Outline
Introduction
Examples
To Do
Summary
Bibliography

"Developing" a PhD-Thesis in Perl, R and LATEX - and all the code involved -

Emanuel Heitlinger

November 19, 2009



Introduction

Noweb/Sweave Editing a .Rnw file

Examples

An example of R-sweave in a LATEX beamer class An example involving perl A Next Gen example Caching data

To Do

Summary

Bibliography



R, Noweb/Sweave and Emacs-ESS

▶ Donald E. Knuth[1]: "The main idea is to treat a program as a piece of literature, addressed to human beings rather than to a computer"

R, Noweb/Sweave and Emacs-ESS

- ▶ Donald E. Knuth[1]: "The main idea is to treat a program as a piece of literature, addressed to human beings rather than to a computer"
- ► Friedrich Leisch: Integration into R -> Sweave[2]

R, Noweb/Sweave and Emacs-ESS

- ▶ Donald E. Knuth[1]: "The main idea is to treat a program as a piece of literature, addressed to human beings rather than to a computer"
- ► Friedrich Leisch: Integration into R -> Sweave[2]
- ► Emacs-ESS[3] is a very powerful editor for R-scripts and the .Rnw files of Sweave

Editing a .Rnw (sweave) file in Emacs

```
version 2.9.2 (2009-08-24)
             \frametitle(Basic concepts)
                                                                                                                                                                                                                                                                                                                                                                                                                                                          K version 2.9.2 [2009-00-24]
Copyright [C] 2009 The R Foundation for Statistical Computing
             \testing it has core code is real \text{\testing item continuous to a fine continuous to the source code is real \text{\testing item continuous to redistribute it under certain conditions.} \text{\text{\testing item continuous to redistribute it under certain conditions.} \text{\text{\testing item continuous to redistribute it under certain conditions.} \text{\text{\text{\testing item continuous to redistribute it under certain conditions.}} \text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\text{\tex
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 Natural language support but running in an English locale
      (subsection(What Emacs-ESS does for you)
   \frame {
\frame!\text{\text{frame}} \text{\text{frame}} \text{\tex
                                                                                                                                                                                                                                                                                                                                                                                                                                                       R is a collaborative project with many contributors
                                                                                                                                                                                                                                                                                                                                                                                                                                                          Type "contributors()" for more information and
                                                                                                                                                                                                                                                                                                                                                                                                                                                              'citation()' on how to cite R or R packages in publications.
                \itemsl-> Send code from the script to the session
             \item=2-> View output in the session
\item<3-> All in one system
                                                                                                                                                                                                                                                                                                                                                                                                                                                       Type 'demo()' for some demos, 'help()' for on-line help, or 
'help.start()' for an HTML browser interface to help. 
Type 'qi)' to quit R.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 options(STERM='iESS', editor='emacsclient')
   \section(Examples)
\unkertion(Editing a .Rnw file)
\frame {
\frame itle {Editing a .Rnw (sweave) file in emacs (directly in terminal -n\)
                                                                                                                                                                                                                                                                                                                                                                                                                                                              y <- x + rnorm(100,0,1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                              - X = 10109

- y = X + roorm[100,0,1]

- summary(x)

Mim. 1st Ou. Median Mean 3rd Ou. Max.

1.00 25.75 50.50 50.50 75.25 100.00
   w outlant)
                \includegraphics(beamer-ess.png)
      \subsection(An example of R-sweave in a LaTeX beamer class)
   \frame [containsverbatim]{
\frame[containsverbatim]{
\frame(frametitle{R code printed on on slides})
   x <- 1:100
y -- x + rnorm(160,0,1)
d tidef
Reads in the code
             Aframetatle(The resulting figure)
      Print the plot seemingly direct in the slide
         section(Summary)
   \frame {
\frametitle{Summary}
   | Megin(Intender) | Megin(Inte
```

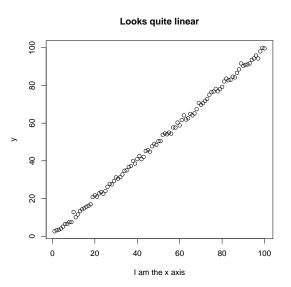


R code printed on on slides

Reads in the code



The resulting figure



```
> S <- as.data.frame(read.delim(pipe("./pilot.pl"),
+ sep = ",", header = FALSE, as.is = TRUE))</pre>
```

