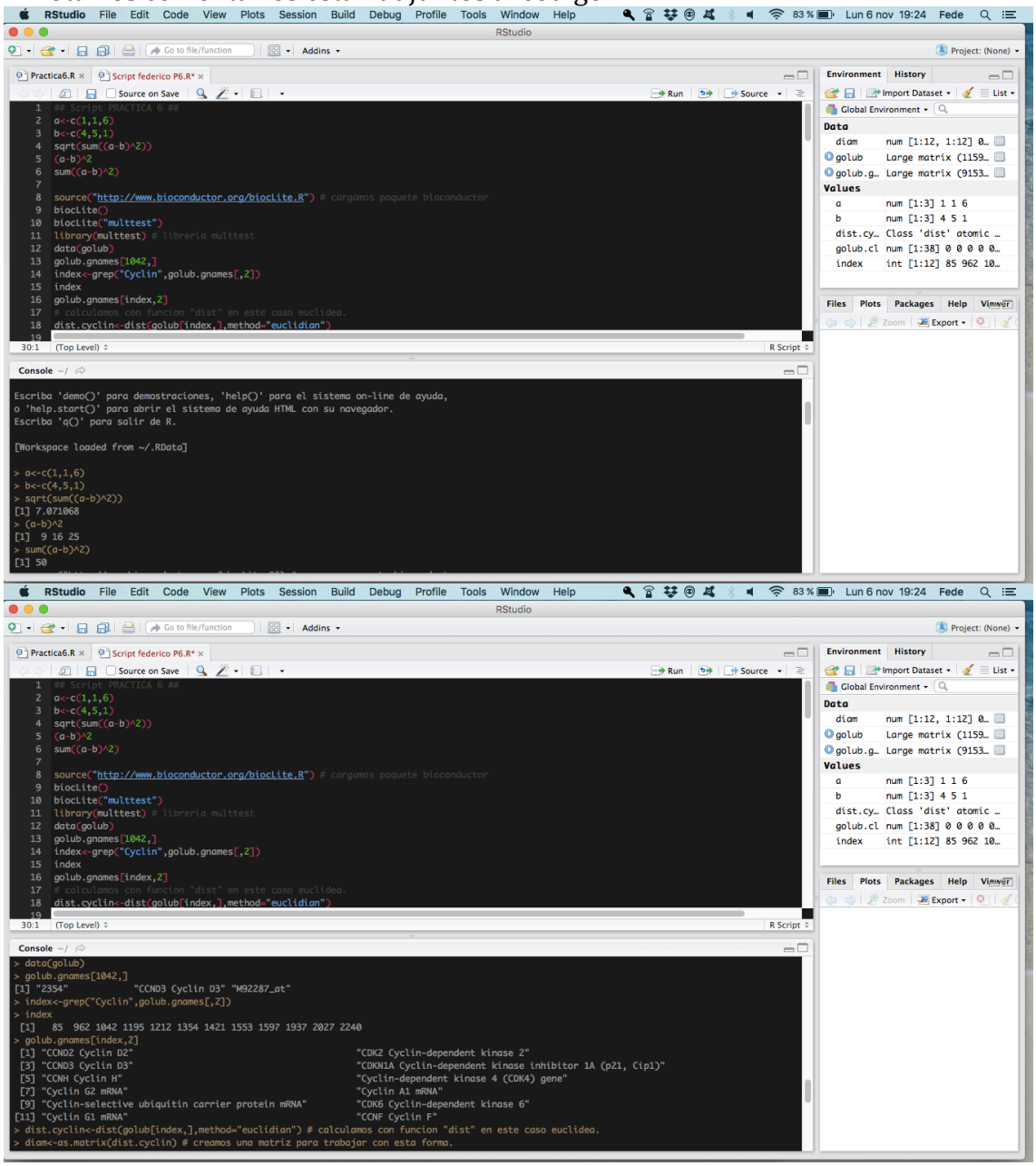
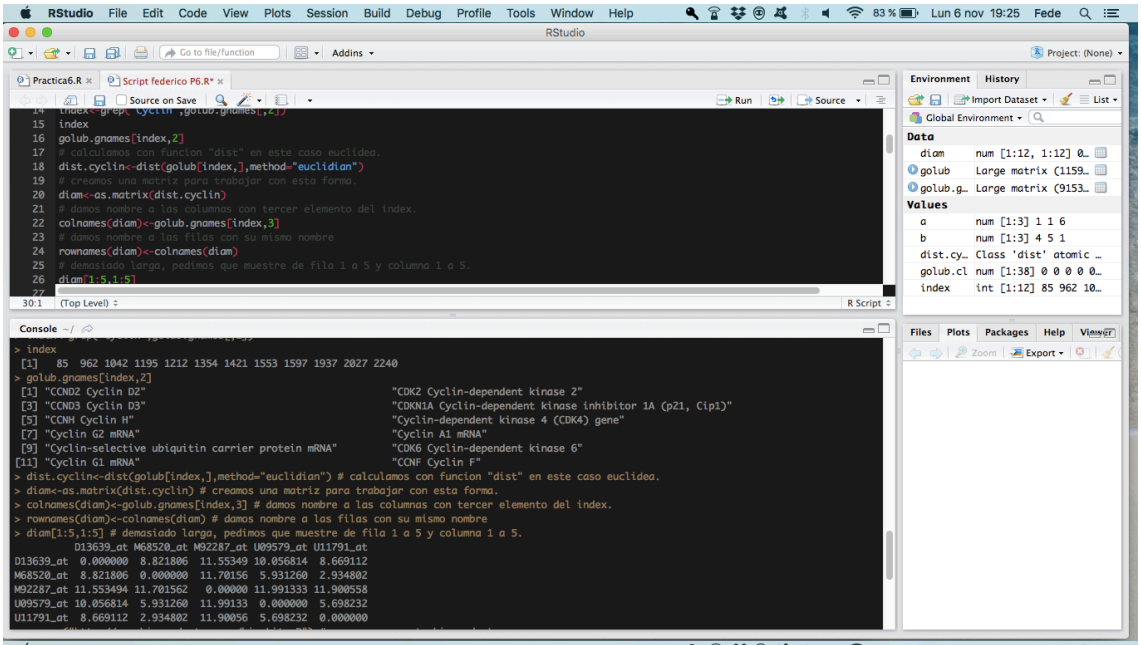
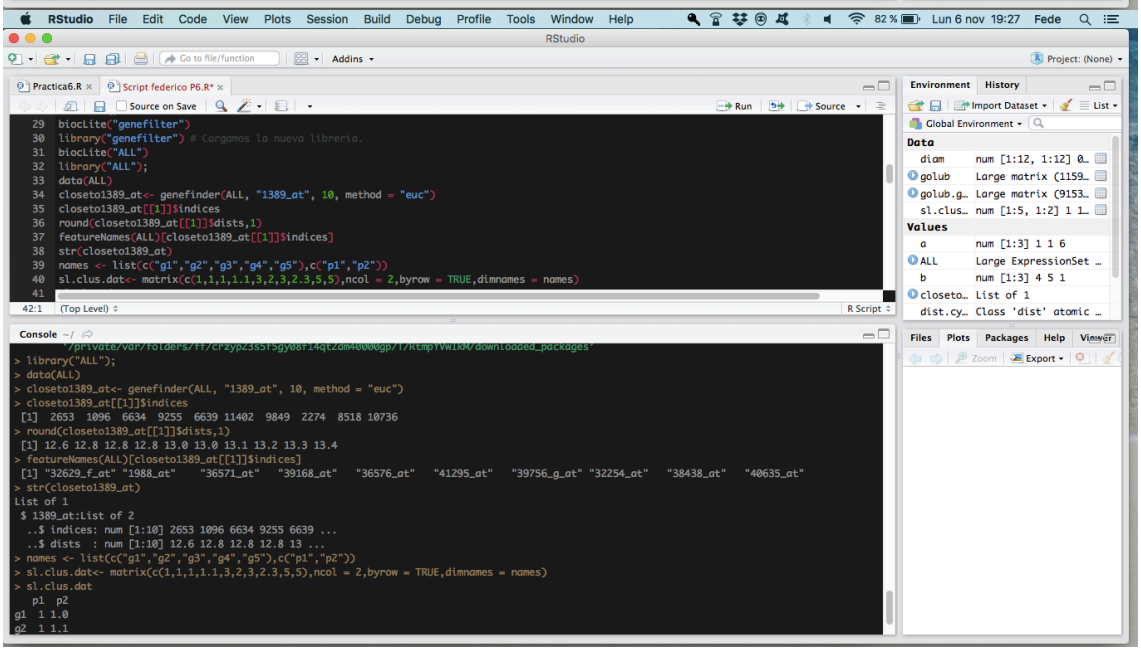
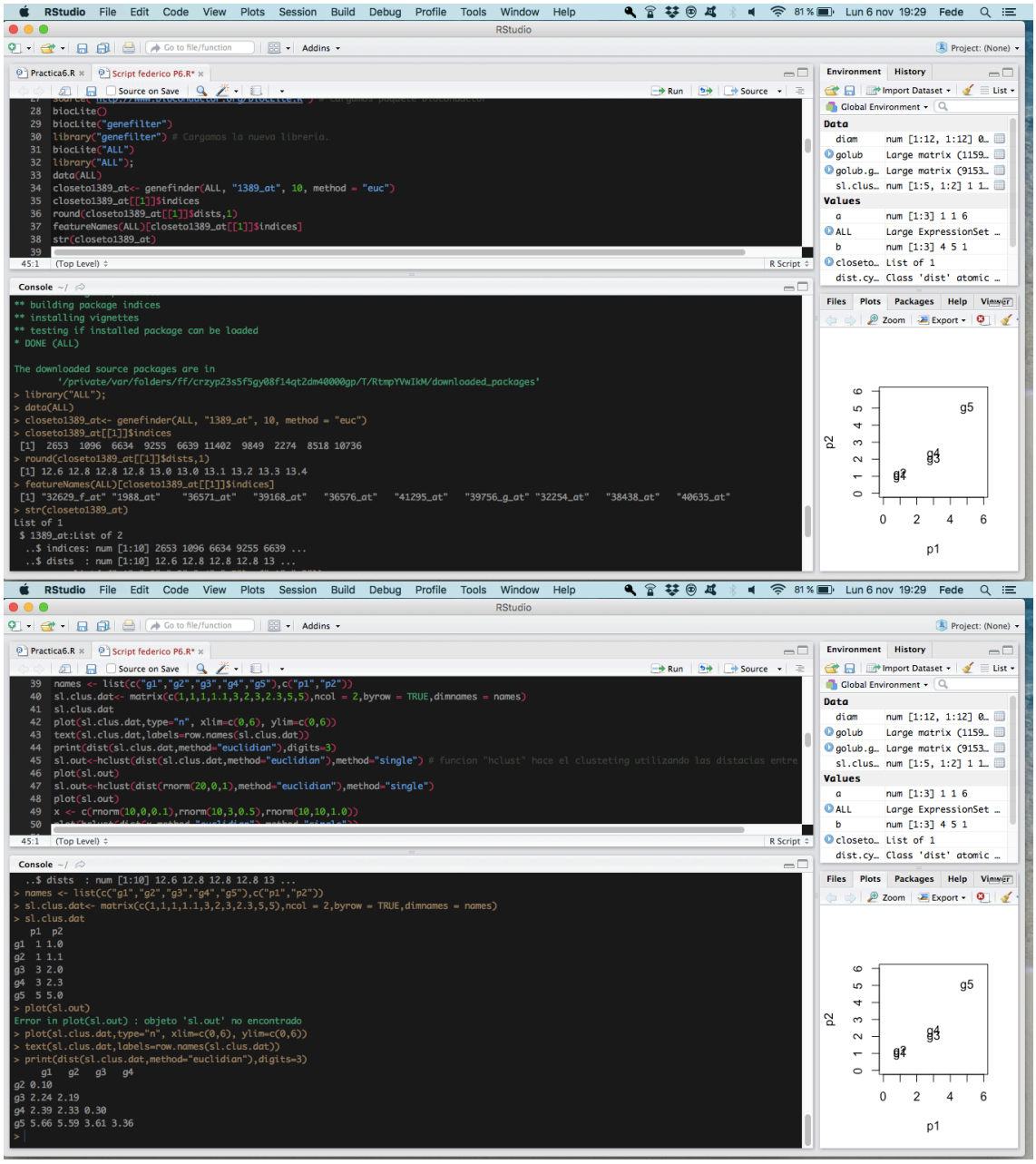
PRACTICA 6 PROGRAMACION EN R:

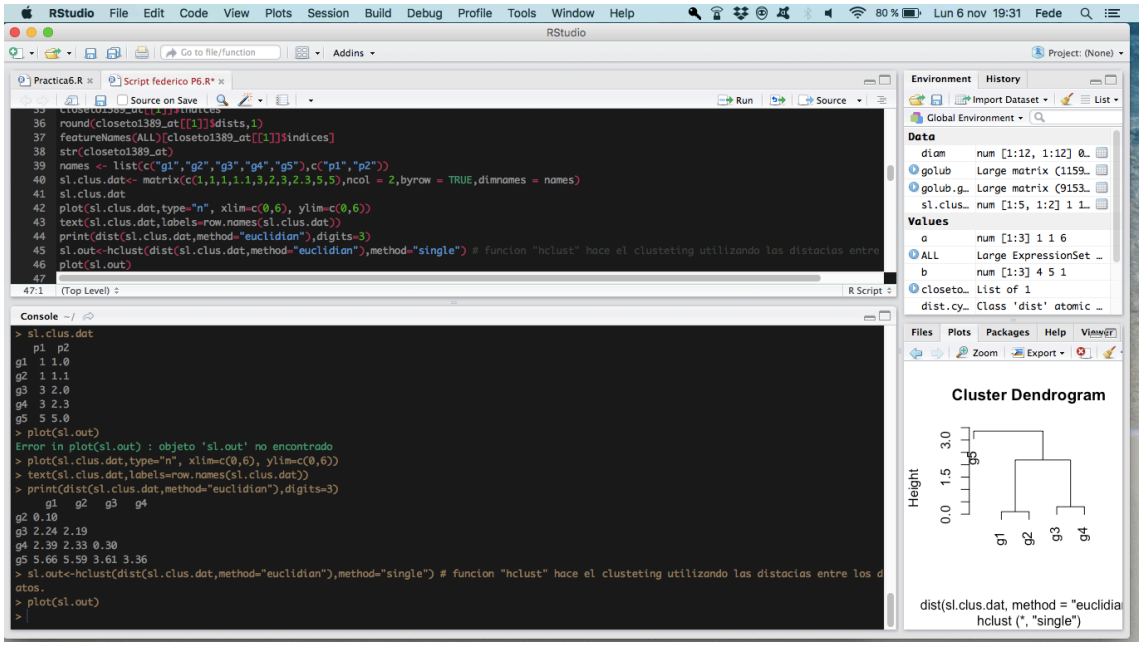
**Nota: los comentarios están adjuntos al código.



