

# 問題解決技法入門

## 4. Data Analysis

### 4. scatter plot matrix using R

堀田 敬介

# R commanderで散布図行列描画

## ① データの準備 : csv ファイル

– [bb2019.csv](#) 2019年プロ野球 セ・パ成績 (出典: Yahoo Japan! Sports navi)

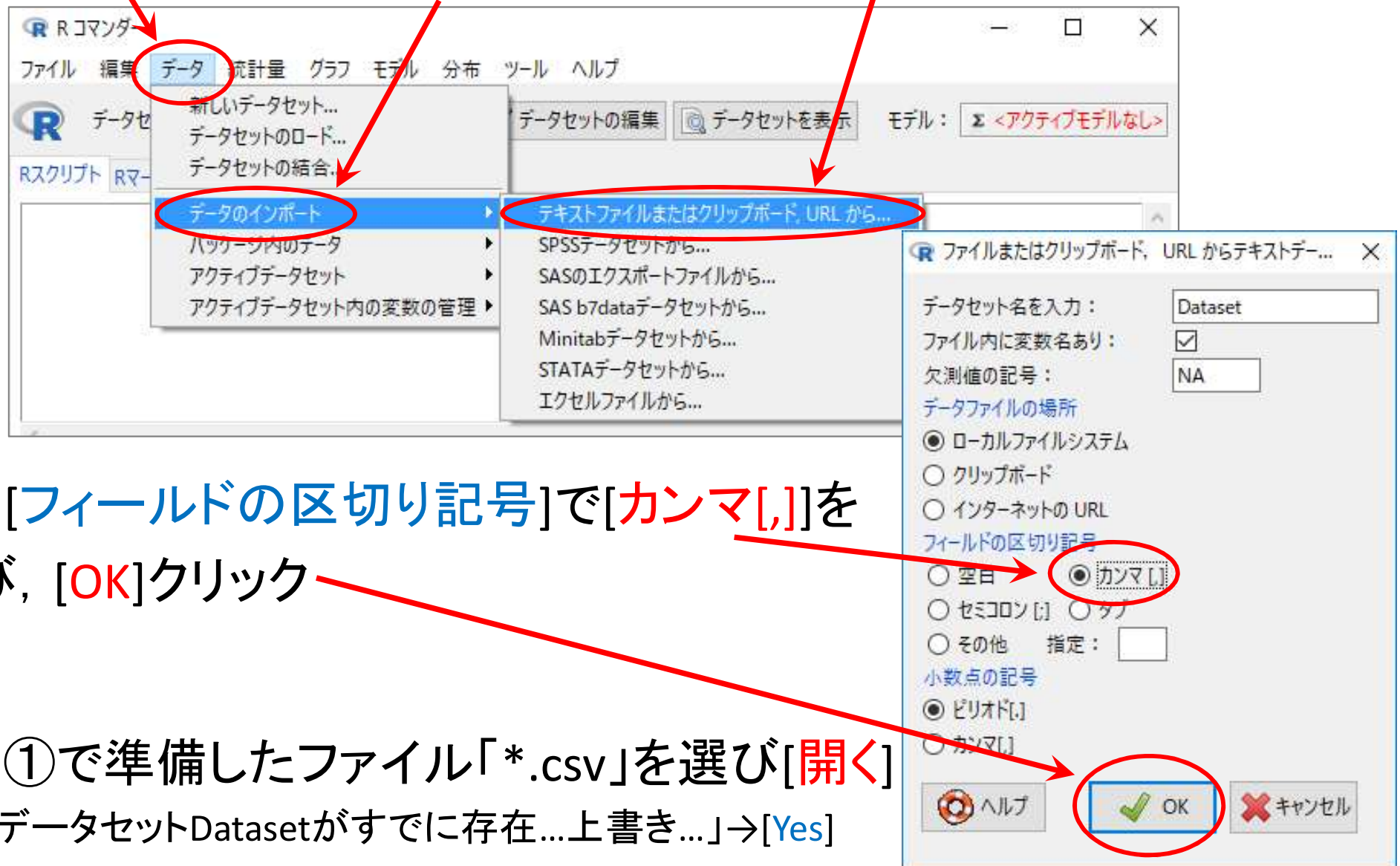
## ② R / R commanderの起動

- 「R [x64](#) [X.X.X](#)」or「R [i386](#) [X.X.X](#)」を選択
- 「[R Console](#)」コマンドプロンプト上 (> [library\(Rcmdr\)](#) [\[Enter\]](#) )

