Chapter 6 Upper Murrumbidgee Cod Nests Distribution

Alan Couch

Thu Aug 06 8:28:10 PM 2015

## Loading required package: knitr  
## Loading project configuration  
## Autoloading helper functions  
## Running helper script: helpers.R  
## Running helper script: otoPartChem.R  
## Autoloading data  
## Loading data set: allOtoChemData  
## Loading data set: CladeNamesToMerge  
## Loading data set: cnData  
## Loading data set: CopyOfDMac14.1567snps  
## Loading data set: DMac14.1567DistMatrix  
## Loading data set: DMac14.1567snps  
## Loading data set: qslAgeData  
## Loading data set: qslGeneticsForNestChapter  
## Loading data set: qslLarvaeAgePlus  
## Loading data set: siteGroupings  
## Munging data  
## Running preprocessing script: 01MungeGeneticsData.R  
## Running preprocessing script: 02MungeChemAverages.R

This document includes methods, results and possibly some discussion dot points for the nest distribution chapter.

source("http://addictedtor.free.fr/packages/A2R/lastVersion/R/code.R")# load code of A2R function  
library(ggplot2)  
library(ggdendro)  
library(ape)  
library(dendextend)  
library(Hmisc)  
library(ade4)

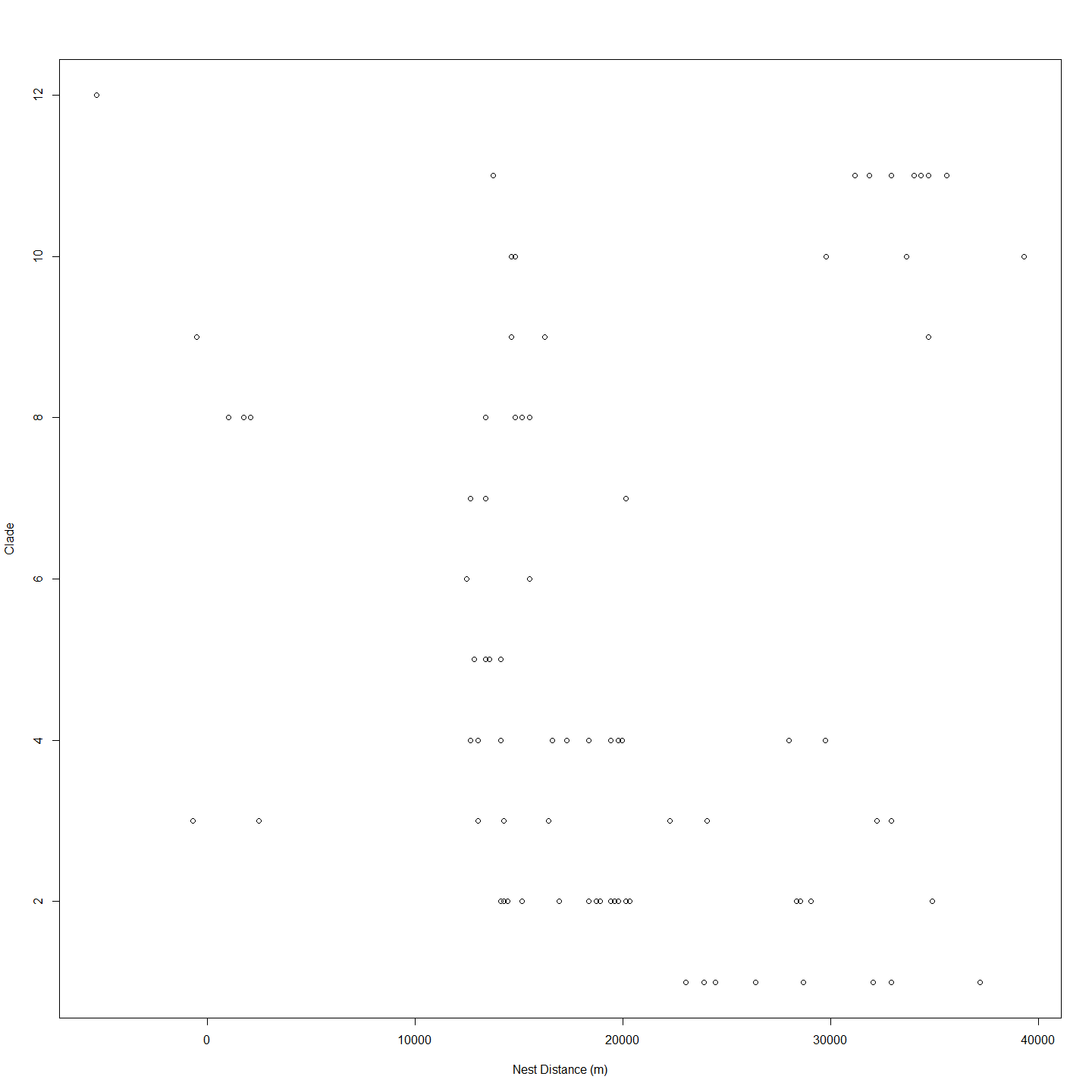
## Methods

## Results

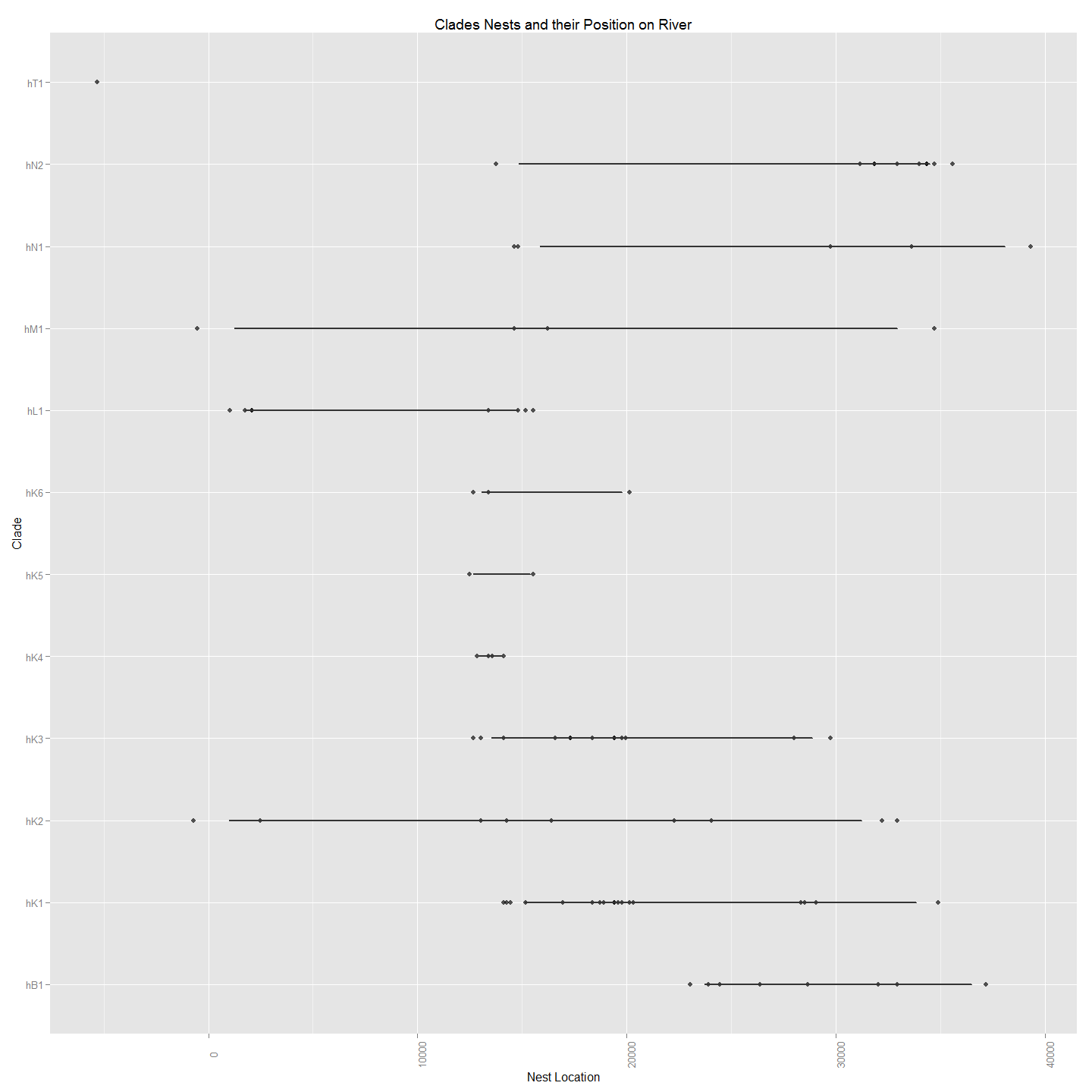
### Look at Clades Over the River Reach

The clades distribution over the river reach suggests some structure.

dispersalVelocity<-1200  
  
larv1<-merge(larv,qslGeneticsForNestChapter, by.x="LarvalID", by.y="LarvaID")  
rownames(CladeNamesToMerge) <- CladeNamesToMerge[,1]  
CladeNamesToMerge$Label<-rownames(CladeNamesToMerge)  
larv2<-merge(larv1,CladeNamesToMerge, by="Label")  
larv2<-merge(larv2,qslAgeData, by.x="LarvalID", by.y="LarvaID")  
# To create a distance using the previously calculated best estimate of drift velocity (m/d available since leaving brood care)  
larv2$nestdist<-larv2$Distance.to.Angle.Crossing..m.-(dispersalVelocity\*(larv2$Day.of.Year-(larv2$hatchDoY+7))) #this should come from DB perhaps  
  
plot(larv2$nestdist,larv2$clade, xlab="Nest Distance (m)", ylab = "Clade")

