

Procvičování 8 - řešení

1. Naimportujte do **R** křestní jména z listu *jména* excelového souboru *dat_2013.xls*. Použijte funkci `scan()`.

```
# Do R posleme prikaz
jména <- scan(what = "", sep = "\t")
# a nasledne zkopirujeme a do R vlozime cely sloupec jmen (bez zhlavi). Po
# poslednim odklepeme Enter. Take bychom mohli neprve sloupec zkopirovat
# do schranky a pri scanovani specifikovat, ze R ma data brat ze schranky:
jména <- scan("clipboard", what = "", sep = "\t")
```

2. Naimportujte do **R** počty výskytů těchto jmen v ČR (vedlejší sloupec). Opět použijte funkci `scan()`.

```
# stejny postup jako v prvni bode, jen what= specifikujeme, ze importujeme
# cisla, pripadne argument what= vynechame.
pocety <- scan(what = 0, sep = "\t")
pocety <- scan(sep = "\t")
```

3. Naimportujte do **R** taxonomické zařazení pakomárů do podčeledí (list *tax*). Opět použijte funkci `scan()`. Jedná se o stejné druhy pakomárů, jako jsou v listu *spe*.

```
# postup stejny jako v bode 1.
tax <- scan(what = "", sep = "\t")
```

4. Naimportujte do **R** do dataframu *env1* nám již známé proměnné prostředí (list *env*). Použijte funkci z rodiny `read.()` a pro přesun dat použijte datovou schránku "clipboard". Při importu specifikujte, že jména řádků má **R** vzít z prvního sloupce *sample*. Po importu zkontrolujte strukturu dataframu.

```
# Tohle se muze lisit na ruznych pocitacich, zalezi na nastaveni desetinného
# oddelovace (ja mam tecku '.'). A moznosti je samozrejme vic. Nejprve v
# Excelu oznacime celou oblast, kterou chceme importovat, zkopirujeme ji do
# schranky (Ctrl + C) a posleme do R jeden z nasledujících prikazu. V prvni
# uvadim vsechny dulezite argumenty, viz napoveda ?read.table.
env1 <- read.table("clipboard", row.names = 1, header = T, sep = "\t", dec = ".")
env1 <- read.delim("clipboard", row.names = 1)
# read.delim() pro pocitace, jez maji tecku '.' jako desetinný oddelovac
env1 <- read.delim("clipboard", row.names = "sample")
env1 <- read.delim2("clipboard", row.names = 1)
# read.delim2() pro pocitace, jez maji carku ',' jako desetinný oddelovac
```

```
str(env1)
```

```
## 'data.frame': 27 obs. of 5 variables:
## $ hab      : Factor w/ 5 levels "Ep","Ep_CPOM",...: 1 1 1 3 3 3 4 4 4 5 ...
## $ hydr     : Factor w/ 2 levels "pool","rifle": 1 1 1 1 1 1 2 2 2 2 ...
## $ depth    : num  0.395 0.422 0.496 0.291 0.32 0.328 0.278 0.213 0.243 0.353 ...
## $ velocity: num  0.274 0.358 0.31 0.078 0.092 0.126 0.49 0.618 0.508 0.758 ...
## $ froude   : num  0.1392 0.176 0.1405 0.0462 0.0519 ...
```

5. Naimportujte do **R** do dataframu *env2* ty stejné proměnné prostředí (list *env*) a stejně jako v předchozím kroku. Navíc při importu specifikujte, že nechcete, aby proměnná *hydr* byla převedena na faktor. Po importu zkontrolujte strukturu dataframu.

```
env2 <- read.delim("clipboard", row.names = 1, as.is = "hydr")
```

```
str(env2)
```

```
## 'data.frame': 27 obs. of 5 variables:
## $ hab      : Factor w/ 5 levels "Ep","Ep_CPOM",...: 1 1 1 3 3 3 4 4 4 5 ...
## $ hydr     : chr  "pool" "pool" "pool" "pool" ...
## $ depth    : num  0.395 0.422 0.496 0.291 0.32 0.328 0.278 0.213 0.243 0.353 ...
## $ velocity: num  0.274 0.358 0.31 0.078 0.092 0.126 0.49 0.618 0.508 0.758 ...
## $ froude   : num  0.1392 0.176 0.1405 0.0462 0.0519 ...
```

6. Uložte tabulku abundancí larev pakomárů listu *spe* do samostatného textového souboru (Tab delimited).

```
# V Excelu klikneme na 'Ulozit jako...' v zalozce 'Soubor' a jako typ
# souboru vybereme 'Text oddeleny tabulatory'.
```

7. Naimportujte do **R** do dataframu *spe* abundance larev pakomárů z nově vytvořeného textového souboru. Při importu specifikujte, že nechcete, aby **R** kontrolovalo a upravovalo druhová jména.

```
# Zde bud musime specifikovat celou cestu k danemu souboru, nebo si
# nastavime pracovni adresar. Data jsem si ulozil na D:/predmety/uvod do
# R/2014, zmenim si pracovni adresar:
setwd("D:/predmety/uvod do R/2014")
spe <- read.delim("spe1.txt", row.names = 1, check.names = F)
# argument check.names= F zajisti, ze R nebude kontrolovat jmena promennych
# - sloupcu.
```

8. Najděte funkci, která z druhových jmen vytváří zkratky (je v balíku *vegan*) a zkratky z druhových jmen vytvořte. (anglicky zkratka je abbreviate)

```
# Prvni odkaz Googlu pri hledani 'abbreviate species names vegan' mi vratil
# funkci vegan::make.cepnames(). Napoveda k dane funkci rika, ze jsem
# nasel, co jsem hledal.
library(vegan)
```

```
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.0-9

zkratky <- make.cepnames(names(spe))
```

