**American International University-Bangladesh**

Amena Akter

Id: 20-42761-1

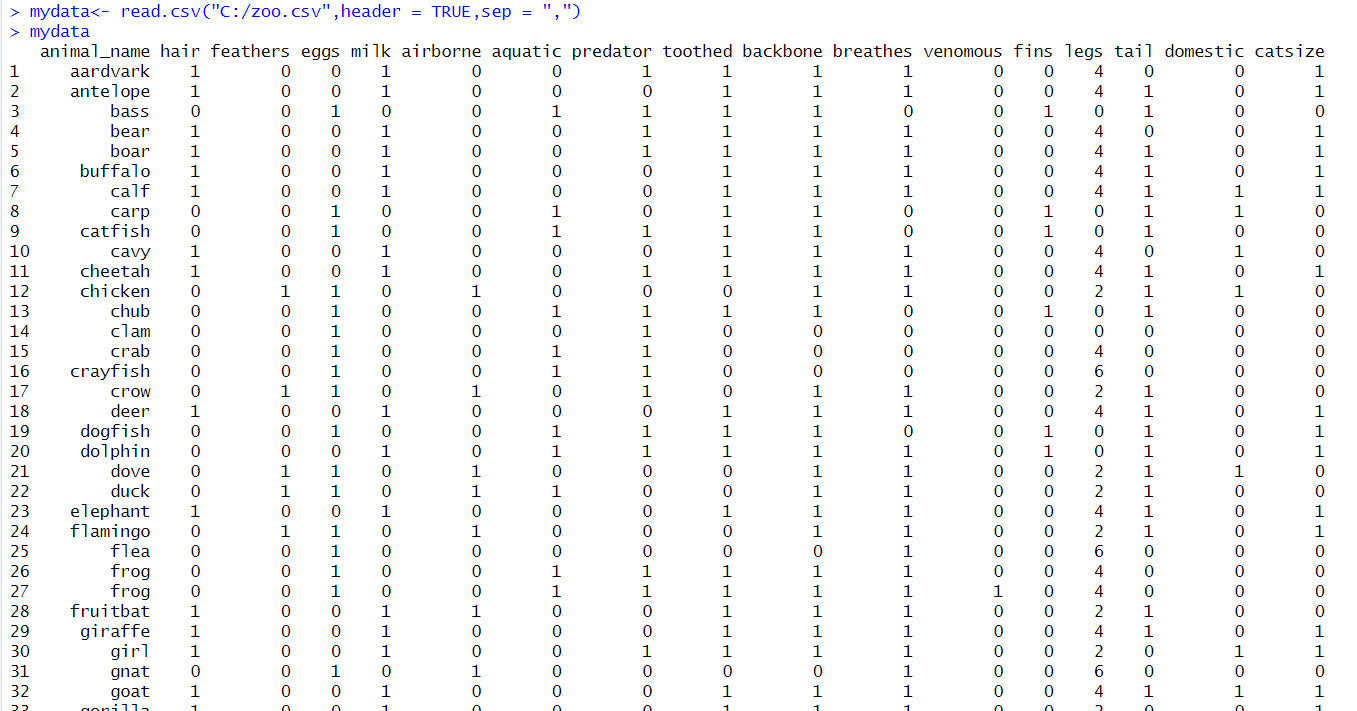
Data Science(D)

Dataset Sources:

https://www.kaggle.com/search?q=numeric+dataset

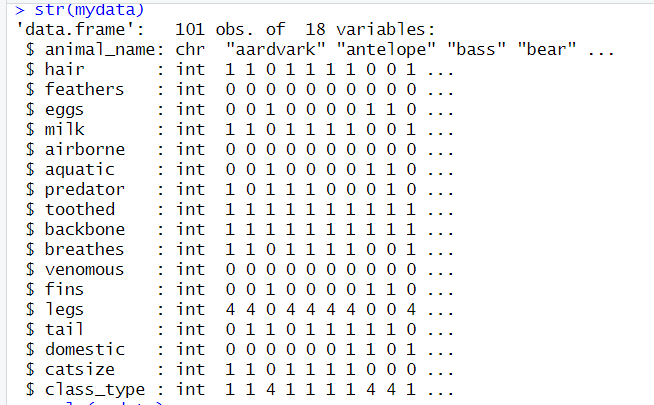
**Import Dataset**

mydata<- read.csv("C:/zoo.csv",header = TRUE,sep = ",")

