**American International University-Bangladesh**

**Arham Chowdhury Kiass**

**ID: 20-42809-1**

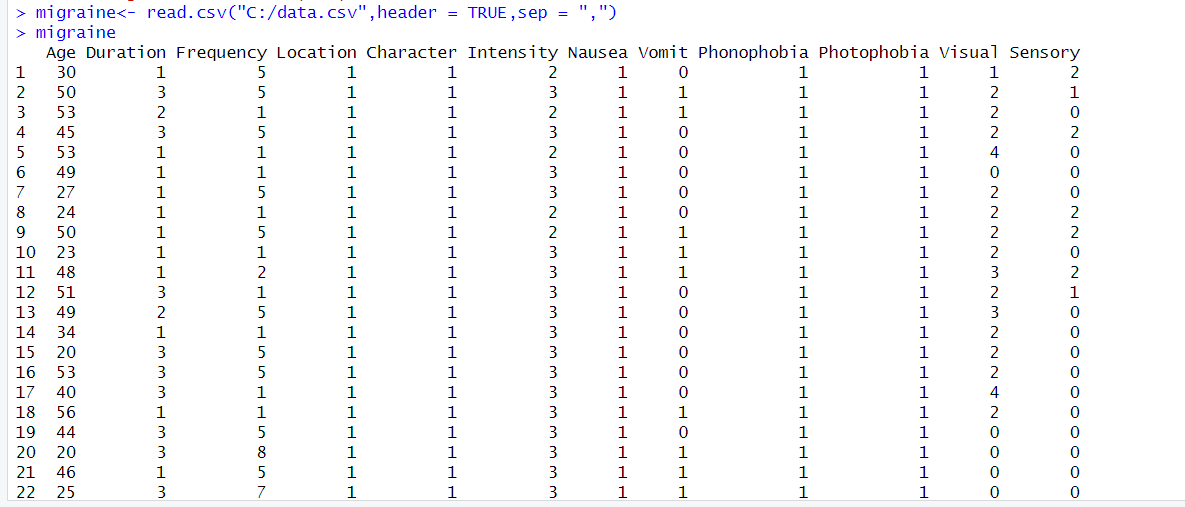
**Data Science(D)**

**Dataset Sources**

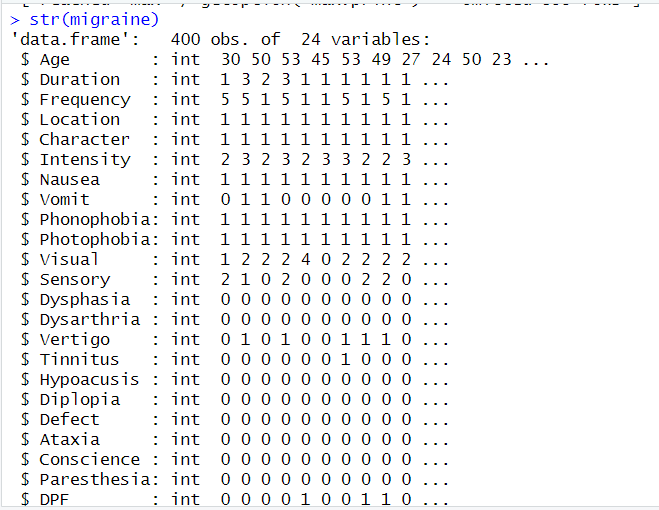
[Salary Prediction dataset | Kaggle](https://www.kaggle.com/datasets/rkiattisak/salaly-prediction-for-beginer)

