St. Jude Children's Research Hospital | DNB Bioinformatics core

Automation and Reproducibility in Computational Biology: "Best practices"

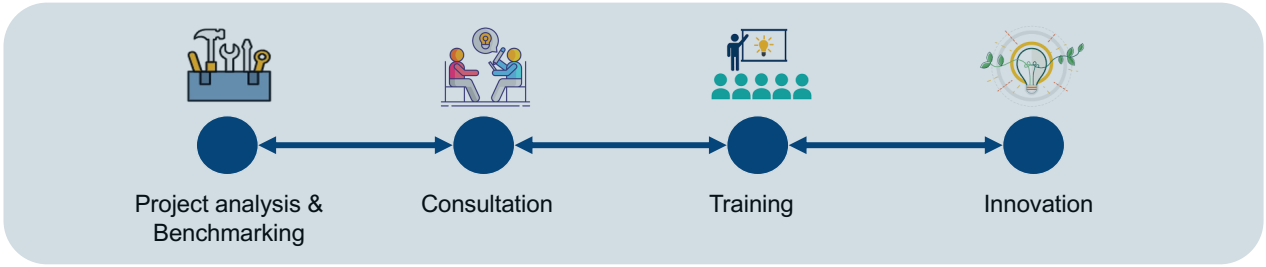
Antonia (Tonia) Chroni, PhD
Senior Bioinformatics Research Scientist
DNB Bioinformatics Core

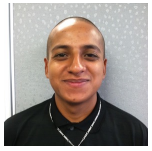
October 10, 2024

Bioinformatics core, Department of Developmental Neurobiology

GitHub

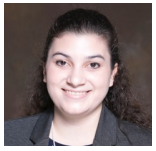
Providing advanced bioinformatic services for investigators to leverage omics data





Cody Alexander Ramirez, PhD
Senior Bioinformatics Research Scientist
Core Director
Boston, Massachusetts

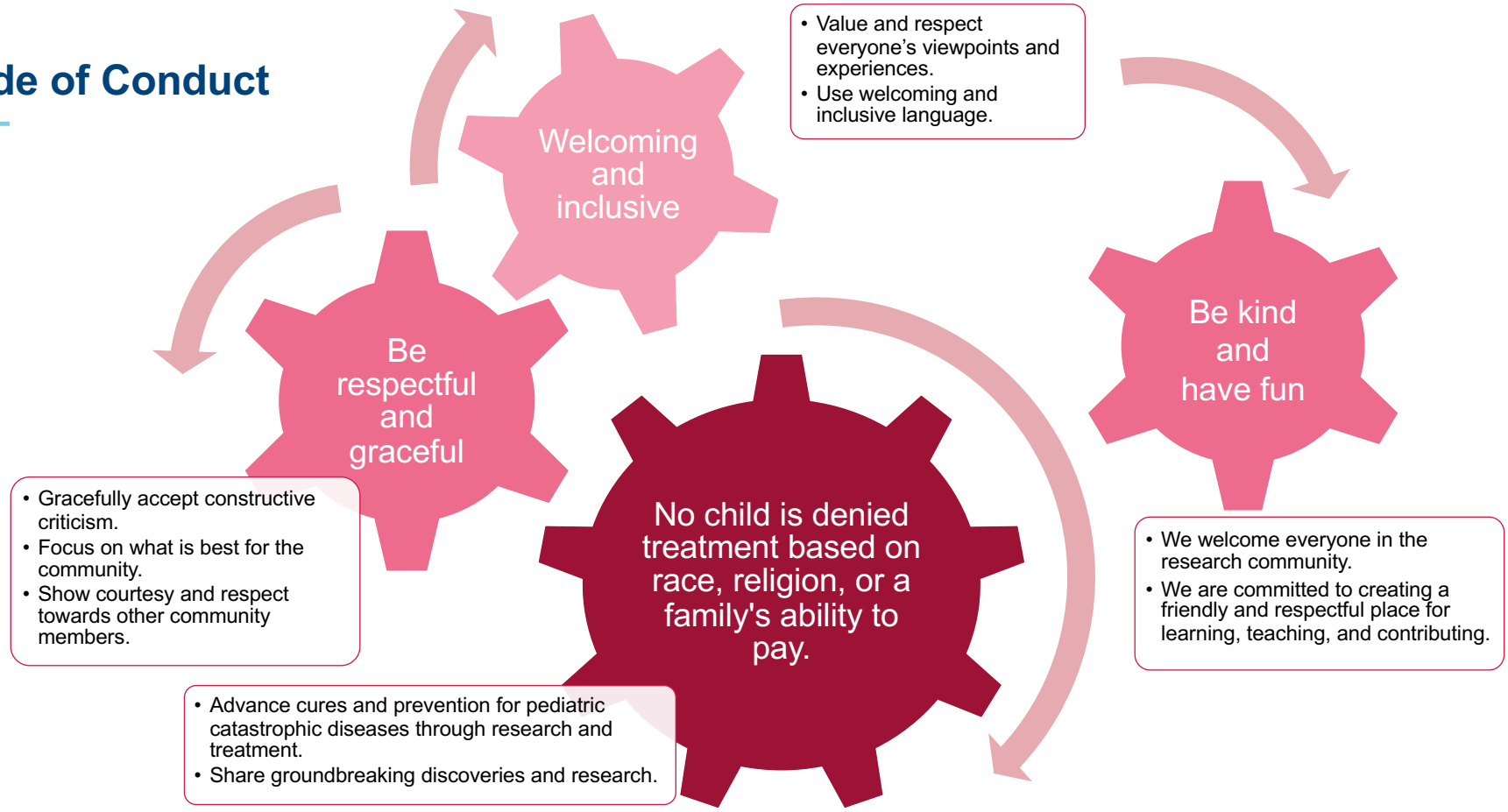
Antonia Chroni, PhD
Senior Bioinformatics Research Scientist
New York, New York

