assigmnet3

QUANG HUY NGUYEN - 45768161

29/10/2021

Question 1

a)

```
library(readxl)
magnes <- read_excel("magnes.xls", col_types = c("numeric", "numeric", "numeric", "skip", "numeric", "nu
```

```
## New names:
## * `Magnesium (y)` -> `Magnesium (y)...1`
## * `Time (x)` -> `Time (x)...2`
## * Treatment -> Treatment...3
## * `Magnesium (y)` -> `Magnesium (y)...4`
## * `Time (x)` -> `Time (x)...5`
## * ...
```

```
magnes <- data.frame(Magnesium = c(magnes[[1]],magnes[[4]]),Time = c(magnes[[2]],magnes[[5]]),Treatment
magnes$Treatment = as.factor(magnes$Treatment)
magnes$Time = as.numeric(magnes$Time)
head(magnes)
```

```
##   Magnesium Time Treatment
## 1      2.08    0         1
## 2      2.07    0         1
## 3      2.09    0         1
## 4      2.06    0         1
## 5      2.17    1         1
## 6      2.15    1         1
```

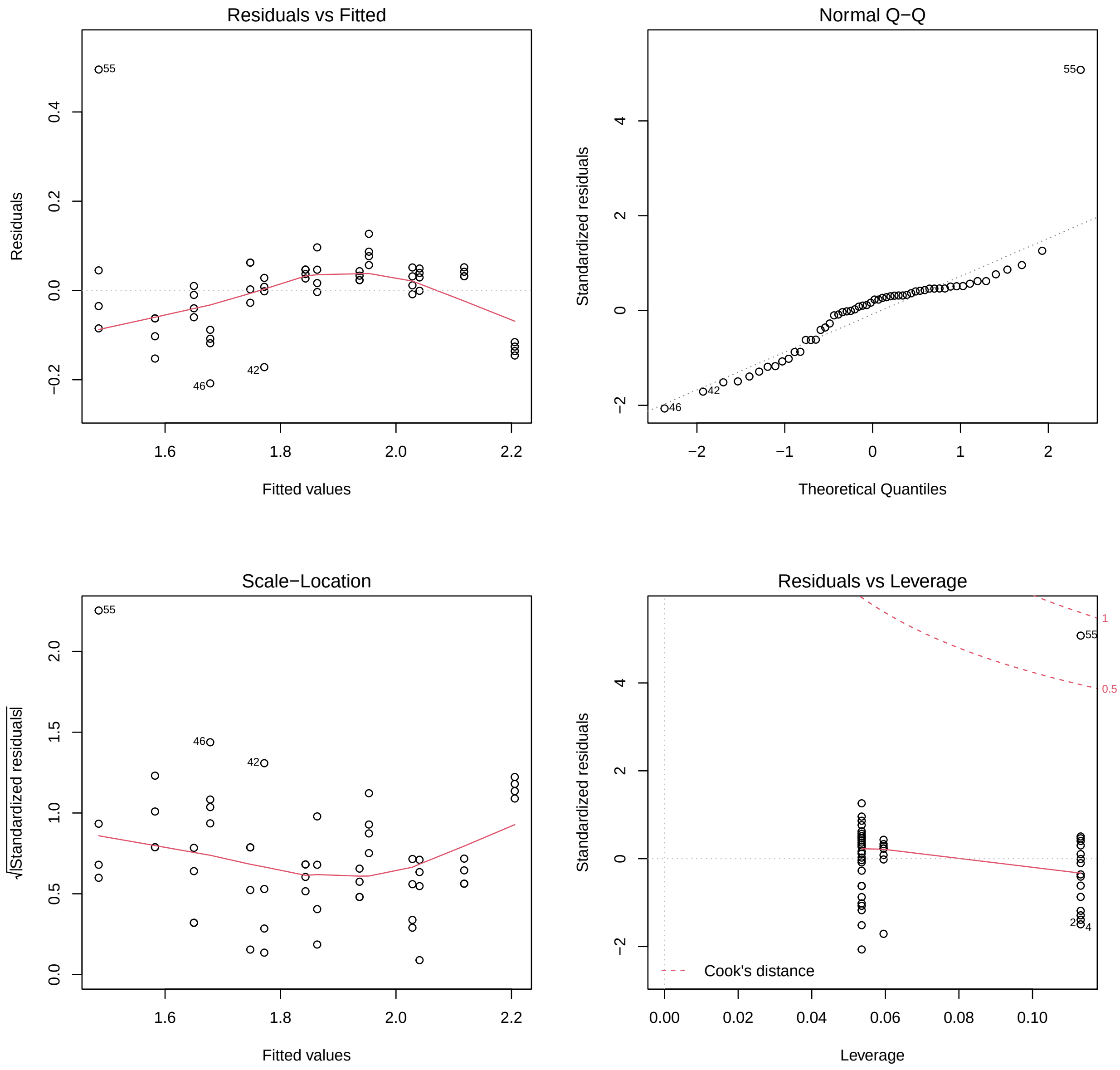
```
fit = lm(Magnesium ~ Time + I(Time^2) + Treatment, data = magnes)
summary(fit)
```

```
##
## Call:
## lm(formula = Magnesium ~ Time + I(Time^2) + Treatment, data = magnes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20821 -0.06062  0.01984  0.04540  0.49503
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.205774   0.034819  63.350  < 2e-16 ***
## Time        -0.086652   0.024942  -3.474  0.00104 **
## I(Time^2)   -0.000997   0.003994  -0.250  0.80386
## Treatment2  -0.165000   0.027671  -5.963  2.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1035 on 52 degrees of freedom
## Multiple R-squared:  0.8052, Adjusted R-squared:  0.794
## F-statistic: 71.64 on 3 and 52 DF,  p-value: < 2.2e-16
```

