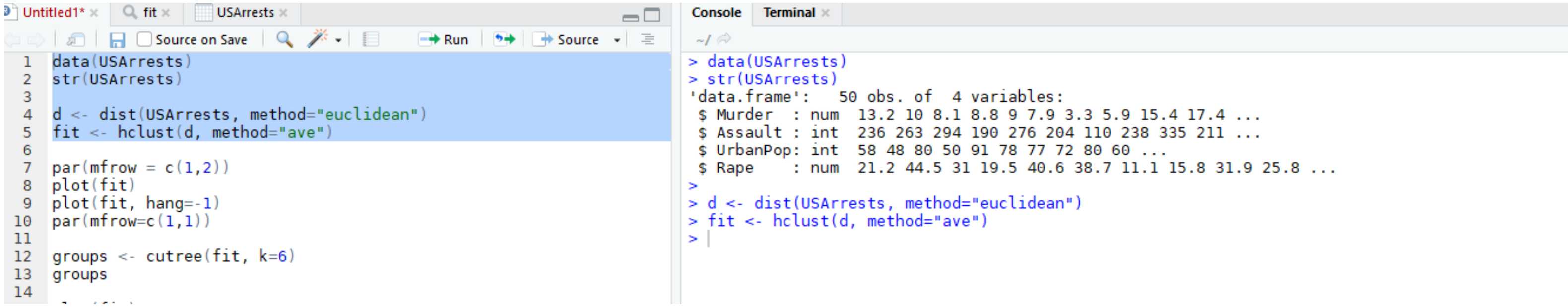
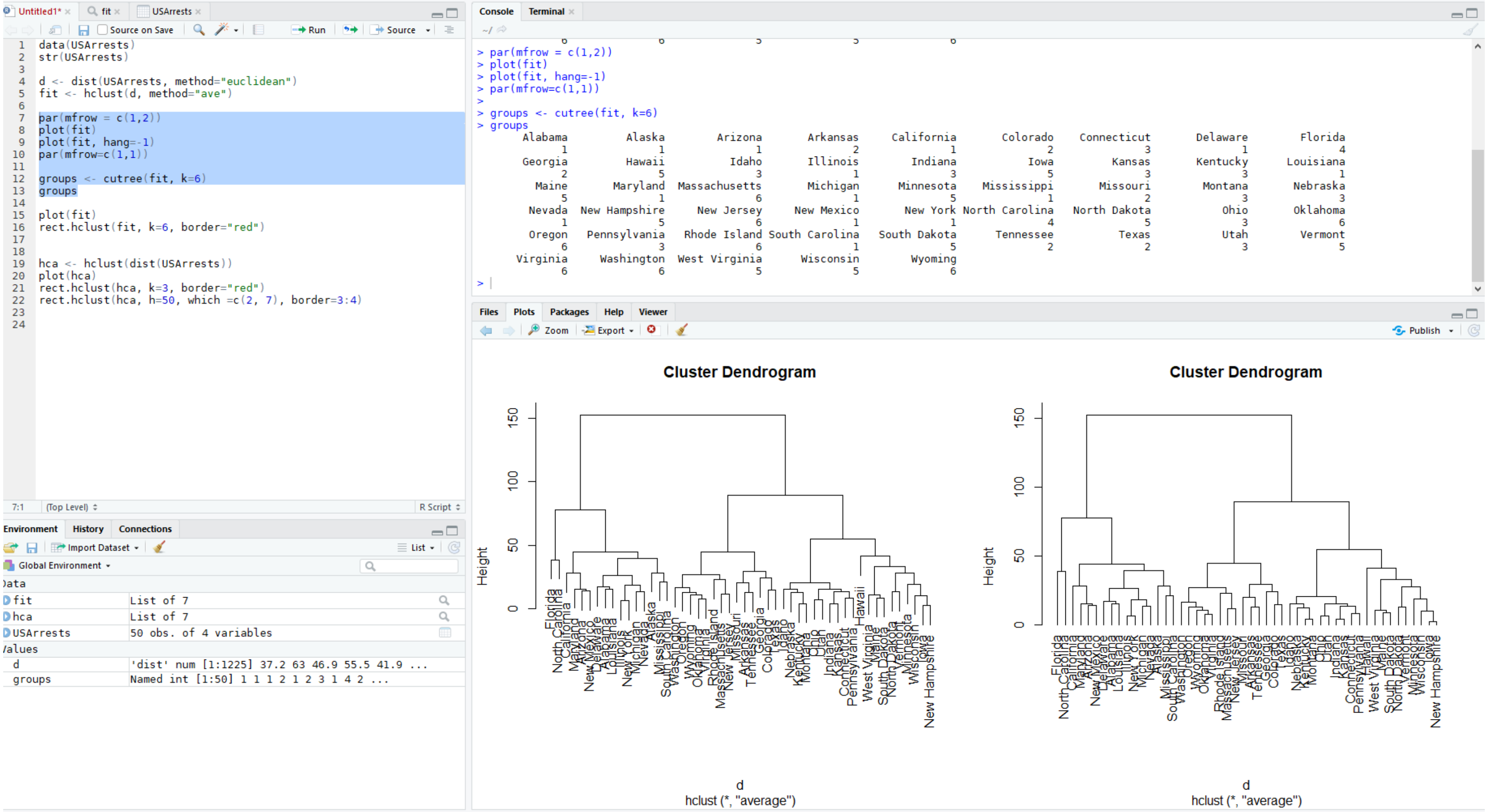


```r
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

