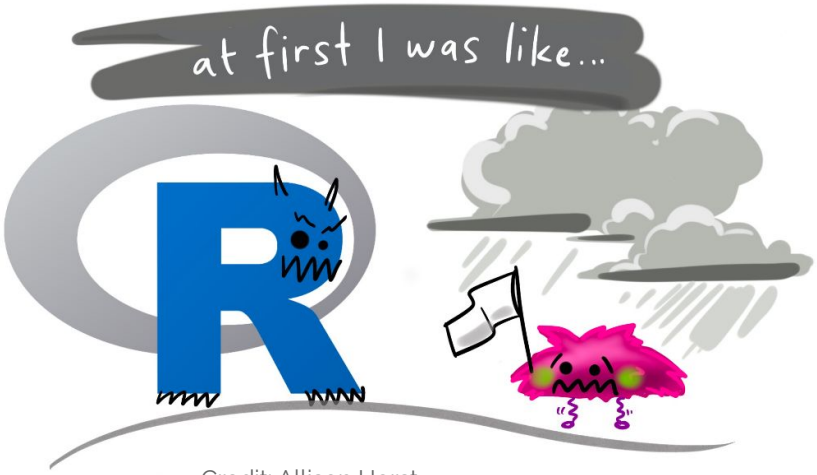
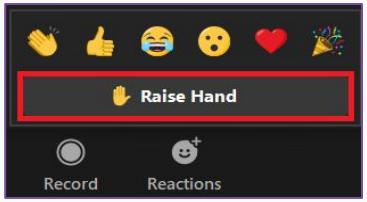
لغة الآر من الصفر إلى الإحتراف



Credit: Allison Horst

| **Twitter**:@batool664 | **Web**: https://batool-almarzouq.netlify.app/ | د.بتول المرزوق



قواعد الدورة

مرحبا في كورس "لغة الأر (R) من الصفر إلي الإحتراف"

المدربة: د. بتول المرزوق

الأسبوع: الأول

التاريخ: ٢٢ يونيو ٢٠٢٢

وصف الجلسة:

في هذه الجلسة سنتعرف على الخطة التعليمية مع بيان أهمية استخدام لغة آر (R) وإبراز دورها المميز في علم البيانات. لغة آر (R) هي لغة مفتوحة المصدر تُستخدم في التمثيل المرئي للبيانات وتعدينها، بالإضافة إلى النمذجة. وقد احتلت أعلى المراتب في تصنيفات لغات البرمجة المستخدمة في علوم البيانات على مدار عدة سنوات، مما أدى إلى تزايد استخدامها.

تسجيل الأسماء

الاسم / المدينة / حسابات وسائل التواصل الإجتماعي لمن يرغب (twitter و GitHub وما إلى ذلك)

-
-
-
-
-
-
-
-

- كتابة الاسم في Google Doc
  - الإجابة على Icebreaker question
- إطفاء الجوال
- يُفضل فتح الكاميرا
- لاتتردد في طرح **أي سؤال** خلال الجلسة
- **Slack**

@OpenSciSaudi, @batool664

www.osc-ksa.com

batool@liverpool.ac.uk

خطة الدورة التدريبية

الأسبوع	نوع المحاضرة	اليوم	التاريخ	الساعة (بتوقيت السعودية)	المدة بالساعات
الرابع	محاضرة مباشرة	الخميس	28 يوليو 2022	من ١:٣٠ إلى ٣:٣٠ مساءً	1
الخامس	محاضرة مباشرة	الثلاثاء	2 أغسطس 2022	من ٣:٣٠ إلى ٥:٣٠ مساءً	2
	محاضرة مباشرة	الأربعاء	3 أغسطس 2022	من ٢:٣٠ إلى ٤:٣٠ مساءً	2
	ساعة مكتبية	الثلاثاء	2 أغسطس 2022	من ٢:٣٠ إلى ٣:٣٠ مساءً	1
السادس	محاضرة مباشرة	الثلاثاء	9 أغسطس 2022	من ٣:٣٠ إلى ٥:٣٠ مساءً	2
	محاضرة مباشرة	الأربعاء	10 أغسطس 2022	من ٢:٣٠ إلى ٤:٣٠ مساءً	2
	ساعة مكتبية	الأربعاء	10 أغسطس 2022	من ٤:٣٠ إلى ٥:٣٠ مساءً	1
السابع	عرض المشروعات النهائية	الأربعاء	17 أغسطس 2022	من ٢:٣٠ إلى ٤:٣٠ مساءً	2
	عرض المشروعات النهائية	الخميس	18 أغسطس 2022	من ١٢:٣٠ إلى ٢:٣٠ مساءً	2

Catherine Kim, PhD
@fishiintheC

Afternoon recapping #git and #github with #RStudio with the legendary @JennyBryan.

Nice to hear from the pros that experiencing agony is normal 😅

And hex stickers make an appearance! #RStudioConf2022

Jenny Bryan standing in front of a slide on comparing workflows without and with using version control with git.

Slide from Jenny Bryan's git a d github with R presentation about the agony of git with the program's github, gitkracken, and rstudio that promote flow.

Hexagon 'hex' stickers for the dev tools, testthat, rlang, usethis, and fs r packages.

5:13 AM · Jul 26, 2022 · Twitter for Android

Jess Butler
@JessButler284

Slides and code from this year's workshops at #RStudioConf2022

Causal Inference
Teaching Data Science
Getting Started with Quarto
Production-Quality Shiny Apps
Package Development Masterclass
Data Science for SysAdmins
Art from Code

github.com
rstudio::conf(2022)
Workshops at rstudio::conf 2022. rstudio::conf(2022) has 21 repositories available. Follow their code on GitHub.