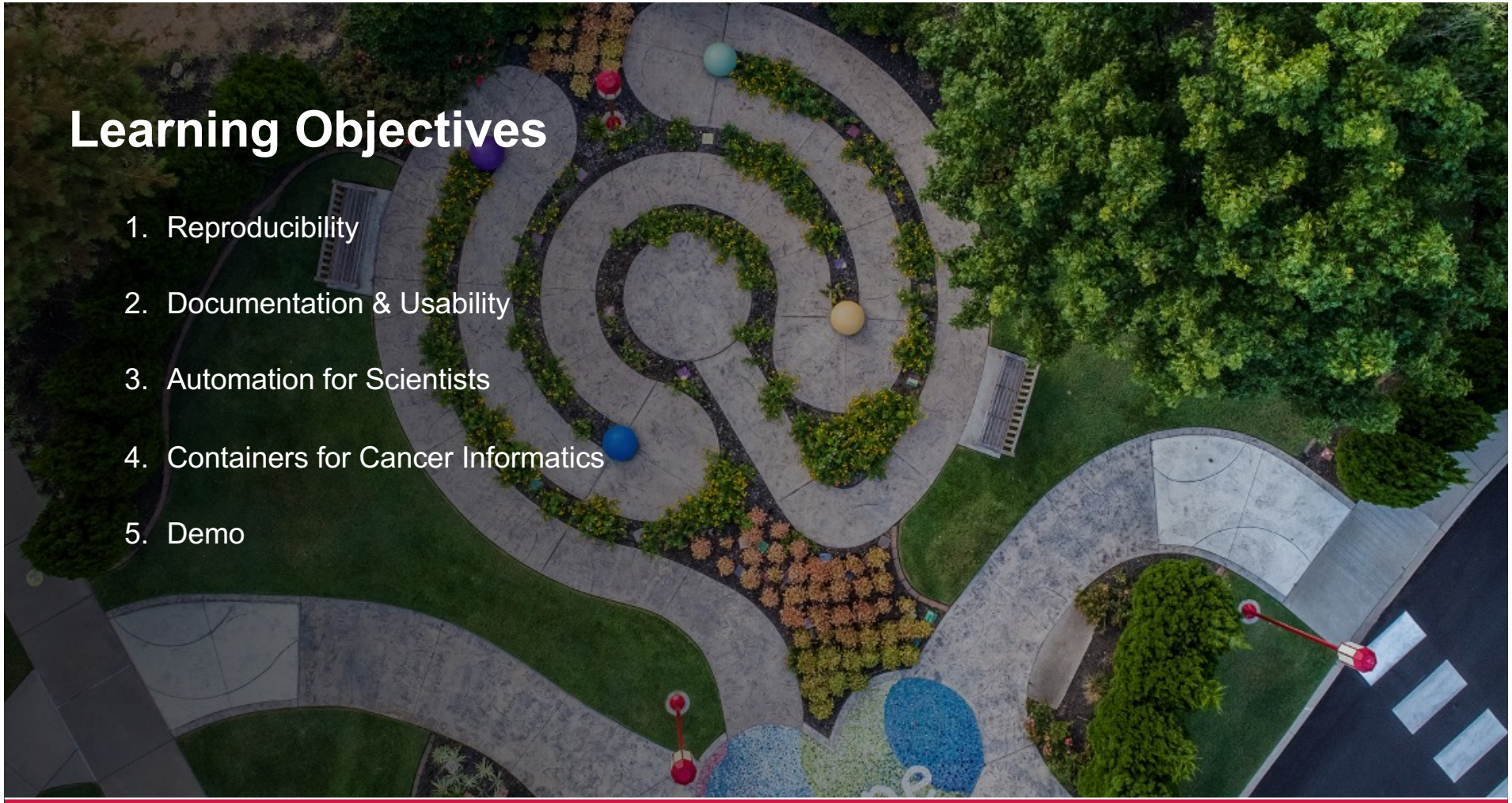


Sharon Freshour, PhD
Bioinformatics Research Scientist
St. Louis, Missouri

Asha Jacob Jannu, PhD
Bioinformatics Research Scientist
Indianapolis, Indiana

