

Zajęcia 7

Analiza wielowymiarowa

Zadanie 1. Zbiór danych `USArrests`, zawiera informacje dotyczące ilości: morderstw, napadów, gwałtów przypadających na 100 tys. osób w poszczególnych stanach USA w roku 1973 oraz procent ludności mieszkającej w miastach.

- (a) Wykonaj analizę składowych głównych. Ile jest wszystkich składowych głównych? **Wskazówka:** `princomp()`

```
Standard deviations:
  Comp.1    Comp.2    Comp.3    Comp.4
82.890847 14.069560  6.424204  2.457837

4 variables and 50 observations.
```

- (b) Jaki procent zmienności (całkowitej wariancji) wyjaśniają dwie pierwsze składowe? **Wskazówka:** `summary()`

```
Importance of components:
              Comp.1      Comp.2      Comp.3      Comp.4
Standard deviation  82.8908472 14.06956001 6.424204055 2.4578367034
Proportion of Variance 0.9655342 0.02781734 0.005799535 0.0008489079
Cumulative Proportion 0.9655342 0.99335156 0.999151092 1.0000000000
```

- (c) Narysuj wykres osypiska i podaj jego interpretację. **Wskazówka:** `plot()` **Odpowiedź:** patrz Rysunek 1

Zadanie 2. Rozważamy zbiór danych `USArrests` (patrz Zadanie 1).

- (a) Wykonaj hierarchiczną analizę skupień dla tych danych. **Wskazówka:** `hclust()`, `dist()`

```
Cluster method : complete
Distance       : euclidean
Number of objects: 50
```

- (b) Jaka liczba skupień wydaje się najbardziej sensowna? Narysuj dendrogram wraz z proponowanym podziałem na skupienia. **Wskazówka:** `plot()`, `rect.hclust()` **Odpowiedź:** patrz Rysunek 2
- (c) Przeprowadź analogiczną analizę używając odległości taksówkowej jako odległości pomiędzy obserwacjami i metody średniego wiązania jako metody łączenia skupień. Czy wyniki się istotnie zmieniły? **Odpowiedź:** patrz Rysunek 3

```
Cluster method : average
Distance       : manhattan
Number of objects: 50
```

Zadanie 3. Zbiór danych `iris` zawiera informacje na temat czterech cech trzech gatunków irysa.

- (a) Wykonaj liniową analizę dyskryminacyjną. **Wskazówka:** `lda()` z pakietu `MASS`

```
Prior probabilities of groups:
  setosa versicolor virginica
0.3333333 0.3333333 0.3333333

Group means:
      Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa           5.006         3.428         1.462         0.246
```

```

versicolor      5.936      2.770      4.260      1.326
virginica       6.588      2.974      5.552      2.026

```

Coefficients of linear discriminants:

```

              LD1      LD2
Sepal.Length 0.8293776 0.02410215
Sepal.Width  1.5344731 2.16452123
Petal.Length -2.2012117 -0.93192121
Petal.Width  -2.8104603 2.83918785

```

Proportion of trace:

```

      LD1      LD2
0.9912 0.0088

```

- (b) Jaki błąd klasyfikacji (metoda ponownego podstawiania) uzyskujemy? **Wskazówka:** `predict()`

```
0.02
```

- (c) Do którego gatunku i z jakim prawdopodobieństwem a posteriori należy zaklasyfikować kwiat, dla którego `Sepal.Length = 5.1`, `Sepal.Width = 3.5`, `Petal.Length = 1.3`, `Petal.Width = 0.3`? **Wskazówka:** `predict()`

