

SPRAWOZDANIE

Zajęcia: Analiza Procesów Uczenia

Prowadzący: prof. dr hab. inż. Vasyl Martsenyuk

| | |
|---|---|
| Laboratorium Nr 1 Data 11.11.2020 Temat: "Podstawy języka R" Wariant 4 | Patryk Starostka Informatyka II stopień, niestacjonarne, 1. semestr, gr.1a |
|---|---|

Repozytorium: <https://github.com/pstarostka/ath-analiza-procesow-uczenia/tree/master/zadanie01>

1. Do zmiennej a podstaw wartość wyrażenia $4 * \sin(\pi)$. Do zmiennej b podstaw potrojną wartość zmiennej a. Wywołaj funkcję sprawdzając, która z wartości zmiennych jest większa.

```
a <- 4 * sin(pi)
b <- 3 * a

compare <- function(a, b) {
  ifelse(a > b, "a jest większe", "b jest większe")
}
print(compare(a, b))
```

```
[1] "b jest większe"
```

2. Uruchom i poczytaj dokumentację dla funkcji max().

```
12 #(b)
13
14 myMax <- max(1, 3)
15 print(myMax)
```

```
[1] "b jest większe"
[1] 3
[1] 10562.5
```

3. Stwórz wektor a zawierający liczby od 90 do 115. Policz średnia kwadratów liczb zawartych w wektorze.

```
17 #(c)
18 myVector <- seq(90, 115, by = 1)
19 squaredVector <- myVector ^ 2
20 print(sum(squaredVector) / length(squaredVector))
21
```

```
[1] 3
[1] 10562.5
[1] "cummax"
```

4. Wyświetl wszystkie funkcje zawierające frazę max w swojej nazwie

```
22 #(d)
23 print(apropos("max", mode = "function"))
```

```
[1] 10562.5
[1] "cummax" "max" "max.col" "mem.maxNSize" "mem.maxVSize"
[6] "pmax" "pmax.int" "promax" "varimax" "which.max"
[11] "łódówka z największą pojemnością"
```

5. Ustaw dowolny katalog roboczy. Następnie stwórz zmienną a zawierającą łańcuch znaków „łódówka z największą pojemnością”. Zapisz zmienną a z obszaru roboczego do pliku w katalogu roboczym. Następnie usuń zmienną a. Sprawdź wartość zmiennej a. Na końcu wczytaj plik ze zmienną a i sprawdź jej wartość

```
29 setwd("/Users/patryk/ath/Magister/Semestr I/apu/zadanie01/workdir")
30 eStr <- "łódówka z największą pojemnością"
31 save(eStr, file = "file.Rdata")
32 remove(eStr)
33 print(eStr)
34 load("file.Rdata")
35 print(eStr)
```

6. Zainstaluj i załaduj pakiet gridExtra, który umożliwia m.in. ładną wizualizację danych tabelarycznych. Następnie przy pomocy dokumentacji pakietu znajdź funkcję do wizualizacji danych tabelarycznych. Użyj jej na pierwszych 10 wierszach zbioru danych rivers.

```
37 #(f)
38 #install.packages("gridExtra")
39 library(gridExtra)
40 grid.table(rivers[1:10])
41
```

```
735
320
325
392
524
450
1459
135
465
600
```

7. Stwórz wektor zawierający ciąg liczb 1000, 998, 996, ... 850.

```
42 #(g)
43 gVector <- seq(1000, 850, -2)
44 print(gVector)
45
```

8.

Stwórz wektora a z liczbami od 30 do 5 oraz wektor b z liczbami od 11 do 23. Utwórz nowy wektory d będący połączeniem wektora b i a (w takiej kolejności). Wyświetl go.

```
46  #(h)
47  hVectorA <- seq(30, 5, -1)
48  hVectorB <- seq(11, 23, 1)
49  hVectorD <- c(hVectorB, hVectorA)
50  print(hVectorD)
51
```

9.

