Phylogenetically Independent Constrasts

PACKAGES

```
library(ape)
library(phytools)
```

```
## Loading required package: maps
```

CHALLENGE PROBLEM 1

Fit a OLS (non-phylogenetic) regression model for y~x and a phylogenetic regression model using contrasts. What do you find? Why?

OLS regression model:

```
obj<-read.csv("pic-exercise-data.csv",row.names=1); head(obj)
```

```
##               x         y
## t64  0.8420183 -1.375677
## t4  -0.7992242 -3.132304
## t14 -0.7434011 -3.211733
## t18 -0.9669659 -3.737145
## t92 -0.7723297 -3.386689
## t35 -0.1333343 -2.110647
```

```
fit.ols<-lm(y~x,data=obj); fit.ols
```

```
##
## Call:
## lm(formula = y ~ x, data = obj)
##
## Coefficients:
## (Intercept)            x
##    -1.95546     -0.02282
```

```
summary(fit.ols)
```

```
##
## Call:
## lm(formula = y ~ x, data = obj)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.70042 -0.18339  0.04348  0.41484  0.95401
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.95546    0.12017 -16.272   <2e-16 ***
## x           -0.02282    0.06582  -0.347     0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6289 on 98 degrees of freedom
```

```
## Multiple R-squared:  0.001225,    Adjusted R-squared:  -0.008967
## F-statistic: 0.1202 on 1 and 98 DF,  p-value: 0.7296
```

```r
plot(obj[,c("x","y")])
abline(fit.ols,lwd=2,lty="dashed",col="red")
```
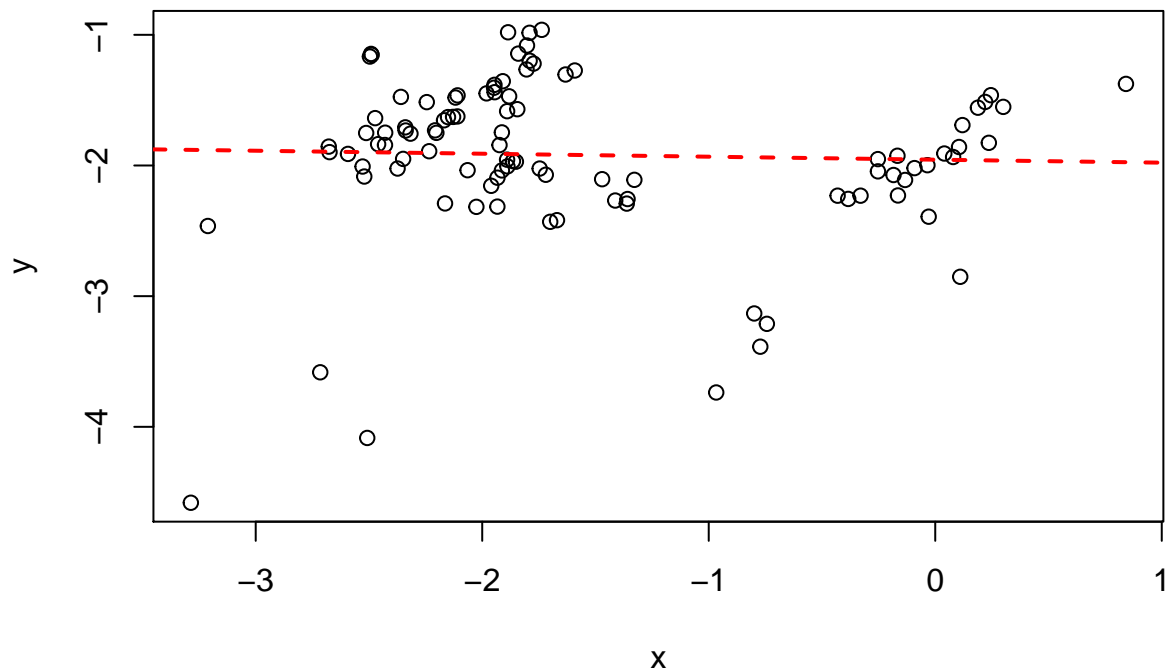

