Group_12_Clustering_PPA.R

Souvik

2020-12-08

```
setwd("C:/Users/Souvik/Downloads/PPA")

library(cluster)
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
SECB <- read.csv("PPA SEC B 2020.csv", row.names = 1, stringsAsFactors = TRUE)
summary(SECB)
```

```
##  Age..in.Years.  Body.Weight..in.Kg. Body.Height..in.cm. Drink
##  Min.   :21.00   Min.   :42.00       Min.   :151.0       No :25
##  1st Qu.:23.00   1st Qu.:62.00       1st Qu.:165.0       Yes:33
##  Median :24.00   Median :70.00       Median :172.7
##  Mean   :24.57   Mean   :70.16       Mean   :171.8
##  3rd Qu.:25.00   3rd Qu.:79.75       3rd Qu.:178.0
##  Max.   :35.00   Max.   :95.00       Max.   :189.0
##          Personality.Trait Food.Preference Grade
##  Agreeableness    :17      Non Veg:34      A :31
##  Conscientiousness:16      Veg    :24      A-: 4
##  Extraversion     : 6                      A+:23
##  Neuroticism      : 3
##  Openness         :16
##
```

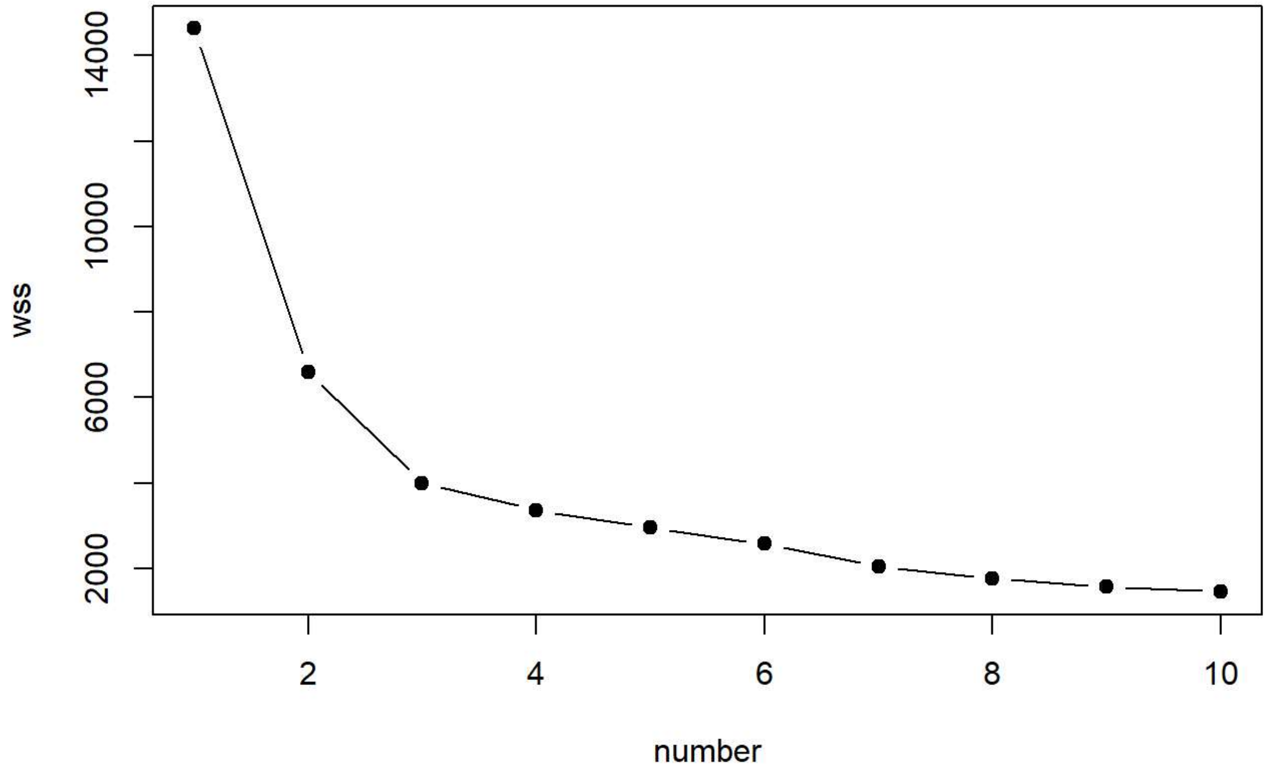
```
table(SECB$Personality.Trait)
```

```
##
##     Agreeableness Conscientiousness      Extraversion       Neuroticism
##                17                16                 6                 3
##          Openness
##                16
```

