**Department of Decision and Computing Sciences**

17MDC46 - Predictive Analysis

PROJECT WORK

Submitted by

Thilak P.L.[2033037]

Santhosh K V[2033030]

Devamuge K U[2033007]

Submitted to

Associate Professor

Department of Computing

A picture containing text

Description automatically generated

**COIMBATORE INSTITUTE OF TECHNOLOGY**

**(Government Aided Autonomous Institution)**

**Coimbatore-641014**

**ANALYSIS OF THE STUDY ON STUDENTS PERCEPTION IN OFF - LINE CLASSES**

**1.ABSTRACT:**

The study aims at analyzing the perception of teachers and students about offline classes. The work tries to explain the opinions of students as regards the impact of offline courses, their comfortability in its usage, and the support received from teachers in offline classes along with teachers' opinions on efficiency, teaching practice followed and training received for an offline class.

**2.PROBLEM STATEMENT:**

Academic institutions continuously run course experience surveys among students to know more about their feelings towards the classes. There should be an assurance that students enjoy the every tasks performed in class.

**3.INTRODUCTION:**

Student feedback is essential for teachers and academic institutes to improve continuously. If you work in academia, it is surely a great idea to know the perceptions and opinions of students. Schools run surveys at the start, the middle, or the end of the academic year. Growing up, every child needs a productive ecosystem. Schools and universities are among the most influential parts of a child’s ecosystem, and support of classmates/friends matters to each child. Every student is unique. Their achievements will be different, and each of these achievements should be celebrated.

**4.OBJECTIVES:**

* To push students boundaries so that they perform well academically as well in terms of extracurricular activities with constant encouragement
* To understand if students face difficulties with the professor.
* To minimize or eliminate the stress of homework.
* To provide students with adequate sports facilities to keep their minds strong.
* To boost a student’s confidence in making their own academic decisions.

**5.METHODOLOGY USED:**

* Elbow method
* Hierarchical cluster – Complete Linkage
* Average linkage
* Silhouette method
* K – means clustering

**6.SOFTWARE USED:**

* R – STUDIO SOFTWARE – It is used to built code for the survey.
* CSV FILE – It is the dataset of the survey.
* Google forms – It is used for collecting the responses.

**7.EXPERIMENTAL ANALYSIS:**

**7.1.DATASET USED:**

* The dataset used for this survey is collected by distribution of the google forms to school and college students through social media like whatsapp, instagram, etc,…
* The questionnaire consist of 16 question which is divided into three sections namely students perception about the class, students perception about the teacher and agreeability towards teachers.
* Finally 50 responses have been reached within in two weeks.
* This is converted into csv file and it is used as the dataset.

**7.2.SOURCE CODE:**

data<-read.csv("pacat.csv")

data<-data[,1:16]

summary(data)

unique(data$AGE)

data[data==""]<-NA

data<-data[complete.cases(data), ]

library(dplyr)

dataset<-data %>%select\_if(is.numeric)

character<-data %>%select\_if(is.character)

library("fastDummies")

character<-dummy\_cols(character, remove\_most\_frequent\_dummy=TRUE)

dataset[,1:16] <-scale(dataset[,1:16])

library ("factoextra")

fviz\_nbclust(dataset, kmeans, method="wss")

labs(subtitle="Elbow Method")

a<-read.csv("pacat.csv")

plot(AGE~ACTIVITY,a)

with(a,text(AGE~ACTIVITY,labels=AGE,pos=4,cex=.6))

m<-apply(a,2,mean)

s<-apply(a,2,sd)

z<-scale(a,m,s)

distance<-dist(z)

print(distance,digits=3)

hc.c<-hclust(distance)

plot(hc.c, labels= a$AGE)

plot(hc.c, hang=-1)

hc.a<-hclust(distance,method="average")

plot(hc.a,hang=-1)

member.c<-cutree(hc.c,3)

member.a<-cutree(hc.a,3)

table(member.c,member.a)

aggregate(z,list(member.c),mean)

aggregate(a[,-c(1,1)], list(member.c),mean)

library(cluster)

plot(silhouette(cutree(hc.c,3),distance))

wss<-(nrow(z)-1)\*sum(apply(z,2,var))

