

لغة الآر من الصفر إلى الإحتراف

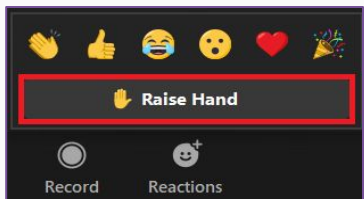
at first I was like...



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 المزوق

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 الاجتماعي لمن يرغب (twitter و GitHub وما إلى ذلك)

-
-
-
-
-
-
-



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk

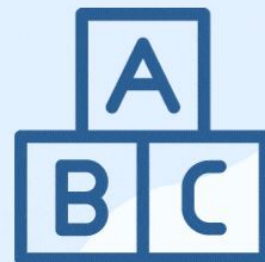


البشر تتكلم بالحروف والأرقام

لكن كيف **تتكلم** لغة R؟

تصنف لغة R أي معلومة إلى أحد هذه الأنواع:

النوع	التعريف	مثال
numerical	رقم متواصل كالطول	12.5
integer	رقم غير متواصل كالأعمار	23
string	كلمات كالأسماء	إيمان، بسمة
factor	أصناف أو فئات أو طبقات كأنواع الأزهار	توليب
logical	هل هي خاطئة أم صائبة	TRUE/FALSE



- Tidy Data
- Base R X Tidyverse
- Dataset
- CSV
- Structure of the files
- Path
- Fetch dataset from GitHub



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

Credit: Allison Horst

Tidy Data

لقراءة المزيد

messy

	treatmenta	treatmentb
John Smith	—	2
Jane Doe	16	11
Mary Johnson	3	1

	John Smith	Jane Doe	Mary Johnson
treatmenta	—	16	3
treatmentb	2	11	1

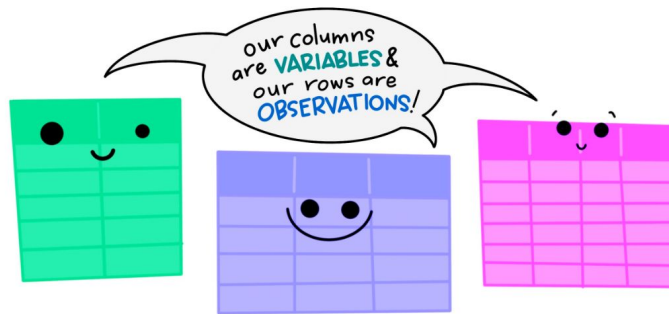
tidy

name	trt	result
John Smith	a	—
Jane Doe	a	16
Mary Johnson	a	3
John Smith	b	2
Jane Doe	b	11
Mary Johnson	b	1



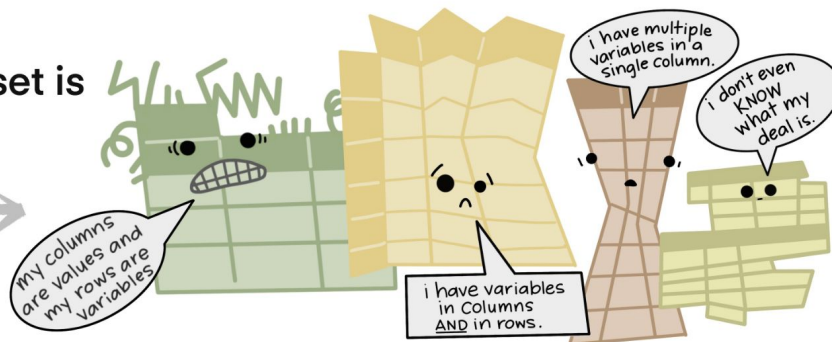
Tidy Data

The standard structure of tidy data means that "tidy datasets are all alike..."



"...but every messy dataset is messy in its own way."

