

# Úvod, význam a výhody jazyka R, instalace a motivace, proč se ho učit — Úvod do jazyka R

Lubomír Štěpánek<sup>1, 2</sup>



<sup>1</sup>Oddělení biomedicínské statistiky & výpočetní techniky  
Ústav biofyziky a informatiky  
1. lékařská fakulta  
Univerzita Karlova v Praze



<sup>2</sup>Katedra biomedicínské informatiky  
Fakulta biomedicínského inženýrství  
České vysoké učení technické v Praze

(2017) Lubomír Štěpánek, CC BY-NC-ND 3.0 (CZ)



Dílo lze dále svobodně šířit, ovšem s uvedením původního autora a s uvedením původní licence. Dílo není možné šířit komerčně ani s ním jakkoliv jinak nakládat pro účely komerčního zisku. Dílo nesmí být jakkoliv upravováno. Autor neručí za správnost informací uvedených kdekoli v předložené práci, přesto vynaložil nezanedbatelné úsilí, aby byla uvedená fakta správná a aktuální, a práci sepsal podle svého nejlepšího vědomí a svých „nejlepších“ znalostí problematiky.

# Obsah

- 1 Plán dnešního semináře
- 2 Motivace
- 3 Úvod do jazyka R
- 4 Instalace
- 5 Začínáme s R
- 6 Literatura

# Plán dnešního semináře

- dopolední část (mezi 8:00 – 11:00)
  - ① úvod, význam a výhody jazyka R, instalace a motivace, proč se ho učit
  - ① přehled datových typů a struktur, jednoduché operace
  - ② načítání dat do R z externích zdrojů a jejich ukládání
  - ③ funkce v R a R jako programovací jazyk
  - ④ analytické možnosti jazyka R

# Plán dnešního semináře

- odpolední část (mezi 13:00 – 16:00)
  - 1 základy statistiky v R
  - 2 aplikace jazyka R v (bio)medicině
  - 3 grafické výstupy, diagramy v R
  - 4 práce s textem v R

# Prezentace a skript

- prezentace a skript lze nalézt na adrese

[https://github.com/LStepanek/uvod\\_do\\_jazyka\\_R](https://github.com/LStepanek/uvod_do_jazyka_R)

# Proč bych se právě já měl učit s R

- jsem pregraduální student
  - chtěl bych se věnovat postgraduálnímu doktorskému studiu nebo vědě, kde jistojistě budu potřebovat statistiku
- jsem postgraduální student
  - jednou bych chtěl dokončit postgraduální doktorské studium, k čemuž tu zatracenou statistiku opravdu potřebuji
  - komerční statistické programy pro mě nejsou dostupné, nebo nejsou dobře použitelné

# Proč bych se právě já měl učit s R

- jsem vědec, „postdoc“ či lékař
  - s vědou (nejen) v biomedicině to myslím vážně
  - publikuji v odborných časopisech a rád bych do statě článku s Methodology and Statistical Analysis pravdivě psal, že „... *all statistical analyses were performed using R language for statistical computing and graphics*...“, protože existují důkazy, že citování R či jiných volných statistických nástrojů mnohdy zvyšuje pravděpodobnost citování takového článku
  - tuším, že věda 2.0 v biomedicině se bude provozovat nejen formou experimentů na živém (in vivo) či v laboratořích (in vitro), ale budou ji tvořit ze značné části počítačové modely a simulace (in silico)
  - hledám konečně efektivní nástroj pro analýzy svých výzkumů



# Proč bych se právě já měl učit s R

- s vědou nemám nic společného
  - sbírám opakovaně data stejného charakteru a rád bych si jejich (před)zpracování automatizoval
  - zpracování dat v tabulkových procesorech a spoléhání se jen na ně je spjato s různými problémy, chybné výstupy z tabulkových procesorů dokonce vyvolaly některé vědecké skandály
  - data již nějakou dobu (sám) analyzuji a přemýšlím, který nástroj pro analýzu (s kvalitní dokumentací a živou podporou a komunitou) se začít učit
  - uvědomuji si, že MS Excel v základním rozhraní neumí doted' vykreslit krabicový diagram a nestačí i jinde

# Co je R



- R je interpretovaný programovací jazyk
- kombinuje několik paradigmat
  - imperativní
  - funkcionální
  - objektové
- R je *domain specific language* – je určen pro statistickou analýzu dat a jejich grafické zobrazení
- R je open-source, konkrétně *free-as-in-beer* a *free-as-in-speech*

