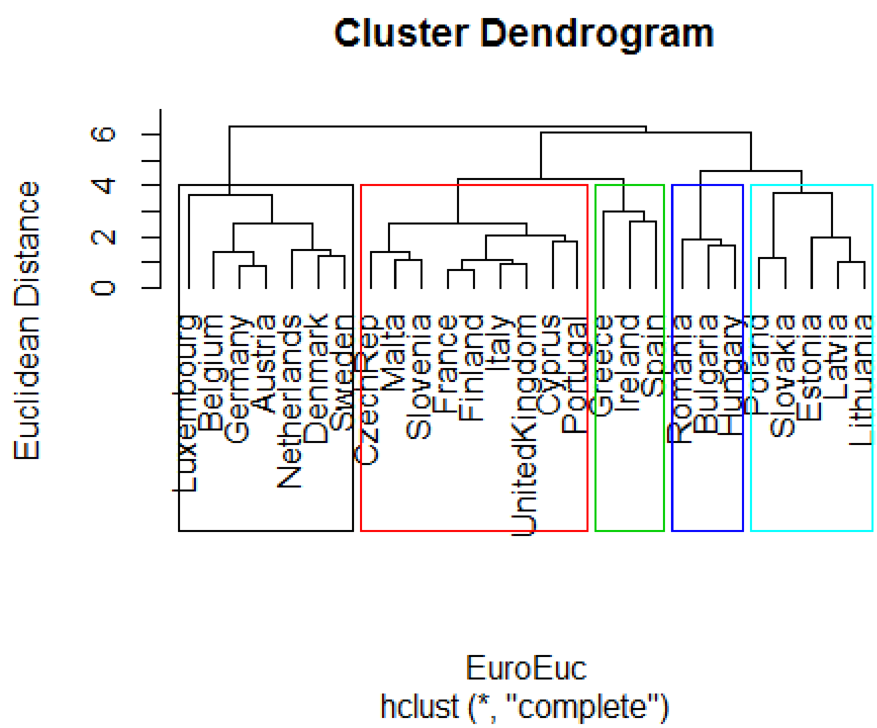
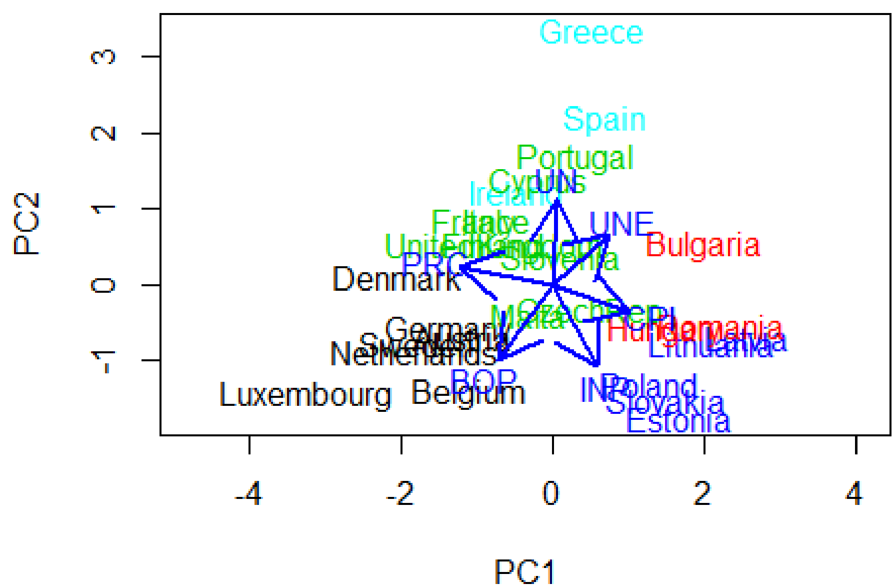
```
### Exercise 1

remove(list = ls())

euro  = read.table(file = "Europe.txt", header = T, dec =".")