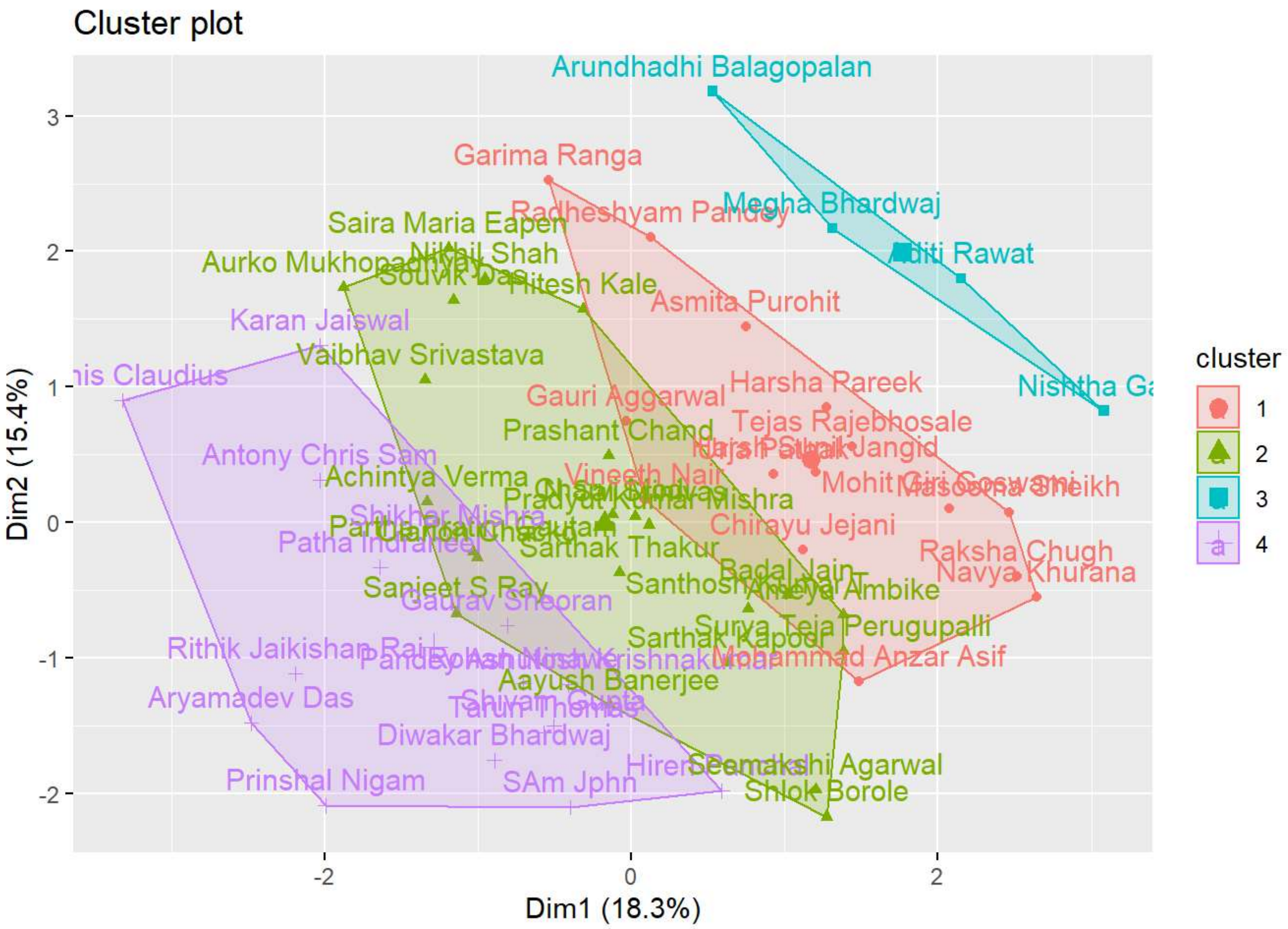
```
### Preparing the dataset for clustering
SECB$Drink<- ifelse(SECB$Drink == "Yes", 1,0)
SECB$Food.Preference <- ifelse(SECB$Food.Preference == "Non Veg", 1,0)
SECB$Personality.Trait_A <- ifelse(SECB$Personality.Trait == "Agreeableness", 1,0)
SECB$Personality.Trait_C <- ifelse(SECB$Personality.Trait == "Conscientiousness", 1,0)
SECB$Personality.Trait_E <- ifelse(SECB$Personality.Trait == "Extraversion", 1,0)
SECB$Personality.Trait_N <- ifelse(SECB$Personality.Trait == "Neuroticism", 1,0)
SECB$Personality.Trait_O <- ifelse(SECB$Personality.Trait == "Openness", 1,0)
SECB$Grade=ifelse(SECB$Grade=='A+',1,(ifelse(SECB$Grade=='A',2,3)))

## Creating a new dataset removing one column ###
SECB_PPA = subset(SECB, select = -c(Personality.Trait))



### k-means clustering ####
##### ELBOW METHOD #####

number <- 1:10
wss <- 1:10

for (i in 1:10)
{
  wss[i] <- kmeans(SECB_PPA,i)$tot.withinss
}
wss
```

```
