

Błażej Domagała - WED

Lab 1

Zadanie 1

a) Nazwa zbioru: Wine

```
> library(datasets)
>
> path = "C:\\users\\petioff\\Desktop\\repos\\uo\\rok 3\\wprowadzenie do eksploracji danych\\lista1\\zadanie2" # używając podwójnych ukośników
> setwd(path) ## ustawienie ścieżki
>
> ## zmiana nazwy kolumn:
>
> # Załadowanie danych
> # Spróbuj wczytać dane z innym separatorem
> wine <- read.csv('wine\\wine.data', header=FALSE)
```

b) Krótki tekstowy opis zbioru

Zbiór danych "Wine" zawiera wyniki analizy chemicznej win wyprodukowanych w określonym regionie we Włoszech przez trzech różnych producentów. Analiza chemiczna dotyczy 13 różnych składników zawartych w winach.

```
> ## zmiana nazwy kolumn:
>
> # Załadowanie danych
> # Spróbuj wczytać dane z innym separatorem
> wine <- read.csv('wine\\wine.data', header=FALSE)
>
> # zmień nazwy kolumn
> names(wine) <- c('class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids', 'Nonflavanoid phenols', 'Proanthocyanins', 'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline')
>
> ##
```

c) Liczba obserwacji w zbiorze: 178

```
> view(wine)
>
> ## Liczba obserwacji w zbiorze
> nrow(wine)
[1] 178
>
```

d) Liczba kolumn: 14 (13 atrybutów + 1 kolumna identyfikująca klasę)

```
> ## Liczba kolumn
> print(names(wine))
[1] "class"                "Alcohol"              "Malic acid"           "Ash"
[5] "Alcalinity of ash"    "Magnesium"            "Total phenols"        "Flavanoids"
[9] "Nonflavanoid phenols" "Proanthocyanins"      "Color intensity"      "Hue"
[13] "OD280/OD315 of diluted wines" "Proline"
```

e) Zmienna celu:

- Nazwa kolumny z klasą: Class
- Liczba klas: 3 (Klasy 1, 2 i 3)

```
> ## Zmienna celu
> unique(wine$class)
[1] 1 2 3
>
```

f) Wykaz i opis cech:

1. **Class:** Zmienna katégoryczna. Klasa wina.
2. **Alcohol:** Zmienna ilościowa. Zawartość alkoholu.
3. **Malic acid:** Zmienna ilościowa. Zawartość kwasu jabłkowego.
4. **Ash:** Zmienna ilościowa. Zawartość popiołu.
5. **Alcalinity of ash:** Zmienna ilościowa. Zasadowość popiołu.
6. **Magnesium:** Zmienna ilościowa. Zawartość magnezu.
7. **Total phenols:** Zmienna ilościowa. Całkowita zawartość fenoli.
8. **Flavanoids:** Zmienna ilościowa. Zawartość flawonoidów.
9. **Nonflavanoid phenols:** Zmienna ilościowa. Zawartość fenoli nieflawonoidowych.
10. **Proanthocyanins:** Zmienna ilościowa. Zawartość proantocyjanidyn.
11. **Color intensity:** Zmienna ilościowa. Intensywność koloru.
12. **Hue:** Zmienna ilościowa. Odcień.
13. **OD280/OD315 of diluted wines:** Zmienna ilościowa. Stosunek absorbancji przy 280 nm do 315 nm w rozcieńczonych winach.
14. **Proline:** Zmienna ilościowa. Zawartość prolina.

```
> summary(wine)
  class      Alcohol      Malic acid      Ash      Alcalinity of ash      Magnesium      Total phenols      Flavanoids      Nonflavanoid phenols
Min.   :1.000   Min.   :11.03   Min.   :0.740   Min.   :1.360   Min.   :10.60   Min.   : 70.00   Min.   :0.980   Min.   :0.340   Min.   :0.1300
1st Qu.:1.000   1st Qu.:12.36   1st Qu.:1.603   1st Qu.:2.210   1st Qu.:17.20   1st Qu.: 88.00   1st Qu.:1.742   1st Qu.:1.205   1st Qu.:0.2700
Median :2.000   Median :13.05   Median :1.865   Median :2.360   Median :19.50   Median : 98.00   Median :2.355   Median :2.135   Median :0.3400
Mean   :1.938   Mean   :13.00   Mean   :2.336   Mean   :2.367   Mean   :19.49   Mean   : 99.74   Mean   :2.295   Mean   :2.029   Mean   :0.3619
3rd Qu.:3.000   3rd Qu.:13.68   3rd Qu.:3.083   3rd Qu.:2.558   3rd Qu.:21.50   3rd Qu.:107.00   3rd Qu.:2.800   3rd Qu.:2.875   3rd Qu.:0.4375
Max.   :3.000   Max.   :14.83   Max.   :5.800   Max.   :3.230   Max.   :30.00   Max.   :162.00   Max.   :3.880   Max.   :5.080   Max.   :0.6600
Proanthocyanins color intensity      Hue      OD280/OD315 of diluted wines      Proline
Min.   :0.410   Min.   : 1.280   Min.   :0.4800   Min.   :1.270   Min.   : 278.0
1st Qu.:1.250   1st Qu.: 3.220   1st Qu.:0.7825   1st Qu.:1.938   1st Qu.: 500.5
Median :1.555   Median : 4.690   Median :0.9650   Median :2.780   Median : 673.5
Mean   :1.591   Mean   : 5.058   Mean   :0.9574   Mean   :2.612   Mean   : 746.9
3rd Qu.:1.950   3rd Qu.: 6.200   3rd Qu.:1.1200   3rd Qu.:3.170   3rd Qu.: 985.0
Max.   :3.580   Max.   :13.000   Max.   :1.7100   Max.   :4.000   Max.   :1680.0
```

