

Řazení datasetů

V této kapitole používám dataframy *spe* a *env* a vektor korelací druhů pakomárů s Froudeho číslem. Všechna data najdete v učebních materiálech v Excelovém souboru ve složce devátého cvičení¹.

Mnohdy potřebujeme data nějak seřadit. Třeba si chceme zběžně prohlédnout rozložení hodnot (hlavně jejich “ocasy”), nebo hledáme patterny mezi dvěma proměnnými, nebo se chceme blíže podívat na část dat, která je nejvíc nebo nejmíň něco, nebo jen chceme zjistit nějaké informace o několika “nej” objektech. Důvodů může být hromada. Zde najdete stručný návod, jak objekty řadit.

Funkce spojené s řazením hodnot jsou tři:

- `sort()`
- `rank()`
- `order()`

`sort()`

Funkce `sort()` **seřadí vlastní hodnoty** objektu podle jejich velikosti, případně abecedy. Můžeme třeba seřadit hodnoty Froudeho čísla pro jejich zběžné prohlédnutí, abychom se rychle podívali na nejnižší a nejvyšší hodnoty. Nejprve je ale zaokrouhlím, ať jsou přehlednější.

```
env$frR<- round(env$fr, 2)
sort(env$frR)

## [1] 0.00 0.02 0.04 0.05 0.05 0.07 0.07 0.13 0.13 0.14 0.14 0.14 0.17 0.18
## [15] 0.18 0.22 0.24 0.30 0.30 0.33 0.41 0.43 0.48 0.51 0.52 0.54 0.56
```

Řazení funguje nejen na numerických hodnotách, ale i na logických a textových. Na logických to ale asi nemá valný smysl.

```
sort(env$frR>0.23)

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE TRUE
## [23] TRUE TRUE TRUE TRUE TRUE

sort(rownames(env))

## [1] "s1" "s10" "s11" "s12" "s13" "s14" "s15" "s16" "s17" "s18" "s19"
## [12] "s2" "s20" "s21" "s22" "s23" "s24" "s25" "s26" "s27" "s3" "s4"
## [23] "s5" "s6" "s7" "s8" "s9"
```

`rank()`

Funkce `rank()` vrátí **pořadí hodnot, v jakém jsou** v daném objektu. Tuto funkci použijeme např. při neparametrických testech, kdy chceme pracovat s pořadími hodnot, nebo chceme-li zjistit pořadí nějakého měření vzhledem k ostatním.

¹Opět, jedná se o abundance larev pakomárů na ploškách dna v řece Svratce (dataframe *spe*). Pro každou plošku je zanamennán typ habitatu (5 kategorií), hloubka vody a rychlost proudu a vypočítáno Froudeho číslo (hydraulický parametr). Vektor *korelace* obsahuje Spearmanovy korelace abundancí druhů s Froudeho číslem.

```

env$frR

## [1] 0.14 0.18 0.14 0.05 0.05 0.07 0.30 0.43 0.33 0.41 0.22 0.52 0.07 0.18
## [15] 0.30 0.17 0.14 0.24 0.04 0.48 0.51 0.13 0.13 0.56 0.54 0.00 0.02

sort(env$frR)

## [1] 0.00 0.02 0.04 0.05 0.05 0.07 0.07 0.13 0.13 0.14 0.14 0.14 0.17 0.18
## [15] 0.18 0.22 0.24 0.30 0.30 0.33 0.41 0.43 0.48 0.51 0.52 0.54 0.56

rank(env$frR)

## [1] 11.0 14.5 11.0 4.5 4.5 6.5 18.5 22.0 20.0 21.0 16.0 25.0 6.5 14.5
## [15] 18.5 13.0 11.0 17.0 3.0 23.0 24.0 8.5 8.5 27.0 26.0 1.0 2.0

```

Pokud se stane, že se nějaká hodnota v objektu opakuje, její (vlastně jejich) pořadí je průměrem jejich pozice v seřazeném objektu. Například hodnota 0,14 je ve vektoru Froudových čísel třikrát: na první, třetí a sedmnácté pozici. Když vektor seřadíme, dostanou se na pozice 10 - 12. Průměr z hodnot 10, 11 a 12 je 11, proto pořadí (rank) hodnoty 0,14 je 11.

order()

