**Implement Hierachical clustering (R / Orange )**

1. Single linkage
2. Complete linkage
3. Average Linkage

**Analysis Visualization:**

Visualized with hclust, geom\_point, scatterplot, heatmap, corrplot::corrplot.

**Dataset:**

**Play.csv (Cluster of paly depend on different weather);**

**Iris.csv (3 Species clustering depend on features);**

**Animal.csv (Animal clustering depend on couple features).**

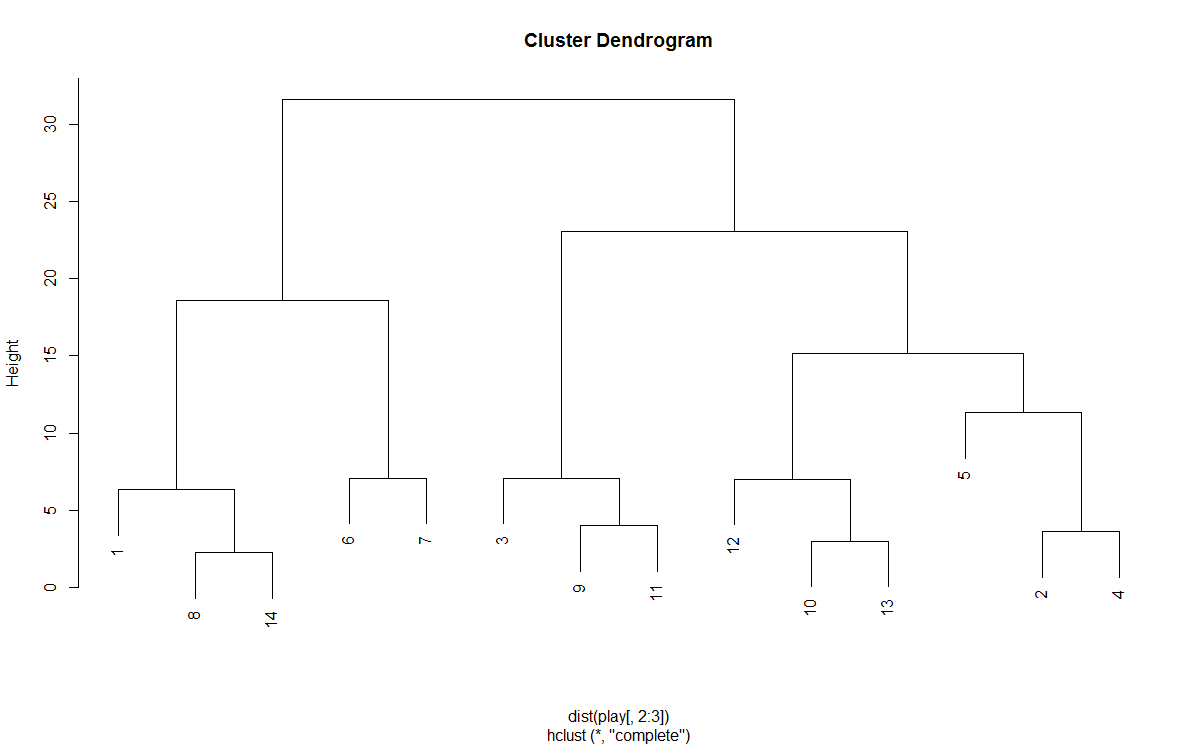
**Play clustering**

> play <- readWorksheetFromFile("C:/Users/Wanwan Zhang/Desktop/2016FALL/ADS/4/play.xlsx", sheet = 1)

> # complete linkage

> clusters <- hclust(dist(play[, 2:3]))

> plot(clusters)



> clusterCut <- cutree(clusters, 2)

> table(clusterCut, play$Decision)

clusterCut dont\_play play

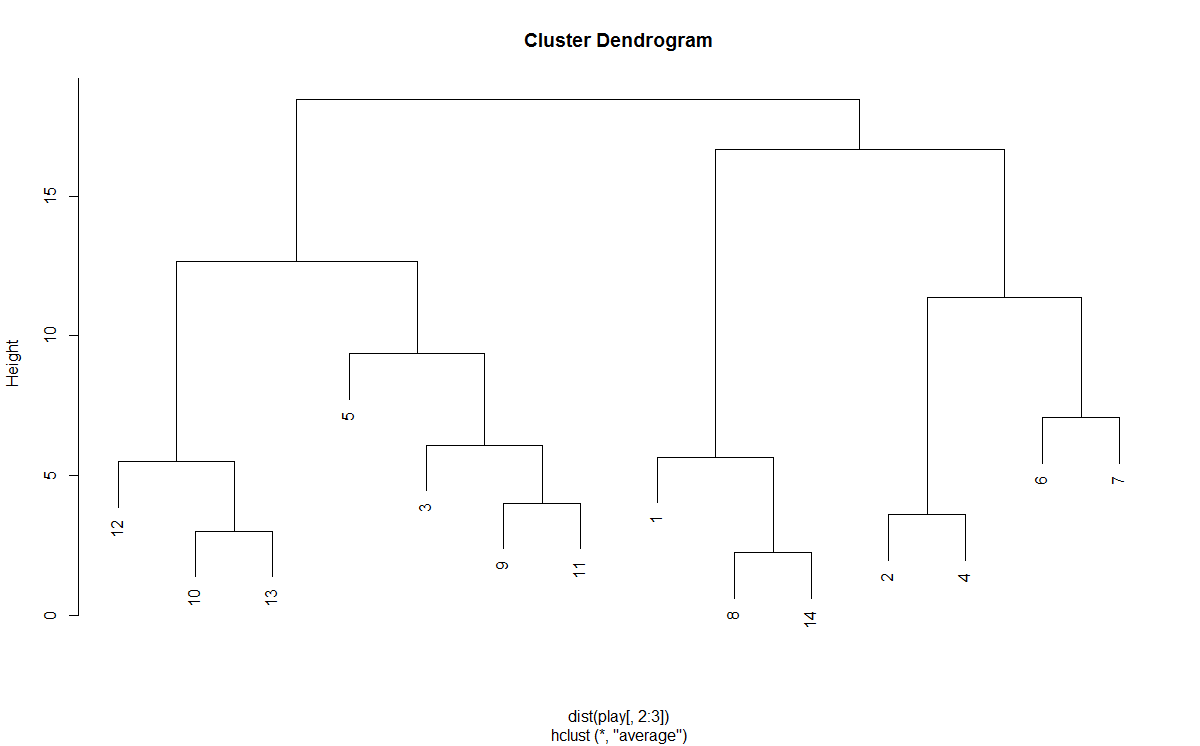
1 3 2

2 2 7

> # mean/average linkage

> clusters <- hclust(dist(play[, 2:3]), method = 'average')