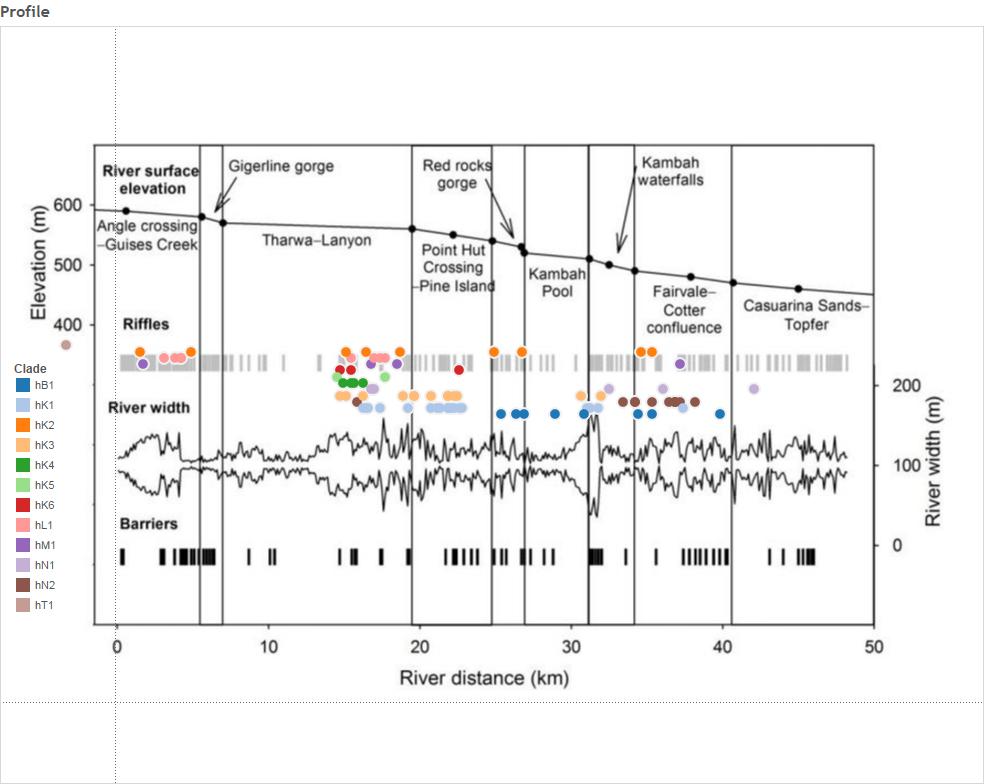


plot2 <- ggplot(larv2, aes(nestdist,clade))  
plot2 + geom\_point(alpha = 2/3) +geom\_boxplot()+labs(title = "Clades Nests and their Position on River")+ labs(x="Nest Location") +labs(y = "Clade")+ theme(axis.text.x=element\_text(angle=90))