9. Přejmenujte jména sloupců dataframu *spe* vytvořenými zkratkami.

```
names(spe) <- zkratky
```

10. Při importu se prázdné buňky vyplnily NA hodnotami. Nahraďte NA hodnoty dataframu *spe* nulami.

```
spe[is.na(spe)] <- 0
```

11. Zjistěte, kolik druhů je v dataframu *spe* (odpovědí **R** bude jedno číslo, a to počet sloupců dataframu).

```
ncol(spe)
```

```
## [1] 54
```

12. Představte si, že vzorky byly při determinaci děleny, aby determinátor nezešlél z množství pakomárů (kde jich bylo moc, byla determinována jen polovina, kde jich bylo ještě víc, jen čtvrtina). Abychom získali přibližnou představu o abundancích pakomárů na lokalitách, musíme abundance zpět vynásobit obráceným podílem zpracovaného vzorku (tedy dvojkou nebo čtyřkou). Ve skutečnosti tomu tak nebylo, proto podíly nasimulujeme. Řekněme, že vzorky s méně než 200 pakomáry pochází ze čtvrtiny původních vzorků (a je třeba jejich abundance vynásobit čtyřmi), a vzorky s méně než 400 pakomáry pochází z poloviny původního vzorku (tedy je musíme vynásobit dvěma).

- Vytvořte vektor násobků, kterými je potřeba abundance vynásobit.
- Vytvořte dataframe *spe.nas* s vynásobenými abundancemi pakomárů.
- Přesvědčte se, že podíly lokalitních sum abundancí vynásobeného a původního dataframu odpovídají vektoru násobků.

```
# (a) - samozřejmě je víc možností, tu jsou dvě:
nasobky <- c(4, 2, 1)[factor((rowSums(spe) >= 200) + (rowSums(spe) >= 400))]
table(nasobky)

## nasobky
## 1 2 4
## 11 5 11

nasobky <- c(4, 2, 1)[cut(rowSums(spe), c(0, 200, 400, max(rowSums(spe))), right = F,
  include.lowest = T)]
table(nasobky)
```

```
## nasobky
## 1 2 4
## 11 5 11

# (b)
spe.nas <- spe * nasobky
# (c)
all(rowSums(spe.nas) == (rowSums(spe) * nasobky))

## [1] TRUE
```

13. Exportujte dataframu *spe.nas* jako tabulátorem oddělený text tak, aby se dal otevřít v excelu. V excelu jej otevřete a zkontrolujte.

```
write.table(spe.nas, file = "spe.nas.txt", sep = "\t", row.names = T, col.names = NA)
# argumenty row.names= T a col.names= NA zajisti, ze v zhlavi bude pridana
# jedna prazdna bunka, jak je zvykem v excelovych spreadsheetech (viz
# ?write.table)
```

14. V některých analýzách chceme srazit vliv hojných druhů. Potom můžeme přistoupit k relativizaci nebo standardizaci druhových abundancí. Vytvořte dataframu *spe.rel* v němž budou zrelativizované druhové abundance tak, že pro každý druh je vydělíte sumou jeho abundancí. Pozor na násobení matic vektory - hodnoty v maticích bere **R** po sloupcích.

```
# Tady potrebujeme roztahnout vektor sum abundanci druhu tak, aby se kazda
# jedna hodnota opakovala tolikrat, kolik mame lokalit. Bud muzeme vektor
# zreplikovat pomoci funkce rep(), nebo pouzijeme jako subscript poradí
# sloupcu, který získáme pomoci funkce col():
sumy <- rep(colSums(spe), each = nrow(spe))
sumy <- colSums(spe)[col(spe)]
spe.rel <- spe/sumy
```

15. Exportujte dataframu *spe.rel* jako tabulátorem oddělený text tak, aby se dal otevřít v excelu. V excelu jej otevřete a zkontrolujte.

```
write.table(spe.rel, file = "spe.rel.txt", sep = "\t", row.names = T, col.names = NA)
```

16. Vytvořte dataframu *stat* s proměnnými *abund* a *freq* obsahujícími součty abundancí a počty výskytů jednotlivých druhů (počet řádků dataframu *stat* bude odpovídat počtu druhů, přičemž pro každý druh bude uvedeno, kolik jedinců celkem bylo zaznamenáno a na kolika lokalitách se vyskytoval).

```
stat <- data.frame(abund = colSums(spe), freq = colSums(spe > 0))
```

17. V dataframu *stat* vytvořte proměnnou odlišující běžné druhy od vzácných s úrovněmi: *vzacny* (druh se vyskytoval na max. 3 lokalitách), *bezny* (výskyt na 4 - 9 lokalitách), a *hojny* (výskyt na 10 a více lokalitách).

```
stat$status <- cut(stat$freq, c(0, 3, 9, max(stat$freq)), labels = c("vzacny",  
"bezny", "hojny"))
```

18. Zjistěte, kolik druhů je vzácných, běžných a hojných (podle kritérií v předchozím bodu).

```
table(stat$status)
```

```
##  
##  vzacny  bezny  hojny  
##      15     11     28
```

19. Exportujte dataframe *stat* jako tabulátorem oddělený text tak, aby se dal otevřít v excelu. V excelu jej otevřte a zkontrolujte.

```
write.table(stat, file = "stat.txt", sep = "\t", row.names = T, col.names = NA)
```