

Aula 2

Segunda à tarde

Curadoria de dados

Como organizar os dados em diretórios no seu computador

Curadoria de dados

- Como nomear arquivos
- Como organizar os arquivos em pastas
- Diretórios de trabalho
- Projetos do RStudio

Melhores práticas para nomear arquivos

- Faça algo descritivo
- Não seja genérico
- Faça algo que tenha um tamanho razoável
- Seja consistente

Quem arquivou melhor?

- PLPP_EvaluationData_Workshop2_2014.xlsx
- MyData.xlsx
- publiclibrarypartnershipsevaluationdata
taworkshop22014CummingsHelenaMontana.
xlsx



Melhores práticas para nomear arquivos

- Arquivos devem incluir apenas letras, números, e sublinhados
- Sem caracteres especiais (%@#*?!)
- Sem espaços
- Minúscula ou intercale minúscula e maiúsculas (LikeThis)
- Nem todos os sistemas são “case sensitive”.
Assuma que **isso**, **ISSSO**, e **iSsO** são a mesma coisa.

Datas e numeração...

1. Comece com zeros para ganhar escala

001

002

009

019

999

2. Se estiver usando datas, faça no formato
AAAAMMDD

June2015 = Ruim!

06-18-2015 = Ruim!

20150618 = ótimo!

2015-06-18 = Isso tá ok também 😊

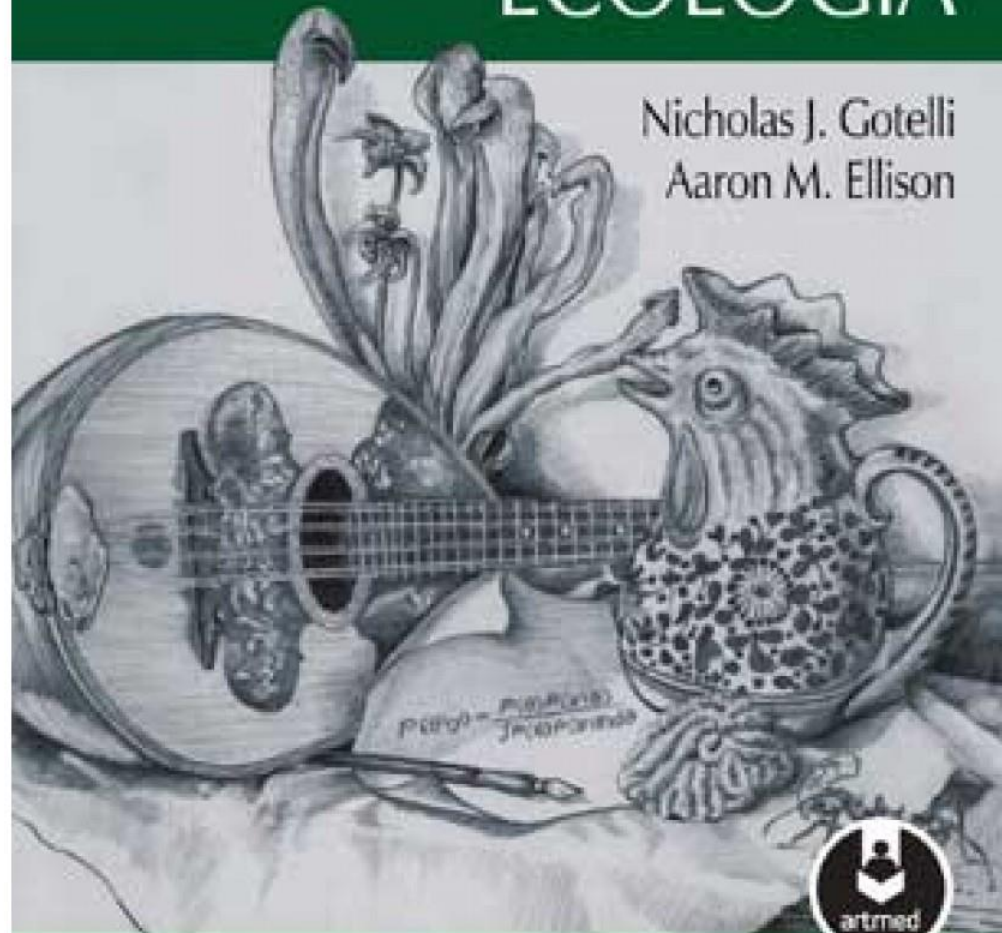
Quem arquivou melhor?

