Chapter 6 Upper Murrumbidgee Cod Nests Distribution

Alan Couch

Mon Oct 19 3:29:34 PM 2015

## Loading required package: knitr  
## Loading project configuration  
## Autoloading helper functions  
## Running helper script: otolithTransectPlot.R  
## Running helper script: otoPartChem.R  
## Autoloading cache  
## Autoloading data  
## Loading data set: allOtoChemData  
## Loading data set: allOtoChemDataOLD  
## Loading data set: DMac14.1567DistMatrix  
## Loading data set: DMac14.1567snps  
## Loading data set: DMac14.1567snps2  
## Loading data set: qslAllLarvaInfo  
## Loading data set: qslAllLarvaInfoold  
## Loading data set: qslAllLarvaInfoold2  
## Loading data set: qslDartCovariates  
## Loading data set: qslLarvaeAgePlus  
## 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 and larval dispersal chapter.

## Packages Required to be Loaded

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

In the previouschapter the mean larval dispersal velocity was used with a value of 1320 metres per day. This chapter will detail how that figure was determined and will use Genetic and dispersal data to examine nest locations and larval dispersal in the Upper Murrumbidgee.

## Results

## Upper Murrumbidgee Larval Murray Cod Genetic and Geographic Distances

This test correlates the physical distances of the nests of the larvae with genetic distance of larvae from the Murrumbidgee collected in 2011,2012 and 2013 from 6 sites. Most analysis is done for each year but if it is not explicitly stated it is 2013 result or all years.

By iterating the mantel test using distance matrices generated for nest distances based on a range of larval dispersal velocities and the time available to those larvae for dispersal, we can identify the most highly correlated - the peak - which then allows an estimate of the mean larval dispersal velocity. In turn the mean distance a nest site is from the collection site. The iterations of the Mantel test used a larval dispersal velocity range from -2000 metres (upstream dispersal) to 5000 metres (downstream dispersal) per day. The highest correlation represents the the mean velocity of the dispersing larvae. While the current assumption is that dispersal is downstream, not upstream, we test for this. The curve produced from the estimations is as follows.

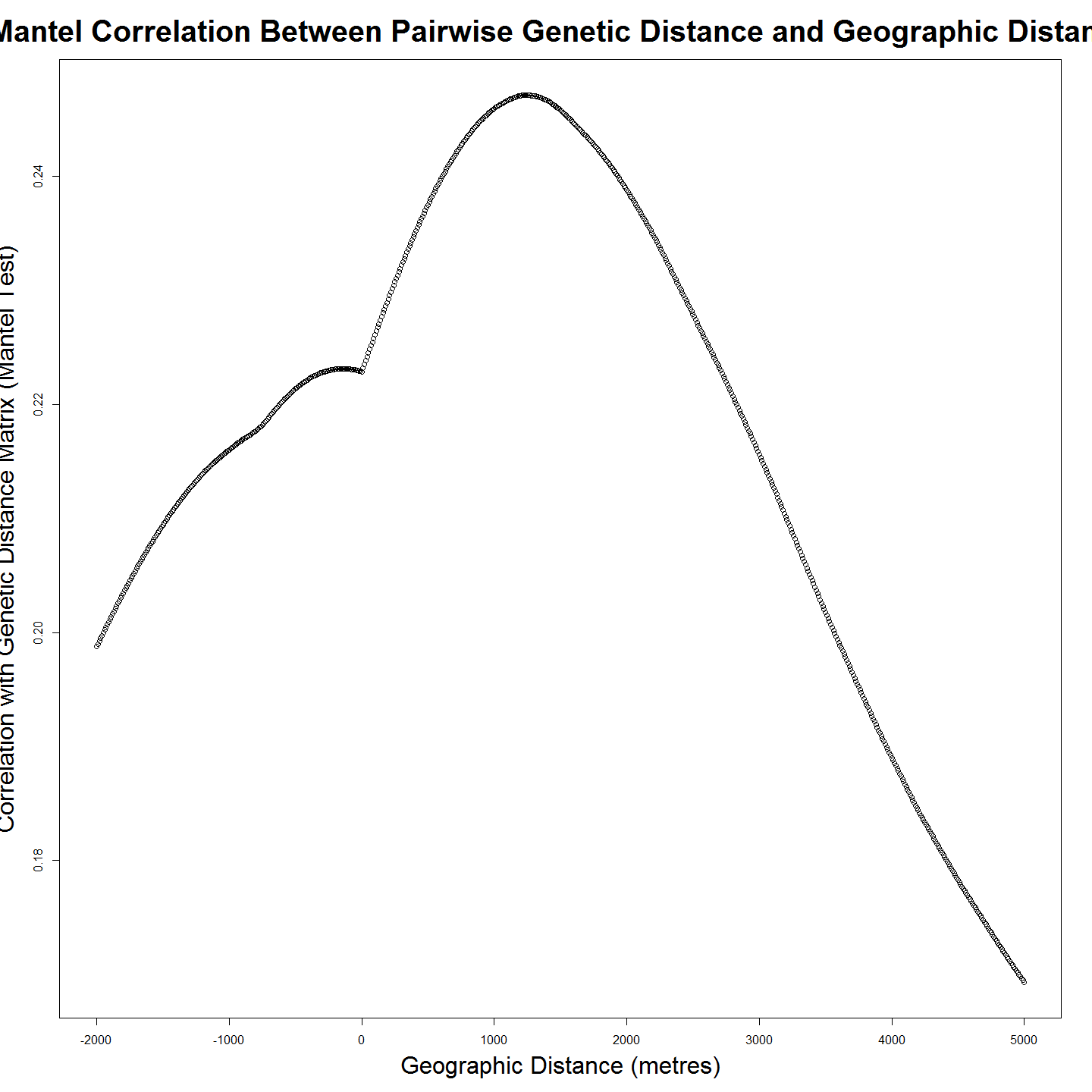
This can only work within years as the idea of geographic distances looses all meaning between years.

### Iterate through Mantel test using a range of possible larval dispersal velocities.

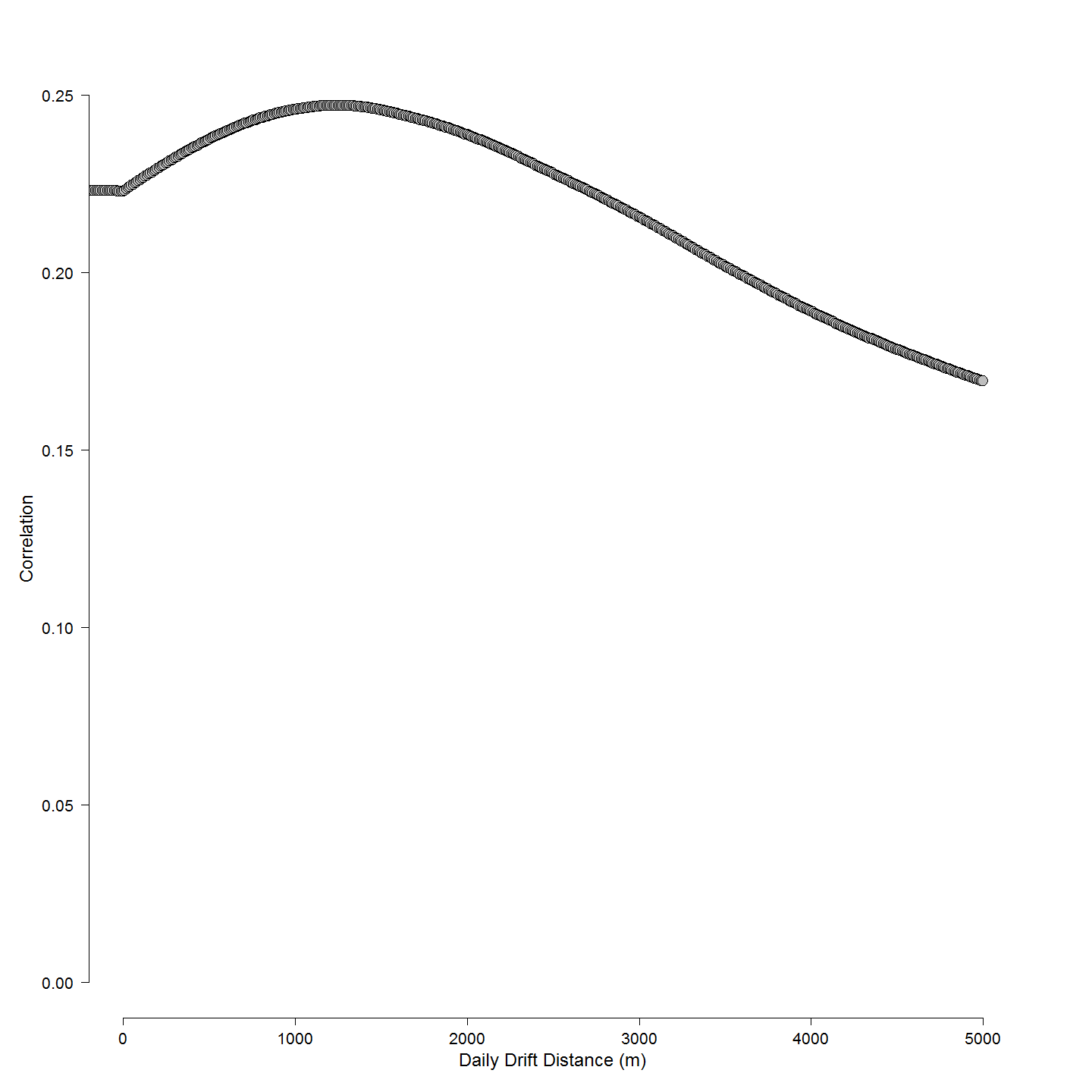
library(Hmisc)  
library(ade4)  
  
