

Metodologias Informacionais com R

Módulo I e II: Revisão

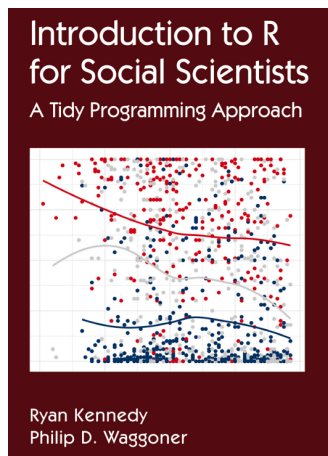
Telmo dos Santos Klipp (telmo.klipp@inpe.br)

Informações Gerais sobre o Curso

- Materiais disponibilizados via **Classroom**;
- O aprendizado requer a prática que será constante nas aulas;

Bibliografia Básica:

- Kennedy, R., & Waggoner, P. D. (2021). Introduction to r for social scientists: a tidy programming approach. CRC Press.

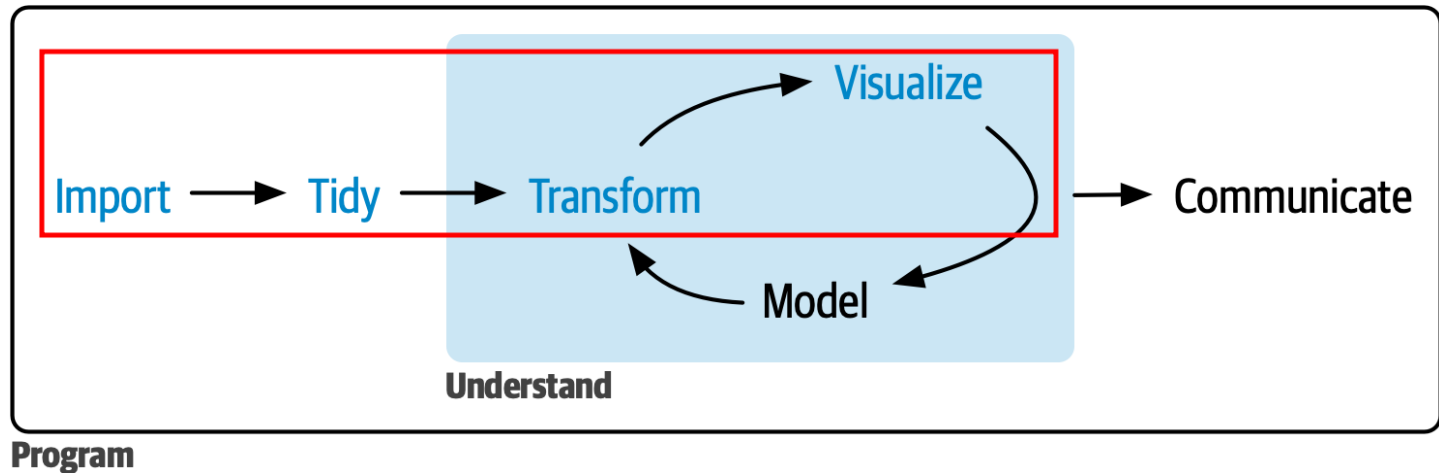


Bibliografia Complementar:

- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). R for data science (2e): import, tidy, transform, visualize, and model data. "O'Reilly Media, Inc.". Disponível em: <https://r4ds.hadley.nz/>. Acesso em: 14 de junho, 2023. (Online)
- Damiani, A. et. al., (2022). Ciência de Dados em R. Curso-R. Disponível em: <https://livro.curso-r.com>. Acesso em: 12 de maio, 2023. (Online)
- de Aquino, J. A. (2014). R para cientistas sociais. Editora da UESC (editus). Disponível em: <http://www.uesc.br/editora/>. Acesso em: 12 de maio, 2023.
- de Oliveira, P. F., Guerra, S., McDonnell, R. (2018). Ciência de Dados com R: Introdução. Editora IBPAD. Disponível em: <https://cdr.ibpad.com.br/index.html>. Acesso em: 12 de maio, 2023. (Online)

Ciclo da ciência dos dados – "whole game"

Créditos: Adaptado de Wickham, et al (2023).