—HADLEY WICKHAM



Credit: Allison Horst



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



BY

- Tidy Data
- Base R X Tidyverse
- Dataset
- CSV
- Structure of the files
- Path
- Fetch dataset from GitHub

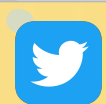


base VS tidyverse VS data.table

Summary of key functions

Environment	base	tidyverse	data.table
Supported data class(es)	<code>data.frame</code>	<code>data.frame</code> , <code>tibble</code>	<code>data.table</code>
Reading data	<code>read.csv</code>	<code>read_csv</code>	<code>fread</code>
Subset by column	<code>[, ...]</code>	<code>select()</code>	<code>[,... ,]</code>
Subset by rows	<code>[... ,]</code>	<code>filter()</code>	<code>[... , ,]</code>
Create new column	<code>df\$y = ...</code>	<code>mutate(tb, y = ...)</code>	<code>[, y := ...,]</code>
Delete a column	<code>df\$y = NULL</code>	<code>select(tb, -y)</code>	<code>[, y := NULL,]</code>
Summarize	<code>apply(df[, y], 2, ...)</code>	<code>summarise()</code>	<code>[, ... (y),]</code>
Grouping	<code>aggregate()</code>	<code>group_by()</code>	<code>[, , by = ...]</code>
Pivot to long	<code>reshape()</code>	<code>pivot_longer()</code>	<code>melt()</code>
Pivot to wide	<code>reshape()</code>	<code>pivot_wider()</code>	<code>dcast()</code>
Joining tables	<code>merge()</code>	<code>left_join()</code>	<code>DT1[DT2, on = ...]</code>





@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



نقاط مهمة:

- Tidy Data
- Base R X Tidyverse
- Dataset
- CSV
- Structure of the files
- Path
- Fetch dataset from GitHub



SAFI Dataset

SAFI (Studying African Farmer-Led Irrigation) is a project which is looking at farming and irrigation methods. This is survey data relating to households and agriculture in Tanzania and Mozambique.

The survey covered such things as; household features (e.g. construction materials used, number of household members), agricultural practices (e.g. water usage), assets (e.g. number and types of livestock) and details about the household members.

https://figshare.com/articles/dataset/SAFI_Survey_Results/6262019



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



SAFI Dataset

	A	B	C	D	E	F	G	H	I	J	K
1	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms	memb_assoc	affect_conflicts	liv_count	items_owned
2	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1	NULL	NULL	1	bicycle
3	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1	yes	once	3	cow_cart
4	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1	NULL	NULL	1	solar_torch
5	4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1	NULL	NULL	2	bicycle
6	5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1	NULL	NULL	4	motorcyle
7	6	God	2016-11-17T00:00:00Z	3	3	muddaub	1	NULL	NULL	1	NULL
8	7	God	2016-11-17T00:00:00Z	6	38	muddaub	1	no	never	1	motorcyle
9	8	Chirodzo	2016-11-16T00:00:00Z	12	70	burntbricks	3	yes	never	2	motorcyle
10	9	Chirodzo	2016-11-16T00:00:00Z	8	6	burntbricks	1	no	never	3	television
11	10	Chirodzo	2016-12-16T00:00:00Z	12	23	burntbricks	5	no	never	2	cow_cart
12	11	God	2016-11-21T00:00:00Z	6	20	sunbricks	1	NULL	NULL	2	radio
13	12	God	2016-11-21T00:00:00Z	7	20	burntbricks	3	yes	never	2	cow_cart
14	13	God	2016-11-21T00:00:00Z	6	8	burntbricks	1	no	never	3	bicycle
15	14	God	2016-11-21T00:00:00Z	10	20	burntbricks	3	NULL	NULL	3	bicycle
16	15	God	2016-11-21T00:00:00Z	5	30	sunbricks	2	yes	once	3	bicycle



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



نقاط مهمة:

- Tidy Data
- Base R X Tidyverse
- Dataset
- **CSV**
- Structure of the files
- Path
- Fetch dataset from GitHub



CSV Dataset

- We can see that the CSV file is imported now.

```
sample - Notepad
File Edit Format View Help
Name,Company,Designation
Rohit,ABC Ltd,Manager
Rohan,ABC Ltd,QA
Kavita,Bluesoftware,Senior Engineer
Tanya,Simplesoft,Analyst
Mahesh,Pinacle,Team Leader
```

