

問題解決技法入門

Rでデータの視覚化

堀田 敬介

Rでデータの視覚化

- csv ファイルをデータとして利用
 - 「マイドキュメント(Y:)」に「R」フォルダをつくり中に保存

bb2015.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		リーグ	試合数	勝数	負数	引分数	勝率	得点	失点	本塁打	盗塁	打率	防御率
2	巨人	セ	144	82	61	1	0.573	596	552	144	102	0.257	3.58
3	阪神	セ	144	75	68	1	0.524	599	614	94	55	0.264	3.88
4	広島	セ	144	74	68	2	0.521	649	610	153	96	0.272	3.79
5	中日	セ	144	67	73	4	0.479	570	590	87	75	0.258	3.69
6	DeNA	セ	144	67	75	2	0.472	568	624	121	76	0.253	3.76
7	ヤクルト	セ	144	60	81	3	0.426	667	717	139	62	0.279	4.62
8	ソフトバンク	パ	144	78	60	6	0.565	607	522	95	124	0.28	3.25
9	オリックス	パ	144	80	62	2	0.563	584	468	110	126	0.258	2.89
10	日本ハム	パ	144	73	68	3	0.518	593	569	119	134	0.251	3.61
11	ロッテ	パ	144	66	76	2	0.465	556	642	96	64	0.251	4.14
12	西武	パ	144	63	77	4	0.45	574	600	125	74	0.248	3.77
13	楽天	パ	144	64	80	0	0.444	549	604	78	64	0.255	3.97

※) 2015年プロ野球セ・パ成績
(Yahoo Japan! Sports naviより)

- ファイルの読み込み

※1行目にheaderあり ※各行の名称は列1に

```
> dfbb <- read.csv("Y:/R/bb2015.csv", header=T, row.names=1)
```

※ファイルのフルパス

例) YドライブのRフォルダ内にあるbb2015.csvという名前のファイル

Rでデータの視覚化

● 棒グラフを作成

※色指定用のベクトル生成. "royalblue"を6回 repeat し,
"violetred"を6回repeatしたベクトルをつくり cc に代入

```
> cc <- c(rep("royalblue",6), rep("violetred",6))  
> barplot(dfbb[,3], names.arg=row.names(dfbb), col=cc, xlab="チーム名", ylab="勝数")
```

dfbb[,3] ... data.frameである dfbb の3列目 ("勝数")を棒グラフのデータとして使用
names.arg ... それぞれの棒に対応する名称