mydata

**Find data for attributes**

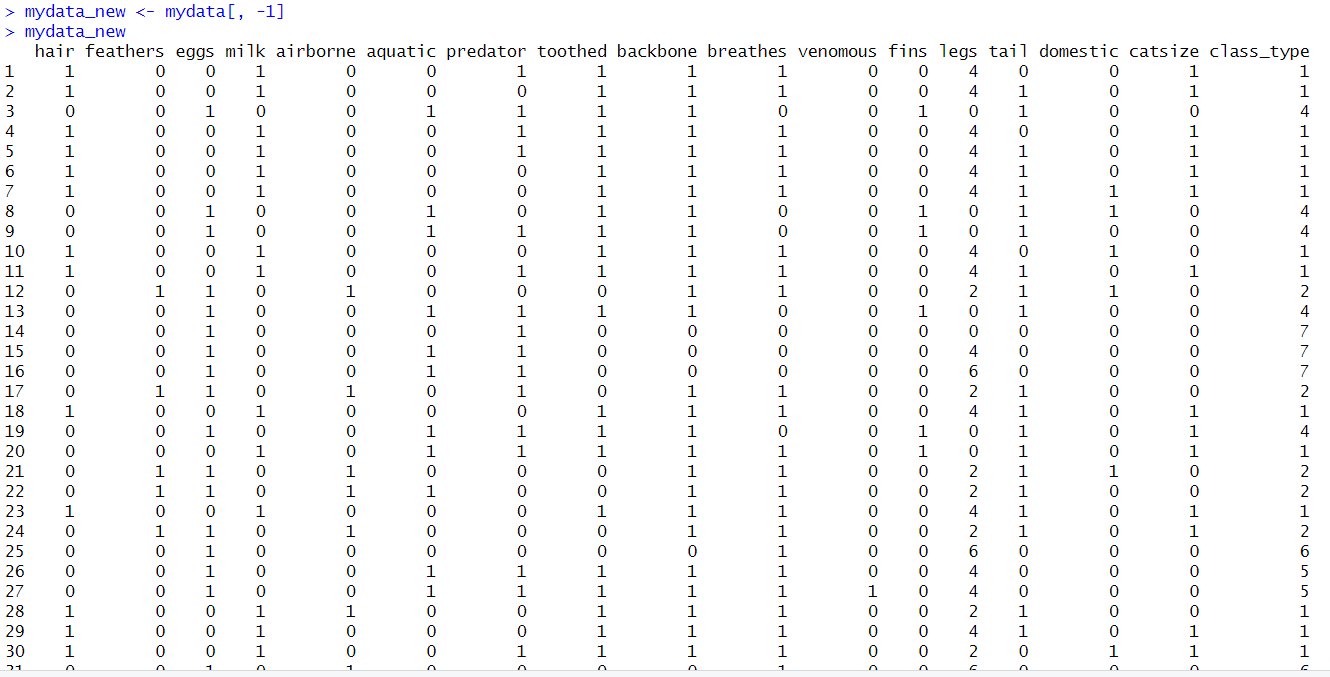
str(mydata)