	A	B	C
1	Name	Company	Designation
2	Rohit	ABC Ltd	Manager
3	Rohan	ABC Ltd	QA
4	Kavita	Bluesoftware	Senior Engineer
5	Tanya	Simplesoft	Analyst
6	Mahesh	Pinacle	Team Leader
7			

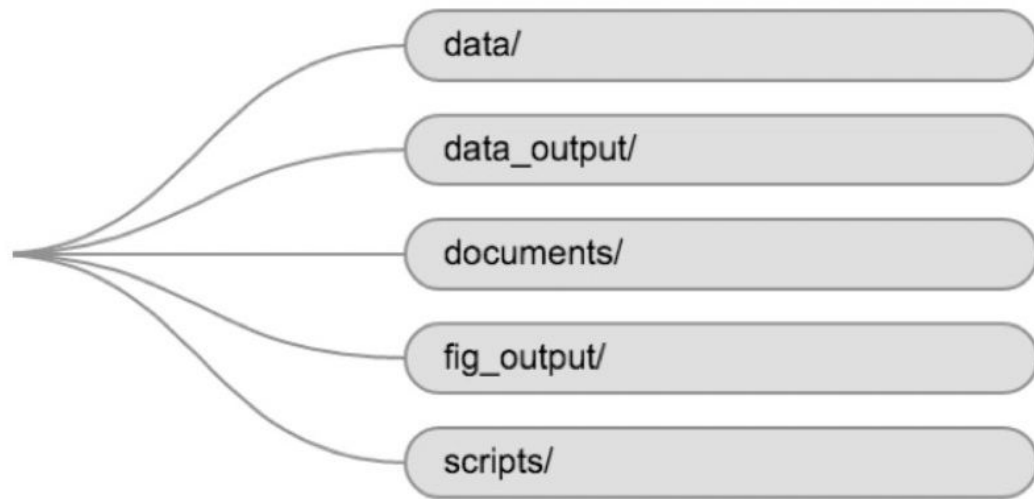


نقاط مهمة:

- Tidy Data
- Base R X Tidyverse
- Dataset
- CSV
- **Structure of the files**
- Path
- Fetch dataset from GitHub



Structure of the Project



نقاط مهمة:

- Tidy Data
- Base R X Tidyverse
- Dataset
- CSV
- Structure of the files
- Path
- Fetch dataset from GitHub



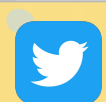
كيفية قراءة البيانات في R

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

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

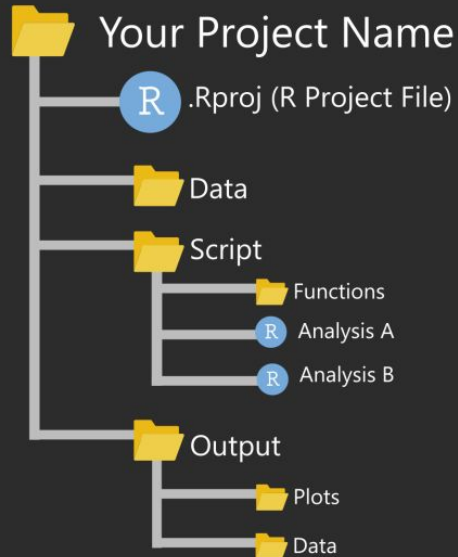
```
read_csv(here("folder", "file"))
```