col ... 棒の色指定

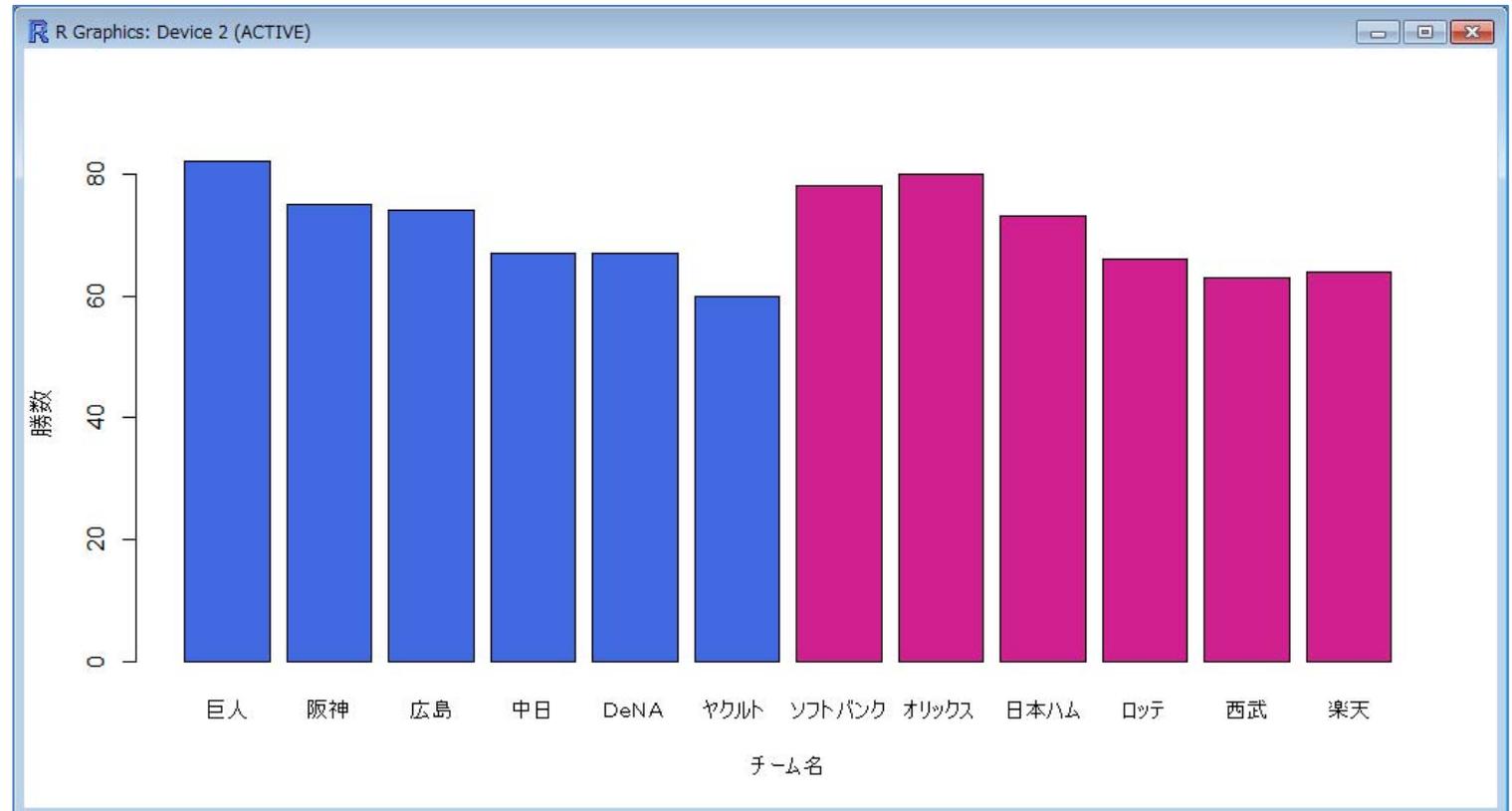
xlab ... x軸のラベル

ylab ... y軸のラベル

● Tips !

```
> colors()
```

※Rで使える657色の名称リスト表示

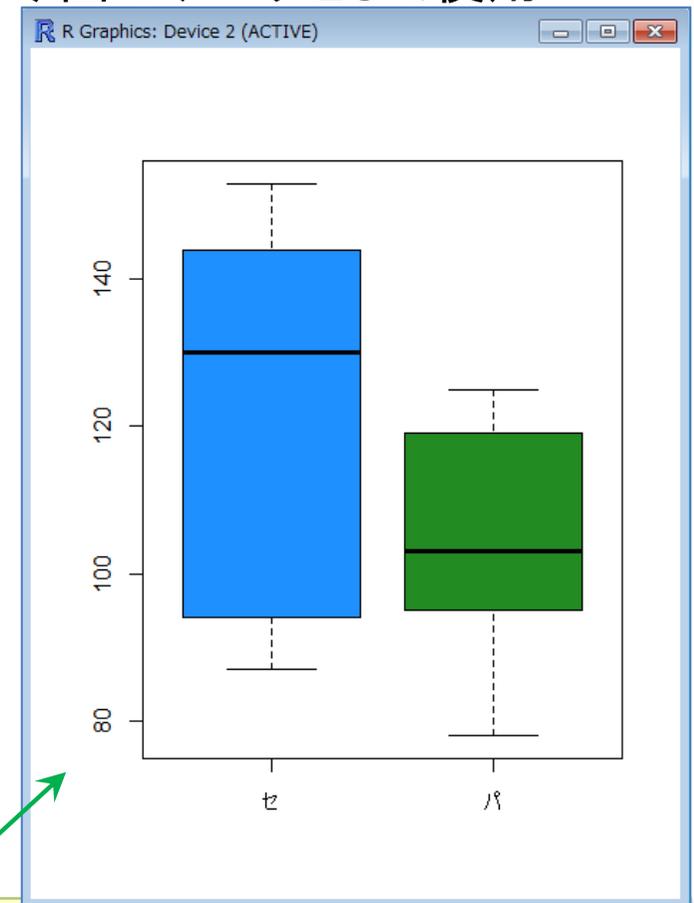
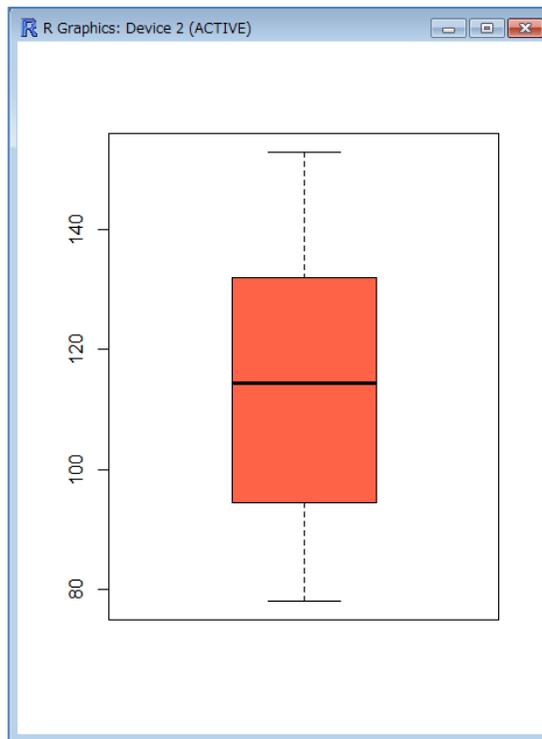


Rでデータの視覚化

- 箱ひげ図を描画

```
> boxplot(dfbb[,9], col="tomato")
```

dfbb[,9] ... data.frameである dfbb の9列目 ("本塁打") を箱ひげ図のデータとして使用
col ... 箱の色指定



- グループ毎に箱ひげ図をかく

```
> boxplot(dfbb[,9]~dfbb[,1], col=c("dodgerblue","forestgreen"))
```

