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

```
> mydata<- read.csv("C:/zoo.csv",header = TRUE,sep = ",")
> mydata
   animal_name hair feathers eggs milk airborne aquatic predator toothed backbone breathes venomous fins legs tail domestic catsize
1     aardvark    1        0    0    1        0       0        1       1        1        1        0    0    4    0        0       1
2     antelope    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
3         bass    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0
4         bear    1        0    0    1        0       0        1       1        1        1        0    0    4    0        0       1
5         boar    1        0    0    1        0       0        1       1        1        1        0    0    4    1        0       1
6      buffalo    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
7         calf    1        0    0    1        0       0        0       1        1        1        0    0    4    1        1       1
8         carp    0        0    1    0        0       1        0       1        1        0        0    1    0    1        1       0
9      catfish    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0
10        cavy    1        0    0    1        0       0        0       1        1        1        0    0    4    0        1       0
11     cheetah    1        0    0    1        0       0        1       1        1        1        0    0    4    1        0       1
12     chicken    0        1    1    0        1       0        0       0        1        1        0    0    2    1        1       0
13        chub    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0
14        clam    0        0    1    0        0       0        1       0        0        0        0    0    0    0        0       0
15        crab    0        0    1    0        0       1        1       0        0        0        0    0    4    0        0       0
16    crayfish    0        0    1    0        0       1        1       0        0        0        0    0    6    0        0       0
17        crow    0        1    1    0        1       0        1       0        1        1        0    0    2    1        0       0
18        deer    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
19     dogfish    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       1
20     dolphin    0        0    0    1        0       1        1       1        1        1        0    1    0    1        0       1
21        dove    0        1    1    0        1       0        0       0        1        1        0    0    2    1        1       0
22        duck    0        1    1    0        1       1        0       0        1        1        0    0    2    1        0       0
23    elephant    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
24    flamingo    0        1    1    0        1       0        0       0        1        1        0    0    2    1        0       1
25        flea    0        0    1    0        0       0        0       0        0        1        0    0    6    0        0       0
26        frog    0        0    1    0        0       1        1       1        1        1        0    0    4    0        0       0
27        frog    0        0    1    0        0       1        1       1        1        1        1    0    4    0        0       0
28    fruitbat    1        0    0    1        1       0        0       1        1        1        0    0    2    1        0       0
29     giraffe    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
30        girl    1        0    0    1        0       0        1       1        1        1        0    0    2    0        1       1
31        gnat    0        0    1    0        1       0        0       0        0        1        0    0    6    0        0       0
32        goat    1        0    0    1        0       0        0       1        1        1        0    0    4    1        1       1
33     gorilla    1        0    0    1        0       0        0       1        1        1        0    0    2    0        0       1
```

Find data for attributes

str(mydata)

```
> str(mydata)
'data.frame':   101 obs. of  18 variables:
 $ animal_name: chr  "aardvark" "antelope" "bass" "bear" ...
 $ hair       : int  1 1 0 1 1 1 1 0 0 1 ...
 $ feathers   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ eggs       : int  0 0 1 0 0 0 0 1 1 0 ...
 $ milk       : int  1 1 0 1 1 1 1 0 0 1 ...
 $ airborne   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ aquatic    : int  0 0 1 0 0 0 0 1 1 0 ...
 $ predator   : int  1 0 1 1 1 0 0 0 1 0 ...
 $ toothed    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ backbone   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ breathes   : int  1 1 0 1 1 1 1 0 0 1 ...
 $ venomous   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ fins       : int  0 0 1 0 0 0 0 1 1 0 ...
 $ legs       : int  4 4 0 4 4 4 4 0 0 4 ...
 $ tail       : int  0 1 1 0 1 1 1 1 1 0 ...
 $ domestic   : int  0 0 0 0 0 0 1 1 0 1 ...
 $ catsize    : int  1 1 0 1 1 1 1 0 0 0 ...
 $ class_type : int  1 1 4 1 1 1 1 4 4 1 ...
```

