Experiment - 7

Student Name: Rajdeep Jaiswal **UID:** 20BCS2761
Branch: CSE **Section/Group:** 20BCS-DM-902/(B)
Semester: 6th **Subject Code:** 20CSP-376
Subject Name: Data Mining Lab

1. Aim:

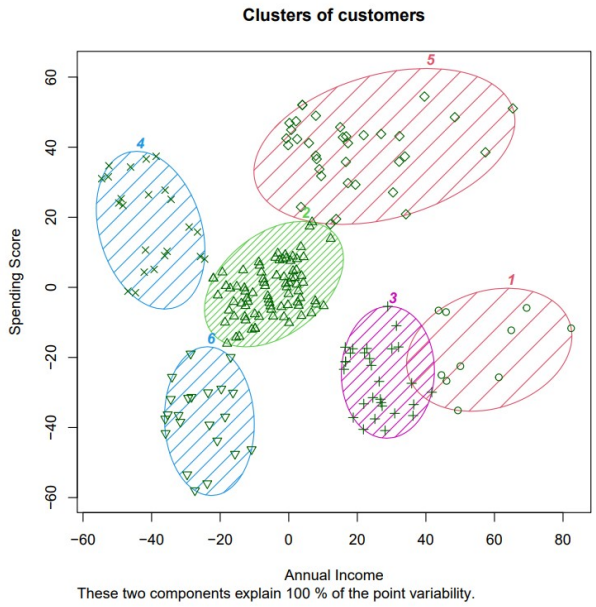
- To perform the cluster analysis by k-means method using R

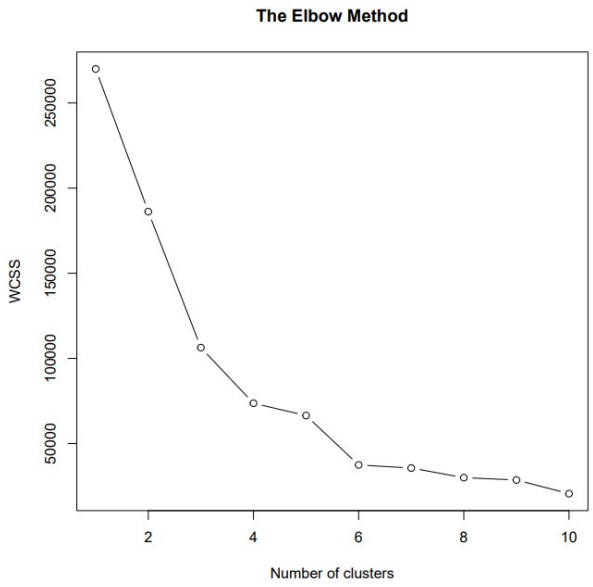
2.CODE:

```
# K-Means Clustering

setwd("C:\\Users\\Documents\\RWA")

# Importing the dataset dataset
= read.csv('mall.csv') X =
dataset[4:5]

# Using the elbow method to find the optimal number of clusters set.seed(6)
wcss = vector()

#$within ss: is the within cluster sum of squares. So it results in a vector with a number for each cluster.
for (i in 1:10) wcss[i] = sum(kmeans(X, i)$withinss)

#Initate PDF File pdf("elbow-graph.pdf",
paper="a4")

plot(x = 1:10,
y = wcss,
type = 'b',
    main = 'The Elbow Method',
xlab = 'Number of clusters',      ylab
= 'WCSS')
```

```
#Close PDF file
dev.off()


# Fitting K-Means to the dataset
set.seed(29)
kmeans = kmeans(x = X,
                centers = 6,
iter.max = 300,
                nstart = 10)
# Visualising the cluster
library(cluster)

# Initate PDF File
pdf("clusterplot.pdf", paper="a4")

clusplot(x = X,          clus =
kmeans$cluster,          lines = 0,
shade = TRUE,          color =
TRUE,          labels = 4,
plotchar = TRUE,          span =
TRUE,          main = 'Clusters of
customers',          xlab = 'Annual
Income',
         ylab = 'Spending Score')

#Close PDF file
dev.off()
```

3.OUTPUT:


Clusters of customers
Spending Score
Annual Income
These two components explain 100 % of the point variability.
1
2
3
4
5
6


The Elbow Method
WCSS
Number of clusters