



OSSConf 2012: 115–120

## SWEAVE: SVÁZÁNÍ $\LaTeX$ U S JAZYKEM R

PAVEL STRÍŽ (CZ)

**Abstrakt.** Autor si váží toho, že zrovna on vám můžeme představit program Sweave {čtete ten či ta svív}, který elegantním způsobem, s nastavením parametrů vzoru jazyka *noweb*, propojil  $\LaTeX$  s výpočetním prostředím R. Jako je Open Source Software  $\TeX$  přijímán komunitou typografů a jejich přátel, tak je R přijímán řadou matematiků a statistiků ve světě výpočetním, viz <http://www.r-project.org/>.

Sweave umožňuje a umí to, že výstupy R z RNW souboru dokáže vpasovat do  $\LaTeX$  dokumentu. Tím se dokument stává dynamickým a grafy a statistické analýzy jsou aktualizované z pohledu vstupních dat. Program si představíme, nainstalujeme a nastavíme pod Microsoft Windows, v Linuxu je to obdobné. Naučíme se s programem Sweave pracovat a ukážeme si typické příklady. Kvůli rozsahu je plná verze článku na webu OSSConf2012.

**Klíčová slova.** Výpočty, Sweave, R,  $\TeX$ ,  $\LaTeX$ .

## SWEAVE: MIXING $\LaTeX$ WITH THE R LANGUAGE

**Abstract.** The article introduces professor Leisch's package Sweave from the R language and computing environment. The Sweave tool puts together  $\LaTeX$  and the R language in a simple way. The authors present their experiences with the installation and functionality testing together with setting the new user-defined batch files under Microsoft Windows.

The author also presents examples of chunks for basic R computing with variable handling and printing. Also table creation and graph production in R which includes graphic output into the  $\LaTeX$  file is presented. Enjoy Sweaveing! The full version of the article is available at the official OSSConf2012 website.

**Keywords.** Computing, Sweave, R,  $\TeX$ ,  $\LaTeX$ .

## Úvod

Otcem i matkou Sweave je profesor Friedrich Leisch [4]. Domovská stránka programu je <http://www.stat.uni-muenchen.de/~leisch/Sweave/>, kde je manuál, otázky a odpovědi (FAQ), články i několik ukázek. Sweave je součástí základního balíčku *utils* jazyka R.

R je oblíbený mezi statistiky a matematiky. Je to Open Source Software, programovatelný a s možností grid grafiky. R je založený na jazyku S-PLUS, blíže např. [3]. Závěrečná kapitola této knihy si všímá rozdílu mezi R a S-PLUS.

Pan Hlávka z MFF UK v Praze dodává [2], že oba jazyky jsou téměř shodné, ale nejsou kompatibilní, mohou se lišit jména některých příkazů a některé konstrukce mohou fungovat trochu jinak.

## Instalace

Začněmež! Sedíme na počítači s Microsoft Windows XP, máme nainstalován  $\text{\TeX}$  Live 2011. Pokud sedíte na jiném operačním systému, máte výhodu, protože poinstalační Sweave zlobí právě pod Microsoft Windows, jak zmiňuje i referenční zdroj jazyka R [6].

Z oficiálních stránek <http://www.r-project.org/> si stáhneme binárky pro Microsoft Windows. Potřebujeme administrátorská práva. Nainstalujeme si.

Věc, kterou chceme nastavit je cesta (PATH) ke spustitelným souborům. Lze využít Start→Ovládací panely→Systém, karta Upřesnit, Proměnné prostředí, Systémové proměnné a PATH upravujeme. Děláme to z toho důvodu, abychom mohli R spouštět bez vypisování adresáře. Při například verzi 2.7.1 (září 2008) to znamená přidat za současné nastavení (středník na začátku, bez tečky na konci): ;C:\Program Files\R\R-2.7.1\bin\. Dále už jen OK. OK. OK.

Přes ikonku „R 2.7.1“ spouštíme RGui s oknem R Console. Kdo je v R nový, může vyzkoušet příkazy potvrzené klávesou Enter: `demo()`, `demo(graphics)`, ...

K hlubšímu studiu jazyka R doporučujeme:

<a href="http://zoonek2.free.fr/UNIX/48_R/all.html">http://zoonek2.free.fr/UNIX/48_R/all.html</a>	{En, Fr}
<a href="http://bm2.genes.nig.ac.jp/RGM2/">http://bm2.genes.nig.ac.jp/RGM2/</a>	{Galerie}
<a href="http://addictedtor.free.fr/graphiques/allgraph.php">http://addictedtor.free.fr/graphiques/allgraph.php</a>	{Galerie}

Poněvadž již víme, že Sweave je součástí základní instalace, tak vyzkoušejte: `help(Sweave)`, či `example(Sweave)`.

