R Statistics Training

Sin

2022-09-26

Chapter 8

Required Package(s)

- car

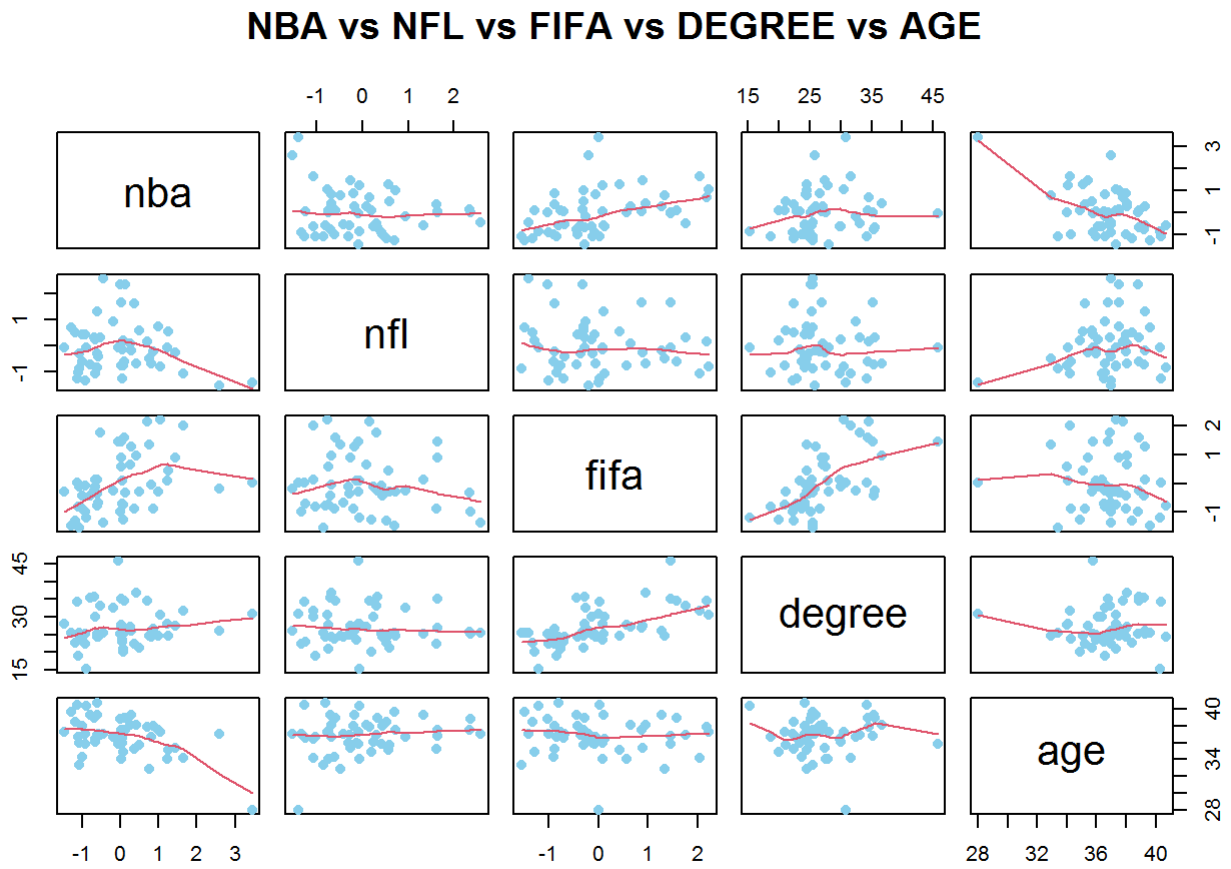
Importing the required dataset

```
search <- read.csv("C:/Users/Sin/Documents/SearchData.csv", header = T)
data <- search[,c(2,3,4,8,9)]
head(data)
```

```
##     nba   nfl  fifa degree  age
## 1 -0.90 -1.34 -0.99   22.3 37.0
## 2 -1.08 -0.86 -1.54   25.5 33.4
## 3  1.23 -0.06  0.09   28.0 34.1
## 4 -1.10 -1.04 -0.83   18.8 36.6
## 5  1.65 -1.08  2.03   31.7 34.2
## 6 -0.68  0.32 -0.27   35.5 38.9
```