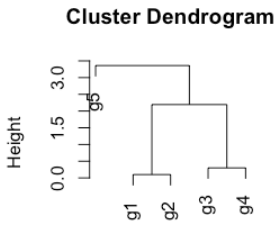

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Lun 6 nov 19:25 Fede
Practica6.R
Script federico P6.R*
Source on Save
Run
Source
index
golub.gnames[index,2]
# calculamos con funcion "dist" en este caso euclidea.
dist.cyclin<-dist(golub[index,],method="euclidian")
# creamos una matriz para trabajar con esta forma.
diam<-as.matrix(dist.cyclin)
# damos nombre a las columnas con tercer elemento del index.
colnames(diam)<-golub.gnames[index,3]
# damos nombre a las filas con su mismo nombre
rownames(diam)<-colnames(diam)
# demasiado larga, pedimos que muestre de fila 1 a 5 y columna 1 a 5.
diam[1:5,1:5]
30:1
(Top Level)
R Script
Console
> index
[1] 85 962 1042 1195 1212 1354 1421 1553 1597 1937 2027 2240
> golub.gnames[index,2]
[1] "CCND2 Cyclin D2" "CDK2 Cyclin-dependent kinase 2"
[3] "CCND3 Cyclin D3" "CDKN1A Cyclin-dependent kinase inhibitor 1A (p21, Cip1)"
[5] "CCNH Cyclin H" "Cyclin-dependent kinase 4 (CDK4) gene"
[7] "Cyclin G2 mRNA" "Cyclin A1 mRNA"
[9] "Cyclin-selective ubiquitin carrier protein mRNA" "CDK6 Cyclin-dependent kinase 6"
[11] "Cyclin G1 mRNA" "CCNF Cyclin F"
> dist.cyclin<-dist(golub[index,],method="euclidian") # calculamos con funcion "dist" en este caso euclidea.
> diam<-as.matrix(dist.cyclin) # creamos una matriz para trabajar con esta forma.
> colnames(diam)<-golub.gnames[index,3] # damos nombre a las columnas con tercer elemento del index.
> rownames(diam)<-colnames(diam) # damos nombre a las filas con su mismo nombre
> diam[1:5,1:5] # demasiado larga, pedimos que muestre de fila 1 a 5 y columna 1 a 5.
D13639_at M68520_at M92287_at U09579_at U11791_at
D13639_at 0.000000 8.821806 11.55349 10.056814 8.669112
M68520_at 8.821806 0.000000 11.70156 5.931260 2.934802
M92287_at 11.553494 11.701562 0.00000 11.991333 11.900558
U09579_at 10.056814 5.931260 11.99133 0.000000 5.698232
U11791_at 8.669112 2.934802 11.90056 5.698232 0.000000
Environment
History
Import Dataset
List
Global Environment
Data
diam num [1:12, 1:12] 0…
golub Large matrix (1159…
golub.g… Large matrix (9153…
Values
a num [1:3] 1 1 6
b num [1:3] 4 5 1
dist.cy… Class 'dist' atomic …
golub.cl num [1:38] 0 0 0 0 0…
index int [1:12] 85 962 10…
Files
Plots
Packages
Help
Viewer
Zoom
Export


