

مقدمة في لغة الأَر R وممارسات العلوم المفتوحة

Beginners Guide for R through Open Data Science Practices

1- Why R?



مرحبا في كورس "مقدمة في لغة الأ R وممارسات العلوم المفتوحة"

التاريخ: ٢٠ جولي ٢٠٢٢

تسجيل الأسماء

الأسم / المدينة / حسابات وسائل التواصل الاجتماعي لمن يرغب (twitter و GitHub وما إلى ذلك)
• بتول المرزوق / الشرقية / GitHub: BatooolMM / Twitter: @batoool664, @OpenSciSaudi

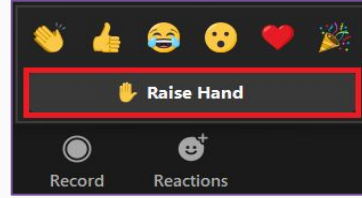
-
-
-
-
-
-
-
-

سؤال عام لرفع التوتر وكسر الجليد

ماهو اجمل كتاب أو مقالة قرأته؟

• بتول: Teaching Tech Together

-
-
-



قواعد الدورة

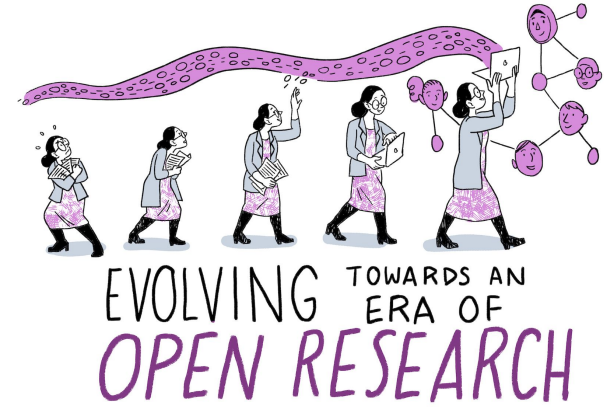
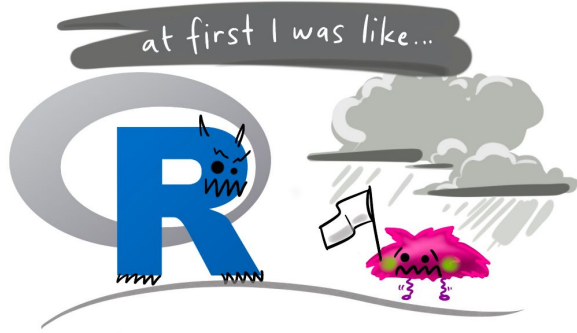
- كتابة الاسم في Google Doc
- الإجابة على Icebreaker question
- إطفاء الجوال
- القواعد السلوكية
- لا تتردد في طرح أي سؤال خلال الجلسة



@batoool664, @OpenSciSaudi



ما هو هدف الكورس؟



Scriberia

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



@batool664, @OpenSciSaudi



من المُنظمين للكورس؟



Open Science Community Saudi Arabia

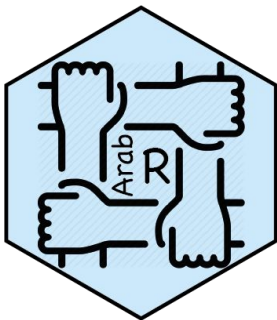
مجتمع العلوم المفتوحة في المملكة العربية السعودية



KSAU-HS

KING SAUD BIN ABDULAZIZ
UNIVERSITY FOR HEALTH SCIENCES

A University for the Nation's Health



THE
CARPENTRIES



@batool664, @OpenSciSaudi



ماهو الكورس؟



Schedule

Day 1

Before starting	Pre-workshop survey
10:00	Housekeeping rules and intro of the communities
10:30	Introduction
11:00	R & RStudio, R Markdown
11:45	Visualizing data with ggplot2
12:45	Data transformation with dplyr
2:00	End

Day 2

10:00	Version control with git
11:00	Collaborating with you and others with Github
12:00	Break
12:10	Become a champion of open (data) science
13:00	ReproHack
End	Post-workshop survey



ماهو الكورس؟



Schedule

	Setup	Download files required for the lesson
00:00	1. Introduction	What will I learn during this workshop? What are the tools that I will be using? What are the tidy data principles? What is working in a more open way beneficial?
00:30	2. R & RStudio, R Markdown	How do I orient myself in the RStudio interface? How can I work with R in the console? What are built-in R functions and how do I use their help page? How can I generate an R Markdown notebook?
01:30	3. Visualizing data with ggplot2	How can I make publication-grade plots with ggplot2? What are the key concepts underlying ggplot2 plotting? What are some of the visualisations available through ggplot2? How can I save my plot in a specific format (e.g. png)?
03:00	4. Data transformation with dplyr	How do I perform data transformations such as removing columns on my data using R? What are tidy data (in opposition to messy data)? How do I import data into R (e.g. from a web link)? How can I make my code more readable when performing a series of transformation?
04:00	5. Version control with git	What is version control? How do I use it? What is the difference between git and Github? What benefits does a version control system bring in for my research?
05:00	6. Collaborating with you and others with Github	How can I develop and collaborate on code with me or another scientist? How can I give access to my code to another collaborator? How can I keep code synchronised with another scientist? How can I solve conflicts that arise from that collaboration? What is Github?
06:00	7. Become a champion of open (data) science	
07:30	Finish	



