Lab 5: PCA and FA

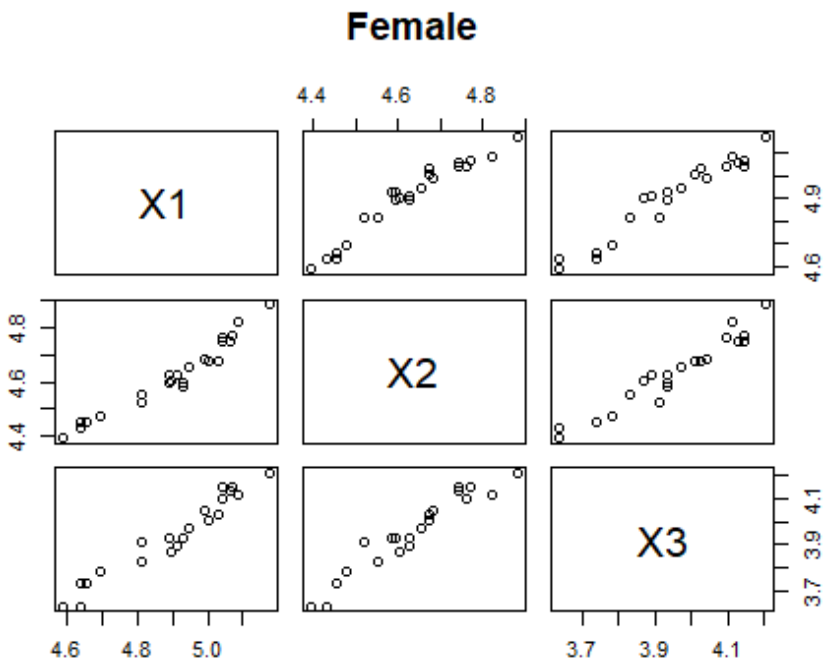
Spring 2018 - Multivariate Data Analysis

Example 1 : Female Turtle

- Data on three dimensions of female turtle carapace(shells) :
  - X1 = log(carapace length)
  - X2 = log(carapace width)
  - X3 = log(carapace height)

*Since the measurements are all on the same scale, we extracted the PCs from the sample covariance matrix S

```
female<-read.csv("./data/ex8-4-female.csv")
female<-log(female)
pairs(female,main="Female")
```




```
var(female)
```

```
##             X1          X2          X3
## X1 0.02640563 0.02011195 0.02491758
## X2 0.02011195 0.01619045 0.01942430
## X3 0.02491758 0.01942430 0.02493980
```

```r
female.PC<-princomp(female)
female.PC
```

```
## Call:
## princomp(x = female)
##
## Standard deviations:
##     Comp.1     Comp.2     Comp.3
## 0.25192540 0.02709137 0.02283700
##
##  3  variables and  24 observations.
```

```r
prcomp(female)
```

```
## Standard deviations (1, .., p=3):
## [1] 0.25734377 0.02767404 0.02332817
##
## Rotation (n x k) = (3 x 3):
##          PC1        PC2         PC3
## X1 0.6266648 -0.5525704 -0.54950625
## X2 0.4878158 -0.2717450  0.82957243
## X3 0.6077228  0.7879217 -0.09925955
```

