

# hw7\_Clustering

201511646\_

2017 11 27

## HW7\_Clustering\_201511646

1. In this example, we have distances between ten American cities based on the flying mileages between them. The objective is to see if we can define clusters of these cities based on the distances.

```
Atlanta<-c(0,587,1212,701,1936,604,748,2139,2182,543)
Chicago<-c(587,0,920,940,1745,1188,713,1858,1737,597)
Denver<-c(1212,920,0,879,831,1726,1631,949,1021,1494)
Houston<-c(701,940,879,0,1374,968,1420,1645,1891,1220)
LA<-c(1936,1745,831,1374,0,2339,2451,347,959,2300)
Miami<-c(604,1188,1726,968,2339,0,1092,2594,2734,923)
Newyork<-c(748,713,1631,1420,2451,1092,0,2571,2408,205)
Sanfrancisco<-c(2139,1858,949,1645,347,2594,2571,0,678,2442)
Seattle<-c(2182,1737,1021,1891,959,2734,2408,678,0,2329)
WashingtonDC<-c(543,597,1494,1220,2300,923,205,2442,2329,0)
```

```
cities<-matrix(c(Atlanta,Chicago,Denver, Houston,LA, Miami, Newyork,Sanfrancisco, Seattle, WashingtonDC),
cities
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    0  587 1212  701 1936  604  748 2139 2182   543
## [2,]  587    0  920  940 1745 1188  713 1858 1737   597
## [3,] 1212  920    0  879  831 1726 1631  949 1021  1494
## [4,]  701  940  879    0 1374  968 1420 1645 1891  1220
## [5,] 1936 1745  831 1374    0 2339 2451  347  959  2300
## [6,]  604 1188 1726  968 2339    0 1092 2594 2734   923
## [7,]  748  713 1631 1420 2451 1092    0 2571 2408   205
## [8,] 2139 1858  949 1645  347 2594 2571    0  678  2442
## [9,] 2182 1737 1021 1891  959 2734 2408  678    0  2329
## [10,]  543  597 1494 1220 2300  923  205 2442 2329    0
```

```
Tcities=t(cities)
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

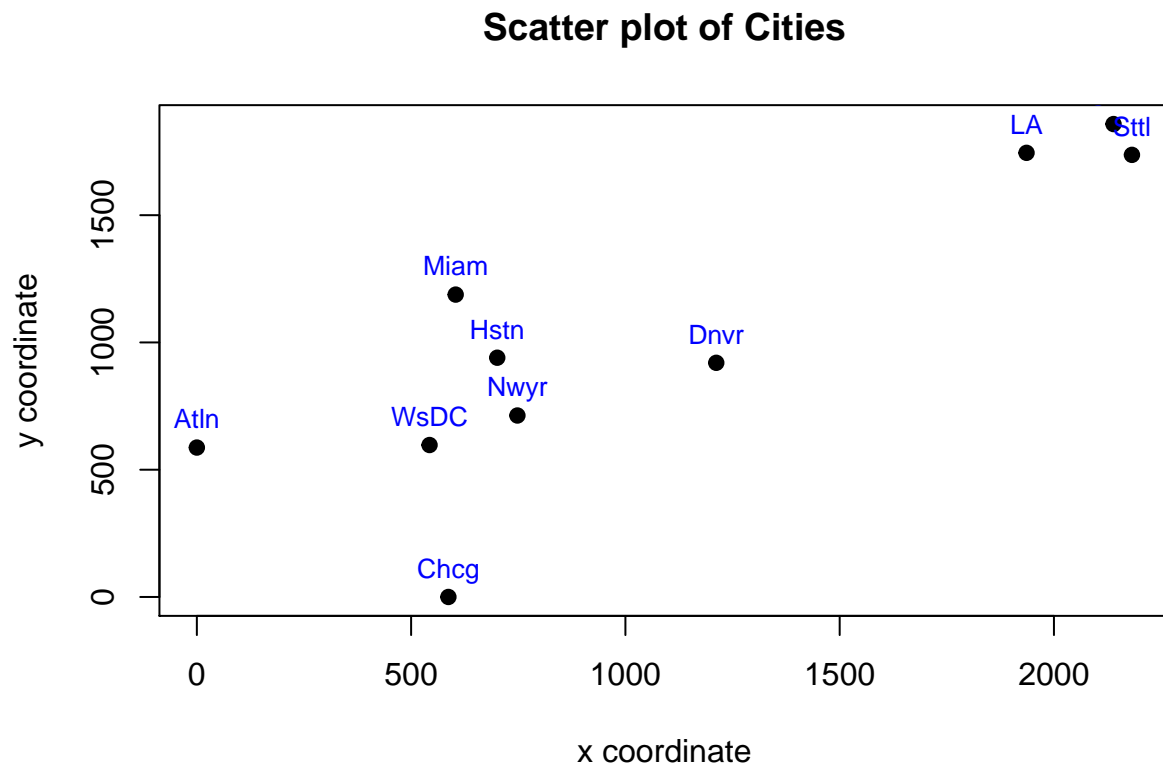
```
##      intersect, setdiff, setequal, union
```

```
all_equal(cities, Tcities)
```

```
## [1] TRUE
```

```
scatter plot
```

```
colnames(cities)<-c("Atlanta","Chicago","Denver", "Houston","LA", "Miami", "Newyork","Sanfrancisco", "S  
rownames(cities)<-c("Atlanta","Chicago","Denver", "Houston","LA", "Miami", "Newyork","Sanfrancisco", "S  
plot(cities,pch = 19, xlab = c("x coordinate"), ylab = c("y coordinate"),main = "Scatter plot of Cities  
text(cities, labels = abbreviate(colnames(cities)), cex = 0.8, pos = 3, col = "blue") # pos=1 : at the
```



- 1) Do the cluster analysis using (1) single linkage, (2) average linkage and (3) the centroid method.

(1)single linkage

```
hc1_cities<-hclust(dist(cities, method="euclidian")^2, method="single")  
hc1_cities
```

```
##
```

```
## Call:
```

```
## hclust(d = dist(cities, method = "euclidian")^2, method = "single")
```

```
##
```

```
## Cluster method : single
```

```
## Distance : euclidean
```

```
## Number of objects: 10
```

```
rev(hc1_cities) #height :      / merge:
```

```
## $dist.method
## [1] "euclidean"
##
## $call
## hclust(d = dist(cities, method = "euclidian")^2, method = "single")
##
## $method
## [1] "single"
##
## $labels
## [1] "Atlanta"      "Chicago"      "Denver"      "Houston"
## [5] "LA"           "Miami"        "Newyork"     "Sanfrancisco"
## [9] "Seattle"      "WashingtonDC"
##
## $order
## [1] 9 5 8 3 4 6 7 10 1 2
##
## $height
## [1] 272544 560711 1435040 1490187 1581207 2363192 2905270 4037129 4895001
##
## $merge
##      [,1] [,2]
## [1,]  -7  -10
## [2,]  -5   -8
## [3,]  -9    2
## [4,]  -1   -2
## [5,]   1    4
## [6,]  -6    5
## [7,]  -4    6
## [8,]  -3    7
## [9,]   3    8
```