



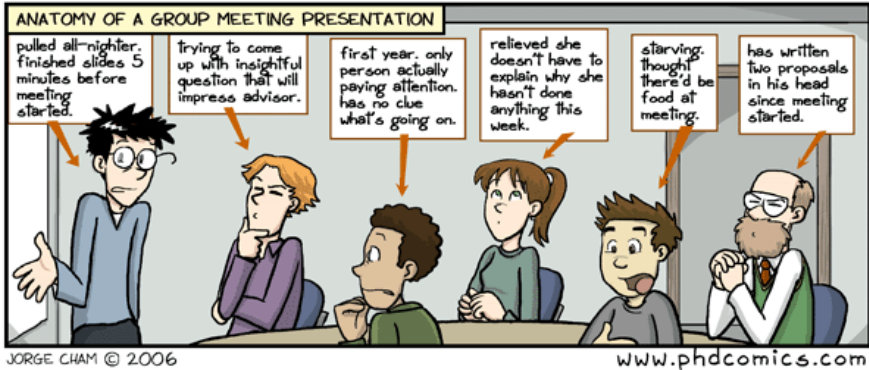
UCI-SeminaR

Sweave: Combining R and \LaTeX

Vinh Q. Nguyen

Department of Statistics
University of California, Irvine

May 7, 2009



“Piled Higher and Deeper” by Jorge Cham
www.phdcomic.com



“Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to humans what we want the computer to do.” -- Donald E. Knuth, 1984



Outline

- 1 Introduction
Motivation
- 2 Sweave
Overview
Examples
Miscellaneous



Outline

1 Introduction Motivation

2 Sweave Overview Examples Miscellaneous



Assumptions

- Familiarity with \LaTeX
 - Write a *.tex file, and compile it into a *.pdf/*.ps file
- Familiarity with R
- Have used or will use results from R in a \LaTeX document
- Some of my examples are illustrated through *nix platform;
Windows follows in parallel



Have you ever...

- Used R to obtain some results (numbers, numbers for a table, plot, etc) for a homework assignment or paper
- Wrote up your assignment or paper in \LaTeX
- Pasted or typed in your results (especially for a table) from R into your \LaTeX document
- Made everything pretty in \LaTeX – document looks super-professional



Then realize....

- You made a mistake in your R code (some parameter was wrong, etc), so the results are all wrong!
- Re-run R code – easy
- Paste/type everything in again and make everything pretty – AGAIN?!?!
- Try this a few times on an assignment/paper and you'll just want to quit!



Other scenarios...

- 1 You publish a paper, and months/years later, an interested reader sends you an email because he/she can't reproduce your results based on how you described it on paper
 - You, however, have moved on to other interesting problems
- 2 You work for a company, and after several months/years, your boss asks you to re-run some analyses on some new data, and produce the same tables and plots
- 3 Or better yet, you have left the company, and someone else has to reproduce these results
- 4 Or even worse, you come to a company and have to reproduce results that someone else did



Solution

$$\text{Sweave} = \text{R} + \text{L}^{\text{A}}\text{T}^{\text{E}}\text{X}$$

- Leisch (2002)
- “Literate programming” (Donald Knuth) for R
 - Programs are useless without descriptions
 - Descriptions should be *literate*, not just comments in code
 - Code in descriptions should work
- A package for “reproducible research”
 - Anything in a scientific paper should be reproducible by the reader!
- Automatic Report Generation
 - In many instances, e.g. the corporate world, reports are “ran” periodically



Introduction
Sweave

Overview
Examples
Miscellaneous



Outline

1 Introduction
Motivation

2 Sweave
Overview
Examples
Miscellaneous



Installing Sweave

- Comes with every R installation in the `utils` package
- By default, loaded into every R session
- Main functions are `Sweave()` and `Stangle()`



Work flow

Start with an *.Rnw/*.Snw file, i.e., a *.tex file with embedded R code, and end up with the results file:





Work flow

- 1 Weaving R/S results into a *.tex file
 - Sweave:
 - R session: `> Sweave(file='foo.Rnw')`
 - *nix shell: `$ R CMD Sweave foo.Rnw`
 - *.tex file will have `\usepackage{Sweave}` added to preamble
 - Make sure `Sweave.sty` is in your `texmf` path (or in your working directory)
 - Can put `export`
`SWEAVE_STYLEPATH_DEFAULT="TRUE"` in your `.bashrc`
for R CMD Sweave to put hardcoded path to `Sweave.sty`
- 2 Compile your *.tex file like usual (e.g., `pdflatex`, `latex`)