Code of Conduct

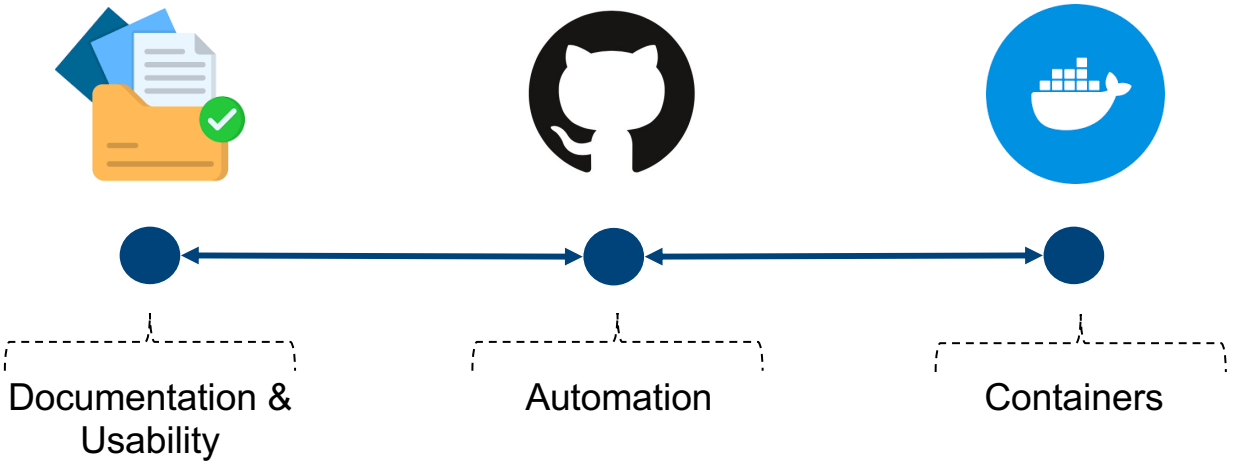




Learning Objectives

1. Reproducibility
2. Documentation & Usability
3. Automation for Scientists
4. Containers for Cancer Informatics
5. Demo