Zadanie 2

- a) Zmiana nazw kolumn (pierwszą kolumnę – zmienną celu – proszę nazwać „Class”; nazwy pozostałych kolumn – atrybutów – są podane w pliku „wine.names”)

```
R 4.3.1 C:/Users/petitoff/Desktop/repos/UO/rok 3/Wprowadzenie do eksploracji danych/lista1/zadanie2
> library(datasets)
>
> path = "C:\\Users\\petitoff\\Desktop\\repos\\UO\\rok 3\\Wprowadzenie do eksploracji danych\\lista1\\zadanie2" # używając podwójnych ukośników
> setwd(path) ## ustawienie ścieżki
>
> ## zmiana nazwy kolumn:
>
> # Załadowanie danych
> wine <- read.csv("wine\\wine.data", header=FALSE)
>
> # Zmień nazwy kolumn
> names(wine) <- c('Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoids', 'Nonflavanoid phenols', 'Proanthocyanins', 'color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline')
>
```

- b) Polecenie View (fragment print-screena z tabelką)

```
> ## =====
> view(wine)
>
```

RStudio interface showing a dataset named 'wine' with 178 entries and 14 columns. The columns are: Class, Alcohol, Malic acid, Ash, Alkalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, and Proline.

The console shows the following R code:

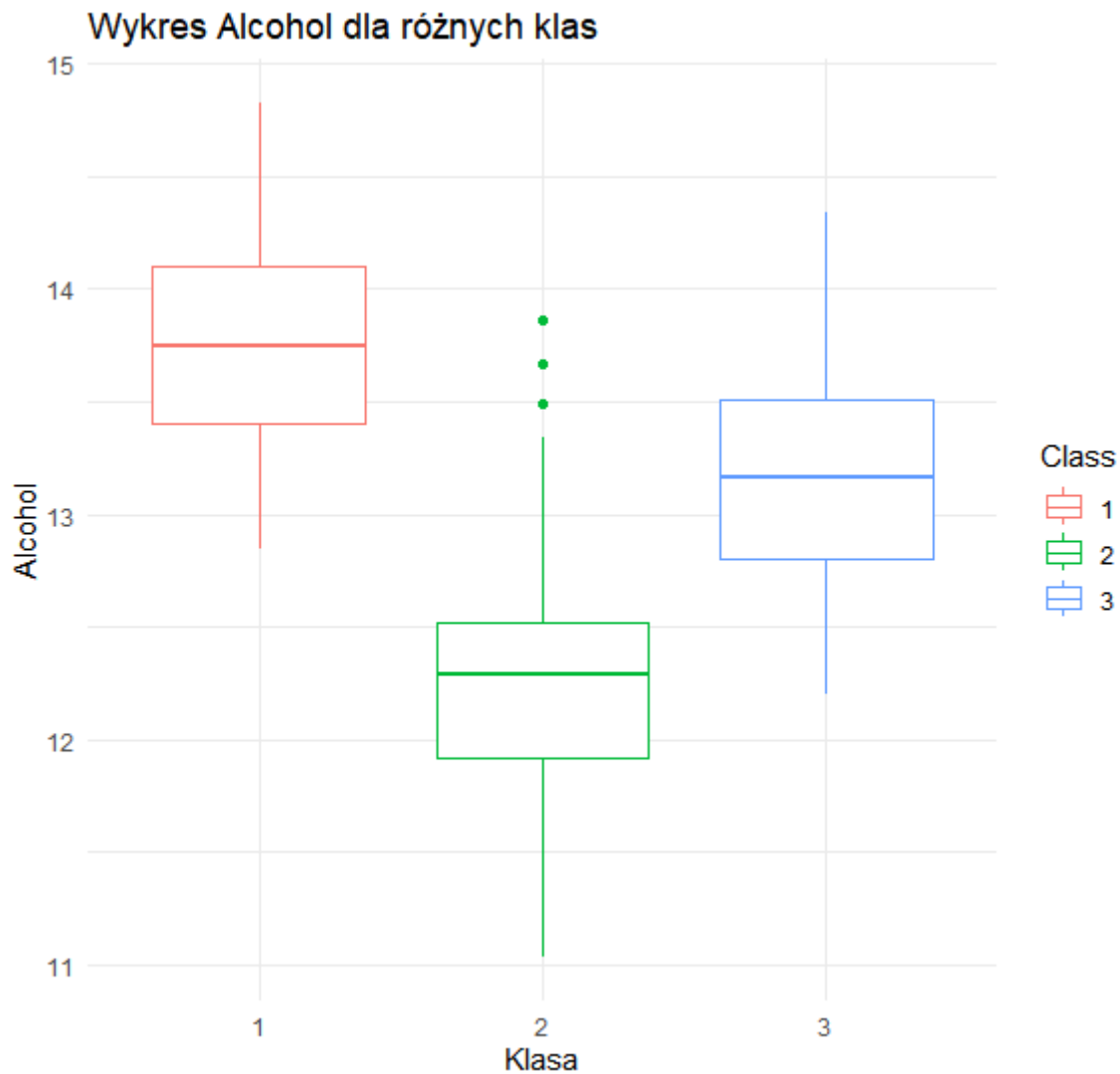
```
R 4.3.1 - C:/Users/petioff/Desktop/repos/UO/rok 3/Wprowadzenie do eksploracji danych/lista1/zadanie2/
> library(datasets)
>
> path = "C:\\Users\\petioff\\Desktop\\repos\\UO\\rok 3\\Wprowadzenie do eksploracji danych\\lista1\\zadanie2" # używając podwójnych ukośników
> setwd(path) ## ustawienie ścieżki
>
> ## zmiana nazwy kolumn:
```