10:09 AM · Jul 27, 2022 · Twitter Web App

92 Retweets **5** Quote Tweets **376** Likes

Link

مراجعة

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk


tidyverse
readr
ggplot2
forcats
tidyr
tibble
dplyr
stringr
purrr

كيفية قراءة البيانات في R

```
read_csv("path")
```

In relative to the working directory using `setwd()`

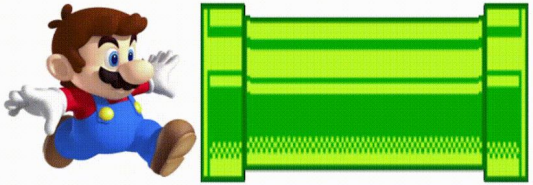
```
read_csv(here("folder","file"))
```

```
read_csv(here("data","data.csv"))
```

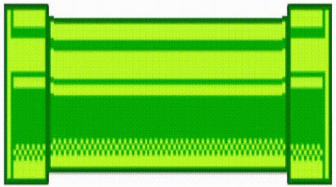
معنى هذا الرمز في لغة R

%>% = (و ثُم)

```
dataset %>%
select(column)
```



Mario %>%



%>%

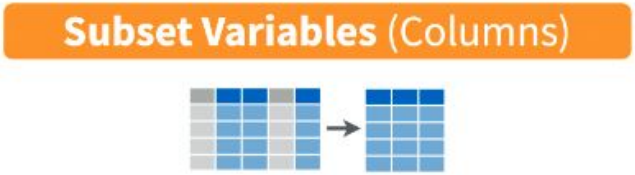
```
filter(data, row == "Mexico")
```

```
filter(data, row == "Mexico")
```

```
Data %>%
filter(row == "Mexico")
```

- `filter()` : pick observations by their values

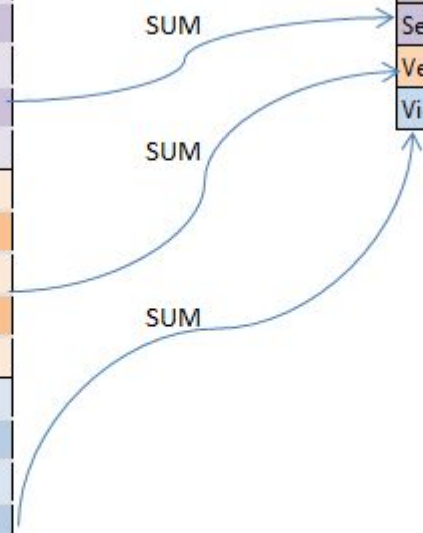
- `select()` : pick variables by their names



- `mutate()`: create new variables with functions of existing variables

- `summarise()`: collapse many values down to a single summary

sepal.length	sepal.width	variety
5.1	3.5	Setosa
4.9	3	Setosa
4.7	3.2	Setosa
4.6	3.1	Setosa
5	3.6	Setosa
7	3.2	Versicolor
6.4	3.2	Versicolor
6.9	3.1	Versicolor
5.5	2.3	Versicolor
6.5	2.8	Versicolor
6.3	3.3	Virginica
5.8	2.7	Virginica
7.1	3	Virginica
6.3	2.9	Virginica
6.5	3	Virginica
7.6	3	Virginica
4.9	2.5	Virginica



variety	sepal.length	sepal.width
Setosa	24.3	16.4
Versicolor	32.3	14.6
Virginica	44.5	20.4

Data transformation with dplyr : : CHEAT SHEET

dplyr functions work with pipes and expect **tidy data**. In tidy data:

Each **variable** is in its own **column**

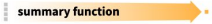
Each **observation**, or **case**, is in its own **row**

pipes

x %>% f(y) becomes **f(x, y)**

Summarise Cases

Apply **summary functions** to columns to create a new table of summary statistics. Summary functions take vectors as input and return one value (see back).



summarise(.data, …**)**
Compute table of summaries.
summarise(mtcars, avg = mean(mpg))

count(.data, …, wt = NULL, sort = FALSE, name = NULL**)** Count number of rows in each group defined by the variables in … Also **tally()**.
count(mtcars, cyl)

Group Cases

Use **group_by(**.data, …, .add = FALSE, .drop = TRUE**)** to create a "grouped" copy of a table grouped by columns in ... dplyr functions will manipulate each "group" separately and combine the results.

```
mtcars %>%
  group_by(cyl) %>%
  summarise(avg = mean(mpg))
```

Use **rowwise(**.data, …**)** to group data into individual rows. dplyr functions will compute results for each row. Also apply functions to list-columns. See tidyr cheat sheet for list-column workflow.

```
starwars %>%
  rowwise() %>%
  mutate(film_count = length(films))
```

ungroup(x, …**)** Returns ungrouped copy of table.
ungroup(g_mtcars)

Manipulate Cases

EXTRACT CASES

Row functions return a subset of rows as a new table.

filter(.data, …, .preserve = FALSE**)** Extract rows that meet logical criteria.
filter(mtcars, mpg > 20)

distinct(.data, …, .keep_all = FALSE**)** Remove rows with duplicate values.
distinct(mtcars, gear)

slice(.data, …, .preserve = FALSE**)** Select rows by position.
slice(mtcars, 10:15)

slice_sample(.data, …, n, prop, weight_by = NULL, replace = FALSE**)** Randomly select rows. Use n to select a number of rows and prop to select a fraction of rows.
slice_sample(mtcars, n = 5, replace = TRUE)

slice_min(.data, order_by, …, n, prop, with_ties = TRUE**)** and **slice_max()** Select rows with the lowest and highest values.
slice_min(mtcars, mpg, prop = 0.25)

slice_head(.data, …, n, prop**)** and **slice_tail()** Select the first or last rows.
slice_head(mtcars, n = 5)

Logical and boolean operators to use with filter()

==	<	<=	is.na()	%in%	|	xor()
!=	>	>=	!is.na()	!	&	

See **?base::Logic** and **?Comparison** for help.

