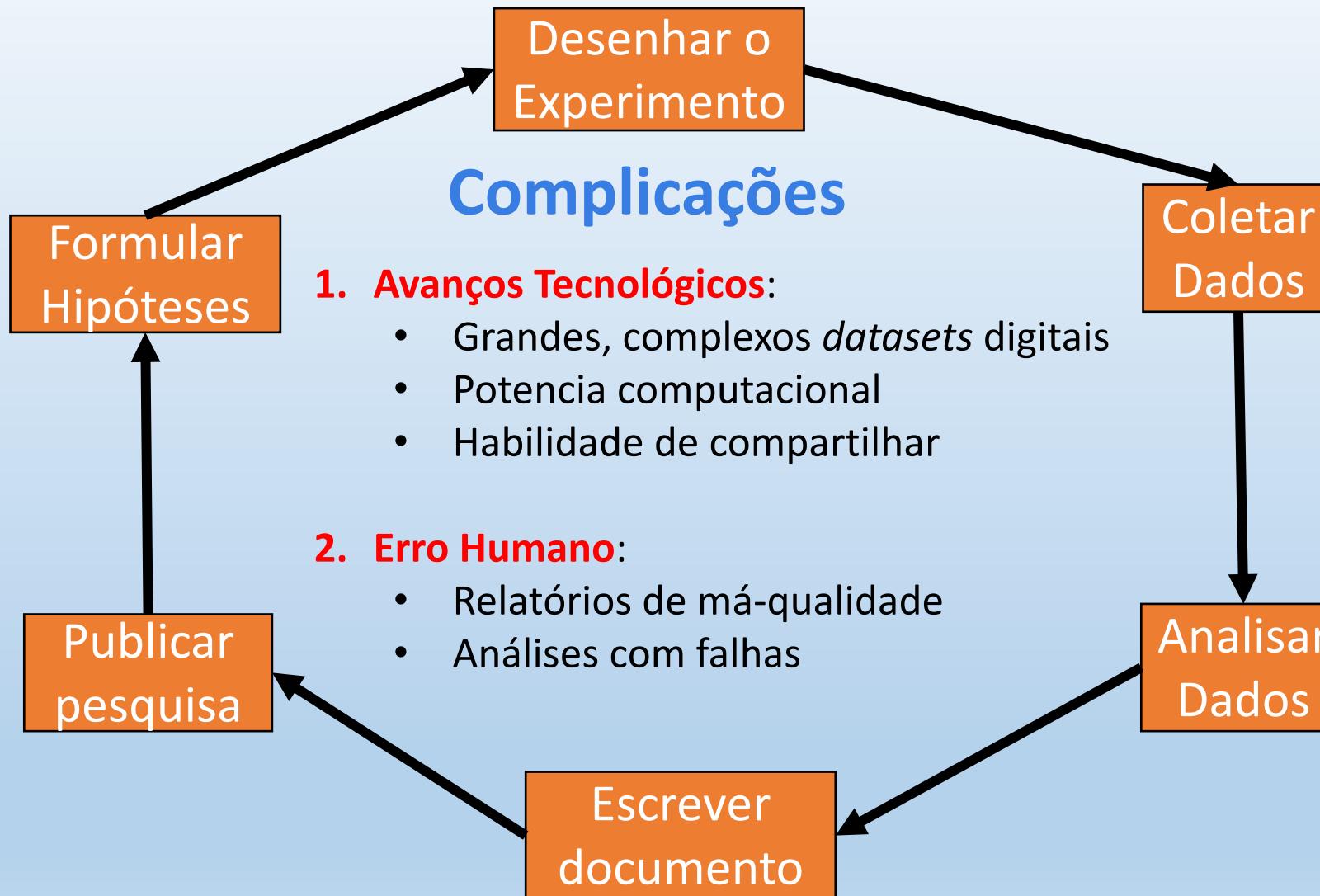


Pesquisa Reprodutível: Teoria

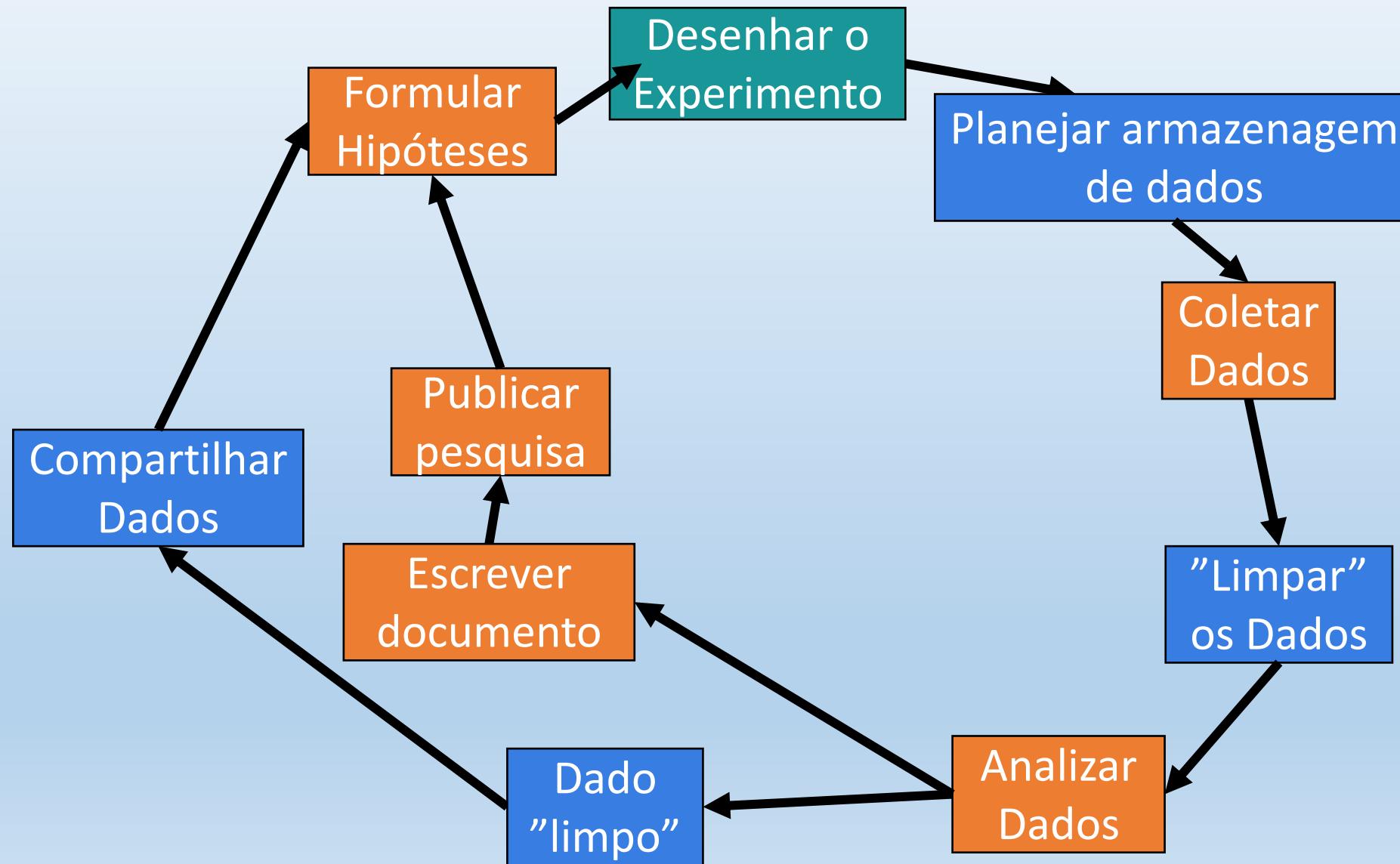
Tobin Magle, PhD
Bioinformationist
Health Science Library
University of Colorado Anschutz
Medical Campus



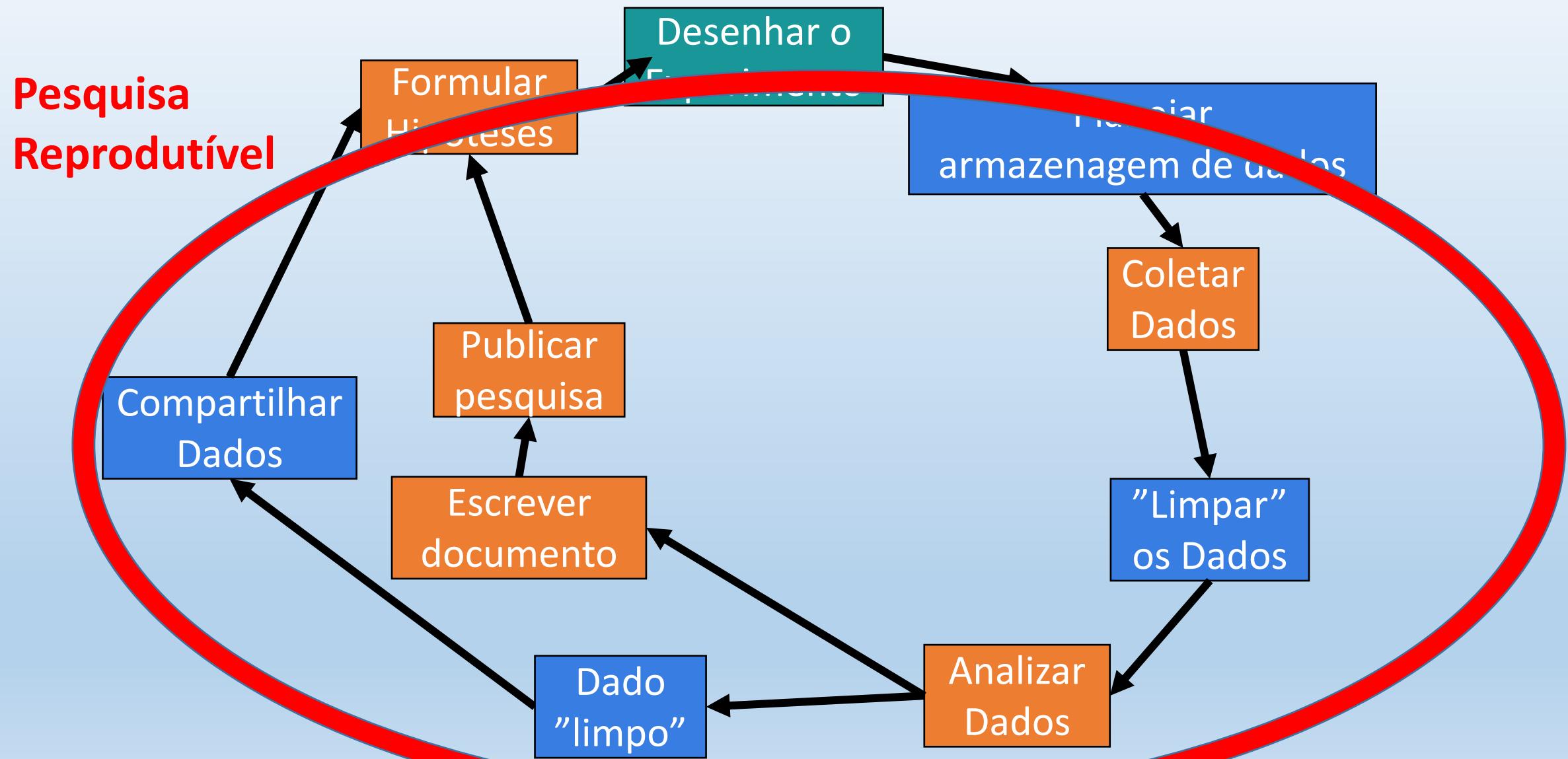
Ciclo de vida de pesquisa:



Ciclo de vida de pesquisa complexo



Ciclo de vida de pesquisa complexo



Reprodutibilidade

é a prática de distribuir todos os dados, o código-fonte do software e as ferramentas necessárias para reproduzir os resultados discutidos em uma publicação de pesquisa.

