

# Raport 4

## Eksploracja danych

Mikołaj Langner, Marcin Kostrzewa  
nr albumów: 255716, 255749

2021-05-28

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>1</b>
1.1	b) . . . . .	2
<b>2</b>	<b>Zadanie 2</b>	<b>6</b>
2.1	Wizualizacja wyników grupowania ( $K = 3$ ) . . . . .	6
2.2	Ocena jakości grupowania . . . . .	10

## 1 Wstęp

Niniejszy raport zawiera rozwiązania zadań z listy 4.

W zadaniu pierwszym zastosujemy zaawansowane metody klasyfikacji:

- bagging,
- boosting,
- random forest,
- metodę wektorów nośnych (SVM),

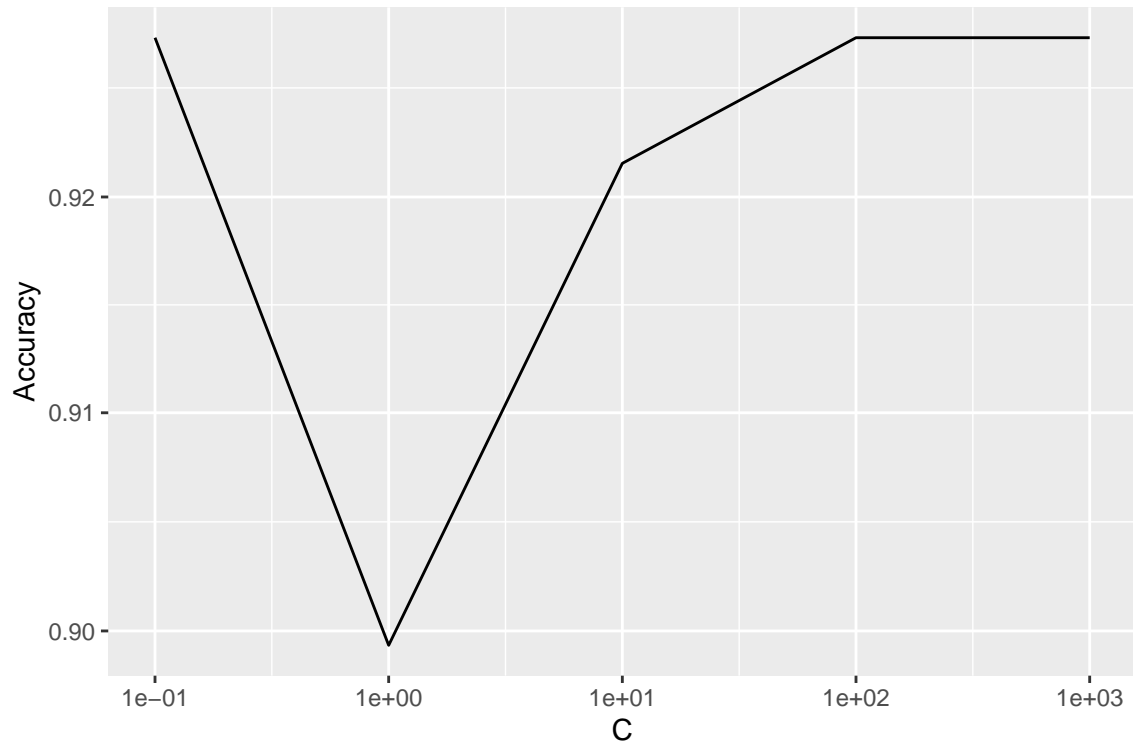
W zadaniu drugim badamy jakość

### 1.0.1 Wnioski

uwagii, tabelka, wniosek, że klasyfikatory wzmocnione radzą sobie lepiej