* **Dataset Input**

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

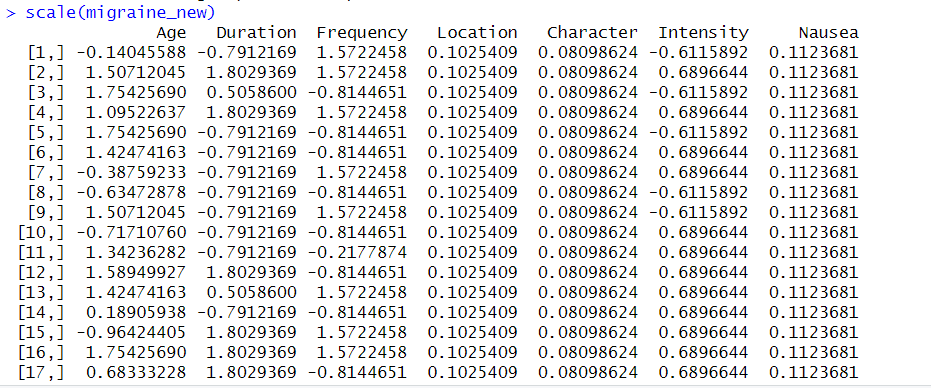
migraine

* **str(**migraine**)**



* **Dataset Scalling**

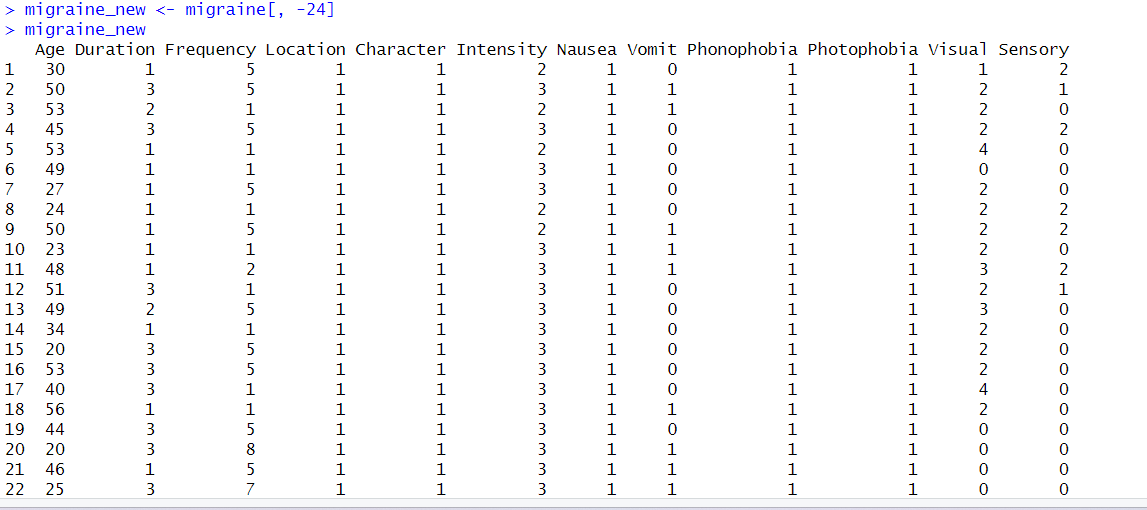
scale(migraine\_new)



* **Remove Type**

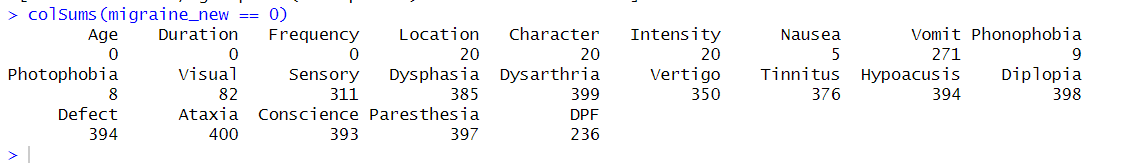
migraine\_new <- migraine[, -24]

migraine\_new



* **Finding zero from numeric dataset**

colSums(migraine\_new == 0)

