

Unos podataka u programsko okruženje R i struktura objekata

Lucija Kanjer, e-mail: lucija.kanjer@biol.pmf.hr

2024-09-30

Sadržaj praktikuma

- Uvod u rad u programskom okruženju R i osnovne funkcije, instaliranje programskih paketa
- *Unos podataka u programsko okruženje R, struktura objekata*
- Rad s objektima i podacima te definiranje bioloških varijabli u R-u
- Grafički prikaz bioloških podataka i testiranje razdiobe podataka u R-u
- Primjeri osnovnih statističkih analiza kategoričkih i numeričkih varijabli u biološkim istraživanjima u R-u
- Regresije i korelacije, linearni modeli bioloških podataka – primjeri u R-u
- Primjena parametrijskih statističkih testova bioloških podataka u R-u
- Primjena neparametrijskih statističkih testova bioloških podataka u R-u
- Primjeri multivarijatnih analize bioloških podataka u R-u - linearni modeli, klaster analize i ordinacijske analize

Sadržaj današnje vježbe

- Kreiranje objekata u R-u - ručni unos
- Varijable, vektori, matrice i podatkovni okviri (*data frame*)
- Provjera tipa podataka
- Provjera strukture objekata
- Učitavanje datoteka u R (tekstualne, CSV i Excel datoteke)
- Pristupanje pojedinačnim stupcima i retcima iz tablica
- Izvoz R objekata u datoteke

Ručni unos objekata u R-u

Variable

```
### Kreiranje objekata u R-u - ručni unos
```

```
# 1. Variable
```

```
x <- 3.14          # numeric - brojčana vrijednost
```

```
y <- "DNA"         # character - tekst (mora ići u navodnike)
```

```
z <- TRUE          # logical - logička vrijednost (TRUE/FALSE)
```

Provjera tipa i strukture podataka

```
# Provjera tipa podataka  
class(x) # prikazuje samo tip varijable
```

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

```
str(x) # struktura: prikazuje tip varijable i što ona sadrži
```

```
## num 3.14
```

Provjerite tip podataka varijabli y i z!

```
# Provjerite tip podataka varijabli y i z!  
class(y)
```

```
## [1] "character"
```

```
class(z)
```

```
## [1] "logical"
```

Vektori

Vektor je niz elemenata iste vrste (npr. svi brojevi ili svi tekstovi).

```
# 2. Vektori
```

```
geni <- c("BRCA1", "TP53", "MYC") # tekstualni vektor gena  
print(geni)
```

```
## [1] "BRCA1" "TP53"  "MYC"
```

```
duljine <- c(1863, 1179, 1584) # numerički vektor duljina gena u parovima baza (bp)  
print(duljine)
```

```
## [1] 1863 1179 1584
```

Matrice

Matrica je 2D struktura podataka iste vrste (redovi x stupci).

```
# 3. Matrice - skup vektora istog tipa
# Definiranje matrice
matrix_mikrobi <- matrix(c(85, 47, 37, 10, 65, 50, 28, 8),
                        nrow = 4, # postavljanje broja redaka u matrici
                        byrow = TRUE) # definiranje popunjavanja matrice po retcima

# Dodavanje naziva redaka i stupaca
rownames(matrix_mikrobi) <- c("Bakterija1", "Bakterija2", "Gljivica1", "Gljivica2")
colnames(matrix_mikrobi) <- c("CFU_kuhinja", "CFU_kupaonica")
```



```
print(matrix_mikrobi)
```

```
##           CFU_kuhinja CFU_kupaonica
## Bakterija1          85           47
## Bakterija2          37           10
## Gljivica1           65           50
## Gljivica2           28            8
```

Podatkovni okvir (*data frame*)

Data frame je struktura podataka koja može sadržavati vektore varijabli raličitih vrsta (brojevi, tekst, logičke vrijednosti).