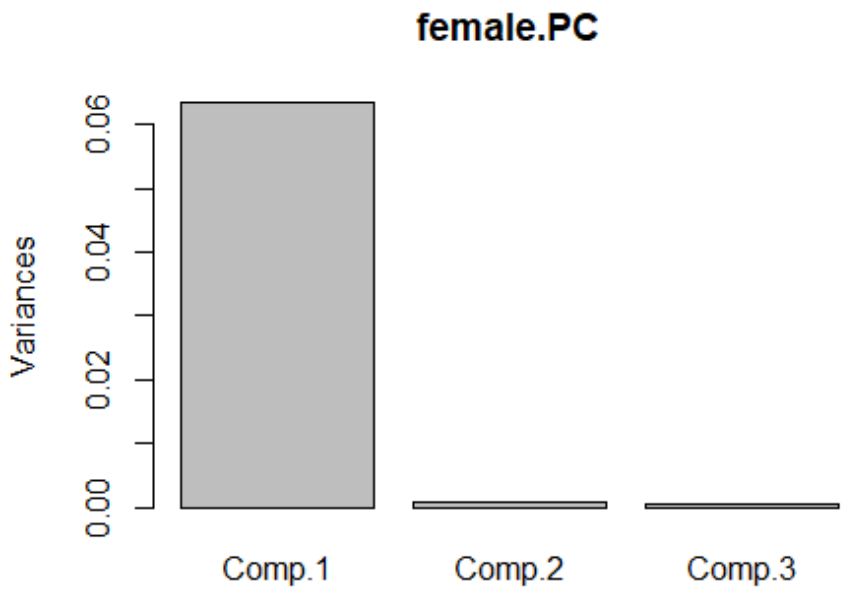
```r
summary(female.PC)
```

```
## Importance of components:
##                            Comp.1     Comp.2      Comp.3
## Standard deviation     0.2519254 0.02709137 0.022836998
## Proportion of Variance 0.9806021 0.01133994 0.008057994
## Cumulative Proportion  0.9806021 0.99194201 1.000000000
```

```r
loadings(female.PC)
```

```
##
## Loadings:
##    Comp.1 Comp.2 Comp.3
## X1 -0.627 -0.553  0.550
## X2 -0.488 -0.272 -0.830
## X3 -0.608  0.788
##
##                Comp.1 Comp.2 Comp.3
## SS loadings     1.000  1.000  1.000
## Proportion Var  0.333  0.333  0.333
## Cumulative Var  0.333  0.667  1.000
```

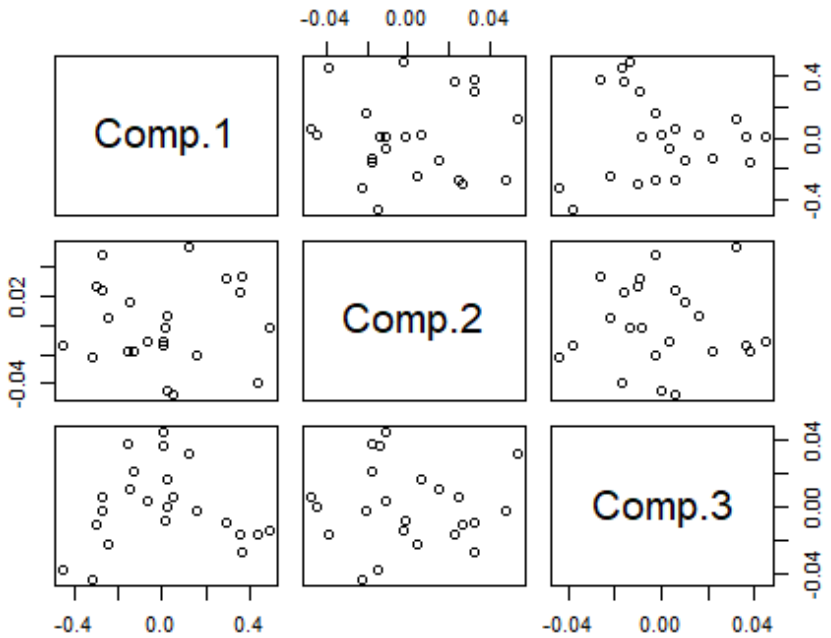
```r
plot(female.PC)
```




```
predict(female.PC)
```

```
##             Comp.1       Comp.2        Comp.3
##  [1,]  0.493236696 -0.001978796 -1.399655e-02
##  [2,]  0.444312197 -0.039358259 -1.682188e-02
##  [3,]  0.372010654  0.033105374 -2.640789e-02
##  [4,]  0.359959036  0.022478693 -1.584014e-02
##  [5,]  0.297043663  0.032226292 -9.749373e-03
##  [6,]  0.121948174  0.054098775  3.246375e-02
##  [7,]  0.156967889 -0.020319242 -2.432252e-03
##  [8,]  0.025158400  0.006582748  1.654778e-02
##  [9,]  0.010595653 -0.001529645 -8.217417e-03
## [10,]  0.010595653 -0.001529645 -8.217417e-03
## [11,]  0.052404509 -0.048054959  6.308895e-03
## [12,]  0.020929572 -0.045376218  6.882489e-05
## [13,]  0.006984099 -0.011050763  4.524919e-02
## [14,]  0.002031612 -0.013809619  3.682706e-02
## [15,] -0.063525729 -0.011374505  3.650546e-03
## [16,] -0.147600256  0.015271775  1.040223e-02
## [17,] -0.129824283 -0.017810680  2.199975e-02
## [18,] -0.157375903 -0.018251992  3.834557e-02
## [19,] -0.272267077  0.047781769 -2.641700e-03
## [20,] -0.251027010  0.004653566 -2.178792e-02
## [21,] -0.274556418  0.024582024  6.304102e-03
## [22,] -0.300796398  0.026704668 -1.000434e-02
```