@batool664, @OpenSciSaudi





Exercises and **assignments**



@batool664, @OpenSciSaudi





This is **NOT JUST** a coding course. We'll teach
you **best practices** and **reproducibility!**



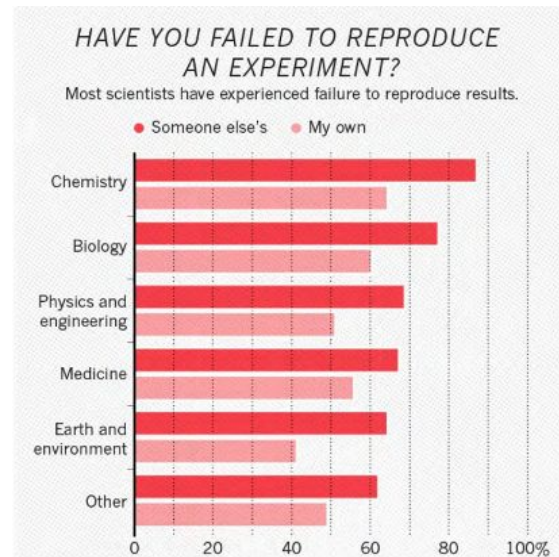
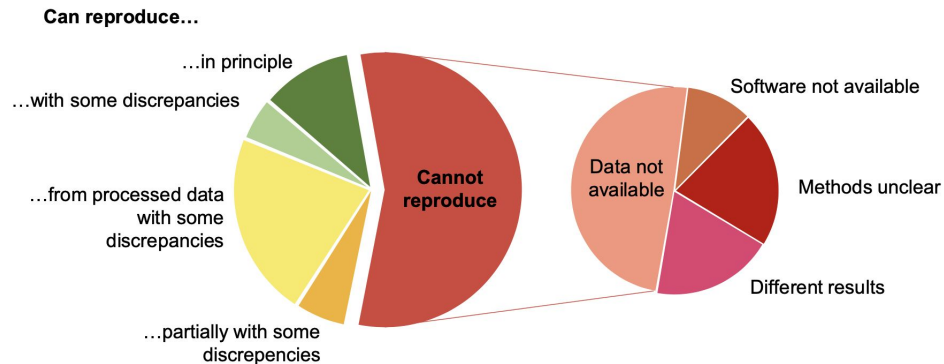
@batool664, @OpenSciSaudi



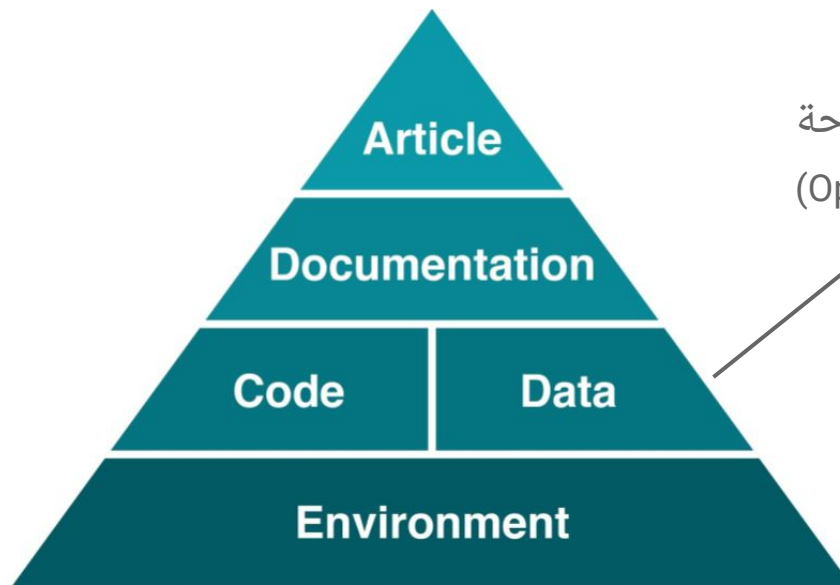
Half of top cancer studies fail high-profile reproducibility effort

Barriers to reproducing preclinical results included unhelpful author communication, but critics argue that one-time replication attempts don't tell the whole story.

Replication of data analyses in 18 articles on microarray-based gene expression profiling published in Nature Genetics in 2005–2006:



لماذا العلوم المفتوحة؟



مبادئ العلوم المفتوحة
(Open Science practices)

Credit: [coderefinery](https://coderefinery.org/)



@OpenSciSaudi, @batool664



www.osc-ksa.com



batool@liverpool.ac.uk

RrproHack

"an article about computational result is advertising, not scholarship.

The actual scholarship is the full software environment, code and data, that produced the result."

John Claerbout

Methodology article | [Open Access](#) | [Published: 06 September 2019](#)

Visualization methods for differential expression analysis

[Lindsay Rutter](#) , [Adrienne N. Moran Lauter](#), [Michelle A. Graham](#) & [Dianne Cook](#)

[BMC Bioinformatics](#) **20**, Article number: 458 (2019) | [Cite this article](#)

50k Accesses | **5** Citations | **33** Altmetric | [Metrics](#)

[Link](#) to the paper



J. Matthew Rhett, PhD

@DrJMatthewRhett

Replying to [@Thatsregrettab1](#) [@oncogenejournal](#) and [@MassGeneralNews](#)

The reproducibility crisis is the biggest open secret in science, and has been long known to be an extraordinarily big problem in cancer research. [nature.com/articles/48353...](#)

No alt



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models. Credit: S. GSCHMEISSNER/SPL