4. Podatkovni okvir (data frame) - skup vektora različitih tipova

```
df_mikrobi <- data.frame(  
  CFU = c(85, 47, 37, 10, 65, 50, 28, 8), # numeručki vektor  
  mikrorganizam = c("Bakterija", "Bakterija", "Gljivica", "Gljivica"), # tekstualni vektor  
  lokacija = c("kuhinja", "kupaonica", "kuhinja", "kupaonica") # tekstualni vektor  
)
```

```
# Uvid u data frame  
print(df_mikrobi) # ispis u konzoli
```

```
##   CFU mikrorganizam  lokacija  
## 1  85      Bakterija  kuhinja  
## 2  47      Bakterija  kupaonica  
## 3  37      Gljivica   kuhinja  
## 4  10      Gljivica   kupaonica  
## 5  65      Bakterija  kuhinja  
## 6  50      Bakterija  kupaonica  
## 7  28      Gljivica   kuhinja  
## 8   8      Gljivica   kupaonica
```

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

	CFU	mikroorganizam	lokacija
1	85	Bakterija	kuhinja
2	47	Bakterija	kupaonica
3	37	Gljivica	kuhinja
4	10	Gljivica	kupaonica
5	65	Bakterija	kuhinja
6	50	Bakterija	kupaonica
7	28	Gljivica	kuhinja
8	8	Gljivica	kupaonica

Showing 1 to 8 of 8 entries, 3 total columns

Učitavanje podataka iz datoteka

```
### Učitavanje datoteka
```

```
# Podsjetimo se naredbi za provjeru i postavljanje radnog direktorija!
```

```
getwd()
```

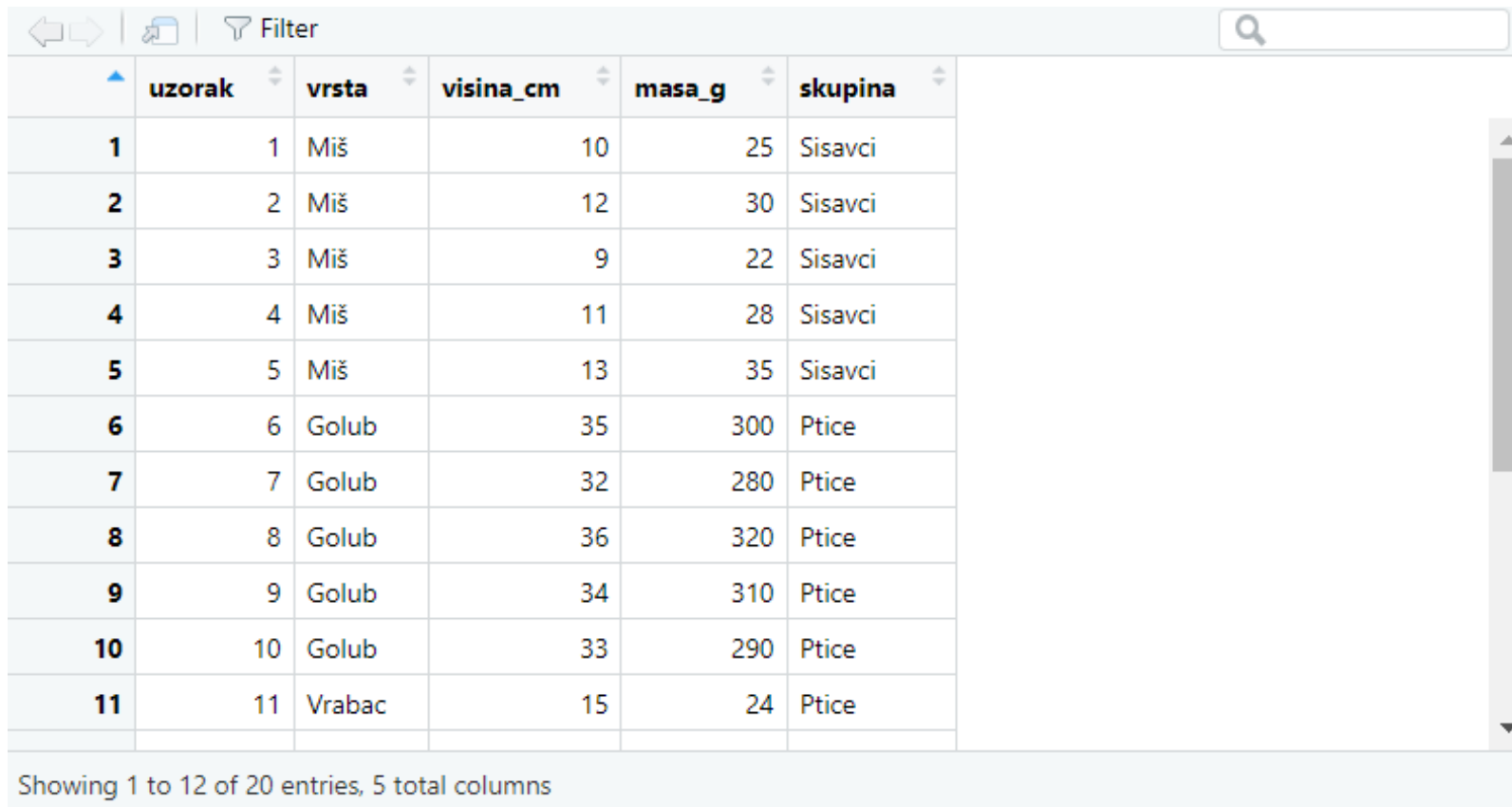
```
## [1] "C:/Users/lucij/Documents/APUBI/02_Unos_podataka"
```

```
setwd("C:/Users/lucij/Documents/APUBI/02_Unos_podataka") #izmjeniti za vaše računalo!
```