> plot(clusters)



> clusterCut <- cutree(clusters, 2)

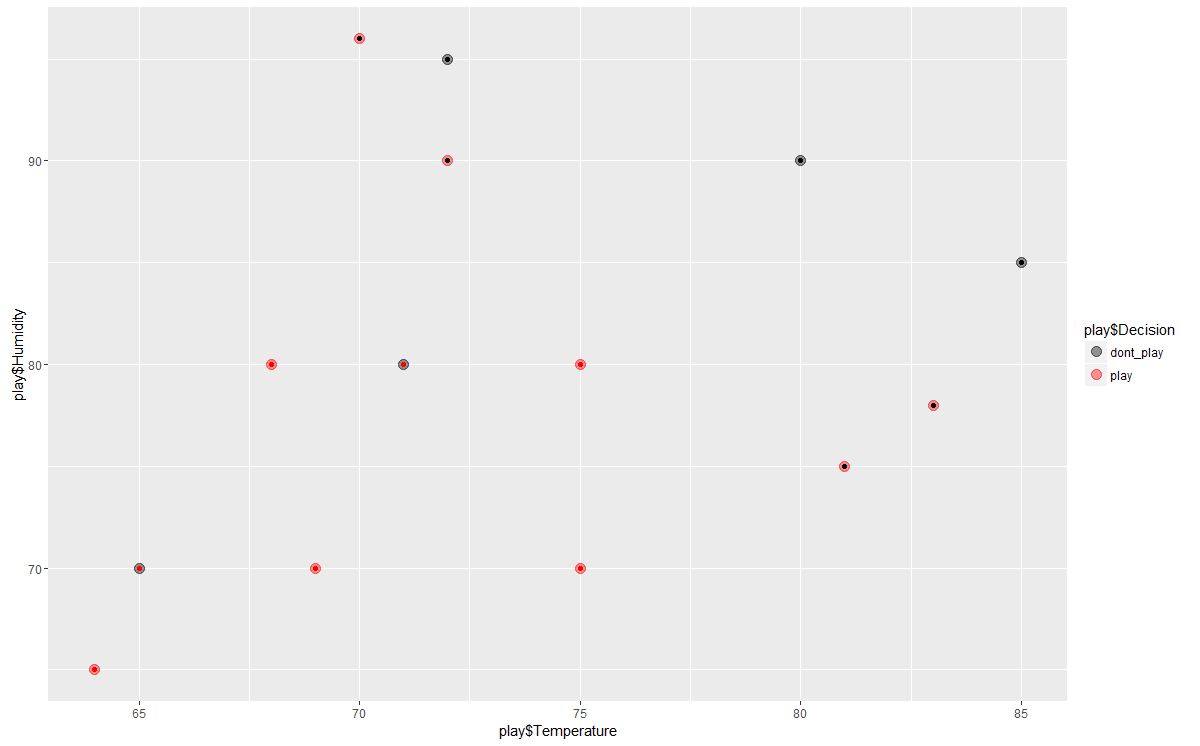
> table(clusterCut, play$Decision)

clusterCut dont\_play play

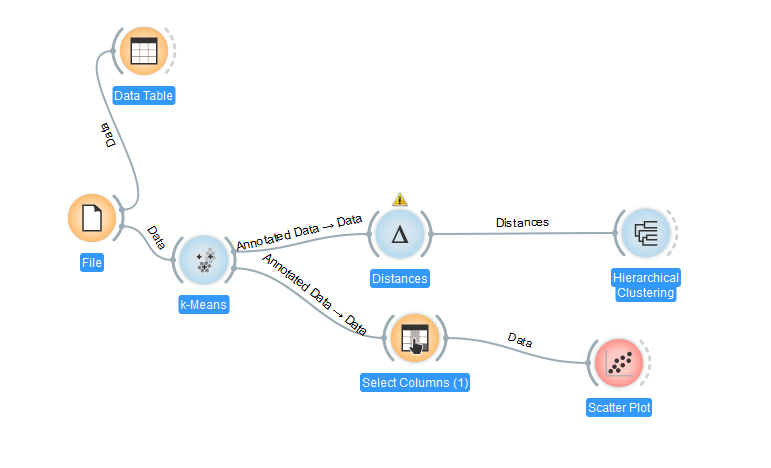
1 3 4

2 2 5

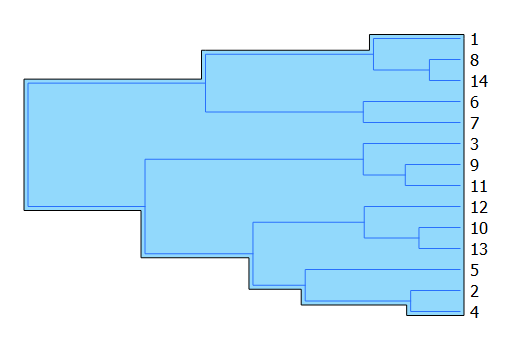
|  |
| --- |
| > ggplot(play, aes(play$Temperature, play$Humidity, color = play$Decision)) +  + geom\_point(alpha = 0.4, size = 3.5) + geom\_point(col = clusterCut) +  + scale\_color\_manual(values = c('black', 'red', 'green')) |
|  |
|  |