- The model's R-squared is 0.794, which means 79% observation of Magnesium can be explain by 3 above variables.
- Time^2 is not significant.

```
par(mfrow = c(2, 2))
plot(fit)
```



- Linear assumption is not satisfied.
- Evident of Heteroscedasticity.
- Q-Q plot show normality.
- There is 1 extreme points (55) that is exceed 4 standard deviations.

b) When combine two given models, we have:

$$E(y_{(56\times1)}) = X_{(56\times3)}\beta_{(1\times3)}$$

With:

$$X = \begin{bmatrix} 1 & 0 & x_1^1 & 0 & x_1^3 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_{56}^1 & 0 & x_{56}^3 & 0 \\ 0 & 1 & 0 & x_1^2 & 0 & x_1^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & x_{56}^2 & 0 & x_{56}^4 \end{bmatrix}$$

And

$$\beta = \begin{bmatrix} \beta_0 \\ \gamma_0 \\ \beta_1 \\ \gamma_1 \\ \beta_2 \\ \gamma_2 \end{bmatrix}$$

Assume that:

$$y_{n\text{Ö}1} \sim N_{56}(\mu, \sigma^2 I_n)$$

With

$$\mu = X\beta.$$

$$H_0 : C\beta_{(1\times 6)} = 0$$

With

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

c) T2 test:

```
fit1 = lm(Magnesium ~ Time + I(Time^2), data = magnes)
b <- fit1$coef
X <- model.matrix(fit1)
C <- rbind(c(0,1,0), c(0,0,1))
sig2 <- sum(fit1$res^2)/(dim(magnes)[1]-3)
V <- sig2*C%*%solve(t(X)%*%X)%*%t(C)
d<-t(C%*%b)%*%solve(V)%*%C%*%b
d
```

```
##          [,1]
## [1,] 108.5795
```

```
cv <- 2*qf(.95, 2, dim(magnes)[1]-3)
cv
```

```
## [1] 6.343252
```

- The T value is greater than critical value. Hence, reject the hypothesis.

d)

```r
rssf <-sum(fit1$res^2)

fit2 <- lm(Magnesium ~ Time + I(Time^2), data = magnes[28:56,])
rssr <- sum(fit2$residuals^2)
# test statistic:
f = ((rssr - rssf)/(3-1)) / (rssf/(dim(magnes)[1]-3))
f
```

```
## [1] -15.51847
```

```r
cv <- 2*qf(.95, 2, dim(magnes)[1]-3)
cv
```

```
## [1] 6.343252
```

