

# Klasyfikacja klientów ze względu na roczny przychód oraz punkty oceny klienta

Tomasz Śmiech ;Alan Kubit;Kaja Miśkowiec

05.03.2019

## 1 Opis danych

Dane zostały pobrane z serwisu kaggle.com. Zawierają dane o klientach zebrane w tabelę o kolumnach :

ID klienta	płeć	wiek	średni roczny przychód	punkty klienta
------------	------	------	------------------------	----------------

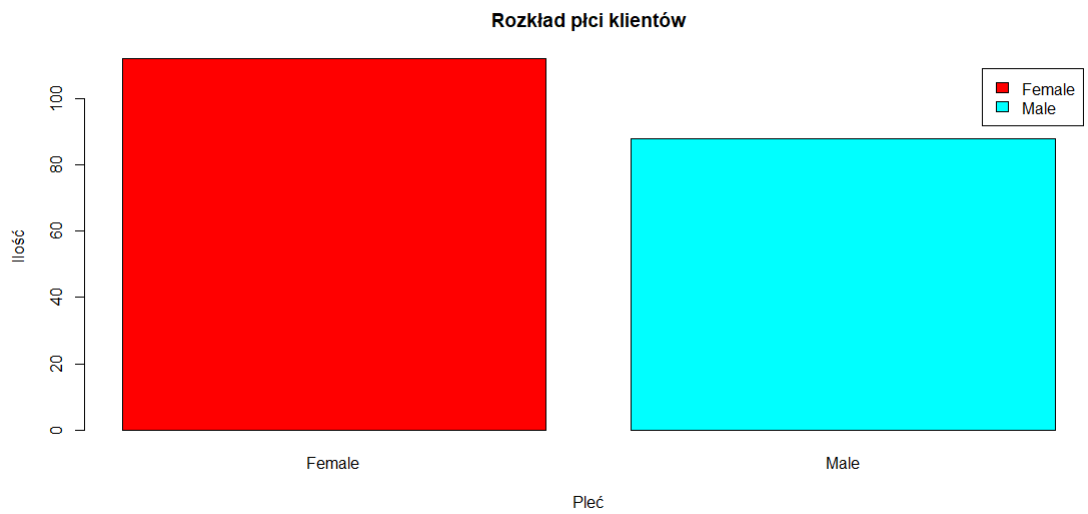
gdzie: punkty klienta - zostały nadawane ze względu na ilość wydanych pieniędzy oraz zachowanie klienta;  
średni dochód klienta - dochód w tysiącach dolarów amerykańskich;

```
head(customer_data)
```

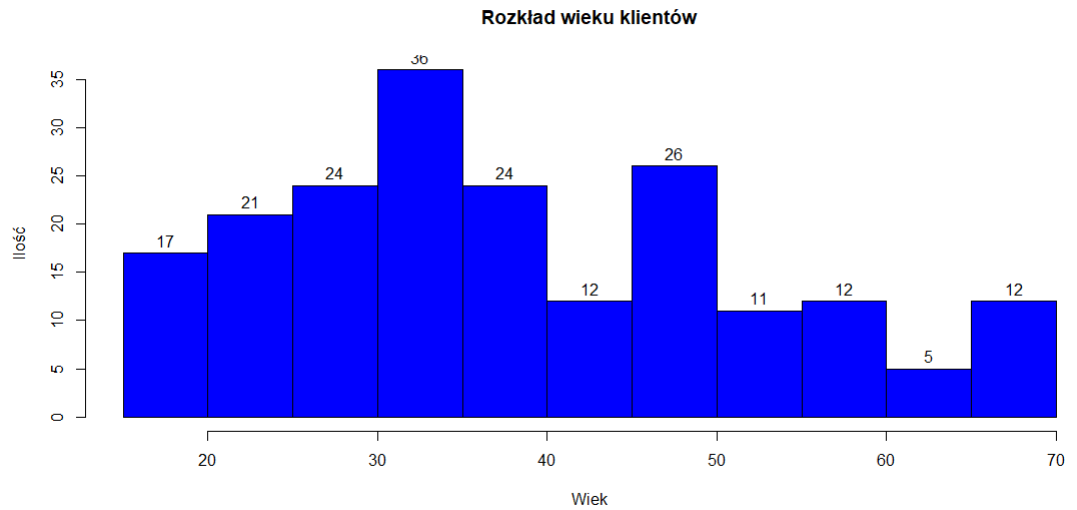
```
CustomerID Gender Age Annual.Income..k.. Spending.Score..1.100.  
1          1  Male  19          15          39  
2          2  Male  21          15          81  
3          3 Female  20          16           6  
4          4 Female  23          16          77  
5          5 Female  31          17          40  
6          6 Female  22          17          76
```

