



THE
CARPENTRIES

مدخل في لغة الآر (R) والعلوم المفتوحة لعلماء المعلوماتية الحيوية

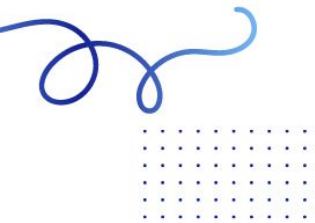
Batool Almarzouq
Monah Abou Alezz
Annajiat Alim Rase
Abdulrahman Dallak
Mona Alsharif



Open Science Community Saudi Arabia

مجتمع العلوم المفتوحة في المملكة العربية السعودية





Day 1

اليوم الأول



@OpenSciSaudi

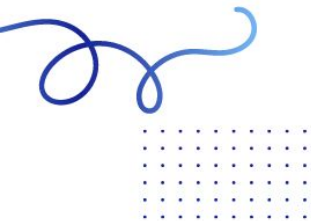


<https://github.com/Open-Science-Community-Saudi-Arabia>

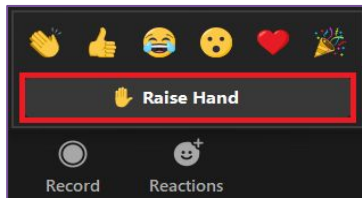


<https://osc-ksa.com/>





قواعد الدورة



Welcome to the Intro to R and Open Science Practices for Biologists . Workshop!

This is a 3-days hands-on workshops organised by Open Science Community Saudi Arabia (OSCSA) to sFDA staff.

The workshop will introduce data wrangling and visualisation with Tidyverse and ggplot2 and how to use packages from Bioconductor to analyse biological data. It'll also introduce you to best practices in open science and Reproducibility using git and GitHub.

Date: 31st of October - 2nd of November (11:00 am - 2:00pm Riyadh time).

Workshop Website: <https://open-science-community-saudi-arabia.github.io/sFDA-Careerries-Workshop/>
Workshop GitHub: <https://github.com/Open-Science-Community-Saudi-Arabia/sFDA-Careerries-Workshop>

Zoom Call:

<https://liverpool-ac-uk.zoom.us/j/92080791938?pwd=H2IjWDdmVnRlUjBwVWVxYjFkNFNkK090dDQ2>

Meeting ID: 920 8079 1938
Passcode: +9qWwS**

● كتابة الاسم في Etherpad

● إطفاء الجوال

● القواعد السلوكية

● لا تتردد في طرح **أي سؤال** خلال الجلسة



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>



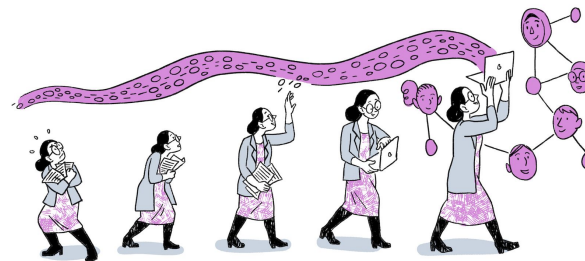
ما هو هدف الكورس؟



at first I was like...



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



EVOLVING TOWARDS AN ERA OF OPEN RESEARCH

Scriberia

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



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>



من المُنظمين للكورس؟



Open Science Community Saudi Arabia

مجتمع العلوم المفتوحة في المملكة العربية السعودية



THE
CARPENTRIES



الهيئة العامة للأدوية والدواء
Saudi Food & Drug Authority



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>

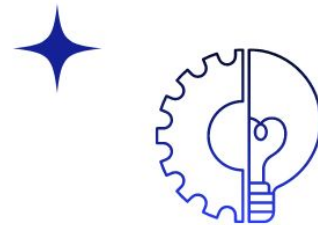
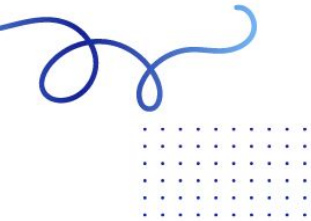


