

# Coleta de dados automatizada e integração contínua utilizando GitHub Actions:

## Exemplos com o Pacote Mananciais

Beatriz Milz

International Seminar on Statistics with R

Junho de 2021



# Beatriz Milz

- 2019 - \* - Doutoranda em Ciência Ambiental - PROCAM/IEE/USP
- 2016-2018 - Mestrado em Ciências - Programa de Pós-Graduação em Análise Ambiental Integrada - UNIFESP
- 2012-2015 - Bacharelado em Gestão Ambiental - EACH/USP

## R: Comunidades, eventos, pacotes, livros..



# Objetivos



- Introduzir o conceito de GitHub Actions:
  - Para que serve?
  - Quais são os pré-requisitos para usar?
  - Onde buscar mais informações?
- O pacote Mananciais servirá de exemplo!
- O objetivo **não** é virar ninja em 20 minutos:





# Pacote Mananciais

- Disponibiliza dados do mananciais que abastecem a Região Metropolitana de São Paulo (RMSP)
- Importância dos dados:
  - RMSP - mais de 21 milhões de habitantes ([fonte](#)) + crise hídrica em 2014/2015
  - O volume armazenado nos reservatórios é um dado muito usado por pessoas pesquisadoras e pela mídia.



# Pacote Mananciais

- Dados são originalmente disponibilizados no Portal dos mananciais da Sabesp:  
<https://mananciais.sabesp.com.br/> . Algumas dificuldades de acesso:
  - Página "Situação dos mananciais" não apresenta opção para fazer o download da base de dados
  - Download é possível na página "Dados dos Sistemas Produtores", selecionando "valores agregados para o sistema";
  - Mais dificuldades: só é possível fazer a consulta de um sistema por vez; ao fazer consulta de longos períodos o site trava; formato do arquivo baixado não é o ideal.
- Os dados são atualizados todos os dias de manhã.



# Pacote Mananciais

- Disponibiliza os dados atualizados diariamente da página "Situação dos mananciais":
  - em `.xlsx` e em `.csv`
  - com a função:

```
mananciais <- mananciais::dados_mananciais()
dplyr::glimpse(mananciais)
```

```
## Rows: 47,994
## Columns: 8
## $ data      <date> 2021-05-09, 2021-05-09, 2021-05-09, 2021-05-09, 2...
## $ sistema   <chr> "Cantareira", "Alto Tietê", "Guarapiranga", "Cotia..."
## $ volume_porcentagem <dbl> 49.6, 58.5, 69.4, 77.8, 88.5, 71.0, 77.8, 49.8, 58...
## $ volume_variacao  <dbl> -0.2, -0.1, -0.2, -0.2, 0.0, 0.1, 0.2, -0.1, -0.1,...
## $ volume_operacional <dbl> 487.50590, 327.96921, 118.82658, 12.83817, 99.3076...
## $ pluviometria_dia   <dbl> 0.0, 0.0, 0.2, 0.0, 0.2, 0.0, 2.6, 0.0, 0.4, 0.2, ...
## $ pluviometria_mensal <dbl> 0.3, 1.6, 0.6, 1.0, 1.6, 28.6, 6.0, 0.3, 1.6, 0.4,...
## $ pluviometria_hist  <dbl> 77.4, 75.4, 58.6, 64.6, 80.4, 133.3, 89.3, 77.4, 7...
```

<https://beatrizmilz.github.io/mananciais/>



# Como isso é feito?

- Dados da SABESP estão disponíveis em uma API, e é atualizado todos os dias de manhã.
- O pacote mananciais apresenta uma função que atualiza a base de dados que está armazenada no pacote, buscando os dados atualizados na API: `mananciais::atualizar_dados()`

| Como manter essa base atualizada diariamente?



# GitHub Actions

| ... nos ajuda a automatizar tarefas. (fonte)

É a coisa mais útil/interessante que eu aprendi nos últimos meses 🚀





# Conhecimentos prévios

Git e GitHub 

Pacotes em R 

Testes unitários  \*

- [Nos slides finais](#), tem referências de cada um destes assuntos! 

