APPENDIX B.

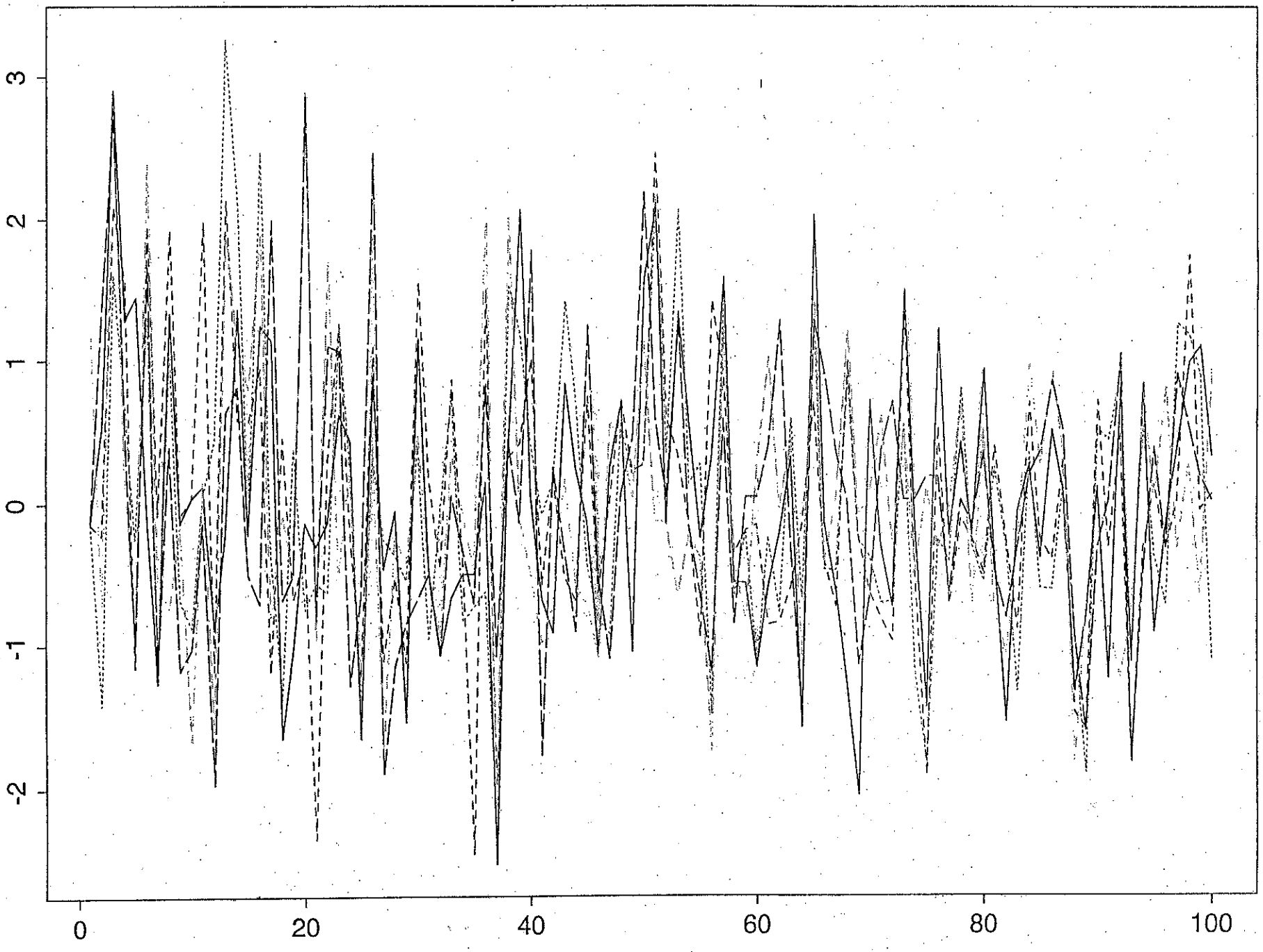
Example 8.1 STOCK PRICE DATA

TABLE 8.4 STOCK-PRICE DATA (WEEKLY RATE OF RETURN)

