

進捗報告

2016/10/07

築山文香

進捗報告

- ・ **今までやったこと**

- Rockstarのフォロワー分析と分類

問題点しかないような気がするので
相談させてください

- ・ **今週やったこと**

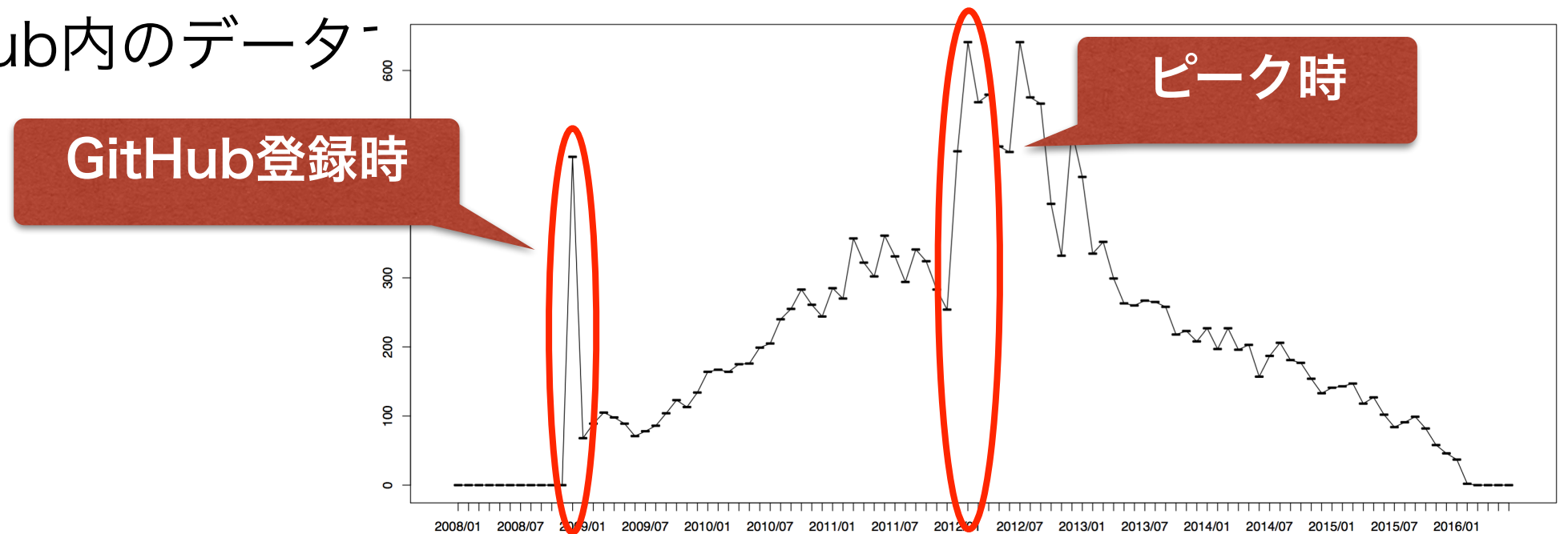
- 散布図行列
- 並行座標プロット
 - ・ **RockstarとnonRockstarを対象にした**

- ・ **来週までにやること**

- RockstarがRockstarになるまでにフォローされたRockstarを見たい
- 他にまだやることがある気がするけど...

Rockstarのフォロワー分析

- Rockstarを月毎のフォロワー獲得数から, 3つの型に分類した
 - GitHubに登録時と, フォロワー獲得数ピーク時に注目
 - どちらが多いかで分類
 - GitHub登録時が多い: GitHub利用前から注目度が高かった, または登録時からそれほど注目度が上がっていないと仮定
 - ピーク時が別にある: GitHub利用後に注目度が上がったと考えられる (GitHub内のデータ-る)



今回の研究は**フォロワー機能**を中心にRockstarを分析したい

注目 (フォロワー) されたきっかけを知りたい

(フォロワー機能がある) GitHub内で分析できないと意味がない

Rockstarの分類結果

- ・ 著名人型：384人
 - GitHub登録時がフォロワー獲得数ピークの開発者
- ・ 早成型：72人
 - フォロワー獲得数がGitHub登録時と別にあり、フォロワー獲得数のピーク時が登録時の**2倍未満**
- ・ 晩成型：66人
 - フォロワー獲得数がGitHub登録時と別にあり、フォロワー獲得数のピーク時が登録時の**2倍以上**

