

Escrevendo manuscritos acadêmicos usando rmarkdown

Latin-R 2020 - Conferencia Latinoamericana
sobre Uso de R en Investigación y
Desarrollo

Andrea Sánchez-Tapia & Sara Ribeiro Mortara

¡liibre!

30-09-2020

apresentação

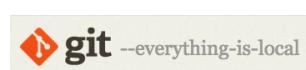


- **Andrea Sánchez-Tapia** Bióloga (UNAL - Colômbia), MSc. Ecologia (UFRJ), Dra. em Botânica (ENBT-JBRJ). Pós-Doc do Núcleo de computação científica do JBRJ. Ecologia quantitativa, informática da biodiversidade, ENM/SDM
- **Sara Ribeiro Mortara** Bióloga (ESALQ/USP), MSc. Ecologia e Conservação da Biodiversidade (UESC) e Dra. em Ecologia (USP). Pesquisadora do Instituto Internacional para a Sustentabilidade (IIS). Modelagem estatística
- **iliibre!** Informática da biodiversidade, ciência aberta e reproduzível, ética na ciência de dados, ciência de dados feminista

Disciplina *Projetos de análise de dados usando R*



Boas práticas em análise de dados



GitHub



LATEX **BIBTEX** **zotero**



esta oficina



- Manuscrito acadêmico: tese/dissertação, artigos acadêmicos, relatórios técnicos
- Fluxo de trabalho: como se organizar?*
- Formatação de figuras e tabelas*
- Citação*
- Formatos finais, templates e modelos

nosso fluxo de trabalho nunca é linear

- Análise de dados é um processo iterativo
- A redação de manuscritos também!
- A gente está acostumada com word e similares... editores de texto
WYSIWYG - What you see is what you get
- **Muito fácil errar** no processo de copiar e colar
- **Paginação, formatação do documento e localização de figuras** não são triviais --
- O resultado gráfico é massomeno :/

um fluxo de trabalho mais amigável

- Escrever *durante* o processo
- Integrar o código e o texto
- Apesar de separar escrita das análises
- **Atualizar automaticamente** se houver mudanças
- Utilizar as vantagens de trabalhar com **scripts + controle de versões (git)** para dar transparência e controlar a qualidade do trabalho
- Facilitar o processo de **citação, formatação e submissão**

"typesetting": formatação, composição tipográfica

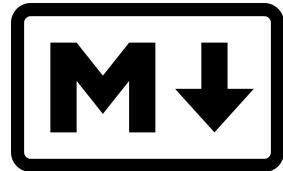


o pioneiro: *LATEX*

- Sistema de preparação de documentos com **alta qualidade tipográfica**
- Baseado em **arquivos de texto** (`.tex`). Formato de saída: **PDF**
- Pode ser usado em **controle de versões** como **git**
- Lida corretamente com **notação** matemática, científica, algebraica etc.
- Muito poderoso com **paginação** e **posicionamento de figuras**
- Arquivos de *LATEX* podem ser **submetidos** como manuscritos para algumas revistas
- "**Templates**" de teses e dissertações em várias instituições

para aprender mais: **overleaf** *LATEX* online, sem instalar nada localmente.

markdown



- Mais simples do que *LATEX* - e para muitas coisas isso é suficiente
- Baseado em arquivos de texto `.md`, formato de saída **HTML**
- Aceita tags de html e planilhas de CSS
- Facilmente interpretado e renderizado por várias plataformas - inclusive GitHub (**README.md**)