Removing names from dataset

mydata_new <- mydata[, -1]

mydata_new

```
> mydata_new <- mydata[, -1]
> mydata_new
   hair feathers eggs milk airborne aquatic predator toothed backbone breathes venomous fins legs tail domestic catsize class_type
1     1        0    0    1        0       0        1       1        1        1        0    0    4    0        0       1          1
2     1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1          1
3     0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0          4
4     1        0    0    1        0       0        1       1        1        1        0    0    4    0        0       1          1
5     1        0    0    1        0       0        1       1        1        1        0    0    4    1        0       1          1
6     1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1          1
7     1        0    0    1        0       0        0       1        1        1        0    0    4    1        1       1          1
8     0        0    1    0        0       1        0       1        1        0        0    1    0    1        1       0          4
9     0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0          4
10    1        0    0    1        0       0        0       1        1        1        0    0    4    0        1       0          1
11    1        0    0    1        0       0        1       1        1        1        0    0    4    1        0       1          1
12    0        1    1    0        1       0        0       0        1        1        0    0    2    1        1       0          2
13    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0          4
14    0        0    1    0        0       0        1       0        0        0        0    0    0    0        0       0          7
15    0        0    1    0        0       1        1       0        0        0        0    0    4    0        0       0          7
16    0        0    1    0        0       1        1       0        0        0        0    0    6    0        0       0          7
17    0        1    1    0        1       0        1       0        1        1        0    0    2    1        0       0          2
18    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1          1
19    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       1          4
20    0        0    0    1        0       1        1       1        1        1        0    1    0    1        0       1          1
21    0        1    1    0        1       0        0       0        1        1        0    0    2    1        1       0          2
22    0        1    1    0        1       1        0       0        1        1        0    0    2    1        0       0          2
23    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1          1
24    0        1    1    0        1       0        0       0        1        1        0    0    2    1        0       1          2
25    0        0    1    0        0       0        0       0        0        1        0    0    6    0        0       0          6
26    0        0    1    0        0       1        1       1        1        1        0    0    4    0        0       0          5
27    0        0    1    0        0       1        1       1        1        1        1    0    4    0        0       0          5
28    1        0    0    1        1       0        0       1        1        1        0    0    2    1        0       0          1
29    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1          1
30    1        0    0    1        0       0        1       1        1        1        0    0    2    0        1       1          1
```

Scalling Dataset

scale(mydata_new)

```
> scale(mydata_new)
           hair  feathers       eggs       milk   airborne    aquatic    predator    toothed   backbone   breathes    venomous       fins
 [1,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515  0.8919727  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
 [2,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515 -1.1100105  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
 [3,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041  0.8919727  0.8057576  0.4633792 -1.9421137 -0.2918387  2.2118440
 [4,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515  0.8919727  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
 [5,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515  0.8919727  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
 [6,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515 -1.1100105  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
 [7,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515 -1.1100105  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
 [8,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041 -1.1100105  0.8057576  0.4633792 -1.9421137 -0.2918387  2.2118440
 [9,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041  0.8919727  0.8057576  0.4633792 -1.9421137 -0.2918387  2.2118440
[10,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515 -1.1100105  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
[11,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515  0.8919727  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
[12,] -0.8567607  2.002474  0.8395333 -0.8225373  1.7822928 -0.740515 -1.1100105 -1.2287803  0.4633792  0.5098049 -0.2918387 -0.4476351
[13,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041  0.8919727  0.8057576  0.4633792 -1.9421137 -0.2918387  2.2118440
[14,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198 -0.740515  0.8919727 -1.2287803 -2.1366929 -1.9421137 -0.2918387 -0.4476351
[15,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041  0.8919727 -1.2287803 -2.1366929 -1.9421137 -0.2918387 -0.4476351
[16,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041  0.8919727 -1.2287803 -2.1366929 -1.9421137 -0.2918387 -0.4476351
[17,] -0.8567607  2.002474  0.8395333 -0.8225373  1.7822928 -0.740515  0.8919727 -1.2287803  0.4633792  0.5098049 -0.2918387 -0.4476351
[18,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515 -1.1100105  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
[19,] -0.8567607 -0.494438  0.8395333 -0.8225373 -0.5555198  1.337041  0.8919727  0.8057576  0.4633792 -1.9421137 -0.2918387  2.2118440
[20,] -0.8567607 -0.494438 -1.1793445  1.2037132 -0.5555198  1.337041  0.8919727  0.8057576  0.4633792  0.5098049 -0.2918387  2.2118440
[21,] -0.8567607  2.002474  0.8395333 -0.8225373  1.7822928 -0.740515 -1.1100105 -1.2287803  0.4633792  0.5098049 -0.2918387 -0.4476351
[22,] -0.8567607  2.002474  0.8395333 -0.8225373  1.7822928  1.337041 -1.1100105 -1.2287803  0.4633792  0.5098049 -0.2918387 -0.4476351
[23,]  1.1556307 -0.494438 -1.1793445  1.2037132 -0.5555198 -0.740515 -1.1100105  0.8057576  0.4633792  0.5098049 -0.2918387 -0.4476351
```

