

Podstawy języka R - zadania (2)

Tomasz Owczarek, Mateusz Naramski

2024/2025, semestr letni

Wybór elementów wektora

16. Utwórz wektor $x = [6, 4, 2, 7, 1, 9, 0, 6]$. Następnie wyświetl wektor logiczny zawierający wartości TRUE na pozycjach, w których wektor x ma wartości:

- a) równe 6,
- b) nierówne 6,
- c) większe niż 4,
- d) mniejsze niż 3 lub większe niż 6.

17. Żeby wyświetlić wartości wektora, które spełniają jakieś warunki, należy użyć wektora logicznego wewnątrz nawiasów kwadratowych, np. zapis `x[x > 6]` wyświetli te wartości wektora `x`, które są większe niż 6. Pamiętając o tym, wyświetl wartości wektora x z poprzedniego zadania, które

- a) są mniejsze od 5,
- b) są mniejsze od 2 lub większe od 7,
- c) są większe od 4 i mniejsze od 9.

18. Korzystając z wektora `rivers` wyświetl:

- a) długości rzek dłuższych niż 1000,
- b) liczbę rzek krótszych niż 300 (*skorzystaj z sumowania wektora logicznego spełniającego warunki polecenia*),
- c) liczbę rzek powyżej średniej długości,
- d) udział rzek poniżej średniej długości w liczbie wszystkich rzek (*podziel liczbę rzek poniżej średniej przez liczbę wszystkich rzek - liczbę wszystkich rzek możesz wydobyć sprawdzając długość wektora `rivers`*),
- e) liczbę rzek o długości równej lub mniejszej od I kwartyla długości wszystkich rzek (*skorzystaj z funkcji `quantile` z zadania 14.*).

Wyniki:

```
## [1] 1459 1450 1243 2348 1171 3710 2315 2533 1306 1054 1270 1885 1100 1205 1038
## [16] 1770
```

```
## [1] 29
```

```
## [1] 47
```

```
## [1] 0.6666667
```

```
## [1] 36
```

19. Wykonaj poniższy kod

```
set.seed(42)
x <- rnorm(100000, mean = 100, sd = 15) # 100tys. liczb z rozkładu normalnego
```

a następnie:

- oblicz średnią x ,
- oblicz odchylenie standardowe x ,
- oblicz, jaki procent elementów x jest poniżej średniej (policz ile liczb jest poniżej średniej, podziel to przez długość x i pomnóż przez 100),
- oblicz, jaki procent elementów x jest mniejszy o więcej niż jedno odchylenie standardowe od średniej x ,
- oblicz, jaki procent elementów x znajduje się nie dalej od średniej x niż dwa odchylenia standardowe.

Wyniki:

```
## [1] 99.93811
```

```
## [1] 15.04837
```

```
## [1] 49.901
```

```
## [1] 15.948
```

```
## [1] 95.463
```

Podstawy pracy z ramką danych

20. Pracuj na ramce `mtcars`. Wyświetl:

- wagę najcięższego samochodu (zmienna `wt`),
- średnią wagę samochodów,
- liczbę samochodów z podziałem na automatyczną i manualną skrzynią biegów (zmienna `am`, skorzystaj z funkcji `table`),
- procent samochodów z automatyczną i manualną skrzynią biegów (podziel wynik z poprzedniego punktu przez liczbę wszystkich samochodów i pomnóż przez 100),
- średnie spalanie (`mpg`) samochodów z czterema biegami (`gear`),
- średni czas przejechania 1/4 mili (`qsec`) samochodów, których waga jest mniejsza niż 2.

Wyniki:

```
## [1] 5.424
```

```
## [1] 3.21725
```

```
##
```

```
## 0 1
```

