Experiment 7

Student Name: Kanishk Soni **UID: 20BCS9398**
Branch: B.E CSE **Section/Group: 20_DM_708_B**
Subject Name: DM Lab

1. Aim:

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

2. Theory:

K-Means clustering groups the data on similar groups. The algorithm is as follows:

- Choose the number K clusters.
- Select at random K points, the centroids (Not necessarily from the given data).
- Assign each data point to the closest centroid that forms K clusters.
- Compute and place the new centroid of each centroid.
- After final reassignment, name the cluster as Final cluster.

3. Code and Output:

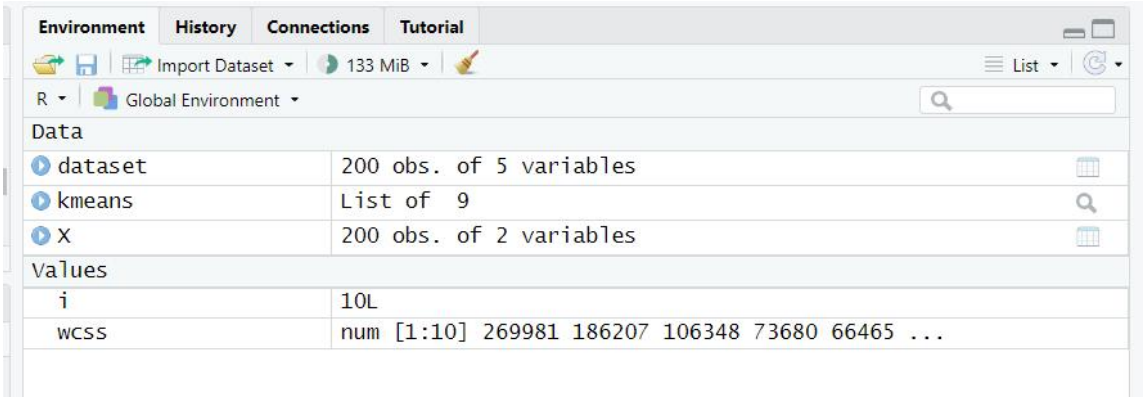
PROGRAM

```
# K-Means Clustering


setwd("D:\\semester\\Semester 6\\Data_Mining\\Experiments\\K-means clustering")


# 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()
for (i in 1:10) wcss[i] = sum(kmeans(X, i)$withinss)


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

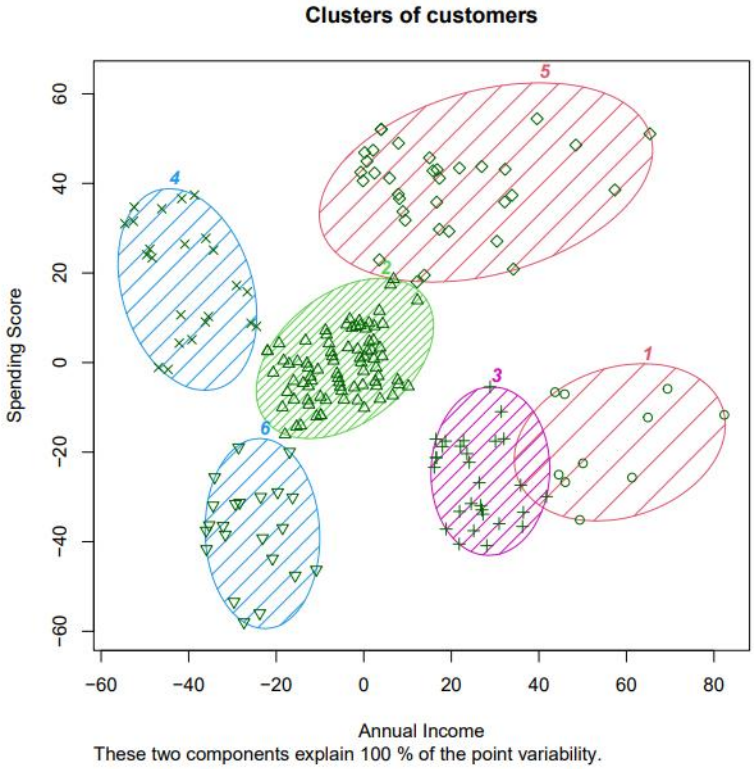
```
plot(x = 1:10,
     y = wcss,
     type = 'b',
     main = 'The Elbow Method',
     xlab = 'Number of clusters',
     ylab = 'WCSS')

#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()
```

OUTPUT



File: clusterplot.pdf



File: elbow-graph.pdf