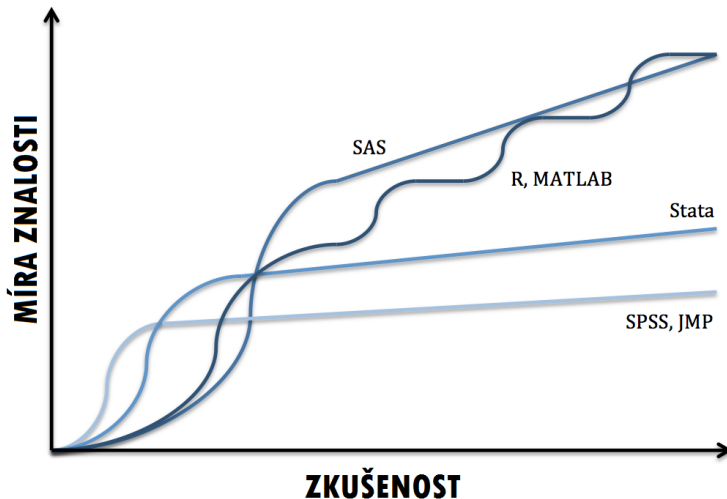
# Stručná historie R

- R vychází z jazyka a prostředí S, které vyvinuto v Bellových laboratořích v letech 1975-1976 (prof. John Chambers)
- přerod v R v roce 1992 na Acklandské univerzitě na Novém Zélandu (prof. Ross Ihaka a Robert Gentleman)
- R je tedy akcentem S
- v roce 1994 první verze prostředí R pro volné použití, poté postupně vzniká řada dalších verzí, v listopadu 2017 poslední verze R 3.4.3 („Kite-Eating Tree“)
- za otce moderního R považován Hadley Wickham (leader analyst v RStudios, adjunct professor na Aucklandské univerzitě)

# Výhody jazyka R

- ① běžné výhody společné všem skriptovacím jazykům
  - „nekonečné“ možnosti zpracování a analýzy dat omezené pouze představivostí uživatele
- ② výborně zapracovaná explorativní analýza dat
- ③ velmi oceňované, pěkné a libovolně customizovatelné grafické výstupy
- ④ R je zdarma
- ⑤ R má pravděpodobně vůbec nejživější komunitu a softwarovou podporu

# Křivka učení se jazyka R



(zdroj: volně dle <https://guides.nyu.edu/quant/statsoft>)

# Stažení a instalace jádra R

- na stránkách R-projectu

`https://www.r-project.org/`

postupně **download R**, vyberme českou doménu a stáhněme desktopově

- poté instalujme dle instrukcí do předvolené složky

# Stažení a instalace RStudio

- RStudio je jedním z grafických IDE (Integrated Development Environment) jazyka R
- na stránkách RStudio

<https://www.rstudio.com/>

postupně **Products > RStudio > Desktop > Open Source Edition > Free > Download**, stáhněme desktopově