```

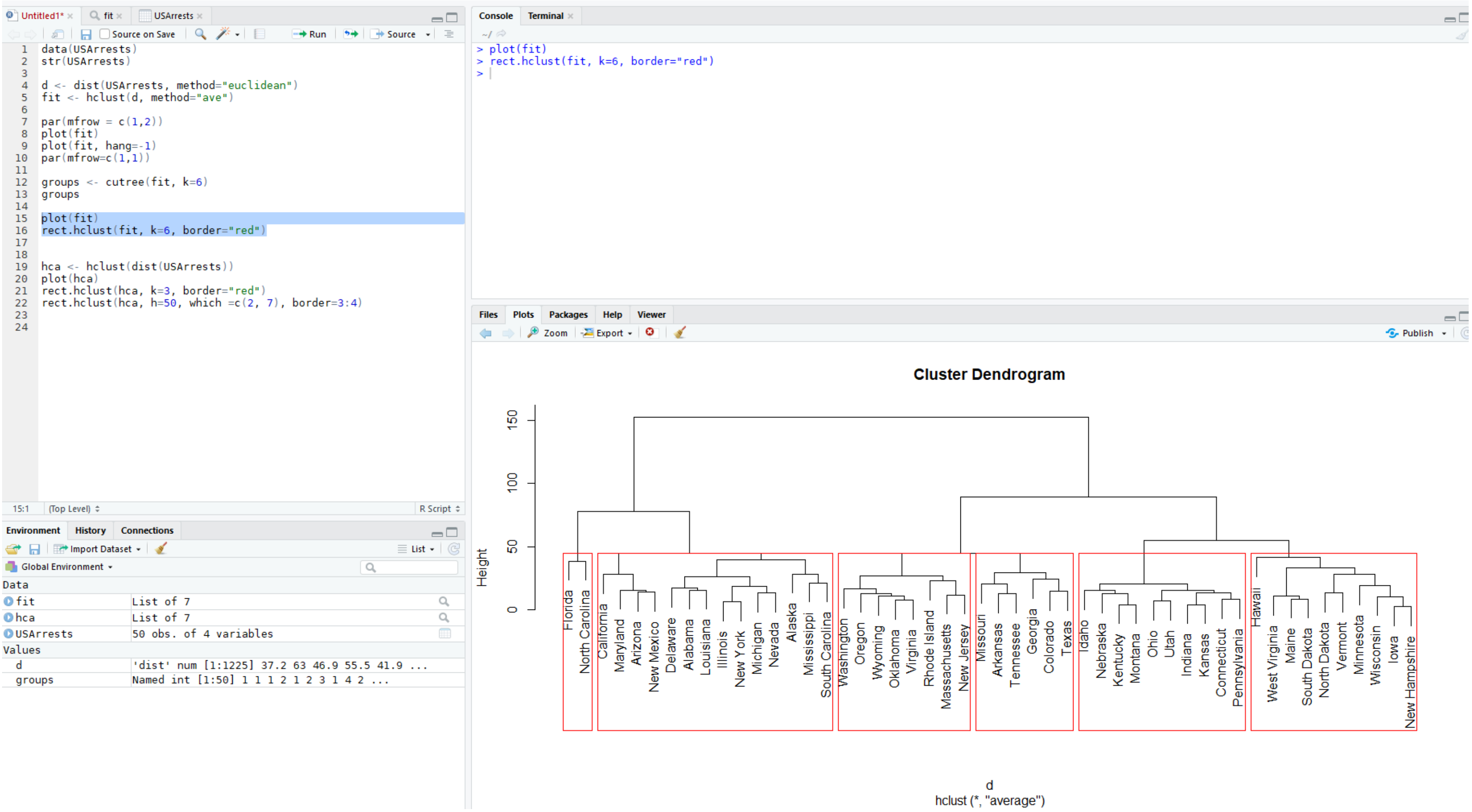
```
> data(USArrests)
> str(USArrests)
'data.frame':	50 obs. of  4 variables:
 $ Murder  : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
 $ Assault : int  236 263 294 190 276 204 110 238 335 211 ...
 $ UrbanPop: int  58 48 80 50 91 78 77 72 80 60 ...
 $ Rape    : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
>
> d <- dist(USArrests, method="euclidean")
> fit <- hclust(d, method="ave")
> 
```



```r
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)
```

```
             6              6              5              5              6
> par(mfrow = c(1,2))
> plot(fit)
> plot(fit, hang=-1)
> par(mfrow=c(1,1))
>
> groups <- cutree(fit, k=6)
> groups
       Alabama         Alaska        Arizona       Arkansas     California       Colorado    Connecticut       Delaware        Florida
             1              1              1              2              1              2              3              1              4
       Georgia         Hawaii          Idaho       Illinois        Indiana           Iowa         Kansas       Kentucky      Louisiana
             2              5              3              1              3              5              3              3              1
         Maine       Maryland  Massachusetts       Michigan      Minnesota    Mississippi       Missouri        Montana       Nebraska
             5              1              6              1              5              1              2              3              3
        Nevada  New Hampshire     New Jersey     New Mexico       New York North Carolina   North Dakota           Ohio       Oklahoma
             1              5              6              1              1              4              5              3              6
        Oregon   Pennsylvania   Rhode Island South Carolina   South Dakota      Tennessee          Texas           Utah        Vermont
             6              3              6              1              5              2              2              3              5
      Virginia     Washington  West Virginia      Wisconsin        Wyoming
             6              6              5              5              6
> |
```