attach(euro)
europe = data.frame(euro)
europe = na.omit(europe)
Country = ï..Country
numbers = cbind(CPI, UNE, INP, BOP, PRC, UN)

# a)
Euro = scale(numbers, center = TRUE, scale = TRUE)
EuroEuc = dist(Euro, method = "euclidean", diag = TRUE, upper = TRUE)
reshclust = hclust(EuroEuc, method = "complete")
plot(reshclust, hang = -1, labels = Country, ylab = "Euclidean Distance")
rect.hclust(reshclust, k = 5, border = c(1,2,3,4,5))
```



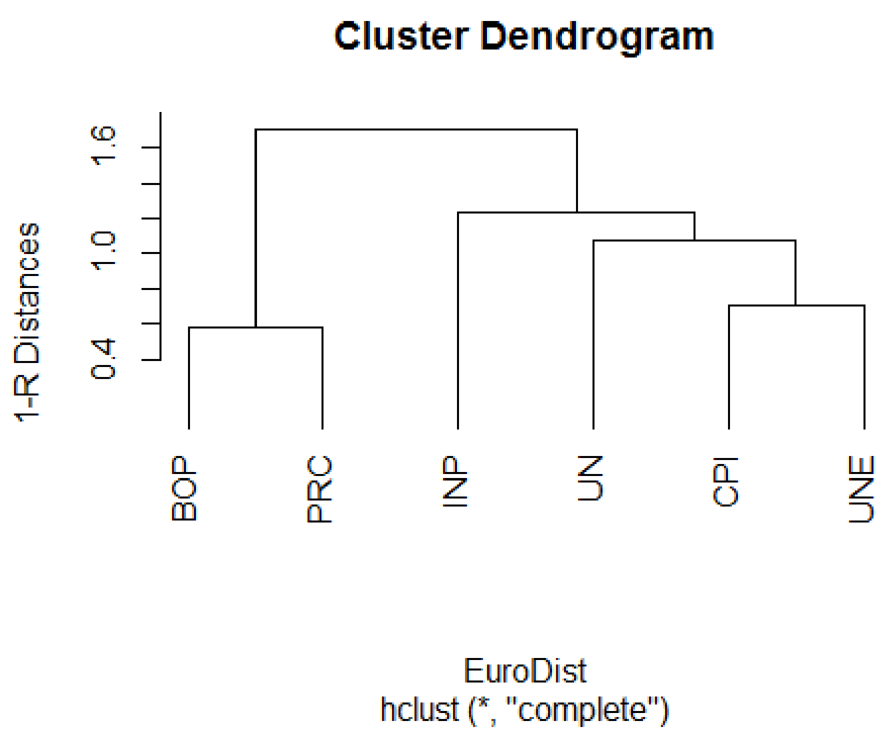

```
# b)
S1 = cov(Euro)
lam1 = eigen(S1)$values[1:2]
e1 = eigen(S1)$vectors[,1:2]
Yhat = Euro%*%e1
plot(-Yhat[,1],Yhat[,2], type = "n", asp = 1, ylab = "PC2", xlab = "PC1")
text(-Yhat[,1],Yhat[,2], Country,  col = cutree(reshclust, k = 5))
```

```
arrows(0,0,-2*e1[1,1],2*e1[1,2], col = "blue", lwd = 2)
text(-2.5*e1[1,1],2.5*e1[1,2], labels = variable.names(europe[2]), col = "blu
e")
arrows(0,0,-2*e1[2,1],2*e1[2,2], col = "blue", lwd = 2)
text(-2.5*e1[2,1],2.5*e1[2,2], labels = variable.names(europe[3]), col = "blu
e")
arrows(0,0,-2*e1[3,1],2*e1[3,2], col = "blue", lwd = 2)
text(-2.5*e1[3,1],2.5*e1[3,2], labels = variable.names(europe[4]), col = "blu
e")
