

Software livre R

Pacotes Tidyverse

Michelle Almeida/Maurício Lordêlo

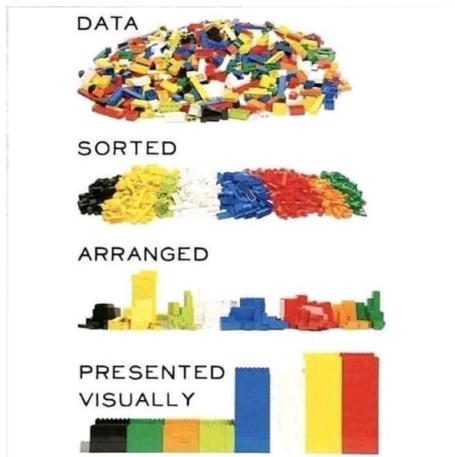
Universidade Estadual de Feira de Santana - Bahia - Brasil

17 de setembro de 2022

Manipulação e visualização de dados

- Entendimento dos dados
- Conhecimento do processo e suas etapas
- Dominar a tecnologia para execução
- Linguagens de programação oferecem uma série de vantagens: reproduzível, extensível, escalonável, etc
- O ambiente R: desde o princípio oferece recursos para manipulação de dados
- O formato *data.frame* é a base para organizar os dados tabulares.
- R sempre possuiu recursos para importar, transformar, modificar, filtrar, agregar os *data.frames*
- Porém, existem “algumas imperfeições”
- *Tidyverse* entrou para apresentar melhorias

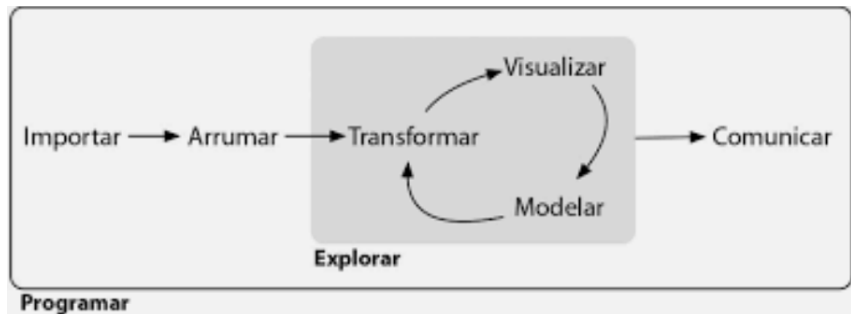
Processo de transformação de dados em informação



Fonte:

(<https://twitter.com/africadatasch/status/1301026743516639237>)

Ciência de dados (*Data Science*)



Pacotes *Tidyverse*

- Conjunto de pacotes (*framework*) para manipulação de dados
- Funciona como um ambiente integrado dos pacotes
- Oferece uma reimplementação e extensão das funcionalidades para manipulação e visualização.
- É uma coleção 8 de pacotes R (núcleo) que operam em harmonia.

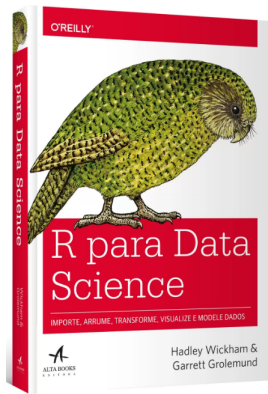
```
library(tidyverse)
tidyverse_packages()

## [1] "broom"      "cli"
## [3] "crayon"     "dbplyr"
## [5] "dplyr"      "forcats"
## [7] "ggplot2"    "haven"
## [9] "hms"        "httr"
## [11] "jsonlite"   "lubridate"
## [13] "magrittr"    "modelr"
## [15] "pillar"     "purrr"
## [17] "readr"      "readxl"
## [19] "reprex"     "rlang"
## [21] "rstudioapi" "rvest"
## [23] "stringr"    "tibble"
## [25] "tidyr"      "xml2"
## [27] "tidyverse"
```