<https://www.ctspedia.org/do/view/CTSpedia/ReproducibleResearchStandards>

reprodutibilidade

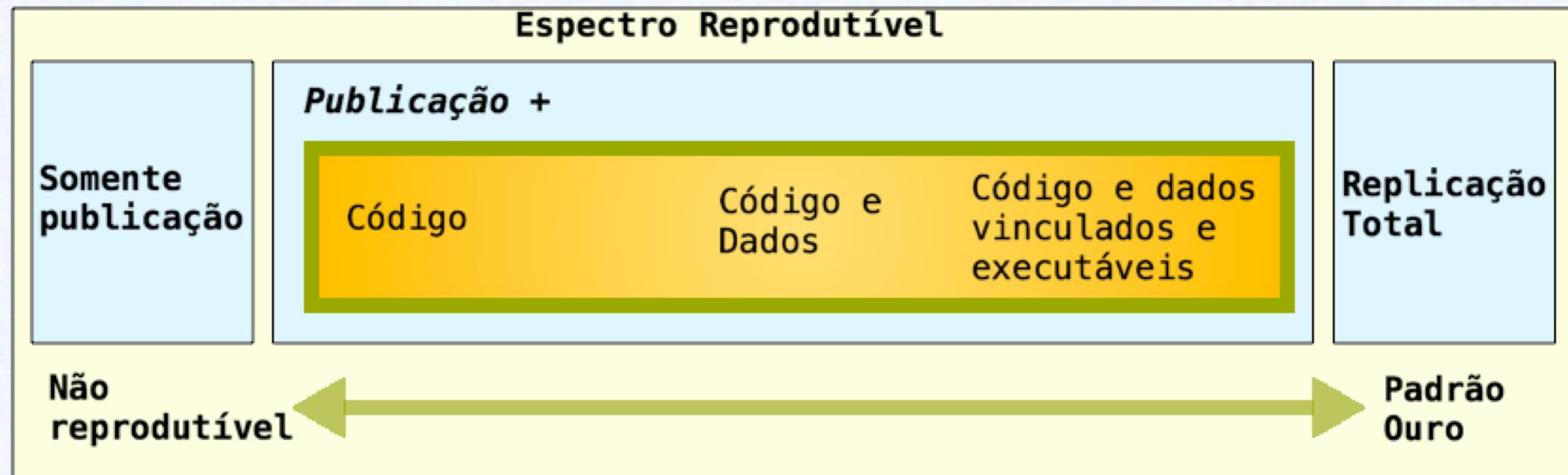
Qualidade de reproduzível, que se pode reproduzir; que se pode exibir ou mostrar novamente.

Taxa de reprodução; velocidade com que algo pode ser produzido inúmeras vezes a partir de uma mesma matriz.

Replicação versus Reprodutibilidade

- **Replicação:** A confirmação dos resultados e conclusões de um estudo, obtido de forma independente em outro, é considerado o padrão ouro científico.
 - “Again, and Again, and Again ...” BR Jasny et. al. Science, 2011. 334(6060) pp. 1225 DOI: 10.1126/science.334.6060.1225
- **Alguns estudos não podem ser replicados:** muito grandes, muito dispendiosos, muito demorados, um evento único, amostras raras...
- **Reprodutibilidade:** padrão mínimo para avaliar o valor das afirmações científicas, particularmente quando a replicação totalmente independente de um estudo não é viável
 - “Reproducible Research in Computational Science”. RD Peng Science, 2011. 334 (6060) pp. 1226-1227 DOI: 10.1126/science.1213847

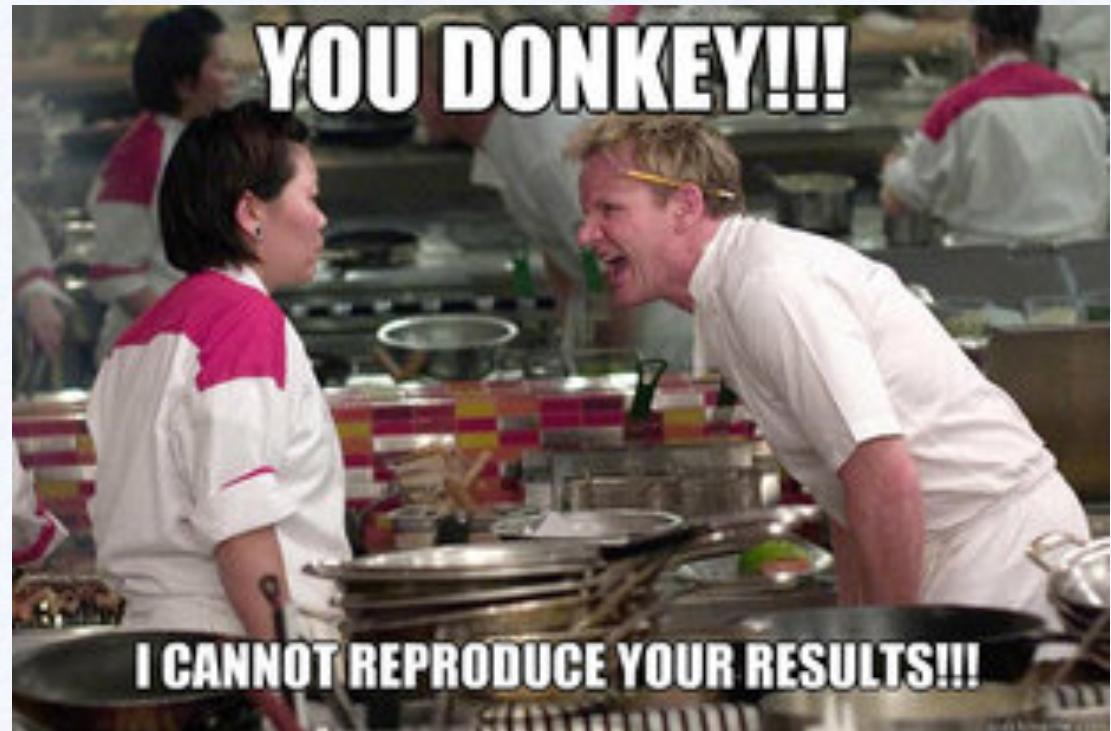
Espectro de reproduzibilidade



"Reproducible Research in Computational Science". RD Peng Science, 2011. 334 (6060) pp. 1226-1227 DOI: 10.1126/science.1213847

Por que fazer pesquisa reproduzível?

- **Bom para o público:** transparência
- **Bom para você:** você é o futuro usuário de seus dados
- **Está começando a ser recomendado** (Journal of Biostatistics, The FAIR Data Principles)
- **Provavelmente será exigido em breve**



<http://campus.murraystate.edu/academic/faculty/cmecklin/RWebpage.html>

É bom para o público...

- Falta de transparência atualmente:
 - Somente 1 de 441 papers forneceram “protocolo completo”, 0 tinham dados
 - Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA (2016) Reproducible Research Practices and Transparency across the Biomedical Literature. PLoS Biol 14(1): e1002333. doi:10.1371/journal.pbio.1002333
- Interrompem ensaios clínicos potencialmente prejudiciais

TREATMENTS

f

Scientists Question Cancer Gene Trials At Duke University

July 20, 2010 · 6:31 AM ET

g+

GEOFFREY BRUMFIEL

Full lecture by Keith Baggerly, Bioinformatician (University of Texas, MD Anderson Cancer Center)
<https://www.youtube.com/watch?v=7gYIs7uYbMo>

<http://www.npr.org/sections/health-shots/2010/07/19/128630549/scientists-question-cancer-gene-trials-at-duke-university>

É bom pra você...:

- Você é o futuro usuário dos seus dados
 - Controle de versão – reverter para versões anteriores
 - Economize tempo ao escrever suas seções de métodos

Começam a ser recomendados...

SCIENTIFIC DATA | COMMENT [OPEN](#)



The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

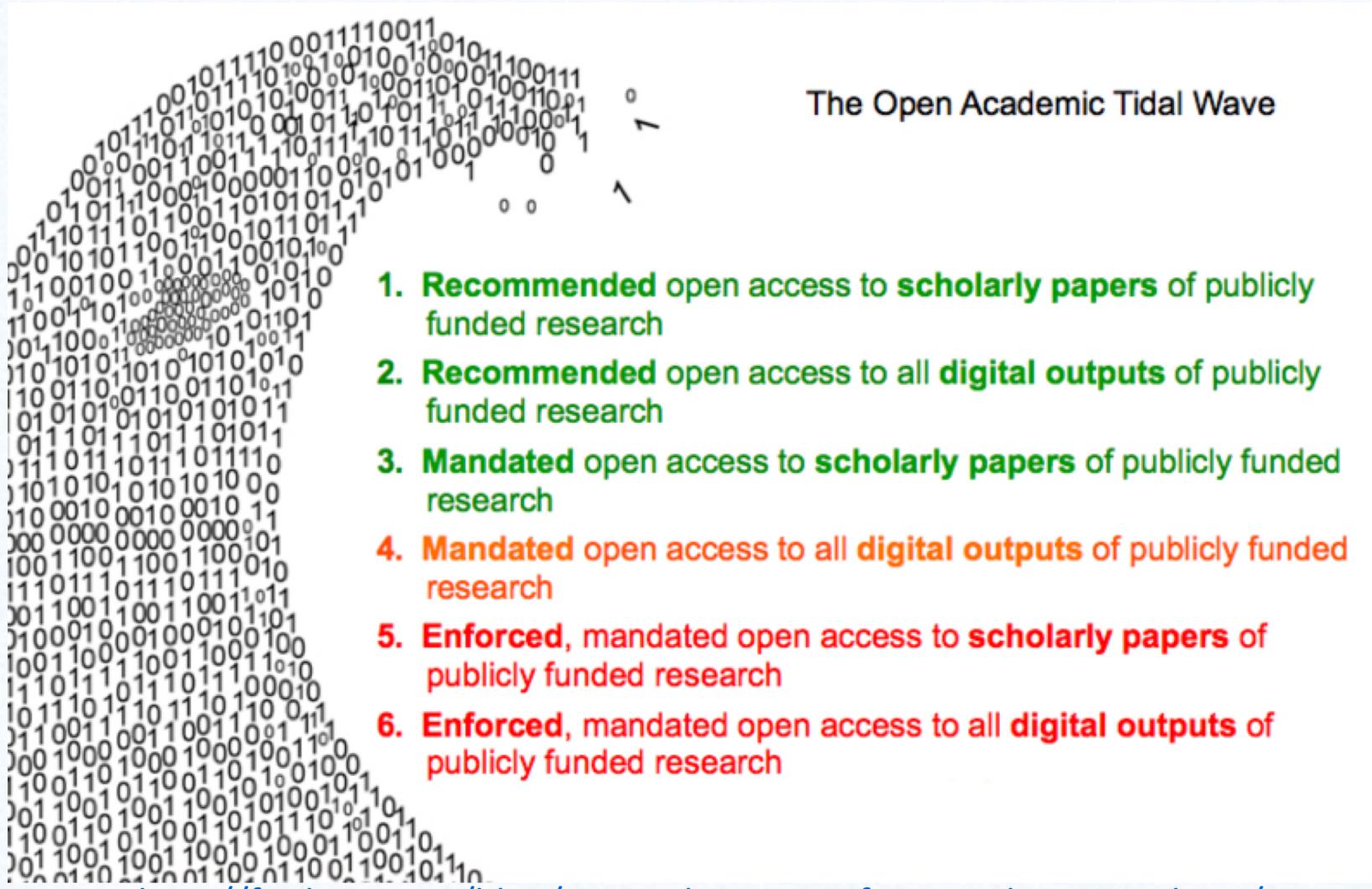
Scientific Data 3, Article number: 160018 (2016) | doi:10.1038/sdata.2016.18

Received 10 December 2015 | Accepted 12 February 2016 | Published online 15 March 2016

Editorial

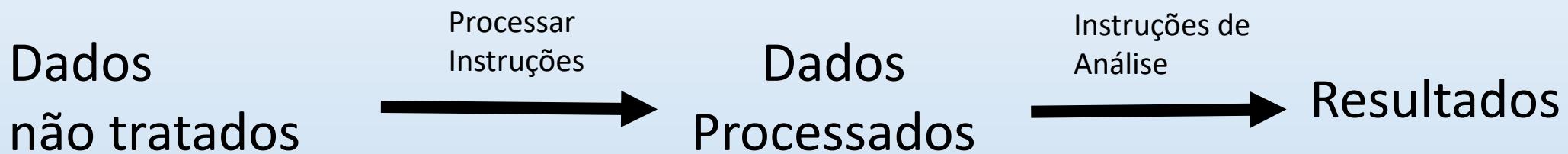
As coeditors of *Biostatistics*, we wish to encourage the practice of making research published in the journal reproducible by others. The following invited piece by Roger Peng sets out our policy on this; Roger will be assuming the role of Associate Editor for reproducibility as set out in his piece.

