Machine learning final

2022-12-06

hr <- read.csv("~/Downloads/human-rights-vdem.csv")

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.4   
## ✔ tibble 3.1.6 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0   
## ✔ readr 2.1.2 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(factoextra)

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

library(ISLR)  
library(flexclust)

## Loading required package: grid  
## Loading required package: lattice  
## Loading required package: modeltools  
## Loading required package: stats4

library(knitr)  
library(stats)  
library(cluster)

summary(hr)

## Entity Code Year civ\_libs\_vdem\_owid  
## Length:32969 Length:32969 Min. :1789 Min. :0.0090   
## Class :character Class :character 1st Qu.:1870 1st Qu.:0.2040   
## Mode :character Mode :character Median :1931 Median :0.3790   
## Mean :1923 Mean :0.4393   
## 3rd Qu.:1977 3rd Qu.:0.6500   
## Max. :2021 Max. :0.9780   
## civ\_libs\_vdem\_high\_owid civ\_libs\_vdem\_low\_owid  
## Min. :0.0130 Min. :0.0030   
## 1st Qu.:0.2450 1st Qu.:0.1540   
## Median :0.4360 Median :0.3240   
## Mean :0.4828 Mean :0.3947   
## 3rd Qu.:0.7070 3rd Qu.:0.5970   
## Max. :0.9910 Max. :0.9720

hr1 <- hr[,c(3:6)]  
summary(hr1)

## Year civ\_libs\_vdem\_owid civ\_libs\_vdem\_high\_owid  
## Min. :1789 Min. :0.0090 Min. :0.0130   
## 1st Qu.:1870 1st Qu.:0.2040 1st Qu.:0.2450   
## Median :1931 Median :0.3790 Median :0.4360   
## Mean :1923 Mean :0.4393 Mean :0.4828   
## 3rd Qu.:1977 3rd Qu.:0.6500 3rd Qu.:0.7070   
## Max. :2021 Max. :0.9780 Max. :0.9910   
## civ\_libs\_vdem\_low\_owid  
## Min. :0.0030   
## 1st Qu.:0.1540   
## Median :0.3240   
## Mean :0.3947   
## 3rd Qu.:0.5970   
## Max. :0.9720

USA <- hr[c(31024:31084),c(1:6)]  
summary(USA)

## Entity Code Year civ\_libs\_vdem\_owid  
## Length:61 Length:61 Min. :1961 Min. :0.8160   
## Class :character Class :character 1st Qu.:1976 1st Qu.:0.9100   
## Mode :character Mode :character Median :1991 Median :0.9360   
## Mean :1991 Mean :0.9212   
## 3rd Qu.:2006 3rd Qu.:0.9490   
## Max. :2021 Max. :0.9510   
## civ\_libs\_vdem\_high\_owid civ\_libs\_vdem\_low\_owid  
## Min. :0.8590 Min. :0.7750   
## 1st Qu.:0.9400 1st Qu.:0.8860   
## Median :0.9580 Median :0.9190   
## Mean :0.9476 Mean :0.9001   
## 3rd Qu.:0.9690 3rd Qu.:0.9330   
## Max. :0.9730 Max. :0.9380

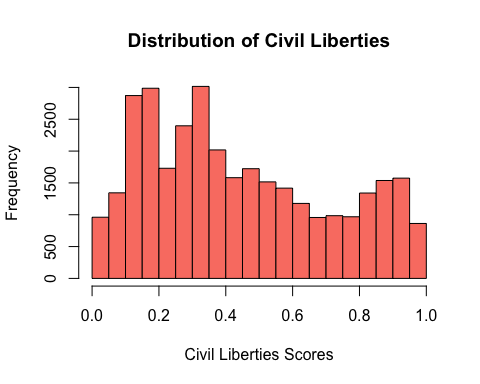
USA1 <- hr[c(31024:31084),c(3:6)]  
summary(USA1)