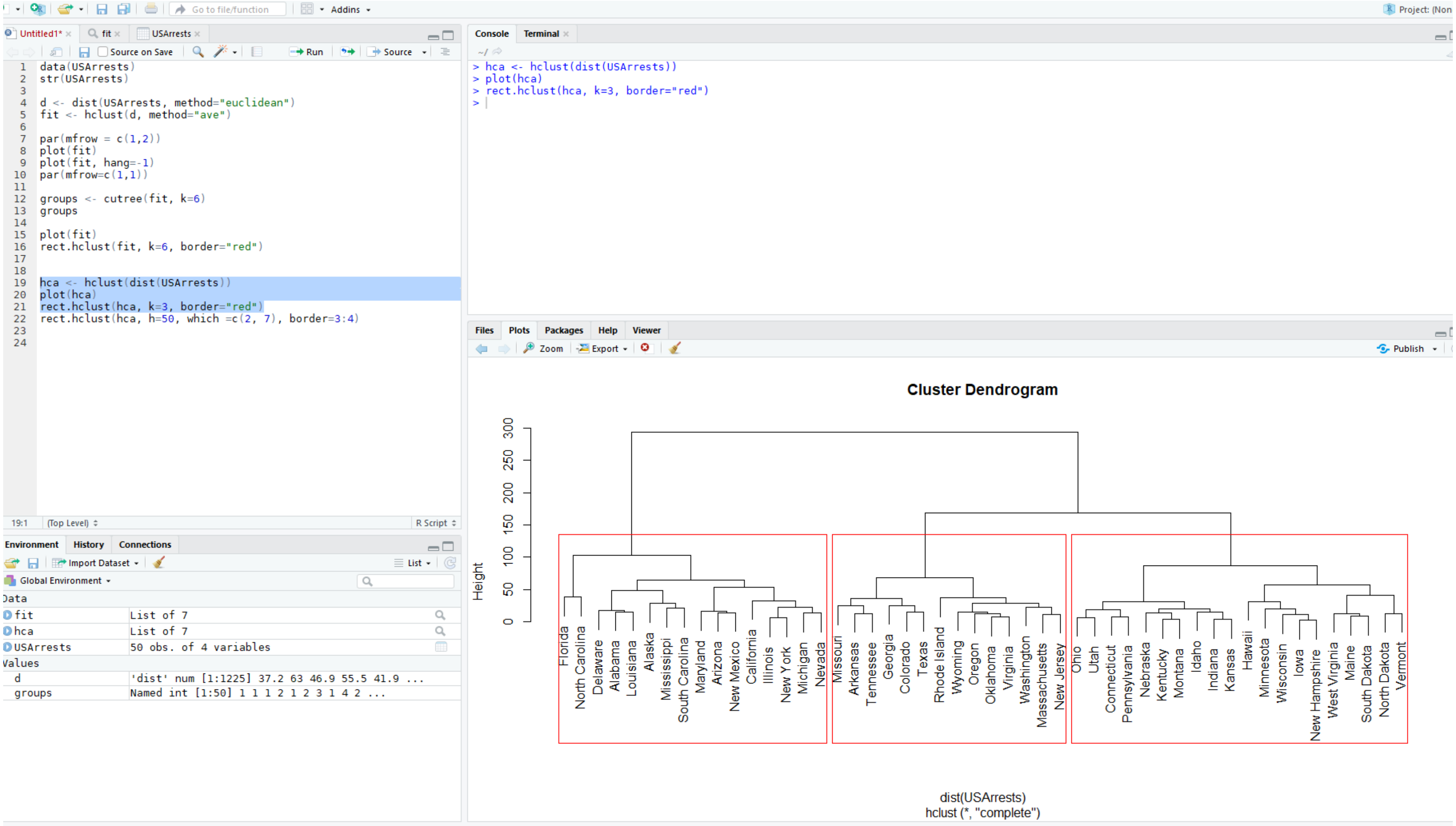
```
Data
fit          List of 7
hca          List of 7
USArrests    50 obs. of 4 variables
Values
d            'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
groups       Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...
```



```
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)
```

```
> plot(fit)
> rect.hclust(fit, k=6, border="red")
> 
```

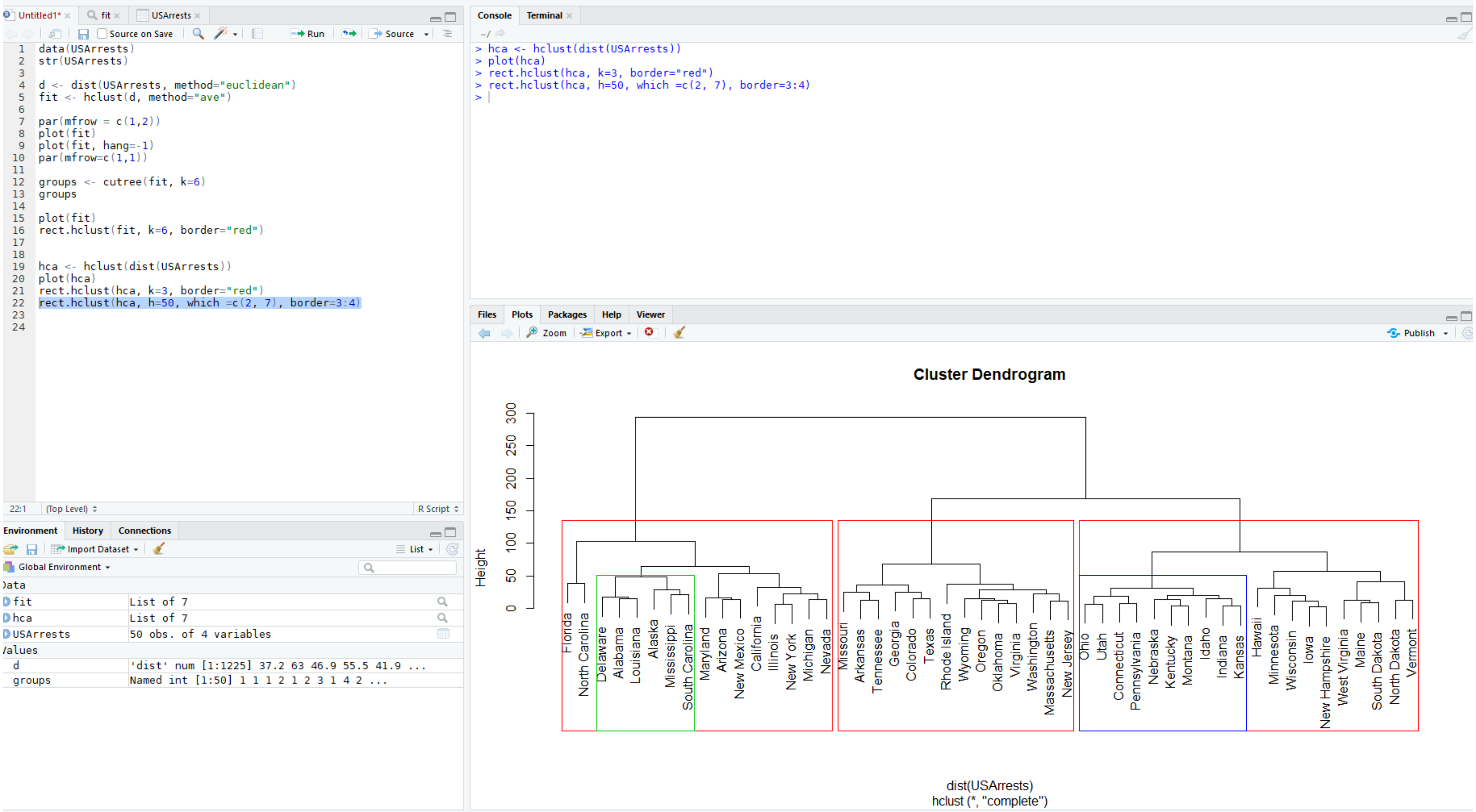
```
Data
fit          List of 7
hca          List of 7
USArrests    50 obs. of 4 variables
Values
 d           'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
 groups      Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...
```



```
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)
```

```
> hca <- hclust(dist(USArrests))
> plot(hca)
> rect.hclust(hca, k=3, border="red")
> 
```

