Exploratory Factor Analysis of the Diversity Data*

Data include answers to

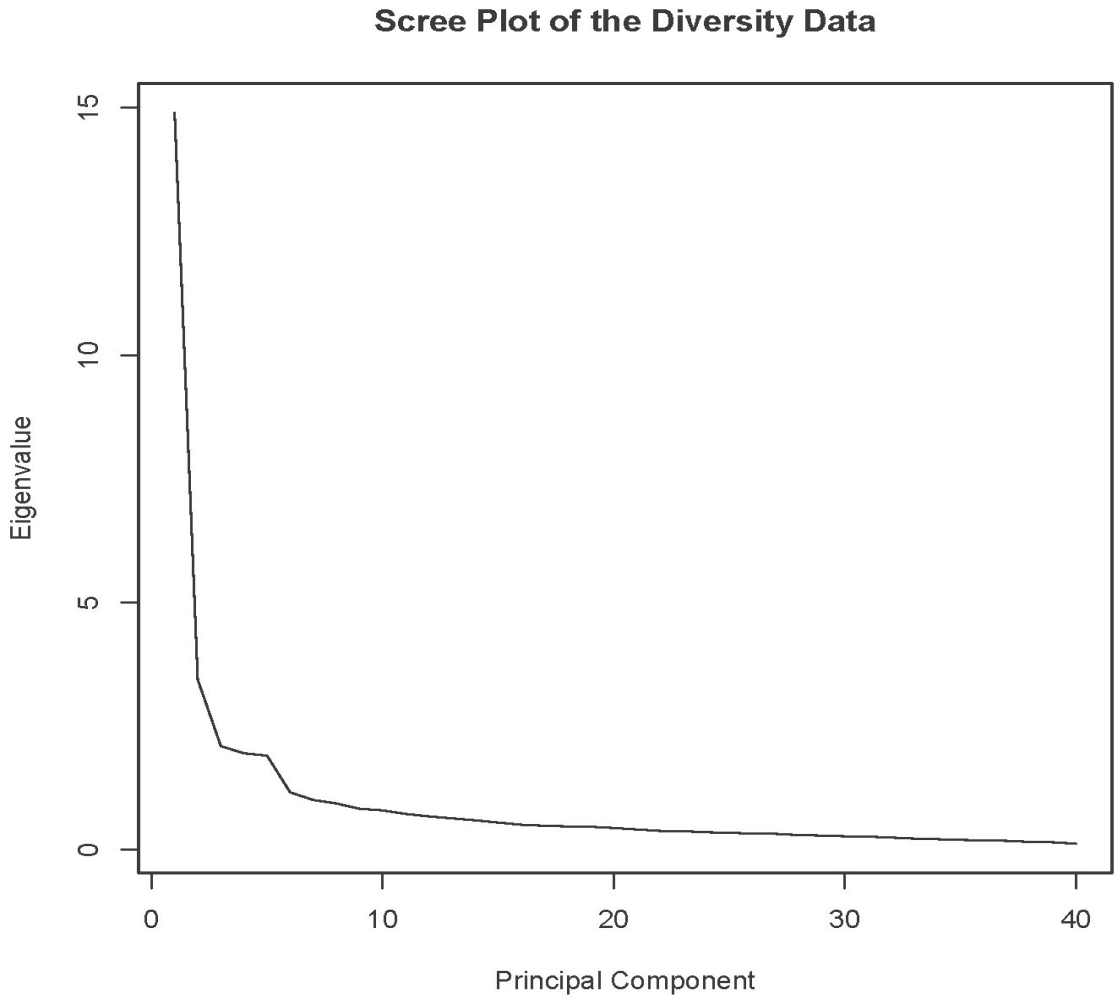
- Ten questions about commitment to the organization, with higher numbers indicating more commitment.
- Five questions about relations with colleagues at work, with higher numbers indicating better relations.
- Twelve questions about relations with management, in particular the respondent's immediate boss. Higher numbers indicate better relations.
- Six questions about fair opportunities for advancement, with higher numbers indicating more fairness.
- Four questions about job satisfaction, with higher numbers indicating more satisfaction.
- Three questions about senior management's commitment to diversity, with higher numbers indicating more commitment. These seem to be on a six-point scale instead of five.

