

2023年11月16日(木) 13:40 – 14:40

AJACSオンライン18「オンラインで使える統計解析ツールを知って・学んで・使う」

オンラインツール **PlotsOfData** による箱ひげ図の作成

講師：

別所-上原 学

名古屋大学高等研究院YLC特任助教

兼 理学研究科生命理学領域 器官機能学グループ

コンテンツ

自己紹介

セクション1: データ視覚化の重要性

セクション2: PlotsOfDataの概要

セクション3: PlotsOfDataの基本的な使い方

セクション4: その他のグラフ描画ツールの紹介

自己紹介



氏名：別所-上原 学
所属機関：名古屋大学
役職：特任助教
専門：発光生物学

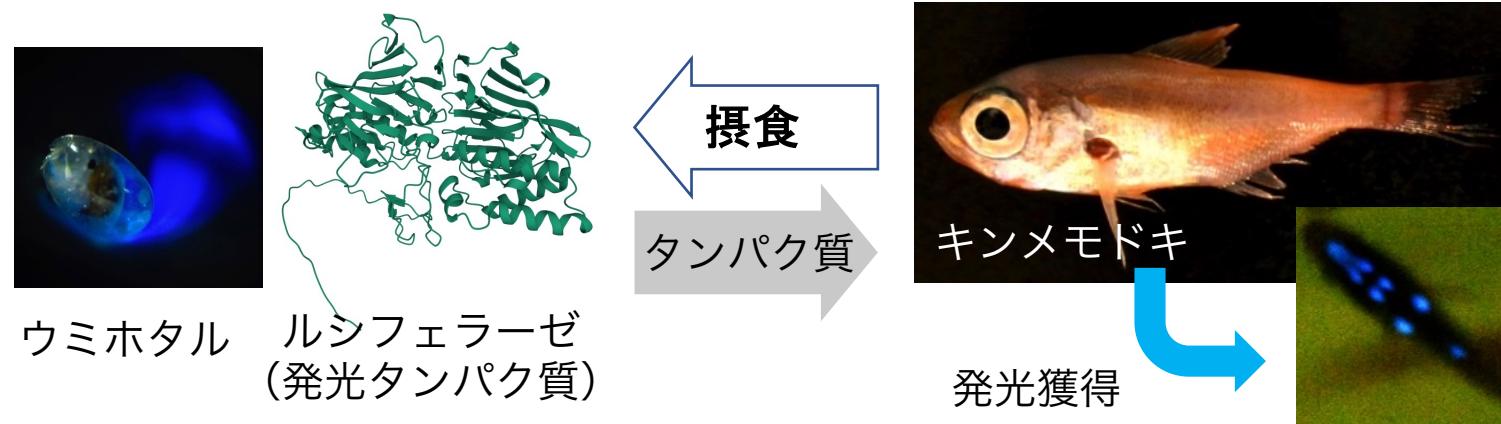
べっしょ-うえはら まなぶ

発光生物の進化

ホタルなどはどうして光るようになったのか？庭先から深海まであらゆるところに生息する光る生き物を調べて、その生命誌を明らかにするため、日々研究しています。

工サ生物の能力を盗む生き物のふしき

キンメモドキは発光に必要なルシフェラーゼ遺伝子をもたずウミホタルからタンパク質を盗むことで、餌生物がもつ機能を獲得します。「盗タンパク質」として知られるこの現象はキンメモドキでのみ見つかっています。消化・分解されるはずのタンパク質が捕食者の体内で機能する現象は生物学にとって革新的な概念です。今後は、盗タンパク質をもつ発光生物を新たに発見し、餌由来のタンパク質を取込む共通原理の解明を目指します。



<略歴>

2012年名古屋大学農学部卒業、2014年名古屋大学生命農学研究科博士前期課程修了、2017年名古屋大学生命農学研究科博士後期課程修了、同年同所属 博士（農学）学位取得、2017年名古屋大学研究員、同年育児により退職、2017年中部大学研究員、2018年モントレー湾水族館研究所ポスドクフェロー、2020年より名古屋大学YLC特任助教

ウミホタルから光を盗むキンメモドキ
新しい進化の戦略 盗タンパク質

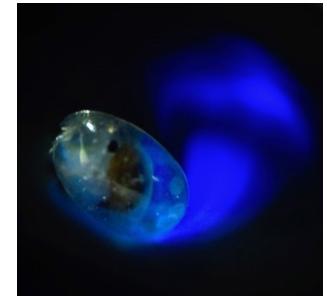
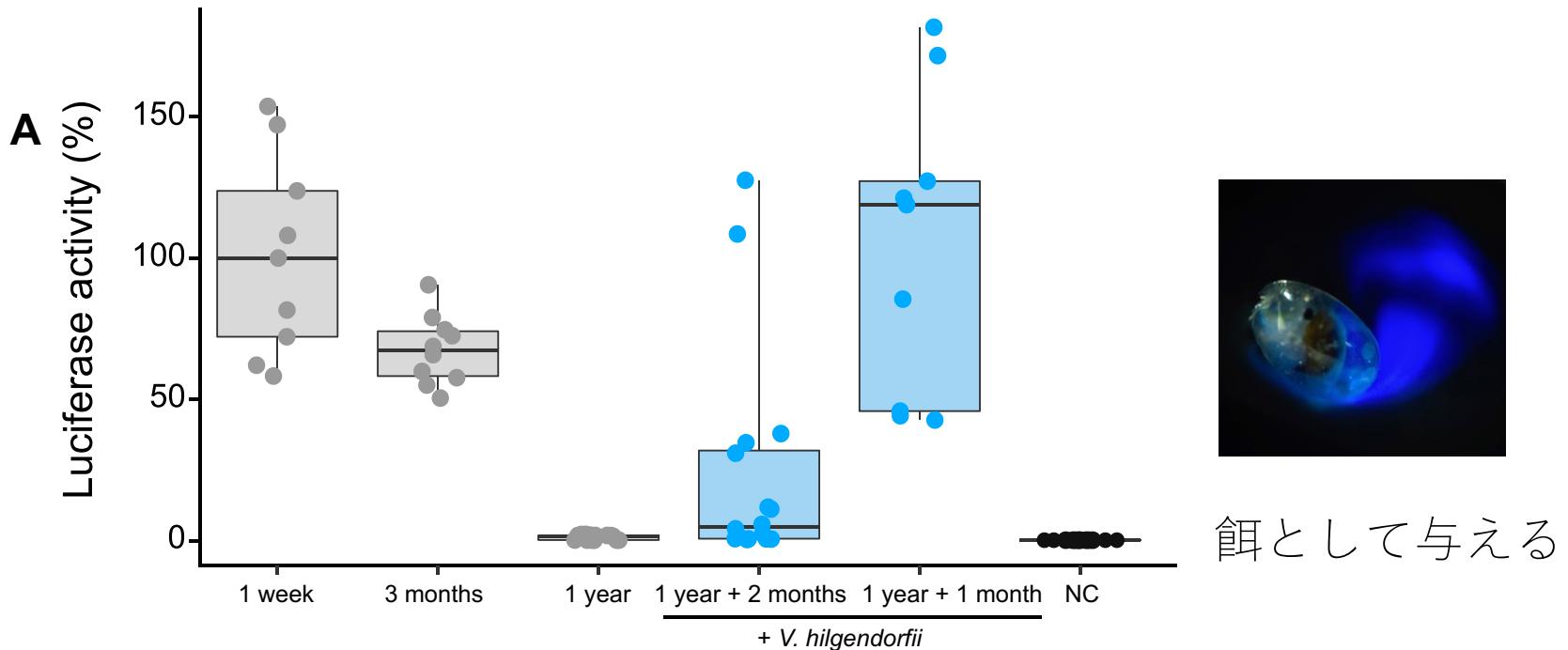