Contents of an *.Rnw file

R codes are embedded inside code *chunks* using the *noweb* syntax of Norman Ramsey:

```
\documentclass[a4paper]{article}
\title{Sweave Example 1}
\author{Vinh Nguyen}
\begin{document}
\maketitle
.
.
.
<<>>=
R code
@
.
.
.
\end{document}
```



Introduction
Sweave

Overview
Examples
Miscellaneous



Outline

1 Introduction
Motivation

2 Sweave
Overview
Examples
Miscellaneous



Example 1

```
\documentclass[a4paper]{article}
\title{Sweave Example 1}
\author{Vinh Nguyen}
\begin{document}
\maketitle
Here is a linear regression example:
<<>>=
set.seed(123) ##reproducibility
n <- 100 ; x <- rnorm(n, sd=2)
y <- 1 -.5*x + rnorm(n)
fit <- lm(y ~ x) ; summary(fit)
@
Sweave embeds the results!
\end{document}
```

Run Sweave ('Ex1.Rnw') then `pdflatex Ex1.tex`, and we get:



Example 1

Sweave Example 1

Vinh Nguyen

May 7, 2009

Here is a linear regression example:

```
> set.seed(123)
> n <- 100
> x <- rnorm(n, sd = 2)
> y <- 1 - 0.5 * x + rnorm(n)
> fit <- lm(y ~ x)
> summary(fit)
```

```
Call:
lm(formula = y ~ x)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.9073 -0.6835 -0.0875  0.5806  3.2904
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.89720    0.09755   9.197 6.69e-15 ***
x           -0.52624    0.05344  -9.847 2.60e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.9707 on 98 degrees of freedom
Multiple R-squared: 0.4974,    Adjusted R-squared: 0.4922
F-statistic: 96.97 on 1 and 98 DF,  p-value: 2.595e-16
```

Sweave embeds the results!



Options for the R chunks: `« . . . »=`

- `name`: *user-specified* – Specify name of chunk
- `eval`: **TRUE**, FALSE – Evaluate chunk?
- `echo`: **TRUE**, FALSE – Show S code?
- `term`: **TRUE**, FALSE – Emulates R session? I.e., include whatever is printed to R console
- `results`: **verbatim**, `tex`, `hide` – What are we outputting? Chunks with 'tex' will be treated as latex code
- `fig`: **TRUE**, **FALSE** – Chunk produce graphics? NOTE: Only one plotting device per chunk!
- `include`: **TRUE**, FALSE – Automatically generate `\includegraphics` statement

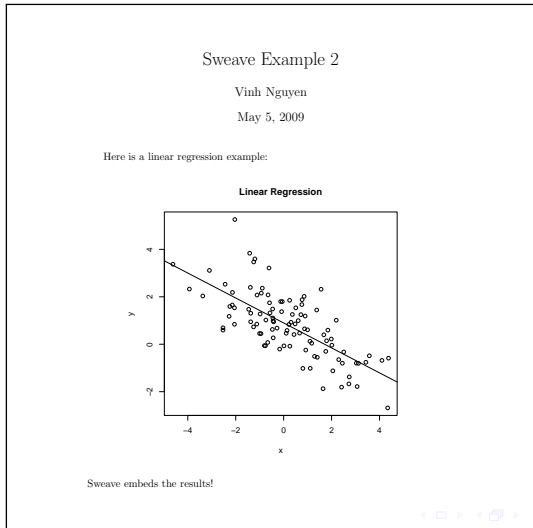


Example 2: Including Plots

```
...  
\begin{center}  
<<LinearRegression, eval=TRUE, echo=FALSE  
  , term=FALSE, fig=TRUE, include=TRUE>>=  
set.seed(123) ##reproducibility  
n <- 100 ; x <- rnorm(n, sd=2)  
y <- 1 -.5*x + rnorm(n)  
fit <- lm(y ~ x) ; summary(fit)  
plot(x,y, main="Linear Regression")  
abline(fit)  
@  
\end{center}  
...
```



Example 2: Plots (`fig=TRUE`, `include=TRUE`)





More on plots

- By default, Sweave produces both a *.pdf and *.eps file when `fig=TRUE`, unless we set `pdf/eps=FALSE`
- Naming convention: `RnwFilename-NameOfChunk.pdf/eps`
- The regression plot produced in `Ex2.Rnw` will be named `Ex2-LinearRegression.pdf/eps`
- `Height` and `Width` options in the chunk are for the size of the output file, not how it appears in our *.tex file
- To change size manually, do a manual `\includegraphics`



