Pemanfaatan R untuk Statistik dan Visualisasi Data

Tailored by: Dr. Adi Wijaya, MKom

Webinar UPT Perpustakaan dan Undip Press

Semarang, 16 April 2025

Short Bio

Dr. Adi Wijaya, MKom

Affiliation: Universitas Indonesia Maju, Jakarta
HP/WA: +62-838-789-19-456
Email: adiwjj@gmail.com | adiwjj@uima.ac.id | adi@metlit.net

I am a Lecturer and Research Fellow at Universitas Indonesia Maju (UIMA), Jakarta, as well as an IT professional specializing in data science, data governance, enterprise architecture, and software engineering. Holding a Doctor in Electrical Engineering with a specialization in Information Technology from Universitas Gadjah Mada (2021), my research focuses on information processing, data mining, machine learning, and health informatics, including brain-computer interfaces.

Expertise:

Data Mining/Data Science/Business Intelligence, Artificial Intelligence/Machine Learning, Secondary Study (Bibliometric Analysis, Scoping Review, Systematic Review, Meta-analysis), Applied Statistics, Enterprise Architecture, Software Engineering

Contents

Sekilas tentang R

R IDE/GUI

R library untuk Uji Statistik

R library untuk Visualisasi Data

Publication quality R plots – a use case

Contents

Sekilas tentang R

R IDE/GUI

R library untuk Uji Statistik

R library untuk Visualisasi Data

Publication quality R plots – a use case

Sekilas tentang R

Karakteristik
Open source dan cross-platform (Windows, macOS, Linux) (Das et al., 2023; Denton, 2011)

Sejarah
Dikembangkan pertama kali oleh Ross Ihaka dan Robert Gentleman pada tahun 1993 (Tollefson, 2022; Core, 2014)

Aplikasi
Digunakan secara luas dalam statistik, bioinformatika, epidemiologi, dan machine learning (Tuimala & Kallio, 2013; Hadi, 2022)

R untuk Statistik dan Data Visualisasi

Komprehensif

R menawarkan ribuan paket untuk analisis statistik dari dasar hingga lanjutan (Giorgi et al., 2022)

Fleksibel

R mendukung pembersihan, transformasi, analisis, dan visualisasi data dalam satu platform (Yi, 2023; Lizana, 2020)

Visualisasi Kuat

R memiliki paket seperti ggplot2, plotly, dll untuk visualisasi data yang dapat disesuaikan dan siap publikasi (Nordmann, 2022; Lüdecke et al., 2021)
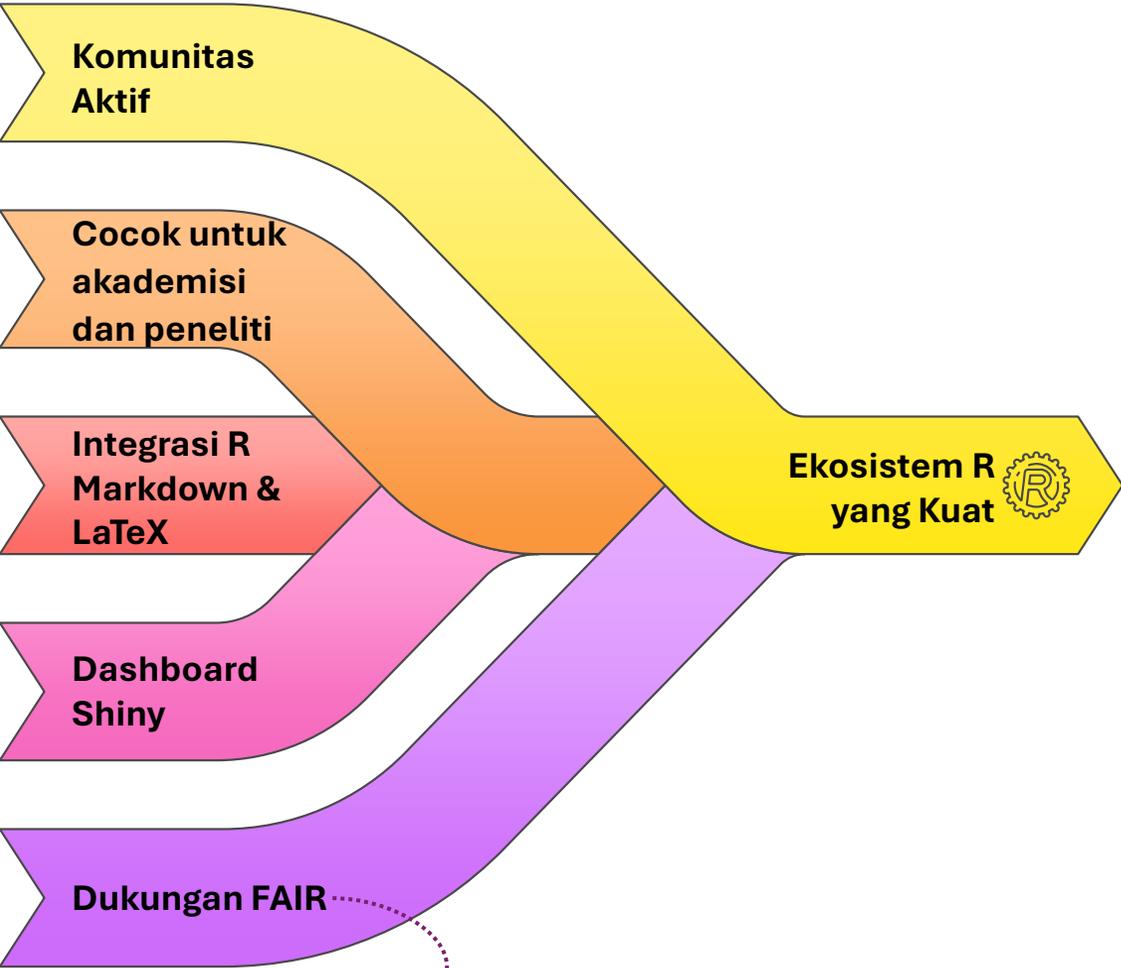
Penelitian Reproduksibel

R banyak digunakan di akademik untuk menghasilkan penelitian yang dapat direproduksi (Ponce & Sandhel, 2021; Giorgi et al., 2022)

Kelebihan & Ekosistem R

1 Dukungan dari CRAN, StackOverflow, dan R-bloggers mempercepat proses belajar dan problem solving (Jones et al., 2023; Giorgi et al., 2022)

2 Digunakan luas dalam bioinformatika, paleobiologi, ilmu sosial, hingga klinis (Johnson, 2020)

3 Dokumen interaktif, reproducible (Konkol et al., 2020)

4 Dashboard web interaktif untuk riset & uji klinis (Golafshar et al., 2021)

5 GitHub dan CRAN memfasilitasi kolaborasi, dokumentasi, dan akses terbuka (Tso et al., 2022; Vuorre & Crump, 2021; Guzura, 2023)

Komunitas Aktif

Cocok untuk akademisi dan peneliti

Integrasi R Markdown & LaTeX

Dashboard Shiny

Dukungan FAIR

Ekosistem R yang Kuat

Findable, Accessible, Interoperable, and Reusable

Contents

R IDE/GUI

RStudio
IDE R paling populer dengan fitur lengkap (Wilson, 2012; Giorgi et al., 2022)

Posit Cloud
Versi RStudio berbasis cloud untuk kolaborasi (Giorgi et al., 2022)

Jupyter Notebook
Menggabungkan kode R, teks, dan visualisasi (Giorgi et al., 2022)

VSCode
Mendukung scripting R melalui ekstensi (Giorgi et al., 2022)

R Commander | **jamovi** | **RKWard**
Antarmuka ramah pengguna untuk non-programmer (Valero-Mora & Ledesma, 2012; Wilson, 2012), Edukasi serta eksplorasi lanjutan (Rödiger et al., 2012; Alcazer, 2020)

Cara Install R (Windows/Linux/Mac)

R 4.4.0

4 Buka R GUI

Luncurkan antarmuka grafis pengguna R setelah instalasi selesai.

Ru

3 Ikuti Instruksi

Jalani proses instalasi dengan mengikuti petunjuk di layar.

