

R Markdown para Comunicação Estatística

Markus Stein

em colaboração com Márcia Barbian, Rodrigo dos Reis
e Silvana Schneider

JORNADA DE MATEMÁTICA E ESTATÍSTICA: 60 ANOS IME/UFRGS

Departamento de Estatística

2020-11-20



Instituto de
MATEMÁTICA
E ESTATÍSTICA

60 anos

UFRGS

Visão Geral

- R e Markdown
- Comunicação Estatística
- Pacote rmarkdown
- Conceitos básicos
- Estendendo documentos .Rmd
- Extras*

O que você vai precisar

- Pacotes do R:
 - knitr e rmarkdown, pacotes auxiliares dplyr, tidyr;
 - Extras: hugo, blogdown, bookdown, xaringan,
- Material do curso
 - Na aba Terminal do RStudio digite:

```
git clone https://github.com/markus-stein/Rmarkdown_workshop
```
 - (Slides com conceitos básicos: https://datathon-ufrgs.github.io/Pintando_e_Bordando_no_R/#1)
- Guia rápido de R Markdown (Cheat Sheet)
 - Pasta cheatsheet

O que é ?

(Por que R???)



Fonte: *The Art of R Programming*

E por que RStudio?

O que é Markdown?

Por que Markdown?

Markdown is a text-to-HTML conversion tool for web writers. Markdown allows you to write using an easy-to-read, easy-to-write plain text format, then convert it to structurally valid XHTML (or HTML).

John Gruber

- [Markdown] é uma [linguagem de marcação] simples para escrever textos
- O texto pode ser lido sem nenhum processamento, ou seja, da maneira como está escrito
- Outras linguagens de marcação como HTML e $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ requerem um grande número de *tags* para formatar o texto, muitas vezes dificultando a leitura do código-fonte
- A proposta do Markdown é que o escritor se concentre no texto e não na formatação
- Pode ser convertido para **vários outros formatos** além de HTML

O que é Markdown?

- Um documento Markdown possui a extensão `.md` (não a única possível)
- Veja o arquivo de exemplo `exemplo-rmarkdown-0.md`
- Para converter um documento Markdown em HTML (ou outro formato) é necessário um **conversor**.
- O Pandoc é um conversor extremamente versátil, diversos formatos, incluindo Markdown para HTML.

Se o Pandoc estiver instalado no seu sistema (Unix), converter o documento na linha de comando (Terminal) com

```
pandoc -f markdown -t html exemplo-rmarkdown-0.md -o exemplo-rmarkdown-0.html
```

O pacote `knitr` possui a função `pandoc()` que é um *wrapper* para executar o programa `pandoc` no sistema.

```
# install.packages("knitr")  
# library(knitr)  
pandoc(input = "exemplos/exemplo-rmarkdown-0.md", format = "html")
```

Em ambos os casos, o resultado é o arquivo `exemplo-rmarkdown-0.html` no navegador.

Comunicação Estatística

- Comunicação em vários sentidos
 - teorias e metodologias entre os pares da área;
 - metodologias e aplicações a outras áreas;
 - comunicar resultados ao público em geral.