```
> # Exploratory Factor Analysis of the Diversity data
> rm(list=ls())
> # install.packages("readxl", dependencies = TRUE) # Only need to do this once
> library(readxl)
> # Download the data and put it in your working directory.
> # https://www.utstat.toronto.edu/brunner/data/legal/DiversityExplore.xlsx
> # The replication data are in DiversityReplic.xlsx
> ddata = read_excel("DiversityExplore.xlsx") # Read local copy

> ddata = as.data.frame(ddata) # Instead of a "tibble"
> dim(ddata); # head(ddata)
[1] 500  47

> quest = as.matrix(ddata[,2:41]); head(quest)

     Com1 Com2 Com3 Com4 Com5 Com6 Com7 Com8 Com9 Com10 RelC1 RelC2 RelC3 RelC4 RelC5
[1,]    4    4    5    3    4    2    3    3    2     3     4     4     4     2     4
[2,]    5    5    5    5    5    4    5    5    4     5     5     5     5     5     5
[3,]    5    5    5    5    5    5    5    5    5     5     5     5     5     5     5
[4,]    4    4    4    4    4    4    4    5    2     4     3     2     2     2     1
[5,]    5    4    4    3    3    3    5    5    4     4     3     4     4     4     5
[6,]    2    4    4    4    2    1    3    4    3     4     5     5     5     4     5
     RelM1 RelM2 RelM3 RelM4 RelM5 RelM6 RelM7 RelM8 RelM9 RelM10 RelM11 RelM12 Fair1
[1,]     2     5     4     5     5     5     5     4     4      4      5      5     4
[2,]     3     5     5     5     5     5     3     4     4      5      5      5     5
[3,]     4     5     3     5     3     2     3     5     5      5      5      5     5
[4,]     1     2     2     3     3     1     1     2     3      3      3      3     3
[5,]     3     4     4     4     4     4     3     3     5      5      5      5     4
[6,]     3     3     3     3     4     2     3     3     3      2      4      4     4
     Fair2 Fair3 Fair4 Fair5 Fair6 Sat1 Sat2 Sat3 Sat4 SM1 SM2 SM3
[1,]     1     4     1     2     1     2     2     3    4   3   1   2