c) Podsumowanie cech (summary)

```
> ## Podsumowanie cech
> summary(wine)
      Class      Alcohol      Malic acid      Ash      Alkalinity of ash      Magnesium      Total phenols      Flavanoids      Nonflavanoid phenols
Min.   :1.000   Min.   :11.03   Min.   :0.740   Min.   :1.360   Min.   :10.60   Min.   : 70.00   Min.   :0.980   Min.   :0.340   Min.   :0.1300
1st Qu.:1.000   1st Qu.:12.36   1st Qu.:1.603   1st Qu.:2.210   1st Qu.:17.20   1st Qu.: 88.00   1st Qu.:1.742   1st Qu.:1.205   1st Qu.:0.2700
Median :2.000   Median :13.05   Median :1.865   Median :2.360   Median :19.50   Median : 98.00   Median :2.355   Median :2.135   Median :0.3400
Mean   :1.938   Mean   :13.00   Mean   :2.336   Mean   :2.367   Mean   :19.49   Mean   : 99.74   Mean   :2.295   Mean   :2.029   Mean   :0.3619
3rd Qu.:3.000   3rd Qu.:13.68   3rd Qu.:3.083   3rd Qu.:2.558   3rd Qu.:21.50   3rd Qu.:107.00   3rd Qu.:2.800   3rd Qu.:2.875   3rd Qu.:0.4375
Max.   :3.000   Max.   :14.83   Max.   :5.800   Max.   :3.230   Max.   :30.00   Max.   :162.00   Max.   :3.880   Max.   :5.080   Max.   :0.6600

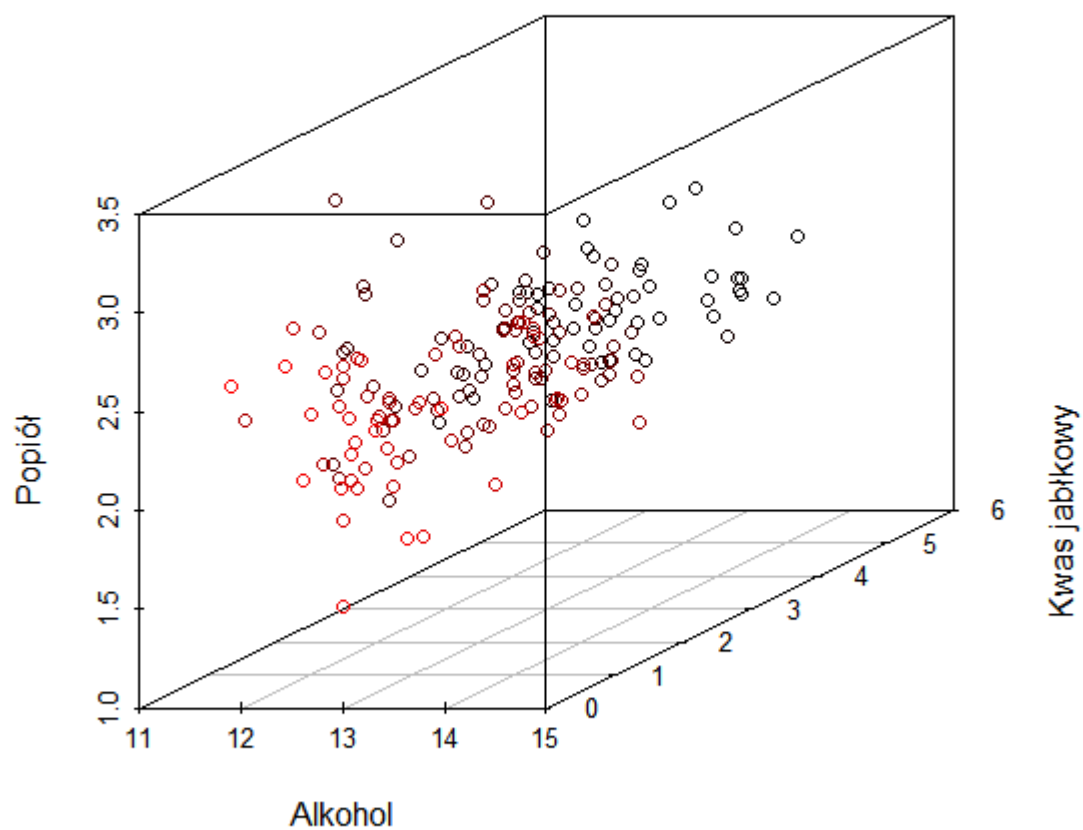
      Proanthocyanins      Color intensity      Hue      OD280/OD315 of diluted wines      Proline
Min.   :0.410   Min.   : 1.280   Min.   :0.4800   Min.   :1.270   Min.   : 278.0
1st Qu.:1.250   1st Qu.: 3.220   1st Qu.:0.7825   1st Qu.:1.938   1st Qu.: 500.5
Median :1.555   Median : 4.690   Median :0.9650   Median :2.780   Median : 673.5
Mean   :1.591   Mean   : 5.058   Mean   :0.9574   Mean   :2.612   Mean   : 746.9
3rd Qu.:1.950   3rd Qu.: 6.200   3rd Qu.:1.1200   3rd Qu.:3.170   3rd Qu.: 985.0
Max.   :3.580   Max.   :13.000   Max.   :1.7100   Max.   :4.000   Max.   :1680.0
```

d) Wykres 2D ilustrujący wybraną cechę dla różnych klas



```
> ## wykres 2D:
> data(wine, package = "datasets")
Komunikat ostrzegawczy:
W poleceniu 'data(wine, package = "datasets")':
zbiór danych 'wine' nie został znaleziony
> # Załadowanie biblioteki do tworzenia wykresów
> library(ggplot2)
> # wybór cechy do przedstawienia na wykresie - np. "Alcohol"
> feature <- "Alcohol"
> # konwersja zmiennej 'class' na faktor
> wine$class <- as.factor(wine$class)
> # Stworzenie wykresu 2D z użyciem ggplot2
> p <- ggplot(wine, aes(x = class, y = !!sym(feature), color = class)) +
+   geom_boxplot() +
+   labs(title = paste("Wykres", feature, "dla różnych klas"),
+        x = "Klasa",
+        y = feature) +
+   theme_minimal()
> # Wyświetlenie wykresu
> print(p)
> ## wykres 3D:
> library(scatterplot3d)
> # Stwórz wykres 3D dla wybranych cech
> scatterplot3d(wine[,c("Alcohol", "Malic acid", "Ash",
+   xlab="Alcohol", ylab="Kwas jabłkowy", zlab="Popiół",
+   highlight.3d=TRUE, angle=30)
> # Zapis do pliku
> ggsave("C:\\Users\\petitoff\\desktop\\repos\\00\\rok 3\\wprowadzenie do eksploracji danych\\lista1\\zadanie2\\wykres.png", plot = p, width = 10, height = 6, dpi = 300)
>
```

