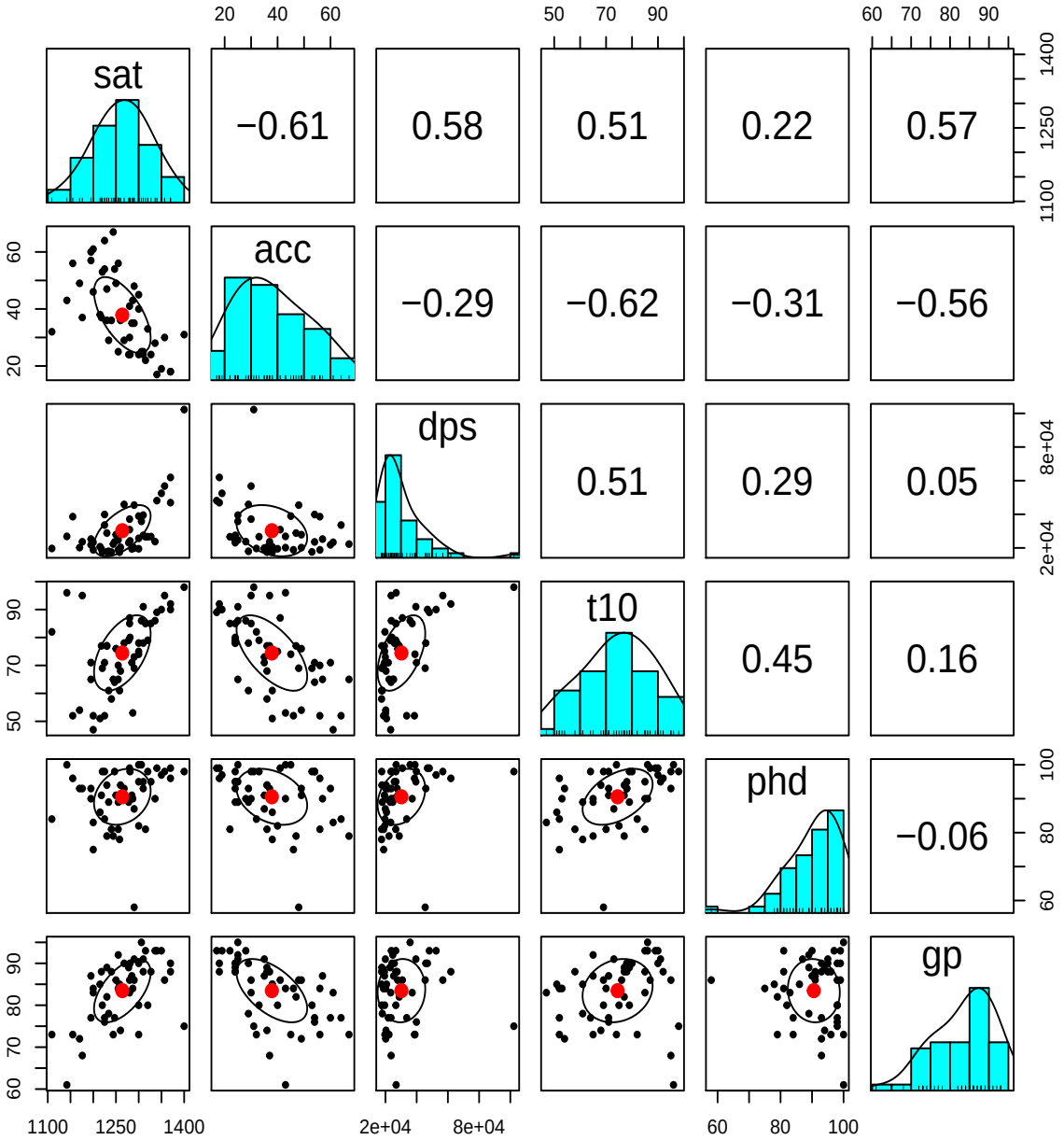
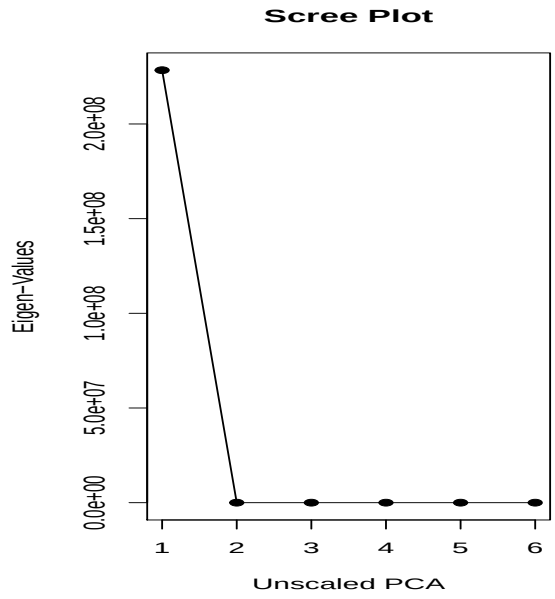
Illustration of Principal Component Analysis