Phylogenetic regression model using contrasts:

```r
ex.tree<-read.tree("pic-exercise-tree.tre")
x.data<-setNames(obj[,"x"],rownames(obj))
y.data<-setNames(obj[,"y"],rownames(obj))
pic.x<-pic(x.data,ex.tree) #categorize buccal length by phylogeny
pic.y<-pic(y.data,ex.tree) #categorize gape width by phylogeny
fit.pic<-lm(pic.y~pic.x+0) #fit linear model with phylogeny considered
fit.pic
```

```
##
## Call:
## lm(formula = pic.y ~ pic.x + 0)
##
## Coefficients:
##  pic.x
## 0.7039
```

```r
summary(fit.pic)
```

```
##
## Call:
## lm(formula = pic.y ~ pic.x + 0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5957 -0.7257  0.2020  0.7100  2.3120
##
## Coefficients:
```

```
##          Estimate Std. Error t value Pr(>|t|)
## pic.x  0.70387    0.09648   7.296 7.81e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9944 on 98 degrees of freedom
## Multiple R-squared:  0.352,  Adjusted R-squared:  0.3454
## F-statistic: 53.23 on 1 and 98 DF,  p-value: 7.808e-11
```

Plot linear model over scatterplot with phylogeny considered:

```
plot(pic.x,pic.y,xlab="PICs for x",ylab="PICs for y")
abline(fit.pic,lwd=2,lty="dashed",col="red")
```
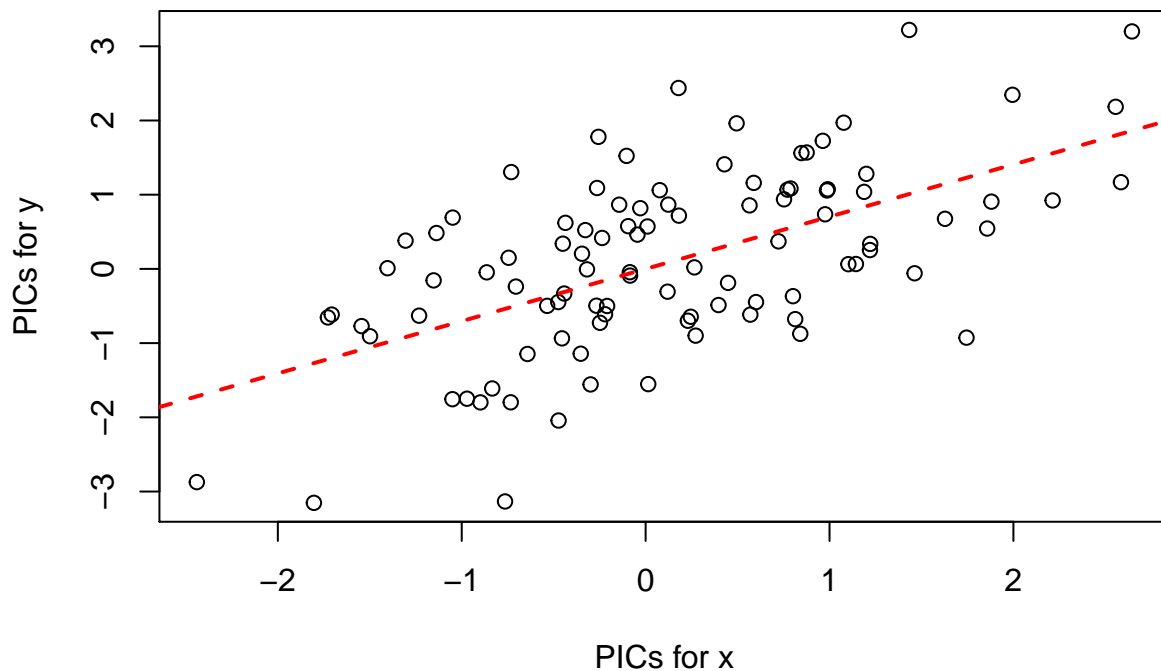

TUTORIAL

Linear Model Basics

Fitting linear regressions to first gape width and buccal length. View Data:

```
obj<-read.csv("Centrarchidae.csv",row.names=1); head(obj)
```

```
##                      feeding.mode gape.width buccal.length
## Acantharchus_pomotis         pisc      0.114        -0.009
## Lepomis_gibbosus              non     -0.133        -0.009
## Lepomis_microlophus           non     -0.151         0.012
## Lepomis_punctatus             non     -0.103        -0.019
## Lepomis_miniatus              non     -0.134         0.001
## Lepomis_auritus               non     -0.222        -0.039
```

