

# USArrest\_clustering

```
library(graphics)
library(magrittr)
library(MASS)
library(psych)
library(cluster)
```

```
#install.packages("factoextra")
data("USArrests")
head(USArrests)
```

```
##           Murder Assault UrbanPop Rape
## Alabama      13.2     236      58 21.2
## Alaska       10.0     263      48 44.5
## Arizona       8.1     294      80 31.0
## Arkansas      8.8     190      50 19.5
## California    9.0     276      91 40.6
## Colorado      7.9     204      78 38.7
```

```
str(USArrests)
```

```
## 'data.frame':   50 obs. of  4 variables:
## $ Murder   : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
## $ Assault  : int  236 263 294 190 276 204 110 238 335 211 ...
## $ UrbanPop: int  58 48 80 50 91 78 77 72 80 60 ...
## $ Rape     : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```

```
# 4가지 변수의 척도가 다르므로 scale 함수를 적용해 표준 점수로 변환 시킨다.
USArrests <- USArrests %>% scale
dim(USArrests)
```

```
## [1] 50  4
```

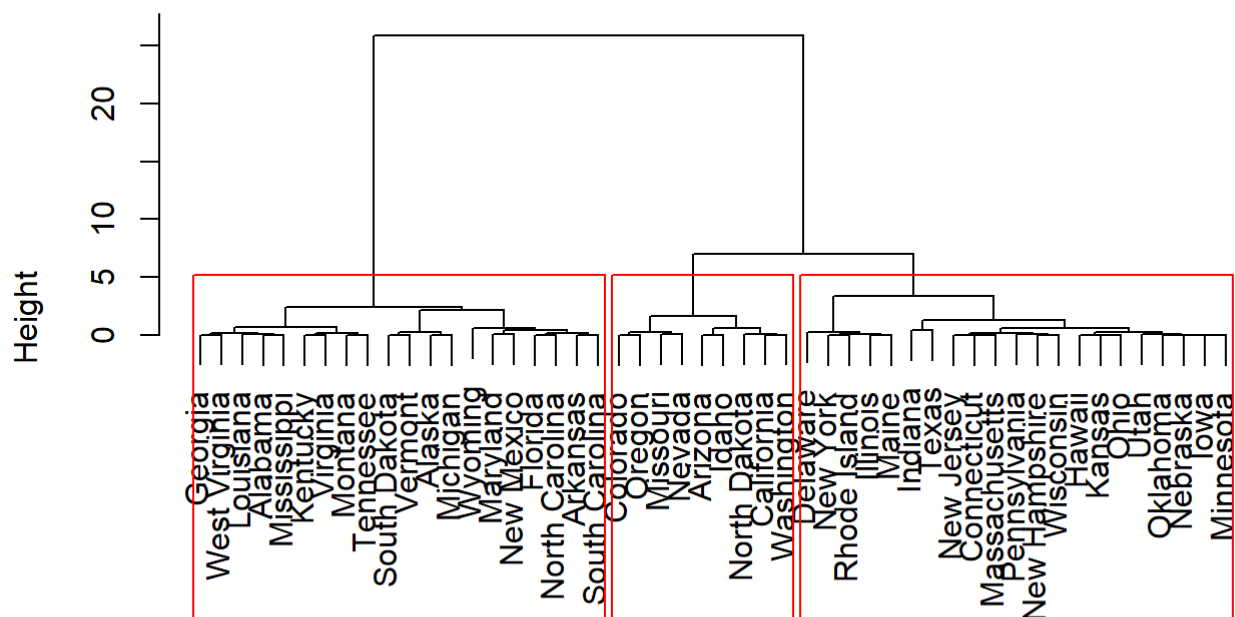
```
# Hierarchical Clustering
# 군집 간 거리를 측정하여 가장 유사한 군집끼리 묶음, 하나의 군집이 형성될 때까지 반복

