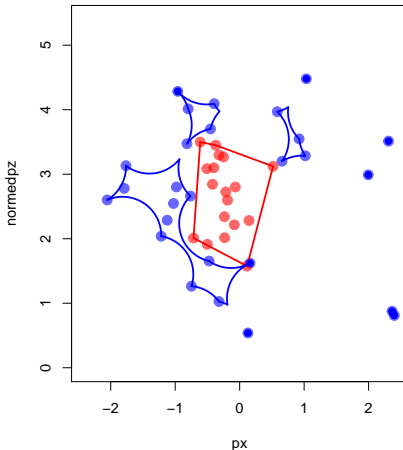
Inconsistency Index Metric

S = established strike zone (convex hull of strikes)

X = established ball zone (α-hull of balls)

$$\text{inconsistency} = \frac{\text{balls in } S + \text{strikes in } X}{\text{total calls}}$$

- This is a *per game* metric.
- Compute separately for right-hand batters and left-hand batters and add.

Inconsistency Index Metric

S = established strike zone (convex hull of strikes)

X = established ball zone (α-hull of balls)

$$\text{inconsistency} = \frac{\text{balls in } S + \text{strikes in } X}{\text{total calls}}$$

- This is a *per game* metric.
- Compute separately for right-hand batters and left-hand batters and add.

Choosing α

- α too small: Ball zone has lots of components
- α too large: Seems unfair to umpires
- α way too large: Ball zone becomes simply connected
- $\alpha = 0.5539$: 1/3 rule-book zone width

Correlations: observations

	Games	ZoneSize	Accuracy	BBrate	Krate	Incon
Games	1.00	-0.05	0.05	-0.03	0.30	-0.02
ZoneSize	-0.05	1.00	-0.49	-0.39	0.09	0.34
Accuracy	0.05	-0.49	1.00	-0.01	0.27	-0.65
BBrate	-0.03	-0.39	-0.01	1.00	-0.24	0.09
Krate	0.30	0.09	0.27	-0.24	1.00	-0.14
Incon	-0.02	0.34	-0.65	0.09	-0.14	1.00

- ▸ Walk rate uncorrelated to inconsistency, accuracy.
- ▸ Smaller zones tend to be more accurate and consistent.
- ▸ Inconsistency is moderately correlated with accuracy.

Correlations: observations

	Games	ZoneSize	Accuracy	BBrate	Krate	Incon
Games	1.00	-0.05	0.05	-0.03	0.30	-0.02
ZoneSize	-0.05	1.00	-0.49	-0.39	0.09	0.34
Accuracy	0.05	-0.49	1.00	-0.01	0.27	-0.65
BBrate	-0.03	-0.39	-0.01	1.00	-0.24	0.09
Krate	0.30	0.09	0.27	-0.24	1.00	-0.14
Incon	-0.02	0.34	-0.65	0.09	-0.14	1.00

- ▶ Walk rate uncorrelated to inconsistency, accuracy.
- ▶ Smaller zones tend to be more accurate and consistent.
- ▶ Inconsistency is moderately correlated with accuracy.

Correlations: observations

|	Games	ZoneSize	Accuracy	BBrate	Krate	Incon
Games	1.00	-0.05	0.05	-0.03	0.30	-0.02
ZoneSize	-0.05	1.00	-0.49	-0.39	0.09	0.34
Accuracy	0.05	-0.49	1.00	-0.01	0.27	-0.65
BBrate	-0.03	-0.39	-0.01	1.00	-0.24	0.09
Krate	0.30	0.09	0.27	-0.24	1.00	-0.14
Incon	-0.02	0.34	-0.65	0.09	-0.14	1.00

- ▶ Walk rate uncorrelated to inconsistency, accuracy.
- ▶ Smaller zones tend to be more accurate and consistent.
- ▶ Inconsistency is moderately correlated with accuracy.

Accuracy vs. Inconsistency ($r = -0.65$)
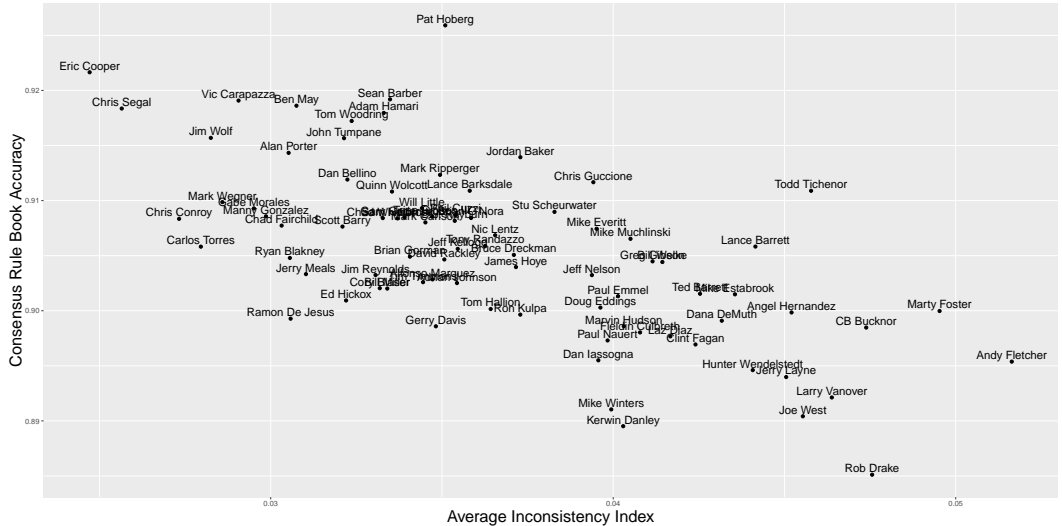
Principal Component Analysis

```
Standard deviations (1, .., p=5):
[1] 1.43 1.21 0.87 0.66 0.54

Rotation (n x k) = (5 x 5):
            PC1    PC2    PC3    PC4    PC5
ZoneSize   0.495  -0.43   0.101  -0.642   0.39
Accuracy  -0.626  -0.13   0.025   0.066   0.77
BBrate    -0.089   0.68  -0.478  -0.536   0.10
Krate     -0.179  -0.56  -0.784  -0.019  -0.21
Incon      0.568   0.16  -0.383   0.543   0.46
```

References

- Data from MLBAM.
- R packages: alphahull, pitchrx, baseballr
- Source code on GitHub: `djhunter/inconsistency`