Klasteryzacja

Przemysław Adam Chojecki

Michał Wdowski

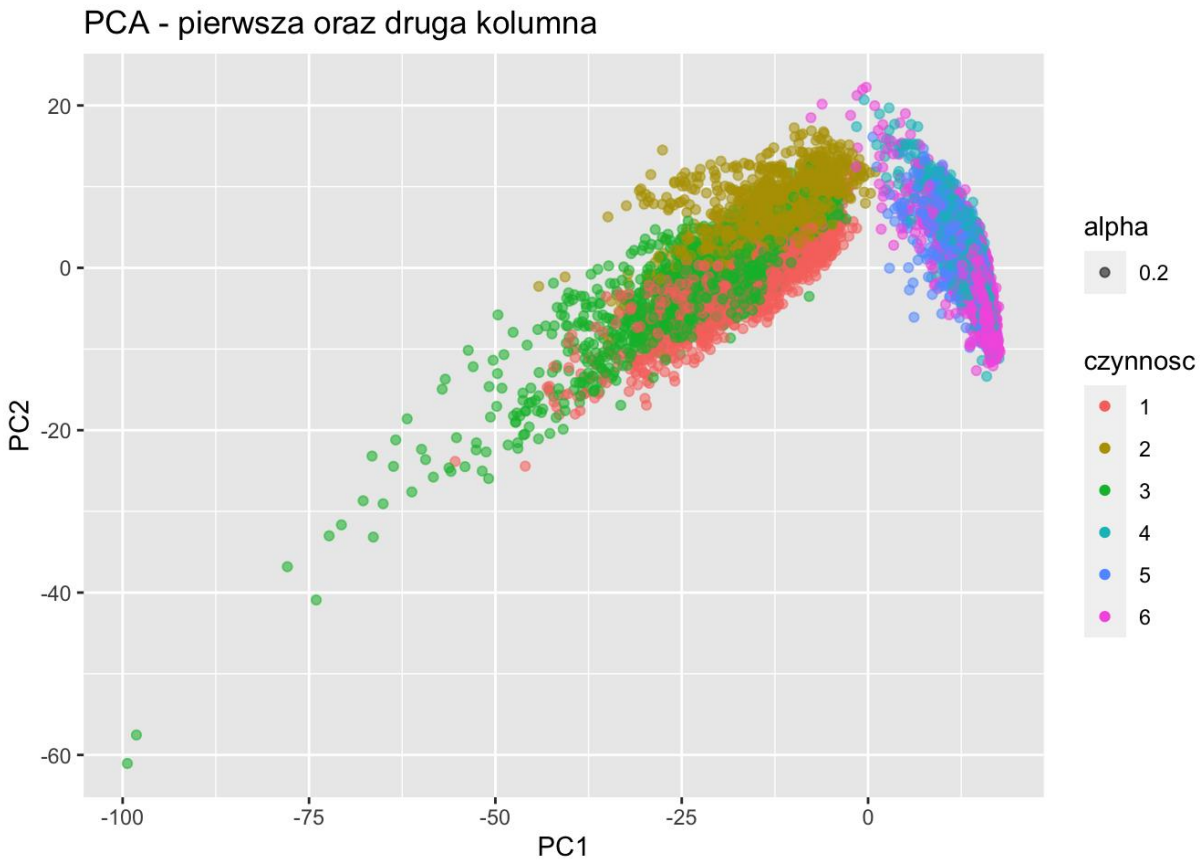
Dane

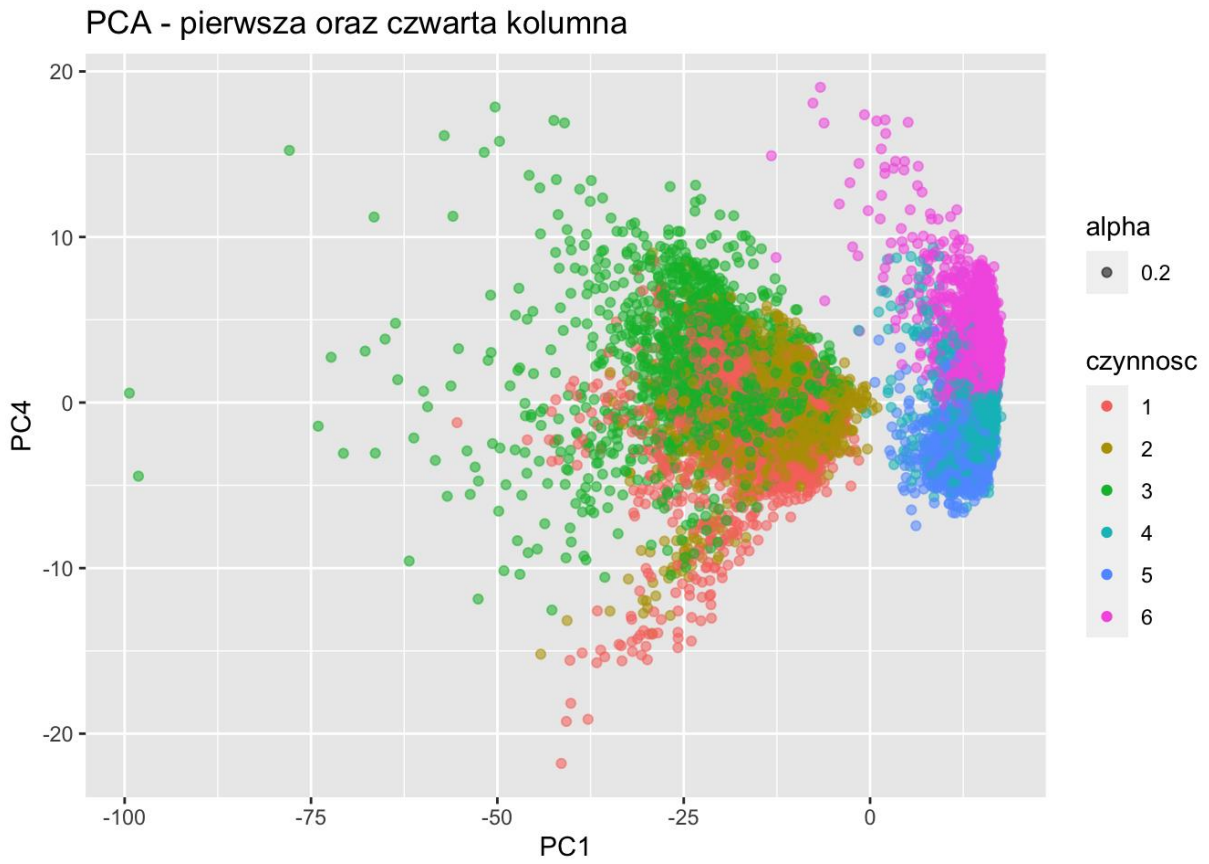
Surowe dane ➔

561 kolumn – 10299 obserwacji:

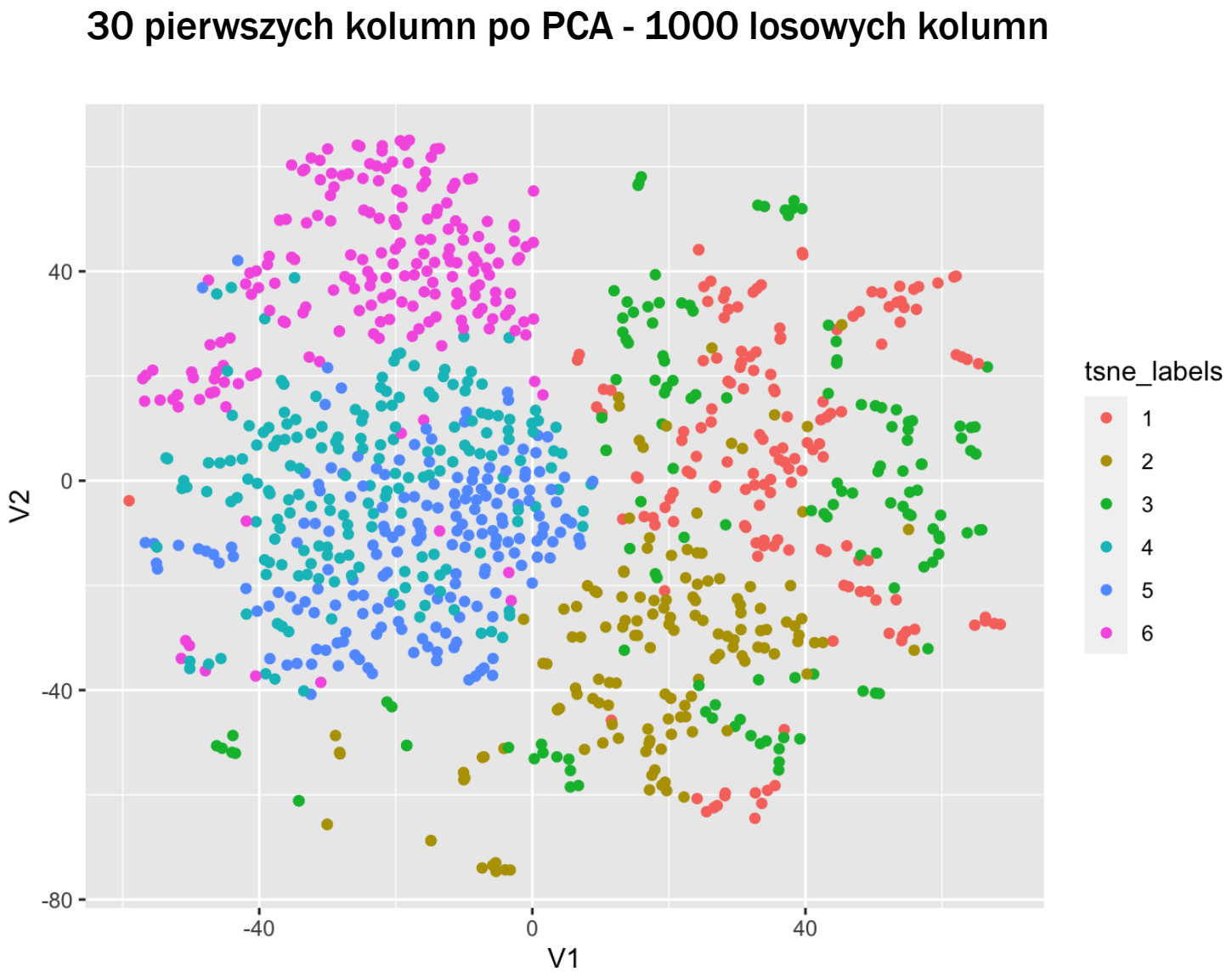
1. Chodzenie po płaskiej powierzchni
2. Wchodzenie po schodach
3. Schodzenie po schodach
4. Siadanie
5. Wstawanie
6. Kładzenie się

PCA

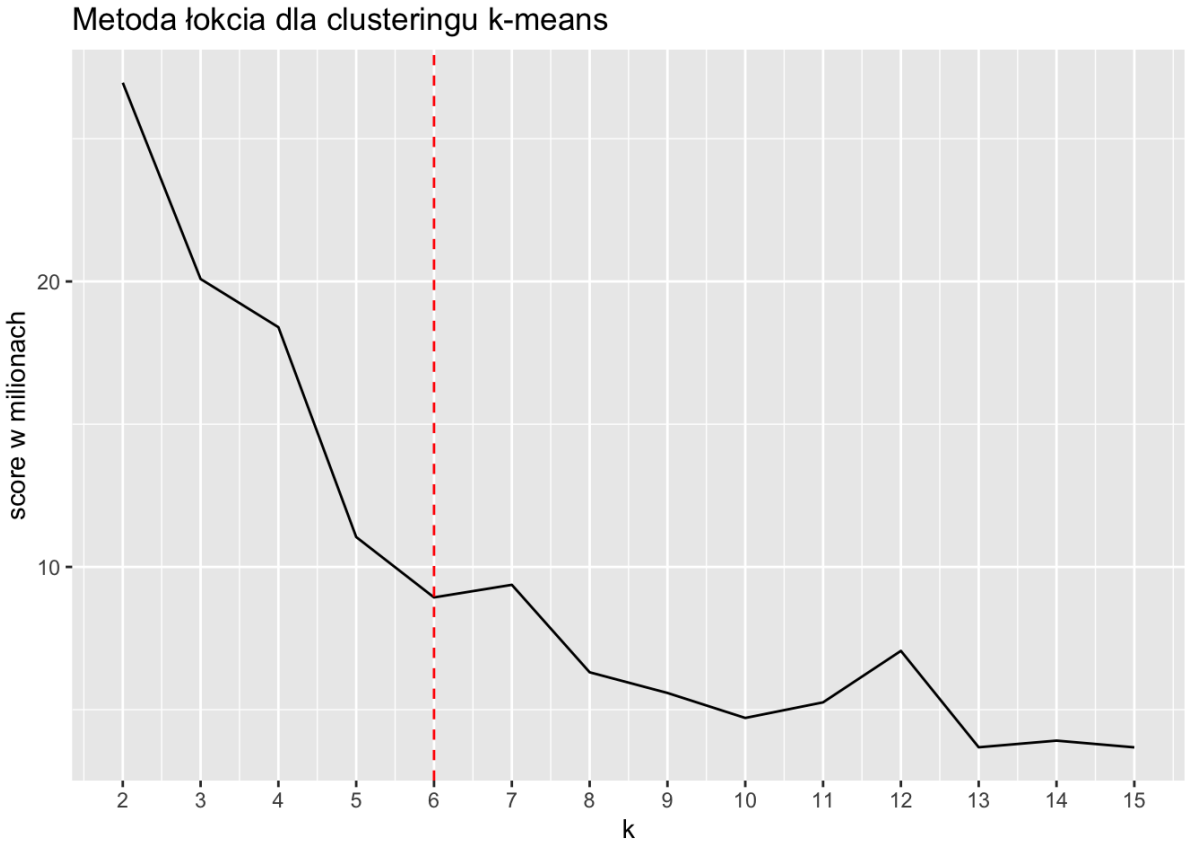

PCA - pierwsza oraz druga kolumna
20
0
-20
-40
-60
PC2
-100
-75
-50
-25
0
PC1
alpha
0.2
czynnosc
1
2
3
4
5
6

