



Visual R Platform を利用した プロ野球データの解析

東海大学大学院理学研究科

八木 圭太

東海大学理学部

河野 克哉

山本 義郎(指導教員)

はじめに

- 野球の打者の評価として
 - チャンスに強い
 - 左投手に強いなど、印象による評価をされる場合がある。
- そのような選手の特徴を数値化し、その数値で分類をする。

はじめに

- どんな指標で打者を評価するか？

日本のプロ野球では打率や打点、ホームラン数などで打者の評価を行うことが多い。

日本野球機構
<http://www.npb.or.jp/>



2012年度 セントラル・リーグ



リーダース(打撃部門)



2012年度 パシフィック・リーグ



リーダース(打撃部門)

■ 全日程終了

■ 全日程終了

打率

1 阿部 慎之助	(巨)	340
2 坂本 勇人	(巨)	311
3 大島 洋平	(中)	310
4 長野 久義	(巨)	301
5 ミレッジ	(ヤ)	300
5 ラミレス	(デ)	300
7 川端 慎吾	(ヤ)	298
8 和田 一浩	(中)	285
9 井端 弘和	(中)	284
10 中村 紀洋	(デ)	274

打率

1 角中 勝也	(口)	312
2 中島 裕之	(西)	311
3 糸井 嘉男	(日)	304
4 内川 聖一	(ノ)	3001
5 田中 賢介	(日)	2997
6 秋山 翔吾	(西)	293
7 稲葉 篤紀	(日)	290
8 栗山 巧	(西)	289
9 陽 岱綱	(日)	287
10 李 大浩	(才)	286

打点

日本野球機構
<http://www.npb.or.jp/>

2012年度 セントラル・リーグ

リーダーズ(打撃部門)

■ 全日程終了

打点		
1 阿部 慎之助	(巨)	104
2 バレンティン	(ヤ)	81
3 ラミレス	(テ)	76
4 坂本 勇人	(巨)	69
5 ミレッジ	(ヤ)	65
5 ブランコ	(中)	65
7 和田 一浩	(中)	63
8 中村 紀洋	(テ)	61
9 長野 久義	(巨)	60
10 鳥谷 敬	(神)	59



2012年度 パシフィック・リーグ

リーダーズ(打撃部門)

■ 全日程終了

打点		
1 李 大浩	(オ)	91
2 中村 剛也	(西)	79
3 中田 翔	(日)	77
4 ベーニャ	(ソ)	76
5 中島 裕之	(西)	74
6 角中 勝也	(口)	61
6 稲葉 篤紀	(日)	61
8 ヘルマン	(西)	60
8 井口 資仁	(口)	60
10 松田 宣浩	(ソ)	56
10 T-岡田	(オ)	56

本塁打数

日本野球機構

<http://www.npb.or.jp/>

2012年度 セントラル・リーグ



リーダーズ(打撃部門)

■ 全日程終了

本塁打

1	バレンティン	(ヤ)	31
2	阿部 慎之助	(巨)	27
3	ブランコ	(中)	24
4	ミレッジ	(ヤ)	21
5	ラミレス	(デ)	19
6	坂本 勇人	(巨)	14
6	長野 久義	(巨)	14
6	堂林 翔太	(広)	14
9	畠山 和洋	(ヤ)	13
10	村田 修一	(巨)	12
10	ブラゼル	(神)	12



2012年度 パシフィック・リーグ



リーダーズ(打撃部門)

■ 全日程終了

本塁打

1	中村 剛也	(西)	27
2	李 大浩	(オ)	24
2	中田 翔	(日)	24
4	ペーニャ	(ソ)	21
5	ホフパワー	(日)	14
6	中島 裕之	(西)	13
7	井口 資仁	(口)	11
8	稲葉 篤紀	(日)	10
8	T-岡田	(オ)	10
8	バルデリス	(オ)	10

評価に使う指標

- 得点圏打率があるように、日本では特に打率が使われることが多いように思われるので、打率を用いて選手の特徴の数値化を行ってもよいが...
- しかしながら、打率では本塁打と安打との区別がない。同じ状況でも単打と本塁打とでは当然印象は異なる。
- そこで打撃の総合指標であるOPSに着目した。

OPSとは...

- OPS(On base Plus Slugging percentage)とは、野球を統計的に分析する手法であるセイバーメトリクスの指標の一つである。
- OPSは長打率(Slug)と出塁率(OB)の和で定義される。

$$\text{OPS} = \text{Slug} + \text{OB}$$

ここで

$$\text{Slug} = \frac{\text{塁打数}}{\text{打数}}, \text{OB} = \frac{\text{安打数} + \text{四死球数}}{\text{打数} + \text{四死球数} + \text{犠飛打数}}$$

である。

OPSを用いた新たな指標を考える。

状況別OPS

- 打者の特徴をとらえるために**状況別OPS**という指標を考える。
- 状況別OPSとは、ある特定の状況の打撃結果からOPSを算出するもの。
 - 例えば対戦投手が右投げの場合のOPSである OPS_{H_R} は以下の様に定義する。