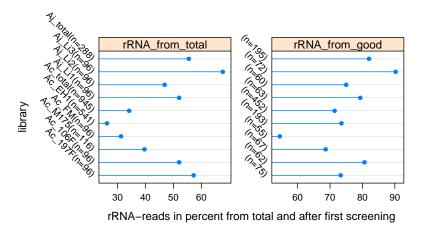
This R function is calling a "custom for this analysis" perl-script and stores the output in a R-object. (The pilot.pl scipt calls istself multi purpose perl-scripts and blastall)

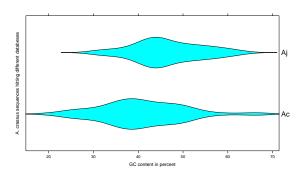
After some R-code (not included here but in the source-file) we can do

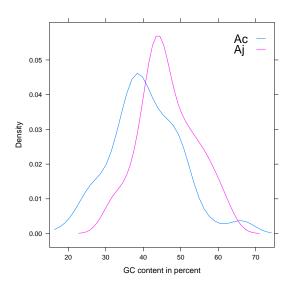
```
> xtable(An, caption = "Screening statistics",
     display = c("s", rep("d", times = 5)),
     label = "tab:num")
+
```

•					
	short	poly	rRNA	fishpep	good2
Ac_197F(n=96)	4	17	55	4	16
$Ac_106F(n=96)$	25	9	50	2	10
Ac_M175(n=116)	30	19	46	2	18
$Ac_FM(n=96)$	12	29	30	5	20
Ac_EH1(n=541)	297	51	142	15	36
$Ac_{total}(n=945)$	368	125	323	28	100
Aj_Li1(n=96)	10	23	50		13
Aj_Li2(n=96)	10	26	45		15
Aj_Li3(n=96)	9	15	65		6
Aj_total(n=288)	29	64	160		34





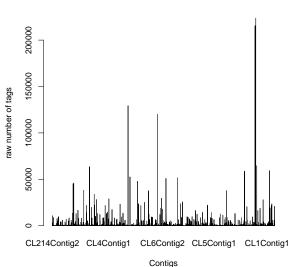




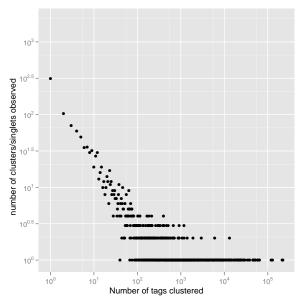
 Much more data, but everything less complicated because R-bioconductor instead of perl is used

- ► Much more data, but everything less complicated because R-bioconductor instead of perl is used
- ▶ I just read in 6201930 tags from a tag-sequencing experiment. 3494662 are mapped to 4077 tgicl- cotigs from 454 using maq. 2836611 to the plus strand 658051 to the minus strand

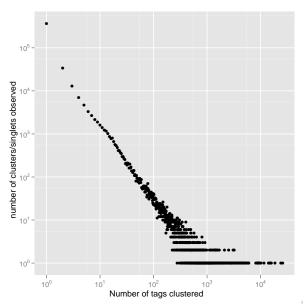
- Much more data, but everything less complicated because R-bioconductor instead of perl is used
- ▶ I just read in 6201930 tags from a tag-sequencing experiment. 3494662 are mapped to 4077 tgicl- cotigs from 454 using maq. 2836611 to the plus strand 658051 to the minus strand
- ▶ All numbers in this slide (expect 454;)) are literally coming from the programm! This is possible because one can include not only figures and tables but als short R-expressions compiling to numbers in text.



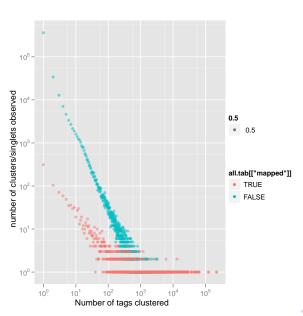
Of the 4077 contigs 2123 had at least one tag mapping 1954 had no tag.



