**Homework 9**

**Yuan Ling Chen**

Use the dominant copepod species data (from HW1).

Perform cluster analysis of stations based on percent composition data of the dominant species, and tell your story about these copepod data.

Tasks:

Compare the results based on different distance measures.

Compare the results based on different cluster algorithms.

Determine the final number of clusters and describe the differences among them.

You are allowed to use any built-in functions in Matlab or R, such as kmeans.m, clusterdata.m, etc.

copepod\_data <- read.table("copepod\_composition.txt", header = TRUE)  
cop\_density <- read.table("cop\_density.txt", header = TRUE)  
species\_list <- read.table("copepodSPlist.txt", fill=T, sep = "\n")  
  
row.names(copepod\_data) <- species\_list$V1  
  
  
# Prepare an empty matrix to store density data for all species  
all\_species\_density <- matrix(0, nrow = nrow(copepod\_data), ncol = ncol(copepod\_data))  
row.names(all\_species\_density) <- row.names(copepod\_data)  
colnames(all\_species\_density) <- colnames(copepod\_data)  
  
# Loop through each species (row) and each station/sample (column) to calculate density  
for (i in 1:nrow(copepod\_data)) {  
 for (j in 1:ncol(copepod\_data)) {  
 # Species proportion × corresponding total density at that station  
 all\_species\_density[i, j] <- copepod\_data[i, j] / 100 \* cop\_density[j, 1]  
 }  
}  
  
#find dominant species  
# Set threshold for dominant species (5%)  
threshold <- 5 # 5%  
  
# Find dominant species based on the condition  
dominant\_species <- rownames(all\_species\_density)[apply(all\_species\_density, 1, function(x) any(x >= threshold))]  
  