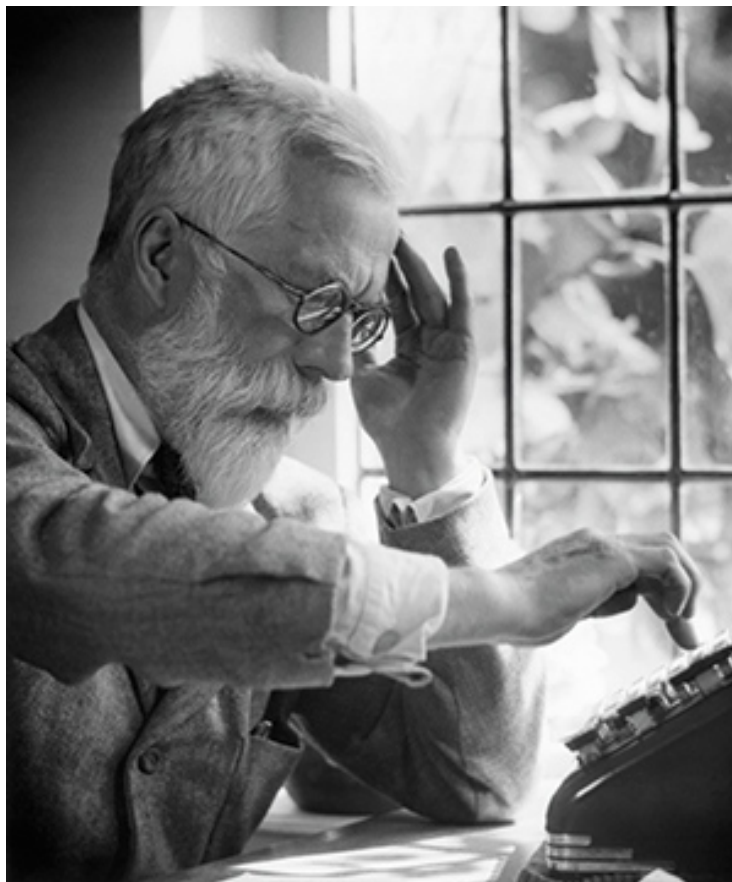
Relatórios, artigos, livros, slides, pôster, websites, blogs, dashboards, ...

Teoria e Prática

- Exemplo: Teoria de prob. + tabelas de dists. = metodologia

Comunicação Estatística

Fisher + milionária = "Statistical Methods for Research Workers"



Comunicação Estatística

Comunicação Estatística

TABLE 32

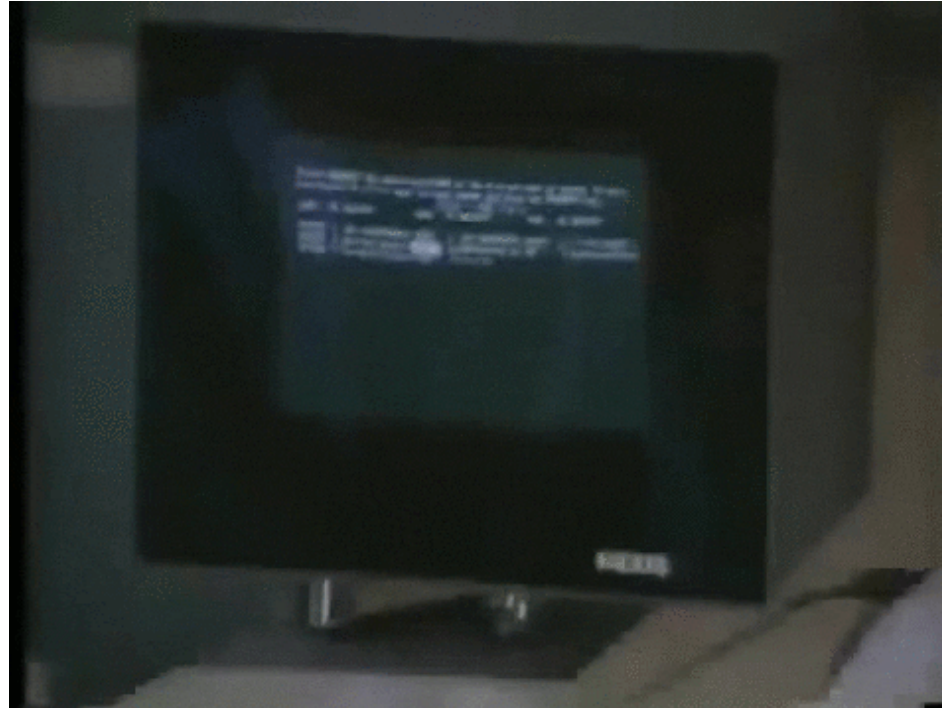
Daughters.				Fathers.				Total for Daughters.	Product.
Deviation.	Frequency.			Deviation.	Frequency.				
-11	5	5.5	60.5	-9	2	18	162	- 8.75	+ 78.75
-10	5	5	50	-8	4.5	36	288	- 15.25	+ 122
-9	-	-	-	-7	7.5	52.5	367.5	- 19	+ 133
-8	1	8	64	-6	14.5	87	522	- 23	+ 138
-7	4.5	31.5	220.5	-5	45	225	1125	-108.75	+ 543.75
-6	14.5	87	522	-4	51.5	206	824	- 81	+ 324
-5	15.5	77.5	387.5	-3	92.5	277.5	832.5	- 76.25	+ 228.75
-4	48.5	194	776	-2	155	310	620	- 88.50	+ 177
-3	99	297	891	-1	178	178	178	-131.25	+ 131.25
-2	141.5	283	566	0	175	1390	..	+ 15.5	..
-1	190.5	190.5	190.5	1	199.5	199.5	199.5	+ 183.25	+ 183.25
0	212	-1179	..	2	166	332	664	+ 197.25	+ 394.5
1	198.5	198.5	198.5	3	135	405	1215	+ 245	+ 735
2	159.5	319	638	4	82.5	330	1320	+ 174.75	+ 699
3	142.5	427.5	1282.5	5	36.5	182.5	912.5	+ 105.25	+ 526.25
4	77.5	310	1240	6	20	120	720	+ 71.5	+ 429
5	36	180	900	7	6.5	45.5	318.5	+ 25.25	+ 176.75
6	19.5	117	702	8	4.5	36	288	+ 14.5	+ 116
7	9.5	66.5	465.5						
8	4	32	256						
9	1	9	81						
	1376	+1659.5			1376	+1650.5		480.5	
		-1179				-1390			
	Total	+480.5	9491.5		Total	+260.5	10556.5	Total	+5136.25
Correction for mean			-167.8	Correction for mean			-49.3	Correction for mean	-90.97
			9323.7				10507.2		
Sheppard's correction			114.7	Sheppard's correction			114.7		+5045.28
			9209.0				10392.5		

