

Reprodutibilidade em Pesquisa

Amanda Pena

11/03/2025

Grupo de Estudos em Políticas Públicas (GEPP)



Introdução

Princípios Básicos de Programação Estatística

Organização do projeto

Carregamento de Pacotes

Importar e Limpar os Dados

Análise de Dados

Execução Completa do Projeto

Introdução

Clique aqui para assistir:
**Será que há crise da
possibilidade de reprodução na
ciência?**


- Replicabilidade vs. Reprodutibilidade
- Reprodutibilidade: Obter os mesmos resultados com os mesmos dados e código.
- Replicabilidade: Obter resultados semelhantes com novos dados e métodos.
- Como isso se aplica às ciências sociais aplicadas? Garantir a correção e confiabilidade dos resultados.
- O código em R pode ser utilizado como uma espécie de mapa, que mostra como você chegou nos seus resultados.
- Código em R como Produto.
- E por que em R? Por que num software como o Excel você faz alterações diretas nos dados, já no R você constrói as instruções de como pegar o dado original e produzir a análise final.

Repositório de Reprodutibilidade do Banco Mundial

Reproducible Research Repository

Search in 209 packages

Search

 Browse Catalog

Getting started

The Reproducible Research Repository is a one-stop shop for reproducibility packages associated with World Bank research. The catalogued packages provide the analytical scripts, documentation, and, where possible, the data needed to reproduce the results in the associated paper.

Read more

Featured



Policy Research Working Papers

This collection is a set of reproducibility packages associated with papers published in the World Bank's Policy Research Working Paper (PRWP) series.



Journal Articles

This collection is a set of reproducibility packages associated with papers published by World Bank staff or consultants in academic journals.



World Bank Reports & Flagships

This collection is a set of reproducibility packages associated with World Bank reports and flagship publications.

Princípios Básicos de Programação Estatística

Princípios Básicos de Programação Estatística (Parte 1)

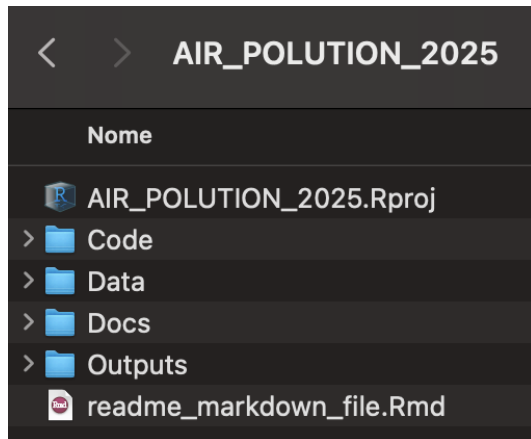
1. Automatize tudo que possa ser automatizado.
2. Escreva um único script que execute todo o processo do início ao fim.
3. Armazene códigos e dados sob controle de versão.
4. Organize os diretórios de acordo com suas funções específicas.
5. Organize os diretórios por função, separando arquivos de entrada e saída, e garantindo portabilidade entre diferentes ambientes.
6. Armazene dados limpos em tabelas com chaves únicas e não faltantes.

10. Abstraia para eliminar redundância e melhorar a clareza. Caso contrário, evite abstrações desnecessárias.
 - No contexto de programação, abstrair significa criar funções ou módulos que encapsulam funcionalidades comuns, permitindo reutilização do código.
11. Escreva códigos claros e autoexplicativos, utilizando abstrações apenas quando realmente melhorarem a clareza e evitarem redundâncias.
12. Utilize nomes de variáveis que sejam intuitivos e fáceis de entender.
13. Verifique manipulações algébricas e teste o script para garantir que os resultados sejam precisos.