# Display the dominant species  
dominant\_species

## [1] "Acartia bifilosa"   
## [2] NA   
## [3] "Acartia negligence"   
## [4] "Acartia omori"   
## [5] "Acartia pacifica"   
## [6] "Acartia sp"   
## [7] NA   
## [8] NA   
## [9] NA   
## [10] NA   
## [11] "Metacalanus auvivillii"   
## [12] NA   
## [13] "Calanoides carinatus"   
## [14] "Calanus sinicus"   
## [15] "Canthocalanus pauper"   
## [16] "Cosmocalanus darwinii"   
## [17] "Mesocalanus tenuicornis"   
## [18] "Nannocalanus minor"   
## [19] "Neocalanus gracilis"   
## [20] "Undinula vulgaris"   
## [21] "Candacia bipinnata"   
## [22] NA   
## [23] NA   
## [24] "Candacia discaudata"   
## [25] "Candacia elongata"   
## [26] NA   
## [27] NA   
## [28] NA   
## [29] "Candaciidae copepodid"   
## [30] "Centropages furcatus"   
## [31] "Centropages orsinii"   
## [32] "Centropages copepodid"   
## [33] "Clausocalanus arcuicornis"   
## [34] "Clausocalanus brevipes"   
## [35] "Clausocalanus furcatus"   
## [36] "Clausocalanus ingens"   
## [37] NA   
## [38] NA   
## [39] "Clausocalanus mastigophorus"   
## [40] "Clausocalanus minor"   
## [41] NA   
## [42] "Clausocalanus pergens"   
## [43] "Clausocalanus sp"   
## [44] "Ctenocalanus vanus"   
## [45] "Pareucalanus attenuatus"   
## [46] NA   
## [47] "Rhincalanus rostrifrons"   
## [48] "Subeucalanus crassus"   
## [49] "Subeucalanus longiceps"   
## [50] "Subeucalanus mucronatus"   
## [51] "Subeucalanus pileatus"   
## [52] "Subeucalanus subcrassus"   
## [53] "Subeucalanus subtenuis"   
## [54] "Subeucalanus copepodid"   
## [55] "Euchaeta concinna"   
## [56] "Euchaeta indica"   
## [57] "Euchaeta plana"   
## [58] "Euchaeta rimana"   
## [59] "Euchaeta spinosa"   
## [60] "Euchaeta copepodid"   
## [61] NA   
## [62] NA   
## [63] "Lucicutia flavicornis"   
## [64] NA   
## [65] NA   
## [66] "Mecynocera clausi"   
## [67] "Pleuromamma abdominalis"   
## [68] "Pleuromamma borealis"   
## [69] NA   
## [70] NA   
## [71] "Metridicidae copepodid"   
## [72] "Acrocalanus gibber"   
## [73] "Acrocalanus gracilis"   
## [74] "Acrocalanus monachus"   
## [75] "Calocalanus contractus"   
## [76] "Calocalanus gracilis"   
## [77] "Calocalanus monospinus"   
## [78] "Calocalanus pavo"   
## [79] "Calocalanus pavoninus"   
## [80] "Calocalanus plumulosus"   
## [81] "Calocalanus styliremis"   
## [82] NA   
## [83] NA   
## [84] "Paracalanus aculeatus"   
## [85] "Paracalanus pavus"   
## [86] "Paracalanus serrulus"   
## [87] "Paracalanus sp"   
## [88] "Parvocalanus crassirostris"   
## [89] "Parvocalanus sp"   
## [90] NA   
## [91] NA   
## [92] NA   
## [93] NA   
## [94] NA   
## [95] NA   
## [96] "Calanopia elliptica"   
## [97] "Calanopia minor"   
## [98] "Lobidocera copepodid"   
## [99] NA   
## [100] "Pontellopsis tenuicauda"   
## [101] NA   
## [102] NA   
## [103] NA   
## [104] NA   
## [105] NA   
## [106] "Scolecithricella longispinosa"   
## [107] NA   
## [108] "Scolecithrix danae"   
## [109] NA   
## [110] "Temora discaudata"   
## [111] "Temora stylifera"   
## [112] "Temora turbinata"   
## [113] "Temoropia mayumbaensis"   
## [114] NA   
## [115] "Tortanus forcipatus"   
## [116] "Oithona atlantica"   
## [117] "Oithona attenuata"   
## [118] "Oithona brevicornis"   
## [119] "Oithona decipiens"   
## [120] "Oithona fallax"   
## [121] "Oithona fragilis"   
## [122] "Oithona longispina"   
## [123] "Oithona plumifera"   
## [124] "Oithona rigida"   
## [125] NA   
## [126] "Oithona similis"   
## [127] "Oithona simplex"   
## [128] NA   
## [129] "Oithona sp"   
## [130] "Paroithona sp"   
## [131] "Clytemnestra rostrata"   
## [132] "Clytemnestra scutellata"   
## [133] NA   
## [134] "Microsetella rosea"   
## [135] "Euterpina acutifrons"   
## [136] "Macrosetella gracilis"   
## [137] "Mormonilla minor"   
## [138] "Corycaeus (Agetus) flaccus"   
## [139] "Corycaeus (Agetus) typicus"   
## [140] NA   
## [141] "Corycaeus ( Corycaeus) speciosus"   
## [142] "Corycaeus (Ditrichocorycaeus) affinis"   
## [143] "Corycaeus (Ditrichocorycaeus) andrewsi"   
## [144] NA   
## [145] "Corycaeus (Ditrichocorycaeus) dahli"   
## [146] "Corycaeus (Ditrichocorycaeus) erythraeus"  
## [147] "Corycaeus (Ditrichocorycaeus) lubbocki"   
## [148] "Corycaeus (Ditrichocorycaeus) subtilis"   
## [149] "Corycaeus (Onychocorycaeus) agilis"   
## [150] "Corycaeus (Onychocorycaeus) catus"   
## [151] "Corycaeus (Onychocorycaeus) giesbrechti"   
## [152] "Corycaeus (Onychocorycaeus) pacificus"   
## [153] NA   
## [154] NA   
## [155] "Corycaeus ( Urochocorycaeus) lautus"   
## [156] "Corycaeus ( Urochocorycaeus) longistylis"  
## [157] NA   
## [158] "Farranula gibbula"   
## [159] "Farranula rostrata"   
## [160] "Corycaeidae sp"   
## [161] "Corycaeidae copepodid"   
## [162] NA   
## [163] NA   
## [164] "Oncaea conifera"   
## [165] "Oncaea media"   
## [166] "Oncaea mediterranea"   
## [167] "Oncaea minuta"   
## [168] "Oncaea similis"   
## [169] "Oncaea venusta"   
## [170] "Oncaea sp1"   
## [171] NA   
## [172] "Copilia mirabilis"   
## [173] NA   
## [174] NA   
## [175] NA   
## [176] "Sapphirina iris"   
## [177] NA   
## [178] NA   
## [179] "Sapphirina stellata"   
## [180] NA   
## [181] "Ratania sp"

Compare the results based on different distance measures.