Učitavanje tekstualnih datoteka

```
# Učitavanje tekstualnih datoteka  
zivotinje <- read.table("zivotinje/zivotinje.txt",  
                        header = TRUE)
```

View(zivotinje) # ili klik na objekt u environmentu



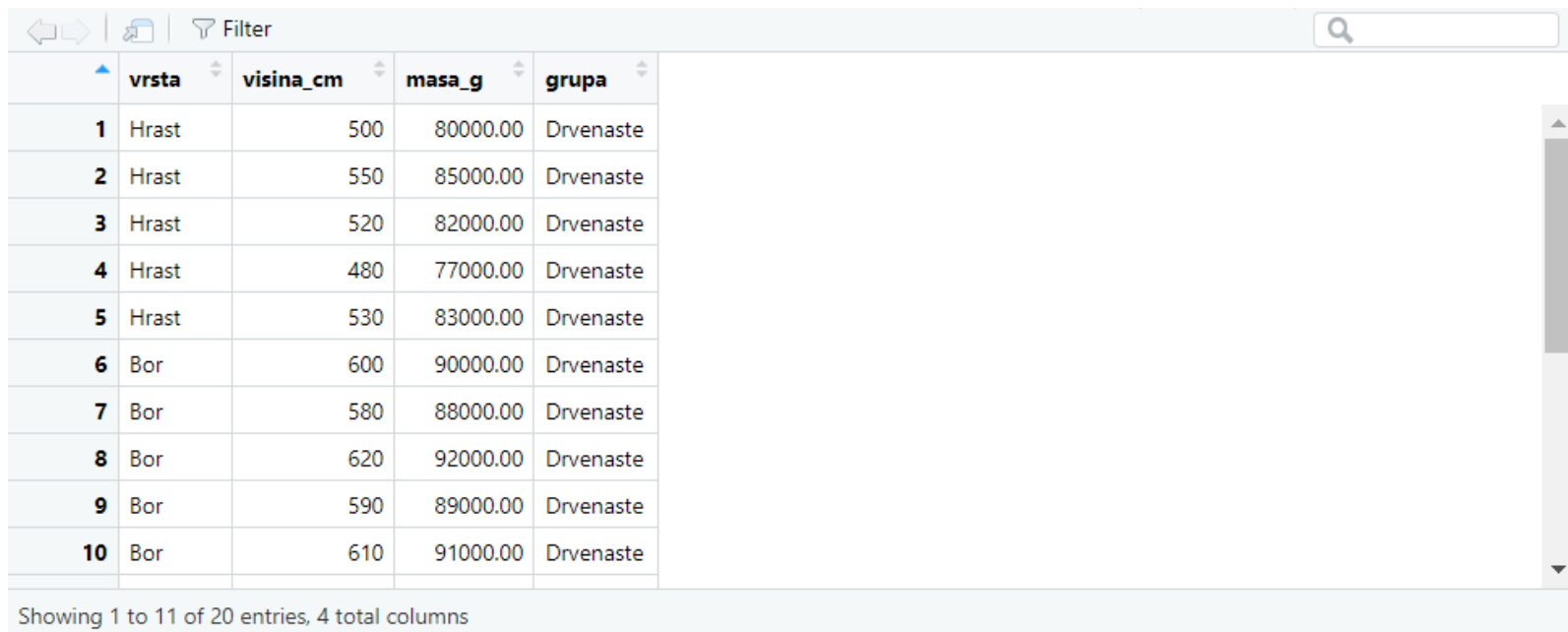
	uzorak	vrsta	visina_cm	masa_g	skupina
1	1	Miš	10	25	Sisavci
2	2	Miš	12	30	Sisavci
3	3	Miš	9	22	Sisavci
4	4	Miš	11	28	Sisavci
5	5	Miš	13	35	Sisavci
6	6	Golub	35	300	Ptice
7	7	Golub	32	280	Ptice
8	8	Golub	36	320	Ptice
9	9	Golub	34	310	Ptice
10	10	Golub	33	290	Ptice
11	11	Vrabac	15	24	Ptice

