

Zadanie 1

a) Nazwa zbioru: Wine

```
> 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
> # 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 kategoryczna. 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 absorpcji 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”)

```
R43.1 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:

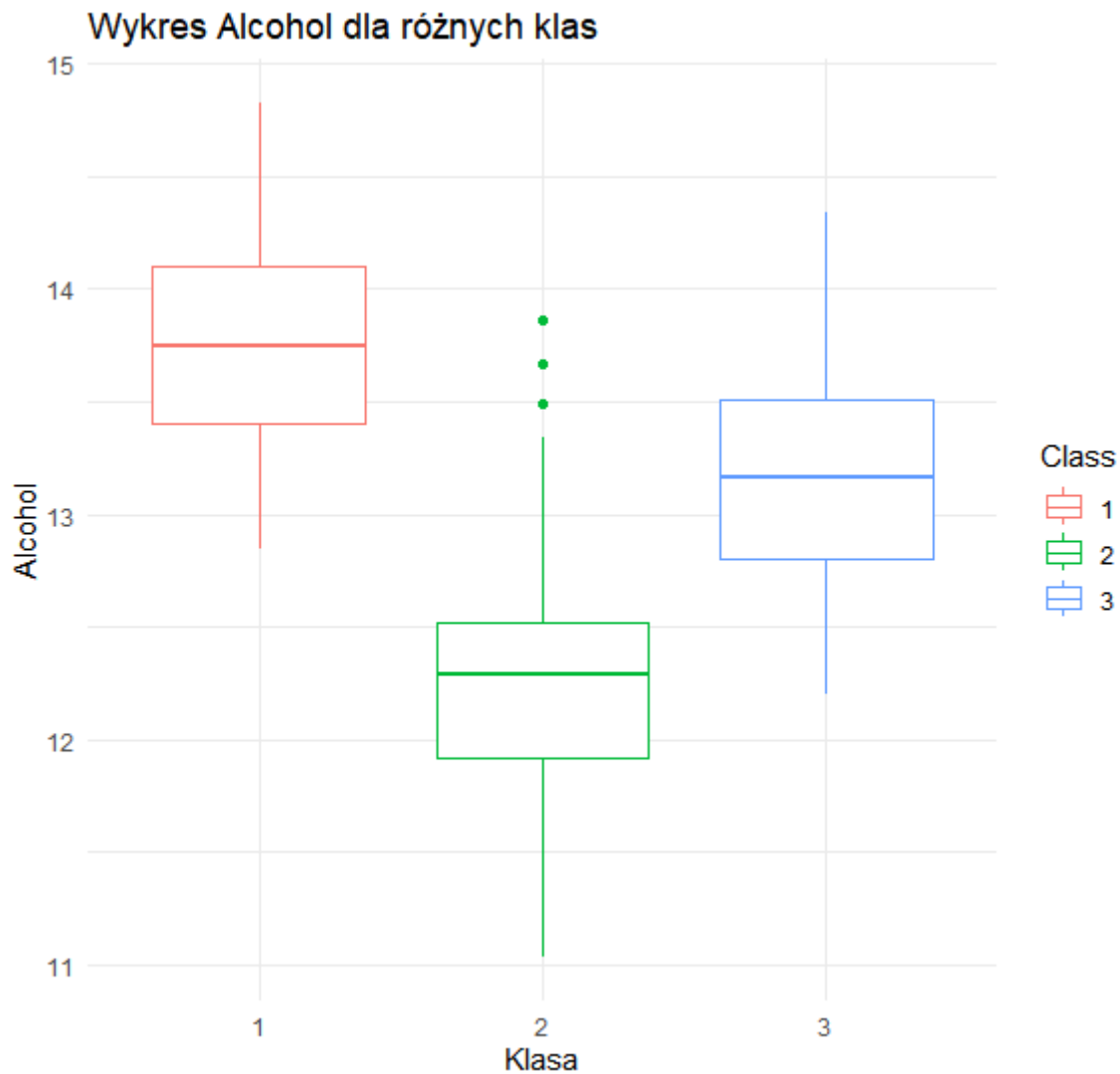
```
R 4.3.1 - C:/Users/petioff/Desktop/repos/UO/rok 3/Wprowadzenie do eksploracji danych/lista1/zadanie2/
> library(datasets)
>