$$OPS_{H_R} = Slug_{H_R} + OB_{H_R}$$

ここで $Slug_{H_R}$ と OB_{H_R} はそれぞれ対戦投手が右投げのときの打席結果から算出した長打率と出塁率である。

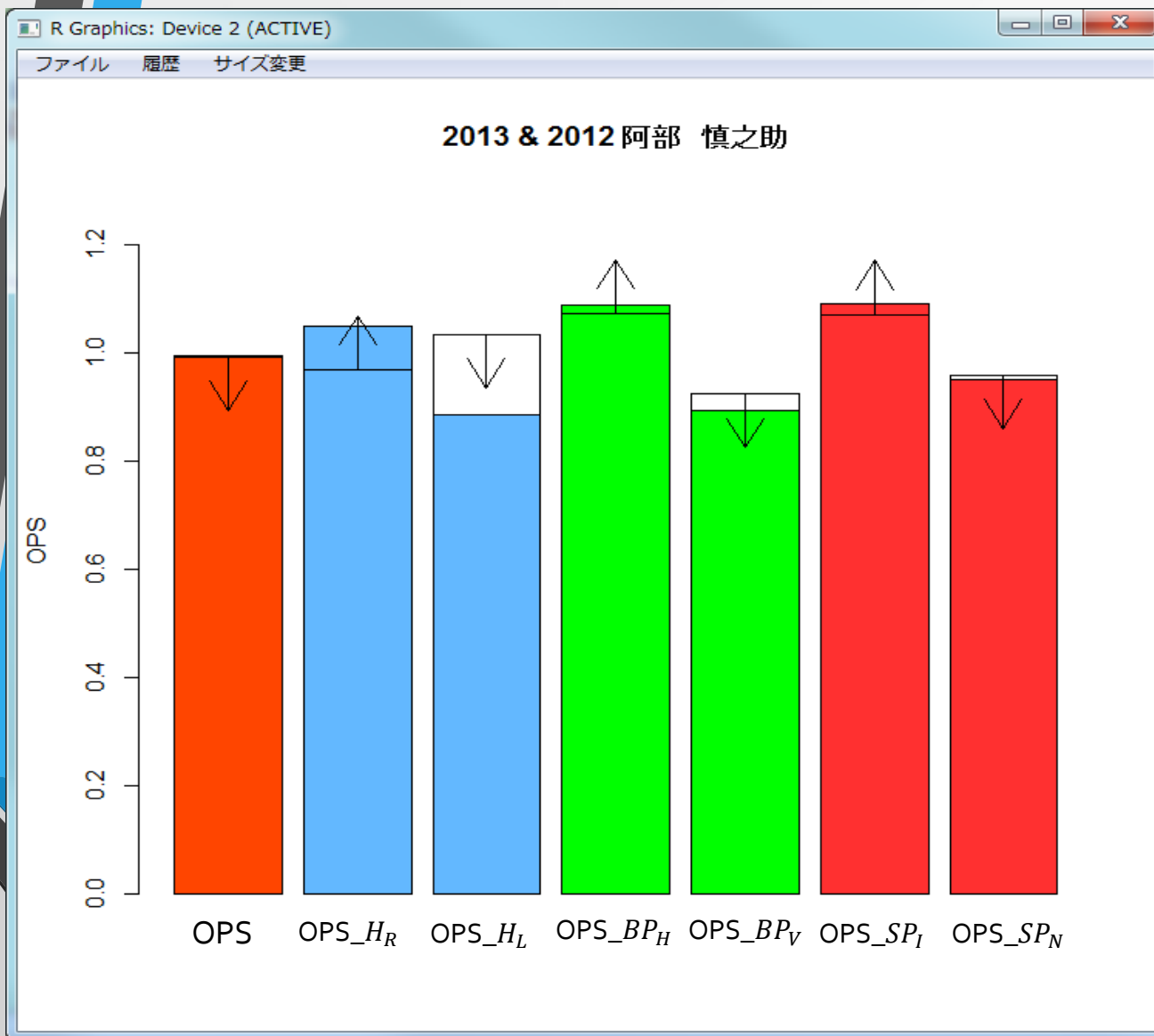
打席の状況

- 今回考える打席の状況
 1. 相手投手が右投げ(OPS_{H_R})か左投げ(OPS_{H_L})か
 2. その試合が打者にとってホーム(OPS_{BP_H})かビジター(OPS_{BP_V})か
 3. 得点圏に走者がいる(OPS_{SP_I})かいない(OPS_{SP_N})か
- 通常のOPSと6個の状況別OPSで選手の特徴を考える。

