

# Introdução ao Tratamento e Análise de Dados em R

## Organizando a Atividade de Tratamento de dados com R

Sérgio Rivero

PPGE-UFPA

23 de maio de 2019



# Sumário

- 1 Objetivo da Aula
- 2 Ciclo de tratamento de dados
- 3 Fluxo do tratamento de dados
- 4 Trabalhando em Equipe
- 5 Aprendendo mais sobre R
- 6 Exercícios

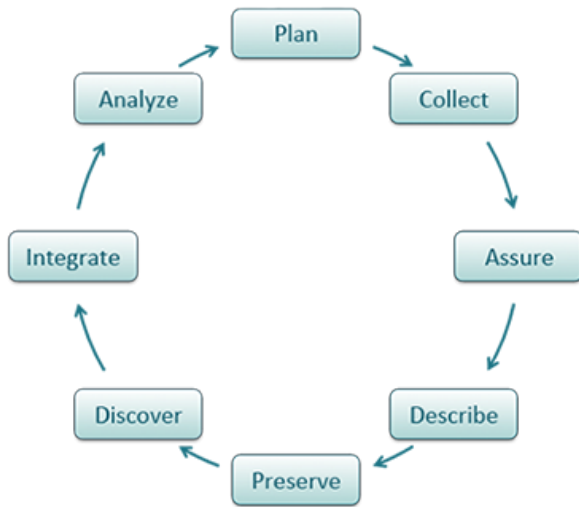


# Objetivo da Aula

- 1 Discutir as estratégias para organizar o processo de tratamento de dados
- 2 Mostrar ferramentas de trabalho colaborativo com dados
- 3 Mostrar algumas das principais fontes de informação sobre o R



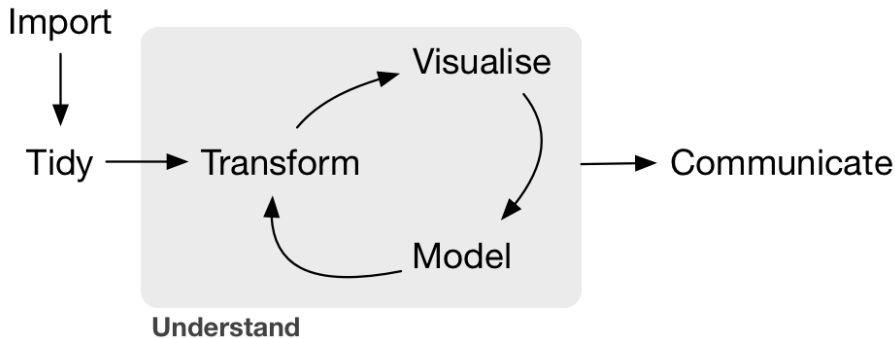
# Relembrando o Ciclo de tratamento de dados



<https://www.dataone.org/data-life-cycle>



# O ciclo de tratamento de dados e modelagem



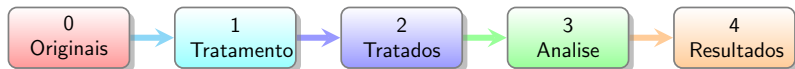
Tratamento de Dados Estatísticos



# Fluxo do tratamento de dados



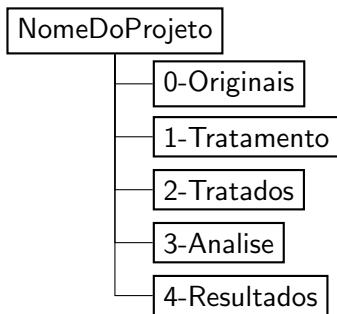
# Uma estrutura sugerida de arquivos adequada para isso



- ① **Originais** vai conter seus arquivos originais (json, csv, txt, etc). Os arquivos a partir dos quais você fazer o tratamento dos dados.
- ① **Tratamento** terá os programas R (ou de outro tipo) que vão gerar seus dados limpos, prontos para a análise.
- ② **Tratados** terá seus arquivos (em geral, dataframes) tratados.
- ③ **Analise** terá seus programas R (ou rmd ou Rnw) para a análise dos dados e geração de resultados (pode ser 2 pastas)
- ④ **Resultados** terá seus arquivos de resultados e os relatórios apresentados (pode ser 2 pastas)