Rockstarの分類結果

- ・ 著名人型：384人
 - GitHub登録時がフォロワー獲得数ピークの開発者

問題点

数が多すぎる

登録時, ピーク時がかなり近い開発者もいる

Rockstarになるまでの期間も考慮したほうがいい？

(中央値が31か月 [2年8か月])

- ・ 早成
 - フ
 - の
- ・ 晩成
 - フ
 - の

のピーク時が登録時の**2倍以上**

Rockstarの分類結果

- ・ 著名な型別 204人
 - GitHub登録時と比べてピーク時にどれくらい注目度が上がったかを分けたかったけど…
- ・ 早成型：72人
 - フォロワー獲得数がGitHub登録時と別にあり、フォロワー獲得数のピーク時が登録時の**2倍未満**
- ・ 晩成型：66人
 - フォロワー獲得数がGitHub登録時と別にあり、フォロワー獲得数のピーク時が登録時の**2倍以上**

分け方が雑？
分ける必要性ない？

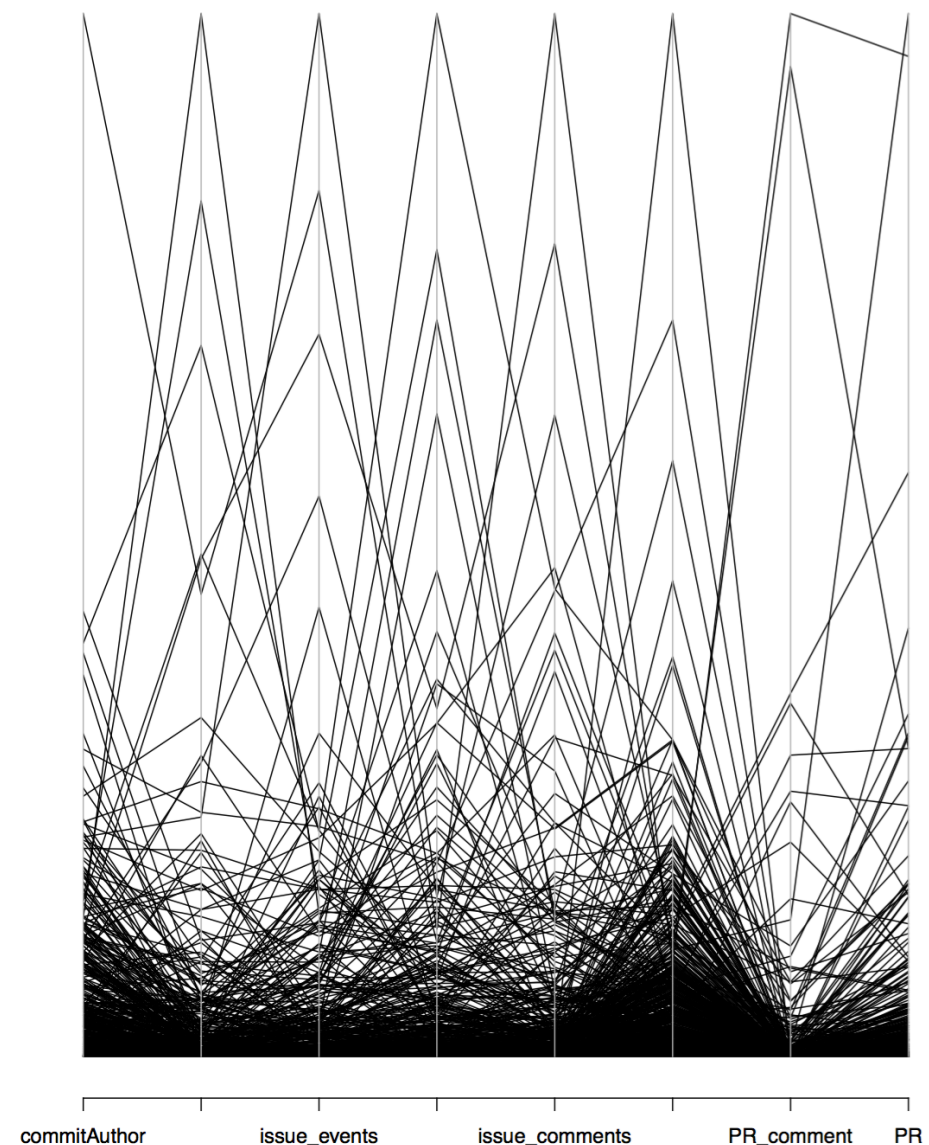
今週やったこと