## 2 Wstępna analiza danych

```
a=table(customer_data$Gender)
barplot(a,main="Rozkład płci klientów",
        ylab="Ilość",
        xlab="Płeć",
        col=rainbow(2),
        legend=rownames(a))
```

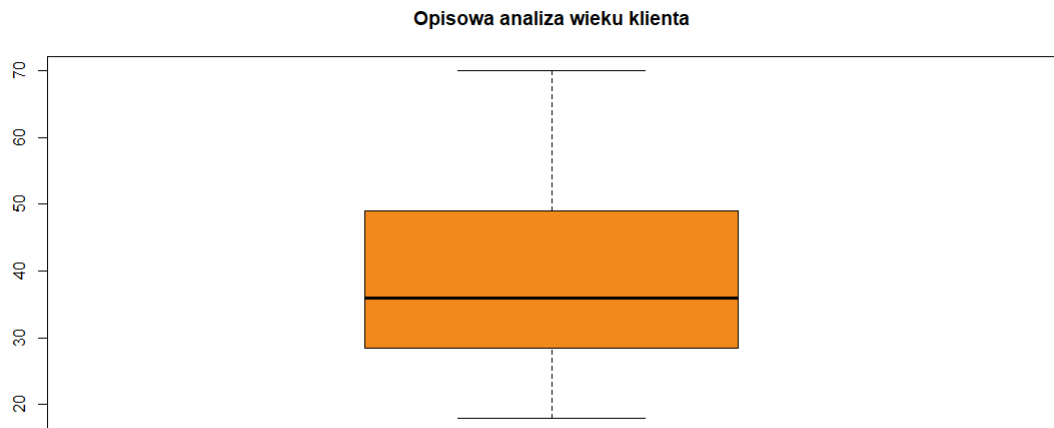


```
hist(customer_data$Age,
      col="blue",
      main="Rozkład wieku klientów",
      xlab="Wiek",
      ylab="Ilość",
      labels=TRUE)
```



```
summary(customer_data$Age)
```

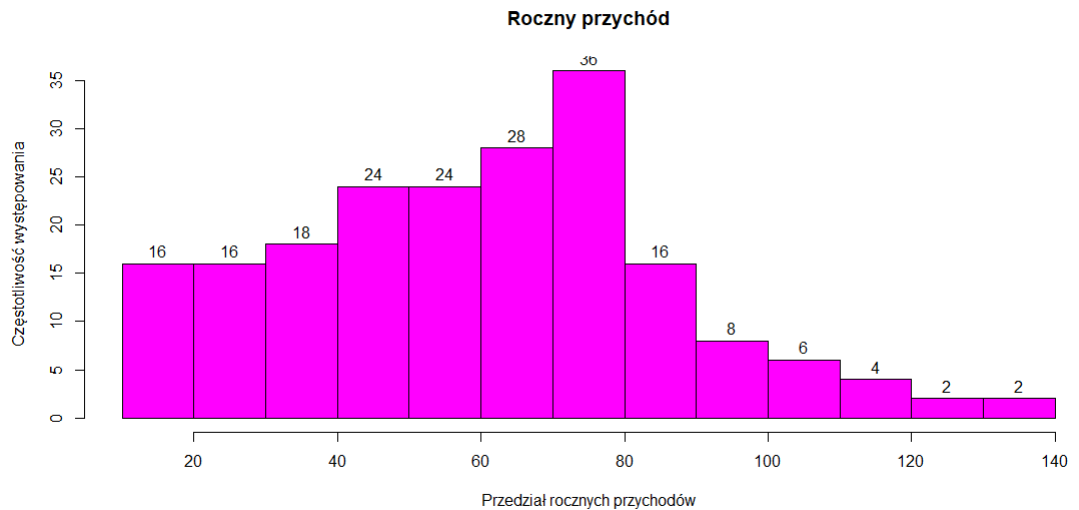
```
Min. 1st Qu. Median Mean 3rd Qu. Max.
 18.00  28.75  36.00  38.85  49.00  70.00
```



```
summary(customer_data$Annual.Income..k..)
```

