Úvod do předmětu Aplikovaná statistika I

- **Přehled základních matematických objektů**
  - **Proměnná** = označení objektu (číslo, vektor, matice, tabulka, funkce)
    - Číslo: a $\leftarrow$ 3
    - Vektor: **a** $\leftarrow$ (3, 6, 9)
    - Matice: $\mathbf{A} \leftarrow \begin{pmatrix} 2 & 3 & 7 \\ 8 & 4 & 5 \end{pmatrix}$
    - Tabulka: **Tab** $\leftarrow$

	s_1	s_2	s_3
r_1	2	3	7
r_2	8	4	5

    - Funkce: **funkce()**
      - Objekt, do kterého vložíme vstup (IN) a získáme výstup (OUT)
      - **sum**(vektor) $\rightarrow$ číslo
      - **matrix**(vektor, 2, 2) $\rightarrow$ matice
      - **mean**(matice) $\rightarrow$ číslo

- **Úvod do syntaxe programovacího jazyka**
  - Jazyk je tzv. **case sensitive** $\rightarrow$ názvy sum, Sum, SUM, sUm, sUM, ... znační různé objekty
  - **Vytvoření proměnné**
    - Číslo:

```
a <- 3
```

    - Vektor:

```
aa <- c(1.2, 5.3, 6.4)
```

    - Matice:

```
A <- matrix(c(1, 2, 3, 4, 5, 6), nrow = 2, ncol = 3, byrow = T)
```

    - Datová tabulka:

```
tab <- data.frame(A, row.names = c('r1', 'r2'))
names(tab) <- c('s1', 's2', 's3')
```

  - **Základní operace**
    - +, -, *, /

```
3 + 2 - 6 * 9 / (5 + 7)
```

    - Operace s čísly

```
a <- 25
b <- 5
a / b
```

    - Operace s vektory

```
x <- c(1, 2, 3)
y <- c(3, 2, 1)
x - y
x + y

z <- c(0, 1, 2, 3)
x + y + z
# !Pozor, vektor z je delsi nez vektory x a y.
#  R sice napise varovnou hlasku, ale vypocet i tak provede!
```

* Operace s maticemi

```
B <- matrix(c(1, 1, 1, 1, 1, 1), nrow = 2, ncol = 3)
A - B

C <- matrix(c(1, 1, 1, 1, 1, 1), nrow = 3, ncol = 2)
A - C
# !Matice A ma rozmer 2x3, matice C ma rozmer 3x2.
#  Nyni jiz R vypocet neprovede, pouze zahlasi chybu.
```

– **Pokročilé operace**

* Délka vektoru a dimenze matice

```
length(z)
dim(A)
```

* Minimální a maximální hodnota vektoru a matice

```
min(z)
max(A)
```

* Součet hodnot a aritmetický průměr vektoru a matice

```
sum(z)
mean(A)
```

* Mocniny a odmocniny

```
(odmocnina <- sqrt(2))
odmocnina ^ 2
```

* Zakrouhlování

```
round(odmocnina, digits = 3)
```

– **Vytváření posloupností**

* Posloupnosti čísel se vzdáleností 1

```
(x <- 1:10)
(y <- 50:55)
```

* Posloupnosti čísel s libovolnou ale ekvidistantní vzdáleností

```
(pst <- seq(from = 0, to = 1, length = 12))
(pst2 <- seq(from = 0, to = 1, by = 0.09))
```

* Posloupnosti opakujících se čísel

```
vaha <- c(58, 61, 57, 59, 60, 54, 64, 71, 66, 70)
divky <- rep(1, 6)
chlapci <- rep(2, 4)
(pohlavi <- c(divky, chlapci))
```

* Přídání nového řádku (resp. sloupce) ke stávajícímu vektoru

```
(data.r <- rbind(vaha, pohlavi))
(data.c <- cbind(vaha, pohlavi))
```

* Přídání nového řádku (resp. sloupce) ke stávající matici

```
diabetes <- rep(0, 10)
(data.r2 <- rbind(data.r, diabetes))
(data.r3 <- cbind(data.r2, c(62, 2, 1)))
```

– **Podmnožiny vektorů a matic**
  * Výběr konkrétních hodnot z vektoru

```
vaha[c(2, 3, 4)]
vaha[2:4]
```

  * Výběr konkrétních řádků z matice

```
data.r3[1, ] # vyber prvniho radku
data.r3[2, ] # vyber druheho radku
```

  * Výběr konkrétních sloupců z matice

```
data.r3[ , 4]
data.r3[ , c(3, 5, 8)]
data.r3[1:2, 5:7]
```

– **Práce s datovým souborem**
  * Zjištění absolutní cesty k aktuální složce a výpis všech souborů z této složky

```
getwd()
dir()
```

  * Načtení datového souboru a vypsání prvních tří řádků z datového souboru (tabulky)

```
data <- read.delim('Zaznam teploty.txt', sep = '\t', dec = '.')
```

```
head(data, n = 3)
```