### 1.1 b)

```
wine <- wine %>% select(c(Type, Alcohol, Flavanoids))
```

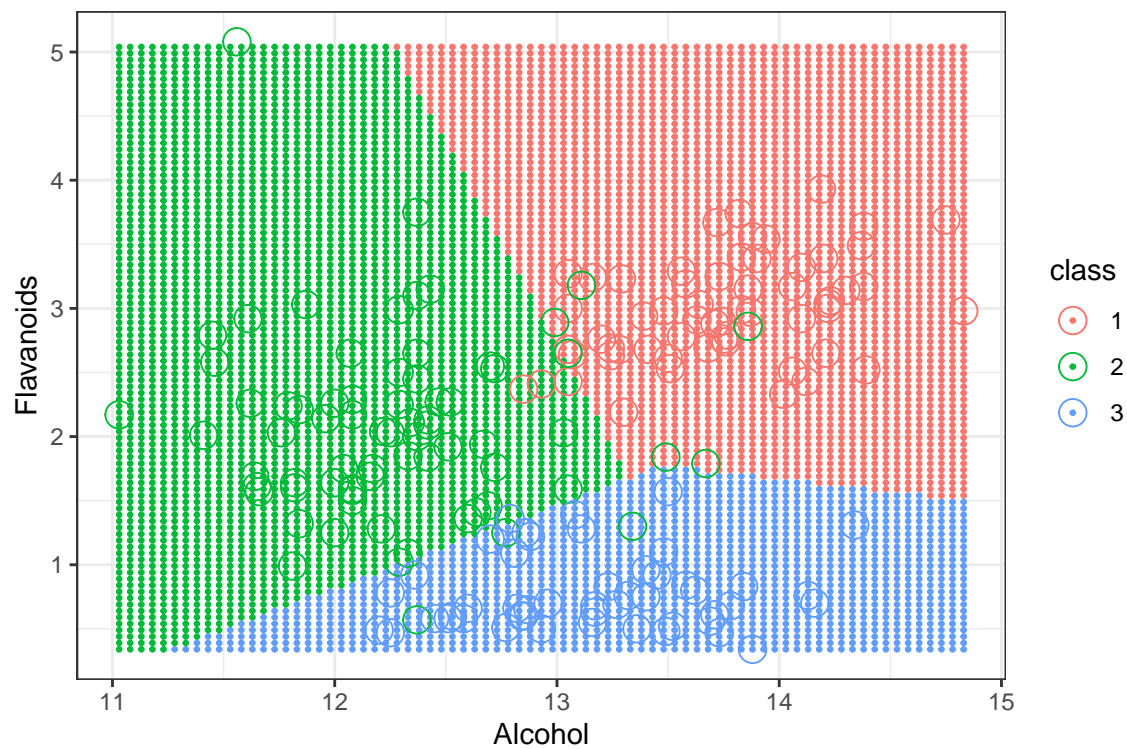


Rysunek 1: Dokładność klasyfikatora od parametru kosztu

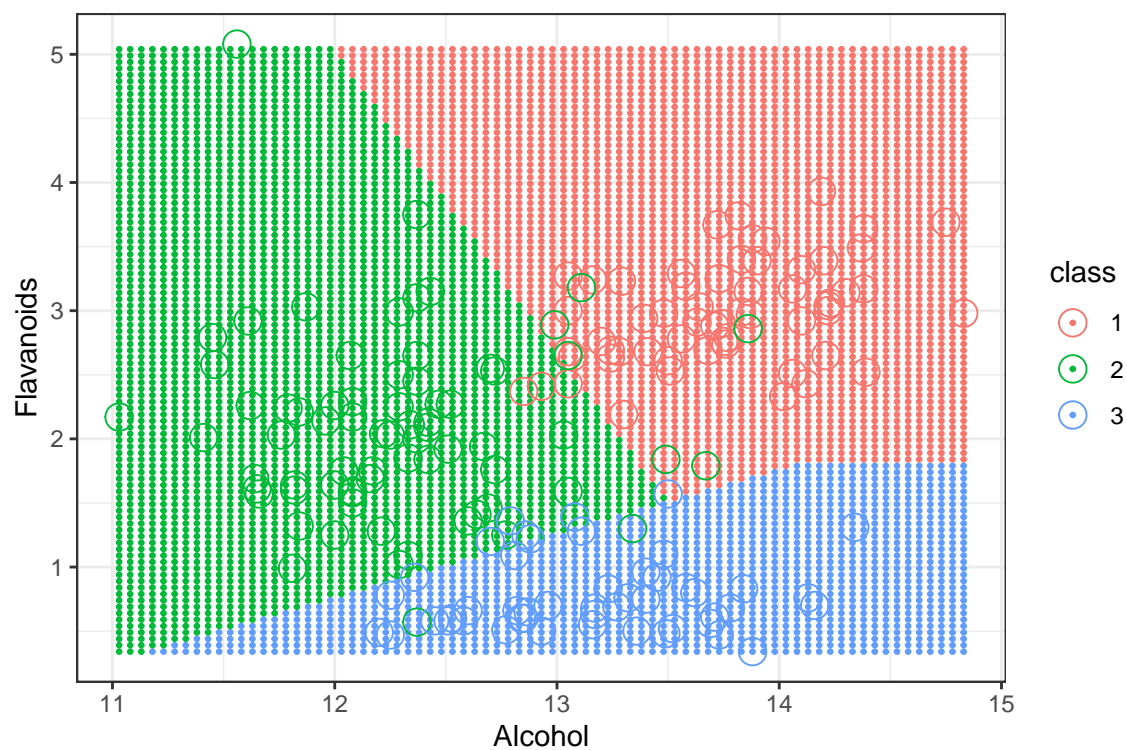
```
## Setting default kernel parameters
## Setting default kernel parameters
## Setting default kernel parameters
## Setting default kernel parameters
## Setting default kernel parameters
```

linear	polynomial	radial
0.926	0.933	0.949

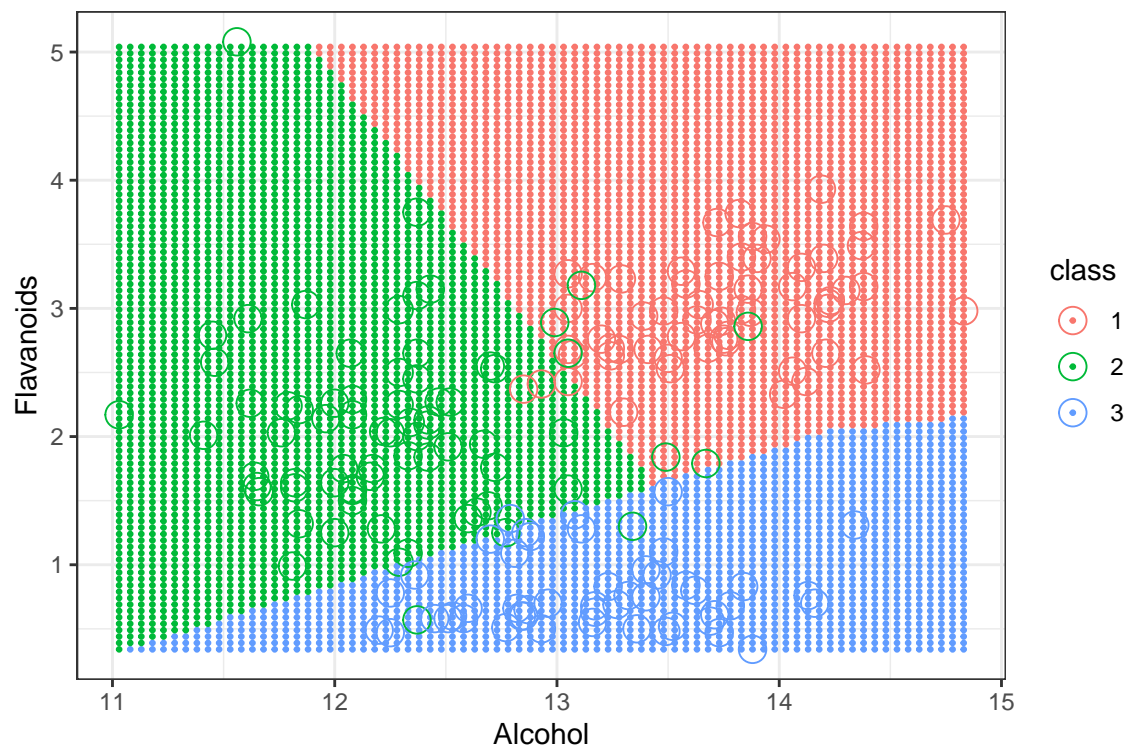
Tabela 1: Porównanie klasyfikatorów dla różnych jąder



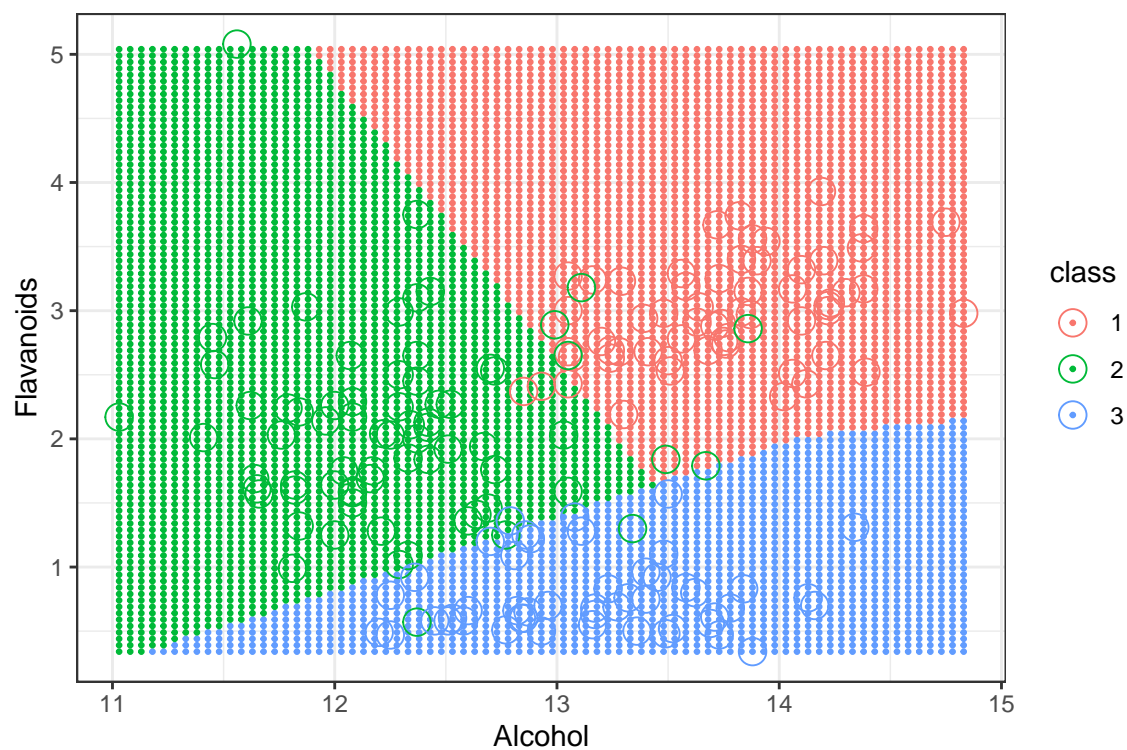
Rysunek 2: Obszary decyzyjne dla  $C = 0.1$