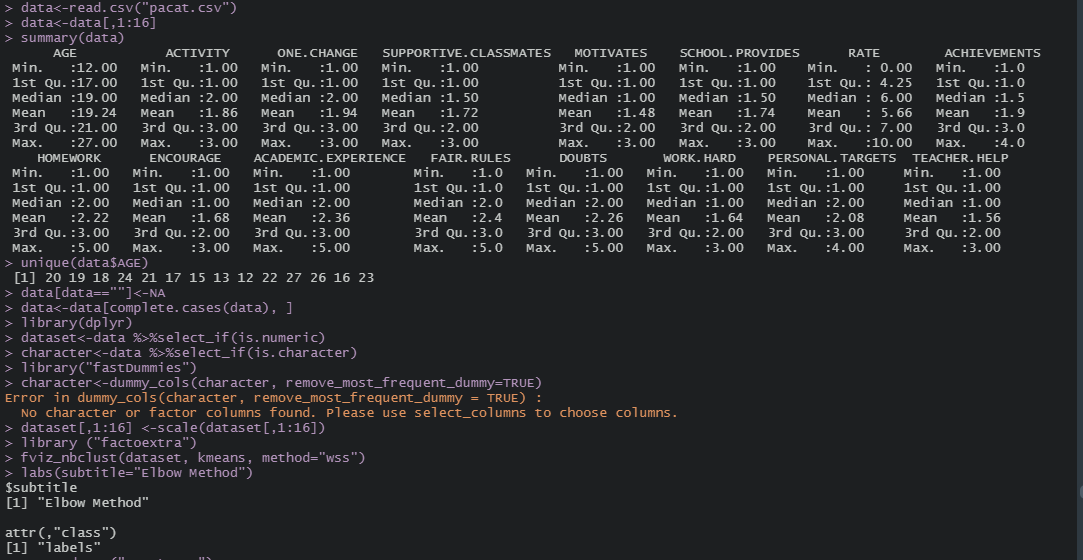
for(i in 2:20) wss[i] <- sum(kmeans(z,centers=i)$withinss)

plot(1:20, wss, type="b", xlab="Number of Clusters", ylab="within Group SS")

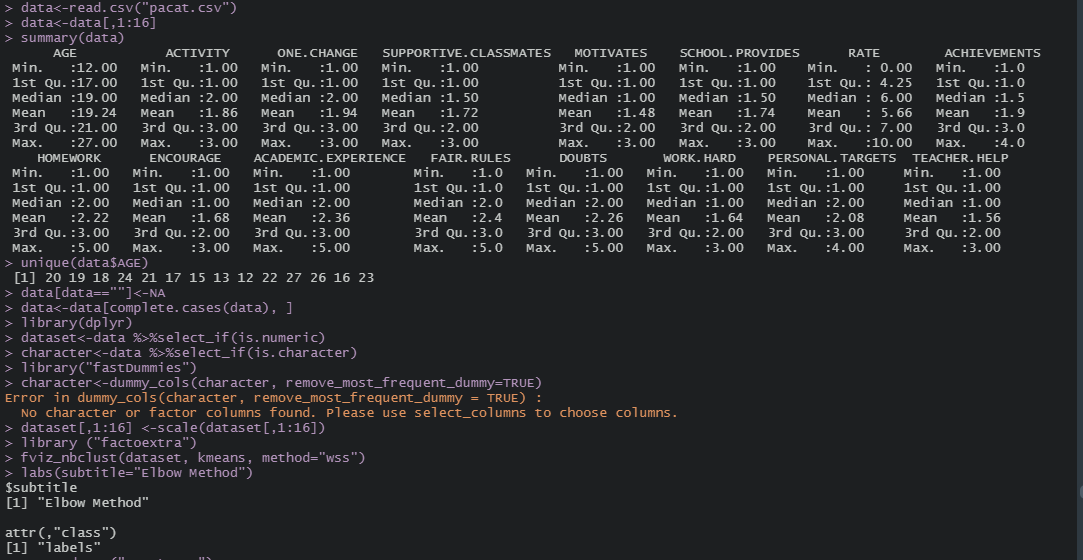
kc<-kmeans(z,3)