More on plots

- We can manually input include the graphics by creating the plot (`fig=TRUE`) but not letting Sweave include it (`include=FALSE`)
- For example:

```
...  
\begin{center}  
<<LinearRegression, eval=TRUE, echo=FALSE  
  , term=FALSE, fig=TRUE, include=FALSE>>=  
Plotting code  
@  
\includegraphics[width=3in,height=3in]  
  {Ex2-LinearRegression}  
\end{center}  
...
```



Multiple Plots in One Chunk

- Currently, Sweave only allows one graphic device per chunk
- If you need to create multiple graphic files from one chunk, say in a loop, use `postscript()` and `pdf()` to create the files
- Include them manually or in your output (`tex=TRUE`)

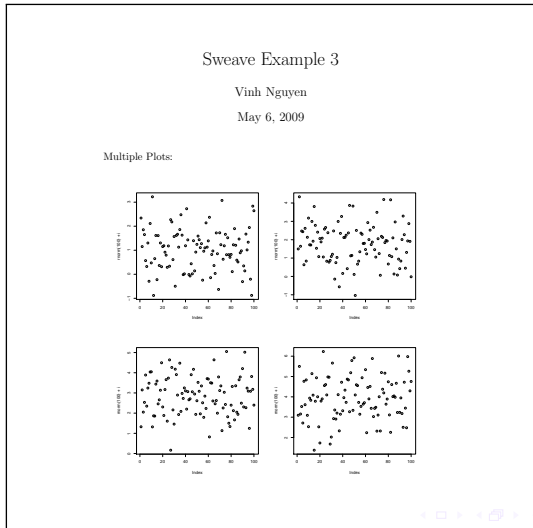


Example 3: Generating Multiple Plots in One Chunk

```
\begin{center}
<<results=tex, echo=FALSE>>=
set.seed(672)
for(i in 1:4){
  file=paste("myfile", i, ".pdf", sep="")
  pdf(file=file, paper="special", width=6, height=6)
  plot(rnorm(100)+i)
  dev.off()
  cat("\\includegraphics[height=2in, width=2in]{",
      , file, "}\n", sep="")
}
@
\end{center}
```



Example 3: Generating Multiple Plots in One Chunk





Re-using Named Chunks

```
<<a>>=  
x <- 10  
@  
  
<<b>>=  
x+y  
@  
  
<<c>>=  
<<a>>  
y <- 20  
<<b>>  
@
```

This is a simple example; I've done simulations where I've used this method to change parameters and re-run simulations.



Scalar Results and `tex` Results

- To get scalar results in your text, use `\Sexpr{Rscalar}`
- `xtable` package will return \LaTeX code of various R objects (matrix, dataframes, lm summary objects, etc)
- In scientific papers, tables tend to be big and complicated (not a typical R object output)
- I typically write a function to return a row for the table (using `paste()` to add in `&`)
- To get the table, I output each row through a loop in an R chunk

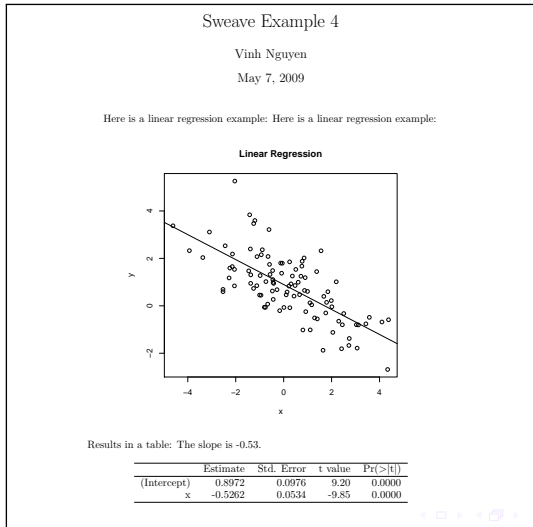


Example 4: Scalar Results and `tex` Results

```
...
Results in a Table:
<<regresults, echo=FALSE, term=TRUE, results=tex>>=
library(xtable)
xtable(summary(fit))
@
The slope is  $\text{\Sexpr{round(fit$coef[2], 2)}}$ .
\end{document}
```



Example 4: Scalar Results and `tex` Results





Introduction
Sweave

Overview
Examples
Miscellaneous



Outline

1 Introduction
Motivation

2 Sweave
Overview
Examples
Miscellaneous



Misc.

