

BioC for HTS - PDCB topic Bioconductor 02

LCG Leonardo Collado Torres
lcollado@wintergenomics.com – lcollado@ibt.unam.mx

September 22nd, 2010

Previous Exercise

Install weaver

L^AT_EX

Sweave

Stangle

weaver package

Exercise

- ▶ Modify the `randomSeq` and `calcGC` functions to include parameter checks using conditionals (`if, ...`), and functions such as `stop` and `warning`
- ▶ Test the functions for cases with Ns and/or Us. Do they affect the distribution of GC?
- ▶ Create again the package `GCcalc` and add simple descriptions to all the help files.
- ▶ Try building and checking the package (`R CMD build` and `R CMD check`)¹. Do they work?

¹Only available through command line

Installing weaver

- ▶ I missed this package last class, luckily it's small :)
 - > `source("http://bioconductor.org/biocLite.R")`
 - > `biocLite("weaver")`

Why learn it?

Yes, L^AT_EX has a steep learning curve, but

- ▶ it's easy to include R code and the result
- ▶ has a lot of functionality to typeset formulas
- ▶ you don't have to worry about the format, only about the content :)
- ▶ L^AT_EX files are simple text files (no encoding), therefore it's easy to use a control version system
- ▶ it's easy to include bibliography
- ▶ it's easy to link different parts of your file
- ▶ it's free

Setup

Installing L^AT_EX is one of the hard parts.

- ▶ Windows: install Miktex <http://miktex.org/2.8/setup>
- ▶ Linux: install texlive <http://tug.org/texlive/> by using:
sudo apt-get install texlive
sudo apt-get install texlive-base
sudo apt-get install texlive-common
- ▶ Mac: install MacT_EX: <http://tug.org/mactex/>

L^AT_EXeditors

Once you have the basic T_EXutilities you might need an editor

- ▶ Windows: Emacs and/or WinEdt²
`http://vgoulet.act.ulaval.ca/en/ressources/emacs/windows`
`http://www.winedt.com/winedt.html`
- ▶ Linux: Emacs
`sudo apt-get install emacs`
`sudo apt-get install ess`
- ▶ Mac: Emacs and/or Aquamacs
`http://vgoulet.act.ulaval.ca/en/ressources/emacs/mac`
`http://aquamacs.org/`

²There is a trial period after which you can choose to buy it.

A little about Emacs

- ▶ It's useful for programming in R because in 1 windows you can have your code and in the other one your R session. Meaning that you can keep your code clean easily :)
- ▶ It highlights the basic L^AT_EX commands
- ▶ Some versions of Emacs also integrate other L^AT_EX commands
- ▶ Check the *Emacs reference card* for more info on the keyboard shortcuts.

Our First \LaTeX file

Now that we have \LaTeX installed, lets start using it :)

- ▶ Create a file *test.tex*
- ▶ Copy the following text into it:

```
\documentclass{article}  
\begin{document}  
Test  
\end{document}
```

- ▶ Now lets create the PDF file *test.pdf*. To do so in your terminal (or command prompt) type:
`pdflatex test.tex`

Our First L^AT_EXfile

- ▶ Now we can open the file using your default PDF viewer:
acroread test.pdf
open test.pdf³
- ▶ What does it contain?

³For Windows users

Basic T_EX

- ▶ That was just a basic T_EXfile in the article format.
- ▶ L^AT_EXfiles are ASCII files. The only non-visible character that might be included is the end of line.
- ▶ To *control* the text (and figures) in a file, we use *commands* which are preceded by a backslash.

Our second file

- ▶ Open the **latsamp0.txt** file
- ▶ Which commands do you identify?
- ▶ Do spaces matter?
- ▶ As it says there, once you create a couple of L^AT_EX files you'll get used to the commands :)
- ▶ How does the PDF look?

3rd file

- ▶ Now that we have some basic notions of L^AT_EX files, are there any special characters?
- ▶ The full answer is in the **latsamp1.txt** file.
- ▶ What do the new commands do?

Math in L^AT_EX

- ▶ We already included a fraction in one of our files. How do we include an equation?
- ▶ There are two main commands to do so. **latsamp2.txt** will teach us how to do it :)
- ▶ Which new commands enable us to link parts different parts of our files?
- ▶ Typesetting math formulas can be very useful :)

L^AT_EX Arrays

- ▶ Take a look at **latsamp3.txt**.
- ▶ What is a L^AT_EX array?
- ▶ Is it useful for math formulas?