Stwórz wektor nazwa zawierający nazwy 10 lodówek. Potem stwórz wektory pojemność_użytkowa_chłodziarki, pojemność_użytkowa_zamrażarki, cena, liczba_opinii zawierające kolejno dane 10 lodówek. Następnie stwórz ramkę danych lodówki złożoną z wektorów pojemność_użytkowa_chłodziarki, pojemność_użytkowa_zamrażarki, cena, liczba_opinii. Wylicz średnią cenę lodówek.

```
52  #(i)
53  lodowki <- c("lodowka 1", "lodowka 2", "lodowka 3", "lodowka 4", "lodowka 5", "lodowka 6", "lodowka 7", "lodowka 8", "lodowka 9", "lodowka 10")
54  pojemnosc_uzytkowa_chlodziarki <- c(10, 20, 30, 40, 50, 60, 70, 80, 90, 100)
55  pojemnosc_uzytkowa_zamrazarki <- c(100, 90, 80, 70, 60, 50, 40, 30, 20, 10)
56  cena <- c(1000, 1220, 3200, 3120, 1020, 3001, 2000, 4000, 4340, 2330)
57  liczba_opinii <- c(10, 30, 123, 1, 230, 120, 40, 53, 20, 11)
58  tabela <- data.frame(lodowki, pojemnosc_uzytkowa_chlodziarki, pojemnosc_uzytkowa_zamrazarki, cena, liczba_opinii)
59
60  print(tabela)
61  print(mean(tabela$cena))
62
```

```
11      10
[1] 2416.455
lodowki pojemnosc_uzytkowa_chlodziarki
```

10.

Do stworzonej w poprzednim zadaniu ramki danych lodówek dodaj wpis zawierający dane nowej lodówki. Wylicz średnią ceny ponownie.

```
63 #()
64 lodowka <- c("lodowka 11")
65 pojemnosc_uzytkowa_chlodziarki <- c(110)
66 pojemnosc_uzytkowa_zamrazarki <- c(15)
67 cena <- c(1350)
68 liczba_opinii <- c(10)
69 nowa_lodowka_tabelka <- data.frame(lodowka, pojemnosc_uzytkowa_chlodziarki, pojemnosc_uzytkowa_zamrazarki, cena, liczba_opinii)
70 tabelka <- rbind(tabelka, nowa_lodowka_tabelka)
71 print(tabelka)
72 print(mean(tabelka$cena))
```

| | lodowka | pojemnosc_uzytkowa_chlodziarki | pojemnosc_uzytkowa_zamrazarki | cena |
|----|------------|--------------------------------|-------------------------------|------|
| 1 | lodowka 1 | 10 | 100 | 1000 |
| 2 | lodowka 2 | 20 | 90 | 1220 |
| 3 | lodowka 3 | 30 | 80 | 3200 |
| 4 | lodowka 4 | 40 | 70 | 3120 |
| 5 | lodowka 5 | 50 | 60 | 1020 |
| 6 | lodowka 6 | 60 | 50 | 3001 |
| 7 | lodowka 7 | 70 | 40 | 2000 |
| 8 | lodowka 8 | 80 | 30 | 4000 |
| 9 | lodowka 9 | 90 | 20 | 4340 |
| 10 | lodowka 10 | 100 | 10 | 2330 |
| 11 | lodowka 11 | 110 | 15 | 1350 |

11.

Korzystając z ramki danych lodówki dodaj nową kolumnę określając ocenę klientów. Wpisz do kolumny odpowiednio oceny w skali od 0 do 5 krok 0.5. Dodana kolumna powinna się automatycznie przekonwertować do cech jakościowych (tzw. factors). Wylicz średnią ceny każdej oceny.

