

Ciência Reprodutível para Experimentos em Computação de Alto Desempenho

Pedro Bruel, Lucas Schnorr, Alfredo Goldman

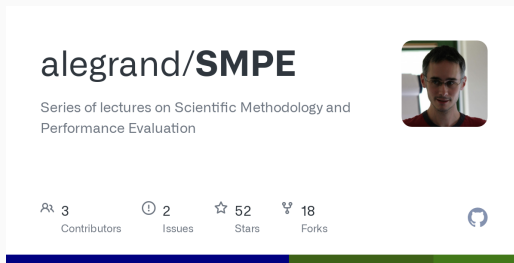
phrb@ime.usp.br

1 de maio de 2021

Introdução

Agradecimentos e Crédito

A **Arnaud Legrand** e seu curso:



<https://github.com/alegrand/SMPE>



Dependências e outros Recursos

- Site com **instruções** e mais **recursos**:

Ciência Reprodutível - ERADSP 2021

Ciência Reprodutível para Experimentos em Computação de Alto Desempenho (ERADSP 2021)

[View the Project on GitHub](#)
phrb/reprodutibilidade-eradsp-2021

Ciência Reprodutível para Experimentos em Computação de Alto Desempenho

Este é o site do minicurso apresentado na ERADSP 2021.

Conteúdo

- [Exercício Prático](#)
- [Recursos](#)

<https://phrb.github.io/reprodutibilidade-eradsp-2021>

- Temos uma imagem **Docker** com Jupyter Notebook, R, pacotes, e dados:

```
git clone https://github.com/phrb/reprodutibilidade-eradsp-2021.git
cd reprodutibilidade-eradsp-2021/exercicio_pratico && ./build.sh -b
```

- Jupyter Docker Stacks: github.com/jupyter/docker-stacks

Roteiro

O que é Ciência Reprodutível?

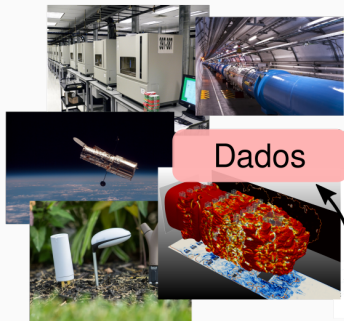
Desafios para se fazer Ciência Reprodutível

Mão na Massa: Ferramentas para Reprodutibilidade

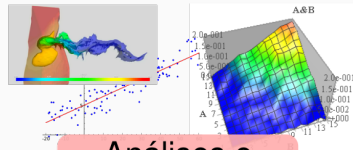
O que é Ciência Reprodutível?

O que sobra do Trabalho Científico?

Quem Produz



Dados



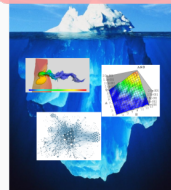
Análises e Visualizações



Quem Lê

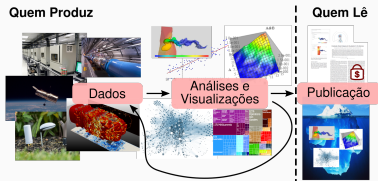


Publicação



O que é Ciência Reprodutível?

Trabalhar de forma **transparente** para diminuir a distância entre **quem produz** e **quem lê**



Trabalhar de forma transparente?

- Caderno de laboratório e metodologia
- Ambientes de software, controle de versão
- Plataformas de compartilhamento, colaboração, e arquivamento

Definições

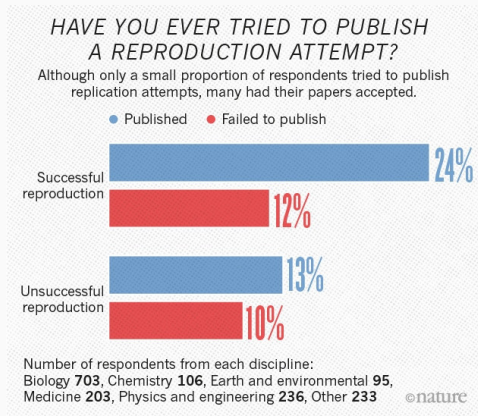
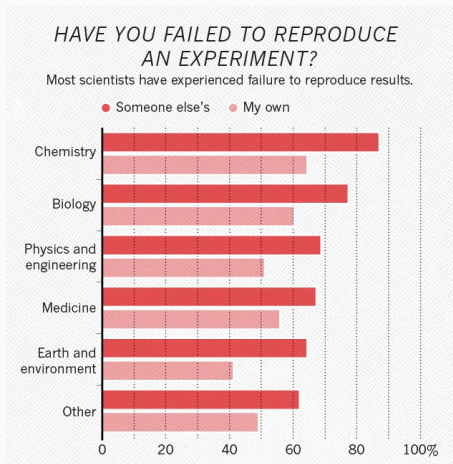
- **Vocabulário Internacional de Metrologia (VIM)**

Distingue entre **resultados** e **conclusões** que podem ser reproduzidos:

- Pela mesma equipe, nas mesmas condições experimentais: *Repetibilidade*
- Por uma equipe diferente, nas mesmas condições experimentais: *Replicabilidade*
- Por uma equipe diferente, em condições experimentais diferentes: **Reprodutibilidade**

Há uma Crise de Reprodutibilidade?