※「リーグ(セ・パ)」毎に描画するよう指定

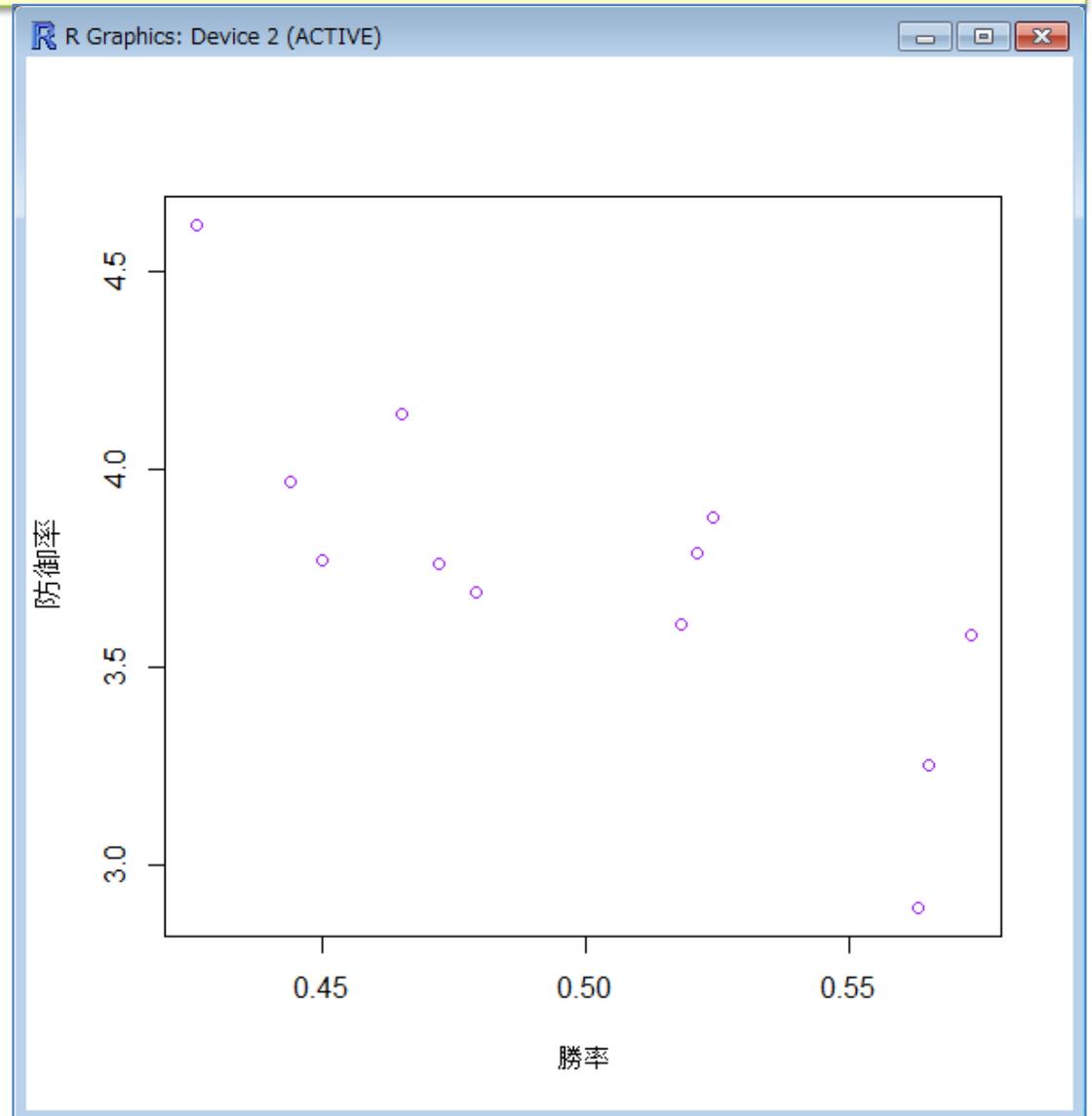
Rでデータの視覚化

- 散布図を作成(1)

```
> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", col="purple")
```

x軸を dfbb[,6]="勝率"
y軸を dfbb[,12]="防御率"
のデータを用い散布図を作成

xlab ... x軸ラベルの指定
ylab ... y軸ラベルの指定
col ... プロットする点の色指定



Rでデータの視覚化

● 散布図を作成(2)

