

# Dados do estudo de caso AMAZON

Luiz C. E. Rodriguez, Eric B. Gorgens, Humberto T. Menecheli Filho, Nathan O. Barreto, João V.L. M.

2023-05-12

---

## Floresta*R*

O projeto **Floresta*R*** reúne professores, estudantes e profissionais dedicados à redação da série **Floresta*R*** de livros sobre o uso do **R** como ferramenta de análise de dados florestais.

---

## Sobre os autores

Os autores da série de livros **Floresta*R*** são:

- **Luiz Carlos Estraviz Rodriguez**, professor titular da Escola Superior de Agricultura Luiz de Queiroz (ESALQ) da Universidade de São Paulo
  - **Eric Bastos Gorgens**, professor do Departamento de Engenharia Florestal da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM)
  - **Humberto Tadeu Menecheli Filho**, engenheiro florestal analista de dados da ForLiDAR
  - **Nathan de Oliveira Barreto**, graduando de engenharia florestal da ESALQ/USP
  - **João Victor Lopes Marinho**, graduando de engenharia florestal da ESALQ/USP
  - **Arthur Kaufmann Sanchez**, graduando de engenharia florestal da ESALQ/USP
  - **José Jorge Monteiro Junior**, engenheiro florestal mestrando do programa de pós-graduação em Recursos Florestais da ESALQ/USP
- 

## Sobre os livros

A coleção está em constante evolução, e no momento é constituída pelos seguintes volumes:

- **Vol\_01** Análise exploratória de dados florestais
  - **Vol\_02** Inferência estatística com dados florestais
  - **Vol\_03** Amostragem para fins de inventário florestal
  - **Vol\_04** Processamento de dados LiDAR para monitoramento florestal
- 

## Conjuntos de dados

Os dados dos estudos de caso usados nos exercícios da série de livros **Floresta*R*** estão organizados em pastas:

Pasta	Conteúdo
1_AMAZON	Árvores inventariadas na Amazônia
2_RESTAU	Medições de árvores de espécies nativas coletadas em área de restauração de corredor ecológico
3_VOLIND	Árvores de <i>Eucalyptus</i> cubadas para modelagem de volume individual
4_INVENT	Medições de árvores de <i>Eucalyptus</i> em parcelas de inventário convencional
5_LIDARF	Nuvens de pontos LiDAR sobre plantio florestal para mapeamento de biomassa

## Dados do estudo de caso AMAZON

A pasta **AMAZON** contém as medições de árvores inventariadas em 2010 para fins de manejo florestal. Implantado às margens da BR 364, no município de Porto Velho (RO) em área próxima à divisa com o Acre, o plano se refere à UPA 4 (Unidade de Produção Anual número 4). Todas árvores (*censo 100%*) com mais de 40cm de diâmetro à altura do peito (DAP) nessa área foram identificadas, geolocalizadas, medidas e destinadas para compor um dos seguintes grupos:

Código	Destinação
APP	Em área de preservação permanente
CF	Corte futuro (árvore com DAP entre 40 e 50 cm)
CPL	Corte protegido por lei
PAB	Para abate
PS	Porta semente
RAR	Rara
REM	Remanescente

O acesso aos dados deste repositório é mantido aberto para maximizar acessibilidade e permitir a reprodutibilidade dos exemplos publicados nos livros da série **FlorestaR**. Os dados foram estruturados em tabelas tidy e shape para facilitar o uso das funções **tidyverse** e **sf** do **R**.

As medições na UPA4 geraram dados sobre 20.108 árvores. Esses dados foram armazenados em planilhas CSV e XLSX, e na tabela de atributos de um **shape** de pontos georeferenciados que identifica a precisa localização dessas árvores.

### . Formato CSV

O conteúdo da planilha `Censo_UPA04.csv` pode ser importado para uma *tibble* `dados`, da seguinte forma:

```
library(tidyverse)
gitOnde <- "https://github.com/FlorestaR/dados/blob/main/1_AMAZON"
gitNome <- "Censo_UPA04.csv"
gitArqv <- file.path(gitOnde, gitNome) %>% paste0("?raw=true")

dados <- read_csv(gitArqv,                               # importação
                 locale = locale(encoding = "latin1"), # garante acentuação
                 show_col_types = FALSE) %>% tibble()  # desabilita msgs

dados$dia <- dados$dia %>% as.Date("%d/%m/%Y") # garante formato date
```

`gitOnde` define a pasta dentro do repositório, `gitNome` define o arquivo de dados e `gitArqv` concatena esses dois termos com o sufixo `?raw=true` para criar o devido URL. Em seguida o URL é passado para a função `read_csv()` do pacote **readr** do **tidyverse**.

A função `read_csv()` do pacote **readr** permite que um usuário R conectado à internet leia diretamente os dados mantidos em repositórios github.