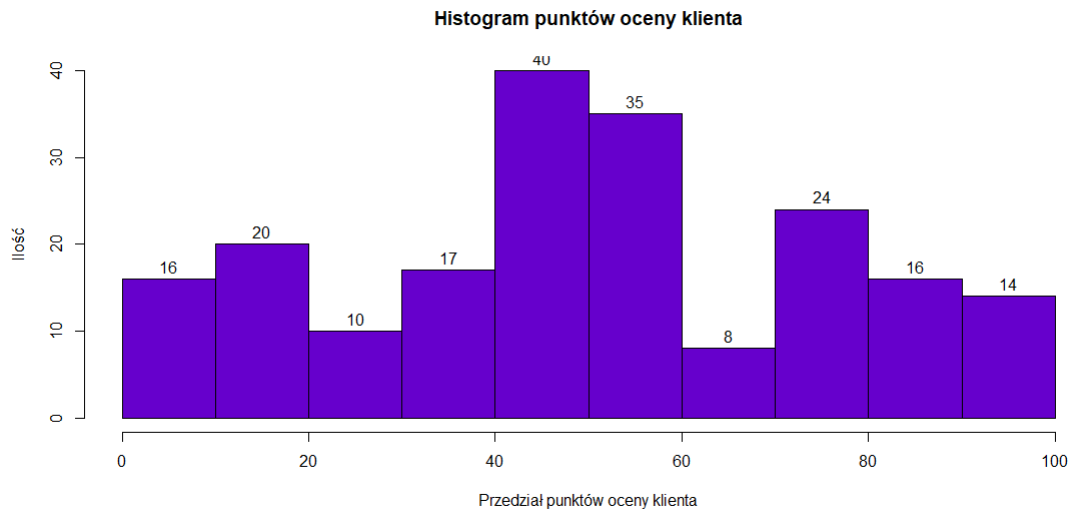
```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 15.00   41.50   61.50   60.56   78.00  137.00
```

```
hist(customer_data$Annual.Income..k..,
      col="#660033",
      main="Roczny przychod",
      xlab="Przedzial rocznych przychodow",
      ylab="Czestotliwosc wystepowania",
      labels=TRUE)
```



```
summary(customer_data$Spending.Score..1.100.)
```

```
hist(customer_data$Spending.Score..1.100.,
      main="Histogram punktow oceny klienta",
      xlab="Przedzial punktow klienta",
      ylab="Ilosc",
      col="#6600cc",
      labels=TRUE)
```



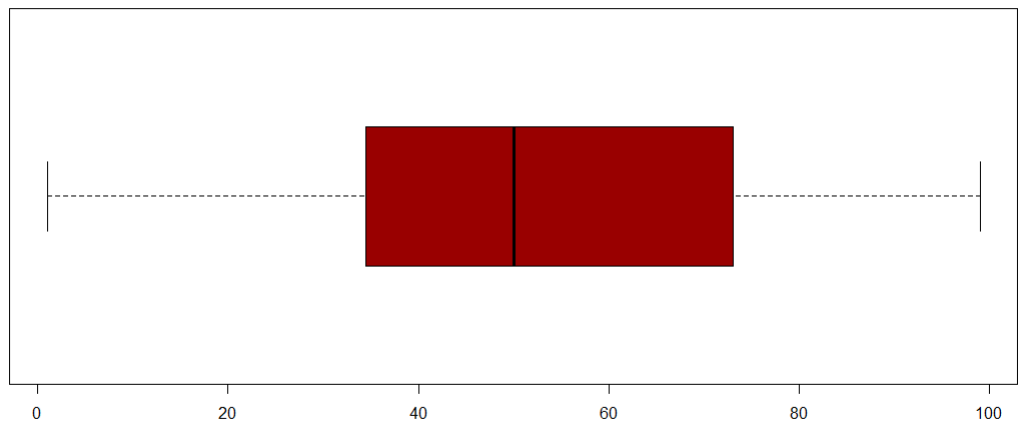
```
summary(customer_data$Spending.Score..1.100.)
```

```

Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.00  34.75   50.00   50.20  73.00   99.00