Será exigido em breve

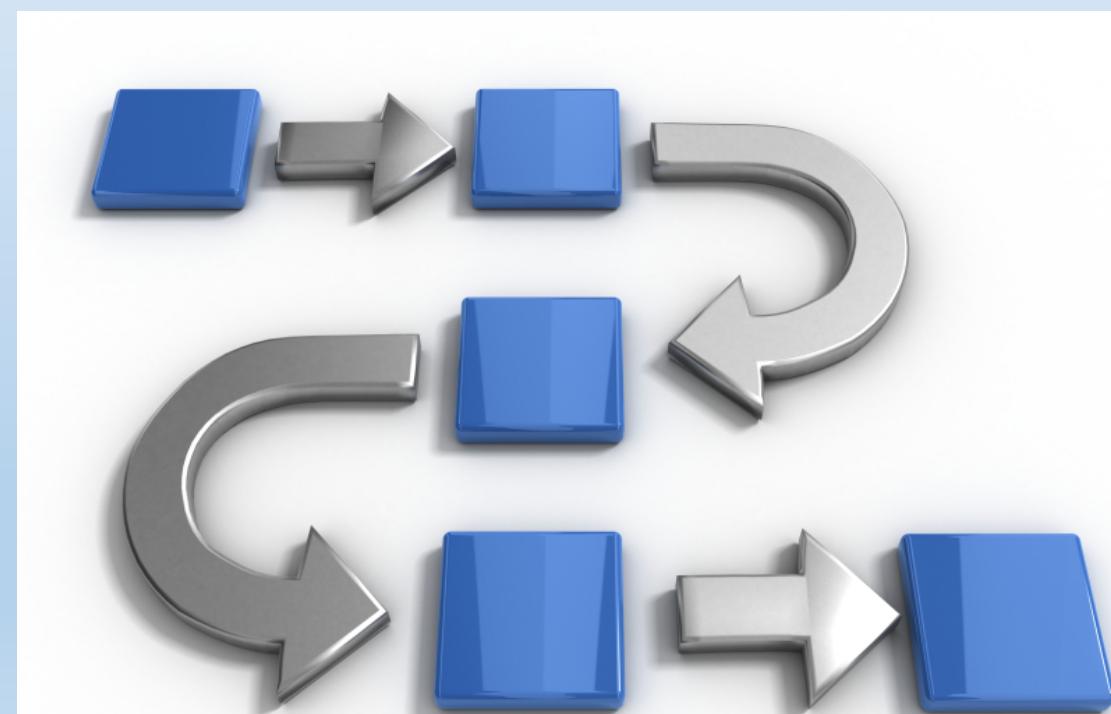


1. Recommended open access to **scholarly papers** of publicly funded research
2. Recommended open access to all **digital outputs** of publicly funded research
3. Mandated open access to **scholarly papers** of publicly funded research
4. Mandated open access to all **digital outputs** of publicly funded research
5. Enforced, mandated open access to **scholarly papers** of publicly funded research
6. Enforced, mandated open access to all **digital outputs** of publicly funded research

Workflow / Pipeline / Linha de produção



- **Ótimo:** inclui todo o fluxo de trabalho
- **Mínimo:** Dados processados para os resultados
- **Ótimo:** as instruções devem ser um arquivo de script automatizado (ou seja, "código")
- **Mínimo:** instruções que permitem a reprodução completa de sua análise



Reprodutibilidade =

Dados

+

Código Fonte/Ferramentas

Tipos de Dado

- **Dados brutos** – o que é registrado
 - Exportado de Ressonância Magnética
 - Leituras de máquinas
 - Resultados de surveys (pesquisas)
- **Dados tratados** – “limpos”
 - Prontos para análises
 - Formatado adequadamente

Aderentes aos
FAIR Data
Principles

<https://www.force11.org/group/fairgroup/fairprinciples>

FAIR Data Principles

A screenshot of a Firefox browser window. The title bar shows 'Firefox' and various tabs. The main content area displays the FORCE11 website. The header features the 'FORCE11' logo and navigation links for 'ABOUT', 'COMMUNITY', 'GROUPS', 'RESOURCES', 'NEWS + BLOGS', 'EVENTS', 'PUBLICATIONS', 'MEDIA', and 'DONATE'. A search bar is also present. The main section is titled 'THE FAIR DATA PRINCIPLES' and includes a 'Preamble' and sections for 'TO BE FINDABLE' and 'TO BE ACCESSIBLE'. A 'BACK TO TOP' button is visible in the bottom right corner.

THE FAIR DATA PRINCIPLES

Preamble

One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows. Here, we describe FAIR - a set of guiding principles to make data **Findable, Accessible, Interoperable, and Re-usable**. The FAIR principles have now been [published](#).

TO BE FINDABLE:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

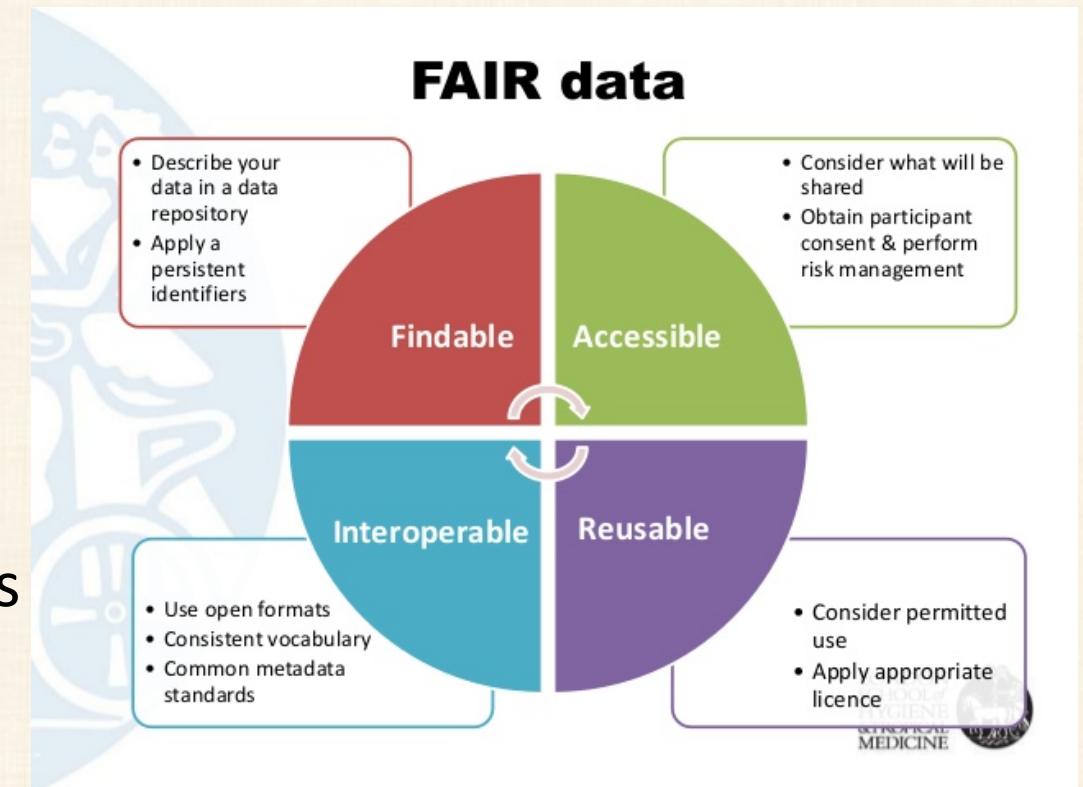
TO BE ACCESSIBLE:

<https://www.force11.org/group/fairgroup/fairprinciples>

FAIR principles

- **Encontrável:** URLs, bons metadados, em um endereço/local pesquisável
- **Acessível:** acesso rápido aos metadados (e também aos dados)
- **Interoperável:** uso de terminologias bem estabelecidas, sempre que possível
- **Reutilizável:** bons metadados, licença de uso, procedência, atende aos padrões comunitários estabelecidos

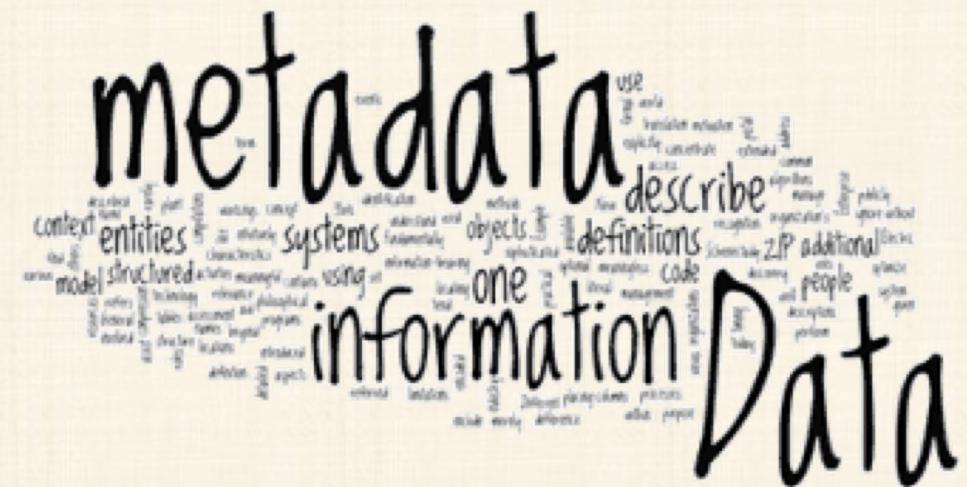
<http://www.nature.com/articles/sdata201618>



<https://pt.slideshare.net/lshtm/preparing-data-for-sharing-the-fair-principles>

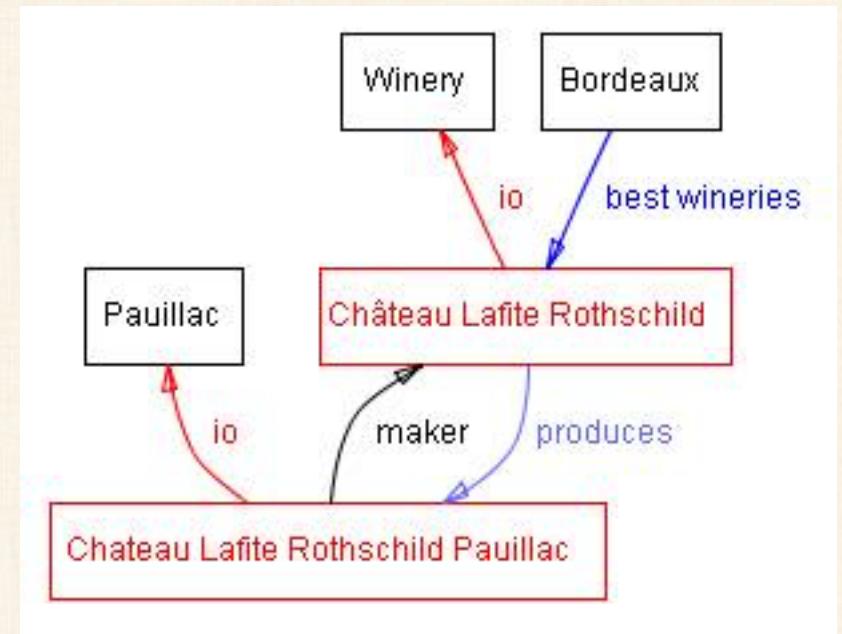
Metadata / Metadado

- Metadata descreve os dados da pesquisa:
 - Format depends on what type of data you have
- Dublin Core: <http://dublincore.org/documents/dcmi-terms/>
 - Pode ser aplicado em qualquer área
- Existem vários padrões de metadados
 - INDA: <http://wiki.dados.gov.br/Padroes-de-metadados.ashx>
 - ePMG: https://www.governoeletronico.gov.br/documentos-e-arquivos/PMG%20Versao%201_1.pdf
 - OBAA - <http://www.portalobaa.org/padrao-obaa>
 - INDE: <http://www.metadados.inde.gov.br/geonetwork/srv/por/main.home>
 - Search for other standards: <https://biosharing.org/standards/>