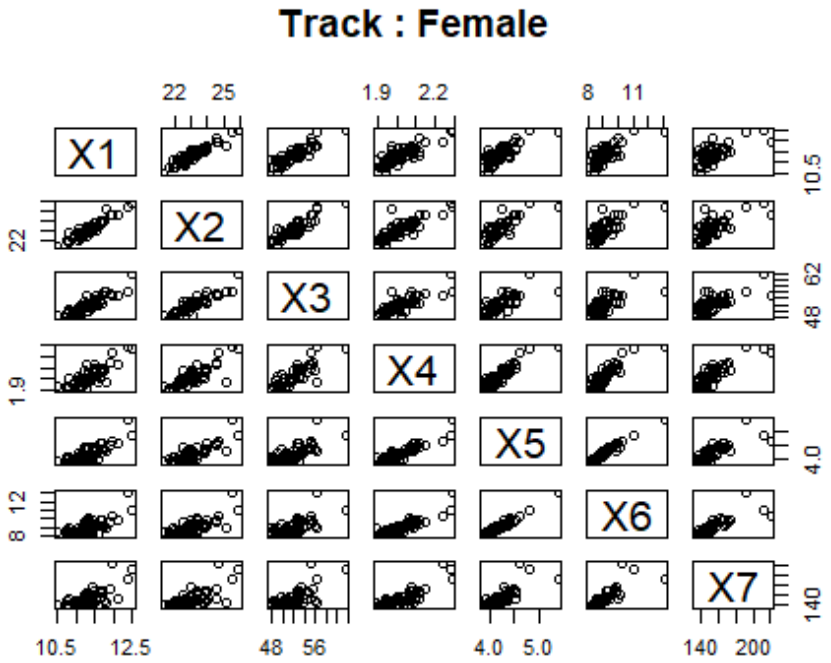
```
## [23,] -0.317098612 -0.022520805 -4.407935e-02
## [24,] -0.460106121 -0.014520553 -3.797147e-02

pairs(predict(female.PC))
```



Example 2 : National Track Records - Female

- The national track records for women in 54 countries
- 7 variables :
  - X1 : 100m (sec)
  - X2 : 200m (sec)
  - X3 : 400m (sec)
  - X4 : 800m (min)
  - X5 : 1500m (min)
  - X6 : 3000m (min)
  - X7 : Marathon (min)

```
track<-read.csv("./data/pca-1.csv")
rownames(track)<-track[,1]
track<-track[,-1]
pairs(track,main="Track : Female")
```




```
round(var(track),3)
```

```
##        X1     X2     X3    X4    X5     X6      X7
## X1 0.155  0.345  0.891 0.028 0.084  0.234   4.334
## X2 0.345  0.863  2.193 0.066 0.203  0.554  10.385
## X3 0.891  2.193  6.745 0.182 0.509  1.427  28.904
## X4 0.028  0.066  0.182 0.008 0.021  0.061   1.220
## X5 0.084  0.203  0.509 0.021 0.074  0.216   3.540
## X6 0.234  0.554  1.427 0.061 0.216  0.665  10.706
## X7 4.334 10.385 28.904 1.220 3.540 10.706 270.270
```

```
round(cor(track),3)
```

```
##        X1    X2    X3    X4    X5    X6    X7
## X1 1.000 0.941 0.871 0.809 0.782 0.728 0.669
## X2 0.941 1.000 0.909 0.820 0.801 0.732 0.680
## X3 0.871 0.909 1.000 0.806 0.720 0.674 0.677
## X4 0.809 0.820 0.806 1.000 0.905 0.867 0.854
## X5 0.782 0.801 0.720 0.905 1.000 0.973 0.791
## X6 0.728 0.732 0.674 0.867 0.973 1.000 0.799
## X7 0.669 0.680 0.677 0.854 0.791 0.799 1.000
```

```
track.pc<-princomp(track,cor=TRUE)
summary(track.pc)
```