### . Formato XLSX

Para ler a versão `xlsx` dos mesmos dados, uma alternativa é usar a função `import()` do pacote **rio**:

```
library(tidyverse)
gitOnde <- "https://github.com/FlorestaR/dados/blob/main/1_AMAZON"
gitNome <- "Censo_UPA04.xlsx"
gitArqv <- file.path(gitOnde, gitNome) %>% paste0("?raw=true")

library(rio)

dados <- import(file = gitArqv) %>% tibble() # importação
dados$dia <- dados$dia %>% as.Date()        # garante formato date
```

Nesse caso, a acentuação nos caracteres especiais já é importada corretamente, havendo necessidade apenas de garantir que a coluna `dia` registre os valores como `date`.

### . A *tibble* dados

A função `str()` nos permite ver detalhes da estrutura da *tibble* dados:

```
str(dados)

## tibble [20,108 x 17] (S3: tbl_df/tbl/data.frame)
## $ codARV : num [1:20108] 40100001 40100002 40100003 40100004 40100005 ...
## $ codUT : num [1:20108] 1 1 1 1 1 1 1 1 1 1 ...
## $ codUCA : num [1:20108] 1 1 1 1 1 1 1 1 1 1 ...
## $ codLIN : num [1:20108] 1 1 1 1 1 1 1 1 1 1 ...
## $ X : num [1:20108] 5 23 43 47 40 14 37 22 15 31 ...
## $ Y : num [1:20108] 13 11 4 7 12 21 25 25 40 36 ...
## $ dia : Date[1:20108], format: "2010-03-03" "2010-03-03" ...
## $ nome : chr [1:20108] "taxi" "caucho" "ucuuba" "pamã" ...
## $ sp : chr [1:20108] "Sclerolobium paniculatum" "Castilla ulei" "Virola surinamensis" "Pseudol...
## $ dap : num [1:20108] 70.8 50.1 43.5 52.8 41.1 46.1 50.2 91.7 57 73 ...
## $ altura : num [1:20108] 15 12 17 10 15 12 15 18 10 12 ...
## $ notaFUST: num [1:20108] 1 2 1 1 2 1 1 1 2 2 ...
## $ notaCIPO: num [1:20108] 1 2 2 2 1 2 2 2 1 2 ...
## $ clasDAP : chr [1:20108] "070-080" "050-060" "040-050" "050-060" ...
## $ areaBAS : num [1:20108] 1.575 0.789 0.594 0.876 0.531 ...
## $ volume : num [1:20108] 4.13 1.66 1.77 1.53 1.39 ...
## $ destino : chr [1:20108] "REM" "PS" "CF" "RAR" ...
```

### . Formato *shape*

Os dados disponibilizados nos formatos `CSV` e `XLSX` se encontram também disponíveis na tabela de atributos do `shape` **Arvores**, com a localização geográfica de cada árvore. O *download* desse, e de outros `shapes` complementares, permite reproduzir os exercícios de exploração de dados espacializados.

Os exercícios com as árvores georeferenciadas usam funções do pacote **sf** do **R**. A função `read_sf()` desse pacote permite ler camadas vetoriais (*layers*) de informação SIG armazenadas no formato `shape`.

Para `download` dos `shapes` e leitura da tabela de árvores georeferenciadas, sugere-se o seguinte procedimento:

```
library(tidyverse)
```

```

gitOnde <- "https://github.com/FlorestaR/dados/blob/main/1_AMAZON/"
gitNome <- "shapes.zip"
gitArqv <- file.path(gitOnde, gitNome) %>% paste0("?raw=true")

tmpd <- tempdir(check = TRUE) # diretório temporário
zipf <- file.path(tmpd, "shapes.zip") # arquivo temporário

if(!file.exists(zipf)) # garante download de dados binários (wb)
  download.file(gitArqv, mode="wb", destfile = zipf)

unzip(zipf, exdir = tmpd) # shape é unzipped no diretório temporário
unlink(zipf) # deleta o arquivo zipado

library(sf)

shpArq <- paste0(tmpd, "/shapes/Arvores.shp") # shape com árvores
dadosComGeo <- sf::read_sf(shpArq) # df completo com geom
dadosSemGeo <- tibble(st_drop_geometry(dadosComGeo)) # df sem geom

dadosSemGeo %>% # Cinco castanheiras mais altas
  filter(nome=="castanheira") %>%
  select(codARV, sp, codUT, codUCA, dap, altura) %>%
  arrange(desc(altura)) %>%
  head(5) %>%
  knitr::kable(caption = "Lista das 5 maiores castanheiras")

```

Table 3: Lista das 5 maiores castanheiras

codARV	sp	codUT	codUCA	dap	altura
40401295	Bertholetia excelsa	4	6	165.6	30
40501461	Bertholetia excelsa	5	7	71.6	30
40501504	Bertholetia excelsa	5	15	92.0	30
40501700	Bertholetia excelsa	5	12	90.6	30
40300157	Bertholetia excelsa	3	13	132.2	29

