

Uvod u R i RStudio

Analize podataka u biološkim istraživanjima

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

2024-10-14

“Koliko volim R i statistiku?”

- Otvorite poveznicu: <https://forms.office.com/e/FPkBKCs02J>
- ili **skenirajte QR kod** i ispunite kratku anketu upoznavanja :)



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

- Upoznavanje s RStudio sučeljem
- Izvršavanje naredbi iz skripte
- različiti načini ispisa
- definiranje varijabli
- pisanje komentara
- osnovne matematičke operacije
- postavljanje radnog direktorija
- traženje pomoći
- instalacija i učitavanje paketa
- vizualizacija tablica
- primjer izrade grafičkog prikaza

Ishodi učenja praktikuma

- Upoznati se s R okruženjem
- Premostiti strah od „programiranja“
- Proći „energiju aktivacije“ koja je potrebna za korištenje R-a i rada na vlastitim podacima i projektima
- Moći izraditi vlastite analize uz pomoći svih danas dostupnih alata

Što nije ishod ovog praktikuma?

- Sveobuhvatno znanje o R-u
- Sveobuhvatno znanje o svim statističkim metodama
- Tečaj programiranja
- Učenje napamet naredbi R-a

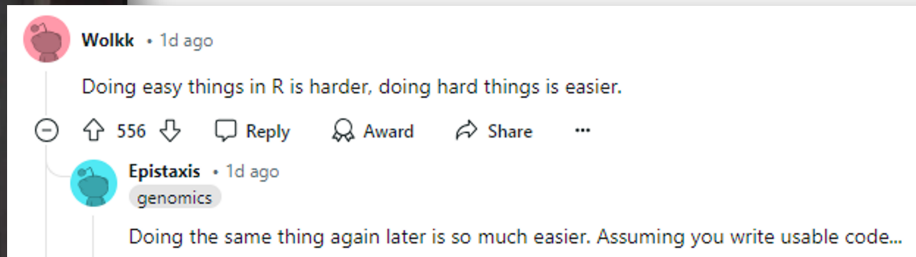
Prednosti R-a

1. Besplatan softver otvorenog koda.
2. Velika zajednica korisnika i podrška.
3. Prilagođen akademskoj zajednici (i biolozima).
4. Fleksibilnost i prilagodljivost - veliki broj paketa za specijalizirane analize.
5. Napredne statističke i analitičke metode.
6. Vizualizacija podataka.
7. Reproducibilnost.

Mane R-a u odnosu na *point and click* programe

- Strma krivulja učenja.
- Manje intuitivno sučelje.

Me spending an an hour making a
“publication quality” histogram in ggplot2
knowing full well I could make the same
thing in Excel in 5 minutes



Instalacija R-a

R se može preuzeti sa CRAN-a (Comprehensive R archive network) <https://cran.r-project.org/>

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, y

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to install R for the first time .
contrib	Binaries of contributed CRAN packages (for R $\geq$ 4.0.x).
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 4.0.x).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

Instalacija RStudio-a

RStudio je dostupan za preuzimanje na Posit web stranici:

- <https://posit.co/download/rstudio-desktop/>

1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

R is not a Posit product. By clicking on the link below to download and install R, you are leaving the Posit website. Posit disclaims any obligations and all liability with respect to R and the R website.

DOWNLOAD AND INSTALL R

2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 265.55 MB | [SHA-256: 513216FE](#) | Version: 2024.09.0+375 |
Released: 2024-09-23

RStudio - sučelje

The screenshot displays the RStudio interface with four main panels, each labeled with a letter in a black box:

- A (Source Panel):** Shows R code for reading metadata, filtering samples, and loading world map data. The code includes comments and function calls like `read.delim`, `filter`, and `world[world$name %in% countries, ]`.
- B (Environment Panel):** Displays the current environment with variables `carapace_samples` (28 obs. of 16 variables) and `metadata` (33 obs. of 16 variables). The `Values` section shows the `countries` variable as a character vector: `chr [1:5] "Croatia" "Slovenia" "Hungary" "Bosnia and Herz." "Italy"`.
- C (Console Panel):** Shows the output of the R script, listing sample names and locations (e.g., `Susak north`, `MaliLosinj north`, `Losinj north`, `MaliLosinj north`, `VeliLosinj north`, `MaliLosinj north`, `Losinj north`, `south`, `DugiOtok south`, `Medulin north`, `MaliLosinj north`, `MaliLosinj north`, `Vis south`, `Susak north`, `Trstenika north`, `MaliLosinj north`). It also shows an error message: `Error in ne_countries(scale = "large", returnclass = "sf"): could not find function "ne_countries"`.
- D (Plots Panel):** Currently empty, showing the `Files`, `Plots`, `Packages`, `Help`, `Viewer`, and `Presentation` tabs.