[2,]     4     5     5     5     5     5     4     5    5   4   5   5
[3,]     2     5     1     5     2     5     4     4    1   6   6   3
[4,]     3     3     1     4     3     2     3     3    3   6   6   6
[5,]     5     4     2     3     2     4     5     5    5   3   3   3
[6,]     4     4     3     2     3     4     4     4    4   4   3   3
```

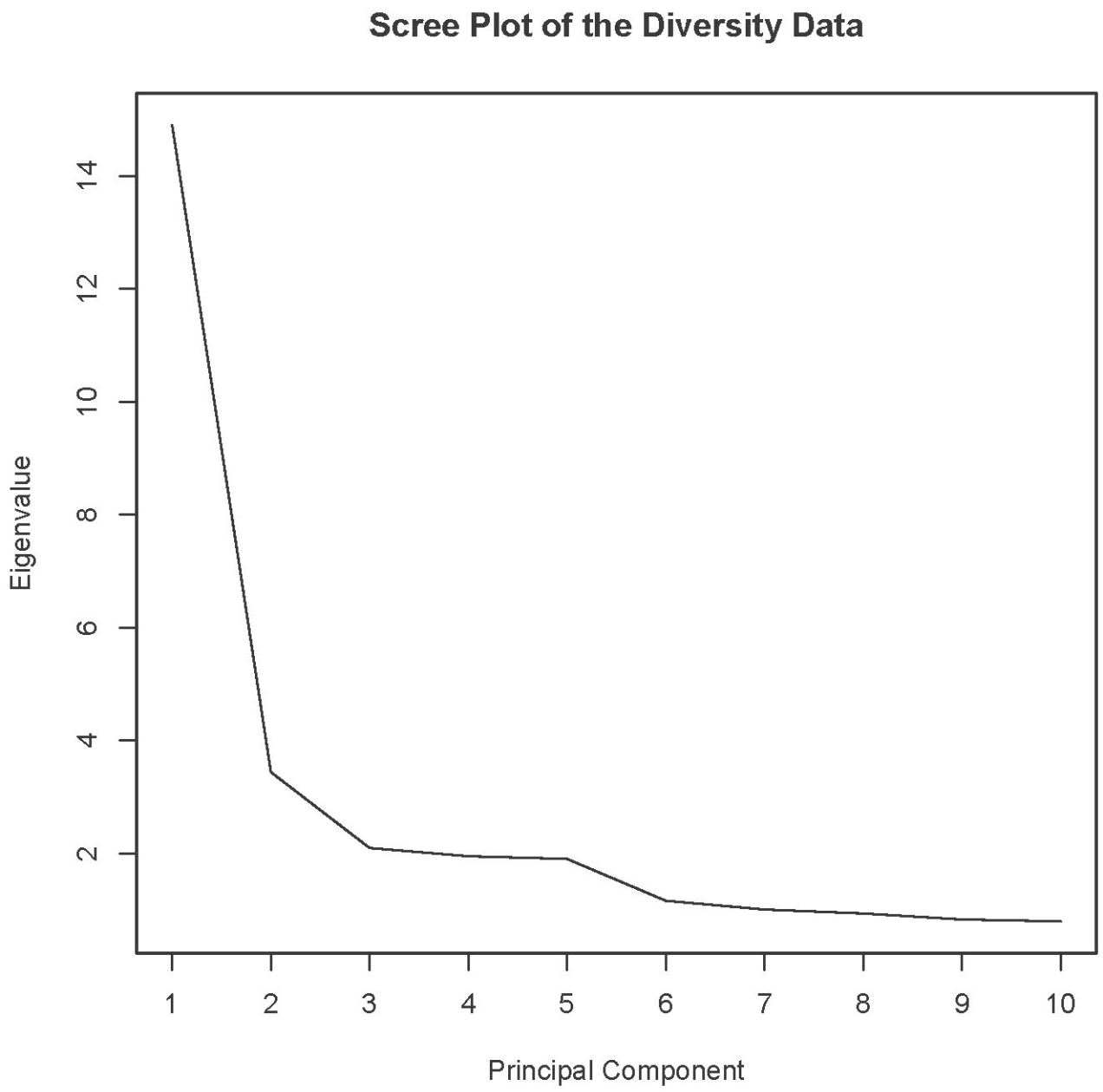
*This handout was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The OpenOffice.org document is available from the course website:

http://www.utstat.toronto.edu/brunner/oldclass/431s23

```
> # Check eigenvalues and scree plot
> pc = prcomp(quest, scale = T)
> # ls(pc) # What's in pc?
>     # center has the sample means before standardization
>     # rotation is C-hat
>     # scale has the standard deviations before standardization
>     # sdev has estimated standard deviations of the components
>     # x is a matrix of the principal components: Y-hat
>
> Eigenvalue = pc$sdev^2; Eigenvalue
 [1] 14.9024012  3.4419990  2.0981709  1.9522247  1.9047772  1.1634482  1.0084527
 [8]  0.9387470  0.8322382  0.7987125  0.7269172  0.6780117  0.6376820  0.5990323
[15]  0.5528033  0.5087344  0.4880584  0.4747871  0.4679052  0.4437166  0.4131693
[22]  0.3845121  0.3784929  0.3567195  0.3457392  0.3305095  0.3247698  0.2996803
[29]  0.2900113  0.2752323  0.2681413  0.2498753  0.2285067  0.2161325  0.2016001
[36]  0.1932987  0.1838430  0.1618372  0.1535005  0.1256088

> # Seven eigenvalues greater than one

> # Scree plot
> plot(1:40,Eigenvalue, xlab = "Principal Component", type = "l",
+      main = "Scree Plot of the Diversity Data")
```



```
> # Zoom in on the first ten
> plot(1:10,Eigenvalue[1:10], type = "l",
+      xaxp = c(1,10,9), # Tick marks on x axis
+      xlab = "Principal Component", ylab = "Eigenvalue",
+       main = "Scree Plot of the Diversity Data")
```




```
> # Scree plot suggests 6 factors.