arrows(0,0,-2*e1[4,1],2*e1[4,2], col = "blue", lwd = 2)
text(-2.5*e1[4,1],2.5*e1[4,2], labels = variable.names(europe[5]), col = "blu
e")
arrows(0,0,-2*e1[5,1],2*e1[5,2], col = "blue", lwd = 2)
text(-2.5*e1[5,1],2.5*e1[5,2], labels = variable.names(europe[6]), col = "blu
e")
arrows(0,0,-2*e1[6,1],2*e1[6,2], col = "blue", lwd = 2)
text(-2.5*e1[6,1],2.5*e1[6,2], labels = variable.names(europe[7]), col = "blu
e")
```



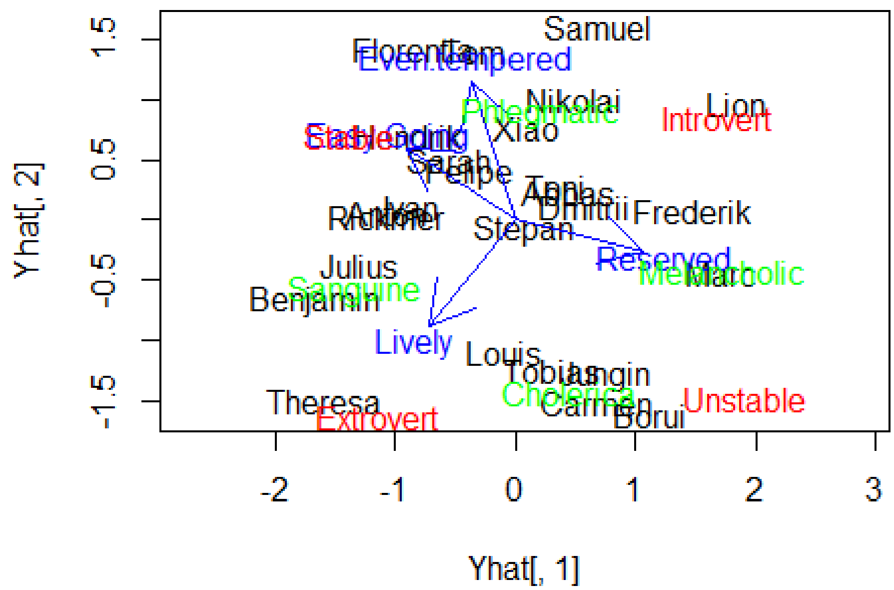

```
# c)

# d)
R1 = cor(Euro)
round(R1, digits = 2)
```

```
##        CPI   UNE   INP   BOP   PRC    UN
## CPI  1.00  0.29  0.21 -0.11 -0.71 -0.08
## UNE  0.29  1.00 -0.10 -0.40 -0.33  0.03
## INP  0.21 -0.10  1.00  0.04 -0.51 -0.24
## BOP -0.11 -0.40  0.04  1.00  0.42 -0.31
## PRC -0.71 -0.33 -0.51  0.42  1.00 -0.02
## UN  -0.08  0.03 -0.24 -0.31 -0.02  1.00

EuroDist = as.dist(1-R1)
EuroClus = hclust(EuroDist, method = "complete")
plot(EuroClus, hang = -1, labels = variable.names(europe[2:7]), ylab = "1-R D
istances")
```

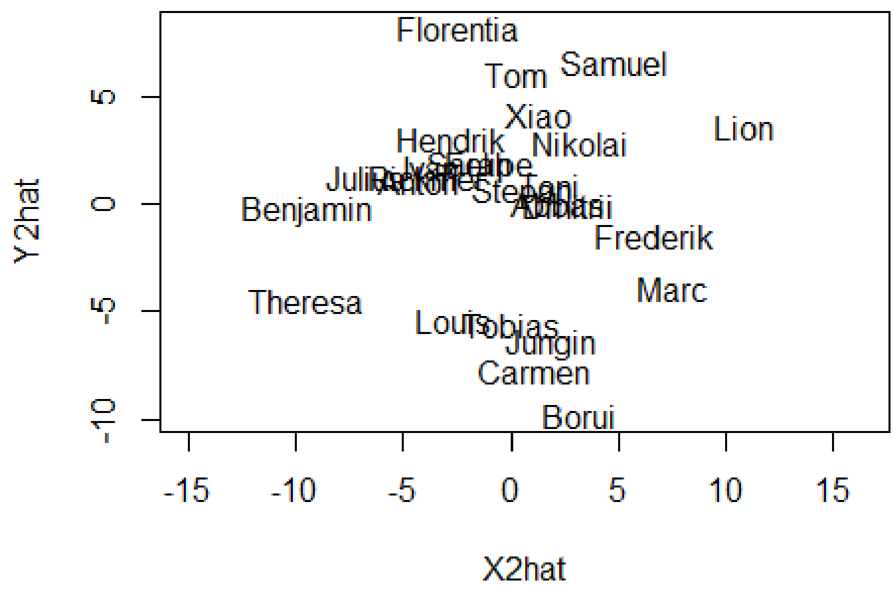


```
### Exercis 2

remove(list = ls())

pers = read.table(file = "Personality2019.txt", header = T, dec =".")