Unquestionably, a significant contributor to failure in oncology trials is the quality of published preclinical data. Drug development relies heavily on the literature, especially with regards to new targets and biology. Moreover, clinical endpoints in cancer are defined mainly in



@batool664, @OpenSciSaudi



تعريف بالمجتمعات العلمية والمنظمين



@batool664، @OpenSciSaudi



من المُنظمين للكورس؟



Open Science Community Saudi Arabia

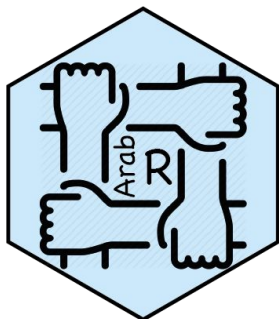
مجتمع العلوم المفتوحة في المملكة العربية السعودية



KSAU-HS

KING SAUD BIN ABDULAZIZ
UNIVERSITY FOR HEALTH SCIENCES

A University for the Nation's Health



THE
CARPENTRIES



@batool664, @OpenSciSaudi



بتول المرزوق

Community Lead



Subject Matter Experts (SME)



Committee member



Honorary research Fellow



UNIVERSITY OF
LIVERPOOL



Core contributor



Education
Committee



@batool664, @OpenSciSaudi



The International Network of Open Science Communities (INOSC)



Overview of the current Open Science Communities. As of March 2022, the International Network of Open Science Communities (INOSC) comprises Open Science Communities at 12 locations in the Netherlands, 1 in Ireland, 1 in Sweden, 1 in Lisbon, 1 in Serbia and 1 in Saudi Arabia. Image by Anita Eerland, licensed under CC BY-ND 3.0.

رؤية 2030

المملكة العربية السعودية
KINGDOM OF SAUDI ARABIA

OSCSA
Open Science Community Saudi Arabia



مجتمع العلوم المفتوحة في المملكة العربية السعودية تم إنشاؤه تماشياً مع رؤية المملكة العربية السعودية لعام 2030، والتي تركز على تثبيت القيم و تعزيز المعرفة وتحسين تكافؤ فرص الحصول على التعليم

The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)



@OpenSciSaudi, @batool664



www.osc-ksa.com



admin@openscisaudi.com

Membership in OSCSA



- Learn about Open Science practices through **our Workshops**.
- OSCSA helps you to **get connected with like-minded** people from other communities that adopts Open Science practices (**e.g. Turing Way, Carpentries, ArabR, ...**).
- Support to **lead your own project**, and provide more visibility and discoverability for your work (e.g. conferences, newsletter, ...)
- **Nominate** yourself in the and **vote** for the international International Network of Open Science & Scholarship (**INOSC**) **Board**

You can join through [this link](#)



@OpenSciSaudi, @batool664



www.osc-ksa.com



admin@openscisaudi.com

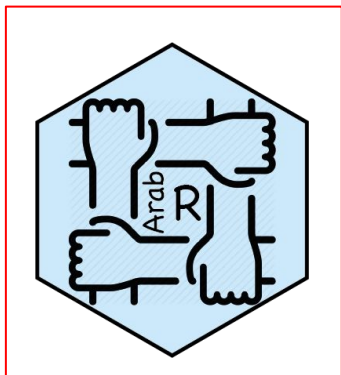
من المُنظمين للكورس؟



+



KSAU-HS
KING SAUD BIN ABDULAZIZ
UNIVERSITY FOR HEALTH SCIENCES
A University for the Nation's Health



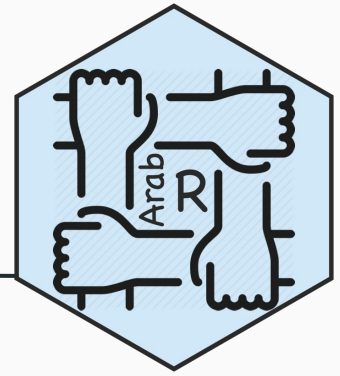
**THE
CARPENTRIES**



@batool664, @OpenSciSaudi



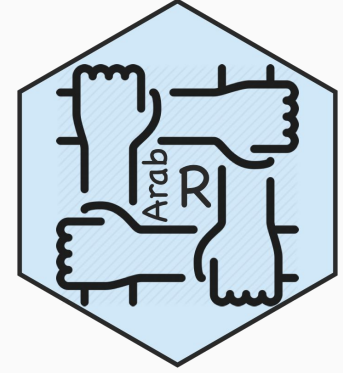
ArabR



- التمثيل العربي لمستخدمي لغة R **منخفض جدا** في المؤتمرات الدولية.
- عدد التجمعات التدريبية للغة R في الشرق الأوسط **محدودة جدا**.
- **لا يوجد منصة** تربط بين مستخدمي وممارسي لغة R في الوطن العربي.



تتمثل **رؤية ArabR** في بناء قدرات مستدامة لمستخدمي R في الوطن العربي **ودعم مستخدمي R** من خلال التواصل **وزيادة ظهورهم** على المستوى الدولي والدفع لإنشاء المزيد من مجموعات useR / R-Ladies في الوطن العربي.



Hussain
Alsalman



Amal Tlili



Iman Al-Hasani



Fahad
Almsned



Haifa Ben
Messaoud



Kamila
BENADROUCHE



Mohammed
El Fodil



Mouna
BELAID



@batool664, @OpenSciSaudi





Hussain Alsalman

@Arabian_Analyst

حسين يتمتع **بخبرة تزيد عن 10 سنوات** في إدارة مؤشرات أداء الأعمال في مجال تكنولوجيا المعلومات مع أرامكو السعودية. لديه ماجستير في إدارة الأعمال. يستمتع بالبرمجة بلغة R. تنضح مدونته الشخصية بالعديد من حالات الاستخدام للتحليلات باستخدام لغة R.