## Poinstalační zásahy

Vše by mělo fungovat, s tím rozdílem, že bychom si přáli něco věcí vylepšit.

V dokumentaci se dočteme, že na tabulky lze užívat R balíček `xtable`. Nainstalujeme si jej. V grafickém R klikáme Packages→Set CRAN mirror... Vybereme si jeden z nich. OK. Pokračujeme Packages→Install packages(s)... , zvolíme `xtable`. OK. Kdo nerad kliká, necht' zapíše na jednom řádku přímo:

```
install.packages("xtable",
  contriburl=contrib.url("http://cran.r-project.org"))
```

Jak píše referenční příručka, bez chybějícího `\usepackage{Sweave}` v RNW se snaží Sweave přidat své nastavení, které pod Microsoft Windows zlobí (mezery, názvy delší než 8 znaků), resp. zlobí v  $\text{\TeX}$ u u aktivních znaků. Proto v RNW na balíček Sweave nezapomeneme. V opačném případě bychom nastavili cestu k `Sweave.sty`, nebo jej nakopírovali do lokálního adresáře.

Také si přejeme Sweave spouštět mimo R. Přes příkaz `sh sweave.sh` to pod Windows nepůjde samo o sobě.

Zkušený linuxový uživatel možná uvítá komplexní řešení<sup>1</sup> či jiný styl psaní o Sweave<sup>2</sup>. Komentáře a zkušenosti se Sweave pod Linuxem a se stylem APA jsou v článku [7]. Jednoduše lze přes balíček R2HTML užít Sweave do HTML<sup>3</sup>.

Sweave dělá to, že vezme náš soubor (RNW) a generuje T<sub>E</sub>Xový soubor (TEX), a to tak, že vnitřní pasáže (chunk, tj. cvalíky) spouští v R a vrací nám výpočty, zdrojové kódy či grafy (PDF a zároveň EPS; T<sub>E</sub>X načte jeden z nich dle přepínače). Běžné T<sub>E</sub>X zdroje během generování nemění, jen slepě kopíruje.

První úkol zařídíme takto. Připravíme si soubor `mySweave.r`:

```
library(utils)
args <- commandArgs(TRUE)
Sweave(args)
# Stangle(args)
```

Tento soubor si načte R balíček `utils`. Vstupní parametry uloží do proměnné `args` a to vše prožene Sweave. Vstupní argument bude název zpracovávaného RNW souboru. Pokud smažeme `#` z posledního řádku, tak ovladač `Stangle` vám cvalíky vyextrahuje do speciálního R souboru.

Vedle toho si připravíme soubor `mySweave.bat`:

```
Rscript mySweave.r %1
R CMD pdflatex.exe %1.tex
R CMD pdflatex.exe %1.tex
```

Tento dávkový soubor nám `mySweave.r` prožene R. Zpracuje RNW na TEX. Pak by stačilo spustit běžné T<sub>E</sub>X prostředí. Problém je, že cestu k `Sweave.sty` zná jen R. Proto doplníme R `CMD`, což nám zajistí přístup do standardního adresáře `C:/Program Files/R/R-2.7.1/share/texmf/`.

V příkazové řádce spouštíme, nebo v našem oblíbeném editoru nastavíme, dávkové spuštění `mySweave myfile`, kde `myfile` je myšlen `myfile.Rnw`.

Mějte tento článek, kde je řídicí soubor, který T<sub>E</sub>Xujete, a článkové soubory v podadresáři, tak `mySweave.bat` může vypadat takto (konkrétně podadresář Pavla Stríže `/PSsweave/` s článkem `PSsweave.Rnw`; argument `%1` je soubor řídicí):

```
cd PSsweave
Rscript mySweave.r PSsweave.Rnw
cd ..
R CMD lualatex.exe %1.tex
bibtex %1
R CMD lualatex.exe %1.tex
R CMD lualatex.exe %1.tex
```

<sup>1</sup><http://sites.google.com/site/gregorgorjanc/software/Sweave.sh>

<sup>2</sup>[stat.epfl.ch/webdav/site/stat/shared/Regression/EPFL-Sweave-powerdot.pdf](http://stat.epfl.ch/webdav/site/stat/shared/Regression/EPFL-Sweave-powerdot.pdf)

<sup>3</sup><http://biostat.mc.vanderbilt.edu/twiki/pub/Main/SweaveConvert/htmlWeave.pdf>

Na tomto místě dodáváme, že EXE je nutností, jinak by R hledal vnitřní balíčky, ne spustitelný soubor. Také za pozornost stojí, že `bibtex` stojí osamocen bez R CMD. Je to z toho důvodu, že nepotřebuje knihovny a balíčky od R.