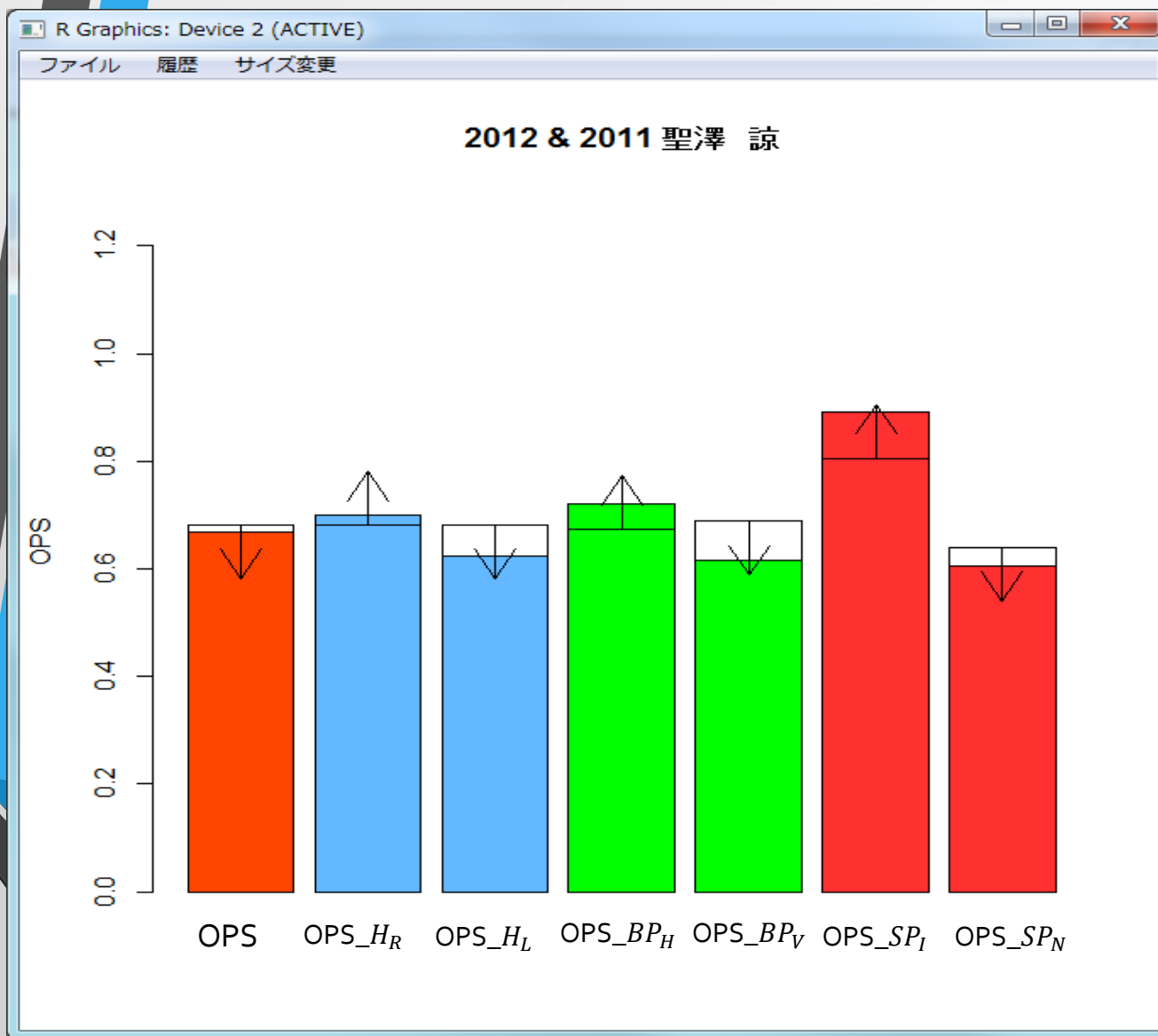
2012年の200打席到達選手の状況別OPS

	OPS	OPS _{HR}	OPS _{HL}	OPS _{BP_H}	OPS _{BP_V}	OPS _{SP_I}	OPS _{SP_N}
1	阿部 (巨人) 0.9938	ホワイトセル (ロッテ) 0.9939	バレンティン (ヤクルト) 1.1579	阿部 (巨人) 1.0718	阿部 (巨人) 0.9240	バレンティン (ヤクルト) 1.1544	阿部 (巨人) 0.9587
2	バレンティン (ヤクルト) 0.9584	阿部 (巨人) 0.9676	阿部 (巨人) 1.0339	ホワイトセル (ロッテ) 1.0306	バレンティン (ヤクルト) 0.9098	清田 (ロッテ) 1.0795	新井良 (阪神) 0.8986
3	ホワイトセル (ロッテ) 0.8856	ブランコ (中日) 0.9208	坂本 (巨人) 0.9945	バレンティン (ヤクルト) 1.0088	李 大浩 (オリックス) 0.88969	阿部 (巨人) 1.0702	バレンティン (ヤクルト) 0.8720
4	ミレッジ (ヤクルト) 0.8646	バレンティン (ヤクルト) 0.9004	ラミレス (横浜) 0.9625	ミレッジ (ヤクルト) 0.9597	ラミレス (横浜) 0.88964	T-岡田 (オリックス) 1.0009	ホワイトセル (ロッテ) 0.8671
5	ブランコ (中日) 0.8510	柘田 (楽天) 0.8919	新井良 (阪神) 0.9619	ブランコ (中日) 0.9151	ペーニャ (ソフトバンク) 0.8884	丸 (広島) 0.9717	中島 (西武) 0.8501

この年大変活躍された巨人の阿部選手はどの状況別OPSも高い値になっている。同じく活躍したバレンティン選手はOPS_{SP_I}が極めて高く、勝負どころで特に力を発揮した選手であるといえる。



ある選手について2年間の成績を比較するプロットを作成した。この図は阿部選手についてのプロットで、色のついたグラフは新しい年の成績になっており、この場合は2013年の成績である。2012年と比べ、左投手に対するOPSは下がっているが右投手に対するOPSは上がっている。しかしながら全体的に高いOPSを示しているので状況に関わらず高い成績を残せたことがわかる。



聖澤選手の2012年と2011年のプロットである。聖澤選手はOPSの高い選手とは言えないが、2012年の OPS_{SP_I} が他の状況別OPSと比べてとても高い。得点圏にランナーがいる場合に力を発揮する勝負強い選手であるとわかる。



中村(紀)選手の2013年と2012年のプロットである。中村選手は2012年に比べ下がってしまっているが左投手を得意とする打者であることがわかる。2012年と比べ2013年はチャンスでのOPSが下がってしまっており、期待されるようなバッティングができなかったと思われる。

- 印象に近い数値化ができたので、この数値を用いてクラスター分析で分類を行う。
- 状況別OPSは現在対になっているため、状況別OPS差を定義し、分析変数として用いる。

状況別OPS差

それぞれの状況別OPS差は以下のように定義する。

1. 利き腕OPS差： $DOPS_{hand}$

$$DOPS_{hand} = \begin{cases} OPS_{H_R} - OPS_{H_L} & \text{(右打者)} \\ OPS_{H_L} - OPS_{H_R} & \text{(左打者)} \end{cases}$$