RStudio - sučelje

- Panel A: **Skripta** - dio u koji pišemo naredbe i komentare koje možemo izvršavati i spremiti za ponovnu upotrebu.
- Panel B: **Konzola** (engl. console) - dio u kojemu se naredbe izvršavaju (interaktivno okruženje), ali se ne može spremiti za ponovnu upotrebu; također i dio za terminal, render i pozadinske poslove.
- Panel C: Okoliš (engl. **environment**) - dio u kojem vidimo stvorene R objekte; također i dio za povijest, veze; git sinkronizaciju i tutoriale.
- Panel D: **plots** - dio za vizualizaciju grafova; **packages** - dio za vizualizaciju i učitavanje dostupnih paketa; **help** - u kojem se prikazu tražene informacije o paketima i naredbama; **files** - vizualizacija trenutnog radnog direktorija (nije za postavljanje novog radnog direktorije).

Rad u R-u i RStudio-u

- Otvorite RStudio i krenimo s prvim naredbama!
- Izvršavanje naredbi: **Ctrl + Enter** u skripti.

```
# Prva naredba: ispis teksta iz skripte u konzolu  
print("Hello world!")
```

```
## [1] "Hello world!"
```

Komentari

- Sve što se piše iza znaka “#” u skripti, predstavlja **komentar** i neće biti izvršeno kao naredba.
- Komentari se pišu u skripti kako bi naš kod bio pregledan i reproducibilan.

```
# Ovo je primjer komentara cijelog reda  
print("Dobar dan svijete!") # ovo je komentar nakon naredbe
```

```
## [1] "Dobar dan svijete!"
```

Definiranje varijabli

- Varijable su pojedinačne vrijednosti.
- U R-u, svaka varijabla može pohranjivati brojeve, tekstualne vrijednosti ili logičke vrijednosti.
- Definiraju se pomoći znakova "<-" ili "="

```
# Definiranje objekata (varijabli)
```

```
a <- 5
```

```
b = 3
```

- prvi način je mnogo češće korišten u praksi jer se znak "=" koristi za i u druge svrhe

Ispis

Ispis tj. prikaz R objekata, može se izvršiti na više načina:

```
# Različiti načina ispisa u konzoli  
# Različiti načina ispisa  
print(a) # ispis u konzoli
```

```
## [1] 5
```

```
a # ispis u konzoli
```

```
## [1] 5
```

```
print(a + b) # ispis rezultata u konzoli
```

```
## [1] 8
```

Još načina ispisa:

```
cat(a) # ispisu u konzoli, ali ima više mogućnosti u odnosu na print()
```

```
## 5
```

```
View(a) # ispis u novom tab-u RStudio-a
```

- Naredba **View()** se koristi uglavnom za tablice.
- pazite na veliko slovo V!

Zadatak

Napravite varijablu “visina” te u nju spremite vrijednost svoje visine, te varijablu “godina” te u nju spremite vrijednost svoje godine rođenja. Ispišite svaku varijablu na različiti način.

```
# Primjer: Definirajmo varijablu svoje visine i godine rođenja  
visina <- 170  
godina <- 1995  
# Ispišimo ove varijable!  
print(visina)
```

```
## [1] 170
```

```
cat (godina)
```

```
## 1995
```

Rad s osnovnim matematičkim operacijama

```
# Zbroj  
zbroj <- 5 + 3  
print(zbroj)
```

```
## [1] 8
```

```
# Umnožak  
umnožak <- 4 * 7  
print(umnožak)
```

```
## [1] 28
```

Zadatak

Na isti način kao za zbroj i umnožak izračunajmo i ispišimo rezultate oduzimanja(razliku) i dijeljenja brojeva (kvocijent).