Finding zero

colSums(mydata == 0)

```
> colSums(mydata == 0)
animal_name        hair    feathers        eggs        milk    airborne     aquatic    predator     toothed    backbone    breathes
          0          58          81          42          60          77          65          45          40          18          21
   venomous        fins        legs        tail    domestic     catsize  class_type
         93          84          23          26          88          57           0
>
```

Remove zero

constant_cols <- which(colSums(mydata == 0) == nrow(mydata))

mydata_new <- mydata[,-constant_cols]

mydata_new

```
> constant_cols <- which(colSums(mydata == 0) == nrow(mydata))
> mydata_new <- mydata[,-constant_cols]
>
> mydata_new
data frame with 0 columns and 101 rows
```

Install Packages

install.packages("ClusterR")

install.packages("cluster")

install.packages("factoextra")

Library Packages

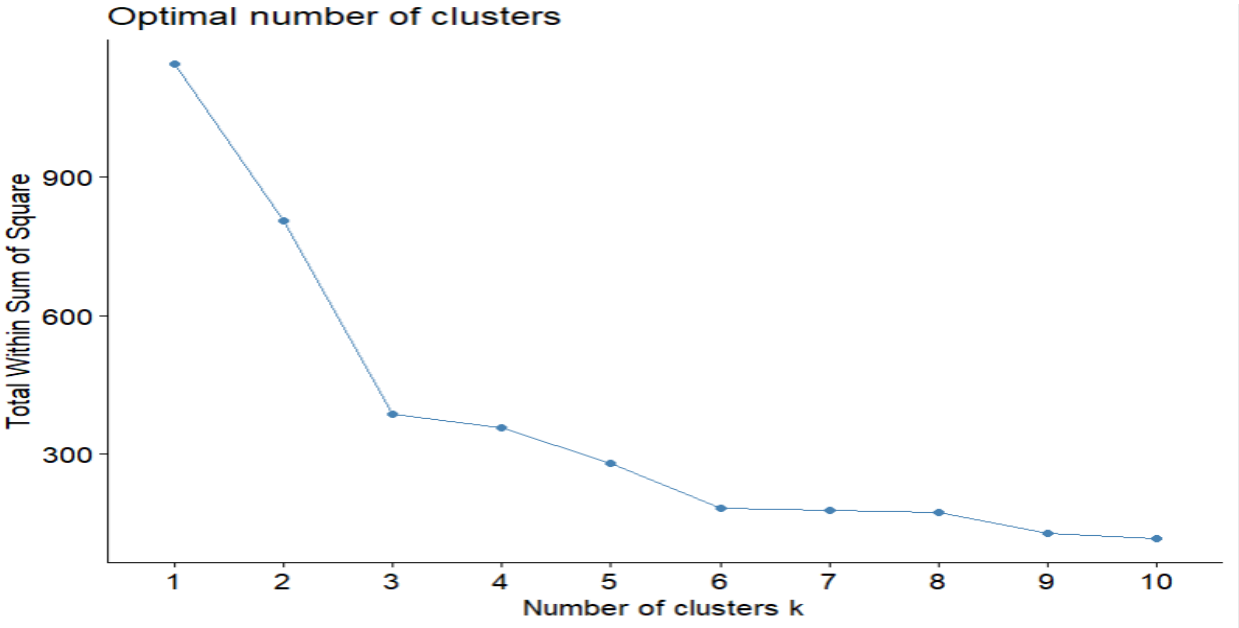
library("ClusterR")

