

PROGRAMAÇÃO

5. Geração de Relatórios

Rita P. Ribeiro

Mestrado em Informática Médica

Curso de Especialização em Informática da Saúde

2016/2017



FACULDADE DE MEDICINA
UNIVERSIDADE DO PORTO



FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Programa

1. Conceitos Básicos do R
2. Manipulação de Dados
3. Sumarização de Dados
4. Visualização de Dados
5. Geração de Relatórios

Reporting: a Abordagem Tradicional

- ▶ O Processo Usual de Análise de Dados
 - ▶ Importar os dados para a nossa ferramenta de análise de dados
 - ▶ Levar a cabo a nossa análise dados
 - ▶ Fazer *copy+paste* de partes da análise e/ou resultados para um programa de processamento de texto e/ou produção de apresentações, juntando algum texto explicativo
 - ▶ Frequentemente é necessário repetir todo o processo de novo!

Reporting: a Abordagem Tradicional

Algumas Desvantagens

- ▶ Demasiados passos manuais → grande potencial para erros humanos
- ▶ Processo com muito esforço humano e frequentemente em tarefas repetitivas e aborrecidas, por exemplo de comunicação entre as diferentes ferramentas
- ▶ Todo o processo é dificilmente registável para o futuro, uma vez que grande parte envolve interação com um interface gráfico
- ▶ Pequenas alterações nos dados originais exigem a repetição de todo o processo!
- ▶ A análise e o reporting estão separados pelo que é preciso grande cuidado na “sincronização” destas duas componentes
- ▶ É difícil passar o trabalho para outra equipa e aproveitar o trabalho feito para o aplicar a problemas semelhantes

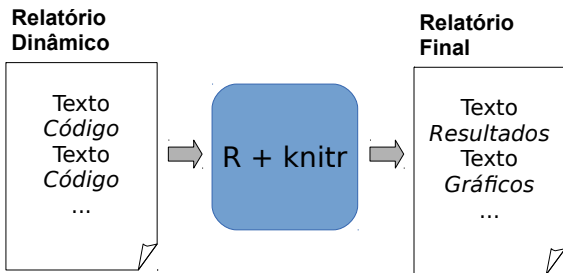
Lista inspirada em *Dynamic Documents with R and knitr* por Yihui Xie

Reporting com Documentos Dinâmicos

- ▶ Em que consistem os Documentos Dinâmicos?
 - ▶ Documentos que misturam passos da análise com a narrativa do relatório
 - ▶ Documentos que são “compiláveis” para o documento final
 - ▶ O documento final é portanto produzido por um computador a partir do código da análise de dados e da narrativa escrita pelo autor
- ▶ Os Documentos Dinâmicos resolvem a maioria dos problemas referidos atrás!

Reporting com Documentos Dinâmicos

R+knitr - uma implementação da ideia



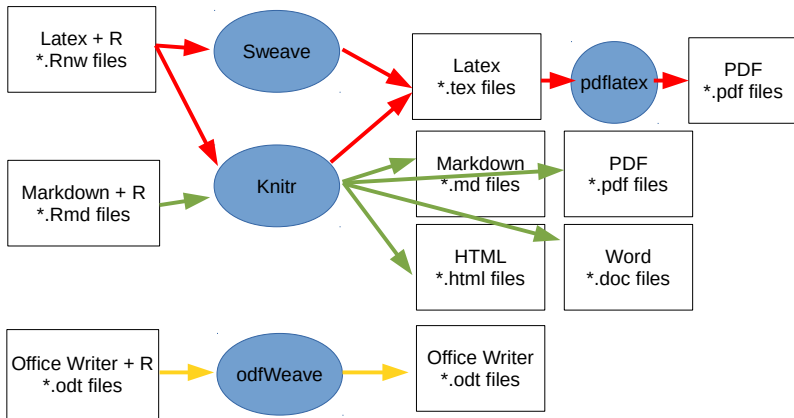
Reporting com Documentos Dinâmicos

Passos da implementação

- ▶ A ideia está relacionada com o conceito de *literate programming*
- ▶ Implementado em 3 passos principais:
 1. Inspeccionar o relatório dinâmico e separar o código das narrativas
 2. Executar o código e guardar os resultados de cada bloco de código
 3. Introduzir esses resultados de novo no meio das narrativas gerando o documento final
- ▶ NOTA: todo o processo é feito sem intervenção humana!