Comunicação Estatística

Surgimento dos computadores pessoais



Comunicação Estatística



Fonte: ghipy.com

Como combinar diferentes tecnologias?

- "Novos instrumentos trazem novos desafios".

Comunicação Estatística



Fonte: ghipy.com

Computadores falicitaram muito o nosso trabalho, mas os problemas continuam complexos.

- *Big data* , modelos e delineamentos complexos
- Aprendizado de máquinas e inteligência artificial

R e Markdown

O pacote `rmarkdown` e a função `render()`

- A função `render()` é uma *wrapper*,
 - internamente chama `knitr::knit()`
 - então converte o documento para `.html` usando o Pandoc.

Conceitos básicos

- Slides da apresentação <https://github.com/datathon-ufrgs/website>
 - Prof. Rodrigo dos Reis



O que é o rmarkdown¹?



- **R** (códigos) + **Markdown** (linguagem **simples** de marcação para geração de **texto**)
 - **Pandoc** (conversor universal de documentos)

[1] Allaire, J.J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J. e Chang, W. (2018). *rmarkdown: Dynamic Documents for R*. R package version 1.10.

O fluxo do rmarkdown?



- Quando compilado, o **R Markdown** alimenta o arquivo .Rmd para **knitr**, que executa todos os fragmentos de código e cria um novo documento markdown (.md) que inclui o código e sua saída.
- O arquivo markdown gerado pelo **knitr** é então processado pelo **pandoc** que é responsável pela criação do formato final.
- Isso pode parecer complicado, mas o **R Markdown** torna extremamente simples encapsulando todo o processamento acima em uma única função de renderização.

Por que rmarkdown?



- Vamos a um exemplo!

Vamos lá!