Plot effect of buccal length on gape width

```
plot(obj[,c("buccal.length","gape.width")])
```
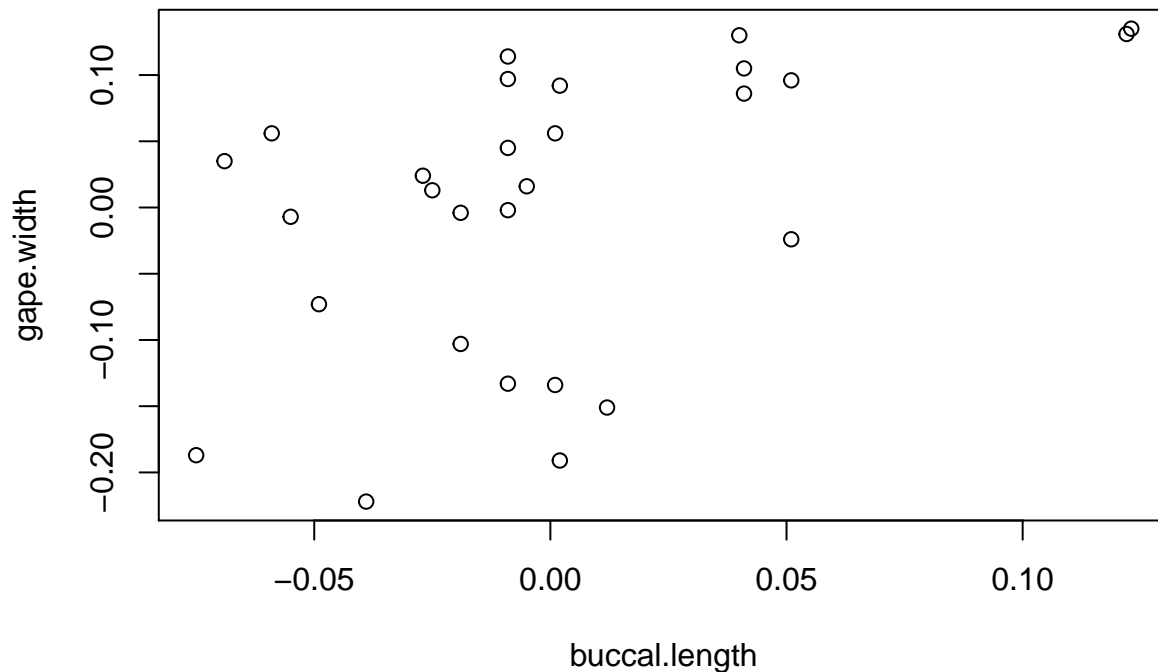
Fit OLS regression model:

```
fit.ols<-lm(gape.width~buccal.length,data=obj); fit.ols
```

```
##
## Call:
## lm(formula = gape.width ~ buccal.length, data = obj)
##
## Coefficients:
##   (Intercept)  buccal.length
##     -3.817e-05      1.069e+00
```

```
summary(fit.ols)
```

```
##
## Call:
## lm(formula = gape.width ~ buccal.length, data = obj)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19310 -0.07951  0.03057  0.05653  0.12366
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.817e-05  1.832e-02  -0.002  0.99835
## buccal.length  1.069e+00  3.843e-01   2.781  0.00995 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09695 on 26 degrees of freedom
## Multiple R-squared:  0.2292, Adjusted R-squared:  0.1996
## F-statistic: 7.733 on 1 and 26 DF,  p-value: 0.009951
```