ARRANGE CASES

arrange(.data, …, .by_group = FALSE**)** Order rows by values of a column or columns (low to high), use with **desc()** to order from high to low.
arrange(mtcars, mpg)
arrange(mtcars, desc(mpg))

ADD CASES

add_row(.data, …, .before = NULL, .after = NULL**)** Add one or more rows to a table.
add_row(cars, speed = 1, dist = 1)

Manipulate Variables

EXTRACT VARIABLES

Column functions return a set of columns as a new vector or table.

pull(.data, var = -1, name = NULL, …**)** Extract column values as a vector, by name or index.
pull(mtcars, wt)

select(.data, …**)** Extract columns as a table.
select(mtcars, mpg, wt)

relocate(.data, …, .before = NULL, .after = NULL**)** Move columns to new position.
relocate(mtcars, mpg, cyl, .after = last_col())

Use these helpers with select() and across()
e.g. select(mtcars, mpg:cyl)

contains(match)	**num_range**(prefix, range)	**:**, e.g. mpg:cyl
ends_with(match)	**all_of**(x)/**any_of**(x, …, vars)	**-**, e.g. -gear
starts_with(match)	**matches**(match)	**everything()**

MANIPULATE MULTIPLE VARIABLES AT ONCE

across(.cols, .funs, …, .names = NULL**)** Summarise or mutate multiple columns in the same way.
summarise(mtcars, across(everything(), mean))

c_across(.cols**)** Compute across columns in row-wise data.
transmute(rowwise(UKgas), total = sum(c_across(1:2)))

MAKE NEW VARIABLES

Apply **vectorized functions** to columns. Vectorized functions take vectors as input and return vectors of the same length as output (see back).

mutate(.data, …, .keep = "all", .before = NULL, .after = NULL**)** Compute new column(s). Also **add_column(), add_count(),** and **add_tally().**
mutate(mtcars, gpm = 1 / mpg)

transmute(.data, …**)** Compute new column(s), drop others.
transmute(mtcars, gpm = 1 / mpg)

rename(.data, …**)** Rename columns. Use **rename_with()** to rename with a function.
rename(cars, distance = dist)

RStudio

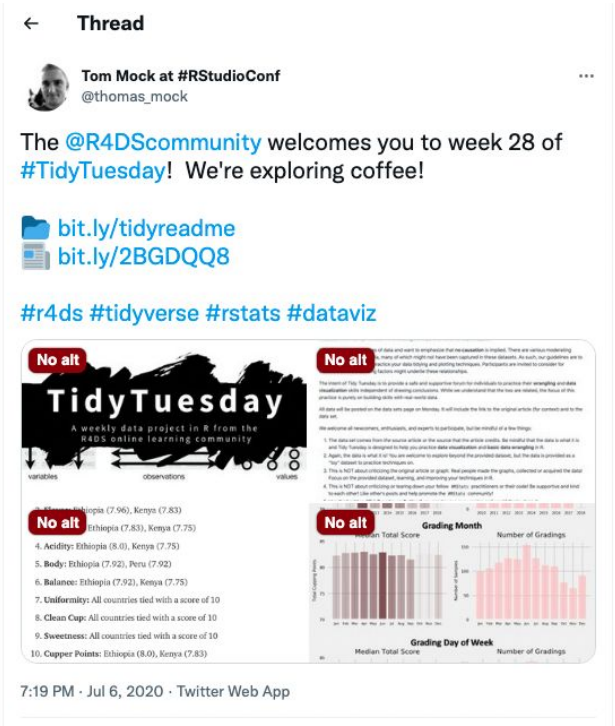
@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

أمثلة من TidyTuesday

TidyTuesday هو مشروع بيانات أسبوعي يستهدف مُجتمع R الدولي وهو مدعوم من R4DS Online Learning Community و R for Data Science ، ويركز على فهم كيفية تلخيص البيانات وترتيبها باستخدام ggplot2 و tidyr و dplyr وغيرها من الأدوات في نظام لغة الآر.

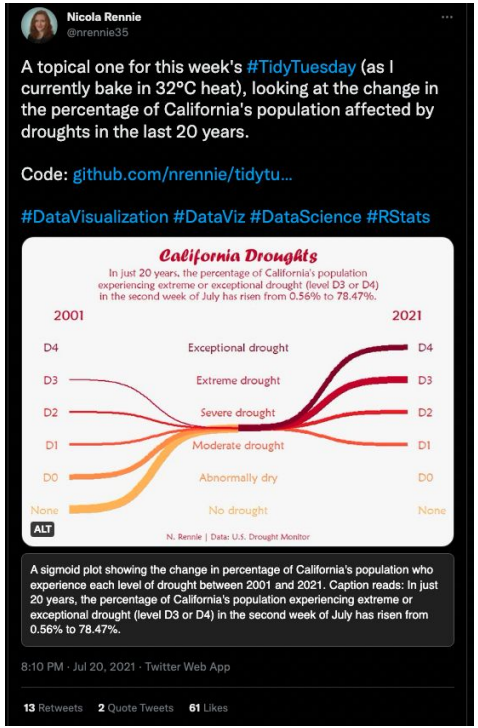
@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