e) Wykres 3D dla trzech wybranych cech (bez klasy)



```
> # Stwórz wykres 3D dla wybranych cech
> scatterplot3d(wine$Alcohol, wine$`Malic acid`, wine$Ash,
+               xlab="Alkohol", ylab="Kwas jabłkowy", zlab="Popiół",
+               highlight.3d=TRUE, angle=30)
> |
```

Lab 2

- a) Dla wybranej cechy wyświetlić wybrane wartości stosując: zakres wartości, sekwencję indeksów (np. co dziesiąty indeks), indeksy ujemne, warunki logiczne.

Kod:

```
path = "C:\\Users\\petitoff\\Desktop\\repos\\UO\\rok 3\\Wprowadzenie do eksploracji  
danych\\lista2"
```

```
setwd(path) ## ustawienie ścieżki
```

```
# Załadowanie danych
```

```
wine <- read.csv('wine\\wine.data', header = FALSE)
```

```
# Zmień nazwy kolumn
```

```
names(wine) <-
```

```
c(
```

```
  'Class',
```

```
  'Alcohol',
```

```
  'Malic acid',
```

```
  'Ash',
```

```
  'Alcalinity of ash',
```

```
  'Magnesium',
```

```
  'Total phenols',
```

```
  'Flavanoids',
```

```
  'Nonflavanoid phenols',
```

```
  'Proanthocyanins',
```

```
  'Color intensity',
```

```
  'Hue',
```

```
  'OD280/OD315 of diluted wines',
```

```
  'Proline'
```

```
)
```

a) Dla wybranej cechy wyświetlić wybrane wartości stosując: zakres wartości, sekwencję indeksów (np. co dziesiąty indeks), indeksy ujemne, warunki logiczne.

```
# Zakres wartości
```

```
wine$Alcohol[5:15]
```

```
# Sekwencja indeksów
```

```
wine$Alcohol[seq(1, nrow(wine), 10)]
```

```
# Indeksy ujemne
```

```
wine$Alcohol[-(1:5)]
```

```
# Warunki logiczne
```

```
wine$Alcohol[wine$Alcohol > 14]
```

```
> path = "C:\\Users\\petitoff\\Desktop\\repos\\U0\\rok 3\\wprowadzenie do eksploracji danych\\lista2"
> setwd(path) ## ustawienie ścieżki
>
> # Załadowanie danych
> wine <- read.csv('wine\\wine.data', header=FALSE)
>
> # Zmień nazwy kolumn
> names(wine) <- c('Class', 'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium', 'Total phenols', 'Flavanoi
ds', 'Nonflavanoid phenols', 'Proanthocyanins', 'color intensity', 'Hue', 'od280/od315 of diluted wines', 'Proline')
>
> # head(wine)
>
> # a) Dla wybranej cechy wyświetlić wybrane wartości stosując: zakres wartości, sekwencję indeksów (np. co dziesiąty i
ndeks), indeksy ujemne, warunki logiczne.
> # Zakres wartości
> wine$Alcohol[5:15]
[1] 13.24 14.20 14.39 14.06 14.83 13.86 14.10 14.12 13.75 14.75 14.38
>
> # Sekwencja indeksów
> wine$Alcohol[seq(1, nrow(wine), 10)]
[1] 14.23 14.10 14.06 13.73 13.56 13.05 12.33 12.29 12.00 12.08 12.08 11.46 11.45 12.86 12.93 13.50 12.36 12.20
>
> # Indeksy ujemne
> wine$Alcohol[-(1:5)]
[1] 14.20 14.39 14.06 14.83 13.86 14.10 14.12 13.75 14.75 14.38 13.63 14.30 13.83 14.19 13.64 14.06 12.93 13.71
[19] 12.85 13.50 13.05 13.39 13.30 13.87 14.02 13.73 13.58 13.68 13.76 13.51 13.48 13.28 13.05 13.07 14.22 13.56
[37] 13.41 13.88 13.24 13.05 14.21 14.38 13.90 14.10 13.94 13.05 13.83 13.82 13.77 13.74 13.56 14.22 13.29 13.72
[55] 12.37 12.33 12.64 13.67 12.37 12.17 12.37 13.11 12.37 13.34 12.21 12.29 13.86 13.49 12.99 11.96 11.66 13.03
[73] 11.84 12.33 12.70 12.00 12.72 12.08 13.05 11.84 12.67 12.16 11.65 11.64 12.08 12.08 12.00 12.69 12.29 11.62
[91] 12.47 11.81 12.29 12.37 12.29 12.08 12.60 12.34 11.82 12.51 12.42 12.25 12.72 12.22 11.61 11.46 12.52 11.76
[109] 11.41 12.08 11.03 11.82 12.42 12.77 12.00 11.45 11.56 12.42 13.05 11.87 12.07 12.43 11.79 12.37 12.04 12.86
[127] 12.88 12.81 12.70 12.51 12.60 12.25 12.53 13.49 12.84 12.93 13.36 13.52 13.62 12.25 13.16 13.88 12.87 13.32
[145] 13.08 13.50 12.79 13.11 13.23 12.58 13.17 13.84 12.45 14.34 13.48 12.36 13.69 12.85 12.96 13.78 13.73 13.45
[163] 12.82 13.58 13.40 12.20 12.77 14.16 13.71 13.40 13.27 13.17 14.13
>
> # Warunki logiczne
> wine$Alcohol[wine$Alcohol > 14]
[1] 14.23 14.37 14.20 14.39 14.06 14.83 14.10 14.12 14.75 14.38 14.30 14.19 14.06 14.02 14.22 14.21 14.38 14.10
[19] 14.22 14.34 14.16 14.13
> |
```