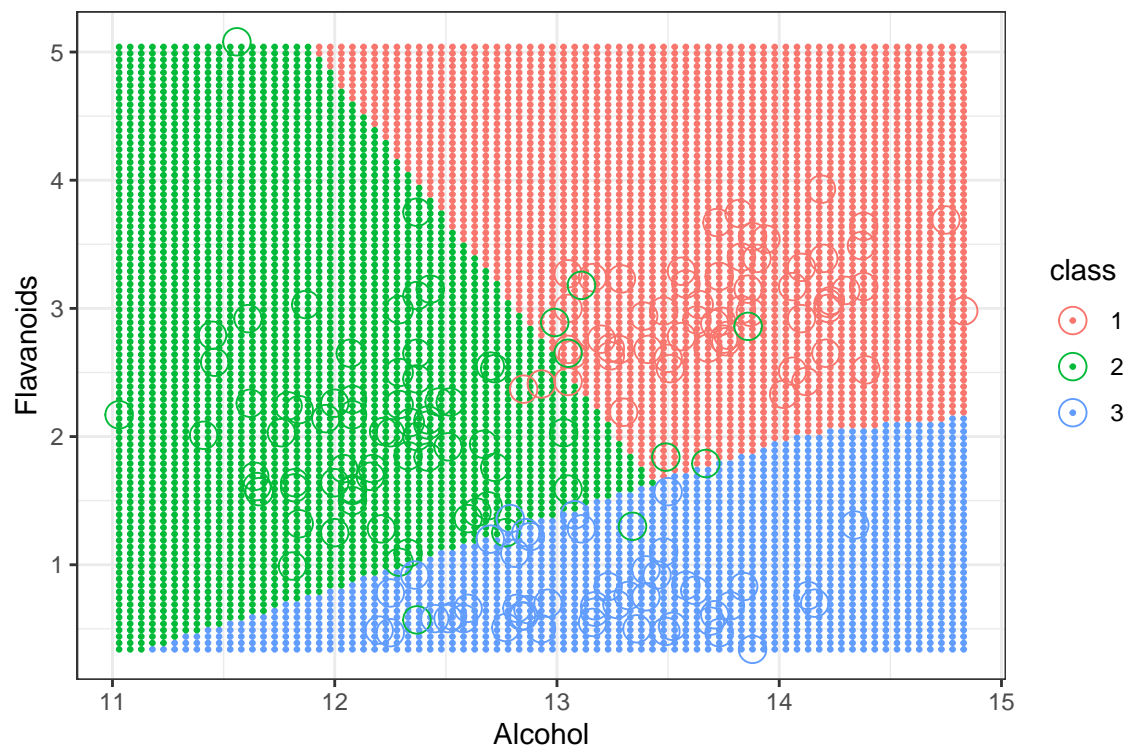
Rysunek 3: Obszary decyzyjne dla  $C = 1$



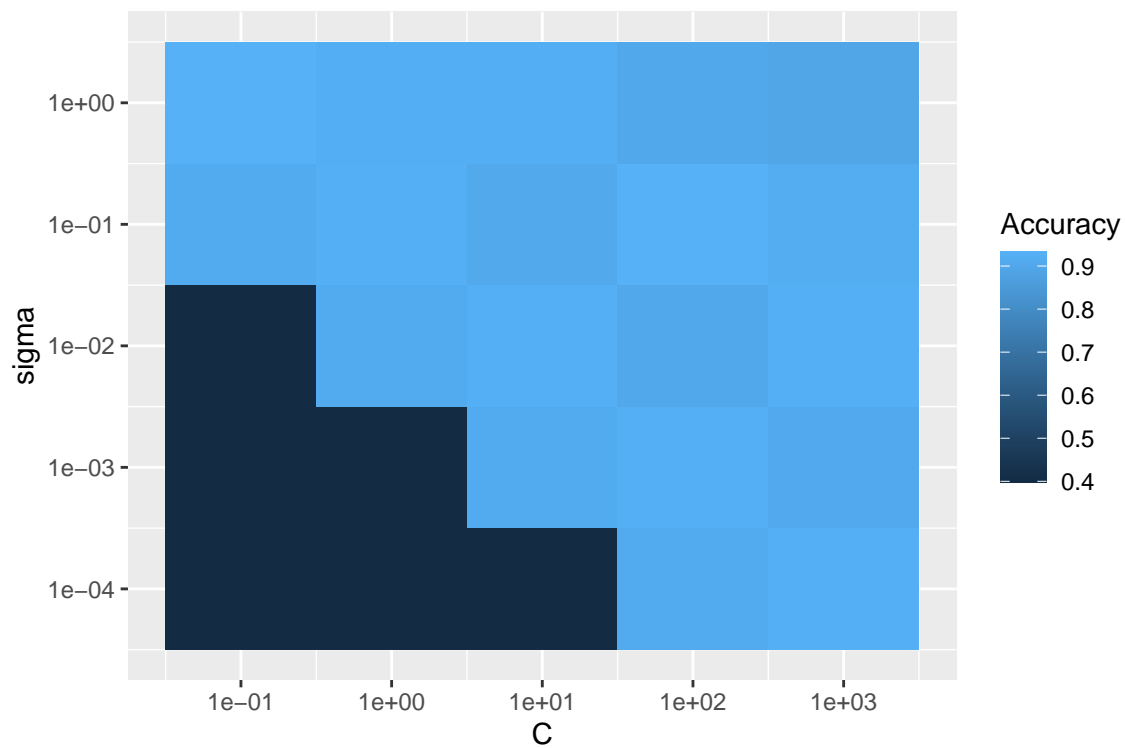
Rysunek 4: Obszary decyzyjne dla  $C = 10$



Rysunek 5: Obszary decyzyjne dla  $C = 100$



Rysunek 6: Obszary decyzyjne dla  $C = 1000$



Rysunek 7: Mapa ciepła dokładności klasyfikatora

sigma	C
1.00	0.10

Tabela 2: Parametry dla najlepszego klasyfikatora

## 2 Zadanie 2

W tym zadaniu zastosujemy algorytmy analizy skupień do wyznaczenia klastrów dla zbioru `wine`, ocenimy ich skuteczność i porównamy je ze sobą. Sięgnijmy po dwa algorytmy: PAM i AGNES.

### 2.1 Wizualizacja wyników grupowania ( $K = 3$ )

Najpierw wyznaczmy macierz niepodobieństwa dla naszych danych.

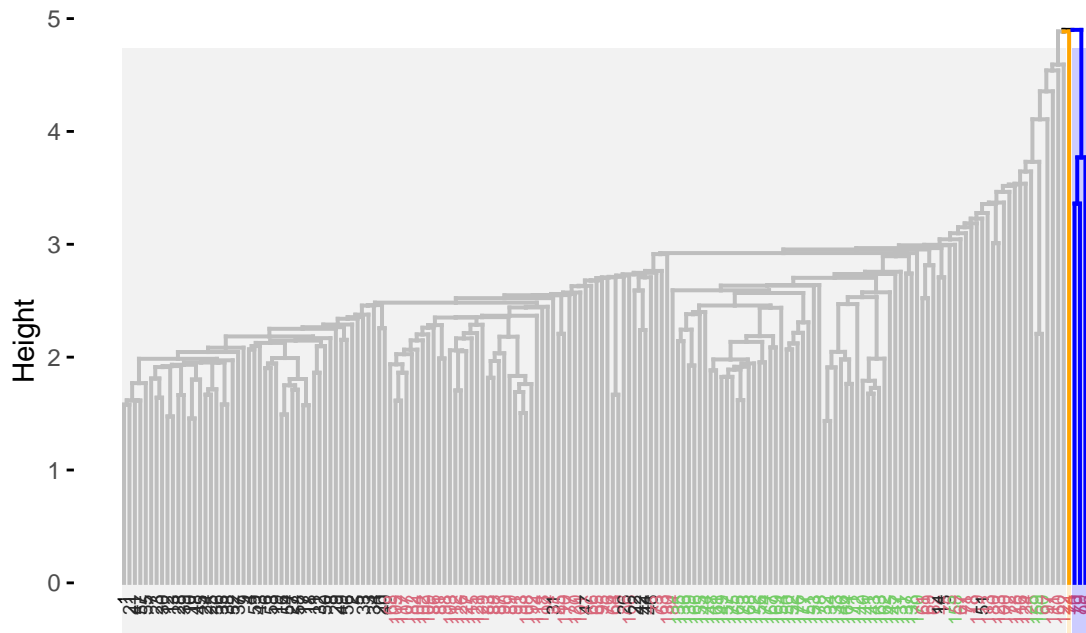
```
data(wine)
wine.subset = wine[, -1]
diss.matrix <- daisy(wine.subset, stand=TRUE) %>% as.matrix
group.colors <- as.numeric(wine$Type)
```