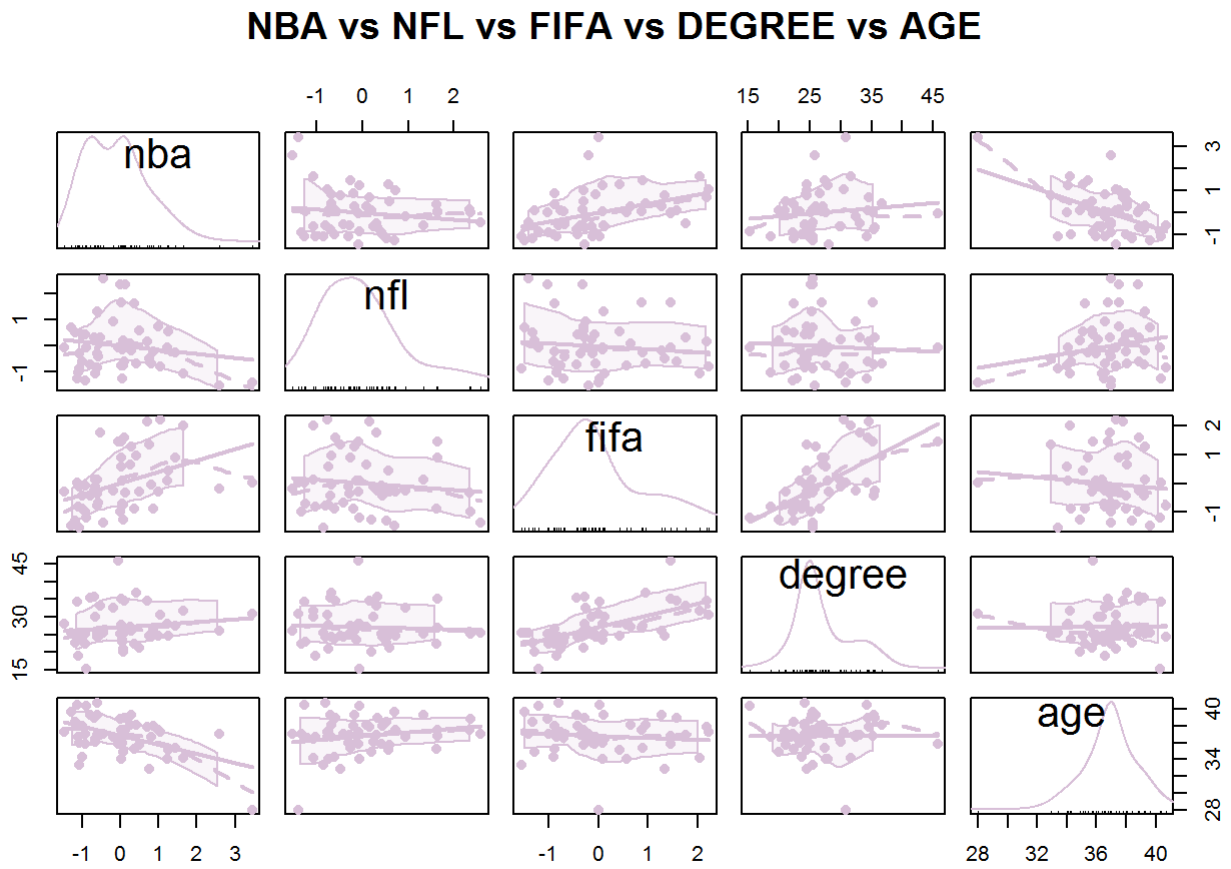
Scatterplot Matrix using base R

```
pairs(data,
      main = "NBA vs NFL vs FIFA vs DEGREE vs AGE",
      pch = 16,
      panel = panel.smooth,
      col = "skyblue")
```



Scatterplot matrix using *car* package

```
scatterplotMatrix(data,
                  main = "NBA vs NFL vs FIFA vs DEGREE vs AGE",
                  pch = 16,
                  col = "thistle"
                  )
```



Chapter 9

Importing the required dataset

```
state <- read.csv("C:/Users/Sin/Documents/StateClusterData.csv", header = T)
rownames(state) <- state$state
state <- state[-1]
head(state)
```

```
##            modern.dance   xbox ice.fishing runny.nose  prius escalade college
## Alabama          -1.186  0.405      -0.621      0.538 -0.749    0.018   0.913
## Alaska           -0.695  0.172       0.491     -0.638 -0.619    0.140  -2.178
## Arizona           0.166  0.122      -0.585      0.688  1.223    0.847  -0.031
## Arkansas         -1.280  1.268      -0.584      0.968 -0.324    1.245  -0.696
## California        0.111 -1.414      -0.614      0.030  3.492   -0.401  -0.514
## Colorado         -0.422 -0.832      -0.033     -0.237  0.285   -0.133  -0.560
##            retirement  X401k deep.fried  jello  vegan
## Alabama         1.589  0.231     -0.754 -0.842 -1.362
## Alaska          2.650  0.008      3.049  0.107  0.504
## Arizona         0.761  0.446     -0.284 -0.617  0.505
## Arkansas        0.075 -0.849     -0.171 -0.250 -0.934
## California     -2.139 -0.176     -0.626 -1.334  1.024
## Colorado       -0.449  0.633     -0.308 -0.601  0.707
```

State Cluster Analysis

```
diss <- dist(state)
state.clust <- hclust(diss)
state.clust
```

```
##
## Call:
## hclust(d = diss)
##
## Cluster method   : complete
## Distance         : euclidean
## Number of objects: 51
```

```
summary(state.clust)
```