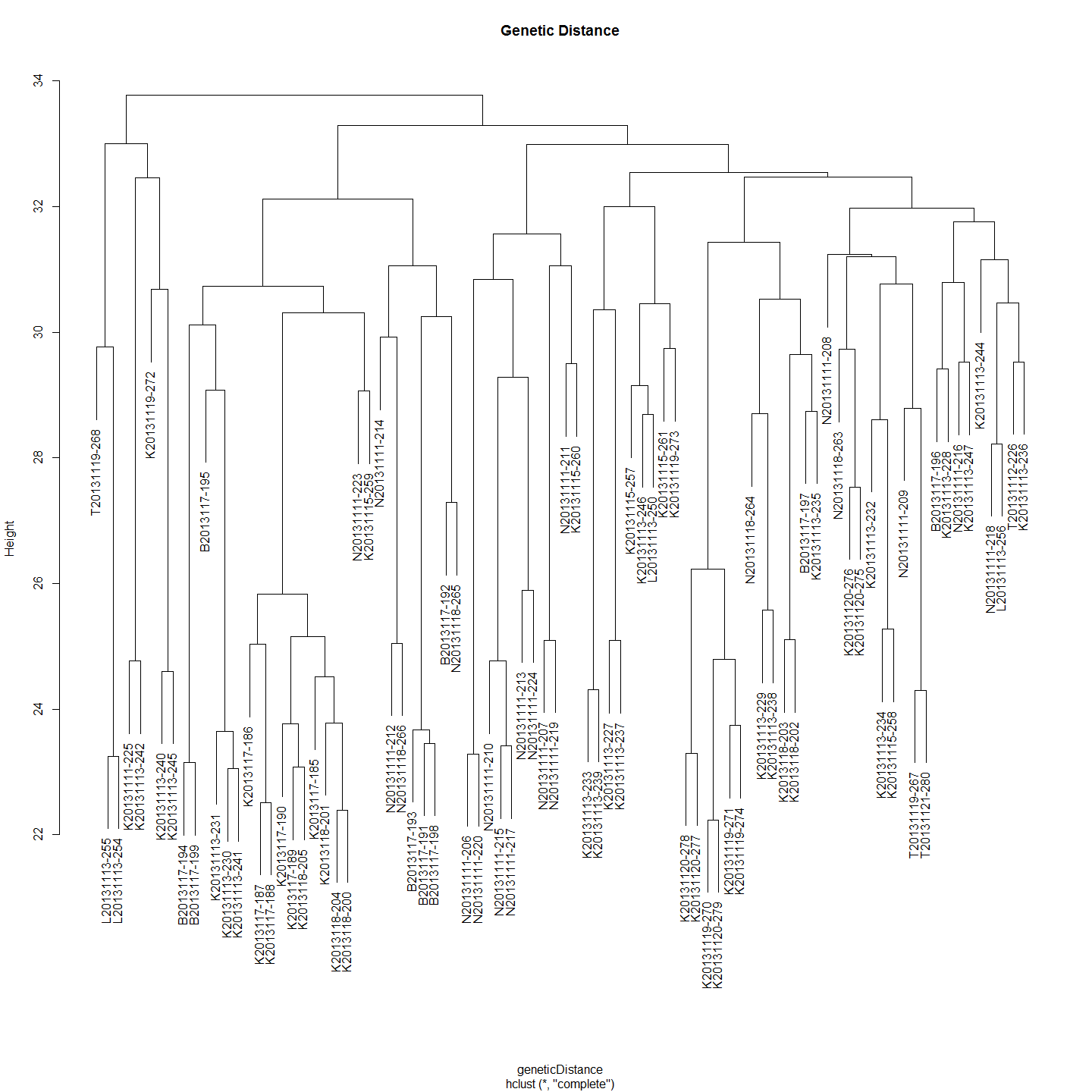


## Discussion (points only)

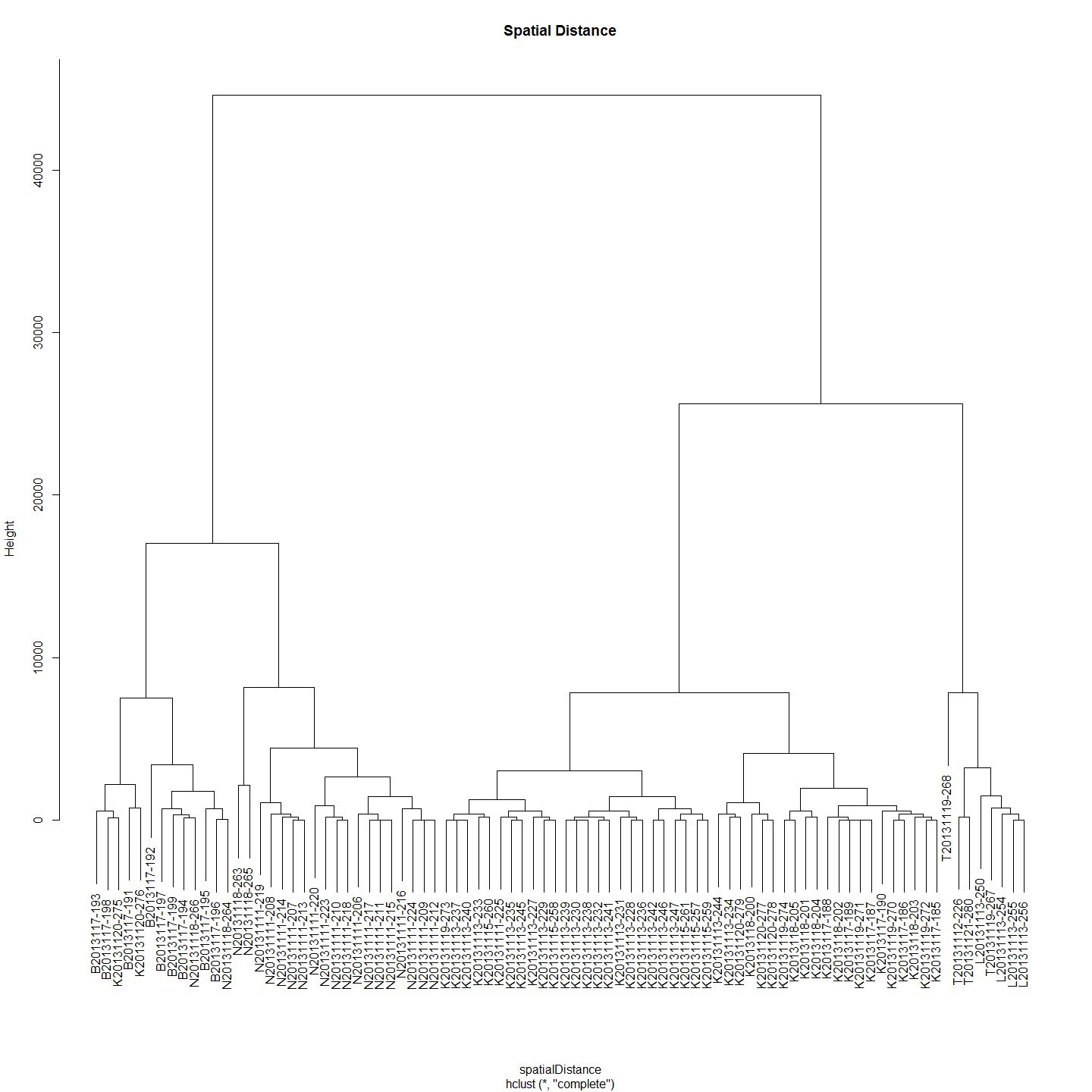
* This suggests that all clades except one (hT1) exist below a barrier around 10000m but that only three clades (hM1,hL1 and hK2) exist above and below this barrier.
* One possible explanation that might be inferred from this is that adults migrating upstream for spawning are prevented from doing so by a barrier between the 5000 and 15000m mark but that larvae produced above the