Przyjrzyjmy się najpierw jakie wyniki daje nam zastosowanie algorytmu PAM.

Zobaczmy teraz, jak poradził sobie algorytm AGNES z single-linkage.

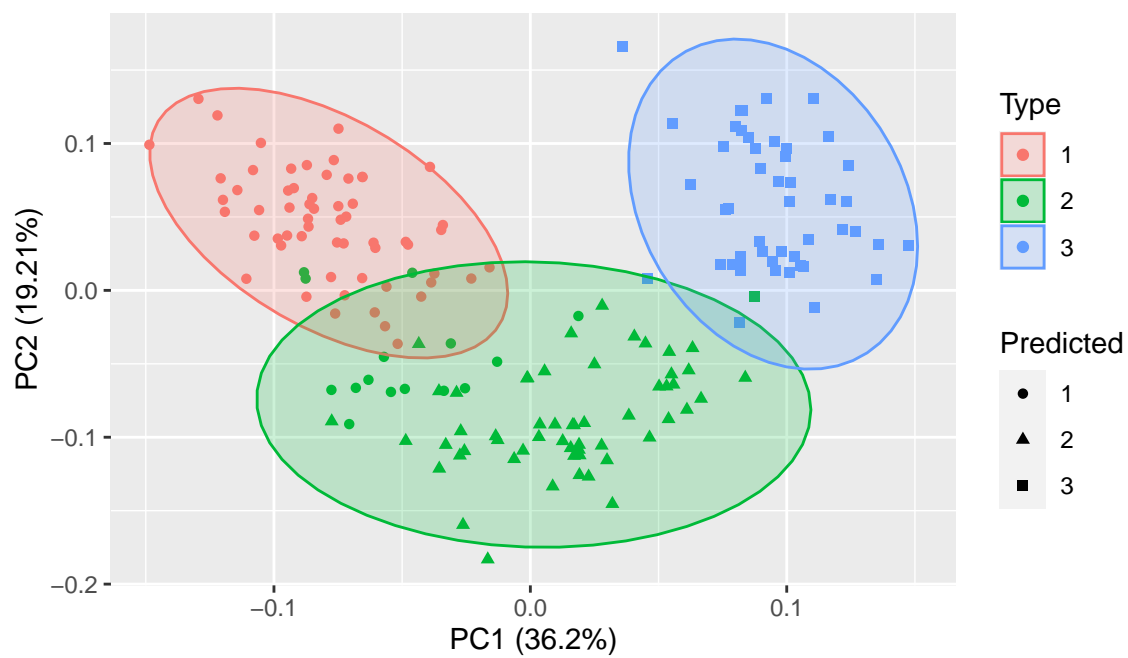
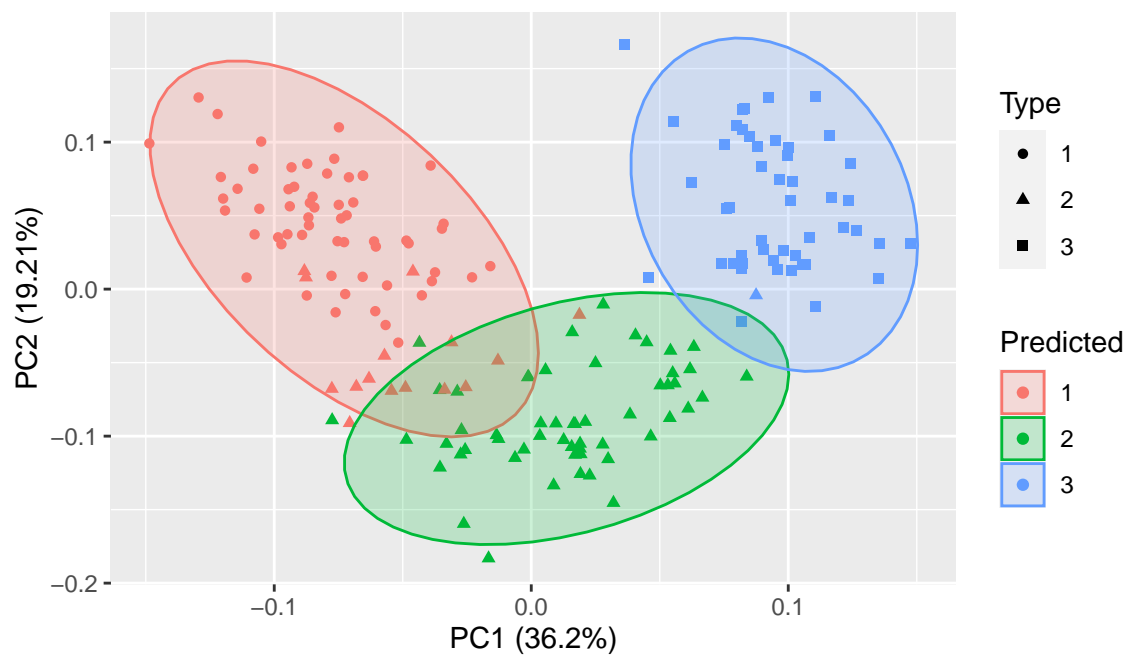
Zobaczmy jak wygląda dendrogram dla tego modelu.