- July 24 2014_SoilSamples%_v6
- 20140724_NSF_SoilSamples_Cummings
- SoilSamples_FINAL



PRINCÍPIOS DE ESTATÍSTICA EM ECOLOGIA

Nicholas J. Gotelli
Aaron M. Ellison



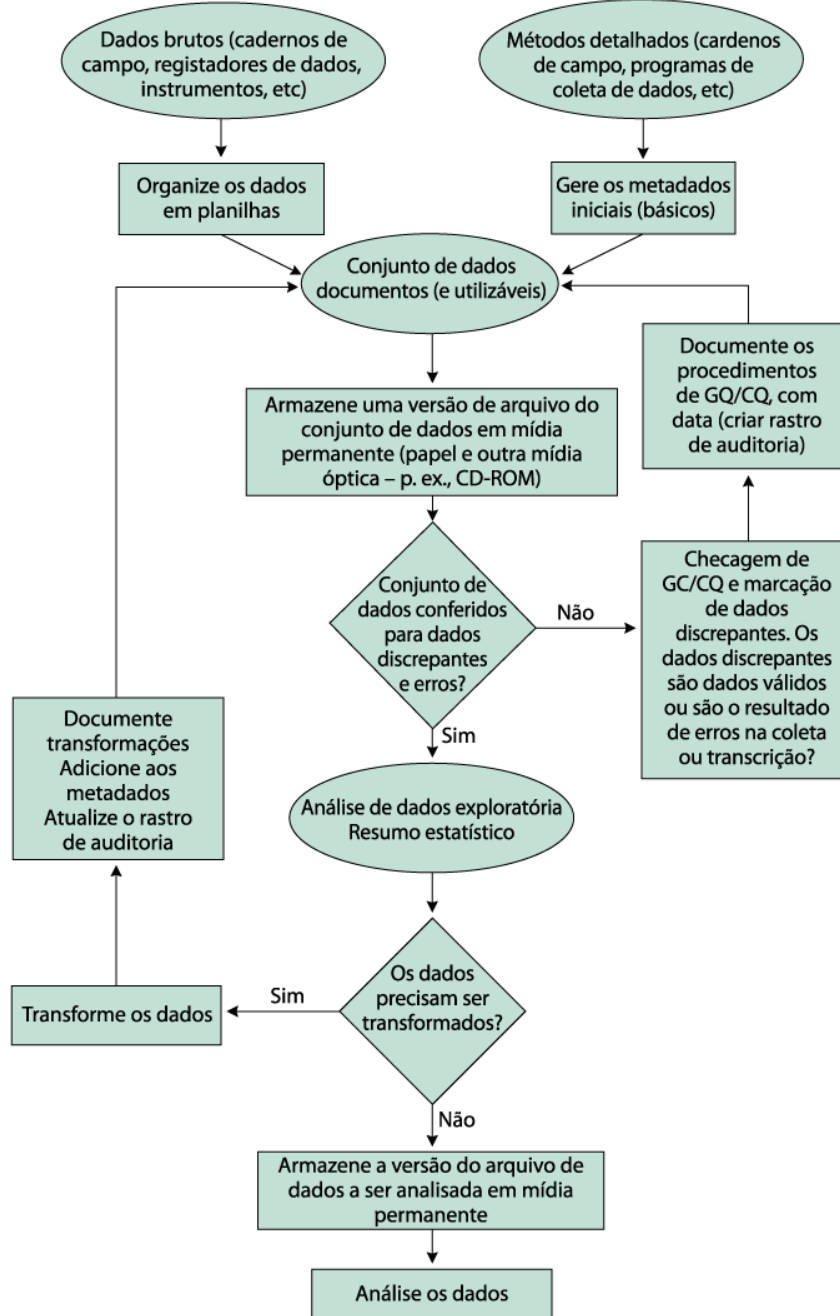
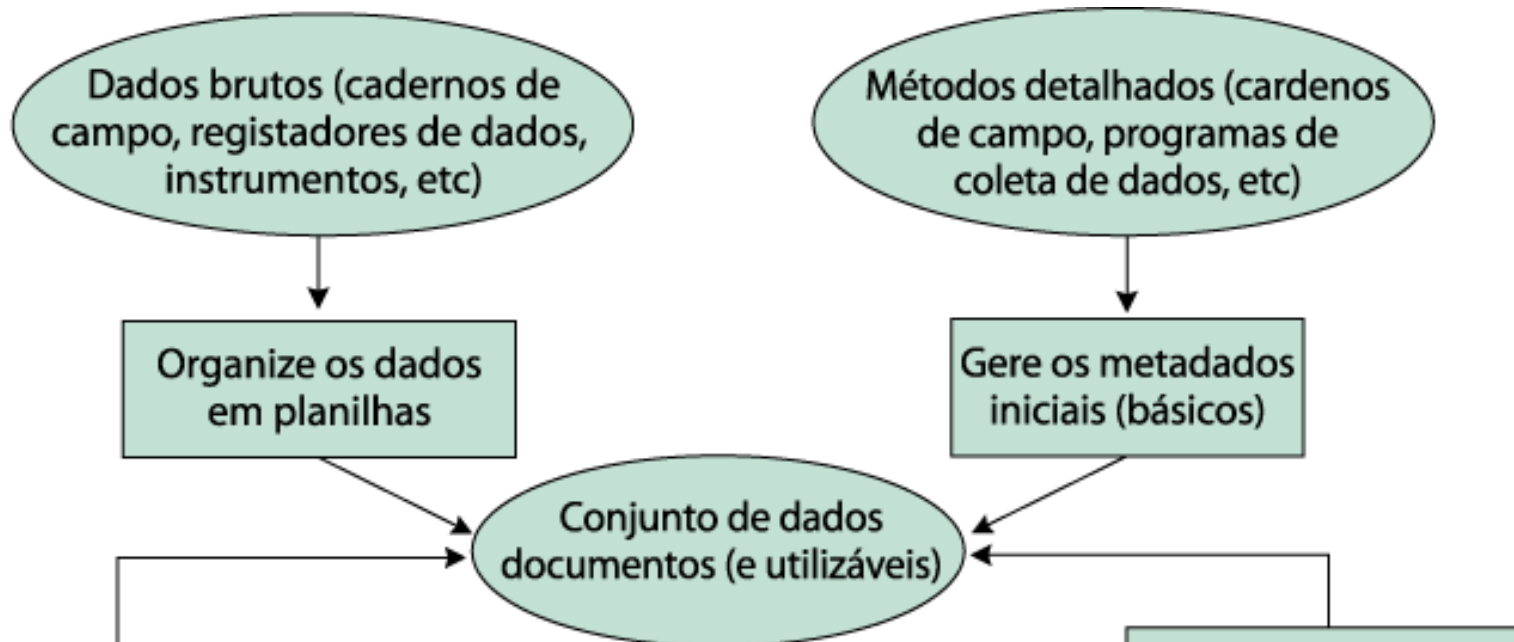


Figura 8.7 Fluxograma de gerenciamento de dados. Esse fluxograma descreve os passos para a gestão, a curadoria e o armazenamento de dados. Todos estes passos devem ser seguidos antes de a análise formal dos dados começar. A organização de metadados é especialmente importante para garantir que o conjunto de dados ainda será acessível e utilizável em um futuro distante.