論文で使用した図の例：Rで作成

SCIENCE ADVANCES | RESEARCH ARTICLE



野外個体の飼育を続けると、
発光活性が落ちる



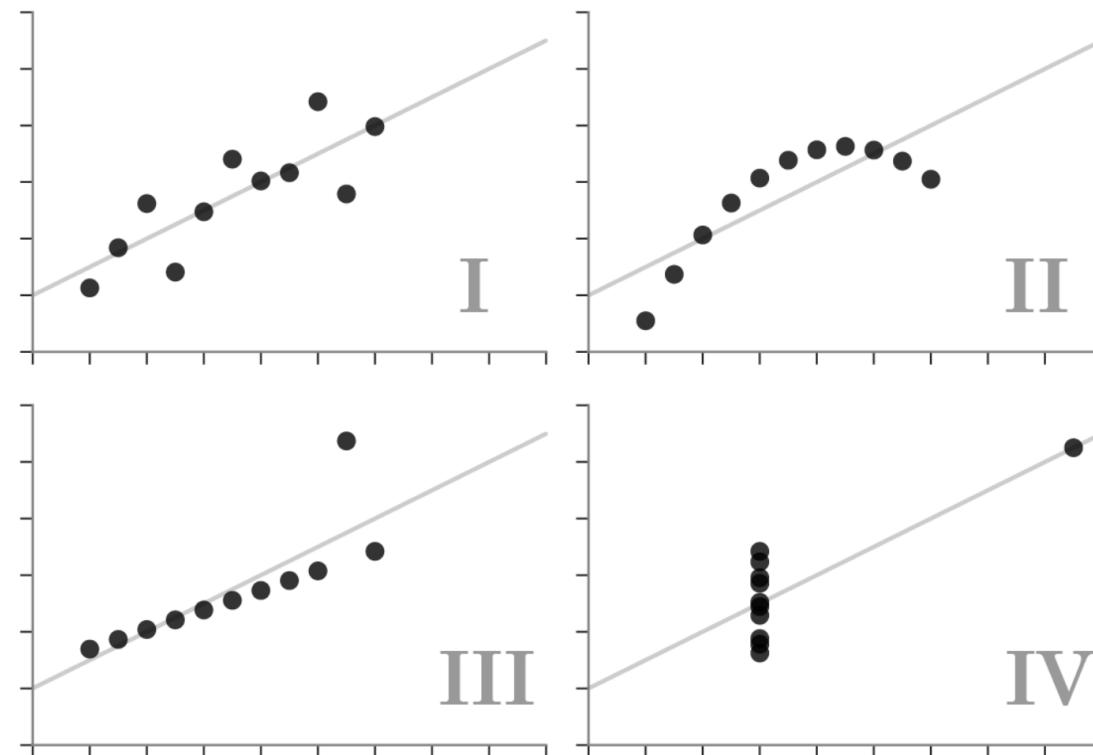
餌として与える

活性が回復した

Same Stats, Different Graphs

✓ Anscombe's Quartet アンスコムの数値例

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.