Organização do projeto

- Estrutura de diretórios:
 - **data/** - Dados originais e limpos
 - **scripts/** - Scripts R para coleta, limpeza e análise
 - **results/** - Resultados e gráficos
 - **docs/** - Documentação e relatórios

Organização do projeto



Carregamento de Pacotes

Script para Carregamento de Pacotes

```
# Carrega pacotes básicos necessários, instalando se ausentes  
pacotes <- c("dplyr", "readr", "ggplot2") # Pacotes essenciais  
  
if (!require("pacman")) install.packages("pacman")  
pacman::p_load(pacotes, character.only = TRUE)
```

Importar e Limpar os Dados

Script para Importar a Base de Dados

```
# Importando a base de dados que vamos utilizar em csv -----  
# Caminho do arquivo com a função here  
path <- here("data", "raw", "consulta_cand_2024_BRASIL.csv")  
# Usando read_delim com o caminho obtido pelo here  
candidatos_2024 <- read_delim(path,  
                                delim = ";",  
                                escape_double = FALSE,  
                                locale = locale(encoding = "ISO-8859-1"),  
                                trim_ws = TRUE)
```


Script para Limpeza de Dados

```
# Limpando e preparando os dados para os cálculos -----
candidatos_cleaned <- candidatos_2024 %>%
  # A função filter é usada para selecionar apenas as linhas onde SG_UF é igual a
  # O pipe `%>%` passa o resultado (as linhas filtradas) para a próxima operação.
  filter(SG_UF == "ES") %>%
  # A função select é usada para escolher as colunas desejadas.
  select(SG_UE, DS_CARGO, SQ_CANDIDATO, CD_GENERO, DS_GENERO)

# Salvar a base limpa em um arquivo CSV.
write.csv(candidatos_cleaned,
          file = here("data", "clean", "candidatos_cleaned.csv"),
          row.names = FALSE)
```

Análise de Dados

Script para Análise de Dados

```
# Análise de mulheres candidatas ao cargo de prefeita -----
candidaturas_prefeito <- candidatos_cleaned %>%
  filter(DS_CARGO == "PREFEITA")
# Total de candidaturas a prefeita
total_candidaturas <- nrow(candidaturas_prefeito)
# Filtrando as mulheres candidatas a prefeita
mulheres_prefeito <- candidaturas_prefeito %>%
  filter(DS_GENERO == "FEMININO")
# Total de mulheres candidatas a prefeita
total_mulheres <- nrow(mulheres_prefeito)
# Calculando a proporção de mulheres candidatas a prefeita
proporcao_mulheres <- total_mulheres / total_candidaturas
cat("Proporção de mulheres candidatas a prefeita: ", proporcao_mulheres, "\n")
```

Execução Completa do Projeto

Script Principal para Rodar Todo o Projeto

```
#-----  
##Project Name: Workshop GEPP  
##Description: Main Script for reproducibillity package  
##Date: 10/03/2025  
##Author: Amanda Pena  
#-----  
source(here("scripts", "01_carregar_pacotes.R")) # 1. Carregamento de Pacotes  
source(here("scripts", "02_importar_dados.R")) # 2. Coleta e Limpeza de Dados  
source(here("scripts", "03_limpeza_dados.R"))  
source(here("scripts", "04_analise_dados.R")) # 3. Análise de Dados  
# 5. Gerar Relatório/Trabalho final. Essa parte será ensinada amanhã.
```

Referências

- Bjärkefur, K., de Andrade, L. C., Daniels, B., & Jones, M. R. (2021). *Development research in practice: The DIME analytics data handbook*. World Bank Publications.
- Gentzkow, M., & Shapiro, J. M. (2014). *Code and Data for the Social Sciences: A Practitioner's Guide*. University of Chicago mimeo.
- Innovations for Poverty Action. (2015). *Best Practices for Data and Code Management*. Recuperado de <https://poverty-action.org/sites/default/files/publications/IPA-Best-Practices-for-Data-and-Code-Management-Nov-2015.pdf>
- The Abdul Latif Jameel Poverty Action Lab. (n.d.). *Research Transparency and Reproducibility*. Recuperado de <https://www.povertyactionlab.org/page/research-transparency-and-reproducibility>

Obrigada!
aamandacarla@gmail.com