2 Pilih OS

Pilih sistem operasi yang sesuai untuk instalasi R.

1 Kunjungi Situs Resmi

Akses situs web R resmi untuk memulai proses instalasi

https://cran.r-project.org
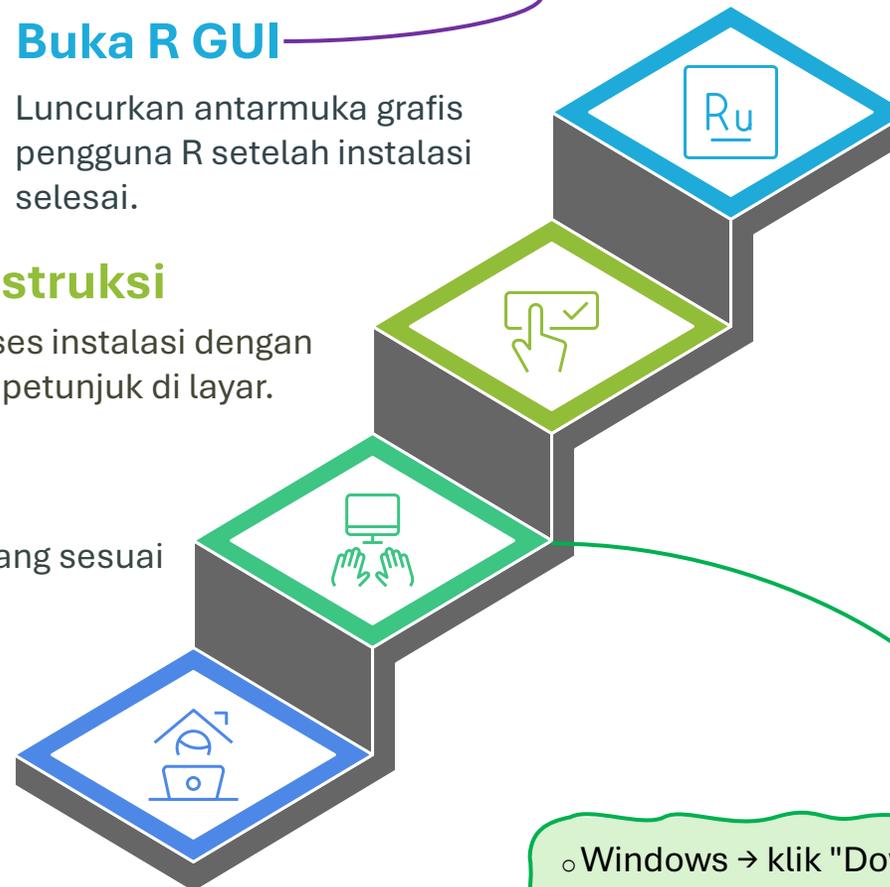
- Windows → klik "Download R for Windows" → "base"

- Mac OS → klik "Download R for (Mac) OS X"

- Linux → pilih distribusi Linux (Debian/Ubuntu/RedHat)

Cara Install RStudio

RStudio ▶ R RStudio

Jalankan RStudio

R base **harus sudah terinstal terlebih dahulu**

Instal Perangkat Lunak

Pilih Versi

Kunjungi Situs Web

https://posit.co/download/rstudio-desktop/
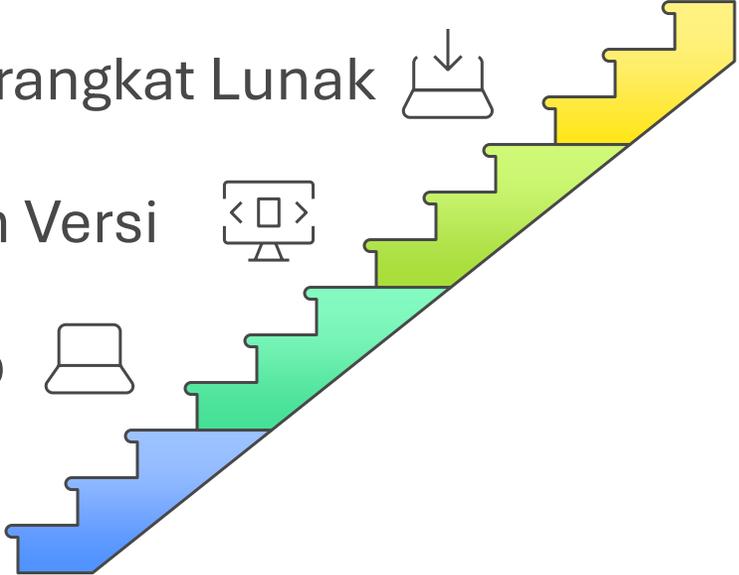
Install Anaconda: Langkah Awal untuk Jupyter dan R

◆ **Apa itu Anaconda?**

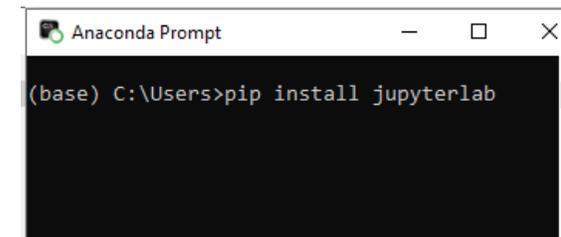Distribusi Python & R yang sudah termasuk:

- ▶ JupyterLab/Notebook

- ▶ Manajer paket: pip atau conda

- ▶ Paket sains/data: numpy, pandas, matplotlib, dll.

◆ **Cara Install:**

1. Download di: anaconda.com

2. Pilih installer sesuai OS (Windows/macOS/Linux)

3. Jalankan installer → Ikuti instruksi (Next → Finish)

◆ **Install JupyterLab (via pip, jika tidak pakai Navigator):**

```
pip install jupyterlab
```

Menjalankan R di JupyterLab

🔁 Integrasi R dengan Jupyter

◆ Install IRkernel di R: Buka R (atau RStudio)

◆ lalu jalankan:

```R
install.packages("IRkernel")
IRkernel::installspec()  # untuk mendaftarkan ke Jupyter
```

Buka **JupyterLab**, lalu buat notebook baru dengan kernel R.
1. Buka **Anaconda Prompt**
   ◆ Klik Start → cari "Anaconda Prompt" → buka
   ◆ Ini akan membuka jendela command line (CMD)
2. Jalankan **JupyterLab**, ketik: `jupyter-lab`
3. Browser akan terbuka otomatis → Akses JupyterLab

Contents

Library/Package R untuk Statistik

▶ **stats (default bawaan R)**: regresi, ANOVA, uji normalitas (Vandever, 2020)

▶ **car**: uji asumsi regresi, multikolinearitas (Salmerón-Gómez et al., 2021)

▶ **psych**: statistik deskriptif, reliabilitas, EFA (Khojasteh, 2022)

▶ **lmtest, sandwich**: uji model linear dan heteroskedastisitas (Chen et al., 2022)

▶ **nortest, tseries**: uji normalitas dan stasioneritas (Shalabh, 2022)

▶ *dan banyak lagi yang lainnya...*

Uji Statistik yang Banyak Dipakai di Paper Ilmiah dengan R

▶ Uji normalitas → `shapiro.test(), ks.test()`

» penting untuk validasi asumsi uji parametrik (Erlichman, 2023)

▶ Uji chi-square → `chisq.test()`

» untuk hubungan antar variabel kategorik (Erlichman, 2023)

▶ Korelasi Pearson dan Spearman → `cor.test()`

» mengukur kekuatan dan arah hubungan variabel kontinu (McDonald & Levine, 2022)

▶ Regresi linear dan logistik → `lm(), glm()`

» untuk hubungan variabel kontinu dan biner (Fusfeld, 2022)

Uji Statistik yang Banyak Dipakai di Paper Ilmiah dengan R

▶ Uji t dan ANOVA → `t.test(), aov()`

  » untuk perbandingan rata-rata antar kelompok (Akbulut, 2023; Rahrig, 2024)

▶ Uji Mann-Whitney U dan Kruskal-Wallis → `wilcox.test(), kruskal.test()`

  » alternatif non-parametrik untuk uji beda (Rahrig, 2024; Ergin & Koşkan, 2023)