Primeiros passos

- Passar informações de caderno de campo, relatório de equipamentos etc para o computador
- Organização de planilhas
 - Cada observação numa linha
 - Unidade amostral, ou pseudoréplica
 - Cada variável numa coluna
 - Cada célula um único valor

Primeiros passos

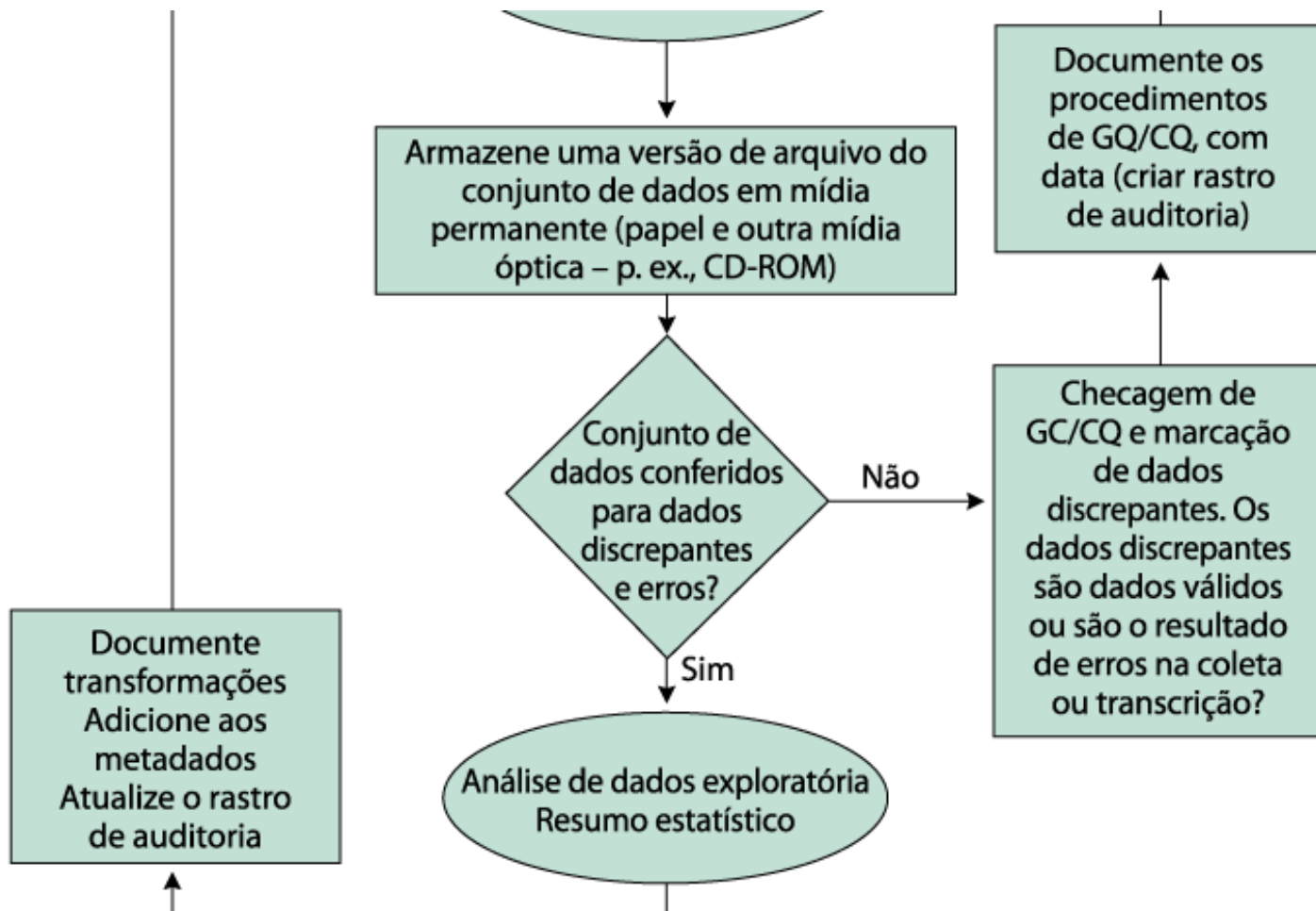
- Sempre mantenha cópias extras (backup) em fontes diferentes: computadores, ou armazenamento em nuvem
- Dados faltantes?
 - Utilize NA (Not Available)
- Evite usar espaços, use _ para separar palavras
- Assim que terminar de planilhar os dados, confira novamente (double check) e contraste com a fonte original (e.g., caderno de campo)

