

WorkFlow

Manoel

2025-08-28

Introdução

- ▶ Em um projeto, a menor parte do tempo é utilizada fazendo a análise dos dados
- ▶ Maior parte é gasta na preparação dos dados e na comunicação dos resultados.
- ▶ Objetivo:
 - ▶ Como organizar um workflow para garantir replicabilidade.
 - ▶ Como organizar a parte empírica de um paper.

Organização do Workflow em R

- ▶ Pastas e convenções:

```
/raw_data: dados originais, nunca modificados.  
/transformed_data: dados limpos, prontos para análise.  
/scripts: código R (01_clean.R, 02_analysis.R etc.).  
/outputs: tabelas, figuras, modelos exportados.  
/paper: texto em RMarkdown/Bookdown.
```

- ▶ Pipeline ideal: uso de targets ou make para documentar dependências.
- ▶ Pipeline pragmático: rodar cada script manualmente, na ordem da documentação
- ▶ Boas práticas: nomes consistentes de arquivos; manter scripts curtos e modulares.

Escrevendo com RMarkdown e Zotero

- ▶ RMarkdown: integra código, tabelas e texto.
- ▶ Elimina inconsistências.
- ▶ Referências:
- ▶ Exportar biblioteca do Zotero em .bib.
- ▶ Citar com [@autor_ano].
- ▶ Vantagem: documento sempre atualizado com resultados da análise.

Replication Package

- ▶ README.md: Título do paper; instruções de reprodução (versão do R, pacotes, ordem dos scripts); descrição das bases (com links para fontes públicas).
- ▶ Template: https://social-science-data-editors.github.io/template_README/ Exemplos de README.md: <https://dataverse.harvard.edu/file.xhtml?fileId=10247078&version=1.0&toolType=PREVIEW>
- ▶ <https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/NDPXND/TCM8JI&version=1.0>
- ▶ Dados (se possível, ou instruções de acesso); Scripts de análise; Arquivo do paper (Rmd + PDF).
- ▶ Boas práticas: sempre rodar `sessionInfo()` no final.
- ▶ Replication Guidelines PA:
https://www.cambridge.org/core/services/aop-file-manager/file/678f5f7189ee789cdb3a1df2/replication-guidelines-2025-v1.pdf?utm_source=chatgpt.com

Repositório de Dados

- ▶ Dataverse é o repositório padrão em ciência política
- ▶ Plataforma padrão em ciência política e outras áreas.
- ▶ Permite DOI de banco de dados ou replication package, garante preservação e facilita replicação.
- ▶ No Brasil, existe o repositório da Scielo (Scielo Data).

Organização da Parte Empírica do Paper

- ▶ Ponto central: convencer que a evidência é compelling.
- ▶ Dados (5% do conteúdo do paper)
- ▶ Estratégia de Identificação (se causal)/Métodos. 10-15% do conteúdo do paper.
- ▶ Resultados (20% do conteúdo do paper)

Dados (detalhes)

- ▶ Dados: fontes, variáveis, tratamento.
- ▶ Destaque variáveis mais relevantes. Variável resposta tem precedência. Controles os menos relevantes.
- ▶ Discuta quaisquer limitações dos dados
- ▶ Questões menores vão para notas de rodapé
- ▶ Apresente estatística descritiva das principais variáveis
- ▶ Apresente gráficos que descrevam a principal relação estimada no modelo

Métodos (detalhes)

- ▶ Por que o leitor deve acreditar em seus achados?
- ▶ Descreva o desenho de pesquisa (experimento, RDD, Dif in Dif etc.)
- ▶ Modelo estatístico (equação de regressão)
- ▶ Suposições de identificação causal
- ▶ Ameaças à identificação causal
- ▶ Indique quais testes de robustez (mas não apresente os resultados)

Resultados (detalhes)

- ▶ Resultados principais: apresentados de forma clara (tabela + figura).
- ▶ Use nomes de variáveis fáceis de interpretar. Nada de x_1 ou *educ*. Use educação.
- ▶ Concentre nas variáveis relevantes. Controles podem não ser apresentados.
- ▶ Após resultados principais, apresente testes de robustez.
- ▶ Eventualmente testes suplementares no apêndice.

Conclusão e Dicas Finais (2 min)

- ▶ Organização é parte da ciência.
- ▶ Transparência gera credibilidade.
- ▶ A parte empírica deve ser lida como evidência convincente a favor do argumento.