library("cluster")

library("factoextra")

Optimal number of Clustering

fviz_nbclust(mydata_new, kmeans, method = "wss")



K-Means Clustering Algorithm Applied

K-means Clustering Algorithm

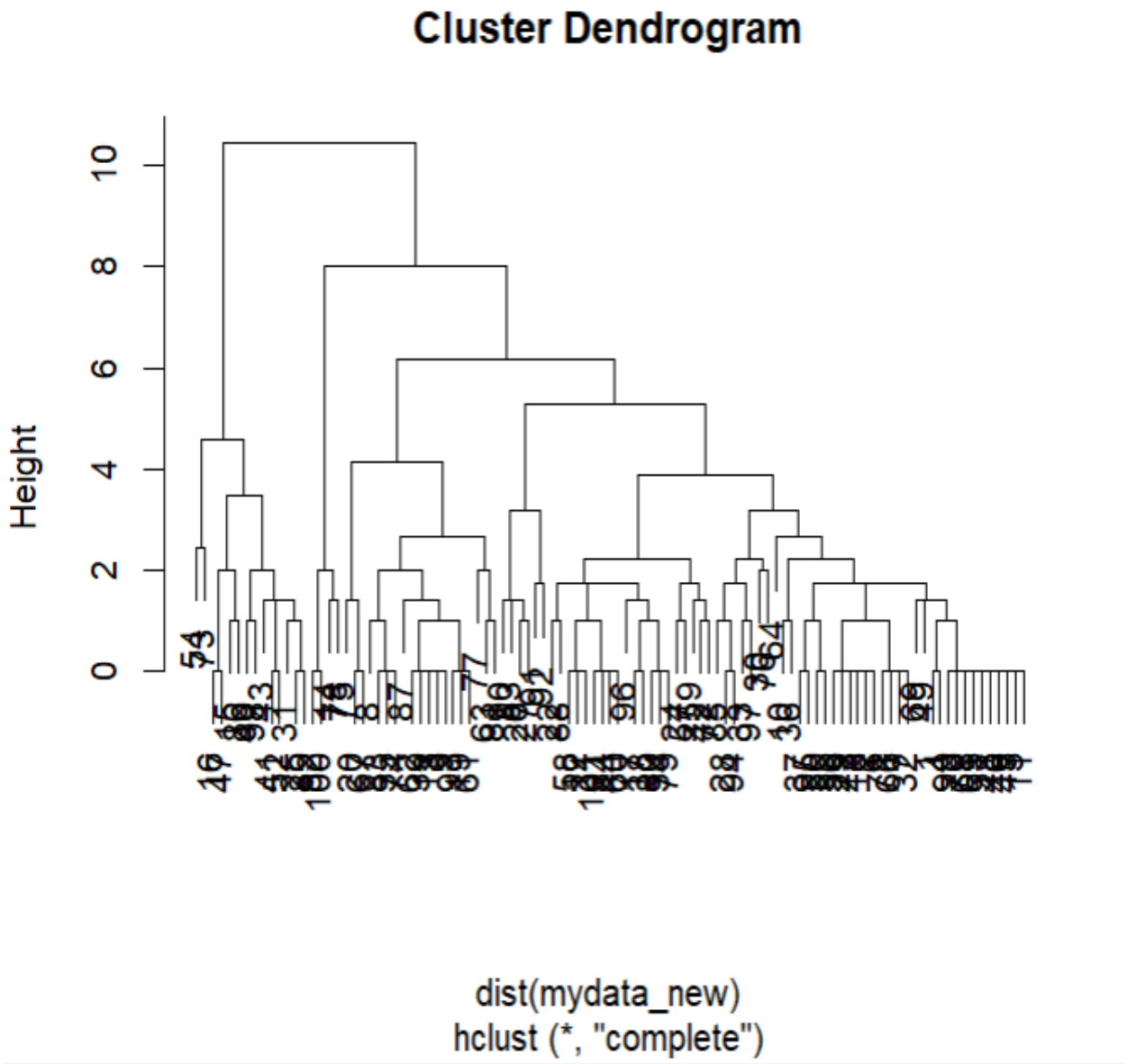
km <- kmeans(mydata_new, centers = 5, nstart = 25)