Reporting com Documentos Dinâmicos

Diferentes implementações do conceito

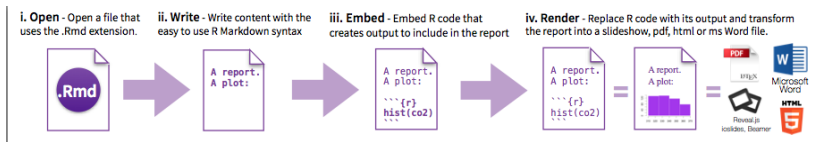


Reporting com Documentos Dinâmicos

R+knitr

- ▶ O knitr é uma plataforma que permite:
 - ▶ Usar diferentes formatos de documentos
 - ▶ $\text{\LaTeX}$, HTML e Markdown
 - ▶ Usar diferentes linguagens de programação para implementar a análise
 - ▶ R, Python, awk, C++ e *shell scripts*

Vamo-nos focar na combinação Markdown + R



Markdown

Breve Introdução

- ▶ **Markdown** é uma linguagem muito simples que foi desenhada para facilitar a criação de conteúdo para a Internet
 - ▶ Muito mais simples que a linguagem nativa - HTML
 - ▶ Existe uma ferramenta que pode ser usada para converter Markdown para HTML
- ▶ O formato é texto simples pelo que não é preciso nenhuma ferramenta especial para criar um ficheiro Markdown

Markdown

Formatação

- ▶ A formatação é conseguida à custa de pequenas anotações
- ▶ Exemplos:

```
*Nota* or _Nota_
```

resulta em *Nota* (i.e. itálico)

```
**Nota** or __Nota__
```

resulta em **Nota** (i.e. bold)

Markdown

Secções e sub-secções

- ▶ Um cardinal (#) antes de algum texto indica um *Heading 1*
- ▶ Usando dois ou três cardinais obtemos *Headings 2 e 3*

```
# Illustrations on the use of Markdow in Dynamic Reports
```

```
## Introduction
```

```
R comes with a series of data sets. We may use the function  
**data** to load them, as show in the following example:
```

Illustrations on the use of Markdow in Dynamic Reports

Introduction

R comes with a series of data sets. We may use the function **data** to load them, as show in the following example:

Markdown

Listas de Items

```
# OM prototype - Admin User

+ Stage 1
  - Define topics of interest
  - Define web sources
    * create crawlers (**Potentially Challenging**)
    * collect data
+ Stage 2
  - Obtain tagged data (**Potentially Challenging**)
  - Obtain, evaluate and select models
```

OM prototype - Admin User

- Stage 1
 - Define topics of interest
 - Define web sources
 - create crawlers (Potentially Challenging)
 - collect data
- Stage 2
 - Obtain tagged data (Potentially Challenging)
 - Obtain, evaluate and select models

Markdown

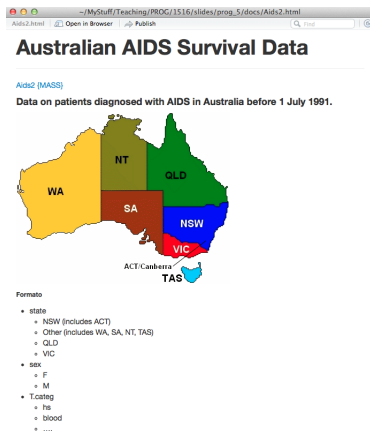
Imagens embebidas e links

```
# Australian AIDS Survival Data
***
[Aids2 {MASS}](https://stat.ethz.ch/R-manual/R-devel/library/MASS...)

#### Data on patients diagnosed with AIDS in Australia before ...

![Australia](australia.jpg "Australia Map")

##### Formato
- state
  - NSW (includes ACT)
  - Other (includes WA, SA, NT, TAS)
  - QLD
  - VIC
- sex
  - F
  - M
- T.categ
  - hs
  - blood
  - ....
```



R Markdown

Package `rmarkdown`

- ▶ `rmarkdown` é uma package do R que permite integrar o R com a linguagem Markdown usando o **knitr** e o **pandoc**
- ▶ **Pandoc** é uma espécie de canivete suíço da conversão de formatos de documentos
- ▶ **R markdown** é uma linguagem de *literate programming* que permite integrar R dentro de documentos markdown



R Markdown

Embeber código R nos documentos

```
# The Distribution of Patients by State
```

An histogram of the variable
can be obtained with:

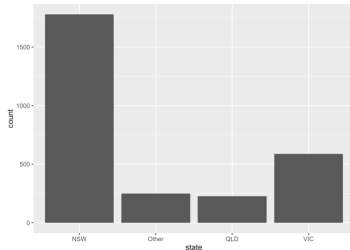
```
```{r eval=FALSE}
ggplot(Aids2,aes(x=state)) + geom_bar()
```
```

```
```{r echo=FALSE, warning=FALSE}
library(ggplot2)
data(Aids2,package="MASS")
ggplot(Aids2,aes(x=state)) + geom_bar()
```
```