\*(caso queira executar os testes, e fazer a avaliação da cobertura dos mesmos)



# GitHub Actions - Eventos (events)

**Eventos:** Um evento é uma atividade que aciona um fluxo de trabalho.

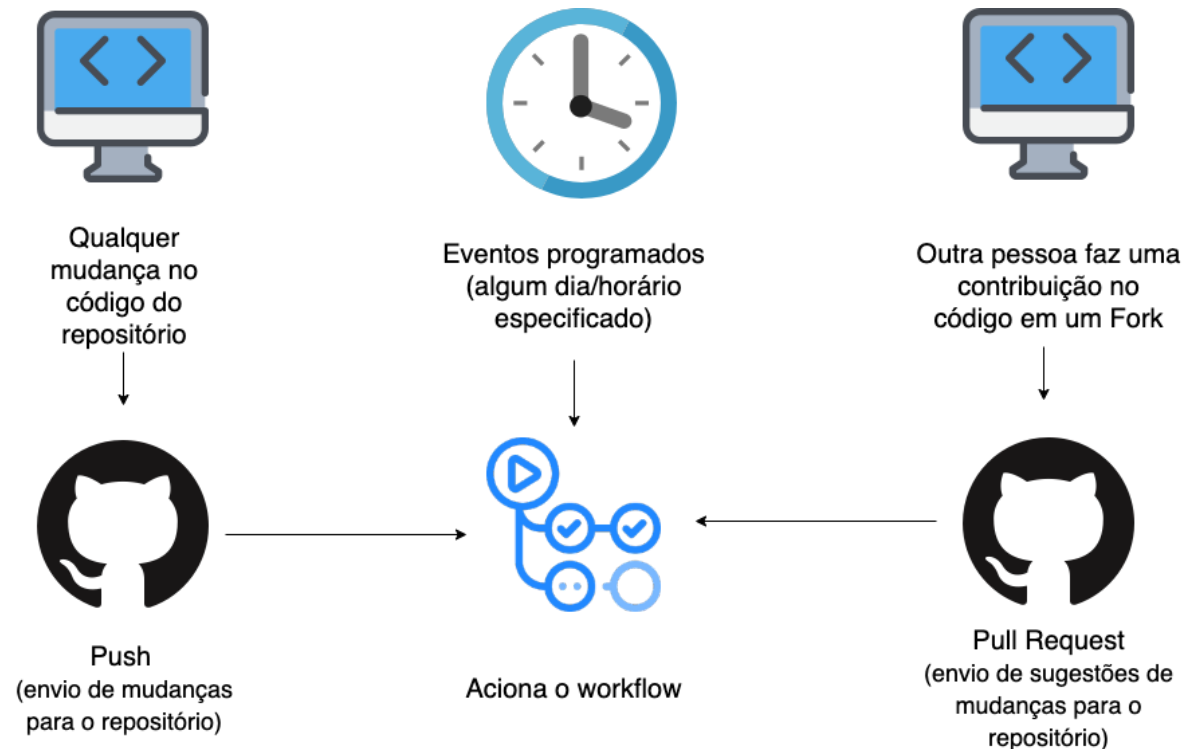


Diagrama feito por Beatriz Milz utilizando o site [draw.io](https://draw.io).