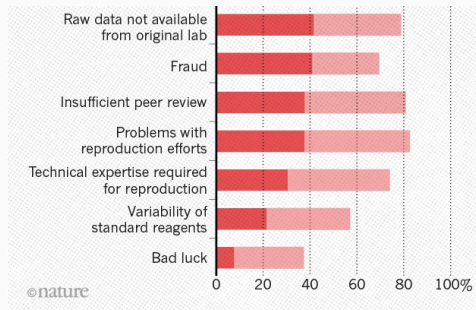
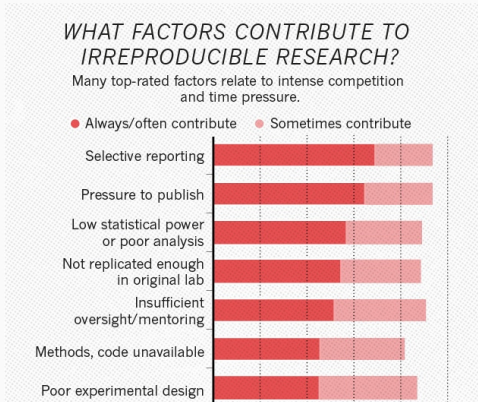
Resultados de um questionário com 1.500 cientistas:



(1,500 Scientists Lift the Lid on Reproducibility, Nature, Maio de 2016)

O que dificulta a Reprodutibilidade?

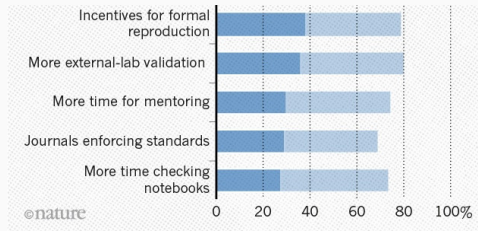
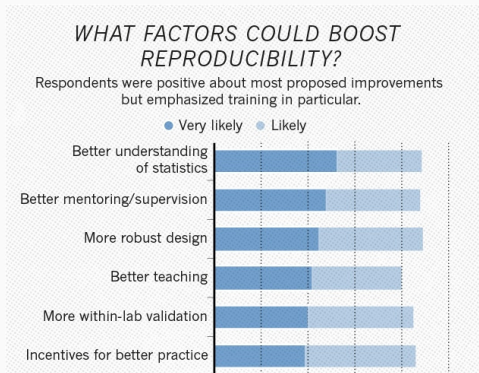
Resultados de um questionário com 1.500 cientistas:



(1,500 Scientists Lift the Lid on Reproducibility, Nature, Maio de 2016)

O que pode promover a Reprodutibilidade?

Resultados de um questionário com 1.500 cientistas:



(1,500 Scientists Lift the Lid on Reproducibility, Nature, Maio de 2016)

Abordagens da ACM

Desafios para se fazer Ciência Reprodutível

Computação de Alto Desempenho: Desafios

- Arquiteturas Específicas
 - Custo elevado
 - Como acessar?
- Software complexo
 - Dependências
- Consumo de Energia

Computação de Alto Desempenho: Abordagens

- Virtualização
- Controle de versão
- Hospedagem pública
- Gerenciadores de dependências
- Simuladores
- Hardware aberto
- Nuvem?

Ciência Fechada: Desafios

- Como acessar publicações?
- Como compartilhar resultados?
- Software fechado
- Custo para acessar arquivos
- Análises de dados, figuras: código

Ciência Fechada: Abordagens

- Ciência aberta
- Revisão por pares aberta
- Documentos computacionais: software livre
- Arquivamento
- Gerenciadores de dependências
- Hospedagem pública
- Arnaud: Changing academic practices
- Arnaud: What About Open Science?

Estatística: Desafios

- Datasaurus Dozen
 - <https://cran.r-project.org/web/packages/datasauRus/>
- O que é incerteza?
- Como planejar experimentos?
- Como analisar resultados?
 - O que é Machine Learning?

Estatística: Abordagens

- Documentos computacionais
- Desenho de Experimentos
- Controle de versão
- Gráficos, antes de qualquer análise
- Fazer análises mais simples primeiro: mas fáceis de interpretar

É Possível Garantir a Reprodutibilidade?

Mão na Massa: Ferramentas para Reprodutibilidade

Controle de Versão

Hospedagem

Arquivamento

Conclusão

É possível fazer Ciência (mais) Reprodutível!

Ciência Reprodutível para Experimentos em Computação de Alto Desempenho

Pedro Bruel, Lucas Schnorr, Alfredo Goldman

phrb@ime.usp.br

1 de maio de 2021