pandoc <https://pandoc.org/>

Um canivete suíço: conversor entre **muitos** tipos de arquivo

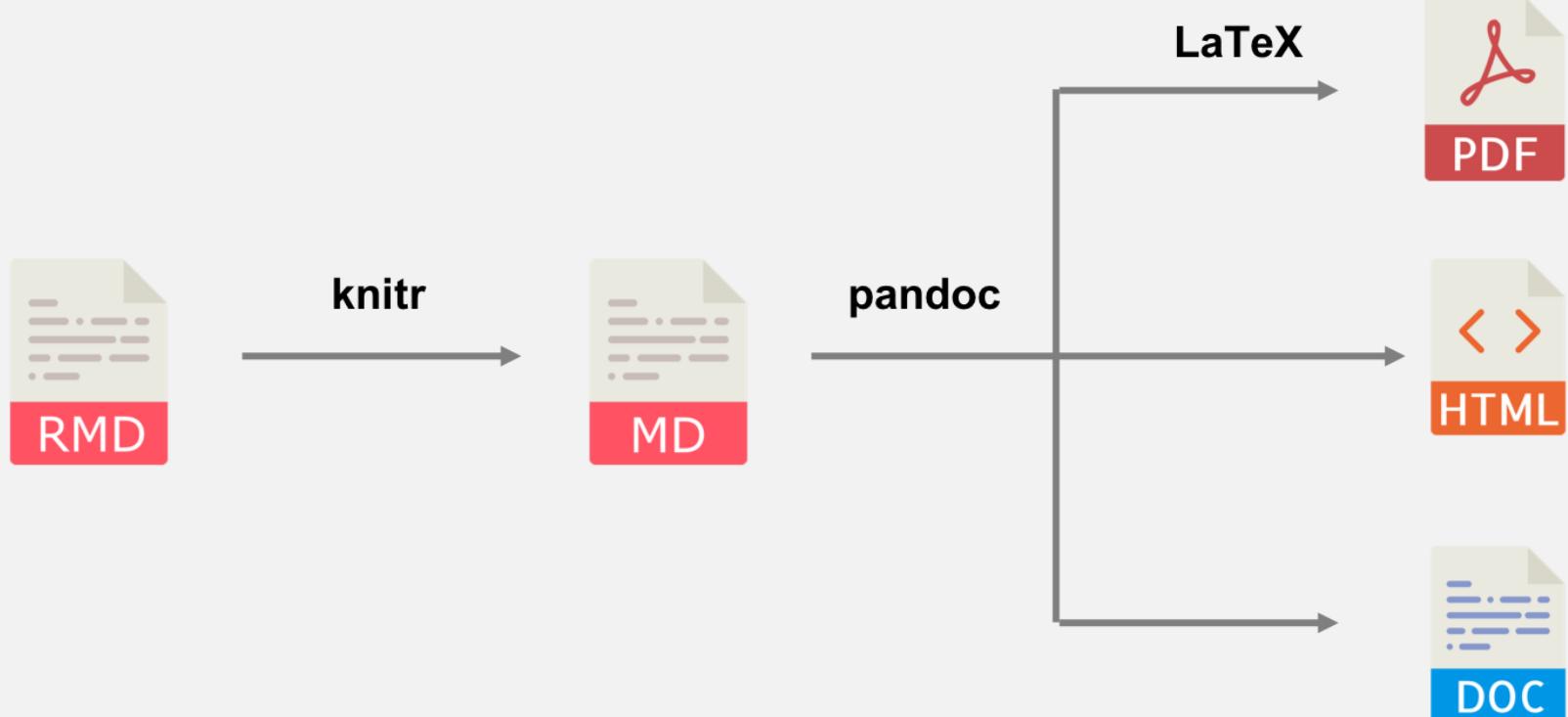
html, rtf, docx,

queremos incluir código!



- *Literate programming* (programação "letrada", Knuth 1984)
- Integrar texto e código
 - *LATEX* e R: **SWEAVE** (**.Rnw**). Output em PDF.
 - markdown e R: **knitr** (**.Rmd**). Output em markdown **.md**.
 - **knitr** também sabe lidar com **.Rnw** e gerar .tex *LATEX*

O pacote **rmarkdown** usa **knitr** para executar o código de R e gerar **pandoc** e consegue gerar muitos formatos de saída.



<https://bookdown.org/yihui/rmarkdown-cookbook/rmarkdown-process.html>

**estrutura de um documento de
rmarkdown**

ajuda para digitar: o pacote **remedy**

remedy é instalado como um **Addin** de RStudio



Você pode modificar os atalhos de teclado para que fique mais fácil escrever:

menu > ferramentas > modificar atalhos de teclado

- **ctrl + b** -> negrito
- **ctrl + i** -> italicico
- **ctrl + k** -> link
- **ctrl + 1** -> h1 (cabeçalho 1)
- **ctrl + 2** -> h1 (cabeçalho 2)
- **ctrl + r** -> insertar *chunk* de código!

para incluir plots

- Gerar no *chunk* com código puro (vai ficar salvo em uma pasta **files**)
- Inserir uma figura externa:
 - Em sintaxe de markdown: ``
 - Em sintaxe de html: ``
 - Em sintaxe de *LATEX*: `\includegraphics{}`

O único que serve com todos os formatos é

...

```
knitr::include_graphics(figura)
```

(dentro de um *chunk*)

incluir tabelas

Formato padrão

a b c
- : - : - :
1 2 3
2 3 4

Resultado

a	b	c
1	2	3
2	3	4

- Para gerar tabelas desde os chunks de R:
 - a função **kable()** de **knitr** e o pacote **kableExtra** que dá mais opções
- Para incluir qualquer tipo de output (por exemplo uma tabela de anova), o pacote **pander**
- Existem serviços para gerar tabelas com este formato, como o **markdown table generator**