Vamos fazer gestão, manipulação e visualização de dados considerando **R project**, *Data Wrangling*, observação de tabelas no RStudio e geração de gráficos.

Fundamentos do R – projeto

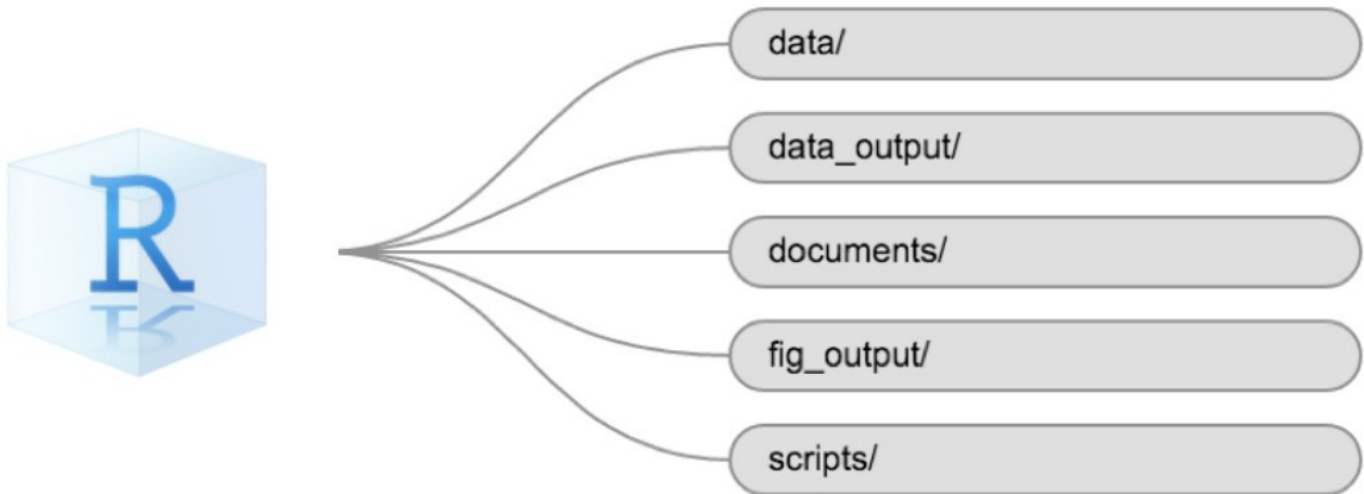
Projetos permitem organização, seja em algum trabalho, seja na vida no geral. No R isso não é diferente. **O que pretendemos organizar?**

- Diferentes tipos de arquivos (ex: .R, .csv, .png ...).
- Diferentes tipos de dados, salvos em arquivos.
- Arquivos especiais do R (ex: .RData e .RHistory ...).
- Outros arquivos com informações relativas ao projeto.

Ganhos?

- Gestão e controle das etapas do projeto.
- Gestão e controle dos recursos do projeto (ex: dados, *R scripts*).
- Produtividade e gerenciamento de tempo.

Fundamentos do R – projeto



Créditos: [Erin Becker \(2023\)](#).

Tidyverse

Coleção de pacotes que possuem filosofia de design, gramática e estrutura de dados em comum e, permitem trabalho conjunto, clareza de código, reprodutibilidade, dentre outros benefícios.