# GitHub Actions - Trabalhos (jobs)

**Trabalho:** O trabalho usa etapas para controlar a ordem em que as ações são executadas

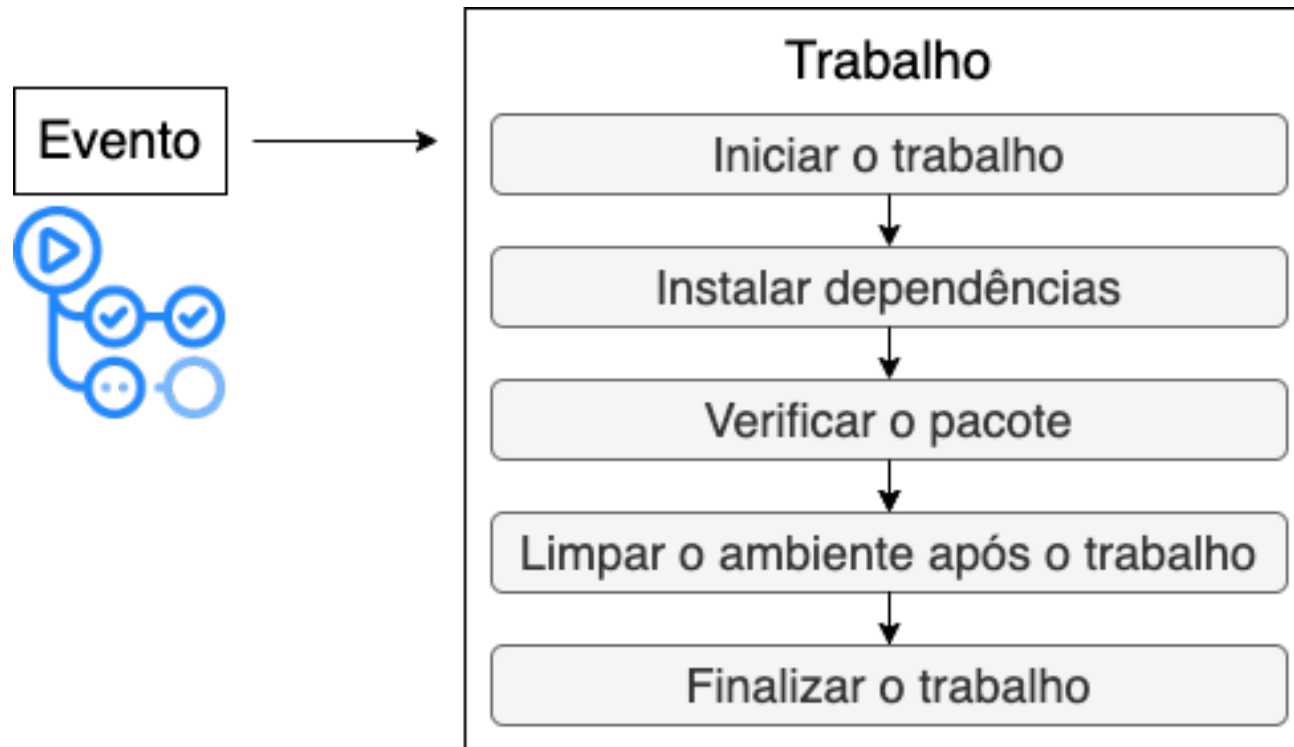


Diagrama feito por Beatriz Milz utilizando o site [draw.io](https://draw.io).