#select dominant species only  
dominant\_comp <- copepod\_data[rownames(copepod\_data) %in% dominant\_species, ]   
  
library(vegan) # for Bray-Curtis

## 載入需要的套件：permute

## Warning: 套件 'permute' 是用 R 版本 4.4.2 來建造的

## 載入需要的套件：lattice

## This is vegan 2.6-8

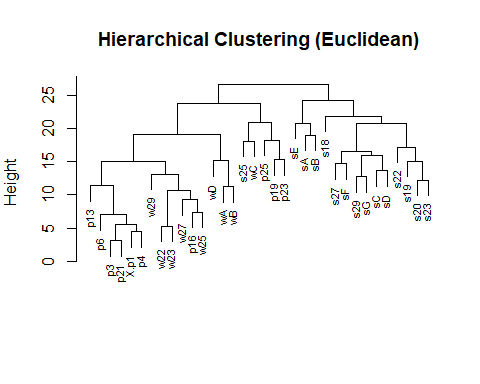
library(cluster) # for clustering visualization  
library(factoextra) # for visualizing clusters

## Warning: 套件 'factoextra' 是用 R 版本 4.4.3 來建造的

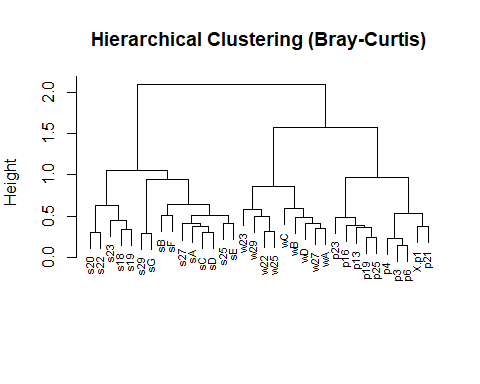
## 載入需要的套件：ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

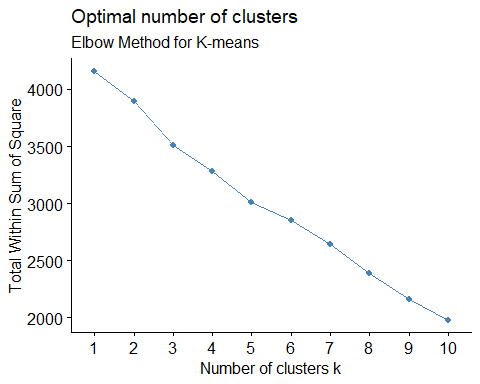
# Transpose: clustering is based on stations (columns), not species (rows)  
dominant\_t <- t(dominant\_comp)  
  
# Standardize data (optional, based on range)  
dominant\_scaled <- scale(dominant\_t)  
  
# Different distance measures  
dist\_euclidean <- dist(dominant\_scaled, method = "euclidean")  
dist\_bray <- vegdist(dominant\_t, method = "bray")  
  
# Hierarchical clustering  
hc\_euclidean <- hclust(dist\_euclidean, method = "ward.D2")  
hc\_bray <- hclust(dist\_bray, method = "ward.D2")  
  
# Plot dendrograms  
par(mar = c(5, 4, 4, 2))  
plot(hc\_euclidean, main = "Hierarchical Clustering (Euclidean)", xlab = "", sub = "", cex=0.7)



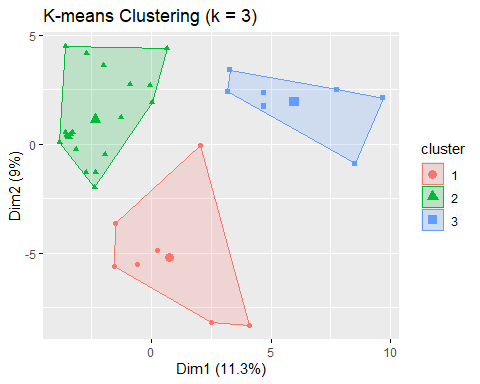
plot(hc\_bray, main = "Hierarchical Clustering (Bray-Curtis)", xlab = "", sub = "", cex=0.7)

 Compare the results based on different cluster algorithms.