b) Wyświetlić wybrane wiersze i kolumny z tabeli.

```
# b) Wybranie wierszy i kolumn:
```

```
selected_rows_columns <- wine[1:10, c("Alcohol", "Malic acid")]
```

```
print(selected_rows_columns)
```

```
# indeksy
```

wine[5:15,]

```
> # b) wybranie wierszy i kolumn:
> selected_rows_columns <- wine[1:10, c("Alcohol", "Malic acid")]
> print(selected_rows_columns)
  Alcohol Malic acid
1    14.23      1.71
2    13.20      1.78
3    13.16      2.36
4    14.37      1.95
5    13.24      2.59
6    14.20      1.76
7    14.39      1.87
8    14.06      2.15
9    14.83      1.64
10   13.86      1.35
>
> # indeksy
> wine[5:15, ]
  Class Alcohol Malic acid Ash Alkalinity of ash Magnesium Total phenols Flavonoids Nonflavanoid phenols
5      1   13.24      2.59 2.87          21.0      118          2.80      2.69          0.39
6      1   14.20      1.76 2.45          15.2     112          3.27      3.39          0.34
7      1   14.39      1.87 2.45          14.6      96          2.50      2.52          0.30
8      1   14.06      2.15 2.61          17.6     121          2.60      2.51          0.31
9      1   14.83      1.64 2.17          14.0      97          2.80      2.98          0.29
10     1   13.86      1.35 2.27          16.0      98          2.98      3.15          0.22
11     1   14.10      2.16 2.30          18.0     105          2.95      3.32          0.22
12     1   14.12      1.48 2.32          16.8      95          2.20      2.43          0.26
13     1   13.75      1.73 2.41          16.0      89          2.60      2.76          0.29
14     1   14.75      1.73 2.39          11.4      91          3.10      3.69          0.43
15     1   14.38      1.87 2.38          12.0     102          3.30      3.64          0.29
  Proanthocyanins Color intensity Hue OD280/OD315 of diluted wines Proline
5          1.82          4.32 1.04          2.93      735
6          1.97          6.75 1.05          2.85     1450
7          1.98          5.25 1.02          3.58     1290
8          1.25          5.05 1.06          3.58     1295
9          1.98          5.20 1.08          2.85     1045
10         1.85          7.22 1.01          3.55     1045
11         2.38          5.75 1.25          3.17     1510
12         1.57          5.00 1.17          2.82     1280
13         1.81          5.60 1.15          2.90     1320
14         2.81          5.40 1.25          2.73     1150
15         2.96          7.50 1.20          3.00     1547
> |
```

c) Dodać do tabeli nową kolumnę z wartościami obliczonymi na podstawie innych wybranych kolumn.

c) Dodanie nowej kolumny:

Sprawdzanie, czy kolumna istnieje i zawiera dane

```
if ("Total phenols" %in% names(wine) &&
```

```
    !all(is.na(wine$`Total phenols`))) {
```

```
  # Dodanie nowej kolumny
```

```
  wine$Total.phenols.squared <- wine$`Total phenols` ^ 2
```

```
  head(wine)
```

```
} else {
```

```
  cat("Column 'Total phenols' does not exist or is empty.")
```

```
}
```



```

> # c) Dodanie nowej kolumny:
> # Sprawdzanie, czy kolumna istnieje i zawiera dane
> if ("Total phenols" %in% names(wine) && !all(is.na(wine$`Total phenols`))) {
+   # Dodanie nowej kolumny
+   wine$Total.phenols.squared <- wine$`Total phenols`^2
+   head(wine)
+ } else {
+   cat("Column 'Total phenols' does not exist or is empty.")
+ }
  Class Alcohol Malic acid Ash Alkalinity of ash Magnesium Total phenols Flavanoids Nonflavanoid phenols
1      1    14.23      1.71 2.43          15.6      127      2.80      3.06              0.28
2      1    13.20      1.78 2.14          11.2     100      2.65      2.76              0.26
3      1    13.16      2.36 2.67          18.6     101      2.80      3.24              0.30
4      1    14.37      1.95 2.50          16.8     113      3.85      3.49              0.24
5      1    13.24      2.59 2.87          21.0     118      2.80      2.69              0.39
6      1    14.20      1.76 2.45          15.2     112      3.27      3.39              0.34
  Proanthocyanins Color intensity Hue OD280/OD315 of diluted wines Proline Total.phenols.squared
1          2.29          5.64 1.04          3.92    1065          7.8400
2          1.28          4.38 1.05          3.40    1050          7.0225
3          2.81          5.68 1.03          3.17    1185          7.8400
4          2.18          7.80 0.86          3.45    1480          14.8225
5          1.82          4.32 1.04          2.93     735          7.8400
6          1.97          6.75 1.05          2.85    1450          10.6929
>