Razlika i kvocijent

```
# Razlika  
razlika <- 10 - 4  
print(razlika)
```

```
## [1] 6
```

```
# Kvocijent  
kvocijent <- 38 / 7  
print(kvocijent)
```

```
## [1] 5.428571
```

Rad s datotakama

Radni direktorij, engl. **working directory**

```
# Provjeravanje radnog direktorija (working directory)  
getwd()
```

```
## [1] "C:/Users/Hrvoje/Documents/APUBI/01_Uvod"
```

```
# Postavljanje radnog direktorija
```

```
#("C:/Users/Lucij/Documents/APUBI/01_Uvod/zivotinje") # moj Laptop  
setwd("C:/Users/Hrvoje/Documents/APUBI/01_Uvod/zivotinje") # poslovno računalo
```

```
# Iznad je još jedan primjer korištenja komentara:  
# za naredbe koje ne želimo da se trenutno izvršavaju.
```

Traženje pomoći

```
help(setwd)
```

```
## starting httpd help server ... done
```

```
?setwd
```

```
??setwd
```

Učitavanje datoteke s podacima

```
# Učitavanje datoteke s podacima u objekt "zivotinje"  
zivotinje <- read.table("C:/Users/Hrvoje/Documents/APUBI/01_Uvod/zivotinje/zivotinje.txt",  
                        header = TRUE)
```

Kako pogledati svoje podatke?

```
print(zivotinje) # ispis cijele tablice u konzoli
```

##	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
## 12	12	Vrabac	14	22	Ptice
## 13	13	Vrabac	16	26	Ptice
## 14	14	Vrabac	15	23	Ptice
## 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

```
# Vizualizacija podataka - kratki oblik  
head(zivotinje) # prikaz "glave" tablice
```

```
##   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
```

View(zivotinje) # prikaz u novom tabu pored skripte

	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

Paketi

- Instalacijom dodatnih paketa moguće je prilagoditi svoje analize, a izbora je stvarno puno!
- Npr. Za vizualizaciju podataka najčešće se korsiti dodatni paket **ggplot2** koji nije uključen u temeljni R (*base R*) te ga je potrebno instalirati.
- Instalacija paketa izvršava se samo jednom, a učitavanje paketa mora se izvršiti prilikom svakog otvaranja RStudio-a, tj. početkom svake nove sesije.

Paketi

```
# Za dodatne analize moramo instalirati dodatne R pakete  
# Instalacija paketa - samo jednom
```