Add OLS regression line to graph:

```
plot(obj[,c("buccal.length","gape.width")])
abline(fit.ols,lwd=2,lty="dashed",col="red")
```
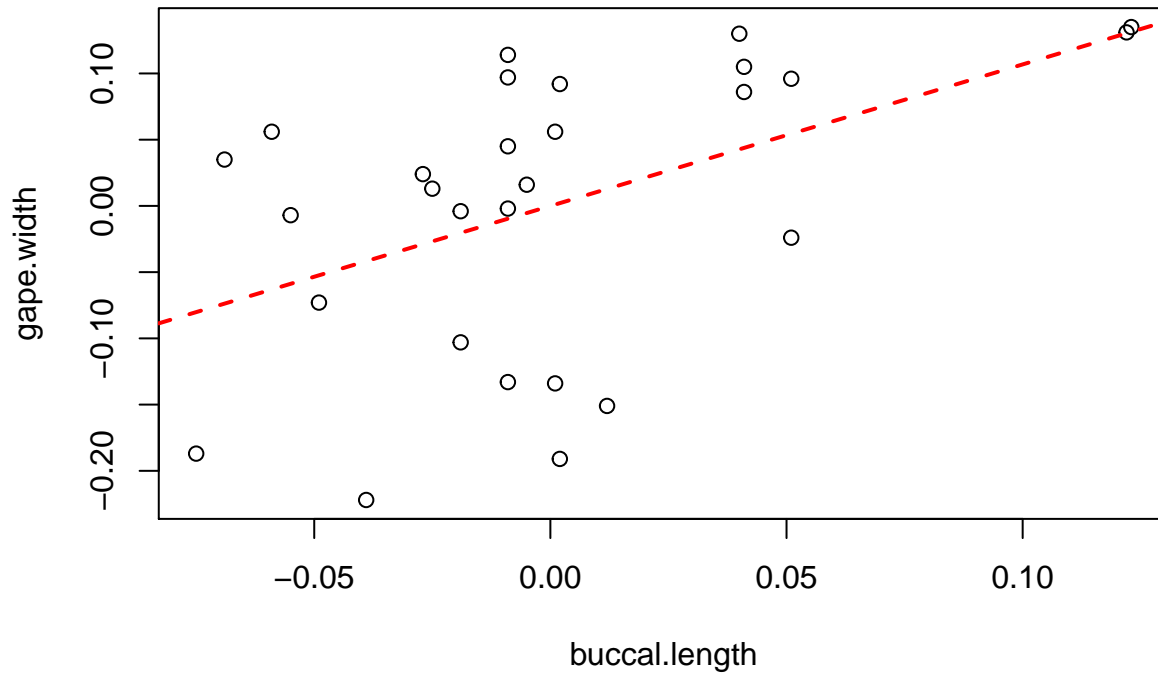

Fit PIC to take into account phylogeny:

```
cent.tree<-read.tree("Centrarchidae.tre")
buccal.length<-setNames(obj[,"buccal.length"],rownames(obj))
gape.width<-setNames(obj[,"gape.width"],rownames(obj))
pic.bl<-pic(buccal.length,cent.tree) #categorize buccal length by phylogeny
pic.gw<-pic(gape.width,cent.tree) #categorize gape width by phylogeny
fit.pic<-lm(pic.gw~pic.bl+0) #fit linear model with phylogeny considered
fit.pic
```

```
##
## Call:
## lm(formula = pic.gw ~ pic.bl + 0)
##
## Coefficients:
## pic.bl
## 0.5932
```

```
summary(fit.pic)
```

```
##
## Call:
## lm(formula = pic.gw ~ pic.bl + 0)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.58401 -0.22945 -0.03814  0.12947  0.99784
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## pic.bl   0.5932     0.2556   2.321   0.0284 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3189 on 26 degrees of freedom
## Multiple R-squared:  0.1716, Adjusted R-squared:  0.1398
## F-statistic: 5.387 on 1 and 26 DF,  p-value: 0.0284
```