[رابط لمدونة المحلل العربي](#)



@Arabian_Analyst








Haifa Ben Messaoud

Haifa Ben messaoud is **an engineer in statistics and data analysis** graduated from the Higher school of statistics and data analysis of Tunis.

She is a **Data Scientist and AWS Machine Learning developer in Cimpres group based in San Diego - USA with 6 years experience.**

She has experience in many fields like pharma, banking and e-commerce. She also taught Artificial Intelligence in ESSAI and **also taught R at many universities in Tunis.** She is a member of the administrative board of ATIS. She is interested in Data visualisation and Dashboarding and loves sharing her knowledge with the community, the reason why **she co founded R-Ladies Tunis**

 benmessaoud.haifa@hotmail.com
 Haifa Ben Messaoud
 @haifa.bm

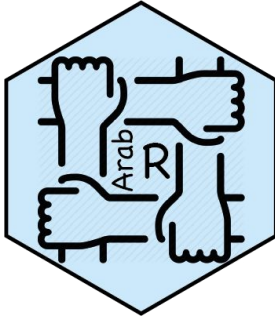
من المُنظمين للكورس؟



+



KSAU-HS
KING SAUD BIN ABDULAZIZ
UNIVERSITY FOR HEALTH SCIENCES
A University for the Nation's Health



**THE
CARPENTRIES**

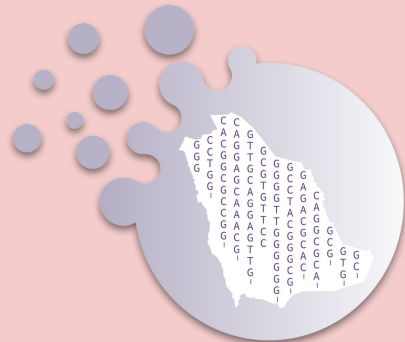


@batool664, @OpenSciSaudi



الجمعية السعودية للمعلوماتية الحيوية وعلم الأحياء الحاسوبي

Saudi Society for Bioinformatics and Computational Biology



SSBCB



@S_Bioinformatics



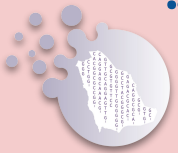
ssbcg.github.io



@SSBIOINFORMATICS

الجمعية السعودية للمعلوماتية الحيوية وعلم الأحياء الحاسوبي

Saudi Society for Bioinformatics and Computational Biology



عبدالرحمن العسيري



عبدالرحمن السواجي



@S_Bioinformatics



UMC Utrecht



ssbcb.github.io

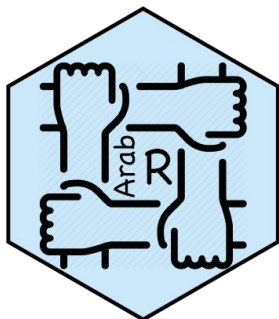
من المُنظمين للكورس؟



+



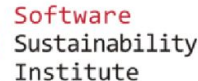
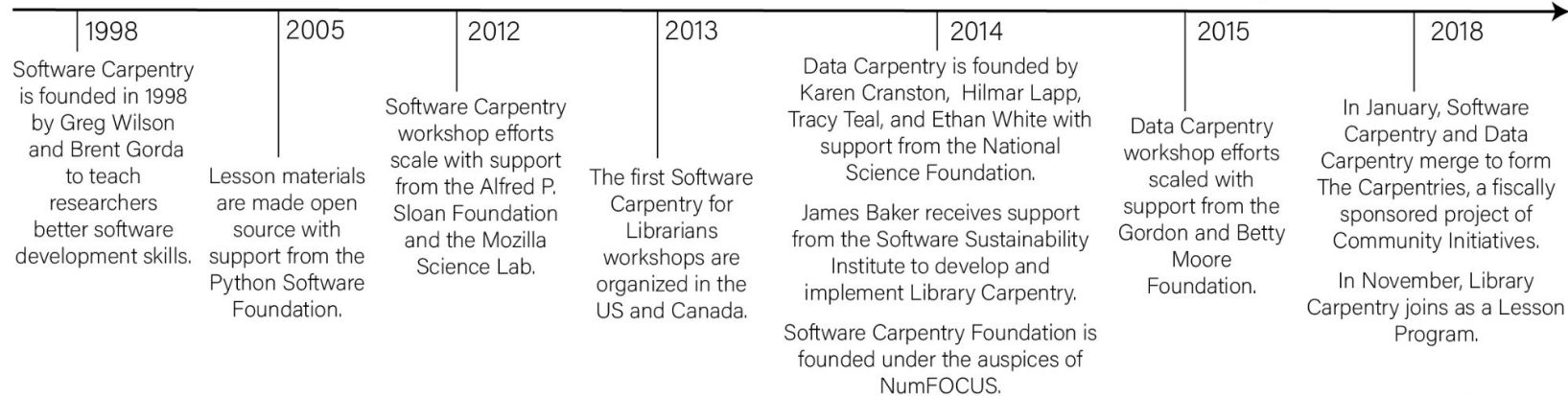
KSAU-HS
KING SAUD BIN ABDULAZIZ
UNIVERSITY FOR HEALTH SCIENCES
A University for the Nation's Health