Instalando e carregando o rmarkdown

- Instalando o pacote **rmarkdown**

```
install.packages("rmarkdown")
```

- Carregando o pacote **rmarkdown**

```
library(rmarkdown)
```

Criando um documento R markdown

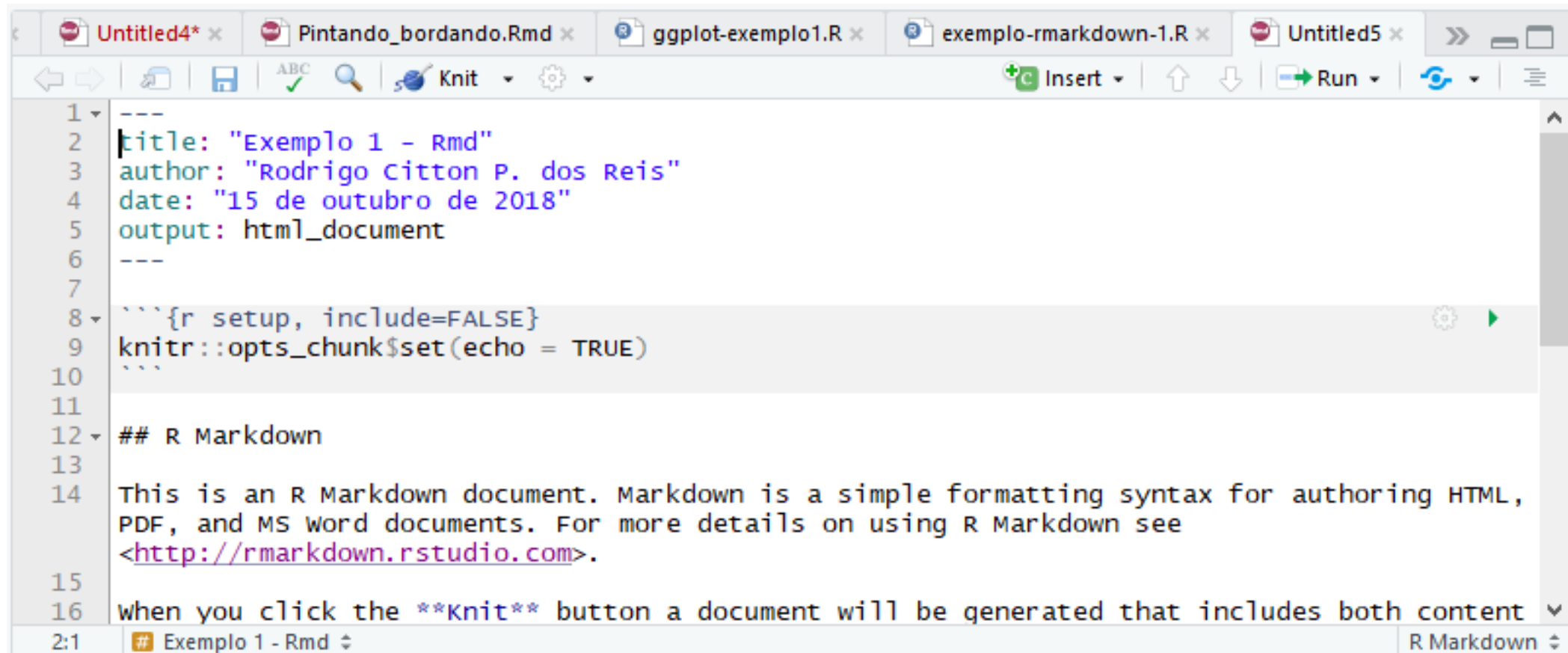
- Clique em `File > New File > R Markdown`

Criando um documento R markdown

- Agora você deve ver uma caixa de diálogo como mostrado abaixo.
- Selecione "Document" no painel à esquerda e preencha o campo de título e autor e clique em "OK".