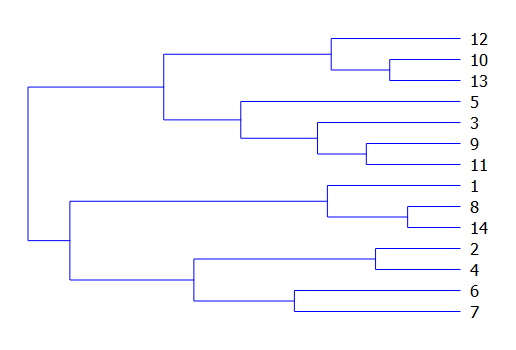
Implement in Orange：



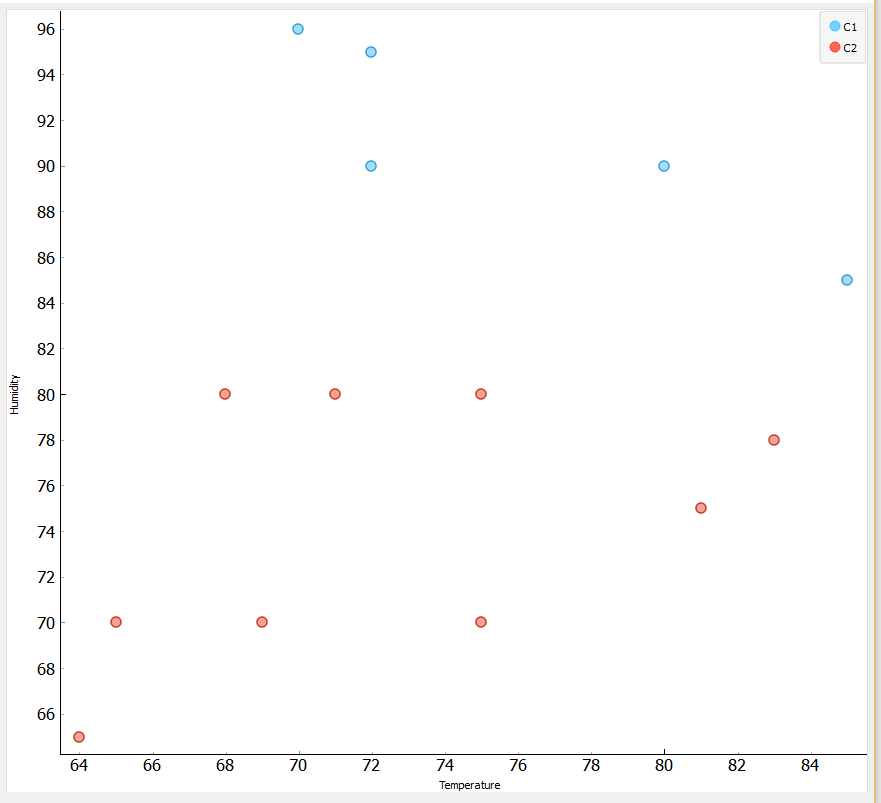
Complete linkage:



Average Linkage:



K-means to set the 2 clusters and get the scatterplot to show the different features and number of each cluster.



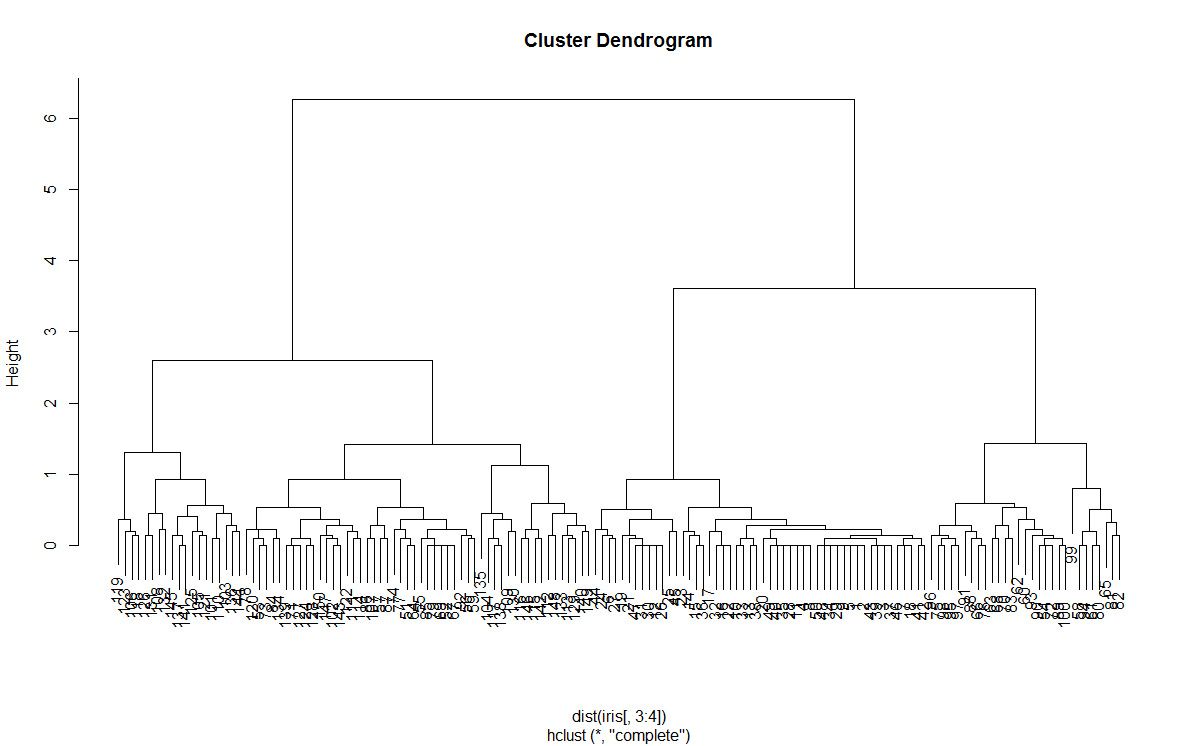
**Iris hierarchical clustering**

> library(ggplot2)