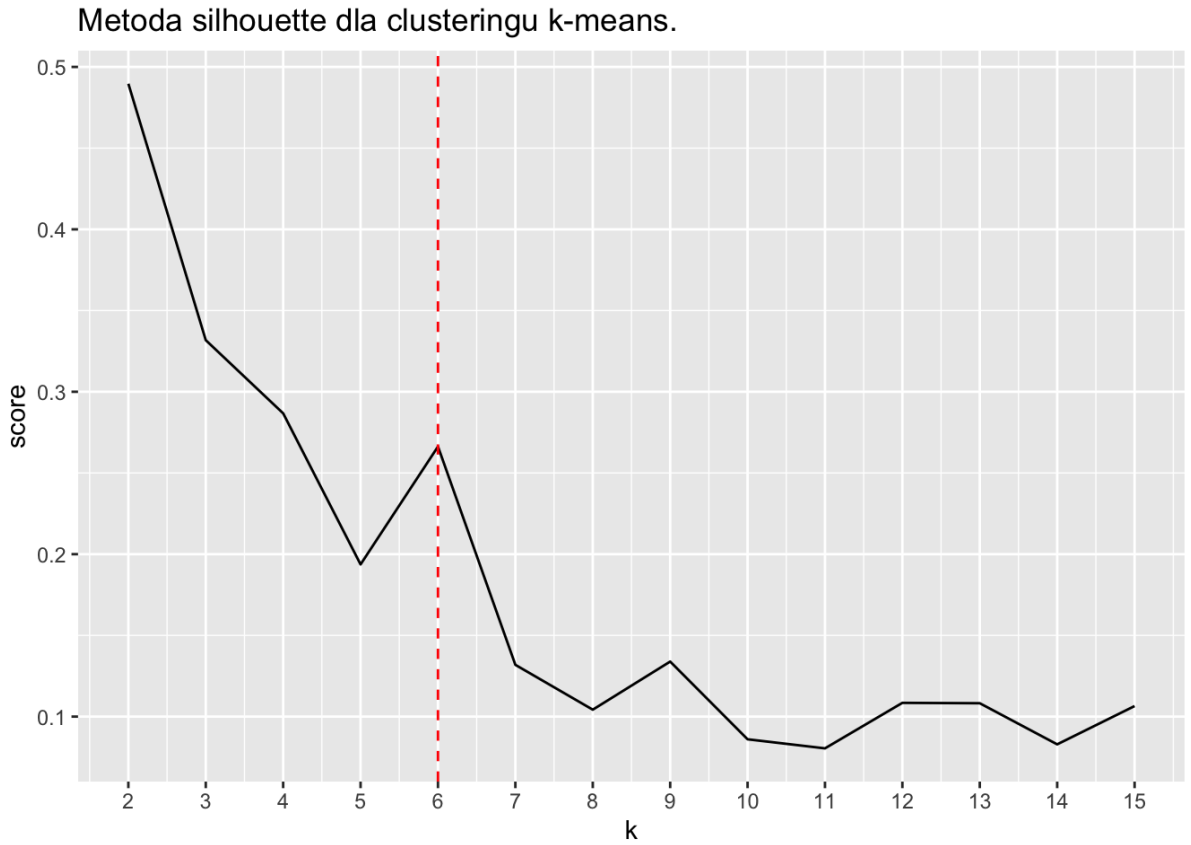

PCA - pierwsza oraz czwarta kolumna
20
10
0
-10
-20
PC4
-100
-75
-50
-25
0
PC1
alpha
0.2
czynnosc
1
2
3
4
5
6

t-SNE

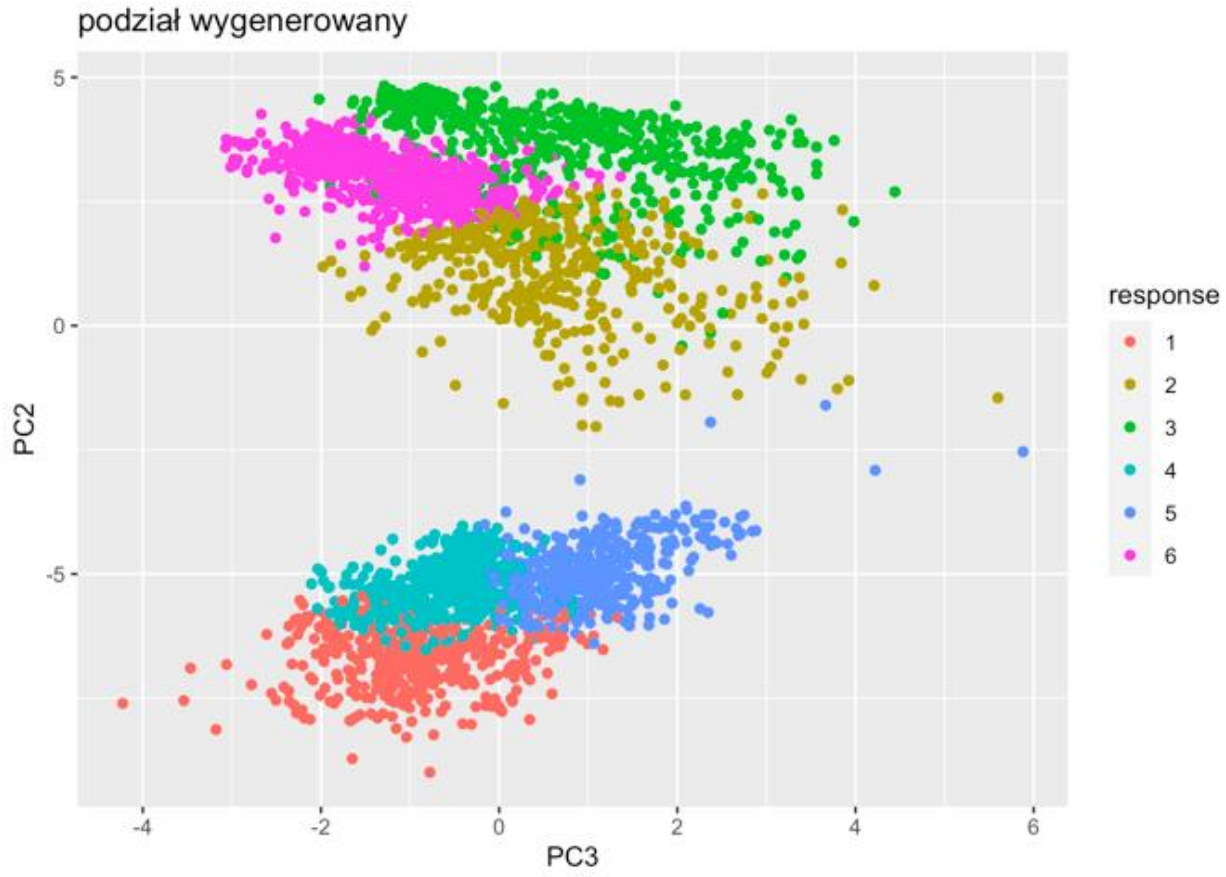

30 pierwszych kolumn po PCA - 1000 losowych kolumn
40
0
-40
-80
V2
-40
0
40
V1
tsne_labels
1
2
3
4
5
6