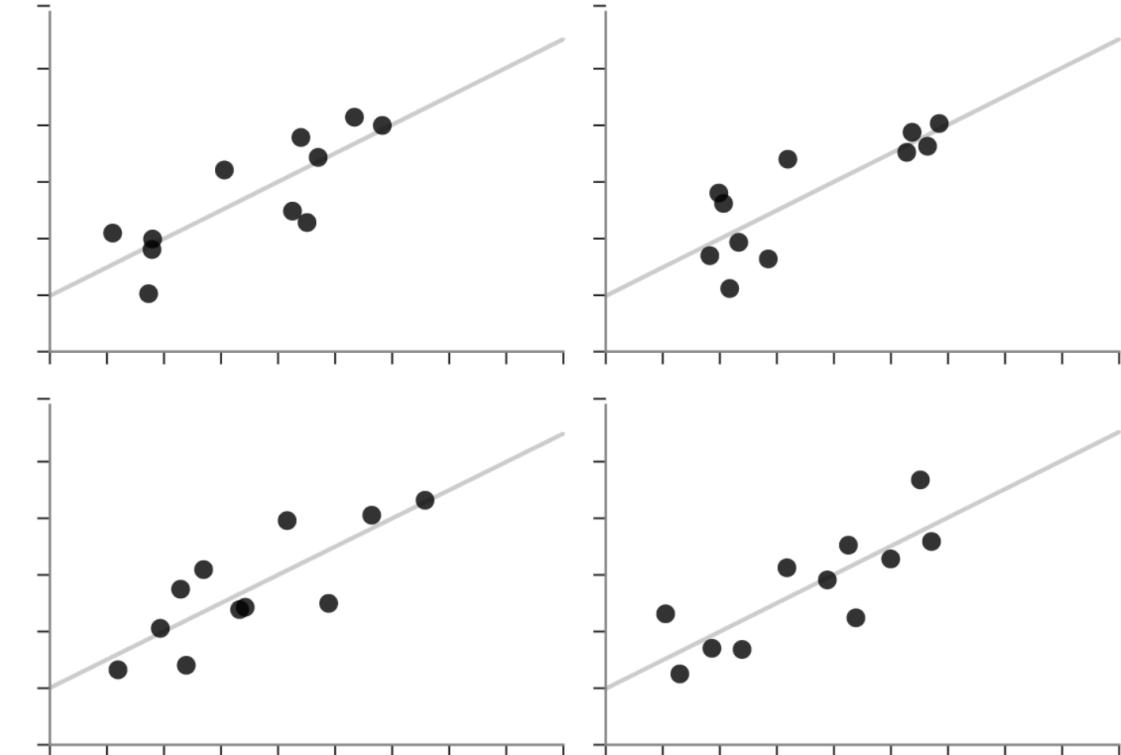


<https://www.research.autodesk.com/publications/same-stats-different-graphs/>

要約統計量だけを表示する危険性

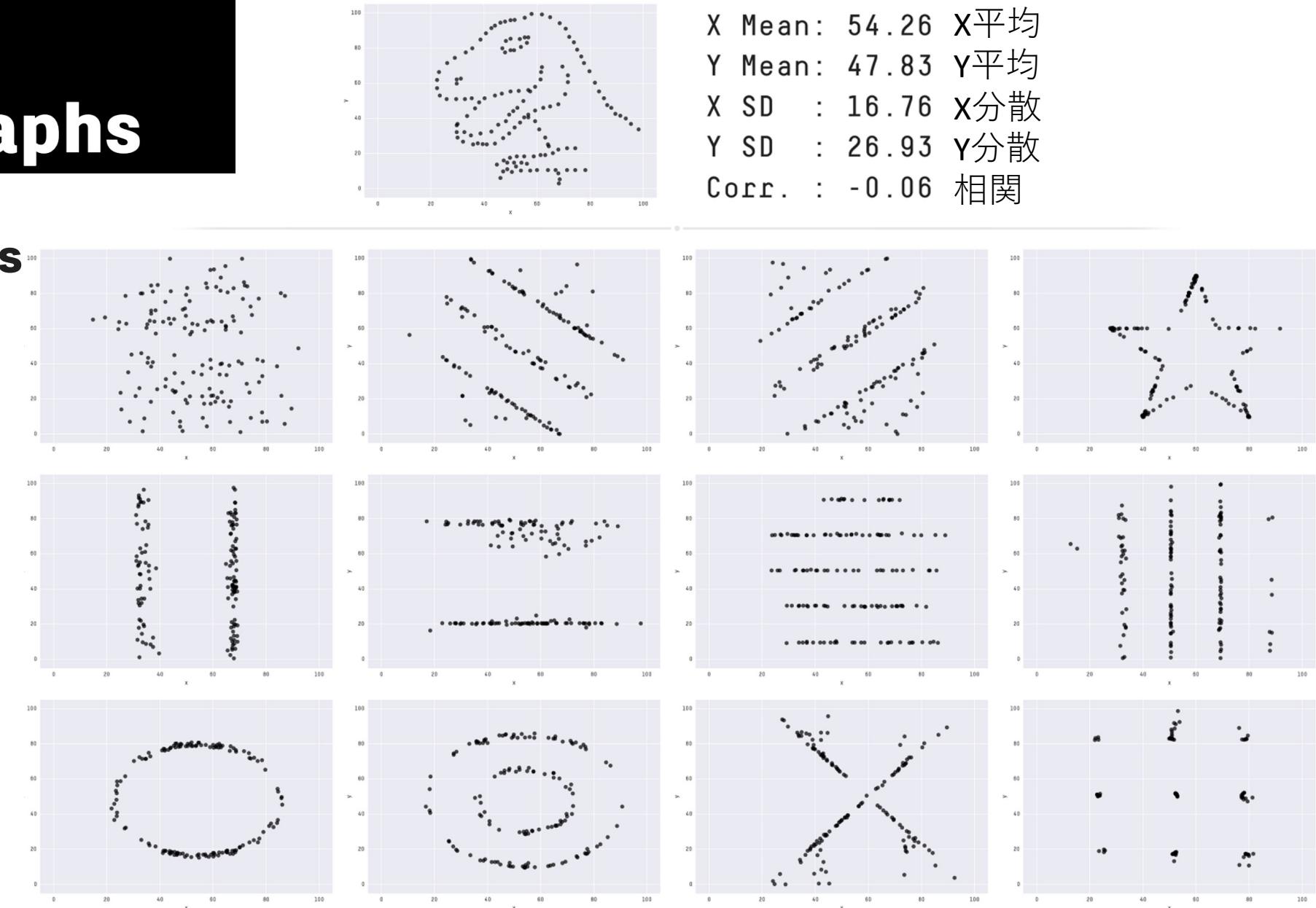
✗ Unstructured Quartet

Each dataset here also has the same summary statistics. However, they are not *clearly different* or *visually distinct*.



Same Stats, Different Graphs

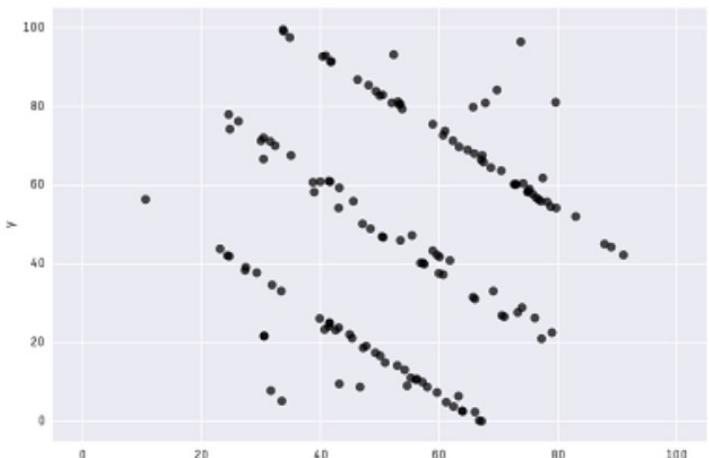
**“never trust
summary statistics
alone;
always
visualize your data.”**
- Albert Cairo



セクション1: データ視覚化の重要性

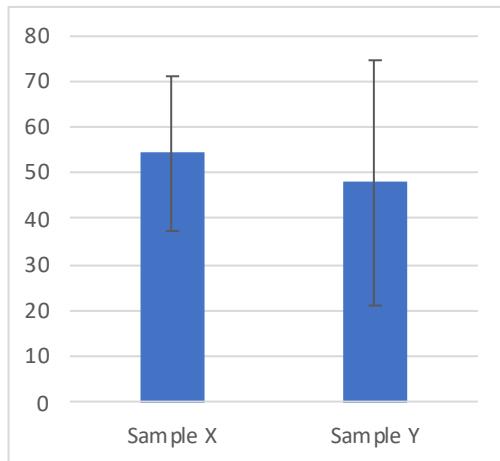
X Mean: 54.26 X平均
Y Mean: 47.83 Y平均
X SD : 16.76 X分散
Y SD : 26.93 Y分散
Corr. : -0.06 相関