# Determine number of clusters with Elbow method  
fviz\_nbclust(dominant\_scaled, kmeans, method = "wss") +   
 labs(subtitle = "Elbow Method for K-means")



# Perform K-means with optimal k (say k = 3)  
set.seed(123)  
km\_res <- kmeans(dominant\_scaled, centers = 3, nstart = 25)  
  
# Plot clustering results  
fviz\_cluster(km\_res, data = dominant\_scaled, geom = "point",  
 main = "K-means Clustering (k = 3)")

 Determine the final number of clusters and describe the differences among them.

# Add cluster labels to data  
cluster\_groups <- km\_res$cluster  
  
# Check species composition differences by cluster  
dominant\_with\_cluster <- cbind(dominant\_t, cluster = factor(cluster\_groups))  
  
# Convert the matrix to a data frame  
dominant\_df <- as.data.frame(dominant\_with\_cluster)  
  
# Add the cluster result from K-means as a new column  
dominant\_df$cluster <- factor(km\_res$cluster)  
  
# Calculate the average species composition for each cluster  
cluster\_summary <- aggregate(. ~ cluster, data = dominant\_df, FUN = mean)  
  
  
print(cluster\_summary)

## cluster Acartia bifilosa Acartia negligence Acartia omori Acartia pacifica  
## 1 1 0.04285714 1.535714 0.00000000 0.3414286  
## 2 2 0.00000000 0.096000 0.00000000 5.2350000  
## 3 3 0.00000000 1.574286 0.03714286 0.3014286  
## Acartia sp Metacalanus auvivillii Calanoides carinatus Calanus sinicus  
## 1 0.0000000 0.0000 0.08 8.4614286  
## 2 0.0755000 0.0185 0.00 4.8945000  
## 3 0.1085714 0.0000 0.00 0.1757143  
## Canthocalanus pauper Cosmocalanus darwinii Mesocalanus tenuicornis  
## 1 3.280000 0.9485714 0.10428571  
## 2 1.438000 0.0520000 0.00000000  
## 3 7.132857 0.2414286 0.04428571  
## Nannocalanus minor Neocalanus gracilis Undinula vulgaris Candacia bipinnata  
## 1 0.4200000 0.0000000 0.3871429 0.00000000  
## 2 0.0710000 0.0155000 0.2665000 0.00000000  
## 3 0.1771429 0.2485714 1.9385714 0.03857143  
## Candacia discaudata Candacia elongata Candaciidae copepodid  
## 1 0.1114286 0.0000 0.2071429  
## 2 0.0500000 0.0185 0.1130000  
## 3 0.0000000 0.0000 0.6442857  
## Centropages furcatus Centropages orsinii Centropages copepodid  
## 1 0.00000000 0.00 0.0700  
## 2 0.01550000 0.00 0.0275  
## 3 0.07285714 0.35 0.0800  
## Clausocalanus arcuicornis Clausocalanus brevipes Clausocalanus furcatus  
## 1 0.40142857 0.0000 4.045714  
## 2 0.08850000 0.0185 0.313000  
## 3 0.03857143 0.0000 3.398571  
## Clausocalanus ingens Clausocalanus mastigophorus Clausocalanus minor  
## 1 0.0000000 0.3428571 2.8500000  
## 2 0.0000000 0.0395000 0.7670000  
## 3 0.1071429 0.2028571 0.7428571  
## Clausocalanus pergens Clausocalanus sp Ctenocalanus vanus  
## 1 0.0000 0.0000 0.000  
## 2 0.0715 0.0285 0.037  
## 3 0.0000 0.0000 0.000  
## Pareucalanus attenuatus Rhincalanus rostrifrons Subeucalanus crassus  
## 1 0.4 0.4371429 0.42571429  
