Cluster Summary

Require the following packages:

library(tidyverse)  
library(gridExtra)

Load in data. growth.rates. csv has the standardised observations (tr0obs) for each species for the years 1976-2014 (Not all species start in 1976.). col.ind.prev.year is the tr0obs value for the previous year. Growth.rate.dif (differential) is the tr0obs value minus the previous year's value. It is therefore a measure of the differnece in population growth rate year to year.

growth.rate <- read.csv(  
 "C:\\Users\\dp005352\\Dropbox\\PhD\\BMS\_Data\_Analysis\\Outputs\\growth.rates.csv",   
 header = T  
 )

Select appropriate rows from the data. Only require the species number, the year and the growth.rate.dif

cluster <- select(growth.rate, species, year, growth.rate.dif)

species.codes is a list of all the species numbers matched with the species common and latin names

codes <- read.csv(  
 "C://Users//dp005352//Dropbox//PhD//BMS\_Data\_Analysis//Synchrony\_Data//species\_codes.csv",   
 header = T  
 )

Merge the two data sets together so that each species has a name and number

cluster <- merge(cluster, codes, by.x = "species", by.y = "species.code")

Remove unnecessary columns. Only need growth rate differential, year and name.

cluster <- select(cluster, growth.rate.dif, year, common.name)

Create long form table with each species as its own column and remove year column as no longer required.

cluster <- spread(cluster, common.name, growth.rate.dif)  
cluster <- select(cluster, -year)

Create a correlation matrix. cor function correlates each column with every other column in the data set. For this reason we removed the year column.

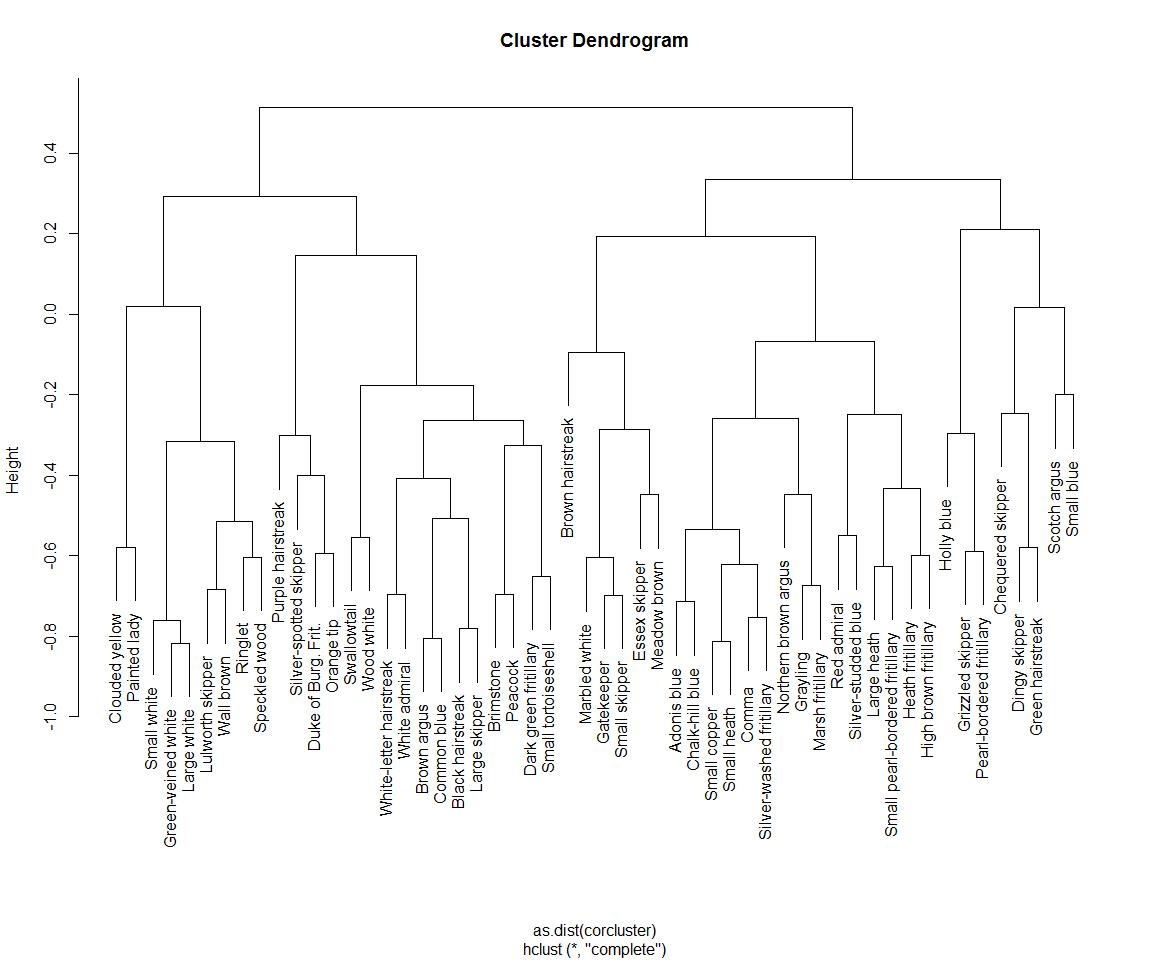
corcluster <- cor(cluster, use = "pairwise.complete.obs")