2. ホームビジターOPS差： $DOPS_{HV}$

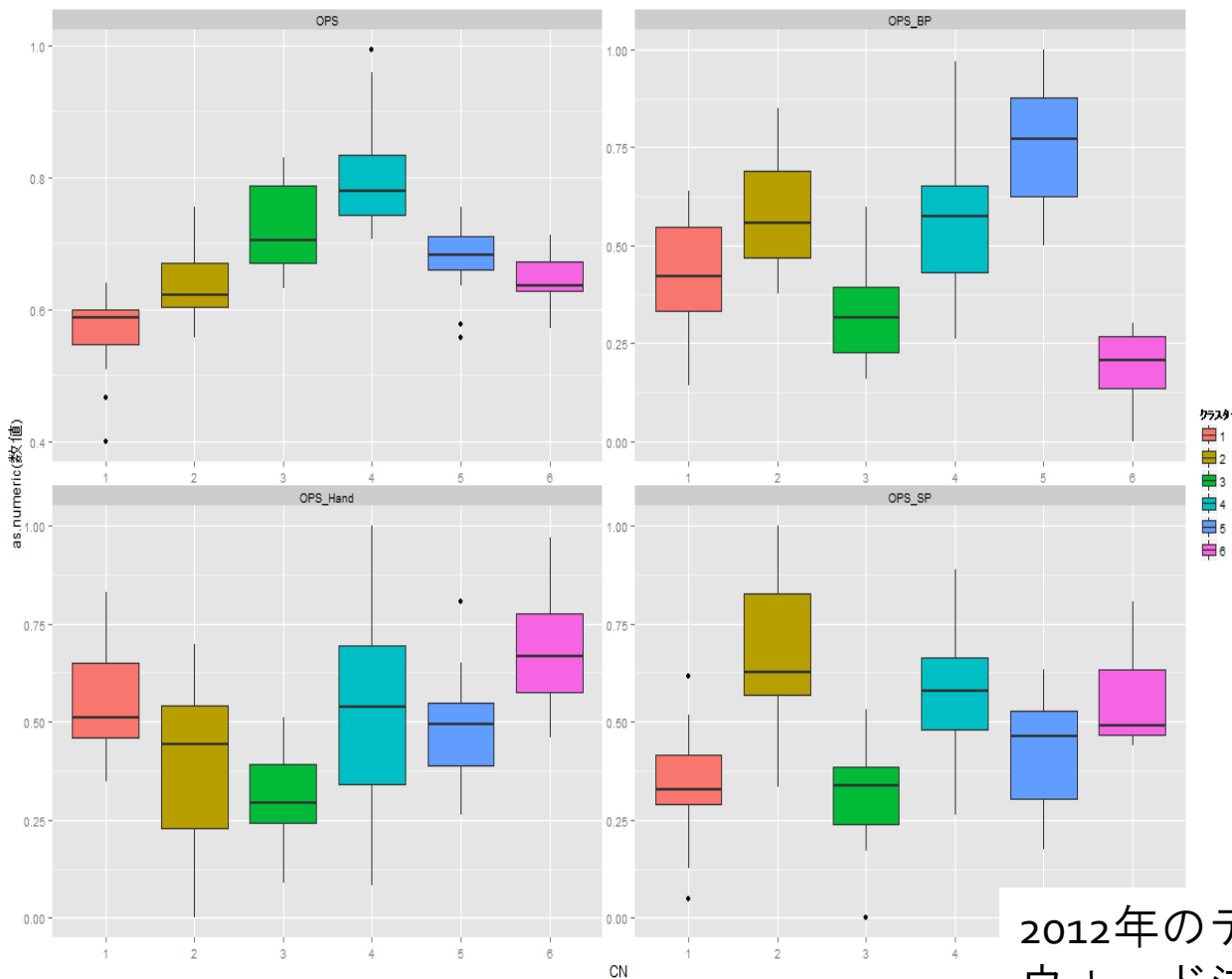
$$DOPS_{HV} = OPS_{BP_H} - OPS_{BP_V}$$

3. 得点圏OPS差： $DOPS_{SP}$

$$DOPS_{SP} = OPS_{SP_I} - OPS_{SP_N}$$

この3変数に通常のOPSを加えた4変数を基準化したものを分析変数に用いてクラスター分析を行う。ただし通常のOPSはそもそも成績が良いかどうか判断するのに重要なので基準化したものを2倍にする。

変数ごとにボックスプロット

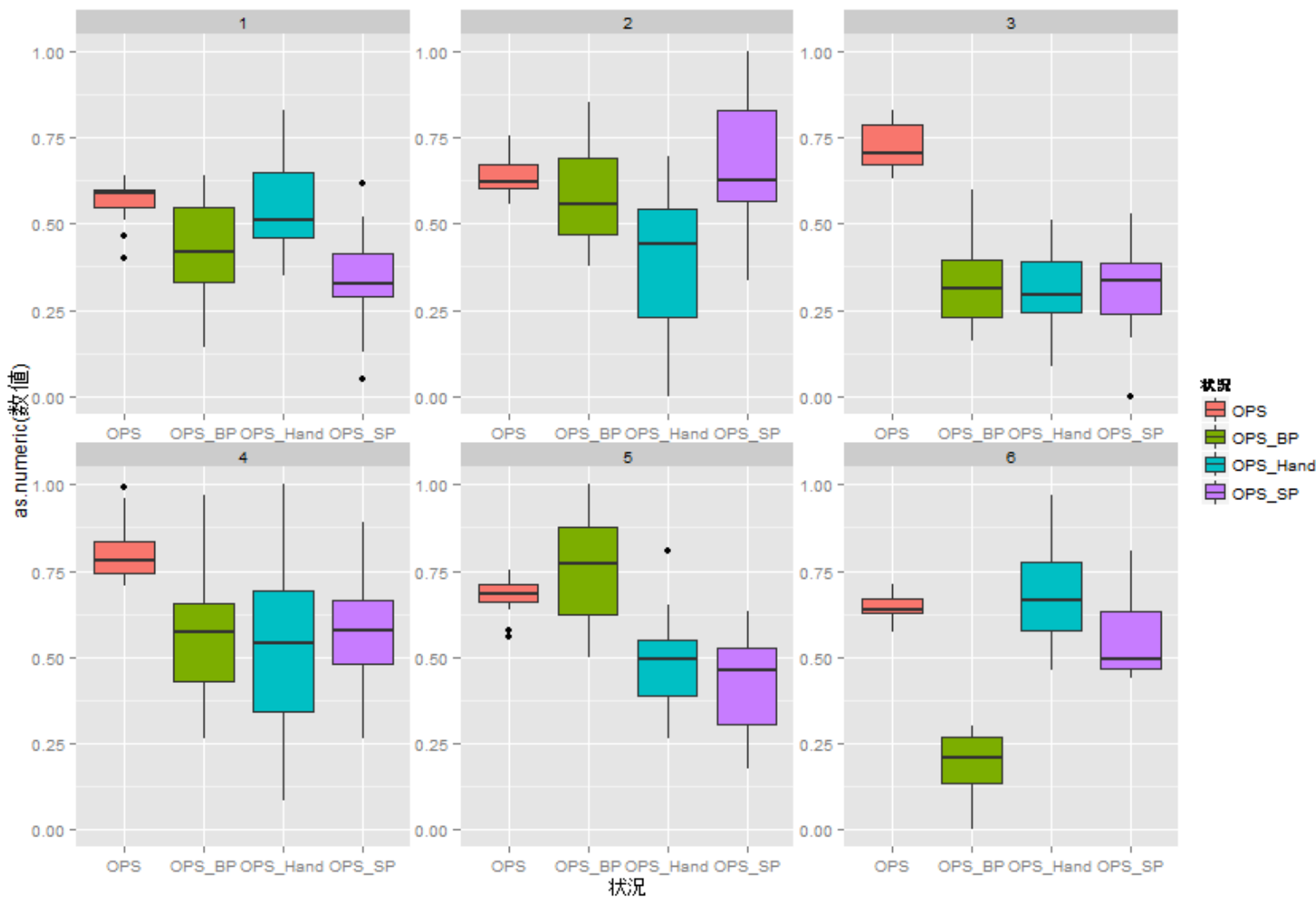


各変数について、クラスター毎のボックスプロットである。分類変数としてはOPSを基準化したものを2倍にしたものを用いているが、表示は基準化したものであることに注意をする。