```

```
> # help(factanal)
> fa1 = factanal(quest,factors=6) # rotation="varimax" is the default
> fa1
Call:
factanal(x = quest, factors = 6)

Uniquenesses:
  Com1   Com2   Com3   Com4   Com5   Com6   Com7   Com8   Com9  Com10  RelC1  RelC2
 0.342  0.602  0.705  0.378  0.253  0.306  0.358  0.507  0.727  0.772  0.259  0.207
 RelC3  RelC4  RelC5  RelM1  RelM2  RelM3  RelM4  RelM5  RelM6  RelM7  RelM8  RelM9
 0.447  0.530  0.626  0.447  0.344  0.309  0.275  0.438  0.347  0.407  0.438  0.283
RelM10 RelM11 RelM12  Fair1  Fair2  Fair3  Fair4  Fair5  Fair6   Sat1   Sat2   Sat3
 0.408  0.347  0.331  0.440  0.444  0.541  0.752  0.551  0.329  0.541  0.210  0.121
  Sat4    SM1    SM2    SM3
 0.552  0.192  0.328  0.516

Loadings:
       Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
Com1     0.186   0.766           0.105   0.127
Com2     0.103   0.571   0.188   0.105   0.105
Com3     0.130   0.506
Com4     0.240   0.714           0.160           0.111
Com5     0.181   0.816                   0.185
Com6     0.189   0.779           0.140   0.166
Com7     0.174   0.755           0.143   0.119
Com8     0.156   0.660   0.109
Com9             0.434   0.164           0.164   0.147
Com10            0.452                           0.103
RelC1    0.159   0.185   0.811   0.101           0.102
RelC2    0.206   0.175   0.834   0.121
RelC3    0.294   0.233   0.605   0.142   0.161
RelC4    0.285   0.151   0.545           0.240
RelC5    0.321   0.148   0.446   0.129   0.176
RelM1    0.656   0.122                   0.289
RelM2    0.737   0.213   0.134           0.183
RelM3    0.783   0.173   0.135   0.114   0.123
RelM4    0.809   0.188   0.122           0.119
RelM5    0.679   0.202   0.134   0.101           0.151
RelM6    0.739   0.116           0.155   0.238
RelM7    0.683   0.120   0.104   0.203   0.237
RelM8    0.684   0.169   0.118   0.207
RelM9    0.755   0.195   0.212   0.212           0.126
RelM10   0.701   0.119   0.157   0.158   0.183
RelM11   0.720   0.217   0.276
RelM12   0.748   0.187   0.194   0.172
Fair1    0.420   0.310   0.183   0.269   0.396   0.159
Fair2    0.312   0.314   0.132   0.268   0.505   0.123
Fair3    0.328   0.224   0.276   0.199   0.415   0.111
Fair4    0.162   0.135   0.144   0.237   0.356
Fair5    0.297   0.143   0.105   0.183   0.541
Fair6    0.358   0.285   0.164   0.272   0.594
Sat1     0.240   0.243   0.112   0.546   0.175
Sat2     0.226   0.207   0.167   0.781   0.218   0.104
Sat3     0.277   0.160   0.121   0.849   0.184
Sat4     0.272   0.183   0.143   0.520   0.178   0.136
SM1      0.136   0.161                   0.109   0.865
SM2      0.173   0.174           0.107   0.143   0.757
SM3              0.113                           0.669

               Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings      7.853   5.463   2.906   2.712   2.115   2.042
Proportion Var   0.196   0.137   0.073   0.068   0.053   0.051
Cumulative Var   0.196   0.333   0.406   0.473   0.526   0.577

Test of the hypothesis that 6 factors are sufficient.
The chi square statistic is 1506.73 on 555 degrees of freedom.
The p-value is 6.94e-89
```

```
> # Loadings below 0.1 are blanked out by default
> L1 = fa1$loadings; print(L1,cutoff=0.3, sort="True")

Loadings:
       Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
RelM1   0.656
RelM2   0.737
RelM3   0.783
RelM4   0.809
RelM5   0.679
RelM6   0.739
RelM7   0.683
RelM8   0.684
RelM9   0.755
RelM10  0.701
RelM11  0.720
RelM12  0.748
Com1            0.766
Com2            0.571
Com3            0.506
Com4            0.714
Com5            0.816
Com6            0.779
Com7            0.755
Com8            0.660
RelC1                   0.811
RelC2                   0.834
RelC3                   0.605
RelC4                   0.545
Sat1                            0.546
Sat2                            0.781
Sat3                            0.849
Sat4                            0.520
Fair2   0.312   0.314                    0.505
Fair5                                    0.541
Fair6   0.358                            0.594
SM1                                              0.865
SM2                                              0.757
SM3                                              0.669
Com9            0.434
Com10           0.452
RelC5   0.321           0.446
Fair1   0.420   0.310                    0.396
Fair3   0.328                            0.415
Fair4                                    0.356

               Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings      7.853   5.463   2.906   2.712   2.115   2.042
Proportion Var   0.196   0.137   0.073   0.068   0.053   0.051
Cumulative Var   0.196   0.333   0.406   0.473   0.526   0.577
> # Check SS loadings = 7.853 for Factor One
> sum(L1[,1]^2)
[1] 7.852818

> # This is such a pretty picture. What if no rotation?