Terminologias estabelecidas

- Linguagem utilizada para descrever os dados
- **Exemplos:**
 - Airport codes
 - MeSH: <http://www.ncbi.nlm.nih.gov/mesh>
- **Ontologias:**
 - Técnica de organização de informações
 - OntoBacen: Ontologia para Gestão de Riscos
 - Gene Ontology: <http://geneontology.org>
 - GT Ontologias: <http://www.w3c.br/GT/GrupoOntologias>
 - SNOMED:
<http://bioportal.bioontology.org/ontologies/SNOMEDCT>
 - Busca por ontologias relevantes:
[https://pt.wikipedia.org/wiki/Ontologia_\(ci%C3%A3ncia_da_computa%C3%A7%C3%A3o\)](https://pt.wikipedia.org/wiki/Ontologia_(ci%C3%A3ncia_da_computa%C3%A7%C3%A3o))



http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

Formatos de arquivos

Formatos proprietários

- .doc, docx
- .ppt, pptx
- .xls, .xlsx

Formatos não proprietários*

- .txt
- .jpeg
- .tiff
- .CSV

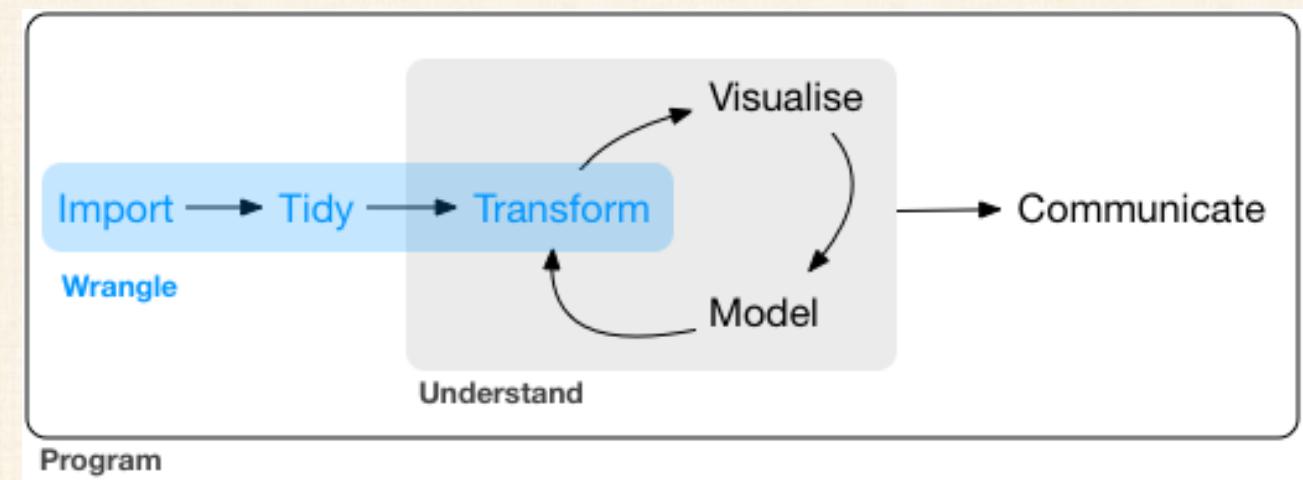
Outros formatos

- **JSON** (JavaScript Object Notation - Notação de Objetos JavaScript)
- **XML** (eXtensible Markup Language)
- **RDF** Resource Description Framework

* Os formatos de arquivo não proprietários são os mais adequados para garantir acesso aos dados no futuro!

Software/Ferramentas

- Como ir dos dados *brutos* para os dados *tratados*?
- Qual software utilizar?
 - O software possui arquivos de log / scripts?
 - Inclui informações e configurações de versões
 - O que o software precisa para "rodar"?
 - Arquitetura do computador
 - SO/Software/ferramentas/complementos (add ons – bibliotecas, pacotes)
 - Bancos de dados externos
- Automatizar o máximo possível
 - Escrever scripts ou salvar arquivos de log
 - Se está escrevendo à mão, registrar exatamente como está fazendo



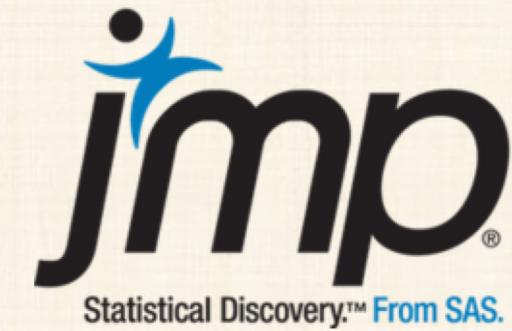
Ferramentas de codificação de pesquisa reproduzível:



Até as interfaces (GUI) podem ser reproduzíveis...



<https://galaxyproject.org>



Controle de versão

- Salvar versões dos seus dados / código
 - Tal que se possa retroceder quando um desastre acontecer...
- Moda antiga (old school): manuscript**V3**.doc
- Moderno: sistemas de controle de versão
 - Permite colaboração

GitHub



git

Lista de verificação para pesquisa reproduzível

- Pense em toda a linha de produção: todas as peças são reproduzíveis?
- O processo de limpeza / análise é automatizado? - Garante a reproduzibilidade?
- Você está fazendo coisas "à mão"? Editando tabelas / figuras; Dividindo / Reformatando dados
- Seu software trabalha com arquivos de log ou scripts? Se não, você possui uma descrição detalhada do seu processo?
- Você utiliza controle de versão?
- Você mantém controle sobre o software (atualizações, upgrades, etc.)?
- E quanto à arquitetura/configuração de seu computador:
 - OS / Software / ferramenta / add ons (bibliotecas / pacotes) / bancos de dados externos
 - Números de versão para tudo (quando disponível)
- Você salva os arquivos certos ?: se não é reproduzível, não vale a pena salvar
- Salve os dados e o código
 - Dados + Código = Saída

Adapted from:

https://github.com/DataScienceSpecialization/courses/blob/master/05_ReproducibleResearch/Checklist/Reproducible%20Research%20Checklist.pdf

Ameaças à Pesquisa Reprodutível



**HARKing – Hypothesizing After the Results are Known

**Elaborando hipóteses após os resultados serem conhecidos

Ameaças à
Pesquisa
Reprodutível!



Wonkblog

An alarming number of scientific papers contain Excel errors

By [Christopher Ingraham](#)

August 26, 2016



(Jewel Samad/AFP/Getty Images)

An advertisement for TestRail. It features the TestRail logo at the top, followed by the text "Modern Test Case Management Software for QA and Development Teams". Below this is a large orange button with the text "Start Testing Now". At the bottom, there is a screenshot of the TestRail software interface showing a dashboard with various charts and data tables.

Onde
encontrar
dados?

The screenshot shows a web browser window with several tabs open at the top. The active tab is 'dados.gov.br'. The page itself is the homepage of the Brazilian Open Data Portal. It features a green header with the text 'dados.gov.br' and 'PORTAL BRASILEIRO DE DADOS ABERTOS'. Below the header is a search bar with the placeholder 'Pesquisar conjuntos de dados...' and a 'PESQUISAR' button. To the right of the search bar is a sidebar titled 'Publicações mais recentes' (Recent publications) listing five entries. The main content area contains four sections: 'Conjuntos de dados em destaque' (Featured datasets), 'Compras públicas do governo federal' (Government procurement), 'Orçamento Federal - Orçamento Fiscal e da Seguridade Social' (Federal Budget - Fiscal and Social Security Budget), and 'Serviços Públicos Federais para o Cidadão' (Federal Public Services for Citizens). Each section has a brief description and a link to more information.

Conjunto de dados	Data	Responsável
Dinheiro em Circulação	06/09/2018	Banco Central do Brasil - BCB
Núcleo de Educação à Distância - NEAD	05/09/2018	Universidade Federal de Roraima - UF...
Diárias e passagens	05/09/2018	Universidade Federal do Oeste da Bah...
Outorgas de Serviços de Radiodifusão - TV Digital	03/09/2018	Ministério da Ciência, Tecnologia, I...
Outorgas de Serviços de Radiodifusão - DADOS DE DOCUMENTOS DE AUTORIA E DO	03/09/2018	Ministério da Ciência, Tecnologia, I...

Onde encontrar dados?

A screenshot of a Firefox browser window. The address bar shows the URL: <https://toolbox.google.com/datasetsearch/search?query=boston+education+data&docid=%2FR2hjicXe2XR6x1AA/>. The search query "boston education data" is entered in the search bar. The results page is from Google Dataset Search. The first result is for the ICPSR dataset, which includes links to www.icpsr.umich.edu, www.researchconnections.org, and search.datacite.org. The second result is for "Highest level of Education" from townfolio.co, last updated in January 2018. The third result is for "BOSTON EDUCATION SKILLS & TRAINING CORP, fiscal year ending June 2016" from projects.propublica.org. The fourth result is for "Massachusetts Public Schools Data" from kaggle.com, last updated on Aug 22, 2017. The fifth result is for "CPSR" (Inter-university Consortium for Political and Social Research) with links to the same three websites.

Dataset Search

boston education data

Data from: Political Power in Boston, Massachusetts and Charleston, South...
www.icpsr.umich.edu
www.researchconnections.org
+1mais

Highest level of Education
townfolio.co
Atualizado em Last Updated: January 2018

BOSTON EDUCATION SKILLS & TRAINING CORP, fiscal year ending June 2016
projects.propublica.org

Massachusetts Public Schools Data
www.kaggle.com
Atualizado em Aug 22, 2017

CPSR

Data from: Political Power in Boston, Massachusetts and Charleston, 1828-1843
Artigo relacionado

www.icpsr.umich.edu | www.researchconnections.org | search.datacite.org

Conjunto de dados fornecido por
Inter-university Consortium for Political and Social Research

Autores
Pease, Jane H.; Pease, William H.