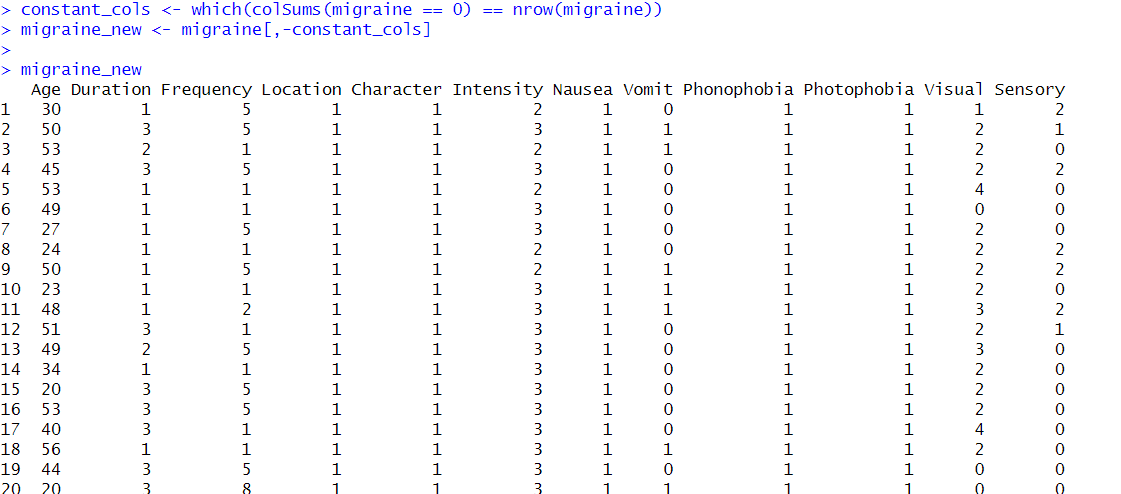


* **Removing Zero**

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

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

migraine\_new



**installing packages**

install.packages("ClusterR")

install.packages("cluster")

install.packages("factoextra")

**Loading packages**

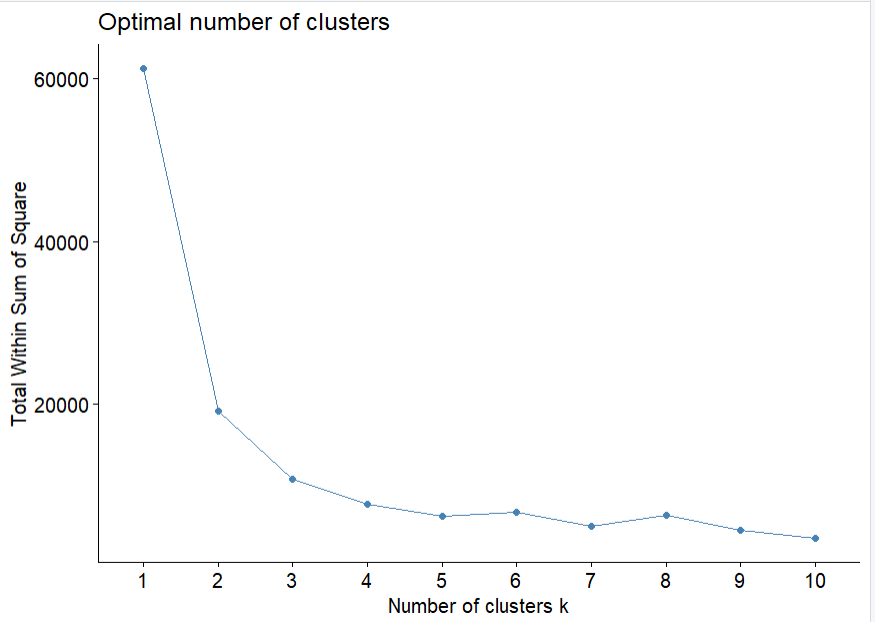
library(ClusterR)

library(cluster)

library(factoextra)