#This file is to use the mantel test to look for correlation between genetic distance and geographic distance. In this case by iterating the geographic distance from -2000 to 5000 metres at 10m increments, to find the mean distance that provides the best correlation between the two distance matrices. Given Isolation by Distance, the optimal correlation provides an estimate of the mean dispersal velocity.  
  
# Firstly remove larvae from larv that do not have genetic analysis done.  
# Create a MCsnps set with row names as a column.  
MCchecklist<-row.names(MCsnps)  
MCchecklist<-as.data.frame(MCchecklist)# 93 records  
  
#remove a few more anomolies. These first seven rows were test larvae that are not for inclusion.  
MCchecklist1 <- as.data.frame(MCchecklist[-c(1:7), ])  
  
# Keep every record in larv that is also in MCchecklist (i.e., the intersection).  
larv\_intersection <- larv[larv$Label %in% MCchecklist$MCchecklist,]  
#Thanks: https://heuristically.wordpress.com/2009/10/08/delete-rows-from-r-data-frame/  
  
larv<-larv\_intersection  
rm(larv\_intersection)  
  
# Create GenDist from code in the Murray Cod SNPS table  
MCdm<-MCsnps[-c(1:7),] #remove non-numeric variables  
MCdm <- dist(MCdm) # Create a Murray Cod distance matrix  
MCdm<-as.matrix(MCdm)  
MCdm<-as.data.frame(MCdm)  
#This is one part of the Mantel test given Isolation by Distance. It does not vary inside the iteration (Genetic distance is fixed) while actual geographic distance (dispersal velocity) is the unknown and is recalculated within the iteration.  
  
#First sort MCdm - just to be sure that the two matrices are ordered identically before mantel test.  
MCdm<-as.data.frame(MCdm)  
MCdm$sort<-row.names(MCdm)  
MCdm <- MCdm[order(MCdm$sort),]#sort row order  
MCdm$sort<-NULL  
MCdm<-MCdm[,order(names(MCdm))]#sort column order  
MCdmForSibs<-as.data.frame(MCdm) #need this matrix for sibling analysis .r  
MCdm<-as.matrix(MCdm)  
  
itmant <- matrix(nrow=7001, ncol=3) #Is a 7002 row DF to store result but NA generated in for loop below are omited later.   
  
#Iteration begins here:  
ptm <- proc.time()  
for (nd in seq(-2000,5000, by=10)){# Posible dispersal from 2km upstream to 5 km downstream, 10m increments   
   
 larv$tmpNestDist<-larv$Distance.to.Angle.Crossing..m.-(nd\*(larv$Day.of.Year-(larv$hatchDoY+7)))   
   
 #Create Geographic Distance Matrix using Nest Distance  
 geodist<-data.frame(larv$Label,larv$tmpNestDist)  
 row.names(geodist)<-geodist[,1]  
 geodist$larv.Label<-NULL  
 geodist<-na.omit(geodist)  
 GeoDistMat<-dist(geodist)  
 GeoDistMathm <- as.matrix(GeoDistMat)  
   
 #make sure GeoDist matrix is in correct order - rows and cols  
 GeoDistMathm<-as.data.frame(GeoDistMathm)  
 GeoDistMathm$sort<-row.names(GeoDistMathm)  
 GeoDistMathm <- GeoDistMathm[order(GeoDistMathm$sort),]#sort row order  
 GeoDistMathm$sort<-NULL  
 GeoDistMathm<-GeoDistMathm[,order(names(GeoDistMathm))]#sort column order  
 GeoDistMathm<-as.matrix(GeoDistMathm)  
   
 mant<-mantel.rtest(as.dist(GeoDistMathm), as.dist(MCdm), nrepet = 99)  
 ndr<-nd+2001 #to keep in positive for row numbers in itmant data frame  
 itmant[ndr,] <- c(nd, mant$obs, mant$pvalue)  
}  
  
larv$tmpNestDist<-NULL #tidy up temporary variables  
rm(ndr)  
rm(geodist)  
rm(GeoDistMat)  
rm(GeoDistMathm)  
rm(mant)  
  
proc.time() - ptm

## user system elapsed   
## 53.18 0.10 53.95

itmant<-na.omit(itmant) #remove NAs   
itmantdf<-as.data.frame(itmant) #make into a data frame  
plot(itmantdf$V1,itmantdf$V2,main = "Mantel Correlation Between Pairwise Genetic Distance and Geographic Distance", xlab = "Geographic Distance (metres)", ylab = "Correlation with Genetic Distance Matrix (Mantel Test)",cex.main = 2.5,cex.lab=2)



rm(itmant)  
  
###########  
require(plotrix)  
op <- par(cex.main = 1.5, mar = c(5, 6, 4, 5) + 0.1, mgp = c(3.5, 1, 0), cex.lab = 1.5 , font.lab = 2, cex.axis = 1.3, bty = "n", las=1)  
plot(itmantdf$V1, itmantdf$V2, col="black", pch=21, bg = "grey", cex = 2,  
 xlim=c(0,5000), ylim=c(0,.25), ylab="", xlab="", axes=F)  
axis(1)  
axis(2)   
reg1 <- lm(itmantdf$V2~itmantdf$V1)  
ablineclip(reg1, lwd=2,x1 = .9, x2 = 1.2)   
par(las=0)  
mtext("Daily Drift Distance (m)", side=1, line=2.5, cex=1.5)  
mtext("Correlation", side=2, line=3.7, cex=1.5)



###########  
  
dispersalVelocity<- subset(itmantdf, V2==max(V2) , select = V1)  
dispersalVelocity<-as.numeric(dispersalVelocity)

The mean larval dispersal velocity in a downstream direction from the nest is 1240 metres per day. The distance above the collection site as the larvae disperse at 1240 metres per day between leaving the nest and being collected at sampling site also allows an estimate of the Murray cod pelagic larval drift duration to be made.