## 2 0.0 0.0000000 0.12750000  
## 3 0.0 0.0000000 0.08857143  
## Subeucalanus longiceps Subeucalanus mucronatus Subeucalanus pileatus  
## 1 0.00000000 0.6314286 1.1700000  
## 2 0.00000000 0.0615000 0.6320000  
## 3 0.07714286 0.0000000 0.4142857  
## Subeucalanus subcrassus Subeucalanus subtenuis Subeucalanus copepodid  
## 1 0.7857143 0.00000000 1.3357143  
## 2 0.4860000 0.00000000 0.2170000  
## 3 0.1257143 0.05428571 0.6242857  
## Euchaeta concinna Euchaeta indica Euchaeta plana Euchaeta rimana  
## 1 0.4471429 0.00000000 0.04428571 0.40142857  
## 2 0.3530000 0.00000000 0.08100000 0.00950000  
## 3 0.0000000 0.04428571 0.00000000 0.08714286  
## Euchaeta spinosa Euchaeta copepodid Lucicutia flavicornis Mecynocera clausi  
## 1 0.04571429 6.3757143 0.3400 0.0000  
## 2 0.00000000 7.5570000 0.1215 0.0495  
## 3 0.00000000 0.4342857 0.0000 0.0900  
## Pleuromamma abdominalis Pleuromamma borealis Metridicidae copepodid  
## 1 0.1128571 0.1228571 0.0000000  
## 2 0.0000000 0.0000000 0.0000000  
## 3 0.0000000 0.0000000 0.1242857  
## Acrocalanus gibber Acrocalanus gracilis Acrocalanus monachus  
## 1 2.621429 0.3785714 0.0000000  
## 2 1.617000 0.1005000 0.0000000  
## 3 5.491429 1.1457143 0.2314286  
## Calocalanus contractus Calocalanus gracilis Calocalanus monospinus  
## 1 0.1114286 0.5585714 0.10285714  
## 2 0.0000000 0.0295000 0.00000000  
## 3 0.1928571 0.2428571 0.07714286  
## Calocalanus pavo Calocalanus pavoninus Calocalanus plumulosus  
## 1 0.04000000 0.1542857 0.3200000  
## 2 0.01350000 0.0155000 0.0330000  
## 3 0.08142857 1.3685714 0.4557143  
## Calocalanus styliremis Paracalanus aculeatus Paracalanus pavus  
## 1 0.2185714 6.044286 20.31000  
## 2 0.0140000 4.623000 28.71100  
## 3 0.8271429 1.200000 11.42714  
## Paracalanus serrulus Paracalanus sp Parvocalanus crassirostris  
## 1 0.3671429 0.000 2.398571  
## 2 0.8000000 0.012 7.276500  
## 3 2.7528571 0.000 10.577143  
## Parvocalanus sp Calanopia elliptica Calanopia minor Lobidocera copepodid  
## 1 0.04571429 0.0400000 0.2814286 0.1300000  
## 2 0.00000000 0.0135000 0.0000000 0.0095000  
## 3 0.00000000 0.2428571 0.1814286 0.3385714  
## Pontellopsis tenuicauda Scolecithricella longispinosa Scolecithrix danae  
## 1 0.00000000 1.44285714 0.3042857  
## 2 0.00000000 1.51300000 0.0110000  
## 3 0.04571429 0.04428571 0.0000000  
## Temora discaudata Temora stylifera Temora turbinata Temoropia mayumbaensis  
## 1 0.1642857 1.0171429 2.201429 0.0000  
## 2 0.0360000 0.0890000 3.035000 0.0365  
## 3 0.0800000 0.1728571 7.137143 0.0000  
## Tortanus forcipatus Oithona atlantica Oithona attenuata Oithona brevicornis  
## 1 0.0000 0.20714286 0.170000 0.04285714  
## 2 0.0185 0.07900000 1.909500 1.14850000  
## 3 0.0000 0.04571429 3.424286 0.42571429  
## Oithona decipiens Oithona fallax Oithona fragilis Oithona longispina  
## 1 0.1814286 