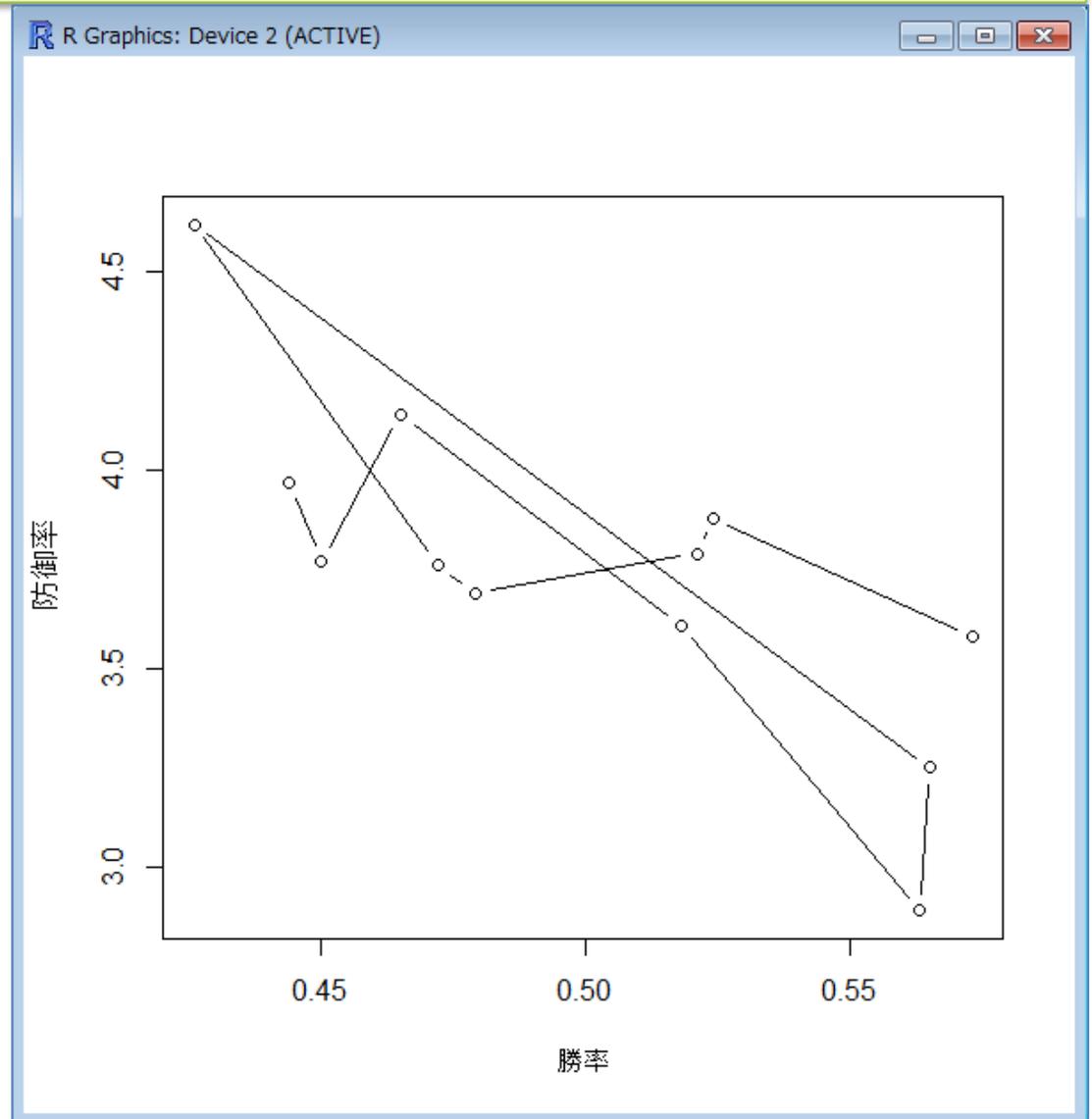
```
> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", type="b")
```

x軸を dfbb[,6]="勝率"
y軸を dfbb[,12]="防御率"
のデータを用い散布図を作成

xlab ... x軸ラベルの指定
ylab ... y軸ラベルの指定

type ... 描画点の種類

- "p" ... points 点 (default)
- "l" ... lines 線分
- "b" ... both 点と線分 両方
- "c" ... "b" から点を抜いたもの
- "o" ... overplotted
- "h" ... histogram
- "s" ... stair steps
- "n" ... no plotting 点をかかない