barrier are able to disperse and so are represented downstream. It is noteable that the barrier and the big gap in apparent nests from 6000m to 14000m corresponds with the Tharwa sand slug - a long stretch of sand that has previously been the subject of remedial work because it has been believed to be a barrier to cod migration (ref).

 # needs updating in Tableau to ensure commensurate with r.

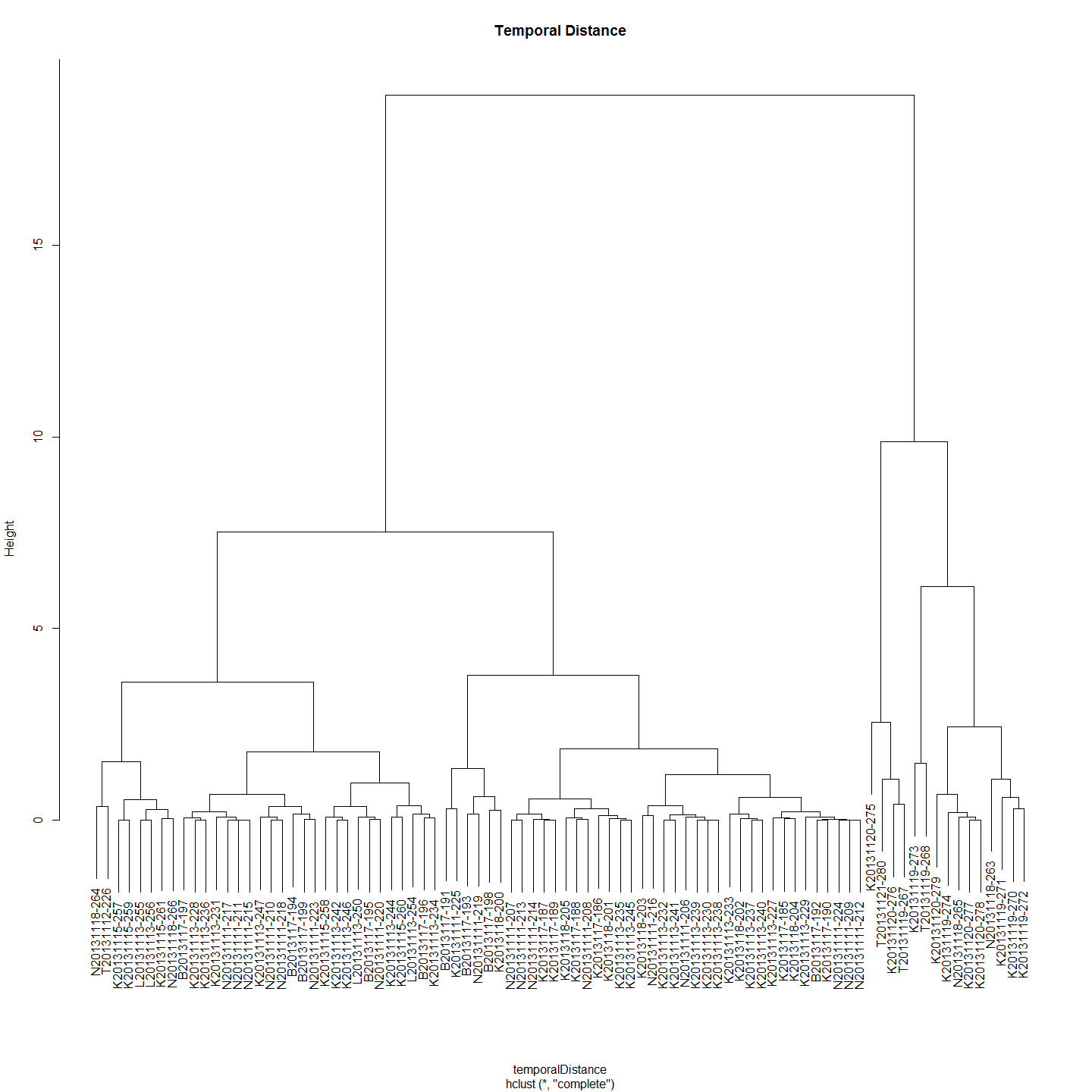
require(clusterSim) # to allow data normalization  
  