> # complete linkage

> clusters <- hclust(dist(iris[, 3:4]))

> plot(clusters)



> clusterCut <- cutree(clusters, 3)

> table(clusterCut, iris$Species)

clusterCut setosa versicolor virginica

1 50 0 0

2 0 21 50

3 0 29 0

> # mean/average linkage

> clusters <- hclust(dist(iris[, 3:4]), method = 'average')

> plot(clusters)

> clusterCut <- cutree(clusters, 3)

> table(clusterCut, iris$Species)

clusterCut setosa versicolor virginica

1 50 0 0

2 0 45 1

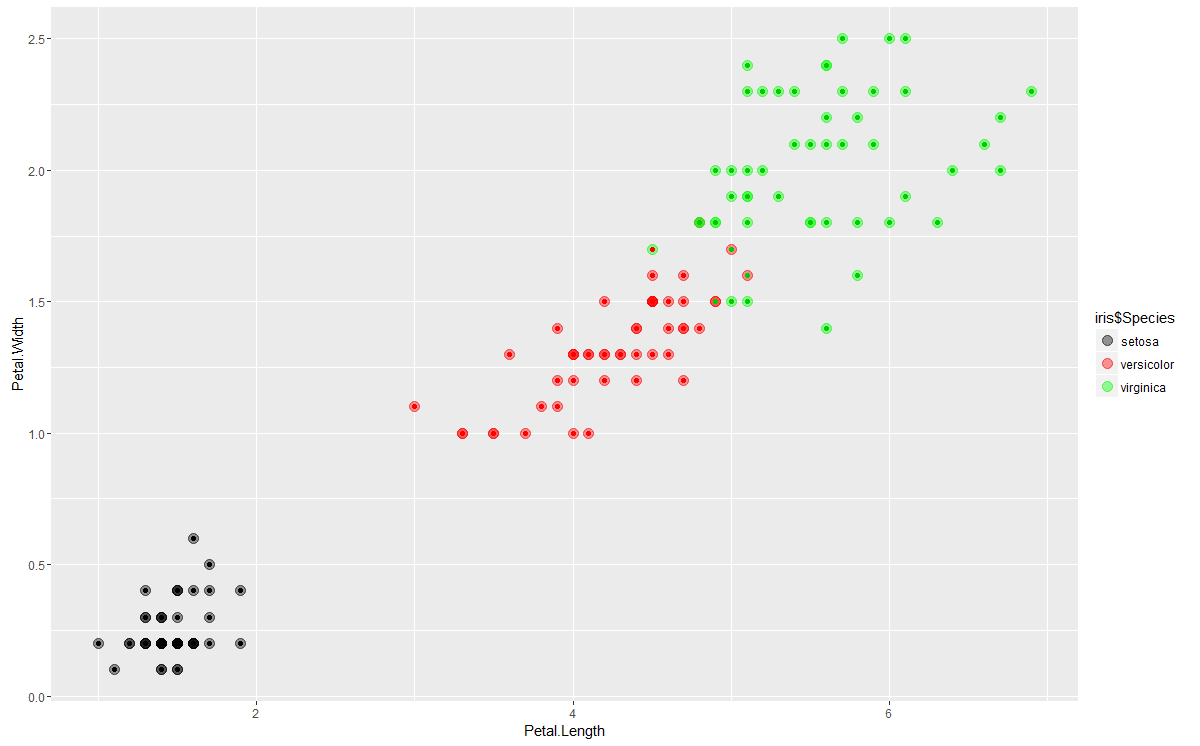
3 0 5 49

> ggplot(iris, aes(Petal.Length, Petal.Width, color = iris$Species)) +

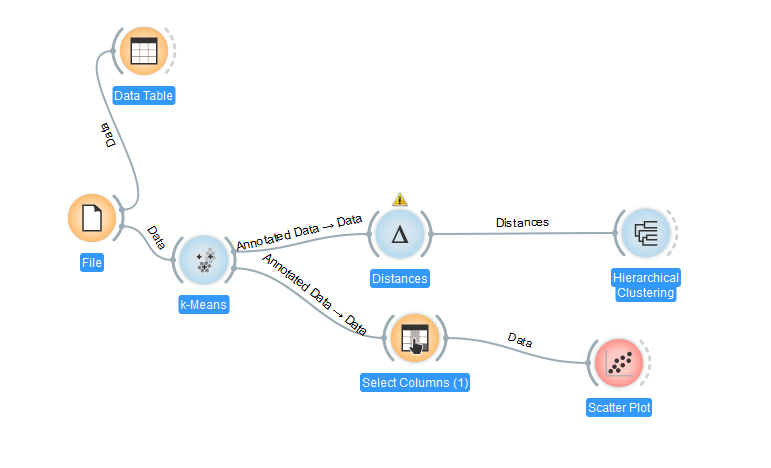
+ geom\_point(alpha = 0.4, size = 3.5) + geom\_point(col = clusterCut) +

+ scale\_color\_manual(values = c('black', 'red', 'green'))