## Year civ\_libs\_vdem\_owid civ\_libs\_vdem\_high\_owid  
## Min. :1961 Min. :0.8160 Min. :0.8590   
## 1st Qu.:1976 1st Qu.:0.9100 1st Qu.:0.9400   
## Median :1991 Median :0.9360 Median :0.9580   
## Mean :1991 Mean :0.9212 Mean :0.9476   
## 3rd Qu.:2006 3rd Qu.:0.9490 3rd Qu.:0.9690   
## Max. :2021 Max. :0.9510 Max. :0.9730   
## civ\_libs\_vdem\_low\_owid  
## Min. :0.7750   
## 1st Qu.:0.8860   
## Median :0.9190   
## Mean :0.9001   
## 3rd Qu.:0.9330   
## Max. :0.9380

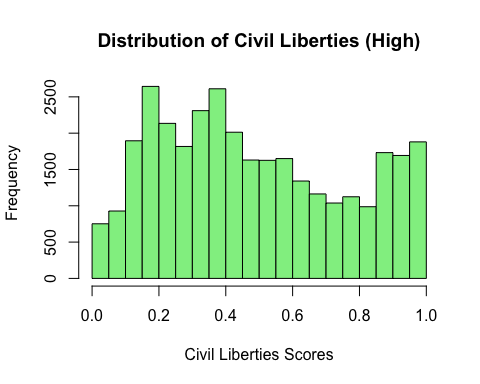
USA2 <- hr[c(30853:31084),c(3:6)]  
summary(USA2)

## Year civ\_libs\_vdem\_owid civ\_libs\_vdem\_high\_owid  
## Min. :1790 Min. :0.5210 Min. :0.574   
## 1st Qu.:1848 1st Qu.:0.5660 1st Qu.:0.615   
## Median :1906 Median :0.7580 Median :0.807   
## Mean :1906 Mean :0.7411 Mean :0.784   
## 3rd Qu.:1963 3rd Qu.:0.8197 3rd Qu.:0.868   
## Max. :2021 Max. :0.9510 Max. :0.973   
## civ\_libs\_vdem\_low\_owid  
## Min. :0.4770   
## 1st Qu.:0.5080   
## Median :0.7145   
## Mean :0.6992   
## 3rd Qu.:0.7845   
## Max. :0.9380

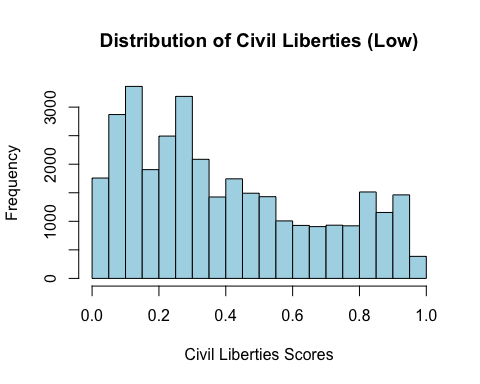
hist(hr$civ\_libs\_vdem\_owid, col= "salmon" , border = 'black', main="Distribution of Civil Liberties", xlab="Civil Liberties Scores")



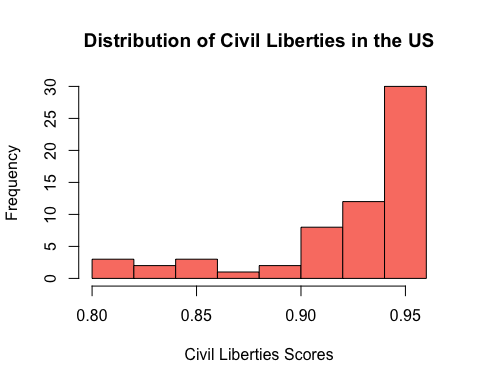
hist(hr$civ\_libs\_vdem\_high\_owid, col = "light green" , border = 'black', main="Distribution of Civil Liberties (High)", xlab="Civil Liberties Scores")



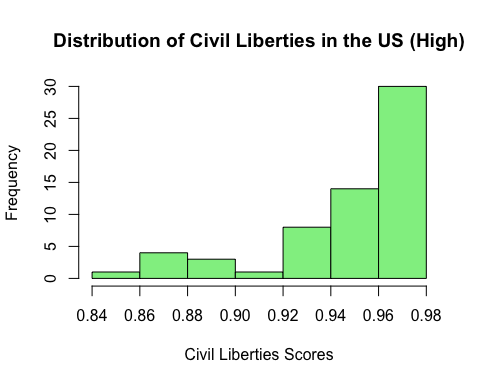
hist(hr$civ\_libs\_vdem\_low\_owid, col ="light blue", border = 'black', main="Distribution of Civil Liberties (Low)", xlab="Civil Liberties Scores")