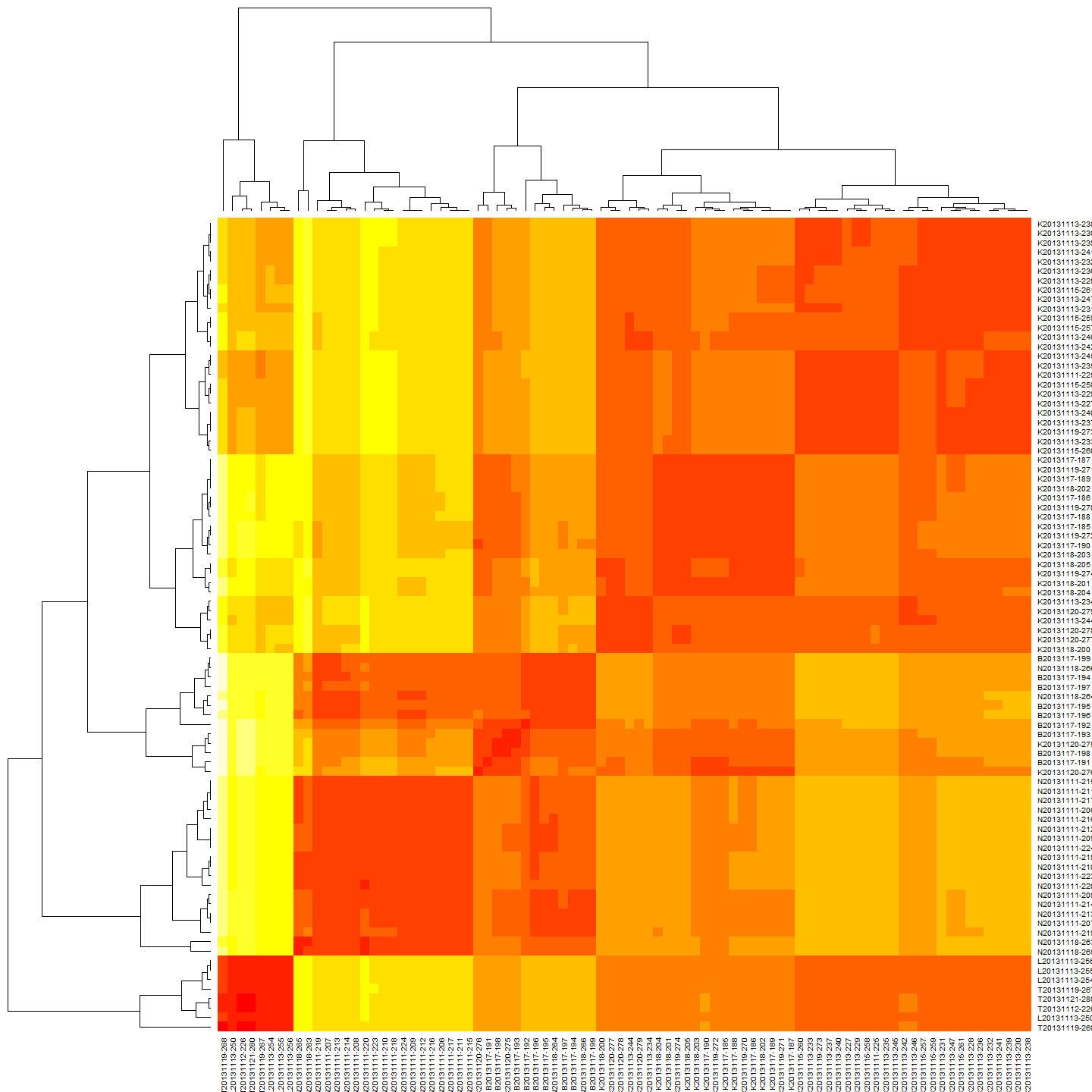
This mean larval dispersal velocity (LDV) value is a mean for larvae along the entire river reach. It tells us nothing about the variance of that dispersal (or can it)??? It is a summary statistic of the velocity of larvae dispersing from the nest, before dispersing larvae are no longer detected in the water column.

The obvious thing to do would be to take the approach on a site by site basis over a number of years, because the river speeds vary at each site and between years but at present there are too few samples from most sites and other years to use such an approach with confidence.

#Recreate the geographic distance matrix with the mean larval dispersal velocity as calculated.  
larv$nestDist<-larv$Distance.to.Angle.Crossing..m.-(dispersalVelocity\*(larv$Day.of.Year-(larv$hatchDoY+7)))  
geodist<-data.frame(larv$Label,larv$nestDist) # Populate databse with this.  
row.names(geodist)<-geodist[,1]  
geodist$larv.Label<-NULL  
geodist<-na.omit(geodist)  
  
GeoDistMat<-dist(geodist)  
GeoDistMathm <- as.matrix(GeoDistMat)  
geoclust<-hclust(GeoDistMat)

Now that these various matrices, class 'dist' objects are created we can make a heat map.

heatmap(GeoDistMathm)



Clearly there is a pattern in the geographic distance matrix. In it self this is unsuprising given the spatial distribution of the sampling regime.

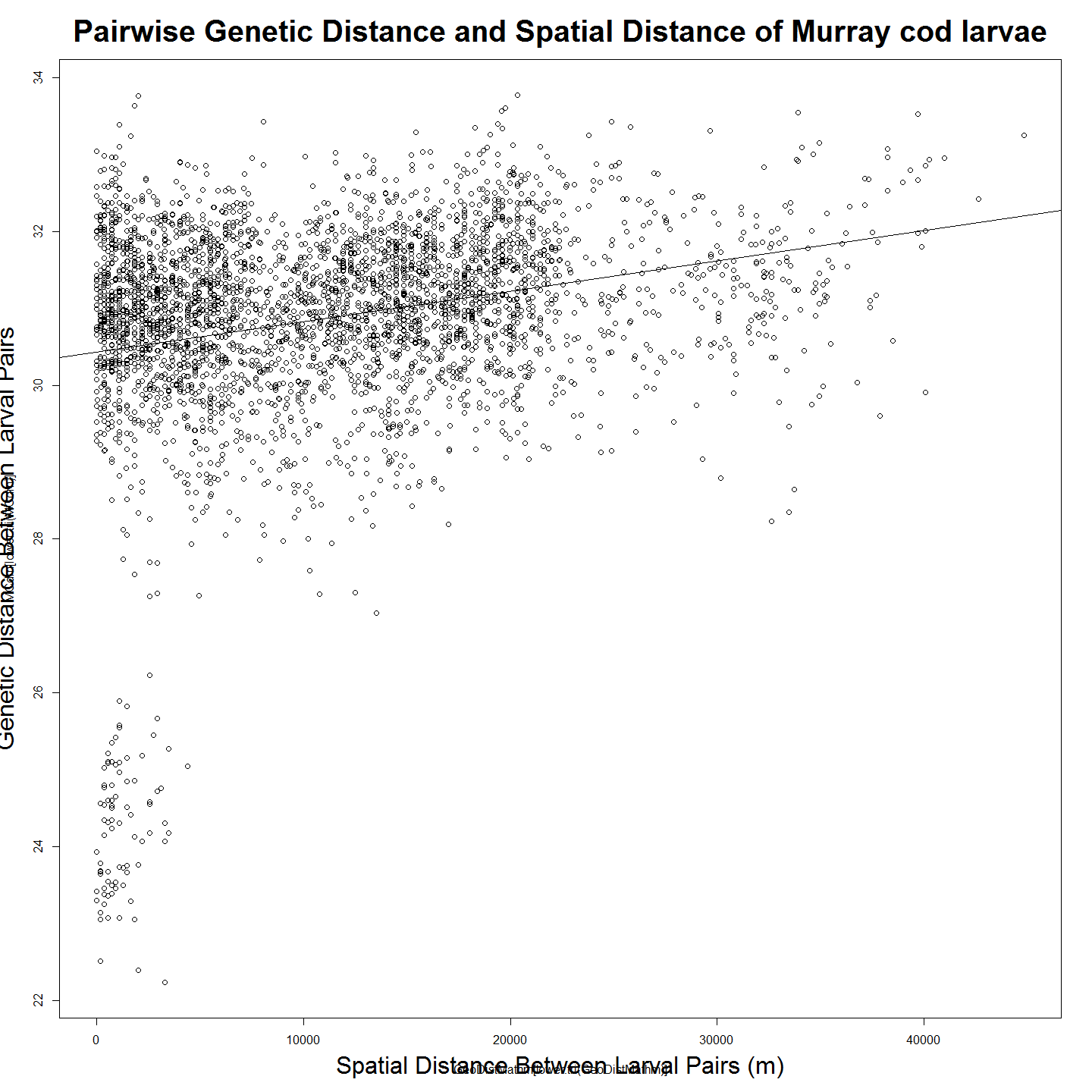
### Plot and Correlate Genetic and Geographic Distance Matrices for Mean Dispersal Velocity

First a regression model is calculated and then the plot:

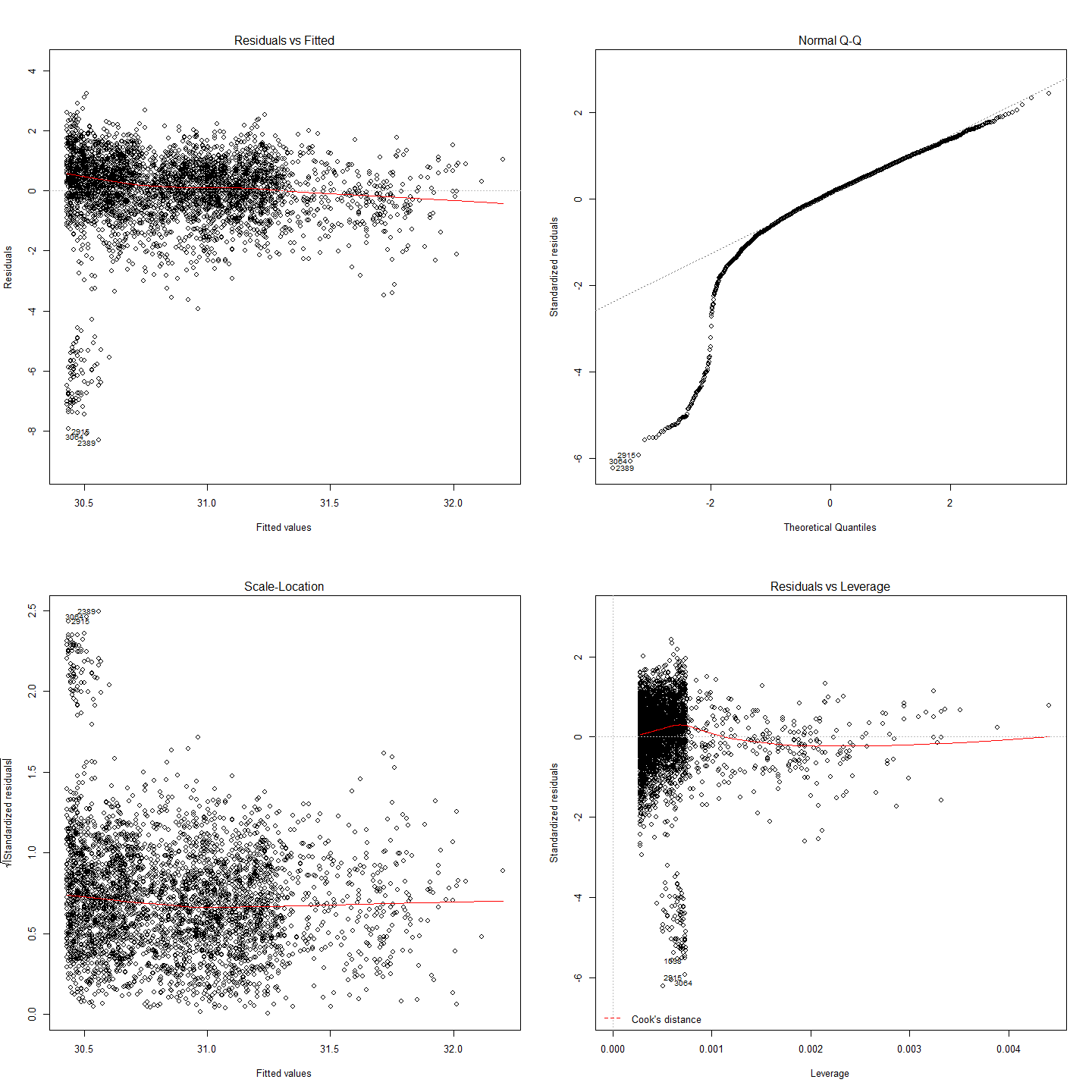
#Linear Regression Model  
reg<-lm(MCdm[lower.tri(MCdm)]~GeoDistMathm[lower.tri(GeoDistMathm)])  
summary(reg) #does xtable do something nice with this?

##   
## Call:  
## lm(formula = MCdm[lower.tri(MCdm)] ~ GeoDistMathm[lower.tri(GeoDistMathm)])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.3228 -0.4726 0.1836 0.7577 3.2562   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.043e+01 3.641e-02 835.82 <2e-16 \*\*\*  
## GeoDistMathm[lower.tri(GeoDistMathm)] 3.949e-05 2.562e-06 15.41 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.338 on 3653 degrees of freedom  
## Multiple R-squared: 0.06106, Adjusted R-squared: 0.06081   
## F-statistic: 237.6 on 1 and 3653 DF, p-value: < 2.2e-16

plot(GeoDistMathm[lower.tri(GeoDistMathm)],MCdm[lower.tri(MCdm)])  
abline(reg)  
title(main="Pairwise Genetic Distance and Spatial Distance of Murray cod larvae", xlab="Spatial Distance Between Larval Pairs (m)", ylab="Genetic Distance Between Larval Pairs", cex.main=2.5, cex.lab=2)



# Correlations with significance levels  
#rcorr(GeoDistMathm[lower.tri(GeoDistMathm)],MCdm[lower.tri(MCdm)])#(x, type="pearson") # type can be pearson or spearman  
  
#ANOVA  
#anova(reg)  
#Plot residuals  
par(mfrow=c(2,2))  
plot(reg)



So there is some small but significant correlation between pairwise genetic distance and pairwise geographic distance in the Murray cod sampled. Isolation by distance is evident, even over this very small spatial scale.

### Mantel Test

mant<-mantel.rtest(as.dist(GeoDistMathm), as.dist(MCdm), nrepet = 999)  
mant

## Monte-Carlo test  
## Observation: 0.2471118   
## Call: mantel.rtest(m1 = as.dist(GeoDistMathm), m2 = as.dist(MCdm),   
## nrepet = 999)  
## Based on 999 replicates  
## Simulated p-value: 0.001

Based on these results, we can reject the null hypothesis that these two matrices, pairwise spatial distance and genetic distance, are unrelated with alpha = 0.001. The observed correlation, 0.2471118, suggests that the matrix entries are positively associated. This means that smaller differences in genotype are generally seen among pairs of larvae that are from nests geographically close to each other, rather than nests which are estimated to be further away from each other. Note that since this test is based on random permutations, the code will always arrive at the same observed correlation but rarely at exactly the same p-value.

The most likely positon of nests can be estimated based on the larval dispersal velocity can now be mapped onto the river topology.

AMOVA: compare catch, nets sites.

## Siblings and What they can tell us about Larval Dispersal

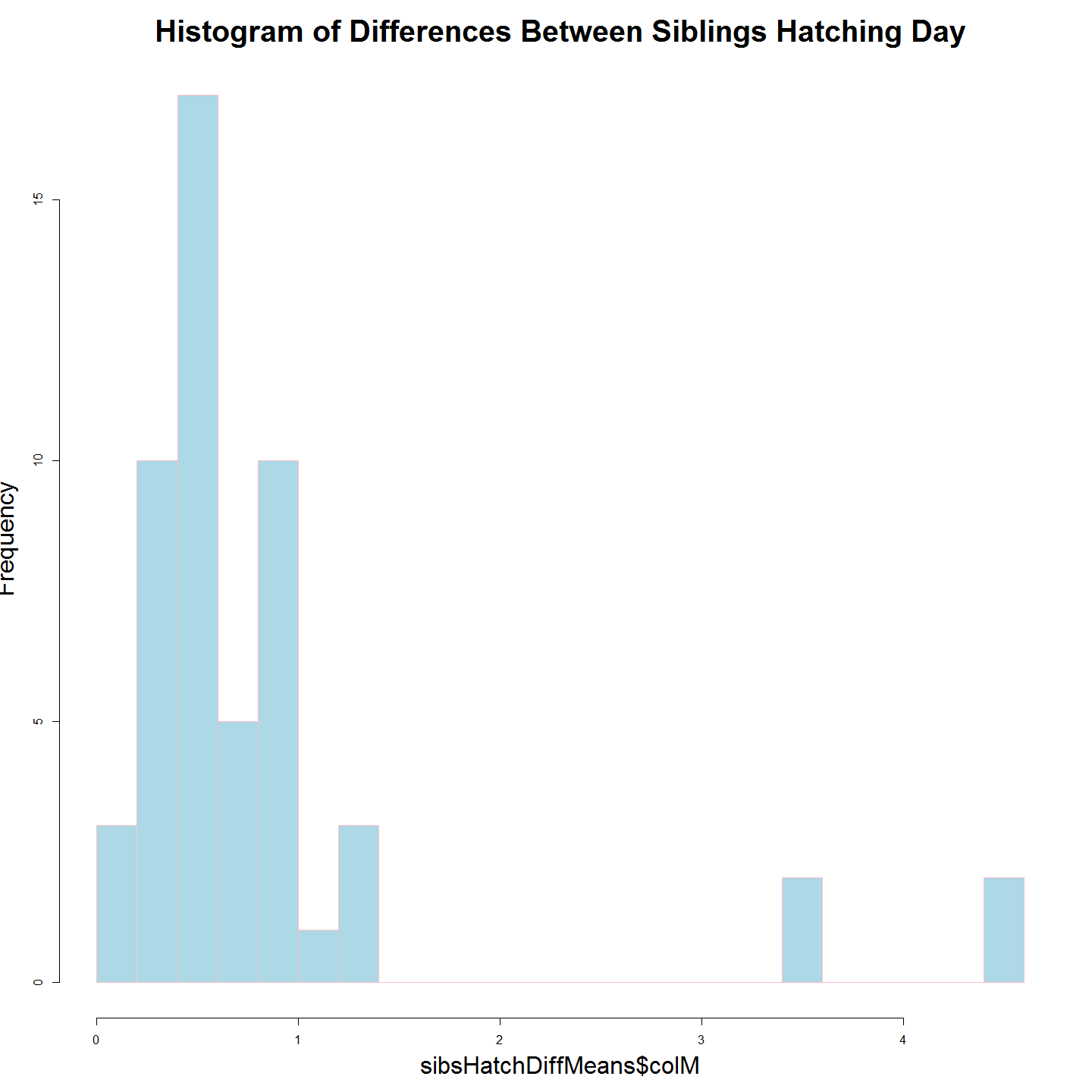
The identification of various sibling pairs in the data (they can be seen in the bottom left of the genetic distance v. distance scatterplot) with a small genetic distance (FST) can potentially be useful in exploring some angles of larval dispersal. For example:

* do siblings dispere in schools or alone?
* refine nest distance estimate to a mean of the known siblings
* How long does the hatch period in the wild last? It is known in hatcheries to last x days but not in the wild. Siblings also facilitate accurate allocation of larvae to nest because, by definition, they come from the same nest. If siblings appear between years then this would suggest repeat adult pairings over years. To date we do not know if this happens.

