

Clustering Example 4: Hierarchical Clustering of Wine Data

Amy Wagaman, Nathan Carter

July 2020

Load necessary libraries.

```
library(mosaic)
library(cluster)
```

This requires you to have access to the `winedata.txt` file. It is available in the book's GitHub repository at the following URL.

<https://github.com/ds4m/ds4m.github.io/tree/master/chapter-5-resources/winedata.txt>

If you run this R code, place the data file in the same folder as the code file.

```
winedata <- read.csv("winedata.txt")
```

For summaries of the variables in the wine data, see the Clustering Example 3 file in the same folder as this one.

Compute a distance matrix for the wine data.

```
wine.dist.scale <- dist(scale(winedata[, -c(1)]))
```

Apply hierarchical clustering with single linkage.

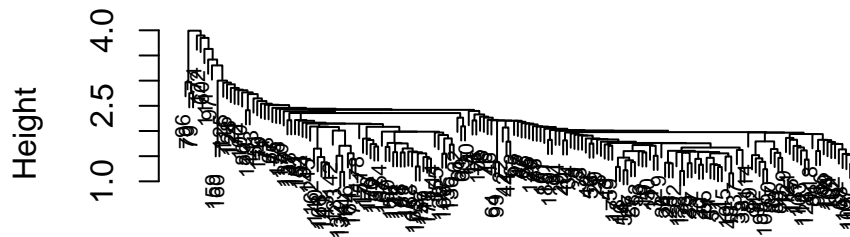
```
hcsingle <- hclust(wine.dist.scale, method = "single")
list(hcsingle) # reminds you of properties of the solution, if desired
```

```
## [[1]]
##
## Call:
## hclust(d = wine.dist.scale, method = "single")
##
## Cluster method   : single
## Distance         : euclidean
## Number of objects: 178
```

Plot the resulting dendrogram.

```
plot(hcsingle, cex = 0.7)
```

Cluster Dendrogram



```
wine.dist.scale
hclust (*, "single")
```

Extract a single clustering solution from the dendrogram.

Here we specify the height h that determines the clustering. We display just the number of observations in each cluster.

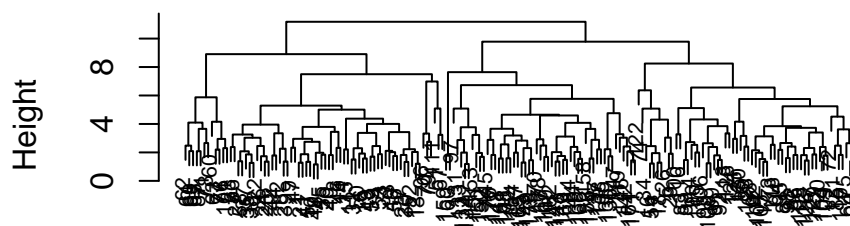
```
singleSol <- (cutree(hcsingle, h = 2.25)) #cluster labels are numeric, h= height
summary(as.factor(singleSol)) #as factor to get table
```

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
## 98  2  1  1  2  1  1  1  1  1  1  1  1  1  1  1  1 43  1  1  1  1  1  1  1  1
## 27 28 29 30 31 32 33 34 35
##  1  1  1  1  1  2  1  2  1
```

Apply hierarchical clustering with complete linkage and plot dendrogram.

```
hccomp <- hclust(wine.dist.scale, method = "complete")
plot(hccomp, cex = 0.7)
```

Cluster Dendrogram



```
wine.dist.scale
hclust (*, "complete")
```

Extract a single clustering solution from the dendrogram.

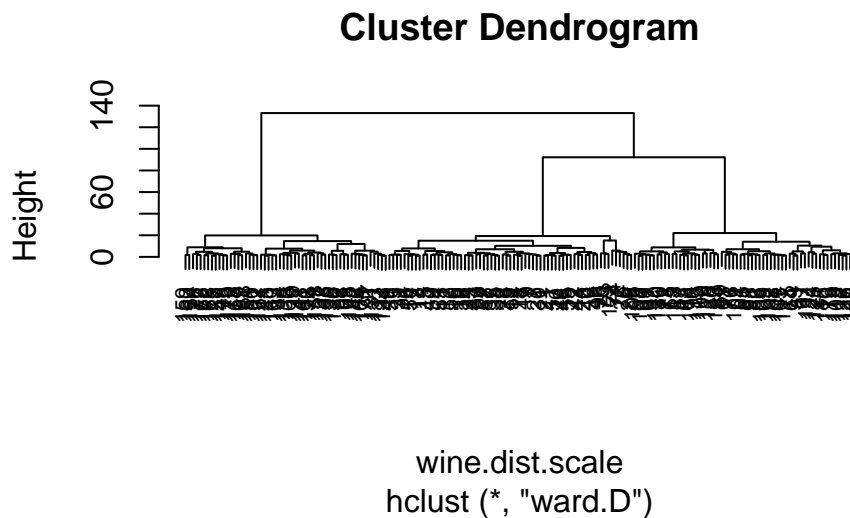
Here we specify the number k of clusters that we desire in the solution. We display just the number of observations in each cluster.

```
compSol <- (cutree(hccomp, k= 4)) #k= # clusters, k=3 also reasonable  
summary(as.factor(compSol)) #as factor to get table
```

```
## 1 2 3 4  
## 57 58 12 51
```

Apply hierarchical clustering with Ward's method and plot dendrogram.

```
hcward <- hclust(wine.dist.scale, method = "ward.D")  
plot(hcward, cex = 0.7)
```



Extract a single clustering solution from the dendrogram.

Here we specify the number k of clusters that we desire in the solution. We display just the number of observations in each cluster.

```
wardSol <- (cutree(hcward, k= 3)) #k= # clusters, k=3 also reasonable  
summary(as.factor(wardSol)) #as factor to get table
```

```
## 1 2 3  
## 65 59 54
```