

Introdução ao Tratamento e Análise de Dados em R

Aula 2 - Produzindo Relatórios Usando R Markdown

Sérgio Rivero

PPGE-UFGA

14 de maio de 2019



Sumário

Objetivos da Aula

Markdown: Origens

Sintaxe de Markdown

Exercícios

Referências



Objetivos da Aula

Nesta Aula queremos:

1. Descobrir as funcionalidades implementadas no R para a produção de relatórios dinâmicos, execução de comandos R e apresentação de resultados utilizando uma linguagem simples chamada **Markdown**;
2. Depois veremos como estes resultados podem ser apresentados em relatórios, utilizando exemplos práticos;
3. Esta aula permitirá que façamos os exercícios já utilizando o *Markdown*;



De onde veio o Markdown?

R Markdown é um formato de arquivo que permite a produção documentos dinâmicos utilizando R. Um documento R Markdown é escrito na linguagem **Markdown** e contém partes de código R incorporado em seções específicas do arquivo, chamadas *chunks*:

Markdown *é duas coisas:*

1. *uma sintaxe de formatação de texto simples;*
2. *uma ferramenta de software, escrita em Perl, que converte a formatação de texto simples em HTML.*

A meta de design predominante da sintaxe de formatação do Markdown é torná-la o mais legível possível.



Markdown - Objetivos

- O uso da linguagem **Markdown**, quando se analisa dados utilizando o R, permite uma transição suave e contínua entre:
 - Análise → Relatório → Apresentação
- Um arquivo *Rmarkdown* (Rmd) funciona como uma espécie de *bloco de notas*
- Pode evoluir para o formato de um relatório final
- Em um conjunto de formatos (Html, PDF, docx, etc.)
- Isto economiza tempo e esforço.



O que é o Rmarkdown?

- O *Rmarkdown* (RMD) é a implementação do **Markdown** em R.
- RMD é um formato de documento introduzido inicialmente no pacote *knitr*.
- O Knitr, suporta um conjunto grande de linguagens, como \LaTeX , HTML, AsciiDoc, reStructuredText e Textile.
- A ideia por trás do RMD é embutir comandos em uma linguagem qualquer (no nosso caso, o R) dentro de um documento.
- Assim, um arquivo .Rmd teria, ao mesmo tempo:
 - comandos que gerariam resultados como gráficos e tabelas,
 - que seriam gerados num arquivo de resultados
 - e compilados posteriormente num documento PDF, HTML, Word, etc.
- o Código R está em partes específicas chamadas *chunks*



RMD no RStudio

The screenshot displays the RStudio interface with an R Markdown document open. The editor shows R code for plotting the relationship between Sepal.Length and Petal.Length from the iris dataset using ggplot2. The code includes comments in Portuguese. The right-hand pane shows the rendered HTML output, which includes a calculation of $1 - (1 + 10^{-15})$ and a scatter plot titled "Relação entre tamanho da Pétala e o tamanho da Sépala". The plot shows a positive correlation between the two variables. The bottom console shows the command-line output of the R Markdown rendering process.

```
23
24 # Mostrando Resultados de Gráficos
25
26 ```{r, echo=FALSE, fig.dims=5}
27 library(ggplot2)
28 data("iris")
29 g1 <- ggplot(iris, aes(x = Sepal.Length, y =
30   Petal.Length)) +
31   geom_point() +
32   ggtitle("Relação entre tamanho da Pétala e o tamanho da
33     Sépala") +
34   labs(x="Tamanho da Sépala", y="Tamanho da Pétala")
35   plot(g1)
36   ...
```

Environment History Connections
Files Plots Packages Help Viewer
Publish

Calcule as seguintes operações:
a. $1 - (1 + 10^{-15})$

$1 - (1 + 10^{-15})$

[1] -1.118223e-15

Mostrando Resultados de Gráficos

Registered S3 methods overwritten by 'ggplot2':
method from
[.quosures rlang
c.quosures rlang
print.quosures rlang

Relação entre tamanho da Pétala e o tamanho da Sépala

Tamanho da Pétala

26:29 Chunk 2 : R Markdown

Console Terminal R Markdown
...:basefauxExemplo_Rmd_Sergio_Rivero_Original.Rmd
cii_identifiers+tex_math_single_backslash --output Exemplo_Rmd_Sergio_Rivero_Original.html --smart --email-obfuscation none --self-contained --standalone --section-divs --table-of-contents --toc-depth 2 --template /home/rivero/R/x86_64-pc-linux-gnu-library/3.6/rmarkdown/rmd/h/default.html --highlight-style tango --variable 'theme:cosmo' --include-in-header /tmp/Rtmp7txV4S/rmarkdown-str3e693312cbb4.html --mathjax --variable 'mathjax-url:https://mathjax.rstudio.com/latest/MathJax.js?config=TeX-AMS-MML_HTMLorMML' output file: Exemplo_Rmd_Sergio_Rivero_Original.knit.md

