



Experiment 7

Student Name: Husanpreet Kaur

Branch: BE-CSE

Semester: 6

Subject Name: Data Mining Lab

UID: 20BCS9512

Section/Group: 20BCS-DM-714/A

Subject Code: 20CSP-376

Date of Performance: 02-05-2023

1. Aim/Overview of the practical: To perform the cluster analysis by k-means method using R.

2. Tools used: RStudio

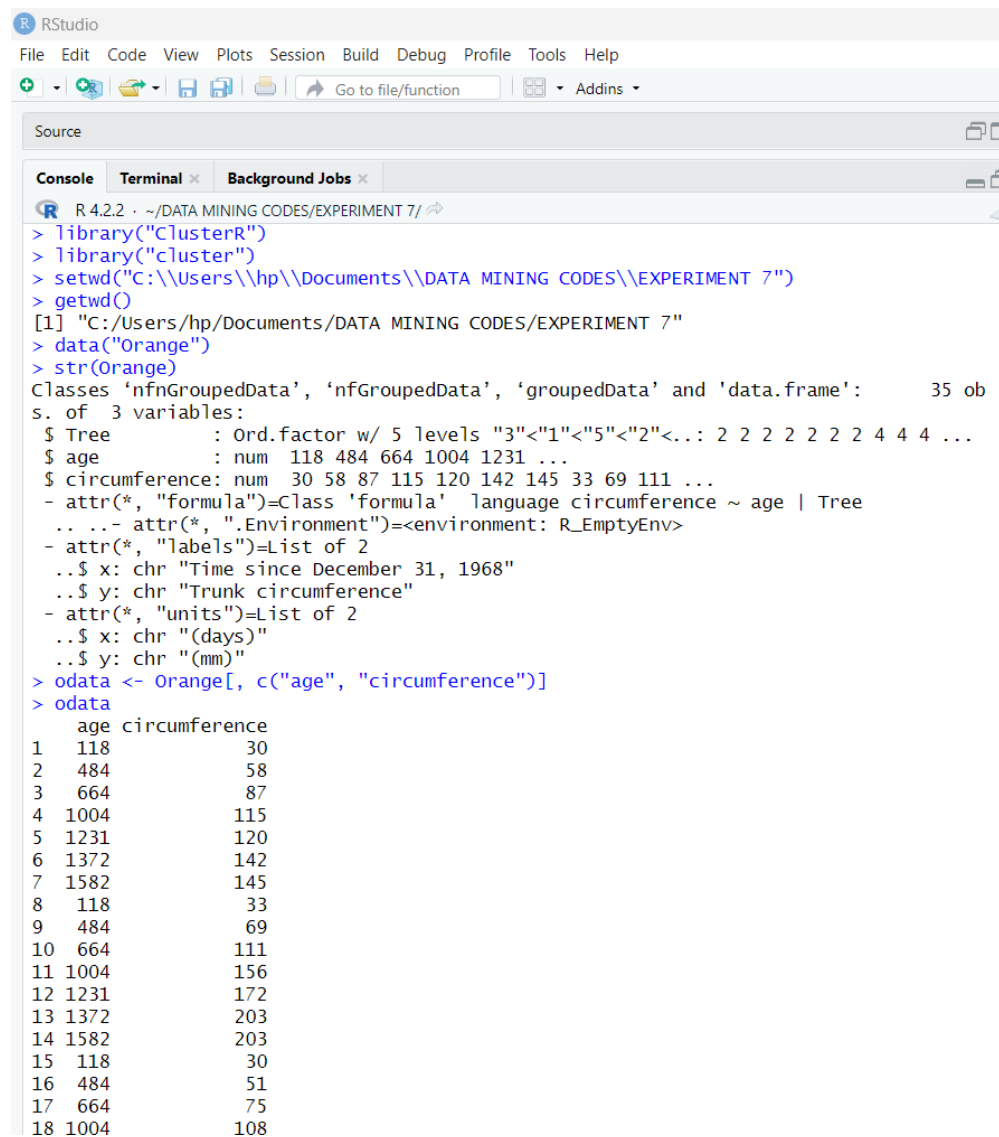
3. Code:

```
library("ClusterR")
library("cluster")
setwd("C:\\Users\\hp\\Documents\\DATA MINING
CODES\\EXPERIMENT 7")
getwd()
data("Orange")
str(Orange)
odata <- Orange[, c("age", "circumference")]
odata
set.seed(123)
k <- kmeans(odata, centers = 3, nstart = 10)
k
k$cluster
confusionmatrix <- table(Orange$Tree, k$cluster)
confusionmatrix
plot(odata[c("age", "circumference")])
plot(odata[c("age", "circumference")], col = k$cluster)
plot(odata[c("age", "circumference")], col = k$cluster, main = "K-means
with 3 clusters")
```

```
k$centers
k$centers[, c("age", "circumference")]
points(k$centers[, c("age", "circumference")], col = 1:3, pch = 8, cex = 3)
y_kmeans <- k$cluster
clusplot(odata[, c("age", "circumference")], y_kmeans, lines = 0, shade =
TRUE, color = TRUE, labels = 2, plotchar = FALSE, span = TRUE, main =
paste("Cluster Orange"), xlab = 'age', ylab = 'circumference')
```