# A estrutura de pastas (sugerida)

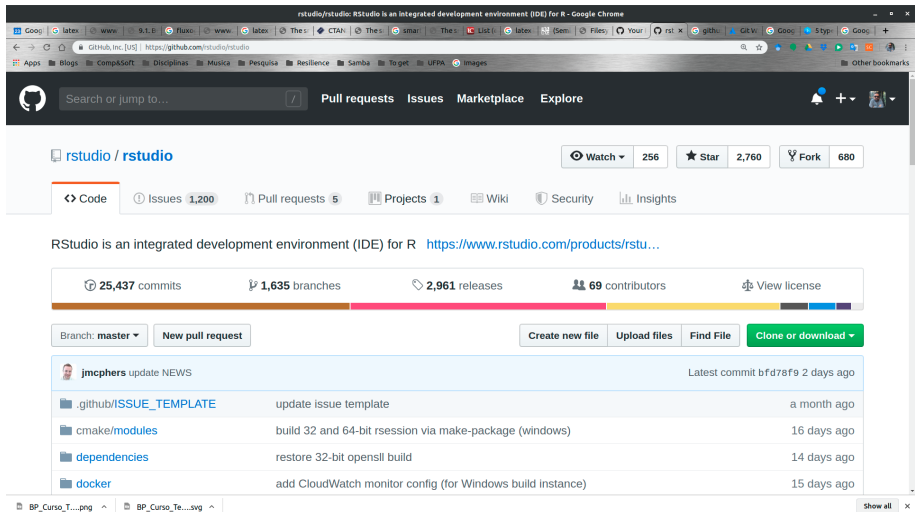


- 1 O fluxo segue sempre em apenas um sentido.
- 2 A correção e alteração dos dados é feita sempre a partir de um programa R
- 3 Erros eventuais exigirão a reexecução de um programa R para sua correção
- 4 Lembre-se, que repete a execução é o computador, não você.
- 5 Alterações erradas nos dados de origem produzem problemas difíceis de corrigir





# Trabalhando em Equipe - O GitHub



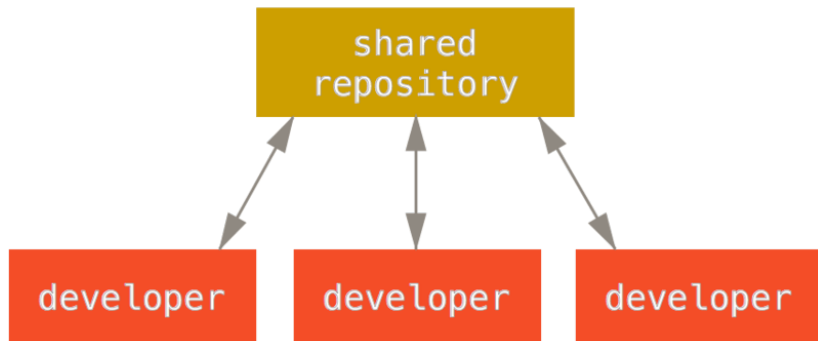
The screenshot displays the GitHub interface for the `rstudio/rstudio` repository. At the top, the browser address bar shows the URL `https://github.com/rstudio/rstudio`. The repository header includes the GitHub logo, a search bar, and navigation links for Pull requests, Issues, Marketplace, and Explore. Below the header, the repository name `rstudio / rstudio` is shown, followed by statistics: 256 Watch, 2,760 Star, and 680 Fork. The main content area features a tabbed interface with Code, Issues (1,200), Pull requests (5), Projects (1), Wiki, Security, and Insights. A summary bar indicates 25,437 commits, 1,635 branches, 2,961 releases, and 69 contributors. Below this, a table lists recent commits by user `jmcphers`, including updates to the issue template, build scripts, dependencies, and Docker configuration.

Commit	Message	Time
<code>jmcphers</code>	update NEWS	Latest commit bfd78f9 2 days ago
<code>.github/ISSUE_TEMPLATE</code>	update issue template	a month ago
<code>cmake/modules</code>	build 32 and 64-bit rsession via make-package (windows)	16 days ago
<code>dependencies</code>	restore 32-bit openssl build	14 days ago
<code>docker</code>	add CloudWatch monitor config (for Windows build instance)	15 days ago



<https://github.com/rstudio/rstudio>

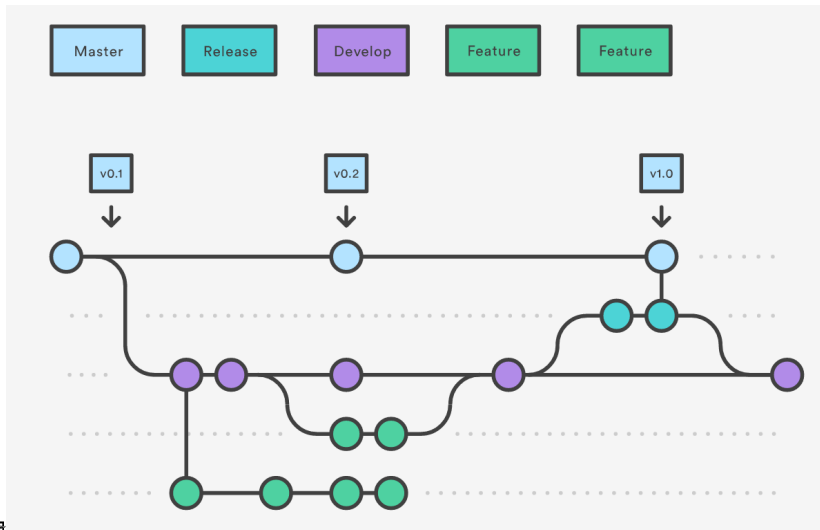
# A estrutura de trabalho no GitHub



Atividades Paralelas - Repositório único



# Fluxo de trabalho no GitHub



Ramos (branches) e junções (merges)

**FACECON**  
Faculdade de Ciências Econômicas



# Onde procurar informação?

- <https://cran.r-project.org/>
- <https://www.rstudio.com/>
- <https://www.rdocumentation.org/>
- <https://www.bioconductor.org/>
- <https://www.r-bloggers.com/>
- <https://r4ds.had.co.nz/>
- [https://en.wikibooks.org/wiki/R\\_Programming](https://en.wikibooks.org/wiki/R_Programming)
- <https://stats.idre.ucla.edu/r/>
- <http://www.cookbook-r.com/>
- <http://www.cookbook-r.com/Graphs/>



# Exercícios

## Exercícios