2012年のデータ
ワード法で6クラスターに分けたとき

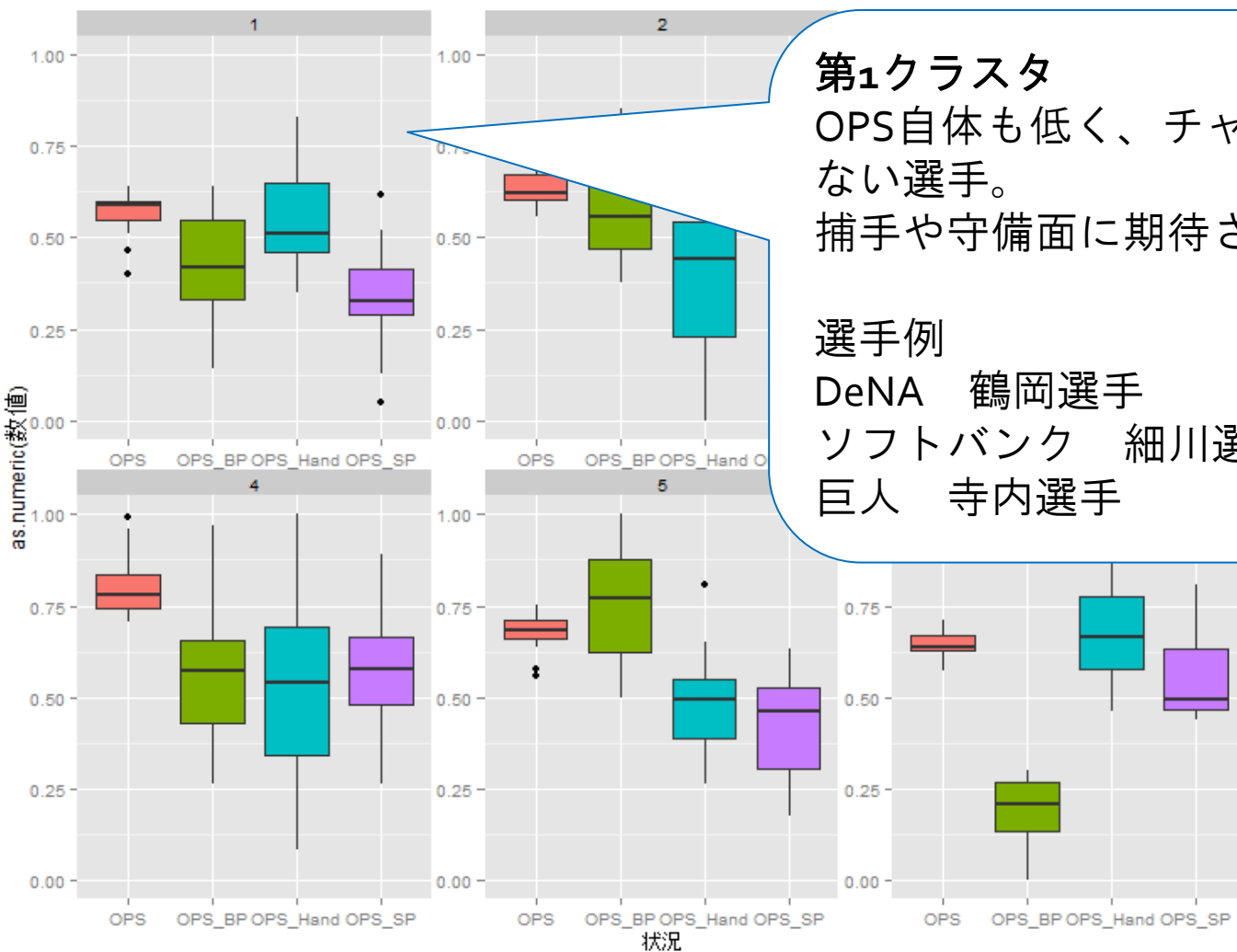
クラスターごとにボックスプロット

2012年のデータ
ワード法で6クラスターに分けたとき



クラスター毎の特徴を見るため、先ほどのプロットを今度は各クラスター毎にまとめた。
この図について考察をしていく。

クラスターごとに各変数を並べる



第1クラスター

OPS自体も低く、チャンスにもあまり打てない選手。
捕手や守備面に期待される選手が多い。

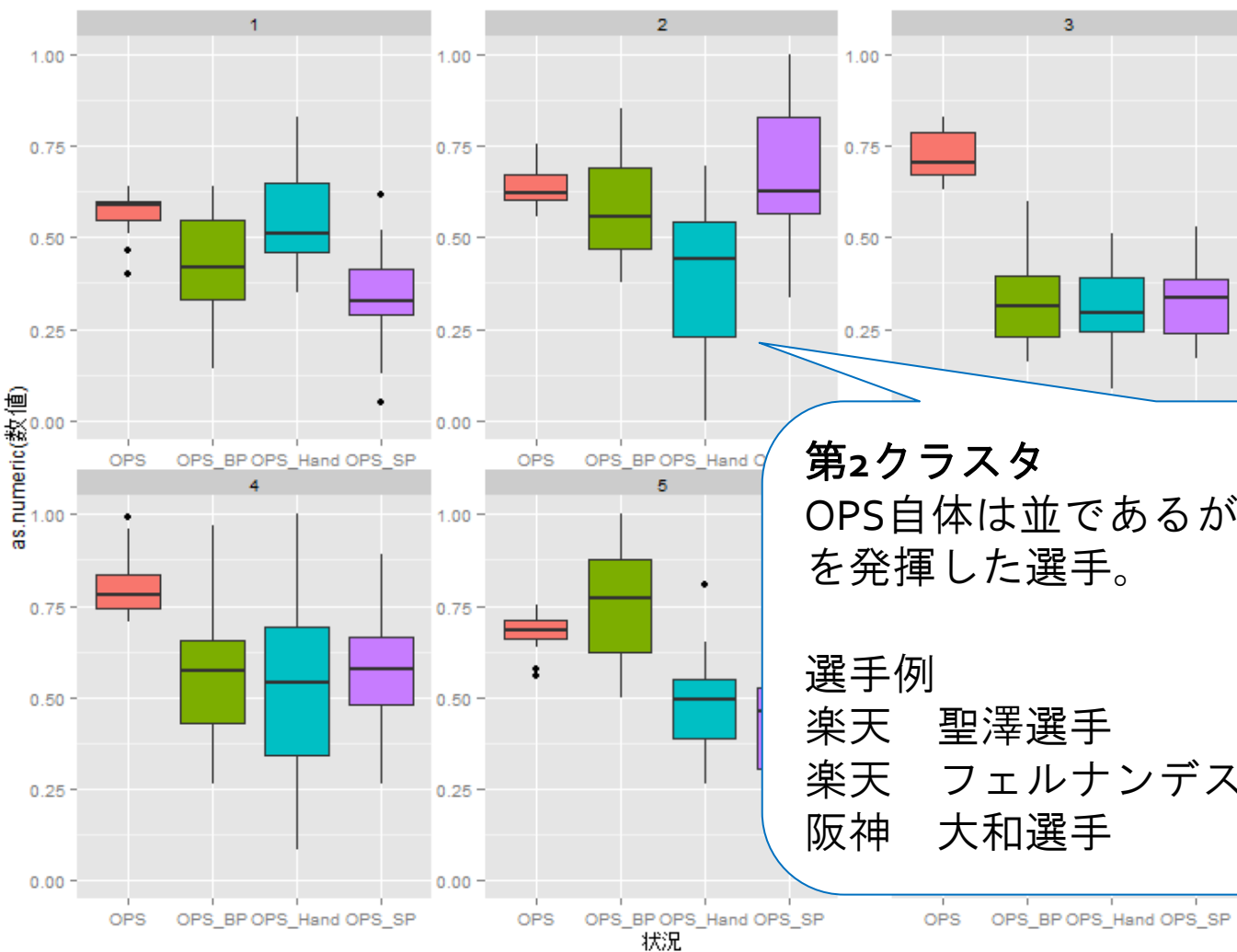
選手例

DeNA 鶴岡選手

ソフトバンク 細川選手

巨人 寺内選手