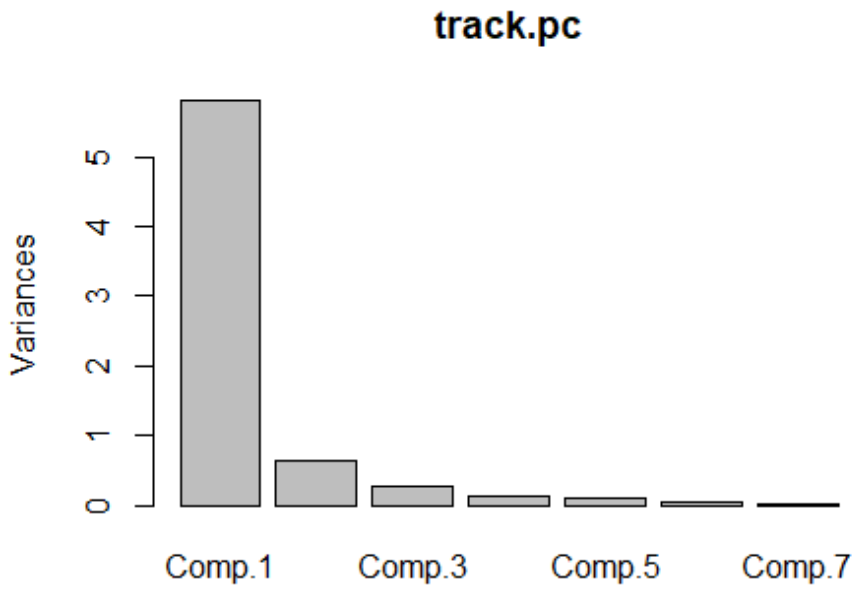
```
## Importance of components:
##                           Comp.1     Comp.2     Comp.3     Comp.4
```

```
## Standard deviation     2.4099013 0.79290190 0.52852112 0.35292310
## Proportion of Variance 0.8296606 0.08981335 0.03990494 0.01779353
## Cumulative Proportion  0.8296606 0.91947398 0.95937892 0.97717245
##                            Comp.5      Comp.6      Comp.7
## Standard deviation     0.30161522 0.233492660 0.119592075
## Proportion of Variance 0.01299596 0.007788403 0.002043181
## Cumulative Proportion  0.99016842 0.997956819 1.000000000
```

```
loadings(track.pc)
```

```
##
## Loadings:
##    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## X1 -0.378 -0.407 -0.141  0.587 -0.167 -0.540
## X2 -0.383 -0.414 -0.101  0.194         0.745 -0.266
## X3 -0.368 -0.459  0.237 -0.645  0.327 -0.240  0.127
## X4 -0.395  0.161  0.148 -0.295 -0.819        -0.195
## X5 -0.389  0.309 -0.422                0.189  0.731
## X6 -0.376  0.423 -0.406         0.352 -0.240 -0.572
## X7 -0.355  0.389  0.741  0.321  0.247
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var  0.143  0.143  0.143  0.143  0.143  0.143  0.143
## Cumulative Var  0.143  0.286  0.429  0.571  0.714  0.857  1.000
```

```
plot(track.pc)
```




```
track.pc1<-princomp(track)
summary(track.pc1)
```

```
## Importance of components:
##                             Comp.1     Comp.2       Comp.3       Comp.4
## Standard deviation     16.409811 1.98552239 0.5131175570 0.3352983644
## Proportion of Variance  0.984153 0.01440805 0.0009622529 0.0004108832
## Cumulative Proportion   0.984153 0.99856109 0.9995233458 0.9999342290
##                               Comp.5       Comp.6       Comp.7
## Standard deviation     1.218369e-01 5.046340e-02 2.460388e-02
## Proportion of Variance 5.425166e-05 9.306974e-06 2.212396e-06
## Cumulative Proportion  9.999885e-01 9.999978e-01 1.000000e+00
```

```
loadings(track.pc1)
```

```
##
## Loadings:
##    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## X1        -0.115 -0.173  0.292  0.933
## X2        -0.290 -0.387  0.795 -0.354
## X3 -0.108 -0.938  0.226 -0.238
## X4                                0.377 -0.925
## X5               -0.268             0.883  0.370
## X6               -0.834 -0.471      -0.265
## X7 -0.992  0.119
##
```

```
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var  0.143  0.143  0.143  0.143  0.143  0.143  0.143
## Cumulative Var  0.143  0.286  0.429  0.571  0.714  0.857  1.000
```