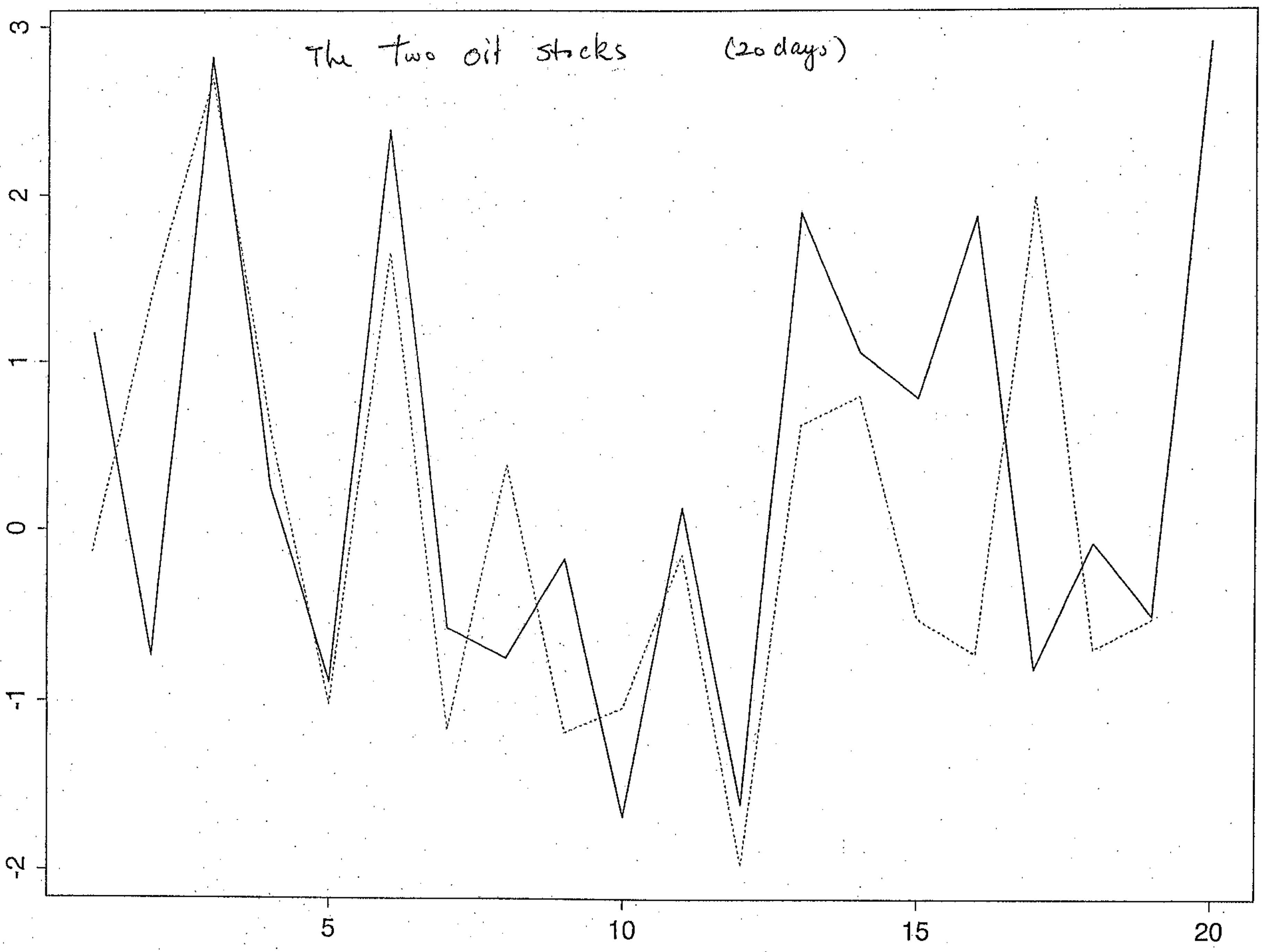
Week	Allied Chemical	Du Pont	Union Carbide	Exxon	Texaco
1	.000000	.000000	.000000	.039473	−.000000
2	.027027	−.044855	−.003030	−.014466	.043478
3	.122807	.060773	.088146	.086238	.078124
4	.057031	.029948	.066808	.013513	.019512
5	.063670	−.003793	−.039788	−.018644	−.024154
6	.003521	.050761	.082873	.074265	.049504
7	−.045614	−.033007	.002551	−.009646	−.028301
8	.058823	.041719	.081425	−.014610	.014563
9	.000000	−.019417	.002353	.001647	−.028708
10	.006944	−.025990	.007042	−.041118	−.024630
⋮	⋮	⋮	⋮	⋮	⋮
91	−.044068	.020704	−.006224	−.018518	.004694
92	.039007	.038540	.024988	−.028301	.032710
93	−.039457	−.029297	−.065844	−.015837	−.045758
94	.039568	.024145	−.006608	.028423	−.009661
95	−.031142	−.007941	.011080	.007537	.014634
96	.000000	−.020080	−.006579	.029925	−.004807
97	.021429	.049180	.006622	−.002421	.028985
98	.045454	.046375	.074561	.014563	.018779
99	.050167	.036380	.004082	−.011961	.009216
100	.019108	−.033303	.008362	.033898	.004566

