Authoring R Markdown vignettes

Andrzej Oleś, Wolfgang Huber and Martin Morgan

Last update: 21 October, 2017

Package

systemPipeR 1.10.0

Contents

1 Prerequisites

Bioconductor R Markdown format is build on top of *R* package *bookdown*, which in turn relies on *rmarkdown* and pandoc to compile the final output document. Therefore, unless you are using RStudio, you will need a recent version of pandoc ($>=$ 1.17.2). See the pandoc installation instructions for details on installing pandoc for your platform.

2 Getting started

To enable the *Bioconductor* style in your R Markdown vignette you need to (H Backman and Girke 2016):

- Edit the `DESCRIPTION` file by adding

  ```
  VignetteBuilder: knitr
  Suggests: BiocStyle, knitr, rmarkdown
  ```

- Specify `BiocStyle::html_document` or `BiocStyle::pdf_document` as output format and add vignette metadata in the document header:

  ```
  ---
  title: "Vignette Title"
  author: "Vignette Author"
  package: PackageName
  output:
    BiocStyle::html_document
  vignette: >
    %\VignetteIndexEntry{Vignette Title}
    %\VignetteEngine{knitr::rmarkdown}
    %\VignetteEncoding{UTF-8}
  ---
  ```

The `vignette` section is required in order to instruct *R* how to build the vignette.[1] The `package` field which should contain the package name is used to print the package version in the output document header. It is not necessary to specify `date` as by default the document compilation date will be automatically included. See the following section for details on specifying author affiliations and abstract.

BiocStyle's `html_document` and `pdf_document` format functions extend the corresponding original *rmarkdown* formats, so they accept the same arguments as `html_document` and `pdf_document`, respectively. For example, use `toc_float: true` to obtain a floating TOC as in this vignette.

[1] `\VignetteIndexEntry` should match the `title` of your vignette

2.1 Use with R markdown v1

Apart from the default markdown engine implemented in the *rmarkdown* package, it is also possible to compile *Bioconductor* documents with the older markdown v1 engine from the package *markdown*. There are some differences in setup and the resulting output between these two engines.

To use the *markdown* vignette builder engine:

- Edit the `DESCRIPTION` file to include

```
VignetteBuilder: knitr
Suggests: BiocStyle, knitr
```

- Specify the vignette engine in the `.Rmd` files (inside HTML comments)

```
<!--
%% \VignetteEngine{knitr::knitr}
-->
```

- Add the following code chunk at the beginning of your `.Rmd` vignettes

```
```{r style, echo = FALSE, results = 'asis'}
BiocStyle::markdown()
```
```

The way of attaching CSS files when using *markdown* differs from how this is done with *rmarkdown*. In the former case additional style sheets can be used by providing them to the `BiocStyle::markdown` function. To include `custom.css` file use

```
```{r style, echo = FALSE, results = 'asis'}
BiocStyle::markdown(css.files = c('custom.css'))
```
```

3   Document header

3.1   Author affiliations

The `author` field allows for specifing author names along with affiliation and email information.

In the basic case, when no additional information apart from author names is provided, these can be entered as a single character string

```
author: "Single Author"
```

or a list

```
author:
- First Author
- Second Author
- Last Author
```

which will print as "First Author, Second Author and Last Author".

Author affiliations and emails can be entered in named sublists of the author list. Multiple affiliations per author can be specified this way.

```
author:
- name: First Author
  affiliation:
```

```
    - Shared affiliation
    - Additional affiliation
- name: Second Author
  affiliation: Shared affiliation
  email: corresponding@author.com
```

A list of unique affiliations will be displayed below the authors, similar as in this document.

