KnitR + \LaTeX \rightarrow \text{paper}

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

kbroman.org
github.com/kbroman
@kwlbroman
Course web: kbroman.org/AdvData
R Markdown for everything

- bookdown for books (or book-like things)
- blogdown for websites
- pkgdown for package websites
- xaringan for slides
- pagedown for CVs, resumes, and letters
- posterdown for posters
\documentclass[12pt]{article}
\usepackage{graphicx}
\title{An example document}
\author{Karl Broman}
\begin{document}
\maketitle
\thispagestyle{empty}
\section{A section}

This is a simple example of a \LaTeX\ document for an article. Here's some in-line math: $y = \beta_0 + \beta_1 x + \epsilon$.

And here's a display equation:

$$\hat{\beta} = (X'X)^{-1} X'y$$

\end{document}
What I actually do

\documentclass[12pt]{article}
\setlength{\headheight}{10pt}
\setlength{\headsep}{15pt}
\setlength{\topmargin}{-25pt}
\setlength{\topskip}{0in}
\setlength{\textheight}{8.7in}
\setlength{\footskip}{0.3in}
\setlength{\oddsidemargin}{0.0in}
\setlength{\evensidemargin}{0.0in}
\setlength{\textwidth}{6.5in}

\begin{document}
\begin{center}
\textbf{\large An example document}
\vspace{10mm}
Karl Broman
\end{center}
\vspace{30mm}
\textbf{A section}
Why \LaTeX{}?

- Fine control of document appearance
- Transparency of how that was achieved
- Version control (diff/merge)
- Typesetting equations
- Markdown’s not quite ready, or sufficiently rich
  (but see the R package \texttt{rticles})
simple $\leftrightarrow$ flexible
simple $\longleftrightarrow$ flexible

\centerline{\Large simple $\quad \longleftarrow\rightarrow \quad $\longleftrightarrow \quad \quad \text{flexible}}
Modify your desires to match the defaults.

Focus your compulsive behavior on things that matter.
Stuff I use a lot

\usepackage{palatino}
\usepackage{times}

\setlength{\rightskip}{0pt plus 1fil} % makes ragged right

\newcommand{\LOD}{\text{LOD}}

\usepackage{setspace}
\setstretch{2.0}

\addtocounter{framenumber}{-1}

% make figures S1, S2, ...
\renewcommand{\thefigure}{\textbf{S\arabic{figure}}}
\renewcommand{\figurename}{\textbf{Figure}}

% bigger space between rows in tables
\renewcommand{\arraystretch}{1.5}

% paragraphs not indented but have space between
\setlength{\parskip}{6pt}
\setlength{\parindent}{0pt}
\documentclass[12pt]{article}
\title{An example Rnw document}
\author{Karl Broman}
\begin{document}
\maketitle
<<load_library, echo=FALSE, results="hide">>=
library(broman) # used for myround()
@
<<example_chunk>>=
x <- rnorm(100)
y <- 5*x + rnorm(100)
lm.out <- lm(y ~ x)
plot(x,y)
abline(lm.out$coef)
@

The estimated slope is $\text{Sexpr{myround(lm.out$coef[2], 1)}}$.
\end{document}
The estimated slope is \Sexpr{myround(lm.out$coef[2], 1)}.
5.1.6 Operators with Limits

Sum $\sum$ and integral $\int$ operators are very often decorated with limits. These limits can be entered in LyX by entering them as you would enter a super- or subscript, directly after the symbol. The sum operator will automatically place its “limits” over and under the symbol in displayed formulas, and on the side in inline formulas. Such as $\sum_{i=0}^{\infty} \frac{1}{2^i} = e$, versus $\sum_{i=0}^{\infty} \frac{1}{2^i} = \infty$.

Integral signs, however, will place the limits on the side in both formula types.

All operators with limits will be automatically re-sized when placed in display mode. The placement of the limits can be changed by placing the cursor directly behind the operator and hitting M-err or using the menu Edit ⇒ Math ⇒ Change Limit Type.

Certain other mathematical expressions have this “moving limits” feature as addition, such as $\int_{x}^{y} f(x) dx$.

which will place the $x \to \infty$ underneath the “lim” in display mode. In inline formulas it looks like this: $\lim_{x \to \infty} f(x)$.