要約統計量は実際のデータを理解するのあまり助けない



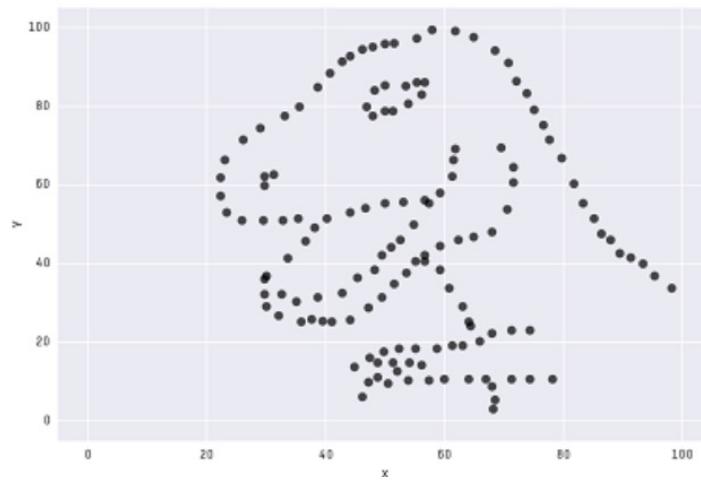
実際のデータ

グラフ作成
→



要約データ

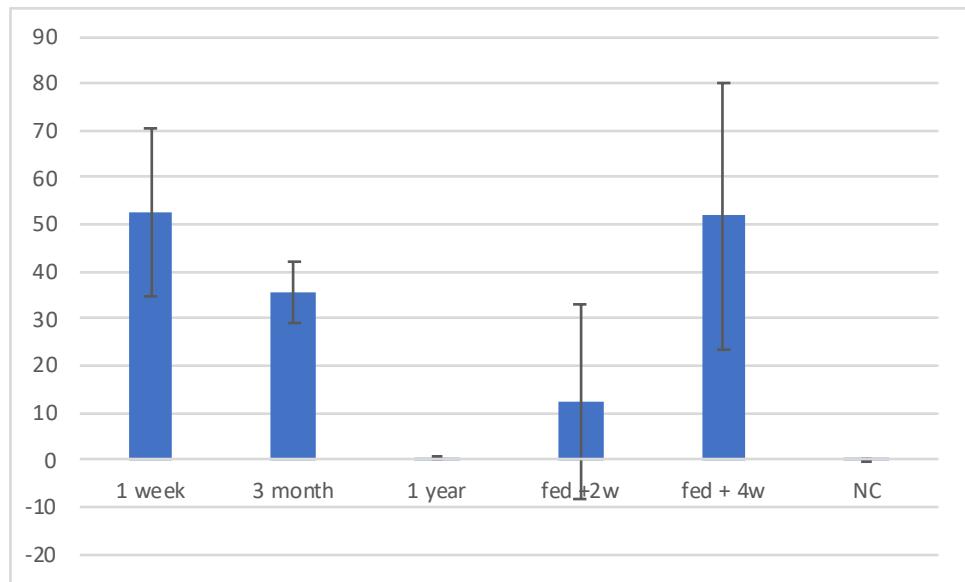
読者の解釈
→



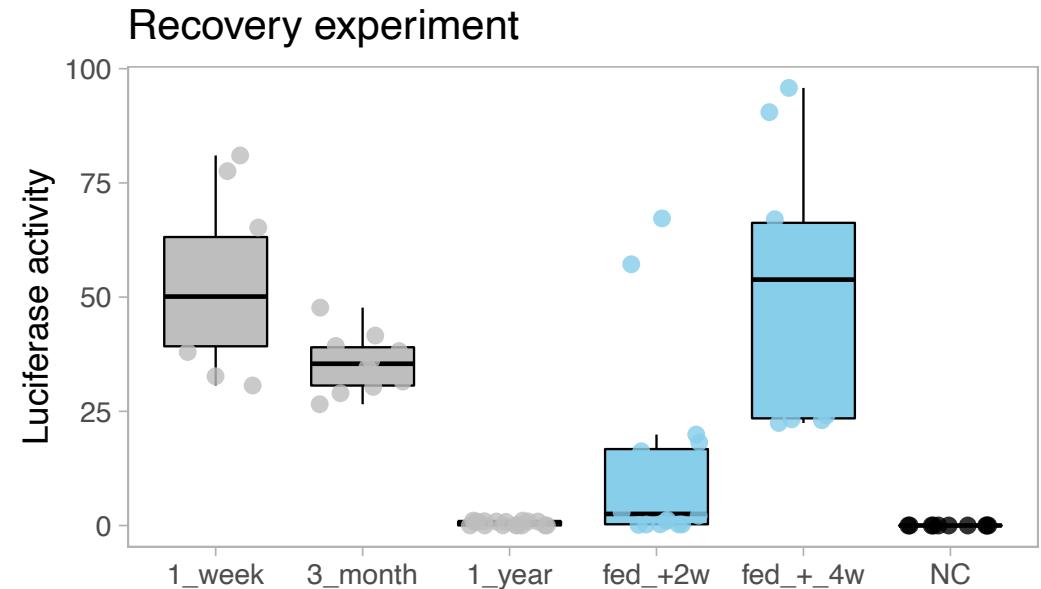
推定されるデータ

Excelだと直感的に操作できるが・・・ Rを使った描画の方が自由度が高い。・・・でも難しい

Excelで作成



PlotsOfDataで作成

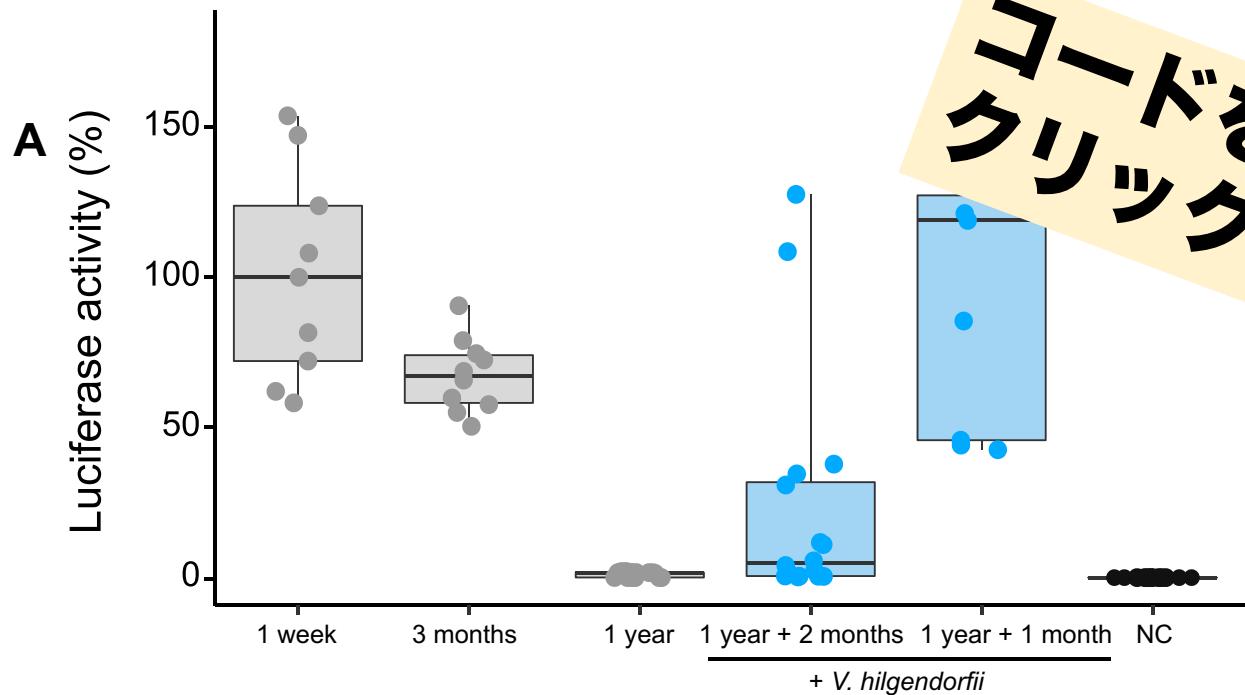


せいぜい棒グラフ + エラーバー

実際のデータのばらつきが一目瞭然
*生物学の生データはばらつきが大きい

論文で使用した図の例：Rで作成

SCIENCE ADVANCES | RESEARCH ARTICLE



Bessho-Uehara et al., Sci. Adv. 2020; 6: eaax4942

コードを書くのが難しい
クリックだけでできないの？

```

1 library(grid); library(ggplot2)
2 ~
3 ~
4 dat <- read.table("RecoveryExp.txt", header = TRUE)
5 head(dat)-
6 typeof(dat)-
7 # is.factor(dat)-
8 # is.list(dat)-
9 ~
10 ~
11 ~
12 #length(dat)-
13 #class(dat)-
14 #str(dat)-
15 ~
16 XLabel <- c("1W", "3M", "1Y", "V2W", "V4W", "buffer")-
17 ~
18 condition<-dat[,6]
19 dat<-dat[,1:5]
20 dat<-cbind(dat,condition)
21 dat<-dat %>% mutate(condition= factor(condition, levels = c("1W", "3M", "1Y", "V2W", "V4W", "buffer")))
22 ~
23 ~
24 ~
25 ~
26 ~
27 ~
28 ~
29 ~
30 p1<-ggplot(dat, aes(x=condition, y=Light_Intensity))
31 B2 <- ggplot(dat, aes(x=condition, y=Light_Intensity))
32 B2<-B2+geom_boxplot(alpha = I(1 / 1.5),aes(color=condition)) + scale_color_manual(values=c("#BBBBBB", "#56B4E9", "#56B4E9"))
33 print(B2)-
34 ~
35 #draw scatter plot on box plot-
36 C <- ggplot(dat2, aes(condition, Light_Intensity))
37 C<-C+geom_jitter(size=2, alpha = I(1 / 1.5),aes(color=condition), position=position_jitter(0.15))
38 print(C)-