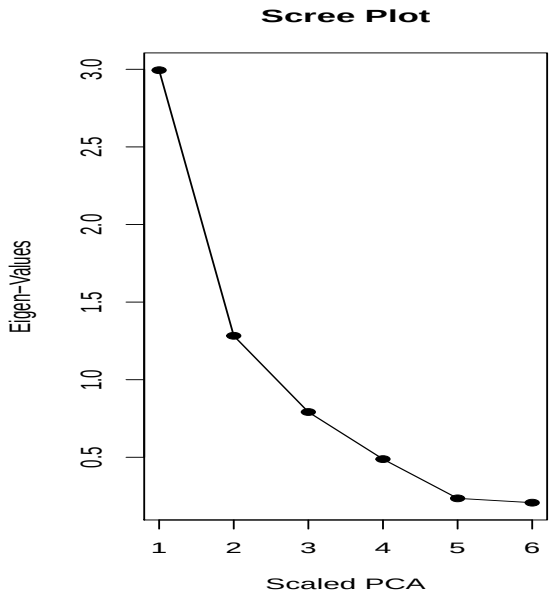
Example 1: The first example pertains to the same `colloeges.data`, we'd analysed for illustrating Hotelling's T^2. There we lacked proper visualisation tools. Here we use PCA for that purpose.

```
> d1<-read.table("colleges.data",header=T)
> X1<-matrix(c(d$sat,d$acc,d$dps,d$t10,d$phd,d$gp),ncol=6,dimnames=list(d$name,
+ c("sat","acc","dps","t10","phd", "gp")))

> library(psych)
> pairs.panels(X1,smooth=F)
```



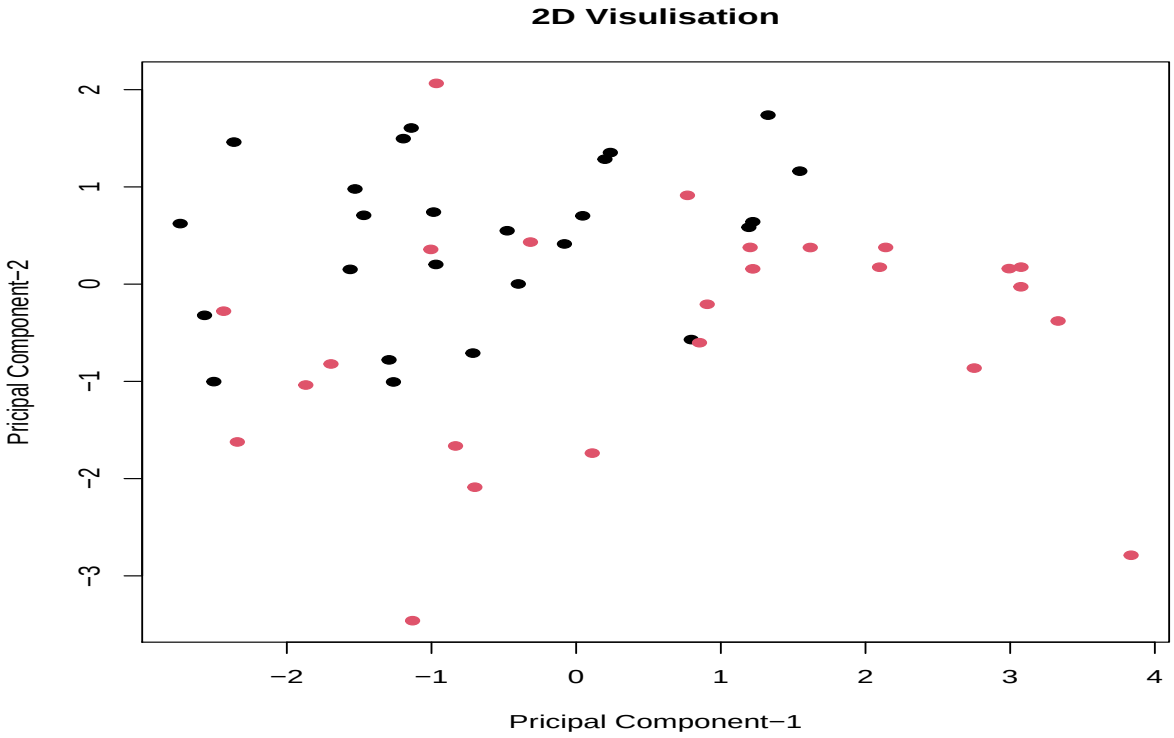
```
> pca1<-princomp(X1)
> summary(pca1)
Importance of components:
                             Comp.1      Comp.2       Comp.3       Comp.4
Standard deviation     1.511278e+04 51.20569316 1.343131e+01 8.273341e+00
Proportion of Variance 9.999871e-01  0.00001148 7.898453e-07 2.996866e-07
Cumulative Proportion  9.999871e-01  0.99999861 9.999994e-01 9.999997e-01
                             Comp.5       Comp.6
Standard deviation     6.849597e+00 4.615603e+00
Proportion of Variance 2.054166e-07 9.327436e-08
Cumulative Proportion  9.999999e-01 1.000000e+00

> pca1.cor<-princomp(X1,cor=T)
> summary(pca1.cor)
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     1.7303978 1.1326824 0.8897564 0.69872679 0.48512370
Proportion of Variance 0.4990461 0.2138282 0.1319444 0.08136986 0.03922417
Cumulative Proportion  0.4990461 0.7128743 0.8448188 0.92618861 0.96541278
                           Comp.6
Standard deviation     0.45554730
Proportion of Variance 0.03458722
Cumulative Proportion  1.00000000

> plot(1:6,pca1$sd^2,pch=19,xlab="Unscaled PCA",ylab="Eigen-Values",main="Scree Plot")
> lines(c(1:6),pca1$sd^2)
> plot(1:6,pca1.cor$sd^2,pch=19,xlab="Scaled PCA",ylab="Eigen-Values",main="Scree Plot")
> lines(c(1:6),pca1.cor$sd^2)
```