**Removing names from dataset**

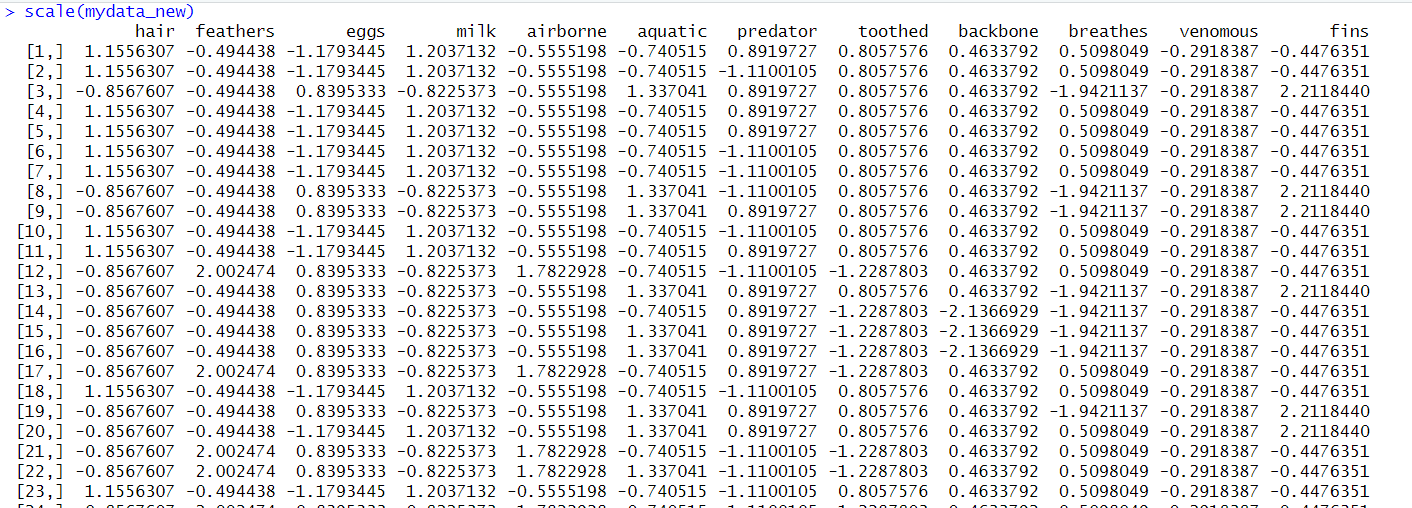
mydata\_new <- mydata[, -1]

mydata\_new



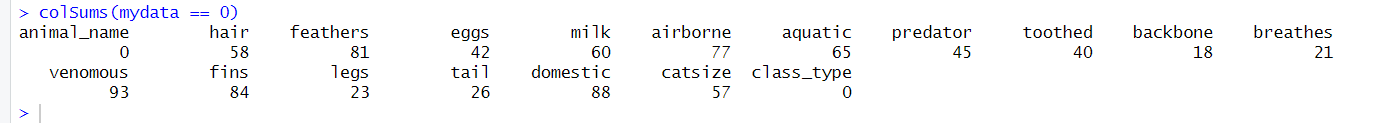
**Scalling Dataset**

scale(mydata\_new)



**Finding zero**

colSums(mydata == 0)

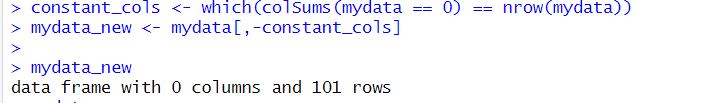


**Remove zero**

constant\_cols <- which(colSums(mydata == 0) == nrow(mydata))

mydata\_new <- mydata[,-constant\_cols]

mydata\_new



**Install Packages**

install.packages("ClusterR")

install.packages("cluster")

install.packages("factoextra")