##  [1] 14634.552  6603.969  3997.685  3355.503  2946.663  2572.128  2051.394
##  [8]  1756.937  1567.129  1449.635
```

```
plot(number,wss,type = "b", pch=19)
```



```
###Taking Optimal number of cluster = 4
#judging from the from the elbow method####
km <- kmeans(SECB_PPA,4)
fviz_cluster(km, data=SECB_PPA)
```



```
str(km)
```

```
## List of 9
##  $ cluster     : Named int [1:58] 4 2 1 2 2 4 4 4 4 1 ...
##   ..- attr(*, "names")= chr [1:58] "Patha Indraneel" "Nikhil Shah" "Mohit Giri Goswami" "C
harvi Modi" ...
##  $ centers     : num [1:4, 1:11] 24.3 24.4 23.5 25.2 58.2 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4] "1" "2" "3" "4"
##   .. ..$ : chr [1:11] "Age..in.Years." "Body.Weight..in.Kg." "Body.Height..in.cm." "Drink"
...
##  $ totss       : num 14635
##  $ withinss    : num [1:4] 967.8 995.4 82.8 1219.5
##  $ tot.withinss: num 3265
##  $ betweenss   : num 11369
##  $ size        : int [1:4] 15 23 4 16
##  $ iter        : int 3
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