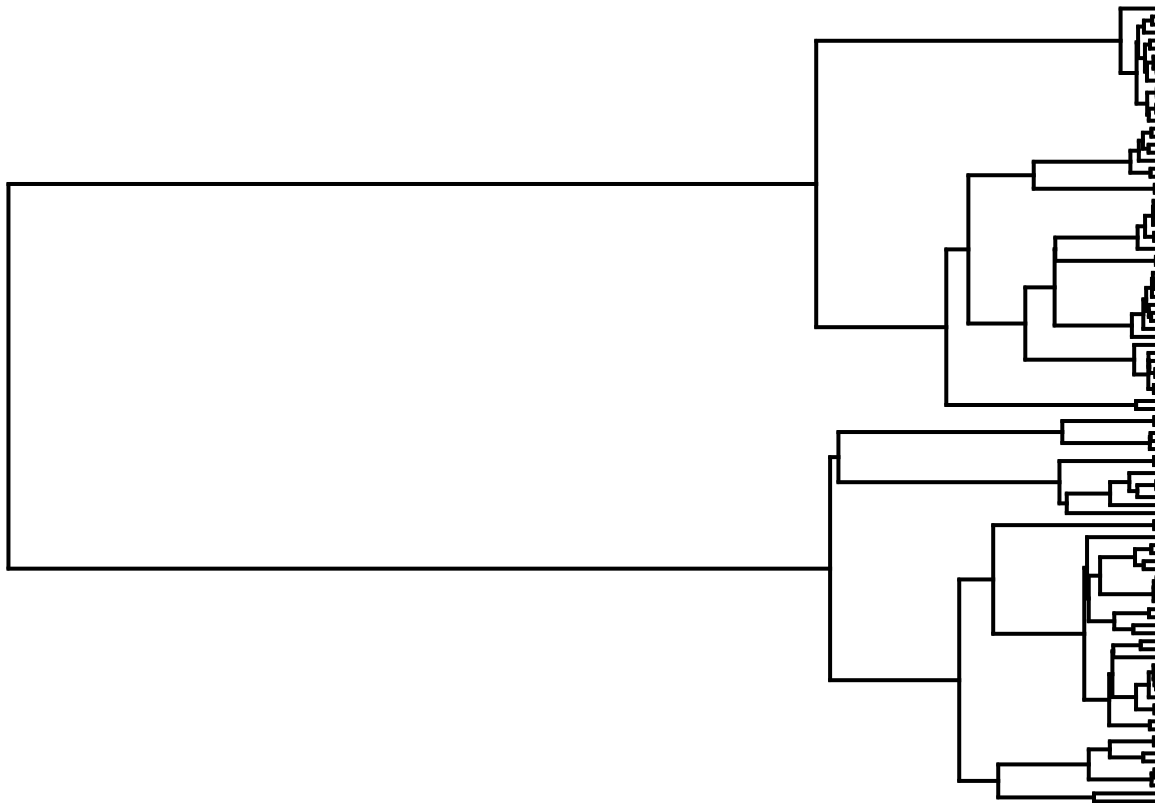Plot linear model over scatterplot with phylogeny considered

```r
plot(pic.bl,pic.gw,xlab="PICs for buccal length",ylab="PICs for gape width")
abline(fit.pic,lwd=2,lty="dashed",col="red")
```
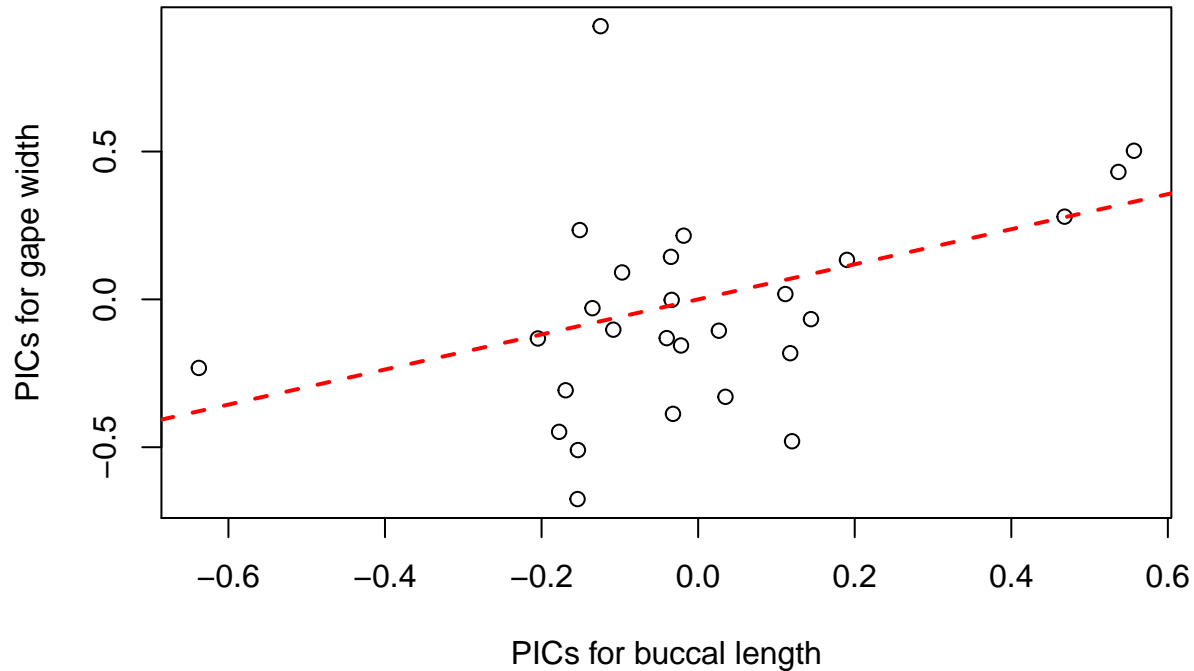