For clarity, compactness, and to avoid errors, repeated nodes in YAML header can be initially denoted by an anchor entered with an ampersand &, and later referenced with an asterisk *. For example, the above affiliation metadata is equivalent to the shorthand notation

```
author:
- name: First Author
  affiliation:
    - &id Shared affiliation
    - Additional affiliation
- name: Second Author
  affiliation: *id
  email: corresponding@author.com
```

3.2   Abstract and running headers

Abstract can be entered in the corresponding field of the document front matter, as in the example below.

```
---
title: "Full title for title page"
shorttitle: "Short title for headers"
author: "Vignette Author"
package: PackageName
abstract: >
  Document summary
output:
  BiocStyle::pdf_document
---
```

The `shorttitle` option specifies the title used in running headers instead of the document title.[2]

[2]only relevant to PDF output

4   Style macros

BiocStyle introduces the following macros useful when referring to *R* packages:

- `Biocpkg("IRanges")` for *Bioconductor* software, annotation and experiment data packages, including a link to the release landing page or if the package is only in devel, to the devel landing page, *IRanges*.

- `CRANpkg("data.table")` for *R* packages available on CRAN, including a link to the FHCRC CRAN mirror landing page, *data.table*.

- `Githubpkg("rstudio/rmarkdown")` for *R* packages available on GitHub, including a link to the package repository, *rmarkdown*.

- `Rpackage("MyPkg")` for *R* packages that are *not* available on *Bioconductor*, CRAN or GitHub; *MyPkg*.

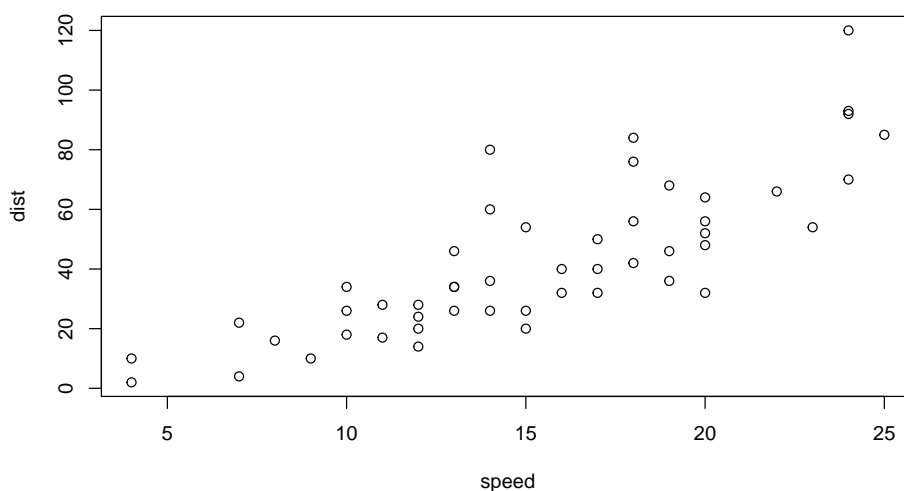These are meant to be called inline, e.g., `'r Biocpkg("IRanges")'`.

5   Code chunks

The line length of output code chunks is set to the optimal width of typically 80 characters, so it is not neccessary to adjust it manually through `options("width")`.

6   Figures

BiocStyle comes with three predefined figure sizes. Regular figures not otherwise specified appear indented with respect to the paragraph text, as in the example below.

```
plot(cars)
```


Figures which have no captions are just placed wherever they were generated in the *R* code. If you assign a caption to a figure via the code chunk option `fig.cap`, the plot will be automatically labeled and numbered[3], and it will be also possible to reference it. These features are provided by *bookdown*, which defines a format-independent syntax for specifying cross-references, see Section 9. The figure label is generated from the code chunk label[4] by prefixing it with `fig:`, e.g., the label of a figure originating from the chunk `foo` will be `fig:foo`. To reference a figure, use the syntax `\@ref(label)`[5], where `label` is the figure label, e.g., `fig:foo`. For example, the following code chunk was used to produce Figure 1.

```
```{r plot, fig.cap = "Regular figure. The first sentence...", echo = FALSE}
plot(cars)
```
```