Gestão e Manipulação de dados

Crie um **R project**. Vamos trabalhar com dados da PNAD contínua do IBGE relativos à população brasileira. Para isso precisaremos:

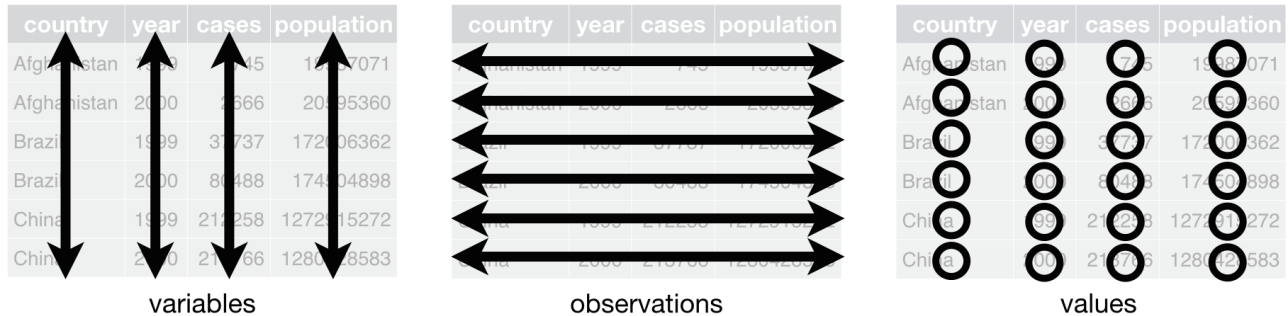
- Criar uma pasta para dados e outra para *scripts* na pasta principal (raiz) do **R project**.
- Baixar os dados (use a tecla **P** para obter informações).
- Obs: Parte dos dados que serão usados foram baixados por meio do pacote **basedosdados**. Existem outras alternativas de dados disponíveis nessa base, sendo necessário seguir alguns passos para o uso do pacote conforme o **tutorial de uso** e **documentação específica**.

Os principais pacotes do **tidyverse** para ler e manipular dados são **readr**, **dplyr** e **tidyr**.

Gestão e Manipulação de dados

- Leia os dados da população brasileira.
- Os dados estão no formato **tidy**? Em caso negativo arrume-os.

Créditos: Wickham, et al (2023).



The diagram shows three tables representing the same data from different perspectives. The first table, labeled 'variables', has columns for country, year, cases, and population, with vertical double-headed arrows indicating the relationship between variables. The second table, labeled 'observations', has the same columns but with horizontal double-headed arrows indicating the relationship between observations. The third table, labeled 'values', has the same columns but with circles around each cell, indicating the relationship between individual values.

country	year	cases	population
Afghanistan	1999	18145	19787071
Afghanistan	2000	2366	20595360
Brazil	1999	31737	172406362
Brazil	2000	80488	174504898
China	1999	211258	1272015272
China	2000	216736	128042583

variables

country	year	cases	population
Afghanistan	1999	18145	19787071
Afghanistan	2000	2366	20595360
Brazil	1999	31737	172406362
Brazil	2000	80488	174504898
China	1999	211258	1272015272
China	2000	216736	128042583

observations

country	year	cases	population
Afghanistan	1999	18145	19787071
Afghanistan	2000	2366	20595360
Brazil	1999	31737	172406362
Brazil	2000	80488	174504898
China	1999	211258	1272015272
China	2000	216736	128042583

values

"Conjunto de dados organizados (tidy) são todos iguais, mas todo conjunto bagunçado é bagunçado à sua maneira."

— Hadley Wickham

- Salve em novos arquivos.

Visualização de dados

Gere visualizações dos dados:

- Comece por gráficos básicos usando o **ggplot2**.
- Tente personalizar os gráficos.
- Use o operador **pipe**.
- Tente produzir um plot interativo com **plotly**.

Salve os passos em um **R script**.

Salve as figuras produzidas em uma nova pasta.

Referências

Wickham, H., M. Çetinkaya-Rundel, and G. Grolemund (2023). *R for data science (2e): import, tidy, transform, visualize, and model data*. Disponível em: <https://r4ds.hadley.nz/>, Acesso em: 14 de junho de 2023. O'Reilly Media, Inc.

Becker, E. (2023). *R for Social Scientists*. Data Carpentry. Disponível em: <https://datacarpentry.org/r-socialsci/index.html>, Acesso em: 14 de junho de 2023.

Metotologias Informacionais com 

Muito Obrigado pela Atenção!