

日本疫学会学術総会プレセミナー2023
2023年2月1日 16:00-18:00(日本時間)



Rで実践！ 美しい Figure & Table を作成しよう

企画：藤井亮輔・佐藤俊太郎

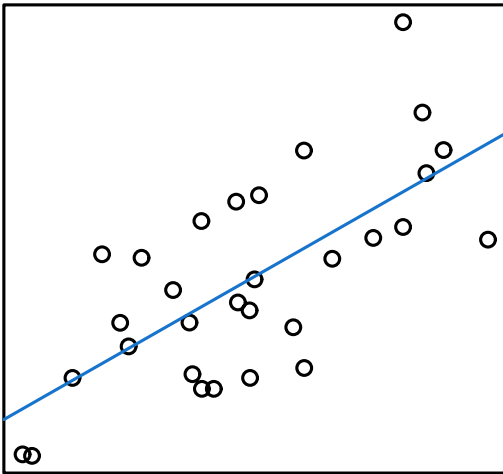
講師：浦和幸・川添百合香・坪井良樹・中島誉也・三橋利晴

資料提供：森岡裕

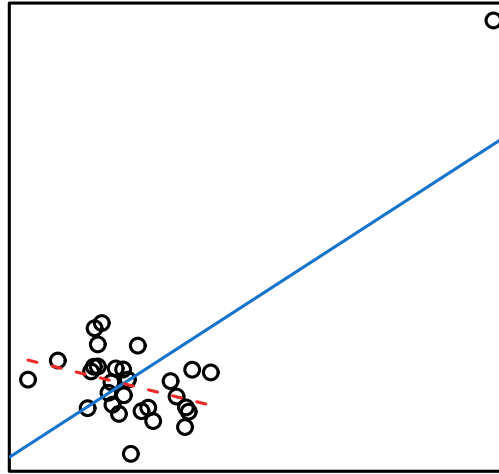
※本発表に関して、開示すべきCOI関係はありません。

「相関係数:0.7」と言われて、 どんな散布図をイメージしますか？

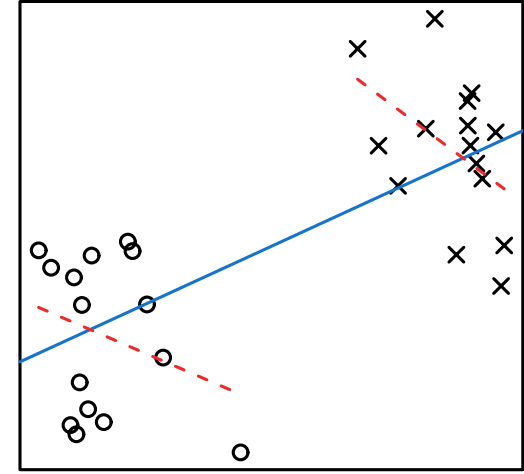
(1) Normal x, normal residuals



(11) A single positive outlier



(14) Two groups



データを実際に見なければ、気づかないことがたくさんある。