[3]for PDF output it will be placed in a floating figure environment

[4]for cross-references to work the chunk label may only contain alphanumeric characters (a-z, A-Z, 0-9), slashes (/), or dashes (-)
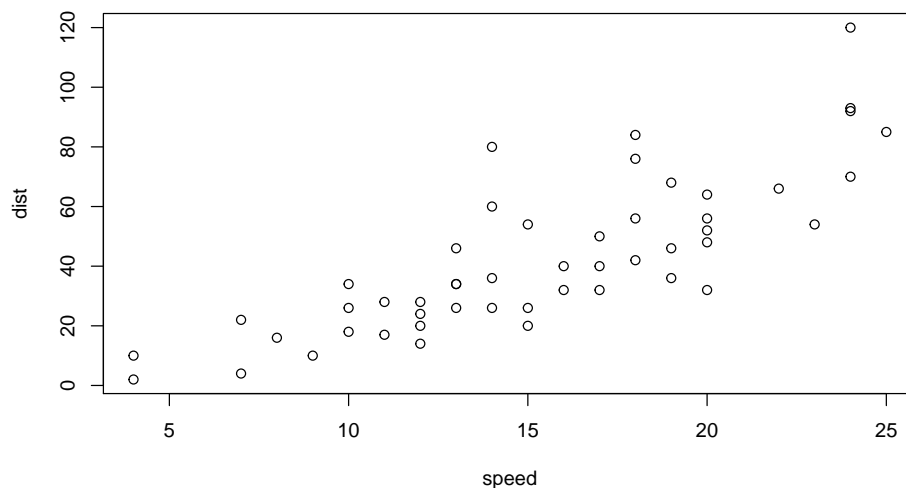
[5]do not forget the leading backslash!

Figure 1: Regular figure

The first sentence of the figure caption is automatically emphasized to serve as figure title.

In addition to regular figures, *BiocStyle* provides small and wide figures which can be specified by `fig.small` and `fig.wide` code chunk options. Wide figures are left-aligned with the paragraph and extend on the right margin, as Figure 2. Small figures are meant for possibly rectangular plots which are centered with respect to the text column, see Figure 3.
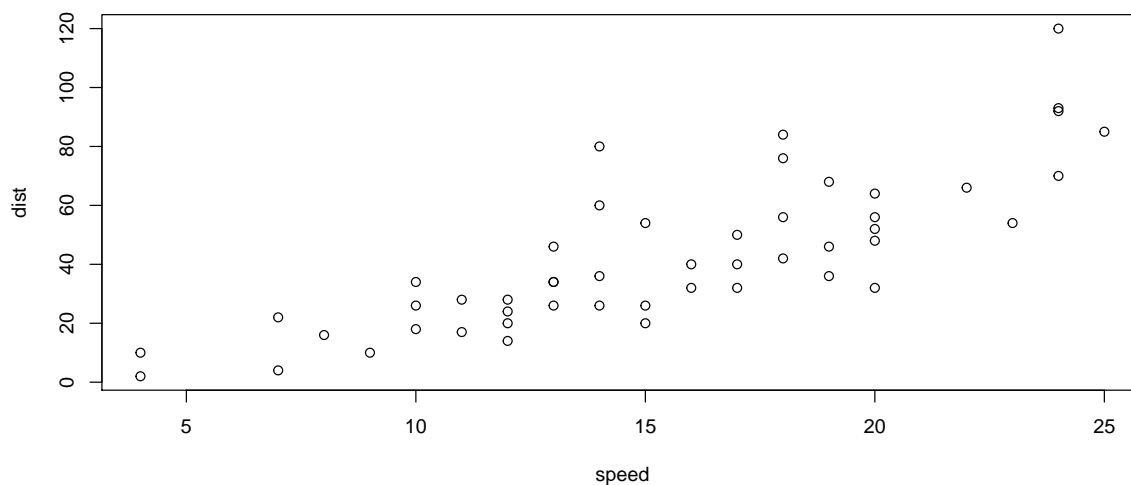