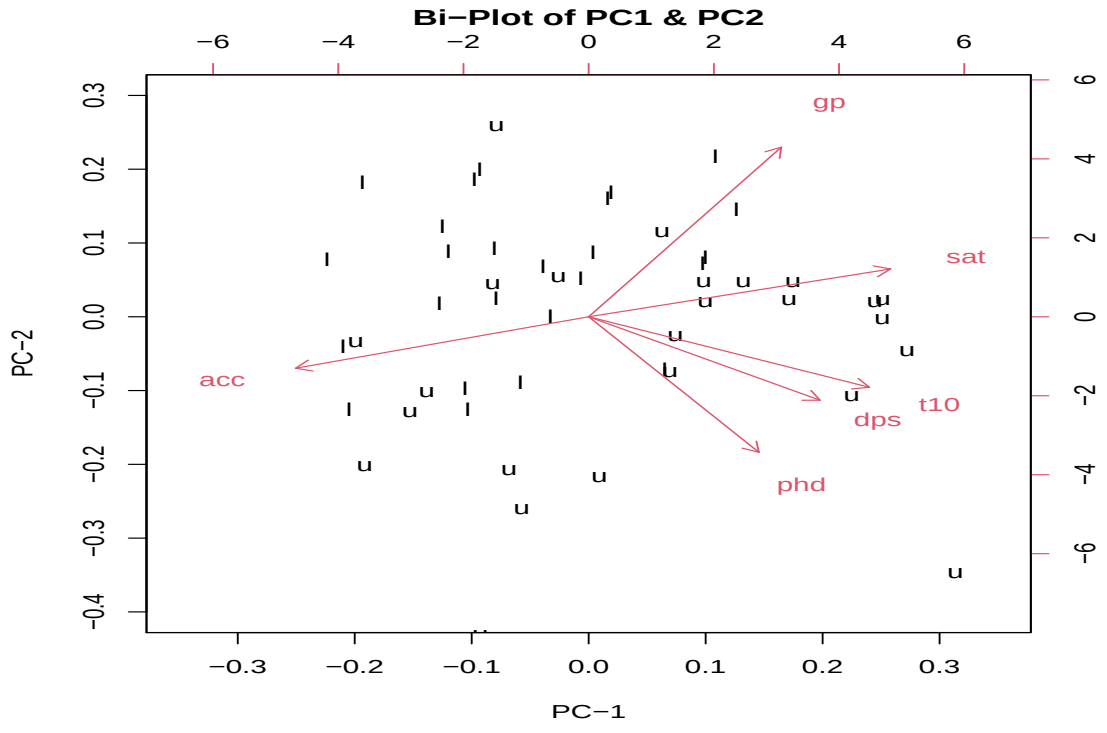



```
> pca1.cor$sd^2
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6
2.9942766 1.2829695 0.7916665 0.4882191 0.2353450 0.2075233
```

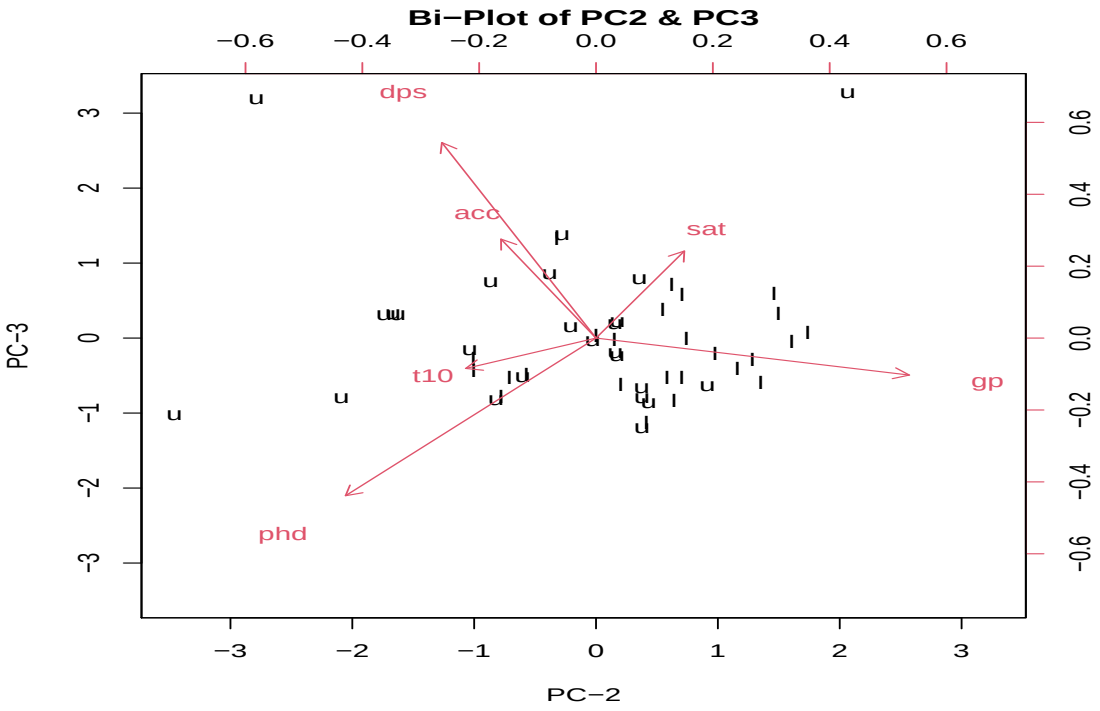
```
> plot(pca1.cor$scores[,1],pca1.cor$scores[,2],col=c(rep(1,25),rep(2,25)),
+ xlab="Pricipal Component-1",ylab="Pricipal Component-2",pch=19,main="2D Visulisation")
```



