

Table of Contents

<i>O que é necessário ter instalado em seu computador</i>	1
<i>Dados a serem utilizados na produção do relatório dinâmico</i>	1
<i>Passo 1 - Abrir o RStudio e criar um novo arquivo R Markdown</i>	2
<i>Passo 2 - Exemplo de resumo da coleta</i>	3
<i>Passo 3 - Modifiquei os dados. E agora?</i>	4
<i>Passo 4 - Produção de um relatório dinâmico</i>	5

Relatórios dinâmicos com R Markdown

Uma rotina de trabalho muito comum no meio acadêmico se resume em ir ao campo, coletar dados, fazer sumários estatísticos, escrever relatórios.

Porém, isso acaba se tornando maçante uma vez que fazemos isso frequentemente. E se fôssemos novamente a esse local em que coletamos os dados apresentados acima e coletássemos mais dados? Teríamos que refazer o texto?

A resposta é '**NÃO!**', caso façamos uso de relatórios dinâmicos.

Os pacotes [R Markdown](#) e [knitr](#) providenciam uma excelente maneira de produzir tais relatórios, isto é, atualizados conforme a mudança dos dados. Isso é extremamente útil em todas as áreas que mexem com dados, isto é, TODAS!

Ao se trabalhar com essas ferramentas, não precisamos gastar energia e tempo atualizando o texto conforme a mudança dos dados. Isso resulta em produtividade e mais tempo para se debruçar sobre as análises dos dados e a discussão dos trabalhos.

Eu pretendo neste breve tutorial dar um breve exemplo de como criar um relatório dinâmico utilizando dados hipotéticos de um estudo de diversidade arbórea em parcela permanente.

O que é necessário ter instalado em seu computador

- R instalado e atualizado;
- R Studio instalado;
- Pacotes [R Markdown](#) e [knitr](#) instalados;
- Para rodar o latex e gerar arquivos .pdf é necessários ter instalado alguma distribuição TeX:
 - Linux - instalem o [TeXLive](#);
 - Windows - instalem o [MikTeX](#)
 1. Baixe o "setup-2.9.6100-x64.exe", execute-o e escolha a opção "Download MikTeX" -> "Complete MikTeX" -> escolha o espelho -> e baixe os arquivos. Este processo é um pouco demorado.
 2. Após baixados execute novamente o "setup-2.9.6100-x64.exe" -> "Install MikTeX" -> "Complete MikTeX" -> selecione a pasta com os arquivos baixados na passo anterior -> instale o MikTeX na pasta recomendada (Arquivos de programas). Este processo é um pouco demorado.
 - Mac OS - instalem o [MacTeX](#)

Dados a serem utilizados na produção do relatório dinâmico

Simulei dois arquivos de dados para acompanhar este tutorial. Tratam de hipotéticas coletas de dados de angiospermas arbóreas em um determinado local. Clique nos links abaixo para poder obter os dois conjuntos de dados hipotéticos:

- [dados1](#)
- [dados2](#)

Os dados acima estão separados por tabulação e codificados em 'UTF-8'.

Certifique-se de que seus dados ESTEJAM NA PASTA DE TRABALHO em que você vai EXECUTAR AS ATIVIDADES ABAIXO.

Passo 1 - Abrir o RStudio e criar um novo arquivo R Markdown

Faremos uso do RStudio pois ele já é integrado aos pacotes necessários para criar o relatório dinâmico desta postagem. Portanto, abra o programa e crie um novo arquivo do tipo 'R Markdown', clicando na aba superior **File > New File > R Markdown**.

Uma janela se abrirá com várias opções. Escolha a opção '**Document**' e o formato de saída, definido como **Default Output Format**. Para nosso exemplo, escolha o formato [HTML](#).

Acrescente as informações de autoria e título, e aperte o botão '**OK**'.

O arquivo será aberto contendo um cabeçalho mais ou menos como o abaixo. Em meu caso, coloquei o nome 'Meu nome' como autor, e o título se chama 'Meu relatório dinâmico'. O restante das informações vem atrelado à sua escolha de formato de saída e data; em nosso caso, o formato de saída é [HTML](#).

```
---
title: "Meu relatório dinâmico"
author: "Meu nome"
date: "`r format(Sys.Date(), '%d/%B/%Y')`"
output: html_document
---
```

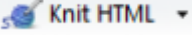
⚠ Esta parte do documento é chamada de YAML é separada do resto do texto por duas linhas com três traços (- - -). É nela que devemos indicar algumas peculiaridades do documento, por exemplo, em um artigo talvez seja necessário colocar outros autores, sua filiação (aqui a filiação vem no rodapé do documento), literatura utilizada (arquivo .bib em bibliography, que deve estar salvo na pasta de trabalho), formato da revista (arquivo .cls em cls, que deve estar salvo na pasta de trabalho) e abstract.