39 ~
40 D <- ggplot(dat2, aes(condition, Light_Intensity, fill=condition)) +
41 geom_boxplot(alpha = I(1 / 1.8), coef=10) +
42 scale_fill_manual(values = c("#BBBBBB", "#BBBBBB", "#BBBBBB", "#56B4E9", "#56B4E9", "#BBBBBB")) +
43 geom_jitter(size=2, alpha = I(1),aes(color=condition), position=position_jitter(0.15)) +
44 scale_color_manual(values = c("#999999", "#999999", "#999999", "#00abff", "#00abff", "#111111")) +
45 theme_classic()
46 print(D)-
47 ~
48 #export figure as .pdf format.-
49 pdf("Fig4A.pdf", width = 6, height = 3)-
50 print(D)
51 dev.off()-
52 ~

```

PlotsOfData

- Microsoft Office Excelでは描画が困難なグラフを簡単に描くことができる無料Webツール
- R言語（プログラミング）を未習得でも、マウス操作だけで使用できる
- 作成できる図の種類：箱ひげ図box plot、バイオリンプロットviolin plot、実測値の散布図jitter plot

PlotsOfData - Plots all Of the Data

Data upload

Example 1 (wide format)
 Example 2 (tidy format)
 Upload file
 Paste data
 URL (csv files only)
 These data are Tidy

Select and order:

Download in tidy format (csv)

Show information on data formats

Data upload Plot Data Summary About

Data as provided

Show 100 entries Search:

	B_lo	A_hi	C_med
	0.807567	1.04029	0.846366
	0.707803	1.00078	0.918764
	0.871154	1.21597	0.912502
	0.831443	1.21458	1.0055
	0.983511	1.39795	0.891115
	0.766772	0.828419	1.14641
	0.720085	1.11681	0.866341
	0.76409	1.52753	0.964837
	0.633214	0.983924	1.18332
	0.834569	0.876815	0.730115
	0.809078	0.716905	0.850897

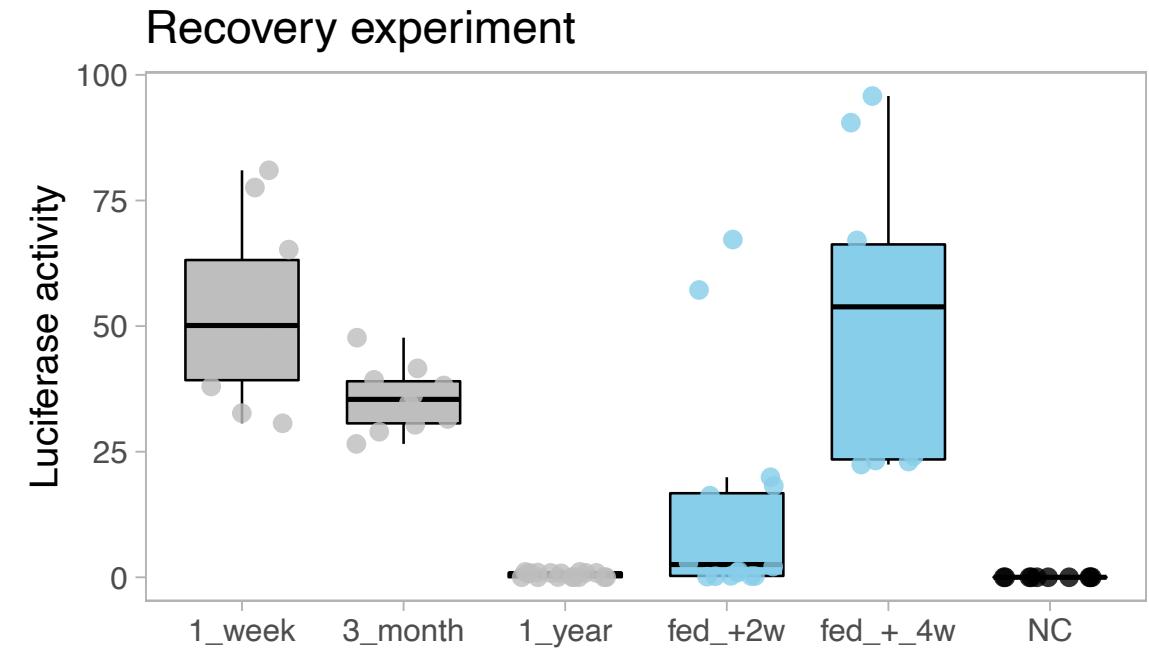
<https://huygens.science.uva.nl/PlotsOfData/>