Criando um documento R markdown

- Agora você deve ter um documento que parece com isso



The screenshot shows the RStudio IDE with several open files. The active file, 'exemplo-rmarkdown-1.R', contains an R Markdown document. The document starts with a YAML header for metadata, followed by an R code chunk to configure Knitr, and then a text block explaining R Markdown. The status bar at the bottom indicates the current position is at line 2, column 1, and the document type is R Markdown.

```
1 ---
2 title: "Exemplo 1 - Rmd"
3 author: "Rodrigo Citton P. dos Reis"
4 date: "15 de outubro de 2018"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML,
15 PDF, and MS word documents. For more details on using R Markdown see
16 <http://rmarkdown.rstudio.com>.
17
18 when you click the **knit** button a document will be generated that includes both content
```

2:1 | # Exemplo 1 - Rmd | R Markdown

YAML

- Um conjunto de opções que definem o arquivo de saída
- Este documento gera um arquivo html

```
---  
title: "Exemplo 1 - Rmd"  
author: "Rodrigo Citton P. dos Reis"  
date: "15 de outubro de 2018"  
output: html_document  
---
```

- Este documento gera um arquivo word

```
---  
title: "Exemplo 1 - Rmd"  
author: "Rodrigo Citton P. dos Reis"  
date: "15 de outubro de 2018"  
output: word_document  
---
```

Markdown: sintaxe básica

```
texto simples  
_italico_  
__negrito__  
[Datahon](https://www.ufrgs.br/datathon)  
# Título 1  
## Título 2  
### Título 3
```

texto simples

italico

negrito

Datahon

Título 1

Título 2

Título 3

Markdown: sintaxe básica

- Lista não ordenada
- item 2

- + sub-item 1

- + sub-item 2

1. Lista ordenada

2. item 2

- + sub-item 1

- + sub-item 2

Cabeçalho tabela	Segundo cabeçalho
------------------	-------------------

-----	-----
-------	-------

Célula tabela	Célula 2
---------------	----------

Célula 3	Célula 4
----------	----------

- Lista não ordenada
- item 2
 - sub-item 1
 - sub-item 2

1. Lista ordenada

2. item 2

- sub-item 1

- sub-item 2

Cabeçalho tabela	Segundo cabeçalho
Célula tabela	Célula 2
Célula 3	Célula 4

0 banco de dados

- Carregando o banco de dados **gapminder**

```
# install.packages("gapminder")  
# library(gapminder)  
gapminder
```

```
## # A tibble: 1,704 x 6  
##   country      continent  year lifeExp      pop gdpPercap  
##   <fct>        <fct>    <int>  <dbl>    <int>    <dbl>  
## 1 Afghanistan Asia      1952   28.8  8425333    779.  
## 2 Afghanistan Asia      1957   30.3  9240934    821.  
## 3 Afghanistan Asia      1962   32.0 10267083    853.  
## 4 Afghanistan Asia      1967   34.0 11537966    836.  
## 5 Afghanistan Asia      1972   36.1 13079460    740.  
## 6 Afghanistan Asia      1977   38.4 14880372    786.  
## 7 Afghanistan Asia      1982   39.9 12881816    978.  
## 8 Afghanistan Asia      1987   40.8 13867957    852.  
## 9 Afghanistan Asia      1992   41.7 16317921    649.  
## 10 Afghanistan Asia      1997   41.8 22227415    635.  
## # ... with 1,694 more rows
```

0 banco de dados

- `gdpPercap`: **renda per capita** ("PIB/Pop").
- `lifeExp`: **expectativa de vida** ao nascer (número de anos aproximados que se espera que um grupo de indivíduos nascidos no mesmo ano irá viver).
- `year`: 1952 a 2007 em incrementos de 5 anos.

```
# install.packages(dplyr)
# library(dplyr)
gapminder <- gapminder %>%
  mutate(pop_m = pop/1e6)
gapminder07 <- gapminder %>%
  filter(year == 2007)
```

Chunk!