## This file is to examine the siblings collected that are known to be siblings from DaRT sequencing. The idea is to learn:  
# 1. What duration is hatching in wild nests?   
# 2. Create a list of larvae with siblings (withSibs) and maybe (withoutSibs?) for checking  
# 3. Create a logical matrix of siblinf relationships to re-use (sibsLogical)  
  
#These data can then be used to inform Nest assignments in cases where clade and location are the same.  
  
#Line to extract subset of DF which include at least one sibling in the dataframe  
sibsSubset<-subset(MCdmForSibs, apply(MCdmForSibs, 1, function(MCdmForSibs){any(MCdmForSibs > 0 & MCdmForSibs < 26.5)}))  
#The siblings have a dissimilarity index of less than 26.5. This can be seen in the dendrogram and in the scatterplot for IBD.  
  
sibsLogical<-sibsSubset<26.5 #a logical DF to record T and F for sibling pairs  
sibsSubset<-sibsLogical\*sibsSubset #This turns to 0 all the genetic similarity above 26.5 (no-sibs)  
  
sibsSubset<-sibsSubset[!sapply(sibsSubset, function(x) all(x == 0))] #now remove those with no sibs  
sibsLogical<-sibsSubset>0  
  
#Make a list (data frame) with names of those with sibs (same nest)  
withSib<-row.names(sibsSubset) #make a list of those with sibs  
withSibDF<-as.data.frame(withSib)  
row.names(withSibDF)<-withSibDF[,1] # Make sensible row names (larva labels)  
withSibDF<-as.data.frame(withSibDF) #make back to a df  
  
#Elsewhere I have Make a list (data frame) with names of those withOUT sibs (as each means must be a different NestID)  
  
##Now to see how many days between hatch of sibs.  
  
sibHatchPeriod<-subset(larv, select=c(Label,hatchDoY)) #get labels and hatch DoY  
sibHatchPeriod<-sibHatchPeriod[sibHatchPeriod$Label %in% withSib,] #reduce df to the siblings only  
  
row.names(sibHatchPeriod)<-sibHatchPeriod[,1] #make label row names for making distance matrix  
sibHatchPeriod$Label<-NULL #tidy up redundant  
  
#Then to make a distance matrix.  
sibHatchPeriodDistMatrix<-dist(sibHatchPeriod)  
sibHatchPeriodDistMatrix<-as.matrix(sibHatchPeriodDistMatrix)  
  
###sort rows and columns of both before we can do matrix algebra  
sibHatchPeriodDistMatrix<-as.data.frame(sibHatchPeriodDistMatrix)  
sibHatchPeriodDistMatrix$sort<-row.names(sibHatchPeriodDistMatrix)  
sibHatchPeriodDistMatrix <- sibHatchPeriodDistMatrix[order(sibHatchPeriodDistMatrix$sort),]#sort row order  
sibHatchPeriodDistMatrix$sort<-NULL  
sibHatchPeriodDistMatrix<-sibHatchPeriodDistMatrix[,order(names(sibHatchPeriodDistMatrix))]#sort column order  
sibHatchPeriodDistMatrix<-as.matrix(sibHatchPeriodDistMatrix)  
  
sibsLogical<-as.data.frame(sibsLogical)  
sibsLogical$sort<-row.names(sibsLogical)  
sibsLogical <- sibsLogical[order(sibsLogical$sort),]#sort row order  
sibsLogical$sort<-NULL  
sibsLogical<-sibsLogical[,order(names(sibsLogical))]#sort column order  
sibsLogical<-as.matrix(sibsLogical)  
####  
  
sibsHatchDiff<-sibsLogical\*sibHatchPeriodDistMatrix # need half the triangle here? as mean lowered otehrwise?  
sibsHatchDiff[sibsHatchDiff == 0] <- NA   
sibsHatchDiffMeans<-as.data.frame(colMeans(sibsHatchDiff, na.rm=TRUE))  
  
#Get rid of NAs  
sibsHatchDiffMeans<-na.omit(sibsHatchDiffMeans)  
  
#Output something  
write.csv(format(sibsHatchDiffMeans), file="sibsHatchDiffMeans.csv")  
  
#Describe and Plot Summary Statistics regading the difference in hatch day-of-year between siblings  
describe(sibsHatchDiffMeans)

## sibsHatchDiffMeans   
##   
## 1 Variables 53 Observations  
## --------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------  
## colMeans(sibsHatchDiff, na.rm = TRUE)   
## n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95   
## 53 0 28 1 0.8488 0.1923 0.2958 0.4030 0.5916 0.8874 1.3311 3.4505   
##   
## lowest : 0.1479 0.2218 0.2958 0.3780 0.3858, highest: 1.0564 1.3311 1.3874 3.4505 4.5142   
## --------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#sibsHatchDiffMeans # was just printed for validation  
names(sibsHatchDiffMeans)[names(sibsHatchDiffMeans)=="colMeans(sibsHatchDiff, na.rm = TRUE)"] <- "colM"  
hist(sibsHatchDiffMeans$colM,breaks = 25, main="Histogram of Differences Between Siblings Hatching Day", col = "lightblue", border = "pink",cex.main = 2.5,cex.lab=2)



#So the three objective have been achieved.   
# 1. What duration is hatching in wild nests? Description above showns mean etc.  
# 2. Create a list of larvae with siblings (withSibsDF) done  
# 3. Create a logical matrix of siblinf relationships to re-use (sibsLogical) done  
  
#Clean Up  
#rm(sibsSubset)  
rm(sibsHatchDiff)  
rm(sibsHatchDiffMeans)  
#rm(sibsLogical)  
rm(withSib)  
rm(sibHatchPeriod)  
rm(withSibDF)  
rm(sibHatchPeriodDistMatrix)  
rm(MCdmForSibs)

The period over which the hatch occurs for siblings varies from 0 to 4.5 days with a mean of 0.85 (should make this inline r code though)