▶ Uji Multivariat: PCA dan EFA → `prcomp(), factanal()`

  » untuk reduksi dimensi dan eksplorasi struktur laten (Mwambi, 2022)

▶ *dan banyak lagi yang lainnya...*

Contoh...

```r
# Uji Korelasi Pearson: disp vs hp
pearson_result <- cor.test(mtcars$disp, mtcars$hp, method = "pearson")

# Tampilkan hasil
cat("Korelasi Pearson antara disp dan hp:\n")
print(pearson_result)
```

```
Korelasi Pearson antara disp dan hp:

        Pearson's product-moment correlation

data:  mtcars$disp and mtcars$hp
t = 7.0801, df = 30, p-value = 7.143e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6106794 0.8932775
sample estimates:
      cor
0.7909486
```

Pilihan method:
pearson, spearman, kendall

P-value = 7.143×10^{-8}
Kecil sekali, di software statistik biasanya disederhanakan menjadi: <0.001

Koefisien korelasinya = 0.791

Contoh...

```r
# Regresi linier: mpg + disp --> hp
model <- lm(hp ~ mpg + disp, data = mtcars)

# Tampilkan ringkasan model regresi
summary(model)
```

Regresi linier
hp = mpg + disp

hp = 172.2204 – (4.2732 * mpg) + (0.2614 * disp)

```
Call:
lm(formula = hp ~ mpg + disp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
 -48.70  -17.67  -10.16   10.12  148.19

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  172.2204    69.9014   2.464   0.0199 *
mpg           -4.2732     2.3027  -1.856   0.0737 .
disp           0.2614     0.1120   2.335   0.0267 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.01 on 29 degrees of freedom
Multiple R-squared:  0.6653,    Adjusted R-squared:  0.6423
F-statistic: 28.83 on 2 and 29 DF,  p-value: 1.279e-07
```

Contents

Library/Package R untuk Visualisasi Data

▶ **ggplot2**: standar emas visualisasi berbasis grammar of graphics (Nordmann, 2022; Cramer et al., 2022)

▶ **lattice**: visualisasi multivariat berbasis formula (Nordmann, 2022)

▶ **plotly**: interaktif dan integrasi dengan ggplot2 (Nordmann, 2022)

▶ **highcharter**: visualisasi interaktif berbasis Highcharts (Nordmann, 2022)

▶ **corrplot**: visualisasi korelasi (Chevalier, 2023)

▶ **gganimate**: animasi visualisasi dari ggplot2 (Nordmann, 2022)

▶ **tidyplots**: grafik eksploratif yang bersih dan konsisten (Engler, 2025)

▶ *dan banyak lagi yang lainnya...*

Library/Package R untuk Visualisasi Data

▶ **caret**: framework ML yang simpel dan konsisten (Janiesch et al., 2021)

▶ **mlr3**: modular, powerful, pengganti mlr (Janiesch et al., 2021)

▶ **randomForest, xgboost, e1071** (SVM): model ML klasik (Janiesch et al., 2021)

▶ **keras & tensorflow**: binding ke Python untuk deep learning (Li & Bezemer, 2022)

▶ **h2o**: scalable ML dan DL untuk big data (Janiesch et al., 2021; Chaves et al., 2023)

▶ **mlf-core**: framework untuk model deterministik dan reproducible (Heumos et al., 2023)

▶ *dan banyak lagi yang lainnya...*

Contoh Visualisasi Data Populer dalam Paper Ilmiah dengan R

Beberapa jenis visualisasi yang umum digunakan dalam R meliputi:

◆ Boxplot, Histogram, Density Plot → *untuk menampilkan distribusi dan sebaran data (Nordmann, 2022)*

◆ Scatter Plot dan Bubble Plot → *menggambarkan hubungan antara dua atau lebih variabel (Nakagawa et al., 2023)*

◆ Heatmap dan Korelogram → *visualisasi matriks dan hubungan korelasi (Berker et al., 2022)*

◆ Bar Plot dan Pie Chart → *perbandingan proporsi kategori; perlu kehati-hatian pada pie chart (Nguyen et al., 2021)*

◆ Line Plot (Tren/Time Series) → *pengamatan pola dari waktu ke waktu (Gerela et al., 2022)*

◆ PCA Biplot, Dendrogram → *untuk analisis dimensi dan klasterisasi data kompleks (Nordmann, 2022)*

◆ Kaplan-Meier Curve → *visualisasi dalam analisis survival, umum dalam studi klinis (Andrade, 2023)*

Visualisasi data berperan penting dalam menyampaikan informasi kompleks secara ringkas dan intuitif, terutama dalam konteks publikasi ilmiah

(Berker et al., 2022; Setzer & Cui, 2022)

R untuk NLP dan LLM

📌 NLP dengan R:

▶ tm, quanteda, textclean → *untuk pembersihan dan analisis awal teks* *(Prasad et al., 2022)*

▶ Tidytext → *mengadopsi paradigma tidyverse untuk text mining yang efisien* *(Prasad et al., 2022)*

▶ text2vec, udpipe → *mendukung embedding dan NLP klasik, cocok untuk sistem retrieval dan chatbot akademik* *(Prasad et al., 2022)*

🤖 LLM dan AI:

▶ Reticulate → *integrasi R dan Python, memungkinkan akses ke GPT-3/4 dan model HuggingFace* *(Serdaliyev, 2023; Goar et al., 2023)*

▶ chatgpt (unofficial) → *API untuk memanfaatkan ChatGPT dalam aplikasi seperti asisten virtual dan penulisan otomatis* *(Hariri, 2023)*

R menyediakan alat yang semakin kuat untuk aplikasi **Natural Language Processing (NLP)** dan integrasi dengan **Large Language Models (LLM)**, memungkinkan eksplorasi teks dan pengembangan sistem AI percakapan yang canggih.

Contents

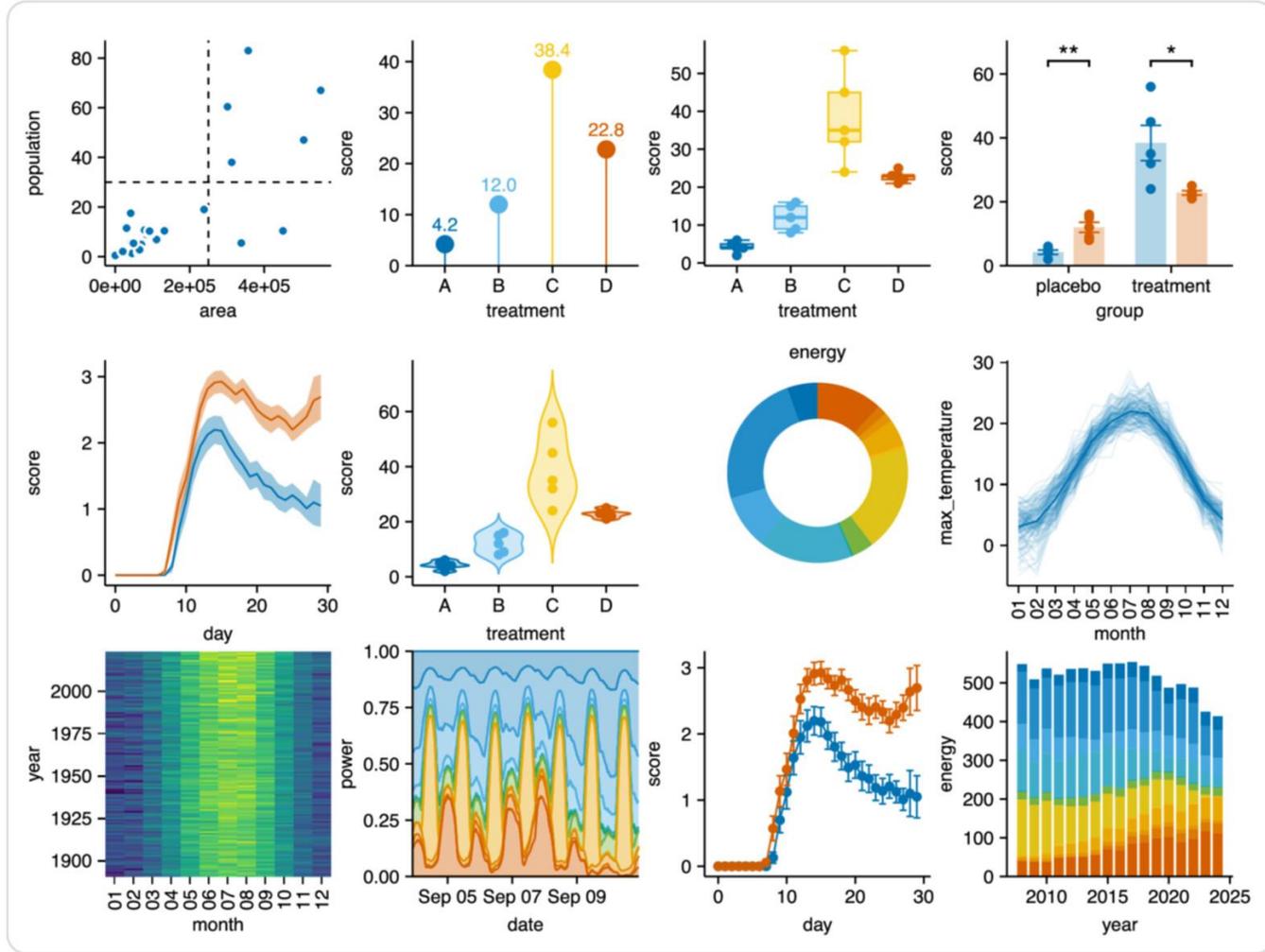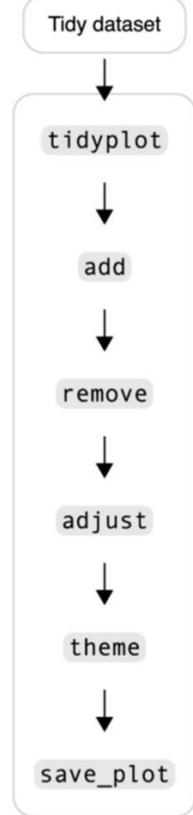# Tidyplots - https://jbengler.github.io/tidyplots/index.html