- ・ 散布図行列分析, 並行座標プロット
 - 活動の変数について, それぞれの関係を全部可視化

散布図行列分析



並行座標プロット



ざっくりこんな感じ

散布図行列

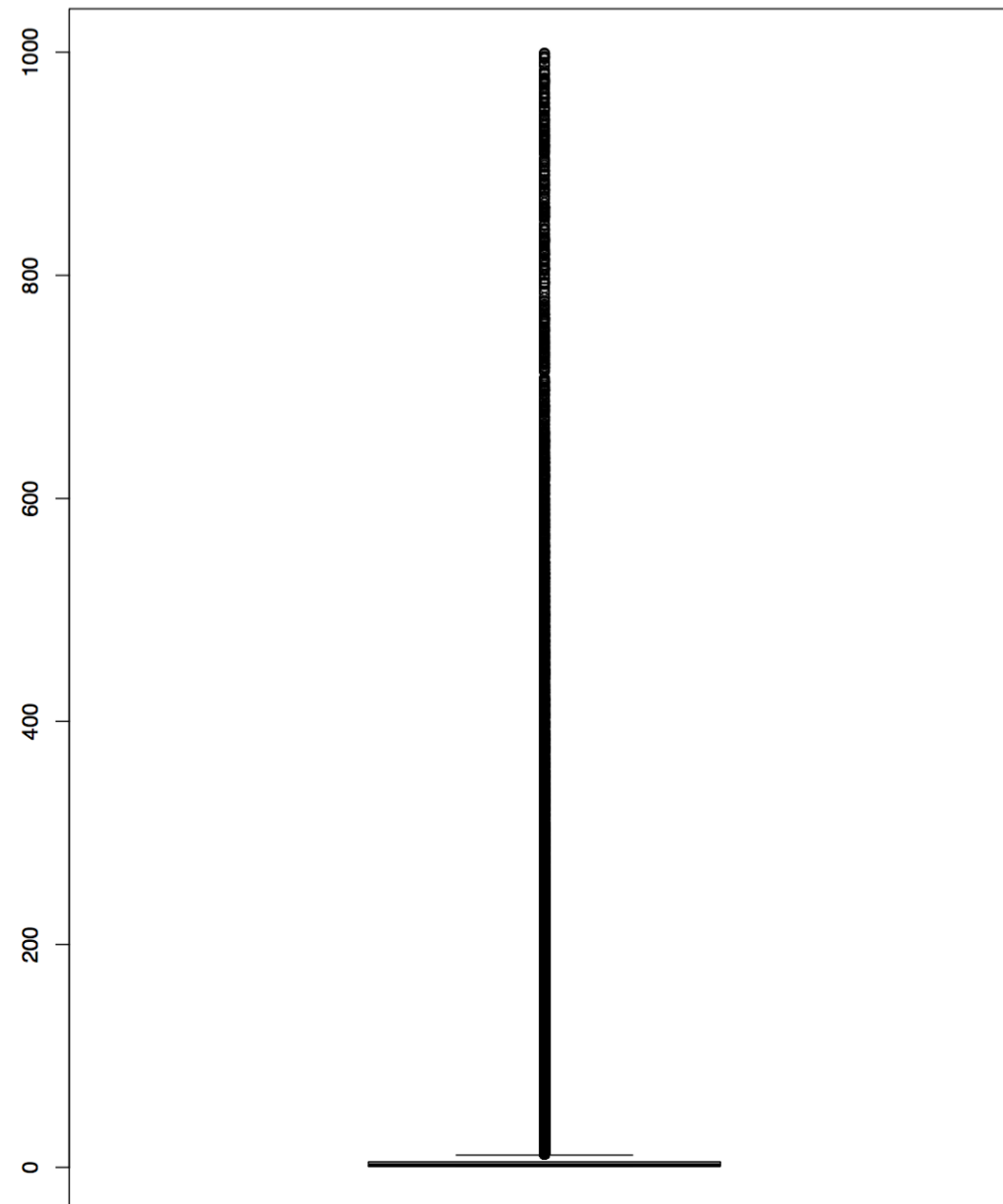
- Rockstarとnon_Rockstarに分けて, 比較
 - Rockstar
 - フォロワーが1000人以上
 - nonRockstar
 - Rockstarではない
 - 1人以上フォロワーがいる開発者
 - $1,790,539 \text{人} - 522 \text{人 (Rockstar)} = 1,790,017 \text{人}$: 約180万人
 - 全ユーザー登録者数約1,300万人の中約13%
 - 登録して31ヶ月（2年8か月）以上経過している
 - 31ヶ月: Rockstarになる開発者の中央値
 - 2016/06/16のデータセットなので, 2016年5月までのデータを取得すると考えて, 2013年9月中までに登録した開発者
 - 1,074,172人 : 約100万人