Rでデータの視覚化

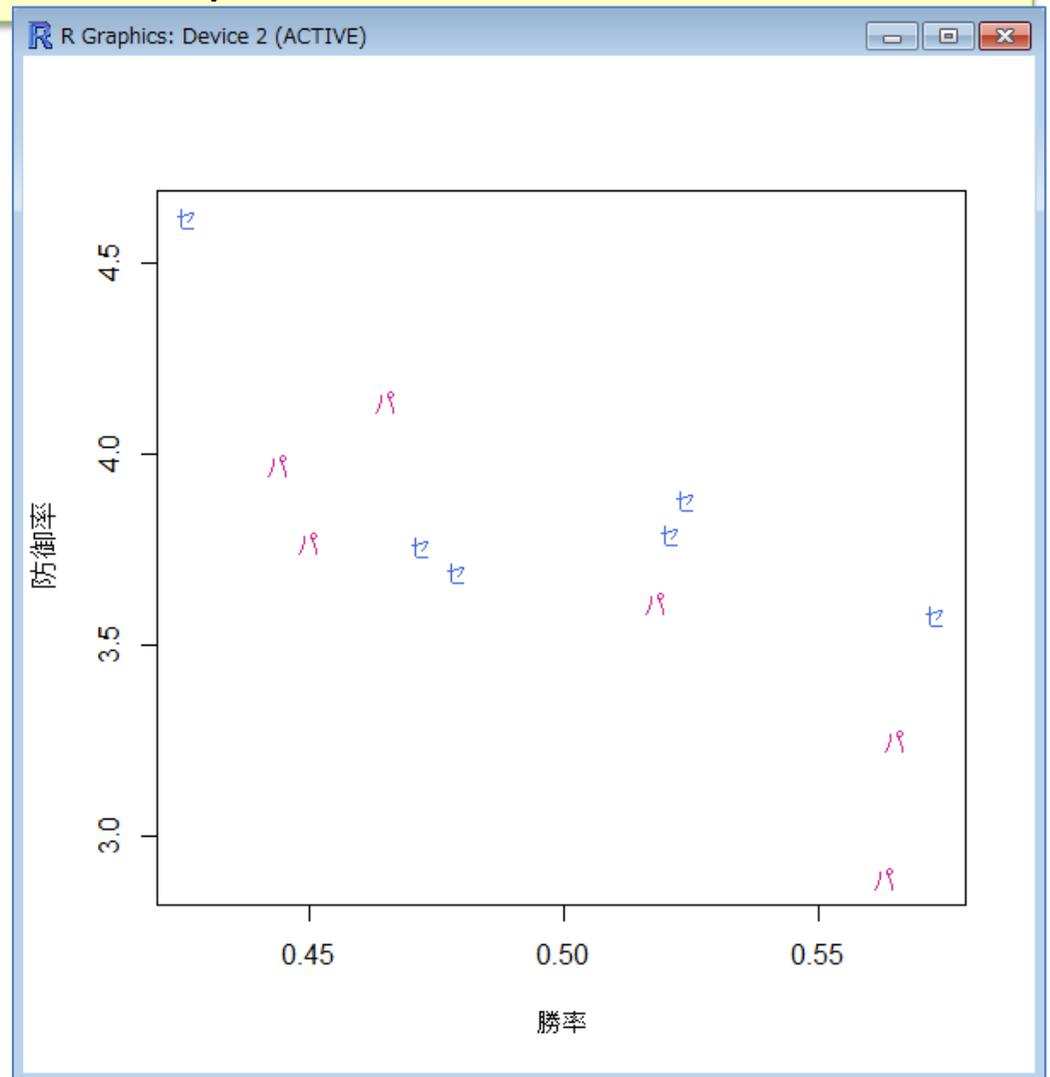
● 散布図を作成(3)

※プロットはせずに、枠・軸だけを描画

```
> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", type="n")  
> text(dfbb[,6], dfbb[,12], dfbb[,1], col=cc)
```

※リーグ名称をプロット点としてかく
(data.frame である dfbb の1列目に
リーグ名を入れたことを思いだそう！)

※col=cc は色設定を ccにすること
(cc はリーグ毎の色設定用ベクトル
として作ったことを思いだそう！)