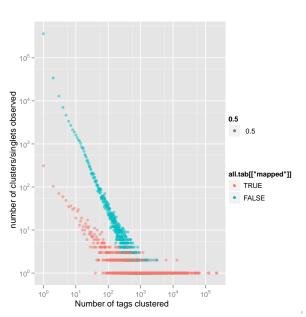
Of the 4077 contigs 2123 had at least one tag mapping 1954 had no tag.



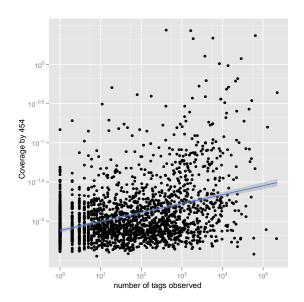
From the 2707268 sequences without hits in the 454 contigs 361468 were singletons. The remaining formed 91470 clusters.



► Of the 4077 contigs 2123 had at least one tag mapping 1954 had no tag.



- ➤ Of the 4077 contigs 2123 had at least one tag mapping 1954 had no tag.
- ► From the 2707268 sequences without hits in the 454 contigs 361468 were singletons. The remaining formed 91470 clusters.



Caching data

Caching[4] of computations allows R-objects to be stored "on disk". If the object does not change it does not have to be computed again.

▶ Minor issues with syntax highlighting etc...

- ▶ Minor issues with syntax highlighting etc...
- ► Investigate ways to work on client but execute on server (would allow the editor to be used in X-window)

- ▶ Minor issues with syntax highlighting etc...
- Investigate ways to work on client but execute on server (would allow the editor to be used in X-window)
- Caching of R-objects produced by perl-scripts must be made aware of changes in perl-scripts

- ▶ Minor issues with syntax highlighting etc...
- Investigate ways to work on client but execute on server (would allow the editor to be used in X-window)
- Caching of R-objects produced by perl-scripts must be made aware of changes in perl-scripts

- ▶ Minor issues with syntax highlighting etc...
- Investigate ways to work on client but execute on server (would allow the editor to be used in X-window)
- ► Caching of R-objects produced by perl-scripts must be made aware of changes in perl-scripts
- Investigate sharing of Cache (between machines and files) -Maybe not even desirable

► Literate programming using R sweave can be used to clue togther code- and markup- language

- ► Literate programming using R sweave can be used to clue togther code- and markup- language
- ▶ When the system is used as a general working environment it needs some commitment

- ► Literate programming using R sweave can be used to clue togther code- and markup- language
- ► When the system is used as a general working environment it needs some commitment
- ► Caching of computations opens the way for large projects, where R code executes other time consuming scripts

- ► Literate programming using R sweave can be used to clue togther code- and markup- language
- ► When the system is used as a general working environment it needs some commitment
- ► Caching of computations opens the way for large projects, where R code executes other time consuming scripts

- ► Literate programming using R sweave can be used to clue togther code- and markup- language
- ▶ When the system is used as a general working environment it needs some commitment
- ► Caching of computations opens the way for large projects, where R code executes other time consuming scripts
- Caching in combination with a VCS may make possible a whole PhD-thesis using Next-Gen being written in Noweb/Sweave?



Knuth DE: Literate Programming. Stanford, California: Center for the Study of Language and Information 1992,
[http://www-cs-faculty.stanford.edu/~knuth/lp.html].



Leisch F: Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis. In Compstat 2002 — Proceedings in Computational Statistics. Edited by Härdle W, Rönz B, Physica Verlag, Heidelberg 2002:575–580, [http://www.stat.uni-muenchen.de/~leisch/Sweave]. [ISBN 3-7908-1517-9].



Rossini A, Heiberger R, Sparapani R, Maechler M, Hornik K: Emacs Speaks Statistics: A Multiplatform, Multipackage Development Environment for Statistical Analysis. Journal of Computational and Graphical Statistics 2004, 13:247–261, [http://ess.r-project.org/index.php?Section=home].



Falcon S: Caching code chunks in dynamic documents. *Computational Statistics* 2009, **24**(2):255–261,

[http://www.springerlink.com/content/55411257n1473414]