この人たちをnon_Rockstarの対象にしようかなと

nonRockstar全体

- ・ 1,074,172人
 - 1人以上フォロワー数がある
 - 登録して31ヶ月以上が経過している

- ・ フォロワー数



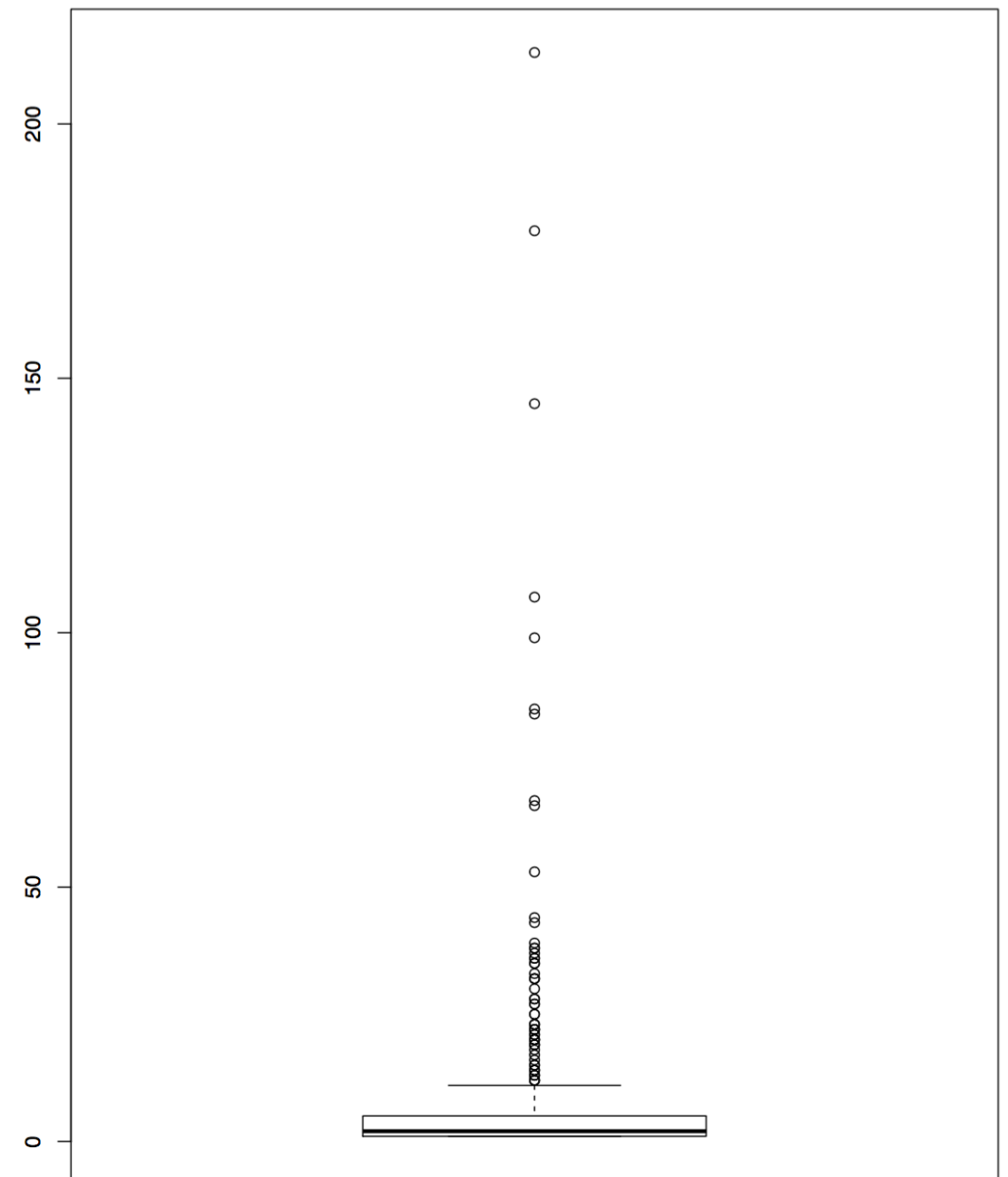
中央値	平均	範囲
2	7.74	1～999

見方が…

今回のnonRockstar対象

- nonRockstarからランダムに522人抽出