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Lun 6 nov 19:27 Fede
Practica6.R
Script federico P6.R*
Source on Save
Run
Source
bioclite("genefilter")
library("genefilter") # Cargamos la nueva libreria.
bioclite("ALL")
library("ALL");
data(ALL)
closeto1389_at<- genefinder(ALL, "1389_at", 10, method = "euc")
closeto1389_at[[1]]$indices
round(closeto1389_at[[1]]$dists,1)
featureNames(ALL)[closeto1389_at[[1]]$indices]
str(closeto1389_at)
names <- list(c("g1","g2","g3","g4","g5"),c("p1","p2"))
sl.clus.dat<- matrix(c(1,1,1,1.1,3,2,3,2.3,5,5),ncol = 2,byrow = TRUE,dimnames = names)
42:1
(Top Level)
R Script
Console
> library("ALL");
> data(ALL)
> closeto1389_at<- genefinder(ALL, "1389_at", 10, method = "euc")
> closeto1389_at[[1]]$indices
[1] 2653 1096 6634 9255 6639 11402 9849 2274 8518 10736
> round(closeto1389_at[[1]]$dists,1)
[1] 12.6 12.8 12.8 12.8 13.0 13.0 13.1 13.2 13.3 13.4
> featureNames(ALL)[closeto1389_at[[1]]$indices]
[1] "32629_f_at" "1988_at" "36571_at" "39168_at" "36576_at" "41295_at" "39756_g_at" "32254_at" "38438_at" "40635_at"
> str(closeto1389_at)
List of 1
$ 1389_at:List of 2
..$ indices: num [1:10] 2653 1096 6634 9255 6639 ...
..$ dists : num [1:10] 12.6 12.8 12.8 12.8 13 ...
> names <- list(c("g1","g2","g3","g4","g5"),c("p1","p2"))
> sl.clus.dat<- matrix(c(1,1,1,1.1,3,2,3,2.3,5,5),ncol = 2,byrow = TRUE,dimnames = names)
> sl.clus.dat
p1 p2
g1 1 1.0
g2 1 1.1
Environment
History
Import Dataset
List
Global Environment
Data
diam num [1:12, 1:12] 0…
golub Large matrix (1159…
golub.g… Large matrix (9153…
sl.clus… num [1:5, 1:2] 1 1…
Values
a num [1:3] 1 1 6
ALL Large ExpressionSet …
b num [1:3] 4 5 1
closeto… List of 1
dist.cy… Class 'dist' atomic …
Files
Plots
Packages
Help
Viewer
Zoom
Export