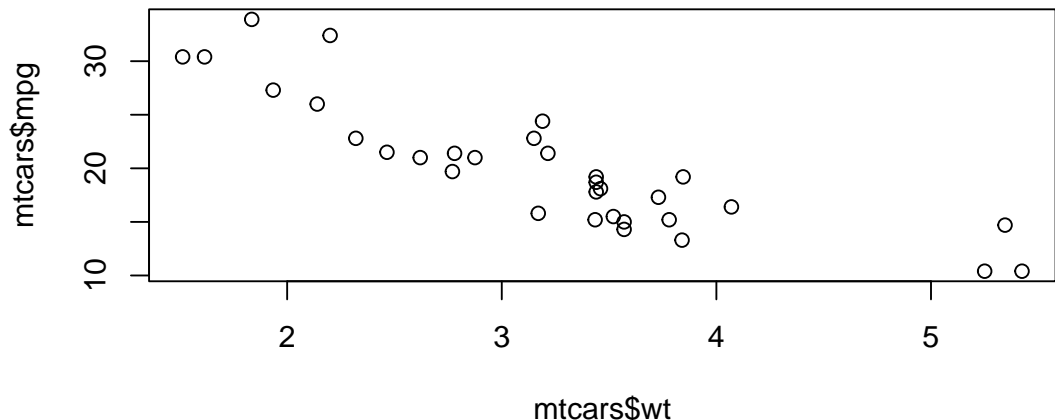
```
## 19 13
```

```
##
##      0      1
## 59.375 40.625

## [1] 24.53333

## [1] 18.555
```

21. Korzystając z ramki `mtcars` przedstaw na wykresie punktowym zależność między wagą samochodu (`wt`) a spalaniem (`mpg`).



Czy ta zależność ma sens? Sprawdź w pomocy ramki `mtcars` co oznacza zmienna `mpg`.

22. Z ramki `mtcars` wyświetl tylko:

- 4 pierwsze pełne rekordy,
- rekordy, dla których liczba gaźników (`carb`) jest równa 1,
- rekordy, dla których waga jest większa niż 4.
- rekordy, dla których liczba koni mechanicznych jest większa niż 150 i waga jest większa niż 5

Wyniki:

```
##      mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46 0  1   4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02 0  1   4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61 1  1   4    1
## Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44 1  0   3    1
```

```
##      mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Datsun 710     22.8   4 108.0  93 3.85 2.320 18.61 1  1   4    1
## Hornet 4 Drive 21.4   6 258.0 110 3.08 3.215 19.44 1  0   3    1
## Valiant        18.1   6 225.0 105 2.76 3.460 20.22 1  0   3    1
## Fiat 128       32.4   4  78.7  66 4.08 2.200 19.47 1  1   4    1
## Toyota Corolla 33.9   4  71.1  65 4.22 1.835 19.90 1  1   4    1
## Toyota Corona  21.5   4 120.1  97 3.70 2.465 20.01 1  0   3    1
## Fiat X1-9      27.3   4  79.0  66 4.08 1.935 18.90 1  1   4    1
```

```
##      mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Merc 450SE     16.4   8 275.8 180 3.07 4.070 17.40 0  0   3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98 0  0   3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82 0  0   3    4
## Chrysler Imperial 14.7   8 440.0 230 3.23 5.345 17.42 0  0   3    4
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Cadillac Fleetwood 10.4  8  472 205 2.93 5.250 17.98 0 0   3   4
## Lincoln Continental 10.4  8  460 215 3.00 5.424 17.82 0 0   3   4
## Chrysler Imperial  14.7  8  440 230 3.23 5.345 17.42 0 0   3   4
```

23. Z ramki `mtcars` wyświetl:

- kolumny `wt`, `hp`, `cyl` i `disp` dla samochodów o wadze mniejszej niż 2,2,
- kolumny `hp`, `cyl` i `mpg` dla samochodów z liczbą koni mechanicznych mniejszą niż 80,
- kolumny `qsec` i `am` dla samochodów z manualną skrzynią biegów (sprawdź w pomocy `mtcars`, która to wartość `am`) i czasem na 1/4 mili poniżej 16.

Wyniki:

```
##           wt  hp cyl  disp
## Honda Civic   1.615  52  4  75.7
## Toyota Corolla 1.835  65  4  71.1
## Fiat X1-9     1.935  66  4  79.0
## Porsche 914-2 2.140  91  4 120.3
## Lotus Europa  1.513 113  4  95.1
```

```
##           hp cyl  mpg
## Merc 240D    62  4 24.4
## Fiat 128     66  4 32.4
## Honda Civic   52  4 30.4
## Toyota Corolla 65  4 33.9
## Fiat X1-9    66  4 27.3
```