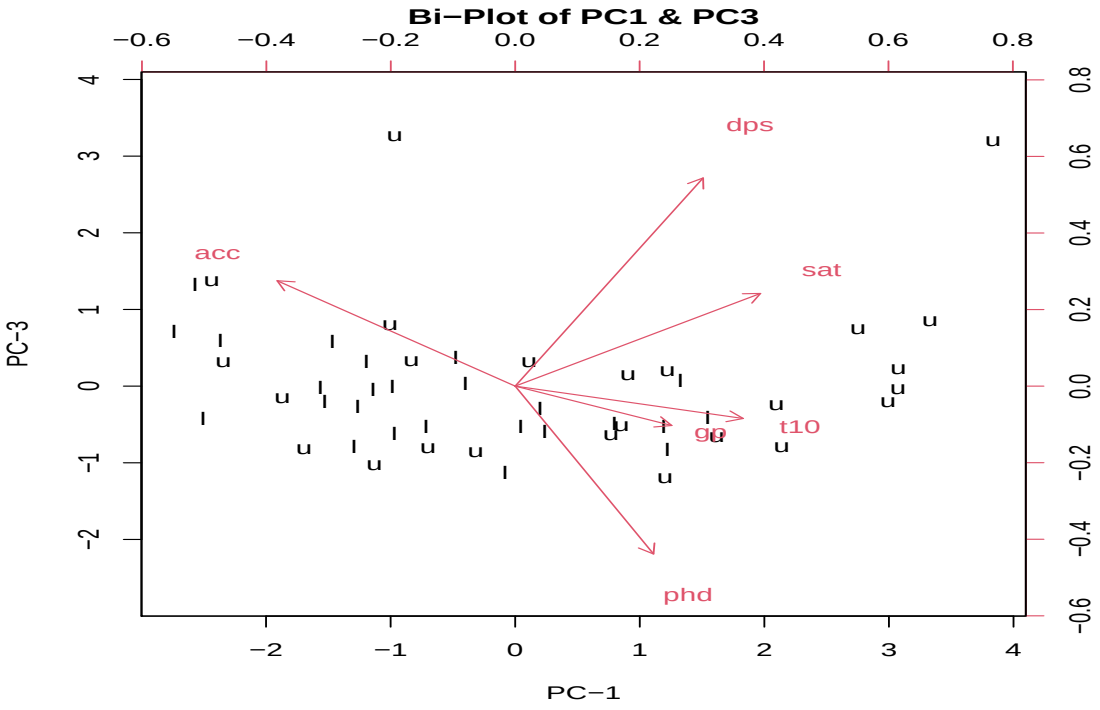

```
> biplot(pca1.cor,main="Bi-Plot of PC1 & PC2",xlabs=c(rep("l",25),rep("u",25)),
+ xlim=c(-0.35,0.35), ylim=c(-0.4,0.3),xlab="PC-1",ylab="PC-2",expand=1.4)
```



```
> biplot(matrix(pca1.cor$scores[,2:3],ncol=2),matrix(pca1.cor$loadings[,2:3],ncol=2),
+ xlabs=c(rep("l",25),rep("u",25)), ylabs=c("sat","acc","dps","t10","phd", "gp"),
+ xlab="PC-2",ylab="PC-3",main="Bi-Plot of PC2 & PC3")
```