# R commanderで散布図行列描画

## ③ データの読込

- 「データ」ー「データのインポート」ー「テキストファイル...」を選択



- [フィールドの区切り記号]で[カンマ[,]]を選び, [OK]クリック

- ①で準備したファイル「\*.csv」を選び[開く]

注)「データセットDatasetがすでに存在...上書き...」→[Yes]

# R commanderで散布図行列描画

## ③ データの読込（読み込んだファイルの確認）

- [データセットを表示]ボタンをクリックし、内容を確認
  - 注1) 氏名の項目名が「X」であることを確認（もとのファイルに項目名がないデータは自動的に「X」となる）
  - 注2) ケース名（左端）が通し番号（1,2,...,12）となっていることを確認



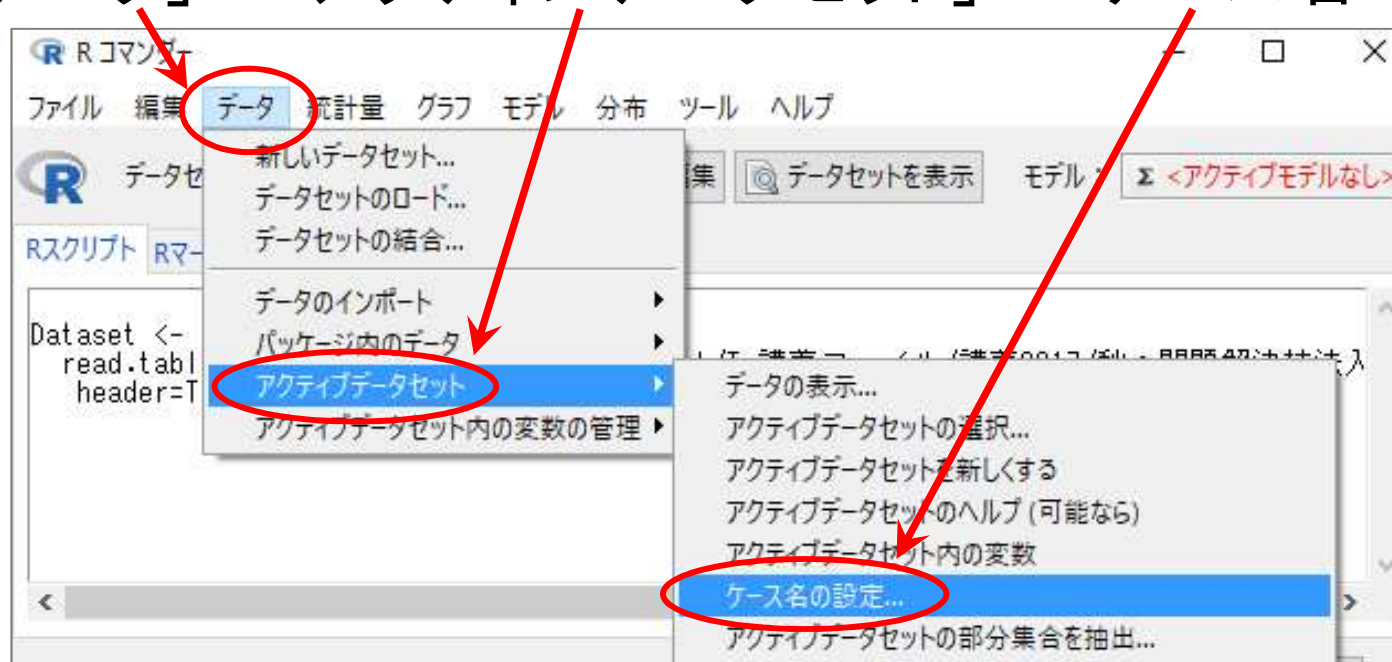
|    | X      | リーグ | 試合数 | 勝数 | 負数 | 引分数 | 勝率    | 得点  | 失点  | 本塁打 | 盗塁  | 打率    | 防 |
|----|--------|-----|-----|----|----|-----|-------|-----|-----|-----|-----|-------|---|
| 1  | 広島     | セ   | 143 | 89 | 52 | 2   | 0.631 | 684 | 497 | 153 | 118 | 0.272 |   |
| 2  | 巨人     | セ   | 143 | 71 | 69 | 3   | 0.507 | 519 | 543 | 128 | 62  | 0.251 |   |
| 3  | DeNA   | セ   | 143 | 69 | 71 | 3   | 0.493 | 572 | 588 | 140 | 67  | 0.249 |   |
| 4  | 阪神     | セ   | 143 | 64 | 76 | 3   | 0.457 | 506 | 546 | 90  | 59  | 0.245 |   |
| 5  | ヤクルト   | セ   | 143 | 64 | 78 | 1   | 0.451 | 594 | 694 | 113 | 82  | 0.256 |   |
| 6  | 中日     | セ   | 143 | 58 | 82 | 3   | 0.414 | 500 | 573 | 89  | 60  | 0.245 |   |
| 7  | 日本ハム   | パ   | 143 | 87 | 53 | 3   | 0.621 | 619 | 467 | 121 | 132 | 0.266 |   |
| 8  | ソフトバンク | パ   | 143 | 83 | 54 | 6   | 0.606 | 637 | 479 | 114 | 107 | 0.261 |   |
| 9  | ロッテ    | パ   | 143 | 72 | 68 | 3   | 0.514 | 583 | 582 | 80  | 77  | 0.256 |   |
| 10 | 西武     | パ   | 143 | 64 | 76 | 3   | 0.457 | 619 | 618 | 128 | 97  | 0.264 |   |
| 11 | 楽天     | パ   | 143 | 62 | 78 | 3   | 0.443 | 544 | 654 | 101 | 56  | 0.257 |   |
| 12 | オリックス  | パ   | 143 | 57 | 83 | 3   | 0.407 | 499 | 635 | 84  | 104 | 0.253 |   |