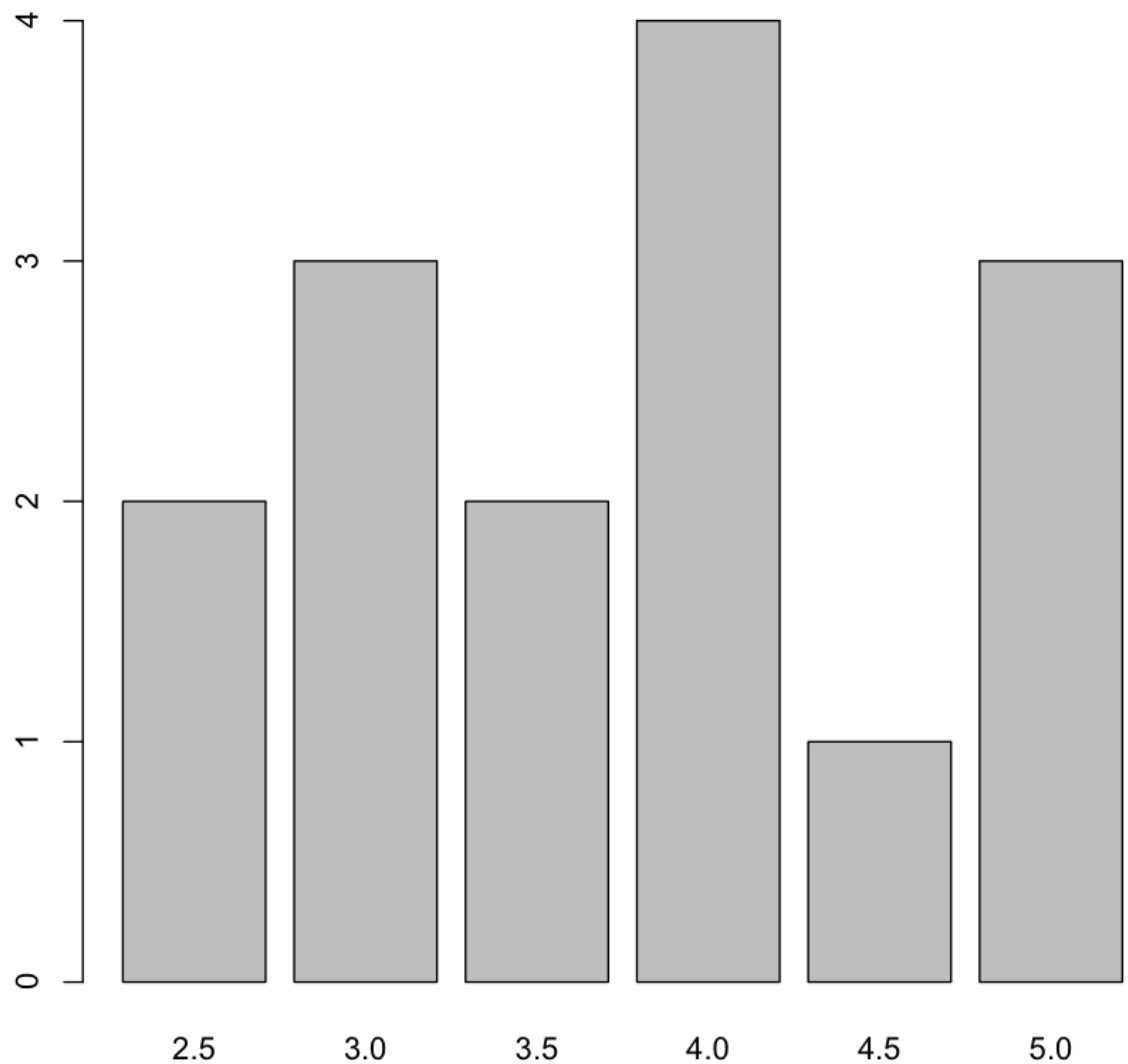
```
74 #()
75 oceny <- c("2.5", "3.0", "4.0", "5.0", "4.5", "4.0", "3.5", "5.0", "3.5", "3.0", "4.0")
76 tabelka <- cbind(tabelka, oceny)
77 print(tabelka)
78 print(tapply(tabelka$cena, tabelka$oceny, mean))
```

| oceny | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
|-------|------|------|------|------|------|------|
| mean | 1000 | 1775 | 3170 | 2517 | 1020 | 3560 |

12.