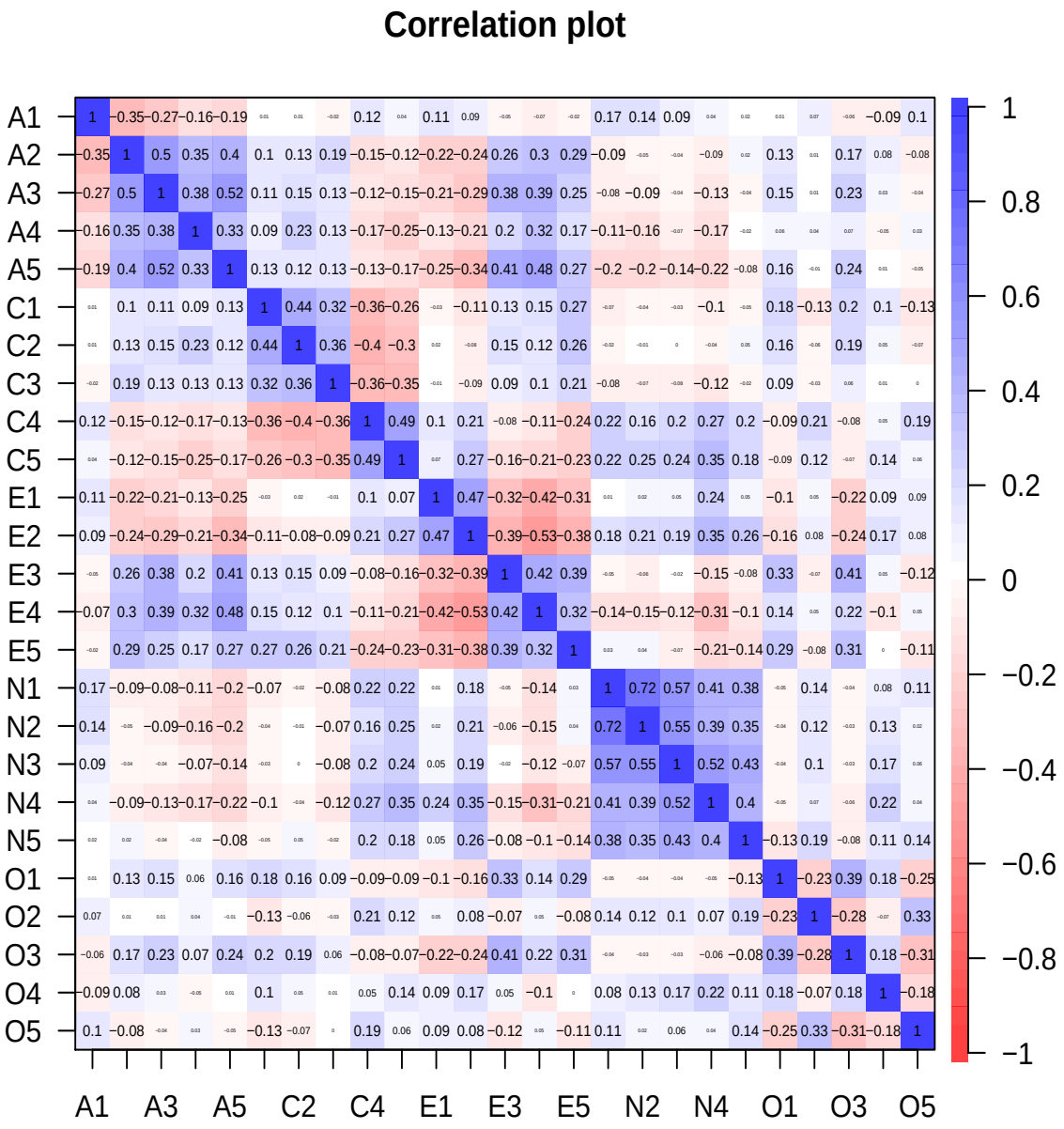

```
> biplot(matrix(pca1.cor$scores[,c(1,3)],ncol=2),matrix(pca1.cor$loadings[,c(1,3)],ncol=2),
+ xlabs=c(rep("l",25),rep("u",25)), ylabs=c("sat","acc","dps","t10","phd","gp"),
+ xlab="PC-1",ylab="PC-3",main="Bi-Plot of PC1 & PC3")
```



```
> pca1.cor$loadings

Loadings:
    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
sat  0.493  0.189  0.303  0.278  0.295  0.683
acc -0.478 -0.203  0.344  0.317  0.692 -0.179
dps  0.378 -0.330  0.680  0.137 -0.338 -0.392
t10  0.458 -0.279 -0.106 -0.625  0.527 -0.180
phd  0.278 -0.536 -0.548  0.576
gp   0.314  0.670 -0.129  0.285  0.203 -0.559

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

Example 2: The second example uses a data-set containing responses to 25 Personality Test Items taken by 2800 (probably neurotic) individuals. The data-set comes with the `R` package `psych` (more precisely the `psychTools` package, which anyway gets automatically installed when the `psych` package is installed with the `dependencies=T` option). The data-set is called `bfi` and can be loaded into the `R` work-space by issuing the `data(bfi)` command. However as such `bfi` is not clean (it contains lots of `NA`s hindering smooth analysis). After cleaning them (I had to do it by hand by writing a small piece of code, but I am not including it here to avoid digression) 2436 records are found with complete responses to all the 25 items. These are next copied into a 2436×25 raw data matrix called `X2`, which we analyse. Brief descriptions of the 25 personality items may be learnt by issuing the `help(bfi)` command. However a quick working description of the 25 features involving human personality traits is provided in `bfi.keys`, which is as follows:

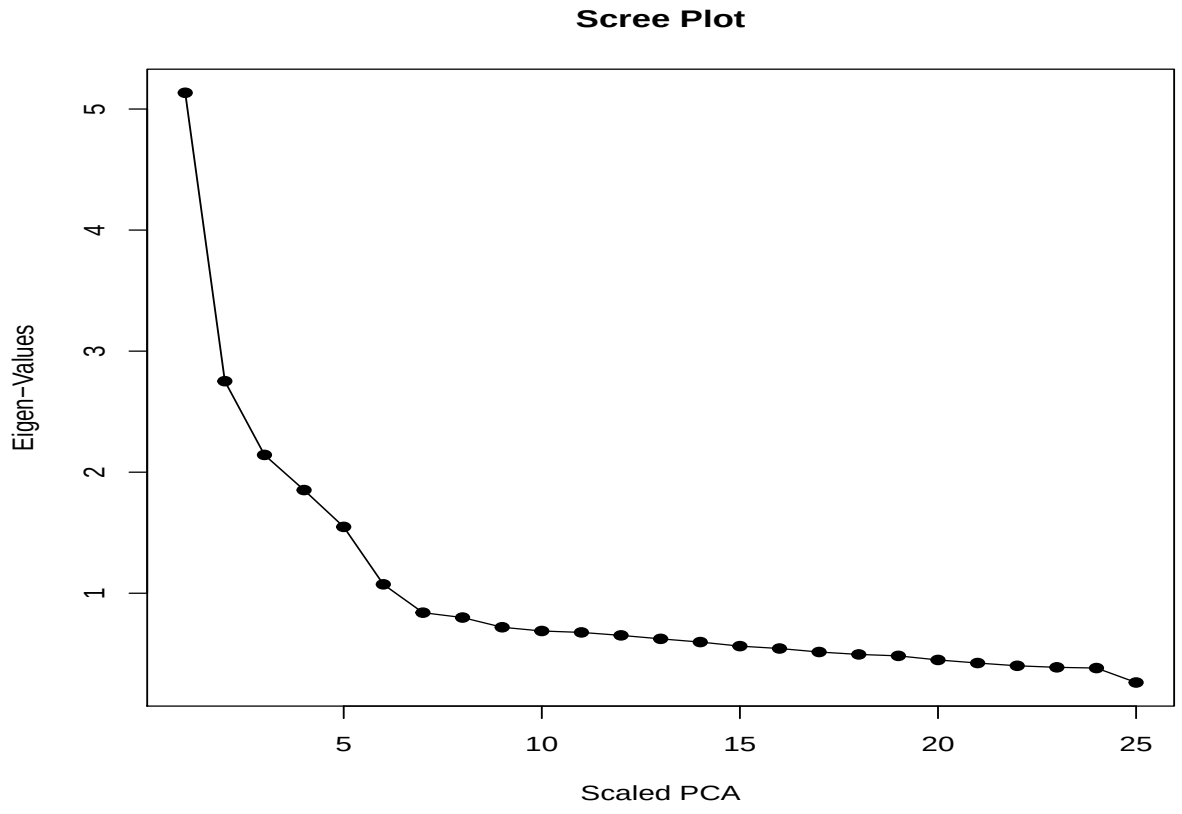
```
> bfi.keys
$agree
[1] "-A1" "A2"  "A3"  "A4"  "A5"