Question 2

a)

```r
air <- read_excel("air.xls", col_types = c("numeric",
    "numeric", "numeric", "numeric", "numeric",
    "numeric", "numeric"))
R <- cor(air)
R
```

```
##                         Wind (x1) Solar radiation (x2)    CO (x3)      NO (x4)
## Wind (x1)               1.0000000          -0.10144191 -0.1938032 -0.26954261
## Solar radiation (x2) -0.1014419           1.00000000  0.1827934 -0.07356907
## CO (x3)              -0.1938032           0.18279338  1.0000000  0.50215246
## NO (x4)              -0.2695426          -0.07356907  0.5021525  1.00000000
## NO2 (x5)             -0.1098249           0.11573199  0.5565838  0.29689814
## O3 (x6)              -0.2535928           0.31912373  0.4109288 -0.13395214
## HC (x7)               0.1560979           0.05201044  0.1660323  0.23470432
##                         NO2 (x5)    O3 (x6)    HC (x7)
## Wind (x1)            -0.1098249 -0.2535928 0.15609793
## Solar radiation (x2)  0.1157320  0.3191237 0.05201044
## CO (x3)               0.5565838  0.4109288 0.16603235
## NO (x4)               0.2968981 -0.1339521 0.23470432
## NO2 (x5)              1.0000000  0.1666422 0.44776780
## O3 (x6)               0.1666422  1.0000000 0.15445056
## HC (x7)               0.4477678  0.1544506 1.00000000
```

```r
er <- eigen(R)
v <- er$values
id <- er$vectors
v
```

```
## [1] 2.3367826 1.3860007 1.2040659 0.7270865 0.6534765 0.5366888 0.1558989
```

```
cumsum(v)/sum(v)
```

```
## [1] 0.3338261 0.5318262 0.7038356 0.8077051 0.9010589 0.9777287 1.0000000
```

- The first values will explain ~80% of the total variability. Hence select 4 common factors.

c)

```
S <- var(air)
#assume m = 2
#v[c(1,2)]/sum(v) #variance explained by two factors
Lam = id[, c(1,2)]%*%diag(sqrt(v[c(1,2)])) #factor loading matrix lambda
Psi = diag(S-Lam%*%t(Lam)) #specific variances
resi = S-(Lam%*%t(Lam)+diag(Psi)) #residual matrix
resi
```

```
##                       Wind (x1) Solar radiation (x2)     CO (x3)    NO (x4)
## Wind (x1)            0.00000000           -2.4634939 -0.07044746 -0.4221944
## Solar radiation (x2) -2.46349386            0.0000000  3.63970915 -1.2508892
## CO (x3)              -0.07044746            3.6397092  0.00000000  0.1914649
## NO (x4)              -0.42219439           -1.2508892  0.19146495  0.0000000
## NO2 (x5)             -0.38685926            6.6696447  1.67531156  0.5284662
## O3 (x6)              -1.83329179           30.2212114  2.39841560 -0.7554919
## HC (x7)               0.22866952            0.6949832 -0.26670989 -0.2907905
##                        NO2 (x5)    O3 (x6)    HC (x7)
## Wind (x1)            -0.3868593 -1.8332918  0.2286695
## Solar radiation (x2)  6.6696447 30.2212114  0.6949832
## CO (x3)               1.6753116  2.3984156 -0.2667099
## NO (x4)               0.5284662 -0.7554919 -0.2907905
## NO2 (x5)              0.0000000  2.9058818  0.5871817
## O3 (x6)               2.9058818  0.0000000  0.5943629
## HC (x7)               0.5871817  0.5943629  0.0000000
```

```
Psi
```

```
##            Wind (x1) Solar radiation (x2)              CO (x3)
##            2.2614844          300.0325639            0.8123373
##              NO (x4)             NO2 (x5)              O3 (x6)
##            0.5872636           10.7286506           30.2868302
##              HC (x7)
##            0.1087371
```