Engler, Jan Broder. 2025. " Tidyplots Empowers Life Scientists With Easy Code-Based Data Visualization." *iMeta* e70018. https://doi.org/10.1002/imt2.70018
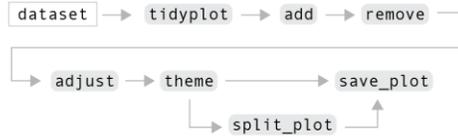
cheatsheet

You can install the released version of tidyplots from [CRAN](#) with:

```r
install.packages("tidyplots")
```

tidyplots cheatsheet
version 0.2.1

@jbengler

Workflow

dataset → tidyplot → add → remove

adjust → theme → save_plot

split_plot

Dataset

Variables

treatment <chr>	group <chr>	dose <chr>	age <dbl>	sex <chr>	score <dbl>
A	placebo	high	23	female	2
A	placebo	high	45	male	4
B	placebo	low	23	female	9
B	placebo	low	45	male	4
C	treatment	high	23	female	32
C	treatment	high	45	male	35
D	treatment	low	23	female	23
D	treatment	low	45	male	25
...	...	...	...	...	...

Observations

Code

```r
study |>
  tidyplot(x = group, y = score, color = dose) |>
  add_mean_bar(alpha = 0.4) |>
  add_sem_errorbar() |>
  add_data_points_beeswarm() |>
  add_title("My study") |>
  add_caption("Petterson et al.")
```

Plot

My study ◄ Title
◄ Axis line
◄ Axis ticks
Height
score
Legend
Legend title
dose
● high
● low
Color labels
placebo   treatment ◄ Axis labels
group ◄ Axis title
Petterson et al. ◄ Caption
Width
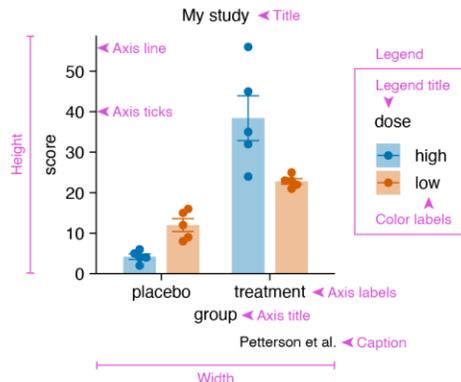
Links

website docs help use-cases

Copyright © 2025 Jan Broder Engler

Add

Raw data   link

add_data_

points()
labels()

Example:
add_data_points()

Summary statistics   link

add_count_
add_sum_
add_mean_
add_median_

bar()
dash()
dot()
value()
line()
area()

Examples:
add_count_bar()
add_sum_line()
add_mean_value()

Dispersion   link

add_sem_
add_sd_
add_range_
add_ci95_

errorbar()
ribbon()

Example:
add_sem_errorbar()

Comparison   link

add_test_

pvalue()
asterisks()

Example:
add_test_pvalue()

Distribution   link

add_boxplot()   add_violin()

Proportion   link

add_barstack()   add_pie()
add_areastack()   add_donut()

Annotation   link

add_title()   add_annotation_text()
add_caption()   add_annotation_rectangle()
                     add_annotation_line()
                     add_reference_lines()

ggplot2 code   link

add()

add(ggplot2::geom_point())

More

add_heatmap()   add_curvefit()

add_histogram()

Adjust

Plot   link

adjust_size()   adjust_description()
adjust_colors()   adjust_padding()
adjust_font()   adjust_legend_title()
                     adjust_legend_position()

Axes   link

adjust_x_axis()   adjust_y_axis()
adjust_x_axis_title() adjust_y_axis_title()

Data labels   link

rename_
reorder_
sort_
reverse_

color_labels()
x_axis_labels()
y_axis_labels()

Examples:
rename_color_labels()
sort_x_axis_labels()
reorder_y_axis_labels()

Remove

Axes   link

remove_ → x_axis_ → line()
          y_axis_   ticks()
                      labels()
                      title()

Examples:
remove_x_axis()
remove_x_axis_title()

More

remove_title()   remove_legend()
remove_caption()   remove_legend_title()
remove_padding()

More

Color schemes   link

colors_discrete_
  friendly
  seaside
  apple
colors_continuous_
  viridis
  plasma
  mako
colors_diverging_
  blue2brown
  blue2red
  spectral