中央値	平均	範囲
2	6.83	1～214



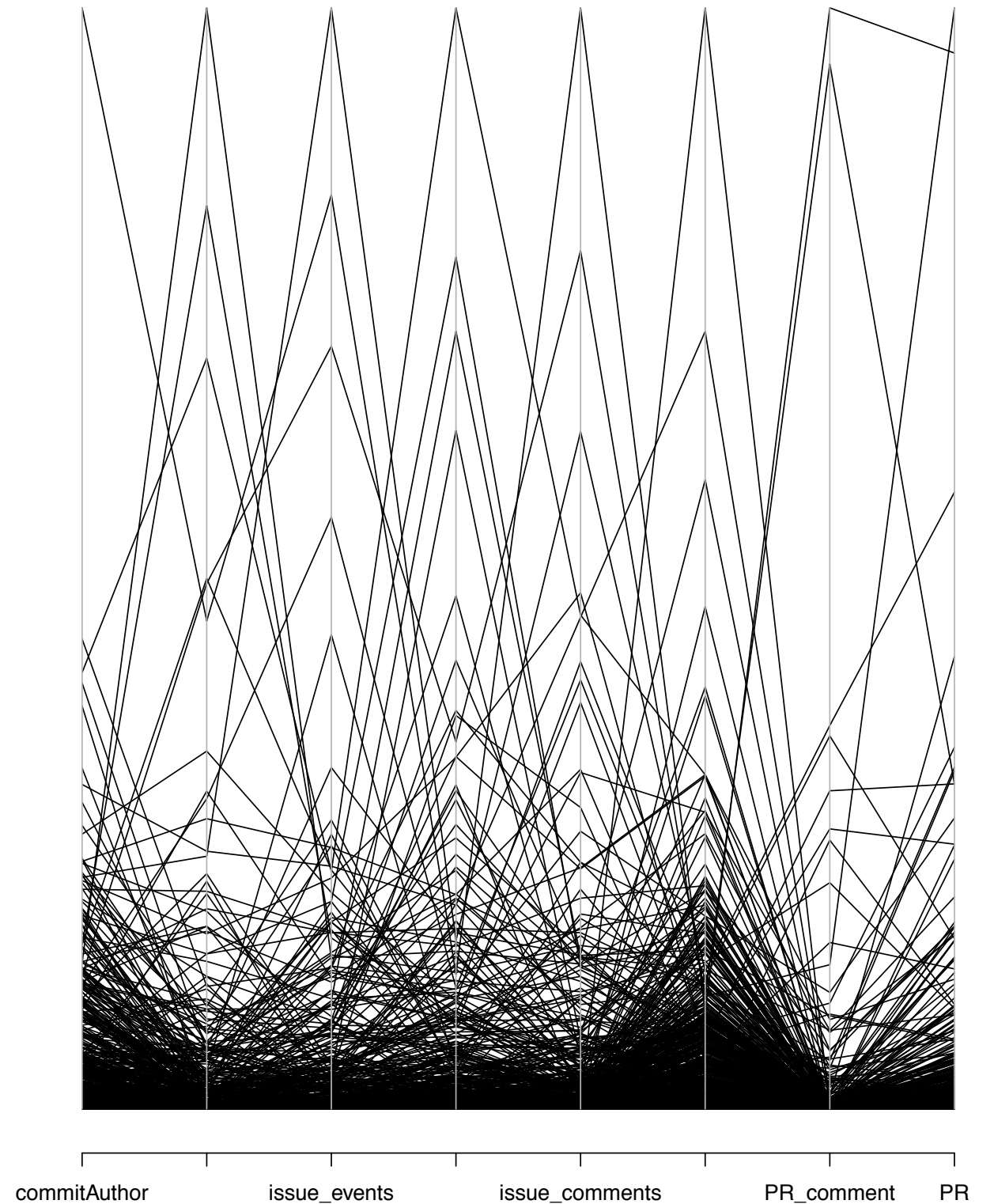
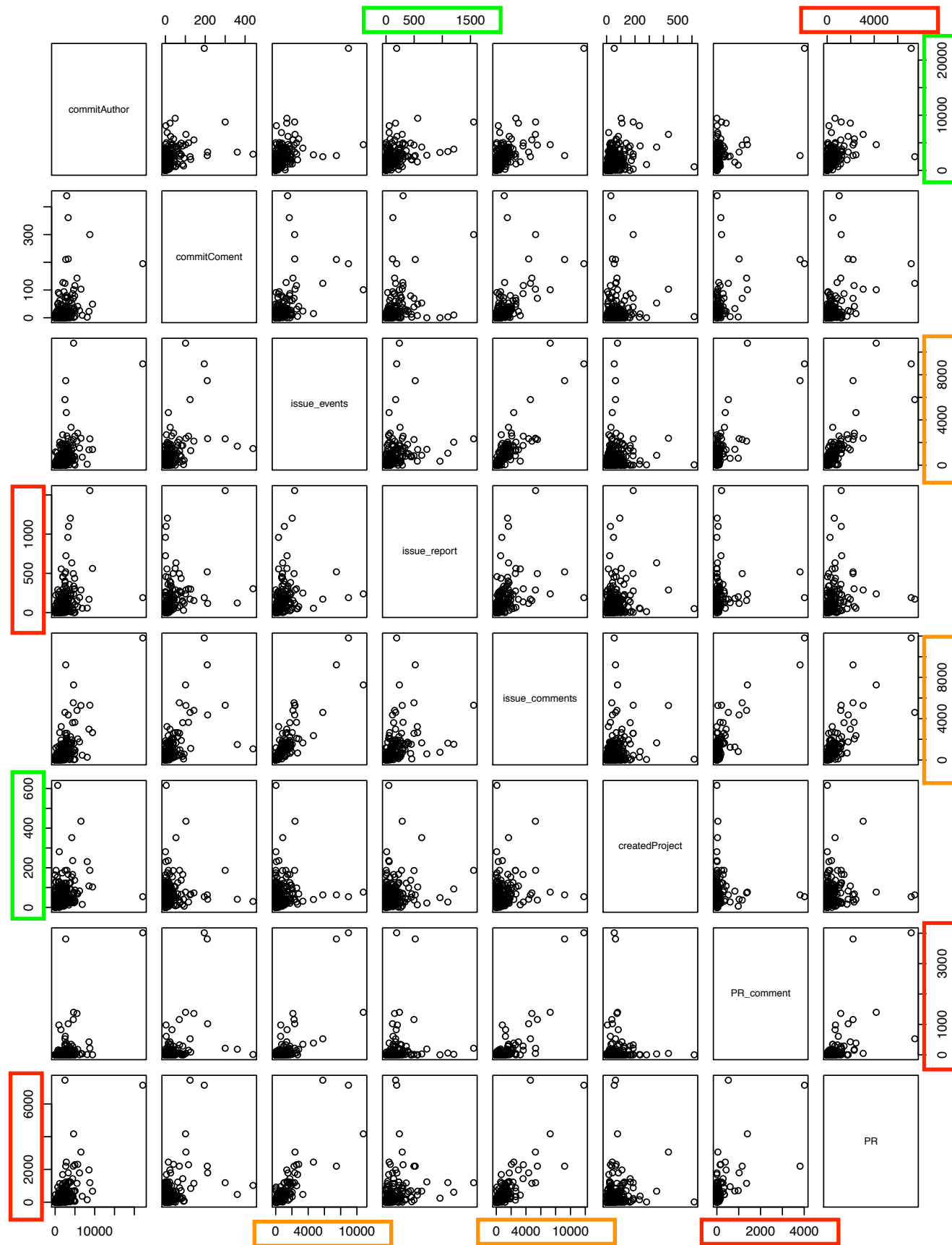
散布図行列