Licença
<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/8653/terms>

Formatos de download disponibilizados pelos provedores
ASCII

Período de cobertura 1828 - 1843

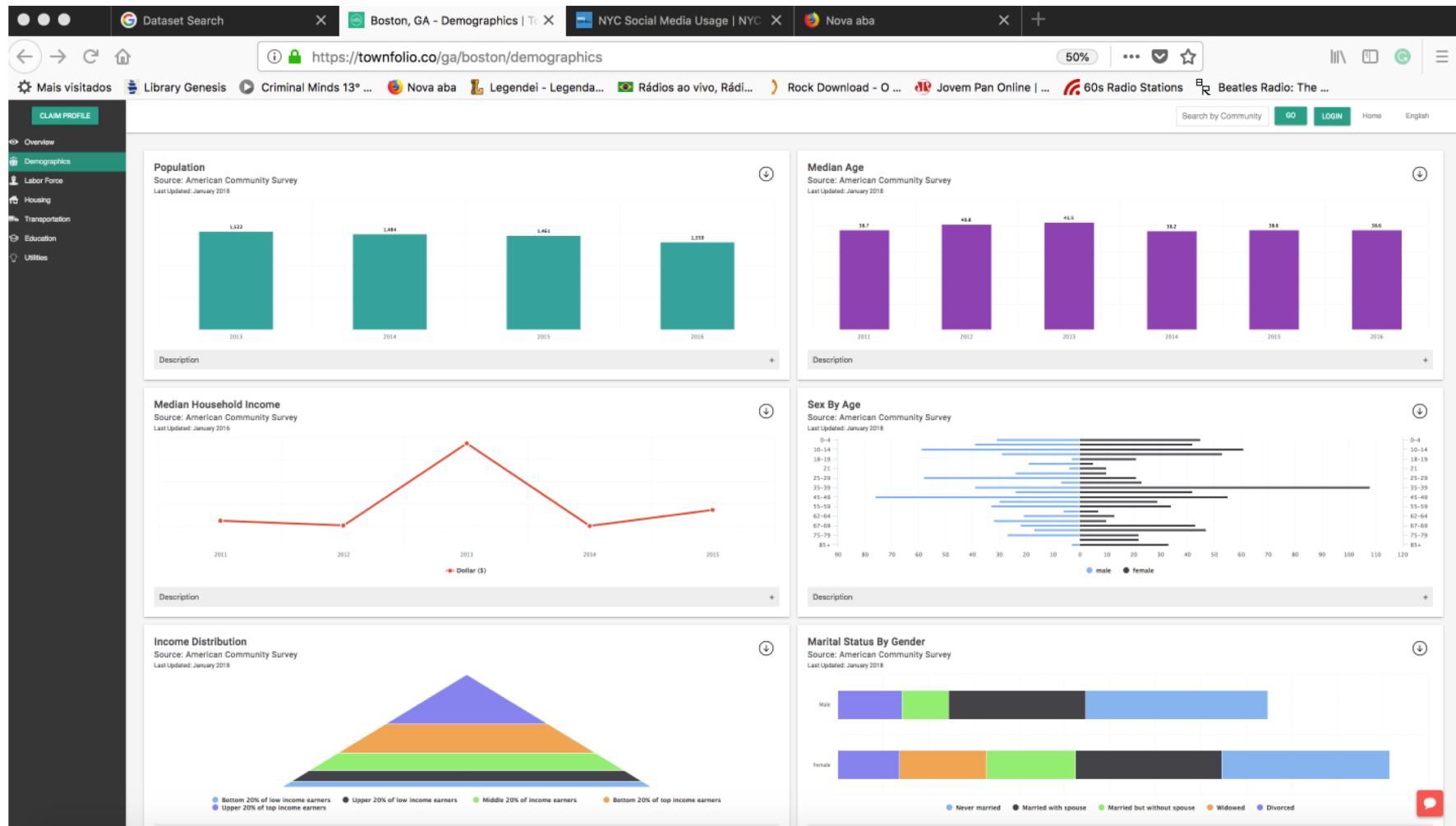
Onde
encontrar
dados?

The screenshot shows a web browser window with multiple tabs open. The active tab is for the Open Data Brasil website at <https://opendatabrasil.com.br/solucao-dados-abertos/>. The page features the Open Data Brasil logo and navigation links for HOME, QUEM SOMOS, SOLUÇÕES, PARCEIROS, BLOG, and CONTATO. A central section discusses the "Publicação de forma autônoma pela área de negócios" (Self-publication by the business area), explaining how it creates a data hub for sharing data between government and society. It highlights user preparation tools for various profiles, from non-technical to advanced users. To the right, there is a diagram illustrating a data hub. It shows a central white cloud containing a server icon and several document icons (Word, Excel, PDF). Three arrows point from three stylized human figures (blue, grey, and black) towards the cloud, representing data sharing and interactivity.

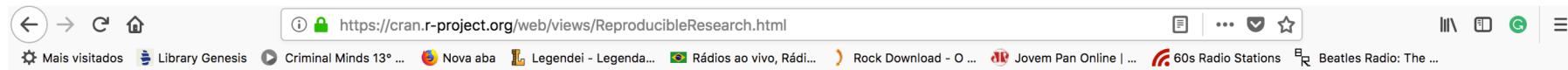
Publicação de forma autônoma pela área de negócios: possibilita a criação de um data hub para compartilhar os dados entre o governo e a sociedade. Possui funcionalidades de preparação de dados pelos usuários de negócios. Pensada para ser utilizada por vários perfis de usuários desde os não técnicos até os usuários avançados, facilitando o trabalho de preparação, publicação, reutilização e compartilhamento dos dados, podendo ser no formato de tabelas, gráficos, mapas, de maneira interativa e sem limite de acesso.

Agora, gestores e sociedade poderão ter os dados e informações na ponta dos dedos de maneira comprehensível e rápida.

Fontes de dados e gráficos



Pesquisa Reprodutível em R



CRAN Task View: Reproducible Research

Maintainer: Max Kuhn

Contact: mxkuhn at gmail.com

Version: 2018-07-30

URL: <https://CRAN.R-project.org/view=ReproducibleResearch>

The goal of reproducible research is to tie specific instructions to data analysis and experimental data so that scholarship can be recreated, better understood and verified. Packages in R for this purpose can be split into groups for: literate programming, package reproducibility, code/data formatting tools, format convertors, and object caching.

Literate Programming

The primary way that R facilitates reproducible research is using a document that is a combination of content and data analysis code. The `sweave` function (in the base R `utils` package) and the `knitr` package can be used to blend the subject matter and R code so that a single document defines the content and the analysis. The `brew` and `R.rsp` packages contain alternative approaches to embedding R code into various markups.

The resources for literate programming are best organized by the document type/markup language:

LaTeX

Both `sweave` and `knitr` can process LaTeX files. `lazyWeave` can create LaTeX documents from scratch.

Object Conversion Functions:

- *summary tables/statistics* : `Hmisc`, `NMOF`, `papeR`, `quantreg`, `rapport`, `reporttools`, `sparktex`, `tables`, `xtable`, `ztable`
- *tables/cross-tabulations* : `Hmisc`, `lazyWeave`, `knitLatex`, `knitr`, `reporttools`, `ztable`
- *graphics* : `animation`, `Hmisc`, `grDevices:::pictex`, `sparktex`, `tikzDevice`
- *statistical models/methods* : `apsrtable`, `memisc`, `quantreg`, `rms`, `stargazer`, `suRtex`, `texreg`, `xtable`, `ztable`
- *bibtex* : `bibtex` and `RefManageR`
- *others* : `latex2exp` converts LaTeX equations to `plotmath` expressions.

Miscellaneous Tools

- `Hmisc` contains a function to correctly escape special characters. `resumer` creates resumes. Standardized exams can be created using the `exams` package.

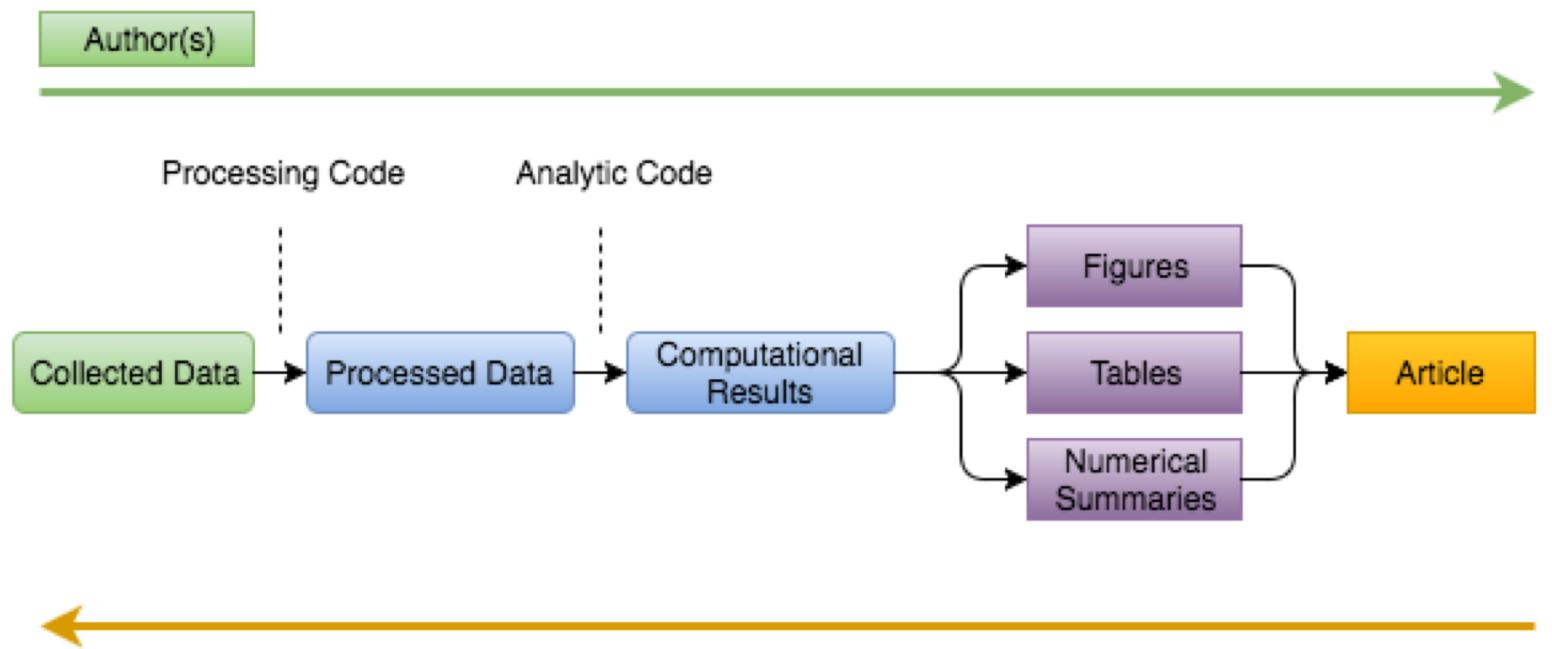
HTML

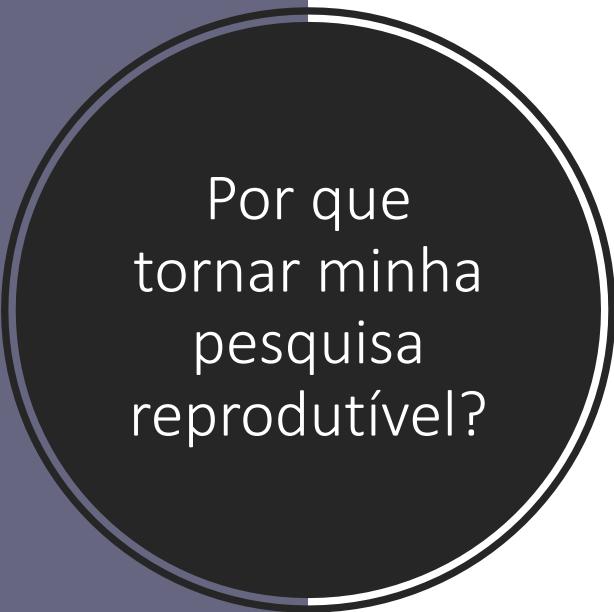
The `knitr` package can process HTML files directly. `sweave` can also work with HTML by way of the `R2HTML` package. `lazyWeave` can create HTML format documents from scratch.

Object Conversion Functions:

Por que
tornar minha
pesquisa
reprodutível?

- **Workflow:** overview do trabalho
- **Dados:** origem, tamanho e normalizações
- **Código:** scripts
- **Documentação:** etapas do trabalho
- **Ambiente:** versão das bibliotecas e softwares

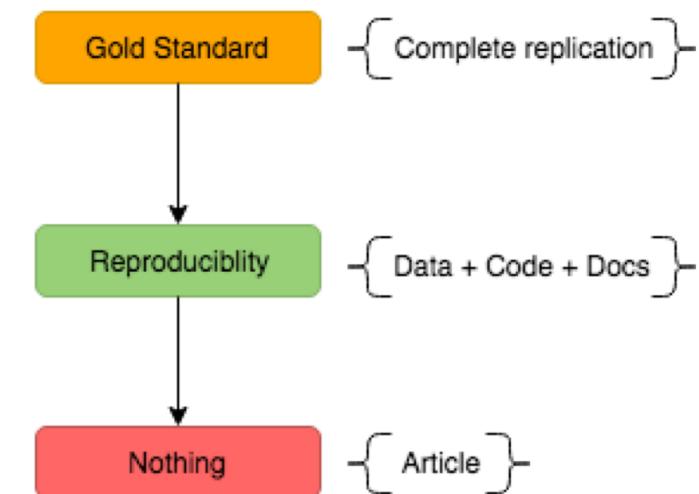




Por que
tornar minha
pesquisa
reprodutível?