d)

```
air.mle <- factanal(air, factors=2) #mle factor analysis
lam.mle <- air.mle$load #loadings
psi.mle <- air.mle$unique # psi's
resi.mle <- R-lam.mle%*%t(lam.mle)-diag(psi.mle) #residuals resi.mle
resi.mle
```

```
##                             Wind (x1) Solar radiation (x2)       CO (x3)
## Wind (x1)                3.500560e-06        -1.248810e-02  4.391254e-02
## Solar radiation (x2) -1.248810e-02        -2.645875e-07  1.461983e-02
## CO (x3)                4.391254e-02         1.461983e-02 -3.934246e-07
## NO (x4)               -1.854901e-01        -6.274426e-02  1.077987e-02
## NO2 (x5)               3.410222e-02         3.445287e-02  1.762457e-02
## O3 (x6)               -9.004738e-04        -2.838523e-04  1.028851e-04
## HC (x7)                2.370172e-01        -8.669908e-03 -9.158393e-02
##                             NO (x4)      NO2 (x5)       O3 (x6)       HC (x7)
## Wind (x1)             -1.854901e-01  3.410222e-02 -9.004738e-04  0.2370172209
## Solar radiation (x2) -6.274426e-02  3.445287e-02 -2.838523e-04 -0.0086699079
## CO (x3)                1.077987e-02  1.762457e-02  1.028851e-04 -0.0915839333
## NO (x4)                8.726995e-07 -9.644292e-02 -1.903072e-04  0.0834567416
## NO2 (x5)              -9.644292e-02  9.544024e-07 -5.462750e-04  0.2742092840
## O3 (x6)               -1.903072e-04 -5.462750e-04 -2.081216e-06  0.0009741271
## HC (x7)                8.345674e-02  2.742093e-01  9.741271e-04 -0.0000016162
```

```r
psi.mle
```

```
##            Wind (x1) Solar radiation (x2)              CO (x3)
##            0.9070224            0.8953343            0.2126417
##              NO (x4)             NO2 (x5)              O3 (x6)
##            0.4983564            0.6144170            0.0050000
##              HC (x7)
##            0.9152467
```

PCA is only a linear transformation of data. No assumptions are made about the form of covariance matrix. In contrast, FA considers transformations from underlying unobservable factors to observed data.

e)

- Weighted least squares

```r
air.faw <- factanal(air, factors=2, scores="Bartlett")
air.faw$scores
```

```
##            Factor1     Factor2
##  [1,]  1.66971580 -0.28817832
##  [2,]  0.08581892 -0.79442341
##  [3,] -0.22848152 -0.60641538
##  [4,] -0.19197249  1.01277795
##  [5,] -0.53656357  0.12321642
##  [6,]  0.28944848  0.46211309
##  [7,]  2.04764802  0.95233044
##  [8,]  2.03002937  0.77057097
##  [9,] -0.06270817  0.29465612
## [10,]  0.51200591 -0.08552353
## [11,]  1.20823548 -1.18728795
## [12,] -0.10784582 -0.42760989
## [13,]  2.59309173  0.03792662
## [14,] -0.28714348 -0.42755154
## [15,] -0.93387229  0.13261646
```

```
## [16,] -0.84169640  0.13125206
## [17,] -0.76792099 -0.40902234
## [18,]  1.28121896 -1.00744595
## [19,]  0.15565889 -1.33757046
## [20,] -0.64541236 -0.78282754
## [21,]  1.09748771 -1.00463530
## [22,] -0.44516166 -0.60023517
## [23,] -0.25115382  0.29298022
## [24,] -0.64470830 -1.32267007
## [25,] -0.71565859  2.46973364
## [26,] -1.07810156 -0.58586019
## [27,] -0.41305379  0.29808060
## [28,] -2.25378622  0.16436643
## [29,] -0.09172319 -0.25437072
## [30,] -1.20752420 -1.30373565
## [31,] -0.58099722 -0.41474895
## [32,] -0.41338495 -0.24032574
## [33,]  1.97334337 -1.03003478
## [34,] -1.52069818  2.66674917
## [35,] -1.11050337 -0.04639066
## [36,] -0.89643781  0.13362696
## [37,] -1.78173876  0.51621727
## [38,]  1.21830756  1.52151855
## [39,]  1.15061034  2.77313381
## [40,]  0.23480991 -0.61572849
## [41,]  0.71776938  0.81159180
## [42,] -0.25695112 -0.79286656
```