```

```
> fa2 = factanal(quest,factors=6, rotation="none"); L2 = fa2$loadings
> print(L2,cutoff=0.3, sort="True")

Loadings:
       Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
Com1    0.580   0.464
Com4    0.620   0.410
Com5    0.606   0.507
Com6    0.600   0.486
Com7    0.578   0.467
Com8    0.511   0.408
RelC3   0.595                           0.412
RelC4   0.542                           0.349
RelC5   0.527
RelM1   0.647
RelM2   0.725
RelM3   0.718  -0.354
RelM4   0.722  -0.362
RelM5   0.673
RelM6   0.704  -0.345
RelM7   0.690  -0.305
RelM8   0.673
RelM9   0.773
RelM10  0.686  -0.328
RelM11  0.698
RelM12  0.731  -0.314
Fair1   0.719
Fair2   0.658                                   0.312
Fair3   0.619
Fair5   0.526                                   0.398
Fair6   0.705                                   0.394
Sat1    0.570                  -0.336
Sat2    0.678                  -0.556
Sat3    0.688                  -0.621
Sat4    0.583                  -0.321
SM1     0.394           0.650          -0.434
SM2     0.448           0.562          -0.357
SM3                     0.531
RelC1   0.519           0.344           0.586
RelC2   0.555           0.337           0.602
Com2    0.468   0.375
Com3    0.384   0.305
Com9    0.397   0.300
Com10   0.311   0.304
Fair4   0.406

               Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings     14.427   2.984   1.657   1.628   1.608   0.788
Proportion Var   0.361   0.075   0.041   0.041   0.040   0.020
Cumulative Var   0.361   0.435   0.477   0.517   0.558   0.577
>
> # Varimax is much nicer.
```

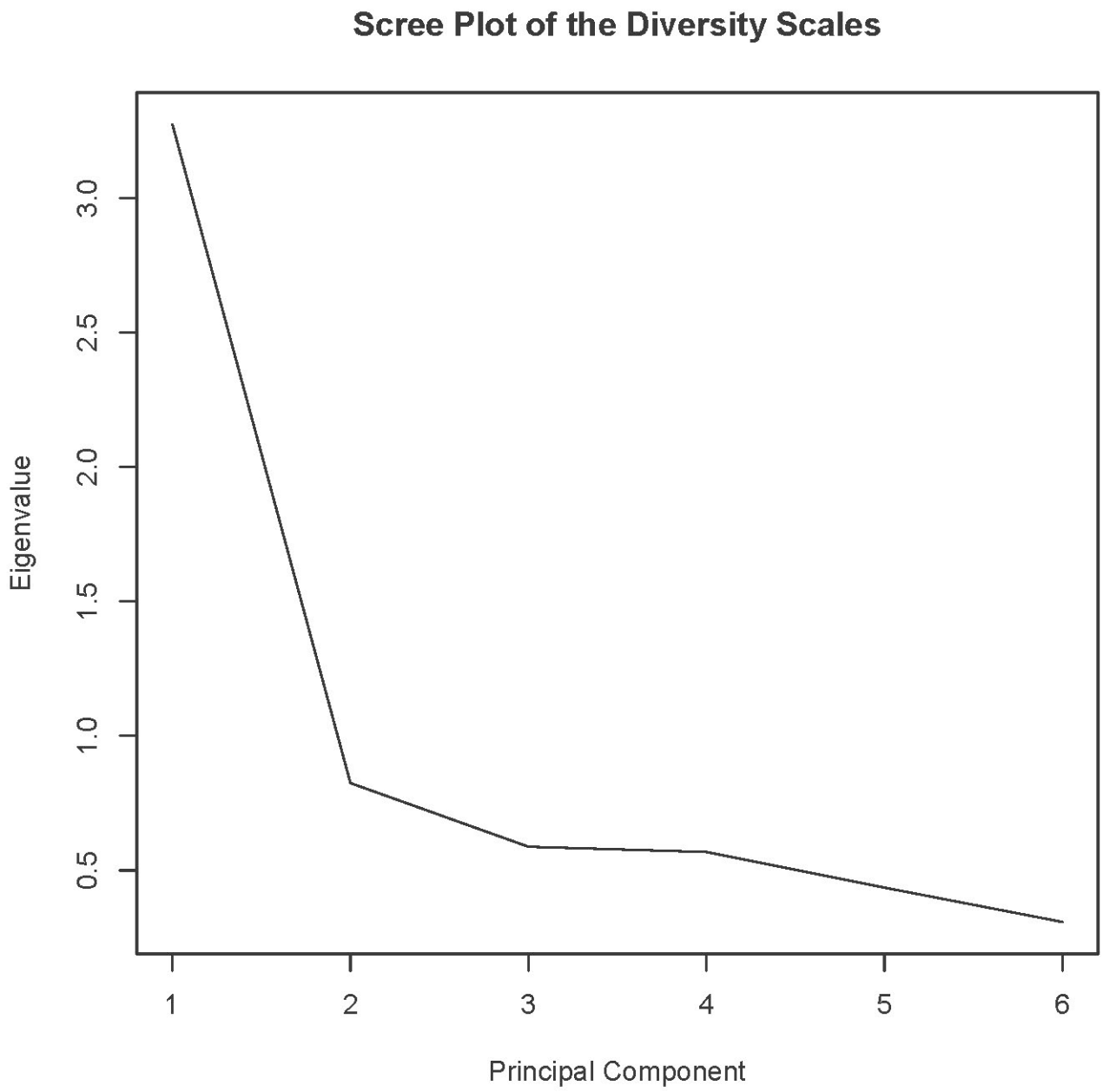
```
> # It's common to specify a rotation (or accept the default) in the initial fit, but one
can also fit a model without rotation, and then rotate the factors as a separate step.
>
> varimax(L2)
$loadings

Loadings:
       Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
Com1     0.186   0.766           -0.105           0.127
Com2     0.103   0.571           -0.105   0.188   0.105
Com3     0.130   0.506
Com4     0.240   0.714   0.111   -0.160
Com5     0.181   0.816                            0.185