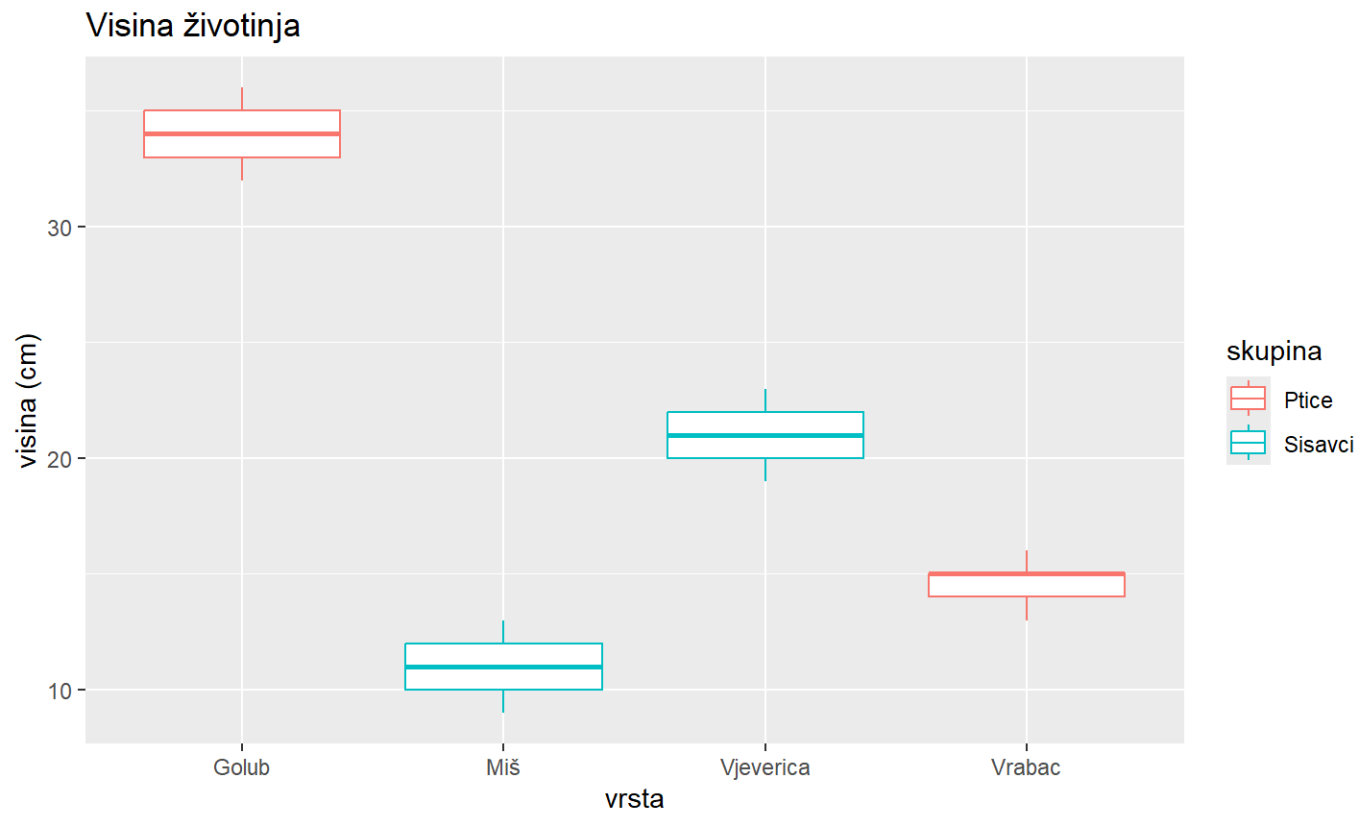
```
install.packages(ggplot2)
```

```
# Učitavanje paketa u radno okruženje - pri svakom pokretanju R-a  
library(ggplot2)
```

Primjer izrade grafa

```
# izrada objekta "graf_visina"  
graf_visina <- ggplot(data = zivotinje, aes(x = vrsta, y = visina_cm)) +  
  geom_boxplot(aes(color = skupina)) +  
  labs(title = "Visina životinja",  
        x = "vrsta",  
        y = "visina (cm)")
```

```
# ispis objekta "graf_visina"  
graf_visina
```