* **Optimal Number of Clusters**

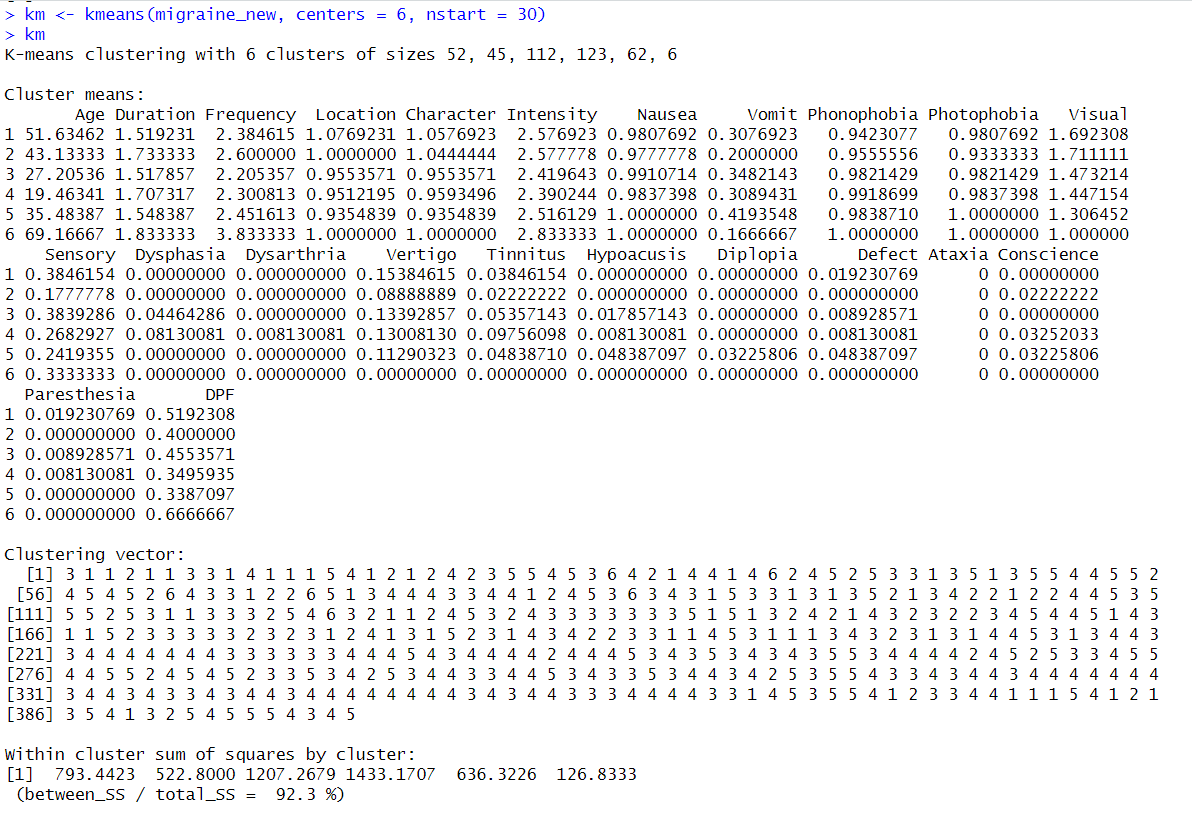
fviz\_nbclust(migraine\_new, kmeans, method = "wss")



* **K-means Clustering Algorithm**

km <- kmeans(migraine\_new, centers = 6, nstart = 30)