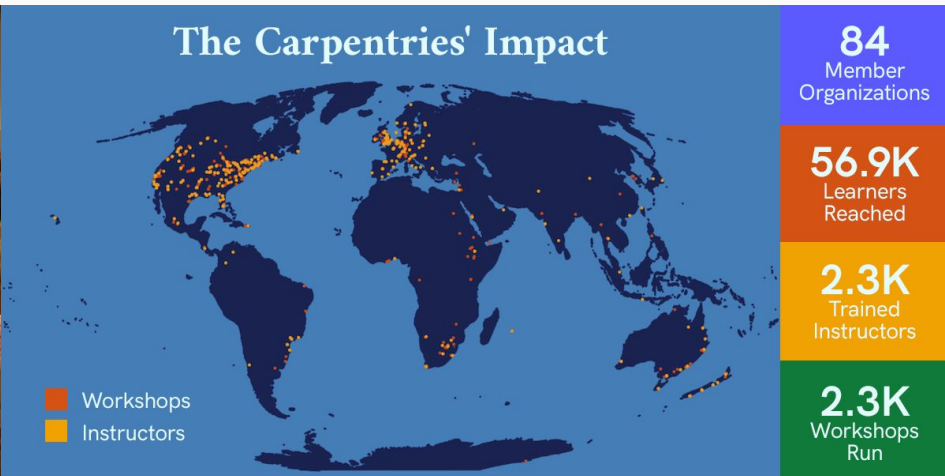


@batool664, @OpenSciSaudi



A VERY BRIEF HISTORY OF









Monah is a postdoc researcher in gene therapy at the San Raffaele Hospital in Milan, Italy and part of the Bioinformatics Core working on analysis of different -omics data from patients and cellular models.

He has been a Bioinformatics tutor at the University of Pavia, Italy for 4 years teaching Masters student lessons about R, python and shell programming languages. He joined recently The Harvard Stem Cell Institute in Boston, USA as a Bioinformatics trainer.

Since 2019, Monah is a certified instructor for the Carpentries organization where he has taught several workshops in Europe and the US both in-person and online. He also joined recently the Carpentries as a lessons Maintainer.



Dr. Monah Abou Alezz

- ✉ Email: aboualezz.monah@hsr.it
-  LinkedIn: Monah Abou Alezz
-  Twitter: @MonahAbouAlezz
-  Github: @monahton
-  Website: monahton.github.io

Carpentry Instructors



عمار الخالدي شريك مؤسس في PSMap، مدرب مع
الهيئة السعودية للمهندسين و Data Carpentry.

متخصص في تحليل وحل المشكلات، يهتم ويدرب في
مجال تحليل البيانات، حل المشكلات، النمذجة المالية،
الاعتمادية وإدارة المخاطر التشغيلية.

Ammar Alkhalidi



Ammar_alkhalidi



Ammaralkhalidi



psmap.io



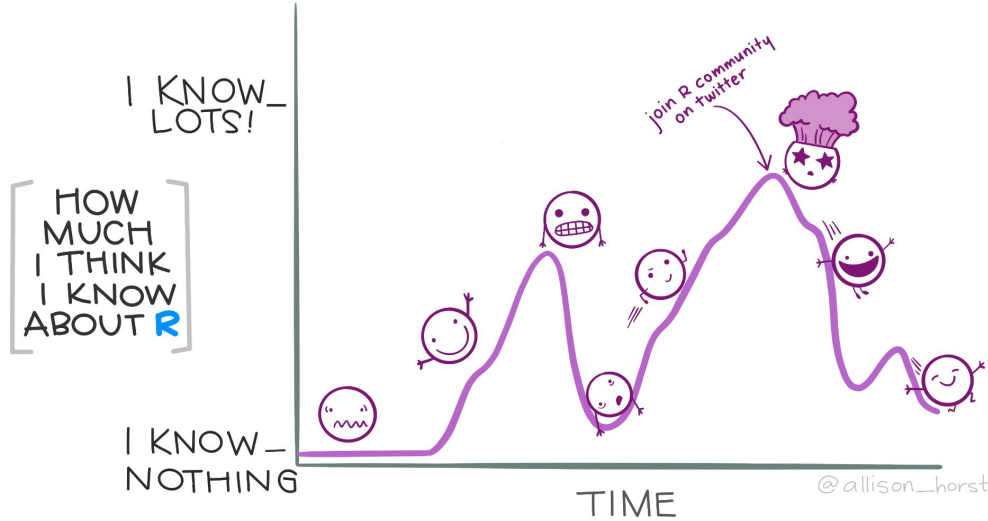
Most of the materials are
developed by the **Carpentries**