> path = "C:\\Users\\petioff\\Desktop\\repos\\UO\\rok 3\\Wprowadzenie do eksploracji danych\\lista1\\zadanie2" # używając podwójnych ukośników
> setwd(path) ## ustawienie ścieżki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

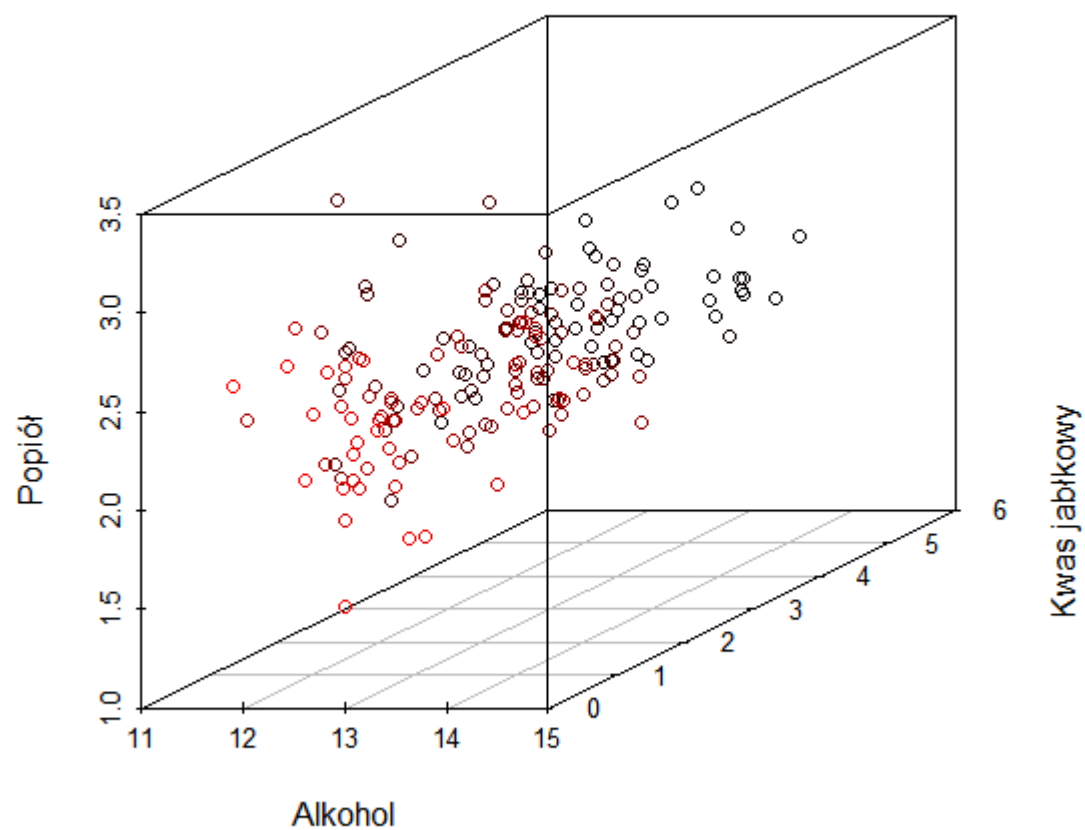
      Proanthocyanins      Color intensity      Hue      OD280/OD315 of diluted wines      Proline
Min.   :0.410   Min.   : 1.280   Min.   :0.4800   Min.   :1.270   Min.   : 278.0
1st Qu.:1.250   1st Qu.: 3.220   1st Qu.:0.7825   1st Qu.:1.938   1st Qu.: 500.5
Median :1.555   Median : 4.690   Median :0.9650   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)
> |
```