# This code is to take the three distances matrices; (genetic distance, spatial distance, temporal distance) and add them and then cluster to give a better approximation of the nest an individual is likely to belong to.   
#However, on relfection I dont think this is valid. Not on dist matrices. should just be on three columns of values clade, nestdist, hatchdoy  
  
#First need to create temporal distance from the larv data set so need label and hatch day of the year.  
  
#Create Hatch DoY and then Temporal Distance Matrix  
  
temporalDistance<-data.frame(larv2$Label,(larv2$Day.of.Year-(larv2$ageOL)))  
colnames(temporalDistance)[2] <- "hatchDoY"  
colnames(temporalDistance)[1] <- "label"  
row.names(temporalDistance)<-temporalDistance$label  
  
#Need to have only those which have genetic (and spatial distance) equivalents  
geneticDistance<-MCsnps[-c(1:7),] # so there are 86 entries.  
df<-merge(temporalDistance, geneticDistance, by="row.names")  
df<-df[,2:3]  
row.names(df)<-df$label  
df$label<-NULL  
temporalDistance<-df  
  
#Make the Actual Temporal Distance Matrix  
temporalDistance<-dist(temporalDistance)  
  
# Create Spatial Distance Matrix based on Best Nest Estimate as previously determined via Iterated Mantel.  
#spatialDistance<-data.frame(larv$Label,(larv$Distance.to.Angle.Crossing..m.-(bestNestEst\*(larv$Day.of.Year-((larv$Day.of.Year-(74.308\*larv$Mean.Otolith.Length.is.in.Millimetres.for.comparison.with.Adults-4.44361)))+7))))  
  
spatialDistance<-data.frame(larv2$Label,(larv2$Distance.to.Angle.Crossing..m.-(dispersalVelocity\*(larv2$Day.of.Year-((larv2$Day.of.Year-(larv2$ageOL)))+7))))  
  
geneticDistance<-MCsnps[-c(1:7),] # so there are 86 entries.  
row.names(spatialDistance)<-spatialDistance[,1]  
df<-merge(spatialDistance, geneticDistance, by="row.names")  
df<-df[,2:3]  
row.names(df)<-df$larv2.Label  
df$larv2.Label<-NULL  
spatialDistance<-df  
  
#Make the Actual Spatial Distance Matrix  
spatialDistance<-dist(spatialDistance)  
  
#Make the Actual Genetic Distance Matrix  
geneticDistance <- dist(geneticDistance)  
  
plot(hclust(geneticDistance), main="Genetic Distance")