@batool664, @OpenSciSaudi



ماهي صعوبة اللغة R؟

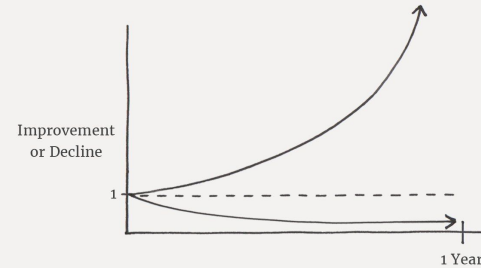


Credit: Allison Horst

The Power of Tiny Gains

$$1\% \text{ better every day } 1.01^{365} = 37.78$$

$$1\% \text{ worse every day } 0.99^{365} = 0.03$$



JamesClear.com



@batool664, @OpenSciSaudi



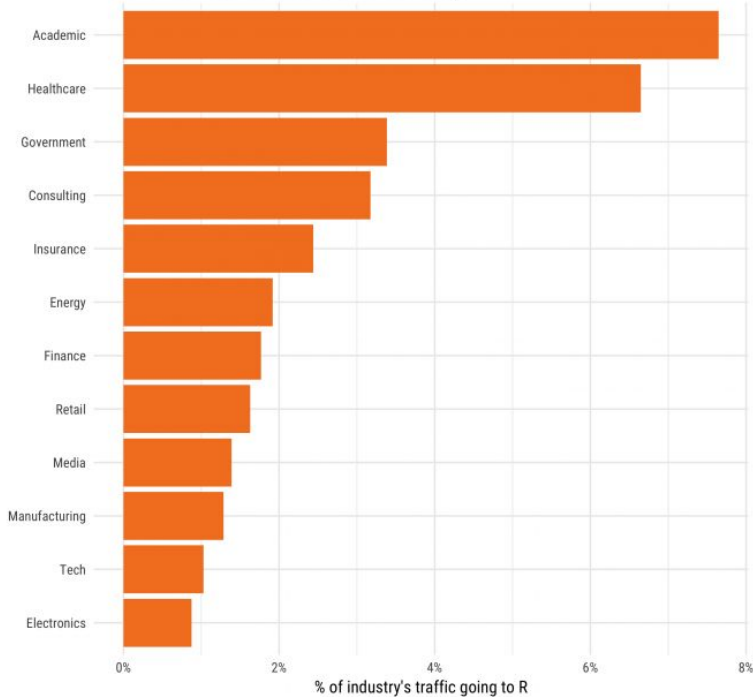
ماهي لغة الآر (R)



أهمية لغة R ؟

Visits to R by industry

Based on visits to Stack Overflow questions from the US/UK in January-August 2017.
The denominator in each is the total traffic from that industry.