- 注3) 確認後は、必ず「Dataset」の「×」をクリックして「閉じる」こと

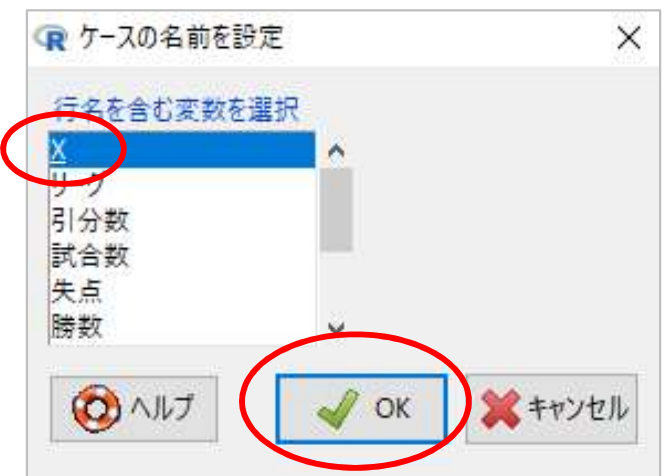
# R commanderで散布図行列描画

## ④ データにケース名を設定する

- 「データ」－「アクティブデータセット」－「ケース名の設定」選択



- [行名を含む変数を選択]で[X]を選び[OK]