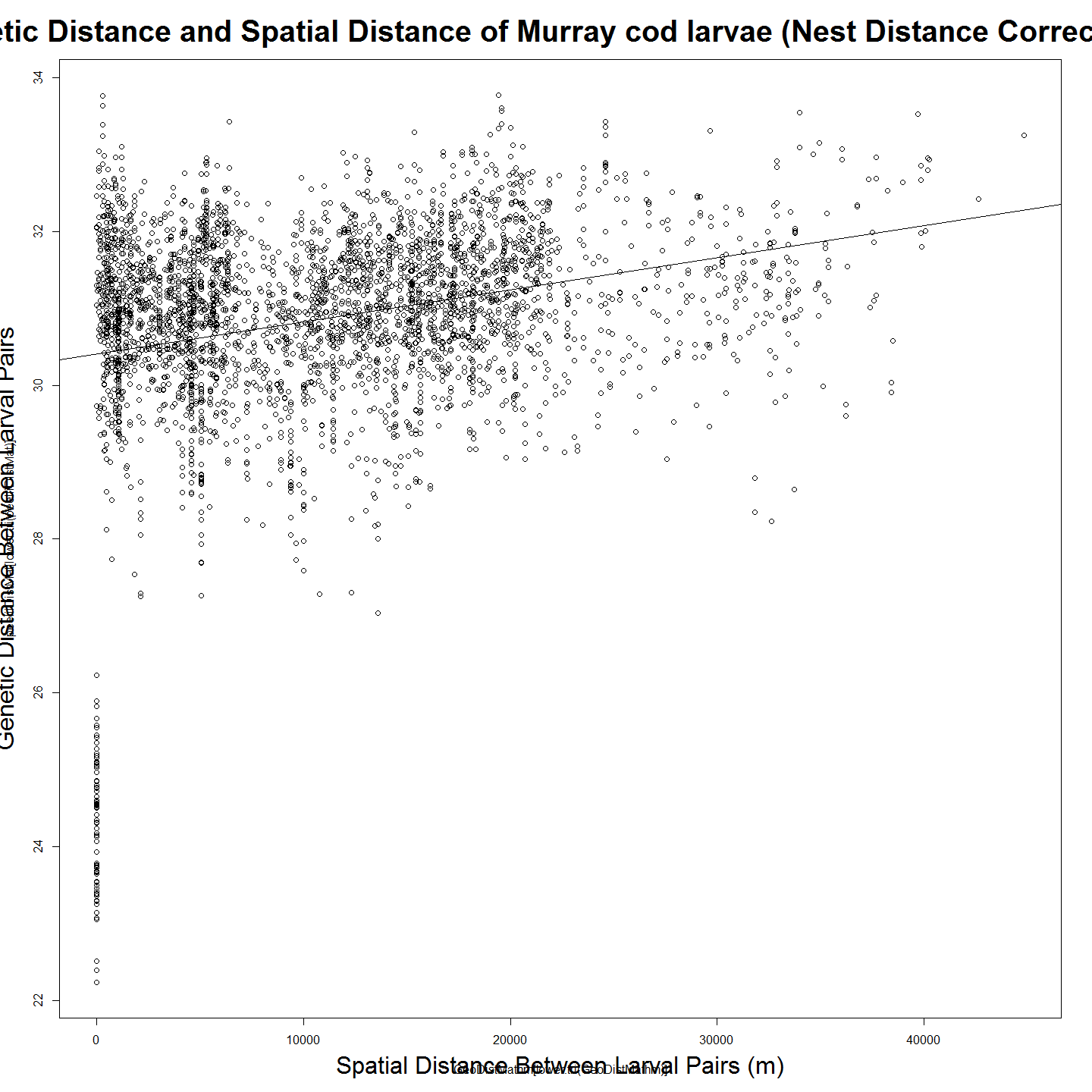
##Calculate a family group  
families<-as.data.frame(cutree(peeliiDendro, h = 26.5))  
row.names(larv)<-larv$Label  
larv<-merge(larv,families, by="row.names", all.x=TRUE)  
larv$Row.names<-NULL  
  
##Create a data frame from which to develop nestID names (not yet combining siblings)  
larv$roundedHatchDoY<-round(larv$hatchDoY)  
larv$nestSeg<-round(larv$nestDist/100)  
larv$siteLetter<-substr(larv$Label, 1, 1)  
names(larv)[names(larv)=="cutree(peeliiDendro, h = 26.5)"] <- "family"  
  
famNo<-unique(larv$family)  
famNo<-na.omit(famNo)  
  
##Initialise a data frame for use in for loop  
final <- data.frame(Date=as.Date(character()),  
 File=character(),   
 User=character(),   
 stringsAsFactors=FALSE)   
  
##Loop through calculating means (corrected nestDist)  
for (nn in famNo){  
 newsub<-subset(larv,family==nn, select=c(LarvalID,Label,nestDist,hatchDoY,incubTime))  
 newsub$sibCorrectedNestDist<-paste(mean(as.numeric(newsub$nestDist)))  
 final<-rbind(final,newsub)  
 }

### Using a sibsCorrectedNestDistance. Can this give us an idea of variance back calculating to dispersal velocity?

# Calculate Mantel based on Sibling Corrected Distance  
  
geodist<-data.frame(final$Label,final$sibCorrectedNestDist)  
row.names(geodist)<-geodist$final.Label  
geodist$final.Label<-NULL  
geodist<-na.omit(geodist)  
  
GeoDistMat<-dist(geodist)  
GeoDistMathm <- as.matrix(GeoDistMat)  
#heatmap(GeoDistMathm)  
geoclust<-hclust(GeoDistMat)  
#plot(geoclust)  
  
#make sure both matrices are in correct order - rows and cols  
#First sort MCdm  
  
peeliiDistMat<-as.data.frame(peeliiDistMat)  
peeliiDistMat$sort<-row.names(peeliiDistMat)  
peeliiDistMat <- peeliiDistMat[order(peeliiDistMat$sort),]#sort row order  
peeliiDistMat$sort<-NULL  
peeliiDistMat<-peeliiDistMat[,order(names(peeliiDistMat))]#sort column order  
peeliiDistMat<-as.matrix(peeliiDistMat)  
  
#Second sort GeoDist  
GeoDistMathm<-as.data.frame(GeoDistMathm)  
GeoDistMathm$sort<-row.names(GeoDistMathm)  
GeoDistMathm <- GeoDistMathm[order(GeoDistMathm$sort),]#sort row order  
GeoDistMathm$sort<-NULL  
GeoDistMathm<-GeoDistMathm[,order(names(GeoDistMathm))]#sort column order  
GeoDistMathm<-as.matrix(GeoDistMathm)  
  
###Plot and Correlate genetic and geographic distance matrices  
  
#First a regression model is calculated and then the plot:  
  
#Linear Regression Model  
reg<-lm(peeliiDistMat[lower.tri(peeliiDistMat)]~GeoDistMathm[lower.tri(GeoDistMathm)])  
summary(reg)

##   
## Call:  
## lm(formula = peeliiDistMat[lower.tri(peeliiDistMat)] ~ GeoDistMathm[lower.tri(GeoDistMathm)])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.1704 -0.4767 0.1852 0.7554 3.3464   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.041e+01 3.605e-02 843.50 <2e-16 \*\*\*  
## GeoDistMathm[lower.tri(GeoDistMathm)] 4.171e-05 2.544e-06 16.39 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.333 on 3653 degrees of freedom  
## Multiple R-squared: 0.06851, Adjusted R-squared: 0.06825   
## F-statistic: 268.7 on 1 and 3653 DF, p-value: < 2.2e-16

plot(GeoDistMathm[lower.tri(GeoDistMathm)],peeliiDistMat[lower.tri(peeliiDistMat)])  
abline(reg)  
title(main="Pairwise Genetic Distance and Spatial Distance of Murray cod larvae (Nest Distance Corrected for Siblings)", xlab="Spatial Distance Between Larval Pairs (m)", ylab="Genetic Distance Between Larval Pairs", cex.main=2.5, cex.lab=2)



# Correlations with significance levels  
#rcorr(GeoDistMathm[lower.tri(GeoDistMathm)],peeliiDistMat[lower.tri(peeliiDistMat)])#(x, type="pearson") # type can be pearson or spearman  
  
###Mantel Test Using Sibling Corrected Nest Distance  
mant<-mantel.rtest(as.dist(GeoDistMathm), as.dist(peeliiDistMat), nrepet = 9999)  
  
##This results in a slightly better r value and Monty Carlo test so it seems appropriate that sibling corrected nest distances are used in the iterated mantel rather than the uncorrected ones. That said the difference, if only, is expected to only be slight.

### Variance in Estimation of Nest Distance Using Larval Dispersal Velocity

Because we know siblings Nest Distance is the same, that is a value of the mean nest distance for the siblings as calculated. But because we know there is a zero difference in actual nest distance between siblings we can calculate an error associated with the estimate derived from the above method. This error is likely to be a suitable proxy for the error of the non-siblings whose nest distances were estimated with the same method.

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## 1 1 53 727.8 637.29 458.48 650.4 407.84 0 2292.39 2292.39 0.98 -0.31 87.54

### Use freq histogram of age to adjust dispersal velocity?

ie: if most fish are sampled at 12 days, few at 7 or 20 then most of the dispersal happens then (doesn’t it?) so dispersal may go up from 1200 in near mean age, and below 1200 at ends of the curve. Therefore the distance calculations can – at least in theory – be adjusted based on their age (proportionally). If I do that does it change the variance in the distribution of (sibs?) larvae nest positions? Ie: does the estimate improve?

### Now to calculate the largest potential settling distance if they moved at mean larval dispersal velocity (LDV) for more of their remaining drift days.