Example 3 : Stock price data

- The weekly rates of return for five stocks (Allied Chemical, du Pont, Union Carbide, Exxon, and Texaco) listed on the NY stock exchange were determined for the period January 1975 through December 1976.
- The observations in 100 successive weeks appear to be independently distributed, but the rates of return across stocks are correlated.
  - X1 : Allied Chemical (AlliedChem)
  - X2 : du Pont (DuPont)
  - X3 : Union Carbide (UnionCarb)
  - X4 : Exxon
  - X5 : Texaco

```
stocks<-read.csv("./data/stocks.csv")
head(stocks)
```

```
##   AlliedChem    DuPont UnionCarb     Exxon    Texaco
## 1   0.000000  0.000000  0.000000  0.039473  0.000000
## 2   0.027027 -0.044855 -0.003030 -0.014466  0.043478
## 3   0.122807  0.060773  0.088146  0.086238  0.078124
## 4   0.057031  0.029948  0.066808  0.013513  0.019512
## 5   0.063670 -0.003793 -0.039788 -0.018644 -0.024154
## 6   0.003521  0.050761  0.082873  0.074265  0.049504
```

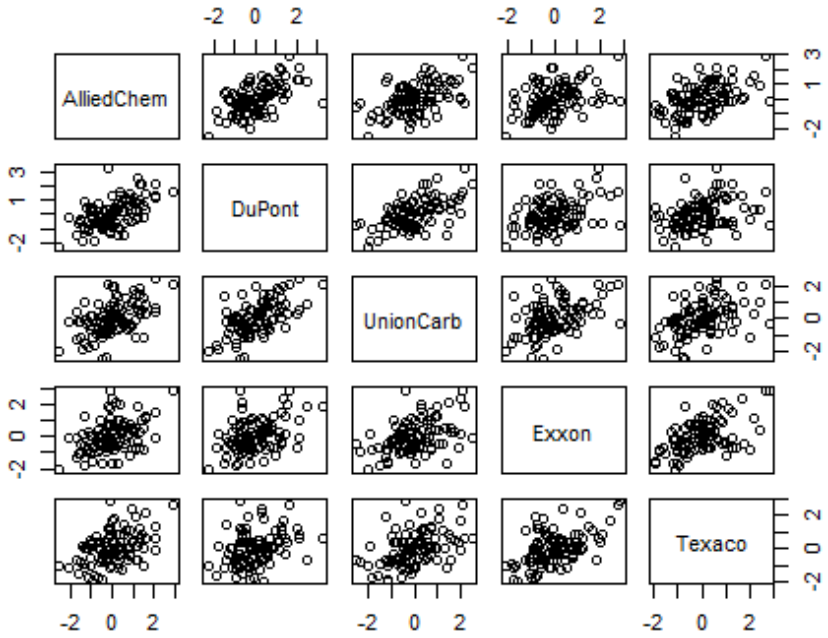
```
stockss<-scale(stocks, center=T,scale=T)
apply(stockss,2,mean)
```

```
##    AlliedChem        DuPont     UnionCarb         Exxon        Texaco
## -3.314162e-17 -7.122124e-18 -1.505415e-18 -1.189370e-17 -1.655865e-18
```

```
apply(stockss,2,sd)
```

```
## AlliedChem     DuPont  UnionCarb      Exxon     Texaco
##          1          1          1          1          1
```

```
pairs(stockss)
```




```
s.cor<-var(stockss)
s.cor
```

```
##             AlliedChem    DuPont UnionCarb     Exxon    Texaco
## AlliedChem  1.0000000 0.5769308 0.5086555 0.3867206 0.4621781
## DuPont      0.5769308 1.0000000 0.5983817 0.3895188 0.3219545
## UnionCarb   0.5086555 0.5983817 1.0000000 0.4361014 0.4256266
## Exxon       0.3867206 0.3895188 0.4361014 1.0000000 0.5235293
## Texaco      0.4621781 0.3219545 0.4256266 0.5235293 1.0000000
```

```
s.pc1<-princomp(stocks,cor=TRUE)
s.pc1
```

```
## Call:
## princomp(x = stocks, cor = TRUE)
##
## Standard deviations:
##    Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## 1.6901150 0.8995104 0.7348756 0.6718246 0.5856636
##
##  5  variables and  100 observations.
```

```
s.pc2<-prcomp(stocks,scale=T,center=T)
s.pc2
```

```
## Standard deviations (1, .., p=5):
## [1] 1.6901150 0.8995104 0.7348756 0.6718246 0.5856636
```

```
##
## Rotation (n x k) = (5 x 5):
##                  PC1        PC2        PC3        PC4        PC5
## AlliedChem 0.4635414 -0.2408580  0.6133475 -0.3813591  0.4533066
## DuPont     0.4570769 -0.5090981 -0.1778906 -0.2113173 -0.6749813
## UnionCarb  0.4699796 -0.2605708 -0.3370501  0.6641056  0.3957057
## Exxon      0.4216766  0.5252677 -0.5390141 -0.4728058  0.1794471
## Texaco     0.4213290  0.5822399  0.4336136  0.3812200 -0.3874650
```

```
s.pc2$rotation # loadings(s.pc1)
```

```
##                  PC1        PC2        PC3        PC4        PC5
## AlliedChem 0.4635414 -0.2408580  0.6133475 -0.3813591  0.4533066
## DuPont     0.4570769 -0.5090981 -0.1778906 -0.2113173 -0.6749813
## UnionCarb  0.4699796 -0.2605708 -0.3370501  0.6641056  0.3957057
## Exxon      0.4216766  0.5252677 -0.5390141 -0.4728058  0.1794471
## Texaco     0.4213290  0.5822399  0.4336136  0.3812200 -0.3874650
```