PlotsOfDataを実際に使ってみましょう

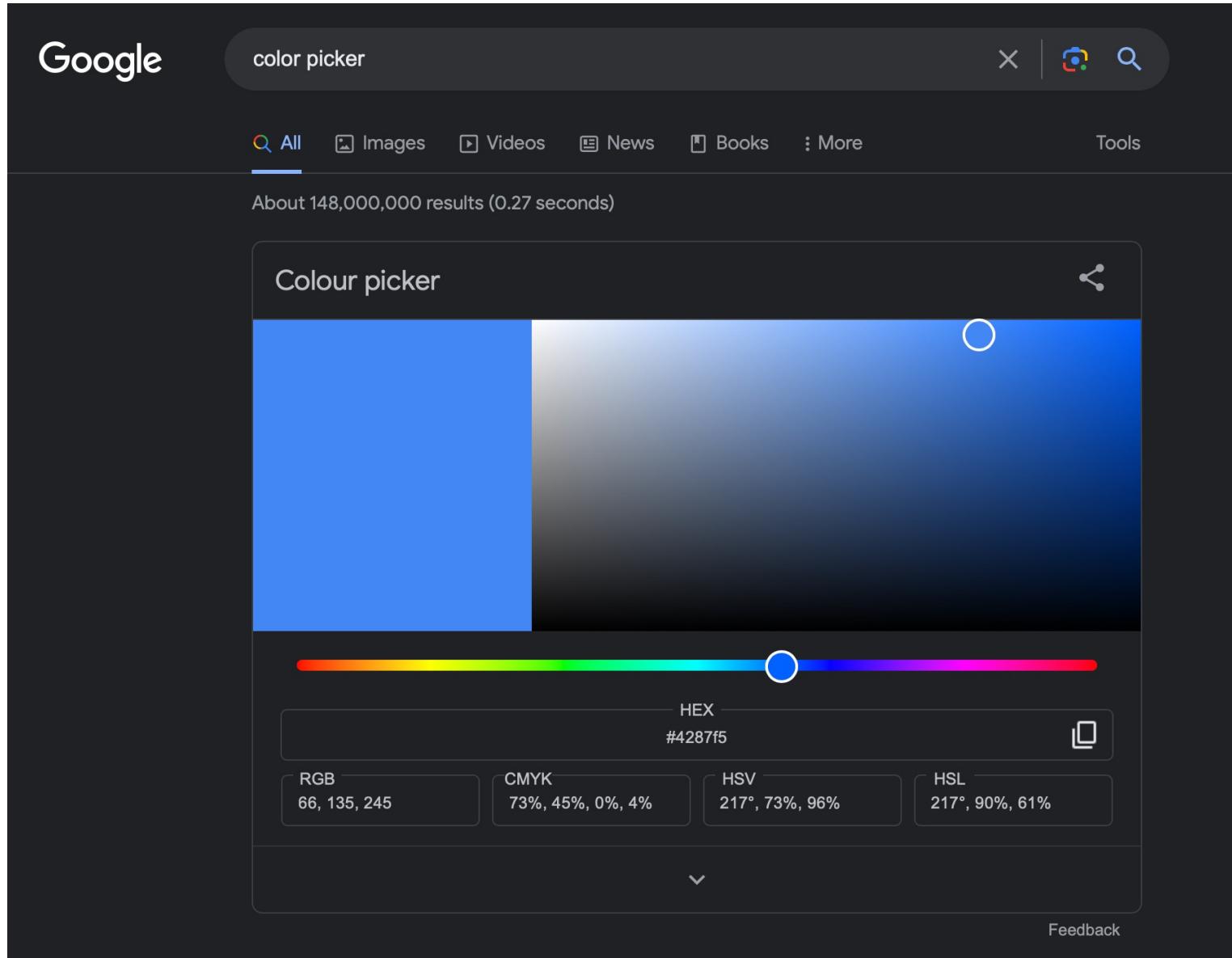
PlotsOfDataを実際に使って下の図をつくってみましょう

生データ

1 week	3 month	1 year	fed +2w	fed + 4w	NC
37.9856303	36.175131	0.86791909	0.21979109	63.8636832	0.00277214
42.9780064	38.1808444	0.87062374	0.24207158	95.7910653	0.00300237
30.6555355	39.300252	0.89172509	5.80183126	23.2423428	0.00316542
32.6745388	41.5833715	1.07218894	6.16609942	24.1048887	0.00317884
52.6857797	47.694693	1.08353552	0.12884839	90.4911199	0.00447882
77.577645	26.5478596	0.74185849	2.13533489	22.4424631	0.00476185
81.0031189	28.9693109	0.88108609	0.94729841	45.0118385	0.00498712
56.9227783	30.3478445	0.01061569	1.23322514	62.6548496	0.00606317
65.2273337	31.5113204	0.012089	57.1837947	67.0505591	0.00639668
47.5489895	34.647752	0.04909602	67.2238963	23.0405714	0.00364603
		0.05065803	16.2508374		
		0.79891713	18.2009412		
		0.94243687	19.9121511		
		0.01438257	0.24790926		
		0.01640409	0.2830323		
		0.02896217	2.94673483		
		0.03361725			



いろんな色を使ってみよう



Color picker
好きな色を選ぶと
HEXやRGBなどでの
表記を教えてくれる。

Office365からExcelでひげ図も作れるようになった

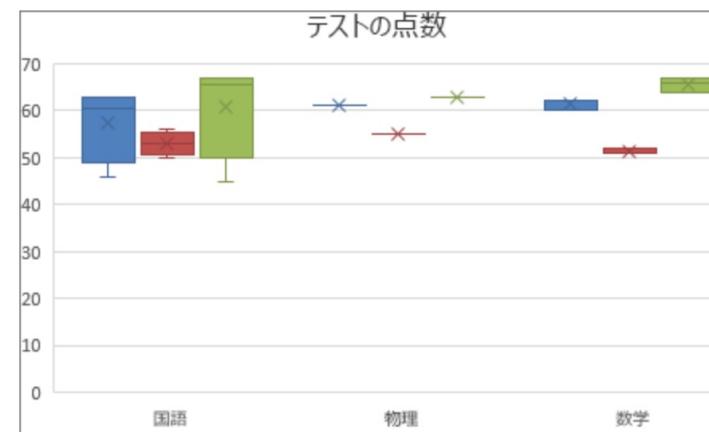


箱ひげ図を作成する

Excel for Microsoft 365, Word for Microsoft 365, Outlook for Microsoft 365, [その他...](#)

箱ひげ図では、データが四分位に分けて示され、平均と特異ポイントが強調されます。箱には、「ひげ」と呼ばれる垂直方向に伸びる線が付いていることがあります。これらの線で、第1四分位数と第3四分位数の外側のばらつきを示し、これらの線、つまり、ひげよりも外側にある点はすべて特異ポイントと見なされます。

箱ひげ図は、統計分析で最も一般的に使用されます。たとえば、箱グラフとウィスカーグラフを使用して、医療試験の結果や教師のテストスコアを比較できます。



リボンの[挿入]タブをクリックし、[統計グラフアイコン]をクリックし、[ボックス]と[ウィスカー]を選択します。

PlotsOfDataの使い方の解説

TO GO TV

[Top](#) [About](#) [Videos](#) [Illustrations](#) [Training](#) [Contact](#) [My page](#)


