

**American International University - Bangladesh (AIUB)**

**INTRODUCTION TO DATA SCIENCE [E]**

**Name: Shanto Kumar Basak**

**ID: 20-42945-1**

**Faculty Name: Tohedul Islam**

**Date of Submission: 30th April 2023**

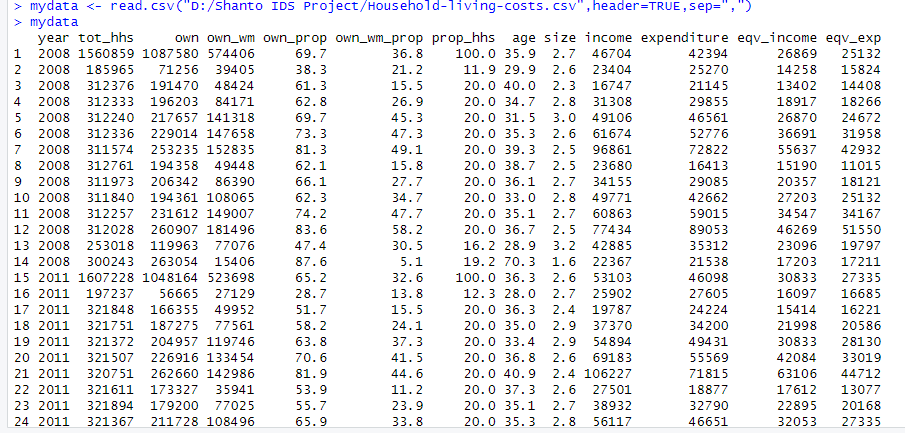
**Final Term Project (Applying K-means)**

**Introduction:** The straightforward and widely used unsupervised machine learning approach K-means clustering. Unsupervised algorithms often draw conclusions from datasets using just the input vectors and no knowledge of the known, or labeled, results. Household Living Cost dataset collected from <https://www.stats.govt.nz/large-datasets/csv-files-for-download/> this site.

1. **Observing the Dataset**

mydata <- read.csv("D:/Shanto IDS Project/Household-living -costs.csv",header=TRUE,sep=",")

mydata

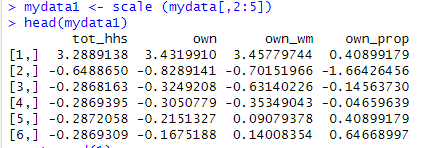


1. **Standarized the Data**

mydata1 <- scale (mydata[,2:5])

head(mydata1)

set.seed(1)

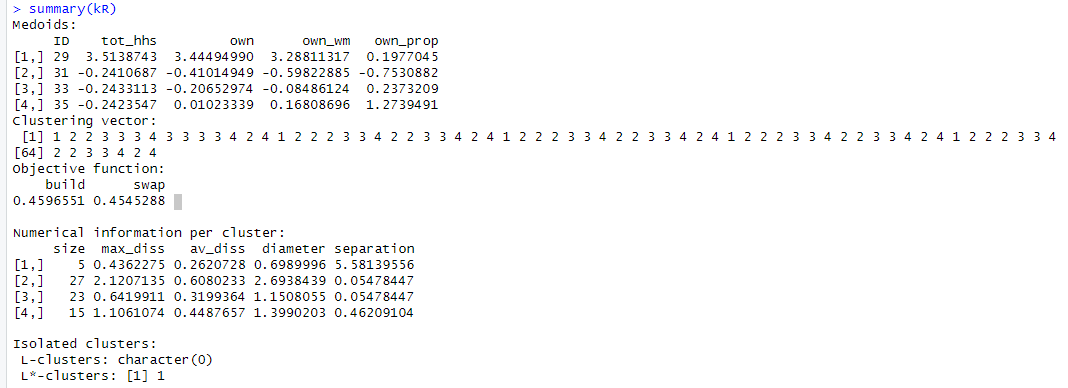


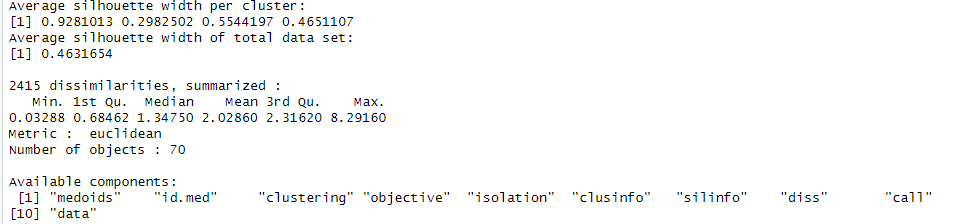
1. **Clustering Result**

kR<- pam(mydata1,k=4)

summary(kR)