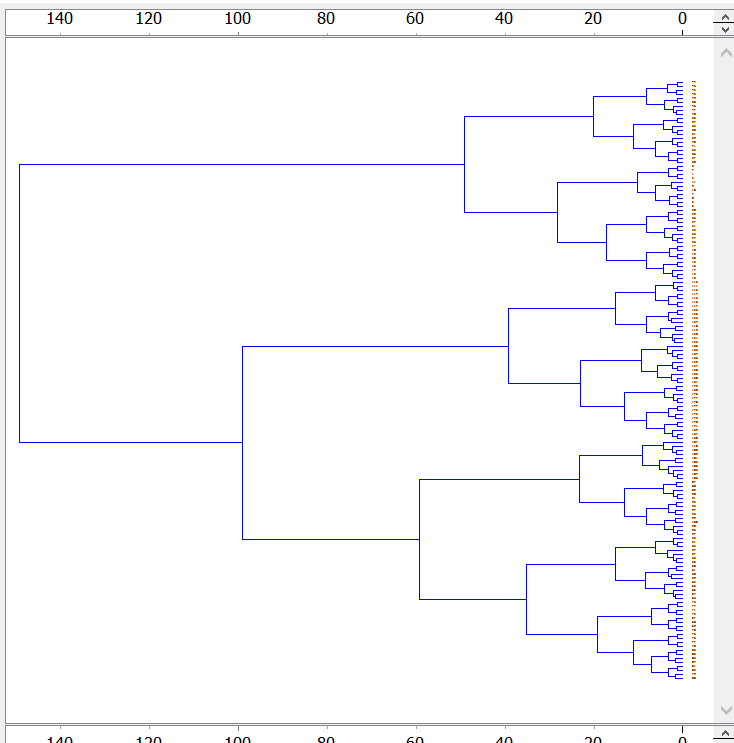
>



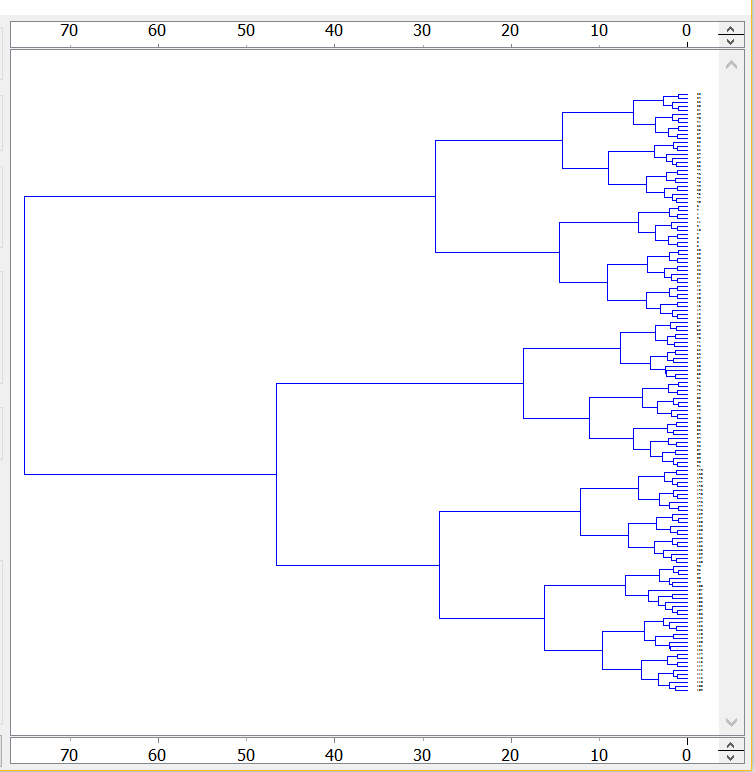
Implement in Orange :



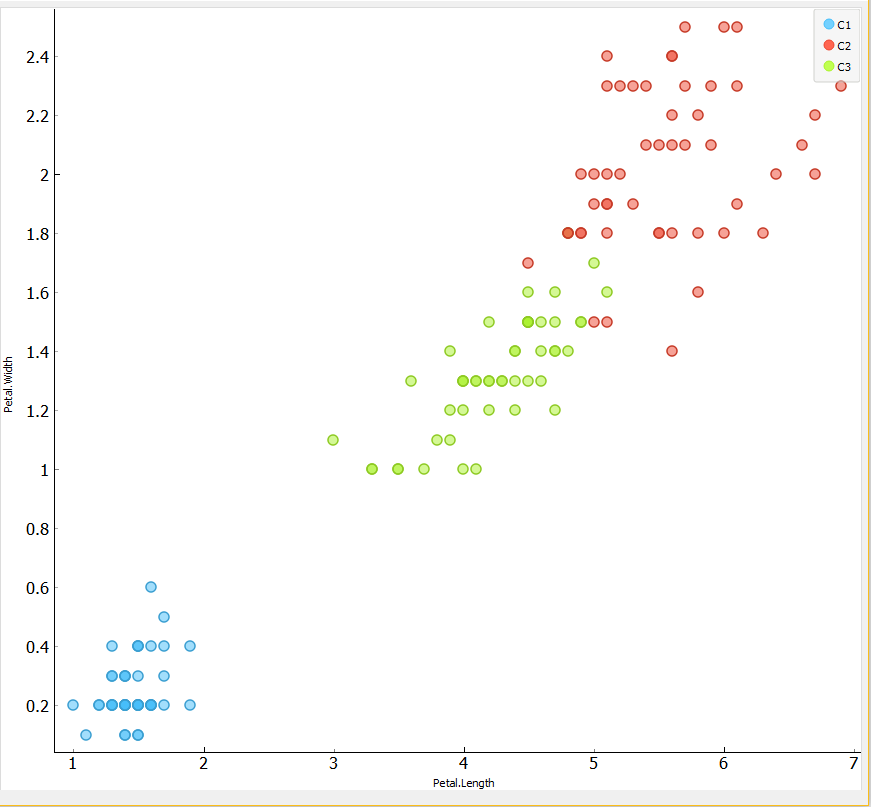
Complete linkage:



Average linkage:



K-means to set the 2 clusters and get the scatterplot to show the different features and number of each cluster.