Pers = as.matrix(pers[,2:33])
pers$Name

##  [1] Toni      Lion      Borui     Sarah     Abbas     Marc      Dmitrii
##  [8] Ivan      Benjamin  Julius    Samuel    Theresa   Rickmer   Florentia
## [15] Anton     Frederik  Carmen    Louis     Felipe    Stepan    Nikolai
## [22] Tom       Tobias    Xiao      Hendrik   Jungin
## 26 Levels: Abbas Anton Benjamin Borui Carmen Dmitrii Felipe ... Xiao

# a)

X = scale(Pers, center = TRUE, scale = FALSE)
S = cov(X)
lam = eigen(S)$values[1:2]
e = eigen(S)$vectors[,1:2]
Yhat = X%*%e
plot(Yhat[,1],Yhat[,2], type = "n", asp = 1)
text(Yhat[,1],Yhat[,2], pers$Name)
Expl = cumsum(eigen(S)$values)/sum(eigen(S)$values)
Expl[2]

## [1] 0.3017799

# The first two principal components explain 30.18% of the total variance.

# b)
lengthOfAxes = sqrt(e[,1]^2+e[,2]^2)
o = order(lengthOfAxes, decreasing = TRUE)
arrows(0,0,3*e[o[1],1],3*e[o[1],2], col = "blue")
text(3.5*e[o[1],1],3.5*e[o[1],2], labels = variable.names(pers[o[1]+1]), col
= "blue")
arrows(0,0,3*e[o[2],1],3*e[o[2],2], col = "blue")
text(3.5*e[o[2],1],3.5*e[o[2],2], labels = variable.names(pers[o[2]+1]), col
= "blue")
arrows(0,0,3*e[o[3],1],3*e[o[3],2], col = "blue")
text(3.5*e[o[3],1],3.5*e[o[3],2], labels = variable.names(pers[o[3]+1]), col
= "blue")
arrows(0,0,3*e[o[4],1],3*e[o[4],2], col = "blue")
text(3.5*e[o[4],1],3.5*e[o[4],2], labels = variable.names(pers[o[4]+1]), col
= "blue")

# c)
Chol = c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0)
San = c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0)
```

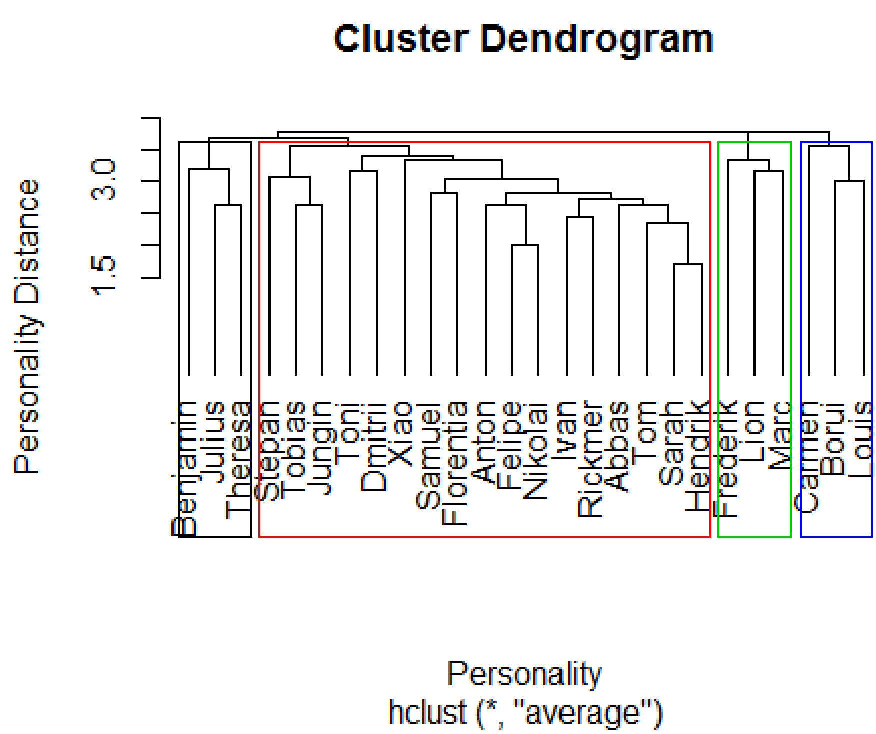
```
Mel = c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0)