Note that the lim-function was entered as the function macro $\lim$. Have a look at section, Ref: sub:Functions for an explanation of function macros.
Also

- Overleaf
- Authorea
Flavors of \LaTeX

- \LaTeX
- pdflatex
- xelatex
- lualatex
Getting help

- Google
- tex.stackexchange.com
- Ask a friend
- Look at others’ documents
- Resign yourself to something less-than-ideal
Figure captions and floats

\begin{figure}
\includegraphics{figure/fig_with_caption}
\caption{Scatterplot of $y$ vs $x$\label{fig:fig_with_caption}}
\end{figure}

```r
<<fig_with_caption, fig.cap="Scatterplot of $y$ vs $x$">>=
x <- rnorm(100)
y <- 5*x + rnorm(100)
lm.out <- lm(y ~ x)
plot(x,y)
abline(lm.out$coef)
@
```

\begin{figure}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
\begin{verbatim}
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
Tables in \LaTeX

\begin{tabular}{rrrrr} \hline
& Estimate & Std. Error & t value & Pr(>|t|) \\ \hline
(Intercept) & 0.04 & 0.11 & 0.4 & 0.69 \\ x & 0.98 & 0.10 & 10.0 & 0.00 \\ \hline
\end{tabular}
```r
<<generate_and_fit>>=
x <- rnorm(100)
y <- x + rnorm(100)
lm.out <- lm(y ~ x)
@

<<table, results="asis">>=
library(xtable)
xtable(lm.out, digits=c(0,2,2,1,2))
@

% a non-floating version
<<table, results="asis">>=
library(xtable)
xtab <- xtable(lm.out, digits=c(0,2,2,1,2))
print(xtab, floating=FALSE)
@
```
Read page proofs carefully

As submitted

\[
\Pr(g_1 = i, g_2 = j) = \begin{cases} 
\frac{1-r}{8(1+6r)} & \text{if } i = j \\
\frac{r}{8(1+6r)} & \text{if } i \neq j
\end{cases}
\]

As printed

\[
\Pr(g_1 = i, g_2 = j) = \begin{cases} 
\frac{1-r}{8(1+6r)} & \text{if } i = j \\
\frac{r}{2(1+6r)} & \text{if } i \neq j
\end{cases}
\]

Table 4 Two-locus haplotype probabilities at generation $f_1$ in the formation of four-way RIL by sibling mating

<table>
<thead>
<tr>
<th>Chr.</th>
<th>Individual</th>
<th>Prototype</th>
<th>No. states</th>
<th>Probability of each</th>
</tr>
</thead>
</table>
| A    | Random     | AA        | 4          | \[
\frac{1}{4} \left( \frac{6r^2 - 7r + 3s}{4(1 + 6r)} + \frac{6r^2 - 7r + 3s}{4(1 + 6r)} + \frac{1 - 2r + s}{4} + \frac{1 - 2r - s}{4} \right)\] |
|      |            | AB        | 4          | \[
\frac{r}{2} \left( \frac{10r^2 - r + \frac{s}{2}}{4(1 + 6r)} + \frac{10r^2 - r + \frac{s}{2}}{4(1 + 6r)} + \frac{1 - 2r + s}{4} + \frac{1 - 2r - s}{4} \right)\] |
|      |            | AC        | 8          | \[
\frac{r}{2} \left( \frac{2r^2 + 3r + s}{4(1 + 6r)} + \frac{2r^2 + 3r + s}{4(1 + 6r)} + \frac{1 - 2r + s}{4} + \frac{1 - 2r - s}{4} \right)\] |
| X    | Female     | AA        | 2          | \[
\frac{1}{2} \left( \frac{2r^2 + 6r^2 - 2r^2 + r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{2r^2 + 6r^2 - 2r^2 + r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
|      |            | AB        | 2          | \[
\frac{r}{2} \left( \frac{2r^2 + 6r^2 - 2r^2 + r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{2r^2 + 6r^2 - 2r^2 + r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
|      |            | AC        | 4          | \[
\frac{r}{2} \left( \frac{9r^2 + 5r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{9r^2 + 5r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
|      |            | CC        | 4          | \[
\frac{1}{2} \left( \frac{9r^2 + 5r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{9r^2 + 5r + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
| X    | Male       | AA        | 2          | \[
\frac{1}{2} \left( \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
|      |            | AB        | 2          | \[
\frac{r}{2} \left( \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
|      |            | AC        | 4          | \[
\frac{r}{2} \left( \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |
|      |            | CC        | 4          | \[
\frac{1}{2} \left( \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{2r^2 + 2r^2 - r^2 + \frac{s}{2}}{4(1 + 5r + \frac{s}{2})} + \frac{1 - r + \frac{s}{2}}{4} + \frac{1 - r - \frac{s}{2}}{4} \right)\] |