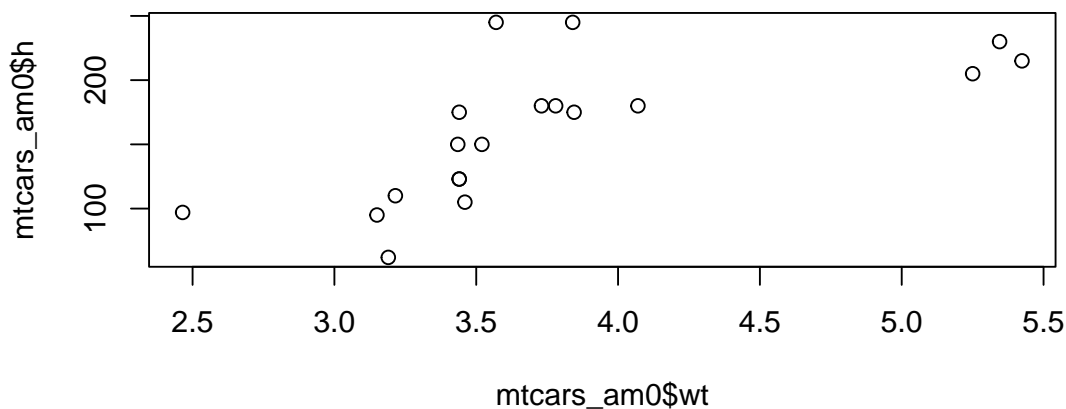
```
##           qsec am
## Ford Pantera L 14.5  1
## Ferrari Dino   15.5  1
## Maserati Bora  14.6  1
```

24. Utwórz ramkę `mtcars_am0` zawierającą wszystkie dane samochodów z automatyczną skrzynią biegów. Następnie:

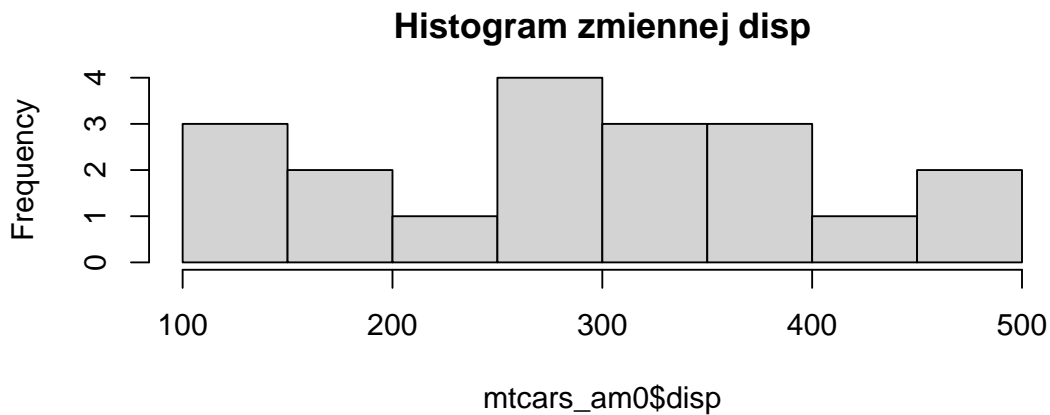
- sprawdź wymiary tej ramki (*funkcja `dim`*),
- utwórz wykres punktowy przedstawiający zależność między zmienną `wt` i `hp`.

Wyniki:

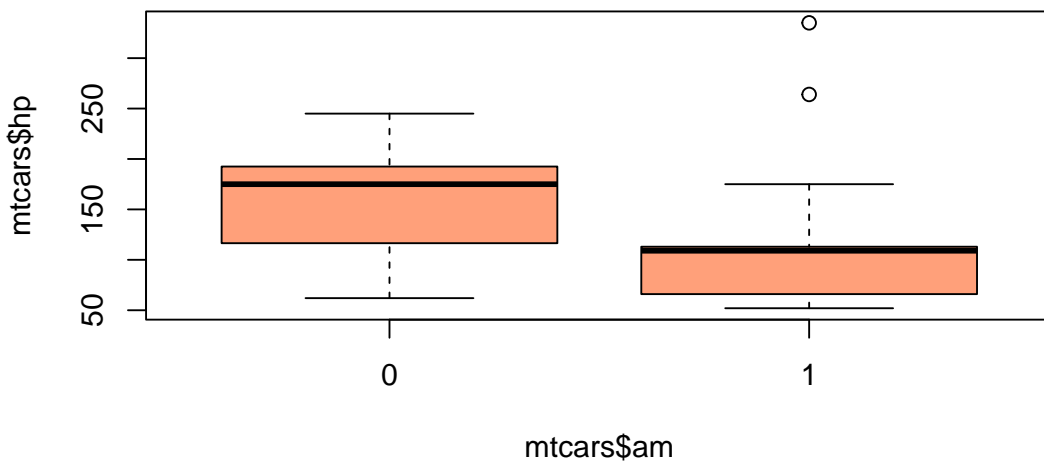
```
## [1] 19 11
```



25. Utwórz histogram zmiennej `disp` z ramki `mtcars_am0` (zmień tytuł wykresu tak jak poniżej).



26. Korzystając z danych z ramki `mtcars` utwórz poniższy wykres pudełkowy (kolor wypełnienia to `lightsalmon`).



Zadania różne

27. Ramka danych `InsectSprays` zawiera dane z eksperymentu o skuteczności 6 preparatów przeciwko insektom.

- Sprawdź strukturę tej ramki, czyli wymiary, nazwy i typy kolumn i pierwsze wartości (*funkcja `str`*).
- Sprawdź, ile rekordów odpowiada każdemu preparatowi (*funkcja `table`*).
- Oblicz średnią liczbę insektów, na których zadziałał preparat "A".
- Oblicz średnią liczbę insektów, na których zadziałał preparat "C".
- Przedstaw na wykresach pudełkowych skuteczność poszczególnych preparatów (opisz odpowiednio osie).

Wyniki:

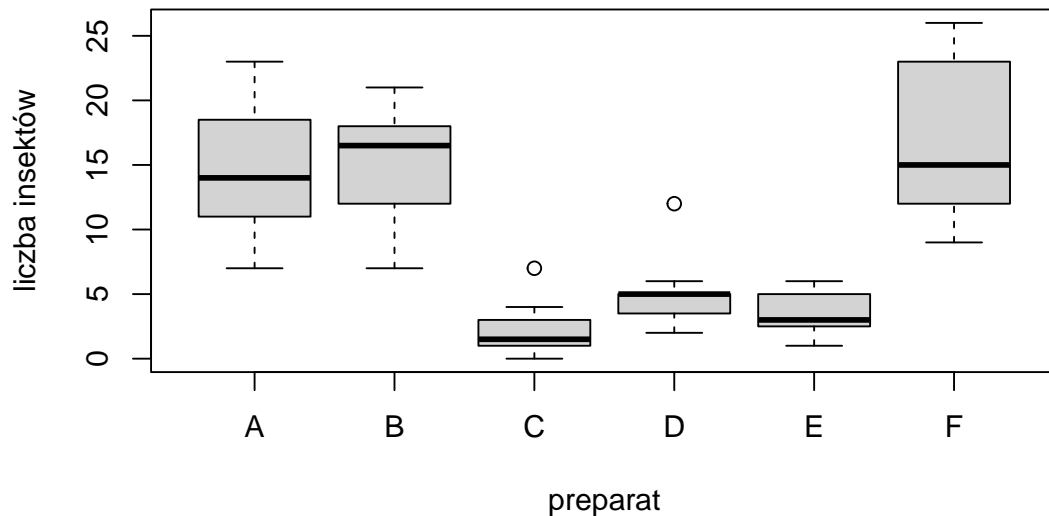
```
## 'data.frame':    72 obs. of  2 variables:
##  $ count: num  10 7 20 14 14 12 10 23 17 20 ...
##  $ spray: Factor w/ 6 levels "A","B","C","D",...: 1 1 1 1 1 1 1 1 1 1 ...

##
```

```
## A B C D E F
## 12 12 12 12 12 12
```

```
## [1] 14.5
```

```
## [1] 2.083333
```



28. Ramka danych `USArrests` zawiera informacje o liczbie aresztowań (na 100tys. mieszkańców) w wyniku popełnionych przestępstw (*Murder, Assault, Rape*) w poszczególnych stanach USA w 1973 roku oraz procent ludności zamieszkujących tereny miejskie w tym stanie (*UrbanPop*). Nazwy stanów stanowią nazwy wierszy.