ماهو الكورس؟



	Setup	Download files required for the lesson
09:00	1. Why care about open (data) science?	What is Open Science? What is Open and Reproducible Research Practices?
09:50	2. Introducing R and RStudio IDE	Why use R? Why use RStudio and how does it differ from R?
10:35	3. Collaborating with Github	How can I develop and collaborate on code with 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 are Github
12:05	4. R Basics	What will these lessons not cover? What are the basic features of the R language? What are the most common objects in R?
13:25	5. Introduction to the example dataset and file type	What data are we using in the lesson? What are VCF files?
13:40	6. R Basics continued - Factors and data frames	How do I get started with tabular data (e.g. spreadsheets) in R? What are some best practices for reading data into R? How do I save tabular data generated in R?
15:10	7. Using packages from Bioconductor	How do I use packages from the Bioconductor repository?
15:23	8. Data Wrangling and Analyses with Tidyverse	How can I manipulate data frames without repeating myself?
16:18	9. Data Visualization with ggplot2	What is ggplot2? What is mapping, and what is aesthetics? What is the process of creating a publication-quality plots with ggplot in R?





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



@OpenSciSaudi

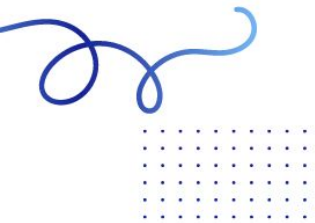


<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>





Exercises and assignments



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>



Reproducibility Crisis in **all** disciplines! ✨



The \$28 Billion a Year Research Reproducibility Crisis

📅 October 19, 2015

The biggest crisis in the future of healthcare has nothing to do with the kind of issues that politicians talk about, such as payment systems and accessibility. Since cost-effective solutions to tragic and costly conditions like heart disease, dementia, diabetes, and cancer depend on bringing to market new treatments that rely upon genomics, nanotechnology, and molecular biology, the *biggest* problem is that at least half of the research money we invest in them is wasted. How can this be? What are the implications? What are the solutions? We'll show you.

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

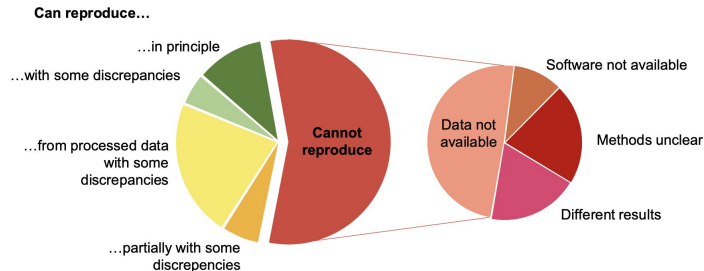
NEWS | 09 December 2021

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:



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

J. P. A. Ioannidis et al., "Repeatability of published microarray gene expression analyses," Nat. Genet., vol. 41, no. 2, pp. 149–155, Feb. 2009, doi: 10.1038/ng.295



@OpenSciSaudi



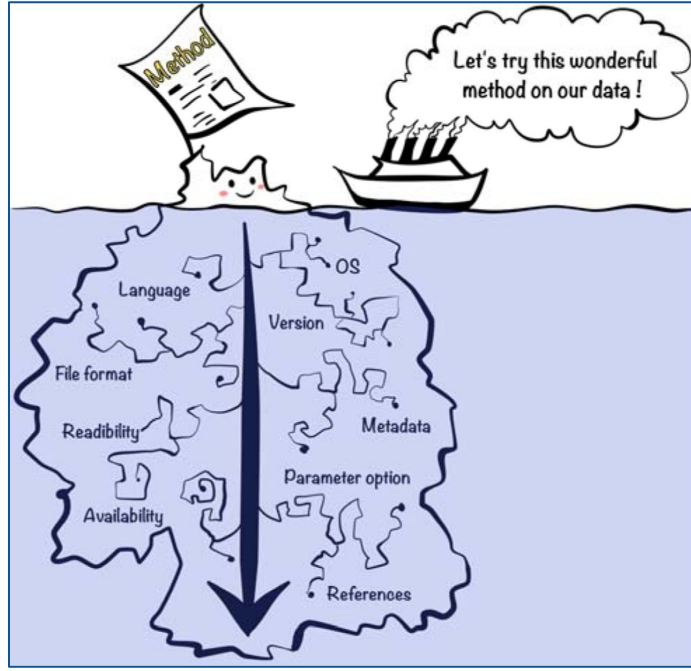
<https://github.com/Open-Science-Community-Saudi-Arabia>