Com6     0.189   0.779           -0.140           0.166
Com7     0.174   0.755           -0.143           0.119
Com8     0.156   0.660                   0.109
Com9             0.434   0.147           0.164   0.164
Com10            0.452   0.103
RelC1    0.159   0.185   0.102   -0.101   0.811
RelC2    0.206   0.175           -0.121   0.834
RelC3    0.294   0.233           -0.142   0.605   0.161
RelC4    0.285   0.151                    0.545   0.240
RelC5    0.321   0.148           -0.129   0.446   0.176
RelM1    0.656   0.122                            0.289
RelM2    0.737   0.213                    0.134   0.183
RelM3    0.783   0.173           -0.114   0.135   0.123
RelM4    0.809   0.188                    0.122   0.119
RelM5    0.679   0.202   0.151   -0.101   0.134
RelM6    0.739   0.116           -0.155           0.238
RelM7    0.683   0.120           -0.203   0.104   0.237
RelM8    0.684   0.169           -0.207   0.118
RelM9    0.755   0.195   0.126   -0.212   0.212
RelM10   0.701   0.119           -0.158   0.157   0.183
RelM11   0.720   0.217                    0.276
RelM12   0.748   0.187           -0.172   0.194
Fair1    0.420   0.310   0.159   -0.269   0.183   0.396
Fair2    0.312   0.314   0.123   -0.268   0.132   0.505
Fair3    0.328   0.224   0.111   -0.199   0.276   0.415
Fair4    0.162   0.135           -0.237   0.144   0.356
Fair5    0.297   0.143           -0.183   0.105   0.541
Fair6    0.358   0.285           -0.272   0.164   0.594
Sat1     0.240   0.243           -0.546   0.112   0.175
Sat2     0.226   0.207   0.104   -0.781   0.167   0.218
Sat3     0.277   0.160           -0.849   0.121   0.184
Sat4     0.272   0.183   0.136   -0.520   0.143   0.178
SM1      0.136   0.161   0.865                    0.109
SM2      0.173   0.174   0.757   -0.107           0.143
SM3              0.113   0.669

                Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings       7.853   5.463   2.042   2.712   2.906   2.115
Proportion Var    0.196   0.137   0.051   0.068   0.073   0.053
Cumulative Var    0.196   0.333   0.384   0.452   0.524   0.577
```

Continued on next page ...