クラスターごとに各変数を並べる



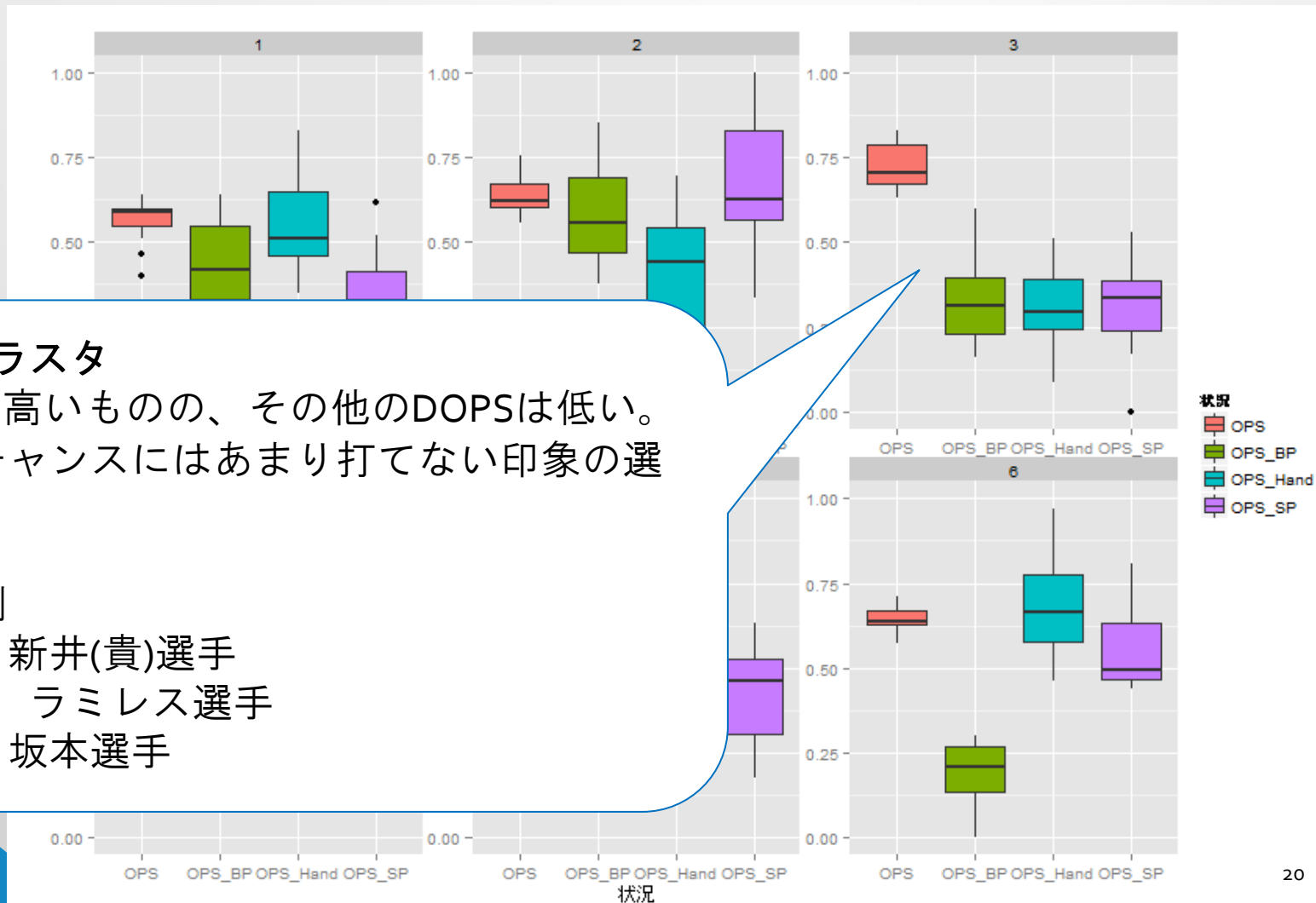
第2クラスター

OPS自体は並であるが、チャンスにその力を発揮した選手。

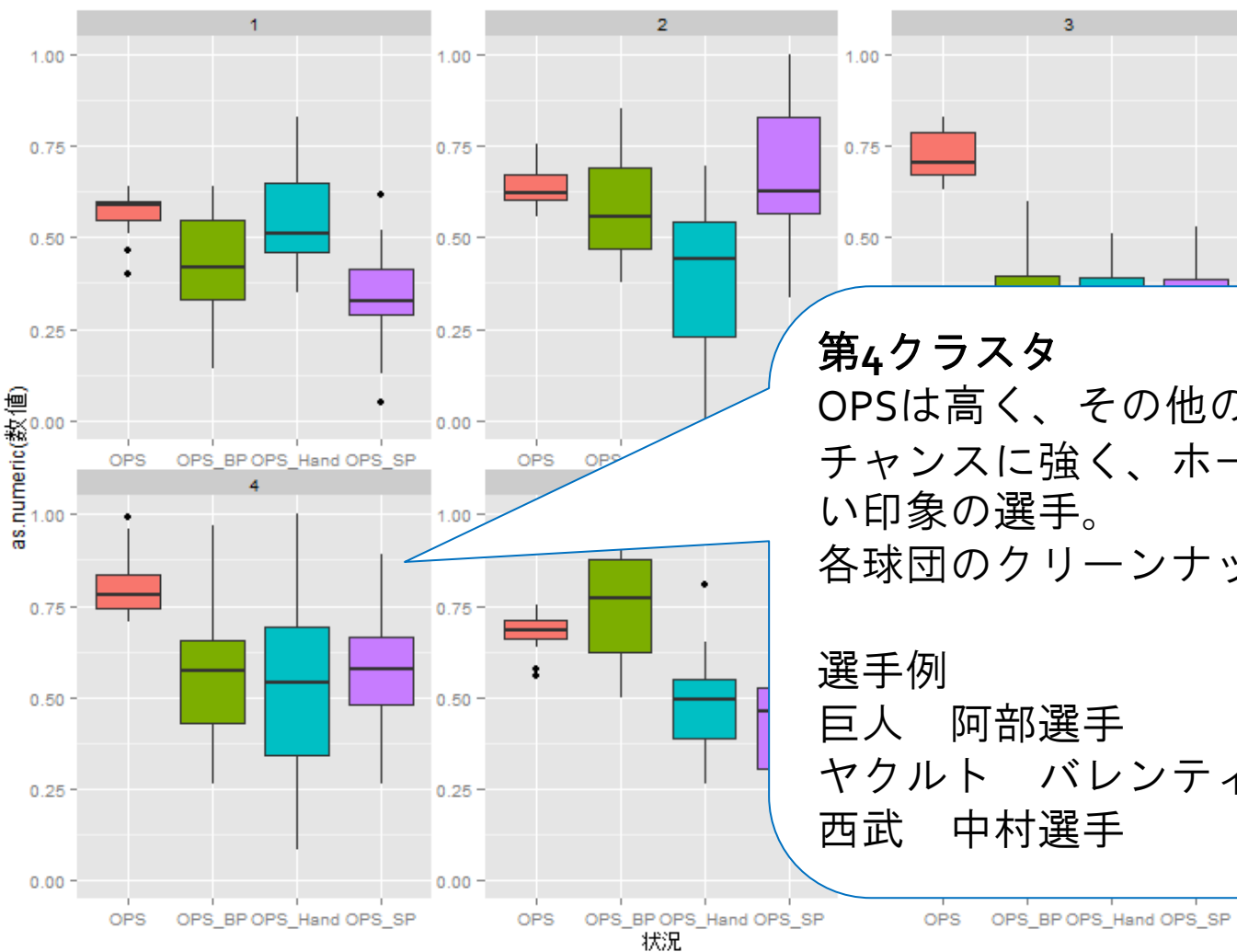
選手例

楽天 聖澤選手
 楽天 フェルナンデス選手
 阪神 大和選手

クラスターごとに各変数を並べる



クラスターごとに各変数を並べる



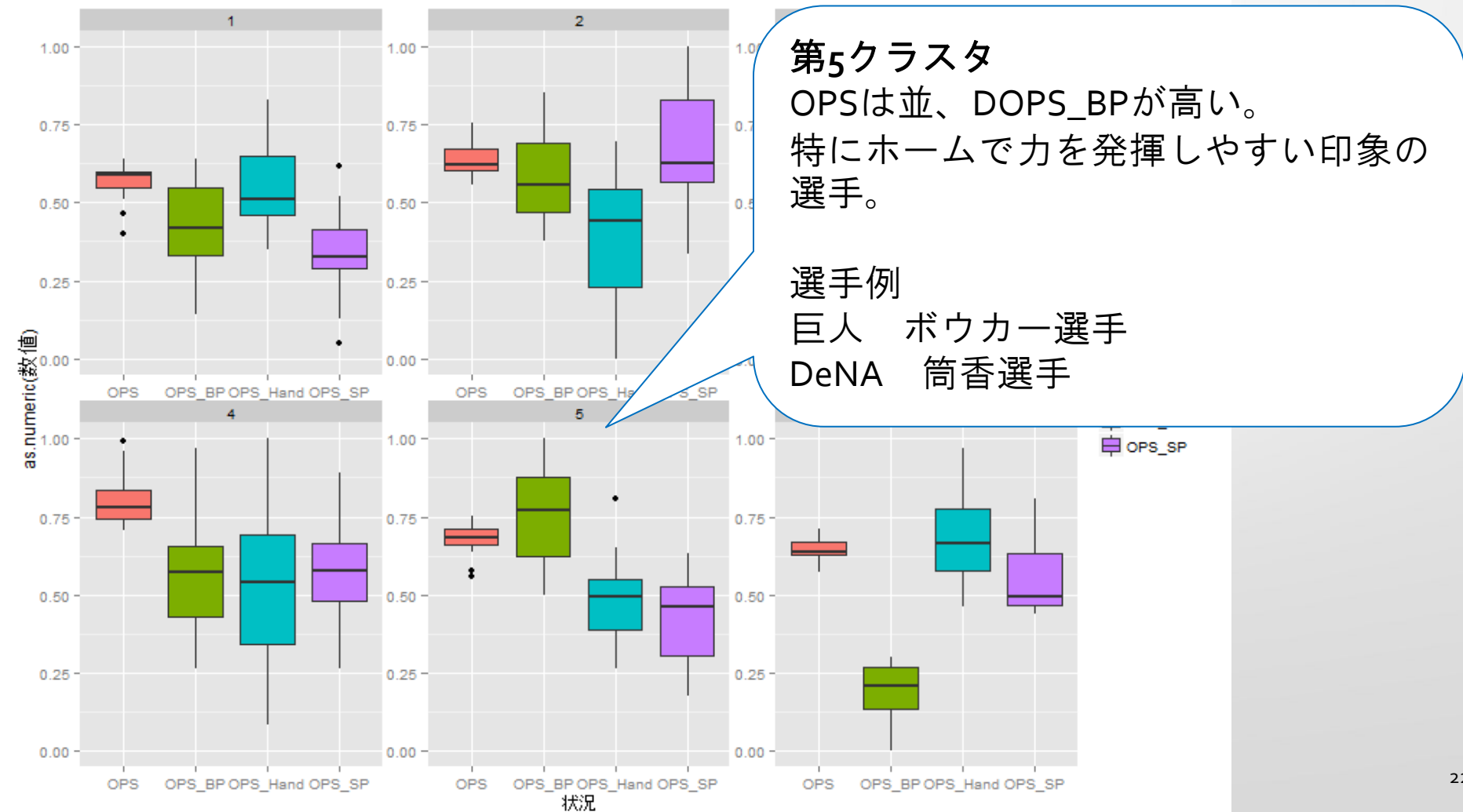
第4クラスター

OPSは高く、その他のDOPSも高い。
チャンスに強く、ホームで力を発揮しやすい印象の選手。
各球団のクリーンナップの選手が多い。