Veja mais comandos para documentos em [HTML](#)! Para documentos ".doc" ! Para documentos ".pdf" !

```
---
title: 'Título do artigo'
author:
- Autor Um^[INPA]
- Autor Dois^[UFAM]
date: "`r format(Sys.Date(), '%d/%B/%Y')`"
abstract: |
  Escreva aqui seu abstract.
output: word_document
bibliography: Lauraceae.bib
```

```
cs1: phytotaxa.cs1  
---
```

O novo arquivo R Markdown também vem com exemplos de disposição do texto e o código no documento. Salve-o em sua pasta de trabalho, e execute-o apertando:

- Usuários Windows e Linux devem apertar as teclas 'CTRL + SHIFT + K';
- Usuários de MAC OS devem apertar 'COMMAND + SHIFT + K';
- Ou apertando o botão do knit  em cima da janela do R markdown no R Studio.

Você deve obter um arquivo '.html' gravado dentro de sua pasta de trabalho, que poderá ser visualizado dentro de qualquer navegador (e.g., Firefox, Safari).

Abra-o e veja o resultado!

Passo 2 - Exemplo de resumo da coleta

O próximo passo é escrever uma frase que contenha um resumo das informações contidas na tabela (dados1). Vamos responder às perguntas abaixo:

- Quantas espécies foram coletadas?
- Qual é a diversidade de gêneros inventariados?
- Qual é o número de parcelas e N amostral?
- Quais são os valores de média e desvio padrão do DAP?

Tradicionalmente, recorreríamos aos dados, anotaríamos os valores e escreveríamos uma frase com essas informações. Abra um R script, copie e cole os comando abaixo, e execute-os para verificar os valores referentes ao número de famílias, gêneros e espécies, além do número de parcelas e indivíduos amostrados, e média e desvio padrão do DAP geral:

```
dados <- read.table('01_blog_tutorial_rmarkdown_dados_01.csv', header = T,  
as.is = T, sep = '\t')  
#numero de familias  
n.fam <- length(unique(dados$FAMILIA))  
print(n.fam)  
#numero de especies  
n.spp <- length(unique(dados$ESPECIE))  
print(n.spp)  
# quantos generos inventariados  
n.gen <- length(unique(dados$GENERO))  
print(n.gen)  
# qual e o numero de parcelas  
n.parcela <- length(unique(dados$PARCELA))  
print(n.parcela)  
# sumario estatistico  
media <- mean(dados$DAP_CM)  
print(media)  
d.pad <- sd(dados$DAP_CM)  
print(d.pad)
```

```
n.amostral <- nrow(dados)
print(n.amostral)
```

Após verificar as respostas de nossas perguntas, pegaríamos esses dados e faríamos uma frase resumindo os resultados como a disposto abaixo:

```
---
title: "Meu relatório dinâmico"
author: "Meu relatório"
date: "`r format(Sys.Date(), '%d/%B/%Y')`"
output: html_document
---

### Resultados
Foram coletadas 14 espécies pertencentes a 5 famílias e 5 gêneros. A amostragem incluiu 2 parcelas e 30 indivíduos. A média de DAP foi 28.933 cm e o desvio padrão foi 13.13572 cm.
```

Abra seu arquivo R Markdown recém criado ".rdm" e COPIE E COLE TODA a informação disposta no bloco de código acima, EXATAMENTE COMO EXPOSTO. COLE essas informações POR CIMA das que estavam anteriormente.

Execute o knit e veja o resultado.

Passo 3 - Modifiquei os dados. E agora?

Copie e cole os comandos do bloco de código abaixo no R script simples. Vejam como ficou nossa tabela de dados após mais uma coleta hipotética (dados2). Execute os comandos e vá comparando com os dados da coleta anterior (dados1). Reparem o quanto mudou após mais uma rodada de coletas.

```
dados2 <- read.table('02_blog_tutorial_rmarkdown_dados_02', header = T,
as.is = T, sep = '\t')
dados2
#numero de familias
n.fam <- length(unique(dados2$FAMILIA))
print(n.fam)
#numero de especies
n.spp <- length(unique(dados2$ESPECIE))
print(n.spp)
# quantos generos inventariados
n.gen <- length(unique(dados2$GENERO))
print(n.gen)
# qual e o numero de parcelas
n.parcela <- length(unique(dados2$PARCELA))
print(n.parcela)
# sumario estatistico
```

```
media <- mean(dados2$DAP_CM)
print(media)
d.pad <- sd(dados2$DAP_CM)
print(d.pad)
n.amostral <- nrow(dados2)
print(n.amostral)
```

Como adendo, vamos fazer um gráfico simples para visualizar um pouco dos dados. Executem o código abaixo e vejam o que acontece. É uma maneira diferente de arrumar os dados para plotar gráficos, diferente do que foi apresentado até então neste curso.