**Library Packages**

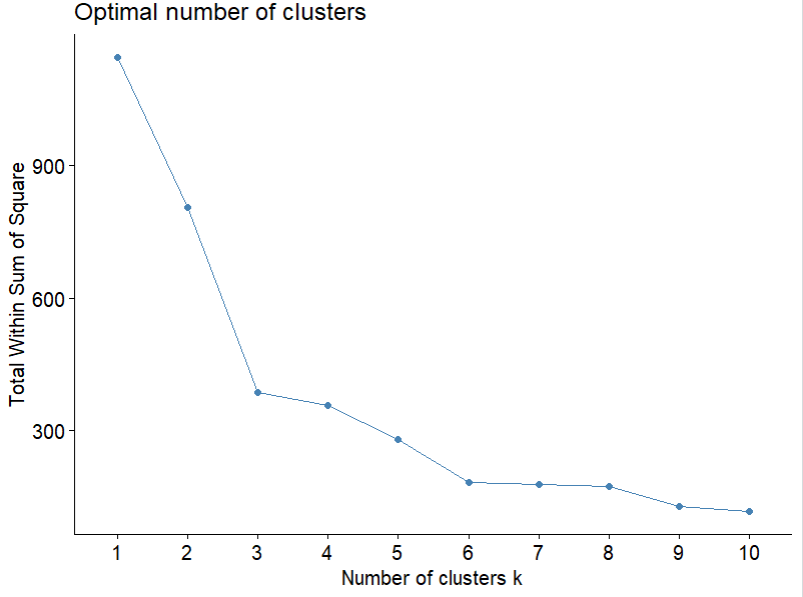
library("ClusterR")

library("cluster")

library("factoextra")

**Optimal number of Clustering**

fviz\_nbclust(mydata\_new, kmeans, method = "wss")

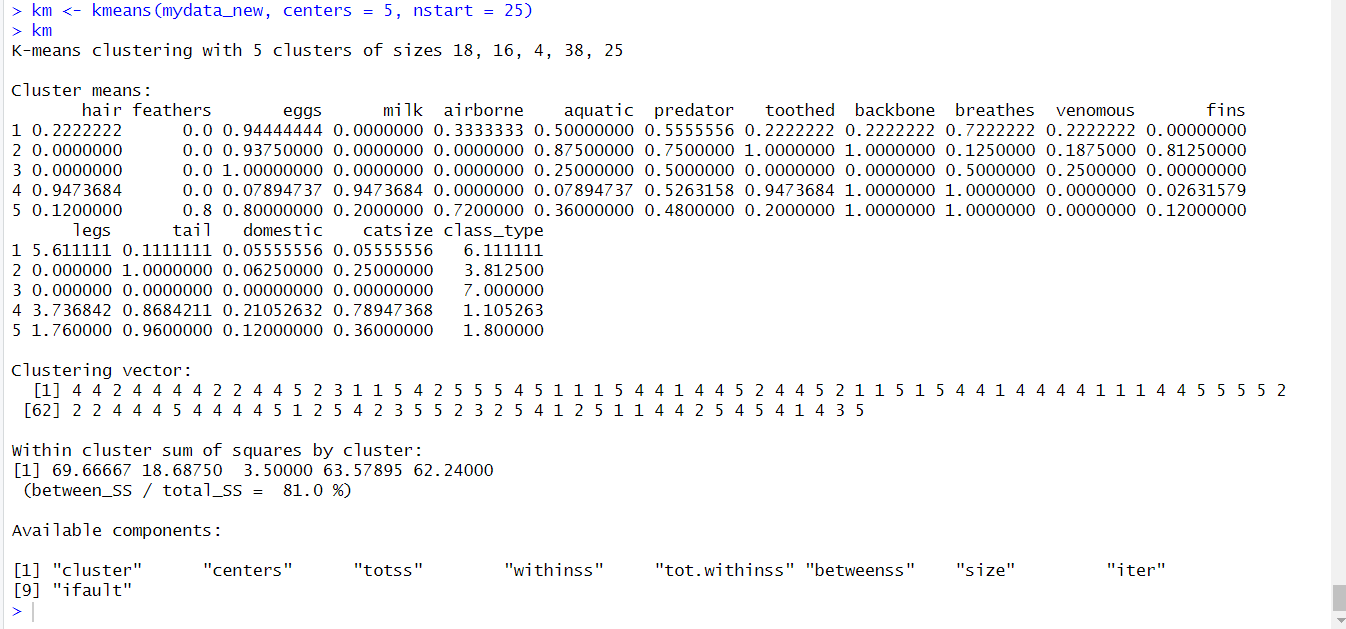


**K-Means Clustering Algorithm Applied**

**K-means Clustering Algorithm**

km <- kmeans(mydata\_new, centers = 5, nstart = 25)