km

```
> km <- kmeans(mydata_new, centers = 5, nstart = 25)
> km
K-means clustering with 5 clusters of sizes 18, 16, 4, 38, 25

Cluster means:
       hair feathers       eggs      milk  airborne    aquatic  predator   toothed  backbone  breathes  venomous       fins
1 0.2222222      0.0 0.94444444 0.0000000 0.3333333 0.50000000 0.5555556 0.2222222 0.2222222 0.7222222 0.2222222 0.00000000
2 0.0000000      0.0 0.93750000 0.0000000 0.0000000 0.87500000 0.7500000 1.0000000 1.0000000 0.1250000 0.1875000 0.81250000
3 0.0000000      0.0 1.00000000 0.0000000 0.0000000 0.25000000 0.5000000 0.0000000 0.0000000 0.5000000 0.2500000 0.00000000
4 0.9473684      0.0 0.07894737 0.9473684 0.0000000 0.07894737 0.5263158 0.9473684 1.0000000 1.0000000 0.0000000 0.02631579
5 0.1200000      0.8 0.80000000 0.2000000 0.7200000 0.36000000 0.4800000 0.2000000 1.0000000 1.0000000 0.0000000 0.12000000
      legs      tail   domestic    catsize class_type
1 5.611111 0.1111111 0.05555556 0.05555556   6.111111
2 0.000000 1.0000000 0.06250000 0.25000000   3.812500
3 0.000000 0.0000000 0.00000000 0.00000000   7.000000
4 3.736842 0.8684211 0.21052632 0.78947368   1.105263
5 1.760000 0.9600000 0.12000000 0.36000000   1.800000

Clustering vector:
  [1] 4 4 2 4 4 4 4 2 2 4 4 5 2 3 1 1 5 4 2 5 5 5 4 5 1 1 1 5 4 4 1 4 4 5 2 4 4 5 2 1 1 5 1 5 4 4 1 4 4 4 4 1 1 1 4 4 5 5 5 5 2
 [62] 2 2 4 4 4 5 4 4 4 4 5 1 2 5 4 2 3 5 5 2 3 2 5 4 1 2 5 1 1 4 4 2 5 4 5 4 1 4 3 5

Within cluster sum of squares by cluster:
[1] 69.66667 18.68750  3.50000 63.57895 62.24000
 (between_SS / total_SS =  81.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"         "iter"
[9] "ifault"
>
```