Times dataset by minus 1 as need to look at similarity rather than distance.

corcluster <- corcluster \* -1

Create dendrogram for correlations. The algorithm works as follows: Put each data point in its own cluster. Identify the closest two clusters and combine them into one cluster. Repeat the above step till all the data points are in a single cluster.

cluster.plot <- hclust(as.dist(corcluster))  
plot(cluster.plot)



Next stage is to plot growth rate differential vs year for each species, split into clusters using the clusters shown in the dendrogram. To do this requires adding names to species code.

growth.rate <- merge(growth.rate, codes, by.x = "species", by.y = "species.code")

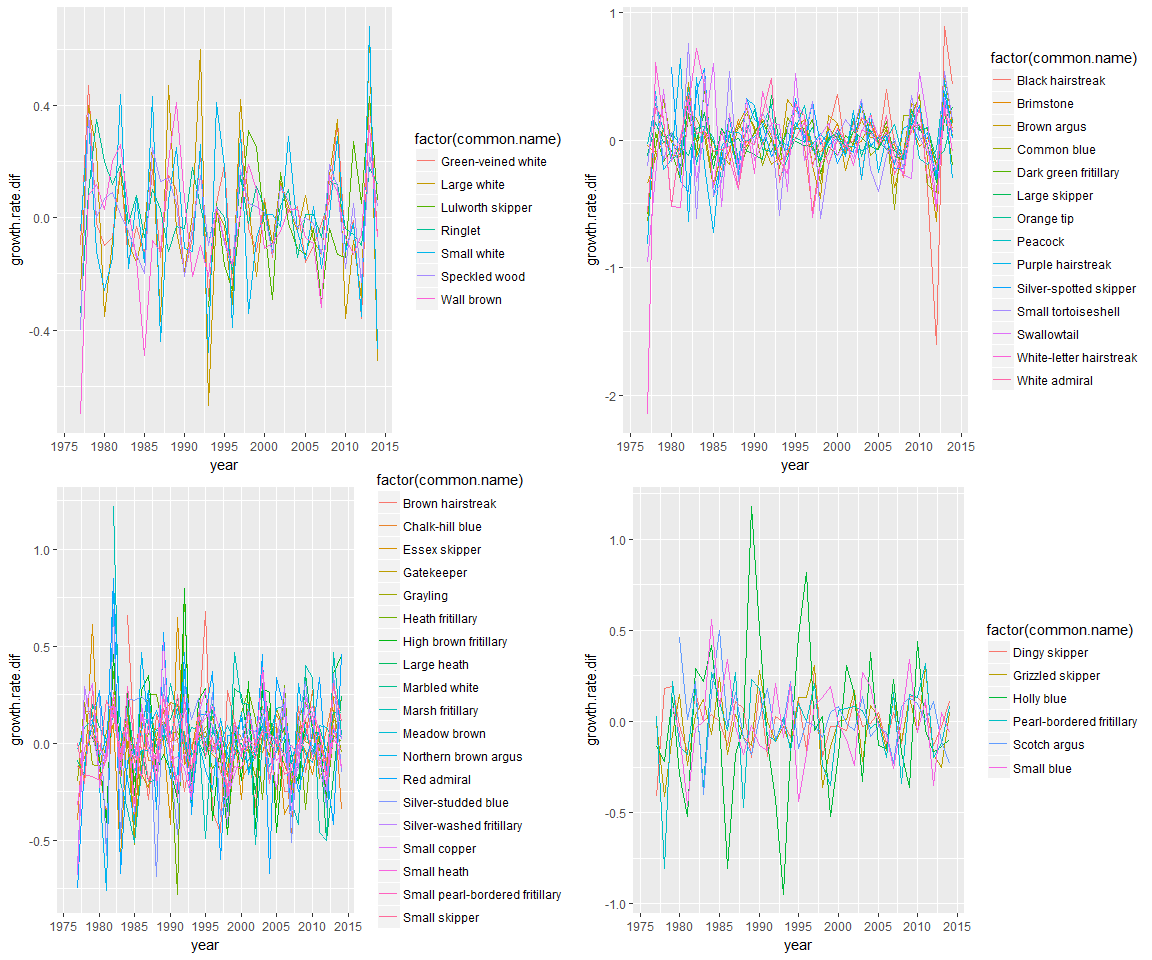
Remove unnecessary columns. Only need growth rate differential, year and name