$s = \sqrt{4r^2 - 15r + 5}$ and $t = \sqrt{r^2 - 10r + 5}$; the autosomal haplotype probabilities are valid for $r < \frac{1}{2}$.

A number of investigators have developed methods for identifying such sample mix-ups \citep{Westra2011, Schadt2012, Lynch2012, Ekstrom2012}, and a similar approach was applied by \citet{Baggerly2008, Baggerly2009} in their forensic...

\bibliographystyle{genetics}
\renewcommand*{\refname}{\centerline{\normalsize\sffamily\textbf{Literature Cited}}}
\bibliography{samplemixups}

@article{Baggerly2008,
  author = {Baggerly, Keith A. and Coombes, Kevin R.},
  journal = {J. Clin. Oncol.},
  pages = {1186--1187},
  title = {Run batch effects potentially compromise...},
  volume = {26},
  year = {2008} }

% bibliography format
\usepackage[authoryear]{natbib}
\bibpunct{(}{)}{;}{a}{}{,}
Organizing analyses

- Directory for the main analysis project
  ~/Projects/Blah

- Directory for a paper
  ~/Docs/Papers/Blah

- Paper directory may have an analysis directory
  ~/Docs/Papers/Blah/Analysis

- Symbolic links to `.RData` files
  
  ```
  ln -s ~/Projects/Blah/DerivedData/blah.RData .
  ```

- Each part well organized and fully reproducible.

- R Markdown reports documenting different aspects.

- Analysis with the paper may be re-done “properly.”
Make every number reproducible.

<<define_numbers, echo=FALSE>>=
numbers <- c("one", "two", "three", "four", "five", 
           "six", "seven", "eight", "nine", "ten")
cap <- function(vec) paste0(toupper(substr(vec, 1, 1)), 
                      substr(vec, 2, nchar(vec)))
Numbers <- cap(numbers)
n <- sample(1:10, 1)

Then if I want to talk about a number, like \Sexpr{n}, I can
refer to it by name: \Sexpr{numbers[n]}. And I can start a
sentence with it. \Sexpr{Numbers[n]} grasshoppers walked into a
bar\dots

But be careful about singular vs. plural, and so write
\Sexpr{Numbers[n]} grasshopper\Sexpr{ifelse(n>1, "s", ")}}
walked\dots
# simple make file

mypaper.pdf: mypaper.tex Figs/fig1.pdf Figs/fig2.pdf
    pdflatex mypaper

Figs/fig1.pdf: R/fig1.R
    cd R; R CMD BATCH fig1.R fig1.Rout

Figs/fig2.pdf: R/fig2.R
    cd R; R CMD BATCH fig2.R fig2.Rout

\clearpage
\includegraphics{Figs/fig1.pdf}

\clearpage
\includegraphics{Figs/fig2.pdf}
Version Control

- Your manuscript is under version control, right?
Version Control

- Your manuscript is under version control, right?
- Local or private repository for the whole thing
  - including reviewers’ reports and my response
  - PDF of submitted and final manuscript
- Snapshot of the final version as a public repository
  - I don’t really want to show the whole history
With papers led by a collaborator, I’m usually stuck with Word.  
But my analyses and figures are fully reproducible.  
Create an R Markdown document with the detailed results.
Summary

- LaTeX is brilliant for fine control and for equations
- Floating figures and tables can be a pain
- You use KnitR with LaTeX much the same way as you’d used it with Markdown.
- Ensure that every statistic, figure, and table in your paper are fully reproducible.
- Use xtable to make tables.
- Separate out the code for the figures.
- Use version control!