

# Writing reproducible reports

Knitr with R Markdown

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

`biostat.wisc.edu/~kbroman`

`github.com/kbroman`

`@kwbroman`

Course web: [bit.ly/tools4rr](https://bit.ly/tools4rr)

# Knitr in a Knutshell

`bit.ly/knitrknutshell`

# Data analysis reports

- ▶ Figures/tables + email
- ▶ Static  $\text{\LaTeX}$  or Word document
- ▶ Knitr/Sweave +  $\text{\LaTeX}$   $\rightarrow$  PDF
- ▶ Knitr + Markdown  $\rightarrow$  Web page

What if the data change?

What if you used the wrong version of the data?

# KnitR code chunks

## Input to knitr:

```
We see that this is an intercross with `r nind(sug)`  
individuals. There are `r nphe(sug)` phenotypes, and genotype  
data at `r totmar(sug)` markers across the `r nchr(sug)`  
autosomes. The genotype data is quite complete.
```

```
```{r summary_plot, fig.height=8}  
plot(sug)  
```
```

## Output from knitr:

```
We see that this is an intercross with 163  
individuals. There are 6 phenotypes, and genotype  
data at 93 markers across the 19  
autosomes. The genotype data is quite complete.
```

```
```r  
plot(sug)  
```
```

```
![plot of chunk summary_plot](RmdFigs/summary_plot.png)
```

# html

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset=utf-8"/>
  <title>Example html file</title>
</head>

<body>
<h1>Markdown example</h1>

<p>Use a bit of <strong>bold</strong> or <em>italics</em>. Use
backticks to indicate <code>code</code> that will be rendered
in monospace.</p>

<ul>
<li>This is part of a list</li>
<li>another item</li>
</ul>

</body>
</html>
```

# CSS

```
ul,ol {  
  margin: 0 0 0 35px;  
}  
  
a {  
  color: purple;  
  text-decoration: none;  
  background-color: transparent;  
}  
  
a:hover  
{  
  color: purple;  
  background: #CAFFFF;  
}
```

[Example]

# Markdown

```
# Markdown example
```

```
Use a bit of bold or italics. Use backticks to indicate  
`code` that will be rendered in monospace.
```

- This is part of a list
- another item

```
Include blocks of code using three backticks:
```

```
```\n x <- rnorm(100)\n```
```

```
Or indent four spaces:
```

```
    mean(x)\n    sd(x)
```

```
And it's easy to create links, like to  
[Markdown](http://daringfireball.net/projects/markdown/).
```



# R Markdown

- ▶ R Markdown is a variant of Markdown, developed at [RStudio.com](https://RStudio.com)
- ▶ Markdown + KnitR + extras
- ▶ A few extra marks
- ▶  $\text{\LaTeX}$  equations
- ▶ Bundle images into the final html file

# Code chunks, again

```
```{r knitr_options, include=FALSE}  
opts_chunk$set(fig.width=12, fig.height=4, fig.path='Figs/',  
               warning=FALSE, message=FALSE)  
set.seed(53079239)  
```
```

### Preliminaries

Load the R/qtl package using the ``library`` function:

```
```{r load_qtl}  
library(qtl)  
```
```

To get help on the `read.cross` function in R, type one of the following:

```
```{r help, eval=FALSE}  
help(read.cross)  
?read.cross  
```
```

# Chunk options

|                               |                                |
|-------------------------------|--------------------------------|
| <code>echo=FALSE</code>       | Don't include the code         |
| <code>results="hide"</code>   | Don't include the output       |
| <code>include=FALSE</code>    | Don't show code or output      |
| <code>eval=FALSE</code>       | Don't evaluate the code at all |
| <code>warning=FALSE</code>    | Don't show R warnings          |
| <code>message=FALSE</code>    | Don't show R messages          |
| <code>fig.width=#</code>      | Width of figure                |
| <code>fig.height=#</code>     | Height of figure               |
| <code>fig.path="Figs/"</code> | Path for figure files          |

There are **lots of chunk options**.

# Global chunk options

```
```{r knitr_options, include=FALSE}
opts_chunk$set(fig.width=12, fig.height=4, fig.path='Figs/',
               warning=FALSE, message=FALSE,
               include=FALSE, echo=FALSE)
set.seed(53079239)
```

```{r make_plot, fig.width=8, include=TRUE}
x <- rnorm(100)
y <- 2*x + rnorm(100)
plot(x, y)
```
```

- ▶ Use global chunk options rather than repeat the same options over and over.
- ▶ You can override the global values in specific chunks.

# Package options

```
```{r package_options, include=FALSE}  
opts_knit$set(progress = TRUE, verbose = TRUE)  
```
```

- ▶ It's easy to confuse global **chunk options** with **package options**.
- ▶ I've not used package options.
- ▶ So focus on **opts\_chunk\$set()** not **opts\_knit\$set()**.