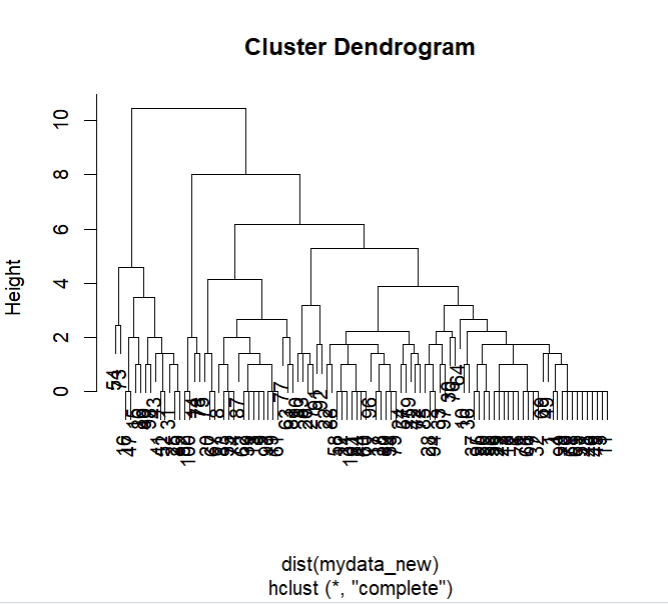
km



**Hierarchical Clustering**

hc <- hclust(dist(mydata\_new))