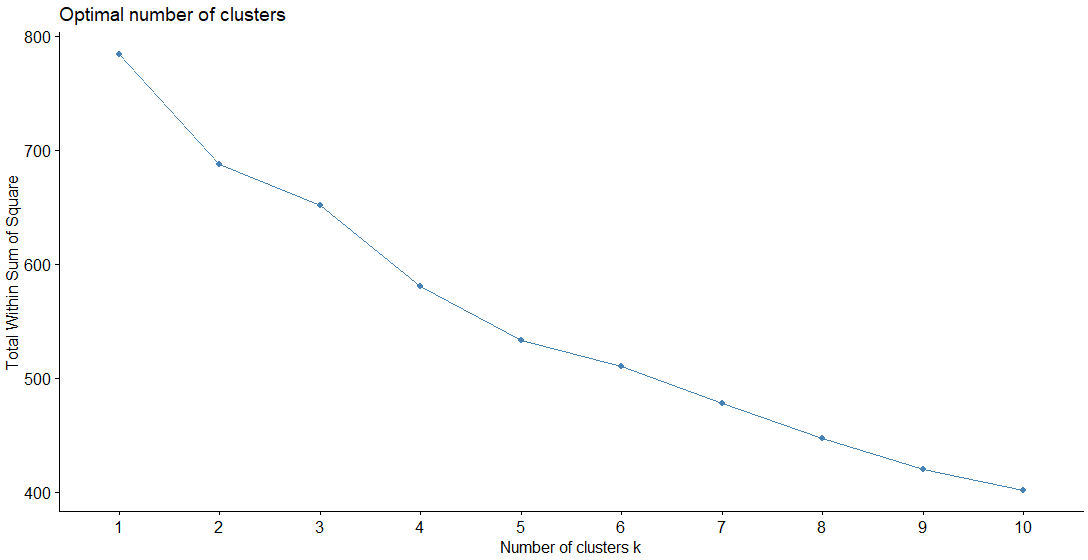
kc

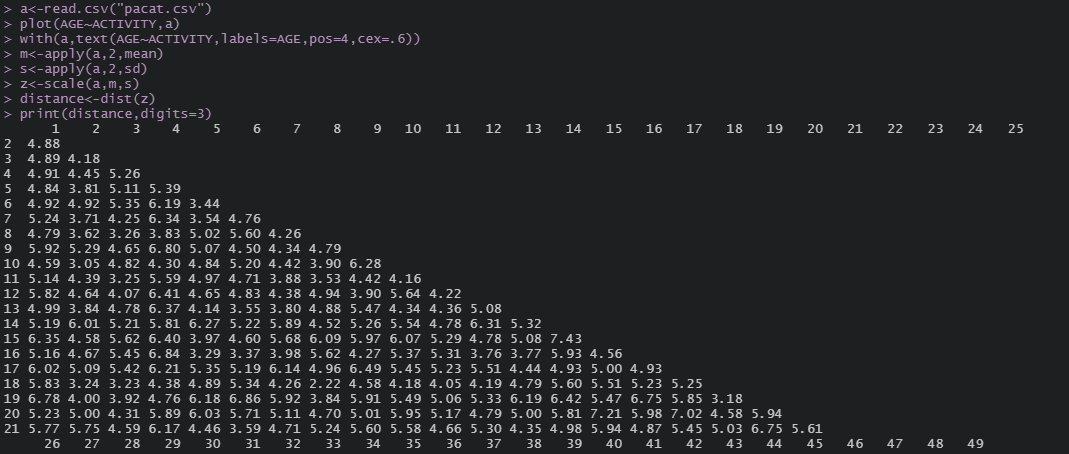
plot(ONE.CHANGE~SUPPORTIVE.CLASSMATES,a,col=kc$cluster)