- Use `Stangle()` to extract chunks from `foo.Rnw` to `foo.R`
- Use `Sweave('foo.Rnw', debug=TRUE)` to have all inputs and outputs printed in `console`
- Use `\Sweaveinput{AnotherFile.Rnw}` to “include” the file in the `Rnw` file – analogous to \LaTeX 's `\input` and `\include`
- To keep comments in code, use `\SweaveOpts{keep.source=TRUE}` in `Rnw` preamble
- Also, use `\SweaveOpts{}` to change default options of `« . . . »=` for the entire file



Misc.

- Large simulation study or costly computations?
 - Leave `eval=FALSE` when polishing Rnw file
 - Run it in a separate R script, save the workspace or objects, and load them into the chunks (**not** everything is in a single file anymore – it will suffice until the final run)
 - Check out `cacheSweave` package



cacheSweave

- Install `cacheSweave` package from CRAN
- Add `cache=TRUE` to computationally expensive chunks
- In an R session:

```
Sweave('foo.Rnw',  
driver=cacheSweaveDriver())
```
- After the first run, objects from the cached chunks will be stored in a hash table
- Subsequent Sweaves will not evaluate these chunks
- Cached chunks should be computations only – NO FIGURES

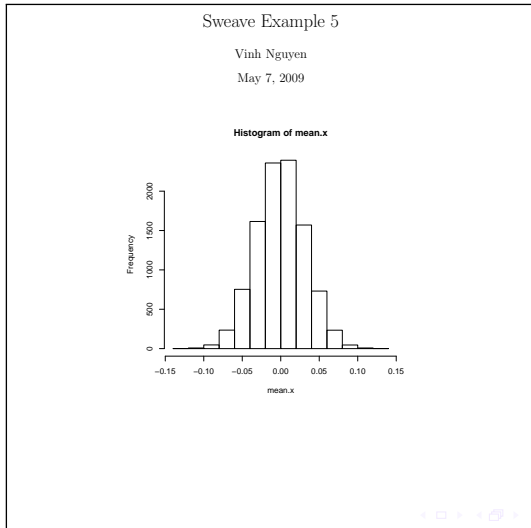


Example 5: cacheSweave

```
<<cacheSweave, eval=TRUE, echo=FALSE, term=FALSE  
  , cache=TRUE>>=  
library(cacheSweave)  
set.seed(23)  
mean.x <- numeric(10000)  
for(i in 1:10000){  
  x <- rnorm(1000)  
  mean.x[i] <- mean(x)  
}  
@  
<<plot, echo=FALSE, term=FALSE, fig=TRUE>>=  
hist(mean.x)  
@
```



Example 5: cacheSweave





Recommendation: Emacs

- I personally use Emacs for all my scientific typing and computing (Linux, Mac, Windows)
- AucTeX – great for editing \LaTeX files
- ESS: Emacs Speaks Statistics
 - Run R within Emacs, and send R code to it
 - ESS automatically recognize Rnw files, distinguishing between \LaTeX code and R code (can send to R)
 - When an *.Rnw file is open with an inferior R session running:
 - M-n s to Sweave file
 - M-n l to run latex
 - M-n p to make and display postscript file
 - M-n P to make and display pdf file



What if I don't use \LaTeX ?

- Check out `odfWeave` (Max Kuhn) for weaving R results into the Open Document Format files (XML-based)
 - Very similar to Sweave (based on `noweb` syntax)
 - Plots will be saved as `*.png` files
 - Tables will be saved in `xml`
 - We can now edit the final results with our choice of any compatible WYSIWYG word-processor, e.g., OpenOffice
- `html`?
 - Export R results into `html`
 - Does NOT use the `noweb` syntax



References I

- 1 <http://www.statistik.lmu.de/~leisch/Sweave/>
- 2 <http://www.statistik.lmu.de/~leisch/Sweave/Sweave-manual.pdf>
- 3 <http://www.stat.umn.edu/~charlie/Sweave/>
- 4 http://www.bias-project.org.uk/Rpackages_course/practical_Sweave.pdf
- 5 <http://www.johndcook.com/talkb.pdf>
- 6 <http://www.stat.auckland.ac.nz/~stat782/downloads/Sweave-customisation.pdf>
- 7 <http://cran.r-project.org/web/packages/cacheSweave/vignettes/cacheSweave>



References II

- 8 <http://cran.r-project.org/web/packages/cacheSweave/vignettes/cacheSweave>