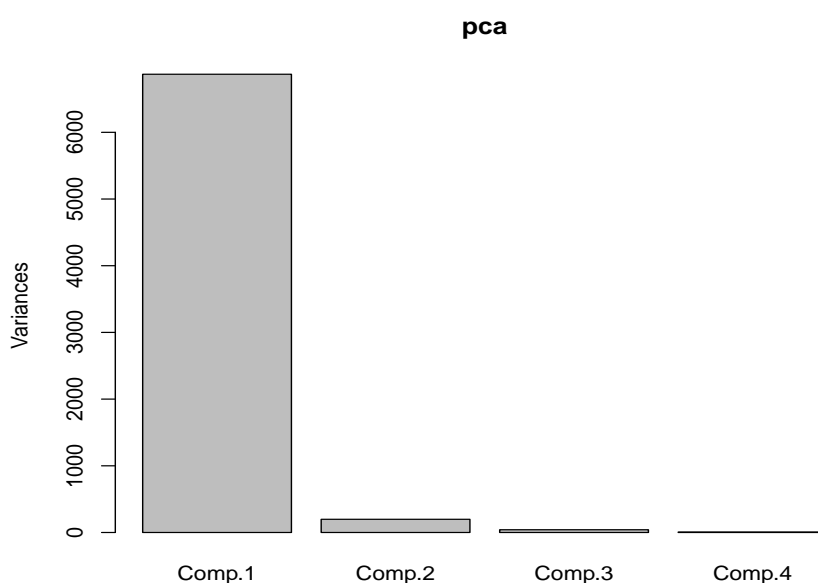
```

$class
[1] setosa
Levels: setosa versicolor virginica

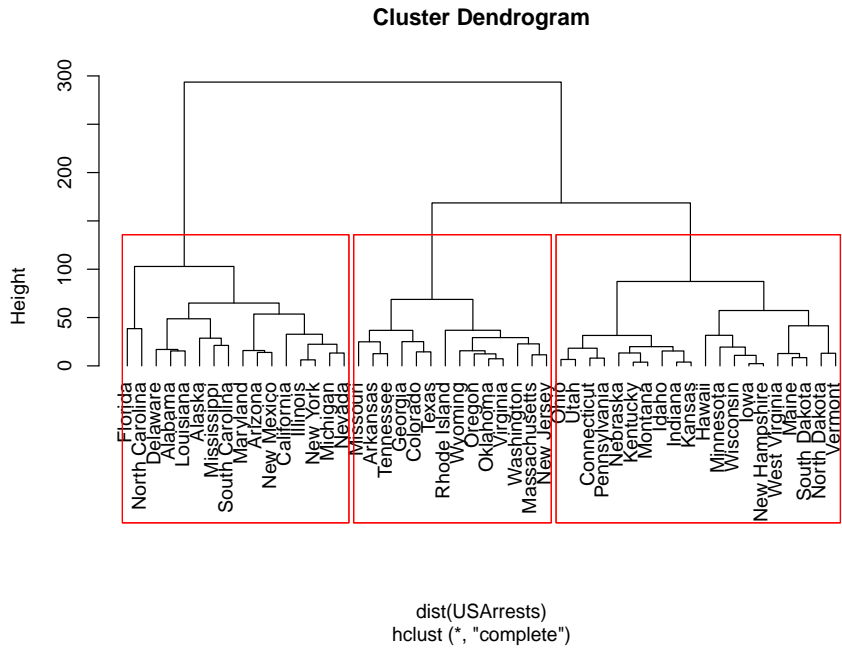
$posterior
      setosa  versicolor   virginica
1          1 4.850575e-22 6.605032e-42

$x
      LD1      LD2
1 8.000875 0.6775315

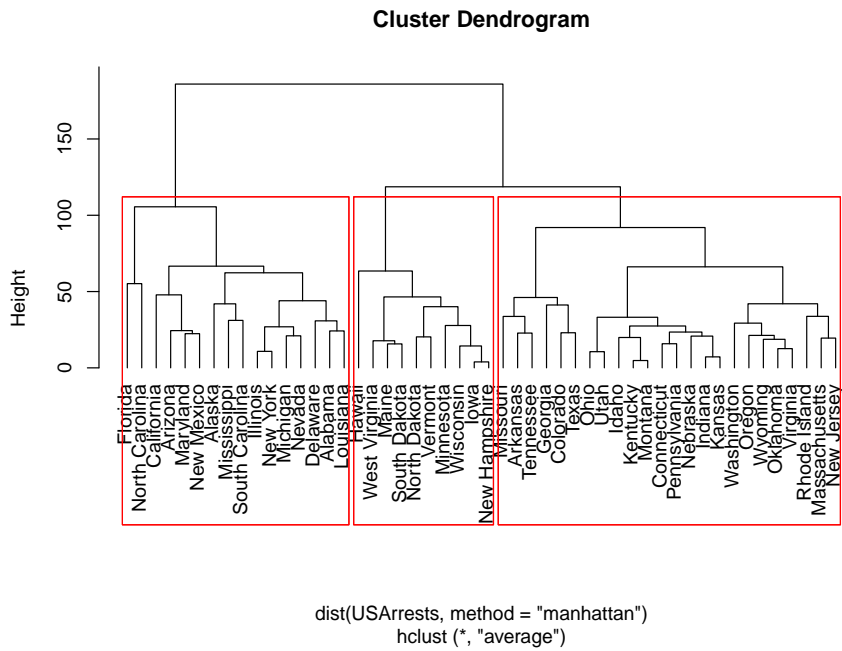
```



RYSUNEK 1. Wykres do Zadania 1 (c).



RYSUNEK 2. Wykres do Zadania 2 (b).



RYSUNEK 3. Wykres do Zadania 2 (c).