#### Partycja na 3 skupienia a rzeczywiste klasy – single-linkage

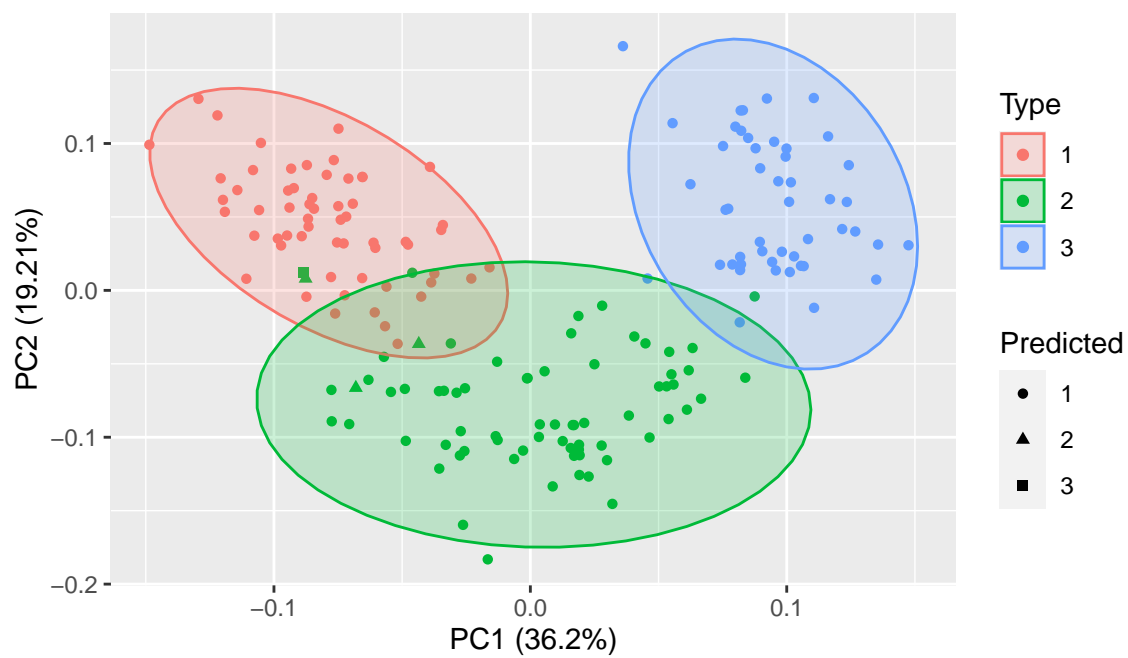
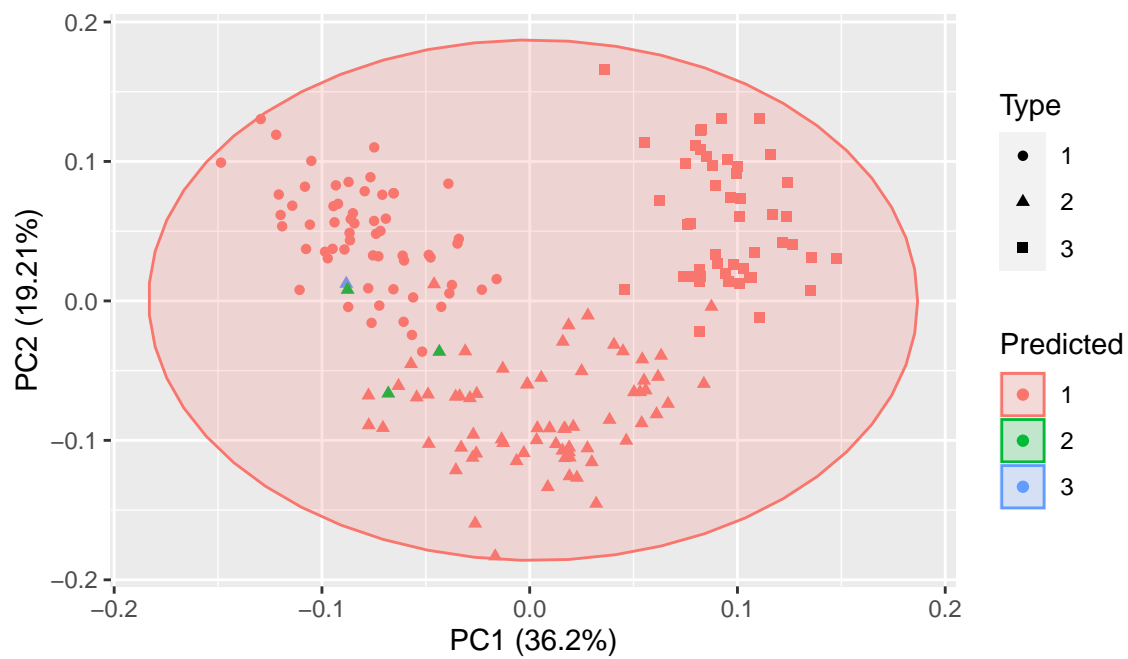


żej wyniki dla algorytmu AGNES z complete-linkage.



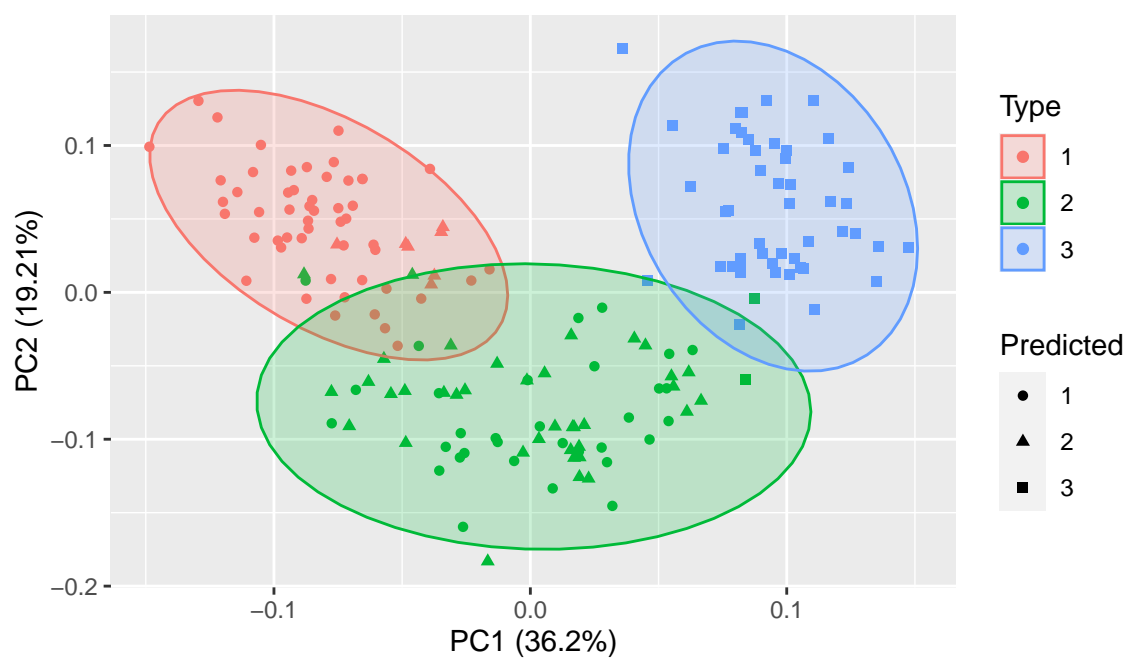
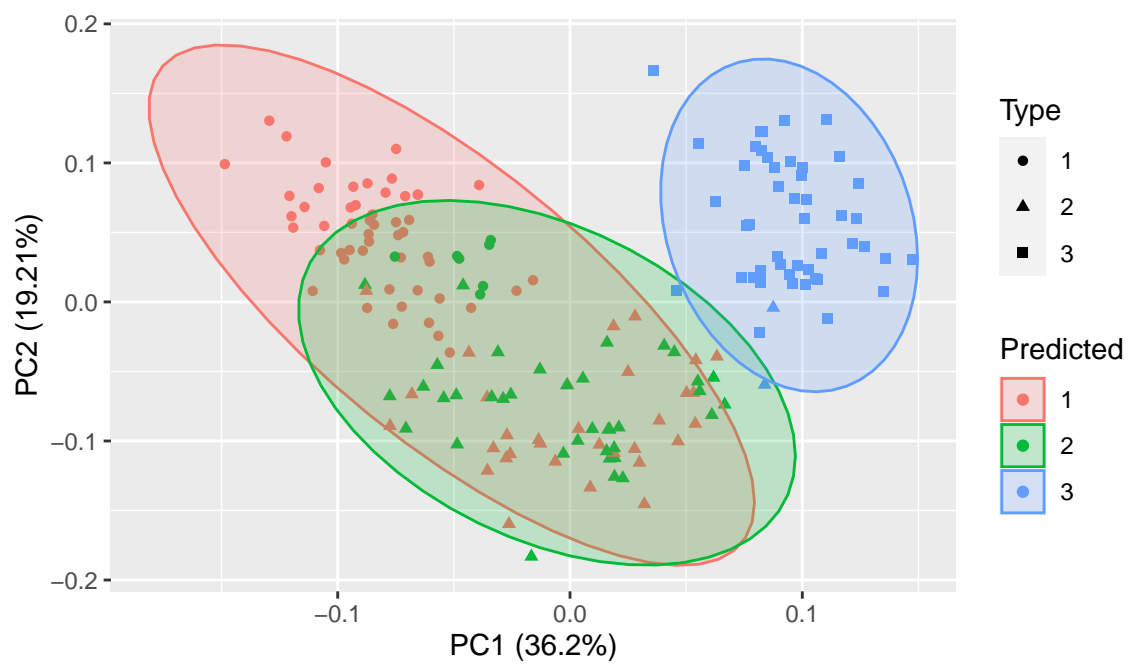