- Regression approach

```
fit <- factanal(air, 2, rotaion="varimax", scores = "regression")
fit$loadings
```

```
##
## Loadings:
##                      Factor1 Factor2
## Wind (x1)            -0.176  -0.249
## Solar radiation (x2)          0.319
## CO (x3)               0.797   0.391
## NO (x4)               0.692  -0.152
## NO2 (x5)              0.602   0.152
## O3 (x6)                       0.997
## HC (x7)               0.251   0.147
##
##                Factor1 Factor2
## SS loadings      1.573   1.379
## Proportion Var   0.225   0.197
## Cumulative Var   0.225   0.422
```

```
fit$scores
```

```
##            Factor1     Factor2
```

```
##  [1,]  1.36720562 -0.27657826
##  [2,]  0.06555171 -0.78978747
##  [3,] -0.19099579 -0.60465611
##  [4,] -0.15126214  1.00636800
##  [5,] -0.43916657  0.11933013
##  [6,]  0.24010732  0.46147058
##  [7,]  1.68456974  0.95979015
##  [8,]  1.66902454  0.77886587
##  [9,] -0.04962899  0.29274972
## [10,]  0.41926064 -0.08198119
## [11,]  0.98340941 -1.17382229
## [12,] -0.09100788 -0.42604693
## [13,]  2.12622781  0.05342610
## [14,] -0.23800832 -0.42707417
## [15,] -0.76485119  0.12627654
## [16,] -0.68928720  0.12547715
## [17,] -0.63207123 -0.41155113
## [18,]  1.04433497 -0.99447041
## [19,]  0.11952373 -1.32969733
## [20,] -0.53389272 -0.78267783
## [21,]  0.89371624 -0.99278646
## [22,] -0.36860794 -0.59981948
## [23,] -0.20414009  0.28994184
## [24,] -0.53658315 -1.31971878
## [25,] -0.57179773  2.45260428
## [26,] -0.88744939 -0.58935018
## [27,] -0.33684615  0.29403581
## [28,] -1.84681690  0.14987254
## [29,] -0.07674080 -0.25360785
## [30,] -0.99790444 -1.30428921
## [31,] -0.47885269 -0.41611662
## [32,] -0.34037664 -0.24158266
## [33,]  1.61165031 -1.01275276
## [34,] -1.23063303  2.64372602
## [35,] -0.91074926 -0.05287217
## [36,] -0.73415366  0.12750839
## [37,] -1.45766957  0.50275749
## [38,]  1.00806364  1.52100889
## [39,]  0.96013676  2.76572885
## [40,]  0.18878660 -0.61111665
## [41,]  0.59339019  0.81173102
## [42,] -0.21546576 -0.79031348
```

Question 3

a)

- Correlation matrix S and the sample principal components.

```
stock <- read_excel("stock.xls")
S <- cor(stock)
S
```

```
##                   Allied Chemical   Du Pont Union Carbide     Exxon    Texaco
## Allied Chemical       1.0000000 0.5769244     0.5086555 0.3867206 0.4621781
## Du Pont               0.5769244 1.0000000     0.5983841 0.3895191 0.3219534
## Union Carbide         0.5086555 0.5983841     1.0000000 0.4361014 0.4256266
## Exxon                 0.3867206 0.3895191     0.4361014 1.0000000 0.5235293
## Texaco                0.4621781 0.3219534     0.4256266 0.5235293 1.0000000
```

```r
s.s <- eigen(S)
s.s$vectors
```

```
##            [,1]       [,2]       [,3]       [,4]       [,5]
## [1,] -0.4635405  0.2408499  0.6133570 -0.3813727 -0.4532876
## [2,] -0.4570764  0.5090997 -0.1778996 -0.2113068  0.6749814
## [3,] -0.4699804  0.2605774 -0.3370355  0.6640985 -0.3957247
## [4,] -0.4216770 -0.5252647 -0.5390181 -0.4728036 -0.1794482
## [5,] -0.4213291 -0.5822416  0.4336029  0.3812273  0.3874672
```

```r
s.s$values
```

```
## [1] 2.8564869 0.8091185 0.5400440 0.4513468 0.3430038
```

b)

