

SPRAWOZDANIE

Zajęcia: Analiza Procesów Ucznienia

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium 01

25.02.2019

Temat: „Podstawy języka R”

Wariant 6

Mateusz Stępień
Informatyka II stopień
stacjonarne
1 semestr
Gr A

1. Polecenie:

Zadanie dotyczy tworzenia danych, które będą wykorzystywane na kolejnych zajęciach w celu podejmowania decyzji przy kupowaniu urządzeń RTV

Wariant 6:

- (a) Do zmiennej a podstaw wartość wyrażenia $20 \cdot \log_{10}(100)$. Do zmiennej b podstaw potęgową wartość zmiennej a. Wywołaj funkcję sprawdzającą, która z wartości zmiennych jest mniejsza.
 - (b) Uruchom i poczytaj dokumentację dla funkcji `sqrt()`.
 - (c) Stwórz wektor a zawierający liczby od 80 do 105. Policz sumę kwadratów liczb zawartych w wektorze.
 - (d) Wyświetl wszystkie funkcje zawierające frazę `round` w swojej nazwie.
 - (e) Ustaw dowolny katalog roboczy. Następnie stwórz zmienną a zawierającą łańcuch znaków "pralka". Zapisz zmienną a z obszaru roboczego do pliku w katalogu roboczym. Następnie usuń zmienną a. Sprawdź wartość zmiennej a (powinno jej brakować). Na końcu wczytaj plik ze zmienną a i sprawdź jej wartość.
 - (f) Zainstaluj i załaduj pakiet `gridExtra`, który umożliwia ładną wizualizację danych tabelarycznych. Następnie przy pomocy dokumentacji pakietu znajdź funkcję do wizualizacji danych tabelarycznych.
- 10
- Użyj jej na pierwszych 10 wierszach zbioru danych `trees`.
- (g) Stwórz wektor zawierający ciąg liczb 1400, 1395, 1390, . . . 1200.
 - (h) Stwórz wektora a z liczbami od 15 do 3 oraz wektor b z liczbami od 24 do 53. Utwórz nowy wektor d będący połączeniem wektora b i a (w takiej kolejności). Wyświetl go.
 - (i) Stwórz wektor nazwa zawierający nazwy 15 pralek ze sposobem załadunku od przodu. Potem stwórz wektory pojemność, prędkość wirowania, cena, liczba opinii zawierające kolejno dane 15 pralek. Następnie stwórz ramkę danych pralki z łączoną z wektorów pojemność, prędkość wirowania, cena, liczba opinii. Wylicz średnią cenę pralek.
 - (j) Do stworzonej w poprzednim zadaniu ramki danych pralek dodaj wpis zawierający dane nowej pralki. Wylicz średnią cenę ponownie.
 - (k) Korzystając z ramki danych pralki dodaj nową kolumnę określającą 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 cenę każdej oceny.
 - (l) Do ramki danych pralki dodaj kolejne 4 pralki. Narysuj na wykresie słupkowym liczebność reprezentantów każdej z ocen klientów.
 - (m) Wykorzystując ramkę danych pralki pokaż procentowy udział każdej oceny przy pomocy wykresu kołowego oraz wachlarzowego.
 - (n) Do ramki danych pralki dodaj nową kolumnę status opinii z wartościami:

"nie ma", "mniej 50 opinii", "50-100 opinii", "więcej 100 opinii" w zależności od liczby opinii. Zamieść dodaną kolumnę na cechy jakościowe. Następnie przy pomocy wykresu kołowego wyrysuj procentowy udział pralek o konkretnym statusie opinii.

(o) Wykorzystując ramkę danych pralki stwórz zdanie o każdej z pralek postaci: nazwa + " ma ocenę klientów " + ocena_klientów + "

11

bo ma liczbę opinii" + liczba_opinii. Plus oznacza konkatencjęłańcuchów i wartości.

(p) Zachować ramkę danych w pliku .csv. Załadować ramkę danych z pliku .csv

Dane (20 pralek ze sposobem załadunku od przodu) pobrać ze strony <http://www.euro.com.pl>