Figure 2: Wide figure

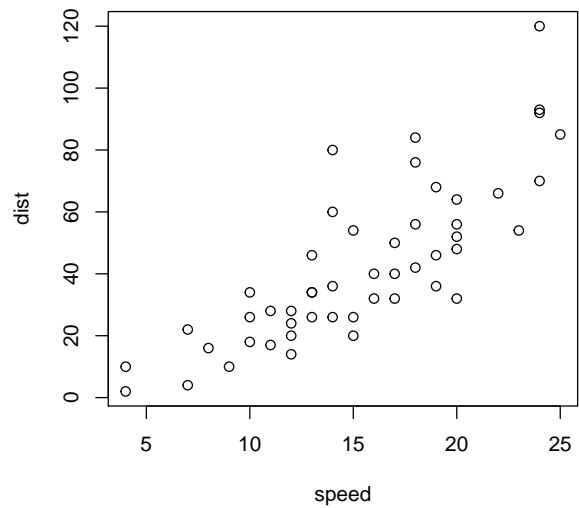A plot produced by a code chunk with option 'fig.wide = TRUE'.

Figure 3: Small figure
A plot produced by a code chunk with option 'fig.small = TRUE'.

7 Tables

Like figures, tables with captions will also be numbered and can be referenced. The caption is entered as a paragraph starting with `Table:`[6], which may appear either before or after the table. When adding labels, make sure that the label appears at the beginning of the table caption in the form `(\#tab:label)`, and use `\@ref(tab:label)` to refer to it. For example, Table 1 has been produced with the following code.

```
Fruit   | Price
------- | -----
bananas | 1.2
apples  | 1.0
oranges | 2.5


: (\#tab:table) A simple table. With caption.
```

Table 1: A simple table
With caption.

Fruit	Price
bananas	1.2
apples	1.0
oranges	2.5

The function `knitr::kable()` will automatically generate a label for a table environment, which is the chunk label prefixed by `tab:`, see Table 2.

```
knitr::kable(
  head(mtcars[, 1:8], 10), caption = 'A table of the first 10 rows of `mtcars`.'
)
```

Table 2: A table of the first 10 rows of 'mtcars'

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0

[6]or just `:`

```
\begin{equation}
  f\left(k\right) = \binom{n}{k} p^k\left(1-p\right)^{n-k}
  (\#eq:binom)
\end{equation}
```

renders the equation below.

$$f\left(k\right) = \binom{n}{k} p^k \left(1-p\right)^{n-k}$$ **1**

You may then refer to Equation **1** by `\@ref(eq:binom)`. Note that in HTML output only labeled equations will appear numbered.

9 Cross-references

Apart from referencing figures (Section 6), tables (Section 7), and equations (Section 8), you can also use the same syntax `\@ref(label)` to reference sections, where `label` is the section ID. By default, Pandoc will generate IDs for all section headers, e.g., `# Hello World` will have an ID `hello-world`. In order to avoid forgetting to update the reference label after you change the section header, you may also manually assign an ID to a section header by appending `{#id}` to it.

When a referenced label cannot be found, you will see two question marks like **??**, as well as a warning message in the R console when rendering the document.

10 Margin notes

Footnotes are displayed as side notes on the right margin[8], which has the advantage that they appear close to the place where they are defined.

[8] this is a side note entered as a footnote

Session info

Here is the output of `sessionInfo()` on the system on which this document was compiled:

```
## R version 3.4.2 (2017-09-28)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.5 LTS
##
## Matrix products: default
## BLAS: /usr/lib/libblas/libblas.so.3.0
## LAPACK: /usr/lib/lapack/liblapack.so.3.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
```

```
##  [5] LC_MONETARY=en_US.UTF-8     LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8        LC_NAME=C
##  [9] LC_ADDRESS=C                LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  utils     datasets  grDevices base
##
## other attached packages:
## [1] BiocStyle_2.5.41
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.4.2  backports_1.1.0 bookdown_0.4    magrittr_1.5
##  [5] rprojroot_1.2   tools_3.4.2     htmltools_0.3.6 yaml_2.1.14
##  [9] Rcpp_0.12.11    stringi_1.1.5   rmarkdown_1.6   knitr_1.16
## [13] methods_3.4.2   stringr_1.2.0   digest_0.6.12   evaluate_0.10.1
```

References

H Backman, Tyler W, and Thomas Girke. 2016. "systemPipeR: NGS workflow and report generation environment." *BMC Bioinformatics* 17 (1): 388. doi:10.1186/s12859-016-1241-0.