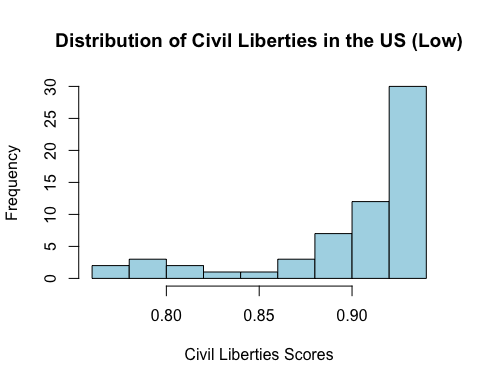
hist(USA1$civ\_libs\_vdem\_owid, col= "salmon" , border = 'black', main="Distribution of Civil Liberties in the US", xlab="Civil Liberties Scores")



hist(USA1$civ\_libs\_vdem\_high\_owid, col = "light green" , border = 'black', main="Distribution of Civil Liberties in the US (High)", xlab="Civil Liberties Scores")



hist(USA1$civ\_libs\_vdem\_low\_owid, col ="light blue", border = 'black', main="Distribution of Civil Liberties in the US (Low)", xlab="Civil Liberties Scores")



k4 <-kmeans(hr1, centers = 5, nstart = 10)  
k4$center

## Year civ\_libs\_vdem\_owid civ\_libs\_vdem\_high\_owid civ\_libs\_vdem\_low\_owid  
## 1 1814.753 0.2927468 0.3393770 0.2409616  
## 2 1862.836 0.3918773 0.4402326 0.3408283  
## 3 2000.560 0.6269055 0.6667359 0.5893506  
## 4 1914.839 0.4067982 0.4511226 0.3613799  
## 5 1958.209 0.4044416 0.4456628 0.3624159

k4$size

## [1] 4724 5349 7759 7316 7821

fviz\_cluster(k4, data = hr1)



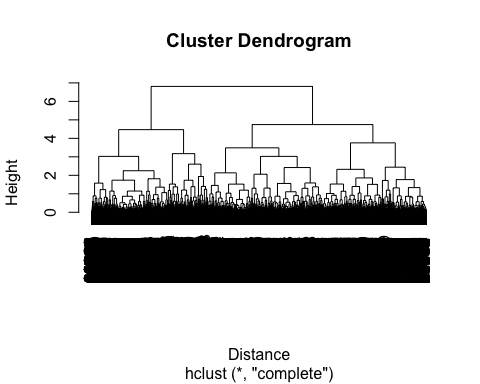
k4 = kcca(hr1, k=5, kccaFamily("kmedians"))  
k4

## kcca object of family 'kmedians'   
##   
## call:  
## kcca(x = hr1, k = 5, family = kccaFamily("kmedians"))  
##   
## cluster sizes:  
##   
## 1 2 3 4 5   
## 7617 5055 7267 5414 7616

library(knitr)  
library(stats)  
library(cluster)

numericaldata <- data.frame(hr[,3:6])  
OmitMissing <- na.omit(numericaldata)  
Normalise <- scale(OmitMissing)

Distance <- dist(Normalise, method = "euclidean")  
  
Clustering\_heirarchial <- hclust(Distance, method = "complete")  
plot(Clustering\_heirarchial)



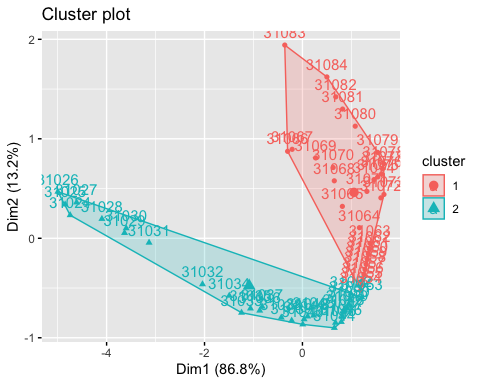
k4 <-kmeans(USA1, centers = 2, nstart = 10)  
k4$center