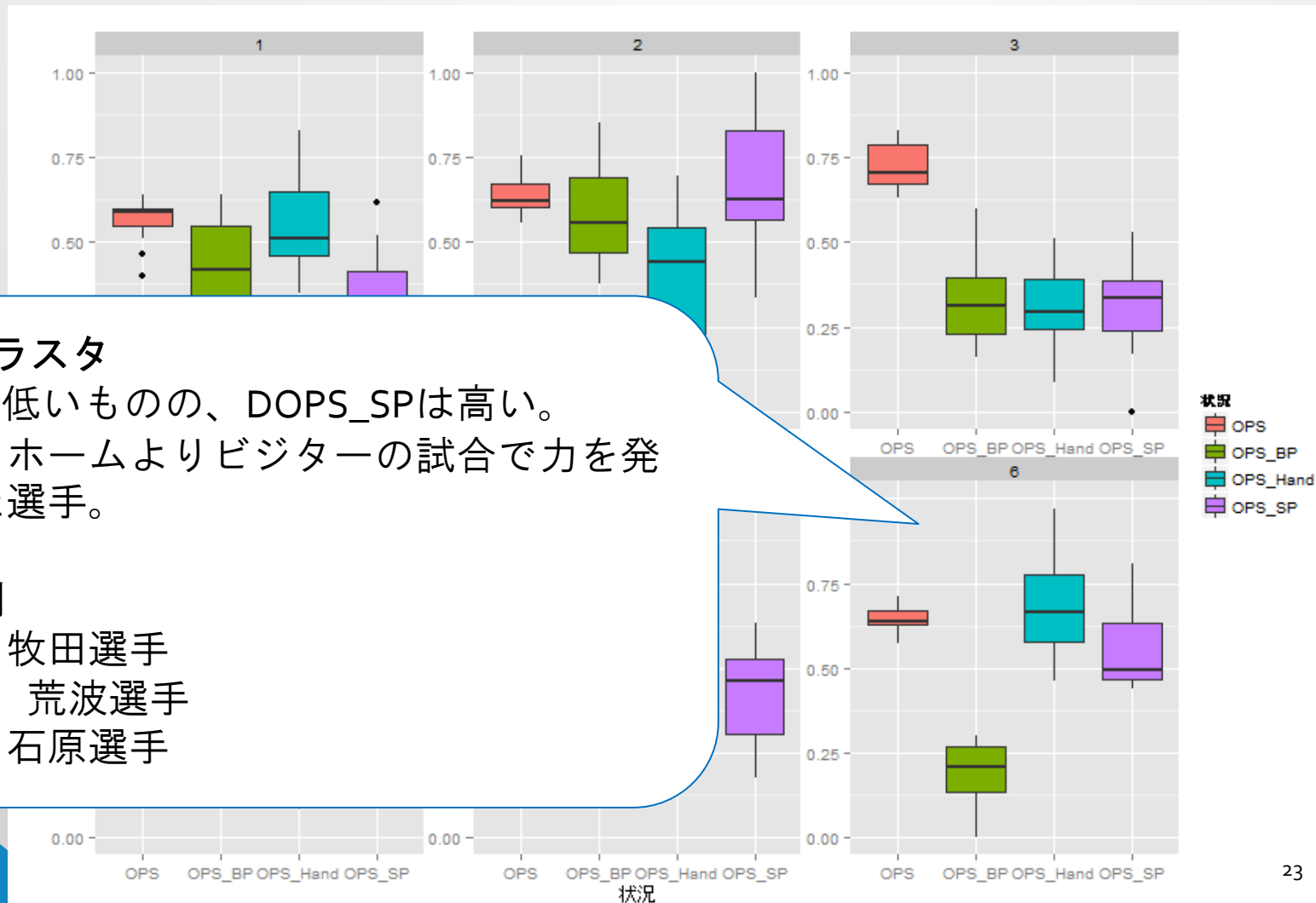
選手例

巨人 阿部選手
ヤクルト バレンティン選手
西武 中村選手

クラスターごとに各変数を並べる



クラスターごとに各変数を並べる



- 状況別OPSによって各選手の特徴を数値化し、視覚的にみることのできるアプリケーションを作成した。
- DOPS(状況別OPS差)を定義し、クラスター分析することで選手の特徴について分類することができた。
- 以前Rを用いて作成していたアプリケーションを、データを入れ替えるだけで作成者でない人にも編集しやすくなった。

参考文献

- Jim Albert, Jay Bennet (2013). Curve Ball: Baseball, Statistics, and the Role of Chance in the Game.
- 船山貴光, 八木圭太, 宗像昌平, 池上拓人, 亀岡瑤, 宮前貴基, 山本義郎(2014). 選手の特徴および調子の波を把握する為の修正OPSとその活用例. 統計数理研究所共同研究リポート314, p51-54
- 山本義郎, 飯塚誠也, 藤野友和(2013). Rで学ぶデータサイエンス 12 統計データの視覚化 (金明哲, 編). 共立出版.
- Winston Chang (2013). Rグラフィックス クックブック ggplot2によるグラフ作成レシピ集(石井弓美子, 河内崇, 瀬戸山雅人, 古畠敦, 訳). オーム社