phle = c(rep(1,32))
Phle = phle-(Chol+San+Mel)
Choleric = Chol-colMeans(Pers)
C = Choleric%*%e
text(C[1,1],C[1,2], labels = "Cholerica", col = "green")
Melancholic = Mel-colMeans(Pers)
M = Melancholic%*%e
text(M[1,1],M[1,2], labels = "Melancholic", col = "green")
Sanguine = San-colMeans(Pers)
Sa = Sanguine%*%e
text(Sa[1,1],Sa[1,2], labels = "Sanguine", col = "green")
Phlegmatic = Phle-colMeans(Pers)
P = Phlegmatic%*%e
text(P[1,1],P[1,2], labels = "Phlegmatic", col = "green")

# d)
Intro = Phle+Mel
Unst = Mel+Chol
Extro = Chol+San
Stabel = San+Phle
Introvert = Intro-colMeans(Pers)
Int = Introvert%*%e
text(Int[1,1],Int[1,2], labels = "Introvert", col = "red")
Unstable = Unst-colMeans(Pers)
Uns = Unstable%*%e
text(Uns[1,1],Uns[1,2], labels = "Unstable", col = "red")
Extrovert = Extro-colMeans(Pers)
Ex = Extrovert%*%e
text(Ex[1,1],Ex[1,2], labels = "Extrovert", col = "red")
Stabele = Stabel-colMeans(Pers)
St = Stabele%*%e
text(St[1,1],St[1,2], labels = "Stable", col = "red")
```



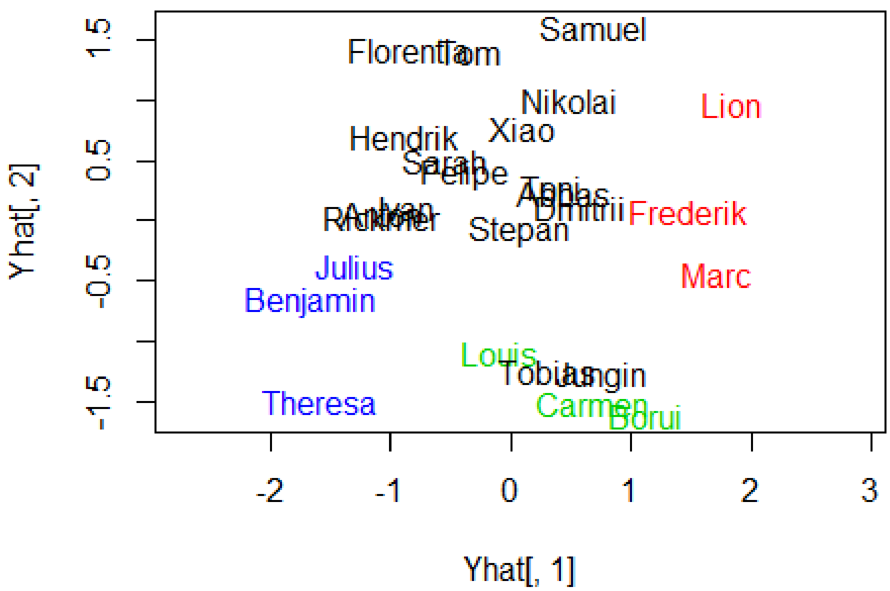

```
# e)
Delta = as.matrix(dist(Pers, method = "manhattan", upper = TRUE, diag = TRUE)
)
Deltastar = Delta^2
I  = diag(1, nrow = 26, ncol = 26)
J = matrix(c(rep(1, 26*26)), nrow = 26, ncol = 26)
n = 1/26
H = I-n*J
B = -0.5*H%*%Deltastar%*%H
reseigen = eigen(B)
lambda = round(reseigen$values, digits = 4)
lambda

##  [1] 580.1522 507.7782 333.5213 296.7573 194.7462 123.1985 103.0154
##  [8]  85.5694  76.6256  52.2152  45.5587  34.9044  26.5002   8.5767
## [15]   5.2770   0.0000  -7.3348 -11.6107 -14.1745 -20.3273 -32.1523
## [22] -35.9884 -50.6483 -54.9865 -74.0686 -86.2971