## Year civ\_libs\_vdem\_owid civ\_libs\_vdem\_high\_owid civ\_libs\_vdem\_low\_owid  
## 1 2006.0 0.9390000 0.9611613 0.9213226  
## 2 1975.5 0.9027333 0.9336000 0.8782333

k4$size

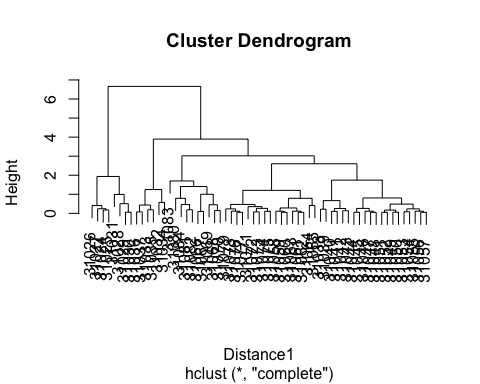
## [1] 31 30

fviz\_cluster(k4, data = USA1)

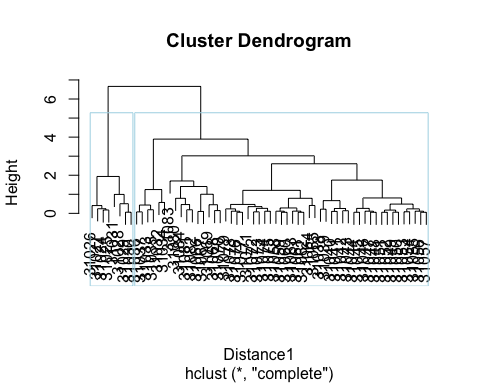


numericaldata1 <- data.frame(USA[,3:6])  
OmitMissing1 <- na.omit(numericaldata1)  
Normalise1 <- scale(OmitMissing1)

Distance1 <- dist(Normalise1, method = "euclidean")  
  
Clustering\_heirarchial1 <- hclust(Distance1, method = "complete")  
plot(Clustering\_heirarchial1)



plot(Clustering\_heirarchial1)  
rect.hclust(Clustering\_heirarchial1, k = 2, border = "lightblue")



singleCH <- agnes(Normalise1, method = "single")  
completeCH <- agnes(Normalise1, method = "complete")  
averageCH <- agnes(Normalise1, method = "average")  
wardCH <- agnes(Normalise1, method = "ward")  
  
print(singleCH$civ\_libs\_vdem\_owid)

## NULL

print(completeCH$civ\_libs\_vdem\_owid)

## NULL

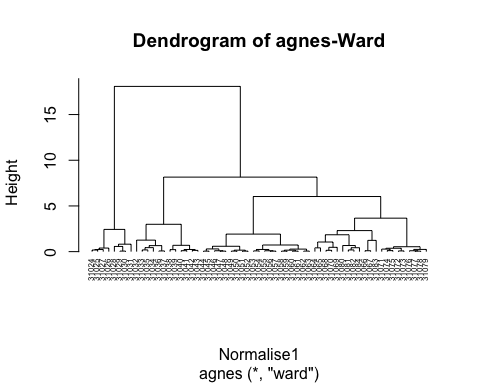
print(averageCH$civ\_libs\_vdem\_owid)

## NULL

print(wardCH$civ\_libs\_vdem\_owid)

## NULL

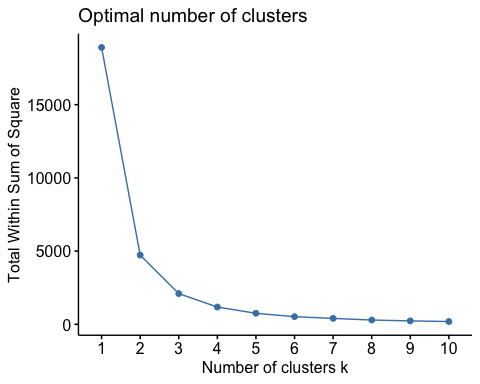
pltree(wardCH, cex = .5, hang = -2, main = 'Dendrogram of agnes-Ward')



cutree(wardCH, k = 4)

## [1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
## [39] 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4

fviz\_nbclust(USA1, kmeans, method = "wss")



fviz\_nbclust(USA1, kmeans, method = "silhouette")

