

# Unos podataka i struktura objekata

Lucija Kanjer

11.09.2024

## Unos podataka u R

Podatke možete unositi ručno ili iz različitih izvora, kao što su: tekstualne datoteke, CSV datoteke, Excel, baze podataka ili izravno s interneta.

### 1. Struktura osnovnih objekata u R

```
# Varijable su osnovni objekti u R-u.
a <- 5           # integer - cijeli broj, dodati "L"
x <- 3.14        # numeric - decimalni broj (realni)
y <- "DNA"       # character - tekst/string
z <- TRUE        # logical - logička vrijednost (TRUE/FALSE)

# Provjera tipa podataka
class(a)
```

#### 1.1 Kreiranje varijabli

```
## [1] "numeric"
```

**1.2 Kreiranje vektora** Vektor je niz elemenata iste vrste (npr. svi brojevi ili svi tekstovi/stringovi).

```
geni <- c("BRCA1", "TP53", "MYC") # vektor gena
duljine <- c(1863, 1179, 1584)    # vektor duljina gena u parovima baza (bp)
```

**1.3 Kreiranje matrica** Matrica je 2D struktura podataka iste vrste (redovi x stupci). Primjer: Matrica ekspresije gena.

```
ekspresija <- matrix(c(2.1, 3.5, 5.0, 1.8, 4.3, 6.2), nrow = 2, byrow = TRUE)
rownames(ekspresija) <- c("Uzorak1", "Uzorak2")
colnames(ekspresija) <- c("BRCA1", "TP53", "MYC")
print(ekspresija)
```

```
##           BRCA1 TP53 MYC
## Uzorak1    2.1  3.5 5.0
## Uzorak2    1.8  4.3 6.2
```

**1.4 Data frame** Data frame je struktura podataka koja može sadržavati vektore varijabli različitih vrsta (brojevi, tekst, logičke vrijednosti)

```
geni_df <- data.frame(  
  Gen = c("BRCA1", "TP53", "MYC"),  
  Duljina_bp = c(1863, 1179, 1584),  
  Funkcija = c("tumor suppressor", "tumor suppressor", "oncogene")  
)  
geni_df # ispis u konzoli
```

```
##      Gen Duljina_bp      Funkcija  
## 1 BRCA1      1863 tumor suppressor  
## 2 TP53      1179 tumor suppressor  
## 3 MYC      1584      oncogene
```

```
print(geni_df) # ispis u konzoli
```

```
##      Gen Duljina_bp      Funkcija  
## 1 BRCA1      1863 tumor suppressor  
## 2 TP53      1179 tumor suppressor  
## 3 MYC      1584      oncogene
```

```
View(geni_df) # vizualizacija tablice u novom listu
```

## 2. Unos podataka u R

**2.1 Provjeravanje i postavljanje radnog direktorija** Provjera trenutnog radnog direktorija (mape) moguća je izvršavanjem naredbe `getwd()`

```
getwd()
```

```
## [1] "C:/Users/lucij/Documents/APUBI/rmd"
```

Postavljanje novog radnog direktorija (potrebno navesti put do mape) Primjer (Windows): `setwd("C:/Users/ImeRacunala/Documents")`  
Primjer (Mac/Linux): `setwd("/Users/ImeRacunala/Documents")`

```
# setwd("C:/Users/lucija/Documents")
```

**2.2 Učitavanje podataka iz tekstualnih datoteka** Učitavanje podataka iz tekstualnih datoteka i CSV-a.

Kreirajte tekstualnu datoteku "biljke.txt" za 5 jedinki biljaka, sa varijablama za visinu, širinu i vrstu. Redovi su jedinke/uzorci, a stupci su varijable. Između stupaca upisujete razmak "Tab".

Učitavanje podataka iz tekstualne datoteke:

```
#data_txt <- read.table("biljke.txt", header = TRUE)  
#data_txt
```

**2.3 Učitavanje podataka iz CSV datoteke** Kreirajte CSV datoteku "zivotinje.csv" sa sličnim sadržajem.

```
#data_csv <- read.csv("zivotinje.csv", header = TRUE)
#data_csv
```

## 2.4 Učitavanje podataka iz Excel datoteka Instalacija paketa “readxl” ako nije već instaliran.

```
# install.packages("readxl")
#library(readxl)
#palmerpenguins_original <- read_excel() # nadopisati
#View(palmerpenguins_original)
```

```
library(palmerpenguins)
data("penguins")

# Pregled prvih nekoliko redaka podataka
head(penguins)
```

## 2.5 Učitavanje podataka iz paketa

```
##   species   island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
## 1  Adelie Torgersen      39.1         18.7           181          3750
## 2  Adelie Torgersen      39.5         17.4           186          3800
## 3  Adelie Torgersen      40.3         18.0           195          3250
## 4  Adelie Torgersen      NA           NA           NA           NA
## 5  Adelie Torgersen      36.7         19.3           193          3450
## 6  Adelie Torgersen      39.3         20.6           190          3650
##      sex year
## 1   male 2007
## 2 female 2007
## 3 female 2007
## 4   <NA> 2007
## 5 female 2007
## 6   male 2007
```

```
# Tipovi podataka - provjera strukture
str(penguins)
```

```
## 'data.frame':   344 obs. of  8 variables:
## $ species      : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ island       : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ bill_length_mm : num  39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
## $ bill_depth_mm : num  18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
## $ flipper_length_mm: int  181 186 195 NA 193 190 181 195 193 190 ...
## $ body_mass_g    : int  3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
## $ sex           : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
## $ year          : int   2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

```
# Pristupanje specifičnim stupcima i redovima
vrsta <- penguins$species # stupac
prvi_redak <- penguins[1, ] # redak
```

## Zadaci

1. Učitajte podatke o pingvinima i provjerite strukturu podataka.
2. Pristupite kolonama `species` i `island` te ispišite prvi redak podataka.
3. Kreirajte vektor s 5 gena i njihovim duljinama (u baznim parovima).
4. Učitajte CSV datoteku koja sadrži podatke o ekspresiji gena.
5. Ispitajte strukturu učitano`g` data frame-a koristeći funkciju `str()`.
6. Dodajte novi redak u data frame s informacijama o novom genu.
7. Ispisujte prva 3 retka i zadnja 3 retka učitano`g` data frame-a koristeći funkcije `head()` i `tail()`.