growth.rate <-  
 select(growth.rate, growth.rate.dif, year, common.name, species)

Dendrogram can be split initially into four large clusters Each set of code creates a plot for an individual cluster.Graphs brought together using grid.arrange

cluster4.1 <-  
 growth.rate %>% filter(  
 common.name %in% c(  
 #"Clouded yellow", #"Painted lady",  
 "Small white", "Green-veined white", "Large white",  
 "Lulworth skipper", "Wall brown", "Ringlet", "Speckled wood"  
 )  
 )  
  
clusterplot4.1 <-  
 ggplot(cluster4.1, aes(  
 x = year,  
 y = growth.rate.dif,  
 colour = factor(common.name)  
 )) +   
 geom\_line() +   
 scale\_x\_continuous(breaks = seq(1975, 2015, 5))  
  
  
cluster4.2 <-  
 growth.rate %>% filter(  
 common.name %in% c(  
 "Purple hairstreak", "Silver-spotted skipper",  
 "Duke of Burg. Frit.", "Orange tip", "Swallowtail",  
 "Wood white", "White-letter hairstreak", "White admiral",  
 "Brown argus", "Common blue", "Black hairstreak",  
 "Large skipper", "Brimstone", "Peacock",  
 "Dark green fritillary", "Small tortoiseshell"  
 )  
 )  
  
clusterplot4.2 <-  
 ggplot(cluster4.2, aes(  
 x = year,  
 y = growth.rate.dif,  
 colour = factor(common.name)  
 )) +   
 geom\_line() +   
 scale\_x\_continuous(breaks = seq(1975, 2015, 5))

cluster4.3 <-  
 growth.rate %>% filter(  
 common.name %in% c(  
 "Brown hairstreak", "Marbled white", "Gatekeeper",  
 "Small skipper", "Essex skipper", "Meadow brown",  
 "Adonis blues", "Chalk-hill blue", "Small copper",  
 "Small heath", "Comma", "Silver-washed fritillary",  
 "Northern brown argus", "Grayling", "Marsh fritillary",  
 "Red admiral", "Silver-studded blue", "Large heath",  
 "Small pearl-bordered fritillary", "Heath fritillary",  
 "High brown fritillary"  
 )  
 )  
  
clusterplot4.3 <-  
 ggplot(cluster4.3, aes(  
 x = year,  
 y = growth.rate.dif,  
 colour = factor(common.name)  
 )) +   
 geom\_line() +   
 scale\_x\_continuous(breaks = seq(1975, 2015, 5))  
  
  
cluster4.4 <-  
 growth.rate %>% filter(  
 common.name %in% c(  
 "Holly blue", "Grizzled skipper",  
 "Pearl-bordered fritillary", #"Chequered skipper",  
 "Dingy skipper", "Green hairstreak",  
 "Scotch argus", "Small blue"  
 )  
 )  
  
clusterplot4.4 <-  
 ggplot(cluster4.4, aes(  
 x = year,  
 y = growth.rate.dif,  
 colour = factor(common.name)  
 )) +   
 geom\_line() +   
 scale\_x\_continuous(breaks = seq(1975, 2015, 5))  
  
grid.arrange(clusterplot4.1,  
 clusterplot4.2,  
 clusterplot4.3,  
 clusterplot4.4,  
 ncol = 2,  
 nrow = 2)



Can then be split into 11 clusters. 