Showing 1 to 12 of 20 entries, 5 total columns

Učitavanje CSV tablica

```
# Učitavanje CSV tablica  
biljke <- read.csv("biljke/biljke.csv",  
                  header = TRUE)
```


View(biljke) # ili klik na objekt u environmentu



	vrsta	visina_cm	masa_g	grupa
1	Hrast	500	80000.00	Drvenaste
2	Hrast	550	85000.00	Drvenaste
3	Hrast	520	82000.00	Drvenaste
4	Hrast	480	77000.00	Drvenaste
5	Hrast	530	83000.00	Drvenaste
6	Bor	600	90000.00	Drvenaste
7	Bor	580	88000.00	Drvenaste
8	Bor	620	92000.00	Drvenaste
9	Bor	590	89000.00	Drvenaste
10	Bor	610	91000.00	Drvenaste

Showing 1 to 11 of 20 entries, 4 total columns

Paket readxl

- Za Excel prvo moramo učitati paket readxl
- `install.packages("readxl")`

```
library(readxl) # učitavanje paketa
```

```
# Učitavanje Excel tablice pingvini
```

```
pingvini <- read_excel("pingvini/pingvini.xlsx")
```

```
View(pingvini) # ili klik na objekt u environmentu
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
1	Adelie	Torgersen	39.1	18.7	181	3750	male	2007
2	Adelie	Torgersen	39.5	17.4	186	3800	female	2007
3	Adelie	Torgersen	40.3	18.0	195	3250	female	2007
4	Adelie	Torgersen	NA	NA	NA	NA	NA	2007
5	Adelie	Torgersen	36.7	19.3	193	3450	female	2007
6	Adelie	Torgersen	39.3	20.6	190	3650	male	2007
7	Adelie	Torgersen	38.9	17.8	181	3625	female	2007
8	Adelie	Torgersen	39.2	19.6	195	4675	male	2007
9	Adelie	Torgersen	34.1	18.1	193	3475	NA	2007
10	Adelie	Torgersen	42.0	20.2	190	4250	NA	2007

Showing 1 to 11 of 344 entries, 8 total columns

Pregled i struktura data frame-ova

```
# Pregled prvih redaka podataka  
head(zivotinje)
```

```
##   uzorak vrsta visina_cm masa_g skupina  
## 1      1  Miš      10      25 Sisavci  
## 2      2  Miš      12      30 Sisavci  
## 3      3  Miš       9      22 Sisavci  
## 4      4  Miš      11      28 Sisavci  
## 5      5  Miš      13      35 Sisavci  
## 6      6 Golub     35     300  Ptice
```

```
# Pregled zadnjih redaka podataka  
tail(zivotinje)
```

```
##      uzorak      vrsta visina_cm masa_g skupina  
## 15      15      Vrabac      13      21      Ptice  
## 16      16 Vjeverica      20     400 Sisavci  
## 17      17 Vjeverica      22     420 Sisavci  
## 18      18 Vjeverica      21     410 Sisavci  
## 19      19 Vjeverica      19     380 Sisavci  
## 20      20 Vjeverica      23     430 Sisavci
```

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

```
## 'data.frame':  20 obs. of  5 variables:  
## $ uzorak    : int  1 2 3 4 5 6 7 8 9 10 ...  
## $ vrsta     : chr  "Miš" "Miš" "Miš" "Miš" ...  
## $ visina_cm: int  10 12 9 11 13 35 32 36 34 33 ...  
## $ masa_g   : int  25 30 22 28 35 300 280 320 310 290 ...  
## $ skupina  : chr  "Sisavci" "Sisavci" "Sisavci" "Sisavci" ...
```

Izvlačenje pojedinačnih stupaca data frame-a