Chunk

- Os códigos em **R** são passados para o arquivo .Rmd por meio de fragmentos de código (*chunk codes*).
- Um exemplo de chunk:

```
```{r}
summary(gapminder07$pop_m)
```
```

- Saída:

```
summary(gapminder07$pop_m)
```

| | | | | | | |
|----|--------|---------|---------|---------|---------|-----------|
| ## | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| ## | 0.1996 | 4.5080 | 10.5175 | 44.0212 | 31.2100 | 1318.6831 |

Chunk

```
```{r, echo=FALSE}  
summary(gapminder07$pop_m)
```
```

- Saída:

| ## | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|----|--------|---------|---------|---------|---------|-----------|
| ## | 0.1996 | 4.5080 | 10.5175 | 44.0212 | 31.2100 | 1318.6831 |

Chunk

```
` `{r, eval=FALSE}  
summary(gapminder07$pop_m)  
` `
```

- Saída:

```
summary(gapminder07$pop_m)
```

Chunk

```
```{r, echo=FALSE, results='asis'}  
library(knitr)
mod1 <- lm(lifeExp ~ gdpPercap, data = gapminder07)
kable(summary(mod1)$coef, format = "html")
```
```

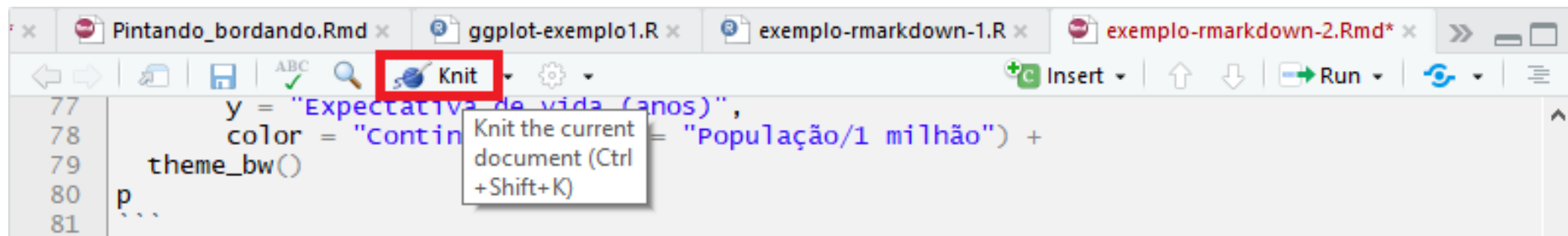
- Saída:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|------------|------------|----------|-----------|
| (Intercept) | 59.5656501 | 1.0104086 | 58.95204 | 0 |
| gdpPercap | 0.0006371 | 0.0000583 | 10.93340 | 0 |

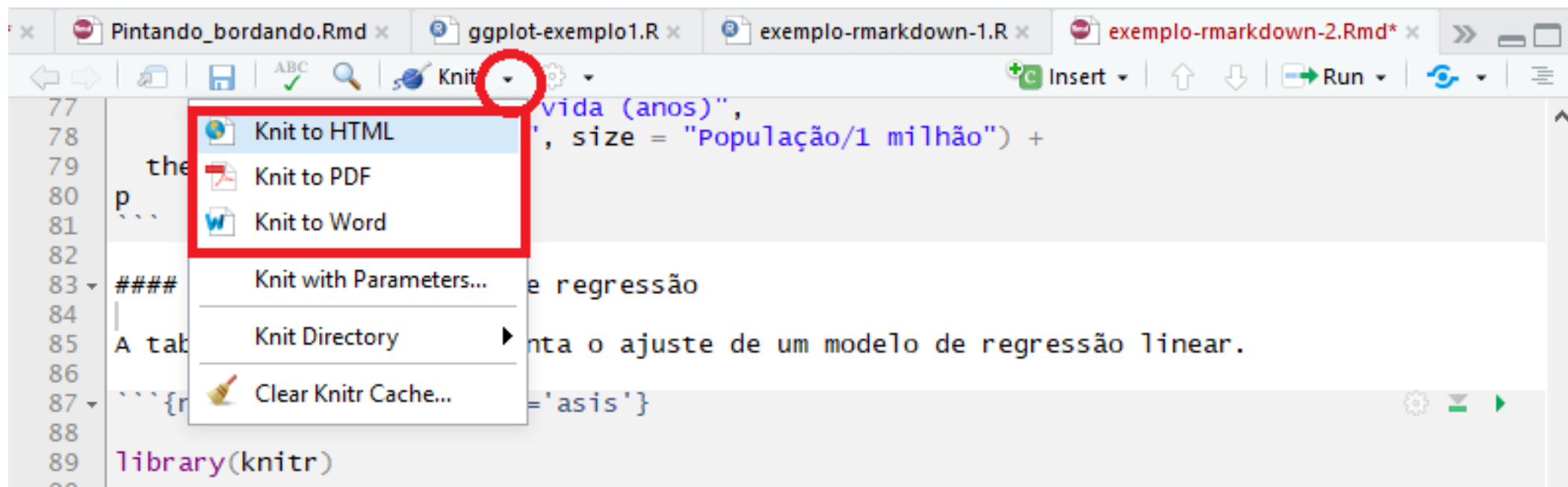
Exercício 1!