...out the effects

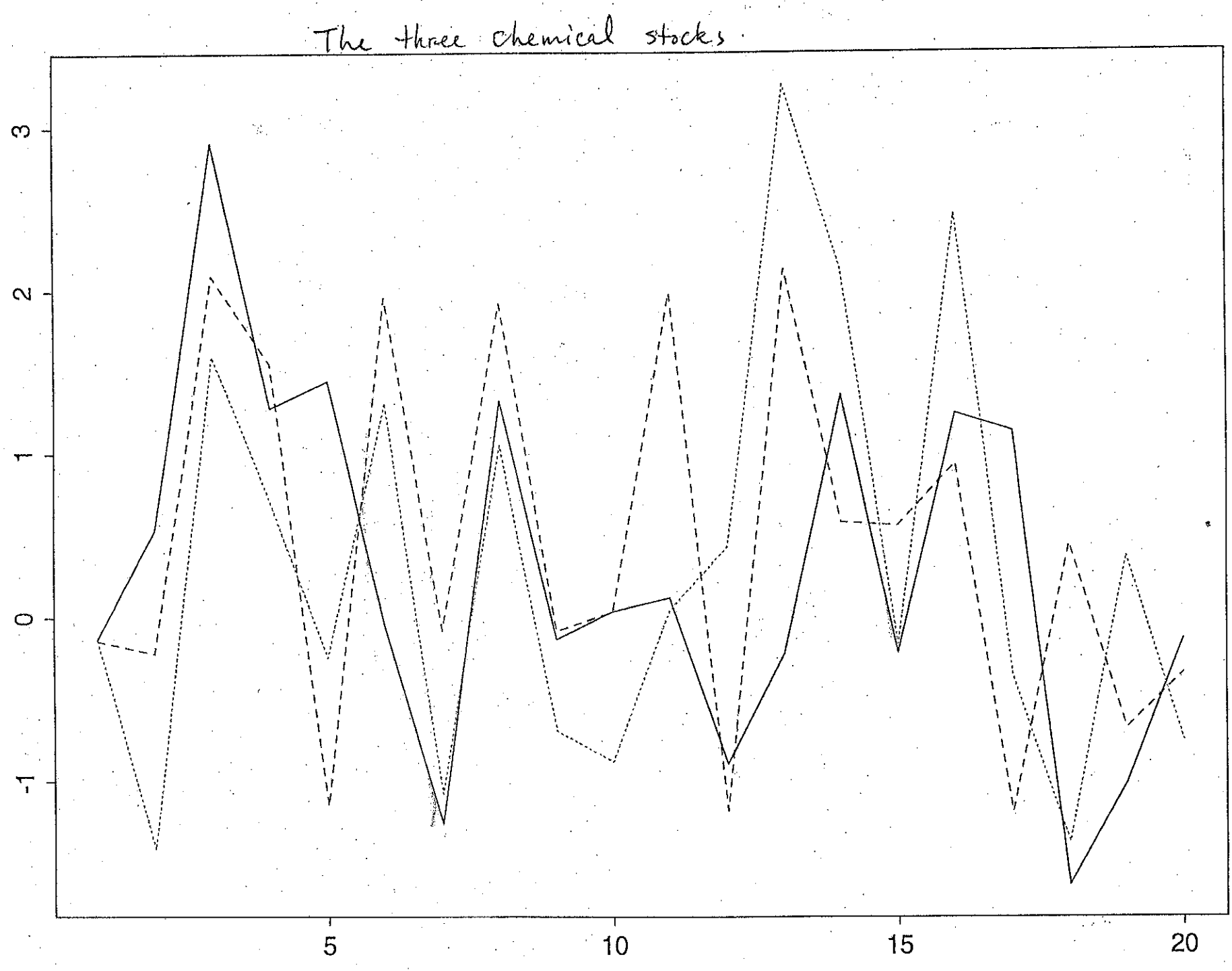
All five stocks.



B



B4



B5

means

[1] 0.00543372 0.00482707 0.00565418 0.00629143 0.00370861

Correlation Matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.5769244	0.5086555	0.3867206	0.4621751
[2,]	0.5769244	1.0000000	0.5983841	0.3895191	0.3219535
[3,]	0.5086555	0.5983841	1.0000000	0.4361014	0.4256269
[4,]	0.3867206	0.3895191	0.4361014	1.0000000	0.5235278
[5,]	0.4621751	0.3219535	0.4256269	0.5235278	1.0000000

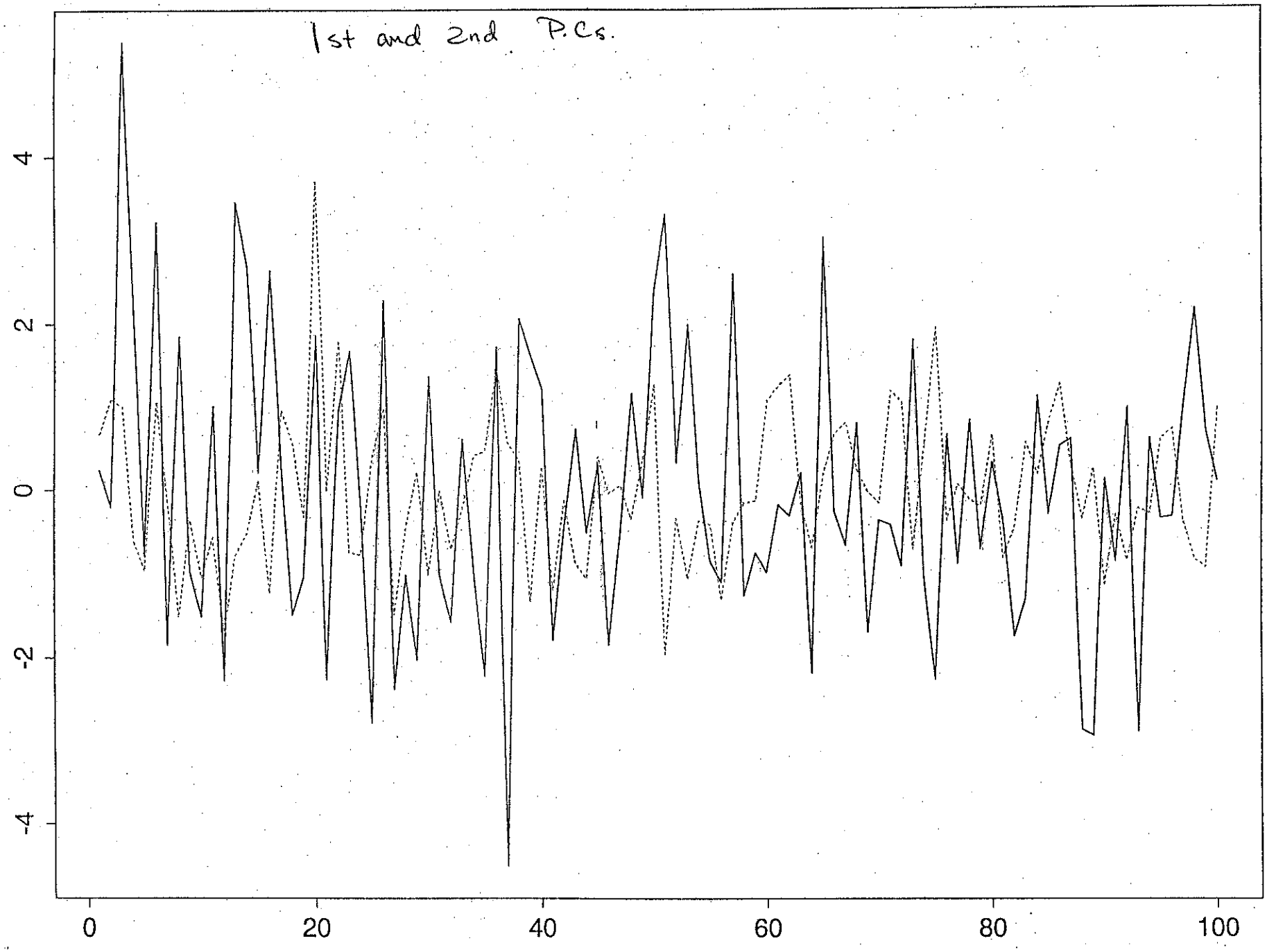
Eigen-values and eigen-vectors:

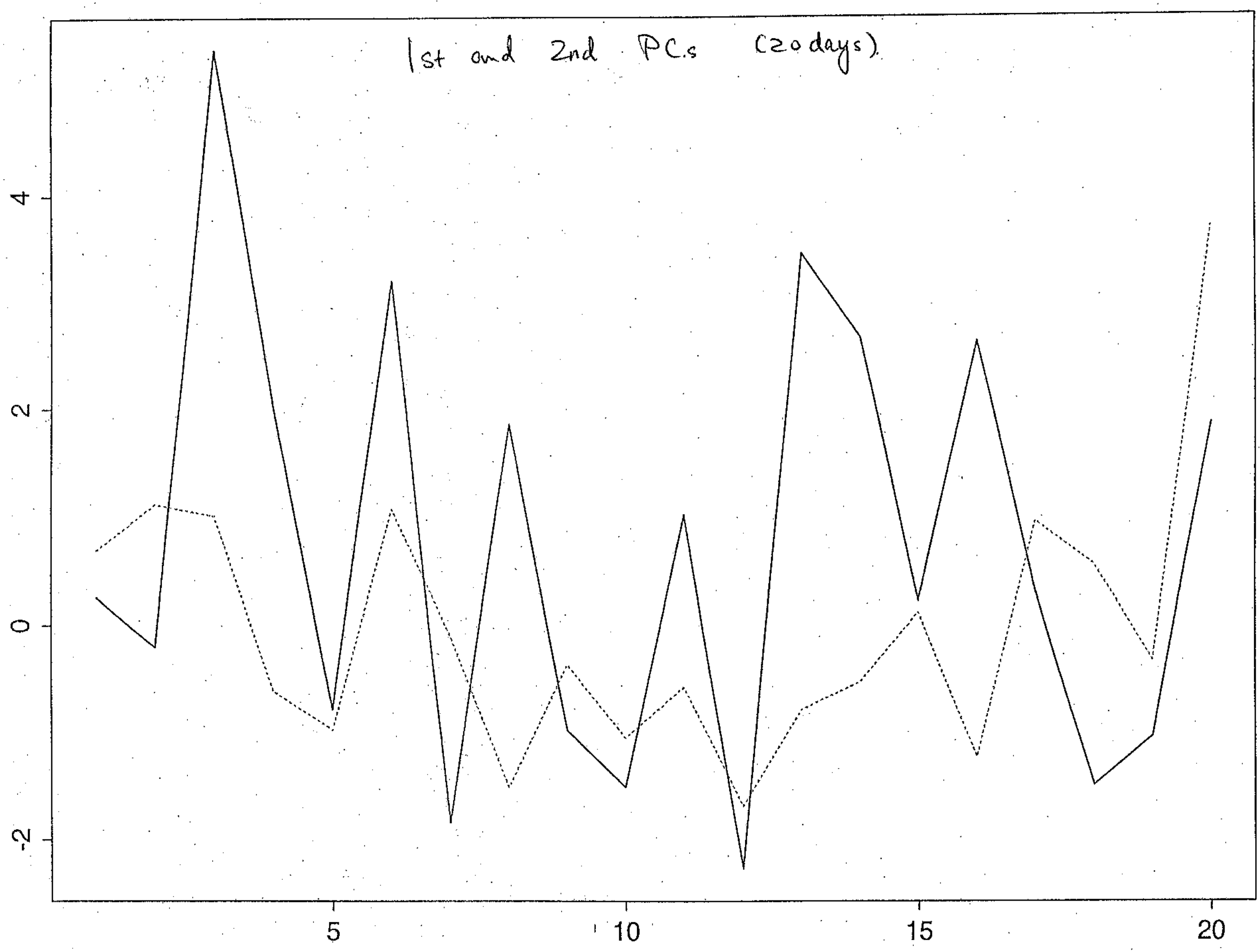
$values:
[1] 2.8564853 0.8091183 0.5400430 0.4513483 0.3430051