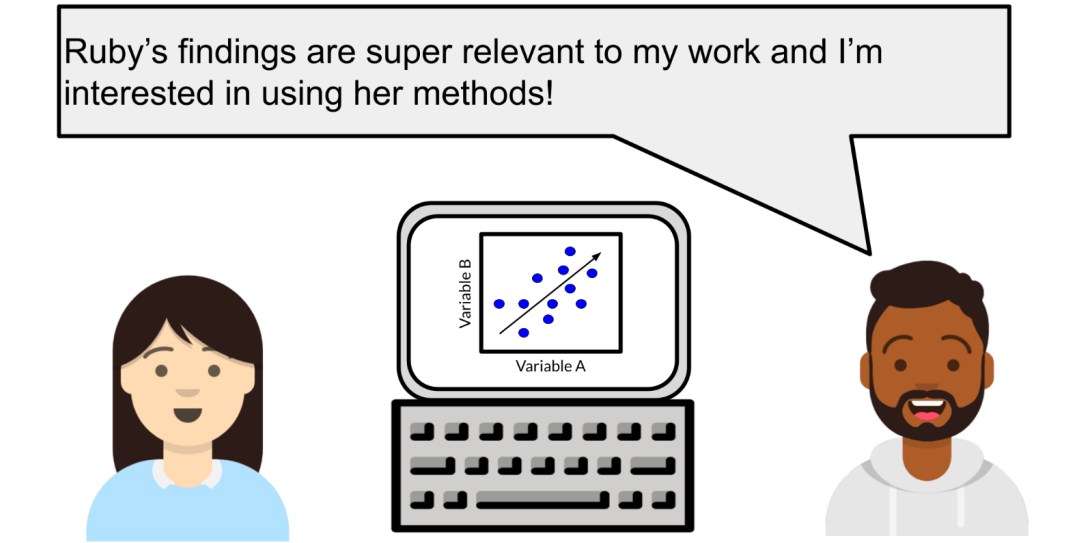
Best practices for reproducibility 🎉



Have you ever had problems reproducing…

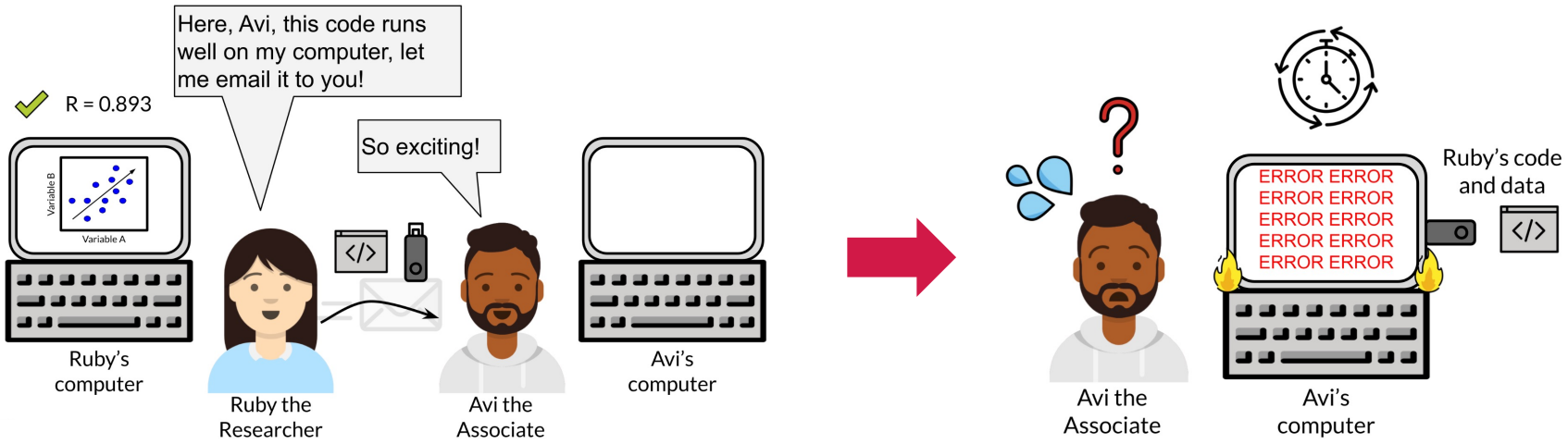
- Someone else's research?
- Your research?
- Both?