Output created: Exemplo_Rmd_Sergio_Rivero_Original.html

Um exemplo de saída de um .Rmd no Rstudio



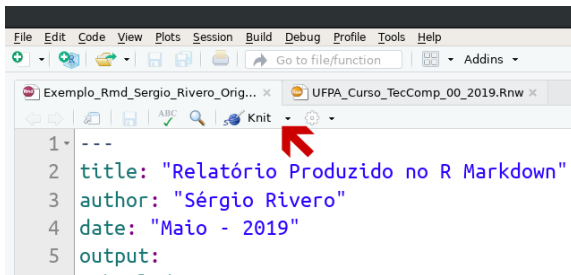
Instalando o RMD

- Instalação do Rmarkdown → *install.packages()*.
- Ou instalar a versão de desenvolvimento do *GitHub*. Para isso é necessário utilizar o pacote *devtools*.
- Para conseguir gerar o pdf no windows a partir de um arquivo RMD é necessário instalar o \LaTeX . O \LaTeX é um conjunto de pacotes que permite a edição e geração de textos formatados de alta qualidade.



Como o Rmd funciona?

O processo de geração de documentos com o RMD é feito a partir da pressão do botão **Knit** no RStudio. Para isso, você deve ter habilitado e configurado o Markdown no app. A partir daí, o arquivo `.rmd` (Rmarkdown) tem seus códigos R executados e gera um código `.md` com os resultados da execução do R (figuras, tabelas, etc) que é então processado pelo pandoc. Esse processamento, gera o arquivo de resultado desejado (HTML, DOC, PDF, etc)



O processo de geração de documentos no RMD



O processo de geração de documentos no RMD

A estrutura de um arquivo .Rmd

Um arquivo .Rmd tem basicamente, 3 partes.

1. Um cabeçalho escrito em YAML
2. YAML é um formato de serialização de dados legíveis por humanos inspirado em linguagens como XML, C, Python, Perl, assim como o formato de correio eletrônico especificado pela RFC 2822. YAML foi proposto por Clark Evans em 2001 em conjunto com Ingy döt Net e Oren Ben-Kiki (Wikipedia). Separados por " - - - " no início e no final do bloco.
3. Textos e comandos em Linguagem Markdown
4. *Chunks* em uma linguagem específica (no nosso caso, Comandos em R) - Separados por " ' ' ' " no início e no final do bloco.



O Cabeçalho de um arquivo .Rmd

```
---  
title: "Relatório Produzido no R Markdown"  
author: "Sérgio Rivero"  
date: "Maio - 2019"  
output:  
  html_document:  
    highlight: tango  
    theme: cosmo  
    toc: yes  
    toc_depth: 2  
subtitle: Exemplos de Documentos Produzidos no R Markdown  
---
```

Um exemplo de Cabeçalho de um arquivo .Rmd



Arquivos que podem ser gerados por um Rmd



output value

creates

html_document	html
pdf_document	pdf (requires Tex)
word_document	Microsoft Word (.docx)
odt_document	OpenDocument Text
rtf_document	Rich Text Format
md_document	Markdown
github_document	Github compatible markdown
ioslides_presentation	ioslides HTML slides
slidy_presentation	slidy HTML slides
beamer_presentation	Beamer pdf slides (requires Tex)



Sub-opções no cabeçalho

```
---  
output:    
  html_document:  
    code_folding: hide  
    toc_float: TRUE  
---  
  
# Body
```