- **Sobre os dados:**

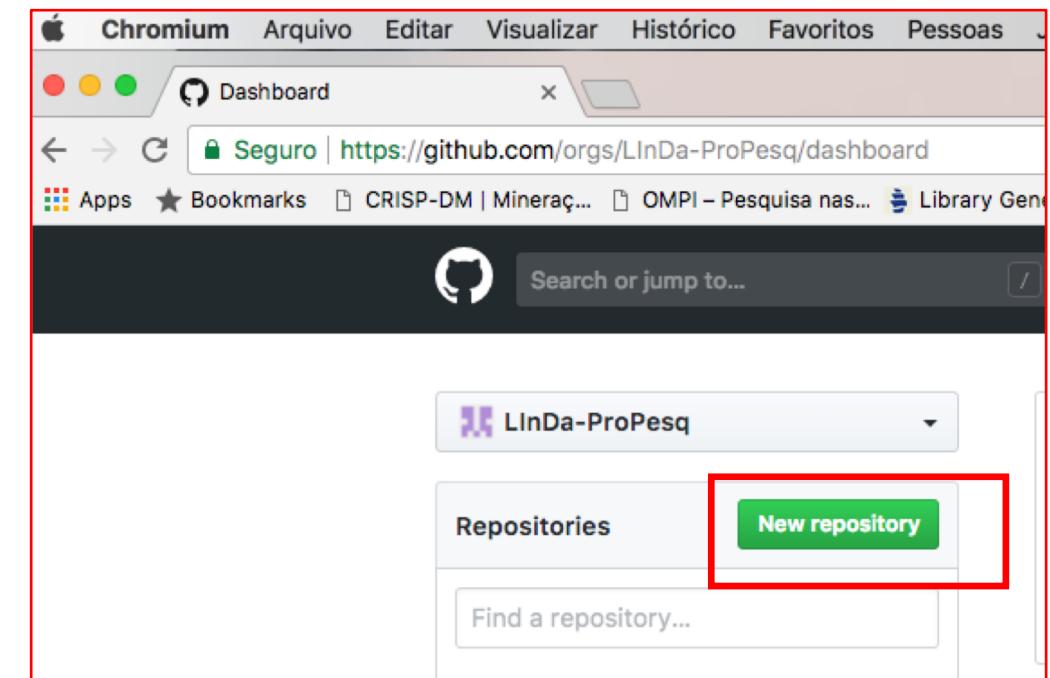
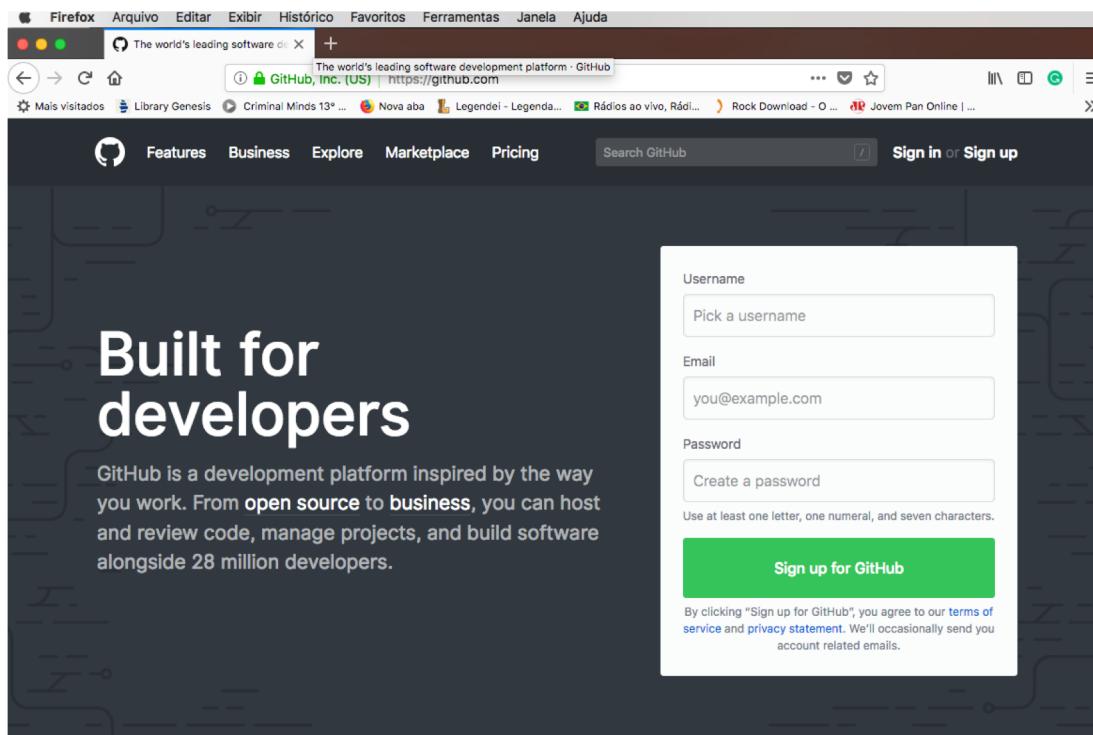
- **Brutos ou Normalizados**
- **Dados Normalizados:** incluir *dados brutos e scripts*
- **Automatizar tarefas:** dados consistentes e evita erros
- **Decisões do Projeto:** motivo e tipo de normalização / remoção de dados / técnicas utilizadas
- **Procedência dos dados:** origem dos dados / transparência



Exemplo de pesquisa reprodutível com R

Iniciando o processo:

- 1. Verificar se o Git está instalado
- 2. Criar uma conta do [Github](#), caso não tiver
- 3. Configurar a [autenticação do Github](#) na máquina.
- Ciar um novo repositório remoto para o novo projeto

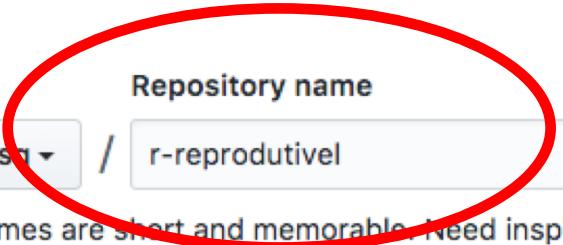


Passo 1: Preenchendo os dados no GitHub

Create a new repository

A repository contains all the files for your project, including the revision history.

Owner / 

Repository name 

Great repository names are short and memorable. Need inspiration? How about [upgraded-adventure](#).

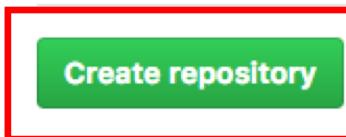
Description (optional)
Pesquisa reproduzível com a linguagem R

 Public
Anyone can see this repository. You choose who can commit.

 Private
You choose who can see and commit to this repository.

Initialize this repository with a README
This will let you immediately clone the repository to your computer. Skip this step if you're importing an existing repository.

Add .gitignore: Add a license: 

 **Create repository**

Chromium Arquivo Editar Visualizar Histórico Favoritos Pessoas Janela Ajuda

LInDa-ProPesq/r-reprodutivel: x Seguro | https://github.com/LInDa-ProPesq/r-reprodutivel

João Pedro

Apps Bookmarks CRISP-DM | Mineração... OMPI – Pesquisa nas... Library Genesis Importados do Safari Using R for statistica... Plataforma Brasil Microsoft DreamSpark Outros favoritos

Search or jump to... Pull requests Issues Marketplace Explore

LInDa-ProPesq / r-reprodutivel Unwatch 1 Star 0 Fork 0

Pesquisa reproduzível com a linguagem R

Manage topics

1 commit 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

jpalbino Initial commit Latest commit 96e9b22 just now

README.md Initial commit just now

README.md

r-reprodutivel

Pesquisa reproduzível com a linguagem R

Novo repositório remoto criado ! <https://github.com/LInDa-ProPesq/r-reprodutivel>

Passo 2: Clonar o repositório

1. No GitHub, navegar até a página principal do repositório
2. Sob o nome do repositório, clique em Clonar ou fazer o download
3. Na seção Clonar com HTTPS, clique para copiar o URL clone do repositório.

The screenshot shows a GitHub repository page for 'LInDa-ProPesq / r-reprodutivel'. A large red arrow on the left points from the text 'Passo 2:' to the repository name. Step 1 is highlighted with a red box around the repository name 'LInDa-ProPesq / r-reprodutivel'. Step 2 is highlighted with a red box around the 'Clone or download' button. Step 3 is highlighted with a red box around the 'Clone with HTTPS' URL field and its copy icon.

Passo 2:
Clonar o
repositório

1 LInDa-ProPesq / r-reprodutivel

2 Clone or download

3 https://github.com/LInDa-ProPesq/r-reprodutivel

Clone with HTTPS

Use Git or checkout with SVN using the web URL.

https://github.com/LInDa-ProPesq/r-reprodutivel

Open in Desktop

Download ZIP

Pesquisa reproduzível com a linguagem R

Passo 3: Clonar o diretório no computador

1. Abrir o Terminal.
2. Alterar o diretório de trabalho atual para o local do diretório a ser clonado
3. Digite **git clone** e cole o URL copiado.

```
MBP-de-Joao-2:projetos jpalbino$ git clone https://github.com/LInDa-ProPesq/r-reproductivel.git
Cloning into 'r-reproductivel'...
remote: Counting objects: 3, done.
remote: Compressing objects: 100% (2/2), done.
remote: Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (3/3), done.

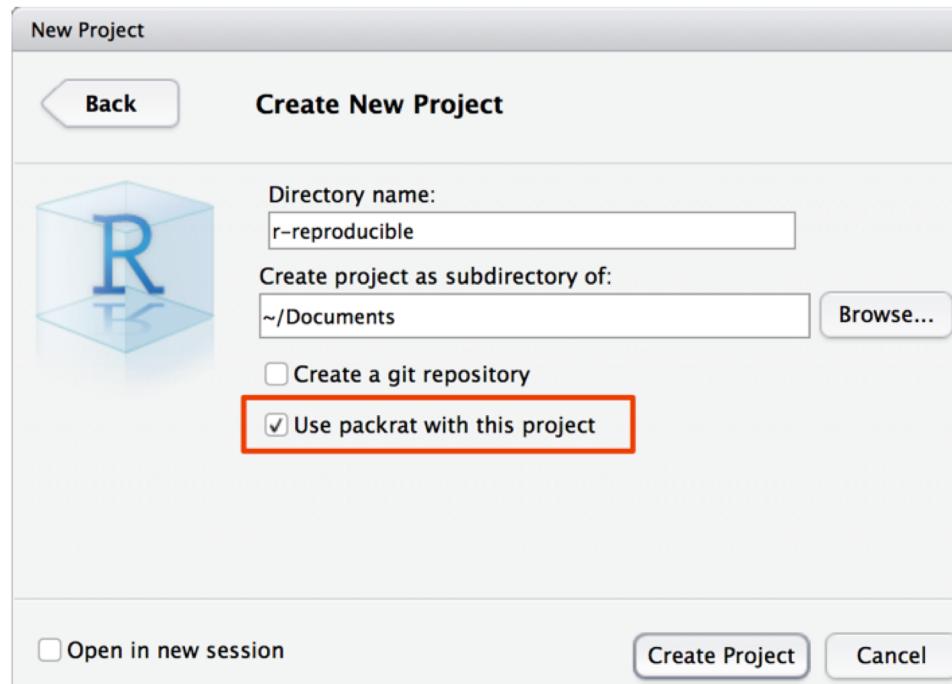
```

Exemplo de pesquisa reprodutível com R

Passo 4: Configuração do Ambiente: instalar globalmente o packrat no computador.

```
install.packages("packrat")
```

Passo 5: criar um novo projeto no RStudio em no diretório clonado pelo git



Fonte: <http://www-bcf.usc.edu/~gareth/ISL/data.html> e8496

Fonte: <http://t-redactyl.io/blog/2016/10/a-crash-course-in-reproducible-research-in-r.html>

INICIALIZANDO PACKRAT E CRIANDO PROJETO

```
Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,  
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
```

```
Digite 'q()' para sair do R.
```

```
[Workspace loaded from ~/OneDrive/Projetos 2018/Ciencia de Dados/Introducao a Ciencia de Dados/Projetos/Analysis of Los Angeles Crime with R/.RData]
```

```
Initializing packrat project in directory:
```

```
- ~/OneDrive/Projetos 2018/Ciencia de Dados/Introducao a Ciencia de Dados/Projetos/r-reproductivel
```

```
Packrat mode off. Resetting library paths to:
```

```
- "/Library/Frameworks/R.framework/Versions/3.5/Resources/library"
```

```
Adding these packages to packrat:
```

```
  -  
    packrat  0.4.9-3
```

```
Fetching sources for packrat (0.4.9-3) ... OK (CRAN current)
```

```
Snapshot written to '/Users/jpalbino/OneDrive/Projetos 2018/Ciencia de Dados/Introducao a Ciencia de Dados/Projetos/r-reproductivel/packrat/packrat.lock'
```