# 주어진 행렬을 거리 행렬로 변환
dist.mat <- as.dist(1 - cor(t(USArrests)))

h1 <- hclust(dist.mat, method = "ward.D")
plot(h1)

#cutree : 군집의 인덱스를 구하는 함수
clusterCut1 <- cutree(h1, k=3)
rect.hclust(h1, k = 3)
```

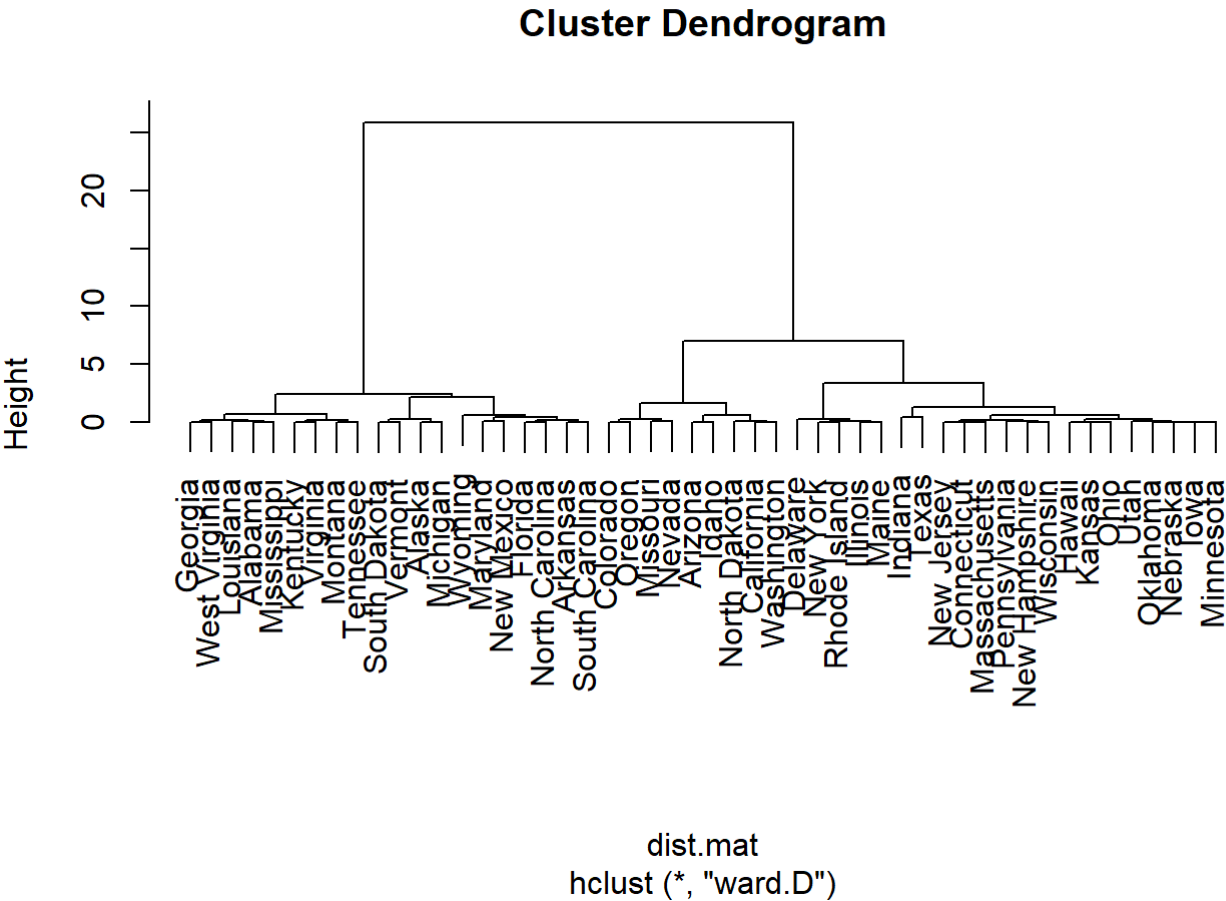
## Cluster Dendrogram



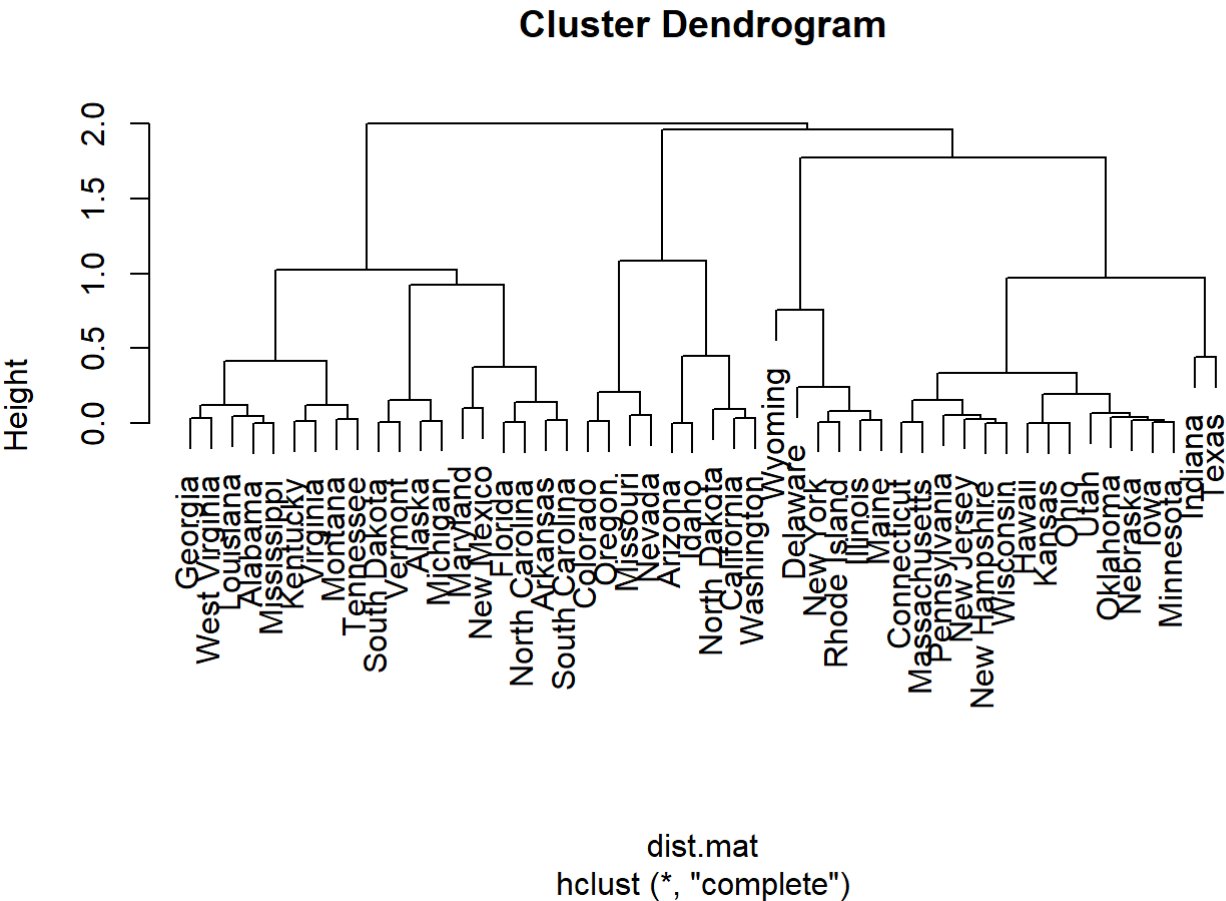
dist.mat  
hclust (\*, "ward.D")

