ch5-2

Harry Woo

2020 6 16

rm(list = ls())  
library(dplyr)  
library(ggplot2)  
library(ggrepel)  
library(smacof)

# source("readMatrix.r")  
# city <- readMatrix()  
# E:/KNOU-MVA/citydist.txt  
# 2  
# 10  
# city\_name <- scan("E:/KNOU-MVA/cityname.txt", what = "")  
# colnames(city) <- city\_name  
# rownames(city) <- city\_name  
# city  
  
data(UScitiesD)  
str(UScitiesD)

## 'dist' int [1:45] 587 1212 701 1936 604 748 2139 2182 543 920 ...  
## - attr(\*, "Labels")= chr [1:10] "Atlanta" "Chicago" "Denver" "Houston" ...  
## - attr(\*, "Size")= int 10  
## - attr(\*, "call")= language as.dist.default(m = t(cities.mat))  
## - attr(\*, "Diag")= logi FALSE  
## - attr(\*, "Upper")= logi FALSE

attributes(UScitiesD)

## $Labels  
## [1] "Atlanta" "Chicago" "Denver" "Houston"   
## [5] "LosAngeles" "Miami" "NewYork" "SanFrancisco"   
## [9] "Seattle" "Washington.DC"  
##   
## $Size  
## [1] 10  
##   
## $call  
## as.dist.default(m = t(cities.mat))  
##   
## $class  
## [1] "dist"  
##   
## $Diag  
## [1] FALSE  
##   
## $Upper  
## [1] FALSE

city <- UScitiesD  
city[44] <- 2422  
city[44]

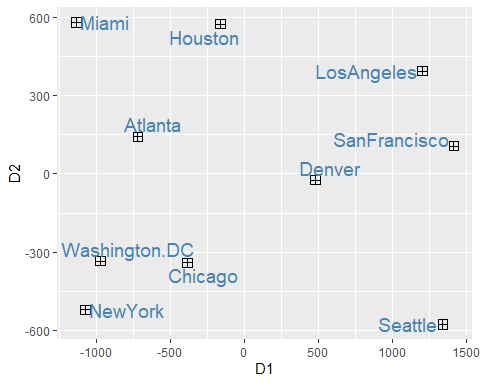
## [1] 2422

city <- as.matrix(city)

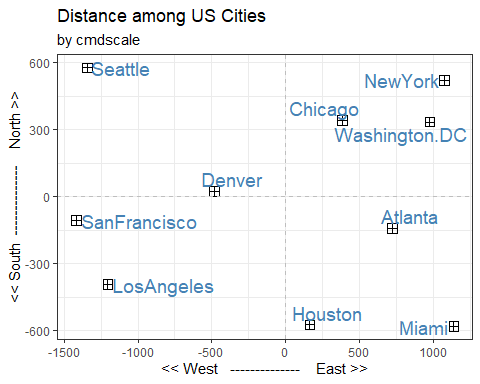
# cmdscale 실행 (2차원, 지도상의 거리)  
MDS1 <- cmdscale(city, k = 2)  
MDS1

## [,1] [,2]  
## Atlanta -719.5645 142.2177  
## Chicago -382.4325 -341.2018  
## Denver 481.7394 -23.6247  
## Houston -162.0259 573.4239  
## LosAngeles 1204.2465 393.5545  
## Miami -1134.8011 580.5064  
## NewYork -1073.0511 -520.9490  
## SanFrancisco 1416.4273 106.6886  
## Seattle 1342.6991 -576.6157  
## Washington.DC -973.2371 -333.9998

data.frame(D1 = MDS1[,1], D2 = MDS1[,2]) %>%   
 ggplot(aes(x = D1, y = D2, label = row.names(MDS1))) +   
 geom\_point(pch = 12, size = 3) +   
 geom\_text\_repel(size = 5, colour = "steelblue")