Citação e gestão de bibliografias em rmarkdown (e *LATEX*)

gestão de bibliografia

- **Gestor de bibliografias**: armazenam as citações, (opcionalmente os arquivos das publicações) e assistem a citação e criação de bibliografias
- **Zotero** e **Mendeley**, EndNote, Reference Manager
- Para citar em *LATEX* e markdown: BibTeX, formato: **.bib**

```
## @Manual{,
##   title = {vegan: Community Ecology Package},
##   author = {Jari Oksanen and F. Guillaume Blanchet and Michael Friendly and Roeland Kindt and others},
##   year = {2019},
##   note = {R package version 2.5-6},
##   url = {https://CRAN.R-project.org/package=vegan},
## }
```

- Formatação automática: arquivo de **estilo de citação *.csl**

usando zotero ou mendeley

- Uma **extensão no navegador** faz *download* da citação e do pdf se ele estiver disponível
- Armazenamento e gestão de referências em um aplicativo local
- Criar coleções específicas por temas/projetos
- Um **conector** para word e libreoffice (**Zotero**: Funcionalidade no google drive porém lentaaaa)
- Um *site* que cria as citações a partir da URL ou o DOI. <https://zbib.org/>
- Uma página de **estilos de citação**: Zotero style repository. Arquivos **.csl**
- **Zotero**: o addin **betterbibtex** cria a chave de citação, exporta coleções em **.bib** e as mantém atualizadas
- **Mendeley**: criar .bib para cada coleção, manter atualizado

em mendley

Preferences

Document Details File Organizer Watched Folders **BibTeX** Zotero Connection

BibTeX Export Preferences

Escape LaTeX special characters (#{}%& etc.)
 Use Journal Abbreviations

BibTeX Syncing

BibTeX syncing keeps one or several BibTeX files up to date with the documents in your library.
Documents in the 'Needs Review' collection will not be exported.

Enable BibTeX syncing
 Create one BibTeX file for my whole library
 Create one BibTeX file per group
 Create one BibTeX file per document

Path:

Citation Keys

Citation keys for documents are automatically generated in the format [AuthorYear]. To edit citation keys manually, enable the 'Citation Key' field on the Document Details tab

em Zotero

The screenshot shows the Zotero library interface at the top and a 'Exportar...' (Export) dialog box in the foreground.

Zotero Library:

Título	Autor	Citekey
Structural size measurements and isotopic signatures of foraging among adult male and female Chinstrap penguins... LTER e Gorman	LTER e Gorman	lter_structural_2020b
Structural size measurements and isotopic signatures of foraging among adult male and female Gentoo penguins... LTER e Gorman	LTER e Gorman	lter_structural_2020a
Structural size measurements and isotopic signatures of foraging among adult male and female Adélie penguins... LTER e Gorman	LTER e Gorman	lter_structural_2020
Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus P... Gorman et al.	Gorman et al.	gorman_ecological_2014
Data Feminism	D'ignazio e Klein	dignazio_data_2020
allisonhorst/palmerpenguins: v0.1.0	Allison M Horst et al.	allisonmhorst_allisonhorst_2020

Exportar... Dialog Box:

Formato: Better BibTeX

Opções do tradutor

- Exportar notas
- Exportar arquivos
- Usar abreviação do periódico
- Manter atualizado

Cancel OK

para citar em documentos de rmarkdown

- Se `refs.bib` e `estilo.csl` estiverem na pasta `/citation`:
- Incluir o `.bib` e o `.csl` da bibliografia no cabeçalho YAML do documento:
 - `bibliography: ../citation/refs.bib`
 - `csl: ../citation/nature.csl`
- Citar usando o formato: `[@sobrenome_primeirapalavra_ano]` (olhar no zotero ou no mendeley a **chave de citação**)

citação no rmarkdown

- No **texto**, a citação é marcada assim: `@sobrenome_palavrachave_ano`
 - Ex. `[@dignazio_data_2020]` -> (D'Ignazio & Klein 2020)
 - Ex. `@dignazio_data_2020` -> D'Ignazio & Klein (2020)
 - Ex. `[@noble_algorithms_2018, @dignazio_data_2020]` -> (Noble 2018, D'Ignazio & Klein 2020)
- O formato final depende do arquivo de estilo `.csl`
- Pré-determinado: as referências completas são inseridas no final do texto
- Se precisar colocar em outro lugar (por ex. porque tem apêndices, figuras, tabelas) Incluir:
`<div id="refs"></div>`

formatos finais

exportando para .docx

- Um documento de referência pro estilo: "template_word.docx" com os estilos desejados modificados (**à mão desde word**)
- **No cabeçalho YAML:**

```
word_document:  
  reference_docx: templates/template_word.docx
```