Companies that use R for Analytics



@batool664, @OpenSciSaudi



معنى البرمجيات المفتوحة؟

البرامج المجانية \neq البرمجيات المفتوحة المصدر

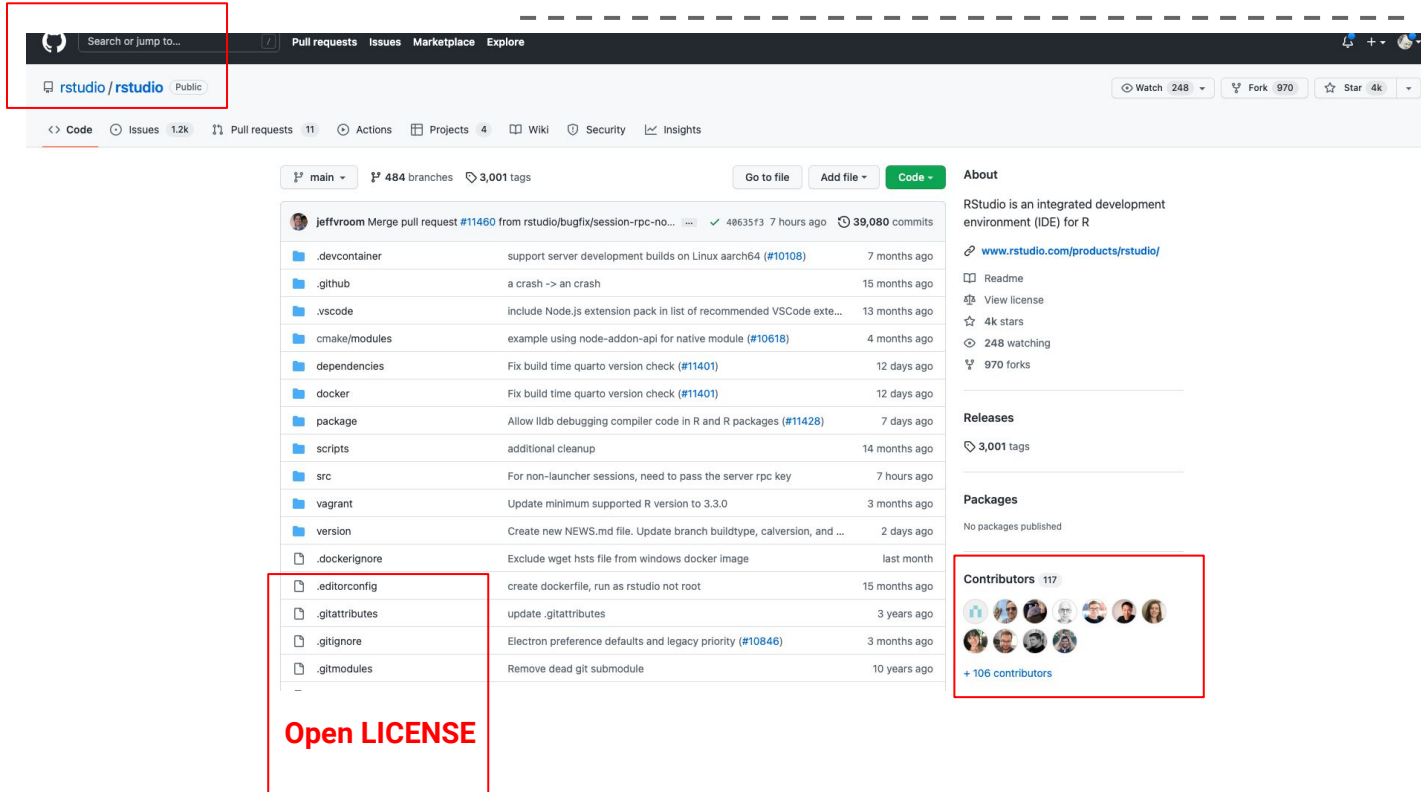
Free Software \neq Open Source



أمثلة للبرمجيات المفتوحة

Examples for Open Source

معنى البرمجيات المفتوحة



The screenshot shows the GitHub repository for RStudio. The repository name 'rstudio / rstudio' is highlighted with a red box. The file list on the left includes .devcontainer, .github, .vscode, cmake/modules, dependencies, docker, package, scripts, src, vagrant, version, .dockerignore, .editorconfig, .gitattributes, .gitignore, and .gitmodules. The contributors section on the right shows 117 contributors, with a red box highlighting the list of contributor avatars. A red box at the bottom left contains the text 'Open LICENSE'.

Search or jump to...

rstudio / rstudio Public

Watch 248 Fork 970 Star 4k

Code Issues 1.2k Pull requests 11 Actions Projects 4 Wiki Security Insights

main 484 branches 3,001 tags

Go to file Add file Code

About

RStudio is an integrated development environment (IDE) for R

www.rstudio.com/products/rstudio/

Readme View license 4k stars 248 watching 970 forks

Releases 3,001 tags

Packages No packages published

Contributors 117

+ 106 contributors

jeffvroom Merge pull request #11460 from rstudio/bugfix/session-rpc-no... 40635f3 7 hours ago 39,080 commits

.devcontainer	support server development builds on Linux aarch64 (#10108)	7 months ago
.github	a crash -> an crash	15 months ago
.vscode	include Node.js extension pack in list of recommended VSCode exte...	13 months ago
cmake/modules	example using node-addon-api for native module (#10618)	4 months ago
dependencies	Fix build time quarto version check (#11401)	12 days ago
docker	Fix build time quarto version check (#11401)	12 days ago
package	Allow libd debugging compiler code in R and R packages (#11428)	7 days ago
scripts	additional cleanup	14 months ago
src	For non-launcher sessions, need to pass the server rpc key	7 hours ago
vagrant	Update minimum supported R version to 3.3.0	3 months ago
version	Create new NEWS.md file. Update branch buildtype, calversion, and ...	2 days ago
.dockerignore	Exclude wget hsts file from windows docker image	last month
.editorconfig	create dockerrfile, run as rstudio not root	15 months ago
.gitattributes	update .gitattributes	3 years ago
.gitignore	Electron preference defaults and legacy priority (#10846)	3 months ago
.gitmodules	Remove dead git submodule	10 years ago

Open LICENSE



@batool664, @OpenSciSaudi



Track your work with GitHub



Batool Almarzouq
BatoolMM

Ph.D. in Biochemistry. Experience in cancer bioinformatics, RNA-Seq, network biology, pathway enrichment, and integrated -omics analyses, Docker, HPC (She/her).

Edit profile

73 followers · 97 following

University of Liverpool

<https://batool-almarzouq.netlify.app/>

@batool664

Achievements



Pinned

Customize your pins

exonR Public

R package for generating beautiful and customizable plots for the exons and introns of any gene

R ☆ 10

DeepSulfation Public

It uses deep learning to predict sulfation sites from protein sequence.

Jupyter Notebook ☆ 4

PlantAI Public

This is a simple application which use a convolutional neural network (CNN) to characterise the plants in my home and help in taking care of them.

Jupyter Notebook ☆ 4

Collaborating-on-Open-Data-Science-Projects Public

This is the materials for a workshop titled "Collaborating in Open Data Science Projects", part of WIDS2021.

HTML 🗑️ 1

An-Open-Science-Approach-to-Machine-Learning Public

"An Open Science Approach to Machine Learning in Medical and Biological Research talk" in Saudi Data community