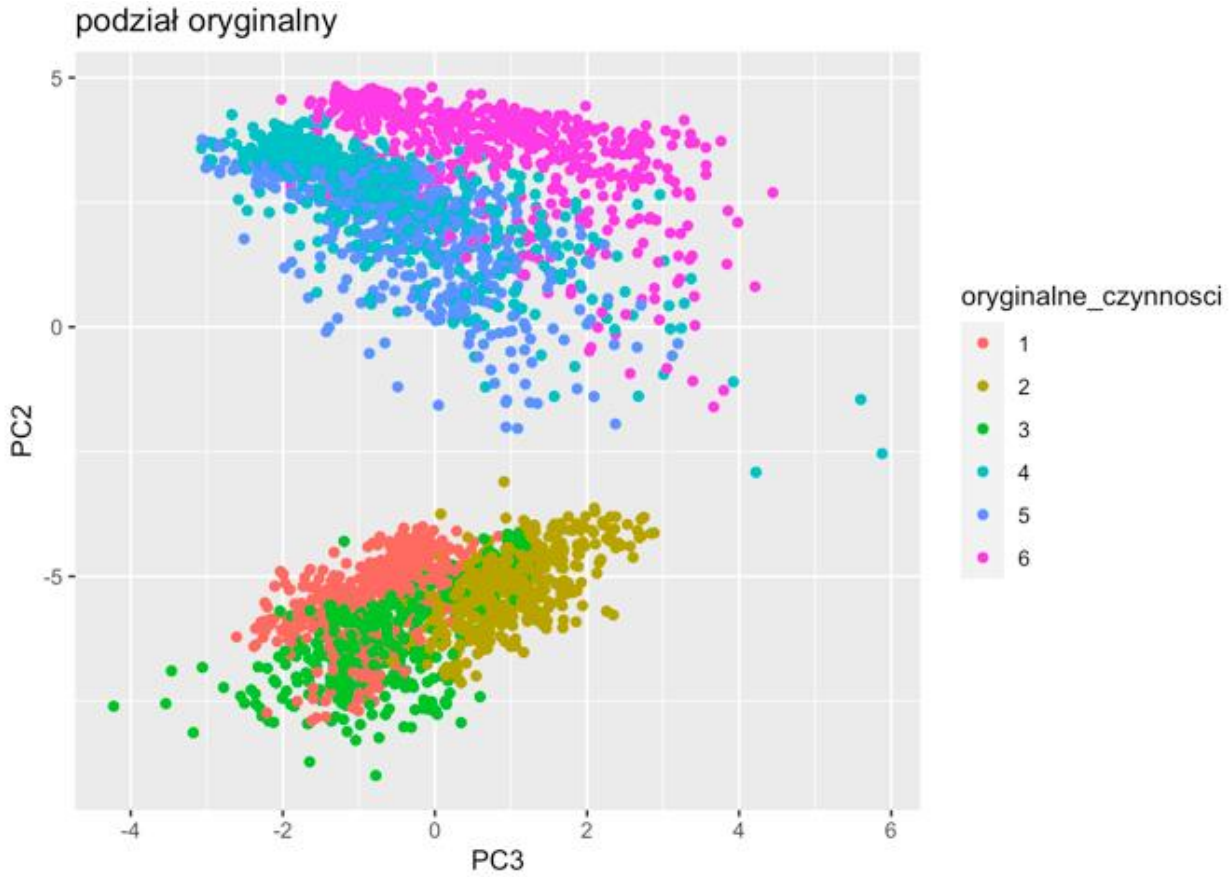
K-means – dobór liczby klastrów





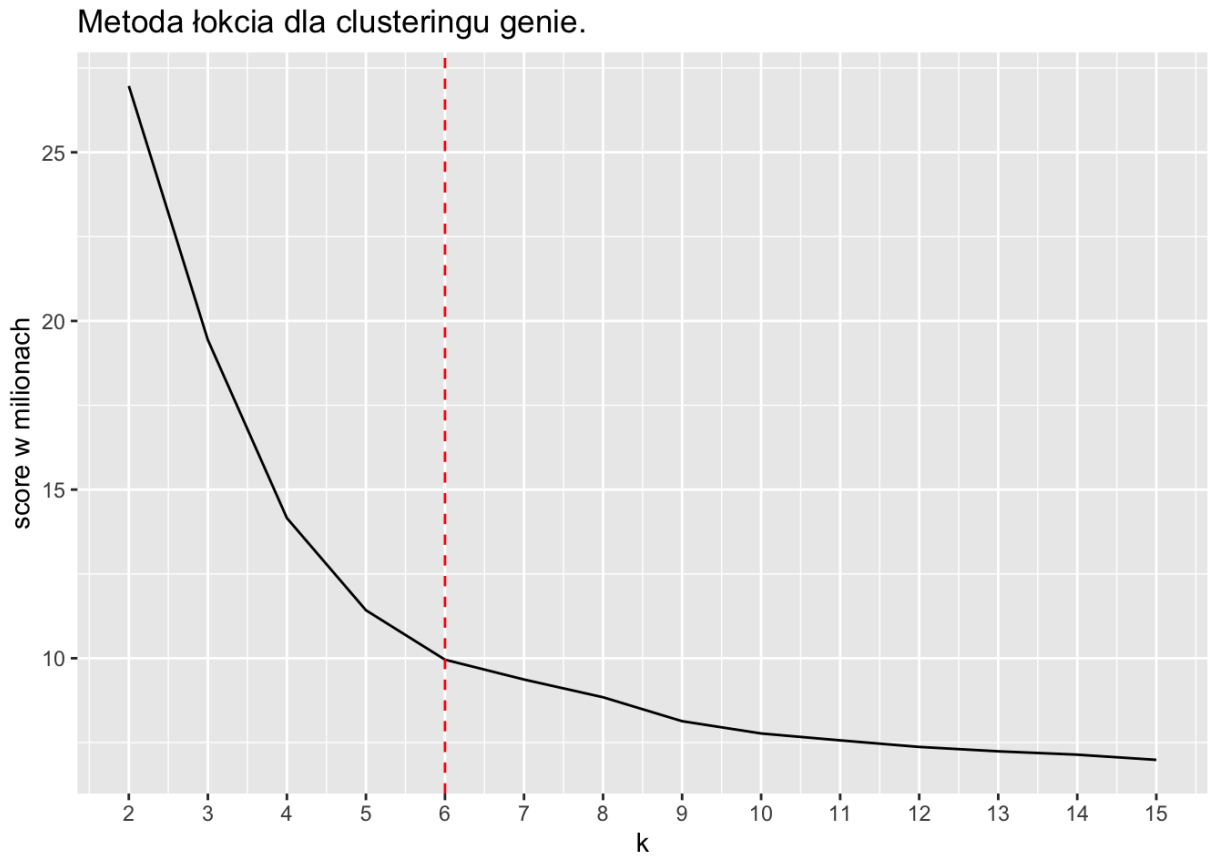
K-means – podział na 6 klastrów