Primeiros passos

- NUNCA, repito, NUNCA faça gráficos junto com a planilha de dados
- Evite utilizar mais que uma aba do Excel (ou qq programa de planilhas)
- Salve o arquivo de preferência em formato .txt ou .csv (ou algum outro que seja de código aberto), evite .xls (a pessoa tem de ter o Excel instalado para abrir o arquivo)

Primeiros passos

- Dê nomes que façam sentido ao arquivo
 - 20180108 –
Composicao_especies_cerradinho_UFMS.txt
 - Planilha de espécies.xls
- Separe os arquivos em pastas para cada projeto
- Se possível comece a usar controle de versão logo no início do projeto



Metadados

- Armazene em um arquivo separado os métodos que utilizou para obter os dados, deixe esse arquivo na mesma pasta
 - Data de coleta, coletor, modelo e marca do equipamento, etc

Contents ?

Title

[Data Tables](#)

[Creators](#)

[Contacts](#)

[Associated Parties](#)

[Metadata Providers](#)

[Abstract](#)

[Keywords](#)

[Intellectual Rights](#)

[Geographic Coverage](#)

[Temporal Coverage](#)

[Taxonomic Coverage](#)

[Maintenance](#)

[Publisher](#)

[Publication Info](#)

[Methods](#)

[Project](#)

[Other Entities](#)

[Data Package ID](#)

[Check Metadata](#) ●

[Send to EDI](#)

[Send to Colleague](#)

Title

Enter a title for the data package:

Title *

Save and Continue

Reset Changes

?

About ezEML

Release 2022.08.10

ezEML is a form-based online application designed to streamline the creation of metadata in the [Ecological Metadata Language \(EML\)](#). If you're not sure what we mean by metadata, or if you've never heard of EML, please see the [Some Terminology](#) section of the [User Guide](#).

Since the EML standard was designed to handle an enormous variety of data scenarios, EML is complex and the learning curve to creating it can be steep. But many data scenarios require only a relatively small subset of EML, and in such scenarios ezEML can greatly simplify the process of creating EML, especially for users who are new to EML or use it only infrequently.

ezEML can be used as a "wizard" leading the user through EML document creation step by step, or it can be used in a more user-directed fashion. Among other things, it supports checking the EML for correctness and completeness, importing EML content from other ezEML documents, uploading data tables and inferring most of their characteristics, and downloading the finished EML document as an EML XML file for downstream use in a data workflow.

ezEML was created by the [Environmental Data Initiative \(EDI\)](#). It employs EDI's Metatype library, which is a general-purpose framework for creating and validating metadata, along with Metatype's implementation of validation rules for the EML 2.2.0 standard.

ezEML is aimed at the "long tail" of scientists and others who want to prepare their dataset for submission to a data repository but are not themselves proficient in the details of EML editing. ezEML is designed to give such users a do-it-yourself tool that can handle the great majority of typical datasets.

We hope you find it useful, and we welcome your feedback at support@edirepository.org.

[Back](#)



CONTRIBUTIONS

Publishing Ecological Data in a Repository: An Easy Workflow for Everyone

Kristin Vanderbilt¹ , Jon Ide², Corinna Gries² , Susanne Grossman-Clarke², Paul Hanson² , Margaret O'Brien³ , Mark Servilla¹ , Colin Smith² , Robert Waide¹ , and Kyle Zollo-Venecek² 

?