**Animal hierarchical clustering**

**animals---attributes-of-animals**

> animals <- cluster::animals

>

> colnames(animals) <- c("warm-blooded",

+ "can fly",

+ "vertebrate",

+ "endangered",

+ "live in groups",

+ "have hair")

>

> dend\_r <- animals %>% dist(method = "man") %>% hclust(method = "ward.D") %>% as.dendrogram %>% ladderize %>%

+ color\_branches(k=4)

>

> dend\_c <- t(animals) %>% dist(method = "man") %>% hclust(method = "com") %>% as.dendrogram %>% ladderize%>%

+ color\_branches(k=3)

>

>

> # some\_col\_func <- function(n) rev(colorspace::heat\_hcl(n, c = c(80, 30), l = c(30, 90), power = c(1/5, 1.5)))

> # some\_col\_func <- colorspace::diverge\_hcl

> # some\_col\_func <- colorspace::sequential\_hcl

> some\_col\_func <- function(n) (colorspace::diverge\_hcl(n, h = c(246, 40), c = 96, l = c(65, 90)))

>

>

> # par(mar = c(3,3,3,3))

> # library(gplots)

> gplots::heatmap.2(as.matrix(animals-1),

+ main = "Attributes of Animals",

+ srtCol = 35,

+ Rowv = animals,

+ Colv = t(animals),

+ trace="row", hline = NA, tracecol = "darkgrey",

+ margins =c(6,3),

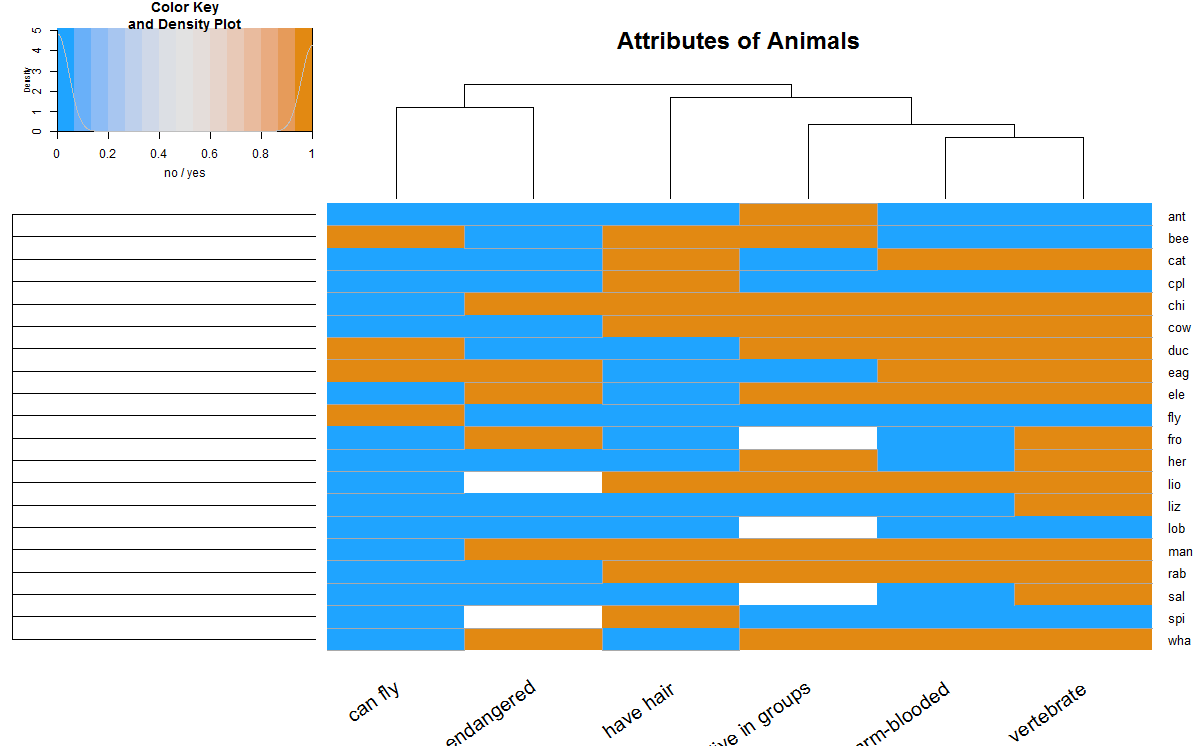
+ key.xlab = "no / yes",

+ denscol = "grey",

+ density.info = "density",

+ col = some\_col\_func

+ )



> hclust\_methods <- c("ward.D", "single", "complete", "average", "mcquitty",

+ "median", "centroid", "ward.D2")

> animals\_dendlist <- dendlist()

> for(i in seq\_along(hclust\_methods)) {

+ tmp\_dend <- animals %>% dist(method = "man") %>%

+ hclust(method = hclust\_methods[i]) %>% as.dendrogram

+ animals\_dendlist <- dendlist(animals\_dendlist, tmp\_dend)