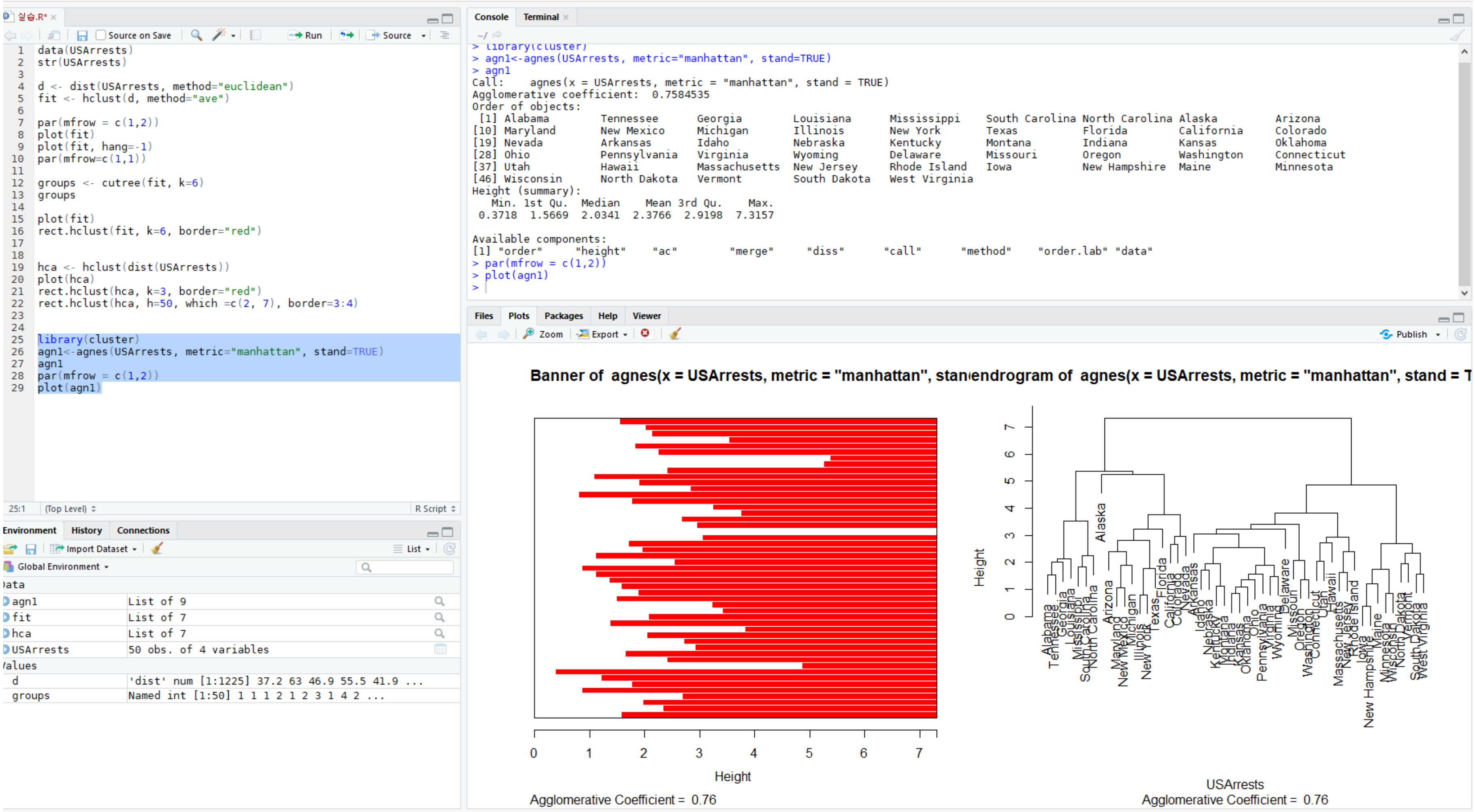
```
Data
fit         List of 7
hca         List of 7
USArrests   50 obs. of 4 variables
Values
d           'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
groups      Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...
```



```
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)
```

```
> hca <- hclust(dist(USArrests))
> plot(hca)
> rect.hclust(hca, k=3, border="red")
> rect.hclust(hca, h=50, which =c(2, 7), border=3:4)
> 
```

```
Data
fit         List of 7
hca         List of 7
USArrests   50 obs. of 4 variables
Values
  d         'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
  groups    Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...
```



```r
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)


library(cluster)
agn1<-agnes(USArrests, metric="manhattan", stand=TRUE)
agn1
par(mfrow = c(1,2))
plot(agn1)
```