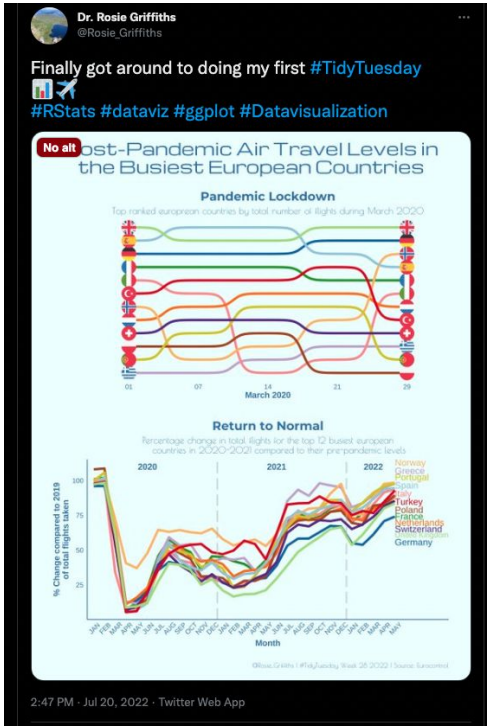
أمثلة من TidyTuesday



@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

أمثلة من TidyTuesday



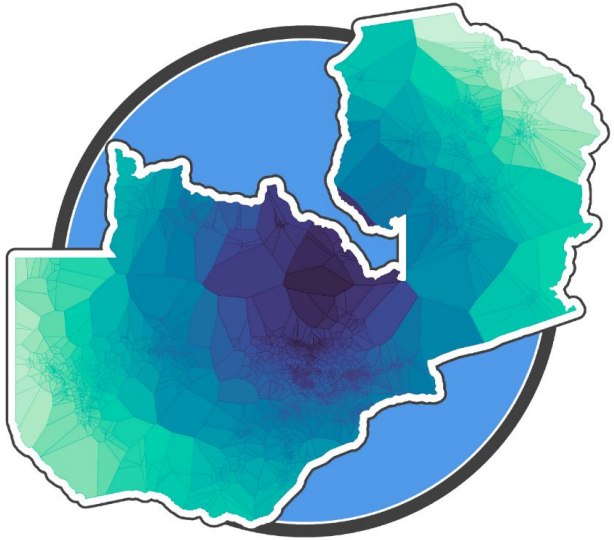


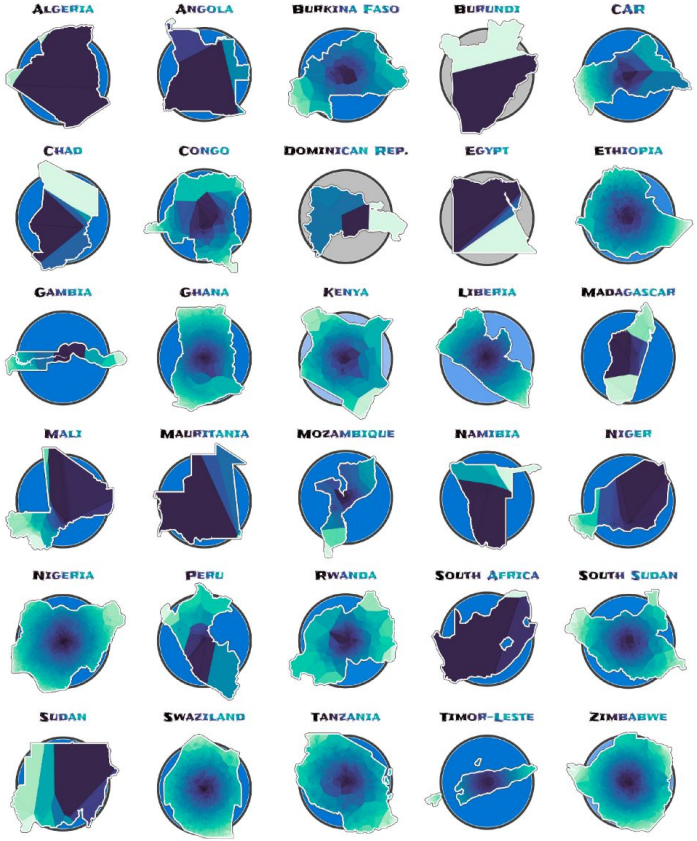
@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

† I extracted all functions starting with *geom* or *stat* from my Rmd files containing the code for all my #TidyTuesday contributions (thanks Georgios for the idea and script). For the contributions from 2019 (n = 26) and 2020 (n = 40) I calculated the frequency of usage per year for each geom|stat as times used divided by the number of contributions. Note that some geom's which usually appear together (e.g. *treemapify::geomtreemap* functions) or behave very similarly (e.g. *ggforce::geommark* functions) were grouped together.