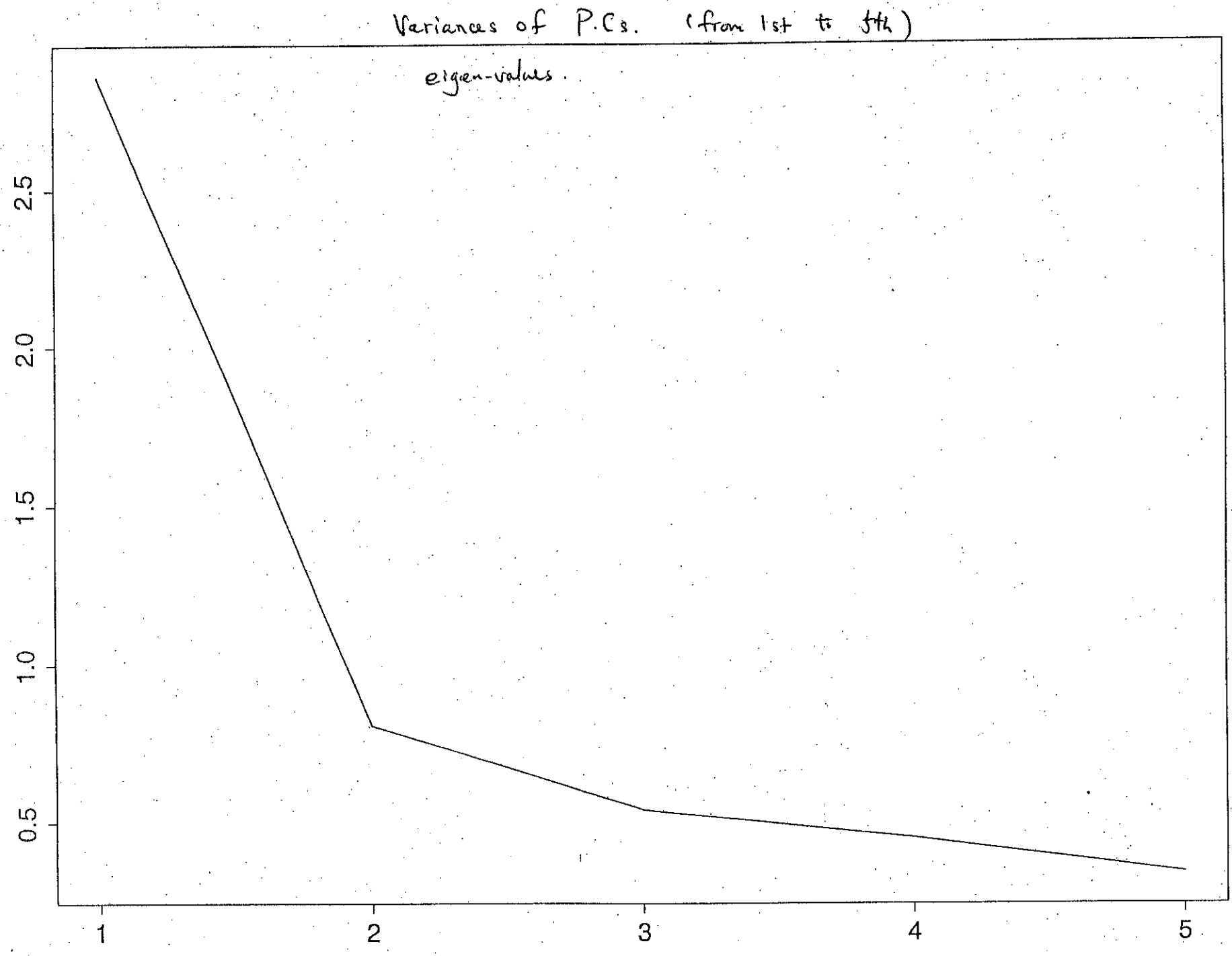
$vectors:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0.4635403	-0.2408533	0.6133541	-0.3813807	-0.4532830
[2,]	0.4570767	-0.5090983	-0.1779009	-0.2112957	0.6749853
[3,]	0.4699808	-0.2605749	-0.3370334	0.6640958	-0.3957323
[4,]	0.4216770	0.5252647	-0.5390189	-0.4728050	-0.1794424
[5,]	0.4213286	0.5822425	0.4336071	0.3812284	0.3874606