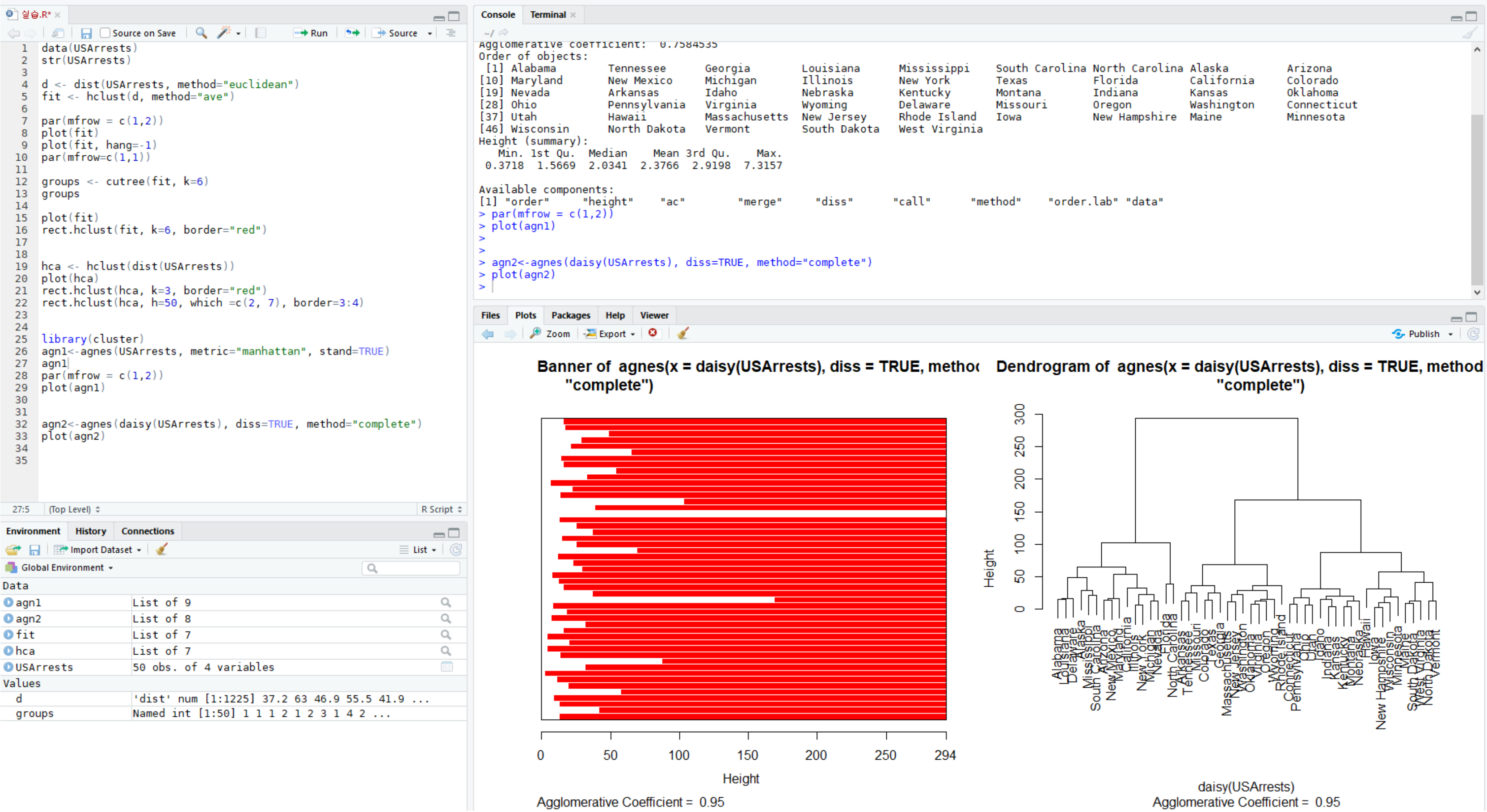
```
> library(cluster)
> agn1<-agnes(USArrests, metric="manhattan", stand=TRUE)
> agn1
Call:    agnes(x = USArrests, metric = "manhattan", stand = TRUE)
Agglomerative coefficient:  0.7584535
Order of objects:
 [1] Alabama        Tennessee      Georgia        Louisiana      Mississippi    South Carolina North Carolina Alaska         Arizona
[10] Maryland       New Mexico     Michigan       Illinois       New York       Texas          Florida        California     Colorado
[19] Nevada         Arkansas       Idaho          Nebraska       Kentucky       Montana        Indiana        Kansas         Oklahoma
[28] Ohio           Pennsylvania   Virginia       Wyoming        Delaware       Missouri       Oregon         Washington     Connecticut
[37] Utah           Hawaii         Massachusetts  New Jersey     Rhode Island   Iowa           New Hampshire  Maine          Minnesota
[46] Wisconsin      North Dakota   Vermont        South Dakota   West Virginia
Height (summary):
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3718  1.5669  2.0341  2.3766  2.9198  7.3157

Available components:
[1] "order"     "height"    "ac"        "merge"     "diss"      "call"      "method"    "order.lab" "data"
> par(mfrow = c(1,2))
> plot(agn1)
> 
```

Data

agn1	List of 9	
fit	List of 7	
hca	List of 7	
USArrests	50 obs. of 4 variables	

Values

d	'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
groups	Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...