- 8個の活動の変数のうち, 2つの活動の変数を総当たりで可視化
 - Rコマンド : `pairs(metrics[,3:10])` でグラフが出てくる

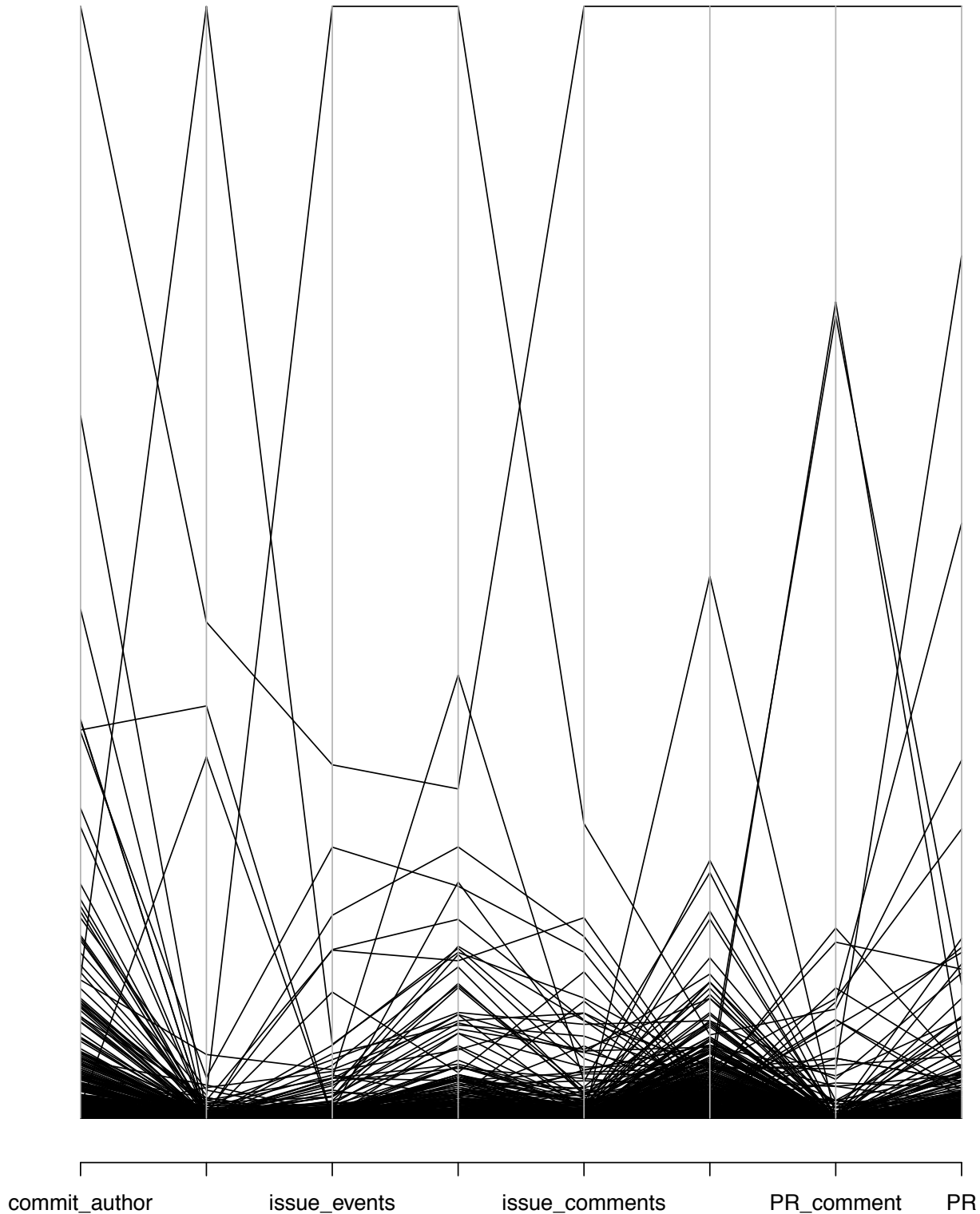
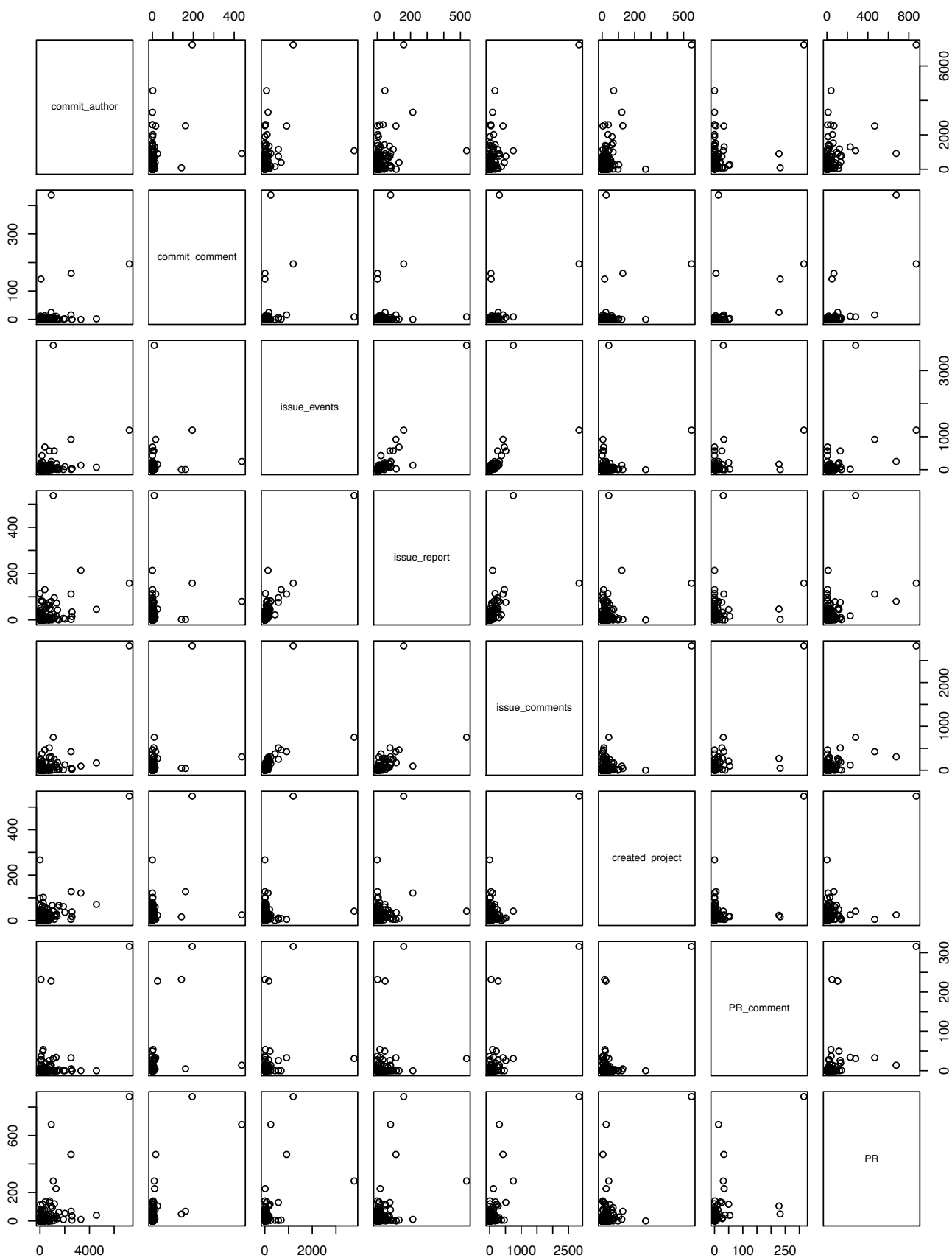
- 各変数

DB (table.column)	内容
commits.author_id	コミットされた投稿の著者
commit_comments.user_id	コミットコメント
issue_events.actor_id	パッチ投稿
issues.reporter_id	不具合報告
issue_comments	不具合コメント
projects.owner_id	プロジェクト作成
pull_request_history.actor_id	プルリクエスト
pull_request_comments.user_id	プルリクエストコメント

Rockstar : 散布図並行行列, 並行座標プロット



nonRockstar：散布図並行行列，並行座標プロット



感想

- ・ RockstarとnonRockstarでは, 活動量が全然違うものも多い
- ・ nonRockstarの中央値が2って…
 - 比較対象を考え直したほうがいい？
 - 活発だけどRockstarになれなかった人とか…？

来週すること

・ ...