

# R + L<sup>A</sup>T<sub>E</sub>X = Sweave

A. Blejec

andrej.blejec@nib.si

Nacionalni inštitut za biologijo

19. maj 2008

# "Klasična" priprava poročil

- priprava podatkov
- analiza
- priprava tabel
- priprava slik
- priprava besedila
- tabele in slike "nalepimo" v besedilo ...

Slike, tabele in drugi rezultati zgubijo povezavo z besedilom, izvorne slike in tabele ter besedilo so razpršene na več koncih.

# "Klasično" popravljanje poročil

Skušamo se spomniti, kje je kaj ...

- morda popravimo podatke
- ponovna analiza
- nove tabele
- nove slike
- stare slike zberemo iz besedila in vstavimo nove

# Priprava poročil s Sweave

bla.Rnw  $\rightarrow$  Sweave (R )  $\rightarrow$  bla.tex  $\rightarrow$  L<sup>A</sup>T<sub>E</sub>X  $\rightarrow$  bla.pdf

- Dokumentacija in program v eni datoteki
- Vse povezave so vključene v tej datoteki
- Ob spremembi ali potrebi po ponovni analizi, cikel ponovimo

Pridobimo preglednost in ponovljivost analize ('reproducible report')

# Ideja WEB in noweb

## Ideja 'literate' programiranja (D. Knuth)

Literate programming is the art of preparing programs for human readers.

V eni datoteki so prepletene

- informacije namenjene računalniku  
(ukazi, ki omogočajo izvedbo)
- informacije namenjene človeku  
(dokumentacija, ki omogoča razumevanje)

Vse na enem mestu!

<http://en.wikipedia.org/wiki/WEB> in

<http://en.wikipedia.org/wiki/Noweb>

# Struktura vhodne datoteke: P01.Rnw ...

```
\documentclass{article}
  \usepackage[cp1250]{inputenc}
  \oddsidemargin -0.5in \topmargin -0.75in
\usepackage{Sweave}
\begin{document}\large{

%% Besedilo in ukazi v eni datoteki
Izračunajmo vrednost izraza:

<<>>=
#R ukazi
# 2 pomnožimo z dva, če ste pripravljeni razumeti šalo
2*2
@

Rezultat računa ni pretirano presenetljiv.}
\end{document}
```

## ... in vsebina končne datoteke P01.pdf

Izračunajmo vrednost izraza:

```
> 2 * 2
```

```
[1] 4
```

Rezultat računa ni pretirano presenetljiv.

# Sweave (F. Leisch, 2002)

## What is Sweave?

Sweave is a tool that allows to embed the R code for complete data analyses in latex documents.

The purpose is to create dynamic reports, which can be updated automatically if data or analysis change.

Instead of inserting a prefabricated graph or table into the report, the master document contains the R code necessary to obtain it.

When run through R, all data analysis output (tables, graphs, etc.) is created on the fly and inserted into a final latex document.

The report can be automatically updated if data or analysis change, which allows for truly reproducible research.

<http://www.statistik.lmu.de/~leisch/Sweave/>



# Dokumentacijski in programski koščki

Dokument je sestavljen iz koščkov (chunk):

## Dokumentacijski košček

## Documentation chunk

Dokumentacijski košček se začne z vrstico, ki ima na prvem mestu znak 'at' (@), za njim pa presledek ali konec vrstice. Preostanek vrstice ni pomemben in je lahko komentar

Dokumentacijski košček je besedilo poročila, pripravljeno v skladu z  $\text{\LaTeX}$ (ali HTML).

## Programski košček

## Code chunk

Programski košček se začne z `<<ime>>=` na začetku vrstice. Tudi tu je preostanek vrstice vzet kot komentar

Programski koščki so sestavljeni iz R ukazov za izvedbo analize

# Struktura vhodne datoteke: P01.Rnw ...

```
\documentclass{article}
  \usepackage[cp1250]{inputenc}
  \oddsidemargin -0.5in \topmargin -0.75in
\usepackage{Sweave}
\begin{document}\large{

%% Besedilo in ukazi v eni datoteki
Izračunajmo vrednost izraza:

<<>>=
#R ukazi
# 2 pomnožimo z dva, če ste pripravljeni razumeti šalo
2*2
@

Rezultat računa ni pretirano presenetljiv.}
\end{document}
```

# Preoblikovanje vhodne datoteke

Vhodno datoteko lahko preoblikujemo z dvema operacijama

`tangle` (Stangle)

zavozlamo

izluščimo le programske koščke - namenjeno računalniku  
v našem primeru so to R ukazi  
⇒ datoteka \*.R

`weave` (Sweave)

stkemo

združimo dokumentacijske koščke in posebej pripravljene rezultate  
programskih koščkov v datoteko, ki je primerna za oblikovanje -  
namenjeno bralcu  
v našem primeru je to datoteka primerna za  $\text{\LaTeX}$  ⇒ \*.tex

Datoteka programskih koščkov

```
#####  
### chunk number 1: prvi primer  
#####  
#R ukazi  
# 2 pomnožimo z dva, če ste pripravljeni razumeti šalo  
2*2
```