1. Crie um arquivo .Rmd.
2. Acrescente texto e a análise do arquivo **exemplo-rmarkdown-1.R**
3. Clique em knit para gerar o arquivo de saída.



- Experimente gerar diferentes formatos de saída.



Por que rmarkdown?



- Reprodutibilidade
- Dinamismo
- Eficiência
- Velocidade



Um último exemplo!





Perguntas!?!

Meus exemplos de R Markdown

- Planos de aula
 - abra o arquivo `Exemplo_rmd_logo_PlanoAula.Rmd` na pasta `estendendo_rmd_exemplos`
 - clique no botão knit
 - se ocorrer erro com o *L^AT_EX* para gerar `.pdf` tente instalar o pacote `tinitex` do R
- Relatório técnico com shiny
 - abra o arquivo `Exemplo_rmd_shiny_MCMLnotes.Rmd`
 - clique no botão > Run Document
- Slides com `ioslides`
 - 1º Datathon UFRGS - ufrgs.br/datathon
 - Arquivos do projeto RStudio em <https://github.com/datathon-ufrgs/website>

Importante: grandes projetos necessitam maior organização!!!

- Projetos em R e `.Rproj`
- Ferramentas Git e Github?

Arquivos . Rmd e *templates*

Para experimentar novos formatos de documentos . Rmd tente:

1. Instale o pacote com o formato desejado;
2. No menu `File > New File > R Markdown > From Template` selecione o formato e clique em OK;

*Note as diferenças no YAML do arquivo . Rmd criado com base.

1. Use o botão para *knit* seu documento e você terá seu novo documento.
- Tente os pacotes ao lado =>



articles



Slides - pacote xaringan

Autor Yiuhi Xie

```
install.packages("xaringan")
```

- Por exemplo, crie um novo documento xaringan em
File > New File > R Markdown > From Template > Ninja Presentation {xaringan}

Outro exemplo:

- Esses slides aqui!!!

Curriculum - pacote vitae

Autor Mitchell O'Hara-Wild

```
install.packages('vitae')
```

<https://github.com/mitchelloharawild/vitae>

<https://docs.ropensci.org/vitae/>

Artigos - pacote r ticles

Autor Yiuhi Xie

```
install.packages("rticles")
```

<https://github.com/rstudio/rticles>

Livros - pacote bookdown

Livro

- bookdown: Authoring Books and Technical Documents with R Markdown
 - <https://bookdown.org/yihui/bookdown/>

Exemplos Juliana: bookdown e learnr

Websites - pacote blogdown

Livro

- blogdown: Creating Websites with R Markdown
 - <https://bookdown.org/yihui/blogdown/>

Github repositories:

- <https://github.com/rbind/blogdown-demo>
- <https://github.com/apreshill/blogdown-workshop>