Exemplos de subopções

sub-option	description	html	pdf	word	odt	rtf	md	github	ioslides	slidy	beamer
citation_package	The LaTeX package to process citations, natbib, biblatex or none	X					X				X
code_folding	Let readers to toggle the display of R code, "none", "hide", or "show"	X									
colortheme	Beamer color theme to use										X
css	CSS file to use to style document	X						X	X		
dev	Graphics device to use for figure output (e.g. "png")	X	X				X	X	X	X	X
duration	Add a countdown timer (in minutes) to footer of slides									X	
fig_caption	Should figures be rendered with captions?	X	X	X	X			X	X	X	X
fig_height,fig_width	Default figure height and width (in inches) for document	X	X	X	X	X	X	X	X	X	X
highlight	Syntax highlighting: "tango", "pygments", "kate", "zenburn", "textmate"	X	X	X						X	X
includes	File of content to place in document (in_header, before_body, after_body)	X	X	X			X	X	X	X	X
incremental	Should bullets appear one at a time (on presenter mouse clicks)?								X	X	X
keep_md	Save a copy of .md file that contains knitr output	X		X	X	X		X	X		
keep_tex	Save a copy of .tex file that contains knitr output		X								X
latex_engine	Engine to render latex, "pdflatex", "xelatex", or "lualatex"		X								X
lib_dir	Directory of dependency files to use (Bootstrap, MathJax, etc.)	X						X	X		
mathjax	Set to local or a URL to use a local/URL version of MathJax to render	X							X	X	
md_extensions	Markdown extensions to add to default definition or R Markdown	X	X	X	X	X	X	X	X	X	X
number_sections	Add section numbering to headers	X	X								
pandoc_args	Additional arguments to pass to Pandoc	X	X	X	X	X	X	X	X	X	X
preserve_yaml	Preserve YAML front matter in final document?						X				
reference_docx	docx file whose styles should be copied when producing docx output			X							
self_contained	Embed dependencies into the doc	X						X	X		
slide_level	The lowest heading level that defines individual slides										X
smaller	Use the smaller font size in the presentation?								X		
smart	Convert straight quotes to curly, dashes to em-dashes, ... to ellipses, etc.	X						X	X		
template	Pandoc template to use when rendering file	X	X	X						X	X
theme	Bootswatch or Beamer theme to use for page	X									X
toc	Add a table of contents at start of document	X	X	X		X	X	X			X
toc_depth	The lowest level of headings to add to table of contents	X	X	X		X	X	X			
toc_float	Float the table of contents to the left of the main content	X									



Sintaxe de Markdown

O texto em um documento R Markdown é gravado com a sintaxe Markdown. Precisamente falando, é o Markdown do Pandoc. Há muitos sabores de Markdown inventados por pessoas diferentes, e o sabor do Pandoc é o mais abrangente para o nosso conhecimento. Você pode encontrar a documentação completa do Markdown do Pandoc em <https://pandoc.org/MANUAL.html>. É altamente recomendável que você leia esta página pelo menos uma vez para conhecer todas as possibilidades com o Markdown da Pandoc, mesmo que você não use todas elas. Esta seção é adaptada da Seção 2.1 de Xie (2016) e abrange apenas um pequeno subconjunto da sintaxe Markdown do Pandoc.



Formatação em linha

O texto embutido será em itálico se estiver cercado por sublinhados ou asteriscos, por exemplo, `_text_` ou `*text*`. O texto em negrito é produzido usando um par de asteriscos duplos (`**text**`). Um par de tís (`~~`) transformar texto para um subscrito (por exemplo, `H~3~P~O~4~` torna H_3PO_4). Um par de circunflexos (`^^`) produz um sobrescrito (por exemplo, `Cu^2+^` renderiza Cu^{2+}).



Formatação (Ainda)

Para marcar o texto como código *inline*, use um par de backticks, por exemplo ``code``

Os hiperlinks são criados usando a sintaxe `[text](link)`, por exemplo `[RStudio](https://www.rstudio.com)`.

A sintaxe das imagens é semelhante: basta adicionar um ponto de exclamação, por exemplo

`![alt text or image title](path/to/image)`,

As notas de rodapé são colocadas dentro dos colchetes após um acento circunflexo `^[]`, por exemplo `^[This is a footnote.]`.



Elementos de nível de bloco

Cabeçalhos de seção podem ser escritos após um número de sinais de libra, por exemplo,

First-level header

Second-level header

Third-level header

Se você não quiser que um determinado título seja numerado, você pode adicionar {-} ou {.unnumbered} após o cabeçalho, por exemplo,

Preface {-}



Listas de itens no Rmd

Os itens da lista não ordenada começar com *, -ou +, e você pode aninhar uma lista dentro de outra lista pelo recuo da sub-lista, por exemplo,

- um item
- um item
- um item
 - mais um item
 - mais um item
 - mais um item

A saída é:

- um item
- um item
- um item
 - mais um item
 - mais um item
 - mais um item



Listas Ordenadas no Rmd

Os itens da lista ordenada começam com números (você também pode aninhar listas dentro de listas), por exemplo,

1. o primeiro item
2. o segundo item
3. o terceiro item
 - um item não ordenado
 - um item não ordenado

A saída não parece muito diferente com a fonte Markdown:

1. o primeiro item
2. o segundo item
3. o terceiro item
 - um item não ordenado
 - um item não ordenado