JavaScript

EV11 Public

Study the phosphorylation of EV11 through MD

Python ☆ 2

22,425 contributions in the last year

Contribution settings ▾



2022

2021

2020

2019



@batool664, @OpenSciSaudi



بيانات مفتوحة

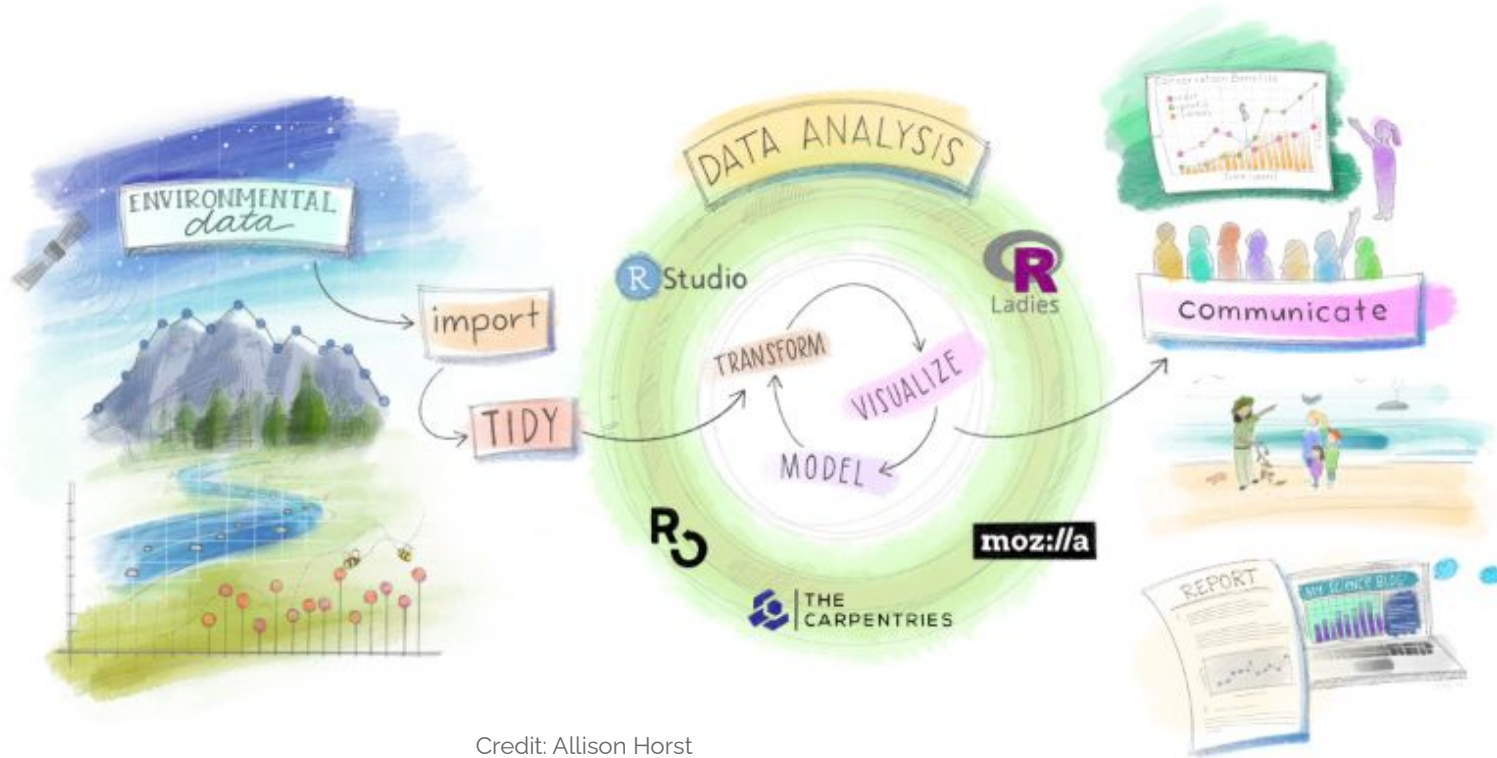


Credit:opensocialclusters.eu



@batool664, @OpenSciSaudi





Credit: Allison Horst



@batool664, @OpenSciSaudi



Rmarkdown

TEXT. CODE. OUTPUT.
(GET IT TOGETHER, PEOPLE.)



Credit: Allison Horst



@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

O RLY?

@ThePracticalDev

The internet will make those bad words go away



Essential

Googling the
Error Message

O RLY?

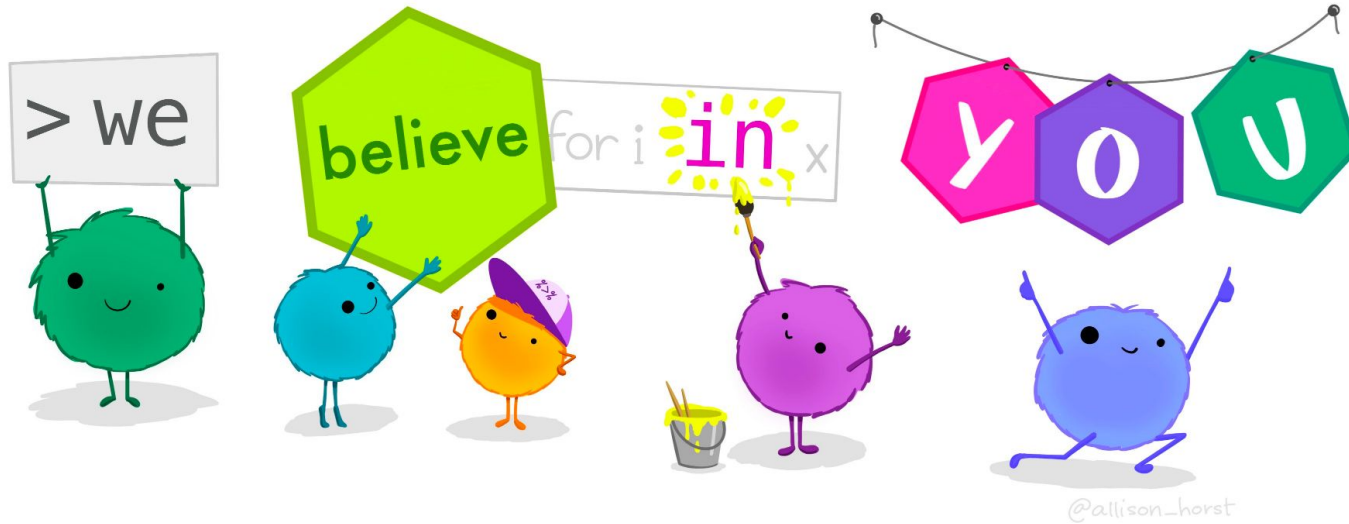
The Practical Developer
@ThePracticalDev



@batool664, @OpenSciSaudi



remote R learners,



Credit: Allison Horst



@batool664, @OpenSciSaudi





Feedback