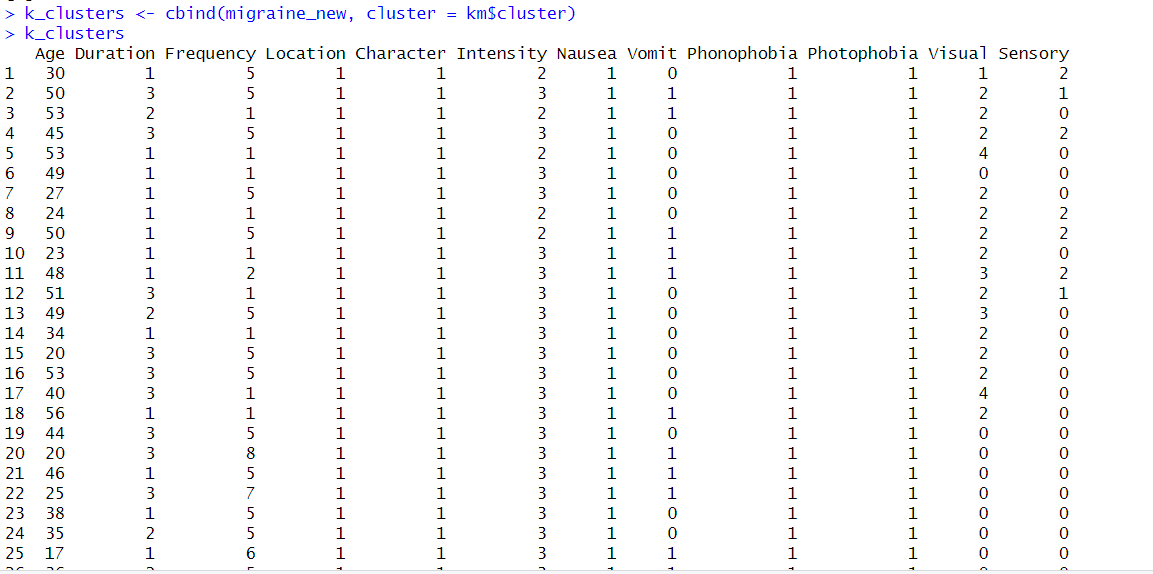
km



* **Visualize the output of K-means Clustering Algorithm**

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

k\_clusters

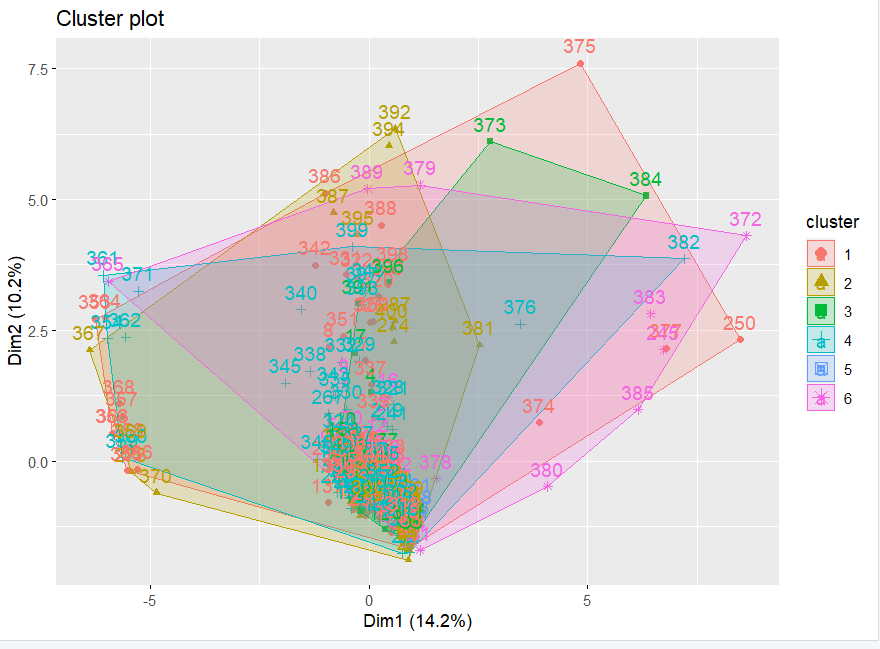


* **Visualize the output of K-means Clustering Algorithm**

library(dplyr)

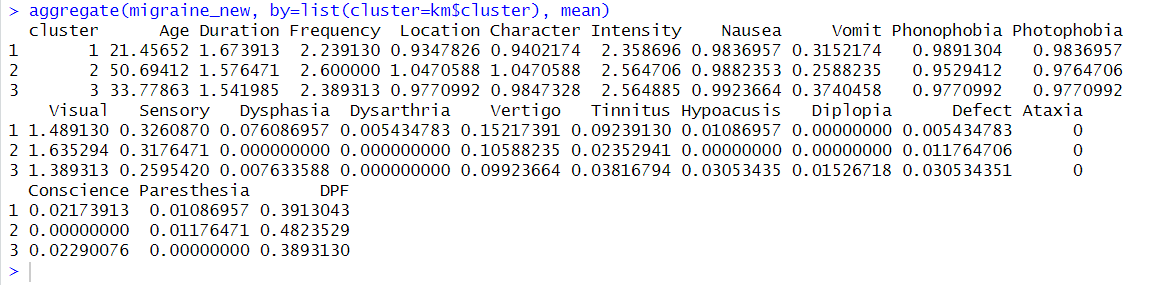
migraine\_new <- select\_if(migraine\_new, function(x) !all(x == 0))

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



* **Find means of each cluster**

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



#hierarcical clustering

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

plot(hc)