Rysunek 8: Skupienia dla metody PAM



Rysunek 9: Skupienia dla metody AGNES z single-linkage

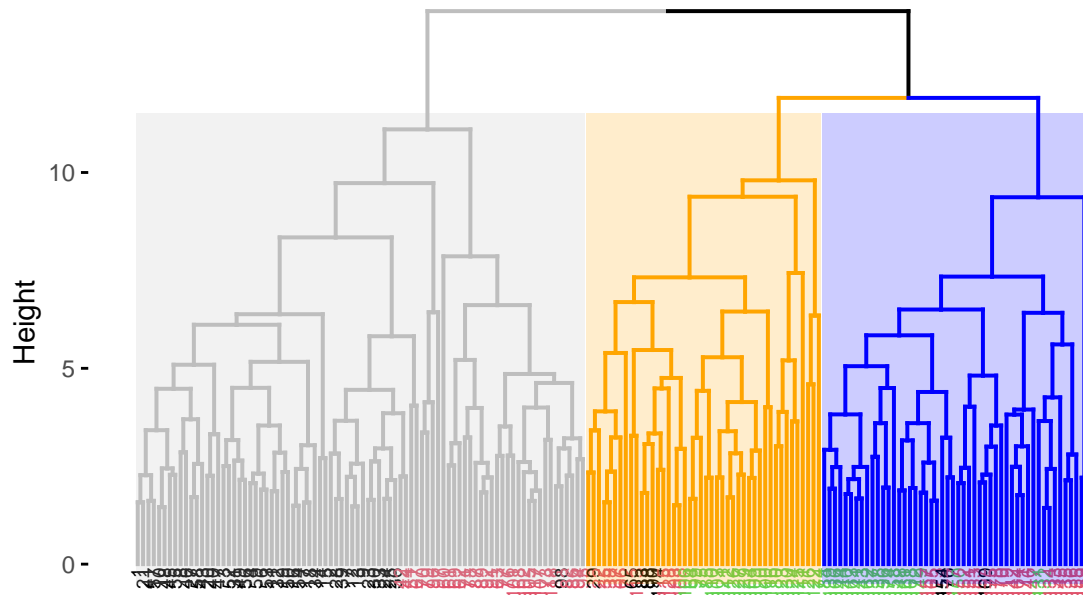




Rysunek 10: Skupienia dla metody AGNES z complete-linkage

Zobaczmy jak wygląda dendrogram w tym przypadku.

### Partycja na 3 skupienia a rzeczywiste klasy – complete-linkage



Rysunek 11: Dendrogram dla complete-linkage.

## 2.2 Ocena jakości grupowania

W tej części zadania, porównamy ze sobą algorytmy, jakość uzyskanego dzięki nim grupowania w zależności od przyjętej ilości skupień. Wykorzystamy wskaźniki wewnętrzne, jak i zewnętrzne.

