

Načítání dat do R a ukládání dat na pevný disk

17VSADR – Skriptování a analýza dat v jazyce R

Lubomír Štěpánek^{1, 2}



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



²Katedra 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 Pracovní složka
- 2 Import a export dat
- 3 Literatura

Pracovní složka

- zjištění, která složka je pracovní

```
|| getwd()
```

- výpis obsahu pracovní složky formou vektoru

```
|| dir()
```

- nastavení pracovní složky

```
|| setwd("C:/.../my_working_directory")
```

- nastavení pracovní složky dialogovým oknem

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

Import a export volného textu

- pomocí funkcí `readLines()` a `writeLines()`
- lze tak nahrát libovolná data, která mají textovou reprezentaci

```
my_html <- readLines(  
  con = paste(  
    "https://predmety.fbmi.cvut.cz",  
    "cs/17VSADR",  
    sep = "/"  
  ),  
  encoding = "UTF-8"  
)
```

Import a export volného textu

- uložení a načtení desktopového dokumentu

```
writeLines(                # ukládám textový dokument
  text = paste(
    "One R to rule them all",
    "one R to find them",
    "one R to bring them all",
    "and in the darkness bind them",
    sep = "\n"             # separátor typu nový řádek
  ),
  con = "my_text.txt"
)

my_loaded_text <- readLines(
  con = "my_text.txt",
  encoding = "UTF-8"
)                          # načítám textový dokument
```

Import a export dat tabulky

- zāklaní funkcí je `read.table()` a `write.table()`
- obě funkce mají řadu wrapperů (`read.csv()` a `write.csv()`, `read.delim()` a `write.delim()` a další)

```
write.table(                                     # ukládám data.frame
  x = mtcars,
  sep = ";",
  row.names = FALSE,
  file = "mtcars.csv" # anebo "mtcars.txt"
)

my_mtcars <- read.table(
  file = "mtcars.csv",
  sep = ";",
  header = TRUE
)                                                # načítám data.frame
```

Import a export dat tabulky

- funkce `read.table()` má spoustu užitečných argumentů

```
write.table(iris, "iris.txt")
my_iris <- read.table(
  file = "iris.txt",
  sep = " ",
  header = TRUE,
  stringsAsFactors = FALSE,
  nrows = 1,      # načte jen první řádek;
                  # může se hodit pro odhad koerce,
  check.names = FALSE,
                  # vynechá kontrolu korektnosti
                  # popisků sloupců
  colClasses = "character"
                  # přetypuje všechny sloupce
                  # na textové proměnné
)
```


Import a export dat tabulky z MS Excel® (.xlsx)

- vhodný je například balíček openxlsx
 - má výhodu, že narozdíl od balíčku např. xlsx nepotřebuje Java Tool Kit, takže dokáže najednou nahrát více souborů MS Excel®
- uložení tabulky do excelového formátu (.xlsx)

```
browseURL(paste(
  "https://raw.githubusercontent.com/LStepanek",
  "17VSADR_Skriptovani_a_analyza_dat_v_jazyce_R",
  "master/export_dat_do_ms_excelu.R", sep = "/"
))
```

- načtení excelové tabulky

```
my_data<- read.xlsx(
  xlsxFile = "moje_tabulka_je_ted_v_excelu.xlsx",
  sheet = 1,      # anebo jméno listu
  colNames = TRUE
)
```

Import bitmapového obrázku

- vhodný je například balíček png, jpeg či raster

```
my_picture <- readJPEG(  
  "__03_landmarky__.jpg"  
)
```

- výsledkem je array o třech rozměrech
 - svislá souřadnice
 - vodorovná souřadnice
 - barevné kanály

Export konzolového výpisu do souboru

- pomocí `sink()` - `sink()` anebo `capture.output()`

```
(muzi <- rnorm(100, mean = 175, sd = 10))
(zeny <- rnorm(100, mean = 160, sd = 10))

t.test(muzi, zeny)

# výpis z konzole do textového souboru
capture.output(
  t.test(muzi, zeny),
  file = "t_test.txt"
)

# anebo
sink("tohle_je_taky_t_test.txt")
t.test(muzi, zeny)
sink()
```

Export smysluplného R-kového objektu do TeXového kódu

- pomocí balíčku xtable

```
library("xtable")

my_linear_model <- lm(mpg ~ hp + cyl,
                      mtcars)

print(
  xtable(my_linear_model,
         digits = 4),
  floating = FALSE,
  tabular.environment = "tabular",
  hline.after = NULL,
  include.rownames = TRUE,
  include.colnames = TRUE
)
```

Import „exotických“ souborů do R

- naším přítelem je balíček `foreign`
- podporuje načítání dat z formátů
 - Epi Info
 - Minitab
 - S
 - SAS, SPSS, STAT, Systat, Weka

```
library("foreign")

# import dat z SPSS
my_data <- read.spss(
  file = "du1_30.sav",
  to.data.frame = TRUE
)
```

Intermezzo

- jak importovat data z MS Word®?
- jak načíst do R datovou strukturu list?

Literatura



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



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

Děkuji za pozornost!

lubomir.stepanek@lf1.cuni.cz

lubomir.stepanek@fbmi.cvut.cz

github.com/LStepanek/17VSADR_Skriptovani_a_analyza_dat_v_jazyce_R