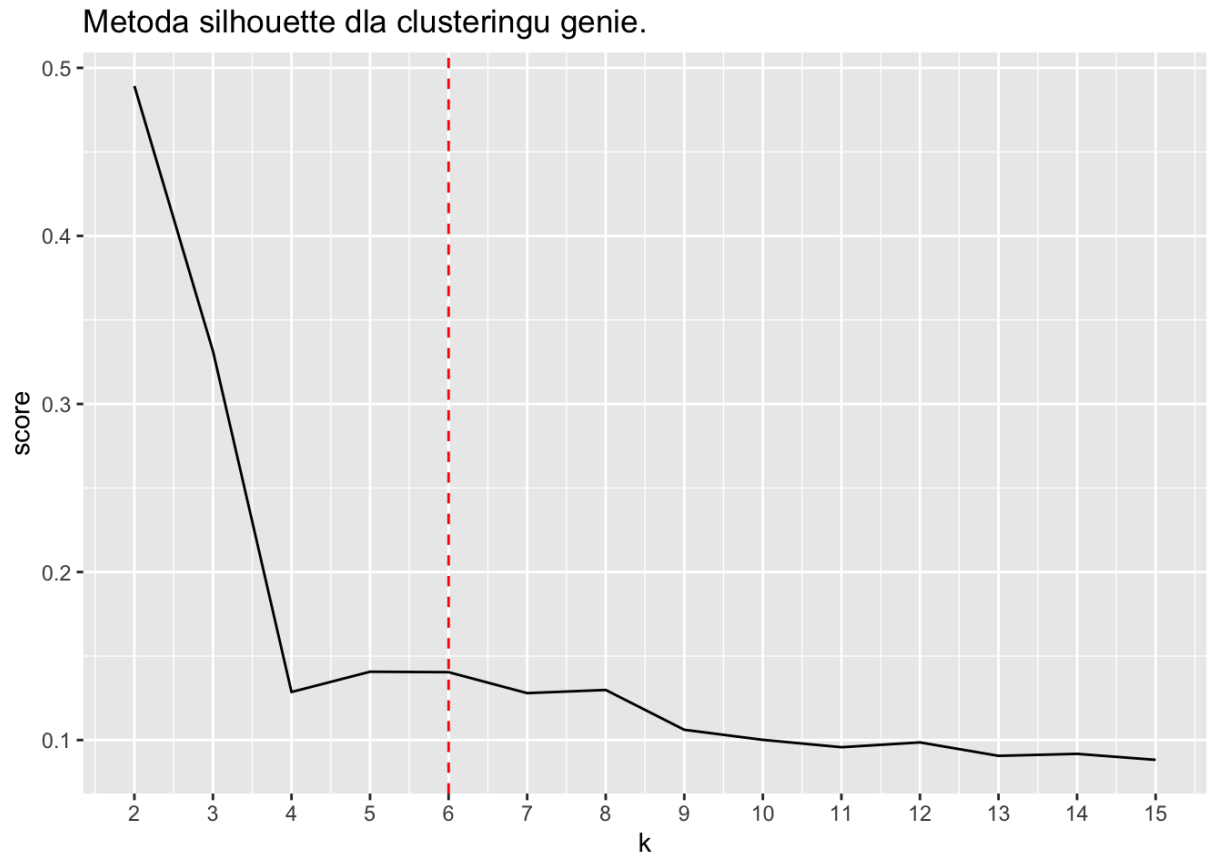




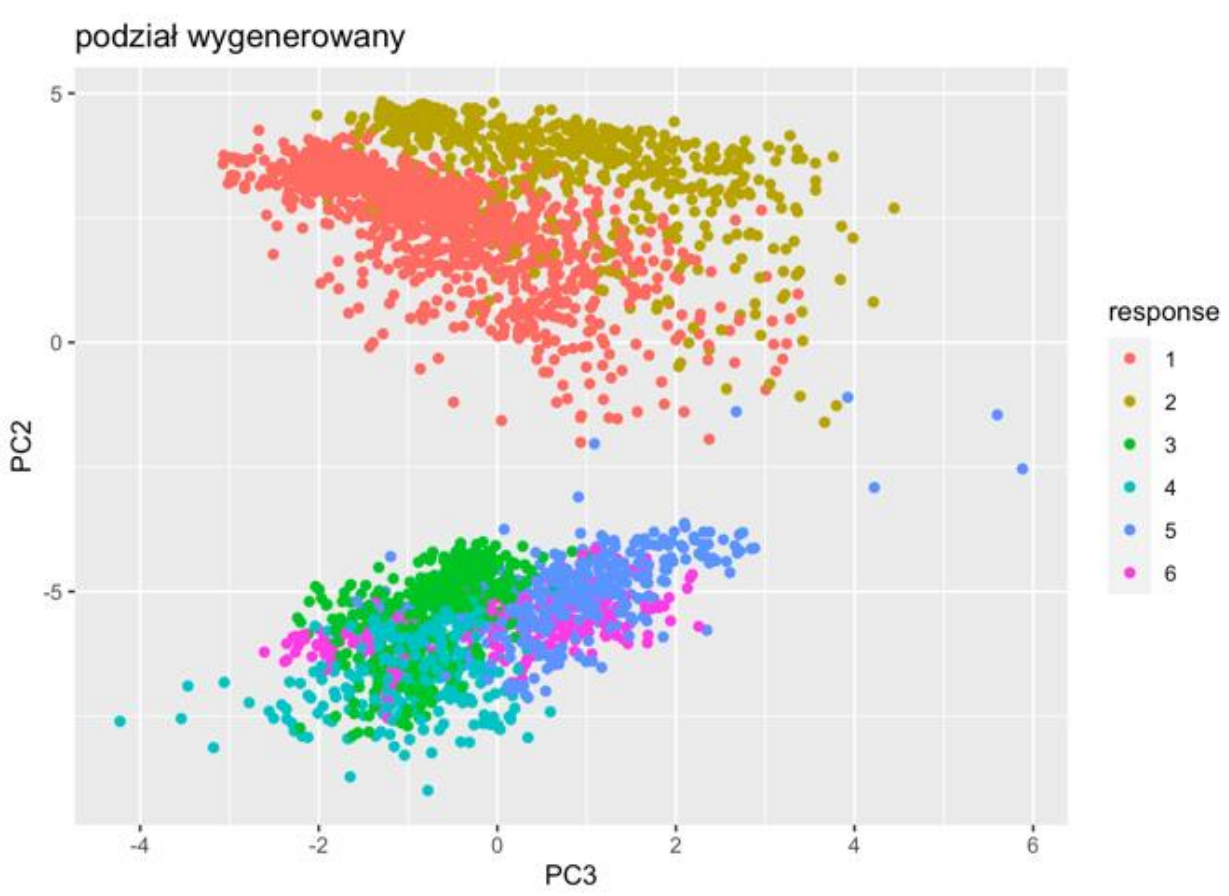
silhouette	DB	Dunn	FM	AR
0.121	2.387	0.135	0.619	0.541