1. **Cluster Structure**

mydata2 <-data.frame(mydata,kR$clustering)

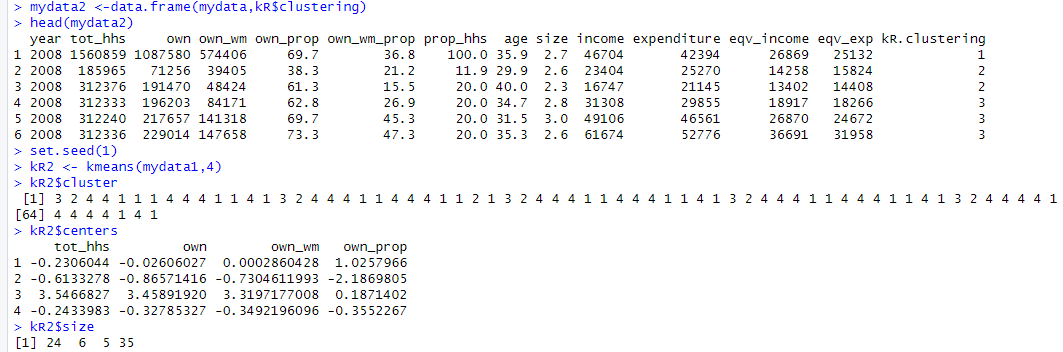
head(mydata2)

set.seed(1)

kR2 <- kmeans(mydata1,4)