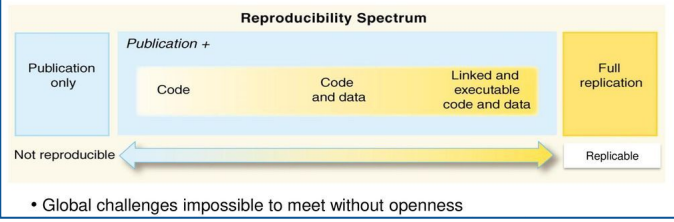
<https://osc-ksa.com/>



Reproducibility Crisis in **all** disciplines! ✨



Reproducibility requires openness

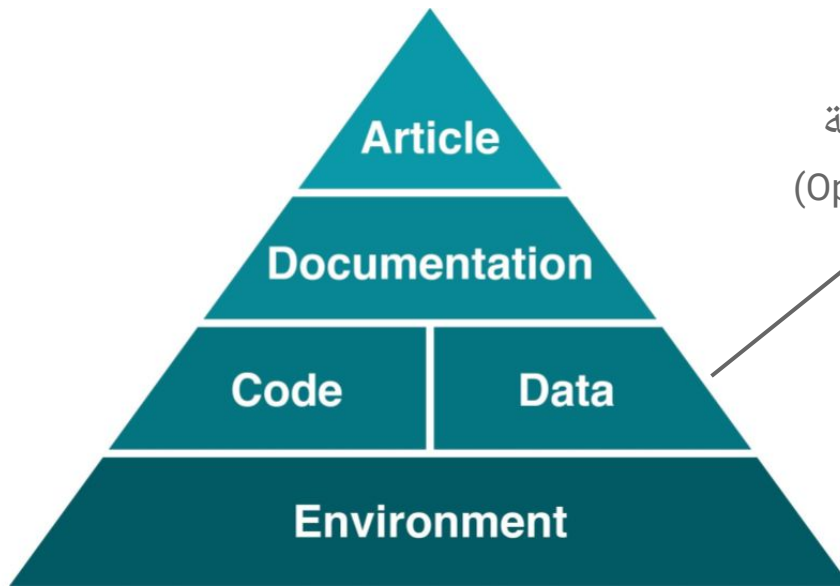


[Peng \(2011\)](#)

Credit: Yang-Min Kim, Jean-Baptiste Poline, Guillaume Dumas, Experimenting with reproducibility: a case study of robustness in bioinformatics, *GigaScience*, Volume 7, Issue 7, July 2018, giy077, <https://doi.org/10.1093/gigascience/giy077>



لماذا العلوم المفتوحة؟



مبادئ العلوم المفتوحة
(Open Science practices)

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



@OpenSciSaudi

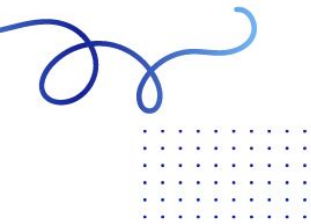


<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>





■ تعريف بالمجتمعات العلمية والمُنظمين



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>



- OSCSA aims to **provide a place** where newcomers and experienced peers interact, inspire each other to embed OS practices and values in their workflows and **provide feedback** on infrastructures, and support services.

- Our community is part of the International Network of Open Science & Scholarship Communities (INOSC).