B6




1st and 2nd PCs (20 days)
4
2
0
-2
5
10
15
20


Variances of P.Cs. (from 1st to 5th)
eigen-values.
2.5
2.0
1.5
1.0
0.5
1
2
3
4
5

Example 8.2 TESTSCORES DATA

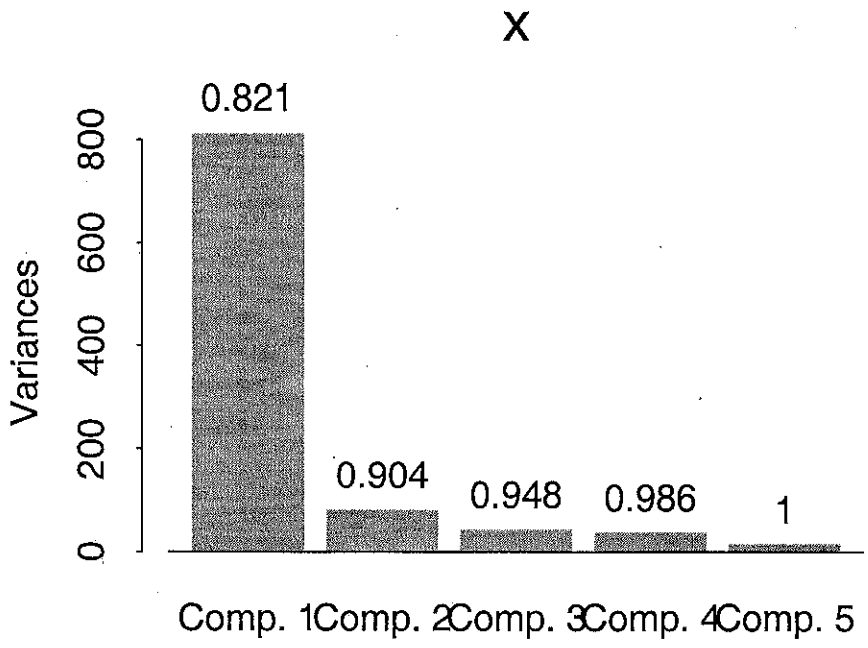
```
> ##
> ##
> ## testscores is a data set stored in Splus containing
> ## scores of 2 closed book exams (diffgeom & complex)
> ## and three open book exams (algebra, reals, statstics)
> ## for 25 students.
> ##
>
> #1. The data
>
> testscores
   diffgeom complex algebra reals statistics
 1       36      58      43    36         37
 2       62      54      50    46         52
 3       31      42      41    40         29
 4       76      78      69    66         81
 5       46      56      52    56         40
 6       12      42      38    38         28
 7       39      46      51    54         41
 8       30      51      54    52         32
 9       22      32      43    28         22
10        9      40      47    30         24
11       32      49      54    37         52
12       40      62      51    40         49
13       64      75      70    66         63
14       36      38      58    62         62
15       24      46      44    55         49
16       50      50      54    52         51
17       42      42      52    38         50
18        2      35      32    22         16
19       56      53      42    40         32
20       59      72      70    66         62
21       28      50      50    42         63
22       19      46      49    40         30
23       36      56      56    54         52
24       54      57      59    62         58
25       14      35      38    29         20
> dimnames(testscores)
[[1]]:
 [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14" "15"
[16] "16" "17" "18" "19" "20" "21" "22" "23" "24" "25"

[[2]]:
[1] "diffgeom"   "complex"    "algebra"    "reals"      "statistics"

>
> # 2. the main function of "princomp".
>
> testprc <- princomp(testscores)
> testprc
Standard deviations:
  Comp. 1  Comp. 2  Comp. 3  Comp. 4  Comp. 5
 28.48968 9.035471 6.600955 6.133582 3.723358

The number of variables is 5 and the number of observations is 25

Component names:

 "sdev" "loadings" "correlations" "scores" "center" "scale" "n.obs" "call"


 "factor.sdev" "coef"

Call:
princomp(x = testscores)
>
```