Při řazení objektů nejčastěji sáhneme k funkci `order()`. Ta nám vrátí **pořadí hodnot, jak by měly být** poskládané, **aby byly seřazené** podle velikosti nebo podle abecedy. Toto pořadí následně můžeme použít jako numerický subscript pro seřazení jakýchkoliv i jiných objektů podle dané proměnné, případně proměnných.

```

env$frR

## [1] 0.14 0.18 0.14 0.05 0.05 0.07 0.30 0.43 0.33 0.41 0.22 0.52 0.07 0.18
## [15] 0.30 0.17 0.14 0.24 0.04 0.48 0.51 0.13 0.13 0.56 0.54 0.00 0.02

order(env$frR)

## [1] 26 27 19 4 5 6 13 22 23 1 3 17 16 2 14 11 18 7 15 9 10 8 20
## [24] 21 12 25 24

```

Výsledek funkce `order()` říká: “Aby ten vektor byl popořadě, dej na první pozici 26. hodnotu, pak 27., pak 19., 4., ... a nakonec 24.”. Opravdu, 26. hodnota je nejnižší (0,00), 27. druhá nejnižší (0,02), ... a nakonec 24. nejvyšší (0,56).

Tedy

```

sort(env$frR)

## [1] 0.00 0.02 0.04 0.05 0.05 0.07 0.07 0.13 0.13 0.14 0.14 0.14 0.17 0.18
## [15] 0.18 0.22 0.24 0.30 0.30 0.33 0.41 0.43 0.48 0.51 0.52 0.54 0.56

```

nám vrátí stejný výsledek jako

```

env$frR[order(env$frR)]

## [1] 0.00 0.02 0.04 0.05 0.05 0.07 0.07 0.13 0.13 0.14 0.14 0.14 0.17 0.18
## [15] 0.18 0.22 0.24 0.30 0.30 0.33 0.41 0.43 0.48 0.51 0.52 0.54 0.56

```

Podle jedné proměnné můžeme tedy seřadit proměnnou jinou, třeba abundance druhu *Microtendipes chloris*, nebo řádky dataframu *env*:

```
spe$micrchgr[order(env$frR)]

## [1] 287 126 25 14 30 31 213 8 5 25 31 1 0 31 0 0 0
## [18] 0 9 0 0 0 0 0 0 0 0

env[order(env$frR),]

##      gr_env depth  vel      fr frR
## s26 Ep_CPOM 0.124 0.002 0.001813 0.00
## s27 Ep_CPOM 0.478 0.044 0.020319 0.02
## s19      Ep 0.162 0.050 0.039662 0.04
## s4  Ep_FPOM 0.291 0.078 0.046165 0.05
## s5  Ep_FPOM 0.320 0.092 0.051925 0.05
## s6  Ep_FPOM 0.328 0.126 0.070242 0.07
## s13 Ep_CPOM 0.184 0.088 0.065500 0.07
## s22      Ep 0.340 0.246 0.134698 0.13
## s23      Ep 0.344 0.234 0.127380 0.13
## s1      Ep 0.395 0.274 0.139193 0.14
## s3      Ep 0.496 0.310 0.140536 0.14
## s17      Ep 0.454 0.286 0.135520 0.14
## s16      Ep 0.501 0.372 0.167799 0.17
## s2      Ep 0.422 0.358 0.175951 0.18
## s14      Ep 0.155 0.224 0.181655 0.18
## s11 Er_VEG 0.236 0.334 0.219511 0.22
## s18 Er_VEG 0.200 0.340 0.242733 0.24
## s7      Er 0.278 0.490 0.296715 0.30
## s15 Er_VEG 0.279 0.496 0.299809 0.30
## s9      Er 0.243 0.508 0.329023 0.33
## s10 Er_VEG 0.353 0.758 0.407331 0.41
## s8      Er 0.213 0.618 0.427528 0.43
## s20      Er 0.290 0.804 0.476675 0.48
## s21 Er_VEG 0.138 0.590 0.507082 0.51
## s12 Er_VEG 0.245 0.810 0.522477 0.52
## s25      Er 0.139 0.630 0.539509 0.54
## s24 Er_VEG 0.300 0.958 0.558432 0.56
```

A nemusíme se omezovat na řazení podle jedné proměnné, ale můžeme jich použít, kolik chceme. Třeba seřadíme dataframe *env* podle typu habitatu a uvnitř habitatu podle Froudeho čísla:

```
env[order(env$gr_env, env$frR),]

##      gr_env depth  vel      fr frR
## s19      Ep 0.162 0.050 0.039662 0.04
## s22      Ep 0.340 0.246 0.134698 0.13
## s23      Ep 0.344 0.234 0.127380 0.13
## s1      Ep 0.395 0.274 0.139193 0.14
## s3      Ep 0.496 0.310 0.140536 0.14
## s17      Ep 0.454 0.286 0.135520 0.14
## s16      Ep 0.501 0.372 0.167799 0.17
## s2      Ep 0.422 0.358 0.175951 0.18
```

```
## s14      Ep 0.155 0.224 0.181655 0.18
## s26 Ep_CPOM 0.124 0.002 0.001813 0.00
## s27 Ep_CPOM 0.478 0.044 0.020319 0.02
## s13 Ep_CPOM 0.184 0.088 0.065500 0.07
## s4  Ep_FPOM 0.291 0.078 0.046165 0.05
## s5  Ep_FPOM 0.320 0.092 0.051925 0.05
## s6  Ep_FPOM 0.328 0.126 0.070242 0.07
## s7      Er 0.278 0.490 0.296715 0.30
## s9      Er 0.243 0.508 0.329023 0.33
## s8      Er 0.213 0.618 0.427528 0.43
## s20     Er 0.290 0.804 0.476675 0.48
## s25     Er 0.139 0.630 0.539509 0.54
## s11 Er_VEG 0.236 0.334 0.219511 0.22
## s18 Er_VEG 0.200 0.340 0.242733 0.24
## s15 Er_VEG 0.279 0.496 0.299809 0.30
## s10 Er_VEG 0.353 0.758 0.407331 0.41
## s21 Er_VEG 0.138 0.590 0.507082 0.51
## s12 Er_VEG 0.245 0.810 0.522477 0.52
## s24 Er_VEG 0.300 0.958 0.558432 0.56
```

A samozřejmě lze řadit i sestupně, od největšího po nejmenší. K tomu buď můžeme použít argument `decreasing=` (funguje i u funkce `sort()`), nebo znaménko mínus -. Budu šetřit místo a seřadím abundance druhu *Microtendipes chloris* podle Froudeho čísla, tentokrát sestupně.

```
spe$micrchgr[order(env$frR, decreasing=T)]

## [1] 0 0 0 0 0 0 0 0 0 9 0 0 31 0 0 25 31
## [18] 1 8 5 31 213 14 30 25 126 287

spe$micrchgr[order(-env$frR)]

## [1] 0 0 0 0 0 0 0 0 0 9 0 0 31 0 0 25 31
## [18] 1 8 5 31 213 14 30 25 126 287
```

Výběr “nej” elementů

Řazení můžeme využít i k vybrání “nej” elementů. Jednoduše objekt podle nějaké proměnné seřadíme, a pak z něj vybereme prvních x elementů. Dostaneme tak podsoubor x “nej” elementů. Například se můžeme zeptat na jména deseti nejpočetnějších druhů:

```
names(spe)[order(-colSums(spe))][1:10]

## [1] "Synosemi" "Orthobum" "rheotasp" "micrchgr" "Orthrubi" "tvetbaca"
## [7] "thieGrGe" "tvetdive" "mictrasp" "corysp."

# nebo
names(spe)[order(-colSums(spe))][1:10]

## [1] "Synosemi" "Orthobum" "rheotasp" "micrchgr" "Orthrubi" "tvetbaca"
## [7] "thieGrGe" "tvetdive" "mictrasp" "corysp."
```

Nebo si můžeme vypsát abundance sedmi (aby se mi vešli na stránku) druhů, které nejlíp korelovaly s Froudovým číslem:

```
spe[,order(-abs(korelace))[1:7]]
```

```
##      micrchgr tvetdive tvetbaca cromussp orththie thieGrGe cricbigr
## s1         25         4         4         0         0         5         2
## s2         31         7        11         0         0        15         7
## s3         31         5         8         0         0         8         0
## s4         14         1         2         0         0        17         0
## s5         30         0         2        10         0        59         0
## s6         31         0         0         4         0        61         1
## s7          0         9         5         0         0         5         2
## s8          0         4         0         0         1         0         1
## s9          0         5         4         0         2         1         7
## s10         0        17        25         0         0         0         6
## s11         0        12        11         0         0         1         7
## s12         0        32        38         0         1         1        13
## s13        213         0         2        20         0        72         0
## s14          0         2         4         1         0         3         2
## s15          9        64        83         0         1        17         3
## s16          0         1         2         0         0         1         2
## s17          1         0         0         0         1         4         0
## s18          0        53        86         0         1         3         9
## s19         25         2         3         1         0        14         0
## s20          0        41        25         0         0        13         0
## s21          0        87       127         0         3        10         2
## s22          8         5         2         6         0        13         0
## s23          5         0         0         7         0         6         0
## s24          0       116       177         0         2         6         3
## s25          0         10        10         0         1         1         1
## s26       287         1         1         5         0       127         0
## s27       126         2         0         2         0        32         0
```