特に、外れ値・ノイズ・欠測値などである。

なぜ、(データを)可視化する？

可視化は、データの取得・分析・発見・コミュニケーションと関係している。



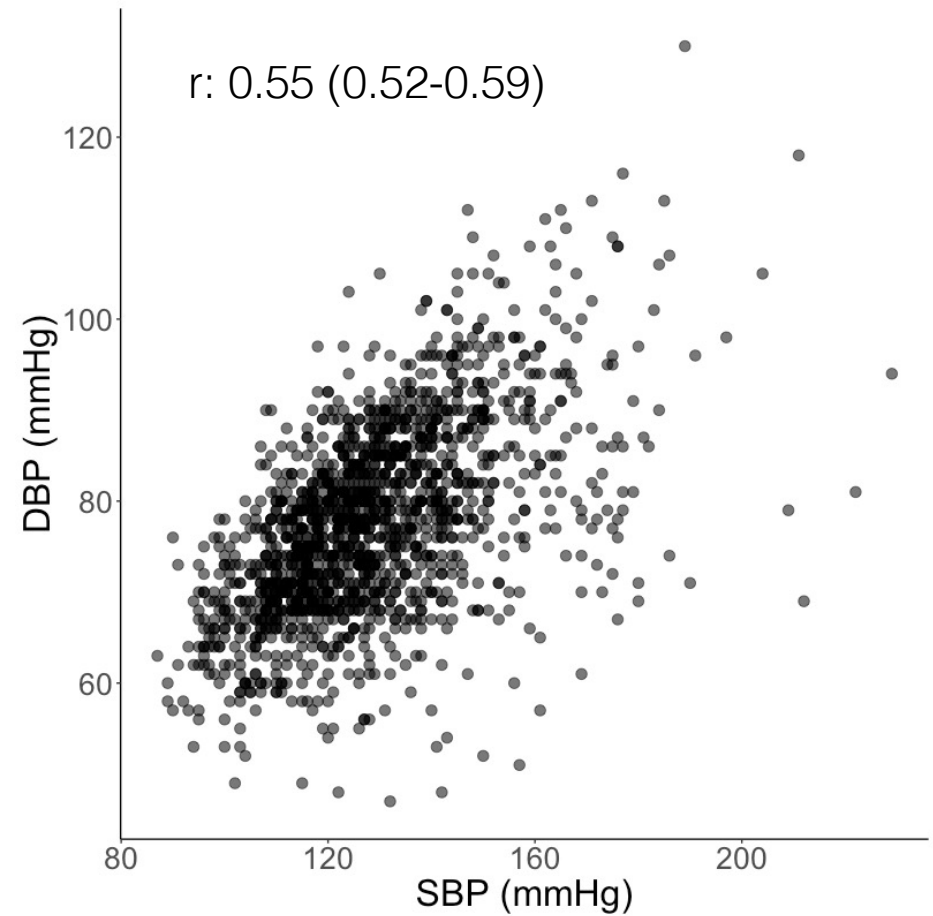
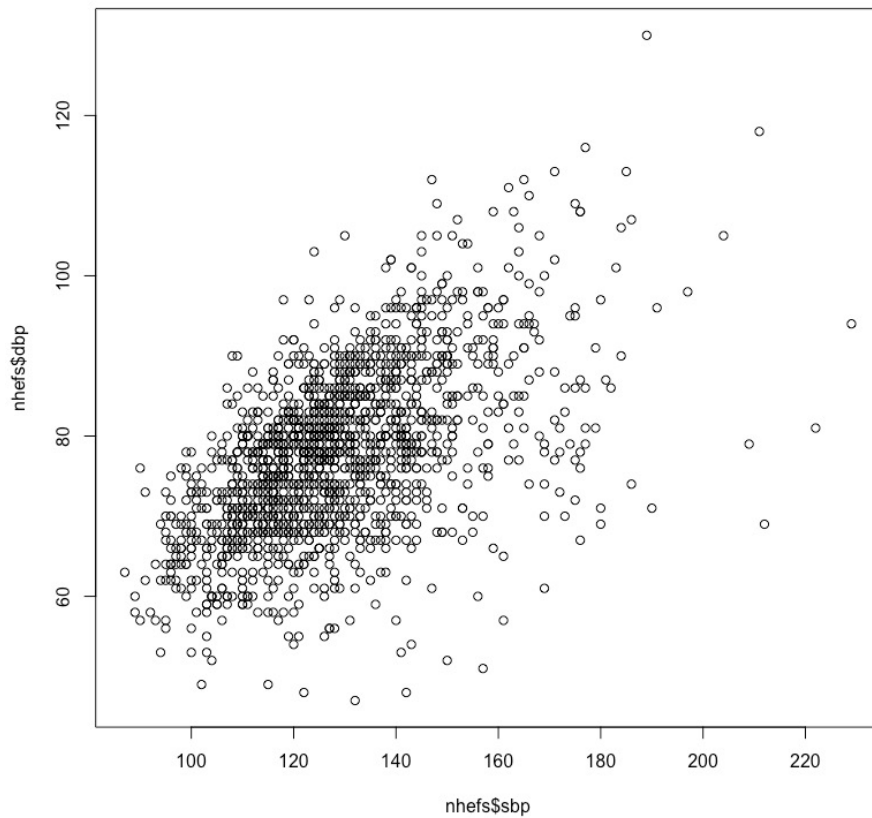
AR O'Donoghue SI, et al. 2018.
Annu. Rev. Biomed. Data Sci. 1:275-304

コミュニケーションだけでなく、
色々なプロセスで可視化は重要な役割を果たしている。

どちらが分かりやすいでしょうか？

4

Scatter plot of SBP and DBP among NHEFS participants (n = 1,548)



SBP (systolic blood pressure); DBP (diastolic blood pressure)