```
h1 <- hclust(dist.mat, method = "ward.D")
h2 <- hclust(dist.mat, method = "complete")
h3 <- hclust(dist.mat, method = "average")
h4 <- hclust(dist.mat, method = "single")
```

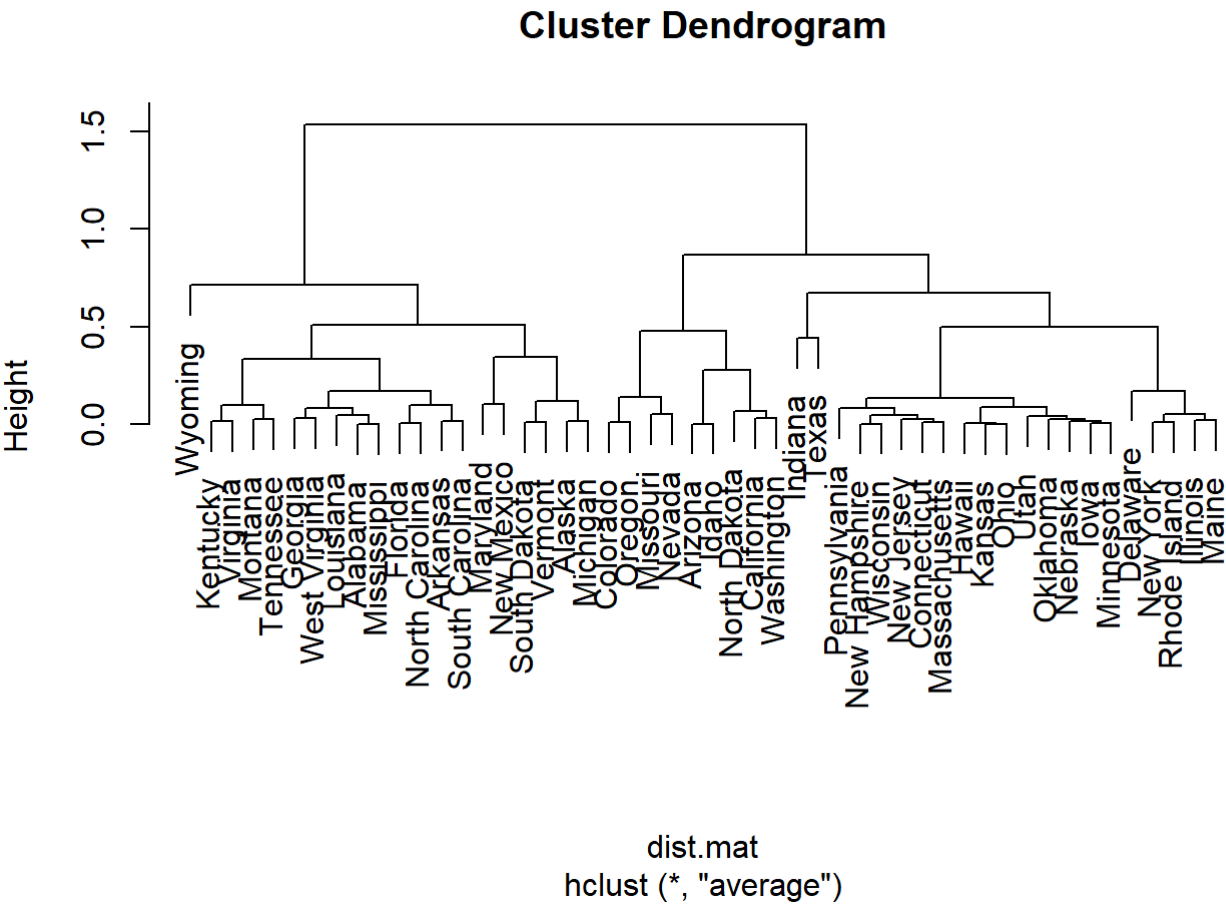
```
plot(h1)
```



```
plot(h2)
```



```
plot(h3)
```



```
plot(h4)
```

Cluster Dendrogram