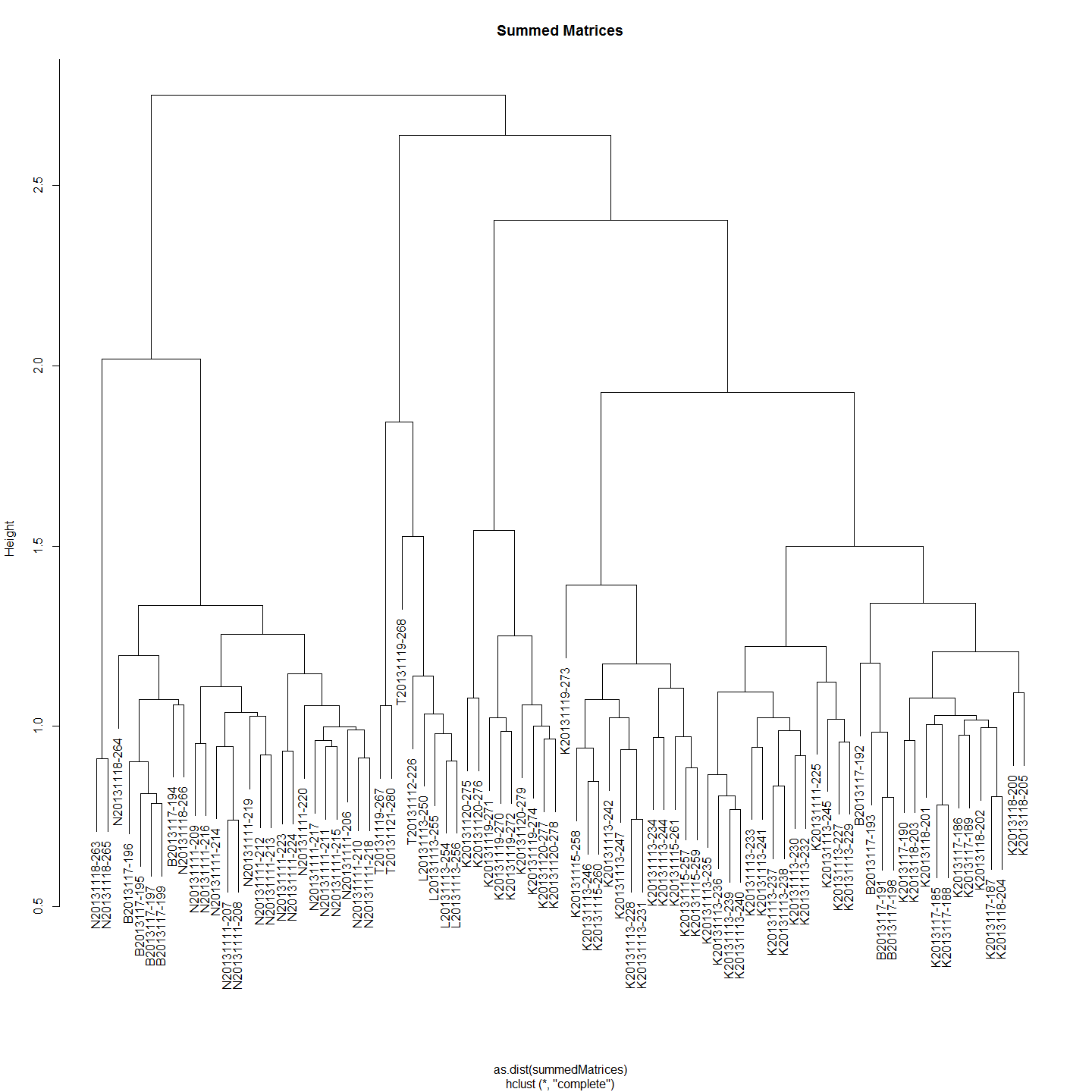
plot(hclust(spatialDistance), main="Spatial Distance")



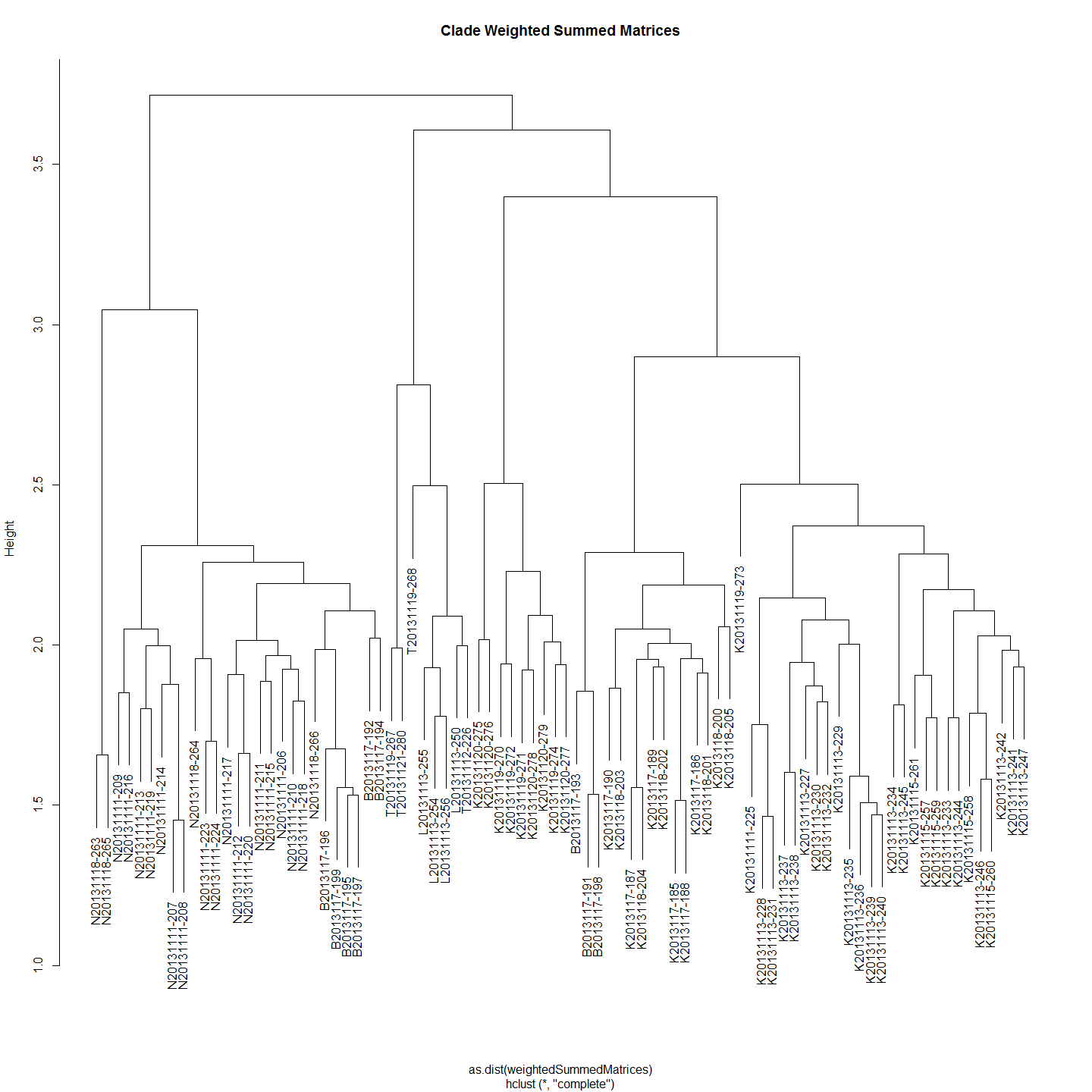
plot(hclust(temporalDistance), main="Temporal Distance")