なぜ、(データを)可視化する？

5

可視化は、データの取得・分析・発見・コミュニケーションと関係しているが、**誰に、何を目的とするか、**によって大きな違いがある。

例えば、散布図を描きたい…

分析やラボミーティング

使用する技法は洗練され、訴求性に優れたものである必要はない。

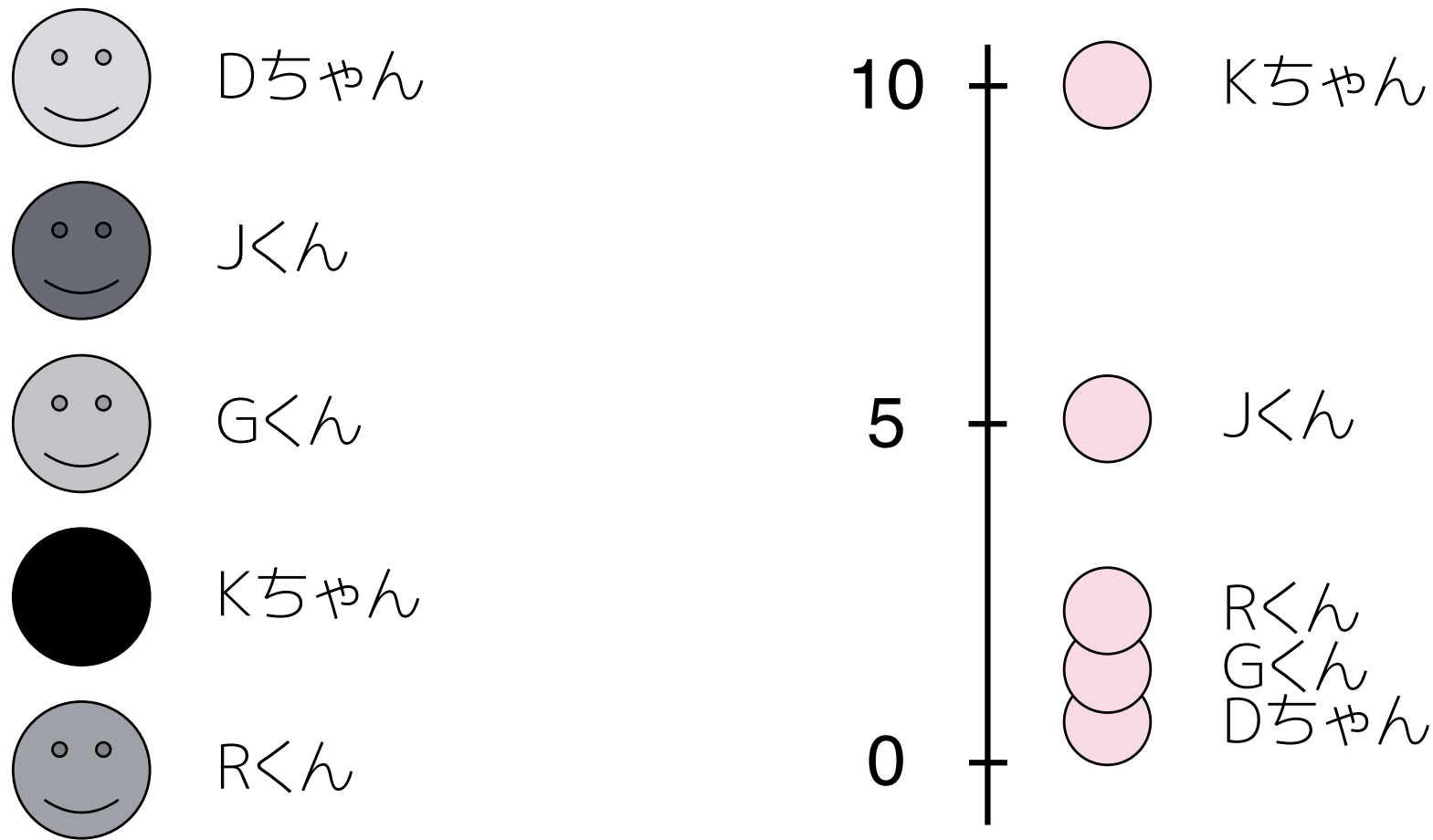
VS.

論文や学会発表

読者や聴衆の興味を惹きつつ、広くの人に理解できる内容でなければならない。

可視化の場面や目的を意識する。

色が濃い/値が大きい順番に、各人の名前を教えてください。



どんな可視化が望ましい？

受け手が、**正確に**情報を**脱符号化**できる表現が良い。



- **符号化**で、目的とデータに合わせた効率的な表現を選択する
- **脱符号化**で、識別しやすい表現を選択する

*特に、前頁の中での「communication(コミュニケーション)」に当たる可視化を想定しています。