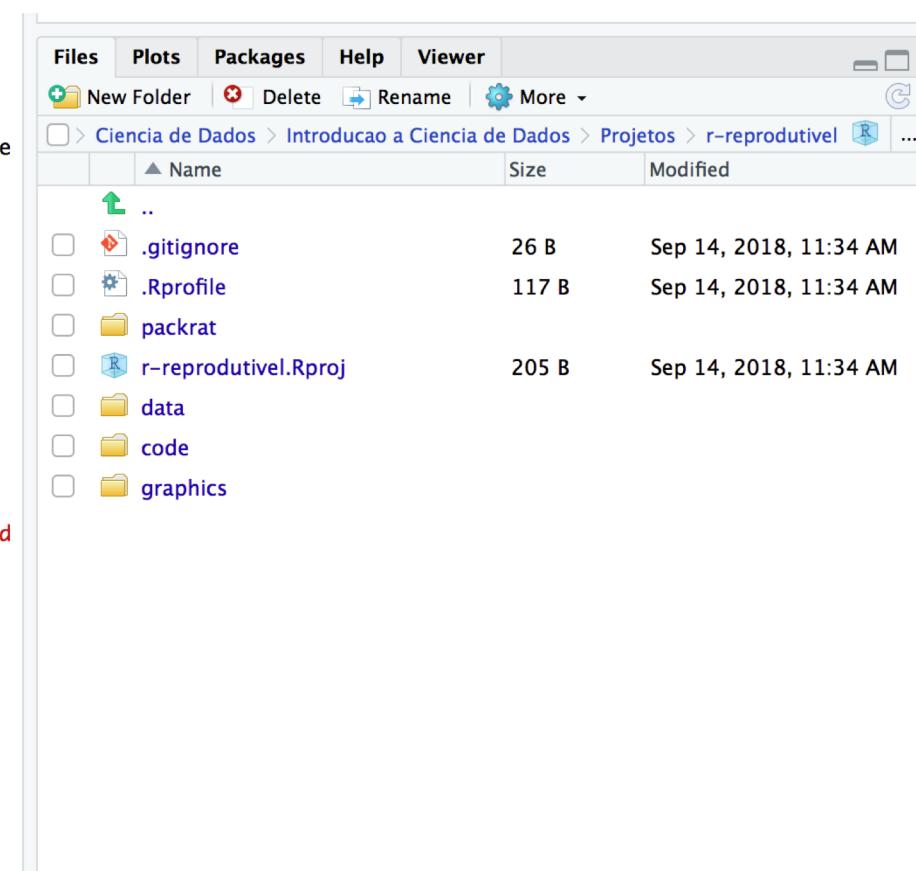
```
Installing packrat (0.4.9-3) ...
```

```
  OK (downloaded binary)
```

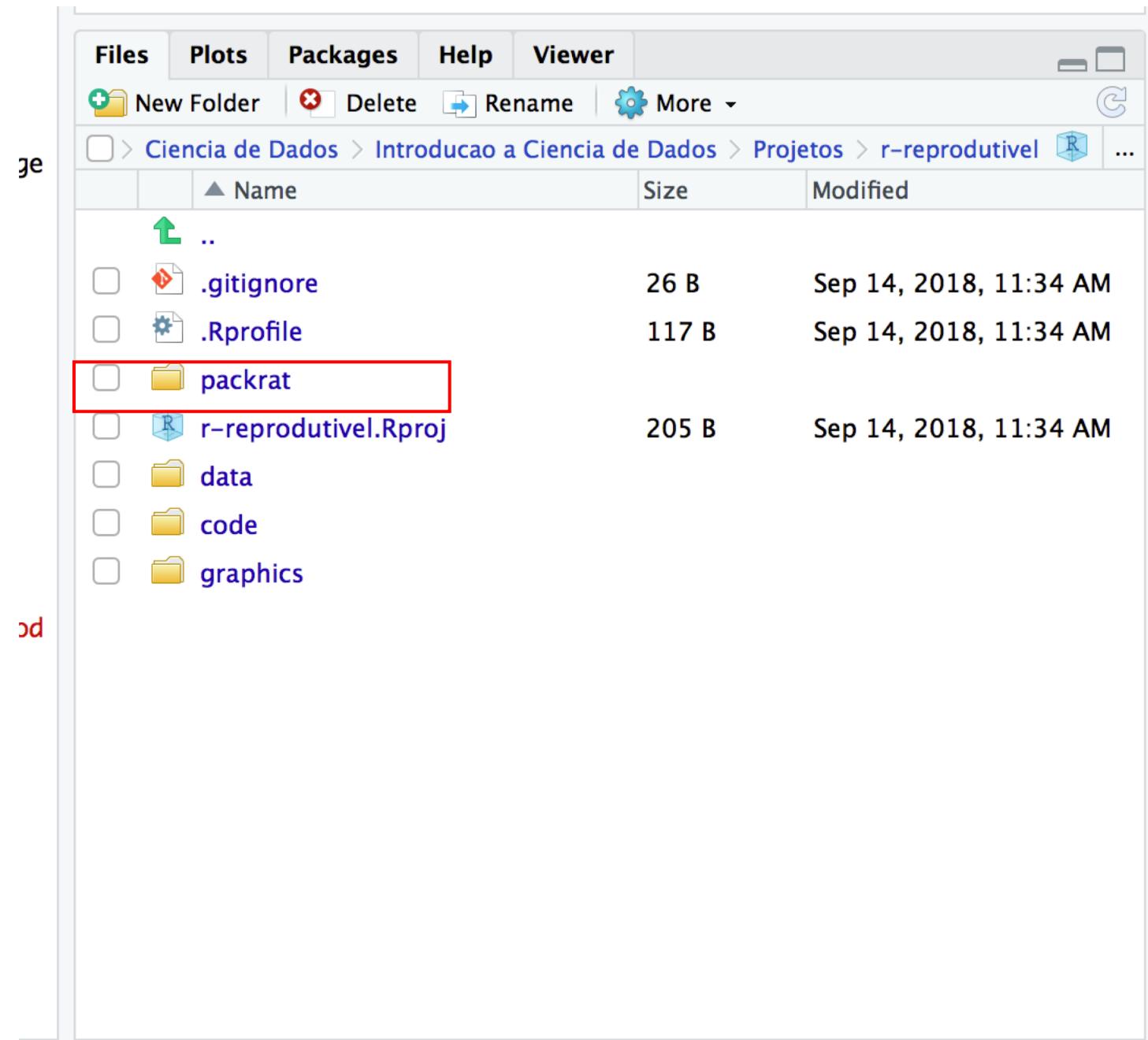
```
Initialization complete!
```

```
> setwd("~/OneDrive/Projetos 2018/Ciencia de Dados/Introducao a Ciencia de Dados/Projetos/r-reproductivel")
```

```
>
```



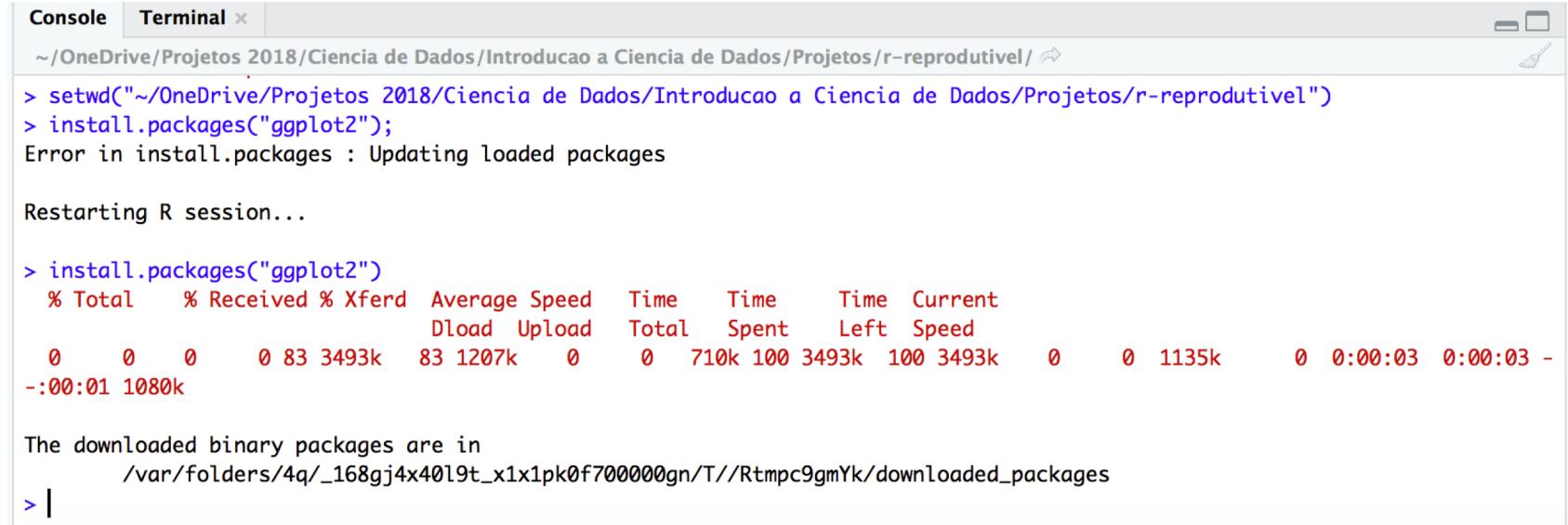
Novo folder
chamado
'packrat'
dentro do
folder
clonado



Exemplo de pesquisa reproduzível com R

Passo 6: Instalar as bibliotecas necessárias.

```
install.packages("ggplot2"); library(ggplot2)
```