  * Výběr konkrétních řádků z tabulky

```
data[4:6, ]
```

  * Výběr konkrétních sloupců z tabulky (více možností)

```
data[ , 1]
data[ , 'Hodina']
data$Hodina
subset(data, select = Hodina)

hodiny  <- data$Hodina
teploty <- data$Teplota
```

– **Logické operátory**
  * Rovnost ==, větší >, menší <, větší nebo rovno >=, menší nebo rovno <=

```
(teploty == 37.7) * 1
sum((teploty == 37.7) * 1)

(teploty < 37.7) * 1
(teploty <= 37.7) * 1
(teploty > 37.7) * 1
```

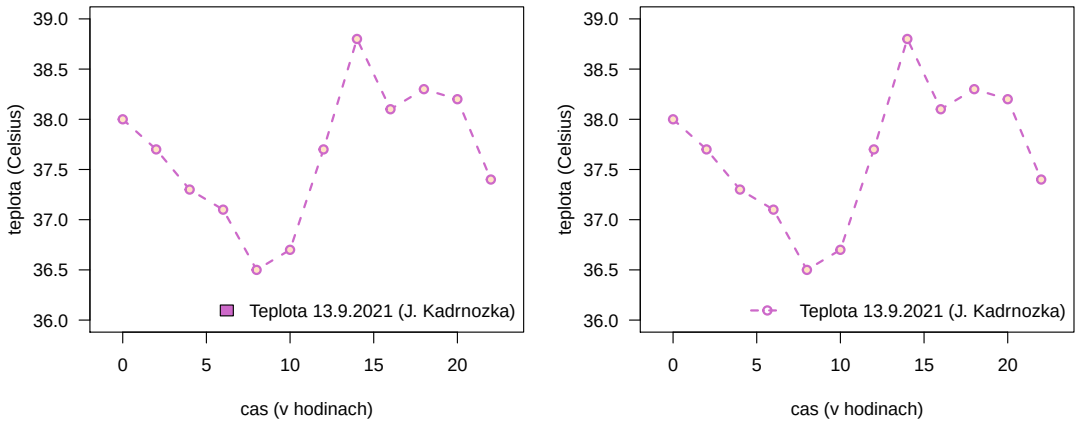
  * Výběr řádků s konkrétní vlastností

```
data[data$Teplota == 37.7, ]
data[data$Teplota == 37.7, 'Hodina']
subset(data, subset = Teplota == 37.7, select = Hodina)
```

– **Tvorba základních grafů příkazem plot(x, y)**

```
par(mar = c(4, 4, 1, 1))
plot(hodiny, teploty,
     main = '', xlab = 'cas (v hodinach)', ylab = 'teplota (Celsius)',
     col  = 'orchid3', type = 'o', pch  = 21, lwd  = 2, lty  = 2, bg = 'bisque',
     xlim = c(-1, 23), ylim = c(36, 39), las = 1)
```

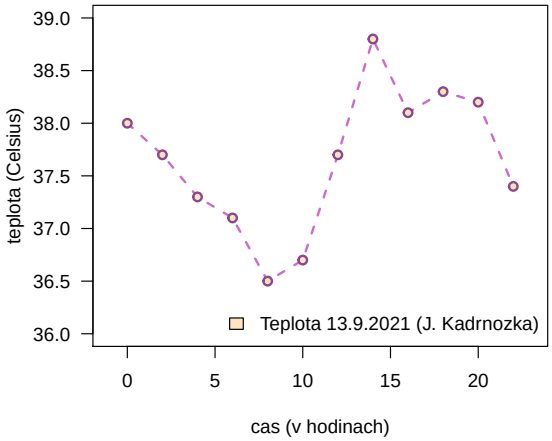
```r
legend('bottomright', fill = c('orchid3'), legend = c('Teplota 13.9.2021 (J. Kadrnozka)'),
       bty = 'n')

legend('bottomright', col = c('orchid3'), pch = c(21), lwd = c(2), lty = c(2),
       legend = c('Jonatan Kadrnozka'), bty = 'n')
```



– **Export grafů do png souboru**

  * **Ruční export: Multifunkční okno → Plots → Export → Save as Image → Maintain aspect ratio: odškrtnout → Save**
  * **Export posloupností příkazů:**

```r
png('Nazev grafu.png')
plot(1:5, 1:5)
dev.off() # prikaz dev.off() projizdime, dokud se v konzoli neobjevi hlaska null device 1
```

Tip na domácí procvičení: Vytvoření analogického grafu příkazy plot(), lines() a points()

```r
par(mar = c(4, 4, 1, 1))
plot(hodiny, teploty, main = '', xlab = 'cas (v hodinach)',
     ylab = 'teplota (Celsius)', type = 'n', xlim = c(-1, 23), ylim = c(36, 39), las = 1)
lines(hodiny, teploty, lwd = 2, lty = 2, col = 'orchid3')
points(hodiny, teploty, pch = 21, col = 'orchid4', bg = 'bisque', lwd = 2)
legend('bottomright', fill = 'bisque', legend = c('Teplota 13.9.2021 (J. Kadrnozka)'),
       bty = 'n')
```