- Para melhores resultados, é melhor **knitar uma vez** sem template, **modificar** o documento resultante e **usar esse documento resultante como template**.
- **Alguns problemas com tabelas, páginas horizontais, vão persistir** checar bem!
- Para documentos **.odt** igual:
reference_odt: my-styles.odt

exportando para PDF

- Tudo relacionado a PDF tem a ver com *LATEX*.
- **rmarkdown** aceita marcação de *LATEX* no corpo do texto (ex. `\newpage`)
- Templates e opções de *LATEX* no YAML
- Permite guardar o arquivo `.tex` intermediário

```
output:  
  pdf_document:  
    template: quarterly-report.tex  
    keep_tex: true  
header_includes:  
  - \
```

- Se for um *template* complexo pode ser mais fácil escrever diretamente em *LATEX*. **Mesmo fluxo de trabalho**

Lidando com capítulos

- O mais simples: um .Rmd com documentos "filhos"

ex. `child = 'introducao.Rmd'` nas opções de um chunk vazio

- Permite manter estrutura de pastas/capítulos organizada
- Funciona com os formatos de saída
- Requer de **atenção com relação aos cabeçalhos YAML!** (depende do formato de saída e depende da estrutura dos documentos)

Vamos para `Tese.Rmd` na pasta docs

pacotes bookdown e thesisdown

- Permite juntar vários **.Rmd** em um PDF único e incluir índice, referências...
- **thesisdown**: Pacotes específicos para templates das universidades : variações são outros pacotes, ex. **coppedown**, **ufscdown**
- **.Rmd** com cabeçalhos e templates de *L^AT_EX*

pagedown

- paginação (separar em páginas) dos documentos de html
- um **passo direto de html a PDF**
- formata com **templates CSS**, e modelos **.html**
- para documentos acadêmicos: **pagedown::thesis_paged()**

precisa editar o html e o CSS para fazer documentos em outra língua

outros formatos pré-determinados de pagedown

The screenshot shows the RStudio interface with a R Markdown document open. The code editor at the top contains the following R code:

```
12 title: "A Capitalized Title: Something about a Package **foo**"  
13 shorttitle: "***foo***: A Capitalized Title"  
14 abstract: |  
15 .. The abstract of the article.  
16 # at least one keyword must be supplied  
17 keywords: [keywords, not capitalized, "[Java]{.proglang}"]  
18 links-to-footnotes: true  
19 paged-footnotes: true  
20 bibliography: index.bib
```

The status bar indicates the file is an R Markdown document. Below the editor, the RStudio menu bar shows 'Files', 'Plots', 'Help', and 'Viewer'. The 'Viewer' tab is active, displaying the rendered output. The output includes a header with the title 'A Capitalized Title: Something about a Package foo', a decorative image of Salvador Dalí's 'The Persistence of Memory', and the journal information 'Journal of Statistical Software' and 'MMMMMM YYYY, Volume VV, Issue II. doi: 10.18637/jss.v000.i00'. The bottom of the viewer panel displays the rendered title 'A Capitalized Title: Something about a Package foo'.

pacote **rticles**

- O pacote **rticles** tem vários templates de *L^AT_EX* aceitos pelas revistas (Elsevier, Frontiers, etc.)
- Temas novos são adicionados à versão de desenvolvimento:

```
remotes::install_github("rstudio/rticles")
rticles::journals()
```

- Só PDF e depende da disponibilidade de templates

conclusões

- **Resista a tentação** de fazer um documento só knitando o texto com todo o código (**Poder fazer ≠ dever fazer**):
 - *Scripts* e *outputs* separados do documento
 - Vários capítulos: várias pastas
 - Não escreva num **.Rmd** único!
 - **Flexibilidade** na hora de fazer a formatação final
- Entenda os **formatos finais requeridos** e sua disponibilidade em diferentes plataformas
- **Não escolha muito cedo mas não deixe para o final** a decisão sobre formatação

as boas práticas ainda aplicam

- Cuidar da **estrutura de pastas**
- **Modularizar** o código de cada manuscrito/capítulo
- Usar **caminhos relativos**
- Usar **git** para controle de versões

referências

- Rosanna van Hespen: Writing your thesis with R
<https://www.rosannavanhespen.nl/2016/02/16/writing-your-thesis-with-r-markdown-1-getting-started/>
- R Markdown: The Definitive Guide <https://bookdown.org/yihui/rmarkdown/>
- R markdown cookbook: <https://bookdown.org/yihui/rmarkdown-cookbook/>
- <https://rmarkdown.rstudio.com>
<https://bookdown.org/yihui/rmarkdown/interactive-documents.html>

¡Gracias!

 @mortarasara  @SanchezTapiaA  @rladiesrio



 R-Ladies+ Rio de Janeiro