

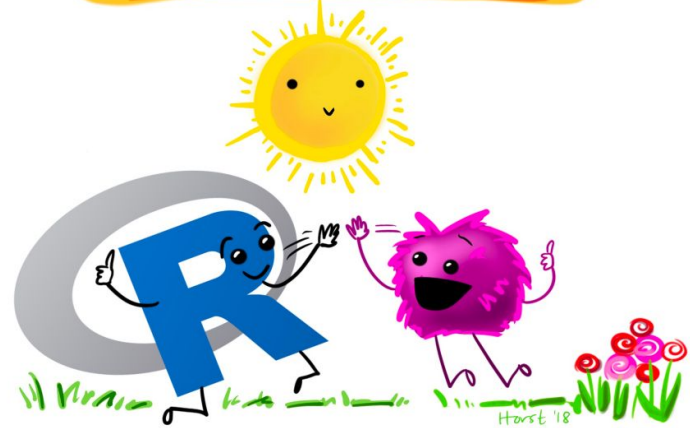
لغة الآر من الصفر إلى الإحتراف

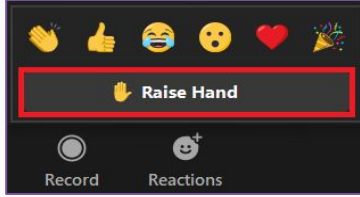
at first I was like...



Credit: Allison Horst

...but now it's like...





قواعد الدورة

● كتابة الاسم في Google Doc

○ الإجابة على Icebreaker question

● إطفاء الجوال

● يُفضل فتح الكاميرا

● لا تتردد في طرح أي سؤال خلال الجلسة

● Slack

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

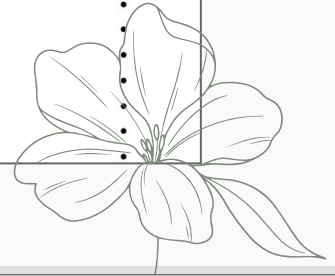
المدرية: د بتول المرزوق
الأسبوع: الأول
التاريخ: ٢٢ يونيو ٢٠٢٢

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 وتعدديتها، 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)



@OpenSciSaudi, @batool664



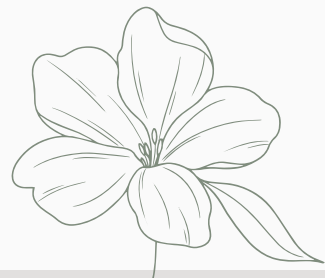
www.osc-ksa.com



batool@liverpool.ac.uk



ملاحظة مهمة



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



How to actually learn any new programming concept



Essential

Changing Stuff and
Seeing What Happens

○ RLY?

@ThePracticalDev

The internet will make those bad words go away



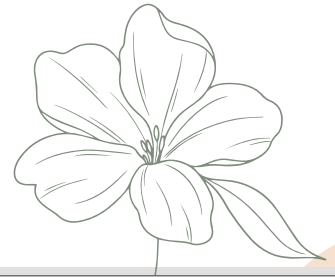
Essential

Googling the
Error Message

○ RLY?

The Practical Developer
@ThePracticalDev

Credit: Carpentries



@OpenSciSaudi, @batool664



www.osc-ksa.com



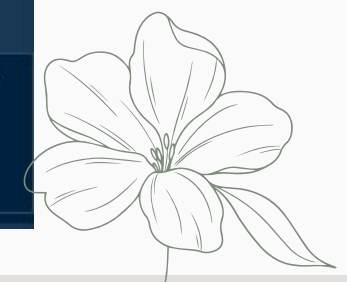
batool@liverpool.ac.uk



```
19 {r}
20 install.packages("ggplot2",dependencies=TRUE)
21 library(ggplot2)
22 install.packages("ggExtra",dependencies=TRUE)
23 library(ggExtra)
24 ...

25 {r}
26 df<-data.frame(sim.speeds,wx$WDSP)
27 p<-ggplot(df,aes(x=sort(df$sim.speeds),y=sort(df$wx.WDSP)))+geom_point(
28 )
29 ggMarginal(p+theme_gray(),type="histogram",
30 fill="steelblue",col="darkblue")
31 # Have downloaded the ggplot2 and ggExtra package but the ggplot
32 function does not seem to work/be found, do not know what the problem
33 is here.
34 ...

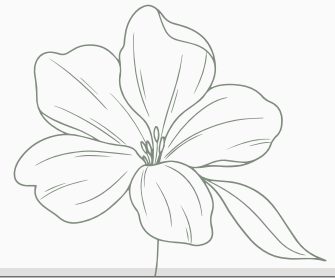
Error in ggplot(df, aes(x = sort(df$sim.speeds), y =
sort(df$wx.WDSP))) :
could not find function "ggplot"
```





```
> library(Hmisc)
```

```
Error: package or namespace load failed for 'Hmisc' in loadNamespace(j <- i[[1L]], c(lib.loc,  
.libPaths()), versionCheck = vI[[j]]):  
there is no package called 'xfun'
```