Split plot   link
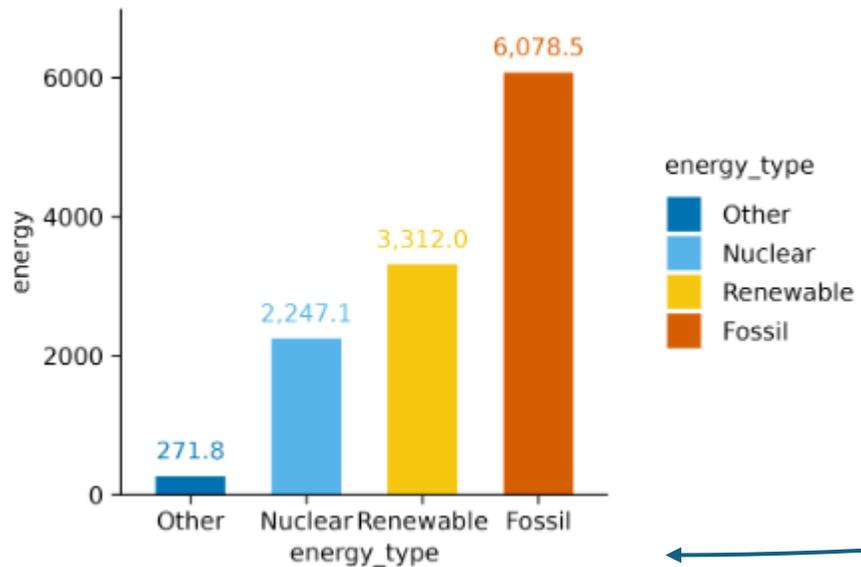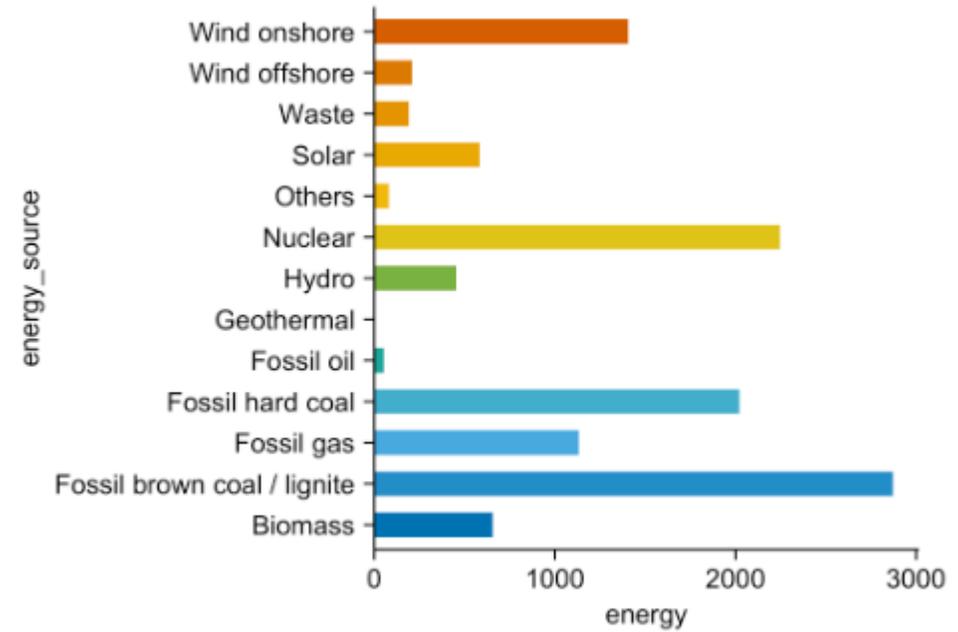
split_plot(by, ncol, nrow, ...)

Output   link

save_plot("my_plot.pdf")
view_plot()

Datasets   link

animals   energy_week
climate   gene_expression
dinosaurs   spendings
distributions   study
energy   time_course
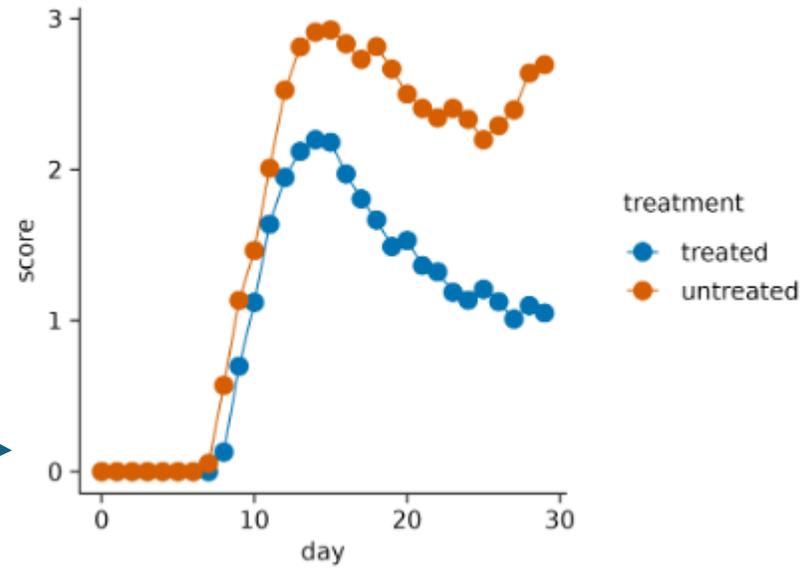
barplot

```
energy |>
  tidyplot(x = energy, y = energy_source, color = energy_source) |>
  add_sum_bar() |>
  remove_legend()
```
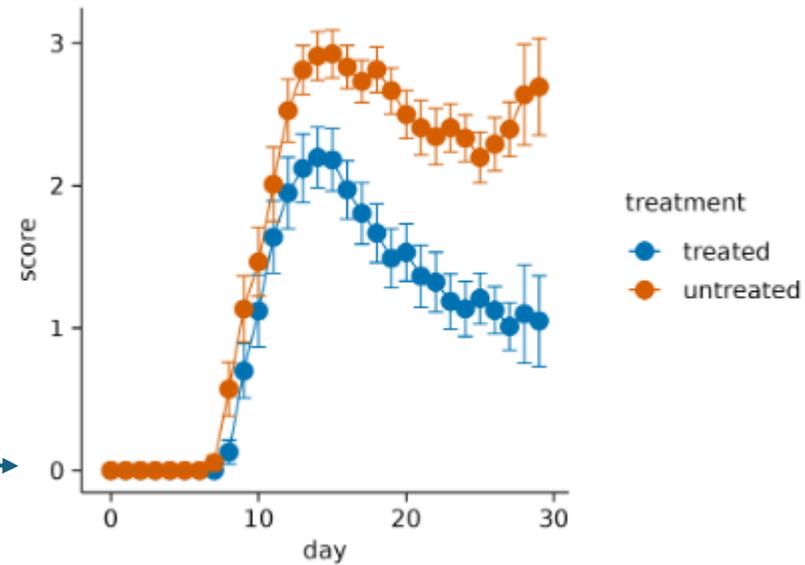


```
energy |>
  tidyplot(x = energy_type, y = energy, color = energy_type) |>
  add_sum_bar() |>
  add_sum_value() |>
  sort_x_axis_labels()
```