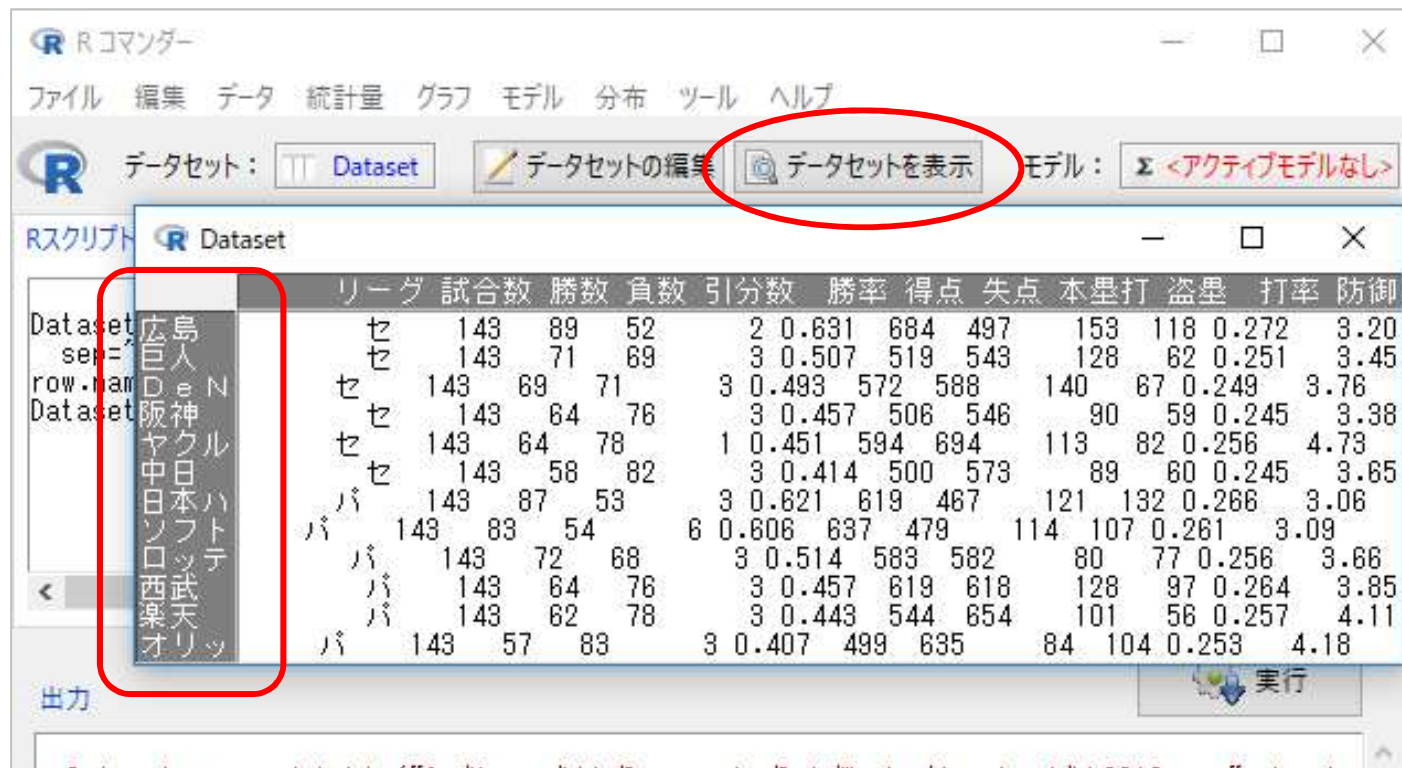


# R commanderで散布図行列描画

## ④ データにケース名を設定する(設定確認)

➤ [データセットを表示]ボタンをクリックし内容を確認

➤ 注1) 指定した変数がケース名になっていることを確認



➤ 注2) 確認後は、必ず「Dataset」の「×」をクリックして「閉じる」こと