- poté instalujme dle instrukcí do předvolené složky

# Další software

- může se časem hodit
  - obecný textový editor, např. Notepad++

<https://notepad-plus-plus.org/>

- univerzální konvertor dokumentů Pandoc

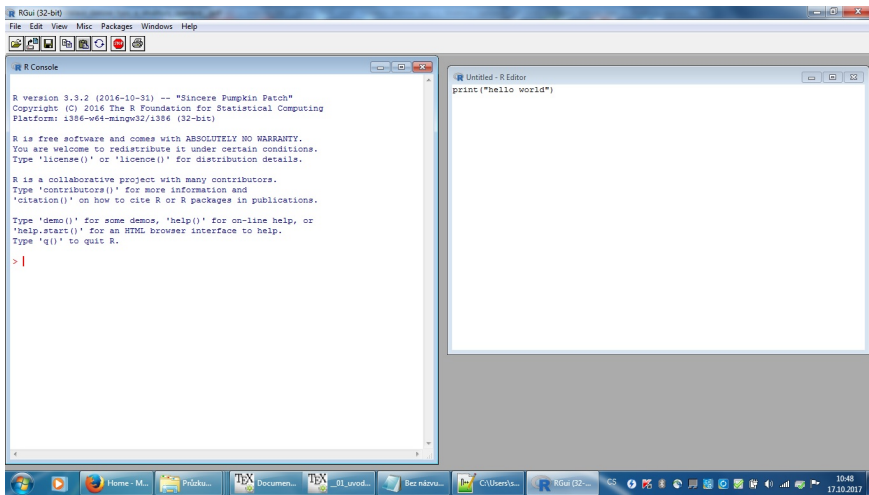
<https://pandoc.org/>

- typografický sázeč T<sub>E</sub>X

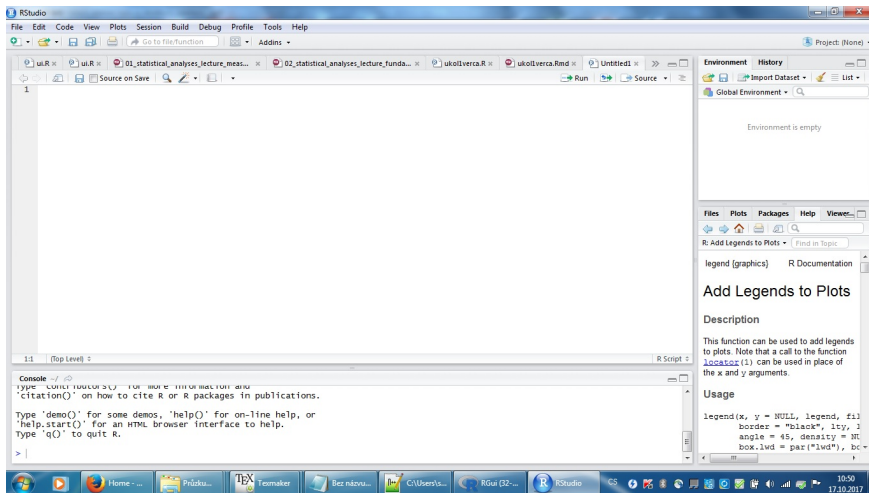
<https://miktex.org/>



# První spuštění R jádra



# První spuštění RStudio



# Ahoj světe!

- do skriptu či konzole napíšeme

```
1 ||      print("hello world")
```

- dostaneme

```
1 ||      [1] "hello world"
```

# Práce s nápovědou

- nápovědu pro funkci či objekt získáme pomocí příkazu `help()`, kde argumentem je název funkce či objektu

```
1 ||      help(print)
```

- nebo předsazením symbolu `?` před název funkce či objektu

```
1 ||      ?print
```

- předsazením symbolů `??` před název funkce či objektu prohledáme veškeré dokumenty nápovědy

```
1 ||      ??print
```

- vždy je zavolán HTML soubor s volnotextovou nápovědou

# Přístup k práci v R

- ① interaktivní práce mezi skriptem a konzolí (*shell programming*)
- ② (dávkové) volání skriptů
  - (i) vytvoření funkčního skriptu s příponou .R
  - (ii) uložení skriptu do pracovní složky, kterou zjistíme pomocí příkazu `getwd()`
  - (iii) zavolání a exekuce skriptu v konzoli pomocí příkazu `source(nazev_skriptu.R)`

# Intermezzo

- ❶ uložíme následující kód jako `scitam_dve_cisla.R`

```
1 ||      a = 3
2 ||      b = 4
3 ||      print(a + b)
```

- ❷ skript `scitam_dve_cisla.R` přesuňme do složky, kterou vrátí příkaz

```
1 ||      getwd()
```

anebo vhodně nastavme pracovní složku pomocí příkazu

```
1 ||      setwd(choose.dir())
```

- ❸ do konzole nyní napíšme

```
1 ||      source("scitam_dve_cisla.R")
```

# Instalace nadstavbových balíčků

- nejlépe pomocí příkazu

```
1 | install.packages(  
2 |     "nazev_balicku",  
3 |     dependencies = TRUE,  
4 |     repos = "http://cran.us.r-project.org"  
5 | )
```

- poslední argument repos není nutný, ale je šikovný, vybere doménu, z které balíček stáhnout, aniž by ji musel uživatel vybrat kliknutím v pop-up okně (vhodné pro úplnou automatizaci kódu)
- zkusme nainstalovat balíčky ShinyItemAnalysis a swirl!

# Literatura



ZVÁRA, Karel. *Základy statistiky v prostředí R*.  
Praha, Česká republika: Karolinum, 2013. ISBN 978-80-246-2245-3.



WICKHAM, Hadley. *Advanced R*.  
Boca Raton, FL: CRC Press, 2015. ISBN 978-1466586963.



Děkuji za pozornost!

lubomir.stepanek@lf1.cuni.cz

lubomir.stepanek@fbmi.cvut.cz