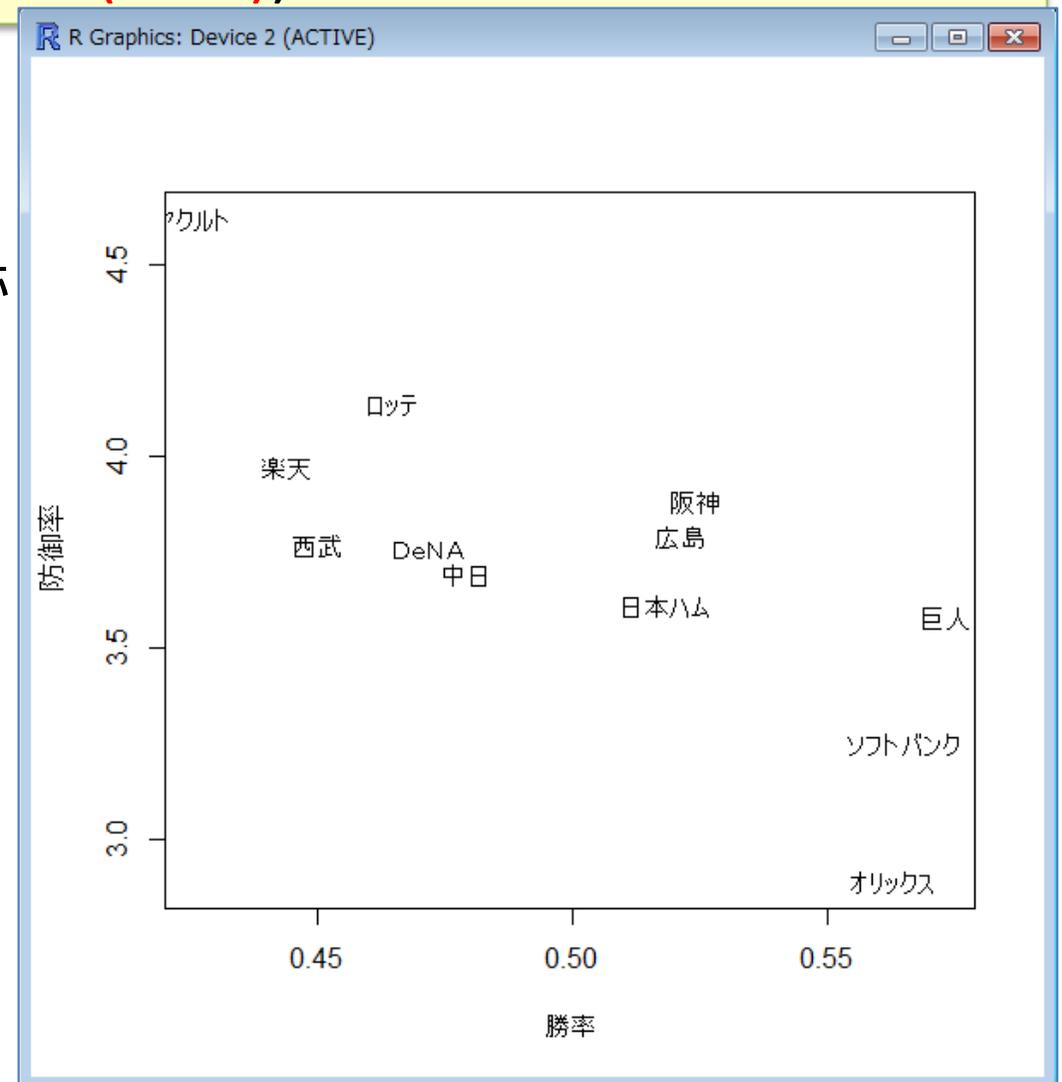
Rでデータの視覚化

● 散布図を作成(4)

※プロットはせずに、枠・軸だけを描画

```
> plot(dfbb[,6], dfbb[,12], xlab="勝率", ylab="防御率", type="n")  
> text(dfbb[,6], dfbb[,12], row.names(dfbb))
```

※チーム名称をプロット点としてかく
(read.csvでcsvファイルを読み込んだ時に、row.namesとして1列目のチーム名称を指定したことを思いだそう!)