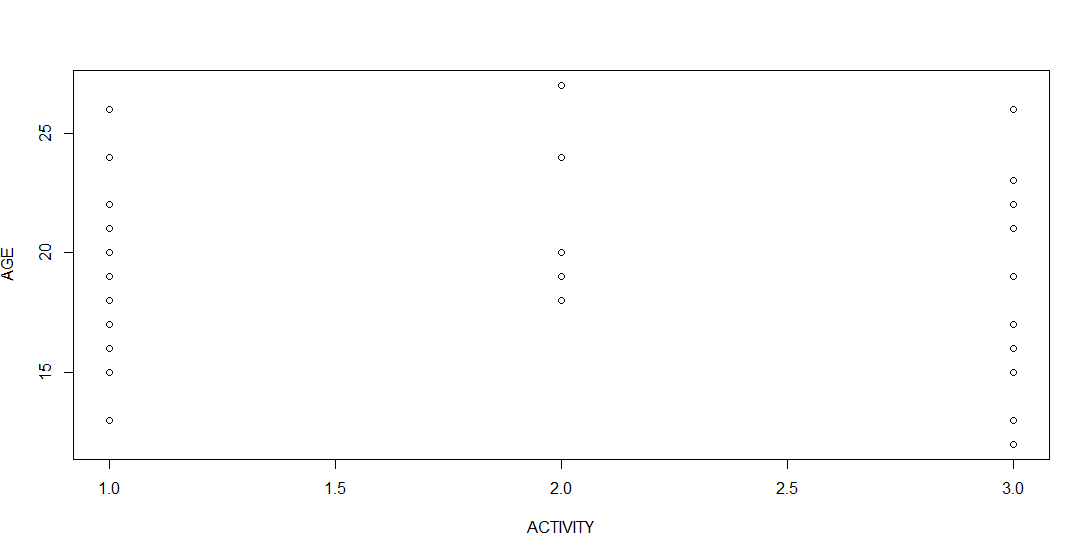
**7.3.RESULTS:**

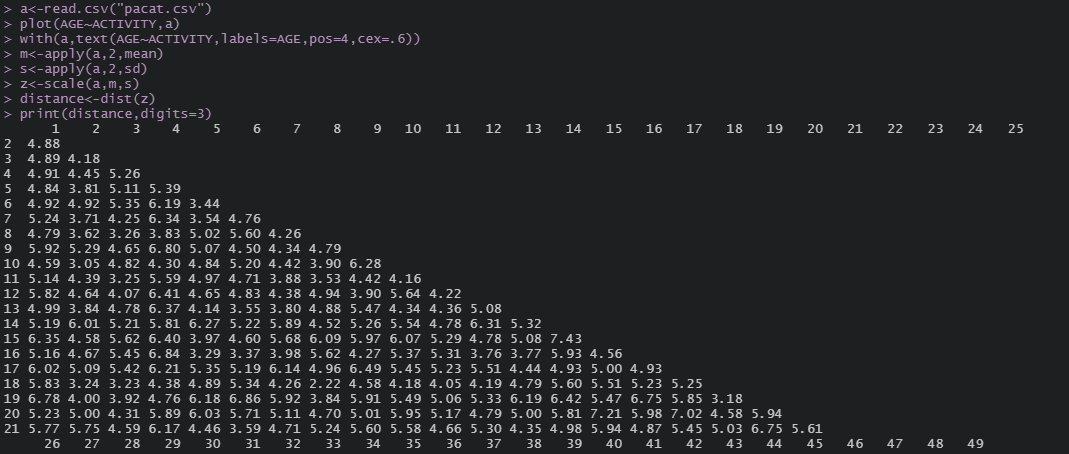
C:\Users\HP\Downloads\GG.PNG

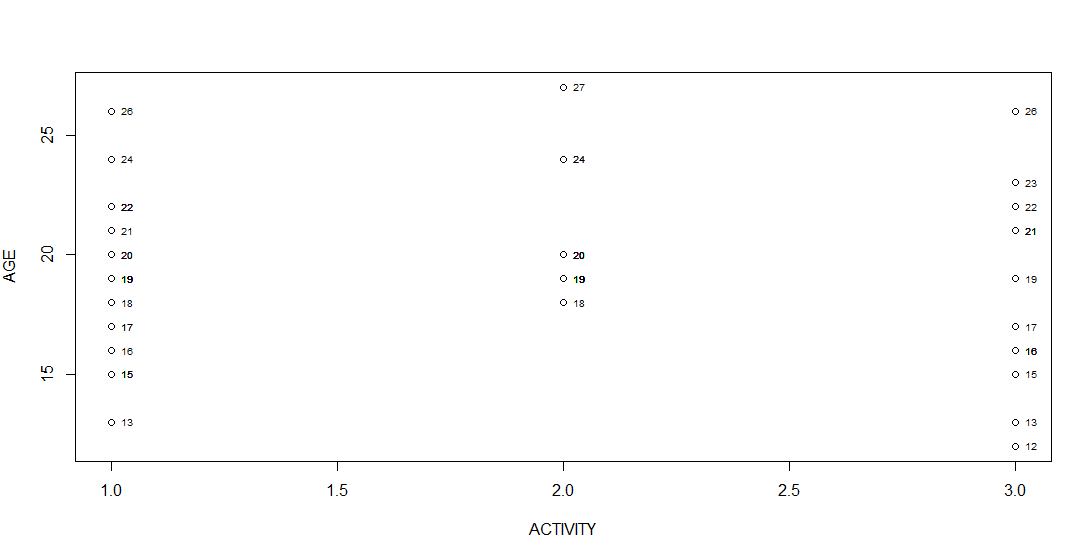


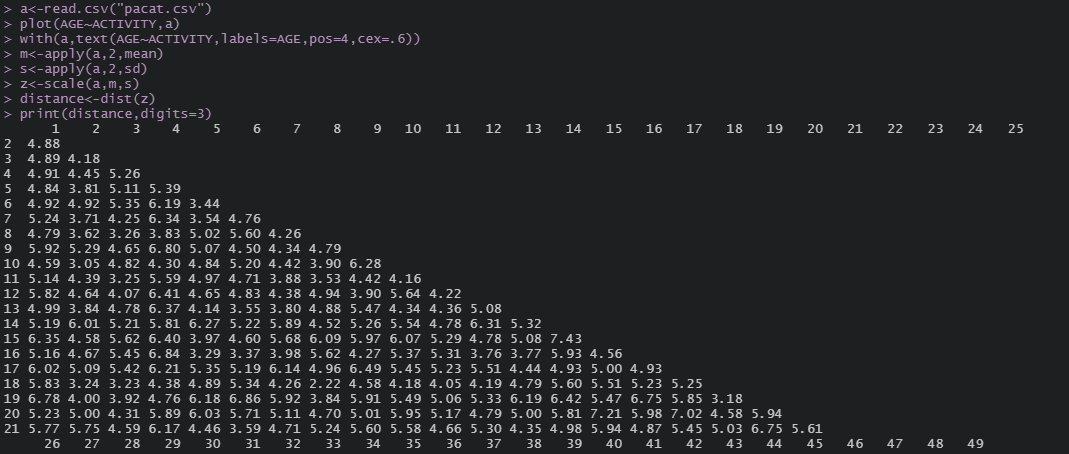




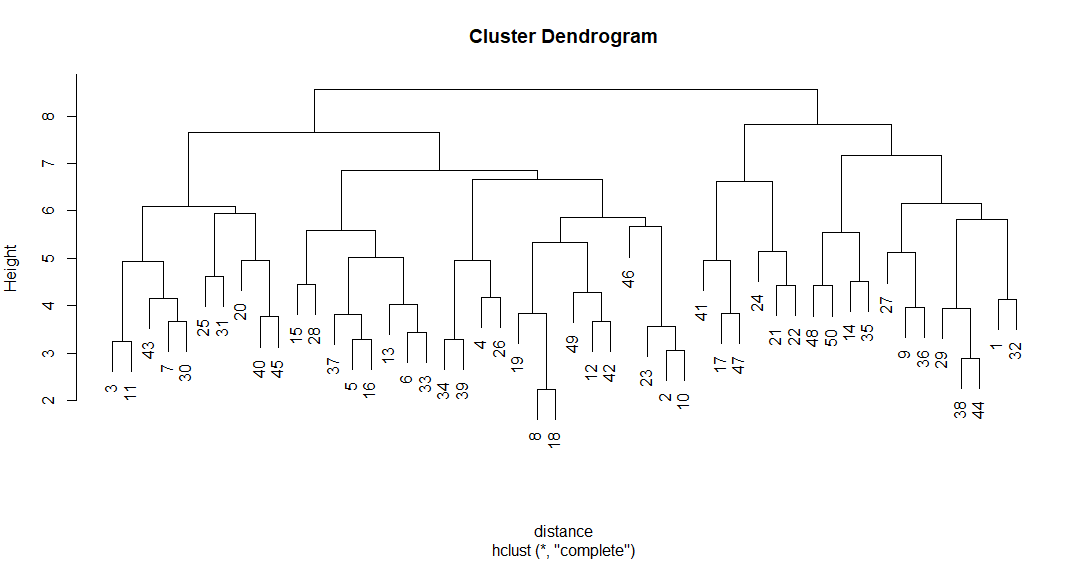




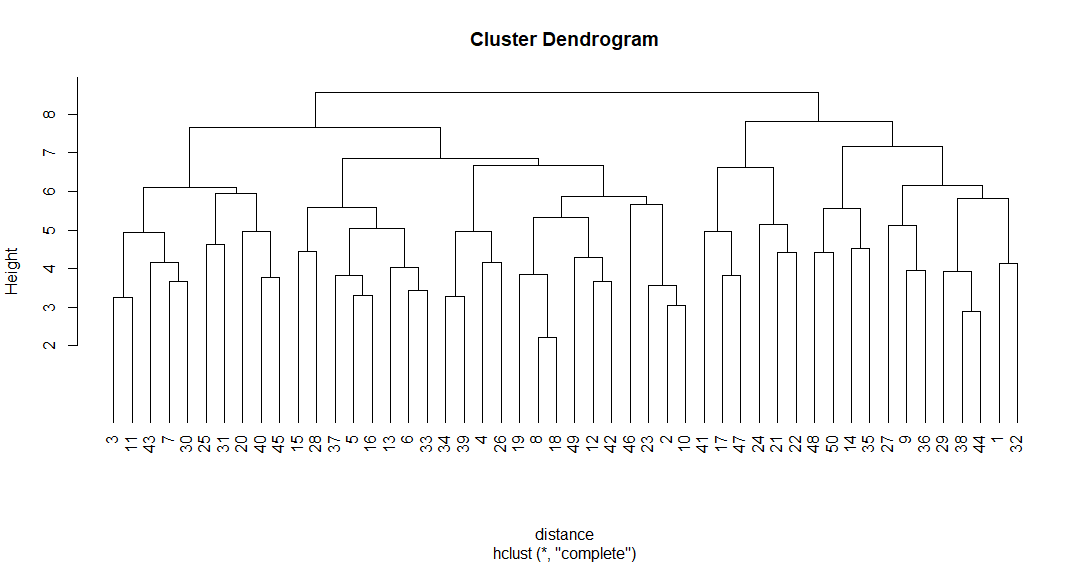




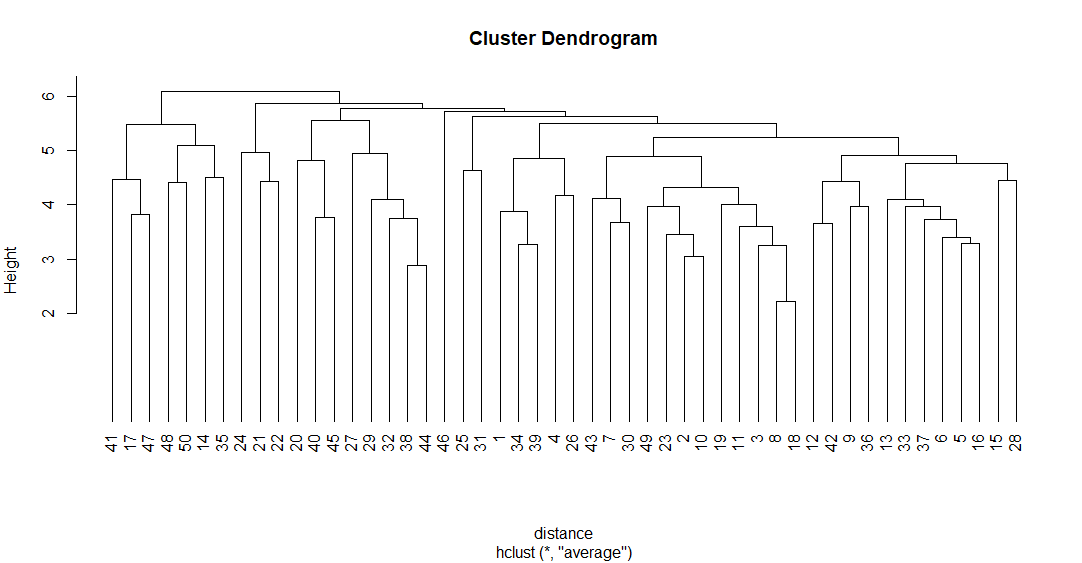




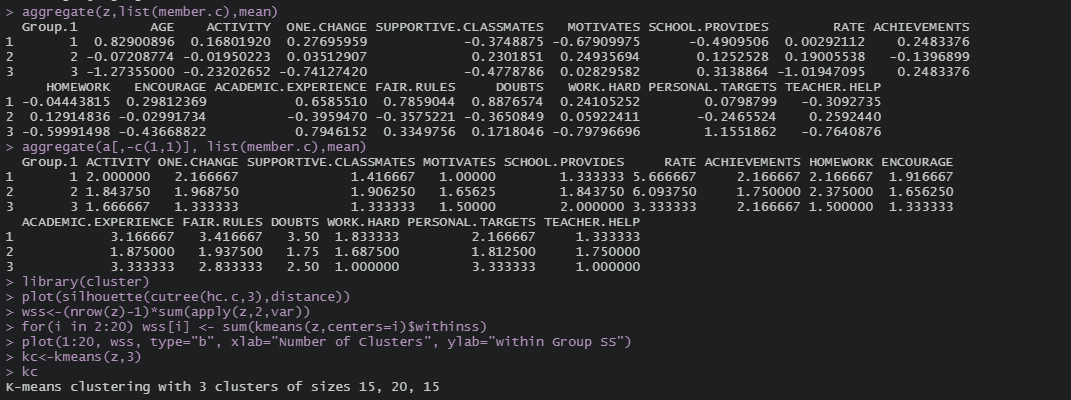


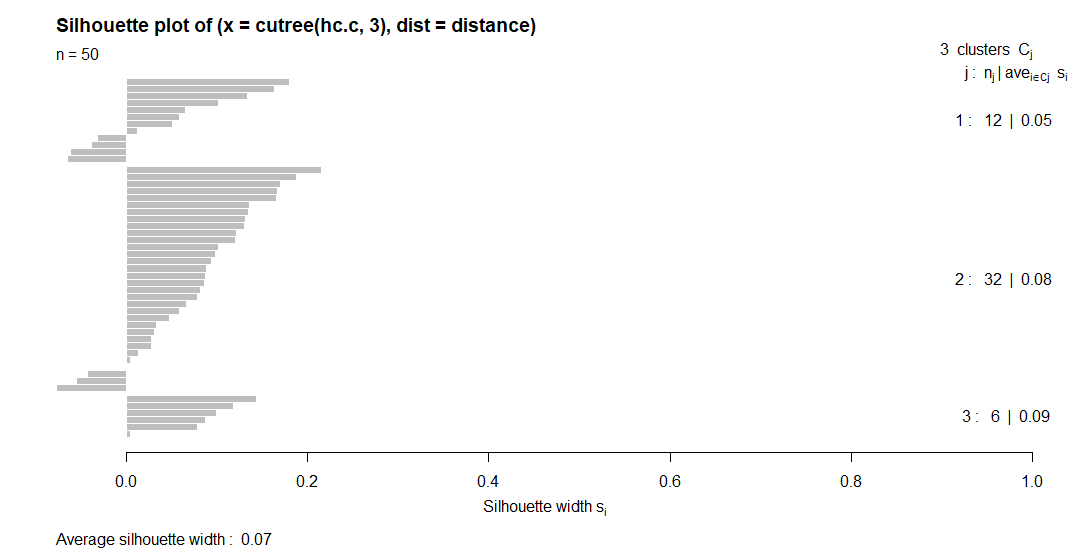


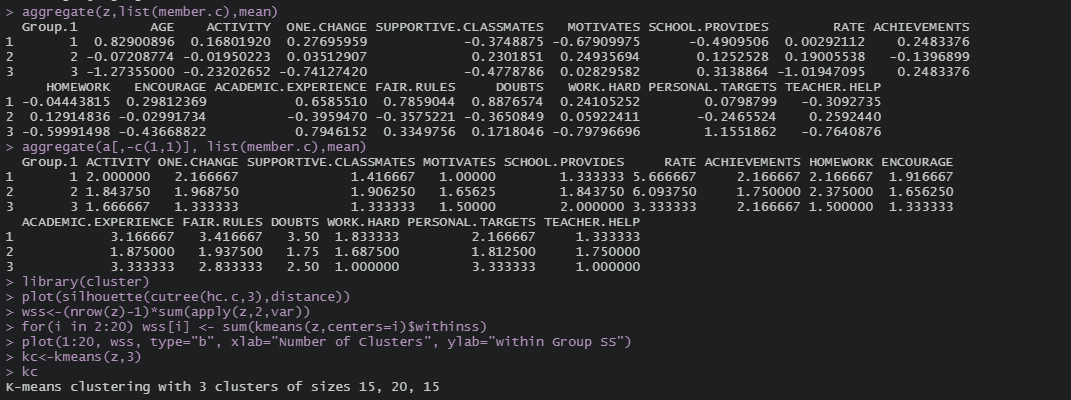


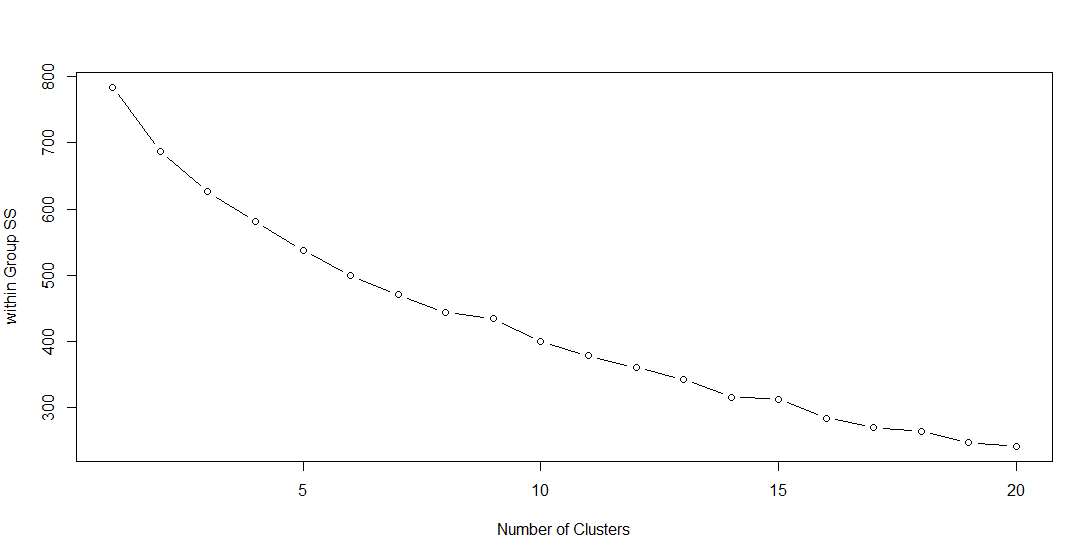


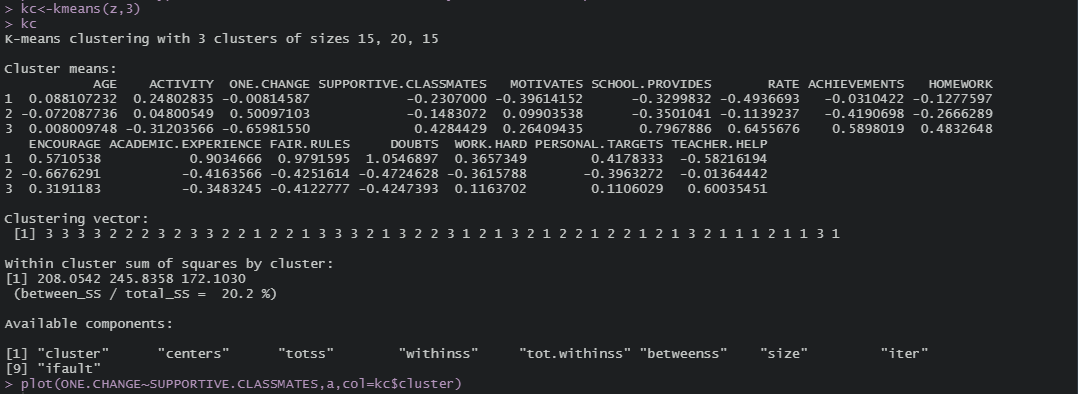












**7.4.RESULT INTERPRETATION:**

* **UNIQUE OF THE DATASET:** The unique function is used to remove the duplicates from the dataset and display all the ages only one time.
* **CHECK OF NUMERICAL VARIABLES:**The check of the dataset whether it has numeric or not, it summarized that the dataset has 50 observations with 16 numerical variables.