2. Wykorzystane komendy:

```
##A
```

```
a = 20*log10(100)
```

```
b = 3*a
```

```
pmin(a,b)
```

```
##C
```

```
x <- 80:105
```

```
suma = sum(x^2)
```

```
##D
```

```
apropos("round")
```

```
##E
```

```
setwd("F:/MAGAZYN/ATH/MAGISTERKA/Uczenie maszynowe/Repo/Lab01")
```

```
a <- "pralka"
```

```
write(a, file='pkt_e.csv')
```

```
rm(a)
```

```
a
```

```
a <- read.csv(file = 'pkt_e.csv', sep = ',')
```

```
a
```

```
#f)
```

```
install.packages("gridExtra")
```

```
library("gridExtra")
```

```
g <- tableGrob(trees)
```

```
grid.arrange(g)
```

```
##G
```

```
y <- seq(from = 1400, to = 1200, by = -5)
```

```

##H
h1 <- 15:3
h2 <- 24:53
h3 <- c(h1,h2)
h3

##I
nazwy <- c('Bosch WAE2026EP', 'Candy Smart GVS34', 'Whirlpool FSCR 80421',
           'Electrolux EW7F348SP PerfectCare', 'Miele WDB 030 WCS Eco', 'Indesit IWSC 51052 C ECO PL',
           'Bosch WAB20264PL Serie 2 VarioPerfect', 'Bosch WLT24460PL Serie 6 VarioPerfect', 'Samsung
AddWash Slim WW60K42109W',
           'Amica DAW 6103 LSW', 'Electrolux EW6S306SPX', 'Samsung WW70J4273MW', 'Candy Smart
CS4 1061D3', 'Indesit Eco Time IWD 61052 C ECO PL', 'Bosch WAB24264PL Serie 2 VarioPerfect')
pojemnosc <- c(7, 6, 8, 8, 7, 5, 6,6,6,5,6,6,6,7,6)

predkosc_wirowania <- c(1000, 1200, 1400, 1400, 1400, 1000, 1000, 1200, 1200,
1200,1000,1000,1200,1000,1200)
cena<- c(1199, 1399,1599, 2099, 3689, 799, 999, 1349, 1649, 1649,1099,1599,1299,999,799)
liczba_opinii <- c(154, 162, 10, 44, 18, 549, 250, 250, 525, 82,115,13,90,156,287)
pralki <- data.frame(nazwy, pojemnosc, predkosc_wirowania, cena, liczba_opinii)

mean(pralki$cena)

##J
newRow <- data.frame(nazwy = 'LG FE0J5NN3W', pojemnosc=6, predkosc_wirowania=1000, cena=
1299, liczba_opinii=43)

pralki <- rbind(pralki, newRow)
mean(pralki$cena)

#K
pralki$ocena <- c('5', '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5')
aggregate(pralki$cena, list(pralki$ocena), mean)

#L
newRow <- data.frame(nazwy = 'LG FE0J5NN3W', pojemnosc=6, predkosc_wirowania=1000, cena=
1299, liczba_opinii=114, ocena='5')

pralki <- rbind(pralki, newRow)
newRow <- data.frame(nazwy = 'LG FE0J5NN3W', pojemnosc=7, predkosc_wirowania=1200, cena=
1399, liczba_opinii=43, ocena='4')

pralki <- rbind(pralki, newRow)
newRow <- data.frame(nazwy = 'LG FE0J5NN3W', pojemnosc=5, predkosc_wirowania=1000, cena=
1799, liczba_opinii=72, ocena='5')

pralki <- rbind(pralki, newRow)
newRow <- data.frame(nazwy = 'LG FE0J5NN3W', pojemnosc=5, predkosc_wirowania=1400, cena=
1299, liczba_opinii=34, ocena='4,5')

pralki <- rbind(pralki, newRow)

```

```
newRow <- data.frame(nazwy = 'LG FE0J5NN3W', pojemnosc=6, predkosc_wirowania=1200, cena=
1499, liczba_opinii=18, ocena='5')
```

```
pralki <- rbind(pralki, newRow)
```

```
dane <- aggregate(pralki$liczba_opinii, list(pralki$ocena), sum)
barplot(dane[,2], names.arg = dane[,1], main = 'Liczba reprezentantow kazdej z ocen')
```

```
#M
```

```
labels <- round(dane[,2]/sum(dane[,2]) * 100, 1)
labels <- paste(labels, "%", sep="")
pie(dane[,2], radius = 1, col = rainbow(length(dane[,2])),
    labels = labels)
legend(1.6, 0.8, dane[,1], cex=0.8, fill=rainbow(length(dane[,2])))
```

```
#N)
```

```
new_column <- ifelse(pralki$liczba_opinii>100,'wiecej 100 opinii',
    ifelse(pralki$liczba_opinii>=50, '50-100 opinii',
        ifelse(pralki$liczba_opinii>0, 'mniej 50 opinii', 'nie ma')))
pralki['status_opinii'] <- factor(new_column)
```

```
#wykres kolowy
```

```
pie(table(pralki$status_opinii), radius = 1, col = rainbow(length(pralki$status_opinii)))
```

```
#O)
```

```
for (i in 1:length(pralki$nazwy)){
    print(paste(pralki$nazwy[i], 'ma ocene klientow', pralki$ocena[i], 'bo ma liczbe opinii',
        pralki$liczba_opinii[i]))
}
```

```
#P)
```

```
write.csv(pralki, 'pralki.csv')
dane <- read.csv('pralki.csv')
```