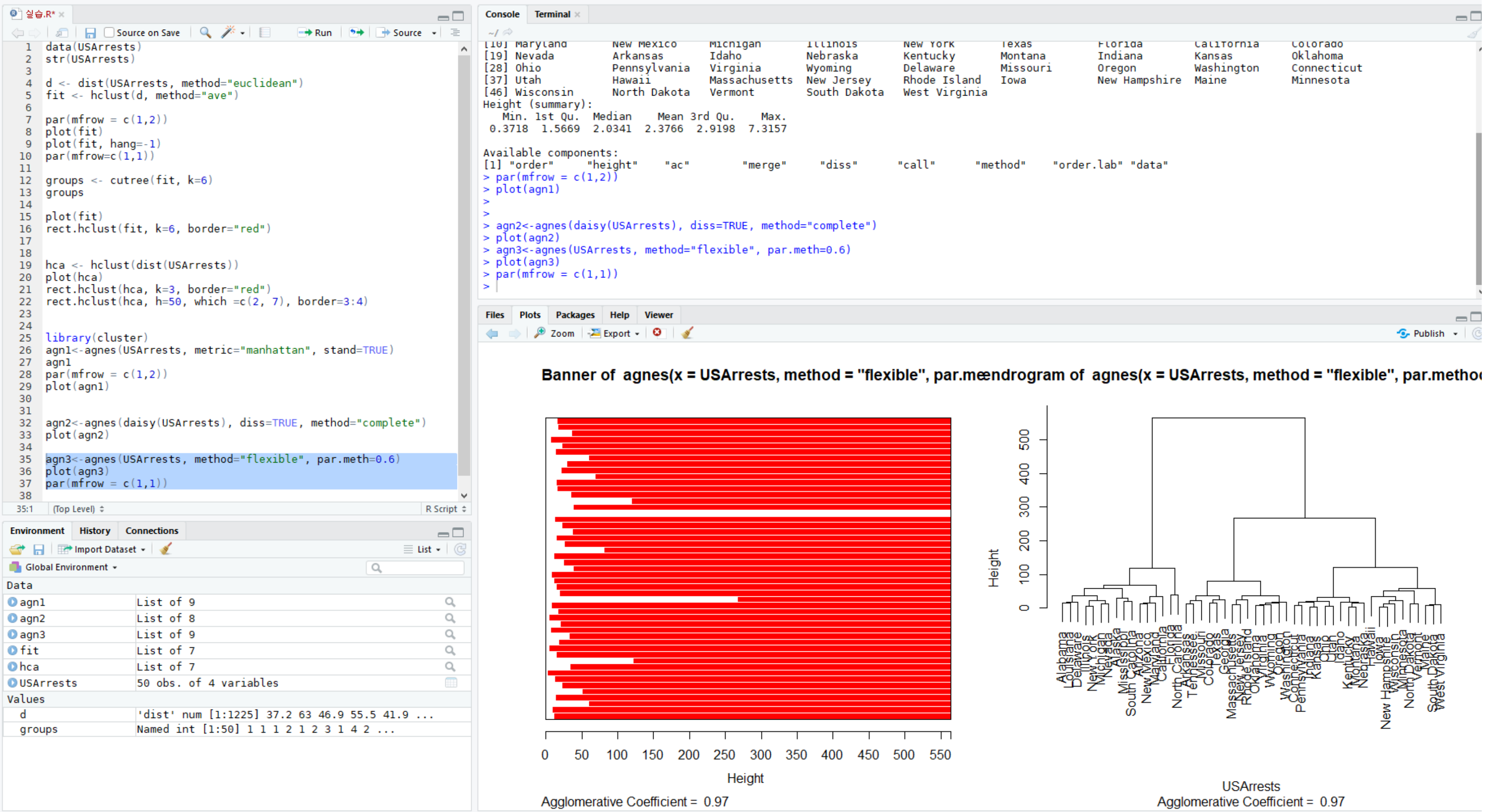
```r
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)


library(cluster)
agn1<-agnes(USArrests, metric="manhattan", stand=TRUE)
agn1
par(mfrow = c(1,2))
plot(agn1)


agn2<-agnes(daisy(USArrests), diss=TRUE, method="complete")
plot(agn2)
```

```
Agglomerative coefficient:  0.7584535
Order of objects:
 [1] Alabama        Tennessee      Georgia        Louisiana      Mississippi    South Carolina North Carolina Alaska         Arizona
[10] Maryland       New Mexico     Michigan       Illinois       New York       Texas          Florida        California     Colorado
[19] Nevada         Arkansas       Idaho          Nebraska       Kentucky       Montana        Indiana        Kansas         Oklahoma
[28] Ohio           Pennsylvania   Virginia       Wyoming        Delaware       Missouri       Oregon         Washington     Connecticut
[37] Utah           Hawaii         Massachusetts  New Jersey     Rhode Island   Iowa           New Hampshire  Maine          Minnesota
[46] Wisconsin      North Dakota   Vermont        South Dakota   West Virginia
Height (summary):
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.3718  1.5669  2.0341  2.3766  2.9198  7.3157 

Available components:
[1] "order"     "height"    "ac"        "merge"     "diss"      "call"      "method"    "order.lab" "data"     
> par(mfrow = c(1,2))
> plot(agn1)
> 
> 
> agn2<-agnes(daisy(USArrests), diss=TRUE, method="complete")
> plot(agn2)
> 
```

```
Data
agn1        List of 9
agn2        List of 8
fit         List of 7
hca         List of 7
USArrests   50 obs. of 4 variables
Values
d           'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
groups      Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...
```