Reproducibility



Source by Advanced Reproducibility in Cancer Informatics

Reproducibility



Source by Advanced Reproducibility in Cancer Informatics

Reproducibility vs replication

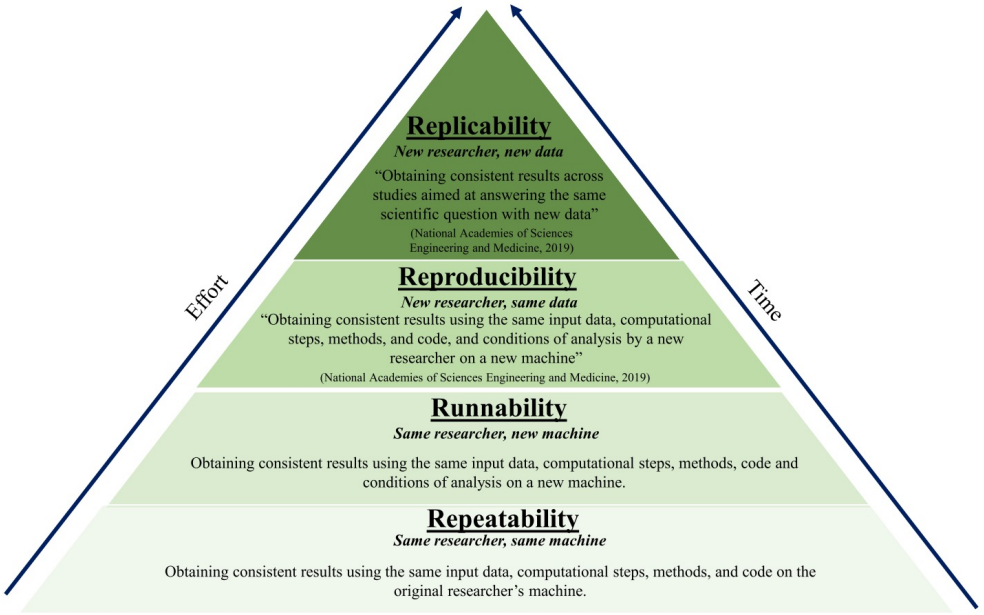
- Reproducibility
  - Authors provide all the necessary data and computer code to re-run the analysis and re-create the results
  - The exact same data/code are used to re-derive the exact same results
- Replication
  - A separate study arrives at the same scientific findings as another study
  - New data/code and analyses are performed that identify consistent results with previous work

	Same data	**Different data**
Same methods	Reproducibility	Replicability
Different methods	Robustness	Generalizability

Source by Data Lab Reproducibility Workshop

Why do we like reproducible science?

- Reproducibility supports…
  - You!
  - Your collaborators and team!
  - Your community!
  - The scientific endeavor!
- Reproducibility makes your funders and journals happy.



Essawy *et al.*, 2020. Environmental Modelling & Software.

Source by Data Lab Reproducibility Workshop

Reproducibility is a tortoise's game – it is an incremental and slow process, but it has high payoffs!