More math in L^AT_EX

- ▶ To end the series of examples by Matchett, lets look at **latsamp4.txt**
- ▶ What new goodies are we adding to this document?
- ▶ What is an eqnarray?
- ▶ All these examples are available at <http://www.uwlax.edu/faculty/matchett/late/late.htm>
- ▶ Check <http://www.wikihow.com/Typeset-Formulas-in-LaTeX>
How would you add math equations next to text?
- ▶ Later on you might want to review the basic steps with <http://www.wikihow.com/Use-LaTeX-for-Text-Formatting>

A much longer file

- ▶ Open *syllabus.tex* and *syllabus.Rnw*
- ▶ What are the differences between the two files?
- ▶ Which new commands did I use in those files?
- ▶ How do we add the list of topics in a L^AT_EX file?
- ▶ What is the difference between the PDF derived from *syllabus.tex* and the PDF file in the course site?

In summary

We now know all basic parts related to \LaTeX files. Of course, if you need some help check

- ▶ The Not so Short Guide to \LaTeX
<http://tobi.oetiker.ch/lshort/lshort.pdf>
- ▶ The Comprehensive \LaTeX Symbols List
<http://www.math.toronto.edu/mathit/symbols-letter.pdf>
- ▶ The Beamer User Guide
http://faq.ktug.or.kr/wiki/uploads/beamer_guide.pdf
- ▶ A *small* list of \LaTeX symbols
<http://www.artofproblemsolving.com/Wiki/index.php/LaTeX:Symbols>

In summary

- ▶ A guide on how to insert figures in \LaTeX
<http://www.hep.manchester.ac.uk/u/jenny/jcwdocs/latex/figures.html>
- ▶ Adding accents in \LaTeX (change francais to spanish)
<http://fontignie.blogspot.com/2006/04/accents-in-latex.html>
- ▶ A guide on how to use Bib \TeX
<http://amath.colorado.edu/documentation/LaTeX/reference/faq/bibstyles.html>

Sweave origins

- ▶ Now that we know the basics of \LaTeX files, let's start to use Sweave
- ▶ Check the help:
> ``?` (Sweave)`
- ▶ *The basic idea is to replace the S code with its output, such that the final document only contains the text and the output of the statistical analysis.*
- ▶ Sweave comes from the S days. Remember that R is an implementation of S.

R in L^AT_EX

So, how do we add R code?

- ▶ We use *code chunks*
- ▶ For each code chunk we can specify several tags: do we want to print the result? do we want to evaluate the code chunk? add a figure?
- ▶ Lets make a simple test file called **test.Rnw**

```
\documentclass{article}  
\begin{document}  
\begin{Schunk}  
\begin{Sinput}  
> rnorm(5)  
\end{Sinput}  
\begin{Soutput}
```

R in L^AT_EX

```
[1] 1.126848e-01 4.664422e-05
[3] 1.179701e+00 -5.364950e-01
[5] 1.378251e+00
\end{Soutput}
\end{Schunk}
\end{document}
```

- ▶ From R we can run Sweave like this:
`> Sweave(file = "test.Rnw")`
- ▶ Now we have a file called **text.tex** with the following info⁴:

R in L^AT_EX

```
\documentclass{article}
\usepackage{Sweave}
\begin{document}
\begin{Schunk}
\begin{Sinput}
> rnorm(5)
\end{Sinput}
\begin{Soutput}
[1]  1.6319330  -0.2694958  -0.8268343   1.0493980
\end{Soutput}
\end{Schunk}
\end{document}
```

R in L^AT_EX

- ▶ Next, from outside R use `pdflatex` and open it:
`pdflatex test.tex`
`acroread test.pdf`

⁴The actual content will vary because we are getting 5 random numbers!!!