@OpenSciSaudi, @batool664

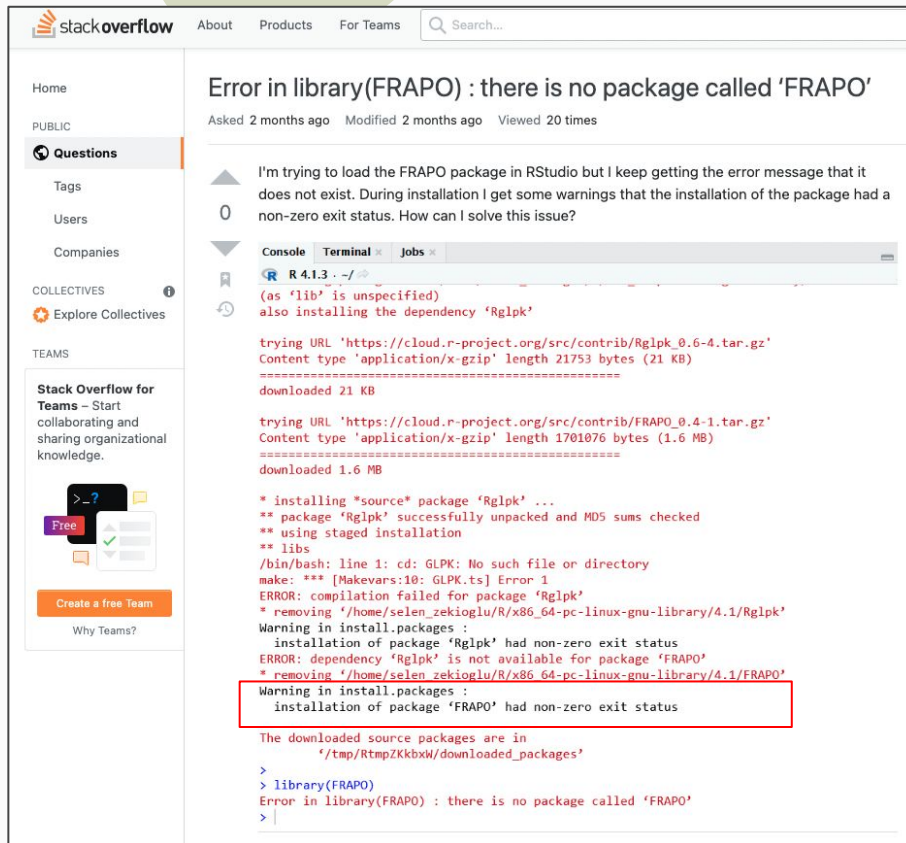


www.osc-ksa.com



batool@liverpool.ac.uk





The screenshot shows a Stack Overflow question page. The title is "Error in library(FRAPO) : there is no package called 'FRAPO'". The question was asked 2 months ago, modified 2 months ago, and viewed 20 times. The user asks for help with an RStudio error message. The console output shows the installation of the 'Rglpk' package and its dependency 'FRAPO', which fails with a non-zero exit status. The error message is highlighted with a red box.

```
R 4.1.3 ~- /~>
(as 'lib' is unspecified)
also installing the dependency 'Rglpk'

trying URL 'https://cloud.r-project.org/src/contrib/Rglpk_0.6-4.tar.gz'
Content type 'application/x-gzip' length 21753 bytes (21 KB)
=====
downloaded 21 KB

trying URL 'https://cloud.r-project.org/src/contrib/FRAPO_0.4-1.tar.gz'
Content type 'application/x-gzip' length 1701076 bytes (1.6 MB)
=====
downloaded 1.6 MB

* installing *source* package 'Rglpk' ...
** package 'Rglpk' successfully unpacked and MD5 sums checked
** using staged installation
** libs
/bin/bash: line 1: cd: GLPK: No such file or directory
make: *** [Makevars:10: GLPK.ts] Error 1
ERROR: compilation failed for package 'Rglpk'
* removing '/home/selen_zekioglu/R/x86_64-pc-linux-gnu-library/4.1/Rglpk'
Warning in install.packages :
  installation of package 'Rglpk' had non-zero exit status
ERROR: dependency 'Rglpk' is not available for package 'FRAPO'
* removing '/home/selen_zekioglu/R/x86_64-pc-linux-gnu-library/4.1/FRAPO'
Warning in install.packages :
  installation of package 'FRAPO' had non-zero exit status

The downloaded source packages are in
  '/tmp/RtmpZKkbxM/downloaded_packages'

>
> library(FRAPO)
Error in library(FRAPO) : there is no package called 'FRAPO'
> |
```