Poslední věc, která nám může vadit je standardní nastavení pro výpis zdrojových kódů. Sweave, konkrétně jsou to poslední čtyři řádky v souboru `Sweave.sty`, pracuje s  $\text{T}_{\text{E}}\text{X}$  balíčkem `fancyvrb`, který užívá v základním nastavení.

Možnosti `fancyvrb` jsou poměrně bohaté a lze si upravit. Autor dává na zdrojové kódy přednost balíčku `listings`. Jeho zanesení a užití přes Sweave je běžná  $\text{T}_{\text{E}}\text{X}$ ová operativa. Záměrně zůstaneme u Sweave nastavení a prostředí `verbatim`.

## Cvalíkujeme: užití chunks (cvalíků kódu)

Nahlédněte prosím na webové stránky OSSConf2012 pro plnou verzi článku.

## Souhrnná ukázka

Tato ukázka je jedna z našich oblíbených R kódů pro studenty. Mají se najít dvě střední hodnoty vrcholů smíšených gaussových křivek nad zadanými daty.

Vstupní data jsou 1, 2, 2, 2, 1, 3, 2, 2, 2, 7, 8, 9, 9, 9, 9, 9, 10, 9, 9, 14, 8, 8, 9, 9, 9, 10, 11, 0, 2, 2, 4 a 5.

Na tento problém si v R doinstalujeme R balíček `mclust`, více viz [1], případně `help(Mclust)` zapsané v R po instalaci `mclust`. Je to balíček, který není licencován mimo studijní účely. Užijme k instalaci například kanadského zrcadla CRANu.

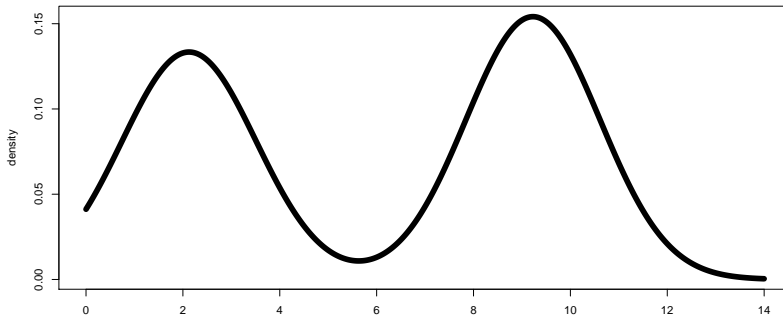
```
install.packages("mclust",
  contriburl=contrib.url("http://cran.stat.sfu.ca/"))
```

Řešení dle Bayesova informačního kritéria BIC (Bayes Information Criterion). První vrchol má souřadnici složky  $x$  2.1273, druhý pak 9.2273.

Výpočet a výpis předchozího odstavce byl v `.Rnw` zapsán takto. Příkaz `rm()` vyčistí předchozí proměnné. Výstupní obrázek a jeho zdroj následují.

```
<<echo=FALSE,results=HIDE>>=
  rm(list=ls(all=TRUE))
  library(mclust)
  y<-c(1,2,2,2,1,3,2,2,2,7,8,9,9,9,10,9,9,14,
      8,8,9,9,9,10,11,0,2,2,4,5)
  yModel <- mclustModel(y, mclustBIC(y, G=2))
  yModel$parameters$mean
  @
```

Řešení dle BIC. První vrchol má souřadnici složky  $x$  `\Sexpr{round(yModel$parameters$mean[1], digits=4)}`, druhý pak `\Sexpr{round(yModel$parameters$mean[2], digits=4)}`.



Obrázek 1. Odhad hustoty součtem dvou gaussových křivek

```

\begin{figure}[hbt]
  \centering
  <<echo=FALSE, fig=TRUE, width=12>>=
    mclust1Dplot(y, parameters = yModel$parameters,
      z = yModel$z, what="density", identify = FALSE,
      ngrid = length(y)*15, lwd=8)
  @
  \caption{Odhad hustoty součtem dvou gaussových křivek}
  \label{hustota}
\end{figure}

```

## Závěr

Myslíte-li to s programem Sweave vážně, tak si dejte pozor na to, že editujete soubor RNW a TEX je v tomto případě generován. Budete-li omylem editovat TEX soubor, tak si při stejného názvu RNW spuštěním Sweave TEX soubor přepíšete.