line+dot+error plot

```
time_course |>
    tidyplot(x = day, y = score, color = treatment) |>
    add_mean_line() |>
    add_mean_dot()
```
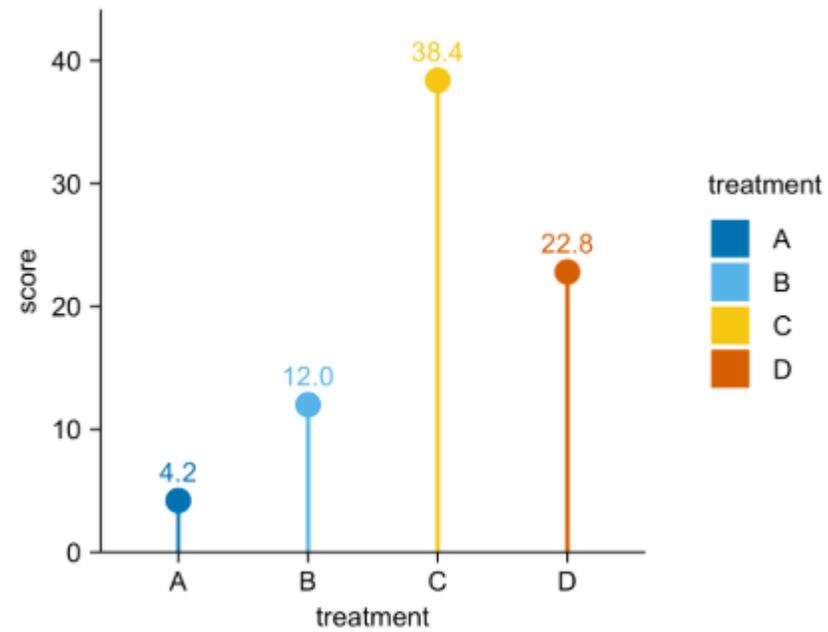


```
time_course |>
    tidyplot(x = day, y = score, color = treatment) |>
    add_mean_line() |>
    add_mean_dot() |>
    add_sem_errorbar(width = 2)
```
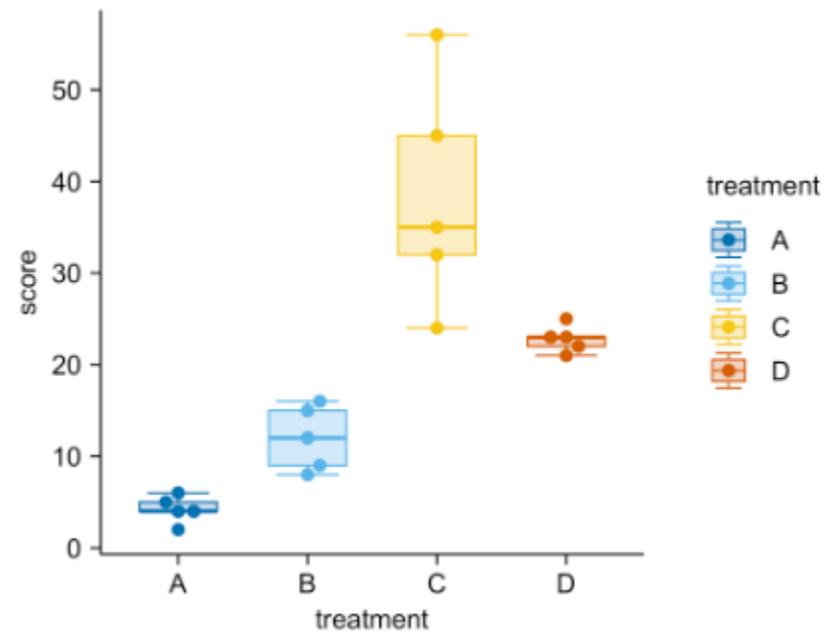
lollipop plot

```
study |>
  tidyplot(x = treatment, y = score, color = treatment) |>
  add_mean_dot(size = 2.5) |>
  add_mean_bar(width = 0.03) |>
  add_mean_value()
```
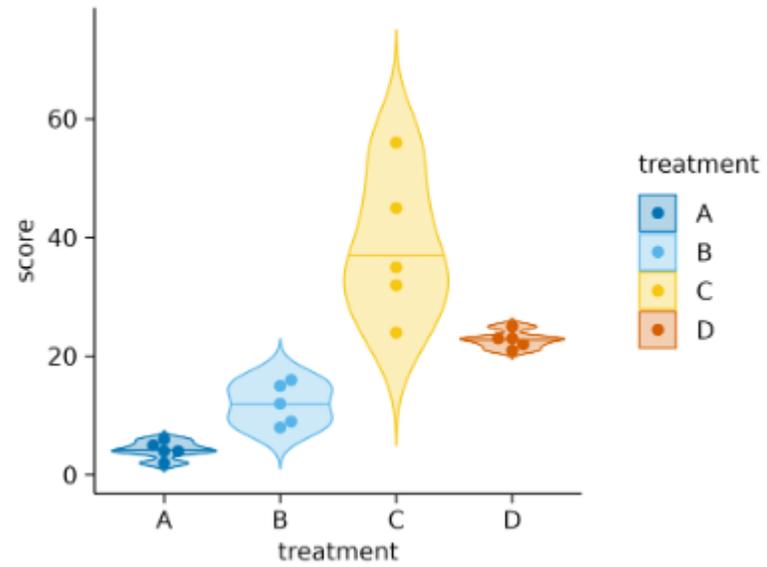
boxplot

```
study |>
  tidyplot(x = treatment, y = score, color = treatment) |>
  add_boxplot() |>
  add_data_points_beeswarm()
```
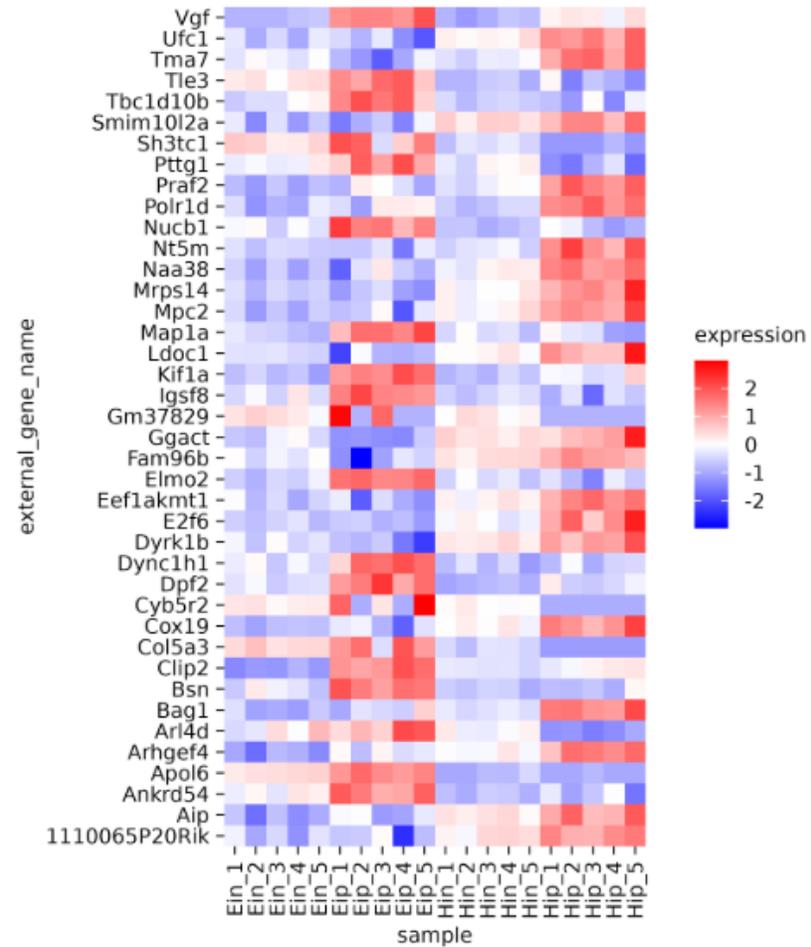
violin plot

```
study |>
    tidyplot(x = treatment, y = score, color = treatment) |>
    add_violin(draw_quantiles = 0.5) |>
    add_data_points_beeswarm()
```

heatmap

```
gene_expression |>
    tidyplot(x = sample, y = external_gene_name, color = expression) |>
    add_heatmap(scale = "row") |>
    adjust_size(height = 100)
```

bar/dash+point + statistical comparison

```
study |>
  tidyplot(x = group, y = score, color = dose) |>
  add_mean_bar(alpha = 0.3) |>
  add_sem_errorbar() |>
  add_data_points() |>
  add_test_asterisks(hide_info = TRUE)
```



```
study |>
  tidyplot(x = dose, y = score, color = group) |>
  add_mean_dash() |>
  add_sem_errorbar() |>
  add_data_points() |>
  add_test_pvalue()
```
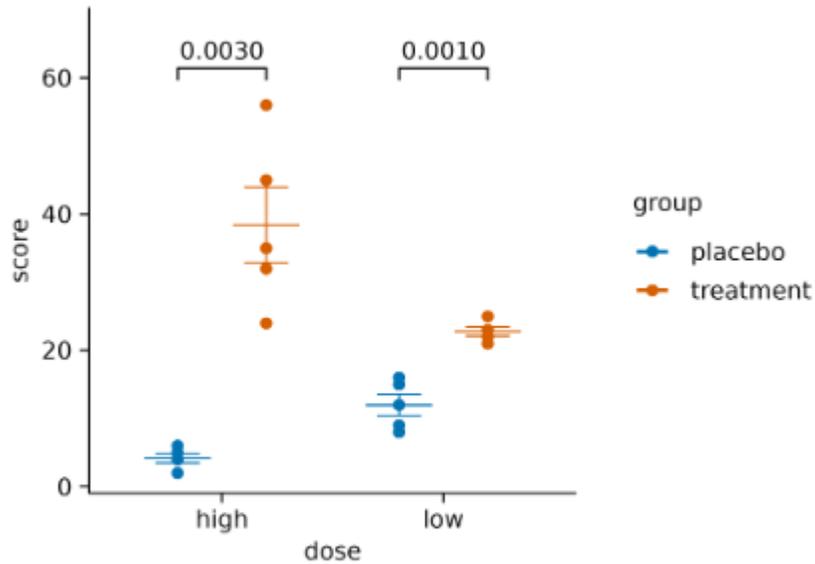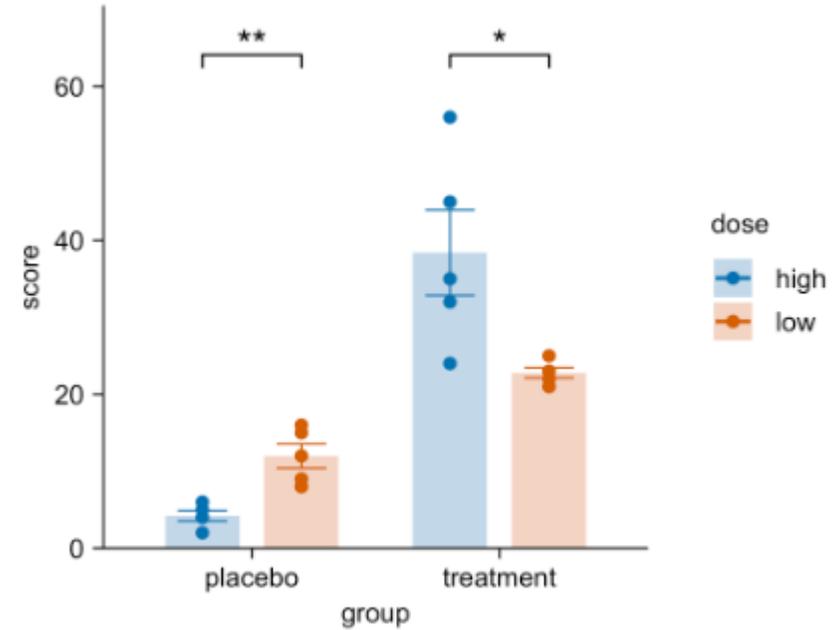