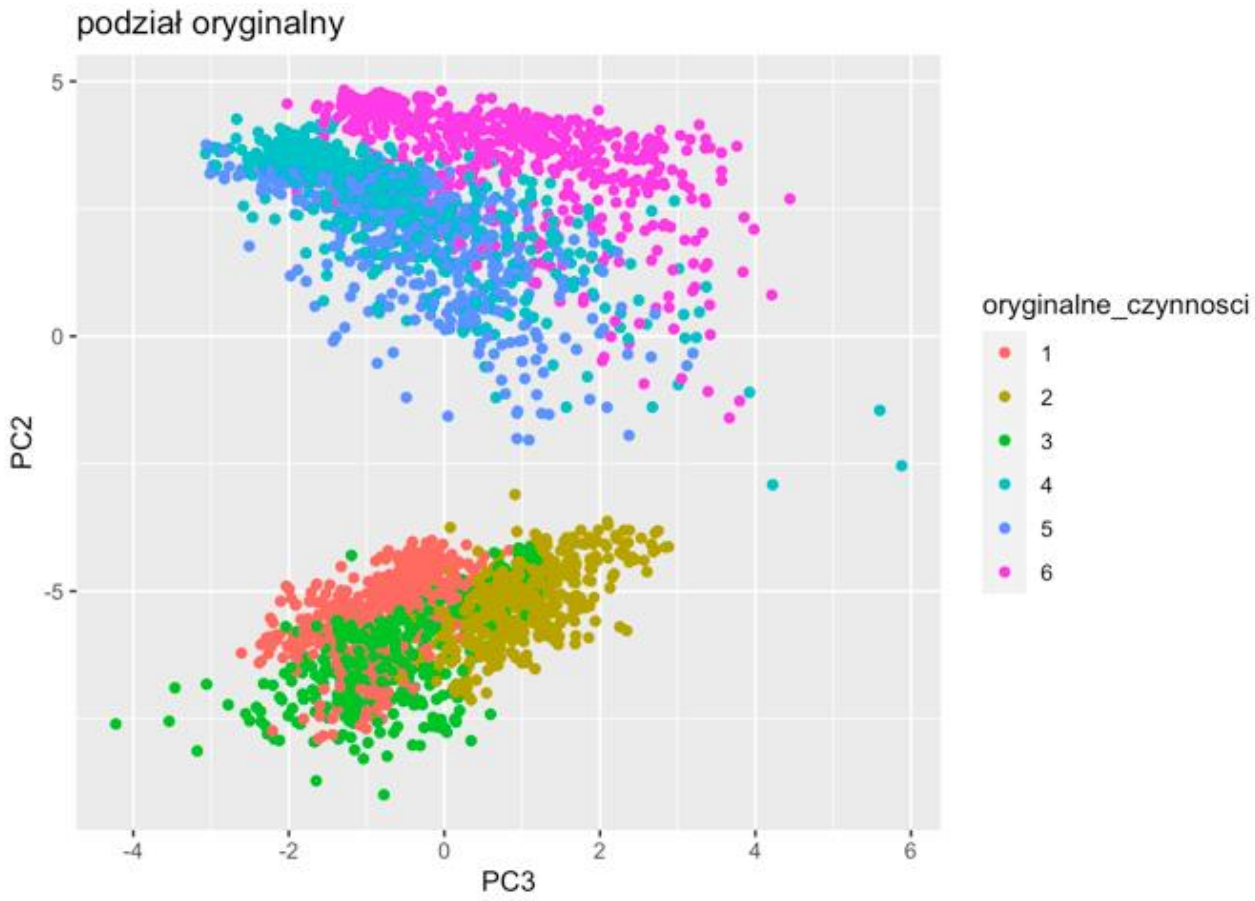
genie – dobór liczby klastrów





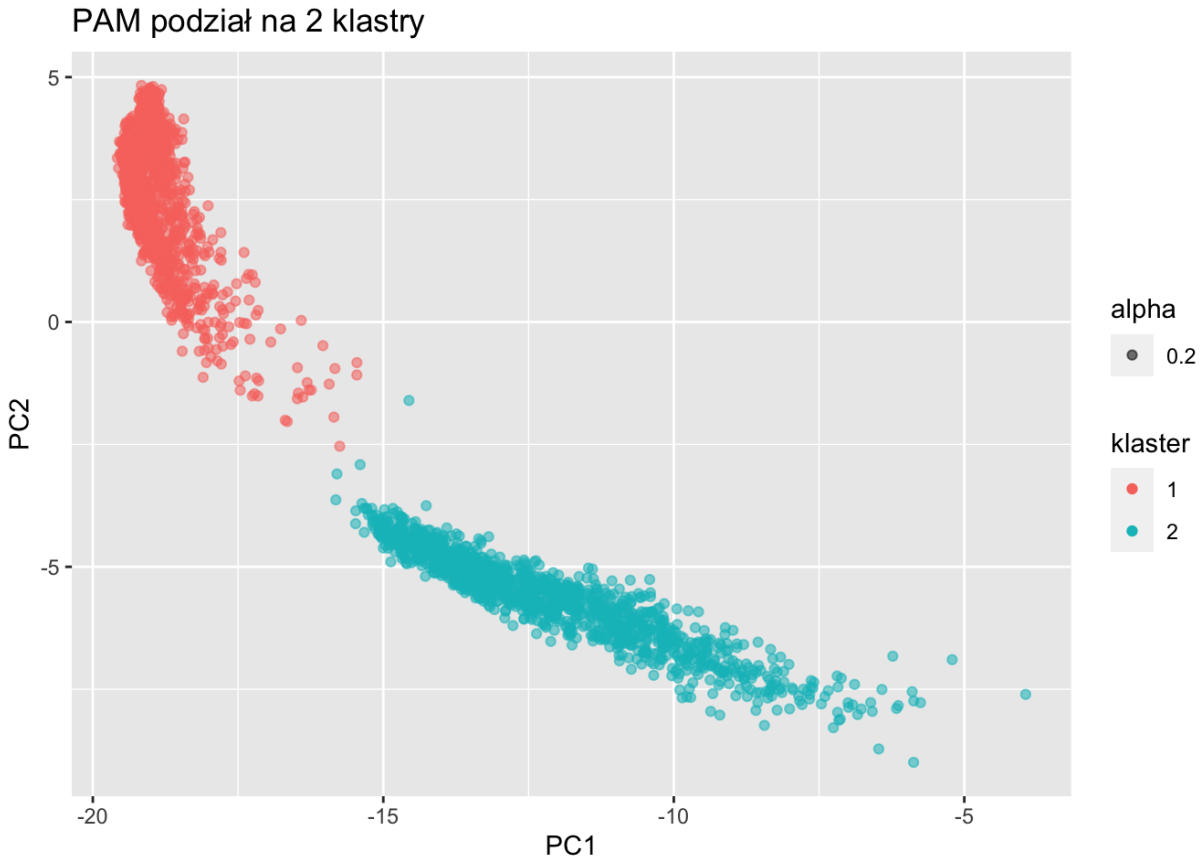
genie – podział na 6 klastrów

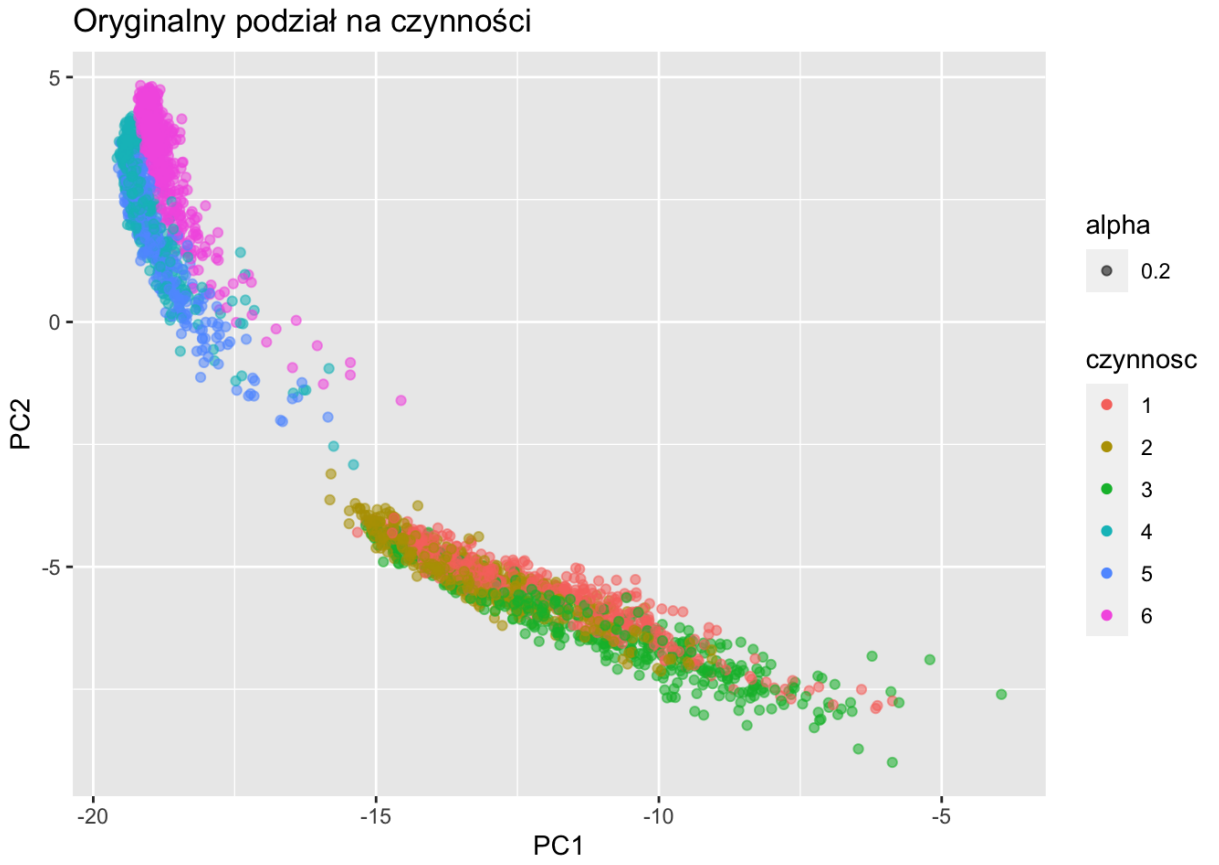




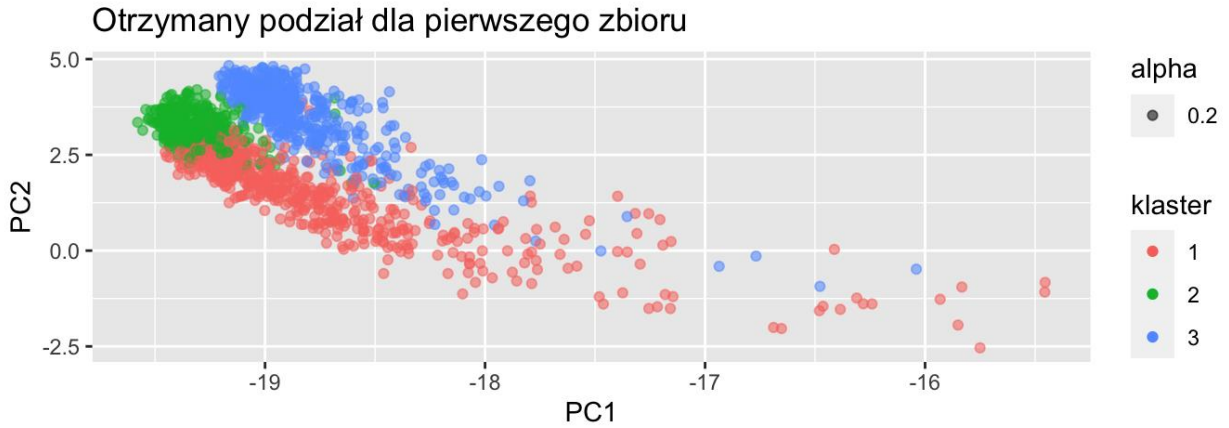
silhouette	DB	Dunn	FM	AR
0.140	2.854	0.220	0.715	0.644