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

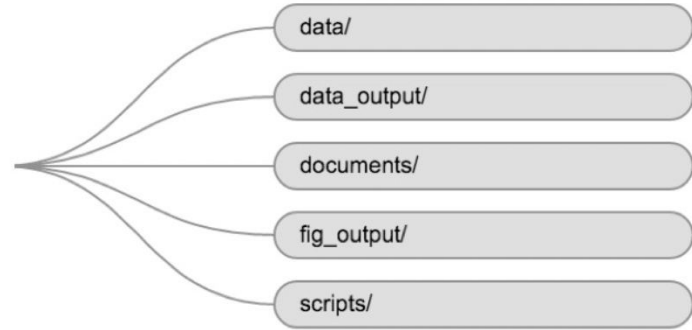


CSV Dataset

A basic R project set up



<https://martinctc.github.io>



@OpenSciSaudi, @batool664



www.osc-ksa.com



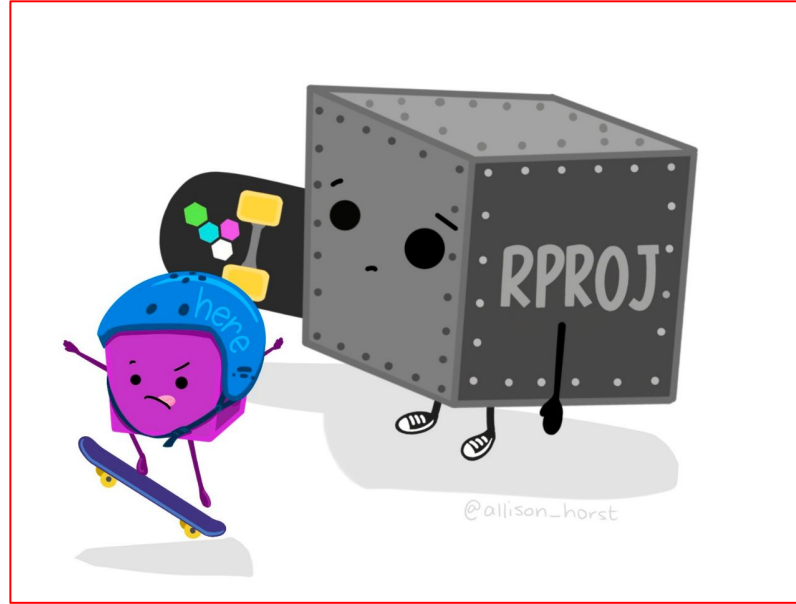
batool@liverpool.ac.uk



كيفية قراءة البيانات في R



Credit: Allison Horst



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



كيفية قراءة البيانات في R



Credit: Allison Horst



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



نقاط مهمة:

- Tidy Data
- Base R X Tidyverse
- Dataset
- CSV
- Structure of the files
- Path
- Fetch dataset from GitHub



كيف احصل على البيانات

BatoolMM / from-Zero-to-hero-in-r Public Pin Unwatch 1

generated from [Open-Science-Community-Saudi-Arabia/Template](#)

<> **Code** Issues Pull requests 1 Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code

BatoolMM Merge branch 'main' of [https://github.com/BatoolMM/from-Zero-t...](https://github.com/BatoolMM/from-Zero-to-...) ✓ 96b4776 16 hours ago ⌚ 43 commits

📁 .github/workflows	Initial commit	7 days ago
📁 Untitled	add session 5	16 hours ago
📁 bin	Initial commit	7 days ago
📁 data	add session 5	16 hours ago
📁 docs	add session 5	16 hours ago
📁 fig	add session 5	16 hours ago
📁 imgsgif	add session 5	16 hours ago
📁 results	Initial commit	7 days ago
📄 .DS_Store	add session 5	16 hours ago
📄 .gitignore	Initial commit	7 days ago
📄 CITATION.md	Initial commit	7 days ago

About Settings

An Arabic course to introduce R

[tutorial](#) [course-materials](#) [rstats](#)
[tutorial-exercises](#)

Readme
CC-BY-4.0 license
Code of conduct
Cite this repository
2 stars
1 watching
13 forks

Releases

No releases published
[Create a new release](#)