$conscientious
[1] "C1"  "C2"  "C3"  "-C4" "-C5"

$extraversion
[1] "-E1" "-E2" "E3"  "E4"  "E5"

$neuroticism
[1] "N1" "N2" "N3" "N4" "N5"

$openness
[1] "O1"  "-O2" "O3"  "O4"  "-O5"

> cor.plot(X2) #library(psych)
```



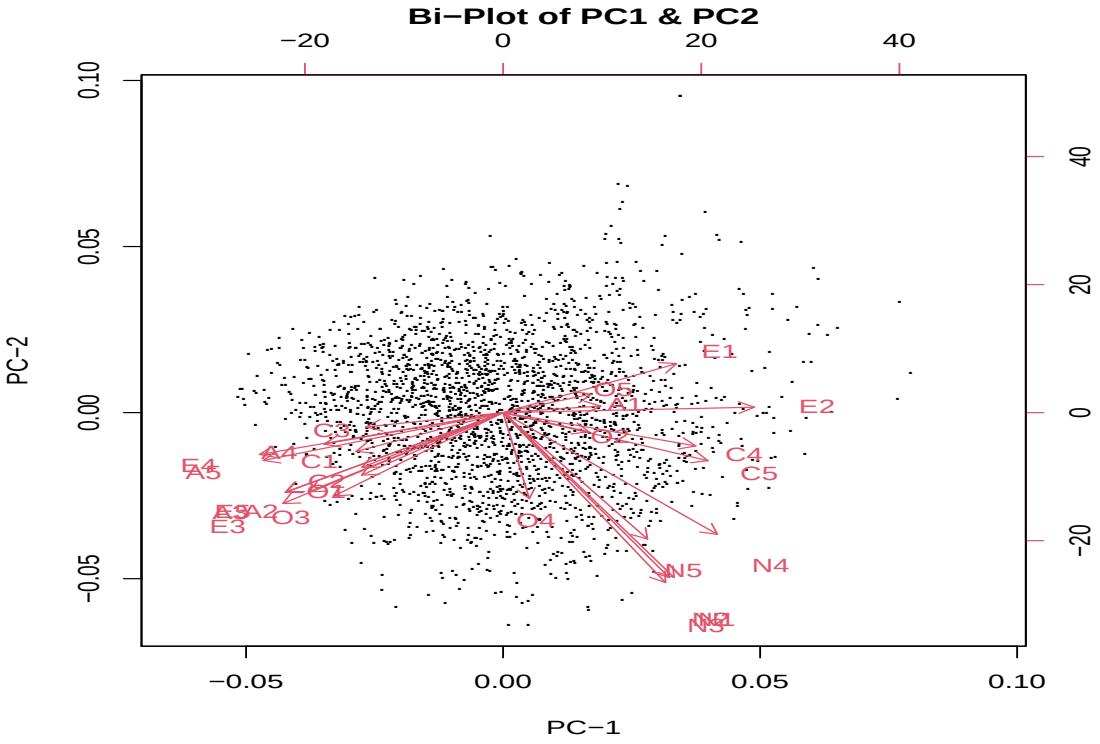

```
> pca2.cor<-princomp(X2,cor=T)

> summary(pca2.cor)
Importance of components:
                          Comp.1     Comp.2     Comp.3     Comp.4     Comp.5
Standard deviation     2.2659018 1.6588811 1.46379710 1.3610024 1.24425192
Proportion of Variance 0.2053724 0.1100755 0.08570808 0.0740931 0.06192651
Cumulative Proportion  0.2053724 0.3154479 0.40115599 0.4752491 0.53717561
                          Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
Standard deviation     1.0361382 0.91626357 0.89398332 0.84793230 0.82951117
Proportion of Variance 0.0429433 0.03358156 0.03196825 0.02875957 0.02752355
Cumulative Proportion  0.5801189 0.61370047 0.64566871 0.67442828 0.70195183
```