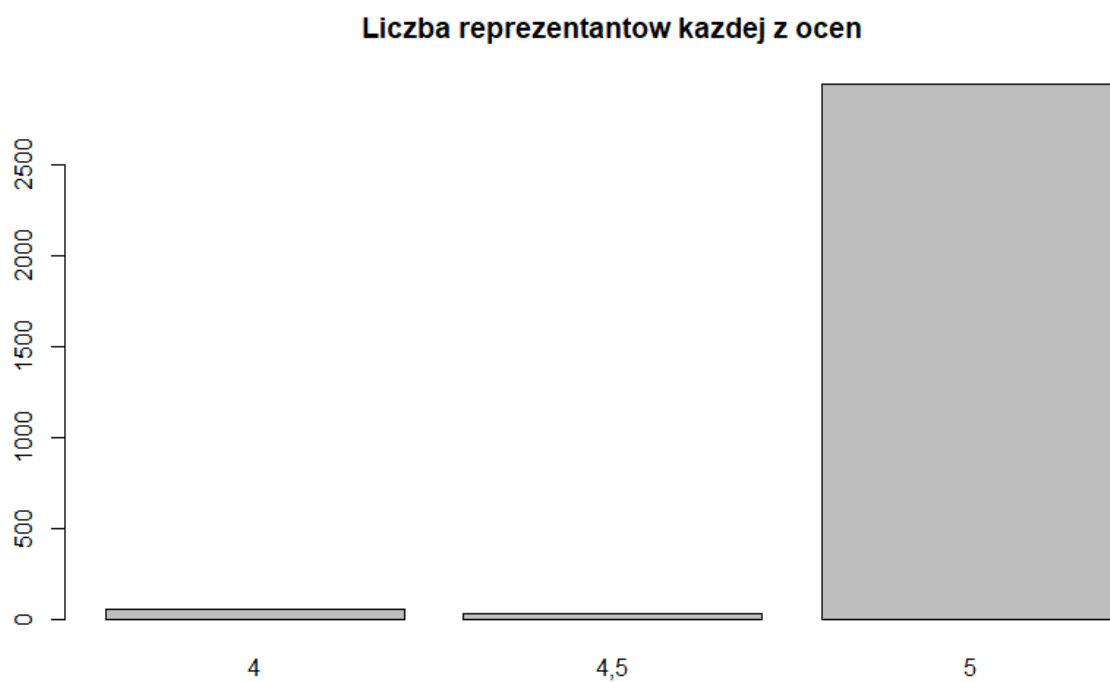
3. Wynik działania:

Link do repozytorium:

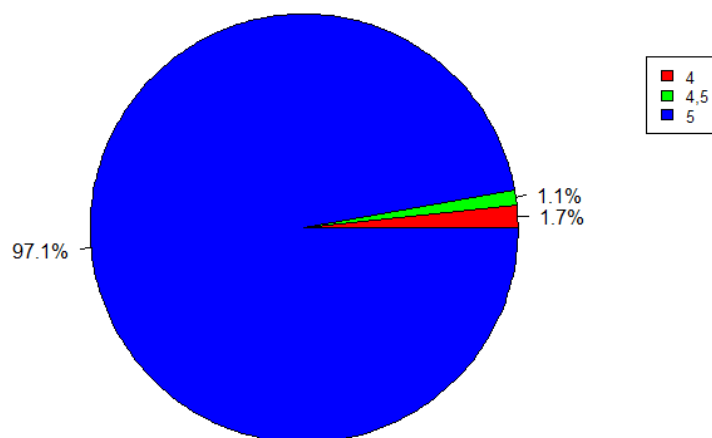
Wyświetlenie 10 wierszy danych trees:

	Girth	Height	Volume
1	8.3	70	10.3
2	8.6	65	10.3
3	8.8	63	10.2
4	10.5	72	16.4
5	10.7	81	18.8
6	10.8	83	19.7
7	11	66	15.6
8	11	75	18.2
9	11.1	80	22.6
10	11.2	75	19.9

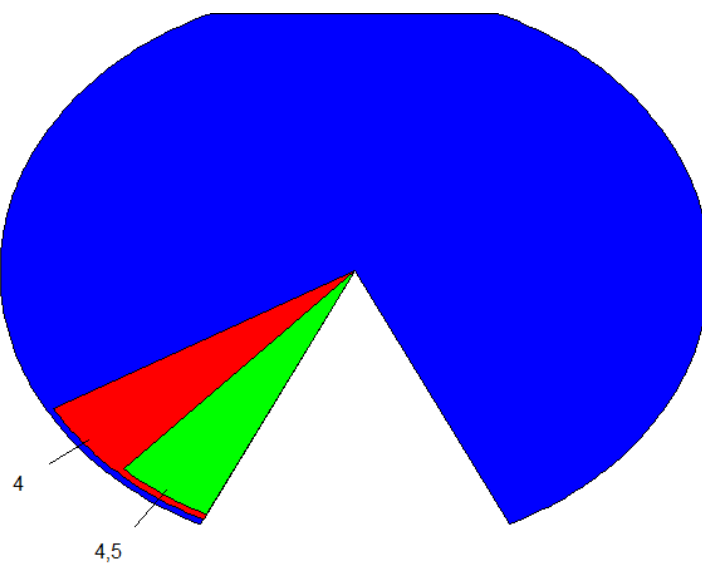
Wykres słupkowy przedstawiający liczebność reprezentantów każdej z ocen pralek:



Procentowy udział każdej z ocen pralek
- wykres kołowy



-wykres wachlarzowy



Procentowy udział pralek o konkretnym statusie opinii