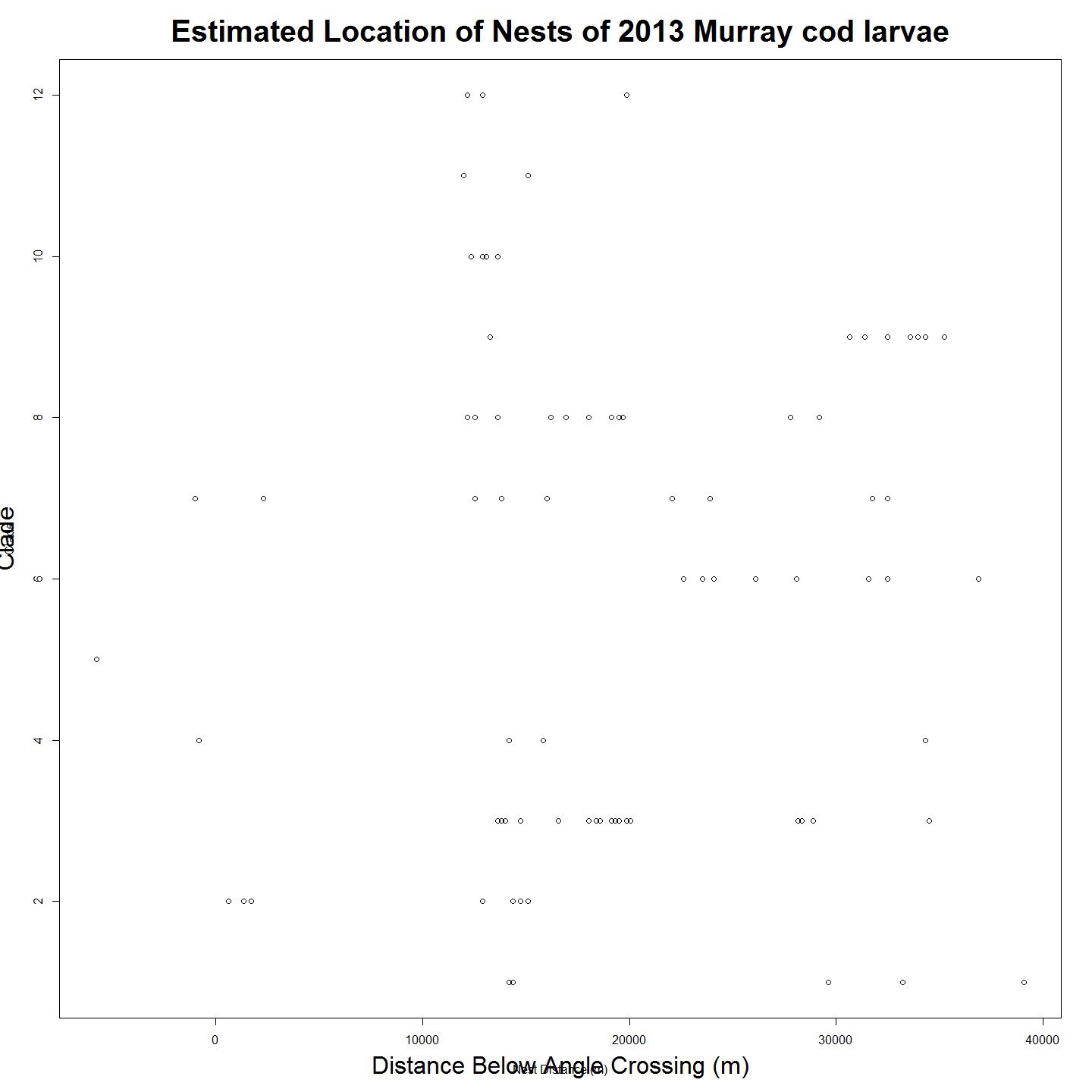
larv$potentialSettlingDistance<-larv$Distance.to.Angle.Crossing..m.+(20-larv$ageOL)\*dispersalVelocity  
summary(larv$potentialSettlingDistance)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10360 29790 34970 35930 44860 55220

### Look at Clades Over the River Reach

The clades distribution over the river reach suggests some structure.

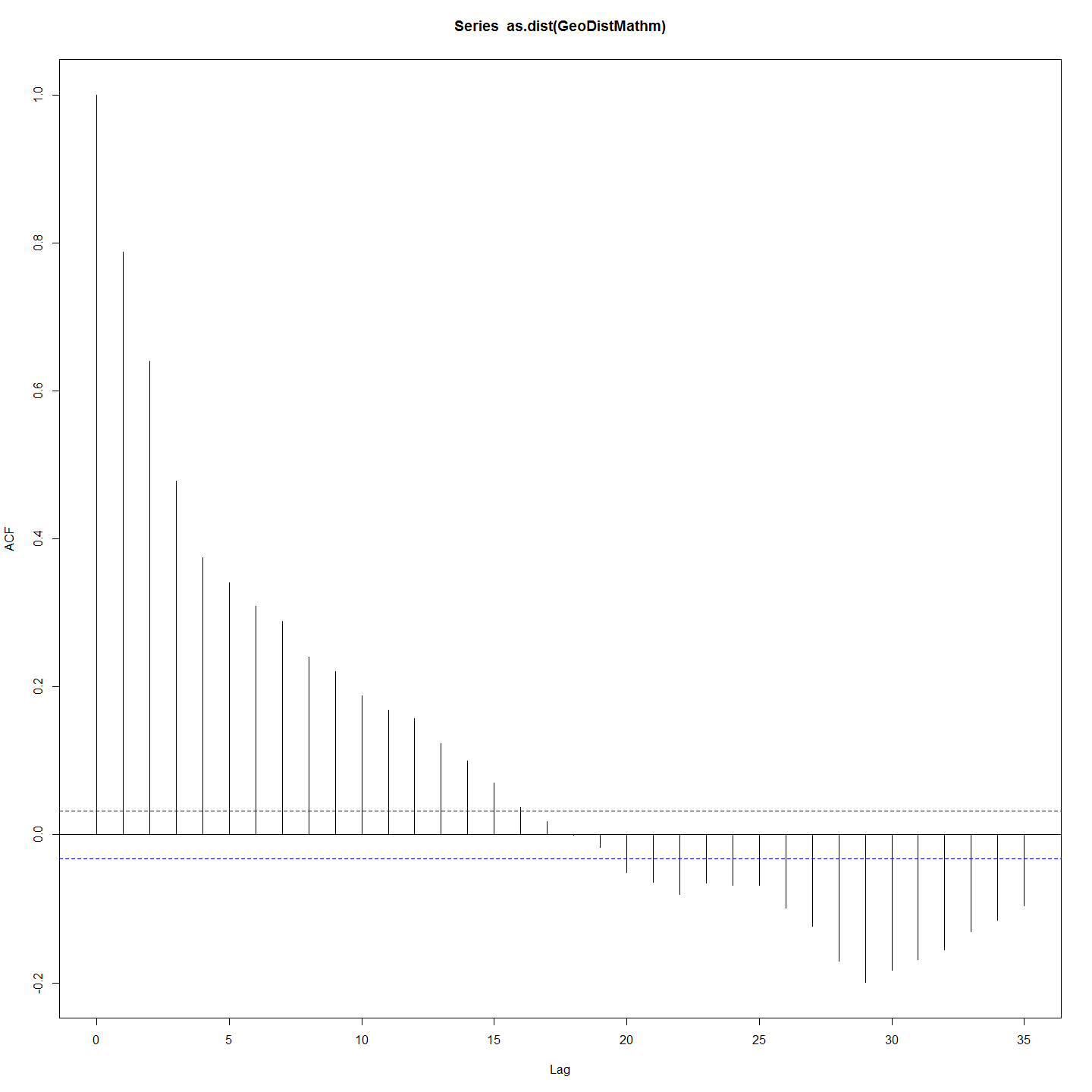
plot(larv$nestDist,larv$cladeNumber, xlab="Nest Distance (m)", ylab = "Clade")  
title(main="Estimated Location of Nests of 2013 Murray cod larvae", xlab="Distance Below Angle Crossing (m)", ylab="Clade", cex.main=2.5, cex.lab=2)