```

d) Podać wartości podstawowych statystyk dla wybranej kolumny: zakres, średnia, mediana,

d) Statystyki podstawowe dla wybranej kolumny, np. "Alcohol"

```
cat("Statystyki dla kolumny 'Alcohol':", "\n")
```

```
cat("Zakres: ", min(wine$Alcohol), " - ", max(wine$Alcohol), "\n")
```

```
cat("Średnia: ", mean(wine$Alcohol), "\n")
```

```
cat("Mediana: ", median(wine$Alcohol), "\n")
```

```
cat("Odchylenie standardowe: ", sd(wine$Alcohol), "\n")
```

```
cat("Kurtoza: ", moments::kurtosis(wine$Alcohol), "\n")
```

```
cat("Skośność: ", moments::skewness(wine$Alcohol), "\n")
```

```
cat("Kwantyle: ", quantile(wine$Alcohol, probs = c(0.25, 0.5, 0.75)), "\n")
```

```

> # d) Statystyki podstawowe dla wybranej kolumny, np. "Alcohol"
> cat("Statystyki dla kolumny 'Alcohol':", "\n")
Statystyki dla kolumny 'Alcohol':
> cat("Zakres: ", min(wine$Alcohol), " - ", max(wine$Alcohol), "\n")
Zakres: 11.03 - 14.83
> cat("Średnia: ", mean(wine$Alcohol), "\n")
Średnia: 13.00062
> cat("Mediana: ", median(wine$Alcohol), "\n")
Mediana: 13.05
> cat("odchylenie standardowe: ", sd(wine$Alcohol), "\n")
odchylenie standardowe: 0.8118265
> cat("Kurtoza: ", moments::kurtosis(wine$Alcohol), "\n")
Kurtoza: 2.13774
> cat("skośność: ", moments::skewness(wine$Alcohol), "\n")
skośność: -0.05104747
> cat("Kwantyle: ", quantile(wine$Alcohol, probs = c(0.25, 0.5, 0.75)), "\n")
Kwantyle: 12.3625 13.05 13.6775
>

```

e) Wyznaczyć i zilustrować na wykresie macierz korelacji dla wybranych pięciu zmiennych.

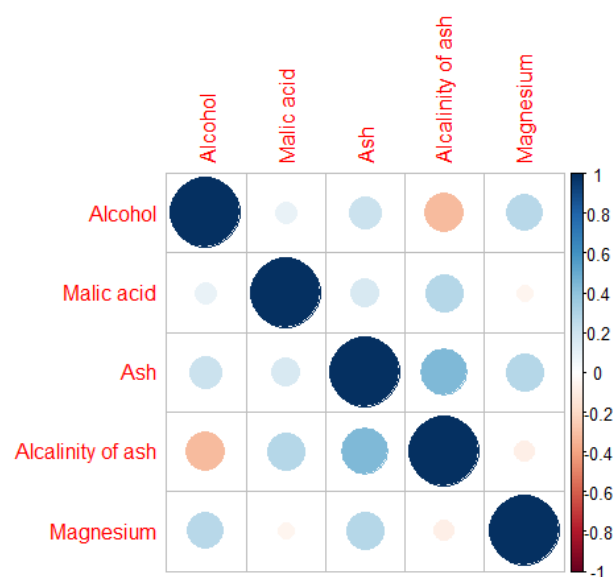
e) Macierz korelacji dla wybranych pięciu zmiennych i jej wizualizacja

```
selected_vars <-
  wine[, c('Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium')]
cor_matrix <- cor(selected_vars)
```

```
library(corrplot)
```

```
corrplot(cor_matrix, method = "circle")
```

```
# e) Macierz korelacji dla wybranych pięciu zmiennych i jej wizualizacja
selected_vars <- wine[, c('Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash', 'Magnesium')]
cor_matrix <- cor(selected_vars)
```



f) Wydrukować histogramy dla trzech różnych zmiennych, przedyskutować wyniki.

f) Histogramy dla trzech różnych zmiennych

```
par(mfrow = c(1, 3)) # ustawienie layoutu na 1 wiersz i 3 kolumny
```

```
hist(wine$Alcohol,
```

```
  main = 'Alcohol',
```

```
  xlab = "",
```

```
  col = 'skyblue')
```

```
hist(
```

```
  wine$`Malic acid`,
```

```
  main = 'Malic acid',
```

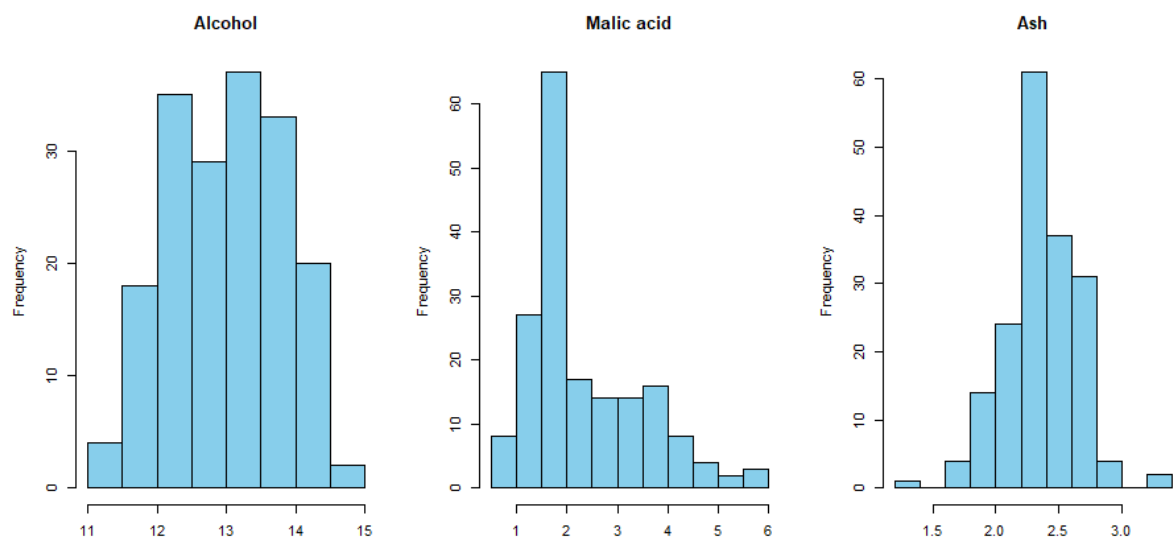
```
  xlab = "",
```

```
  col = 'skyblue')
```

)

```
hist(wine$Ash,  
     main = 'Ash',  
     xlab = '',  
     col = 'skyblue')
```

```
60  
61 # f) Histogramy dla trzech różnych zmiennych  
62 par(mfrow=c(1,3)) # ustawienie layoutu na 1 wiersz i 3 kolumny  
63 hist(wine$Alcohol, main='Alcohol', xlab='', col='skyblue')  
64 hist(wine$Malic acid, main='Malic acid', xlab='', col='skyblue')  
65 hist(wine$Ash, main='Ash', xlab='', col='skyblue')
```



Lab 3

a) Usuń kolumny z wartościami nominalnymi (identyfikatory, itp.) – jeżeli są, inne niż zmienna celu.