4. Output:

RStudio:

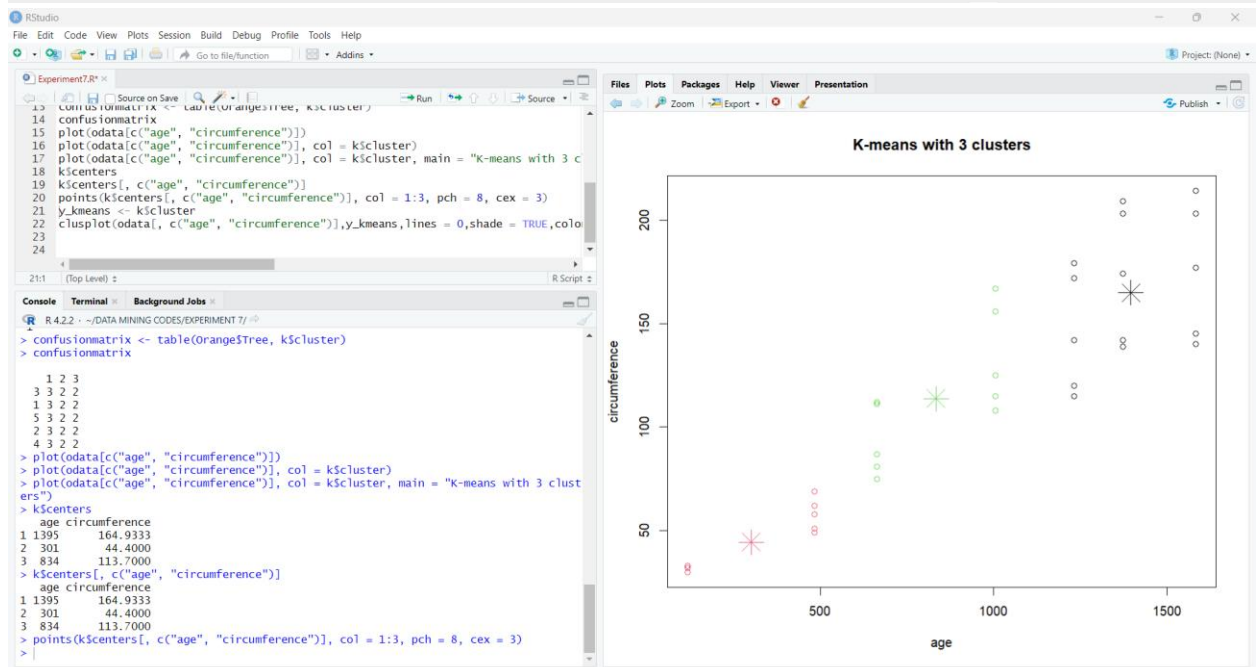
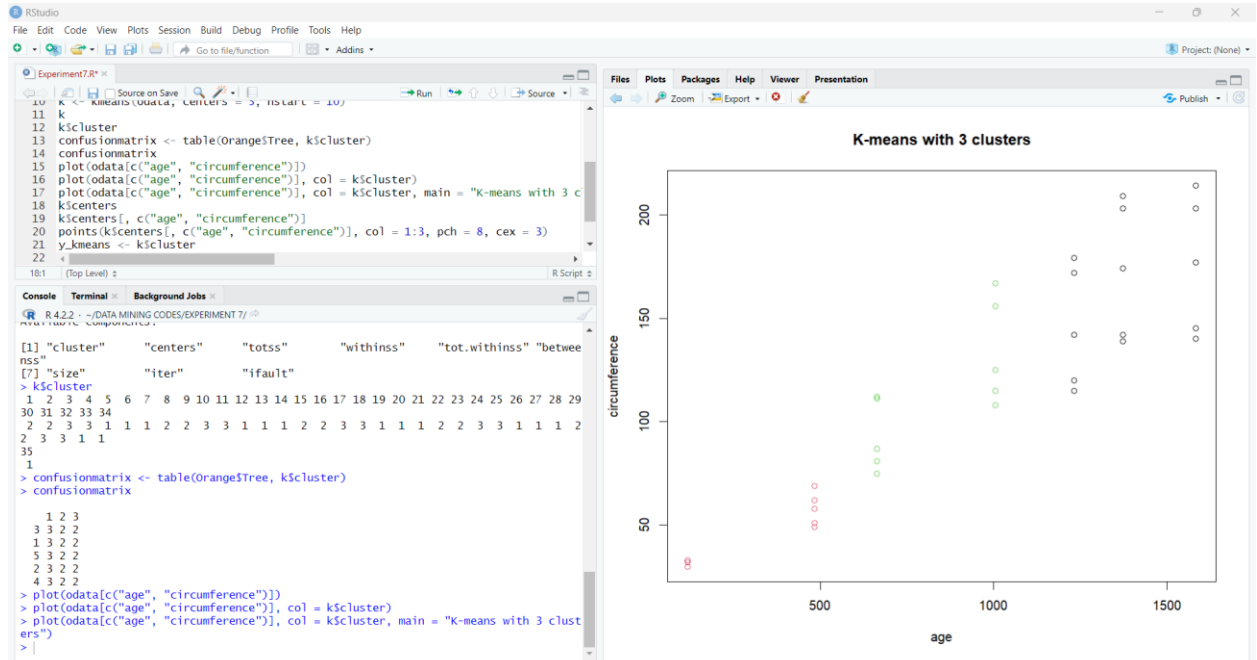