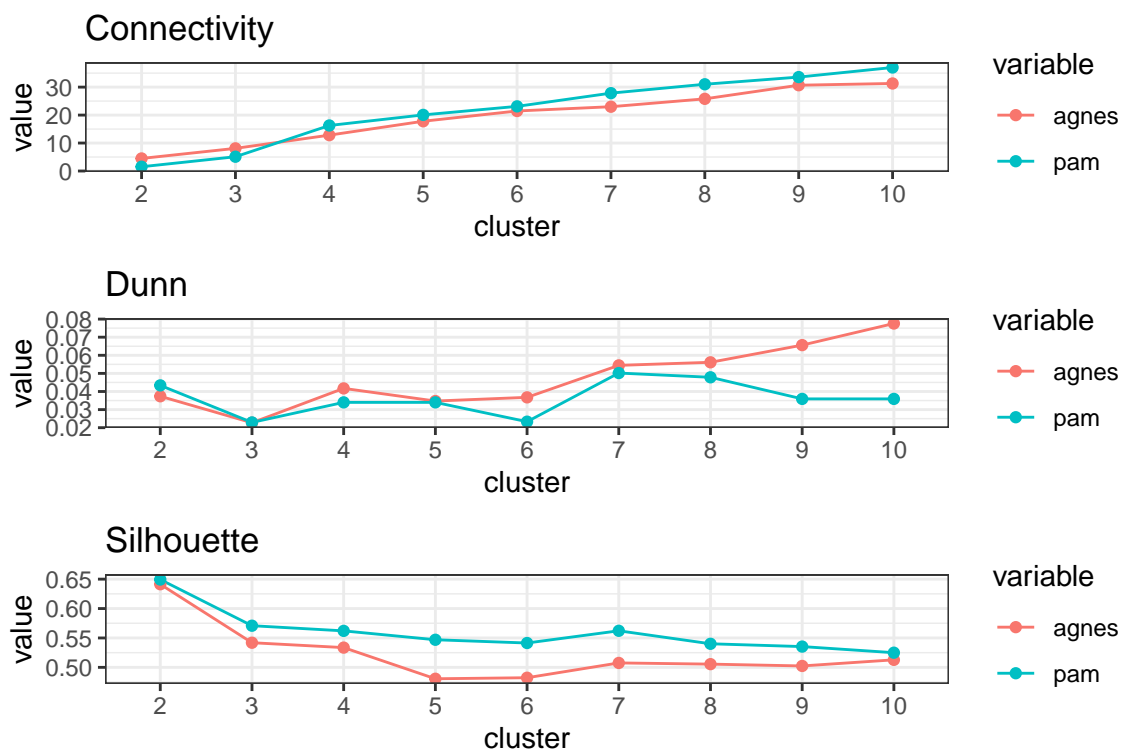
### 2.2.1 Wskaźniki wewnętrzne

### 2.2.2 Wskaźniki zewnętrzne

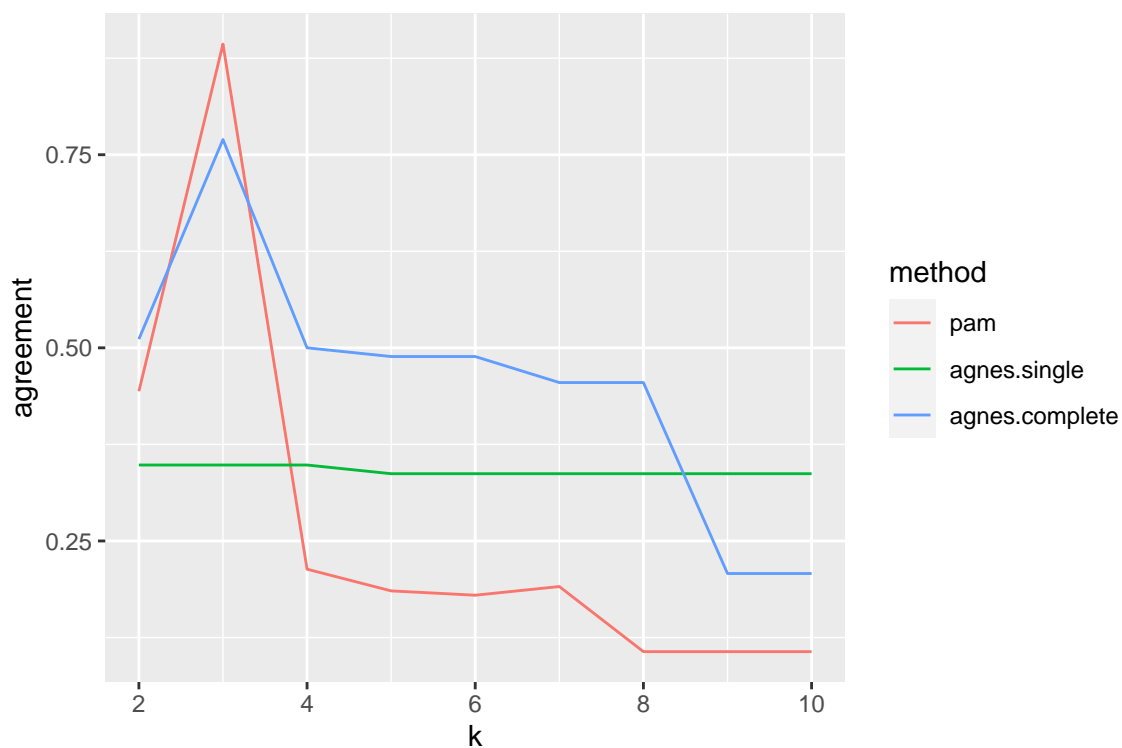
## Cases in matched pairs: 80.9 %

## 1 2 3

## 1 2 3



Rysunek 12: Wskaźniki wewnętrzne dla PAM i AGNES z complete-linkage



Rysunek 13: Porównanie wskaźników zewnętrznych

	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8	K = 9	K = 10
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.93
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.93	1.75
3	0.00	0.00	0.00	0.00	0.00	0.00	1.93	1.75	1.75
4	0.00	0.00	0.00	0.00	0.00	1.64	1.75	1.75	2.05
5	0.00	0.00	0.00	0.00	1.64	1.64	1.75	1.94	2.03
6	0.00	0.00	0.00	1.64	1.64	1.66	1.94	1.94	2.17
7	0.00	0.00	1.64	1.64	1.71	1.94	1.94	2.35	1.93
8	0.00	2.21	1.64	1.71	1.71	2.35	2.35	1.93	2.03
9	2.15	2.14	2.21	1.71	1.93	1.93	1.93	4.36	4.90
10	2.15	2.14	2.76	2.76	1.93	1.93	1.93	1.93	1.93

Tabela 3: Seperacja w skupiskach

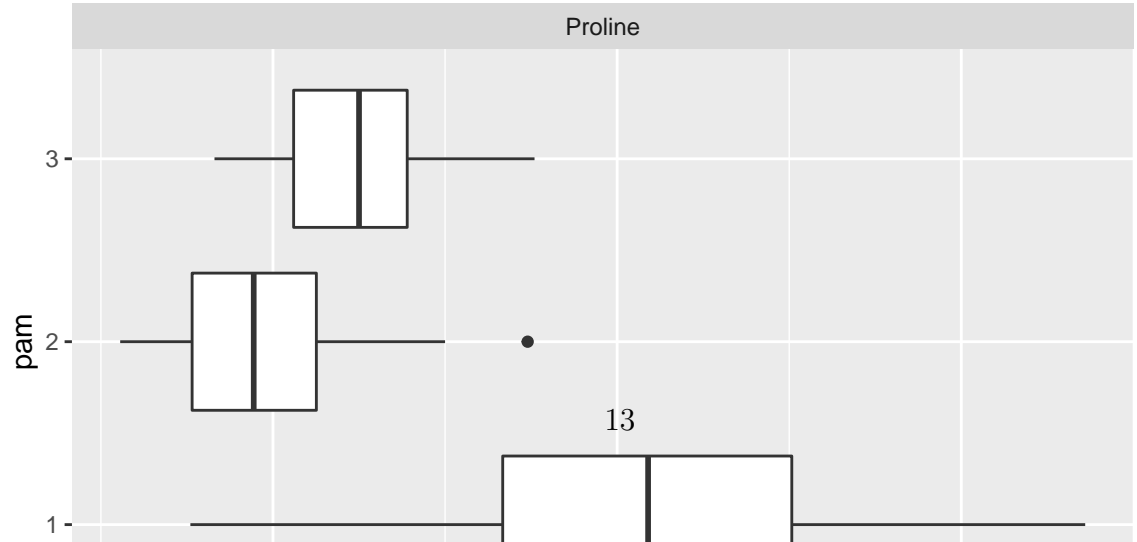
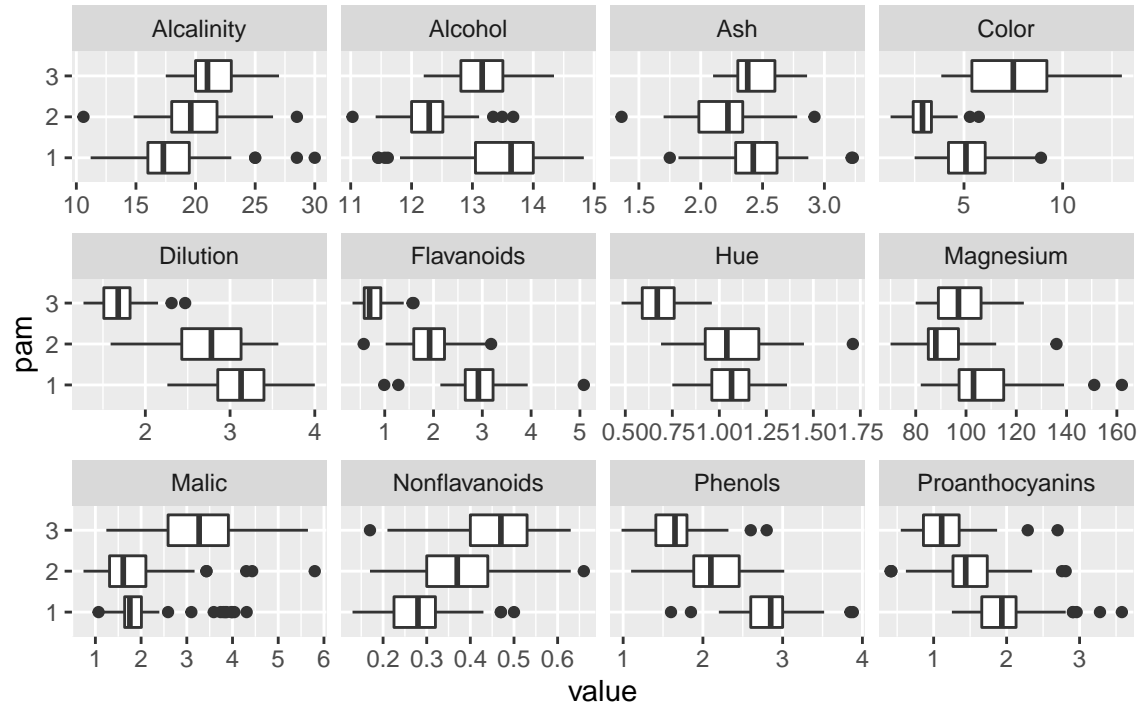
	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8	K = 9	K = 10
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.02
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.02	6.35
3	0.00	0.00	0.00	0.00	0.00	0.00	8.87	6.35	7.08
4	0.00	0.00	0.00	0.00	0.00	9.16	6.35	8.22	9.80
5	0.00	0.00	0.00	0.00	9.16	9.52	8.22	6.57	8.91
6	0.00	0.00	0.00	9.16	9.46	9.85	6.57	8.91	7.64
7	0.00	0.00	9.16	9.86	10.44	6.57	9.85	6.87	7.12
8	0.00	11.26	9.86	10.44	7.64	6.87	6.87	7.12	4.49
9	11.37	10.90	10.90	7.64	7.12	7.12	7.12	7.06	4.13
10	11.17	9.03	9.03	9.37	7.98	7.98	7.98	7.98	7.98