@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

ZAMBIA

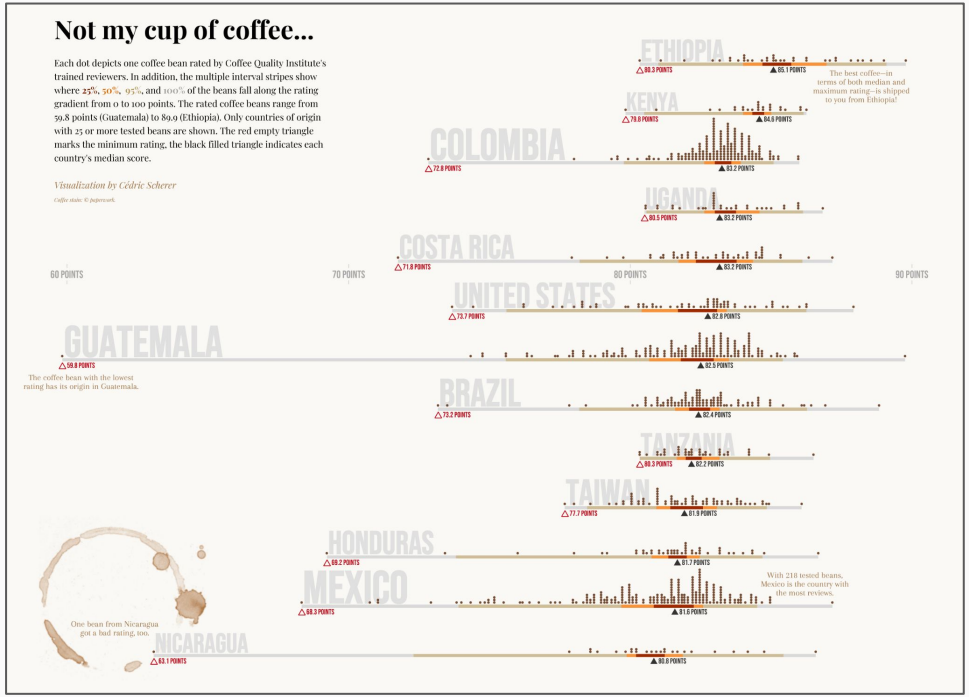


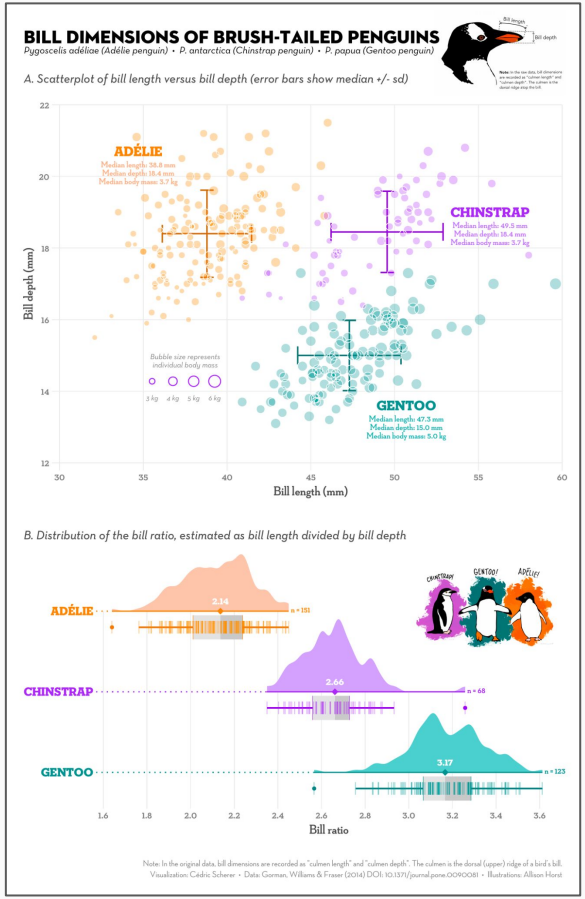


The voronoi map shows water points monitored over time by governments and development partners. The data is collected by the **Water Point Data Exchange (WPDx)** which aims to unlock the potential of water point data to improve decision-making and increase rural water access. Voronoi maps are based on the minimal distance needed to reach a landmark—*here reported water points*—by using tessellation techniques that partition a plane into regions closest to these points. The darker the color of the region, the closer it is to the country's middle point. The coloring of the circle indicates the proportion of improved versus unimproved water sources: the more intense the blue, the more improved water points; grey means no information).

Visualization: Cédric Scherer • Data: Water Point Data Exchange (WPDx), May 2021

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk





@OpenSciSaudi, @batool664
www.osc-ksa.com
batool@liverpool.ac.uk

كيف أرسم البيانات في لغة R باستخدام ggplot2 ؟

1. تعيين البيانات (data)
2. ربط المتغيرات بالصفات الجمالية أو الإحداثيات aes()
3. تحديد نوع geom

What are the building blocks in ggplot2?