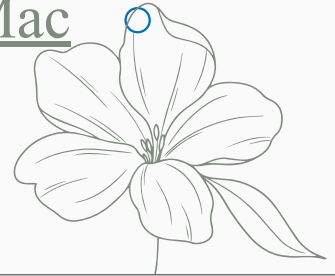
You can't push and pull in GitHub if you didn't config your machine

- تنزيل git في الجهاز (الخطوات شوية معقدة لكن نحتاج نكملها فقط مرة وحدة)


Windows ○

Add an SSH Key ■

Mac ○





 [Browse](#) [Log in](#) [Sign up](#)

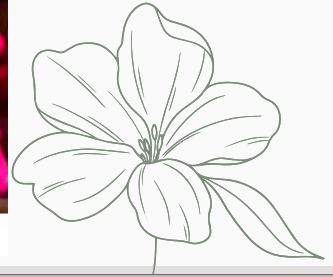
store, **share**, discover **research**

get more citations for all of the outputs of your academic research
over 80,000 citations of **figshare** content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

"figshare wants to open scientific data to the world" **WIRED**

The background figure: Comparative model of novel coronavirus 2019-nCoV... by Christian Gruber in Virology.



مستودعات رقمية (Repositories)

The screenshot shows the re3data.org website interface. At the top, there is a search bar with 'plant sciences' entered and a 'Search' button. Below the search bar, there are navigation links for '← Previous', a page indicator '1' (highlighted), and 'Next →'. A 'Sort by' dropdown menu is also visible. The main content area displays two search results:

GabiPD
GABI Primary Database

Subject(s): Plant Genetics, Plant Sciences, Biology, Life Sciences

Content type(s): Scientific and statistical data formats, Raw data, Structured graphics, Databases

Country: Germany

GABI, acronym for "Genomanalyse im biologischen System Pflanze", is the name of a large collaborative network of different plant genomic research projects. Plant data from different 'omics' fronts representing more than 10 different model or crop species are integrated in GabiPD.

PMN
Plant Metabolic Network

Subject(s): Plant Sciences, Plant Biochemistry and Biophysics, Plant Cell and Developmental Biology, Plant Genetics, Biology, Life Sciences

Content type(s): Standard office documents, Databases, Images, Scientific and statistical data formats, Raw data, Plain text, Software applications, Audiovisual data

Identify mandated / disciplinary repository:

- Institutionally specified data repository
- Domain or discipline-specific data repository

Find & compare disciplinary repositories through Repository of Research Data Repositories <https://www.re3data.org/>

Credit: Repro4Everyone



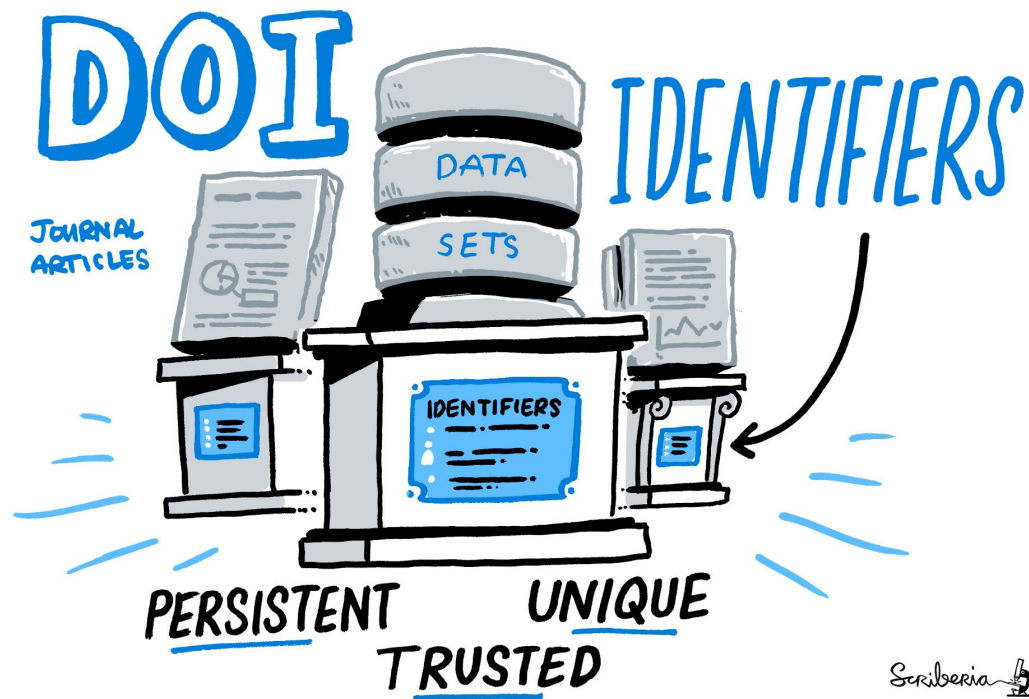
> Neuroinformatics. 2018 Jan;16(1):51-63. doi: 10.1007/s12021-017-9348-7.