The screenshot shows an R console window with two tabs: "Console" and "Terminal". The "Console" tab is active, displaying the command `install.packages("ggplot2")`. The output shows an error message: "Error in install.packages : Updating loaded packages". It then restarts the R session and shows the progress of the package download. The download table includes columns for % Total, % Received, % Xferd, Average Speed, Time Dload, Upload, Total, Spent, Left, and Current Speed. The final message indicates that the downloaded binary packages are located in a specific directory.

```
Console Terminal ×
~/OneDrive/Projetos 2018/Ciencia de Dados/Introducao a Ciencia de Dados/Projetos/r-reprodutivel/ ↗
> setwd("~/OneDrive/Projetos 2018/Ciencia de Dados/Introducao a Ciencia de Dados/Projetos/r-reprodutivel")
> install.packages("ggplot2");
Error in install.packages : Updating loaded packages

Restarting R session...

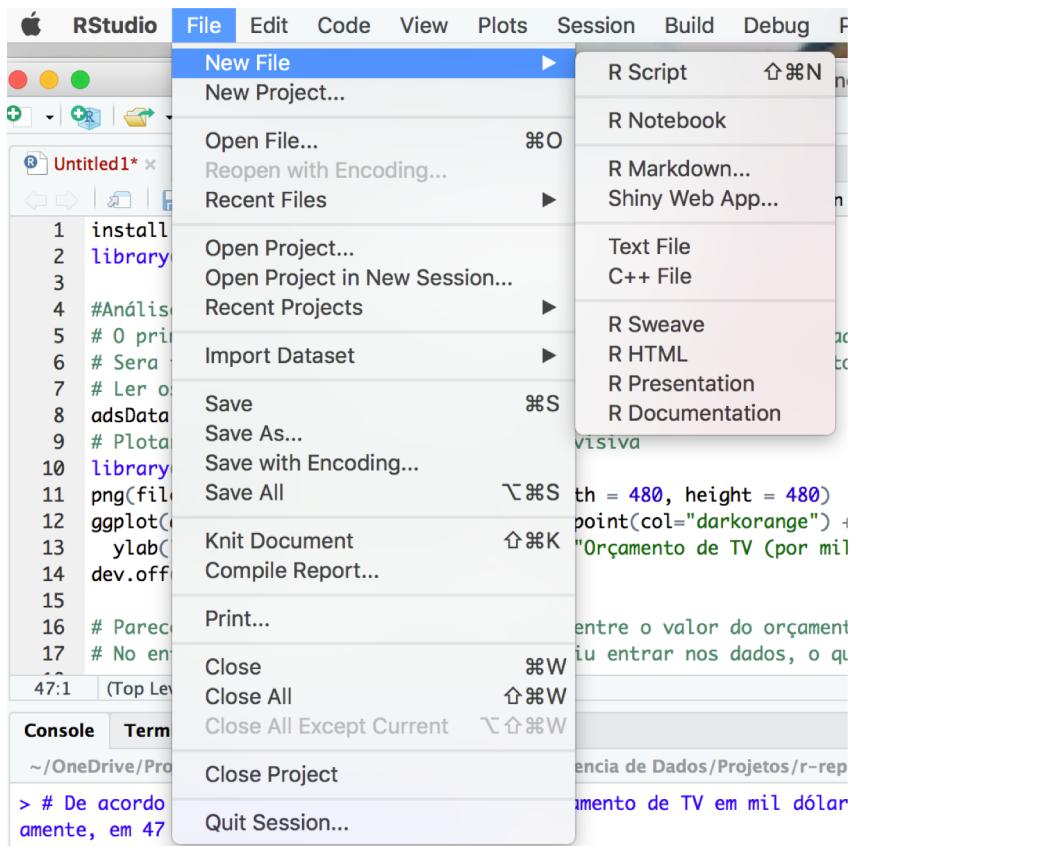
> install.packages("ggplot2")
  % Total    % Received % Xferd  Average Speed   Time   Time     Time  Current
  Dload  Upload Total   Spent   Left  Speed
  0      0      0  83 3493k  83 1207k   0      0  710k 100 3493k  100 3493k   0      0 1135k      0  0:00:03  0:00:03 -
-:00:01 1080k

The downloaded binary packages are in
  /var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T//Rtmpc9gmYk/downloaded_packages
> |
```

Fonte: <http://www-bcf.usc.edu/~gareth/ISL/data.html> e8496

Fonte: <http://t-redactyl.io/blog/2016/10/a-crash-course-in-reproducible-research-in-r.html>

Passo 7: Criar e digitar o script em R



The screenshot shows the RStudio interface with the code editor pane active. The code editor contains the following R script:

```
1 install.packages("ggplot2");
2 library(ggplot2)
3 
4 #Análise exploratória de dados
5 # O primeiro passo de uma análise geralmente envolve explorar os dados.
6 # Sera feita a exploração para o arquivo Advertising2.csv (orçamentos de TV).
7 # Ler os dados
8 adsData <- read.csv("data/Advertising2.csv")
9 # Plotar vendas por mil de publicidade televisiva
10 library(ggplot2)
11 png(filename = "graphics/publi_tv.png", width = 480, height = 480)
12 ggplot(adsData, aes(x=TV, y=Sales)) + geom_point(col="darkorange") +
13   ylab("Vendas (por mil unidades)") + xlab("Orçamento de TV (por mil dólares)")
14 dev.off()
15 
16 # Parece haver uma clara relação positiva entre o valor do orçamento e suas vendas.
17 # No entanto, um orçamento negativo conseguiu entrar nos dados, o que é mais provável que s
47:1 (Top Level) <-->
```

The code editor has a status bar at the bottom indicating 'R Script'. To the right of the code editor are the Environment, History, Connections, and Global Environment panes. The Environment pane shows variables like 'adsData', 'lmFit', and 'outlierIndex'. The Global Environment pane shows the same variables. At the bottom, the Terminal pane shows the command 'RStudioGD' followed by two question marks.

Passo 8: Navegar até o diretório no computador

```
[MBP-de-Joao-2:r-reprodutivel jp Albino]$ ls
[code                      graphics                  r-reprodutivel.Rproj
data                      packrat
[MBP-de-Joao-2:r-reprodutivel jp Albino]$ ls -lia
total 48
33422834 drwxr-xr-x 13 jp Albino  staff   442 14 Set 14:08 .
29034539 drwxr-xr-x@ 14 jp Albino  staff   476 14 Set 11:34 ..
33426240 -rw-r--r--@  1 jp Albino  staff  10244 14 Set 14:14 .DS_Store
[33439737 -rw-r--r--  1 jp Albino  staff    0 14 Set 14:08 .Rhistory
33422870 -rw-r--r--  1 jp Albino  staff   117 14 Set 11:34 .Rprofile
[33422969 drwxr-xr-x  4 jp Albino  staff   136 14 Set 11:34 .Rproj.user
[33422836 drwxr-xr-x  9 jp Albino  staff   306 14 Set 14:07 .git
33422873 -rw-r--r--  1 jp Albino  staff    26 14 Set 11:34 .gitignore
33423884 drwxr-xr-x  4 jp Albino  staff   136 14 Set 14:19 code
[33423846 drwxr-xr-x  3 jp Albino  staff   102 14 Set 12:03 data
33424101 drwxr-xr-x  5 jp Albino  staff   170 14 Set 14:09 graphics
33422867 drwxr-xr-x  9 jp Albino  staff   306 14 Set 11:34 packrat
33422835 -rw-r--r--  1 jp Albino  staff   205 14 Set 11:34 r-reprodutivel.Rproj
[MBP-de-Joao-2:r-reprodutivel jp Albino]$
```

Passo 7: Verificar arquivos que o Git está acompanhando

```
[MBP-de-Joao-2:r-reproductivel jpalbino$ git status
On branch master
Your branch is up to date with 'origin/master'.

Untracked files:
  (use "git add <file>..." to include in what will be committed)

    .RData
    .Rhistory
    .Rprofile
    .gitignore
    code/
    data/
    graphics/
    packrat/
    r-reproductivel.Rproj

nothing added to commit but untracked files present (use "git add" to track)
MBP de Joao 2:a:r-reproductivel jpalbino$ git add ..
```

Passo 9: Preparar para transferir arquivos ao remoto

```
MBP-de-Joao-2:r-reproductivel jpalbino$ git add *
```

```
MBP-de-Joao-2:r-reproductivel jpalbino$ [MBP-de-Joao-2:r-reproductivel jpalbino$ git status  
On branch master  
Your branch is up to date with 'origin/master'.  
  
Changes to be committed:  
(use "git reset HEAD <file>..." to unstage)  
  
    new file:   code/ajuste_linear_propaganda_tv.R  
    new file:   code/lmFitAdvertising.R  
    new file:   data/Advertising2.csv  
    new file:   graphics/publi_tv.png  
    new file:   graphics/relacao_orcamento_tv.png  
    new file:   graphics/relcao_orcamento_tv.png  
    new file:   packrat/init.R  
    new file:   packrat/packrat.lock  
    new file:   packrat/packrat.opts  
    new file:   packrat/src/R6/R6_2.2.2.tar.gz  
    new file:   packrat/src/RColorBrewer/RColorBrewer_1.1-2.tar.gz  
    new file:   packrat/src/Rcpp/Rcpp_0.12.18.tar.gz  
    new file:   packrat/src/assertthat/assertthat_0.2.0.tar.gz  
    new file:   packrat/src/cli/cli_1.0.0.tar.gz  
    new file:   packrat/src/colorspace/colorspace_1.3-2.tar.gz  
    new file:   packrat/src/crayon/crayon_1.3.4.tar.gz  
    new file:   packrat/src/digest/digest_0.6.16.tar.gz  
    new file:   packrat/src/fansi/fansi_0.3.0.tar.gz  
    new file:   packrat/src/ggplot2/ggplot2_3.0.0.tar.gz  
    new file:   packrat/src/glue/glue_1.3.0.tar.gz  
    new file:   packrat/src/gtable/gtable_0.2.0.tar.gz  
    new file:   packrat/src/labeling/labeling_0.3.tar.gz
```

Obs: Incluindo “todos” os arquivos!

```
Untracked files:
  (use "git add <file>..." to include in what will be committed)

  .RData
  .Rhistory
  .Rprofile
  .gitignore

[MBP-de-Joao-2:r-reproductivel jpalbino$ git add .RData
[MBP-de-Joao-2:r-reproductivel jpalbino$ git add .Rhistory
[MBP-de-Joao-2:r-reproductivel jpalbino$ git add .Rprofile
[MBP-de-Joao-2:r-reproductivel jpalbino$ git add .gitignore
[MBP-de-Joao-2:r-reproductivel jpalbino$ git status
On branch master
Your branch is up to date with 'origin/master'.

Changes to be committed:
  (use "git reset HEAD <file>..." to unstage)

  new file:  .RData
  new file:  .Rhistory
  new file:  .Rprofile
  new file:  .gitignore
  new file:  code/ajuste_linear_propaganda_tv.R
  new file:  code/lmFitAdvertising.R
  new file:  data/Advertising2.csv
  new file:  graphics/publi_tv.png
  new file:  graphics/relacao_orcamento_tv.png
  new file:  graphics/relcao_orcamento_tv.png
  new file:  packrat/init.R
  new file:  packrat/packrat.lock
  new file:  packrat/packrat.opts
  new file:  packrat/src/R6/R6_2.2.2.tar.gz
  new file:  packrat/src/RColorBrewer/RColorBrewer_1.1-2.tar.gz
  new file:  packrat/src/Rcpp/Rcpp_0.12.18.tar.gz
  new file:  packrat/src/assertthat/assertthat_0.2.0.tar.gz
  new file:  packrat/src/cli/cli_1.0.0.tar.gz
  new file:  packrat/src/colorspace/colorspace_1.3-2.tar.gz
  new file:  packrat/src/crayon/crayon_1.3.4.tar.gz
  new file:  packrat/src/digest/digest_0.6.16.tar.gz
```

O Git colocou todos os arquivo **em uma fila** para serem confirmados (comitt)
É preciso explicitamente executar o “**comitt**” com uma mensagem

Finalizando as operações!

```
[MBP-de-Joao-2:r-reprodutivel jp Albino$ git remote -v  
origin https://github.com/LInDa-ProPesq/r-reprodutivel.git (fetch)  
origin https://github.com/LInDa-ProPesq/r-reprodutivel.git (push)  
[MBP-de-Joao-2:r-reprodutivel jp Albino$ git commit -m "Finalizado"  
[master 57fb940] Finalizado  
 42 files changed, 795 insertions(+)  
 create mode 100644 .RData  
 create mode 100644 .Rhistory  
 create mode 100644 .Rprofile  
 create mode 100644 .gitignore  
 create mode 100644 code/ajuste_linear_propaganda_tv.R  
 create mode 100644 code/lmFitAdvertising.R  
 create mode 100644 data/Advertising2.csv  
 create mode 100644 graphics/publi_tv.png  
 create mode 100644 graphics/relacao_orcamento_tv.png  
 create mode 100644 graphics/relcao_orcamento_tv.png  
 create mode 100644 packrat/init.R  
 create mode 100644 packrat/packrat.lock  
 create mode 100644 packrat/packrat.opts  
 create mode 100644 packrat/src/R6/R6_2.2.2.tar.gz  
 create mode 100644 packrat/src/RColorBrewer/RColorBrewer_1.1-2.tar.gz  
 create mode 100644 packrat/src/Rcpp/Rcpp_0.12.18.tar.gz  
 create mode 100644 packrat/src/assertthat/assertthat_0.2.0.tar.gz  
 create mode 100644 packrat/src/ansi/ansi_1.0.0.tar.gz
```

```
[MBP-de-Joao-2:r-reprodutivel jp Albino$ git push  
Enumerating objects: 78, done.  
Counting objects: 100% (78/78), done.  
Delta compression using up to 8 threads.  
Compressing objects: 100% (52/52), done.  
Writing objects: 100% (77/77), 29.25 MiB | 516.00 KiB/s, done.  
Total 77 (delta 0), reused 0 (delta 0)  
To https://github.com/LInDa-ProPesq/r-reprodutivel.git  
 96e9b22..57fb940 master -> master
```

Pesquisa Reprodutível:

