

M0130 – 2. PRAKTIKUM : M0130pr02 (*Práce s daty v R*)

A. Načtení vstupních dat:

I. Čtení dat pomocí příkazu `scan`PŘÍKLAD 1

O městě Veselí máme za roky 1967–2005 k dispozici následující údaje

sňatky 65, 94, 79, 84, 84, 70, 98, 86, 80, 88, 82, 83, 87, 89, 99, 88, 92, 81, 106, 103
76, 102, 111, 109, 81, 83, 78, 66, 59, 55, 76, 80, 72, 52, 47, 53, 59, 51, 45
rozvody 7, 11, 13, 19, 17, 24, 18, 13, 31, 18, 26, 28, 25, 31, 27, 32, 31, 28, 25, 31
23, 31, 24, 37, 16, 20, 42, 18, 46, 41, 37, 43, 23, 23, 27, 23, 30, 41, 35

Nejprve zkopírujeme do schránky první řádek týkající se sňatků. Pomocí funkce `scan` načteme data ze schránky do proměnné `snatky1`.

```
> snatky1 <- scan(file = "clipboard", sep = ",")
```

Obdobně zkopírujeme do schránky data, která se týkají druhého řádku, opět použijeme funkci `scan`.

```
> snatky2 <- scan(file = "clipboard", sep = ",")
```

Stejně pokračujeme i v případě rozvodů, tj. nejprve do schránky zkopírujeme první řádek dat pro rozvody, následně použijeme funkci `scan`.

```
> rozvody1 <- scan(file = "clipboard", sep = ",")
```

Nakonec do schránky zkopírujeme druhý řádek vztahující se k rozvodům, pak použijeme funkci `scan`.

```
> rozvody2 <- scan(file = "clipboard", sep = ",")
```

Nyní vytvoříme proměnné `snatky` a `rozvody`.

```
> snatky <- c(snatky1, snatky2)
> rozvody <- c(rozvody1, rozvody2)
```

Z proměnných `snatky` a `rozvody` vytvoříme datový rámeček (datovou tabulku, datový soubor) pomocí `data.frame`. Jeho strukturu zjistíme pomocí příkazu `str` a vypíšeme si jeho začátek a konec.

```
> Veseli <- data.frame(rok = 1967:2005, snatky, rozvody)
> str(Veseli)
```

```
,data.frame,: 39 obs. of 3 variables:
 $ rok      : int  1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 ...
 $ snatky   : num  65 94 79 84 84 70 98 86 80 88 ...
 $ rozvody  : num   7 11 13 19 17 24 18 13 31 18 ...
```

```
> head(Veseli)
```

```
   rok snatky rozvody
1 1967     65        7
2 1968     94       11
3 1969     79       13
4 1970     84       19
5 1971     84       17
6 1972     70       24
```

```
> tail(Veseli)
```

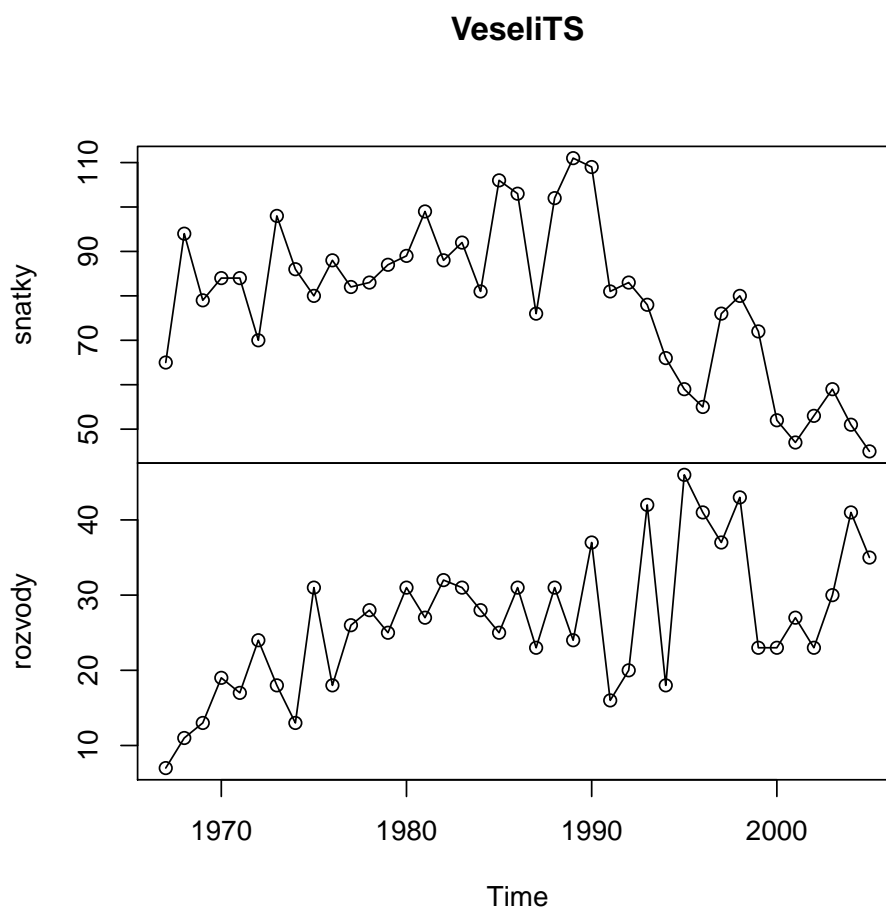
	rok	snatky	rozvody
34	2000	52	23
35	2001	47	27
36	2002	53	23
37	2003	59	30
38	2004	51	41
39	2005	45	35

Při práci s časovými řadami je lepší využívat jiný datový typ (*time series*), který lze vytvořit pomocí příkazu `ts`. Přes příkaz `str` si prohlédneme strukturu a pomocí příkazu `plot` si obě časové řady graficky znázorníme.

```
> VeseliTS <- ts(Veseli[, -1], start = 1967, frequency = 1)
> str(VeseliTS)
```

```
mts [1:39, 1:2] 65 94 79 84 84 70 98 86 80 88 ...
- attr(*, "dimnames")=List of 2
..$ : NULL
..$ : chr [1:2] "snatky" "rozvody"
- attr(*, "tsp")= num [1:3] 1967 2005 1
- attr(*, "class")= chr [1:2] "mts" "ts"
```

```
> plot(VeseliTS, type = "o")
```



Obrázek 1: Počty sňatků a rozvodů ve městě Veselí v letech 1967–2005

II. Čtení dat pomocí příkazu `read.table`

PŘÍKLAD 2

Načteme datový soubor pomocí příkazu `read.table`. Vzhledem k tomu, že datový soubor v prvním řádku obsahuje názvy proměnných, v příkazu `read.table` nesmíme zapomenout nastavit `header=TRUE`.

```
> fileDat <- paste(data.library, "MorKrasNavstevnost.txt", sep = "")
> MorKras <- read.table(fileDat, header = TRUE)
```

Příkazem `str` vypíšeme strukturu datového rámce, příkazem `header` se vypíše prvních šest řádků a příkazem `tail` posledních šest řádků.

```
> str(MorKras)

,data.frame,: 34 obs. of 6 variables:
 $ Rok          : int  1999 1999 1999 1999 1999 1999 1999 1999 1999 ...
 $ Mesic        : Factor w/ 12 levels "březen","červen",...: 6 11 1 4 5 2 3 10 12 9 ...
 $ Punkevni     : int  2002 1694 4927 14186 29207 32158 39025 38629 19381 13625 ...
 $ SloupskoSosuvske: int  0 240 314 1498 4136 5343 12306 11725 2993 1629 ...
 $ Katerinska   : int  0 292 508 2141 7510 8822 18225 19681 3669 2212 ...
 $ Balcarka     : int  0 198 231 1265 4310 7445 10353 11243 1752 1400 ...
```

```
> head(MorKras)
```

	Rok	Mesic	Punkevni	SloupskoSosuvske	Katerinska	Balcarka
1	1999	leden	2002	0	0	0
2	1999	únor	1694	240	292	198
3	1999	březen	4927	314	508	231
4	1999	duben	14186	1498	2141	1265
5	1999	květen	29207	4136	7510	4310
6	1999	červen	32158	5343	8822	7445

```
> tail(MorKras)
```

	Rok	Mesic	Punkevni	SloupskoSosuvske	Katerinska	Balcarka
29	2001	květen	33353	6731	8421	5031
30	2001	červen	34478	8183	8786	5486
31	2001	červenec	40996	13478	17030	8505
32	2001	srpen	40223	13911	15881	7909
33	2001	září	20858	4685	4288	1694
34	2001	říjen	12614	2017	2410	806

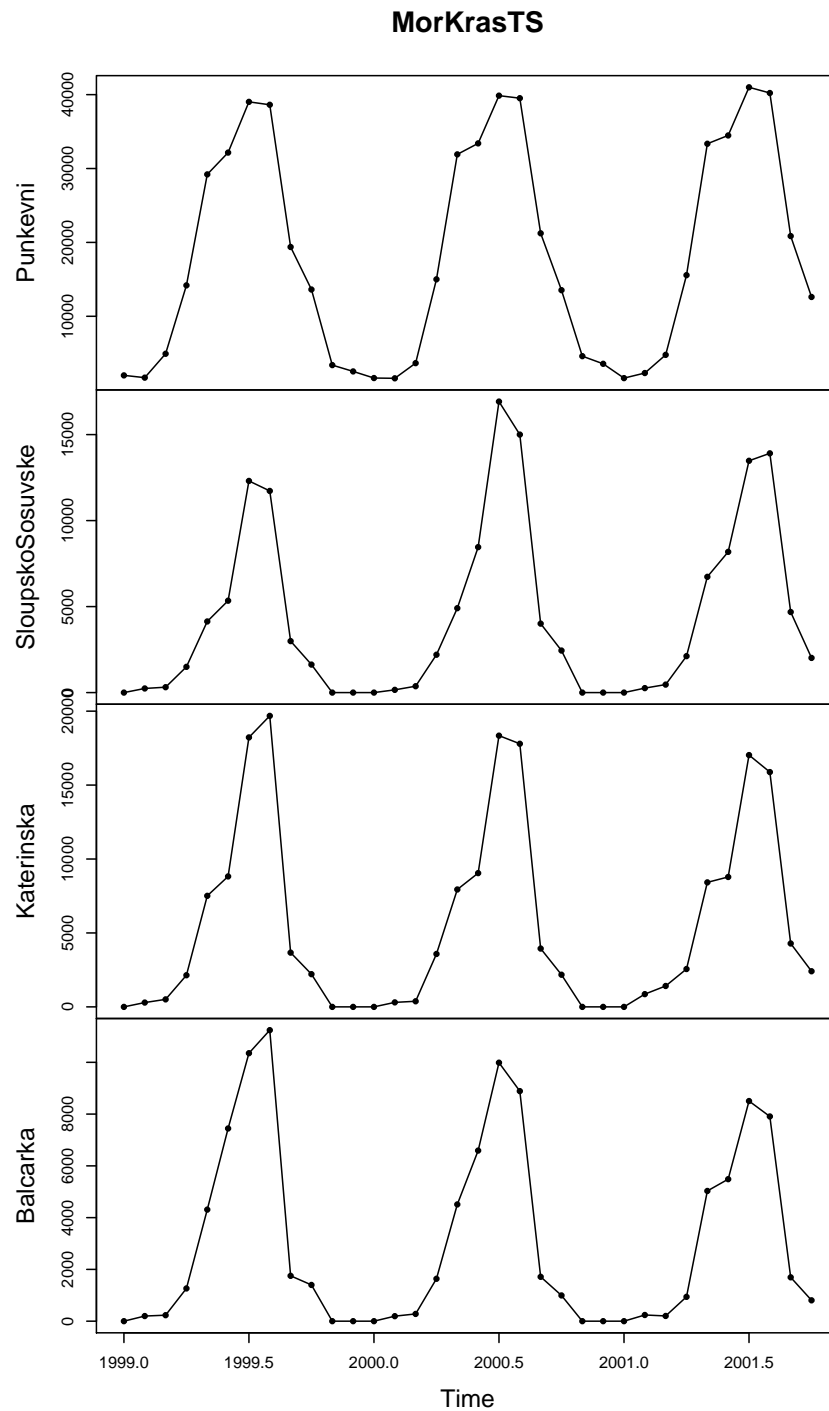
Vedle datového rámce vytvoříme i vícerozměrnou časovou řadu, přes příkaz `str` se podíváme na její strukturu.

```
> MorKrasTS <- ts(MorKras[, -(1:2)], start = 1999, frequency = 12)
> str(MorKrasTS)
```

```
int [1:34, 1:4] 2002 1694 4927 14186 29207 32158 39025 38629 19381 13625 ...
- attr(*, "dimnames")=List of 2
..$ : NULL
..$ : chr [1:4] "Punkevni" "SloupskoSosuvske" "Katerinska" "Balcarka"
- attr(*, "tsp")= num [1:3] 1999 2002 12
- attr(*, "class")= chr [1:2] "mts" "ts"
```

Časovou řadu opět vykreslíme pomocí příkazu `plot`.

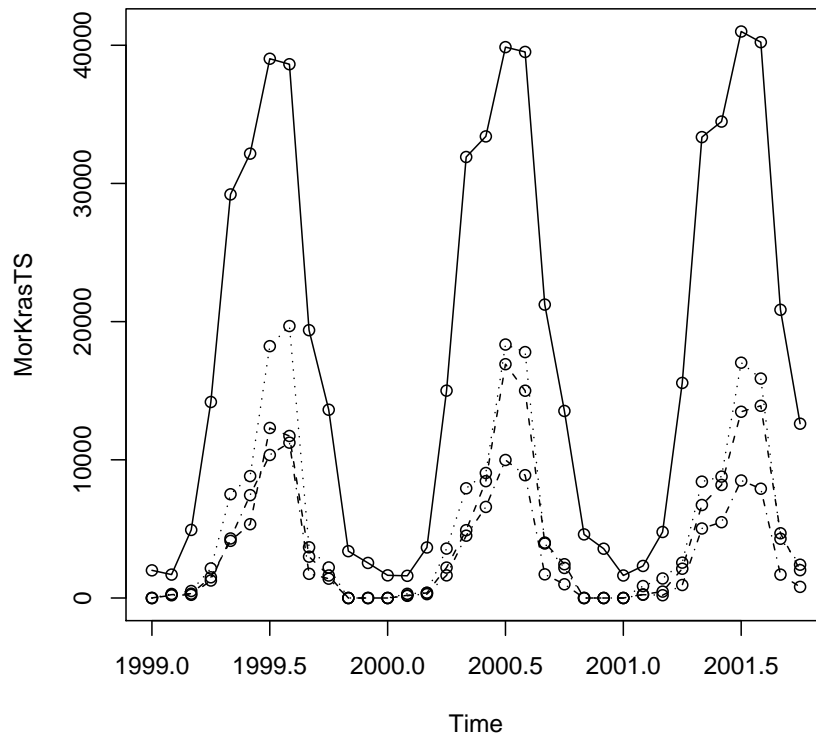
```
> plot(MorKrasTS, type = "o", pch = 20, cex = 3)
```



Obrázek 2: Počty návštěvníků ve vybraných jeskyních Moravského Krasu v letech 1999–2001

Pokud bychom chtěli vykreslit všechny časové řady do jediného panelu, použijeme příkaz

```
> k <- dim(MorKrasTS) [2]
> plot(MorKrasTS, type = "o", plot.type = "single", lty = 1:k)
```



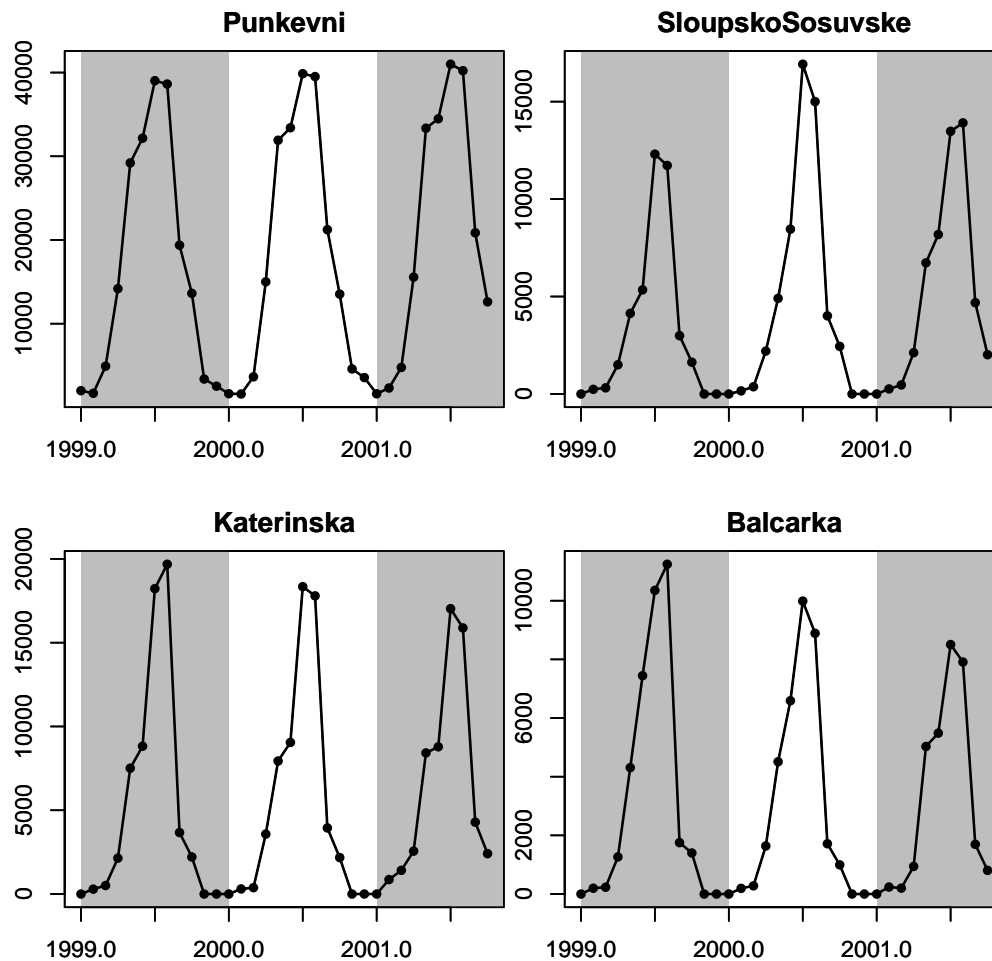
Obrázek 3: Počty návštěvníků v Moravském Krasu v letech 1999–2001 – jediný panel

Abychom mohli používat graf, ve kterém budou zřetelně odděleny roční bloky dat, nejprve vytvoříme funkci

```
> plot.band <- function(x, lwd = 1.5, type = "o", pch = 20, ...) {
  plot(x, ...)
  a <- time(x)
  i1 <- floor(min(a))
  i2 <- ceiling(max(a))
  y1 <- par("usr")[3]
  y2 <- par("usr")[4]
  if (par("ylog")) {
    y1 <- 10^y1
    y2 <- 10^y2
  }
  for (i in seq(from = i1, to = i2 - 1, by = 2)) {
    polygon(c(i, i + 1, i + 1, i), c(y1, y1, y2, y2), col = "grey",
            border = NA)
  }
  par(new = T)
  plot(x, lwd = lwd, type = type, pch = pch, ...)
}
```

Tuto funkci pak použijeme pro grafické oddělení jednotlivých let.

```
> k <- dim(MorKrasTS)[2]
> par(mfrow = c(2, 2), mar = c(3, 2, 2, 0) + 0.1)
> for (ik in 1:k) plot.band(MorKrasTS[, ik], ylab = "", main = colnames(MorKrasTS)[ik])
```



Obrázek 4: Počty návštěvníků v Moravském Krasu v letech 1999–2001 – roční bloky

B. Úkol:

Na internetu najdete několik časových řad (roční, měsíční, čtvrtletní). Načtěte je a vykreslete.