* **CHECK OF CHARACTER VARIABLES:**The check of the dataset whether it has any characters or not, it summarized that the dataset has 50 observations with 0 character variables.
* **DUMMY NUMERIC:**The dummy\_cols() function is used to fill dummy numeric to the column and it is inferred from the error statement that the dataset has no character or factor column.
* **ELBOW METHOD GRAPH:**The graph decreases till the number of clusters is 10 at 400 totla within sum of squares.
* **SCATTER PLOT:**The scatter plot is just simply plotted to have a general idea of the further clusters.It shows that there might be three possible clusters.
* **DENDROGRAM:**The dendrogram is generally used to combine individual clusters and form a single cluster. In hierarchical clustering of complete linkage, the single cluster is achieved at the height of 9. In hierarchical clustering of average linkage, the single cluster is achieved at the height of 6.
* **TABLE VALUES OF AVERAGE AND COMPLETE LINKAGE**: 

As per the complete linkage, the first cluster has 8+4 observations, second cluster has 32 observations and the third cluster has 3+3 observations. In the same way, the first cluster of average linkage has 8+32 observations, the second cluster has 4+3 observation and the third cluster has 3 observations.It is also seen that the observations which lie on the diagonal of the table will perfectly fit in the particular cluster. Here 8 and 3 observations are perfectly fitted in the first and third clusters respectively of both the average and complete linkage.

* **AGGREGATION:** From the aggregate values we can find the questions which have high value will belong to the corresponding cluster. Here from the supportive classmates question the highest aggregate value is 0.2301851 and belongs to the second cluster.