1. A mapping of variables in data to
2. aes() thetic attributes of
3. geom_etric objects.

مثال

```
ggplot(data = mydata,
aes(x =column1,
y=column2) +
geom_point())
```



@OpenSciSaudi, @batool664


www.osc-ksa.com


batool@liverpool.ac.uk

Theme

Coordinates

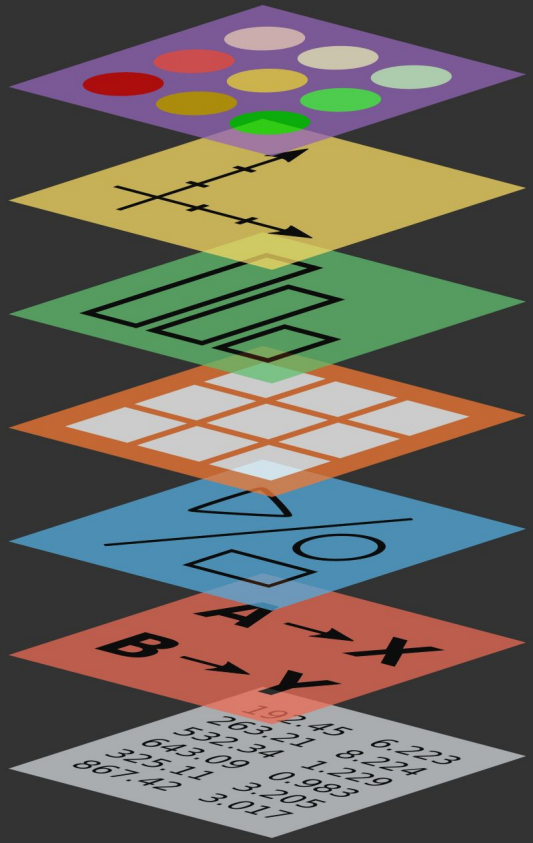
Statistics

Facets

Geometries

Aesthetics

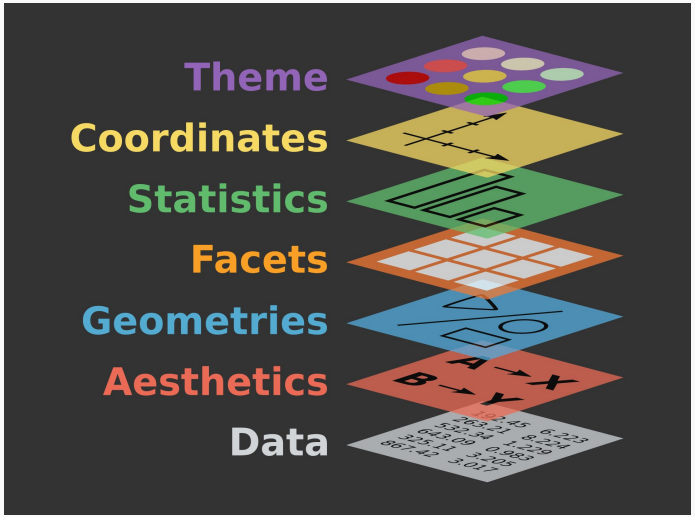
Data



@OpenSciSaudi, @batool664

www.osc-ksa.com

batool@liverpool.ac.uk




```
ggplot(data, aes(x=col1, y = col2)) +
  geom_point(alpha = 0.5) +
  geom_smooth() +
  facet_wrap(~ col) +
  coord_flip() +
  theme_bw() +
  theme(legend.position = "none")
```

1- Read data from this link
https://raw.githubusercontent.com/ScienceParkStudyGroup/r-lesson-based-on-ohi-data-training/gh-pages/data/se.csv

2- plot the data using `geom_jitter` with the following conditions:

1. x = park_name
2. y = visitors
3. color = park_name
4. alpha = 0.5
5. theme_bw
6. Change the coordinator using `coord_flip()`

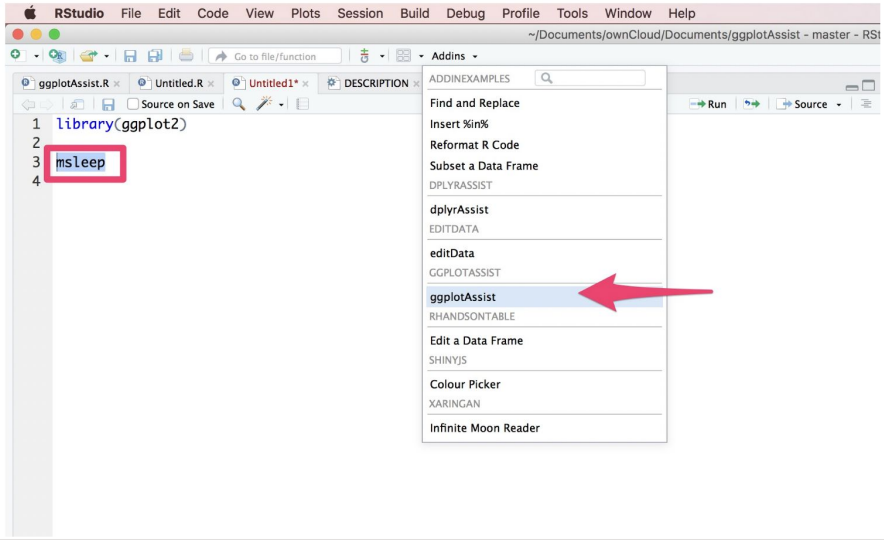
ggplotAssist package

[Link to docs in GH]

Usage: As an RStudio Add-in

This addin can be used to interactively generate a `ggplot` using `ggplot2` package. The intended way to use this is as follows:

1. Highlight a symbol naming a `data.frame` or a `tibble` in your R session, e.g. `msleep` (1). Execute this addin(arrow), to interactively manipulate it.



plotly

حفظ التمثيل البياني في الجهاز

```
ggsave("name_of_file.png", my_plot, width = 15, height = 10)
```

ggplot2 theme elements reference

Set `minimal` as the baseline theme:

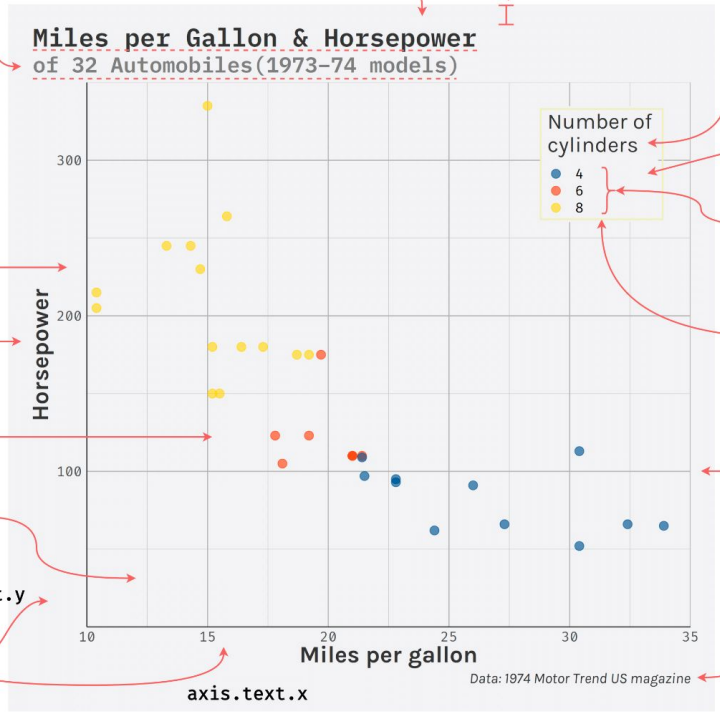
```
theme_minimal() +
theme(theme.element = element_type())
```

Use `element_blank()` to **remove an element**

Axis titles, **text**, **ticks**, and **lines** can be specified per axis using theme inheritance by putting `.x`/`.y` at the end of the theme element.

```
plot.title.position = "plot"
plot.caption.position = "plot"
```
"plot" means that they will be aligned to the entire plot (instead of the panel)

```
plot.title = element_text()
plot.subtitle = element_text()
plot.margin = margin(25, 25, 25, 25)
```




```
legend.title
= element_text()
legend.background
= element_rect()
legend.text
= element_text()
legend.position
= c(.85,.85) / "none"/
 "left"/ "right"/
 "bottom"/ "top"
plot.background
= element_rect()
plot.caption
= element_text()
axis.line.y = element_line()
axis.title.y = element_text()
panel.grid.major = element_line()
panel.grid.minor = element_line()
axis.text.y
axis.text = element_text()
axis.text.x
```

`text = element_text()` ← modifications will be applied to all text elements

Full list of elements at ggplot2.tidyverse.org/reference/theme

isabella-b

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

YAML Header in R Markdown File

```yaml
---
title: "Example"
author: "Zachary M. Smith"
date: "November 16, 2018"
output: html_document
---
```

Rendered YAML Header

Example

Zachary M. Smith

November 16, 2018

YAML Header in R Markdown File

```yaml
---
title: "Example"
author: "Zachary M. Smith"
date: "November 16, 2018"
output:
  html_document:
    toc: true
    number_sections: true
---
```

Rendered YAML Header

Example

Zachary M. Smith

November 16, 2018

- 1 R Markdown
- 2 Including Plots

Ref

YAML Header in R Markdown File

```yaml
---
title: "Example"
author: "Zachary M. Smith"
date: "November 16, 2018"
output:
  html_document:
    toc: true
    number_sections: true
    toc_float: true
---
```

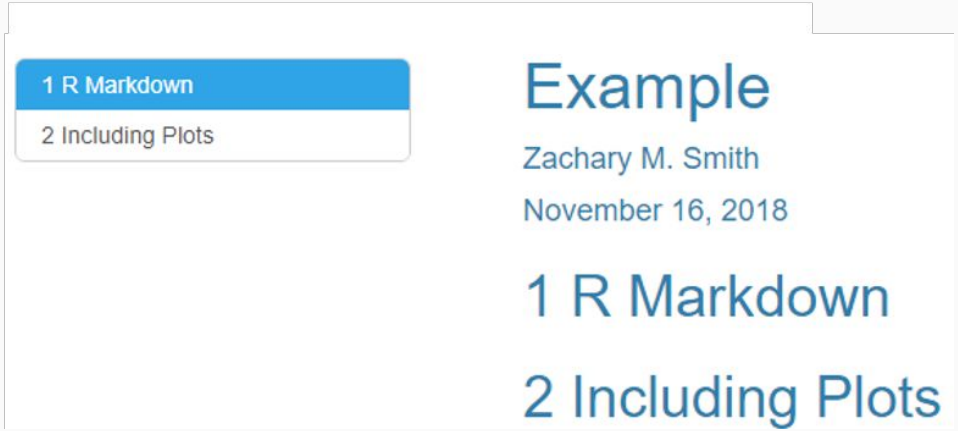
Rendered YAML Header

@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

YAML Header in R Markdown File

```yaml
---
title: "Example"
author: "Zachary M. Smith"
date: "November 16, 2018"
output:
  html_document:
    toc: true
    number_sections: true
    toc_float: true
    theme: cerulean
---
```