Simulating a Tree to Test PIC

Simulate a coalescent tree:

```r
set.seed(21)
## simulate a coalescent shaped tree
tree<-rcoal(n=100)
plotTree(tree,ftype="off")
```

Simulate uncorrelated Brownian evolution:

```
x<-fastBM(tree) #simulate trait 1
y<-fastBM(tree) #simulate trait 2
par(mar=c(5.1,4.1,2.1,2.1)) #set plot dimensions
plot(x,y)
fit<-lm(y~x) #fit linear regression to data
abline(fit,lwd=2,lty="dashed",col="red") #add lm line
```
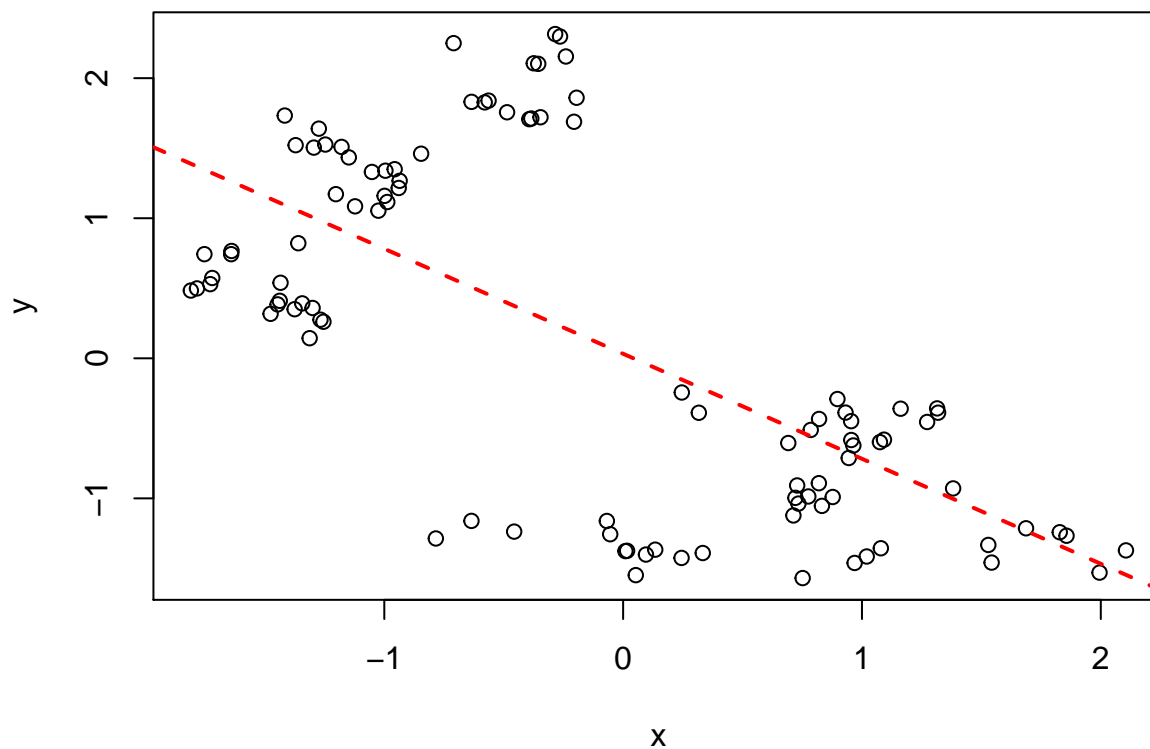
Return descriptions of fit:

```
fit
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##      0.0322      -0.7496
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.90636 -0.68689  0.03199  0.53566  2.06929
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.03220    0.09378   0.343    0.732
## x           -0.74963    0.08651  -8.665 9.46e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9294 on 98 degrees of freedom
## Multiple R-squared:  0.4338, Adjusted R-squared:  0.428
## F-statistic: 75.09 on 1 and 98 DF,  p-value: 9.458e-14
```

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x          1 64.857  64.857   75.09 9.458e-14 ***
## Residuals 98 84.644   0.864
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observe that phylogeny can very easily cause a type I error. The close taxonomical relationship of the data points within the two clusters causes the traits to appear correlated even when they are not.