* **SILHOUETTE PLOT:** If the cluster formation is good or the members in the clusters are closer to each other then the SI values will be high or else it will be low.This kind of visiualization helps to identify the clusters visually.if we have negative SI values obviously that member in the group is sort of outlier, that will not belong to that group.
* **SCREE PLOT:**It gives the over view of the all possible clusters and within group sum of squares. Here after three points there is no big difference in the within group sum of squares. So it is good to have three clusters.
* **K – MEANS:** The size of three clusters from the k – means clustering is found to get fitted are 15, 20 and 15 questions.
* **CLUSTERING VECTOR**: From the clustering vector it is found that the first question should belong to the third cluster, second to third and so on. 15 questions belongs to the first cluster, 20 questions belongs to second cluster and 15 questions belongs to third cluster.
* **WITHIN CLUSTER SUM OF SQUARES:** The within cluster variability for the first, second and third cluster with 15,20 and 15 questions is found to be 208.0542, 245.8358 and 172.1030 respectively. The third cluster with 15 questions has the lower variability which means the questions are closer to each other in terms of distance.It is also found that the between sum of square is divided by total sum of square is 20.2%.

**8.CONCLUSION:**

Teachers are generally aware of the most loved classroom activities, but knowing it directly from the students is an assurance. A student’s opinion is always unadulterated; it feels like a breath of fresh air in teaching monotony. The survey is essential for the overall development of the student. It helps them build a strong mind. Ensure to provide students with adequate sports facilities to keep their minds strong.

**9.REFERENCE:**

1. <https://www.questionpro.com/blog/student-survey/amp/>

2. <https://www.youtube.com/watch?v=5eDqRysaico&feature=youtu.be>

3. <https://www.youtube.com/watch?v=Zuq1jh8PWpo>