- *Tidyverse* sozinho não tem funcionalidade específica

Principal referência

Apresenta o uso coordenado de todos os pacotes: importar, arrumar, transformar, visualizar, modelar e comunicar



(<https://r4ds.had.co.nz/>)

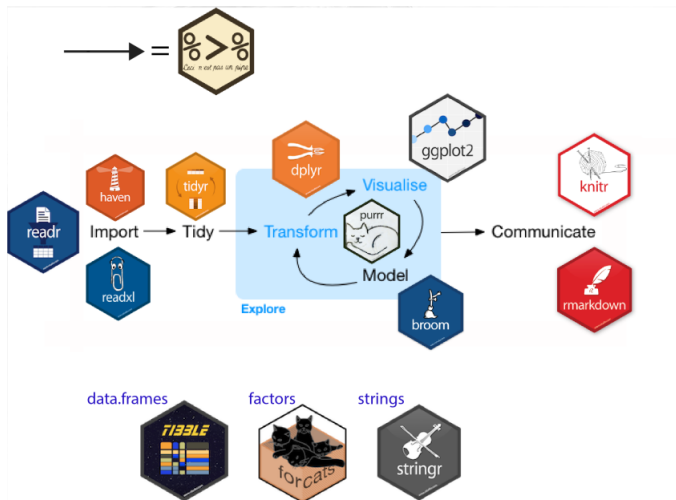
Pacotes *Tidyverse*

- Eles foram planejados e construídos para trabalhar em conjunto.
- Possuem gramática, organização, filosofia e estruturas de dados mais clara.
- Maior facilidade de desenvolvimento de código e portabilidade.
- Outros pacotes harmonizam muito bem com o *tidyverse*.
- Pacotes: <https://www.tidyverse.org/packages/>.

Pacote *tidyverse*



Ciência de dados



- Reimplementação do *data.frame*
- Apresentação de forma mais resumida e informativa
- Mais consistente para seleção e modificação de conteúdo
- Mais fácil conversão de outros formatos para *tibble* .
- Colunas/células podem representar objetos mais complexos.
- Documentação <https://tibble.tidyverse.org/>.

- tem recursos para importação de dados retangulares na forma de texto pleno.
- Importação de dados de forma mais rápida
- Leitura/escrita de dados tabulares: csv, tsv, fwf.
- Recursos "inteligentes" que determinam tipo de variável,
- Exemplo: importar campos de datas como datas
- Documentação <https://readr.tidyverse.org/>.

- Pacote para arrumação
- dados no formato *tidy* (tabular): cada variável está em uma coluna, cada observação é uma linha e cada valor é uma célula.
- Mudar disposição dos dados: *long* (empilhar) e *wide*(esparramar)
- Lidar com valores ausentes
- Partir/concatenar variáveis
- Documentação <https://tidyr.tidyverse.org/>.

- Mais importante: recursos principais para transformação de dados
- É a gramática para manipulação de dados.
- Tem um conjunto consistente de verbos para atuar sobre tabelas
- *mutate()*, *select()*, *filter()*, *arrange()*, *summarise()*, *slice()*, *rename()*, etc.
- Sufixos: *_at()* , *_if()* , *_all()*, etc.
- Agrupar e desagrupar: *group_by()* e *ungroup()*
- Junções: *inner_join()*, *full_join()*, *left_join()* e *right_join()*
- Documentação <https://dplyr.tidyverse.org/>.

- Criação de gráficos baseado no *The Grammar of Graphics* (Wilkinson et. al, 2013)
- Documentação <https://ggplot2.tidyverse.org/>.

Outros pacotes

- *purrr*: programação funcional
<https://purrr.tidyverse.org/>
- *forcats*: para variáveis categóricas/fatores
<https://forcats.tidyverse.org/>
- *stringr*: expressões regulares (detectar, contar, partir, extrair, etc)
<https://stringr.tidyverse.org/>
- *magrittr*: operador *pipe* %>%
<https://magrittr.tidyverse.org/>
- *lubridate* e *hms*: manipulação de dados cronológicos
<https://lubridate.tidyverse.org/>
<https://hms.tidyverse.org/>