Do ramki danych aparaty dodaj kolejne 4 lodówki. Narysuj na wykresie słupkowym liczebność reprezentantów każdej z ocen klientów.

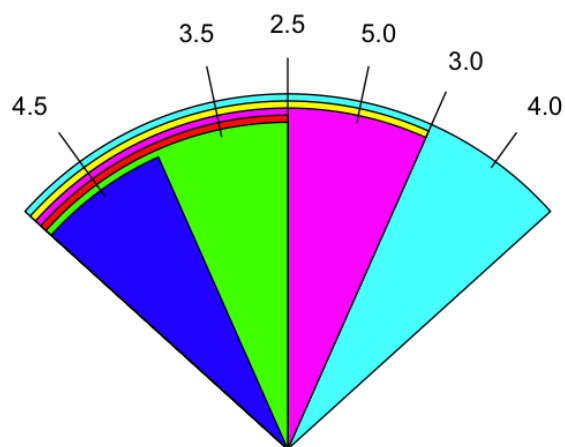
```
80 #()
81 lodowki <- c("lodowka 91", "lodowka 92", "lodowka 93", "lodowka 94")
82 pojemnosc_uzytkowa_chlodziarki <- c(10, 20, 30, 40)
83 pojemnosc_uzytkowa_zamrazarki <- c(100, 90, 80, 70)
84 cena <- c(1000, 1220, 3200, 3120)
85 liczba_opinii <- c(10, 30, 123, 1)
86 oceny <- c("2.5", "3.0", "4.0", "5.0")
87 dodatkowa_tabelka <- data.frame(lodowki, pojemnosc_uzytkowa_chlodziarki, pojemnosc_uzytkowa_zamrazarki, cena, liczba_opinii, oceny)
88 tabelka <- rbind(tabelka, dodatkowa_tabelka)
89 liczebosc <- table(tabelka$oceny)
90 barplot(liczebosc)
```



13.

Wykorzystując ramkę danych lodówki pokaż procentowy udział każdej oceny przy pomocy wykresu kołowego oraz wachlarzowego.

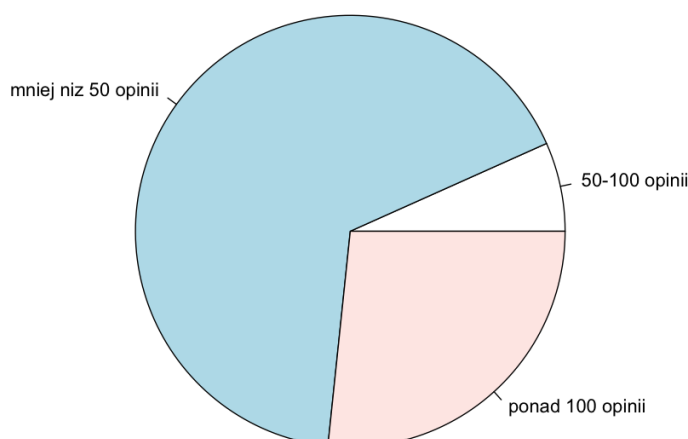
```
92  #(m)
93  #install.packages("plotrix")
94  procenty <- liczebnosc / sum(liczebnosc)
95  pie(procenty)
96  library(plotrix)
97  fan.plot(liczebnosc, labels = names(liczebnosc))
98
```



14.

Do ramki danych lodówki dodaj nową kolumnę `status_opinii` z wartościami: "nie ma", "mniej 50 opinii", "50-100 opinii", "więcej 100 opinii" w zależności od liczby opinii. Zamień dodaną kolumnę na cechy jakościowe. Następnie przy pomocy wykresu kołowego wyrysuj procentowy udział lodówek o konkretnym statusie opinii.

```
99   #(n)
100  tabela[, "status_opinii"] <-
101  ifelse(tabela$liczba_opinii >= 100, 'ponad 100 opinii',
102  ifelse(tabela$liczba_opinii >= 50, '50-100 opinii',
103  ifelse(tabela$liczba_opinii > 0, 'mniej niz 50 opinii', 'nie ma'))
104  )
105  print(tabela)
106  udzialy <- table(tabela$status_opinii)
107  pie(udzialy)
```



15.