A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus

Žiga Lesjak¹, Alfiia Galimzianova², Aleš Koren³, Matej Lukin³, Franjo Pernuš², Boštjan Likar⁴, Žiga Špiclin²

Affiliations + expand

PMID: 29103086 DOI: 10.1007/s12021-017-9348-7



The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



General Purpose Repositories

In addition to a specified data repository, you can make a deposit to a general purpose repository:

- DataDryad <http://datadryad.org/> (curated digital repository; free to access, \$120 to publish dataset up to 20GB)
- Figshare <https://figshare.com/> (free digital repository, 5GB per file limit)
- Zenodo <https://zenodo.org/> (free digital repository; 50GB per dataset limit)

Credit: Repro4Everyone

re3data.org



zenodo



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



Get started

1 Flip the switch

Select the repository you want to preserve, and toggle the switch below to turn on automatic preservation of your software.

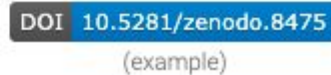


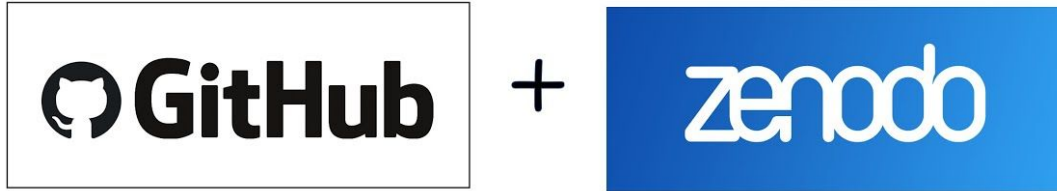
2 Create a release

Go to GitHub and [create a release](#). Zenodo will automatically download a .zip-ball of each new release and register a DOI.

3 Get the badge

After your first release, a DOI badge that you can include in your GitHub README will appear next to your repository below.





= Citable Code



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



March 17, 2022

Presentation Open Access

Practices to Improves Visibility and Outputs for Postgraduate Students in Saudi UK network!

By Batool Almarzouq


This talk was presented for the Saudi UK Network on 17-3-2022 - it can also be also accessed through this link. It introduces open science practices to postgraduate students living in the UK.

You are welcome to join Open Science Community in Saudi Arabia (OSCSA) through this link. OSCSA welcome everyone regardless of their knowledge on Open Science

The Saudi UK alumni network is a growing network of people who have studied or are still studying at UK universities and higher education institutions. It aims to provide members with a forum for networking and sharing experiences.

Preview

Page: 1 of 66 Automatic ZoomF



Practices to Improves Visibility and Outputs for Postgraduate Students!

By Batool Almarzouq (She/her)
Computational Biologist @batool664

OSCSA
Open Science Community Saudi Arabia

Files (10.6 MB)

Edit

New version

Communities

Open Science Community Saudi Arabia [Remove](#)

314

views

133

downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

March 17, 2022

DOI:

[10.5281/zenodo.6366449](https://doi.org/10.5281/zenodo.6366449)

Keyword(s):

open science CC-BY open source communities
vision 2030 DOI zenodo

Communities:

Open Science Community Saudi Arabia

License (for files):

[Creative Commons Attribution 4.0 International](#)

Share



Cite as

Batool Almarzouq, (2022, March 11). Reflection on Open Science practices and Research Software in the Kingdom in alignment with Saudi Arabia's Vision 2030. Zenodo.
<https://doi.org/10.5281/zenodo.6345895>

Start typing a citation style...



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



مراجعة



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



Tidy Data

“**TIDY DATA** is a standard way of mapping the meaning of a dataset to its structure.”