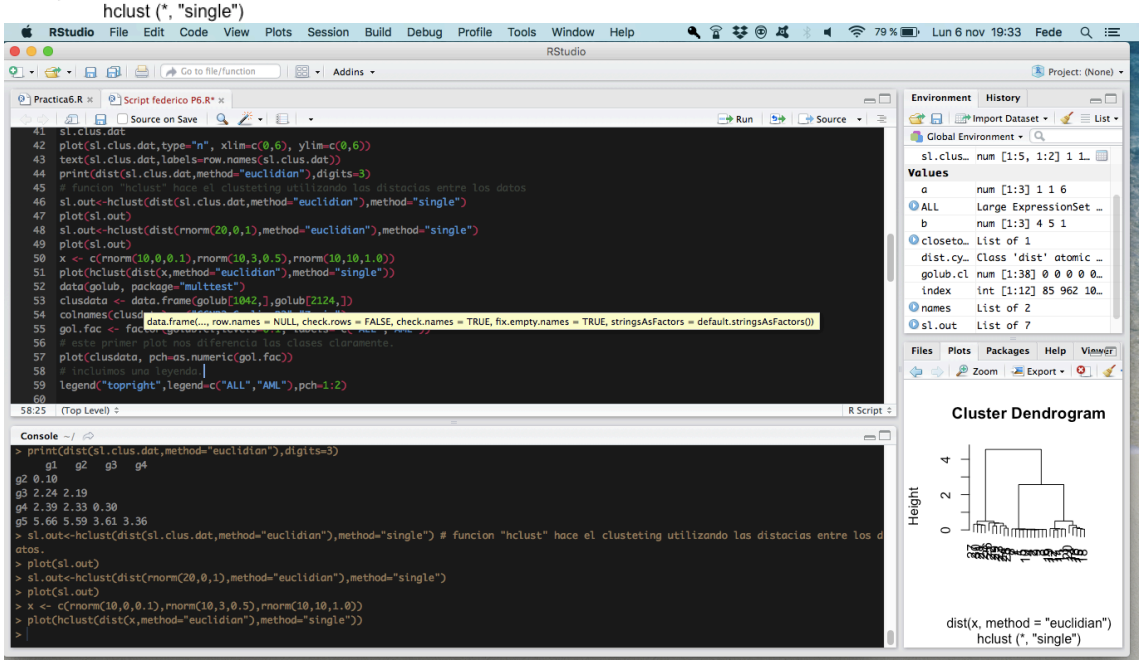

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Lun 6 nov 19:29 Fede
RStudio
Practica6.R
Script federico P6.R*
Source on Save
Run
Source
bioclite()
bioclite("genefilter")
library("genefilter") # Cargamos la nueva libreria.
bioclite("ALL")
library("ALL");
data(ALL)
closeto1389_at<- genefinder(ALL, "1389_at", 10, method = "euc")
closeto1389_at[[1]]$indices
round(closeto1389_at[[1]]$dists,1)
featureNames(ALL)[closeto1389_at[[1]]$indices]
str(closeto1389_at)
45:1 (Top Level) R Script
Console ~/
** building package indices
** installing vignettes
** testing if installed package can be loaded
* DONE (ALL)
The downloaded source packages are in
'/private/var/folders/ff/crzyp23s5f5gy08f14qt2dm40000gp/T/RtmpYYwIkM/downloaded_packages'
> library("ALL");
> data(ALL)
> closeto1389_at<- genefinder(ALL, "1389_at", 10, method = "euc")
> closeto1389_at[[1]]$indices
[1] 2653 1096 6634 9255 6639 11402 9849 2274 8518 10736
> round(closeto1389_at[[1]]$dists,1)
[1] 12.6 12.8 12.8 12.8 13.0 13.0 13.1 13.2 13.3 13.4
> featureNames(ALL)[closeto1389_at[[1]]$indices]
[1] "32629_f_at" "1988_at" "36571_at" "39168_at" "36576_at" "41295_at" "39756_g_at" "32254_at" "38438_at" "40635_at"
> str(closeto1389_at)
List of 1
$ 1389_at:List of 2
..$ indices: num [1:10] 2653 1096 6634 9255 6639 ...
..$ dists : num [1:10] 12.6 12.8 12.8 12.8 13 ...
Environment History
Import Dataset
List
Global Environment
Data
diam num [1:12, 1:12] 0…
golub Large matrix (1159…
golub.g… Large matrix (9153…
sl.clus… num [1:5, 1:2] 1 1…
Values
a num [1:3] 1 1 6
ALL Large ExpressionSet …
b num [1:3] 4 5 1
closeto… List of 1
dist.cy… Class 'dist' atomic …
Files Plots Packages Help Viewer
Zoom
Export
p2
p1
g5
g4
g3
g2
g1
0 2 4 6
0 1 2 3 4 5 6
RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Lun 6 nov 19:29 Fede
RStudio
Practica6.R
Script federico P6.R*
Source on Save
Run
Source
names <- list(c("g1","g2","g3","g4","g5"),c("p1","p2"))
sl.clus.dat<- matrix(c(1,1,1,1.1,3,2,3,2.3,5,5),ncol = 2,byrow = TRUE,dimnames = names)
sl.clus.dat
plot(sl.clus.dat,type="n", xlim=c(0,6), ylim=c(0,6))
text(sl.clus.dat,labels=row.names(sl.clus.dat))
print(dist(sl.clus.dat,method="euclidian"),digits=3)
sl.out<-hclust(dist(sl.clus.dat,method="euclidian"),method="single") # funcion "hclust" hace el clusteting utilizando las distacias entre
plot(sl.out)
sl.out<-hclust(dist(rnorm(20,0,1),method="euclidian"),method="single")
plot(sl.out)
x <- c(rnorm(10,0,0.1),rnorm(10,3,0.5),rnorm(10,10,1.0))
45:1 (Top Level) R Script
Console ~/
..$ dists : num [1:10] 12.6 12.8 12.8 12.8 13 ...
> names <- list(c("g1","g2","g3","g4","g5"),c("p1","p2"))
> sl.clus.dat<- matrix(c(1,1,1,1.1,3,2,3,2.3,5,5),ncol = 2,byrow = TRUE,dimnames = names)
> sl.clus.dat
p1 p2
g1 1 1.0
g2 1 1.1
g3 3 2.0
g4 3 2.3
g5 5 5.0
> plot(sl.out)
Error in plot(sl.out) : objeto 'sl.out' no encontrado
> plot(sl.clus.dat,type="n", xlim=c(0,6), ylim=c(0,6))
> text(sl.clus.dat,labels=row.names(sl.clus.dat))
> print(dist(sl.clus.dat,method="euclidian"),digits=3)
g1 g2 g3 g4
g2 0.10
g3 2.24 2.19
g4 2.39 2.33 0.30
g5 5.66 5.59 3.61 3.36
>
Environment History
Import Dataset
List
Global Environment
Data
diam num [1:12, 1:12] 0…
golub Large matrix (1159…
golub.g… Large matrix (9153…
sl.clus… num [1:5, 1:2] 1 1…
Values
a num [1:3] 1 1 6
ALL Large ExpressionSet …
b num [1:3] 4 5 1
closeto… List of 1
dist.cy… Class 'dist' atomic …
Files Plots Packages Help Viewer
Zoom
Export
p2
p1
g5
g4
g3
g2
g1
0 2 4 6
0 1 2 3 4 5 6

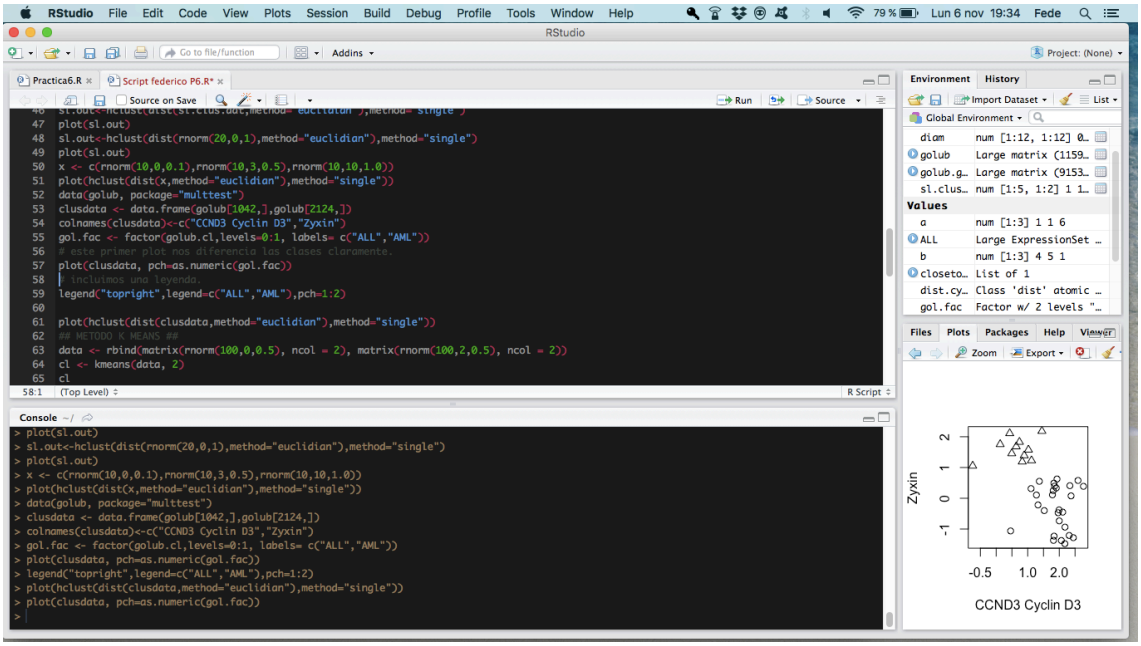


Obtenemos la siguiente grafica:

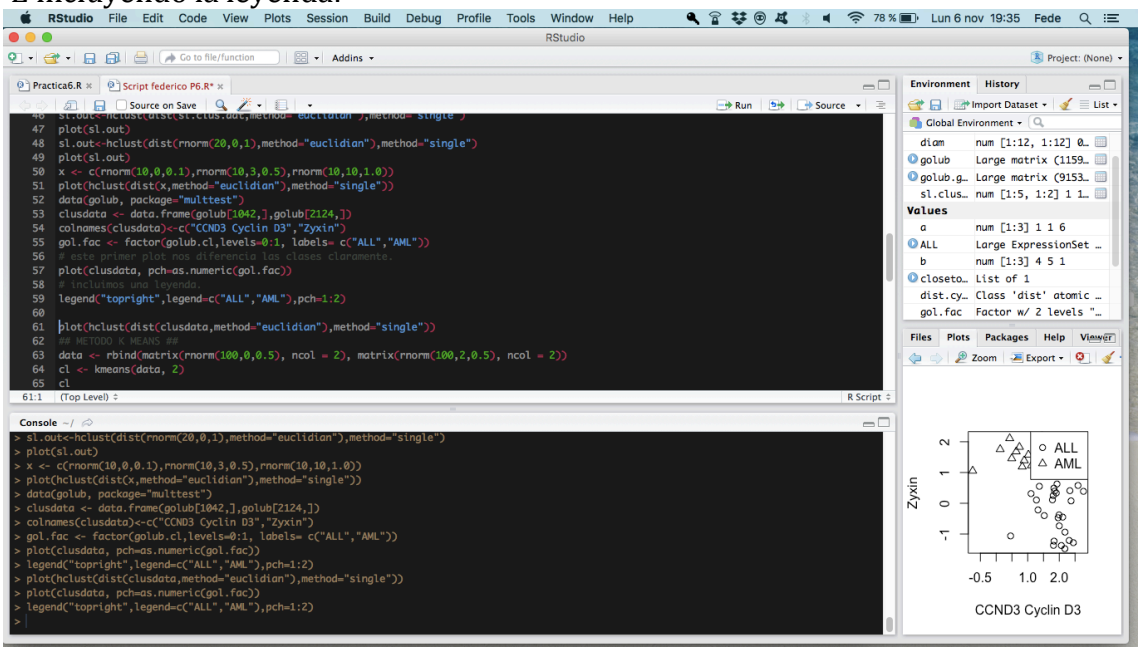


dist(sl.clus.dat, method = "euclidia
hclust (*, "single")

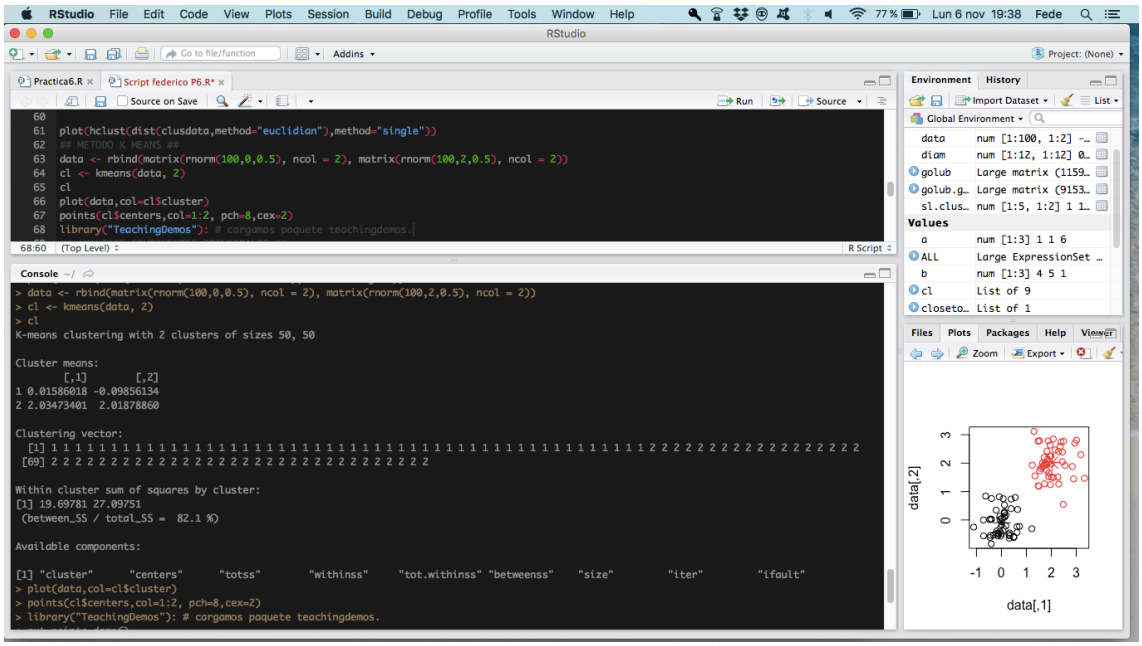




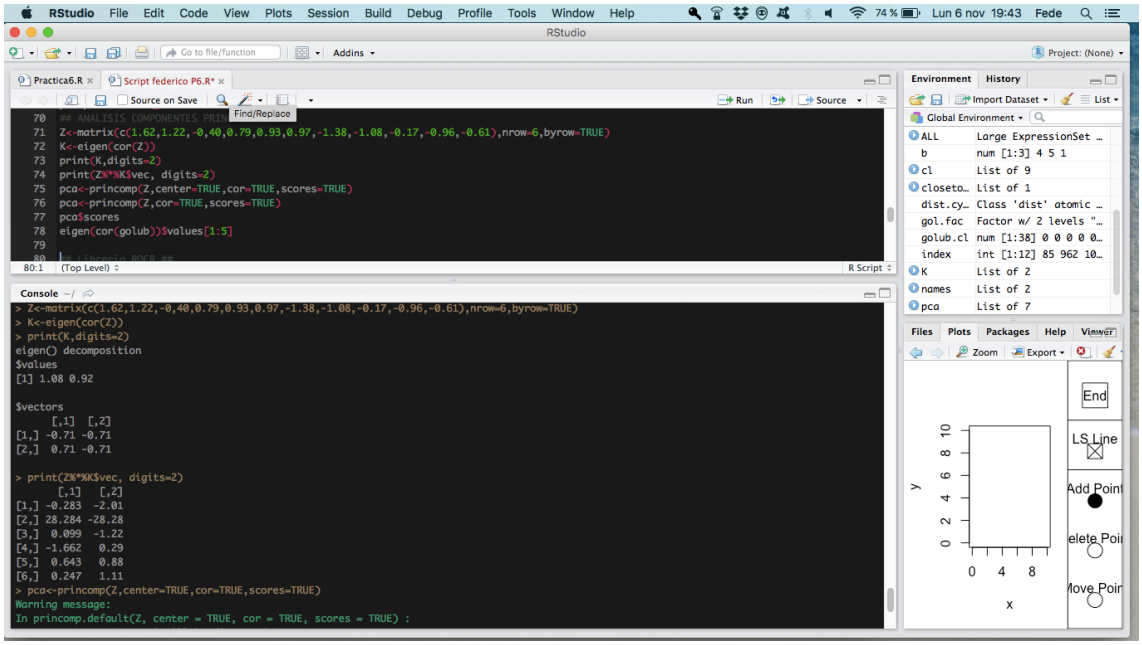
E incluyendo la leyenda:

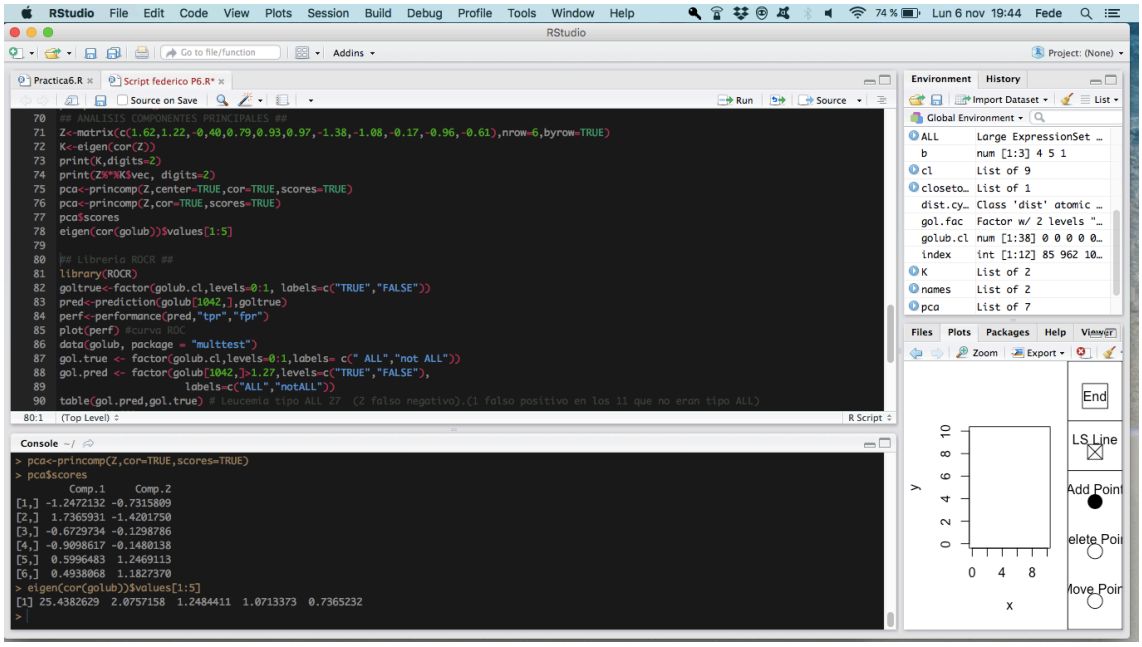


Metodo K means;