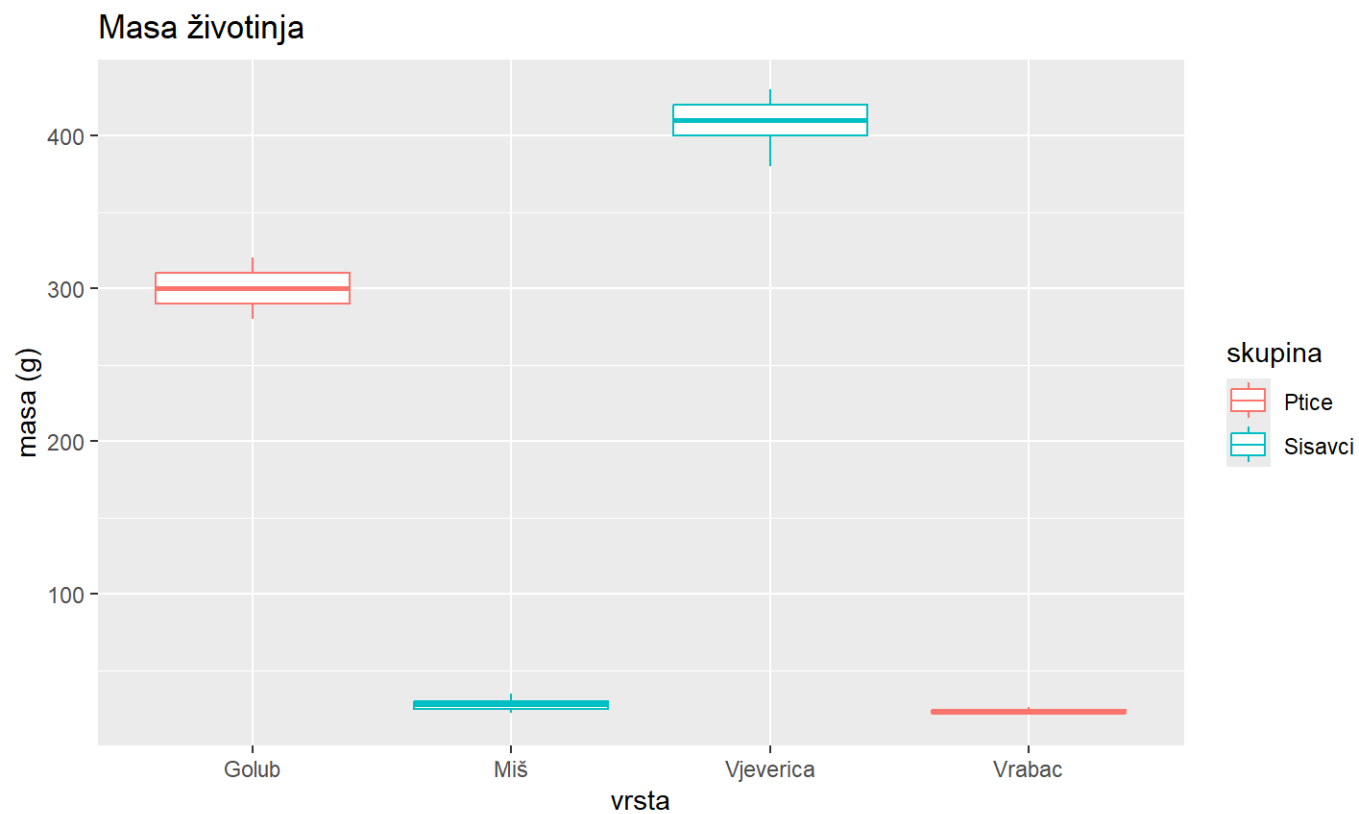


Izmjenite gornji graf tako da se umjesto visine prikazuju podaci za masu!

Pazite da izmjenitie naslov grafa i nazive osi da odgovaraju novom prikazu.

```
# izrada objekta "graf_masa"  
graf_masa <- ggplot(data = zivotinje, aes(x = vrsta, y = masa_g)) +  
  geom_boxplot(aes(color = skupina)) +  
  labs(title = "Masa životinja",  
        x = "vrsta",  
        y = "masa (g)")
```

```
# ispis objekta "graf_masa"  
graf_masa
```



Samostalni zadaci

1. Kreirajte 3 varijable za visine 3 različite životinje.
2. Izračunajte i ispišite umnožak visine životinja koje ste upisali.
3. Napravite novu mapu naziva "biljke" te ju postavite kao novi radni direktorij.
4. Instalirajte paket "tidyverse" i učitajte ga u radno okruženje..

Rješenje zadatka 1

1. Kreirajte 3 varijable za visine 3 različite životinje.

```
miš <- 5
```

```
sova <- 58
```

```
ris <- 75
```

Rješenje zadatka 2

```
# 2. Izračunajte i ispišite umnožak visine životinja koje ste upisali.  
zbroj_visina <- miš + sova + ris  
print(zbroj_visina)
```

```
## [1] 138
```

Rješenje zadatka 3

- u Pregledniku datoketa napravite novu mapu (desni klik -> Novo -> Mapa -> nazovite ju "biljke")

```
# 3. Napravite novu mapu naziva "biljke" te ju postavite kao novi radni direktorij.  
# Postavljanje radnog direktorija naredbom setwd()  
setwd("C:/Users/Hrvoje/Documents/APUBI/01_Uvod/biljke")
```

Rješenje zadatka 4

4. Instalirajte paket "RColorBrewer" i učitajte ga u radno okruženje.

```
install.packages(RColorBrewer)
```

```
library(RColorBrewer)
```

Kviz

Otvorite poveznicu: <https://forms.office.com/e/cG5bh8WvPk>

ili skenirajte QR kod i ispunite kratki kviz danas naučenoga! :)

Kviz iz prve vježbe