Rでデータの視覚化

● 箱ひげ図と散布図を作成(1) -scatterplot()-

```
> install.packages("car")  
> library(car)
```

※scatterplot() の使用準備

← package "car" のインストール

← package "car" の読み込み

```
> scatterplot(dfbb[,4], dfbb[,8], xlab="負数", ylab="失点")
```

x軸を dfbb[,4]="負数"

y軸を dfbb[,8]="失点"

のデータを用い散布図を作成

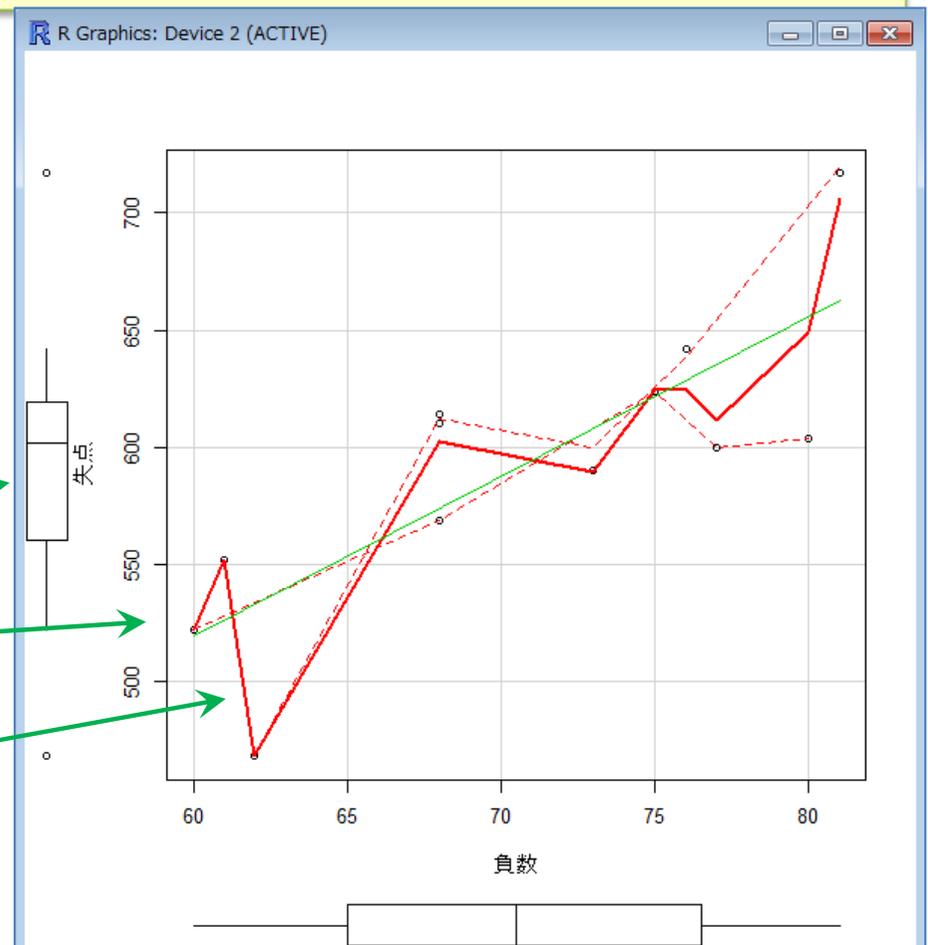
xlab ... x軸ラベルの指定

ylab ... y軸ラベルの指定

※それぞれの軸に、それぞれのデータの箱ひげ図が描かれる

※緑線は回帰直線 regression line

※赤線・赤点線は平滑化線とspan



Rでデータの視覚化

● 箱ひげ図と散布図を作成(2) -scatterplot()-

```
> install.packages("sp")  
> install.packages("maptools")  
> library(sp)  
> library(maptools)
```

※pointLabel() の使用準備

packages "sp","maptools"のインストール

packages "sp","maptools"の読み込み

(注:必ず sp → maptools の順!)

－ 点とチーム名を両方プロットする

```
> scatterplot(dfbb[,4], dfbb[,8], xlab="負数", ylab="失点", reg.line=F,  
smooth=F)  
> pointLabel(x=dfbb[,4], y=dfbb[,8], labels=row.names(dfbb))
```

※平滑化線は描かない

※散布図の点のラベルを
row.names(dfbb)として書く

※回帰直線 regression line
は描かない(FはFalseの意)

Rでデータの視覚化