# Since there are negative eigenvalues, B is not n.n.d.

# f)
QE = reseigen$vectors
Qlam = reseigen$values
QLam = diag(Qlam[1:2])
QYhat = QE[,1:2]%*%sqrt(QLam)
plot(-QYhat[,1],QYhat[,2], asp = 1, type = "n", ylab = "Y2hat", xlab = "X2hat
")
text(-QYhat[,1],QYhat[,2], pers$Name)
```




```
# The biplot differs from the MDS-plot.

# g)
Personality = dist(X, method = "euclidean", upper = TRUE, diag = TRUE)
PersALM = hclust(Personality, method = "average")
plot(PersALM,  hang = -1, labels = pers$Name, ylab = "Personality Distance")
rect.hclust(PersALM, k = 4, border = c(1,2,3,4))
```

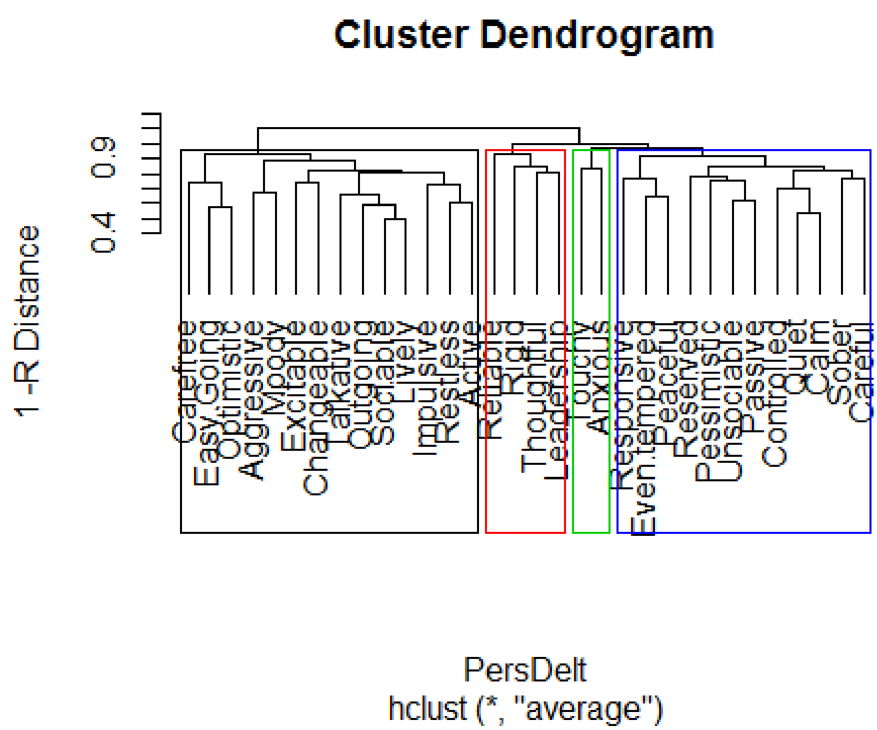


```r
plot(Yhat[,1],Yhat[,2], type = "n", asp = 1)
text(Yhat[,1],Yhat[,2], pers$Name, col = cutree(PersALM, k = 4))
```



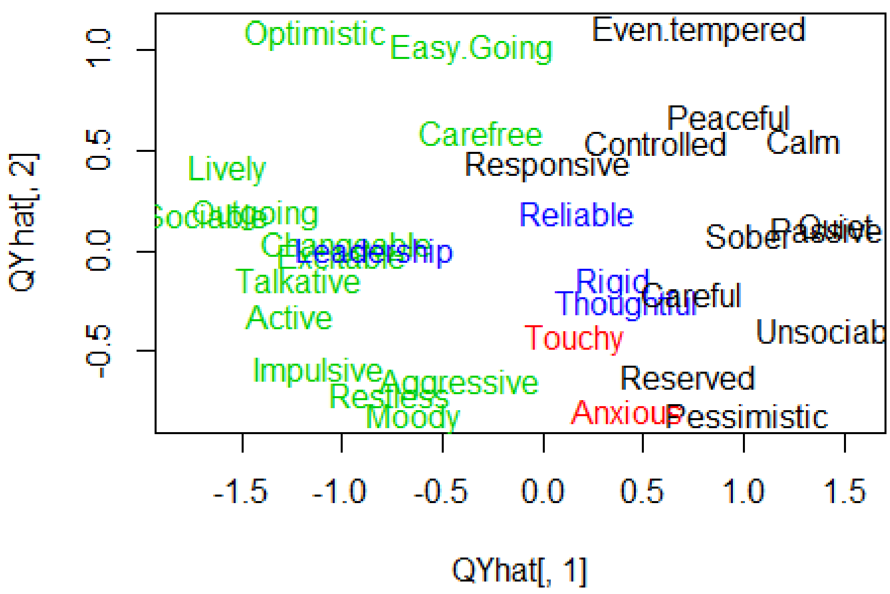

```r
# The cluster does correspond more or less to the biplot except the observati
ons Tobias and Jungin.
```

```
### Exercise 3

remove(list = ls())

pers3 = read.table(file = "Personality.txt", header = T, dec =".")

# a)
Pers3 = scale(pers3, scale = FALSE, center = TRUE)
R = cor(pers3)
PersDelt = as.dist(1-R)
Persreshc = hclust(PersDelt, method = "average")
plot(Persreshc,  hang = -1, labels = pers3$Name, ylab = "1 -R Distance")
rect.hclust(Persreshc, k = 4, border = c(1,2,3,4) )
```




```
Group = sort(cutree(Persreshc, k = 4))
Group[1:12]