Agglomerative Coefficient = 0.95

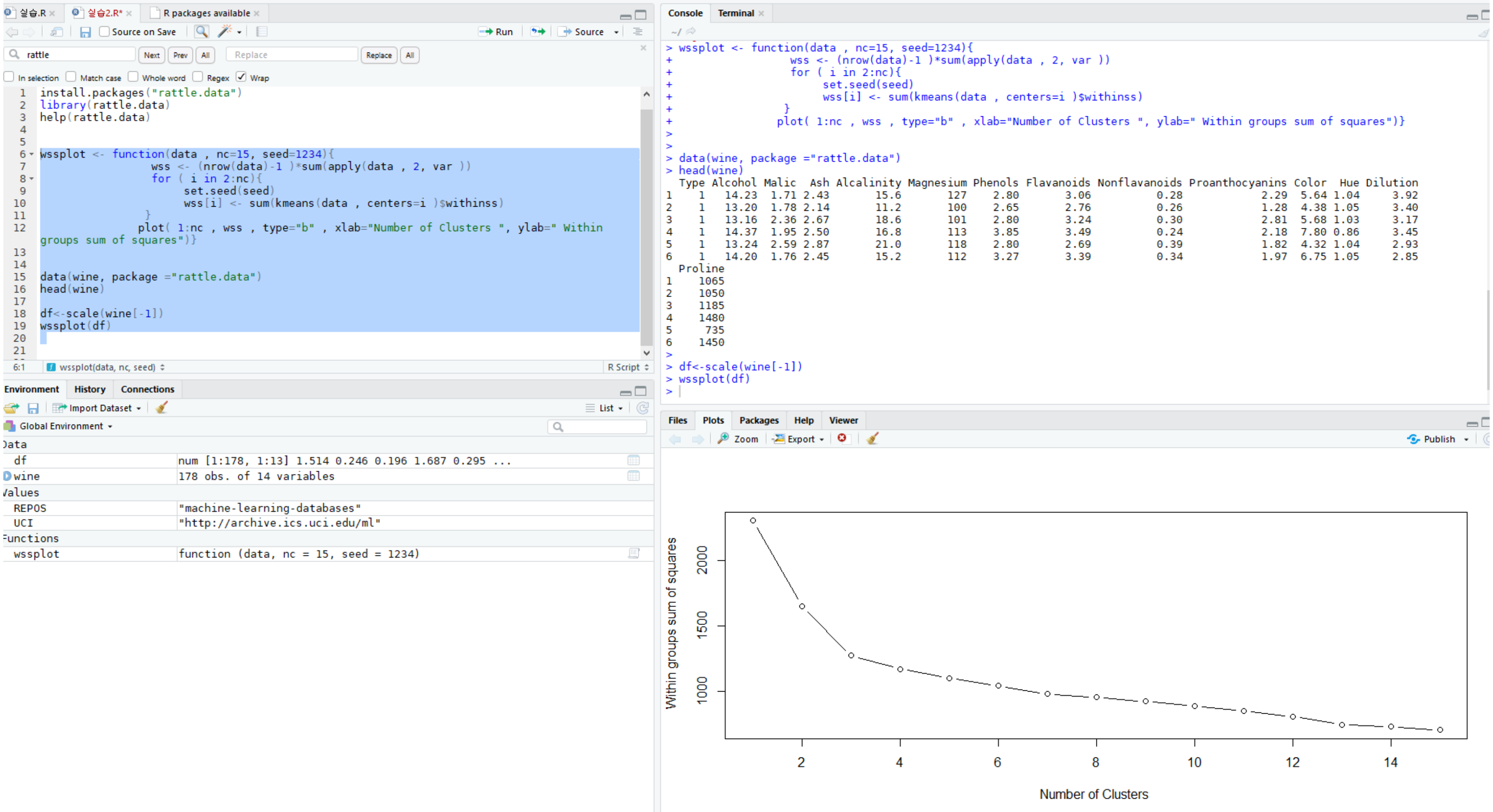
Agglomerative Coefficient = 0.95



```r
data(USArrests)
str(USArrests)

d <- dist(USArrests, method="euclidean")
fit <- hclust(d, method="ave")

par(mfrow = c(1,2))
plot(fit)
plot(fit, hang=-1)
par(mfrow=c(1,1))

groups <- cutree(fit, k=6)
groups

plot(fit)
rect.hclust(fit, k=6, border="red")


hca <- hclust(dist(USArrests))
plot(hca)
rect.hclust(hca, k=3, border="red")
rect.hclust(hca, h=50, which =c(2, 7), border=3:4)


library(cluster)
agn1<-agnes(USArrests, metric="manhattan", stand=TRUE)
agn1
par(mfrow = c(1,2))
plot(agn1)


agn2<-agnes(daisy(USArrests), diss=TRUE, method="complete")
plot(agn2)

agn3<-agnes(USArrests, method="flexible", par.meth=0.6)
plot(agn3)
par(mfrow = c(1,1))

```

```
[10] Maryland       New Mexico     Michigan       Illinois       New York       Texas          Florida        California     Colorado
[19] Nevada         Arkansas       Idaho          Nebraska       Kentucky       Montana        Indiana        Kansas         Oklahoma
[28] Ohio           Pennsylvania   Virginia       Wyoming        Delaware       Missouri       Oregon         Washington     Connecticut
[37] Utah           Hawaii         Massachusetts  New Jersey     Rhode Island   Iowa           New Hampshire  Maine          Minnesota
[46] Wisconsin      North Dakota   Vermont        South Dakota   West Virginia
Height (summary):
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.3718  1.5669  2.0341  2.3766  2.9198  7.3157 

Available components:
[1] "order"     "height"    "ac"        "merge"     "diss"      "call"      "method"    "order.lab" "data"     
> par(mfrow = c(1,2))
> plot(agn1)
> 
> 
> agn2<-agnes(daisy(USArrests), diss=TRUE, method="complete")
> plot(agn2)
> agn3<-agnes(USArrests, method="flexible", par.meth=0.6)
> plot(agn3)
> par(mfrow = c(1,1))
> |
```