- It was established inline with vision 2030.



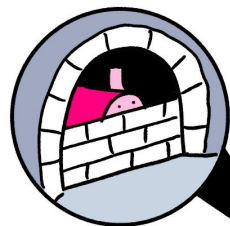
The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807)



Get Involved!

All our materials are open through GitHub and Zenodo.

- Share your ideas
- Re-use our resources for no-cost
- Contribute to our resources
- Promote OSCSA in your community
- Join OSCSA

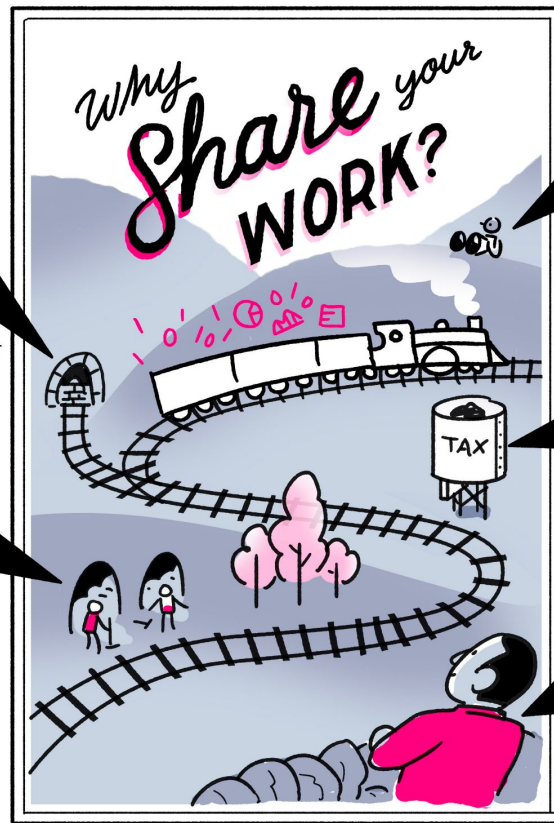


RESEARCH IS USELESS IF IT SITS BEHIND A PAYWALL



AVOID DUPLICATION OF EFFORT AND FUNDING

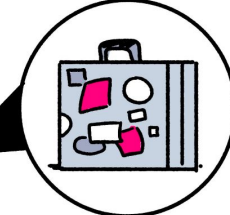
Scriberia



OPEN SOURCE WORK IS MORE LIKELY TO BE REVIEWED AND REUSED



PUBLICLY FUNDED WORK SHOULD BE PUBLICLY AVAILABLE



BENEFITS FOR AUTHORS INCLUDING MORE CITATIONS

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



<https://github.com/Open-Science-Community-Saudi-Arabia>



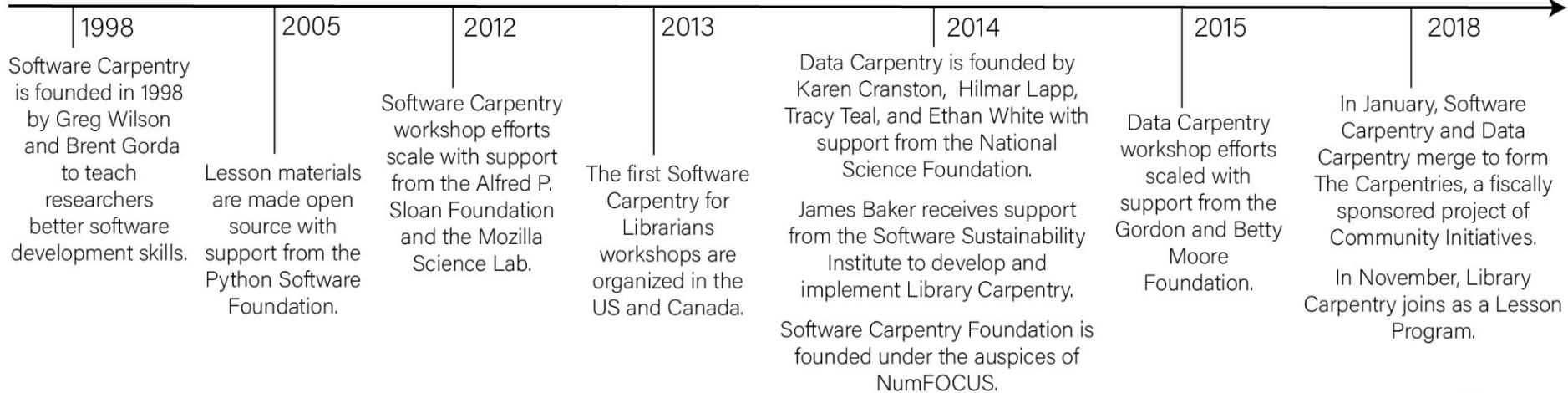
<https://osc-ksa.com/>



A VERY BRIEF HISTORY OF



THE CARPENTRIES



Software Sustainability Institute

GORDON AND BETTY MOORE FOUNDATION

NUMFOCUS

COMMUNITY INITIATIVES
in service to great ideas

Batool Almarzouq



Computational Biologist

Honorary research Fellow

Founder



UNIVERSITY OF
LIVERPOOL



Education Committee

Core Contributor

Global Team

Subject Matter Experts (SME)



@OpenSciSaudi




<https://github.com/Open-Science-Community-Saudi-Arabia>




<https://osc-ksa.com/>






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






Annajiat is a Sr. researcher at the Department of Computer Science and Engineering (CSE) under School of Data and Sciences (SDS), Brac University, Dhaka, Bangladesh.



He usually teaches problem-solving skills using **Java and Python** while trying to explore the exciting **fields of NLP and Parallel, Distributed, and HPC**. He has been a part of different Carpentries workshops and is trying to gain more experience. He is excited to be a part of this workshop as an additional learning opportunity.