Hierarchical Clustering

```
hc <- hclust(dist(mydata_new))
plot(hc)
```



Clustering Output Visualization

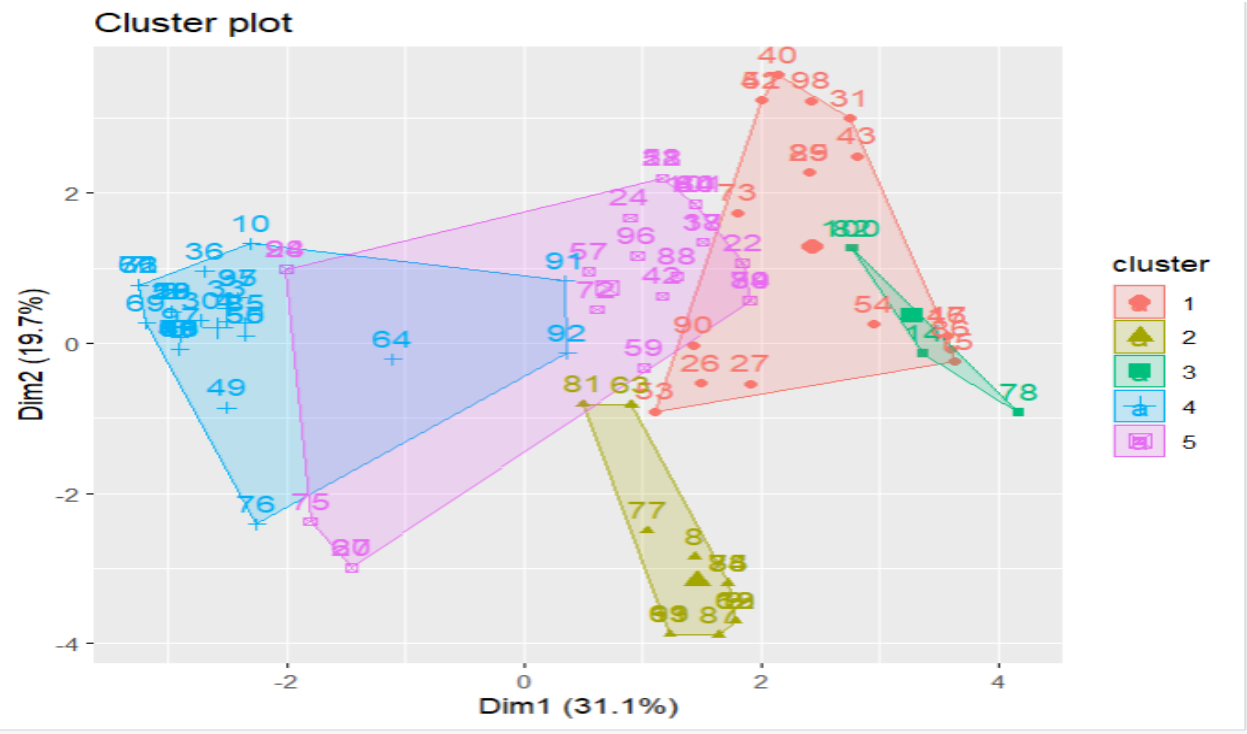
k_clusters <- cbind(mydata_new, cluster = km$cluster)