```
$rotmat
           [,1]        [,2]        [,3]         [,4]        [,5]        [,6]
[1,]  0.6530884  0.45458701  0.19404997 -0.385736238  0.31744582  0.28214626
[2,] -0.6433623  0.73250417  0.17215414 -0.121446030 -0.03415892  0.06300467
[3,] -0.1701312 -0.30272251  0.82282254  0.093713442  0.43733619 -0.04829448
[4,]  0.2579775  0.37841673  0.03807809  0.872710531  0.08827413 -0.13922091
[5,] -0.2107186 -0.02240009 -0.50031317 -0.002964388  0.83599693 -0.07670065
[6,] -0.1402183 -0.14644587 -0.06250933  0.256989782  0.01073220  0.94277574

> fa1$rotmat
           [,1]        [,2]         [,3]        [,4]        [,5]        [,6]
[1,]  0.6530884  0.45458701  0.385736238 -0.19404997  0.31744582  0.28214626
[2,] -0.6433623  0.73250417  0.121446030 -0.17215414 -0.03415892  0.06300467
[3,] -0.2579775 -0.37841673  0.872710531  0.03807809 -0.08827413  0.13922091
[4,]  0.1701312  0.30272251  0.093713442  0.82282254 -0.43733619  0.04829448
[5,] -0.2107186 -0.02240009  0.002964388  0.50031317  0.83599693 -0.07670065
[6,] -0.1402183 -0.14644587 -0.256989782  0.06250933  0.01073220  0.94277574
>
> varimaxL2 = varimax(L2); print(varimaxL2$loadings, cutoff=0.3, sort="True")

Loadings:
       Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
RelM1   0.656
RelM2   0.737
RelM3   0.783
RelM4   0.809
RelM5   0.679
RelM6   0.739
RelM7   0.683
RelM8   0.684
RelM9   0.755
RelM10  0.701
RelM11  0.720
RelM12  0.748
Com1            0.766
Com2            0.571
Com3            0.506
Com4            0.714
Com5            0.816
Com6            0.779
Com7            0.755
Com8            0.660
SM1                     0.865
SM2                     0.757
SM3                     0.669
Sat1                            -0.546
Sat2                            -0.781
Sat3                            -0.849
Sat4                            -0.520
RelC1                                   0.811
RelC2                                   0.834
RelC3                                   0.605
RelC4                                   0.545
Fair2   0.312   0.314                           0.505
Fair5                                           0.541
Fair6   0.358                                   0.594
Com9            0.434
Com10           0.452
RelC5   0.321                                   0.446
Fair1   0.420   0.310                           0.396
Fair3   0.328                                   0.415
Fair4                                           0.356

                Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings       7.853   5.463   2.042   2.712   2.906   2.115
Proportion Var    0.196   0.137   0.051   0.068   0.073   0.053
Cumulative Var    0.196   0.333   0.384   0.452   0.524   0.577
```

```
> # Scales
>
> # I am thinking that the a priori scales are pretty good. In fact, they are amazing.
>     # Are they really uncorrelated?
>     # What factor(s) underlie them?