The Distribution of Patients by State

A barplot of the variable can be obtained with:

```
ggplot(Aids2,aes(x=state)) + geom_bar()
```



No RStudio: Chunk > Insert Chunk

R Markdown

Blocos de código em R

- ▶ Os blocos de código são partes do documento que deverão ser executadas pelo R
- ▶ Em R Markdown elas são indicadas da forma seguinte:

```
```${r}```  

```\n
```

- ▶ Qualquer coisa dentro do bloco será executada pelo R
- ▶ O resultado da execução será incluído no documento final
- ▶ Também podemos ter código no meio de uma frase fazendo:

```
0 resultado de 2 + 2 = `r 2+2`
```

R Markdown

Opções dos Blocos de Código

- ▶ A forma como os blocos são interpretados pelo **knitr** pode ser controlada por uma série de opções dos blocos
- ▶ Sem qualquer opção os blocos são:
 1. Executados
 2. O código é incluído no documento final
 3. Os resultados também são incluídos no documento final
- ▶ As opções dos blocos permitem ajustar estes *defaults*
- ▶ Isto pode ser feito bloco a bloco ou globalmente para todo o documento

R Markdown

Algumas opções comuns

eval (TRUE) - permite controlo se o código no bloco é ou não executado pelo R. No seguinte exemplo o código aparece no documento final mas não é executado pelo R

```
```{r eval=FALSE}  
ggplot(Aids2,aes(x=state)) + geom_bar()
```
```

echo (TRUE) - permite esconder o código R do documento final se escolhermos o valor FALSE, fazendo com que só os resultados da execução do código sejam visíveis (p.ex. uma figura)

- ▶ Qualquer bloco pode incluir várias opções separadas por vírgulas
- ▶ Existem muitas outras opções! Ver uma lista exhaustiva e explicações em <http://yihui.name/knitr/options>

Gerar o Documento Final

Usando a consola do R

- ▶ A função `render()` da package `rmarkdown` recebe como input um ficheiro markdown e produz o documento final usando o R para executar os blocos de código:

```
library(rmarkdown)
render("Relatorio.Rmd")
```

- ▶ O tipo de documento que é gerado por ser controlado através de alguma meta-informação que pode ser incluída no documento markdown,

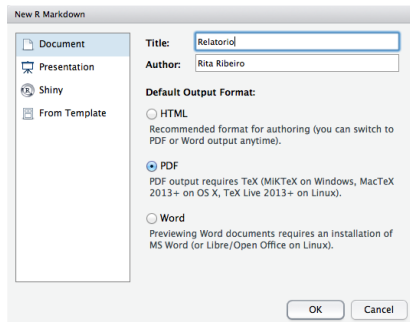
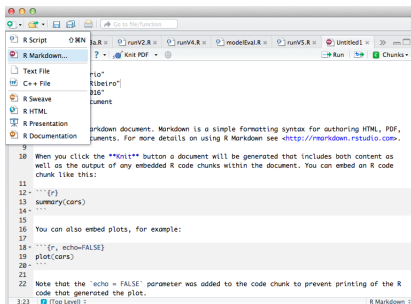
```
---
title: "Relatorio"
author: "Rita Ribeiro"
date: "9 Mar 2016"
output: pdf_document
---
```

Gerar o Documento Final

Usando o interface do RStudio

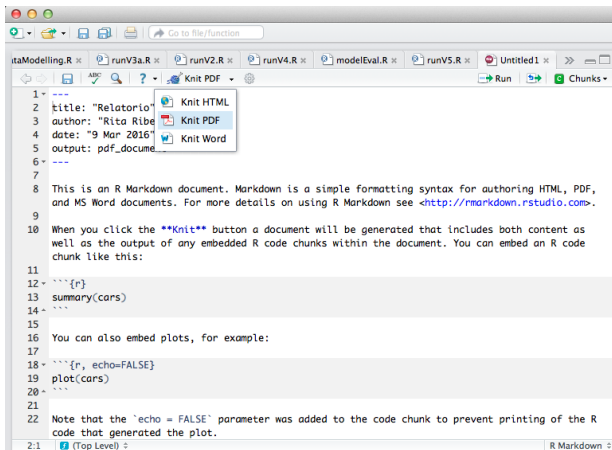
As versões mais recentes do **RStudio** facilitam a criação de documentos dinâmicos

File > New File > R Markdown



Gerar o Documento Final

Usando o interface do RStudio (2)



Aprender mais

Bastante mais informação está disponível em
<http://rmarkdown.rstudio.com/>