#Add the distance matrices together after scaling. Need to scale of Distance dominates with its big numbers.  
summedMatrices<- (data.Normalization (as.matrix(spatialDistance),type="n4",normalization="column"))+(data.Normalization (as.matrix(geneticDistance),type="n4",normalization="column"))+(data.Normalization (as.matrix(temporalDistance),type="n4",normalization="column"))  
  
  
plot(hclust(as.dist(summedMatrices)), main="Summed Matrices")



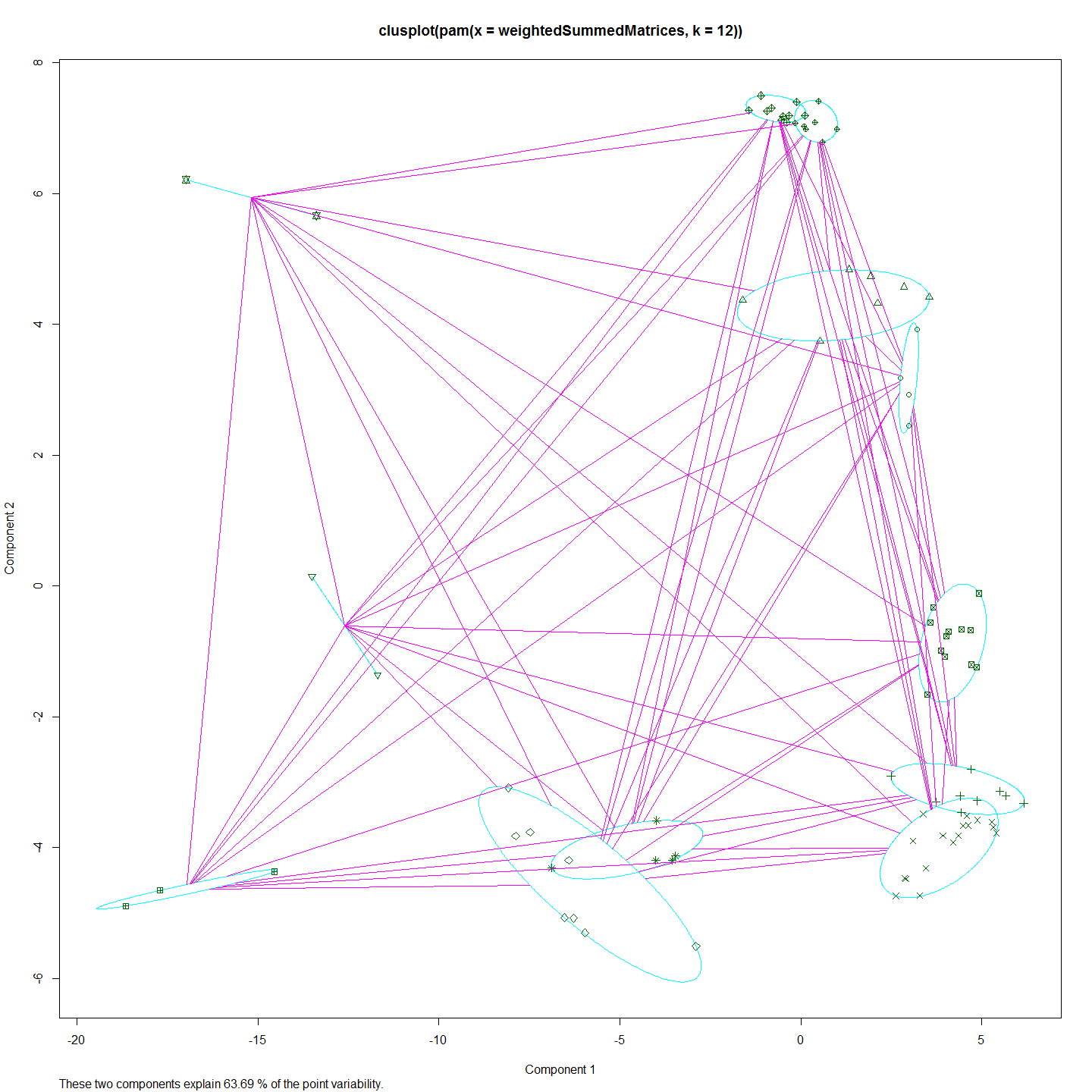
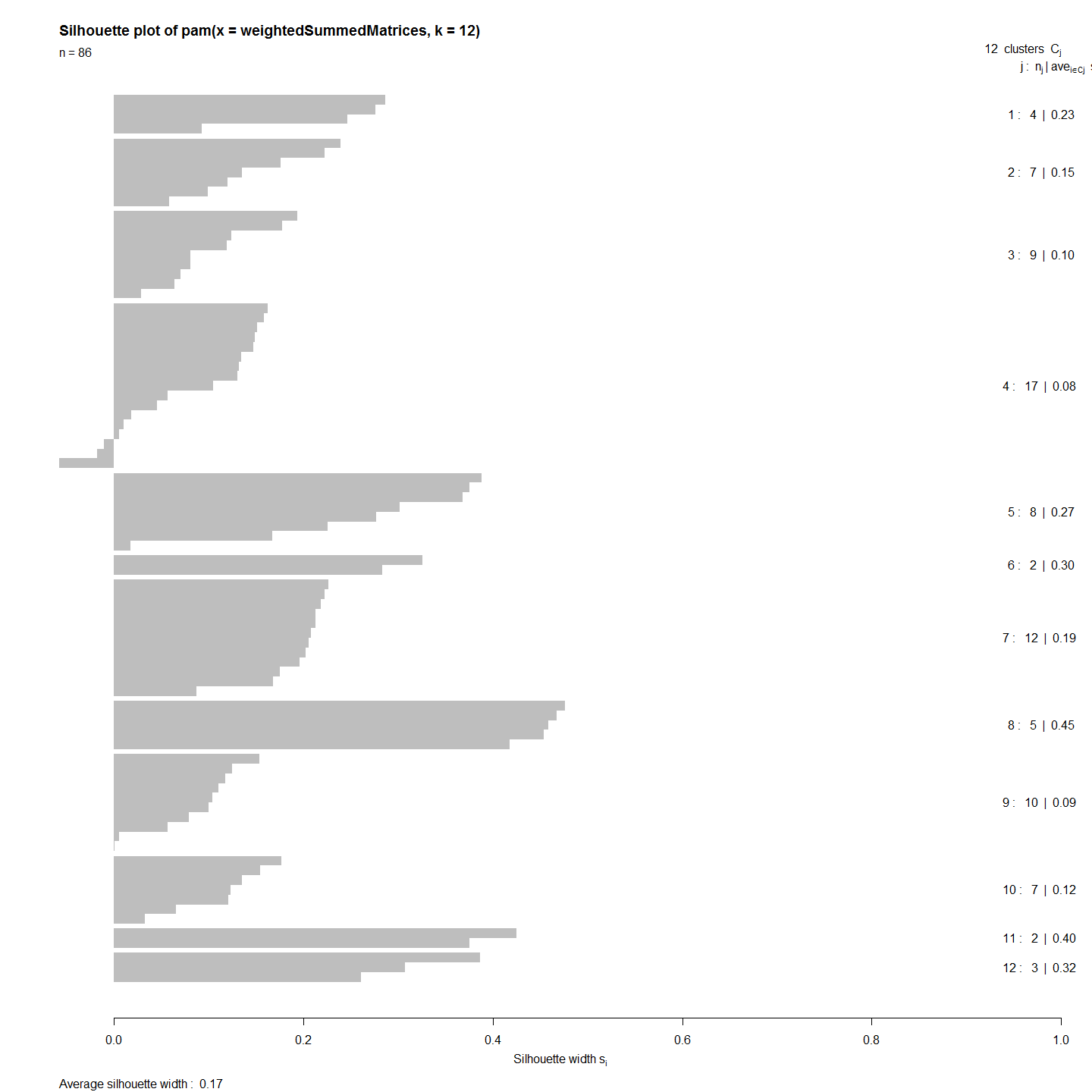
#Weight Genetic Distance two times more than Spatial and Temporal distances.  
weightedSummedMatrices<- (data.Normalization (as.matrix(spatialDistance),type="n4",normalization="column"))+2\*(data.Normalization (as.matrix(geneticDistance),type="n4",normalization="column"))+(data.Normalization (as.matrix(temporalDistance),type="n4",normalization="column"))  
  