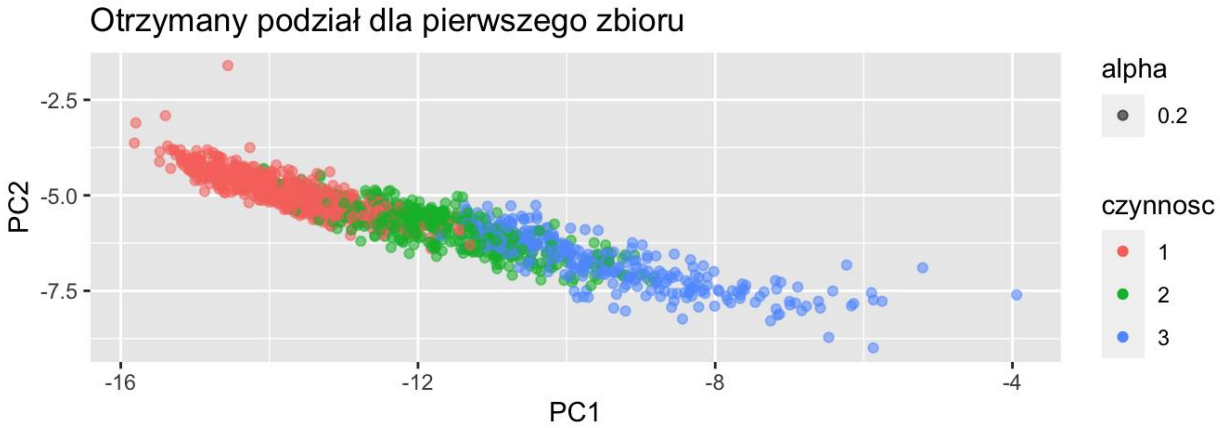
Podział na 2 klastry

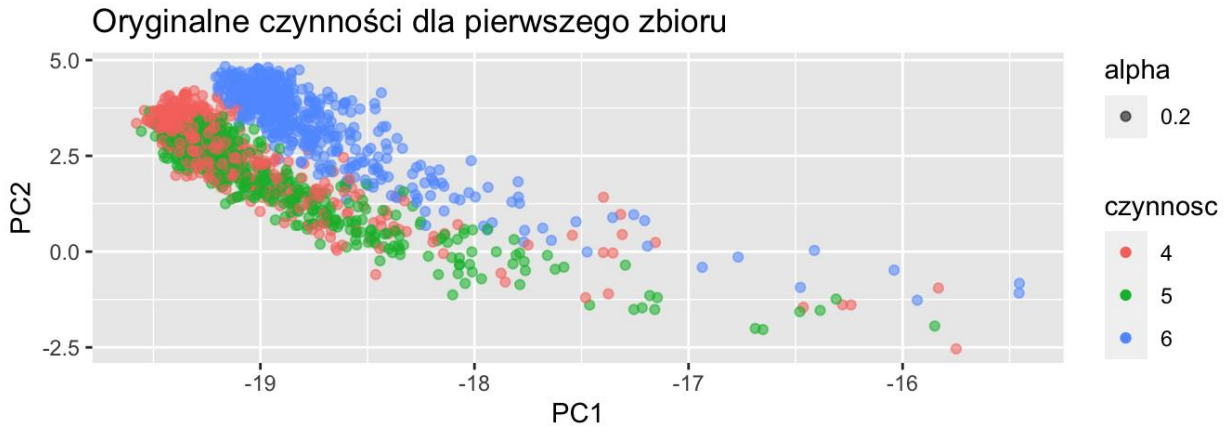


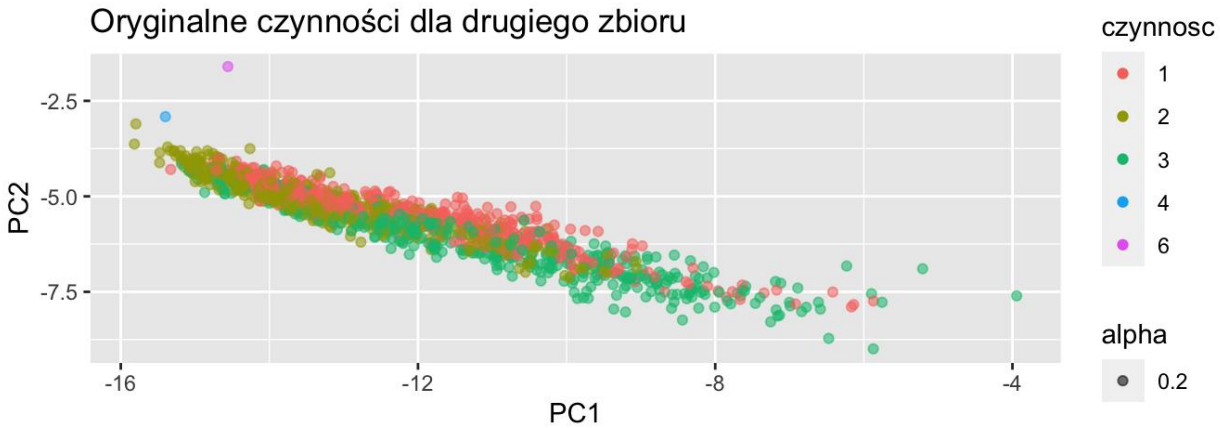


Podział na 2 klastry

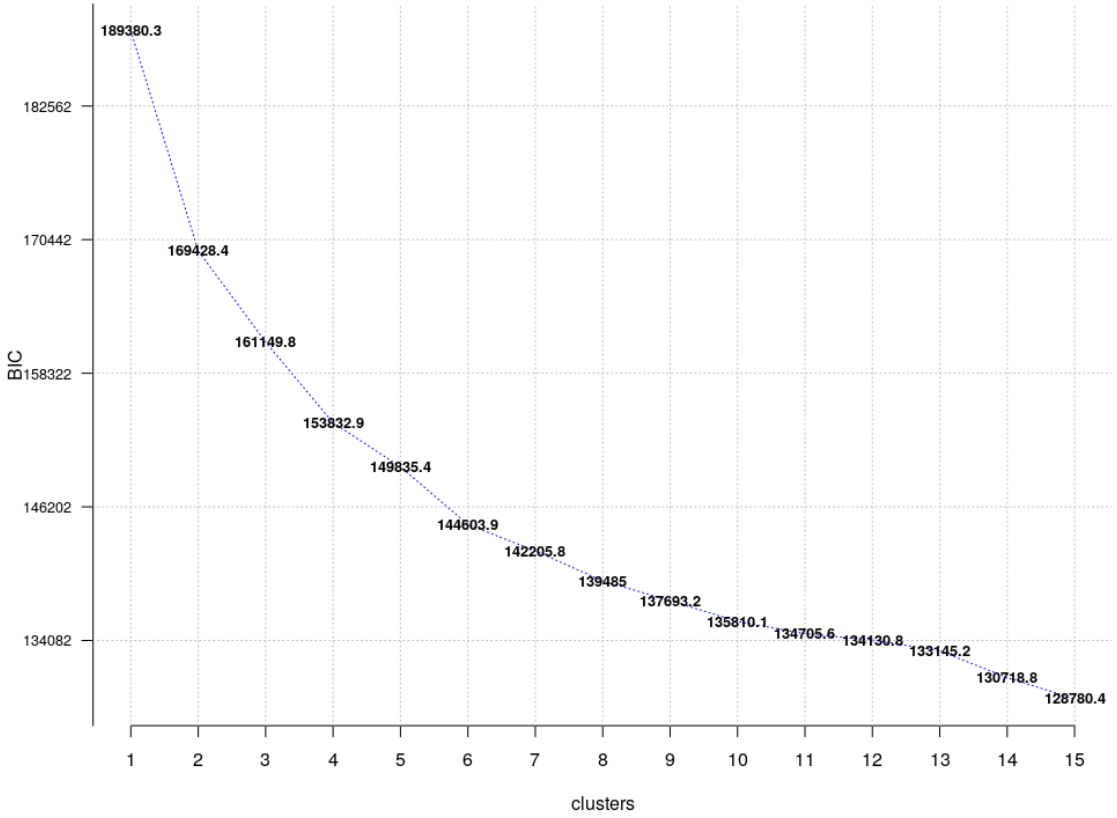


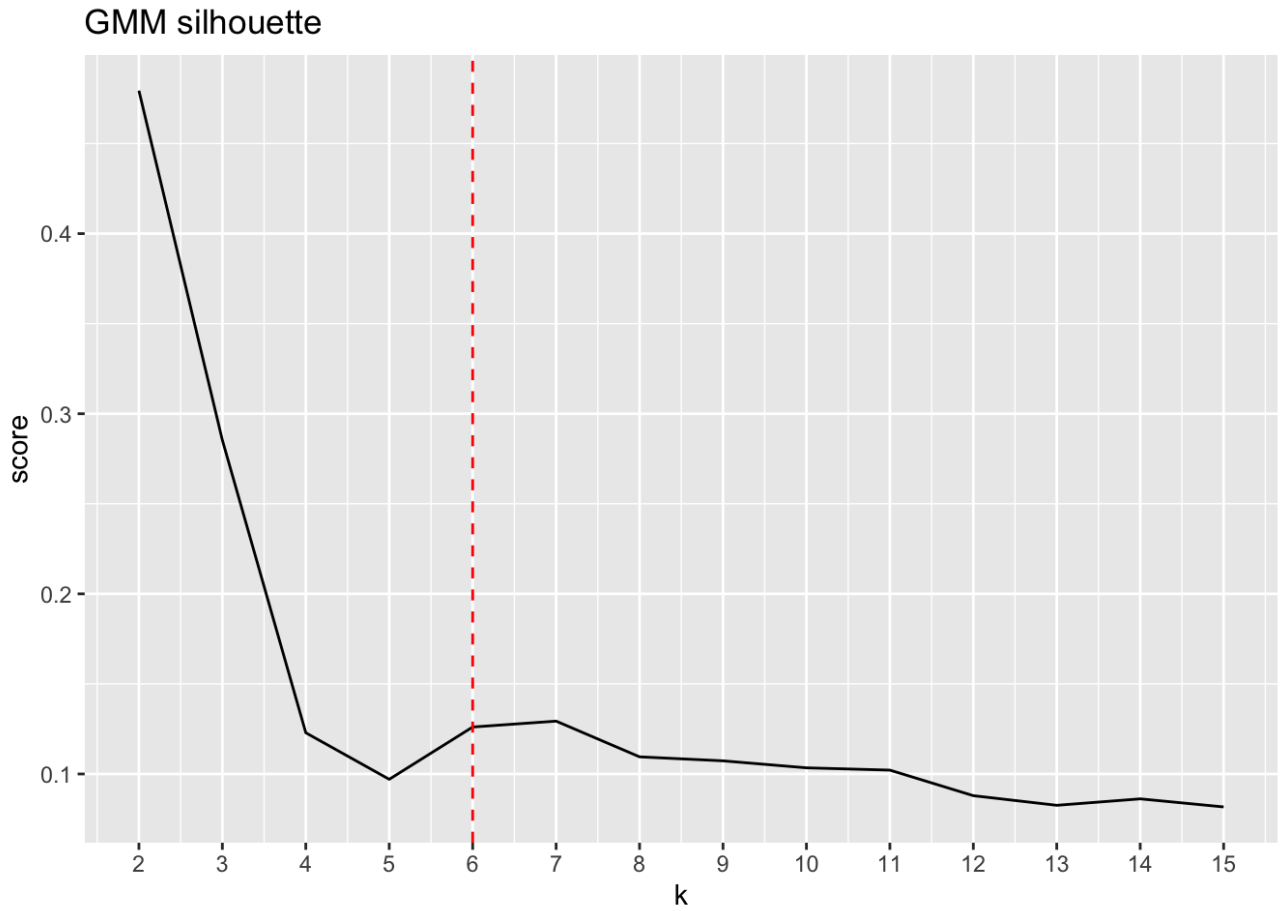






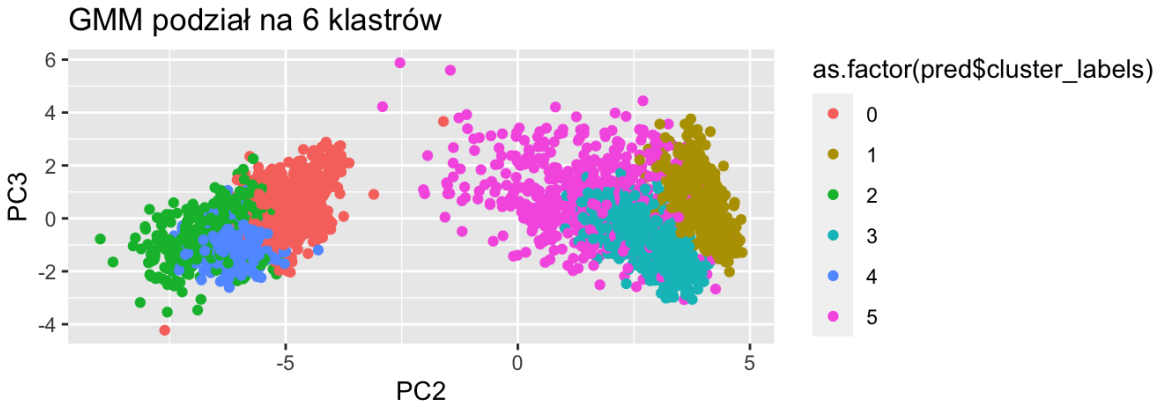
gmm – dobór liczby klastrów