Wykonaj poniższy kod, który utworzy nową ramkę `arr` z tymi samymi danymi oraz z dodatkową kolumną `State`, która będzie zawierała nazwy stanów:

```
arr <- USArrests

# nowa kolumna
arr$State <- rownames(USArrests) # funkcja rownames() zwraca nazwy wierszy
```

Pracując na tej nowej ramce:

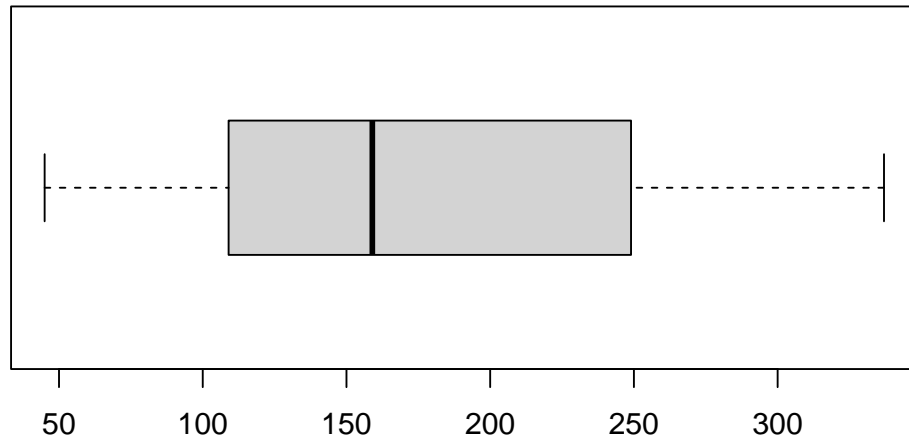
- wyświetl nazwy stanów, w których ludność w większości mieszka na wsi,
- wyświetl nazwę stanu z największą liczbą aresztowań za gwałt,
- wyświetl nazwę stanów z najmniejszą liczbą aresztowań za morderstwo,
- liczbę stanów, w których aresztowania za napad przekraczają 200,
- utwórz wykres pudełkowy liczby aresztowań za napad (wykres ma być poziomy, znajdź w pomocy funkcji `boxplot` odpowiedni argument).

```
## [1] "Alaska" "Mississippi" "North Carolina" "North Dakota"
## [5] "South Carolina" "South Dakota" "Vermont" "West Virginia"
```

```
## [1] "Nevada"
```

```
## [1] "North Dakota"
```

```
## [1] 19
```



Funkcja `is.na`

29. Specyficzną wartością w R jest `NA` (wartość nieznana, brakująca). Funkcją, która sprawdza, czy element wektora jest `NA`, jest funkcja `is.na()`. Sprawdź działanie tej funkcji wykonując poniższy kod:

```
v <- c(5, NA, 2, NA, 7)
is.na(v)
!is.na(v)
```

Następnie wykonaj poniższe 2 linie kodu:

```
set.seed(10) # ziarno generatora liczb losowych
grades <- sample(c(5, 4.5, 4, 3.5, 3, NA), 58, replace = TRUE) # symulacja 58 ocen
```

oraz

- wyświetl wektor logiczny zawierający wartości `TRUE` na pozycjach, na których w `grades` znajdują się `NA` i `FALSE` na pozostałych,
- liczbę wystawionych ocen (czyli liczbę pozycji, na który w wektorze `grades` nie ma `NA`),
- wyświetl wystawione oceny, bez wartości `NA`,
- wyświetl liczbę ocen 5.0*,
- wyświetl średnią wszystkich ocen (przy obliczaniu średniej pomiń wartości `NA`)*.

* W obu tych przypadkach warto zerknąć do pomocy funkcji `sum` i `mean` i zmienić odpowiedni argument tych funkcji - to będzie sprytne rozwiązanie.

Wyniki:

```
## [1] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE
## [13] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE

## [1] 49
```

```
## [1] 4.0 5.0 4.5 3.5 4.0 4.5 4.5 4.5 3.0 4.0 4.5 3.0 3.0 3.0 5.0 3.5 4.5 3.0 4.5
## [20] 3.0 4.5 5.0 4.5 4.5 5.0 3.5 3.0 4.5 4.0 4.0 4.5 4.0 5.0 4.5 3.5 4.0 4.5 3.0
## [39] 5.0 5.0 5.0 3.0 4.5 5.0 5.0 4.5 3.0 3.5 4.0
```

```
## [1] 10
```

```
## [1] 4.112245
```

Zadanie dodatkowe

30! Skorzystaj z ramki `arr` utworzonej w zadaniu 28. Utwórz poniższy wykres przedstawiający dane 12 stanów z największym udziałem populacji miejskiej. Dopasuj wielkość i odległość napisów od punktów, żeby wszystko było czytelne.