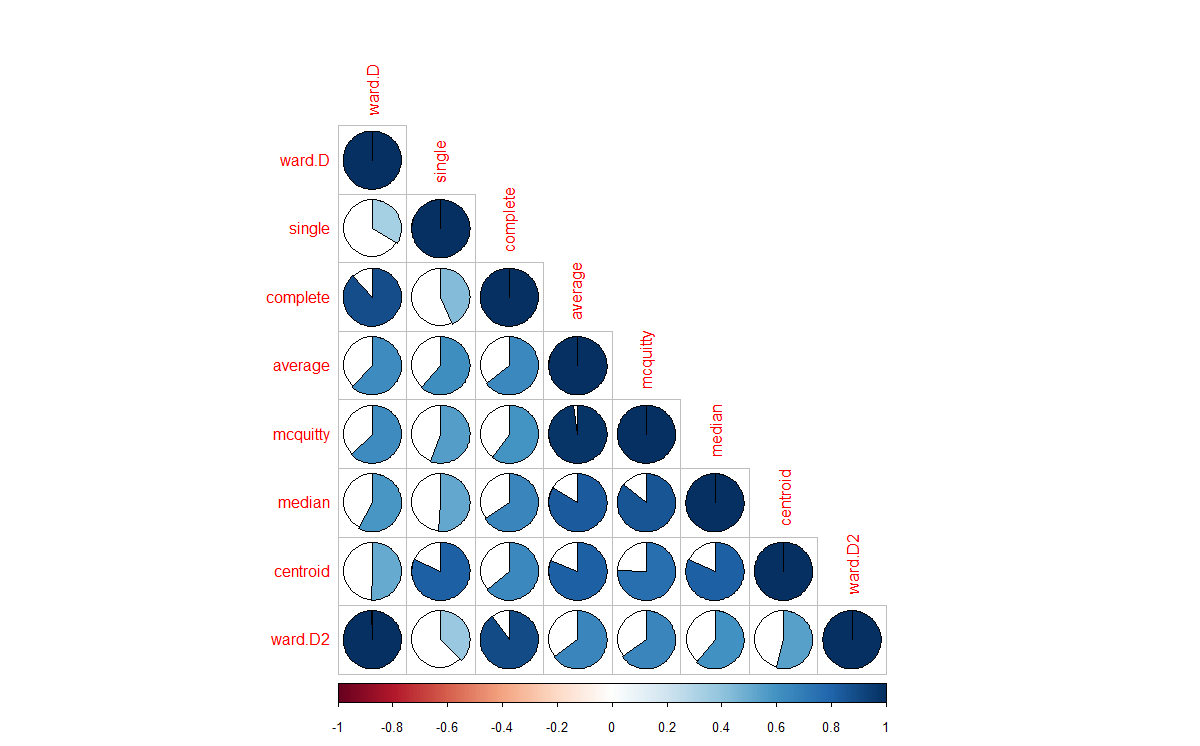
+ }

> names(animals\_dendlist) <- hclust\_methods

> # votes.repub\_dendlist

> cophenetic\_cors <- cor.dendlist(animals\_dendlist)

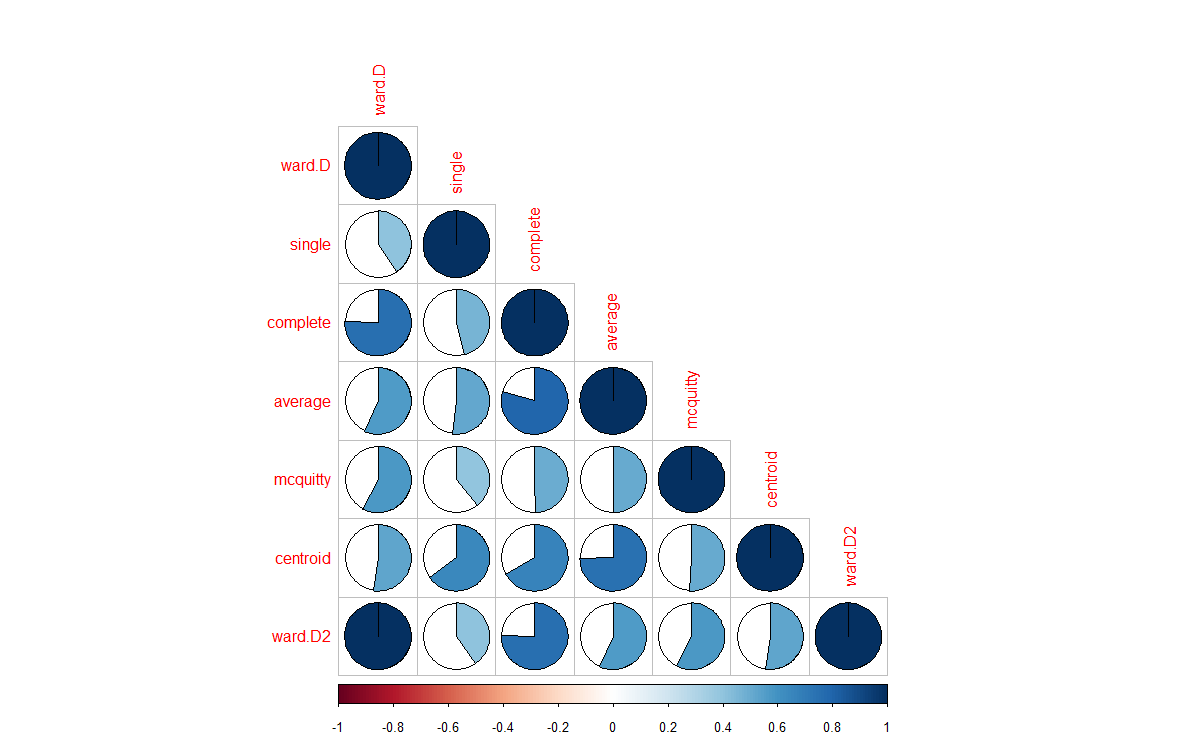
> corrplot::corrplot(cophenetic\_cors, "pie", "lower")



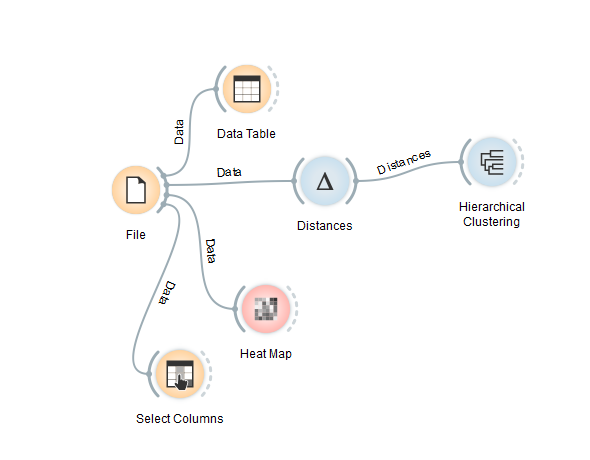
> remove\_median <- dendlist(animals\_dendlist, which = c(1:8)[-6] )

> FM\_cors <- cor.dendlist(remove\_median, method = "FM\_index", k = 4)

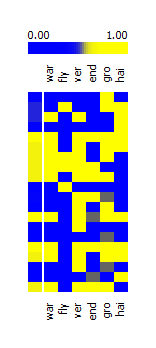
> corrplot::corrplot(FM\_cors, "pie", "lower")