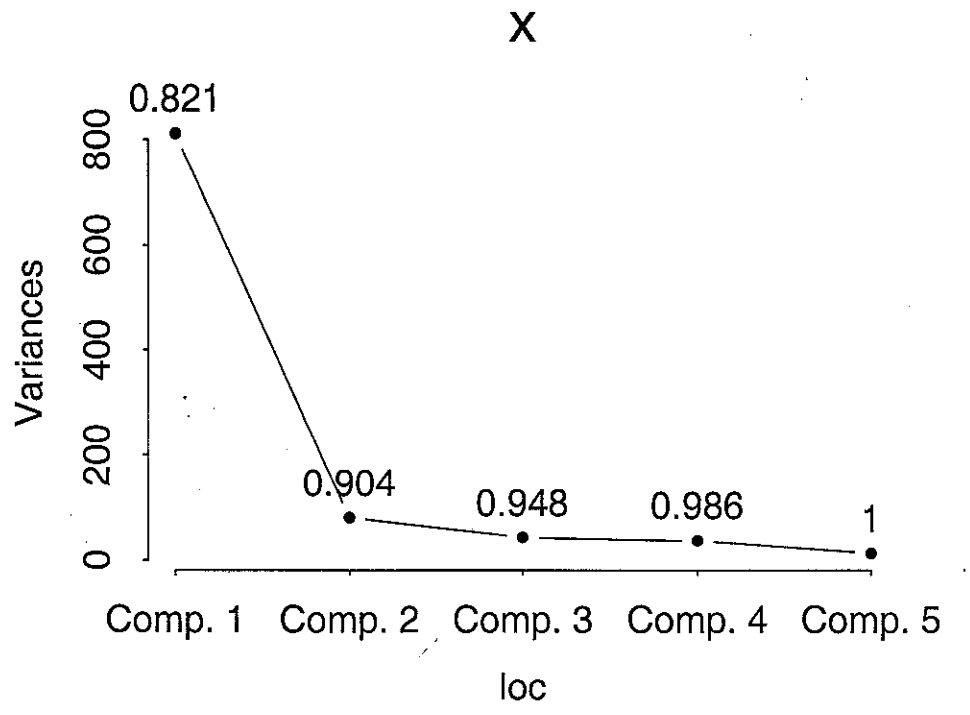
```
> summary(testprc)
Importance of components:
                          Comp. 1    Comp. 2    Comp. 3    Comp. 4    Comp. 5
    Standard deviation 28.4896795 9.03547104 6.60095491 6.13358179 3.72335754
Proportion of Variance  0.8212222 0.08260135 0.04408584 0.03806395 0.01402668
 Cumulative Proportion  0.8212222 0.90382353 0.94790936 0.98597332 1.00000000
>
> ##viewing also the loadings
>
> summary(testprc, loadings=T)
Importance of components:
                          Comp. 1    Comp. 2    Comp. 3    Comp. 4    Comp. 5
    Standard deviation 28.4896795 9.03547104 6.60095491 6.13358179 3.72335754
Proportion of Variance  0.8212222 0.08260135 0.04408584 0.03806395 0.01402668
 Cumulative Proportion  0.8212222 0.90382353 0.94790936 0.98597332 1.00000000

Loadings:
           Comp. 1 Comp. 2 Comp. 3 Comp. 4 Comp. 5
  diffgeom  0.598  -0.675  -0.185  -0.386
   complex  0.361  -0.245   0.249   0.829  -0.247
   algebra  0.302   0.214   0.211   0.135   0.894
     reals  0.389   0.338   0.700  -0.375  -0.321
statistics  0.519   0.570  -0.607          -0.179
>
> plot(loadings(testprc))
>
>
> ## 3. perform principal component analysis on standardized variables,
> ## that is to say, on the correlations.
>
> testprc1 <- princomp(testscores, cor=T)
> testprc1
Standard deviations:
  Comp. 1   Comp. 2  Comp. 3   Comp. 4   Comp. 5
 2.020188 0.6114409 0.465352 0.4525298 0.3516318

The number of variables is 5 and the number of observations is 25

Component names:

 "sdev" "loadings" "correlations" "scores" "center" "scale" "n.obs" "call"


 "factor.sdev" "coef"

Call:
princomp(x = testscores, cor = T)
>
> summary(testprc1)
Importance of components:
                          Comp. 1    Comp. 2   Comp. 3    Comp. 4    Comp. 5
    Standard deviation 2.0201880 0.61144085 0.4653520 0.45252975 0.35163180
Proportion of Variance 0.8162319 0.07477198 0.0433105 0.04095664 0.02472898
 Cumulative Proportion 0.8162319 0.89100388 0.9343144 0.97527102 1.00000000
>
> summary(testprc1, loadings=T)
Importance of components:
                          Comp. 1    Comp. 2   Comp. 3    Comp. 4    Comp. 5
    Standard deviation 2.0201880 0.61144085 0.4653520 0.45252975 0.35163180
Proportion of Variance 0.8162319 0.07477198 0.0433105 0.04095664 0.02472898
 Cumulative Proportion 0.8162319 0.89100388 0.9343144 0.97527102 1.00000000

Loadings:
           Comp. 1 Comp. 2 Comp. 3 Comp. 4 Comp. 5
  diffgeom  0.443  -0.473   0.352   0.615  -0.279
   complex  0.437  -0.616  -0.368  -0.424   0.339
```

```
   algebra  0.462   0.276  -0.210  -0.338  -0.743
     reals  0.445   0.470  -0.470   0.476   0.367
statistics  0.449   0.317   0.690  -0.318   0.347
>
> # the results are similar.
>
> ## 4. Using covariance or corrleation matrix instead of the
> ## raw data to perform principal component analysis.
>
>  covtest <- var(testscores)
> covtest
           diffgeom   complex   algebra    reals statistics
  diffgeom 348.7733 181.69167 137.54500 176.8850   233.6583
   complex 181.6917 145.75000  91.28333 108.4750   142.5000
   algebra 137.5450  91.28333  95.39333 106.1383   135.1833
     reals 176.8850 108.47500 106.13833 166.9567   167.5500
statistics 233.6583 142.50000 135.18333 167.5500   272.6667
>
> princomp(covtest)
Standard deviations:
  Comp. 1  Comp. 2  Comp. 3  Comp. 4 Comp. 5
 94.32082 35.68043 18.58679 9.669862       0

The number of variables is 5 and the number of observations is 5

Component names:

 "sdev" "loadings" "correlations" "scores" "center" "scale" "n.obs" "call"


 "factor.sdev" "coef"

Call:
princomp(x = covtest)
>
> covtest <- list(cov=covtest, center=rep(0, 5))
> covtest
$cov:
           diffgeom   complex   algebra    reals statistics
  diffgeom 348.7733 181.69167 137.54500 176.8850   233.6583
   complex 181.6917 145.75000  91.28333 108.4750   142.5000
   algebra 137.5450  91.28333  95.39333 106.1383   135.1833
     reals 176.8850 108.47500 106.13833 166.9567   167.5500
statistics 233.6583 142.50000 135.18333 167.5500   272.6667

$center:
[1] 0 0 0 0 0

> princomp(covlist=covtest)
Standard deviations:
  Comp. 1  Comp. 2  Comp. 3  Comp. 4  Comp. 5
 29.07716 9.221789 6.737071 6.260061 3.800136

The number of variables is 5 and the number of observations is unknown.

Component names:

 "sdev" "loadings" "correlations" "center" "scale" "call" "factor.sdev" "coef"

Call:
princomp(covlist = covtest)
>
> ##the above procedure also works for correlation matrix.
>
> ## 5. the Cutoff to retain-drop the PCs.
>
```