## Sweave

P01.Rnw → P01.tex

Datoteka z dokumentacijo in rezultati

```
\documentclass{article}
  \usepackage[cp1250]{inputenc}
  \oddsidemargin -0.5in \topmargin -0.75in
\usepackage{Sweave}
\begin{document}\large{

%% Besedilo in ukazi v eni datoteki
Izračunajmo vrednost izraza:

\begin{Schunk}
\begin{Sinput}
> 2 * 2
\end{Sinput}
\begin{Soutput}
[1] 4
```

# Nadzor nad programskimi koščki

Programskim koščkom lahko predpišemo oblike delovanja (options)

`echo` izpis programskega koščka

`results` način izpisa rezultatov

`print` izpis vmesnih rezultatov

`eval` odloženo izvajanje

`fig` priprava slike (ena slika na košček)

`prefix.string` predpona imenom slik

`eps, pdf` format slike

`width` širina slike

`height` višina slike

# Option:

Vhodna .Rnw datoteka

```
<<>>=  
x <- c(1,2,2,4,3,6,5,3,2)  
length(x)  
table(x)  
@
```

Izhodna .pdf datoteka

```
> x <- c(1, 2, 2, 4, 3, 6, 5  
> length(x)  
[1] 9  
> table(x)  
  
x  
1 2 3 4 5 6  
1 3 2 1 1 1
```

# Option: echo=false

Vhodna .Rnw datoteka  
Izpis kode izključen

```
<<echo=false>>=  
x <- c(1,2,2,4,3,6,5,3,2)  
length(x)  
table(x)  
@
```

Izhodna .pdf datoteka  
Izpis kode izključen

```
[1] 9  
x  
1 2 3 4 5 6  
1 3 2 1 1 1
```



# Option: results

Vhodna .Rnw datoteka

```
<<>>=  
options(width=70)  
library(Hmisc)  
library(xlsReadWrite)  
2+2  
@
```

Izhodna .pdf datoteka

```
> options(width = 70)  
> library(Hmisc)  
> library(xlsReadWrite)
```

```
xlsReadWrite version 1.3.2 (  
Copyright (C) 2007, Hans-Peter
```

```
This package can be freely d  
purpose. It comes with ABSOL  
xlsReadWrite has been writte  
code from a 3rd party library
```

```
Please refer to http://treet  
bugreports, donations, updat
```

```
> 2 + 2
```

```
[1] 4
```

# Option: results=hide

Vhodna .Rnw datoteka

```
<<results=hide>>=  
options (width=70)  
library (Hmisc)  
library (xlsReadWrite)  
2+2  
@
```

Izhodna .pdf datoteka

```
> options (width = 70)  
> library (Hmisc)  
> library (xlsReadWrite)  
> 2 + 2
```

# Option: results=tex

Vhodna .Rnw datoteka

```
<<results=tex,echo=false>>=  
x <- c(1,2,2,4,3,6,5,3,2)  
tbl <- as.data.frame(table(x))  
latex(tbl,file="")  
@
```

Izhodna .pdf datoteka

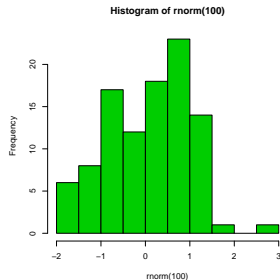
tbl	x	Freq
1	1	1
2	2	3
3	3	2
4	4	1
5	5	1
6	6	1

# Option: figs=true

Vhodna .Rnw datoteka

```
<<echo=false, fig=true>>=  
hist(rnorm(100), col=3)  
@
```

Izhodna .pdf datoteka



# Splošne nastavitve

## SweaveOpts

Izbiram lahko spremenimo privzete vrednosti:

```
\SweaveOpts{eps=false,prefix.string=./figs/SwPres}
```

Velikost slike nadziramo z Gin parametrom

LaTeX paketa **graphicx**

```
\setkeys{Gin}{width=0.8\textwidth}
```

## Vključitev rezultatov v besedilo: Sexpr . . .

```
> x <- rnorm(20, mean = 10)
> y <- rnorm(15, mean = 10.5)
> primerjava <- t.test(x, y)
> (p <- round(primerjava$p.value, 5))

[1] 0.15923
```

Verjetnost za napako prve vrste  $p = 0.15923$

# SweaveInput

Ta rezultat je vključen iz datoteke 'vključen1.Rnw'

```
> 3 * 3
```

```
[1] 9
```

# Kaj potrebujemo?

- Sweave (je že vključen v R )
- LaTeX (npr. MiKTeX)
- kak urejevalnik besedil (npr. WinEdt A. Simonič)



# Kako poženem Sweave iz R ?

```
Sweave(' ../doc/P02.Rnw')  
shell('latex ../doc/P02.tex')  
shell.exec('C:/_Y/R/SweaveP/doc/P02.pdf')
```

# Sweave.bat

Uporabimo lahko paketno datoteko Sweave.bat:  
`shell('Sweave ../doc/sd.Rnw')`

Paketne datoteke najdemo na

<http://cran.at.r-project.org/contrib/extra/batchfiles/>  
ali pa

<http://batchfiles.googlecode.com/>

# Primeri

Nekaj primerov in poskusne datoteke najdete na  
<http://ablejec.nib.si/R>