Alternatives

- ▶ Besides opening R, we can also use Sweave like this:

```
R CMD Sweave test.Rnw  
pdflatex test.tex
```

- ▶ Or like this:

```
R -e "Sweave('test.Rnw')"  
pdflatex test.tex
```

- ▶ The last way is specially useful in Windows (R will be R.exe and you need to add it to your PATH by modifying your environment variables). More info at http://www.johndcook.com/troubleshooting_sweave.html

Sweave.sty

- ▶ Normally pdflatex will complain that it cannot find the file **Sweave.sty**.
- ▶ What is the solution? The simple one is to copy that file to your current working directory. It might be a **pain** but it's easy to do.

Sweave Options

Note that in Rnw files we can add the following \LaTeX command:

It will set the default options for our code chunks. It's specially useful if you want to save the figures in a special directory and/or add a given prefix.

Adding R plots

How do we do it? Simple, we use the flag **fig=TRUE**.

Create a new directory with the file *test2.Rnw* and create the PDF file. What do you notice?

```
\documentclass{article}
```

```
\begin{document}
```

We create x

```
\begin{Schunk}
```

```
\begin{Sinuput}
```

```
> x <- rnorm(5)
```

```
\end{Sinuput}
```

```
\end{Schunk}
```

We now plot x:

```
\medskip
```

Adding R plots

```
\includegraphics{plots/fig-007}  
\end{document}
```

GenomicRanges vignette

- ▶ Now that we can add R code into our \LaTeX files, lets reproduce a BIG vignette.
- ▶ Download <https://hedgehog.fhcrc.org/bioconductor/trunk/madman/Rpacks/GenomicRanges/inst/doc/GenomicRangesIntroduction.Rnw>
- ▶ Using Sweave and pdflatex create the PDF file.
- ▶ Did something unexpected happen? If so, what was it?

What do we use it for?

- ▶ Simple, this R function extracts all the R code chunks from a given file. For example:

```
> Stangle(file = "test2.Rnw")
```

- ▶ produces the file **test2.R**:

```
#####
```

```
### chunk number 1:
```

```
#####
```

```
#line 5 "test2.Rnw"
```

```
x <- rnorm(5)
```

```
#####
```

```
### chunk number 2:
```

What do we use it for?

```
#####  
#line 10 "test2.Rnw"  
barplot(x)
```


Overview

- ▶ The weaver Bioconductor package is a rather easy to use infrastructure package.
- ▶ It's useful in case parts of your R code take a lot of evaluation time.
- ▶ Lets look at the help and the vignette:
 - > `library(weaver)`
 - > `help(package = weaver)`
 - > `browseVignettes("weaver")`

Wrap up

- ▶ As they emphasize, you should normally only have ONE Rnw file per directory to avoid any confusion. This is also true for T_EXfiles!
- ▶ L^AT_EX itself is VERY useful if you are writing a long document (30 or more pages). In conjunction with weaver now you have nothing to worry about with long files.
- ▶ However, please do not use weaver for package vignettes! Why? Because the package has to build in less than 5 minutes and you should avoid slow computations.
- ▶ To enhance reproducibility avoid using the *EVAL=FALSE* tag in code chunks!
- ▶ Finally, take a look at the files *sw.bat* and *toPDF.sh*. They might be useful for you as well :)

Exercises

- ▶ Complete the doc1 and doc2 weaver examples.
- ▶ Create short PDF file using Sweave where you'll include a small abstract, a table of contents, some calculations in section 1, a plot in section 2.

Session Information

```
> sessionInfo()
```

```
R version 2.12.0 Under development (unstable) (2010-09-08 r52880)  
Platform: x86_64-unknown-linux-gnu (64-bit)
```

```
locale:
```

```
[1] LC_CTYPE=en_US.utf8  
[2] LC_NUMERIC=C  
[3] LC_TIME=en_US.utf8  
[4] LC_COLLATE=en_US.utf8  
[5] LC_MONETARY=C  
[6] LC_MESSAGES=en_US.utf8  
[7] LC_PAPER=en_US.utf8  
[8] LC_NAME=C  
[9] LC_ADDRESS=C  
[10] LC_TELEPHONE=C  
[11] LC_MEASUREMENT=en_US.utf8  
[12] LC_IDENTIFICATION=C
```

```
attached base packages:
```

Session Information

```
[1] tools      stats      graphics  
[4] grDevices  utils      datasets  
[7] methods    base
```

other attached packages:

```
[1] weaver_1.15.0  codetools_0.2-2  
[3] digest_0.4.2
```