kR2$cluster

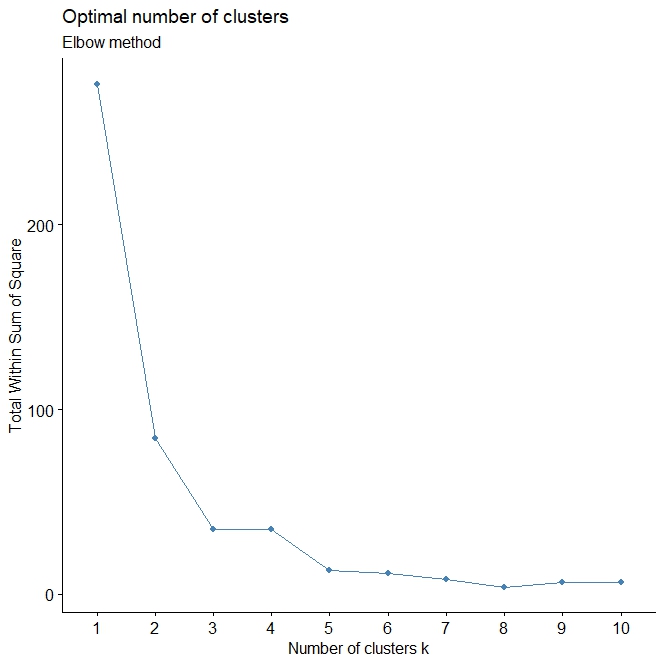
kR2$centers



**Elbow Method:**

fviz\_nbclust(mydata1, kmeans, method = "wss", diss=NULL) +

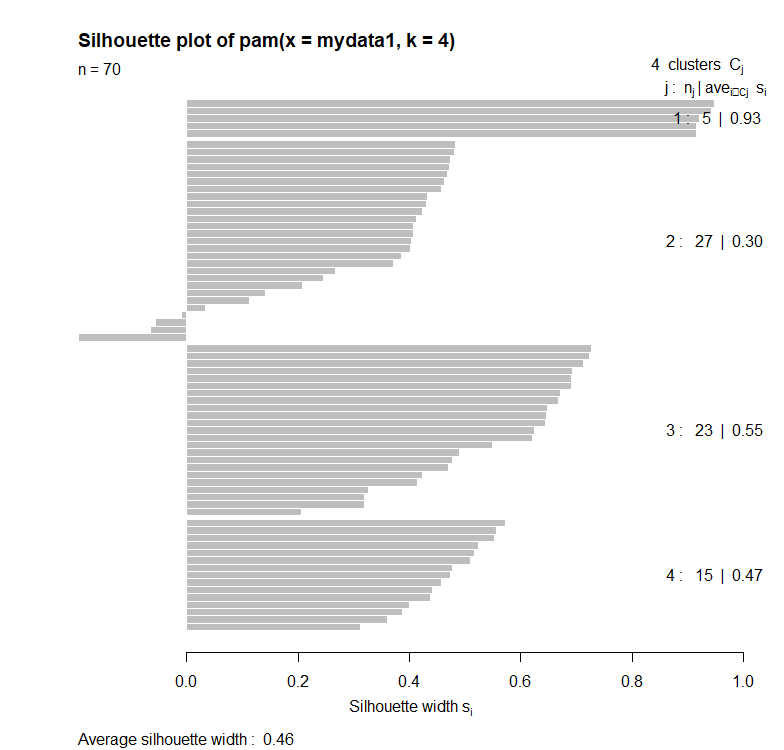
labs(subtitle = "Elbow method")



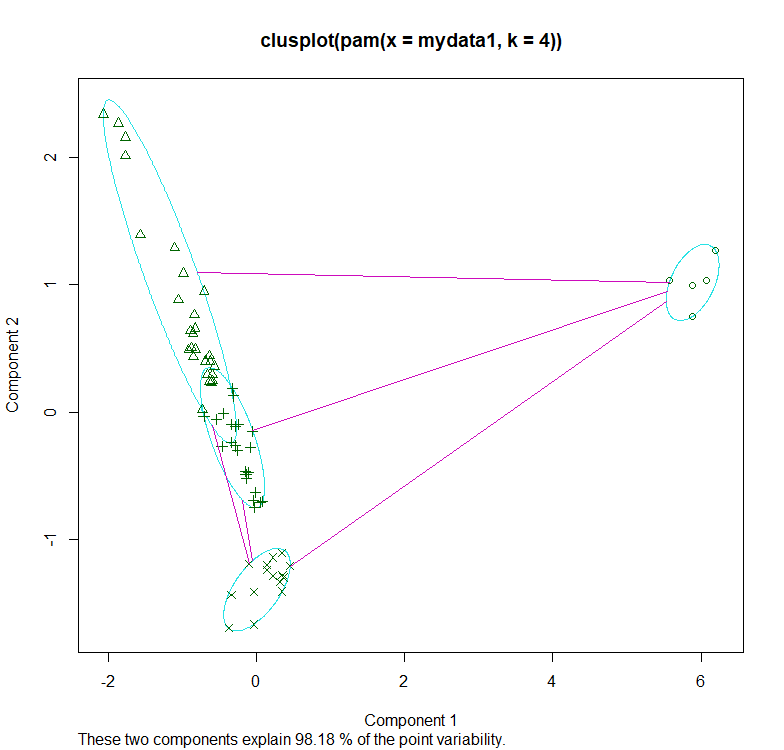
1. **Cluster and Silhouette Plot**

plot(kR)

**Silhouette Plot:**



**Cluster Plot:**



**Conclusion:** K-means clustering is an unsupervised machine learning method that is a component of a vast array of data approaches and operations in the field of data science. Data points are categorized using kmeans into unique, non-overlapping groupings. It is very easy to put into practice. Cluster generalization for various sizes and forms.

**References:**

1. <https://www.stats.govt.nz/large-datasets/csv-files-for-download/>
2. <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>
3. <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
4. <https://www.geeksforgeeks.org/k-means-clustering-introduction/>
5. <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>
6. <https://www.analyticsvidhya.com/blog/2021/11/understanding-k-means-clustering-in-machine-learningwith-examples/>