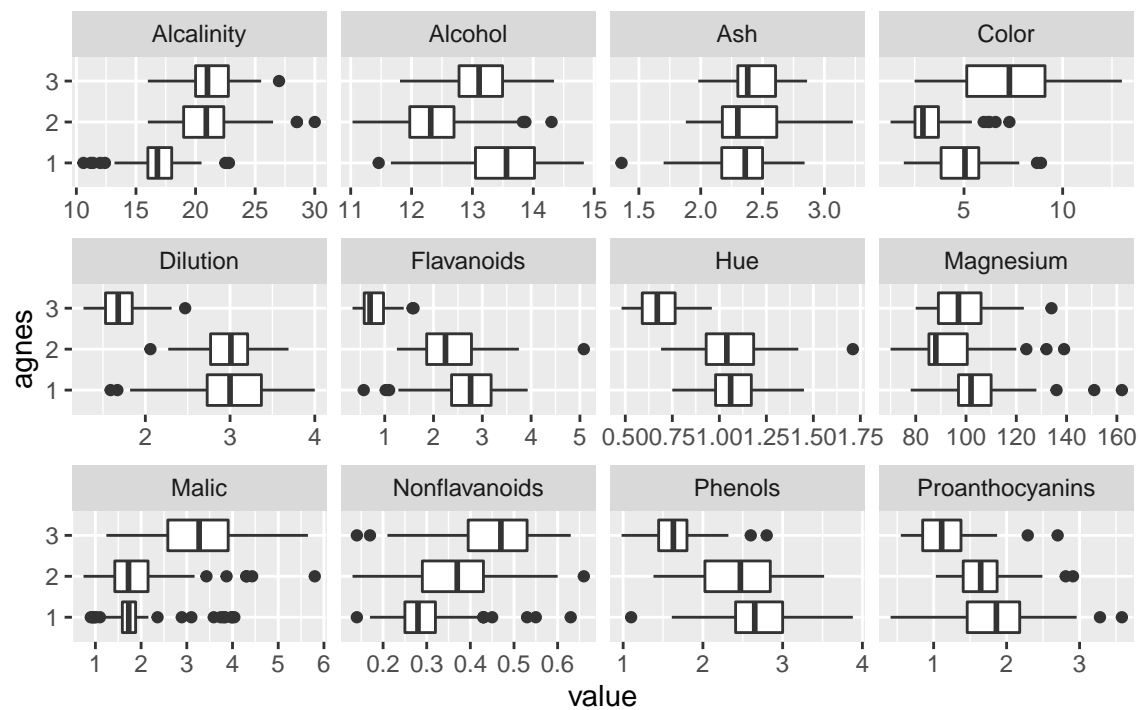
Tabela 4: Srednice skupisk

	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8	K = 9	K = 10
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	19.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	19.00	20.00
3	0.00	0.00	0.00	0.00	0.00	0.00	22.00	20.00	23.00
4	0.00	0.00	0.00	0.00	0.00	32.00	20.00	24.00	12.00
5	0.00	0.00	0.00	0.00	32.00	33.00	24.00	24.00	16.00
6	0.00	0.00	0.00	32.00	29.00	20.00	24.00	18.00	19.00
7	0.00	0.00	32.00	31.00	41.00	24.00	19.00	16.00	23.00
8	0.00	75.00	40.00	41.00	22.00	16.00	16.00	22.00	12.00
9	111.00	54.00	58.00	25.00	23.00	22.00	22.00	4.00	3.00
10	67.00	49.00	48.00	49.00	31.00	31.00	31.00	31.00	31.00

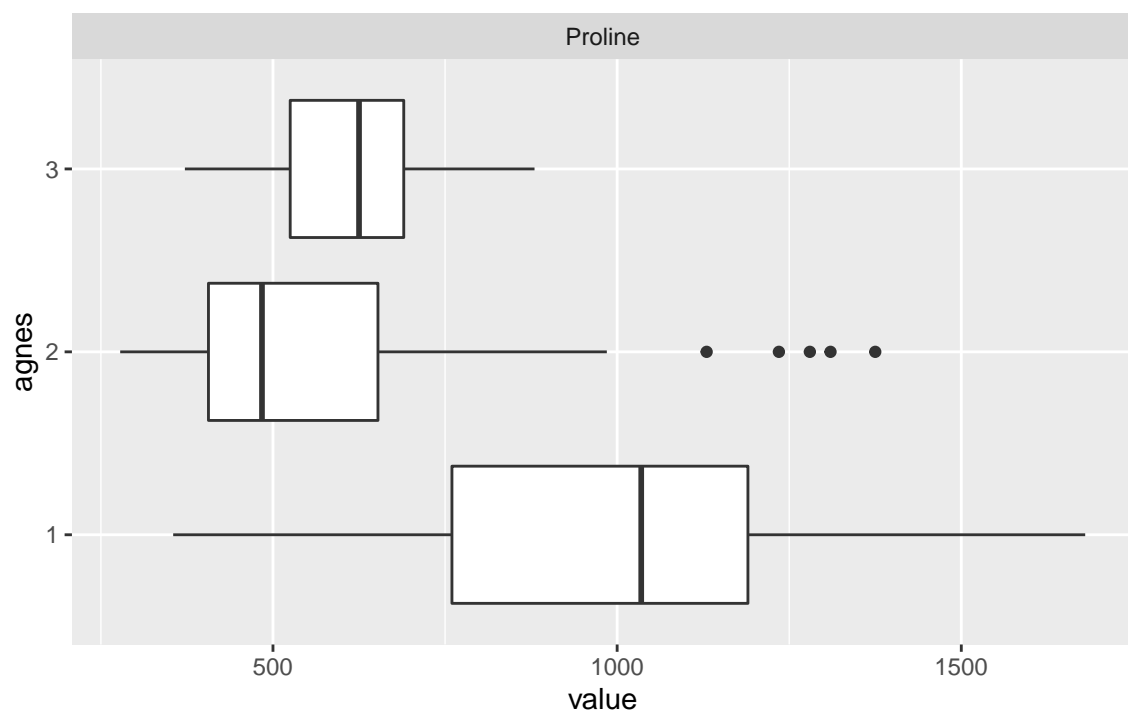
Tabela 5: Rozmiary skupisk

### 2.2.3 Ocena otrzymanych rezultatów





Page 1



Page 2

	1	2	3
Alcohol	0.59	-0.92	0.39
Malic	-0.47	-0.54	0.81
Ash	0.16	-0.90	0.05
Alcalinity	0.30	-0.15	0.60
Magnesium	0.02	-1.38	-0.54
Phenols	0.65	-1.03	-0.58
Flavanoids	0.95	0.00	-1.27
Nonflavanoids	-0.82	0.07	0.71
Proanthocyanins	0.47	0.07	-0.60
Color	0.02	-0.72	1.45
Hue	0.36	0.19	-1.78
Dilution	1.21	0.79	-1.40
Proline	0.55	-0.75	-0.31

Tabela 6: Medoidy dla metody PAM przy K=3