- Highlights**
- 00:10 1. PlotsOfDataの概要
 - 00:30 2. 対応するデータのフォーマット
 - 00:34 3. Wide data formatのグラフの作成
 - 00:41 4. 使用するファイルの説明
 - 00:51 5. ファイルのアップロード
 - 01:33 6. データのプロットの方法の指定
 - 02:03 7. データのプロット点の透明度の調整
 - 02:24 8. 統計量の指定
 - 03:10 9. 統計量表示の透明度の調整
 - 03:28 10. サンプルの順の変更
 - 03:39 11. グラフの向きや大きさの変更
 - 04:17 12. グラフのサイズ変更
 - 05:35 13. 説明文の追加
 - 04:27 14. ラベルやタイトルの追加
 - 05:50 15. 説明文のコピー
 - 06:04 16. グラフのダウンロード
 - 06:32 17. データの要約統計量の表示
 - 07:09 18. 有効数字の変更
 - 07:20 19. Tidy data formatのグラフの作成

Tags for this video

ウェブツール 可視化 R

Download the video file
220314_PlotsOfData.mov

<https://togotv.dbcls.jp/en/20220314.html>

PlotsOfDataの使い方の解説



2019-04

02 データを可視化するwebツール PlotsOfData

HTS (NGS) 関連のインフォマティクス情報についてまとめています。

2019/4/11 twitterリンク追記

2020/3/2 関連リンク追加

近年、いくつかのグループが、データの要約ではなく実際のデータをグラフで表示することを提唱している[ref.1-4]。生データは、ヒストグラムやドットプロットなど、さまざまな方法で視覚化できる。データ解釈を補助するにはデータ要約が表示されてもよい。さらに、95%信頼区間（95CIs）が与えられている場合、異なるカテゴリー/条件の直接比較は「視覚的推論」によって行うことができる[ref.5,6]。

データとその要約を引き出すために、いくつかの市販のソフトウェアパッケージが利用可能である。しかしながら、理想的には、そのようなツールはオープンソースで、自由に利用可能で、そしてユーザによる貢献または修正を可能にするべきである。生データと要約の組み合わせをプロットするための無料のオープンソースWebベースアプリケーションの一例はBoxPlotR（<http://shiny.chemgrid.org/boxplotr/>）である。このWebベースのアプリは非常によく引用されている論文[ref.7]に記載されている。その人気は、publication品質のデータ視覚化を生成する使いやすいアプリケーションに対する要求を反映している。ただし、この人気のあるオンラインツールはデータの要約としてボックスプロットに偏っており、データと要約の組み合わせ表示をカスタマイズするためのオプションはほとんどない。さらに、プロットは外観がかなりベーシックになっている。

Rのパッケージggplot2を使用すると、最先端のデータの視覚化が可能になる。これは、「grammar of graphics」のアイデアを使用して、複数のデータ層を使用してグラフィックを生成する[ref.8]。マルチレイヤー アプローチにより、個々のコンポーネントからグラフを作成することができる。各コンポーネントは個別に調整できる。データレイヤーに透明度を適用するオプションは、柔軟性を増す。それでも、ggplot2によって提供される高品質のデータ視覚化には、コーディングスキルときちんとしたデータの概念を理解することが必要になる[ref.9]。

生データの統計サマリーを伴った最先端のデータ視覚化方法を民主化するために、PlotsOfDataと名付けたWebツールを作成した。このWebツールは通常のスプレッドシート（ワード）formatおよびtidy data formatでデータを処理し、ggplot2を使用してグラフを作成する。PlotsOfDataを使ってグラフを作成するのにコーディングスキルは必要なく、ggplot2によって提供される高品質のデータ視覚化は誰にでも利用できる。

PlotsOfDataは<https://huygens.science.uva.nl/PlotsOfData/>でオンラインで利用できる。アプリはshinyパッケージを使用し、R (<https://www.r-project.org>) とRstudio (<https://www.rstudio.com>) を使用して、Rで書かれている。いくつかの無料で入手可能なパッケージ（shiny、ggplot2、dplyr、readr、magrittr、

- プロフィール

上坂一馬 (id:kazumaxneo)
最終更新: 31日前
+ 読者になる (176) [@kazumachackをフォロー](#)

このブログについて
- 検索

記事を検索
- 最新記事

MAGの株レベルでの定量を可能にする MAGinator
ショートリードアセンブリからプラスミドコンティグを同定する PLASMe
ロングリードを用いた De Novo ゲノムアセンブリとマッピングのためのツールキット ntLink
パンゲノムグラフの構築と探索を行う pantoools
ゲノムのGFF3アノテーションファイルを扱う AEGeAn Toolkit
- 月別アーカイブ

▶ 2023 (171)
▶ 2022 (247)
▶ 2021 (329)
▶ 2020 (344)
▼ 2019 (352)

 - 2019 / 12 (27)
 - 2019 / 11 (31)
 - 2019 / 10 (31)
 - 2019 / 9 (29)
 - 2019 / 8 (24)
 - 2019 / 7 (31)
 - 2019 / 6 (31)
 - 2019 / 5 (28)
 - 2019 / 4 (30)
 - 2019 / 3 (33)
 - 2019 / 2 (28)
 - 2019 / 1 (29)

▶ 2018 (459)
▶ 2017 (259)

PlotsOfDataの使い方の解説

 羊土社
生命科学と医学の専門出版社

書籍を探す 検索ボタン

FAQ 新規登録 ログイン カート

新刊 実験医学 レジデントノート 教科書 書籍・雑誌付録特典

**実験医学増刊 Vol.40 No.17
バイオDBとウェブツール ラボで使える最新70選**
知る・学ぶ・使う、バイオDX時代の羅針盤
小野浩雅／編
2022年10月20日発行 B5判 248ページ ISBN 978-4-7581-0406-7

定価：5,940円（本体5,400円+税） カートに入れる
在庫：あり

本書を一部お読みいただけます

序にかえて

バイオデータベースとバイオインフォマティクスツールを知り、学び、使いこなし、研究成果をつかむために
小野浩雅
(情報・システム研究機構データサイエンス共同利用基盤施設ライフサイエンス統合データベースセンター (DBCLS))

はじめに

生命科学・医学研究の現場において、バイオデータベースとバイオインフォマティクスツール（DBとツール）は研究インフラとして欠かせない役割を担っている。数多あるDBとツールを取捨選択したり、それらを的確に組合せて活用する技術の習熟度合いが、日々の研究活動の効率的な進展に影響するだけでなく、研究プロジェクト全体の可能性を左右するといつても過言ではない。一方で、生命科学・医学分野の実験手法や技術革新は目覚ましく、またインターネット技術の驚異的な進歩とともに、その種類や対象、データ量は拡大の一途をたどっている。