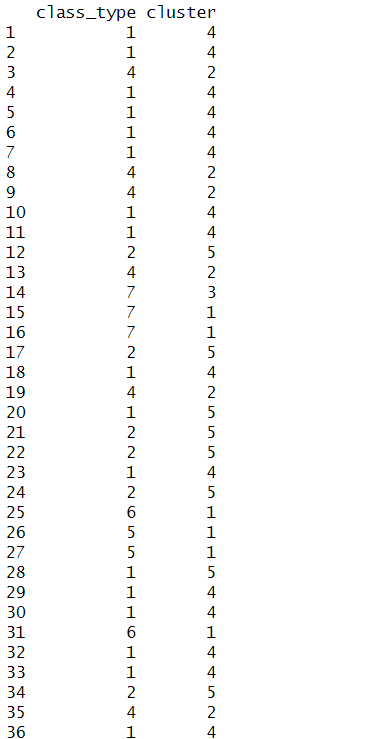
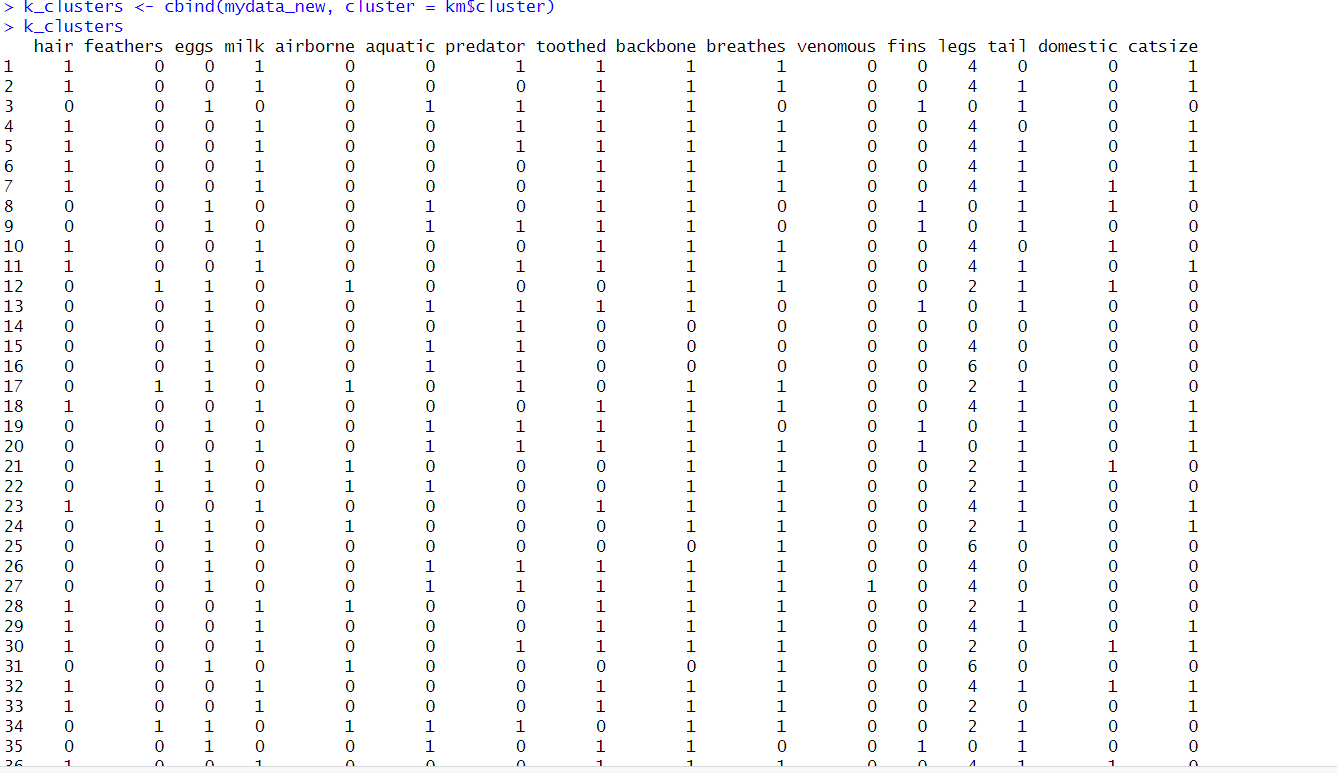
plot(hc)



**Clustering Output Visualization**

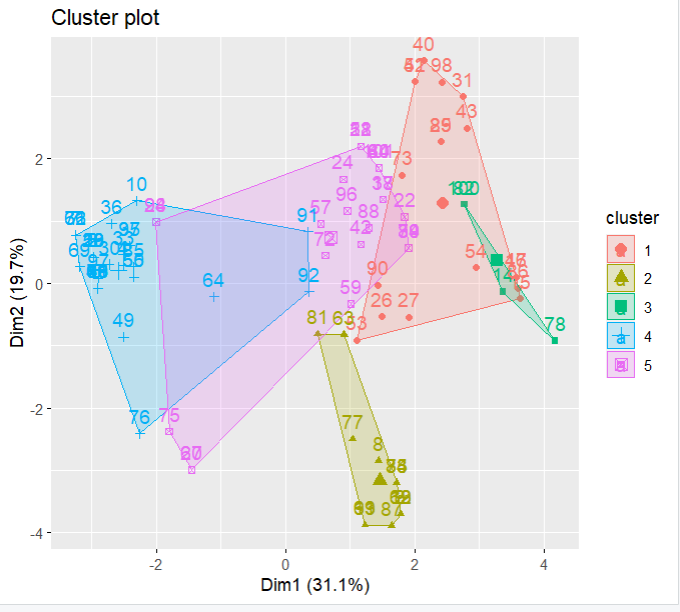
k\_clusters <- cbind(mydata\_new, cluster = km$cluster)

k\_clusters



**Visualize clustering Plot output**

fviz\_cluster(km, data = mydata\_new)



**Find means of each cluster**

aggregate(mydata\_new, by=list(cluster=km$cluster), mean)