```
Accuracy <- km$betweenss/km$totss
Accuracy
```

```
## [1] 0.7768645
## Save Cluster in Original dataset ##

SECB_PPA$cluster <- km$cluster

### Profiling of Clusters ####

cmeans <- aggregate(SECB_PPA, by=list(SECB_PPA$cluster),mean)
cmeans
```

```
##   Group.1 Age..in.Years. Body.Weight..in.Kg. Body.Height..in.cm.     Drink
## 1       1       24.33333            58.20000            165.9592 0.3333333
## 2       2       24.43478            70.30435            173.0974 0.6956522
## 3       3       23.50000            48.50000            153.0000 0.7500000
## 4       4       25.25000            86.56250            179.9375 0.5625000
##   Food.Preference   Grade Personality.Trait_A Personality.Trait_C
## 1       0.4000000 1.80000           0.4666667           0.2000000
## 2       0.6521739 1.73913           0.2608696           0.2608696
## 3       0.5000000 1.50000           0.0000000           0.5000000
## 4       0.6875000 1.50000           0.2500000           0.3125000
##   Personality.Trait_E Personality.Trait_N Personality.Trait_O cluster
## 1          0.06666667          0.00000000           0.2666667       1
## 2          0.08695652          0.08695652           0.3043478       2
## 3          0.00000000          0.00000000           0.5000000       3
## 4          0.18750000          0.06250000           0.1875000       4
```

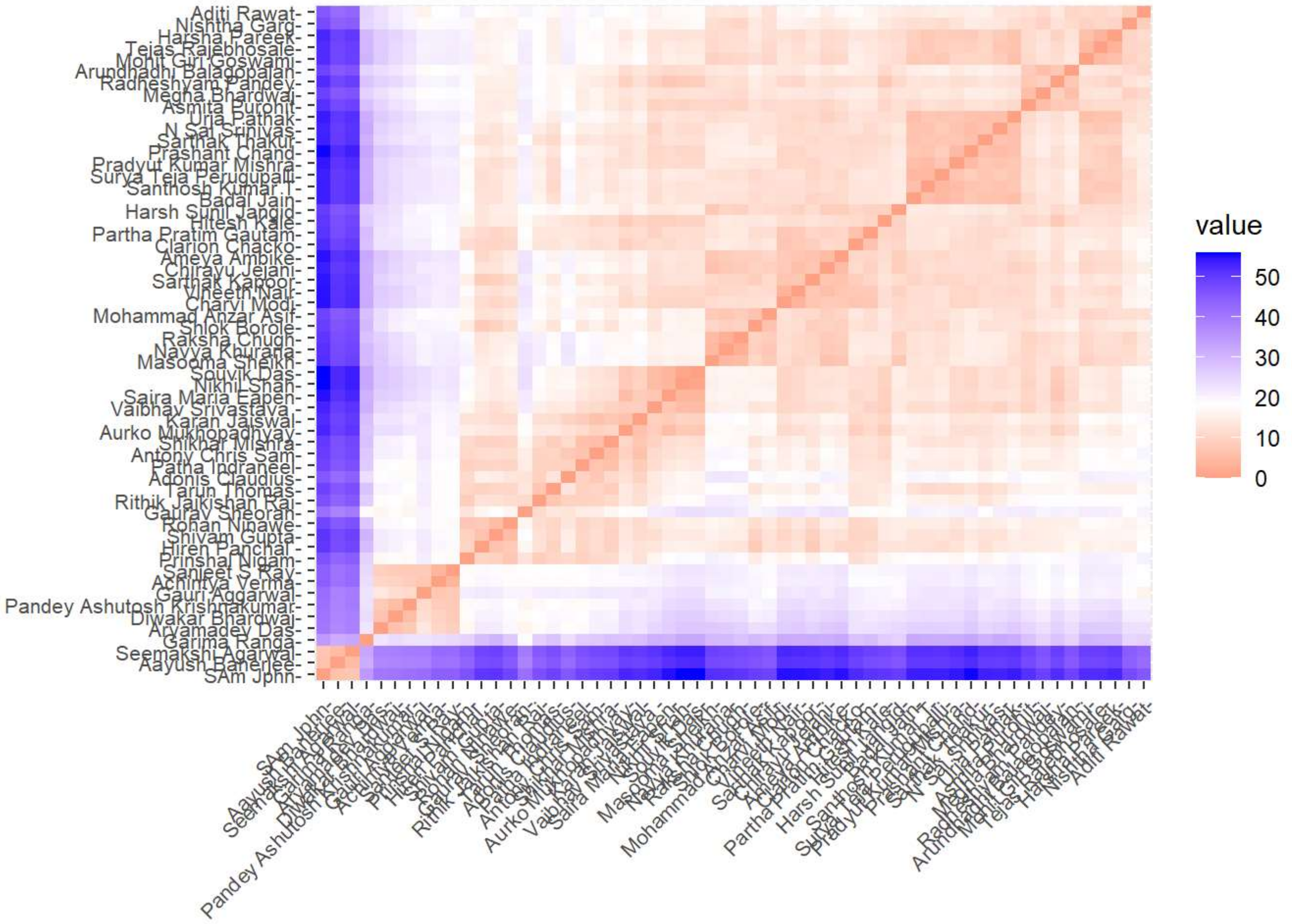
```
#### Hierarchical Clustering ###