##         Quiet          Calm    Unsociable Even.tempered      Reserved
##             1             1             1             1             1
##   Pessimistic    Responsive    Controlled         Sober      Peaceful
##             1             1             1             1             1
##       Careful       Passive
##             1             1

# Cluster 1
Group[13:14]
```

```
##   Touchy Anxious
##        2       2

# Cluster 2
Group[15:28]

##    Sociable   Restless   Outgoing Aggressive  Talkative  Excitable
##           3          3          3          3          3          3
## Changeable Easy.Going  Impulsive     Lively Optimistic   Carefree
##           3          3          3          3          3          3
##       Moody     Active
##           3          3

# Cluster 3
Group[29:32]

##    Reliable      Rigid Thoughtful Leadership
##           4          4          4          4

# Cluster 4

# b)
X3 = R
Q = X3%*%t(X3)
QE = eigen(Q)$vectors
Qlam = eigen(Q)$values
QLam = diag(Qlam[1:2])
QYhat = QE[,1:2]%*%sqrt(QLam)
plot(QYhat[,1],QYhat[,2], asp = 1, type = "n")
text(QYhat[,1],QYhat[,2], variable.names(pers3), col = cutree(Persreshc, k =
4))
```



```
# c)
Chol = c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0)
San = c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0)
CSan = 2*San
Mel = c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0)
CMel = 3*Mel
phle = c(rep(1,32))
Phle = phle-(Chol+San+Mel)
CPhle = 4*Phle
Color = CPhle+Chol+CSan+CMel
plot(QYhat[,1],QYhat[,2], asp = 1, type = "n")
text(QYhat[,1],QYhat[,2], variable.names(pers3), col = Color)
```