Para ilustrar a vantagem de usar a versão georeferenciada de dados, apresenta-se um exemplo que exibe uma sobreposição das camadas APP e UTs com as castanheiras.

```

library(tidyverse)

gitOnde <- "https://github.com/FlorestaR/dados/blob/main/1_AMAZON"
gitNome <- "shapes.zip"
gitArqv <- file.path(gitOnde, gitNome) %>% paste0("?raw=true")

tmpd <- tempdir(check = TRUE) # diretório temporário
zipf <- file.path(tmpd, "shapes.zip") # arquivo temporário

if(!file.exists(zipf)) # garante download de dados binários (wb)
  download.file(gitArqv, mode="wb", destfile = zipf)

unzip(zipf, exdir = tmpd) # shape é unzipped no diretório temporário
unlink(zipf) # deleta o arquivo zipado

```

```

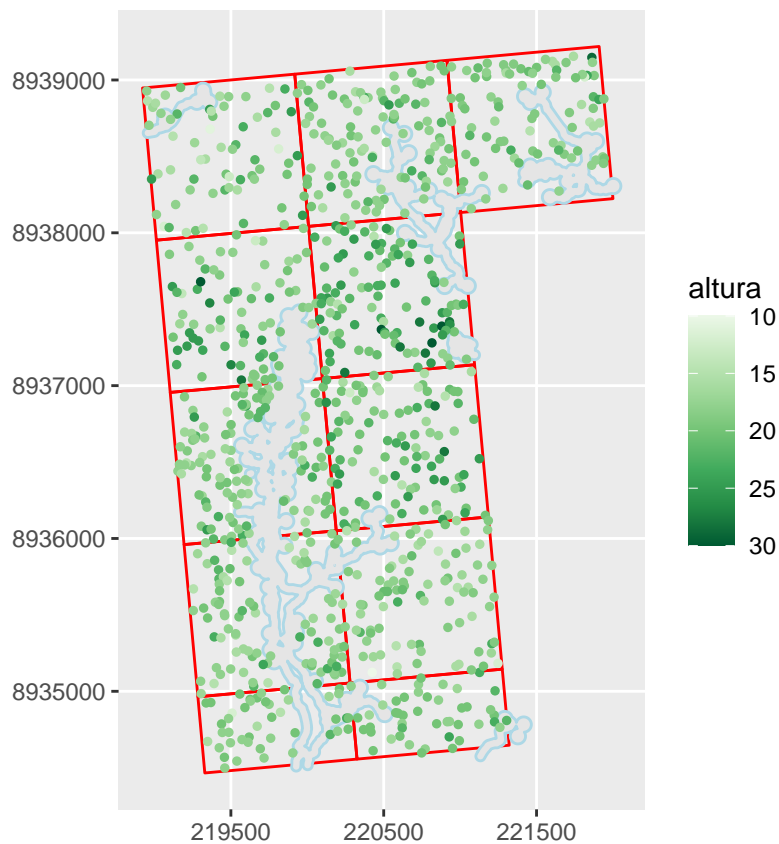
library(sf)

# leitura das camadas
dadosArv <- paste0(tmpd, "/shapes/Arvores.shp") %>% read_sf()
dadosUca <- paste0(tmpd, "/shapes/UCAs.shp") %>% read_sf()
dadosUts <- paste0(tmpd, "/shapes/UTs.shp") %>% read_sf()
dadosApp <- paste0(tmpd, "/shapes/APP.shp") %>% read_sf()
dadosHid <- paste0(tmpd, "/shapes/Hidrografia.shp") %>% read_sf()
dadosEst <- paste0(tmpd, "/shapes/Estrada.shp") %>% read_sf()
dadosRam <- paste0(tmpd, "/shapes/Ramal.shp") %>% read_sf()
dadosPat <- paste0(tmpd, "/shapes/Patios.shp") %>% read_sf()

Especie <- dadosArv %>% # novo df só com castanheiras
  filter(nome=="castanheira") %>%
  select(codARV,altura,dap,volume)

ggplot() + # plot das UTs, APPs e castanheiras (col por altura)
  geom_sf(data = dadosUts, colour = "red", fill=NA) +
  geom_sf(data = dadosApp, colour = "lightblue") +
  geom_sf(data = Especie, aes(colour = altura), size = 1) +
  scale_color_distiller(palette = "Greens", trans = "reverse") +
  coord_sf(datum=st_crs(29190)) + # Especifica sistema de coord.
  scale_x_continuous(breaks = seq(from = 218500, to = 223000, by = 1000))

```



Caso encontre alguma inconsistência nas informações deste repositório, ou queira registrar um comentário,

clique em **issues** no menu do repositório [github.com/LuizEstraviz/FlorestaR\\_dados](https://github.com/LuizEstraviz/FlorestaR_dados) e deixe a sua contribuição clicando no botão **New issue**. Agradecemos o seu interesse por esta iniciativa.

Disclaimer: Os dados são disponibilizados para fins exclusivamente educativos e não devem ser usados para outro propósito. O uso fora do contexto para o qual estão sendo disponibilizados, mesmo que apenas parcialmente, além de indevido, produzirá resultados impróprios, pois alguns dos atributos originais desses dados foram propositalmente modificados para simplificar a realidade. Os autores não se responsabilizam pelo uso indevido que, além de infringir códigos científicos e éticos de conduta, extrapole o caráter ilustrativo dos exercícios que apresentamos na série de livros **FlorestaR**.