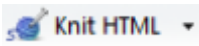
```
#primeiro chama os pacotes
libs <- c('ggplot2', 'magrittr', 'dplyr')
lapply(libs, library, character.only = T, logical.return = T)

#arruma e plota os dados
as.data.frame(count(dados2, FAMILIA)) %>%
  ggplot(aes(x = reorder(FAMILIA, -n), y = n )) + geom_bar(stat =
'identity', width=0.5) +
  coord_flip() + theme_minimal() +
  xlab('Famílias de angiospermas, segundo APG III (2009)') +
ylab('Quantidade de amostras') -> plot.taxa
plot.taxa
```

Isso fica como um brinde e como uma demonstração de como o ambiente R oferece múltiplas ferramentas (= PACOTES!) para visualização de nossos dados.

Passo 4 - Produção de um relatório dinâmico

O próximo passo é escrever o texto e o script no R Markdown para obter esses dados dinamicamente.

Cole os dados do bloco do código abaixo, CONFORME EXPOSTOS, em seu arquivo '.rmd', substituindo o que estava anteriormente. COPIE E COLE TUDO! Ele inclui todas as informações que trabalhamos neste tutorial. Execute-o clicando no botão knit .

```
---
title: "Meu relatório dinâmico"
author: "Meu Nome"
date: "`r format(Sys.Date(), '%d/%B/%Y')`"
output: html_document
---

````{r medidas para relatório, echo = F, eval = T, prompt = T, cache = T,
results = 'hide'}
dados2 <- read.table('02_blog_tutorial_rmarkdown_dados_02', header = T,
as.is = T, sep = '\t')
#numero de familias
n.fam <- length(unique(dados2$FAMILIA))
print(n.fam)
```

```
#numero de especies
n.spp <- length(unique(dados2$ESPECIE))
print(n.spp)
quantos generos inventariados
n.gen <- length(unique(dados2$GENERO))
print(n.gen)
qual e o numero de parcelas
n.parcela <- length(unique(dados2$PARCELA))
print(n.parcela)
sumario estatistico
media <- mean(dados2$DAP_CM)
print(media)
d.pad <- sd(dados2$DAP_CM)
print(d.pad)
n.amostrat <- nrow(dados2)
print(n.amostrat)
````
```

Resultados

Foram coletadas `r n.spp` espécies pertencentes a `r n.fam` famílias e `r n.gen` gêneros. A amostragem incluiu `r n.parcela` parcelas e `r n.amostrat` indivíduos. A média de DAP foi `r media` cm e o desvio padrão `r d.pad` cm.

A figura abaixo mostra quantidade de amostras coletadas ordenadas da menos abundante para a mais abundante.

```
````{r graficos, echo = F, eval = T, prompt = T, comments = '', message = F,
warnings = ''}
library(ggplot2)
library(magrittr)
library(dplyr)
as.data.frame(count(dados2, FAMILIA)) %>%
 ggplot(aes(x = reorder(FAMILIA, -n), y = n)) + geom_bar(stat =
'identity', width=0.5) +
 coord_flip() + theme_minimal() +
 xlab('Famílias de angiospermas, segundo APG III (2009)') +
 ylab('Quantidade de amostras') -> plot.taxa
plot.taxa
````
```

Viram em que resultou o uso da notação `r FORMULA` no texto? Não é uma maneira prática para mesclar texto e código do chunk?

Repararam também que fizemos uso de outro tipo de notação, demonstrado abaixo:

```
````{r NOME, PARÂMETROS}
COMANDO1
COMANDO2
````
```


Isso é chamado de 'code chunk', ou bloco de código literalmente. Essa notação é outra forma que permite mesclar código e texto, porém que se apresenta em blocos.

🚨 Caso esteja curioso, visite o site do criador do pacote 'knitr', [Yihui Xie](#). Lá há muito material para que se possa compreender melhor o uso dessa notação.

O resultado desta ação após a compilação dos dados deve ser exatamente igual ao arquivo [aqui disponível](#).

From:

<http://www.botanicaamazonica.wiki.br/labotam/> - **Ecologia e Evolução de Plantas Amazônicas**

Permanent link:

<http://www.botanicaamazonica.wiki.br/labotam/doku.php?id=bot89:precurso:rmarkdown:inicio>

Last update: **22/50/2016 00:50**