method = t_test
label = {format_p_value(p.adj, 0.0001)}
p.adjust.method = none
hide.ns = FALSE
ref.group = NULL

barplot in action

```r
library(tidyverse)
library(tidyplots)

# Baca data CSV
data <- read_csv("scopus-eeg-review-2017-sd-now.csv", show_col_types = FALSE)

# Ringkas data: hitung jumlah & rata-rata tahun
summary_data <- data %>%
  group_by(`Source title`) %>%
  summarise(
    count = n(),
    avg_year = mean(Year, na.rm = TRUE)
  ) %>%
  slice_max(order_by = count, n = 9)  # ambil 9 teratas

# Visualisasi pakai tidyplots
summary_data |>
  tidyplot(x = count, y = `Source title`, color = avg_year) |>
  add_mean_bar(width = 0.6) |>
  sort_y_axis_labels(.reverse = FALSE) |>
  adjust_colors(new_colors = colors_continuous_viridis) |>
  add_sum_value(fontsize = 14, accuracy = 1) |>
  adjust_x_axis(title = "Number of Docs", limits = c(0, 15)) |>
  remove_y_axis_title() |>
  adjust_font(fontsize = 14) |>
  adjust_legend_title("Avg Year") |>
  adjust_size(width = NA, height = 100)
```
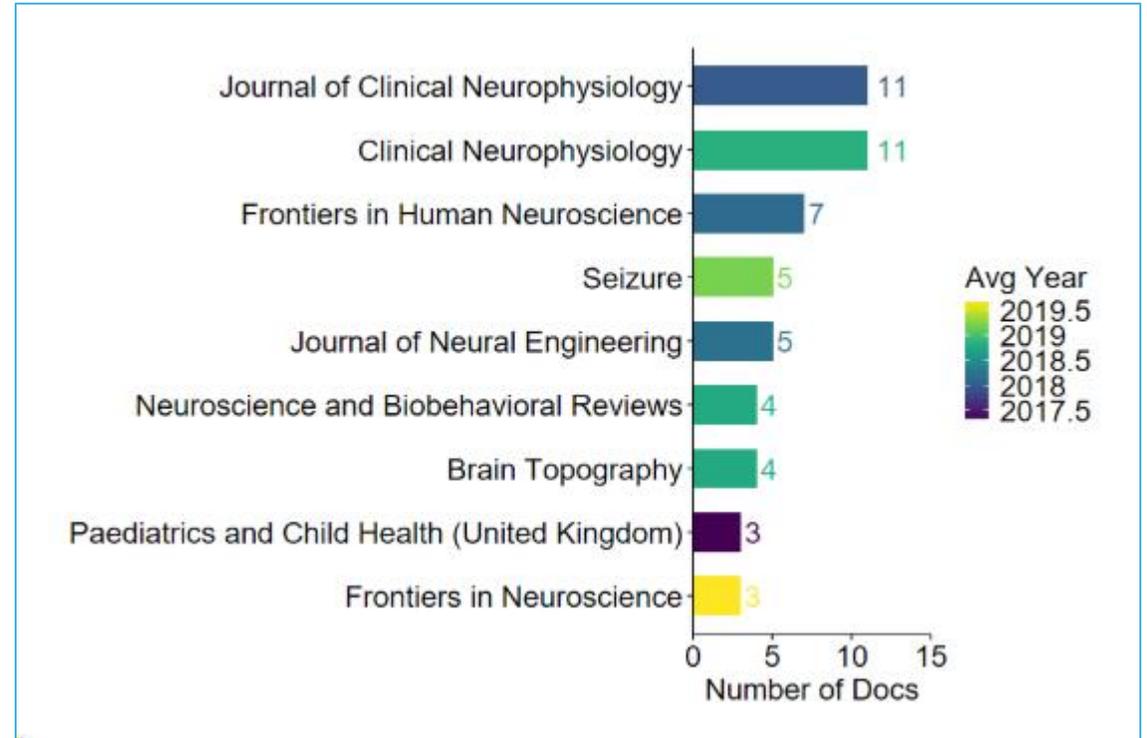
Reference

Akbulut, Ö. (2023). Inferential statistics suggested to report in scientific articles. *Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi*. https://doi.org/10.18016/ksutarimdoga.vi.1282850

Alcazer, V. (2020). StatAid: An R package with a graphical user interface for data analysis. *The Journal of Open Source Software, 5*(54), 2630. https://doi.org/10.21105/JOSS.02630

Andrade, C. (2023). Survival analysis, Kaplan-Meier curves, and Cox regression: Basic concepts. *Indian Journal of Psychological Medicine, 45*, 434–435. https://doi.org/10.1177/02537176231176986

Bennett, M. J., & Hugen, D. L. (2016). The R language for statistical computing (pp. 7–22). Cambridge University Press. https://doi.org/10.1017/CBO9781316584460.003

Berker, Y., Muti, I. H., & Cheng, L. L. (2022). Visualizing metabolomics data with R. *NMR in Biomedicine, 36*(4). https://doi.org/10.1002/nbm.4865

Chaves, A. J., Martín, C., & Diaz, M. A. (2023). The orchestration of machine learning frameworks with data streams and GPU acceleration in Kafka-ML: A deep-learning performance comparative. *Expert Systems*. https://doi.org/10.1111/exsy.13287

Chen, Y. T., Williamson, B. D., Okonek, T., Wolock, C. J., Spieker, A. J., Hee Wai, T. Y., Hughes, J. P., Emerson, S. S., & Willis, A. D. (2022). rigr: Regression, inference, and general data analysis tools in R. *The Journal of Open Source Software, 7*(80), 4847. https://doi.org/10.21105/joss.04847

Chevalier, R. (2023). GlobeCorr: Interactive globe-based visualization for correlation datasets, *3*(1). https://doi.org/10.1093/bioadv/vbac099

Core, R. (2014). R: A language and environment for statistical computing. *MSOR Connections, 1*(1).