dmatrix <- daisy(SECB_PPA, metric = c("euclidean"), stand = TRUE)
```

```
## Warning in daisy(SECB_PPA, metric = c("euclidean"), stand = TRUE): binary
## variable(s) 4, 5, 7, 8, 9, 10, 11 treated as interval scaled
```

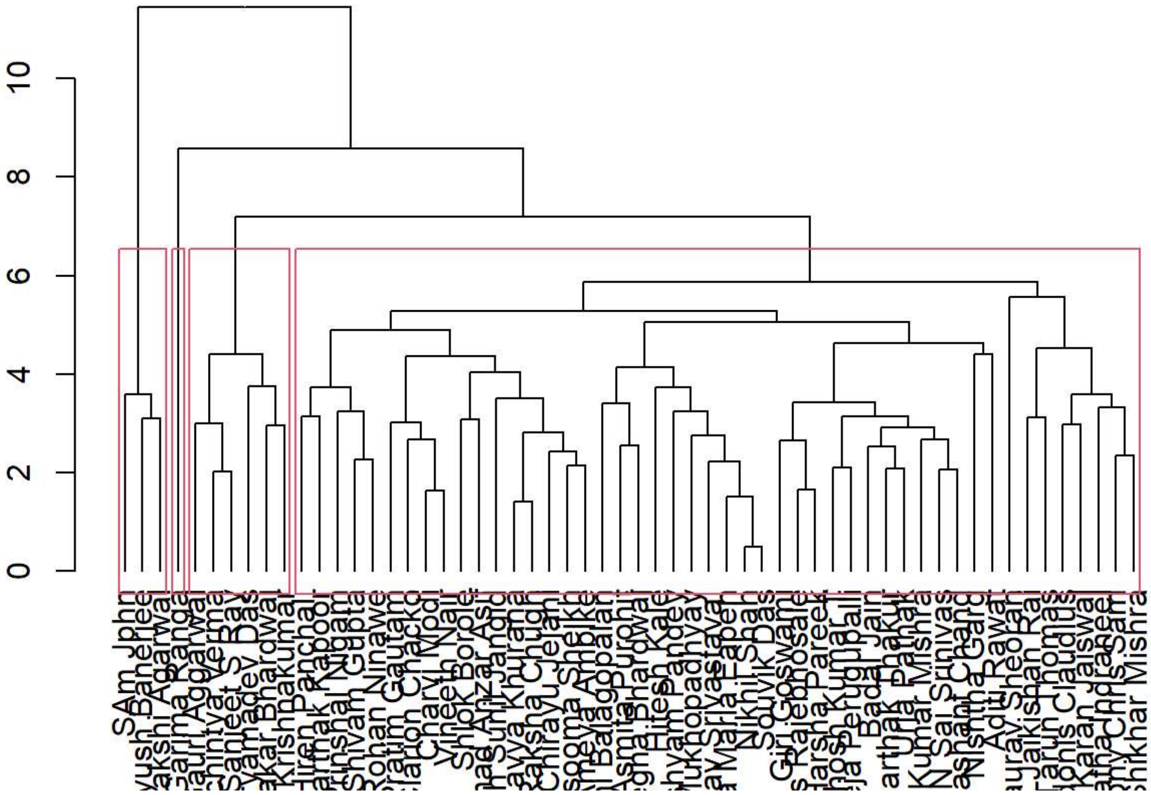
```
class(dmatrix)
```