# R commanderで散布図行列描画

## ⑤ 散布図行列を描く

➤ 「グラフ」ー「散布図行列」を選択

➤ 『散布図行列』で以下を設定

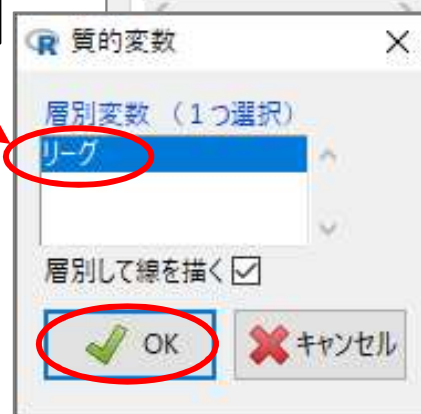
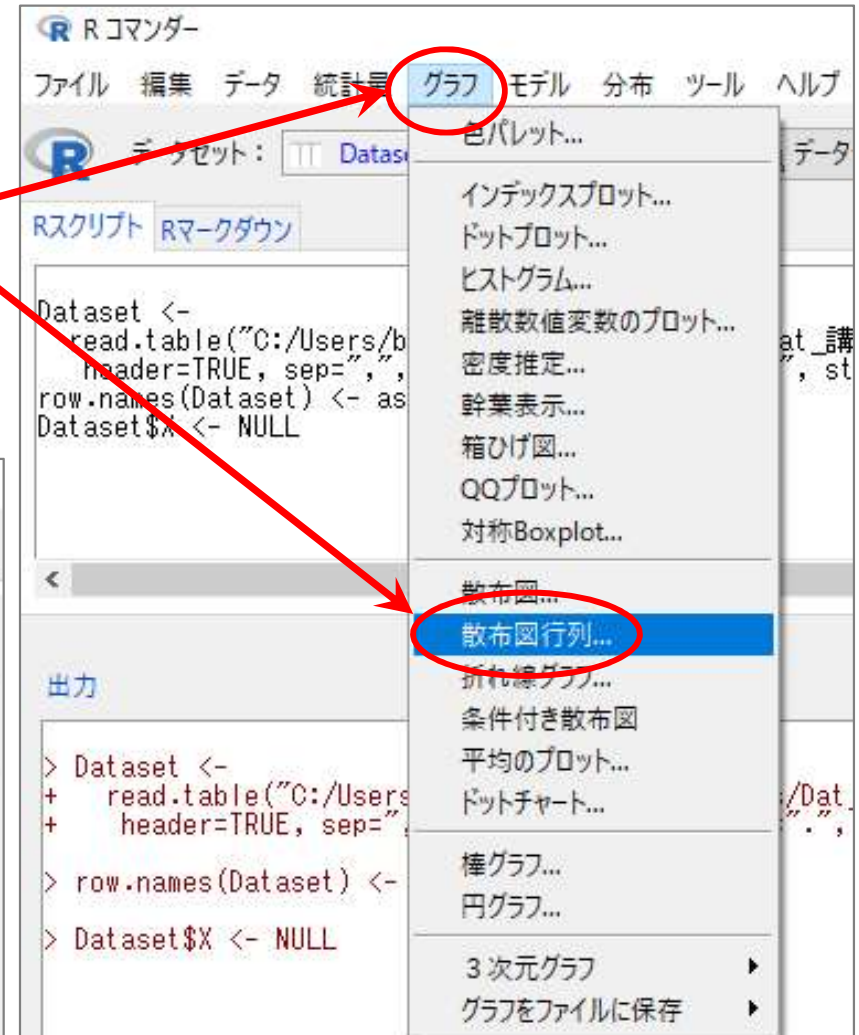
➤ [変数(3つ以上)]で[失点][勝利数][打率][本塁打] 4つを選択

(※複数選択は[Ctrl]キーを押しながら選択)