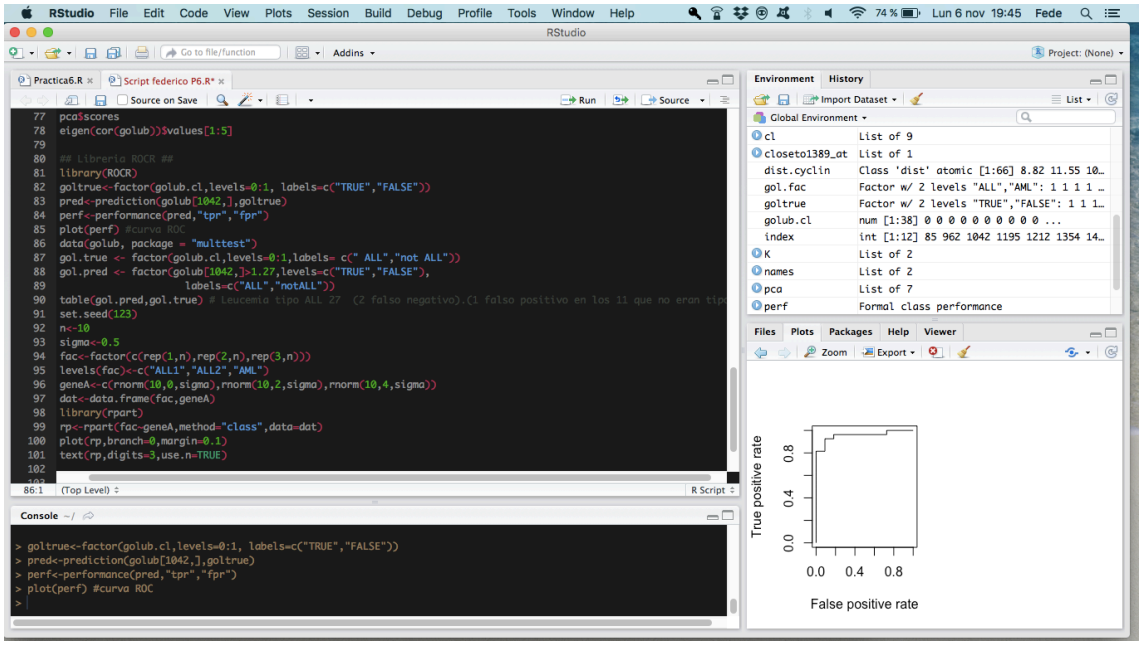


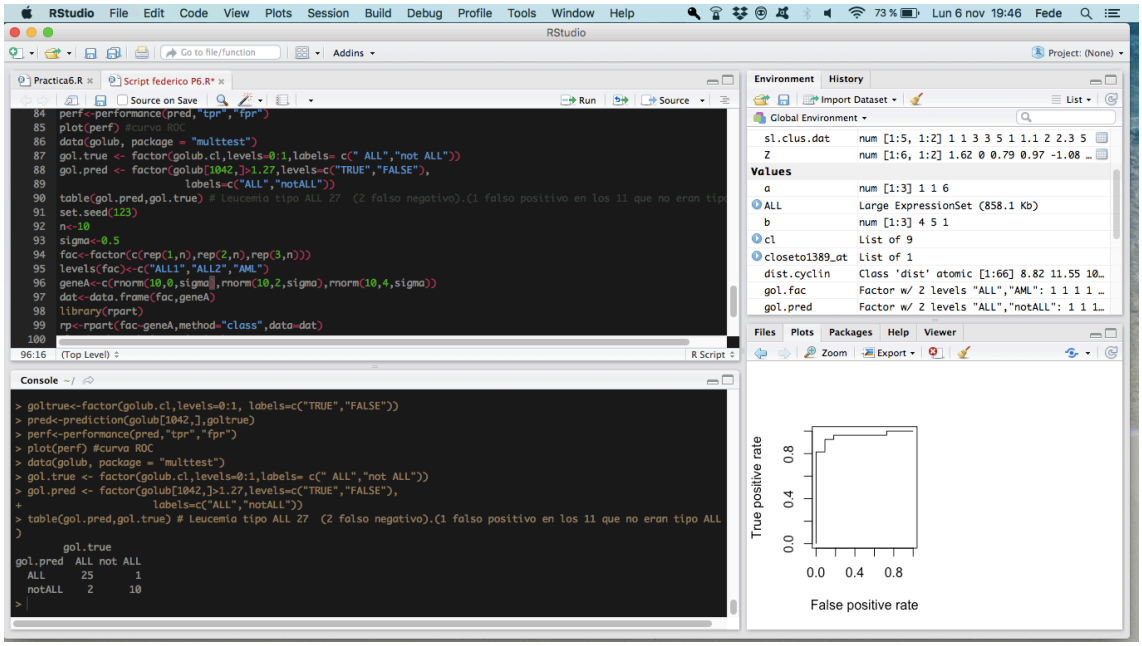
Analisis de componentes principales:




RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Lun 6 nov 19:44 Fede
Practica6.R
Script federico P6.R*
Source on Save
Run
Source
ANALISIS COMPONENTES PRINCIPALES
Z<-matrix(c(1.62,1.22,-0,40,0.79,0.93,0.97,-1.38,-1.08,-0.17,-0.96,-0.61),nrow=6,byrow=TRUE)
K<-eigen(cor(Z))
print(K,digits=2)
print(Z%*%K$vec, digits=2)
pca<-princomp(Z,center=TRUE,cor=TRUE,scores=TRUE)
pca<-princomp(Z,cor=TRUE,scores=TRUE)
pca$scores
eigen(cor(golub))$values[1:5]
Libreria ROCR
library(ROCR)
goltrue<-factor(golub.cl,levels=0:1, labels=c("TRUE","FALSE"))
pred<-prediction(golub[1042,],goltrue)
perf<-performance(pred,"tpr","fpr")
plot(perf) #curva ROC
data(golub, package = "multtest")
gol.true <- factor(golub.cl,levels=0:1,labels= c(" ALL","not ALL"))
gol.pred <- factor(golub[1042,]>1.27,levels=c("TRUE","FALSE"),
labels=c("ALL","notALL"))
table(gol.pred,gol.true) # Leucemia tipo ALL 27 (2 falso negativo).(1 falso positivo en los 11 que no eran tipo ALL)
80:1 (Top Level)
R Script
Console ~/
> pca<-princomp(Z,cor=TRUE,scores=TRUE)
> pca$scores
Comp.1 Comp.2
[1,] -1.2472132 -0.7315809
[2,] 1.7365931 -1.4201750
[3,] -0.6729734 -0.1298786
[4,] -0.9098617 -0.1480138
[5,] 0.5996483 1.2469113
[6,] 0.4938068 1.1827370
> eigen(cor(golub))$values[1:5]
[1] 25.4382629 2.0757158 1.2484411 1.0713373 0.7365232
Environment History
Import Dataset
List
Global Environment
ALL Large ExpressionSet …
b num [1:3] 4 5 1
cl List of 9
closeto… List of 1
dist.cy… Class 'dist' atomic …
gol.fac Factor w/ 2 levels "…
golub.cl num [1:38] 0 0 0 0 0…
index int [1:12] 85 962 10…
K List of 2
names List of 2
pca List of 7
Files Plots Packages Help Viewer
Zoom
Export
End
LS Line
Add Point
elete Poi
love Poir
y
x

Curvas Roc:


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Lun 6 nov 19:45 Fede
Practica6.R
Script federico P6.R*
pca$scores
eigen(cor(golub))$values[1:5]
Libreria ROCR
library(ROCR)
goltrue<-factor(golub.cl,levels=0:1, labels=c("TRUE","FALSE"))
pred<-prediction(golub[1042,],goltrue)
perf<-performance(pred,"tpr","fpr")
plot(perf) #curva ROC
data(golub, package = "multtest")
gol.true <- factor(golub.cl,levels=0:1,labels= c(" ALL","not ALL"))
gol.pred <- factor(golub[1042,]>1.27,levels=c("TRUE","FALSE"),
labels=c("ALL","notALL"))
table(gol.pred,gol.true) # Leucemia tipo ALL 27 (2 falso negativo).(1 falso positivo en los 11 que no eran tipo
set.seed(123)
n<-10
sigma<-0.5
fac<-factor(c(rep(1,n),rep(2,n),rep(3,n)))
levels(fac)<-c("ALL1","ALL2","AML")
geneA<-c(rnorm(10,0,sigma),rnorm(10,2,sigma),rnorm(10,4,sigma))
dat<-data.frame(fac,geneA)
library(rpart)
rp<-rpart(fac~geneA,method="class",data=dat)
plot(rp,branch=0,margin=0.1)
text(rp,digits=3,use.n=TRUE)
86:1 (Top Level)
R Script
Console ~/
> goltrue<-factor(golub.cl,levels=0:1, labels=c("TRUE","FALSE"))
> pred<-prediction(golub[1042,],goltrue)
> perf<-performance(pred,"tpr","fpr")
> plot(perf) #curva ROC
Environment History
Import Dataset
List
Global Environment
cl List of 9
closeto1389_at List of 1
dist.cyclin Class 'dist' atomic [1:66] 8.82 11.55 10…
gol.fac Factor w/ 2 levels "ALL","AML": 1 1 1 1 …
goltrue Factor w/ 2 levels "TRUE","FALSE": 1 1 1…
golub.cl num [1:38] 0 0 0 0 0 0 0 0 0 0 ...
index int [1:12] 85 962 1042 1195 1212 1354 14…
K List of 2
names List of 2
pca List of 7
perf Formal class performance
Files Plots Packages Help Viewer
Zoom
Export
True positive rate
0.0
0.4
0.8
0.0 0.4 0.8
False positive rate



Leucemia tipo ALL 27 (2 falso negativo).(1 falso positivo en los 11 que no eran tipo ALL)