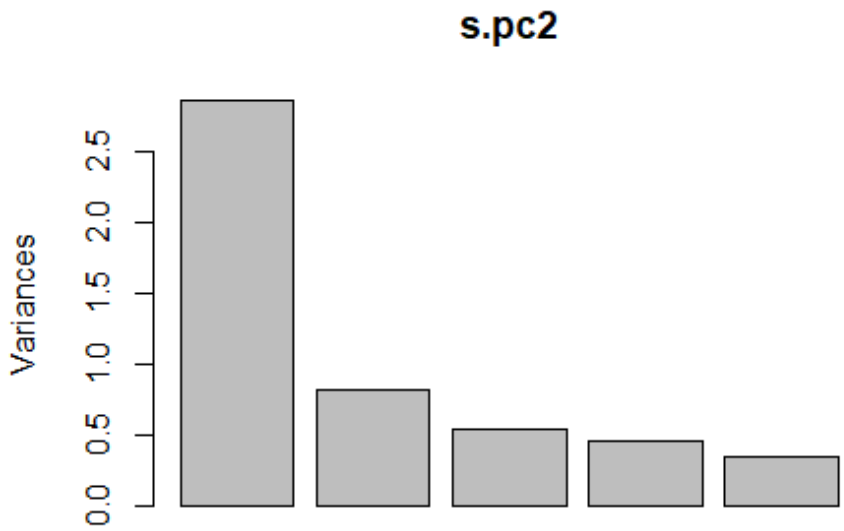
```
head(s.pc2$x) #s.pc1[[6]]
```

```
##             PC1        PC2        PC3        PC4        PC5
## [1,]  0.2445666  0.6767813 -0.6995513 -0.6199602  0.2375697
## [2,] -0.2038992  1.1056186  1.6753621  0.8461387  0.4208396
## [3,]  5.3881865  0.9980249  0.4445948 -0.3616822  0.5280784
## [4,]  1.9977363 -0.6085699  0.2452619  0.4889273  0.5326387
## [5,] -0.7825190 -0.9734526  1.3526432 -1.2325779  0.5979402
## [6,]  3.2092353  1.0628920 -1.4944745  0.5404242 -0.3447039
```

```
head(predict(s.pc1))
```

```
##          Comp.1     Comp.2     Comp.3     Comp.4     Comp.5
## [1,] -0.2457987 -0.6801908  0.7030755 -0.6230835 -0.2387665
## [2,]  0.2049264 -1.1111885 -1.6838023  0.8504014 -0.4229597
## [3,] -5.4153312 -1.0030528 -0.4468346 -0.3635043 -0.5307388
## [4,] -2.0078006  0.6116357 -0.2464975  0.4913904 -0.5353220
## [5,]  0.7864612  0.9783567 -1.3594576 -1.2387874 -0.6009525
## [6,] -3.2254028 -1.0682467  1.5020033  0.5431468  0.3464405
```

```
plot(s.pc2)
```




```
F1<-factanal(stocks,factors=1)
F1

##
## Call:
## factanal(x = stocks, factors = 1)
##
## Uniquenesses:
## AlliedChem     DuPont  UnionCarb      Exxon     Texaco
##      0.474      0.458      0.439      0.648      0.651
##
## Loadings:
##            Factor1
## AlliedChem 0.726
## DuPont     0.736
## UnionCarb  0.749
## Exxon      0.593
## Texaco     0.591
##
##                Factor1
## SS loadings      2.330
## Proportion Var   0.466
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 15.49 on 5 degrees of freedom.
## The p-value is 0.00847
```

```r
F2<-factanal(stocks,factors=2)
F2
```

```