# GitHub Actions - Fluxos de trabalho (workflows)

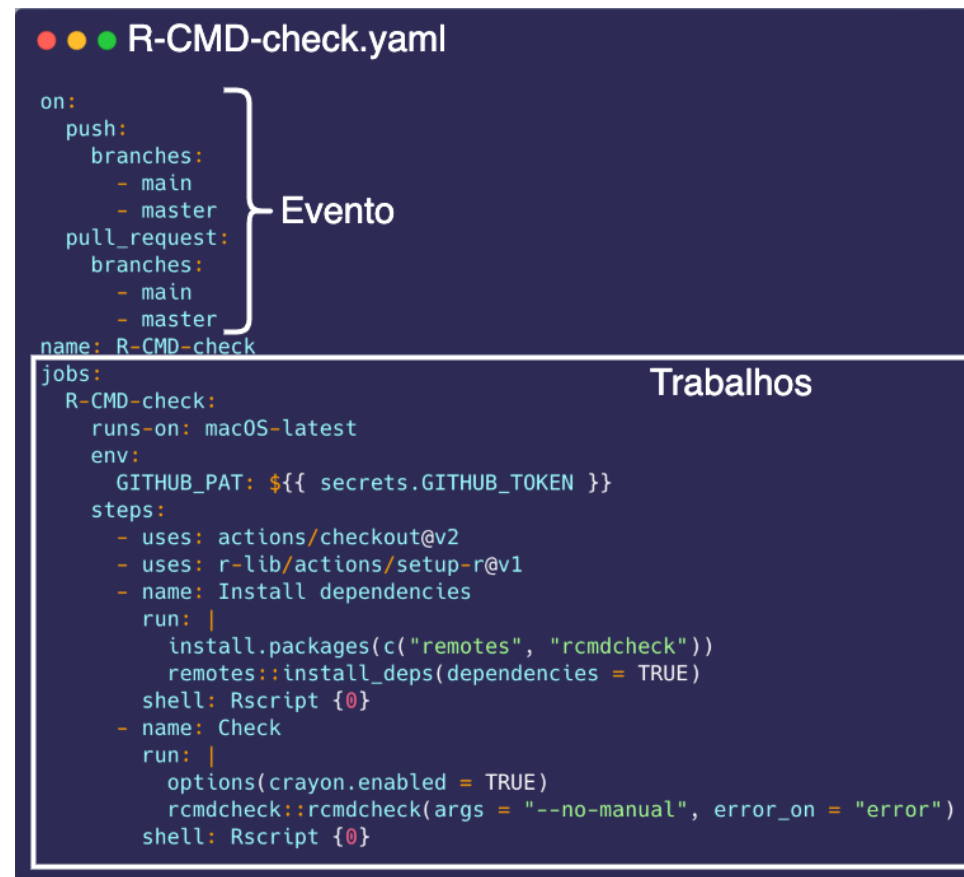


Diagrama feito por Beatriz Milz utilizando o site [draw.io](https://draw.io) e [Carbon](https://carbon.now.sh). Código de [r-lib/actions](https://github.com/r-lib/actions).



# Exemplo com GHA

## Atualização da base de dados

- Confira o workflow no [Actions](#)
- Como configurar? Adapte um dos exemplos disponíveis no [r-lib/actions](#)
- Evento:
  - Todos os dias, por volta de 9h15 da manhã + Sempre que houver um novo push no repositório
- Trabalho:
  - [Esse código](#) será executado: verifica se é necessário atualizar a base, e se for: atualiza a base, disponibiliza a mesma no pacote e exporta em [.csv](#) e [.xlsx](#). Atualiza o README e atualiza a página inicial do site do pacote.



# Exemplo com GHA

## Checagem do pacote e execução de testes

- Confira o workflow no [Actions](#)
- Como configurar? A função abaixo irá configurar utilizando o código disponível no [r-lib/actions](#).

```
usethis::use_github_action("check-release")
```

- Evento:
  - Sempre que houver um novo push no repositório
- Trabalho:
  - Similar ao `devtools::check()`: vai verificar o pacote e também executar todos os testes.



# Exemplo com GHA

## Avaliação da cobertura de testes

- Confira o workflow no [Actions](#)
- Como configurar? A função abaixo irá configurar utilizando o código disponível no [r-lib/actions](#).

```
usethis::use_github_action("test-coverage")
```

- Evento:
  - Sempre que houver um novo push no repositório
- Trabalho:
  - Executa `covr::codecov()` e faz a avaliação de cobertura de testes unitários.

# Conclusões

- Não é algo simples de aprender, mas facilita nosso trabalho :)
- Veja a palestra do [Julio Trecenti](#), ele mostrará outros exemplos com GHA!
- Caso tenha dúvidas, envie neste fórum:  
<https://discourse.curso-r.com/>
  - Antes de postar a dúvida, recomendo ler:  
[Dicas para escrever uma boa pergunta](#)



- É normal não acertar de primeira :) Provavelmente vai dar errado da primeira, segunda, terceira, ... , décima vez. Perseverança!





# Referências/recomendações de consulta

- Git e GitHub
  - Materiais da R-Ladies SP sobre a Hacktoberfest 2020
  - Apresentação sobre pacotes em R da Curso-R - slides sobre Git e GitHub
- Pacotes
  - Apresentação sobre pacotes em R da Curso-R
  - Capítulo 2 - Livro R Packages
  - Livro Zen do R
- GitHub Actions
  - Documentação do GitHub Actions
  - Exemplos de workflows no repositório r-lib/actions
- Testes unitários
  - Pacote {testthat}
  - Pacote {covr}

# Agradecimentos





# Muito obrigada!



Slides criados usando os pacotes em R:

**xaringan**

[gadenbuie/xaringanthemer](#)  
[eventoserooficial/templates.SER](#)

O chakra vem de [remark.js](#), **knitr**, e R Markdown.



[🏠 beatrizmilz.com](#)

[🐦 @BeaMilz](#)

[✉ milz.bea@gmail.com](#)