plot(hclust(as.dist(weightedSummedMatrices)), main="Clade Weighted Summed Matrices")



#pam medoid clustering  
#http://www.stat.berkeley.edu/~spector/s133/Clus.html for interpretation  
  
require(cluster)  
wsm.pam <- pam(weightedSummedMatrices,12)  
table(wsm.pam$clustering)

##   
## 1 2 3 4 5 6 7 8 9 10 11 12   
## 4 7 9 17 8 2 12 5 10 7 2 3

plot(wsm.pam)

## Code Chunks in this Document

## [1] "Project\_Template\_and\_Knitr" "Set\_Global\_Options" "LoadLibraries" "Clades\_and\_Location"   
## [5] "sumMatricesToFindNests" "sumMatricesToFindNestsGo" "Include\_Chunk\_Labels\_and\_Session Information"

## R version 3.1.1 (2014-07-10)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
##   
## locale:  
## [1] LC\_COLLATE=English\_Australia.1252 LC\_CTYPE=English\_Australia.1252 LC\_MONETARY=English\_Australia.1252 LC\_NUMERIC=C LC\_TIME=English\_Australia.1252   
##   
## attached base packages:  
## [1] splines grid stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] clusterSim\_0.44-2 MASS\_7.3-34 cluster\_1.15.2 ade4\_1.6-2 Hmisc\_3.14-4 Formula\_1.1-2 survival\_2.37-7 lattice\_0.20-29   
## [9] dendextend\_0.18.3 ape\_3.1-4 ggdendro\_0.1-15 ggplot2\_1.0.0 ProjectTemplate\_0.5-1 knitr\_1.6   
##   
## loaded via a namespace (and not attached):  
## [1] class\_7.3-11 colorspace\_1.2-4 digest\_0.6.4 e1071\_1.6-4 evaluate\_0.5.5 formatR\_1.0 gtable\_0.1.2 htmltools\_0.2.6 labeling\_0.3   
## [10] latticeExtra\_0.6-26 magrittr\_1.0.1 munsell\_0.4.2 nlme\_3.1-117 plyr\_1.8.1 proto\_0.3-10 R2HTML\_2.3.1 RColorBrewer\_1.0-5 Rcpp\_0.11.2   
## [19] reshape2\_1.4 rgl\_0.95.1201 rmarkdown\_0.2.54 scales\_0.2.4 stringr\_0.6.2 tools\_3.1.1 whisker\_0.3-2 yaml\_2.1.13