Zakładając, że zmienna celu to V1, a pozostałe kolumny to wartości liczbowe nie ma kolumn z wartościami nominalnymi do usunięcia.

Jeśli jednak byłyby takie kolumny, można by je usunąć za pomocą polecenia subset.

Przykład: Jeśli kolumna V2 była by zmienną nominalną, można by ją usunąć następująco:

```
wine <- subset(wine, select = -V2) # Usuń kolumnę V2 z ramki danych wine
```

```
View(wine) # Wyświetl dane po usunięciu kolumny V2
```

b) Zmień nazwy kolumn na nazwy w języku polskim. Nowe nazwy powinny być: krótkie, znaczące, bez polskich znaków i spacji. Wyświetl dane poleceniem View.

```
library(datasets)
```

```
path = "C:\\Users\\petitoff\\Desktop\\repos\\UO\\rok 3\\Wprowadzenie do eksploracji  
danych\\lista3\\" # używając podwójnych ukośników
```

```
setwd(path) ## ustawienie ścieżki
```

```
# Załadowanie danych
```

```
wine <- read.csv('wine\\wine.data', header = FALSE)
```

```
# Oto kod zmieniający nazwy kolumn na nazwy w języku polskim:
```

```
nowe_nazwy <- c(
```

```
  "Klasa",
```

```
  'Alkohol',
```

```
  'Kwas jabłkowy',
```

```
  'Popiół',
```

```
  'Alkalność popiołu',
```

'Magnez',
 'Całkowite fenole',
 'Flawonoidy',
 'Fenole nietrwałe',
 'Proantocyjanidy',
 'Intensywność koloru',
 'Odcień',
 'Stężenie odwiedlane win',
 'Prolina'

The screenshot displays the RStudio interface. The main window shows a data frame with 178 rows and 14 columns. The columns are: Klasa, Alkohol, Kwas jabłkowy, Popiół, Alkalność popiołu, Magnez, Całkowite fenole, Flawonoidy, Fenole nietrwałe, Proantocyjanidy, Intensywność koloru, and Odcień. The console shows the following commands and output:

```
R 4.3.1 - C:\Users\petitoff\Desktop\repos\U003\Wprowadzenie do eksploracji danych\lista11.R #
> names(wine) <- nowe_nazwy
> view(wine)
> nowe_nazwy
```

c) Zmienne o wartościach logicznych (jeżeli są) zapisz jako logiczne (polecenie as.logical).

The screenshot shows the RStudio interface with the 'wine' dataset loaded. The main window displays a data frame with 13 columns: Klasa, Alkohol, Kwas jabłkowy, Popiół, Alkalność popiołu, Magnez, Całkowite fenole, Flavonoidy, Fenole nietrwałe, Proantocyjanidy, Intensywność koloru, and Odcień. The Environment pane on the right shows the 'wine' object with 178 observations and 14 variables. The Console pane at the bottom shows R code for loading the data and converting logical values to numeric.

```
wine$Klasa <- as.logical(wine$Klasa)
```

d) Upewnij się, że zmienne o wartościach liczbowych są typu liczbowego, a jeżeli nie są, to zapisz je jako numeryczne (as.numeric)

```
wprowadzone <- c('Alkohol', 'Kwas jabłkowy', 'Popiół', 'Alkalność popiołu', 'Magnez', 'Całkowite fenole', 'Flawonoidy', 'Fenole nietrwałe', 'Proantocyjanidy', 'Intensywność koloru', 'Odcień', 'Stężenie odwiedlane win', 'Prolina')
```

```
for (zmienna in wprowadzone) {
  if (class(wine[[zmienna]]) != "numeric") {
    wine[[zmienna]] <- as.numeric(wine[[zmienna]])
  }
}
```

e) Zmienną celu zapisz jako mającą wartości nominalne (polecenie as.factor).

```
wine$Klasa <- as.factor(wine$Klasa)
```

f) Policz brakujące wartości. Jeżeli są, to dla kolumn o wartościach liczbowych zastąp je wartościami średnimi dla kolumn.

```
for (kolumna in kolumny_numeryczne) {  
  brakujace <- is.na(wine[[kolumna]])  
  if (sum(brakujace) > 0) {  
    srednia <- mean(wine[[kolumna]], na.rm = TRUE)  
    wine[[kolumna]][brakujace] <- srednia  
  }  
}
```

```
write.csv(wine, file = 'wyniki.csv', row.names = FALSE)
```

[illegible]

Lab 4

a) Obliczy i narysuje macierz korelacji zmiennych (bez zmiennej celu wyznaczającej klasy)

```
library(ggplot2)
```

```
library(reshape2)
```

```
correlation_matrix <- function(df, threshold) {
```

```
  # Obliczanie macierzy korelacji
```

```
  cor_matrix <- cor(df)
```

```
  # Rysowanie macierzy korelacji
```

```
  melted_cor_matrix <- melt(cor_matrix)
```

```
  plot <- ggplot(data = melted_cor_matrix, aes(x=Var1, y=Var2, fill=value)) +
```

```
    geom_tile() +
```

```
    scale_fill_gradient2(low = "blue", high = "red", mid = "white",
```

```
                        midpoint = 0, limit = c(-1,1), space = "Lab",
```

```
                        name="Pearson\nCorrelation") +
```

```
    theme_minimal() +
```

```
    theme(axis.text.x = element_text(angle = 45, vjust = 1,
```

```
        size = 12, hjust = 1),
```

```
        axis.text.y = element_text(size = 12)) +
```

```
    coord_fixed()
```

```
  # Zapisywanie rysunku do pliku
```

```
  ggsave("correlation_matrix.png", plot)
```

```
  # Wypisywanie par zmiennych o korelacji większej niż zadany próg
```

```
  cor_pairs <- subset(melted_cor_matrix, abs(value) > threshold & Var1 != Var2)
```

```
  # Usuwanie powtórzeń
```