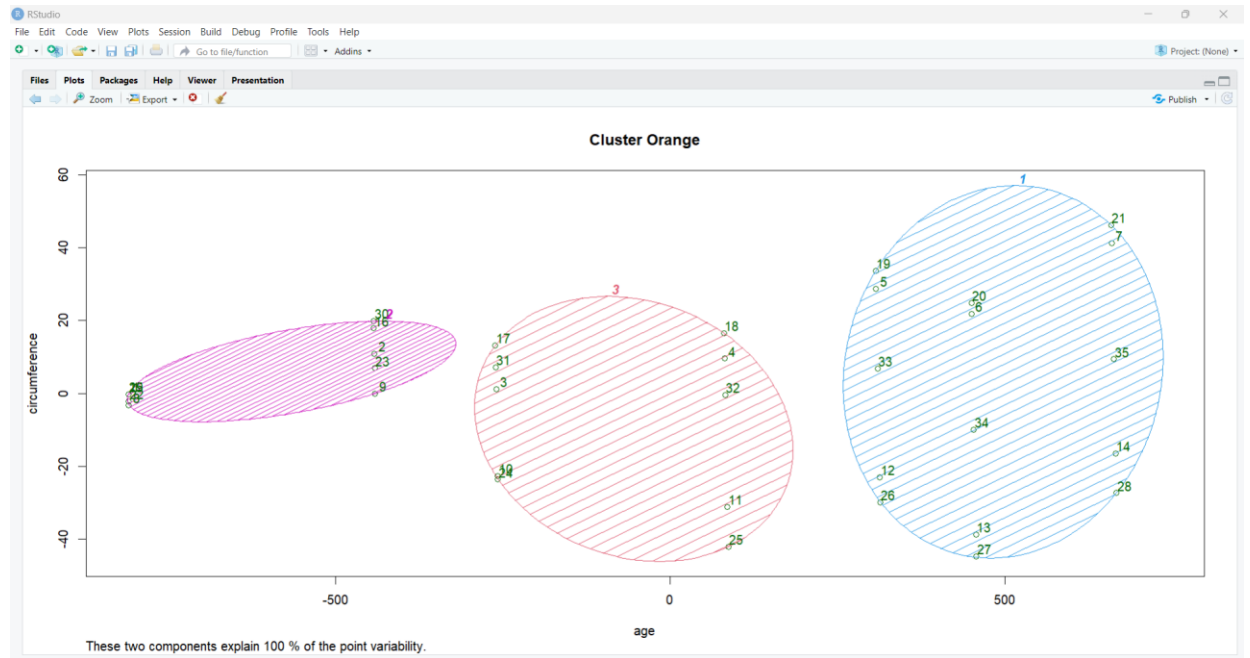
```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ - [Icons] Go to file/function [Grid] Addins
Source
Console Terminal Background Jobs
R 4.2.2 · ~/DATA MINING CODES/EXPERIMENT 7/
> library("ClusterR")
> library("cluster")
> setwd("C:\\Users\\hp\\Documents\\DATA MINING CODES\\EXPERIMENT 7")
> getwd()
[1] "C:/Users/hp/Documents/DATA MINING CODES/EXPERIMENT 7"
> data("Orange")
> str(Orange)
Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':   35 ob
s. of 3 variables:
 $ Tree      : Ord.factor w/ 5 levels "3"<"1"<"5"<"2"<...: 2 2 2 2 2 2 2 4 4 4 ...
 $ age       : num  118 484 664 1004 1231 ...
 $ circumference: num  30 58 87 115 120 142 145 33 69 111 ...
- attr(*, "formula")=Class 'formula' language circumference ~ age | Tree
.. ..- attr(*, ".Environment")=<environment: R_EmptyEnv>
- attr(*, "labels")=List of 2
..$ x: chr "Time since December 31, 1968"
..$ y: chr "Trunk circumference"
- attr(*, "units")=List of 2
..$ x: chr "(days)"
..$ y: chr "(mm)"
> odata <- Orange[, c("age", "circumference")]
> odata
  age circumference
1  118             30
2  484             58
3  664             87
4 1004            115
5 1231            120
6 1372            142
7 1582            145
8  118             33
9  484             69
10 664            111
11 1004            156
12 1231            172
13 1372            203
14 1582            203
15  118             30
16 484             51
17 664             75
18 1004            108
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.





5. Observation:

- Learnt how to use R and create a file in Rstudio.
- Learnt how to install packages in Rstudio.
- Learnt how to make clusters and plot data.
- Learnt how to load dataset Orange in Rstudio.
- Learnt the use of cluster and ClusterR libraries.
- Learnt how to create table and confusion matrix.