##
## Call:
## factanal(x = stocks, factors = 2)
##
## Uniquenesses:
## AlliedChem     DuPont  UnionCarb      Exxon     Texaco
##      0.497      0.252      0.474      0.610      0.176
##
## Loadings:
##            Factor1 Factor2
## AlliedChem 0.601   0.378
## DuPont     0.849   0.165
## UnionCarb  0.643   0.336
## Exxon      0.365   0.507
## Texaco     0.207   0.884
##
##                Factor1 Factor2
## SS loadings      1.671   1.321
## Proportion Var   0.334   0.264
## Cumulative Var   0.334   0.598
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 0.58 on 1 degree of freedom.
## The p-value is 0.448
```

```r
F2.1<-factanal(stocks,factors=2,rotation="none")
F2.1
```

```
##
## Call:
## factanal(x = stocks, factors = 2, rotation = "none")
##
## Uniquenesses:
## AlliedChem     DuPont  UnionCarb      Exxon     Texaco
##      0.497      0.252      0.474      0.610      0.176
##
## Loadings:
##            Factor1 Factor2
## AlliedChem  0.683   0.192
## DuPont      0.692   0.519
## UnionCarb   0.680   0.251
## Exxon       0.621
## Texaco      0.794  -0.439
##
##                Factor1 Factor2
## SS loadings      2.424   0.567
```

```
## Proportion Var   0.485   0.113
## Cumulative Var   0.485   0.598
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 0.58 on 1 degree of freedom.
## The p-value is 0.448
```

```
F2.2<-factanal(stocks,factors=2,scores="regression")
F2.2
```

```
##
## Call:
## factanal(x = stocks, factors = 2, scores = "regression")
##
## Uniquenesses:
## AlliedChem     DuPont  UnionCarb      Exxon     Texaco
##      0.497      0.252      0.474      0.610      0.176
##
## Loadings:
##            Factor1 Factor2
## AlliedChem 0.601   0.378
## DuPont     0.849   0.165
## UnionCarb  0.643   0.336
## Exxon      0.365   0.507
## Texaco     0.207   0.884
##
##                Factor1 Factor2
## SS loadings      1.671   1.321
## Proportion Var   0.334   0.264
## Cumulative Var   0.334   0.598
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 0.58 on 1 degree of freedom.
## The p-value is 0.448
```