Annajiat Alim Rasel



Email: annajiat@gmail.com



LinkedIn: Annajiat Alim Rasel



Twitter: @annajiat



GitHub: @annajiat



Website: <http://annajiat.googlepages.com>



@OpenSciSaudi






<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>





 Email: dallak-99@hotmail.com
 Twitter: [@amdallak](https://twitter.com/amdallak)
 Github: [@Dallak](https://github.com/Dallak)



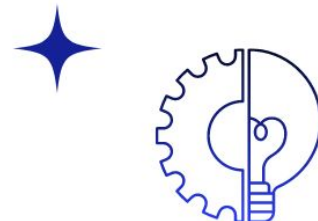
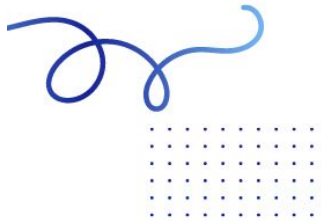
Abdulrahman Dallak is a 4th year PhD student at Newcastle University – UK. He is a phonetician working on speech-related data where he investigates how people produce and perceive Arabic speech sounds.



He has volunteered for teaching data science and statistics for MA and PhD students at Newcastle University, and has **7 years of experience in R**. His workflow includes **Python, Matlab, and Shell** but mostly works on speech-related data. Beyond frequentist statistics, he has an experience in **Bayesian analysis, Generalised Additive Mixed Models (GAMMs), and estimation statistics**.

He is a lecturer in the English Language Institute at Jazan University - Saudi Arabia.





Mona Al-Sharif

Lecturer at King Faisal University, department of Agricultural Biotechnology, KSA

Master of Science degrees in Plant Biotechnology, Leeds, UK

Research experience: **Plant DNA damage and repair**

✉ Email: malsharef@kfu.edu.sa

🐙 Github: Mona-2022



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>





Feedback



@OpenSciSaudi



<https://github.com/Open-Science-Community-Saudi-Arabia>



<https://osc-ksa.com/>