DBとツールの種類や対象が日々増加するなかで、初学者にとっては、どのような場面で活用できるのか、まずは何を選ぶべきか、どう使い、組合わせることができるのか、などについて、その全体像を俯瞰し客観的に取捨選択することが難しくなってきている。また、DBとツールの活用についてその重要性を認識しているような経験者であっても、最初に教わったDBとツールを使い続けてしまいかで、新たなそれらに手を伸ばしづらいという声も多い。

本増刊号は、読者が多種多様なDBとツールを知り、学び、使いこなすため、そして研究という大海原を進んでいくための「羅針盤」となるべく企画したものである。各分野で目的別に厳選した最新のDBとツールを、生命科学・医学を専門とする広いレベルの研究者に向けて紹介している。特にバイオインフォマティクスに明るくない実験系研究者であっても今すぐ使えることを重視し、基本的にコーディングを必要とせず

<https://www.yodosha.co.jp/yodobook/book/9784758104067/3.html>

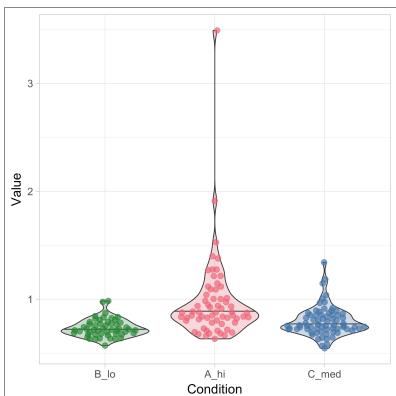
Huygensらが提供するクリックだけで使えるウェブツール

Shiny apps – Molecular Cytology / LCAM

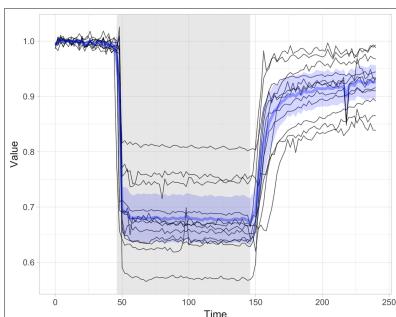


Introduction

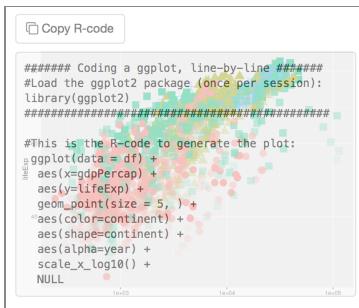
On this page you will find the links (just click on the image) for a number of web-based #DataViz tools (Shiny apps) with a short description. More information on the purpose of the app can be found in the "about" tab after running the tool. Some more information can be found at the bottom of this page.



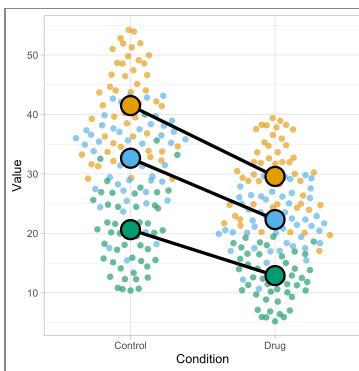
[PlotsOfData](#) visualizes (unpaired) data for different conditions together with a choice of statistical summaries. Also, a customizable output table with statistics is generated. The importance of depicting the actual data (instead of summaries only) is discussed in [this blog](#)



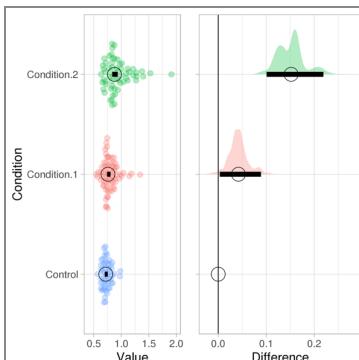
[PlotTwist](#) was developed for visualizing data from time-lapse imaging but can be used for any type of continuous data. It features several visualizations, including the small multiple and heatmap. See also [this blog](#)



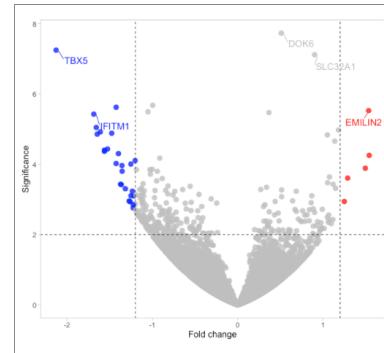
[ggPloteR](#) was developed to demonstrate how ggplot2 can be used to encode a data visualization, line-by-line. After generating the plot, the R-code can be retrieved and directly used in R to (re-)create the plot.



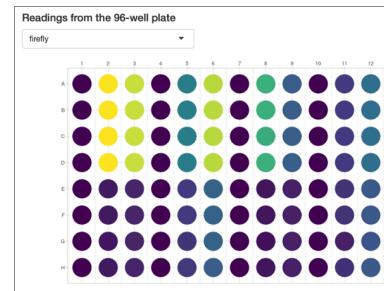
[SuperPlotsOfData](#) is an extension of PlotsOfData that enables the identification data from replicate experiments and their analysis. The summary data from the replicates is used for the quantitative comparison of conditions. The difference to a reference is calculated to give an 'effect size', which has advantages over null-hypothesis tests for comparison of samples, see [this blog](#)



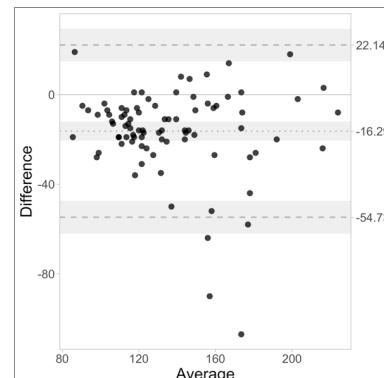
[PlotsOfDifferences](#) is an extension of PlotsOfData that plots the differences between the mean or median of (several) conditions and the mean or median of a selected reference condition. The difference is a type of effect size and has some advantages over null-hypothesis tests for comparison of samples, see [this blog](#)



[VolcaNoseR](#) is a Shiny app for generating and nosing around volcano plots. The app allows user-defined cut-offs for the fold-change and significance. Lists with top candidates are shown and these can be annotated in the plot. In addition, user-selected candidates can be added as labels to the plot.



[PlotXpress](#) is a Shiny app for processing and visualizing data from dual-luciferase expression assays. The app accepts data from the Promega GloMax reader. It uses the renilla signal for normalization of expression data and plots the relative expression grouped per condition and for different treatments.



[BA-ploteR](#) is a Shiny app for generating Bland-Altman plots. These plots are used to evaluate the quantitative difference between measurements. The input is paired data from repeated measurements. When two different methods are used for measurements on the same subject, the Bland&Altman plot is a useful way of depicting their differences (bias).

お疲れ様でした。

講義当日は、お手持ちのデータをつかって
PlotsOfDataを使っていただいた時の質問などお受けします。