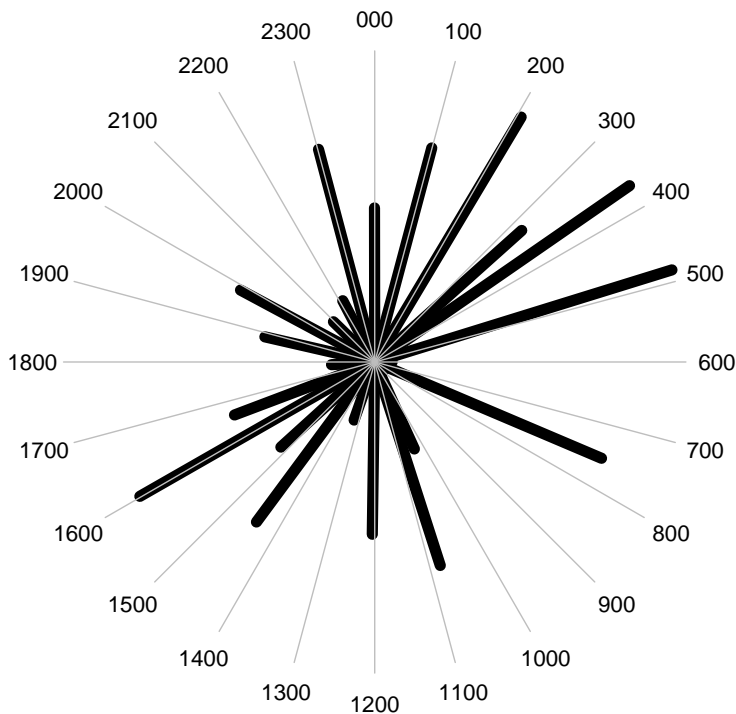
Profesor Leisch vytvořit také pomůcku vignettes [5], která rozšiřuje některé možnosti přípravy a prezentování dokumentace programů R a BioConductor.

V závěru Sweave manuálu profesor píše, že pracuje od nuly na balíčku S4weave, který by měl Sweave nahradit a překonat v nabízených možnostech. Máme se na co těšit, i když nekonečné potřeby smrtelníků bohatě pokrývá potenciál Sweave.

Zbytek zkoumání necháváme na vás. S vámi i se všemi cvalíky se loučí autor!

## Reference

- [1] Fraley C., Raftery A. E.: *Mclust version 3 for R: Normal mixture modeling and model-based clustering*, Technical Report 504, University of Washington, Department of Statistics, September 2006.
- [2] Hlávka Z.: *Emailová odpověď velkého pražského eRkaře autorovi na vznesený vznešený dotaz*, 28. dubna 2008.



**Obrázek 2.** Grafický potenciál jazyka R je pro techniky neskutečný

- [3] Krause A., Olson M.: *The Basic of S-PLUS*, First edition, ISBN 0-387-26109-5.
- [4] Leisch F.: *Sweave*: Dynamic generation of statistical reports using literate data analysis, In Wolfgang Härdle and Bernd Rönz, editors, *Compstat 2002—Proceedings in Computational Statistics*, 575–580, Physica Verlag, Heidelberg, 2002, ISBN 3-7908-1517-9.
- [5] Leisch F.: *Sweave*, part II: Package vignettes, *R News*, 3(2): 21–24, October 2003.
- [6] The R Development Core Team: *R: A Language and Environment for Statistical Computing*, Reference Index. *Sweave*: Automatic Generation of Reports. The *utils* package, Version 2.7.1 (2008-06-23), ISBN 3-900051-07-0.
- [7] Zahn I.: *Learning To Sweave in APA Style*, The *PracTeX* Journal, (1), 2008, <http://tug.org/pracjourn/2008-1/zahn>.

## Kontaktní adresa

**Ing. Pavel Stríž, Ph.D.**, Fakulta managementu a ekonomiky, Univerzita Tomáše Bati ve Zlíně,  
T. G. Masaryka 5555, 760 01 Zlín, Česká republika,  
*E-mailová adresa*: [striz@fame.utb.cz](mailto:striz@fame.utb.cz)