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



كيف احصل على البيانات

Local Repository



Your Fork



Batool's repo



@OpenSciSaudi, @batool664



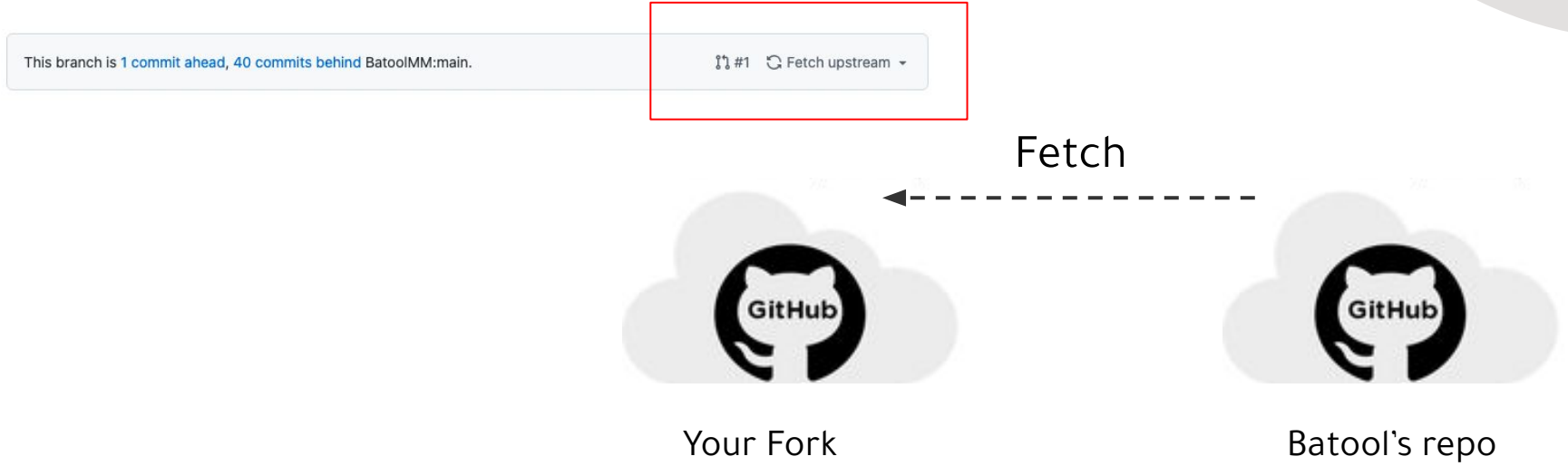
www.osc-ksa.com



batool@liverpool.ac.uk



كيف احصل على البيانات



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



كيف احصل على البيانات



@OpenSciSaudi, @batool664



www.osc-ksa.com

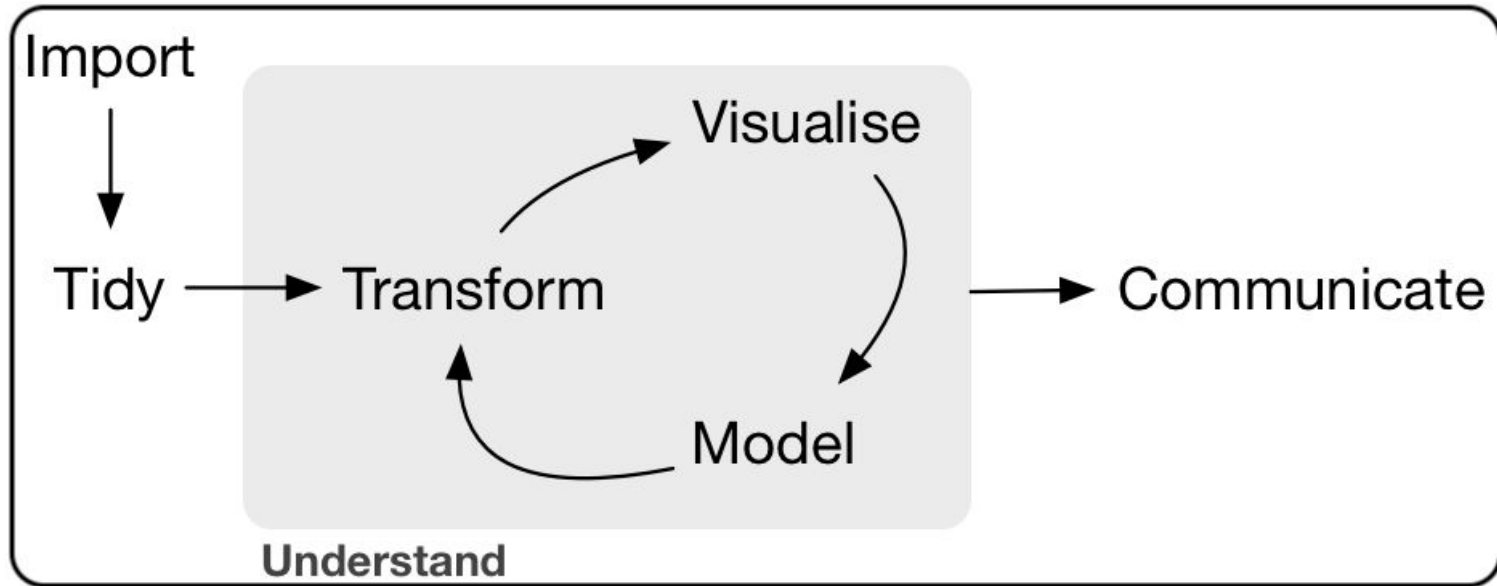


batool@liverpool.ac.uk