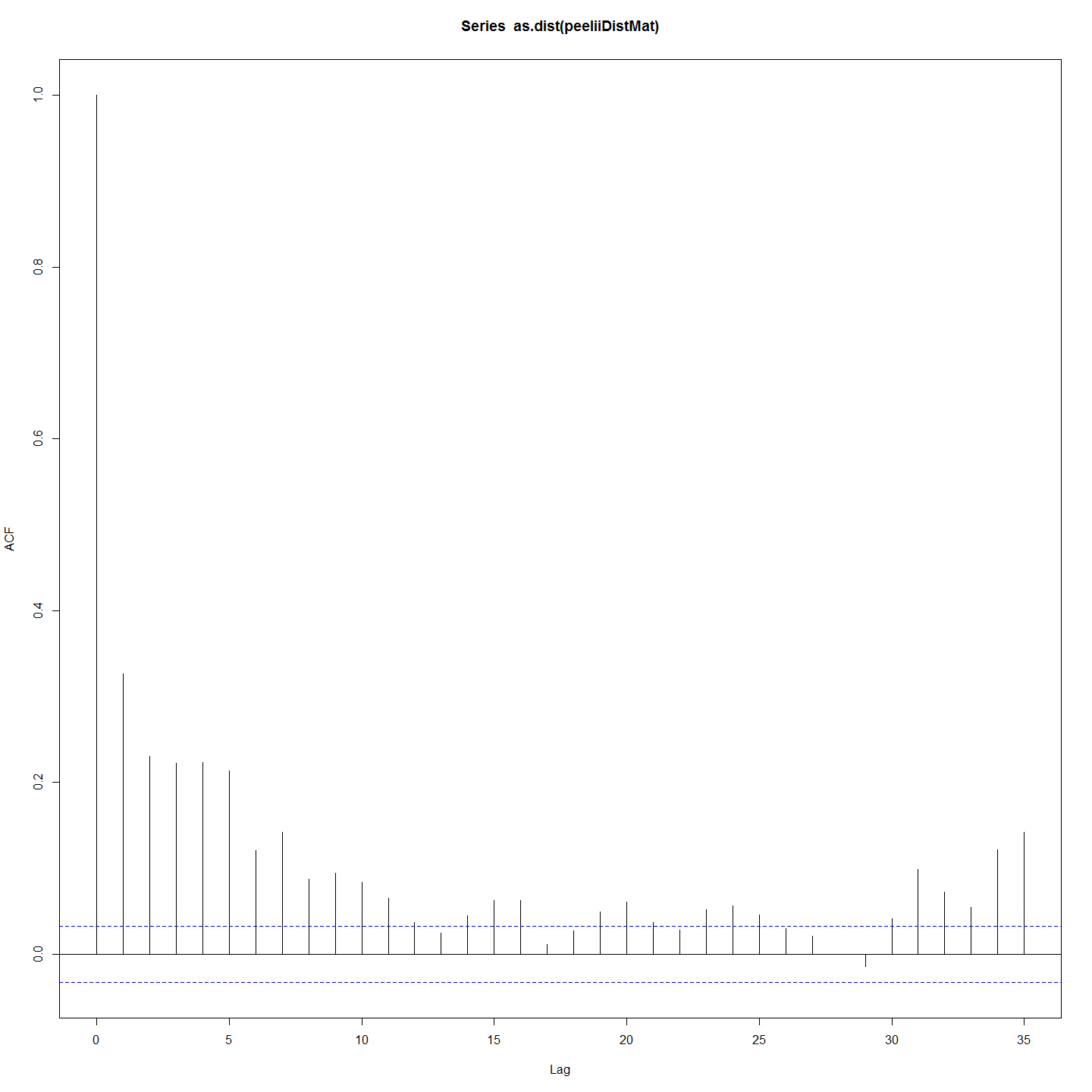
#plot2 <- ggplot(larv1, aes(nestDist,CladeName))  
#plot2 + geom\_point(alpha = 2/3) +geom\_boxplot()+labs(title = "Clades Nests and their Position on River")+ labs(x="Nest Location") +labs(y = "Clade")+ theme(axis.text.x=element\_text(angle=90))

# Autocorrelation Test? How to interpret this?

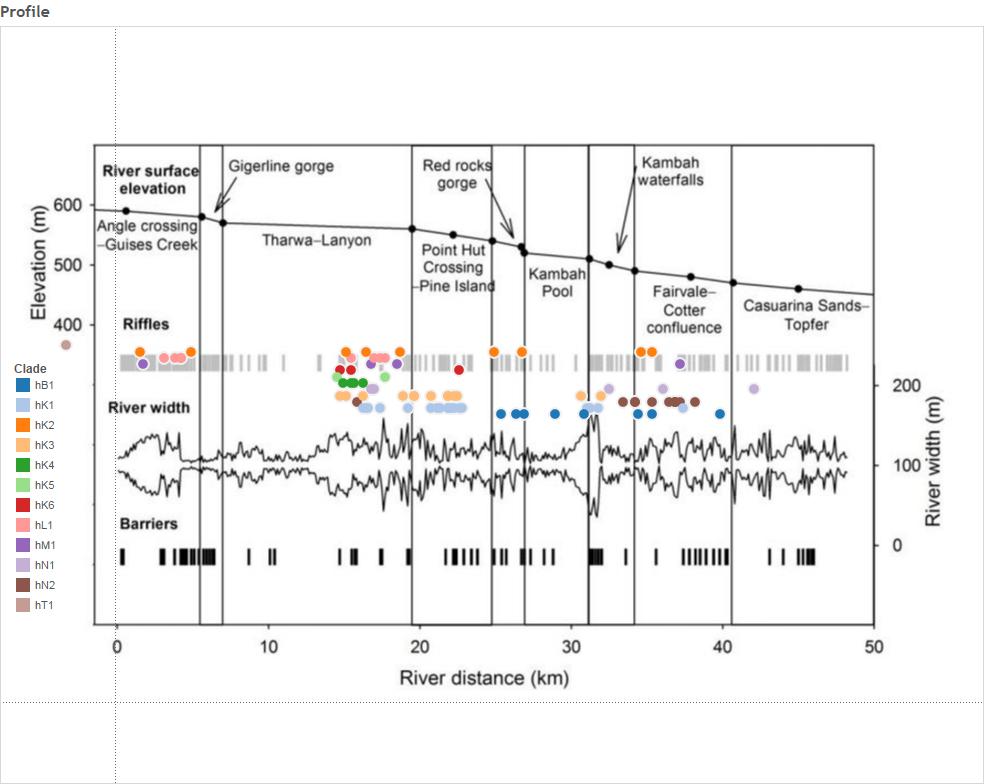
acf(as.dist(GeoDistMathm))



acf(as.dist(peeliiDistMat))

 ##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).
* So nest site is a much better 'site' than the capture site which confounds age and other dispersal variables. Nest will be used in subsequent chapters.



needs updating in Tableau to ensure commensurate with r.

## Code Chunks in this Document

## [1] "Project\_Template\_and\_Knitr" "Set\_Global\_Options" "LoadLibraries"   
## [4] "unnamed-chunk-1" "unnamed-chunk-2" "Iterated Mantel"   
## [7] "Distance\_Matrices\_and\_Ordering" "Spatial Distance Matrix Heatmap" "Plots\_and\_Correlation"   
## [10] "Mantel\_Test" "siblingAnalysisGo" "Dispersal Distances and Nest Location Variables"  
## [13] "unnamed-chunk-3" "nestDistanceDifferenceError" "unnamed-chunk-4"   
## [16] "Clades\_and\_Location" "Autocorrelation" "Include\_Chunk\_Labels\_and\_Session Information"

## R version 3.2.2 (2015-08-14)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 7 x64 (build 7601) Service Pack 1  
##   
## 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] grid stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] tidyr\_0.3.1 psych\_1.5.8 ade4\_1.7-2 Hmisc\_3.17-0 Formula\_1.2-1 survival\_2.38-3 lattice\_0.20-33 dendextend\_1.1.0 ape\_3.3   
## [10] ggdendro\_0.1-17 ggplot2\_1.0.1 ProjectTemplate\_0.6 knitr\_1.11   
##   
## loaded via a namespace (and not attached):  
## [1] Rcpp\_0.12.1 formatR\_1.2.1 RColorBrewer\_1.1-2 plyr\_1.8.3 tools\_3.2.2 rpart\_4.1-10 digest\_0.6.8 evaluate\_0.8 nlme\_3.1-122   
## [10] gtable\_0.1.2 DBI\_0.3.1 parallel\_3.2.2 yaml\_2.1.13 proto\_0.3-10 gridExtra\_2.0.0 dplyr\_0.4.3 stringr\_1.0.0 cluster\_2.0.3   
## [19] nnet\_7.3-11 R6\_2.1.1 foreign\_0.8-66 rmarkdown\_0.8 latticeExtra\_0.6-26 reshape2\_1.4.1 magrittr\_1.5 whisker\_0.3-2 scales\_0.3.0   
## [28] htmltools\_0.2.6 MASS\_7.3-44 splines\_3.2.2 assertthat\_0.1 mnormt\_1.5-3 colorspace\_1.2-6 stringi\_0.5-5 acepack\_1.3-3.3 lazyeval\_0.1.10   
## [37] munsell\_0.4.2