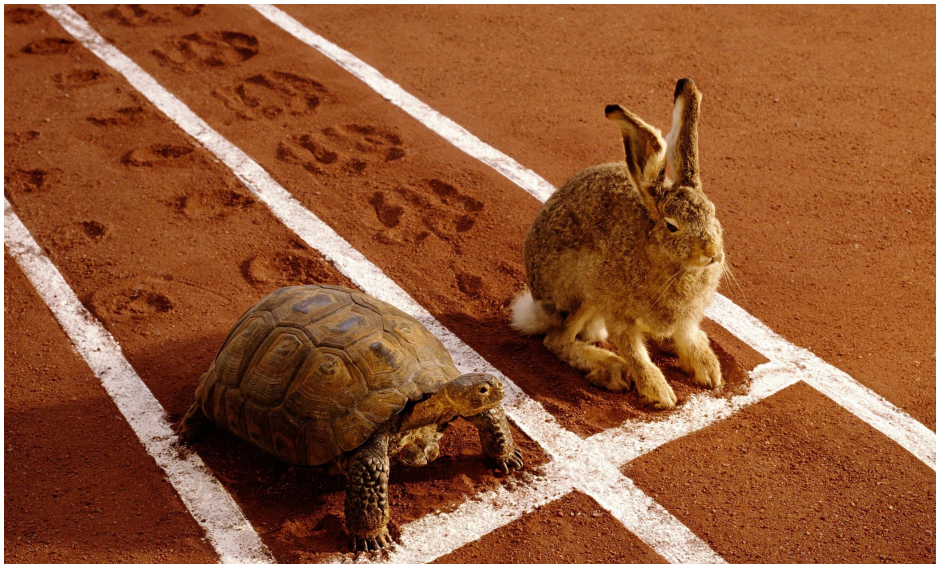
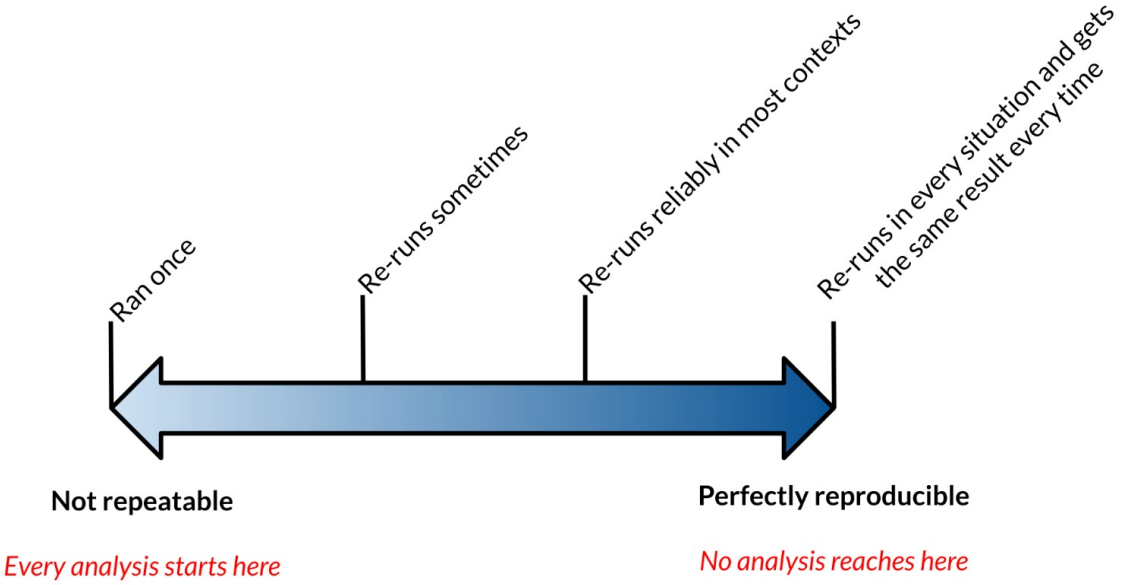


Figure from https://edwinsjournal.com/the-rabbit-and-the-tortoise-reimagined/

Reproducibility exists on a continuum!



Source by Advanced Reproducibility in Cancer Informatics

More resources 🎉

- Advanced Reproducibility in Cancer Informatics
- Elements of Style
- Building reproducible analytic pipelines in R
- Working reproducibly with others on OpenScPCA