كيف احصل على البيانات





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) = % > %

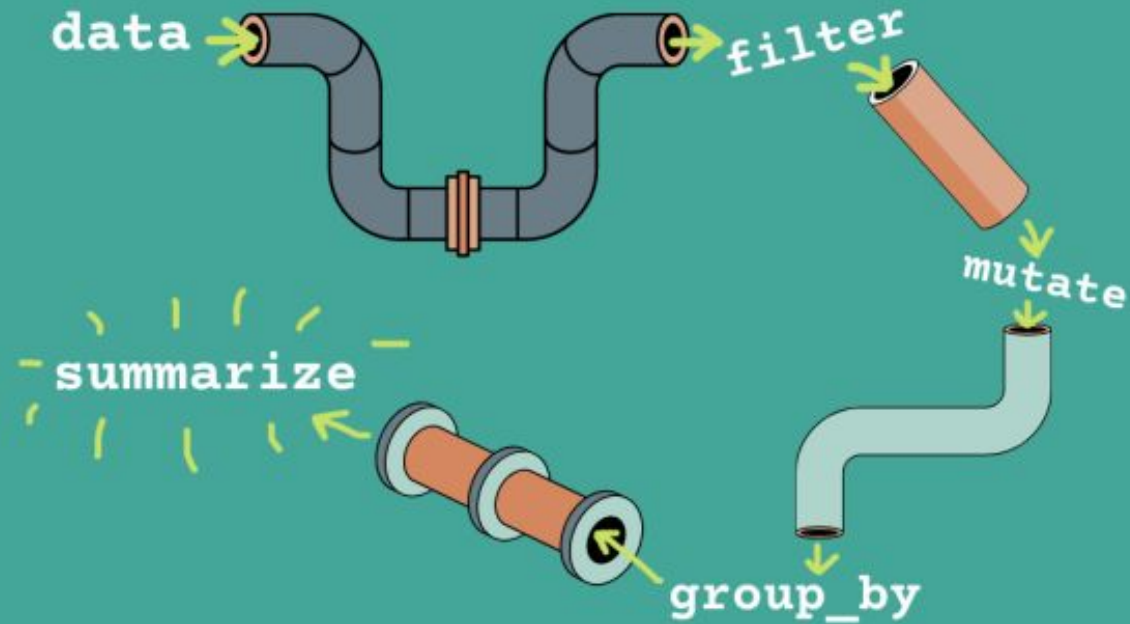
معنى هذا الرمز في لغة R

(و تُم) = %>%

dataset %>%

select(**column**)

Using pipes in



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk



NA, NULL, Inf & NaN

- **NA** # missing
- **NULL** # undefined
- **Inf** # infinite 3/0
- **NaN** # Not a number Inf/Inf



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؟