- The proportion and the cumulative proportion of the total variance

```r
s.s$values/sum(s.s$values)
```

```
## [1] 0.57129738 0.16182370 0.10800880 0.09026936 0.06860076
```

```r
cumsum(s.s$values)/sum(s.s$values)
```

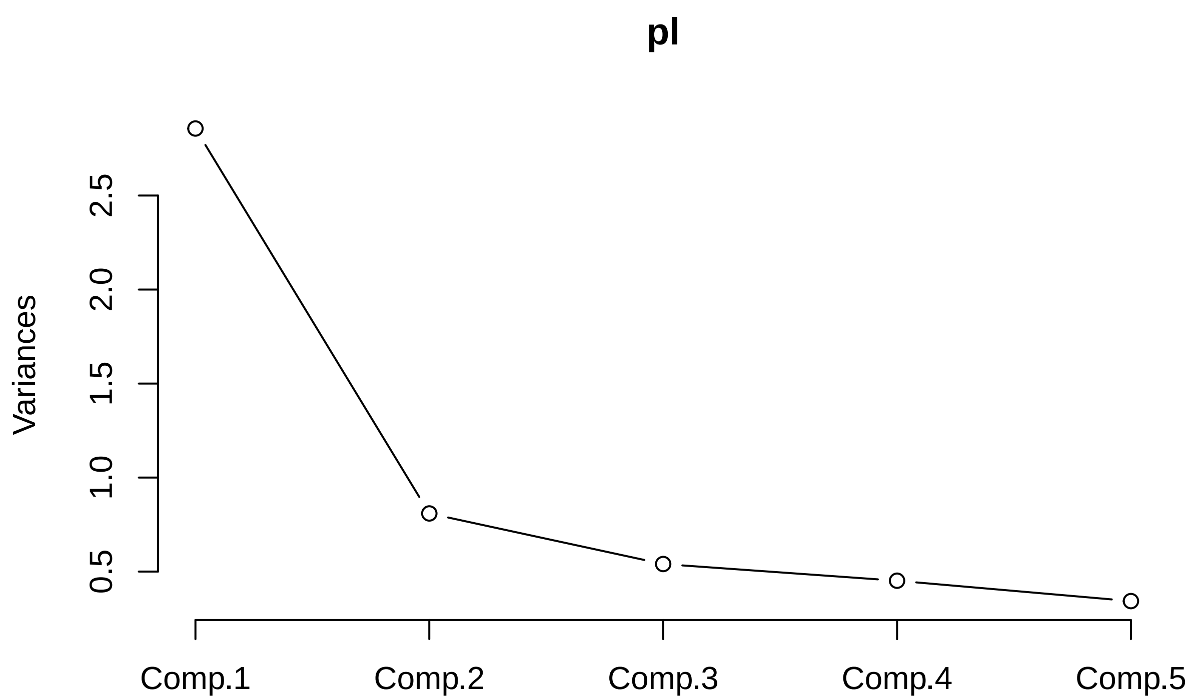
```
## [1] 0.5712974 0.7331211 0.8411299 0.9313992 1.0000000
```

c)

```r
pl=princomp(~stock$`Allied Chemical`+stock$`Du Pont`+stock$`Union Carbide`
               +stock$Exxon+stock$Texaco, cor=TRUE, data=stock)
pl$loadings
```

```
##
## Loadings:
##                           Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## stock$`Allied Chemical`  0.464  0.241  0.613  0.381  0.453
## stock$`Du Pont`          0.457  0.509 -0.178  0.211 -0.675
## stock$`Union Carbide`    0.470  0.261 -0.337 -0.664  0.396
## stock$Exxon              0.422 -0.525 -0.539  0.473  0.179
## stock$Texaco             0.421 -0.582  0.434 -0.381 -0.387
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings       1.0    1.0    1.0    1.0    1.0
## Proportion Var    0.2    0.2    0.2    0.2    0.2
## Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

```r
plot(pl, type="lines")
```



- Base on the graph, we should reducing dimension from 7 to 2.