Análise exploratória de dados

- Por fim, importe os dados, visualize os dados crus para verificar discrepâncias
- Plote os dados utilizando o gráfico mais adequado para o tipo de dado
 - Scatter plot para duas variáveis contínuas
 - Boxplot para um tratamento (preditora) e uma resposta contínua
 - Um histograma nunca é demais

Análise exploratória de dados

- Faça e documente qualquer alteração nos dados utilizando o script, mas evite mexer diretamente na planilha

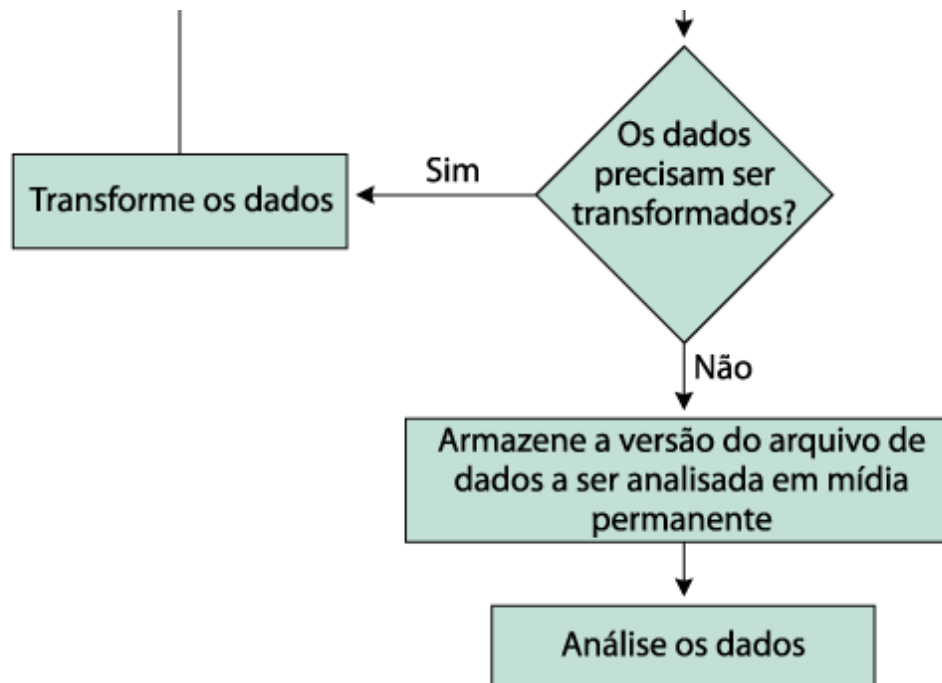


Figura 8.7 Fluxograma de gerenciamento de dados. Esse fluxograma descreve os passos para a gestão, a curadoria e o armazenamento de dados. Todos estes passos devem ser seguidos antes de a análise formal dos dados começar. A organização de metadados é especialmente importante para garantir que o conjunto de dados ainda será acessível e utilizável em um futuro distante.

Como escrever e organizar um script do R

- NiceR Code
 - <https://nicercode.github.io/guide.html>
- Google R style guide
 - <https://google.github.io/styleguide/Rguide.html>
- Tidyverse style guide
 - <https://style.tidyverse.org>

Melhores práticas em organizar um script

1. Comece colocando a data e o título do projeto em que está trabalhando
2. Carregue os pacotes que vá usar
3. Comente as linhas dizendo porque está usando
4. Separe elementos da linha de código
5. Use `<-` para atribuir, e não `=`
6. Quebre a linha, se ela for muito longa
7. Evite dar nome de objetos que já existem

Melhores práticas em organizar um script

1. Escreva TRUE e não T
2. Separe o código em tópicos/partes

Quais funções você aprendeu hoje?