Rendered YAML Header



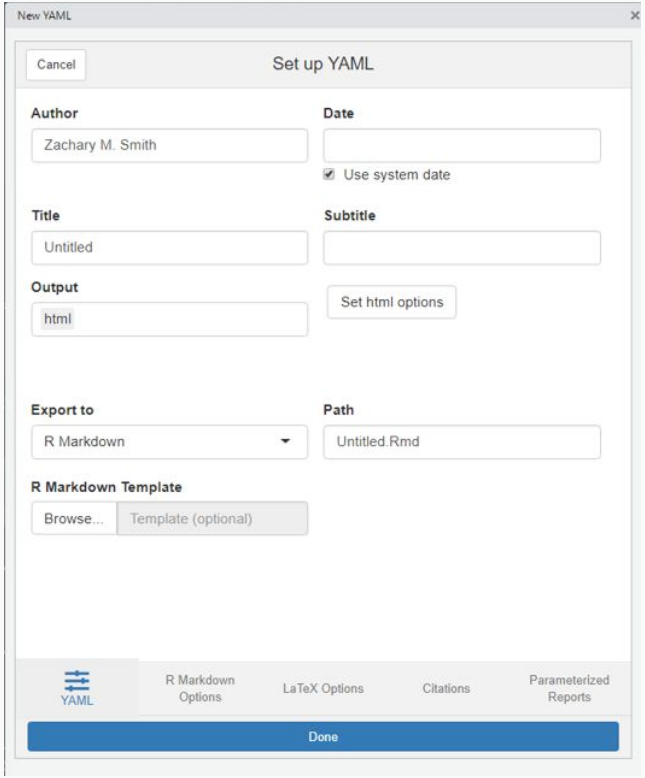
@OpenSciSaudi, @batool664 www.osc-ksa.com batool@liverpool.ac.uk

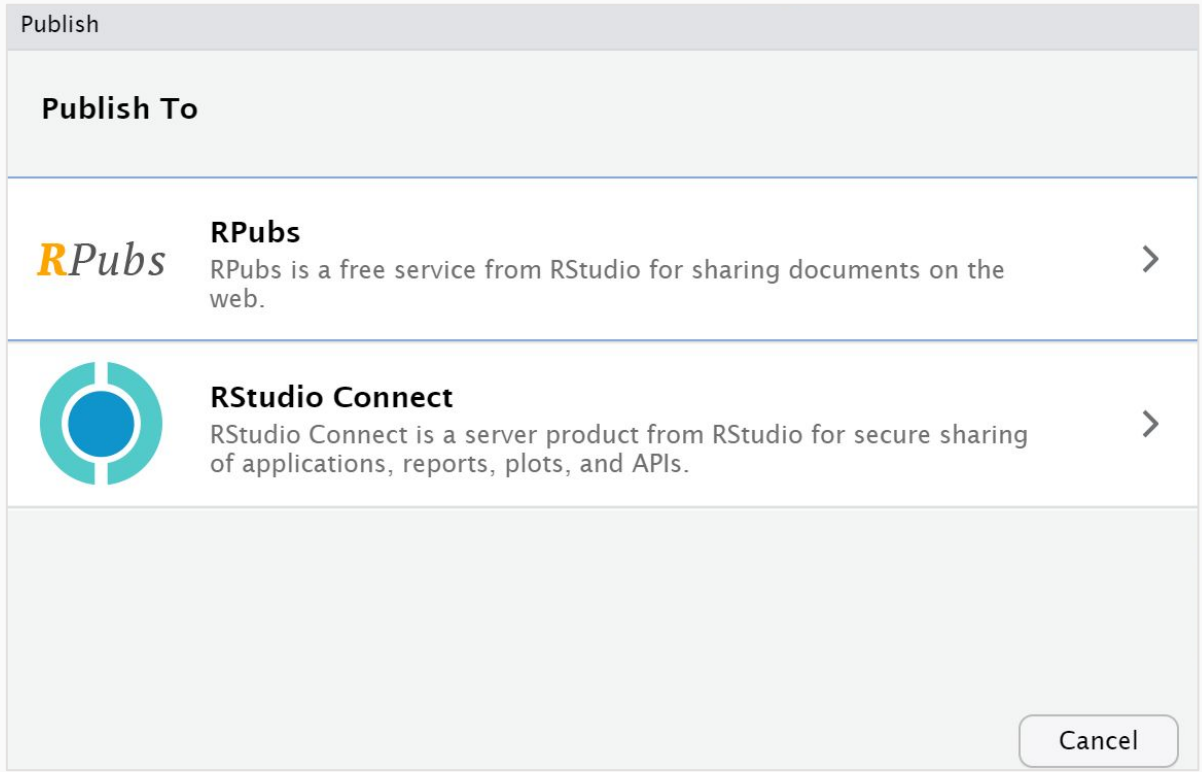
```yaml
---
title: "Example"
author: "Zachary M. Smith"
date: "`r Sys.Date()`"
output:
  word_document: default
  html_document: default
---
```

```yaml
---
title: "Example"
author: "Zachary M. Smith"
date: "`r Sys.Date()`"
output:
  word_document:
    reference_docx: word_template.docx
---
```

ymlthis is an R package intended to make it easier for you to generate YAML headers

```
ymlthis:::launch_yaml_addin().
```




Publish
Publish To
RPubs
RPubs
RPubs is a free service from RStudio for sharing documents on the web.
RStudio Connect
RStudio Connect is a server product from RStudio for secure sharing of applications, reports, plots, and APIs.
Cancel

Install **prettydoc package and change the YAML header to:**

```
output:
 prettydoc::html_pretty:
    theme: cayman
    highlight: github
    math: katex
```

Publish your website?

الواجب

- حمل حُزمة tidytuesdayR
- راجع بيانات الأسبوع ٢٨ هنا
  - https://twitter.com/thomas_mock/status/1280174453410934784
  - غرد plot جديد وسيط مبني على هذه البيانات
  - مراجعه
    - https://www.youtube.com/watch?v=-1x8Kpyndss

Feedback

- ماهو **أفضل** شيء تعلمته اليوم؟
- ماهو **أصعب** مفهوم تعرضت له في هذه الجلسة؟
- ماهو الشيء الذي تحب ان **يتحسن** في الجلسات القادمة؟