Visualize the relatedness of the data points in the scatterplot:

```
phylomorphospace(tree,cbind(x,y),label="off",node.size=c(0.5,0.7))
abline(fit,lwd=2,lty="dashed",col="red")
```
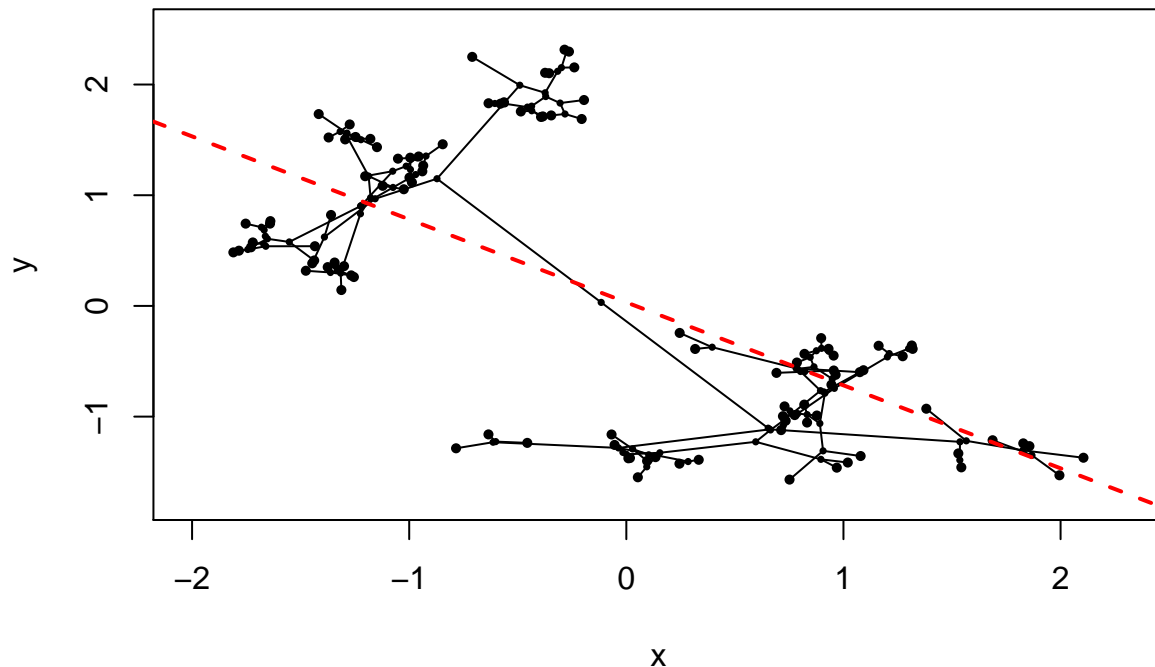

Use Felsenstein's algorithim to weight the contrasts between species:

```
ix<-pic(x,tree)
iy<-pic(y,tree)
fit<-lm(iy~ix-1) ## we have to fit the model without an intercept term
fit
```

```
##
## Call:
## lm(formula = iy ~ ix - 1)
##
## Coefficients:
##       ix
## -0.03771
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = iy ~ ix - 1)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -2.9265 -0.8937 -0.2124  0.4106  2.7002
##
## Coefficients:
##     Estimate Std. Error t value Pr(>|t|)
## ix -0.03771    0.09814  -0.384    0.702
##
## Residual standard error: 1.043 on 98 degrees of freedom
## Multiple R-squared:  0.001504,   Adjusted R-squared:  -0.008685
## F-statistic: 0.1476 on 1 and 98 DF,  p-value: 0.7017
```

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: iy
##           Df Sum Sq Mean Sq F value Pr(>F)
## ix         1   0.16 0.16046  0.1476 0.7017
## Residuals 98 106.53 1.08703
```