Das, B. K., Jha, D. N., Sahu, S. K., Yadav, A. K., Raman, R. K., & Kartikeyan, M. (2023). Introduction to R software. In *Concept building in fisheries data analysis*. Springer. https://doi.org/10.1007/978-981-19-4411-6_12

Denton, W. (2011). Getting started with R. https://macsphere.mcmaster.ca/handle/11375/14377

Engler, J. B. (2025). Tidyplots empowers life scientists with easy code-based data visualization. *iMeta*, *e70018*. https://doi.org/10.1002/imt2.70018

Ergin, M., & Koşkan, Ö. (2023). Comparison of Student-t, Welch's t, and Mann–Whitney U tests in terms of Type I error rate and test power. *Selcuk Journal of Agriculture and Food Sciences*. https://doi.org/10.15316/sjafs.2023.022

Erlichman, C. (2023). Use of appropriate statistical tools in biomedical research. *Indian Journal of Medical Research, Publish Ahead of Print*. https://doi.org/10.4103/ijmr.ijmr_809_20

Fusfeld, Z. H. (2022). Statistical inference via data science: A modern dive into R and the Tidyverse. *Biometrics, 78*(1), 410–412. https://doi.org/10.1111/biom.13620

Gerela, P. S., Mishra, P., & Vipat, R. (2022). Study on data visualization. *International Journal of Health Sciences*, 6(3), 6298–6305. https://doi.org/10.53730/ijhs.v6ns3.7393

Giorgi, F. M., Ceraolo, C., & Mercatelli, D. (2022). The R language: An engine for bioinformatics and data science. *Reproductive and Developmental Biology, 12*(5), 648. https://doi.org/10.3390/life12050648

Goar, V., Yadav, N., & Yadav, P. S. (2023). Conversational AI for natural language processing: An review of ChatGPT. *International Journal on Recent and Innovation Trends in Computing and Communication, 11*(3s), 109–117. https://doi.org/10.17762/ijritcc.v11i3s.6161

Golafshar, M. A., Petersen, M., Vargas, C., Samadder, N. J., Kunze, K. L., McCormick, N., Watkin, S. A., Maleyeva, D., Cheng, T. W., Vargas, M., & DeWees, T. A. (2021). Utilizing open-source platforms to build and deploy interactive patient-reported quality of life tracking tools. *Quality of Life Research, 30*(11), 3189–3197. https://doi.org/10.1007/S11136-020-02617-Z

Guzura, T. (2023). Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.14108

Hadi, F. (2022). Introduction to R-statistics (pp. 263–277). CRC Press eBooks. https://doi.org/10.1201/9781003364689-29

Hariri, W. (2023). Unlocking the potential of ChatGPT: A comprehensive exploration of its applications, advantages, limitations, and future directions. *arXiv*. https://doi.org/10.48550/arXiv.2304.02017

Heumos, L., Ehmele, P., Kuhn Cuellar, L., Menden, K., Miller, E., Lemke, S., ... & Nahnsen, S. (2023). mlf-core: A framework for deterministic machine learning. *Bioinformatics, 39*(4). https://doi.org/10.1093/bioinformatics/btad164

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, 31*(3), 685–695. https://doi.org/10.1007/S12525-021-00475-2

Johnson, P. E. (2020). R Markdown: The definitive guide. *The American Statistician, 74*(2), 209–210. https://doi.org/10.1080/00031305.2020.1745577

Jones, L. A., Gearty, W., Allen, B., Eichenseer, K., Dean, C., Galván, S., ... & Chiarenza, A. A. (2023). palaeoverse: A community-driven R package to support palaeobiological analysis. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.14099

Khojasteh, J. (2022). Review of *An Introduction to Multivariate Statistics R SPSS SAS*. *Multivariate Behavioral Research, 30*(2), 336–338. https://doi.org/10.1080/10705511.2022.2143782

Konkol, M., Nüst, D., & Goulier, L. (2020). Publishing computational research—a review of infrastructures for reproducible and transparent scholarly communication, *5*(1), 1–8. https://doi.org/10.1186/S41073-020-00095-Y

Li, H., & Bezemer, C.-P. (2022). Studying popular open source machine learning libraries and their cross-ecosystem bindings. *arXiv*. https://arxiv.org/abs/2201.07201

Lüdecke, D., Patil, I., Ben-Shachar, M. S., Wiernik, B. M., Waggoner, P. D., & Makowski, D. (2021). An R package for visualizing statistical models. *The Journal of Open Source Software, 6*(64), 3393. https://doi.org/10.21105/JOSS.03393

McDonald, S., & Levine, D. (2022). Commonly used statistics in occupational therapy literature. *American Journal of Occupational Therapy, 76*(Suppl. 1), 7610505093p1.

Mwambi, H. (2022). Statistical methods for global health and epidemiology: Principles, methods and applications. *Biometrical Journal*. https://doi.org/10.1002/bimj.202200006

Nakagawa, S., Lagisz, M., O'Dea, R. E., Pottier, P., Senior, A. M., Yang, Y., & Noble, D. W. A. (2023). orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.14152

Nguyen, V. T., Jung, K., & Gupta, V. (2021). Examining data visualization pitfalls in scientific publications. *Visual Computing for Industry, Biomedicine, and Art, 4*(1), 27. https://doi.org/10.1186/S42492-021-00092-Y

Nordmann, E. (2022). Data visualization using R for researchers who do not use R. *Advances in Methods and Practices in Psychological Science, 5*(2), 251524592210746. https://doi.org/10.1177/25152459221074654

Ponce, M., & Sandhel, A. (2021). covid19.analytics: An R package to obtain, analyze and visualize data from the 2019 coronavirus disease pandemic. *The Journal of Open Source Software, 6*(60), 2995. https://doi.org/10.21105/JOSS.02995

Prasad, P. S. S., Parvathi, O., Sri, N., Nandini, S., Priya, G., & Rahul, R. (2022). Weaponising AI for natural language processing: Novel perspectives. *International Journal of Advanced Research in Science, Communication and Technology*, 407–411. https://doi.org/10.48175/ijarsct-2465

Rahrig, R. R. (2024). ANOVA F-test and Kruskal-Wallis test performance comparison under varying distributions, variance heterogeneity, sample sizes, and noncentrality structures. *Journal of Statistics and Management Systems, 27*(8), 1569–1587. https://doi.org/10.47974/jsms-1116

Rödiger, S., Friedrichsmeier, T., Kapat, P., & Michalke, M. (2012). RKWard: A comprehensive graphical user interface and integrated development environment for statistical analysis with R. *Journal of Statistical Software, 49*(1), 1–34. https://doi.org/10.18637/JSS.V049.I09

Salmerón-Gómez, R., García-García, C., & García-Pérez, J. (2021). A guide to using the R package "multiColl" for detecting multicollinearity. *Computational Economics, 57*(2), 1–8. https://doi.org/10.1007/S10614-019-09967-Y

Serdaliyev, Y. (2023). Applications of ChatGPT: Where it can be used and what we can solve with ChatGPT. *Қ.А. Ясауи Атындағы Халықаралық Қазақ-Түрік Университетінің Хабарлары, 24*(1), 129–138. https://doi.org/10.47526/2023-1/2524-0080.12

Setzer, J. C., & Cui, Z. (2022). Communicating measurement outcomes with (better) graphics. *Educational Measurement: Issues and Practice, 41*(3), 5–13. https://doi.org/10.1111/emip.12519

Tollefson, M. (2022). Downloading R and RStudio and setting up a file system (pp. 3–14). Apress eBooks. https://doi.org/10.1007/978-1-4842-7924-3_1

Tso, C.-H. M., Hollaway, M., Killick, R., Henrys, P. A., Monteith, D., Watkins, J. D., & Blair, G. (2022). Advancing reproducible research by publishing R Markdown notebooks as interactive sandboxes using the *learnr* package. *R Journal, 14*(1), 255–263. https://doi.org/10.32614/rj-2022-021

Tuimala, J., & Kallio, A. (2013). R, programming language. https://link.springer.com/content/pdf/10.1007%2F978-1-4419-9863-7_530.pdf

Valero-Mora, P. M., & Ledesma, R. D. (2012). Graphical user interfaces for R. *Journal of Statistical Software, 49*(1), 1–8. https://doi.org/10.18637/JSS.V049.I01

Vuorre, M., & Crump, M. J. C. (2021). Sharing and organizing research products as R packages. *Behavior Research Methods, 53*(2), 792–802. https://doi.org/10.3758/S13428-020-01436-X

Wilson, J. (2012). Statistical computing with R: Selecting the right tool for the job—R Commander or something else? *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(6), 518–526. https://doi.org/10.1002/WICS.1228

Yi, X. (2023). Basic statistics and functions using R. *International Series in Management Science/Operations Research*, 7–32. https://doi.org/10.1007/978-3-031-21480-6_2

That's all, thank you...

Further discussion

Adi Wijaya | +62838-7891-9456 | adiwjj@gmail.com | adiwjj@uima.ac.id