k_clusters

```
> k_clusters <- cbind(mydata_new, cluster = km$cluster)
> k_clusters
   hair feathers eggs milk airborne aquatic predator toothed backbone breathes venomous fins legs tail domestic catsize
1     1        0    0    1        0       0        1       1        1        1        0    0    4    0        0       1
2     1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
3     0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0
4     1        0    0    1        0       0        1       1        1        1        0    0    4    0        0       1
5     1        0    0    1        0       0        1       1        1        1        0    0    4    1        0       1
6     1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
7     1        0    0    1        0       0        0       1        1        1        0    0    4    1        1       1
8     0        0    1    0        0       1        0       1        1        0        0    1    0    1        1       0
9     0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0
10    1        0    0    1        0       0        0       1        1        1        0    0    4    0        1       0
11    1        0    0    1        0       0        1       1        1        1        0    0    4    1        0       1
12    0        1    1    0        1       0        0       0        1        1        0    0    2    1        1       0
13    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       0
14    0        0    1    0        0       0        1       0        0        0        0    0    0    0        0       0
15    0        0    1    0        0       1        1       0        0        0        0    0    4    0        0       0
16    0        0    1    0        0       1        1       0        0        0        0    0    6    0        0       0
17    0        1    1    0        1       0        1       0        1        1        0    0    2    1        0       0
18    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
19    0        0    1    0        0       1        1       1        1        0        0    1    0    1        0       1
20    0        0    0    1        0       1        1       1        1        1        0    1    0    1        0       1
21    0        1    1    0        1       0        0       0        1        1        0    0    2    1        1       0
22    0        1    1    0        1       1        0       0        1        1        0    0    2    1        0       0
23    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
24    0        1    1    0        1       0        0       0        1        1        0    0    2    1        0       1
25    0        0    1    0        0       0        0       0        0        1        0    0    6    0        0       0
26    0        0    1    0        0       1        1       1        1        1        0    0    4    0        0       0
27    0        0    1    0        0       1        1       1        1        1        1    0    4    0        0       0
28    1        0    0    1        1       0        0       1        1        1        0    0    2    1        0       0
29    1        0    0    1        0       0        0       1        1        1        0    0    4    1        0       1
30    1        0    0    1        0       0        1       1        1        1        0    0    2    0        1       1
31    0        0    1    0        1       0        0       0        0        1        0    0    6    0        0       0
32    1        0    0    1        0       0        0       1        1        1        0    0    4    1        1       1
33    1        0    0    1        0       0        0       1        1        1        0    0    2    0        0       1
34    0        1    1    0        1       1        1       0        1        1        0    0    2    1        0       0
35    0        0    1    0        0       1        0       1        1        0        0    1    0    1        0       0
36    1        0    0    1        0       0        0       1        1        1        0    0    4    1        1       0
   class_type cluster
1           1       4
2           1       4
3           4       2
4           1       4
5           1       4
6           1       4
7           1       4
8           4       2
9           4       2
10          1       4
11          1       4
12          2       5
13          4       2
14          7       3
15          7       1
16          7       1
17          2       5
18          1       4
19          4       2
20          1       5
21          2       5
22          2       5
23          1       4
24          2       5
25          6       1
26          5       1
27          5       1
28          1       5
29          1       4
30          1       4
31          6       1
32          1       4
33          1       4
34          2       5
35          4       2
36          1       4
```

Visualize clustering Plot output

fviz_cluster(km, data = mydata_new)



Find means of each cluster

aggregate(mydata_new, by=list(cluster=km$cluster), mean)

```
> aggregate(mydata_new, by=list(cluster=km$cluster), mean)
  cluster      hair feathers       eggs      milk  airborne    aquatic  predator   toothed  backbone  breathes
1       1 0.2222222       0.0 0.94444444 0.0000000 0.3333333 0.50000000 0.5555556 0.2222222 0.2222222 0.7222222
2       2 0.0000000       0.0 0.93750000 0.0000000 0.0000000 0.87500000 0.7500000 1.0000000 1.0000000 0.1250000
3       3 0.0000000       0.0 1.00000000 0.0000000 0.0000000 0.25000000 0.5000000 0.0000000 0.0000000 0.5000000
4       4 0.9473684       0.0 0.07894737 0.9473684 0.0000000 0.07894737 0.5263158 0.9473684 1.0000000 1.0000000
5       5 0.1200000       0.8 0.80000000 0.2000000 0.7200000 0.36000000 0.4800000 0.2000000 1.0000000 1.0000000
   venomous       fins     legs      tail   domestic     catsize class_type
1 0.2222222 0.00000000 5.611111 0.1111111 0.05555556 0.05555556   6.111111
2 0.1875000 0.81250000 0.000000 1.0000000 0.06250000 0.25000000   3.812500
3 0.2500000 0.00000000 0.000000 0.0000000 0.00000000 0.00000000   7.000000
4 0.0000000 0.02631579 3.736842 0.8684211 0.21052632 0.78947368   1.105263
5 0.0000000 0.12000000 1.760000 0.9600000 0.12000000 0.36000000   1.800000
>
```