```

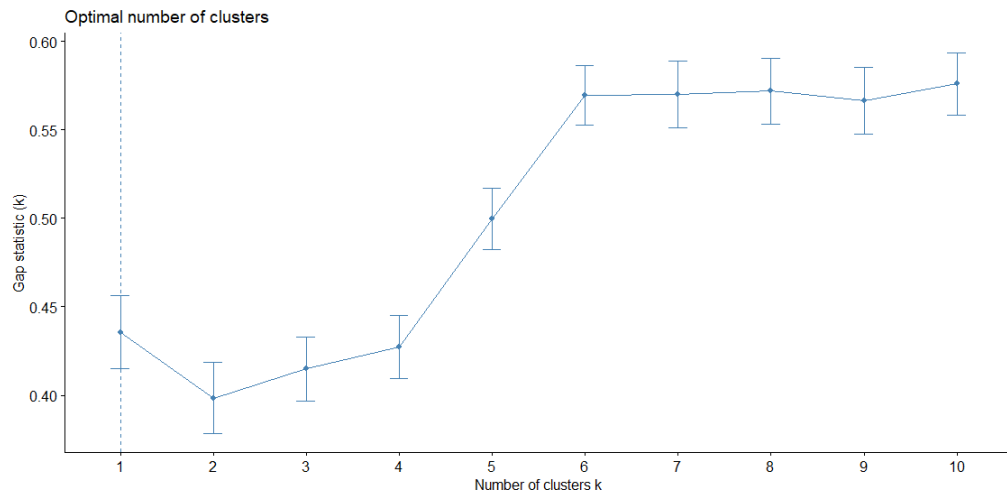
Opisowa analiza punktów oceny klienta



### 3 Metoda Centroidów

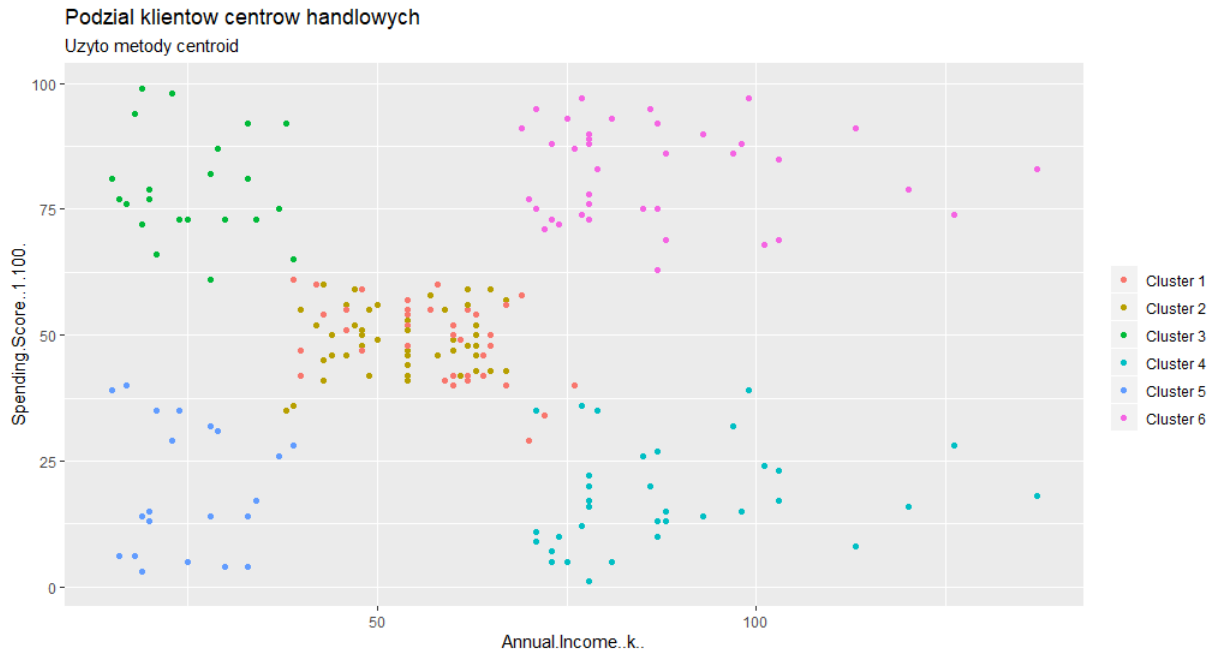
1. Ustalamy liczbę skupień(klastrów);
2. Ustalamy wstępne środki skupień tzw. centroidy;
3. Obliczamy odległości obiektów od środków skupień;
4. Przypisujemy obiekty do skupień;
5. Ustalamy nowe środki skupień(Punkt, którego współrzędne są średnią arytmetyczną współrzędnych punktów należących do danego skupienia);
6. Wykonujemy kroki 3, 4, 5 do czasu, aż warunek zatrzymania zostanie spełniony(np. ilość iteracji zadana na początku lub brak przesunięć obiektów)

```
stat_gap <- clusGap(customer_data[,3:5], FUN = kmeans, nstart = 25,  
                  K.max = 10, B = 50)  
fviz_gap_stat(stat_gap)
```



```
k6<-kmeans(customer_data[,3:5],  
           6,iter.max=100,nstart=50,algorithm="Lloyd")  
k6
```

## 4 Podsumowanie



K-means clustering with 6 clusters of sizes 38, 45, 22, 35, 21, 39

Cluster means:

	Age	Annual.Income..k..	Spending.Score..1.100.
1	27.00000	56.65789	49.13158
2	56.15556	53.37778	49.08889
3	25.27273	25.72727	79.36364
4	41.68571	88.22857	17.28571
5	44.14286	25.14286	19.52381
6	32.69231	86.53846	82.12821