Abychom v tom něco dokázali vykukat, můžeme seřadit řádky podle Froudeho čísla:

```
spe[order(env$fr),order(-abs(korelace))[1:7]]
```

```
##      micrchgr tvetdive tvetbaca cromussp orththie thieGrGe cricbigr
## s26       287         1         1         5         0       127         0
## s27       126         2         0         2         0        32         0
## s19         25         2         3         1         0        14         0
## s4         14         1         2         0         0        17         0
## s5         30         0         2        10         0        59         0
## s13        213         0         2        20         0        72         0
## s6         31         0         0         4         0        61         1
## s23         5         0         0         7         0         6         0
## s22         8         5         2         6         0        13         0
## s17         1         0         0         0         1         4         0
## s1         25         4         4         0         0         5         2
## s3         31         5         8         0         0         8         0
## s16         0         1         2         0         0         1         2
## s2         31         7        11         0         0        15         7
## s14         0         2         4         1         0         3         2
## s11         0        12        11         0         0         1         7
```

```
## s18      0      53      86      0      1      3      9
## s7       0       9       5      0      0      5      2
## s15      9      64      83      0      1     17      3
## s9       0       5       4      0      2      1      7
## s10      0      17      25      0      0      0      6
## s8       0       4       0      0      1      0      1
## s20      0      41      25      0      0     13      0
## s21      0      87     127      0      3     10      2
## s12      0      32      38      0      1      1     13
## s25      0      10      10      0      1      1      1
## s24      0     116     177      0      2      6      3
```

A sloupce podle odpovědi druhů (jestli jeho abundance s Froudeho číslem klesá nebo stoupá):

```
# to je lepsi rozdelit do nekolika kroku.
# nejprve omezime dataframe spe na 7 druhu, ktere nejlip koreluji a seradime jeho radky podle Froudeho
spe.vyb<- spe[order(env$fr),order(-abs(korelace))[1:7]]
# ted stejne omezime vektor korelace a logicky vektor odlisujici druhy s kladnou a zapornou korelaci
kor.vyb<- korelace[order(-abs(korelace))[1:7]]
vetsi.vyb<- (korelace>0)[order(-abs(korelace))[1:7]]
# a nakonec omezeny dataframe seradime podle omezenych vektoru:
spe.vyb[, order(vetsi.vyb, -abs(kor.vyb))]
```

```
##      micrchgr cromussp thieGrGe tvetdive tvetbaca orththie cricbigr
## s26      287         5      127         1         1         0         0
## s27      126         2       32         2         0         0         0
## s19       25         1       14         2         3         0         0
## s4        14         0       17         1         2         0         0
## s5        30        10       59         0         2         0         0
## s13      213        20       72         0         2         0         0
## s6        31         4       61         0         0         0         1
## s23        5         7        6         0         0         0         0
## s22        8         6       13         5         2         0         0
## s17        1         0        4         0         0         1         0
## s1       25         0        5         4         4         0         2
## s3       31         0        8         5         8         0         0
## s16        0         0         1         1         2         0         2
## s2       31         0       15         7        11         0         7
## s14        0         1         3         2         4         0         2
## s11        0         0         1        12        11         0         7
## s18        0         0         3        53        86         1         9
## s7         0         0         5         9         5         0         2
## s15        9         0       17        64        83         1         3
## s9         0         0         1         5         4         2         7
## s10        0         0         0       17        25         0         6
## s8         0         0         0         4         0         1         1
## s20        0         0       13        41        25         0         0
## s21        0         0       10        87       127         3         2
## s12        0         0         1       32        38         1        13
## s25        0         0         1       10        10         1         1
## s24        0         0         6     116       177         2         3
```