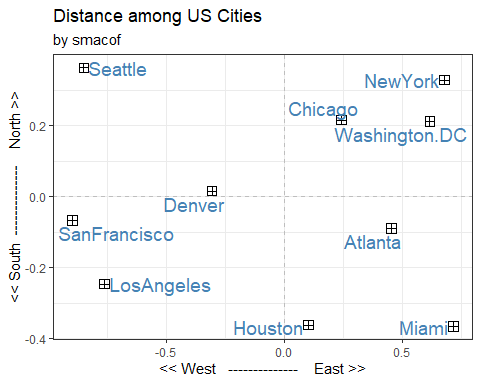
# Plot  
MDS1 <- -1 \* MDS1 # 실제 위치에 맞추기 위해 rotate  
  
data.frame(D1 = MDS1[,1], D2 = MDS1[,2]) %>%   
 ggplot(aes(x = D1, y = D2, label = row.names(MDS1))) +   
 geom\_point(pch = 12, size = 3) +   
 geom\_hline(yintercept = 0, lty = 2, col = "gray") +  
 geom\_vline(xintercept = 0, lty = 2, col = "gray") +  
 geom\_text\_repel(size = 5, colour = "steelblue") +  
 xlab("<< West -------------- East >>") +  
 ylab("<< South -------------- North >>") +  
 ggtitle("Distance among US Cities", subtitle = "by cmdscale") +  
 theme\_bw()



# smacofSym 활용  
MDS2 <- smacofSym(city, ndim = 2, type = "ordinal")  
attributes(MDS2)

## $names  
## [1] "delta" "dhat" "confdist" "iord" "conf" "stress"   
## [7] "spp" "ndim" "weightmat" "resmat" "rss" "init"   
## [13] "model" "niter" "nobj" "type" "call"   
##   
## $class  
## [1] "smacofB" "smacof"

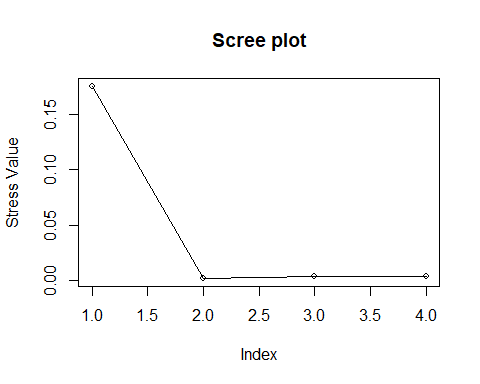
# Plot  
MDS2$conf <- -1 \* MDS2$conf # 실제 위치에 맞추기 위해 rotate  
  
data.frame(D1 = MDS2$conf[,1], D2 = MDS2$conf[,2]) %>%   
 ggplot(aes(x = D1, y = D2, label = row.names(MDS2$conf))) +   
 geom\_point(pch = 12, size = 3) +   
 geom\_hline(yintercept = 0, lty = 2, col = "gray") +  
 geom\_vline(xintercept = 0, lty = 2, col = "gray") +  
 geom\_text\_repel(size = 5, colour = "steelblue") +  
 xlab("<< West -------------- East >>") +  
 ylab("<< South -------------- North >>") +  
 ggtitle("Distance among US Cities", subtitle = "by smacof") +  
 theme\_bw()



# Stress Value : 0.05~0.10 만족, 0~0.05 매우좋음  
# smacof 버전의 차이로 보임, 최신 버전에는 stress.m 이 없음 (m : metric)  
MDS2$stress # 매우 좋음

## [1] 0.000127482

# Scree plot  
MDS2.1 <- smacofSym(city, ndim = 1)  
MDS2.2 <- smacofSym(city, ndim = 2)  
MDS2.3 <- smacofSym(city, ndim = 3)  
MDS2.4 <- smacofSym(city, ndim = 4)  
  
stress\_value = c(MDS2.1$stress, MDS2.2$stress, MDS2.3$stress, MDS2.4$stress)  
plot(stress\_value, type = "l", main = "Scree plot", ylab = "Stress Value")  
points(stress\_value, cex = 0.9)



# 적합도  
# z01dist <- (MDS2$delta - min(MDS2$delta))/(max(MDS2$delta) - min(MDS2$delta))  
  
plot(MDS2$delta, MDS2$confdist,  
 xlab = "Observed distance", ylab = "Configuration distance")