>
> # head(quest)
>
> Com = apply(quest[,1:10],MARGIN=1, FUN=sum) # Commitment to Organization
> RelC = apply(quest[,11:15],MARGIN=1, FUN=sum) # Relations with Colleagues
> RelM = apply(quest[,16:27],MARGIN=1, FUN=sum) # Relations with Management
> Fair = apply(quest[,28:33],MARGIN=1, FUN=sum) # Fair opportunities for advancement
> Sat = apply(quest[,34:37],MARGIN=1, FUN=sum) # Job satisfaction
> SMdiv = apply(quest[,38:40],MARGIN=1, FUN=sum) # Sr. Man commitment to diversity
>
> # Checking
> # head(cbind(quest[,1:10],Com))
> # head(cbind(quest[,11:15],RelC))
> # head(cbind(quest[,16:27],RelM))
> # head(cbind(quest[,28:33],Fair))
> # head(cbind(quest[,34:37],Sat))
> # head(cbind(quest[,38:40],SM))
>
> scaledat = cbind(Com, RelC, RelM, Fair, Sat, SMdiv)
> head(scaledat); dim(scaledat)
     Com RelC RelM Fair Sat SMdiv
[1,]  33   18   53   13  11     6
[2,]  48   25   54   29  19    14
[3,]  50   25   50   20  14    15
[4,]  39   10   27   17  11    18
[5,]  40   20   49   20  19     9
[6,]  31   24   37   20  16    10
[1] 500   6

> corrmat = cor(scaledat); corrmat
            Com       RelC       RelM       Fair       Sat      SMdiv
Com   1.0000000 0.4242217 0.4566834 0.5448548 0.4491691 0.3054170
RelC  0.4242217 1.0000000 0.5368269 0.5203092 0.4414932 0.2497628
RelM  0.4566834 0.5368269 1.0000000 0.6510840 0.5348077 0.2961944
Fair  0.5448548 0.5203092 0.6510840 1.0000000 0.6379985 0.3119867
Sat   0.4491691 0.4414932 0.5348077 0.6379985 1.0000000 0.2799530
SMdiv 0.3054170 0.2497628 0.2961944 0.3119867 0.2799530 1.0000000
> # Under H0: rho = 0, t = r * sqrt(n-2) / sqrt(1-r^2) ~ t(n-2)
> critval = qt(0.975,498); critval
[1] 1.964739
> # Calculate all the t statistics
> n = 500; tmat = corrmat * sqrt(n-2) / sqrt(1-corrmat^2)
> round(tmat,3)
         Com   RelC   RelM   Fair    Sat SMdiv
Com      Inf 10.454 11.456 14.500 11.219 7.158
RelC  10.454    Inf 14.199 13.597 10.980 5.756
RelM  11.456 14.199    Inf 19.143 14.124 6.920
Fair  14.500 13.597 19.143    Inf 18.489 7.328
Sat   11.219 10.980 14.124 18.489    Inf 6.508
SMdiv  7.158  5.756  6.920  7.328  6.508   Inf
```

```
> # Factor analysis: How many factors?
>
> pc2 = prcomp(scaledat, scale = T)
> Eigenvalue = pc2$sdev^2; Eigenvalue
[1] 3.2740205 0.8249742 0.5880816 0.5685478 0.4360184 0.3083576
>
> # Scree plot
> plot(1:6,Eigenvalue, xlab = "Principal Component", type = "l",
+      main = "Scree Plot of the Diversity Scales")
```




```
> # Either one factor or two
```

```
> # Either one factor or two
>
> one = factanal(scaledat,factors=1); L.1 = one$loadings
> print(one,cutoff=0.3, sort="True")

Call:
factanal(x = scaledat, factors = 1)

Uniquenesses:
  Com  RelC  RelM  Fair   Sat SMdiv
0.599 0.591 0.418 0.272 0.476 0.849

Loadings:
[1] 0.633 0.639 0.763 0.853 0.724 0.389

               Factor1
SS loadings      2.794
Proportion Var   0.466

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 16.42 on 9 degrees of freedom.
The p-value is 0.0586

>
> two = factanal(scaledat,factors=2)
> print(two,cutoff=0.3, sort="True")

Call:
factanal(x = scaledat, factors = 2)

Uniquenesses:
  Com  RelC  RelM  Fair   Sat SMdiv
0.607 0.005 0.438 0.235 0.472 0.854

Loadings:
      Factor1 Factor2
Com   0.576
RelM  0.671   0.335
Fair  0.835
Sat   0.690
RelC  0.330   0.941
SMdiv 0.356

               Factor1 Factor2
SS loadings      2.191   1.199
Proportion Var   0.365   0.200
Cumulative Var   0.365   0.565

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6 on 4 degrees of freedom.
The p-value is 0.199
```