```
## [1] "dissimilarity" "dist"
```

```
dmatrix1 <- dist(dmatrix)
fviz_dist(dmatrix1, lab_size = 8)
```



```
d <- as.matrix(dmatrix1)
write.csv(d, "D_MATRIX.csv")

hc <- hclust(dmatrix,method = "average")
plot(as.dendrogram(hc))

cluster <- rect.hclust(hc,4)
```


10
8
6
4
2
0
SAm Jphn
Banerjee
Agarwal
Sanjeet S Ray
Mahadev Das
Bhardwaj
Panchal
Kapoor
Nigam
Shivam Gupta
Rohan Ninawe
Gautam
Chacko
Charvi Modi
Vineeth Nair
Shlok Borole
Khurana
Chugh
Chirayu Jelan
Shelke
Ambekar
Purohit
Hitesh Kale
Pandey
Mukhopadhyay
Srivastava
Nikhil Shah
Souvik Das
Giri Goswami
Ghosale
Pareek
Badal Jain
Thakur
Mishra
N Sai Srinivas
Nishtha Garg
Aditi Rawat
Sheoran
Thomas
Claudius
Karan Jaiswal
Shikhar Mishra

```
cluster
```

```
## [[1]]
##   Aayush Banerjee Seemakshi Agarwal          SAm Jphn 
##                24                37                46 
## 
## [[2]]
## Garima Ranga 
##           23 
## 
## [[3]]
##               Achintya Verma                Sanjeet S Ray 
##                            5                           11 
##             Diwakar Bhardwaj Pandey Ashutosh Krishnakumar 
##                           12                           13 
##                Aryamadev Das               Gauri Aggarwal 
##                           18                           27 
## 
## [[4]]
##        Patha Indraneel            Nikhil Shah     Mohit Giri Goswami 
##                      1                      2                      3 
##            Charvi Modi           Shivam Gupta   Rithik Jaikishan Rai 
##                      4                      6                      7 
##       Antony Chris Sam           Rohan Ninawe         Asmita Purohit 
##                      8                      9                     10 
##             Badal Jain         N Sai Srinivas             Souvik Das 
##                     14                     15                     16 
##         Gaurav Sheoran     Aurko Mukhopadhyay           Shlok Borole 
##                     17                     19                     20 
##           Vineeth Nair   Pradyut Kumar Mishra         Clarion Chacko 
##                     21                     22                     25 
##         Chirayu Jejani          Navya Khurana      Tejas Rajebhosale 
##                     26                     28                     29 
##      Radheshyam Pandey         Sarthak Thakur      Saira Maria Eapen 
##                     30                     31                     32 
##         Hiren Panchal         Adonis Claudius   Partha Pratim Gautam 
##                     33                     34                     35 
##          Harsha Pareek         Prashant Chand            Urja Pathak 
##                     36                     38                     39 
##            Hitesh Kale         Shikhar Mishra         Masooma Sheikh 
##                     40                     41                     42 
##     Harsh Sunil Jangid           Nishtha Garg       Santhosh Kumar T 
##                     43                     44                     45 
##           Ameya Ambike    Vaibhav Srivastava             Aditi Rawat 
##                     47                     48                     49 
##    Mohammad Anzar Asif         Sarthak Kapoor Surya Teja Perugupalli 
##                     50                     51                     52 
##           Tarun Thomas           Raksha Chugh          Karan Jaiswal 
##                     53                     54                     55 
## Arundhadhi Balagopalan         Prinshal Nigam         Megha Bhardwaj 
##                     56                     57                     58
```