```
# Stupac predstavlja varijable  
vrsta_zivotinja <- zivotinje$vrsta  
print(vrsta_zivotinja)
```

```
## [1] "Miš"      "Miš"      "Miš"      "Miš"      "Miš"      "Golub"  
## [7] "Golub"    "Golub"    "Golub"    "Golub"    "Vrabac"    "Vrabac"  
## [13] "Vrabac"   "Vrabac"   "Vrabac"   "Vjeverica" "Vjeverica" "Vjeverica"  
## [19] "Vjeverica" "Vjeverica"
```

Izvlačenje pojedinačnih redaka data frame-a

```
# Redak predstavlja uzorke  
zivotinja_1 <- zivotinje[1, ]  
print(zivotinja_1)
```

```
##   uzorak vrsta visina_cm masa_g skupina  
## 1      1   Miš       10     25 Sisavci
```


Izvoz R objekata u datoteke

- koristimo naredbu `write.table()`
- nova datoteka će se stvoriti u radnom direktoriju u kojem se trenutno nalazimo

```
# Izvoz R objekata u datoteke
```

```
write.table(df_mikrobi, # definiramo koji objekt želimo eksportati
```

```
  file = "mikrobi.csv", # definiramo naziv datoteke
```

```
  sep = ",", # definiramo separatore stupaca
```

```
  row.names = FALSE) # definiramo da R ne zapiše broj redaka u zasebni stupac
```

Zadaci

1. Kreirajte 2 varijable, u jednu spremite ime vrste po vašem izboru, a u drugu broj jedinki te vrste.
2. Ispištite tip objekta za novokreirane varijable.
3. Kreirajte 2 vektora, u jednom pohranite tekstove, a u drugom brojeve.
4. Kreirajte data frame s biološkim podacima po vašem izboru, mora sadržavati 2 numeričke i 2 kategoričke varijable.
5. Ispištite strukturu novokreiranog data frame-a.
6. Iz vašeg kreiranog data frame-a izolirajte jedan stupac i spremite ga u novi vektor.

Rješenje zadatka 1

1. Kreirajte 2 varijable, u jednu spremite ime vrste po vašem izboru, a u drugu broj jedinki te vrste.

```
vrsta <- "glavata zelva"
```

```
brojnost <- 26
```

Rješenje zadatka 2

2. Ispištite tip objekta za novokreirane varijable.

```
class(vrsta)
```

```
## [1] "character"
```

```
class(brojnost)
```

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

Rješenje zadatka 3

#3. Kreirajte 2 vektora, u jednom pohranite tekstove, a u drugom brojeve.

```
vektor_kornjača <- c("glavata želva", "zelena želva", "sedmopruga usminjača")
```

```
vektor_jedinki <- c(27, 17, 7)
```

Rješenje zadatka 4

```
# 4. Kreirajte data frame s biološkim podacima po vašem izboru,  
# mora sadržavati 2 numeričke i 2 kategoričke varijable.  
df_kornjače <- data.frame(  
  vrsta = c("glavata želva", "zelena želva", "sedmopruga usminjača"),  
  brojnost = c(27, 17, 7),  
  duljina = c(58, 39, 96),  
  lokacija = c("Jadransko more", "Jadransko more", "Tirensko more")  
)
```

Rješenje zadatka 5

```
# 5. Ispištite strukturu novokreiranog data frame-a.  
str(df_kornjače)
```

```
## 'data.frame':    3 obs. of  4 variables:  
##  $ vrsta      : chr  "glavata želva" "zelena želva" "sedmopruga usminjača"  
##  $ brojnost: num  27 17 7  
##  $ duljina  : num  58 39 96  
##  $ lokacija: chr  "Jadransko more" "Jadransko more" "Tirensko more"
```

Rješenje zadatka 6

6. Iz vašeg kreiranog data frame-a izolirajte jedan stupac i spremite ga u novi vektor.

```
lokacija_kornjače <- df_kornjače$lokacija  
print(lokacija_kornjače)
```

```
## [1] "Jadransko more" "Jadransko more" "Tirensko more"
```