➤ [層別のプロット]クリック

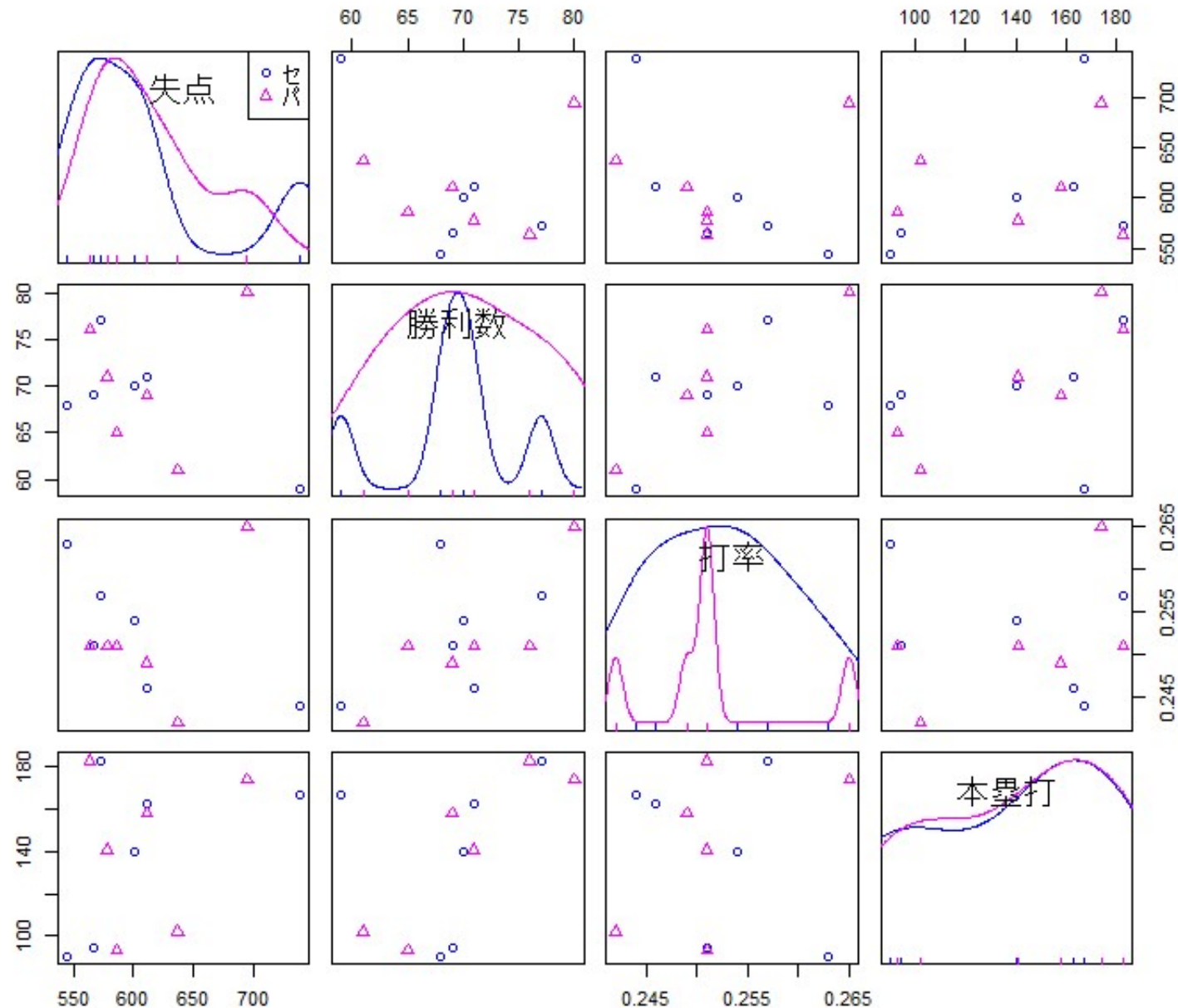
➤ [層別変数(1つ選択)]で[リーグ] 選択し [OK]

➤ [散布図行列]に戻ったら [OK]