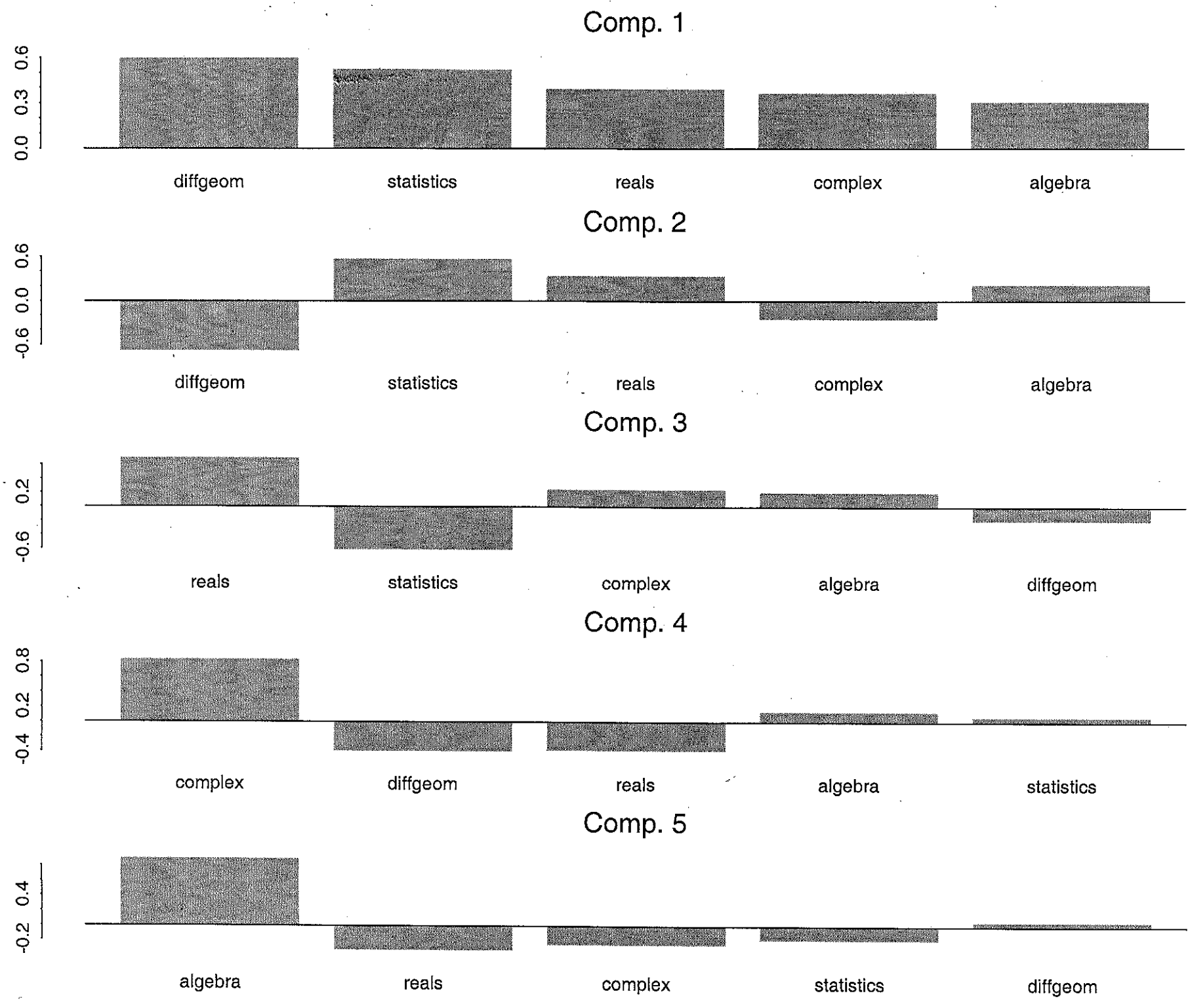
```
> # 1). screeplot method.
>
> plot(testprc)
[1] 0.7 1.9 3.1 4.3 5.5
> plot(testprc, style="lines")
[1] 1 2 3 4 5
> # From the picture, we keep PC 1 and 2.
>
> # 2). Keeping enough p-c to explain 90% of the variances.
>
> testprc$sdev^2
  Comp. 1  Comp. 2  Comp. 3  Comp. 4  Comp. 5
 811.6618 81.63974 43.57261 37.62083 13.86339
> testprc$sdev^2/sum(testprc$sdev^2)
    Comp. 1    Comp. 2    Comp. 3    Comp. 4    Comp. 5
 0.8212222 0.08260135 0.04408584 0.03806395 0.01402668
> #this method keeps PC 1 and 2.
>
> # 3). Keep the pcs that with variance above the mean variance.
> mean(testprc$sdev^2)
[1] 197.6717
> # this method keeps PC 1.
>
>
```

B13


X
0.821
0.904
0.948
0.986
1
Variances
0
200
400
600
800
Comp. 1
Comp. 2
Comp. 3
Comp. 4
Comp. 5


X
0.821
0.904
0.948
0.986
1
Variances
0
200
400
600
800
Comp. 1
Comp. 2
Comp. 3
Comp. 4
Comp. 5
loc

B14


Comp. 1
0.0
0.3
0.6
diffgeom
statistics
reals
complex
algebra
Comp. 2
-0.6
0.0
0.6
diffgeom
statistics
reals
complex
algebra
Comp. 3
-0.6
0.2
reals
statistics
complex
algebra
diffgeom
Comp. 4
-0.4
0.2
0.8
complex
diffgeom
reals
algebra
statistics
Comp. 5
-0.2
0.4
algebra
reals
complex
statistics
diffgeom