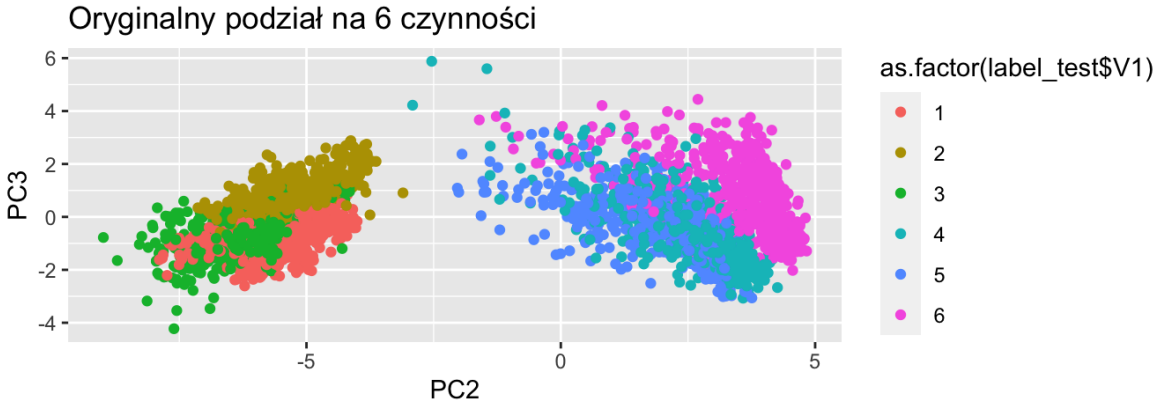




gmm – najlepsze podziały

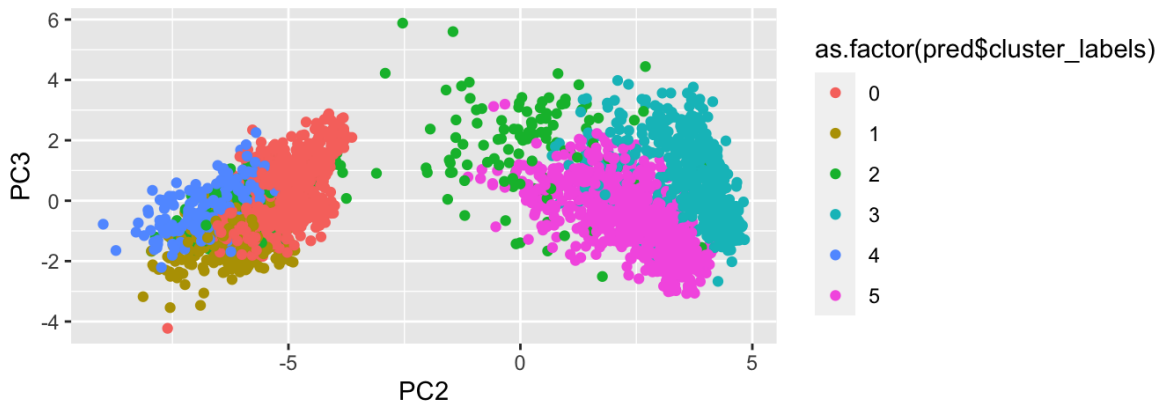
Random spread + metryka euklidesowa

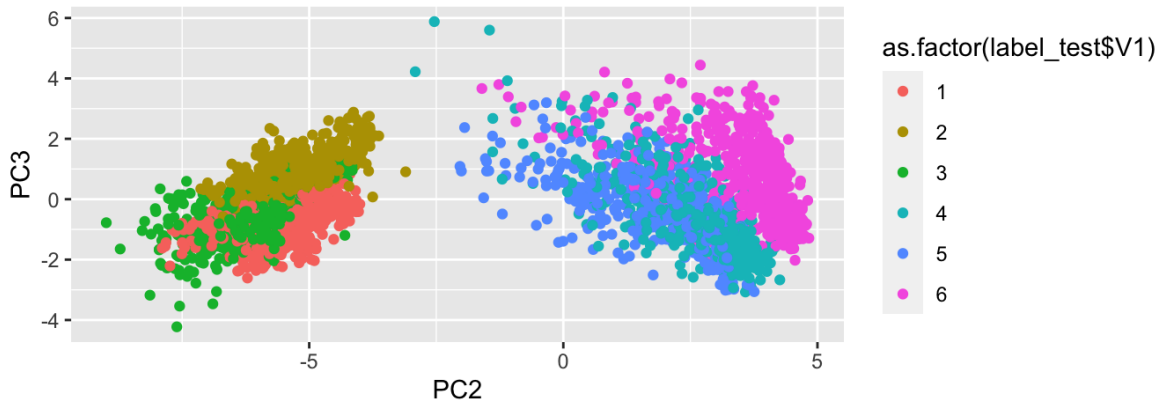




silhouette	DB	Dunn	FM	AR
0.121	2.531	0.128	0.491	0.382

Static subset + metryka Manhattan





silhouette	DB	Dunn	FM	AR
0.148	2.179	0.117	0.573	0.469

Podsumowanie