```
                          Comp.11    Comp.12    Comp.13    Comp.14    Comp.15
Standard deviation     0.82241921 0.80734122 0.78946371 0.77237481 0.75039378
Proportion of Variance 0.02705493 0.02607199 0.02493012 0.02386251 0.02252363
Cumulative Proportion  0.72900677 0.75507876 0.78000888 0.80387139 0.82639503
                          Comp.16   Comp.17    Comp.18    Comp.19    Comp.20
Standard deviation     0.73709248 0.7172988 0.70320918 0.69472262 0.67001567
Proportion of Variance 0.02173221 0.0205807 0.01978013 0.01930558 0.01795684
Cumulative Proportion  0.84812724 0.8687079 0.88848807 0.90779365 0.92575049
                          Comp.21    Comp.22    Comp.23    Comp.24    Comp.25
Standard deviation     0.65066590 0.63298614 0.62273950 0.61794562 0.51238561
Proportion of Variance 0.01693464 0.01602686 0.01551218 0.01527427 0.01050156
Cumulative Proportion  0.94268513 0.95871199 0.97422417 0.98949844 1.00000000
```

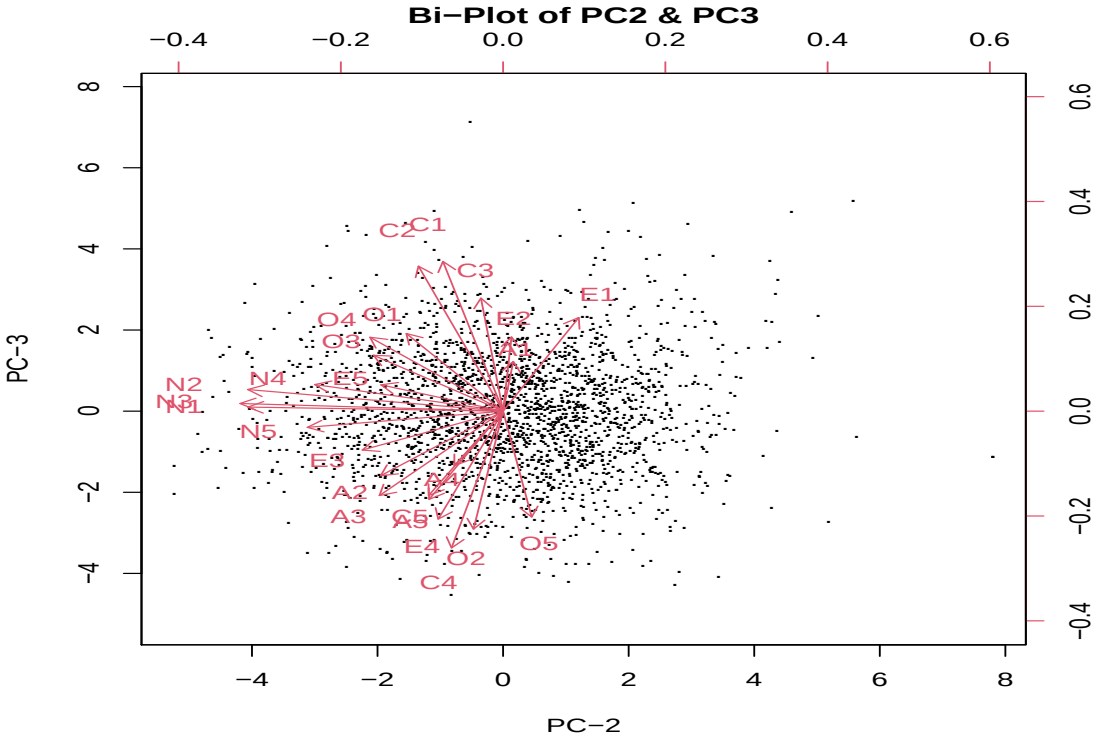
```
> plot(1:25,pca2.cor$sd^2,pch=19,xlab="Scaled PCA",ylab="Eigen-Values",
+ main="Scree Plot")
> lines(c(1:25),pca2.cor$sd^2)
```



```
> biplot(pca2.cor,xlabs=rep(".",2436),xlab="PC-1",ylab="PC-2")
```




```
> biplot(matrix(pca2.cor$scores[,2:3],ncol=2),matrix(pca2.cor$loadings[,2:3],
+ ncol=2),xlabs=rep(".",2436),ylabs=colnames(X2),xlab="PC-2",ylab="PC-3",
+ main="Bi-Plot of PC2 & PC3")
```



```
> pca2.cor$loadings

Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
A1  0.110         0.118         0.490  0.422  0.194         0.195  0.207
A2 -0.219 -0.189 -0.154        -0.333 -0.163 -0.159  0.242
A3 -0.246 -0.191 -0.201        -0.255         0.127  0.164        -0.119
A4 -0.202        -0.130  0.241 -0.189         0.503  0.119  0.395  0.114
A5 -0.270 -0.114 -0.210        -0.151  0.187
C1 -0.166         0.357  0.153         0.125 -0.204 -0.330  0.208 -0.426
C2 -0.159 -0.130  0.346  0.281         0.148  0.101 -0.168        -0.258
C3 -0.155         0.269  0.324                -0.277  0.118 -0.461  0.311
C4  0.217        -0.326 -0.205         0.315
C5  0.231 -0.115 -0.201 -0.258         0.113 -0.210 -0.104        -0.410
E1  0.196  0.117  0.223  0.115 -0.270  0.343  0.184  0.327        -0.141
E2  0.283         0.177        -0.280  0.166         0.108
E3 -0.248 -0.217        -0.152  0.125  0.221                -0.166  0.138
E4 -0.275        -0.257         0.175  0.112        -0.267
E5 -0.244 -0.188                0.247        -0.256  0.344        -0.212
N1  0.193 -0.393                0.227 -0.180         0.215
N2  0.187 -0.393                0.184 -0.229         0.214  0.143
N3  0.183 -0.405                                     0.148
N4  0.242 -0.290                -0.172         0.105        -0.180 -0.106
N5  0.163 -0.301         0.199 -0.110         0.118 -0.424 -0.257  0.173
O1 -0.159 -0.149  0.185 -0.309         0.289         0.313 -0.184
O2               -0.281  0.337         0.271 -0.394         0.271
O3 -0.192 -0.200  0.133 -0.348         0.135  0.104 -0.129 -0.150
O4        -0.205  0.176 -0.210 -0.332  0.207 -0.393         0.348  0.490
O5               -0.252  0.373  0.121  0.290                -0.336 -0.145
   Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18 Comp.19
A1  0.231   0.196   0.393                   0.230   0.126   0.198   0.271
A2          0.134   0.296                   0.253           0.279   0.304
A3  0.315                                  -0.137           0.126   0.386
A4 -0.328  -0.126   0.271          -0.172          -0.124  -0.298
A5  0.432                           0.255           0.245          -0.437
C1  0.180  -0.128          -0.134   0.217          -0.385  -0.258   0.226
C2 -0.261          -0.105   0.113  -0.244           0.399   0.199  -0.137
C3  0.131   0.157   0.357          -0.297  -0.241  -0.101  -0.182  -0.146
C4                  0.197   0.150                  -0.428  -0.118  -0.132
C5          0.107   0.363          -0.215  -0.130   0.335  -0.142
E1  0.246          -0.241                   0.126  -0.275   0.186  -0.290
E2          0.102           0.251   0.308  -0.136   0.194  -0.313
E3                 -0.380          -0.112   0.152   0.130  -0.538   0.218
E4         -0.147          -0.142          -0.119  -0.113   0.218  -0.290
E5 -0.160                   0.209           0.528          -0.138  -0.288
N1                          0.120          -0.230
N2  0.130                   0.118   0.115  -0.263
```