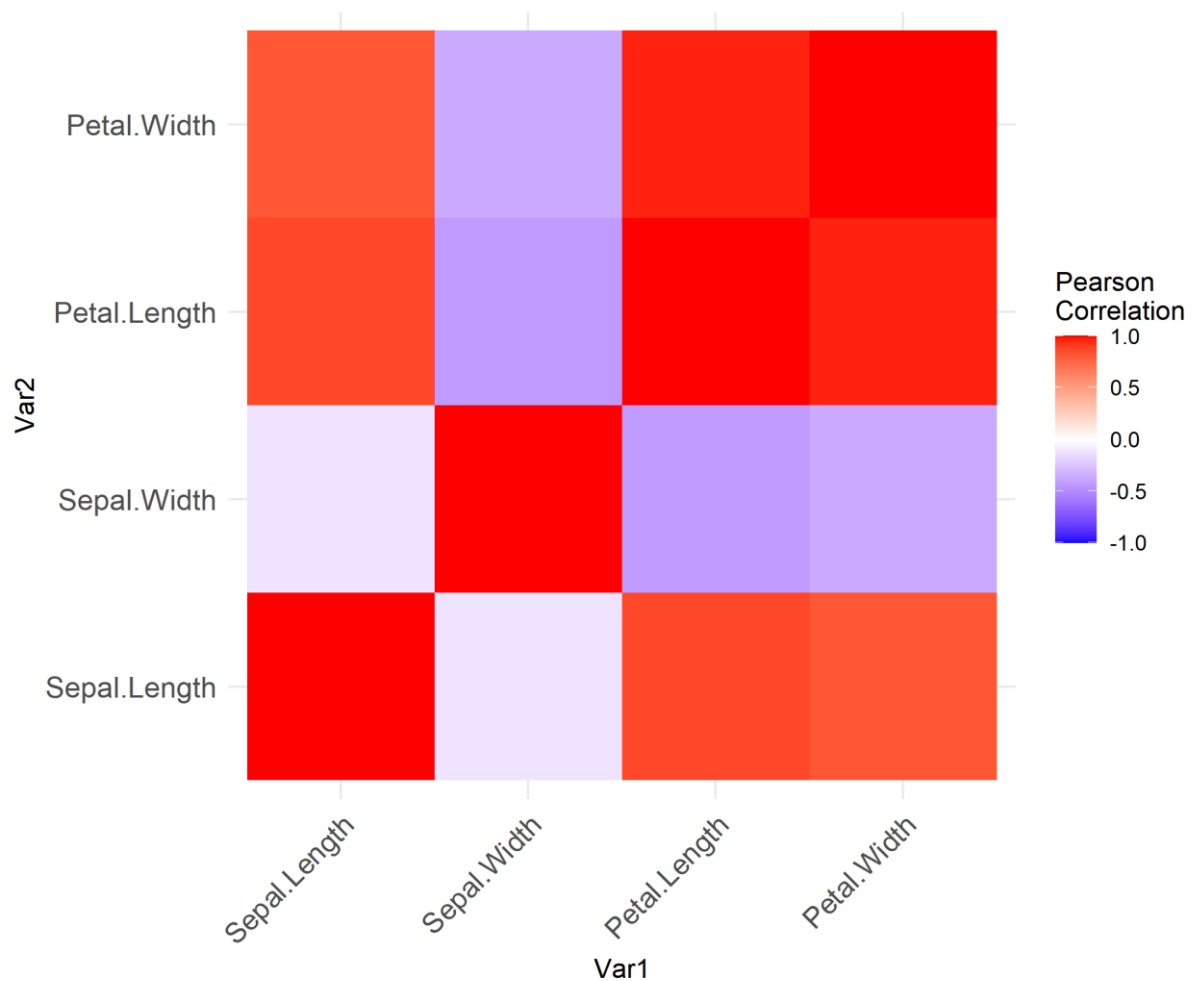
```
cor_pairs <- cor_pairs[!duplicated(t(apply(cor_pairs[,c("Var1", "Var2")], 1, sort))),]  
  
return(cor_pairs)  
}
```

```
# Załadowanie zestawu danych iris  
data(iris)
```

```
# Usunięcie kolumny Species (bo to jest nasza zmienna celu)  
df <- iris[,-5]
```

```
# Użycie funkcji na df z progiem 0.5  
correlation_matrix(df, 0.5)
```

b) zapisze rysunek macierzy korelacji do pliku



c) wypisze pary (nazwy) zmiennych o korelacji większej niż zadany próg oraz odpowiadające im wartości korelacji (wartość progu powinna być argumentem funkcji).

- Proszę uwzględnić ujemne wartości korelacji; czyli przyjmujemy, że np. korelacja równa -0.95 jest powyżej progu 0.9, bo jest to silna korelacja, tylko ujemna (wraz ze wzrostem wartości jednej cechy następuje spadek wartości drugiej cechy).
- Pary proszę wypisać bez powtórek (czyli jeżeli mamy już korelację cechy x z cechą y, to nie wypisujemy korelacji cechy y z x).

Wyniki z konsoli:

```
Saving 7 x 7 in image
      Var1      Var2      value
3 Petal.Length Sepal.Length 0.8717538
4  Petal.Width Sepal.Length 0.8179411
12 Petal.Width Petal.Length 0.9628654
>
```