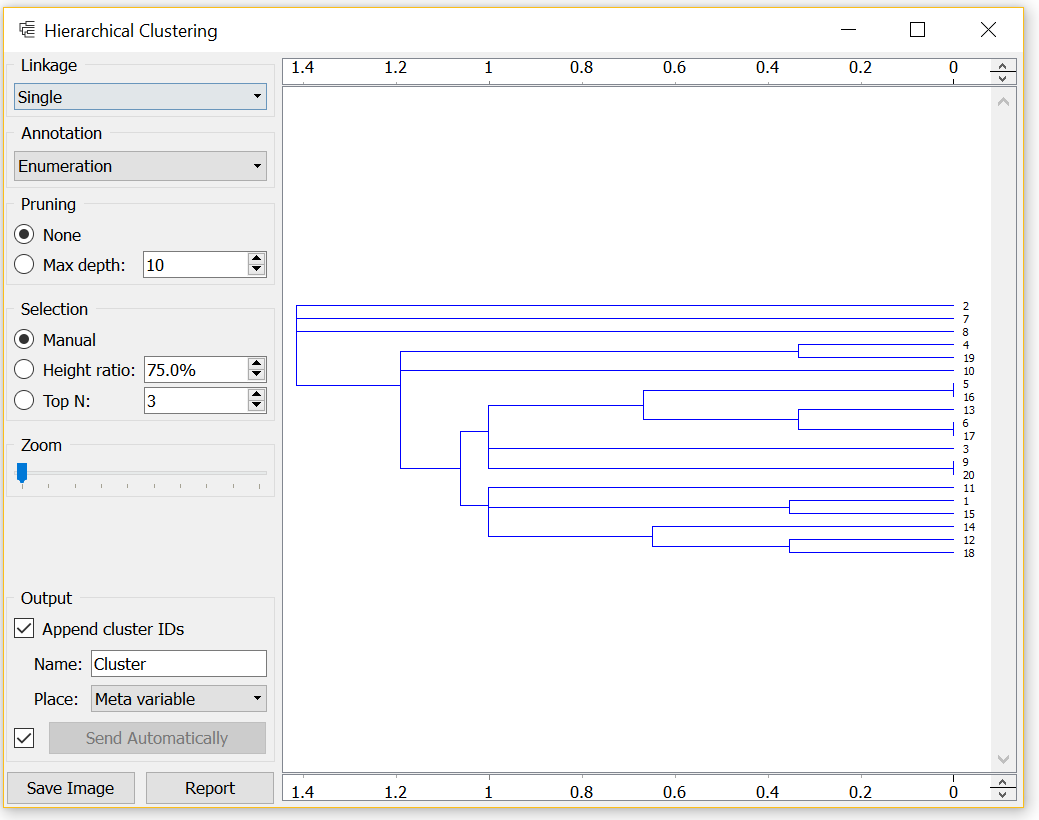
Implementing in Orange



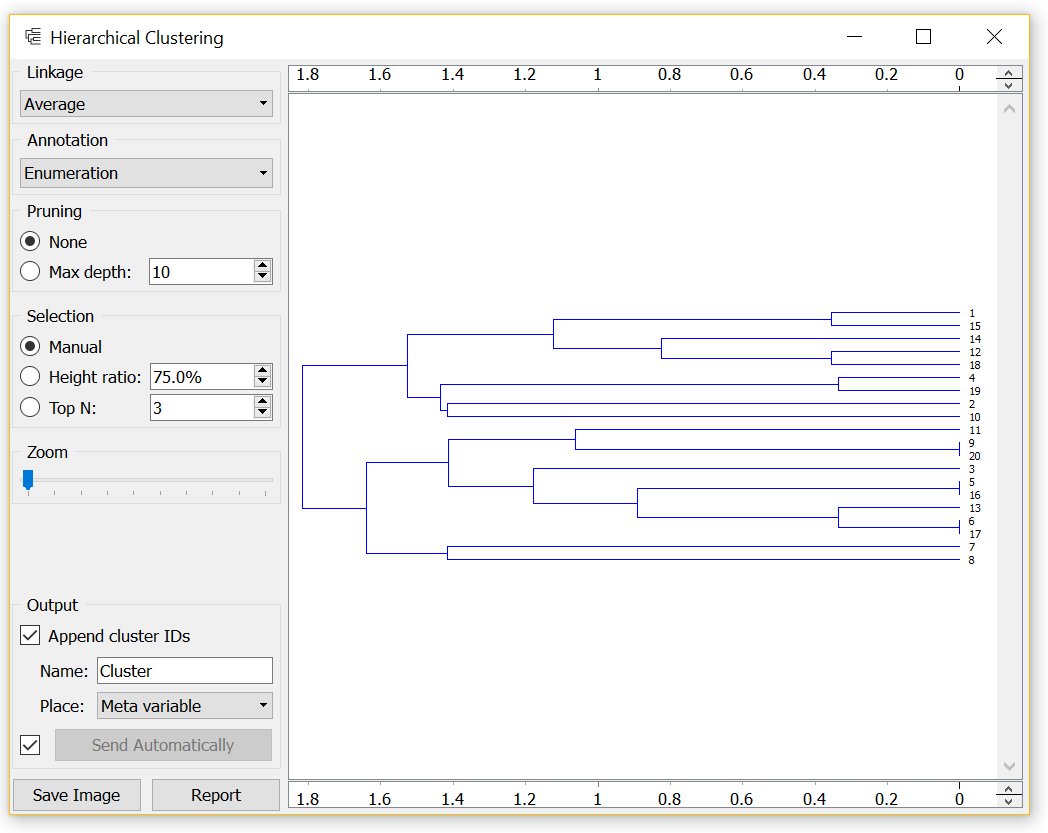
Get colorspace heatmap



Method:Single



Average linkage：



Complete linkage