Pôster - pacote `pagedown`

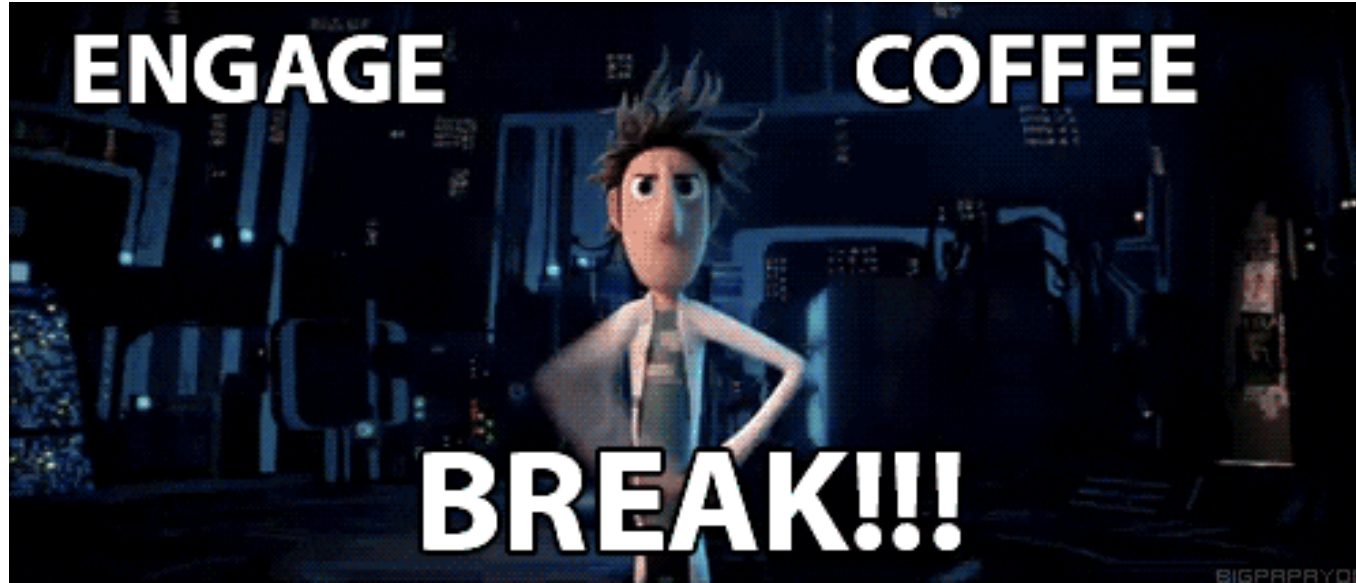
Tutoriais - pacote `learnr`

Provas - pacote `rexams`

Referências (e eventos)

- Workshop UseR!2019 - <https://github.com/rstudio-education/communicate-rmd-workshop>
 - Datathon (Rmarkdown + Shiny): <https://user2019.r-project.org/datathon/>
 - Winner's work: https://user2019.r-project.org/static/uploads/winner_datathon2019.html
- Rstudio Conf 2019 - <https://github.com/dtkaplan/shinymark>
- https://datathon-ufrgs.github.io/Pintando_e_Bordando_no_R/#1
- http://cursos.leg.ufpr.br/prr/capMarkdown.html#escrevendo_um_documento_em_markdown
- *curso-r website*

Thanks!



Fonte: ghipy.com

https://github.com/markus-stein/Rmarkdown_workshop

markus.stein@ufrgs.br

...em breve...

Desafio Datathon UFRGS " *Templates* para o Dest"

"Criar um template de monografias (e dissertações?) para o Dest".

- Os três participantes que mais contribuírem em uma única proposta (ou o grupo com a melhor proposta) serão premiados, além de terem seu trabalho divulgado para o IME e sendo utilizado por muitas pessoas.
 - .Rmd ou *L^AT_EX*?
 - Avaliaremos estilo e normas ABNT, qual será a melhor proposta?
- Temos três canecas para os participantes que trabalharem no desenvolvimento da(s) proposta(s) escolhida(s).

Top secret: Durante o minicurso tivemos a oferta de contribuição com alguma premiação por parte de um dos participantes. ;-)