The 2000 World's ten largest companies (shown in the table) are available in file `"10largest.txt"` on Blackboard. In this example we find principal components `PC1, PC2` of variables `sales, profit`

World's Ten Largest Companies

Company	Sales	Profits	Assets
Citigroup	108.28	17.05	1484.10
GE	152.36	16.59	750.33
American Intl. Group	95.04	10.91	766.42
Bank of America	65.45	14.14	1110.46
HSBC Group	62.97	9.52	1031.29
Exxon Mobil	263.99	25.33	195.26
Royal Dutch/Shell	265.19	18.54	193.83
BP	285.06	15.73	191.11
ING Group	92.01	8.10	1175.16
Toyota	165.68	11.13	211.15

Forbes, The Forbes Global 2000

a) Find principal components (PC1, PC2) when the data is centered and not scaled. Find their covariance matrix. Verify that their eigenvalues do not add up to one.

b) Find principal components when the data is centered and scaled. Verify that their eigenvalues add up to one.

```
d0=read.table("10largest.txt")
names(d0)=c("sales","profit","assets")

# only sales and profit
d1=d0[,c(1,2)]

# centering, no scaling
#==================================================================

pr1=prcomp(d1)
names(pr1)
# [1] "sdev"     "rotation" "center"   "scale"    "x"

# rotation matrix
rot = pr1$rotation
rot
#              PC1         PC2
#sales  0.99917338  0.04065165
#profit 0.04065165 -0.99917338

# columns are eigenvectors (orthogonal)

d1b = scale(d1,scale=F)
#        sales profit
# [1,] -47.323  2.346
# [2,]  -3.243  1.886
# [3,] -60.563 -3.794
# [4,] -90.153 -0.564
# [5,] -92.633 -5.184
# [6,] 108.387 10.626
# [7,] 109.587  3.836
# [8,] 129.457  1.026
# [9,] -63.593 -6.604
#[10,]  10.077 -3.574

colMeans(d1b)
#        sales        profit
#-4.263256e-15 -7.105427e-16
```

```
# transformed data (score vectors)

as.matrix(d1b)%*%rot
#             PC1        PC2
# [1,] -47.188513 -4.2678188
# [2,]  -3.163650 -2.0162743
# [3,] -60.667170  1.3288779
# [4,] -90.101405 -3.1013344
# [5,] -92.767166  1.4140305
# [6,] 108.729370 -6.2111060
# [7,] 109.652353  0.6220633
# [8,] 129.391697  4.2374888
# [9,] -63.808896  4.0133806
#[10,]   9.923381  3.9806923

pr1$x
#             PC1        PC2
# [1,] -47.188513 -4.2678188
# [2,]  -3.163650 -2.0162743
# [3,] -60.667170  1.3288779
# [4,] -90.101405 -3.1013344
# [5,] -92.767166  1.4140305
# [6,] 108.729370 -6.2111060
# [7,] 109.652353  0.6220633
# [8,] 129.391697  4.2374888
# [9,] -63.808896  4.0133806
#[10,]   9.923381  3.9806923


#  prcomp() agrees with eigen()

# eigenvalues
pr1$sdev^2
# [1] 7488.80605   13.83751

eigen(var(d1))
# $values
# [1] 7488.80605   13.83751

# $vectors
#             [,1]        [,2]
# [1,] -0.99917338  0.04065165
# [2,] -0.04065165 -0.99917338

# eigen(var(d1b)) same results
```

```
d1x = pr1$x
var(d1x)
#             PC1          PC2
#PC1 7.488806e+03 1.275985e-15
#PC2 1.275985e-15 1.383751e+01

# PCs uncorrelated
# but sum of eigenvals do not add up to p=2

# centering and scaling
#==================================================================

pr3=prcomp(d1,scale=T)
d3x = pr3$x
var(d3x)
#             PC1          PC2
#PC1 1.686136e+00 1.387252e-16
#PC2 1.387252e-16 3.138640e-01

sum(diag(var(d3x)))   # [1] 2
```