Os *chunks*

Um *chunk* é uma parte de código (no nosso caso, R) incluído num arquivo **Rmarkdown**. Chunks iniciam com ````{opcoes do chunk}` e terminam com `````. Na figura abaixo podemos ver um exemplo de um chunk típico gerando um gráfico em R.

```
```{r, echo=FALSE,fig.dim=.5}  
library(ggplot2)
data("iris")
g1 <- ggplot(iris, aes(x = Sepal.Length, y = Petal.Length)) +
 geom_point() +
 ggtitle("Relação entre tamanho da Pétala e o tamanho da Sépala") +
 labs(x="Tamanho da Sépala", y="Tamanho da Pétala")
plot(g1)
```
```

Um *chunk*



Opções de *chunks*

Algumas dicas importantes sobre as opções em um *chunk* ¹:

1. Opções de *chunk* devem ser escritas em uma linha;
2. não são permitidas quebras de linha dentro das opções de *chunk*;
3. Evite espaços e pontos em rótulos de partes e nomes de diretórios;
4. Se sua saída for um documento TeX, esses caracteres podem causar problemas (em geral, é recomendável usar caracteres alfabéticos com palavras separadas por - e evitar outros caracteres), por exemplo, `setup-options` é um bom label, enquanto `setup.options` e `chunk 1` são ruins;
5. Todos os valores de opções devem ser expressões R válidas, assim como escrevemos argumentos de função;
6. para opções lógicas, deve-se usar TRUE ou FALSE ou T ou F mas, importante, "true" e "false" não funcionam



¹O texto original pode ser encontrado em

Abaixo falamos de algumas opções relevantes do cabeçalho de um chunk

Primeiro o código pode ou não ser executado. O opção `eval = FALSE` faz com que o código seja avaliado mas não executado. Sem nenhuma opção, o código é avaliado.

Algumas opções que configuram os padrões de saída de um *chunk* são:

- **echo:** (TRUE) se deve incluir o código-fonte R no arquivo de saída; além de TRUE / FALSE que liga / desliga completamente o código-fonte, também podemos usar um vetor numérico para selecionar quais expressões R ecoar em um trecho, por exemplo, `echo = 2:3` significa apenas ecoar a segunda e terceira expressões, e `echo = -4` significa excluir a quarta expressão;
- **results:** ('markup'; character) pega esses valores possíveis:
 - **markup:** marque os resultados usando o gancho de saída, por ex. colocar resultados em um ambiente especial do \LaTeX ;
 - **asis:** saída como é, ou seja, grava os resultados brutos de R no documento de saída;



Mais opções do *chunk*

- **warning:** (default é VERDADEIRO;) preserva os avisos (produzidos por `warning()`) na saída como se executássemos código R em um terminal (se for FALSO, todos os avisos serão impressos no console ao invés do documento de saída);
- **error:** (default é VERDADEIRO) preserva os erros (de `stop()`). A avaliação não será interrompida mesmo em caso de erros!! Se quisermos que R pare nos erros, precisamos definir essa opção como FALSO;
- **message:** (default é VERDADEIRO) preserva as mensagens emitidas por `message()` (semelhante ao `warning`).

Um detalhamento maior das opções de chunks no Rmarkdown podem ser achados em <https://yihui.name/knitr/options/>



Tabelas (veremos isso mais tarde)

Table suggestions

Several functions format R data into tables

Table with kable

| eruptions | waiting |
|-----------|---------|
| 3.600 | 79 |
| 1.800 | 54 |
| 3.333 | 74 |
| 2.283 | 62 |

eruptionswaiting

| | | |
|---|------|-------|
| 1 | 3.60 | 79.00 |
| 2 | 1.80 | 54.00 |
| 3 | 3.33 | 74.00 |
| 4 | 2.28 | 62.00 |

Table with xtable

Table with stargazer

| | eruptionswaiting |
|---|------------------|
| 1 | 3.600 79 |
| 2 | 1.800 54 |
| 3 | 3.333 74 |
| 4 | 2.283 62 |

```
data <- faithful[1:4, ]
```

```
```\r results = 'asis'\nknitr::kable(data, caption = "Table with kable")\n```\n
```

```
```\r results = "asis"\nprint(xtable::xtable(data, caption = "Table with xtable"),\n      type = "html", html.table.attributes = "border=0")\n```\n
```

```
```\r results = "asis"\nstargazer::stargazer(data, type = "html",\n  title = "Table with stargazer")\n```\n
```

Learn more in  
the **stargazer**,  
**xtable**, and  
**knitr** packages.



# Agradecimentos

Obrigado!



# Exercício



# Referências I

<https://daringfireball.net/projects/markdown/>  
[https://rmarkdown.rstudio.com/authoring\\_quick\\_tour.html](https://rmarkdown.rstudio.com/authoring_quick_tour.html)  
<https://www.markdownguide.org/getting-started/>  
<https://ourcodingclub.github.io/2016/11/24/rmarkdown-1.html>  
<https://github.com/yihui/knitr-book>  
<https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>  
<https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf>