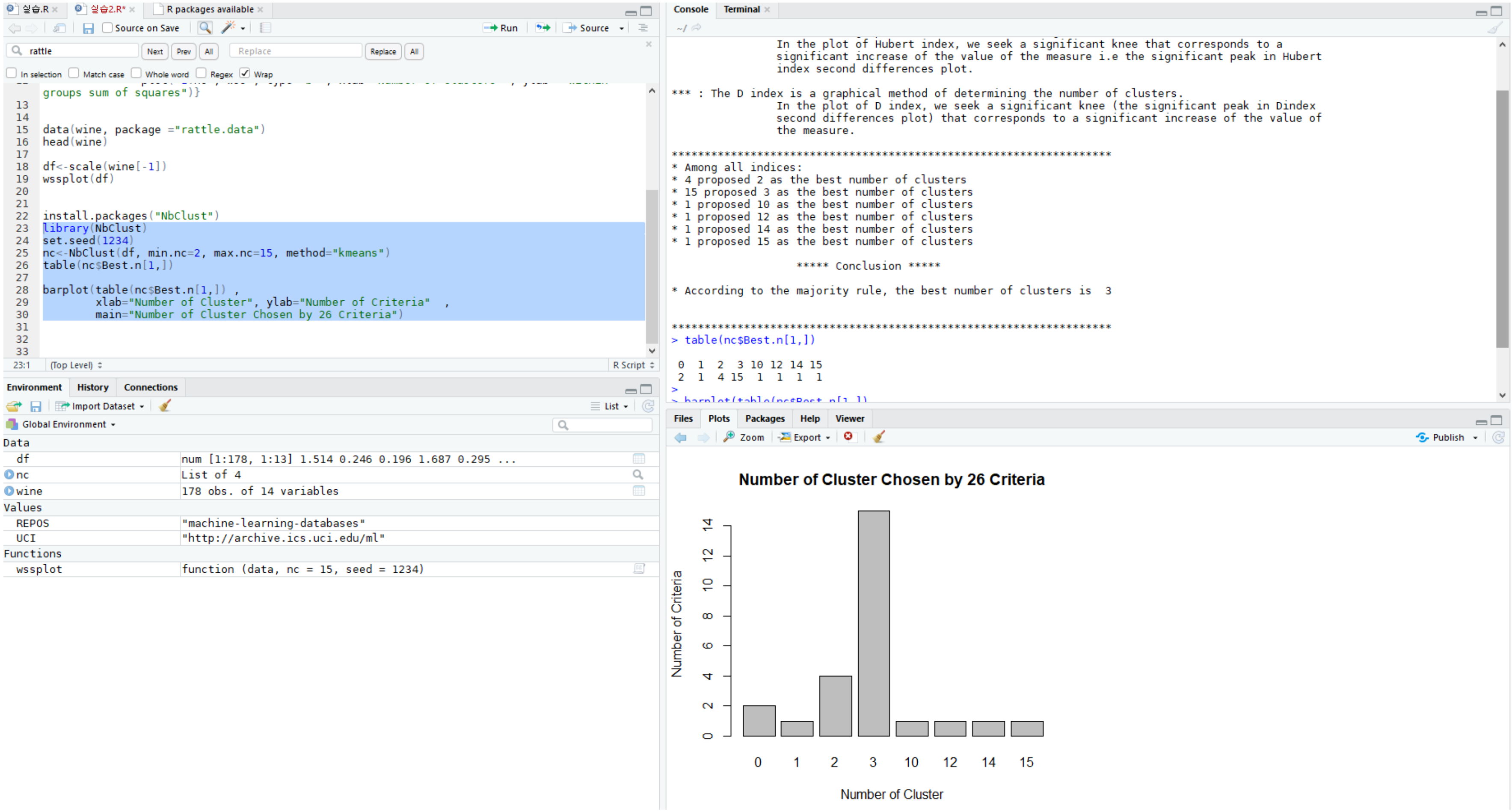
```
Data
 agn1        List of 9
 agn2        List of 8
 agn3        List of 9
 fit         List of 7
 hca         List of 7
 USArrests   50 obs. of 4 variables
Values
 d           'dist' num [1:1225] 37.2 63 46.9 55.5 41.9 ...
 groups      Named int [1:50] 1 1 1 2 1 2 3 1 4 2 ...
```



```r
install.packages("rattle.data")
library(rattle.data)
help(rattle.data)


wssplot <- function(data , nc=15, seed=1234){
                wss <- (nrow(data)-1 )*sum(apply(data , 2, var ))
                for ( i in 2:nc){
                     set.seed(seed)
                     wss[i] <- sum(kmeans(data , centers=i )$withinss)
                }
              plot( 1:nc , wss , type="b" , xlab="Number of Clusters ", ylab=" Within
groups sum of squares")}


data(wine, package ="rattle.data")
head(wine)

df<-scale(wine[-1])
wssplot(df)
```

```
> wssplot <- function(data , nc=15, seed=1234){
+                 wss <- (nrow(data)-1 )*sum(apply(data , 2, var ))
+                 for ( i in 2:nc){
+                      set.seed(seed)
+                      wss[i] <- sum(kmeans(data , centers=i )$withinss)
+                 }
+               plot( 1:nc , wss , type="b" , xlab="Number of Clusters ", ylab=" Within groups sum of squares")}
>
>
> data(wine, package ="rattle.data")
> head(wine)
  Type Alcohol Malic  Ash Alcalinity Magnesium Phenols Flavanoids Nonflavanoids Proanthocyanins Color  Hue Dilution
1    1   14.23  1.71 2.43       15.6       127    2.80       3.06          0.28            2.29  5.64 1.04     3.92
2    1   13.20  1.78 2.14       11.2       100    2.65       2.76          0.26            1.28  4.38 1.05     3.40
3    1   13.16  2.36 2.67       18.6       101    2.80       3.24          0.30            2.81  5.68 1.03     3.17
4    1   14.37  1.95 2.50       16.8       113    3.85       3.49          0.24            2.18  7.80 0.86     3.45
5    1   13.24  2.59 2.87       21.0       118    2.80       2.69          0.39            1.82  4.32 1.04     2.93
6    1   14.20  1.76 2.45       15.2       112    3.27       3.39          0.34            1.97  6.75 1.05     2.85
  Proline
1    1065
2    1050
3    1185
4    1480
5     735
6    1450
>
> df<-scale(wine[-1])
> wssplot(df)
>
```

```
Data
  df          num [1:178, 1:13] 1.514 0.246 0.196 1.687 0.295 ...
  wine        178 obs. of 14 variables
Values
  REPOS       "machine-learning-databases"
  UCI         "http://archive.ics.uci.edu/ml"
Functions
  wssplot     function (data, nc = 15, seed = 1234)
```



```r
groups sum of squares")}


data(wine, package ="rattle.data")
head(wine)

df<-scale(wine[-1])
wssplot(df)


install.packages("NbClust")
library(NbClust)
set.seed(1234)
nc<-NbClust(df, min.nc=2, max.nc=15, method="kmeans")
table(nc$Best.n[1,])

barplot(table(nc$Best.n[1,]) ,
        xlab="Number of Cluster", ylab="Number of Criteria"  ,
        main="Number of Cluster Chosen by 26 Criteria")
```

```
                In the plot of Hubert index, we seek a significant knee that corresponds to a
                significant increase of the value of the measure i.e the significant peak in Hubert
                index second differences plot.

*** : The D index is a graphical method of determining the number of clusters.
                In the plot of D index, we seek a significant knee (the significant peak in Dindex
                second differences plot) that corresponds to a significant increase of the value of
                the measure.

*******************************************************************
* Among all indices:
* 4 proposed 2 as the best number of clusters
* 15 proposed 3 as the best number of clusters
* 1 proposed 10 as the best number of clusters
* 1 proposed 12 as the best number of clusters
* 1 proposed 14 as the best number of clusters
* 1 proposed 15 as the best number of clusters

                   ***** Conclusion *****

* According to the majority rule, the best number of clusters is  3


*******************************************************************
> table(nc$Best.n[1,])

 0  1  2  3 10 12 14 15
 2  1  4 15  1  1  1  1
>
> barplot(table(nc$Best.n[1,])
```

Data

df	num [1:178, 1:13] 1.514 0.246 0.196 1.687 0.295 ...
nc	List of 4
wine	178 obs. of 14 variables

Values

REPOS	"machine-learning-databases"
UCI	"http://archive.ics.uci.edu/ml"

Functions

wssplot	function (data, nc = 15, seed = 1234)

Number of Cluster Chosen by 26 Criteria