# R commanderで散布図行列描画

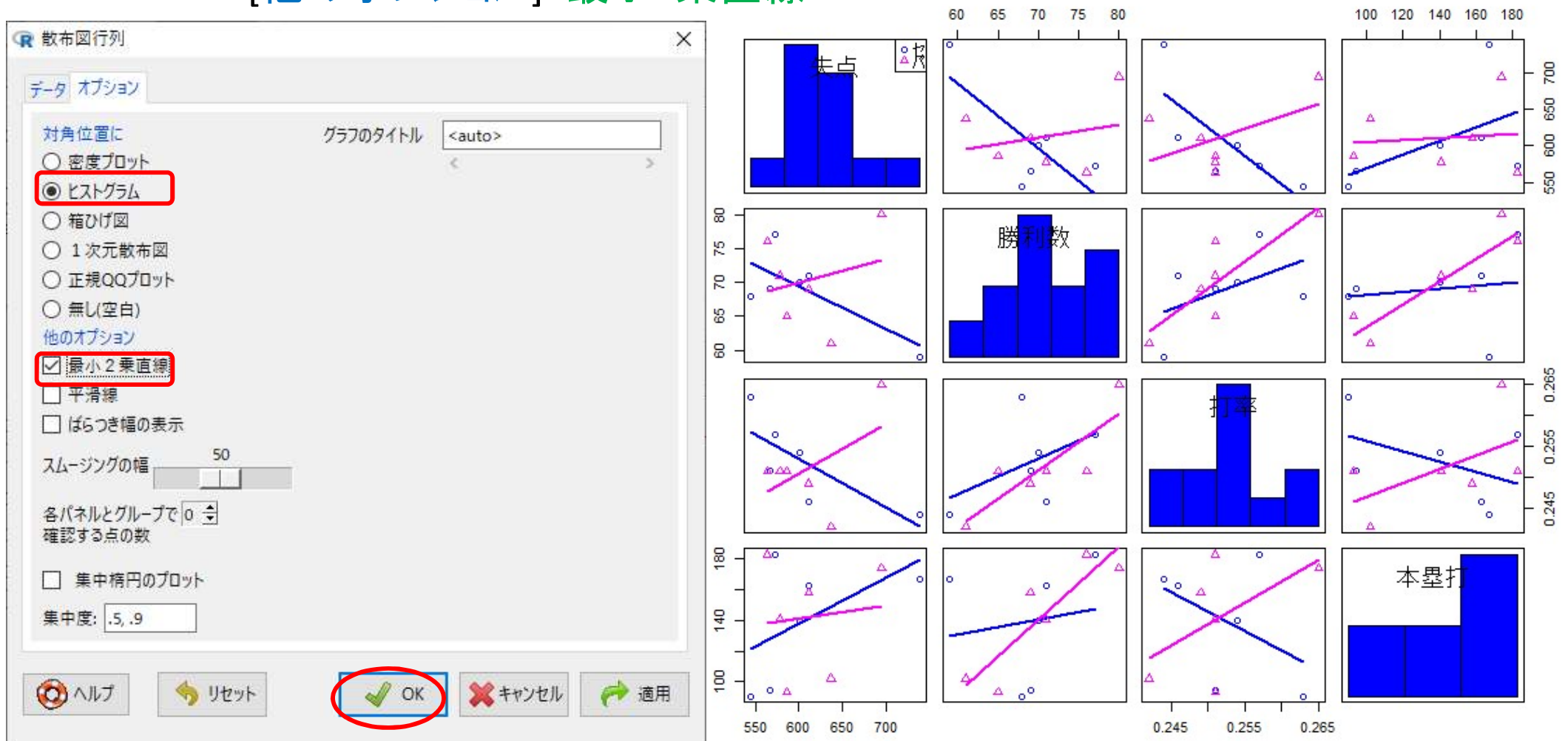
## ⑤ 散布図行列を描く1 【完成1】



# R commanderで散布図行列描画

## ⑥ 散布図行列を描く2 【完成2】

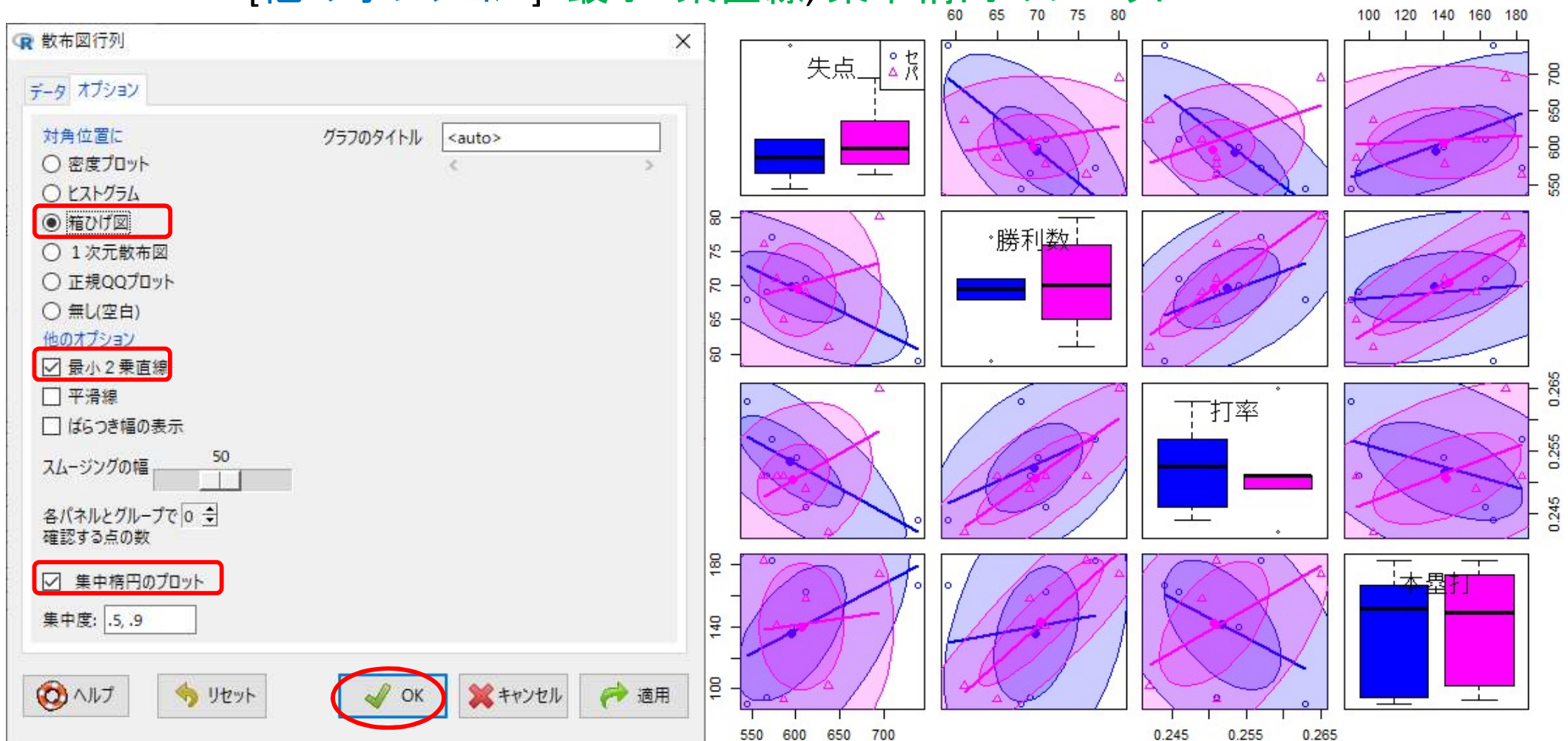
- 再度「グラフ」－「散布図行列」選択し, [オプション]タブで設定
  - [対角位置に]=ヒストグラム
  - [他のオプション]=最小2乗直線



# R commanderで散布図行列描画

## ⑦ 散布図行列を描く3 【完成3】

- 再度「グラフ」－「散布図行列」選択し, [オプション]タブで設定
  - [対角位置に]=箱ひげ図
  - [他のオプション]=最小2乗直線, 集中楕円のプロット



# 参考文献

- ◆ 山本他『Rで学ぶデータサイエンス12統計データの視覚化』共立出版(2013)
- ◆ 奥村晴彦『Rで楽しむ統計』共立出版(2016)
- ◆ J. P. Lander『みんなのR』マイナビ(2015)
- ◆ W. Chang『Rグラフィックス クックブック』オライリー(2013)
- ◆ 青木繁伸『Rによる統計解析』オーム社(2009)
- ◆ 荒木孝治『RとRコマンダーではじめる多変量解析』日科技連(2007)
- ◆ 金明哲『Rによるデータサイエンス』森北出版(2007)
- ◆ 新納浩幸『Rで学ぶクラスタ解析』オーム社(2007)

## もっと知りたい人へ

- 関連する経営学科の授業
  - 「統計の見方」(1/2セメ)
  - 「統計の分析と利用」(2セメ)
  - 「データ処理Ⅱ」(2/3セメ)
  - 「統計データの扱い方」(3/4セメ)
  - 「多変量の統計データ解析」(4セメ)