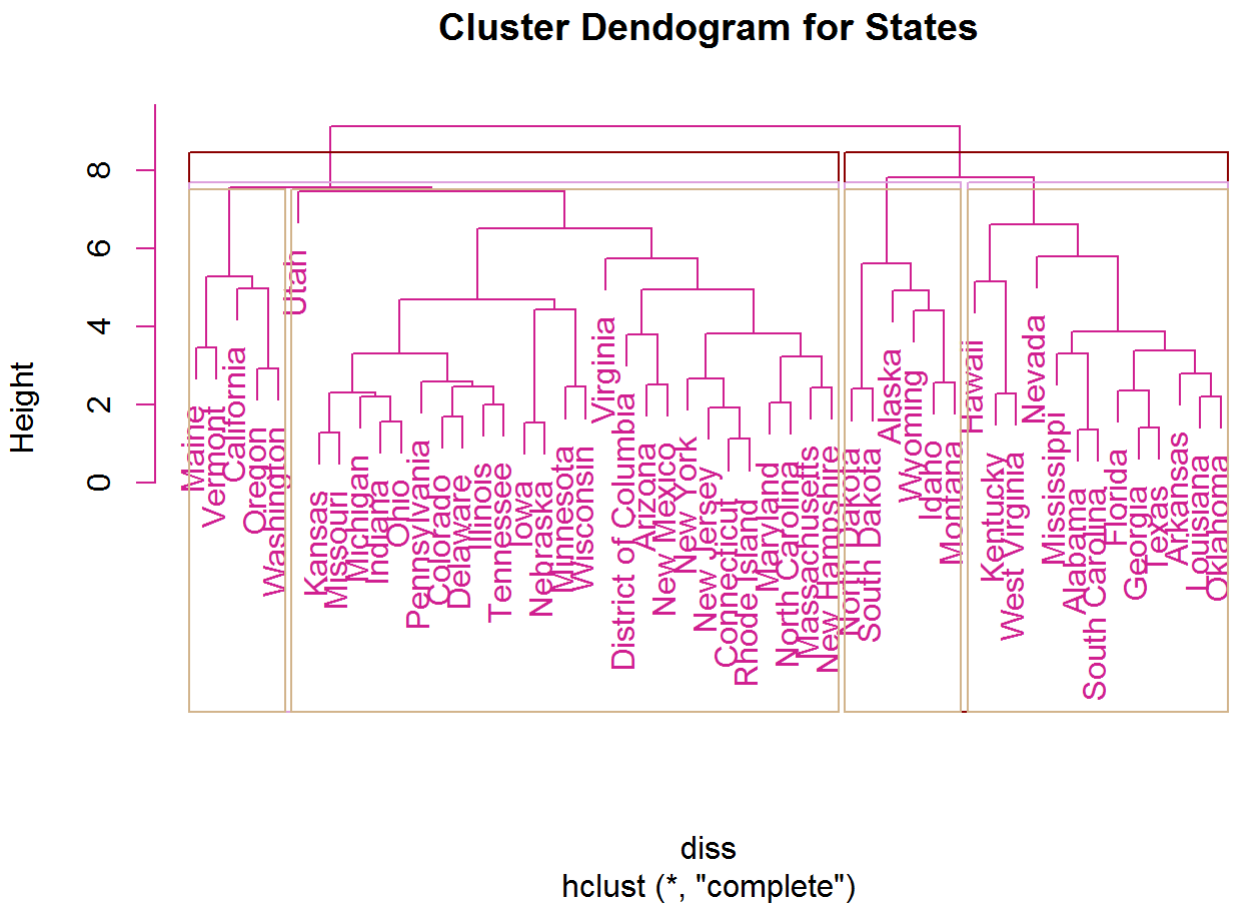
```
##             Length Class  Mode
## merge       100    -none- numeric
## height       50    -none- numeric
## order        51    -none- numeric
## labels       51    -none- character
## method        1    -none- character
## call          2    -none- call
## dist.method   1    -none- character
```

Ploting a Cluster Dendogram and drawing boxes around the clusters.

```
state.dend <- plot(state.clust,
                   col = "violetred",
                   main = "Cluster Dendogram for States"
                   )
state.dend
```

```
## NULL
```

```
rect.hclust(state.clust, k = 2, border = "darkred")
rect.hclust(state.clust, k = 3, border = "plum")
rect.hclust(state.clust, k = 4, border = "tan")
```


Cluster Dendogram for States
Height
0
2
4
6
8
Maine
Vermont
California
Oregon
Washington
Utah
Kansas
Missouri
Michigan
Indiana
Ohio
Pennsylvania
Colorado
Delaware
Illinois
Tennessee
Iowa
Nebraska
Minnesota
Wisconsin
Virginia
District of Columbia
Arizona
New Mexico
New York
New Jersey
Connecticut
Rhode Island
Maryland
North Carolina
Massachusetts
New Hampshire
North Dakota
South Dakota
Alaska
Wyoming
Idaho
Montana
Hawaii
Kentucky
West Virginia
Nevada
Mississippi
Alabama
South Carolina
Florida
Georgia
Texas
Arkansas
Louisiana
Oklahoma
diss
hclust (*, "complete")