```
N3  0.149  -0.112          -0.285  -0.160          -0.105
N4                         -0.336  -0.359   0.333
N5 -0.284   0.203          0.135   0.455   0.215
O1 -0.357                  -0.542   0.330  -0.235
O2 -0.209   0.437  -0.359  -0.153          -0.144  -0.108
O3 -0.131          -0.107   0.472  -0.217  -0.245  -0.257   0.282
O4          -0.387          0.105
O5 -0.118  -0.636          0.130                   0.150           0.174
   Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25
A1
A2 -0.111   0.451  -0.230          0.152
A3 -0.144  -0.552   0.194         -0.255
A4  0.158           0.150
A5  0.363          -0.159          0.360
C1                          0.115   0.118
C2 -0.329  -0.116  -0.332          0.107
C3
C4 -0.295  -0.296  -0.441          0.126  -0.103
C5          0.149   0.409  -0.146
E1 -0.236   0.219   0.306
E2  0.196   0.155  -0.337         -0.495
E3 -0.287   0.296           0.121
E4 -0.206   0.276                 -0.610
E5  0.151  -0.210          -0.154  -0.230
N1                          0.224   0.128   0.704
N2          0.161           0.198          -0.668
N3                 -0.195  -0.735
N4  0.306                   0.512  -0.174
N5 -0.133           0.327
O1
O2  0.218
O3  0.395   0.136
O4
O5  0.156

                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
SS loadings       1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00
Proportion Var    0.04   0.04   0.04   0.04   0.04   0.04   0.04   0.04   0.04
Cumulative Var    0.04   0.08   0.12   0.16   0.20   0.24   0.28   0.32   0.36
                Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17
SS loadings        1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
Proportion Var     0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
Cumulative Var     0.40    0.44    0.48    0.52    0.56    0.60    0.64    0.68
                Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25
SS loadings        1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
Proportion Var     0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
Cumulative Var     0.72    0.76    0.80    0.84    0.88    0.92    0.96    1.00
```

```
> pca2.bfi<-principal(X2,nfactors=25)  #library(psych)

> sqrt(pca2.bfi$values)
 [1] 2.2659018 1.6588811 1.4637971 1.3610024 1.2442519 1.0361382 0.9162636
 [8] 0.8939833 0.8479323 0.8295112 0.8224192 0.8073412 0.7894637 0.7723748
[15] 0.7503938 0.7370925 0.7172988 0.7032092 0.6947226 0.6700157 0.6506659
[22] 0.6329861 0.6227395 0.6179456 0.5123856

> pca2.bfi$loadings

Loadings:
   RC2    RC6    RC8    RC7    RC5    RC9    RC10   RC14   RC4    RC3    RC12
A1                            0.973
A2                     0.148 -0.177
A3                     0.165 -0.128
A4                     0.948
A5                     0.125
C1                                   0.131                       0.943
C2                                   0.157                       0.197
C3                                   0.954                       0.127
C4                                  -0.155                      -0.154
C5                    -0.101        -0.154
E1          0.941
E2          0.219  0.111
E3         -0.126                                    0.141
E4         -0.187         0.128
E5         -0.123                                    0.120         0.105
N1  0.414          0.163
N2  0.899          0.137
N3  0.247          0.189
N4  0.148  0.101   0.175                       0.102
N5  0.132          0.941
O1                                                   0.951               -0.101
O2                                                          0.965         0.146
O3                                                   0.170 -0.133        -0.139
O4                                                 0.974
O5                                                          0.148         0.962
   RC23   RC13   RC17   RC16   RC11   RC19   RC18   RC20   RC15   RC21   RC22
A1                                   -0.145                      -0.100
A2                     0.111  0.140  0.913                       0.200
A3                            0.209  0.214  0.142                0.894
A4                            0.107  0.133                       0.139
A5                            0.897  0.147  0.154                0.207
C1                0.188                                                -0.135
C2         -0.110  0.929                                                -0.160
```

```
C3        -0.136  0.145                                                  -0.132
C4         0.209 -0.173                                                   0.905
C5         0.916 -0.115                                  0.128            0.202
E1                       -0.112               -0.111
E2                       -0.147 -0.103        -0.138         0.116
E3                        0.145  0.152         0.903  0.167         0.138
E4                        0.105  0.192         0.154        -0.108  0.133
E5                        0.915         0.111  0.141  0.109
N1  0.249                                                   0.150
N2  0.224                                                   0.132
N3  0.892                                                   0.218
N4  0.221  0.139                                            0.891
N5  0.163                                                   0.148
O1                        0.106               0.121  0.154
O2                                                -0.116
O3                        0.106               0.157  0.922
O4
O5                                                -0.123
   RC24   RC1    RC25
A1
A2
A3  0.125
A4  0.102
A5  0.178
C1
C2
C3
C4
C5
E1 -0.150  0.176
E2 -0.205  0.879
E3  0.141 -0.128
E4  0.876 -0.208
E5        -0.131
N1                0.825
N2                0.263
N3                0.191
N4 -0.104  0.112  0.120
N5                0.111
O1
O2
O3
O4
O5
```

```
> biplot.psych(pca2.bfi,choose=c(1,2,3),pch=".",main=""){
# Should do at least "choose=c(1:5)" (and probably even upto
"choose=c(1:7)") but doing "choose=c(1:3)" for graphical clarity
```