# In-line code

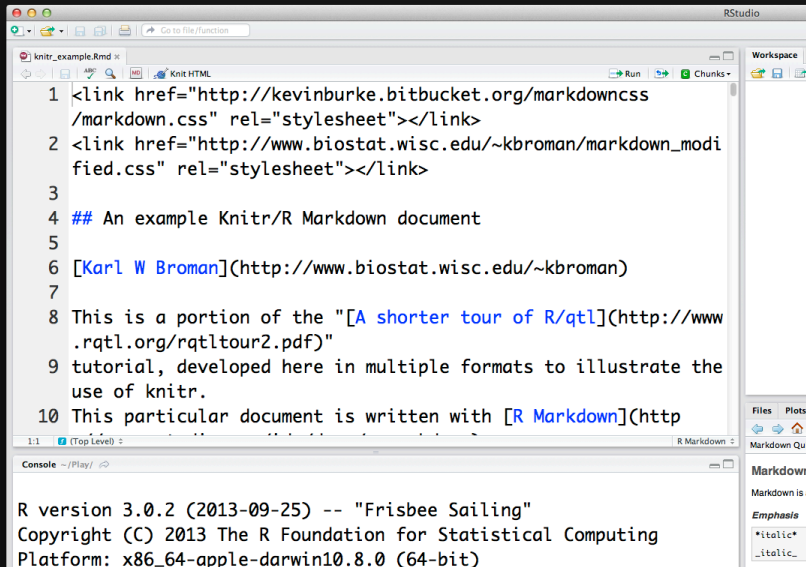
```
We see that this is an intercross with `r nind(sug)`  
individuals. There are `r nphe(sug)` phenotypes, and genotype  
data at `r totmar(sug)` markers across the `r nchr(sug)`  
autosomes. The genotype data is quite complete.
```

- ▶ Each bit of in-line code needs to be within one line; they can't span across lines.
- ▶ I'll often precede a paragraph with a code chunk with `include=FALSE`, defining various variables, to simplify the in-line code.
- ▶ Never hard-code a result or summary statistic again!

# Rounding

- ▶ `cor(x,y)` might produce `0.8992877`, but I want `0.90`.
- ▶ `round(cor(x,y), 2)`, would give `0.9`, but I want `0.90`.
- ▶ You could use `sprintf("%.2f", cor(x,y))`, but `sprintf("%.2f", -0.001)` gives `-0.00`.
- ▶ Use the `myround` function in my [R/broman](#) package.
- ▶ `myround(cor(x,y), 2)` solves both issues.

# R Markdown → html, in RStudio



The screenshot shows the RStudio application window. The main editor pane displays an R Markdown document named 'knitr\_example.Rmd'. The document content is as follows:

```
1 <link href="http://kevinburke.bitbucket.org/markdowncss
2 /markdown.css" rel="stylesheet"></link>
3
4 ## An example Knitr/R Markdown document
5
6 [Karl W Broman](http://www.biostat.wisc.edu/~kbroman)
7
8 This is a portion of the "[A shorter tour of R/qtl](http://www
9 .rqt1.org/rqt1tour2.pdf)"
10 tutorial, developed here in multiple formats to illustrate the
    use of knitr.
11 This particular document is written with [R Markdown](http
```

The status bar at the bottom of the editor indicates '1:1 (Top Level)' and 'R Markdown'. Below the editor is the Console pane, which shows the output of the R session:

```
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0 (64-bit)
```

On the right side of the RStudio window, the 'Workspace' pane is visible, showing a list of files and plots. The 'Files' pane shows a list of files, including 'knitr\_example.Rmd'. The 'Plots' pane shows a list of plots, including 'R Markdown Quickstart'.



# R Markdown → html, in R

```
> library(knitr)
> knit("knitr_example.Rmd")
> library(markdown)
> knit("knitr_example.md", "knitr_example.html")
```

```
> library(knitr)
> knit2html("knitr_example.Rmd")
```

# R Markdown → html, GNU make

```
knitr_example.html: knitr_example.Rmd  
  R -e 'library(knitr);knit2html("knitr_example.Rmd")'
```

# Reproducible knitr documents

- ▶ Don't use absolute paths like `~/Data/blah.csv`
- ▶ Keep all of the code and data in one directory (and its subdirectories)
- ▶ If you **must** use absolute paths, define the various directories with variables at the top of your document.
- ▶ Use R `--vanilla` or perhaps  
`R --no-save --no-restore --no-init-file --no-site-file`
- ▶ Use GNU make to document the construction of the final product (tell future users what to do)
- ▶ Include a chunk with `sessionInfo()` at the end
- ▶ For simulations, use `set.seed` in your first chunk.

# Controlling figures

```
```{r test_figure, dev.args=list(pointsize=18)}  
par(cex.axis=1.5, cex.lab=1.5)  
x <- rnorm(100)  
y <- 2*x + rnorm(100)  
plot(x,y)  
```
```

- ▶ The default is for KnitR/R Markdown is to use the `png()` graphics device.
- ▶ Use another graphics device with the chunk option `dev`.
- ▶ Pass arguments to the graphics device via the chunk option `dev.args`.
- ▶ To increase the size of axis labels, I seem to need `par(cex.lab=1.5)` as well as `dev.args=list(pointsize=18)`

# Tables

```
```{r kable, results="asis"}
x <- rnorm(100)
y <- 2*x + rnorm(100)
out <- lm(y ~ x)
kable(summary(out)$coef, format="html",
       digits=2)
```
```

```
```{r xtable, results="asis"}
library(xtable)
tab <- xtable(coef_tab, digits=c(0, 2, 2, 1, 3))
print(tab, type="html")
```
```

# Important principles

Modify your desires to match the defaults.

Focus your compulsive behavior on things that matter.