—HADLEY WICKHAM

كل متغير هو عمود

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

كل عينة هي صف



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



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"))
```



كيفية قراءة البيانات في R



Credit: Allison Horst



@OpenSciSaudi, @batool664

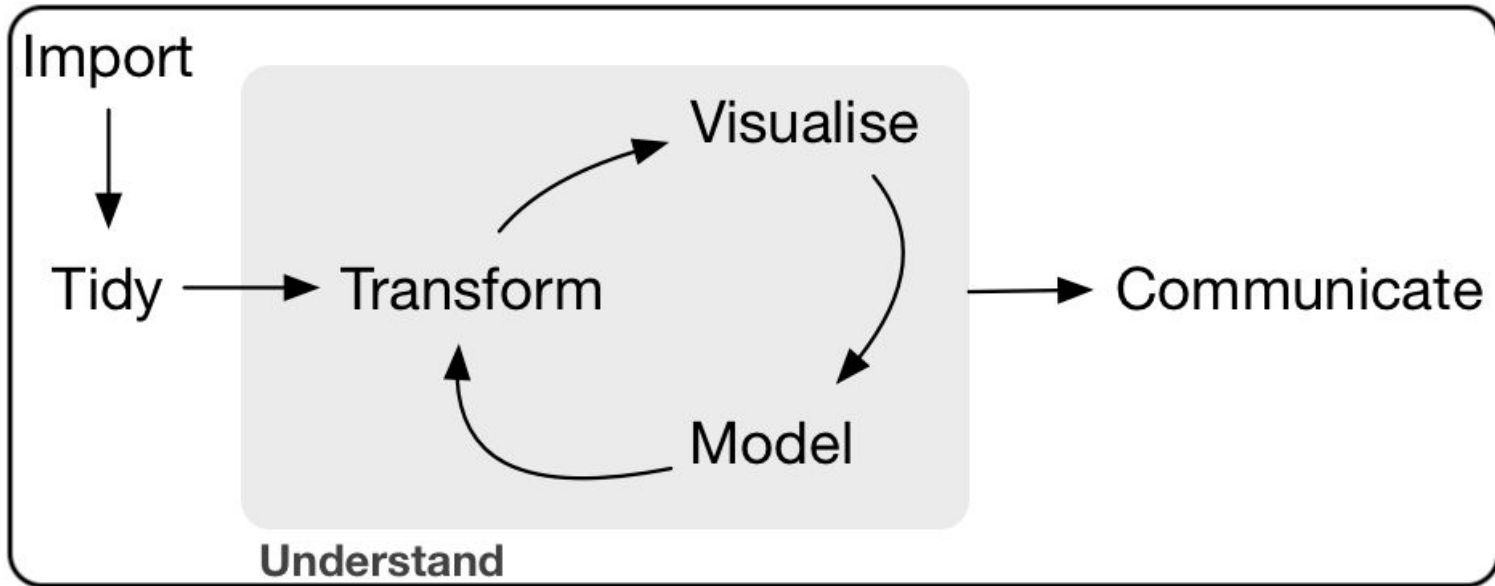


www.osc-ksa.com



batool@liverpool.ac.uk





DATAFRAME

\$ OPERATOR **COLUMN TO
SELECT**

`dataf$NewData`

```
[1] "A" "A" "A" "A" "A" "A" "A" "A" "A" "A"
```



معنى هذا الرمز في لغة R

(و تُم) = % > %

معنى هذا الرمز في لغة R

(و تُم) = %>%

dataset %>%

select(**column**)

Operators



Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x y	x OR y
x & y	x AND y
isTRUE(x)	test if X is TRUE





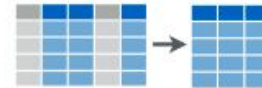
- `filter()` : pick observations by their values

Subset Observations (Rows)



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

Subset Variables (Columns)



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

SUM

SUM

SUM

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





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

Make New Variables



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

Summarise Data



Combine Data Sets

a			b		
x1	x2		x1	x3	
A	1	+	A	T	=
B	2		B	F	
C	3		D	T	

Mutating Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NA

dplyr::left_join(a, b, by = "x1")

Join matching rows from b to a.

x1	x3	x2
A	T	1
B	F	2
D	T	NA

dplyr::right_join(a, b, by = "x1")

Join matching rows from a to b.

x1	x2	x3
A	1	T
B	2	F

dplyr::inner_join(a, b, by = "x1")

Join data. Retain only rows in both sets.

x1	x2	x3
A	1	T
B	2	F
C	3	NA
D	NA	T

dplyr::full_join(a, b, by = "x1")

Join data. Retain all values, all rows.



Feedback



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