1.1128571 0.05142857 0.0000  
## 2 0.0545000 0.4390000 0.09650000 0.1235  
## 3 0.0000000 0.8742857 0.00000000 0.0000  
## Oithona plumifera Oithona rigida Oithona similis Oithona simplex Oithona sp  
## 1 5.180000 0.0000 0.6185714 0.0000 0.00000000  
## 2 0.876500 0.2665 5.0725000 0.1315 0.00000000  
## 3 2.942857 0.3900 0.1457143 0.0000 0.03714286  
## Paroithona sp Clytemnestra rostrata Clytemnestra scutellata  
## 1 0.0000000 0.08285714 0.08857143  
## 2 0.0210000 0.02700000 0.26300000  
## 3 0.1214286 0.00000000 0.04428571  
## Microsetella rosea Euterpina acutifrons Macrosetella gracilis  
## 1 0.0000000 2.3385714 0.1942857  
## 2 0.1435000 3.2105000 0.1570000  
## 3 0.1742857 0.9271429 0.2385714  
## Mormonilla minor Corycaeus (Agetus) flaccus Corycaeus (Agetus) typicus  
## 1 0.00000000 0.06428571 0.13428571  
## 2 0.00000000 0.00000000 0.00000000  
## 3 0.08142857 0.03857143 0.03857143  
## Corycaeus ( Corycaeus) speciosus Corycaeus (Ditrichocorycaeus) affinis  
## 1 0.15428571 0.27142857  
## 2 0.00000000 6.19800000  
## 3 0.07571429 0.04571429  
## Corycaeus (Ditrichocorycaeus) andrewsi Corycaeus (Ditrichocorycaeus) dahli  
## 1 0.0000000 0.5900  
## 2 0.1470000 0.1925  
## 3 0.4371429 2.4300  
## Corycaeus (Ditrichocorycaeus) erythraeus  
## 1 0.07428571  
## 2 0.13200000  
## 3 0.08285714  
## Corycaeus (Ditrichocorycaeus) lubbocki Corycaeus (Ditrichocorycaeus) subtilis  
## 1 0.000 0.3428571  
## 2 1.965 0.1285000  
## 3 0.200 0.6042857  
## Corycaeus (Onychocorycaeus) agilis Corycaeus (Onychocorycaeus) catus  
## 1 0.09714286 0.1328571  
## 2 0.00000000 0.0550000  
## 3 0.20714286 0.7085714  
## Corycaeus (Onychocorycaeus) giesbrechti Corycaeus (Onychocorycaeus) pacificus  
## 1 0.1971429 0.04571429  
## 2 0.3195000 0.07250000  
## 3 1.3585714 0.27285714  
## Corycaeus ( Urochocorycaeus) lautus Corycaeus ( Urochocorycaeus) longistylis  
## 1 0.000 0.14857143  
## 2 0.039 0.00000000  
## 3 0.000 0.05428571  
## Farranula gibbula Farranula rostrata Corycaeidae sp Corycaeidae copepodid  
## 1 0.4042857 0.00000000 0.000 0.7100  
## 2 0.0470000 0.06950000 0.012 0.5505  
## 3 1.0128571 0.07714286 0.000 0.0000  
## Oncaea conifera Oncaea media Oncaea mediterranea Oncaea minuta Oncaea similis  
## 1 0.6428571 1.091429 0.9071429 0.0000 0.1385714  
## 2 0.6820000 0.046500 0.0840000 0.0265 0.0295000  
## 3 4.9585714 0.000000 0.2385714 0.2800 0.1271429  
## Oncaea venusta Oncaea sp1 Copilia mirabilis Sapphirina iris  
## 1 5.338571 0.00000000 0.07285714 0.00000000  
## 2 3.110000 0.00950000 0.00000000 0.00000000  
## 3 12.497143 0.04571429 0.00000000 0.04571429  
## Sapphirina stellata Ratania sp  
## 1 0.1300 0.000  
## 2 0.0185 0.012  
## 3 0.0000 0.000

Cluster 1 was dominated by Calanus sinicus, Cluster 2 had a higher proportion of Euchaeta copepodid, and Cluster 3 was also dominated by Calanus sinicus.