Wykorzystując ramkę danych lodowki stwórz zdanie o każdej z lodówek postaci: nazwa + " ma ocenę klientów " + ocena_klientów + " bo ma liczbę opinii" + liczba_opinii. Plus oznacza konkatencję łańcuchów i wartości.

```
109 #()
110 res <- paste(tabela$lodowki, "ma ocene klientow", tabela$oceny, " bo ma liczbe opinii", tabela$liczba_opinii)
111 print(res)
```

```
[1] "lodowka 1 ma ocene klientow 2.5 bo ma liczbe opinii 10"
[2] "lodowka 2 ma ocene klientow 3.0 bo ma liczbe opinii 30"
[3] "lodowka 3 ma ocene klientow 4.0 bo ma liczbe opinii 123"
[4] "lodowka 4 ma ocene klientow 5.0 bo ma liczbe opinii 1"
[5] "lodowka 5 ma ocene klientow 4.5 bo ma liczbe opinii 230"
[6] "lodowka 6 ma ocene klientow 4.0 bo ma liczbe opinii 120"
[7] "lodowka 7 ma ocene klientow 3.5 bo ma liczbe opinii 40"
[8] "lodowka 8 ma ocene klientow 5.0 bo ma liczbe opinii 53"
[9] "lodowka 9 ma ocene klientow 3.5 bo ma liczbe opinii 20"
[10] "lodowka 10 ma ocene klientow 3.0 bo ma liczbe opinii 11"
[11] "lodowka 11 ma ocene klientow 4.0 bo ma liczbe opinii 10"
[12] "lodowka 91 ma ocene klientow 2.5 bo ma liczbe opinii 10"
[13] "lodowka 92 ma ocene klientow 3.0 bo ma liczbe opinii 30"
[14] "lodowka 93 ma ocene klientow 4.0 bo ma liczbe opinii 123"
[15] "lodowka 94 ma ocene klientow 5.0 bo ma liczbe opinii 1"
```


16.

Zachować ramkę danych w pliku .csv. Załadować ramkę danych z pliku .csv

Dane (15 lodówek) pobrać ze strony <http://www.euro.com.pl>

```
[15] lodowka 37 ma 3 opinie klientow 370 30 ma 11220 opinii 1
```

| | nazwa | cena | ocena | liczba_opinii | pojemnosc_chlodziarki |
|----|-----------------------|------|-------|---------------|-----------------------|
| 1 | Whirlpool W7 9210 OX | 1599 | 5 | 52 | 264 |
| 2 | Samsung RB37J501MB1 | 2399 | 5 | 137 | 255 |
| 3 | LG GBB72MCDFN | 2899 | 5 | 440 | 277 |
| 4 | Sharp SJ-BA05DMXLF-EU | 1499 | 5 | 168 | 194 |
| 5 | Samsung RS50N3913BC | 4299 | 5 | 43 | 357 |
| 6 | Amica FK2695.4FTHAA | 1699 | 5 | 205 | 161 |
| 7 | Samsung RS50N3513SA | 3299 | 5 | 50 | 357 |
| 8 | Samsung RB31FERNDBC | 1799 | 5 | 240 | 212 |
| 9 | LG GSJ361DIDV | 4399 | 5 | 219 | 394 |
| 10 | Samsung RB31FDRNDSA | 1890 | 5 | 291 | 210 |
| 11 | Samsung RB29FSRNDSA | 1699 | 5 | 497 | 192 |
| 12 | Amica FK200.4 | 900 | 5 | 63 | 109 |

```
113 #(p)
114 #install.packages('here')
115 csvLodowki <- read.csv(file = here('zadanie01/lodowki.csv'))
116 print(csvLodowki)
```